

Statistiques descriptives à 2 variables

I Nuage de points et point moyen

Vocabulaire

Sur une même population, on considère une **série statistique à 2 variables** X et Y de même dimension n
 X et Y sont les **caractères** étudiés

Les valeurs prises par le caractère X sont notés $(x_1; x_2; \dots; x_n)$

Les valeurs prises par le caractère Y sont notés $(y_1; y_2; \dots; y_n)$

Cette série statistique peut être définir par le tableau ci-dessous :

X	x_1	x_2	...	x_n
Y	y_1	y_2	...	y_n

Définition 1

L'ensemble des points M_i de coordonnées $(x_i; y_i)$ est appelé le **nuage de points** associé à la série statistique à 2 variables

Le point G de coordonnées $(\bar{x}; \bar{y})$ est appelé le **point moyen** du nuage de points, avec :

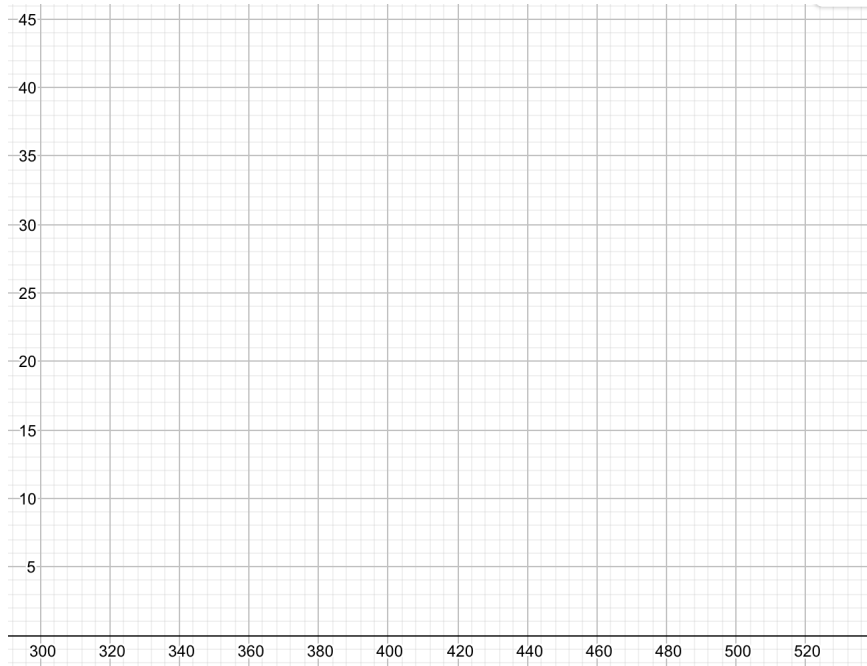
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{et} \quad \bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$$

Exercice 1

Une enquête a été menée auprès de huit représentants d'entreprises différentes de vente de photocopieurs. Le tableau suivant donne, pour chaque représentant, le chiffre d'affaires annuel, en milliers d'euros.

Chiffre d'affaires annuel (en milliers d'euros)	340	390	440	620	470	360	460	520
Revenu annuel (en milliers d'euros)	25	26	33	41	36	28	35	38

- 1) Dans le repère orthogonal suivant, représenter le nuage de points de cette série statistique
- 2) Il y a deux parties dans le salaire d'un représentant : un fixe et une commission proportionnelle au chiffre d'affaires réalisé. Pourquoi le nuage de points montre t'il que la rémunération est approximativement la même dans ls huit entreprises?
- 3) Déterminer les coordonnées du point moyen G du nuage de point puis le placer sur le graphique



II Ajustement affine

Le nuage de points permet parfois de visualiser un lien possible entre les deux variables.

Dans l'exercice précédent, le nuage de points semble correspondre à une droite.

On cherche alors à déterminer cette droite (de la forme $y = ax + b$) qui résume approximativement l'ensemble des points.

On dit qu'on réalise un **ajustement affine de y en x**

1) Droite d'ajustement affine (ou de régression linéaire)

Définition 2

*Lorsque les points d'un nuage de points sont sensiblement alignés, on peut construire une droite, appelée **droite d'ajustement (ou droite de régression)** passant "au plus près" de ces points*

Il existe plusieurs méthodes permettant d'obtenir une telle droite :

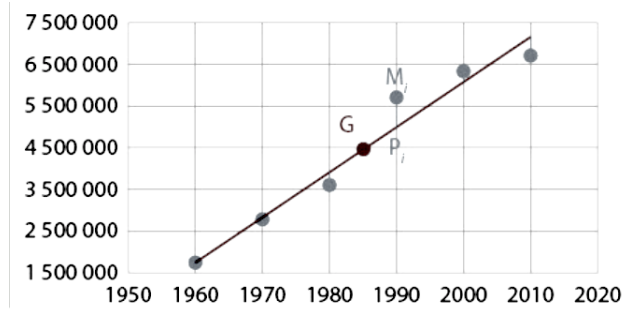
La méthode "au jaugé", qui consiste à tracer une telle droite sans méthode particulière. Cette méthode est rapide mais pas assez précise.

La méthode des points moyens qui consiste à faire passer la droite par 2 points moyens (on divise en deux le nuage de points, on calcule pour chaque groupe le point moyen associé, on place les deux points moyens et on trace la droite ainsi formée)

La méthode des moindres carrés qui est celle que l'on va utiliser le plus souvent car plus précise.

Principe de la méthode des moindres carrés

On souhaite que l'écart entre le point M_i du nuage et le point P_i de la droite ayant la même abscisse soit le plus petit possible. On cherche alors à ce que la somme $S = M_1P_1^2 + M_2P_2^2 + \dots + M_nP_n^2$ soit minimale.



Théorème 1

Cette droite passe par le point moyen G

Calculs de a et b à la calculatrice TI :

Touches : **STATS**, puis **Edite**, saisir les valeurs de x_i dans $L1$ et les valeurs de y_i dans $L2$

Appuyer à nouveau sur **STATS**, puis **CALC** et **RegLin(ax + b)**

Saisir $L1, L2$

Exemple 1

En reprenant l'exemple de l'exercice 1, donner la droite d'ajustement affine (arrondir au dixième) et tracer cette droite sur le repère précédent.

2) Interpolation et extrapolation

Définition 3

A partir de l'ajustement obtenu, on pourra faire des estimations :

On **extrapolera** pour des valeurs à l'extérieur de l'intervalle des valeurs connues

On **interpolera** pour des valeurs à l'intérieur de l'intervalle des valeurs connues

Exemple 2

En reprenant l'exercice 1,

1) Si un neuvième représentant d'entreprise déclare gagner 530000 euros de chiffre d'affaires annuel, estimer son revenu annuel graphiquement, puis retrouver ce résultat par le calcul

2) S'agit t'il d'un extrapolation ou d'une interpolation?