TD3 : Régression linéaire, χ^2 et corrélations des rangs

Exercice 1

Nous souhaitons établir une relation entre l'âge a d'un individu et sa taille t. Pour cela, nous avons mesuré les membres d'une population et nous avons regroupé les résultats par tranches d'âge identifiées par leurs moyennes. Cela donne les observations uniformément pondérées suivantes :

k	1	2	3	4	5
a_k	18	20	22	24	26
t_k	176.1	178.1	178.8	179.9	181.2

1. Calculer $\mathrm{Var}(a)$, $\mathrm{Var}(t)$ et $\mathrm{Cov}(a, t)$.
2. Comment s'interprète le signe de $\mathrm{Cov}(a, t)$?
3. Donner l'équation de la droite de régression de t en fonction de a.
4. Prédire la taille d'un individu de 30 ans et celle d'un individu de 75 ans. Commenter.

Exercice 2

Le responsable d'une chaîne de magasins de bricolage pense qu'il y a une relation entre le nombre de personnes qui s'installent dans la région et le chiffre d'affaire des magasins. Il a noté, pour chacune des 10 dernières années, la valeur n_k du nombre de personnes ayant déménagé pour s'installer dans la région pendant l'année et c_k le chiffre d'affaire cumulé de l'ensemble des magasins de la chaîne lors de l'année. Le tableau suivant reproduit ces valeurs :

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
n_k	5.2	4.6	7.3	8.2	6.4	7.8	3.6	4.9	2.6	3.7
c_k	25.85	28.30	31.68	36.98	31.89	34.59	24.14	23.11	18.60	24.72

1. Représenter le nuage de points associé à ces observations.
2. Vu le graphique, vous semble-t-il y avoir une relation entre les deux variables ? De quelle nature ? Justifier votre réponse en calculant une certaine quantité et en la commentant.
3. Déterminer la droite de régression linéaire et tracer la sur le graphique. Est-ce que le modèle linéaire vous semble correct ?
4. En 2011, la région prévoit une valeur de n égale à 4.9. A quelle valeur de c le responsable peut-il s'attendre. Comparer votre réponse avec le chiffre d'affaire obtenu en 2008 et commenter la différence.

Exercice 3

Les observations de deux variables qualitatives sur une population de 12 individus ont donné les résultats suivants :

B	A	C	C	C	A	A	C	B	C	B	C
X	X	Y	X	X	Y	X	Y	X	Y	X	X

1. Représenter ces données par un tableau d'effectifs n_{ij}.

2. Déterminer le tableau théorique n^*_{ij} de données indépendantes donnant les mêmes effectifs en lignes et colonnes.
3. Calculer le χ^2 d'écart à l'indépendance. Commenter.

Exercice 4

Nous nous intéressons à l'existence d'une relation entre le sexe d'une personne (♂ pour un homme et ♀ pour une femme) et la couleur de ses cheveux (Br pour brun, Bl pour blond et R pour roux). Pour réaliser cette étude, nous demandons à 19 individus uniformément pondérés ces deux informations et nous obtenons les résultats suivants :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
♀	♀	♂	♂	♂	♂	♂	♀	♀	♀	♂	♀	♀	♀	♂	♀	♀	♂	♀
Br	R	Bl	R	Br	Bl	Bl	Bl	Br	Br	Bl	R	Bl	R	Bl	Br	Bl	Br	Bl

1. Donner la table de contingence des effectifs associés à ces observations.
2. En vous basant sur les marges observées, déterminer la table de contingence des effectifs théoriques en cas d'indépendance entre les variables.
3. Calculer la distance du χ^2 à l'indépendance et commenter le résultat.

Exercice 5

Nous nous intéressons à l'existence d'une relation entre le QI d'une personne et la taille de sa tête. Pour cela, nous avons calculé le QI de 5 individus et nous avons mesuré leurs crânes. Voici les rangs associés à ces mesures :

Personne	A	B	C	D	E
Rang pour le QI	1	2	3	4	5
Rang pour la taille de la tête	3	5	2	1	4

1. Calculer le coefficient de corrélation ρ_S de Spearman.
2. Calculer le coefficient de corrélation ρ_K de Kendall.

Exercice 6

Cinq produits sont soumis à une batterie de dix tests. Dans le tableau ci-dessous, on retrouve leurs prix et le nombre de tests auxquels ils ont resité.

Prix	100	80	500	2000	50
Tests	6	2	10	9	4

1. Calculer le coefficient ρ_S de corrélation de Spearman en représentant le tableau des rangs centrés et en calculant la variance et la covariance des rangs.
2. Calculer ρ_S par la formule

$$\rho_S = 1 - \frac{6}{n(n^2-1)} \sum_{k=1}^{n} (r_k - s_k)^2$$

3. Calculer le coefficient ρ_K de corrélation des rangs de Kendall.

Exercice 7

En classant les valeurs de deux variables observées sur n individus, on obtient des rangs :

$$r_1 = 1,\ r_2 = 2,\ \ldots,\ r_n = n$$

et

$$s_1 = n,\ s_2 = 1,\ s_3 = 2,\ \ldots,\ s_n = n - 1\ .$$

1. Calculer ρ_S. Vérifier le cas $n = 5$ en calculant directement la covariance avec les rangs centrés.
2. Calculer ρ_K.

Exercice 8

En introduisant deux ex-aequo, on obtient des rangs

$$r_1 = 1,\ r_2 = 2.5,\ r_3 = 2.5,\ r_4 = 4,\ r_5 = 5,\ \ldots,\ r_n = n\ .$$

Pour calculer $\mathrm{Var}(r)$ avec des poids uniformes normalisés $p_k = 1/n$, on s'aide du calcul de $\mathrm{Var}(s) = (n^2 - 1)/12$ où $s_k = k$. On considère un premier groupe avec les deux ex-aequo $\{2, 3\}$ et un second groupe avec tous les autres indices.

1. Donner les poids q_1 et q_2 des deux groupes.
2. Comparer les moyennes dans le premier groupe $\bar{r}_1$ et $\bar{s}_1$. Qu'en est-il pour les moyennes $\bar{r}_2$ et $\bar{s}_2$ dans le second groupe ? En déduire que les variances inter-groupes de r et s sont égales.
3. Calculer les variances dans le premier groupe $\sigma_1^2(r)$ et $\sigma_1^2(s)$. Expliquer pourquoi les variances $\sigma_2^2(r)$ et $\sigma_2^2(s)$ dans le second groupe sont égales. Que pouvez-vous déduire sur les variances intra-groupes de r et s.
4. Calculer $\mathrm{Var}(r)$.

Attention : les seuls calculs de cet exercice sont ceux de $\sigma_1^2(r)$ et $\sigma_1^2(s)$. Ne vous lancez pas dans des calculs fastidieux et inutiles...