

Statistiques descriptives

Examen final

DURÉE : 2H

Documents interdits, calculatrice UPPA autorisée.

*La qualité de la rédaction sera prise en compte dans la note. Les réponses devront être justifiées.
On donnera les résultats à 10^{-4} près sauf pour les pourcentages où on travaillera à 10^{-2} près.*

Exercice 1 (5 points) Soit $\{(x_1, y_1), \dots, (x_n, y_n)\}$ une série statistique bivariée de deux variables quantitatives.

- (a) Définir la moyenne, la variance et l'écart-type de la variable x à partir de la série.
(b) Énoncer le Théorème de Koenig-Huygens pour la variance de la variable x .
- Définir la covariance de x et y .
- Définir le coefficient de corrélation linéaire $\rho_{x,y}$ entre x et y . Quels sont les valeurs possibles pour ce coefficient ?

Exercice 2 (6 points) Suite à une étude mise à jour¹ sur certains personnages de la série TV américaine *Game of Thrones* (aussi appelée *Le Trône de Fer*), trois variables ont été prises en compte : "Genre" (modalités : "homme" ou "femme"), "État du personnage" (modalités : "mort" ou "vivant") et "Famille" (modalités : "Stark" ou "Lannister"). Cela a conduit au tableau de contingence à trois dimensions donné ci-dessous :

	Stark		Lannister	
	Mort	Vivant	Mort	Vivant
Homme	20	15	15	17
Femme	7	6	3	5

- Quelle est la population ? Quelle est la nature des variables étudiées ? Justifiez votre réponse. Quelle est la taille de la population ?
- Pour l'ensemble des personnages, quel est la fréquence de la modalité "Mort" (pour la variable "État du personnage") ?
- Construire les tableaux des effectifs marginaux et des fréquences marginales croisant les variables "État du personnage" et "Famille".
- Calculer le coefficient de Pearson Φ^2 pour les variables "État du personnage" et "Famille". Peuvent-elles être considérées comme indépendantes ?
- Calculer la fréquence conditionnelle de la modalité "Mort" (pour la variable "État du personnage") sachant que le personnage appartient à la famille Stark. Calculer la fréquence conditionnelle de la modalité "Mort" (pour la variable "État du personnage") sachant que le personnage appartient aux Lannister. Comparer.

Exercice 3 (10 points) On considère une automobile roulant sur une route mouillée à une certaine vitesse x . Cette automobile freine brusquement pour simuler un freinage d'urgence. On note y la distance de freinage qui est fonction de la vitesse x . Les données recueillies $\{(x_1, y_1), \dots, (x_{10}, y_{10})\}$ sont les suivantes :

Vitesse x (Km/h)	50	60	70	80	90	100	110	110	120	130
Distance y (m)	10	21	34	49	66	85	106	129	154	181

- Quelles sont les natures des deux variables étudiées ? Justifiez votre réponse.
- Tracer le nuage de points. Quel est le signe du coefficient de corrélation linéaire (sans effectuer de calculs) ? Justifiez votre réponse.
- Déterminer la droite de régression linéaire des distances en fonction des vitesses par la méthode des moindres carrés (c.-à-d. de la forme $y = ax + b$ avec $a, b \in \mathbb{R}$). Tracer la droite sur le nuage de points.

1. Analyse sur <http://regressing.deads핀.com/valar-morghulis-a-statistical-guide-to-deaths-in-game-1618282560>

4. Déterminer le coefficient de détermination R^2 de ce modèle linéaire. Peut-on considérer qu'il existe une dépendance linéaire forte entre y et x ?
5. Quelle est la distance estimée de freinage si la vitesse est de 95 Km/h ? 150 Km/h ?
6. On suppose qu'il existe une relation quadratique simple entre les deux variables, i.e. une relation de la forme $y = \alpha x^2 + \beta$ avec $\alpha, \beta \in \mathbb{R}$.
- (a) On donne les résultats numériques suivants à partir des données :

$$\frac{1}{10} \sum_{i=1}^{10} x_i^2 = 9100, \quad \frac{1}{10} \sum_{i=1}^{10} x_i^4 = 103558000 \quad \text{et} \quad \frac{1}{10} \sum_{i=1}^{10} x_i^2 y_i = 1008540.$$

Déterminer α et β par la méthode des moindres carrés.

- (b) Calculer le coefficient de détermination R^2 pour ce modèle quadratique. Quel modèle faut-il retenir ? Expliquer pourquoi.