S 3 statistical hypothesis testing – independent samples

hypotheses testing for the means of two independent files

1. T – test for independent samples, if the variances are equal ($\sigma_1^2 = \sigma_2^2$)

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n-1} + \frac{s_2^2}{n-1}}}$$

2. T – test for independent samples, if the variances are not equal ($\sigma_1^2 \neq \sigma_2^2$) (Welsh test)

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

The general characteristics of the individual stages:

a) assessment of the meaningfulness of the application of statistical methods
b) corresct formulation of the null hypothesis H_0
c) select the significance level (convention 0,05)
d) calculate value of the statistical test
e) calculate the value of p (*the probability of outcomes that still speaks against H_0*) *if the value is greater than or equal to 0,05 so H_0 cannot be rejected (software, including Jamovi)*
f) the assessment of statistical significance (if that is our goal)
g) assessment practical significance
h) interpretation of the results

EXAMPLE 1

Using hand grip dynamometry, we measured strength of the hand with two independent **samples** of these mans: teachers ($n_1 = 20$) and non-teachers ($n_2= 30$) groups. Do comparison of these two groups. Calculate these values (from the source data from the subsequent table)

$n_1 = 20$ $\quad$ $\bar{x}_1 = 60$ $\quad$ $s_1 = 3{,}2$

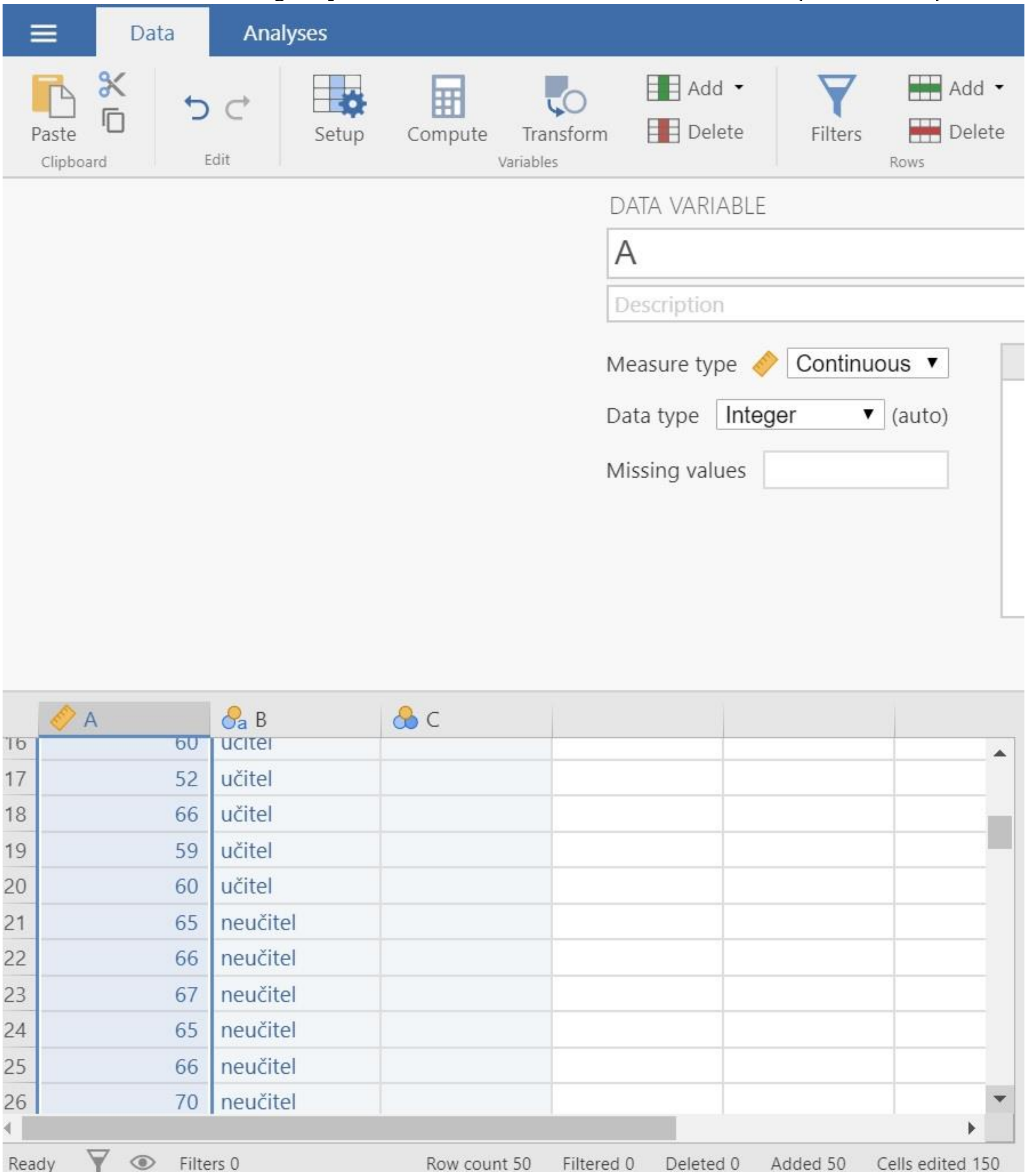
$n_2 = 30$ $\quad$ $\bar{x}_2 = 66$ $\quad$ $s_2 = 3{,}6$

Source data

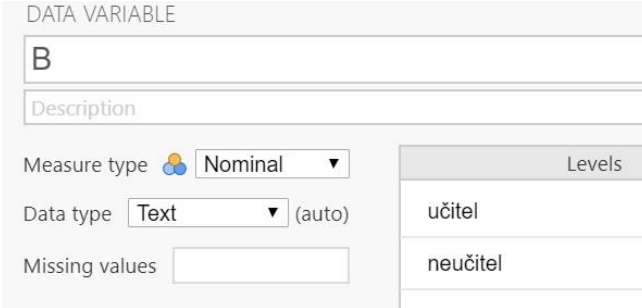
n_1	Power	Stud.group	n_2	Power	Stud.group
1	60	teacher.	1	65	not a teacher.
2	62	teacher.	2	66	not a teacher.
3	58	teacher.	3	67	not a teacher.
4	61	teacher.	4	65	not a teacher.
5	60	teacher.	5	66	not a teacher.
6	55	teacher.	6	70	not a teacher.
7	65	teacher.	7	64	not a teacher.
8	62	teacher.	8	59	not a teacher.
9	61	teacher.	9	63	not a teacher.
10	57	a teacher.	10	72	not a teacher.
11	64	teacher.	11	70	not a teacher.
12	60	teacher.	12	66	not a teacher.
13	61	teacher.	13	61	not a teacher.
14	59	teacher.	14	72	not a teacher.
15	58	teacher.	15	65	not a teacher.
16	60	teacher.	16	67	not a teacher.
17	52	teacher.	17	67	not a teacher.
18 of	66	teacher.	18	63	not a teacher.
19	59	teacher.	19	65	not a teacher.
20	60	teacher.	20	66	not a teacher.
			21	59	not a teacher.
			22	70	not a teacher.
			23	67	not a teacher.
			24	68	not a teacher.
			25	60	not a teacher.
			26	61	not a teacher.
			27	66	not a teacher.
			28	69	not a teacher.
			29	70	not a teacher.
			30	70	not a teacher.

Procedure of calculation in the program Jamovi.

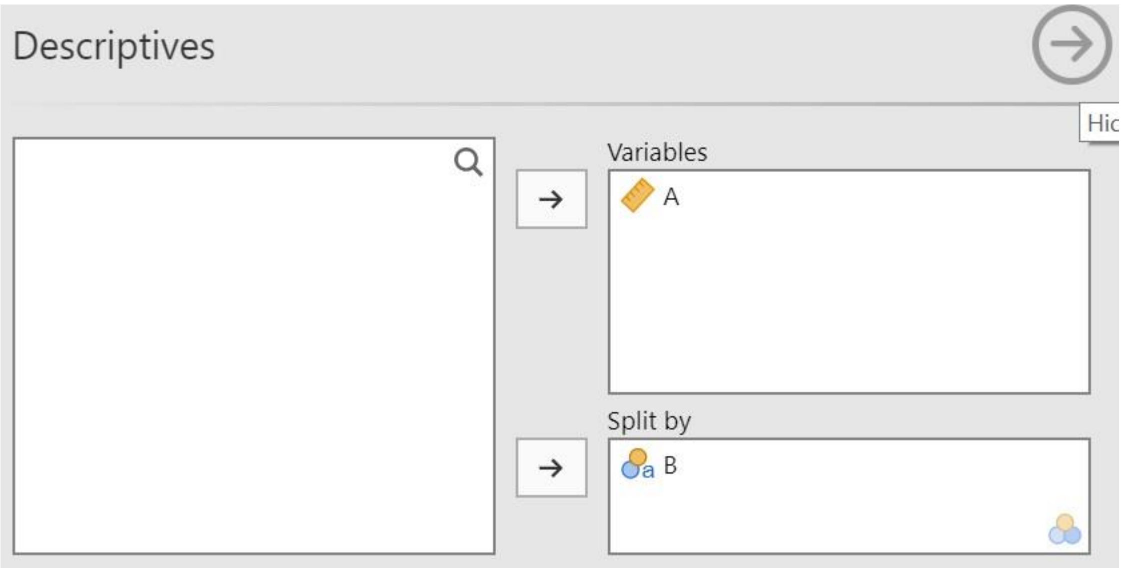
Put the data from both groups in column a, we denote it as a metric (*Continuous)*



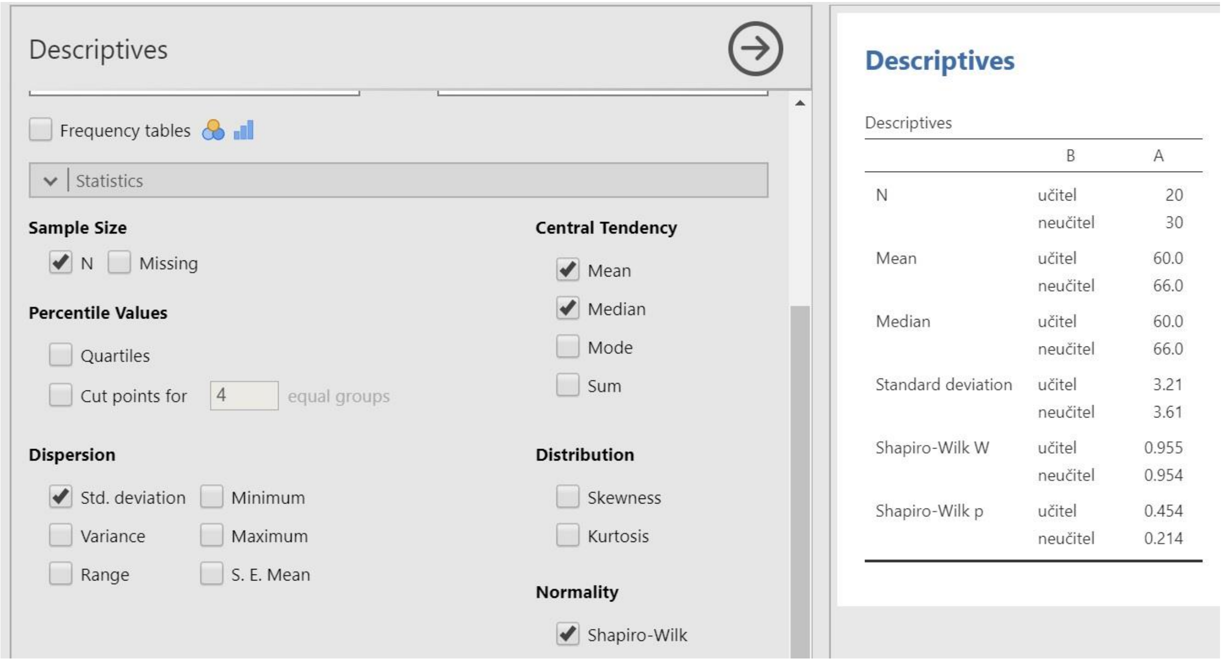
In column B, enter the study group. The Data should be marked as *Nominal,* the data type - *Text.*



Using the functions of the *Da* → *Exploration* → *Descriptives* assign the data to the item *Variables* (the variables in our example, the individual performances) and class (in our examples the teacher/not a teacher) to items *(Split would)*



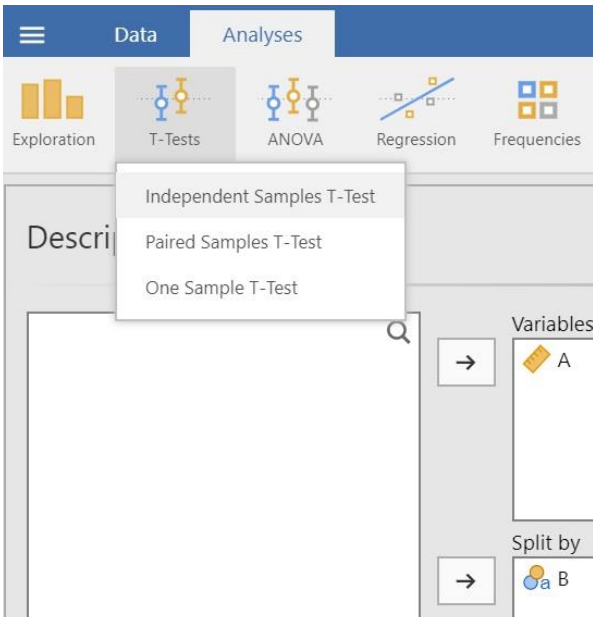
In the tab *Statistics* we choose the subsequent items descriptive statistics including a test of normality – *Shapiro- Wilk test.*



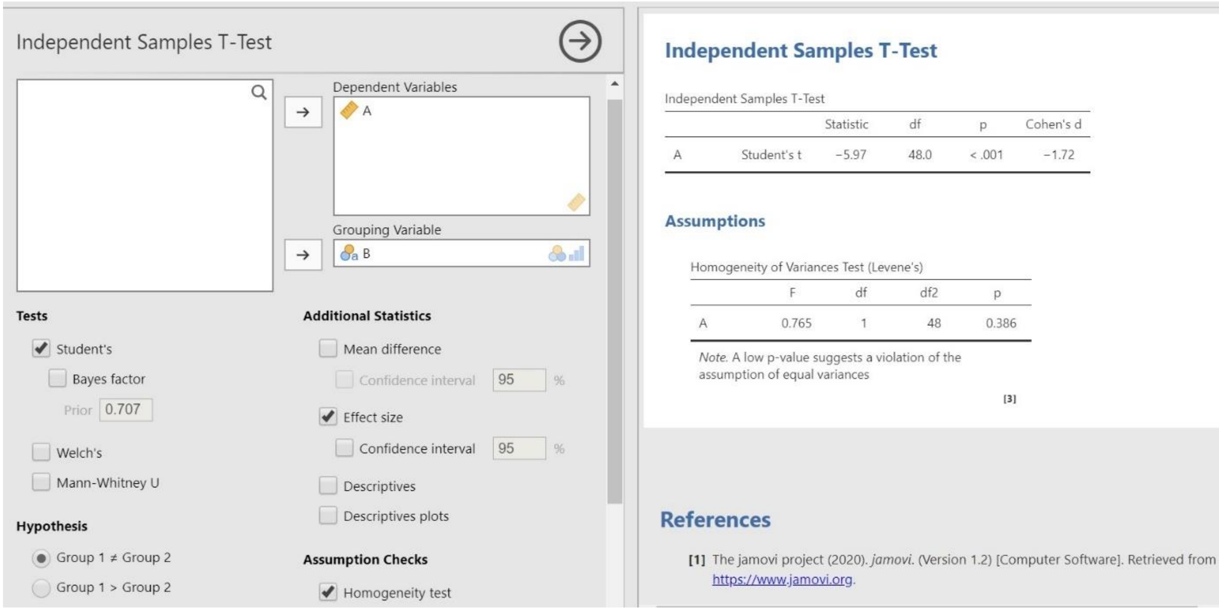
In the results in the right part of the check the resulting value of *p* Shapiro- Wilk test. If it is higher than 0.05 for both groups, we see a normal distribution frequency, which allows us to use the intended parametric T test for independent samples. In the opposite case, we used the non-parametric equivalent of the T test – thus the Mann- Whitney U test (more in chapter 9).

Note. The assumption that both selections come from a normal distribution, does not have to be followed. According to Sebera (2014), the T-test because he works with the averages of the two samples, and those already at the extent of selection in the order of tens have approximately a normal distribution due to the central limit theory.

Calculate the value of the t test for independent samples: moving through the menu *Analyses → T – tests → Independent Samples T - test*



We choose the subsequent items and in the right window, the results will be displayed:



The resulting value of p for the *T test* is less than 0,001. By comparing the values of p (with the level of significance of 0.05) so we're rejecting the null hypothesis H_0 and conclude on the statistically significant difference between the both gropus.

By ticking the options *Homogenity test*, we test whether the variances of the two files are equal, whose not. In this case, the value of p (0,386) for Leven test is higher than 0.05 which indicates the equality of variances. The use of the T test for independent files with equal variances was correct. In the case of the opposite, i.e., if the value of p was less than 0.05 we state the inequality of variances. In this case, for testing hypotheses about the average used T test for independent files with unequal variances, the so-called Welsch test (see the choice in the previous image). This variant of the us approach example 2 in this seminar.

Theory

Practical significance.

So far, researchers evaluated the practical the significance of the exclusively in measured units eg. in cm, seconds, points, etc., which remains necessary. Concurrently, we use statistical coefficients so-called "effect size" These are the coefficients, which we deem significant in relation to the other unmonitored effect and they are usually given in percentages.

For the assessment of practical significance, we have available a minimum of three tools available:

1. The statistical significance of the as a rule, $p = 0{,}05$
2. Logical judgement, when we set the minimum value of the size in units of measurement (prior)
3. Determination of the percent of the "effect size"

Processed co according to Blahuše, (2000) The procedure of computing the **(practical)** significance (**efect size, ω^2**)

it is calculated according to the formula: $\omega^2 = \frac{t^2-1}{t^2+n_1+n_2-1}$

$$\omega^2 = \frac{-5{,}97^2 - 1}{-5{,}97^2 + 20 + 30 - 1} = 0{,}409$$

The result is greater than 0.1, and therefore, the observed difference is practically significant. This means that the difference in performance between the two groups is 41 % influenced to study groups. Other, as a rule, the unknown factors is influenced 59 % of the difference.

Among other of the applicable coefficients of substantive significance include Cohen's *d*. The calculation of this coefficient is part of the program Jamovi. In the menu for the calculation of the value of the T test to check the *Effect size.* The value of Cohen's *d* is then listed in the results section. See the previous picture of the program Jamovi. The final reported value is interpreted as an absolute value, therefore, in the previous particular example is a value of 1.72. The difference between the groups according to the following scheme we evaluate verbally as very large.

The size of the effect of	*d*	**Source**
of Very small	0,01	Sawilowsky, 2009
Small	0,20	Cohen, 1988
Medium	0,50	Cohen, 1988
Big	0,80	Cohen, 1988, A
Very large	1,20	Sawilowsky, 2009
Huge	2,00	Sawilowsky, 2009

EXAMPLE 2

Random selected of women's study groups Tv-Cze ($n_1 = 20$) and Tv-Geo ($n_2 = 30$) achieved these performance in the vertical jump (Sargentův test).

$n_1 = 20$ $\bar{x}_1 = 62{,}2$ $s_1 = 9{,}2$

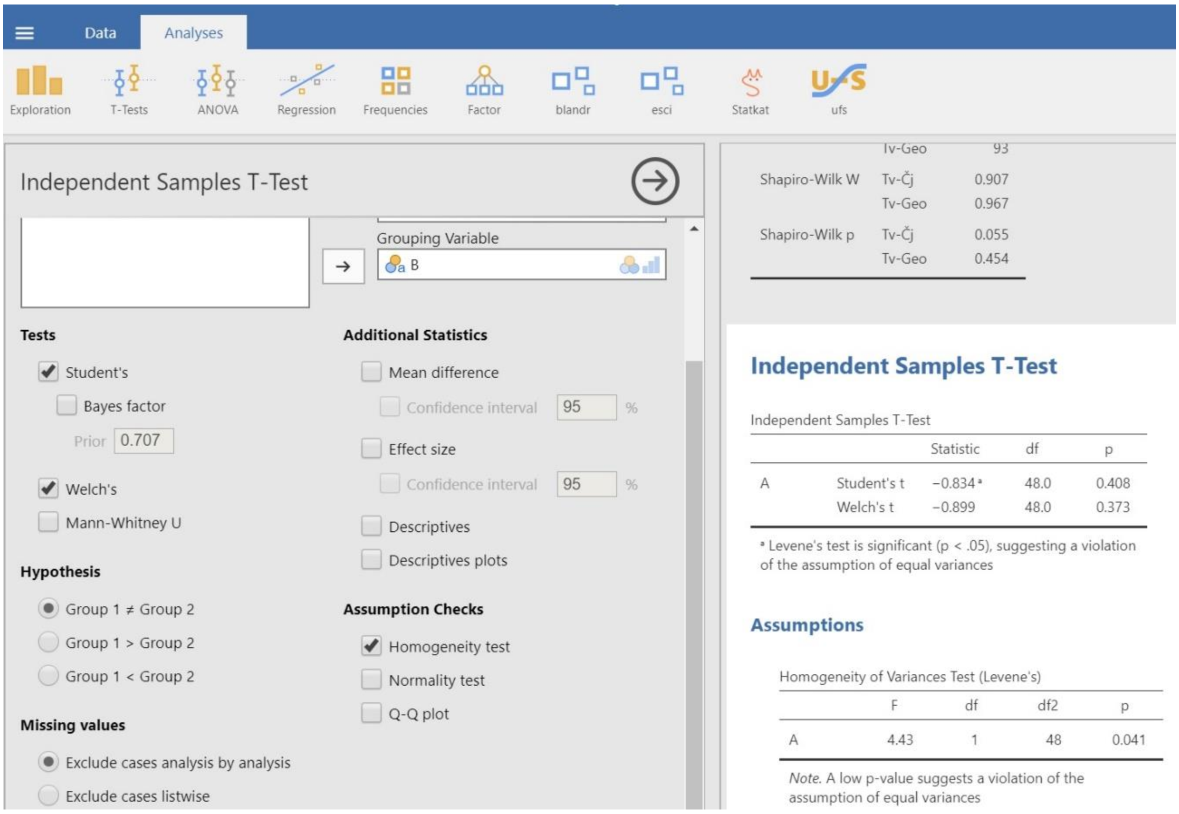
$n_2 = 30$ $\bar{x}_2 = 65{,}2$ $s_2 = 13.6$

Do comparison of the two groups. The source data are listed in the subsequent table.

n_1	Performance	Field testing of	n_2	Performance	Field of specialization
1	60	Tv- Cze	1	71	Tv-Geo
2	65	Tv- Cze	2	62	Tv-Geo
3	62	Tv- Cze	3	60	Tv-Geo
4	45	Tv- Cze	4	58	Tv-Geo
5	62	Tv-Cze	5	60	Tv-Geo
6	80	Tv-Cze	6	77	Tv-Geo
7	65	Tv-Cze	7	77	Tv-Geo
8	65	Tv-Cze	8	77	Tv-Geo
9	65	Tv-Cze	9	77	Tv-Geo
10	65	Tv-Cze	10	77	Tv-Geo
11	65	Tv-Cze	11	56	Tv-Geo
12	70	Tv-Cze	12	63	Tv-Geo
13	70	Tv-Cze	13	85	Tv-Geo
14	73	Tv-Cze	14	55	Tv-Geo
15	65	Tv-Cze	15	45	Tv-Geo
16	56	Tv-Cze	16	65	Tv-Geo
17	55	Tv-Cze	17	44	Tv-Geo
18	68	Tv-Cze	18	36	Tv-Geo
19	45	Tv-Cze	19	45	Tv-Geo
20	43	Tv-Cze	20	81	Tv-Geo
			21	55	Tv-Geo
			22	60	Tv-Geo
			23	77	Tv-Geo
			24	65	Tv-Geo
			25	75	Tv-Geo
			26	80	Tv-Geo
			27	45	Tv-Geo
			28	70	Tv-Geo
			29	93	Tv-Geo
			30	Of 64	The Tv-Geo

In the calculation in the Jamovi software we proceed identically as in the previous case. In the results section (see fig.) we can see the warning, that the variances of the two files are not equal. (*Leven*'s *test is significant (p < .05), suggesting , and the violation of the assumption of equal variances).* This fact we can, where appropriate, verify and selecting the option *Homogeneity test* with the result of the values p = 0,041.

In which, where appropriate, Student's t test we choose Welsch test. p value = 0,373 is greater than 0.05 we can reject the null hypothesis; the files are no different. For this reason, we do not sompute the practical significance.



TASK

Is there a statistically significant difference in the values of the starting reaction top sprinters? (Is the value of the starting reaction influenced by sex?)

As input data, we use the starting reaction racers in preliminary on the athletic championship in Doha 2019. Do a random selection of 15 mans and 15 women. Data are available at:

https://www.worldathletics.org/competitions/world-athletics-championships/iaaf-world-athletics-championships-doha-2019-7125365/timetable/bydiscipline

Even in case of detection of deviations from normality use of the training grounds T test.