

STATISTIQUE 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 vitesses (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	2014	2015	2016	2017	2018	2019	2020	2021
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

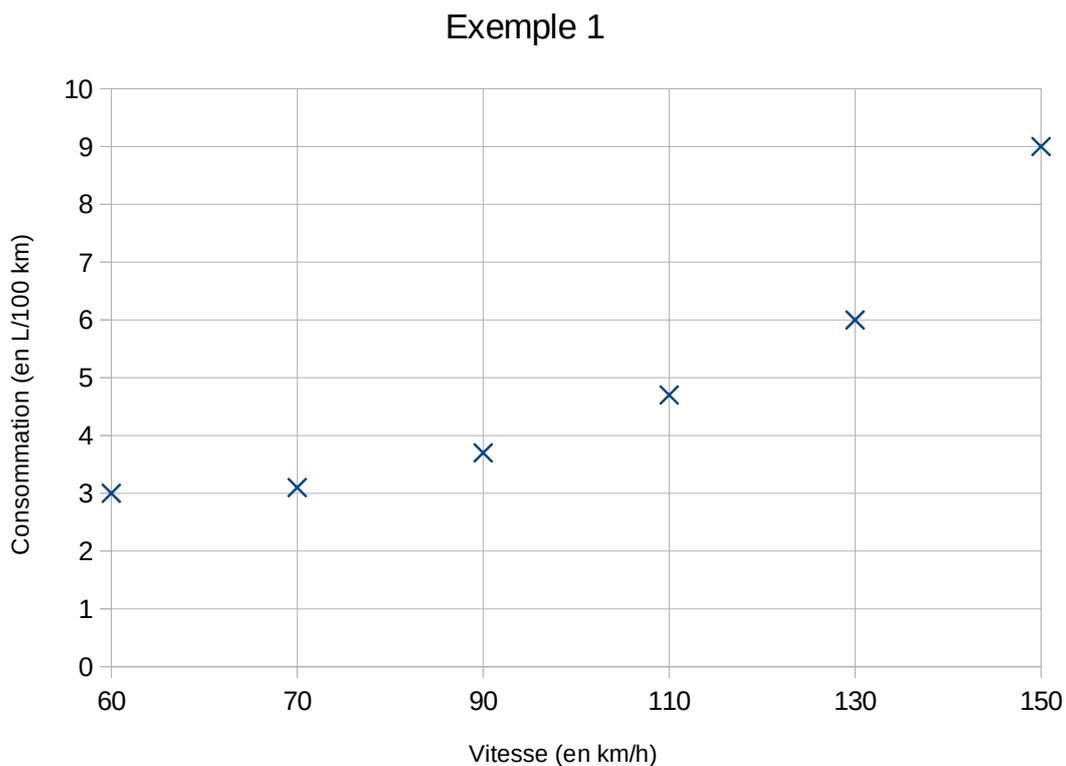
2°) Nuage de points :

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.

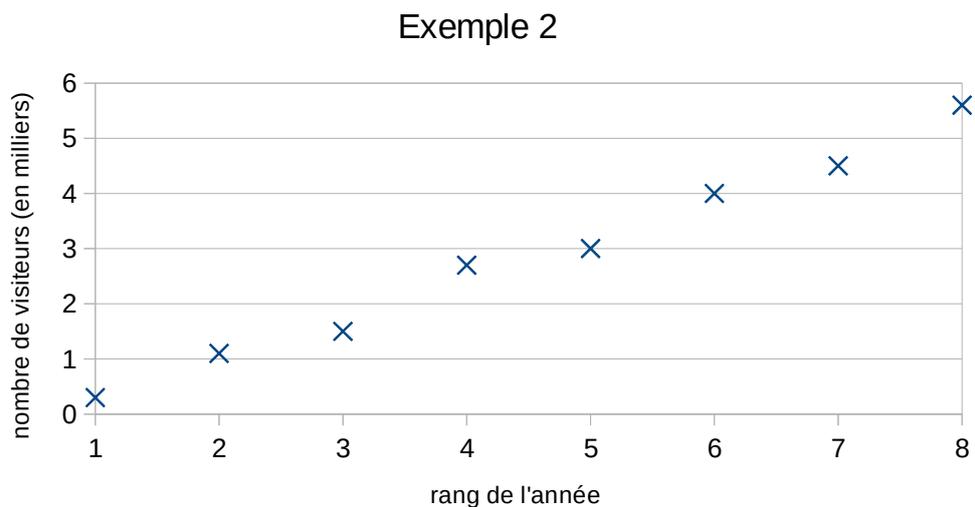
Exemple 1 :

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



Exemple 2 :

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



Nuage de points à la calculatrice :

Pour les TI :

- Accéder au menu Stats , EDIT. Saisir les valeurs du caractère abscisse dans L₁ et celles du caractère ordonnée dans L₂.

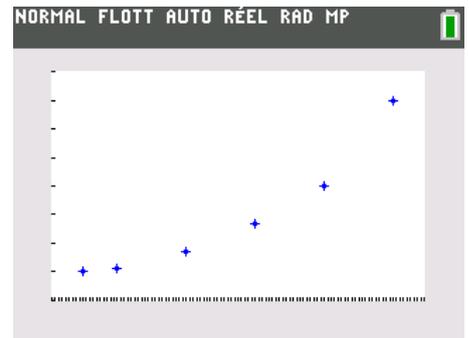
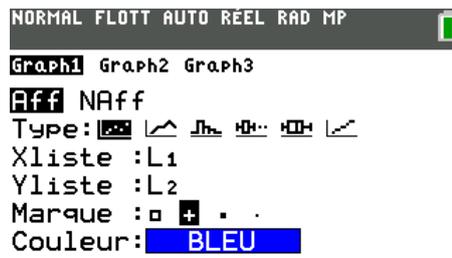
- Accéder au graphique statistique : 2^{de} graphstats, entrer dans graph1, sélectionner Aff, type nuage (1^{er} type), ListeX : L₁, ListeY : L₂, puis zoom 9 : Zoomstat

Exemple :

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

L1	L2	L3	L4	L5	2
60	3				
70	3.1				
90	3.7				
110	4.7				
130	6				
150	9				

L2(?)=



Tutos en vidéo (pour tout le cours) :

Pour TI : [ici](#) ou



pour casio : [ici](#) ou



4°) Point moyen :

Définition :

Le point moyen G d'un nuage de N points $M_i(x_i; y_i)$, est le point de coordonnées $G(x; y)$, où x et y sont

les moyennes des séries à une variable (x_i) et (y_i) ; c'est-à-dire :

$$\begin{cases} \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \\ \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \end{cases}$$

Exemples 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 $G(4,5 ; 2,84)$.

Remarque :

Les coordonnées du point moyen peuvent être obtenues à la calculatrice (voir tutos précédents) :

Pour les TI :

- Accéder au menu Stats , EDIT. Saisir les valeurs du caractère abscisse dans L₁ et celles du caractère ordonnée dans L₂.
- Puis Stats , CALC, 2:Stats 2-Var, indiquer L₁ pour Xlist, L₂ pour Ylist, et rien dans ListeFréq, puis entrer.
- On obtient les paramètres des deux séries, en particulier leurs valeurs \bar{x} et \bar{y} moyennes respectives des x_i et y_i .

II. Ajustement affine :

1°) Définition :

Définition :

Si les points du nuage sont sensiblement alignés (« forme étirée du nuage autour d'une ligne »), 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.

Remarques :

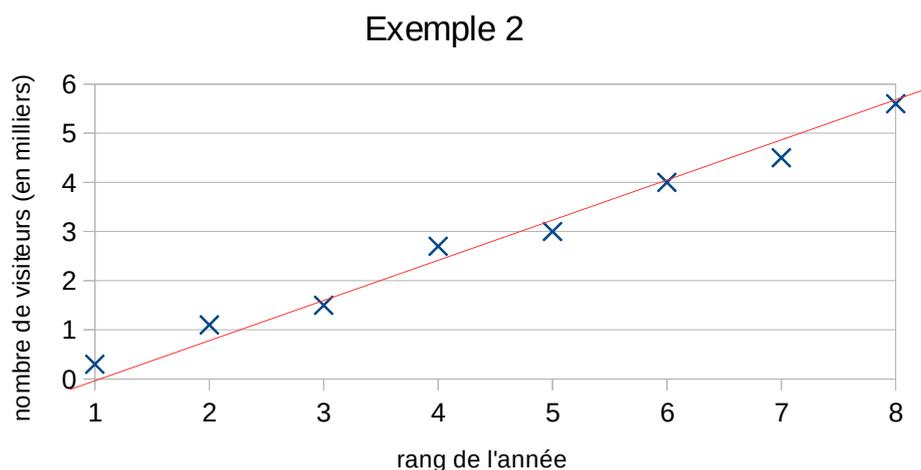
- Un ajustement affine n'est pas unique : suivant la méthode utilisée, on ne trouvera pas la même droite d'ajustement.
- En supposant que la droite d'ajustement soit un modèle pour le lien entre les variables, on peut effectuer des estimations et des prévisions grâce à la droite d'ajustement, soit par la graphique, soit par le calcul à l'aide de l'équation de la droite.

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 . C'est la droite d'équation $y = ax + b$ réalisant le minimum de $\sum_i (y_i - (ax_i + b))^2$ pour le nuage de points (x_i, y_i) .

Si nous considérons l'équation réduite de la droite (de la forme $y = ax + b$), la calculatrice calculera a et b (voir tutos précédents) :

Pour les TI :

- Accéder au menu Stats , EDIT. Saisir les valeurs du caractère abscisse dans L_1 et celles du caractère ordonnée dans L_2 .
- Puis Stats , CALC, 4 : RégLin(ax+b), indiquer L_1 pour Xlist, L_2 pour Ylist, et rien dans ListeFréq, puis entrer.
- 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 :

L1	L2	L3	L4	L5	2
1	0,3	-----	-----	-----	
2	1,1	-----	-----	-----	
3	1,5	-----	-----	-----	
4	2,7	-----	-----	-----	
5	3	-----	-----	-----	
6	4	-----	-----	-----	
7	4,5	-----	-----	-----	
8	5,6	-----	-----	-----	

L2(9)=

Allons ensuite chercher la fonction :

NORMAL FLOTT AUTO RÉEL RAD MP	
ÉDIT	CALC TESTS
1:	Stats 1 Var
2:	Stats 2 Var
3:	Med-Med
4:	RégLin(ax+b)
5:	RégDeg2
6:	RégDeg3
7:	RégDeg4
8:	RégLin(a+bx)
9:	RégLn

Appuyons sur entrer et complétons :

NORMAL FLOTT AUTO RÉEL RAD MP	
	RégLin(ax+b)
Xliste:	L1
Yliste:	L2
ListeFréq:	
Enr réglEQ:	Y1
Calculer	

Enfin on appuie sur entrer pour lancer le calcul et on obtient :

NORMAL FLOTT AUTO RÉEL RAD MP	
	RégLin
y=ax+b	
a=	0.7369047619
b=	-0.4785714286
r ² =	0.9899496449
r=	0.9949621324

Dans notre exemple, l'équation de la droite de régression est donc environ $y = 0,74x - 0,48$ (avec les coefficient arrondis au centième).

Remarque :

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

III. Estimations, 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 de 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.