1 LA STATISTIQUE DESCRIPTIVE

1.1 Définition

On appelle statistique descriptive l'ensemble des méthodes utilisées pour obtenir des renseignements sur une population à partir des renseignements sur un échantillon de cette population.

Le but de toute statistique est de donner une image concise et simplifiée de la réalité.

Pour être soumis à un traitement statistique, des données doivent comporter une variable de nature aléatoire.

1.2 Les séries statistiques à une dimension :

1.2.1 Représentation numériques des données :

Une série de données peut être résumé par quelques valeurs numériques appelées caractéristiques des séries statistiques, lassées en quatre grandes catégories :

- Les caractéristiques de tendance centrale
- Les caractéristiques de dispersion
- Les caractéristiques de forme
- Les caractéristiques de concentration

A Les caractéristiques de tendance centrale :

Qu'elles soient non groupées ou au contraires groupées par valeurs ou par classes, les variables quantitatives peuvent être utilement résumées par des caractéristiques dites de « tendance centrale ». Ces nombres résumés sont ainsi appelés car ils privilégient les valeurs principales de la distribution, au détriment par exemple de ceux qui caractérisent la dispersion ou la concentration des valeurs d'une série.

Ces valeurs centrales sont les moyennes, la médiane et le mode. Nous exposerons leur mode de calcul et leur signification en distinguant pour chacune d'elles le cas des données non groupées et le cas des données regroupées (soit par valeurs, soit par classes).

A.1 La moyenne arithmétique :

C'est le résumé le plus connu de l'information. La formule pour le cas continu est :

$$\overline{x} = \frac{1}{n}\sum_{i}^{n} x_i$$

A.1.1 Les propriétés de la moyenne :

La moyenne arithmétique permet de résumer par un seul nombre la série statistique elle prend en compte toutes les valeurs de la série et elle est facile à calculer.

La moyenne ou somme des écarts (différences algébriques) à la moyenne est nulle. Ses inconvénients sont d'être sensible aux valeurs extrêmes et de fournir un très mauvais résumé des données dans le cas de distributions très dispersées ou dissymétriques.

Exemple 1:soit les valeurs du ph pour des analyses d'eau :

Ph	**8.26**	**8.41**	**8.26**	**8.29**

La moyenne est de : $\overline{\boldsymbol{x}} = \mathbf{8.305}$

A.2 La moyenne géométrique :

Soit $\{x_1, x_2, \ldots, x_n\}$ une série de chiffre ,la formule de la moyenne géométrique de cette série est égale a :

$$G = \left[\prod_{i=1}^{n} \boldsymbol{x_i}\right]^{\left(\frac{1}{n}\right)}$$

Dans la pratique, la moyenne géométrique est utilisée essentiellement pour calculer la moyenne de ratios, ou plus particulièrement la moyenne d'indices.

Comme la moyenne arithmétique, la moyenne géométrique prend en compte chaque observation individuellement. Toutefois, elle diminue l'effet des très grands nombres. C'est une raison pour laquelle elle est parfois préférée à la moyenne arithmétique c'est un des aspects importants de la moyenne géométrique.

Exemple 2 : l'analyse de la Production de tomate dans la wilaya de Relizane à donner les resultats suivant pour trois années successives :

Année	Production (Quintal)
2015	18400
2016	31100
2017	103750

..

Si l'on s'intéresse aux taux moyen d'accroissement de la production entre chaque intervalle de temps nous aurons :

$$\boldsymbol{Taux1 = \frac{31100}{18400} = 1.69} \quad , \boldsymbol{Taux2 = \frac{103750}{31100} = 3.336}$$

Le taux moyen d'accroissement de la série des taux {1.69 , 3.336} est égal à leur moyenne géométrique :

$$\boldsymbol{G = [1.69 * 3.336]^{\left(\frac{1}{2}\right)}} \textbf{=2.375}$$

Temps t	Quantité de bactérie	Taux d'accroissement	Quantité de bactérie en appliquant le G=4
2015	18400		18400
		=31100/18400=1.69	
2016	31100		=2.375*18400=43692
		=103750/31100=3.336	
2017	103750		=2.375*43692=103750
		G=2.375	

A.3 MOYENNE MOBILE

A.3.1 Définition

On appelle moyenne mobile d'ordre k la moyenne arithmétique calculée sur k valeurs successives d'une **série chronologique**. La suite de ces moyennes arithmétiques donne alors la série des moyennes mobiles d'ordre k. Le processus qui consiste à remplacer la série initiale par cette nouvelle série est appelé lissage des séries chronologiques.

Quand la série chronologique consiste en données annuelles ou mensuelles, les moyennes prennent respectivement les noms de moyennes mobiles sur k années ou k mois.

Lors de données mensuelles, par exemple, en calculant une moyenne mobile sur 12 mois, les resultats obtenus correspondent au milieu de la période considérée, à savoir à la fin du sixième mois (ou au premier jour du septième mois), au lieu d'être affectée au milieu d'un mois comme les données d'origine. On corrige ceci en effectuant une moyenne mobile centrée sur 12 mois.

En général, une moyenne mobile centrée d'ordre k est calculée en effectuant la moyenne mobile d'ordre 2 de la moyenne mobile d'ordre k de la série chronologique initiale.

A.3.2 SERIE CHRONOLOGIQUE

Une série chronologique est constituée par une succession d'observations, sur un même sujet ou sur un même phénomène, régulièrement espacées dans le temps. Les séries chronologiques sont mensuelles, trimestrielles ou annuelles, parfois hebdomadaires, journalières, voire horaires (étude de trafic routier, trafic téléphonique) ou au contraire biennales, décennales (recensement de la population).

L'analyse des séries chronologiques est un outil statistique de prévision parmi ceux dont dispose le conjoncturiste pour planifier et faire face au changement.

Les séries chronologiques peuvent être décomposées en quatre composantes, chacune exprimant un aspect particulier du mouvement des valeurs de la série chronologique.

Ces quatre composantes sont :

- la tendance séculaire, qui traduit le mouvement à long terme ;
- les variations saisonnières, qui représentent des changements saisonniers ;
- les fluctuations cycliques, qui correspondent à des variations périodiques mais non saisonnières ;
- les variations irrégulières, qui sont les autres sources non aléatoires de variations de la série.

L'analyse d'une série chronologique consiste à faire une description mathématique des éléments qui la composent, c'est-à-dire a estimer séparément les quatre composantes.

A.3.3 ASPECTS MATHEMATIQUES

Etant donné la série chronologique : $Y_1, Y_2, Y_3, \ldots, Y_N$, on définit la série des moyennes mobiles d'ordre k par la suite des moyennes arithmétiques :

$$\frac{Y_1 + Y_2 + \cdots + Y_k}{k}, \frac{Y_2 + Y_3 + \cdots + Y_{k+1}}{k}, \frac{Y_{N-k+1} + Y_{N-k+2} + \cdots + Y_N}{k}$$

Les sommes aux numérateurs sont appelées sommes mobiles d'ordre k.

On peut utiliser des moyennes arithmétiques pondérées, avec des poids spécifiés à l' avance ; la suite ainsi obtenue est appelée série des moyennes mobiles pondérées d'ordre k.

A.3.4 DOMAINES ET LIMITATIONS

Les moyennes mobiles sont utilisées dans l'étude des séries chronologiques, pour l'estimation de la tendance séculaire, des variations saisonnières et des fluctuations cycliques.

On prendra soin que chaque moyenne successive soit affectée à l'une des dates d'observation pour la série chronologique, ce qui est le cas lorsque l'on fait une moyenne mobile d'ordre impair. Cependant lorsque l'ordre est pair, la moyenne mobile est affectée au centre de l'intervalle séparant deux dates consécutives d'observation.

A.3.5 Propriétés

L'opération « moyenne mobile » appliquée a une série chronologique permet :

- D'éliminer les variations saisonnières dans la mesure où celles-ci sont rigoureusement périodiques
- De lisser les variations irrégulières;
- De conserver approximativement le mouvement extra-saisonnier.

Citons les inconvénients de la méthode des moyennes mobiles :

- Les données de début et de fin de série sont perdues ;
- Les moyennes mobiles peuvent engendrer des cycles ou d'autres mouvements qui n'étaient pas présents dans les données d'origine;
- Les moyennes mobiles sont fortement affectées par les valeurs aberrantes ou accidentelles;

Lorsque le graphique de la série chronologique étudiée présente une tendance séculaire exponentielle ou plus généralement une forte courbure, la méthode des moyennes mobiles ne donne pas des resultats très précis pour l'estimation du mouvement extra-saisonnier.

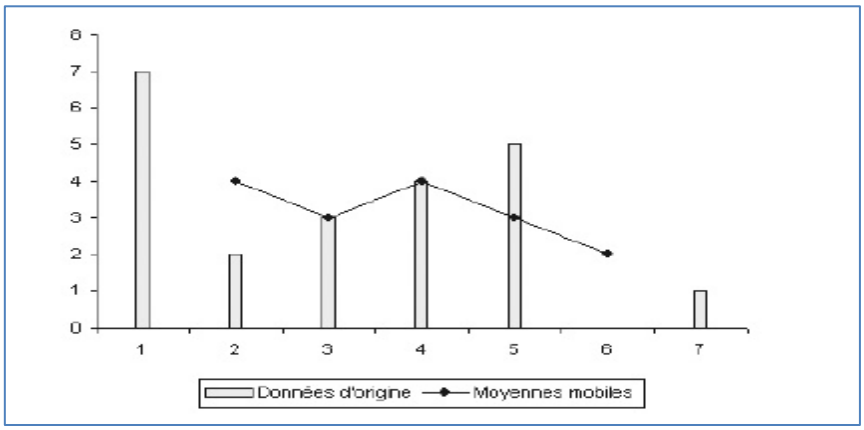
A.3.6 EXEMPLE 3

Etant donne la série chronologique 7, 2, 3, 4, 5, 0, 1,

On obtient la série des moyennes mobiles d'ordre 3 :

$$\frac{7+2+3}{3}, \frac{2+3+4}{3}, \frac{3+4+5}{3}, \frac{4+5+0}{3}, \frac{5+0+1}{3}$$

C'est-a-dire 4, 3, 4, 3, 2.



A.4 La médiane :

A.4.1 Définition

La médiane est une mesure de tendance centrale d´définie comme la valeur qui se trouve au centre d'un ensemble d'observations lorsque celles-ci sont rangées par ordre croissant ou décroissant.

Nous trouvons donc 50 % des observations de chaque côté de la médiane :

Si nous avons un nombre impair d'observations, la médiane correspond donc à la valeur de l'observation du milieu.

Si, en revanche, nous avons un nombre pair d'observations, il n'existe pas une observation unique du milieu. La médiane sera donnée par la moyenne arithmétique des valeurs des deux observations du milieu.

A.4.2 Aspects Mathématiques :

Lorsque les observations nous sont données de façon individuelle, le processus de calcul de la médiane est simple :

1. Classer les n observations par ordre de grandeur.
2. Si le nombre d'observations est impair :
   - repérer l'observation du milieu (n+1)/2
   - la médiane est égale à la valeur de l'observation du milieu.

Exemple 4:

X(note)	5	7	8	9	10

La médiane : $i = 1 \dots 5, n = 5\ (impair); \frac{i+n}{2} = 3\ donc\ M_e = 8$

3. Si le nombre d'observations est pair :
   - repérer les 2 observations du milieu n/2 et (n /2)+1 ;
   - la m´médiane est égale à la moyenne arithmétique des valeurs de ces deux observations.

Exemple 5 :

X(note)	5	7	8	9	10	11

La médiane : $i = 1 \dots 6, n = 6\ (pair); x_{\left(i=\frac{n}{2}=3\right)} = 8; x_{\left(i=\frac{n}{2}+1=4\right)} = 9\ donc\ M_e = \frac{8+9}{2} = 8{,}5$

A.4.3 Propriétés :

- Le calcul de la médiane est rapide.
- Les observations aberrantes n'influencent donc pas la médiane comme elle neutralise l'effet des valeurs extrêmes.

B Les caractères de Dispersion et de concentration :

Ainsi, une fois la moyenne connue, on peut compléter la connaissance d'une série pour apprécier dans quelle mesure les données sont dispersées ou au contraire concentrées autour de la moyenne.

B.1 L'Intervalle De Variation « l'étendue »

L'intervalle, ou « spread » c'est la différence entre la plus grande valeur et la plus petite valeur de la variable.

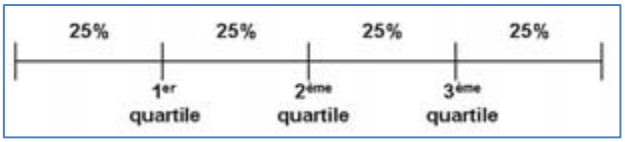
L'intervalle de variation est influencé par les valeurs extrêmes.

Exemple 6 : soit la série suivante {1016, 774, 1008, 8, 1001, 999, 1100} l'intervalle de variation est égale a 1100-1016=84

B.2 Les Quartiles :

B.2.1 Définition :

Les quartiles sont des mesures de position d'une distribution d'observations.
Nous appelons quartiles les quantiles qui partagent une distribution en quatre parties.
Nous aurons donc trois quartiles pour une distribution donnée. Entre chaque quartile se trouvent 25 % des observations :



Notons que le 2éme quartile est égal à la médiane.

B.2.2 Aspects Mathématiques :

Le processus de calcul du quartile est similaire à celui de la médiane.

Lorsque nous possédons toutes les observations brutes, le processus de calcul des quartiles est le suivant :

1. Les **n** observations doivent être organisés sous forme d'une distribution de fréquences.

2. Les quartiles correspondent aux observations pour lesquelles la fréquence relative cumulée dépasse respectivement 25 %, 50 % et 75 %.

B.2.3 DOMAINES ET LIMITATIONS

- Le calcul des quartiles n'a de sens que pour une variable quantitative pouvant prendre des valeurs sur un intervalle déterminé.
- Il permet d'obtenir des informations relatives aux intervalles dans lesquels se situent les quarts successifs de l'ensemble des observations.
- La notion de quartile est similaire à la notion de médiane. Elle est aussi basée sur le rang des observations plutôt que sur leur valeur. Une observation aberrante n'aura donc que peu d'influence sur la valeur des quartiles.

Exemple 7:

Dans un institut technique, on étudie l'arôme d'un jus. Dans cette étude, La qualité des arômes est indiquée par une note fournie par un jury de dégustation selon une échelle croissante de 0 à 10.

Nous avons la série suivante de 31 notes : {1.0 ; 1.0 ;1.5,2.0 ;2.5 ;3.0 ;3.0 ;3.5 ;3.5 ;4.0 ;4.5 ;4.5 ;4.5 ;5.0 ;5.0 ;5.5 ; 5.5 ; 5.5 ;6.0 ; 6.0 ; 6.0 ; 6.0 ; 6.0 ;6.5 ;6.5 ;7.0 ; 7.0 ; 7.0 ;8.0 ; 8.0 ;9.0}

N=31 ; n/4=31/4=7,7 on prend la valeur 7 et 8 pour $Q_1 = \frac{3.0+3.5}{2} = 3.25$

n/2=31/2=15.5 on prend la valeur 15 et 16 pour $Q_2 = \frac{5.0+5.5}{2} = 5.25$ c'est la mediane

3n/4=(3*31)/4=23.25 on prend la valeur 23 et 24 pour $Q_3 = \frac{6.0+6.5}{2} = 6.25$

Cela veut dire :

25% des sondés ont donné une note supérieur à 6.25

25% des sondés ont donné une note entre 5.25 et 6.25

25% des sondés ont donné une note entre 3.25 et 6.25

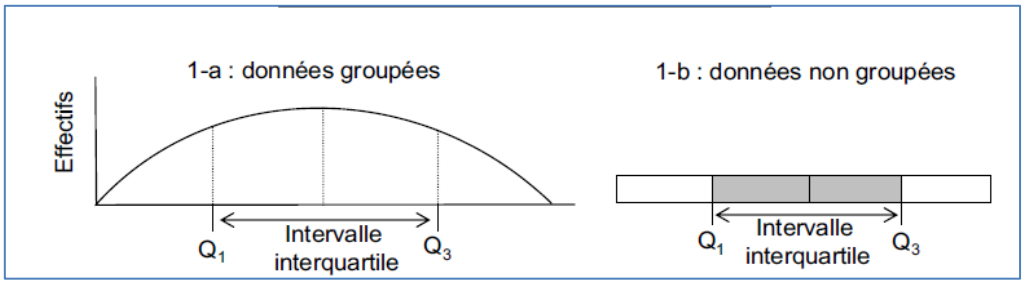
25% des sondés ont donné une note inferieur a 3.25.

B.3 L'INTERVALLE INTERQUARTILE

Sa définition est simple : l'intervalle interquartile mesure l'étendue des 50% de valeurs situées au milieu d'une série de données classées.
Le calcul de l'intervalle interquartile se fait par la formule :

$$\mathbf{IQ = Q_3 - Q_1}$$

L'intervalle interquartile est une mesure de variabilité qui ne dépend pas du nombre d'observations. De plus, cette mesure est nettement moins sensible aux observations aberrantes.



Exemple 8:

Une usine fabrique des boites de conserve de poids μ voici les poids d'une série de vingt boites de conserves :

163.5	165.1	165.4	165.6	166.0	166.4	166.8	168.5	168.8	168.8
169.9	170.0	171.1	171.5	171.8	173.6	173.8	174.4	174.5	174.6

$$IQ = Q_3 - Q_1 = 173.6 - 166.0 = 7.6\ g$$

Nous pouvons donc conclure que 50 % des observations centrée se situent dans un intervalle de longueur 7.6 gr.

B.4 VARIANCE, ÉCART-TYPE ET COEFFICIENT DE VARIATION

La variance, l'écart-type et le coefficient de variation sont les indicateurs les plus fréquemment utilisés pour mesurer la dispersion d'une série. Ces indicateurs renseignent sur la dispersion des données autour de la moyenne.
Plus les données sont concentrées autour de la moyenne, plus les valeurs de ces trois indicateurs sont faibles. Inversement, plus les données sont dispersées autour de la moyenne, plus ces trois indicateurs sont élevés.

B.4.1 La variance et l'écart type:

1 Définition et formule

La variance est généralement designee par S^2 lorsqu'elle est relative a un zchantillon, et σ^2 lorsqu'elle est relative a une population.

Soit une population de N observations relatives a une variable quantitative X. Selon la definition, la variance d'une population se calcule comme suit (Continu):

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

$$\boldsymbol{ou\ \sigma^2 = (\frac{1}{n}\sum_{i=1}^{n} {x_i}^2) - (\overline{x}^2) \quad (Formule\ simplifiée)}$$

2 Propriétés :

L'ecart type caracterise la dispersion d'une serie de données , plus σ est petit plus les données sont regroupés autour de la moyenne arithmetique et plus la population est homogene.

L'ecart type permet de trouver le pourcentage de la population appartenant a un intervalle centré sur l'esperance mathematique (moyenne arithmetique) :
L'intervalle $[\overline{x} - \sigma_x, \overline{x} + \sigma_x]$ contient environ $\frac{2}{3}$ des valeurs.

B.4.2 Le coefficient de variation :

1 Définition

Le coefficient de variation est une mesure relative de dispersion :rarament utilisée ,il a cependant des propriétés importantes.

Il s'exprime sous la forme d'un pourcentage , par l'expresion suivante :

$$CV = \frac{\sigma}{\overline{x}} * 100$$

Proprietes

- le coefficient de variation ne depend pas des unités choisis .
- Il permet d'apprecier la representativité de la moyenne par rapport a l'ensemble des données .cependant si la moyenne est voisine de zéro , ce coefficient tend vers l'infini, il est donc sensibles aux leegres variations de la moyenne .
- Il permet d'apprciere l'homogeneite d'une serie d'observations : une valeur du coefficient de variation inferieur a 15% traduit une bonne homogeineité de cette distribution .
- Il peut etre utilisé pour comparer deux series d'observation dont les moyennes ont des ordres de grandeur trés diffeerntes car, dans ce cas , la comparaison des variance peut conduire a des conclusions erronées .

Exemple 9 :

Nous avons les notes de trois étudiants et cela pour trois examens
Calculer la variance et l'écart type :

	Etudiant1	Etudiant2	Etudiant3	Moyenne $\bar{x}$	Variance	Ecart type
Examen1	10	10	10	10	0	0
Examen2	9	10	8	10	0.666	0.816
Examen3	0	10	20	10	66.666	8.16