Statistiques

Christophe ROSSIGNOL*

Année scolaire 2017/2018

Table des matières

Table des figures

Activité : TP 22 page 280[1] [TransMath]

1 Quelques rappels sur la moyenne

Définition : On considère la série statistique suivante :

valeur du caractère	x_1	x_2	x_3	$\cdots$	x_p
effectif	n_1	n_2	n_3	$\cdots$	n_p

L'effectif total est : $N = n_1 + n_2 + n_3 + \ldots + n_p = \sum_{i=1}^{p} n_i$.

La moyenne de la série est :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{N} = \frac{1}{N}\sum_{i=1}^{p} n_ix_i$$

Remarques : 1. On peut aussi calculer la moyenne à partir des fréquences :

$$\overline{x} = f_1x_1 + f_2x_2 + \ldots + f_px_p$$

2. Pour une série statistique simple (non regroupée suivant les effectifs) x_1, x_2,…,x_n la formule de la moyenne est plus simplement : $\overline{x} = \frac{x_1+x_2+\ldots+x_n}{n}$.
3. Pour une série dont les valeurs sont regroupées en classes, on utilise le centre de chaque classe comme valeur de x_i dans le calcul de la moyenne.
4. La moyenne est très sensible aux valeurs extrêmes.

Exercices : 1, 2 page 266 et 31 page 282[2] ; 18 page 278[3] ; 23 page 281[4] ; 26, 27, 28 page 282[5] [TransMath]

2 Médiane, quartiles, diagramme en boîte

2.1 Médiane

Définition : On considère une série statistique dont les valeurs du caractère étudié ont été rangés dans l'ordre croissant :

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$$

On appelle médiane la valeur centrale de cette série, c'est-à-dire celle qui la sépare en deux parties de même effectif.
On la note : **Me**.

Remarques : 1. Si l'effectif total est impair, la médiane correspond à la valeur centrale.
Si l'effectif total est pair, la médiane correspond à la demi-somme des deux valeurs centrales.

2. Au moins 50 % des valeurs de la série sont inférieures (ou égales) à la médiane et au moins 50 % des valeurs de la série lui sont supérieures (ou égales).
3. La médiane est beaucoup moins sensible aux valeurs extrêmes que la moyenne.

Exemple 1 : On considère la série statistique suivante :

valeur du caractère	50	45	30	60	61
effectif	2	3	2	2	2

1. Utilisation de la calculatrice.
2. Calculs de moyenne.
3. Effet de structure
4. Utilisation d'un tableur.
5. Utiliser la notation $\sum$.

L'effectif total est $N = 11$, il est impair. La médiane est donc la $6^{\text{ème}}$ valeur, classées dans l'ordre croissant.
La médiane est **Me** $= 50$.

Exemple 2 : On considère la série statistique suivante :

valeur du caractère	2	9	6	8	5
effectif	3	2	1	3	3

L'effectif total est 12, il est pair. La médiane est donc située entre la $6^{\text{ème}}$ et la $7^{\text{ème}}$ valeur, classées dans l'ordre croissant.
La médiane est **Me** $= \frac{5+6}{2} = 5,5$.

2.2 Quartiles

Définition : On considère une série statistique dont les valeurs du caractère étudié ont été rangés dans l'ordre croissant :

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$$

1. Le premier quartile est la plus petite valeur Q_1 de la liste telle qu'au moins un quart des valeurs de la liste sont inférieures ou égales à Q_1.
2. Le troisième quartile est la plus petite valeur Q_3 de la liste telle qu'au moins les trois quart des valeurs de la liste sont inférieures ou égales à Q_3.

Exemple : On reprend les données de l'exemple 1 du 2.1.
- L'effectif total est $N = 11$.
- $\frac{N}{4} = \frac{11}{4} = 2,75$. Comme au moins un quart des valeurs doit être inférieure à Q_1, Q_1 est donc la $3^{\text{ème}}$ valeur (classée dans l'ordre croissant...). On a donc $Q_1 = 45$.
- $\frac{3N}{4} = \frac{33}{4} = 8,25$. Comme au moins les trois quart des valeurs doit être inférieure à Q_3, Q_3 est donc la $9^{\text{ème}}$ valeur (classée dans l'ordre croissant...). On a donc $Q_3 = 60$.

Remarques : 1. On a donc partagé la série en quatre parties de même effectif, comme indiqué sur le schéma suivant :

$$a_{min}\underbrace{\cdots\cdots\cdots\cdots}_{25\ \%\text{de l'effectif}}\underbrace{\mathbf{Q_1}\underbrace{\cdots\cdots\cdots\cdots}_{25\ \%\text{de l'effectif}}\mathbf{Me}\underbrace{\cdots\cdots\cdots\cdots}_{25\ \%\text{de l'effectif}}\mathbf{Q_3}}_{50\ \%\text{de l'effectif}}\underbrace{\cdots\cdots\cdots\cdots}_{25\ \%\text{de l'effectif}}a_{max}$$

- 25 % de l'effectif a une valeur du caractère comprise entre a_1 et $\mathbf{Q_1}$;50 % de l'effectif a une valeur du caractère comprise entre $\mathbf{Q_1}$ et $\mathbf{Q_3}$;
- 25 % de l'effectif a une valeur du caractère comprise entre $\mathbf{Q_3}$ et a_n.

2. On peut aussi déterminer la médiane et les quartiles d'une série statistique à partir du diagramme des fréquences cumulées croissantes (f.c.c.) ou décroissantes (f.c.d.).

Exercices : 3, 4 page 266[6] [TransMath]

2.3 Diagramme en boîte

On peut résumer ces caractéristiques par un diagramme en boîte, en faisant apparaître sur un axe gradué les valeurs extrêmes, les quartiles et la médiane (voir figure 1).

Exemple : On reprend l'exemple 1 du 2.1.
Le diagramme en boîte est représenté sur la figure 1.

Remarques : 1. On peut aussi représenter ces diagrammes verticalement.

2. Il est possible d'obtenir des diagrammes en boîte sur les calculatrices graphiques (voir figure 2, capture d'écran d'une T.I. 89 Titanium et page 280 [TransMath]).
3. Représenter sur un même graphique plusieurs diagrammes en boîte peut permettre de comparer plusieurs séries statistiques (voir figure 2).

6. Médiane, quartiles.

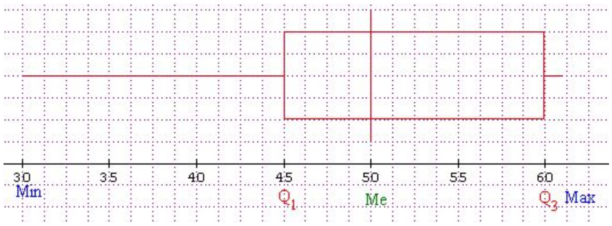

FIGURE 1 – Diagramme en Boîte

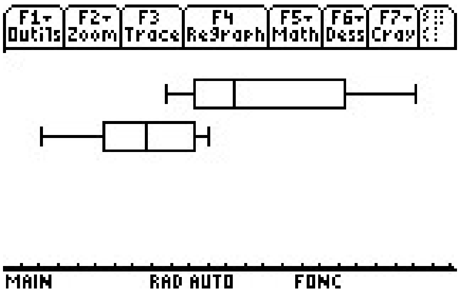

FIGURE 2 – Diagrammes en Boîte sur une TI 89 TITANIUM

Exercices : 1, 2, 3, 4 page 272 [7] ; 6, 7 page 274 et 30 page 282 [8] ; 44, 45 page 286 [9] [TransMath]

3 Paramètres de dispersion

3.1 Étendue et écart interquartile

Définitions :
1. On appelle étendue la différences entres les valeurs extrêmes de la série.
   On a donc $e = x_{max} - x_{min}$.
2. On appelle intervalle interquartile l'intervalle $[\mathbf{Q_1}\ ;\ \mathbf{Q_3}]$.
   On appelle écart interquartile la quantité : $E_i = \mathbf{Q_3} - \mathbf{Q_1}$.

Remarques :
1. L'étendue est très peu utilisée car cet indicateur n'est pas très précis. Il sera amélioré dans la sous-section 3.2.
2. L'intervalle interquartile contient donc les 50 % de l'effectif dont les valeurs sont « les plus proches » de la médiane. L'écart interquartile, qui est une mesure de la longueur de cet intervalle, est donc une mesure de la dispersion des données autour de la médiane :
   — plus il est grand, plus les données sont dispersées autour de la médiane ;
   — plus il est petit, plus les données sont proches de la médiane.

3.2 Variance, écart-type

Activité : Activité 2 page 267 [10] [TransMath]

7. Construire et interpréter un diagramme en boîte.
8. Comparaison de séries.
9. Déciles, diagrammes en boîte à moustaches.
10. Approche d'une mesure de dispersion.

Définition : On considère la série statistique suivante :

valeur du caractère	x_1	x_2	x_3	$\cdots$	x_p
effectif	n_1	n_2	n_3	$\cdots$	n_p

— L'effectif total est : $N = n_1 + n_2 + n_3 + \ldots + n_p = \sum_{i=1}^{p} n_i$.

— La variance V est donné par :

$$V = \frac{n_1 (x_1 - \overline{x})^2 + n_2 (x_2 - \overline{x})^2 + \ldots + n_p (x_p - \overline{x})^2}{N} = \frac{1}{N} \sum_{i=1}^{p} n_i (x_i - \overline{x})^2$$

— L'écart-type σ est : $\sigma = \sqrt{V}$.

L'écart-type est une mesure de dispersion autour de la moyenne.

Remarques :
1. Pour une série statistique simple (non regroupée suivant les effectifs) x_1, x_2,...,x_n la formule de la variance est plus simplement : $V = \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\ldots+(x_n-\overline{x})^2}{n} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2$.
2. Pour une série dont les valeurs sont regroupées en classes, on utilise le centre de chaque classe comme valeur de x_i dans le calcul de la variance.
3. On peut aussi calculer la variance à partir des fréquences :

$$\overline{x} = f_1 (x_1 - \overline{x})^2 + f_2 (x_2 - \overline{x})^2 + \ldots + f_p (x_p - \overline{x})^2 = \sum_{i=1}^{p} f_i (x_i - \overline{x})^2$$

4. La variance mesure la moyenne du carré des écarts à la moyenne de la série.
5. On peut utiliser le mode Statistiques de la calculatrice pour déterminer l'écart-type.

Propriété : (admise)

La variance V peut aussi s'écrire :

$$V = \frac{1}{N}\left(n_1 x_1^2 + n_2 x_2^2 + \ldots + n_p x_p^2\right) - \overline{x}^2 = \frac{1}{N}\sum_{i=1}^{p} n_i x_i^2 - \overline{x}^2$$

Exercices : 9, 10, 11, 12 page 276 et 32 page 282 [11] – 35 page 283 [12] – 36 page 283 et 38, 39 page 284 [13] – 40, 41 page 284 [14] – 46 page 286 et 52 page 288 [15] – 47 page 287 [16][TransMath]

Références

[TransMath] transMATH 1reS, édition 2011 (Nathan)

2, 3, 4, 5

11. Calculs de variance et d'écart-type.
12. Une propriété de la variance.
13. Choisir un résumé adapté.
14. Comparaison de séries.
15. Effet d'un changement linéaire ou affine.
16. Algorithmique.