

12

Statistiques à deux variables

1. Nuage de points et point moyen

On s'intéresse à deux caractères d'une population, notés x et y .

On obtient deux listes de nombres, qui sont deux séries statistiques.

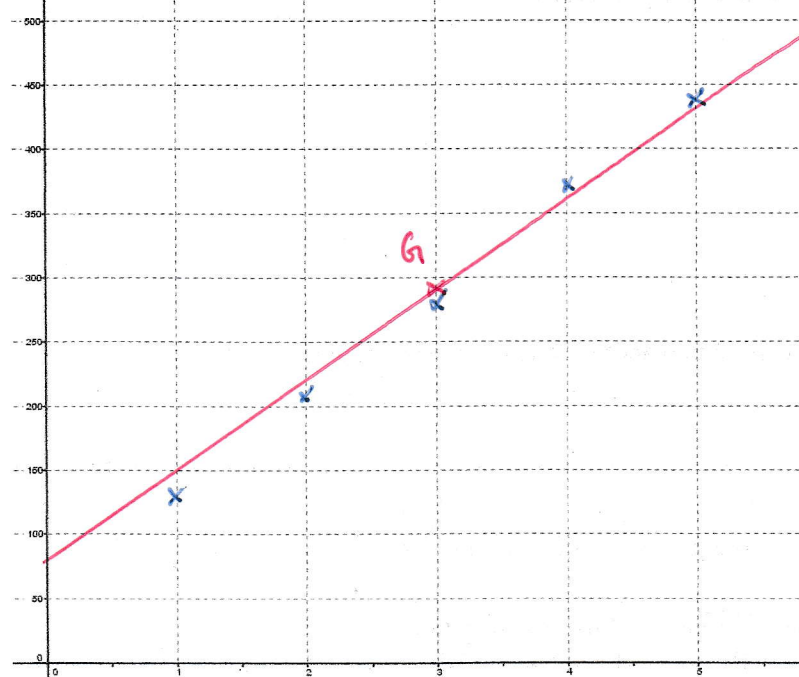
On peut alors tracer un **nuage de points** en prenant en abscisse les valeurs x_i de la première série statistique et en ordonnée les valeurs y_i de la deuxième.

A. Nuage de points

Exercice 12.1 Une salle de musculation a ouvert ses portes en 2008.

Le tableau suivant donne le nombre d'adhérents chaque année :

Année	2008	2009	2010	2011	2012
Rang de l'année (x_i)	1	2	3	4	5
Nombre d'adhérents (y_i)	126	201	275	375	438



Pour les plus rapides : Tracer une droite qui passe "à peu près" par les points obtenus, et essayer de donner une équation approximative de cette droite (par lecture graphique).

$y = 75x + 50$

Définition 12.1 Les points obtenus s'appellent un **nuage de points**.

B. Point moyen

Exercice 12.2 1. Avec les deux séries de données (x_i et y_i) de l'exercice précédent, calculer :

(a) La moyenne de la série (x_i), la noter $\bar{x}$

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3$$

(b) La moyenne de la série (y_i), la noter $\bar{y}$

$$\bar{y} = \frac{126+201+275+375+438}{5} = 283$$

$G(3; 283)$

2. Placer le point de coordonnées ($\bar{x}$; $\bar{y}$) sur le graphique précédent, noter ce point G.

Pour les plus rapides : Modifier légèrement l'équation de la droite trouvée précédemment pour qu'elle passe par G.

Définition 12.2 Le point G de coordonnées ($\bar{x}$; $\bar{y}$) s'appelle le **point moyen**.

C. Exercices bilan

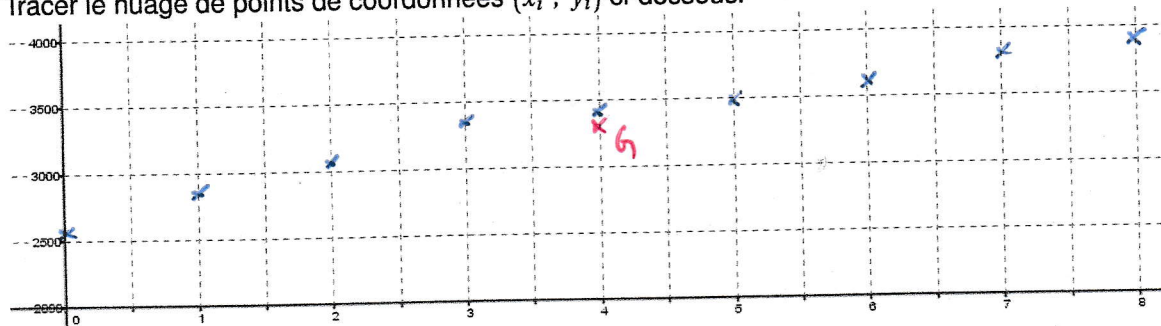
Exercice 12.3 On s'intéresse à l'évolution du prix des appartements neufs en France métropolitaine.

Le tableau ci-dessous indique le prix des appartements neufs en France métropolitaine, en euros par m^2 , entre 2004 et 2012.

Année	2004	2005	2006	2007	2008	2009	2010	2011	2012
Rang de l'année : x_i	0	1	2	3	4	5	6	7	8
Prix de l'appartement (e/m^2) : y_i	2563	2852	3071	3276	3344	3368	3571	3773	3861

Source Insee

1. Tracer le nuage de points de coordonnées (x_i ; y_i) ci-dessous.



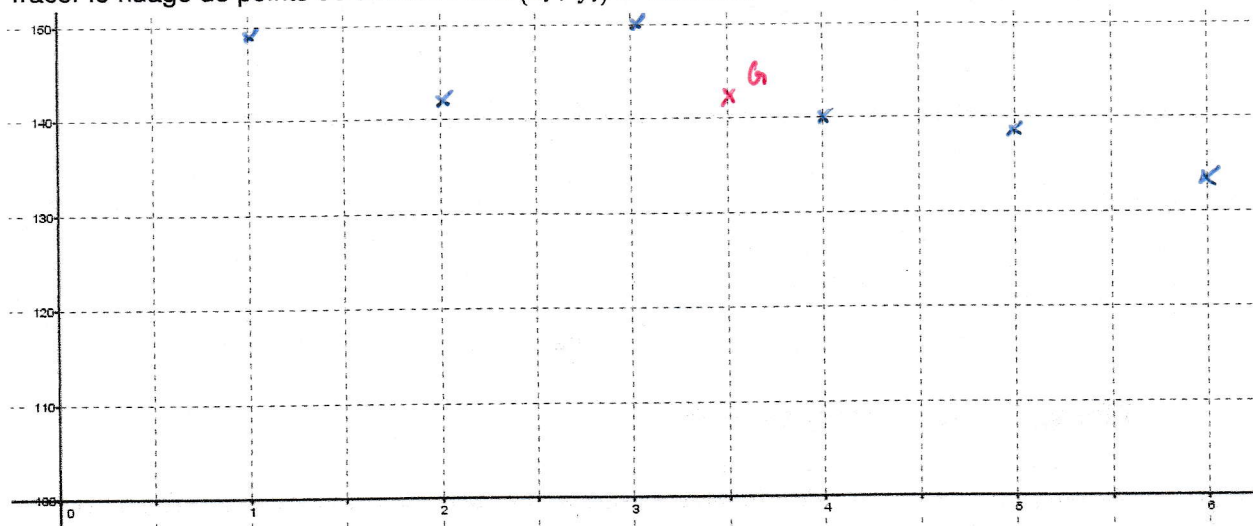
2. Calculer les coordonnées du point moyen, et le placer en couleur sur le graphique ci-dessus.

$$\bar{x} = 4 \text{ et } \bar{y} = 3297,7$$

Exercice 12.4 Le tableau ci-dessous donne le nombre de voitures neuves (en milliers) vendues en France durant les six premiers mois de l'année 2013.

Mois	Janvier	Février	Mars	Avril	Mai	Juin
Rang du mois x_i	1	2	3	4	5	6
Nombre de ventes (en milliers) y_i	149	144	150	140	139	135

1. Tracer le nuage de points de coordonnées $(x_i ; y_i)$ ci-dessous.



2. Calculer les coordonnées du point moyen, et le placer en couleur sur le graphique ci-dessus.

.....
 $\bar{x} = 3,5$, $\bar{y} = 142,83$

2. Ajustement affine

Lorsque les points du nuage de points sont "presque alignés", on peut chercher une droite D d'équation $y = ax + b$ qui passe "le plus près possible" des points du nuage.

On dit alors que l'on fait un **ajustement affine**.

Il existe plusieurs façons d'obtenir cette droite :

- On peut tracer la droite "au jugé"
- On peut utiliser la **méthode des moindres carrés** détaillée dans ce cours
- On peut utiliser d'autres méthodes, qui sont disponibles sur certains tableurs.

A. Covariance de deux séries statistiques

On rappelle la formule de la variance : $V(X) = \frac{x_1^2 + x_2^2 + x_3^2 + \dots}{N} - \bar{x}^2$

La covariance d'une série statistique à deux variables se calcule ainsi :

$$\text{cov}(x, y) = \frac{x_1 y_1 + x_2 y_2 + x_3 y_3 + \dots + x_n y_n}{n} - \bar{x} \bar{y}$$

Exemple :

Année	2008	2009	2010	2011	2012
Rang de l'année (x_i)	1	2	3	4	5
Nombre d'adhérents (y_i)	126	201	275	375	438

On a vu précédemment que $\bar{x} = 3$ et que $\bar{y} = 283$

La covariance est :

$$\begin{aligned} \text{cov}(x, y) &= \frac{1 \times 126 + 2 \times 201 + 3 \times 275 + 4 \times 375 + 5 \times 438}{5} - 3 \times 283 \\ &= \frac{5043}{5} - 3 \times 283 = 1008,6 - 3 \times 283 = 159,6 \end{aligned}$$

Exercice 12.5 Calculer la covariance de la série statistique à deux variables ci-dessous :

Année	2004	2005	2006	2007	2008	2009	2010	2011	2012
Rang de l'année : x_i	0	1	2	3	4	5	6	7	8
Prix de l'appartement (e/m^2) : y_i	2563	2852	3071	3276	3344	3368	3571	3773	3861

$$\text{cov}(x, y) = 1005,222$$

Exercice 12.6 Calculer la covariance de la série statistique à deux variables ci-dessous :

Mois	Janvier	Février	Mars	Avril	Mai	Juin
Rang du mois x_i	1	2	3	4	5	6
Nombre de ventes (en milliers) y_i	149	144	150	140	139	135

$$\bar{x} = 3,5$$

$$\bar{y} = 142,8$$

$$\text{cov}(x, y) = (1 \times 149 + 2 \times 144 + 3 \times 150 + 4 \times 140 + 5 \times 139 + 6 \times 135) - 6 \times 3,5 \times 142,8$$

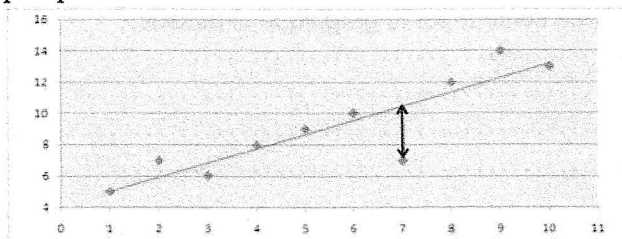
$$= 692 - 3,5 \times 142,8 = -7,8$$

(la covariance peut être négative)
 - mais les 2 vides ne se compensent pas
 "ps peut" us-à-us de leurs moyennes respectives

B. Méthode des moindres carrés

Cette méthode donne l'équation $y = ax + b$ d'une droite D .

C'est la droite pour laquelle les carrés des distances (mesurées verticalement) entre les points du nuage et D sont les plus petits.



L'équation de cette droite s'obtient ainsi :

- Le coefficient directeur est $a = \frac{\text{cov}(x, y)}{V(x)}$, où $\text{cov}(x, y)$ est la covariance, et où $V(x)$ est la variance de la série statistique x_i
- L'ordonnée à l'origine b est choisie de telle manière que le point moyen $G(\bar{x}; \bar{y})$ appartienne à la droite.

Exemple :

Année	2008	2009	2010	2011	2012
Rang de l'année (x_i)	1	2	3	4	5
Nombre d'adhérents (y_i)	126	201	275	375	438

On a vu que la covariance est $\text{cov}(x, y) = 159,6$, que $\bar{x} = 3$ et que $\bar{y} = 283$.

On rappelle la formule de la variance : $V(X) = \frac{x_1^2 + x_2^2 + x_3^2 + \dots}{N} - \bar{x}^2$ (les $n_1, n_2, n_3, \dots$ n'apparaissent pas car dans la série statistique x_i , chaque valeur n'apparaît qu'une fois).

Déterminer $V(X)$:

$$V(X) = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2}{5} - 3^2 = 2$$

$$(V(Y) = 127,85, 2)$$

Déterminer la covariance $\text{cov}(x, y)$:

$$\text{On a vu que } \text{cov}(x, y) = 159,6$$

Déterminer le coefficient directeur a de la droite des moindres carrés :

$$a = 159,6 / 2 = 79,8$$

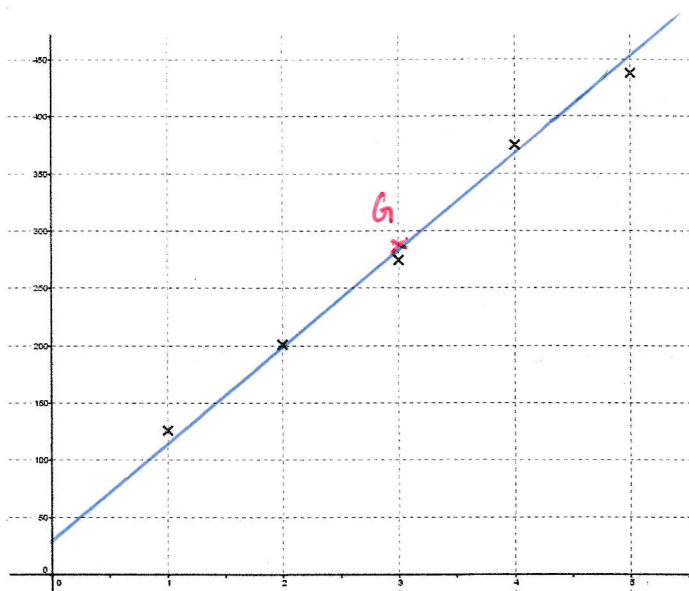
Déterminer l'ordonnée à l'origine b de la droite des moindres carrés, en utilisant les coordonnées du point moyen :

$$\text{On a } y = 79,8x + b, \text{ avec les coordonnées de } G \text{ l'entraînant : } \bar{y} = 79,8 \bar{x} + b$$

$$\text{i.e. } 283 = 79,8 \times 3 + b, \text{ donc } 283 = 239,4 + b,$$

$$\text{donc } b = 283 - 239,4 = 43,6$$

Tracer cette droite sur le graphique ci-dessous, placer également le point moyen G et vérifier que la droite obtenue passe par G :



Finalement, l'équation de la droite des moindres carrés est :

$$y = 79,8x + 43,6$$

Exercice 12.7 On reprend la série statistique à deux variables :

Année	2004	2005	2006	2007	2008	2009	2010	2011	2012
Rang de l'année : x_i	0	1	2	3	4	5	6	7	8
Prix de l'appartement (e/m^2) : y_i	2563	2852	3071	3276	3344	3368	3571	3773	3861

On a vu que la covariance est $cov(x, y) = 1005,09$, que $\bar{x} = 4$ et que $\bar{y} = 3297,7$.

On rappelle la formule de la variance : $V(X) = \frac{x_1^2 + x_2^2 + x_3^2 + \dots}{N} - \bar{x}^2$ (les $n_1, n_2, n_3, \dots$ n'apparaissent pas car dans la série statistique x_i , chaque valeur n'apparaît qu'une fois).

Déterminer $V(X)$:

$$V(X) = \frac{0^2 + 1^2 + 2^2 + \dots + 8^2}{9} - 4^2 = \frac{204}{9} - 16 = 22,7 - 16 = 6,7$$

Déterminer le coefficient directeur a de la **droite des moindres carrés** :

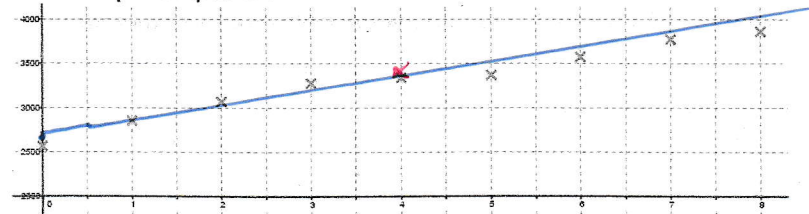
$$a = \frac{cov(x, y)}{V(X)} = \frac{1005,09}{6,7} \approx 150,01$$

Déterminer l'ordonnée à l'origine b de la **droite des moindres carrés**, en utilisant les coordonnées du point moyen :

$$y = 150,01x + b ; 3297,7 = 150,01 \times 4 + b ; 3297,7 = 600,04 + b$$

$$\text{Donc } b = 3297,7 - 600,04 = 2697,7$$

Tracer cette droite sur le graphique ci-dessous, placer également le point moyen G et vérifier que la droite obtenue passe par G :



N.B. : Un graphique est toujours "approximatif".

C. A la calculatrice

Heureusement, l'équation de cette droite des moindres carrés peut être obtenue à la calculatrice.

Un exemple de vidéo sur Youtube illustrant cette utilisation :

<https://youtu.be/DQOLw13rk88>

... mais il y a beaucoup d'autres tutoriels à ce sujet.

Pour les calculatrices TI

- Après avoir appuyé sur la touche *stats*, sélectionner *EDIT*
- Entrer alors les données x_i dans la colonne *L1*, et les données y_i dans la colonne *L2*
- Appuyer sur la touche *stats*; sélectionner *CALC*, puis *RégLin(ax+b)*
- Vérifier que la calculatrice a bien sélectionné *XList :L1* et *YList :L2* (ne rien écrire sur les lignes suivantes), puis valider : on obtient les coefficient a et b .

Représenter le nuage de points avec une TI

- Configurer la calculatrice pour les graphiques statistiques : *2nd STAT PLOT* ou *2nd graph stats* puis *Enter* et valider les choix représentés ci-dessous :

```

2nd STAT PLOT
1:Plot1...On
  L1 L2
2:Plot2...Off
  L1 L2
3:Plot3...Off
  L1 L2
4:PlotsOff
  Type:
  Xlist:L1
  Ylist:L2
  Mark:
  
```

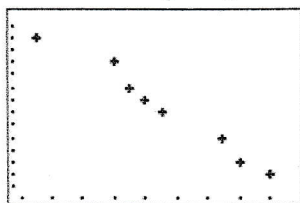
- Régler la fenêtre d'affichage : *WINDOW*, puis *Zoom 9 : Zoomstats*

```

WINDOW
Xmin=1
Xmax=12
Xscl=1
Ymin=10
Ymax=30
Yscl=2
Xres=1

2nd ZOOM MEMORY
3:Zoom Out
4:ZDecimal
5:ZSquare
6:ZStandard
7:ZTrig
8:ZInteger
9:ZoomStat
  
```

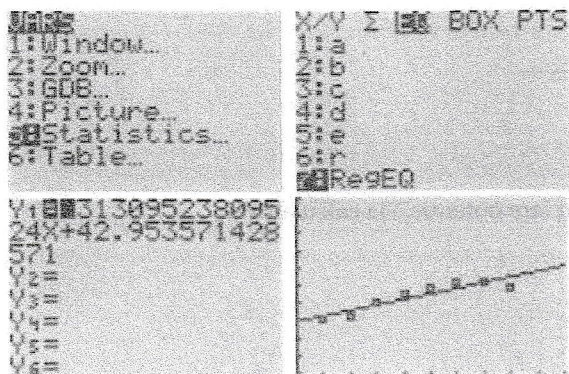
- Afficher le nuage de points : *GRAPH*



Représenter la droite des moindres carrés avec une TI

Avec les mêmes réglages que ci-dessus pour le nuages de points, et après avoir calculé les coefficient a et b qui donnent l'équation de la droite (voir ci-dessus).

- Appuyer sur la touche $f(x)$ en haut à gauche
- Se placer sur $Y1=$
- Appuyer sur *var*, choisir 5 : *Statistiques*
- Sélectionner l'onglet *EQ*, puis la ligne *EqReg*
- La calculatrice revient automatiquement à l'écran " $Y1 = \dots$ ", et a écrit l'équation de la droite des moindres carrés sur la ligne $Y1$.
- Appuyer sur *graphe*



Pour les calculatrices Casio

- Sélectionner le menu *Stat*
- Entrer les données x_i en *List1*, et les données y_i en *List2*
- Vérifier le réglage de la calculatrice : *CALC*, puis *SET* : *List1* doit être écrit devant *2Var XList*, et *List2* doit être écrit devant *2Var YList*
- Appuyer deux fois sur *EXIT* pour revenir aux listes
- Pour obtenir les coefficient a et b de l'équation de la droite : *CALC*, *REG*, *X* puis $ax + b$.

Pour les calculatrices Numworks

- Sélectionner le menu *Regression*
- Entrer les données x_i en *X1*, et les données y_i en *Y1*
- Aller dans l'onglet *Graph*, vous obtenez le graphique, et en-dessous, les valeurs de $\bar{x}$, $\bar{y}$ (point moyen), a , et b (coefficients de l'équation $ax + b$ de la droite des moindres carrés).

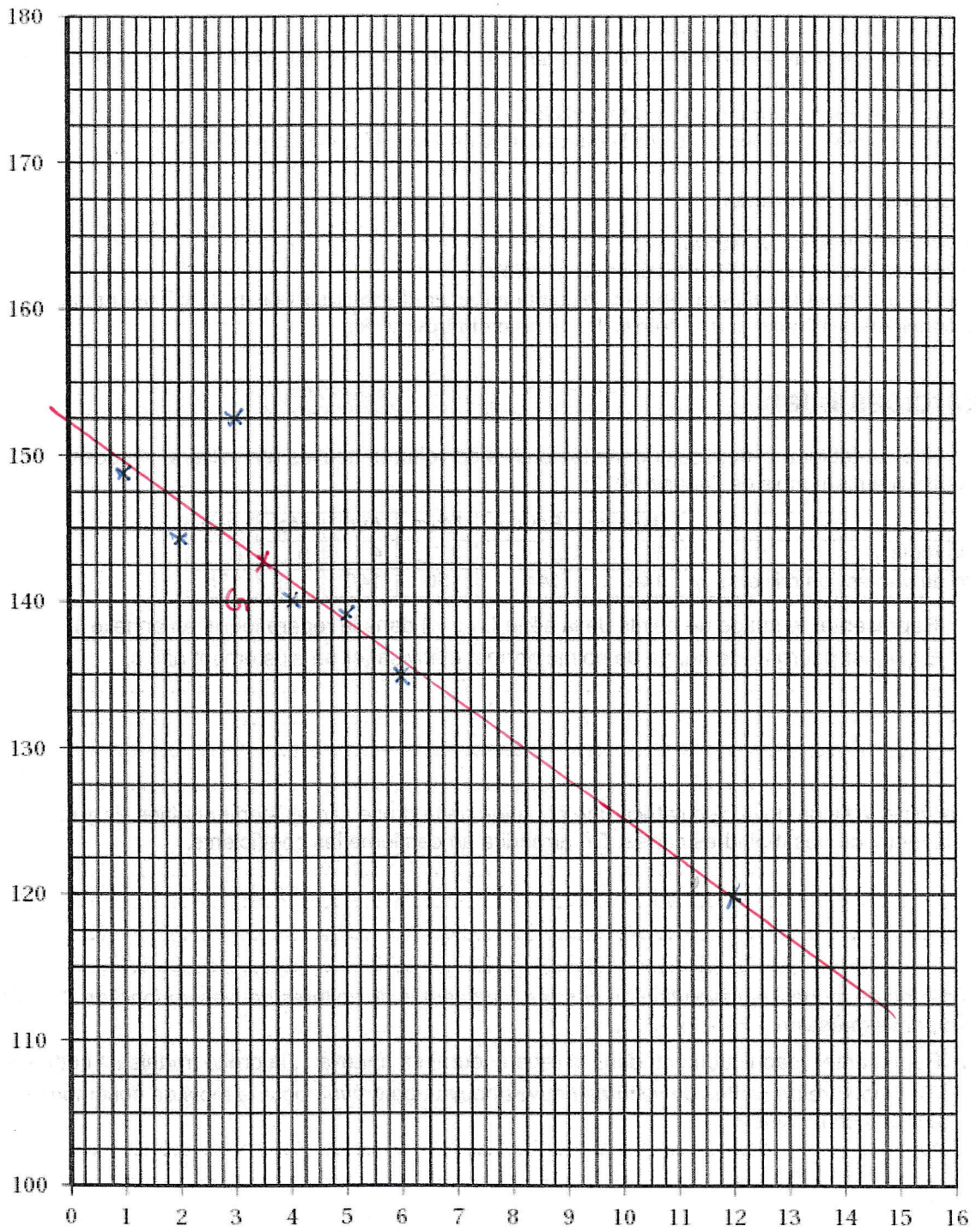
3. Exercices bilan

Cfns. Exercices, juin 2016, ex 2 -
Exercice 12.8 Le tableau ci-dessous donne le nombre de voitures neuves (en milliers) vendues en France durant les six premiers mois de l'année 2013.

Mois	Janvier	Février	Mars	Avril	Mai	Juin
Rang du mois x_i	1	2	3	4	5	6
Nombre de ventes (en milliers) y_i	149	144	150	140	139	135

- Représenter le nuage de points de la série $(x_i ; y_i)$ dans le repère fourni en annexe 1.
 - Expliquer pourquoi ce nuage de points permet d'envisager un ajustement affine.
les points semblent être répartis autour d'une droite
(à peu près alignés, sauf un)
- Déterminer à l'aide de la calculatrice une équation de la droite D d'ajustement affine de y en x obtenue par la méthode des moindres carrés. On arrondira au centième les coefficients.
 $y = -2,71x + 152,33$ $\begin{pmatrix} \bar{x} = 3,5 \\ \bar{y} = 142,83 \end{pmatrix}$
- On décide de modéliser l'évolution du nombre y de ventes de voitures neuves en fonction du rang x du mois par l'expression $y = -2,7x + 152$.
 - Représenter graphiquement, dans le repère fourni en annexe 1, la droite traduisant cette évolution.
 - Quel nombre de ventes de voitures neuves pouvait-on prévoir pour le mois de décembre 2013 en utilisant ce modèle ?
Pour $x = 12$, $y \approx 120$ donc 120 000 voitures
 - À partir de quel mois pouvait-on prévoir que le nombre de voitures neuves en France serait strictement inférieur à 130000 véhicules ?
*Par lecture graphique, ce sera le cas pour $x = 9$,
à partir de septembre 2013*

Annexe 1 :



Polygone juin 2014, ex 3.

Exercice 12.9 Une entreprise de livraison de colis à domicile demande à un cabinet comptable de réaliser une étude sur son activité.

Une partie des données concerne les bénéfices (en milliers d'euros) réalisés chaque année depuis 2007.

Ces informations sont résumées dans le tableau ci-dessous.

Année	2007	2008	2009	2010	2011	2012	2013	2014	2015
Rang de l'année : x_i	1	2	3	4	5	6	7	8	9
Bénéfice en milliers d'euros : y_i	10,2	12,8	13,8	14,4	16,7	17,5			

Les données du tableau ci-dessus sont représentées par le nuage de points en **annexe 1** à rendre avec la copie.

- À l'aide de la calculatrice, déterminer pour cette série statistique une équation de la droite d'ajustement de y en x obtenue par la méthode des moindres carrés. Arrondir les coefficients à 0,01 près.

La droite des moindres carrés est $y = 1,39x + 9,35$

- Pour les deux questions suivantes, on prendra comme ajustement affine la droite d'équation $y = 1,4x + 9,4$.

(a) Tracer cette droite sur l'annexe 1 de l'exercice.

(b) On suppose que cet ajustement restera valide jusqu'en 2015. Déterminer le bénéfice en euros que l'on peut prévoir pour l'année 2015.

En 2015, $x = 9$. On peut compléter les colonnes du tableau par le rang de l'année 2015, d'où $y = 1,4 \times 9 + 9,4 = 22$.

Le bénéfice prévu en 2015 est de 22 milliers d'euros (22 000 €).

Annexe 1 :

