

# Statistiques

## 1 Rappels

Soit  $X$  une série statistique composée de  $n$  nombres, numérotés  $x_1, x_2, \dots, x_n$ .

**Définition 1** Moyenne d'une série

La moyenne de la série  $X$  est le nombre :  $\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$

**Remarque 1** Cette définition illustre ce que l'on fait pour calculer la moyenne d'un élève alors que toutes les notes sont au même coefficient.

**Remarque 2** On peut également noter la moyenne à l'aide d'un autre symbole mathématique.

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

Le symbole  $\Sigma$  (sigma dans l'alphabet grec, le S de notre alphabet, S comme somme), signifiant que l'on remplace d'abord  $i$  par 1, on écrit ensuite  $x_1$ , puis on écrit  $+$ . On remplace ensuite  $i$  par 2, on écrit alors  $x_2$  puis  $+$ , etc jusqu'à  $i$  égale à  $n$ .

**Définition 2** Médiane d'une série statistique

Pour trouver la médiane d'une série statistique contenant  $n$  éléments  $x_1, x_2, \dots, x_n$ , on ordonne ces  $n$  valeurs :

- Si  $n$  est impair, alors la médiane est la valeur du nombre  $x_i$  tel que  $i = \frac{n+1}{2}$
- Si  $n$  est pair, alors la médiane est la valeur du centre de l'intervalle  $[x_{\frac{n}{2}}; x_{\frac{n}{2}+1}]$

**Remarque 3** Pourquoi s'intéresse-t-on à la médiane ?

La médiane, comme la moyenne, est un paramètre d'une série. Intuitivement la médiane est le nombre qui permet de couper la population étudiée en deux groupes contenant le même nombre d'éléments.

Prenons un exemple en étudiant les salaires bruts des salariés travaillant en France. La connaissance du salaire moyen brut ne donne aucun renseignement sur la répartition des salaires. Or, il y a de tels écarts entre les salaires qu'il est intéressant de connaître le salaire "médian". En 1998 il était de 150 000 F par an. Ainsi la connaissance de ce salaire fictif permet de dire que 50% des salariés gagnaient moins de 150 000 F brut par an et 50% gagnaient plus. On peut d'ailleurs chercher à mieux étudier la répartition des salaires, en les partageant non pas en deux groupes de même effectifs, mais quatre, ou dix ou plus.

**Définition 3** Étendue d'une série statistique

On appelle étendue d'une série statistique la différence entre sa plus grande valeur et sa plus petite.

**Définition 4** Variance d'une série statistique

Soit  $X = (x_1, x_2, \dots, x_n)$  une série statistique de moyenne  $\bar{x}$ . La variance de cette série notée  $V(x)$  est le nombre :

$$V(x) = \frac{1}{n}((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)$$

**Remarque 4** On a donc, avec la notation en sigma, que :  $V(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

**Définition 5** Écart-type d'une série statistique

On note  $\sigma$  l'écart-type d'une série statistique  $X = (x_1, x_2, \dots, x_n)$ . On a la formule suivante :

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

**Remarque 5** En d'autres termes, l'écart-type est la racine carrée de la variance.

**Remarque 6** La formule de la variance est à rapprocher de la formule permettant de calculer une distance dans un repère orthonormé :  $AB^2 = (x_B - x_A)^2 + (y_B - y_A)^2$ .

La variance mesure donc la distance carrée moyenne (car on divise par  $n$ ) par rapport à la moyenne  $\bar{x}$ .

L'écart-type mesure donc l'écart moyen par rapport à la moyenne. Voyons cela sur exemple.

**Exemple 1**

Considérons deux élèves, l'un appelé  $X$  et l'autre  $Y$ , ayant obtenu les notes suivantes à quatre interrogations de même coefficient :  $X = (12, 8, 11, 9)$  et  $Y = (18, 19, 2, 1)$ .

On remarque tout d'abord qu'ils ont la même moyenne :  $\bar{x} = \frac{1}{4}(12 + 8 + 11 + 9) = 10$  et  $\bar{y} = \frac{1}{4}(18 + 19 + 2 + 1) = 10$ .

Mais sont-ils le même genre d'élève ? La réponse est évidemment non, au vu de leur note, mais l'outil qui va nous permettre de répondre mathématiquement est l'écart-type.

Calculons, la variance puis l'écart-type de chacune de ces séries :

$$V(X) = \frac{(12 - 10)^2 + (8 - 10)^2 + (11 - 10)^2 + (9 - 10)^2}{4} = 2,5 \text{ et alors } \sigma = \sqrt{2,5} \approx 1,58$$

$$V(Y) = \frac{(18 - 10)^2 + (19 - 10)^2 + (2 - 10)^2 + (1 - 10)^2}{4} = 72,5 \text{ et alors } \sigma = \sqrt{72,5} \approx 8,51$$

On voit donc que l'élève  $X$  est en moyenne à 1,58 de sa moyenne, alors que  $Y$  y est en moyenne à 8,51.

## 2 Nuage de points-Covariance

Dans la partie précédente les séries statistiques ne possédaient qu'un caractère. Par exemple la note obtenue à une interrogation. Dans la suite nous allons étudier des séries à deux caractères. Par exemple, la note obtenue à l'interrogation et la température corporelle pendant l'interrogation. Peut-être pouvons-nous découvrir un lien entre la note et la température du corps ???

Le but du programme de terminale est donc de chercher une corrélation (un lien) entre deux caractères d'une série statistique, et pour cela nous devons définir des outils, et faire des dessins !

Considérons une série statistique sur une population (pas forcément humaine ou vivante) comportant  $n$  individus. On note  $X$  le premier caractère étudié sur cette série, et  $Y$  le deuxième. A chaque nombre  $x_i$  de  $X$  correspond un nombre  $y_i$  de  $Y$ . (Par exemple  $X$  peuvent être les notes obtenues par un élève à une interrogation, et  $Y$  la température corporelle de cette élève ;  $x_3$  étant la note obtenue à la 3ème interrogation et  $y_3$  sa température corporelle lors de la 3ème interrogation.)

On peut représenter les données d'une telle série statistique à deux variables dans un tableau :

$X$	$x_1$	$x_2$	$\dots$	$x_n$
$Y$	$y_1$	$y_2$	$\dots$	$y_n$

### Définition 6 Nuage de points

Dans un repère orthogonal, on trace les points  $A_i$  de coordonnées  $(x_i; y_i)$ . L'ensemble de points obtenu est appelé le **nuage de points** associé à cette série statistique à deux variables.

### Définition 7 Point moyen

On note  $\bar{x}$  la moyenne des  $x_i$  et  $\bar{y}$  la moyenne des  $y_i$ .

Le point  $G$  de coordonnées  $(\bar{x}, \bar{y})$  est appelé le **point moyen** du nuage de points associé à cette série statistique à deux variables.

### Définition 8 Covariance

On appelle **covariance** de  $X$  et  $Y$  le nombre noté  $cov(X, Y)$  et défini par :

$$cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

**Remarque 7** Une autre expression de la covariance, beaucoup plus commode pour les calculs est :

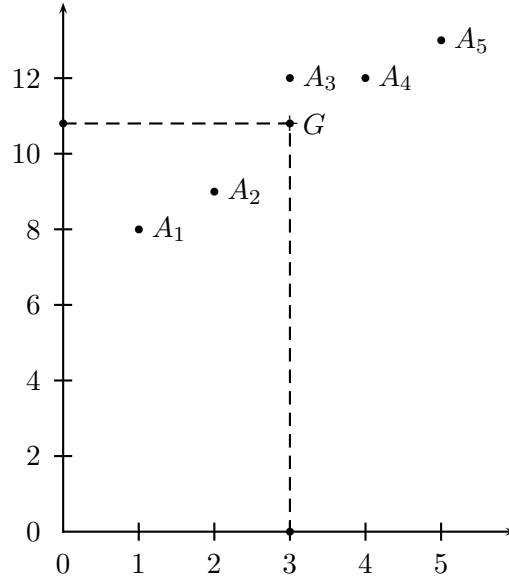
$$cov(X, Y) = \left( \frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x} \bar{y}.$$

**Exemple 2**

La série statistique double suivante indique les notes mensuelles d'un élève au cours des cinq premiers mois de l'année scolaire numérotés de 1 à 5.

Mois $x_i$	1	2	3	4	5
Note $y_i$	8	9	12	12	13

Le point moyen  $G$  du nuage représenté ci-dessous a pour coordonnées  $(3; 10,8)$ .



En effet, on a bien que :  $\bar{x} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$  et  $\bar{y} = \frac{8+9+12+12+13}{5} = \frac{54}{5} = 10,8$ .

De plus :  $cov(X, Y) = \frac{1}{5}((1-3)(8-10,8) + (2-3)(9-10,8) + (3-3)(12-10,8) + (4-3)(12-10,8) + (5-3)(13-10,8))$ .

Donc  $cov(X, Y) = \frac{13}{5} = 2,6$ .

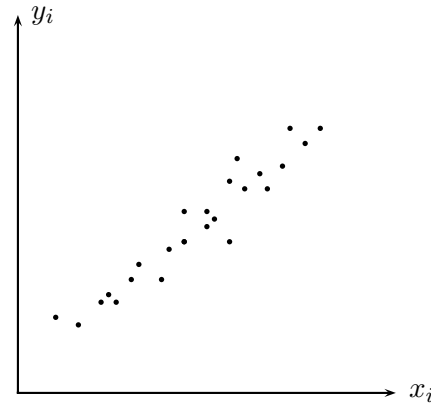
**Remarque 8** Le calcul de la covariance peut-être fait directement par les calculatrices usuelles.

### 3 Ajustement affine par moindres carrés

#### Problématique

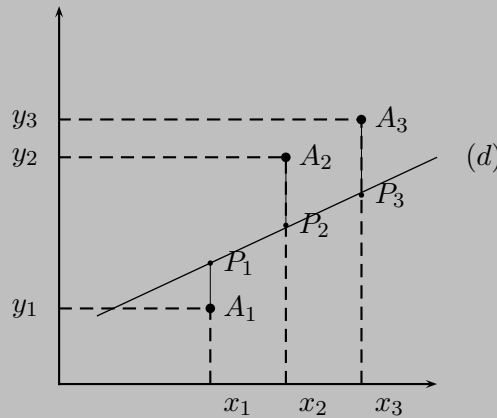
Parfois, il se peut que le nuage associé à une série statistique à deux variables ait une forme "allongée" : il semble que l'on peut tracer une droite (même plusieurs) autour de laquelle sont situés les points du nuage. On dit que chacune de ces droites réalise un **ajustement affine** du nuage.

On peut se demander alors : "Quelle droite tracer?", "y en a-t-il une meilleure que les autres?", "... meilleure, selon quel critère?". La méthode des moindres carrés apporte une certaine réponse à ces questions.



#### **Théorème 1** Méthode des moindres carrés

On considère un nuage de points  $A_i(x_i; y_i)$ , une droite  $(d)$  passant au milieu de ce nuage de point, et  $P_i$ , le projeté de  $A_i$  sur  $(d)$  parallèlement à l'axe des ordonnées.



Il existe une unique droite associée au nuage de points  $A_i(x_i; y_i)$ , telle que la somme  $S$  des  $A_i P_i^2$  soit minimale.

○ Cette droite passe par le point moyen  $G(\bar{x}; \bar{y})$ .

○ Son coefficient directeur  $a$  est égal à  $\frac{cov(X; Y)}{V(X)}$ .

Cette droite s'appelle **droite des moindres carrés**.

#### **Remarque 9**

Pour aller un peu loin, on peut dire que la droite des moindres carrés a pour équation  $y = ax + b$ , avec :

$$a = \frac{\text{cov}(X; Y)}{V(X)}$$

$$b = \bar{y} - \frac{\text{cov}(X; Y)}{V(X)} \bar{x}.$$

Mais ces formules ne doivent pas vous affoler. La calculatrice fait tous les calculs seule si on lui demande gentiment.

## 4 Utilisation de la calculatrice

Nous allons apprendre, dans cette partie, à utiliser les calculatrices TI 83 et CASIO GRAPH 35. Ceux qui possèdent d'autres calculatrices s'inspireront de ce qui est proposé ici, où se référeront à leur manuel d'utilisation.

### 4.1 La situation


Le tableau suivant indique en millions d'euros, le chiffre d'affaires d'une entreprise.

Années $x_i$	1984	1986	1988	1990	1998	2000	2001	2005
Chiffre d'affaires $y_i$	30	33	34,2	38,1	51	53	54,5	60

### 4.2 Affichage du nuage de points

Nous allons afficher sur la calculatrice le nuage de points associé à cette série statistique. On vérifiera que pour celle-ci le graphique nous permet d'envisager un ajustement affine.

Avec la TI 83

Instructions	Commentaires
<div style="border: 1px solid black; padding: 5px; display: inline-block;">Stat</div> <div style="border: 1px solid black; padding: 5px; display: inline-block; background-color: black; color: white;">EDIT</div> <div style="border: 1px solid black; padding: 5px; display: inline-block; margin-left: 10px;">1</div>	Entrer les valeurs de la série $X_i$ dans la colonne $L_1$ , puis les valeurs de la série $Y_i$ dans la colonne $L_2$ .
<div style="border: 1px solid black; padding: 5px; display: inline-block; margin-bottom: 5px;">2<sup>nd</sup></div> <div style="border: 1px solid black; padding: 5px; display: inline-block; margin-bottom: 5px;">STAT PLOT</div> <div style="border: 1px solid black; padding: 5px; display: inline-block; margin-left: 10px;">1</div>	Choisir alors l'option "On", puis le type de graphique  qui permet d'afficher un nuage de points. Pour "Xlist" taper $L_1$ , pour "Ylist" taper $L_2$ .
<div style="border: 1px solid black; padding: 5px; display: inline-block; margin-bottom: 5px;">ZOOM</div> <div style="border: 1px solid black; padding: 5px; display: inline-block; margin-left: 10px;">9</div>	Un repère adapté est choisi par la calculatrice et le nuage de points est affiché.

Avec la CASIO GRAPH 35

Instructions	Commentaires
<p> <span>MENU</span> <span>STAT</span> <span>EXE</span>   <span>F1</span> <span>GRPH</span> <span>F6</span> <span>SET</span>   <span>EXIT</span> <span>F1</span> <span>GRPH</span> </p>	<p>Le menu Statistiques permet d'entrer des séries statistiques, de calculer divers paramètres et d'obtenir des représentations graphiques.</p> <p>Dans la colonne "List1", entrer les valeurs de la série <math>x_i</math>, puis dans la colonne "List2", entrer la série <math>y_i</math>.</p> <p>Cette séquence de touches permet de choisir de représenter un nuage de point.</p> <p>Ceci provoque l'affichage du nuage de points.</p>

### 4.3 Equation et affichage de la droite des moindres carrés

Avec la TI 83

Instructions	Commentaires
<p> <span>Stat</span> <span>CALC</span> <span>4</span>   <span>VARS</span> <span>Y-VARS</span> <span>1</span>  <span>1</span> <span>Enter</span>   <span>GRAPH</span> </p>	<p>Ceci permet de demander l'équation de la droite des moindres carrés.</p> <p>On obtient ainsi les coefficients <math>a</math> et <math>b</math> de la droite des moindres carrés et l'équation de cette droite est stockée dans <math>Y_1</math>.</p> <p>La droite des moindres carrés et le nuage de points s'affichent à l'écran.</p>

Avec la CASIO GRAPH 35

Instructions	Commentaires
<p> <span>F2</span> <span>MED</span>   <span>F6</span> <span>DRAW</span> </p>	<p>A partir de l'écran du nuage de points, on obtient ainsi l'affichage des coefficients <math>a</math> et <math>b</math> de la droite des moindres carrés.</p> <p>La droite des moindres carrés et le nuage de points s'affichent à l'écran.</p>