Université de Tlemcen
Faculté des Sciences
Département de Mathématiques
1ère Année LMD-MI
Année Universitaire 2021/2022.

Examen Final-Statistique descriptive-

Exercice 1 (Questions de cours) (8 Pts)

I) Soit X un caractère discret : $X\ \Omega \to \{x_1, \ldots\ldots, x_n\}$. (Avec $x_1 < x_2 < \cdots, < x_n$ et $12 \leq n \leq 100$).

On suppose que pour tout i, $f_i = \frac{1}{n}$, la fréquence partielle de la modalité x_i.

1) Montrer que si $x_{k_0} \notin [\bar{x} - 2\,\sigma_X , \bar{x} + 2\,\sigma_X]$ alors $|x_{k_0} - \bar{x}| \geq 2\,\sigma_X$ (1 Pt)

2) Montrer que si p modalités du caractère X sont dans l'intervalle $[\bar{x} - 2\,\sigma_X , \bar{x} + 2\,\sigma_X]$, alors (1,5 Pts)
$\sum_{k=1}^{n}(x_k - \bar{x})^2 \geq 4\,(n-p)\,\sigma_X{}^2$

3) En déduire qu'au moins trois quarts des modalités sont dans l'intervalle $[\bar{x} - 2\,\sigma_X , \bar{x} + 2\,\sigma_X]$ (1,5 Pts)

II) Pour $i = 1, \ldots, k$ et $j = 1, \ldots l$, le couple *de caractère* (S, Y) prend les valeurs (s_i, y_j) avec les fréquences f_{ij}. On pose $f_{i.} = \sum_{j=1}^{l} f_{ij}$ et $f_{.j} = \sum_{i=1}^{k} f_{ij}$

a) Que représente $f_{i.}$ pour le caractère S et $f_{.j}$ pour le caractère Y (0,5 Pt + 0,5 Pt)

b) Donner la définition de Covariance (S, Y) (1 Pt)

c) On suppose que $\forall\, i, j$; $f_{ij} = f_{i.} \cdot f_{.j}$. Montrer que Covariance $(S, Y) = 0$ (2 Pts)

NB : Les parties I) et II) sont indépendantes

Exercice 2 (6 Pts)

On note Y le caractère continu égal à la taille (en cm) d'une population d'adolescents. Les valeurs prises par Y réparties en classes avec leurs effectifs cumulés respectifs, sont données par le tableau suivant :

Classe C_i	[140 ; 150[	[150 ; 160[	[160 ; 170[	[170 ; 180[
N_i	20	35	75	100

1) Compléter le tableau : (1 Pt)

Classe C_i	n_i	f_i	F_i	$f_i\,c_i$

2) Déterminer la classe modale C_{M_o} et calculer le mode M_o du caractère Y. (0,5 Pt + 1 Pt)

3) Calculer les bornes de l'intervalle $[Q_1 , \bar{y}]$ (1 Pts + 0,5 Pt)

4) En déduire le nombre d'adolescents dont la taille n'est pas dans l'intervalle. $[Q_1 , \bar{y}]$ (2 Pts)

Exercice 3 (6 Pts)

Les valeurs (z_i, t_j) prises par couple de caractère (Z, T) avec leurs fréquences f_{ij} sont données par le tableau suivant :

z_i \ t_j	-1	1	3	5
0	0,12	0,24	0,18	0,06
2	0,04	α	0,06	0,02
3	0,04	0,08	0,06	0,02

1) Déterminer la valeur de α (0,5 Pt)

2) Calculer les moyennes marginales $\bar{z}$ et $\bar{t}$ des caractères marginaux Z *et* T. (1,5 Pt +1,5 Pt)

3) Calculer la covariance du couple (Z, T) (1,5 Pt)

4) Calculer ρ, le coefficient de corrélation linéaire du couple de caractère (Z, T). (1 Pt)

- Juin 22 -

_ Corrigé de l'examen final _ LMD - MI -
\- Module de Statistique descriptive -

(8 pts)

Exercice 1 : Questions de cours

I/ 1°) $x_{k_0} \notin [\bar{x}-2\sigma_x, \bar{x}+2\sigma_x] \Rightarrow x_{k_0} \leq \bar{x}-2\sigma_x$ ou $x_{k_0} \geq \bar{x}+2\sigma_x$

$\Rightarrow |x_{k_0}-\bar{x}| \geq 2\sigma_x$ (1 pt)

$x_{k_0}-\bar{x} \geq 2\sigma_x$ ou $\Rightarrow \bar{x}-x_{k_0} \geq 2\sigma_x$

2°) $$\sum_{k=1}^{n}(x_k-\bar{x})^2 = \sum_{k:\, x_k \in [\bar{x}-2\sigma_x;\bar{x}+2\sigma_x]}(x_k-\bar{x})^2 + \sum_{k:\, x_k \notin [\bar{x}-2\sigma_x;\bar{x}+2\sigma_x]}(x_k-\bar{x})^2$$

$$\geq \sum_{k:\, x_k \notin [\bar{x}-2\sigma_x;\bar{x}+2\sigma_x]}(x_k-\bar{x})^2 \geq \sum_{k:\, x_k \notin [\bar{x}-2\sigma_x;\bar{x}+2\sigma_x]} 4\sigma_x^2$$

or il y a $(n-p)$ modalités hors de l'intervalle $[\bar{x}-2\sigma_x;\bar{x}+2\sigma_x]$

d'où $$\sum_{k=1}^{n}(x_k-\bar{x})^2 \geq 4(n-p)\sigma_x^2$$ (1,5 pt)

3°) Par ailleurs, on a $$\sum_{k=1}^{n}(x_k-\bar{x})^2 = n\left[\frac{1}{n}\sum_{k=1}^{n}(x_k-\bar{x})^2\right] = n\sigma_x^2$$

d'où $n\sigma_x^2 \geq 4(n-p)\sigma_x^2 \Rightarrow p \geq \frac{3}{4}n$. (1,5 pt)

II/ a) $f_{i.}$ est la fréquence de s_i (0,5 pt) ; $f_{.j}$ est la fréquence de y_j (0,5 pt)

b) $\mathrm{Cov}(S,Y) = \sum_i \sum_j f_{ij} s_i y_j - \bar{s}\,\bar{y}$ (1 pt)

c) On suppose que $f_{ij} = f_{i.} \cdot f_{.j}$ et on sait que :

$\bar{s} = \sum f_{i.} s_i$ et $\bar{y} = \sum f_{.j} y_j$

d'où $\mathrm{Cov}(S,Y) = \sum_i \sum_j f_{ij} s_i y_j - \bar{s}\,\bar{y} = \sum_i f_{i.} s_i \sum_j f_{.j} y_j - \bar{s}\,\bar{y}$

$\Rightarrow \bar{s}\,\bar{y} - \bar{s}\,\bar{y} = 0 \Rightarrow \mathrm{Cov}(S,Y) = 0$ (2 pts)

Exercice 2 : (6 pts)

1) complétons le tableau (1 pt)

classe	[140 ; 150[	[150 ; 160[	[160 ; 170[	170 ; 180[
n_i	20	15	40	25
$f_i=\frac{n_i}{N}$	0,20	0,15	0,40	0,25
F_i	0,20	0,35	0,75	1,00
f_iC_i	145x0,20=29	23,25	66	43,75

$\sum n_i = 100 = N$

C_i est le centre de la Classe C_i

2.) $C_{M_o} = [160 ; 170[$ (car elle a la plus grande frequence partielle) (0,5 pt)

D'après le cours $M_o = a_i + h\,\frac{\Delta_1}{\Delta_1+\Delta_2}$ (0,5 pt)

où $a_i = 160$; $\Delta_1 = 0,25$; $\Delta_2 = 0,15$ et $h = 10$

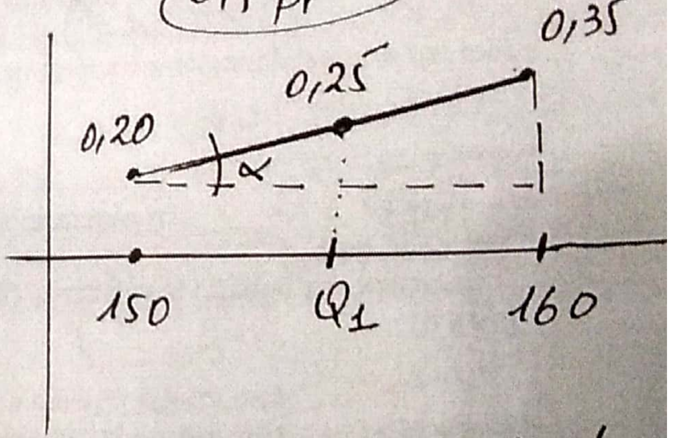
d'où $\underline{M_o = 166,25}$ (0,5 pt)

3°) $\bar{y} = \sum f_iC_i = \underline{162}$ (0,5 pt)

d'après le tableau ($F_X(Q_1) = 0,25$)

or $Q_1 \in [150 ; 160[$ (0,5 pt)

On a $tg\alpha = \frac{0,35-0,20}{10} = \frac{0,25-0,20}{Q_1-150}$



$\Rightarrow \underline{Q_1 = 153,33}$ (0,5 pt)

d'où $[Q_1 ; \bar{y}] = [153,33 ; 162]$

4°) On sait que le pourcentage des adolescents dont ~~le poids~~ taille est dans l'intervalle $[Q_1 ; \bar{y}]$ est égal à $F_X(\bar{y}) - F_X(Q_1)$ (0,5 pt)

Calculons $F_X(162)$

Comme $162 \in [160 ; 170[\Rightarrow F_X(162) = 0,35 + \frac{0,40}{10}(162-160)$

$= 0,43$ (0,5 pt)

$\Rightarrow 0,43 - 0,25 = 0,18 =$ % des adolescents dont la taille est dans l'intervalle $[Q_1 ; \bar{y}]$ (0,5 pt)

conclusion : le pourcentage des adolescents dont la taille n'est pas dans l'intervalle $[Q_1 ; \bar{y}]$ est 82%. (0,5 pt)
$\Rightarrow$ il y a 82 adolescents

Exercice 3 (6 pts)

1) on sait $\sum_i \sum_j f_{ij} = 1 \Rightarrow \alpha = 0{,}08$ (0,5 pt)

2°) on a

z_i	0	2	3
$f_{i\cdot}$	0,60	0,20	0,20

$\bar{z} = \sum f_{i\cdot} z_i = 1$. (1,5 pt)

et

t_j	-1	1	3	5
$f_{\cdot j}$	0,20	0,40	0,30	0,10

$\bar{t} = \sum f_{\cdot j} t_j = 1{,}6$ (1,5 pt)

3°) $cov(Z, T) = \sum f_{ij} z_i t_j - \bar{z}\bar{t}$ (0,5 pt)

$\Rightarrow cov(Z,T) = \Big[0[(-1\times 0{,}12) + \dots + (5\times 0{,}06)] + 2[(-1\times 0{,}04) + (1\times 0{,}08) + (3\times 0{,}06) + (5\times 0{,}02)] + 3[(-1\times 0{,}04) + (1\times 0{,}08) + (3\times 0{,}06) + (5\times 0{,}02)] \Big] - (1\times 1{,}6)$

$\Rightarrow cov(Z,T) = [0 + 2(0{,}32) + (3\times 0{,}32)] - 1{,}6$

$= 1{,}6 - 1{,}6 \Rightarrow cov(Z,T) = 0$ (1 pt)

4°) or $\rho_{X,Y} = \dfrac{cov(Z,T)}{\sigma_T \cdot \sigma_Z}$ (0,5 pt)

$\Rightarrow \rho_{X,Y} = 0$ (0,5 pt)

Scanné avec CamScanner