

Vocabulaire statistique : Définitions

Qu'est ce qu'une Population ?

ensemble des unités statistiques ou individus sur lesquels on effectue une analyse statistique

Qu'est ce que l'effectif totale ?

est le nombre d'individus constituant la population, ou bien c'est la somme des fréquences. Il est généralement noté par n .

Qu'est ce qu'une Echantillon ?

est un sous-ensemble de la population.

Qu'est ce qu'un Caractère ?

c'est l'aspect particulier commun à tous les individus de la population ou de l'échantillon.

Vocabulaire statistique : Définitions

Qu'est ce que les Modalité ?

Ce sont les différentes valeurs que peut prendre un caractère qualitatif.

Qu'est ce qu'une Série statistique ?

C'est une suite des valeurs prises par une variable X notées $(x_1 ; x_2 ; x_3 ; \dots)$. Ces valeurs ne sont pas nécessairement des nombres.

Qu'est ce qu'une Fréquence ?

est le nombre d'individus présentant la même modalité dans la population. Ou bien, c'est le nombre d'occurrences d'une valeur.

Qu'est ce qu'une Fréquence relative ?

est le quotient d'un effectif par celui de la population.

Qu'est ce qu'une Pourcentage ?

c'est la fréquence relative exprimée en pour cent. Il est égal à $100 * fr_i$.

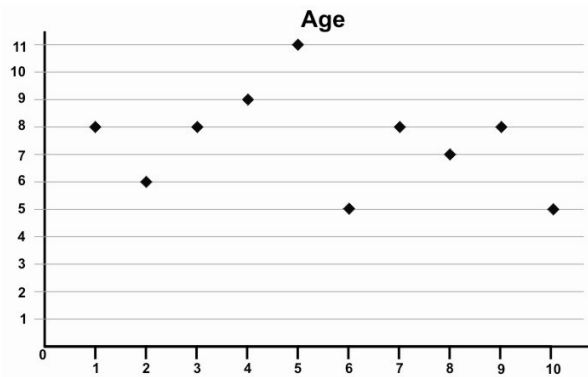
Analyse descriptive de données

Une série peut être représentée par une liste de valeurs, tableau ou un graphique.

Age des enfants : 8, 6, 8, 9, 11, 5, 8, 7, 8, 5

Leur couleurs préférées : vert, rouge, rouge, bleu, jaune, bleu, bleu, vert, marron, bleu

Enfant	Khaled	Ahmed	Omar	Zaki	Nabil	Ali	Samir	Bachir	Farid	Yahia
Age	8	6	8	9	11	5	8	7	8	5
Couleur	vert	rouge	rouge	bleu	jaune	bleu	bleu	vert	marron	bleu



Analyse descriptive de données

Vocabulaire statistique : Définitions

Qu'est ce qu'une Série temporelle ?

est un ensemble de données collectées sur plusieurs périodes de temps différentes.

Qu'est ce qu'une Etendue ?

c'est la différence entre les valeurs extrêmes d'une série.

Pour la calculer, il suffit de soustraire la plus grande valeur (maximum) et la plus petite valeur (minimum) de la série.

Les unités statistiques d'une population peuvent être représentées sous forme d'une série simple ou regroupées. Lorsqu'elles sont regroupées on les appelle des distributions.

Vocabulaire statistique : Définitions

Qu'est ce qu'une Classe ?
est un groupe de données dont les caractéristiques sont très similaires.

Qu'est ce que l'Amplitude d'une classe ?
est la différence entre la valeur supérieure et la valeur inférieure de la classe.

Qu'est ce que le centre d'une classe ?
est égal à l'amplitude divisée par deux.

Résumé des concepts

Population	Échantillon	Caractères					
		Niveau scolaire			Situation matrimoniale		
		Modalités	fréquence	fréquence relative	Modalités	fréquence	Fréquence relative
Wilaya d'El-Oued	2500 employé	primaire	71	71/2500	M	2196	2196/2500
		moyen	156	156/2500	C	214	214/2500
		secondaire	631	631/2500	V	28	28/2500
		universitaire	1642	1642/2500	D	62	62/2500
		Effectif Total	2500				2500

Analyse descriptive de données

Remarque

Pour ne pas confondre les vocabulaires, il est noté que certains auteurs utilisent les termes précédent de manière différente, par exemple:

Terme	Autres termes
Fréquence	Effectif
	Fréquence absolue
Fréquence relative	Fréquence
	Pourcentage

Analyse descriptive de données

Tabulation et présentation des données

Les données brutes, telles qu'elles résultent d'une enquête, un examen ou un sondage, sont parfois impossibles à assimiler.

Par conséquent, il est nécessaire de les rassembler, les résumer et les présenter sous une forme compréhensible (tableaux appropriés),

afin de faciliter leur consultation et leur comparaison avec d'autres données, déduire de leurs résultats, et avoir un premier aperçu des informations qu'elles contiennent.

Généralement, il existe deux manières basiques pour représenter et tabuler les données statistiques sont :

Analyse descriptive de données

1. Tableau de fréquences (Distribution de fréquences) :

La distribution des données brutes peuvent être :

- 1. par valeurs : Lorsque le critère de regroupement est numérique.
- 2. par modalités : Lorsque le critère de regroupement n'est pas numérique.

Pour une variable considérée, la distribution de fréquences présente au moins les fréquences pour chacune des modalités de cette variable (Tableau de fréquences simple).

La distribution peut aussi inclure la fréquence relative et le pourcentage de chaque modalité. (on parle donc sur : un Tableau de fréquences.

Modalité x_i	Fréquence f_i	Fréquence relative $fr_i = n_i / n$	Poucentage $p_i = fr_i * 100 \%$
x_1			
...			
x_k			
Total	n	1	100 %

Analyse descriptive de données

Exemple 1 :

Les données suivantes représentent les notes d'examen du module d'analyse de données en première année master Informatique pour l'année universitaire 2010/2011, obtenues par 20 étudiants.

Mettez ces données dans un tableau itératif simple.

10	17	12	14	14	11	16	14	16	19
16	17	15	16	15	10	15	11	11	14

Solution

Note	Fréquence
10	2
11	3
12	1
14	4
15	3
16	4
17	2
19	1

Analyse descriptive de données

Exemple 2 :

Les données suivantes sont les appréciations de 20 étudiants durant l'année universitaire 2010/2011.

Il est demandé de mettre ces données dans un tableau de fréquence.

Passable	Bien	Bien	Excellent	Passable	Bien	Très bien	Passable	Bien	Très bien
Excellent	Bien	Passable	Très bien	Passable	Bien	Excellent	Excellent	Bien	Bien

Solution

Appréciation	Fréquence	Fréquence relative	Pourcentage
Passable	5	5/20	25 %
Bien	8	8/20	40 %
Très bien	3	3/20	15 %
Excellent	4	4/20	20 %

Analyse descriptive de données

Limites :

1. Lorsque le nombre de modalité est très grand (plus de 20) dans un tableau de distribution de fréquence
2. Le cas d'un caractère quantitatif continu



On aura besoin d'effectuer des regroupements/répartitions en Classes (ou catégories) de valeurs ou de modalités afin de minimiser la taille du tableau.

Généralement, les données statistiques peuvent être classées selon l'un des critères suivants:

- 1- Classification géographique
2. Classification historique ou chronologique.
- 3- Classification qualitative ou descriptive.
- 4- Classification quantitative.

Analyse descriptive de données

Classification quantitative :

1. Données discrètes

Nous mentionnons la valeur minimale et la valeur maximale de chaque classe comme indiqué dans le tableau suivant:

Classes (n)	Fréquence
[\min_1 - \max_1]	5
[\min_2 - \max_2]	15
...	..
[\min_n - \max_n]	18

Remarque

Cette méthode est défectueuse car elle n'est pas valide pour les données contenant des fractions (quantitatives continues).

Analyse descriptive de données

Classification quantitative :

2. Données continues

Nous ne mentionnons que la valeur minimale de la classe et mettons ensuite un tiret et énonçons la première classe, par exemple (du 10 à moins de 20).

Classes (n)	Fréquence		Classes (n)	Fréquence
10 -	5		[10 - 20 [5
20 -	15	ou bien	[20 - 30 [15
30 -	10		[30 - 40 [10
40 -	20		[40 - 50 [20



Détermination du nombre de classes

En générale, il est recommandé que le nombre de classe doit être entre 5 et 20

Etapes de construction d'un tableau de fréquence en classes

- 1- Calculer l'étendue $R =$ la plus grande valeur - la plus petite valeur
- 2- Calculer le nombre de classes k
 - La formule de Sturge : $k = 3.322 * \log(n)$
 - La formule de Yule : $k = 2.54 * \sqrt[4]{n}$
- 3- Calculer l'amplitude de la classe $h =$ étendue / nombre de classes
- 4- Précisez les intervalles des classes, par exemple le début de la première classe sera égal à la valeur la plus basse des données ou légèrement inférieur.
- 5- Construire la table et marquer les fréquences.

Remarques

1. Il peut arriver que vous devez arrondir les résultats de calcul de k et h , mais il faut que vous prendre en considération que :

$$\mathbf{R \leq k * h}$$

2. Même s'il est préférable de construire des classes d'amplitudes égales, il est parfois plus naturel d'utiliser des classes larges pour les grandes valeurs et des classes étroites pour les petites valeurs.

2. Représentation graphique

La représentation graphique nous permet voir de manière visuelle et rapide les caractéristiques d'une distribution ainsi que les différences pouvant exister entre deux populations.

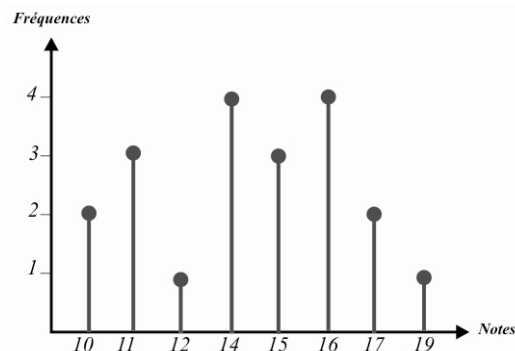
Pratiquement, il existe plusieurs forme graphique pour représenter les données, mais presque toute distribution peut avantageusement être présentée sous une forme graphique plus adéquate que d'autres.

A) Variables quantitaives

1. Diagramme en bâtons :

ce diagramme est formé d'un ensemble de bâtons ayant des tailles proportionnelle à la fréquence des valeurs x_i . Il s'applique aux variables discrètes où les valeurs discrètes x_i sont placées sur l'axe des abscisses, et les fréquences sur l'axe des ordonnées.

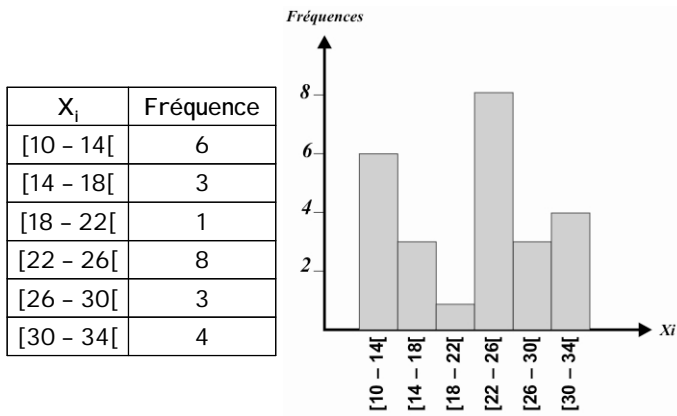
Note	Fréquence
10	2
11	3
12	1
14	4
15	3
16	4
17	2
19	1



Analyse descriptive de données

2. Histogramme :

est un graphique formé d'un ensemble de rectangles adjacents. Chaque rectangle est associé à une classe et il a une surface proportionnelle à la fréquence des classes qui est reporté ordonnée si les amplitudes des classes sont identiques.



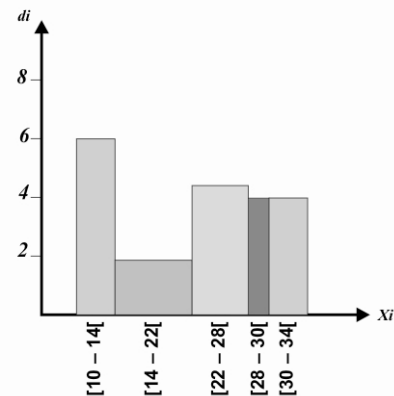
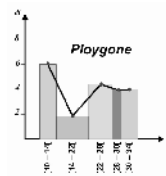
Analyse descriptive de données

2. Histogramme :

Si les amplitudes sont différentes, on reporte en ordonnée la densité d_i des données c'est-à-dire, la fréquence divisée par l'amplitude de la classe (ou bien, par les unités de largeur) pour que la surface de chaque rectangle soit proportionnelle à la fréquence.

X_i	Fréquence
[10 - 14[6
[14 - 18[3
[18 - 22[1
[22 - 26[8
[26 - 30[3
[30 - 34[4

X_i	Fréquence	Densité
[10 - 14[6	6
[14 - 22[4	2
[22 - 28[9	4.5
[28 - 30[2	4
[30 - 34[4	4

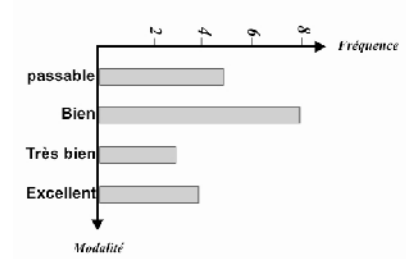
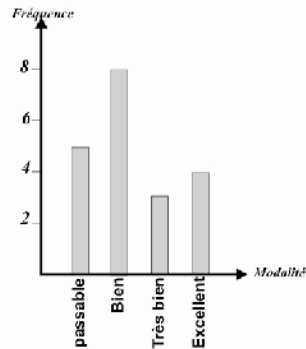


Analyse descriptive de données

3. Diagramme en barre (à bandes):

il est similaire au diagramme en bâtons sauf que les barres ont une certaine épaisseur pour qu'il ne soit pas confondues avec les bâtons.

Appréciation	Fréquence
Passable	5
Bien	8
Très bien	3
Excellent	4



Analyse descriptive de données

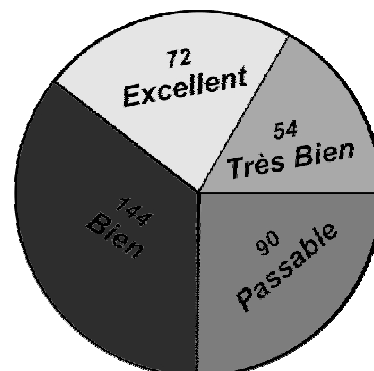
4. Diagramme circulaire (en secteurs) :

L'effectif total est représenté par un disque où chaque modalité est représentée par un secteur circulaire dont la surface (l'angle d'un secteur) est proportionnelle à la fréquence correspondante.

L'angle de chaque modalité se calcule par la formule suivante :

$$\text{Angle} = \frac{\text{fréquence}}{\text{effectif totale}} * 360^\circ$$

Appréciation	Fréquence	Angle
Passable	5	90
Bien	8	144
Très bien	3	54
Excellent	4	72



Analyse descriptive de données

