Chapitre 8 : Statistiques

Axel Carpentier

Terminale technologique :

Tronc commun

Table des matières

1. Statistiques à deux variables

2. Exercice bilan

Statistiques à deux variables

1. Statistiques à deux variables

2. Exercice bilan

Statistiques à deux variables

Pour étudier une population selon deux caractères quantitatifs, on considère une série statistique à deux variables x et y composée de n observations $(x_1; y_1)$, $(x_2; y_2)$, ..., $(x_n; y_n)$, chaque x_i correspondant au caractère x et chaque y_i correspondant au caractère y.

Exemple 1:

Le tableau suivant donne l'évolution du nombre d'adhérents d'un club de rugby de 2001 à 2006.

Année	2001	2002	2003	2004	2005	2006
Rang x_i	1	2	3	4	5	6
Nombre d'adhérents y_i	70	90	115	140	170	220

Le but est d'étudier cette série statistique à deux variables (le rang et le nombre d'adhérents) afin de prévoir l'évolution du nombre d'adhérents pour les années suivantes.

Statistiques à deux variables

Définition 1.

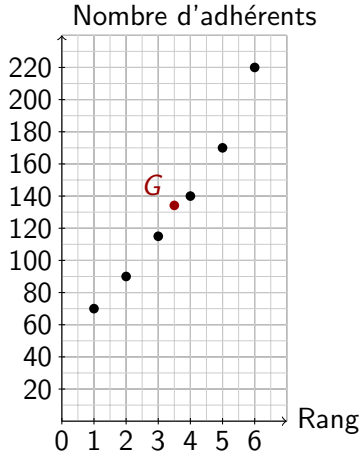
A une série statistique on associe :

- Un **nuage de points**, qui est l'ensemble des n points $M_i(x_i; y_i)$ dans un repère du plan.
- Un **point moyen** dans le même repère : $G(\overline{x}; \overline{y})$ où $\overline{x}$ et $\overline{y}$ sont respectivements les moyennes des séries x et y.

Statistiques à deux variables

Exemple 2:

En reprenant l'exemple précédent on a donc le nuage de point suivant :



En calculant les moyennes des séries statistiques taille (x) et poids (y), on obtient respectivement :

$$\bar{x} = 3,5 \quad \text{et} \quad \bar{y} = 134,2$$

On a donc le point moyen $G(3,5; 134,2)$.

Statistiques à deux variables

Définition 2.

Le principe de **l'ajustement** est de chercher un lien éventuel et simple entre x et y.
Dans le cadre d'un ajustement affine, on cherche à lier x et y par une relation de la forme:

$$y = ax + b$$

On obtient alors une droite d'ajustement censée représenter le nuage de points.

Remarque 1.

Si les points ne sont pas à peu près alignés, un ajustement affine n'a pas d'intérêt.
Cependant, on peut faire un changement de variable (c'est-à-dire transformer les valeurs de x ou y) pour obtenir un nuage de points à peu près alignés.

Statistiques à deux variables

On remarque qu'un ajustement affine ne semble pas très approprié pour ce nuage de points à partir de 2006, on se propose de déterminer un ajustement plus juste.
On effectue un changement de variable en posant $z_i = \ln y_i$. On obtient alors le tableau suivant :

Rang x_i	1	2	3	4	5	6
z_i	$4,248$	$4,500$	$4,745$	$4,942$	$5,136$	$5,394$

On est ainsi ramené à un problème d'ajustement affine, par la méthode des moindres carrés on obtient la droite de régression :

$$z = 0,224x + 4,045$$

Statistiques à deux variables

La variable x_i représente le rang de l'année et y_i le nombre d'adhérents. La variable z_i n'est définie que par $z_i = \ln y_i$ et donc ne représente rien de réel. On cherche donc à se ramener à une relation entre y et x :

$$\begin{cases} z &= 0,224x + 4,045 \\ z &= \ln y \end{cases} \qquad \text{donc :} \quad \ln y = 0,224x + 4,045$$

On a donc :

$$y = e^{0,224x+4,045}$$

Exercice bilan

1. Statistiques à deux variables

2. Exercice bilan

Exercice bilan

On a relevé la population d'une grande métropole sur 50 ans tous les 5 ans. Les résultats sont présentés dans le tableau suivant :

x_i	0	5	10	15	20	25	30
y_i	$19,4$	$19,4$	$27,6$	$40,3$	50	59	69

1. Représenter le nuage de points dans un repère.
2. On effectue le changement de variable $z = \ln y$.
   - 2.1 Réaliser un nouveau tableau présentant les valeurs prises par les variables x et z.
   - 2.2 A l'aide de la calculatrice, déterminer une équation de la droite d'ajustement.
3. - 3.1 En déduire la relation qui lie y et x.
   - 3.2 En admettant que le modèle mathématiques reste valable en dehors du domaine d'étude, extrapoler le nombre d'habitant 5 ans après l'étude.