

Filière : GC3 & GI3

Année universitaire : 2018 - 2019

Elément de module : Analyse des données

Enseignant : M. Derouch

Fiche TD N° : 2

**Exercice 1. :**

1. Montrer que :  $|\rho(X, Y)| \leq 1$ .
2. On considère deux séries statistiques  $(x_i)$  et  $(y_i)$  de taille  $n$   
 Soient  $\alpha_i$  et  $\beta_i$  deux séries statistiques liées aux séries statistiques  $(x_i)$  et  $(y_i)$   
 par les relations suivantes :  

$$\forall i \alpha_i = \frac{x_i - c}{d} \text{ avec } d \neq 0, c, d \in \mathbb{R}$$

$$\forall i \beta_i = \frac{y_i - c'}{d'} \text{ avec } d' \neq 0, c', d' \in \mathbb{R}$$
 Montrer les propriétés suivantes :  
 i)  $cov(\alpha, \beta) = \frac{1}{dd'} cov(x, y)$  et ii)  $\rho(\alpha, \beta) = \frac{|dd'|}{dd'} \rho(x, y)$

**Exercice 2. :**

Le tableau suivant représente des âges de patients X et les pressions systoliques Y de 9 malades.

L'âge X	56	42	72	36	63	47	55	49	38
Tension artérielle Y	147	125	160	118	149	128	150	145	115

1. Représenter le nuage de points  $M(x_i; y_i)$  dans le repère orthogonal ci-dessous.
2. Calculer la moyenne et l'écart-type de chacun des deux caractères X et Y.
3. Placer le point  $G(\bar{X}, \bar{Y})$  dans le repère précédent.
4. Calculer la covariance et le coefficient de corrélation du couple (X,Y). Que peut-on conclure ?
5. Trouver la droite de régression de X en Y.
6. Lorsque l'âge est 75 ans, quelle Tension artérielle Y peut-on prévoir ?

**Exercice 3. :**

sur un échantillon de 100 étudiants, on relevé la taille X en centimètre, ainsi que le poids Y en kilogrammes comme l'indique le tableau suivant

X \ Y	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
[150, 160[	10	3	1	0	0
[160, 170[	2	12	6	7	2
[170, 180[	1	7	11	17	4
[180, 190[	0	2	2	4	9

1. Calculer la moyenne et l'écart-type de chacun des deux caractères X et Y
2. Calculer la covariance et le coefficient de corrélation du couple (X,Y). Que peut-on conclure?
3. Trouver la droite de régression de Y en X.

**Exercice 4. :**

On dispose pour un secteur industriel donné et sur une période de 8 années du nombre de salariés Y (en milliers) et du chiffre d'affaires X (en dizaines de milliards) :

Année	1	2	3	4	5	6	7	8
X	3	4	5	6	8	9	11	13
Y	3.5	4.2	5	5.5	6	6.5	6.7	7.2
Ln(X)	1,1	1,4	1,6	1,8	2,1	2,2	2,4	2,6

1. Représenter le nuage de points  $(x_i, y_i)$ .
2. Calculer la moyenne et l'écart-type de chacun des deux caractères X et Y.
3. Calculer la covariance du couple  $(X, Y)$ .
4. a) Donner le coefficient de corrélation linéaire  $\rho(X, Y)$  de la série statistique  $(x_i, y_i)$ . Un ajustement affine est-il justifié?  
b) Ecrire une équation de la droite de régression  $D$  de Y en X. Représenter  $D$  dans le repère précédent
5. Calculer la moyenne et l'écart-type de variable Z.
6. Calculer la covariance du couple  $(Z, Y)$ .
7. a) Donner le coefficient de corrélation linéaire  $\rho(Z, Y)$  de la série statistique  $(z_i, y_i)$ . Un ajustement affine est-il justifié?  
b) Ecrire une équation de la droite de régression  $\Delta$  de Y en Z.
8. En l'an 2010, on prévoit pour le secteur étudié un chiffre d'affaires de 400 milliards.  
i) Utiliser les droites  $(D) : Y = aX + b$  et  $(\Delta) : Y = a'Z + b'$  pour proposer deux prévisions du nombre d'employés de ce secteur à l'horizon 2010.  
ii) Quelle prévision vous semble la plus appropriée