Série 1

Statistique à deux dimensions

Flière IID

Exercice 1

Un zootechnicien s'est proposé d'étudier la relation qui existe entre l'injection d'hormone de croissance (GH) en UL/kg et le gain pondéral en kg chez un groupe de 20 moutons. Des mesures ont été prises sur chaque mouton. Les résultats de ces mesures ont été reportés dans le tableau suivant :

$Y \backslash X$	25	30	35	40
20	4	2	1	0
25	5	1	0	0
30	3	2	1	1

1. Calculer les moyennes marginales des variables X et Y.
2. Calculer les variances marginales et les écarts-types marginaux de X et Y.
3. Calculer la moyenne conditionnelle et l'écart-type conditionnel de la variable X lorsque $Y = 25$.
4. Calculer la covariance entre ces deux variables.
5. Calculer le coefficient de corrélation linéaire.

Exercice 2

On va étudier l'évolution du poids corporel (kg) chez les agneaux simples après le sevrage. Le poids, que l'on appellera X (kg), est considéré comme une variable continue et l'âge Y (jours) est considéré comme une variable discrète. Les données sont dans le tableau suivant :

$Y \backslash X$	150	170	190
$[30, 33[$	3	4	0
$[33, 36[$	0	3	2
$[36, 39[$	2	1	0
$[39, 42[$	3	0	2

1. Les mêmes questions que l'exercice précédent.
2. Calculer la moyenne conditionnelle et l'écart-type conditionnel de la variable X lorsque $Y = 190$.

Exercice 4

Dans cet exercice, tous les résultats seront donnés par leur valeur approchée arrondie à 10^{-3}. On a étudié la durée de vie d'un certain nombre d'équipements mécaniques identiques. Dans le tableau suivant, t_i représente la durée de vie (exprimée en heures) et $R(t_i)$ est le pourcentage d'équipements encore en service à la date t_i. Par exemple, pour $t_i = 100$, il reste 80 % des équipements en service puisque $R(t_i) = 0,80$.

t_i	100	200	300	400	500	600	750	1 000	1 500
$R(t_i)$	0,80	0,64	0,52	0,40	0,32	0,28	0,20	0,12	0,04

1. Représenter graphiquement le nuage de points de coordonnées $(t_i, R(t_i))$ dans un repère. Quelle forme prend le nuage de points ? Est-ce pertinent de faire de la régression linéaire ? Calculer tout de même le coefficient de corrélation linéaire de cette série.
2. Posons $y_i = \ln R(t_i)$ (on ne conservera que 2 chiffres après la virgule). Représenter graphiquement le nuage de points de coordonnées (t_i, y_i) dans un repère. Comparer avec la question précédente.
3. Calculer le coefficient linéaire de la série statistique de variables t et y et comparer avec celui obtenu à la première question.
4. Déterminer par la méthode des moindres carrés, une équation de la droite de régression D de y en t sous la forme $y = at + b$.
5. En déduire qu'il existe deux nombres réels positifs k et λ tels que, pour tout élément $t \in [100; 1\,500]$, $R(t) = ke^{-\lambda t}$.
6. Dans cette équation, on prend $k = 1$ et $\lambda = 0,002$. Déterminer le pourcentage d'équipements encore en service au bout de 900 heures de fonctionnement.

Exercice 5

Une épidémie s'est déclarée dans une ville de 200 000 habitants. On a d'abord supposé que chaque malade peut contaminer 5 personnes par jour.

1. Combien faut-il de temps pour que tous les habitants de la ville soient touchés ?
2. On a enregistré chaque jour le nombre de cas qui se sont déclarés. Au septième jour, le tableau des résultats réels a été comme suit :
   X_i représente le numéro du jour.
   Y_i représente le nombre de cas enregistrés.

X	1	2	3	4	5	6	7
Y	4	13	38	106	330	965	2920

   Ajuster la variable Y par la variable X à l'aide d'une fonction exponentielle de la forme $Y = B \cdot A^X$.
3. Si la capacité hospitalière de la ville est de 7 000 lits, à quel jour les services hospitaliers seront-ils dépassés ?
4. Combien de jours faut-il pour que tous les habitants de la ville soient atteints, si aucune mesure n'est prise pour stopper cette épidémie ?

Exercice 6

Une épidémie de typhoïde s'est déclarée dans une certaine région et chaque jour on compte le nombre de nouveaux malades. Le tableau suivant réunit les données des dix premiers jours :

X	1	2	3	4	5	6	7	8	9	10
Y	4	12	35	109	320	3	10	27	81	243

1. Calculer les moyennes arithmétiques des deux variables X et Y.
2. Calculer la variance de X.
3. Calculer la covariance entre X et Y.
4. Ajuster la variable Y par la variable X à l'aide d'une équation de la forme $Y = aX + b$.
5. Quel est le nombre de nouveaux malades (suivant le modèle linéaire) que nous devons attendre le 20$^{\text{e}}$ jour après le déclenchement de l'épidémie ?
6. Ajuster cette fois-ci la variable Y par la variable X à l'aide d'une fonction exponentielle de la forme $Y = B \cdot A^X$.
7. Quel est alors, suivant cette fonction exponentielle, le nombre de nouveaux malades que nous devons attendre le 20$^{\text{e}}$ jour après le déclenchement de l'épidémie ?

Fin