IV. Appropriateness of the Statistical Test

Juan O. Talavera, Rodolfo Rivas-Ruiz

When we observe the difference between two therapies or the association of a risk factor or prognostic indicator with its outcome, we have to assess the certainty of the result. This assessment is based on a judgement that uses information related with the design of the study and the statistical handling of the information. In this article, the relevance of the selected statistical test is specifically mentioned. Statistical tests are chosen based on two features: the objective of the study and the type of variables. The objective can be divided in three groups of tests: *a)* those in which showing differences between groups or in a same group before and after a maneuver is wanted; *b)* those in which showing a relationship between variables is wanted; *c)* those in which predicting an outcome is pretended. As for the types of variables, we have two: quantitative (continuous and discontinuous) and qualitative (ordinal and dichotomous). For example, if we want to demonstrate age differences (quantitative variable) between patients with systemic lupus erythematosus, with and without neurological involvement (two groups), the adequate test is Student's *t*-test for independent samples; but if what is being compared in those same groups is the frequency of females (binomial variable), then the relevant statistical test is the chi-square test (χ^2).

Key words
biomedical research
research projects
statistics and quantitative data

This article was originally published in Rev Med Inst Med Seguro Soc 2011; 49 (3): 289-294 and it has been reviewed for this issue.

Introduction

When we observe the difference between two therapies or the association of a risk factor or a prognostic indicator with its outcome, a question arises: Is the result real? Deciding if it is real requires two complementary judgments:

1. The planning and development of the process that document such difference or association are free of errors, or at least these are of a minor magnitude, which does not modify the sense of the difference or association (i.e., appropriate design and adequate execution).
2. The size of the sample is sufficient to maintain the stability of data and the statistical test is suitable for the objective.

The planning and development of the process have been mentioned in the three previous chapters of this series. On the other hand, data stability will be discussed in detail in a subsequent article when the size of the sample and the *p*-value are addressed.

In this article, we will discuss the relevance of the selected statistical test. Undoubtedly, this knowledge will allow for us to understand more precisely the results obtained in clinical research studies and, of course, it will increase our ability to make an adequate use of them.

Study Objective and Type of Variable

Statistical tests are selected based on two features: the objective of the study and the type of variables. Within the study objectives we can identify three:

1. Demonstrating differences between groups or differences in a same group before and after a maneuver (e.g., treatment with drug *A* reduces high blood pressure in a greater proportion than treatment with drug *B*).
2. Showing relationships (correlation) between variables (e.g., serum creatinine rises as renal function decreases).
3. Predicting an outcome (e.g., the likelihood for the subject with sedentary life and overweight of developing type 2 diabetes mellitus).

Frequently, the models overlap, and thus, models initially identified to predict an outcome are sometimes used to demonstrate differences between two groups. This happens especially when the principal maneuver has to be adjusted (drug *A* versus drug *B*) for multiple factors (age, sex, body mass index, etc.). But the oppo-

site phenomenon also happens when looking to predict an event that will occur in the future but there are only one or two predictors available; in this case, a test to demonstrate differences is used.

It is important to clarify that the correlation basically is useful for seeing the magnitude of the association between variables, although it should remain clear that it does not establish causality. As a matter of fact, no statistical test can. This requires covering a number of principles described by Sir Austin Bradford Hill.

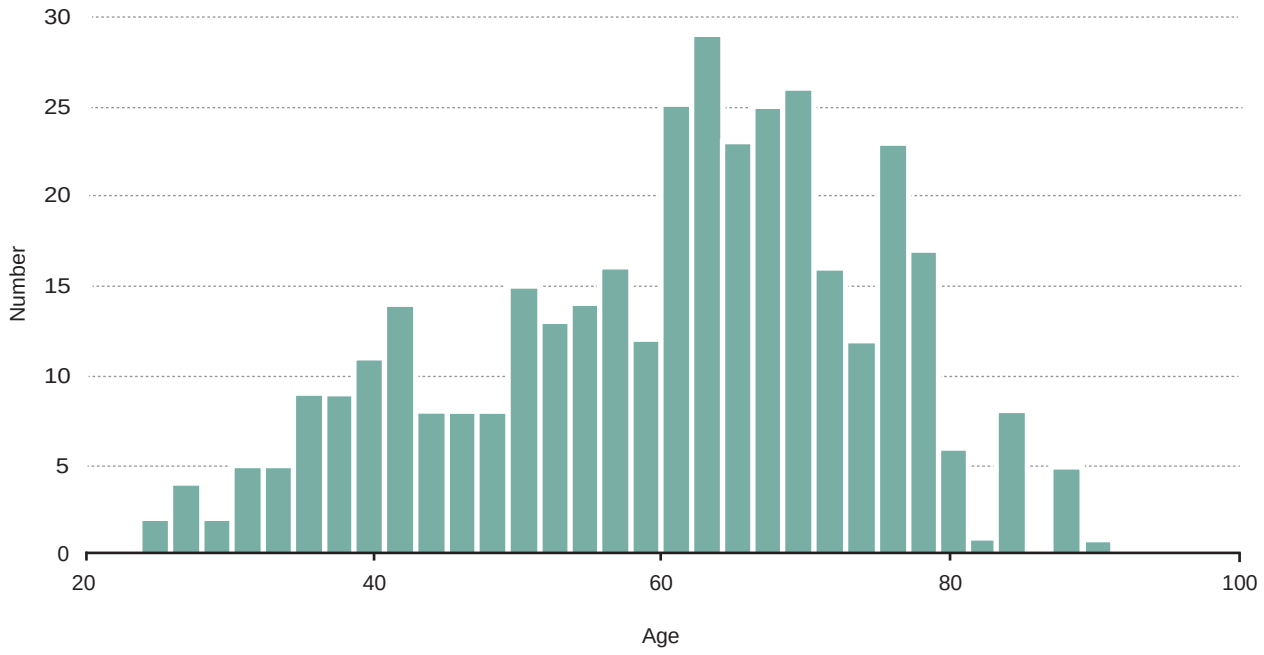
Defining the *type of variable* is relevant because it is the axis in the selection of the appropriate test depending on the desired objective. Within the types of variables there are two groups:

1. Quantitative: continuous and discontinuous or discrete. The former are characterized because they can take any value throughout a continuum (for example, 1.75 m height). On the other hand, discontinuous or discrete variables use exclusively whole numbers (parity, 1, 2, 3...). In both instances, the distance between one unit and another throughout its scale is equidistant.
2. Qualitative: these include the ordinal and the dichotomous variables. The ordinal variable allows for the characteristic under study to be ordered and, unlike what happens in quantitative variables, the distance between two categories is not equidistant (e.g., heart failure grades I to IV). Dichotomous variables, as their name indicates, are those with only two categories, which can be binomial (one option or another, e.g., male or female) or nominal (it refers to the presence or absence of the feature, e.g., alive at six months, yes or no).

Table I Weight of subjects studied under two therapeutic regimens

Group A			Group B		
77			65		
78			69		
80			77		
82			78		
85	83.5	Average	85	83.5	Average
85	85.0	Median	85	85.0	Median
85	85.0	Mode	89	85.0	Mode
86			93		
88			96		
89			98		

Central tendency measurements are equal, but the dispersion of data is different



Mean: 59.79
Standard deviation: 13.882. Two standard deviations at either side of the mean reflect 95 % of the population
Average: 59.79, 95 % CI = 32.03-87.55

Figure 1 Histogram

It is important to mention the handling that the type of variable will suffer during the analytical process, starting with the collection of "crude" data, which means that this is only a collection of information from a group of subjects. In order for these data to have a useful meaning, they have to be organized and summarized. The simplest organization method is the frequency distribution tables; however, sometimes it is easier to understand their graphic representation through a histogram or frequency polygon. Regardless of the usefulness of this information, collected data are required to provide quantitative information, i.e., numerical indices reflecting different probability distributions are required, whose primary function is to *model* the behavior of a large variety of biological phenomena. These numerical indices include the measures of central tendency and the measures of dispersion.

1. Measures of central tendency (Table I and Figure 1).

a) *Mean:* it is the sum of a set of data divided by its total number. The symbol to represent the mean of a population is the Greek letter mu (μ), and the mean of a sample is represented by $\bar{x}$. It is the most widely used summary measure for quantitative variables.

b) *Median:* it is the value located exactly in the middle of the entire set of data. The median divides a distribution of data ordered exactly in two equal parts. The advantage of the median as a measure of central tendency is that it is not affected by the value of extreme data, a phenomenon that does occur with the mean. It is the type of summary measure most widely used for quantitative variables not following a normal distribution and for ordinal variables.

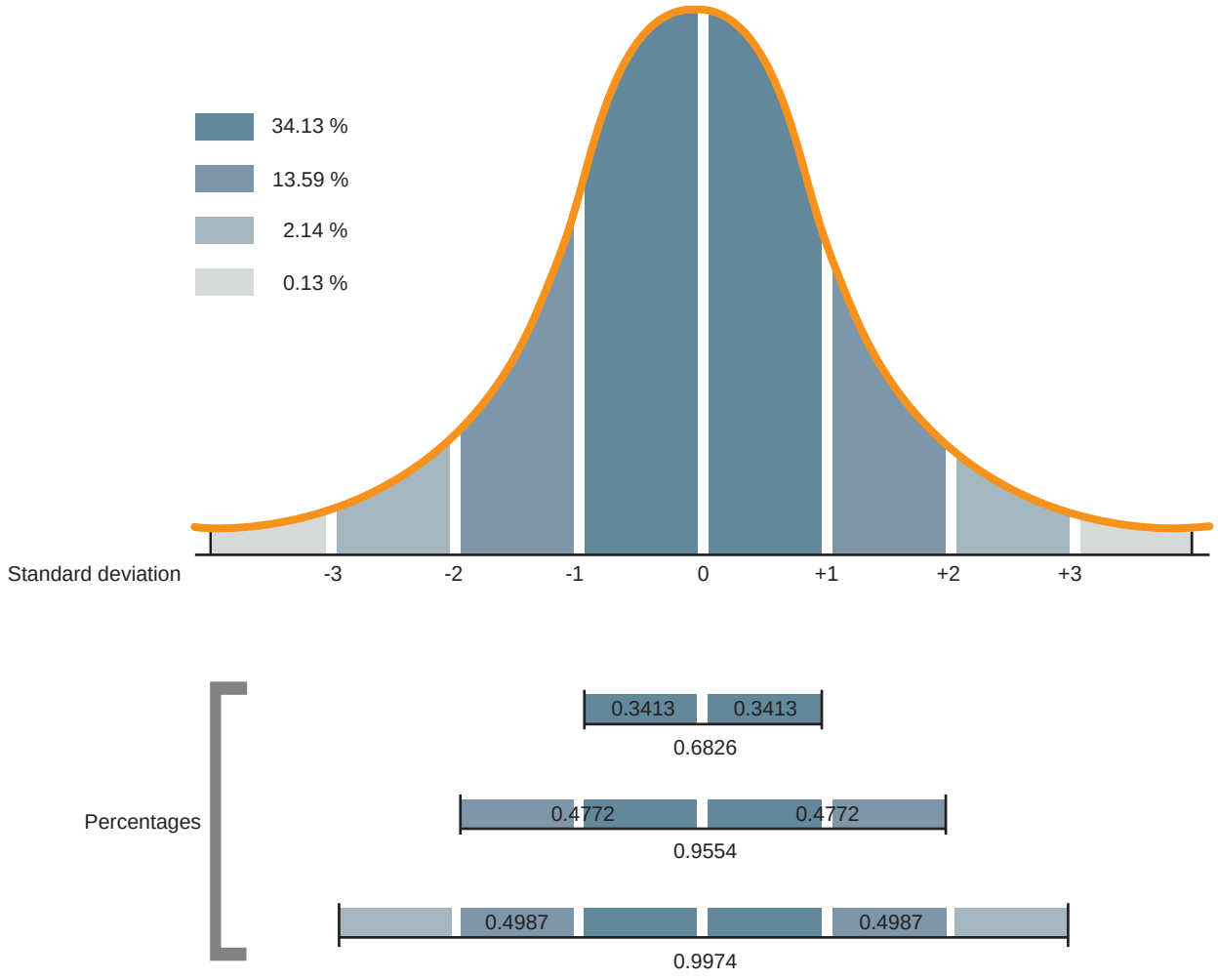


Figure 2 Normal distribution curve

c) *Mode:* it refers to the most repeated value in a distribution. This measure is hardly used in medicine.

2. Most common measures of dispersion.

a) *Standard deviation:* it reflects the variation between the whole data set and it is used when these follow a normal distribution.
b) *Percentile*: it describes the position of a value of the distribution. It is used for quantitative variables not following a normal distribution and for ordinal variables.
c) *Range:* it is the difference between the highest and the lowest value of the distribution.
d) *Interquartile ranges:* these are referred to the values of the first and third quartile.

In clinical research, as in many other real-life phenomena, the most commonly analyzed data are quantitative, which in most cases show a Gaussian distribution, also known as normal distribution, which is characterized for having a bell-like shape, for being symmetric with regard to its mean, for having decreasing frequency values as they move away from the mean, and for never reaching zero (asymptotic). The mode and the median are equal to the mean; about 68 % of data are within ± 1 standard deviation from their mean and 95 % within ± 2 standard deviations (Figure 2). Thus, if the set of data is quantitative with a normal distribution, its summary measure will be the mean, and its dispersion measure, the standard deviation. However, if its distribution is not Gaussian, same as it is for an ordinal-type variable, its summary measure will be the median, and its dispersion measure, the percentile or rank. Generally, these variables do not have dispersion measures and when they are used, 95 % confidence intervals are preferred.

Appropriateness of the Statistical Test

Once we know our objective and the characteristics of our data (type of variable), we can consider the appropriateness of the statistical test (Table II). However, there are two more considerations when the objective is to demonstrate difference:

1. If it is a study in which the value of a data item is compared before and after a maneuver, either observational or experimental, it is known as related samples test, but if it involves the comparison of data between different groups, it is called unrelated samples test.
2. If it consists in a comparison between different groups, it is necessary to establish if it is going to be between two or more.

Table II Selection of the statistical test according to the objective and type of variable

		To demonstrate difference		To show relationship[&]	To predict 1 variable[‡]
Type of variable	Type of sample	Two groups	Three groups	Two variables	Outcome variable
Quantitative	NR	*Student's t**	1 factor ANOVA	Pearson	Linear regression
(normal distribution)	R	*Student's t* **	1 factor ANOVA		
Qualitative ordinal	NR	*Mann-Whitney U*	Kruskal-Wallis	Spearman	
(free distribution)	R	Wilcoxon	Friedman		
Qualitative dichotomous	NR	χ^2 (or Fisher exact test)	χ^2 (of linear tendency)	Phi coefficient	Logistic regression
R	McNemar			Survival curves	

NR = not related; R = related; R = measure of the variable in the same subject at two different time-points
* Student's *t* for independent samples
** Student's *t* for related samples
& For the correlation between 2 variables, the test of that at the lower scale is used (actually, no scale is lower; however, variables have been ordered from quantitative continuous to dichotomous, by way of quantitative discontinuous and ordinal variables).
‡ The predictor can be quantitative, dichotomous or ordinal (with these last transformed into *dummy*-like variables)

With the information already complete, with Table II we can verify if the selection of the statistical test was appropriate according to the variable and the objective. For example, if age is compared (quantitative variable with normal distribution in this case) between patients with systemic lupus erythematous, with and without neurological involvement (two groups), the appropriate test is Student's *t*-test for independent samples. But if that what is being compared between these same patients is the frequency of females (binomial variable), then the appropppriate statistical test is the chi-square (χ^2) test. If that what is being compared between both groups is their degree of lupus-like activity (ordinal scale), the appropriate statistical test is the Mann-Whitney *U*-test. On the other hand, if that what we are shown is the magnitude of association (relationship) between age (quantitative variable with normal distribution) and the degree of lupus-like activity (ordinal variable), the relevant test is Spearman's *r*. Finally, if that what is sought to be predicted is the weight of a child (quantitative variable) based on age (quantitative variable), type of nutrition (ordinal variable: good, fair or poor) and sex (dichotomous), the appropriate test is the linear regression. But if that what is wanted to be predicted is the probability of infarction (dichotomous nominal) over the next 10 years based on age (quantitative), atherogenic risk (ordinal, low, moderate and high) and sex (dichotomous binomial) the relevant test is the multiple logistic regression.

Finally, we hope this article allows for the reason of the selection of the most widely used statistical tests in health research to be understood and, at the same time, to serve as a guideline to those who are taking their first steps in statistics. It is not sufficient for establishing if the obtained results are real; it will be necessary to take into consideration the design and execution of the study and the stability of the information, but this last issue deserves to be discussed in another section. The next chapters of this series will further address Student's *t*, Mann-Whitney *U* (with which we will address how to select the type of distribution of quantitative variables) and chi-square tests.

Bibliography

1. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed. Massachusetts, MA: Blackwell Publishing; 2002.
2. Bland M. Introduction to medical statistics. 3rd ed. Oxford: Oxford University Press; 2003.
3. Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia, PA: W.B. Saunders; 1985.
4. Feinstein AR. Multivariable analysis: an introduction. New Haven, CT: Yale University Press; 1996.
5. Feinstein AR. Principles of medical statistics. New York, NY: Chapman and Hall/CRC; 2002.
6. Le Chap T. Introductory biostatistics. Hoboken, NJ: New Jersey: John Wiley and Sons; 2003.
7. Peat J, Barton B. Medical statistics. A guide to data analysis and critical appraisal. Malden, MA: Blackwell Publishing; 2005.
8. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Saddle River, NJ: Pearson/Prentice Hall; 2009.