REVIEW ARTICLE

Importance of using basic statistics adequately in clinical research☆

CrossMark

Célio Fernando de Sousa Rodrigues[a], Fernando José Camello de Lima[b], Fabiano Timbó Barbosa[b],*

[a] Universidade Federal de Alagoas, Morfologia, Maceió, AL, Brazil
[b] Universidade Federal de Alagoas, Ciência da Saúde, Maceió, AL, Brazil

Abstract
Background and objective: The inadequate use of basic statistics is the main responsible for scientific article misinterpretation. The purpose of this review article was to review some basic statistical topics to alert authors and readers about the importance of basic statistics proper reporting.
Content: A bibliographical and cross-sectional study was carried out, which analyzed publications in books and articles in the following databases: SciELO (Scientific Electronic Library Online) and PubMed (Available from the National Center for Biotechnology Information). Medical research is not free from the risk of false positive and false negative results due to the choice of statistical tests and presence of small sample sizes.
Conclusion: Understanding the correct use of basic statistics leads to fewer errors in reporting the results of studies performed and in the interpretation of their conclusions.
© 2017 Published by Elsevier Editora Ltda. on behalf of Sociedade Brasileira de Anestesiologia. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Importância do uso adequado da estatística básica nas pesquisas clínicas

Resumo
Justificativa e objetivo: O uso inadequado da estatística básica é o maior responsável pelo erro de interpretação dos artigos científicos. O objetivo deste artigo de revisão foi rever alguns tópicos básicos de estatística para alertar autores e leitores sobre a importância do relato adequado da estatística básica.

☆ Study performed at Universidade Federal de Alagoas.
* Corresponding author.
  E-mail: fabianotimbo@yahoo.com.br (F.T. Barbosa).

Conteúdo: Foi feita pesquisa bibliográfica e transversal que analisou publicações em livros, artigos nas bases de dados SciELO (*Scientific Electronic Library Online*) e PubMed, do *National Center for Biotechnology Information*. Pesquisas na área médica não estão livres do risco de resultados falso positivos e falso negativos devido à escolha dos testes estatísticos e à presença de pequenos tamanhos de amostra.

Conclusão: A compreensão acerca do uso adequado da estatística básica propicia menores erros nos relatos dos resultados de estudos executados e na interpretação das suas conclusões.

Introduction

Statistics is a science that uses data analysis to test statistical hypotheses, assess the strength of clinical evidence and thus whether there are true associations or phenomena between groups.[1] Researchers must formulate hypotheses, note the biological phenomena occurring in the population and draw a sample from that population to test their hypotheses. The sample similarity with the original population allows the results of data analysis to be more reliable in the elucidation of hypotheses.[1]

Statistical analysis, which is present in scientific researches and reported in original articles, allows readers, patients and health providers to interpret the information derived from data collected during a research and use it for the benefit of society.[2] Concerns about adequately reporting the results of biomedical researches have been present in the world literature since past decades.[3]

The frequency of the adequate use of statistical tests can be seen in a number of medical fields, such as oncology, radiology, surgery, and anesthesiology.[2,4–6] Consequences can be serious if the scientific content analysis is inadequate, such as false results with unwarranted assumptions and conclusions lacking biological support.[3]

There are several guidelines available for data reporting and statistical measurements that have been published by various authors of scientific articles, indicating which items are important to be used in scientific research reports.[7,8] Despite the availability of such guidelines, errors in the use of statistics in research reports still exist and are due to both the use of basic statistics and advanced statistics, but the greater frequency of errors occurs with the use of basic statistics, contrary to what one might think.[2,9]

The present review is an attempt to make anesthesiologists aware of the various aspects of statistical methods used in clinical research, as well as to attempt through this narrative review to reduce as much as possible the statistical errors that are still committed in basic statistics. The objective of this paper was to review some basic statistical topics to alert scientific research authors and readers about the importance of adequately reporting basic statistical data.

Method

A bibliographic and transversal search of books and scientific articles published in electronic media was carried out in the following databases: SciELO (Scientific Electronic Library Online) and PubMed (National Center for Biotechnology Information). The following meSH terms were used: ''biostatistics'', ''anesthesia'', ''and sample size''.

Literature review

Basic concepts of descriptive statistics

Clinicians should be able to make the best decisions before the patient in their routine practice, and acquiring new knowledge will only be possible if they are able to read and critically analyze articles published in scientific journals. Descriptive statistics is a part of statistics that helps researchers and readers understand the information of data collected through its organization and summarization.[10] Descriptive statistics is the only statistic used in descriptive works and some epidemiological studies.[10] The use of raw data in scientific articles, that is, data as collected in the survey, is uncommon and may impair its interpretation and make reading uninteresting.

Descriptive statistics is used to describe data using numbers or statistical measures that may best represent all data collected during a research. It is considered an initial step for the appropriate choice and use of statistical tests of hypotheses.[11] It is essential to know which statistic is most appropriate for the most different levels of measurement.[12] The most used in published health articles may be seen in Table 1.

Descriptive statistics can be divided into measures of central tendency and dispersion.[13] The first uses a value that represents what is more typical and may be used to represent all other values collected in a research.[13] The second uses a value that reveals how data varies around that value that is more typical.[11] The main measures of central tendency are: mean, mode, and median.[13] The main measures of dispersion are: variance, standard deviation and interquartile range.[11]

Table 1 Summary of the most commonly used descriptive statistics in health publications.[12]

Descriptive statistics			
Shape and normality	Central tendency	Dispersion or variation	Percentile and quartile
Symmetry	Mode	Range	Percentile
Kurtosis	Median	Variance	Interquartile range
	Mean	Standard deviation	

The mean is an important measure because it incorporates the score from every subject in the research study.[12] The required steps for its calculation are: count the total number of cases—referred in statistics as *n*; add up all the scores and divide by the total number of cases.[13] This advantage of the mean is also its problem, as it is affected by high or low discrepant scores that distort the desired information one wishes to transmit about the analyzed data.[12]

Median differs from mean because it is the middle value in distribution when the values are arranged in ascending order.[14] If we take random values, such as 88, 89, 90, 91 and 92, we will have a median of 90.

Mode is the value that occurs most often and does not provide an indication of all the values collected in a research, but rather it expresses the most repeated value.[13] If we take random values, such as 88, 88, 90, 91 and 92, we will have a mode of 88.

Median and quartiles are values that represent the position, in percentage scale, of the values distributed in ascending order. The median represents the 50% position in the distribution scale.[14] To find out where the median position is, simply divide the total score of cases by two.[12] A simple way to find out the numerical value sorting the values in ascending order, gradually eliminating the extreme values, and then identifying the value that remained in the center.[12] This value will be the median. In some cases, all the end values are eliminated and no central value remains. When this occurs, one should average the last two values and thus calculate the central value.[12] The median is not influenced by the discrepant values and should be preferred when they are present.[14] If we take random values, such as 85, 89, 90, 91 and 97, we will have 90 as median.

Measures of central tendency have their applicability. Table 2 shows the indication for the application of each measure. Taking two sets of random values, the first being 88, 89, 90, 91 and 92 and the second 30 + 70 + 90 + 120 + 140, we will have two sets of 90 as the mean. By noting only the mean, one does not perceive the information about the rest of the values and, therefore, it is necessary to use dispersion measures to realize that data from groups are not equal.

Table 2 Applicability of measures of central tendency.[12]

Measures of central tendency			
Characteristics	Mean	Median	Mode
Interval and scalar data	Yes	Yes	Yes
Ordinal data	No	Yes	Yes
Nominal data	No	No	Yes
Distortion with discrepant values	Yes	No	No

Values may be near or far from the mean and this distance from the value to the mean is known as discrepancy.[12] The sum of all discrepancies can be zero, so in order to use these discrepancies it is advisable to square each discrepancy value before using it mathematically.[12] The average square of these values is known as variance.[12] The analyzed variable unit of measurement is also square, so in some cases it is difficult to understand its meaning.[12]

Standard deviation is one of the most commonly used statistical measures to demonstrate data variability.[15] It estimates the degree to which the value of a particular variable deviates from the mean.[12] Mathematically, the square root of the variance is the standard deviation.[12] The unit of measure of a variable remains in its original form.[12]

Total range of a variable is the distance between the highest and lowest values.[12] It is calculated by subtraction between the highest and lowest values from a set of data.[12] The measure does not tell whether the values are evenly distributed, if there are groups of values near each other, or if there are absent groups of values between the collected data.[12]

Interquartile range is a measurement position related to the median.[12] Quartiles represent the 25% and 75% positions in the scale, so that the first quartile represents the value corresponding to the first quartile of the distribution (25% of the values below that position) and the third quartile represents the value corresponding to the third quartile of the distribution (75% of the values above that position).

Dispersion measures have their applicability. Reanalyzing the two sets of previous random values, we can note that for the first set the mean was 90, standard deviation 1.15, and total range 88–92 and for the second set the mean was 90, standard deviation 43.01, and total range 30–140. By the use of dispersion measures, we can note that the sets of values are different. Indication of where each measure may be used is in Table 3.

Mean and standard deviation are best used when data are normally and symmetrically distributed, as well as median and interquartile range for data asymmetrically distributed.[12] One way to identify whether symmetry occurs in data distribution is to create the histogram graph and observe its shape.[12] Graph creation begins with the distribution of the number of cases on the *y*-axis and the level of the analyzed variable on the *x*-axis (Fig. 1). If the shape resembles a bell, there is already a strong indication that data is normally distributed.

Data distribution may also be statistically assessed by comparing the curve formed by the distribution of data collected in a research and the normal curve. Calculation can be done by computer applications such as BioEstat version 5.0, STATA, EpiInfo, and others.

Table 3 Applicability of dispersion measures.[11–13]

	Dispersion measures		
Characteristics	Range	Interquartile range	Standard deviation
Interval and scalar data	Yes	Yes	Yes
Ordinal data	Yes	Yes	No
Sample variability description	Yes	Yes	Yes
Statistical inference participation	No	No	Yes

Basic concepts of inferential statistics

Inferential statistics is used to draw conclusions and make inferences after analyzing data collected in surveys.[13] Inferential statistics include hypothesis tests and estimation to make comparisons and predictions and draw conclusions that will serve populations based on sample data.[1] Statistical inferences may be: bivariate analysis and multivariate analysis.[1] The first analyzes the relationship between a dependent variable and an independent variable. The second analyzes the relationship between a dependent variable and multiple independent variables and verifies the potential confounding effect of the latter on the first.[1]

Statistical inference is only possible after testing the statistical hypotheses.[16] Hypothesis is a numerical statement about an unknown parameter to the researcher.[16] The two statistical hypotheses are: null and alternative hypotheses.[16] Null hypothesis refers to the absence of effect or association.[1] Alternative hypothesis states that there is a difference between at least two populations studied and when positive states that there is difference between the groups analyzed.[16]

Researchers may have two errors when they rely on these two hypotheses to formulate conclusions: type I and type II errors.[1] Type I error refers to a false positive result; that is, rejecting the null hypothesis when in fact it is true.[1] Type II error refers to a false negative result; that is, accepting the null hypothesis when in fact it is false.[1] The probability of making a type I error is known as level of significance or alpha.[1] The acceptable level of significance most used in

the health field is 5%.[1] Statistical hypothesis tests calculate the probability of an event occurring assuming that the null hypothesis is true.[17] This probability is known as p-value.[1] If the p-value is less than the level of significance, the null hypothesis may be rejected and the alternative hypothesis that says there is a difference or association between the analyzed groups may be accepted.[1] This reasoning applies to superiority clinical trials. The most common error among readers is to believe that p-value represents the probability that a null hypothesis is true.[17] Non-inferiority or equivalence clinical trials assess the exact opposite, the logic of interpretation is opposite, as null hypothesis represents the difference between the observed values.

Kurichi et al.[2] performed a research in 2006, analyzing publications in several scientific journals in the field of surgery and demonstrated that Student's t and chi-square tests were the most used for hypotheses. This finding is supported by other researches in other areas of medicine.[4–6]

Student's t-test is a parametric test which compares the mean of two samples.[18] The use of this test requires some conditions[18]: sampled population must have symmetrical distribution, sample variances must be equal or approximate, and samples must be independent.[18] This test statistics may be obtained according to the following steps: calculate the sample means and the respective standard deviations, find the difference between the two sample means, calculate the standard error and divide the value of the difference between means by the standard error value.[19] Once the t-value is found, a table of critical values of the t-statistic must be consulted according to the degrees of
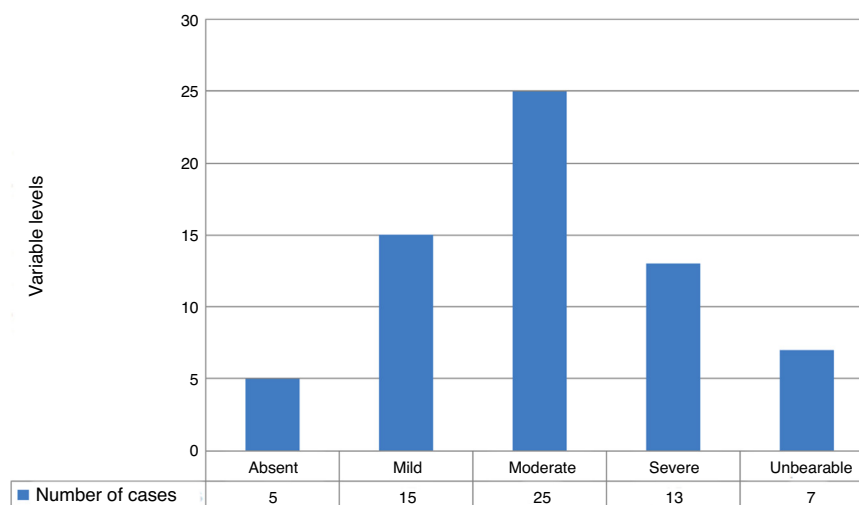

Figure 1 Distribution of number of cases in y-axis and level of the analyzed variable in x-axis.

freedom appropriate to each case.[18] If the *t*-value found is greater than or equal to the tabulated *t*-value, the null hypothesis may be rejected.[18] The statistical *t*-value can also be converted to *p*-value.[19] If *p*-value is less than the level of significance adopted for the study, the null hypothesis should be rejected.[19]

Medical research generally involves more than two groups. The Anova test is used to simultaneously test the equality between more than two groups.[20] The Anova various forms of testing are: one-way for an independent variable, two-way for two independent variables, and repeated measures for patients serving as their own controls.[20] The use of Anova requires some conditions: sample should have symmetrical distribution, samples should be chosen randomly, and homoscedasticity should be evaluated. Variance represents the dispersion of data that will be analyzed. Homoscedasticity represents the variance homogeneity and is a presupposition that must be observed to perform the test.[20]

Chi-square test is a nonparametric test used to answer study questions involving rates, ratios, or frequencies.[21] It does not require data symmetrical distribution.[21] There are two tests: chi-square of independence and goodness of fit.[21] The independence test is the most commonly used and assess the frequency of data from two or more groups.[21] The goodness of fit test is used to compare sample data with data from known populations.[21]

Chi-square test statistics for two samples may be obtained according to the following steps: calculate the sample proportions, find the difference between these two proportions, calculate the general sample proportion that will be used in the standard error calculation, calculate the standard error and divide the value of the difference between the proportions by the standard error value.[19] The null hypothesis may be rejected if the *p*-value is less than the significance level adopted in the study or if the value found is greater than or equal to a tabulated value as is occurs in *t*-test.[19]

The use of non-parametric statistical tests has increased over the years.[2] A study analyzing publications in the field of surgery found that in Archives of Surgery there was an increase from 0% in 1985 to 33% in 2003 and Annals of Surgery from 12% in 1985 to 49% in 2003.[2] Non-parametric methods are used for data with asymmetric distribution or from ordinal and nominal scales.[21] The most common and their indications are: chi-square and Fisher's exact test for proportions or frequencies; Mann–Whitney *U*, Wilcoxon, Kruskal–Wallis and Friedman tests for ordinal data; and Kruskal–Wallis and Friedman for intergroup comparison.[21] Data from small total sample size may be better evaluated with non-parametric tests.[1]

The professional training of physicians generally offers a basic knowledge in statistics, but many are not apt to use this knowledge in the interpretation of the data.[1] The decision of which test to use for each particular situation requires clarification of some points: data measurement scale; number of groups; relationship between participants, that is, whether the groups are independent or related and the researcher's intention to establish a difference or relationship between groups.[22] A hypothetical example would be to evaluate complications in the post-anesthetic recovery unit. The first step is to count the event of interest and

divide it by the total number of patients to find the proportion and, by multiplying that proportion by 100, we have the percentage. Then, the gender difference can be verified by the chi-square test, or the amount of anesthetic used by each patient can be checked and the mean extracted. A general guide for choosing tests is shown in Table 4.

How to calculate the sample size

Statistics are used for comparisons between groups and to make predictions for populations using the sample data, as it is generally not feasible to analyze the data of all members of a population.[1] The hypothesis is formulated by observing the population tested in the sample. An adequate number of participants should be calculated before the study.[23] If the sample size is smaller than necessary, the actual effect analyzed may be neglected by the researcher, and if this size is too large, resource and animal waste will occur if it is an experimental research.[23]

Sample size calculation can be performed through computer applications (apps). There are some free online apps that use the statistical power method. Some examples are: http://www.openepi.com/Menu/OE_Menu.htm; http://www.biomath.info/power/index.htm; http://homepage.stat.uiowa.edu/~rlenth/Power/#Download_to_run_locally; http://statpages.org; http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize; http://tinyurl.com/timbocalculo.

Common errors in anesthesia

Statistical error identification was investigated in the bibliographic material of the Anesthetic Research Society.[24] The categories pointed out in this research were: method presentation or choice of statistical test, variability and probability. The most common errors were: no identification of inferential statistics tests, inadequate presentation of data to allow *p*-value interpretation, and inadequate presentation of mean standard deviation.

Common errors found in anesthesia are[3]: wrong choice of a hypothesis test that disregards data distribution; wrong choice of a hypothesis test that disregards the clinical hypothesis, which leads to type I error during analyzes of significance; use of chi-square when the expected frequency of a cell is <5; use of chi-square without Yates' correction in small sample; use of *t*-test for paired samples in unpaired samples and pair samples to analyze with *t*-test.

Final considerations

Proper use of basic statistics allows the clinician to feel more confident with the study results and thus implant new interventions or drugs in clinical practice.

The main recommendations to minimize errors in scientific reports are[7,8]: describe the research hypothesis; conceptualize the variables used in the research; summarize the variables data using descriptive statistics; describe the methods used in the analysis of each variable and relate the statistical methods used; check data distribution before performing the analyzes and report the test or

Table 4 General guide for choosing statistical tests.[1,20,21]

Hypothesis test	Test indications
Student's *t*	Compare means of two groups whose data have normal distribution Independent samples or related samples
Anova	Compare mean of more than two groups whose data have normal distribution Independent samples or related samples
Chi-square	Analyze nominal data of more than 40 participants regardless of data distribution Independent samples
Fisher exact	Analisar dados nominais de até 40 participantes independentemente da distribuição dos dados Independent samples
Mann–Whitney *U*	Analyze scalar and ordinal data from two groups regardless of data distribution Independent samples
Wilcoxon signed-ranks	Analyze scalar and ordinal data from two groups regardless of data distribution Related samples
Kruskal–Wallis	Analyze scalar and ordinal data from more than two groups regardless of data distribution Independent samples
Kolmogorov–Smirnov	Check if data is from the same population Independent samples

technique used; describe the adjustment methods used for multiple comparisons; describe how discrepant values were treated; describe the level of significance; describe the parameters used to for sample size calculation so that calculations can be repeated; describe the software or statistical package used in the analysis; use mean and standard deviation for data with normal distribution; use median and interquartile range for asymmetrically distributed data; do not replace standard deviation by standard error.

The greatest errors in the interpretation of data from scientific studies are due to the inadequate use of the basic statistics addressed in this narrative review. Health professionals should be able to critically evaluate the results of studies so that information in the literature can positively influence patient care. Understanding the validity of the conclusions favors the findings applicability to patients.

Understanding the proper use of basic statistics provides less error in reporting the results of studies and interpreting their findings.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Windish DM, Diener-West M. A clinician-educator's roadmap to choosing and interpreting statistical tests. J Gen Intern Med. 2006;21:656–60.
2. Kurichi JE, Sonnad SS. Statistical methods in the surgical literature. J Am Coll Surg. 2006;202:476–84.
3. Bajwa SJ. Basics, common errors, and essentials of statistical tools and techniques in anesthesiology research. J Anaesthesiol Clin Pharmacol. 2015;31:547–53.
4. Barbosa FT, Sousa DA. Frequency of the adequate use of statistical tests of hypothesis in original articles published in the Revista Brasileira de Anestesiologia between January 2008 and December 2009. Rev Bras Anestesiol. 2010;60:528–36.
5. Hokanson JA, Luttman DJ, Weiss GB. Frequency and diversity of use of statistical techniques in oncology journals. Cancer Treat Rep. 1986;70:589–94.
6. Goldin J, Zhu W, Sayre JW. A review of the statistical analysis used in papers published in Clinical Radiology and British Journal of Radiology. Clin Radiol. 1996;51:47–50.
7. Altman DG, Gore SM, Gardner MJ, et al. Statistical guidelines for contributors to medical journals. Br Med J (Clin Res Ed). 1983;7:286–93, 1489.
8. Bailar JC, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. Ann Intern Med. 1988;108:266–73.
9. Kim M. Statistical methods in arthritis & rheumatism: current trends. Arthritis Rheum. 2006;54:3741–9.
10. McHugh ML. Descriptive statistics. Part I: level of measurement. J Spec Pediatr Nurs. 2003;8:35–47.
11. Twycross A, Shields L. Statistics made simple. Part 2. Standard deviation, variance and range. Paediatr Nurs. 2004;16:24.
12. McHugh ML. Descriptive statistics. Part II: most commonly used descriptive statistics. J Spec Pediatr Nurs. 2003;8:111–6.
13. Twycross A, Shields L. Statistics made simple. Part 1: mean, medians, and mode. Paediatr Nurs. 2004;16:32.
14. Gaddis GM, Gaddis ML. Introduction to biostatistics. Part 2: descriptive statistics. Ann Emer Med. 1990;19:309–15.
15. Clegg F. Introduction to statistics I: descriptive statistics. Br J Hospital Med. 1987;37:356–67.
16. Gaddis GM, Gaddis ML. Introduction to biostatistics. Part 3, sensitivity, specificity, predictive value, and hypothesis testing. Ann Emerg Med. 1990;19:591–7.
17. Biau DJ, Jolles BM, Porcher R. P value and the theory of hypothesis testing: an explanation for new researchers. Clin Orthop Relat Res. 2010;468:855–62.

18. Gaddis GM, Gaddis ML. Introduction to biostatistics. Part 4: statistical inference techniques in hypothesis testing. Ann Emerg Med. 1990;19:820–5.

19. Rumsey D. Testes de hipóteses mais utilizados. In: Rumsey D, editor. Estatística para Leigos. Rio de Janeiro: Alta Books; 2010. p. 237–47.

20. Gaddis ML. Statistical methodology: IV. Analysis of variance, analysis of covariance, and multivariate analysis of variance. Acad Emerg Med. 1998;5:258–65.

21. Gaddis GM, Gaddis ML. Introduction to biostatistics. Part 5: statistical inference techniques for hypothesis testing with nonparametric data. Ann Emerg Med. 1990;19:1054–9.

22. Twycross A, Shields L. Statistics made simple. Part 4: choosing the right test. Paediatr Nurs. 2004;16:24.

23. Charan J, Kantharia ND. How to calculate sample size in animal studies? J Pharmacol Pharmacother. 2013;4:303–6.

24. Goodman NW, Hughes AO. Statistical awareness of research workers in British anaesthesia. Br J Anaesth. 1992;68:321–4.