

COURS DE STATISTIQUES DESCRIPTIVES

La **Statistique Descriptive** est l'étude de la collecte de données à une date donnée, leur analyse, leur traitement, l'interprétation des résultats et leur présentation afin de rendre les données compréhensibles par tous. C'est à la fois une science, une méthode et un ensemble de techniques.



I / VOCABULAIRE ET EXEMPLE

Exemple : On étudie le prix d'un article dans différents magasins et on obtient les résultats suivants : 35 ; 37 ; 40 ; 35 ; 36 ; 36 ; 39 ; 36 ; 35 ; 37 ; 37 ; 38 ; 36 ; 35 ; 40 ; 37 ; 36 ; 36 ; 38 ; 39 ; 35 ; 36 ; 37 ; 38 ; 36.

- L'ensemble des valeurs est appelé **série statistique**.
- L'ensemble sur lequel porte l'étude statistique s'appelle la **population**.
Ici la population correspondra à l'ensemble des prix.
- Un élément x_i de cet ensemble X est un **individu** ou une **valeur**.
Chaque prix est un individu ou une valeur.
- La **caractéristique** d'une série statistique est la propriété étudiée sur chaque individu.
On étudie la différence entre les prix.
- Lorsque les modalités du caractère ne sont pas mesurables (c'est-à-dire qu'ils ne sont pas représentés par des nombres), le caractère est dit **qualitatif** (couleur des voitures...)
- Lorsque les modalités du caractère sont mesurables (c'est-à-dire qu'ils sont représentés par des nombres), le caractère est dit **quantitatif** (âge, prix...)
- La **variable** X est dite **quantitative discrète** si x_i est une valeur numérique.

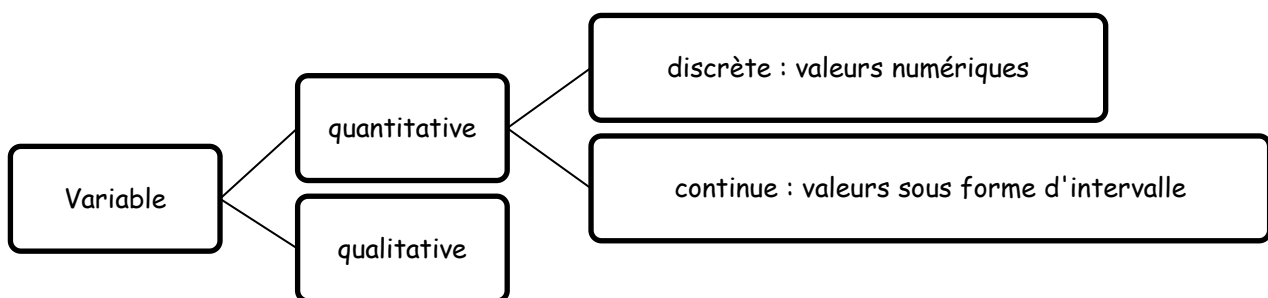
On peut mettre la série sous forme d'un tableau de valeurs discrètes :

Prix constatés en € (x_i)	35	36	37	38	39	40
Nombres de magasins (n_i)	5	8	5	3	2	2

- Le caractère est dit **quantitatif continu** si x_i est mis sous forme d'intervalle.

On peut mettre la série sous forme d'un tableau de valeurs continues :

Prix constatés en € (x_i)	[35 ; 37[	[37 ; 39[	[39 ; 41[
Nombres de magasins (n_i)	13	8	4



II / PARAMETRES GENERAUX DES SERIES QUANTITATIVES DISCRETES**1. Effectif.**

- Le nombre d'individus pour lesquels la variable x_i prend une valeur donnée est l'**effectif** n_i de cette valeur.
- La somme des effectifs de toutes les valeurs est l'**effectif total**, $n = \sum_{i=1}^p n_i = n_1 + n_2 + \dots + n_p$.

Individus x_i	x_1	x_2	...	x_p
Effectifs n_i	n_1	n_2	...	n_p
Effectifs cumulés croissants	n_1	$n_1 + n_2$	...	$n = \sum_{i=1}^p n_i = n_1 + n_2 + \dots + n_p$

2. Fréquence :

La fréquence d'une valeur est le quotient de l'effectif de la valeur par

l'effectif total, $f_i = \frac{n_i}{n}$

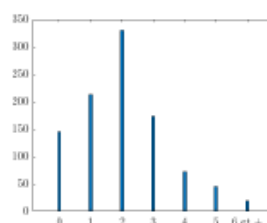
Pourcentage ou Fréquence en pourcentage : $P_i = \frac{n_i}{n} \times 100$

3. Mode : On appelle **mode** d'une série statistique, la valeur x_i ou la classe qui a le plus grand effectif.

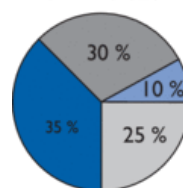
4. Etendue : On appelle **étendue** la différence entre la plus grande et la plus petite valeur du caractère.
Étendue = $x_{\max} - x_{\min}$

6. Représentation graphique des séries statistiques discrètes**a. Diagramme en bâtons :**

On représente sur l'axe des abscisses les différentes valeurs du caractère x_i et, sur l'axe des ordonnées, les effectifs.

**b. Diagramme circulaire :**

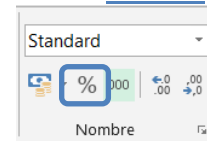
L'angle d'ouverture de chaque secteur est proportionnel à l'effectif.

**Avec EXCEL**

• L'effectif total se calcule avec la fonction **SOMME**.

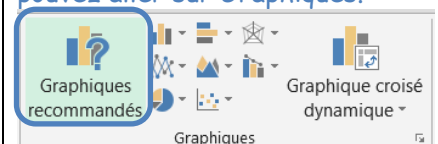
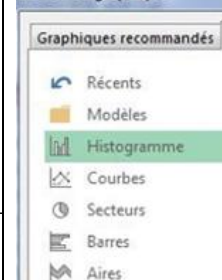
• L'effectif cumulé se calcule aussi avec la fonction **SOMME** : Il faut « bloquer » la première composante de la plage afin qu'en étirant la formule, celle-ci donne un résultat exact.
=SOMME(\$B\$2:B2)

Sur Excel, la cellule peut être mise en pourcentage : Dans l'onglet **ACCUEIL**, vous pouvez aller sur **Nombre**.



Sur Excel, les maximums et minimums se calculent avec les fonctions **MAX** et **MIN**.

Dans l'onglet **INSERTION**, vous pouvez aller sur **Graphiques**.

**Insérer un graphique**

Le diagramme en bâtons es appelé à tort histogramme sur Excel.

III / AUTOUR DE LA MOYENNE

Individu x_i	x_1	x_2	...	x_p
Effectifs n_i	n_1	n_2	...	n_p

a. Paramètre de position : La moyenne arithmétique.

$$\bar{x} = \frac{\sum_{i=1}^{i=p} n_i x_i}{N} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

b. Paramètres de dispersion autour de la moyenne

Variance : La variance V d'une série statistique de moyenne $\bar{x}$ dont les valeurs du caractère sont $x_1, x_2, x_3, \dots, x_k$ et les effectifs correspondants sont $n_1, n_2, n_3, \dots, n_k$ est égale à :

$$V(X) = \frac{\sum_{i=1}^{i=p} n_i x_i^2}{N} - \bar{x}^2$$

Autre expression utilisée dans les livres (et pas sur Excel) :

$$V(X) = \frac{\sum_{i=1}^{i=p} n_i (x_i - \bar{x})^2}{N} = \frac{n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_p (x_p - \bar{x})^2}{N}$$

Ecart-type: $\sigma(X) = \sqrt{V(X)}$

Remarque : L'écart-type possède la même unité que les valeurs de la série.

c. Comparaison de deux séries :

La série ayant le plus petit écart-type est dite plus concentrée autour de la moyenne.
La série ayant le plus grand écart-type est dite plus dispersée autour de la moyenne.

Avec EXCEL

- La formule de la moyenne (=MOYENNE) calcule une moyenne sans effectif (sans n_i).

$$\bar{x} = \frac{\sum_{i=1}^{i=p} x_i}{N}$$

- Pour calculer une moyenne pondérée (formule générale avec des n_i), il va falloir procéder par étape :

- 1^{ère} étape : Calculer $n_i * x_i$
2^{ème} étape : Calculer la somme $\sum_{i=1}^{i=p} n_i x_i$
3^{ème} étape : Diviser par l'effectif total.

- Les formules de la variance (=VARIANCE) ou de l'écart-type (=ECARTYPE.PEARSON) calculent une variance ou un écart-type sans effectif (sans n_i).

- Pour calculer une variance pondérée (formule générale avec des n_i), il va falloir procéder par étape :

- 1^{ère} étape : Calculer $n_i * x_i^2$
2^{ème} étape : Calculer la somme $\sum_{i=1}^{i=p} n_i x_i^2$
3^{ème} étape : Diviser par l'effectif total.
4^{ème} étape : Soustraire par la moyenne au carrée.

- Le calcul de l'écart-type se fait alors avec la formule de la racine (=RACINE).

IV / AUTOUR DE LA MEDIANE**a. La médiane, paramètre de position :**

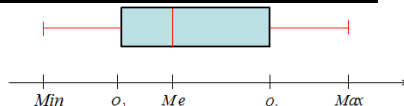
Définition : La **médiane**, notée Me , d'une série statistique partage cette série en deux parties de telle sorte que :

- au moins la moitié des valeurs sont inférieures ou égales à la médiane ;
- au moins la moitié des valeurs sont supérieures ou égales à la médiane.

Remarque : La médiane est la valeur « centrale » qui partage la série en deux groupes de même effectif.

b. Les quartiles, paramètre de dispersion :

- Le **premier quartile** d'une série statistique, noté Q_1 , est le premier nombre tel que au moins **25 %** des données soient inférieures à ce nombre.
- Le **troisième quartile** d'une série statistique, noté Q_3 , est le premier nombre tel que au moins **75 %** des données soient inférieures à ce nombre.

c. Ecart interquartile : $Q_3 - Q_1$ **d. Diagramme en boîte ou boîte à moustache****Interprétation du diagramme en boîte**

- La médiane : 50 % des valeurs sont situées en-dessous de la Médiane
50 % des valeurs sont situées au-dessus de la Médiane
- Les quartiles : 50 % des valeurs sont situées dans l'écart interquartiles.
25 % des valeurs sont situées en-dessous de Q_1 .
75 % des valeurs sont situées au-dessus de Q_3 .

Comparaison de deux diagrammes en boîte

On compare les deux médianes ; les quartiles ; l'écart interquartile ; l'étendue...

Avec EXCEL

- La formule
`=QUARTILE.INCLURE(plage ; 2)`
calcule une médiane sans effectif (sans n_i).

- Pour calculer une médiane pondérée (avec des n_i), il va falloir procéder par étape :

1^{ère} étape : Ranger les valeurs x_i par ordre croissant.

2^{ème} étape : Choisir la valeur « centrale » si l'effectif total est impair ou la moyenne des valeurs « centrales » si l'effectif total est pair.

- Les formules
`QUARTILE.INCLURE(plage ; 1)` et
`QUARTILE.INCLURE(plage ; 3)`
calculent les quartiles sans effectif.

- Méthode pour trouver les quartiles avec des effectifs :

1^{ère} étape : Ranger les valeurs x_i par ordre croissant.

2^{ème} étape : Calculer les fréquences associées aux effectifs.

3^{ème} étape :

- Q_1 : Première valeur des x_i ayant une fréquence $> 25\%$
- Q_3 : Première valeur des x_i ayant une fréquence $> 75\%$

Le diagramme à moustache n'est pas faisable sur Excel.

A faire à la main sur votre copie...

V / EXERCICES :

Le fichier EXCEL nommé « TD5 Ressource » contient toutes les données nécessaires.
Ouvrez-le et enregistrez-le dans votre espace personnel sous le nom « TD5_NOM »

Exercice 1 :

La feuille nommée « Exe1A » donne les notes de 95 étudiants. Les notes vont de 1 à 10.

1. Calculer à l'aide des formules EXCEL la moyenne, l'écart-type, le minimum, le maximum, l'étendue, la médiane et les quartiles. Arrondir vos résultats au centième.
2. Ouvrir une nouvelle feuille Excel nommée « Exe1B ».

Recopier le tableau suivant :

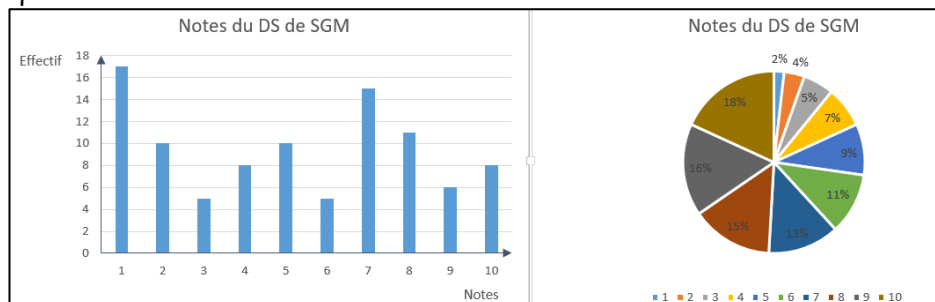
NOTES (xi)	1	2	3	4	5	6	7	8	9	10
Effectifs (ni)										
Effectifs cumulés croissants										
Fréquences										
Fréquences cumulées croissantes										
Effectif total										

A l'aide d'une formule que vous étirerez et de la fonction NB.SI, calculez les effectifs correspondants aux différentes notes.

A l'aide d'une formule, calculez l'effectif total.

A l'aide de formules que vous étirerez, calculez l'effectif cumulé croissant, la fréquence et les fréquences cumulées croissantes.

3. Faire les diagrammes en bâtons et circulaires des notes par rapport aux effectifs correspondants. Ils doivent être identiques à ceux-ci-dessous :



4. A l'aide de la ligne des fréquences cumulées croissantes, retrouvez la médiane et les quartiles.

5. Ouvrez une nouvelle feuille Excel nommée « Exe1C ».

Recopiez et complétez à l'aide des formules adéquates que vous étirerez les tableaux suivants :

NOTES (xi)	1	2	3	4	5	6	7	8	9	10
Effectifs (ni)										
ni*xi										
ni*xi²										
Effectif total										
Moyenne										
Variance										
Ecart-type										

Exercice 2 : On étudie le prix d'un article dans différents magasins et on obtient les résultats suivants :

Prix constatés en € (x_i)	35	36	37	38	39	40
Nombres de magasins (n_i)	5	8	5	3	2	2

1. Tracer le diagramme en bâtons de la série statistique suivante.

(Ne pas oublier le titre, les légendes sur les axes, les flèches)

2. Quel est l'effectif total ?

3. Quel est le mode de cette série ?

4. Quelle est l'étendue de la série précédente ?

5. En vous aidant de l'exercice précédent, calculez la moyenne de cette série.

6. Quel est son écart-type ? (Arrondir à 0,1 près)

7. Les valeurs de la série sont-elles dispersées ou concentrées ?

8. Calculer, sur le fichier Excel, les fréquences de la série statistiques.

9. Calculer, sur le fichier Excel, les fréquences cumulées croissantes.

10. En déduire la médiane et les quartiles.

11. Faire le diagramme en boîte de la série statistique.

12. Interpréter le diagramme en boîte.

Exercice 3 : La série du tableau ci-dessous présente le nombre de buts par match durant la Coupe du monde de football de 2010 :

Nombre de buts x_i	0	1	2	3	4	5	6	7
Nombre de matchs n_i	7	17	13	14	8	6	0	1

1. Quel est l'effectif total ?

2. Faire le tableau des effectifs cumulés croissants et des pourcentages.

3. Quel est le mode de cette série ?

4. Faire le diagramme en bâton de cette série.

5. Quelle est l'étendue de la série précédente ?

6. Quels sont la médiane et les quartiles de cette série ?

7. Quel est l'écart interquartile de la série précédente ?

8. Faire le diagramme en boîte de la série précédente.

9. Interpréter les résultats.

10. Sur une nouvelle feuille Excel, calculez la moyenne arithmétique de cette série.

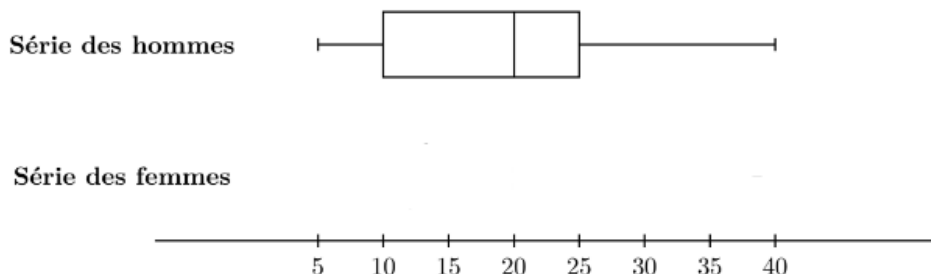
11. Quel est son écart-type ?

12. Interpréter les résultats.

Exercice 4 :

Dans une entreprise, on a dénombré 59 femmes et 130 hommes fumeurs. L'entreprise souhaite proposer à ses employés plusieurs méthodes pour diminuer, voire arrêter, leur consommation de cigarettes.

Une enquête est menée parmi les fumeurs, femmes et hommes, pour déterminer la quantité approximative de cigarettes fumées sur une journée. Des informations ont été relevées sur le diagramme ci-dessous.



1. La moyenne des femmes fumeuses est 15,17 et celles des hommes fumeurs est de 19,08. Quelle est la moyenne globale ?

2. Calculer l'étendue de la série des hommes.

3. Le diagramme en boîte de la série des femmes a été effacé. Tracez-le avec les informations ci-dessous.

Minimum : 5

Maximum : 40

Médiane : 15

Premier quartile : 10

Troisième quartile : 20

4. a. Parmi les fumeurs, environ la moitié des hommes fument au plus 20 cigarettes par jour. Vrai Faux
(justifier)

4. b. Parmi les fumeurs, environ la moitié des femmes fument entre 10 et 20 cigarettes par jour. Vrai Faux
(justifier)

5. Comparer les séries A et B. Il faut au moins deux interprétations.

Exercice 5 :

Une étude sur le budget consacré aux vacances d'été auprès de ménages a donné les résultats suivants :

Budget X	[800, 1000[	[1000, 1400[	[1400, 1600[	[1600, 1800[	[β , 2400[	[2400, α [
Milieu de l'intervalle x_i						
Effectifs						
Fréquences						
Fréquences cumulées croissantes	0,08	0,18	0,34	0,64	0,73	1

Certaines données sont manquantes.

1. Calculer la borne α manquante sachant que l'étendue de la série est égale à 3200.

2. Calculer les fréquences dans le tableau (sur Excel).

3. En prenant comme effectif total 200, calculer les effectifs de chaque classe de cette série (sur Excel).

4. En prenant le centre de l'intervalle comme caractère x_i , calculez la borne manquante β si le budget moyen est égal à 2013.

Exercice 6 : On donne la série statistique suivante :

Intervalle de temps en min	[10;50[	[50;80[	[80;120[	[120;300[	Total
Effectifs n_i	14	18	20	8	
Fréquences					
Fréquence cumulée croissante					

1. Quel est l'effectif total ?

2. Quelle est la classe modale de cette série ?

3. Quelle est l'étendue de la série précédente ?

4. En prenant le centre de l'intervalle comme caractère x_i , donnez la moyenne arithmétique de cette série ?

5. Quelle est son écart-type ?

6. Interpréter ces résultats.

7. Quels sont la médiane et les quartiles de la série ?

8. Quel est l'écart-interquartile ?

9. Faire le diagramme en boîte.

10. Interpréter les résultats.

Exercice 7 : Voici un échantillon dans lequel manquent plusieurs variables :

Valeurs du caractère x_i	5	a	b	c	d
Effectifs n_i	3	1	2	3	1

Déterminer a, b, c et d sachant que la médiane vaut 17, la moyenne vaut 15, que l'étendue vaut 21 et que le troisième quartile vaut 21.

Exercice 8 : Une association de consommateur a relevé le prix en euros d'un même article dans 80 magasins d'une grande ville. Malheureusement les données ont été effacées.

Prix en euros	20	22	23	a	30
Nombre de magasins	17	x	x	14	9

1. Quelle est la valeur de x manquante dans le tableau ?

2. Le prix moyen observé par cette association est de 23,25 €. Déterminer la valeur de a.

3. Tracer le diagramme à moustache de cette série.

4. Depuis que les prix ont été relevés, un nouveau magasin a ouvert et vend ce même article. En l'ajoutant dans son listing, l'association a remarqué que le prix moyen (sur les 81 magasins) montait alors à 23,50 €. Quel est le prix de vente dans ce nouveau magasin ?