

Statistiques

I Adaptons le vocabulaire !

A une épreuve de mathématiques au baccalauréat professionnel, on a relevé les notes suivantes :

Notes x_i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Effectifs n_i	5	5	4	5	8	10	5	3	3	8	6	4	6	9	4	5	2	6	3	1

La statistique descriptive a pour objet de recueillir et de classer des faits mesurables (ou identifiables) relatifs aux éléments d'un ensemble donné.

Aujourd'hui, la statistique est indispensable dans de nombreux domaines : étude des populations, économie, biologie, sociologie, industrie, ...

Une étude statistique porte sur un ensemble appelé **population** (ensemble des individus sur lequel vont porter les observations). Les éléments de cet ensemble sont appelés **individus** ou **unités statistiques**. On parle aussi parfois d'**échantillon** (partie de la population).

Le nombre total d'individus est l'**effectif** de la population. La propriété étudiée (mesurable ou identifiable) correspond au caractère statistique.

Ce caractère peut être :

qualitatif s'il n'est pas mesurable (couleur ...) on parle alors de nature du caractère

quantitatif s'il est mesurable, le nombre est alors la valeur du caractère.

Remarque : Si la valeur possible du caractère ne peut-être qu'isolée (ex : une note, un âge , ...) on dit que la variable est **discrète** ou discontinue

Si le caractère peut prendre toutes les valeurs d'un intervalle, on dit que le caractère est **continu**.

Une **classe** est un groupe de valeurs voisines prises par un caractère et définie par un intervalle **[a ; b[** où a est la borne inférieure

b est la borne supérieure

b – a est l'amplitude

Notion de **fréquence** : on appelle fréquence de la valeur x_i du caractère et on note n_i le rapport de l'effectif partiel de cette valeur à l'effectif total de la population

$$f_i = \frac{n_i}{N}$$

Remarques :

On appelle **effectif cumulé croissant** de la valeur x_i , la somme des effectifs de toutes les valeurs du caractère inférieures ou égales à x_i .

On appelle **effectif cumulé décroissant** de la valeur x_i , la somme des effectifs de toutes les valeurs du caractère supérieures ou égales à x_i .

On parle également de fréquence cumulée croissante ou décroissante.

II Les principales représentations graphiques

le diagramme circulaire ou semi-circulaire (diagramme à secteurs) : cette représentation est utilisée lorsque l'étude porte sur un caractère qualitatif : la mesure du secteur angulaire est proportionnelle à l'effectif.

le diagramme en bâtons (diagramme à barres ou à tuyaux d'orgue) : cette représentation est utilisée lorsque le caractère de l'étude est discret : la longueur du segment ou de la barre est proportionnelle à l'effectif.

l'histogramme : représentation utilisée lorsque l'étude porte sur un caractère continu : les rectangles ont pour bases les amplitudes des classes et leurs aires sont proportionnelles aux effectifs. Si les classes ont même amplitude, tous les rectangles ont une même base (largeur) ; si les classes n'ont pas la même amplitude, la hauteur du rectangle est déterminée par le rapport de l'effectif et de l'amplitude de la classe.

Le polygone : le polygone des effectifs cumulés croissants est une ligne joignant les points :

Dont l'abscisse est la borne supérieure de la classe

Dont l'ordonnée est l'effectif cumulé croissant de la classe.

Etude de situation :

1°) La répartition des salaires mensuels des employés d'une entreprise est la suivante :

Salaires	Effectifs	ECC	ECD	Fréquences	FCC	FCD
[6000 ; 6500 [20					
[6500 ; 7000 [35					
[7000 ; 8000 [45					
[8000 ; 9000 [30					
[9000 ; 10000 [25					
[10000 ; 12000 [25					
[12000 ; 14000 [15					
[14000 ; 15000 [5					

Compléter le tableau

Tracer l'histogramme des effectifs

Tracer le polygone des ECC et des ECD dans un même repère.

2°) Soit le tableau suivant :

Branche d'activité	Nombre d'emplois	Fréquences en %	Angle du secteur
Agriculture	1700000		
Pêche	60000		
Activités hippiques	140000		
Filière forêt-bois	500000		
Tertiaire agricole	20000		
Industrie agro-alimentaire	400000		

Tracer le diagramme à secteurs correspondant

III Les caractéristiques de position

1°) Mode d'une série statistique

Cas d'une série à caractère discret : dans ce cas , **le mode** de la série est la valeur du caractère statistique ayant le plus grand effectif.

Cas d'une série à caractère continu : dans ce cas on définit **la classe modale** qui est alors la classe qui le plus grand effectif par intervalle unitaire.(le centre d'une classe modale est appelé le mode)

Exemples : La répartition du personnel de deux entreprises A et B suivant leurs salaires mensuels en F est donnée par les tableaux suivants :

Tracer l'histogramme correspondant

Quelle est la classe modale de la série ? Quel est son mode ? Quel est le salaire le plus fréquent ?

Entreprise A

Salaires en F	Effectif n_i
[5000 ; 7000 [180
[7000 ; 9000 [240
[9000 ; 11000 [280
[11000 ; 15000 [140
[15000 ; 20000 [60

Entreprise B

Salaires en F	Effectif n_i
[5000 ; 7000 [310
[7000 ; 9000 [250
[9000 ; 11000 [180
[11000 ; 15000 [150
[15000 ; 20000 [90
[20000 ; 30000 [20

Remarque : On constate que les deux modes des distributions A et B sont très différents, bien que comme nous le verrons plus tard, les salaires moyens soient proches. Pour un caractère quantitatif, il faut rester très prudent dans l'utilisation du mode.

C'est un indicateur intéressant dans le cas d'un caractère qualitatif.

2°) La médiane

La **médiane** est la valeur du caractère qui partage l'effectif total en deux parties de même effectif. cas d'un caractère discret : si l'effectif total est impair, la médiane est la valeur du caractère situé au milieu de la série et, si l'effectif est pair, la médiane est la demi-somme des deux valeurs centrales

Cas d'un caractère continu : une méthode graphique consiste à tracer dans un même repère le polygone des ECC et des ECD, l'abscisse du point d'intersection est alors la valeur de la médiane.

Exemple :

Le tableau suivant donne la répartition des clients d'un magasin sur une journée d'après l'importance de leurs achats, déterminer la médiane de cette série.

Classe	n_i	ECC	ECD
[0 ; 100 [65		
[100 ; 250 [83		
[250 ; 500 [22		
[500 ; 1000 [18		
[1000 ; 1500 [12		

3°) La moyenne

Calculer les salaires moyens correspondant aux entreprises A et B du 1°)

Les salaires moyens des deux entreprises sont peu différents, pourtant la répartition des salaires n'est pas identique.

La **moyenne** est donc insuffisante à elle seule pour caractériser une distribution. La comparaison des écarts-type permet d'apprécier la dispersion des salaires autour des salaires moyens.

Définition : la moyenne, notée \bar{x} , d'une série statistique est un paramètre de position qui donne une tendance centrale de la série. Elle se calcule en utilisant la formule :

$$\bar{X} = \frac{\sum n_i X_i}{N} \quad \text{avec } n_i \text{ effectif de la } i^{\text{ème}} \text{ classe et } x_i \text{ valeurs du caractère (ou centres de classes)}$$

IV les caractéristiques de dispersion

1°) L'étendue

L'**étendue** d'une série statistique est la différence entre la plus grande valeur et la plus petite valeur du caractère.

Calculer les étendues des séries statistiques étudiant les salaires

2°) L'écart-type

L'**écart-type** d'une série statistique est un paramètre de dispersion qui mesure la dispersion des valeurs de la variable autour de leur moyenne

Plus l'écart-type est grand et plus les valeurs de la variable sont dispersées.

$$\text{Variance} \quad V = \frac{\sum n_i (x_i - \bar{x})^2}{N} \quad V = \frac{\sum n_i x_i^2}{N} - \bar{x}^2$$

$$\text{Ecart-type} \quad \hat{\sigma} = \sqrt{V}$$

Calculer les écarts-type des deux séries étudiées