Practical work 8

- Analysis of variance (ANOVA) aims to test significant differences between means.
- If we are comparing only two means, ANOVA will yield the same results as a t-test for independent samples (to compare two different groups of observations).
- With a single dependent variable and a discrete explanatory variable (also known as categorical or nominal, such as gender, socioeconomic category, etc.), we use one-way analysis of variance.
- With a single dependent variable and multiple discrete explanatory variables, we use two-way analysis of variance.
- When there are multiple dependent variables to be considered simultaneously, instead of conducting several analyses of variance (one per dependent variable), we perform a multiple analysis of variance (MANOVA).
- Assumptions of the analysis of variance test:
    - Groups are independent and randomly sampled from their respective populations.
    - Population values are normally distributed.
    - Population variances are equal.

Example

Groupe1	Groupe2	Groupe3
15	16	14
15,4	16,4	14,4
15,8	16,8	14,8
16,2	17,2	15,2
16,6	17,6	15,6

Are there significant differences in the mean scores of the statistics module among the three groups

1. ***Formulation of hypotheses:***

Null hypothesis: No significant differences in the statistics module among the three groups.

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

Alternative hypothesis: There are significant differences in the statistics module among the three groups.

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$$

2. ***Calculation of the function F:***
    a. Calculation of means
    b. Calculation of the total sum of squares (SSTot)

G1	G2	G3
15	16	14
15,4	16,4	14,4
15,8	16,8	14,8
16,2	17,2	15,2
16,6	17,6	15,6
les moyennes		
79/5=15.8	84/5=16.8	74/5=14.8

	x_i	$\overline{X_G}$	$(X_i - \overline{X_G})^2$
G1	15	15,8	0,64
M=15.8	15,4	15,8	0,16
	15,8	15,8	0
	16,2	15,8	0,16
	16,6	15,8	0,64
G2-G3	16	15,8	0,04
M=16.8	16,4	15,8	0,36
	16,8	15,8	1
	17,2	15,8	1,96
	17,6	15,8	3,24
G3	14	15,8	3,24
M=14.8	14,4	15,8	1,96
	14,8	15,8	1
	15,2	15,8	0,36
	15,6	15,8	0,04
		SSTot=	14,8

c. Calculation of the intra-variance (SSW - within sum of squares):

	x_i	$\overline{X_A}$	$(X_i - \overline{X_A})^2$
G1	15	15,8	0,64
M=15.8	15,4	15,8	0,16
SD=0.632455	15,8	15,8	0
	16,2	15,8	0,16
	16,6	15,8	0,64
G2-G3	16	16,8	0,64
M=16.8	16,4	16,8	0,16
SD=0.632455	16,8	16,8	0
	17,2	16,8	0,16
	17,6	16,8	0,64
G3	14	14,8	0,64
M=14.8	14,4	14,8	0,16
SD=0.632455	14,8	14,8	0
	15,2	14,8	0,16
	15,6	14,8	0,64
		SSW=	4,8

d. Calculation of the inter-variance (SSB - between sum of squares):

	$\overline{X_A}$	$\overline{X_G}$	$(\overline{X_A} - \overline{X_G})^2$
G1	15,8	15,8	0
M=15.8	15,8	15,8	0
	15,8	15,8	0
	15,8	15,8	0
	15,8	15,8	0
G2-G3	16,8	15,8	1
M=16.8	16,8	15,8	1
	16,8	15,8	1
	16,8	15,8	1
	16,8	15,8	1
G3	14,8	15,8	1
M=14.8	14,8	15,8	1
	14,8	15,8	1
	14,8	15,8	1
	14,8	15,8	1
		SSB=	10

e. The construction of the ANOVA table

La somme des carrés(SS)	Le degré de liberté	MS(Mean Squares)	Calcul F
SSB=10	k-1=2	MSB=SS/dd=10/2=5	F=MSB/MSW =5/0,4=12,5
SSW=4,8	N-k=15-3=12	MSW=4,8/12=0,4	
SSTot=14,8	N-1=14		

3. *Extract the tabular value*

DISTRIBUTION F (SEUIL DE SIGNIFICATION 5%)

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10
10	4.97	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.97	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.56	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38

4. *Decision making*

The F-value of 12.5 at significance level 0.05 (with degrees of freedom 2, 12) is greater than the critical value of 3.89. Therefore, we accept the alternative hypothesis (H_1) and reject the null hypothesis (H_0). There are significant differences among the means in the statistics module among the three groups.

The Post hoc tests

The analysis of variance test only informs us whether the null hypothesis is rejected or not. It does not specify where the differences lie among groups. Therefore, additional tests need to be conducted to identify these differences.

These tests are called ***post-hoc*** or a ***posteriori tests***. They specifically identify where the differences exist.

The Tukey test

$$HSD_{\alpha} = q_{\alpha}\sqrt{\frac{MS_W}{N_A}} = 3,77\sqrt{\frac{0,4}{5}} = 1,06631702$$

N_A is the sample size for each group. The value 3.77 is the critical value of the Tukey test at a significance level of α=0.05, with k=3 groups and degrees of freedom=12.

G1-G2	15.8-16.8=-1
G1-G3	15.8-14.8=1
G2-G3	16.8-14.8=2

So, 2 > 1.066317026, significant differences are related to the contrast between groups 2 and 3; the others do not indicate any statistically significant difference.