

CHAPITRE 1 STATISTIQUES A DEUX VARIABLES

I. Activité d'introduction :

Une personne court sur un tapis roulant dont la vitesse peut être modifiée.

On a relevé sa fréquence cardiaque en battements par minute selon l'intensité du travail fourni, exprimée en kilojoules. Voici les résultats obtenus :

Intensité du travail fourni en kilojoules : x_i	10	12	15	22	34	45	53	60
Fréquence cardiaque en battements.min ⁻¹ : y_i	69	80	88	97	114	126	145	158

- 1) Construire le nuage de points représentant cette série statistique. On prendra comme unités graphiques en abscisses 1 unité pour 10kJ et en ordonnées 1 unité pour 20 battements par minute.



- 2) Déterminer la valeur de l'intensité moyenne et la valeur de la fréquence cardiaque moyenne.

Ces deux valeurs sont les coordonnées du point moyen G du nuage. Placer ce point.

- 3) Quelle est la forme du nuage ? Alongé ? parabolique ?

- 4) On se propose de trouver une droite qui passe au plus près de tous les points du nuage.

Ce type de droite réalise un ajustement affine du nuage.

Il existe plusieurs type d'ajustements affine :

une droite tracée au jugé, la droite de Meyer et la droite des moindres carrés.

Première technique : On trace une droite au jugé, du mieux que l'on peut.
Elle doit passer par le point moyen.
On détermine ensuite son équation.

Deuxième technique : La droite de Meyer.
On divise le nuage de points en deux et on calcule les coordonnées des deux points moyens G_1 et G_2 . La droite de Meyer est la droite (G_1G_2) .
On détermine ensuite son équation.

Troisième technique : La droite des moindres carrés.
Cette droite minimise la somme des carrés des distances entre chaque point du nuage et le point de la droite ayant la même abscisse.
Son équation est donnée par la calculatrice. Tracer ensuite cette droite.
On constate qu'elle passe par le point moyen G .

II. STATISTIQUES A DEUX VARIABLES (vocabulaire et méthodes) :

On appelle **série double** un ensemble de couples $(x_i ; y_i)$ avec x_i et y_i des valeurs de deux caractères quantitatifs discrets concernant une même population.

1) Représentation graphique:

Une série double sera représentée dans un repère orthogonal par un **nuage de points** de coordonnées $(x_i ; y_i)$.

2) Ajustement affine :

a) Recherche d'une courbe de régression :

On recherche une courbe connue (droite, parabole ...) qui approche , d'aussi près que possible, le nuage de points.

b) Principe de la méthode :

Lorsque les points du nuage paraissent presque alignés, on va chercher **une droite qui passe au plus près de tous les points du nuage. On effectue alors un ajustement linéaire.**

Une telle droite permettrait notamment de faire des **prévisions**.

Pour mesurer la qualité d'une telle droite, on considère, pour chaque valeur x_i , la différence entre la valeur observée, c'est à dire y_i , et la valeur calculée par l'équation, c'est à dire $m x_i + p$.

On souhaite que toutes les différences : $y_i - m x_i - p$ appelées **erreurs**, ou **résidus**, soient les plus petites possible.

Il existe trois méthodes largement utilisées : la droite au jugé, la droite de Meyer et la droite des moindres carrés.

3) La droite au jugé :

Cette méthode consiste à tracer du mieux que l'on peut une droite qui approche au plus près les points du nuage.

Cette droite passe par le point moyen du nuage, G dont l'abscisse est la moyenne des abscisses des points du nuage et l'ordonnée est la moyenne des ordonnées des points du nuage.

On notera $G(\bar{x} ; \bar{y})$.

Une fois cette droite tracée, il suffit de lire son équation.

4) La droite de Meyer :

Cette méthode consiste à diviser le nuage en deux nuages de même taille.

On calcule les coordonnées des points G_1 et G_2 , points moyens du premier et du deuxième nuage.

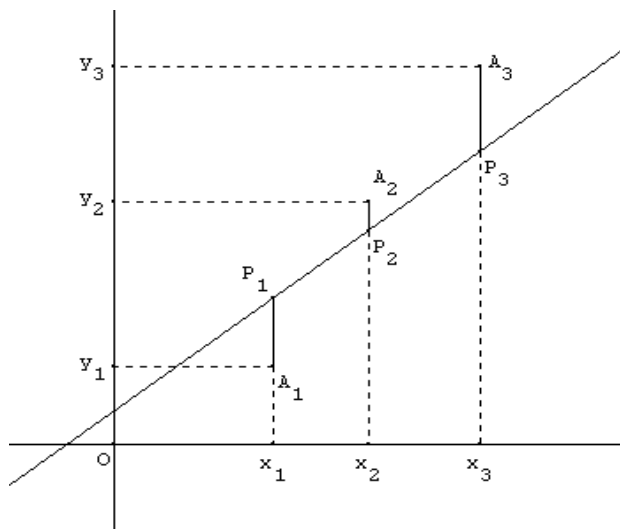
La droite de Meyer est la droite $(G_1 G_2)$. Elle passe par G.

A l'aide des coordonnées des points G_1 et G_2 on peut déterminer son équation.

5) La droite des moindres carrés :

Cette méthode consiste à choisir m et p de façon que la **somme des carrés des résidus soit la plus**

petite possible. On cherche à minimiser $\sum_{i=1}^{i=n} (y_i - m x_i - p)^2$.



On utilisera la calculatrice pour déterminer l'équation de cette droite.

Texas Instrument (TI – 83)

[STAT] 1 : Edit... permet d'entrer les valeurs de x dans L1, puis celles de y dans L2

[STAT] CALC 3 : LINREG(aX+b) puis

(2nd **[L1]** , 2nd **[L2]**) **[ENTER]**

Ceci nous donne les nombres a et b qui représentent m et p .

Casio Graph 35

Dans le menu STAT, entrer les valeurs de x dans List 1, puis celles de y dans List 2.

[CALC]

[SET], entrer dans 2VarXList : List 1
2VarYList : List 2

[EXE] puis Calc Reg 1-Linear

Ceci nous donne les nombres a et b qui représentent m et p .

Exemple :

Considérons la série statistique à deux variables

x_i	10 500	10 590	10 750	10 845	10 963	11 020
y_i	880	822	783	697	632	640

- Placer, dans un repère orthogonal, le nuage de points $A_i(x_i; y_i)$.
On choisira 1 unité pour 100 sur l'axe des abscisses et 1 unité pour 100 sur l'axe des ordonnées.
On choisira comme origine le point O (10 000 ; 600).
- Constater que la forme « allongée » du nuage justifie un ajustement affine.
- Déterminer l'équation la droite d des moindres carrés à l'aide de la calculatrice puis tracer d .