UNIVERSITE MOHAMED KHIDER-BISKRA
DEPARTEMENT DE GENIE CIVIL ET D'HYDRAULIQUE

NOTIONS DE STATISTIQUES

Analyse fréquentielle: "Approche classique"

Le tableau ci-dessous représente une série de 60 pluies annuelles (mm).

An	Pluie	An	Pluie	An	Pluie	An	Pluie	An	Pluie	An	Pluie
1922	626	1932	161	1942	358	1952	510	1970	404	1980	159
1923	411	1933	443	1943	388	1953	386	1971	545	1981	414
1924	537	1934	576	1944	562	1954	350	1972	321	1982	440
1925	553	1935	737	1945	371	1955	509	1973	680	1983	475
1926	472	1936	661	1946	274	1956	507	1974	606	1984	238
1927	579	1937	648	1947	722	1957	559	1975	547	1985	380
1928	[illegible]	1938	701	1948	707	1958	310	1976	583	1986	835
1929	[illegible]	1939	496	1949	522	1959	519	1977	448	1987	350
1930	[illegible]	1940	455	1950	650	1960	575	1978	416	1988	553
1931	[illegible]	1941	473	1951	488	1969	916	1979	682	1989	823

- avec un intervalle de classe égal à 100mm:

Construire l'histogramme des fréquences

Construire la courbe des fréquences cumulées

Calculer la médiane M par :

- la formule suivante:

$$\boxed{\frac{M-a}{b-a} = \frac{N/2 - ECC_1}{n}}$$

- Interpolation

Tableau 1: Série de pluie ordonnée par ordre croissant

Rang	Pluie	Rang	Pluie	Rang	Pluie	Rang	Pluie	Rang	Pluie	Rang	Pluie
1	159	11	380	21	455	31	511	41	575	51	661
2	161	12	386	22	472	32	519	42	576	52	680
3	238	13	388	23	473	33	522	43	579	53	682
4	274	14	404	24	475	34	537	44	582	54	701
5	310	15	411	25	488	35	545	45	583	55	707
6	321	16	414	26	496	36	547	46	606	56	722
7	350	17	416	27	499	37	550	47	626	57	737
8	350	18	440	28	507	38	553	48	648	58	823
9	358	19	443	29	509	39	559	49	650	59	835
10	371	20	448	30	510	40	562	50	658	60	916

Tableau 2: Effectifs de chaque classe (distribution des fréquences)

Numéro de classe	Classes les pluies	Effectif ou Fréquence Absolue
1	$130 \le P < 230$	2
2	$230 \le P < 330$	4
3	$330 \le P < 430$	11
4	$430 \le P < 530$	16
5	$530 \le P < 630$	14
6	$630 \le P < 730$	9
7	$730 \le P < 830$	2
8	$830 \le P < 930$	2
somme		60

Chapitre II

Notions de Statistiques et de probabilités

objectif : La statistique regroupe l'ensemble des méthodes permettant d'obtenir et traiter des informations qui aident à la prise de décision.

I - Description statistique

A) Définition.

1. Population : l'ensemble sur lequel porte l'étude
2. Echantillon : C'est un sous-ensemble de la population
3. Statistique : l'ensemble de données sur une variable collectées et mises en ordre

soit : x : variable
x_i : Toute les valeurs que peut prendre la variable x_i.

Exemple : on veut étudier le bilan hydrologique.

$$x = \{x_i\} = \{I, P, R, E\}$$

B) Dépouillement des observations.

L'information doit faire l'objet d'un dépouillement pour s'assurer de sa qualité

* Vérification des données
  - erreur d'echantillonnage
  - erreur de mesure
  - erreur systématique
* Classification des données

✓ Deux types de variables sont distinguées :
- variable quantitative
- variable qualitative.

✓ L'organisation des données peut se faire par des tableaux.

- tableau simple : comprenant 2 colonnes au moins, l'une pour la variable x et sa valeur x_i et l'autre pour les effectifs n_i.

c) Notion de fréquence.

on définit la fréquence par :

$$f(x) = \frac{n_i}{N}$$

- n_i : effectif d'une valeur, il représente le nombre de fois où cette valeur apparaît.
- N : effectif total ; $N = \sum n_i$.

Exemple :

classe des précipitations (mm)	effectif (n_i)	fréquence f_i
50 - 150	6	6/80 = 0,075
150 - 250	12	12/80 = 0,15
250 - 350	27	27/80 = 0,337
350 - 450	28	28/80 = 0,35
450 - 550	7	7/80 = 0,087
	$\sum n_i = N = 80$	

- Quelle est la probabilité d'avoir une précipitation entre 450 et 550 mm.

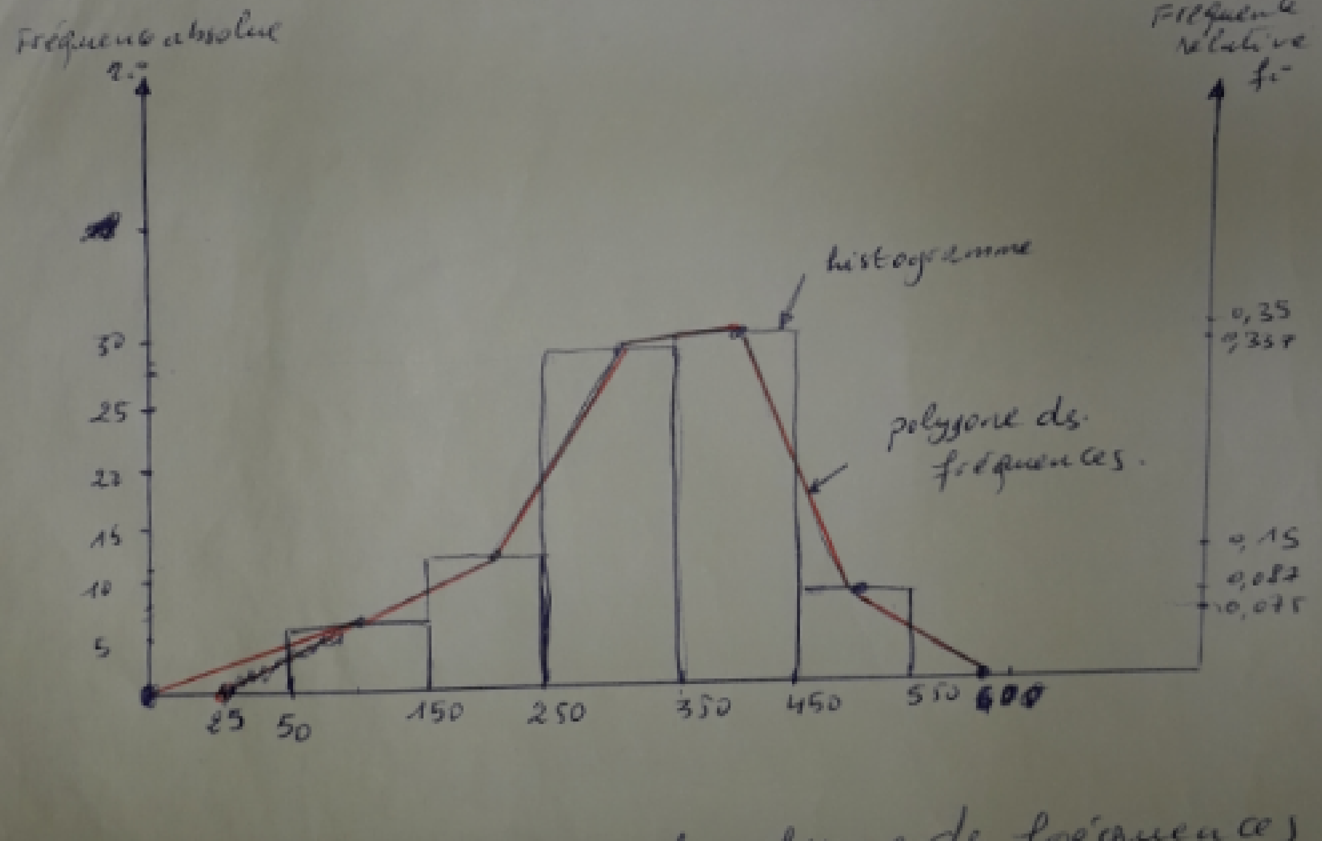
$$f = \frac{n_i}{N} = \frac{7}{80} = 0{,}0875 \ (8{,}75\%)$$

- La probabilité d'avoir un événement entre 450 et 550 est de 8,75 %.

Histogramme et polygone de fréquence

c'est la représentation graphique de la fonction de distribution (n_i, x_i) ou (f_i, x_i)

- un histogramme est une série de rectangle ayant :
  - a- leurs bases sur l'axe des x ; les limites des intervalles et dont la longueur est égale à l'amplitude ou la grandeur de l'intervalle.
  - b. leurs hauteurs sont égales aux fréquences
- le polygone des fréquences est obtenu en joignant les milieux des sommets des rectangles.



Histogramme et polygone des fréquences

(Distribution des fréquences).

✓ Notion de fréquence cumulée F(x) (Fonction de Répartition) :

Le tableau ci-dessous indique les débits maximum annuels d'un oued en m³/s. $\sum n_i = N = 25$

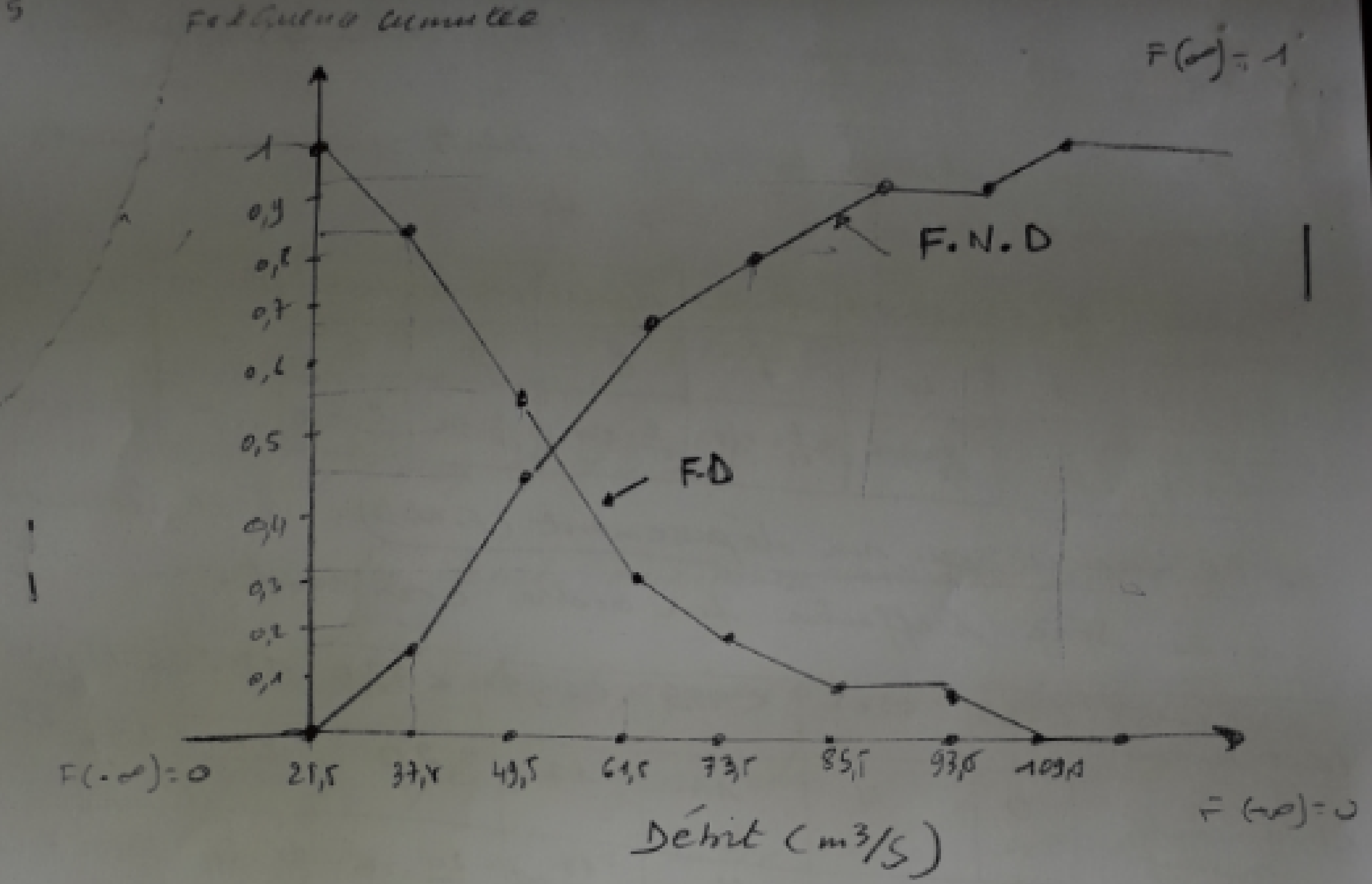
Débit (m³/s)	<25,5	25,5-37,5	37,5-49,5	49,5-61,5	61,5-73,5	73,5-85,5	85,5-97,5	97,5-105,5
Effectif (n_i)	0	4	7	6	3	3	0	2
Fréquence relative (f_i)	0	$\frac{4}{25}=0,16$	$\frac{7}{25}=0,28$	$\frac{6}{25}=0,24$	$\frac{3}{25}=0,12$	$\frac{3}{25}=0,12$	$\frac{0}{25}=0$	$\frac{2}{25}=0,08$

1. Fréquence au non dépassement (F.N.D) : le classement de la série s'effectue par ordre croissant.

Débit (m³/s) (limite supérieure)	<25,5	<37,5	<49,5	<61,5	<73,5	<85,5	<97,5	<105,5
E.C.C effectif cumulé croissant	0	4	11	17	20	23	23	25
F.N.D Fréquence cumulée	0	$\frac{4}{25}=0,16$	$\frac{11}{25}=0,44$	$\frac{17}{25}=0,68$	$\frac{20}{25}=0,80$	$\frac{23}{25}=0,92$	$\frac{23}{25}=0,92$	$\frac{25}{25}=1$

2. Fréquence au dépassement (F.D) : le classement de la série s'effectue par ordre décroissant.

Débit (m³/s) (limite inférieure)	>105,5	>97,5	>85,5	>73,5	>61,5	>49,5	>37,5	>25,5
E.C.D effectif cumulé décroissant	0	2	2	5	8	14	21	25
F.D Fréquence cumulée	0	$\frac{2}{25}=0,08$	$\frac{2}{25}=0,08$	$\frac{5}{25}=0,20$	$\frac{8}{25}=0,32$	$\frac{14}{25}=0,56$	$\frac{21}{25}=0,84$	$\frac{25}{25}=1$



fréquence cumulée : fréquence
Fonction de répartition

soit F.D → monotone De croissante entre 0 et 1
F.N.D → " croissante entre 0 et 1

on suppose que les caractères entre les intervalles sont uniformément répartie

- La somme des valeurs de toutes les fréquences plus petite que la limite supérieure d'un intervalle est appelée Fréquence au non dépassement (F.N.D).

$prob(Q_{max} < 73,5) = 0,8 \ (80\%) \rightarrow F.N.D$

(80% des débits maximum annuels sont $< 73,5 \ m^3/s$)

- La somme des fréquences de toutes les valeurs plus grandes que la limite inférieure d'un intervalle est appelée Fréquence au dépassement (F.D)

$prob(Q_{max} > 73,5) = 0,2 \ (20\%) \rightarrow F.D$

(20% des débits maximum annuels sont $> 73,5 \ m^3/s$]

on constate : $F.N.D + F.D = 80\% + 20\% = 100\%$
$= 0,8 + 0,2 = 1$

- La fonction de répartition $F(x)$ (F.N.D ; F.D) est une fonction intégrale $\underset{échan}{F(x)} = \sum f(x)$

- pour une population $F(x) = \int^{x} f(x).dx$
définie uniquement dans l'intervalle (0,1)

- $f(x)$: fonction de densité de probabilité
- $F(x)$: fonction de probabilité.

7

Notion de période de retour

- La probabilité de dépassement d'une valeur critique x_c est l'inverse de la période de retour.

$$P(x \geq x_c) = \frac{1}{T} = \boxed{F.D}$$

$$\Rightarrow \quad T = \frac{1}{P(x \geq x_c)} = \frac{1}{F.D}$$

T : période de retour

P : risque hydrologique.

- La fiabilité (sureté de fonctionnement)

$$P(x < x_c) = 1 - \frac{1}{T} = \boxed{F.N.D}$$

II. caractéristiques de tendance centrale et de dispersion.

II.1. caractéristique de tendance centrale.

1- Mode mo. ça correspond à la valeur de fréquence maximale (- qu'on rencontre le + souvent, - effectif n_i plus élevé

2- Médiane me. ça correspond à la valeur des caractères qui partage les effectifs en 2 parties égales.

✓ pour les séries continues

$$\boxed{\frac{me - a}{b - a} = \frac{N/2 - ECC_{-1}}{n}}$$

La médiane correspond à l'effectif $N/2$; on repère dans la colonne des effectifs cumulés croissant (E.C.C) à quelle classe appartient cette effectif soit [a,b[cette classe, n son effectif ; ECC_{-1}, l'effectif cumulé croissant de la classe précédente.

3. Moyenne $\bar{x}$

i. moyenne arithmétique

$$\bar{x} = \frac{x_1 + x_2 + \dots x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

_ n : nombre des x.

ii. moyenne pondérée

. c.à-d, on va attribuer un poids à la valeur x.

$$\bar{x}_p = \frac{\sum_{i=1}^{n} W_i \, x_i}{\sum_{i=1}^{n} W_i}$$

W_i : facteur de pondération.

exemple. Les notes d'un étudiants.

Matière	Phy	Math	Fr
notes	12	13	15
coefficient	4	3	2

facteur de pondé

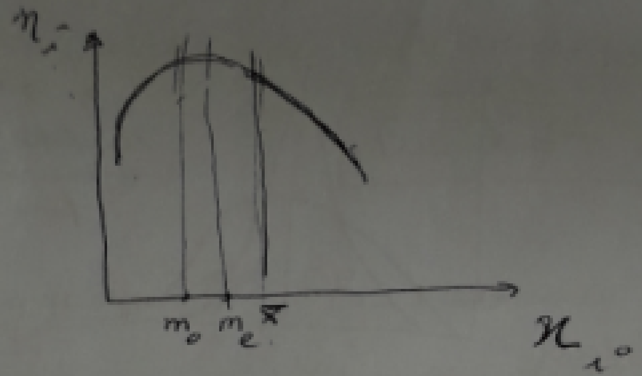
moyenne des no

$$\bar{x}_p = \frac{12 \times 4 + 13 \times 3 + 15 \times 2}{4 + 3 + 2 ,, 9} = 13$$

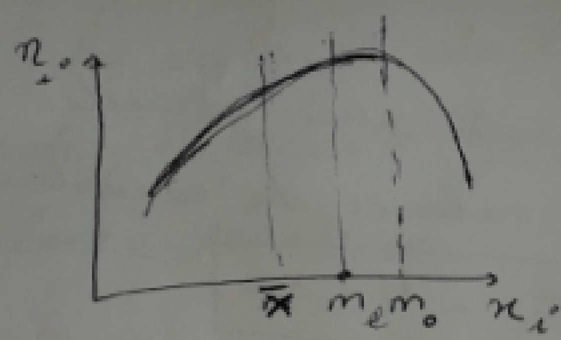
* moyenne arithmétique = moyenne arithmétique quand le facteur de pondération est le même

c'est le quotient de la Σ pondérée de x_i par la somme des poids.

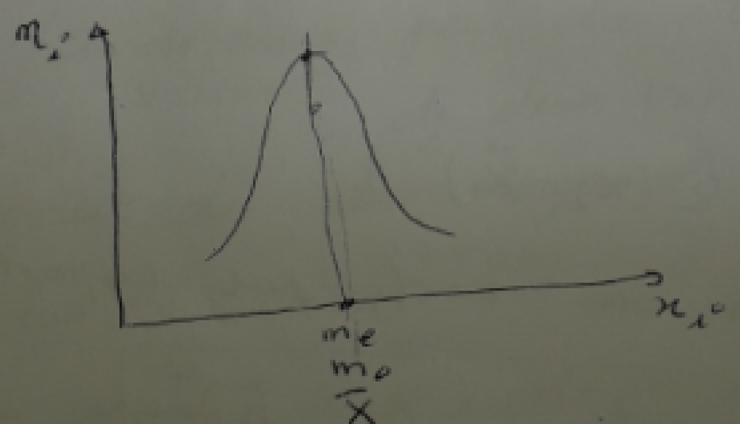
distribution des fréquences



fonction de distribution oblique à gauche.



fonction de distribution oblique à droite.

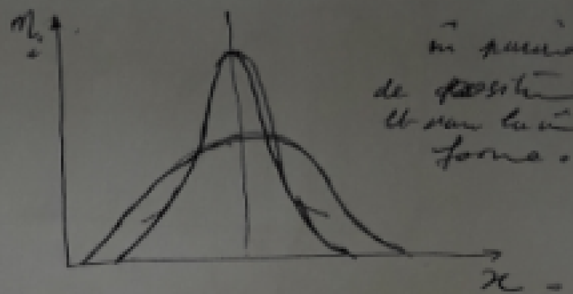


F.D.R symétrique

paramètre de position

II . 2. Caractéristiques de dispersion.

deux séries ont les mêmes paramètres de position. Elles sont de forme différente.



même paramètre de position et non la même forme.

Les paramètres de dispersion permet d'apporter une réponse à cette différence

1. L'étendue

c'est la différence entre la plus grande valeur et la plus petite valeurs du caractère

$$\boxed{e = Max\, x_i - Min\, x_i}$$

Appelé aussi intervalle de variation, c'est un indicateur simple mais insuffisant, il peut prendre des valeurs exceptionnelles.

2- Variance.

$$Var(x) = (\sigma_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2) =$$

c'est la moyenne des écarts carrés par rapport à la moyenne.

$$\boxed{\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

3. <u>Ecart Type</u>. ça correspond à la racine carré de la variance.

$$\boxed{\sigma = \sqrt{\sigma^2} = \sqrt{Var(x)}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

<u>Remarque</u>: la variance et l'ecart type permettent de mesurer la « dispersion » des valeurs de la série autour de la moyenne.

si les valeurs de la série possèdent une unité, l'écart type, s'exprime dans la même unité.

<u>Covariance</u>.

$$cov(x,y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{N} \left(\sum_{i=1}^{N} xy - \bar{x}\,\bar{y} \right)$$

* La covariance peut être vue comme le produit des valeurs de deux variables moins le produit des deux moyennes.
* plus la covariance est faible et plus les séries sont indépendantes et inversement plus elle est élevée et plus les séries sont liées.