1 Introduction to Statistics

1.1 Definitions of Statistics

1. Systematic evaluation and methodological numeric analysis of parametric data collected on a subject or population are called statistics.
2. The conclusive study of statistical methods and principles employed to understand the outcome of a research project is also termed as statistics.
3. Statistics means quantitative data analysis by statistical methods to elucidate the validity and accuracy of a scientific procedure or study.
4. Statistics could be defined as quantitative data affected to a certain extent by multiplicity of causes and evaluated through statistical methods.

1.2 Origin and Development of Statistics

Erstwhile kings used to collect information regarding population and wealth of their people. Collected data was analyzed to plan development of the state and to finance war. Statistics in those days was known as "science of kings." Later the data of diverse nature were obtained for general uses of the government. Students of game of chances also developed certain methods of statistical analysis. Biology and insurance as well as other natural sciences are bright fields for application of statistical methods.

1.3 Concept of Probability

The concept of probability is used in day-to-day life which stands for the probability of occurring or non-occurring of events. The notion of probability is used in social sciences, statistics, economics, industry, business, and engineering. The probability

C. S. Rayat, *Statistical Methods in Medical Research*,
https://doi.org/10.1007/978-981-13-0827-7_1

is an element of uncertainty about the happening of an event. The following statements help in understanding the concept of probability:

1. There may be rain tomorrow.
2. Final examination may take place in the month of July.
3. The chance of winning a lottery is less than 1%.

The examples cited above are uncertain about the associated events. So, these statements are just conjectures. The concept of probability was very first used by Italian mathematician Eardan in his book entitled *The Book on Games of Chance* in 1963. The foundation of the "mathematical theory" of probability was laid by French mathematicians Blaise Pascal and Pierre de Fermat. Afterwards the Swiss mathematician James Bernoulli (1654–1705) contributed to the theory of probability; however, his concepts came to light after his death. Other pioneers associated with the probability are de Moivre (1667–1754), Thomas Bayes (1702–1761), Markov (1856–1922), and Kolmogorov.

The probability has a great role in our day-to-day life. Our personal life, social life, academic life, and even business life are deeply associated with probability. Managerial decisions in production and business are analyzed in the light of theories of probability to calculate risks and uncertainties.

1.4 Definition of Probability

The word probability denotes the likelihood of occurring of an event. It is an intelligent guess regarding happening of an event. The probability of occurring of an event could range between "0" and "1." If the event does not happen, the probability would be "0" (zero). On the other hand, when the event happens, the probability would become "1" (one). As per concept the event can happen in M ways or cannot happen in N ways. So, the total possibilities of probabilities are $M + N$, and the probability of happening can be calculated as

$$P = \frac{M}{M+N} = \frac{\text{Number of favorable cases}}{\text{Total number of likely events}}$$

Definition of Events The observed results of identical experiments are called events. An event may be elementary or composite. Events are denoted by capital letters A, B, C, etc. Each event is classified into success (p) or failure (q). Now, if success is zero (0), failure will have the value as "one" (1) or the other way. As per the rule of probability

$$
\begin{aligned}
p + q &= 1 \\
p &= 1 - q \\
q &= 1 - p
\end{aligned}
$$

1.5 Types of Events

1. *Mutually exclusive events*: All events, happening independent of the other, are called "mutually exclusive events." As in the case of tossing a coin, the head and tail cannot occur simultaneously. Only head or tail can occur. So, the occurrence of one event would exclude the occurrence of the other.
2. *Equally likely events*: Events are said to be "equally likely" if all the cases have an equal chance to occur. As we draw a playing card from a pack of cards, every card out of 52 cards will have the equal chance of being getting drawn. Similarly when a dice is thrown, 1, 2, 3, 4, 5, or 6 equally has a likely chance to occur. So, such events are called "equally likely events."
3. *Exhaustive events*: When all the cases of random experiment are included in the study, the events are called "exhaustive events." Possible outcomes in a dice are 1, 2, 3, 4, 5, or 6, and any of these may appear on the top.
4. *Simple and compound events*: "Simple events" are based on the natural law and without any calculations. An event may occur or may not occur. In "compound events" joint possibility of two or more events is considered. For example, if we have a bag containing five white, six red, and seven green balls, the possibility of drawing two balls at a time may be both white, both red, and both green. Possibility would also be of white and red, white and green, or red and green.
5. *Independent and dependent events*: The event is said to be independent if happening of this is not under the happening of any other event. For example: Drawing of a "King" from a pack of cards would always have a probability as $\frac{4}{52}$. A single card seen and put in the packet again would have probability as $\frac{1}{52}$.

   If the happening of the second event depends on the happening of the first event, we call it "dependent event." For example, if we try to draw a "King" from the pack of cards, its probability will be $\frac{4}{52}$. If we remove a card from the pack at random and try to draw a "King" again, its probability will be $\frac{4}{51}$. In the second case, if the drawn card was a "King" and we keep that aside and try to draw the second "King," then the probability of second King will be $\frac{3}{51}$.
6. *Complementary events*: Events are said to be "complementary events" if one event complements the other event for the failure or success. For example, in a throwing of "dice," if condition is applied that winning player will win if the score is odd number, it would mean that the player will win only if 1 or 3 or 5 comes at the top during the toss of dice.

1.6 Different Theories of Probability

The concept of probability is based on certain laws which may be modified or amended with passage of time. There are four theories which will help us understand the concept and applications of probability. These theories are:

1. Classical or priori probability
2. Relative theory of probability
3. Subjective approach
4. Axiomatic theory of probability

1.6.1 Classical or Priori Probability

The "classical or priori probability" is the earliest approach, and the same was developed by Laplace who also coined the definition of probability as "the ratio of the favorable case to the number of equally likely cases." If random experiment (A) results in N exclusive and equally likely events, out of which m are the favorable outcomes, then the probability (P) of occurring of favorable event is given by the following formula:

$$P(\mathrm{A}) = \frac{m}{N}$$

Example 1 Suppose you throw a dice. The probability of occurring of number "1" would be denoted as

$$P(1) = \frac{\text{No. of favorable cases}}{\text{Total No. of equally likely cases}} = \frac{1}{6}$$

So, classical probability has the following four properties:

(a) $0 \leq m \leq N$ is divided by N; we will have $0 < \frac{m}{N} < 1$, i.e., $P(\mathrm{A}) \leq 1$.
(b) If "A" is an impossible event, then $m = 0$. So, $P(\mathrm{A}) = \frac{0}{N} = 0$.
(c) If "A" is certain event, then $m = N$. So, $P(\mathrm{A}) = \frac{N}{N} = 1$.
(d) If occurrence of "A" implies to the occurrence of "B" and occurrence of "B" may or may not imply to the occurrence of "A," then $P(\mathrm{A}) \leq P(\mathrm{B})$.

1.6.2 Relative Theory of Probability

The "probability" of an event is personal or empirical according to "relative theory of probability." It is proportionate to time under identical circumstances. The probability of happening of an event is determined based on past experiences. For example, if a teacher is asked to give questions for the coming examinations, then his probability would be based on past examinations. The "relative probability" of an event would be expressed as

$$P = \frac{\text{Relative Frequency}}{\text{No. of cases or items}}$$

Example 2 Suppose 10,000 products are manufactured on a machine and as per past experience 300 products were found to be defective items. In this case the probability of defective (D) items would be

$$P(\mathrm{D}) = \frac{300}{10000} = 0.03$$

1.6.3 Subjective Approach

Personalistic or "subjective concept of probability" measures that an individual has truth of a particular proposition. Frank Ramsey developed the subjective approach and published in his book entitled *The Foundations of Mathematics and Other Logical Essays* in 1926. This theory is used by everybody in day-to-day life for making decisions in business especially in those where one man dominates the show. Subjective concept of probability is also used in war where every personal approach varies due to individual instincts. This is the most flexible approach in comparison to other approaches. This requires careful analysis during its applications.

1.6.4 Axiomatic Theory of Probability

The word axiom represents the common saying in the society. The approach includes both "classical" and "empirical" probability to develop this theory. Russian mathematician Kolmogorov AN introduced this theory in 1993. According to this theory, the probability of an event ranges between "0" and "1," and if the events are mutually exclusive, the probability of occurrence of either event "A" or event "B" would be denoted by

$$P(\mathrm{A\ or\ B}) = P(\mathrm{A}) + P(\mathrm{B})$$

1.7 Uses of Probability

Theories of probability have extensive applications in various fields in day-to-day life as enumerated below:

1. Law of statistical regularities.
2. Law of inertia of large numbers. Preparation of various estimates is based on probability of outcome.
3. Various parametric and non-parametric tests like *Z*-test, *t*-test, *F*-test, etc. are based on "theory of probability."
4. Probability is applied to "theories of games" for managerial decisions. The expected values are calculated through probability.
5. Sales managers and production managers apply various approaches of probability to take economic decisions in various situations of risks and market uncertainty.
6. The theory of "subjective probability" is always preferred if it is not possible to calculate each expectation.
7. Theoretical aspects of probability are applied to contest the practical significance of a project, experiment, game, or business.

1.8 Theorems of Probability

Basically, there are two theorems of probability: (1) addition theorem and (2) multiplication theorem.

Addition Theorem

It states that when two events are exclusive, the probability of occurring of event "A" or event "B" would be the sum of individual probability of each event as depicted below:

$$P(\mathrm{A\ \ or\ \ B}) = P(\mathrm{A}) + P(\mathrm{B})$$

Example 3 Suppose a card is drawn from a pack of cards. The probability of occurring of a King (K) or Queen (Q) would be

$$P(\mathrm{K\ \ or\ \ Q}) = P(\mathrm{K}) + P(\mathrm{Q}) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}$$

It denotes that the King and Queen are mutually exclusive events, individuals, words, or outcomes. Mutually exclusive events have been shown in Fig. 1.1.

If the events are not mutually exclusive and the two events would overlap each other as depicted in the Fig. 1.2, the "addition theorem" as postulated for the King or Queen would not apply. The theorem is modified as below:

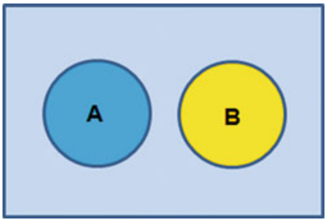

Fig. 1.1 Mutually exclusive events

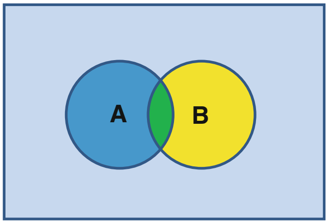

Fig. 1.2 Overlapping events (green area is the overlap)

$$P(\mathrm{A\ \ or\ \ B}) = P(\mathrm{A}) + P(\mathrm{B}) - P(\mathrm{AB})$$

Example 4 There is a chronic ailment to a patient, and there are 75% (¾) chances that doctor "A" can treat it and 80% (⅘) that doctor "B" can treat it. What will be the probability of being cured if the patient gets treatment from doctor "A" and "B" simultaneously?

Doctor "A" treats the patient: $P(\mathrm{A}) = ¾$
Doctor "B" treats the patient: $P(\mathrm{B}) = ⅘$

Probability of A and B collectively:

$$\begin{aligned} P(\mathrm{A\ \ or\ \ B}) &= P(\mathrm{A}) + P(\mathrm{B}) - P(\mathrm{AB}) \\ &= \frac{3}{4} + \frac{4}{5} - \frac{3}{4} \times \frac{4}{5} \\ &= \frac{3}{4} + \frac{4}{5} - \frac{3}{5} \\ &= \frac{15 + 16 - 12}{20} = \frac{19}{20} = 95\% \end{aligned}$$

So, there will be 95% chances that the patient gets treated.

If three events overlap each other as depicted in Fig. 1.3, then their probability would depend on the following theorem:

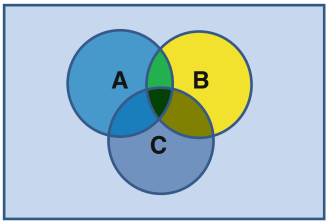

Fig. 1.3 Overlapping events (dark green area is the overlap of three events)

$$P(\mathrm{A\ \ or\ \ B\ \ or\ \ C}) = P(\mathrm{A}) + P(\mathrm{B}) + P(\mathrm{C}) - P(\mathrm{AB}) - P(\mathrm{AC}) - P(\mathrm{BC}) - P(\mathrm{ABC})$$

Multiplication Theorem

The "multiplication theorem" states that if two events are independent, the occurrence of event "A" and event "B" would be the product of individual probabilities of both the events.

$$P(\mathrm{A\&B}) = P(\mathrm{A}) \times P(\mathrm{B})$$

If three independent events are there, then the probability of occurrence of events "A," "B," and "C" would be the product of probabilities of these three events.

$$P(\mathrm{A\&B\&C}) = P(\mathrm{A}) \times P(\mathrm{B}) \times P(\mathrm{C})$$

Example 5 Suppose we ask two persons to accomplish a task. The probability that person "A" can accomplish it is ¾, and probability that B can accomplish it is ⅔. What will be the probability if both the "A" and "B" are assigned the task?

$$\begin{aligned} P(\mathrm{A\&B}) &= P(\mathrm{A}) \times P(\mathrm{B}) \\ &= \frac{3}{4} \times \frac{2}{3} = \frac{1}{2} \quad (\text{By Positive Approach}) \end{aligned}$$

Now if we apply "negative approach" in the same case:

"A" cannot accomplish the task = 1 − ¾=¼
"B" cannot accomplish the task = 1 − ⅔=⅓
A and B cannot accomplish the task = ¼×⅓=1⁄12
Probability that "A and B" can accomplish the task = 1 − 1⁄12=11⁄12

This shows that probability of accomplishing the task by "positive approach" is ½, whereas under "negative approach," it is 11⁄12. So, a modern statistician would suggest "negative approach" for solving a problem as it has improved probability.

Example 6 Let us find out the outcome of the "positive" and "negative" approach in the case of more than two options. Suppose we ask three persons to accomplish a task. The probability that "A" can accomplish it is ¾, that B can accomplish it is ⅔, and that "C" can accomplish it is ½. What will be the collective probability in this case?

Positive Approach

$$P(\text{A\&B\&C}) = P(\text{A}) \times P(\text{B}) \times P(\text{C})$$
$$= \frac{3}{4} \times \frac{2}{3} \times \frac{1}{2} = \frac{1}{4}$$

Negative Approach

"A" cannot accomplish the task = 1 − ¾=¼
"B" cannot accomplish the task = 1 − ⅔=⅓
"C" cannot accomplish the task = 1 − ½=½

Probability that "A and B and C" cannot accomplish the task = ¼×⅓×½=1/24.
Probability that "A and B and C" can accomplish the task = 1 − 1/24=23/24.
This shows that "negative approach" is better than "positive approach."

Conditional Probability
The "multiplication theorem" does not hold well when events are dependent. Two events "A" and "B" are said to be dependent when "B" can occur only when "A" occurs.

Suppose the boss says that he would attend your wedding only when there is no rain. His statement is conditional subject to the condition of "no rain." The "probability" (concern) to such event is termed as "conditional probability." When two events "A" and "B" are dependent, the conditional probability of "B" with given condition of "A" will be

$$P(\text{B/A}) = \frac{P(\text{AB})}{P(\text{A})}$$
$$P(\text{B/A}) = P(\text{B}) \times P(\text{A/B})$$
$$P(\text{A/B}) = P(\text{A}) \times P(\text{B/A})$$

For three events:

$$P(\text{ABC}) = P(\text{A}) \times P(\text{B/A}) \times P(\text{C/AB})$$

Bayes' Theorem
If we do not know the concept of probability, then any task can be accomplished by daily inspection as per new inputs. Bayes' theorem is based on this concept of "revisiting probability" when new information is available. Reverend Thomas

Bayes, a British mathematician, developed the idea of "revisiting probability," and the same was published in 1763. This theorem is applied to ascertain the probability of event "B" when an associated event "A" had occurred. This is just like revising the probabilities based on additional information when the other event has already happened. Accordingly, the "posterior probability" of event "A" for a particular result of event "B" could be computed as

$$P(\mathrm{A/B}) = \frac{P(\mathrm{A})P(\mathrm{B}\ )}{P\ (\mathrm{A})P(\mathrm{B}) + P(\mathrm{A})P(\mathrm{B}) + P(\mathrm{A})P(\mathrm{B})}$$
$$= \frac{\text{Individual Joint Probability of X, Y or Z}}{\text{Sum of Joint Probabilities of X, Y and Z}}$$

This formula for "Bayes' theorem" is based on conditional probability. This could be simplified and expressed in the form of a table as illustrated below as solution for Example 7.

Example 7 There are three salesmen in a store. Salesman "X" issues 40% bills daily, and out of these 1% are faulty; salesman "Y" issues 35% bills daily, and out these 2% are faulty; and salesman "Z" issues 25% bills daily, and out of these 3% are faulty. If a faulty bill is chosen at random, what would be the probability that the same was issued by salesman "X"?

Solution

Salesmen: X, Y, and Z
Event A: Issuing of bills
Event B: Faulty bills
Condition: Bill can be faulty (event B) only after it is issued (event A)

Posteriori probabilities have been computed in the Table 1.1.

After computation we found that probability of faulty bill issued by salesman X would be $\frac{0.004}{0.0185} = \frac{40}{185}$ or roughly $\frac{1}{5}$.

Table 1.1 Posteriori probabilities

1	2	3	4	5
Salesmen	Priori probability P(A)	Conditional probability event B given event A P(B/A)	Joint probability (2) × (3)	Posteriori probability $\frac{(4)}{P(\mathrm{B})}$
X	0.40	0.01	0.004	$\frac{0.004}{0.0185} = 0.216$
Y	0.35	0.02	0.007	$\frac{0.007}{0.0185} = 0.378$
Z	0.25	0.03	0.0075	$\frac{0.0075}{0.0185} = 0.405$
Total	**1.00**	**–**	**P(B) 0.0185**	**1.00**

Note Similar situation can be there in a hospital as illustrated below with the help of Example 8.

Example 8 Suppose renal transplant surgeries are done by three surgeons daily in a hospital. Surgeon "X" does 40% surgeries in a month, and out of these 1% of patients die within 3 months; surgeon "Y" does 35% surgeries in a month, and out these 2% of patients die within 3 months; and surgeon "Z" does 25% surgeries in a month, and out of these 3% patients die within 3 months. If at random a patient dies within 3 months after renal transplant surgery at this hospital, what would be the probability of surgeons X, Y, and Z individually that the patient may have been operated by him?

Solution

Surgeons: X, Y, and Z
Event A: Renal transplant surgery
Event B: Death of patient within 3 months
Condition: Death can occur (event B) only after transplant surgery (event A)

Posteriori probabilities have been computed in the Table 1.2.

After computation we found that probability of postoperative death could be linked to surgeons X, Y, and Z, respectively, as listed below:

1. Surgeon $X = \frac{0.004}{0.0185} = \frac{40}{185} = 21.6\%$
2. Surgeon $Y = \frac{0.007}{0.0185} = \frac{70}{185} = 37.8\%$
3. Surgeon $Z = \frac{0.0075}{0.0185} = \frac{75}{185} = 40.5\%$

Mathematical Expectations

We are always inclined to monetary gain or loss in daily life, may it be availing medical facilities or marketing some products. Mathematical expectations could be used to work out expected value of an even in different possibilities. Suppose "A" is a random variable which could occupy one value in A_1, A_2, A_3, … A_n with

Table 1.2 Posteriori probabilities

1	2	3	4	5
Surgeon	Priori probability $P(A)$	Conditional probability event B given event A $P(B/A)$	Joint probability (2) × (3)	Posteriori I probability $\frac{(4)}{P(B)}$
X	0.40	0.01	0.004	$\frac{0.004}{0.0185} = 0.216$
Y	0.35	0.02	0.007	$\frac{0.007}{0.0185} = 0.378$
Z	0.25	0.03	0.0075	$\frac{0.0075}{0.0185} = 0.405$
Total	**1.00**	–	***P*(B) 0.0185**	**1.00**

corresponding probabilities of $P_1, P_2, P_3, \ldots P_n$. At this the mathematical expectation of "A" would be denoted by

$$E(A) = P_1A_1 + P_2A_2 + P_3A_3 \ldots P_nA_n$$

So, the expected value would be the sum of the products $P_1A_1 + P_2A_2 + P_3A_3 \ldots P_nA_n$.

Following points are kept in mind in mathematical expectations:

1. If various possibilities of the event are positive, then we calculate the sum of every expectation as $E(A) = P_1A_1 + P_2A_2 + P_3A_3 \ldots P_nA_n$.
2. If expected possibility carries positive as well as some negative expectations, then net difference of gain and loss is considered as given below:

$$E(A) = P_1A_1 - P_2A_2 + P_3A_3 \text{ etc.}$$

3. Mathematical expectations are used to evaluate different possibilities under risks involved and select the comparatively better options. Suppose we want to set up a corporate hospital or diagnostic center. We would focus on the beneficial option between the two.

Let us solve an example to learn more about "mathematical expectation":

Example 9 Suppose a private hospital earns Rs. 80,000-00 daily on an average during rainy season and Rs. 95,000-00 daily on an average during rest of the year. What will be the expected earning daily if chances of rainy season are 20% in a year?

Solution

$$\begin{aligned} P_1 &= 20\% = 0.20 \\ A_1 &= 80,000\text{-}00 \\ P_2 &= 80\% = 0.80 \\ A_2 &= 95,000\text{-}00 \end{aligned}$$

$$\begin{aligned} E(A) &= P_1A_1 + P_2A_2 \\ E(A) &= 0.20 \times 80,000 + 0.80 \times 95,000 \\ &= 16,000 + 76,000 = 92,000\text{-}00 \\ &= \textbf{Rs.92, 000-00 Ans}. \end{aligned}$$