NETSCHOOL1
ACADEMY

Mathématiques

Classe : BAC

Chapitre : Statistiques

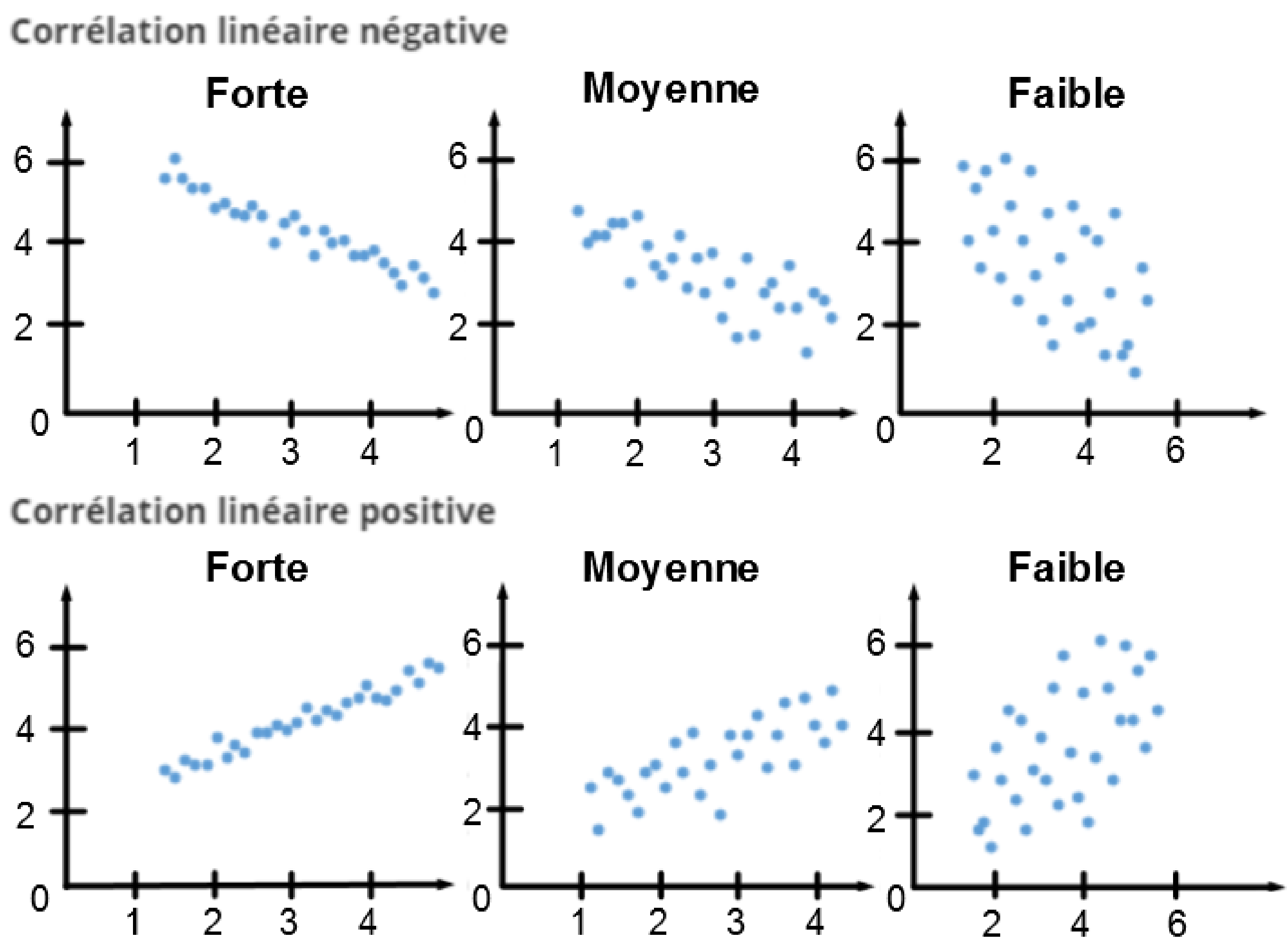
★★★★★★★★★

MAGAZINE DE MATHEMATIQUES

★★★★★★★★★

Profs : *ÉQUIPE* ACADEMIQUE *MATHEMATIQUES*

STATISTIQUES



★★★★★★★★★

RESUME DU COURS

★★★★★★★★★

I-RAPPELS

A- Série statistique double en données individuelles

Paramètres d'une série statistique :

Etant donnée une populations de n individus sur laquelle on étudie deux caractères X et Y. (On dit que (X,Y) est une série statistique double sur un échantillon de taille n)

On désigne par $x_1, x_2, ... x_n$ les variations de X

On désigne par $y_1, y_2, ... y_n$ les variations de Y

X	x_1	x_2	……………………………….	x_n
Y	y_1	y_2	……………………………….	y_n

L'ensemble des points $A_i(x_i, y_i)$ du plan muni d'un repère orthogonal est appelé nuage de points associé à la série statistique double (X,Y)

1°) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ s'appelle moyenne arithmétique de X (espérance de $X = E(X)$)

$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ s'appelle moyenne arithmétique de Y (espérance de $Y = E(Y)$)

Le point $G(\overline{X}, \overline{Y})$ est appelé le point moyen

2°) $V(X) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - \left(\overline{X}\right)^2$ s'appelle variance de X

$V(Y) = \frac{1}{n}\left(\sum_{i=1}^{n} y_i^2\right) - \left(\overline{Y}\right)^2$ s'appelle variance de Y.

3°) $6(X) = \sqrt{V(X)}$ écart type de X ; $6(Y) = \sqrt{V(Y)}$ écart type de Y.

Interprétation de l'écart type :

- L'écart type est un moyen qu'on utilise pour mesurer la dispersion des valeurs d'une variable statistique à variable quantitative autour de la moyenne de cette série.

Un écart type important signifie que les valeurs de la série s'éloignent souvent et de façon importante de la moyenne.

- Pour comparer deux séries statistiques qui n'ont pas le même ordre de grandeur, on peut comparer leurs écarts-type relatifs respectifs plus l'écart type est relatif est faible plus la dispersion au tour de la moyenne est faible.

$\frac{\sigma}{\overline{X}}$ est l'écart type relatif de X $\qquad$ $\frac{\sigma}{\overline{Y}}$ est l'écart type relatif de Y

B- Série statistique double en données groupées

Distribution marginale d'une série statistique double :

On désigne par $x_1, x_2, ..., x_i, ...x_p$ les valeurs de X et par $y_1, y_2, ..., y_i, ...y_q$ les valeurs de Y

Si le couple (x_i, y_i) se répètent plus qu'une fois soit n_{ij} l'effectif associé au couple.

Les effectifs n_{ij} associés aux couples (x_i, y_i) sont représentés à l'aide d'un tableau à double entrée de la forme :

x \ y	y_1	y_2	....	y_j	....	y_q	Distribution marginale de X
x_1	n_{11}	n_{12}		n_{1j}		n_{1q}	$n_1 = \sum_{j=1}^{q} n_{1j}$
x_2	n_{21}						$n_2 = \sum_{j=1}^{q} n_{2j}$
x_j	n_{i1}						
x_p	n_{p1}						n_p
Distribution marginale de Y	$n_1 = \sum_{i=1}^{p} n_{i1}$	n_2	....	n_j	....	n_q	n

D'où les deux tableaux ci-dessous représentent les distributions marginales de X et Y :

X	x_1	x_2	....	x_p
n_i	n_1	n_2	....	n_p

Y	y_1	y_2	....	y_q
n'_j	n'_1	n'_2	....	n'_q

NETSCHOOL1 ACADEMY

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{p} n_i x_i \;;\; \overline{Y}=\frac{1}{n}\sum_{i=1}^{q} n'_i y_i$$

$$V(X)=\frac{1}{n}\left(\sum_{i=1}^{p} n_i x_i^2\right)-\left(\overline{X}\right)^2 \;;\; V(Y)=\frac{1}{n}\left(\sum_{i=1}^{q} n'_i y_i^2\right)-\overline{Y}^2$$

$$\sigma(X)=\sqrt{V(X)} \;;\; \sigma(Y)=\sqrt{V(Y)}$$

$$Cov(X,Y)=\frac{1}{n}\left(\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij} x_i y_i\right)-\overline{X}\,\overline{Y}$$

Remarque1 : Toutes les formules vues dans la partie précédente de ce cours restent valables pour le cas d'une Série statistique double

Remarque2 : Si les caractères sont continus on considère les centres des classes

II-COVARIANCE

1-Définition Soit (x_i,y_i) avec i=1,……..n une série statistique double .On appelle covariance x et y le nombre noté Cov(x,y) et définie par : $Cov(X,Y)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$

2-Théorème

* Soit (x_i,y_i) avec i=1,……..n une série statistique double : $Cov(X,Y)=\frac{1}{n}\sum_{i=1}^{n} x_i y_i-\overline{X}\,\overline{Y}$

* $Cov(X,Y)=Cov(Y,X)$

3-Interprétation de la covariance :

La covariance permet une mesure de la dispersion des points du nuage par rapport au point moyen

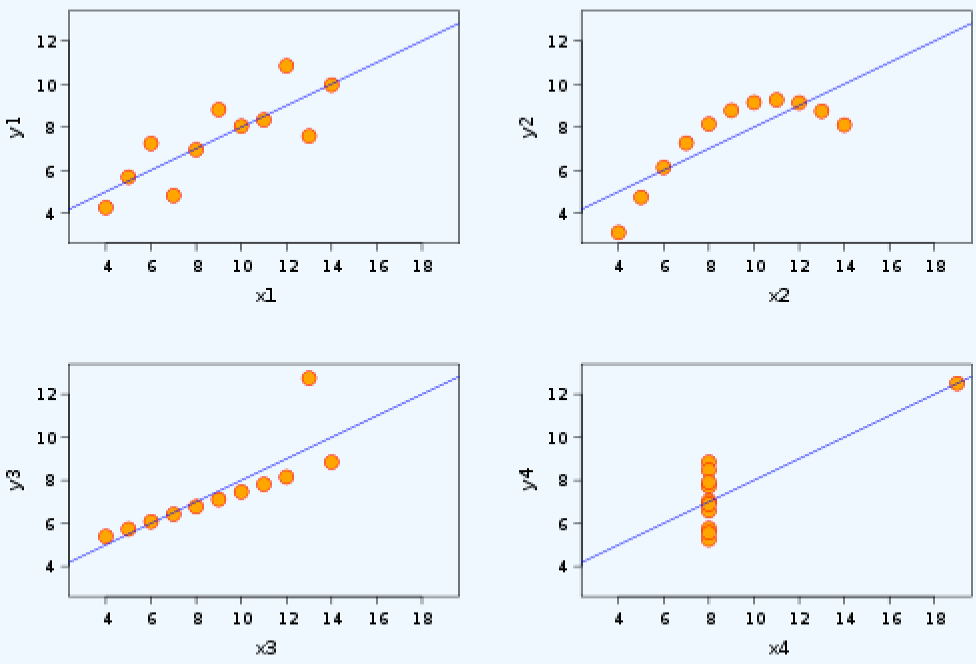
La covariance est positive si X et Y ont tendance à varier dans le même sens

La covariance est négative si X et Y ont tendance à varier en sens contraire.

III-AJUSTEMENT AFFINE

1-Ajustement affine

: La courbe peut être une droite ou une parabole.



ou bien il peut ne pas y avoir de courbe visible :

2- Méthode de Mayer:

Soit (X,Y) une série statistique double et G son point moyen

On scinde le nuage de point de (X,Y) en deux parties contenant à peu prés le même nombre de points obtenant ainsi deux nuages de points.

On désigne par G_1 et G_2 les points moyens de ces deux nuages.

La droite (G_1G_2) passe par le point G et définie un ajustement affine du nuage de points représentant la série statistique double (X,Y).

3- Méthode de Moindre carrés:

a- Principe de la méthode des moindre carrés

Le principe de cette méthode c'est de trouver la droite D « La plus proche possible » des points du nuage, c'est-à-dire que la somme des carrés des écarts entre les points M_i du nuage et les points P_i de la droite D de même abscisse, est la plus petite possible.

On dit qu'on a effectué un ajustement linéaire par la méthode des **moindres carrés**.

b-Théorème (admis)

La droite de régression de y en x est la droite qui passe par G $\left(\overline{X},\overline{Y}\right)$ et de coefficient directeur le réel $a=\dfrac{\operatorname{cov}(X,Y)}{V(X)}$

Remarque 1

- La droite de régression de Y par rapport à X est : $D: y-\overline{Y}=\dfrac{\operatorname{cov}\left(X,Y\right)}{V(X)}(x-\overline{X})$
- La droite de régression de X par rapport à Y est : $D': x-\overline{X}=\dfrac{\operatorname{cov}\left(X,Y\right)}{V(Y)}(y-\overline{Y})$

Remarque2 :

$G(\overline{X},\overline{Y}) \in D \cap D' d'ou$:

$D: y=ax+b \; avec \; a=\dfrac{\operatorname{cov}(X,Y)}{V(X)} et \; b=\overline{Y}-a\overline{X}$

$D': x=a'y+b' \; avec \; a'=\dfrac{\operatorname{cov}(X,Y)}{V(Y)} et \; b'=\overline{X}-a'\overline{Y}$

Remarque3 :

Les deux coefficients a et a' sont de même signe et le coefficient de corrélation r vérifie

4- Coefficient de corrélation linéaire :

a)-Définition : Le coefficient de corrélation linéaire de $\left(X,Y\right)$ est : $\rho_{XY}=\dfrac{\operatorname{cov}\left(X,Y\right)}{6\left(X\right),6\left(Y\right)}$

Remarque :

- On note encore ρ_{XY} par r.
- Le coefficient de corrélation linéaire de $\left(X,Y\right)$ est égal Le coefficient de corrélation linéaire de $\left(X,Y\right)$

- $-1 \leq r \leq 1$
- r est invariant par changement d'unité ou d'origine.

b) Interprétation du coefficient de corrélation linéaire

- Si $|r| \leq \dfrac{\sqrt{3}}{2}$ alors la corrélation linéaire entre x et y est faible
- Si $|r| > \dfrac{\sqrt{3}}{2}$ alors la corrélation linéaire entre x et y est forte
- Les points du nuage de points sont alignés si et seulement si r = 1 ou r = −1.

IV-UTILISATION D'UNE CALCULATRICE

(Casio fx 570 ES ou fx 570 ES ou fx 991 ES plus)

Tous les calculs mentionnés ici s'effectuent dans le mode STAT

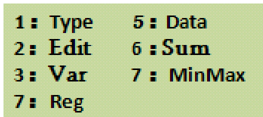
Types de calculs statistiques

Touche	Elément du menu	Calcul statistique
1	1-VAR	Une variable
2	A+BX	Régression linéaire
3	_+CX2	Régression quadratique
4	In X	Régression logarithmique
5	e^X	Régression exponentielle e
6	A•B^X	Régression exponentielle ab
7	A•X^B	Régression de puissance
8	1/X	Régression inverse

Utilisation du menu STAT

Lorsque l'écran de l'éditeur STAT ou l'écran de calcul STAT est affiché, appuyez sur pour afficher le menu STAT.
Le contenu du menu STAT est différent selon qu'une variable ou deux variables sont utilisées pour le calcul statistique actuellement sélectionné.

Statistiques à une variable

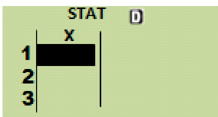


Statistiques à deux variables

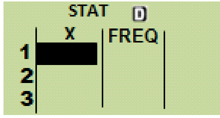
Exemple 1 : On considère la série statistique à une variable :

X	10	14
Effectif	40	20

❑ On passe en mode statistique

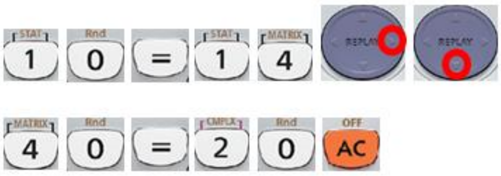


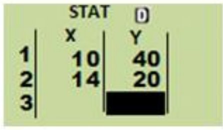
- Afficher la colonne des effectifs 

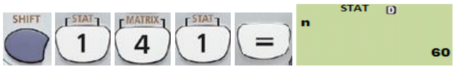
(FREQ) 

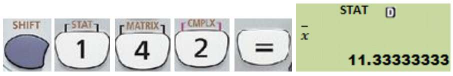
- Introduction des données

X	Effectif
10	40
14	20





- Pour déterminer l'effectif total
  - On trouve : **N = 60** 

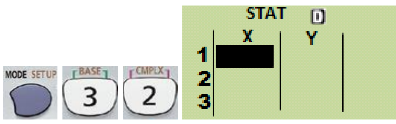
- Pour déterminer la moyenne
  - On trouve la valeur moyenne : $\overline{\mathbf{X}} \approx$ **11,33** 

- Pour déterminer l'écart type
  - On trouve l'écart type : $\sigma \approx$ **1,89** 

- Pour déterminer la variance
  - On trouve la variance : **V** $\approx$ **3,56** 

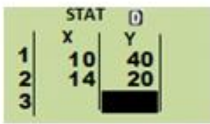
Exemple 2 : On considère la série statistique à deux variables :

X	10	14
Y	40	20

- On passe en mode statistique 

- Introduction des données

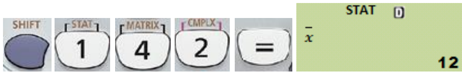
X	Y
10	40
14	20



NETSCHOOL1 ACADEMY

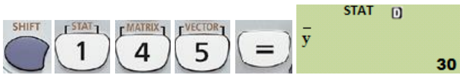
❑ Pour déterminer la moyenne de **X**

☞ On trouve la valeur moyenne : $\overline{\mathbf{X}} = 12$



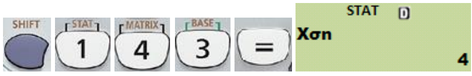
❑ Pour déterminer la moyenne de **Y**

☞ On trouve la valeur moyenne : $\overline{\mathbf{Y}} = 30$



❑ Pour déterminer l'écart type **X**

☞ On trouve l'écart type : $\sigma(\mathbf{X}) = 2$

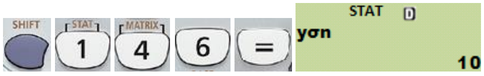


❑ Pour déterminer la variance de **X**

☞ On trouve la variance de **X** : $V(\mathbf{X}) =$

❑ Pour déterminer l'écart type **Y**

☞ On trouve l'écart type : $\sigma(\mathbf{Y}) = 10$

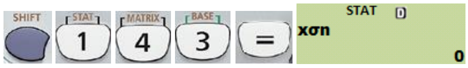


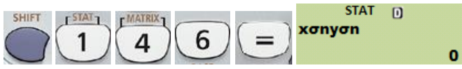
❑ Pour déterminer la variance de **Y**

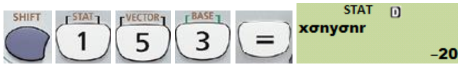
☞ On trouve la variance de **Y** : $V(\mathbf{Y}) = 100$

❑ Pour déterminer la covariance de **(X ,Y)**

☞ On trouve la covariance : $\mathbf{Cov(X ,Y)} = -20$





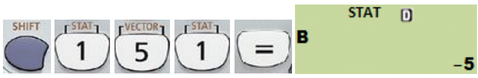


❑ Pour déterminer le coefficient de corrélation linéaire $\mathbf{r_{xy}}$:

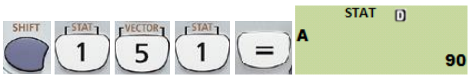
☞ On trouve: $\mathbf{r_{xy}} = -1$

❑ Droite de moindre carrés de **Y** en **X** <u>**ou**</u> droite de régression de **Y** en **X. (Y=BX+A)**

☞ On trouve: **B** = –5



☞ et on trouve: **A** = 90

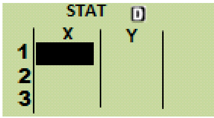


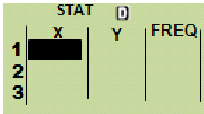
d'où Y = –5X+90

NETSCHOOL1 ACADEMY

Exemple 3 : On considère la série statistique à double entrée :

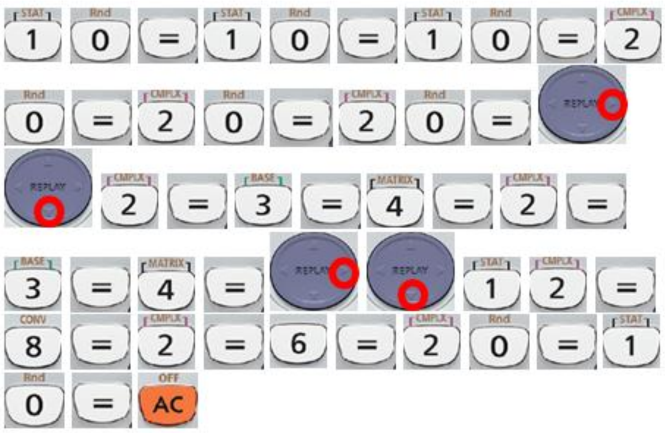
Y / X	2	3	4
10	12	8	2
20	6	20	10

❑ On passe en mode statistique  

❑ Afficher la  colonne des effectifs (**FREQ**) 

❑ Introduction des données

X	Y	FREQ
10	2	12
10	3	8
10	4	2
20	2	6
20	3	20
20	4	10



❑ Pour déterminer la moyenne de **X**

☞ On trouve la valeur moyenne : $\overline{\mathbf{X}} \approx 16.2$ 

❑ Pour déterminer la moyenne de **Y**

☞ On trouve la valeur moyenne : $\overline{\mathbf{Y}} \approx 2.9$ 

❑ Pour déterminer l'écart type **X**

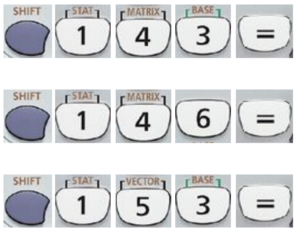
☞ On trouve l'écart type : $\sigma(\mathbf{X}) \approx 4.85$ 

❑ Pour déterminer la variance de **X**

☞ On trouve la variance de **X** : $V(\mathbf{X}) \approx 23.54$ 

❑ Pour déterminer l'écart type **Y**

☞ On trouve l'écart type : $\sigma(\mathbf{Y}) \approx 0.71$ 

- Pour déterminer la variance de **Y**
  - On trouve la variance de **Y** : V(**Y**) ≈ 0.5 
- Pour déterminer la covariance de **(X ,Y)**
  - On trouve la covariance : **Cov(X ,Y)** ≈ 1.33 
- Pour déterminer le coefficient de corrélation linéaire $\mathbf{r_{xy}}$:
  - On trouve: $\mathbf{r_{xy}} = 0.39$