CHAPITRE 9 - STATISTIQUES

A) Rappels

1) Moyenne

Dans une série simple, c'est juste la moyenne des valeurs (total des valeurs divisé par nombre de valeurs).

Dans une série avec effectifs, on pondère chaque valeur par son effectif (on les multiplie), on fait la somme, et on divise le total par le total des effectifs.

2) Médiane

La médiane est une valeur, c'est la valeur se trouvant au milieu de l'effectif après avoir trié par ordre croissant sur les valeurs.

Si l'effectif total est pair, il y aura deux valeurs au milieu. On fait alors la moyenne de ces deux valeurs.

B) Compléments

1) Les quartiles, déciles, centiles

a) Les quartiles

Dans une série statistique **ordonnée**, on appelle premier quartile (Q_1) la plus petite valeur telle que 25% des termes de la série lui soit inférieure ou égale.

Même définition en remplaçant 25% par 75% pour le troisième quartile (Q_3)

Remarques :

. Il n'y a pas de deuxième quartile, car cela correspondrait à la médiane.

. Une autre définition, utilisée souvent dans les calculatrices, donne les premier et troisième quartiles comme les médianes respectives de la première moitié et de la seconde moitié des effectifs.

Sur de grandes séries, le résultat est sensiblement le même.

b) L'écart interquartile

C'est la différence $Q_3 – Q_1$.

Ce nombre est un indicateur de la dispersion de la série, mais contrairement à l'étendue, il fait abstraction des 25% les plus petits et des 25% les plus grands.

c) Déciles et centiles

Ils fonctionnent comme les quartiles, mais en divisant l'effectif en 10 ou en 100 au lieu de le diviser en 4.

Il y a donc 8 déciles (D1 à D4 et D6 à D9), puisque le cinquième serait la médiane.

On utilise souvent les déciles D1 et D9 pour avoir un écart interdécile (= D9 - D1), intermédiaire entre l'étendue et l'écart interquartile.

Question : combien y a-t-il de centiles ?

(réponse : 99 -1 – 8 = 90)

d) Exemples

Déterminer les quartiles pour les séries suivantes des deux façons vues ci-dessus (ici, les séries sont trop petites pour que les déciles aient un intérêt).

a) 1, 4, 4, 4, 5, 7, 9, 21

b) 3, 9, 9, 11, 12, 13, 13, 15, 17

c) 2, 7, 7, 9, 10, 10, 10, 13, 19, 27

2) La boîte à moustaches (ou boîte à pattes)

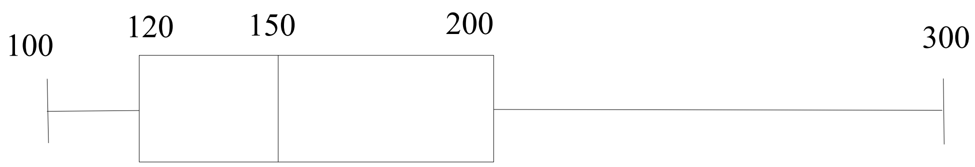
C'est un diagramme très utilisé, qui fait intervenir les quartiles, les extremums et la médiane :

Exemple :

Soit une série statistique où les valeurs vont de 100 à 300, où l'on a comme médiane 150, et où les quartiles sont 120 et 200.

Autrement dit, plus petite valeur = 100, plus grande valeur = 300, M = 150, Q1 = 120, et Q3 = 200.

Boîte à moustaches :



On peut voir d'après ce diagramme que 25 % des effectifs ont une valeur entre 100 et 120, donc assez proche, par contre les derniers 25 % sont étalés entre 200 et 300.

Il y a donc plus de dispersion dans les grosses valeurs que dans les petites.

Cela se voit aussi dans la médiane, qui est plus proche du premier quartile que du troisième.

Règle de construction de la boîte :

La distance entre deux points de la boîte doit être proportionnelle à la différence entre les deux valeurs correspondantes.

Par exemple, si je mets 4cm entre 100 et 120, je dois en mettre 6 entre 120 et 150, 10 entre 150 et 200, et 20cm entre 200 et 300.

Remarque :

Lorsqu'il y a beaucoup de valeurs possibles, on fait aussi des boîtes à moustaches qui s'arrêtent, non pas aux valeurs extrêmes, mais au premier et dernier décile.

Dans ce cas, on rajoute le minimum et le maximum comme de gros points à gauche et à droite des extrémités des moustaches.

Exemple :

Faire la boîte à moustaches pour l'exemple suivant :

Valeur	10	13	15	17	22	27	31	32
Effectif	5	8	4	14	11	2	2	4

C) La variance et l'écart-type

1) Introduction

Au conseil de classe, on veut départager les trois meilleurs élèves, Alain, Bernard et Catherine.

Voici leurs moyennes dans les matières générales :

Matière/ Moyennes	Français	Maths	Physique	SVT	LV1	Hist-Géo	EPS
Alain	15	18	14	11	12	10	18
Bernard	14	17	16	13	15	11	12
Catherine	17	12	15	17	11	13	13

Pour les départager, on aimerait récompenser celui dont le travail est le plus régulier dans toutes les matières.

Calculer pour chacun des élèves, la moyenne générale, puis les écarts entre note et moyenne et remplir le tableau ci-dessous pour chaque élève en notant pour chaque matière l'écart avec sa moyenne générale :

Moyenne	Générale	Français	Maths	Physique	SVT	LV1	Hist-Géo	EPS
Alain								
Bernard								
Catherine								

Ce que l'on cherche donc à mesurer ici, c'est donc la « dispersion » des termes de la série autour de la moyenne.

Que se passe-t-il si l'on fait la somme des différences des notes avec la moyenne ?

Une idée serait donc de faire la moyenne des écarts, pris en **valeur absolue**.

On appelle cette valeur <u>l'écart absolu moyen</u> :

Malheureusement, la valeur absolue se prête mal aux manipulations mathématiques, aussi préfère-t-on utiliser la moyenne des carrés des écarts, qu'on appelle variance.

Calculer pour chaque élève l'écart absolu moyen et la variance.

Peut-on comparer ces deux résultats ?

Non, c'est pourquoi on prendra la racine carrée de la variance, qu'on appelle l'écart-type.

2) Définitions et théorème

a) Définitions

On appelle variance d'une série statistique et on note V la moyenne des carrés des écarts entre chaque valeur et la valeur moyenne de la série.

On appelle écart-type la racine carrée de la variance et on la note σ.

L'avantage de prendre la racine carrée de V est qu'ainsi, l'écart-type aura la même unité que la valeur.

Si on note n_i les effectifs, x_i les valeurs et N le nombre de valeurs, on aura :

$\boxed{V=\frac{1}{N}\sum_{i=1}^{N} n_i(x_i-\bar{x})^2.}$ et $\boxed{\sigma=\sqrt{V}}$.

b) Théorème

La variance est égale à la différence entre la moyenne des carrés des valeurs et le carré de la moyenne :

$$\boxed{V=\frac{1}{N}\sum_{i=1}^{N} n_i x_i \ - \ \bar{x}^2 \ = \ \overline{x^2}-\bar{x}^2.}$$

c) Démonstration :

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

$$\text{d'où } V=\overline{x^2}-2\bar{x}\frac{1}{N}\sum_{i=1}^{N} x_i+\bar{x}^2 = \overline{x^2}-2\bar{x}^2+\bar{x}^2 = \overline{x^2}-\bar{x}^2.$$

d) Exemples

Calculer la moyenne, l'écart absolu moyen la variance et l'écart type de

a) 7, 8, 8, 11, 13, 15, 15, 19

b)

Valeur	10	13	15	17	22	27	31	32
Effectif	5	8	4	14	11	2	2	4

Remarque :

La plupart des calculatrices nous font ces calculs automatiquement.