Statistique à deux variables : L'essentiel du cours

Pour une série :

Caractère x_i	x_1	x_2	$\cdots$	x_n
Caractère y_i	y_1	y_2	$\cdots$	y_n

a) Point moyen

Le point moyen G a pour abscisse la moyenne des x_i et pour ordonnée la moyenne des y_i.

b) Droite des moindres carrés

La droite des moindres carrés de y en x :

- a pour équation $y = ax + b$ où a et b sont donnés par la calculatrice

  ▶ pour CASIO :

  `MENU` → `STAT`

  • Entrée des données : rentrer les valeurs x_i dans la liste 1 et les valeurs y_i dans la liste 2.

  • Affichage des résultats : `CALC` → `SET`

  Pour `2Var XList`, choisir `List 1`

  Pour `2Var YList`, choisir `List 2`

  Pour `2VarFreq`, taper `1`

  Choisir `REG`, puis `X`

  On peut lire a et b dans la liste des résultats

  ▶ pour TI :

  • Entrée des données : `stats` `EDIT` `1:Edite` ; rentrer les valeurs x_i dans L1 et les valeurs y_i dans L2.

  • Affichage des résultats : `stats` `CALC` `4:RegLin(ax+b)`

  `Xlist : L1` et `Ylist : L2`, puis `Calculs`

- doit passer par le point moyen.

▶ *Exemple :*

x_i	1	2	3	4	5
y_i	8	9	12	12	14

• Point moyen : $\bar{x} = \dfrac{1+2+3+4+5}{5} = 3$; $\bar{y} = \dfrac{8+9+12+12+14}{5} = 11$

Donc, on a $G\,(3;11)$

• Droite des moindres carrés : La calculatrice donne $a = 1,5$ et $b = 6,5$. Une équation de la droite des moindres carrés est donc : $y = 1,5x + 6,5$

• Estimation de la valeur de y pour $x = 7$: $y = 1,5 \times 7 + 6,5 = 17$

c) Coefficient de corrélation

Le coefficient de corrélation linéaire r :

- s'obtient à la calculatrice ;
- est toujours compris entre -1 et 1 ;
- plus les points sont « regroupés » autour d'une même droite et plus il est proche de 1 pour des données croissantes et de -1 pour des données décroissantes ;
- ne prouve en rien un lien de cause à effet entre les deux variables de la série statistique.

Statistique à deux variables

▶ Exercice n°1

Calculer les coordonnées du point moyen de la série suivante :

x_i	200	205	208	211	215
y_i	5200	5400	5600	5900	6400

▶ Exercice n°2

Déterminer x et y pour que le point moyen de la série soit de coordonnées (7,5 ; 12,6).

x_i	8,2	7,4	x	6,1	9
y_i	15	12,1	16,3	y	12

Rappel :

Détermination à la calculatrice de la droite des moindres carrés :

▶ pour CASIO :

`MENU` → `STAT`

• Entrée des données : rentrer les valeurs x_i dans la liste 1 et les valeurs y_i dans la liste 2.

• Affichage des résultats : `CALC` → `SET`

Pour `2Var XList`, choisir `List 1`

Pour `2Var YList`, choisir `List 2`

Pour `2VarFreq`, taper `1`

Choisir `REG`, puis `X`

On peut lire a et b dans la liste des résultats

▶ pour TI :

• Entrée des données : `stats` `EDIT` `1:Edite` ; rentrer les valeurs x_i dans L1 et les valeurs y_i dans L2.

• Affichage des résultats : `stats` `CALC` `4:RegLin(ax+b)`

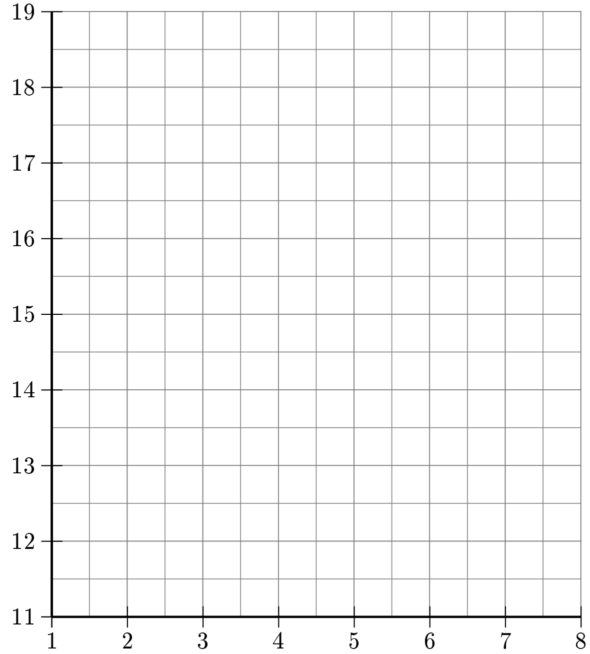
`Xlist : L1` et `Ylist : L2`, puis `Calculs`

▶ Exercice n°3

On considère la série statistique double ci-dessous.

x_i	1	3	4	6	7	8
y_i	11,1	13	14,5	16	16,9	19

1. Représenter le nuage de points $M_i(x_i,y_i)$ dans le repère ci-dessous :



2. Calculer les coordonnées du point moyen et placer ce point dans le repère.
3. Donner une équation de la droite de régression de y en x obtenue par la méthode des moindres carrés et tracer cette droite dans le repère. *(a et b seront arrondis à 0,01 près)*

▶ Exercice n°4

Le tableau suivant donne la moyenne y des maxima de tension artérielle en fonction de l'âge x.

âge x_i	36	42	48	54	60	66	70
tension y_i	12	13	13,6	14,3	15,4	15,8	16

1. Donner une équation de la droite de régression de y en x obtenue par la méthode des moindres carrés. *(a et b seront arrondis à 0,01 près)*
2. Quelle serait la tension théorique d'une personne de 75 ans en utilisant le modèle de la droite des moindres carrés ? *(arrondir le résultat à 0,1 près)*

▶ Exercice n°5

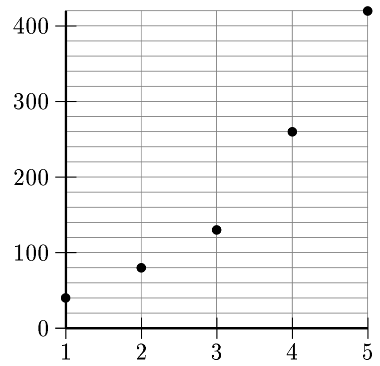
Un nuage de points $A_i(x_i,y_i)$ est tel que $\ln y_i = 2x_i + 5$. Écrire y_i en fonction de x_i. *(on écrira y_i sous la forme $A\,e^{Bx_i}$ en arrondissant A à 0,01 près)*

▶ Exercice n°6

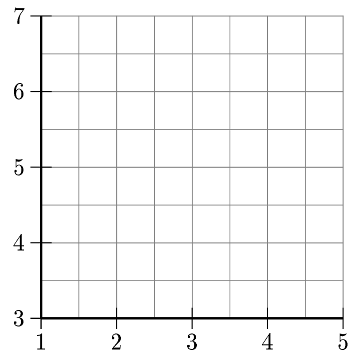
Le tableau suivant donne le nombre d'abonnés à un jeu en ligne.

Année	2014	2015	2016	2017	2018
Rang de l'année x_i	1	2	3	4	5
Nombre d'abonnés y_i (en milliers)	40	80	130	260	420
$z_i = \ln y_i$					

1. Le nuage de points $A_i(x_i,y_i)$ est représenté ci-dessous. Un ajustement affine vous semble-t-il adapté ?



2. Compléter dans le tableau la ligne indiquant $z_i = \ln y_i$. *(arrondir à 0,01 près)*
3. Représenter le nuage de points $B_i(x_i,z_i)$ dans le repère ci-dessous :



4. Donner une équation de la droite de régression de z en x obtenue par la méthode des moindres carrés et tracer cette droite dans le repère. *(a et b seront arrondis à 0,001 près)*
5. En déduire l'expression de y en fonction de x en suivant ce modèle. *(on écrira y sous la forme $A\,e^{Bx}$ en arrondissant A à 0,001 près)*
6. Selon ce modèle, quel serait le nombre d'abonnés en 2022 ? *(donner le résultat au millier près)*

▶ Exercice n°7

Le tableau suivant indique la teneur en CO_2 depuis 1850.

Année	1850	1900	1950	1990
Rang de l'année : x_i	0	50	100	140
Teneur en CO_2 : y_i	275	290	315	360
$z_i = \ln(y_i - 250)$				

1. Compléter dans le tableau la ligne indiquant $z_i = \ln(y_i - 250)$.
2. Donner une équation de la droite de régression de z en x obtenue par la méthode des moindres carrés. *(a et b seront arrondis à 0,01 près)*
3. Selon ce modèle, quel serait le taux de CO_2 en 2050 ? *(donner le résultat à une unité près)*

▶ Exercice n°8

Un étude sur la solubilité d'un médicament en fonction de la température de l'eau a donné les résultats suivants :

Température x_i	20	30	40	50	60	70
Solubilité s_i	10,30	10,59	10,81	11	11,15	11,28
$y_i = e^{s_i}$						

1. Compléter dans le tableau la ligne indiquant $y_i = e^{s_i}$. *(arrondir à l'unité près)*
2. Donner une équation de la droite de régression de y en x obtenue par la méthode des moindres carrés. *(a et b seront arrondis à l'unité près)*
3. En déduire l'expression de la solubilité S en fonction de x en suivant ce modèle.

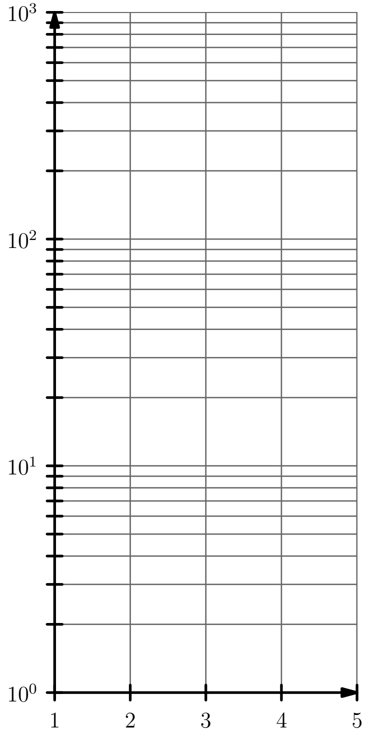
▶ Exercice n°9

On considère la série suivante :

x_i	1	2	3	4	5
y_i	5	20	70	300	800

1. Ces données qui progressent rapidement en ordonnée peuvent difficilement être représentées dans un repère orthogonal classique. Dans ce cas là, on peut utiliser un repère dit semi-logarithmique où un point d'ordonnée y est placé à une « hauteur » $Y = \log y$ (dans l'unité choisie) où log représente le logarithme

décimal.
Placer dans le repère semi-logarithmique ci-dessous le nuage de points A_i.



2. Placer dans ce repère, G, le point moyen de ce nuage de points.
3. Déterminer à la calculatrice le coefficient de corrélation pour les données $(x_i; y_i)$ et comparer le à celui des données $(x_i; \log(y_i))$

Statistique à deux variables

▶ Exercice n°1

$\overline{x} = 207{,}8 \quad ; \ \overline{y} = 700$

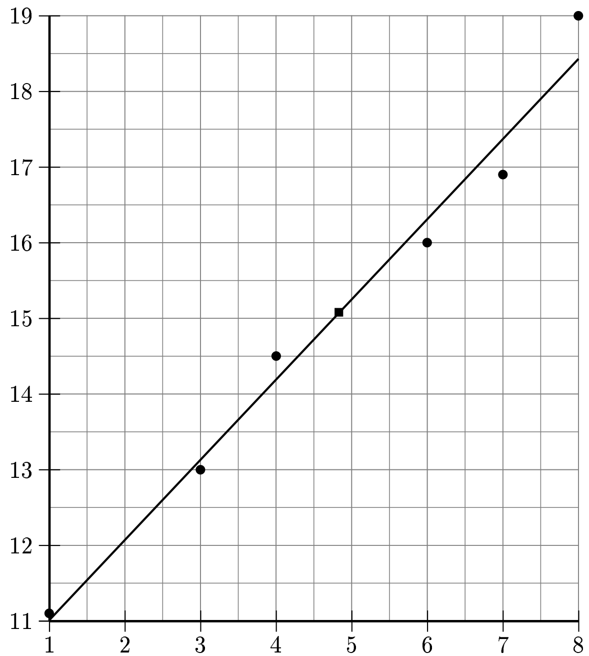
▶ Exercice n°2

On doit avoir $\begin{cases} \dfrac{8{,}2+7{,}4+x+6{,}1+9}{5} = 7{,}5 \\ \dfrac{15+12{,}1+16{,}3+y+12}{5} = 12{,}6 \end{cases}$

$$\Leftrightarrow \begin{cases} \dfrac{30{,}7+x}{5} = 7{,}5 \\ \dfrac{55{,}4+y}{5} = 12{,}6 \end{cases} \Leftrightarrow \begin{cases} 30{,}7+x = 37{,}5 \\ 55{,}4+y = 63 \end{cases} \Leftrightarrow \begin{cases} x = 6{,}8 \\ y = 7{,}6 \end{cases}$$

▶ Exercice n°3

1.



2. $\overline{x} = 4{,}83 \quad ; \ \overline{y} = 15{,}8$

3. $y = 1{,}06x + 9{,}95$

▶ Exercice n°4

1. $y = 0{,}12x + 7{,}84$
2. Si $x = 75$, $y \approx 16{,}8$

▶ Exercice n°5

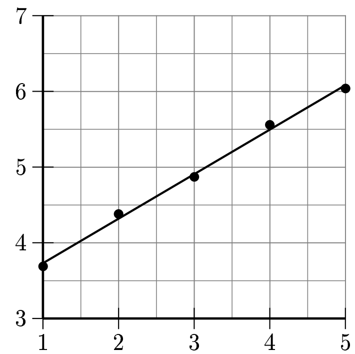
$\ln y_i = 2x_i + 5 \Leftrightarrow y_i = \mathrm{e}^{2x_i+5} = \mathrm{e}^{2x_i} \times \mathrm{e}^5 \approx 148{,}41\,\mathrm{e}^{2x_i}$.

▶ Exercice n°6

1. Non.
2.

Année	2014	2015	2016	2017	2018
Rang de l'année x_i	1	2	3	4	5
Nombre d'abonnés y_i (en milliers)	40	80	130	260	420
$z_i = \ln y_i$	3,69	4,38	4,87	5,56	6,04

3.



4. $z = 0{,}588x + 3{,}144$
5. $\ln y = 0{,}588x + 3{,}144 \Leftrightarrow y = \mathrm{e}^{0{,}588x+3{,}144} = \mathrm{e}^{0{,}588x} \times \mathrm{e}^{3{,}144} \approx 23{,}196\,\mathrm{e}^{0{,}588x}$
6. 2022 correspond à $x = 9$.
   On a alors $y \approx 23{,}196\,\mathrm{e}^{0{,}588\times 9} \approx 4610$ milliers d'abonnés.

▶ Exercice n°7

1.

Année	1850	1900	1950	1990
Rang de l'année : x_i	0	50	100	140
Teneur en CO_2 : y_i	275	290	315	360
$z_i = \ln(y_i - 250)$	3,22	3,69	4,17	4,70

2. $z = 0{,}01x + 3{,}19$
3. 2050 correspond à $x = 200$.
   On a alors $z \approx 0{,}01 \times 200 + 3{,}19 \Leftrightarrow \ln(y - 250) \approx 5{,}19 \Leftrightarrow y - 250 \approx \mathrm{e}^{5{,}19}$
   $\Leftrightarrow y \approx 250 + \mathrm{e}^{5{,}19} \approx 429$

▶ Exercice n°8

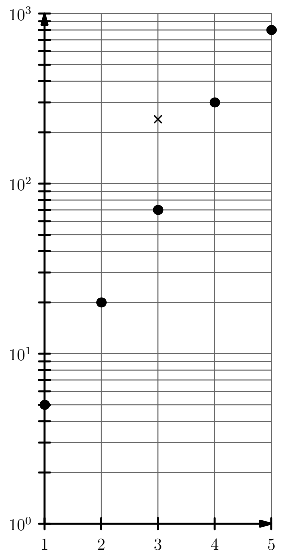
1.

Température x_i	20	30	40	50	60	70
Solubilité s_i	10,30	10,59	10,81	11	11,15	11,28
$y_i = \mathrm{e}^{s_i}$	29733	39735	49513	59874	69564	79221

2. $y = 992x + 9955$
3. $y = 992x + 9955 \Leftrightarrow \mathrm{e}^s = 992x + 9955 \Leftrightarrow s = \ln(992x + 9955)$

▶ Exercice n°9

1.



2. $\overline{x} = 3$; $\overline{y} = 239$
3. Coefficient de corrélation pour $(x_i; y_i)$: $r \approx 0{,}882$
   Coefficient de corrélation pour $(x_i; \log(y_i))$: $r \approx 0{,}998$
   Les points sont beaucoup plus proches d'une droite avec la série $(x_i; \log(y_i))$.