Série statistique à 2 variables

T^le STMG

Table des matières

1 Définitions

1.1 Définition : Série statistique à 2 variables

On appelle **série statistique à 2 variables**, l'étude simultatnée de 2 variables statistiques définies sur une même population.

Exemples :

- Le poids et la taille de nouveaux nés dans une marternité
- La consommation d'un véhicule et sa vitesse
- Le diamètre et la hauteur des arbres d'un forêt

Exemple :

On mesure l'allongement Y d'un ressort en fonction de la masse suspendue X.

Masse (en g)	30	40	50	60	70	80	90	100
Allongement (en mm)	12	19	24	30	37	42	48	55

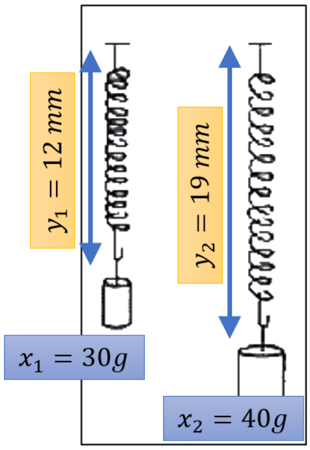


FIGURE 1 – Allongement d'un ressort

1.2 Définition : Nuage de points

Le plan étant muni d'un repère orthogonal, on peut associer chaque couple $(x_i; y_i)$ de la série statistique le point M_i de coordonnées $(x_i; y_i)$. Le graphique obtenue constitue un **nuage de points**.

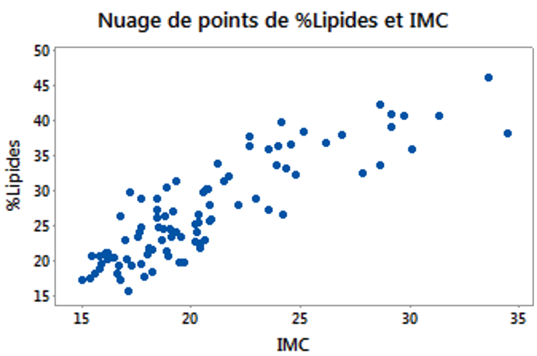
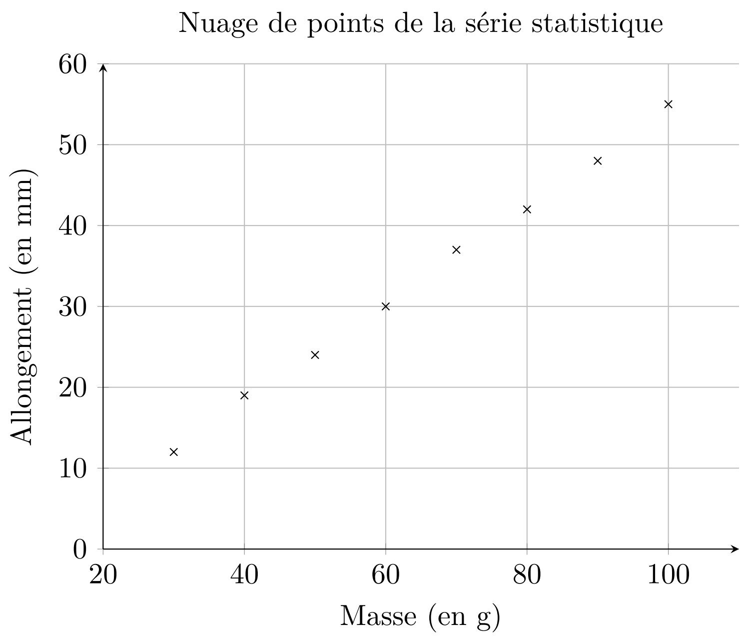


FIGURE 2 – Nuage de points

Exemple :

Masse (en g)	30	40	50	60	70	80	90	100
Allongement (en mm)	12	19	24	30	37	42	48	55



1.3 Définition : Point moyen

On appelle **point moyen** du nuage de points, le point G de coordonnées :

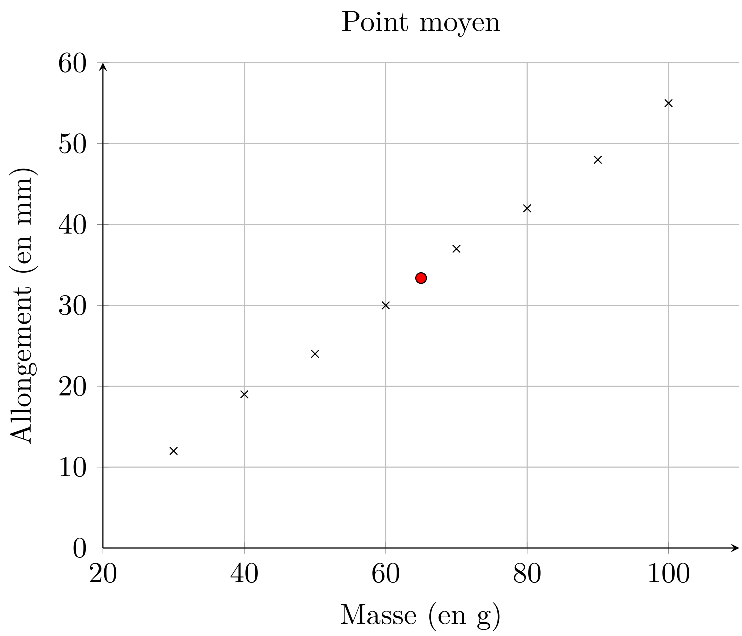
$x_G =$ "moyenne des x_i" et $y_G =$ "moyenne des y_i"

Exemple :

Masse (en g)	30	40	50	60	70	80	90	100
Allongement (en mm)	12	19	24	30	37	42	48	55

Le point G a pour coorodnnées :

$x_G = \dfrac{30 + 40 + ... + 100}{8} = 65$ et $y_G = \dfrac{12 + 19 + ... + 55}{8} = 33.375$



Remarque :

Soient X ey Y, deux variables statistiques définies sur la même population. Dans certains cas, on peut soupçonner l'existence d'une relation entre elles.

Exemple :

— Plus un arbre sera haut, plus son tronc aura un diamêtre important
— Plus un bébé naît grand, plus son poids sera élevé
— Plus le budget publicité d'une entreprise sera élevé, plus les stocks seront faible
— ...

1.4 Définition : Corrélation

Il y a **corrélation** entre deux variables X et Y lorsque X et Y varient dans le même sens (ou sens contraire).

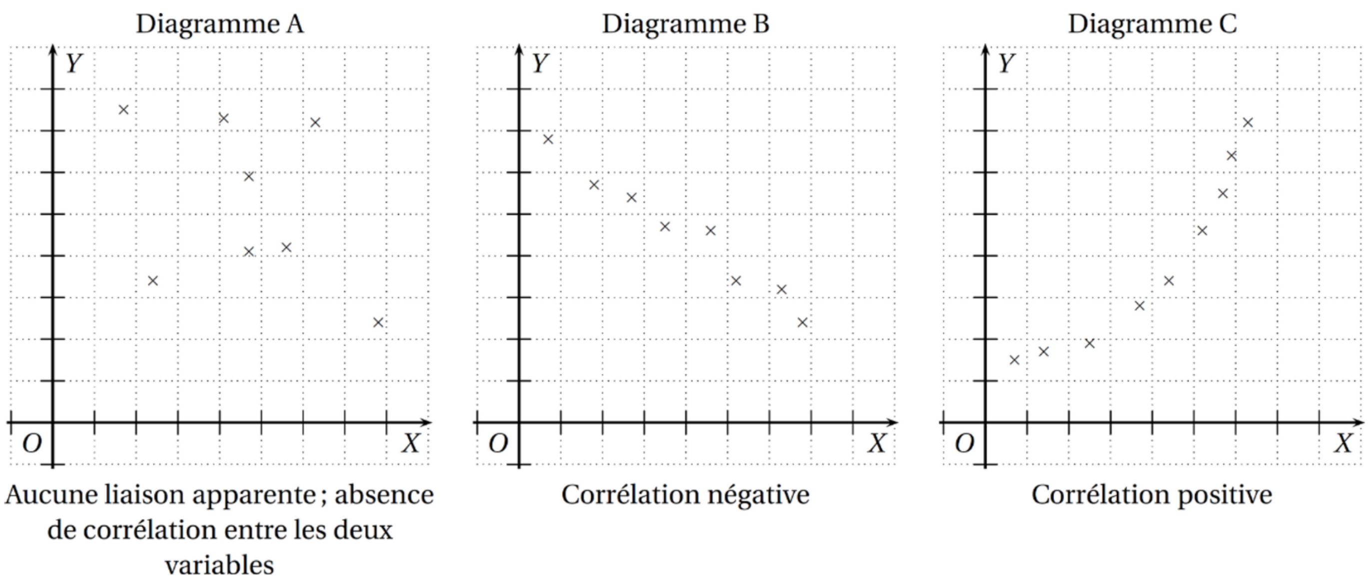
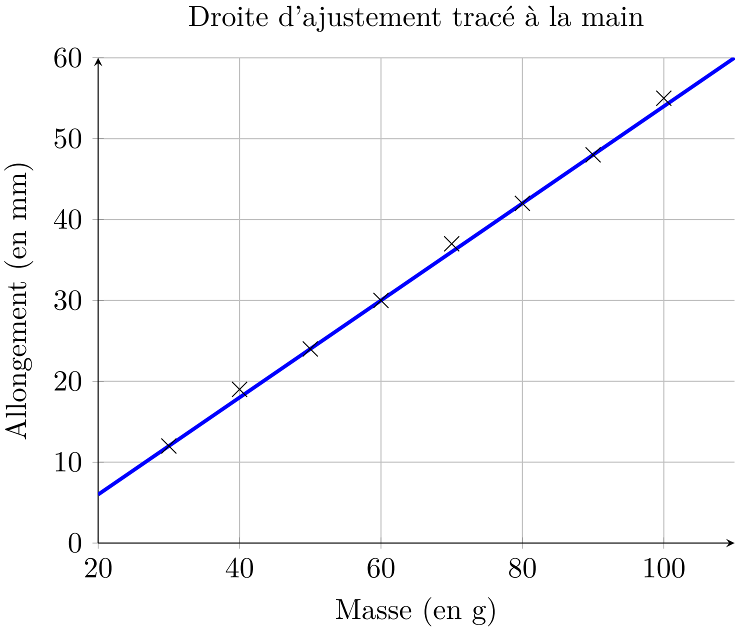


FIGURE 3 – Corrélation (ou pas)

2 Ajustement linéaire (ou affine)

2.1 Définition : droite d'ajustement

Lorsque les points d'un nuage sont sensiblement alignés, on peut tracer une droite, appelée **droite d'ajustement** (ou droite de régression), passant au plus près de ces points.



Rappel

L'équation d'une droite est de la forme $y = ax + b$

2.2 Méthode : Déterminer l'équation de la droite d'ajustement à l'aide de la série statistique ?

C'est la méthode des **moindres carrés** que la *calculatrice* va utiliser pour établir l'équation de la droite.

`Casio Graph 85 SD`

- Menu → Stat : Entrer les valeurs de x_i dans `list 1` et de y_i dans `list 2`
- Aller dans `CALC` → `REG` → `x`
- Vous pouvez lire les valeurs de a et b

Dans notre exemple on peut lire : $a \approx 0.603$ et $b \approx -5.857$

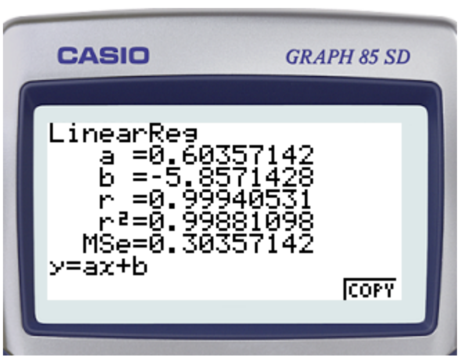
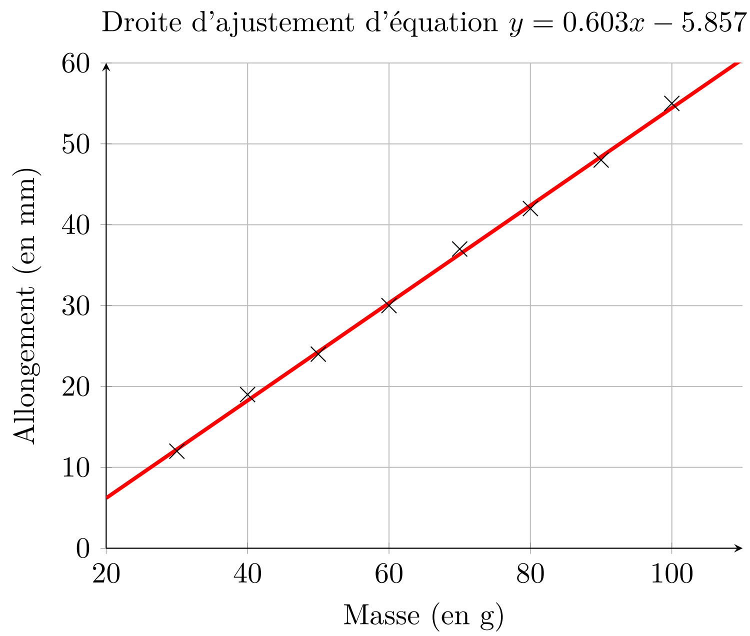


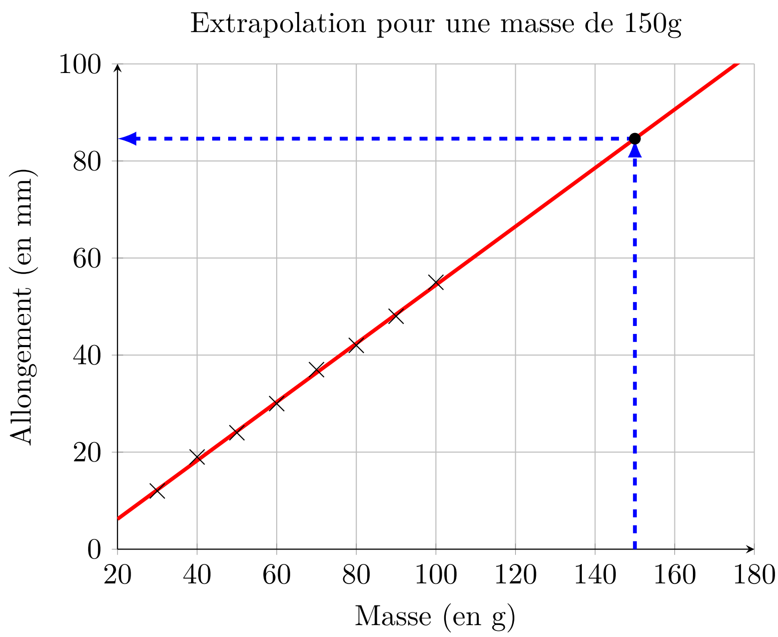
FIGURE 4 – Droite d'ajustement à l'aide de la Casio 85 SD



3 Interpolation - Extrapolation

Remarque :

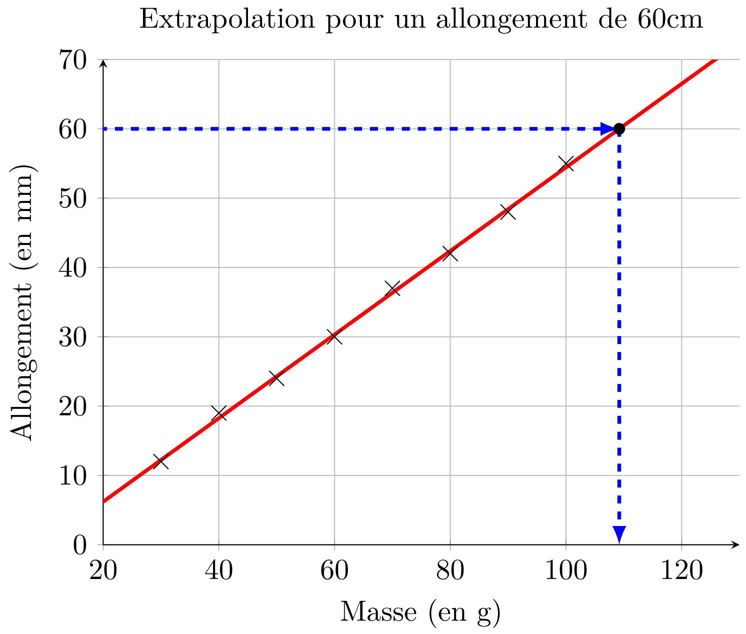
Nous pouvons prévoir l'allongement du ressort par rapport à la masse. Si le masse est de $x = 150$g alors le ressort s'allongera de $y = 0,603 \times 150 - 5,8 \approx 84.65$cm



Remarque :

Si on veut allonger le ressort de 60cm, il suffit résoudre : $60 = 0.603x - 5.857$

$$
\begin{aligned}
0.603x - 5.857 &= 60 \\
0.603x &= 60 + 5.857 \\
x &= \frac{65.857}{0.603} \approx 109.21
\end{aligned}
$$



3.1 Définitions : Extrapolation / Interpolation

- L'interpolation et l'extrapolationsont des méthodes qui consistentà estimer une valeur inconnue dans une série statistique.
  - Pour une interpolation, le calcul est réalisé **dans le domaine d'étude** fourni par les valeurs de la série.
  - Pour une extrapolation, le calcul est réalisé **en dehors du domaine d'étude**.

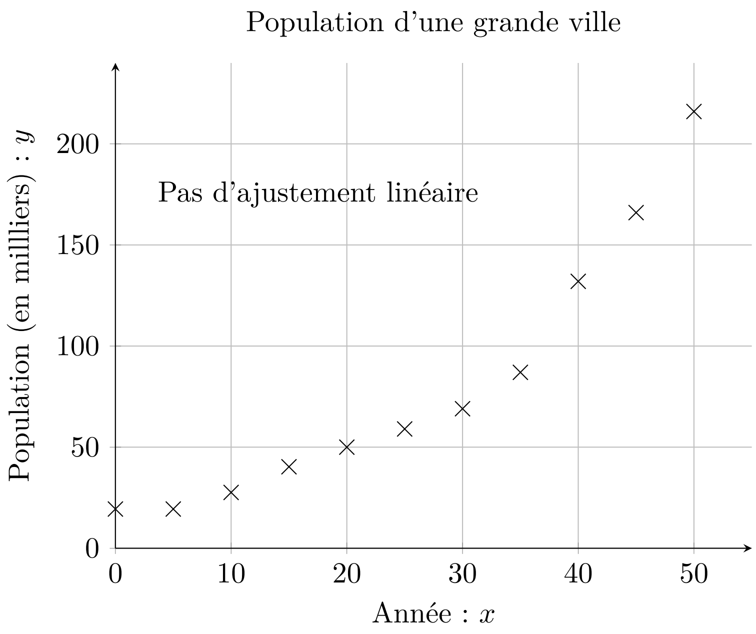
4 Ajustement par changement de variables

Lorsque le nuage de points n'est à priori pas modélisable par une droite, on peut, parfois, réaliser un ajustement linéaire en effectuant un changement de variable.

4.1 Exemple :

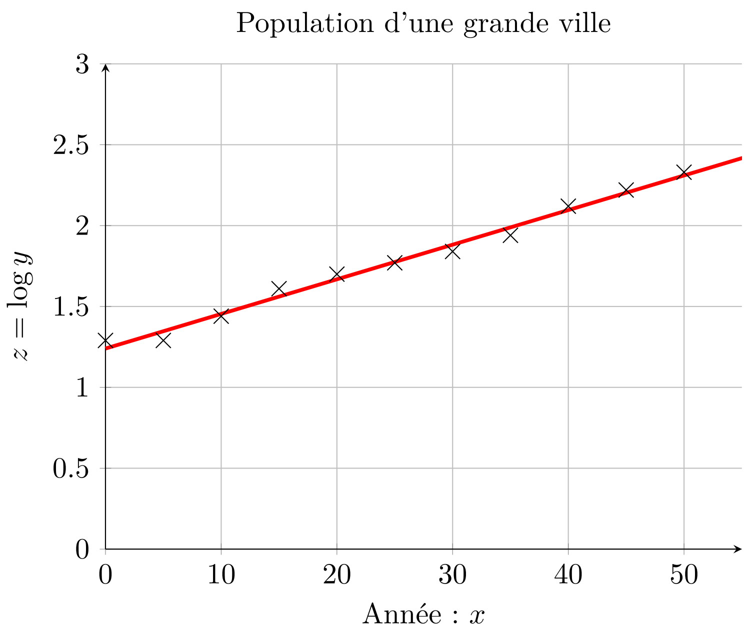
Population d'une grande ville sur 50 ans tous les 5 ans (en milliers).

Année x_i	0	5	10	15	20	25	30	35	40	45	50
Population y_i	19,4	19,4	27,6	40,3	50	59	69	87	132	166	216



On effectue le changement de variable $z = \log y$

Année x_i	0	5	10	15	20	25	30	35	40	45	50
Population y_i	19,4	19,4	27,6	40,3	50	59	69	87	132	166	216
z_i	1,29	1,29	1,44	1,61	1,7	1,77	1,84	1,94	2,12	2,22	$\log 216 = 2,33$



La droite d'ajustement a pour équation : $z = 0,0213x + 1,2427$

On a donc $z = 0,0213x + 1,2427$ et $z = \log y$ donc :

$$\begin{aligned} \log y &= 0,0213x + 1,2427 \\ y &= 10^{0,0213x+1,2427} \\ y &= 10^{0,0213x} \times 10^{1,2427} \\ y &= 10^{0,0213x} \times 17,4864 = 17,4864 \times 10^{0,0213x} \end{aligned}$$

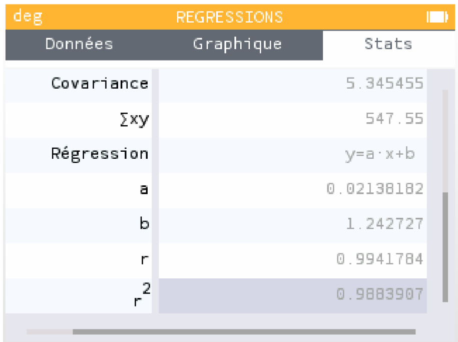


Figure 5 – Droite d'ajustement à l'aide de la NumWorks

Population d'une grande ville : $y = 17,4864 \times 10^{0,0213x}$