T-test in SPSS (Frequentist) - Solutions

Developed by Lion Behrens, Naomi Schalken and Rens van de Schoot
last modified: 20-02-2017

Example Data

***Question**: Write down the null and alternative hypotheses that represent this question. Which hypothesis do you deem more likely?*

H0: PhD recipients with and without children have similar PhD-delays.

H1: PhD recipients with and without children have different PhD-delays.

Preparation - Importing and Exploring Data

***Question**: Have all your data been loaded in correctly? That is, do all data points substantively make sense? If you are unsure, go back to the .csv-file to inspect the raw data.*

The descriptive statistics make sense:

B3_difference_extra: Mean = 9.97, SE=0.791

E4_having_child: Mean= 0.18 (=18%), SE=0.021

T-Test

```
T-TEST GROUPS= E4_having_child (0 1)
 /MISSING=ANALYSIS
 /VARIABLES=B3_difference_extra
 /CRITERIA=CI(.95).
```

Perform an independent samples t-test and interpret the output.

***Question**: Using a significance criterion of 0.05, is there a significant effect?*

The result can be found in the 'Independent Samples Test' table. However, before we can inspect the results, we should have a look at Levene's test of homogeneity of variances. Because Levene's test is significant, F=4.517, p=0.034, we should look at the second row for the results. Based on these results, we are not able to reject H0, and therefore conclude that PhD recipients with and without children are not significantly different in their PhD delays, with t(78.96)=-1.821, p=.072.

Surveys in academia have shown that a large number of researchers interpret the p-value wrong and misinterpretations are way more widespread than thought. Have a look at the article by Greenland et al. (2016) that provides a guide to clear and concise interpretations of p.

***Question**: What can you conclude about the hypothesis being tested using the correct interpretation of the p-value?*

Assuming that the null hypothesis is true in the population, the probability of obtaining a test statistic that is as extreme or more extreme as the one we observe is 7.2%. Because the effect is above our pre-determined alpha level we fail to reject the null hypothesis.

Recently, a group of 72 notable statisticians proposed to shift the significance threshold to 0.005 (Benjamin et al. 2017, but see also a critique by Trafimow, …, Van de Schoot, et al., 2018). They argue that a p-value just below 0.05 does not provide sufficient evidence for statistical inference.

***Question**: How does your conclusion change if you follow this advice?*

The conclusion doesn't change here, but it becomes more obvious that the null hypothesis shouldn't be rejected.

Of course, we should never base our decisions on single criterions only. Luckily, there are several additional measures that we can take into account. A very popular measure is the confidence interval.

***Question**: What can you conclude about the hypothesis being tested using the correct interpretation of the confidence interval?*

The 95% CI [-8.638, 0.385].

The 95% CI's does contain 0, which means, the null hypotheses should not be rejected. A 95% CI means, that if infinitely samples were taken from the population, then 95% of the samples contain the true population value. But we do not know whether our current sample is part of this collection, so we only have an aggregated assurance that in the long run if our analysis would be repeated our sample CI contains the true population parameter.

***Question**: What can you say about the relevance of your results? Focus on the mean difference between the groups and calculate Cohen's D. If you are unsure how to calculate it, consult Cohen (1992).*

Cohen's d can be calculated as follows:

$$\hat{d} = \frac{\overline{X_1} - \overline{X_2}}{s} = \frac{9.22 - 13.35}{s_p} = \frac{9.22 - 13.35}{14.365} = -0.2875$$

Because the SD's are not equal for both groups, we should calculate a pooled SD:

$$s_p = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}} = \sqrt{\frac{(273 - 1)13.910^2 + (60 - 1)16.300^2}{273 + 60 - 2}} = 14.365$$

According to Cohen's definitions, this is a small to medium effect.

d = 0.2 (small)

d = 0.5 (medium)

d = 0.8 (large)

Only a combination of different measures assessing different aspects of your results can provide a comprehensive answer to your research question.

***Question**: Drawing on all the measures we discussed above, formulate an answer to your research question.*

Based on the above measures, we cannot reject the null hypothesis, because p = .072 and 95% CI [-8.638, 0.385]. However, we also couldn't say the null hypothesis is 'true', or the difference between the groups is equal to 0. Besides that, Cohen's d indicates there is a small to medium effect. Therefore, it becomes clear that making a decision only based on a p-value is not sufficient, and other measures should be considered as well.