Examen Semestriel

M2 Biologie Moléculaire **2023-2024**

Exercice 01 : Des prises successives de tensions artérielles ont été faites à une personne donnée (on entend par x la variable indépendante représentant la pression systolique et par y la variable dépendante représentant la pression diastolique), les résultats sont mentionnés dans le tableau n°1 :

x_i	**126**	**128**	**128**	**130**	**130**	**132**
y_i	**78**	**80**	**82**	**82**	**84**	**86**

1. Représenter le nuage de points de cette série de données (on prend le dixième de ces valeurs) ;
2. Trouver l'équation de régression linéaire de y en x ;
3. Représenter cette droite dans le même graphique précédent ;
4. Calculer les coefficients de corrélation et de détermination, Commenter les valeurs obtenues.

Remarque importante *: tous les résultats sont donnés avec 2 chiffres après la virgule.*

Exercice n°2 : étant donné le tableau de l'exercice n°1, procédant à une analyse en composantes principales :

1. Centrer les donner du tableau n°1 pour obtenir la matrice de gravité G ;
2. Calculer ou déduire de l'exercice n°1 la matrice des variances-covariances V ;
3. Peut-on déduire si notre ACP est normée ou non normée ? et si les données du tableau n°1 sont homogènes ou non homogènes ?

On donne :	**Valeurs propres**	**Vecteurs propres**
1	10.08	$\begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}$
2	0.27	$\begin{bmatrix} -0.8 \\ 0.6 \end{bmatrix}$

4. Calculer l'inertie totale ;
5. Combien de composantes principales sont nécessaires pour récupérer 80% de l'inertie totale ?
6. Trouver ces composantes principales ;
7. Tracer le nouveau nuage de points dans le repère orthonormé (CP1, CP2)

À prendre en considération :

- La présentation des courbes sur papier millimétré ;
- L'utilisation d'une calculatrice et non pas du téléphone cellulaire.

Bonne Courage

Corrigé-type

M2 Biologie Moléculaire **2023-2024**

Biostatistiques et Informatique

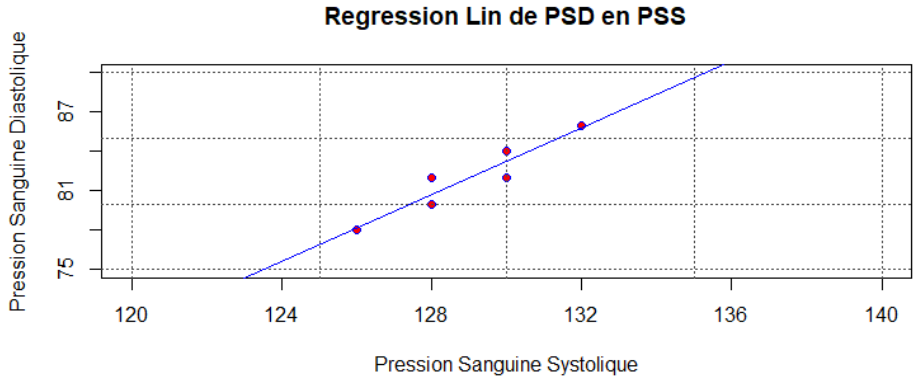
Exercice 01 : (12.00 pts)

							Sommes
x_i	126	128	128	130	130	132	**774**
y_i	78	80	82	82	84	86	**492**
x_i **^2**	15876	16384	16384	16900	16900	17424	**99868**
y_i **^2**	6084	6400	6724	6724	7056	7396	**40384**
x_i **x** y_i	9828	10240	10496	10660	10920	11352	**63496**

Ou bien : (1.75 pts)

$x_i - \overline{x_i}$	-3	-1	-1	1	1	3	Sommes
$y_i - \overline{y_i}$	-4	-2	0	0	2	4	
($x_i - \overline{x_i}$ **) ^2**	9	1	1	1	1	9	**22**
($y_i - \overline{y_i}$ **) ^2**	16	4	0	0	4	16	**40**
($x_i - \overline{x_i}$ **) x (** $y_i - \overline{y_i}$ **)**	12	2	0	0	2	12	**28**

1. Représenter le nuage de points de cette série de données (on prend le dixième de ces valeurs) ; (1.25 pts)



2. Trouver l'équation de régression linéaire de y en x ;

Moyenne de x : $\overline{x} = \frac{1}{N}\sum x_i = \frac{774}{6} = 129$

Moyenne de y : $\overline{y} = \frac{1}{N}\sum y_i = \frac{492}{6} = 82$

Variance de x : $V(x) = \frac{1}{N}\sum x_i^2 - \overline{x}^2 = \frac{99868}{6} - 129^2 = \frac{1}{N}\sum (x_i - \overline{x})^2 = \frac{22}{6} = 3.67$

Variance de y : $V(y) = \frac{1}{N}\sum y_i^2 - \overline{y}^2 = \frac{40384}{6} - 82^2 = \frac{1}{N}\sum (y_i - \overline{y})^2 = \frac{40}{6} = 6.67$

Covariance entre x et y : $\mathrm{cov}(x,y) = \frac{1}{N}\sum x_i \cdot y_i - \overline{x} \cdot \overline{y} = \frac{63496}{6} - 129 \cdot 82 = \frac{1}{N}\sum (x_i - \overline{x}) \cdot (y_i - \overline{y}) = \frac{28}{6} = 4.67$

Le coefficient directeur « a » :
(3.00 pts)

$$a = \frac{\text{cov}(x,y)}{V(x)} = \frac{28}{22} = 1.27$$

L'ordonnée à l'origine « b » :
(3.00 pts)

$$b = \bar{y} - a \cdot \bar{x} = 82 - 1.27 \cdot 129 = -81.83$$

L'équation de la régression linéaire :

$$y = 1.27 \cdot x - 81.83$$

3. Représenter cette droite dans le même graphique précédent ;
   (voir la 1ère figure)
4. Calculer les coefficients de corrélation et de détermination, Commenter les valeurs obtenues.

Le coefficient de corrélation « ρ » :
(2.00 pts)

$$\rho = \frac{\text{cov}(x,y)}{\sqrt{V(x) \cdot V(y)}} = \frac{4.67}{\sqrt{3.67 \cdot 6.67}} = 0.94$$

Une très forte corrélation positive en les deux variables.

Le coefficient de détermination « r » :
(1.00 pts)

$$r = \rho^2 = 0.94^2 = 0.89$$

Notre modèle mathématique peut expliquer 89% de l'évolution de PSD en fonction de PSS.

Exercice n°2 : **(8.00 pts)**

1. **(1.00 pts)** Centrer les donner du tableau n°1 pour obtenir la matrice de gravité G ;

$$G = \begin{bmatrix} x_i - \bar{x} & y_i - \bar{y} \end{bmatrix} = \begin{bmatrix} 126-129 & 78-82 \\ 128-129 & 80-82 \\ 128-129 & 82-82 \\ 130-129 & 82-82 \\ 130-129 & 84-82 \\ 132-129 & 86-82 \end{bmatrix} = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix}$$

2. **(1.00 pts)** Calculer ou déduire de l'exercice n°1 la matrice des variances-covariances V ;
   a. $V = \frac{1}{n} G^T \cdot G = \begin{bmatrix} V(x) & Cov(x,y) \\ Cov(y,x) & V(y) \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 22 & 28 \\ 28 & 40 \end{bmatrix} = \begin{bmatrix} 3.67 & 4.67 \\ 4.67 & 6.67 \end{bmatrix}$

3. **(1.00 pts)**

$$S = \frac{\sigma_{\max}}{\sigma_{\min}} = \frac{\sqrt{6.67}}{\sqrt{3.67}} = 1.35 < 5$$

Donc on procède à une ACP non-normée avec des données homogènes.

On donne :	**Valeurs propres**	**Vecteurs propres**
1	10.08	$\begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}$
2	0.27	$\begin{bmatrix} -0.8 \\ 0.6 \end{bmatrix}$

4. **(1.00 pts)** Calculer l'inertie totale ;

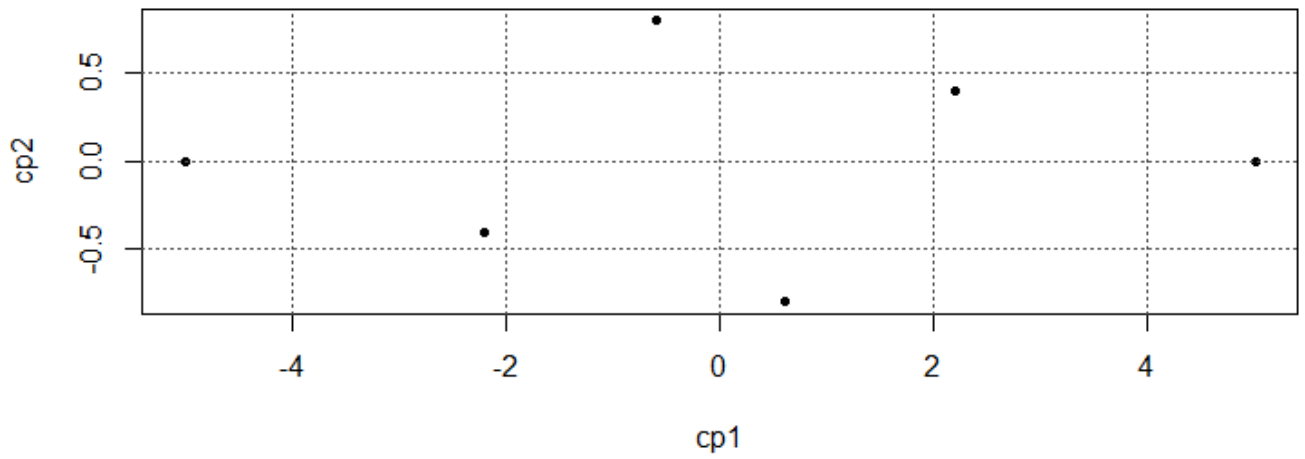
$$I = \sum \lambda_i = 10.08 + 0.27 = 10.35$$

5. **(1.00 pts)**

$$\frac{\lambda_1}{I} \cdot 100 = \frac{10.08}{10.35} \cdot 100 = 97\%$$

Donc une seule composante principale est nécessaire pour récupérer 97% d'information initiale (plus de 80%)

6. **(2.00 pts)**

$$CP = G \cdot VP = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{bmatrix} = \begin{bmatrix} -5 & 0 \\ -2.2 & -0.4 \\ -0.6 & 0.8 \\ 0.6 & -0.8 \\ 2.2 & 0.4 \\ 5 & 0 \end{bmatrix}$$

7. **(1.00 pts)**