Statistiques à deux variables
Ajustements affines

Christophe ROSSIGNOL*

Année scolaire 2011/2012

Table des matières

1 Série statistique à deux variables **2**

1.1 Définition – Nuage de points 2

1.2 Point moyen 2

2 Ajustement affine d'une série statistique à deux variables **3**

3 Ajustement par la méthode des moindres carrés **5**

Table des figures

1 Nuage de points 3

2 Un exemple d'ajustement affine 4

3 Méthode des moindres carrés 6

Liste des tableaux

1 Répartition de notes d'une classe de Terminale STG 2

2 Nombre d'acheteurs potentiels 4

Activités : Activité 1 page 218[1] et activité 2 page 219[2] [Roche-Barny]
Exercices : 3, 4, 6 page 244 et 7 page 245[3] [Roche-Barny]

1 Série statistique à deux variables

1.1 Définition – Nuage de points

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 **??** donne répartition des moyennes de 10 élèves en mathématiques et en comptabilité-gestion d'une classe de terminale STG.

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 STG

Le nuage de points associé à cette série statistique est représenté sur la figure 1.

Remarque : On peut utiliser la calculatrice ou un tableur pour représenter un nuage de points. Voir la feuille annexe et le manuel page 238 [Roche-Barny].

1.2 Point moyen

Définition : Le point moyen d'un nuage de points est le point G de coordonnées $(\overline{x}\,;\,\overline{y})$ où :
- $\overline{x}$ représente la moyenne des x_i :

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- $\overline{y}$ représente la moyenne des y_i :

$$\overline{y} = \frac{y_1 + y_2 + \cdots + y_n}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

1. Les statistiques à une variable.
2. Une approche des séries statistiques à deux variables.
3. Révisions sur les statistiques à une variable.

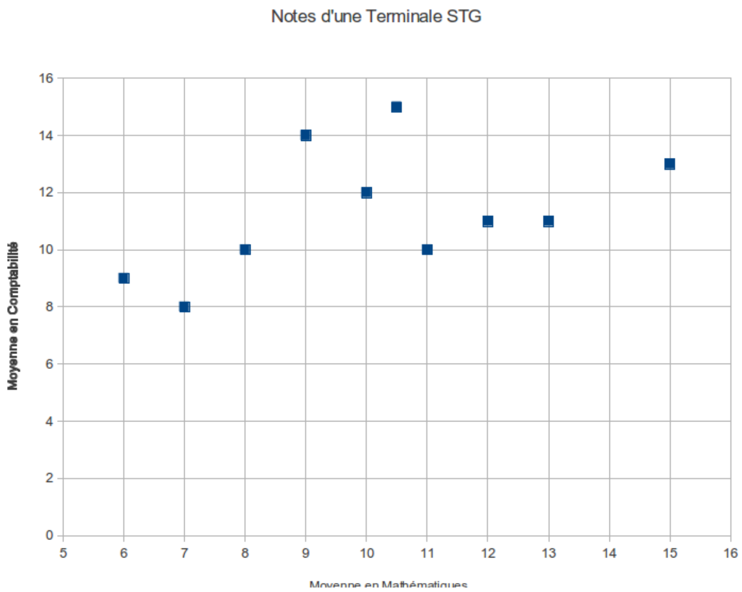

FIGURE 1 – Nuage de points

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

$$\overline{x} = \frac{12+8+11+9+15+10+7+13+10,5+6}{10} = 10,15$$

$$\overline{y} = \frac{11+10+10+14+13+12+8+11+15+9}{6} \simeq 11,3$$

Le point moyen est donc $G\ (10,15\,;\,11,3)$.

Remarques :

1. On fait généralement figurer le point G sur le nuage de points.
2. On peut utiliser la calculatrice ou un tableur pour calculer les coordonnées du point moyen. Voir la feuille annexe et le manuel page 238 [Roche-Barny].

Exercices : 11 page 245 et 13 page 245 [4] [Roche-Barny]

2 Ajustement affine d'une série statistique à deux variables

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 $y = mx + p$. 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.

Propriété : On admettra que, pour que l'ajustement affine soit le meilleur possible, il faut que la droite d'ajustement passe par le point moyen G du nuage de points.

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

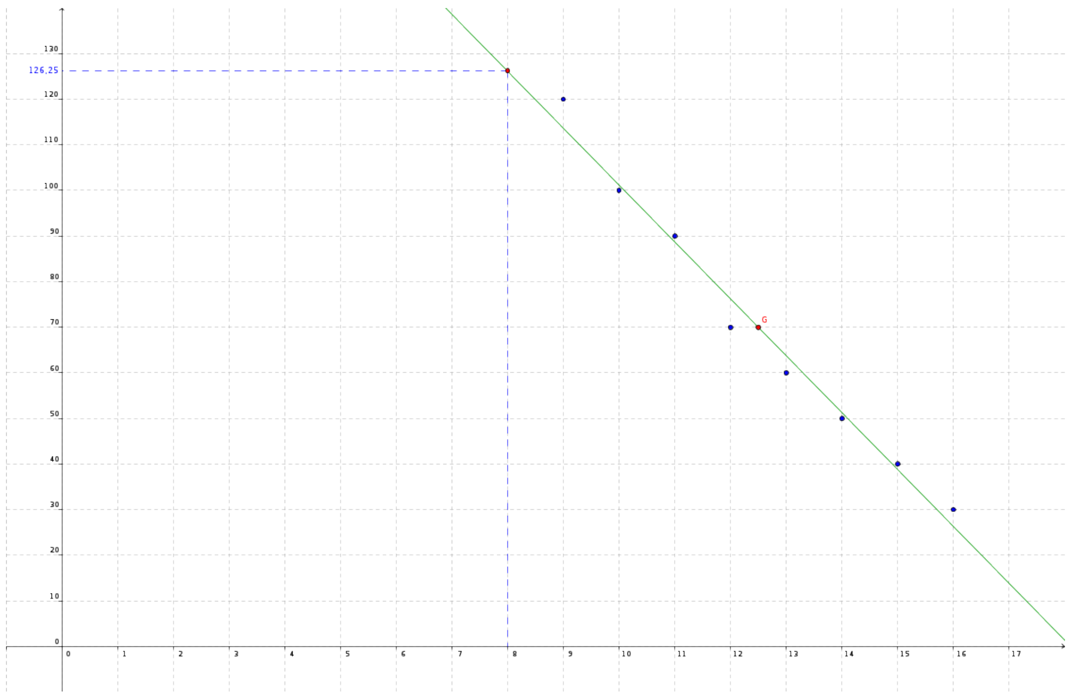

FIGURE 2 – Un exemple d'ajustement affine

Exemple : (tiré de l'exercice 15 page 246 [Roche-Barny])

Le tableau 2 donne le nombre d'acheteurs potentiels d'un produit donné en fonction de son prix de vente. On a représenté le nuage de points correspondant sur la figure 2.
Le point moyen G du nuage a comme coordonnées :
$\overline{x} = \dfrac{9+10+11+12+13+14+15+16}{8} = 12,5$ et $\overline{y} = \dfrac{120+100+90+70+60+50+40+30}{8} = 70$
On choisit d'ajuster ce nuage de point par la droite d'équation $y = -12,5x + 226,25$.
On peut remarquer que cette droite passe par le point moyen G car : $-12,5 \times 12,5 + 226,25 = 70$
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.

Exercices : 10 page 245[5] – 14, 16 page 246 et 17 page 247[6] – 32, 33 page 253[7] [Roche-Barny]
Module : TP3 page 232[8] et TP6 page 235[9] [Roche-Barny]
Exercices : 18 page 247 et 20 page 248[10] [Roche-Barny]

3 Ajustement par la méthode des moindres carrés

Activité : 3 page 221[11] [Roche-Barny]

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 $\sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2$.

Interprétation graphique : (voir figure 3)

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.

Remarques :

1. On peut utiliser la calculatrice ou un tableur pour déterminer l'équation de la droite de régression. Voir feuille annexe et page 238 [Roche-Barny].
2. La droite de régression de y en x. passe par le point moyen G.

Exemple : On reprend l'exemple du 2.

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 : 21, 23 page 249 et 25 page 250[12] – 26 page 250 et 30 page 252[13] [Roche-Barny]
Modules : TP2 page 230[14] – TP5 page 234[15] [Roche-Barny]

Références

[Roche-Barny] Mathématiques Terminale STG, F. Roche et F. Barny, Hachette Éducation, 2006.

2, 3, 5

4. Utilisation de la calculatrice.
5. Choix d'un ajustement affine.
6. Exemples d'ajustements affines.
7. Autres ajustements.
8. Un ajustement affine par la méthode de Mayer.
9. Statistiques et étude de fonction.
10. Méthode de Mayer
11. Des droites d'ajustement.
12. Ajustement par la méthode des moindres carrés.
13. Avec un changement de variable.
14. Utilisation du tableur.
15. Quelques exemples de lissage.

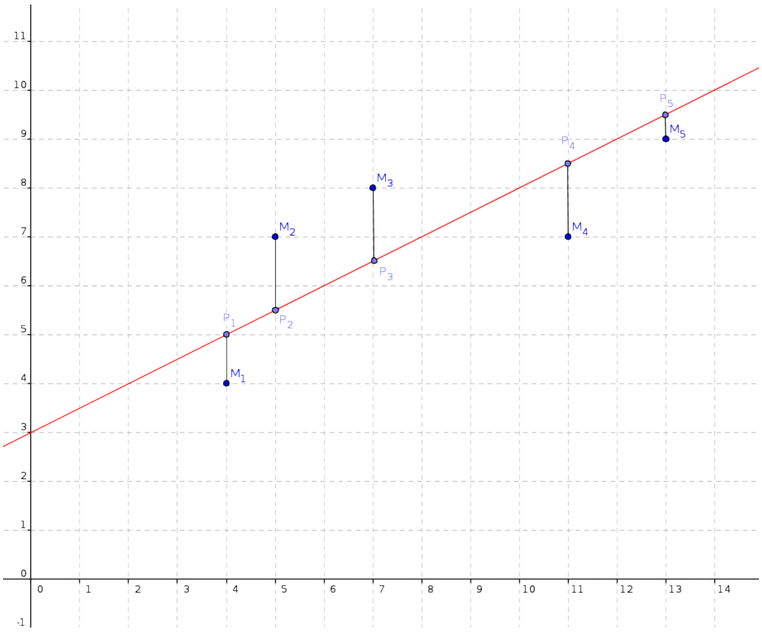

FIGURE 3 – Méthode des moindres carrés