EuroIntervention 2013;9:765-767

Tools and Techniques - Statistics: Analysis of continuous data using the t-test and ANOVA

Isabella Kardys*, MD, PhD; Sanne Hoeks, PhD; Ron van Domburg, PhD; Mattie Lenzen, PhD; Eric Boersma, PhD

Clinical Epidemiology Unit, Department of Cardiology, Erasmus MC, Rotterdam, The Netherlands

Series Editors: Philipp Kahlert[1], MD; Jerzy Pregowski[2], MD; Steve Ramcharitar[3], MD; Christoph Naber[4], MD

1. Department of Cardiology, West German Heart Center Essen, University Duisburg-Essen, Essen, Germany; 2. Institute of Cardiology, Warsaw, Poland; 3. Wiltshire Cardiac Centre, Great Western Hospital, Swindon, United Kingdom; 4. Contilia Heart and Vascular Centre, Elisabeth Krankenhaus, Essen, Germany

Introduction

In clinical research, we often examine continuous variables such as blood pressure, ejection fraction, laboratory values (e.g., cholesterol), and angiographic variables (e.g., percent stenosis). We may, for example, want to compare these measurements between different patients or between different time points. To compare continuous data, parametric or non-parametric tests of significance may be applied. Which of these tests is appropriate depends on several factors, including the nature of the data to be analysed. These data should meet the assumptions that are required for the particular test. This paper describes parametric methods and the circumstances under which these methods are appropriate. Its aim is to give a general overview, without claiming completeness; hence, exceptions are possible in specific situations.

Basic considerations

DISTRIBUTION OF DATA

When continuous (i.e., interval or ratio) data are organised and graphed as a histogram, they take on a shape referred to as a distribution[1]. The most common distribution is the normal curve, which is symmetric and has a shape that resembles a bell **(Figure 1)**. However, a distribution may also be skewed, i.e., not symmetric[2]. For example, a distribution can be positively skewed when most of the measurements occur at the lower end of the distribution.

MEAN AND MEDIAN

The arithmetic mean is the sum of all values, divided by the number of values. The median is the value that divides the distribution in half, i.e., if the observations are arranged in increasing order, the median is the middle observation[2]. Thus, half of the observations are above the median and half are below it. If there is an even number of observations, there is no middle one and the average of the two "middle" ones is taken. To summarise a variable, it is usually recommended that, when a variable follows a normal distribution, the mean and standard deviation (SD) should be reported. When a variable does not follow a normal distribution, the mean may be unrepresentative of the majority of the data[2]. In that case, the median and range (e.g., 25th and 75th percentile) should be reported instead. Furthermore, since the mean is very sensitive to outliers, to which the median is more robust, it is preferable to report the median and range when the distribution shows extreme cases, despite being expected to be normal.

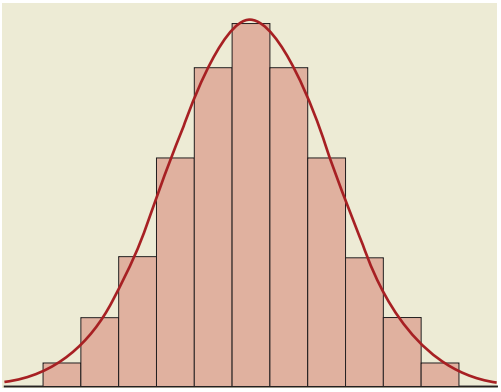

Figure 1. *Histogram and normal distribution.*

STATISTICAL SIGNIFICANCE

In short, hypothesis testing is based on the following. The null hypothesis states that there is no difference between the study groups; for example, it states that mean blood pressure is the same in group 1 and group 2, or that blood pressure is the same before receiving

**Corresponding author: Erasmus MC, Department of Cardiology, room Ba-561, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands. E-mail: i.kardys@erasmusmc.nl*

DOI: 10.4244/EIJV9I6A123

medication and after receiving medication. The null hypothesis can be true or false. Statistical tests determine the probability that the null hypothesis is erroneously rejected; this probability is called a p-value (or "alpha level")[1]. The smaller the p-value, the stronger is the evidence against the null hypothesis. The cut-off point of the p-value is usually set at 0.05 and is called the significance level. There are numerous statistical tests available; in this paper we focus on two commonly used parametric tests.

Choosing between parametric and non-parametric methods

Parametric methods are appropriate when data are measured on the interval or ratio scale and when they are distributed normally. Normality of the data may be examined by visual examination of histograms, box plots or Q-Q plots and by performing tests of normality such as the Kolmogorov-Smirnov test or the Shapiro-Wilk test[3]. If the data are not distributed normally, non-parametric statistical methods are appropriate[4]. Non-parametric methods are also more appropriate when one is dealing with small samples, since in such cases it is often difficult to assess the normality of the distribution of the data[5], and the influence of extreme data points on the mean is larger.

Parametric tests are typically more powerful than non-parametric tests, meaning that if a difference between the study groups truly exists, that difference is more likely to be found using the parametric test[4]. However, because not all data are normally distributed or measured on an interval or ratio scale, in some cases only non-parametric methods can be applied.

Parametric tests

T-TEST

The t-test, also called Student's t-test, is one of the most commonly used methods in clinical research[6]. It is a parametric method that is based on the means and SDs or variances of the data[4]. There are several assumptions for using the t-test: the sample data must be derived from a normally distributed population; for two sample tests, the two populations must have equal variances; and each measurement (or the difference score for dependent data) must be independent of all other measurements[1,3]. In case the assumption of equal variances is not fulfilled, Welch's t-test may be used as described elsewhere[7].

Several types of t-test exist[3,4,6,7]. The one-sample t-test is applied when one study group is examined and may be used to compare the group mean to a theoretical mean. The paired t-test is used to estimate whether the means of two related sets of measurements are significantly different from one another. This test is used when measurements are dependent because they are collected a) from the same participant at different times, b) from different sites on the same person at the same time, or c) from cases and their matched controls[3]. When two study groups are examined and the measurements performed in the groups are independent (which, for example, applies to an [unmatched] control versus experimental group design), an independent two-sample t-test is appropriate.

When a group mean is compared to a theoretical mean (one-sample t-test), the null hypothesis states that the group mean is equal to this theoretical mean. The t-statistic is calculated as the difference between the group mean and the hypothesised value of the group mean, divided by the standard error (SE) of the mean (**Figure 2**, equation 1)[6]. The SE may be substituted by the SD divided by the square root of the number of subjects (equation 1). For example, suppose that in a certain medical centre mean systolic blood pressure in patients treated for stable angina pectoris is 124 mmHg. We want to test whether the mean systolic blood pressure in patients with stable angina in our own centre is different from our hypothesised value of 124 mmHg. Our study population consists of 100 patients. Suppose that we measure a mean systolic blood pressure of 132 mmHg with an SD of 23 mmHg in our study population. In this case, t is equal to $(132-124)/(23/\sqrt{100})=3.48$. The accompanying p-value, as derived from a table of critical values for t, is <0.001. Thus, we conclude that the null hypothesis ("mean systolic blood pressure in our study population is equal to 124 mmHg") may be rejected at the 0.05 level (and in this case even at the 0.001 level). These results can also be easily obtained from statistical software programmes.

To compare the means of two dependent sets of measurements, such as repeated measurements performed on patients from one single group, a paired t-test is used. Suppose we want to test in our study population of 100 patients with stable angina whether the mean systolic blood pressure before receiving certain medication is the same as that after receiving the medication. In this case, t is calculated as the mean difference between the measurements at the two time points, divided by the SE of this difference (**Figure 2**, equation 2)[6]. Again, the SE may be substituted by the SD (in this case, the SD of the mean difference), divided by the square root of the number of subjects. It should be noted that the SD of the mean difference is not merely a combination of the SDs of the means at the two time points, but is calculated from the total number of mean differences. Suppose that in our data we find a mean difference in systolic blood pressure of 5 mmHg with an SD of 4 mmHg. Consequently, t is equal to $5/(4/\sqrt{100})=12.5$. Once more, this renders a p-value <0.001.

To compare the means of two independent sets of measurements (independent two-sample t-test), t is calculated as the difference between the two group means divided by the SE of this difference (**Figure 2**,

(1) One-sample t-test:

$$t=\frac{\bar{X}-\mu_0}{SE}=\frac{\bar{X}-\mu_0}{SD/\sqrt{n}}$$

$\bar{X}$ is the observed mean
μ_0 is the hypothesised value of the mean
SE is the standard error
SD is the standard deviation
n is the number of subjects

(2) Paired t-test for the mean difference:

$$t=\frac{\bar{d}-0}{SE\ (\bar{d})}=\frac{\bar{d}-0}{SD_d/\sqrt{n}}$$

$\bar{d}$ is the mean difference
SD_d is the SD of the mean difference

(3) Independent two-sample t-test:

$$t=\frac{\bar{X}_1-\bar{X}_2}{SE\ (\bar{X}_1-\bar{X}_2)}=\frac{\bar{X}_1-\bar{X}_2}{SD_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$SD_p=\sqrt{\frac{(n_1-1)\ SD_1^2+(n_2-1)\ SD_2^2}{n_1+n_2-2}}$$

$\bar{X}_1$ is the observed mean in group 1
$\bar{X}_2$ is the observed mean in group 2
n_1 is the number of subjects in group 1
n_2 is the number of subjects in group 2
SD_p is the pooled standard deviation

Figure 2. *t-test: calculation of t.*

equation 3)[6]. The SE of the difference can be calculated from the pooled SD and the numbers of subjects of both groups, as depicted in equation 3. The pooled SD, for its part, can be calculated from the SDs of each of the groups and the number of subjects of both groups **(Figure 2)**. Suppose we want to investigate whether in our study population of 100 patients with stable angina systolic blood pressure in men is the same as systolic blood pressure in women. Suppose there are 30 women with a mean systolic blood pressure of 134 mmHg, and 70 men with a mean systolic blood pressure of 131 mmHg, and the pooled SD is 24 mmHg. The SE of the difference then equals $24^* \sqrt{(1/30+1/70)}=5.2$. Thus, $t=(134-131/5.2)=0.58$, and the accompanying p-value is >0.5. We conclude that the null hypothesis may not be rejected at the 0.05 level.

ANOVA

When more than two groups are compared using multiple t-tests, the probability of rejecting a true null hypothesis is increased as the number of comparisons made using independent t-tests increases[4]. Analysis of variance (ANOVA) is the appropriate statistical method to test for differences among three or more groups[3]. The assumptions for ANOVA are similar to those for the t-test (i.e., normal distribution; equal variances; each measurement independent of all other measurements)[3,4,6].

The general theory behind the calculations of ANOVA is based on the following. ANOVA considers the variation in all observations and divides it into: a) the variation between each subject and the subject's group mean, and b) the variation between each group mean and the grand mean[6]. If the group means are quite different from one another, considerable variation will occur between them and the grand mean, compared with the variation within each group. If the group means are not very different, the variation between them and the grand mean will not be much more than the variation among the subjects within each group[6]. The concept of ANOVA can be thought of as an extension of a two-sample t-test but the terminology used is different[3]. Just as the t-test uses calculation of a t-statistic, ANOVA uses calculation of an F-ratio. This F-ratio is defined as (between-groups variance) / (within-group variance), and indicates whether the variability between the groups is large enough compared to the variability of data within each group to justify the conclusion that two or more of the groups differ[3,4,6]. If an ANOVA was being used instead of the t-test to compare two groups, it would be found that $F=t^2$ for these data[4]. After obtaining the F-ratio, it may be compared to the critical F-ratio in order to find the p-value. In our example of patients with stable angina, we would apply ANOVA if, for example, we wanted to test whether systolic blood pressure is the same for current smokers, former smokers and those who have never smoked.

For related measurements, for example blood pressure assessed at three or more time points in the same patients, repeated measures ANOVA may be used[3]. Repeated measures ANOVA can be considered as an extension of the paired t-test. A detailed description of repeated measures ANOVA is beyond the scope of this article.

Conclusions

Parametric methods are used for comparison of continuous data and include the t-test, which is appropriate when the experimental design consists of one or two sample groups, and ANOVA, which may be used when there are three or more groups to compare. The data being analysed should meet the assumptions which apply to the given test. These assumptions are summarised in **Figure 3**.

Other parametric methods for analysing continuous data, including linear regression, as well as non-parametric methods, will be described in future papers within the current series. Moreover, excellent references on these topics are provided by Bland and Altman[5,8].

I. Assumptions required for all t-tests (1 or 2 groups):

The outcome variable must be continuous

The outcome variable should approach a normal distribution in each group

For two-sample tests, the population variance should be the same for both groups (variance is SD^2)

Assumptions which apply specifically to the paired t-test:

The differences between the pairs of measurements should approach a normal distribution

Assumptions which apply specifically to the independent two-sample t-test:

The groups must be independent, i.e., a subject may only be part of one group

The measurements must be independent, i.e., the subject's measurement can be included only once

II. Assumptions required for ANOVA (3 or more groups):

The outcome variable must be continuous

The outcome variable should approach a normal distribution in each group

The population variance should be the same for all groups

The value of one measurement is not related in any way to the value of another measurement

Figure 3. *Assumptions for t-test and ANOVA.*

Conflict of interest statement

The authors have no conflicts of interest to declare.

References

1. Wissing DR, Timm D. Statistics for the nonstatistician: Part I. *South Med J.* 2012;105:126-30.
2. Kirkwood B, Sterne JAC. Essential Medical Statistics. Malden, MA: Blackwell Science; 2003.
3. Peat J, Barton B. Medical statistics: a guide to data analysis and critical appraisal. Malden, MA: Blackwell Publishing; 2005.
4. Gaddis GM, Gaddis ML. Introduction to biostatistics: Part 4, statistical inference techniques in hypothesis testing. *Ann Emerg Med.* 1990;19:820-5.
5. Bland JM, Altman DG. Analysis of continuous data from small samples. *BMJ.* 2009;338:a3166.
6. Dawson B, Trapp RG. Basic & Clinical Biostatistics. New York, NY: McGraw-Hill; 2005.
7. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. Malden, MA: Blackwell Science; 2002.
8. Altman DG. Statistics and ethics in medical research: V--Analysing data. *Br Med J.* 1980;281:1473-5.