

Université Batna 2

Faculté Mathématiques-Informatique (Année 2021-2022)
1^{re} Année LMD

Module : probabilité et statistique

TD1 (statistique descriptive)

Exercice :

i. Les résultats en module algèbre 1, d'un groupe de 40 étudiants étaient :

4	9	11	14	6	2	2	4	15	14
11	4	6	9	4	4	15	9	6	15
13	11	11	11	13	10	10	10	10	13
9	9	9	9	6	6	6	10	10	6

- 1- Indiquer la population statistique et l'unité statistique.
 - 2- Quel est le caractère étudié ? et quel est sa nature ?
 - 3- Dresser le tableau statistique correspond.
 - 4- Quelle est la représentation graphique adéquate ?
 - 5- Quelles est la proportion d'étudiants n'ayant pas eu la moyenne ?
 - 6- Calculer les proportions suivantes : $P(X < 2)$, $P(10 \leq X < 14)$, $P(X \geq 9)$.
 - 7- Quel est le mode de cette série statistique ?
 - 8- En utilisant les fréquences relatives cumulées en %, calculer la médiane de cette série statistique ?
 - 9- Calculer la moyenne arithmétique en fonction des fréquences absolues, ensuite en fonction des fréquences relatives.
 - 10- Est-ce qu'on peut dire que cette moyenne donne une idée juste sur la répartition des notes dans ce groupe ?
 - 11- Calculer La variance et l'écart type.
 - 12- Si les notes au second semestre du module algèbre 2 augmenteraient de 10% par rapport des notes du module algèbre 1, quel serait le taux d'augmentation de la moyenne du module algèbres 2 ? Conclure la valeur de la variance à partir de la question 11.
 - 13- Représenter les données du tableau par un diagramme en bâtons et la courbe des fréquences relatives cumulées croissantes en % des notes des étudiants du module algèbre 1.
- ii. En regroupant les notes du module algèbre 1 dans les intervalles de 3 points, on obtient une série statistique exprimée en classes.
- 1) Dresser le tableau statistique correspond.
 - 2) Dans ce cas, indiquer la population statistique, l'unité statistique, le caractère étudié et sa nature.
 - 3) Quelle est la représentation graphique adéquate ?
 - 4) En utilisant les fréquences relatives cumulées en %, calculer la médiane de cette série statistique ?

- 5) Calculer la moyenne en fonction des fréquences absolues, ensuite en fonction des fréquences relatives.
- 6) Calculer l'écart type.
- 7) En utilisant les fréquences relatives cumulées en %, calculer la médiane de cette série statistique ? et quel est votre commentaire?
- 8) Quelles est la proportion d'étudiants n'ayant pas eu la moyenne ?
- 9) Calculer les proportions suivantes : $P(X < 2)$, $P(10 \leq X < 14)$, $P(X \geq 9)$.
- 10) Tracer l'histogramme, le polygone des fréquences et la courbe des FCC%.

Solution

1) 1 et 2) Population statistique, unité statistique, caractère et sa nature

Population statistique	Unité statistique	Caractère	Nature du caractère
Groupe de 40 étudiants	L'étudiant	Note du module Algèbre 1	Quantitatif discret

3) Le tableau

Notes	2	4	6	9	10	11	13	14	15	La somme
Nombre d'étudiants	2	5	7	7	6	5	3	2	3	40

4) la représentation graphique adéquate c'est « le diagramme en bâton », car les notes sont présentées sous forme de valeurs entières (isolées).

5) La proportion d'étudiants n'ayant pas eu la moyenne est :

REMARQUE pour la calculer il faut calculer la calculer la fréquence cumulé croissante

donc

Notes (x_i)	2	4	6	9	10	11	13	14	15	Total
Nombre d'étudiants (n_i)	2	5	7	7	6	5	3	2	3	N=40
Les fréquences relatives (f_i)	0.050	0.125	0.175	0.175	0.150	0.125	0.075	0.050	0.075	1
F.C.C ($F_i \uparrow$)	0.050	0.175	0.350	0.525	0.675	0.800	0.875	0.925	1	

$$f_i = \frac{n_i}{N} \text{ où } 0 \leq f_i \leq 1 \text{ et } \sum_{i=1}^p f_i = 1,$$

$$N = \sum_{i=1}^p n_i \text{ où } N : \text{ effectif total,}$$

F.C.C : fréquence cumulé croissante,

$$P(X < 10) = f_1 + f_2 + f_3 + f_4 = 0.05 + 0.125 + 0.175 + 0.175 = 0.525,$$

6) -La proportion $P(X < 2)$:

$$P(X < 2) = 0,$$

- La proportion $P(10 \leq X < 14)$

$$P(10 \leq X < 14) = P(X < 14) - P(X < 10) = 0.875 - 0.525 = 0.350,$$

-La proportion $P(X \geq 9)$

$$P(X \geq 9) = P(X \leq 15) - P(X < 9) = 1 - (f_1 + f_2 + f_3) = 0.650.$$

7) Le mode de cette série statistique (**Mo**)

D'après le tableau on a la plus grande des valeurs de l'effectif est **7**, on a deux modalités qui correspondent à cet effectif donc on peut dire c'est une série statistique « **bimodale** » et les modes sont : **Mo=6** et **Mo=9**.

8) Pour calculer le médiane (**Me**), on remarque que c'est une série statistique discrète donc la méthode utilisé dans ce cas c'est « **l'effectif cumulé croissant** », donc :

Notes (x_i)	2	4	6	9	10	11	13	14	15	Total
Nombre d'étudiants(n_i)	2	5	7	7	6	5	3	2	3	N=40
E.C.C	2	7	14	21	27	32	35	37	40	

E.C.C : effectif cumulé croissant

La première étape c'est de trouver le rang de la médiane, dans ce cas **N=40 (pair)**

$$Me \in [x_{\frac{40}{2}}, x_{\frac{40}{2}+1}], Me = \frac{x_{\frac{40}{2}} + x_{\frac{40}{2}+1}}{2} = \frac{x_{20} + x_{21}}{2} = \frac{9+9}{2} = 9,$$

9) Calculer la moyenne arithmétique en fonction des fréquences absolues, ensuite en fonction des fréquences relatives :

$$\bar{X} = \sum_{i=1}^p \frac{n_i x_i}{N} = \sum_{i=1}^p f_i x_i$$

Notes (x_i)	2	4	6	9	10	11	13	14	15	Total
Nombre d'étudiants(n_i)	2	5	7	7	6	5	3	2	3	N=40
$n_i x_i$	4	20	42	63	60	55	39	28	45	356
f_i	0.050	0.125	0.175	0.175	0.150	0.125	0.075	0.050	0.075	1
$x_i f_i$	0.1	0.5	1.05	1.575	1.5	1.375	0.975	0.7	1.125	8.9

Donc : $\bar{X} = 8.9$

10) La moyenne donne une idée juste sur la répartition des notes dans ce groupe, notre réponse est « **non** ».

La moyenne des notes c'est $8.9 \approx 9$, on remarque que la moyenne nous donne pas une idée sur les valeurs extrêmes de notre série, mais on remarque que la valeur de la moyenne pondérée et la médiane sont proche qui nous donne une idée sur la valeur centrale.

11) La variance et l'écart type :

- La variance $v(x)$

$$\begin{aligned} v(x) &= \sum_{i=1}^p \frac{n_i (x_i - \bar{X})^2}{N} = \sum_{i=1}^p f_i (x_i^2 - 2x_i \bar{X} + \bar{X}^2) \\ &= \sum_{i=1}^p f_i x_i^2 - 2 \sum_{i=1}^p f_i x_i \bar{X} + \sum_{i=1}^p f_i \bar{X}^2 \\ &= \sum_{i=1}^p f_i x_i^2 - 2\bar{X}^2 + \bar{X}^2 \sum_{i=1}^p f_i = \bar{X}^2 - \bar{X}^2 \end{aligned}$$

Tel que : $\bar{X}^2 = \sum_{i=1}^p f_i x_i^2, \sum_{i=1}^p f_i = 1$

Notes (x_i)	2	4	6	9	10	11	13	14	15	Total
Nombre d'étudiants (n_i)	2	5	7	7	6	5	3	2	3	N=40
$n_i x_i$	4	20	42	63	60	55	39	28	45	356
f_i	0.050	0.125	0.175	0.175	0.150	0.125	0.075	0.050	0.075	1
$x_i f_i$	0.1	0.5	1.05	1.575	1.5	1.375	0.975	0.7	1.125	8.9
$x_i^2 f_i$	0.2	2	6.3	14.175	15	15.125	12.675	9.8	16.875	92.15

$$v(x) = 92.15 - (8.9)^2 = 12.64.$$

- Ecart type $\sigma(x) = \sqrt{v(x)} \approx 3.5552$

12) Le taux d'augmentation :

$$\bar{Y} = \sum_{i=1}^p \frac{n_i(x_i + 0.1x_i)}{N} = 1.1 \sum_{i=1}^p \frac{n_i x_i}{N} = 1.1 \bar{X} = 1.1 * 8.9 = 9.79.$$

-Conclure la variance :

$$v(y) = \sum_{i=1}^p f_i y_i^2 - \bar{Y}^2 = \sum_{i=1}^p f_i (x_i + 0.1x_i)^2 - (9.79)^2 = 1.1^2 \sum_{i=1}^p f_i x_i^2 - (9.79)^2$$

$$= 1.1^2 * \bar{X}^2 - (9.79)^2 = 15.6574$$

II) On va avoir une nouvelle série statistique quantitatif continue

Le tableau

classes	[2,5[[5,8[[8,11[[11,14[[14,17[total
n_i	7	7	13	8	5	40
Centre des classes C_i	3.5	6.5	9.5	12.5	15.5	
$n_i C_i$	24.5	45.5	123.5	100	77.5	371
f_i	0.175	0.175	0.325	0.200	0.125	1
F.C.C ($F_i \uparrow$)	0.175	0.350	0.675	0.875	1	
F.C.C $\%(F_i \uparrow)$	17.5	35	67.5	87.5	100	
F.C.D ($F_i \downarrow$)	1	0.825	0.650	0.325	0.125	

3) La représentation graphique adéquate, c'est l'histogramme et polygone des fréquences.

4) En utilisant les fréquences relatives cumulées en %, calculer la médiane de cette série statistique (l'interpolation linéaire):

La position de la médiane c'est 50% donc $Me \in [8, 11]$, donc

8	35
Me	50
11	67.5

$$\frac{Me - 8}{11 - 8} = \frac{50 - 35}{67.5 - 35}$$

$$\text{Donc : } Me = 3 \left(\frac{15}{32.5} \right) + 8 = 9.385$$

On remarque que la valeur de la médiane calculée à partir du tableau statistique exprimé en classes est approximativement égale à celle calculée avec les données discrètes. Cette légère différence est due au fait qu'on ignore la distribution effective des données à l'intérieur de chaque classe et on suppose qu'elles sont distribuées uniformément.

- 5) La moyenne en fonction des fréquences absolues, ensuite en fonction des fréquences relatives.

$$\bar{z} = \sum_{i=1}^p \frac{n_i C_i}{N} = \sum_{i=1}^p f_i C_i = 9.275$$

Calculer l'écart type (pour calculer l'écart type il faut qu'on calcule la variance)

$$v(z) = \sum_{i=1}^p \frac{n_i (C_i - \bar{z})^2}{N} = \bar{z}^2 - \bar{z}^2$$

$$\text{Tel que : } \bar{z}^2 = \sum_{i=1}^p f_i C_i^2, \sum_{i=1}^p f_i = 1$$

$$\text{Ecart type } \sigma(x) = \sqrt{v(z)}$$

- 6) La proportion d'étudiants n'ayant pas eu la moyenne

$$P(X < 10) = f_1 + f_2 + \alpha f_3$$

$$\text{Où : } \alpha = \frac{10-8}{11-8} = \frac{2}{3}$$

Donc

$$P(X < 10) = 0.350 + \frac{2}{3} * 0.325 \approx 0.566$$

- 7) Calculer les proportions suivantes : $P(X < 2)$, $P(10 \leq X < 14)$, $P(X \geq 9)$.

- 8) Le mode, on remarque que c'est une série avec classe uniforme (dans le cas contraire on va utiliser l'effectif rectifier) on a la plus grande des effectif **13** la classe modale c'est : **[8,11]**, par conclusion $Mo = \frac{8+11}{2} = 9.5$ (le mode c'est le centre de la classe)