Master math et math-info 2012-2013
Travaux dirigés de Statistique 7
Tests d'un sous modèle linéaire et Analyse de la variance

Exercice 1 *On considère le modèle de régression multiple:*

$$Y_j = a + bx_j + cz_j + \epsilon_j, \quad j = 1 \dots 4n.$$

Où:

- *a, b, c sont des paramètres inconnus.*
- $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{4n} \end{pmatrix} \sim \mathcal{N}_{4n}(0, \sigma^2 I_{4n})$, *la variance σ^2 est inconnue.*
- $x_{4j-3} = -1$, $x_{4j-2} = 1$, $x_{4j-1} = -1$, $x_{4j} = 1$, $z_{4j-3} = 1$, $z_{4j-2} = 1$, $z_{4j-1} = -1$, $z_{4j} = -1$, $(j = 1, \dots, n)$.

1. *Estimer les paramètres a, b, c, σ^2. Dans la suite on note S^2 l'estimateur de σ^2.*
2. *Tester l'hypothèse $a = 0$. Application numérique: $n = 50$, $S^2_{obs} = 2.0$, $\overline{Y}_{obs} = 0.17$.*
3. *Tester l'hypothèse $b = c$.*

Exercice 2 *On considère les modèles de régression :*

$$Y_{ij} = a_i + b_i x_{ij} + \varepsilon_{ij}, \ i \in \{1, 2\} \text{ et } 1 \le j \le n,$$

où $(\varepsilon_{ij})_{i \in \{1,2\},\, 1 \le j \le n}$ est une famille de variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2)$, $\sigma > 0$.

1. *Donner des estimateurs sans biais de a_1, b_1, a_2, b_2 et σ^2. Dans la suite on note S^2 l'estimateur de σ^2.*
2. *Donner la loi de $(\hat{a}_1, \hat{a}_2)^T$ et tester $a_1 = a_2 = 0$.*
3. *Tester $a_1 = a_2$ et $b_1 = b_2$.*

Exercice 3 *Le tableau ci-dessous résume une enquête sur la consommation quotidienne en grammes de confiture dans un certains pays d'europe.*

	France	*Angletere*	*Allemagne*	*Italie*
n_i	30	35	20	10
$X_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$	47.3	50.2	51	46
$\frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - X_{i.})^2$	4.02	3.18	3.96	4.5

On prend comme modèle : $X_{ij} = m_i + \varepsilon_{ij}, \quad 1 \leq j \leq n_i$ *et* $1 \leq i \leq 4$, *où* ε_{ij} *est une suite de variables aléatoires indépendantes de loi* $\mathcal{N}(0, \sigma^2)$.

1. *Quel est l'estimateur du maximum de vraisemblance* $\hat{\theta}$ *de* $\theta := (m_1, m_2, m_3, m_4)^T$. *Vérifier que* $\hat{\theta}$ *est sans biais.*
2. *Quel est l'estimateur du maximum de vraisemblance* $\hat{\sigma}^2$ *de* σ^2. *Vérifier que* $\hat{\sigma}^2$ *est biaisé et donner un estimateur sans biais* s^2 *de* σ^2. *Donner la loi de* $\frac{(n-k)s^2}{\sigma^2}$.
3. *Rappeler brièvement le principe de la méthode de l'analyse de la variance. La consommation de confiture dépend-t-elle du pays.*

Exercice 4 *Le tableau suivant représente des mesures faites dans trois forêts de la hauteur de certains arbres.*

	Forêt 1	*Forêt 2*	*Forêt 3*
	23.4	22.5	18.9
	24.4	22.9	21.1
	24.6	23.7	21.2
	24.9	24.0	22.1
	25.0	24.4	22.5
	26.2	24.5	23.5
	26.3	25.3	24.5
	26.8	26.0	24.6
	26.8	26.2	26.2
	26.9	26.4	26.7
	27.0	26.7	
	27.6	26.9	
	27.7	27.4	
		28.5	
n_i	13	14	10
$\sum Y_{ij}$	337.6	355.4	231.3
$\sum Y_{ij}^2$	8789.36	9062.96	5403.51

Dans la $i^{\text{è}}$ *forêt,* n_i *arbres sont mesurés, de hauteur* $Y_{i1}, \ldots, Y_{in_i}$ *On suppose que la hauteur des arbres se distribue selon le modèle*

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

où m_i *est déterministe, et les* ε_{ij} *sont des variables aléatoires indépendantes normales centrées de variance* σ^2.

1. *Donner des estimateurs non biaisés des moyennes* m_i.
2. *Donner un estimateur sans biais de* σ^2.
3. *Donner un intervalle de confiance au niveau* 0.99 *pour chaque moyenne* m_i.
4. *Faire un test au niveau de risque* 0.05 *pour l'égalité des trois moyennes.*
5. *Peut-on affirmer, avec un niveau de risque de* 0.05*, que deux des trois moyennes sont égales?*