Springer Series in Reliability Engineering

Dimitris N. Chorafas

Quality Control Applications

Springer

Springer Series in Reliability Engineering

Series Editor

Hoang Pham

For further volumes:
http://www.springer.com/series/6917

Dimitris N. Chorafas

Quality Control Applications

Springer

Dimitris N. Chorafas

ISSN 1614-7839
ISBN 978-1-4471-2965-3 ISBN 978-1-4471-2966-0 (eBook)
DOI 10.1007/978-1-4471-2966-0
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2012936761

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Part II Service Quality

Part III Product Reliability

Part IV Statistical Inference

Part V Statistical Quality Control

Introduction

A house is only as strong as its foundation, and the same applies to all man-made devices and systems. The foundation of any product, as well as of workmanship, is quality. Quality is, however, a moving target and simple inspection tools do not suffice for thousands of new and of revamped elder products offered to the market every year.

The need for high quality, reliable products, and systems in industry and the military has set the stage for a great expansion in the field of quality control and reliability. It also brought forward the requirement for a much better understanding, by engineers and projects managers, of the concepts underlying quality, its tools, and the enlarged domain of their implementation.

When it comes to quality control, many good books are available on mathematics written for mathematicians. The gap that this book intends to fill is that of practical applications. The mathematician's approach is absolute rigor and purity. The engineer, on the other hand, has to do compromises. The most perfectly engineered device is an optimum of conditions which, at times, may require less in terms of theories but more of an iron discipline on quality.

By means of practical examples and case studies the book offers insight into how a "high quality/controlled cost" strategy works, and what it takes to put it in place and sustain it. The case of Japan's Fukushima Daiichi nuclear plant had much more to do with disregard of quality control and reliability principles than with the risks associated to nuclear power production.

The same is true of the Macondo incident in the Gulf of Mexico which underscored that safe operations require rigorous quality control systems and procedures. Process quality assurance is fundamental to our ability to open new frontiers in engineering—deep water drilling being an example.

The electronics industry, too, has plenty of quality control problems particularly connected to manufacturing by western firms in developing countries. Quality control is a high riding subject also in mechanical engineering, as exemplified by quality failures such as the massive car recalls by Toyota and other vehicle vendors.

Quality challenges have as well confronted administrative offices and logistics operations. An example is operational risk failures in the course of office work.

The practical examples on quality control applications included in this book can be divided into two main classes. Those relating to quality in the large which offer the broad view of what quality assurance is all about; and those coming directly from the production floor which, though highly important, could be looked at as quality in the small.

Quality in the large starts at the drafting board and its presence dominates throughout manufacturing engineering and field maintenance. Its concepts parallel those of reliability engineering, with experimental design at the core of the product assurance effort.

Based on these notions, Part I concentrates on product quality in the large. Chapter 1 defines the sense of product assurance; Chap. 2 presents the broader concepts underpinning quality control; Chap. 3 explains what is meant by designing for better quality at an affordable cost.

We live in a service economy and this has implications in service assurance. It is absolutely unacceptable that the quality of services is wanting and the results obtained from man-made systems look like the roll of the dice. What is meant by wanting service quality is explained in Part II through two case studies. They both focus on quality in the large in the domain of service assurance.

The theme of Chap. 4 is a detailed step-by-step explanation of quality failures in an alternative energy project due to "I can not care less" policies and an unprecedented imprudence. The dramatic failures in quality assurance and reliability in the Fukushima Daiichi nuclear plants are the background of the case study in Chap. 5 on man-made catastrophes.

Product reliability, Part III subject, is also a theme in the realm of quality in the large. Chapter. 6 first explains, then demonstrates the similarities and differences between quality and reliability assurance. Chapter. 7 documents why reliability assurance is inseparable from lifecycle maintainability.

Though the book's contents have been deliberately tuned to applications, not theories, there is a need for a basic background on quality control and reliability assurance. The domain which they share is statistical inference. To this subject is dedicated Part IV. In addition, stochastic thinking provides the conceptual bridge between quality in the large and quality in the small.

Chapter 8 offers the reader an introduction to stochastic thinking; Chap. 9 explains sampling methods, and principles underpinning them; Chap. 10 concentrates on the decision power of operating characteristics curves; Chap. 11 focuses on experimental design and Latin squares.

The four chapters of Part V concentrate on statistical control at the production floor—hence, quality in the small. Chapter 12 explains the fundamental concept of statistical quality inspection; Chap. 13 offers practical examples on quality control charts by variables; Chap. 14 explains the relative advantages of statistical quality

control by attributes and its applications. Chapter 15 concludes this book by presenting the reader with practical examples of statistical quality control charts for small- and medium-sized enterprises, like lot templates. It also brings to the reader's attention the cultural prerequisites necessary for an implementation of successful statistical quality control—from cultural issues to open communications channels.

Throughout its 15 chapters, the book provides the reader with both principles of and practical examples on quality control applications. Particular attention has been paid to enlarging the perspective of quality and reliability and on demonstrating how quality control solutions can be customized. Cookie-cutter approaches which are often applied to quality problems in business and industry will serve neither quality in the large nor quality in the small.

Both the larger picture of a design for quality control and attention to detail are a premium. Do not listen to what is often said that paying attention to detail means one does not have the brains to look at the bigger picture. Without detail we do not know, much less understand, all the aspects of the quality control system we are working with.

This is counterproductive. If you know yourself and know your opponent you do not need to worry about the outcome of 100 battles, said 2500 years ago by Sun Tzu, the great Chinese statesman and general. Do we really know ourselves and what we can deliver?

I am indebted to a long list of knowledgeable people, and of organizations, for their contribution to the research which made this book feasible. Also to several senior executives and experts for constructive criticism during the preparation of the manuscript. Most particularly Dr. Heinrich Steinmann, Dr. Nelson Mohler, Eva-Maria Binder, Prof. Eike Jessen, and Dean Vijay Dhir.

Let me take this opportunity to thank Anthony Doyle for suggesting this project, Grace Quinn for seeing it all the way to publication, Tanja Jäger and Gayathri Umashankar for the production of the book.

Valmer and Entlebuch, January 2012 Dr. Dimitris N. Chorafas

Part I
Product Quality

Chapter 1
Product Assurance

1.1 Product Assurance Defined

When we talk of technology and of technological progress what we mean is the application of scientific discoveries through engineering. The quest for scientific advances is the object of basic research which works at the frontiers of knowledge. Scientific discoveries are brought forward toward practical implementation by applied research while development, the next step, produces useful products.

It is not easy to say a priori how valuable and beneficial a new product is going to be. For some of them demand is already there; whatever promotes health services in an ageing society is an example. For others the demand must be created. This is the role of marketing which is the driving source of sales.

The driving force of product development is technology. But which technology? The way an old saying has it: The technology you have is not the one you want; the technology you want is not the one you need; the technology you need is not the one you can obtain; the technology you can obtain is not the one you can afford. Real life may not be so negative; but using technology is not the only requirement. A company also has to deliver the right level of product quality.

The management of quality is not an option. It is an obligation for every company, for every designer, and every manufacturing engineer. It is, as well, a basic responsibility of all senior executives. From research and development (R&D) to design, prototyping, production, and field maintenance all personnel, not only that performing specific quality functions, must have sufficient and well-defined:

- Authority,
- Responsibility, and
- Organizational freedom to identify, evaluate, and report quality problems.

This can be effectively done only when the corporate culture is right and when the senior management gets directly involved in reviewing the status and adequacy of the quality embedded in engineering designs, as well as in manufacturing

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering,
DOI: 10.1007/978-1-4471-2966-0_1, © Springer-Verlag London 2013

processes. The accountability connected to product assurance requirements must be specific and detailed. Both organizational units and persons, should be directly responsible for quality.

As Confucius said, in his time,[1] however, the assignment of responsibilities can only then be effective when there exist clear definitions. Today, both in the literature and in the day-to-day practice, terms like quality management, quality control, quality assurance, and product assurance tend to be used interchangeably—while, evidently, they do not mean quite the same thing to all people. This creates a state of diffused responsibility.

Product assurance is the broader term which includes *quality assurance.* As a concept it is a guarantee extending from producer to consume. By itself, quality assurance indicates whether a product (or process) conforms to specifications and is within established engineering tolerances. The tendency is to associate product assurance to engineering design and all the way to after sales service, while quality assurance relates more closely to manufacturing.

Another term widely used since the 1950s is *reliability* (see Chaps. 6 and 7), sometimes confused with product assurance. That is wrong. Reliability is not an ability. It is the probability that under specified environmental conditions and over a given period of time, the engineered product will operate without failure to accomplish a specified mission.

This book provides evidence that quality and reliability complement one another and they must both be designed into a product at the drafting board. Engineering specifications and tolerances should reflect technical requirements related to the intended use of the product, but a simple adherence to them does not guarantee quality or reliability:

- Apart of being controlled at production level, quality will be verified in the course of operations with preventive maintenance playing a critical role, and
- Quite often, the quest for a product's reliability will start with destructive testing, and then follow through at component, subsystem, and system levels by means of physical testing and simulation.

Not only original designs but also manufacturing engineering and the production process itself are great contributors to quality and reliability, as well as to its contrary: Poor quality and unreliability. "I learned everything I could about possible causes of failure," wrote David Packard in his seminal book *The HP Way*, and "I decided to spend most of my time on the factory floor to make sure every step was done properly [1]."

Packard found that his approach significantly improved quality because it promoted personal communications. Person-to-person communications is often necessary to back up written instructions and specifications. He called that *management by walking around*, and it is indeed an integral (and important) part of *quality control* (QC, Chap. 3).

[1] Circa 500 B.C.

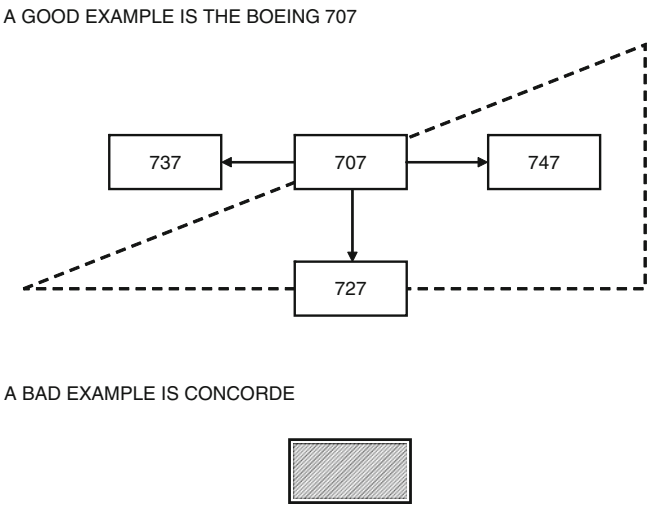

Fig. 1.1 Advanced technological products and the quality associated to them must be projected with the larger picture in mind

Statistical inference is one of the most important methods of effective quality control and, therefore, it should be used to full extent by all practitioners (Chaps. 8–11). The fact that quality control does not always use mathematical statistics[2] where this is feasible. This is indeed a mistake because stochastic thinking helps in promoting quality at two levels of reference:

- *Quality in the small*, which is quality control at factory floor and
- *Quality in the large*, which means total quality management starting with R&D, incorporating design reviews (Chap. 3), integrating SQC, and including field maintenance.

The risk with solutions too much centered *in the small* is that one can be boxed in a corner. Take as an example Boeing's 700 series and the supersonic Concorde. Figure 1.1 helps in visualizing what lies behind this distinction "in the small" and "in the large". The 707 was designed in a way that it could expand toward bigger versions and get redimensioned to smaller sizes. Working in the large means greater flexibility. By contrast, the BAE-Sud Aviation Concorde boxed itself in by being projected as a monolithic airframe. That is a way of working in the small through a narrow-sighted approach.

In product assurance terms the large picture is provided by quality in the large, with quality in the small being its subset. Detail is very important and quality in the small will provide it, but we should not start with a small concept and hope it

[2] For a discussion on statistical quality control (SQC) see Chaps. 12–15.

will grow. Products assurance can be best served if we start with a policy of quality in the large.

Alarms, checks, and balances as well as plans and controls embedded *in the large* can be instrumental in putting a limit to managerial and professional indifference as far as quality is concerned. Indifference to quality of service (see also Chap. 4) has characterized, for example, BP's lack of attention to safety issues in the case of the Texas refinery and in the Gulf of Mexico oil spill. Also, TEPCO's attitude in the nuclear tragedy in Japan (Chap. 5).

- Both TEPCO and BP had poor environmental ratings and
- Both dramatically demonstrated that deliberate indifference can have a very significant negative impact.

A key role of quality in the large is to address high-level quality decisions. Because such decisions lead to policy commitments, they should guide and substantiate company-wide quality initiatives. The remit of quality in the small is tactical, establishing the specifications for expected values of critical and tolerances. Also the observation of these tolerances in the manufacturing process.

- The conceptual framework of product assurance is established by quality in the large.
- The *what*, *how* and *within which range* of variation, is the responsibility of quality in the small.

In conclusion, as far as product and process quality is concerned we are confronted with alternative approaches: We can theorize as to the possible causes of the problem; proceed by trial and error toward what seems to be the most likely cause; or establish a firm product assurance program through quality in the large.

Theorizing brings no deliverables. Quality improvements by trial and error lead quite often to minor improvements in comparison to what is being targeted. Yet, as noted in the opening paragraphs of this section, the effective management of quality is not an option. It is an obligation which has to be effectively met.

1.2 Industrial Leadership and Product Assurance

The way it is defined in Sect. 1.1, product assurance is a holistic process served by the policy of quality in the large, and this should be present since the conception of a new product idea. To promote a spirit of company-wide product assurance, management must not only train all personnel on concepts and methods underpinning high quality, but also establish and maintain quality control databases using knowledge artifacts (agents) [2] to steadily evaluate:

- The product quality of suppliers and
- The firm's own in-process quality as well as that of outgoing products.

This is true not only with traditional engineering design but also with what has been recently called *frugal engineering* practices. (The term "frugal engineering" was first used by Carlos Ghosn, the CEO of Renault-Nissan, to describe making basic, low-cost products through a back-to-basics approach.) *If* a product is worth doing at all, *then* it must be done well.

The engineering may be traditional or frugal, but the producer is always accountable for the quality of the results. General Electric, for example, has applied frugal engineering techniques to develop basic medical scanners in India, which it now sells in the West, but nobody has accused GE of neglecting product quality.[3]

Low cost instruments and other products are welcome as long as their quality is upright. The same is true at the other side of the spectrum of engineering endeavors, miniaturization; miniaturization, too, has presented significant quality challenges which were also met in an able manner.

As transistors got smaller, laying down materials in proximity got harder, with the overriding requirements becoming a breakthrough in physics rather than in cost. New designs incorporated channels thin enough to be unthinkable with already known technology, and new processes had to be developed to radically increase the transistor count on an integrated circuit. As a reminder:

- The first working transistor came in 1947; 13 years later 100 transistors were packed in an integrated circuit through a planar manufacturing process.
- It took 20 more years to develop plasma etching, and another 2 years for UV-laser lithography to bring the transistor count to 100,000.
- The next quantum jump was copper interconnect. It came 15 years down the line, bringing the count to 10 million transistors.
- In 2000, silicon wafers featured 100 million transistors per integrated circuit, and within a short decade (by 2010) high K-gate electronics, liquid immersion and multigate transistors brought the count well beyond one billion in an integrated circuit.

Figure 1.2 gives a snapshot of this technological evolution made possible because product assurance has closely followed developments in physics. At every step, the choice has been breakthrough *or* cost, not both at the same time. With micro-miniaturization the choice was breakthroughs; with subsequent popularization the choice was cost. The customer was the driver for that choice which defined how far should the design go—or, alternatively, how much a producer should trim its products' cost to retain its market.

The market has its own prerogatives which vary with the client organization's weight. In military procurement the government's representative may disapprove a particular quality control procedure, or even reject the overall quality control system, if it fails to provide the necessary product assurance. By contrast, for

[3] Quite to the contrary, GE implemented and popularized Six Sigma (Chap. 12), which was originally invented by Motorola.

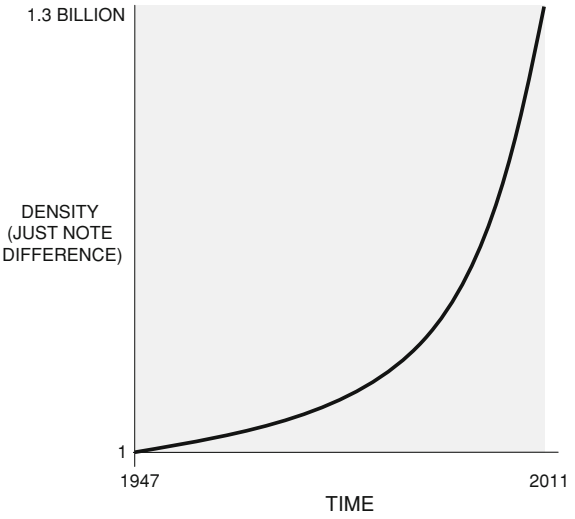

Fig. 1.2 Decades of technological evolution in transistor count on integrated circuits

civilian products the market at large, let alone the individual customers, has neither the authority nor the skills to prescribe a particular quality control solution.

In the longer run, however, other things being equal, quality assurance commensurate with the product will ultimately determine its success or failure in sales. Frugal marketing can push a lower quality product, but this will not last for long as customers discover that its quality leaves much to be wanted.

Highly capable entrepreneurs like the late Steve Jobs, of Apple Computer, know this and appreciate that in the longer run the prime motor in sales is quality. Price alone can go only so far. Jobs demonstrated that for leadership in design one has to be proactive and courageous, learn from his mistakes and assume the responsibility if the product fails. iPod® has been based on an idea which was out-of-the box but could deliver:

- Quality and
- Service.

"Treasure the things that are difficult to attain," urges a Chinese motto. That is what Jobs did, and he appreciated the fact that product assurance is by no means only a matter of inspection at the end of the manufacturing process. The iPod® did embed quality into its specifications at the drafting board. The best approach to product assurance is to examine a project, product, or process both forward and in reverse.

- Studying its flaws,
- Examining its limits, and
- Evaluating the more likely functions or places to break down.

In practical terms this requires that both before and after making an engineering design we should be looking for instances that indicate whether the original

hypothesis[4] and intermediate conclusions have shortcomings. It is much easier and far less costly to correct design flaws early on rather than later, after the specs have been drawn or, even worse, after the manufacturing process has started.

Speaking from personal experience, I have found it to be very important that the company's quality control system provides not only for detection but also for *prevention* of faults and other discrepancies. In quality assurance terms, prevention cannot be achieved unless:

- The overall climate within the enterprise is conductive to *quality consciousness* rather than indifferent to it, and
- Organization-wise are instituted specific and steady measures that make possible the identification of situations or trends that cause nonconformance.

There are challenges in doing so. One of the reasons why the analysis of quality assurance data is not properly done (if it is done at all) is that the complexity of relationships of obtained data discourages further treatment. This is an organizational failure, and experimental design (Chap. 11) helps in correcting it.

In other cases, the required testing program appeared too costly and therefore it is discontinued. When this happens management has no more control over the variables affecting quality. Research engineers know from experience that when the number of quality variables within a product (or project) increases, the amount of work required for an investigation rises in an exponential.

Intermittent tests and qualitative observations of quality output are by no means enough. Steady analyses should be made of the extent to which production, testing, and field maintenance conform to pre-established product assurance standards. To be sustained, such a policy requires both a procedural framework and the appropriate *culture*. Who says culture says people. More than four centuries ago (in 1597) Francis Bacon wisely suggested that if one works with another person he must either:

- Know his nature or fashions, and thus lead him;
- Or, his ends, and so persuade him;
- Or, his weaknesses and disadvantages and so awe him;
- Or, those that have interest in him and so govern him.

The impact of human elements on product assurance is felt all the way from specifications to manufacturing engineering and testing plans, culminating in the ultimate delivery of a product (or service) of required quality. While in terms of organization companies have latitude in devising and selecting methods and procedures, they frequently fail to pay due attention to motivation and skill needed to steadily satisfy quality requirements.

Last but not least, both for organizational and for human resources reasons, the product assurance function must have direct access to top management. It is a sign of industrial leadership that the chief executive officer (CEO) promotes a

[4] A hypothesis is a tentative statement which needs to be tested. More on this in Chap. 8.

company-wide unbiased and objective judgment of product assurance, after having heard first hand about quality problems being encountered and their correction or lack of it.

In conclusion, even if a company claims to abide by the principle of quality, in the end product quality is a pragmatic, and not a theoretical, result. Many beautiful theories of how to assure quality are merely that—beautiful theories. It is only thanks to looking at both the big picture: quality in the large, and by examining in-depth the detail: Quality in the small that we can be in charge of product assurance.

1.3 Real Quality Versus Appearances: A Case Study

While it is important to know the standards characterizing quality policies and procedures, their effects, details, and impacts can only be demonstrated through practical examples. Case studies are the best way to give to such examples both a bone structure and muscles. The present case study comes from one of the best known lamp manufacturing companies, and it starts paying particular emphasis to the quality of Wolfram, the ore from which incandescent lamp wires are made.

This product has been chosen as an example because while demanding considerable expertise lamps and their filaments (which is their most vital component) are simple enough in their description. This allows for processes visibility. With the exception of special lamps, the largest number of lamp types is made of up to 64 parts. Therefore:

- Complexity will not be an issue in this case study and
- A relative simplicity permits to emphasize the effect of policies on product quality.

Quality is the issue. How to make Wolfram wire is thought to be a well-defined process characterized by graduations in assurance of end product quality. This end product comes at different structured levels of quality, depending both on the basic material (Wolfram) and on intended use—but as always the devil is in the detail.

The company behind this case study is one of the better known lamp producers in the global market. I will call it *Omega* (not to be confused with the watch company), since its name cannot be divulged. However, all facts, figures, research results, and discussions the reader finds in this section and in Sects. 1.4 and 1.5 have been real life events.

Contrary to other case studies included in this book, where technology plays a critical role, high technology's impact is not really outstanding in lamp manufacturing. The most widespread and most popular types of lamps date back some 120 years.[5] The fact of dealing with a mature product allows the case study to

[5] The incandescent lamp is a nineteenth century development, one of the many which came from the brilliant mind of Thomas Edison.

concentrate on challenges affecting product assurance which survived the test of time.

Here are the facts. Since the start of the product assurance study which I did in Omega, I got the feeling that quality in the large was not coordinated in a way promoting uniformity of systems and procedures, even if nearly everybody recognized this was necessary. This relative heterogeneity in quality policies was equally present among the factories of the company's country of origin as well as in its operations abroad.

Early enough in my study it became fairly clear that the lack of a total company-wide quality assurance system may well prove to be the weakest links in Omega's organization. And, indeed, it proved to be so, outpacing all other challenges which included data processing, absence of optimization in production planning[6] and scarcity of analytical approaches to sales forecasting which weighted on:

- Inventories and
- Product availability details.

In addition, much of what has been said about the lack of a system supporting quality in the large, can be attributed to the absence of a systematic approach to quality standards which is widespread in the global lamp industry—as I can attest from personal experience with three different lamp manufacturers among the better known names.

Top management is not really conscious of this, and people at lower echelons have other hares to chase. In the beginning of the study I was told that as far as light efficiency, life tests, and geometric measures are concerned, quality control in the home country and abroad follows the same standards. But in reality there were no global corporate quality standards—let alone systems and procedures applicable on a multinational basis.

Also a handicap was the absence of a product assurance database designed and used for tests on outgoing quality level (Chaps. 10 and 13). There were, as well, precisely few projections, extrapolations, and management reports on quality performance. In addition, the rather limited quality control reports which existed were not communicated within the organization on a wider basis. There were a limited number of recipients who kept them close to their chests.

This meant that primarily, and above anything else, the attitude of the management and the internal company culture had to change. This was true both in the use of statistical tools and in internal communications. In more than one instance, the reference was made that statistical quality control is too much of a "theoretical subject" to have practical applications. There is nothing more untrue! In other cases, people responsible for different departments were suggesting that:

---

[6] Lamp manufacturing is notoriously difficult to optimize because lamps are made of glass and a small lots policy is very expensive because of breakage in setup time and the first runs following it.

- The trouble with quality was to be found at their neighbors' courtyard, and
- The responsibility for improving quality was not theirs but that of their neighbor.

In reality, both the trouble and the responsibility for it were to be found at the office of each and every department head, all the way up to the senior management. Walkthroughs were unknown and feedbacks rarely saw the light. Lack of quality feedbacks was hurting product assurance because it would have given warning signals, telling where failures start biting.

Not everything was, however, negative. The organization had pockets of excellence, but these could be found rather low in the pecking order. A practical example has been the experiments made to pinpoint reasons for deviations between factories in materials' quality, and the illusive nature of quality embedded in some of the lamps components.

A very interesting project on product assurance designed along experimental lines, existed in one of the Omega's research laboratories. It aimed at checking the behavior of fluorescent lamps part by part, because, as the responsible research engineer was to remark:

> We know that we do not possess the necessary know-how to identify all influences important to the life of a coil. Since we do not yet have in our hands the true experimental conditions, we are working to unearth facts and features which can be managed to improve quality.[7]

One of the research executives expressed the following opinion as to why developments taking place in the home country were not benefiting the foreign subsidiaries as much as it should have been expected. The focus of his remarks centered on organizational reasons which had not taken a global view, commensurate with the company's operations:

> Today each factory is responsible for the quality of its product. Hence, R&D can only propose improvements. Whether manufacturing follows the proposal is a totally different matter.

Besides this, organization-wise, between R&D and lamp manufacturing stood the functions of the components group. Components quality assurance covered subjects concerning semi-finished products such as glass tubes and bulbs; wire; sheets of tungsten and molybdenum; lamp bases; fluorescent powder, and the like. In addition, there were the usual supplier problems. Quality issues with bought semi-finished goods left their imprint on end product quality.

These were not dark corners that only existed in Omega. What I am writing is present (in its own way) not just in all other lamp companies but throughout the industry at large. The director of one of Omega's subsidiaries abroad[8] was to comment:

---

[7] The specific issue at the time of this meeting was the behaviour of the discharge electrode.

[8] Which had manufacturing facilities but no research lab.

> The trouble with wire is that the problems we confront don't only come from properties of materials but also from many other factors including atmospheric conditions. These affect recrystallization. Bringing a geometrically correct coil in the lamp is also a problem in itself.

It was not the first time that I heard that argument. In the lamp industry at large exists a legion of opinions, some of them contradictions, about the reasons why product assurance is not at its best. Not surprisingly, these different opinions are in no way converging which discourages lamp companies from getting together to overcome them.

The result is that needed cross-industry quality standards have never existed in an unambiguous manner. Neither have technical observations been integrated with practical experience in order to obtain a solid basis for judging quality results. This way, when new demands are placed on lamps and wire, they have to be met through improvisation rather than by means of a continuously evolving quality standard—no matter how much money is spent on R&D.[9]

This is a global problem which, to a considerable extent, has to do with top management policies and organizational reasons rather than being only a technical issue. To better appreciate the effects of lack of global quality standards, the reader should know that, barring specials, lamps are a mass product item.

Theoretically, in mass production it is much easier to observe quality standards because there are available the proper tools for doing so, centered around statistical quality control (see Chaps. 12–15). Practically, mass production is destined for a wide market where cost rather than quality is the overriding issue. Cost is guiding the hand of senior management in decisions being made.

The problem of product quality in mass production can be reduced to an expected value and a variance which must be steadily kept in control. Figure 1.3 shows how this applies with mass-produced incandescent lamps. Typically, the mean life is taken as equal to 1,000 use hours, and the quality assurance goal is to keep the standard deviation under control—which means as small as possible.

- High quality is characterized by a small standard deviation;
- A rising standard deviation is characteristic of low quality and of a process getting out of control.

When I asked during the early fact-finding meetings at Omega if this quality measure was respected, I was told that "There are working instructions to that effect. But they are only descriptive. They contain no information on how to control the manufacturing process by means of statistical quality control charts. Hence, while general instructions exist they do not permit the quality specialists to exercise timely process control."

[9] Among lamp companies this generally stands at about 6% of annual business.

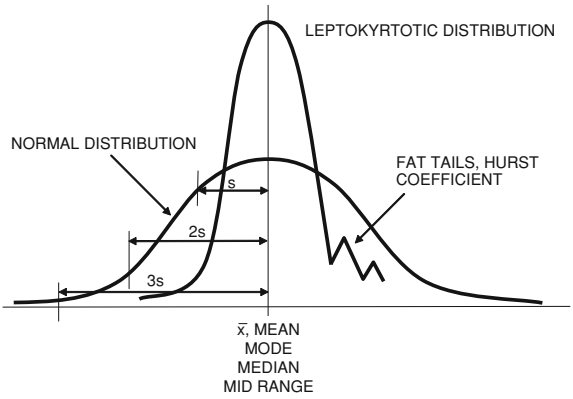

Fig. 1.3 Bell-Shaped normal distribution and leptokyrtotic distribution

1.4 Basic Quality Characteristics: An Example

If the filament of a lamp is its most important component and the one that, other things being equal, will determine its life cycle (Sect. 1.3), then a great deal of attention must be paid to its manufacturing process—starting with Wolfram ore. The ore, ingots, wire, and filament productions are the physical parts of the process. The human resources part includes the skills, methods, and tests being used. All these elements contribute to the quality of deliverables.

Theoretically, but only theoretically, the development of Wolfram ingots and subsequently wire and filaments is not a very difficult process. Practically, however, it requires a lot of expertise and this sees to it that several lamp companies prefer to buy their wire from competitors. As far as product assurance is concerned, this is the wrong policy.

While the reasons for outsourcing vary from one firm to the next, costs play a critical role in such decisions. Another main reason pushing lamp companies to outsource their wire and filament is the lack of the right experimental methodology. The next salient problem is testing, with two predominant issues:

- Absence of precise wire and filament specifications;
- Lack of appropriate wire testing skills (more on this later).

Regarding the first bullet, the typical argument from outside procurement is that wire and filament producers "know what they do". While this might be true, relegating specifications, tolerances and quality control to a third party does not bode well for product assurance.

The way to bet is that when a company's engineering has not written specifications based on appropriate experimentation on filament design and manufacturing including the necessary stress tests.[10] Interestingly enough, this is a problem

[10] As contrasted to the traditional lighting of a very small sample of lamps till burnout.

which also confronts companies making their wire and filament in-house but coordination is wanting.

In Omega's case, every factory was working, so to speak, on its own and regarding wire/filament specs and tolerances—following its own standards. In addition, R&D believed that the responsibility for quality assurance lay with manufacturing, while manufacturing was convinced that this responsibility would be squarely found on R&D's shoulders.

To my query about corporate specifications and standards, the director of engineering said: "We have specs but factories do not work on company specs. Factories observe specifications which they have developed on their own." Such a wrong-way policy was reinforced by the fact that several Omega executives tended to think that the homogeneity of quality and associated tests were not a priority "as long as the quality was good".[11]

It does not need explaining that problems existing within a country's borders become so much bigger in international operations because cultures, languages, and other barriers magnify the issues connected to homogeneous product assurance standards. Globalization has magnified quality problems—and this has more to do with organization and management rather than with technology.

In his seminal book about his years with Ford and Chrysler, Lee Iacocca states that a year after he became Chrysler's CEO, of the company's 31 vice presidents only one remained in its payroll [3]. Hire and fire is an excellent policy to weed out the unwilling, the unable, and the unnecessary but it cannot be universally applied because many countries have stiff labor laws. In other cases, it is the paternalistic internal culture that does not permit it. The president of Omega said:

> I don't want to see blood around. Bring them together and get a sense of an accord.

Such a meeting was indeed organized and it was fruitful as it led to the decision to establish a classification of technological priorities to which subscribed research and development, manufacturing, procurement, sales, and finance—of the home country and of the main foreign subsidiaries. Later I learned that the president had used the grapevine to carry the message that those who disagreed for disagreement's sake would have to go. Iacocca's shadow did the miracle.

Subsequent to this, researchers, manufacturing staff, components experts, as well as the sales and financial brass worked close with quality control to establish priorities in product assurance—starting with clarifying, in research projects, the milestones with an impact on quality. This job was cross-departments and was carried to detail, because the devil of product quality is always in the detail. Here are the highlights which provide an excellent reference on what can be achieved:

[11] Besides the fact that such a policy did not make sense lies the fact that "good", "bad" and "average" are totally subjective.

1. Pre-Products
   1.1 Metals. Main project: quality of wire
      (a) Split in wire problems
      (b) Need for a new method to prepare tungsten wire
      (c) Need for a new analytical methodology, including X-ray fluorescence
      (d) Low temperature work (He 4°)
   1.2 Phosphors. Main project: quality production of phosphors
      (a) New type of phosphors needed for uniform quality
      (b) New methodology to prepare phosphors
      (c) New methods for quality control, starting with establishing pre-criteria
   1.3 Glass
      (a) Rethinking quality criteria connected to quartz
      (b) Research on additives on glass (titanium bioxide)
      (c) Investigation on wire cracks in the glass base
   1.4 Electrodes. Main project: avoiding the blackening of quartz walls
      (a) Reduction in Si0 which results in blackening the glass cover of electrode(s)
   1.5 Ceramics
      (a) Problems with seals
2. Lamps/discharge
   2.1 Low pressure
      (a) Challenges with electrodes-coil; breakage and blackening by emitter material
      (b) Challenges with amalgam for lower pressure lamps
   2.2 High pressure. Main project: prototypes of new lamps
      (a) Iodide problems (circle process)
      (b) New components research necessary with emphasis on simulation
3. Incandescent and solid-state lamps
   3.1 Incandescent
      (a) Circuit process (tungsten) between gas and coil. Simulation
      (b) Filling and control of lamps with iodine
      Notice that until 1958, incandescent lamps presented no particular needs for research. Since the 1960s however, R&D projects have focused on making lamps smaller in size, designing a new generation based on quartz or special glass, and undertaking long-range and medium-range projects for “the lamp of the future”.

3.2 Solid state

(a) Solid-state research
(b) Electroluminescent lamps for special purposes

4. Chemistry and metallurgy
Analytical and synthetic chemistry; work on metals
5 Measuring
Projects focusing on precise measuring of light, color, spectral distribution, temperature and more

Since the late 1960s the development of a new analytical methodology for metals became an urgent matter for research as evidence was provided that wire problems increase with the lamp's sophistication. At Omega research projects identified 11 factors of major importance in wire design and manufacturing: Deposition of lubricant, lubricant material, baking the lubricant, degree of dryness of lubricant, the die, drawing (pulling) the wire, parallelism between the geometric axis of the die and direction of the pull, lateral direction of the pull (in absolute terms), temperature of the interface between wire and die, roundness of the hole of the die, and drawing speed.

A consensus reached by the working group which included senior executives from R&D, manufacturing, procurement, sales, and finance has been the need for improvements in wire product assurance. In common accord, these involved:

- Closer coordination between R&D, manufacturing, procurement, sales, and finance
- Company-wide standardization of quality control practices
- Simplification and accuracy of manufacturing methods
- Coordination and review of interfactory practices (in home country and abroad)
- Establishment of standards for manufacturing equipment and tools
- Free communication of quality records between "producer" and "user" departments
- Development of a first class database of product assurance accessible by all authorized people

The most important outcome of the product assurance project has been that Omega's senior management, and the professionals, became anxious to have efficient solutions to the quality problems facing the firm worldwide. It was also appreciated that an able solution requires not only calls for study and research on product assurance but also a cultural change in the organization with steady evaluation of the fruits which it bears.

Table 1.1 Research needed for checking what is not easy to see

Process(wire manufacturing)	Needed research
Ore control (average quantity received 10–20 tins)	Ongoing analysis of procured ore
Decomposition	Analysis of efficiency and process computations
Separation of Wolfram (washing with water)	Analysis on ferrous calcium and other residuals
Solving and addition of tungsten solution	
Processing over ultra-filter	
Exchanger for + ions	Analysis for manganese
Exchanger for – ions	Analysis for fluorite[a]
Filtering	
Evaporation	Process calculation needed for the evaporation[b]
Separation of ammonium para-tungsten (APT)	
Recrystallization of APT + $5H_2O$ (white powder)	Analysis for H_2O, ammonium, tungsten, trioxide
	Analysis regarding particle (grain) size
	Study of modification of crystals
	Study of the balance of Wolfram + NH_4

[a] In this and the previous research steps, there exist considerable possibilities for human error
[b] In this and in process computation, continuous measurements must be assured during decomposition on: reaction, velocity, temperature, and other variables

1.5 Research Needed for Checking What Is Not Easy to See

The story of wire manufacturing from ore to wire of, say, a 0.2 diameter, goes roughly like this: Wolfram ore is received and undergoes a quality inspection.[12] Classically, tests of procured material have been provided by consulting laboratories with which Omega collaborated over many years. These laboratories essentially assured that the ore meets certain minimal specifications which are, primarily, economically oriented: the Wolfram to be extracted will not be less than a given percentage of the mass of ore.

Curiously enough the independent labs checked impurities only against a ceiling. Yet, as we will see, a quality-bound main worry had to do with the exact amount of different impurities—with details well beyond the observance of a ceiling. Technically speaking, if one knows the *exact* composition of the ore, then these impurities can be taken care of in subsequent chemical processing prior to producing para tungsten.

[12] In Omega's case there have been two main ways for making tungsten from ore: the one is Wolframit with procurement sources Spain (pure), Peru (Sio_2), Borneo (Sio_2, Po, Arsenate). The other is C_aWo_4 with sources: Spain (pure), China (much Si, tin, fluorite), Russia, Sweden, Finland (much fluorite), and Canada.

Table 1.2 Research projects aimed at challenging the "obvious"

Process (wire manufacturing)	Devil's advocate suggestion for needed research
Reduction to blue oxide (treatment through furnace)	Study on temperature limits and effect of their variation
Doping, washing	Research on doping and its effects
Handling through meshes (separation of grain)	Study on optimal grain composition with a view to the (subsequent) pressing operations. Different grain sizes have to be examined
Reduction to tungsten in hydrogen	Research on the proposed process of washing (water, acid) to reduce doping additives to 10%
Bottling into big jars.[a] Storing. Dividing into little jars.	Research on the effects of storing. (A deterioration process can take place in respect to the pressability of the powder)

[a] Jars up to 1 ton constitute the lot

On the contrary product assurance will feature important gaps if an Omega factory is short of impurity details. In this case, wire production will feature impurities emanating from the ore and, over and above those connected to manufacturing. Less cryptically this means that:

- *If* impurities sneak into the wire manufacturing process,
- *Then* product assurance will be out of control.

It follows that for every lamp company an important scientific investigation centers around the existence of impurities in the ore. It targets their exact qualitative and quantitative definition which leads to unambiguous establishment of their variation and of the effects of their presence. Table 1.1 presents, step-by-step, the nature of the research the working group of Omega defined as necessary for checking what otherwise is not detectable in product assurance terms.

The process described in Table 1.1 in the left-hand column has started already in operation in the Omega company. The working group, however, judged that a finer test grid was necessary because more detailed tests are crucial in allowing the process to continue uninterrupted as well as in indicating the need for recrystallization. In addition, individual parameter limits and composite results had to be established in a way that their accuracy is experimentally verified.

The 10 experiments on the right hand column of Table 1.1 were requested by R&D and by components manufacturing. It was unanimously agreed that one of the most basic experiments in this list is testing of the *crystallization hypothesis*. In the background of that decision was the fact that recrystallization was handled differently by the different lamp factories of Omega.

[For starters, the *crystallization hypothesis* states that the grain size of the salt (Ammonium para tungsten) will influence the grain of the powder to be produced later. This hypothesis has to be proved for every type of ore and for every origin of the ore. Short of that, there exist doubts about its exact identity, and these doubts find their way in product quality.]

Neither was recrystallization the only worry as far as product assurance is concerned. Say, as an example, that a given lamp factory disposes three containers,

Table 1.3 Further research projects necessary for improvements in wire manufacturing

Process (wire manufacturing)	Needed research
Pressing of green ingot ("as pressed" ingot)	Research on the result of having the powder washed prior to green ingot
	Examination of the wisdom (and possible outcome) of pressing near round or polyhedral ingots
Sintering bottle (heat treatment; pure hydrogen for 45 min)	Study of alternative time schedules (The current one which based on tradition lacked experimental proof)
	Research on effects of the distribution of densities in green ingot, on the resulting quality after sintering
	Study of feed-forward features capable of being provided by an Ohm test of the sintered bar
	Study on the effects of double length sintering bars to reduce 10% + 10% (2 ends) rejection rates

and that each processing requires 5 h. Because the time between processing intervals should definitely be counted, a lot of ore (for instance 20 tons), needs more than 1 week's work for treatment. Thus, weekends interleave with workdays and the risks on quality arise there and then because of weekend stops.[13]

I deliberately insist on these details to demonstrate that any serious project on product assurance must start where the process itself begins. Testing only at the end is worse than useless; it is misleading. That's why I have pressed the point that procurement of wire from a third party can lead to a truncated and, therefore, inaccurate quality control process. Other quality problems which subsequently sprung are described in Table 1.2.

The reader should appreciate that each of the outlined issues characterizing a consistent and focused quality assurance effort, requires a systematic study to bring the underlying quality variables under control. Critical scientific investigations should be designed to test hypotheses made by the company's professionals. To broaden their horizon, researchers should also be avid readers of tests and results published in the literature emanating from other R&D projects.

In my professional work I have as well found the *devil's advocate* to be a rewarding practice. This is a role to be played by a mature researcher approaching his retirement. At Omega, the devil's advocate made the hypothesis that existing quality controls were not properly focused—because they did not pay full attention to the fact that the most important matter influencing the subsequent quality of the wire is "doping and the reduction process which follows". A small experiment proved him right, having proved the influence of the doping substance which affected both:

- The quality of crystallization, and
- The cracks which may later be presented in the wire.

[13] Philips tried to work on ore on a continuous, non-stop basis but this did not work so well, either.

To reach these conclusions the reduction process was studied not only in absolute terms but also relative to the influence of the furnace and of temperature variation on grain size. The goal was optimal grain composition. The devil's advocate suggested to focus on the mixing process. His thesis was that *if* the grain form is destroyed by mixing, *then* the quality would suffer. Research on wire manufacturing also involved the steps described in Table 1.3.

As this example documents, upholding product assurance requires a well-done homework. And whether the theme is wire manufacturing or any other, details should be an integral part of the picture.

Most of the projects outlined in this and in the preceding sections are a "must" if a lamp company really wishes to master product quality. It should furthermore be appreciated that each of these projects was subject to constraints. For instance, vibration cannot be used for equalization purposes, because it tends to differentiate grain sizes by weeding out the smallest grains.

This is an unwanted consequence because good wire quality presupposes a fair mix of grains (small, big), making unwise mechanical pretreatment. On the other hand, non-equal mass results in an uneven spread of doping material and the cut size itself may vary depending on the quality of the spread. (More on the Omega case study in Chap. 2.)

In conclusion, product assurance is by no means what the French call "a petit travail tranquil".[14] Product assurance has no place for guesswork, unchallenged "obvious" ways of doing things, or fantasy precision. "If you don't keep challenging yourself, you start wasting away," advises Lee Iacocca.[15]

References

1. Packard D (1995) The HP way: how Bill Hewlett and I built our company. Harper Business, New York
2. Chorafas DN (1998) Handbook of agent technology. McGraw-Hill, New York
3. Iacocca L (1988) Talking straight. Bantam Books, New York

[14] A small undemanding job, a characterization usually reserved to the postal service.

[15] Idem.

Chapter 2
Quality Control

2.1 Quality Control is a Flexible, Polyvalent Concept

The standards adopted for product assurance are manifesting themselves not only in the domains of inspection and testing but also in the examination of rejected and borderline units. Inspection, testing, and investigative feedbacks from the contains of throwaway bins are quality in the small (Chap. 1). It is wise as well to appreciate that:

- No complete verification by inspection operations is ever possible;
- A 100% inspection, if and when it can be done, is by no means a 100% protection.

The second bullet underlines the importance of statistical quality control (SQC, Chaps. 12–15), and the basis for inference which it offers. Statistical inference, however, presupposes that we know what we are looking for and what kind of risks we would like to keep under lock and key if and when we can do so.

This emphasis on the relative merits and limitations of statistical inference (see also Chaps. 8–11) is deliberate because a manufacturing organization is a system involving human beings whose attitudes tend to mix objective and subjective factors. On the other hand, the success of any industrial quality control system is largely contingent upon its ability to handle people by providing a solution which:

- They can easily understand;
- They are willing to use it in their day-to-day work without being overburdened by its complexity.

Quality control is in no way intended to be a way of putting people under stress; properly applied it acts as a smart structure endowed with an array of sensors—the workers, supervisors, and engineers at factory floor. Their feedback produces a stream of data to be analyzed to provide continuous monitoring of quality's integrity.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering,
DOI: 10.1007/978-1-4471-2966-0_2, © Springer-Verlag London 2013

This is written in a factory-wide sense. As for the quality control (QC) department, its responsibility is to assure that the production output conforms to the engineering specifications. For this reason QC is usually attached to manufacturing, though it may as well be given a broader mission including incoming material rating.

Indeed, well-managed companies establish a procedure by which QC and purchasing work in close cooperation. This avoids QC problems down the line because of defective materials which were not properly screened out at reception. Let us look at this alternative through a small case study, which exemplifies what companies do for obtaining good quality supplies at minimum cost. A sound procedure has three steps:

- Where past supplier data are available, the QC organization compiles a list of items (using drawing and part numbers) which quality-wise are considered troublesome.
- The purchasing organization prepares a list of problem items in a like manner, including source of procurement; hence supplier service and dependability. Such items are considered troublesome due to quality, price, or service.
- Based on the aforementioned two lists of items in common agreement, QC and purchasing create a single list to be used in the company's procurement and incoming materials rating program.

If there was no past supplier information, reference data will have to be asked from the clients of a given supplier and further enriched with future information as it becomes available. At first instance, in this case a tightened control is advisable. In addition, a supplier's appraisal programs should be scrutinized and compared to our company's quality control plan(s). Company Alpha, a real entity, has been using three factors for determining a source of supply. In order of importance:

1. Quality
2. Price
3. Service

In its evaluation Alpha allocated 50 points to quality; 36 points to price; and 20 points to service. This allocation of weights was flexible enough to suit departmental requirements if necessary, without changing the basis of the overall QC plan or the way in which it worked.

QC and purchasing based their rating on the assumption that incoming lots are either acceptable or rejectable. If all lots were acceptable over a specified period of time during which a given item was rated, such as 1 month, a number of 50 was assigned to *quality*. Otherwise, this number 50 was reduced according to the degree the lots were unacceptable, by:

- Determining the ratio of acceptable lots received as a percentage and multiplying it by 50;
- The resulting number gave the weight to be assigned to quality.

A similar procedure computed was employed for the price factor. The lowest price was assigned a weight of 30. Higher prices were used as denominators in a ratio with the numerator as the lower price. Multiplied by 30, this ratio gave the weight to be assigned to the vendor's offer in terms of the cost factor.

Purchasing based its rating of service quality on the vendor's handholding, timely delivery, promises kept, and similar criteria. A percentage of 100, 90, 80, or some other ratio has been used to adjust the maximum weight of 20. Contrary to the quality quotient which was specific to each item and represented product assurance, service rating labeled the dependability of the vendor.

The sum of the compound rating of these three metrics: Quality, cost, and service were judged as excellent (if it were 100 or 98); good (from 93 to 97) and fair (from 86 to 92). If it were 85 or less, purchasing would eliminate that vendor's offer as being questionable at best.

That is all, with regard to QC relating to incoming materials and devices . In connection to its own production, Alpha has used SQC targeting quality in the small within each factory's walls. There have also been charts for *trend of trends* assisting the management's thinking in terms of corporate-wide quality of manufactured products.

Many of the tools Alpha employed for stochastic inference were its own, though based on generally accepted principles. The underlying concept of experimental inference can be expressed through *if, then* rules which, among themselves, reflect one of the basic scientific laws:

- *If* the experimentation tells me something is erroneous;
- *Then* I have an evidence it is erroneous;
- *If* Experimentation tells me something is *not* erroneous;
- *Then*, I have *no* evidence as to whether it is erroneous or correct. Therefore, I tentatively accept the hypothesis as being right.

In science, SQC is a scientific tool implemented in an industrial environment; we are far more sure when we *reject* than when we *accept* a hypothesis, or a "fact" under test. What we mean by accepting is that we have no evidence for rejecting it, but we know that a "nasty" new fact may pop up in subsequent experimentation, leading to its rejection.

In Alpha company, the trend of trends and specific SQC charts were not filed away and forgotten. The output of SQC enriched the product assurance database just like sensor networks collect information elements all over the environment where they are implemented. This constituted the quality assurance database of the company's corporate memory facility (CMF) [1].

A CMF is an important and integral part of the governance of a modern enterprise. Inter alia, it registers all decisions being taken, the justification supporting them and the names of the decision makers. Quality data is an important part of a company's memory. As David Shenk suggests "Without (memory) all life and thought are an unrelated succession. As gravity holds matter from flying into space, so memory gives stability to knowledge…. It holds us to our family, to our friends. Here a home is possible; hereby only a new fact has value [2]."

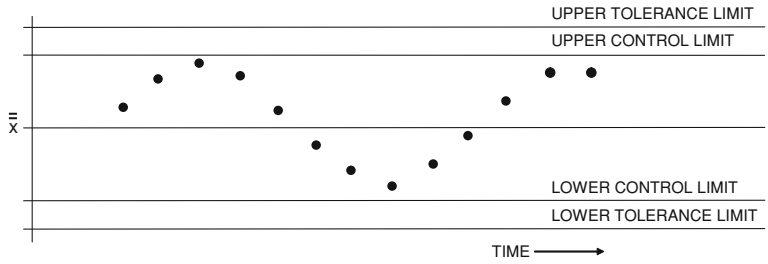

Fig. 2.1 Statistical QCC by variables: A preview

Information elements in the quality assurance database of Alpha's CMF permitted to reconstruct online, on request, statistical charts by variables and by attributes of any operation in any one of the company's factories, at any time. As a preview of the discussion in Chap. 13, Fig. 2.1 gives a glimpse of how an SQC chart by variables looks like.

An additional advantage, in terms of industrial management is that the stream of SQC data permits to identify and challenge chance causes: What makes the normal pattern of variation look the way it does? Can the cause(s) of variation be clearly identified? If yes, can they be eliminated? Should changes to the process pattern of variation be expected? This brings into the picture the need for inference, but it is good to remember that:

- In principle, the process pattern of variation may be predicted if only chance causes operate.
- To the contrary, the process pattern of variation may *not* be predicted if assignable causes of variation hold the upper ground.

In conclusion, the methodology defined in this section helps in identifying the importance of QC at large, and most specifically SQC: A process is said to be in control when it has a stable pattern of variation within prescribed limits and assignable causes are not operating in this process. Control charts tell us when the production process is in a state of statistical control (Chaps. 12 and 15).

2.2 The Expanding Horizon of Industrial Statistical Methods

During nearly seven decades since the end of World War II, statistical methods have progressively become an important adjunct not only to engineering but also to most industrial and financial sectors of the economy. From manufacturing processes, statistical methods expanded into procurement and acceptance inspection (Sect. 2.1); then into applications in finance and banking.

In historical sequence, the first industrial domain where statistical methods demonstrated their importance has been the control of quality in the production

process, including sampling inspection (Chap. 9) and quality reports for management. The performance of experiments in scientific laboratories (Chap. 1) is another example. This has permitted engineers, physicists, and other professionals to:

- Draw objective conclusions from their experiments; and
- Base them on verifiable data rather than subjective considerations.

A contributing factor to the expansion of implementation of statistical methodology has been the war effort in the US, over the 1940–1945 timeframe. Military specification MIL-Q-5923, which specifies the contractor's duties (that he shall develop and sustain an effective and economical QC system)—saw to it that QC has been charged with assuring:

- Adequate inspection coverage, present throughout the entire process of manufacture, including packaging and shipping; and
- Evidence of required inspection to be provided by the contractor prior to submission for acceptance, so that an objective quality evidence is evaluated and verified by the representative of the defense procurement agency.

To the question of what is needed for an objective report on quality, US military specifications stated that experimental evidence is an absolute requirement, documented by means of actual test results. To the question regarding sampling inspection or 100% inspection, WWII US military specifications left no doubt that sampling inspection is required not only when the test destroys the product units but also when it is a more economical or efficient solution, as it is so frequently the case. Other questions revolved around quantity and quality viewpoints in manufacturing (see also Chap. 10 on operating characteristics curves):

- The producer's viewpoint is that he aims for quantity and hopes for the benefits of quality. He demands protection against the rejection of a good product.
- The consumer's viewpoint is that he aims for quality and hopes for the benefits of quantity at reasonable cost. He demands protection against the acceptance of a poor product.

Both viewpoints have weighted on the way SQC methods have been implemented and developed over time. To the question when does SQC best apply? originally the answer was: "When units are produced in quantity," but it has been superseded by the more accurate response: "When the quality measurements with which we are concerned vary." If we take fine enough measurements, all machines and processes exhibit variation. When this variation increases we have a quality problem which can be solved by SQC techniques, as well as through experimentation. As Chap. 1 has made reference to this happening by bringing to the reader's attention that product assurance is a system with several variables having an effect on the result.

This raises the question about the best method for studying these variables' effect. When testing a device, product, or system a number of approaches are available for finding the optimum combination of variables.

- The classical method suggests that one variable be measured while all others are held constant.

This has been the classical approach, but it takes time and there is a risk of missing hidden, essential information resulting from the variables' interaction. An alternative method chooses values for each variable at random (largely by intuition), scattering them about in a way which will expose major trends. Its downside is that data of this sort are not readily analyzed in a mathematical way.

To the contrary, by means of statistical inference (Chap. 8) the analysis can be systematic. All points statistically scattered yield tables of data which are easily reduced to a series of equations.[1] Through analytics, we obtain computed values for the amount of variation. But as we will see in Chap. 11 the experiment must be planned and only a researcher with considerable technical knowledge and judgment can plan experiments and properly interpret their results.

Experimentation and QC have a symbiosis. As QC, practically the first major industrial application of statistics gained ground, the requirements grew for more sophisticated tools and this led to better focused research and experimentation (see also the Omega case study in Chap. 1). Experimentation returned the compliment by promoting more advanced application of statistics in the industry beyond the original confines of:

- Frequency distribution, as visual representation of a pattern of variation (which, in a more coercing manner is also given by a simple histogram).
- Central tendency (or value) about which the other measurements tend to concentrate or spread;
- Pattern of variation of a quality characteristic which describes the value at which this characteristic has been measured or would be measured under certain conditions; and
- Spread of the pattern of variation describing not only the extent to which the values deviate from the central value but also the outliers and long leg of the distribution.

One of the milestones in the expanding horizon of industrial application of statistical methods has been a whole family of *quality control charts* (QCCs) related to a quality characteristic or group of quality characteristics resulting from a production process. Figure 2.1 has presented an example of QCC. Notice that upper and lower control limits are within the engineering tolerances (more on this in Chaps. 13 and 14).

This chart comes from *inspection by variables* where the actual measurement of the quality characteristic is taken and recorded (see Chap. 13). The alternative is *inspection by attributes* where units of product are classified as defective or non-defective with respect to an attribute or set of attributes (see Chap. 14).

Each point on the quality control chart represents one sample drawn from the production process. The information from the sample is summarized into a statistic

[1] However, data which have not been scattered statistically cannot yield statistical information.

whose value determines the vertical location on the chart. The time, the sample was taken determines the horizontal location of the point. The chapters on SQC by variables and by attributes will provide much greater detail on this process.

With experience accumulating throughout business and industry, the sophistication of QCC implementation has grown, and statistical inference got a boost. One of the most interesting applications of SQC thinking has been to help us choose one of two hypotheses about the production process:

H_0: The production process if in a state of statistical control; hence *in control*.

H_1: The production process is not in a state of statistical control; therefore, it is *out of control*.

Figure 2.2 offers a bird's-eye view of the quarter spaces in a test of hypothesis. This is a preview of this most important tool of mathematical statistics which is explained in Chap. 8. The application of statistical inference in business and industry has brought us, as tools, the test of hypothesis and statistical quality control charts. The underlying process can be summed up as follows:

- Identify the production process
- Choose the particular quality characteristic with which the control activity will be used
- Select a random sample of *n* units from the production process
- Inspect the sample by testing or measuring each of the *n* units and record results
- Calculate the value of the statistic and plot the point on the control chart
- Make a decision dictated by the rule governing QC, with two options: *accept* or *reject*[2]

Taken together these six steps add up to a rule which underpins the test of hypothesis: *If* the pattern of measurements falls between the lower and upper control limits, *then* accept the null hypothesis: H_0. By contrast, *if* one or more points fall outside the control limits, then reject H_0; accept the alternative hypothesis H_1 and render the consequences of this decision. When statistical quality control charts are used to arrive at our decision, our options are limited in the four quarter spaces of Fig. 2.2.

2.3 Instituting a Quality Control Program

Experience in quality planning and control indicates that a product assurance program will only succeed when it reports directly to top management, is staffed with the best of brains, and has been given clear objectives. The function of the group whose responsibility is to watch after quality should be to establish, monitor, and evaluate the performance of all planning and control programs affecting quality and reliability.

[2] In Chap. 9, on sampling, we will also examine a three-option system: accept, continue sampling, reject.

		DECISION MADE	
		H_0 : IN CONTROL	H_1 : OUT OF CONTROL
ACTUAL STATE IN PRODUCTION PROCESS	H_0 : IN CONTROL	CORRECT	ERROR*
	H_1 : OUT OF CONTROL	ERROR**	CORRECT

* THIS IS α (TYPE I ERROR, PRODUCER'S RISK). SEE THE DISCUSSION ON OPERATING CHARACTERISTICS CURVES (OCC) IN CHAPTER 10

** THIS IS β (TYPE II ERROR, CONSUMER'S RISK) IN OCC

Fig. 2.2 The test of hypothesis "in control" by using QCC

Since quality planning starts at the drafting board (Chap. 3), the leader and members of the quality assurance group should furnish design engineers, procurement specialists, manufacturing engineers, and sales engineers with methods and tools that assist them in meeting product assurance parameter(s). *Thinking* makes the difference and thinking must be part of every professional's job. One of the most resourceful businessmen, Thomas Watson, Sr., used the logo THINK almost at par with IBM. In his Cambridge lecture on August 31, 1837, Ralph Emerson said:

> *If* one is a true scientist, *then* he is one who THINKS.

Niels Bohr, the nuclear physicist, was teasing his peers, his assistants, and his students by telling them: "You are not thinking, you are only being logical." Great men in history have always appreciated that thinking means challenging the "obvious", therefore, doubting and experimenting. Both *doubting* and *experimenting* are the roots of high quality and of quality-oriented services.

This quality-oriented professional group should delegate teams of specialists from among its members to work with specific engineering, manufacturing, and field service personnel (where applicable). Its quality planning and control expertise must be available wherever needed both for consulting and for quality and reliability auditing reasons.[3]

[3] The work which I have done many times in my professional experience. Consulting and auditing should never be done by the same expert in the same program, project, or product.

For example, the procurement operations may require assistance in controlling the quality of complex supplies, components, equipment, and systems which became more sophisticated and for which current requirements are inadequate to provide the needed quality assurance. Part of the aforementioned group's contribution should be to assure that purchased materials and devices conform to upgraded quality requirements. Another equally important duty is to examine and ascertain interface compatibility between new devices and those already used by the company.

No QC program can be successful unless it employs highly knowledgeable personnel, preferably by developing an in-house experience which presents greater possibilities for continuity, though some outside assistance is also very helpful. This calls for the ability of maintaining key people. (The president of a Silicon Valley company was saying in a meeting that every evening watching his engineers take the elevator to go home was asking him if next morning they will be also taking the same elevator to come to his firm.)

At project level, a team of product assurance specialists should be headed by a group leader who reports directly to an executive vice president, who himself reports to the president. Depending on the project, the EVP may be heading R&D, Manufacturing, or After Sales Service.[4]

This type of direct reporting structure helps in assuring that the members are not placed in the untenable position of having a subordinate status to the people whose work they are expected to plan or control. In the case of technical audits of quality and reliability, reporting to the president is a "must". No A subordinate ?? status will make the technical audit function ineffective. Here are ten challenges I encountered in my technical audits:

1. Predict system product assurance based on paper designs or a breadboard model.
2. Providing information needed by QC in upholding tolerances during pilot production, while manufacturing engineering has not yet completed its homework.
3. Testing product prototype(s) with pilot equipment to determine whether specific quality (or reliability) figures are being met.
4. Suggesting means of improving quality after testing without returning to and redoing the original design.
5. Assisting parts procurement in a mandatory switch of supplier(s) and evaluating their non-standard parts.
6. Suggesting methods of further improving quality (or reliability) through better components or design techniques.
7. Providing an adequate model of possible tradeoff while using a limited number of parameters.
8. During manufacturing audit and improve QC methods, tracking obtained results through statistical quality control charts.

[4] See also the discussion of after sales service in Chap. 4.

9. Assisting QC in setting up production tests and/or administers such test independently of the manufacturing organization.
10. Provide support for good internal communications between design and manufacturing engineering, and between manufacturing engineering and field maintenance (see also Sect. 2.5).

Many of these challenges are present not only in average designs but also in good ones. An example of good design, when product reliability warrants it, is to have two separate automatic control subsystems so that the total system can survive the loss of one and still function. This has been the case of A380 (Qantas Flight QF32) on November 4, 2010. The Airbus superjumbo touched down after suffering a potentially devastating engine failure shortly after takeoff from Singapore, but thanks to control system redundancy the pilots made a safe landing.

As we will see in Chap. 6, the redundancy of critical systems is a design strategy that both presents advantages and has constraints. Originally practiced with weapons systems, it migrated to civil aviation and now the automotive industry is following the aerospace industry's lead. Automotive experts forecast more and more parallels in the automotive world [3].

This evolution in design strategy toward greater product assurance came as automotive engineers watched and learned from potential failures in aerospace which could well materialize in their own business. Neither is design the only area of a transfer in experience—QC concepts, standards, methods, tools, and tests have to be further upgraded because of the so-called "global platform" which gives auto companies the chance to push down production costs,[5] but lacks multi-manufacturer standards—and if lower quality sneaks in—that practice can be very expensive.

It does not need explaining that there are great practical challenges not only in the manufacturing of totally homogeneous platforms in all the countries, an auto company has such facilities, but also in partnerships between motor vehicle firms. Although manufacturers are still trying hard to achieve a truly global quality, in practice this may never be the case as they source most of their parts and materials locally.

These are precisely challenges confronting QC engineers, and can be found in the background of challenges No. 6, 7, and 8 in the list earlier in this section. If raw materials are different from one plant to the other, what kinds of tradeoffs are available to lift product assurance? In practical terms, the task breaks down into a series of control functions:

- Better design control
- More focused procurement control
- System analysis aimed to flash out discrepancies
- Tight inspection in production QC

[5] AT Kearney has estimated that global platforms will become increasingly important, accounting for almost 50% of global production by 2015.

- Cross-factory homogeneity evaluation
- Steady field feedback control.

While these missions are largely technical, they are not solely technical in nature because they also involve financial issues, like cost control and senior management's strategy. Costs matter; to shore up its balance sheet, General Motors in the process of making substantial cost cuts are in prospect too.

For instance, as older workers retire many are replaced on GM's assembly lines by new ones earning half as much, under a two-tier wage structure agreed with the United Auto Workers. This is one reason that the carmaker has closed a cost gap with its Asian rivals that was once $2,000 or more per vehicle [4].

It will however be short-sighted to believe that all the reduction done in wages and in health costs will save GM from another bankruptcy. The longer term answer is greater productivity and much greater attention to quality, because as I never tire repeating there is nothing more expensive than low quality—from breakage and rejected materials to recalls and the loss of reputations which negatively affects the customer base.

2.4 Unwarranted Resistance to Better Methods for Quality Control

Section 2.2 provided the reader with evidence that the application of mathematical statistics in production offers valuable means for the detection of errors involved in the manufactured goods. It also helps in creating a corporate memory facility, most valuable in estimating developments taking place in outgoing quality level because something is happening in the production process.

There is no better evidence of a QC system's product assurance deliverables than the results obtained during the World War II years, particularly in the United States. This strategy of reliance on tough quality measures, assisted through statistics, was further reinforced during most of the decade which followed WWII. Curiously enough, however, it then ebbed and started to decline in the West while it rose in Japan (but by year 2000 it also declined in Japan).

The rise and fall of quality assurance conscience can be approximated by the log-normal curve. Indiscriminate cost cutting is one of the reasons, the other being a new generation's lack of appreciation of the important role played by quality in the company's growth and survival. Reduced skills in mathematical analysis has been a third reason which saw to it that not all firms were ready to apply SQC. Negativism about SQC includes some of the following arguments.

But we have always made money using the system we have. Why change?

That's a very defensive statement and a wrong one. It also shows the lower level of know-how and shop experience existing in the firm. Sharpening up the

whole decision process about accepting or rejecting a product, or a lot, on quality reasons enables a company to go much further in product assurance.

This is a general tool, but my problem is unique.

Although each product and process has its own characteristics, something really "unique" is a very rare bird indeed. This is just an excuse for doing nothing, and it also shows that the person using it does not understand the flexibility and applicability of an SQC methodology. A factory should use SQC to gain competitive advantage. Practically any process is adaptable to this technology.

They tried it at so-and-so's and it did not work; therefore we don't try it.

As Dr. Edward Coleman, my professor of QC at UCLA, used to say: "Just because one cannot make a board smooth with an axe does not mean that the axe is a bad tool. A professional must know a misapplication or failure of a tool or method due to misunderstanding or lack of skill. This, however, does not detract from the validity of the tool when properly used.

We have too many variables in our process to use it.

If it is really so, *then* this is one of the best reasons for using not only SQC but also experimental design (Chap. 11) as basic prerequisite to better management. Properly applied, SQC will help determine which of the many possible causes of past failures is really affecting the quality of the factory's produce.

It is too technical (or too complex).

It is true that the use of tools based on mathematical statistics requires (indeed involves) something of a revolution in the company's culture and skills. But SQC is by no means "too technical" or "too complex". People with a high school or even grade school training have learned to use successfully the simpler SQC methods (see Chap. 15). The basis for SQC is mathematical. However, what lies behind its evolution does not have to be understood any more than one has to be a mechanical engineer to drive a car.

It is fine for long runs, but we only make short runs.

As an argument this, too, is a fake. SQC technology can be applied to as few as 1 (one) piece. This is by excellence the domain of percent defective (Chap. 14). The economics of the situation and the length of run may determine the techniques used, but length of run alone will not determine the applicability of SQC.

While the foregoing six examples of pseudo-reasons for not using SQC are dirty excuses often heard by reactionaries who are unable to change their professional culture, it is no less true that—like the microscope—a more powerful methodology or tool reveals hidden or latent problems in a so far "traditional" process.

This negativistic view is by over 90% defensive. It also documents lack of know-how. A case in point is errors in measurements where one is unable to distinguish between:

- Random errors and
- Systematic errors.

Here is an example. When the same object is measured repeatedly we get a series of measurements which, due to a large number of uncontrollable factors of small importance, vary *at random* about a certain value. They vary in such a way that:

- Really random errors can be considered as normally distributed about zero; and
- The uncertainty of the method being used is characterized by the standard deviation(s).

By contrast, *systematic* errors are usually due to some isolated factors with an impact. They result in a displacement (or translation) of the measurements in one direction, with observations distributed about a "new" value, say, μ different from the expected mean population value μ_0.

- The systematic error is equal to $\delta = \mu - \mu_0$

Sometimes when examining the systematic error of a method of measurement, it is possible to design our approach so that the expected mean value of a measured quantity becomes known. Systematic deviations from the standard indicate that the production process is not functioning as it should. An investigation of the raw materials, of the machines, of workers skills, and of the manner the machines have worked is necessary in order to find the causes of the trouble. In principle, the larger the change in the population mean, the greater the probability of getting warn signals through the control chart.

Another problem associated to product assurance in engineering and manufacturing is labeling. Both for the producer and for the consumer, it is advantageous to identify the quality characteristics of a product through a label attached to the product itself, but there are prerequisites which are not always observed:

- The label must be truthful in its contents; and
- Its contents must be understood and appreciated by its user(s).

If every manufactured product has a "quality characteristics label", then a great deal of repetitive work will be avoided. For example, in the case of Omega (see also Chap. 1) we instituted a policy of quality labeling each individual tungsten ingot, each wire package, and each lot of filaments made by the same wire. Prior to this, there was only a monosyllable description which said practically nothing in real about the quality term to the user. Hence, he had to do all over again the QC routines. In addition, to be successful a quality labeling policy:

- Must start at the lower level of the food chain (in the Omega case the ingot); and
- Continue uninterrupted as step-by-step the product gets transformed, either within the same company or by another entity.

The problem of implementing SQC at all levels of production discussed in the first half of this section (albeit from the negative viewpoint of resistance to change) and that of correct labeling correlate with one another.

- SQC provides most valuable input for the labels; and
- The labels inform the next production stage regarding the quality information it needs, without having to reinvent the wheel.

Quantitative data is important because part of the problem with labels is that their message is not always understood. Take carbon footprint labels as an example; the idea has been that carbon labels would let shoppers identify products with the smallest carbon footprints, just as other labels already indicate bio-production telling the shopper that this merchandize is greener.

However, a survey carried out in 2010 by *Which?*, a British consumer group, found that just 20% of shoppers recognized the carbon footprint label compared with a recognition rate of 54% for organic labeling (which, it should be noticed dates back to the 1970s in Britain) [5].

This has been disappointing because adding a carbon label to a product is a rather complex and costly process that involves tracing its ingredients back up their respective supply chains and through their manufacturing processes, to work out their associated emissions. As if this lack of public recognition was not enough, the current practice is also confusing.

Different carbon foot printing and labeling standards have emerged in different countries, preventing direct comparisons between the various types of labels. This should be seen as a reminder that when we talk of more sophisticated approaches to information relating to industrial production and merchandizing, standardization is a "must". It is even more so in a globalized environment.

In conclusion, at one side the last four decades have seen a regression of QC standards and a relative abandonment of methodologies—like SQC—which time and again have proven their value. On the other side there is a push, largely by politicians toward identifications made in a way not clearly comprehensive as well as suspect, because they lack the needed support by analytical methods, like SQC, which could make them unavailable.

2.5 Open Communications Channels and Quality Improvement

Generally speaking, data acquisition and data presentation for product assurance reasons has not yet reached the level of standardization and sophistication required by a knowledge society, and this is particularly true in certain cases where technological developments continue to run at high pace. Data whose acquisition has not been planned, will be found lacking vital details or even headline information such as:

- Component identification
- Functionality specifics
- Operating conditions for acceptance

- Reliability characteristics
- Performance measurements
- Description of failure mode(s)
- Failure symptoms and aftereffects
- Diagnosis of past failures and their avoidance

In other cases the inadequacy of information content comes from the fact that not only test conditions vary from one outfit to the other, even in the same country, but also there is diversity in approaches to gathering, analyzing, and presenting quality-related data. This is easily observable in the case of reliability studies.

Just like a successful product assurance program must start at the very beginning of a new project, the gathering of quality and reliability data should begin at the product's drafting board. *If* the company's QC procedures are only designed to address problems at the final production level, *then* quality information will be skin-deep and its use will be ineffectual.

Fulfilling adequacy requirements for data analysis is not so difficult because in manufacturing the majority of man-made products go through a number of phases which can be seen as discrete steps. An example is the case of supplier–client relationship(s).

Going back to the case study of the Omega lamp company, which we followed in Chap. 1, the dichotomy created by the existence of discrete steps will characterize the passage from the Wolfram mineral-to-ingot transformation (done in one factory), to the ingot-to-wire fabrication (executed by another manufacturing entity), and finally the wire to filament process.

The first company is a provider to the second, and the second to the third. In every case there is a client who asks for information about quality assurance. Omega engineers suggested that in order to state in a documented way whether or not electric conductivity is of any value in pre-telling the quality behavior of wire and filament, it is necessary to have information about a number of physical tests done by the ore-to-ingot plant.

For example, the measurement of resistance (Ohm) at the sintered ingot level. These data will be so much more dependent if quality measurements were based on statistically valid samples. For this, it is not enough that the appropriate physical study is taking place; it is also important that measurements are taken from samples representative of the population, their labeling is correctly done (Sect. 2.4), and they are databased.

The communication of correct information about product quality can be invaluable not only in connection to the needs of R&D but also in regard to product assurance experiments. In Omega's case, once these product experiments found that *if* a bar of Tungsten has equal conductivity throughout, *then* the resultant wire is good—though none of the experts was able to explain "why" this is so. Said one of the experts: "The future of the wire's quality is embodied in this bar but, as things now stand, we cannot measure it in advance. More research is necessary;" and, I would add, open communication channels which informed all players on the finding some of them have made and the conclusions they reached.

Open communication channels can also carry the message of return on investment. There are research projects that have a significant direct financial impact. An Omega example is a study on the effects of doubling the length of the sintered bar (from 600 mm to 1200 mm). This was intended to reduce the rejection rate, as in a sintered bar 10% has to be dropped from each side.

The aforementioned study was motivated by the fact that prior efforts using trial and error provided no valid solution. Research was needed because doubling the length of the sintered bar presented both a materials and an equipment challenge; a fact which serves as a reminder that in a significant number of cases research on quality assurance involves a dual perspective:

- Materials and
- Equipment.

Open communications let company people know that research results are strong and serious. When a research project on product quality is successful it motivates other people to suggest another one able to provide a better understanding of their problem. There is always a "next step".

The sequential progression toward wire quality described in, Sect. 1.5, did not end with the steps presented in that text. More people came forward with hypotheses to be validated through experimentation. A sample is shown in Table 2.1.

One of the interesting research projects focused on rethinking and re-establishing conditions under which a spool is rejected. In many lamp companies roughly 8% of wire production is thrown away at that very step, in an effort to improve the highly variable (and usually low) outgoing quality. Rejection evidently impacts on outgoing quality level but also affects manufacturing cost/effectiveness.

Altogether, the technical audit brought up 25 necessary research projects brought forward by company people who were previously too timid to ask for them. Most of these were aimed at correcting faulty product and process situations that were carried along by "tradition".

Some of them were counterproductive. When I asked the director of one of the Omega factories how it happens that the wire spools were not labeled for outgoing quality level, I received the answer: "It is company policy that the wire factory controls its quality, but that it should not give out any data" (see also in Sect. 2.4 the discussion on labeling). One should not be polemic but is allowed to ask:

- By whom, when, why, and how was this wrong policy established? and
- How it happens that the high rate rejection at coil manufacturing level has not led to a sharp policy revision?

Among the properties that data coil manufacturing factories wish to have are weight, diameter, resistance, crystallization, tearing features, and fusion. As one of Omega's factory directors put it: "For our operation, it is most important to know this data. It is also vital to be able to trace mistakes to each machine as they happen. Today, we simply cannot do that kind of tracing and until it is done, we

Table 2.1 Research requested by the omega factories to help them meet quality standards

Wire manufacturing process	Needed research
Recrystallization at 8.3 mm	Thorough and documented study on the properties of crystallization
Mechanical treatment to 2.8 mm	Optimization to reduce the resulting materials waste which reaches up to 5% of the weight of the material, but could be reduced to 1.5% (optimistic estimate) or 3% (pessimistic estimate)
Test for Splits (electronically done)	The prevailing 1/10 rule needed rethinking and research aiming to help in reducing the percent rejection of coils
Packaging in spools	A relatively small project was needed to establish the kind of quality information to be forwarded with the spool from "producer" to "user", in accordance with the firm's new product assurance principles

will not be able to control outgoing quality levels." From design to procurement, manufacturing, sales and after sales service, the sharing of information is not just the best policy, it is the only one able to support and promote product assurance.

In conclusion, both analytical approach to product and process quality and open communications should become corporate policy—replacing the silos approach which in most companies characterizes quality and cost information. When this happens, it makes sense to interconnect computer aided design (CAD) workstations into a firm-wide network for sharing R&D, design, and production data on quality. On the contrary,

- *If* secrecy continues being the rule of the day,
- *Then* CAD networking provides only an illusion of something happening, while in reality wrong-way policies prevail and there is no improvement in the results.

CAD workstations or no CAD workstations, an engineering database cannot be established let alone operate under silo conditions. Neither can a factory systematically improve its quality. If a manufacturing process does not control the incoming quality, it cannot control its outgoing quality no matter what sort of wizardries it introduces to its technology and in its inspection operations.

After, and only after, doing away with secrecy and the silos, it becomes highly advantageous to handle *all* product assurance problems within the organization by interconnecting the different design, manufacturing, testing, and other decision steps. It also makes sense to model the cascade of events taking place within the firm and between the firm and the market—to provide ways and means for correcting faults and steadily improving quality.

References

1. Chorafas DN, Steinmann H (1994) Database mining. Lafferty publications, London and Dublin
2. Shenk D (2001) The forgetting. Alzheimer's: portrait of an epidemic. Doubleday, New York
3. www.automotivedesign.eu.com, Sept 2010
4. *The Economist*, Nov 13, 2010
5. *The Economist*, June 4, 2011

Chapter 3
Designing for Quality

3.1 Mission of a Design Engineer

Achieving high quality in a product or process is a goal that must be pursued from the very beginning of a project, while it is still at the drafting board. As Chap. 1 brought to the reader's attention, the design engineer should not only observe the overall product requirements but also carefully consider the impact of each individual component used or to be used in the product, analyzing its capabilities and limitations as well as its dependability, cost, and sources(s) of procurement.[1]

- A sound product design will be strong in components analysis, and
- Its R&D will be supplemented by *an intended applications* study incorporating critical elements of manufacturing.

Other things being equal, higher product quality is promoted by standardization, reproducibility, and what I call "component stability". Standardization, however, presents its own problems as we will see in Sect. 3.5. Changes in component parts may be required to capitalize on new technology, lower cost, or other reasons. But it is not less true that many changes are made just for the sake of changing, and this is not a good practice.

The reproducibility of component parts requires that all engineering specifications and tolerances be known and accounted for by manufacturing engineering. Component stability implies that parts and subsystems being used are free of inherent weaknesses and *if*, for whatever reason, these cannot be avoided *then* they are compensated by careful design.

Between the lines of the foregoing paragraphs lies the fact that design considerations must encompass a number of critical decisions and account for boundary conditions imposed by the mission the product is going to fulfill and the availability of critical components. The evaluation of any and every component part should consider:

[1] See in Chap. 2 the attention paid by the Alpha company on these issues.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering,
DOI: 10.1007/978-1-4471-2966-0_3, © Springer-Verlag London 2013

- Deterioration,
- Intermittent failures,
- Maladjustments, and
- The unavoidable wear-out.

It is therefore important to provide adequate safety margins (see Chap. 7) which allow for the various disturbances that may occur, a case exemplified by the behavior of a circuit during its lifetime. To design highly reliable and stable products, the engineer must appreciate that in their wider utilization they may well be subject to operation under extreme conditions, accelerated deterioration, and sudden failures. As far as electronic circuits are concerned, three criteria have been established to guide the designer's hand:

1. The circuit must be tested if it meets its performance specifications with all components at their worst initial tolerances, and with any one of them at its worst end-of-life tolerance.

This is a stress testing criterion representing a good engineering practice in search of circuit longevity. When a circuit which meets it has been found, the second and third criteria must be considered and met.

2. The circuit must be able to withstand loss of any one of the supply voltages in it, in any circuit connected to its input or output, without component damage.

Murphy's Law is behind this requirement. If anything wrong can happen in the useful life of the circuit, it will! Companies which conveniently forget to test for worst conditions resemble people taking leave of their senses.

3. Since all components degenerate with usage, circuit design should include means for detecting significant changes in component behavior during use of the circuit.

This must be done in steady manner so that an alarm goes up soon enough to assure replacement of the component before failure occurs. The best design policy is the endow critical circuits with knowledge artifacts (agents) designed to test component parts for drifting.

These are basically *quality requirements* which underline the amount of attention to be placed on procurement. Because fully vertically integrated companies are not the rule of the day, and for many of them an amount representing 50% or more of their annual business is purchased material from suppliers, procedures must be in place for control of the quality of supplies and their wares including procedures for:

- Identification, and
- Disposition of supplies of lower quality than that being targeted.

Not only all nonconforming supplies must be diverted from normal material movement channels, but also the engineering database a designer is accessing online from his CAD workstation should be updated. Historical records must reflect supplier dependability or lack of it (see the case study of company Alpha in

Chap. 2). Since registering bad news is not a common practice periodic controls are needed to assure all nonconforming supplies are properly identified to prevent their use.

It is highly advisable that senior management appreciates the effect of this information process on procurement and the fact that weeding out nonconforming supplies constitutes a prime element of a company's quality control effectiveness. The right policy will identify and isolate material containing departures from specified quality requirements.

- *If* such departures from quality standards exist also in the company's own product line,
- *Then* their causes must be fully analyzed and prompt corrective action taken to prevent recurrence.

Let us face it. We are living in an age where the concepts of quality employed by the typical company are inadequate. Engineering failures are not properly analyzed to learn lessons from them. Yet doing so is a professional mark of distinction.

Since the late 1950s, which means more than half a century ago, Dr. Robert Lusser the reliability administrator of the US missile program,[2] had identified 12 traditional concepts of poor electronic design, testing, and reliability control [1]. These were:

- Drifting into complexity,
- Reliance on "bug hunting",
- Reliance on standardized components,
- Reliance on testing to specified limits, and derating
- Reliance in inadequate failure reporting,
- Reliance on inadequate production and environmental testing,
- Reliance on an average level maintenance,
- Inaccurate means for measuring reliability,
- Aversion to failure testing aimed at increasing reliability,
- Reliance in testing entire assemblies, rather than testing component by component,
- Reliance on redundancy as general solution (see Chap. 6), and
- Lack of statistical thinking.

Lusser could have had as well added in his list of negatives a disconnect between design and production, which persists even today despite the fact that computer-aided design and computer-aided manufacturing (CAD/CAM) provides an excellent means for interdepartmental communication. The need for coordination does not concern only manufacturing per se, but as well production scheduling, set-up and pilot runs often characterized by a short period of defects.

[2] And designer of the German V2 rocket during WWII.

In *The HP Way* David Packard gives an excellent example of how he promoted the individual initiative of his engineers, and their focus on product quality, while maintaining unity of command. He set clear objectives for every individual effort. In his words: "…objectives were meant to be evaluated from time to time and, if necessary, modified for the future benefit of the company [2]". Management control was exercised through feedback, a policy which gave excellent results in a variety of HP's engineering missions.

Part of the secret of success of Hewlett-Packard in its early years was that tolerances were required for all engineering products. Everybody in the organization was informed that if this is not done properly, even a relatively simple operation can lead to product failure and he would carry part of the responsibility for it.

3.2 Improving Quality Assurance

Striving to improve the accuracy of specifications, establish strict tolerances and try to ease the manufacturing and assembly work—all of that done at the same time—leads to an increased complexity. This is counterproductive. When it happens, designers are often presented with the problem of choosing between:

- Performance and
- Quality assurance.

The devil is in the details of this difficult choice. *If* performance is placed first, *then* quality may suffer. The opposite is also true, particularly in regard to reduced versatility. This is not a one tantum happening but a frequent event in engineering design. Trying to hit two or three birds with one well-placed stone amounts to "drifting into complexity".

Sometimes the choice of functionality over quality at design level is made on the premise that later on, in the life of a device or system, field tests will provide the needed information for quality improvements, and/or for reducing the level of complexity through simplifications. This approach, however, presents three problems.

- Quality standards introduced postmortem are usually half-baked,
- The simplified system may not work at good performance level because some of its performance was compromised through simplification, and
- The experimenter to whom was given the mission of "higher quality, lower complexity was not able to obtain a statistically valid sample of data for inference.

Some people believe that a statistically valid sample is not important because testing to a specified level of severity will bring out the potential failures. To a certain extent this might be true, but it does not reveal the modes of failure. In most cases the fact that a component has failed cannot be used as a clue as to *why* this failure has occurred. Derating usually reduces the hazard of one parameter, yet there may be many other parameters whose effect is almost as serious as the one eliminated.

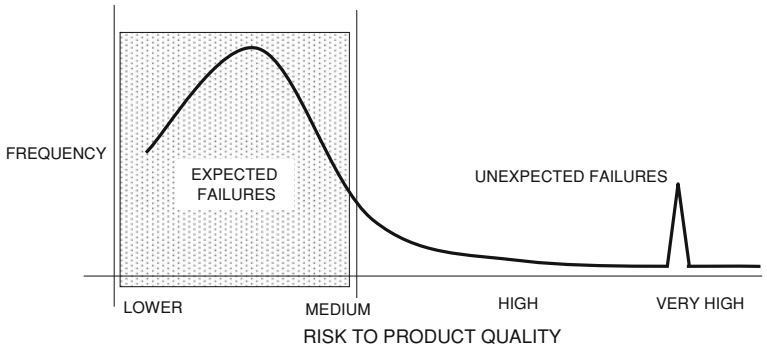

Fig. 3.1 Expected failures are typically normally distributed to the contrary, unexpected failures show up at the long leg of the risk distribution

Forgotten parameters and those whose importance "has been known", and therefore they were downplayed, can create havoc in product quality. Parameters which are under control usually lead to expected failures which are typically normally distributed—with the exception of baby failures and wear out failures.

Not all failures are however normally distributed and as Fig. 3.1 shows the more damaging problems in terms of product quality are unexpected failures, which find themselves at the long leg of the risk distribution. The best policy to flash them out is stress testing, but it is not fail-safe because many unexpected failures are due to so far unknown factors and parameters. In such cases, it is far from clear which one(s) of the design variables should be subjected to stress tests.

An added problem for the designer aiming to improve product assurance is that in spite of advances in failure reporting this practice still has shortcomings. For instance, since checking all of the characteristics of procured components are next to impossible, many defects remain hidden from the testers and therefore cannot be stored in the engineering database for subsequent use. Nonconforming supplies are a headache.

Life tests, too, have constraints beyond those imposed by destructive testing. In electronics, everything has to be tested for life—i.e., for the life of the component. For some elements this is satisfactory. For larger systems that are stationary, this again is acceptable. But in other cases, this approach is not fail-safe and in many cases the benefits on quality and reliability obtained through testing are overemphasized.

- Unreliability is not the cause of short life. On the contrary, short life is the cause of unreliability
- Higher quality standards can be achieved only if engineers from research to design, production, and operation are fully aware of an almost universal tendency to failure among man-made devices.

Many people believe cost issues and the pressure of schedules are responsible for failure-prone characteristic of man-made devices and systems. Superficially, this is true. More fundamentally, however, the cost will be lower in the long run if higher quality is achieved. The principle for engineering design is: If something cannot be done well, it should not be done in the first place.

The reasons behind difficulties in improving product quality explain why design engineers should be given guidance regarding the level of assurance they should target. This will help them in the preparation of their product quality plan, which is of value only when it is specific to the problem confronting the designer and not a standard prodding like "be careful in your work".

While it is always necessary, carefulness in the work one is doing is not enough. Imagination, ingenuity, and other qualities are also needed. The best advice that can be given to a designer is to enlarge his or her culture in a way that allows to be effective and prosper not only by working in big ticket projects but also, if not particularly, by working in *niches* which require much more imagination and ingenuity than mass products. Niches and unique products are where:

- Customers prize quality over cost and
- Are willing to pay the price for an exceptional accomplishment.

In addition, while what the designer does should reflect his company's quality assurance policies, he should be given the freedom to model his solution according to what he understands as being his project's requirements. A counterpart to this is that as he progresses in his work he should provide conclusive evidence that he has complied with corporate policies.

Hand in hand with the development of tolerances and specifications should come the definition of inspection and test procedures regarding the product in design. Good governance requires that the description of quality assurance as well as of all applicable inspection and test procedures is available to senior management on request, and besides that they should be the subject of technical audits (Sect. 3.3) and of design reviews (DRs) of (Sect. 3.4).

In addition, concomitant to the attention paid to product quality and functionality, which lies behind the guidelines and examples given in the preceding paragraphs, design engineering should be keen to preestablish ways and means for change control. The CAD/CAM system organization and that of the engineering database should assure that not only the latest applicable drawing, technical requirement and design change information is available online, but as well the original engineering plans. Such policy permits not only a fallback, in case it is needed, but also walkthroughs and flashbacks. This requires keeping online for authorized access:

- The original design and annotations made to it;
- Identification of all intermediate changes;
- All reasons which led to these changes; and
- All decisions made regarding these changes, along with identification of the person(s) who authorized each change.

Chapter 2 has pressed the point for establishing a basic procedure connected to design engineering, procurement, manufacturing engineering, and field maintenance—but also available to senior management to answer cost, quality, or other queries. This is the role of the *corporate memory facility* (CMF)—brought to the reader's attention in Chap. 2. While all changes must be processed in a timely manner which assures the updating of specifications and design features all the way to manufacturing and inventories, there is great merit in retaining records with references to the history of a product's evolution.

In conclusion, research, development, and engineering jobs must be led toward placing emphasis on product assurance. For every product, this is up to a point conditioned by the industry to which it is addressed, the market's drives and implicitly expressed requirements, as well as the need to be ahead of competition. The company's longer term survival depends much more on the quality of its produce than on other technical factors.

3.3 Technical Audits

Competition is a matter of vital importance to a company because it keeps it fit and running. Competition determines who succeeds and who fails; who advances and who retreats; who profits and who loses. The real battlefield in business is the mind of the clients who must be served with functional products and services, of good quality and at an acceptable cost.

The fact that the board or top management has approved a new project is no assurance that this project will succeed. Or, if it succeeds that quality guidelines will be respected, its functionality will be the one originally intended, the development timetable will be held, and costs will not rise to a point to price it out of the market.

During the period of applied research and development *design reviews* (DR, Sect. 3.4) are now firmly established as a means of controlling progress toward meeting objectives. It is to senior management's interest to promote DRs as a basic engineering parameter, and therefore it must be ready to devote time, money, and effort to DRs. But the useful life of a product must be also carefully reviewed and evaluated. For this reason, well-managed companies promote *technical auditing*.

It has been a deliberate choice that before tackling DRs and their impact on quality assurance, which is a prime area for the exercise of management control, are examined the dynamics of technical auditing which set the stage for all subsequent discussion. These dynamics rest on five pillars:

- Development timetables become faster and faster leading to rapid compression of design schedules. That is one of the results of global competition,
- The obsolescence of products already in the market has increased. Every company tries to develop a "better mousetrap" in terms of lower cost/higher functionality—with quality often paying the bill,

- In other cases, conflicts may arise because of the dual requirement: low cost and high quality. These must be solved almost in real-time to avoid undue delays in development.
- Effective coordination and avoidance of frictions between competing a company's divisions is at a premium, as different divisions clash driven by different criteria and priorities,[3] and
- The proverbial long, hard look is not frequently taken in regard to what happens or might be expected to happen to product assurance and its transition development, to manufacturing and field service.

As far as timetables are concerned, a product that would have taken in the past more than a couple of years to leisurely design must today be ready for marketing in 6 months or less, and this must be achieved within the perspective of hardware and software in full evolution. The test of timetable observance is one of technical audit's most important tasks.

The test of real and fancy obsolescence is another. A growing range of alternatives possibilities and the emergence of highly integrated circuits are making obsolescent what existed so far. In addition, traditional tools and design methods, including CAD, have been mostly geared toward a lower level implementation. They cannot effectively model the behavior of systems which become increasingly complex. Using more or less classical approaches to design complex new products:

- Dramatically increases the risk that the resulting solution will not be competitive,
- Jeopardizes the goal of bringing product innovation to market ahead of competition, and
- Practically assures the result may drift toward high cost/low quality; the opposite of what is wanted.

Everyone of these risks should be addressed by technical audits, Not only to be but also to *remain competitive* a company needs careful and honest technical audits that combine the results of past DRs with problems encountered in production and in field feedback.

Speaking from personal technical audit experience, their pattern is one of internal management control intelligence. Their findings allow design engineers, manufacturing engineers, and field engineers to better appreciate quality problems and achieve a common vision of the product's challenges.

An example of the pattern I am talking about is shown in Fig. 3.2. It identifies the drop in product assurance as the file is transferred from applied research to development and then to manufacturing. A significant drop into customer premises and responsibility for its performance is assumed field maintenance.

[3] A classic example has been that of IBM in the 1980s where the division making microprocessors clashed with the PC division which preferred Intel chips—and top management was not able to decide which way the chips should fall.

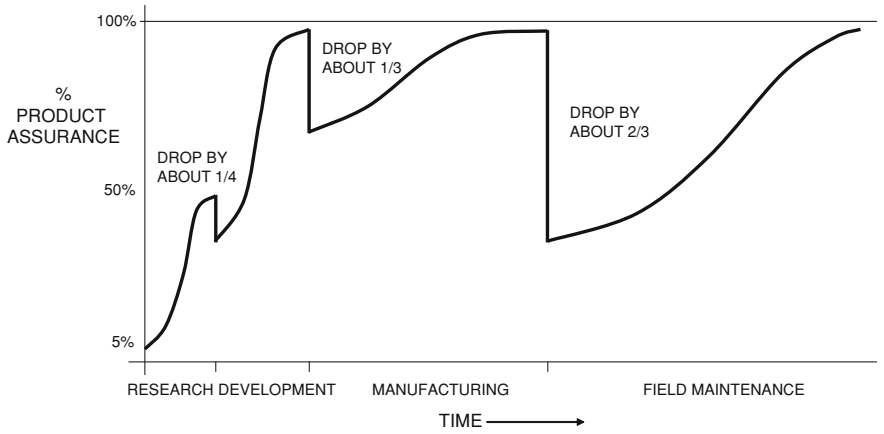

Fig. 3.2 The old way of working in airtight departments of research, development, manufacturing, and field service is detrimental to product quality

The data reflected into this pattern come from three different computer manufacturers: Univac, General Electric Bull, and Olivetti Computers, with whom I had worked as a consultant, often in technical audit. There are reasons for bending of the curve. A new computer coming into the product line can only then be properly maintained when the technical people responsible for this job are properly trained and gain experience with *that* model's failures (see also Chap. 7 on availability). This takes time, and this is detrimental to product quality.

Another domain where technical audits provide needed insight is the convergence of communications and computing which has created design challenges of its own for hardware designers and developers of embedded software with a new technology it is difficult to explore the interactions of a new product's components to their full extent; and the risk of errors in the translation of algorithms into designs is always present. In addition, as a technical audit revealed:

- There existed heterogeneous test benches among different development centers of a firm, all of them working on the same project, and
- There was a lack of integrative view of system design leading to miscommunications, therefore misunderstandings and misapplications.

Added to these factors has been the fact that because the team members of development centers had different backgrounds, design specs were prone to be misinterpreted while design verification, and therefore corrective action, occurred too late. When this happens, the risk of design flaws increases and some of the goals originally enunciated cannot be implemented.

What is necessary, therefore, is a common view of both system and components-level that makes feasible to coordinate technical requirements. Technical auditors should therefore examine if all design stages and procurement practices share common methods and promote seamless interoperability.

Tool barriers must as well be overcome whether these are a classically used CAD solution, algorithms, system diagrams, embedded software, or a hardware description language. Not only aged means of design but also a tool's specificity can be a weakness when it comes to sharing results among labs, within engineering offices or between laboratories—with colleagues who do not use that same tool.

In conclusion, achieving first class results in engineering design requires technical audit to provide intelligence on methods, tools, tests, and timetables. Technical audits are part of a company's internal control, providing invaluable assistance whether their object is quality, performance, costs, or other factors affecting deliverables.

3.4 Design Reviews

For any practical purpose DRs are a process of rigorous auditing while the project is still on a development course. The successful chief executive officer does not sit on his projects. He assigns his people to them and gives them the authority necessary to execute them fast and well. Authority is delegated but this is not true of responsibility which rests squarely on the shoulder of the CEO, his director of engineering and the project leader. Hence, the importance of:

- Handling each project the company undertakes *as if* its future depended on it; a reference equally valid of the status of the project leader and his professionals, and
- Controlling through appropriate milestones the project's progress, including quality, functionality, cost, time spend, and plan versus actual in terms of deliverables.

DRs are the project leader's best tool to assure he is in control of the project. They consist of well-timed, rigorous evaluations of crucial variables on which depends a project's success. The object is to assure compliance to goals and specifications. It is advisable to use both quantitative scores and qualitative comments to express the design review's results. The successful project leader not only plans his work and that of his team meticulously but also controls through milestones:

- The time people depending on him are spending;
- Costs incurred up to the day of the DR versus budget.

It needs no explaining that it is to every's interest shareholder that the project succeeds. Therefore, DRs are no love affairs. They have to be biting. Success cannot be assured by chance, neither just by using a kind word.[4] DRs are of two types:

- Minor, and
- Major.

[4] Al Capone said to a journalist, with a kind word *and* a gun you can go so much further.

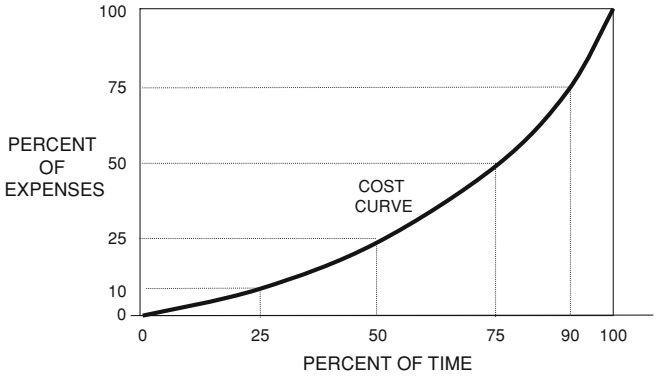

Fig. 3.3 Major DRs should be well-timed, rigorous, and pointing to corrective action or cancelation of the project

Minor DRs are internal to engineering and they should be done weekly, based on focused question, accurate answers to them, as well as database mining, spreadsheets, and modeling. By contrast, *major* DRs are much more demanding and should include technical auditors, risk managers, marketing experts, financial people, and consultants. Also, senior management.

Following the policy and practice I learned in the 1960s with General Electric, *major* DRs should be done at 25, 50, and 75% of budgetary allocation. The person chairing a DR should *not* come from engineering, but should be an acknowledged expert endowed with the authority to kill a project there and then if the results are unsatisfactory (more on this later).

Figure 3.3 presents in a nutshell the timing of major DRs. The coordinates are the project's progress in time toward its goal and budgetary allocation. As the careful reader will observe, there is a nonlinear relationship between time and cost. The abscissa shows milestones to be achieved within 25, 50, 75, and 90% of the assigned time:

- If the first 25% of the project's time is applied research, the way to bet is that costs will progress slowly
- Costs will pick-up somewhat in the next 25% of project time which is CAD-based co-design work
- By the time of prototyping at 75% of schedule the cost curve is steepening, and it steepens even more with series production.

If the design review indicates that the project heads toward an unacceptable quality level, has fallen behind in its time schedule, overrun its budget, or presents other major ills, it should be stopped there and then. It is better to do away with a project when about 25% of the budget is spent, than throw away double that time and the whole budget by waiting till the completion of something that will be short of its targets.

- Projects which do not perform to exacting standards must be stopped cold in their tracks, and
- The persons leading such projects, as well as their supervisor(s), must be made to feel that they are in trouble in terms of their career because they failed to deliver on their mission.

When he was executive vice president of Bankers Trust (a mighty bank in its time) and among other responsibilities he run the institutions' information technology, Dr. Carmine Vona had a policy of major DRs of software in which he himself was present. If he saw that the project's manager had not performed as expected, rather than letting the project drift he will kill it. But *if* this project's functionality was vital to the bank's competitiveness, *then* he will start it again (practically from scratch, to avoid existing mistakes) with a new project manager.

Senior executives can no longer take the attitude that quality of deliverables, costs, and timetables are problems that can be left to run at their pace. This traditional approach, which treats project managers as all-knowing and capable single-handedly of solving project problems without the need for supervisory control is unsound.

Competitive forces, if nothing else, make of project success a contractual obligation which requires that product assurance and a few other key factors are key design parameter pursued through an established well-managed program. Hence the vital need for predicting, measuring, and suggesting means of improving the timely results of developmental designs.

A methodology of DRs must be developed in a way which fits best a company's business, its engineering projects and the requirements of the market to which it appeals—since the product or process under development will be for the enduser. An important part of this methodology is the metrics which will be used. With each DR, the compliance to product specs, quality assurance, and timetables must be expressed as a score.

- Some organizations do so by variables rating perceived quality (at DR level), adherence to deadlines, and so on through rating
- Others value each critical element by attribute on a *go/no go* basis based on written opinion of the evaluator(s).

Senior management, too, should give its input. Well-managed companies see to it that projects and their deliverables are absolutely visible to top management and subject to internal control. Another advice, based on experience, is not to start at the same time too many projects. Some of the companies I was consultant to the board had over 50 running in parallel.

That is how to spend time, money, and manpower for doing nothing. Therefore I used all my power of conviction to change this inefficient research policy. No single person—and projects should be controlled by one senior executive to avoid infighting—can follow 50 projects at the same time. He might do so for half a dozen to a dozen, while the others will be loafing. Neither is it a good idea to institute a hierarchy project management.

- The more intermediate steps there are, the greater the political plots which interfere with sound management, and
- Therefore, the less effective these projects are going to be; as well as the more inconsistent will be the outcome of DRs.

I was consultant to the CEO of General Electric-Bull, the American-French computer company, right after GE got control of Bull. The CEO asked for a technical audit of all projects which had been running under Bull management. There were 56 of them. The next request was to evaluate each one in respect to:

- Its possible contribution to the combined company, and
- Its strengths and weaknesses as a project.

Twenty-one out of 56 passed the test. Twelve among them, those which had the better prospects, were continued. The others, all others, were disconnected from the company's budget.

As this example demonstrates, the better policy is that of limiting the number of projects running in parallel, but seeing to it that those retained have tip-top human resources, are fast-moving and are properly budgeted. This is as well vital for another reason. Today's technology and markets are, economically speaking nonlinear. Exponential curves are much more difficult to master than linear processes, like those which were the rule right after World War II.

In conclusion, DRs are an important dimension of internal control in engineering. They contribute to higher quality, lower cost, and fast development timetables which give to the company clout in a market more hard driving and more demanding than ever. By running faster one can overtake a competitor or near competitor who might be way ahead. Look at Apple and Nokia. That is what global competition is all about.

3.5 The Clash Between Evolution and Standardization

Moore's Law holds that microprocessors double in computing power every 18 months. The effects of Moore's Law have ranged widely as microprocessors spread from computers to telecom equipment, appliances, motor vehicles, and other machines, making the fortune of companies which followed a smart machine strategy—while opening the grave of the laggards.

Take mobile phones as an example. The first generations were purely devices for conversation and text messages. Success depended on designing desirable handsets, manufacturing them cheaply, and selling them to the global market. As microprocessors become increasingly powerful, mobile phones changed into handheld computers, with large part of their value in software and data services.

This quantum leap in functionality had a profound business aftermath. Apple and Google, which knew how to build widely reaching technological platforms,

gained the upper ground. Market-wise the bill has been paid by Motorola, Nokia, and other companies who took time to rethink and revamp their product line.

As far as loss of speed is concerned, at least in marketing drive, an interesting case is that of Nokia which has been working on a touch-screen phone not too different from Apple's iPhone as early as 2004. But turning a predominantly hardware-maker into a provider of software and services has not been easy, and Nokia lost the initiative. Something similar has happened with Cisco in the market for networking equipment.

The lesson to learn from these examples is that business success capitalizing on engineering design is highly dependent on versatility and the ability to turn on a dime. This evidently works against standardization, bringing into perspective the need for a tough decision on whether *standardization* or *evolution*—along with it *change* is the better strategy in today's global industry.

The examples in this chapter presented to the reader weigh on the side of flexibility. It is however proper to notice that with the exception of one-shot products for special engineering and weapons systems (see in Sect. 3.1 Dr. Lusser's list) standardization has merits. Moreover, a novel device can have standard components used by its predecessor in the company's product line.

Communications protocols and basic software are two issues quite sensitive to the question of change (often nick-named "upgrade"), because they affect so many online devices at the same time. Still, because standardization versus flexibility is an important argument, In this section we will see several reasons on why change becomes unavoidable.

Let us start with a historical perspective which allows to look at the concept behind standardization from its early beginning. Efforts directed to the establishment of standards date back to the Nineteenth Century. One of the first cases has been that of the British Post Office which standardized the size of the envelop for a two pennies stamp. In engineering credit for standardization should go to Henry Ford who:

- Was a visionary designer, and
- Perceived early on that standardized components will allow to radically reduce the cost of producing an automobile.

From design to manufacturing Ford's *Model T* excelled in economy and efficiency by means of standardization. Its industrial production could be replicated and scaled up without significant additional costs, provided established standards were observed. In the issuing decades other companies, too, including Ford's immediate competitors, adopted his methods, and industrial organizations became the epicenter of:

- Efficiency studies and
- Standardization practices.

Indeed, looking back in industrial history the contribution of standardization can be hardly overestimated, but this does not mean that there is no downside. The price for religiously observing established standards is that the company may have to forego using new, more efficient components, and technologies. This is particularly true in two cases:

- In an environment of rapid technological evolution and
- When the people deciding on the standard do not really see both sides of the coin affecting their decision.

One of the best examples of ignorance in standards setting was the early 1980s decision by the Council of Ministers in Japan which adopted X.25 as the exclusive communications standard for the Japanese industry. This first packet-switching standard by CCITT, X.25, was good but ephemeral. It was overtaken by a rich array of packet-switching communications protocols. Obliging the whole Japanese industry to abide to it, was tantamount to condemning it to forego new developments.

Opponents of universal or even national standards say that when an industrial sector, a product, or a process is new and in a course of rapid evolution, the role of designers as well as of managers is to push, provoke, propose, and keep the door open to new developments. But there are constraints, such as the need to provide for compatibility.

Look at Intel's microprocessor in the 1970s. Customers loved it or hated it, but they bought it. It was certainly not dull but it was soon overtaken by technology. Intel, however, had to stick to it and it did so not by using to its advantage the solution of "upwards" through the process of upwards compatibility—so that the software which was already developed did not become massively obsolete.

At the same time the big question on the lips of Intel's clients was: "What is next?" Other microprocessor producers gave these same clients different options which reinvented the concept of rapid obsolescence (in style if not in substance) originally developed by Alfred Sloane of General Motors. This time around the focal point was greater integration is microprocessor design. Intel reacted through an agreement with Microsoft. *Wintel* allowed it to have an upcoming software company as an ally.

The notion derived from this example is that internal standardization over a period of time is important, but to become effective it requires a policy of *change control*. This applies not only to design and fabrication processes, but as well as to tests, inspections, and customer handholding. Specifications and technical documentation of standards—even of "standards" which evolve and change overtime - should cover in detail all critical items in detail to:

- Provide a fundamental description of each item, its functionality, and product assurance, and
- Clearly provide critical reviews of changes to existing designs, including stress testing and other inspection requirements.

As Intel's example documents, a company's method of controlling changes and implementing a sound market strategy are critical elements of success. In no instance should changes in standards, indeed in designs, be made without proper study of "pros" and "cons", and without company-wide coordination.

No company can afford to stay still. Take as an example the fairly recent development of *additive manufacturing* (AM) which permits modes and forms not previously possible with traditional manufacturing techniques. To a substantial

degree it removes the restrictions of mechanical tooling which demanded to assure that parts such as castings can be removed from a mold.

Today's technology enables to profit from new design freedoms, such as amalgamating assemblies of components into one part in one instance; saving weight and reducing footprint. Just like 30 years ago CAD/CAM altered the way we looked at engineering design today AM permits to get into the industry's consciousness the huge amount of embodied waste in conventional machining, thereby optimizing material usage. Companies which stay put with casting standards would not have the best future in front of them.

Paying attention to the effect of changes in standards in processes and products already in the market, is by no means an option. It is a vital policy which should cover the entire installed base at customer premises. Mechanical, electrical, and electronic components are cogs in a bigger machine, and engineering must make sure that these cogs are interchangeable without the bigger machine having to be thrown away outright.

- The best concept is not to restrain design changes and options,
- It is to define the most viable choices that meet both technical and marketing objectives balancing benefits against cost, including the cost of being overtaken by competition.

In conclusion, whether in product design or in manufacturing changes to old established standards present challenges which have to be met. To become effective such changes require research and scientific experimentation paying attention to the quality of the resulting process or product. Often, though not always, such experimentation involves (inter alia) a search for little known or so far unknown factors: Finding their origin and cause, examining their behavior, trying to comprehend their likely effect(s), and using them to gain competitive advantage.

References

1. Lusser R (1957) Unreliability of electronics—cause and cure. Redstone Arsenal, Huntsville
2. Packard D (1995) The HP way: how bill hewlett and I built our company. Harper Business, New York

Part II
Service Quality

Chapter 4
Service Assurance: A Case Study

4.1 Alternative Energy Technology and Service Assurance

The specific theme of this case study is the use of technology for environmentally friendly energy production. This is a rather theoretical statement. Practically, technology assisted in the development of more efficient or unconventional energy resources. An example is advanced engineering solutions that provided a clearer picture of the oil fields by means of three-dimensional seismic imaging, which led to breakthroughs in deepwater exploration. It also permitted better extraction from oil reservoirs through multi directional drilling.

At the same time, however, it is no less true that high tech and better living standards have significantly increased energy consumption and this depleted energy resources and/or polluted the environment. Both are negatives, as there are reasons to believe that major leaps forward in oil exploration are past.

While oil companies spend billions of dollars looking for new and so far unknown natural oil resources, they fail to admit that exploration is suffering. Some experts fear that the last big discoveries have already been made and scientific attention should switch to better understanding of the fields already found, which will improve their exploitation.[1]

The way to bet is that in the years to come, together with wringing the last few drops from existing reservoirs, a great deal of attention would be paid to new energy sources. This has already started with renewables like solar and wind, but according to my findings, the results are far from being stellar [1]. The hope is that plenty of effort will also focus on energy conservation. At least that is what one likes to think, but if past experience is used to project into the future, energy preservation is far from being a safe bet.

Those believing that energy conservation and alternative energies will gain the upper ground say that since the late 1980s there has been a significant increase in

[1] For example, in seismic technology geophysicists try to remotely sense in real-time the motion of the fluids using both acoustic and electromagnetic mapping. This combination improves the exploitation of the natural reservoir.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering,
DOI: 10.1007/978-1-4471-2966-0_4, © Springer-Verlag London 2013

environmental awareness, with consumers developing a clear desire to buy green and manufacturers being ready to oblige. This is, at least, what transpires in their advertising campaigns where they promote the greenness of their goals.

Critics answer that behind advertising campaigns are by no means technological breakthroughs; all this is only propaganda while some technological developments are being abandoned because of environmental reasons. Such is the case of CFCs which have been punching holes in the ozone layer. Ben Gurion had once said that two things which are negative do not make one that is positive. Here are the two negatives:

- Public attitudes show a basic disenchantment with materials polluting the environment, and
- Environmental claims about alternative energies and some new technologies are failing to pass real-life tests.

CFCs are a case in point. Virtually all scientists praised them, but some years down the line they agreed that they were responsible for damaging the ozone layer. No laboratory came up with, so to speak, a truly benign alternative. CFCs were simply banned; let us however not forget that they were originally considered among the chemical industry's safest products. Quality assurance has been on leave.

There are plenty of other examples with claims made by known companies which prove to be smoke and mirrors. In 1983, BP claimed its unleaded "Supergreen" petrol caused no pollution to the environment; a short time thereafter, it was accused that Supergreen contained several toxic compounds.

Over the last 30 years, big firms spending a lot "on the environment" were also revealed to be major polluters of it. They had simply jumped on the bandwagon of greenery, or acted out of ignorance that words alone do not make a product environmentally friendly:

- The agent supposed to do that miracle has been technology,
- But technology has not yet conquered the secret of how to do away with pollution (if it ever will).

Bio is a label that sells, but quality control is wanting and therefore everyone uses this term the way it best fits their interests. Not long ago in France, it was found that while a company which marketed bio-products respected the rule of not using fertilizers, the land in which salads were cultivated and cows ate grass was right next to a nuclear power plant.

No wonder therefore that the public is becoming increasingly skeptical of any organizations that get on the green bandwagon, particularly those that have little direct contact with the environment. In addition, while consumers may want to accept social responsibility for environmental protection, few will forego quality in the products they buy.

The person who buys something is looking for —*product assurance*—and this is what reputable firms' technology should deliver. This is perfectly true of

cleantech, bio's alter ego. At New York Stock Exchange, there is even a "Cleantech Index" which correlates with the S&P 500. Energy cleantech includes:

- Biofuels
- Geothermal
- Solar
- Storage
- Waste
- Water, and
- Wind

Applications cover transportation (vehicles, biofuels, batteries); heating (private houses, apartment buildings, office buildings, factories); biology (particularly for clean water); efficiency (which is somewhat of a catch-up category) and other domains. Years ago solar and wind were the main drivers moving in tandem with a 0.95 correlation.

More recently, solar fell from grace and wind is following the same trajectory as heavy state subsidies have been cut back. All of a sudden there is downsizing of investments in cleantech sectors, just like in the late 1990s and first years of this century the public interest in "greener and greener" waned. Critics say there has been something wrong with the substance of the alternative energies movement and its technological underpinning; this saw to it that the rising curve of its popularity bent. What went wrong is not difficult to ascertain:

- Alternative energy, like bio, suffers from a lack of *product assurance* (Chap. 1), and
- It has so far provided no evidence that in the future *service assurance* is at the top of its priorities—the theme of this section.

Service assurance is *quality in the large* in the service industry. In information technology (IT), for example, user organizations ask for availability statistics (Chap. 7). Mean life of man-made devices and systems is closely linked to system design, system integration, and maintainability. What counts is a continuous uninterrupted service at high quality level.

There is a very significant difference between off-the-shelf packaged products, like those with a bio or any other label, and alternative energy installations. To the end user alternative energy, for instance heat pumps (HP), does not come off-the-shelf *it is a service* which must be steady, cost/effective, and high quality.

- That is the sense of *service assurance*, and
- *Risk factors* associated to services, as contrasted to those of stand-alone products, are not being paid the attention they deserve.

A houseowner may buy bio milk to drink or bio cheese to eat, but if he buys a heat pump which burns an inordinate amount of electricity, fails to integrate with the existing heating installation and heats the house in a miserable way (Sects. 4.3 and 4.4) he will regret it 100 times. And he will say so publicly. He may as well initiate a legal procedure to get rid of the damned machine.

Moreover, service assurance may be adversely impacted by changes in government policies and regulations. Energy production and distribution services are designed according to regulations that continually change. Quite often, such changes affect the extent and type of benefits and/or of obligations befalling energy users.

- Added requirements increase the cost of providing alternative energy services.
- Reduction in subsidies have the effect of discouraging investments in alternative energy, and
- Absence of proper training of plumbers and other professionals who install and deliver alternative energy equipment has proved fatal to service assurance.

Governments promoting alternative energy, equipment vendors, and professional associations of companies involved in alternative energy installations fail to appreciate that services assurance is severely damaged by the total absence of education on how to do a first class job. As we will see in this chapter, this has an adverse impact on the vendor's reputation and negative effects on his business.

Another risk confronting the providers and "*installateurs*"[2] of alternative energy machines and services is that they are not able to keep pace with changes in technology. This is counterproductive. To maintain a growth strategy, they must adapt and respond to technological advances and market requirements—which is not necessarily the case. Instead, they rely on strong arm tactics (Sect. 4.2). As we will see in this chapter, Stiebel-Eltron (S-E), the company on whose poor *product and service assurance* on which the following case study is based,

- Let its machines and their control units become outdated, and
- Hurried to cover this failure by another one: mistreating the client and lying to him.

There can be no assurance that a firm will be successful just by being ahead of the curve in product design. A great lot depends on the quality of its management. But *if* it does not upgrade in a timely manner its machines and *if* it does not steadily strive to improve its service assurance, *then* it will fail to provide quality results and this will have an adverse material effect on its business.

4.2 Reinventing the I.T.'s Strong Arm Tactics

There are two domains in which cleantech has promised wonders, but so far deliverables do not justify the public and private investments which have been made in the hope of environmentally friendly breakthroughs. The one is

[2] Generally speaking, these parties are plumbers, but do not need to be only plumbers. The French and German word "installateur" are more appropriate. Therefore, they have been chosen in this text as the appropriate term to this profession.

transportation with the reinvented 100-year-old electric car. The other is home heating through heat pumps, the subject of this and the following sections.

Take as an example, a three-story family house which from the time it was built was heated with oil, to the tune of 7,000 l per year. In 2010, being environment friendly, the new owner wanted to replace the oil with solar and geothermal power. Because the two wells might have been near to a clean water catchment area, the administrative authorities objected to the wells and the geothermal solution was replaced by two heat pumps (HP).

The vendor was S-E. That is where the problems started. To appreciate what has been involved in customer disservice; and evident dissatisfaction, it is wise to look back to the 1960s—IBM's high water mark in the computer industry. Then with computers and now with heat pumps, three issues curiously correlate.

1. Trained specialists able to understand the product and initiate the necessary preparatory steps, are a rare species.

Two half-century-old images due to absence of homework stick to my mind. The first dates back to early 1959 at the data processing center of the National Bank of Greece, in Athens. The bank's executives proudly showed me their computer installation where three IBM 650s were still wrapped up for protection from dust. The answer I got when I asked why these computers were not being used has been: "We did not know computers have to be programmed. It takes time to do that." Two years later, in 1961, I got the same reply from executives of the National Iranian Oil Company in the Abadan refinery, when I observed that four IBM 650s were still wrapped up.

To its credit, IBM attacked the problem of education in system analysis head-on when in the late 1950s, first in the US then in Europe, it created the Applied Science department. Neither S-E nor its competitors, or for that matter government authorities promoting alternative energy, have started a similar effort to educate plumbers and other *installateurs* on the intricacies of alternative energy systems.

2. Selling machines which are not ready for integrated heating systems.

This amounts to reinventing the "broken down cash register" by NCR, a marketing gimmick of the 1910s which periodically returns to life. For instance, in the 1960s with IBM's scientific computer (a *coup* designed to overtake Control Data's leadership), and quite recently with heating pumps which burn lots of electricity with low returns (more on this in Sect. 4.3).

3. Strong arm tactics by salesmen, designed to intimidate the client into submission to the vendor's wrong way policies.

IBM did that by going over the head of the IT manager, in case the latter objected to some of its practises. The computer vendor organized a person-to-person meeting between the CEO of its local subsidiary and the client company's CEO. The latter rarely, if ever, knew anything about IT so IBM's senior executive had an easy time in reverting an unfavorable (to his company) decision by the IT manager—whom he ridiculed in the eyes of the client company's CEO.

In S-E's case there was no client company CEO to trap through the same unethical policy, so an S-E executive wrote to other plumbers an insulting letter to the client to intimidate them from intervening to correct the technical errors of its installation. Here is an extract from that letter:

> Despite instruction and as he so often likes to mention himself possessing a large technical knowledge, Mr. … has unfortunately not understood that the installation is flexible and therefore very efficiently regulates according to outside temperature and only goes to higher temperatures in case there is a warm water request.

Apart from being insulting to the company's client, this statement is as well a big lie. As we will see later on in this case study, the S-E installation was absolutely inflexible. The two heat pumps either worked together or not at all. Worse yet, because of wrong engineering of their controls when they were working, they also activated the oil burners ending up to a failure in terms of both *cost* and *effectiveness*.

By all evidence, the vendor wrote this letter to build up his defenses because he knew very well that his equipment was improperly integrated in the heating system of the private house. The plumber who made the installation said that all he did was to follow the instructions and plan given to him by S-E, which is shown in Fig. 4.1.

Allegedly, this was a cookie-cutter plan made to fit all installations independently of their specific requirements, which is evidently wrong because no two installations need exactly the same solution. The result was an unmitigated disaster.

- The *No. 1 Problem* has been *regulation* of the two heat pumps and oil heater into a smoothly working system.

As it has been revealed postmortem through technical auditing, the S-E machines were hardwired to their control unit, hence impossible to optimize on an individual basis. Incorrectly and without informing its client, S-E also hardwired to its inflexible control unit the existing oil heater, destroying the latter's independent control unit. As a result, all energy sources worked together all of the time with the heat pumps burning 5,000 kWh/month and still unable to heat the house (we will return to this issue).

- The *No. 2 Problem* was the regulation of the heat pumps themselves.

At first sight it was indeed very curious that the two HPs started working at the same moment together no matter how high or low was the outside temperature. The reason for having installed two machines was to use the second one for peak energy demand. Their simultaneous start/stop made optimization virtually impossible.

- The *No. 3 Problem* has been that S-E oversold its boilers. The two of them were in parallel sharing 1,600 l a quantity that was too big for the private house.

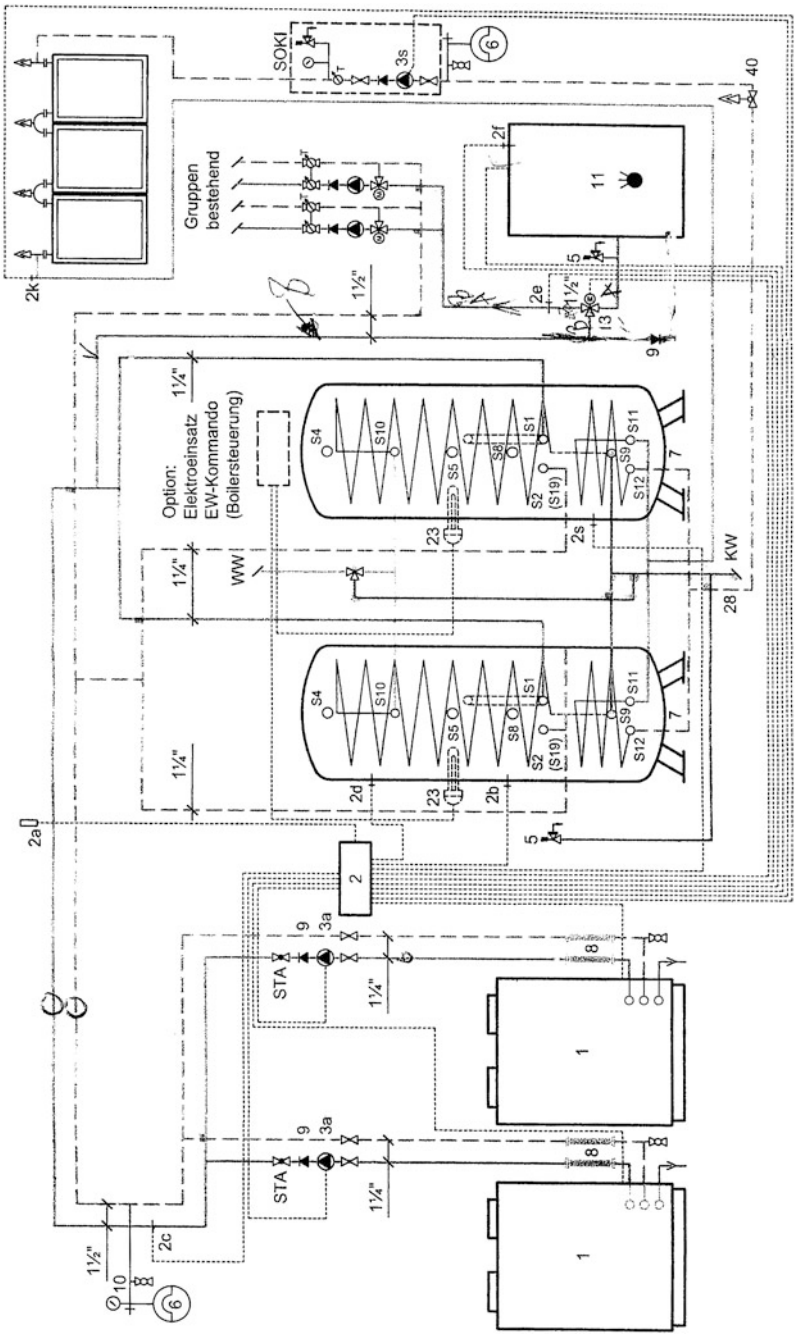

Fig. 4.1 Steibel-Elton instructions and plan

This could be partly corrected by putting them in series with a switch to close out of the circuit one of them, but S-E objected in writing. Its letter said that any change to the installation in its client's house—and therefore in its client's property—will automatically result in ending all maintenance on its behalf.

What this written statement said was illegal. The vendor was contractually obliged to provide maintenance for 2 years, and there was no escape clause in the contract. But in a way similar to that of selling computers in the 1960s the company resulted in scare tactics. This is a not-too-ethical though a rather classical IT marketing tool which S-E used.

- The *No. 4 Problem* was much system inertia. The hot water temperature increased and decreased too slowly, which explains why the heating system was not sensitive to changes in the outside temperature.

Problem Nos. 3 and 4 correlate among themselves. At 1,600 l the mass of water was too large and unnecessary for the house needs. They also correlate to problem Nos.1 and 2 regarding lack of appropriate regulation of the two heat pumps at their level, as well as at system level with the oil burner.

As in all engineering problems, system level integration and optimization required a flexible approach to the control of all heat sources through software. But as it has been already brought to the reader's attention the heat pumps' controller was hardwired. With its "knocked down cash register" policy the vendor had sold machines with characteristics they did not have (more on this in Sect. 4.6).

The absence of the necessary basic technologic features was made worse by the fact that the plumber who acted as *installateur* had no experience with the integration of three energy sources: oil burner, heat pumps, and solar. An expert in alternative energy was therefore asked to solve the Gordian knot of overall failure, with the mission to:

- Proceed with data analysis of the prevailing heating conditions, and
- Come up with a concept on how the current imbalances and heating problems can be corrected.

Without this bolts-and-nuts approach to a problem which had several unknowns and encountered a significant number of uncertainties, there could be no solution worth talking about. The vendor's and *installateur's* attitude in this case raised also legal questions which are themselves subject to inherent uncertainties, including the possibility that their ultimate resolution could have a material adverse effect on the financial position and reputation of the vendor.

To the first important query: Has the cost/effectiveness of alternative energy been attractive? The answer is: There exists no evidence of any savings, but there is plenty of evidence of troubles. To the second important query: Did the vendor care about customer satisfaction, or did he only want to push a machine down the customer's throat? The answer is the second option as documented by the vendor's strong arm tactics. (These answers will be further documented in the following sections.)

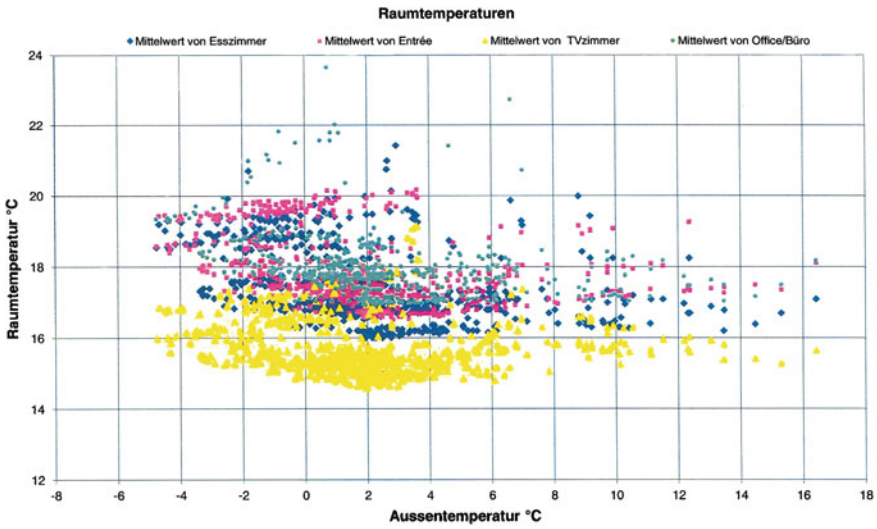

Fig. 4.2 Temperature sensors recording temperature in four rooms: dining, entrance, TV, and office

Let me conclude with an advice to vendors who think of emulating such unwise policies instead of concentrating on service assurance and of doing whatever it takes to guarantee their customer's satisfaction. Hard sales lead to customer dissatisfaction, including disputes or potential disputes related to breach of contract, breach of fiduciary duty, performance-related claims, and other highly negative issues which damage the vendor's profitability and reputation.

4.3 Unfulfilled Promises of Heat Pumps

To uncover the pattern of the heating problem, the independent expert started with data collection. He targeted the way the heat pumps worked and the internal temperature which was delivered. Fig. 4.2 presents the temperatures, the sensors registered in four rooms: dining, entrance, TV, and office for an outside temperature varying between −5°C and slightly over +16°C.

The target temperature which the installation, including heat pumps and oil burner should have delivered was 18–19°C. This is what the owner of the house (*the owner*) had asked on the premise that nothing walks a straight line—and it is an important reference in regard to expected energy consumption. As the reader can see in Fig. 4.2,

- The most heated room was the entrance, up to 20°C.
- The least heated was the TV room, with its temperature varying from 14.5°C to a little over 16°C.

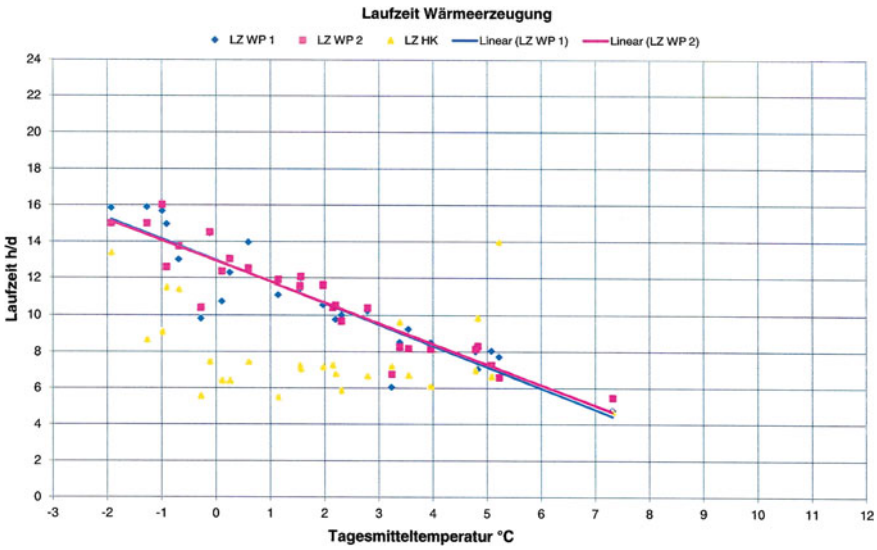

Fig. 4.3 Heat generation period for two heat pumps

This talks volumes of imbalances as well as of low efficiency of the heat pumps keeping in mind that they consumed 5,000 kWh/month and over and above that was the oil burner's consumption. The fact that, most incorrectly the two heat pumps worked in unison is shown in Fig. 4.3.

The blunder is even greater if one keeps in mind that an internal temperature targeted at 18–19°C is favorable condition to a heat pump installation, because HPs work better with lower temperatures. What *the owner* expects to receive in return for costs incurred with heat pumps is greater efficiency in energy usage. In the typical installation:

- 1/3 of energy being delivered comes from electricity, and
- 2/3 of delivered energy comes from external air

This 1/3 ratio of power consumption was far from being the case with the S-E installation where the denominator was 1.4 or less. Therefore the aforementioned ratio was reduced to below $\frac{1}{1.4}$; a pitiful result.

The local power company has a preferential tariff for HP with current provided 18 h/day. Hence, with an outside temperature between −2 and 0°C the heat pumps worked almost continuously while the oil burner also run. But the most striking feature in Fig. 4.3 is that the two heat pumps worked together all of the time, no matter which was the outside temperature.

Shown in this figure, the vendor's sales manager and his technician said that "their company's heat pumps were designed to work always together." The answer he got has been that this is most evidently is a lie. *If* it were true, *then* these machines were very badly designed. In reality the real reason was a deficiency in their control and this was indeed proven some time later on.

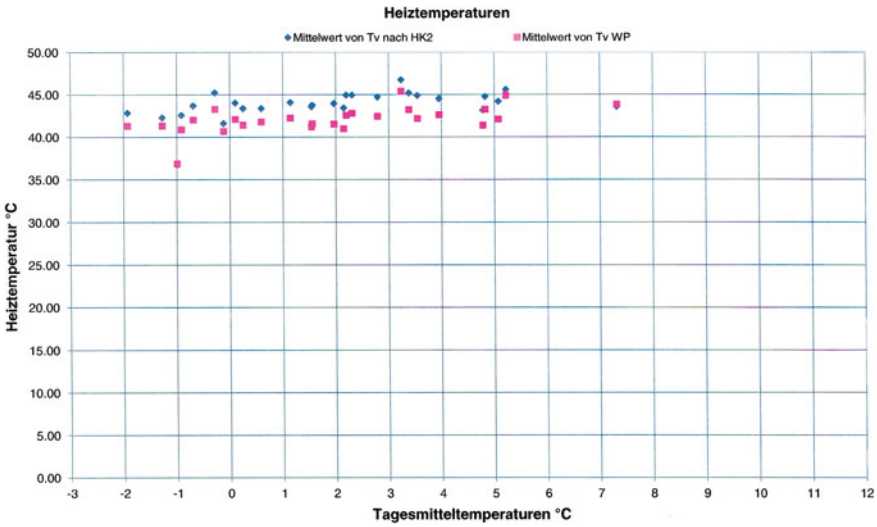

Fig. 4.4 The input heating temperatures from S-E heat pumps was 44°C to 45°C and the output was 42°C to 43°C—leaving less than 2°C difference for the consumption of 5,000 kW per month

Figure 4.4 gives a hint in regard to why the inside house temperatures were so low. What the heat pumps delivered by consuming 5,000 kWh/month was 40–45°C depending on the outside temperature. Because the oil burner's control was destroyed by the S-E's technicians and its functionality made subservient to the HP, its contribution was less than 3°C. The whole installation badly underperformed.

All this has been very upsetting to *the owner* because machines cost money and when they are incapable to deliver the whole idea of using them turns on its head. Which was the party responsible for this heating disaster and unwarranted investment? The machine vendor or the plumber who did the installation? The answer to this query is complex because, in a way very similar to what has happened with IT there were not two but three parties:

- The client,
- The machine vendor, and
- The plumber-*installateur*.

In IT the "plumber" is typically the software provider. In the case of alternative energy the client's contractual relation has been with the plumber, and not with S-E. But because the heating installation has never been completed by the plumber, in spite of successive deadlines fixed by legally required registered letters, that contract had come to an end.

Therefore, when maintenance problems developed with the heat pumps of S-E underperforming, the client had to inform the vendor through registered letter that

an alarm signal by his equipment had to be taken seriously. The meeting held subsequent to this communication provided some more characteristic aspects of the "knocked down cash register".

Stiebel-Eltron's sales manager and his technician were provided with the evidence by the electricity company that in the first 3 months of 2011 their company's heat pumps had burned 5,000 kWh/month. They were also given the evidence that during April, May, and up to mid-June 2011 the S-E HP burned another 7,800 kWh—or over 3,000 kWh/month, at a time when the outside temperature fluctuated between 11 and 29°C.

The S-E representatives were further informed that a heating system study, made by an independent alternative energy consultancy, fully documented that the installation was a failure—and this was not the only bad news as far as S-E's responsibilities were concerned.

The sales manager's excuse has been that the installation plan S-E has given to the plumber was "only a suggestion". But this was not receivable. Not only nowhere in this plan which was printed under the S-E label, was written "suggestion", but also the plumber had said to *the owner* that all he did was to execute the plan S-E gave to him.

Over and above of that, during his June 27, 2011 meeting with *the owner* the sales manager confirmed in person, through a slippage of his tongue that this was *the installation plan* S-E gives to all *installateurs* to use in their clients' premises. An old proverb says that from the kid and the silly person you learn the truth, but there are also other ways for the truth to come out. Sometimes the truth finds a way to come out.

In trying to correct the mistake which he had made, the sales manager said that after all his company was a manufacturer of equipment and had no responsibility in terms of *service assurance*. To this *the owner* answered that in a highly competitive market like alternative energy companies which shy away from:

- Product assurance, and
- Service assurance

will soon find out that they can sell their equipment only to local friends. A company which negates its responsibilities when troubles develop is one which has taken leave of its senses. That attitude of the vendor's sales executive puts in question the firm's quality of top management (Sect. 4.6).

During that same meeting *the owner* reminded the sales manager and his technician that they knew very well the *installateur* of the heat pumps was not reliable and made mistakes. They themselves were allegedly overheard having said in the course of a lunch in a local restaurant: "How are we going to tell the installateur that this is a mistake…".[3]

[3] If a lie detector test was taking place, this is a question the owner wanted to ask the two persons in reference, and he was looking forward to read the polygraph's results.

All this added up to a very negative reaction and the meeting ended short of any result. As the owner explained to the S-E representatives, such a negative attitude across the line by a vendor is big-headed and fully unacceptable. The vendor is not a dictator, and as an American proverb has it: "The customer is always right."

In conclusion, a company lives and prospers because of the trust and continued patronage of its customers and because of the service assurance which it provides them. A policy of "NO!" answers, bad will as well as insults and threats to the client is followed only by companies which do not respect themselves; therefore they are not able to respect their clients.

4.4 Requirements for a Solution to Improve House Heating

Rather than continuing this dialog of the deaf about the disastrous state of the heat pumps' installation, it would have been more rational to find a solution which corrects the failures and hopefully makes the investment in alternative energy worthwhile. This evidently required changes to the current installation, along with prerequisites and conditions summed up in the following terms:

1. Reestablishment at S-E's care and expense of the oil burner's controls, as they were prior to its intervention of December 2010 which altered (with very negative impact) the existing oil burner installation.

The oil burner had to work under its own control as it used to operate prior to the S-E intervention. After this was done, it had to be integrated into a heating system by means of a programmable thermostat.

2. Installation of a reasonably priced thermostat/controller, dimensioned for a private house not for an apartment building or factory.[4] This unit should integrate the oil burner's controls and the HP controls into one, well-functioning heating system.

The need for a programmable thermostat was evident since the beginning and was asked by *the owner* since the time of the negotiations with the different providers of heat pumps and the installateur. The offers he got, however, failed to provide it and S-E choosing instead to manipulate and destroy the controls of the oil burner.

3. Installation of an input/output temperature register, from and to the S-E HPs, to provide measurement on how many °C have been gained by the user of the installation, given the high electricity consumption of the heat pumps.

The proper functioning of this input/output registration presupposed that S-E agrees to these measurements as the sensors had to be established online to and

[4] Which was a curious characteristic of some of the proposals being made.

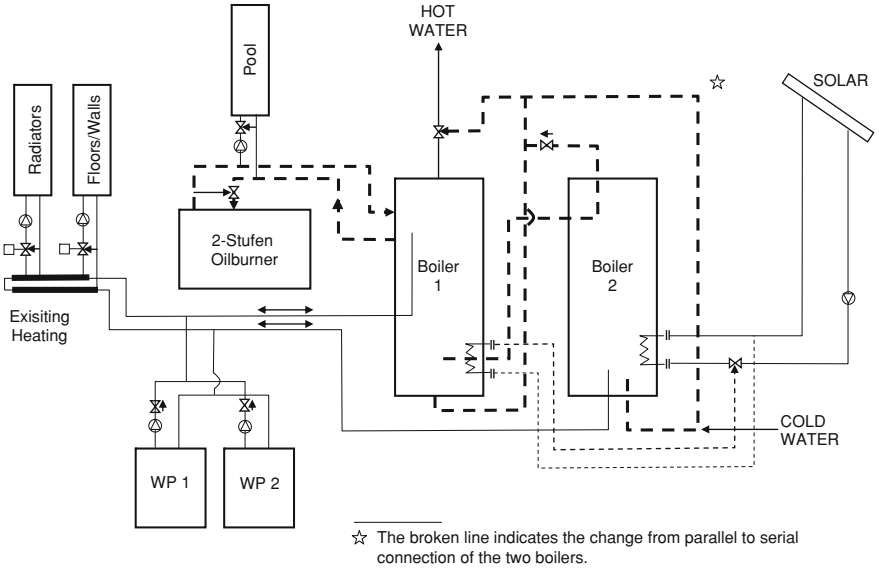

Fig. 4.5 The restructured heating plant

from its HPs, and submits to the owner a factual and documented report on input/output temperatures as well as on improvements to the efficiency ratio (if any).

4. Confirmation that the two boilers which were installed in parallel will be put in series. The general line of this schema is shown in Fig. 4.5.

As it has already been pointed out in Sect. 4.2 the S-E installation had over-dimensioned the boilers. As a demonstration of his goodwill the owner offered to have another company change the boilers connection from parallel to serial, *provided that* in its answer S-E guaranteed for 2 years the maintenance of its equipment, after the aforementioned change had been completed.

5. Regulation of the S-E HP so that only one of them answers the basic heating needs of the private house, while the second serves when peak energy demands present themselves. In addition, the way the HP work had to be effectively coordinated with the oil heater.

The explicit requirement for service assurance is presented in Fig. 4.6. It calls for an outside temperature of °C or higher the heating should be provided by the heat pumps alone. Between 0 and −5°C there should be coordination between the HP and the oil burner. At or below −5°C the oil burner should provide the heating with the HP on standby. The user should have the flexibility to bring this lower limit to −3°C.

This had to be confirmed by a *written* offer documenting the work which had to be done and giving instructions on system management. Quick oral explanations,

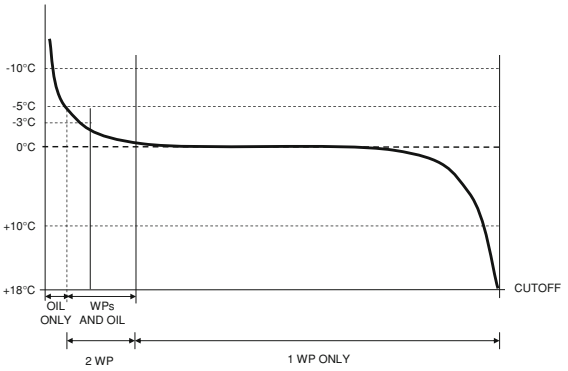

Fig. 4.6 A coordinated operation of heat pumps and oil burner

the policy followed by the S-E representatives over the months of suffering from its equipment's malfunction, were no more acceptable. All explanation of what is done and how it should be managed had to be:

- Clear,
- Comprehensive, and
- Printed on S-E stationary.

"Seat of the pants" offers were not receivable. S-E had in its possession graphs based on measurements by the independent expert. These were provided to its president, Rudolf Sonnemann, by the owner himself in documentation that, with the exception of the hallway, the temperature in the rooms of the house varied between 14.5 and 17°C.

For service assurance purposes, S-E was asked to confirm in writing its acceptance of the independent expert's findings. *The owner* offered the vendor the possibility to conduct an experimental design (Chap. 11) whereby S-E could proceed under the same conditions, with its own technical tests. But he also cautioned the vendor that analysis of past data from other industries shows that the vendors' repeatability is poorer than that of their customers.

The owner reminded the vendor of a principle in business that a company's appeal to the market and its reputation can be negatively affected by its inability to deliver product quality and service assurance. A condition for the experiment was that all defects in the heating installation had to be immediately corrected and a new set of tests made—prior to "ready for use". The vendor refused to follow this course.

Most evidently, *the owner* also asked that S-E excuses itself for the insults to his person written by one of its executives. Also for defamation in the references made to a local company of programmable controllers/thermostats which had called S-E for information on its equipment and was told that it must keep out of this because "*the owner* was not reliable" and anyway another controls firm was already on this project." Stiebel-Eltron executives were old enough to know that:

- In a free market nobody has monopoly, and
- Price fixing by trying to exclude competitors is punishable by the law.

As for the personal insults and defamation, the S-E managers probably never heard of a sound business practice: No client should experience financial loss, harm to reputation, or business interruption because of vendor's faults or bad intentions. If for no other reason, respect for the client is to the vendor's own interest because such events expose the firm to:

- Investigation,
- Litigation,
- Liability,
- Penalties,
- Loss of clients' business,
- Unfavorable impact to the firm's reputation and operations.

It is a silly policy for a company to engage in strong arm tactics which can potentially turn against its own interests. When this happens, it is an evidence of very poor governance which runs afoul of the risks and the uncertainties such a policy involves. It is also the sort of vendor behavior which makes *service assurance* a chimera. The whole company culture has been against it.

4.5 The Vendor Continued to Violate Service Assurance Prerequisites

Stiebel-Eltron was not short of tricks. Borrowing a leaf out of IBM's book of the 1960s, the secretary to the general manager (GM) of its subsidiary called to ask for a meeting. He was ready to visit the owner, the secretary said, and indeed he did so on August 2011 accompanied by his sales manager.

During that meeting the GM started with the statement that the contract *the owner* had signed was with the *installateur* and not with S-E.[5] Therefore, his company had no product assurance responsibility.[6] Then he repeated the cheap statement his sales manager[7] had made a couple of months earlier:

> We only manufacture machines—nothing more than that.

To this the owner responded that while the statement about the plumber's responsibilities was right, the reference "we only manufacture machines" was wrong. A machine manufacturer who does not care about product assurance and

[5] Which legally speaking was true.

[6] Which was a false argument.

[7] Who was present in this meeting.

service assurance—including how well his machines are installed, are operating and are maintained can keep his machines for his museum.

The owner further stated that while the *installateur* was responsible for the quality of his work, saying that Stiebel-Eltron had no responsibility was incorrect. It did not need explaining that:

- Every heating machine attached by the installateur to the system must also be efficient, perform well, and be reliable, and
- Contrary to this, the postmortem study left no doubt that the two S-E HPs underperformed as they delivered only 40–45°C with an outside temperature from −2 to 10°C.

The machines were not up to the required standard, and their controller performed even worse. This was most evidently unacceptable and had to be corrected by S-E. The GM answered that the *installateur* has the responsibility for the functioning of the heating system, adding: "S-E will maintain the installed machines. But nothing more."

The owner answered that apart from the fact that what the plumber did was in the S-E installation diagram, in its letter of December 28, 2010 S-E has clearly stated that *if* any change was made in the heating installation with its equipment, *then* it will no more maintain its machines. Therefore, the GM had to write a letter on continuing maintenance—one which also cancels the S-E December 28, 2010 correspondence emphasizing that maintenance will be discontinued. Such a letter:

- Was orally promised by the GM
- But in a typical S-E fashion it was never sent.

The same violation of the word one has given during a negotiation on service assurance, characterized the second subject discussed during that August 9, 2011 meeting. *The owner* pointed out that the integration of a programmable thermostat/controller for the heating system cannot be performed without the specifications and diagrams of the S-E controller of its HPs. This is a proprietary S-E device on which S-E has given absolutely no information.

The GM answered that S-E will provide all the necessary information as well as detailed documentation, so that the company installing the leading system controller knows what needs to be done in terms of interfacing and connections. This, too, proved to be a fake promise.

As time passed by and neither the detailed specifications nor the maintenance guaranty ever arrived, the owner wrote to the GM of S-E's subsidiary to remind him of his word. In this letter, he also pointed out that the detailed engineering specifications and diagrams, as well as interfaces of the S-E control unit, are absolutely necessary in order to:

- Attach to it the new programmable thermostat/controller, and
- Correcting the disaster all around these machines which had lasted way too long.

To this reminder to the GM of his promises was attached a statement by a control systems specialist who visited the installation and wrote: "The problem

can be solved, if Stiebel releases to us the control on SPS-Saia via pole-free contacts and the hydraulic is adjusted."[8]

It is clear, the owner added in his letter to the GM, that by not being possible to connect the S-E HPs to the control system because of lack of cooperation by S-E, this equipment is useless—and the GM was invited to take back his heat pumps. The GM who had played the dead bug in August suddenly woke up in September 2011. Moreover, in a short letter to *the owner* S-E's GM let it be known that his word had turned to ashes. On the subject of guaranteeing the 2-year legally required maintenance, he said that the general business conditions were included in the letter, but nothing like that *was* included, another trick. Moreover, the GM knew very well that his company's December 22, 2010 letter had to be canceled, and this was not done. Still another trick.

Even more tricky was the GM's answer to the transmission of details of the control unit and specifications of the interface to which the central control of the heating system was to be attached. To this he said, the GM wrote that his lawyers and unidentified other people advised him not to do so. Hence, he was sending nothing. The fact that he had given his word to transmit these documents was totally unimportant to him.

This being the case, he evidently got a reply which started with the fact that as on previous occasions of S-E mail, his last letter's contents were most surprising. On the contrary, the client had thought that after the GM gave his word he would keep it. The contents of his letter proved precisely the opposite. The GM was informed that:

- A serious person not only keeps his promises, but also promises what he knows he can keep.
- "Who told him what" is irrelevant. He gave his word but he failed it.

In this reply was also brought to the GM's memory that as discussed in the course of the August 9, 2011 meeting, the winter of 2010–2011 was very painful to the people living in the house—with temperatures down to 14.5°C because the S-E machines did not work right. It was moreover evident that this could not continue. There had to be an end to this *total lack* of service assurance.

As part of providing services to clients, vendors rely on a number of third-party service providers. These service providers include, but are not limited to plumbers, who have not been trained on how to install alternative energy equipment. Failure by these service providers, to deliver their services:

- Impacts client relations,
- Results in material interruptions to operations, and
- Can lead to significant penalties or liabilities for the firm.

The lesson to retain from this case study is that *service assurance* which should accompany every machine and every system is adversely impacted by the failure

[8] Written on 24.8.2011, 15 days after the aforementioned S-E meeting.

of service providers to perform their functions in an able manner. At the end of the day this reflects most negatively both on the manufacturer and on the *installateur*.

4.6 When Irrelevance Reaches the Top, Service Assurance is Doomed

Chapter 1 has brought to the reader's attention that during the early part of product assurance life-cycle in the design, development, and production phases, considerable effort must be spent in attaining a high level of quality and reliability. As this case study documents, during the implementation phase emphasis must be placed on a service assurance with the aim to attain and support continuous improvement of:

- Operational dependability and
- Client satisfaction.

A vital function of such a service assurance program is operational analysis, based on a scientific approach and fully supported by a rich database. As far as service assurance is concerned, word of mouth is immaterial and company politics play dirty games. What is important is the ability to fundamentally:

- Confirm progress toward attaining service goals,
- Alert on problems which tend to thwart this progress, and
- Solve these problems in a timely manner to improve both client satisfaction and company profitability (the two correlate).

This requires that top management is very sensitive to its responsibilities. Integral part of its planning duties is well-defined procedures for obtaining, measuring, and utilizing service assurance information. There are two basic methods for doing so, and they support each other.

One is laboratory analysis in which cause and effect relating to failure in implementation are critically examined in order to be understood and controlled. The other is field surveillance in which results are observed and systematic procedures are instituted to flash out the cause. The basic steps in using either method are very similar:

- A *technical audit* is done (Chap. 3),
- Service assurance results are *observed* and recorded,
- The observations are studied, with their background identified and *explained*,
- This explanation is *verified* through round-table meetings with clients as well as by means of engineering studies.

Service assurance-wise, it is the senior management's responsibility to sponsor walkthroughs in which vendor and client participate and handhold. This being done, and provided the client is satisfied with the results, the ultimate authority to

decide on the company's side how to right the service assurance balances is its president—since he is personally accountable:

- For management control, and
- For taking corrective action regarding anything happening under his watch.

What has been stated in the previous paragraphs constitutes the kernel of both engineering and market leadership in industrial operations. Some company presidents nevertheless are "too weak", "too busy", or "too uninterested" to look in this way at their responsibilities. Warming an armchair is for them more important than taking action.

Supposing that the president of S-E was the sort of manager who takes care of service assurance failures and corrects them, following the tandem of problems described in Sects. 4.2–4.4,[9] *the owner* asked some common friends to make an introduction to Rudolf Sonnemann, president of S-E. The first contact was through a telephone conversation in which the S-E president was informed of the service assurance failures regarding:

- His company's produce, and
- His subsidiary's illogical and totally negative attitude.

The S-E president was also informed that an independent consultancy in implementation of heat pumps had taken measurements which documented beyond doubt the malfunctioning. *The owner* came forward in offering to the S-E CEO copies of the graphs based on the measurements (see Figs. 4.2, 4.3, 4.4), and it was agreed that S-E's vice-president/engineering will examine them and come up with a concrete service assurance proposal.

The CEO was as well informed that, most improperly, S-E and the heat pump *installateur* destroyed the oil heater's independent control unit, already in place and well-functioning, by making the oil heater subservient to S-E's heat pump control.[10] The CEO was also made aware that the heavy duty heat pumps themselves were poorly regulated as they worked in parallel, while in the original client order one of them was intended to be used as:

- Standby for reliability, and/or
- Employed only in peak heating for very low temperatures.

In his letter, *the owner* further expressed the hope that all these product assurance problems will be corrected without delay, and that this case becomes a first class study on service assurance to be published and honor the CEO's company—rather than a case in court.

---

[9] The problems described in Sect. 4.5 followed the failure of the company's president to intervene.

[10] As the careful reader will remember, the *installateur* said that this was what S-E told him to do, and in fact the destruction was done by still another company to which this destructive action was subcontracted by S-E and the *installateur*.

What happened next is that time passed with no response from S-E's CEO. The company's president was away for 3 weeks with nobody to take his place. Even the executive secretary did not answer the phone, using instead a recording which said that the executive secretary to the boss simply was not there. Only in the afternoon another secretary would answer the phone simply saying the president was absent.

The Null Hypothesis, H_0, was that the S-E CEO was an absentee manager. The alternative hypothesis, H_1, was that he was hiding. The fact has been that he was never found. This led *the owner* to write another letter, which apart from for asking for a long overdue reply to his first letter informed the S-E CEO that in the meantime the local power utility send a bill for electric power consumption at a cost higher than:

- If the oil burner was working non-stop, and
- The S-E heat pumps were never installed.

The S-E president was told in no uncertain terms that *if* the provision of service assurance was not possible, *then* as an alternative the client offered his company to take back its equipment. In the meantime, while the hideaway was still on, a technical audit made by two different independent automatic control companies found the reason for this procrastination.

The controller of the two S-E heat pumps was hardwired; therefore of a medieval concept. Hence, the two machines *had to* work together and also pulled along the oil heater—at triple expense to the user. Stiebel-Eltron sold an equipment that it did not have, falsely presenting its controller as an optimizer (a fact also confirmed by the insulting paragraph of its letter quoted in Sect. 4.2).

When with an inexcusable delay the S-E president came alive, all he said was that his people in the subsidiary told him that everything was alright. "We have thoroughly controlled your information," he said in an e-mail, "with the appropriate sales- and service colleagues.[11] Thereby we found that the S-E machines work correctly. The technical documentation[12] proves that our machines deliver the necessary power."

That is precisely the point: The CEO was incapable of expecting the unexpected. Not only his installation did not deliver the necessary product assurance but also his underlings had their way. The CEO also wrote in his e-mail: "At the time we had advised you to involve a specialist planer. You did not take this suggestion into consideration." That is double talk because, as the reader has already seen,

- S-E had given the *installateur* the implementation document (Fig. 4.1), and
- The installateur applied what S-E had told him to do through this cookie-cutter "plan".

[11] Precisely with those who had created the disaster.

[12] !!!.

The owner answered that he was sorry to see the CEO had been so incorrectly informed by his subordinates. The installation of S-E heat pumps had been a *disaster*. A very bad reference for his company. The answer to the CEO concluded that it is a pity he missed the opportunity to do a thorough audit of the wheeling and dealings of his subsidiary. By believing "what he was told", he did not exercise his prerogatives as chief executive and failed to apply management control. By default the company under his watch was run under a very curious principle: "If a plan is not working, stick to it."

Reference

1. Chorafas DN (2011) Energy, natural resources and business competitiveness in the EU. Gower, London

Chapter 5
Man-Made Catastrophes are the Result of Wanting Service Assurance

5.1 What the Nuclear Plants Damage at Fukushima Meant to Japan and to the World

A vast social, economic, and material damage was created on March 11, 2011 by the 8.9° Richter scale earthquake and the 10-m tsunami which followed it.[1] The aftermath was a disaster for the Daiichi nuclear reactors at Fukushima which were part of a system designed and built to answer Japan's growing need for electric power. Nuclear energy had provided 30% of Japan's power needs. As 2011 came to a close, two-thirds of the reactors were still idle.

Following a long time of meddling through with the catastrophe, the Japanese government fell. Shortly after taking over, Yoshihiko Noda the new prime minister, said his country should cut its reliance on nuclear energy over the longer run. He did not say, however, from where will come the power that a highly industrialized country requires. That is the challenge every western country and many of the developing economies are currently facing.

In "*Energy, Natural Resources and Business Competitiveness in the EU*" [1]. I have outlined the options, and each one of them is far from being brilliant. In a nutshell, the alternatives are gas, oil, hydro, coal, nuclear, wind, solar, and geothermal. Every one of them has advantages and drawbacks. None can by itself accommodate today's demand for a colossal amount of energy which continues to grow by leaps and bounds. The human drama aside, what the nuclear plants damage at Fukushima meant to Japan and the world is that it put squarely on the table the question of energy supplies and their sources. Unfortunately no government, including the Japanese, came forward to ask for a thorough study which can:

- Put on the table in a factual, documented, and objective manner the options and

[1] A day after the earthquake and tsunami Emperor Akihito gave a televised address to express his deep concern. By contrast, the Japanese government and the TEPCO energy company, owner of the nuclear plant, took a long time to put their acts together.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering, DOI: 10.1007/978-1-4471-2966-0_5, © Springer-Verlag London 2013

- Thereafter lead to very difficult and far-reaching decisions with huge impact on the coming generations.

The nuclear plants' damage at Fukushima also brought to attention that with atomic energy, as in so many other domains *service assurance* and the discipline it implies have been lousy. Management by Tokyo Electric Power (TEPCO), the owner and operator of the Daiichi nuclear reactors, has been very imprecise about the accident and its aftereffects (Sect. 5.3) and this ended up in a huge amount of confusion, with staff taking different views about what to do in an emergency. That is what happens when, rather than being regarded as the Number One prerequisite no attention is paid to service assurance.

The fact that TEPCO's top brass was not held accountable for very poor management this has been bad enough, but punishing the whole of mankind because a company, an industry, or a profession has been careless, is even worse. Obliging the economy to pay exorbitant prices for energy is like choosing the worse possible alternative leading to:

- Unemployment and
- Lower standard of living for all (more on this later).

The answer to the query on how to handle the developing energy crisis is in no way self-evident—let alone easy. Coal pollutes and nuclear energy has Fukushima aftereffects—but neither are some of these alternatives pollution free, as the different self-appointed environmentalists want the common citizen to believe.

Hydroelectric dams disrupt the life cycle of other animals in the biosphere. In addition, when they become gigantic, like Assuan, they have a great deal of unexpected consequences which is why many experts now suggest that the Assuan dam should be demolished. Wind power has been opposed in court by nearby living residents for creating migraine, apart from the fact that its cost/effectiveness is nothing to crow about. Solar power can have a very limited coverage of today's big population centers (though it is fine for private houses). The heat pump service assurance disaster has been documented in Chap. 4.

This should be by no means interpreted as a call to ban altogether all alternative energies. They should be used always in a prudent way. But they are not massive solutions. Precisely, the same principle applies to nuclear energy. Whoever says that nuclear plants should be outright banned without a study of the pluses and minuses of *all* power production means, has simply got it wrong.

There is plenty that can be said about risk-associated to nuclear energy, particularly in terms of radioactive waste, but nothing is risk-free. The accident at Fukushima has been a case of extreme mismanagement (Sect. 5.2). In addition, the conception behind the thermal part of nuclear plants is a copycat of coal plants, and therefore obsolete. Yet nuclear energy remains an alternative as it produces more than just electricity for the general grid. It is used for:

- Fresh water by desalination of seawater and
- Hydrogen production as an alternative intended to displace oil and fuel.

Expected technological breakthroughs like nuclear fusion are not yet onboard, and by all evidence it will take some more decades till we have an answer as to whether they are feasible. Nuclear fusion would be an interesting alternative because, unlike fission, it does not produce much in the way of radioactive waste. But so far laboratory models have not been able to run for long enough to turn out a meaningful amount of electricity.

The first attempt to do so was *Zeta*, a machine constructed in Britain in the 1950s. The effect was short-lived. This was followed by the Joint European Torus whose power production lasted for a matter of seconds.

Currently, hopes are pinned on research done by the International Thermonuclear Experimental Reactor (ITER) that can hold 840 m^3 of hot, gaseous fuel. The projected cost is euro 15 billion ($22 billion) and, as with nearly all research projects, the deliverables will be either big or nothing.

When the time comes we shall see what can be expected from fusion, but as things now stand it will be silly to bet the world's energy supply needs on it. In addition, oil experts project that during the coming decades oil and gas production cannot be maintained up to about 100 million barrels per day, this would require either:

- An even greater commitment to nuclear energy, made in a rush, hence very risky or
- *Carte blanche* to extract oil from tar sands, oil shales, very deep ocean drilling, and other unconventional sources which involve many unknowns, plenty of pollution, and other great risks.

A person able to think of consequences would come to the conclusion that we cannot continue the business-as-usual energy policy of starts and stops characterizing the last four decades. A bending of the rising curve of energy sources will lead the world into economic collapse. We have to assure adequate energy supplies for the coming generations, or they will be inheriting a rotten world.

Figure 5.1 shows the three spikes in energy costs to worldwide nominal GDP. The energy has been coal, nuclear, hydro, oil, and gas from over five decades, 1971–2011. These are the top five energy sources. The 1970s experienced two oil crises and in between them stagnation. Extremely high energy costs have also characterized this century and most particularly the last 5 years—another epoch of stagnation at least in the West.

It can be reasonably argued that the world economy cannot afford a cost of energy equal to 9% or more of global GDP. This is an unaffordably high ratio. It is also a prognostication of a meager economy. Energy costs representing 9% of global GDP are a level attained right before the double-dip recession in the early 1980s and then again in the early 2010s.

Are deep recessions our objective? If not, we have to be careful, very careful, about how we choose our energy supplies. As Fig. 5.1 shows, nuclear power production represents only 2% of the cost of energy supplies. Hydroelectric power, too, is very competitive; and so is coal. By contrast, by far the higher energy cost is that of oil, which also exhibits sharp pick-ups in price volatility rewarding the oil producers and penalizing everybody else.

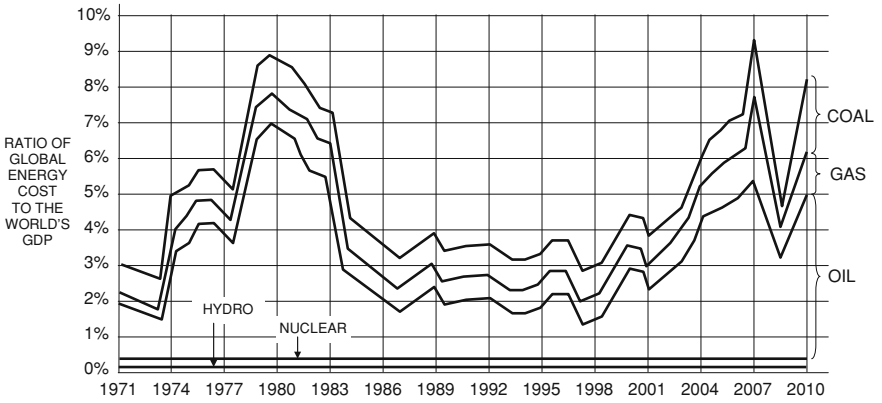

Fig. 5.1 The global economy cannot afford a 9% ratio of energy cost

5.2 At the Daiichi Plants Service Assurance was on Leave

On March 11, 2011, nature struck back and in economic terms the aftereffect is expected to cost well over $100 billion. According to some estimates it will be much higher. Goldman Sachs thinks that it will reach or exceed $200 billion. After the tsunami hit, factories closed all over the world's third largest industrial power.[2] Toyota decided to shut all of its plants to assess the damage and preserve human lives. It took a couple of months till the auto manufacturer decided to open all of them again.

There has been an interesting comparison between the effects of earthquake and tsunami which a few years earlier (in 2004), had hit Sumatra, Indonesia (reaching at far out places like Sri Lanka)—and what happened in Japan. Because the level of technology and industrialization in the Fukushima[3] area was so much higher, the damage was exponentially greater.

This is a fact to be kept in perspective whenever we talk of *service assurance*. Loss of life apart, 6 million homes were initially cut off power supply and this rose to more than 30 million in the evening of March 12, 2011 as Tokyo itself, 250 km south of Fukushima, was in the dark. The problems which first hit two nuclear

[2] This had much to do with earthquake and meltdown. On October 20, 2011 severe floods in Thailand had a similar effect in interrupting the world's supply chain.

[3] While the Fukushima prefecture is the one most often mentioned as having suffered greatly from earthquake and tsunami, the damage extended way up its north into the Miyagi and Innate prefectures. In fact, the earthquake's epicenter has been near the border of the latter two prefectures but Fukushima was more industrialized; and it was also where the Daiichi nuclear power plants were located.

reactors at Fukushima led to the evacuation of a 20-km radius around the factory. This involved 170,000 people. It should also be remembered that this 20-km range was criticized as being too narrow. A 50-km radius would have been better.

Experts were uncertain about the extent of the damage. Some said that what happened looked more like the Three Miles Island, the 1979 nuclear accident in the US. Others answered that in Three Miles Island the problem came from inside while in Fukushima it originated in the outside–and its many unknowns brought a host of stability problems into the picture. At first glance, however, no expert likened it to Chernobyl in the Ukraine in 1986. This comparison surfaced later on as the damage, and the peril from it, continued to deepen.

Japan is on its way to overcoming the challenge of the Daiichi nuclear plants, but at high cost. The cost is far from being irrelevant because following the excesses of the 1970s and 1980s which led to two decades of no growth in the economy, Japan has become the most indebted nation in the world as far as sovereign debt to GDP is concerned. The most basic lesson for sovereigns, companies, and families is not to overextend themselves; when adversity hits one has to have reserves in order to recover.

It is in no way unreasonable that when the crisis hit these worries about accumulated liabilities were eclipsed by the meltdown in the damaged nuclear power plant. Initially, some nuclear experts said the potential danger to human health from the stricken reactors was blown out of proportion, especially when set against the wider spread suffering of the tsunami victims. But there has been a growing sense that Japan brought upon itself this nuclear nightmare because neither the government nor TEPCO was in charge.

According to Dr. Vijay Dhir, dean of UCLA Engineering and a nuclear engineer, service assurance at Daiichi was at its worst not only in the sense of preventing a catastrophe but also after adversity hit. Dhir identifies five major failures in crisis management[4]:

1. The Japanese government and TEPCO did not immediately call for help from the better known nuclear experts around the world. Instead they tried to minimize the accident.
2. TEPCO failed to bring in General Electric, the Daiichi reactors builder. It should have called GE right away and shared with it the responsibility.
3. Following earthquake and tsunami the salient problem was to avoid core meltdown. With the proper effort core meltdown could have been avoided; but service assurance failed in its mission.
4. TEPCO did not immediately bring in emergency replacement generators, as those previously installed were unwisely located near the seashore and swept by the tsunami.
5. The nuclear fuel was stored in an outside pool, which was inadequately protected. This, too, was a severe service assurance failure.

[4] Notes from a meeting we had on October 2, 2011 in Entlebuch, Switzerland.

Taken together these five failures compounded the damage to the nuclear plants from earthquake and tsunami—and led to the much wider catastrophe. The most important duty facing TEPCO's management when adversity hit was self-discipline. It did not do so and this should have brought board members and top executives to court.

Of course, service assurance does not come cost-free, but would you like to try the Daiichi alternative? The cost of a well-studied and rigorous service assurance system at the Fukushima nuclear plants would have been a small portion of the $200 billion or more of destroyed equity because of the sudden great disaster. And this is perfectly true of *all* nuclear plants—as well as of so many others—around the globe.

To be effective, in nuclear power production service assurance requires foresight, insight, and discipline. But is this not true of any other field of human activity? For instance, the function of disciplined movement in battle is to produce:

- In the mind of the foe the belief that he cannot win and
- In the mind of the friend the conviction that he cannot lose.

The No. 1 flaw is indiscipline and mismanagement. We can win the battle for energy production but this requires a great amount of preparation and immense discipline. *If* we do not want, or cannot do so, *then* the alternative is repetitions of the Fukushima Daiichi catastrophe no matter which is the energy source being used.[5]

Precisely because service assurance at the Daiichi nuclear plants has been absent, both Tokyo Electric and the Japanese government were under siege. Critics said that the safety conditions were substandard (more on this in Sect. 5. 3); the pros answered that the safety record has been good (!!), considering the number of plants and their length of service. Length of service, however, correlates with reliability (Chaps. 6, 7), and there exists no evidence that reliability was studied in this context. Instead there is a record of:

- Cover-ups,
- Poor crisis management, and
- An inbred complicity between Japanese regulators and the country's utilities.[6]

Daiichi's 40-year-old nuclear reactors, which faced an unmitigated disaster, should have been taken out of action years ago, particularly as there have been engineering design failures—and their continuing existence compounded the dismal physical event.

Other failures were of a business nature, and they can be found in every country. Accounting for building and operating costs of a nuclear plant, is not enough. It costs however more money and takes more time to decommission a nuclear plant than to build a new one.

---

[5] Just look at BP's sorrows in the Gulf of Mexico.

[6] This is more or less true all over the world.

A mare's nest of failures gave the Japanese public good reason for skepticism about what may come next in the unfolding nuclear crisis. Such a sense of growing worry was also fed by memories from the atomic bombs dropped on Hiroshima and Nagasaki near the end of World War II, and the devastation they had created.

By contrast, many residents among the 35 million of greater Tokyo, the most populous metropolitan area in the world, have shown stoicism as long as the Japanese authorities and media remained calm. This stoicism is most commendable as a notable event, because in parallel to it were:

- Radiation fears,
- Power cuts,
- Train disruptions, and
- Aftershock warnings.

The nuclear accident was also exacerbated by a tandem of public relations failures. With its shares and bonds hammered, Tokyo Electric faced worries about its credit worthiness. Japanese companies selling nuclear energy technology abroad: Toshiba, Hitachi, and Mitsubishi Heavy Industries, were also uncertain about their business future. Was this the *coup de grace* on an industry which had taken decades to establish itself?[7]

This uncertainty has been the aftereffect of another failure by nuclear plant vendors, because they have learned nothing of earlier accidents. Fukushima Daiichi was not the first to be paralyzed by an earthquake among Japanese nuclear plants. It was however the first to be laid low first by what was conceived as the technology's dependence on a ready supply of water for cooling, and then by what was revealed as grand design mistakes associated to:

- Service assurance,
- Technical requirements, and
- The bad luck to find itself in the path of a tsunami.

In respect to technical failures, the six nuclear reactors of Daiichi were near the seashore because of the need for water, but plant designers had failed in projecting what is known as *a defense in depth* which is part of service assurance. Available evidence suggests that different defenses were drawn up haphazardly.

Fukushima's nuclear plant defense consisted of simple barriers unable to hold anything else than small waves of the Pacific. Worse yet, the independent generators—supposed to serve as emergency services in a catastrophe—were located between the nuclear plants and the shore. (And as already mentioned, TEPCO failed to bring new emergency generators in less than 24h.)

The nuclear fuel was stored in an outside pool—another major mistake. The reactor core the fuel rods make up, and the water it sits in, were contained within a steel pressure vessel. That sat within a larger steel structure (primary containment

[7] On March 14 and 15, 2011 shares on the Tokyo exchange fell by over 16%, their worst two-day fall since 1987. Curiously enough, the yen strengthened to record highs against the dollar.

vessel) encased in a steel and concrete secondary containment structure. Reportedly, there were no extraordinary precautions given to the nuclear plant's earthquake exposure and close proximity to the Pacific Ocean's whims.

In conclusion, what the Fukushima Daiichi nuclear plants drama means to Japan and to the world is how easy technologic progress can turn into a man-made disaster (Sect. 5.4). By all evidence, not much thought was given to the fact that in real life it often takes almost forever for things to happen but, when they do,

- They take place quickly, and
- They can have devastating effects.

In every man-made system there are risks which, when taken lightly, can show up instantaneously with a "bang". They are human failures which repeat themselves through the centuries. The "bang" has not been an exception reserved to the Daiichi nuclear plant; its current currency in many man-made systems which have not been under steady watch and turned into an unmitigated disasters. As US Supreme Court Justice Felix Frankfurter once said: "Key to (the) problem is men [2]."

5.3 A Shot-by-Shot Account of the Perfect Storm

After the 2004 tsunami which hit Indonesia and other countries in the South Pacific and Indian Ocean, Vienna-based IAE Commission informed all countries with nuclear energy plants that they should be installing emergency aggregates. At the Fukushima Daiichi nuclear plants Tokyo Electric followed IAE's advice but incorrectly installed the emergency aggregates between the plants which were (unwisely) very near the ocean line and the shore. No wonder that they were wiped out by the 10-m high tsunami.

Nuclear safety experts also debit TEPCO with the failure to immediately bring to Fukushima new emergency aggregates (Sect. 5.2). This was not only terrible crisis management but also a violation of the rule regarding the first action to be taken in the case of such a natural disaster: emergency shutdown.

Emergency shutdown is achieved by thrusting control rods, that sit below the reactor in its pressure vessel, up into the reactor's core. Their expected effect is to soak up the neutrons that mediate the chain reaction which produces most of a reactor's energy, shutting it down.

Shutdown, of course, is easier said than done. Straight after shutdown, other reactions continue for a while and things can get hot pretty quickly if there is no way of keeping it cool. In connection to nuclear power plants, that is probably the most important rule in service assurance. But after the tsunami hit, Fukushima Daiichi lost its cool.

- The fact that emergency generators were out of action led to despair,
- The flooding of electrical switching equipment added to the paralysis, and
- Attempts to get the cooling system working with batteries and generators from elsewhere, failed).

In the end, a natural and a man-made catastrophe (Sect. 5.4) reinforced one another creating the perfect storm. Regretfully, this has also set a very bad precedent on how to manage emergencies in nuclear power plants. Adding to mismanagement, different policies were followed for handling the developing crisis reactor-by-reactor.

Reactor 1 was shut down after the quake. The cooling failure led to partial melting of core; vapor vented; and the building was damaged by hydrogen explosions. Seawater was pumped in an attempt to emulate the cooling procedure. *Reactor 2* was also shut down after the quake. Another cooling failure; fuel rods were fully exposed; vapor vented; and the building housing the reactor was damaged by blast at *Reactor 3*.

Reactor 3 was also shut down after the quake. Still another cooling failure; partial melting of core; vapor vented; seawater was pumped in; the building was damaged by explosion; there was damage to reactor containment vessel; and high-level radiation was measured nearby. Critics said that some of the emergency procedures were hardly worth that name.

In *Reactor 1* and *Reactor 3*, steam contaminated with radioactive elements was allowed out of the pressure vessel and into the larger containment vessel that surrounds it. There, hydrogen in the released gas met a spark and exploded, blowing off the roofs of the buildings but without, it seemed at the time, damaging the containment systems. *Reactor 3* pressure vessel was then treated the same way as, eventually, was *Reactor 2*.

At some point, according to the company's reports, *Reactor 2* boiled dry. Because of an explosion it suffered damage to the donut at the bottom of the containment vessel. At that point in time, *Reactor 2* posed the greatest threat. (Later it was admitted that the containment was also compromised at *Reactor 3*.)

One of the reasons the salvage efforts suffered serious setbacks is that the risks were not sized up *a monte* in the first place. *Reactor 3* required the venting of radioactive gases. It contained highly toxic fuel that included reclaimed plutonium. A problem so complex should have been the subject of scenario analyses and engineering studies well before calamity hits. This was not the case.

Reactor 4 was under maintenance when the quake struck. There was a fire possibly caused by hydrogen explosion at the pool holding spent fuel rods; no water was poured in to cool the pool; there was an abnormal temperature rise in spent fuel storage pool; the water level itself was a matter of uncertainty; and fire was observed.

Reactor 5 and *Reactor 6* were under maintenance when quake struck. At the start of this catastrophic Daiichi nuclear scenario, these two reactors' temperatures were slightly rising in spent fuel storage pools. Eventually these reactors, too, joined the perfect storm.

With information about the disaster released by TEPCO piecemeal and with considerable delay, experts made their own hypotheses about the best and worst cases, as well as the outcome of each of them. With time, the result has been nearer to the more pessimistic projections, as it happens so often with man-made catastrophes (Sect. 5.4).

The fact that TEPCO engineering was wanting—and (to put it mildly) TEPCO management was not in charge of the disaster under its watch—in no way means that the people who worked under perilous conditions exposing themselves to radiation did not show zeal and dedication. They did. The faults which led to a man-made calamity had accumulated over many years:

- From the original grand design decision which put the nuclear plants at sea level near the shore,
- To the allegedly questionable maintenance and lack of a detailed emergency plan on *what* to do and *how* if a natural catastrophe hit.

According to several accounts, a worst-case scenario which is the alter ego of projecting for service assurance, had never seen light. Arguably, Tokyo Electric Power had built the Fukushima Daiichi nuclear power plant to withstand a powerful earthquake, but not one as big as the 9.0-magnitude quake that struck on March 11, 2011. Yet, this was in the realm of possibilities. As for defenses against a tsunami, they were practically non-existent.

Nobody can reasonably state that the probability of such a major calamity was unforeseeable, or that it defied the understanding of exposure engineers and managers should have. Least of all Tokyo Electric could make that argument as in 2007 it had just escaped a disaster at its large Kashiwazaki-Kariwa nuclear power plant, on the opposite side of Japan's main island. Let us recall that:

- Kashiwazaki-Kariwa was damaged by a 6.8-magnitude earthquake and
- That level on the Richter scale was three times as large as what the plant had been designed to withstand. Now think of Daiichi and its six nuclear reactors.

The bigger question regarding this man-made catastrophe has been whether electric power companies, on which society has come to place so much trust, really understand the outsized risks taken with mammoth plants—nuclear or otherwise. It is a mistake to think that all risks are exclusively connected to nuclear energy. I am not at all sure that those who parade down the street against the nuclear understand and appreciate the mega risks involved with all man-made mega systems.

- *If* they do so, and that is an IF,
- *Then* they are totally irresponsible.

Technology being what it is, not all risks are properly identified and appreciated. Worse yet, risk control is way behind technological development. When confronted with total absence of service assurance, the best-case scenario would have been that radiation leaks from damaged parts of the reactor remain small. This, however, would have required a brigade of well-trained troubleshooters.

There was none. Spent fuel pools were refilled by water cannons on trucks, and seawater pumping covered partly melted reactor fuel rods, in a desperate attempt to keep them cool. Eventually it was restored to the plant, and the fires went out leading to the hope that the local area is spared major radiation damage.

But other risks remained. Part of the worst-case scenario was that water in spent fuel pools drained or boiled away. Radiation spikes, fires, or leaks from damaged

parts of the reactor halted seawater pumping; and partly melted fuel rods slumped to bottom of reactor containment vessel. Then, they melted through.

All this does not make happy reading and equally upsetting is the fact that, so far, nobody has been brought to justice for having failed in service assurance. Neither is there evidence of a TEPCO study on how to try to reign over fires from spent fuel rods which pushed long-lived radioactive elements into the air and contaminated an area up to 50 km from plant.

Over the days and weeks following the perfect storm at Fukushima, good news and bad news interleaved. But the former were more or less hopes; while the latter were facts. Some days, the salvage operations appeared to make moderate progress in stabilizing some of the nuclear reactors. On other days, there were:

- New signs of reactor problems and
- Growing radioactive contamination in agricultural produce and livestock.

Contamination from radiation was a particularly feared fallout, and it appeared to be spreading. The government said it was barring all shipments of milk from Fukushima Prefecture and of spinach from Ibaraki Prefecture. The same *communiqué* gave false hopes that there was progress at two of the six damaged reactors considered to be under control in cold shutdown, after engineers finally restored the emergency water pumps.

In conclusion, the most important message this section brought to the reader's attention is that the potential mega risk of a meltdown, triggered by a natural catastrophe and augmented by human failures, was never really appreciated by the power company's top management; or by the Japanese government and its inspectors. There has been a widespread impression of everything being "normal" till the worst continued to worsen.

5.4 Natural Catastrophes and Those Man-Made

History books say that an ancient Greek sage was asking the Gods for three powers: The power to change the things he can, that of accepting the things he cannot change, and the wisdom to know the difference. Among the things we cannot change, though we can take measures to protect against them, are *natural catastrophes.*

The term natural catastrophe means an event caused by the forces of nature, which typically result in a large number of losses. The effect of these losses can be subdued by building up defenses well before adversity hits. One of the most important is that of increasing the effectiveness of disaster control. Natural catastrophes are:

- Floods,
- Storms,
- Hail,

- Hurricanes,
- Earthquakes,
- Tsunamis,
- Droughts,
- Forest fires,
- Heat and frost waves, and
- Other events due to natural reasons.

In 2010, in terms of total damage to society Asia was the hardest hit region. China and Pakistan experienced very high rainfall during the summer, resulting in unprecedented floods which affected much of each country. In China, an estimated 230 million people suffered the consequences, 15 million of whom became homeless.

In Pakistan, flash floods and massive landslides added to the overall damage to dwellings and infrastructure in affected areas. One-fifth of the country's agricultural land was affected, severely impairing the livelihood of over 20 million people—the greater natural disaster in the country's history. In 2011, the worst calamities were again in the Asia-Pacific region, starting with severe floods in northeastern Australia and the aforementioned earthquake and tsunami in Japan.

Natural catastrophes have been happening since the beginning of history and the same is true of those man-made. But while both tend to increase in frequency and impact, man-made catastrophes are ahead of the curve. To err is human. This however does not mean that errors and their aftereffects should be left to drift till they become unprecedented events Fukushima Daiichi style.

Man-made, or technical disasters are major happenings associated with human activities, as well as with inaction when confronted by an impending physical or other man-made disaster. The most glaring example is the population explosion which brought 7 billion people on spaceship Earth, and is now heading for an unaffordable, unsustainable, only half-employed but poorly educated 9 million people.[8] Wars, civil wars, terrorism, and other war-like events are examples of man-made catastrophes, and so are:

- Major fires,
- Explosions,
- Aviation disasters,
- Shipping disasters,
- Rail disasters,
- Mining accidents,
- Collapse of buildings and bridges,
- Nuclear and other plant explosions,
- Major medical errors,
- Uncontrollable birth rates,

[8] The mother of all man-made catastrophes for which nobody really bothers to do anything to stop it.

- Improper use or overuse of limited natural resources,
- Disasters whose origin lies in faulty system design or incomplete integration, and
- Calamities due to wanting product quality and absence of service assurance.

In their majority, man-made catastrophes do not happen by accident. In his book "*Pilot Error*", about the airplane accident of Air France flight from Rio de Janeiro to Paris, Jean-Pierre Otelli writes, the pilots today are no real pilots. They only learn how to manipulate the knobs in the cockpit, since so much is automated while the classical pilot training has taken a leave.

Earlier on, a pilot had not only to undergo 150 h of flight training but he also had to pass tough exams on his ability to be in control of adversities, such as restoring the airframe's stability. Included in the pilot's examination were simulations of tough atmospheric conditions and equipment failures. Today, the shortage of pilots on one side and the automation of many pilot functions on the other, have done away with the old rules, opening the gates to man-made disasters.

As with natural catastrophes, those due to man-made reasons may result in loss of life and are usually accompanied by financial losses in connection to machines, vehicles, buildings, infrastructure, and other areas of activity indispensable to a civilized society. One of the visible risks is business interruption. It is a direct consequence of property damage. The non-delivery of important services may have:

- Economic consequences, like reputational risk,
- Non-economic aftereffects like impaired quality of life, and
- A long list of liabilities toward third parties suing for damages.

Man-made disasters may as well be pharaonic projects which have started with good intentions but with scant attention paid to their aftereffect. At the end of May 2011 China acknowledged "urgent problems" afflicting the Three Gorges Dam, the flagship of its hydroelectric power initiatives. China's State Council noted severe damage to:

- The surrounding region's water supply and geology and
- The life of the 1.3 million people displaced by its construction.

As many people fail to understand, nuclear power is by no means the only devil in a massive electricity production aimed to satisfy demands for greater and greater supply. The reference to the Three Gorges Dam does not mean that hydro is not a good alternative; it is. But as with nuclear, coal, oil, and wind, it requires profound forward-looking studies in:

- Environmental factors and
- Associated mega risks.

Another big dam, Sanmenxia, on the Yellow River was China's pride until its reservoir silted up. Assuan had the same aftermath and Three Gorges may be the next to fall from grace. Already on June 7, 2011 *Shanghai Daily* called it "that monstrous damming project" [3].

Switzerland has so far been a good example of balanced energy supply. It derives some 60% of its electricity from hydro sources, and another 30% from nuclear. If one closes down the nuclear plants, then both consumers and the industry will be strangled because the Swiss hydroelectric potential has already been exploited. Wind and solar will not fill the gap. The only alternative will be coal and its CO_2.

When thinking of unwanted aftereffects one should not be misguided into thinking that power production is the only domain of man-made calamities. Health services, too, particularly those provided "for free" by the State Supermarket are by no means immune to catastrophes or near-catastrophes[9]—most of them man-made.

According to *The Guardian*, Dr Edward Chandarahan, an obstetrician and expert in the domain of classical errors at St George's hospital, in south London, said that maternity staff was making too many mistakes. About 500 babies die in Britain alone because of midwives and similar errors, while an unknown number of others suffer brain damage, such as cerebral palsy [4]..

The irony is that man-made catastrophes turn into a lawyer's paradise. Figures, collated by the Litigation Authority of the British National Health Service (NHS) have shown that the cost of damages paid out by the NHS in 2006, which amounted to £11.8 million in 2006, rose to £85.8 million in 2010. While some of this cost explosion could be accounted for by increasing costs generally and in particular, the cost of caring for a catastrophically injured child, the size of the increase begs the question of whether more mistakes are being made.[10]

Let us face it. We live at a time when not only are there so many more people on Earth, thereby increasing the likelihood of man-made disasters, but also people care less and less about the work they are doing–in case, of course, that they have a job. The same is true of companies as the Fukushima serious nuclear accident documents.

Population explosion, water shortages, and the insatiable quest for electric power have the potential to become the greatest ever man-made catastrophes. Problems connected to water supplies and power generation correlate. In 2003, France was forced to shut down 58 nuclear power stations responsible for supplying nearly 90 % of the nation's electricity because of severe water shortages. This leads to the question: Of the three top uses to which water is massively put:

- Drinking and hygiene,
- Agricultural produce, and
- Power generation,

which one is the most precious? The one we cannot do without? And to which water uses shall we place quantitative consumption limits so that the coming generations do not run dry. Catastrophes, both natural and man-made, will always

[9] According to the World Economic Forum, more than 1 billion people live without clean drinking water and nearly 2 million people die every year from inadequate sanitation.

[10] Clinical Negligence. Midwife Mistakes Cost NHS Millions of Pounds, April 15, 2011.

happen. They cannot be legislated away. But the Fukushima nuclear accident would most likely never have happened if all care was taken in service assurance (Sect. 5.5).

There is no *God ex machina*. The thinking that politicians would always come to the rescue of a misbehaving company and cover the errors of its top brass, creates a moral hazard that fosters reckless behavior. J. Edgar Hoover, the legendary boss of FBI, had enough experience and plenty of good judgment when he expressed his thoughts in six words about those who govern: "Politics itself is Public Enemy Number One [2]."

5.5 Service Assurance for Power Production

We are in the middle of an energy crisis, and as it is true with all other crises it is very important not to cut and run. Managing through a crisis, however, requires a study of the underlying causes and of the drivers behind the salient problem. Once this information is on hand, the hallmark of leading is *bold decisions*. One of the most critical decisions political leaders in all countries have to make in this year and the next is from where will come the energy supplies in a century from now. Such a decision should take into full account the rapid increase in use of energy consumption because of:

- Technology and
- Standard of living.

It would be very silly indeed to cut off one of the most vital providers of electric power because the greens parading down the street and the anarchists burning cars say so. Avoiding nuclear power will lead to dark alleys in terms of supplies and to power cuts which will further damage an already depressed economy. At the same time, however, it is not possible to permit business-as-usual and repetition of Three Mile Island, Chernobyl, and Fukushima Daiichi.

Reliability engineering (Chaps. 6, 7) as well as intensive training and tough supervision can provide the methods and tools necessary to avoid man-made catastrophes in energy production in a power-hungry world. If the use of coal is to taper-off (let alone to be phased out) because of CO_2 and of oil resources (by all accounts) are running short, then "something else" is necessary to steadily provide hefty chunks of mankind's rapidly growing electricity hunger. But it is right to say that nuclear power should be subjected to:

- Much better design through quality control and reliability studies,
- Vastly improved maintenance, including ceaseless upgrade of existing plant facilities,
- Frequent inspections and walkthroughs to sustain a high service assurance level,
- Continuous training of personnel with frequent stress tests,[11]

[11] Based on accumulated real-life experience with man-made catastrophes in nuclear plants.

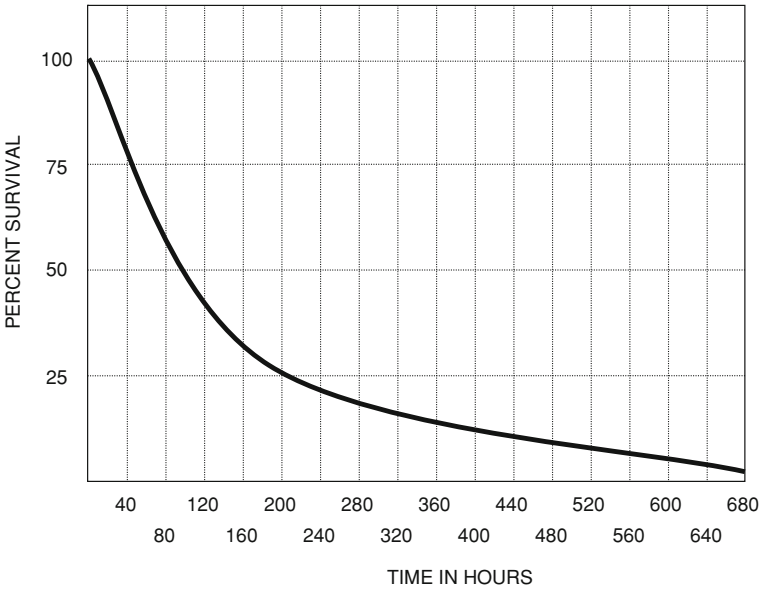

Fig. 5.2 Cumulative distribution of failure-free intervals

- A highly responsible management with severe individual penalties for failures, and
- End-of-life phase out of nuclear plant facilities with service assurance defining the life cycle.

When I say a much better design of nuclear plants I mean both experimental design (Chap. 11) and full-scale application of weapons systems technology for nuclear plant service assurance. Technologists know how to compute cumulative rates of failure-free intervals; an example is given in Fig. 5.1. The dots represent field data from a radar unit. Smooth curve is a theoretic exponential function based on Mean Time Between Failures (MTBF) which has to be carefully studied and controlled.[12]

As we will see in greater detail in Chaps. 6, 7 reliability engineering requires that the life characteristics of critical system components are studied for failure rate. As shown in Fig. 5.2, the prediction life curves must be enriched with confidence intervals equipment which is used repetitively over a long period of time, will take on random-time failure characteristics. When failure rates become approximately constant, that is the equipment's useful life.

To my knowledge, reliability studies are no daily business with nuclear power plants, but the time for business-as-usual is past. The catastrophe at Fukushima has altered the past calculus of nuclear plant safety. Newer models are necessary,

[12] See Chap. 7

having learned a lesson from huge past accidents. The problem is that both energy companies and politicians have given so far no evidence that they learned something and inertia may carry the day.

The real question today is: What can we learn from the different nuclear power debacles? An evident lesson is that nuclear power costs will go well beyond current capital investment and maintenance expenses. Not only service assurance has to be vastly upgraded but also the proper dismantling of the plant should be accounted for. An even more important lesson is that as far as energy supplies are concerned crisis management is not an acceptable option. The fire brigade approach must stop.

Crisis prevention is the keyword, and this is not done by words and pronouncements. At the top of the list are people because with power production the Number one risk is man-made catastrophes. To get out in front of problems and anticipate them we must first and foremost handle people—who are conceivably the weakest link in product assurance.

The persons responsible for power projects and for running power factories must have the foresight to recognize when a small crisis might turn into a major one. This foresight, too, requires intensive training. As we saw in the preceding sections, behind failures are men. Therefore, it is important to understand:

- Their motivation,
- Their strengths, and
- Their weaknesses.

The talk that in the near future power factories will be run only by robots is fantasy. Their most important ingredient will still be men. The sad news is that men, those who should be the keener students, did not learn much from the catastrophe in Fukushima.

On September 15, 2011 Luc Oursel, CEO of Areva one of the foremost nuclear power engineering companies in the world, gave an interview to *Les Echos*, the French financial newspaper. In this, he was more preoccupied with the drying up of new orders for nuclear plants (a projected reduction of 30%), than with what has been learned from Fukushima to avoid similar blow-ups in the future.

Oursel said he is in favor of high security standards, but failed to give any details on what he meant by that. The only example he gave was about the need for an autonomous group to generate power in case of an accident. Fukushima Daiichi had that (Sect. 5.3), but it did not work because it was placed in the most inappropriate location and was overrun by the tsunami.

This answer demonstrated lack of knowledge of basic principles in quality control and reliability engineering. When we develop new products or new systems we look not only at spot values but at ranges that may be the unexpected result of unknown factors. This is shown in Fig. 5.3 taking as an example fixed carbon composition resistor failure rate, from a real life study. Cases involving system design are more complex, but the principle is similar. Neither was this the only dud by the boss (a political appointee).

Asked about the September 12, 2011 accident at Marcoule, a nuclear waste recycling factory in the south of France—where a person was killed and four

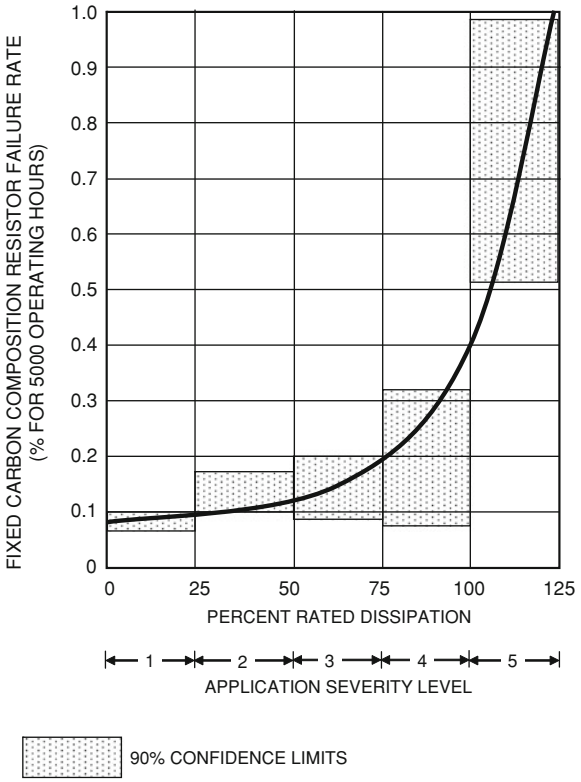

Fig. 5.3 A prediction guide curve. Fixed carbon composition resistor failure rate versus application severity

severely injured—Oursel answered this was an industrial accident and not a nuclear one. Everything happening in a plant is an industrial accident, but:

- Both Fukushima and Marcoule treated nuclear material and
- Both had allegedly a chronic substandard maintenance record.

Since 2000 the Marcoule nuclear recycling plant (which belongs to a subsidiary of EDF) had 18 serious accidents (Level 1). Over that period, there have been successive reports by technical auditors asking for the immediate correction of weakness and faults, but no action was taken. Crisis management had turned on its head, and it became a man-made disaster.

One lesson Oursel should have learned from Fukushima Daiichi but, judging from his interview to *Les Echos*, he did not is that the cost of complex engineering like that of a nuclear plant, is too great to undertake with no supporting reliability program. A well-done reliability program will result in a decrease in the likelihood of accidents as well as of costs—because of an increase in service assurance.

If the attitude of Areva's new CEO is the policy of atomic energy executives, designers, and plant managers, *then* it would seem to me that the next crisis is already programmed. The way an article in the *Economist* had it: "Tokyo Electric Power … (had) sworn blind that their safety records are exemplary and there is no

danger of any meltdowns. This safety mythology has been… tacitly endorsed by the government, media and public at large [5]."

In short, not only the owner of the Fukushima nuclear plant and the government lied, but also they failed to assure disaster preparedness—yet everybody knew that disaster is a possibility even if it may (hopefully) be remote. There is no 100% zero-risk assurance, and if anybody says he can provide it he is lying. However, first class quality control and reliability studies can most significantly reduce the risk of a man-made debacle, and provide for effective damage control if an accident happens.

5.6 Technology and Living Standards Require Greater and Greater Power Supply

If *clean energy*—the uninterrupted supply of power, which for many people in the West has become a sort of second nature—is one of our society's goals, then prior to increase power consumption one has to be sure about the sources of supply. Since the end of World War II the world's economy labored hard to provide the energy necessary to meet growing requirements, first in the developed world and then in developing countries.

Few people would argue about the rapidly growing list of energy demands, but nobody really has a program on how they should be met. By default, everybody looks at fossil fuels as the solution, but can fossil fuels continue to provide the annual increase in energy supplies which for the larger developing countries runs between 6 and 10% per year?

The greens and other anti-nuclear hotheads want all nuclear power plants closed and no new ones built. But they also want to increase energy consumption for greater ease with home appliances, a switch to electric cars, worldwide communications networks, cloud computing [6] and more. This is a highly asymmetric thinking, akin to believing that electric power is produced out of thin air.

Between 2006 and 2010 China doubled its capacity for electricity generation. Between 2010 and 2030 India's electricity consumption is expected to increase 5-fold. It is not only the developing countries which require a massive amount of electricity. In its data centers Google consumes 2.25 billion kwh per year. That much about electric power requirements from the "communicate don't commute" green policies.[13]

For two decades the Internet was heralded as the solution. Today, in terms of energy consumption and environmental damage, it looks as being a problem. In 2000, its worldwide energy consumption stood at about 66 billion kwh per year. In 2010, it has been estimated at about 500 billion kwh. That is a 757% increase roughly divided between servers, cooling, network, and other gear.

[13] The latest craze is sending electronic Christmas cards over the internet because it is more "green"—and never mind the energy being consumed.

Consumers extensively using social network and other Internet services do not think in these terms, but companies do because they have an ever increasing electric power budget for IT. American firms now spend over 10% of their IT budget on electricity consumption and this is expected to rise and equal expenditures for hardware.

There are no quick technologic fixes to the power consumption problem, but neither are people thrifty in their energy consumption. The amount of power consumed by the average European every year is enough to make by a car voyage one and a half times around the earth. Spoilage of energy is unprecedented and no effort, official or by common citizen, is made for conservation in spite of the fact that of the total amount of energy consumed in the European Union:

- Only 42% is produced within its borders.
- The other 58% is being *imported*, making the old Continent a hostage to energy producers.

In terms of tons of oil equivalent, in 2010 America led the world with 2.2 billion—which was matched by China that same year. India and Japan stood at nearly half a billion. Each of Brazil, Canada, France, Germany, and South Korea consumed between 0.25 and 0.30 billion tons of oil equivalents. In electricity generation alone, the global statistics on sources of energy are eye-opening:

- Coal accounts for 40%
- Different renewables[14] 19%
- Nuclear 16%
- Gas 15%
- Oil 10%[15]

Fossil fuels supply a large chunk of the world's energy needs and, as the main driver of China's and India's growth they will remain preeminent for decades. Coal is still running high in satisfying electric power requirements. China is building one coal power plant per week, and this is expected to continue for the next several years.

It does not need explaining that burning coal means lots of CO_2; nevertheless, if you do not like nuclear for massive power production CO_2 from coal plants may be the way to go. If you do not like coal, try oil – but do not forget that conventional oil sources are rapidly depleted.

Rejoice. "Close to 80% of the world's energy supply could be met by renewals by mid-century," says the Intergovernmental Panel on Climatic Change (IPCC). That's not the first time the panel's pronouncements sound hollow. Moreover, conflicts of interest revealed by Steve McIntyre, a blogger, have led to another controversy about IPCC's wheeling and dealing [1].

[14] With hydro power the main component.

[15] Unconventional deep sea exploration promises plenty of repetitions of the 2010 Gulf disaster. This is practically unavoidable because the risks are increasing by so much.

Wheeling and dealing has become commonplace as for the most part, alternative energies are being promoted by state subsidies. One day the taxpayer will be tired of always underwriting the bill of other peoples' follies. For instance, subsidies for the silliest of all plans—that of raising solar energy it in the Sahara and transporting it to Europe; if terrorists allow it of course.

Mid-September 2011 government loans and subsidies to renewable-energy companies came under the spotlight when in the US a congressional committee held a hearing on the fate of Solyndra. Based in Silicon Valley, one of Solyndra's subsidiaries which made solar panels:

- Had an annual business of $100 million,
- But it received a US government loan of $580 million—money which came from the public purse only to file for bankruptcy right after.

Scams apart, so far the contribution of carbon-free electricity credited to alternative energy is modest even after unprecedented annual government subsidies to the tune of $40 billion, according to the International Energy Agency (IEA). It is indeed curious that people do not understand that wind and solar projects are capital intensive and they do not generate power around the clock as required by service assurance in electrical power. After years of effort which produced no impressive results whatsoever, it is proper to recognize that growth in wind, solar, and geothermal:

- Comes at high price and
- Gives questionable results, at best.

The IEA estimates that tripling renewables' share of the global energy by 2035, will require $5.7 trillion in subsidies. Who pays? Statistics say that renewables contribute about 8% of primary energy in the US. That is as much as nuclear, but hydroelectric power and biomass are the bulk of this 8%. (In the US nuclear power still generates about seven times as much as solar, wind, and geothermal combined.)

The problem with the greens and other environmentalists is that they are living in cuckoo land, using artificial rates of return and questionable "What If" assumptions about the technology's potential to produce miracles. But as Dr. Rabii, the Manhattan Project physicist once said: Miracles are those things we do not understand.

References

1. Chorafas DN (2011) Energy, natural resources and business competitiveness in the EU. Gower, London, 2011
2. Gentry C (1991) J. Edgar Hoover. Norton, New York, 1991
3. The Economist, June 11, 2011
4. The Guardian, April 14, 2011
5. The Economist, May 7, 2011
6. Chorafas DN (2011) Cloud computing strategies. Auerbach/CRC, New York, 2011

Part III
Product Reliability

Chapter 6
Reliability Assurance

6.1 Quality and Reliability

The designer who aims to attain functionality, versatility, quality, and reliability at the same time is confronted with a very difficult task. Versatility demands more complex design and complexity is bound to pose many more demands for a strict quality control discipline. It may also adversely affect equipment reliability (see Sect. 6.4 or redundancy in design).

To be attained any quadruple objective invariably introduces some nebulous parameters, which usually lead to an elusive design concept difficult to tie down and specify. Therefore it has been a deliberate choice to leave functionality and versatility out of this discussion, in order to concentrate on the interplay between quality and reliability.

Starting with their differences, and prior to discussing how they correlate, it is proper to remind the reader that quality is a nearly absolute design characteristic and it is decided at the drafting board. By contrast, reliability is a relative factor which largely depends on mission, operational conditions, factors pertaining to the environment, and specified mean time between failures (MTBF). As we will see in Sect. 6.3 MTBF is not the only one of the critical metrics in reliability studies, but it is calculated on the basis of real life data which permit to measure the *mean life* of a device or system.

The careful reader will as well recall from Chap. 1 that reliability is not an ability. It is the probability of providing uninterrupted service assurance over a specified period of time, in a prescribed environment and under properly defined conditions of usage. Other things equal the higher is the quality, the greater will be an item's reliability, but this is not a monotonic function because so many more factors come into play in terms of equipment design and use.

Contrary to quality which is the subject of measurements and controls of what is produced as compared to engineering specifications and tolerances, reliability is a *prediction* is expressed as a probability. Our prediction is based on the number of

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering, 105
DOI: 10.1007/978-1-4471-2966-0_6, © Springer-Verlag London 2013

specific events (for instance, equipment failures) taking place in a given number of trials (identical test runs) which constitute the sample (Chap. 9).

Say that we want to know something about the reliability of a heat pump (like the one discussed in Chap. 4). The contract specifies that during the test period it should run 1000 h (about 40 days), and there are ten interruptions. The mean time between failures is 100 h (a very low statistic for a heat pump). The MTBF is an *empirical probability*.

There are different ways of expressing the mean time between failures, and not always the different modes agree with one another. The most widely used calculation is to simply divide the total hours in operation by the number of failures as was done in the example we just saw. There exist however improvements to this method connected to specific applications. We will return later on to this subject.

Designers who use and appreciate statistical inference may ask: What about an a priori probability? Let's assume the parts we are testing are cubes with drilled holes in all faces but one. Since a failure in delivery has been defined as a cube which comes out with an undrilled face up, that defines a theoretical probability of failure equal to 16.70%. This is a priori *probability*.

Predicted by means of analysis, with an infinite number of trials an a priori probability approaches the empirical probability. Notice that we have a certain degree of confidence when expressing an a priori probability, as it is based on a standard notation and benefits from some useful statistical laws.

As we can see from these very simple examples, quality and reliability complement one another and provide the user with some very useful tools. Each is an essential part of any engineering study. When available and acceptable, reliability measurements upgrade the value of a product by providing added knowledge to general information regarding devices, products, and systems.

The keyword is *metrics*. In terms of science-based approaches it would be meaningless for example to say that a machine such as a motor vehicle has long life, unless it is also stipulated what's its MTBF and under what conditions this life is measured. Was it on heavy duty or for pleasure driving? On a rough, or a smooth road? Under intermittent driving conditions or continuous heavy use? Were the atmospheric conditions mild or adverse when measurements were made?

All of these factors have an effect on the life figure of man-made devices and systems. Therefore on an equal level of quality it is necessary to prescribe the reliability of equipment one contemplates to purchase. Integral part of this information should be the conditions under which the measurements were taken. In principle, such conditions should emulate the actual conditions of use in the intended service of the equipment.

- In life cycle terms, it designed for long-term use, or
- Is it expected to operate in an environment characterized by normal temperature, humidity and altitude, or one with extreme conditions?

Notice that a factor of *time of use* is included in the first bullet. Time is a crucial element of any measurement of reliability. A life cycle approach emphasizes that initial performance alone is not what is desired. The measurement of reliable of

performance must span over a specified period of time. This view represents an important departure from classical quality concepts which stressed quality attributes at the time of manufacture.

Stated in different terms, a traditional quality specification makes no provision for assuring that the prescribed attributes and tolerances are maintained over time. This type of measure is rather static. By contrast, a reliability specification must be dynamic as it addresses the behavior of prescribed attributes and variables *in time*, including environmental factors. Reliability engineering, its:

- Methods, and
- Tools.

has been promoted by weapons systems studies. From there, it spread into a wealth of other applications—computers and networks being an example. Therefore, it will be wrong to think that the concepts described hereby are only applicable with weapons systems. Incandescent and fluorescent lamps designed for the chemical industry provide an example of civilian usage. Their reliability of time is very important. Unreliable lamps can start a fire if they explode in a chemical plant; therefore chemical and other factories use special lamps which have been designed with reliability in mind.

As Chaps. 4 and 5 have documented, it is not sufficient to assure by word of mouth that a device or man-made system—be it a lamp, a heat pump, or a nuclear plant—"operates well" it must as well be documented that it is of high quality, its reliability is beyond doubt and service assurance requirements have been fully observed.

Similarly, while it is necessary it is not enough that a machine or other component operates satisfactorily only when it is tested at the end of the production line. It is just as important that it continues operating satisfactory over time. Therefore, the effect(s) of latent defects have to be carefully and methodically evaluated which is the job of *reliability assurance*—all the way to field maintenance and parts replacement as they approach wear-out conditions. This raises three questions:

- Has the reliability requirements been imbedded at the drafting board?
- Are individual parts truly interchangeable and fitting the machine specifications and tolerances?
- Are the field engineers properly trained to assure maintainability and reliability of this machine or device (see Chap. 7)?

An answer to the first question will be provided by reliability specifications *and* tests to assure that they have been fully observed in this product design. It may well happen that these have to be destructive tests.

In regard to the second question, a sufficiently large sample of component must be available to verify interchangeability and its effect on system reliability. In the case of multisourcing, samples are needed from all manufacturers. An interchangeability specification written by the user organization and based on quality

standards can be very helpful to the people entrusted with procurement. The same is true regarding multisourcing vendor dependability.

If the material device or system will be used in different locations, *then* operational procedures and data collection systems must be on hand, all the way from how interchangeability and reliability data is gathered to how reports are issued in order to be comprehensible. This is also important in connection to the third bullet whose implications are discussed in Chap. 7.

As the reader will appreciate the nature of required actions for reliability assurance is not provided in a couple of words. Many manufacturers have approached it rather light ways. Those most serious have organized reliability engineering groups which used mathematical formulas and a rigorous methodology including destructive testing and conditioning. The latter requires that each part of equipment be operated for a specific time for *burn-in* to assure the elimination of any weak link or component—the so-called *baby failures*.

In conclusion, high reliability is attainable but to reach it we must start early in the design. Moreover, reliability testing should always account for unexpected and often unintended consequences. The more vivid way of explaining them is through a military example. In the 1980s in Afghanistan, America supplied Stinger missiles to help Afghan fighters against Soviet helicopter gunships. Not all of them were, however, used in action and in later years the American military spend time and money to comb the Middle East's and Southern Asia's arms bazaars to buy them back. (According to certain sources some of the Stingers bought in bazaars were then booby-trapped and sold again, to deter anyone who, at the risk of his life, wanted to use them.)

6.2 Designing for Reliability Assurance

The theme of this section is how to design engineering should be reorganized to meet higher reliability goals. Let me start in reverse with technical auditing. As careful reader will recall from Chap. 3 design reviews are an important part of design control, provided that they are performed by a team of mature professionals. *If* a design review includes reliability objectives, *then* three types of analysis must be performed for control purposes:

- The first should take place as soon as the device or system design has been tentatively established on paper.

This first reliability analysis is entirely theoretical, since no hardware has been produced except perhaps, a breadboard or two. Take as an example a circuit design. In this case, the circuit should be examined under conditions of worst case stress to each of its parts. Different scenarios could be made for different severities characterizing the operating environment. From this information, the most likely failure rate of each part can be deduced and summed to predict the inherent reliability of the design.

If this first reliability prediction for each device (or system-wise) meets the goal set for it by engineering, *then* approval can be granted to freeze the design and proceed with the development of engineering pilot models. *If* not, then the right policy is redesigned accompanied by design reviews. This may have to be repeated until reliability prediction meets the assigned goal.

- The second stage of analysis can be performed as soon as hardware similar to the final model is available.

This should roughly emulate the general configuration and embodiment of the projected equipment. Actual measurements under prescribed operational and environmental conditions are necessary to provide a basis for judgment, whose aim is to establish a more refined hardware-based prediction of inherent reliability.

When this figure checks with the first prediction and meets the assigned reliability goals, approval can be granted to manufacture the first series (which may still be prototypes). Since such an analysis considers each part of each circuit, a comparison of this data with the data of the first analysis will quickly identify those areas needing additional development or redesign.

- The third reliability prediction is based on data from special reliability tests run on first series production; or, alternatively, from complete engineering prototypes.

This third phase contrasts to those preceding it because the first prediction is entirely theoretical, while the second prediction is partly theoretical and partly practical as it uses data from tests on preliminary and partial hardware.

The third phase should also provide information on part failure rates necessary for preparation of part procurement specifications. For this purpose, statistical data must be obtained on actual failures and checked on whether they correlate with predictions "one" and "two", as well as with original reliability goals. High failure rate areas which indicate under design, or the use of improper parts, may be quickly located by comparison of test data with figures from the previous analyzes.

When the results of the third reliability prediction correlate with those of previous two predictions, and actual failure rates comply with the initial goals, it is reasonably safe to release the design for production. The major portion of the responsibility assumed so far is to ascertain that the specified parts will exhibit an acceptably low constant failure rate for a predetermined period when stressed to the conditions involved in the design. In other words, engineering qualification must certify that the sample parts:

- Obey the exponential failure law (see Fig. 5.1), and
- Have an acceptable failure rate for a suitable time when used in the planned application conditions.

In the logical sequence of events the process of reliability assurance should be followed by control procedures targeting the manufacturing stage, which can be achieved by means of quality control operation. Although quality control, by itself, can do nothing to improve the reliability of the product above that inherent in

engineering design, it nevertheless has a major role to play by assuring that the manufacturing operation does not reduce the quality—thereby degrading the reliability -below the design's inherent potential.

Engineers, manufacturers and end-users should be aware of the fact that the most mature design accompanied by the best specifications cannot result in a reliable equipment unless production operations brings out the maximum potential of the engineering work. The role of quality control in this context is to assure that the various reliability goals established during the design are met in the factory.

Quality control should as well assure that subsequent production lots retain the same life characteristics and low failure rates. For every complex equipment the problem of procuring parts to a low suitable failure rate is indeed a difficult one. Incoming screening operations and elaborate acceptance tests have to determine the time stability of incoming parts reliability characteristics over time.

Contrary to *initial quality* the so-called *long-term quality* may be reduced by the manufacturing process, and it cannot be improved merely by better maintenance. It is unfortunate but true that most established specifications and part acceptance procedures based on them deal only in terms of initial quality at time zero. Tightening up initial quality is important, but what happens to quality with operation after time zero is the element which must be addressed through focused reliability studies.

A contrarian view of the approach I have just outlined is that although quality is an important manufacturing problem, it is not necessarily the major factor of the ultimate reliability of the system. According to this school, initial quality affects the producibility of components and systems much more than it affects dependability.

Whether one accepts my approach or the contrarian view, the fact remains that in a way closely resembling what has been brought to the reader's attention in Chap. 2 in connection to quality control there exist control loops in a factory reliability program. In fact, they can also be found in different forms in connection to:

- Component procurement,
- Acceptance testing (incoming inspection),
- Production inspection, and
- Production level tests.

The primary objective of these control loops is that parts which do not pass specified tests are weeded out. Their common procedure is that a statistically valid sample of parts (Chap. 9) should be tested under simulated performance including accelerated life conditions. The experimenter should bear in mind that though accelerated testing[1] has certain shortcomings it is also helpful in proving a uniform background in putting under stress various lots of parts.

[1] For instance, using much higher voltage in life testing incandescent lamps.

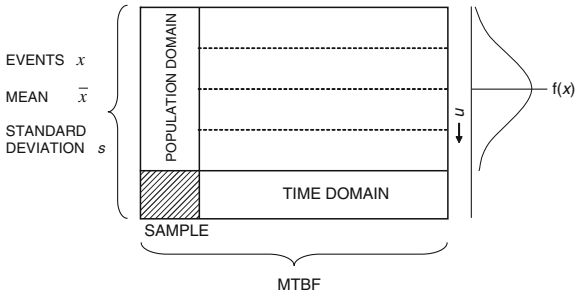

Fig. 6.1 Protocol of a double sampling approach: in the population and in time

Although reliability cannot be tested into equipment in the same way we are testing quality, it can be estimated by a priori approaches. Eventually, these have to be confirmed through an empirical probability which should come from the population of equipments in the field. By using stochastic thinking we can make reliability predictions based on the sampling, and population sampling as Fig. 6.1 suggests. This dial approach is particularly with equipment either requiring destructive testing or presenting some unfeasibility for statistically valid sampling.

In conclusion, classical quality control thinking must be upgraded by placing emphasis on the change of quality with time. Acceptance procedures based on specifications which include little, if any, consideration of random failure rate with time, have little control on the deterioration in quality or unpredictable catastrophic failures which occur during the hours of use. The time factor is assuming particular importance in the study of basic causes for equipment reliability.

Our aim with the double sampling approach: in the population and in time is to weed out *baby failures*, which occur at the very beginning of a product's life, and *wear-out failures* at the end of its useful life. The concept is shown in Fig. 6.2. All man-made devices and systems, as well as all natural ones, are subject to this double whammy of baby failures and wear-out failures. By weeding them out we significantly improve overall reliability.

6.3 Combining Mathematical and Engineering Knowhow

Reliability is *per excellence* a domain mastered by the combination of mathematical and engineering knowledge and techniques. Reliability assurance depends a great deal on cause and effect—and cause cannot be related to effect by mathematical theory alone. Assigning a given cause to a certain effect is not fully justified unless the fine print of why they are related is understood and appreciated.

The question whether the data or the technical knowledge comes first resembles a great deal the query about the chicken and the egg. Both are important as documented by the successful use of statistical methods in the analysis of reliability problems. Because we should always learn from our failures the most interesting engineering aspect of a reliability study is that of failures:

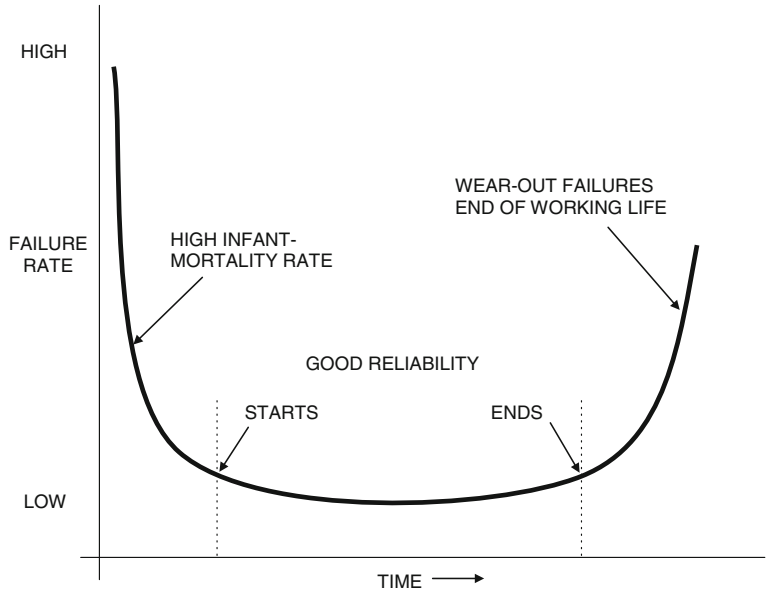

Fig. 6.2 Natural and man-made systems have baby failures and wear-outs. High reliability is found between these two extremes of the life curve

- Their cause(s), and
- Their effects (see also Chap. 12 on cause and effect).

Practical experience helps. Quite often in my practice a comparison of predicted and observed reliability revealed that the system's mean time between failure is less than that which was predicted based on the detailed design of the system. The real reliability statistics are provided by the *observed phenomena*, which does not mean that mathematical analysis is of secondary importance. One of its primary objectives is to explain the difference(s) between a priori and empirical conditions.

Another goal of mathematical analysis is to provide information which can be used to improve the reliability of systems and components by pinpointing the weakest link(s) in the chain. In order to attain analytical objectives, a statistically valid number of failure observations has to be analyzed with the aim to produce:

- Explanations, and
- Predictions.

If the probability of zero failures at unit time is accepted, *then* reliability prediction of man-made equipment, from devices to systems, is expressed as the probability that such equipment will perform satisfactorily for a given period of time t after being put in service. The theoretical function for this relationship is:

$$R(t) = \mathrm{e}^{-t/\overline{T}}$$

Table 6.1 MTBF required for failure-free time

Failure-free time,in hours, at 95% probability (h)	Required MTBF (h)
3	67
10	200
200	4000

where R is the reliability in percent as a function of time,$\overline{T}$ is the mean time between failures (which we have defined in Sect. 6.1 as MTBF), t is the time interval over which we project to use the equipment without failure under defined operational and environmental conditions.

The probability of failure-free life expressed by the foregoing algorithm presupposes that the failure rate is constant per unit of time. The algorithm is developed from the Poisson distribution:

$$R_{(x=x_i)} = \frac{e^{-N}\mu^{x_i}}{x_i!}$$

where:

μ is the universe mean number of failures in unit time,

N is the number of articles in the population under study,

x is the number of failures.

The law of the Poisson distribution in this application states that, for $N > 10$, failures will exhibit a normal distribution about $\overline{T}$. For $x = 0$, which means the probability of zero failures in unit time, the theoretical relationship is the aforementioned $R = e^{-t/\overline{T}}$.

The significance of this exponential life curve is that $\overline{T}$ falls approximately at the 37% probability point. This means that there is only a 37% probability of a particular equipment providing failure-free operation equal to the mean time between failures. With this life characteristic, the MTBF must be many times the required failure-free time. At 95% level of confidence (Chap. 10) this algorithm gives the values presented in Table 6.1.

Not everybody working in reliability studies agrees with the notion $\overline{T}$ = MTBF. As Sect. 6.1 has explained, to calculate mean time between failures we simply divide the total hours in operation by the number of failures. If all observations did stop at a failure, such calculation would be accurate. However, this is not always the case.

In many instances the time to failure (TTF) is longer than the period of observation. When this happens only information on TTF is that it is longer than some known value but we don not exactly know by how much. Hence, the observation is incomplete. A complete observation would cover from time zero, t_0, up to time of failure t_f. In order to correct such a shortcoming, some practitioners use the algorithm:

$$\overline{T} = \frac{\text{closed intervals} \quad + \quad \text{open intervals}}{\text{sample size}}$$

where:

- Closed intervals are the total time between failures,
- Open intervals the total periods of observation where no failure has yet occurred, and
- The total sample size equals the amount of closed and opened intervals.

This is a different, and more accurate way of looking at the mean life of a system. Its $\overline{T}$ reflects the total hours of operation divided by the number of malfunctions, considering each open interval terminated in a failure. Critics of this approach say that it is pessimistic because observations do not terminate at the close of an observation period but extend for an unknown period beyond the time of observation. This is one of the arguments in reliability studies which have not been settled and I doubt it will. I stick to the MTBF, unless conditions specifically warrant otherwise.

Mean life is by no means the only important statistic in reliability engineering, though it is probably the most crucial reference. Table 6.2 presents in a nutshell the important concepts and terms in defining life expectancy for man-made devices and systems. All of them are useful for statistical inference connected to reliability studies—whether they regard equipment in use or lot inspection of goods bought from suppliers.

6.4 Inspecting Lots and Systems

In connection to quality control and reliability studies (as well as in procurement), the term *lot* is used to mean a group of items connected to a process of production, purchasing, or inspection. The lot is a collection of devices or other products manufactured under essentially the same conditions, from which a sample is drawn and tested.

Notice, however, that the same term *lot* is used in regard to a preproduction process whereby one or more units of product are submitted prior to initiation of production. Through testing must be determined compliance with the *acceptability criterion* (see also Chap. 9 on sampling). Whether the subject of our attention is the lot or equipment already in operation, the objective of inspection is to determine the choice of one among three possible decisions:

- The object of our study meets the acceptability criterion,
- It does not meet the acceptability criterion, or
- The evidence is insufficient for either decision and the test must continue.

Inspection is a technical audit (Chap. 3). Our observations and the examination of test results may reveal outliers biasing the projected reliability characteristics with negative effects; for instance, skewing an established failure pattern. An

Table 6.2 Concepts and terms used in defining life expectancy and sampling plans for statistical inference connected to quality and reliability studies (in alphabetic order)

Acceptable failure rate during period of time. This is the maximum failure rate that can be considered satisfactory

Acceptable mean life identifies the mean TTF considered satisfactory

Acceptable proportion of lot failing before specified time. This is the maximum fraction of a lot that may fail before specified time with the lot still being considered satisfactory

Expected number of failures. The number of failures required for decision which have occurred at the time of decision as to lot acceptability

Expected waiting time. The time elapsed from start of the life test to the time decision is reached as to lot acceptability

Instantaneous failure rate, also known as hazard rate is equal to 1/MTBF

Length of life is a term often used interchangeably with TTF to denote the length of time it takes for a product or system to fail after being placed on life test[a]

Life test sampling plan. This is a procedure specifying the number of units of product from a lot which are to be tested, and the criterion for determining lot acceptability

Life test terminated at preassigned time. Termination takes place when a preassigned termination time, τ is reached

Life test terminated upon occurrence of preassigned number of failures

Life test. This is the process of placing a given product under a specific set of test conditions and measuring the time it takes till failure

Mean time between failures (MTBF). The terms "mean time between failures" and "mean life" are often used interchangeably and shall denote the mean length of life of a device or system under test

Proportion of lot failing before specified time, denotes the fraction of the lot that fails before some specified time

Sequential life test. In this life test neither the number of failures nor the time required to reach a decision are fixed in advance. Instead, decisions depend on the accumulated results of the life test

Specifying failure. The state that constitutes a failure must be specified in advance of the life test

[a] This length of time may be expressed in any convenient timescale, such as seconds, hours, days. On other occasion the term *TTF* is used to differentiate from MTBF when the time between failures is longer than the time of observation.

examination of the failure data may show that failures are due to out-of-tolerance component parts or, alternatively, weaknesses in system design.

Whether made for quality control purposes or for reliability studies, inspection implies rules which should characterize observations being made. Such rules may or may not require an experiment or selection of a sample; but experimentation is an increasingly acceptable procedure because of the evidence it provides.

Typically, in the background of quantitatively-oriented inspection rules lies the fact that measurements taken in the course of observations will be subjected to statistical analysis. Quantitative inspections may or may not be accompanied by qualitative inspections—though in a number of cases a combination of the two provides the best assurance.

Many inspections require experimental evidence, particularly when one or more complex variables are involved, because usually (though not always) it is

impossible to solve such problems through intuitive trial-and-error approaches. An analytical examination of relationships is called for, with:

- Carefully selected parameters, and
- The likely effects of environmental factors taken into account.

For each individual case, a technical audit, or inspection, must incorporate pertinent engineering knowledge. *If* both field experience and laboratory data are required to account for the parameters under study and their variability from design to production and field service, then the person(s) in charge of the inspection must be able to provide them. In addition, such data should satisfy the requirements of an observable level of service assurance, as explained in Chap. 4.

Experimentation will be surely required when there are changes in expected reliability of a device or system; and/or when the critical variable is that of external stresses. Changes over a system's useful life is one of the reasons why Sect. 6.2 pressed the point that the probability of failure must be expressed as a function of time. There may as well exist multiple stresses, each being a variable. For instance:

- Vibration,
- Shock,
- Temperature, and so on.

Reliability analysis could be significantly simplified if these stresses could be separated. If so, then the total probability of failure is expressed by means of mutually exclusive events, and their sum must equal one. This can happen, but it is not a frequently encountered case in reliability studies. Hence the wisdom of using factor analysis and Latin squares (Chap. 11) which permit to study the interactions between variables.

A rather common happening is the need to calculate the compound reliability of a system. Systems are usually made up of several components with differing failure characteristics. The way to bet is that the probability distribution of each component will be different. To combine them and find an overall system reliability function, we take the product of the separate reliabilities of the system's components:

$$R = R_1 \cdot R_2 \cdot R_3 \cdot \ldots R_n$$

The product of these probabilities is another confirmation that the most unreliable components create havoc with overall system reliability. Keep this in mind in Sect. 6.5 when we talk of compound reliability curves with components varying from 1to400.

The compound reliability computed by the foregoing algorithm may need to be further reduced because of safety factors. A safety margin is usually expressed in number of standard deviations from the mean for instance six times standard deviations (the Six Sigma method is discussed in Chap. 12).

Testing for critical weaknesses which may require safety margins becomes an important tool when the test's objective is to foretell wear-out failures (see Fig. 6.2). This is assisted by the fact that, quite often, wear-out failures tend to make a normal distribution about a mean lifetime. A way to arrive at an a priori MTBF on a rapid basis is a technique known as *search for critical weakness*. This consists of testing in environments exceeding qualification levels to see:

- How much *margin* the design has, and
- Where the weak spot lies, always remembering the principle that a chain is never stronger than its weakest link.

Another case encountered in the analysis of reliability data is that of constant-hazard-rate failures. These are likely to occur at any time during the life of a product. Say, for example that we have reduced a given type of failure to a minimum of 10 per 100 operating devices every 100 h. Starting off with 100 devices, we lose 10 in the first 100 h, 9 in the second 100 h, a little over 8 (8.1) in the third 100 h, and so on. This plots as an exponential curve.

6.5 Can Reliability be Improved Through Redundancy?

The analysis and interpretation of empirical reliability measurements, particularly when these are below a priori established levels has led many designers toward the use of redundancy—the theme of this section. Experienced designers, however, know that prior to using redundancy they must obtain clarification and (possibly) reestablishment of realistic reliability goals applied to a system and its components.

If redundancy is judged to be necessary, *then* the designer should start with a rough draft of a connection in parallel which allows him to proceed with reliability prediction. The data he uses should be in principle based on component part stress analysis permitting to ascertain whether a sought-after reliability can be met within the realm of his chosen approach.

For specific engineering projects, another crucial step is customer liaison to point out relationships between specified and predicted reliability figures. Also, to explain that while redundancy might improve reliability there will be trade-offs with other characteristics of the design. It is important that the end user knows a priori what kind of trade-offs will (most likely) have to be made.

Speaking in practical terms, redundancy is indeed an effective means for increasing reliability, but in some cases restrictions on weight and space, for example, make this approach unwise (if not almost impossible). In addition, redundancy cannot be applied to the most unreliable components because:

- Sophisticated systems would have to be devised to sense the various malfunctions, and
- These subsystems may be more complex than the original system; itself as well as uneconomic.

There is as well the risk of over redundancy, which is relative to the situation. For an electronic data processing system, for example, redundancy boundaries may be set by economic criteria. To the contrary, for a missile system the criteria would be accuracy on target.

In professional life I have as well found cases where redundancy was an excuse because the real problem behind reliability failures were conceptual. These had much to do with the inability to properly grasp the key aspects of the project.

In other cases, statistical inference and probabilistic thinking were used scarcely if at all. Indeed, there is a widespread lack of statistical thinking which is a pity because stochastic approaches must be integral part of analysis and design. Some higher reliability solutions do not redundancy but:

- Careful analysis of requirements to determine compatibility with field needs,
- Comparison of reliability requirements of the proposed system with case history data from similar systems already in use,
- Detail in system design to determine specific subsystem reliability goals, and
- Prediction of likelihood of achievement of reliability goals, based on parts counts and failure rate information on available parts.

Having said this, it is proper to add that redundancy has a role to play on a design. Let us assume that after having gone through the aforementioned four steps the need for redundancy is still present. There are four alternatives shown in Fig. 6.3, and we should study what they may offer.

As we saw in Sect. 6.4, *if* n elements are connected in series, and the failure of each element results in a system failure, *then* the overall reliability R of the system is equal to $R = r^n$, where r is each component's reliability (on the hypothesis that r is equal for every one of the components[2]; and n is the number of elements in series. As Table 6.3 shows, the relationship $R = r^n$ produces some startling results.

We can as well work the other way around, starting with a target system reliability and for a given number of components, which we know a priori, examine the required reliability of each component in order to reach a target failure figure. Table 6.4 presents the results of this exercise by assuming a system of 500 component parts, each of an equal reliability r.

It needs no explaining that the required component reliability of 0.99996 imposes almost impossible requirements. Rather than improving quality to such high figures, which is hardly feasible, it is advisable to build redundancy into the system.

Take as an example the simple message transmission system AB in Fig. 6.3, and say that we want to send a message from A to B. We can do that with one link (Fig. 6.3a), or with multiple link (Fig. 6.3b). In the latter case, in lieu of sending

[2] r could also be seen as the average component reliability, but it is unwise to use averages in reliability studies.

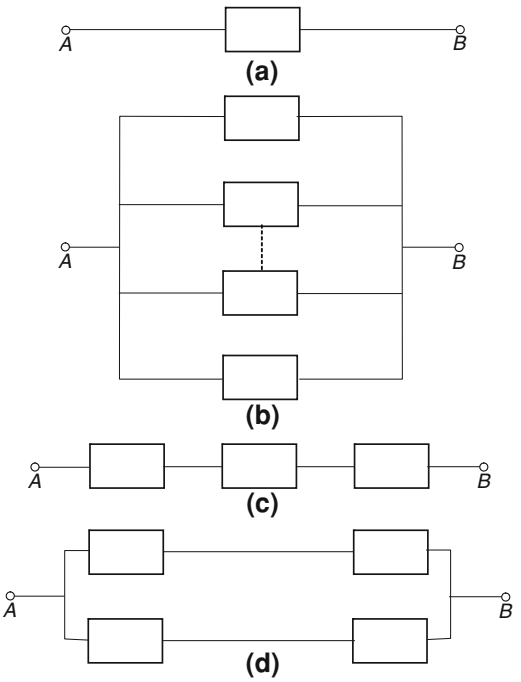

Fig. 6.3 Connections in series and in parallel of critical components

Table 6.3 System reliability as the number of components increases

Number of components in the system,(n)	Mean element reliability (r_i)	Resulting system reliability
10	0.99	0.90
100	0.99	0.40
500	0.999	0.60
1000	0.9999	0.37

Table 6.4 Reliability of a system with 500 component parts and expected number of failures

For R it must be	r resulting in	*Failures equal to*
0.70	0.9929	0.00071
0.80	0.99955	0.00045
0.90	0.99991	0.00009
0.95	0.99996	0.00004

from A to B one message, we send m messages, where m is the number of links. With this approach the probability of success increases.

See in a different way, with a connection in parallel, the probability of failure of the system, Q, is equal to the product of the individual probabilities of failure.

$$Q = Q_1 \cdot Q_2 \cdot Q_3 \cdot \ldots Q_m$$

Table 6.5 Reliability of a system with components in series and in parallel

	$n = 1$	$n = 3$	$n = 20$	$n = 100$
$m = 1$	0.90	0.73	0.12	0.000026
$m = 2$	0.99	0.97	0.81	0.35
$m = 3$	0.999	0.997	0.98	0.90

Hence, the coupling in parallel of transmitting arcs (components) is an effective means of increasing the system reliability. Connection in parallel, however, is not without risks.

In contrast, the probability of success decreased if there is n components in series. Figure 6.3c shows three components, and the failure of each component would result in the failure of the system. Data on the reliability of a system with components in series and in parallel are given in Table 6.5.

Like the example we have seen at end of Sect. 6.4, with a connection in series the probability of success of the system, R is equal to the product of the probabilities of success of the components R_i:

$$R = R_1 \cdot R_2 \cdot R_3$$

As stated in the opening paragraphs of this section improvements in the reliability of a design may require built-in redundancy. But at the same time the more components a system has the greater is its fragility, let alone its cost. Neither are bigger systems small systems which have grown-up. Bigger systems have inherent to them many more design requirements.

Redundancy is no penicillin. Ways and means must be put in action to work at the same time to accomplish the task. Standby redundancy is a better option applicable to systems with alternate means of accomplishing their mission, the redundant unit being switched in by a malfunction sensing device when the primary system fails.

In conclusion, decision as to whether or not to increase reliability through a built-in redundancy, must be based on the operational requirements of a system, seriousness of failure, redundancy cost versus high-quality cost, and redundancy cost versus cost resulting from failure times the probability of failure. The quest for reliability must be seen as one of system parameters, all of which should be properly balanced in an optimum systems design.

6.6 The Case of Cost/Effectiveness

A systematic view of all elements of a reliability assurance program must not only consider the costs of each component but as well system integration expenditures. While reliability and quality objectives should be always upheld, designers will be well advised to search for opportunities to reduce costs. A number of opportunities arise from current practices which are not always rational:

- Imposing specifications which are too rigid for intended use,
- Asking in excess of contractual requirements,
- Requiring tasks to be performed because of habit rather than true need, and
- Calling for excessive documentation which will never be read.

The principle of cost/effectiveness can be best served by creating a questioning attitude throughout the organization, and finding answers to the "Why is *this* necessary?" Such answers may not only help to simplify the system, but may as well motivate a search for less costly components.

Some people think that as far as reliability engineering is concerned the best possible policy is to "pull all the stops". In practical cases, however, there are mitigating factors which constrain an approach to reliability assurance without bounds. Costs and complexities are the more frequent reasons, though they are not the only ones.

Many companies who have followed this policy of questioning "Why is *this* necessary?" have seen dramatic improvements in cost/effectiveness. The better managed have developed case studies by recording the cost improvement in projects which were initiated and conducted by the reliability insurance personnel.

Accomplishing a program of achieving the proper level of reliability at the lowest cost represents a real professional challenge, because any cost/effectiveness program worth its salt will never lose from sight reliability targets. These have always to be confirmed through technical audits and design reviews.

Integral part of greater cost/effectiveness is the review of the feasibility of alternative designs, accompanied by check out of alternative component parts. Sometimes solving and an old problem can be as easy as looking at it in a new light, while avoiding to go to other extremes which might lead downhill.

The devil's advocate in cost optimization should be the company's reliability assurance manager, who it would be wise to support through technical audits by an independent consultant. This requires:

- Analyzing all work elements of a reliability and/or quality program, and
- Coming up with a factual and documented conclusion on whether better cost/effectiveness can be met.

Reliability assurance audit should not be taken in the trap of the appearance of savings at the expense of dependability of the product. Quite to the contrary to what is commonly believed, considerable savings in design as well as in manufacturing can be achieved by:

- Recognizing the difference between appearance and quality, and
- Avoiding the addition of unnecessary costs under the guise of improved product versatility or performance.

Another issue closely related to the effects of "appearance" and its impact on product quality or cost is the tendency to use familiar type specifications regardless of whether they apply to the intended end use of a device or system. Often these familiar specifications and tolerances are excess of contractual requirements.

Creeping complexity is also an element which adds substantially to cost. Like organizations, design plans always tend to grow rather than diminish. If, because of an application problem a particular reliability procedure was introduced, the tendency exists to keep this procedure on the company books and apply it to all subsequent programs or products. To counteract this rather natural tendency, the reliability assurance manager must continuously ask, "*Why?*" and assure himself that:

- Only those characteristics required for a particular device or system are integrated in that product, and
- Company departments are measured on the performance of the product and its cost/effectiveness—not on the size of the quality or reliability assurance group.

Work psychology also plays a role. One of the reasons why sometimes projects never end, particularly in information technology, is that people working on them are afraid to lose their job if they come to completion. In the majority of cases there is no such danger. New problems always arise necessitating the initiation of a new product assurance activity or cost/effectiveness investigation and these new projects will bring along new challenges as well as plenty of opportunities for thinking and for working.

Models can help both in simulation and in investigation, but as far as cost/effectiveness is concerned, there exists no simple mathematical model able to provide its user with optimum reliability assurance policy and program. The crucial element is the management attitude of examining the cost and effectiveness of each task in a way to increase the effectiveness-to-cost ratio.

- The best program is one which supplies the *proper reliability at lowest cost*,
- But attacks one issue at a time: higher quality/reliability or lower cost. Not both at the same time.

Key to this process is to divorce oneself from the past practices and examine all activities from a questioning standpoint. The person doing optimization must have the professional integrity and intellectual courage to question even his own basic premises not only those of other people in his organization.

- Why do we do "this"?
- Who asked for it?
- Why does he want it?
- How much does it cost?
- Is this the cheapest way to do it?
- What would happen if we did not do it?
- How often has this item (or solution) presented a deficiency?
- How much money can we save by not doing this?
- What will be the impact of eliminating this item (or activity) on resulting quality (or reliability)?

I have been asking these questions most frequently in technical audits I have been doing and I am no more surprised when I see that companies cannot supply

valid answers to in-depth questions. The reason for asking them is that by so doing persistently the questioning attitude spreads in the organization.[3]

A consultancy thrusting upon itself this questioning effort should remember that reliability and quality requirements, as well as service assurance programs, vary from company to company in terms of goals and constraints. To the contrary, the basic queries of cost/effectiveness tests described in this section tend to be valid cross organizations and cross products.

It is also important to bring to the reader's attention that whether we talk of quality or of reliability product-assurance programs vary not only with the company's culture, its past accomplishments, the mission and the product but also with timing of for a given action. A project which has been initiated to cure a specific shortcoming at time t will not necessarily provide an answer which is always valid because in the mean time the problems have changed.

[3] Care must be exercised to create a permissive attitude in the course of questioning, and not to push too hard for plausible answers or to reflect discredit on an individual if he or she has not been thinking in these terms and therefore cannot come up with an answer.

Chapter 7
Reliability and Life Cycle Maintainability

7.1 Tests for Reliability Assurance

The theory underlying the development of reliability methods, standards, and objectives, including statistical inference (Chap. 8), life cycle sampling plans (Chap. 9), operating characteristics curves (Chap. 10), and experimental design (Chap. 11)—assumes that length of life measurements are drawn from an exponential distribution. Practical experience suggests that life test sampling plans are not to be used indiscriminately, simply because it is possible to obtain life test data. Only after:

- The reliability test is properly designed
- The exponential assumption is deemed reasonable.

should a sample plan be employed. It is so important to pay attention to the accuracy of a sampling plan, because the accuracy of reliability models is direct function of the accuracy of the data such plan will provide. In addition, the nature of reliability models to be used for quantitative evaluation would often depend on the *type* of reliability with which we are concerned. There are three main types:

1. *Instantaneous*. Fuses are an example of a device whose reliability requirement is that of instantaneous life. Generally, all so-called "one shot deals" have an instantaneous life requirement, but that "instant" may be far apart from the device's production time.
2. *Normal long life*. Normal long life is a characteristic of communications equipment, computing machinery (Sect. 7.3), and fire control systems. For normal life, reliability studies utilize as the unit of measure an average time between failures is in the range of 100–200 h.
3. *Extreme long life*. When reliability requirements for a life span extend from 10 to100 years, we refer to them as intended for extreme long life. Examples of systems with such requirements are underwater telephone cables and repeaters, earth satellites and spacecraft. No truly dependable tools of mathematical statistics have been developed to measure a priori extreme long life of man-made systems.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering, 125
DOI: 10.1007/978-1-4471-2966-0_7, © Springer-Verlag London 2013

Because reliability is in actuality a probability, with all three aforementioned types statistical inference is used for prediction of central tendency and variance, based on measurements of samples from a population. The prediction of future reliability should make use of levels of confidence (Chap. 10). For normal long life, prediction can be:

- Theoretical,
- Experimental, or
- Empirical.

A *theoretical* approach requires considerable work to develop and it is ever subject to being discarded because of a "nasty" new fact or finding. Some years ago, a researcher charted a certain operation and then constructed a mathematical model, tested it and fount it fit for the operation. Three years later, however, the specifications of two key system components changed and the model had to be discarded.

The *experimental* solution challenges theories and dogmas; it also mistrusts poorly controlled empirical results. An experimental approach provides a better basis for reliability inference, and this for two reasons. First, since the model is based on an experiment the experimental findings themselves can serve for documentation on whether it fits a real life situation. Second, once a method has been developed the experiment can be easily repeated every time there is a new component or other type of change.

An *empirical* approach consists of making observations and measurements of an actual operation. Statistical inferences are made on the basis of these measurements often, but by no means always without bothering to check them for pertinence and completeness.

Empirical models can play a constructive role, but they have a downside if changes alter the data on which they have been based. For instance, Olivetti, the Italian business equipment company, asked its lab in Palo Alto to write a model of its manufacturing operations in Ivrea, Italy. The model was an .optimizer and its application identified several conditions which needed correction. When these corrections were done, the model had to be discarded because the data on which it was based had significantly changed.

The existence of *model risk* is not duly appreciated though it is always present. Quite often when a particular equation, or set of equations, is developed to emulate a given operation, this model tends to become an *established* scientific tool. Such a tool, however, often contains paradoxical assumptions even the equations are accepted without much questioning.[1]

Model risk might be reduced when reliability data on components and system performance are collected in statistically significant amounts. It is useless to collect information which provides a weak statistical basis, as we will see in Chap. 10 on operating characteristics curves.

[1] An example is the Black-Scholes option pricing model.

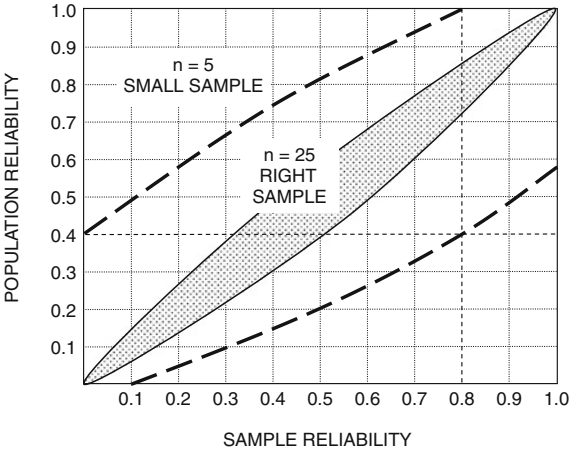

Fig. 7.1 Correlation between sample and population reliability with samples $n = 5$ and $n = 25$

The use of theoretical models which have passed the test of time helps in improving the accuracy of reliability prediction. At the early day's reliability studies, Hughes Aircraft capitalized on the fact that field data showed the failures of airborne radar equipment closely approximating the Poisson distribution (exponential life). Other studies at Vitro Laboratories, based on a very large number of shipboard electronic equipment failures, demonstrated a remarkable agreement with the Poisson life curve, particularly during the middle portion of the Weibull life curve (Sect. 7.2) [1].

When new designs are introduced, the desired information elements must be produced by new test programs. Robert Lusser [2] had made a rule that the reliability test program should be started when the system is in its preliminary design stages; that it should be conducted, at high priority, throughout system development; and it should continue as long as the system remains in production. Lusser's goal was *prevention* of unreliability rather than detection after operations. This policy requires the following steps:

- Make and test a prototype,
- Start laboratory tests with the most basic parts,
- After these have been found to be satisfactory at component level, switch the emphasis to higher-order assemblies,
- Make a full-scale prototype and test it to failure,
- Make reliability predictions based on obtained failure data.

Lusser also insisted on what can be called *sampling reliability*, aiming at predicting the reliability of population through a correct sample size. Too small a sample size may give incomplete information. Too large sample size results in waste of material, time, and effort. Figure 7.1 presents a correlation between sample and population reliability. Notice the difference of results obtained with a very small sample and with a statistically valid sample.

Having said this, it is proper to bring to the reader's attention that in reliability engineering possibilities for sample selection are limited by the small number of units available early in the course of equipment development. This is no good news because the confidence level of test outcome will vary directly with the number of units on test.

There exists as well another essential difference between statistical quality control and reliability engineering due to the fact that in reliability we make inferences in both the population domain and in the time domain. In terms of equipment dependability and performance the time distribution cannot be assigned or assumed. It must be measured.

Because of this time dimension, sampling plans used in reliability studies operate on a continuous basis, albeit at a reduced sampling level as far as no substandard equipment are identified. A reduced inspection procedure, as opposed to normal sampling (and from there to tighten inspection) if defective items are found in the sample.

7.2 Poisson and Weibull Distributions

Chapter 6 defined reliability as the probability that over time t, a device or system will perform without failure of its intended function(s), under well-defined environmental and operational conditions. It has well been stated that *if* the probability of zero failures at unit time is accepted, *then* the theoretical function of reliability is:

$$R(t) = e^{-t/\overline{T}}$$

Chapter 6 also brought to the reader's attention that the reliability of devices connected in series and sharing a common reliability, is this common reliability of each device taken to a power equal to the number of devices. Given that reliability is a probability, it can be expressed by the algorithm:

$$R(t) = p(0, t) = 1 - p(x \geq 1, t)$$

where x, is the number of failures in time t, $p(x \geq 1, t)$, is the probability of one or more failures in time t.

According to statistical theory, if failures occur as discrete, single, independent events in time, the probability of exactly x failures in time is given by the Poisson distribution formula:

$$p(x, t) = \frac{\mu^x e^{-\mu}}{x!}$$

Where μ is the expected number of failures or, essentially, the universe number of failures.

If $\overline{\lambda}$ stands for the average failure rate, or average number of expected failures per unit time, as contrasted to λ which is an assignable fixed function of time, *then*:

$$\mu = \overline{\lambda} t$$

the general failure rate as a function of time t will be:

$$\overline{\lambda} = \overline{\lambda}(t)$$

with the running average failure rate computed by:

$$\overline{\lambda} = \overline{\lambda}(t) = \frac{1}{t} \int_0^t \lambda(t) dt$$

If the expected number of failures μ is a random statistical variable, such as in a normal distribution, *then* the resulting distribution is said to be a generalized Poisson distribution. It is exactly in this sense that it is employed:

- In reliability studies, and
- In risk management.

The Poisson distribution can be considered as a derivative of the binomial distribution, serving as an approximation to the latter when the probability of a given result, out of two possible outcomes per trial, is small. In this sense the Poisson distribution is sometimes called the *law of rare events*. With practice, and because of the potential they present in a growing number of studies—from engineering to finance—Poisson distributions have taken the status of a fundamental, indispensable statistical tool.

In the special case of zero failures, hence, $x = 0$, the Poisson equation gives the reliability time function, known as survival equation:

$$R(t) = p(0, t) = \mathrm{e}^{-\overline{\lambda}\mathrm{t}}$$

In general, $\overline{\lambda}$ changes as t changes. The algorithm of the example which we saw in Chap. 6 with devices connected in series each with reliability R_i becomes:

$$R = R_1 \cdot R_2 \cdot R_3 \cdot \ldots \cdot R_n = (\mathrm{e}^{-\overline{\lambda}\mathrm{t}})_1 (\mathrm{e}^{-\overline{\lambda}\mathrm{t}})_2 \ldots (\mathrm{e}^{-\overline{\lambda}\mathrm{t}})_\mathrm{n} = \mathrm{e}^{-\mathrm{n}\overline{\lambda}\mathrm{t}}$$

where the failure rate is averaged with respect to time in the interval t. In reliability measurements a helpful extension of the inference provided by the Poisson distribution is given by the Weibull distribution. Weibull offers a variant of the reliability algorithm:

$$R(t)\mathrm{e}^{-\mathrm{t}^{\mathrm{a}/\mathrm{b}}}$$

where a and b are parameters. Proposed in 1951, the Weibull distribution was first used for general statistical applications; at least that was the intention of its

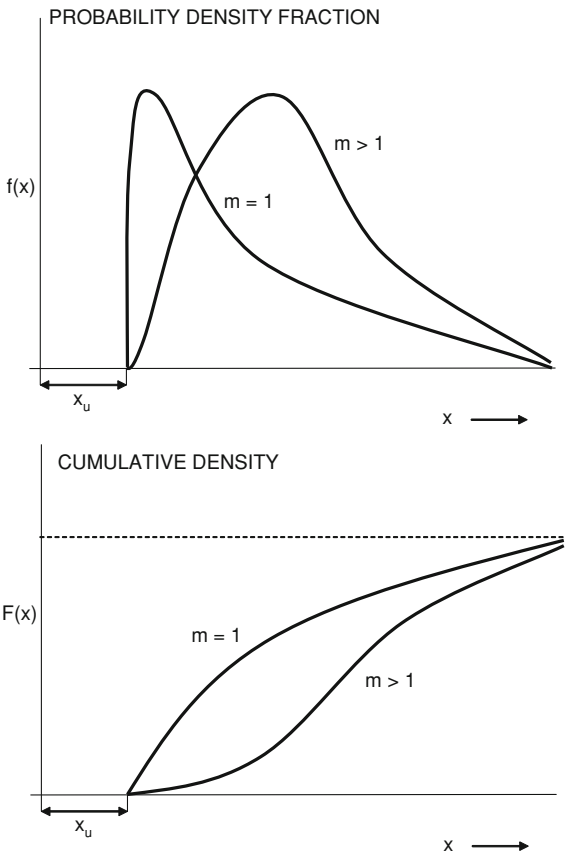

Fig. 7.2 Typical curves for the Weibull distributions as deduced from their formulas

developer, Wallodi Weibull [3]. Over time, its main domain of applications has been in reliability studies.

Reliability data conform to the pattern represented by the Weibull equation if the range of t is great enough. Provided this is so, the failure rate is expressed by:

$$\lambda(t) = \left(\frac{a}{b}\right) t^{a-1}$$

This algorithm can be used with a growing range of applications, for instance, in connection to telephone traffic and associated failure rates. Even if the failure rate changes with time, with certain provisions the probability of x failures in time t, each occurring singly, discretely, and independently is still Poisson-distributed.

Typical forms of Weibull distributions deduced from their formulas are presented in Fig. 7.2. The examples come from engineering projects, where the Weibull algorithm has been applied to fit accumulated failure data on electronic parts subjected to life test [4]. The Weibull probability density function is:

$$f(x) = \frac{m(x - x_u)^{m-1}}{x_0} e^{-\frac{(x-x_u)^m}{x_0}}$$

where m shape parameter, x stochastic variable equal to or greater than x_{u}.

The corresponding cumulative distribution function is:

$$F(x) = \int_{-\infty}^{\infty} f(x)dx = \int_{x_u}^{x} f(x)dx = 1 - e^{-\frac{(x-x_u)^m}{x_0}}$$

In an algorithmic sense, we should watch out for four parameters named for their effect on probability density and cumulative distribution functions:

x_u = location parameter
m = shape parameter
x_0 = scale parameter
t = time parameter

With life tests measurements it is recommended to start at $t = 0$. Therefore in applications to life tests or other aspects of reliability, x_{u} is usually set equal to zero. It in other cases—for instance in the management of financial risk—it is advisable that a similar hypothesis is followed in terms of initial condition, subject to confirmation by test.

7.3 Life Cycle Maintainability

Maintainability has been one of the earliest preoccupations of the engineering profession. Today it is even more vital for the purpose of ensuring reliability in applications with lifecycle perspective. Over the years the way we look at, and mission we give to, maintainability have broadened,

- Become more complex, and
- Demanding a higher degree of expertise, methods, and tools.

Precisely because maintainability can make or break lifecycle reliability, it has to be looked at very carefully. The crucial points where something could go wrong have multiplied, and reliability standards for each of them need to be established reflecting not only the present but also the coming requirements as the use of a device or system continues to expand.

Much of the background of this section is based on practical applications with computers and communications, particularly with real-time systems. The targets of reliability and maintainability within an online environment include but are not limited to:

- Hardware, its topology and serviceability,
- Basic software: new releases, support of past releases,
- Communications lines, nodes, modems, environment,
- Maintenance of applications programs and packages,
- Distributed database management, and
- Online (remote) testing and maintenance capabilities.

In a way quite similar to the other examples we saw in the preceding section, systems reliability must be maintained throughout the life of an online system. There exist two basic standards for evaluating uptime in an objective, factual, and documented manner. The first is system scheduled hours (uptime and downtime). The second is system downtime resulting from different reasons interrupting normal processing. A simple algorithm reflecting this is:

$$\frac{\text{System usage time}}{\Sigma \text{ of interruptions}}$$

Unscheduled interruptions may be due to hardware failures, to new basic software updates, or to other reasons like operator failures. We measure hardware failures according to:

- Mean time between failures, MTBF (Chap. 6), and
- Mean time to repair (MTTR).

With computers and communications systems, MTTR is associated to repair proper (including the time maintenance) engineers take to come to the installation and start working. Over and above that time, computers require recovery procedures which make the strict definition of MTTR inadequate. Therefore, we also use two other metrics:

- Mean time between system interruptions, MTBSI, and
- Mean time of system interrupts, MTOSI.

There exists as well the notion of *reparability*, which identifies the speed with which a component failure can be detected and fully corrected. Reparability is important inasmuch as all types of interruptions create a prejudice and the greater is MTOSI the greater will probably be the prejudice.

The way to bet is that in the majority of cases short interruptions are due to software while long ones are hardware oriented. In either case start-up and recovery time may be significant. It is easy to shut down a computer system, the problem is to start it up again and make it available for useful work.

Therefore, while MTBF and MTTR are useful metrics, MTBSI and MTOSI are much better means of evaluating the service obtained from computers and communications due to causes of unreliability or other interruptions. The mean time between failures and the mean time of system interrupt have a dramatic effect on lifecycle cost. The selling price of equipment at standard reliability levels is a relatively small portion of total system expenditures related to the service to be derived from the system. Other things equal:

- Costs decrease as the reliability increases, and
- They increase very quickly as the reliability level drops.

Systems availability is important both from the end-users viewpoint and from that of cost-effectiveness. For any practical purpose system availability represents the degree of attainable service assurance (Chaps. 4 and 5). Mathematically, *availability* is the probability that a system is running at any point in time during scheduled hours. With computers and communications it is calculated through a relatively simple algorithm:

$$\text{Percent availability} = 100 \cdot \frac{\text{Uptime}}{\text{Scheduled time}}$$

Uptime is the effective system usage time. With computers and communications reliability may also be defined as the extent to which a system or component performs its specified functions without any failures visible to the end user. This notion can be extended to cover the after effects of redundancy on availability.

When in spite of failure of some of its components the computer system continues operating albeit at reduced functionality, due to the embedded redundancy, we are talking of a state of *fail soft*. Within a life cycle perspective *availability* will still address the system running at any point during scheduled time but the algorithm:

$$\text{Availability} = \frac{\text{System usage time}}{\sum \text{ of interruptions}}$$

will need to be modified to account for reduced functionality over some periods of time. This can be easily done by substituting in the numerator a sum of time segments instead of an end-to-end usage time.

$$\sum T_1 \cdot k_1 + T_2 \cdot k_2 + T_3 \cdot k_3 + \ldots + T_n \cdot k_n$$

where

T_i are the time segments,
K_i are the functionalities where $K_i \leq 1$.

This availability algorithm is easily derived from the alternative approach to calculation mean life (instead of MTBF) which has been discussed in Chap. 6.

Including all interruptions independently of their origin or nature, helps in defining the extent to which the system (all components of hardware, software, and documentation provided by the supplier) may be dependent upon to deliver service assurance. In a way closely paralleling the examples we saw in Chaps. 4 and 5, with computers and networks service assurance is critically important.

If the end-user entrusts his data to the database, he must not only be assured that this data will be easily available online when he needs it, but also that system interrupts will be transparent to him. Hence, reliability standards must be

established, based on the expected failure rate of system components while system design pays attention to the criticality of each online application to each end-user.

Fail-soft is the way to go but redundancy costs money. Therefore, each supported service should be costed and the cost compared to the benefit obtained through redundancy in terms of uptime. There is no free lunch.

How much end users are willing to pay for much higher reliability? With the executive vice president of one of the largest and best known global banks we projected a network at $R = 99.99$. This institution had the (right) policy of charging the user department for computers and communications services. When the end-users saw the $R = 99.99$ bill they had to pay, they revolted.

It was therefore explained to them that $R = 99.99$ were what they had asked for. However, if they accept "three 9 s" ($R = 99.9$) and fail-soft, then it was possible to reduce the cost by 20%. That is what they chose. Whether the network is designed with four 9 s, three 9 s, or only two 9 s as a goal (less than $R = 99\%$ will be an aberration), its lifecycle reliability has to be assured but as well costed and has to be managed.

The availability of reliability statistics is essential not only for network control but also for the reconfiguration of reliable online systems to meet evolving applications. A properly designed failure detection and correction mechanism should also include a means for diagnosing the failure reasons, isolating each incident to a specific machine, functional block, or component. A databased failure log with MTBF, MTTR, MTOSI, and MTBSI will be invaluable.

Statistics should be kept on all incidents and actions taken for later analysis to determine if particular components are more susceptible to failure than others. A log of intermittent failures might also be used to determine the progressive weakening of a given component and thus permit its replacement before a failure occurs (preventive maintenance).

In conclusion, effective life cycle maintenance requires the organization of a database so that it is able to provide upon request MTBF, MTOSI, MTTR, and MTBSI for each piece of equipment, with classification by type of failure. To ease corrective action interactive approaches should be chosen with knowledge artifacts tracking failures alerting for repairs, monitoring restarts, and registering recoveries as well as the time these require.

7.4 Safety Factors and Safety Margins

Generally speaking, a system component may be unreliable until every mode of failure is known, understood, measured, and controlled. So long as the mode of failure is unknown, or being known it is not understood, or although known and understood it has not yet been measured, the component is unreliable and therefore neither the designer nor the user can depend on its function.

Chapter 6 has provided evidence that since the probability of success of a complex system is roughly equal to the product of the probabilities of success of all the essential components, in order to achieve acceptable system reliability

component reliability must be raised by order of magnitude above that expected from the whole system.

While reliability testing provides revealing evidence, its results may be limited because of lack of understanding in the determination of the *crucial variables* (of the system under test) and its environmental conditions. According to Robert Lusser, a radically new approach should be taken in respect to specifications and tolerances, based on those factors which truly govern reliability. These are:

- The maximum environmental conditions occurring in service,
- The actual modes of failure inherent to the system and its components,
- The ultimate strength with regard to each of these modes,
- The variability of the strength due to operational conditions, and
- The safety margin between average strength and environmental conditions.

In the past, the practice with all engineering professions has been to specify *safety factors*. These were really multipliers of the actual finding of an engineering study and were expected to provide a safeguard against the "unexpected" as well as a protection against human errors in computation and estimation. But really they were a kind of primitive and largely rested on wishful thinking.

Invented multipliers are essentially factors of ignorance and an escape from analytical work. In civil engineering the usual safety factor, which more or less remained unchanged through the ages was taken equal to two. In guided missiles it was set at 1.5. This was selected from piloted aircraft where originally it was used for the estimation of the bulky parts, and then it was used indiscriminately for any component of a guided missile—including electronics.

It is rather obvious that these have been shortsighted approaches since they neglected consideration of the fact that, in an automatic process, vital components must be made several orders of magnitude more reliable than those of man-manned equipment. Based on the statistical concept of spread in regard to the output of any production process, it can be demonstrated that depending on the variability of the product the safety factor of 1.5, 2.0, or 2.5 may be quite unsatisfactory. This is shown in the statistical quality control chart in Fig. 7.3.

Although the mean of the particular variable under consideration is 1.5, a sizeable number of units are below the *reliability boundary*. This does not mean that the safety factor should be generally increased. The latter would result in a penalty to the producers of units with a low standard deviation therefore of higher quality. The right multiplier can be found through proper inspection of lots and systems (Chap. 6).

In addition, depending on the environment in which the component or system will be working, the safety factor may have to be increased, but still arbitrary numbers will not provide the wanted protection. It is much better to use *safety margins* based on well-known concepts of engineering statistics and statistical quality control (see also Part 4 and Part 5).

Going back to fundamentals, to achieve an absolute level of component reliability it is necessary to start by determining with laboratory tests the characteristic variability of lot quality. This involves testing of sufficiently large samples (Chap. 9 discusses sampling). With the obtained data a statistical plot is made of

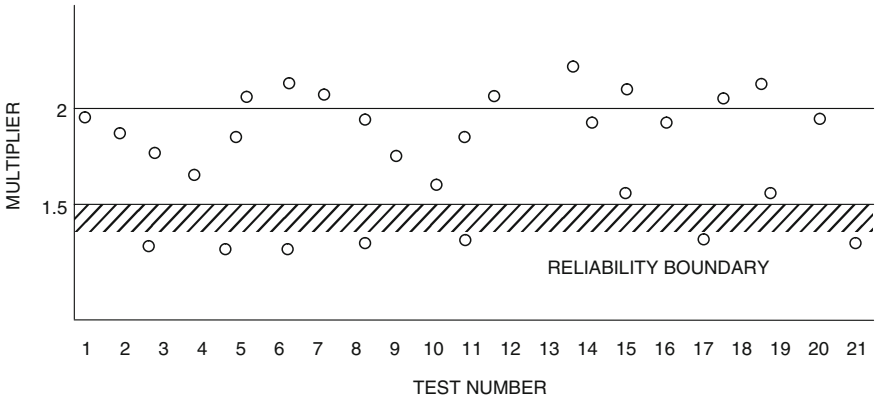

Fig. 7.3 Arbitrary multipliers see to it that the reliability boundary is broken by measurements

strength versus time (test number), basically using the mean and standard deviation.

Figures 7.4 and 7.5 show a sample of alternative cases which may develop in reliability testing. It can easily be seen that a safety factor (multiplier) set by rule of thumb may result in an extremely poor protection and have disastrous consequences in regard to the reliability of the outgoing product. Depending on the population under test, the same safety factor may be:

- Excessive,
- Satisfactory, or
- Plainly inadequate.

In Fig. 7.4 the so-often used factor of 2.0 provides no reliability protection with what is generally considered to be "normal" quality of production. In this particular project (from which come the quality statistics), the safety margin had to be significantly raised. However, in the case of a factory with very tight variance in its produce a factor equal to two is satisfactory.

In the example of Fig. 7.5 even a factor of 4.0 is unsatisfactory for the produce of a manufacturer with lousy production processes and practically non-existent quality control systems. To the contrary, it is excessive for a factory with high quality of production. As shown by these examples. As shown by these examples there is much worth in using of statistical plans rather than arbitrary fixed factors. [As Chap. 12 explains in the discussion on 6σ (Six Sigma) the best safety margin would be equal to six standard deviations from the mean to the reliability boundary.]

As with statistical quality control (Part 5), there is an added advantage in this testing procedure: It offers a good indication of the *trend* and therefore it can be effectively used for predictive purposes. The trend may be toward decreasing reliability, it may indicate no change, or it may be an *improvement trend* like the one shown in Fig. 7.6.

Because safety margins are so important in terms of reliability assurance, the better policy regarding observance of reliability standards is to adopt a test

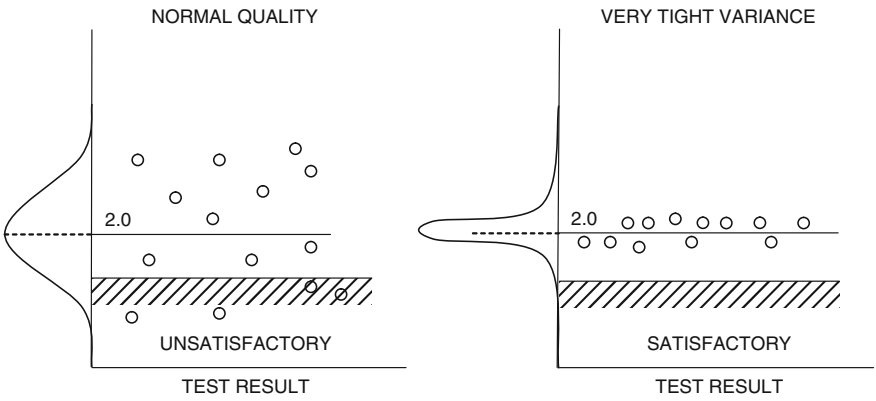

Fig. 7.4 The usual factor of 2.0 provides no reliability protection with "normal" quality

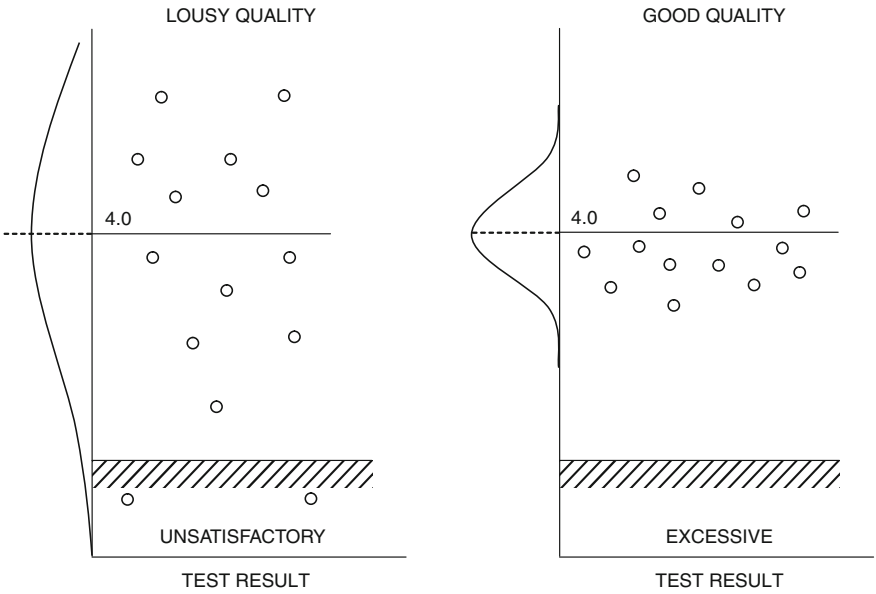

Fig. 7.5 Even a multiplier of four is unsatisfactory with poor quality produce

technique which provides a measure of control over the variability of devices during production. A system of measurements should meet two objectives:

- It must pinpoint processes or products that need corrective action, hence providing assurance that produced devices meet a quantitative reliability goal, and
- It must be reasonably sensitive to favorable or unfavorable trends, answering the question "Is the effort to improve reliability paying off?"

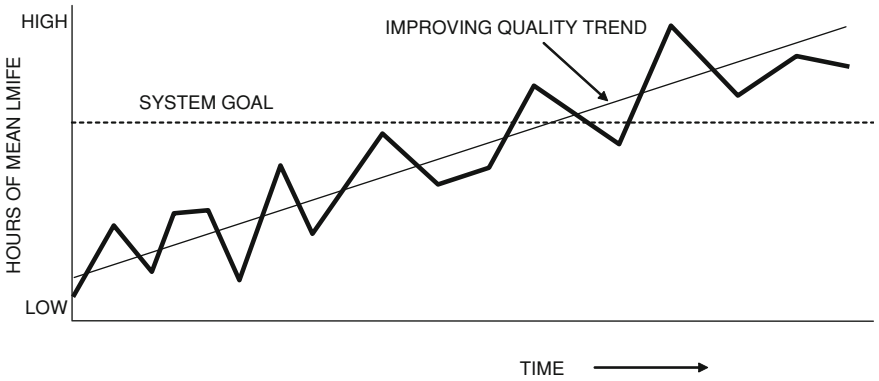

Fig. 7.6 An improving quality trend provides important evidence in reliability engineering

As a general rule, the effort to improve reliability pays off most by concentrating on those units with the lowest mean life value. The measure of mean life and of variance during a time sample can be used for a prediction of the performance that can be expected during subsequent operation, even if this prediction will obviously be an approximation.

A significant difference between reliability testing and the conventional sampling and testing by attributes (Chap. 14) lies in the action taken when the product does not meet acceptance criteria. In the case of a classical quality control chart by attributes the consumer's risk can, at least in theory, be reduced to 0 by 100% screening of the lot—though in practice this is never achieved.

By contrast, because reliability studies imply time sampling, no effective method of screening exists which even theoretically provides 100% assurance. In reliability testing key to the estimate's precision, or sensitivity of the plan, is in the number of failures observed in the time sample.

For reasons already explained, which have to do with the time-sample size, it is necessary to balance the cost of the test against the desire of providing adequate assurance that the required mean life is being maintained. Odd as it may seem, by having a higher reliability less complex units will require a larger time sample for a given precision of estimated performance.

Another fact to bring to the reader's attention is that experience gained by, and results achieved in, time-sampling applications underline the need to reduce potential sources of failure before production begins. This is one of the efforts strengthening the correlation between quality control and reliability studies.

In conclusion, safety factors and safety margins are critical to reliability assurance. Reliability testing is inseparable from time-sampling procedures which provide documentation for reliability specifications. The accuracy of such specifications greatly depends on knowledge of the variability of a production process therefore the use of fixed multipliers for safety margins can be counterproductive.

7.5 Field Feedback for Reliability Assurance

In judging the reliability of components, subsystems and complete systems, specific figures of merit must be named and tests prescribed which enable measurement of compliance at a chosen level of confidence. Much has been stated so far on this subject. What still remains to be discussed is the importance of field feedback to reliability and product assurance.

Let's start with a flashback on system structure. Complex systems can be subdivided into simpler subsystem and elements which are individually analyzed then combined in the overall design. In field use, each of these devices and subsystems, as well as the total system, provide a stream of information which is of fundamental importance to reliability studies. Precious information can be obtained from a well-designed feedback.

Seen from a company's perspective, field feedback is an *internal feedback* with many loops. "Internal" means that its loops never pass outside the boundary of the corporate system—or for that matter of a man-made technological system. A simple example is the temperature control mechanism of the human body. In systems engineering the objective of steady feedback is reliability control and service assurance.

There exist no universal standards on how to structure a field feedback or who exactly should be receiving information elements on quality and reliability. These feedbacks should surely reach the reliability administrator and may also reach the management of the subsidiary, senior management at headquarters or both. Normally, every feedback should be registered in the company's quality and reliability database. This is, however, rarely the case as recipients of feedback reports tend to keep them close to their chest.

This issue of an unwise and counterproductive secrecy regarding quality information has been brought to the reader's attention in Chaps. 2 and 3 While computer aided design (CAD) and computer aided manufacturing (CAM) [5] provide an infrastructure able to promote better communication and integration of effort, they are not always used to promote that objective and the reason is secrecy.

As long as CAD/CAM operates in silos in R&D and manufacturing at the home country and abroad, with field feedbacks filed away instead of being accessible through the corporate memory facility (CMF), their contribution would be minor. Nothing or nearly nothing will be learned from failures—the way it has happened with the Fukushima Daiichi nuclear plants (Chap. 5).

By contrast, first class management will use the field feedbacks to promote quality and reliability thinking as the basic *product assurance* (Chap. 1). All feedbacks are of value whether from product development, manufacturing or field operations emphasizing. Their message can contribute to:

- Issues relating to engineering control,
- Factory-oriented quality assurance control operations,
- Customer service subjects problems from field maintenance, and

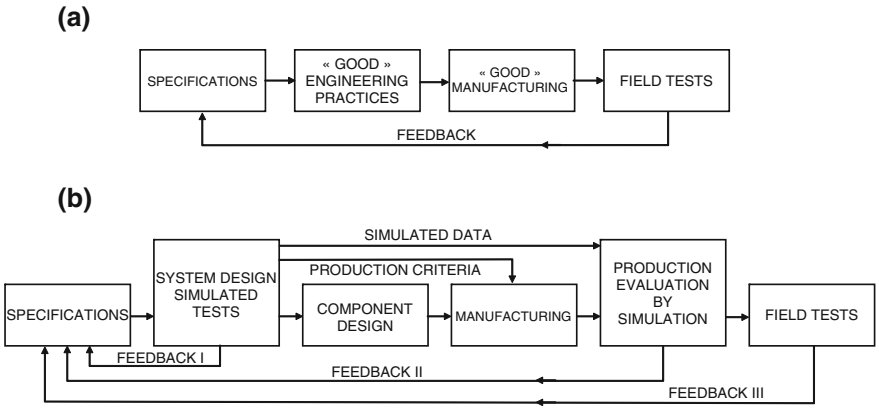

Fig. 7.7 The feedback loop must be much more sophisticated than in old times. **a** A simple feedback with limited effect, **b** A nest of feedbacks with open communications line

- Field maintenance activities, including cases which require follow-up and handholding by marketing.

Cornerstone to the successful implementation of this method is open channels of communication which facilitated the transmission of an unbiased feedback. The notions of a narrow and of a broader view of a feedback for quality and reliability reasons must be reinforced. Feedback loops to development, production, and field service are illustrated in Fig. 7.7.

A so-called *good practice* shown in Fig. 7.7a has been classically employed in many cases in connection to the development and production stages. However, with the advances which have occurred in the state-of-the-art over the past years, the requirements for what constitutes "good practice" in each technical area have changed.

A major improvement in regard to product assurance and reliability is presented in Fig. 7.7b. This is a multiple feedback replacing a lot of full redesign cycles by multiple prediction of the probable results of tentative design changes. Generally speaking a modern solution may feature two different feedback types:

- A *tight feedback* which rapidly corrects any deviations or errors. This should be directed to the executive in charge of product and service assurance.
- A *loose feedback* permitting marked deviations from steady state (defined by a safety margin) before initiating a control.

Tight feedback is necessary for technical personnel. Loose feedback is usually oriented to general management, and the increased lead time to control action is the result of consultations prior to decision. Notice, however, that the concepts of feedback and reinforcement of action are related.

A documented corrective action necessitates feedback of the operational result of a system's, subsystem's, or device's behavior. Although feedback theories generally concentrate on input/output relationships, the case of reinforcing

reliability is different because it incorporates environmental and other operational factors. Data on most vital *reliability measurement* come from:

- Structured failure reports,
- Failure data by first-hand field observation,
- Correlated data from two or more sources for checking failure report validity,
- Correction factors for application to failure reporting, and
- Analyses of secondary failures (failure of a part caused by malfunction of some other part).

Another group of vital feedback data is *severity level factors*. These include the severity level of all parts, devices, subsystems; severity levels versus failure rates for each parts category; and evaluation of correlations between levels of severity and failure rates. Severity level curves for each parts category tend to show the same or similar operating characteristics:

- Failure rates increase with severity level,
- A danger signal is a sharp rise in failures, as the severity level approaches 100% of rating, and
- The failure rate flattens as the severity level approaches zero; thus indicating that the failure rate of parts would not drop to zero even when operated at 0% rating.

Corrective action aside, another objective to be reached by means of appropriate field feedbacks is *reliability prediction*. From drawing-board design, or measurement in experimental model(s), we need to determine severity level of each part. From previously determined severity level versus failure-rate curves, we can estimate the failure rate for each part. Failure rates of each individual part must be integrated to obtain the overall predicted equipment failure rate. From the predicted equipment failure rate we calculate the predicted mean time between failures at a given level of approximation.

To my knowledge, this method has been originally developed in the mid- to late 1950s by the Vitro Corporation under contract to the US Navy, Bureau of Ships. Its goal has been measuring and predicting reliability of shipboard electronic equipment; its application, however, is much wider.

References

1. Chorafas DN (1960) Statistical processes and reliability engineering. Butterworth, London
2. Robert Lusser (1952) Reliability administrator at redstone arsenal and designer of the V2 flying bomb rocket of WWII. Planning and conducting reliability test programs for guided missiles, Bureau of Aeronautics, Technical Memorandum No. 70, GSR No. 30
3. Weibull W (1951) A statistical distribution function of wide applicability. J Appl Mech 18(3):292 [This article also indicates a means of estimating Weibull parameters by eye by a double logarithmic (natural) transformation]
4. Chorafas DN (1960) Statistical processes and reliability engineering. D. Van Nostrand, Princeton
5. Chorafas DN (1987) Engineering productivity through CAD/CAM. Butterworth, London

Part IV
Statistical Inference

Chapter 8
A Brief Introduction to Stochastic Thinking

8.1 Probabilities

The word *probability* has a Latin root. It is a combination of *probare*, which means to test, prove, or approve, and *ilis* which stands for "able to be". In the late seventeenth century, Jacob Bernoulli, the Swiss mathematician, said that what characterizes probability is the *degree of certainty*. Bernoulli added that this differs from absolute certainty as the part differs from the whole—a definition which is at the heart of *stochastic thinking*.

Jacob Bernoulli also contributed to *sampling* theory and the process of *conjecture*—which aims at estimating the whole from the parts. Gottfried Leibniz commented to Bernoulli's approach by saying that nature has established patterns originating in the return of events, but only for the most part ($\omega\sigma\ \varepsilon\pi\grave{\imath}\ \pi o\lambda\upsilon$). Jacob Bernoulli's theorem for a posteriori calculation of probabilities is known as Law of Large Numbers.[1]

Another of Jacob Bernoulli's contributions has been the principle underlying *scientific experimentation*. At the core of it has been the repetitiveness of experiments. To use his words: "We must assume that under similar conditions, the occurrence or non-occurrence of an event will follow the same pattern observed in the past" [1].

The early eighteenth century saw another mathematical genius: Daniel Bernoulli, a descendent of Jacob. One of his important contributions has been the notion of *expected value*; another one, his statement that *utility* is inversely related to the quantity of goods previously possessed. (Daniel Bernoulli has as well been a precursor of concepts related to *human capital*. His was the first ever expressed thesis, outside the ancient world, that tangible assets and financial claims are less valuable than productive capacity).

What a difference a century makes. In the seventeenth century Jacob Bernoulli spoke about the degree of certainty—and by consequence about uncertainty, the staple food for thought of a stochastic thinker. In the eighteenth century Daniel

[1] Not to be confused with the Law of Averages.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering, DOI: 10.1007/978-1-4471-2966-0_8, © Springer-Verlag London 2013

Bernoulli applied his ancestor's notions to derive a concept of *utility*, proceeding from there toward tangible assets, productive capacity, and financial claims.

Certainty and uncertainty are not absolute concepts; they are relative to a situation, event, or decision. The variation in degree of certainty is created by aleatory, hazard, or chance events. *Aleatory* is also of Latin origin, coming from *alea* a word referring to games of chance, like those of an honest dice.[2] The word *hazard* derives from the Arabic *zahr* which means "dice".

- An honest dice is a game of *chance* and
- Games of chance should not be confused with other games where the skill of players weights on the outcome.

As these historical references demonstrate, the concept of an event's probability (more on this later) and its practical notion—that of stochastic calculations—preceded by a score the notion of *statistics* and its two meanings. Statistics (noun, singular) stands for collection, compilation, analysis, evaluation, and interpretation of data. Statistics (noun, plural) refers to a collection of data stated in numbers and arranged in tables. In this book we will be concerned with statistics noun, singular.

Books on the history of mathematics credit Giambattista Cardano, a physician who lived in the sixteenth century, with compilation and analysis of statistics to derive intelligence. Cardano published "*Ars Magna*" (Great Art) in 1545, and he is the first person know to have put black on white the concept of probability.

Cardano is famous for his statement: "A man is nothing but his mind. If that be out of order, all is amiss; and if that be well, the rest is at ease." His interests centered on games of chance. While many other people had similar interests they did not put their thoughts in writing. It is remarkable that it took nearly a century till the next notable book on probability was published, written by Christian Huygens.

The seventeenth century was rich in probabilistic mathematics. The roster of great names includes: Blaise Pascal, Pierre de Fermat, Chevalier de Méré, Gottfried Leibniz, and Isaac Newton. Pascal and Fermat provided a systematic method for calculating the probability of future events. One of their associates, Antoine Arnaud, is credited with:

- The idea of developing a *hypothesis* (Sect. 8.5) from a limited set of facts and
- The test of hypothesis opened the way to *statistical inference*, inferring a global estimate from a sample of data.

Typically, though not necessarily always, statistical inference is *causal inference* (Chap. 12). In the analysis of statistical data we are very frequently searching for cause and effect. Statistical methods do not eliminate chance variation, though sometimes analytical findings permit to develop means for controlling it.

Control demands knowledge and in statistics (noun, singular) this knowledge is acquired through *stochastic thinking*. William Petty was an Englishmen who,

[2] One which is not loaded and, therefore, its six sides have equal probability of showing up.

in mid-seventeenth century contributed a great deal to stochastic thinking. Like Cardano, Perry was a physician by training; he also was surveyor of Ireland and professor of Anatomy. His book *Political Arithmetic* is generally considered as having set the foundations of economics and political science.

Still another of the great names in stochastic thinking is Edmund Halley, who at age 35 was one of England's most distinguished astronomers. (He is also credited of having persuaded Isaac Newton in 1684 to publish his *Principia*.) It is remarkable that Oxford University has turned Halley down for a professorship because he held "materialistic views" which did not match the university's religious orthodoxy. Little did the Oxford mainstreamers appreciate that:

- The probability of an event transcends religious orthodoxy, and
- The calculus of probabilities requires the existence of materialistic views, as shown by both definitions of the term.

The more general probability definition states that if a large number of trials are made under the same condition, the ratio of the number of trials in which a given event may happen to the total number of trials approaches a limit as the number of trials is indefinitely increased. This *limit* is the probability that the event will happen under that condition. Under this definition the probability of an event is determined by making a series of experiments and therefore it is never resolved firmly.

- As more trials are made, the relative frequency may change, and,
- *If* the relative frequency changes, *then* our estimate of the probability will change.

The advantage of using this definition is the absence of a need to make the assumption that the events are equally likely. Take as an example the probability of obtaining a head from flipping a coin. Say that one performs $n = 100$ trials, and the number of heads is 51. The probability is given by the relative frequency $\frac{51}{100}$, or 51%. As more trials are run, the relative frequency will approach the true probability value.

The alternative definition of probability is more specific, based on the algebra of sets. A population consists of black cows "B" and red cows "R". These two are mutually exclusive and equally likely to occur. The probability P of randomly selecting a red cow is $\frac{P_R}{P_R+P_B}$. This is the ratio of the number of red cows to the total number of cows. What underpins this definition is the "equally likely" and "mutually exclusive requirements".

Notice that while it does not explicitly make that assumption, the "events" in the first definition: "heads" and "tails" are also mutually exclusive. A coin cannot fall on both sides at once. In the case of an honest coin they will as well be equally likely. There is an equal probability a head can occur, $P_{h,}$ and an equal probability a tail will occur, P_t. Hence the probability of flipping a head is $\frac{P_h}{P_h+P_t} = \frac{1}{2} = 0.50$.

In the example with this probability definition, with $n = 100$ trials, the probability of heads was equal to 0.51 only because the number of trials was relatively

small. In the long run with an honest coin the probability of number of heads to be expected is 0.50. Because, however, this is not necessarily the outcome with a given sample (Chap. 9) it can be said that the second probability definition is a special case of the first, while because it involves uncertainty the first is nearer to stochastic thinking.

In conclusion, probability describes the outcome of a large number of events essentially the same; not the outcome of a single event. The probability of an event is always between 0 and 1. In addition, the probability of an event not happening is one minus the probability of the event happening: $P_{\text{not A}} = 1 - P_{\text{A}}$.

8.2 Stochastic Method

Stochastic thinking underpins the method of science and, by consequence, of all analysis which is factual and free of prejudice. There have existed, however, over the years different ways of looking at science—I will call them "right" and "wrong"—as well as its methods and its tools.

A wrong way of looking at science is to believe that scientific proof is a matter of showing formal consistency with a set of *self-evident* definitions, axioms, and postulates of a given system of thought. Not only this is the false way, but also the effect of believing in such a procedure of unchallenged formality, impacts in a negative way the mind of scientists because it leads to denying the existence of anything outside the bounds of that formal system.

The right way of looking at science, as well as at analysis which is the most powerful scientific tool, is to abstain from reference to self-evident definitions, axioms, and postulates depending on experiments. The experimental method requires that we scientifically validate our hypotheses regarding cause and effect or physical principles.

In the background of the scientific method and its experimental foundations, lies the fact that in the real universe there are no fixed sets of "self-evident" truths, definitions, axioms, or postulates. Researchers typically operate on the basis of hypotheses (Sect. 8.5), which they have assumed to be sufficient up to a point but, as we already saw in Part I in connection to product assurance, these researchers are:

- Eager to challenge the "obvious", and
- Open-minded about discovering that some of their assumptions might be false, while other principles they had not known are the determining notions.

As a basic principle of scientific research—and therefore of invention and discovery—the most important class of investigation is that of being alert to evidence of changes needed in our assumptions. Many beautiful theories have been dropped over the centuries because of the discovery of one or more facts which contradicted them.

The open mind is a most precious asset of stochastic thinking. To inspire confidence, scientific analysis should be designed and executed with great

accuracy and objectivity. The existence of bias destroys it. Discovery through experiments often brings up data which show unexpected relationships. It is not by chance that the scientific method's two pillars are:

- At one side, stochastic thinking which makes possible investigation, and
- At the other, a systematic collection and classification of observations on facts and occurrences as the ground for inference.

As an elementary example of the work of a statistical analyst consider a process of statistical quality control (SQC, Chap. 12). A random selection of 20 lamps is made from the manufacturer's daily production lot and they are subjected to destructive testing. The test measures the *mean life* of the sample and this is 1,110 h. However, it is not unlikely that the sample from tomorrow's production will have a different mean life (sampling procedures, as well as the opportunities and risks associated to them are explained in Chap. 9.)

In industrial practice not only the sample mean but also the distribution of life test values has to be critically examined, with questions arising about the best estimate of *outgoing quality*. The *variance* of the sample—which is an estimate of the variance of the underlying population—helps in establishing the *limits* within which the estimate is likely to vary

- Typically, these limits are taken at ± 3 standard deviations from the mean, a measure of dispersion deriving its legitimacy from the hypothesis of normal deviation of lamp lives.
- Because of a random impact on the process of lamp production, however, the distribution may be skew or have a long leg with outliers at 5, 10, or 15 standard deviations from the mean.

Therefore, prior to providing himself with a ground for inference on the lamps' life, based on prior samples and the most recent one the analyst must test the hypothesis of normal distribution. Then, using an *operating characteristics* (OC) curve (Chap. 10), he must select the level of confidence to characterize his inference. Moreover, in case he is uncertain about the shape of the distribution and/or the key factors influencing it, he will be well advised to proceed with *experimental design* (Chap. 11).

There is no shortage of well-established statistical methods. But while statistical methods may be employed to advantage, the way the mind of the scientist, experimenter or analyst works will usually impact on the finding. Objective results are frequently influenced by somewhat subjective considerations. Contrary to what is written in typical books of statistics, and sometimes of science:

- The inferences being made may resemble a fuzzy set.
- They are not a priori objective clear-cut answers supported by firm numbers.

This is important inasmuch as stochastic thinking goes beyond probabilities to include asymmetries and other notions (Sect. 8.3). The important concept underpinning this extension is *uncertainty* which has created the base of several new

concepts, including the proposition that in the minimal regions of space–time a given particle will not have a precise location.

To find the beginning of this approach one must return to the early 1930s when, under the influence of quantum mechanics, the idea of discontinuity of space and time took a new form though propositions which excited the imagination of scientists for centuries. For instance, the S metric of Werner Heisenberg, the physicist, is an operator that permits us to describe the state of a system after diffusion and do so effectively, provided we know the system's state before diffusion.

This emphasis on *initial conditions* provides the starting point to the study of a transient process characterized by *uncertainty*. Inference based on *belief* functions [2] utilizes a method of treating essentially subjective information—therefore, ill-defined and ambiguous. Such information is typical of human cognition and reasoning. Belief functions, expressed through *possibilities* should not be confused with probabilities.

- Probabilities are based on objective measurements (as shown in Sect. 8.1), and they are crisp. The sum of probabilities of two events P_R and of its opposite P_B is equal to one: $P_R + P_B = 1$.
- On the contrary, with possibility theory an event may have a possibility equal to 1 and its contrary event can also have a possibility equal to 1 (or any other value from 0 to 1), while the sum of them can be greater than 1.

Belief functions are most important in analysis and have been used in evaluating a *hypothesis* given an *event*. A measure of belief and one of disbelief are key to nearly all decisions with the inference made empirically, without the formal testing of hypothesis which is done in probability theory.

With belief functions *low*, *high*, and *reasonable* are qualifiers. *Very*, *quite*, and *about* are hedges that can be used with qualifiers. Noise words are phrases such as *should be* that dilute the impact of policy statements without necessarily affecting their general meaning. Make no mistake, however, plausibility concepts are closely linked to the mathematical infrastructure designed to deal with a more realistic world in which:

- Information need not be manipulated so precisely and
- Judgement and/or evaluation can be only approximate.

By employing fuzzy logic, the resulting *heuristics* (noun, singular) has a much greater chance of accurately following a perception prevailing in the real world which is a product of subjectivity and objectivity at various degrees. Such approximate reasoning relies on qualifiers, linguistic variables, plausibility concepts, and a new mathematical infrastructure (more on this in Sect. 8.3).

Stochastic thinking based on possibility theory is a scientific method. Its application is most important when we cannot describe the process of a weighted decision in crisp terms. Fuzzy engineering works by *qualifiers* which are not black or white but shades of gray that is intermediate values. In many cases in quality

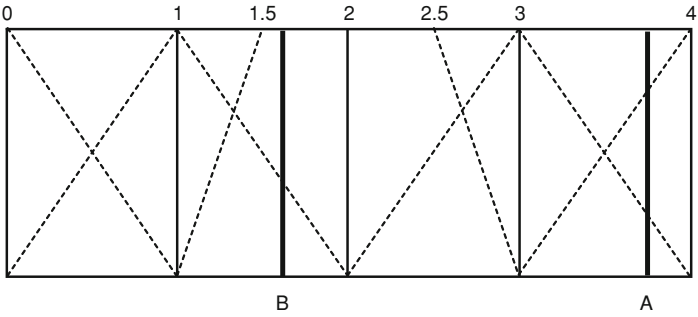

Fig. 8.1 Grading students A and B through fuzzy

control (QC) we only possess information knowledge which can be fuzzy. Figure 8.1 presents an example from education. In grading the final exams of students A and B.

- A grading system is based on causal evidence: in education this is the exam; in QC, it is the statistical chart or result of a test.
- Analysis should be able to capitalize on uncertainty. It will be a failure if its effort is full of monocausal, linear thinking.

As a tool of analysis, fuzzy-set thinking is best suited to flashing out complex causes. Precisely because many tests and decisions thought to be objective have a subjective quotient, fuzzy engineering proved so successful in many areas of activity. The Japanese were the first to develop fuzzy chips with which the equipped lots of equipment from house appliances range to learning a car owner's driving habits in order to optimize gas consumption.

In conclusion, the horizon of stochastic thinking has been enriched with a new mathematical tool: Fuzzy engineering, which has found a wide domain of applicability. Learning how to work with fuzzy sets is easy business for an open mind. The reader must however be aware that there exist conceptual differences when fuzzy engineering is compared to statistics, as well as the mathematical difference inherent in *possibilities versus probabilities*:

- An event which has a probability equal to 1 is considered certain, in which case the contrary event has to have a probability of 0.
- This is not true of an event which has a possibility equal to 1 since the contrary event can also have a possibility equal to 1.

In a way, probability theory is a subset of possibility theory in which the range of variation in outcome is fixed a priori. If an event has been given a belief, necessity, or certitude equal to 1, *then* it can be considered certain. The necessity of the opposite event is then set to 0. A measure of belief and one of disbelief are key to QC and reliability studies when data is not crisp. In those instances, belief functions can be effectively used in evaluating a *hypothesis* given an *event*.

8.3 Asymmetries, Fractals, and Complex Numbers

Not all engineering schools teach their student stochastic thinking. Rather they focus on conditions characterized by the normal distribution (Sect. 8.4). Translated in statistical terms, this means that they are strongly in favor of the theory which says that events in life follow a bell-shaped curve. This is a questionable policy because "normal" cases and "normal" distributions of events are just that, theories. They do not usually happen in real life. What is known as the "normal" curve is based on two premises:

- That the central tendency of the distribution has the highest frequency and
- That the distribution of events or measurements is symmetric around this expected value.

Neither is a law of nature, or for that matter of the behavior of man-made systems and of social aggregates. Dr. Milton Friedman, the economist, once said that the average expected return on higher education may be high, but there is wide variation about the average (central tendency).

This wide variation has important implications. For instance, in education it suggests that the broad distribution of outcomes raises doubts about the post-World War II principle of higher education for all. Moreover, the distribution (of returns from higher education or of any other variables we care to measure) may not be symmetric but skew or kyrtotic as certain factors are weighting much more than others.

In a statistical sense, there is no *a priori* reason to believe that a distribution of events, measurements, or results will be symmetric; therefore, bell-shaped. Symmetry will be the exception, because the prevailing rule in nature, as well as in obtained outcomes of human effort, is *asymmetry*—another of the characteristic qualities of *stochastic thinking*.

"If we envisage all of nature's creations in the mineral, animal and vegetable worlds, and we also consider man-made artifacts," said Louis Pasteur, a molecular physicist, biologist, researcher, and one of the greatest scientists of modern times, "we will see that they belong to two great classes: some have a sense of symmetry, while others don't" [3]. This has a profound meaning for human activities—from physics and engineering to finance.

Pasteur took as an example of objects exhibiting mathematical symmetry: the human body, a dice, a table. He then pointed out that there are other objects and parts of objects which lack these characteristics. Taken as a whole the human body exhibits symmetry *if* a vertical plan passes through the middle of the nose, but the parts themselves, which constitute either side of such symmetric aggregate, lack symmetry.

The image of objects lacking symmetry is not superimposable to reality. If a chair is placed in front of a mirror, the image being reproduced is symmetric and can be superimposed to the chair. The same is true of the human body as an aggregate. But the mirror image of a hand is not superimposable to the hand,

because it is not symmetric. This is a fundamental design principle affecting the quality assurance of man-made devices and systems.

The reference made in the preceding paragraph particularly impacts products with moving parts. A storage tank is symmetric. A heat pump's subsystems are not. A boiler may be seen as symmetric (though this is not necessarily true in all cases). An automobile is not symmetric. The driver's seat is on one side (the left side in America and in continental Europe, the right side in Britain and Japan).

The mineral and artificial (man-made) products which are *nature morte* present symmetry. By contrast, vegetal and animal entities formed under the influence of life are atomically asymmetric, and as we have seen the same is true of artifacts characterized by motion. That lack of symmetry has in its background what Pasteur called *the force of deviation* of the polarization plan:

- Dead nature is symmetric.
- Objects under the creative influence of *becoming*,[3] have an internal asymmetry.

What might be the reasons for such most significant design differences? Pasteur maintained that they can be found in nature's molecular forces which are present and act. For instance, in vegetables under the sun's influence. Also, quite probably, they are due to certain asymmetric phenomena of the universe even if these, themselves, may be dissymmetric.

It is indeed difficult to find a more profound separation of living matter from dead nature, than this asymmetry in part of the world around us which is absent in non-living objects. Precious lessons can be learned from molecular biology which has greatly benefited from Pasteur's ingenious analytics. An asymmetry in the internal arrangement of a chemical substance manifests itself in its external properties, which are capable of asymmetry.

Such a most fundamental issue, which the molecular physicist has discovered as a dichotomy dividing the world of minerals and artifacts from that of plants and animals has not been studied in a deep sense as to the way in which it affects engineering design and quality assurance. Yet, there are precious lessons to be learned from it because it derives from millennia of evolution and, compared to them, experience in engineering design is only a trifle. *Designing for quality* can be largely improved by thinking about:

- How to capitalize on asymmetries and
- How to make the best use of *becoming* events by means of stochastic thinking.

Chaos theory [2] teaches an excellent lesson in this regard. The tone is given by chaos theory's inventor Jules-Henri Poincaré (1854–1912), who is famous for his saying: "What is chance for the ignorant is not chance for the scientist. Chance is only the measure of our ignorance." But as Louis Pasteur had it "chance favors the prepared mind."

[3] *Du devenir* in Louis Pasteur's words.

Mathematician by training and inventor of *fractals* theory, chaos theory's alter ego, Benoit Mandelbrot, made one of the most interesting contributions with a domain of applicability ranging from science and engineering to economics. He believed that not only coastlines and the shape of clouds but as well a wide range of events in other domains, like financial market movements, have fractals form. They are not falling under the familiar bell shape of the normal distribution.

Therefore,

- From specifications and tolerances of engineered products,
- To trading practices, and all sorts of financial models,

designs based on the assumption of a normal distribution are wrong. Indeed, the 2007–2012 economic and financial earthquakes proved that Mandelbrot was right. His set of fractal happenings (or of their measurements) is a collection of points in a complex number plane. The formula is:

$$z_{n+1} = z_{n2+c}$$

where

c is a complex number,
n represents the digits $1–\infty$, and counts the number of times the calculation has been performed.[4]

The fractals domain is a complex number plane. In "Alice in Wonderland", Louis Carroll (pseudonym of Charles Ludwig Dodgson, a young Oxford don and a mathematician), asked Alice to believe as many as six impossible things before breakfast. Mathematically speaking these are in the complex plane space on which all numbers, real, imaginary, and combinations of the two, can be plotted.[5]

"Normal", let alone "obvious", events do not exist in this complex plane; but one should not be discouraged by the fact that the challenges associated to our understanding of such an environment are so great. The good news is that these challenges:

- Help in shaping up a creative mind, and
- In their domain may lay solutions to complex problems associated to quality assurance and reliability.

Mathematicians have worked for centuries with the question of what multiplied by itself gives the answer -1, till Leonhard Euler, suggested that the best way to deal with the problem was to invent a new symbol, i and work on the consequences.

---

[4] z starts as any number one likes to choose, and changes with each calculation. The value of z_{n+1} is used as z_n the next time round.

[5] A real number is the familiar sort from normal arithmetic. An imaginary one is a multiple of the square root of -1.

Carl Friedrich Gauss, Euler's successor, discovered that if we plot real numbers on one axis of a graph and imaginary ones on the other, we create a plane that represents both sorts of numbers. This permitted the plotting of complex numbers, which have a real and an imaginary part—a concept which largely extended the scientist's knowledge frontiers.[6]

The references just made to conceptual and mathematical complexity should not just be skipped over because ignorance, plain ignorance, can be found behind the majority of quality assurance problems. Engineers are not sufficiently taught about the need to challenge the "obvious" (or are somehow afraid of doing so). To really challenge the "obvious" one must all way leading to use stochastic thinking.

For instance, what many people consider to be an unchallengeable "obvious" principle, is that the more information they have the more certain they are about their decisions and actions. Kenneth Arrow, one of the few people with the intelligence to appreciate that information needs are often overestimated, says that this is nonsense. In his words, vast ills have followed a belief in certainty, whether:

- Historical inevitability,
- Grand diplomatic and military designs, or
- Extreme views in economic policy and investments

The same is true about engineering designs and the search for product assurance. In Arrow's opinion, long range forecasts are no better than numbers pulled out of a hat. For his part Peter Drucker, one of the fathers of modern management, advised that the best contribution of a forecast is that it offers an opportunity to judge the future impact of present-day decisions.

In conclusion, asymmetries, fractals, and complex numbers are the real models of living and inanimate nature. Therefore they constitute basic mathematical tools in the quest for quality and reliability. They are the true approximations to reality rather than the familiar bell shapes of normal distributions that Gauss first described. But as we will see in the following section there are reasons why we are still working with the normal distribution even if we know its shortcomings.

8.4 Which Might be the "Normal" Case?

Compared to asymmetries, fractals, and complex numbers working with an orderly symmetric distribution of measurements or events—like the one shown in Fig. 8.2—is an example of simplicity.[7] But it lacks the intellectual challenge, an

[6] Which, characteristically, do not lie on either axis.

[7] The population of events within this bell shaped curve is distributed *approximately* as follows: 68% within plus and minus 1 standard deviation (s) from the mean $\bar{x}$; 95% within plus and minus 2 s from $\bar{x}$; and over 99% within plus and minus 3 s from $\bar{x}$.

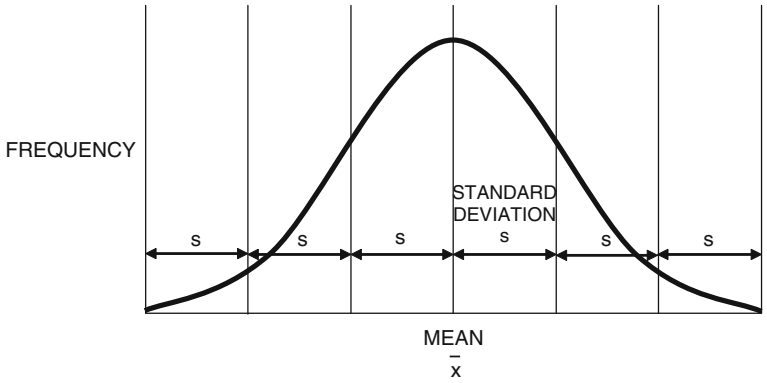

Fig. 8.2 The normal distribution is an approximation of a distribution of events encountered in real life

engineering mind should actively look after and it is only an approximation of a situation encountered in real life.

In the course of the analysis of the problem he is confronted with, the designer or quality expert may decide to use the normal distribution curve rather than a more sophisticated approach like stochastic thinking. The choice frequently depends on company culture, skills, and circumstances than on a priori reasons—provided one appreciates the impact an approximation has on quality assurance.

Some people might say that by working on the hypothesis of normality the problem is on its way to be solved. But the opportunity may be lost. "In life we do not have problems we have opportunities," Harold D. Koontz, my professor of business policy at UCLA taught his students. There exist however cases, and therefore reasons, why the normal distributions curve is helpful.

Historically, the concept underpinning the normal distribution curve has more than one father. One of the notable minds is Abraham de Moivre who in 1725 drew attention to the fact that, when traced out as a curve, the distribution of events shows the highest frequency clustered to the center close to the mean—then it slopes symmetrically downwards ("symmetrically" is, of course, an approximation).

De Moivre's work significantly contributed to the mathematical description of "normal" cases. Among other things he is credited for having established the standard deviation, s, as the square root of variance. A distribution's first momentum is its mean or central tendency: $\overline{x}$, where x_i are individual measurements. Its second momentum is its variance, necessary for measuring the dispersion of events or observations around the mean.

Carl Friedrich Gauss is another of the normal distribution's fathers. He lived in the early to mid-nineteenth century and his contribution to statistics saw to it that the bell-shaped curve of the normal distribution is known as Gaussian. A contemporary to, though 29 years elder than Gauss, was Pierre Simon de Laplace who in 1812 published his "*Théorie Analytique des Probabilités*". Laplace was first to

assume that there are laws, similar to the physical laws, governing everything else, and to his judgment this included events as well as human behavior.

Like Isaac Newton who lived a century earlier, Laplace was free of prejudices. He had a clear and independent mind, and this reflected itself in his work. A proof is that when he presented his "*Mécanique Céleste*" to Napoléon, the emperor asked him why he did not use a reference to God and Laplace answered: "I did not need that hypothesis". Years later, in Victorian England, following on the work of Laplace and Gauss, Francis Galton[8] developed the statistical tools for distinguishing between:

- Measurable risk characterized by normally distributed events and
- Uncertainty, which involves guesses about what future events will be and does not follow a bell-shaped curve.

Galton's work increased the sophistication of a distribution of measurements or events by introducing the element of uncertainty. One of his famous sayings is "whenever you can count". His policy has been that of testing his ideas through experimentation, which led to new advances in statistical theory. Two other of Galton's famous statements worth recording are:

- Eminence does not last long and
- Mediocrity always outnumbers talent.

It takes talent to develop new departures and that is what Laplace, Gauss, Galton, and others contributed in the eighteenth and nineteenth centuries moving forward the science of statistics. In these early days, however, events obeying the laws of the normal distribution seemed to fit within ± 3 standard deviations from the mean (Fig. 8.2). Today we know that this is far from being the general case. Many distributions, particularly those most challenging from an analytical perspective are not "normal", while the bell-shaped curve is a special case.

Figure 8.3a provides a practical example from risk analysis in a financial project. This distribution of risk events had a long leg with spikes. Figure 8.3b comes from a QC project. The distribution is bi-modal and manufacturing took some time till to make the necessary adjustments so that produced goods observe engineering tolerances. Based on real life events Fig. 8.4 expands of the pattern shown in Fig. 8.3a, b by emphasizing the long leg of events adjunct to the normal distribution:

- The better known events with which we are familiar might be approximated by a bell-shaped curve, falling within ± 3 *s*.
- Relatively unknown events or factors whose existence we contemplated, or on which we experimented might fall within, say, ± 10 *s* from the mean.

[8] Francis Galton and Charles Darwin shared the same grandfather, Erasmus Darwin.

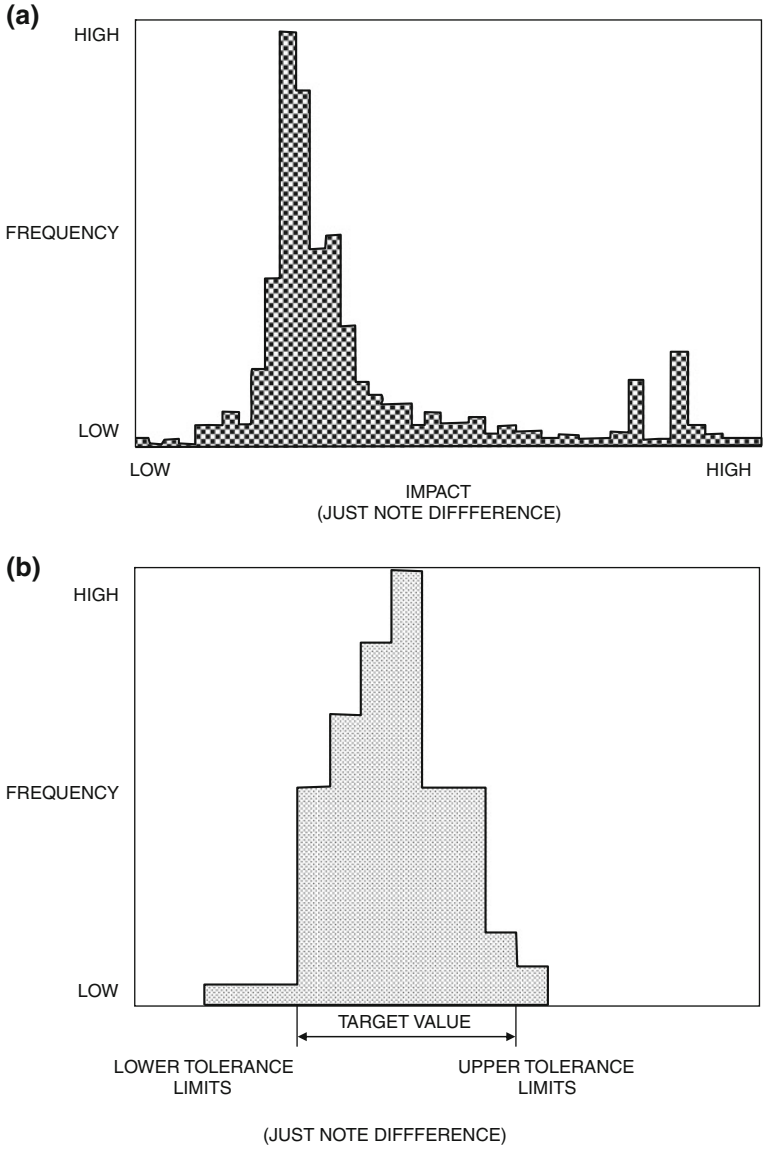

Fig. 8.3 **a** Practical example from real life risk analysis: distribution of risk events with long leg and spikes. **b** A practical example from quality control measurements, biomodal distribution with measurements falling outside tolerances

- But what Donald Rumsfeld, the former US Defense Secretary, called "unknown unknowns" would by all likelihood fall within $\pm 25\ s$ or further out. The October 1987 market panic was a 14.5 *s* event; 21 years later, in 2008, we experience 25 *s* events.

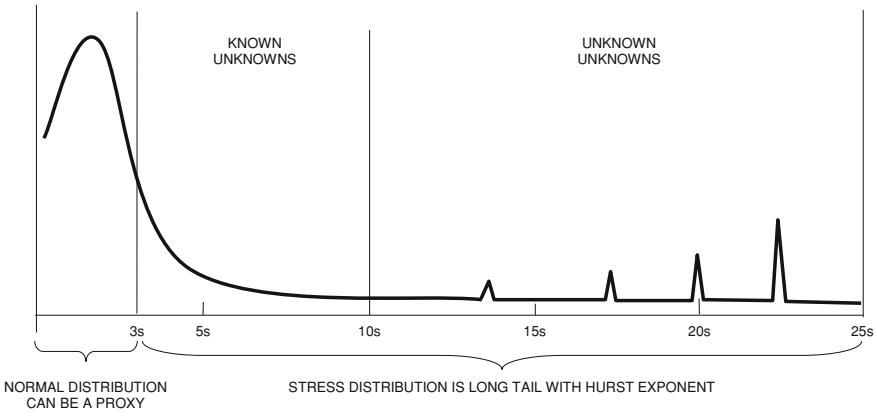

Fig. 8.4 The normal distribution is a proxy valid only in connection to low impact events. The distribution of high impact events is not normal

Stress tests [4, 5] permit to guestimate the distribution's long legs. In principle, the further a measurement falls from the central tendency of the bell-shaped curve the greater is the number and impact of unknown and/or the higher the degree of complexity. In Fig. 8.4 the challenge is found in the outliers beyond 15 *s* from the mean.

That extraordinary events do not fall under the bell curve was not an alien concept to the fathers of the normal distribution. Some of them had pointed out that we cannot understand current phenomena without systematic analysis and evaluation of earlier events which (a) affected the present and (b) continue to exercise profound effects on the future. By applying the founders' principle, we see that we have no other option than to deepen and amplify the use of existing statistical tools in appreciation of the fact that:

- The problems we are confronted with continue changing and
- Because of innovation, designing for quality has many more variables than ever before.

This means that the tools we use must be commensurate to the challenge, which diminishes by nothing the importance of past breakthroughs. The hypothesis that past events are normally distributed, hence they obey a standard form, helped mathematicians, physicists, and engineers who followed in this track in developing invaluable statistical tables. Another reason for using the normal distribution curve, while we do know that it only provides an approximation, is the culture of an orderly approach which developed over the centuries in scientific analysis.

An orderly procession presents advantages for friends and foes. Here is an example. During World War II the Allies were worried that a new German tank could keep them from invading continental Europe. Intelligence reports about the number of tanks were incomplete at best, and at worse contradictory. Therefore,

statisticians were asked if they could make a contribution to these uncertain estimates.

The statisticians started by assuming that, as methodologically behaving people, the Germans had numbered their tanks in the order they were produced. Based on this hypothesis they used the serial numbers of captured new generation tanks to estimate the total production:

- The number they came up with, 256 a month, was low enough for the Allies to go ahead.
- After the war ended and German records became available, it was thumbs up for the statisticians. The records showed that number to be 255 [6].

Apart from statistical tables which permit to know the area under the curve once we know the distance from central tendency, as well as the culture of an orderly processing, the flexibility of the statistical method permits us to correct the bell curve's shortcomings. An example is the kyrtotic distribution allowing for similar events repeating each other in a way nonconforming to the principle underpinning the bell-shaped symmetric curve; such as:

- Floods happening year after year or
- Sharp price swings exhibiting a tendency to cluster.

Known as Hurst-coefficient, after the English engineer who at the beginning of the twentieth century studied the floods of the Nile, kyrtosis sees to it that events at the end of the tail—where under the "normal" hypothesis they should have a minimal probability of appearance—repeat themselves with great frequency. Such a distribution, which does not follow the bell shape, may be leptokyrtotic or platokyrtotic. (Kyrtosis is the fourth momentum of a distribution.)

In conclusion, the opportunity for *creative thinking*[9] comes from the fact the area under the curve is never stable, and outliers tend to multiply. To make matters more complex, these may consist of a larger and larger number of daily occurrences and they must be evidently taken into account. Uncertainty is the most frequent reason behind the kyrtosis of the distribution of events. The commonly used standard deviation around the mean is not capable of describing uncertainty leads to asymmetries and fractals.

8.5 Test of Hypothesis

The *test of hypothesis* (see also Chap. 10) is integral part of the science of *statistical inference*. Like any other test, it bases itself on estimations in order to reach conclusions. Say, as an example, that we evaluate two samples about whether or

[9] Not to be confused with creative accounting, which is a scam.

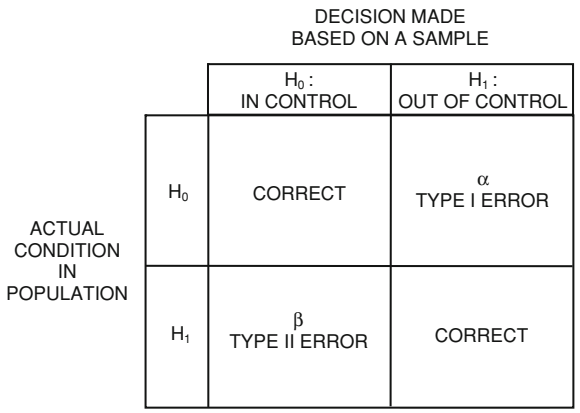

Fig. 8.5 The test of hypothesis: H_0 vs. H_1

not they have come from the same population. Sample A has a mean $\bar{x}_A$; and sample B, $\bar{x}_B$. The population's mean is equal to μ which is a parameter:

- The null hypothesis, H_0, states that there is no difference.
- The alternative hypothesis, H_1, states that there is a difference.

Under either hypothesis, the mean $\bar{x}$ is a statistic computed from observations. A sample mean far removed from μ (the population mean) will rarely occur if the null hypothesis is true. By contrast, a bull hypothesis will be rejected if a value $\bar{x}$ occurred which would be expected very rarely if the null hypothesis was valid. Just how rare this "rarely" is, can be computed through the level of significance (see also Chap. 10 on operating characteristics curves).

The pattern followed in the test of hypothesis has been explained in Fig. 2.2. As a reminder it is again shown in Fig. 8.5. The four quarter-spaces map the alternative possibilities of a correct decision and of an error. The test of hypothesis involves six steps:

1. Define the population and the sample
2. State the level of acceptability
3. Perform "n" random trials
4. Classify each trial as success of failure
5. Count the number of failures
6. Reach a decision: accept or reject the null hypothesis using statistical inference

The level of significance (or confidence) is denoted by α which in an operating characteristics curve is known as producer's risk (Type I error). Typically, we choose α the 0.01, 0.05 or 0.1 levels of confidence, corresponding, respectively, to 99, 95, or 90% of all measurements under the normal distribution curve. This choice should be made before we start testing.

In addition to the possible Type I error of rejecting the null hypothesis when it is true, there is also the possibility of accepting the null hypothesis when it is false. This is represented by β and it is known as Type II error or consumer's risk. (It should not be confused with β, the volatility metric.) α and β can be used to:

- Indicate the type of error involved in a test of hypothesis and
- Tell about the chance of making the one, the other, or both types of error.

However, while for a given number of observations n (sample size) we can choose the level of α, the β is statistically determined. For a fixed n a decrease in α will increase β; for a given α a bigger sample will decrease β. On this simple notion rests the whole theory of OC curves. Basically, operating characteristics curves constitute a statistical *power analysis* very helpful in establishing the probability that a statistical test will correctly reject a false null hypothesis. The integration of statistical power analysis into an experimental design (Chap. 11) is a relatively simple process, involving three basic components:

1. The existence of a phenomenon and its effects.
   By "effects" is meant the degree to which a given phenomenon is present in the population. *If* everything else is constant, *then* the larger the effect's size the greater the probability it will be detected and the null hypothesis will be rejected.
2. The chosen level of significance α.
   As briefly mentioned in the preceding paragraphs, this is the risk of committing a Type I error, also known as producer's risk, and gives the probability of incorrectly rejecting the null hypothesis (H_0) when performing a test of significance.
3. The sample size n in absolute number.
   The larger is the sample size in absolute number, the steeper the OC curve, hence the smaller the error, the greater the accuracy, and the higher the power of the test. The absolute number of n is more important than its size relative to the population's size N.

These three factors and the power level of a statistical test correlate. The sample size is an important feature of every statistical study (Chap. 9). If n is small, then the test has inadequate power. The degrees of freedom and the power level of the test are directly dependent upon the sample size. As n increases, the probability of error decreases.

The notion outlined in the preceding paragraphs are applicable with all observations, whether embedded in time series or coming from experimental data. One of the problems with observational information is that of scanty knowledge about the properties of residual variation, which is only a minor issue when dealing with experimental data. This difference is due to the fact that methods for hypothesis testing:

- Have been designed in line with the tradition of self-contained experiments and
- The statistical decision mechanism is based solely on data under investigation.

As far as testing tools are concerned, tests for frequencies and for expected values were already in use in the nineteenth century. The twentieth century brought great rigor through the introduction of tests of significance (Chap. 10) as well as by demanding exactness in the treatment of tests of hypothesis. The assessment of levels of significance and of power functions, requires detailed specification of the stochastic assumptions made by the analyst.

Among the testing tools of a hypothesis, the Student's *t*-test is popular. Its use assumes that the distribution of rule returns is normal and the events are independent. The t-statistic follows an asymptotic normal distribution and can be used either as a one-tail test of hypothesis, or a two-tail test; for instance of zero defects against non-zero defects.

The application of the *t*-test to problems involving the significance of difference between the means of two independent samples, is not the only way of validation. We are in many cases interested in testing the hypothesis that several independent samples have been drawn at random from a common population. The appropriate method is known as *analysis of variance*.

Fisher is responsible for the development of the analysis of variance as a power statistical tool. He was an English statistician with training in biological and agricultural research, who dealt extensively with intimate relationships between factors by means of experimental design. Fisher described the analysis of variance not as a mathematical theorem but as a convenient method of arranging the arithmetic on:

- How to find the greatest common measure and
- How to bring to attention a pattern of a mass of statistical data, so that the logical content in the whole is really appreciated.

Apart from aiding the process of logical thinking, the results of analysis of variance help in validating and reducing to a common form all the tests of significance we may want to apply. Nearly always, Dr. Fisher had once suggested, we can, if we choose, put our data in other forms and other languages. We can also structure them into patterns of which the bell-shaped curve is just one example.

Sampling, patterning, the test of hypothesis, as well as other techniques of mathematical statistics have helped in verifying data, evaluating assumptions, and detecting errors impossible to see through conventional means. But besides the tools of statistics we need knowledge artifacts capable of discovering anomalies by steadily collecting, screening, analyzing, and reporting information available from many sources.

A different way of making this statement is that statistical power analysis can be served through expert systems which help in increasing the probability that an effect is found in the experimental study. A higher power level means that the test we are performing has an improved probability of producing a statistically significant result.

References

1. Bernoulli J (1713) Ars Conjectandi
2. Chorafas DN (1994) Chaos theory in the financial markets. Probus, Chicago
3. Pasteur M (Late 19th Century. Exact date unknown) Histoire d'un Savant par un Ignorant. Hetzel et Cie J Paris
4. Chorafas DN (2007) Stress testing for risk control under Basel II. Elsevier, Oxford and Boston
5. Chorafas DN (2003) Stress testing: risk management strategies for extreme events. Euromoney, London
6. *The Economist*, 1 Jan 2011

Chapter 9
Sampling Methods

9.1 Sampling Defined

The researcher, designer, or quality control engineer must learn not only the mechanics of mathematical statistics, but also the concepts that lie behind them. One of the most fundamental is the method of sampling and plotting the results of sample tests. Too often sampling is not properly planned and this weights heavily on the results of tests as well as on findings.

Sampling is often left to a subordinate who draws samples that are neither random nor representative of the desired population, in total disrespect of the fact that sampling is one of the pillars of statistics and great attention should be exercised in its design (Sect. 9.2). The best way to start is by appreciating that no matter what our profession is, we constantly use samples to make inferences about the population from which they have been derived. This is a process involving a trilogy of human knowledge:

$$\text{Sample} \rightarrow \text{StatisticalInference} \rightarrow \text{Population}.$$

- A *population*, or universe, consists of all events, measurements, or other issues of interest in our present work.
- A *sample* is a subgroup of this population drawn under proper rules for specific reasons of study and research.

Closely associated to the notion of a population is its consistence. Ideally, it consists of a homogeneous mass of measurements, events, manufactured items or other assets whose nature, behavior, or some other variable is of interest to us. To be of value in a scientific study,

- Both the population and its sample have to be clearly defined, and
- The sampling method which we use must be carefully chosen.

There are many reasons why we use samples. One of them is that the population may be so scattered that it is impossible to reach and test all of it. Also, the tests we do may be

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering, 165
DOI: 10.1007/978-1-4471-2966-0_9, © Springer-Verlag London 2013

Table 9.1 A bird's eye view of symbols for parameters and statistics

	Population	Sample
Mean	μ	$\bar{x}$
Standard deviation	σ	s

$\bar{x}$, s are statistics; μ, σ are parameters

destructive; or studying a representative sample of it may be much more effective; or it may be difficult to comprehend the significance of large quantities of ungrouped data.

For any one of these reasons, it is desirable to find a few numbers of mathematical expressions which describe the relevant properties of the entire field of data. These numbers are called the *parameters* of the population: We have spoken of μ for population's mean, and σ the population's standard deviation—which we study through their proxies derived from the sample, respectively $\bar{x}$ and s.

$\bar{x}$ and s are the *statistics* of the *sample*. A relatively small number of individual items can be selected, for instance at random, from the population (see Sect. 9.2 on sampling procedures). By analyzing sample data, we derive statistics which describe the properties of the sample. Inferences concerning the population parameters can then be drawn from these sample statistics. Table 9.1 gives a bird's-eye view of parameters and statistics. In this book we use the symbols μ and σ for mean and standard deviation respectively.

Sampling implies procedural rules based on probability theory which, however, are not always observed. Quite often, people make serious mistakes by sampling data that are not independent or by using small samples. Both errors lead to biased samples and false conclusions.

The population from which a sample is drawn, often called the *universe*, can be finite or it can be infinite. In practical applications, the way to bet is that the population will be finite but large, a reason why we may treat it as though it were infinite. By contrast, the samples we draw will be finite.

The size of a population is usually represented by the letter N. This is the number of events, measurements, individuals, or some other factor under study in the universe. The accuracy of a sample statistic always depends upon the size of the sample that has been determined. The size of a sample is usually represented by n.

A complete set of observations upon which a study, analysis, or experiment is based is called a *sample of n*, where n refers to the number of observations. *If* a group of observations involves different sets, *then* the sum of all sets of n observations in the group will be equal to $\sum n_i$, where i stands for each sample.

- The measurements, events, individuals, and so on in the population will form a distribution whose mean is μ and the variance is σ^2.
- The measurements or other factors in the sample will also form a distribution which will have a mean denoted by $\bar{x}$ and a variance s^2.

The sample's $\bar{x}$ and s^2 which we actually measure are expected to give us information about μ and σ^2, whose values are usually not known. Notice however that $\bar{x}$ and s^2 tend to be different from sample to sample. By contrast, for a particular population μ and σ^2 are constant.

The sample's statistics $\overline{x}$ and s^2 have a distribution, respectively known as sampling distribution of the mean and sampling distribution of the variance. By examining the sampling distribution of the mean we can tell how frequently the $\overline{x}$ of samples will fall in the interval we wish to fall. This is very important in statistical quality control (SQC) by variables, as we will see in Chap. 13. In a nutshell the quality control procedure which is typically followed consists of:

- Selecting a sample of the units of a manufactured product from the population.
- Observing and recording a quantitative test value of the appropriate statistic(s), and
- On the basis of the computed value of the statistic, and of the underlying quality rule, accept or reject the statement that the population possesses acceptable quality.

The reader should appreciate that the use of statistical methods including those employed for quality control, does not eliminate chance variation, but avails means for watching over it. By the use of statistical inference, scientific analysis replaces guesswork. However, chance variation is still present, and therefore a sound statistical plan should account for it.

For example, in a manufacturing process, SQC procedures should consider both the producer and the consumer viewpoint. The producer aims for quantity and hopes for the benefits of quality. He demands protection against the rejection of good product. The consumer aims for quality and hopes for the benefits of quantity. He demands protection against the acceptance of poor product. Thus:

- The perspectives of the producer and the consumer are not diametrically opposed.
- What happens is that they do demand protection against different undesirable events.

In conclusion, the complete set of observations upon which a statistical analysis is based is a *sample of n*, where n refers to the number of observations, measurements, events or other factors. The sample of observations is usually assumed to be representative of a much larger number of possible events or measurements based on observations or experimental conditions. This larger group of potential observations is the *population*.

Measurements both of the population and of the sample are distributed in some way from a minimum value to a maximum value. We refer to the plotting of this distributed data as a *distribution* which has a central tendency and a variance. To characterize this distribution we need to measure its expected value and dispersion which is done through $\overline{x}$ and s for the sample; and μ and σ for the population.

9.2 Principles Underpinning Sampling Plans

"Sampling" as a process is probably as old as mankind, but the concept of using inference based on a sampling plan to eliminate rule of thumb, bias, or some obviously unacceptable assumptions, is relatively new. The goal is an objective

estimate in contrast to the biased and subjective approaches that are used so often. Given that it is not feasible, or it is not economically acceptable, to test a whole population we work on a sample.

Say that the officer in charge of procurement has to decide whether a population of springs is acceptable or non-acceptable. One way to screen the lot would be that of testing every single spring in it. *If* the testing was destructive, *then* after it was complete there would not remain any useful springs. If the testing was not destructive, the cost of this 100% quality control procedure might have been the prohibitive factor. So we may be better off by:

- Selecting a sample,
- Examining carefully every unit included in this sample, and
- Basing our decision on the outcome of this examination.

Statistical theory has established criteria for sound sampling procedures. One of them is that the properties of the sample should correspond as closely as possible to those of the population. If the sampling is repeated a number of times, the mean of the samples means should approximate nearer and nearer the population's expected value. In the general case, there are two methods of drawing a sample:

- Random sampling and
- Representative sampling.

The assumption with random sampling is that all possible choices are equally probable. A table of random numbers helps in the selection of items from the population. The keywords in random sampling are "without bias", which conditions the sampling plan. *If* some objects in the population are more likely to be chosen than others, *then* the sample is said to be *biased*. Typically, subjective methods of selection from a population lead to biased samples.

There exist different plans representative of sampling. Random is often the best option. Another more interesting one involves stratification of the population to provide more focused information pertinent to each stratum. For stratification purposes the population is subdivided into several parts, or strata, and the number of observations in the sample is apportioned among these strata.

Stratified sampling can also be the result of a *proportional testing strategy* employing a partition and *allocation scheme* in which test cases fall into subdomains. Some studies have demonstrated that, under certain conditions, stratification is an option at least as effective as random sampling and testing.

The problem with stratification is that the steps leading to it may include bias. In addition, stratification must have a valid reason. It would be wrong if it is dictated by past laboratory practises that have outlived their life cycle. Another shortcoming of stratification is that while the data represent the desired stratum, but are not representative of some critical factor(s) which are not evenly distributed among the strata.

The different sampling plans we will study in this and the following section are only to a small degree alternatives. Each case has a sampling plan which fits it best, and misleading results can occur if samples are not taken correctly.

An approach leading toward a better sampling procedure starts with the realization that "sampling", essentially means *data sampling*, and these data must have a reason for being.

Consider as an example a case in the processing industry. The chief engineer makes the prediction of uniformity regarding certain characteristics of the firm's main chemical product. The key variables described in the specifications are constant mean and constant spread.

The prediction is plotted as a function of time. If the production engineer uses $\bar{x}$ and R SQC charts, where $\bar{x}$ is the mean and R stands for the range, he can readily see if and when his process gets out of control. Say that in the beginning the process is in control, subsequently however:

- An $\bar{x}$ point above the upper control limit (UCL) indicates a shift in the distribution,
- An R point above the UCL indicates a larger spread, which is also an out-of-control situation.

As this brief example documents, SQC charts are interactive tools, easy to establish and their pattern is well understood (see also Chaps. 13 and 14). The challenge is that of choosing and implementing the proper sampling method, which is also a vital part pf the design of:

- Experiments and
- Control conditions.

Take as an example of control conditions that of a factory installing a plan for statistical inspection of manufactured items. Certain specifications for the factory's product are stipulated in the client's contract which also states that prior to delivery the produced goods shall be inspected to assure that only a small portion of the items could fail to meet specifications.

The quality control engineer can use this "small portion" as consumer's risk β and produce an operating characteristics curve which defines sample size after α has been chosen, usually by senior management. He could also use some practical findings; mathematically sound data samples permit us to do a much better job than is possible with 100% inspection.

We can make a smaller number of measurements and estimate the true value from samples by applying *the principle of persistence* of small numbers: *If*, in a group of quantitative phenomena selected without bias, a small proportion of the group deviates sharply from the characteristics of the remainder of the group, *then* this tendency will persist:

- No matter how large the group may be made, and
- Irrespective of the number of samples selected.

It is often said that it is better to take a large number of small samples. This is true in two cases: sequential sampling (Sect. 9.3) and SQC by variables (Chap. 13). Even then there is a limit on how small a sample could be, and there is as well a counterargument. Small samples can have severe aftermath on the accuracy of

statistical inference. Estimated approaches based on statistical tools at or near their limits involve:

- A high variance, and
- Significant margins for error.

Examples are statistical estimates with a sample of four or five, and projections simply based on averages without accounting for differences among the members of a population—whether people, manufacturing applications, events, or other observations. A sample size of five is acceptable in SQC by variables in manufacturing because of the orderly way in which successive samples are taken. On the contrary, in experimental studies small samples can lead to unreliable results.

Statistically insignificant samples may be one of the reasons why in social science studies—as contrasted to manufacturing—estimating approaches tend to involve high variance and high margins of error. I recently heard an argument that social scientists in a given study considered a sample of one as being statistically valid, and based on it their conclusions.

Another principle which impacts on sample size, is that of *decreasing variation*. As a large and larger proportion of a group of observations, measurements, events or phenomena is selected from the population by means of successive unbiased samples, the characteristics of each enlarged sample—such as the central value and variance—will differ less and less from the characteristics of the population. This makes it possible to determine the size of the sample in proportion to the whole of data.

A third principle underwriting sampling procedures is *statistical regularity*. If a reasonably large sample is selected without bias from a population, the characteristics of this sample will differ only a little from those of the universe. Some statisticians consider the *principle of large numbers* as the alter ego of that of statistical regularity. *If* an event may happen in only one of two ways and is observed to happen under the same essential conditions a large number of times, *then* the ratio of the number of times that it happens in one way to the total number of trials appears to approach a definite limit.

What is known as a *stability test* of sampling may be used as a rough check upon the adequacy of a sample. This test consists of division of the original sample into two equal samples by means of random sampling; or selection from the original data of a new unbiased sample equal in size to the sample already taken. The next step is noting whether or not the characteristics of the newly selected sample, differ materially from the characteristics of the original sample.

Another principle underpinning sampling plans concerns the fact that the degree of variation permissible between the sample and the whole cannot change arbitrarily since this depends upon the use that is to be made of the sample. The necessary sample size, i.e., the proportion represented by the sample, varies with different problems. The required number in a sample increases as:

- The variation in the individual items increases and
- There is a need for greater accuracy of the results.

Notice, however, that the accuracy of a sample does not increase directly in proportion to the size; it increases with the square root of the size of the sample increases. Hence, to double the accuracy, we must quadruple the size of the sample, and to treble the accuracy we should increase the number of items to nine times the former number.

9.3 Practical Examples with Sampling Plans

Theoretically, in the case of a production process every individual item chosen should be measured and returned to the population before another selection is made. This however is not practical at the production floor, apart from the fact that some samples undergo destructive testing. If the population is large compared with the sample size, very little error will result from the procedure of not returning each individual to the population.

In SQC and other implementation domains, plans permitting more than one number of samples are described as *multiple*. A double sampling plan can be considered as a special case of a multiple sampling plan. A multiple plan involves a finite number of samples, and this contrasts to what is known as a *sequential* sampling plan, which may permit a virtually unlimited number of samples until a quality decision is reached.

Sequential sampling is the rule in SQC because we deal with plural samples in the population domain. Its objective is to reduce the overall number of required observations by making subgroup observations in sequence to each other. However, the best quality control practice is to carefully evaluate the plan to be chosen, otherwise quality control objectives may not be attained.

Usually, sequential sampling plans can be designated having operating characteristics curves (Chap. 10) closely similar to the OC curves of a single sampling plan. The calculation of the operating characteristics curves for sequential sampling plans usually follows the pattern, which is discussed later in this chapter in the context of multiple sampling.

A test known as the *sequential ratio test* is designed to distinguish between alternative hypotheses, based on the likelihood ratio. The latter consists of the independent sample values being measured assuming that the hypothesis of no difference, H_0, is true.

Multiple sampling plans can be double, triple, etc., and they are adopted for two reasons: because they can be more flexible and because (as stated) in the long run they require less inspection. To properly design the multiple sampling plan one should carefully project the series of samples sizes, with associated acceptance and rejection numbers to be used in determining the acceptability of the lot. Tables, for instance, the double sampling plans are usually based on the Dodge–Roming tables, providing information for:

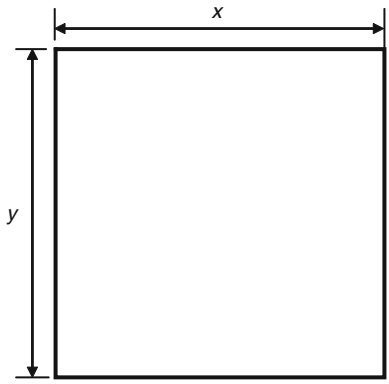

Fig. 9.1 Use of x, y coordinates of a box for sample section

- Sample sizes and
- Accept–reject levels.

For a numerical example on alternative sampling plans, say that the Eastern Electronics Laboratories has been requested to develop alternative SQC plans for inspection: by single sampling, by double sampling, and by triple sampling. According to MIL-STD-105A, developed by the US Military for procurement during WWII, for lot size $N = 300$ the following sample sizes are required by inspection level:

Severity of inspection level	I	II	III
Sample size	15	35	75

Let us suppose that it is decided to tail inspection level II, which requires a sample size $n = 35$ for single sampling procedure. The items in the lots are being delivered in a box and they are arranged so that the inspector can select any unit. Since the inspection plan requires random sampling, a table of random numbers has been used, with two columns of two digits each for each dimension. Hence, selection was made on the basis of x, y coordinates as in Fig. 9.1.

Say that finally the *acceptable quality level* (AQL) equal to 4 has been selected (see Table 9.2). *If* a single sampling plan is used, *then* with inspection severity II the inspection department should take a sample of 35 devices and accept the lot if three or less items are defect, or reject the lot if four or more are defects (Table 9.3). With a double sampling plan the first sample would be equal to 25. For two or less defects the lot would be accepted; for five or more it would be rejected. For three or four defects the inspector proceeds with a second sample of 50. In different terms, if the sample has three or four defects the inspector cannot reach an immediate decision; he should take and test another sample before accepting or rejecting the lot.

The same procedure is used with the multiple sampling plan in Fig. 9.2. This example comes from the financial industry; a loan's application known as

Table 9.2 Outgoing quality level single sampling

	Tightened inspection	Normal inspection	Reduced inspection
Outgoing quality level (AQL)	0	4	6
Accept	0	3	5
Reject	1	4	6

Table 9.3 Outgoing quality level double sampling

	Sample size:	A[a] 4	R[b]	A 6.5	R	A 10	R
First	25	≤2	5	≤3	7	5	11
Second	50	≤4	5	≤6	7	10	11

[a] Accept
[b] Reject

Risk-Adjusted Return on Capital (RAROC) which uses operating characteristics curves for judging creditworthiness. RAROC was developed in the late 1980s by Dr. Carmine Vona who was a nuclear engineer by training, and the boss of information technology at Bankers Trust of New York.

Usually, loans are given on an accept–reject basis, at one shot reflecting the client's creditworthiness. Vona thought that this is the wrong approach because it is possible to take an insurance on the risk assumed by the client. That is exactly what RAROC does with multiple sampling. Every time the test moves to the right the interest rate offered to the client is increased by the corresponding amount of insurance for the greater counterparty risk being assumed by the bank.

Vona's concept follows the principle of classification of defects in sampling inspection in a way emulating the testing procedure by the Eastern Electronics Laboratories. The sophistication of the method can be increased if the more critical defects are considered separately from, and are usually given lower AQL values than, less critical defects. Quite often in the industry the quality control department proposes a double classification system which weights the criticalness of defects.

As an example, say that the SQC-classification system in question defines the criticality of defects in a manner distinguishing "major", "midway", and "minor" criteria—hence three thresholds. A code letter is assigned to each inspection point by engineering and cannot be changed by the inspector. The classes are defined in terms of assemblies whose failure would result in:

- *Major*. Serious injury to personnel or loss of the manufactured unit—for instance, in an aircraft the weakened root cord structure
- *Midway*. Failure of the final product to function as intended, but without the loss of personnel or of the unit, as for example defective hydraulic line connection
- *Minor*. Interference with subsequent assembly or repair operations, or reduced quality of the end product, as for instance massive rivets

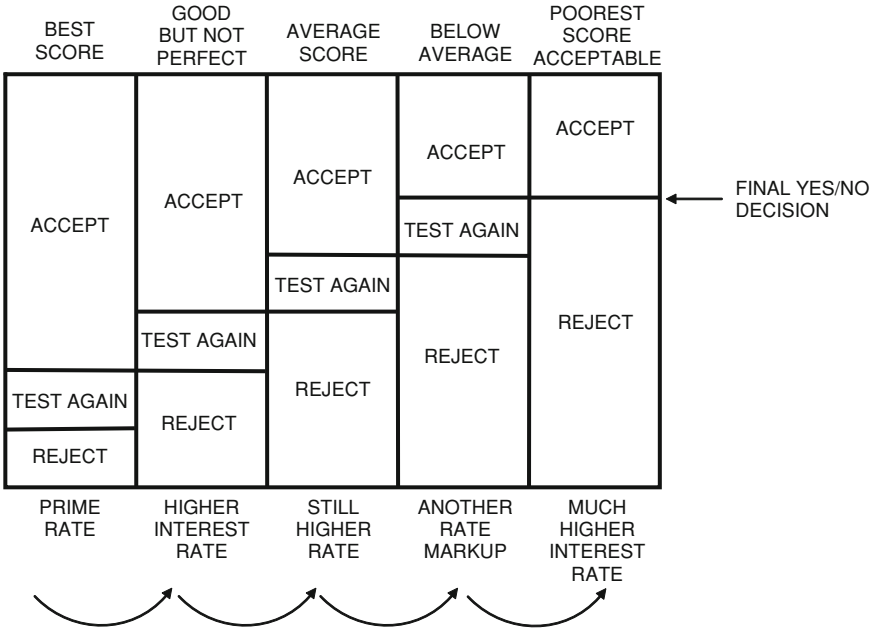

Fig. 9.2 A sequential sampling plan permits to avoid inflexible yes/no decisions on loans, by taking a reinsurance for higher risk

When a defect is found in any class, it is assigned a code number indicating its seriousness; i.e., the probability that this defect will eventually cause failure of the part. This number is usually determined by the inspector by means of criteria such as:

- A discrepancy which will prevent an assembly or installation from operating properly or will cause intermediate failure.
- A discrepancy that may not cause a malfunction or possible failure, but is a deviation from specifications and acceptable workmanship standards.

A sound system to use in connection with quality control is *chargebacks*, also known as *demerits*. In my practice, I have found it top class in keeping inspection personnel on its toes because laxity is penalized postmortem. Here is, in a few words, how it works.

In the course of production a defect may be found by device level, subassembly, system inspection, or post-inspection personnel. If a defect is found which should have been cleared previously by another department, it becomes a chargeback to that department—or, if a fine grid is used, to the person responsible for laxity.

If a production worker observes a defect prior to inspection, he points it out so that inspection sends it to be reworked, if possible, to specifications. If this is not possible, the defect is recorded in the production pickup section of the inspection book, with the classification code letter and number. If, however, a defective item

passes through to, say, the assembly operations and is subsequently flashed out, *then* there is a chargeback.

Inspection-found defects are coded and recorded on regular pickup sheets. Defects missed by inspection and detected later will be treated as demerits. Each SQC checkmark may contain several defects, such as missing rivets, oversize holes, and so on. The various classes of defects should be considered in separate categories and given different AQL values. Alternatively, rather than creating a large number of categories, these defects can be weighted according to code letter and number as demerits.

Quality control tables give the demerits assigned to each combination of code letter and AQL requires. *Chargebacks* are weighted *double* in consideration of the increased risk of not finding the defect prior to delivery, while those checks of defects made by the production line are given in the tabulated value. When all the above factors have been considered, a numerical total of demerits is available as a measure of quality of the lot (or of a major unit).

This reference to a quality control plan comes from a real-life application. Among the advantages it provided have been that conditions causing the process to go out of control were investigated sooner; adverse conditions affecting a small number of units were readily apparent; and differences between shifts showed up as out of control points in a quality control chart which tracked the sample's spread (*R* chart for range of measurements, Chap. 13).

Some inference rules were established to facilitate accept–reject decisions because of out of control points on SQC charts. Attention was paid as to whether these had to be attributed to a lower quality being produced, poorer inspection techniques, or an assignment of code numbers in borderline cases influenced by the inspector.

9.4 Discovery Sampling[1]: A Case Study

Discovery sampling is a step forward in acceptance sampling. Its value lies in the simplicity with which it yields a prescribed measurable result, by introducing the concept that products from a group of machines and workers have a process average that may be used as a measure of quality by taking into account three factors:

- Some lots do not actually contain defectives.
- Lots that do contain defectives are likely to contain a small percent of them.
- The average percent of defectives is defined at the level of the bank of products between production floors (intermediate stocks).

[1] Originally developed by Lockheed Aircraft, discovery sampling has been a little known but powerful SQC tool useful to all sorts of enterprises—both big and small.

In procedural terms discovery sampling starts with inspection of 10 items from a lot. The lot is accepted if no defectives are found. Above zero percent there is a graduation in decisions. In a practical implementation, for a given lot size the average percent defective in stock was less than 0.25%, and the average product quality was better than 99.75%. Given the AQL specifics of this application, the lot was accepted.

The notion underpinning this procedure is that a lot that is 100% conforming to specifications will be accepted in any case, while a lot that is totally unacceptable will be discovered if only one part is inspected. The entire sampling risk is confined to that category of lots in which both good and defective items occur; such a lot is defined as *partially defective*. The fraction of partially defective lots presented for acceptance is:

$$A = \frac{\text{Partially Defective Lots}}{\text{Good Lots} + \text{Partially Defective Lots}}$$

In the implementation in reference for individual departments "A" was found rarely to exceed 20%; it was less than 10% in most cases. The value of "A" was reasonably stable at less than a value prescribed by quality control. This indicated a constant cause system.

Because partially defective lots are at the core of discovery sampling, their fraction should be verified before installation of that method, as well as at intervals thereafter because quality wise the production pattern may change. A verification is essentially based on measurement of the process average and dispersion trends.

Attention should be paid to the distribution of partially defective lots within the group. A study of 20,000 lots made in different inspection areas by a major American manufacturing firm documented that they had the same or very similar distribution curve. The *average percent defective* in stock was found to be less than:

$$100 \cdot \frac{A}{4n + 2}$$

where

A = the fraction of partially defective lots presented for acceptance

n = sample size or the number of items inspected per lot.

Content with the results of discovery sampling, the management formalized its procedure through the organization into the following steps: Start with assigning intervals for p, the fraction defective. For instance:

0–5
6–10
11–15

The fraction defective p is computed for each partially defective lot by dividing the defectives found by the total lot quantity. When all of the partially defective lots are distributed within the aforementioned intervals, the quality control procedure computes the decimal fraction for the lots found within each interval. The probability of acceptance for a lot that contains defectives when the sample size is 10 and no defectives are allowed in the sample, is:

$$P_a = (1 - p)^n$$

where n is the sample size. Assigned intervals for p, quantity of lots found within the internal and computed statistics should be entered in a 6-column table as follows:

1 Assigned intervals for p	2 Quantity of lots found within the interval	3 Fraction of lots occurring within the interval	4 Probability of acceptance with a sample of 10, no defectives allowed	5 Mid-point of the interval	6 Fraction of defectives in stock for partially defective lots

The fraction of defectives in stock for partially defective lots for each interval is the product of Columns 3–5. This product is entered in Column 6. The meaning of fraction defectives in stock partially for defective lots is that "so many lots" (typically a small to very small fraction) occurred for each interval, and accepted by the plan.

The sum of Column 6 from assigned interval for p from interval 0–5 to interval 96–100, is the total contribution for defectives from all of the lots in all of the intervals of the partially defective group. The product of the sum of Column 6 and the percentage of partially defective lots $100A$, gives the average percent defective (APD) in stock.

If a frequency density function or operating characteristics curve are desired, each value in Column 3 must be multiplied by the number of intervals—in the present example 20—and plotted over the range for each interval. This gives a frequency density histogram. A smooth curve representing the histogram is a frequency density function for material presented for acceptance.

The example which we have seen has been discovery sampling by attributes; an inspection part is either good or bad with no measure of the degree to which this is true (Chap. 14). If this degree can be measured, then it is preferable to use an SQC plan by variables (Chap. 13).

Say that five parts were selected at random from a lot are arranged in order of size. The dimension of the part is specified as 1 cm with tolerance of ± 0.010. These five parts are within tolerances characterized by the following distribution.

	Dimension
Upper limit of tolerance	1.010
	1.006
	1.002
Ordered sample	1.001
	0.998
	0.993
Lower limit of tolerance	0.990

If either the largest or smallest value exceeded the tolerance the lot should be held for disposition. Otherwise, discovery sampling discards the highest and lowest measurements—in this case 1.0006 and 0.993—and calculates the difference of the remaining extreme measurements.

	1.002
	−0.998
Difference	0.004

Then, it subtracts this difference from the smallest remaining measurement:

	0.998
	−0.004
Difference	0.994

If this number is greater than the lower tolerance, the lot is accepted as being above the minimum. The next step is to add the same difference to the largest remaining measurement.

	1.002
	+0.004
Difference	1.006

If this number is less than the upper tolerance, the lot is accepted as being below the upper tolerance. If either of the above numbers exceed the corresponding engineering tolerance, the lot is held for disposition.

In essence, discovery sampling by variables excludes the extremes of the sample and works with the extremes of the middle measurements (in this example three out of five). If the sample contained ten parts, then the upper and lower two are discarded—essentially the two largest and the two smallest measurements.

The rest of the discovery sampling procedure works in a similar manner to that of the example presented to the reader.

9.5 Sampling Errors

In 2000, as the dot.com bubble had burst, many Internet companies went against the wall taking along with them some of the established names in information technology and in business systems. One of them was Xerox. In September 2000, its cash on hand was a razor-thin $154 million for a total debt of $17 billion, including a $7 billion credit line which was projected to be exhausted by the end of that year.

A new president took over and one of the first things he did was order up a review of the economics of the existing Xerox product line. He was presented with charts showing that Xerox was "world class" in terms of manufacturing and development costs. But the company's profit and loss (P&L) statements told a different story and his response was: How do you know?

It turned out that Xerox staffers had relied on a sample of 1994 market data, so limited as to exclude most of Xerox's Japanese competitors. The new CEO ordered them back to the drawing board. Weeks later, he finally was presented with evidence that Xerox had failed to maintain its hard-won parity with the Japanese.

The proper sample of competitors and challenges they presented made it clear that Xerox was at a large and material cost disadvantage against the Japanese across the copier market. The staffers had been very imprudent by using a sample of obsolete statistics. Besides that, the sample of competitors they had taken was too narrow leaving out their most formidable challengers. The result was:

- Loss of market share and
- A critical condition in the company's finances.

At about the same time, a study by the Information Technology Association of America (ITAA) suggested that companies seeking to hire information specialists in the next few years will be faced with a severe shortage. ITAA put the deficit at 191,000 professionals. Such a large number of a deficit was derived from a study of responses from:

- 149 out of 1,000 IT companies in a sample, and
- 122 out of 1,000 in another sample of non-IT companies.

In both cases the 1,000 companies were the sample targets, but that number was never reached. Those responses that were received indicated that there was an average projected vacancy level of 33 positions per IT company employer and between four and five per non-IT company employer. The guestimate of 191,000 IT vacancies seems to have been made by multiplying these figures [1].

It does not require great ingenuity to appreciate that the samples were substandard. The survey data was inadequate and the sampling methodology itself was wanting. The flaw in the sampling methodology has been that such small number of responses make it quite likely that the survey had a significant bias.

- The sample was small and
- There has been a very low level of response.

Both factors led to statistical bias. In addition, there was a failure to statistically validate the estimates being made by a follow-up survey. The estimate of vacancies is a tricky subject. A specific problem is that data from employers is only a small part of the information necessary to determine supply–demand imbalances and therefore level of projected vacancies.

This is by no means an exceptional case. Many sampling inspections and surveys fail to account for the number of applications, as well as the number of individuals interviewed. Yet, both the number of offers made and of salary levels offered (accepted or rejected) are important data for an accurate estimate of vacancies. A good statistical supplement would have been:

- The number of hires,
- The number of reductions in workforce, and
- The people who leave voluntarily.

None of this information was provided in the survey and therefore in the report. For instance, vacancies can balloon if employers offer lower salaries than those prevailing in the industry. In this case they should not be taken as indicators of real demand for a given profession. The lesson to be learned from this and similar cases is that not only the size of samples but also their composition, and the questions being asked, are very important in a statistically valid analysis leading to an estimate or projection.

Another example of defective sampling methodology which came to my attention in the technical auditing of a financial institution was that of poorly documented credit ratings. Without a well-documented history of defaults, including many instances of successful and unsuccessful debt service, it is not possible to be sure that indicators typically used by credit officers will pick up future problems.

Similar cases exist in quality control. This is a problem connected to focusing on one's aims prior to using sampling approaches, and it is indeed inherent in the analysis of all types of phenomena for which only small samples exist. What I just stated about an erroneous credit sampling and screening methodology has been confirmed by the careful study of past instances of default. Defaults have engulfed entities with relatively:

- Low as well as high debts,
- Good as well as poor management,
- Long-established product lines as well as more recently launched innovative instruments.

A stratified sampling of low debt levels, poor management practises, and classical type loans will provide incomplete and most likely misleading evidence for management decision. A counterparty's credit behavior should not only be unearthed by focusing on high debts and poor governance but as well—and most importantly—by concentrating on permutations of risk factors.

Another example where sampling methods have been wanting is that of unreliable statistical samples in opinion polls. Take politics as a first example. In the US from 1992 to 2010 the number of presidential polls has more than quadrupled. Theoretically, that means a lot of more information available to the public. But the proliferation of rapid polls — known as down and dirty—makes much of that information unreliable.

According to expert pollsters, as a minimum opinion surveys should be taken over a period of at least 3 days and include at least 1,000 voters. By contrast, there has been plenty of single-day polls based on samples as small as 250 people. There are also one-night polls which are totally irresponsible and verging on product liability—as a veteran pollster has it.

In conclusion, like any other activity which is worth doing sampling has its rules. These rules are not always observed and what happens is a misuse of the term *sampling* which can end up in liability. The reason behind such false sampling plans is human error. Therefore, it is not enough to hear that "this result (or conclusion) has been based on sampling." It is also necessary to know how the sampling was done.

9.6 Product Innovation Requires More Sophisticated Sampling Plans

While the science of statistics moves relatively slowly, innovation in man-made products is characterized by rapid advances which in turn require more sophisticated sampling plans. Old concepts are still precious as evidence on the transition in the implementation of statistical methods and tools, but they are not a guide to new applications which require plenty of imagination as well as detail in the way a sampling plan works.

There are plenty of examples on product liability due to sampling errors, and not only in connection to political opinion polls. Detroit, for instance, has made major blunders in projecting the type of motor vehicle American and international clients will require. This wrong-way strategy has been based on the narrow perspective of opinions by clients biased toward SUVs.

One might have thought that Detroit's Big Three employ the most effective methods a technology of prognostication can provide. Practically, this is far from being the case as documented by the fact that the formerly mightiest vehicle manufacturer in the world, General Motors, brought itself to its knees and Chrysler (the smaller of the Big Three) went along.

While the insistence to continue designing and manufacturing big gas-consuming automobiles when the price per gallon hit $4, was a top management blunder, other failures were directly debited to the engineering of Detroit firms. An article in Automotive Design put it in these terms: "To Europeans, US domestic products were deemed to be relatively crudely engineered, of poor build quality, using low-grade materials and of questionable design—never mind indifferent dynamics and thirsty engines" [2].

It took the ordeal of going in-and-out of bankruptcy for GM and Chrysler to change their product management's perspective and their method for sampling potential customers to unearth needed changes in design in a way that engineering can be ahead of the curve. After GM's bankruptcy, its engineering division has been restructured to take advantage of regional centers of excellence. A new top management decided the company had to design:

- Small and mini cars in South Korea,
- Compact models in Europe,
- Global mid-size truck platforms in Brazil, and
- Mid-sized platforms and trucks mainly in North America.

Among themselves these centers aim to cater for all markets around the world, by providing a global engineering competence. So far so good, but for such laconic classification of regional competence, is it enough to define the mission given to design engineers? I doubt it, because it is too general and therefore ineffectual.

The big car–small car debate is more than 30 years old; it is not a newcomer in automotive competition. Packing the car with a great deal of *in-car* electronics[2] is a dozen years old design feature. Therefore, it is no more an indicator of future competitive advantages. The new generation of competitive advantages work under the impact of novel customer inputs motivated, among other reasons, by:

- Death statistics due to car accidents and
- How automotive manufacturers could contribute to the reduction in fatal accidents.

While the weakest link in car driving is the man behind the wheel, the equipment's features too play a role. This is an engineering challenge closely connected to *product assurance* (Chap. 1) and to *service assurance* (see Chap. 4). However, as we will see in the following paragraphs the widely held notion that the reasons behind car accidents have a universal bearing is wrong.

- Major countries tend to have an individual pattern of background reasons for auto accidents, and
- Samples taken indiscriminately from the global population will be biased, because the global population of drivers is by no means homogeneous.

[2] *In-car* electronics has more or less reached a saturation point, with new offers distracting rather than helping the driver.

Let us start with statistics. In mid-2011 it was estimated that worldwide in that year there were 5.5 million car crashes. Precisely because the global population of drivers is not homogeneous, the designer's challenge is that, more or less, every one of these care accidents will be different. "The only time you get two crashes the same is in a crash lab", wrote Thomas M. Kowalick in Automotive Engineering "and that's not real-world data. If you could take just one day's worth of crashes in America —which is 20,000 tow-aways—that would be more data than you collect in one year in the lab" [3].

An upcoming device in the automotive industry, and a "plus" in competition, is the *event data recorder* (EDR)[3]—a so-called 'black box'—which by all evidence will play an important role in establishing causes and culpability in crash events. The intent of such information is to help in revealing:

- Driver and component failures,
- Accident patterns, and
- Injury risks by type of accident.

It could also be instrumental in detecting vehicle defects, directly contributing to future vehicle development and design. In addition, because the EDR is a car component its information will help in making more accurate statistics regarding accidents with motor vehicles which vary widely from country-to-country and over time. When security in auto transport escapes the control of authorities and accidents mount, these statistics convey the message of a crisis.

The wider spread of EDR is a long delayed innovation in private car features and it comes on the heels of another innovation—that of automatic breaking by car electronics to avoid a crash. Another similar advance is that of an automatic pilot which can park the car in a narrow space at a signal from a smart phone. In both cases:

- A system of microprocessors activates the car's engine, gearbox, steering and brakes and
- Sensors alert the car's command about the risk of bumping into other cars or people.

At least theoretically, this increases the security of the car's passengers and of the surroundings.

Pilotless cars, such as the Volkswagen Sharan are still laboratory models. Many people at the mid-September 2011 Frankfurt Motor Show were asking not only how the cars of the future will be powered, but what kind of security control will feature and how they could ease the driver's job. That is an issue where behavioral patterns are important, and such patterns vary from one country to another (more on this later).

[3] Back in 1996, in the US, Ford and GM began installing the early predecessors of EDRs in their cars.

"Where does the car end and the phone begin?" asked Chris Anderson, the editor of *Wired* magazine, at a brainstorming session organized by Audi, the carmaker [4]. By all likelihood, a future car will be more like a computer on wheels, networked with the surrounding infrastructure and other vehicles. This looks like science fiction, but it is at least an innovation—which cannot be said of the 110-year-old electric car.

It is appropriate to recognize that like many other man-made devices—for instance airplanes—motor vehicles are both an opportunity and a risk. In Chinese, the first syllable of the word *weiji*, for crisis, is *wei* and means danger; by contrast the meaning of the second syllable: *ji* is opportunity. The challenge is that of designing a better mousetrap, which has always been a guiding light in business.

In the case of auto accidents the "better mousetrap" is electronic devices specifically developed to prewarn the driver of an impending accident, which is within the each of current technology while Sharan is still in the lab. Within this perspective the mission of the design engineer is particularly influenced by two factors: product assurance and functionality. An optional balance is doable provided that there is:

- A wealth of statistics *personalized* at the drivers level, and
- Sensors sensitive enough to differentiate one type of accident from another.

As already mentioned in preceding paragraphs accidents happen in a variety of ways, and they have no unique pattern around the globe. After classifying them into seven main categories (plus a minor class of "other kinds") Automotive Design has provided some most interesting auto accident statistics from America, Germany, and Japan [2].

- Collision with fixed object or vehicle leaving the carriageway, has a frequency of 46% in the US, 32% in Germany, and only 17% in Japan
- At 34% accident with a pedestrian is the highest in Japan, followed by Germany at just 14% and the US at 12%
- At a 21% likelihood, collision with a vehicle which turns into or crosses a road is also the highest in Japan, but the US at 18% and Germany at 15% are not far behind
- Germany leads in parallel vehicle accidents with 21% of all crashes; Japan and the US, respectively stand at 11 and 10%, have half the German score
- There is a nearly equal probability of accident with a vehicle moving ahead. It stands at 6% of all cases in Germany, 5% in the US, and 5% in Japan
- The likelihood of an accident with a vehicle moving laterally in the same direction is 3% in Germany, 2% in the US, and 2% in Japan
- Worst of all, the head-on collision with an oncoming vehicle is 2% in Japan, 1% in Germany, and 1% in the US—this still being a big percentage for this type of accident

Developed for all major markets of motor vehicles such statistics can be instrumental in promoting product assurance in tomorrow's competitive markets. They can be used to advantage from car design to the development of sophisticated

devices which respond to specific safety requirements country-by-country and eventually driver-by-driver. Since accident patterns are not universal a "good for everybody" design makes no sense.

The mission of the designer is that of analyzing accident patterns in various regions and in different countries, understanding and formalizing the foremost safety needs and incorporating service quality features to be activated/deactivated on driver's sign-on and sign-off. This has been done for nearly two decades with fuzzy chips which:

- Learn the car owner's driving profile and
- Optimize the use of fuel in accordance with this driving pattern.

In conclusion, significantly improving product assurance through higher security is a quantum leap in competition. Big achievements are made when engineering designers are given the goal to put muscle to well-defined breakthroughs, seeing them through in an evolutionary way from an ideal to a model and from there to a product appealing to the customer.

Such a goal, however, must be focused and the chosen course documented through studies based on correct sampling of the user population. Years ago at UCLA, my professor of quality control mentioned an old saying about errors with statistics. It went like this:

> There's a great text on errors in estimation.
> Once you trip on it, entails
> Twenty-nine distinct damnations.
> One sure, if another fails.

References

1. Electronic Design, 2 Sept 1997
2. www.automotivedesign.eu.com, Sept 2010
3. www.automotivedesign.eu.com, July/Aug 2011
4. *The Economist*, 17 Sept 2011

Chapter 10
Operating Characteristics Curves

10.1 Operating Characteristics and Power Curves

Typically, for every sampling plan there is an operating characteristics (OC) curve which shows how the plan will perform as lots of different quality levels are submitted to it. In the background of OC curves is statistical inference which helps in determining critical points corresponding to the risk, or risks, under study.

The seminal work in the sense of World War II which led to practical applications of OC curves has left a legacy of tables, the best among them being MIL-STD-105A [1]. Say, as an example that a sample of size n is taken and inspected. Depending on the value of the percent defective, p (Chap. 14), the lot is:

- Accepted if there are up to l defective items, and
- Rejected if there are more than l, which is the *acceptance number*.

An application with acceptance numbers has been brought to the reader's attention in Chap. 9. Evidently $l \leq n$. Let us choose for l the values of 0, 1. When the number in the lot is large compared with that in the sample, the probability of acceptance can be computed from the theoretic sampling distribution (more on this later).

The contribution of an OC curve is to indicate the likelihood of rejection of a lot with acceptable percent defective, for instance $l = 1$, while it should have been accepted. Also the likelihood of acceptance of a lot with, say, $l = 3$, while it should have been rejected. Because of the dynamics of sampling inspection:

- The lot under inspection might be rejected while overall it is of acceptable quality.

This is known as Type I error or producer's risk. It is shown in Fig. 10.1 as α. The statistic α is the measure of our confidence that something will happen; it is a threshold permitting the quantification of risk (see also Sect. 10.4).

- The lot under inspection might be accepted while overall its quality is not acceptable; hence, it should have been rejected.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering, 187
DOI: 10.1007/978-1-4471-2966-0_10, © Springer-Verlag London 2013

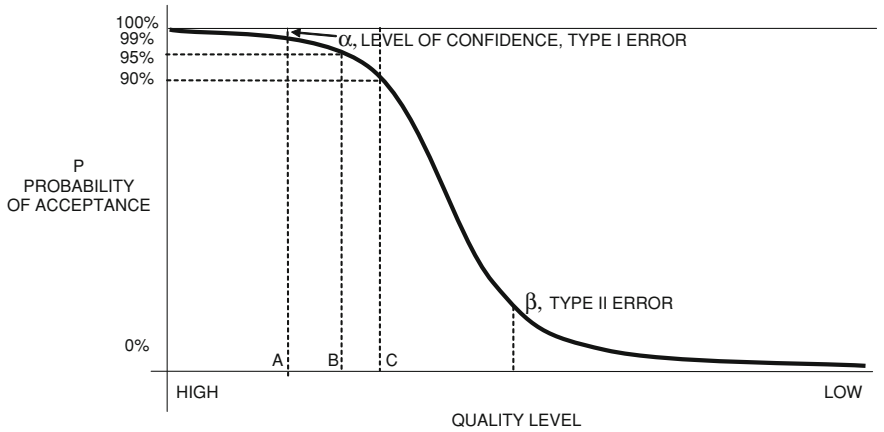

Fig. 10.1 The OC curve of a sampling plan's statistical distribution

This is known as Type II error or consumer's risk. It is shown in Fig. 10.1 as β.[2] Beta is another type of error committed when an existing effect remains undetected in spite of having defined the acceptance/rejection threshold. *Detection* is the keyword and the power of a statistical test is defined as the probability that it will correctly accept or reject the null hypothesis H_0 (see Chap. 8). The rejection of the null hypothesis is represented by $1-\beta$.

Type I error is embedded into the stochastic system. α is usually set at the level of 0.01, 0.05, or 0.10. With $\alpha = 0.01$ there is a 1 in 100 chance of incorrectly rejecting H_0. With $\alpha = 0.05$ this probability of rejection of H_0 increases to 5%. With $\alpha = 0.10$ the probability of rejecting of H_0, while it should have been accepted rises to 10%. In scientific research $\alpha = 0.10$ values border the ridiculous, still they are widely employed in economies and finance which talks volumes of the seriousness of studies providing a level of significance of only 90%.

In Fig. 10.1 the Type I error is shown at 99, 95 and 90% level with corresponding projections A, B, C on the abscissa which identifies the quality level. As a visual inspection can confirm the 90% level on the OC curve and its projection C on the abscissa is far out toward a lower quality level—which, as we will see later on, is the acceptable quality level (AQL) of a production process.

A weak point in power curve analysis is that β is often ignored by researchers. This is wrong because by doing so one disregards the important message conveyed by the OC curve. If β is properly considered, it can assure that a statistical test will have sufficient power to detect whether the phenomenon being examined is characterized by a large Type II error (most often because the sample size n is too small).

[2] Not to be confused with β which stands for volatility.

The reader should as well appreciate that α and β are not independent of one another. They are connected by the OC curve. With α set, the value of β will be constrained which affects the power of a test. A good criterion linking the Type I to Type II error is: β/α.

- If $\alpha = 0.05$ and $\beta = 0.30$, then $30/5 = 6$.
- Hence, the rejection of H_0 is six times more likely than erroneously accepting it.

In this example, the *power* of the test is $1-\beta = 1-0.30 = 0.70$. The power of the statistical test becomes particularly important when the null hypothesis H_0 is not rejected. In principle, the lower the power of the test the less likely H_0 is accepted correctly. The test results are ambiguous because the effect which is being examined has not been fully demonstrated.

If we test for the hypotheses of no difference between two populations with respectively mean parameters μ_1 and μ_2, but a common parameter σ for standard deviation, then the *sample size effect* can be computed as:

$$\gamma = \frac{\mu_2 - \mu_1}{\sigma}$$

where γ is an index.

Power analysis of experimental data permits an estimation of the effect of a size index, which can be used for calculating the power unit of the dependent variable by dividing it by the standard deviation of the measures in their respective populations. The null hypothesis H_0 assures $\mu_1 = \mu_2$ while the alternative hypothesis H_1 assures $\mu_2 > \mu_1$ for one-tailed distributions. Correspondingly for two-tailed distributions (nondirectional test) with two independent samples having the same standard deviation the algorithm is:

$$\gamma = \frac{|\mu_1 - \mu_2|}{\sigma}$$

where μ_1 and μ_2 are the means of the populations; and σ is the standard deviation of either population (assuming they are equal). Furthermore, the computation of a power level requires the values of α and of sample size n. When these values are available, the power can be easily calculated through the formula:

$$\delta = \gamma \,.\, \mathrm{f(n)}$$

δ combines the size effect and sample size into a single index that can be used with α to obtain the power level from statistical tables. The symbols γ and δ are used in this text in connection to statistical tests. They should not be confused with γ and δ respectively for second and first derivatives of underlying functions in the study of derivatives.

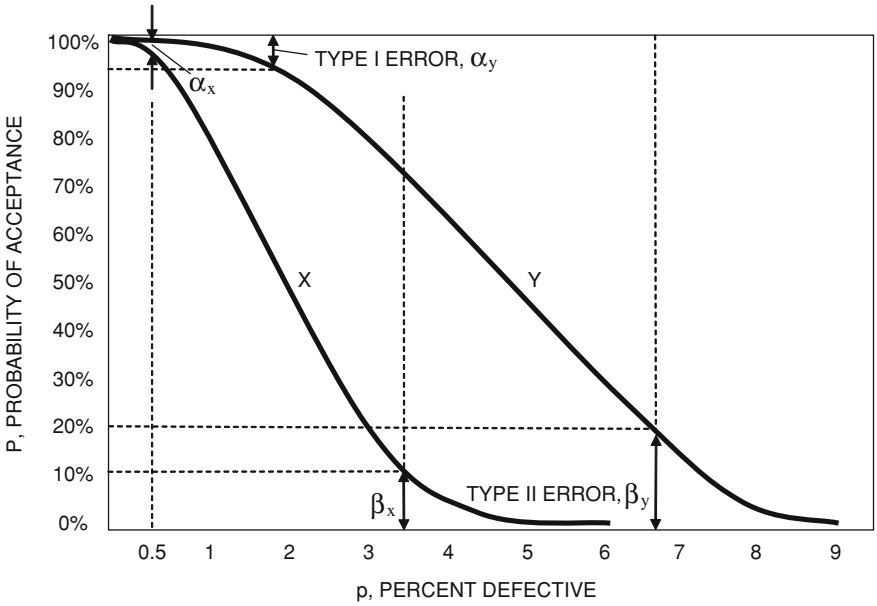

Fig. 10.2 OC curves for two different sampling plans X and Y

10.2 Improving the Shape of an OC Curve

A good way of improving the operating characteristic of the test is to decrease the standard deviation of the sample's statistics being tested. Different plans exist for this purpose; one of them is to take a lot size half as large as the original. Another, which ends up to the same result in terms of percentages is to double the size of a sample. Notice, however, that:

- The OC curve will steepen and β will shrink,
- But there will always be present α and β, albeit smaller ones.

Figure 10.2 compares the OC curves of two sampling plans for percent defective: P, X and Y. The OC curve of X is steeper. $\alpha_x < \alpha_y$ and $\beta_x < \beta_y$. In regard to both Type I and Type II error, sampling plan X is better than sampling plan Y in terms of producer's risk and consumer's risk.

Nevertheless, as Sect. 10.1 brought to the reader's attention, the Type I and Type II errors continue to exist because the percent defective in a sample may be more (or less) than the actual proportion of defective items in the lot. Given this variation, any lot-by-lot inspection plan based on sampling will include a certain amount of risk. What is important is that:

- With statistical quality control the errors are quantifiable and known.

- By contrast, with 100% the errors are present but unknown; therefore, it is like inspecting by the seat of the pants.

In the opinion of people who do not believe in the power of statistical tests, the existence of Type I and II errors in connection to sampling poses a question about the advantages of such a method as contrasted to a 100% inspection. Cost, fatigue of the inspectors (hence errors), and other reasons see to it that a 100% inspection would not provide 100% assurance.

Another significant advantage of OC curves is that they help in calibrating *sample size* in regard to *lot size*, by visualizing the effects of lot size and sample size on OC. People with experience in statistical quality control (SQC) know how to calibrate the sample n when they know (or decide) the size of the population N. The task is straightforward: it is possible to reduce the variance by:

- Decreasing N, while holding n constant
- Increasing n, while holding N constant
- Increasing n and N, while holding constant the ratio n/N.

In the general case, the effect of varying the sample size n is more important than the effect of varying N. In addition, the absolute size of the sample is more important than its size relative to the size of the population. These are two easy rules that should be always remembered.

Once the right SQC plan has been established, we are much better in control of outgoing quality. In addition, this plan becomes integral part of procedures put in place for quality assurance. The prerequisites to SQC are by no means complex, and the same is true of the aftermath. With sampling, the number of lots that would be accepted and the number that would be rejected would depend upon both:

- The nature of the inspection plan used, and
- The actual percentage of defective items in the submitted lots.

Therefore, in selecting a sample plan among alternatives it is good to have specific knowledge of how each of the available plans differentiates between good and bad lots (see also Chap. 9 on discovery sampling). Such information can best be presented as an operating characteristic curve where each point shows on the ordinate, the frequency of accepted lots giving the corresponding rating on the abscissa.

In the majority of cases, this study of alternatives should include the assumption that lots rejected by the sampling plan will be sorted out for control action. The fact of facilitating management control makes the SQC plan an important instrument in *quality analysis*, and it can provide a precious feedback to engineering design (Chap. 7). This feedback can be quantified by using two notions very important in quality control:

- AQL, (which the careful reader will recall from previous references), and
- Lot tolerance percent defective (LTPD, see also Sect. 10.3).

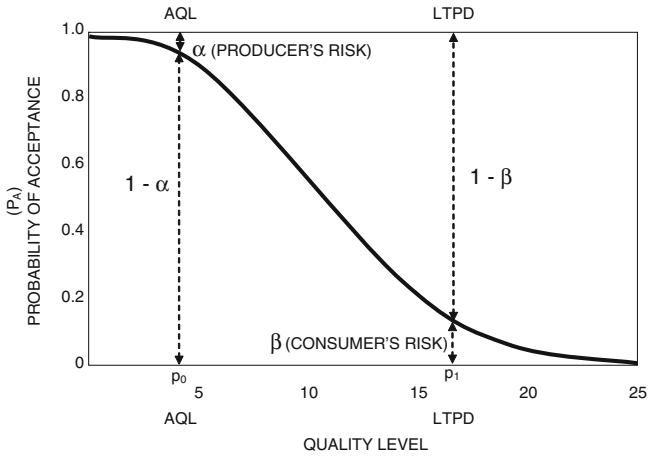

Fig. 10.3 Acceptable quality level and lot tolerance percent defective

As shown in Fig. 10.3, the AQL is equal to $1-\alpha$, and it identifies the percent of defective items in an inspection lot which are considered below the level of lot rejection. An $\alpha = 0.01$ at AQL gives a 1% chance of rejecting a submitted good lot containing a 0.5% defective. This corresponds to the 99% level of confidence.

- *If* the lot acceptance plan accomplishes its full purpose, *then* the process average will become at least as small as p_0. The probability of this happening is α.
- AQL contrasts to LTPD which is equal to $1-\beta$ and identifies a lot of sufficient bad quality that we do not wish to accept more often than a small portion over time.

The way to interpret the $1-\beta$ in Fig. 10.3 is that if lots of 0.5% defective are submitted to this sampling plan, the consumer has a 12% risk that bad lots will pass. This 12% level, which is admittedly unacceptable, can be improved by steepening the OC curve—which, as we have seen, can be done by increasing the sample for the same population. Two more terms need to be defined in regard to practical applications of OC curves.

- Average outgoing quality (AOQ), and
- Average outgoing quality limit (AOQL).

AOQ is the expected fraction defective after substituting good items for bad ones in rejected lots (or correcting the identified errors), and in samples taken from accepted lots. AOQL represents the value of AOQ for lots that result in the largest average outgoing quality. Or, the best average quality that can result over a period of time under the chosen sampling plan. AQL and AOQL correlate.

The flexibility in terms of control action afforded with sampling plans and OC curves is another reason why SQC presents a better protection; one which is

measurable and costs less than 100% inspection. The cases examined in this section help to confirm that the range between outgoing quality level and LTPD is an excellent quality assurance solution.

While these examples come from manufacturing where there exists today a very significant experience on the implementation of OC curves, the concepts underpinning them are just as applicable in finance. For instance in loans, like the example with RAROC in Chap. 9.

Whether in industrial applications or in banking, a hard-hitting successful program using OC curves must be carefully planned, simple, and clear-cut. It should as well be accompanied by the understanding and appreciation of opportunities and limits of statistical testing. Another "must" is to train in stochastic thinking the personnel and have its positive participation in the implementation of any statistical plan.

10.3 Using OC Curves: A Methodology

On many occasions, we are interested in comparing the performance of several acceptance sampling plans over a range of different (or likely) quality levels of submitted products. The use of OC curves is one of the best ways possible for attaining this goal. This opportunity comes from the fact that:

- OC curves serve in estimating the probability of accepting lots from a flow of products with fraction defective p, and
- For any given fraction defective p in a submitted lot, the OC curve shows the probability P_A that such a lot will be accepted by a given sampling plan.

The plan in Fig. 10.3 (Sect. 10.2) has been devised from statistics taken from a production line of a manufacturing firm. The quality of submitted inspection lots, p, in percent defective, is shown on the abscissa. The probability of accepting a lot if quality p is shown on the ordinate. This probability P_A presents the percent of lots accepted by the chosen sampling plan when many lots of quality p are submitted. The curve may be checked at a few points to determine the system's behavior.

A lot with no defectives should always be accepted. From the OC curve, we see that when $p = 0$, the probability of acceptance is equal to 1 ($P_A = 1$), indicating that all lots would be accepted. A lot "all defective" should never be accepted. Indeed, for $p = 1$, $P_A = 0$, the OC curve tells that none of the lots would be accepted. Between these two extreme points, there is a certain risk which should be taken into account.

The company whose statistics have been used in this case study did not employ SQC only for its own production lines. When giving a contract to a vendor, this contract specified in quantitative terms the quality level which the vendor must meet: For instance, all lots must be of quality p_0, or better. Lots of quality p_0, or

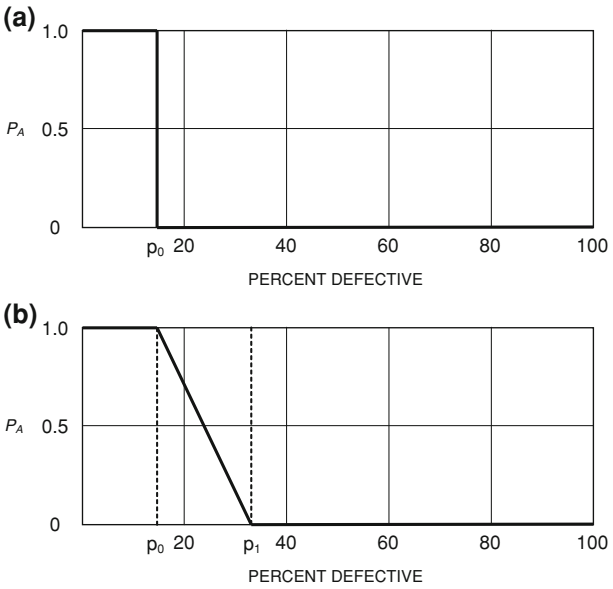

Fig. 10.4 Ideal OC Curves

better, were accepted by this plan all the time—a policy described by the OC curve in Fig. 10.4a.

A different firm, which also used sampling plans and OC curves for purchased goods, had the policy that it will:

- Buy all lots of quality p_0, or better;
- Buy none of the lots of quality p_1, or worse, and
- Accept a portion of the lots whose quality is between p_0 and p_1.

This policy is described by the sampling plan in Fig. 10.4b. Notice that the straight lines also represent an ideal condition not attainable whether with sampling or 100% inspection. One can easily observe that the OC curve in Fig. 10.3 is better than the "ideal" one of Fig. 10.4b. The power curve based on a SQC leads to the acceptance of:

- More lots when p is low, and
- Less lots as the percent defective increases.

By applying an OC curve, most of the lots with quality p_0, or better, are accepted. Most of the lots with quality p_1, or worse, are rejected. A portion of the lots with quality worse than p_0 but better than p_1 are rejected.

By specifying p_0 (AQL), p_1 (LTPD), α (alpha) and sample size which impacts on β (beta), the OC curve is specified. In choosing the sampling plan, the person in charge of quality control should remember that his critical decisions beyond p_0, p_1,

and α are: n, the sample size, and c (specified by statistical tables) the limit of defective items, c is the *acceptance number*.

A quality control plan should never be adopted prior to being tested for the behavior of its OC curve. This is necessary to assure that it has the wanted characteristics.

For any practical purpose, an SQC plan is a *quality assurance* plan based on statistical inference. In its simplest form, a sample of given size, for instance n, is taken and inspected. As we have seen in a previous reference, depending on the value of the percent defective, p, the lot is accepted or rejected, the probability of acceptance being P_A.

$$P_A = (1-p)^n, \quad \text{for} \quad c = 0,$$

$$P_A = (1-p)^n + np(1-p)^{n-1}, \quad \text{for} \quad c = 1,$$

$$P_A = (1-p)^n + np(1-p)^{n-1} + \frac{n(n-1)}{2}p^2(1-p)^{n-2}, \quad \text{for} \quad c = 2,$$

On the basis of n and c, can be calculated the OC curves. Some sample curves are presented in Fig. 10.5, for $\alpha = 0.05$. In Fig. 10.5a, the sample size is kept constant and c takes values c_1, c_2, c_3, where $c_1 > c_2 > c_3$. Conversely in Fig. 10.3b the acceptance number c is kept constant and the sample size takes the values n_1, n_2, n_3, where $n_1 > n_2 > c_3$.

One of the interesting possibilities provided by this methodology is that by means of statistical analysis based on test data, quality assurance information can move upwards the manufacturing hierarchy—and from there all the way to the design source (we have discussed this issue in Chaps. 6 and 7 in connection to reliability engineering). However, the reader should be aware of the fact that while statistical theory provides measures for errors which give guidance,

- It does not remove uncertainty.
- To the contrary, statistical inference is based on uncertainty and the reader should learn to live with it.

Take a simple situation as an example. A producer offers a lot which the consumer either accepts or rejects. This action is the result of inspection and it is often seen as being a simple choice between the alternatives of acceptance and rejection. This is, however, the wrong way of thinking.

As documented by the practical example in this section, the factors underpinning product assurance (Chap. 1) are much more complex than what is revealed by superficial approach to quality control in manufacturing. A dry number of rejected devices or lots provides no way for understanding whether the production process is *in control*.

Many advantages in quality assurance can be derived from the fact that the OC curve of a sampling plan offers a complete statistical description of the consequences of variation in outgoing quality.

The probability of accepting a lot of items can be read directly from the diagram we have seen in the preceding figures. If the population mean and sample

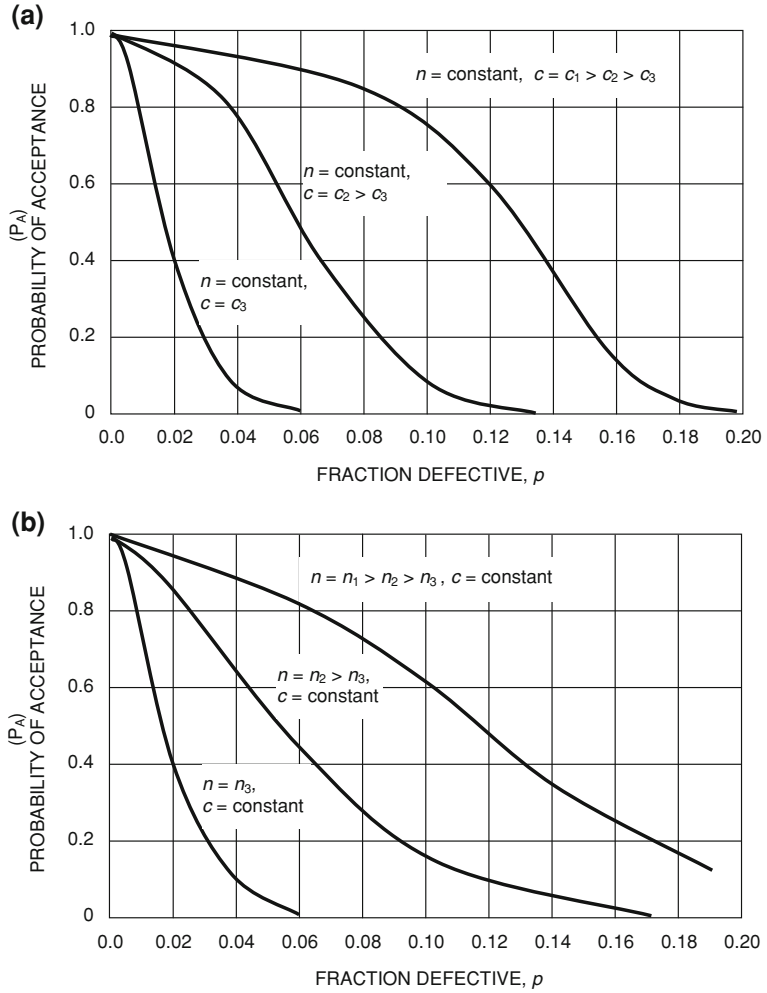

Fig. 10.5 **a** Alternative OC curves by keeping sample size constant, while varying the acceptance number. **b** Alternative OC curves by keeping the acceptance number constant and varying the sample size

size are known; the probability of rejection is one minus the probability of acceptance. Moreover, this complete statistical description can be invaluable in reliability engineering.[3] As it has been already discussed in Chaps. 6 and 7, reliability should not be confused with quality control.

[3] Nevertheless, in considering OC curves it would be of advantage to give an example from reliability engineering.

As with other quality control problems, in reliability engineering there exist two statistical risks. The first risk is that good equipment will be considered bad (producer's risk). The second risk is that bad equipment will be considered good (consumer's risk). Product acceptability as judged by a sampling plan is commonly established by statistically estimating the fraction of the total lot which is defective.

There exist as well limits of the accuracy of measurement often referred to as "confidence intervals" which is unfortunate since they may be confused with statistical confidence limits (Sect. 10.4). For instance, in response to the request by its customer on mean life of its equipment, a company stated that the mean time between failures (MTBF)[4] was estimated to be 600 h ±15% accuracy of measurement. Such a statistic was evidently unacceptable (see in Sect. 10.4 the discussion on accuracy).

A different way of looking at this issue is to consider consumer's risk β in terms of how long shall the test continue when the equipment MTBF is so inaccurate that the incorrect decision might be made from a short test (For instance, one made under stress conditions). Precisely for this reason, in reliability practice certain limitations have been established with the objective of optimizing test procedures.

As an example in one of the projects I participated, it was decided to require by contract that the minimum acceptable MTBF should be by 50% greater than the actually desired minimum. This value is associated with a level of confidence $\alpha = 0.05$ (Sect. 10.4) and β which led to a recommended sampling plan.

Going back to the fundamentals of inspection, the effect of errors in manufacturing and in acceptance of purchased material is the likelihood of a region of poor discrimination among the lots which should be accepted and those which should be rejected. The greater the errors in inspection, the poorer the discrimination would be. This is of particular importance in quality testing because the number of samples and time available for failure rate testing is usually severely limited.

10.4 Level of Confidence

Level of confidence is the degree of protection observed in statistical inference against movements in the underlying measurements or observations, in regard to characteristics of a population under study. To appreciate the fine print of this definition we should return to what was stated about the normal distributions as well as asymmetries in Chap. 8.

The development of an OC curve is based on the hypothesis of the normal distribution which, as the careful reader will remember, is an approximation of real life situations. In other terms, the level of confidence we define (more exactly, we

[4] For a definition of MTBF see Chap. 7.

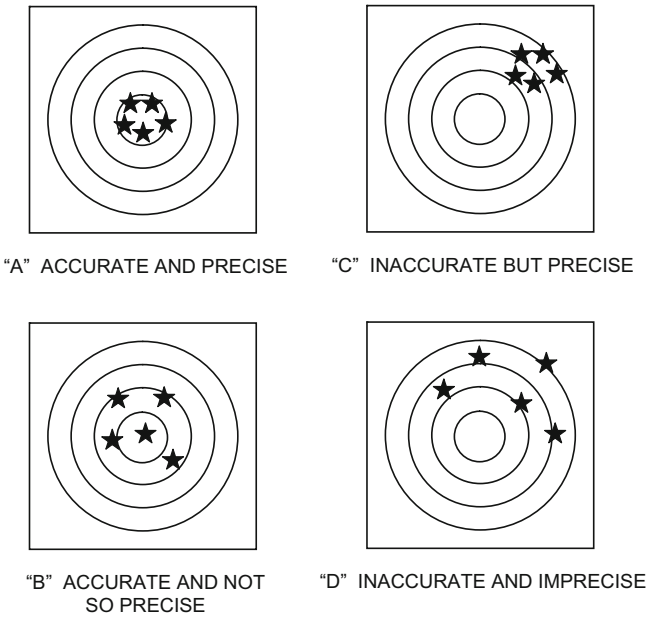

Fig. 10.6 Accuracy and precision are not at all the same thing. **a** Accurate and precise. **b** Accurate and so not precise. **c** Inaccurate but precise. **d** Inaccurate and imprecise

seek to define) will not be precise—but in a large number of cases it will be accurate enough for our job.

Many people, as well as some technical articles, tend to confuse *accuracy* and *precision*. This is wrong. Not only these two terms have different meanings but also they bite into one another. Maybe we like that our measurements are both accurate and precise, but usually we cannot have both accuracy and precision simultaneously:

- Something is *accurate* if it is correct and error-free. In statistics an accurate measurement is close to the expected value.
- *Preciseness* has different meanings which range from exactness to a state of being meticulous, critical, scrupulous, unambiguous, and unbending.

Figure 10.6 presents a graphical definition. Five people are shooting at the same target. In terms of outcome, "A" is highly skilled. His results are both accurate and precise; "B" is accurate though not so precise; "C" is inaccurate but precise; "D" is both inaccurate and imprecise.

In science, and most particularly in engineering, *if* we cannot have the results of "A" *then* we will go for those of "B". In other terms, accuracy is more important than precision. In terms of accuracy in measurements the three higher order non-zero digits will do. We usually, albeit not always, do not need a 7 or 10 digit precision. Interestingly enough, this is also true in business.

- The president of a $100 billion corporation should think in billions and hundreds of millions. If he counts down to cents his attention will be misdirected and his company will go to the rocks.

Accuracy relates to *materiality*. An amount of $1 million is not material for the $100 billion company, but it is highly material to the local firm which makes $10 million per year.

- By contrast, accounting must be precise all the way to dollars and cents. That is what the law demands in most lands, and the letter of the law has to be observed.

That holds all the way to statistical inference. When they are accurate, even if not quite precise, statistical confidence levels are an excellent way to reflect on the likelihood of events, observations, or measurements which will (or will not) occur with a specified degree of confidence.

Accuracy is necessary to be in charge of the variation of a given process, whether in engineering, finance, or other fields. When this is the case, we can utilize a confidence level in order to be certain, in terms of percentages, that a given event will not exceed a particular amount in the envelope of the level of confidence. Through confidence levels, an engineer or other scientist (as well as a financial analyst) is in a position of determining the differential between expected and unexpected events, observations, or measurement. An example is provided in Fig. 10.7 with 95, 99, and 99.9% confidence intervals. The statistics come from a financial study on the change of correlation co-efficients between product lines as a function of volatility. The variable in the abscissa is time. Or, more precisely, the change of volatility over time identified by time series.

As shown in Fig. 10.7, an upper confidence limit is a value larger than the statistic of the central tendency. The opposite is true for a lower confidence limit. The criterion is that in the long run a specified portion of observations, measurement, or test results—hence of the actual population values which interest us—will fall within the so-defined envelope.

- For a two-tailed test of confidence, as the one in Fig. 10.7, the interval between upper and lower limit is the *confidence interval*.
- But the test may also be one tailed. For instance we may be interested *only* in the upper or *only* in the lower confidence limit.

It needs no explaining that the expected population distribution is always important, and so is the sampling procedure. In addition, as it has been underlined on several occasions, we must make sure that the sample is statistically valid, the data are drawn from the same population (a fact which concerns us greatly in connection to confidence intervals), and this continues being so as the number of observation increases.

As with every case of statistical inference, the right sampling procedures, associated to the analytical study we are doing, increases the dependability of the level of confidence and its implied intervals. In the longer run this level assures that a specified proportion of the distribution will fall between the expected value

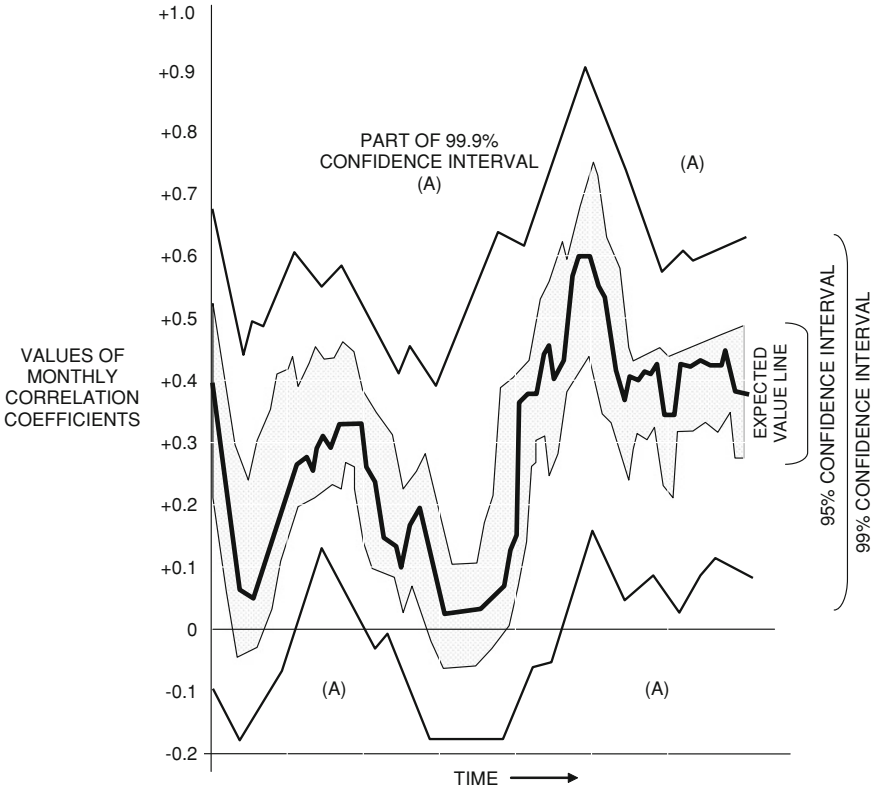

Fig. 10.7 Confidence intervals of correlation co-efficients of two principal variables of a financial risk model

(mean, central tendency) and the limit implied by the level of significance. Only the Type I error, α, will fall outside the so-created boundary condition.

- $\alpha = 0.1$ means that the confidence interval is 90%, and 10% of all cases may fall outside this envelope.
- $\alpha = 0.01$ means that the confidence interval is 99%, and the outliers are only 1%.

The level of confidence is a modeling tool, and as these examples demonstrate one of the major benefits we obtain through modeling is the proper identification and definition of *boundaries*. In all scientific studies, a significant part of the importance played by boundary conditions lies in the ambivalent role of dividing and connecting at the same time.

- Boundaries are places which mark the transition between different conditions, regimes, or functions.

- With this marking, they define different characteristics of the underlying, or its parts, and also reflect the tension which may exist in the limits.

A basic rule of *boundary conditions*, and processes, is fencing off, sealing off what is included in the boundary envelope. There is more homogeneity between points within the boundary, for instance being part of the 95% level of confidence, than across the boundary.

The background concept resembles to one of the famous paradoxes developed by the Greek philosopher Zeno of Elea in the fifth century B.C.: To go from point A to point B, a runner must first reach the midpoint between A and B, then the midpoint of the remaining distance, and so on ad infinitum. Because the process involves an infinite number of steps, Zeno argued, the runner will never reach the destination. The infinite sum 1/2 + 1/4 + 1/8 + 1/16 + … converges to the finite limit 1 but is not equal to 1.

For another examples, if we assume that domestic and foreign letters have on the average the same number of bits, then statistics on transmitted letters tell a story. The ratio of domestic to foreign mail tends to vary between 3 and 87, always significantly more than 1 [2].

In conclusion, it is important to appreciate the meaning of confidence intervals and of boundary conditions. "The 99.9% level is more prudent than the 95%," said a risk management officer in the course of a study, "because with 99% limits are considerably larger than with 95%." But another risk management officer was of the opposite opinion when he stated that "with 95% confidence level traders are more careful since they know the worst case will be exceeded with a frequency of 5%." The latter statement talks volumes about illiteracy in statistical inference, as well as about the wrong psychology associated to boundary conditions.

10.5 Tests of a System

Inspection sampling methods encounter cases where the homogeneity of samples drawn from a given population, or simply supposed to exist, comes into question. A frequently encountered challenge is that involving vendor and consumer. The former submits to the latter the result of a test based on a sample from the lot it delivers. But the consumer is not convinced that this is accurate; hence he, too, takes a sample from that lot and makes a test.

Essentially, what producer and consumer do is to test a system of exchange (goods versus money) from two different viewpoints identified by the now familiar Type I and Type II errors: α and β. The test of a system, of any system, involves the following seven steps:

1. Define the system
2. State the hypothesis (Chap. 8)
3. Select a typical portion: Random sample or representative part
4. Administer the appropriate experiment

5. Observe and record quantitative results
6. Subject these results to a statistical test procedure
7. Decide on the basis of outcome: Whether the system is or is not operating at an AQL.

As this and previous chapters have explained, there are several problems associated to the test of a system. They are ranging from sample size to the homogeneity of the background population under study and the exact methodology used in the two tests (producer and consumer).

- *If* the homogeneity is questionable,
- *Then* the quality level of the inspection is reduced and it may even be questionable.

This is the theme treated in the present section. Let us start with the assumption that two independent samples are drawn at random from a lot. The inspection is by attributes (Chap. 14). Each item in the samples is classified as go/no–go, which stands for conforming/nonconforming. The comparison of fractions defective p_1 and p_2 in the two samples is a *test of homogeneity* of the two samples.

Provided that sampling inspection and testing are uniformly accomplished, the concern will be whether the percentages of defective being observed would be occurring by chance selection on the reason is nonhomogeneity. The question to be answered; therefore, is whether the differences in sampling inspection results between vendor and consumer can be attributed to:

- The luck of the draw in selecting sample units at random from the lot, or
- Real differences distinguishing the two samples, or
- The difference finds its reason in varying inspection practice between vendor and consumer.

For instance, in the latter case among background factors may be improper use of inspection aids, misinterpretation of inspection standards, or failure to select random samples. (It is a sound practice to regard the inspection performed by the consumer as the standard against which the performance of the supplier will be judged).

The test of the system and of homogeneity of its contents is based on critical values for indicating discrepancies which should not be confused with the rejection numbers of sampling plans in determining acceptance of supplies. The decision regarding the dependability of inspection results is distinct from the decision to accept or reject a lot for quality reasons, even if the latter decision may be contingent upon the former.

In the case the test concerns difference in quality inspection practises between supplier and consumer, the consumer must ascertain an *action number* associated with the number of defectives observed and recorded by the vendor. He can then compare the number of defectives he found with that action number. If the number of defects observed equals or exceeds the action number, the consumer's inspector

should adopt a course based on the premise that the discrepancy actually exists in the vendor's inspection system.

Notice that independently of issues regarding methodology there exists the case that the size of the sample used by vendor and consumer differs. The ratio of sample sizes used by the two parties may be equal to 1, 2, 3, or higher. However, for simplicity in the OC curve to be examined we will take this ratio equal to 1 (the two sample sizes are equal).

The frequency rate of the probability of *not* accepting the hypothesis of homogeneity is set with the aid of statistical tables [3]. These show the correspondence between likelihood of acceptance of verification tests. OC of the test for homogeneity are shown in Fig. 10.8 for two equal size samples by vendor and consumer. These are analogous to the OC curves of acceptance sampling plans which we have already studied.

- The ordinate is the probability of acceptance at consumer's side.
- The abscissa gives the ratio of fraction defective, and
- The key variable is the expected number of defects in the vendor's samples.

These curves demonstrate the relationship between a range of apparent quality differences brought about by differences in the vendor–consumer inspection systems and the probability of accepting the hypothesis of homogeneity. *If* the consumer can specify the tolerable ratio of the quality which should be detected as frequently as possible when it exists, *then* the appropriate sample size ratio can be selected—provided the expected number of defectives is in the vendor's samples can be estimated from his:

- Sample size, and
- Process average.

Operating characteristics curves can give good approximation of true probability of acceptance associated with the test for homogeneity. *If* the samples of the supplier and the consumer risk depleting the lot, *then* a special arrangement is required to permit valid comparison of the respective inspection results. For instance, the vendor retains his sample and does not return it to the lot purified of the defectives it contains until the consumer has drawn an independent sample.

In addition, since the incidence of defectives is a small lot is very low, the results from consecutive lots must be pooled until the expected number of acceptance within the desired range for required quality. Alternatively, the consumer can rely in part upon an engineering check of the quality control and inspection system of the vendor.

A double or multiple sampling procedure (Chap. 9) is also possible instead of single sampling. When the supplier elects to use it, some minor modifications are necessary in the verification methodology described in the preceding paragraphs. Check ratings are obtained only for the first sample from each lot, but the critical values of single sampling are applicable to each sample individually or collectively as predetermined.

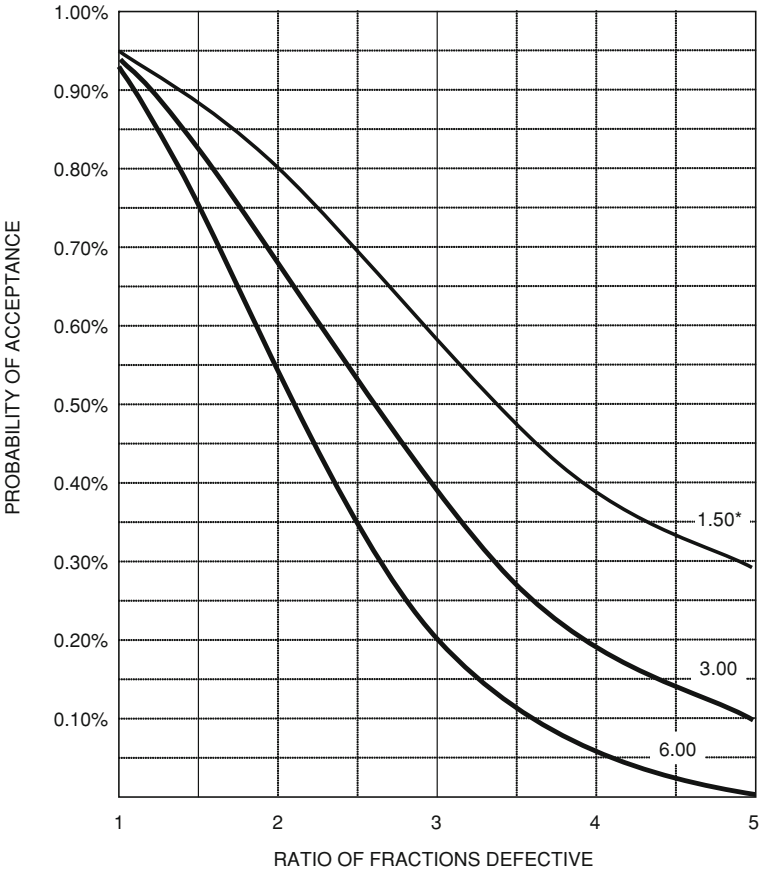

Fig. 10.8 OC curves of two-sample test for homogeneity with variable expected number of defects. * Expected number of defects in vendor's sample

Resubmitted lots may require a larger verification sample because the relative incidence of defects may be smaller than usual, or because the tolerable quality discrepancy ratio and associated risk may be modified to protect the consumer. Whether to pool results of resubmitted lots or not will depend upon the number to be submitted at every case and the expected number of defectives in the vendor's samples.

The test for homogeneity can be instrumental in avoiding arguments between suppliers and consumers, their inspectors and other affected personnel. It is therefore to the advantage of everyone involved, especially in large organizations, to proceed with well-planned statistical tests after standardizing form, content, defect definition, lot and other issues as well as AQL.

Generally speaking the estimate of a lots acceptability is subject to errors. Therefore, there are benefits to be derived from a scientific method. The

methodology described in this section was designed primarily for use in receiving inspection but it is useful in all manufacturing operations. Among other advantages it assists in reducing inspection costs while assuring dependable results.

10.6 Controlling the Hazard of Guesswork Through Experiments

As the reader will remember, some of the case studies in Chaps. 1 and 2 came from the Omega lamp manufacturing company and they mainly concerned wire quality. The provision of comprehensive information on quality is a recurrent problem, usually addressed by tons of paper. As Omega's vice president of Engineering said: "I have one kilo heavy reports. They don't provide me with any clue on quality problems. It is important for me to see in one page the exception. Working in the traditional way it is difficult to decide what is an exception, and that's why I keep on getting these one kilo reports".

Omega's CEO looked at this same problem under a somewhat different light. He wanted that his company establishes a unique quality control system which can be valid through its global operations, and which at the same time observes both international norms and specifications demanded by major local customers. In his words: "A subject like quality control can never be spoken of too much. Our aim should be to have a uniform quality, consistent with sales objectives and with the standards of our manufacturing equipment. This consistence will be the real mark of our products' high quality".

The VP Engineering interpreted the CEO's wish as fulfilling the company's marketing argument: "We are more expensive in our product but this is the best lamp one can find in the market". To make this argument stick, he outlined a complete list of factors which influence wire drawing quality and constancy by addressing the physical and chemical properties of the end product:

1. Wire drawing: Regularity of spooling as a condition for the regularity of wire drawing back tension and uniformity.
2. Elements of wire guidance: Error-free run, size of the angle.
3. Deposition of lubricant: Binding element between wire and lubricant, thickness of oxide layer; porosity of the oxide; sticking of oxide of the wire (more on this later).

Both the lubricant and its baking were signaled out as being important and calling for more attention than it received that far. This greater amount of attention included the lubricant's chemical composition, dispersion of graphite form and size of graphite particles, temperature of the lubricant, method of deposing, and mechanical reliability of the system. In connection to baking of the lubricant, critical factors have been length of furnace, average temperature of furnace, temperature profile in furnace, drawing speed.

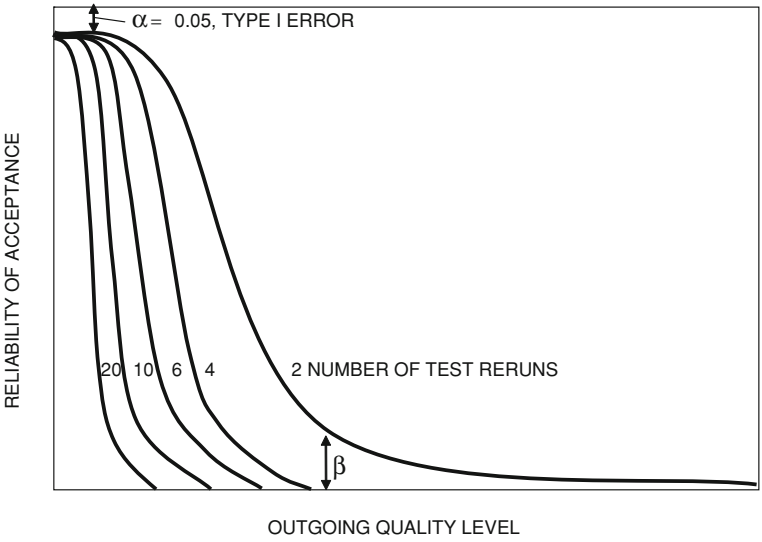

Fig. 10.9 OC curves helped in choosing the number of test reruns (sample size)

For problem No. 3 (the lubricant) was decided a level of significance $\alpha = 0.01$ and consumer's risk β (which cannot be communicated). Four variables were selected as being the most important:

- Binding element
- Thickness of oxide layer
- Porosity of oxide
- Sticking of the oxide of the wire.

The levels of each were fixed by the nature of the test. The big question has been: What level of incremental change, δ, justifies rejection of the null hypothesis? The question was answered by engineering which conducted experiments to document the level of critical difference.

4. Drawing process: Wire approach to the die, including temperature of wire, degree of dryness of lubricant, thickness of oxide layer, thickness of graphite layer.

Conditions specific to the die were: Geometric shape of the die, polishing, length of deforming part relative to diameter, heat transfer between diamond and casing, parallel between the geometric axis of the die and the direction of lateral transfer, lateral transfer of the draw itself, temperature of the interface between wire and die, roundness of the hole. Crucial in regard to drawing conditions were: Amount of reduction and drawing speed. Quality of spooling has been studied in regard to spooling tension, accuracy of spools, and mechanical stability of spools.

All the foregoing factors were proven to have important bearing on end quality, and so did the methodology selected for establishing an orderly approach to the

study and analysis of the aforementioned critical factors. The chosen methodology has been experimental design (Chap. 11) and OC curves.

Knowing the number of variables and the range at which each can vary, facilitated the design of experiments. The total number of retests in an experiment was established by selecting the best OC curve among those shown in Fig. 10.9. Though 20 experiments were estimated to produce the steepest OC curve, it was judged that the accuracy of ten is acceptable.

This is not a choice to be made likely. In all experiments, the most important element is to discover main effects of those variables acting independently. Some experiments also called for an evaluation of interaction effects by variables which, in combination, affect the output. Both issues were present in this research which by means of experimental design (Chap. 11) thrust upon itself the goals of:

- Determining repeatability's accuracy
- Analyzing main effects, and
- Evaluating interaction effects.

Thickness of oxide layer and porosity of oxide were examined for interaction effects. Two-way tables gave the better answer through a comparison of average readings for a combination of two levels of two variables. Similar tables were made for each of the other combinations of variables. Variables which were found to be working together to change the output were treated through Latin squares (Chap. 11). The results were satisfactory as they went well beyond the change that would occur with either variable considered alone.

References

1. Military Standard Sampling Procedures and Tables for Inspection by Attributes (MIL-STD-105A) (1950) US Government Printing Office, Washington, DC
2. Miller JG (1978) Living systems. McGraw-Hill, New York
3. Sampling Procedures and Tables for Life and Reliability Testing (1960) Office of the Assistant Secretary of Defense (Supply and Logistics), Washington DC

Chapter 11
Experimental Design and Latin Squares

11.1 Experimental Design In-the-Large

Practitioners in engineering, as well as in management, face an increasing need for experiments which go beyond the classical notion of laboratory research. What is needed is to plan and administer experimental solution to test hypotheses about alternative designs, manufacturing processes, and other issues where the implications of seat of the pants decisions may be huge.

For instance, in production the industrial engineer has to study wholesale changes due to robotics and generally the automation of work posts, which involve jobs, skills, investments, and organizational problems. He has also to develop a methodology which permits that the results of more limited studies are evaluated against the broader picture of the solution we have adopted. As Albert Einstein once said, the significant problems we face cannot be solved at the same level of thinking we were at when we created them.

Experimental design makes feasible a systematic search for underlying reason of variations and provides documented evidence connected with them. Extensively used in several branches of science, experimental design is a mathematically based approach to the role of the devil's advocate, taking fully into consideration an event's or system's crucial variables. Analytical solutions worth their salt must actively search for weaknesses all the way:

- From hypotheses being made,
- To measurement, test mechanisms, and benchmarks.

Experimental design originated in applied psychology, where it became popular and then spread in engineering design, quality control problems, qualitative decision issues in management practice. It also found good domains of implementation in physics, particularly in cases involving causal inference. If we look in a holistic manner at the problem of causal inference we observe the existence of four quarter spaces which define the confines of our work.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering, 209
DOI: 10.1007/978-1-4471-2966-0_11, © Springer-Verlag London 2013

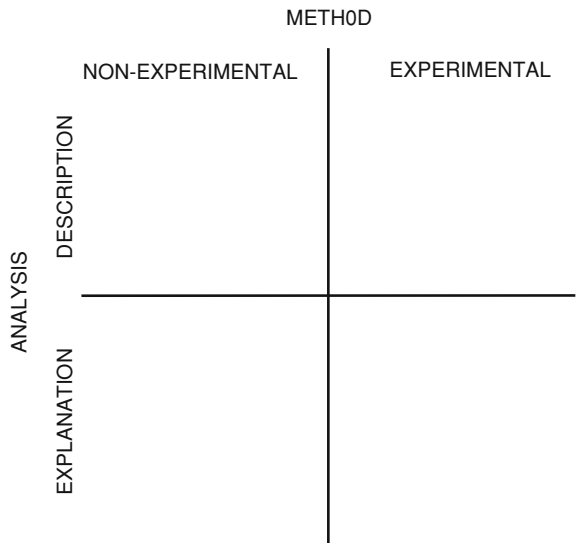

Fig. 11.1 The quarter spaces of causal inference

- The method which we choose to apply can be "experimental" or "nonexperimental" and
- Our choice of analysis may focus on "description" or on "explanation".

As shown in Fig. 11.1, based on this a mainstream approach to scientific reasoning we can draw a matrix in which the rows are description and explanation, the columns experimental and nonexperimental methods. Experimental design resides in the lower part of the right column, a space characterized by the:

- Experimental approach and
- Quest for explanation.

As a method, experimental design is primarily concerned with separating the volatility in interpretations associated with uncontrolled variables from that due to the influence(s) of variables in which we are interested. As such, it provides the experimenter with powerful analytical means helping in avoidance of paths taken by researchers in their work often characterized by arbitrary choices of variables in relation to a single dependent variable or concept.

Without an experimental methodology it is not unlikely that the researcher or analyst selects the variables that will best test the hypothesis he has in mind on the basis of his own hunches. This can be counterproductive in terms of an objective evaluation of data obtained from experiments.

Variations due to experimental conditions, to experimental material, and to errors of measurement, have to be separated into distinct groups. In most cases, the purpose of an experiment is to compare the effects of experimental treatments. By *experimental treatment* we mean the combination of experimental factors which are to be compared.

The notion underpinning experimental treatments is so vital to science that it should be second nature to engineers and physicists. By all evidence, however, it is not yet so. Several experts blame for this shortcoming poor laboratory practices at both preparatory school and university level. Usually the student performs some unexciting, routine experiments because doing so saves time and effort while it needs only standard apparatus.

The problem is that by so doing, the student does not appreciate that experiments which answer "yes" or "no" to a question result from many preliminary (and major) experiments made by competent people over a long period of time. What these competent people were doing was experimental design in the large.

The reduction of the experimental method to a routine has done away with this broader perspective. It is counterproductive for the student to enter the lab to find that the required apparatus is set out on the table for him, along with:

- A sheet or two of instructions and
- A log sheet for the record he or she should keep.

If the main factors entering into an experimental approach are all determined in advance *then* the student and eventually the graduate engineer, does not know the answers to *how* or *why*. As a result, their experience and technical knowledge will be inadequate. By contrast, with instruction in the design of experiments they will learn to:

- Plan and analyze their tests,
- Carry through the experimental operations methodically, and
- Reach conclusions which are factual and documented.

In addition the science student must be taught, and the graduate engineer must appreciate, that a statistically designed experiment cannot completely remove the hazard. What it does is to let the researcher know magnitude of the risks and the steps that can be taken to minimize them. Theoretically, one can learn all that "later on"—but in reality if he has not learned them in his schooling he may never do so.

What has been stated in more general terms about the researcher fits hand-in-glove the work the quality control engineer must do, as well as the person working on reliability studies. Unless the primary factors affecting quality are flashed out, analyzed, and controlled the system will be prone to miscalculations.

This flashing-out is done by means of experiments. To this, there are two prerequisites: The basic reason for an experiment must be thoroughly stated, and the handling and presentation of experimental results must be comprehensible. Such results are usually reported in terms of frequencies, ratios, means, standard deviations, correlation coefficients, and other statistical measures. On the basis of data recorded during the administration of the experiment, the engineer must be able to draw inferences, and conclusions. The inferences' validity would depend on:

- Adequacy of the experimental design,
- Correct conduct of the experiment, and
- Proper administration of statistical techniques.

There exists a solid theory on the study of individual differences by test methods including correlations, analysis of variance, and factor analysis. These are brought together under one comprehensive methodology through the experimental method which, moreover, allows the researcher to make an event occur under the conditions he wishes to study, as well as to separate the key variable in his observations from other events. He can increase his accuracy by:

- Describing his conditions systematically,
- Analyzing his measurements mathematically,
- Repeating his observation under the same conditions, for verification (see Sect. 10.6),
- Enabling other experimenters to *duplicate them*, and
- Allowing an *independent check* on his result.

In conclusion, through the use of experimental design experimenters make a quantum jump in stochastic inference. They can *vary* the conditions under study and note the variation in results in a factual manner. They may follow the old concept of the rule of one variable, in which they holds all conditions constant except for one factor, the *experimental variable*, which they regard as being responsible for the observed variation vary; or, most importantly, they can vary two or three crucial variables in the same run and study their interdependence.

11.2 Simultaneous Variation and Randomization

Section 11.1 concluded with the reference that experimental design permits the simultaneous variation of a problem's two or more key factors. The tool is a disciplined statistical inference; disciplined in the sense that the study is so designed that the experimenter can observe the effects of each variable, as well as its possible interaction with other variables—the theme of this section (see also Sect. 11.4 on factorial design).

A steady, close observance of the behavior of crucial factors in an experiment is critical in most scientific studies because quite often the variation of uncontrolled factors affects the outcome. In this sense, one of the contributions of experimental design is that of helping in properly planning the experiment. A benefit of proper planning is that we can use obtained measurements in a more effective manner in our attempt to reach valid conclusions.

Experiments connected with simultaneous variation of two, three, or more crucial factors have in common the search for a response in a way that may confirm (or, alternatively, demolish) established theoretical systems accounting for observed phenomena. The study of simultaneous variation, and its results, also help in better focusing on our problem. An important contribution by experimental design is that of broadening or deepening an investigation regarding:

- Assumptions,
- Choice of variables, and
- Hypotheses regarding correlations.

The sense of experimentation is promoted by our ability to simultaneously address the variation of two or more variables (see also Sect. 11.6 on Latin squares), searching for correlations or other issues of interdependence. This can also be seen as an audit of past, simpler approaches by providing the researcher with the ability to review and revise yesterday's choices among different theories or options.

One of the frequent failures in the variable-by-variable experimentation is that the definition of inputs is not sufficiently crisp and comprehensive. The same is true of the outputs. In finance, for instance, this leads to misinterpretation of information on exposure, or plain negligence in looking at the accumulated amount of assumed risk.

Problems with two, three, four, or more variables changing simultaneously have been frequently confronting researchers, but the scarcity of powerful tools saw to it that past approaches have quite frequently opted for oversimplification. Today the more powerful, by a margin, is experimental design.

Experience teaches that the choice of oversimplification is made for several reasons. One of them is limited imagination. When people encounter a new phenomenon they try to fit it into an existing framework. Until enough experiments have been conducted, they do not know whether there is really a difference between *this* problem and "other" problems; neither are they clear in regard to the best way to examine it and to report it.

Experimental design, however, requires tools. Latin squares (Sects. 11.5 and 11.6) are a very useful tool. When the number of experimental conditions increases, Latin squares may be used to advantage as a means for careful programming of requested tests in a way to reduce time, test material, and costs—as well as research manpower requirements.

The use of Latin squares improves the researcher's ability to investigate complex relationships at the expense of some accuracy in the variance. (It is impossible to obtain the same statistical dependability from Latin square as from a design which includes in all possible combination the involved factors.)

Another power tool of analysis is *operating characteristics* (OC, Chap. 10). OC curves do not fall precisely under the heading of experimental design, but they can be employed in an advantageous way in conjunction with it. As we saw in Chap. 10 the use of OC curves ranges widely.

Many experimental conditions can be structured in a way emulating producer's risk, α, and consumer's risk, β. An example from banking, therefore outside the engineering and manufacturing domains where OC have been originally employed, is credit risk. Risk adjusted return on capital (RAROC) has been an excellent tool which significantly enlarged the perspective of creditworthiness in banking loans.

When this emulation of producer's risk and consumer's risk is achieved, the use of power curves becomes not only possible but also highly recommended. We should always search for interdisciplinary applications because cross-fertilization is one of the most vital forces in science—as well as a means for promoting objective approaches in quality control problems.

Another tool borrowed from the domain of statistical inference which assist in experimental design is *systematic randomization* of the number of measurements, events, or subjects (Sect. 11.3) needed to fill out the plot of an experimental plan. (Systematically randomization should not be employed if there is reason to believe that one condition transfers to a second condition more than the second transfers to the first.)

As a variation of the systematic randomization, can be employed complete randomization of conditions. The latter involves assignment of order, position, or task based on a table of random numbers. If a large number of measurements, events, or subjects is used, in the long run, each condition will occur about equally often at each stage. This method has a wider applicability. Applications in technology can range from wire production for lamp filaments, the quality control subject discussed in Chaps. 1, 2 and 10, to:

- Circuit boards,
- Nuclear technology applications, and
- High precision mechanical engineering.

Chapter 10 has already provided the reader with a practical example where an experimental design application in lamp manufacturing uses operating characteristics curves to provide a meaningful method for judging the quality level and for improving quality. In experimental terms, three concepts lie behind this approach.

The first, and most important, of the three is the recognition that an element of chance exists in every test result. Every process has different sources of variability which show up in experiments. If not recognized, these can cause errors in the readings. For instance the sources may be:

- The experimenter who conducts the test and collect the data,
- The test equipment,
- The test conditions, and
- The tested issue, entity, or device.

Variations caused by the first two of these factors can be minimized by repeating some tests—using different experimenters and, if possible, different test equipment. Variations resulting from test conditions can be reduced by randomizing or rotating the sequence of the tests. The rotation in Fig. 11.2 is a systematic alternative to randomization (Page 217).

In the case experimental design has been projected to test different devices, the trouble source can be minimized by carefully choosing the sample to be tested (see Chap. 9 on sampling). The reader should notice that part of the challenge in every experiment is measuring the effect on the output when an input (or several inputs) is (are) changed. In addition, it is not only what is measured, but as well how the measuring is done that counts.

In conclusion, the classical way of conducting experiments is that of holding one variable as constant while the other variable is changed. When a point of, say, maximum output is found the "other variable" is held constant and the first one changed to determine if maximum output can be improved. This has many drawbacks such as:

- It is frequently impossible to hold one variable constant over a sufficient period of time.
- In many real life instances two variables acting together have more effect on output than the effect of single variables acting independently on the output.
- The experiment is usually run in a nonsystematic manner, thereby permitting changes in test conditions to bias the targeted true maximum output.

Experimental design in the large permits to overcome such downsides. It is as well advisable to choose a level of confidence as well as to employ operating characteristics curves. In terms of level of confidence $\alpha = 0.99$ is the lowest value to be used; $\alpha = 0.999$ is better.

11.3 Experimental Design Methods in the Small

Deliberately Sect. 11.1 treated as nearly synonymous the terms experimental design and the design of experiments. This is not precise, but it has been necessary in order to introduce the reader to the concept of statistical inference in the large connected to experimental approaches.

Experimental design in the small, which is the more typical use of the term, is based on established statistical methods. The goal of this section is to provide the reader with a snapshot of the four main methods used in experimental design. While particular attention has been placed on cases involving human operators (*subjects*),[1] the same treatment can be done in designing experiments in respect to tensile strengths, chemical compounds, classes of polymers, dosages of drugs and medicines, particular characteristics on machine design, evaluation of electronic components, nuclear engineering issues, economics policies, and more.

Starting with the fundamental notions underpinning experimental design in the small, when the research worker wishes to find the differences in performance produced by two conditions A and B, he has two ways of approaching the problem:

- Using a different group of subjects for each condition and
- Using all subjects under both conditions.

Using a different group of subjects for each condition can be subdivided into two distinct approaches of experimental design. They are known as Design Method I, *random groups*; Design Method II, *matched groups*.

With Design Method I, two separate samples are taken; one group or sample serving under condition A, the other under condition B. For more than two conditions additional samples should be considered. The basic assumption with this

[1] Because, as already stated, experimental design finds its origin in applied experimental psychology—and so do its methods and tools.

method is that random assignment of subjects to groups will result in groups which do not differ significantly on any variable affecting the measured response.

Inherent in Method I, however, is a danger that an observed difference in results could be due to differences in the groups, as such, and not to differences produced by experimental conditions. Therefore Design Method I is not recommended for small groups. Design Method II is better, in this case, and it can be subdivided into:

- Groups matched on the basis of statistical means and variability and
- Matched pairs.

With the approach of matching groups through statistical means, the experimenter should give all subjects practice on a given task and, on the basis of these scores, form two groups. In this way, mean performance and variability will (by all likelihood) not differ significantly.

With matched pairs, the subjects are separated in such a manner that the scores of a member from each group are the same or nearly the same with the scores of the other members of the pair. Obviously, this matching requires the existence of quantitative data on the performance of the subjects. There are two kinds of pretests which have been used as matching tasks.

One is that of giving to the subject different, but highly correlated, tasks. In this case, the research worker must be sure that the task from which he derives scores of matching is highly correlated with the task to be used in the experiment. The alternative is to use matching data from the initial performance on the experimental task in question—which amounts to a prerun.

It is often convenient, as well as feasible, to give all subjects a few trials or, alternatively, a short period of practice on the task to be used in the experiment. Matching criteria can be based on the results of these trials. That much about the subjects; for instance, subjects used in evaluating statistical quality control assignments. But there are, as well, other criteria for choosing a Design Method.

In the case of a critical analysis, it is often better to use Design Method II instead of Design Method I. While theoretically there are no absolute advantages with either method practical experience weights toward Method II. In the general case, both methods have relative advantages, and the decision of which one to use must be based on:

- The requirements of the problem under analysis and on
- Existing limiting factors.

There exist as well Design Methods III and IV considered being more intricate but also more widely used in experiments. Design Methods III and IV employ a single group of subjects for all conditions; in other words, each member of the group is subjected to all experimental treatments but in a varying order.

Because in Design Method III a subject serves in all conditions, this method is applicable with a two-condition experiment, the second condition being influenced by a practice effect, whereas the first is not. It follows that the analyst needs some means that will enable him to equalize the practice effect so that both conditions are on a level field.

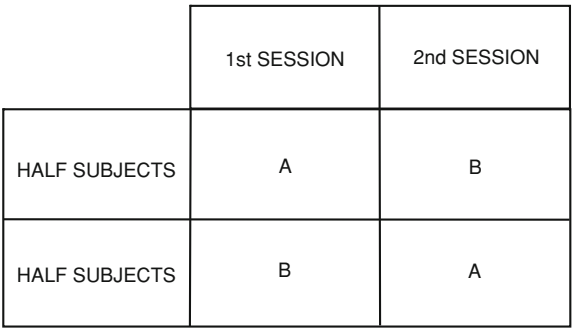

Fig. 11.2 Rotation of conditions A and B to cancel a possible practice effect

As an example, say that the two conditions are A and B. *If* the practice effects were linear, the ABBA order could be used in order to equalize the practice effect. However, experience shows that usually the practice effects are not linear. In this case the researcher must select among the following alternatives:

- Use ABBA for half the subjects, and BAAB for the other half of the subjects,
- Practice all subjects well before starting,
- Make the rotation shown in Fig. 11.2, which is considered by many as being the "proper" Method III for experimental design.

With this procedure, each condition occurs equally often at each level of practice. Therefore, the influence of the practice effect is the same for both conditions. It is, nevertheless, wise that the experimenter always remembers that "counterbalancing" the effect of "this" or "that" factor does not eliminate the practice effect. It only distributes the effect of practice equally over all conditions for all subjects combined.

A brief example helps in better explaining this concept. Let us assume that the experiment involves three conditions A, B, C. When this is the case, the researcher must make sure that each condition occurs equally often at each stage of practice. Say that there are six subjects. The following matrix suggests a way for randomization.

	Session 1	Session 2	Session 3
S1	A	B	C
S2	A	C	B
S3	B	A	C
S4	B	C	A
S5	C	A	B
S6	C	B	A

If the experimenter has available, for instance, 30 subjects, *then* such a three-conditional design can be repeated five times. Design Method III is used with less than or equal to four conditions (see also the discussion on Latin squares in Sects. 11.6 and 11.7). For five or more conditions, Design Method IV is preferred. In the

manufacturing industry and in quality control counterbalancing can be employed to advantage with three kinds of experiments:

- Experiments in which the experimenter varies the nature of the material,
- Those in which the experimenter varies the conditions under which the material is presented, and
- Those in which the experimenter studies practice effects, for instance connected to the use of recently acquired equipment.

Design Method IV, also known as *systematic randomization of conditions* has been mostly used with experiments in which five or more conditions are desired, and in which all subjects are to serve in all conditions. This approach has three basic characteristics: Each condition occurs equally often at each stage of practice, each condition precedes and follows all other conditions an equal or approximately equal number of timers; and all possible orders of conditions are not being employed.

11.4 The Use of Factorial Design

A great deal of experiments in engineering and physics are concerned with two or more variables, each of which may be varied in several ways. The experiment is said to be in *factorial design* when the variables involved in it are studied for all possible combinations (see also the discussion in Sect. 11.2).

Factor analysis starts with a set of observations obtained from a sample by means of a priori measures. It is a method of studying these observations as well as their interaction, to determine whether the variations represented can be accounted for adequately by a number of basic categories fewer than those with which the investigation began.

By means of factor analysis, data obtained with a large number of a priori measures may be explained in terms of a smaller number of reference variables. Take the intelligence of a group of students as an example. The factor analysts may be interested in determining whether the individual differences represented by intelligence test scores:

- Are attributable to a single source of variation, or
- Are they a combination of several mental traits.

The variables involved come in combinations and include reasoning, verbal skill, and numerical ability which individuals may possess in varying degrees. Factor analysis helps to determine the level of association and also selects the essential wholes among the influences at work. Factor analysis groups the numerous possible variables into the fewest possible *single wholes* or their influences. It captures difference of means which may constitute so slight a degree of association as to seem scientifically insignificant.

A radical departure in the concept underpinning this approach is that it does not accept arbitrary choices regarding the important variables, in any field. Furthermore,

it does not simply answer *yes* or *no* in deciding whether a change in any one variable is associated with a change in another. Instead, it:

- Shows how some variables might be grouped together because they behave the same way, and
- Delineates new, underlying factors which may be responsible for these groupings.

The analytical procedure aims to discover and deal with the more massive functional whole, rather than a large number of subjectively conceived variables. Interpretations are done by observing which tests fall on a given dimension and inferring what these tests have in common that is absent from tests not falling on the dimension. Tests correlate to the extent that they measure common traits.

By observing and analyzing the pattern of correlations, the operation of one or more underlying traits or other sources of common variance is inferred. This is done by establishing basic sources of variance in the field of investigation, determining the nature of each measure in terms of basic categories. In this sense, a proper order might be to first use a set of a priori measures in the field of investigation, factor analyzing them to determine the basic traits, or other sources of variance; and, then to study these factors. *Analysis of variance* (Sect. 11.7) can be used to determine:

- How events are affected by different experimental conditions, or
- How they vary among groups that differ with respect to pertinent background variables.

In analysis of variance, a single measure may be administered over a series of occasions and conditions to determine the significance of group differences. Individual differences determined by this approach are often used as an error variance against which to evaluate group differences of events. Like correlation coefficients, the analysis of variance is an important tool allowing to study these events experimentally under carefully controlled conditions.

Individual differences represented by a large number of measures that are given to a single population are studied to detect possible common sources of variation (or variance). The reader should, however, appreciate that factors are not eternal truths. They simply represent metrics of underlying variation in data observed under a specified set of conditions, for given events.

In other words, factor analysis has limitations and to obtain meaningful results who applies it should be skillful in experimental design.

With these notions in mind let us look into a simple factorial design involving experiments with n variables changing in two possible ways. This variation results in 2^n experimental conditions. Among such designs the simplest one would be a 2×2. If a third dimension is added, the experiments become more complex: $n \cdot m \cdot l$; the simplest being a $2 \cdot 2 \cdot 2$ experiment.

As an example of a 2×2 factorial design, say that we are interested in a certain property "A" of a chemical compound. One of the variables concerns the analogy of a component "B", which can vary in two ways. Another variable concerns the

Table 11.1 Statistics on property "A"; high "B"/low "B" immediate, and delayed tests (two conditions of "C")

		High "B"		Low "B"
	Immediate	Delayed	Immediate	Delayed
	4	9	7	6
	5	6	8	5
	6	2	9	4
	3	4	2	6
	7	6	5	7
Total	25	27	31	28
Standard deviation	1.73	3.0	3.0	1.29

time "C" at which testing takes place: An immediate test is an alternative to a delayed test which happens 10 min later. Thus:

- We have $2^2 = 4$ possible combinations of experimental variables, and
- Each of these combinations constitutes one of the experimental conditions.

Suppose that 20 experiments are made, five with each experimental condition, with every other variable than "B" and "C" kept constant; for example, room temperature, and so on. In order to avoid any effect of practice, boredom, etc., each experiment was randomly allocated to each condition. Measurements followed each experiment; the statistics as to property "A" are given in Table 11.1. We compute:

$$\begin{aligned}
\text{Total sum of squares} &= 4^2 + 5^2 + \ldots 7^2 + 8^2 + \ldots + 6^2 + 7^2 - \frac{111^2}{20} \\
&= 695 - 618 = 77 \\
\text{Between groups} &= \frac{25^2}{5} + \frac{27^2}{5} + \frac{31^2}{5} + \frac{28^2}{5} - \frac{111^2}{20} = 620 - 618 = 2 \\
\text{Within groups} &= 77 - 2 = 75
\end{aligned}$$

Table 11.2 presents the results of analysis of variance. The hypothesis of no difference is accepted. The standard deviation, s, and the variance of each group are as follows:

	I	II	III	IV
s	1.73	3.00	3.00	1.29
s^2	3.00	9.00	9.00	1.67

Say that we want to test the variance of the four groups as to homogeneity. Results of the test for homogeneity of variance are given in Table 11.3. We compute the ratio of the sum of standard deviation to the number of groups (or conditions, v). In the present case, $v = 4$.

Table 11.2 Analysis of variance

Source of variation	Sum of squares	f^a	Mean square	Ratio $\frac{v_x}{v_r}$
Between	2.0	3	0.666	$v^2_{0.95}(3.16) = 8.69$
Within	75.0	16	4.68	$v^2 = \frac{s_0^2}{s_x^2} = 0.142$
Total	77.0	19		

[a] f is degrees of freedom

Table 11.3 Test for homogeneity of variance

Group	n	$n-1$	s^2	v^2
I	5	4	3.00	0.47712
II	5	4	9.00	0.95424
III	5	4	9.00	0.95424
IV	5	4	1.67	0.22272
			22.67	2.60836

$$\frac{\Sigma s^2}{v} = \frac{22.67}{4} - 5.68; \quad \log\frac{\Sigma s^2}{v} = 0.75435$$

$$v_x\left(\log\frac{\Sigma s^2}{v}\right) - 4 \times 0.75435 = 3.02$$

The difference, D, is given by the algorithm.

$$D = v\left(\log\frac{\Sigma s^2}{v}\right) - \Sigma \log s^s = 3.02 - 2.61 = 0.41$$

The Chi-square distribution is:

$$\chi^2(f = v - 1) = ln\,10 \cdot (n-1) \cdot D = 2.3026 \cdot 4 \cdot 0.41 = 3.77$$

ln is the natural algorithm. In the χ^2 tables we find that at the level of significance $\alpha = 0.05$, $\chi_p(3) = 7.81$.

The probability of χ^2 being less than 7.81 is 95%: $P(\chi^2 < 7.81) = 95\%$ for three degrees of freedom: $(f = 3)$ and one-tailed test. For a two-tailed test with the same f, it is $P(\chi^2 < 7.81) = 90\%$ (see Chap. 8 on the normal distribution and its tails).

Chi-square tests are essentially a statistical analysis of variance. In every system, the measure of dispersion, therefore variance, is most crucial in establishing performance and quality. As it cannot be repeated too often, small variance is denoting high quality and big variance low quality results.

In the example we just saw, since the computed χ^2 was found equal to 3.77, the hypothesis of no difference is accepted. We may conclude that the variation of the four values of s^2 is within the limits of random variation, sampling from a population with a common variance.

The sum of squares *between* groups can be further partitioned into as many component parts as there are degrees of freedom. For instance, in the present

Table 11.4 $2 \cdot 2 \cdot 2$ factorial design with temperatures 70 and 80°F

70°F				80°F			
High "B"		Low "B"		High "B"		Low "B"	
Immediate	Delay	Immediate	Delay	Immediate	Delay	Immediate	Delay
x_{1i}	x_{2i}	x_{3i}	x_{4i}	x_{5i}	x_{6i}	x_{7i}	x_{8i}
			$(i = 1, 2, 3, 4, 5)$				
$\sum_i x_1$	$\sum_i x_2$	$\sum_i x_3$	$\sum_i x_4$	$\sum_i x_5$	$\sum_i x_6$	$\sum_i x_7$	$\sum_i x_8$

example, between groups there are three degrees of freedom, ($f = 3$). Thus we can analyze it into three meaningful parts, each having one degree of freedom. Namely:

- One based upon the difference in component "B" (high-low),
- One based upon the timing difference "C" in testing, and
- One based upon the interaction of the two variables, level of "B," and timing.

If we were also varying the temperature, say for example that we were making these tests at 70°F and at 80°F, we would have been faced with a $2 \cdot 2 \cdot 2$ factorial design. In other terms, we would have eight experimental conditions. In this case, a number of 40 experiments would be required and they will be distributed as shown in Table 11.4.

The "total," "between," and "within" sums of squares can be found in the usual way, and the v^2 can be computed as before and evaluated against the tabulated v^2 fractile at the α level of significance. The degrees of freedom for between and within groups are 7 and 31, respectively. The Latin squares method (Sects. 11.6 and 11.7) helps in reducing the necessary number of tests.

11.5 Full Factorial Design and Predetermined Confounding

As Sect. 11.1 has brought to the reader's attention that experimental design in the large is the alter ego of modern industrial research. Section 11.3 has brought explained what is meant by experimental design methods in the small. For purposes of statistical inference, the analysis of data derived from research is generally done through two methods:

- Factorial design which can be full factorial or fractional and
- Latin squares, a name which derives from an ancient puzzle.[2]

[2] The puzzle was that of determining in how many different ways Latin letters may be arranged in a square table, so that each letter appears once, and only once, in each row and each column.

Both factorial design and Latin squares are prized methods of analysis. The need for developing advanced methods lies the fact that very frequently in industrial research the need arises for designing an experiment where we seek to find the individual effect of several factors upon an end product. This is a frequently encountered case in *product assurance*, where we need a method which is both powerful and applicable to a wide variety of quality problems.

Typically, though not necessarily exclusively, analysis of variance is the better method to numerically evaluate a critical factor of the experiment, an experimental error, or an important but often neglected detail. It assists in comparing the effect of each variable, for instance, with experimental error to determine the significance of each variable.

- This must be done in a way permitting to determine in a logical manner existing relationships and
- The method's flexibility is important because quality-oriented experiments may involve the simultaneous study of up to eight or nine variables even if typically we will consider only 2, 3, or 4.

A dependable analysis of experimental data is so important because a better product at an equal or lower cost is the final objective of most industrial research. In nearly every case a better understanding of the processes involved is required to accomplish this goal and the research is, therefore, directed toward a determination of the influence:

- Each variable and
- Each individual operating condition has upon the end result.

In many cases, the complexity of the relationships has discouraged needed research on significant quality improvements because the required testing program appeared too complex or too costly to be justifiable. Therefore, a method which reduces the research effort and manpower requirements, hence time and cost; or which improves the ability to investigate complex relationships, is of general interest.

When each of the possible combinations of the levels regarding the chosen variables are tested, the method is known as *full factorial*.[3] For an example on the analysis of quality assurance data consider Fig. 11.3 with four variables: *W*, *X*, *Y*, *Z*. The tests required to assess the main effects and interactions should address the 16 combinations

Full factorial design is a major improvement over the elder method discussed in Sect. 11.1, which calls for changing only one variable at a time—assessing the effects of *W*, *X*, … and so on, by means of repeats. We have to account for interactions and their effect on product quality,

- Enabling quality effects to be estimated with precision, and

[3] We will not be concerned, in this text, with fractional factorial design but will discuss other methods for experimental simplification.

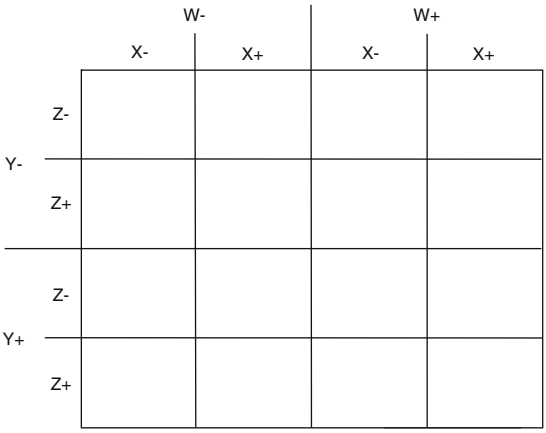

Fig. 11.3 Four variables *W*, *X*, *Y*, *Z*, and resulting 16 test combinations. X- and X+ etc are the two levels of each variable. In this case there are 4 variables and 2 conditions. The algorithm is $2n = 2^4 = 16$ test conditions

- Tracking experimental error for the purpose of assessing the significance of the effects.

A full factorial design not only provides these results but also allows to determine confidence limits. The problem arises when the number of variables significantly increases, even if the levels of each variable are kept steady. As already discussed in Sect. 11.4, the number of test conditions is equal to 2^n. For ten variables there will be 1,024 test conditions; and for 30 variables 1,074,755.824 test conditions.

The explosion of test conditions can be contained by what is known as *predetermined confounding*, a process by which unimportant interactions are deliberately left aside. The researcher is concentrating on main effects and important interactions. The confounding should be chosen intelligently because the analysis is based upon the assumption of negligibility of chosen interactions not taken into account—while we still want to obtain maximum information,

- *If* for whatever reason the confounded interactions are not negligible,
- *Then* this will be revealed when the accuracy of our conclusions is tested.

In that case, then ongoing analysis must concentrate on separating the confounded effects by means of additional testing, their number depending upon the extent of confounding. The usual procedure is to continue the search for all main effects and interactions, advancing in a series of short steps each:

- Involving relatively few observations,
- Being as efficient as possible, and
- Permitting analysis of present data in order to assess ongoing progress, thereby determining the necessity for further testing.

A relatively simple analysis of variance reveals the experimental error in a quantitative manner. Subsequently, a comparison between the variance ascribed to

each variable and the magnitude of the experimental error permits an estimation of the significance of the effect of each variable. For instance, whether the determined effect should be considered real or attributed to experimental error.

The total variance ascribed to each variable may be broken down into components: linear, quadratic, cubic, and so on, limited only by the number of levels originally set for each variable (in the example in Sect. 11.2) and the necessity for going to higher order. (Five levels of a factor permit analysis up to and including the quartic component.)

After we have evaluated our experimental error, we can make a comparison between each of the separate components of variance and the experimental error, thereby establishing whether each component is significant. Subsequent to this, we will have to decide on most suitable way for representing, then for estimating or predicting, the effect.

The downside of this method is that it introduces judgmental, therefore subjective, issues of choice in the middle of data analysis. Even so, however, it is preferable to the widespread method whereby interactions between variables—the condition where the effect of one factor is dependent upon the value or level of one or more other factors—are assumed to be of no importance.

In conclusion, in many quality problems interactions are of primary interest. Effects on quality must, therefore, be efficiently evaluated by means of a sound data analysis methodology. Even an orderly reduction of test, through Latin squares provides an efficient way for planning the experiment. It also leads to an orderly study scheme confined to relatively basic mathematical tools, and endowed with an effective method of presenting the results in a comprehensive form.

11.6 Latin Squares

Latin squares are a statistical methodology for analysis of experimental results where it can be assumed that each variable has an independent effect on the outcome. By contrast, when variables interact among themselves it may be preferable to use full factorial design. This being said, Latin squares could also be worked up to measure certain preselected interactions.

Latin squares are a very important tool for engineers and other professionals involved in data analysis following observations or experiments with many variables, that need to cut experimental time and expense. A typical Latin square application will concern a problem with three variables. A fourth variable (all the way to ninth) can be handled through the Greco-Latin squares which is an extension of the same technique by assigning Greek letters to the fourth variable combining it with the Latin-lettered three variables.

Experimental design by Latin squares is generally applicable to a wide variety of engineering problems and may result in substantial savings. However, as it has already been said, the experimenter should keep in mind that for these savings he pays a certain price: the impossibility of obtaining the same statistical

dependability from a Latin square as from a design which includes all possible combinations of the involved variables, or factors—for instance, a full factorial test (Sect. 11.4).

Some years ago, for example, a Greco-Latin Square solved a problem in receiver circuitry for Texas Instruments in 275 measurements that would have required an inordinate time and effort by the old "one variable at a time" approach. The importance of Latin and Greco-Latin Squares is attested by the fact that they have been used as tools for predicting failure hazard, with implementation extending in quality control, and reliability programs.

Consider as a practical example a test of three variables, each varying at four levels. The variance of a treatment mean for a $4 \cdot 4$ Latin square is $\sigma^2/4$. If all 64 combinations had been run, the variance of a treatment mean would be $\sigma^2/16$, which is better than the one offered by the Latin square design.

The 16 combinations[4] used in the Latin square make up one fourth of the possible 64 combinations. It is obvious that ¼ the combinations (those in the Latin square) does not give as much information as the total. The advantage is that one can test for the significance of main effects (rows, columns, and treatments) more economically, even if at somewhat reduced statistical dependability. However,

- *If* there are no interactions among the factors,
- *Then* the Latin square solution is sufficient to carry out all important tests (again subject to the aforementioned constraints).

Let us assume that in planning our experiment we have chosen the experimental factors and levels of these factors. The experimental factors may refer to particular environmental conditions, to timing, to the nature of material that has been used, or to the technique that has been applied.

Say, for example, that a civil engineer is testing two different mixtures of reinforced concrete, and for each mixture he is using two different qualities of cement. Obviously, he is faced with a $2 \cdot 2$ experiment. If he wants to expose his mixtures in two different time intervals, he introduces another independent variable—time. The result is a $2 \cdot 2 \cdot 2$ experiment.

For his experiment, the civil engineer will need eight samples, four from each quality of cement, and he will end with only one piece of information for each quality-mixture-timing interaction. If he needs a larger sample, he will have to perform as many experiments as eight times the sample size he requires. If he was testing three different mixtures, for three different qualities of cement, in three different time intervals, he would end with $3 \cdot 3 \cdot 3$ experiment.

In the following example, say that A represents quality of cement (thus A_1, A_2, A_3 represent the three different qualities); B type of mixture (B_1, B_2, B_3, the three types); and C timing (C_1, C_2, C_3 the three time intervals). The result is 27 possible

[4] Remember that in the example in Sect. 11.5 these were four variables each varying in two states (+,−) and $2^4 = 16$. In the present case there are four variables and four levels. It is always $4^4 = 64$.

Table 11.5 Twenty-seven combinations of experimental treatment

$A_1B_1C_1$	$A_1B_1C_2$	$A_1B_1C_3$
$A_2B_1C_1$	$A_2B_1C_2$	$A_2B_1C_3$
$A_3B_1C_1$	$A_3B_1C_2$	$A_3B_1C_3$
$A_1B_2c_1$	$A_1B_2C_2$	$A_1B_2C_3$
$A_2B_2C_1$	$A_2B_2C_2$	$A_2B_2C_3$
$A_3B_2C_1$	$A_3B_2C_2$	$A_3B_2C_3$
$A_1B_3C_1$	$A_1B_3C_2$	$A_1B_3C_3$
$A_2B_3C_1$	$A_2B_3C_2$	$A_2B_3C_3$
$A_3B_3C_1$	$A_3B_3C_2$	$A_3B_3C_3$

combinations of experimental treatments shown in Table 11.5. If the civil engineer wanted to have complete information on the interaction of his key variables, he should perform all 27 combinations. In this case he would end with one piece of information for each A–B–C interaction.

To give some muscle to this example, suppose that because of some technical, economic or other reason the civil engineer was unable to perform all these experiments and he reduces his experiment by considering only the first two variables: A (quality of cement), and B (type of mixture). Each can vary in three ways:

	A_1	A_2	A_3
B_1			
B_2			
B_3			

A Latin square solution allows him to reinstate independent variables of timing by distributing C (timing) so that no two same C-treatments would appear in the same row or in the same column:

	A_1	A_2	A_3
B_1	C_1	C_2	C_3
B_2	C_3	C_1	C_2
B_3	C_2	C_3	C_1

In this way when comparing the A (column) means, the A-comparisons would seem to be balanced so far as B or C effects are concerned, since each of the three B treatments and each of the three C treatments appear only once in each column. The same is true in respect to the B and C comparisons.

This is a good example of Latin squares contribution to the analysis of experimental data. Let me repeat what I have already stated. A full factorial design would have required 27 experimental treatments. Latin squares and it asks for only nine combinations of experimental treatment instead of the 27 which we have seen in Table 11.5.

Regarding the "price" to be paid for the simplification, the reader should notice that, when making the A comparisons, in Latin square matrix we have for A_1 the interactions: B_1C_1, B_2C_3, B_3C_2 while for A_2 we have the interactions: B_1C_2, B_2C_1, B_3C_3. In that sense,

- The effects of the B and C treatments are balanced as for A_1, A_2.
- However, on the other hand, the interactions between B and C may or may not be different.

B_1C_1 may have some inherent interaction which would give results other than the pure effect of B_1 and of C_1. A given interaction may be nonexistent in the case of B_1C_2. A certain type of mixture may respond quite differently in a short time interval than in a long time interval, and this response may not be counterbalanced by the other combinations. Thus Latin squares may give a "defective" response unless one can assume that there are no interactions of the factors involved.[5]

The fourth paragraph of this section made reference to an application of Greco-Latin squares at Texas Instruments, and it was stated that they constitute a technique able to take care of additional variables or chance variations which may affect the experiment. Generally speaking, a Greco-Latin square is an arrangement of r Greek and r Latin letters in an $r \times r$ square, so that each Greek and each Latin letter appears once and only once in every row and column.

D δ	C β	B α	A γ
C γ	D α	A β	B δ
B β	A δ	D γ	C α
A α	B γ	C δ	D β

With such an arrangement, a matrix of four independent variables may be tested using only 16 observations. A number of independent variables can be studied in an experiment with relatively few tests, at the price of a less dependable analysis of variance (Sect. 11.7).

The question is sometimes asked how to deal with variables which do not have discrete levels, but are continuous functions—which sees to it that potentially there is an infinite number of combinations. When confronted with this situation, the experimenter may start by stratifying the continuous measurement, assigning a number of arbitrary discrete levels. He then combines them in a Latin square or factorial design to get a rough preliminary estimate.

The obtained result can be subsequently refined by regression analysis. A plot of three variables will give a 3-dimensional surface, known as *response surface*, representing the relationships between the variables. The exact mathematical shape of this surface can be found by means of data reduction from the preliminary experiment, leading to a set of regression equations which show the best combination of values for the variables.

[5] A fact already brought to the reader's attention.

A particularly challenging problem is the use of full-fledged factorial analysis or Latin squares for statistical prediction of stress conditions or environments. These cannot be predicted particularly well even in destructive or life tests as it is often done for reliability evaluation. Sometimes it is difficult or outright impossible to duplicate environments in a laboratory, but even in borderline cases factorial design presents interesting possibilities which should be fully exploited.

11.7 Latin Squares and Analysis of Variance

A critical issue (if not *the* critical issue), connected to every experimental design is the analysis of variance. As a reminder, in statistics *variance* is a measure of the variation, or scatter, about an expected (mean) value. However, in connection to experimentation and plenty of other applications the term "variance" is used rather loosely. As far as a Latin squares methodology is concerned, variance is looked at as synonymous to "sums of squares," or "mean squares" used for analysis. A large value for variance indicates that individual observations:

- Cover a relatively large range,
- Or come from a population with huge spread, hence low quality.

An important mathematical characteristic of variance is that it is additive. The total variance is the arithmetic sum of the variance associated with factor "A", the variance associated to factor "B" and so on, plus the variance pertaining to experimental error. This makes it possible to quantitatively evaluate experimental error once the total variance is known.

$$v_{\text{error}} = v_{\text{total}} - v_{\text{A}} - v_{\text{B}} - v_{\text{C}} \ldots$$

Having evaluated the variance due to experimental error, the variance associated with each factor may be compared with it through a fairly simple algorithm:

$$\text{variance ratio for "A"} = \frac{v_{\text{A}}}{v_{\text{error}}}$$

From knowledge of this variance ratio, the experimenter is able to decide whether or not v_{A} is significant. It is obvious that v_{A} must be fairly large in comparison with v_{error} before we are justified in deciding that factor "A" really does have an effect upon the end product. (It is also important to be sure that v_{A} is not the result of experimental error.) We must, as well, determine the level of significance of the obtained results.

In the sense of the test we have been obtaining, variance is the sum of squares of the differences from the mean, divided by the observations n, minus one. This quantity $n-1$ is the number of *degrees of freedom* (df). The degrees of freedom are mathematically equal to the number of information elements for a given value of variable minus one. The algorithm for the variance is:

$$v = \frac{\Sigma(x_1 - m_1)^2}{n - 1}$$

where
x_1 is observed value of end product; and
m_1 is mean value of end product.

In daily practice it is often convenient to modify the above equation into:

$$\text{Mean squares} = \frac{\Sigma(x_1^2) - \frac{(\Sigma x_1)^2}{n}}{n - 1}$$

Mean squares is a term closely related to variance, particularly applied to that property of variance employed for determining the variance ratio: F. The *sums of squares* is a term applied to the quantity $\Sigma(x_1^2) - \frac{(\Sigma x_1)^2}{n}$, which is the sums of the mean squares of the deviations from the mean.

The analysis of variance is a test of significance for variation in experimental conditions. It will tell the researcher whether the assignable causes of variation are present or not. In undertaking an experiment, the engineer usually makes a hypothesis about the existence or the nonexistence of certain conditions. What the researcher knows in advance is that the obtained data will show variation. The question to which he seeks an answer is whether:

- This variation is significant
- Or is due to chance effects on the various controlled factors and/or experimental errors.

As an example, consider conducting an experiment to test the effect of various water temperatures on a certain material. Say that four people make readings of the water temperatures, using for thermometers. We would like to know whether there is an assignable cause of variation because of differences in the people making the reading or in the thermometers. The procedure is simple:

- We formulate a null hypothesis and
- The analysis of variance gives us a means for testing this hypothesis.

The null hypothesis states that there is no difference (more precisely that there is no significant difference in effect). The alternative hypothesis states that there is a difference. *If* from the statistical analysis of the experimental results we conclude that the null hypothesis should be sustained, this would obviously mean that the outcome of the experiment provides no statistical basis to prove that:

- The effects of the two distinct treatments differ in a significant manner.
- But in no way it will mean we are sure there is no difference.

All we can say is that this experiment has provided no evidence there is a difference at stated level of confidence. To make this example with four

thermometers and four people a little more interesting, and more involved, say that in addition we also have four tanks of different materials. The effect of material differences may affect our results and mask what we are interested in knowing about the people and the thermometers.

A Latin square design can provide us with a means of setting up a test while eliminating any effect due to the difference in material of the tanks. The arrangement is presented in the following matrix with rotation of the letters A, B, C, D indicating a method of assigning tanks so that their effect can be eliminated.

Thermometers \ People	1	2	3	4
1	A	B	C	D
2	D	A	B	C
3	C	D	A	B
4	B	C	D	A

The advantage of this arrangement lies in the fact that it weeds out possible correlation effects due to types of material in the tanks. If we could be sure there were no such influences at work and that the tanks were entirely the same, it would not matter which tanks were used.

The flexibility of experimentation using Latin squares allows us to arrive at a result concerning people and thermometers by accepting that no tank appears more than once in the same row or column. By doing so, we remove the horizontal and vertical effect of different tank materials in our comparison.

The foregoing example made use of three independent variables: people, temperatures and tanks, varying atfour different manners each. This required the determination of total degrees of freedom: $4 \cdot 4-1 = 16-1 = 15$. In the Latin square nine degrees of freedom account for rows, columns, and tanks. The residual is six degrees of freedom.

Associated with these six degrees of freedom is a remainder, or residual sum of squares, which may also be obtained by subtraction: Total minus rows, columns, tanks. The residual mean square based upon six degrees of freedom may be used for testing the significance of the mean for rows, columns, and tanks:

Source of variation	Sum of squares	f	Mean square	Ratio $\frac{v_x}{v_r}$
Rows	...	3	...	...
Columns	...	3	...	...
Tanks	...	3	...	...
Residual	...	6	...	...
Total	...	15	...	...

Part V
Statistical Quality Control

Chapter 12
Fundamentals of Statistical Quality Inspection

12.1 Measured Quality of Manufactured Products

Measured quality of manufactured products is always subject to a certain amount of variation as a result of chance. A system of chance causes is inherent in any particular scheme of production as well as of inspection. Variation within this pattern is inevitable. The reasons for variation outside this pattern should, however, be discovered and corrected.

Nothing walks on a straight line, said Werner Heisenberg, the physicist. This is reflected in statistical quality control systems (SQC) and procedures—the theme of this and of the next three chapters. SQC systems and procedures can be defined as the use of probability theory in problems of quality inspection and product assurance. The charts associated to them are the visualization of a quality control analysis.

- Instrumental in inspecting, improving, or restructuring production systems
- Able to provide secure information to be used in establishing more effective inspection and/or acceptance procedures
- Coming forward with evidence to hunt for causes of variation, and taking action intended to correct them
- Providing a basis for current decisions on acceptance or rejection of manufactured or purchased products, and
- Capable of familiarizing personnel with the importance and use of control charts, as well as information on product quality and process accuracy derived from them.

Both products and also manufacturing processes have tolerances. Through the disclosure of natural tolerances and control limits of a production process, the control chart permits better decisions on engineering tolerances and improved comparisons between alternative designs and/or between alternative production methods.

Through improvement of conventional acceptance procedures, SQC can provide a significantly higher quality assurance at lower inspection cost. It can as well

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering, 235
DOI: 10.1007/978-1-4471-2966-0_12, © Springer-Verlag London 2013

tell in a documented manner the process behaves as expected and, therefore, it should be left alone.

The first conscience of the need of a well established procedure for quality control goes back to the nineteenth century, when industry became aware of the impact of product quality—placing responsibility for it on the line organization. At that time, however, there were no inspectors as we conceive them today and few designers appreciated the need and impact of quality assurance. A major step forward was made in the 1920 and 1930s with:

- The production line becoming an independent department,
- Advent of scientific management, and the realization that inspection and was a specialized job which work differed from production work.

With this, inspection gradually emerged a recognized self-standing function, called to screen the bad stuff from the good products prior to shipment. The number of inspectors grew as factories became bigger and their tools sharpened up. Statistical methods were developed during World War II, making quality control a sophisticated enterprise.

With these evolutionary steps in mind and the growing emphasis on product assurance quality control, and most particularly SQC should be viewed as a tool which initiates and documents—therefore influences—decisions about product acceptance. This is in a nutshell a definition of what *quality control* is all about, and it can have two interpretations:

- A narrow one centered on measured quality of manufactured products which in Part I we called quality control in the small, and
- A much wider interpretation which promotes quality and reliability from engineering design to production, inspection, and field maintenance which has been labeled quality control in the large.

Sustaining quality is a demanding process (and so are, incidentally the famed apple pie and motherhood). Whether in the small or in the large product assurance confronts two problems: errors in measurement and faulty procedures. Both can be seen as opportunities for action. Errors, John von Neumann has said:

- Are not adverse or unwanted aftermaths of the study we do, or the method we are using.
- They are integral part of our work, and we should be using their existence to our advantage.

In control systems, for example, errors make feasible to use of *feedback* which has opened wide perspectives in engineering design and product development. Figure 12.1a presents the simple feedback most frequently found in applications from servomechanisms to automatic controls. In a more sophisticated implementation, which capitalizes on expert systems, the feedback uses knowledge engineering and learns as it goes along. This is the example of the block diagram in Fig. 12.1b.

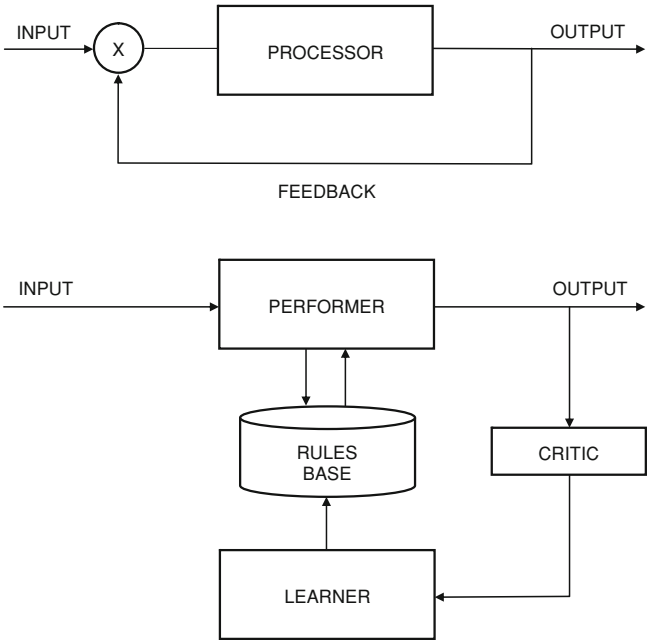

Fig. 12.1 Feedback can be simple as in a servomechanism or sophisticated based on knowledge engineering

A sophisticated approach to feedback and inspection is a leap forward from the classical theory of measurements and their errors, which has assumed that they were independent and normally distributed. That near-sighted approach was developed mainly as a model for the distribution of errors of measurements occurring in astronomy, where sources of variability existed in the nonuniformity of experimental techniques.

By contrast, the modern, scientific way of looking at errors is that they are produced by a system of chance causes which exist not only in manufacturing and the provision of services, but as well in any type of experiment, or measurement, we are doing. The causes may be quite diverse. Errors might find their origin in:

- Heterogeneity of experimental materials, or
- Variations of manufacturing conditions not being under control.

In many cases of quality control endeavors we practically have no clear-cut assignable causes, yet we must deal with the existence of such errors. To our advantage, we have some clues of cause and effect. For instance, the heterogeneity of the experimental materials might show a *systematic* and relatively large variation of experimental results. But random errors do not behave that way.

Errors can have many causes, some of them transparent. There is indeed a significant difference between errors due to instruments and methods, and those

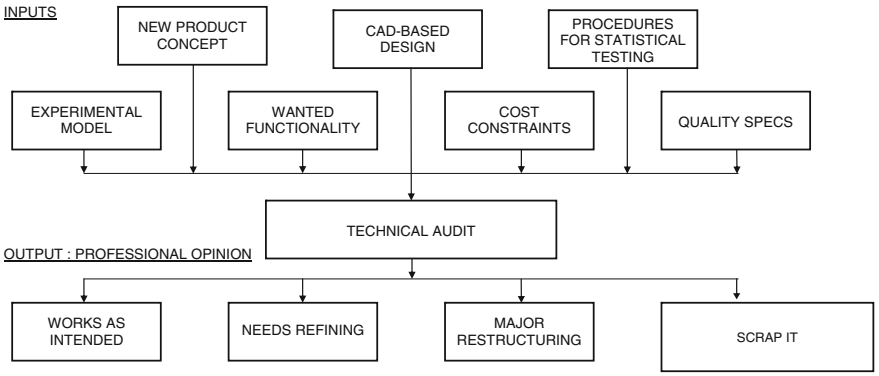

Fig. 12.2 Inputs to a technical audit and professional opinion provided as output

resulting from the experimental factors themselves. The principle in scientific experimentation is that:

- *If* variations caused by nonexperimental factors are not in a state of statistical control,
- *Then* the experiment should not be administered.

Factors which bring the experiment outside the range in which statistical experimental control is effective, have first to be corrected or compensated. Only then can valid conclusions be drawn by means of statistical methods. (The first known compensation has been to allot the treatments to the experimental units at random. R.A. Fisher, the British statistician, was the first to produce randomized experiments). As we saw in Chap. 11 factor analysis and Latin squares are the modern way for dealing with this problem.

A similar statement can be made about quality control in manufacturing. As the opening paragraph of this chapter brought to the reader's attention measured quality of manufactured products is always subject to a certain amount of variation due to chance. Ideally this is the stuff to which, ideally SQC addresses itself. Gross errors may as well be due to organizational reasons. When this happens, responsible for their correction is senior management.

12.2 Organization for Quality Control

Based on the principle that quality has to be embedded into the product at the drafting board, which has been brought to the reader's attention in Part I, the block diagram in Fig. 12.2 defined the inputs and the outputs of the quality equation in an enterprise. The variables of technological development have to be enlarged to reflect the use of statistical tools.

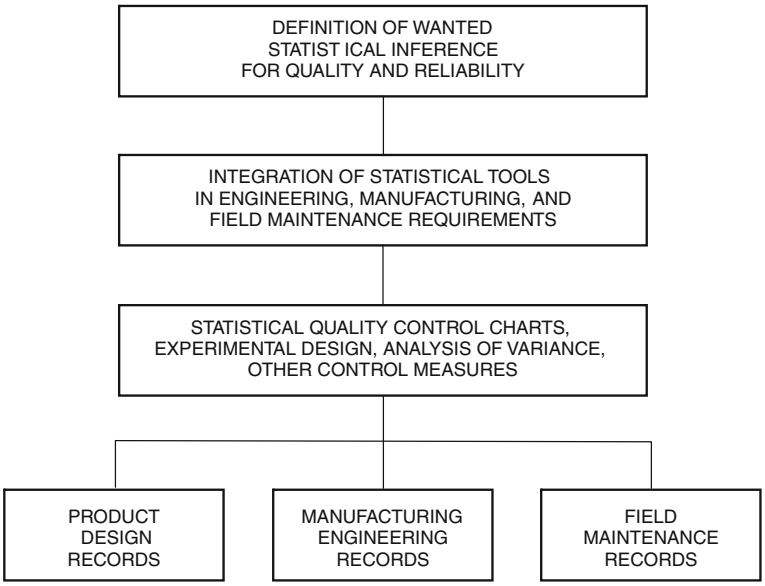

Fig. 12.3 Senior management should audit whether means for statistical inference have been incorporated since the drafting board

The purpose of this simplified graph is to aid in recognition of the level of product design where procedures for statistical testing must join the product assurance premises already in product design. Quality assurance will be significantly assisted by statistical analysis of data, keeping well in mind that statistical methods useful in engineering design are far from having been exhausted.

As far as quality control in the large is concerned design engineers can perform so much better if they are familiar with, and incorporate in their work, a statistical testing procedure. Later on, this will become a milestone in the company's inspection system enabling it to assure that all components and finished products (as well as all supplies and services) conform to specification requirements.

Though participants to my seminars do not particularly like to hear me saying so, I never fail pointing out that in several cases the quality control systems and practices that manufacturing organizations are using to guarantee quality levels for their products are of the 1920 and the 1930s. This is in dissonance to:

- Higher quality demands by customers, and
- The use of more technologically sophisticated materials.

Taken together, these background quality testing conditions suggest that in many plants quality assurance practices are now obsolete, or fast becoming so, which explains why quality costs have increased astronomically in the last decades. In many companies they are between 7 and 10% of sales. To bend the cost curve, senior management should be keen to audit whether means for statistical inference have been incorporated since the drafting board.

Figure 12.3 gives a concise view of the steps involved in this function, which is needed to substantiate the emergence of holistic quality control as a technically recognized full-time function. In a way quite similar to that of risk management in a financial firm, the quality control in the large responsibility in a department of a manufacturing enterprise should be standing high up in the organization.

There are several reasons for this structural position, foremost among them being the fact that independence of opinion in quality control is a "must." Therefore the quality control department should depend neither from engineering (a rare case) nor from manufacturing (its usual remit).

It should be free of conflicts of interest to perform inspections and tests required to substantiate product conformance in terms of drawings, specifications, and other requirements. A quality in the large function must also perform all inspections and tests required of products and processes, informing in the most timely manner engineering, manufacturing, and maintenance of their deliverables.

Quality control is promoted if the results of testing and inspection are prescribed in a clear, complete, and unambiguous way. The results of tests inspections and audits should be readily available through online access. The management of adequate records of all inspections and tests is not an option; it is a responsibility including the:

- Nature and number of observations made by product or process,
- Number and type of deficiencies found,
- Quantities approved and rejected, and
- Nature of appropriate corrective action being taken.

Corrective action is the alter ego of quality control. The responsible department must always be ready to correct assignable conditions which resulted, or could result, from testing or inspection. Database mining can be instrumental in defining patterns for repeated events which time and again handicap product performance.

Another requirement is that the company's inspection system also audits whether procedures are in place to assure that drawings, specifications and instructions required by production are always up to date and accessible online. Such procedures should be integral part of the inspection system with certification regarding the timeliness, accuracy, and condition of updates.

Like well-planned experiments, quality inspections require the maintenance of detailed records throughout all stages of product performance. In fact, not only data on inspections and tests *per se*, but also checks made to assure accuracy of inspection and testing people and equipment should be data based. In short, all quality control related records should be available for consultation and review, and this will also be useful in the next quality control study.

The objective of never relinquishing the quality assurance responsibility is to obtain a high degree of exactness and efficiency which may also contribute to reducing the cost of the inspection function while assuring product quality. Example of an area where investigation may provide welcome inputs are the sampling plans we examine in Chaps 13– 15 which supply rational techniques, for:

- Mass observations, and
- Assessment of frequencies, means, variances and other elements of individual, and collective description.

Quality control will be often asked by senior management explanations based on observational data, particularly in connection to quality costs which are mainly:

- Prevention costs, incurred in keeping defects from occurring in the first place. Included here are such costs as quality control engineering quality training, and quality maintenance.
- Failure costs caused by materials and products that do not meet quality specifications—scrap, spoilage, field complaints which can rise above 25% of total costs, and
- Appraisal costs including expense of maintaining company quality levels by means of evaluations of product quality. They cover such elements as inspection and laboratory acceptance testing.

Costs of any type, and to this quality costs are in no way an exception, should always be under scrutiny. In the connection, a statistical quality control plan can contribute in reduction of overall costs by a cut in failure costs; swamping defects; improving the general level of product quality; lower appraisal costs; and improving inspection and test methods. An example is the replacement of routine inspections by fewer but more effective SQC inspections.

12.3 Responsibilities in Quality Control

Section 12.2 made the point that though with modern techniques we try to minimize the sources of variation, variation still exists, and will continue doing so. Therefore, the manufacturer must have a sound quality control system in order to be sure of what he is going to offer to the market is a product whose quality leaves nothing to be desired.

This in no way means that every product has to hit an unprecedented quality level. As a matter of principle, too high quality may be just as bad as too low, because it prices the product out of the market. The outgoing product quality has to be balanced and correspond to that which is demanded by the market and offered by the product. This being said, it is a wise strategy to assure that (other things equal) *our* product is of somewhat better quality than that of competitors. This involves five basic quality control responsibilities aimed to uphold quality while swamping costs:

1. *New design control*. This involves analyzing the "quality ability" of a new product or process, as well as "debugging" quality problems to give a defect-free outcome. Part of this mission is the planning of inspections and tests to be carried on when production is under way on a new product, incremented by establishing continuous control of in-process quality. Another responsibility, of

which the reader is already aware, is the design of inspection and testing equipment that can be integrated into manufacturing processes to permit them to check their own work.

2. *Incoming materials control*. In performing this duty quality control engineering must assist in the establishment of sound quality relationships with vendors by means of planning the periodic rating of quality performance of present suppliers, evaluating the quality capability of potential suppliers; working with vendors in understanding the quality requirements; and establishing quality-certification programs for all purchased materials, component parts, and equipment.

3. *Cost control*. The exercise of this responsibility requires that quality control engineering carries on the cost measurement and quality cost reduction activity required for cost optimization. This requires handholding with both design engineering and manufacturing, enriching each product or process with checks and tests against subjective evidence. Evidently also integrating statistical methods into the decision process.

4. *Special process studies*. This includes analysis of complex in-process quality problems that have been fed back by engineering, manufacturing, field maintenance, or special inspection. The core of this work should be experimental design techniques (Chap. 11) which allows to objectively examining quality control premises, tearing down the citadel of medieval practices which are still around—largely untouchable because their need is "obvious."

5. *Outgoing product control*. In this the most obvious mission is that of making sure that the customers are supplied with dependable quality and that the outgoing product meets specifications and tolerances. Well-managed organizations appreciate that outgoing product control will be so much more effective if the different company departments are not "silos" but collaborate in the development of testing procedures, sampling plans, outgoing quality database, and timely examination of feedback data for corrective action.

As we have already seen quality inspection is promoted by both observational and experimental data. Observational inspection is done by variables or by attributes (respectively Chaps 13 and 14). In attributes sampling, randomly selected samples are inspected to determine whether the quality characteristics of the item are within specification limits through a "go, no go" approach. By contrast, in variables inspection a specific quality characteristic is actually measured, sample means plotted, and inspected whether they fall within control limits.

It is irrational not to use effective sampling plans developed over the years. These are associated to normal, reduced, or tightened inspection. The latter is invoked when the process average exceeds the statistically based control limits for a specified quality level. Existing systems and procedures make it feasible to proceed with frequent periodic computations of mean values which enable to detect, at an early stage downward trends in quality.

As we have seen in Chap. 10, when quality levels and sampling risks (producer's and consumer's risk) associated to valid statistical procedures are employed, the result is a dependable estimate of quality levels and risk values. Normally, these must be based on full knowledge of technical design requirements of the product and of its manufacturing process. Experience in SQC is a major "plus."

A mark of distinction in quality control is to be ahead of the curve of adverse events and bend it before it raises its head. This cannot be achieved without experimentation, and the standardization of reporting procedures. Chapter 11 brought to the reader's attention that it is sound practice to carry out tests under identical conditions, even if it is difficult to obtain such uniformity, due to the natural variability of test equipment which (bought over a period of time) does not follow always the same standard.

The effect of human elements conducting the study, procedural changes happening over time, and environmental influence add to the variability. Its planned elimination, whenever this is possible, is welcome because in a way it broadens the experienced conclusions rendering our data more applicable. Within this perspective, it is advisable to randomize:

- Sequence of testing, and
- Selection of test materials.

Having randomized with respect to test sequence (subjects, material, and time), we are more confident that if time, material, or operators have a definite effect, this too will be randomly scattered throughout the results. The downside of this procedure is that it might increase experimental error, but at the same time it eliminates possible bias. In a well-planned experiment on product assurance, all sources of variation should be considered in order to reach objective conclusions.

Objective conclusions are one of the quality control's most basic responsibilities, and this underlines the importance of statistical methods in conjunction to sound engineering practices. "Systematize, then mechanize" is a sound guideline in quality control, because what is most important is not technically elaborate quality control equipment, but skill on behalf of those who choose them and use them in a systematic, well-ordered manner.

The automation of part of quality control assurance work requires far better procedures for determining the quality capability of new designs prior to production, than that which was required in conjunction to manual intervention. The preparatory work demands skill and detail and has to take place ahead of the use of complex test equipment. This policy requires that senior management:

- Emphasizes the need for on planning and designing the quality inspection system;
- Evaluates through tests and inspections the quality capability of current and potential suppliers, and
- Gears quality control to detect problems before they become costly like the not-so-infrequent recalls by auto manufacturers of their motor vehicles.

Section 12.2 has underlined the need for quality control records to be utilized not only for in-process but as well for incoming product inspections. Inspection during manufacturing and final assembly will be half-baked if the procurement sources are not subjected to quality control. A similar statement is valid of the calibration of inspection equipment and certification of special processes even if one tantum.

When statistical process controls are utilized as quality evidence for determining acceptability of a lot, product or process, control charts and frequency distributions, are important inspection records. In addition, it is imperative that a manufacturing firm maintains suitable records of the quality performance of its suppliers. Integral part of supplier quality records, particularly valuable for corrective action is the evaluation of:

- How rapidly the supplier responds,
- How thorough and complete is the corrective action, and
- Whether that action is limited to particular lots or provides basic correction to prevent repetition of unwanted events or errors in the production of subsequent lots.

Whether they come from the production line or from examinations and tests, quality control information elements provide much of the objective quality evidence required to determine the acceptability of the quality of *our* company's products. In the general case, quality control records are considered adequate if they identify quality characteristics as well as inspection and test results. Such records can also be used to provide evidence that required inspection has been performed, when and by whom.

12.4 Six Sigma: A Quality Culture

Quality control's responsibility is in no way limited to the observance of engineering specifications, though most evidently this is very important. The responsibility is much wider and it includes the creation of a *quality control culture* in the organization, as well as a permanent environment for product sustenance with a steady stream of quality improvements.

A steady stream of quality improvements requires experimentation able to answer deeper quality questions such as: Why? How? How much? (Sect. 12.5). The lessons learned from study and analysis, however, may be fast forgotten as the responsibilities of people change or the quality control process drifts. Without appropriate organization (Sect. 12.2) even the most splendid improvements may be lost. Two issues are outstanding in this connection:

- The development and maintenance of any culture is not done just by word of mouth. It requires an appropriate structure, and

- In quality control terms the best available structural example is Six Sigma (6σ), the theme of this section.

It needs no explaining that σ stands for the standard deviation of the population; s is the standard deviation of the sample. Correspondingly, the notation of the mean is: μ for the population and $\bar{x}$ for the sample. μ and σ are parameters; $\bar{x}$ and s is statistics. In this methodology σ is used as a statistic, not as a parameter.

Originally the Six Sigma methodology was developed by Motorola, but it was made famous through its high-profile implementation at General Electric (GE). The best way to look at 6σ is as a structured quality control program that has been successfully applied in a number of companies—all the way from manufacturing to finance.

A keyword with Six Sigma is *collaboration* between all quality priests and factors of technology. As I never tire repeating the time when every designer, every industrial engineer, every maintenance specialist, and each enduser were living in an environment of their own without sharing experience, is past. Today, if they wish to prosper they will have to collaborate in regard to quality assurance and cost control. Failing to exploit the potential provided by collaboration means that we are not working for but against the best interests of *our* company.

What has been stated about the collaboration of internal departments is also true regarding *our* company and the market. Cornerstone to the implementation of Six Sigma methodology is the ability to understand and appreciate customer needs and expectations in quality terms. These are defined by the customers themselves when they set basic requirements and standards, as well as target values and tolerances (more on this later).

Well-governed producers do appreciate that defects are sources of customer irritation and they are costly to clients, not only to the manufacturers themselves and other service providers. Vigilance in regard to quality and costs should exist at all management and workmen levels, still it is the board's responsibility to instill an organizational culture whereby everybody looks at quality as being his or her personal responsibility.

With the preceding references in mind it is not difficult to see that the best way to look at GE's consistent effort in quality improvement through 6σ has been by appreciating the benefits of a disciplined process. That helped company management to focus on developing and delivering error free products and services, while at the same time it provided an excellent training ground of all employees.

Six Sigma's foremost goal is *total quality management* through high level statistical analysis aimed to drive out both defects and unnecessary costs. Low quality is too expensive. As the preceding sections underlined, the synergy between a quality assurance policy and a cost control program is critical to business success—but it is not made in a vacuum.

- It starts with decisions by the board on the necessity of both high quality and low cost, and
- It is followed by a definition in depth of the methodology, tools, and standards to be used to reach such goal.

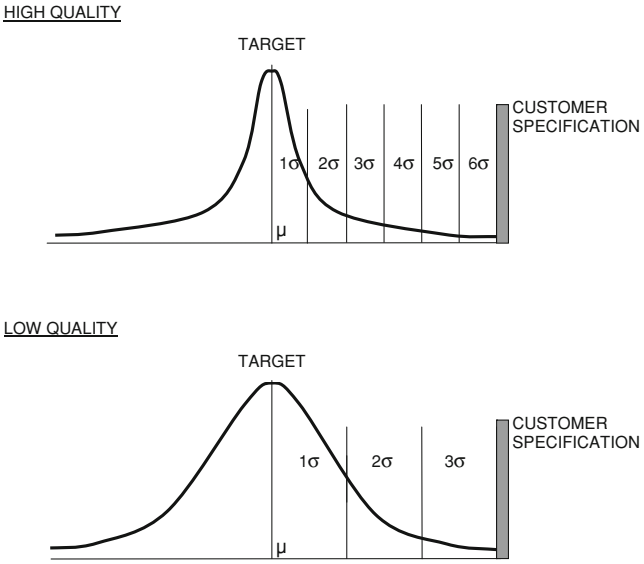

Fig. 12.4 The challenge is to fit six or more standard deviations between mean value (or target) and customer specifications

In so many occasions it has been said that a small standard deviation is evidence of quality. That is also a basic principle of Six Sigma. As Fig. 12.4 shows, the challenge is to fit six or more standard deviations between mean value (the *target*), which represents the manufacturer's tolerances, and customer specifications.

As a visual inspection of Fig. 12.4 confirms 6σ between company specs and customer specs is high quality; by contrast, 3σ is low quality. Six Sigma is not attained only by establishing firm guidelines on product assurance (Chap. 1), doing timely supervision and attracting new talent impacts on quality results—it also requires a high quality oriented company culture.

Speaking from personal experience, high quality is a catalyst to an attractive work environment, which adds to reputation. Motivated technical talent wants to work on high quality projects, because the work one is doing defines his or her self worth. That is high quality's psychological side.

As manufacturers prepare to introduce the next generation of products, and these days of rapid innovation products follow one-another in quick succession, everybody working for the company must not only adapt to changes in the methodology and technology, but also improve cost and quality performance. Firms which would not or could not do so, are not going to be around for long as independent entities.

Internal control can effectively reinforce the culture of quality assurance if and only if, top management's goal is one of dramatic improvements in quality, and if the CEO looks at it as being a crucial organizational performance parameter. Usually, firms find difficulties in adopting this stand because they have not taken

care in a comprehensive form of the quality of their produce. For such firms; Six Sigma is an excellent means for policy change, provided internal organization studies:

- Define which quality target is needed, and how it must be reached,
- Which information collected at design level, production floor and in field maintenance usable in deriving a quality pattern, and
- How data collection and analysis should be done to have a reasonable degree of confidence in the results of analysis.

In conclusion, product quality, cost containment, fast time to market, reputation and a challenging work environment lead to improved job performance and create a virtuous cycle which increases personal satisfaction. It is not secret that successful companies are careful to populate their work groups with people who are creative, decisive, and productive and care for quality and costs. By doing so, they gain an advantage over their competitors.

12.5 Using Six Sigma

An excellent example of practical results obtained through Six Sigma is GE's Medical System Performix CT X-ray tube when it was a new product, just introduced to the market. The stated goal has been 0% *dead on arrival* (DOA) which happens with several electronic products, including computers, as one of their components malfunction, and this does not permit the vendor to tell the customer the equipment is *ready for use*.

Reaching the objective of ready for use is vital both to the producer of the equipment (who gets his money earlier) and to the user (the clinic or hospital) as there are no patient rescheduling. Other goals targeted at GE through 6σ have been guaranteed tube availability, and an order of magnitude reduction in what the company calls *unquality cost*.

Every one of these benefits is food for thought to practically every firm in every industry. The sequential steps in reaching such goals are dramatized by the torrent of normal distributions in Fig. 12.5, from a practical Six Sigma implementation at general electric. Since nothing walks on a straight line, there will be a variance around the mean:

- At the top of the graph this variance is too wide resulting in 6.6% defects.
- Six Sigma has been instrumental in reducing the variance; but this did not happen overnight.
- With consistent effort the quality of production improved so much, that six standard deviations separated the mean from customer's specification.

Let me repeat the reference. As this general electric application has documented, it is not possible to go from 6.6% defects to 0.0% defects overnight. Such

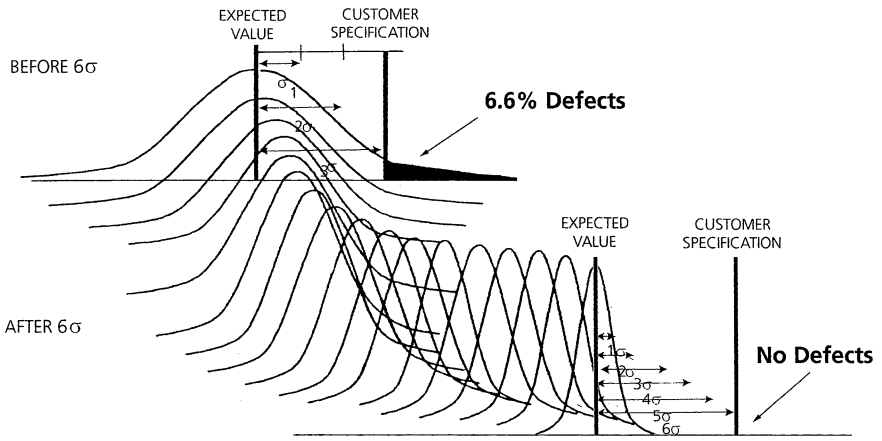

Fig. 12.5 The application of the Six Sigma (6σ) method led from low quality to high quality results

an improvement is doable over a period of time, *if* there is a will, a method and a well-trained team. This is precisely the reason why the 6σ methodology:

- Starts by identifying, qualifying, and quantifying all factors Critical To Quality (CTQ).
- Co involves the customers in pinpointing those aspects of product or service deemed critical from a quality perspective, and
- Establishes a process which permits steady quality improvements till the goal of no defects is reached.

These basic steps to Six Sigma deliverables are valid not only with proactive quality control at the factory floor but also for service quality reasons, at sales outlets or the backoffice. The Morgan Bank has implemented Six Sigma through 300 projects that, squeezed costs out of widely ranging process channels while improving the deliverables.

Companies which adopted Six Sigma have been active in developing and implementing a whole methodology around it. They also claim that its implementation has changed their DNA in everything they do and in every product or service they design and market. The basic tools underpinning GE's methodology, for example, include:

- *Statistical process control* methods analyzing data and helping to monitor process quality, capability, and performance.
- *Control charts* which monitor variance in a process over time, and alert to unexpected jumps in variance which may cause defects.
- *Process mapping* illustrating how things get done, by visualizing entire processes, their strength, and weaknesses.

- *A tree diagram* which graphically shows goals broken into levels of detailed actions, encouraging creative solutions.
- *A defect measurement method* accounting for number and frequency of defects that hit product or service quality.
- *Chi-square testing* evaluating the variance between two different samples, detecting statistical differences in case of variation.
- *Experimental design* permitting to carry out methodologically *t*, *z,* and Chi-square tests of the null hypothesis.
- *Root cause analysis* targeting original reason for noncompliance or nonconformance, aiming at their elimination.
- The *Dashboard mapping* progress towards customer satisfaction. Targets may be: percent defective, billing accuracy, and more.
- *A Pareto diagram* exhibiting relative frequency and/or size of events; searching for the 20% of the sources which cause 80% of problems.

At GE, the entire approach is guided by *Design for Six Sigma* (DFSS), a systematic methodology utilizing training tools and measurements to produce at high quality level, keep costs in check, and meet customer expectations. Associated to this is a process for continued improvement known as *Define, Measure, Analyze, Improve, and Control* (DMAIC).

DMAIC is a closed-loop process, characterized by its developers as systematic, scientific, and fact based. It helps to eliminate unproductive steps, focus on in-process measurements, and apply the best available technology for improvement. Through DMAIC, 6σ provide a vision of quality permitting people to strive for better results. The underlying *quantitative* approach is to:

- Multiply the number of units processed by the number of potential defects per unit, which gives the *opportunities for error*, and
- Then divide the number of actual defects by the number of these opportunities for error. Multiplying the result by 10^6 provides an impressive number which dramatizes the need for action.

In Six Sigma methodology, each combination of adjustments and quality correction measures becomes a system of equations that can be solved as a matrix. *Experimental design* (Chap. 11) allows users to efficiently test a significant number of variables, and hypotheses connected to them, in a dependable manner. As with other mathematical tools experimental design is at everyone's disposal, but the organization aiming to employ it must become so familiar with the method that it can take advantage of it in its daily work.

In conclusion, a methodology like Six Sigma must be exploited for maximum impact. In addition, the performance of a quality assurance process should be audited. Internal control should bring both matter-of-course quality reporting and technical auditing results to senior management's attention. This can be done in an able manner *if* there is the will to be ahead of the curve—which in turn requires increasing the product assurance sensitivity of the firm as a whole.

12.6 Evaluating Results from Observation and Experimentation

Since 1953, in my graduate studies at UCLA I have been exposed to a going debate on whether the statistician needs a thorough knowledge of the field in which he or she performs data analysis and/or an evaluation of experimental results. To my mind, whether he deals with experimental or observational data the statistician must either be an expert in the field in which he is working or collaborate very closely with an expert in that field.

For instance, with design engineers who work on quality but are not necessarily expert statisticians or experimental design professionals. Hence, GE's policy of close interdepartmental collaboration (Sect. 12.4).

Another issue to be brought to the reader's attention is the question of what sort of background in statistics assists in statistical inference. This is a matter of some controversy. Should not it be that the design engineer is also a statistician in order to enlarge the knowledge he acquires in the field in which he is working?

Opinions vary widely in regard to these queries. To my mind, even if a statistical investigation may have no purpose other than description, data analysis requires a statistical background. Not only is this needed for evaluation reasons, but also obtained results affect the way of collecting and providing data for subsequent causal analysis (Sect. 12.7). A complete statistical evaluation process involves four phases:

- Data collection satisfying specific criteria such pertinence to the problem being investigated and sample size (if based on sampling).
- Description including primary measurements, tabular and graphic presentation of information, and assessment of collective characteristics.
- Explanation which may involve specification of causal hypotheses and their testing against empirical data.
- Application, which might incorporate some element of prediction, or inferences, as in conclusions primarily drawn from observed data.

Extending facts in time and/or space through statistical inference is an important goal of experimentation for quality control and other reasons. We can put muscle to it by developing scenarios based on *what if* questions. The inverse is also true; rewards from inference underpin and expand our interest in experimentation.

Whether taken in a broad or narrow sense successful experimentation requires planning and control from the side of the experimenter. In a narrow sense, for example, it may be a *controlled experiment* which we specify by requiring conditions to be fulfilled, more, or less rigorously.

- Replications of the experiment are made under similar conditions, so as to yield an internal measure of uncontrolled variation.
- Such replications are mutually independent and uncontrolled variation in them is subjected to randomization (Chap. 11).

The causal hypothesis of an explanatory experiment frequently takes the form of a relationship in which the variable under investigation is expressed as a function of causal factors. Some of the latter are controlled in the sense of being subjected to systematic variation while others are uncontrolled. In this case randomization helps in making the controlled variables independent of the uncontrolled variables.

In contrast to an experimental input, the empirical data may be observational, as in the case of time series; or assumptions being adopted may be stochastic. In the case of industrial quality control these assumptions may be connected to an SQC chart which is under investigation on whether it is or it is not under *statistical control*. Shifts of emphasis can be examined, for instance, with regard to the:

- Phrasing of hypotheses,
- Estimation procedures,
- Hypothesis testing, and
- Tracking a prevailing trend in quality of production.

With observational data which consists of direct input on quality variables from the production line, the statistical analysis becomes dependent on good co-ordination with the subject under investigation. A tandem of samples and associated statistical graphics are needed to test against past evidence. The benefits being obtained from a consistent inspection are:

- A clear view of trend in quality of production, and
- The integrated picture provided by statistical methodology

While an explanatory approach can be carried out by the use of averages, frequency distributions or other elementary devices, this is not the way to go. Neither should the significance of results be judged just by common sense[1] without refined techniques. Statistical evidence requires tests which go beyond causal assumptions by involving experimental tools able to bring an uncontrolled variation under statistical control. Equally critical are the assessment of:

- Confidence intervals for parameter estimates,
- Significance levels in hypothesis testing, and.
- Models which aim to substantiate causal inference.

The test of hypothesis can be based on either or both experimental and observational data. What should be carefully considered *a priori* is the relative importance of the causal and stochastic assumptions, degree of complexity of the causal assumptions and systematic coordination of these two sets of assumptions. The technical control of causal factors enables the experimenter to breakdown a problem into subproblems, each explored by a separate series of experiments.

A critical element of the work described in these paragraphs, whether we talk of experimental or observational approaches, is the unambiguous specification of

[1] Which, as a saying has it, is widely available—and that is why each one of us has so little.

what we are targeting by the test of hypothesis. We can generally distinguish between two mutually supporting parts which can also work independently of one-another, but whether stand-alone or in unison form the basis of statistical data treatment.

- *Causal assumptions* forming the main explanatory approach.

Causal assumptions are typically obtained from prior experience. In the case of experimental situations, the design of the experiment must be made so as to test or demonstrate the underlying causal assumption.

- *Stochastic assumptions*, which are part of the content of the study inasmuch as they give an interpretation of deviations between observed and theoretical or expected values.

In a variety of cases, stochastic assumptions supplement the causal assumptions in providing the rationale of the statistical approaches for hypothesis testing. Statistical treatment of experimental data is typically done by the *t*-test for means[2] or by analysis of variance (Chap. 11).

In this case, the causal assumption is embodied in the parameter which indicates the varying mean for the different treatments of a quality factor. On the null hypothesis the parameter takes the same value for all treatments.

In conclusion, stochastic assumptions specify the frequency distributions of various treatment effects. Replications are assumed to be mutually independent and the distributions are taken to be normal with the same variance and with means given by the treatment parameters. As we have already seen, this is an approximation, but at the same time it is an important tool in quality assurance.

12.7 Cause, Effect, and Causal Inference

In his "*Essai-Philosophique sur les Probabilités*" Pierre–Simon de Laplace, the 19th mathematician, astronomer and philosopher, stated:"Present events are connected to preceding ones by a tie based upon the evident principle that a thing cannot occur without a cause that produces it." In a similar way, a great lot of future events are based on the after effect of present causes—which may be events or decisions.

Mathematicians look at causal inference as interplay between empirical and theoretical analysis which, sometimes, involves stochastic thinking. Stochastic assumptions play an essential part in establishing the efficiency of statistical techniques, accuracy of estimation, and of hypothesis testing.

[2] The *t*-test developer is Walter A. Shewhart, an American statistician, who in the 1930's published his algorithm under the pseudonym student.

Engineers consider causation as part of the more general concept of association. Many practical problems in causal inference can be solved with recourse to the theory of association. Techniques associated to the search of cause and effect is the result of thinking along the following lines:

- Every effect has one or more contributing causes.
- These usually known causes do not contribute equally to the effect. One or a few causes have a much higher impact than others.

According to Pareto's[3] principle, a very large part of the total effect almost invariably comes from relatively few causes. These few causes which quite often may be only *one* cause are not constant in their presence and impact. Frequently enough, the aim of analysis is to provide valid clues as to the nature (or concise area) of the important cause or causes.

Not every method employed in the analysis of cause and effect comes up with the results we would like to have. This obliges us to search for new tools and there should be no limit in developing more powerful tools and methods. Today, one of the better procedures permitting to test for causal inference is based on three steps:

- Examine cause-and-effect underpinning data sets, time series, observational, or experimental information.
- Project on likely aftermath, taking stock of hypotheses, and observable effect(s). The hypotheses are proxies for causes.
- "Forecast the past." for instance, by turning back a time series; choosing a tranche (e.g., 10 years); applying the pattern being developed or tested to that tranche; and comparing the emulated results with real-life statistics.

Plenty of evidence can be provided by examining how close a model's results approximate statistics and other data from the years following the selected tranche. The advantage of testing by "forecasting the past" is that it benefits from real-life data which permit to evaluate how well the model's output fits with events which have already taken place.

Among the main factors conditioning the efficiency of statistical techniques used in causal analysis are observational errors and sampling errors. The most frequent *observational error* is inaccuracy in measurement. Corrections for errors related to inaccuracy in measurements are made on the assumption that:

- Such errors are independent of the error-free variables.
- The error variances are known *a priori*.

In the general case observational errors are of larger order of magnitude than sampling errors. *Sampling errors* in statistical estimates are considered to be of the order of $\frac{1}{\sqrt{n}}$, where n is the size of the sample. Hence in large samples they tend to

[3] Vilfredo Pareto was a Swiss mathematician and economist, professor at the University of Lausanne in the late 19th.

be smaller. However, because plenty of work in statistical inference is made with relatively smaller samples, it is a mistake to ignore the sampling errors.

A type of sampling error affecting the results of causal inference is failure to take into consideration the lack of randomization in a causal interpretation of an interdependent variable. This result in inconsistency, bias, and inefficiency of estimates obtained through analysis.

For instance, because of lack of randomization the statistical interpretation may lead to an underestimate of the uncontrolled variation, resulting in an underestimate of the standard errors of the estimates. The need for correct randomization illustrates the close correspondence which exists between experimental design (Chap. 11) and model building.

At the origin of *observational data errors* are measurements and transcriptions. Observational errors tend to cluster, but they may also be characterized by diffusion patterns.

One of the tools which worth considering in trying to make sense out of test results or observational data is known as *meta-analysis*—a discipline whose impact has grown up over the past several years. Originally invented in 1948, it blossomed more recently as a way of:

- Extracting statistically meaningful information which can be used in modeling from lots of smaller trials.
- Doing so, even if the tests have been conducted in ways that makes it rather difficult to compare obtained results.

The downside of meta-analysis is that its output is valid only *if* all trials are included, not only those with positive results. *If* the negative trials are left out, *then* the output may be too optimistic—something that often happens with the interpretation of experimental data.

An interesting feature of interdependent modeling exercises is that they involve identities which are introduced by way of assumptions and of definitional relations. Such identities form a constraint for the effect variables. This implies that the relations under study are not autonomous in causal sense because there is an interdependent system.

Identities may as well represent instantaneous feedbacks. In this case, the downside lies in questions raised with regard to the operational significance of the model. Nevertheless, according to a dictum based on experience from mathematical modeling in banking and the financial industry: "All models are wrong, but some are useful." This usefulness comes from the fact that:

- Fulfilling the modeling prerequisites acts as an eye-opener.
- Different types of data can be treated on the basis of the same theoretical principles underpinning model development.
- The rationale of the statistical procedure can be examined by subjecting the empirical results to rigorous tests.
- Empirical results can be studied for agreement with theoretical relationships, and vice versa.

One of the issues working to the advantage of modeling is that it helps in searching for patterns. This is important in causal and statistical interpretation. A mathematically correct casual inference opens the way for applying statistical methods for controlling quality. Causal analysis is a powerful tool for quality-assurance-by objectives based on the laws of probability.

Chapter 13
Quality Control Charts by Variables

13.1 Variables and Attributes Defined

The most common type of varying number that arises in production or inspection is the *measurement*. Measurements are made on quantities as length of a hub, inside or outside diameter of a cylinder, electrical characteristics, such as capacitance, temperature, weight, tensile strength of a material, or the time taken by a certain operation.

Whether the measuring instrument is simple or complex, accurate or inaccurate, the measurement provides a single number descriptive of the characteristic being studied. In the terminology of statistics, when the quality of an item is indicated by a dimension it is said, to be expressed *by variables*.

Inspection is the process of measuring, examining, testing, gaging, or otherwise comparing a unit of product with applicable requirements.[1] In inspection by variables a specified quality characteristic on a given product is measured on a continuous scale. Such measurement is recorded for use in the decision to be made whether a lot, or process, is in control in product assurance terms.

Often it is either uneconomical or impossible actually to measure at great degree of precision a given characteristic of an item. All we can do is to determine whether or not the item conforms to more or less carefully defined specifications. For example, an item may have a crack or it may not, it may be of a certain color or it may not, a motor may run or it may not, the item may have been welded in all of the required places or it may not have been welded.

In other words, an item may "go" or it may "not go" according to rules set by inspection. When inspection, or testing, is done in a manner similar to those just mentioned, the testing or inspecting is said to be done *by attributes*. Both variables and attributes are tools of quality control.

Whether by variables or by attributes, statistical quality control charts present their user with a simple, comprehensive but dynamic picture of product assurance

[1] The unit of product is the entity inspected to determine its measurable quality characteristic. This may or may not be the same as the unit of purchase, production or shipment.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering,
DOI: 10.1007/978-1-4471-2966-0_13,

characterizing a given process. They make possible to state in a factual and documented manner whether this process is *in control*, is drifting out of control, or is already *out of control*. Statistical quality control operates through four main elements, three of which are charts developed in the 1930s by Walter A. Shewhart, an American statistician:

1. Control chart for measurable quality characteristics.

This includes charts for variables, typically for the mean $\bar{x}$, range R, and standard deviation s.

2. Control chart for fraction defective, also called p chart.
3. Control chart for number of defects per unit, or c chart, and
4. Sampling theory which deals with quality protection given by any specified sampling acceptance procedure (Chap. 9).

As the careful reader will recall, a vital subject in statistical quality inspection is the size of the sample. Chance fluctuations in percent defective will be greater with small lots than with large lots—but inspection of large lots is costly, and if pushed to extremes it can eliminate the competitive advantages of SQC. There are two pertinent questions:

- What is the optimum size of a sample, and
- How many samples should be drawn before a control chart is plotted?

In regard to the first query, over several decades of practice was to consider that a sample size of $n = 5$ is the best. More recently, quality control engineers believe that the optimum size of n lies 7 or 8, but $n = 5$ is still widely used because it facilitates the computation of the mean.

Regarding the second query, the plotting of SQC charts depends on continued sampling of materials or devices passing through the production line. The issue of control limits for these charts is discussed in Sect. 13.3, and the practical implementation of SQC charts for product assurance purposes is presented in Sect. 13.4.

Statistical quality control charts impose procedural requirements necessary to substantiate an SQC plan. For instance, a minimum of 25 samples should be drawn before a SQC chart is plotted, and this is equally valid of quality control by variables and by attributes.

A question often asked in my seminars is whether the better method is by variables or by attributes. The answer is that each and both of these methods have advantages. Although inspection by variables will almost always give more specific information regarding the control of a process, the method of attributes will be somewhat faster and easier.

The choice made over the years by industrial engineers is also an important reference. While the method by variables is more essential, at times it appears that most quality control applications strive to reduce their procedures to that of inspection by attributes. This may be attested by the number of "go-no-go" snap gages, or maximum-minimum depth and angle gages, used in quality control processes.

Table 13.1 Advantages of variables and attributes in statistical quality control

Variables	Attributes
1. Smaller samples	1. Less time, less skill
2. No problems in borderline cases	2. Less quality control equipment
3. Better guidance on quality	3. Less paperwork
4. Visual picture of quality, easier to interpret	4. Less arithmetic
5. More sensitive to changes	5. Easy to implement
6. Points faster toward corrective action	6. Easy to explain to the layman

If the quality can be controlled by the information contained about an item by means of attributes, it is preferable to use this method because of the all-important time element. Thus, more samples may be tested per unit length of time and this enables a better control of the process. The factor of final cost must also be taken into account before a decision on whether to use "attributes" or "variables" is made.

It is however; true that inspection by variables has some distinct advantages. Under some conditions variables inspection can give a better look at the product with less total inspection. Furthermore, a quality control chart by variables enables inspectors to judge practically in real-time whether a process is drifting out of control.

Stated in different terms, where exact measurements of critical inspection characteristics are made, we can get much more information with SQC by variables than would be possible with attribute plans. Yet variables inspection has one major disadvantage: the often extensive calculation required to determine product variability and estimated percentage outside of tolerance requirements. The relative merits of each method are outlined in Table 13.1.

Whether by variables or by attributes, the quality inspection of a process is made against *control limits*. Since no process works on a straight line we accept that it can fluctuate within limits. Such limits are quality characteristics a production process should meet, and they must not be confused with *tolerances*. Tolerances established at the drawing board, too, have to be met. To assure that the process is in control, its control limits must fall within the tolerance limits.[2] They are usually two of them:

- An upper control limit, and
- A lower control limit.

[2] Originally there have been three pairs of limits: (1) specifications connected to engineering, (2) so-called "tolerance", reflecting process capabilities through an algorithm, and (3) control pertaining to operations. This abundance of limits does not provide clarity in the production line with clear guidance.In this book, the terms *tolerance* and *specifications limits* are used interchangeably and they reflect values and other conditions set by engineering design. By contrast, *control limits* pertain to the production process, its machines, their fine-tuning, the skills of people and results from operation.

But as it has been discussed in connection to sampling, there exist as well cases with a single limit. A double limit is not an absolute requirement.

Integral part of a statistical quality control plan by variables or by attributes is a method of classifying defects. Such a method may call for enumeration of defects—of a stream of products through the manufacturing line, or of the unit of product—classified according to their importance.

- In the technical language of inspection, an article is defective if it fails to conform to specifications.
- A defect is a deviation tolerance of other requirements of specifications, drawings, purchaser descriptions, and more.

However, in many manufacturing plants it is common for some purchased lots to be rejected initially under the regular acceptance procedures and then finally to be accepted under some sort of material-review procedures. Whether this is or is not the case, it is always proper to remember that a 100% inspection does not offer better protection than a well set statistical quality control plan.

As the reader is already aware, even where the necessary inspections are not destructive *inspection fatigue* steps in to prevent 100% inspection from providing 100% insurance of conformance to specification requirements. Defects typically belong to one of the three classes, though other classes might also be necessary on certain occasions:

- *Catastrophic or critical defects.*

A critical defect is one that judgment and experience indicate could result in hazardous or unsafe conditions for parties using or maintaining such a product. For big end units of product, such as aircraft or ships, and for units with special mission (for instance in defense), a critical defect is one that could prevent performance of their mission. Medical equipment which is defective on delivery is known as "dead on arrival" (see also the discussion on Six Sigma in Chap. 12).

- *Major defects.*

A major defect, other than critical, is one that could result in failure, short MTBF, or materially reduce the usability of the unit of product for its intended purpose. There is no unique description of what may be a major defect and the best definition for it will consider both the product and the function it will perform in its intended implementation.

- *Minor defects.*

A minor defect is one that does not materially reduce the usability of the product for its intended purpose. For example, it may be a not so significant departure from established standards having very little or no bearing on the effective use or operation of the product assigned to its intended use.

For any practical purpose, the likelihood of defects whose probability is not far from one makes indispensable the definition of *acceptable quality level* (AQL), of which many references have been made in preceding chapters. AQL is a nominal

value. In quality control by attributes it is expressed in terms of percent defective specified for a single quality characteristic.

When a range of AQL values is specified, it must be treated as if it was equal to the value of AQL for which sampling plans are furnished and which is included within the AQL range. When the specified AQL is a particular value other than those for which sampling plans are provided to the vendor by the consumer (for instance a merchandizing company or the military), the particular AQL value to be used for a single quality characteristic of a given product must be clearly specified.

13.2 Sampling Inspection by Variables and Modes of Inspection

The considerations guiding the hand of the design engineer in establishing specification limits may be classified into three groups: (1) those related to the service needs of the unit (and its components) for which specifications are being written; (2) those related to the capabilities of the production process to produce given specification limits; and (3) those connected to the means to be used for determining whether the specifications are actually met by the product.

A sampling plan should be designed to facilitate the observance of tolerances. In general, the selection of a sampling plan must governed by the risk of accepting products with various percentage failing to conform to design specifications. This is fundamental even if the use of a sampling plan implies the willingness to take a chance on passing some products outside specification limits.

In SQC, this sampling plan and its method of implementation will reflect themselves in the control chart providing evidence about quality in the production process. They do so by visualizing in an objective, quantitative manner the choice inspectors should do about the production process; a choice embedded in the test of the null hypothesis (Chap. 8).

H_0: The production process is in an acceptable state of statistical control
H_1: The production process is not in an acceptable state of statistical control

For every sample drawn from the production process a point is plotted on the control chart. The abscissa of the point represents the time the sample was taken; the ordinate is the value of the statistic computed from the sample. Typically, what we plot is the mean $\bar{x}$, but it may also be the range R or standard deviation s. In a general classification, we distinguish three types of SQC by variables with the following criteria:

- *Known-sigma plans*, where the decision on acceptance or rejection of a lot is based on the sample average alone,
- *Unknown-sigma plans*, where the decision is based on the sample average, in combination with a measure of sample dispersion,
- *The lot plot*, where the decision depends in some way on the frequency distribution of the sample as proxy of the population.

What these three types have in common is that the criteria for acceptance or rejection of in-process devices, or of a series of lots, is based on the evidence of the control chart. Hence the need to define clearly unit of product, quality characteristic(s), quality standard(s), and production process under consideration. All of them impact on the type of control chart to be used for SQC of the production process.

If the production process is *in control*, more precisely in a state of statistical control, *then* as already mentioned 25 successive sample means $\overline{x_i}$ can be pooled together to yield a grand average value $\overline{\overline{x}}$. This is a good estimate of the expected value of the production process for the statistic in reference. This value is plotted on the control chart as the center line of the SQC chart by variables. However:

- Before the $\overline{x}$ and $\overline{R}$ may be plotted, a scale must be chosen for the chart, and
- The choice of a scale should make it easy to place all $\overline{x}$ and $\overline{R}$ values which are anticipated without having to redraw the SQC chart.

Notice that the sample averages $\overline{x_i}$ have a pattern of variation which is different from, but related to, the pattern of variation of individual measurements $\overline{x_j}$. A comparison related to the frequency distribution of sample averages with the frequency distribution of individual measurements will help to point out some of these differences.

In a two-tailed distribution the SQC chart needs upper control limit (UCL) and lower control limit (LCL); Sect. 13.3 explains how to calculate them. Theoretically, these are two values on the scale of the control chart situated $\pm 3s$ from the mean of the means $\overline{\overline{x}}$, so that approximately 99.7% of all possible values of the production process for the statistic under study would lie between them. These values of UCL and LCL are also known as three-sigma limits.

If the production is lousy, this will be attested from whether the tolerances fall inside or outside the UCL and LCL. If the production process is not in a state of statistical control, *then* the upper and lower tolerances will fall inside the UCL and LCL. Therefore prior to proceeding with SQC implementation it is necessary to tune the process reducing its variance. Otherwise, the SQC chart will indicate that the process is in control while the tolerances are not met. Figure 13.1 shows two quality control charts by variables:

- One for $\overline{\overline{x}}$, the mean of sample means
- The other for $\overline{R}$, the mean range

We need to plot both $\overline{\overline{x}}$ and $\overline{R}$ when, as the preceding discussion has explained, this is an unknown-sigma plan. To improve the reader's understanding, let me repeat step by step the mechanics of how this statistical quality control chart has been established:

- We start by defining what we wish to control: the process which will produce a given product of acceptable quality level.
- For charting purposes we select a random sample of n units, typically 5, from the production process.

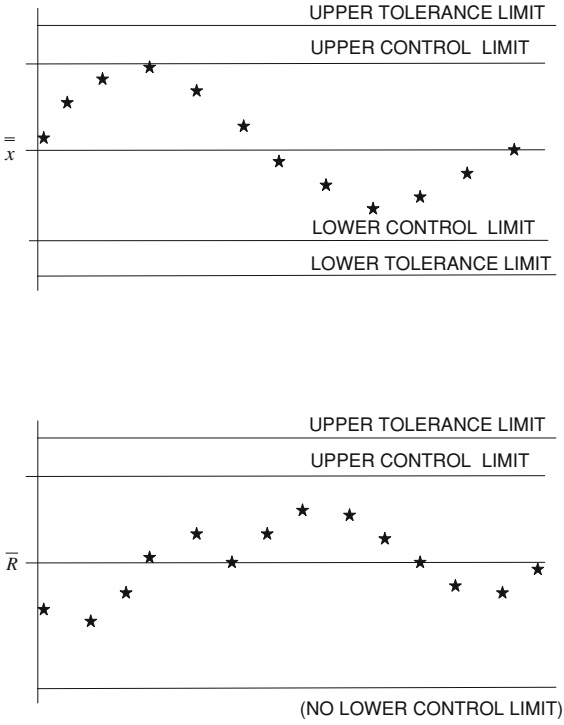

Fig. 13.1 Control charts by variables for mean of means $\bar{\bar{x}}$ and $\bar{R}$

- We inspect the sample by measuring or testing each of the n units and record such observations.
- We take 25 such samples to establish the center lines $\bar{\bar{x}}$ and $\bar{R}$.
- We check whether the tolerances fall outside UCL and LCL. If not, we tune the production process.
- We continue sampling and compute the values of statistics $\bar{x}$ and R, plotting point on the two control charts.

We do so on a permanent basis, computing $\bar{x}$ and R values from the production line, plotting them and observing whether they fall within the control limits have been established on the chart. The decision which we make on whether the production process is or is not *in control* is essentially a test of hypothesis.

In regards the actual state of the production process, *if* the plotted points $\bar{x}$ and R fall between the control limits, *then* we accept the statement that the process is producing the desired acceptable quality. By contrast, *if* the $\bar{x}$ and/or R points fall outside the control limits, *then* we reject the statement that the process is producing acceptable quality.

Processes have trends, and a statistical quality control chart reflects such trends. A rule of thumb is that if three successive $\overline{x_i}$ or R point in the same direction, there is a probability that the next $\overline{x_i}$ or R will fall outside the control limit (the upper or lower, whichever is in the direction of the three successive $\overline{x_i}$ in the trend).

Whatever has been discussed so far in connection to SQC charts is a straight process of cause and effect (Chap. 12), hence of statistical inference. A plotted point falling outside the control limits indicates that the production process must be investigated for assignable causes of the loss of a state of statistical control.

By combining the recorded measurements into a frequency distribution, the quality control engineer obtains a factual picture of the process pattern of variation as revealed by the samples. Taken in order of production, this pattern can be also profitably used for two other purposes:

- To make a comprehensive control chart on the production process, and
- To document whether or not the mode of inspection is satisfactory or it should be changed.

The second bullet is important in connection to the way statistical quality control is exercised. There are essentially three modes of quality control inspection: *Normal*, *tightened*, and *reduced*. Associated to them are sampling plans.

At start of an inspection aimed at statistical quality control we use what is typically called *normal inspection* which is more or less a reflection of practices prior to SQC—unless otherwise designated. During the longer term of inspection procedures, however, normal inspection should be used when inspection conditions are such that tightened or reduced inspection is not required.

- *Tightened inspection* must be instituted when the estimated process average computed from the preceding ten samples, or such other number of samples that has been specified, is greater than the AQL.

Another criterion in connection to tightened inspection is when more than a certain number M of lots under inspection have estimates of the percent defective exceeding the AQL. Different companies tend to apply their own rules. My personal rule is when the process gets out of control by breaking one of the control limits, as shown in Fig. 13.2.

Normal inspection is reinstated if the estimated process average of samples under tightened inspection is less than the AQL. Many companies add to this other criteria specific to their case.

- *Reduced inspection* may be instituted when a number of conditions established by engineering and quality control are fully satisfied.

For instance, a specified number of samples, as designated, have been under normal inspection and none has been rejected. Or, the estimated percent defective for each of preceding lots which were tested is a number well below the applicable lower limit of defects in AQL. For certain sampling plans there is as well the requirement that the estimated lot percent defective is equal to zero for a specified number of consecutive lots.

Still another, often encountered, condition for reduced inspection is that of a production at steady rate. Normal inspection must however be reinstated if any one of the following events occurs under reduced inspection: Production becomes

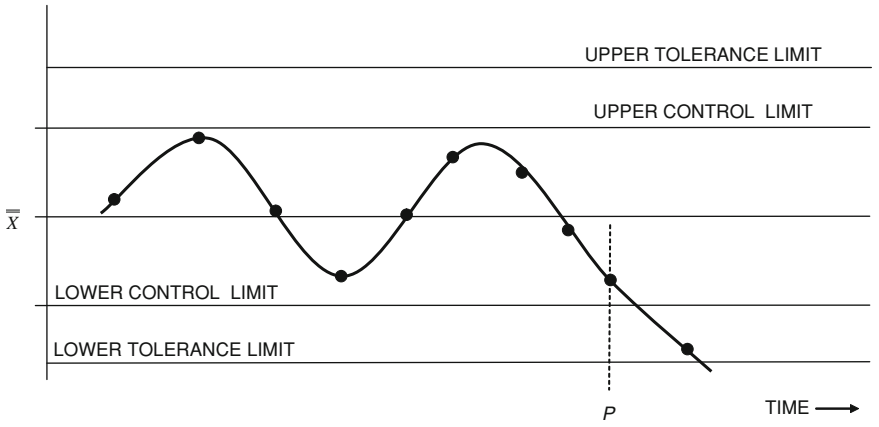

Fig. 13.2 A statistical quality control chart where the sample breaks the control limits

irregular or delayed. A lot is rejected; or the estimated process average is greater than the AQL. Such conditions warrant that normal inspection should be reinstated.

13.3 The Calculation of Control Limits

When we say that a given process is *in control*, what we essentially mean is that it is governed by a constant cause system. This is practically equal to stating that the process is at a satisfactory level with respect to mean and dispersion; therefore, it is sufficiently uniform.

The upper and lower control limits: UCL, LCL allows us to keep an eye on the variance. As long as they remain valid, they constitute standard values able to providing the boundaries of variability in current production. Indeed, the reason of taking periodic samples of the production process is to detect any significant changes in the process by means of $\overline{\overline{x}}$, $\overline{R}$, UCL and LCL.

- If the control chart indicates that the cause system has changed,
- Then it is up to the quality control specialist and process engineer to discover the cause of the change, and to correct the trouble.

Criteria must be established for effectively detecting lack of control, keeping in mind that one of the merits of an SQC the assistance it provides in controlling the ongoing production process. Such criteria must account for the fact that since almost any unusual pattern of variability may be a cause of suspicion, the subjective element involved in interpreting control charts cannot be entirely eliminated. Hence, the interest in defining specific patterns which are generally recognized as being warning signals. Examples of such signals are as follows:

- *A point outside its control limits*. Provided the population is normal, the probability of a mean going outside of a three-sigma control limit when the process is in control is minimal. If this happens, it is a warning.
- *Several points near together* are a reason to worry. This is especially true of several successive points that are close to a control limit, or beyond some secondary limit such as a two-sigma limit (used by some firms).
- *A run* of successive points *on one side of the central value*, or a cluster of a larger number of points, most of which are on one side of the central value. Also important is the trend characterizing these points.

Let us however always keep in mind that once computed, control limits are not forever; they have to be reevaluated. It is also good to remember the difference between the initial setting of control limits based on the analysis of historical data through which UCL and LCL are drawn, and their reevaluation while using statistical control charts during production.

1. The control limits UCL and LCL for an $\bar{x}$ control chart by variables represent the two values which 99.7% of the $\bar{x}$ values will fall if the process is in a state of statistical control, and

The formulas used to calculate the control limits when the process mean μ (which is the population mean of which $\bar{\bar{x}}$ is a proxy) and process standard deviation σ are known, are:

$$\text{UCL} = \mu + 3\sigma$$
$$\text{LCL} = \mu - 3\sigma$$

Basically, the process mean, μ, and process standard deviation, σ, are almost always unknown, but we do know $\bar{\bar{x}}$ and $\overline{R}$ where (as already stated) $\bar{\bar{x}}$ is the mean of sample averages, and $\overline{R}$ is the average of sample ranges. $\bar{\bar{x}}$ is an approximate but still best available estimate of μ based on the information we gather. The standard deviation of the population, $\sigma_x{}'$ is approximately equal to:

$$\sigma \cong \frac{\overline{R}}{d_2}$$

where d_2 is a factor found in Table 13.2 for sample sizes up to 10. If the process mean and process standard deviation are unknown, the following algorithms can be used to calculate the control limits:

$$\text{UCL}_x = \bar{\bar{x}} + 3\frac{\overline{R}}{d_2}$$

$$\text{LCL}_x = \bar{\bar{x}} - 3\frac{\overline{R}}{d_2}$$

It needs no explanation that it is important to examine the relation of control limits to engineering tolerance limits. In principle, tolerances must be compared

Table 13.2 For small samples the d_2 divisot provides a rapid calculation of s from R

Sample size n	d_2
2	1.128
3	1.693
4	2.059
5	2.326
6	2.534
7	2.704
8	2.847
9	2.970
10	3.078

with control limits for x_j and not with control limits for $\overline{x_i}$. In addition, both location and spread of the control limits in relation to engineering specification limits (tolerances) are important.

2. Based on engineering specifications we establish tolerance limits as the values within which the $\overline{x}$ values must fall to be acceptable.

Let us recall that by answering the question: Is the process in statistical control? The control chart tells us what is happening to the pattern of variation in the production process. For this reason when the process is in control, the control limits indicate:

- The boundaries of the pattern of variation.
- Whether we manufacture our product within engineering tolerances.
- Amount of improvement needed to make product within tolerances.

Borrowing a leaf from Six Sigma (Chap. 12) *if* the sample standard deviation is small, which can be rapidly tested by calculating the sample's R and using Table 13.2, we can assume that $\overline{x_i}$ is a proxy of x_j in the sample of, say, five units. Alternatively, if this hypothesis is rejected we better reexamine the process and recalibrate it, as well as our metrics. This is not always done in statistical quality control.

Let us take a practical example. In Table 13.3 are listed 25 observations on the first grind of edge widths x of piston rings. The observations were recorded in inches to the nearest 0.0001 inch. This data has been used in a statistical control chart.

From Table 13.2 we see that the divisor of $\frac{\overline{R}}{d_2} = 2.236$. The largest $\overline{R}$ in these five samples is equal to 0.0010; $\frac{\overline{R}}{d_2} = \frac{0.0010}{2.236} = 0.000443$; and $2.98 \times 0.000443 = 0.00133$. The $\overline{x}$ of this small sample of 5 is equal to 0.13966. It does not look unreasonable to take $\overline{x}$ as proxy of x_i.

Ad hoc tests have not been part of the SQL mainline. Nevertheless because every industry and every factory have not only general quality assurance problems but also special ones, ad hoc solutions have been developed which observe the general rules but also have some added value. Improvements should be welcome.

Table 13.3 Five samples of five observations, $\bar{x}$ and R

	0.1397	0.1399	0.1402	0.1394	0.1393
	0.1400	0.1398	0.1399	0.1397	0.1395
	0.1401	0.1393	0.1398	0.1397	0.1395
	0.1393	0.1393	0.1392	0.1400	0.1398
	0.1394	0.1396	0.1392	0.1395	0.1393
$\bar{x}$	0.1397	0.1396	0.13966	0.13948	0.3954
R	0.0007	0.0006	0.0010	0.0006	0.0005

They may for example indicate that quality control needs to be tightened and/or the technology of the production process ameliorated, in order to meet specifications because the SQC charts ate steadily in the borderline.

Alternatively, an experimental design (Chap. 11) might reveal that the production process is of a higher quality than necessary, or quality control procedures are unduly severe. For stochastic thinking to be a valuable means of assistance in process control, such studies should be done periodically. This is akin to auditing the statistical quality control plan already adopted.

There are as well other conditions to be accounted for; for instance, the case when a controlled process should meet not one but two specification limits on individual values, $x_{\max}$ and $x_{\min}$ In that case, all possible situations may be grouped into three general classes:

- The spread of the process $(6s)$[3] is less than the difference between the specification limits $(x_{\max} - x_{\min})$;
- The spread of the process is greater than the difference between specification limits.
- The spread of the process is approximately equal to the difference between the specification limits.

A merit of the control chart is that under a variety of conditions it helps in quality decisions because it tells when to leave a process alone as well as when to take action to correct the trouble. The elimination of the assignable causes of erratic fluctuation is described as bringing the process *in control* and is responsible for the many savings in cost resulting from statistical quality control.

To gain the most from stochastic inference, the reader should appreciate that the variable chosen for control charts must be one feasible to measure and express quantitatively. In order to be manipulated in an effective manner, the observation must be divided into what Shewhart has called *rational subgroups* for control charts.

One important consideration in sub grouping is *order of production*, since control charts are used to detect shifts in the process average that may come out due to a change in manufacturing procedures or settings. Whether by variables or by attributes (Chap. 14) when control charts indicate that a process if out of

[3] $6\ s$ should not be confused with $6\ \sigma$ in Chap. 7. In the present case it comes from $\bar{x} \pm 3s$.

control, it is vital to establish through appropriate and timely research what is necessary to permit the manufacturing process to make a product that meets established specifications for quality.

The careful implementer of SQC will also remembered, however, that estimates of means, variances, and percent defectives are subjected to sampling errors. Therefore any conclusions obtained from a short period of time, such as time is tentative. This is as well true of the 25 samples—the recommended minimum number of samples on hand before a control chart is plotted and analyzed. In my practice I stress the point that, contrary to tolerance limits, control limits must be regarded as:

- Tentative and
- subject to confirmation.

Experience teaches that change comes as the control chart is implemented over a period of time, therefore as more evidence becomes available. Moreover, the construction and implementation of quality control charts is just a step in the chain of manufacturing control. It is a means, not the ultimate objective (see also Chap. 15).

In conclusion, in establishing control lines the culture of stochastic thinking is a big "plus" because after control chart data has been collected it must be interpreted and corrective steps taken whenever necessary. Deliberately in this book no fixed rules are laid down regarding the appropriate action based on the interpretation of statistical qualities of control chart data. The person who takes decisions about the appropriate action must understand both the:

- Process under investigation.
- General principles underlying statistical control chart analysis.

While the calculation of control limits and their upkeeping, the themes of this section, require skill and experience there are also other issues to which must be placed attention. The good news is that all of them blend into manufacturing engineering and major strides can be made in product assurance if design engineering and manufacturing collaborate all the way from tolerances to statistical control limits and beyond.

13.4 Using SQC Charts by Variables

Section 13.2 presented the reasons why we use statistical quality control charts for in-process quality assurance. They help in keeping track of the tendency of every manufacturing process to drift outside engineering tolerances. As we have seen through practical examples, control is done in a visual manner and in a comprehensible way, so that corrective action can be taken whenever needed for it is indicated.

The analysis of control charts, like those in Figs. 13.1 and 13.2, is reduced to a problem of telling when a high or low point on the graph is an indication of the presence of nonrandom variations due to a change in the manufacturing process; or, it is due only to the random variation within that process.

Working with the problem of quality control in industry, Walter A. Shewhart was the first to realize that the solution to questions of evaluating a dynamic process like manufacturing lay in the field of engineering statistics. He made the hypothesis that it would be possible to set limits upon the natural variation of a given process so that:

- Fluctuations within these limits could be explained by chance causes, and
- Variations outside these limits would indicate deviations in the process which should be sought out and corrected.

As we have already seen, statistical quality control of a manufacturing process by the method of variables can be exercised through a control chart for means, $\bar{x}$, and a control chart for ranges (R). Or, alternatively, through a control chart for $\bar{x}$ and a control chart for standard deviations, s. The reason only one chart, for instance only $\bar{x}$ or only R, is not enough for sound SQC procedures should be found in the fact that one statistic cannot fully describe a statistical distribution:

- $\bar{x}$ is a measure of central tendency,
- R and s are measures of spread.

A process may not change in central tendency, yet it may spread outside the specification limits, as in Fig. 13.3a. Or, it may keep a constant range but shift in central value, as in Fig. 13.3b. A sampling inspection plan taking account of only the sample mean of the variable under study is often misleading, as the mean might well be within control limits while the individual observations might be located above or below these control limits.

To avoid errors of this type, R or s charts are plotted along with $\bar{x}$ charts for each sample. As it has been explained, if the variance is known the s chart is not needed, but the way to bet is that in most cases the variance is not known. These charts, $\bar{x}-R$ or $\bar{x}-s$, can give an accurate indication both of the central value and of the variability, of the variable under consideration, within the sample.

In addition, a chart showing directly the sample mean and range of values gives a good indication of the trends and stability within the manufacturing process. There are, however, some special cases where it is recommended to use the $\bar{x}-s$ instead of the $\bar{x}-R$ charts.

For example, when sample sizes are fairly large or when each measurement is comparatively expensive, and therefore we wish to extract all possible information from the measurement, the $\bar{x}-s$ chart is desirable. Its advantage is that it gives an estimate of the theoretical standard deviation of the variable which is subject to a smaller average error than the estimate obtained from the range.

The reader should also notice that in some plants it is considered preferable to plot on the control chart the sum of n measurements in each subgroup rather than

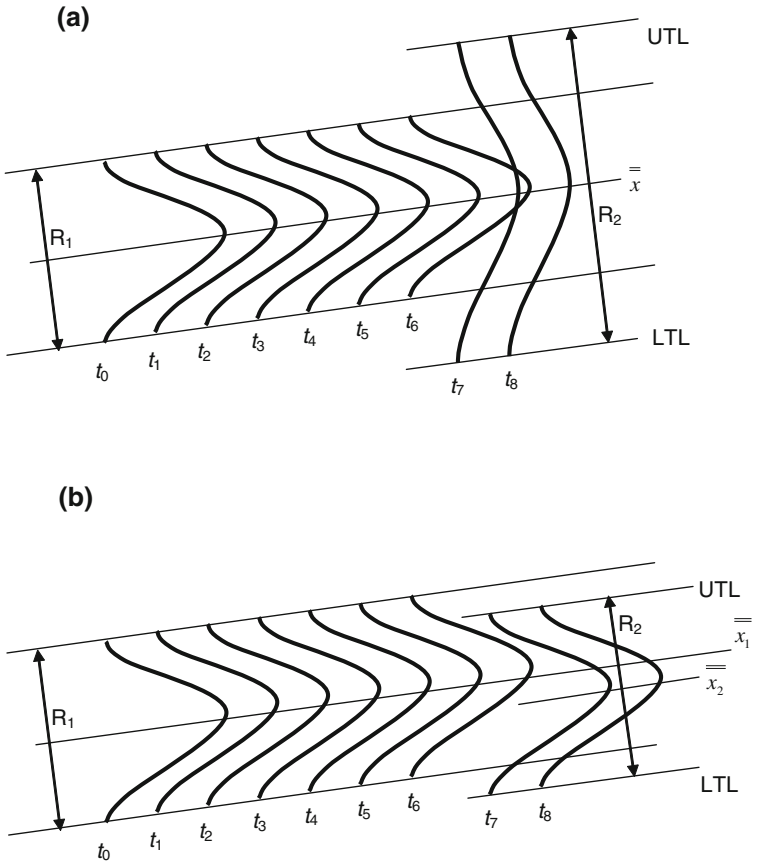

Fig. 13.3 One statistic is not enough to accurately describe the behavior of a process (Upper tolerance limit, lower tolerance limit of the widening and shifting distributions)

the average of these measurements. This procedure might simplify the necessary computations, but sometimes it may result in a misleading interpretation of the charts.

There is some truth in this argument. By using control limits for standard deviations rather than ranges we gain somewhat in efficiency. What this discussion suggests us that there is no lack of opinions on the best way to proceed with SQC, as far as practical applications are concerned.

Where the different opinions expressed by implementers converge is that for valid estimates of quality assurance parameters, the data on which SQC will be based must come from a process which is *in control*. *If* control limits are computed from estimates made from a process that is out of control, *then*, this data will most likely bear little relationship to what is needed for corrective action. Neither does this data

represent what can be expected in the future when the process is in control. When they are based on unreliable data, statistical quality control charts may result in:

- False alarms, and
- Failures to activate corrective procedures when trouble is developing.

I had cases where manufacturing people asked for changes in engineering tolerances when they were not able to meet them. Their argument was that the company's manufacturing process, and its machines, are not at the level assumed by the engineers. According to their opinion, this was the reason for SQC alarm.

To my book, such arguments are defensive and have little value, if any. Engineering specifications are "the master" and should be attained by production unless it can be a thorough study documents that the specs are really not attainable because of a retrograde manufacturing establishment.

People with practical experience from applications of statistical quality control charts come up with several suggestions in improvements, some of which worth recording. One of them regards criteria for judging the control chart on the basis of evidence of a high degree of control when no more than *1* out of *35* points (or no more than 2 out of 100) fall outside the upper and lower control limits.

One should try to capitalize on this observation, though it may not apply to all cases. A common problem with different suggestions, particularly those involving a sort of general criteria, is that they do not have the mathematical background which can be properly tested. Because of this, there is a risk of rejecting the hypothesis of a controlled process while the process is in control. Similarly, tests determined to suit the needs of a particular type of application do not answer necessarily more general requirements.

On the other hand, it may be helpful to plot in time the variation in UCL and LCL as documented by the output of the process under quality control (Fig. 13.4). This helps in identifying possibilities for quality improvements, or cost reduction, after the reasons for such variation in control limits have been established.

A rule of thumb is that the calculation of a 5-day moving average is simplified by carrying a *moving total* to which is added each day the algebraic difference between the value of today and the value of 5 days ago. I have tried it but I am not convinced of its value in statistical inference.

Still another suggestion coming out of practical applications of SQC, is that a control chart's role can be enhanced by evaluating past patterns of process behavior and deriving from them certain specific rules. For instance, that a strange pattern of the $\bar{x}$ and R points, even though within (but close) to a control limit, may be an indication that the process is getting out of control.

"A production is not probable because we think so… (But also) there is little likelihood of our discovering a method of recognizing particular probabilities, without any assistance whatever from intuition or direct judgment," said John Maynard Keynes, the British economist. The same is true of technical developments adjunct to statistical quality control theory, based on practical experience. They should be given a chance while being carefully watched. Many of them have to do with variance.

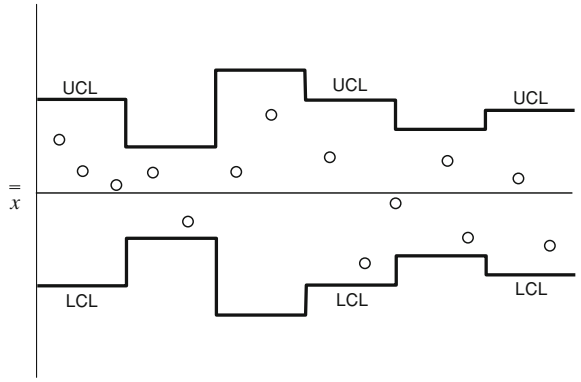

Fig. 13.4 Variation of upper and lower control limits on a $\bar{\bar{x}}$ SQC

Provided that the hypothesis of population homogeneity is upheld, whether the background process which we aim to control obeys or does not obey the normal distribution, the variation in the measurement of its deliverables is very important. Our best means of control are based on the observed pattern of such variation which we hope to keep under control. The fact that change is visualized through charting helps in building quality into manufacturing processes. As a reminder:

- Control charts reflect the rule that variation will follow a stable pattern as long as the system of chance causes remains the same, and
- Once a stable system of chance causes is established, the quality control limits for the resulting pattern of variation can be determined.

When these conditions prevail, future data streams can be expected to fall within the same limits as past data, unless there is a change in the system of causes. This change will by all likelihood have to do with the dynamics of the production process, but there may also be a case of change in the specs. Tolerances are established not only through the engineering department's initiative but also by specific client requests. A new order by a client (for instance, the military) may require new more stringent tolerances which the production facility *as is* cannot meet.

Outside of these conditions, when a process has been stabilized and then suddenly data fall outside the upper and lower control limits, an investigation must be made to determine the cause of the change in pattern. *If* the job stream and/or the manufacturing environment changed, *then* both the central tendency of the chart and the limits must be adjusted.

In conclusion, statistical quality control techniques are easy to adapt to developing conditions because they have a rich inventory of tools from which to choose. Different types of control charts serve different functions. A reference to what this flexibility means for the end user is that while quality control limits are established from *past and present performance* they serve as a guide to *future performance*.

13.5 The Statistical Quality Control Chart Is a Model

A statistical quality control chart is the model of a real life situation; namely, the quality assurance provided by current production. Whether by variables or by attributes, it lays down the relationships which exist between a given product, its specifications, available production capabilities, skills, and resulting quality level.

Models are working analogies models emulating prevailing conditions. One of the simplest forms of a model is the scale drawing; a common design tool for graphically analyzing, for instance, a piece of equipment or the motion of what may be a complex linkage.

As long as it is accurate (not necessarily also precise), a good model represents motion of the linkage and serves as a helper that can be put to work on a practical problem. The SQC does so with the control of quality. Most importantly, a model is an instrument for thinking. *Thinking* is, so to speak, movement without movement. We contemplate then simulate a move,

- Trying to guess what it involves, and
- If possible, project the aftereffect of the movement.

This may sound simple, but it is a complex process. Millions of years of evolution have been devoted to make the process of thinking, as we know it, a reality. The diagram which controls the brain functions and whose exact pattern is still unknown (including matters of behavior) is still a black box. A great deal of effort is today expanded in reverse engineering this process, along the hypothesis that the nature of learning and of all sorts of human activity, including aggression, is the key to understanding behavior.

A model like SQC has no claims of establishing the universal truths. Its domain is limited to quality yet it requires thinking—even if, some people believe that it is safer not to think at all (while others are afraid of what they might think).

The problem with this attitude is that a thought is like a child inside a woman's body. It has to be born. If it dies inside us, we die too. Therefore, *thinking* must be part of everybody's education but this is far from being generally true. In his Cambridge lecture on August 31, 1837, Ralph Emerson said:

> *If* one is a true scientist, *then* he is one who THINKS.[4]

Since 1953, when I learned to apply SQC charts, I look at them as an opportunity for thinking. Niels Bohr, the nuclear physicist, was teasing his peers and his students by telling them: "You are not thinking, you are only being logical." Great thinkers in history have always appreciated that thinking means:

- Challenging the "obvious",
- Hence, doubting and experimenting.

[4] Emphasis added.

Because it is based on thinking, experimentation is the mother of all sciences. Thinking should be guiding young and aspiring students when deciding about their first big investment: higher education. An ambitious young person about to take the first steps toward his or her professional life should reasonably examine, indeed speculate, about what sort of qualities and skills the future employer might be looking for. Employers have goals in their mind:

- When they hire the new generation of employees, and
- When they promote people to senior positions, up the organization ladder.

It is quite fortunate that models made through statistical quality control procedures prod to thinking, because science without thinking is no science at all. It is therefore all for the better that a statistical quality control charts helps the people who implement it and use it in thinking. As long as it does so, it is a scientific tool. Its continuous presence and dynamic behavior offer distinct advantages given that:

- Researchers are as good as their last experiment, and
- If they do not continue exercising their mind scientists become incompetent.

Not everybody appreciates the creative role of scientific thinking. In her book "How the Laws of Physics Lie", published at Oxford in 1983, Nancy Cartwright advances the thesis that science does not describe a profound physical reality. It only advances phenomenal models, valid only in a limited space or conditions—which, therefore, are fictitious. But at the same time they can be creative.

Engineers and physicists are not merchants of dry facts or purveyors of fabricated conclusions. Neither are they, by profession, prone to see through the prisms of a narrow discipline. The best professionals are those with the more investigative mind and a great deal of critical spirit.

Both the scientist and the artist attempt to free themselves from the beaten path. They do so by stepping outside convention to embark on a voyage uniquely their own—and their creativity leaves its mark on all of us. Facing a blank canvas, a great painter cannot fully anticipate the completed work.

- Which turns his imagination will take?
- What colors will compose the painting?
- What technique will produce the most thought-provoking impression?

Similarly, "a scientist beginning a research project may have an idea, a hunch. But where will the adventure lead? To what discoveries?" [1] There is no way of knowing until the work develops to the point that it gets an interpretation momentum. In manufacturing engineering the concept of such a momentum cannot be better explained than through the SQC charts.

The role of SQC charts is as well the role of a model: that role has been that of creating new knowledge, not just interpreting information by adding to it an ever greater but uncertain detail. Science involves much more than collecting facts and data and then trying to interpret them.

- It begins where what one knows is too limited, obscure, or incomplete, and
- By trying to bring light into the darkness, it takes off to new frontiers.

This does not always succeed. There are many reasons why we may not reach our goal in scientific study and discovery. The problem might have been much bigger, or its variables more interdependent than we thought. The method we used was too limited. We failed to use the best of our abilities in investigative thinking.

Or, we might have been confronted with *model risk*. Not every model is the right one for the work we are doing, let alone being a construct well adapted and complete. Moreover, what many people forget is that models must be tested, and the test of a model should be done not at one but at five levels:

1. Test of hypothesis, and associated assumptions
2. Test of whether the problem approximates the normal distribution, or there are major outliers
3. Stress test of the model. Can it stand up to extreme events? At 5, 10, 15, 20, 25 standard deviations?
4. Implementation test, based on its deliverables and on their acceptance by the end user(s)
5. Post-mortems and walk-through, including auditing by an independent expert, to ascertain the quality of the model's mechanisms and results

This is part of *meta-analysis*, a statistical technique for rigorous testing, also used for extracting information from small trials that are not by themselves statistically reliable. The latter is a good way of looking at an SQC chart and the message it conveys to its user, given that each $\overline{x_i}$ is based on a small sample of 5–8 items.

In conclusion, when we look at the SQC chart as a model, *patterning* is the keyword. Successful prosecutors know that a case is made much stronger if a pattern of misdeeds can be established. It is easier to cast doubt over evidence related to one isolated incident than to a whole pattern of incidents which have been proven. It is as well possible to overlook an isolated case as an outlier, but a whole pattern cannot be cast aside.

Reference

1. Weizmann Institute of Science (2003), Annual report 2003

Chapter 14
Quality Control Charts by Attributes

14.1 Fraction Defective

"I learned that quality requires minute attention to every detail," says David Packard, "that everyone in the organization must want to do the job, that written instructions are seldom adequate, and that personal involvement is essential [1]". Precisely for this reason, Packard instituted in his company *management by walking around* (MBWA).

When the manager is walking around the plant (or the office) to watch over quality of deliverables, he does not take it as a social duty or as a promenade. He wants to see evidence, and this evidence can be provided by variables or by attributes. Control charts by either method give a graphic presentation of the quality of the process. The average percent defective (Sect. 14.3) is an example.

The sense of quality control by attributes has been explained in, Sect. 13.1. The *go/no-go* approach is applicable both with lot inspection of purchased goods and of original inspection[1]—that is, the first inspection of units coming out of a production process. Two metrics are the most important with quality control by attributes:

- Fraction defective p
- Number of defects per unit, c

With fraction defectives we are interested in the process mean which must be estimated from the results of inspection of samples drawn from the production line or from a specified number of lots undergoing inspection. Computation-wise this statistic is the arithmetic mean of the estimated lot percent defectives computed from the inspection results of the preceding ten lots but it may as well be otherwise designated.

[1] The term *original inspection* is also used in connection to quality control submitted for acceptability as distinguished from inspection of a product which has been resubmitted after prior inspection.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering,
DOI: 10.1007/978-1-4471-2966-0_14, © Springer-Verlag London 2013

In line with this definition, the percent defective for a quality characteristic of a given lot of products is the number of units of products defective, for a precisely defined product characteristic, divided by the total number of units of products and multiplied by one hundred. The algorithm is very simple:

$$P = \text{percent defective} = \frac{\text{Number of defectives} \times 100}{\text{Number of units}}$$

In assembly operations and in procurement a unit of product is an assembled unit of product. Say, however, as an example that these products are cylinders and our mission is quality control through the method of attributes. The precisely defined product characteristic is their internal diameter.

By use of a go/no-go gage we develop a list of defects as each unit of product is classified as a defective or a non-defective. This can be applied to a sample taken from a lot or to a sample of units coming out of a production line. The algorithm for the fraction defective parameter p is[2]:

$$p = \frac{\text{Number of defective units in production process}}{\text{Total number of units in production process}}$$

The production process will be in an acceptable quality status if p is (1) small and (2) constant during production operation. We will see how this works in Sect. 14.2 on sampling plans for control for attributes, and in Sect. 14.3 where we use the sample fraction defective control chart to accept or reject a given quality status of the production process.

Different happenings could bias the results of a statistical quality control (SQC) plan by attributes. For instance, lack of uniformity in the population of cylinders in the lot whose conformity to specifications is tested by means of go/no-go based on samples. This might have been due to variations in the quality of raw materials used for the cylinders or other reasons.

Change in the internal dimension of the cylinder may as well result from tool wear or a change in the setting of the tool. But they may also be the result of one or more faulty gages. The latter problem is relevant with SQC by attributes, though it is not often present with SQC by variables. Biases with measuring instruments are not at all welcome because what we are after is a trend in points; particularly:

- Points beyond one of the control limits,
- Points clustering close to one of the limits, or
- A succession of points most of which are on one side of the chart.

A way to avoid gage bias is to frequently calibrate the gages. Another way is to use a different gage for additional measurement to be taken immediately in order to confirm or deny the evidence of defects.

[2] The group distribution of sample fraction defective (p) values approximates a binomial distribution.

The acceptance of a lot submitted for inspection is determined by use of one of the sampling plans associated with a specified value of acceptable quality level (AQL). The sampling plan may be based on known or unknown variability depending on the population's characteristics. If it is unknown, there exist two alternative methods for quality control. One of them is based on the estimate of lot standard deviation and the other on the mean range of the sample. These are respectively known as:

- Standard deviation method and
- Range method.

With quality control by attributes the AQL is a nominal value expressed in terms of percent defective or defects per hundred units, whichever is applicable. If the AQL is specified as a range it is usually treated as if it were equal to that value of quality level, for which sampling plans are furnished and which is included within the range.

The classification of defects associated to SQC by attributes is similar to the one we have seen in Chap. 8 in connection to SQC by variables. *Catastrophic* (or critical) defects, are those that judgment and/or experience indicate could result in hazardous or unsafe conditions.

Major defects, other than critical, could result in failure or materially reduce the usability of the unit of product for its intended use. By contrast, *minor* defects do not materially reduce the usability of the unit but are departures from design standards which have to be corrected even if they have no significant bearing on the effective use of the product.

Note that there is permeability between these three defect classes. Minor defects can graduate into major, and major into catastrophic—if they are not promptly corrected. All of the above categories represent deviation from specifications and standards. They are events falling outside tolerances. To flash them out, we should sample our production process and test for defects, identifying them, classifying them, plotting them in quality control chart(s) and taking corrective action.

Corrective action is a "must" because we want to be able to offer our customers the best quality at an affordable cost. For any company, big or small, the job which it does is to satisfy the requirements of its customers. In today's globalized industrial world the contribution a company can make—and its way to compete—is not just innovation but also quality, ease of use, and affordability.

Chapter 13 made the reference that up to a point QC by variables and by attributes are alternatives. There are however processes where SQC by attributes is the only possible solution. An example is office work.

There exist as well cases where SQC by variables and SQC by attributes are equally applicable alternatives. In that case the decision factor becomes affordability of the final product while protecting quality. Like any other activity quality control costs money, the readers should however notice that while SQC costs $<100\%$ inspection and gives better results—this does not mean that it does not have a price tag associated to it.

There is always a price tag for everything we do. The cost of applying statistical tests varies with the operation under control, its requirements, and the skills available to fulfill these requirements. At the same time, the consequences of being unaware of or failing to correct quality defects in the product line can be a high multiple of SQC costs.

In conclusion, in cases quality control by variables and by attributes are alternatives, the bill to be paid for quality assurance varies with the nature of the product, with the frequency and severity of deviation(s) in manufacturing and with the method used to track (then correct) trouble. Percent defective allows inspectors to be in charge of quality and (quite often) of costs. Such a double goal requires a plan, constant vigilance, and self-discipline.

14.2 Sampling Plan for Control by Attributes

A methodology for quality control by attributes means three things at the same time: method of inspection, way of sampling, and sample size. What has been discussed about sampling at large in Chap. 9 and about samples in connection to SQC by variables remains valid with attributes but there exist as well some additional remarks to bring to the reader's attention.

With quality control by attributes we can use a simple, double, or multiple sampling plan to determine whether the lot shall be accepted with respect to a particular AQL value. With *single sampling* plans, if the number of defectives found in the sample is equal to or less than the acceptance number in the sampling plan, the lot from which the sample was drawn is accepted. If the number of defectives is greater than the acceptance number the lot is rejected.

With *double sampling* plans, when the number of defectives found in the first sample is equal to or less than the first acceptance number of that sampling plan, the lot from which the samples were drawn is accepted. When the number of defectives found in the first sample is equal to or greater than the first rejection number, the lot is rejected.

By contrast, when the number of defectives found in the first sample is between the first acceptance and rejection numbers, a second sample of the size indicated in the sampling plan shall be examined. This being done, the number of defectives found in the first and second samples are accumulated.

- *If* the cumulative number of defectives is equal to or less than the second acceptance number of the sampling plan, the lot is accepted.
- *If* it is equal to or greater than the second rejection number, the lot is rejected.

With *multiple sampling* plan the procedure is similar to that of the double sampling plan, except that the number of successive samples required to reach a decision is more than two. The similitude to sampling plan for SQC by variables is also present in regard to sampling plans for *normal*, *tightened*, and *reduced* inspection, as well as with respect to the criteria determining them. As a reminder:

- *Normal* inspection is employed when the estimated process average is not outside the applicable upper and lower limits of the SQC plan by attributes (Sects. 14.3 and 14.4).
- *Tightened* inspection is instituted when the estimated process average exceeds the applicable upper or lower limit of the chart.
- *Normal* inspection is reinstated if the estimated process average is equal to or less than the AQL while tightened inspection is in effect, and
- *Reduced* inspection may be adopted if a number of quality assurance conditions are satisfied. This "number" is typically defined by the company for its products and processes under SQC.

In actual practice, several companies defined their conditions for reduced inspection in the following way: production is at a steady rate; the preceding 10 lots have been under normal inspection and none have been rejected; and the estimated process average is less than the applicable lower limit.

Correspondingly, the policy for reinstating normal inspection followed a reduced inspection usually rests on the occurrence of the following conditions: production becomes irregular or delayed; a lot is rejected; the estimated process average is greater than AQL, or for some other reason connected to the controlled process management deems that normal inspection (or even tightened) should be reinstated.

The concept of tightened inspection as an alternative to normal inspection is essential in statistical sampling procedures, whereas economies can be realized by permitting reduced inspection when the quality history is first class. This has, however, prerequisites—both cultural and numeric. A systematic record of *quality history* is an important aspect in statistical acceptance procedures, it becomes an absolute "must" if and when reduced inspection is contemplated.

In a way similar to that discussed in Chap. 13 about SQC by the method of variables, a QC plan by attributes requires the computation of process mean, even if in this case it is the limits rather than the central tendency that are the more important. The process average is:

- The mean percent defective,[3] or
- Average number of defects-per-hundred units of product submitted for original inspection.

Estimating the process mean is typically done by an arithmetic mean computed from the results of sampling inspection of the preceding 10 lots or other quantity of production. This poses no mathematical problem. The problem is procedural, specifically one of carefully excluding from the estimated process mean, the results of inspection of product manufactured under conditions deemed as "not typical" of usual production. With this precaution in mind, a systematic plan for single sampling requires that three numbers be specified.

[3] Usually known as "average percent defective", but "average" is a term which should be avoided.

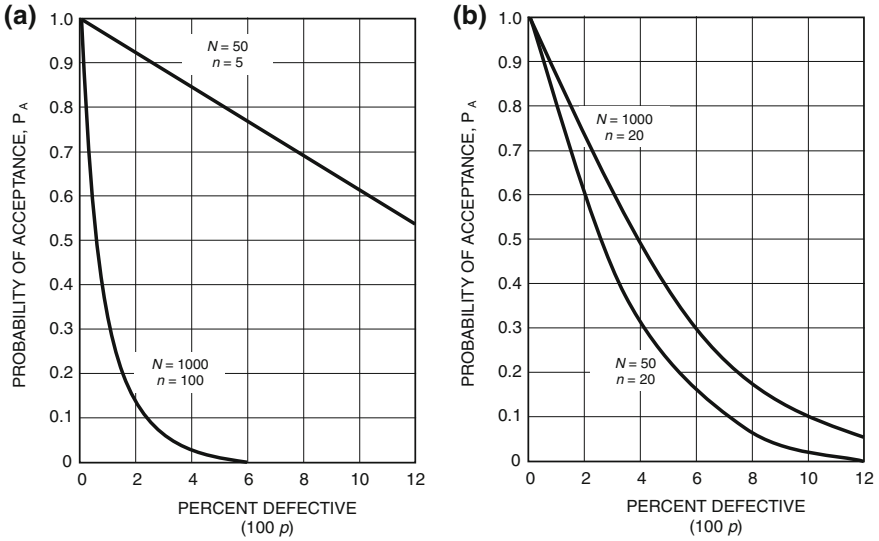

Fig. 14.1 The power of a sampling plan lies in the absolute size of the sample

- N, the number of articles in the population from which the sample will be drawn,
- n, the number of articles in the random sample drawn from the lot, and
- Acceptance number of defectives to be used in decisions regarding quality assurance.

Another issue calling for attention is that sampling acceptance plans with the *same percent* samples can give very different quality protection. As underlined in Chap. 10 in regard to OC curves, it is the absolute size of the sample rather than its relative size that determines the level of quality assurance by an acceptance sampling plan.

This is shown in Fig. 14.1 by means of operating characteristics curves. In Fig. 14.1a the same size n is "same percent" whether the population N is 50 or 100. With $n = 5$ the OC is a chimera, but there is a nice looking (discriminating) OC curve with $n = 100$.

By contrast, in Fig. 14.1b the sample size has been kept constant, equal to 20, whether the population is equal to 50 or 1,000. The OC curve of $N = 50$ is steeper than the one of $N = 1{,}000$, but the latter is still acceptable. Sampling procedures have a price tag and the lesson is that an economic evaluation of the different alternative plans must be done before deciding which one will be selected for a given situation.

Designers, experimenters, manufacturing engineers, and quality controllers who do not pay full attention to the shape of the operating characteristics curve will live to regret it. As it cannot be repeated too often, the operating characteristics curve of an acceptance sampling plan shows the ability of that plan to distinguish between:

- Good and
- Bad lots.

For any given fraction defective p in a submitted lot, the OC curve shows the probability P that such a lot will be accepted by the given sampling plan. In quality control engineering three points of the OC curve have been given particular importance:

- The lot (or process) quality for which $P_A = 0.95$ (or $p_{0.95}$)
- The lot quality for which $P = 0.50$, or $p_{0.50}$
- The lot quality for which $P = 0.10$, or $p_{0.10}$

The corresponding 100 $p_{0.10}$ is called *the lot tolerance percent defective* (LTPD). It is a crucial quantity to be respected by everybody—from design, to manufacturing and field maintenance. For any plan it is also feasible to compute the maximum possible value of the average percent defective in the outgoing product. This maximum figure is referred to as the *average outgoing quality limit* (AOQL). This, too, is a key factor in product assurance.

People with experience in both engineering and business appreciate that the choice among various possible plans for acceptance of manufactured products is essentially dual: technical and economic. An important element in the selection of an acceptance inspection plan should be the probable contribution of the plan to quality improvement, which also has a dual aftermath:

- Technical and
- Economic.

A major assistance in SQC planning is the Dodge-Romig Tables. They can be distinguished in the following four sets: (a) single sampling lot tolerance tables, (b) double sampling lot tolerances tables, (c) single sampling AOQL tables, (d) double sampling AOQL tables. The double sampling AOQL tables have proved the most useful. In many cases, the saving in inspection due to double sampling usually exceeds 10%.

Another useful set of sampling plans is offered by the statistical research group of Columbia University. Broad areas in SQC decisions have been defined by military standards, like the American MIL-STD-105A developed during WWII but still able to provide guidance and useful tables in establishing and managing a sampling plan.

14.3 Plotting Percent Defective in SQC

In SQC terms, percent defective is managed through p charts. In many industrial applications percent defective maps provide a good measure of quality performance, as they measure manufacturing process quality. Applied in an office environment, the p measures the accuracy of the administrative or other type of work.

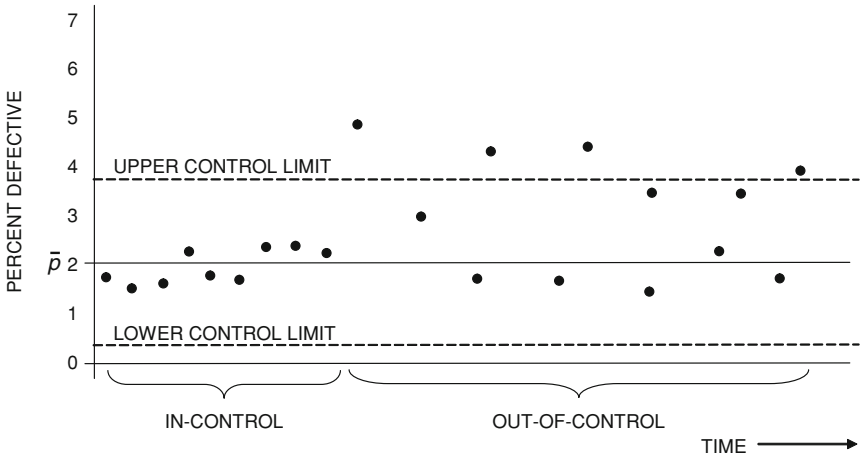

Fig. 14.2 Percent defective chart monthly performance

The cost of collecting data for p charts is likely to be less than that for $\overline{x}$ and $\overline{R}$ charts. As already brought to the reader's attention, this is one of the relative merits of SQC by attributes over SQC by variables. The extent of nonconformance of a product is expressed in terms of the number of defects per one hundred units.

The plot in Fig. 14.2 is a $\overline{p}$ chart for office work.[4] The variable being measured is a critical function characterizing a large back office operation. Information about quality becomes available as lots are tested upon completion. At negligible cost, information from such checks serves in eliminating claims by customer on erroneous data and costs for auditing such claims. These audits were necessary for identifying errors which had gone undetected.

Like the control charts by variables, a control chart by attributes may have one or more of the following purposes: (1) to discover the average proportion of defective articles or parts submitted for inspection over a period of time, (2) to bring the attention of management to any changes to this average quality level, (3) to discover those out-of-control high spots that call for action of identifying and correcting causes of bad quality.

The use of p charts assists in discovering those out-of-control low spots that indicate relaxed inspection standards or erratic causes of quality improvement. The decision procedure for sampling inspection of percent defective is as follows:

- Define the inspection lot.
- State level of acceptability: $H_0 : \mathrm{p} < p_0$ where p_0 is a quality constant.
- Select the appropriate sampling plan, which specifies the acceptance number based on AQL and sample size n.

[4] Precisely, the backoffice of a major credit institution.

- Select a random sample of n items from the lot under inspection.
- Inspect all items of the sample, and classify each as a nondefective or as a defective.
- Compute the number d of defective items found in the sample.
- If d is less than or equal to the acceptance number, accept the inspection lot; if d is greater than the acceptance number reject the inspection lot.

In terms of the percent defective charts applicability, there exist certain prerequisites of which the reader will be well advised to note. For example, the production process on which the chart is to be made must produce a relatively large number of units. In addition, inspection of selected units of product must be accomplished by the method of attributes, with units classified as either conforming or non-conforming to engineering specifications. More precisely, they are non-defective or defective in regard to the quality criterion being measured.

Another prerequisite, which can be found with most SQC plans, is the appreciation of the fact that what we are after are chance factors operating in the production process. Their result is that of producing a certain proportion of defective units in every set, or nearly every set, of units produced. The measure of this characteristic of defective units being produced is what was defined in Sect. 14.1 with the algorithm.

$$p = \frac{\text{Number of defective units in production process}}{\text{Total number of units in production process}}$$

where p is the production process' fraction defective. This measure of fraction defective does not tell which parts coming from the process will be the defective ones, but it does given the *probability* of drawing a defective unit.

- *If* the probability of drawing a defective unit is constant from trial to trial,
- *Then* the production process is in a state of statistical control.

The p chart helps to decide when the process is in a state of statistical control, which means when it operates essentially the same way, from sample to sample. Timed to respond to quality criteria on daily basis, the fraction defective *each day* is the number of parts rejected, divided by the number of parts inspected in the same day.

At the end of the month the mean fraction defective $\overline{p}$ is computed. The correct way to calculate $\overline{p}$ is to divide the total number of defectives by the total number of parts inspected during the period in reference—for instance a month—not as the average of mean value of subperiods or samples.

If n_1 is the number of units in first random sample drawn from the production process, and d_n is the number of defective units contained in first random sample of n units, *then* the fraction defective of the sample is:

$$p_1 = \frac{d_1}{n_1}$$

The classical frequency distribution of fraction defective values is computed from samples being drawn. Say that we take 500 samples, counting and recording the number of defectives in each sample; grouping samples by fraction defective at sample level; and counting the number of samples in each group. The result will be a frequency distribution of $\overline{d_s}$ which present a picture of the pattern of variation.

- *If* the process is in statistical control,
- *Then* the sample fraction defective values will exhibit the same pattern of variation from day to day.

The reader should however notice that with p charts the frequency distribution is not symmetric in all instances (see Chap. 8 on asymmetries). The standard deviation of the frequency distribution is the best estimate of process variation, and the mean of a frequency distribution is the best estimate of process fraction defective, p.

14.4 Upper and Lower Control Limits

Prior to calculating the upper and lower control limits of a p chart, we draw samples from the production process. The units in the sample are inspected and results recorded. From this information will be computed the center line and control limits. The mean fraction defective is:

$$\overline{p} = \frac{\sum p_i}{n_i}$$

where $\overline{p}$ is the center line, p_i the number of defectives and n_1 the size of the i sample. An estimate of process variation is provided by the standard deviation s_p.

$$s_p = \sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

The expected value of the population's p distribution can be estimated using $\overline{p}$ as a proxy. *If* all points fall within the trial control limits, *then* the expected value of the μ_p parameter may be assumed to be equal to the statistic $\overline{p}$. Similarly, the standard deviation of a fraction defective distribution, in other words σ_p of the population, uses s_p as proxy.

The control limits for the p chart are computed from the mean $\overline{p}$ and standard deviation s_p. The upper and lower control limits are, respectively, at $+3s$ and at $-3s$.

$$\mathrm{UCL_P} = \text{Upper control limit} = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{\sqrt{n}}}$$

$$\text{LCL}_\text{P} = \text{Lower control limit} = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{\sqrt{n}}}$$

To establish the framework for the p chart of a given process, we take a sample containing 50 units ($n = 50$), drawn at random from a lot or an ongoing process. We continue drawing till we have 20 samples, recording results and calculate limits (as explained in the following paragraphs). We draw additional samples and use the p chart to make decision according to the rule:

- If a point falls between UCL_p and LCL_p, the process is in a state of statistical control.
- If the point falls above UCL_p or below LCL_p, the process is not in a state of statistical control.

When the process is in statistical control, between UCL_p and LCL_p will be included over 99.7% of all sample fraction defectives values. As with $\overline{x}$ and $\overline{R}$ charts, these two limit lines become the quality production guideposts when placed on a chart. The method is very similar to the one employed in placing limit lines on the pattern of variation that was used with the normal distribution of $\overline{x}$'s in charting events by variables (Chap. 13).

An elegant way of thinking about the process underpinning fraction defective limits, for different probabilities of $\overline{p}$, is presented in Fig. 14.3. After computations have been made, limits may be converted into percent defective. If the sample size n is not the same from sample to sample, control limits must be calculated for each sample size. *If* the computed LCL_p is a negative number, *then* it is set at zero since negative values for p are impossible.

To use the facility provided by Fig. 14.3, we start at lower left-hand corner of the chart and move right along the horizontal axis to sample size, n, for which limits are to be obtained. From this point, we move vertically to the diagonal line corresponding to $\overline{p}$. Then we move horizontally to the left and read the number which is equal to $3s_p$. To obtain the UCL_p and LCL_p, this number must be added to and subtracted from $\overline{p}$ expressed in percent defective.

Notice that because the distribution of sample fraction defective values is not always symmetric, the limit lines placed at $3s_p$ from the mean will not always enclose the same proportion of area defined by $3s_l$ on the normal curve. Usually, however, this difference is small enough to be neglected.

New limits should be calculated as improvements are made to the production process and the percent defective changes. Incorrectly, however, many companies follow the policy that even if a quality condition improves or worsens the old limits are maintained—hoping that it is possible to continue without extra effort to use the previous system and its limits. This is the wrong way to control quality.

As the preceding discussion documents, $\overline{p}$ charts are flexible instruments, but they should not be abused. In each case, a chart can be plotted for whatever time unit is practical. In office work or group performance it is usually plotted daily.

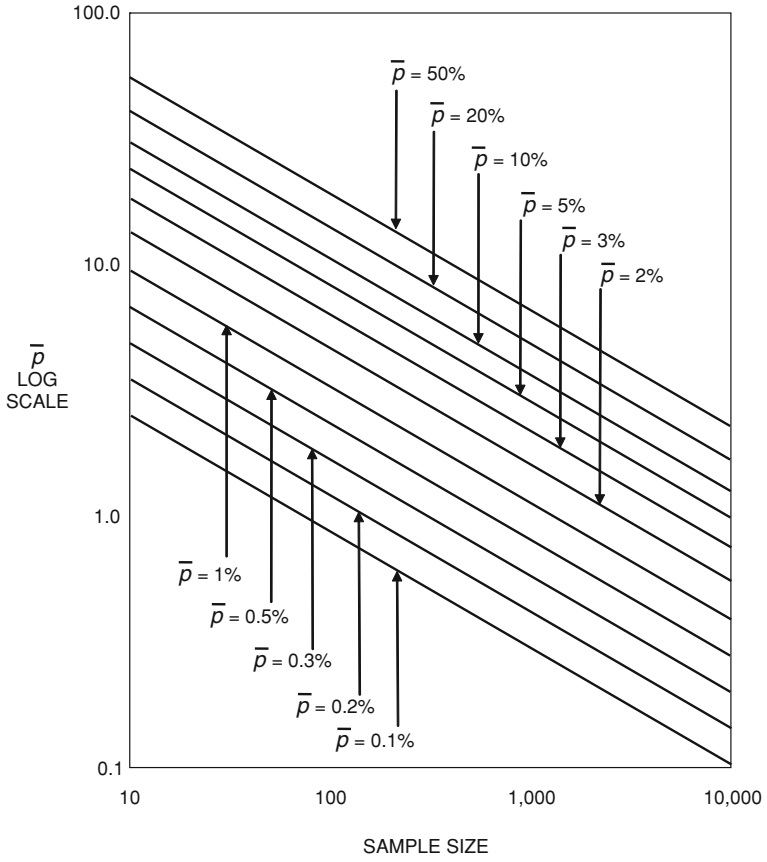

Fig. 14.3 Upper and lower control limits for the fraction defective chart

Where it is imperative to exercise close control of quality, plotting percent defective should be done intraday. For instance, $\overline{p}$ charts on the quality of production of individual parts of assemblies are often plotted hourly, and a similar principle is applicable to plenty of cases in office work.

One of the best industrial uses I have seen of $\overline{p}$ charts is where they serve as early warning signals flashing out abnormal percent defective situations. Outliers indicate out-of-control conditions and the best thing management can do is to immediately correct the causes and get back the process to the *in control* pattern. One of the worst things management can do is to forget about warning signals and let the process go even deeper out of SQC till action is finally taken.

Very useful as quality management devices are operating characteristics curves associated to p charts. As it is seen in Fig. 14.4 this OC resembles a Poisson distribution. The statistics behind the curve shown in this figure come from the computer industry.

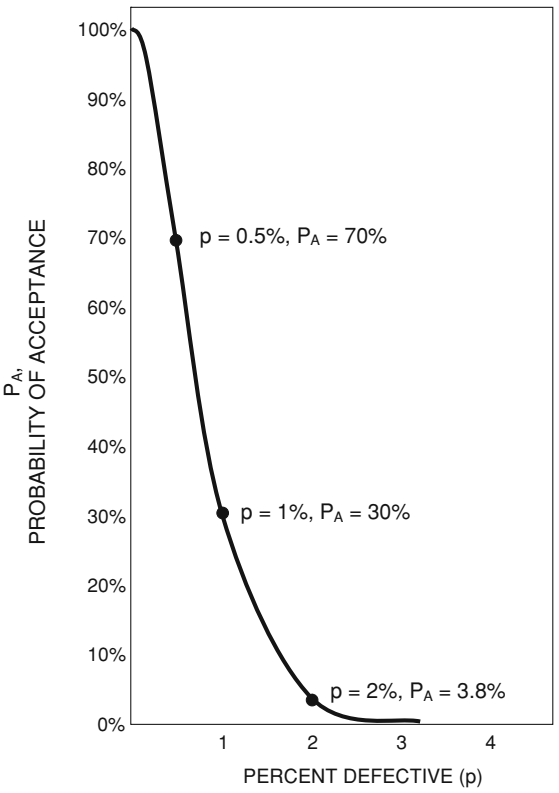

Fig. 14.4 Operating characteristics curve for percent defective

- The abscissa is the quality of lots coming into the sampling plan, expressed in percent defective, p.
- The ordinate is the probability of acceptance of each lot, P_A.

The $\overline{p}$ chart can also be used as a gage on the consistency of inspection. Companies which employ QC charts for percent defective found them to be excellent yardsticks on whether or not quality of inspection respects high standards. In many cases, $\overline{p}$ charts prove to be the ideal tool for improving or maintaining a steady quality level because:

- They show an abnormal variation in percent defective, where it exists; therefore, depicting trends, and
- They act as a strong psychologic tool on personnel for maintaining an above average quality level.

In the financial industry, one of the major fields where institutions have been successful with the implementation of percent defective charts is accounting and auditing—which is, after all, financial inspection. In other cases, results obtained from the implementation of percent defective charts have strengthened the

company's position in areas like financial planning, budgetary control, overhead analysis, and cost control [2].

14.5 Quality Control Charts for the Number of Defects Per Unit

Like $\overline{p}$ charts c charts, for number of defects per unit, work by attributes and serve in the inspection of quality levels. But while $\overline{p}$ charts target percent defective in a process, c charts map defects per major production unit. As such, they have a wide area of applications from manufacturing to the service industry, being of value when quality is determined in terms of quality of work in the large. There are several differences between a c chart and a $\overline{p}$ chart:

- On the c chart are plotted defects typically found in a big unit, but it may also be a sample of a few big units
- On the $\overline{p}$ chart, are plotted defective units found in sample divided by the sample's size
- The c chart has limits based on the Poisson Distribution
- The $\overline{p}$ chart limits are based on the binomial distribution.

With the c chart is even more important to make a distinction between *defectives* and *defects*, than with the p chart. A defective is an article that in some way fails to conform to one or more given specifications. Every defective contains one or more defects. Each instance of the article's lack of conformity to specifications is a defect. Moreover,

- The c chart need not be restricted for a single type of defect and
- It may be used for the total of many different kinds of defects observed in one unit.

The c chart is of particular value in connection to assemblies and completed big products. The aircraft industry has used c charts to good advantage by focusing on defects per plane in order to control plane production quality. Other successful applications of c charts have been in the automobile industry regarding defects per hour's production.

To establish the frequency distribution of defects found in a statistically valid number of samples (each being one major production unit), we choose that number of samples; count and recording the defects in each of them; group samples by number of defects found in each one; count the number of samples in each group; and plot a graph of defects per sample. The frequency distribution:

- Will be non-symmetric
- Its mean $\overline{c}$ is an estimate.

When a standard value of the mean number of defects per unit is not used, then $\overline{c}$ may be estimated as equal to the arithmetic mean.

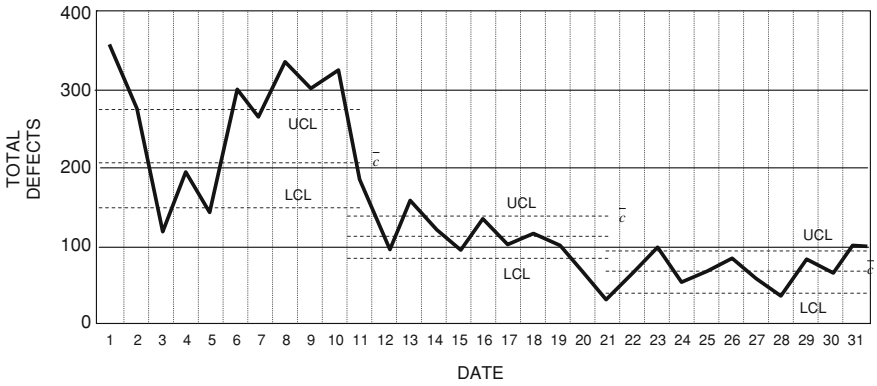

Fig. 14.5 Centerline UCL and LCL for fraction defective c

- The spread of such frequency distribution is an estimate of process spread, and
- The standard deviation of the c distribution is equal to $\sqrt{\overline{c}}$.

The limits of the frequency distribution from a process in SQC become the c chart control limits. Over 99.7% of all samples will fall within these limits. For a c chart, the control limits are:

$$\text{UCL} = \overline{c} + 3\sqrt{\overline{c}}$$

$$\text{LCL} = \overline{c} - 3\sqrt{\overline{c}}$$

For UCL and LCL to be calculated, samples must be drawn from the production process, the units constituting the sample inspected, and the results recorded. From this information are computed the center line and control limits. Some plans in manufacturing call for setting new limits every 2 weeks if the process remains in control. The problem with pre-fixed intervals is that the process may be in control but the centerline has changed, as in the example in Fig. 14.5 with the $\overline{c}$ chart.

Figure 14.5 presents a $\overline{c}$ chart with upper and lower control limits, spread over 31 days. Notice that UCL and LCL change as a function of time as the total number of defects changes. This application and its statistics come from final inspection of automobiles.

Both the mean $\overline{c}$ and the spread are decreasing as a function of time. Furthermore, although in the first place the manufacturing process was mostly out of statistical control by the end of the plot, it came in control. This example emphasizes once again the great assistance that quality control charts can provide to engineering and management.

The c chart shown in Fig. 14.6 comes from an industrial application. Its objective has been to map into a graphic form the number of hourly adjustments on the production floor. In the implementation from which this figure is derived, the quality control limits were based on an average value taken from previous data

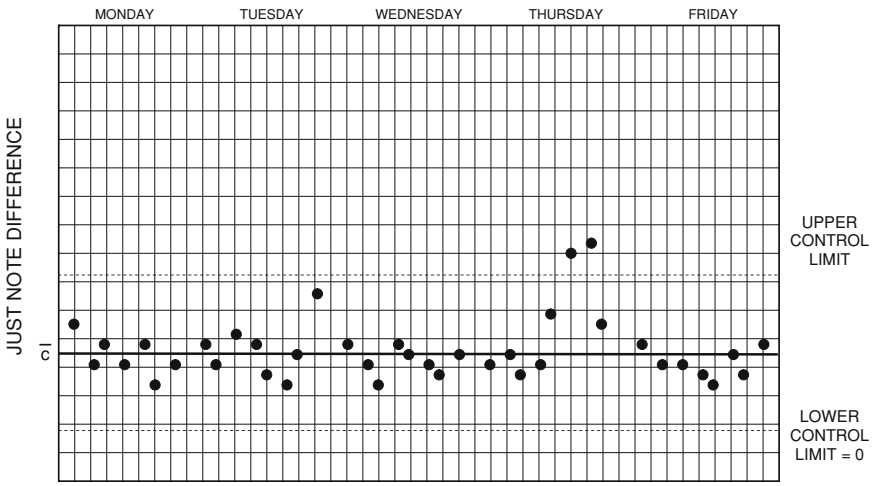

Fig. 14.6 Quality control chart for defects per unit

collected at the same production floor. The application tracked quality *intraday*, $\overline{c}$ was equal to 22.1 adjustments per hour. With this statistic, the resulting upper and lower limits are:

$$\text{Upper control limits} = 22.1 + 3\sqrt{22.1} = 36.2$$

$$\text{Lower control limit} = 22.1 - 3\sqrt{22.1} = 8.0$$

For c control chart purposes, when there is no 1 unit being inspected by a subgroup consisting of more than 1 unit, some companies divide defects c by units n. The ratio c/n is often represented by the symbol u.

Similar charts can be plotted for a range of applications in office work, including the execution of customer orders. In principle, fraction defective points on the c chart should fall inside control limits based on past performance as long as the system of chance causes remains the same. If points fall outside, then the system is out of statistical control. In the opinion of most experts, the use of the c charts is most appropriate if:

- The opportunities for a defect in a production unit are nearly infinite.
- The probability of a defect at any point (component, subassembly) of the unit is rather small and fairly steady.
- The area where defects can show up, that is the size of the unit, is constant.

Other experts believe that contrary to some of the statements made in the preceding paragraphs, the c chart has a more restricted use than the p chart. The reason they give for this statement is that each subgroup of the c chart consists of a

single article (albeit a big one), and the variable c consists of the number of defects observed in that one article. This is not an unchallengeable argument.

c charts are so important because they can be applied to a wide range of cases. An interesting implementation is that in connection to the number of accidents occurring on a turnpike during high density of traffic from 4:30 to 5:00 pm.

- The number of automobiles passing a given auto route section represents the number of trials.
- An accident on that section represents a defect.

As with p charts, the general principles for c chart implementation are to: (1) have a process which is put under statistical control; (2) state the null hypothesis: "This process is in a state of statistical control"; (3) take a sample of size n (remembering that when using a c chart, the sample size must be constant); (4) inspect that sample for all chosen quality characteristics; recording the number of defects found; (5) plot on chart; and (6) take action dictated by the c chart in accord with SQC rules.

To avoid the obsolescence and sometimes irrelevance of centerlines as well as of UCL and LCL, it is wise that computations regarding these statistics are made at periodic intervals and quality control undertakes appropriate tests to assure the process has been in control. Whenever the process changes and control is restabilized at a new centerline level, new control limits should be computed based on a fresh accumulation of data.

Moreover, since a major advantage of the SQC chart is its psychologic effect on workers and inspectors, it is important that SQC charts are well constructed. Several companies allow for notes to be recorded below the chart, indicating that action has been taken at various times as well as failures in taking a timely action.

References

1. Packard D (1995) The HP way. Harper Business, New York
2. Chorafas DN (2000) Reliable financial reporting and internal control: a global implementation guide. Wiley, New York

Chapter 15
The Culture of Statistical Quality Control

15.1 Fulfilling the Prerequisites: Culture and Expertise

The purpose of this chapter is to explain how small and medium enterprises (SMEs) can gain the advantages statistical quality control (SQC) can offer. The text is written in a way that bigger firms, too, may use it, particularly those who do not yet have but contemplate installing a statistical system for quality assurance.

On the hypothesis that many small to medium enterprises have not in operation SQC, I start by outlining the prerequisites, foremost among them the change in internal culture regarding product assurance. Skill, too, is important but it can be acquired from outside the firm both to help in starting an SQC system and to train the company's people on how to use and manage it.

Culture cannot be acquired from the outside. Section 15.2 provides the reader with a real-life case study on how much cultural change counts. SQC calls for discipline and in a number of companies this is in short supply. To be successfully implemented SQC also requires homogeneous systems and procedures throughout the organization. Frequently, this is assumed to be the case, but is not documented by the facts.

In addition, because one of the important characteristics of SQC implementation is the change from 100% inspection to sampling inspection, the company's management and all of its professionals must appreciate statistical inference. This cultural change is not as easy as might seem because not all schools teach causal inference as they should do.

Whether by attributes or by variables, statistical plans imply randomization and they include causative factors which assist in data interpretation. Causal analysis may be complicated by the fact that in many cases there are several causes, some of which may be interrelated—hence the need for experimental design discussed in Chap. 11.

In the background of the approach to quality control in the large and in the small this book described, lies the fact that, in the real universe, there are no fixed sets of *self-evident* truths or theoretical definitions, axioms, and postulates. Researchers, and

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering, 295
DOI: 10.1007/978-1-4471-2966-0_15, © Springer-Verlag London 2013

indeed all scientists, typically operate on the basis of assumptions which they have accepted as being sufficient up to a point, but:

- They continue being eager to challenge the "obvious" and
- They are open-minded about discovering that some of their assumptions might be false, or that some other principles they had not appreciated are the determining factors.[1]

I bring all these notions to the reader's attention because while SQC is a powerful scientific tool, it will not be effective if its implementation clashes with the company's culture. The wrong way of looking at science is to believe that scientific proof is a matter of showing formal consistency with a set of what are treated as being *self-evident* truths because they are based on prevailing theories. Not only is this false, but also the effect of believing in it:

- Impacts in a negative way the mind of people and
- Leads to denying the existence of anything outside the bounds of that system whose "laws" make it practically deterministic.

By contrast, the concept of causal inference is synonymous to a state of mind characterized by the quest for answers which are *stochastic* rather than "certain". In real life the notion of being "certain" answers rarely if ever research requirements. Expecting things to come your way is an illusion. Stochastic answers involve uncertainty about a measurement, an observation, an event, an outcome. But that is how science (and to a considerable extent business) works. What I just said is valid whether the answer concerns:

- A description,
- An observation, or
- An explanation.

The existence of an open mind able to think in terms of statistical inference is so fundamental that it underpins the entire field of statistics. This statement is valid whether we talk of tests of hypothesis, sampling, $\overline{x}$ plots, $\overline{R}$ plots, percent defective p, fraction defective c, tests of significance, statistical tables, or a graphical exposition of trends and conditions.

Measures of central tendency and of dispersion, frequency curves, regression analysis, covariance, correlation, and a long list of statistical tests, or presentations, evidently including SQC are all based on causal inference. This is also true whether our measurements are experimental or observational, but in quest of description and explanation.

The problems to which these notions apply are not only industrial or scientific. They may as well be economic, demographic, or many other fields where causal inference is at a premium. The reader who went through the first 14 chapters of this

[1] Everything changes, said Herodotus the ancient Greek philosopher.

book will appreciate that the outlined approach to the analysis of cause and effect rests on three pillars:

- Statement of the hypothesis,
- Estimation of parameters, and
- Testing of the hypothesis.

To analyze data statistically the hypothesis must be expressed in a mathematical form, as a model with two components: assumptions and random events influencing the subject under study. Assumptions are outlined in a cause and effect duality. Random events are essentially responsible for deviations between observed and expected values—unless there is a bias in the system.

The test of hypothesis takes the form of testing the results of statistical analysis against other knowledge some of which is derived from causal inference. In the medium to longer run, however, confrontation with new data derived from operations is the ultimate test of hypotheses. Even when it is verified, the hypothesis remains a tentative statement which might be upset by new facts.

Any SQC plan, including the simplest ones, which does not start with the statement of hypothesis and follows up with testing will not worth its salt. Sampling plans, too, have rules which must be observed. In addition, the right methodology must be established for an SQC implementation.

These are, in a nutshell, the elements of the new culture a company will need to acquire in order to successfully implement scientific tools in its management, SQC being one of them. This leads to the next prerequisite: How well trained are its human resources in terms of stochastic inference and statistical tests.

A company which starts with implementing SQC will require an experienced person, usually an outsider to train and to provide technical assistance to the internal quality control group. My recommendation is that this SQC specialist starts his work in collaboration not only with manufacturing but also with engineering, to assure that the inherent quality of a design is fully observed in the manufacturing operations.

Consultants, however, should be on tap not on top. The company will need to develop internal SQC resources, and to appoint an able person to be in charge. His *job description* should include checking all parts used in fabrication, to ascertain that all components have correct tolerances and acceptable failure rates. This evidently requires establishing and carrying out a program of acceptance testing for all incoming parts; as well as a program of production level tests to be applied to the fabrication lines.

Together with the SQC consultant, design engineers and manufacturing engineers (also field maintenance personnel, if field maintenance chores are applicable) the SQC manager should initiate a program to study the company's specific mechanism of failure. The details of this effort will be carried out at the functional group level. More often than not, the aforementioned effort will require a thorough reliability evaluation of the production equipment to be made through special tests.

- These tests should yield an accurate measure of both the inherent equipment reliability and accuracy in observing engineering specifications and
- Because manufacturing may have conflicts of interest in evaluating its present machines, such tests should be performed by an independent, unbiased group with some outside assistance.

Also part of the SQC manager's *job description* must be the investigation of the current state of the art of testing for quality. By monitoring failure reports the SQC manager is in a position to inspect, at frequent intervals, the behavior of equipment used for production. To perform these tasks in an able manner, the SQC manager will do well to follow the advice by the father of Admiral Chester W. Nimitz to his son:

- "The sea—like life itself—is a stern taskmaster."
- "The best way to get along with either is to learn all you can."
- "Then you do your best and don't worry—especially over things you have no control [1]."

If the word "sea" is replaced by quality and reliability, *then* what Nimitz' father said fits hand in glove stochastic inference—and therefore the premises necessary to successfully implement an SQC system.

15.2 Restructuring System and Procedures

The president of GAMMA, a mid-sized manufacturing company, expressed as follows his opinion on the projected implementation of SQC: "We are concerned with the problem of quality inspection, but somehow it has been difficult to introduce statistical tools in our firm. Discussions along the line of a statistical methodology have taken the last 4 years. We now wish to introduce the necessary systems and procedures, but both our factories must be in accord on common standards."

In the opinion of the executive VP/Manufacturing, GAMMA was missing an analytical culture. The way he put it: "What we do today is very largely empirical. We do not really know 'why' in an analytical sense, we just do so. In some aspects we lack coordination in quality control. But our competitors, too, are confronted by this problem."

A preliminary study showed that the lack of common standards on quality inspection prevailed not only between the company's two factories, but also between different shifts in the same plant. The second shift, for example, had higher percent defective than the first because it employed only foremen. There were no engineers present in the second shift as they all worked in the first shift. Quality simply received a different handling. (From my experience with other companies I can say that the second shift almost always gives lower quality results.)

Because several incompatible standards could be found in the fine tissue of the manufacturing process, it was important to account for them at a certain degree of precision in order to establish the appropriate quality control methodology. Factory I, for example, had installed at the production floor a quality control template which permitted the supervisor to easily decide whether adjustment is necessary.

Factory II was not following this (recommendable) practice. What is more, there existed certain confusion between what was called "screening test" and fully fledged quality control. That screening test had little if anything to do with inspection at the production floor. Quality control was done on a 100% basis at the "banks" between two production lines where semi-manufactured goods were stored.

The differences between the two entities went down to a level of detail. Factory I did electrical tests for resistance. Within Factory I operations, however, there existed substantial disagreements on the importance of the electrical resistance tests, as opposed to weight and diameter measurements.

Factory II did not believe in these tests. At the headquarters, some of the manufacturing engineers like one approach; others follow exactly the opposite one. Testing methods varied so much from one factory to the other that products which passed the test in one factory could be rejected in the other, and vice versa.

Rejection meant that the product will not be outright scrapped but will be (probably) sold at a discount. Or, in case the factory was out of stock of products which passed the test, rejected products will be reworked and sold.

Another example has been that of tests destined to give indications on minimum and maximum weight which is important for GAMMA's second product line of manufactured goods. In Factory II, ten units were taken out of a box and controlled for weight. Factory I took a sample of five units out of each box for quality inspection purposes. From this sample of five, Factory I computed the mean weight. Corrective steps were taken if the mean weight exceeded specified limit.[3]

The flaws in the two factories' quality control procedures do not end there. The inspectors assigned to the job were not properly trained. They had a cookbook in which a quarter of the values of control ranges were missing. Going through the cookbook I found some control data were written in pencil. The answer to my question *why* has been that they were tentative (!).

To make matters worse, the inspectors were thought as being able to know by heart the different control ranges. Out of five tests one of them did in front of me, he passed one test but failed the four others he could not exactly remember. This is how GAMMA built quality into its products.

Even when and where mathematics of sorts was used, the methodology necessary to improve inspection procedures and practices was missing. In one of the production lines of Factory I, sampled lots were checked through the kind of tests which were disavowed by Factory II as being irrelevant; and when sampling inspection was practiced, the sample size varied.

---

[3] This, however, was characterized as "tentative".

The statistical plan in that particular product line of Factory I, where a semblance of SQC was used, run as follows: If one, two, or three defects were found, the lot was accepted. If four or five defects were found, the inspector drew another sample. For six defects or more the lot returned for 100% inspection. Three things can be said about this approach:

- Superficially, it looked as if it were a reasonable sampling plan. In reality, this sampling was invalid because no study has taken place to document its mathematical/statistical foundations.

As a result, the acceptance/rejection of lots worked like a roll of the dice, without confidence limits being placed on the (accept/reject) decisions.

- The quality control plan lacked even a tentative corporate acceptance by senior management.

The director of marketing said that *if* this were a sound quality assurance program, it would have been adopted throughout GAMMA's operations. Marketing wanted to have a homogeneous basis for quality judgment, but then several of the company's engineers challenged the validity of that plan.

- The difference of opinions among manufacturing engineers and design engineers, at headquarters, was such that no sound SQC plan could be developed.

The contradictions characterizing the different quality control opinions at the home office were such that wherever quality records existed they were incompatible with one another, hence impossible to examine them in a dependable manner. It comes therefore as no surprise that the mission given to the consultant was to come up with a firm plan for SQC able to integrate the different practices and assure that sound quality control principles are first established, then followed by a thorough revamping of systems and procedures.

Three alternative sampling plan solutions were elaborated by a working group which included designers, manufacturing engineers and the consultant. Testing each one of them, the group came to the conclusion that, given the company's product line, the better method to adopt was sequential sampling with lot templates (Sects. 15.4 and 15.5). A cost/benefit analysis was made prior to presenting this quality control plan to GAMMA's management board for approval.

Integral part of the proposal has been an intensive training program involving key personnel from headquarters, and the two factories. Both engineers and marketing people participated in an effort to induce a cross-departmental cultural change.

Regarding the implementation mechanics of the quality control system, the first options of starting with one of the factories was disregarded because of the risk of inter-factory friction. Instead, it was decided that the better approach was to be implemented SQC simultaneously at Factory I and Factory II at properly selected production lines.

Based on lot templates and sequential inspection, the methodology of the SQC plan which was adopted, allowed switching from normal to tightened inspection

procedures as quality worsened. Then, back to normal and eventually to reduced inspection as quality improved.

The consultant insisted on the fact that the new systems and procedures being elaborated for quality control, as well as the statistical methodology underpinning them, should be followed throughout GAMMA's operations without "ifs", "buts", and deviations. This required extending the original training program beyond quality inspectors and salesmen to include management personnel at headquarters and the factories. The aim was to assure:

- The full understanding of the SQC method and
- The need to observe all prerequisites of an effective statistical quality assurance.

The consultant further suggested a number of experimental studies aimed to establish the exact nature of defects and their background reasons. Apart from the benefits these studies were expected to bring in better understanding of quality problems, the proposed plan bet on a reduction of inspection time by eliminating repetitive work. One of the plant managers had provided evidence that in the course of a month 23.403 h were spent on inspection and control.

In conjunction with the SQC methodology, the question was raised about defect patterns by the plant. In that same factory, the cumulative rejection rate for all types of products has varied between 12.6 and 26.2%. Worse yet, 8% of these rejections precisely were non-identified—a vague reason being given was: "early processing".

The cultural change effort progressed in parallel to the SQC methodology. As one of the factory directors was to comment "I am many years in manufacturing. There are many things which are *very* necessary. We have to do them. In deciding 'what' we must study expected results and check costs against them. The time has come to institute a uniform quality control system in our company."

15.3 Implementation of Sampling Plans in Smaller Firms

The concept and process of sampling have been fully described in Chap. 9, where it was stated that sampling plans are the foundation of every SQC procedure. Table 15.1 summarizes the symbols and abbreviations with which the director of an SME and his SQC manager should become familiar.

What the general manager of a company introducing SQC, and his immediate assistants, should appreciate is that in the background of sampling plans lies the fact they are expected to reveal the level of quality assurance in the population. It needs no explaining that whether we talk of manufactured products or financial accounts, their inherent quality level is very important.

The first step in utilizing sampling plans is to realize that this is a reasonable compromise and not an absolute guarantee of outgoing quality—which cannot even be provided by 100% inspection. Chapter 14 familiarized the reader with the percentage of defective items and the probability of accepting a lot containing that percentage of defects at a given level of significance. The concept of level of

Table 15.1 Table of symbols and abbreviations used in SQC

Widely used	
AQL	Acceptable quality level
AOQ	Average outgoing quality
AOQL	Average outgoing quality limit
LTPD	Lot tolerance percent defective
OC	Operating characteristics
OCC	Operating characteristics curve
P_A	Probability of a lot being accepted by a sampling plan
SQC	Statistical quality control
in Sampling	
N	Number of items in a population
n	Number of items in a sample
in $\bar{x}$ *and R Charts*	
x	Measurement of one item in a sample
$\bar{x}$	Mean value of a sample
$\bar{\bar{x}}$	Mean of mean values
R	Spread, or range, of a sample
$\bar{R}$	Mean range
UCL	Upper control limit
LCL	Lower control limit
UTL	Upper tolerance limit
LTL	Lower tolerance limit
in p charts	
p	Percent defective
$\bar{p}$	Mean percent defective
in c charts	
c	Defects per production unit
$\bar{c}$	Mean defects per production unit

significance, or confidence, has been introduced in Chap. 10. Basic notions in a sampling inspection are:

- A representative sample is one drawn from a uniform population.
- Random sampling could make a "not so uniform population" act *as if* it were nearly uniform.
- The sample size should be large enough to adequately represent the population.
- The size of the population generally does not affect the representativeness of the sample; more important is the sample size.

Typically, sampling plans are classified by sample size, a measure of the amount of inspection required, and acceptable quality level (AQL) of the lot (or sample). A company starting with SQC must appreciate the importance of capturing the characteristics, effects, and interactions of items in a population. This is done by using the sample as proxy.

The bolts and nuts of sampling are simple. Typically, a random sample is taken from a lot, and a decision reached to accept or reject is based on the number of

defectives found in it. There exist different sampling methods and choice among them is usually influenced by intended use. Such choice should take place at the beginning of a quality control program, in full consideration of the dynamics of the statistical inference we are after. A post-mortem choice is like betting in a horserace after the horses run their course.

A standard or normalized sampling inspection procedure is a procedure for selecting and using sampling plans in accordance with statistical theory. With a standard procedure, the range of choices is narrowed to relatively few alternatives, with detailed prescriptions becoming available for choosing among them. A normalized approach leads to quality control decisions through properly established evaluation steps.

Because the mass production of manufactured goods and of services like the handling of transactions and accounts, is the order of the day, it is no longer practical to sort item-by-item and look up each one's quality right after they have been made. I do insist on these matters because statistical inference is so vital to the success of an enterprise—of practically every enterprise.

- A sampling plan works in accordance with statistical laws and its results are always subject to probabilities.
- But over a period of time, a sampling plan will give desired protection, even if there are risks associated to the behavior of "this" or "that" variable.

An interesting similitude between SQC, indeed quality control at large, and business decisions is the importance of identifying and handling the controlling variables. Decisions are not made in the abstract. This identification provides the necessary focus which concentrates the decision maker's mind on what is really important.

Let us consider a case in which it is desired to control the variability of a production process. Statistical relationships connected to random variations add by the sum of squares, rather than by straight addition. *If* one source of variation contributes seven units and another two units, the total variation is not too different from the more important of the two[4]:

$$\sqrt{7^2 + 2^2} = \sqrt{53} = 7.28$$

In effect, this is saying that the vital variable has such an overriding effect that even complete elimination of a secondary source would be insignificant. The lesson to be learned from this example is that the proper combination of vital few variables, identified by the approaches already described, is the best way to be in charge of a process—from decision-making to quality assurance. This is, as well, one of the simplest ways to explain the logic behind SQC.

What the preceding paragraph brought to the reader's attention is the so-called *zero center method*. Say that the manufacturing process has the capability to be

[4] This in managerial parlance may be behind the *salient problem* confronting the CEO.

within engineering tolerances, but *resetting* is required due to tool wear. If so, resetting is the key variable.

The way to be in charge of it uses quality control by attributes (Chap. 14). Companies employ a marked board, laid out in colored areas, in which holes are drilled to accommodate the workpieces which belong to a sample and have been gaged by the operator.

- Parts above the nominal dimension are placed on one side of the board and
- Parts below the nominal dimension are placed on the opposite side.

If one side of the board has been filled and a critical number has not been exceeded on the other side of the board, this simple SQC procedure conveys the message that the process has shifted and needs to be reset.

Just like the operator must be trained to apply this method the SME's CEO must be trained to appreciate it. The operator also needs to be provided with a gage set at the nominal dimension of the process, as well as with a board containing holes that have been equally divided on the left and right sides. The top hole(s) on both sides are painted, say, red to designate the critical number.

The operator's or inspector's work is by no means complex. What he does is to check the parts at random intervals depending upon how fast the process shifts. This resembles the normal, tightened, and reduced inspection of which we spoke in Chaps. 13 and 14. He places these parts on the correct sides of the board.

When either side of the board is filled, the operator looks at the opposite (remaining) side to determine whether or not he has gotten "out of the red". If not, then his machine must be reset.

The message this simple example brings to the reader's attention is that statistical inference does not always require sophisticated models. At operator level the setting may be very simple so that the statistical method is applicable without extensive training. The inspector should however have a deeper knowledge so that he can explain to the operator (and to management) not only how the method works but also *what* went wrong and *why* it did so.

15.4 Lot Templates for Quality Inspection

Lot templates, the theme of this section, is a fairly simple and popular sampling technique from quality inspection that could assure an accuracy comparable to that of conventional sampling plans, particularly where lot sizes are small. The example with the board in Sect. 15.3 is indeed a lot template.

Fundamental to a lot template quality control plan is a histogram which gives information about product quality. The essential activity is no different from that discussed in Chaps. 13 and 14. Samples are taken from ongoing production to be controlled by the template. Our assurance of the end result's dependability comes from the knowledge that:

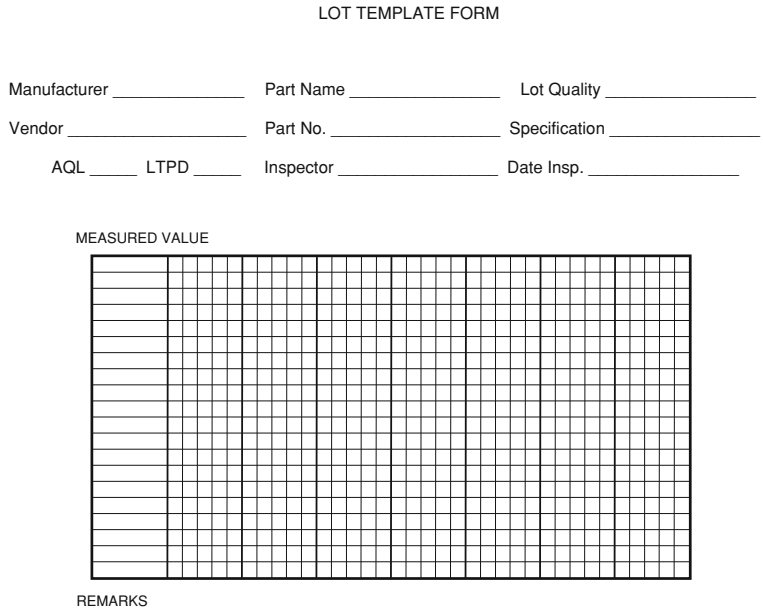

Fig. 15.1 Lot template form

- The variability of the sample is small and
- The histogram is well centered within specification limits.

Contrary to the example in Sect. 15.3 the lot template plans we will study in this section are for inspection by variables and not by attributes. Moreover, the basic procedure is for *acceptance inspection*. Templates are not a tool for control of a production process like the applications we have examined in Chaps. 13 and 14.

To start working in a lot template plan we must obtain a random sample of 50 pieces which will be measured for the specified quality characteristic we wish to control. This must be a random sample. As with every quality inspection plan, the next step is to record the measurements on a *lot template form* (LTF), on which we have set the appropriate specification limits and a suitable scale. An example is shown in Fig. 15.1).

After each measurement is made, a number from 1 to 50 in sequence is placed in the row or cell that corresponds to the value of the measurement. When all 50 readings have been plotted, we have a simple frequency tally or picture of the spread of the product's characteristic that was measured.

Whenever possible it is advisable that the measurement scale is adjusted so that the resulting frequency tally (or histogram) has between 9 and 15 cells. This permits to get a more accurate estimate of the population from which the sample of 50 pieces has been taken. Such an adjustment can be accomplished by estimating

Table 15.2 Determination of lot template scale

Range	Units per cell
3–6	1
7–13	2
14–26	3
20–32	4
27–32	5
33–40	6

the spread of the product from the range of the first seven pieces, driving the scale according to Table 15.2.

The *range R* is computed from the first seven measurements of the random sample of 50, used to construct the plot. *Units per cell* gives the number of units of measurement that must be assigned to each cell to result in plots of approximately 9–15 cells. For range values over 40, or for decimal values, the quality inspector or operator should move the decimal point to the left or right accordingly.

- *If* the *R* value is between 0 and 3 measurement units,
- *Then* the sensitivity of the measuring instrument must be increased.

Notice that while a range of one sample of seven observations may give the most efficient estimate of the scale to be used, the calculation of template values is based on the $\overline{R}/d_2$ relationship which provides a proxy of the standard deviation (see Chap. 2 for the values of d_2). Hence product history and specification limits should also be considered in determining the appropriate scale.

Figure 15.2 presents an example with three templates. The first is a normal template of those most frequently used (we will see how). It has nine cells and a standard deviation (s) of 1.5 cells. The second is also a normal template with 12 cells, with an s of two cells. It has been included to underline that the lot template method does not deal only with normal distributions and their approximations.[5]

When the product being inspected is well centered within the specification limits, this indicates that the process is in control. *If* the product is grouped against one of the specification limits, *then* there is a problem and it may be necessary to follow the procedure for screened lots.

Depending on the application, it may happen that during the definition of inspection by variables it becomes necessary that the measurable range be subdivided into small intervals. *If* the specification is such that we cannot divide its range into a suitable number of intervals, *then* either:

- The specification is unrealistic or
- The measuring instrument is not sufficiently sensitive.

[5] It might happen that because of low product variability, the use of an inadequate measuring device, we cannot divide the range of measurements into sufficiently small intervals. When this happens, we may have a plot only 3 or 4 cells wide.

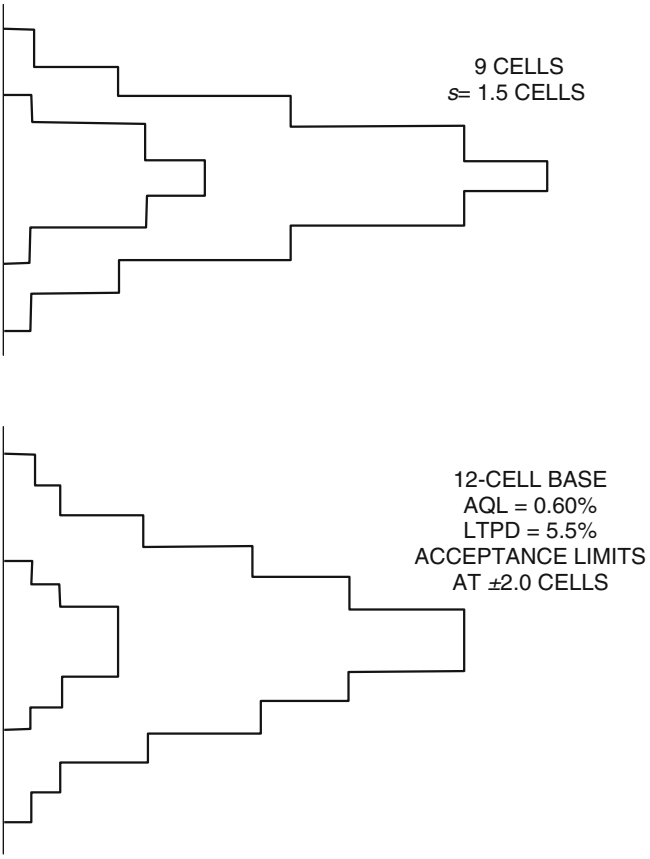

Fig. 15.2 Example with two lot templates for normal distribution

Provided that this setup phase of lot inspection by templates has been successfully completed, we can proceed with the second phase of the plan which consists of testing the plot through the use of transparent guidance templates on which are marked the upper and lower cell limits. An example is given in Fig. 15.3 with skewed templates, designed to be fitted over the sample plot. When this is done, an estimate of the variability is determined in terms of cell units.

As the foregoing discussion suggests, template design is a crucial aspect of the whole process. A company will be well-advised to develop a series of templates having different numbers of cells and able to accommodate plots of varied spread or base width is available.

For an example on how to proceed with implementation, say that we have plotted a random sample of 50 observations from a lot. If this lot is approximately normal, the number of observations in each cell of the sample lot can be expected to fall within the upper and lower cell limits of the template. The determination of *cell limits* is accomplished by binomial expansion using the expected frequency

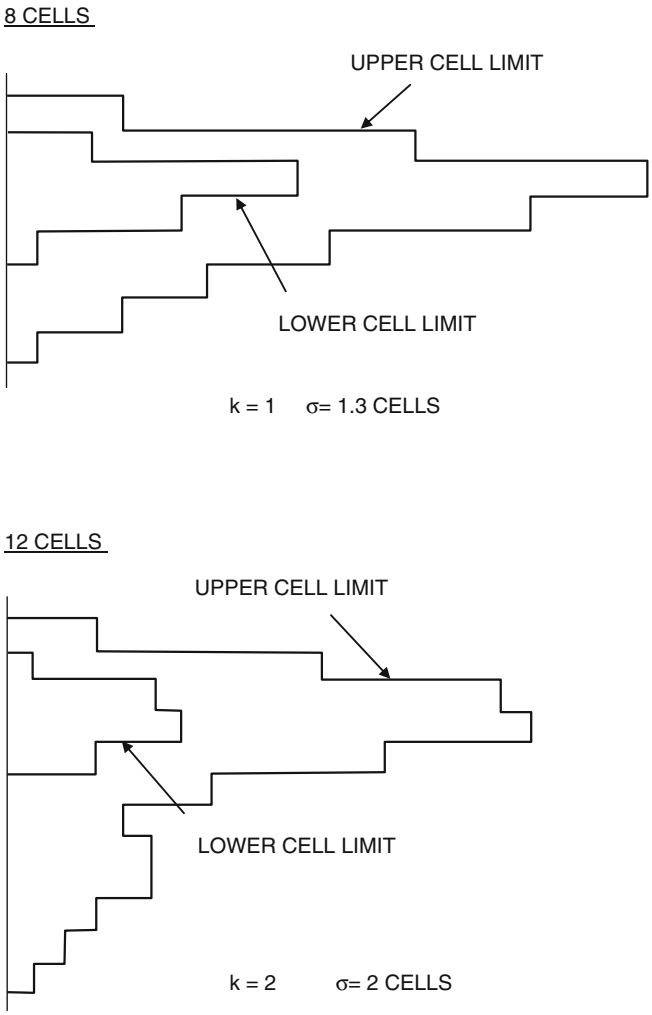

Fig. 15.3 Example with skewed templates

limits for each cell. This should be calculated so that the overall results will fall within 95% acceptance of plots from essentially normal lots, when tested with normal templates.

It is not difficult to appreciate that the implementation of the template method is quite forward. Although it requires training it is not necessary for this to be done at a sophisticated level. On the contrary, the developers have to be versatile in mathematics because behind the foregoing example was the performance of a test similar to the Chi-square (χ^2) for goodness fit.

When the smallest template which fits the plot is found, this becomes the best template estimate of the shape or spread of the parent lot from which the plot was

made. If the normal templates do not fit the plot, the developer(s) should try skewed templates of $k = 1$ or $k = 2$. The developer(s) should also be keen to estimate the mean and variability of a lot. For this they can compare the template estimates with other standard estimates of mean and standard deviation. For instance,

- For instance the template estimate of the mean can be compared with the arithmetic mean of the samples and
- The template estimate of the standard deviation compared with the root-mean-square method.

The distribution of the template means and standard deviations can show whether there is less or greater variability than the distribution of the true averages. If neither is the case at the chosen level of significance, it can be concluded that the template estimates are quite satisfactory.

If the quality inspector does not know whether the lot is or is not normally distributed, it is better to start testing with normal templates and then change according to the obtained results. It is a recommended policy always to try to find the smallest template that fits. As an example, a 9-cell lot is better than a 12-cell lot. The quality engineer should start by trying nine-cell templates rather than jumping higher.

All said, the use of a lot template plan makes possible quality control through meaningful patterns. The producer's and consumer's risk can be reduced by the increased discrimination of the plan. By this is meant that for a given AQL one can reduce both:

- The chance of rejecting acceptable material Type I error, α and
- The chance of accepting products that contain too many defectives (Type II error, β).

Specific solutions for nonnormal distributions can also be provided, with the variables plot often giving a clue to the reasons for defective items. Such conditions as skewness, widespread variability, or off-centeredness will frequently be indicated by the histogram—providing a good basis for bringing the manufacturing process in control.

For procurement activities, too, the lot template method can find many useful applications in acceptance inspection. Indeed, it is particularly well adapted to receiving or final inspection where large quantities of completed material must be handled. Its limitation is that while it is widely applicable with manufacturing goods it cannot help in process control, at least to my knowledge.

In conclusion, as the examples which we have seen demonstrate for the manufacturing industry, and most particularly for the SMEs which do not have an army of quality engineers, a lot template plan offers a relatively simple SQC approach which reduces training and administrative problems while it is adaptable to different inspection situations. The constant sample size and elimination of calculations and/or use of formulas frees the inspector for more hands-on work.

15.5 Acceptance Limits for Lot Templates

An important step in the development of a lot template plan is that of testing the plot for acceptance with the appropriate acceptance limits. These are derived for any degree of protection or AQL desired, by varying the distance of the limits from the center or mean of the template. This must be accomplished with attention because established acceptance limits will be responsible for acceptance or rejection for quality reasons.

The reader should notice that such limits must not be confused with lot limits of the lot plot plan, which are fixed a ±3 sigma. The engineering specifications limits help in establishing the positions at which the lot template's acceptance limits should fall.

Acceptance limits of ±3 sigma will provide a tight plan of AQL 0.05% which can be employed in place of 100% inspection. If the measurements fall by a large majority inside the acceptance limits, and therefore the tolerance limits, the lot should be accepted even if a couple of measurements exceed the acceptance limits but are within specifications.

Figure 15.4 shows a typical a typical plot tested with an acceptance template to determine lot acceptability. The lot is acceptable because the plot falls within upper and lower cell limits, with only two exceptions out of 40 which however fall within upper and lower tolerance limits. These individual units in the sample falling outside of control limits can be rejected, but this does not affect the decision made with regard to the whole lot.

This example demonstrates the simplicity of the lot template plan; it also shows how easy it is to apply it. There are almost no calculations required. The limits are automatically determined by proper choice of a template with the appropriate quality level. The operating instructions can be stated simply as follows:

- Select a random sample of 40 pieces from the lot to be inspected.
- Decide on a suitable scale and plot the observations and the specification limit(s).
- Select the smallest template that will contain the observations within its cell limits from the series of templates of desired quality level.
- Accept the lot if the acceptance limits are inside the specification limits.

Attention should be paid to the fact that many manufacturing processes are subject to slight shifts in central value and/or an extension of one side of the distribution. Because the probability of finding these pieces near the end of a distribution becomes pretty small in a sample of 40 or 50, from time to time tests should be made with larger samples.

Empirical results give an approximate operating characteristics curve for the probability of calling various skewed distributions normal. The operating characteristic (OC) curves in Fig. 15.5 show that this probability of calling a skewed lot normal when tested with normal template is significantly reduced if instead of a single lot three successive lots are inspected.

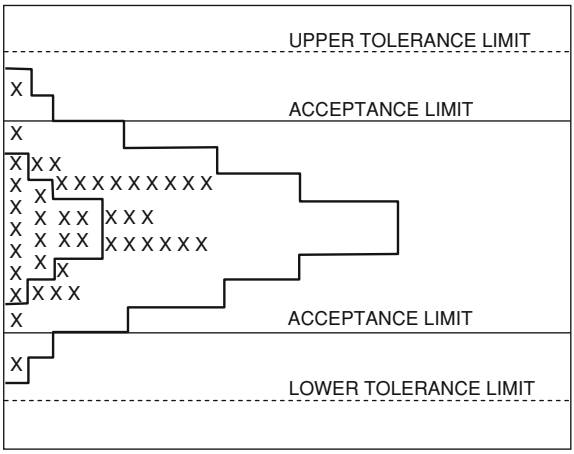

Fig. 15.4 12-cell plot with $\sigma = 2$ cells, acceptance limits at ± 2 s, LTPD = 5.5% and AQL = 0.60%

We have also spoken of the case of using skewed templates. These are typically designed for a skewness of $k = 1$ and $k = 2$ and employed to approximately establish the nature of various non-normal lots when the normal template does not fit.

Acceptance limits for such templates must be determined through careful study. The method is similar to that for normal distribution except that the limits are not equidistant from the center value. Hence, they must be calculated separately for each side of the distribution. In principle, the greater the skewness,

- The lower the probability of fit with the normal acceptance templates and
- The greater the width of the normal templates which do fit the plot.

This is a consequence of the fact that acceptance limits are extended to cover the longer distribution tails. At the same time, the use of wider normal templates is limited by the fact that for significant skewness center cell limits will be exceeded.

Sometimes we may fit a template which is too small and at other times one which is too large. Two different errors which may arise are in the estimate of the variability of standard deviation of the lot and in the estimate of the true mean. A sample of 40–50 observations will not give us these key data with high accuracy.

Another error may arise in rounding off values to the nearest cell interval. This problem is minimized through the use of templates differing from each other by the least amount that is physically possible. Every lot condition will have its own operating characteristics curve, and all of these OC curves will differ slightly from each other.

Among the challenges confronting the quality engineer is that of a production process which has shifted, as well as that of lots that have been mixed. This results in two levels around which the values are clustered. If these two levels are close together, we may find that the distributions overlap or alternatively an increased product spread requires the use of wider templates.

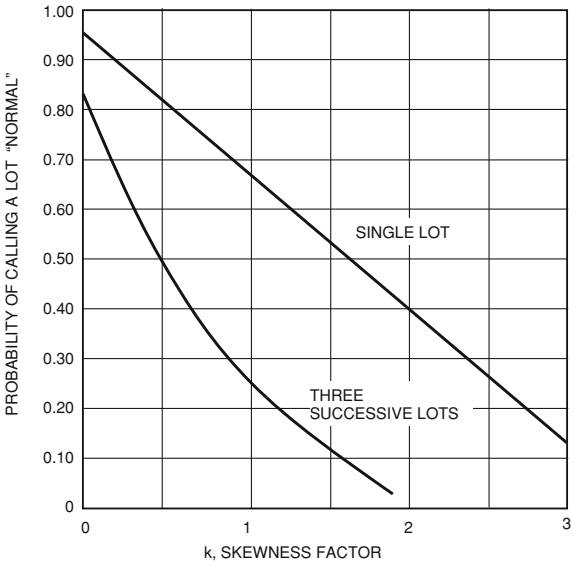

Fig. 15.5 Operating characteristic curve for skewed distributions

When two or more distributions are clearly recognized, it is wise to plot additional observations and test for product acceptance by shifting the template to fit each plot independently. These remarks impact upon the implementation of the lot template method because in industrial production skewness of $k = 2$ or less is not uncommon.

One of the cases that come to mind is electrical characteristics like insulation resistance. It is important however to appreciate that many distributions which are not normal can be adequately treated, for purpose of lot template acceptance, by considering them to be skewed in various amounts.

One of the interesting cases of non-normal conditions is when the manufacturer has screened his product before the lot template method is used. This also happens with procurement. A special condition also arises when the product has overrun one or both of the specification limits, requiring performance of attribute inspection at or near the specification limit to segregate the bad from the good.

Greater protection can also be achieved by attribute inspection of an additional sample, combining these data with the original sample. As the reader is aware from the preceding examples (see also Fig. 15.5) we get greater dependability under double or multiple sampling. Alternatively, we can obtain equivalent protection with smaller additional samples using an acceptance number of zero.

In conclusion, the lot template method will fulfill in an able manner quality control objectives and it is particularly attractive to the smaller firm as well as for bigger companies wanting to acquire the culture of causal inference prior to investing in more complex and more expensive SQC tools and methods. It is however very important that the lot templates plan is not seen as a side issue but is carefully applied and administered under a well planned SQC policy.

15.6 Overcoming Communications Barriers

The reference has been frequently enough made that a common problem confronting companies of all sizes is that quality control information does not circulate freely within the enterprise. Speaking about mistakes in the production process is taboo, and the fact that available manufacturing machinery is not able to meet engineering tolerances is not even allowed to be discussed.

- Management looks the other way and
- Hopes that the problem will disappear by itself.

To a significant degree the background reason is secrecy about a negative evaluation of a production process. But such secrecy is counterproductive because once the problem(s) confronting something important is known, everything follows as an alert management takes measures to upgrade the equipment or the system.

One of the domains where lot templates can contribute to the evaluation of products and processes is that of substantiating the policy of technical auditing brought to the reader's attention in Chap. 3. Rather than being characterized by secrecy, the results of an evaluation should become known to all authorized persons whose work is affected by this technical audit.

A policy of open information channels is rewarding in more than one ways. Operations become so much more efficient *if* the organization takes the initiative to break down the *communications barriers* which exist because of tradition, or of a clash of personalities. Companies do not always appreciate the high cost they are paying by allowing the existence of silos within their borders and even within some of their departments.

Communications barriers are in reality *discipline barriers* as interdepartmental and interpersonal exchange of information breaks down even between teams working within the same engineering office or laboratory. I have seen in my practice more communications barriers being raised when design cycles are longer and/or engineers, within each discipline, are prodded to pursue their own portion of the design independently of work done by their colleagues—leaving integration for later on. By contrast, when design teams are requested to cross disciplinary barriers at the beginning of their design process, they:

- Discover opportunities to optimize their work taking advantage of a wider realm of know-how and
- Eliminate expensive and time-consuming integration errors which have the nasty habit to pop-up only at the last minute.

Such an exchange of inter-disciplinary information will be the more effective if a policy is established (and maintained) that targets the adequacy, completeness and currentness of interfaces between design and manufacturing engineering. Process instructions and industrial instructions relating to a particular design must not only be developed but as well be up to date.

- *If* the lot template method is chosen for SQC.
- *Then* design engineers should not only become familiar with its feedback, but also with the method itself.

This requirement is in direct application of the principle that a sound quality assurance program must provide complete coverage of all information elements necessary to design and produce an article. The double feedback I am suggesting is as well key to complete compliance with contract requirements for proposing, approving, and effecting of engineering changes.

Further still, *if* the company uses contractors *then* what has been stated about the free flow of information inside the organization—in product quality, specs, drawings, costs, and other elements—should include also the contractor's own departments and laboratories. The contractor must be responsible for assuring that this is also true for all supplies and services procured from his suppliers (subcontractors).

Controlled conditions include documented work instructions, adequate skills, first class production equipment, and any special working environment that might be needed. Information about all these issues, on the supplier's behalf, should be an integral part of deciding on acceptable or unacceptable partnership in the sense of a supply chain.

The supplier(s) methods of measuring, monitoring and inspecting should be examined and their suitability demonstrated with a reasonable degree of evidence. The same is true of the supplier(s) statistical methods employed in production and purchasing, including statistical tests, analysis, and quality control procedures.

This underlines the need to break down the communications barriers both in-house and with suppliers. The latter has been done since the late 1990s with enterprise resource planning (ERP) as far as factory scheduling is concerned. The focus of this discussion, however, is not scheduling but wide ranging information from R&D to manufacturing and testing,

- Leading to an information grid and
- Helping management decide on quality versus cost for each produced item.

Evidently, such a system has to be tuned as it requires the development of both qualitative and quantitative criteria of quality assurance, effectively applied in the company's home country, its foreign subsidiaries, and all of the suppliers which it uses in the global market. Wider dissemination of information keeps other projects and departments informed about actual developments and how ongoing programs are progressing toward meeting their goals. I have seen many cases where this policy steered subsequent engineering efforts by revealing the areas needing the most attention, and thereby avoiding that each department had to rediscover the wheel.

If information cannot flow freely within the organization because of mountains of resistance, the only alternative left is the creation of a strong central authority invested with the power to crack the nut of secrecy. Prior to suggesting to the board of companies where I was a consultant in the creation of such a central

authority I tried to open the way by conviction, because I considered that flat orders are a regrettable approach inasmuch as they risk to generate a permanent animosity between departments and projects. When things take this turn, it is not technology and product quality but company politics which take the high ground.

Philosophically speaking, when the politicking which can be found in all enterprises and in nearly all families is discounted, it is not quality in the large and quality in the small that is most interesting. It is what you can accomplish with their results.

Reference

1. Wooley J, Peters G, Gerald R. Ford; Remarks at the U.S.S. Nimitz commissioning ceremony in Norfolk, Virginia. http://www.nimitz-museum.org/nimitzbio.htm

Index

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering,
DOI: 10.1007/978-1-4471-2966-0, © Springer-Verlag London 2013