

Contrôle Terminal
Statistiques - L1, S1
2022

Les calculatrices sont autorisées en mode examen.

Aucun document n'est autorisé.

Le barème est indiqué à titre indicatif et peut faire l'objet de modifications.

Les résultats doivent être arrondis au centième.

Tout calcul nécessite de présenter les formules utilisées.

L'indice et les bornes des sommes doivent être présentés : une simplification de la notation ne sera pas acceptée.

Exercice 1 (8 points) Démontrez les propositions suivantes.

Soit x_i et y_i , $i = 1, 2, \dots, N$, deux variables statistiques.

- (2 points) La moyenne empirique et la variance empirique du z-Score de x , noté $z_i(x)$, sont respectivement égales à 0 et 1.
- (2 points) L'expression de la covariance empirique entre x et y , $s_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$, peut être réécrite de la façon suivante :

$$s_{xy} = \frac{1}{N-1} \sum_{i=1}^N x_i y_i - \frac{N}{N-1} \bar{x} \bar{y}.$$

- (2 points) La covariance empirique entre x et y peut être calculée en utilisant la formule suivante : $s_{xy} = \frac{1}{N-1} \sum_{i=1}^N x_i (y_i - \bar{y})$.
- (2 points de bonus) Supposons maintenant que $\bar{x} = \bar{y} = 0$ et définissons $z_i = x_i + y_i$. Alors, $s_z^2 = s_x^2 + s_y^2 + 2s_{xy}$.

Exercice 2 (6 points) Le tableau ci-dessous présente les valeurs des z-Scores $z_i(x)$, $z_i(y)$ de deux variables statistiques x_i et y_i , $i = 1, \dots, N$.¹

i	1	2	3	4	5	6
$z_i(x)$	0.90	1.06	-0.16	-0.75	-1.51	0.45
$z_i(y)$	0.67	-0.45	1.68	-1.11	-0.36	-0.41

- Que pouvez-vous déduire de ce tableau sur la relation entre x_i et y_i ?

Trouvez et interprétez les statistiques suivantes :

- $b_1(x)$
- $b_2(y)$
- r_{xy}

Exercice 3 (8 points) Le tableau ci-dessous contient les données sur le pourcentage d'entreprises ayant une femme à la tête de la direction (x) dans 62 pays. Après avoir réordonné les données dans l'ordre croissant, nous trouvons les valeurs suivantes : $x_{(1)} = 0.9$; $x_{(7)} = 6.3$; $x_{(8)} = 6.6$; $x_{(16)} = 11.3$; $x_{(17)} = 11.9$; $x_{(31)} = 17.2$; $x_{(32)} = 17.7$; $x_{(46)} = 21.3$; $x_{(47)} = 21.8$; $x_{(55)} = 27$; $x_{(56)} = 27.8$; $x_{(62)} = 43.1$.

- Calculez et interprétez Q_1 , Q_2 , Q_3 .
- Définissons $\tilde{x}_i = x_i - \bar{x}$. On donne les statistiques suivantes :

$$\begin{aligned} \sum_{i=1}^{62} x_i &= 1072.8; & \sum_{i=1}^{62} x_i^2 &= 23092.58; \\ \sum_{i=1}^{62} \tilde{x}_i^3 &= 19982.32; & \sum_{i=1}^{62} \tilde{x}_i^4 &= 1157437.82. \end{aligned}$$

Calculez la moyenne empirique, la variance empirique, m_3 et m_4 de x . Interprétez m_3 . La distribution est-elle symétrique ?

- Tracez le diagramme en boîte. Interprétation.

¹ Les valeurs sont arrondies au centième. Elles satisfont les propriétés d'un z-Score mentionnées dans l'exercice 1, question 1. Pour cet exercice, nous considérons ces propriétés vérifiées. C'est-à-dire, $\overline{z(x)} = \overline{z(y)} = 0$ et $s_{z(x)} = s_{z(y)} = 1$.

TABLE 1 – Pourcentage d'entreprises ayant une femme à la tête de la direction

ALB	18.1	EST	22.5	LTU	30.7	RWA	21.8
ARG	8	GMB	9.6	LUX	7.4	SRB	18.2
ARM	19.1	GEO	16.5	MLT	11.3	SLE	15.9
AZE	16.5	GRC	17.2	MDA	18.6	SVK	22.9
BLR	21.1	GTM	18.5	MNG	38.9	SVN	18.8
BEL	17.9	HUN	22.1	MNE	15	SUR	11.9
BOL	26.3	ITA	15.3	MAR	5.4	TJK	6.6
BIH	16.6	JOR	3.1	MOZ	15.6	TUN	10.4
BGR	28.8	KAZ	26	NER	10.6	TUR	3.9
TCD	12	KEN	18.1	MKD	21.3	UKR	17.7
COL	18.9	XKX	2.7	PRY	19.6	URY	10.6
HRV	27	KGZ	32.9	PER	19.9	UZB	12.4
CYP	8.2	LAO	43.1	POL	27.8	PSE	0.9
CZE	16.1	LVA	32.6	PRT	14	ZMB	12.9
ECU	22.9	LBN	5.9	ROU	17.2		
EGY	6.3	LBR	20.4	RUS	24.3		