



WASP (Write a Scientific Paper) using Excel – 8: *t*-Tests

ARTICLE INFO

Keywords:
Software
Computers
Statistics
Biostatistics

ABSTRACT

t-Testing is a common component of inferential statistics when comparing two means. This paper explains the central limit theorem and the concept of the null hypothesis as well as types of errors. On the practical side, this paper outlines how different *t*-tests may be performed in Microsoft Excel, for different purposes, both statically as well as dynamically, with Excel's functions.

1. Central limit theorem

A parametric statistical test is one that assumes that the data being analysed is normally distributed. The Central Limit theorem states that a substantial sample tends to be normally distributed, and that the mean of all such theoretical samples (were it possible to repeatedly sample the population) will be approximately equal to the population mean, even when the population from where it is taken is not very normally distributed. This allows a parametric test to be carried out in order to compare two means, using the *t*-distribution, as outlined in the previous paper in this series [1]. It is thanks to this fact that a panoply of parametric tests have evolved. Compared to non-parametric tests which do not invoke the Central Limit theorem, parametric tests are more accurate and precise, with higher statistical powers [2].

2. The null hypothesis

The null hypothesis is an a priori statement of impartiality, prior to analysis, which states that when testing two samples, it is assumed that these come from the same population and that they are not statistically significantly different. This is stated as:

$H_0: \mu_1 - \mu_2 = 0$, where μ_1 represents one sample and μ_2 represents the second sample.

The pre-established threshold or cut-off (known as α) to demonstrate a difference in the biological sciences is taken at 5% or significance testing for $p \leq 0.05$, which is the 1.96 standard deviations that was previously discussed [3]. However, this implies that if testing were to be done a hundred times on random samples, 5% would be erroneously shown to be different despite there being any true difference, due to chance alone.

Statistical testing may thus lead to the rejection of the null hypothesis (the statistical test finds a difference, $p \leq 0.05$, i.e. the test generates a 95% likelihood that results do not fit the null hypothesis) when the null hypothesis is true. Or the test may fail to reject the null hypothesis ($p > 0.05$) when it is false.

3. Errors

The statements above imply an inevitable lack of certitude and the possibility of errors, of which there are two types.

A Type 1 error (also known as an “error of the first kind”, an α error) is a false positive result, a statistically significant result which does not represent reality. This thus causes the rejection of the null hypothesis when the null hypothesis is true since testing results in the claim that there is difference between two samples. At the accepted 5% level, this may naturally occur 5 times out of a hundred (tests) simple due to chance, the random nature of sampling. This is thus also referred to as an error of excessive credulity (Table 1). This typically occurs when the sample population is not representative of the population from which it was chosen.

A Type 2 error (also known as an “error of the second kind”, a β error) is a false negative result, a statistically non-significant result which does not represent reality. This thus causes the acceptance of the null hypothesis, when the null hypothesis is false since testing results in the claim that there no difference between two samples (Table 1). Typically type 2 errors occur because the study is “underpowered” or had too few subjects to investigate a particular question.

Table 1
Types of errors.

| | | Reality | |
|--------|--------------|--------------|--------------|
| | | H_0 true | H_0 false |
| Test | Accept H_0 | Accept H_0 | Type 2 error |
| Result | Reject H_0 | Type 1 error | Reject H_0 |

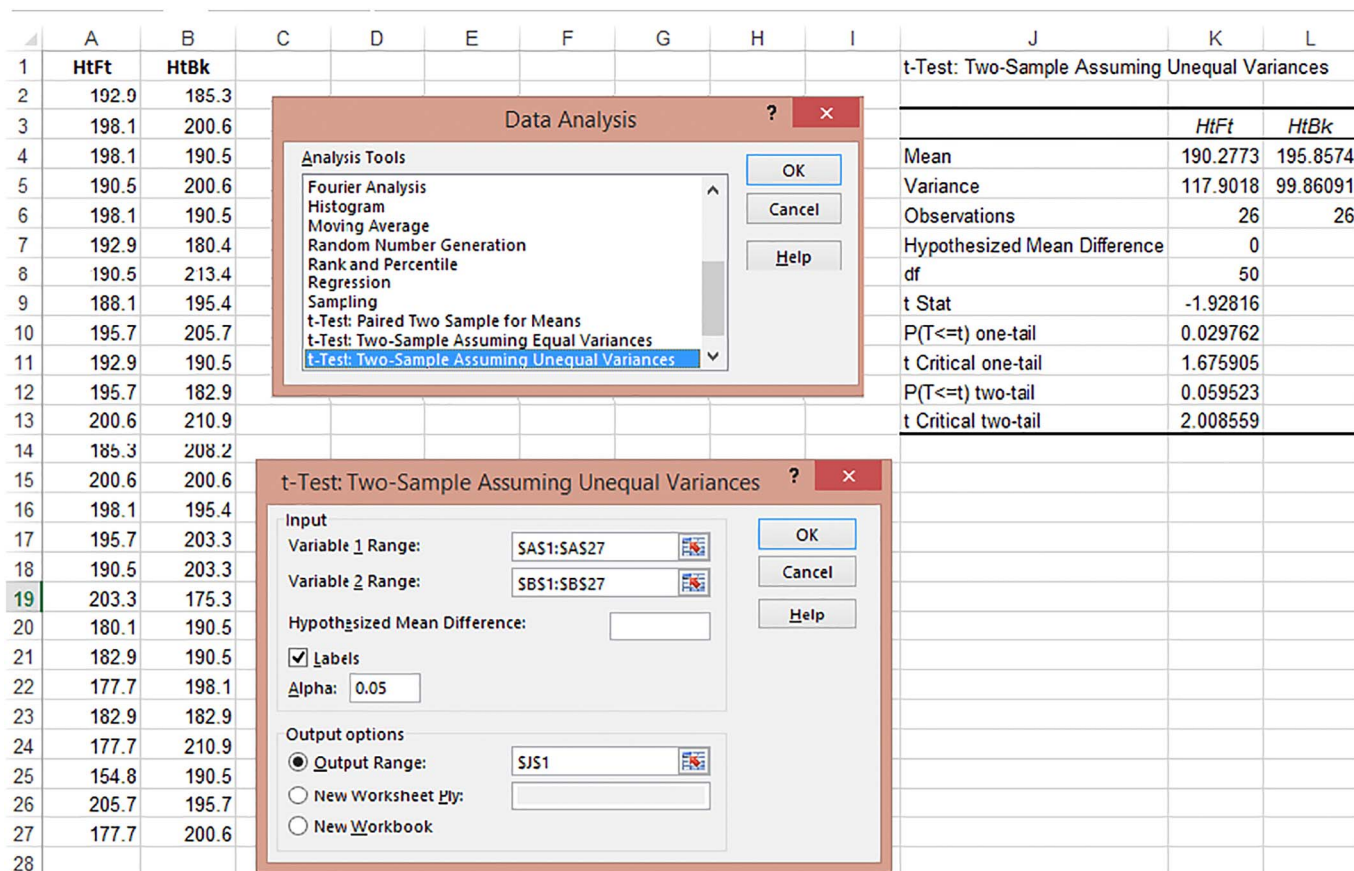


Fig. 1. *t*-Testing American football (HtFt) and basketball players (HtBk) using Excel's Data Analysis Toolpak.

4. Other fields of science

Other science fields are far more stringent. For example, in high-energy physics, the threshold of “evidence of a particle” detection is taken at $p = 0.003$ and “discovery” of a particle at $p = 0.0000003$, equivalent to five standard deviations, such that the test would have to be repeated 3.5 million times for a similar result to occur just once by chance alone [4].

5. *t*-Testing

Testing involves comparing means and since the *t*-test is a parametric test, it assumes that the data is approximately normally distributed. Three types of *t*-tests are available within Excel from the Data Tab-Data Analysis option, as shown in Fig. 1 which displays the analysis of a random selection of athletes from two groups: American football and basketball players. In this example, there are 26 players in each group and the null hypothesis states: $H_0 = \text{height of football players (HtFt)} - \text{height of basketball players (HtBk)} = 0$.

In practice, the variance of the population being studied is very rarely known with any certainty. Most data is analysed using the option “*t*-test: Two-Sample Assuming Unequal Variances”. Clicking on this option opens the dialog box at the bottom of Fig. 1 which should be filled in as shown in order to perform the test. Please note that Labels is ticked since each column contains a header row, and that alpha is preset at the conventional 0.05. The generated results are shown in the last three columns and are static – results will not change if the original data is changed.

The results additionally provide some descriptive statistics: the means of both samples, their variances, and the total number of observations. These are followed by the *t* statistic obtained for the test, the value for a one-tailed and a two-tailed test, and the critical values for both.

6. One- vs two-tailed tests

In this case, p (two-tailed) = 0.0595, so the H_0 is accepted as the test has not reached statistical significance at the 0.05 (5%) level. The two-tailed

Download English Version:

<https://daneshyari.com/en/article/8777639>

Download Persian Version:

<https://daneshyari.com/article/8777639>

[Daneshyari.com](https://daneshyari.com)