UPS TOULOUSE III **Le 15/01/2003**
MAGISTÈRE

EXAMEN
MODÉLISATION STATISTIQUE

Durée 3 heures

PROBLÈME I

6 points

On appelle fréquence seuil d'un sportif amateur, sa fréquence cardiaque obtenue après trois quarts d'heure d'un effort soutenu de course à pied. Elle est mesurée à l'aide d'un cardio-fréquence-mètre. On cherche à savoir si l'âge d'un sportif a une influence sur sa fréquence seuil. On dispose des résultats suivants où x_i représente l'âge du sportif et y_i sa fréquence seuil.

x_i	30	54	29	51	36	41	40	23	49	30
y_i	175	165	169	172	170	170	167	170	166	167

x_i	32	22	22	32	44	34	32	20	46	45
y_i	177	169	172	173	168	169	170	172	175	168

On suppose chaque y_i est une réalisation d'une variable aléatoire Y_i de loi $\mathcal{N}(a + bx_i, \sigma^2)$.

1) Estimer les paramètres inconnus a, b et σ^2.

2) Proposer un intervalle de confiance à 95% pour a et pour b.

3) Tester la significativité de l'âge du sportif sur sa fréquence seuil.

4) Pour une nouvelle valeur $x^* = 65$, calculer la prédiction naturelle $\widehat{Y}^*$ de Y^* et trouver un intervalle de prévision à 95% pour Y^*.

PROBLÈME II

3 points

On considère la régression linéaire multiple

$$Y_i = a + b_1 x_{i1} + b_2 x_{i2} + \varepsilon_i, \qquad i = 1, 2, \cdots, n$$

où (ε_i) est une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$. On lui associe les résultats matriciels suivants: $\| Y \| = 38$,

$$X^t X = \begin{pmatrix} 50 & 0 & 2 \\ 0 & 20 & -5 \\ 2 & -5 & 4 \end{pmatrix} \quad \text{et} \quad X^t Y = \begin{pmatrix} 250 \\ 40 \\ 10 \end{pmatrix}.$$

1) Préciser les lois des estimateurs des moindres carrés $\widehat{\theta} = (\widehat{a}, \widehat{b})$ et $\widehat{\sigma}^2$ de $\theta = (a, b)$ et σ^2 puis calculer les résultats numériques associés à ces estimateurs.

2) Tester l'hypothèse $b_2 = 4$ contre $b_2 < 4$ puis l'hypothèse $a + 2b_1 = 10$ et $4b_1 - b_2 = 8$ contre $a + 2b_1 \neq 10$ ou $4b_1 - b_2 \neq 8$.

PROBLÈME III

4 points

On cherche à savoir si le mode de fabrication d'une ampoule électrique est différent selon la marque en terme de la durée de vie de l'ampoule. On dispose des durées de vie suivantes obtenues pour six marques différentes notées $A_1, A_2, \cdots, A_6$ sur 36 ampoules.

A_1	1602	1615	1617	1624				
A_2	1480	1482	1485	1493	1500	1507	1510	
A_3	1548	1555	1559	1563	1575			
A_4	1435	1438	1448	1449	1454	1458	1467	1475
A_5	1493	1498	1500	1502	1509	1510		
A_6	1596	1599	1602	1604	1612	1620		

On désigne par y_{ij} la durée de vie dans la marque A_i de la j^e ampoule. On suppose que pour $i = 1, 2, \cdots, 6$, y_{ij} est une réalisation d'une variable aléatoire Y_{ij} satisfaisant

$$Y_{ij} = m_i + \varepsilon_{ij}$$

où (ε_{ij}) est une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$.

1) Estimer les paramètres inconnus m et σ^2.

2) Tester l'hypothèse d'égalité de la durée moyenne de vie pour les six marques en dressant le tableau d'analyse de la variance à un facteur.

PROBLÈME IV

7 points

On souhaite étudier l'efficacité de trois types d'aliment A_1, A_2, A_3 et de trois conditions d'élevage B_1, B_2, B_3 sur la gavage des canards. Pour chaque type d'aliment et chaque condition d'élevage, on dispose d'un échantillon de quatre canards et on mesure le poids en grammes des foies gras obtenus après gavage.

Poids des foies	B_1	B_2	B_3
A_1	620 612 615 627	670 631 628 614	704 679 687 707
A_2	586 618 604 614	662 678 597 619	711 678 682 649
A_3	591 622 579 574	623 647 580 607	671 623 597 615

On désigne par y_{ijk} le poids du foie du k^e canard ayant reçu l'aliment A_i sous la condition d'élevage B_j. On suppose que y_{ijk} est une réalisation d'une variable aléatoire Y_{ijk} satisfaisant pour $i = 1, 2, 3$ et $j = 1, 2, 3$

$$Y_{ijk} = \mu + a_i + b_j + c_{ij} + \varepsilon_{ijk} \qquad k = 1, 2, 3, 4$$

où (ε_{ijk}) est une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$. On suppose également que les contraintes standards sur la décomposition de la moyenne sont réalisées.

1) Estimer les paramètres inconnus μ, a, b, c et σ^2.

2) Dresser le tableau d'analyse de la variance à deux facteurs sachant que

$$
\begin{aligned}
S_R &= \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{4}(Y_{ijk} - Y_{ij*})^2 = 15\,918.25, \\
S_T &= \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{4}(Y_{ijk} - Y_{***})^2 = 52\,234.75.
\end{aligned}
$$

2) Tester l'hypothèse selon laquelle il n'y a pas d'interaction.

3) Tester l'hypothèse d'absence d'effet de chacun des 2 facteurs.