Chapitre 3

Statistiques à deux variables

Sommaire

3.1 Activités

Activité 3.1 (Ajustement affine par une méthode graphique).
A la suite d'une visite médicale dans dix entreprises de services informatiques, on a constaté qu'une certaine proportion du personnel travaillant devant un ordinateur souffrait régulièrement de maux de tête ou de troubles de la vision. Ces résultats figurent, par entreprise, dans le tableau ci-dessous dans lequel l'horaire est donné en heures et centièmes d'heures.

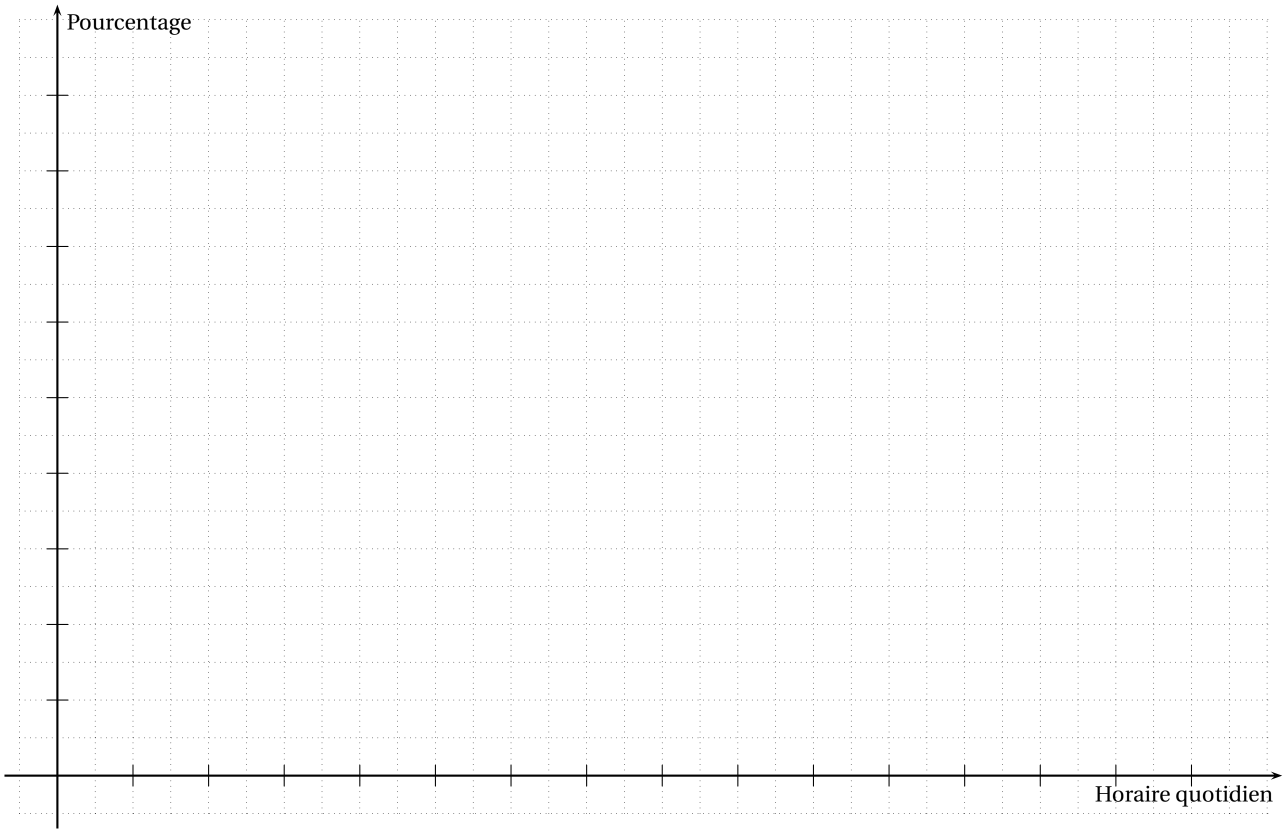
Entreprises	n°1	n°2	n°3	n°4	n°5	n°6	n°7	n°8	n°9	n°10
Horaire quotidien devant un ordinateur x_i	5,5	5,5	6	6,5	6,5	6,5	6,75	7,25	7,25	7,25
Pourcentage du personnel atteint y_i	30	40	45	40	45	50	55	55	60	55

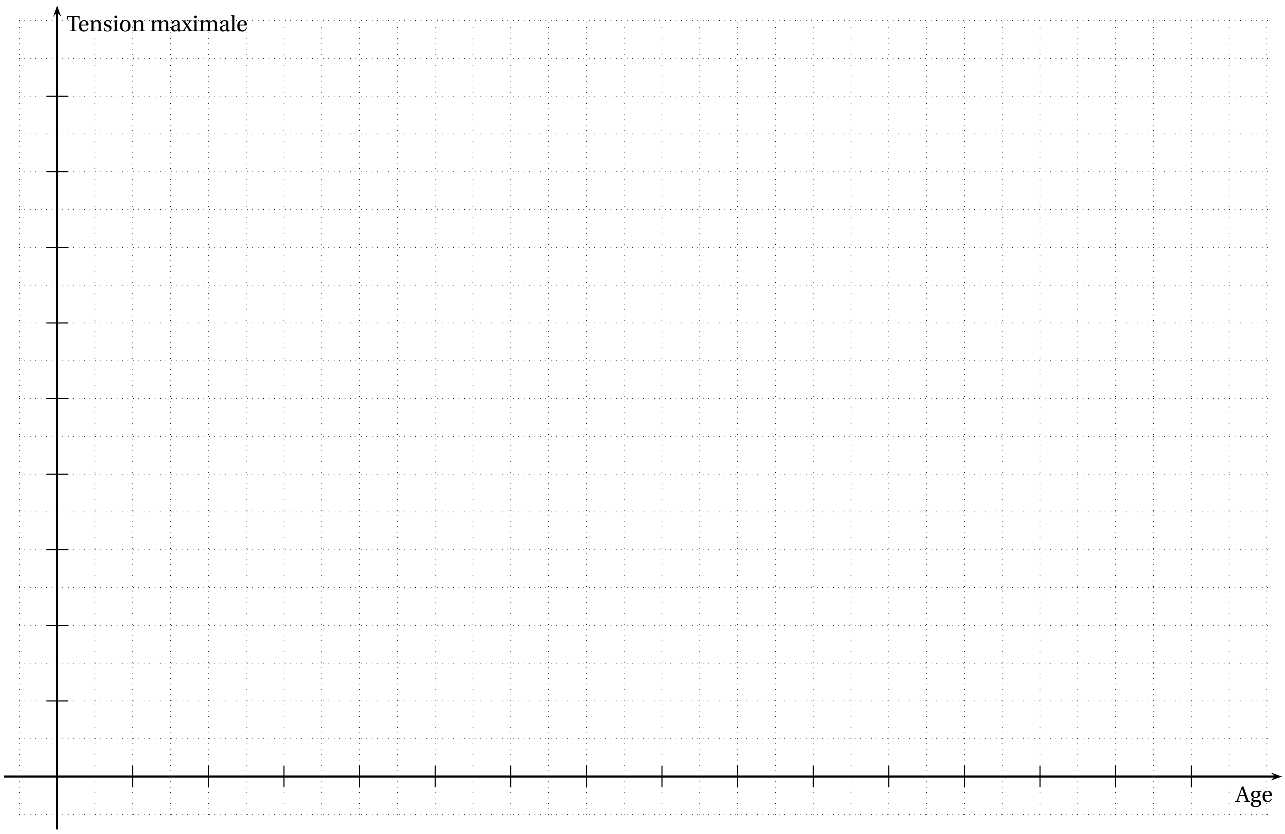
1. Construire dans le repère orthogonal de la figure page 25 le nuage de points associé à ce tableau statistique. On prendra les unités suivantes : en abscisse 1 cm pour 0,25 heure et en ordonnées 1 cm pour 10 %. On graduera l'axe des abscisses à partir de la valeur 5.
2. (a) On note respectivement $\overline{x}$ et $\overline{y}$ les moyennes des séries (x_i) et (y_i). Calculer $\overline{x}$ et $\overline{y}$.
   (b) On note G le point de coordonnées $(\overline{x}\,;\overline{y})$. G est le point moyen du nuage. Placer le point G sur le graphique.
3. On considère les points $A\,(8\,;70)$ et $B\,(8\,;55)$. Construire les droites (GA) et (GB) sur le graphique précédent.
   (a) On se propose de faire un ajustement du nuage par l'une de ces droites. Quelle droite vous semble la plus appropriée ? Expliquer votre choix.
   (b) Déterminer une équation de la droite choisie.
4. En utilisant l'ajustement que vous avez choisi, estimer le pourcentage de personnes atteintes de maux de têtes pour une utilisation moyenne de 7 heures.

Activité 3.2 (Ajustement par la méthode de MAYER et la méthode des moindres carrés).
Le tableau suivant donne dans une population féminine, la moyenne de la tension artérielle maximale en fonction de l'âge.

Rang de l'âge	1	2	3	4	5	6
Age en années : x_i	36	42	48	54	60	66
Tension maximale : y_i	11,8	13,2	14	14,4	15,5	15,1

1. Droite de MAYER
   (a) Représenter graphiquement le nuage de points de coordonnées $(x_i\,;\,y_i)$ de cette série statistique dans le repère orthogonal de la figure page ci-contre. On graduera l'axe des abscisses à partir de 36 et l'axe des ordonnées à partir de 11. De plus, on prendra comme unités 0,5 cm pour une année en abscisse et 2 cm pour une unité de tension en ordonnée.
   (b) Calculer les coordonnées du point moyen G du nuage et placer G sur le graphique.
   (c) On fractionne le nuage précédent en deux parties constituées respectivement par les points numérotés de 1 à 3 et ceux numérotés de 4 à 6. On note G_1 et G_2 les points moyens respectifs de ces deux parties du nuage de points.
      i. Calculer les coordonnées de G_1 et G_2.
      ii. Tracer la droite (G_1G_2).
      La droite (G_1G_2) s'appelle droite de MAYER. On admet qu'elle constitue une bonne droite d'ajustement pour un nuage de points « étiré ».
      iii. Vérifier que la droite (G_1G_2) a pour équation : $y = 19x + 253$.
      iv. Vérifier que le point G appartient à la droite (G_1G_2).
      C'est un résultat général : le point moyen G appartient dans tous les cas à la droite de MAYER (G_1G_2).
   (d) On admet que la droite de MAYER constitue un ajustement convenable du nuage de points précédent.
      i. Déterminer graphiquement, en faisant apparaître les traits de construction utiles, la tension artérielle maximale prévisible pour une personne de 70 ans.
      ii. Vérifier le résultat précédent par le calcul en utilisant l'équation de la droite (G_1G_2).
2. Méthode des moindres carrés
   (a) i. On décide de faire un ajustement affine de cette série à l'aide de la calculatrice. Pour trouver la droite qui passe « au plus près » de tous les points, les calculatrices utilisent une méthode de calcul appelée méthode des moindres carrés. Avec le menu *stat* de la calculatrice (voir le paragraphe 3.2.5 page 30 pour des informations sur les calculatrices), écrire les deux séries et donner l'équation de la droite $\mathscr{D}$.
      ii. Tracer la droite $\mathscr{D}$ dans le repère.
   (b) A l'aide de l'équation de $\mathscr{D}$, déterminer par le calcul une estimation de l'âge d'une personne dont la tension est 16.

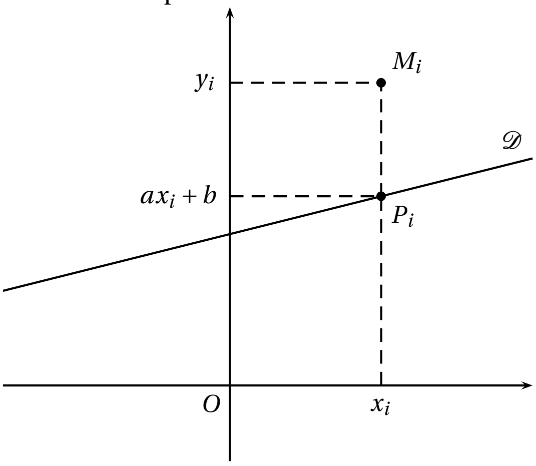

Pourcentage
Horaire quotidien


Tension maximale
Age

Activité 3.3 (Principe de la méthode des moindres carrés).

On peut mesurer la distance d'une droite $\mathscr{D}$ à un nuage de points en calculant la somme des carrés des distances M_iP_i où pour chaque i, le point M_i est un point du nuage et le point P_i est le point de la droite $\mathscr{D}$ ayant la même abscisse que M_i (voir la figure 3.1 de la présente page).

FIGURE 3.1 – Principe de la méthode des moindres carrés



Plus cette somme sera petite et plus la droite sera proche du nuage de points.

On procédera suivant la démarche :

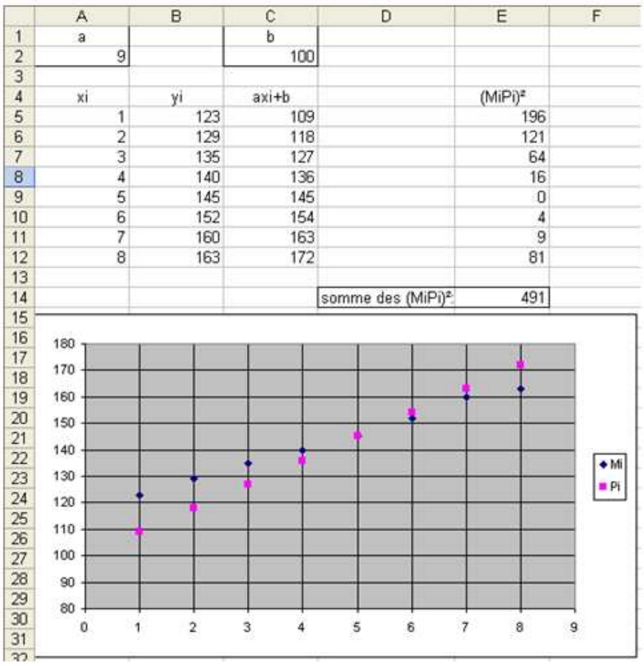
- Première étape : Chercher une équation de droite ($y = ax + b$) qui passe au plus près des points du nuage.
- Deuxième étape : Calculer pour chaque M_i la valeur $M_iP_i^2 = (y_i - ax_i - b)^2$.
- Troisième étape : Chercher à minimiser la somme des $M_iP_i^2$.

Voici la série statistique étudiée :

x_i	1	2	3	4	5	6	7	8
y_i	123	129	135	140	145	152	160	163

1. Saisir la feuille de calcul

   A l'aide d'un tableur, reproduire la feuille de calcul sur le modèle de la figure de la présente page.



	A	B	C	D	E	F
1	a		b			
2	9		100			
3						
4	xi	yi	axi+b		(MiPi)²	
5	1	123	109		196	
6	2	129	118		121	
7	3	135	127		64	
8	4	140	136		16	
9	5	145	145		0	
10	6	152	154		4	
11	7	160	163		9	
12	8	163	172		81	
13						
14				somme des (MiPi)²	491	

- En A2 est inscrit le coefficient directeur et en C2 est inscrit l'ordonnée à l'origine de la droite d'ajustement.
- Pour les lignes 5 à 12, on trouve :
  - dans la colonne A, la série (x_i) des abscisses des points du nuage
  - dans la colonne B, la série (y_i) des ordonnées des points du nuage
  - dans la colonne C, les ordonnées des points P_i de la droite

  · dans la colonne E, $M_iP_i^2 = (\ldots\ldots\ldots)^2$
- Pour la cellule C5, il faut écrire = ……… puis copier-coller cette formule de C6 à C12.

2. Modifier l'ordonnée à l'origine
   D'après le graphique, il semble que la droite d'équation $y = 9x+100$ ne soit pas le meilleur ajustement : l'ordonnée à l'origine n'est pas assez grande.
   (a) Remplacer le contenu de la cellule C2 par 115.
   (b) Essayer par tâtonnement de trouver la valeur de a (au dixième près) qui minimise la somme des $M_iP_i^2$.
   $a \approx \ldots\ldots\ldots$
   (c) Quelle est la somme des $M_iP_i^2$ avec cet ajustement ?
   somme des $M_iP_i^2 \approx \ldots\ldots\ldots$
3. Modifier le coefficient directeur
   (a) Remplacer le coefficient directeur de l'ajustement par 6 (cellule A2).
   (b) Essayer par tâtonnement de trouver la valeur de b (au dixième près) qui minimise la somme des $M_iP_i^2$.
   $b \approx \ldots\ldots\ldots$
   (c) Quelle est la somme des $M_iP_i^2$ avec cet ajustement ?
   somme des $M_iP_i^2 \approx \ldots\ldots\ldots$
4. En appliquant la méthode de MAYER
   (a) En prenant les quatre premiers points du nuage, puis les quatre derniers, déterminer avec le tableur, les coordonnées des deux points moyens G_1 et G_2 des deux sous-séries.
   G_1 (…… ; ……) et G_2 (…… ; ……)
   (b) Déterminer l'équation réduite de la droite (G_1G_2) de MAYER.
   $y = \ldots\ldots x + \ldots\ldots$
   (c) Avec le tableur, quelle est la somme des $M_iP_i^2$ avec cet ajustement ?
   somme des $M_iP_i^2 \approx \ldots\ldots\ldots$
5. En utilisant la méthode des moindres carrés
   (a) A l'aide du tableur, donner l'ajustement affine par la méthode des moindres carrés.
   $y = \ldots\ldots x + \ldots\ldots$
   (b) Quelle est la somme des $M_iP_i^2$ avec cet ajustement ?
   somme des $M_iP_i^2 \approx \ldots\ldots\ldots$

   Remarque. Avec le tableur, on utilisera les fonctions DROITEREG pour déterminer le coefficient directeur et COEFFICIENT.DIRECTEUR pour déterminer le coefficient directeur de la droite d'ajustement.

3.2 Bilan et compléments

3.2.1 Série statistique à deux variables

Définition 3.1. On appelle série statistique à deux variables, l'étude simultanée de deux variables statistiques définies sur une même population.

Exemples.
- Le poids et la taille de nouveaux nés dans une maternité.
- Le volume des ventes et le montant alloué à la publicité dans une entreprise.
- La consommation d'un véhicule et sa vitesse.

En notant $(x_i)_{1\leqslant i\leqslant p}$ et $(y_i)_{1\leqslant i\leqslant p}$ les p valeurs prises par les deux variables statistiques, les données sont données à l'aide d'un tableau d'effectif :

Variable X	x_1	x_2	...	x_p
Variable Y	y_1	y_2	...	y_p

On note $(x_i\,;\,y_i)$ la série statistique double ainsi définie.

Exemple 3.1. Dans toute la suite du chapitre, on se référera à la série statistique suivante : On mesure l'allongement Y d'un ressort en fonction de la masse suspendue X.

Masse (en g)	30	40	50	60	70	80	90	100
Allongement (en mm)	12	19	24	30	37	42	48	55

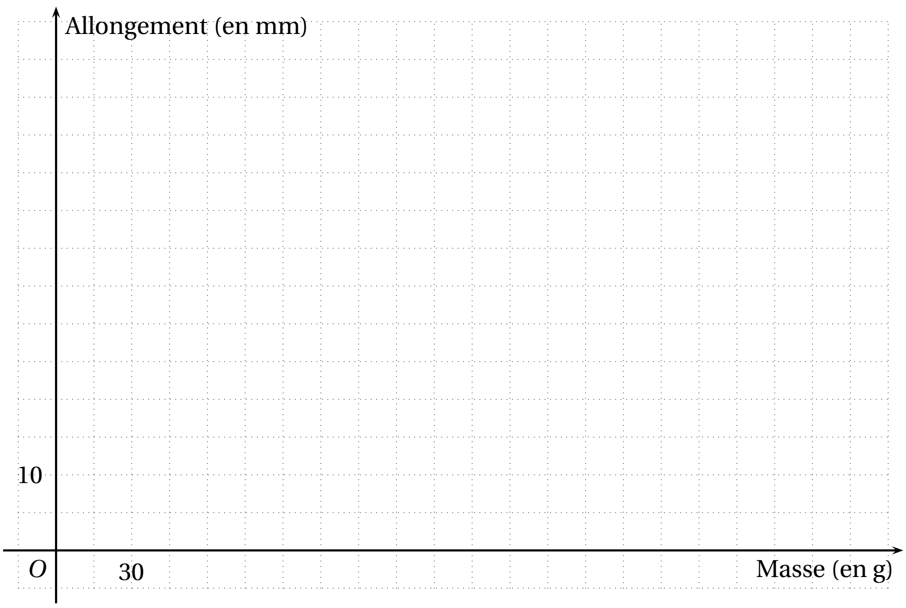
3.2.2 Nuage de points

Définition 3.2. Le plan étant muni d'un repère orthogonal, on peut associer à chaque couple $(x_i\,;\,y_i)$ de la série statistique le point M_i de coordonnées $(x_i\,;\,y_i)$. Le graphique ainsi obtenu constitue un nuage de points.

Exemple. Les points du nuage ont pour coordonnées $M_1\,(30;12)$, $M_2\,(40;19)$, $M_3\,(50;24)$, ..., $M_8\,(100;55)$.

Placer ces points dans le repère de la figure 3.2 de la présente page (Unités graphiques : 1 cm pour 10 g en abscisses et 1 cm pour 10 mm en ordonnées, on graduera l'axe des abscisses à partir de 30).

FIGURE 3.2 – Nuage de points



Définition 3.3. On appelle point moyen du nuage de N points M_i de coordonnées $(x_i\,;\,y_i)$ le point G de coordonnées :

$$x_G = \overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{et } y_G = \overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Exemple. Dans la série précédente, le point moyen G a pour coordonnées :

$$x_G = \dots\dots\dots \text{ et } y_G = \dots\dots\dots \quad \text{donc } G(\dots\dots\dots\,;\dots\dots\dots)$$

3.2.3 Corrélation

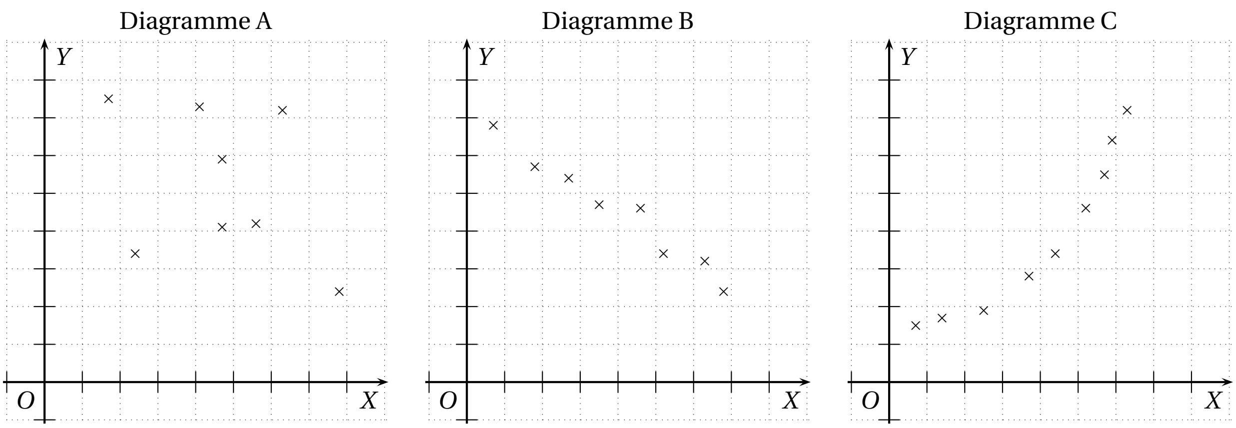
Soient X et Y deux variables statistiques définies sur une même population. Dans certains cas, on peut soupçonner l'existence d'un lien entre ces deux variables. Par exemple, l'allongement du ressort est fonction de la masse suspendue, la taille des enfants peut dépendre de la taille des parents, etc.

Définition 3.4. Il y a corrélation entre deux variables X et Y observées sur les individus d'une même population lorsque X et Y varient dans le même sens ou en sens contraire.

Remarque. L'existence d'une corrélation entre deux variables peut être décelée à l'aide d'un nuage de points.

Exemples. Considérons les diagrammes de la figure 3.3 de la présente page

FIGURE 3.3 – Corrélation



Aucune liaison apparente ; absence de corrélation entre les deux variables — Corrélation négative — Corrélation positive

Lorsque les points du nuage ont tendance à s'aligner comme pour le diagramme B, on parle alors de corrélation linéaire entre les deux variables. On essaye alors d'effectuer un ajustement affine, ce qui consiste à trouver une droite qui rend compte de la forme alignée du nuage en approchant « au mieux » les points qui le constituent.

3.2.4 Ajustement affine

Ajustement à la règle

Exemple. La droite $\mathscr{D}$ contenant les points $M_1\,(30;12)$ et $M_8\,(100;55)$ approche de façon satisfaisante les points du nuage. Déterminer une équation de cette droite $\mathscr{D}$.
En utilisant l'équation de la droite d'ajustement, déterminer quel serait l'allongement du ressort correspondant à une masse de 110 grammes.

Méthode de MAYER

On fractionne le nuage de points en deux nuages partiels de même effectif (à un près si l'effectif de la série est impair) : le premier nuage comprend les points ayant les abscisses les plus petites, et l'autre, les plus grandes.
On détermine ensuite les points moyens G_1 et G_2 des deux nuages partiels.
La droite de MAYER du nuage est la droite (G_1G_2).

Exemple. Le premier nuage comprend les points M_1, M_2, M_3 et M_4. Son point moyen G_1 a pour coordonnées :

$$x_{G_1} = \dots\dots\dots \quad y_{G_1} = \dots\dots\dots \quad \text{d'où } G_1\,(\dots\dots\dots;\dots\dots\dots)$$

Avec les points restants, on obtient un second nuage dont le point moyen G_2 a pour coordonnées :

$$x_{G_2} = \dots\dots\dots \quad y_{G_2} = \dots\dots\dots \quad \text{d'où } G_2\,(\dots\dots\dots;\dots\dots\dots)$$

Placer les points G_1 et G_2 et tracer la droite de MAYER.
Déterminer l'équation de la droite (G_1G_2).

On admettra la propriété suivante :

Propriété 3.1. *La droite de* MAYER *d'un nuage de points passe par le point moyen du nuage.*

Vérifier ce résultat sur le graphique et par le calcul.
En utilisant l'équation de la droite de MAYER, déterminer par le calcul la masse correspondant à un allongement de 60 mm. Retrouver ensuite ce résultat sur le graphique.

Méthode des moindres carrés

L'ajustement affine effectué avec la calculatrice est obtenu par une méthode appelée des moindres carrés qui consiste à trouver une droite qui passe « au plus près » des points du nuage. La calculatrice donne le coefficient directeur m et l'ordonnée à l'origine p.

Avec la calculatrice, déterminer l'équation de la droite d'ajustement et tracer dans le repère.

3.2.5 Utilisation de la calculatrice

On peut retrouver tous ces paramètres statistiques en utilisant les listes d'une calculatrice.

	TI-82	Casio Graph 25
Effacer les anciennes données	STAT 4 : ClrList 4 2^{nd} L1 , 2^{nd} L2 ENTER	Sélectionner le menu STAT F6 DEL-A F4 YES F1 ▷ DEL-A F4 YES F1
Entrer les nouvelles données. On entre les valeurs des x_i dans la première colonne (L1 ou list 1) et les valeurs des y_i dans la deuxième colonne (L2 ou list 2) ;	STAT 1 : Edit ENTER A l'écran : L1 | L2 | L3 30 | 12 40 | 19 50 | 24 60 | 30 … | …	A l'écran | List 1 | List 2 | List 3 1 | 30 | 12 2 | 40 | 19 3 | 50 | 24 4 | 60 | 30 … | … | …
Calculer les paramètres statistiques	CALC ▷ 4 : Linreg $(ax+b)$ ENTER A l'écran : Linreg $(ax+b)$ $y=ax+b$ $a=$ $b=$ $r=$	CALC F2 REG F3 X F1 A l'écran : Linreg $a=$ $b=$ $r=$ $r^2=$ $y=ax+b$

3.3 Exercices

11 p 90, 15 p 91, 26 et 29 p 96, 36 et 40 p 101