Chapitre 8 : Statistiques à deux variables

On étudie simultanément deux caractères sur chaque individu d'une population.

I Variables qualitatives. Fréquence conditionnelle

Définition
On effectue un tir croisé lorsqu'on trie une population suivant deux caractères. Chaque valeur du caractère partage la population en sous-populations.

Remarque
On représente souvent les résultats dans un tableau de contingence (tableau à double entrée) qui donne les effectifs ou les fréquences.

Définition
La fréquence de A sachant B (ou dit aussi de A parmi B), notée $f_B(A)$ est la proportion des individus appartenant à la fois à A et à B par rapport aux individus de la population B.

$$f_B(A) = \frac{\text{effectif de } A \cap B}{\text{effectif de } B} = \frac{n_{A\cap B}}{n_B}$$

Exercice 1
Sur 5000 clients, 80% ont ont bénéficié des conseils d'un vendeur.
70 % des clients qui ont bénéficié des conseils d'un vendeur ont effectué un achat, alors que 20 % des clients qui n'ont pas bénéficié des conseils d'un vendeur ont effectué un achat.

1. (a) Combien de clients ont bénéficié des conseils d'un vendeur ?
   $5000 \times \frac{80}{100} = 4000$.
   $\boxed{\text{4000 clients ont bénéficié des conseils d'un vendeur.}}$

   (b) Montrer que 2800 clients ont bénéficié des conseil d'un vendeur et effectué un achat.
   $4000 \times \frac{70}{100} = 2800$.
   Parmi les 4000 clients qui ont bénéficié des conseils d'un vendeur, 2800 ont effectué un achat.
   $\boxed{\text{Donc 2800 clients ont bénéficié des conseils d'un vendeur et effectué un achat.}}$

   (c) Compléter le tableau :
   $1000 \times \frac{20}{100} = 200$.
   200 clients n'ont pas bénéficié des conseils d'un vendeur et ont effectué un achat.

	Ont effectué un achat	N'ont pas effectué un achat	Total
Ont bénéficié des conseils d'un vendeur	2800	**1200**	4000
N'ont pas bénéficié des conseils d'un vendeur	**200**	**800**	**1000**
Total	**3000**	**2000**	5000

2. Parmi les clients ayant effectué un achat, quelle est la proportion de clients qui ont bénéficié des conseils d'un vendeur ? Exprimer le résultat en pourcentage arrondi à 1% près.
$p = \dfrac{2800}{3000} = \dfrac{28}{30} = \dfrac{14}{15} \approx 0,93.$

Parmi les clients ayant effectué un achat, environ 93 % ont bénéficié des conseils d'un vendeur.

II Variables quantitatives : nuage de points et exemples d'ajustements

Sur une population de n individus, on cherche à étudier simultanément deux caractères quantitatifs X et Y (« quantitatifs » signifie : qui prennent des valeurs numériques).
Pour chacun des n individus (numérotés de 1 à n) de la population, notons x_i et y_i la valeur prise respectivement par les caractères X et Y.
On présente les données de la série statistique à deux variables obtenue sous forme d'un tableau :

Valeurs prises par X	x_1	x_2	$\dots$	x_n
Valeurs prises par Y	y_1	y_2	$\dots$	y_n

Exemple :

Pays	All	Fr	Esp	R.U.	It
Nombre de naissances x_i	715	760	436	696	545
Population totale y_i	82 539	59 901	42 198	59 516	57 804

Définition (Nuage de points)
Dans un repère orthogonal, l'ensemble des points A_i de coordonnées $(x_i; y_i)$, pour i allant de 1 à n, est appelé nuage de points de la série statistique à deux variables.

Définition (Point moyen du nuage)
Notons $\overline{x}$ la moyenne de la série des x_i, et $\overline{y}$ la moyenne des y_i $(1 \leqslant i \leqslant n)$.
Le point G de coordonnées $(\overline{x};\overline{y})$ est appelé le point moyen du nuage de points de la série statistique.
Rappel :

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

$$\overline{y} = \frac{y_1 + y_2 + y_3 + \cdots + y_n}{n}$$

Les rappels sur l'utilisation de la calculatrice sont aux dernières pages du livre.

Propriété (Utilisation de la calculatrice pour les statistiques)
Pour les calculatrices TI (82-83).

1. Rentrer les données de la série dans la calculatrice.

   $\boxed{\text{STAT}}$ $\boxed{\text{EDIT}}$

   Rentrer les x_i dans L_1, et les y_i dans L_2.
   Pour effacer, utiliser $\boxed{\text{CLEAR}}$.
2. Coordonnées du point moyen $G(\overline{x};\overline{y})$.

   $\boxed{\text{STAT}}$ $\boxed{\text{CALC}}$ $\boxed{\text{2-VAR}}$ $\boxed{L_1}$ $\boxed{,}$ $\boxed{L_2}$

Propriété (Utilisation de la calculatrice pour les statistiques)
Pour les calculatrices CASIO (Graph25-35-65).

1. Rentrer les données de la série dans la calculatrice.

   $\boxed{\text{MENU}}$ $\boxed{\text{STAT}}$

   Rentrer les x_i dans $List1$, et les y_i dans $List2$.
   Pour tout effacer, utiliser $\boxed{\text{DEL-A}}$ (delete all).
   Pour effacer petit à petit, $\boxed{\text{DEL}}$.
2. Coordonnées du point moyen $G(\overline{x};\overline{y})$.
   Il faut d'abord s'assurer d'avoir les bons réglages :

   $\boxed{\text{CALC}}$ $\boxed{\text{SET}}$

   Dans ce menu, on doit choisir :

2 Var X List	:	$List1$
2 Var Y List	:	$List2$
2 Var X Freq	:	1
2 Var Y Freq	:	1

   Ensuite, pour afficher les coordonnées du point moyen $\overline{x}$ et $\overline{y}$:

   $\boxed{\text{CALC}}$ $\boxed{\text{2-VAR}}$

III Ajustement affine

III.1 Ajustement

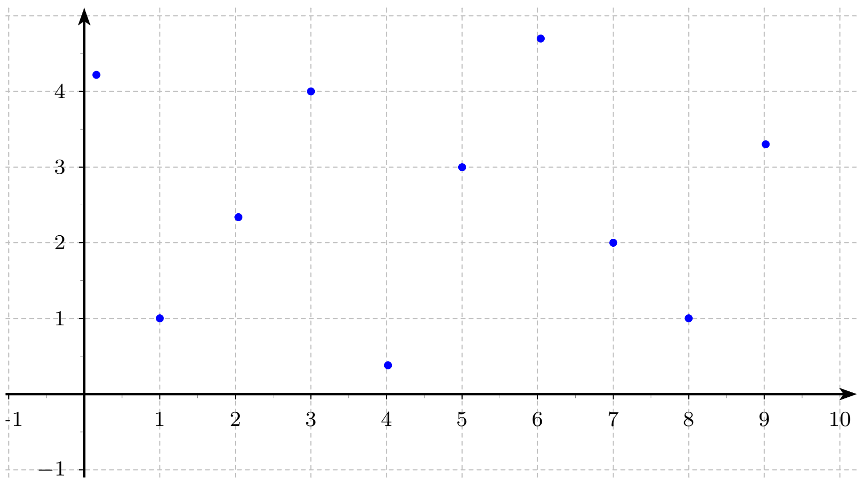
Définition
Effectuer un ajustement de y en x d'un nuage de points, c'est trouver une fonction f dont la courbe représentative passe "au plus près" des points du nuage.
On dit que l'on réalise un ajustement affine lorsque l'ajustement se fait par une fonction affine : $f(x) = ax + b$ et la courbe de f est une droite.

Remarque

— Il est intéressant de faire un ajustement affine lorsque le nuage a une forme plutôt allongée.
— Ces modélisations vont permettre de faire des interpolations et extrapolations (prévisions).
— Il existe d'autres types d'ajustements, c'est-à-dire d'autres fonctions f qui peuvent modéliser un nuage de points.

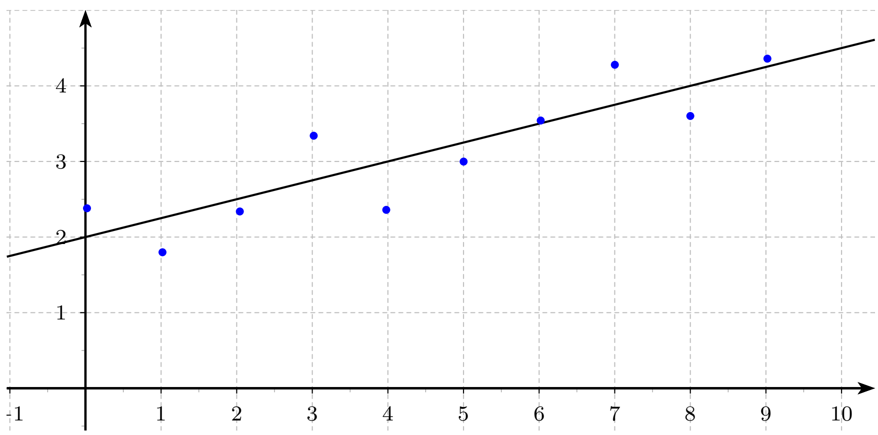
Exemple :

1. Exemple 1



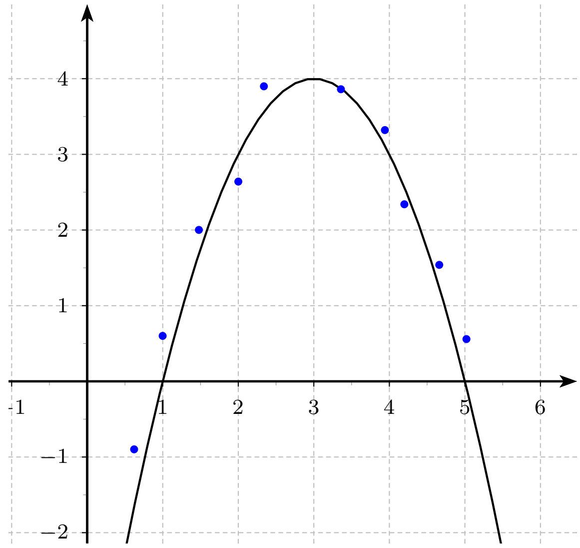
Pas d'ajustement possible avec ce nuage de points.

2. Exemple 2



Le nuage a une forme allongée, on peut chercher un ajustement affine.

3. Exemple 3



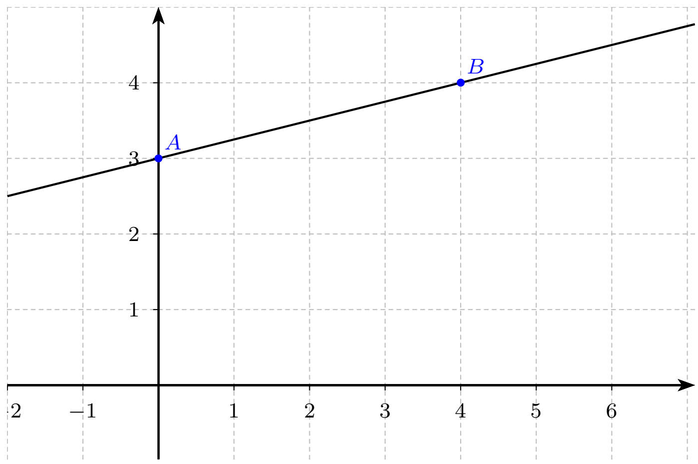
Ajustement du nuage par une fonction trinôme du second degré.

III.2 Rappels sur les équations de droites

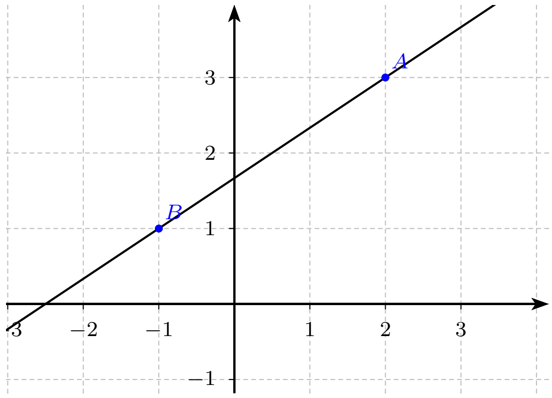
1. Méthode pour tracer une droite connaissant son équation.
   Tracer la droite déquation $y = \dfrac{1}{4}x + 3$. Il suffit de déterminer deux points. On choisit deux valeurs de x.

   On présente les résultats dans un tableau.

x	0	4
y	3	4



2. Tester si un point appartient à une droite dont on connaît une équation.
   Le point $G(2;5)$ appartient-il à la droite (d) déquation $y = 3x + 1$?
   on teste si les coordonnées du point vérifient l'équation en remplaçant x par l'abscisse du point.
   $3x_G + 1 = 3 \times 2 + 1 = 7 \neq y_G$. Donc $G \notin (d)$.
3. Déterminer une équation de droite à partir de deux points.
   Soient $A(2;3)$ et $B(-1;1)$. Déterminer une équation de la droite (AB).



$x_A \neq x_B$, donc (AB) a une équation de la forme $y = mx + p$.

— Calcul du coefficient directeur.
$m = \dfrac{y_B - y_A}{x_B - x_A} = \dfrac{1-3}{-1-2} = \dfrac{2}{3}$.
La droite (AB) a une équation de la forme $y = \dfrac{2}{3}x + p$.

— Calcul de l'ordonnée à l'origine p.
On remplace dans l'équation par les coordonnées d'un des deux points. Avec le point $A(2;3)$, on a

$$\begin{aligned} y &= \frac{2}{3}x + p \\ 3 &= \frac{2}{3} \times 2 + p \\ p &= 3 - \frac{4}{3} \\ p &= \frac{5}{3} \end{aligned}$$

La droite (AB) a pour équation $y = \dfrac{2}{3}x + \dfrac{5}{3}$.