Chapitre 5 : Statistiques à deux variables

I Nuage de points

Définition
On appelle série statistique à deux variables une série statistique dans laquelle deux caractères X et Y sont étudiés simultanément.

Exemples : étude de la consommation d'électricité et de la température extérieure, du prix de vente d'un objet et du nombre d'objets vendus, ...

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 :

Année	2014	2015	2016	2017	2018
Rang de l'année x_i	0	1	2	3	4
Nombre d'habitants y_i (en milliers)	15,2	15,25	15,4	15,5	15,55

Définition (nuage de points)
On considère une série statistique à deux variables définie par l'ensemble des couples $(x_i; y_i)$ où i est un entier variant de 1 à n.
Dans un repère orthogonal, on appelle nuage de points de la série statistique l'ensemble des points A_i de coordonnées $(x_i; y_i)$.
Si la variable X correspond à des dates, la série est dite chronologique.

Définition (Point moyen du nuage)
Notons $\overline{x} = \dfrac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$ la moyenne de la série des x_i,
et $\overline{y} = \dfrac{y_1 + y_2 + y_3 + \cdots + y_n}{n}$ la moyenne des y_i $(1 \leqslant i \leqslant n)$.
Le point moyen du nuage est $G(\overline{x}; \overline{y})$.

Exercice 1
Le tableau suivant présente l'évolution du budget publicitaire et du chiffre d'affaire d'une société au cours des 6 dernières années :

Budget publicitaire x_i (en millliers d'euros)	8	10	12	14	16	18
Chiffre d'affaires y_i (en milliers d'euros)	40	55	55	70	75	95

1. Représenter le nuage de points dans un repère orthogonal.
2. Déterminer les coordonnées du point moyen G, puis placer G.

II Ajustement affine

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 parle d'ajustement affine si l'ajustement se fait par une fonction affine, la courbe de f est une droite.

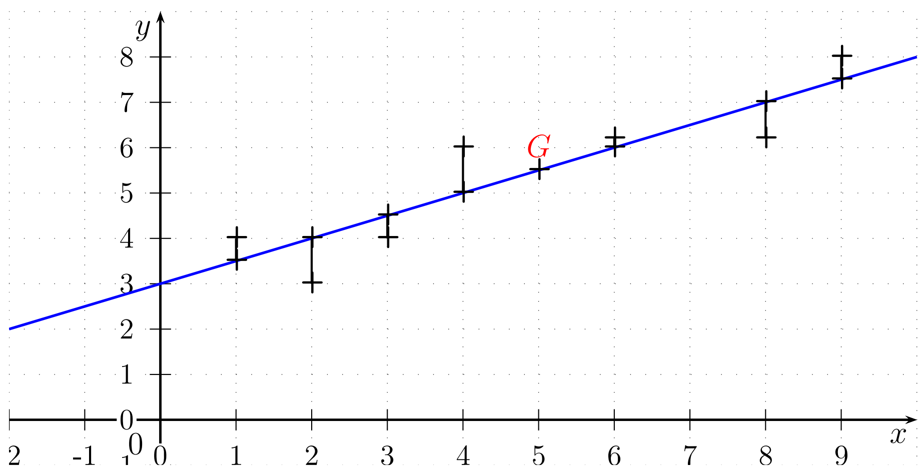
Remarque
- Il est intéressant de faire un ajustement affine lorsque le nuage a une forme plutôt allongée.
- Parfois on peut se contenter de tracer la droite "au jugé" pour qu'elle passe au plus près du nuage.
- Cela va permettre de faire des interpolations et extrapolations.
- Il existe d'autres types d'ajustements, c'est-à-dire d'autres fonctions f qui peuvent modéliser un nuage de points.

Exercice 2 (exemple d'ajutement affine : la méthode de Mayer)
On reprend la série de l'exercice 1.
Soit G_1 le point moyen associé aux trois premiers points du nuage et G_2 le point moyen associé aux trois derniers points du nuage.

1. Calculer les coordonnées de G_1 et G_2.
2. On prend (G_1G_2) comme droite d'ajustement. Tracer cette droite.
3. À l'aide du graphique :
   (a) Estimer le chiffre d'affaire à prévoir pour un budget publicitaire de 22 000 euros.
   (b) Estimer le budget publicitaire qu'il faudrait prévoir pour obtenir un chiffre d'affaire de 100 000 euros.

III Méthode des moindres carrés

Propriété (moindres carrés)
Pour toute série statistique $(x_i; y_i)$ à deux variables, il existe une unique droite d'équation $y = ax + b$ réalisant le minimum de $\sum_{i=1}^{n}(y_i - (ax_i + b))^2$.
Elle est appelée droite de régression de y en x ou droite d'ajustement par la méthode des moindres carrés.
Elle passe toujours par le point moyen G du nuage.



Remarque
On obtient les valeurs de a et b à l'aide de la calculatrice.

Utilisation de la calculatrice

	Casio	Texas	Numworks
Entrer les données	Menu Statistique Rentrer les x_i dans L_1, et les y_i dans L_2	Menu Statistique x_i dans L_1 y_i dans L_2	Menu Regression x_i dans X1 y_i dans Y1
Droite de régression	Calc Set 2VarXlist : list1 et 2VarYList : list2 REG, X, $ax+b$	Stat Calc RegLin($ax+b$) L_1, L_2	Onglet Graphique regression linéaire $ax+b$
Point moyen	Calc 2-Var	Calc 2-Var L_1, L_2	Onglet Graphique

Exercice 3
On considère la série statistique à deux variables donnée dans le tableau suivant :

x_i	5	10	15	20	25	30	35	40
y_i	13	23	34	44	50	65	75	90

1. Dans un repère, représenter le nuage de points $(x_i; y_i)$.
2. À l'aide de la calculatrice, déterminer une équation de la droite d d'ajustement par la méthode des moindres carrés.
3. Représenter la droite d'ajustement de y en x.
4. Donner les coordonnées du point moyen G.
   On peut vérifier que la droite de régression d passe par G.
5. Estimer graphiquement la valeur de x pour $y = 70$.
   Retrouver ce résultat par le calcul.

Exercice 4
Le tableau suivant indique les effectifs de la population en France de 2000 à 2009.

Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Rang x_i	0	1	2	3	4	5	6	7	8	9
Population y_i (millions d'hab.)	58,86	59,27	59,69	60,1	60,51	60,96	61,40	61,80	62,13	62,47

1. Donner les coordonnées du point moyen de la série. Arrondir à 0,01.
2. Donner une équation de la droite d'ajustement par la méthode des moindres carrés. Arrondir les coefficients à 0,01.
3. Utiliser cette droite pour proposer une estimation de la population en 2012.

Exercices sur les statistiques à deux variables

Exercice 1 (comparaison de deux ajustements affines)
Le tableau suivant donne le nombre de clients du téléphone mobile en France atteint à la fin de chaque année.

Année	1998	1999	2000	2001	2002	2003	2004	2005
Rang de l'année x_i	0	1	2	3	4	5	6	7
Nombre de clients en millions y_i	11,5	20,6	29,7	37,0	39,6	41,7	44,5	48,0

Source : ARCEP observatoire des mobiles

1. Déterminer le taux d'évolution du nombre de clients de téléphonie mobile de 1998 à 2005.
   On arrondira le résultat à 0.1 %.
2. Donner les coordonnées du point moyen G du nuage associé à la série statistique. On arrondira les coefficients au dixième. On ne demande pas de justifier.
3. Représenter le nuage de points associée à la série dans un repère orthogonal. Placer le point G.
4. Déterminer une équation de la droite d'ajustement Δ obtenue par la méthode des moindres carrés. (Arrondir les coefficients au centième).
5. Tracer Δ (justifier le tracé).
6. En supposant que ce modèle reste valable pour 2006 et 2007, prévoir le nombre de clients pour la fin de l'année 2007. Indiquer la méthode utilisée.
7. On note $A(0; 11.5)$ et $B(7; 48)$ les points extrêmes du nuage.
   (a) Déterminer une équation de la droite (AB) sous la forme $y = ax + b$.
   Arrondir a et b au centième.
   (b) En déduire une nouvelle estimation du nombre de clients du téléphone mobile pour la fin 2007.
8. À la fin de l'année 2007, on a dénombré 58.7 millions de clients du téléphone mobile. Quelle était la meilleure estimation ?

Exercice 2
En un lieu on relève la pression atmosphérique en fonction de l'altitude.

Altitude x_i (km)	0	0,5	1	1,5	2	2,5	3
Pression y_i (hPa)	1013	955	900	847	797	750	705

1. Construire le nuage de points.
2. Calculer les coordonnées du point moyen G, arrondir l'ordonnée à l'unité.
3. Soit $A(0; 1013)$. On prend la droite (GA) pour droite d'ajustement.
   Déterminer une équation de (GA), et tracer (GA).
4. En utilisant ce modèle, donner une estimation de l'altitude correspondant à une pression de 650 hPa.
5. On a observé une pression de 170 hPa à une altitude de 13 km. L'ajustement affine convient-il à cette altitude ?

Exercice 3

Dans le cadre de cet exercice, on s'intéresse à la consommation d'électricité en France (exprimée en TWh, c'est-à-dire en milliards de kWh) dans le secteur des transports urbains et ferroviaires pour les années $1994 + x_i$ où x_i est un nombre entier naturel.

Année : $1994+x_i$	1995	2000	2004	2005	2006	2007
Rang de l'année : x_i	1	6	10	11	12	13
Consommation : y_i	8,6	10,4	12,2	11,9	12,1	12,2

Source : http ://www.developpement-durable.gouv.fr

On décide d'effectuer un ajustement affine.

1. (a) Représenter le nuage de points associé à la série dans un repère orthogonal. Unités : 0.5 cm en abscisses et 0.5 cm en ordonnées.

   (b) Donner les coordonnées $\overline{x}$ et $\overline{y}$ du point moyen G du nuage. Placer G sur le graphique.

2. Au moyen de la calculatrice, donner une équation de la droite de régression de y en x par la méthode des moindres carrés (arrondir les coefficients à 10^{-3} près).

3. Pour toute la suite de l'exercice, on utilisera la droite d'équation $y = 0,31x + 8,46$ comme droite d'ajustement.
   Tracer cette droite sur le graphique. Justifier par un tableau de valeurs.

4. On considère que cette droite fournit un bon ajustement jusqu'en 2015. Estimer la consommation d'électricité en France pour l'année 2010. Justifier.

5. Estimer à partir de quelle année la consommation d'électricité en France dans le secteur des transports urbains et ferroviaires dépassera $14,5$ TWh.