

Chapitre 5

Statistiques à deux variables

Sommaire

5.1 Série statistique à deux variables	56
5.1.1 Définition	56
5.1.2 Nuage de points, point moyen	56
5.2 Ajustement	57
5.2.1 Corrélation	57
5.2.2 Ajustement de y en x	57
5.2.3 Ajustement affine	58
5.3 Utilisation de la calculatrice	60
5.4 Exercices et travaux dirigés	61
5.4.1 Exercices	61
5.4.2 Travaux dirigés	67

5.1 Série statistique à deux variables

La statistique descriptive à deux variables a pour but de mettre en évidence une relation éventuelle qui peut exister entre **deux variables** d'une population, considérées **simultanément**.

5.1.1 Définition

Définition 5.1. On appelle *série statistique à deux variables* X et Y , le relevé simultané de la valeur de deux caractères statistiques X et Y . Elle est donc constituée d'une liste de couples de nombres $(x_i; y_i)$.

Les données d'une série statistique à deux variables sont la plupart du temps présentées dans un tableau où l'on indique les n valeurs des variables X et Y :

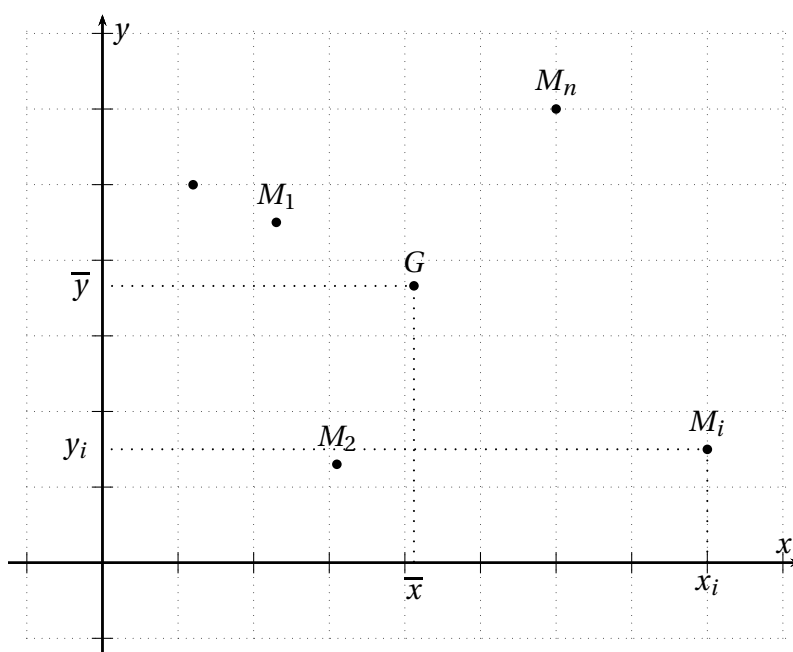
Variable X	x_1	x_2	\dots	x_n
Variable Y	y_1	y_2	\dots	y_n

5.1.2 Nuage de points, point moyen

Définition 5.2. 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 de coordonnées $(x_i; y_i)$. Le graphique ainsi obtenu constitue un *nuage de points*.

Définition 5.3. Le point moyen G du nuage de points est le point de coordonnées $(\bar{x}; \bar{y})$, c'est-à-dire le point dont :

- l'abscisse est la moyenne de la série (x_i) : $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$
- l'ordonnée est la moyenne de la série (y_i) : $\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$



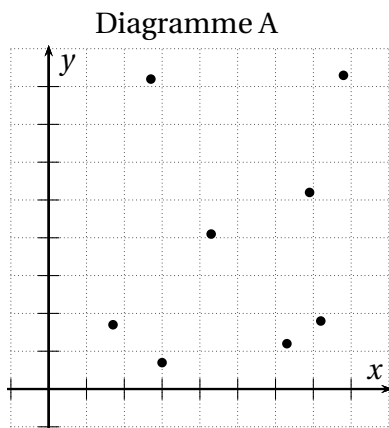
5.2 Ajustement

5.2.1 Corrélation

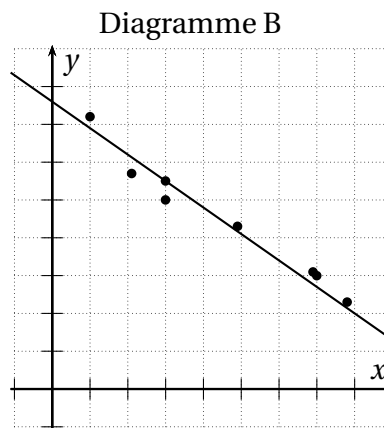
Définition 5.4. Il y a *corrélation* entre deux variables X et Y observées sur les individus d'une même population lorsqu'il y a une relation entre X et Y .

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

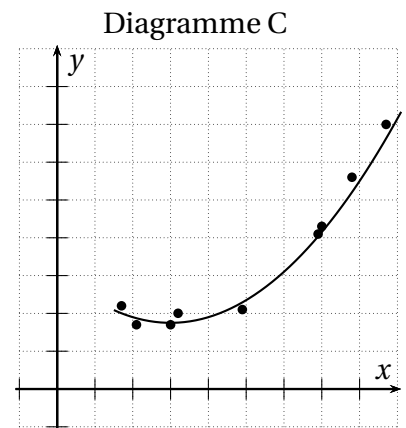
Exemples. Considérons les diagrammes suivants :



Absence de corrélation entre X et Y



Corrélation entre X et Y



Corrélation entre X et Y

5.2.2 Ajustement de y en x

Définition 5.5. Lorsque les valeurs de x sont connues, effectuer un ajustement de y en x d'un nuage de points consiste à trouver une fonction dont la courbe représentative d'équation $y = f(x)$ est la plus « proche » du nuage.

Remarques.

- Un ajustement permet de faire des estimations : *interpolation* (dans l'intervalle d'étude) et *extrapolation* (en dehors).
- Lorsque les points du nuage sont *presque alignés*, comme pour le diagramme B, on recherche une droite qui passe le plus près possibles des points. On effectue alors un *ajustement affine*.
- On verra qu'il existe des ajustements qui ne sont pas affines, comme sur le diagramme C.
- Lorsque les valeurs de y sont connues, on effectue un ajustement de x en y du nuage (on cherche une fonction g dont la courbe d'équation $x = g(y)$ est la plus proche du nuage). C'est le contexte qui donne le type d'ajustement.

5.2.3 Ajustement affine

Une droite d'ajustement affine est une droite qui passe au plus près du nuage de points.

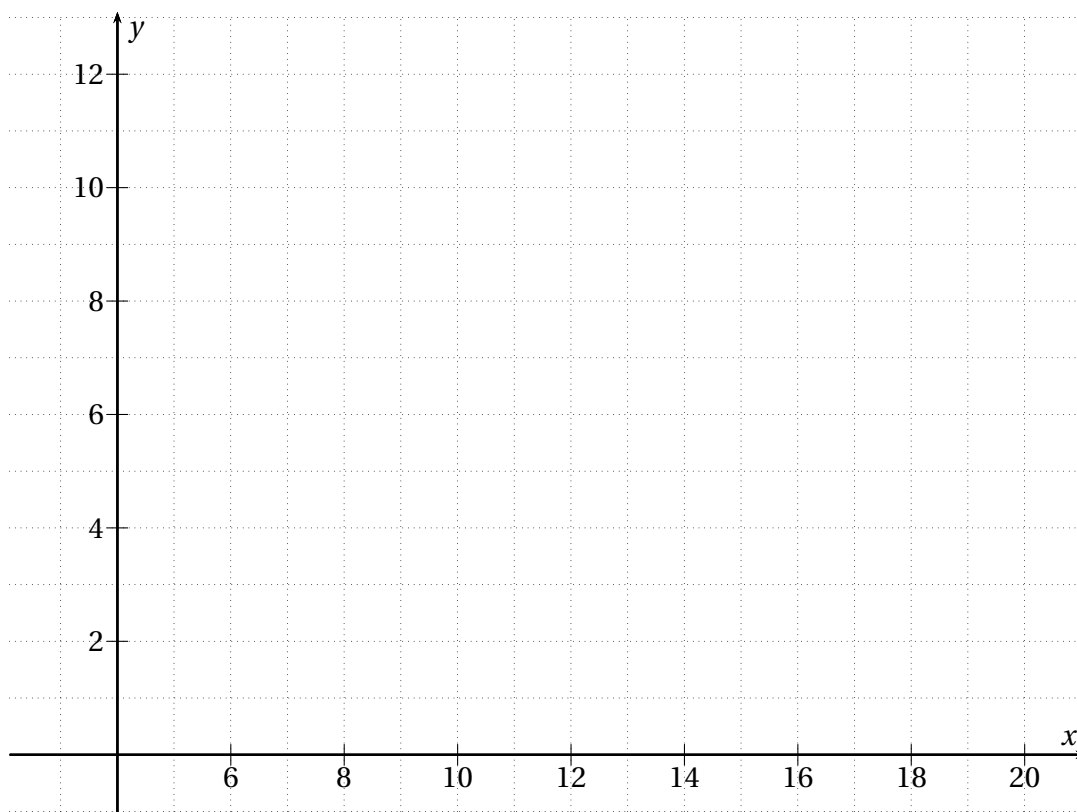
Ajustement au jugé

Le gérant d'un hypermarché, disposant d'un potentiel de vingt-huit caisses enregistreuses, a fait réaliser une étude statistique sur le temps moyen (en minutes) d'attente d'un client à la caisse.

On note x_i le nombre de caisses ouvertes, y_i le temps moyen correspondant en minutes.

x_i	4	5	6	7	8	9	10	11	12	13
y_i	12,25	12	11,5	11,75	10	10	9,75	9	8,25	8

1. (a) Construire dans le repère de la figure ci-dessous le nuage de points associé à la série.



- (b) Calculer les coordonnées du point moyen G du nuage. Placer G dans le repère.
2. Peut-on penser à un ajustement affine? Justifier.
3. On propose comme droite d'ajustement affine la droite \mathcal{D} qui a pour équation $y = -0,5x + 14,5$.
- (a) Tracer la droite \mathcal{D} sur le graphique.
- (b) Le point G appartient-il à la droite \mathcal{D} ? Justifier par un calcul.
- (c) En utilisant cette droite :
- déterminer le temps moyen d'attente d'un client à la caisse lorsque 20 caisses sont ouvertes
 - déterminer le nombre de caisses à ouvrir pour que le temps moyen d'attente d'un client à la caisse soit de trois minutes

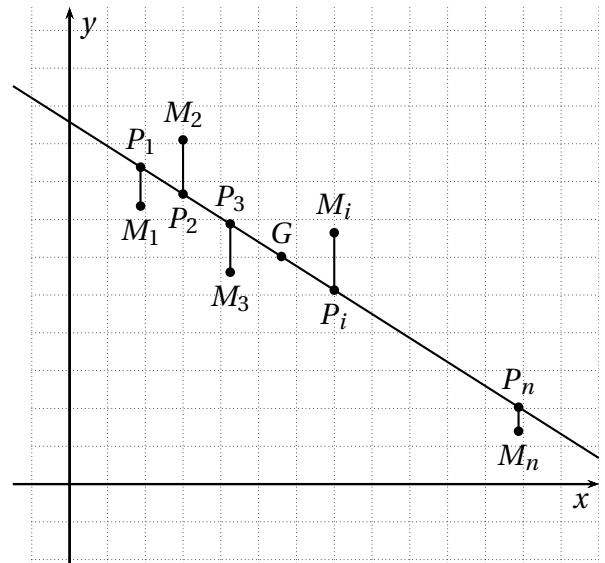
Ajustement par la méthode des moindres carrés

On connaît les valeurs x_i , on cherche à obtenir une droite d'ajustement dont les valeurs y sont les plus proches possibles des y_i « verticalement ».

Les points M_1, M_2, \dots, M_n sont les points du nuage.

Les points P_1, P_2, \dots, P_n sont les points d'une droite \mathcal{D} de mêmes abscisses que, respectivement, M_1, M_2, \dots, M_n , d'équation $y = ax + b$ qui est telle que la somme $S = M_1P_1^2 + M_2P_2^2 + \dots + M_nP_n^2$ soit minimale.

On admet qu'une telle droite existe toujours et on dit que cette droite réalise un ajustement affine du nuage de y en x par la méthode dite *des moindres carrés*. Elle passe par le point moyen G du nuage.



Définition 5.6. La droite \mathcal{D} d'équation $y = ax + b$ obtenue par la méthode des moindres carrés est appelée *droite de régression de y en x* .

Remarques.

- La droite de régression de y en x minimise la somme des carrés des distances en ordonnée
- La droite de régression de x en y minimise la somme des carrés des distances en abscisse. Elle a pour équation $x = a'y + b'$. Elle permet d'estimer les valeurs de X lorsqu'on connaît celles de Y .

Propriété 5.1. Pour apprécier la qualité d'un ajustement affine, on peut déterminer la coefficient de corrélation linéaire, noté r .

On admet que :

- r est toujours compris entre -1 et 1
- r est positif lorsque la droite d'ajustement monte et négatif lorsqu'elle descend
- plus r est proche de -1 ou 1 et meilleure est la qualité de l'ajustement.

Remarque. On détermine a , b et r avec la calculatrice ou le tableur.

Exemple. On reprend l'exemple du gérant d'hypermarché.

1. Avec la calculatrice, déterminer l'équation de la droite d'ajustement Δ obtenue par la méthode des moindres carrés. On arrondira a et b à 10^{-2} .
2. Déterminer le coefficient de corrélation r , arrondi à 10^{-2} . Que peut-on penser de la qualité de cet ajustement?

5.3 Utilisation de la calculatrice

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

Calculatrice Casio

- Effacer les données précédentes : Dans le menu **STAT**, **DELA** (F6), **YES** et **EXE**.
- Saisir les valeurs de X dans la 1^{ère} colonne (LIST1) et celles de Y dans la 2^{ème} colonne (LIST2).

--	--	--

Utilisation du graphique :

- Pour paramétrer le graphique : **GRPH** (F1), **SET** (F6). Choisir **Scatter** pour Graph Type, **List1** pour XList et **List2** pour YList et **EXE**.
- Pour afficher le nuage : **GPH1** (F1).
- Pour afficher les coordonnées du point G : **CALC** (F2), **2VAR** (F2). Faire défiler les valeurs.
- Pour afficher les résultats a, b et r, de la « régression linéaire » (ajustement affine) : **CALC** (F2), **REG** (F3) **X** (F1) et **ax+b** (F1).
- Pour dessiner la droite d'ajustement : **COPY** (F6), puis dans le menu **GRAPH**, la dessiner.

--	--	--

Calcul des paramètres :

- Paramétrer les listes : **CALC** (F2) et **SET** (F6). Choisir **List1** pour 2Var XList et **List2** pour 2Var YList et **List2** pour YList, puis **EXE**.
- Pour afficher les coordonnées du point G : **CALC** (F1) et **2VAR** (F1). Faire défiler avec la touche.
- Pour afficher les résultats a, b et r, de la « régression linéaire » (ajustement affine) : **REG** (F3) et **X** (F1).

Pour obtenir des estimations :

- Pour estimer la valeur de Y pour la valeur X = x₀ : Dans le menu **RUN**, x₀, **OPTN** **STAT** (F5) **ŷ** (F2) et **EXE**.
- Pour estimer la valeur de X pour la valeur Y = y₀ : Dans le menu **RUN**, y₀, **OPTN** **STAT** (F5) **ŷ** (F1) et **EXE**.

Calculatrice TI

- Effacer les données précédentes : **EDIT**, 4:EffacerList **entrer**, L1,L2 (2nde 1, 2nde 2) et **entrer**.
- Dans le menu **STATS** 1:EDIT **entrer**, saisir les valeurs de X dans L1, et celles de Y dans L2 et quitter (2nde Mode).

--	--	--	--

Pour utiliser le graphique :

- Pour paramétrer le graphique : **graph stats** (2nde f(x)), **Graph1** **entrer**.
- Mettre **On** en surbrillance, icône nuage en surbrillance, **L1** pour Xlist et **L2** pour Ylist, etc...
- Pour afficher le nuage : **Graph**. Modifier la fenêtre graphique **fenetre** en prenant en compte les valeurs des x_i et y_i.

--	--	--	--

Pour obtenir les paramètres :

- Paramétrer les listes :
- Pour afficher les coordonnées du point G : **CALC**, 2:STATS 2-Var **entrer** et L1,L2 (2nde 1, 2nde 2). Faire défiler les valeurs.
- Pour afficher les résultats a, b de la « régression linéaire » (ajustement affine) : **CALC**, 4:RégLin(ax+b) **entrer** L1,L2 et **entrer**.
- Pour afficher le coefficient de corrélation r : **VAR**, 5:Statistiques **entrer**, mettre EQ en surbrillance 7:r **entrer** **entrer**.

Pour obtenir des estimations :

- Pour estimer la valeur de Y pour la valeur X = x₀ : **def tbl** (2nde fenetre).
- Compléter avec la valeur de x₀, **table** (2nde Graph). Affichage sur la 1^{ère} ligne de la valeur de y₀.

5.4 Exercices et travaux dirigés

5.4.1 Exercices

EXERCICE 5.1.

On considère les quatre séries statistiques suivantes :

x_i	1	2	3	4	5	6	7	8
y_i	1	1	4	5	3	8	9	8

x_i	1	2	3	4	5	6	7	8
y_i	0	3	8	15	24	35	48	63

x_i	1	2	3	4	5	6	7	8
y_i	0	10	20	30	40	50	60	70

x_i	1	2	3	4	5	6	7	8
y_i	6,7	9,3	10,7	10,2	11,6	12	8,7	8,2

1. Représenter le nuage de points associé à chacune de ces séries sur la calculatrice.
2. Pour lesquels, y a-t-il corrélation entre les variables ? Préciser les nuages pour lesquels un ajustement affine est judicieux.

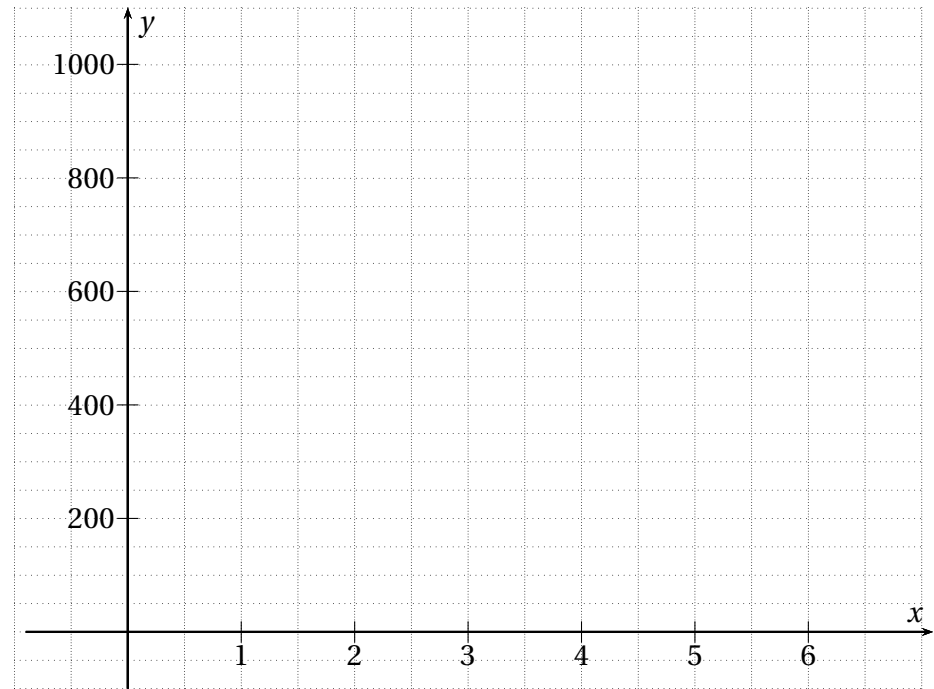
EXERCICE 5.2.

Le service commercial de la société LACOM SA, spécialisée dans la fabrication de fibres optiques destinées aux opérateurs téléphoniques, a mis sur le marché un nouveau produit dont l'évolution des ventes depuis 18 mois est résumée dans la tableau ci-dessous.

Trimestre x_i	1	2	3	4	5	6
Quantités vendues y_i	255	330	435	570	740	960

1. Représenter ci-contre le nuage de points $M_i(x_i; y_i)$, avec $1 \leq i \leq 6$, associé à cette série statistique.

2. Un ajustement affine vous semble-t-il justifié ?



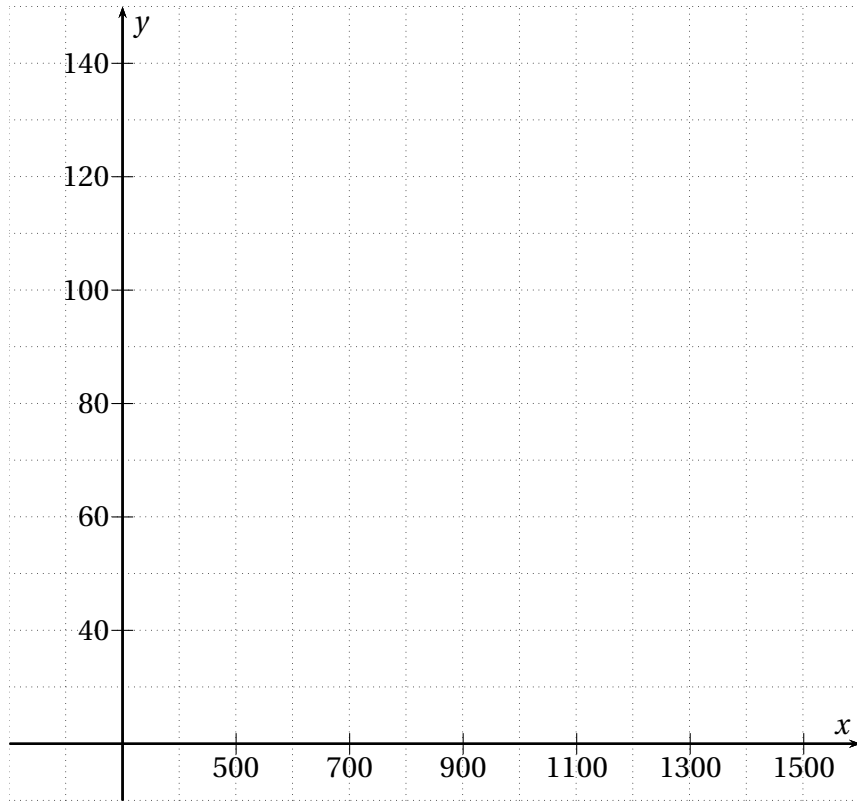
EXERCICE 5.3.

Avant de proposer une croisière, une agence de voyage étudie le nombre de clients qui seraient prêts à réserver ce nouveau produit. Elle obtient les résultats suivants :

Prix en euros x_i	300	750	800	1 050	1 150	1 350	1 450	1 500
Nombre de clients y_i	140	130	125	90	85	50	25	20

1. (a) Représenter le nuage de points dans le graphique ci-contre.
(b) Un ajustement affine vous semble-t-il justifié ?
(c) Calculer les coordonnées du point moyen G et l placer dans le repère.
2. On choisit de réaliser un ajustement du nuage par la droite \mathcal{D} , dite « droite des extrêmes » passant par $M_1(300; 140)$ et $M_8(1 500; 20)$.

- (a) Déterminer l'équation réduite de la droite \mathcal{D} .
- (b) Le point G appartient-il à la droite \mathcal{D} ?
- (c) Tracer la droite \mathcal{D} dans le repère.
Réalise-t-elle un bon ajustement du nuage? Expliquer.
- (d) En utilisant l'ajustement par la droite \mathcal{D} , calculer le prix à partir duquel moins de 100 personnes seraient prêtes à réserver cette croisière.



Année	2001	2002	2003	2004	2005	2006	2007	2008	2009
Rang de l'année x_i	1	2	3	4	5	6	7	8	9
Quantité (millions de tonnes) y_i	11,6	11,4	11,2	10,9	10,7	10,2	9,8	9,1	8,7

EXERCICE 5.4.

Le tableau ci-dessous donne les quantités de super sans plomb livrées et vendues (en millions de tonnes) en France de 2001 à 2009 (source : INSEE).

- (a) Représenter dans un plan muni d'un repère orthogonal le nuage de points $M_i(x_i; y_i)$ avec $1 \leq i \leq 9$, associé à cette série statistique. On prendra pour unités graphiques :

 - 1 cm pour une année en abscisses
 - 2 cm pour un million de tonnes en ordonnées (commencer la graduation à 7 millions de tonnes)

(b) Déterminer les coordonnées du point moyen G du nuage et placer ce point sur le graphique.
- (a) Déterminer, à l'aide de la calculatrice, le coefficient de corrélation linéaire r de la série statistique $(x_i; y_i)$ (arrondir le résultat à 10^{-3}).

(b) Déterminer, à l'aide de la calculatrice, une équation de la droite d'ajustement de y en x , obtenue par la méthode des moindres carrés (aucune justification n'est demandée). Les coefficients de l'équation de la droite seront arrondis à 10^{-2} .

(c) Tracer la droite d'ajustement obtenue.

(d) En supposant que cet ajustement reste valable jusqu'en 2012, déterminer une estimation des quantités de super sans plomb livrées et vendues pour l'année 2012.

Anée x_i	2004	2005	2006	2007	2008	2009	2010
Nombre de PACS en milliers y_i	40,1	60,5	77,3	102	146	174,5	205,6

EXERCICE 5.5.

Le tableau ci-dessus donne l'évolution, depuis 2004, du nombre de pactes civils de solidarité (PACS) conclus en France jusqu'en 2010.

- À l'aide d'une calculatrice, déterminer le coefficient de corrélation linéaire entre ces deux variables (arrondir à 0,001 près).
- Un ajustement affine vous semble-t-il justifié?
- À l'aide d'une calculatrice, déterminer une équation de la droite \mathcal{D} d'ajustement affine de y en x , obtenue par la méthode des moindres carrés (arrondir la valeur du coefficient a de la droite au millième).
- En utilisant cette droite, déterminer :
 - une estimation du nombre de PACS en 2012
 - l'année au cours de laquelle le nombre de PACS devrait dépasser 300 000

EXERCICE 5.6.

Un fournisseur d'accès à Internet souhaite faire une prévision du nombre de ses abonnés pour l'année 2018. Il établit un relevé du nombre des abonnés des années de 2011 à 2015. Il affecte l'indice 100 à l'année 2011 pour établir la statistique des abonnés et consigne les données dans le tableau ci-dessous.

Année	2011	2012	2013	2014	2015
Rang x_i	1	2	3	4	5
Indice y_i	100	112	130	160	200

- (a) Le nombre d'abonnés était de 2 040 pour l'année 2011. Déterminer le nombre d'abonnés en 2015.

- Déterminer le taux d'évolution du nombre d'abonnés (en%) entre 2014 et 2015.
- Représenter ci-contre le nuage de points associé à cette série statistique dans le repère ci-dessous.

2. Un ajustement affine

- Déterminer, à l'aide de la calculatrice, une équation de la droite de régression de y en x par la méthode des moindres carrés sous la forme $y = ax + b$ (arrondir a à 10^{-1} et b à l'unité près).
- Construire la droite dans le repère.
- Déterminer le coefficient de corrélation linéaire r de la série (arrondir à 10^{-3}). Quelle information fournit-il?
- Quelles prévisions du nombre d'abonnés peut-on faire pour les années 2018 et 2020 (arrondir à l'unité)?

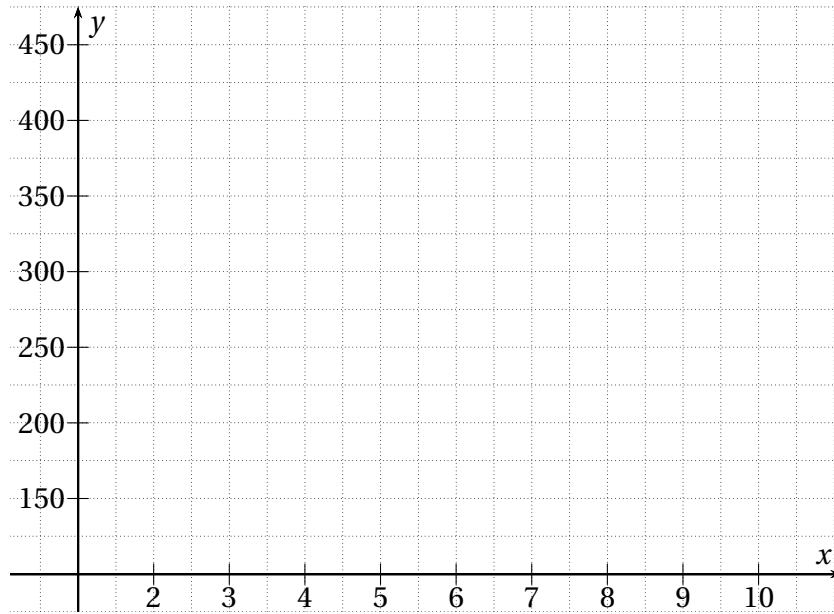
3. Ajustement à l'aide d'un logiciel

En observant le nuage de points, on peut conjecturer qu'une parabole pourrait convenir comme courbe d'ajustement du nuage. Un logiciel de calcul propose d'ajuster le nuage de points à l'aide d'une partie de la courbe \mathcal{C} d'équation $y = 5x^2 - 5x + 100$. On note g la fonction polynôme du second degré associée à la courbe \mathcal{C} .

- Calculer $g'(x)$ et en déduire le tableau de variations de g sur $[0; 10]$.
- Tracer la courbe \mathcal{C} dans le graphique.
- Déterminer, par le calcul, l'ordonnée du point de la courbe \mathcal{C} qui a pour ordonnée 8.

(d) On suppose que le modèle défini par la courbe \mathcal{C} est valable jusqu'en 2018.

Donner, selon ce modèle, la valeur de l'indice et comparer avec le résultat du 2d.



EXERCICE 5.7 (Évolution de l'indice IMVP – Source : Renault).

L'indice IMVP (International Motor Vehicle Program) est un indicateur de référence élaboré par le MIT (Massachusetts Institute of Technology) qui mesure, en heures, le temps de montage moyen d'un véhicule.

Dans une entreprise de construction automobile, on a obtenu le tableau suivant.

Année	2005	2006	2007	2008	2009	2010	2011
Rang de l'année x_i	5	6	7	8	9	10	11
Temps en heure y_i	26,2	23,7	21,4	18,5	16,8	15,4	14,6

Partie A. Le nuage de points M_i associé à la série statistique $(x_i; y_i)$ dans un plan rapporté à un repère orthonormé est donné sur la figure 5.1 page suivante.

Les résultats seront arrondis si nécessaire au centième.

- Calculer les coordonnées du point oyen G et le placer sur le graphique.
- Le nuage de points montre qu'un ajustement affine semble justifié.
Le vérifier à l'aide du coefficient de corrélation linéaire.
- (a) À l'aide de la calculatrice, donner une équation de la droite \mathcal{D} d'ajustement affine de y en x obtenue par la méthode des moindres carrés.
(b) Représenter \mathcal{D} sur le graphique.
- (a) En déduire graphiquement puis par le calcul les prévisions du temps de montage moyen pour l'année 2015 puis l'année 2017.
(b) Calculer la variation en pourcentage de ce temps de l'année 2010 à l'année 2011.

Partie B. On décide d'approcher ce nuage par un arc de parabole; pour cela on pose $z = \sqrt{y}$.

- Donner le tableau des valeurs $(x_i; z_i)$. Les valeurs z_i seront arrondies au millième.
- (a) Donner le coefficient de corrélation linéaire de la série de z en x . L'ajustement affine de z en x est-il justifié?
(b) Donner une équation de la droite d'ajustement affine obtenue par la méthode des moindres carrés (les coefficients seront arrondis au millième).
- (a) En déduire l'expression de y en fonction de x , puis le temps de montage en 2015 et en 2017 arrondis au dixième.
(b) Ces temps sont-ils plus plausibles que ceux obtenus dans la partie A? Expliquer.

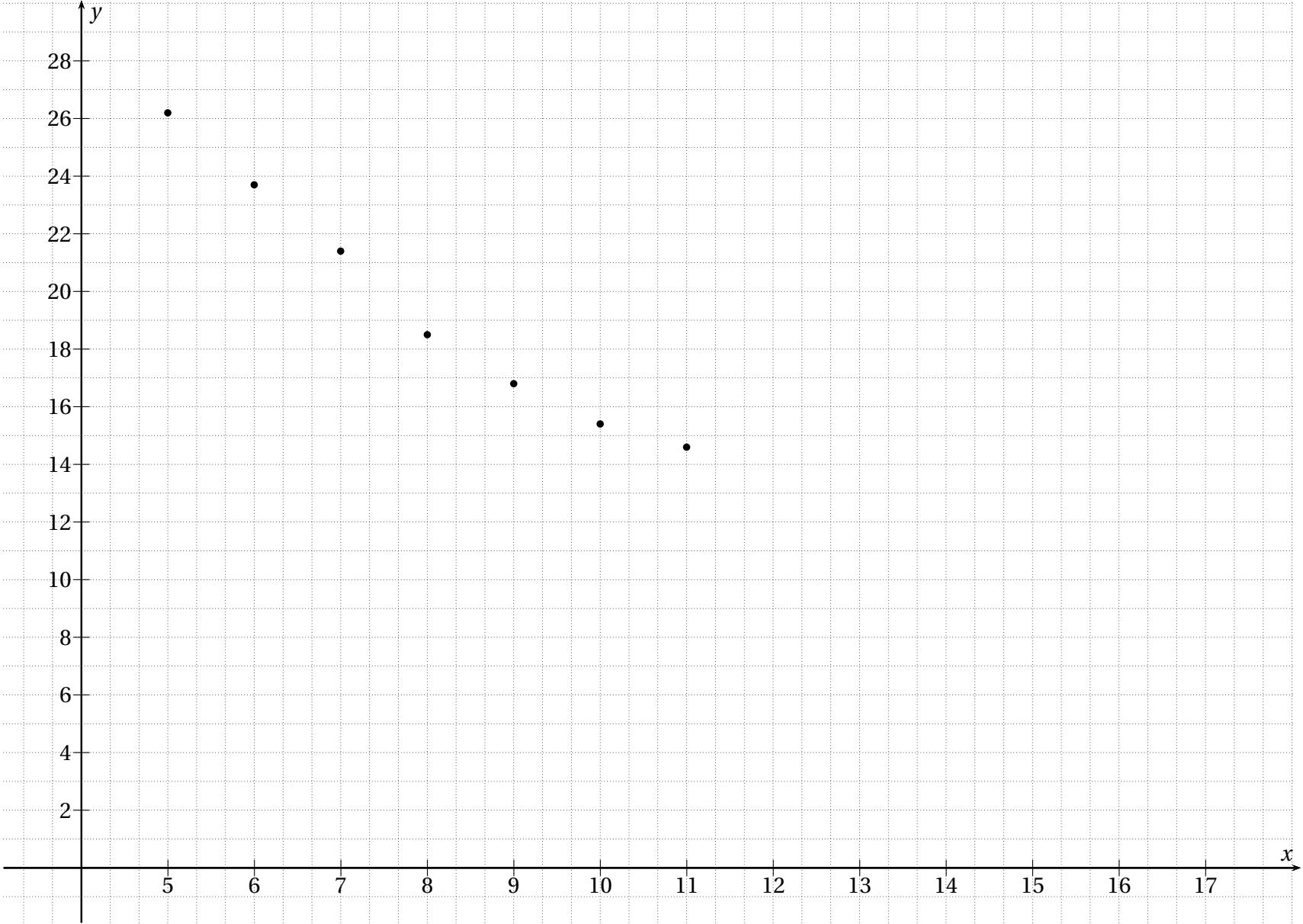


FIGURE 5.1: Figure de l'exercice 5.7

5.4.2 Travaux dirigés

EXERCICE 5.8.

On donne dans le tableau suivant le chiffre d'affaires, en milliards d'euros d'une grande entreprise.

Année	2000	2002	2007	2010	2015
Rang de l'année x_i	1	3	5	8	13
Chiffre d'affaires y_i	28	27,2	30,6	37,6	40,7

On recherche une droite \mathcal{D} d'équation $y = ax + b$, passant par le point moyen G et ajustant au mieux le nuage des quatre années données.

On souhaite obtenir la feuille de calculs ci-dessous.

	A	B	C	D	E	F
1	Année	Rang de l'année x_i	Chiffre d'affaires y_i	$ax + b$	Écarts verticaux	Écarts au carré
2	2000	1	28			
3	2002	3	27,2			
4	2004	5	30,6			
5	2007	8	37,6			
6	2010	13	40,7			
7	2015	15				Somme
8				50		
9						
10	Moyenne					
11						
12	a	1				
13	b					
14	r					

Partie A. Ajustement manuel d'une droite passant par le point moyen.

- Quelle formule peut-on saisir en **B10** et recopier en **C10** pour calculer les coordonnées du point moyen G ?
- Dans cette question, on choisit comme droite d'ajustement du nuage, la droite \mathcal{D} qui passe par G et qui a pour coefficient directeur 1.

Puisque la droite \mathcal{D} passe par le point G alors $\bar{y} = a\bar{x} + b \Leftrightarrow b = \bar{y} - a\bar{x}$.

- Quelle formule doit-on saisir en **B12** pour calculer b ?

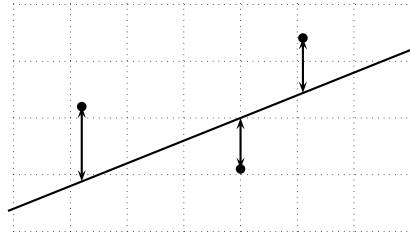
- Quelle formule doit-on saisir en **D2** et recopier vers le bas pour calculer les ordonnées des points de la droite \mathcal{D} ?
- Quelle formule doit-on saisir en **E2** et recopier vers le bas pour calculer les écarts verticaux entre les points du nuage et les points correspondants de la droite \mathcal{D} ?
- Quelle formule peut-on saisir en **F2** et recopier vers le bas pour calculer les carrés de ces écarts?
- Pourquoi la somme des écarts obtenus dans la colonne **E** ne permet-elle pas de savoir si la droite est proche des points du nuage? Quel est l'avantage de considérer la somme des écarts au carré dans la colonne **F**?

- Représenter le nuage de points de la série ainsi que celui donné par la droite d'ajustement.
 - Modifier la valeur du coefficient directeur a de la droite \mathcal{D} en **B12** (essayer les valeurs 1,1, 1,2, 1,3, etc.) de telle sorte que la somme des écarts au carré soit minimale. Observer en même temps la droite \mathcal{D} sur le graphique.
 - Quelle est, à 10^{-1} près, le coefficient a optimal?
 - En utilisant cette droite d'ajustement :
 - quelle estimation du chiffre d'affaire (arrondi à 0,1 milliard près) peut-on faire en **D7** pour l'année 2015?
 - quelle estimation de l'année peut-on faire en **B8** pour obtenir un chiffre d'affaire de 50 milliards d'euros? Indication : utiliser la commande **valeur cible**.

Partie B. Ajustement par la droite des moindres carrés de y en x .

On connaît le rang de l'année x et on cherche à estimer le chiffres d'affaires y .

Le tableur peut afficher directement la droite d'ajustement de y en x souhaitée avec le principe précédent, dit des « moindres carrés », c'est-à-dire en minimisant la somme des carrés des écarts verticaux.

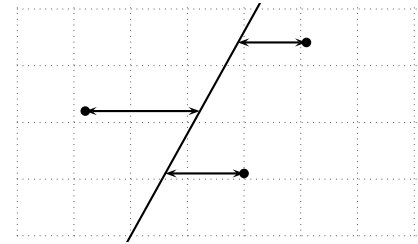


1. (a) Sur le graphique, cliquer droit sur un point du nuage situé à l'écart de la droite et choisir **Courbe de tendance**. Dans la boîte de dialogue choisir **Linéaire** et cocher **Afficher l'équation**.
- (b) Quelle est l'équation de la droite d'ajustement donnée par le tableur ?
- (c) Quelle formule doit-on saisir en **B14** pour calculer le coefficient de corrélation ?
2. On souhaite exploiter l'équation de la droite précédente pour estimer le chiffre d'affaires en 2014 (rang 15).
 - (a) Quelle formule doit-on saisir en **C7** pour calculer une estimation du chiffre d'affaires selon le modèle de la droite de régression ?
 - (b) Quelle est le résultat affiché (arrondir à 10^{-1} milliard près).
 - (c) Comparer le résultat à celui affiché en **D7**.

Partie C. Ajustement par la droite des moindres carrés de x en y .

On connaît le chiffre d'affaires y et on cherche à obtenir le rang de l'année x .

Le tableur peut afficher directement la droite d'ajustement de x en y souhaitée avec le principe précédent, dit des « moindres carrés », c'est-à-dire en minimisant la somme des carrés des écarts horizontaux.



1. recommencer le même travail et donner l'équation de la droite d'ajustement de x en y sous la forme $x = ay + b$ obtenue avec la méthode des moindres carrés.
2. On souhaite exploiter l'équation de la droite pour estimer l'année pour laquelle le chiffre d'affaires sera de 50 milliards d'euros.
 - (a) Quelle formule peut-on saisir dans la cellule **B7** pour obtenir cette estimation ?
 - (b) Quel est le résultat affiché ? (arrondir à l'unité)