Practical work 7 : The t-test

• The t-test is the most widely used method to assess mean differences between one or two groups.

• There are several types of t-tests:

1. One-sample t-test
- Is the average height of female high school students greater than 1.68 meters?
2. Independent two-sample t-test
- Is the average height of female high school students radically different from the average height of male high school students?
3. Dependent two-sample t-test
- If you weigh high school students before and after taking a weight loss pill, is the average weight loss significant enough to conclude that this pill is effective?

Independent Samples t-Test

In this section, we will see how to test the null hypothesis using two means from two independent samples (or subgroups). We will estimate whether two population means are equal based on the comparison between these two samples. The technique employed is called the ***Independent Samples t-Test***.

This technique is used to compare TWO groups created by a categorical variable.

a) Premises of the independent t-test

- The data is normally distributed.
- The data is randomly sampled.
- The variances of the groups are equal (homogeneity of variance) (Levene's test).
- The groups are independent (the same observations cannot be in both groups).

b) Exemple

	Class	Statistics_mean
1	Science	13,00
2	Science	11,60
3	Science	15,20
4	Science	17,00
5	Science	18,00
6	Science	12,00
7	Science	14,00
8	Humanities	10,20
9	Humanities	10,00
10	Humanities	8,40
11	Humanities	8,00
12	Humanities	11,00
13	Humanities	8,00
14	Humanities	12,40
15	Humanities	12,00

- Are there any significant differences in the statistics module for students in the science class and the humanities class?

- **The homogeneity test (Levene's test)**
  1. Formulation of hypotheses

     Null hypothesis : $H_0: \sigma_1^2 = \sigma_1^2$

     The alternative hypothesis : $H_1: \sigma_1^2 \neq \sigma_1^2$

  2. Calculate F

$$F_{Levene} = \frac{(n-k)\sum_{i=1}^{k} n_i(\overline{w_i} - \overline{w})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(w_{ij} - \overline{w_i})^2}$$

As k represents the number of groups (k=2) and ni is the sample size for the ith group (i=1,2); n1 = 7; n2 = 8; n = n1 + n2

	x	x- $\overline{X}$	w_{ij}	$w_{ij} - \overline{w_i}$	$(w_{ij} - \overline{w_i})^2$	$(\overline{w_i} - \overline{w})$	$(\overline{w_i} - \overline{w})^2$
1	13	-1,4	1,4	-0,6	0,36	0,32	0,1024
1	11,6	-2,8	2,8	0,8	0,64	0,32	0,1024
1	15,2	0,8	0,8	-1,2	1,44	0,32	0,1024
1	17	2,6	2,6	0,6	0,36	0,32	0,1024
1	18	3,6	3,6	1,6	2,56	0,32	0,1024
1	12	-2,4	2,4	0,4	0,16	0,32	0,1024
1	14	-0,4	0,4	-1,6	2,56	0,32	0,1024
2	10,2	0,2	0,2	-1,2	1,44	-0,28	0,0784
2	10	0	0	-1,4	1,96	-0,28	0,0784
2	8,4	-1,6	1,6	0,2	0,04	-0,28	0,0784
2	8	-2	2	0,6	0,36	-0,28	0,0784
2	11	1	1	-0,4	0,16	-0,28	0,0784
2	8	-2	2	0,6	0,36	-0,28	0,0784
2	12,4	2,4	2,4	1	1	-0,28	0,0784
2	12	2	2	0,6	0,36	-0,28	0,0784
			25,2		13,76		1,344
	100,8/7=14,4		14		$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(w_{ij} - \overline{w_i})^2$		$\sum_{i=1}^{k} n_i(\overline{w_i} - \overline{w})^2$
	80/8=10		11,2				
	$\overline{X}$ 14,4		$\overline{w_1}$ 2				
	10		$\overline{w_2}$ 1,4				
			$\overline{w}$ 1,68				
V	6,01333333		2,45220989				
	3,05141505		1,74683				

3. Find the value of F(0.025,6,7) from the F-distribution table.

D.of Freedom, v (for replicates)	Degrees of Freedom, u (for treatments)								
	1	2	3	4	5	6	8	12	∞
1	161	200	216	225	230	234	239	244	254
2	18.5	19.0	19.2	19.3	19.3	19.3	19.4	19.4	19.5
3	10.1	9.6	9.3	9.1	9.0	8.9	8.8	8.7	8.5
4	7.7	6.9	6.6	6.4	6.3	6.2	6.0	5.9	5.6
5	6.6	5.8	5.4	5.2	5.1	5.0	4.8	4.7	4.4
6	6.0	5.1	4.8	4.5	4.4	4.3	4.2	4.0	3.7
7	5.6	4.7	4.4	4.1	4.0	3.9	3.7	3.6	3.2
8	5.3	4.5	4.1	3.8	3.7	3.6	3.4	3.3	2.9

4. The decision making:

We reject H_1 and accept H_0, therefore, the two populations are homogeneous.

- **Independent two-sample t-test**

Application of t-test to check for differences between means

1. Formulation of hypotheses:

Null Hypothesis : There is no difference between the means of the two groups in the population. In other words, the difference between the two population means is 0.

$$H_0 : \mu_1 = \mu_2$$

The alternative hypothesis : is that there is a difference between the two means.

$$H_1 : \mu_1 \neq \mu_2$$

2. Calculate t

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad t = \frac{14{,}4-10}{\sqrt{\frac{6{,}01333333*6+3{,}05142857*7}{13}\left(\frac{1}{7}+\frac{1}{8}\right)}} = 4{,}044506344$$

3. Find the tabular value

df	α = 0.05	α = 0.01	α = 0.05 $\frac{\alpha}{2}$ = 0.025	α = 0.01 $\frac{\alpha}{2}$ = 0.005
1	6.31	31.8	12.70	63.65
2	2.92	6.96	4.30	9.92
3	2.35	4.54	3.18	5.84
4	2.13	3.74	2.77	4.60
5	2.01	3.36	2.57	4.03
6	1.94	3.14	2.44	3.70
7	1.89	2.99	2.36	3.49
8	1.86	2.89	2.30	3.35
9	1.83	2.82	2.26	3.25
10	1.81	2.76	2.22	3.16
11	1.7 9	2.71	2.20	3.10
12	1.78	2.68	2.17	3.05
13	1.77	2.65	2.16	3.01
14	1.76	2.62	2.14	2.97
15	1.75	2.60	2.13	2.94

$$t(n_1+n_2-2,\alpha/2)=t(13,0.025)=2.16$$

$$4.044506344>2.16$$

4. The decision making

We accept H1 and reject H0, It is stated that there are differences between the mean of students in science and humanities classes in the statistics module.