Chapitre 3 : Statistique.

I Un peu d'histoire.

La science statistique semble exister dès la naissance des premières structures sociales. D'ailleurs, les premiers textes écrits retrouvés étaient des recensements du bétail, des informations sur son cours et des contrats divers. On a ainsi trace de recensements en Chine au $XX^{ième}$siècle av. J.-C. ou en Égypte au $XVIII^{ième}$ siècle av. J.-C.. Ce système de recueil de données se poursuit jusqu'au $XVII^{ième}$ siècle. En Europe, le rôle de collecteur est souvent tenu par des guildes marchandes, puis par les intendants de l'État.
Mais on attribue à l'histoire des statistiques la date de commencement de 1749, bien que l'interprétation du terme « statistique » a changé au cours du temps. Aux temps plus anciens, cette science ne consistait qu'à la collection d'informations sur les états, d'où l'étymologie du nom, de l'allemand Staatskunde. Plus tard, cette définition est étendue à tout type d'information collectée et, encore plus tard, les sciences statistiques incluent l'analyse et l'interprétation de ces données. En termes modernes, les statistiques incluent les ensembles de données, telles celles de la comptabilité nationale et les registres de températures, ainsi que le travail d'analyse, lequel requiert les méthodes de l'inférence statistique.

(Wikipédia)

II Attendus

- Savoir déterminer "à la main", à partir des données du tableau des effectifs :
  - La médiane.
  - Les quartiles.
  - L'étendue.
  - L'écart inter-quartile.
  - Le diagramme en boîte.
  - La variance.
  - L'écart type.
- Savoir obtenir les éléments précédent à la calculatrice.
- Savoir interpréter les paramètres de position et de dispersion d'une série statistique.

III Un exemple concret :

Exemple 1. On considère une petite évaluation donnée à deux classe (2A et 2B) et noté sur 5 : On obtient les résultats suivants :‘

Résultats dans la classe 2A (série X) :

Indice : i	0	1	2	3	4	5
Notes : y_i	0	1	2	3	4	5
Effectifs : n_i	1	12	2	2	3	10

Résultats dans la classe 2B (Série Y) :

Indice : i	0	1	2	3	4	5
Notes : x_i	0	1	2	3	4	5
Effectifs : n_i	0	2	12	8	4	3

Par exemple l'on peut affirmer qu'en 2A, il y a eu 10 élèves qui ont obtenus une note de 5 sur 5. Dans la série X, on a la valeur de $x_5 = 5$ et l'effectif de cette valeur est $n_i = 10$.

IV Paramètres présents dans la série : quartile et boîte à moustache.

A Paramètres de position.

1 Médiane.

Définition 1

Pour obtenir la médiane d'une série, on range les valeurs de la série dans l'ordre croissant. La médiane est la valeur qui partage la série en deux populations d'effectif égal.

Méthode 1

Pour déterminer la médiane, on peut procéder un "rangeant" dans l'ordre croissant les valeurs de la série. Avec l'exemple ci-dessus, on obtient :

$$\underbrace{0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2}_{14\ notes}\ \underbrace{2\quad 3}_{médiane}\ \underbrace{3\ 4\ 4\ 4\ 5\ 5\ 5\ 5\ 5\ 5\ 5\ 5\ 5\ 5}_{14\ notes}$$

On remarque malheureusement ici qu'il n'y a pas de note "au milieu" :

- Si l'on choisit la $15^{ième}$ qui est 2. On a alors 14 notes en dessous et 15 notes au dessus.
- Si l'on choisit la $16^{ième}$ qui est 2. On a alors 15 notes en dessous et 14 notes au dessus.

On choisit donc de prendre la note moyenne des deux. La médiane est donc $\dfrac{2+3}{2} = 2,5$.

Vidéo 1

Calcul de la médiane.

Exercice 1. Déterminer en reprenant l'exemple précédent la médiane de la série Y. Conclure par une phrase interprétant le résultat.

2 Médiane et effectifs cumulés croissants.

Méthode 2

Si l'on reprend l'exemple précédent (vous compléterez le tableau pour la série Y :

Résultats dans la classe 2A (série X) :

Indice : i	0	1	2	3	4	5
Notes : y_i	0	1	2	3	4	5
Effectifs : n_i	1	12	2	2	3	10
Effectifs cumulés croissants	1	13	15	17	20	30

Résultats dans la classe 2B (Série Y) :

Indice : i	0	1	2	3	4	5
Notes : x_i	0	1	2	3	4	5
Effectifs : n_i	0	2	12	8	4	3
Effectifs cumulés croissants						

Ici ces effectifs cumulés croissant signifient :

- La $1^{ière}$ note est 0.
- De la $2^{ième}$ note jusqu'à la $13^{ième}$, ce sont des 1.
- La $14^{ième}$ et la $15^{ième}$ sont des 2.
- La $16^{ième}$ et la $17^{ième}$ sont des 2.
- De la $18^{ième}$ note jusqu'à la $20^{ième}$, ce sont des 4.
- De la $21^{ième}$ note jusqu'à la $30^{ième}$, ce sont des 5.

Ensuite pour déterminer la médiane :
$1^{ière}$ **étape :** On déterminer l'effectif total :

$$N = \sum x_i = x_0 + x_1 + ... + x_5 = 1 + 12 + ... + 10 = 30$$

$2^{ième}$ **étape :** On détermine le "numéro" de la valeur représentant la médiane.

$$numéro = \frac{N+1}{2} = \frac{30+1}{2} = 15,5^{ième}\ note$$

Attention : N'oubliez surtout pas de faire "$N + 1$"
$3^{ième}$ **étape :** Ici l'on obtient un nombre à virgule. On choisit donc la $15^{ième}$ (c'est-à-dire 2) et la $16^{ième}$ (c'est-à-dire 3) en utilisant le tableau des effectifs croissants et on en prend la moyenne :

$$médiane = \frac{2+3}{2} = 2,5$$

La médiane pour la série X est 2,5. On dira que la note médiane pour la seconde A est 2,5. Son interprétation est : "la moitié des notes de la classe sont en dessous de 2,5 et l'autre moitié des notes sont au dessus de 2,5.

Exercice 2. Déterminer en reprenant l'exemple précédent la médiane de la série Y. Conclure par une phrase interprétant le résultat.

3 Quartile.

Définition 2

- Le premier **Quartile** Q_1 est la plus petite valeur de la série tel qu'au moins 25 % des valeurs soit inférieurs ou égales à Q_1.
- Le troisième **Quartile** Q_3 est la plus petite valeur de la série tel qu'au moins 75 % des valeurs soit inférieurs ou égales à Q_3.

Méthode 3

Résultats dans la classe 2A (série X) :

Indice : i	0	1	2	3	4	5
Notes : y_i	0	1	2	3	4	5
Effectifs : n_i	1	12	2	2	3	10
Effectifs cumulés croissants	1	13	15	17	20	30

Résultats dans la classe 2B (Série Y) :

Indice : i	0	1	2	3	4	5
Notes : x_i	0	1	2	3	4	5
Effectifs : n_i	0	2	12	8	4	3
Effectifs cumulés croissants						

Pour déterminer le 1^{ier} quartile :
$1^{ière}$ **étape :** On détermine le "numéro" de la valeur représentant le premier quartile (on considère que l'on connait l'effectif total $N = 30$) :

$$numéro = \frac{N}{4} = \frac{30}{4} = 7,5^{ième}\ note$$

$2^{ième}$ **étape :** Ici l'on obtient un nombre à virgule. On choisit donc la $8^{ième}$ en utilisant le tableau des effectifs croissants. Le 1^{ier} quartile est donc $Q_1 = 2$.
Remarque Si nous avions arrondi à 7 au lieu de 8, nous aurions 7 valeurs en dessous de du quartile ce qui aurait constitué moins de 25 % des valeurs
On fait de même pour déterminer le $3^{ième}$ quartile.

Vidéo 2

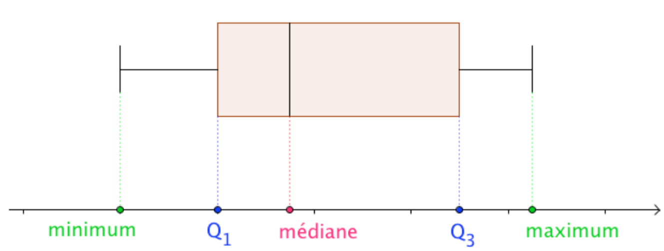
Pour déterminer les quartiles.

Exercice 3. Déterminer en reprenant l'exemple précédent, le troisième quartile de la série X, ainsi que le premier et le troisième quartile de la série Y. Conclure par une phrase interprétant les résultats.

4 Quartile est diagramme en boite.

Définition 3

Digramme en boîte :

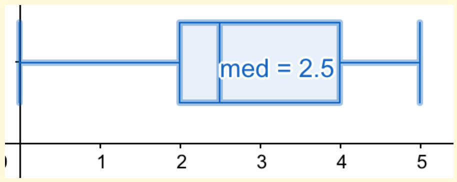


Méthode 4

On considère l'exemple précédent pour la seconde A :
Pour dessiner le diagramme en boîte (ou à moustache), il faut avoir déterminer :

- le minimum : la note 0.
- Le premier quartile : la note $Q_1 = 2$.
- La médiane : $Med = 2,5$.
- Le troisième quartile : la note $Q_3 = 4$.
- Le maximum : la note 5

Ensuite, on le fait apparaitre sous cette forme :



Vidéo 3

Pour dessiner un diagramme en boîte.

Exercice 4. Déterminer en reprenant l'exemple précédent, le diagramme en boîte de la série Y. Comparait les diagrammes en boîte des deux séries.

B Paramètres de dispersions.

1 L'étendue.

Définition 4

L'étendue est différence entre la plus grande valeur est la plus petite valeur.

Exercice 5. En reprenant l'exemple précédent, déterminer les étendues des séries X et Y. Interprétez.

2 L'écart inter-quartiles.

Définition 5

L'écart inter-quartiles est différence entre le premier quartile et le troisième quartile.

Exercice 6. En reprenant l'exemple précédent, déterminer l'écart inter-quartiles des séries X et Y. Interprétez.

Exercices 11 à 19 page 69-70

V Paramètres "calculés" à partir des données de la série.

A Paramètres de positions : La moyenne.

Définition 6

La moyenne est la somme des valeurs divisée par l'effectif total :

$$moyenne = \frac{\sum n_i \times x_i}{N}$$

Méthode 5

Si l'on reprend l'exemple précédent, on obtient pour la seconde A :

$$\bar{X} = \frac{1 \times 0 + 12 \times 1 + ... + 10 \times 5}{30} = 2,8$$

On dira que la moyenne des la seconde A sur cette petite interrogation est de 2,8.

Exercice 7. En reprenant l'exemple précédent, déterminer la moyenne de la série Y. Comparer les résultats des deux secondes et interpréter.

B Paramètres de dispersions : L'écart type.

1 La variance.

L'écart-type exprime la dispersion des valeurs d'une série statistique autour de sa moyenne. Plus il est grand, plus les valeurs sont "éloignées" de la moyenne et donc la "dispersion" est importante.

Définition 7

La variance est la moyenne est écart à la moyenne au carré :

$$V(X) = \frac{\sum n_i \times (x_i - \bar{X})^2}{N}$$

Proposition 1

Pour calculer la variance l'on peut aussi calculer la moyenne des carrés et lui retrancher la moyenne au carré :

$$V(X) = \underbrace{\frac{\sum n_i \times x_i^2}{N}}_{moyenne\ des\ carrés} - \underbrace{\bar{X}^2}_{moyenne\ au\ carré}$$

On utilisera plus volontiers la formule précédente.

Méthode 6

Si l'on reprend l'exemple précédent (vous compléterez le tableau pour la série Y :

Résultats dans la classe 2A (série X) :

Indice : i	0	1	2	3	4	5
Notes : y_i	0	1	2	3	4	5
Effectifs : n_i	1	12	2	2	3	10
Effectifs cumulés croissants	1	13	15	17	20	30

Résultats dans la classe 2B (Série Y) :

Indice : i	0	1	2	3	4	5
Notes : x_i	0	1	2	3	4	5
Effectifs : n_i	0	2	12	8	4	3
Effectifs cumulés croissants						

Pour déterminer la variance de la série X :

$$V(X) = \frac{1 \times 0^2 + 12 \times 1^2 + ... + 10 \times 5^2}{30} - 2,8^2 = 11,2 - 2,8^2 = 3,36$$

2 L'écart type.

Pour que le calcul de la dispersion soit homogène à la série étudier (et non au son carré) l'on prendra la racine de la variance.

Définition 8

L'écart type est la racine carré de la variance :

$$\sigma(X) = \sqrt{V(X)}$$

Méthode 7

Pour déterminer l'écart type, on commencera par déterminer la variance puis l'on calculera sa racine.
Si l'on reprend l'exemple précédent l'on obtient :

$$\sigma(X) = \sqrt{V(X)} = \sqrt{3,36} \simeq 1,83 \text{ à } 10^{-2} \ près.$$

Vidéo 4

Pour déterminer moyenne et écart type d'une série.

Exercice 8. Reprendre l'exemple et déterminer la variance et l'écart type de la série Y. comparer les résultat des deux séries. Interpréter.

Exercices 20 à 26 page 69.

VI A la calculatrice.

Voir la page 205 du livre