

# Statistiques

## I Rappels

### 1 - Série statistique à une variable

#### Définition :

Une **série statistique à une variable** ( $x_i$ ) est constituée d'une liste de valeurs  $x_1 ; x_2 ; \dots ; x_p$  avec les effectifs correspondants  $n_1 ; n_2 ; \dots ; n_p$

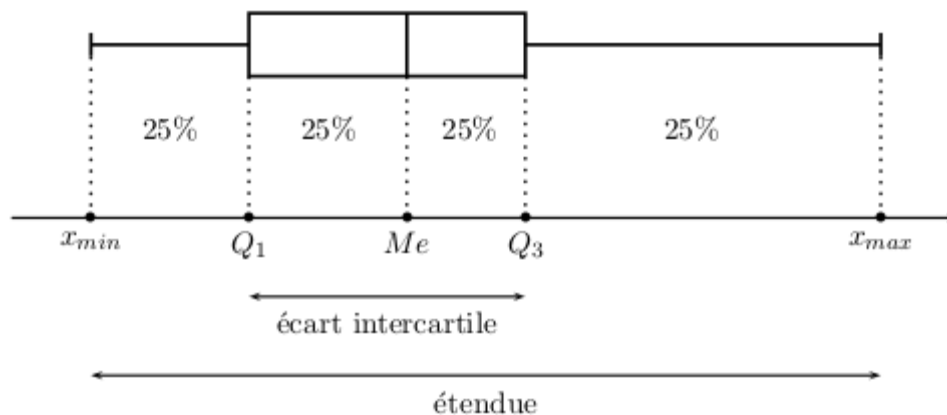
Exemple : Les 31 élèves d'une classe de terminale ont obtenu les notes suivantes à une évaluation de mathématiques.

Valeurs $x_i$ →	Notes	7	8	9	10	11	12	13	14	
	Total	1	5	4	12	5	3	0	1	← Effectifs $n_i$

#### Définition :

Les valeurs de la série étant rangées dans l'ordre croissant, les **quartiles**  $Q_1$  et  $Q_3$  et la **médiane**  $Me$  partagent la série ordonnée en quatre groupes de même effectif.

Schéma : On peut représenter cette répartition par un diagramme en boîte.



#### Paramètres d'une série statistique :

- **La moyenne** :  $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{N}$
- **La médiane**  $Me$  partage la série ordonnée en deux groupes de même effectif.  
Si l'effectif est impair, la médiane est la valeur située au milieu de la série.  
Si l'effectif est pair, la médiane est la moitié de la somme des deux valeurs au milieu.
- **L'étendue** : c'est la différence des valeurs extrêmes :  $e = x_{max} - x_{min}$
- **L'écart-type** mesure la dispersion par rapport à la moyenne. On le note  $\sigma$ .
- **Les quartiles**  $Q_1$  et  $Q_3$
- **L'écart interquartile** [ $Q_1 ; Q_3$ ]

Application : Exercice 1

## 2- Fonction affine et équation réduite de droite

### Définition :

Une **fonction affine**  $f$  est une fonction définie pour tout réel  $x$  par :

$$f(x) = ax + b$$

Le nombre  $a$  est le **coefficient directeur** (ou pente).

Le nombre  $b$  est l'**ordonnée à l'origine**.

### Propriétés :

Si  $a = 0$  alors la fonction est **constante** et s'écrit sous la forme  $f(x) = b$

Si  $b = 0$  alors la fonction est **linéaire** et s'écrit sous la forme  $f(x) = ax$

### Propriété :

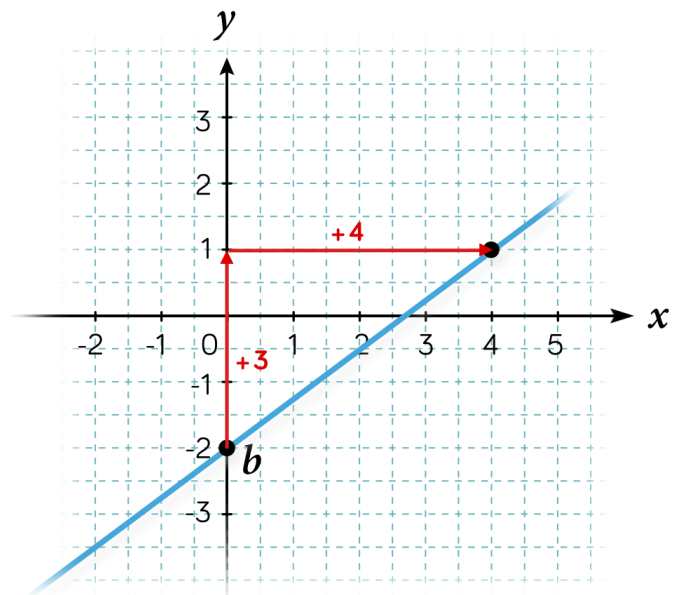
La représentation graphique d'une fonction affine est une droite.

### Méthode : Déterminer graphiquement l'équation de la droite représentative d'une fonction affine

Exemple : Soit  $f$  une fonction affine dont la représentation graphique est donnée ci-contre.

- Pour déterminer  $b$ , il suffit de regarder le point d'intersection entre la droite et l'axe des ordonnées. Ici,  $b = -2$
- Pour déterminer  $a$ , il faut observer le nombre de carreaux dont on se déplace horizontalement et verticalement pour partir de  $b$  et arriver à un point avec des coordonnées « entières ».

$$a = \frac{\text{déplacement vertical}}{\text{déplacement horizontal}} = \frac{3}{4}$$



### Définition :

Soient  $A(x_A; y_A)$  et  $B(x_B; y_B)$  deux points distincts de la droite représentative d'une fonction affine  $f$  telle que  $f(x) = ax + b$ .

Le coefficient directeur  $a$  est égal à :  $a = \frac{f(x_B) - f(x_A)}{x_B - x_A} = \frac{y_B - y_A}{x_B - x_A}$

## Méthode : Déterminer par le calcul l'équation de la droite représentative d'une fonction affine

Soient  $A(1; 2)$  et  $B(6; 8)$  deux points d'une fonction affine  $f$  et  $d$  sa droite représentative.

On veut déterminer par le calcul l'expression de la fonction  $f$ .

- On commence par la valeur du coefficient directeur de la fonction  $f$  :

$$a = \frac{y_B - y_A}{x_B - x_A}$$

$$a = \frac{8 - 2}{6 - 1}$$

$$a = \frac{6}{5}$$

Donc :  $f(x) = \frac{6}{5}x + b$

- On peut ensuite déterminer  $b$  en remplaçant  $x$  et  $f(x)$  dans l'expression de la fonction par les coordonnées d'un point appartenant à la droite  $d$ .

$$f(x_A) = \frac{6}{5}x_A + b$$

$$2 = \frac{6}{5} \times 1 + b$$

$$2 = \frac{6}{5} + b$$

$$2 - \frac{6}{5} = b$$

$$\frac{4}{5} = b$$

Donc :  $f(x) = \frac{6}{5}x + \frac{4}{5}$

Application : Exercice 2

## II Série statistique à deux variables

### 1 - Définitions

**Définition** :

Une **série statistique à deux variables**  $(x_i; y_i)$  est constituée d'une liste de  $n$  couples de valeurs  $(x_1; y_1); (x_2; y_2); \dots; (x_n; y_n)$ .

Le **nuage de points** de cette série est l'ensemble des points du plan de coordonnées  $(x_1; y_1); (x_2; y_2); \dots; (x_n; y_n)$ .

Remarque : Le nuage de points permet de visualiser un lien possible entre deux variables ou de repérer des incohérences dans une série statistique à deux variables.

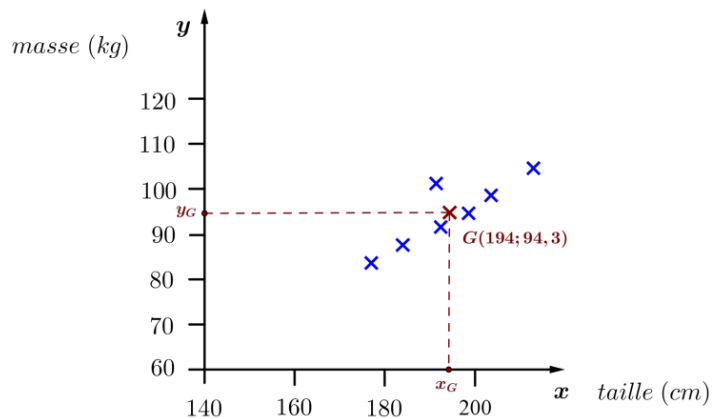
Exemple : Le tableau suivant regroupe la taille et la masse de sept joueurs

Taille (cm)	183	192	177	213	191	199	203
Masse (kg)	88	91	83	105	101	95	97

La représentation graphique est la suivante :

Les coordonnées du point moyen  $G$  sont  $(\bar{x}; \bar{y})$ , moyenne des  $x_i$ , et moyenne des  $y_i$ .

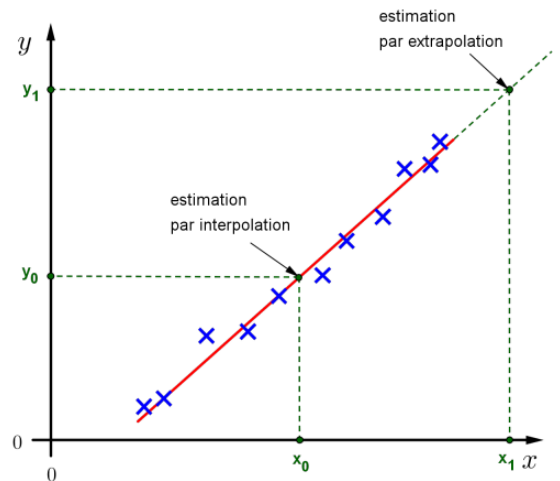
Le nuage de points permet de constater qu'un des joueurs semble être en surcharge pondérale par rapport aux autres.



### Application : Exercice 3

Remarque : Pour estimer une valeur inconnue dans une série statistique, on peut utiliser deux procédés :

- La méthode par interpolation quand le calcul est réalisé dans l'intervalle des valeurs prises par la série statistique.
- La méthode par extrapolation quand le calcul est réalisé en dehors des valeurs prises par la série statistique.



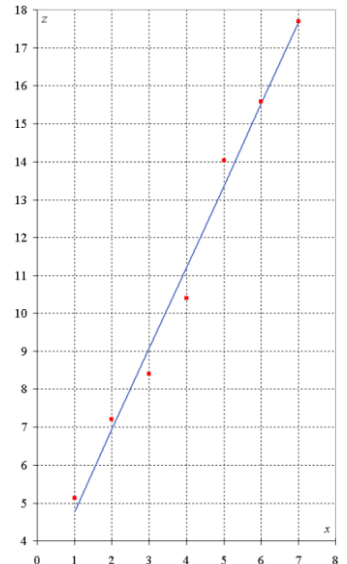
## 2- Droite d'ajustement

### a- Méthode « au jugé »

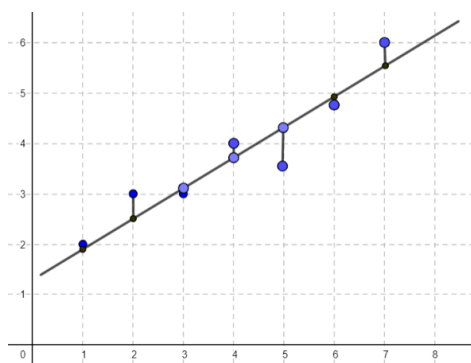
La méthode « au jugé » consiste à tracer une droite passant au plus près des différents points du nuage.

Après avoir tracé cette droite, on peut en déterminer son équation graphiquement ou par le calcul en déterminant les coordonnées de deux points appartenant à la droite (Méthodes précédentes).

### Application : Exercice 4



### b- Méthode des moindres carrés



L'équation de la droite d'ajustement peut être aussi déterminée à l'aide d'une calculatrice ou d'un tableur, par la méthode des moindres carrés.

[Vidéo explicative et tutos Casio/TI](#)

[Tuto Numworks](#)

### C- Méthode des points moyens (ou droite de Mayer)

Cette méthode consiste à déterminer deux points moyens du nuage de points et de déterminer par le calcul l'équation de la droite passant par ces deux points moyens.

Exemple : On considère la série statistique suivante

Superficie $X$	2	2	3	4	5	6	7	7,6
Nombre d'exploitations $Y$	14	26	31	29	44	40	54	50

Les valeurs du caractère  $X$  sont 8 au total. On va donc partager la série en deux séries de 4 valeurs :

#### Série 1

$X$	2	2	3	4
$Y$	14	26	31	29

#### Série 2

$X$	5	6	7	7,6
$Y$	44	40	54	50

Point moyen  $G_1 (\bar{X}_1; \bar{Y}_1)$  :

$$\bar{X}_1 = \frac{2 + 2 + 3 + 4}{4} = 2,75$$

$$\bar{Y}_1 = \frac{14 + 26 + 31 + 29}{4} = 25$$

Point moyen  $G_2 (\bar{X}_2; \bar{Y}_2)$  :

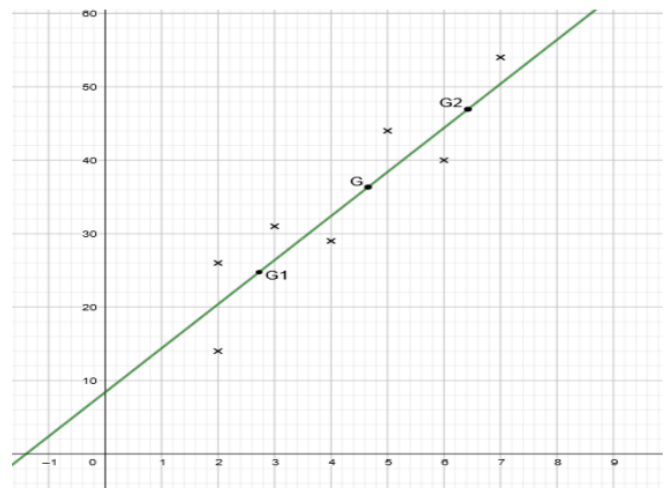
$$\bar{X}_2 = \frac{5 + 6 + 7 + 7,6}{4} = 6,4$$

$$\bar{Y}_2 = \frac{44 + 40 + 54 + 50}{4} = 47$$

On peut ensuite déterminer l'équation de la droite ( $G_1G_2$ ) grâce à la méthode vue précédemment pour déterminer l'équation réduite d'une droite à l'aide des coordonnées des deux points  $G_1$  et  $G_2$  :

$$a = \frac{47 - 25}{6,4 - 2,75} \approx 6 \quad b = 25 - 6 \times 2,75 = 8,4$$

Donc : l'équation de la droite d'ajustement ( $G_1G_2$ ) est  $y = 6x + 8,4$



Application : Exercice 5

### 3- Ajustement par changement de variable

Il arrive que le nuage de points ne soit pas modélisable par une droite. Il faut alors changer de variable pour réaliser un ajustement linéaire.

Exemple : Un centre d'appel comptait 66 employés en 2001. Le tableau ci-dessous donne l'évolution du nombre d'employés en fonction du rang de l'année.

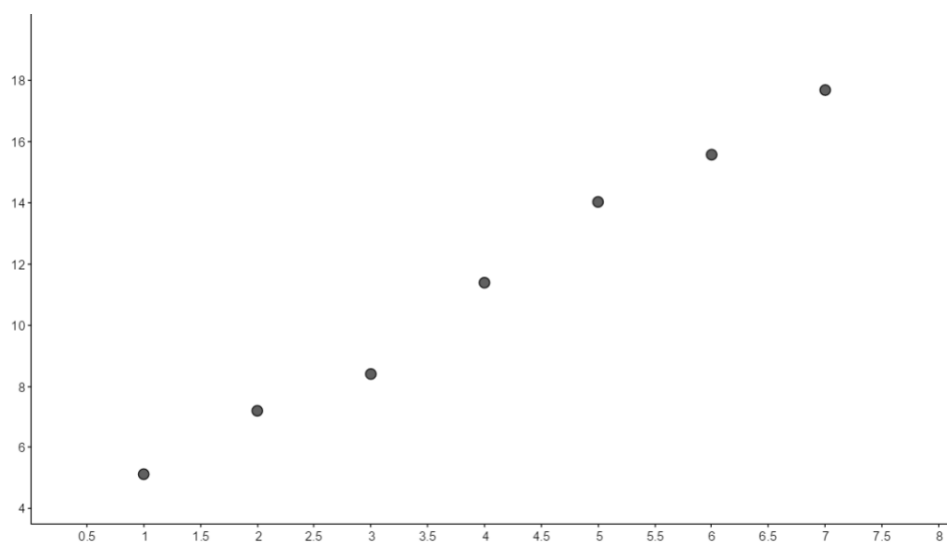
Année	2001	2002	2003	2004	2005	2006	2007
Rang de l'année $x_i$	1	2	3	4	5	6	7
Nombre d'employés $y_i$	66	104	130	207	290	345	428

On cherche à étudier l'évolution du nombre  $y$  d'employés en fonction du rang  $x$  de l'année. Une étude graphique montre qu'un ajustement affine ne convient pas.

On pose alors  $z = \sqrt{y} - 3$

Année	2001	2002	2003	2004	2005	2006	2007
Rang de l'année $x_i$	1	2	3	4	5	6	7
Nombre d'employés $y_i$	66	104	130	207	290	345	428
$z_i$	5,12	7,2	8,4	11,39	14,03	15,57	17,69

On obtient le graphique suivant après le changement de variable. Les points semblent presque alignés, ce qui rend réalisable un ajustement affine (linéaire) par l'une des méthodes vues précédemment.



Application : Exercice 6

# Statistiques (Exercices)

## Exercice 1

On considère les notes suivantes, obtenues à l'épreuve de mathématiques de la dernière session du BTS CG par les 35 candidats d'un centre d'examen.

16,5	13,5	2,5	8,5	17,5
9	16	9,5	10,5	9,5
15	11,5	8,5	6	5,5
6,5	7,5	12	5	7
12,5	7	9,5	5	16
7	16,5	11	11,5	18,5
13,5	15	11,5	15	9

1/ Regrouper et ordonner les valeurs de cette série statistique dans un tableau comme le suivant :

Note	2,5	...
Effectif	1	...

2/ Déterminer la médiane de cette série statistique.

3/ Déterminer la moyenne de cette série statistique (valeur approchée arrondie au dixième près).

4/ Donner la valeur de l'étendue de cette série statistique.

5/ Quel est l'écart interquartile de cette série statistique ?

## Exercice 2

Soit deux points appartenant à la droite représentative d'une fonction affine  $g$  :  $A(0 ; -1)$  et  $B(3 ; 4)$

1/ Déterminer par le calcul l'expression de la fonction  $g$

2/ Tracer la droite  $d$ , représentation graphique de la fonction  $g$

3/ Retrouver graphiquement les valeurs du coefficient directeur et de l'ordonnée à l'origine de la droite  $d$ .

## Exercice 3

1/ Représenter dans un repère le nuage de points de la série statistique regroupée dans le tableau

$x_i$	-3	0	3	5	8
$y_i$	1	-1	-2	5	7

2/ Calculer les coordonnées du point moyen  $G$  et le placer dans le repère.

## Exercice 4

Le tableau ci-dessous présente des données du nombre de personnes malades dans un pays.

Année	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Rang de l'année $x_i$	0	1	2	3	4	5	6	7	8	9	10	11
Nombre de personnes (en milliers) $y_i$	9	11	13	16	20	21	29	35	41	49	57	64

1/ Représenter la série statistique dans un repère avec en abscisse un carreau pour une année, et en ordonnée un carreau pour 10 000 personnes.

2/ Tracer en rouge et « au jugé » une droite d'ajustement du nuage de points.

3/ Calculer l'équation de la droite d'ajustement ainsi tracée.

## Exercice 5

Dans cet exercice, les données sont les mêmes que celles de l'exercice précédent.

1/ Par la méthode des points moyens, déterminer l'équation réduite de la droite d'ajustement linéaire.

2/ Tracer en vert (sur le même repère) les points moyens  $G_1$  et  $G_2$  et la droite  $(G_1G_2)$ .

3/ Quelle est la droite la plus précise ? Justifier

## Exercice 6

On considère les données suivantes représentatives de la distance de freinage en fonction de la vitesse d'un véhicule donné.

Vitesse en km/h $x_i$	0	30	60	90	120	140
Distance de freinage en m $d_i$	0	18	58	120	212	285

1/ Dans un repère, placer les différents points de cette série statistique. On prendra un carreau pour 10 km/h en abscisse et un carreau pour 50 m en ordonnée.

2/ Pourquoi la forme du nuage de points ne permet-elle pas d'envisager un ajustement affine ?

3/ On procède à un changement de variable, soit  $y_i = \sqrt{d_i}$

Reproduire et compléter le tableau suivant :

Vitesse en km/h $x_i$	0	30	60	90	120	140
Distance de freinage en m $d_i$	0	18	58	120	212	285
$y_i = \sqrt{d_i}$						

4/ Déterminer l'équation de la droite de régression (d'ajustement) de  $y$  en fonction de  $x$ .

5/ Estimer par calcul la distance de freinage pour une vitesse de 50 km/h et vérifier la cohérence sur le graphique.

6/ Estimer par calcul la vitesse qui correspond à une distance de 150 m. Vérifier la cohérence sur le graphique.

7/ En déduire de la question 4, l'équation de la droite de régression (d'ajustement) de  $d$  en fonction de  $x$ .

## Exercice 7

Le maire d'une ville a mis en place une politique pour réduire les incivilités sur les voies publiques de sa commune. Durant les 6 dernières années, un bilan a été établi pour comptabiliser le nombre d'incivilités et ces données sont résumées dans le tableau suivant :

Année	Rang de l'année $x_i$	Nombre d'incivilités $y_i$
2011	0	857
2012	1	810
2013	2	720
2014	3	604
2015	4	375
2016	5	273

1/ Le maire annonce à ses concitoyens que sa politique de lutte contre les incivilités a permis de réduire leur nombre de plus de 60% entre 2011 et 2015. A-t-il raison ? Justifier

2/ Placer dans un repère les points de coordonnées  $(x_i ; y_i)$

3/ A l'aide de la calculatrice, donner une équation de la droite qui réalise un ajustement affine du nuage de points de coordonnées  $(x_i ; y_i)$  par la méthode des moindres carrés (arrondir à 0,01 près).

Pour la suite, on prendra comme ajustement affine la droite ( $d$ ) d'équation  $y = -124x + 917$

4/ Tracer la droite ( $d$ ) dans le repère.

5/ Combien d'incivilités ce modèle d'ajustement prévoit-il pour l'année 2018 ?

## Exercice 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 $y_i$	149	144	150	140	139	135

1/ Représenter le nuage de points de la série  $(x_i ; y_i)$  dans un repère.

2/ 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.

3/ On décide de modéliser l'évolution du nombre  $y$  de ventes de voitures neuves en fonction du rang  $x$  du mois par l'expression  $y = -2,7x + 152$ .

a/ Représenter dans le repère précédent la droite traduisant cette évolution.

b/ Quel nombre de ventes de voitures neuves pouvait-on prévoir pour le mois de décembre 2013 ?

c/ A partir de quel mois pouvait-on prévoir que le nombre de voitures neuves en France serait strictement inférieur à 130 000 véhicules ?

## Exercice 9

Dans le cadre de cet exercice, on s'intéresse à la consommation d'électricité en France (exprimée en TWh, c'est-à-dire en milliards de kWh) dans le secteur des transports urbains et ferroviaires pour les années 1994 +  $x_i$ . On décide d'effectuer un ajustement affine.

Année : 1994 + $x_i$	1995	2000	2004	2005	2007
Rang de l'année : $x_i$	1	6	10	11	13
Consommation : $y_i$	8,6	10,4	12,1	11,9	12,2

Source : <http://www.developpement-durable.gouv.fr>

1/ Tracer le nuage de point associé à cette série statistique dans un repère d'unités graphique :

- 1cm pour deux années sur l'axe des abscisses
- 1cm pour un TWh, en commençant à 7 TWh.

2/ Déterminer les coordonnées du point moyen  $G$  du nuage et le placer sur le graphique.

3/ Au moyen de la calculatrice, donner une équation de la droite de régression de  $y$  en  $x$  par la méthode des moindres carrés (arrondir les coefficients à  $10^{-3}$  près).

4/ Pour toute la suite de l'exercice, on utilisera la droite d'équation  $y = 0,31x + 8,43$  comme droite d'ajustement. Tracer cette droite sur le graphique précédent.

*On considère que cette droite fournit un bon ajustement jusqu'en 2015.*

5/ Estimer la consommation d'électricité en France pour l'année 2010.

6/ Estimer à partir de quelle année la consommation d'électricité en France dans le secteur des transports urbains et ferroviaires dépassera 14,5 TWh.