

Statistiques

Christophe ROSSIGNOL*

Année scolaire 2015/2016

Table des matières

1	Série statistique à une variable	2
1.1	Médiane et écart interquartile	2
1.2	Moyenne et écart-type	3
2	Série statistique à deux variables	4
3	Ajustement affine	5
3.1	Définition, ajustement « au jugé »	5
3.2	Ajustement par la méthode des moindres carrés	7

Table des figures

1	Diagramme en Boîte	3
2	Nuage de points	5
3	Un exemple d'ajustement affine	6
4	Méthode des moindres carrés	7

Liste des tableaux

1	Répartition de notes d'une classe de Terminale STMG	4
2	Nombre d'acheteurs potentiels	6

*Ce cours est placé sous licence Creative Commons BY-SA <http://creativecommons.org/licenses/by-sa/2.0/fr/>

En préliminaire au cours :

Exercices : 1, 2 page 67¹ [Intervalle]

1 Série statistique à une variable

1.1 Médiane et écart interquartile

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

$$x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$$

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	30	45	50	60	61
effectif	2	3	2	2	2

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

La médiane est **Me** = 50.

Exemple 2 :

On considère la série statistique suivante :

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

L'effectif total est 12, il est pair. La médiane est donc située entre la 6^{ème} et la 7^{ème} valeur, classées dans l'ordre croissant.

La médiane est **Me** = $\frac{5+6}{2} = 5,5$.

Définition :

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

$$x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$$

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** .
3. L'**écart interquartile** est la différence **$Q_3 - Q_1$** .

Exemple : On reprend les données de l'exemple 1.

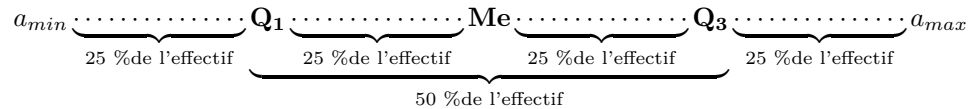
— L'effectif total est $N = 11$.

1. Rappels sur les fréquences relatives.

- $\frac{N}{4} = \frac{11}{4} = 2,75$. Comme au moins un quart des valeurs doit être inférieure à Q_1 , Q_1 est donc la 3^{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^{ème} valeur (classée dans l'ordre croissant...). On a donc $Q_3 = 60$.
- L'écart interquartile est donc $Q_3 - Q_1 = 60 - 45 = 15$.

Remarques :

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



- 25 % de l'effectif a une valeur du caractère comprise entre a_1 et Q_1 ; 50 % de l'effectif a une valeur du caractère comprise entre Q_1 et Q_3 ;
 - 25 % de l'effectif a une valeur du caractère comprise entre Q_3 et a_n .
2. L'écart interquartile est une **mesure de dispersion** autour de la médiane.
 3. On peut utiliser la calculatrice ou le tableur pour déterminer la médiane et les quartiles d'une série statistique à une variable. Voir rabats de couverture du manuel [Intervalle].
 4. 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.

Le diagramme en boîte est représenté sur la figure 1.

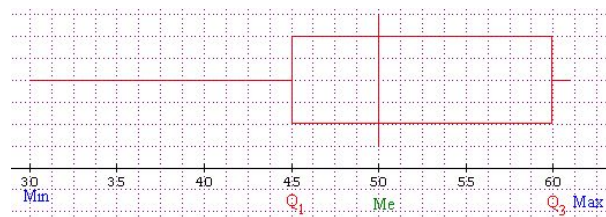


FIGURE 1 – Diagramme en Boîte

Exercices : 10, 11 page 69² [Intervalle]

1.2 Moyenne et écart-type

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 : $N = n_1 + n_2 + n_3 + \dots + n_p$.

La **moyenne** de la série est :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

Exemple 1 :

On reprend l'exemple 1 du 1.1 :

$$\bar{x} = \frac{2 \times 50 + 3 \times 45 + 2 \times 30 + 2 \times 60 + 2 \times 61}{11} \simeq 48,8$$

2. Médiane, quartiles, diagramme en boîtes.

Exemple 2 :

On reprend l'exemple 2 du 1.1 :

$$\bar{x} = \frac{3 \times 2 + 2 \times 9 + 1 \times 7 + 3 \times 8 + 3 \times 6}{12} \simeq 6,1$$

Remarques :

1. Pour une **série statistique simple** (non regroupée suivant les effectifs) x_1, x_2, \dots, x_n la formule de la moyenne est plus simplement : $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$.
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 moyenne.
3. On peut utiliser la calculatrice ou le tableur pour déterminer la moyenne d'une série statistique à une variable. Voir rabats de couverture du manuel [Intervalle].

Définition : L'écart-type d'une série statistique à une variable est un nombre positif, noté σ , qui mesure la dispersion autour de la moyenne.
Ce nombre est obtenu à l'aide de la calculatrice ou d'un tableur. Voir rabats de couverture du manuel [Intervalle].

Exercices : 6, 7, 8 page 68³ [Intervalle]

2 Série statistique à deux variables

Activité : page 60⁴ [Intervalle]

Définition : On appelle **série statistique à deux variables** (ou série statistique doubles) une série statistique où deux caractères sont étudiés **simultanément**.

Remarque : Dans ce chapitre, on n'étudiera que des séries statistiques doubles dont les deux caractères étudiés sont **quantitatifs**.

Si, pour chacun des n individus de la population, on note x_i et y_i les valeurs prises par les deux caractères, on peut alors présenter la série statistique sous la forme d'un tableau :

Caractère x	x_1	x_2	\dots	x_n
Caractère y	y_1	y_2	\dots	y_n

Définition : Dans un repère **orthogonal**, l'ensemble des points M_i de coordonnées $(x_i ; y_i)$ constitue le **nuage de points** associé à la série statistique à deux variables.

Exemple : Le tableau 1 donne répartition des moyennes de 10 élèves en mathématiques et en comptabilité-gestion d'une classe de terminale STMG.

Élèves	Moyenne en Mathématiques (x_i)	Moyenne en Comptabilité (y_i)
Antoine	12	11
Cédric	8	10
Guillaume	11	10
Kevin	9	14
Latifa	15	13
Mohammed	10	12
Pierre	7	8
Sandra	13	11
Stéphanie	10,5	15
Tania	6	9

TABLE 1 – Répartition de notes d'une classe de Terminale STMG

Le nuage de points associé à cette série statistique est représenté sur la figure 2.

3. Moyenne, écart-type.
4. Nuage de points.

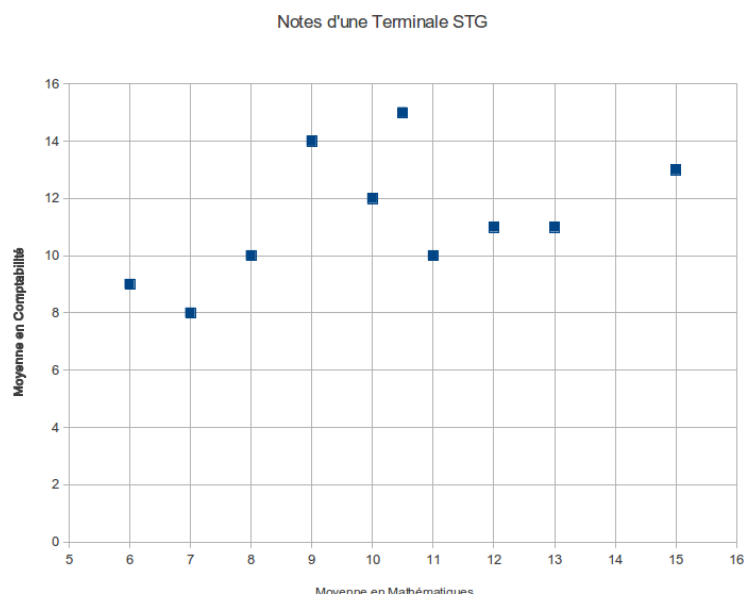


FIGURE 2 – Nuage de points

Remarque : On peut utiliser la calculatrice ou un tableur pour représenter un nuage de points. Voir rabats de couverture du manuel [Intervalle].

Exercices : 13, 14 page 70⁵ – 17 page 70⁶ – 19 page 70⁷ [Intervalle]

3 Ajustement affine d'une série statistique à deux variables

3.1 Définition, ajustement « au jugé »

Activité : page 62⁸ et exercice 20 page 71⁹[Intervalle]

Définition : Effectuer un **ajustement** d'un nuage de points consiste à trouver une fonction dont la courbe représentative « approche » le nuage, c'est-à-dire dont la courbe passe au plus près des points du nuage. Quand le nuage présente une forme « rectiligne », la courbe cherchée est une droite d'équation $y = ax + b$. On parle alors d'**ajustement affine**.

Remarques :

1. Tous les nuages de points ne peuvent pas être approchés par un ajustement affine.
2. Même si le nuage peut être approché par un ajustement affine, il n'y a pas unicité de la droite d'ajustement.

5. Nuage de points.
 6. Utilisation de la calculatrice.
 7. Utilisation du tableur.
 8. Ajustement d'un nuage de points par une droite.
 9. Ajustement affine ou pas ?

Exercice résolu :

Le tableau 2 donne le nombre d'acheteurs potentiels d'un produit donné en fonction de son prix de vente.

1. Représenter le nuage de points correspondant dans un repère orthogonal dont vous aurez choisi judicieusement les unités.
2. On choisit d'ajuster ce nuage de points par la droite d'équation $y = -12,5x + 226,25$.
 - (a) Tracer cette droite sur le graphique précédent.
 - (b) Utiliser cette droite pour déterminer le nombre d'acheteurs potentiels si le prix de vente est fixé à 8 €.

On a représenté le nuage de points correspondant et la droite d'ajustement sur la figure 3.

On peut utiliser cette droite d'ajustement pour déterminer le nombre d'acheteurs potentiels si le prix est fixé à 8 €.

Il sera de : $y = -12,5 \times 8 + 226,25 = 126,25$, soit proche de 126 personnes.

Prix x_i en euros	Nombre y_i d'acheteurs éventuels
9	120
10	100
11	90
12	70
13	60
14	50
15	40
16	30

TABLE 2 – Nombre d'acheteurs potentiels

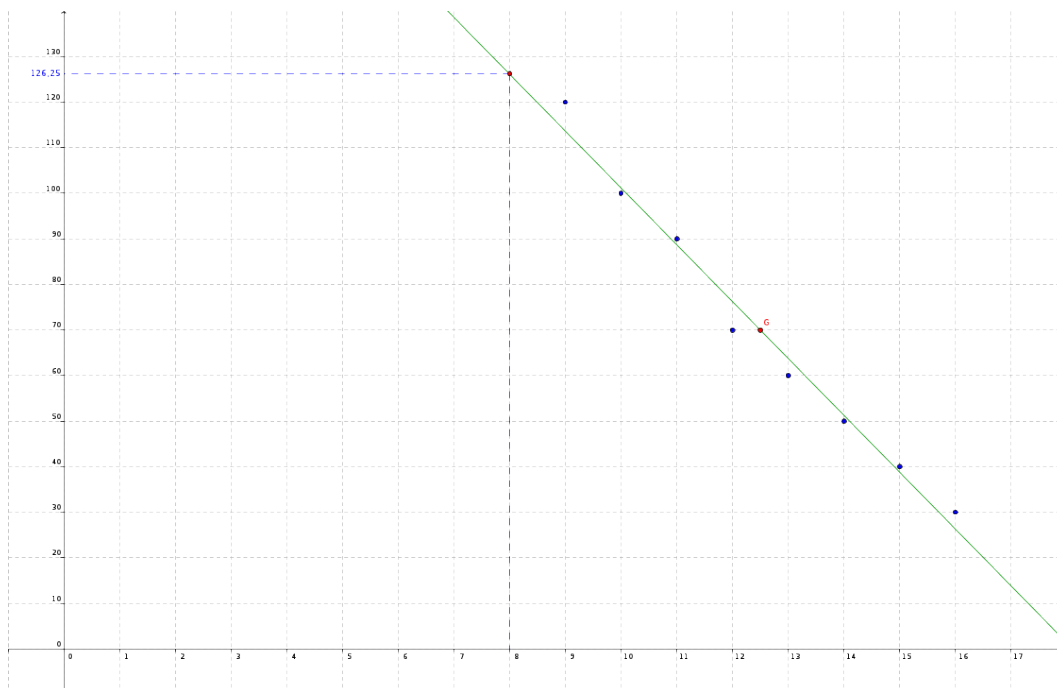


FIGURE 3 – Un exemple d'ajustement affine

Exercices : 21, 23 page 71 et 24 page 72¹⁰ – 35 page 74¹¹ – 43, 44, 45 page 79¹² [Intervalle]

3.2 Ajustement par la méthode des moindres carrés

Effectuer un ajustement de y en x d'un nuage de points par la **méthode des moindres carrés** consiste à trouver la droite d'équation $y = ax + b$ qui minimise la **somme des carrés des écarts** entre les valeurs y_i observées et les valeurs $ax_i + b$ données par la droite.

La fonction f doit donc minimiser l'expression $\sum_{i=1}^n (y_i - (ax_i + b))^2$.

Interprétation graphique : (voir figure 4)

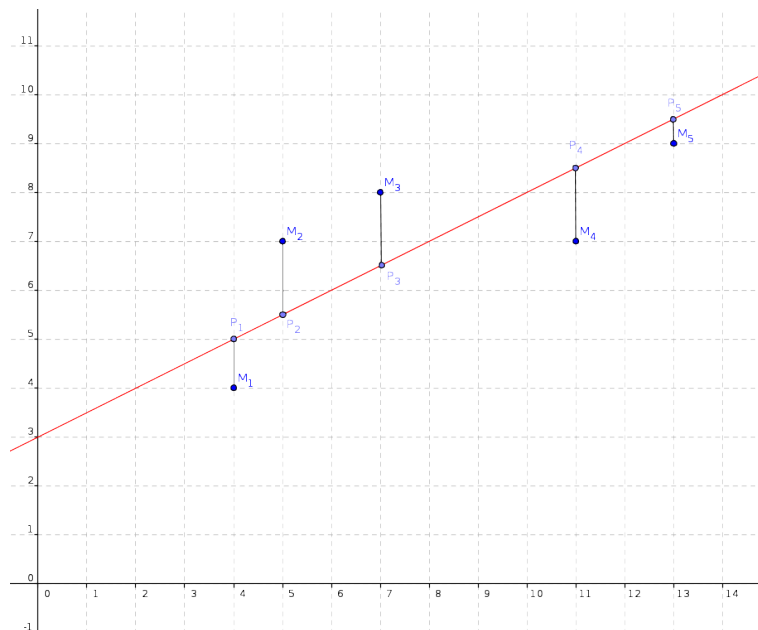


FIGURE 4 – Méthode des moindres carrés

Cela revient à minimiser la somme des carrés des distances « verticales » entre la courbe et les points du nuage :

$$(M_1P_1)^2 + (M_2P_2)^2 + \cdots + (M_nP_n)^2$$

La droite qui minimise cette somme est appelée droite de **régression** de y en x .

Remarque : On utilisera la calculatrice ou un tableur pour déterminer l'équation de la droite de régression. Voir rabats de couverture du manuel [Intervalle].

Exemple : On reprend les données de l'exercice résolu du 3.1.

À l'aide de la calculatrice, on trouve que la droite de régression de y en x admet comme équation $y = -12,6x + 227,7$.

Exercices : 26, 28 page 72 et 33 page 73¹³ – 27 page 72¹⁴ – 34 page 73 et 36 page 74¹⁵ – 42 page 78¹⁶ – 46, 48 page 80 ; 49, 50, 51 page 81 et 53, 55 page 82¹⁷ [Intervalle]

Modules : Exercices 27 page 76¹⁸ ; 38 page 76¹⁹ et 41 page 78²⁰ [Intervalle]

-
- 10. Ajustements affines.
 - 11. Estimations à l'aide d'un ajustement affine.
 - 12. Type BAC.
 - 13. Utilisation de la calculatrice.
 - 14. Utilisation du tableur.
 - 15. Estimations.
 - 16. Q.C.M.
 - 17. Type BAC.
 - 18. Droite d'ajustement par les moindres carrés et point moyen.
 - 19. Comparaison de différents ajustements.
 - 20. Ajustement par une hyperbole.

Références

[Intervalle] Collection Intervalle, Mathématiques, programme 2013, Term STMG, NATHAN, 2013.

2, 3, 4, 5, 7