Biostatistics' Corner

Which one is better to present results of a study? p-value or confidence interval?

Turkish Respiratory Journal, 2000;1:63-64

P value is based on the null hypothesis. In this approach data are examined in relation to a stastical "null hypothesis". Then according to the statistical test results the null hypothesis is rejected or accepted. This dichotomous statements p<0.05 (significant result), p>0.05 (non significant result) give little information about the study. On the other hand the purpose of the most researches in medicine is to determine the magnitude of some factor(s) of interest. In such cases confidence interval (CI) helps us. CI depends on the degree of "confidence" that we want to associate with resulting interval. The main purpose of confidence intervals (CI) is to indicate the precision of the sample study estimates as populations values.

An important point is that there is a close link between the use of a confidence interval and a two sided hypothesis test. CI's are more informative than p–values because they provide a range of possibilities for the population value. It is recommended that if the confidence interval is calculated then the result of the hypothesis test can be interpreted at a selected level of statistical significance as follows:

For ex: The difference between the sample mean FEV1 measurements in asthematics and non-asthematics was 21% with a 95% confidence level from 5 to 32, the t-test statistic was 2.4 with 48 degrees of freedom and associated p-value of p=0.02.

Correspondence: *Nural Bekiroğlu*
Marmara University Hospital
Altunizade, İstanbul

Mean ±SD 21± : 95% CI 5 to 32: t=2.4. d.f.= 48. p=0.02 (p<0.05).

What is the difference between correlation and regressino analysis?
The most common statistical analyses are those analyses the consider the relation between two variables in one group of subjects. For two variables X and Y, regression analysis is used to predict one variable from another and correlation analysis to see if the values of two variables (X and Y) are associated. The purposes of these two analyses are distinct and usually one only should be used.

Correlational studies only establish the direction and the degree of relatedness between variables and do not permit causal inference in and of themselves. Linear regression is concerned with finding the best fitting straight line for a data set and estimating parameters from the sample data. We use sample data to estimate the parameters such as slope, intercept.

The equation for the population regressino line is: Y= A+BX where A is the intercept on the Y-axis (the value of Y when X=0) and B is the slope of the line which estimates the mean change in Y for a unit change in X.

In standard regression analysis it is assumed that the distribution of Y variable at each value of X is Normal with the same variance. but no assumptions are made about the distribution of the X variable.

When you fit a simple linear model to a set of data, a process called a simple regression analysis and it contains one independent variable and the dependent (response) variable. In classical regression analysis there are two or more independent variables and a single dependent variable. The methodology for finding the multivariable predictor is called multiple regression analysis which is fairly complex.

The independent variables are fixed by the researcher and are therefore not random on the other hand the dependent variable is random.

For correlation analyses, either Pearson's product moment correlation coefficient (r) is used as a parametric correlation coefficient (which measures the direction and the degree of linear correlation between two variables x and y) or Spearman's rank correlation coefficient (rs) as a nonparametric correlation coefficient.

X2 Coefficient of Determination

The square of the correlation coefficient. r2, is called the coefficient of determination. r2 indicates a close connection between correlation and regression. r2 reveals the percentage of the variation on Y that is accounted for by variation in the X variable or conversly, the percentage of the variation in X that is accounted for by variations in the Y variable.

An important point is that :r is a measure of linear correlation and that X and Y could be perfectly related by some curvilinear function when the observed value of r is equal to zero which means there is no linear relationship between X and Y.

The primary purpose of this section is to answer to statistical common problems. My intention is that this section be an attractive feature for medical researchers. I hope that this section is to be a valuable feature.