

---

Chapitre 4 : Statistique à deux  
variables

D. Zancanaro      C. Aupérin

2008-2009

---

“J’aimais et j’aime encore les mathématiques pour elles-mêmes  
comme n’admettant pas l’hypocrisie et le vague,  
mes deux bêtes d’aversion”

STENDHAL

Dernière modification : 7 décembre 2008

## Table des matières

<b>1</b>	<b>Séries</b>	<b>1</b>
<b>2</b>	<b>Ajustement affine</b>	<b>2</b>
2.1	Objectif . . . . .	2
2.2	Méthode des moindres carrés . . . . .	3

**COURS : SÉRIE STATISTIQUES À DEUX VARIABLES**

## 1 Séries

*Travail de l'élève* : Hachette Activité 2 p 219 (construire un nuage, son point moyen, observer sa forme)

Les séries statistiques quantitatives à deux variables concernent l'étude simultanée de deux caractères (quantitatifs, ie que les données sont des nombres) d'une population, noté généralement  $(x; y)$ . Ces données sont souvent représentées dans un tableau.

*Exemples* : (poids ; taille), (Nombre d'employés, Chiffre d'affaire), (Année, Chiffre d'affaire) ...

**Définition 1.** L'ensemble des  $N$  couples  $(x_1; y_1), (x_2; y_2), \dots, (x_N; y_N)$  où  $x_1$  et  $y_1, x_2$  et  $y_2, \dots, x_N$  et  $y_N$  sont des valeurs observées de deux caractères  $x$  et  $y$  est appelé **série statistique à deux variables**  $x$  et  $y$  (ou série statistique double).

Le **nuage de points** d'une telle série est l'ensemble des  $N$  points du plan, rapporté à un repère, de coordonnées  $(x_1; y_1), (x_2; y_2), \dots, (x_N; y_N)$ . On note  $\bar{x}$  la moyenne  $N$  des nombres  $x_i$  et  $\bar{y}$  la moyenne des  $N$  nombres  $y_i$ .

Le **point moyen** du nuage est le point  $G$  du plan de coordonnées  $(\bar{x}; \bar{y})$ .

*Exemple* : Deux élèves de BTS ont créé un site internet pendant leur cycle d'études. Ils ont relevé sur le tableau suivant le nombre de visiteurs par mois pendant les huit premiers mois de fonctionnement :

Mois	Rang $x_i$	Nombre $y_i$ de visiteurs
Novembre	1	322
Décembre	2	325
Janvier	3	328
Février	4	327
Mars	5	334
Avril	6	332
Mai	7	335
Juin	8	337

On note par exemple  $(x_5; y_5) = (5; 334)$ .

1. Représenter la série dans un repère où l'on prendra soin de choisir judicieusement le point d'intersection des deux axes, ainsi que les unités sur chacun d'eux.
2. Placer le point moyen de ce nuage de point.

L'étude des séries statistiques à deux variables  $(x; y)$  a pour objectif cette année :

- de tenter d'établir un lien entre les deux variables

- d’effectuer, à l’aide de l’estimation précédente, des prévisions sur les valeurs de la série dans l’avenir, par exemple.

**Exercice 1.1.** Retrouver le tableau à partir du graphique et inversement. Transmath n°23 à 33 p 108.

**Exercice 1.2.** Nuage de points et point moyen. Hachette n°8 + 11 à 13 p 245.

## 2 Ajustement affine

### 2.1 Objectif

**Définition 2.** Lorsque l’on réussit à établir un lien fonctionnel entre les variables  $x$  et  $y$ , on dit que l’on réalise un ajustement.

En fait, effectuer un ajustement de nuage de points consiste à trouver une fonction dont la courbe représentative passe au plus près de chacun des points du nuage.

#### Remarques :

- Cette année, nous nous contenterons d’étudier les ajustements affines, c’est-à-dire que les valeurs  $y_i$  de la série sont estimées à l’aide d’une relation du type  $y_i = ax_i + b$ , avec  $a, b \in \mathbb{R}$ .
- On admettra que, pour que l’ajustement affine soit le meilleur possible, il faut que la droite d’ajustement passe par le point moyen du nuage.

*Exemple :* Faire un graphique avec ajustement affine, parabolique et un pour lequel l’ajustement est impossible.

**Exercice 2.1.** Représenter un nuage de point, tracer une droite d’ajustement donnée ou non, justifier. Transmath n°34 à 41 p 108.

**Exercice 2.2.** Justifier une droite d’ajustement. Hachette n°9 + 10 p 245.

**Définition 3.** Une série statistique à deux variables est appelée série chronologique lorsqu’une des variables est le temps. Elle sera alors choisie comme première variable.

Effectuer un ajustement sur une série chronologique permet alors d’estimer des valeurs de  $y_i$  dans le futur ou le passé.

*Exemple :* Établir une prévision de population pour les années futures, à partir de données statistiques sur les années passées.

Estimer le nombre de malade atteints d’une épidémie, le chiffre d’affaire d’une entreprise ...

**Exercice 2.3.** Prévisions. Transmath n°47 + 48 p 108.

**Exercice 2.4.** Prévisions. Hachette n°14 à 20 p 246.

## 2.2 Méthode des moindres carrés

On suppose que les points d'un nuage sont approximativement alignés.

Pour déterminer une droite d'ajustement de ce nuage, on peut le faire "au jugé", mais cela a l'inconvénient de dépendre totalement de la personne qui réalise le tracé. Il n'y a pas unicité et l'on comprend aisément que cette méthode manque de précision.

*Travail de l'élève* : Hachette Activité 3 p 221 (comparer plusieurs droites qui approchent un nuage)

**Définition 4.** La méthode des moindres carrés est un méthode qui permet de donner une droite d'ajustement, appelée droite de régression (linéaire).

La droite de régression de  $y$  en  $x$  est la droite d'équation  $y = ax + b$  où les paramètres  $a$  et  $b$  ont été calculés de façon à ce que la somme :

$$(y_1 - (ax_1 + b))^2 + (y_2 - (ax_2 + b))^2 + \dots + (y_N - (ax_N + b))^2$$

soit minimale.

Cette droite passe par le point moyen du nuage.

**Remarque** : Les paramètres  $a$  et  $b$  sont données par toute calculatrice graphique ou par tout tableur : cf Hachette p238-239

**Remarque** : La droite de régression linéaire n'est pas la seule droite d'ajustement affine d'un nuage. Par exemple, il y a la méthode de Mayer (cf Transmath exo 40 p 110, Hachette TP3 p 232)

**Exercice 2.5.** Déterminer une équation de la droite de régression linéaire, corrélation, causalité Transmath n°42 à 46 + 49 p 108 + 54 à 57 p117 + type BAC p 119

**Exercice 2.6.** Méthode des moindres carrés, causalité, corrélation, changement de variables ... Hachette n°21 à 25 p 249 + TP 7 p 236 + n°26 à 31 p 250

## Les Annexes

### Casio

- Accéder au menu **Stat, Edit, Entré**
- Rentrer les valeurs de  $x$  dans  $L_1$  et celles de  $y$  dans  $L_2$
- Accéder au menu **Stat, Calc, SET, Lin-Reg(ax+b)**
- Indiquer **List1** et **List2** sur les lignes **Var-X-List** et **Var-Y-List**
- Appuyer sur **Entrée**, choisir **REG, X**.
- Appuyer sur **Entrée** pour voir apparaître les coefficients  $a$  et  $b$ .

### Texas Instrument

- Accéder au menu **Stat, Edit, Entrée**
- Rentrer les valeurs de  $x$  dans  $L_1$  et celles de  $y$  dans  $L_2$
- Accéder au menu **Stat, Calc, Lin-Reg(ax+b)**
- Indiquer  $L_1$  et  $L_2$
- Appuyer deux fois sur **Entrée** pour voir apparaître les coefficients  $a$  et  $b$