



# Statistical Editor's Practical Advice for Data Analysis

# 13

Fikri M. Abu-Zidan

## 13.1 Introduction

Working as a Statistical Editor for international refereed journals for the last 20 years, I find that the majority of serious statistical errors are easy to avoid if the authors take care to follow very basic principles. I hope that this can be achieved if attention is taken when reading this short chapter. I will mainly address the basic statistical analysis when performed by young researchers. I advise readers who are not experts in statistics and want to perform advanced statistical methods like logistic regression, mixed linear models, or general linear models to consult and follow the advice of an experienced statistician. These models require specific assumptions that have to be fulfilled to be reliable [1]. Nevertheless, the majority of basic univariate analysis can be performed with confidence following the recommendations given in this chapter.

### Learning Objectives

- Understand the importance of building the analysis based on the research question.
- Simplify the theoretical background to justify the selection of the analysis.
- Enable the reader to define the rules in which he/she can select the proper statistical method.
- Have a practical map that can direct the analysis process.
- Enforce the learning process through practical applications.

---

F. M. Abu-Zidan (✉)

The Research Office, College of Medicine and Health, Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates

**Table 13.1** Basic questions to be answered before starting the analysis

Sequence	Question
Question 1	What is the objective of the analysis?
Question 2	What is the type of data?
Question 3	Are the data normally distributed?
Question 4	How many groups are compared? (two or more than two)
Question 5	What is the number of subjects in each group?
Question 6	Are the compared data related or unrelated?

There are six questions that have to be answered in sequence before starting the analysis. These are shown in Table 13.1. If answered properly, I hope that the correct statistical methods will be selected. We will go through these questions in sequence.

## 13.2 What Is the Objective of the Analysis?

Statistics is only a tool to summarize and compare data in an informative way. It is essential to define the research question and the objectives of the analysis before even starting it. This is more important when analyzing data retrieved from retrospective studies or large clinical registries. Statistics cannot salvage an inadequate research question or poorly designed study.

Simple descriptive statistics can sometimes be sufficient in high-quality research projects. Collaborators who approach me to perform an advanced statistical analysis get occasionally surprised to see that I used simple descriptive statistics instead of comparative statistical methods because that could address the aim of the study [2]. Statistics is simply a tool to answer the research question, not an aim by itself. Furthermore, the quality of the analysis will depend on the quality of the data. Never start any statistical analysis before getting assured that the data is of good quality and properly coded. If the objective is well defined, the data is accurate, well-understood, and properly coded, you will be surprised to see how the statistical analysis is easy, smooth, and straightforward. The results should then be accepted regardless of the outcome. I personally aim to perform the statistical analysis only once and accept its results even if they were negative.

Unfortunately, it is a common practice that some researchers perform repeated subgroup analysis, fishing for a significant  $p$  value and then retrospectively define the research question to fit the data after the analysis. This is usually difficult to detect. It is erroneous, non-professional, and may even be a research misconduct if not explicitly mentioned in the methods. A clear example for that is the interim analysis of randomized controlled trials, if not declared, which should be transparent as part of the research protocol. It is more difficult in retrospective studies to know whether the results were hypothesis-driven with

**Table 13.2** Mechanism of injury of hospitalized patients involved with road traffic collisions during the pre-COVID-19 and COVID-19 periods, Al-Ain City, United Arab Emirates

Variable	Pre-COVID period <i>n</i> = 750	COVID period <i>n</i> = 499	<i>P</i> value
Mechanism of injury			<0.0001
Motor vehicle collision	540 (72)	302 (60.5)	
Motorcycle	84 (11.2)	116 (23.3)	
Bicycle	42 (5.6)	35 (7)	
Pedestrian	84 (11.2)	46 (9.2)	

This table was reproduced and modified from the study of Yasin et al. [4], which is distributed under the terms of the Creative Commons Attribution 4.0 International License

a clear research question to be answered or whether they stemmed from fishing for a *p* value [3]. This will depend on the conscious and integrity of the researcher.

Table 13.2 gives an example of how defining the research question clearly makes the statistical analysis focused. It shows the mechanism of road traffic collisions in Al-Ain City, United Arab Emirates, before and after the COVID-19 pandemic in one of our recently published papers [4]. The analysis in this scenario will depend on the research question. If the question is: “Is there difference in the mechanism of injury of road traffic collisions before and during the COVID-19 Pandemic?” then Pearson’s Chi-Square test using a  $4 \times 2$  table should be used. This will produce only a single *p* value. The subgroup analysis comparing each mechanism alone between the two groups will increase the chance of getting significance by multiple testing. This will include four comparisons, each with a type I error of 5% of finding statistical significance by chance. Multiple testing can be done as post hoc analysis to explain the significance but not to prove it. If the overall analysis was not significant, then the post hoc analysis should not be performed.

Understandably, the probability of finding statistical significance by chance increases with each additional subgroup analysis [3]. Bonferroni correction can be used to protect against this error by defining the proper *p* value to be 0.05 divided by the number of subgroup pair comparisons [5].

### 13.3 What Is the Type of Data?

The second step is to thoroughly understand the nature and type of the studied variables (Table 13.3). *Categorical (nominal) data* (like eye color or race) do not have an ordered nature nor a measurement of distance between different categories. Even if categorical data are numbered during statistical analysis, these numbers are artificial and just represent the category [6]. *Binary data* is a special type of categorical data that has only two possible options. These are mutually exclusive where one option implies the negation of the other (like dead and alive). If one option is given the probability value of 1 (occurring), the other will be given the probability value

**Table 13.3** Types of data

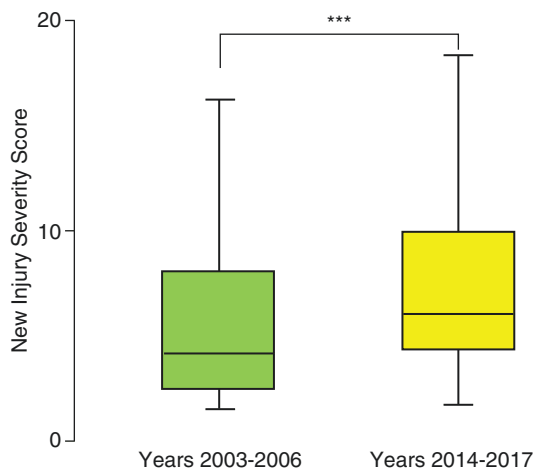
Type of data	Example
Categorical	Eye color
Binomial	Sex
Ordinal	Likert scale
Interval	Number of students
Continuous	Weight

of 0 (not occurring). This makes it possible to perform logistic regression analysis for binary dependent outcome variables. *Ordinal data* has an order of ranks in its nature (like the Likert type questionnaire including very poor, poor, good, very good, and excellent). These can be ordered from 1 to 5. Ordinal data have an ordered nature of three or more levels. Nevertheless, the distances between these levels are not equal. The Anatomical Injury Scale (1–5), Injury Severity Score (1–75), and Glasgow Coma Scale (3–15) are examples of ordinal data. Ordinal data should be presented as median (range or interquartile range (IQR)). *Interval (discrete)* data are real whole numbers (like number of students in a college or number of road traffic collisions). They do not have decimal places. *Continuous data* are numerical or quantitative data that can take any value (like level of serum albumin, height, or stroke volume) and can take decimal places [6].

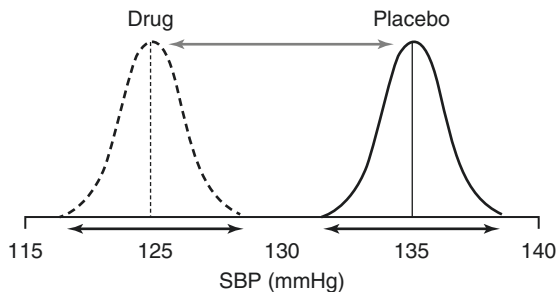
Two common mistakes in statistical analysis are considering ordinal data as continuous data, or changing the continuous data to ordinal data or categorical data in the research protocols. An example is changing the Glasgow Coma Scale to mild, moderate, and severe head injuries. Doing so will weaken the nature and strength of the analysis. It is advised to collect the actual ordinal or continuous data in the research protocols. It is always possible to change the ordinal or continuous data to categorical data during the analysis if needed but not the opposite.

### 13.4 Are the Data Normally Distributed?

It is essential to check for normality of continuous data before the analysis. This can be done by looking into the histograms [6]. Normal distribution should have a bell shape. This is important for deciding the form in which the data will be presented. If the data has a normal distribution, then it can be presented as mean (standard deviation/standard error of the mean) because the mean is the proper point-estimate. If the data are ordinal or do not have a normal distribution, then the median (interquartile range (IQR)) is the proper point-estimate as it lies in the middle of the data. Figure 13.1, which is in one of our recently published papers [7], highlights this point. It compares the New Injury Severity Score (NISS) of two independent groups. Since the data are ordinal, data were presented as box-and-whisker plot. The box represents the 25th to the 75th percentile IQR. Kindly note that the horizontal line within each box, which represents the median, is not in the middle, indicating that the data are not normal and skewed to the right.



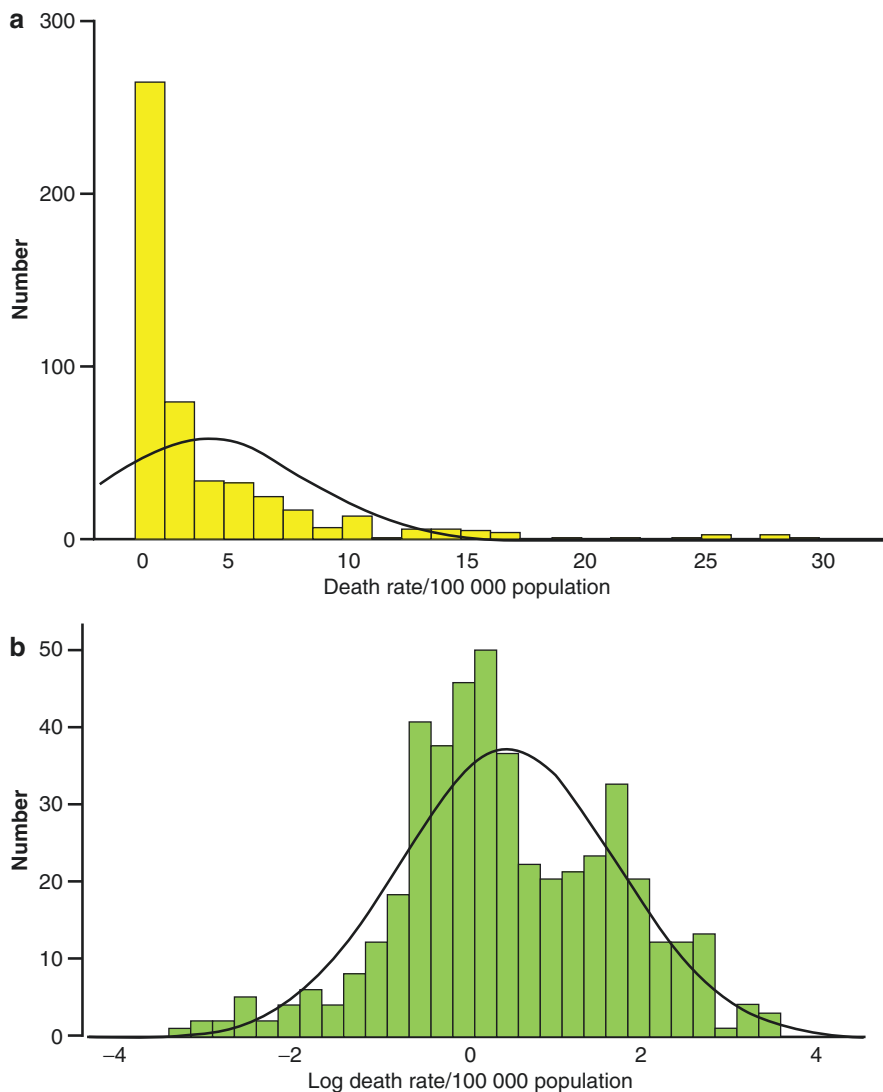
**Fig. 13.1** Box-and-whisker plot of New Injury Severity Score (NISS) for hospitalized trauma patients during the period 2003–2006 ( $n = 2573$ ) and 10 years later ( $n = 3519$ ) during the period 2014–2017, Al-Ain Hospital, Al-Ain, United Arab Emirates. The box represents the 25th to the 75th percentile IQR. The horizontal line within each box represents the median. \*\*\* $p < 0.0001$ , Mann–Whitney  $U$  test. (Reproduced from the study of Alao et al. [7]), which is distributed under the terms of the Creative Commons Attribution 4.0 International License)



**Fig. 13.2** A theoretical example testing a new anti-hypertensive drug. Hypertensive patients were randomized into two groups to receive the drug or a placebo. The data of both groups have a normal distribution and the variance of both groups (the black arrows) is equal. The difference between the means is 10 mmHg (gray arrow). The proper statistical test to use in this situation is the unpaired- $t$  test (student's  $t$  test)

Comparing the continuous data of two groups using parametric methods requires two assumptions: (1) data should have a normal distribution, (2) data should have the same variability. Figure 13.2 is a theoretical example of testing a new anti-hypertensive drug. Hypertensive patients were randomized into two groups to receive the drug or a placebo, each having a sample of 200 subjects. Notice that the data of both groups have a normal distribution and the variance of both groups (the black arrows) is equal. The difference between the means is 10 mmHg (gray arrow). The proper statistical test to use in this situation is the unpaired- $t$  test (student's  $t$  test).

The histogram can demonstrate whether the continuous data is skewed. If the data do not have a normal distribution, then there are two solutions: (1) change the data to normal distribution and then perform the analysis using a parametric method, define the mean of the new data, and then back transform it for reporting or (2) use non-parametric methods. As an example, Fig. 13.3 is retrieved from our recently published paper on the global data of motorcycle related death rates [8]. Kindly observe that the



**Fig. 13.3** Global data of motorcycle related death rates (a) and its log transformation (b) (crude data are from the study of Yasin et al. [8]), which is distributed under the terms of the Creative Commons Attribution 4.0 International License

data of death rates are skewed to the right (Fig. 13.3a) with a skewness value of 3.1 and having a wide peak (kurtosis) of 11.6. The normal values of both Skewness and kurtosis should be between  $-1$  and  $1$  [6]. Log transformation of the data (Fig. 13.3b) has a normal distribution with skewness of  $-0.05$  and kurtosis of  $0.013$ . Accordingly, the log transformed data was used as the outcome variable in the mixed linear model. The outcome variable of mixed linear model should have a normal distribution.

---

### 13.5 How Many Groups Are Compared?

This question looks easy to answer but is sometimes tricky. We need to decide whether the data represent one group, two groups, or more than two groups in order to define the proper statistical method to be used. You should be careful differentiating between studied groups of patients and groups of data. You may measure a variable in one group of patients, give the same group a medication, and then measure the variable again after giving the medication. If the values of the variable are compared before and after the medication, these are two dependent groups of data although they were measured in the same group of patients.

---

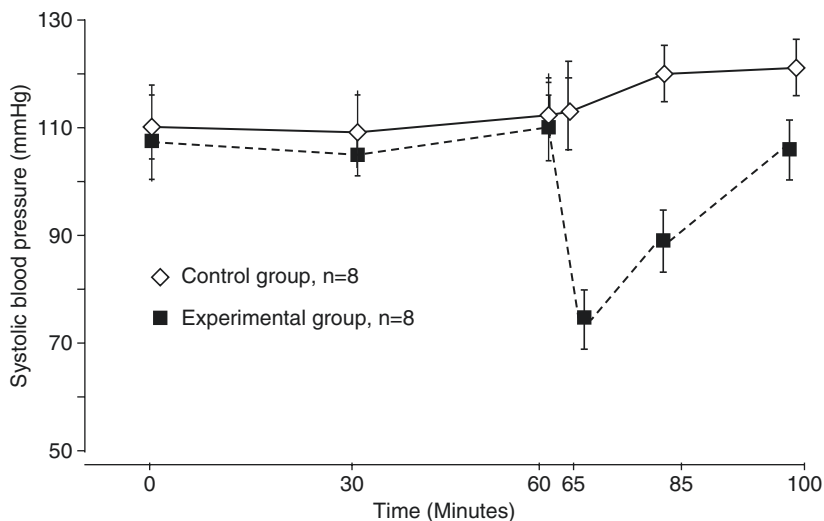
### 13.6 What Is the Number of Subjects in Each Group?

Defining the size of subjects in each group is important to define the statistical methods of analysis. If the number of subjects is less than 20 in each group, it is advised to use non-parametric methods. Non-parametric methods compare the ranks, do not need a normal distribution, are useful in small samples, are more strict than parametric methods, and will not accept significance easily. One approach is to use non-parametric methods all the time, which I practice. There is a risk of missing statistical significance with this approach if parametric methods are not used in normal distributed data (type II statistical error). This may be important when trying to prove harm but not benefit. Kindly note that a significant  $p$  value in comparisons and correlations can be achieved when the sample size is very large. This may not translate to a clinically significant finding as the correlation may be weak or the effect size is small.

---

### 13.7 Are the Compared Data Related or Unrelated?

This question is very important and needs deep thinking to address. When comparing the weight of patients who died and those who survived following road traffic collisions, it is clear that these two groups are completely independent because each subject will be only in one group. In comparison, if we study the effect of bypass surgery on the weight of morbid obese patients, we will measure the weight before surgery and after surgery which enables us to measure weight change in each patient. Weight before surgery and after surgery are related (dependent) data. In the first example the two groups are independent and the weight of the two groups can



**Fig. 13.4** A theoretical animal experiment comparing two groups of anesthetized rats, each consists of eight rats. One group is a control laparotomy group (white diamonds), while the other group is a bowel-ischemia reperfusion group (black square). The data presented are the mean systolic pressure (standard error of the mean) of each group over time. The proper method of statistical analysis in this situation is the repeated measurement analysis of variance

be compared using unpaired *t*-test if other assumptions of using this test were met. In the second example the two groups are dependent and the weight of the two groups can be compared using paired *t*-test. The paired *t*-test has the advantage of comparing each subject with itself which standardizes all variation within the subject and makes it easy to find the statistical significance. This analysis can be used in natural pairs like twins or selected matched pairs of patients.

Let us look into another common example. Figure 13.4 shows a theoretical animal experiment over time comparing two groups of anesthetized rats, each consists of eight rats. One group is a control laparotomy group (white diamonds), while the other group is a bowel-ischemia reperfusion group (black square). Systolic blood pressure (SBP) directly dropped following the small bowel reperfusion. Kindly note the relationship between the collected data of SBP. The data within each group are dependent as it is repeatedly measured in the same animal. In contrast, the data between the two groups are independent as each animal is located within a specific group. The proper method of statistical analysis in this situation is the repeated measurement analysis of variance. This analyzes three components: (1) difference within each group, (2) difference between groups, and (3) the interaction between the two groups to evaluate the direction of change. Each of these factors should have only a single reported *p* value [9–11].



### 13.8 Which Test to Use?

Table 13.4 shows the summary of the recommended statistical methods to be used for analyzing the continuous or ordinal data after answering the previous questions. We have now defined the type of data, number of the groups to be compared, number of the subjects within each group, whether the data have a normal distribution or not, and whether the data are related or not. Non-parametric statistical methods are the proper method when the number of the subjects of the groups are small, data do not have a normal distribution, or data are ordinal in nature. Non-parametric methods are advised in these conditions because they compare the ranks of the groups and a normal distribution is not needed [12, 13].

Let us assume that we are comparing the New Injury Severity Score (NISS) of trauma patients who were admitted during the last year in four different trauma centers in our state. Their numbers range between 750 and 1200 patients. The data are ordinal, the groups are independent, the number of the groups are more than 2. Then, the proper test to use is Kruskal–Wallis test. If the analysis was not significant, then we stop at this stage, and accept that the injury severity of the hospitalized trauma patients is the same between these four hospitals. If we find that there was statistical significance between the hospitals then we proceed with comparisons between each two hospitals using Mann–Whitney  $U$  test, just to explain the finding and not prove it, because the overall test will not be able to show that.

Beware that you should always use two tailed tests which indicate that the difference can go in any direction. This is the standard accepted way for comparison. Do not use a one tailed test. I have never used it in my three decades of intense research activities. One tailed test indicates that the difference between the groups can go only in one direction. This should be decided before the analysis is started, clearly mentioned and justified in the methods section, and clearly reported in the results section.

When comparing categorical data of two or more independent groups, then Pearson's Chi-square can be used. Nevertheless, if the sample size of the groups is small (less than 20), any of the cells is 0, or any of the expected cells is less than 5, then Fisher's Exact test should be used. Advanced Statistical packages (like SPSS, SPSS Inc, Chicago, IL, USA) will give a warning and advise which test is to be used. McNemar's test should be used when comparing matched (related) categorical data [14].

**Table 13.4** Selection of statistical tests for comparison of continuous or ordinal data

	Parametric		Non-parametric	
	2 groups	>2 groups	2 groups	>2 groups
Independent	Unpaired $t$ -test	ANOVA with multiple comparison Bonferroni correction	Mann–Whitney $U$ test	Kruskal–Wallis test
Dependent	Paired $t$ -test	Repeated measurement ANOVA	Wilcoxon Signed rank test	Friedman test

## Do and Don't

- Understand the research question and the type of data thoroughly before starting the analysis.
- Define the number of groups to be compared, number of subjects in each group, and the relationship between the groups.
- Use parametric methods only for normally distributed data. Alternatively use non-parametric methods.
- Do not overuse statistics.
- Do not fish for a  $p$  value.
- Ask for help when needed.

## Take Home Messages

- Basic statistics is easy to perform if well understood.
- There are two main types of statistical comparisons: parametric and non-parametric.
- The correct statistical method will be selected by following the roadmap explained in this chapter.

**Conflict of Interest** None declared by the author.

---

## References

1. Sprent P. Statistics in medical research. *Swiss Med Wkly*. 2003;133(39–40):522–9.
2. Shaban S, Cevik AA, Canakci ME, Kuas C, El Zubeir M, Abu-Zidan F. Do senior medical students meet recommended emergency medicine curricula requirements? *BMC Med Educ*. 2018;18:8.
3. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: the perils of multiple testing. *Perspect Clin Res*. 2016;7:106–7.
4. Yasin YJ, Alao DO, Grivna M, Abu-Zidan FM. Impact of the COVID-19 pandemic on road traffic collision injury patterns and severity in Al-Ain City, United Arab Emirates. *World J Emerg Surg*. 2021;16:57.
5. Hassard TH. Analysis of variance. In: Hassard TH, editor. *Understanding biostatistics*. 1st ed. St. Louis, MO: Mosby Year Book; 1991. p. 75–97.
6. Chan YH. Biostatistics 101: data presentation. *Singapore Med J*. 2003;44:280–5.
7. Alao DO, Cevik AA, Eid HO, Jummani Z, Abu-Zidan FM. Trauma system developments reduce mortality in hospitalized trauma patients in Al-Ain City, United Arab Emirates, despite increased severity of injury. *World J Emerg Surg*. 2020;15:49.
8. Yasin YJ, Grivna M, Abu-Zidan FM. Motorized 2-3 wheelers death rates over a decade: a global study. *World J Emerg Surg*. 2022;17:7.
9. Chan YH. Biostatistics 301. Repeated measurement analysis. *Singapore Med J*. 2004;45:354–68.
10. Munro BH. Repeated measures analysis of variance. In: Munro BH, editor. *Statistical methods for health care research*. 4th ed. New York: Lippincott; 2001. p. 201–21.
11. Ludbrook J. Repeated measurements and multiple comparisons in cardiovascular research. *Cardiovasc Res*. 1994;28:303–11.

12. Chan YH. Biostatistics 102: quantitative data—parametric & non-parametric tests. *Singapore Med J.* 2003;44:391–6.
13. Munro BH. Selected nonparametric techniques. In: Munro BH, editor. *Statistical methods for health care research.* 4th ed. New York: Lippincott; 2001. p. 97–121.
14. Chan YH. Biostatistics 103: qualitative data—tests of independence. *Singapore Med J.* 2003;44:498–503.

---

## Further Reading

- Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Walter S. Basic statistics for clinicians: 1. Hypothesis testing. *CMAJ.* 1995;152:27–32.
- Hassard TH, editor. *Understanding biostatistics.* 1st ed. St. Louis, MO: Mosby Year Book; 1991.
- Munro BH, editor. *Statistical methods for health care research.* 4th ed. New York: Lippincott; 2001.
- Siegel S, Catellan NJ Jr. *Nonparametric statistics for behavioral sciences.* 2nd ed. Singapore: McGraw-Hill International Editions; 1988.