STATISTIQUES À DEUX VARIABLES

I. Statistique à deux variables :

1°) Définition :

Définition :

On appelle série statistique à deux variables (ou série statistique double), une série statistique où deux caractères sont étudiés simultanément.

Exemples :

Le poids et la taille de nouveaux nés dans une maternité.

Le volume des ventes et le montant alloué à la publicité dans une entreprise.

La consommation d'un véhicule et sa vitesse.

Remarque :

Dans ce chapitre, on n'étudiera que des séries statistiques doubles dont les deux caractères étudiés sont quantitatifs. Si, pour chacun des n individus de la population, on note x_i et y_i les valeurs prises par les deux caractères, on peut alors présenter la série statistique sous la forme d'un tableau :

Caractère x	x_1	x_2	x_3	…	x_n
Caractère y	y_1	y_2	y_3	…	y_n

Exemple 1 :

On a relevé, pour un modèle de voiture, la consommation en carburant (en L/100 km) pour différentes vitesse (en km/h) sur le cinquième rapport :

Vitesse x_i (en km/h)	60	70	90	110	130	150
Consommation y_i (en L/100 km)	3	3,1	3,7	4,7	6	9

Exemple 2 :

On a relevé, pour un site internet, le nombre de visiteurs (en milliers) par an pendant les huit premières années de fonctionnement :

Année	2007	2008	2009	2010	2011	2012	2013	2014
Rang x_i	1	2	3	4	5	6	7	8
Nombre de visiteurs y_i (en millier)	0,3	1,1	1,5	2,7	3	4	4,5	5,6

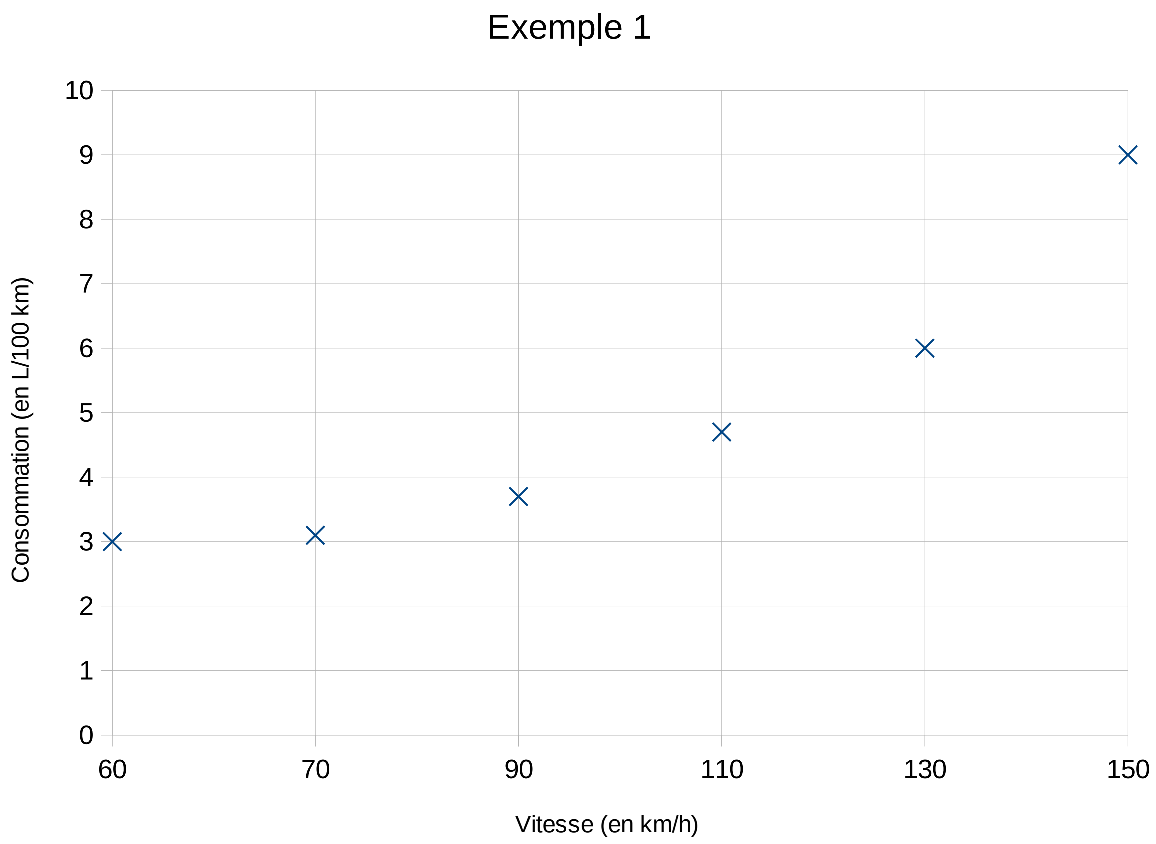
2°) Nuage de points :

Définition :

Le nuage de points d'une série statistique à deux variables est la représentation graphique de cette série dans un repère orthogonal où l'un des caractères est en abscisse et l'autre en ordonnée. Autrement dit, c'est l'ensemble des points $M_i(x_i ; y_i)$.

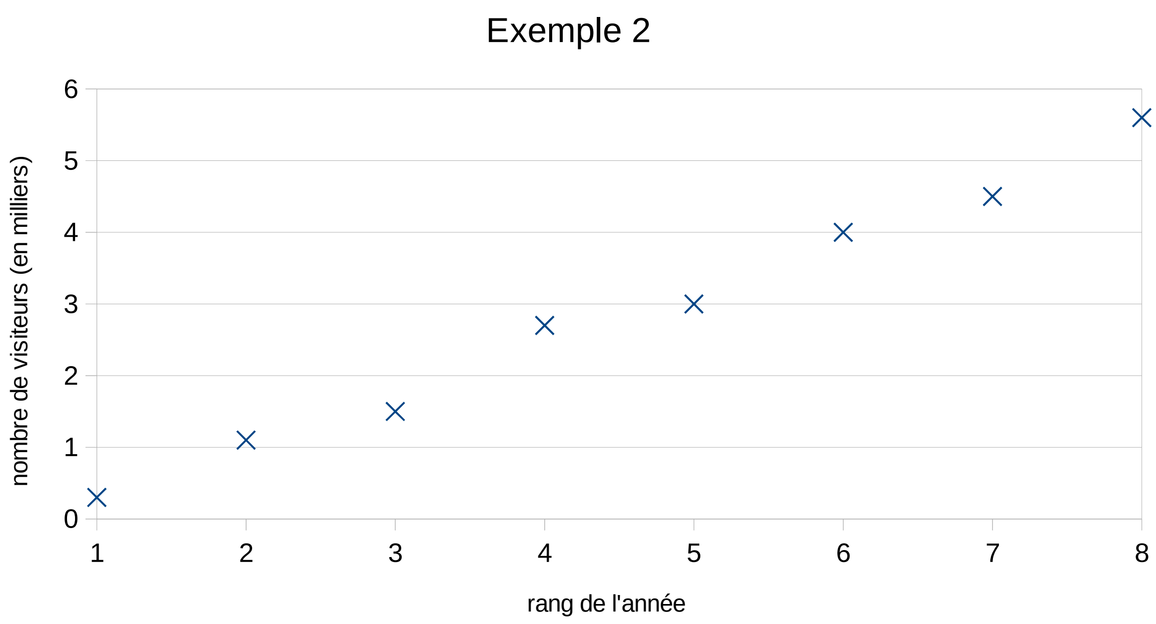
Exemple 1 :

Voici le nuage de points de l'exemple 1 précédent :



Exemple 2 :

Voici le nuage de points de l'exemple 2 précédent :



3°) Point moyen :

Définition :

On appelle point moyen d'un nuage de n points $M_i\,(x_i\,;\,y_i)$, le point G de coordonnées $G(\overline{x}\,;\,\overline{y})$, où $\overline{x}$ et $\overline{y}$ sont les moyennes des séries à une variable (x_i) et (y_j).

C'est-à-dire : $\begin{cases} \overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i \\ \overline{y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i \end{cases}$.

Exemple 1 :

Le point moyen du nuage de points de l'exemple 1 précédent est G(101,7 ; 4,9).

Exemple 2 :

Le point moyen du nuage de points de l'exemple 2 précédent est G(4,5 ; 2,84).

II. Ajustement affine :

Définition :

Si les points du nuage sont sensiblement alignés, alors il existe une liaison affine entre les deux variables x et y. On peut alors tracer une droite qui ne passe pas loin de chacun des points. Cette droite est appelée droite d'ajustement. On dit alors qu'on a effectué un ajustement affine.

Remarque :

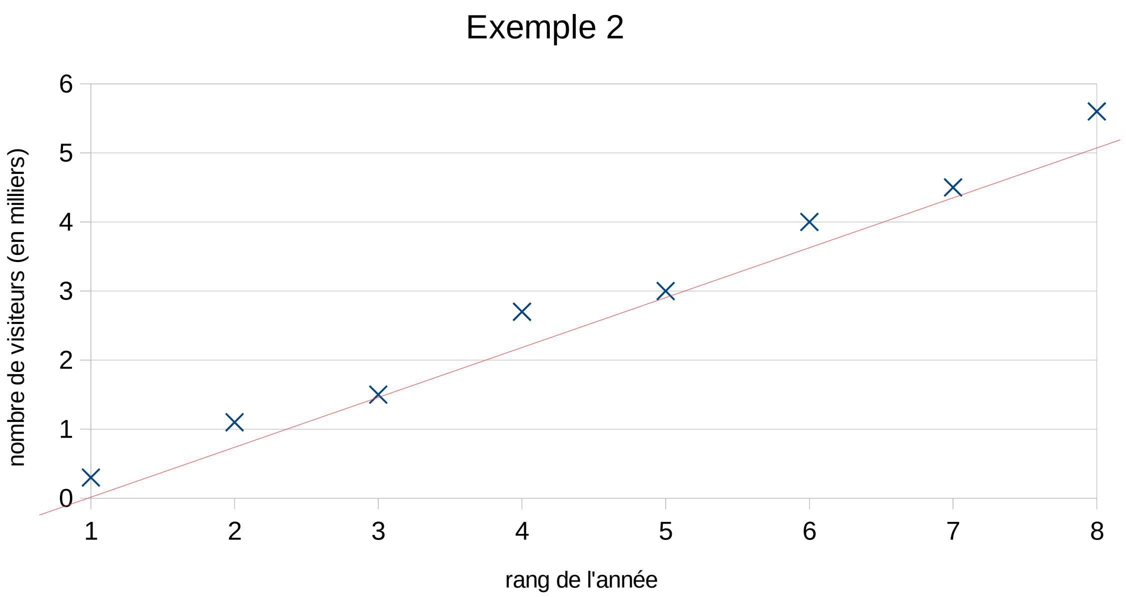
Un ajustement affine n'est pas unique : suivant la méthode utilisée, on ne trouvera pas la même droite d'ajustement.

1°) Ajustement affine graphique :

Sur le nuage de points, on trace une droite passant au plus près de tous les points.

Exemple :

Dans l'exemple 1, il ne semble pas exister de liaison affine entre les deux variables. Dans l'exemple 2 en revanche, les points du nuage sont sensiblement alignés. Traçons une droite d'ajustement :



Remarque :

Par cette méthode, on effectue un tracé « au jugé », nous n'obtiendrons donc pas à chaque fois exactement la même droite.

2°) Ajustement affine par la méthode des moindres carrés :

Définition :

La méthode des moindres carrés est une méthode permettant de déterminer par le calcul une équation d'une droite d'ajustement. La droite obtenue par cette méthode est appelée droite de régression de y en x.

Si nous considérons l'équation réduite de la droite (de la forme $y = ax + b$), la calculatrice calculera a et b via le menu statistique :

Pour les TI :

• Accéder au menu Stat , Edit.

Saisir les valeurs de x dans L_1 et celles de y dans L_2.

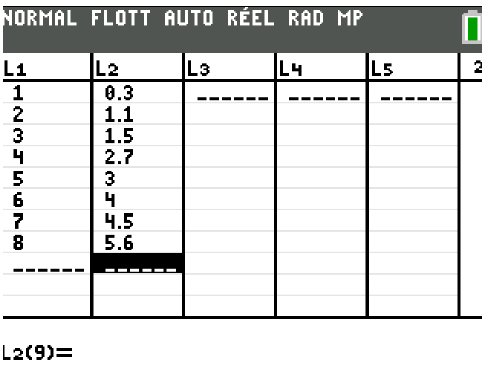
• Accéder au menu Stat, Calc,

4 : RégLin(ax+b) (ou LinReg(ax+b) suivant le modèle) et indiquer L_1 et L_2 pour Xliste et Yliste. Les autres champs doivent rester vides.

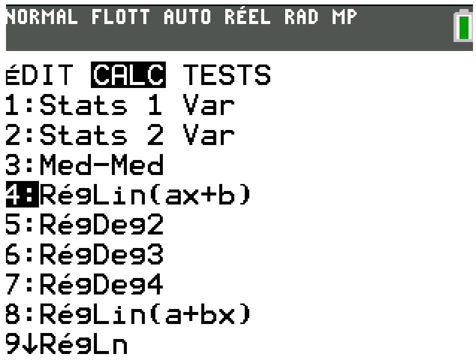
• Aller sur calculer et appuyer sur Entrer pour voir apparaître les coefficients a et b.

Exemple :

Reprenons l'exemple 2. Commençons par saisir les valeurs dans le tableau statistique de la calculatrice (touche stats, puis onglet edit) :



Allons ensuite chercher la fonction (touche stats, onglet calc) :



Complétons :

Enfin on lance le calcul et on obtient :

Dans notre exemple, l'équation de la droite de régression est donc environ $y = 0{,}74x - 0{,}48$.

Remarque :

La droite de régression ainsi trouvée passe par le point moyen.

Remarque :

Nous pouvons également effectuer les calculs des coefficients de la droite de régression à l'aide d'un tableur.

3°) Coefficient de corrélation :

Suivant les séries statistiques étudiées, l'approximation du nuage de points par une droite est plus ou moins précis. Pour apprécier la qualité d'une approximation affine, on introduit un nouveau paramètre.

Définition :

Le coefficient de corrélation r est un nombre tel que $-1 \leqslant r \leqslant 1$. Plus r est proche de 1 ou de -1, plus l'approximation affine est est justifiée.

Remarque :

Sur la calculatrice, le coefficient de corrélation s'affiche sous les coefficients de la droite (attention, c'est bien r qu'on regarde, et non pas r^2).

S'il ne s'affiche pas, il suffit d'aller dans le catalogue (2nde 0) puis de sélectionner CorrelAff dans la liste et de valider avec entrer.

III. Estimation, prévisions :

En supposant que l'évolution reste la même, on peut effectuer des estimations et des prévisions grâce à la droite d'ajustement, soit par la graphique, soit par le calcul à l'aide de l'équation de la droite.

Exemple :

Toujours dans l'exemple 1, supposons que les statistiques le l'année numéro 5 aient été perdues. Nous pouvons estimer alors la fréquentation du site grâce à l'ajustement affine que nous avons fait :

$y = 0{,}74x - 0{,}48$ avec $x = 5$, on obtient $y = 5 \times 0{,}74 - 0{,}48 = 3{,}22$. Nous estimons le nombre de visiteurs à 3 220 la cinquième année.

La valeur cherchée était à l'intérieur du nuage de points, on dit qu'on a fait une interpolation.

Exemple :

Toujours dans l'exemple 1, nous voulons estimer la fréquentation du site au bout de 20 ans de fonctionnement, en considérant que l'évolution reste la même.

$y = 0{,}74x - 0{,}48$ avec $x = 20$, on obtient $y = 20 \times 0{,}74 - 0{,}48 = 14{,}32$. Nous estimons le nombre de visiteurs à 14 320 la vingtième année.

La valeur cherchée était au-delà du nuage de points, on dit qu'on a fait une extrapolation. Il s'agit ici d'une prévision.