

LA STATISTIQUE

I Statistiques à une variable

Vocabulaire

La statistique est une discipline scientifique qui étudie quantitativement les CARACTERES des INDIVIDUS d'une POPULATION.

Le CARACTERE peut – être QUALITATIF ou QUANTITATIF.

Les VALEURS du CARACTERE sont appelées LES VARIABLES ou LES MODALITES.

Définition

Un caractère est QUANTITATIF lorsque les modalités sont numériques

Un caractère est QUALITATIF lorsque les modalités NE sont PAS numériques

Exemple :

Dans une entreprise de construction de bungalows on veut étudier les commandes des clients, et plus particulièrement la structure et la surface des bungalows commandés.

Population : Ensemble des bungalows livrés.

Individu : Un bungalow

Caractère quantitatif : Structure des bungalows

Variables : Béton (parpaings) , béton (voile), béton et bois, bois .

Ici on a obtenu une SERIE QUALITATIVE.

Caractère qualitatif : surface des bungalows

Variables : 20 m² ; 25m² ; 30 m² ; 35 m².

Ici on a obtenu une série QUANTITATIVE DISCRETE.

1°) Etude d'une série statistique quantitative discrète

Exemple :

On étudie les notes de mathématiques de deux classes de BTS obtenues à un examen blanc.

Dans la classe de BTS 1 on a les résultats suivants :

2 , 3, 7, 7, 10, 11, 14, 16, 16, 17

Dans la classe de BTS 2 on a les résultats suivants :

8, 8, 8, 9, 10, 10, 11, 11,11, 11, 16

On voit que dans la classe de BTS 1 les élèves obtiennent 2 fois la note 7, l'effectif de la note 7 est donc 2. Il ya 10 élèves donc l'effectif total est 10,

le quotient $\frac{2}{10}$ est la fréquence de la note 7.

Pour chaque série quantitative on va résumer les résultats dans un tableau :

Classe de BTS 1

Variables x_i	2	3	7	10	11	14	16	17	Total
Effectifs n_i	1	1	2	1	1	1	2	1	10
Fréquences f_i	10%	10%	20%	10%	10%	10%	20%	10%	100%

Exercice : compléter le tableau

Classe de BTS 2

Variables x_i						Total
Effectifs n_i						
Fréquences f_i						

CAS GENERAL

Soit la série statistique S quantitative résumée dans le tableau suivant :

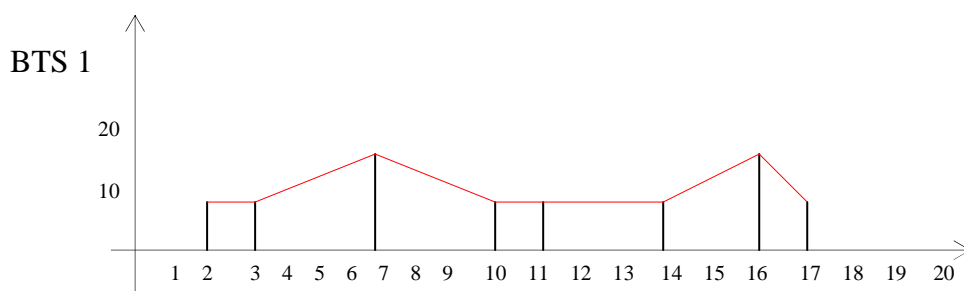
Variables x_i	x_1	x_2	...	x_p	TOTAL
Effectifs n_i	n_1	n_2	...	n_p	N
Fréquences f_i	$f_1 = \frac{n_1}{N}$	$f_2 = \frac{n_2}{N}$...	$f_p = \frac{n_p}{N}$	1

Remarque : $N = n_1 + n_2 + \dots + n_p$

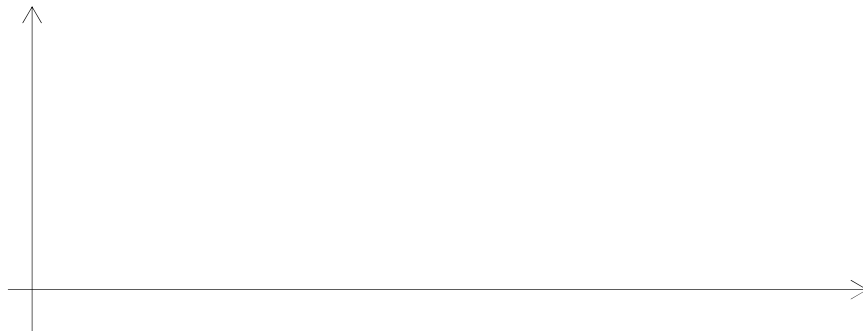
Représentation graphique : diagrammes en bâtons

Chaque donnée est représentée par un bâton dont **la hauteur est proportionnelle à l'effectif ou à la fréquence**. On peut compléter le diagramme en dessinant le polygone des effectifs. Il est formé des segments joignant les sommets des bâtons.

Diagramme des fréquences en %



Exercice : Construire le diagramme en bâtons de la série de BTS 2

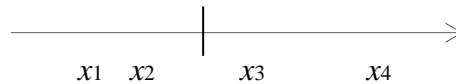


MEDIANE :

Dire que M_e est la médiane d'une série statistique signifie qu'il y a autant de valeurs de la série supérieures à M_e que de valeurs de la série inférieures à M_e .

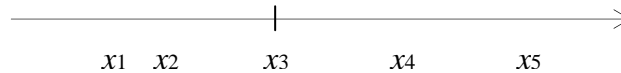
Si la série statistique a pour effectif total N :

- Avec N PAIR : $M_e = (x_2 + x_3) / 2$



La médiane est demi – somme des deux valeurs centrales, $[x_2 ; x_3]$ est l'intervalle médian.

- Avec N IMPAIR : $M_e = x_3$



La médiane est la valeur centrale.

Dans le cas de la série du BTS1 : 2 , 3 , 7 , 7 , 10 , 11 , 14 , 16 , 16 , 17

.....
Dans le cas de la série du BTS 2 : 8 , 8 , 8 , 9 , 10 , 10 , 11 , 11 , 11 , 11 , 16

QUARTILES

La médiane M_e sépare une série statistique en deux sous- séries de même effectif, l'une contenant les plus petites valeurs, l'autre les plus grandes.

LES QUARTILES sont les médianes de ces sous – séries.

Le premier quartile noté Q_1 est la médiane de la sous - série inférieure, on dit aussi que au moins un quart (ou 25%) des valeurs prises sont inférieures ou égales à Q_1 .

Le troisième quartile noté Q_3 est la médiane de la sous – série supérieure, on dit aussi que au moins trois quarts (ou 75%) des valeurs prises sont supérieures ou égales à Q_3 .

A Noter : La médiane M_e est parfois appelée aussi deuxième quartile.

Définition

Le nombre $Q_3 - Q_1$ est appelé ECART INTERQUARTILE.
L'intervalle $] Q_1 ; Q_3[$ est appelé INTERVALLE INTERQUARTILE.

Remarques :

1. Les nombres Q_1 , M_e , Q_3 permettent de couper la population étudiée en quatre groupes contenant chacun le même nombre d'éléments.
2. **NOTION DE DECILES** : on peut définir de façon analogue la notion de déciles qui sont des nombres qui permettent de couper la population en dix groupes contenant chacun le même nombre d'éléments.

On utilise alors **D_1 le premier décile, et D_9 le neuvième décile.**

Au moins un dixième (ou 10%) des valeurs prises sont inférieures ou égales à D_1

Au moins neuf dixièmes (ou 90%) des valeurs prises sont supérieures ou égales à D_9 .

Exemples :

Dans le cas de la série du BTS1 : 2, 3, 7, 7, 10, 11, 14, 16, 16, 17

$$Q_1 = 7 \quad Q_3 = 16$$

Dans le cas de la série du BTS 2 : 8, 8, 8, 9, 10, 10, 11, 11, 11, 16

Exemples pour les déciles :

On considère deux entreprises A et B de 100 salariés dont la grille des salaires est décrite par les séries statistiques ci-dessous :

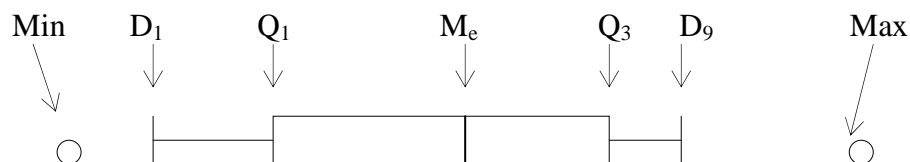
Salaires mensuels En milliers d'euros	1	1.1	1.2	1.3	1.4	1.5	2	2.2	2.4	2.5
Effectifs Entreprise A	25	20	15	15	8	7	6	2	1	1
Effectifs Entreprise B	20	15	12	10	10	7	8	8	6	4

Pour A : $D_1 = Q_1 = 1$ $Q_3 = 1.3$ $D_9 = 1.5$ $M_e = 1.2$

Pour B : $D_1 = 1$ $Q_1 = 1.1$ $Q_3 = 2$ $D_9 = 2.2$ $M_e = 1.3$

DIAGRAMMES EN BOITE

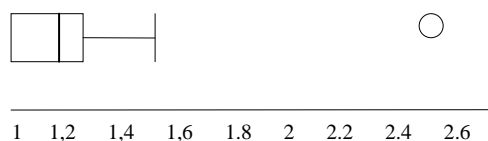
Pour utiliser les quartiles et la médiane on utilise une représentation simple appelée « diagramme en boîte » ou « diagramme à moustaches ».



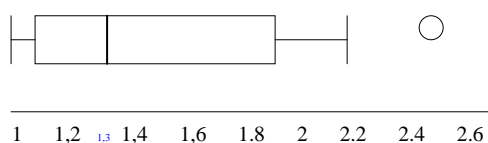
Droite graduée

Dans nos exemples des entreprises A et B :

Entreprise A



Entreprise B



Exercices : Construire les diagrammes en boîtes des séries des BTS 1 et 2

MOYENNE, VARIANCE, ECART - TYPE

Variables x_i	x_1	x_2	...	x_p	TOTAL
Effectifs n_i	n_1	n_2	...	n_p	N
Fréquences f_i	$f_1 = \frac{n_1}{N}$	$f_2 = \frac{n_2}{N}$...	$f_p = \frac{n_p}{N}$	1

Lorsque les valeurs sont nombreuses, le tableau peut – être aussi interprété à l’aide de quelques valeurs caractéristiques numériques. La plus simple et la plus connue est la moyenne.

Définition

La moyenne notée \bar{x} de la série statistique S est :

Formule avec les effectifs :
$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

Où $N = n_1 + n_2 + \dots + n_p$

Formule avec les fréquences :
$$\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_p x_p$$

Exemple : dans les deux classes de BTS 1 la moyenne de mathématiques est $\bar{x}_1 = 10,3$.
 et de BTS 2 $\bar{x}_2 \approx 10,27$.

Définition

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

ou

$$V = \frac{1}{N} \sum_{i=1}^n n_i x_i^2 - \bar{x}^2 = \frac{1}{N} [n_1 x_1^2 + n_2 x_2^2 + \dots + n_p x_p^2] - \bar{x}^2$$

L'écart – type de S est $\sigma = \sqrt{V}$

Exemples :

Donner la variance et l'écart – type des séries du BTS 1 et du BTS2. Quelle conclusion peut – on faire ?

2°) Etude d'une série statistique quantitative continue

Exemple : Lors d'un contrôle cent câbles sont testés . Considérons la série statistique qui à chaque câble, associe sa charge de rupture en g.

711	862	851	912	922	791	825	975	895	758
915	873	926	864	800	931	722	774	903	925
853	700	885	857	844	907	917	786	820	930
789	790	753	910	847	784	936	706	758	887
941	909	784	882	859	903	925	704	792	888
890	925	895	768	869	892	895	912	850	960
763	805	796	759	916	853	789	942	712	764
892	893	915	890	888	865	909	931	710	798
914	794	931	701	772	935	887	880	933	905
889	791	782	713	724	868	842	892	955	792

Le dépouillement de cette série de données brutes consiste à regrouper les valeurs en intervalles appelés CLASSES.

L'effectif n_i est alors le nombre de valeurs observées dans cette classe.

On dit que l'on a une VARIABLE CONTINUE.

On décide dans notre exemple que le regroupement se fera selon les six classes suivantes :

[700 ; 750[[750 ; 800 [[800 ; 850 [[850 ; 900 [[900 ; 950 [[950 ; 1000[

Certaines informations seront perdues (ordre d'acquisition des données, la valeur exacte des effectifs...)

On va résumer les résultats dans un tableau :

Variables x_i	[700 ; 750[[750 ; 800 [[800 ; 850 [[850 ; 900 [[900 ; 950 [[950 ; 1000 [Total
Effectifs n_i	10	23	7	29	26	5	100
Fréquences f_i	10%	23%	7%	29%	26%	5%	1

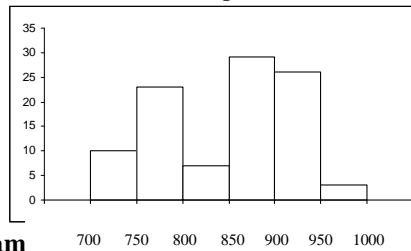
HISTOGRAMME

Dans un histogramme, chaque classe est représentée par un rectangle dont **la largeur est égale à l'amplitude de la classe** et dont **l'aire est proportionnelle à l'effectif de la classe**.

Deux cas se présentent :

- Les classes sont d'amplitudes égales : on prend pour hauteur l'effectif n_i .
- Les classes ne sont pas d'amplitudes égales : On se ramène au cas précédent. En partageant chaque classe en sous – classes de MEME AMPLITUDE. On prend ensuite pour hauteur de chaque rectangle, l'effectif de la sous – classe qu'il représente. L'AMPLITUDE DE CHAQUE SOUS – CLASSE EST L'UNITE D'AMPLITUDE DE L'HISTOGRAMME.

Dans notre exemple : Toutes les classes ont la même amplitude c'est – à – dire 50.



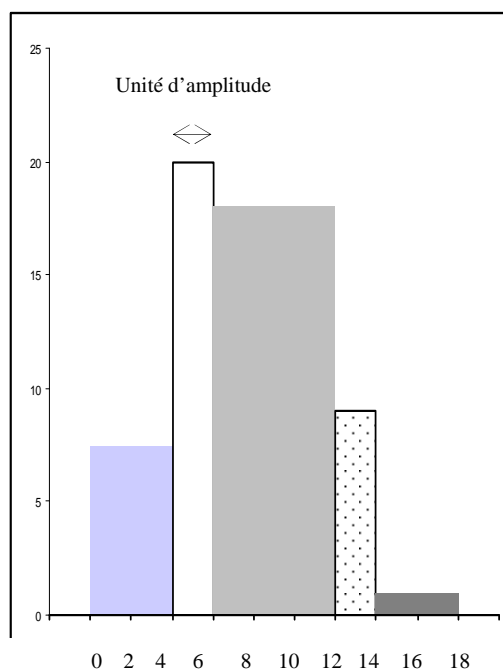
Exemple dans le cas de classes d'am

On considère les notes obtenues à un examen blanc de mathématiques en BTS :

Variables x_i	[0 ; 4[[4 ; 6[[6 ; 12[[12 ; 14[[14 ; 18[Total
Effectifs n_i	15	20	54	9	2	100

On a des classes dont l'amplitude varie de 2 à 6. On va donc prendre ici 2 pour l'amplitude des sous – classes. L'unité d'amplitude est donc 2. On a le tableau suivant :

Variables x_i	[0 ; 4[[4 ; 6[[6 ; 12[[12 ; 14[[14 ; 18[
Effectifs n_i	15	20	54	9	2
Amplitude	4	2	6	2	4
Nombre de sous- classes	2	1	3	1	2
Effectif de Sous – classes	$15/2=7,5$	20	18	9	1



ATTENTION : les calculatrices fournissent souvent des histogrammes faux.

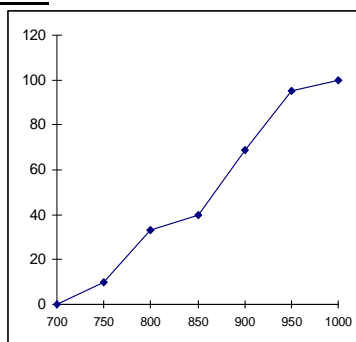
Remarque : le polygone des effectifs complète le plus souvent l'histogramme. Il est obtenu en joignant, par des Segments les points dont les abscisses sont les centres des sous – classes et dont les ordonnées sont les effectifs correspondants. Un sous – classe d'effectif nul est ajoutée à chaque extrémité de l'histogramme.

Effectifs cumulés croissants et fréquences cumulées croissantes

Dans le cas d'une série statistique continue on peut déterminer les effectifs cumulés croissants et Les fréquences cumulées croissantes :

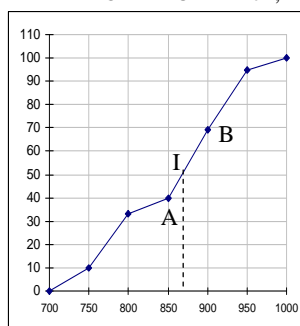
Variables x_i	[700 ; 750[[750 ; 800 [[800 ; 850 [[850 ; 900 [[900 ; 950 [[950 ; 1000 [
Effectifs Cumulés Croissants	10	10 + 23 = 33	40	69	95	100
Fréquences f_i	10%	33%	40%	69%	95%	100%

Polygone des effectifs cumulés croissants



MEDIANE

Méthode graphique : On admet que la Médiane est égale à l'abscisse du point de la représentation graphique Dont L'ORDONNEE est égale à la MOITIE DE L'EFFECTIF TOTAL ici , $100/2 = 50$.On lit $M \approx 870$.

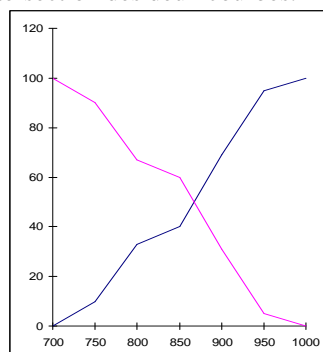


Méthode algébrique :

A(850 ; 40) B(900 ; 70) I (M_e ; 50)

$$\frac{y_I - y_A}{x_I - x_A} = \frac{y_B - y_A}{x_B - x_A} \quad \text{soit} \quad \frac{50 - 40}{M - 850} = \frac{70 - 40}{900 - 850} \quad \text{et } M = 850 + 50/3 \approx 866,67$$

Remarque : En construisant le polygone des effectifs cumulés décroissants on obtient graphiquement la Médiane qui est l'abscisse du point d'intersection des deux courbes.



Variables x_i	[700 ; 750[[750 ; 800 [[800 ; 850 [[850 ; 900 [[900 ; 950 [[950 ; 1000 [
Effectifs Cumulés décroissants	100	90	67	60	31	5
Fréquences f_i	100%	33%	40%	69%	95%	5%

MOYENNE, VARIANCE, ECART – TYPE

Dans le cas d'une série statistique continue **les centres des classes c_i** remplacent les x_i dans les formules données pour une série statistique discrète :

Variables c_i	c_1	c_2	...	c_p	TOTAL
Effectifs n_i	n_1	n_2	...	n_p	N
Fréquences f_i	$f_1 = \frac{n_1}{N}$	$f_2 = \frac{n_2}{N}$...	$f_p = \frac{n_p}{N}$	1

Définition

La moyenne notée \bar{x} de la série statistique S est :

Formule avec les effectifs :
$$\bar{x} = \frac{n_1 c_1 + n_2 c_2 + \dots + n_p c_p}{N}$$

Où $N = n_1 + n_2 + \dots + n_p$

Formule avec les fréquences :
$$\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_p x_p$$

Exemple : Calculer la moyenne pour la série des câbles.

Définition

$$V = \frac{1}{N} \sum_{i=1}^n n_i (c_i - \bar{x})^2 = \frac{1}{N} [n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + \dots + n_p (c_p - \bar{x})^2]$$

ou

$$V = \frac{1}{N} \sum_{i=1}^n n_i c_i^2 - \bar{x}^2 = \frac{1}{N} [n_1 c_1^2 + n_2 c_2^2 + \dots + n_p c_p^2] - \bar{x}^2$$

L'écart – type de S est $\sigma = \sqrt{V}$

Exemples :

Donner la variance et l'écart – type de la série des câbles.

.....

.....

.....

.....