Short Communication

SEJODR

StaTips Part I: Choosing statistical test when dealing with differences

Perinetti Giuseppe *

** Department of Medical, Surgical and Health Sciences, University of Trieste, Italy*

ABSTRACT

When dealing with statistical test hypothesis, one of the most common problems to deal with relates to the difference between or among groups, treatments or time points. Herein, a short guide is provided to chose the proper statistical test according to the nature of the data and study design.

Perinetti G. StaTips. South Eur J Orthod Dentofac Res. 2016;3(1):4-5.

FRAMING THE PROBLEM

When dealing with statistical test hypothesis, one of the most common problems to deal with difference between or among groups, treatments or time points. For instance, what would be the effects of different functional treatments in reducing the ANB angle between two compared groups? Of course, no one single test would be apt for every purpose or situations. Therefore, several statistical tests are available to this goal, even though some of them are most common in medicine. According to different conditions, data may belong to a group of the investigation over different time points, or to a time point of across different groups investigation. That is why the term data set is used to indicate a set of data from all of these situations.

To choose the appropriate test is a simple logical combination of three parameters: i) nature of the data sets (as interval, ordinal and nominal); ii) dependence of the data sets (as independent or paired) and iii) number data sets to compare per time (as 2 or ≥3). By crossing the different categories of each parameter, the appropriate test is easily identified (See Table 1 and worked examples).

Corresponding Author:
Perinetti Giuseppe
Struttura Complessa di Clinica Odontoiatrica e Stomatologica,
Ospedale Maggiore,
Piazza Ospitale 1, 34129 Trieste, Italy.
e-mail: G.Perinetti@fmc.units.it

NATURE OF THE DATA

Interval data refers to quantities on a linear scale. This data may be defined as fully quantitative since intervals between consecutive values are identical. Cephalometric parameters are examples of interval data. Ordinal data refers to quantities that have a natural ordering on a non-linear scale. This data may be defined as semi-quantitative since increasing values denote greater magnitudes, although intervals between consecutive values are not necessarily identical. The adhesive remnant index is an example of ordinal data. Nominal data refers to categories and may be defined as qualitative. Male or female gender is an example of (dichotomous) nominal data. Of note, data may be transformed from one kind to another provided that the only possible way is from interval to ordinal/nominal or from ordinal to nominal. For instance, the (interval) age may be re-classified as ordered (ordinal) clusters as 20-29 years for cluster 1, 30-39 years for cluster 2, and so on.

Once recognised the nature of data, one more step is necessary when dealing with interval data. Indeed, to proper use statistical tests for such a data (also referred as parametric tests), it is required that each data set to be included in the analysis meets some requirements (also referred as assumptions). The most important assumptions are: i) normal distribution of the data sets and ii) equality of variances between/among the data sets. These assumptions may be checked by the Shapiro-Wilk test and Leven test for normal distribution and equality of variance, respectively. Several statistical packages include these and other similar tests. Unless in the case of large data sets (with more than 100 cases), failure of meeting either assumption in any data sets requests interval data to be treated as ordinal (by non-parametric tests).

DEPENDENCE OF THE DATA SETS

The concept of a dependence of the data sets relates to how data is collected. For instance, data sets obtained from two or more different groups of individuals (or in vitro samples) are independent, while data sets deriving from the same group of individuals (or samples) over time are classified as paired (or dependent). Of note, while longitudinal data are paired by nature, cross-sectional data is usually, but not always, independent. Every split-mouth protocol presents with cross-sectional matched data sets collected from the same individual at the very same time and are thus considered as paired.

MULTIPLE COMPARISONS WHEN HAVING 3 OR MORE DATA SETS TO COMPARE

When dealing with 3 or more data sets to compare it is necessary to evaluate the significance of the difference among all data sets as a first attempt, and only if the difference reaches a significant level, subsequent (pairwise) comparisons between data sets may be run. For instance, when dealing with three data sets, A, B and C, provided that the differences among the sets are significant, then pairwise comparisons will follow considering the case of the 2 data sets as A vs. B, A vs. C and B vs. C. In this case, however, each resultant p-value of the pairwise comparisons have to be multiplied by the number of multiple comparisons made (in this case 3). If the corrected P-value still stays below the significant level (usually 0.05), then the corresponding difference may be considered significant. Such a procedure is referred as Bonferroni correction and may be applied to P-value derived from any test for 2 data sets listed in Table 1, irrespective of the nature or dependence of the data sets. Of note, Bonferroni corrections may become too much conservative above 6-8 multiple comparisons requiring other procedures not reported herein.

Table 1. *Most common statistical methods to test the significance of the differences between/among data sets*

Nature of the data	Dependence of the data			
	Independent		Paired	
	2 data sets	≥3 data sets	2 data sets	≥3 data sets
Interval with assumptions	Unpaired t-test	Analysis of variance (ANOVA)	Paired t-test	Repeated measures analysis of variance (ANOVA)
Ordinal (or Interval without assumptions)	Mann-Whitney test	Kruskal-Wallis test	Wilcoxon test	Friedman test
Nominal	Chi-squared test (best referred as Fisher exact test)	Chi-squared test	McNemar test	Cochran Q test

In Table 1, data sets may represent groups or time points. Some worked examples:

1) When testing the significance of the differences between 2 groups represented by independent interval data sets (with assumption), then, an unpaired t-test would be the choice. Alternatively, if assumptions are not meet in either data set, a Mann-Whitney test has to be used
2) If one wishes to test the significance of the differences in ANB angle after a functional treatment in a group of treated Class II patients, then, assuming that the ANB is an interval data, and that before and after recordings are represented by 2 data sets of paired data (with assumptions), the paired t-test has to be used. Alternatively, if assumptions are not meet in either data set, a Wilcoxon test has to be used
3) When comparing different groups of treated/control subjects, it is important to be sure that groups are homogeneous regarding gender distribution. In the case of 3 different compared groups, considering that gender is a nominal and independent (in this case) data, the Chi-squared test has to be employed.

REFERENCES

1. Glantz S. Primer of Biostatistics, 7th ed. Columbus: McGraw-Hill Education. 2011.
2. Kleinbaum DG, Kupper LL, Muller KE, editors. Applied Regression Analysis and Multivariable Methods. Boston: PWS-KENT Publishing Company. 1989.