

Chapitre I. Notion sur les Statistiques

Introduction

La statistique est la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes aléatoires, c'est-à-dire dans lesquels le hasard intervient.

L'analyse des données est utilisée pour d'écrire les phénomènes étudiés, faire des prévisions et prendre des décisions à leur sujet.

En cela, la statistique est un outil essentiel pour la compréhension et la gestion des phénomènes complexes.

Les données étudiées peuvent être de toute nature, ce qui rend la statistique utile dans tous les champs disciplinaires et explique pourquoi elle est enseignée dans toutes les livres universitaires, de l'économie à la biologie en passant par la psychologie, et bien sur les sciences de l'ingénieur.

1. Analyse statistique

Une série statistique est constituée par l'ensemble des valeurs du caractère étudié.

En général, les valeurs ne sont pas organisées donc pour l'analyse il faut procéder comme suit:

- Ranger les valeurs soit par ordre croissant soit par ordre décroissant. La différence entre la plus grande valeur et la plus petite valeur est appelée Amplitude de la série.
- Une valeur est inscrite qu'une seule fois
- On indique le nombre de fois où l'on observe (fréquence absolue n_i)
- On calcule la fréquence relative $f_r = f_a/N$
- On détermine l'amplitude
- On détermine le nombre d'intervalles $k = 1 + (10 \cdot \log N)/3$

2. Histogramme

C'est la représentation graphique de la fonction de répartition.

Un histogramme est une série de rectangles ayant sur l'axe des x les milieux de classe et sur l'axe des y (hauteur) les fréquences (absolue et relative)

3. Courbe de fréquence cumulées ou fonction de répartition

La fonction de distribution d'une variable est composée par l'ensemble de couple (x_i, n_i) ou (x_i, f_i) .

4. Paramètres de position

- Le mode : c'est la valeur dont la fréquence est la plus grande qui se répète le plus souvent

- La moyenne arithmétique X_a

$$\overline{X}_a = \frac{(\sum x_i)}{N}$$

- La moyenne géométrique X_g

$$\overline{X}_g = (x_1 * x_2 * \dots * x_N)^{1/N} = (\prod x_i)^{1/N} = 1/N \sum \ln x_i$$

- La moyenne harmonique X_h

$$\overline{X}_h = N / (\sum 1/x_i)$$

- La moyenne Harmonique X_q

$$\overline{X}_q = \sqrt{\frac{\sum x_i^2}{N}}$$

5. Hiérarchie des moyennes

On appelle moyenne d'ordre α la valeur X_α telle que

$$(X_\alpha)^\alpha = \frac{X_1^\alpha + X_2^\alpha + \dots + X_N^\alpha}{N} = \sum \frac{X_i^\alpha}{N}$$

Si

$\alpha=2$ moyenne quadratique

$\alpha= 1$ moyenne arithmétique

$\alpha = -1$ moyenne harmonique

Si

$\alpha \xrightarrow{\text{tend}} 0$, on montre que la limite est la moyenne géométrique

La hiérarchie des moyennes est donc celle des valeurs de α

$$\overline{X}_h < \overline{X}_g < \overline{X}_a < \overline{X}_q$$

6. La médiane

$$\text{Médiane} = L_1 + \left(\frac{(N/2) - \sum f_i}{f_{\text{médiane}}} \right) * C$$

Où

L_1 = limite inférieure de la classe médiane,

N : nombre de valeur dans l'échantillon,

$\sum f_i$: somme des fréquences absolues de toutes les classes inférieures à la classe médiane,

$f_{\text{médiane}}$: fréquence de la classe médiane

C : grandeur de la classe médiane.

7. Les paramètres de dispersion

- La variance S^2

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{N-1}$$

- L'écart type S est égal à la racine carrée de la variance
- Le coefficient de variation C_v

$$C_v = \frac{S}{\bar{X}}$$

Exemple d'application

Un certain nombre de valeur qui représente des débits maximums annuels d'un oued exprimé en m^3/s ,

Tableau 1. Débits annuels

Années	Q (m^3/s)								
1965	28	1970	36	1975	44	1980	77	1985	58
1966	37	1971	69	1976	49	1981	77	1986	39
1967	52	1972	99	1977	53	1982	59	1987	38
1968	34	1973	77	1978	58	1983	54	1988	103
1969	44	1974	62	1979	64	1984	49	1989	47

Tableau 2. Fréquences absolues et relatives

Variable X_i	f_a	f_r	Variable X_i	F_a	F_r
28	1	0.04	53	1	0.04
34	1	0.04	54	1	0.04
36	1	0.04	58	2	0.08
37	1	0.04	59	1	0.04
38	1	0.04	62	1	0.04
39	1	0.04	64	1	0.04
44	2	0.08	69	1	0.04
47	1	0.04	77	3	0.12
49	2	0.08	99	1	0.04
52	1	0.04	103	1	0.04

Tableau 3. Centres de classe

Numéro de classe (i)	Bornes de classes	Centre de classe (x_i)	f_a	f_r
1	$25.5 < X < 37.5$	31.5	4	0.16
2	$37.5 < X < 49.5$	43.5	7	0.28
3	$49.5 < X < 61.5$	55.5	6	0.24
4	$61.5 < X < 73.5$	67.5	3	0.12
5	$73.5 < X < 85.5$	79.5	3	0.12
6	$85.5 < X < 97.5$	91.5	0	0
7	$97.5 < X < 109.5$	13.5	2	0.08

Histogramme

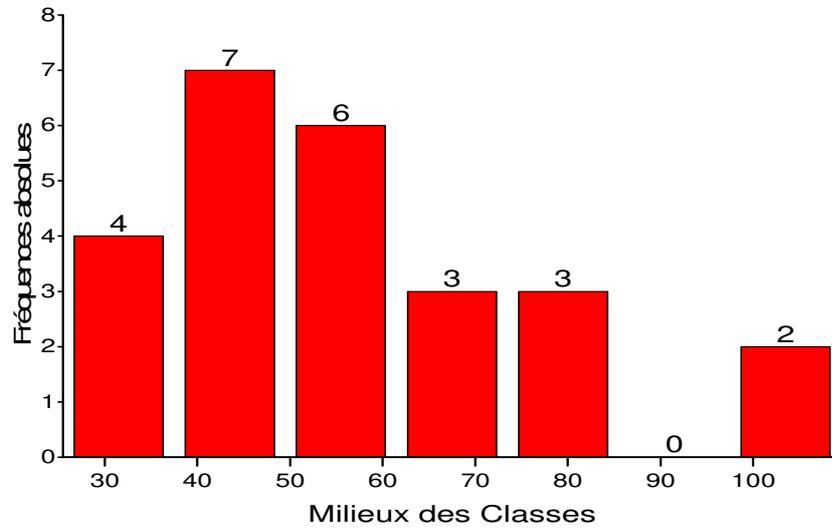


Fig 1. Histogramme

Tableau 4. Fréquences au non dépassement (FND) et Fréquences au dépassement (FD)

Q (m ³ /s)	Effectif cumulé	FND	Q (m ³ /s)	Effectif cumulé	FD
<25.5	0	0	>25.5	25	1
<37.5	4	0.16	>37.5	21	0.84
<49.5	11	0.44	>49.5	14	0.56
<61.5	17	0.68	>61.5	8	0.32
<73.5	20	<u>0.80</u>	>73.5	5	<u>0.20</u>
<85.5	23	0.92	>85.5	2	0.08
<97.5	23	0.92	>97.5	2	0.08
<109.5	25	1	>109.5	0	0

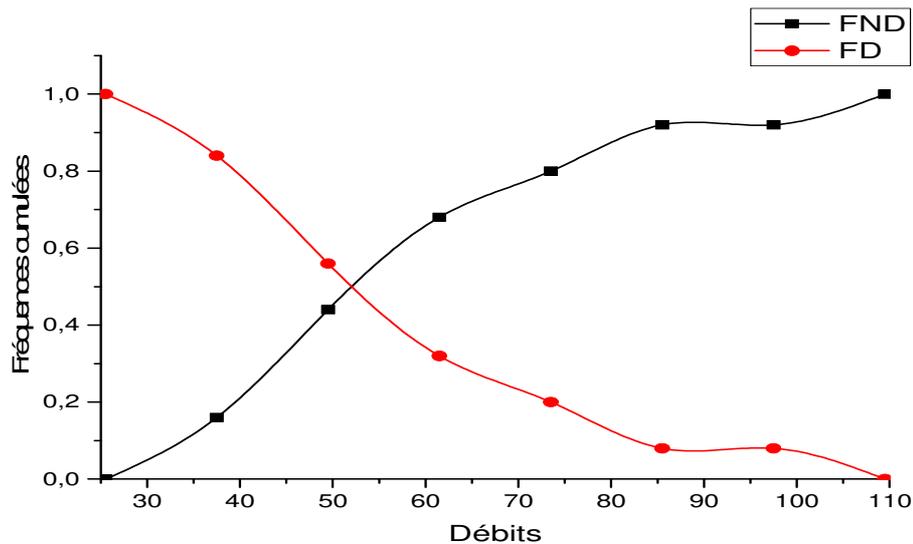


Fig 2. Courbe des fréquences cumulées

Description Statistique

	Valid N	Mean	Confidence – 95%	Confidence - 95%	Geometric - Mean	Harmonic - Mean
Débits	25	56,28	48,37451	64,18549	53,40295	50,76123
Median	Sum	Minimum	Maximum	Variance	Std.Dev.	Coef.Var.
53	1407	28	103	366,7933	19,15185	34,02958

Bibliographie

Abdelwaheb Sari Ahmed., Hydrologie de surface, 2eme édition, avril 2002.

Dagnelie P., Statistique théorique et appliquée, 2eme édition, De Boeck Université, 2007.

Montgomery D.C., Runger G.C., Applied Statistics and Probability for Engineers, 4eme édition, Wiley, 2007.

Saporta G., Probabilités, analyse de données et statistique, 2eme édition, Technip, 2006.