

T.D. XII - Statistiques descriptives

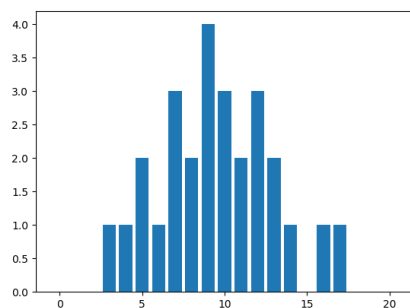
I - Séries statistiques à un caractère

Solution de l'exercice 1.

1. Représentons les notes dans un tableau d'effectifs

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	1	1	2	1	3	2	4	3	2	3	2	1	0	1	1	0	0	0

2. En utilisant le tableau précédent, on obtient le diagramme en bâtons :



3. La modalité 4 a un effectif de 1.

La classe $[10, 14[$ a un effectif de $3 + 2 + 3 + 2 = 10$.

La classe $[0, 5[$ a un effectif de $0 + 0 + 0 + 1 + 1 + 2 = 4$.

4. Le tableau des effectifs cumulés est le suivant :

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	1	2	4	5	8	10	14	17	19	22	24	25	25	26	27	27	27	27

On obtient la courbe suivante :



5. L'effectif total étant de 27 étudiants, on obtient la courbe des fréquences :

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{3}{27}$	$\frac{2}{27}$	$\frac{3}{27}$	$\frac{2}{27}$	$\frac{1}{27}$	0	$\frac{1}{27}$	$\frac{1}{27}$	0	0	0

Ainsi que la courbe des fréquences cumulées :

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{5}{27}$	$\frac{8}{27}$	$\frac{10}{27}$	$\frac{14}{27}$	$\frac{17}{27}$	$\frac{19}{27}$	$\frac{22}{27}$	$\frac{24}{27}$	$\frac{25}{27}$	$\frac{25}{27}$	$\frac{26}{27}$	1	1	1	1

6. La valeur modale est le caractère dont l'effectif est le plus important. Ainsi, la valeur modale est 4.

7. La valeur médiane est la valeur m pour laquelle le nombre d'élèves ayant eu strictement moins de m et le nombre d'élèves ayant eu strictement plus que m sont inférieurs à $\frac{27}{2}$. C'est la valeur pour laquelle les fréquences cumulées dépassent 0,5. Ainsi, la médiane vaut 9.

8. En utilisant un outil pour effectuer les calculs, on obtient que

- * la moyenne vaut environ 9,52,
- * la variance vaut environ 11,73,
- * l'écart-type vaut environ 3,42.

□

Solution de l'exercice 2.

1. La population est un ensemble de 40 téléviseurs.
Le caractère étudié est le nombre de pannes. Il s'agit d'un caractère quantitatif.

2. On obtient le diagramme en bâtons suivant :



3. Les valeurs modales sont les valeurs ayant l'effectif le plus important. Ici, les valeurs modales sont 1 et 2.

4. En utilisant un outil de calcul on obtient :

* La moyenne vaut $\frac{11+2 \times 11+3 \times 8+4 \times 4+5}{40} \simeq 1,95$.

* L'écart-type vaut

$$\sqrt{\frac{5 \cdot 1.95^2 + 11 \cdot 0.95^2 + 11 \cdot 0.05^2 + 8 \cdot 1.05^2 + 4 \cdot 2.05^2 + 3 \cdot 0.5^2}{40}} \simeq 1,26.$$

5. Le tableau des effectifs cumulés est :

0	1	2	3	4	5
5	16	27	35	39	40

Le tableau des fréquences cumulées est :

0	1	2	3	4	5
$\frac{5}{40}$	$\frac{16}{40}$	$\frac{27}{40}$	$\frac{35}{40}$	$\frac{39}{40}$	1

6. La médiane est égale au plus petit caractère pour lequel la fréquence dépasse 0,5. Ainsi, la médiane vaut 2.

7. a) Le pourcentage de téléviseurs ayant eu entre 1 et 3 pannes est égal à

$$\frac{11 + 11 + 8}{40} = \frac{3}{4} = 75\%.$$

b) Le pourcentage de téléviseurs ayant eu au moins une panne est égal à

$$\frac{40 - 5}{40} = \frac{7}{8} = 87,5\%.$$

□

II - Séries statistiques à deux caractères

Solution de l'exercice 3.

1. À l'aide d'un logiciel de calcul, on obtient

- * la moyenne m_X de l'âge : $m_X \simeq 4,43$,
- * la variance de l'âge : $\sigma_X^2 \simeq 6,10$,
- * la moyenne du coût de maintenance : $m_Y \simeq 13,22$,
- * sa variance : $\sigma_Y^2 \simeq 32,84$,
- * la covariance entre X et Y : $\sigma_{X,Y} \simeq 11,83$.

On obtient ainsi une droite de régression linéaire d'équation

$$y = \frac{\sigma_{X,Y}}{\sigma_X^2}(x - m_X) + m_Y = 1,93(x - 4,43) + 13,22.$$

2. Le coefficient de corrélation est égal à

$$r = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \simeq 0,83.$$

□

Solution de l'exercice 4.

1. À l'aide d'un logiciel de calcul, on obtient

- * la moyenne m_t du temps : $m_X \simeq 5,5$,
- * la variance du temps : $\sigma_t^2 \simeq 8,25$,
- * la moyenne du nombre de nuitées : $m_N \simeq 33,1$,
- * sa variance : $\sigma_N^2 \simeq 34,5$,
- * la covariance entre t et N : $\sigma_{t,N} \simeq 15,6$.

On obtient ainsi une droite de régression linéaire d'équation

$$y = \frac{\sigma_{X,Y}}{\sigma_X^2}(x - m_X) + m_Y = 1,9(x - 5,5) + 33,1.$$

2. Le coefficient de corrélation est égal à

$$r = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \simeq 0,92.$$

□