1^{ere} Année Ingénieur

Analyse de la variance et modèle linéaire

Statistiques TD 8

1 Forêts

On observe $X = (X_{ij})_{1\leq j\leq n_i\ ,\ 1\leq i\leq 3}$ où X_{ij} suit une loi $\mathcal{N}(m_i, \sigma^2)$ et représente la hauteur du $j^{\text{ième}}$ arbre de la forêt i. On fournit les données suivantes.

Forêt 1 :	23,4	24,4	24,6	24,9	25,0	26,2	26,3	26,8	26,8	26,9
	27,0	27,6	27,7							
Forêt 2 :	22,5	22,9	23,7	24,0	24,4	24,5	25,3	26,0	26,2	26,4
	26,7	26,9	27,4	28,5						
Forêt 3 :	18,9	21,1	21,2	22,1	22,5	23,5	24,5	24,6	26,2	26,7

Tester l'égalité des moyennes et donner des intervalles de confiance pour les moyennes.

2 Confidence set

Soient $(Y_i)_{1\leq i\leq n}$ des variables aléatoires réelles indépendantes, chacune des Y_i suivant une loi $\mathcal{N}(\alpha+\beta x_i\ ,\ \sigma^2)$. Estimer α, β et σ^2. Donner une région de confiance pour (α, β). Tester $\beta = 0$.

3 Cholestérol

La composition en cholestérol des lipides du système nerveux central de 4 spécimens dans 5 espèces différentes est donnée dans le tableau suivant (en mole/mg de lipide):

espèces	1	2	3	4
Humain	0.7	0.73	0.68	0.72
Boeuf	0.58	0.62	0.65	0.56
Rat	0.68	0.64	0.71	0.67
Lapin	0.63	0.65	0.67	0.60
grenouille	0.64	0.66	0.69	0.67

Analyser les données pour découvrir si il y a une différence réelle dans le taux de cholestérol entre les espèces.

4 Poulets

On a mesuré la consommation d'eau de 20 poulets alors qu'ils étaient soumis à 5 régimes alimentaires différents: Consommation d'eau (ml /h)

Régime A	9.2	8.6	9.3	9.1
Régime B	8.3	8.6	8.7	8.0
Régime C	7.6	7.3	8.2	7.9
Régime D	7.2	6.5	7.6	6.3
Régime E	6.2	7.2	6.8	6.1

Construire une analyse de la variance de ce tableau pour dire si la consommation d'eau dépend du régime alimentaire.

5 Laboratoire

Pour tester la possible différence entres divers laborantins dans un examen de microscopie en laboratoire, on dispose de 4 préparations. Chaque laborantin compte le nombre présent d'un certain type de cellules. On obtient le tableau suivant :

Laborantins	1	2	3	4	5
A	48	51	44	75	63
B	62	58	57	96	75
C	57	54	55	83	78
D	39	50	44	70	66

Y-a-t-il une différence significative entre les laborantins?

6 Modèle linéaire

Soit le modèle $Y = X\theta + \epsilon, \theta \in R^k$. X est une matrice $n \times k$ de rang plein et $\epsilon = (\epsilon_i)_{i=1\ldots n}$ est une suite i.i.d de loi $\mathcal{N}(0, \sigma^2)$.

Quel est l'estimateur du maximum de vraisemblance $(\hat{\theta}, \hat{\sigma}^2)$ de (θ, σ^2)? Que peut-t-on dire de $\hat{\theta}$ et de $\hat{\sigma}^2$ et donner un estimateur sans biais s^2 de σ^2 multiple de $\hat{\sigma}^2$. quelles sont les distributions de : $s^2 \frac{n-k}{\sigma^2}$, $\frac{\|X(\hat{\theta}-\theta)\|}{\sigma^2}$, et $\frac{\frac{\|X(\hat{\theta}-\theta)\|}{k}}{s^2}$. En Déduire une région de confiance pour θ.

7 Régression polynômiale

Soient p et n deux entiers tels que $0 \leq p < n$. Pour tout $1 \leq j \leq n$ on pose $x_j = j - \frac{n+1}{2}$. On considère $\psi_0, \psi_1, \ldots, \psi_p$ polynômes fixés tels que :

- degré$(\psi_i) = i, \ \forall 0 \leq i \leq p$
- $\sum_{j=1}^{n} \psi_i(x_j)\psi_k(x_j) = 0, \ \forall 0 \leq i \neq k \leq p$

On a en vue d'étudier le modèle de régression

$$Y_j = \sum_{i=0}^{p} \lambda_i \psi_i(x_j) + \varepsilon_j, \quad 1 \leq j \leq n$$

où $\varepsilon_1, \ldots, \varepsilon_n$ sont indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$ et $\lambda_0, \ldots, \lambda_p, \sigma^2$ sont des paramètres inconnus.

a) Estimer les paramètres de ce modèle.

b) Ecrire un test de niveau α de "$\lambda_p = 0$" contre "$\lambda_p \neq 0$".

Le revenu net par action de la compagnie Gillette pour les années 57 à 64 est le suivant :

Année (z_j) :	1957	1958	1959	1960	1961	1962	1963	1964
Revenu en \$ ($Y_j(\omega)$) :	0,93	0,99	1,11	1,33	1,52	1,60	1,47	1,33

On pose $x_j = z_j - 1960,5$ puis $\psi_0(x) = 1, \psi_1(x) = 2x$ et $\psi_2(x) = x^2 - 21/4$.

c) Vérifier brièvement que le choix ci-dessus de $x_j, \psi_0, \psi_1, \psi_2$ rentre dans le cadre décrit plus haut.

d) En supposant que $Y_j = \sum_{i=0}^{2} \lambda_i \psi_i(x_j) + \varepsilon_j, \quad 1 \leq j \leq 8$ avec $\varepsilon_1, \ldots, \varepsilon_8$ indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$, estimer $\lambda_0, \lambda_1, \lambda_2$ et σ^2. Tester si $\lambda_2 = 0$.

e) Faire une prévision pour le revenu net par action en 1965.

<u>N.B.</u> Pour faciliter les calculs, on indique les valeurs numériques suivantes :
$\sum_j \psi_1^2(x_j) = \sum_j \psi_2^2(x_j) = 168$; $\sum_j Y_j^2(\omega) = 13,65$; $\sum_j \psi_0(x_j) Y_j(\omega) = 10,28$;
$\sum_j \psi_1(x_j) Y_j(\omega) = 6,86$; $\sum_j \psi_2(x_j) Y_j(\omega) = -4,1$.