I. Introduction :

La statistique est une science ayant pour objet l'étude des phénomènes sociaux surtout ceux donnant lieux à des variations ou ceux ne pouvant être suffisamment maîtrisés que lorsqu'on les étudie dans des ensembles ayant un nombre d'éléments relativement élevé.
On distingue deux branches :

** **La statistique descriptive :** a pour rôle de réunir les observations sur le phénomène à étudier et les grouper dans des tableaux statistiques ou les représenter sur des graphiques.
** **La statistique inductive :** a pour rôle de traiter les informations obtenues en s'appuyant sur des calculs de probabilité, en donnant à ces informations une signification et en faisant des prévisions pour le futur.

II. Vocabulaire statistique :

- **Population :** C'est l'ensemble étudié.
(« population » est employé, ici dans un sens très particulier, elle n'est pas nécessairement humaine.). les éléments de l'ensemble sont appelés : *unités statistiques* ou *individus.*

- **Echantillon :** C'est un sous ensemble quelconque de la population. Si l'échantillon est prélevé au hasard, c'est un échantillon *aléatoire.*

- **Caractère (ou variable) :** C'est l'aspect de l'unité statistique auquel on s'intéresse.
Il peut être *qualitatif :* couleur d'une voiture , etc. ou *quantitatif* : il se traduit alors par un nombre.

- **Valeur statistique** ou **valeur du caractère :** La valeur du caractère est sa mesure lorsqu'on a choisi une unité. On obtient des valeurs de la *variable statistique.*

Exemple :

Population : un ensemble de notes attribuées à 25 élèves dans une classe A.

Tableau de données :

14 15 4 7 11 14 9 10 12
12 16 10 18 13 12 3 11
14 12 10 13 6 12 10 7

Ici le caractère étudié est *quantitatif* (valeur de la note).
On pourrait s'intéresser à la *qualité* :
« être pair » ; « être inférieur à 12 »…

- **Variable discrète (ou discontinue) :** Elle ne prend que des valeurs isolées : $x_1, x_2, ..., x_n$.

Ici variable discrète qui prend treize valeurs.
$x_1 = 3; x_2 = 4; x_3 = 6; ....; x_{13} = 18.$

- **Variable continue :** Elle peut prendre n'importe quelle valeur d'un intervalle [a,b]. Dans ce cas, on peut partager cet intervalle en k intervalles
(partition de [a,b]) :
$a < a_1 < a_2 < .... < a_{k-1} < b$
$[a, a_1[; [a_1, a_2[; ....[a_{k-1}, b[$
Chaque intervalle $[a_i, a_{i+1}[$ est appelé classe ; a_i et a_{i+1} sont les frontières de la classe, $\frac{a_i + a_{i+1}}{2}$ est le centre de la classe.

- **Effectif :** L'effectif de x_i est le nombre n_i d'observations associées à la valeur x_i de la variable statistique, ou l'effectif de la classe $[a_i; a_{i+1}[$.

- **L'effectif total :**
$N = n_1 + n_2 + ... + n_k = \sum_{i=1}^{k} n_i$.

- **Série statistique :** C'est l'ensemble des couples $(x_i; n_i)$ ou $([a_i; a_{i+1}[; n_i)$. On donne souvent cette série sous la forme d'un *tableau statistique.* Ne pas le confondre avec le *tableau de données* (succession de résultats).

Tableau statistique :

x_i	n_i
3	
4	
6	
7	
9	
10	
11	
12	
13	
14	
15	
16	
18	
Total	25

III. Paramètres de position :

- **La dominante ou mode :** C'est la valeur du caractère la plus fréquente. Dans une répartition par classes, on parle de classe modale.
- **La moyenne arithmétique**: C'est le quotient de la somme des mesures par l'effectif total.

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + ... + n_kx_k}{\sum n_i}$$

$$\bar{x} = \sum f_i x_i$$

Dans une répartition par classes, on prend pour x_i le centre de la classe.

- **Médiane :** C'est la valeur Me du caractère telle que l'effectif des individus dont la valeur du caractère est inférieur à Me soit égal à l'effectif des individus dont la valeur du caractère est supérieur à Me.

C'est le réel $x_{\left[\frac{N}{2}\right]+1}$ où $\left[\frac{N}{2}\right]$ est la partie entière de $\frac{N}{2}$

- **Premier quartile :** c'est la valeur Q_1 du caractère telle que au moins 25 % des valeurs lui sont inférieures et au moins 75 % des valeurs lui sont supérieures.

C'est le réel $x_{\left[\frac{N}{4}\right]+1}$ où $\left[\frac{N}{4}\right]$ est la partie entière de $\frac{N}{4}$

- **Troisième quartile :** c'est la valeur Q_3 du caractère telle que au moins 75 % des valeurs lui sont inférieures et au moins 25 % des valeurs lui sont supérieures.

C'est le réel $x_{\left[\frac{3N}{4}\right]+1}$ où $\left[\frac{3N}{4}\right]$ est la partie entière de $\frac{3N}{4}$

Exemple (des notes attribuées aux 25 élèves)

La dominante est …… ; elle est unique. C'est une série *uni modale.*

Calcul de la moyenne arithmétique :

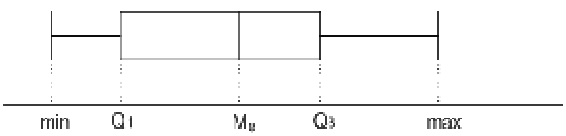
$\bar{x} =$

La médiane est ……….

Le premier quartile est …………………

Le troisième quartile est …………………

Remarque : On peut représenter une série par son diagramme en boîte qui fait intervenir ses valeurs extrêmes (min et max.), ses quartiles **Q_1** et **Q_3** et sa médiane **M_e** :



IV. Paramètres de dispersion :

- **L'étendue :** Différence entre la plus grande et la plus petite des valeurs observées.

L'étendue est ………..

- **L'écart moyen :** $e = \dfrac{\sum_{i=1}^{i=p} n_i \left| x_i - \overline{x} \right|}{\sum_{i=1}^{i=p} n_i}$

- **La variance :** $v = \dfrac{\sum n_i \left(x_i - \overline{x} \right)^2}{\sum n_i}$

V =………

Formule de *Koenig* : $v = \dfrac{\sum n_i x_i^{\,2}}{\sum n_i} - \overline{x}^2$

- **L'écart- type :** $\sigma = \sqrt{v}$

$\sigma = \sqrt{V} \approx ....$

C'est une mesure de dispersion qu'on utilise pour mesurer la dispersion des valeurs d'une série statistique autour de la moyenne de cette série
Un écart type important signifie que les valeurs de la série s'éloignent souvent et de façon importante de la moyenne.

$Q_3 - Q_1 = ...$

- **L'écart interquartile :** c'est la différence $Q_3 - Q_1$ qui représente l'étendue de la distribution sur la quelle se trouve concentrée la moitié des éléments dont les valeurs de X sont les moins différentes de la médiane. On exclut alors les 25 % des valeurs les plus faibles et les 25 % des valeurs les plus élevées.

V. Série statistique à deux variables

1. Introduction

Une série statistique à deux variables, X et Y, est le résultat de l'observation des deux caractères X et Y pour chaque individu d'une population.

Lorsque les caractères sont quantitatifs discrets, on peut associer, à chaque individu i, un couple de nombres réels noté (x_i, y_i).

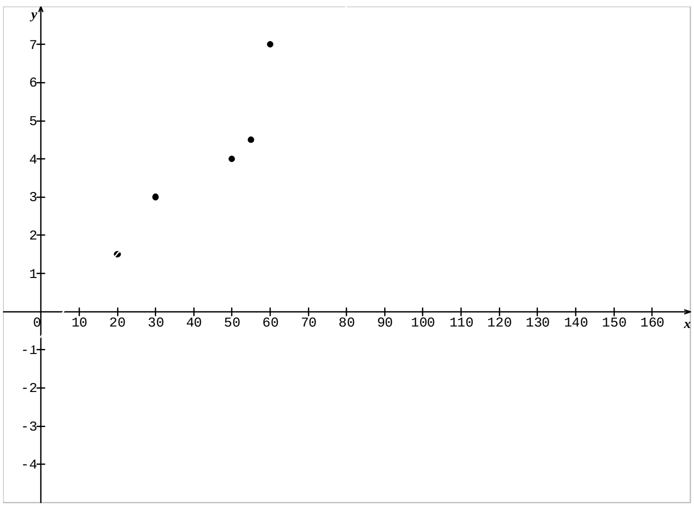
Exemple :

Le tableau suivant donne, en millions de dinars, le chiffre d'affaires x_i et la somme consacrée aux dépenses de publicité y_i pour cinq entreprises :

Entreprises	x_i	y_i
A	30	3
B	55	4,5
C	60	7
E	20	1,5
F	50	4

Définition

Dans un repère orthogonal du plan, le nuage de points associés à la série statistique à deux variables, X et Y, est l'ensemble des points M_i de coordonnées (x_i, y_i) représentatifs de tous les individus i de la population.



2. Point moyen

On note X le caractère : « le chiffre d'affaires de chaque entreprise » et Y : « les dépenses de publicité ».
Calculer la moyenne, la variance et l'écart – type de chaque caractère.
On rappelle que :

$$\overline{x} = \frac{1}{N}\sum_i n_i x_i$$

$$V(X) = \frac{1}{N}\sum n_i\left(x_i - \overline{x}\right)^2 = \frac{1}{N}\sum n_i x_i^2 - \overline{x}^2 \quad \text{et} \quad \sigma(X) = \sqrt{V(X)}.$$

$\overline{x} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots$; $V(X) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$; $\sigma(X) = \ldots\ldots\ldots\ldots\ldots\ldots$

$\overline{y} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots$; $V(Y) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$; $\sigma(Y) = \ldots\ldots\ldots\ldots\ldots\ldots$

Définition

Le point $G\left(\overline{x}, \overline{y}\right)$ est appelé point moyen du nuage de points associé à la série statistique à deux variables X et Y.
Exemple : Placer le point moyen G dans le repère précédent.

3. Liaison entre deux caractères

Droite d'ajustement

Lorsque le nuage a tendance de s'accumuler autour d'une droite, alors on cherche une équation de la droite D qui approche le « mieux possible » les points du nuage, c'est ce qu'on appelle un ajustement linéaire.

On divise le nuage de points en deux parties contenant à peu prés le même nombre de points obtenant ainsi deux nuages de points. On désigne par G_1 et G_2 les points moyens de ces deux nuages.

La droite (G_1G_2) passe par le point G et définit un ajustement affine du nuage de points représentant la série statistique (X, Y).

Tracer dans le repère précédent la droite (G_1G_2).

<u>*Exercices 1 p 212, 9 p 214 et 15 p 217.*</u>