Chapitre PS1 : Statistiques

1 – Série statistique à une variable

● On étudie un seul **caractère** (taille, poids, notes, …) au sein d'une **population** (les français, la classe, …)

● On considère une **série statistique** prenant p **valeurs** différentes $x_1, \dots, x_p$, rangées dans l'ordre **croissant,** avec les **effectifs** $n_1, \dots, n_p$. On note N **l'effectif total** : $N = n_1 + \dots + n_p$.

● Une série statistique peut se résumer par quelques valeurs appelées **paramètres de la série** :

. **Moyenne** : $\bar{x} = \frac{n_1 x_1 + \dots + n_p x_p}{N}$

. **Ecart-type** : $\sigma = \sqrt{V}$ où $V = \frac{n_1(x_1-\bar{x})^2 + \dots + n_p(x_p-\bar{x})^2}{N}$

. **Médiane :** Med = Valeur qui sépare la série en deux parties égales.

. **1e Quartile :** Q_1 = Valeur située au quart (25 %) de la série.

. **3e Quartile :** Q_3 = Valeur située au trois-quarts (75 %) de la série.

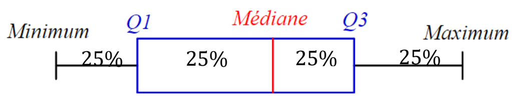


Diagramme en boîte

● L'ensemble de ces paramètres peuvent être obtenues en entrant la série dans la calculatrice.

Exemple 1 : Voici les notes d'une classe de 20 élèves pour un DS de maths.

Données : 14 ; 8 ; 8 ; 10 ; 9 ; 14 ; 10 ; 11 ; 11 ; 11 ; 12 ; 9 ; 8 ; 13 ; 9 ; 11 ; 9 ; 10 ; 14 ; 13 ;

1) Compléter le tableau d'effectifs et fréquences suivant :

Valeurs (x_i)	8	9	10	11	12	13	14	Total (N)
Effectifs (n_i)	3	4	3	4	1	2	3	20
Fréquences (f_i) (en %)	15%	20%	15%	20%	5%	10%	15%	100%

2) Représenter la série à l'aide d'un nuage de point.

3) Calculer la moyenne de la classe :

$\bar{x} = \frac{3\times8+4\times9+\dots+3\times14}{20} = 10{,}7$

4) Déterminer les paramètres statistiques de la série, à l'aide de la calculatrice :

. Moyenne : $\bar{x} = 10.7$

. Ecart-type : $\sigma = 2.00$

. 1er Quartile : $Q_1 = 9$

. Médiane : $Med = 10.5$

. 3e Quartile : $Q_3 = 12.5$

5) Réaliser le diagramme en boite de la série.

Nuage de points :

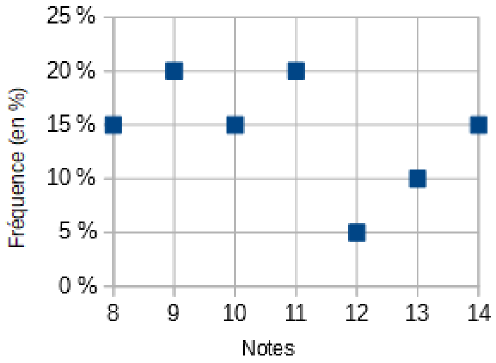
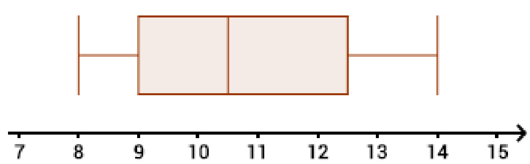


Diagramme en boîte :



2 – Série statistique à deux variables

a. Nuage de points, Point moyen

● On étudie simultanément deux **caractères** au sein d'une même **population**.

● Une **série statistique à deux variables** est constituée de deux listes de n **valeurs** $x_1, \dots, x_n$ et $y_1, \dots, y_n$

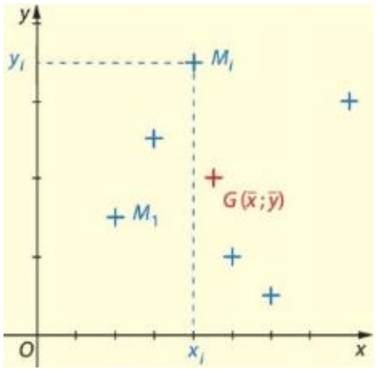
Définition :

● Une série statistique à deux variables se représente par un **nuage de points** de coordonnées $M_i(x_i; y_i)$

● On appelle **point moyen** du nuage le point $G(\bar{x}; \bar{y})$ où :

. $\bar{x}$ est la moyenne de la 1e variable : $\bar{x} = \frac{x_1+\cdots+x_n}{n}$

. $\bar{y}$ est la moyenne de la 2e variable : $\bar{y} = \frac{y_1+\cdots+y_n}{n}$



Exemple 2 : Le tableau suivant donne la superficie et le prix de dix appartements anciens vendus récemment dans le centre d'une petite ville

Superficie (en m²) : (x_i)	32	36	38	42	45	65	70	80	90	110
Prix de vente (en K€) : (y_i)	33	37	40	43	45	66	68	78	85	105

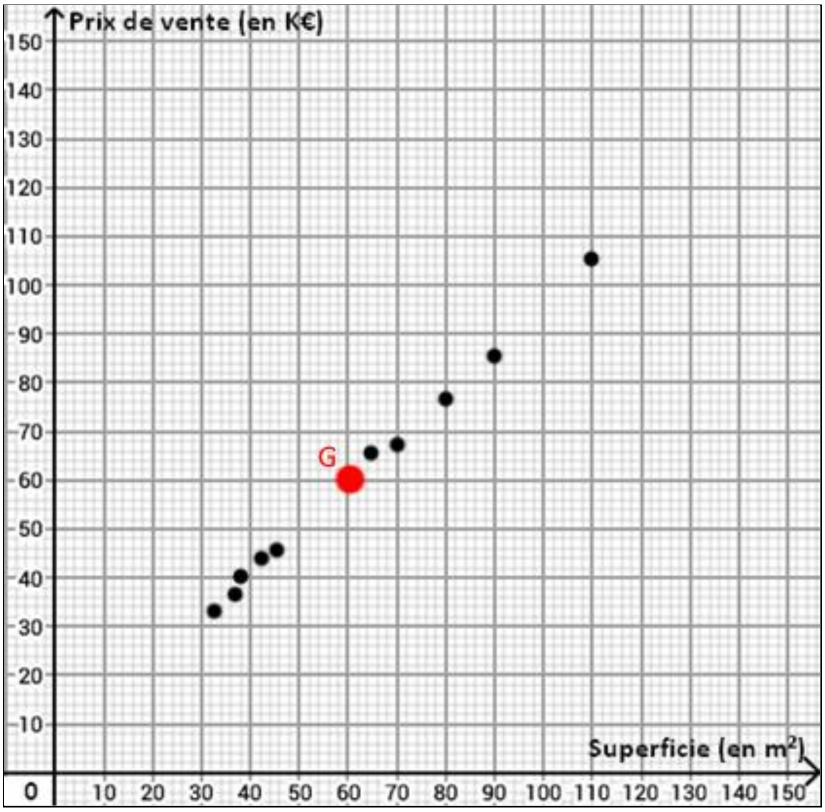
1) Représenter dans le repère ci-dessous, le nuage de points $M_i(x_i; y_i)$ associé à la série ci-dessus.

2) Calculer les coordonnées du point moyen G du nuage, puis placer le dans le repère.

$\bar{x} = \frac{32+36+\cdots+110}{10} = 60.8$ et $\bar{y} = \frac{33+37+\cdots+105}{10} = 60$ donc le point moyen est $G(60.8; 60)$

3) Commenter la forme du nuage de points.

Les points sont presque alignés.



b. Ajustement affine

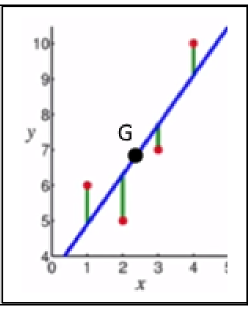
Lorsque les points sont presque alignés, on peut approcher ces points à l'aide d'une **droite d'ajustement**.

Définition :

● La droite d'ajustement qui passe au plus près des points du nuage est la droite de Régression $\mathcal{D}$ d'équation $y = ax + b$ obtenue par la méthode des **moindres carrés :**

On minimise la somme des carrés des distances entre les points du nuage et la droite.

● Cette droite $\mathcal{D}$ passe par le point moyen $G(\bar{x}; \bar{y})$ du nuage.



Remarque :

● Pour déterminer les coefficients a et b de l'équation de cette droite, on utilise le tableur ou la calculatrice.

● Un ajustement permet de faire des estimations :

. Une **interpolation** dans l'intervalle des valeurs connue.

. Une **extrapolation** à l'extérieur de cet intervalle.

Exemple 2 : (Suite)

4) a. A l'aide de la calculatrice déterminer l'équation de la droite $\mathcal{D}$ d'ajustement obtenue par la méthode des moindres carrés (On arrondira les coefficients au centième).

A l'aide de la calculatrice on trouve $a = 0.91$ et $b = 4.62$ donc $\mathcal{D}$ a pour équation $y = 0.91x + 4.62$

b. Tracer cette droite dans le repère.

On calcule les coordonnées de deux points de la droite :

x	0	150
y	4.62	141.12

Si $x = 0$ alors $y = 4.62$ (ordonnée à l'origine). On place $A(0; 4.62)$

Si $x = 150$ alors $y = 0.91 \times 150 + 4.62 = 141.12$. On place $B(150; 141.12)$.

5) Vérifier que cette droite passe par le point moyen du nuage.

$0.91 \times x_G + 4.62 = 0.91 \times 60{,}8 + 4.62 \approx 60 = y_G$ donc $G \in \mathcal{D}$.

6) Réaliser les estimations suivantes en précisant si il s'agit d'une interpolation ou d'une extrapolation :

a. Le prix d'un appartement de 130 m^2 (arrondi au K€ près).

On a $x = 130$ donc $y = 0.91 \times 130 + 4.62 \approx 123$.

Le prix d'un appartement de 130 m^2 est estimé à 130 000 €.

$130 \notin [32; 110]$ donc il s'agit d'une extrapolation.

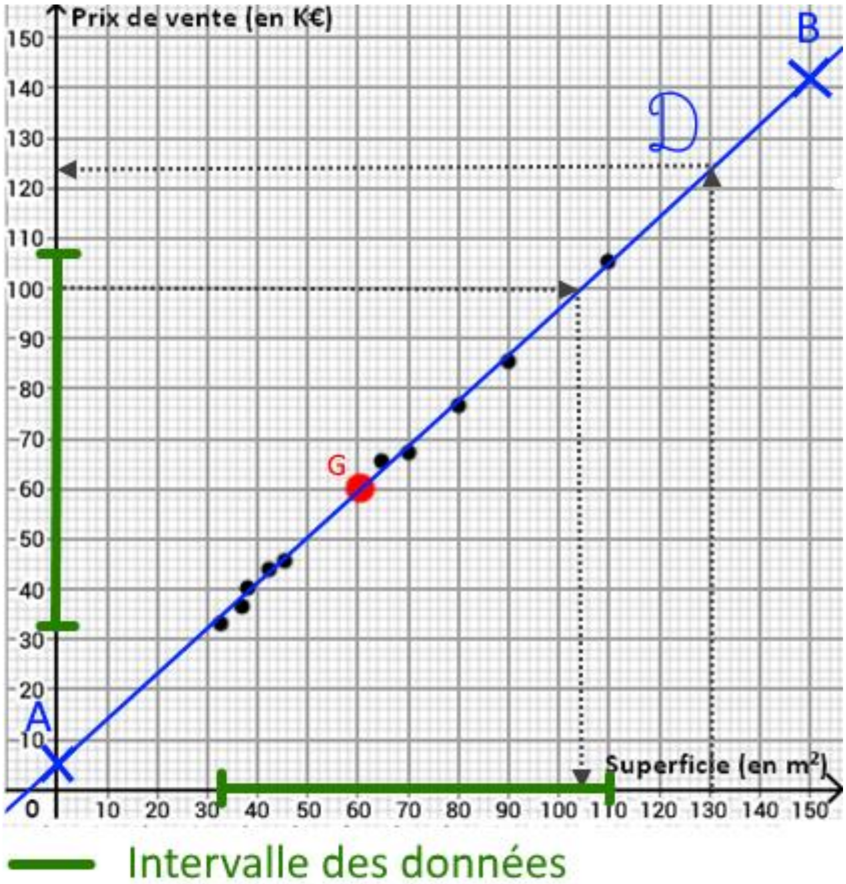
b. La surface d'un appartement coutant 100 000 €.

On a $y = 100$ donc on doit résoudre l'équation :

$$100 = 0.91x + 4.62 \Leftrightarrow 100 - 4.62 = 0.91x \Leftrightarrow 0.91x = 95.38 \Leftrightarrow x = \frac{95.38}{0.91} \approx 105$$

La surface d'un appartement coutant 100 000 € est estimée à 105 m^2

$100 \in [33; 105]$ donc il s'agit d'une interpolation.


Prix de vente (en K€)
Superficie (en m²)
A
B
G
D
Intervalle des données

Statistiques & Calculatrices – Casio 35+

1) Sélection du menu « Statistiques »

Appuyer sur la touche MENU puis sélectionner le menu STAT

2) Statistiques à une variable

- Entrer la série statistique (appuyer sur EXE pour valider) :
  - . $List\ 1$: Valeurs de la série
  - . $List\ 2$: Effectifs (Facultatif)
- Configurer la calculatrice : Appuyer sur F2 (CALC) puis sur F6 (SET) :
  - . $1Var\ XList : List1$ F2 (LIST) / 1
  - . $1Var\ Freq$:
    - → Si une simple liste de valeurs a été saisie : 1 ; Touche F1
    - → Si les effectifs sont dans la Liste 2 : $List2$; Touches F2 (LIST) / 2
  - . Puis, appuyer sur EXIT.
- Affichage des mesures statistiques : Appuyer F1 (1VAR) :
  - . $\bar{x}$: Moyenne . $x\sigma n$: Ecart-type . n : Effectif total
  - . Q_1 : 1er Quartile . Med : Médiane . Q_3 : 3e Quartile

3) Statistiques à deux variables

- Entrer la série statistique (appuyer sur EXE pour valider) :
  - . $List\ 1$: Valeurs de la 1e variable.
  - . $List\ 2$: Valeurs de la 2e variable.
- Configurer la calculatrice : Appuyer sur F2 (CALC) puis sur F6 (SET) :
  - . $2Var\ XList : List1$ F2 (LIST) / 1
  - . $2Var\ YList :\ List2$ F2 (LIST) / 2
  - . Puis, appuyer sur EXIT.
- Droite de régression : F3 (REG) puis F1 (X) puis F1 (ax+b) :
  - . a : Coefficient de directeur . b : Ordonnée à l'origine
  - . r et r^2 : Coefficient de corrélation et son carré.

Statistiques & Calculatrices – TI 82-83

1) Sélection du menu « Statistiques »

Appuyer sur la touche STAT puis sur la touche 1 pour sélectionner $1: Editer$.

2) Statistiques à une variable

- Entrer la série statistique (appuyer sur ENTER pour valider) :
  - . $List\ 1$: Valeurs de la série
  - . $List\ 2$: Effectifs (Facultatif)
- Choix de la fonction : Appuyer sur STAT, puis sur la touche ▶ pour afficher l'onglet $CALC$, puis sur la touche 1 pour sélectionner $Stats\ 1 - Var$:
  - . Si une simple liste de valeurs a été saisie : $Stats\ 1 - Var\ L_1$
    Touches 2nd / 1
  - . Si les effectifs sont dans la Liste 2 : $Stats\ 1 - Var\ (L_1, L_2)$
    Touches (puis 2nd / 1 puis , puis 2nd / 2 puis)
- Affichage des mesures statistiques : On appuie sur ENTER
  - . $\bar{x}$: Moyenne . σx : Ecart-type . n : Effectif total
  - . Q_1 : 1er Quartile . $Méd$: Médiane . Q_3 : 3e Quartile

3) Statistiques à deux variables

- Entrer la série statistique (appuyer sur ENTER pour valider) :
  - . $List\ 1$: Valeurs de la 1e variable.
  - . $List\ 2$: Valeurs de la 2e variable.
- Choix de la fonction : Appuyer sur STAT, puis sur la touche ▶ pour afficher l'onglet $CALC$, puis sur la touche 4 pour sélectionner $RegLin(ax + b)$
- Droite de régression : On appuie sur ENTER
  - . a : Coefficient de directeur . b : Ordonnée à l'origine

Statistiques – Fiche d'exercices

Ex 1 Lors d'une séance de travaux pratique, on a relevé le rythme cardiaque d'une classe de 1$^{\text{ère}}$ au repos, puis après un effort physique.

Au repos

81	73	73	86	69
85	77	76	71	78
82	94	89	91	84
88	71	76	70	91
84	86	83	76	72
77	74	70	86	85

Avec effort

109	94	105	111	89
127	109	126	111	99
128	114	106	119	130
105	107	121	103	139
135	121	89	101	124
112	105	89	103	121

1. A l'aide de votre calculatrice, afficher les principaux paramètres de ces deux séries statistiques.
2. Comparer ces deux séries statistiques.

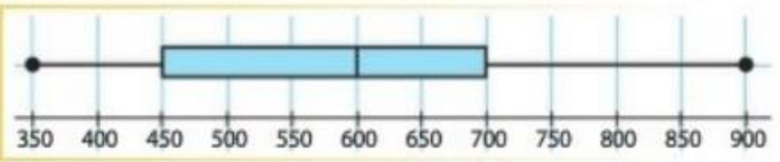
Ex 2 **(Pointure de chaussure)**

1) Construire une série statistique regroupant les pointures de chaussures des élèves de la classe puis regrouper les valeurs dans un tableau d'effectifs.

2) A l'aide de la calculatrice, afficher les principaux paramètres de cette série statistique.

3) Représenter cette série statistique (Nuage de points et diagramme en boîte).

Ex 3 Le diagramme en boîte ci-dessous donne le budget moyen mensuel, en euros, consacré à l'alimentation par 750 familles de quatre personnes.

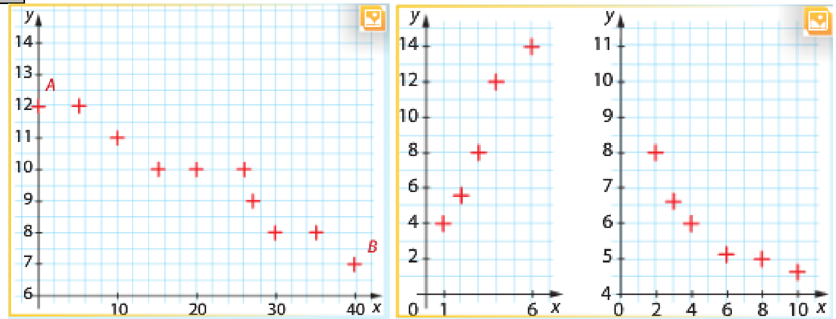


1. Lire les paramètres de cette série statistique indiqués par le diagramme.
2. Quelle est la part de ces familles dont le budget mensuel est compris entre 600 et 700 € ?

Ex 4 De 2000 à 2008, le PIB des Etats-Unis est passé de 9,9 milliards de dollars à 14,2 milliards de dollars. On note x le rang de l'année par rapport à 2000 et y la valeur du PIB en milliard de dollars.

1) Dans un repère orthonormée, placer les points $A(0; 9{,}9)$ et $B(8; 14{,}2)$
2) Calculer le coefficient directeur de la droite (AB).
   Que représente-t-il pour la droite (AB) ?
3) Déterminer l'équation de la droite (AB).
4) A l'aide de cette droite, estimer le P.I.B des Etats-Unis en 2006.

Ex 5 On donne les trois nuages de points suivants



1) Pour quel(s) nuage(s) un ajustement affine serait-il un bon ajustement ?
2) Pour le 1$^{\text{er}}$ nuage on utilise la driote (AB) comme ajustement affine. Déterminer l'équation de cette droite d'ajustement.
3) Pour le 2$^{\text{e}}$ nuage, donner une équation de la droite d'ajustement obtenue par la méthode des moindres carrés.
4) Pour le 3$^{\text{e}}$ nuage, déterminer les coordonnées du point moyen $G(\bar{x}; \bar{y})$.

Ex 6 On considère la série statistique ci-dessous, où x est le prix en euros par kg de fraises et y la quantité vendue, en centaines de kg, au prix x

prix x_i	5	5,7	7,1	8,6	9	9,5	12
quantité y_i	12	11	9	7,2	6,5	6	2

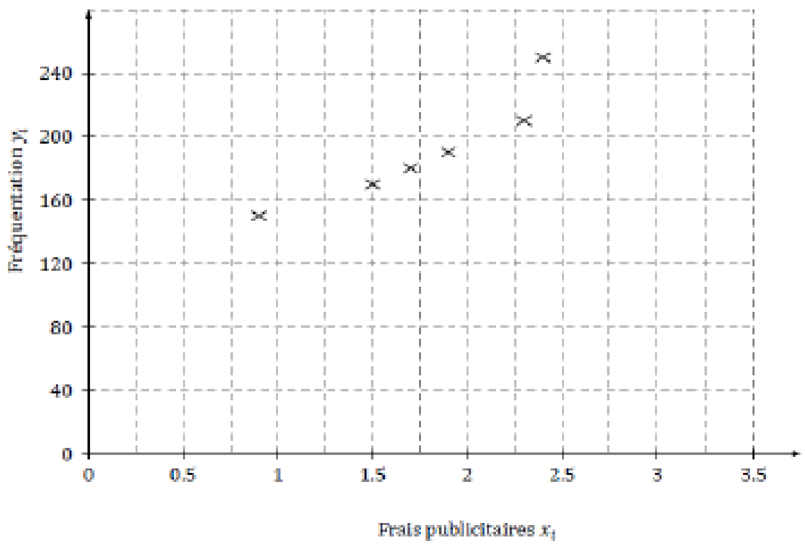
1) Est-on dans une situation de proportionnalité ?
2) Représenter cette série double par un nuage de points $M_i(x_i; y_i)$.
3) Calculer les coordonnées du point moyen $G(\bar{x}; \bar{y})$ et placer ce point.
4) Déterminer l'équation de la droite d'ajustement obtenue par la méthode des moindres carrés, puis tracer cette droite.
5) Estimer la quantité vendue pour un prix de 15€ le kilogramme.
6) Calculer la recette engendrée par la vente de $\bar{y}$ centaines de fraises au prix moyen $\bar{x}$ par kg.

Ex 7 Frais publicitaires (Tiré du bac Pondichery 2017)

Le service marketing d'un centre commercial veut évaluer l'impact des frais engagés en publicité, par mois, sur le nombre de clients.
Pour cela, ce service s'appuie sur les données ci-dessous, relevées sur une période de 6 mois :

Frais publicitaires x_i (en milliers d'euros)	1,9	2,4	1,5	0,9	2,3	1,7
Fréquentation y_i (en milliers de clients)	190	250	170	150	210	180

Le nuage de points de coordonnées (x_i, y_i) est représenté ci-dessous.



1. Donner à l'aide de la calculatrice une équation de la droite réalisant un ajustement affine de ce nuage de points, obtenue par la méthode des moindres carrés.
   On arrondira les coefficients au centième.

2. On décide d'ajuster ce nuage de points par la droite d'équation $y = 58{,}3x + 87{,}6$.
   a. On estime alors que pour 4 000 euros de frais publicitaires engagés, la fréquentation s'élèverait à 321 000 clients. Vérifier la cohérence de l'estimation annoncée.
   b. Quel est le montant des frais publicitaires devant être engagés pour espérer 400 000 clients au cours d'un mois ?
   On arrondira à la centaine d'euros.
   c. Le centre commercial décide d'engager 5 000 euros pour la campagne publicitaire du prochain mois. Lors du bilan, on dénombre 330 000 clients ayant fréquenté le site au cours de ce mois. Comment peut-on analyser ce résultat ?

Ex 8 Evolution du P.I.B (Tiré du bac Polynésie 2015)

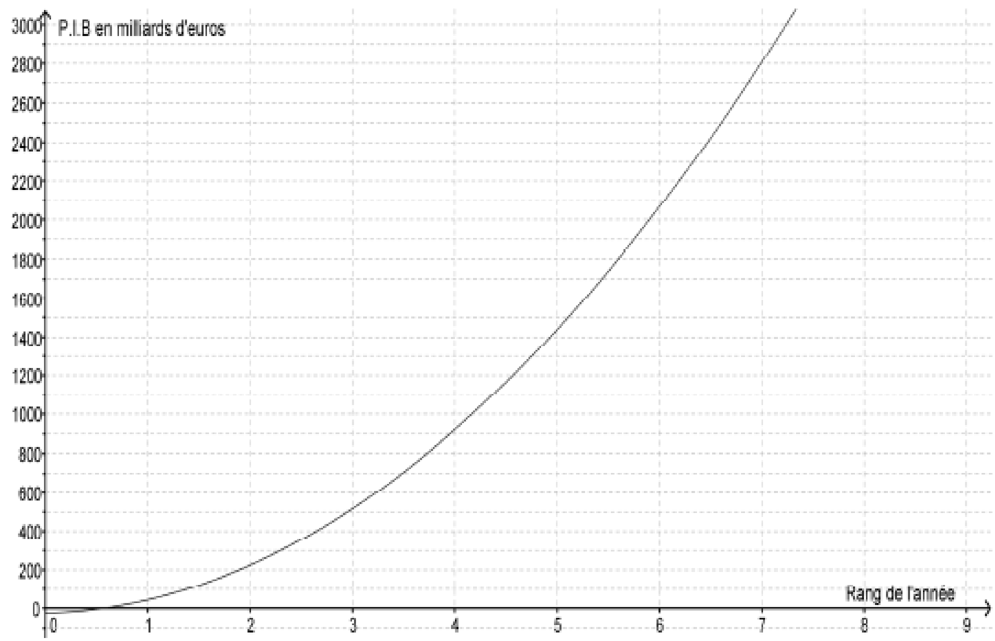
On s'intéresse aux évolutions décennales du P.I.B. en France de 1950 à 2010.

Années	1950	1960	1970	1980	1990	2000	2010
rang de l'année x_i	0	1	2	3	4	5	6
P.I.B. en milliards d'euros y_i	15,5	47,0	126,1	453,2	1 058,6	1 485,3	1 998,5

Source : Comptes nationaux - Base 2010, Insee

1. Dans le graphique en **annexe page 6 à rendre avec la copie**, représenter le nuage de points de coordonnées $(x_i ; y_i)$ pour i variant de 0 à 6.
2. Donner une équation de la droite d'ajustement affine de y en x obtenue par la méthode des moindres carrés en se limitant à la période 1970-2010.
3. On ajuste l'ensemble du nuage avec la droite (D) d'équation $y = 478x - 886$.
   Tracer cette droite sur le graphique en **annexe page 6 à rendre avec la copie.**
4. On se propose d'ajuster ce nuage de points par la parabole, tracée sur le graphique en annexe, d'équation $y = 56x^2 + 12{,}6x - 25$.

   Donner une estimation du P.I.B. en 2020 par la méthode qui vous semble la plus adaptée.



Ex 9 Consommation d'énergie primaire fossile (Tiré du bac Métropole 2018)

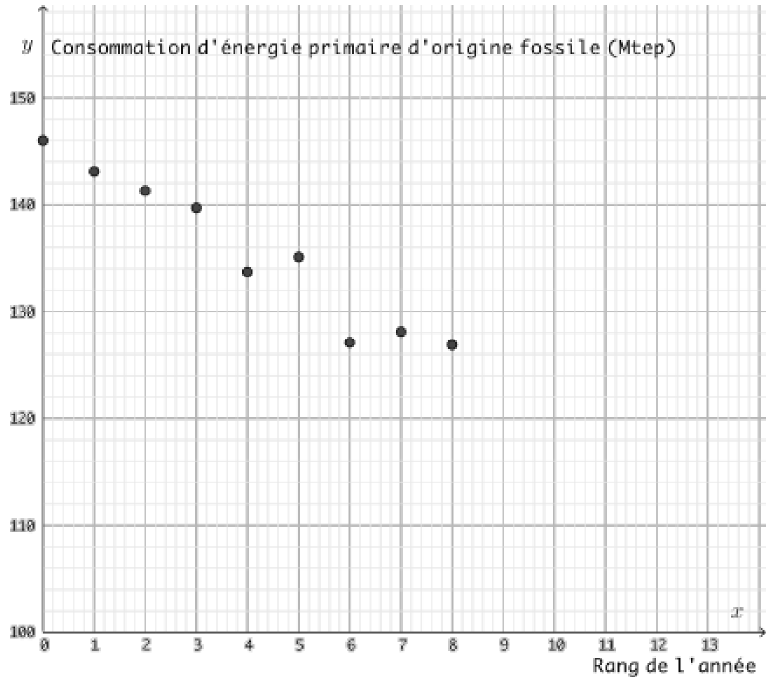
Le tableau ci-dessous donne la consommation d'énergie primaire d'origine fossile (charbon, gaz, pétrole) en France entre 2005 et 2013. Elle s'exprime en million de tonnes équivalent pétrole (Mtep) et est arrondie au dixième.

Année	2005	2006	2007	2008	2009	2010	2011	2012	2013
Rang de l'année : x_i	0	1	2	3	4	5	6	7	8
Consommation d'énergie primaire d'origine fossile (en Mtep) : y_i	146,0	143,1	141,3	139,7	133,7	135,1	127,1	128,1	126,9

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

Une représentation graphique du nuage de points de coordonnées $(x_i\,;y_i)$ est donnée en **annexe**.

1. Donner l'équation réduite de la droite d'ajustement de y en x obtenue par la méthode des moindres carrés. Les coefficients seront arrondis au centième.
2. On décide d'ajuster le nuage de points par la droite $\mathscr{D}$ d'équation $y = -2{,}6x + 146$.
   Tracer la droite $\mathscr{D}$ sur le graphique donné **en annexe, à rendre avec la copie**.
3. La loi de 2015 relative à la transition énergétique fixe à la France l'objectif suivant : avant 2030, réduire de 30 % la consommation en énergie primaire d'origine fossile par rapport à sa valeur en 2012.
   Selon le modèle retenu à la question 2., l'objectif de la loi sera-t-il atteint ? Si oui, au cours de quelle année ? On expliquera la démarche utilisée.



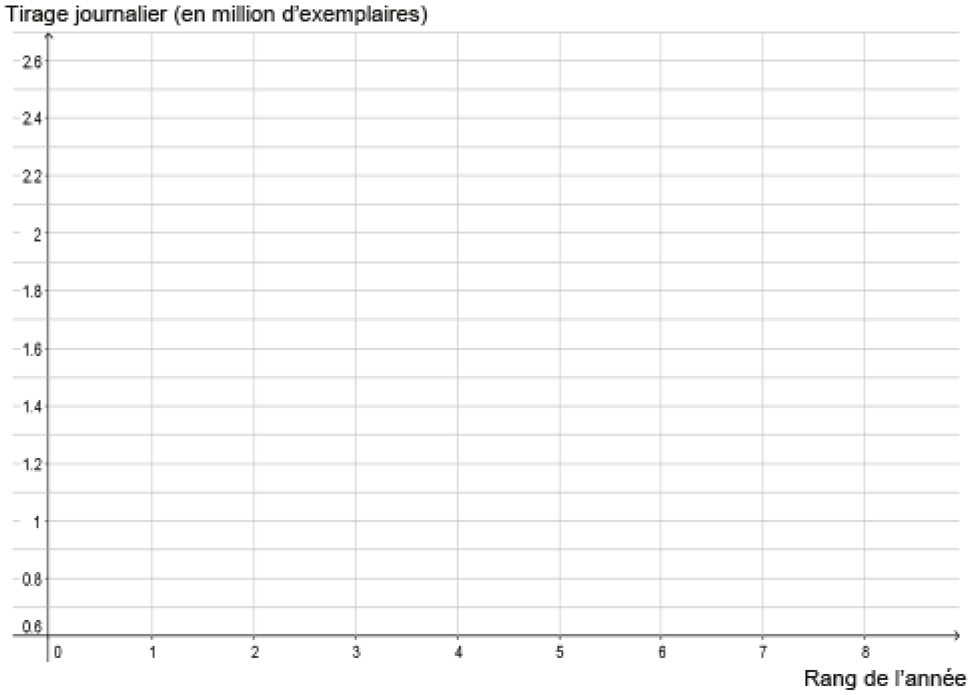
Ex 10 Evolution du tirage de la presse (Tiré du bac Antilles - Guyanne 2018)

Le tableau suivant donne l'évolution du tirage journalier (nombre d'exemplaires imprimés par jour) de la presse quotidienne d'information générale et politique en France.

Année	2010	2011	2012	2013	2014
Rang de l'année : x_i	0	1	2	3	4
Tirage journalier en million d'exemplaires : y_i	1,80	1,73	1,60	1,47	1,36

Source : INSEE

1. Représenter le nuage de points $(x_i\,, y_i)$ associé au tableau ci-dessus dans le repère donné en annexe 1.
2. À l'aide de la calculatrice, déterminer une équation de la droite d'ajustement affine obtenue par la méthode des moindres carrés. Les coefficients seront arrondis à 0,01.
3. Pour les deux questions suivantes, on prendra pour ajustement affine la droite D d'équation $y = -0{,}1x + 1{,}8$.
   a) Représenter la droite D dans le repère donné en annexe 1.
   b) Selon ce modèle, estimer le tirage journalier que l'on peut prévoir pour l'année 2017.



Ex 11 Population française (Tiré du bac Polynésie 2016)

A partir des recensements effectués tous les dix ans, on a établi le tableau suivant qui donne l'évolution de la population française en millions d'individus entre 1851 et 1911. Peu de données sont disponibles pour l'année 1871.

	Population en 1851	Population en 1861	Population en 1881	Population en 1891	Population en 1901	Population en 1911
Rang de la décennie : x_i	0	1	3	4	5	6
Population en millions : y_i	35	37,4	37,7	39,9	39	39,6

Source : INSEE

Partie A : Approximation de la population en 1871.

1. Placer sur le graphique donné en annexe le nuage de points de coordonnées $(x_i ; y_i)$.
2. Donner une équation de la droite d'ajustement affine de y en fonction de x obtenue par la méthode des moindres carrés. Les coefficients seront arrondis au millième.
3. On décide d'ajuster ce nuage de points par la droite (d) d'équation $y = 0{,}7x + 35{,}9$. Tracer cette droite sur ce même graphique.
4. A l'aide de ce modèle, estimer la population en 1871.

Partie B : Évolution de la population après 1911.

1. Les données de l'INSEE (Institut National de la Statistique et des Études Économiques) montrent qu'en 1921 la population française était d'environ 39,2 millions de personnes. Le modèle utilisé dans la partie A prévoyait-il ce résultat ?
2. Sachant qu'en 2011 il y avait 65,2 millions d'habitants en France, pensez-vous que ce modèle reste valable jusqu'à nos jours ? On attend une réponse argumentée.