Online version at http://www.minervamedica.it
Minerva Anestesiologica 2016 August;82(8):821

PREFACE

Three oncoming editorials with some suggestions on statistical analysis for authors and readers

Statistics offer powerful tools to medical researchers to describe their findings (descriptive statistics) and translate them to general population (inferential statistics). Analysis and presentation of data are therefore a cornerstone of the evaluation process of manuscripts submitted for publication. The statistical analysis is checked from different points of view. First, we assess the appropriateness of the tests utilized. This is a major topic because the use of a wrong statistical test makes it necessary to redo the data analysis and to reconsider discussion and conclusions. Second, the presentation of data is evaluated. It should be clear and comprehensive, the distribution between text, tables, and figures well calibrated, and duplicates avoided; to make reading easier, details of secondary importance should be given as electronic supplements. Finally, it is asked to authors to adopt the journal's style for data presentation.

Starting from the current issue and for the next two, we will publish three editorials to provide useful guidelines for authors in order to perform and correctly write the statistical section of their own research article. Prof. B.M. Cesana, a professional statistician and a member of the editorial Board of *Minerva Anestesiologica*, will address the most common issues encountered in the manuscripts submitted to the journal. These notes are not exhaustive (it would take much more space), but they aim to provide basic information to facilitate the authors' task. More complete information and lists of references for some topics will be made available on *Minerva Anestesiologica*'s website.

Franco Cavaliere
Editor-in-Chief
Minerva Anestesiologica

Online version at http://www.minervamedica.it
Minerva Anestesiologica 2016 August;82(8):822-6

EDITORIAL

Basics to perform and present statistical analyses in scientific biomedical reports Part 1

Bruno M. CESANA [1]*, Franco CAVALIERE [2]

[1]Statistics and Biomathematics Unit, Department of Molecular and Transactional Medicine, University of Brescia, Brescia, Italy; [2]Department of Cardiovascular Science, Catholic University of the Sacred Heart, Rome, Italy

*Corresponding author: Bruno M. Cesana, Statistics and Biomathematics Unit, Department of Molecular and Transactional Medicine, University of Brescia, Brescia, Italy, V.le Europa 11, 25143 Brescia, Italy. E-mail: bruno.cesana@unibs.it

Basics to perform and present statistical analysis

Scientific reading and reporting are fundamental activities of medical profession. Statistics represent a crucial tool to comprehend the contents of scientific articles and to analyze and report the results of researches. We would like to offer a brief summary of fundamental aspects and tools to perform and report statistical analyses in scientific manuscripts.

First of all, the Methods section should always include a subsection (Statistics) in which every statistical procedure (descriptive and inferential) and variable transformation are reported, as well as the sample size calculation, the significance level (one or two-tailed), and the statistical software utilized (included the version).[1-3]

Descriptive statistics

This Section mainly regards data presentation. Table I summarizes how many decimal places should be utilized and how to approximate values.

Qualitative (categorical) variables

Nominal variables like gender have two or more categories and no intrinsic order. Ordinal variables, like ASA (American Society of Anesthesiologists) Physical Status (PS) Classification Systen classes, have two or more categories and can be ordered or ranked. Both kinds of variables should be reported as absolute (number) and percent frequencies. Percentages should be rounded to the nearest tenth (25.4%, for example) with the second decimal figure rounded to the upper value if it is equal or greater than 5.[4-7] Conversely, relative frequencies should be rounded to two decimal figures at least;[4] however, their use in manuscripts submitted to MA is discouraged.

Quantitative (continuous) variables

Values should be given with the unit of measure and approximated to one decimal figure more than raw data. Alternatively, authors may choose to report fewer decimal figures according to biological or clinical meanings.[1-7] The arithmetic mean and the standard deviation (SD) should be utilized as measures of position and dispersion if values are symmetrically (Gaussian-like) distributed; otherwise, the median, the first (Q1) and third (Q3) quartiles, and maximum and minimum values should be given. From these values, the reader can eas-

TABLE I.—*How many decimal places should be utilized and how values should be approximated.*

How many decimal places		
Values	Decimal places	Notes
Numbers	None	
Percentages	One	
Measure of position/dispersion	One more than raw data	Consider biological and clinical meanings to decrease places
Statistic tests (*i.e.* χ^2 or Student's *t*-test)	Two	
Odds ratios, relative risks and hazard ratios	Two	
Probabilities	Four	Always report the value, even if it is not statistically significant. The P is upper case and not italicized, and there is no hyphen between "P" and "value"(for instance: P=0.0456). Use P<0.0001 for values lower than 0.0001
How to approximate the values		The last decimal figure reported is rounded to the upper value if the following figure is equal or greater than 5

ily calculate the range (the maximum minus the minimum) and the interquartile range (IQR: the third quartile minus the first one). To evaluate the type of the parent distribution of our sample, whether symmetrical or asymmetrical, is not always easy. As a "rule of thumb", if the difference between the mean and the median is greater than 25% of the lower value, it is better to rely on "robust" statistics, such as median, quartiles, and extremes.[7, 8]

The Coefficient of Variation (CV) is a further absolute measure of dispersion, easily calculated as the ratio (multiplied by 100) between the SD and the mean. It is particularly suitable for judging the sample variability and for comparing the variability of data measured on different scales [7] [Cesana BM. Simulation Studies. Data not shown, 2015].

The Standard Error of the Mean (SEM) should not to be utilized as a measure of dispersion. However, it can be employed in graphs, provided that its use is clearly indicated in the legend.[4-9]

Inferential Statistics: general aspects

Statistical tests for univariate and multivariate (better multivariable) analyses should be described exhaustively and references given, if uncommon. A two decimal place accuracy should be used for reporting test statistics, such as these from chi-squared or Student's *t*-test.[1-3] P values should be always reported (also in the case of not statistically significant results) with at least three decimal figures.[1-3]

In non-randomized studies, multivariable statistical procedures should be preferred in order to: 1) highlight the influence of several factors on the investigated outcome; 2) build models that compensate the effects of baseline imbalance between groups; 3) calculate propensity scores in order to try to avoid biases due to confounding factors.

In the case of missing data, authors should explain how they dealt with them and report the results of a sensitivity analysis on the consistency of their results.[10-15] On this regard, the intervention of a professional statistician is strongly recommended because of the complexity of this topic. Missing data have a different impact on the internal validity of a research whether they are missing "completely at random" (CMAR), "at random" (MAR), or "not at random" (NMAR). In general, however, the presence of missing data should be kept as low as possible regardless of how they have been generated. As a rule of thumb, it can be regarded as acceptable up to 5-10% of missing values in experimental studies, up to 25% of nonresponse in surveys, and up to 5% of lost to follow-up in survival studies (and preferably much less in the case of a low incidence of low failure incidence). Nonetheless, those percentages become un-

acceptable if the pattern of missingness suggests the presence of selection biases. In multivariable analyses, a listwise deletion [16] of subjects with missing data may exclude large percentages of observations and induce relevant selection biases. Conversely, appropriate methods such as multiple imputation, estimation using a maximum likelihood approach, multi-level models for incomplete longitudinal data are available and should be implemented, provided that missing data follow the MAR pattern.[10-15]

Inferential Statistics: checking the assumptions of statistical tests.

A basic requirement for all standard statistical tests is that data are independent of each other. Repeated measurements on the same subject should be analyzed by *ad hoc* procedures (*i.e.* paired Student's *t*-test, Wilcoxon's Signed-Rank Test, ANOVA for repeated measures, Friedman's two-way analysis of variance) in order to take into account the correlations obviously present among recorded data and to avoid fictitious increases of the sample size.

A second aspect to stress is that doing a statistical test if a previous one did not reject the null hypothesis is a bad practice that should be avoided. Indeed, it leads to a "conditional testing" procedure, which interferes on type I error (probability of rejecting a "true" H_0) and type II error (probability of not rejecting a "false" H_0).

Finally, it is generally assumed, particularly by not professional statisticians, that before performing a statistical significance test, the "fulfillment of its assumptions has to be checked". Consequently, the Gaussian distribution and homoskedasticity assumptions are checked by means of statistical tests, which in turn rely on their own assumptions.

Gaussian distribution

Gaussian (Normal) distribution is usually regarded as a condition necessary to perform parametric tests. Generally speaking, the Shapiro-Wilk test is the best choice for assessing the normality of data,[17-25] while the more frequently utilized Kolmogorov-Smirnov (K-S) Test is too conservative because it is highly sensitive to extreme values and does not reject "false" null hypotheses at the chosen significance level.[17-25] Nonetheless, some further aspects have to be highlighted:

1) If the total sample size is larger than 100, the hypothesis that a variable is Gaussian (normally) distributed can be simply and effectively tested by visually inspecting the frequency diagram.[26, 27]

2) Yet, if the sample size is larger than 30, testing the Gaussian distribution may be completely useless because parametric tests compare means rather than raw values, and according to the Central Limit Theorem (CLT), means tend to be "asymptotically" Gaussian distributed, irrespectively of the distribution of the variable. For sample sizes equal or larger than 30, approximation is (generally) good enough for parametric tests to be quite robust to non-Gaussian distributions and, consequently, tests on Gaussian distribution should not be performed and sentences such as "a nonparametric test has been used due to the rejection of the Gaussian distribution of the variable of interest" should be avoided. Of note, the tendency to be "asymptotically" Gaussian distributed applies also to proportions from discrete binomial variables. Indeed, proportions are calculated as the sum of Bernoulli variables, coded as 1 for success and 0 for failure, divided by the number of the trials so that they are, finally, arithmetic means.

3) If the sample size is lower than 30 nonparametric tests are strongly recommended.

To compare proportions, the use of the z-test (or, equivalently, of the χ^2 test) requires that a continuous distribution (z-distribution or χ^2 distribution) approximates the discrete binomial distribution. Such approximation is considered acceptable if the sample size is ≥ 30 and the products of the sample size (n) by the proportion (p) and by (1-p) are equal or greater than five (*i.e.* $np \geq 5$ and $n(1-p) \geq 5$).[6, 7, 28] The fulfillment of this condi-

tion (*i.e.* the validity of using the Gaussian approximation to the binomial distribution) should be reported.

Homoskedasticity (homoscedasticity)

Homoskedasticity or homoscedasticity is the equality of variances of two or more populations and should be evaluated only in non-randomized trials, in which groups may theoretically belong to different populations. Homoscedasticity is very important for the analysis of linear regressions, for the unpaired Student's *t*-test, and for all models of analysis of variance (one-way, two-way, and factorial).[29-32]

To test homoskedasticity, the Levene's test [33] should be preferred to the Bartlett's test.[31] However, both are markedly affected by sample numerosity. The smaller the sample size, the lower the power of the tests to detect any deviation from homoskedasticity (even if large enough to invalid Student's *t*-test or ANOVA). On the other hand, deviations negligible for the validity of statistical tests may appear as statistically significant if the sample size is large. In the end, authors are suggested not to perform a formal testing, but to consider the sample size and the estimated variance of each group of values. As a conservative "rule of thumb", a nonparametric test is strongly recommended if: 1) the sample size is small (<30); 2) there is a large unbalance between groups (the ratio between the greater and the smaller sample size is >1.5, say; or 3) the ratio between the larger and the smaller estimated variances is more than 1.5 (say).[7]

References

1. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. Br Med J 1988;296:401-5.
2. Bailar JC 3rd, Mosteller F. Guidelines for statistical reporting in articles for medical journals: Amplifications and Explanations. Ann Int Med 1988;108:266-73.
3. Reporting Results of Common Descriptive and Inferential Statistics in APA (American Psychological Association) Format [Internet]. Available from http://www.bryanburnham.net/wp-content/uploads/2014/01/Reporting-Statistics-in-APA-Format.pdf [cited 2016, Apr 2].
4. Sokal RR, Rohlf FJ. Biometry: The principles and practice of statistics in biological research. Third edition. New York: WH. Freeman; 1995.
5. Zar JH. Biostatistical Analysis. Fourth edition. Upper Saddle River, NJ: Prentice Hall; 1999.
6. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. Fourth edition. Osney Mead, Oxford, UK: Blackwell Science; 2002.
7. Cesana BM, Antonelli P, Pea G. La statistica per le scienze biomediche. Libreriauniversitaria.it Edizioni; 2012. Prima edizione digitale: marzo 2012.
8. McDonald JH. Handbook of Biological Statistics Standard error of the mean. [Internet]. Available from http://www.biostathandbook.com/standarderror.html [cited 2016, Apr 2].
9. Statistical guidelines - The Royal Society of Medicine (searched through Google) [Internet]. Available from http://www.rsmjournals.com/site/misc/stats.xhtml; [cited 2016, Apr 2].
10. Missing data [Internet]. Available from https://en.wikipedia.org/wiki/Missing_data [cited 2016, Apr 2].
11. Analysis of clinical trials [Internet]. Available from http://en.wikipedia.org/wiki/Analysis_of_clinical_trials; [cited 2016, Apr 2].
12. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; 2010. [Internet]. Available from http://www.nap.edu/read/12955/chapter/1#ii [cited 2016, Apr 2].
13. Yuan Y. Sensitivity Analysis in Multiple Imputation for Missing Data. Paper SAS270-2014. [Internet]. Available from https://support.sas.com/resources/papers/proceedings14/SAS270-2014.pdf [cited 2016, Apr 2].
14. Resseguier N, Giorgi N, Paoletti X. Sensitivity analysis when data are missing not-at-random. Epidemiology 2011;22:282-3.
15. Carpenter JR, Kenward MG. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. Stat Methods Med Res 2007;16:259-75.
16. https://en.wikipedia.org/wiki/Listwise_deletion [Internet]. Available from [cited 2016, Apr 2].
17. Thode HJ. Testing for normality. N.Y.: Marcel Dekker; 2002.
18. Kolmogorov–Smirnov test [Internet]. Available from https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test [cited 2016, Apr 2].
19. Lilliefors HW. On the Kolmogorov–Smirnov test for the exponential distribution with mean unknown. J American Statist Assoc 1969;64:387-9.
20. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965;52:591-611.
21. Shapiro SS, Francia RS. An approximate analysis of variance test for normality. Journal of the American Statistical Association 1972;67:215-6.
22. D'Agostino RB. Tests for the Normal Distribution. In D'Agostino RB and Stephens MA. Goodness-of-Fit Techniques. N.Y.: Marcel Dekker; 1986.
23. D'Agostino RB, Belanger A, D'Agostino RBJr. A suggestion for using powerful and informative tests of normality. The American Statistician 1990;44:316-21.
24. Normality tests I: Brief overview [Internet]. Available from http://brainder.org/2011/07/03/normality-tests-i/ [cited 2016, Apr 2].
25. Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. Int J Endocrinol Metab 2012;10:486-9.
26. Wilk MB, Gnanadesikan R. Probability plotting methods for the analysis of data. Biometrika 1968; 55:1–17. at: http://graphpad.com/faq/file/1872QQ%20normality%20plots.pdf [cited 2016, Apr 2].

27. Gan FF, Koehler KJ, Thompson JC. Probability plots and distribution curves for assessing the fit of probability models. The American Statistician 1991;45:14-21.
28. Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Statist Med 2007;26:3661-75.
29. Glejser H. A new test for heteroskedasticity. J Am Stat Assoc 1969;64:316-23.
30. Cook RD, Weisberg S. Diagnostics for heteroscedasticitiy in regression. Biometrika 1983;70:1-10.
31. Bartlett's Test. [Internet]. Available from http://www.itl.nist.gov/div898/handbook/eda/section3/eda357.htm [cited 2016, Apr 2].
32. PROPHET StatGuide: Do your data violate one-way ANOVA assumptions? [Internet]. Available from: http://www.basic.northwestern.edu/statguidefiles/oneway_anova_ass_viol.html [cited 2016, Apr 2].
33. Levene H. In: Olkin I editor. Contributions to probability and statistics: essays in honor of Harold Hotelling. Stanford University Press; 1960. p. 278-92.

Conflicts of interest.—The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Acknowledgements.—The contribute of Pietro Caironi, MD (Dipartimento di Fisiopatologia Medico-Chirurgica e dei Trapianti, Fondazione IRCCS Ca' Granda – Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy; and Dipartimento di Anestesia, Rianimazione ed Emergenza Urgenza, Fondazione IRCCS Ca' Granda – Ospedale Maggiore Policlinico, Milan, Italy) who read a preliminary version of this paper, added some contributions from the clinical point of view and gave some very useful suggestions for making this contribute usable and understandable as possible to the clinicians is warmly acknowledged.

Article first published online: May 31, 2016. - Manuscript accepted: May 30, 2016. - Manuscript revised: May 18, 2016. - Manuscript received: April 12, 2016.

(*Cite this article as:* Cesana BM, Cavaliere F. Basics to perform and present statistical analyses in scientific biomedical reports. Part 1. Minerva Anestesiol 2016;82:822-6)