Dirigé par : Pr. H.Merzouk *Années universitaires : 2025/2026*

<u>Correction de contrôle S4 :</u>

Matière :Analyse de données et statistique appliquées

<u>SOLUTION EX 01 : 8 points</u>

1- Les valeurs des plages numérique pour le coefficient de corrélation linéaire *r* : 1.5 point

0,5 L'intervalle général des valeurs de (r) 1 se situe entre +1 et -1.
• (r = + 1) : Corrélation positive parfaite. Les points sont alignés sur une droite de pente positive.
0,5 • (r = - 1) : Corrélation négative parfaite. Les points sont alignés sur une droite de pente négative. ·
0,5 • (r = 0) : Aucune relation linéaire n'existe (les variables peuvent être corrélées de manière non linéaire).
ou 0,5 • Positive ($0 < r \leq 1$) : Les deux variables augmentent simultanément.
0,5 • Négative ($-1 \leq r < 0$) : Lorsqu'une variable augmente, l'autre diminue.
0,5 • Non corrélée ($r \leq 0$) : Aucune relation linéaire n'est observée.

2- La définition de la covariance et définition de la variance et les différentes entre les deux : 2 points

0,5 **<u>Variance</u> (V(X)) ou (σ ^2) :** Mesure la dispersion des valeurs d'une variable aléatoire (X) par rapport à sa moyenne. Elle est calculée comme la moyenne des carrés des écarts à la moyenne.

- *Formule* $V(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{N}$

0,5 **<u>Covariance</u>** (Cov(X,Y)) : Quantifie la variabilité conjointe de deux variables aléatoires (X) et (Y). Elle indique si les variables augmentent ou diminuent ensemble (positive) ou dans des directions opposées (négative).

- *Formule :*

$$Cov(X,Y) = \frac{\sum[(x_i - \bar{X})(y_i - \bar{Y})]}{N} \qquad Cov(X,Y) = \frac{\sum(x_i y_i)}{N} - (\overline{XY})$$

➢ **Différences Principales :**

0,25 • **Nombre de variables :** La variance s'applique à une seule variable (X) , tandis que la covariance s'applique à la relation entre deux variables (X) et (Y) .
0,25 • **Signe et interprétation :** La variance est toujours positive ou nulle (mesurant une dispersion). La covariance peut être positive (tendances similaires), négative (tendances opposées) ou nulle (pas de relation linéaire).
0,25 • **Unité de mesure :** La variance est exprimée dans le carré de l'unité des données d'origine. La covariance est exprimée dans le produit des unités des deux variables.
0,25 • **Lien :** La covariance d'une variable avec elle-même est égale à sa variance [Cov(X,X) = V(X)]

3- Acceptable une hypothèse nulle (H_0) et la rejetons : 2 points

On rejette l'hypothèse nulle (H0) si la (p-valeur) est inférieure au seuil de signification (α), généralement fixé à 0,05. Dans ce cas, les résultats sont statistiquement significatifs. En revanche, si la (p-valeur) est supérieure à (alpha), on ne rejette pas (H0) ; les preuves sont insuffisantes pour conclure. 0,5

Règle de décision : La décision repose sur la comparaison entre la (p-valeur) (la probabilité d'obtenir les mêmes résultats que les vôtres si H0 était vraie) et le risque d'erreur (alpha) (votre seuil de tolérance, souvent (5%).Rejet de (H0) : Si (p-valeur < 0,05)Signification : La probabilité d'observer ces données par pur hasard est très faible. 0,5 **Conclusion** : Vous rejetez l'hypothèse nulle. Il existe une différence réelle ou un effet significatif. Vos données soutiennent l'hypothèse alternative (H1).Non-rejet (ou acceptation) de (H0) : Si (p-valeur > ou = 0,05) Signification : Les différences observées peuvent être simplement dues aux fluctuations de l'échantillonnage ou au hasard. 0,5 **Conclusion** : Vous ne rejetez pas l'hypothèse nulle. Il n'y a pas assez d'évidence statistique pour affirmer qu'un effet réel existe 0,5

4- Les différents types des mesures : 1 point

a- Mesure Nominale : **Ex** : numéros de maillots des joueurs. 0,25
b- Mesure Ordinale : **Ex** : ordre d'arrivées des joueurs. 0,25
c- Mesure Graduées : **Ex** : évaluation des performances sur une échelles de rapports 0,25
d- Mesure de Rapport : **Ex** : le temps nécessaire pour terminer la course. 0,25

5- Les différents types des variables selon leurs relations mutuelles : 1.5 point

a- Variable Indépendante (VI) / Explicative. 0,25
b- Variable Dépendante (VD) / Réponse. 0,25
c- Variable Médiatrice (Médiateur). 0,25
d- Variable Modératrice (Modérateur). 0,25
e- Variable de Contrôle / Covariable. 0,25
f- Variables de Confusion. 0,25

SOLUTION EX 02 : 12 points

1-

- ✓ L'hypothèse nulle **(H0) :** il n'y a pas une relation entre la quantité de la matière première utilisée (maïs) kg **X** et la quantité les additifs destinés à améliorer la qualité (g)**Y**. 0,75
- ✓ L'hypothèse alternative (bilatérale) **(H1) :** il y'a une relation entre la quantité de la matière première utilisée (maïs) kg **X** et la quantité les additifs destinés à améliorer la qualité (g) **Y**. 0,75

2- Restructuration des données du tableau :

Les mois (n)	la quantité de la matière première utilisée (maïs) kg **X i**		la quantité les additifs destinés à améliorer la qualité (g)**Y i**		$(x_i - \bar{X})$	$(x_i - \bar{X})^2$	$(y_i - \bar{Y})$	$(y_i - \bar{Y})^2$	$(x_i-\bar{X})(y_i-\bar{Y})$
1	X_1	134	Y_1	44	-11	121	-6	36	66
2	X_2	94	Y_2	45	-51	2601	-5	25	255
3	X_3	92	Y_3	40	-53	2809	-10	100	530
4	X_4	168	Y_4	52	23	529	2	4	46
5	X_5	145	Y_5	48	0	0	-2	4	0
6	X_6	172	Y_6	54	27	729	4	16	108
7	X_7	208	Y_7	64	63	3969	14	196	882
Σ *N*= 7	**1013**		**347**		-2	**10758**	**-3**	**381**	**1887**
					0,5	0,5	0,5	0,5	0,5

2

dll = 7-2 = 5

Dont :

$\bar{X} = \frac{\sum Xi}{N}$ et $\bar{Y} = \frac{\sum Yi}{N}$

3.Calcule de la moyenne :

La taille de l'échantillon N =7 (cas des petits échantillons).

On calcule les moyennes arithmétiques $\bar{X}$ et $\bar{Y}$. On trouve :

$\bar{X}=\Sigma xi / N$ et $\bar{Y} = \frac{\sum Yi}{N}$

$\bar{X}$=1013 / 7 = 144.71 ≈ 145 . 0,5

$\bar{Y}$=347 / 7 = 49.57 ≈ 50 0,5

4. Calcule de la Variance :

$V(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{N}$ **donc :** $V(x)$= **10758** / 7 = 1536.85 0,5 et $V(y)$= **381** / 7 = 54.42 0,5

5. Calcule l'écart type :

$\delta(x) = \sqrt{V(x)}$= 39.20 0,5 et **$\delta(y) = \sqrt{V(y)}$= 7.37** 0,5

6. Calcule de la covariance :

$Cov(X,Y) = \frac{\sum[(x_i - \bar{X})(y_i - \bar{Y})]}{N}$ donc : ***Cov (X, Y)* = 1887 / 7 = 269.57** 0,5

7. Calcule de la coefficient de corrélation:

$r = \frac{Cov(xy)}{\delta_x \delta_y}$ donc : ***r* =269.57 / (39.20× 7.37) = + 0.93** →Corrélation positive forte. 0,5

✓ **Calcule de la coefficient R^2 :**

$r^2 = (0.93)^2 = R^2 \approx 0.87$. $R^2 = (0.87)^2 \times 100 = 75.7$ % . 0,5

8. La droite de régression de Y en X :

Elle est de *la* forme : Y'=*ax*+*b* 0,25

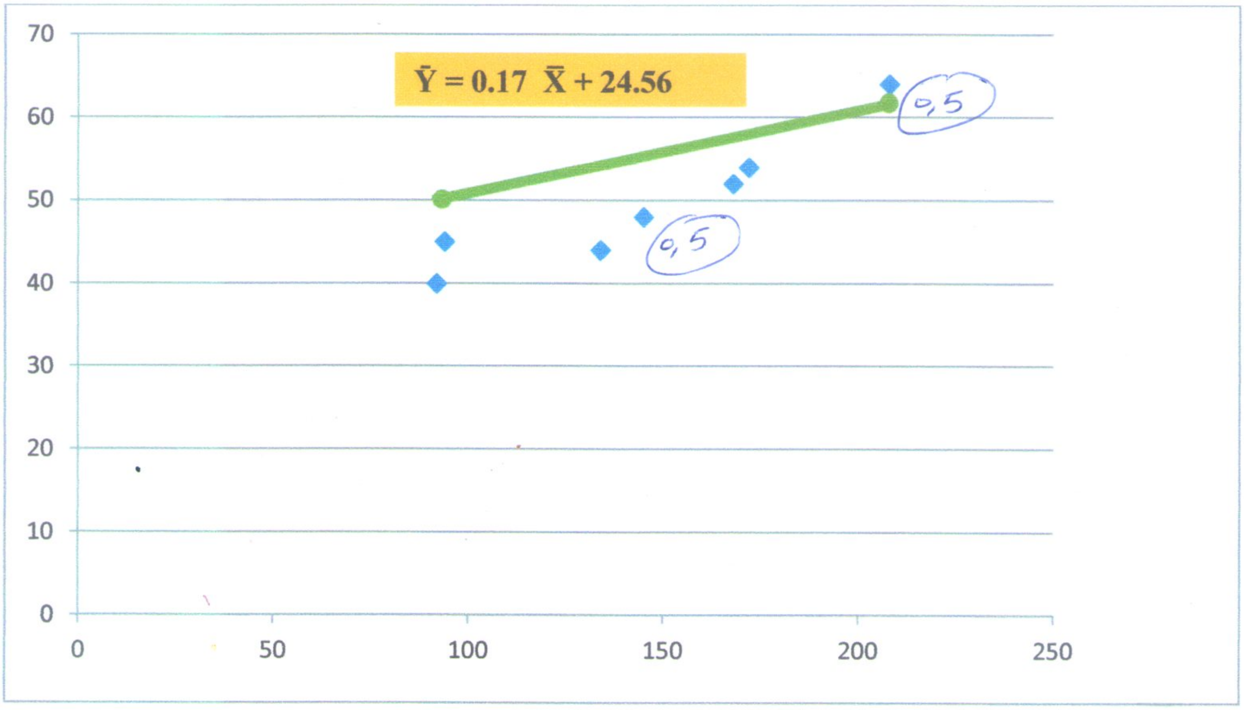
a = *Cov*(X, Y) / *V(x)* = **269.57 / 1536.85 = 0.175** 0,25

b = $\bar{Y}$ - *a*$\bar{X}$→ *b*= **50** - (**0.175**) × (**145**) = **24.56** 0,25

C'est-à-dire : l'équation de droite de régression de $\bar{Y}$:

$\bar{Y}$ = 0.17 $\bar{X}$ + 24.56 0,25

✓ **Tracer le graph de droite:** corrélation entre les dépenses publicitaire (X) et le volume des ventes des bon produits (Y).



9. Calcul t_0 :

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

t_0=[0.93 * √7-2] / √(1-0.87) = [(0.93×2.23)/ √0.13] = 5.75 = t_0 0,25

t_0=5.75 et *dll* (degré de liberté) = *n* – 2 = 7 - 2 = 5. 0,25

DONC : $t_{\alpha = 0.01} < t_0 < t_{\alpha = 0.001}$

10. Conclusion :

Nous pouvons conclure qu'il ya une corrélation positive forte entre la quantité de la matière première utilisée (maïs) kg (X) et de la quantité les additifs destinés à améliorer la qualité (g) (*r* = +0.93). 0,25

Le coefficient de corrélation entre la quantité de matière première utilisée (maïs) en kg (X) et la quantité d'additifs utilisés pour améliorer la qualité (g) (R^2 = 75,7 %) est d'environ 0,87.

Concernant les hypothèses : 0,25

On a : ***t_0=5.75 ; n=7 ; dll=5***

D'après le tableau du T- Student on conclure : **t_0 = 5.75** DONC : $t_{\alpha=0.01} < \mathbf{t_0} < t_{\alpha = 0.001}$

donc :

***α* < 0.05** → accepter **H1** et rejeter **H0.** 0,25

Donc il y a une relation forte entre la quantité de la matière première utilisée (maïs) kg et de la quantité les additifs destinés à améliorer la qualité (g). 0,25