TD n° 7 : ANOVA à 1 facteur

Exercice 1. Une enquête sur la consommation annuelle des ménages est réalisée par l'INSEE régulièrement. Ces ménages sont répartis en 5 grandes catégories suivant leur localisation :

$C1$: ménages en zone rurale,

$C2$: ménages résidant dans une unité urbaine inférieure à 20000 habitants,

$C3$: ménages résidant dans une unité urbaine comprise entre 20000 habitants et 100000 habitants,

$C4$: ménages résidant dans une unité urbaine supérieure à 100000 habitants autre que l'agglomération parisienne,

$C5$: ménages résidant dans l'agglomération parisienne.

Un groupement commercial s'intéresse particulièrement à la consommation annuelle des produits contenus dans la nomenclature 17 de l'INSEE c'est-à-dire, la consommation annuelle en mouton, agneau et chevreau et il souhaite savoir s'il y a un effet "localisation" sur la consommation annuelle moyenne des ménages pour ces produits. Le groupement commercial interroge 5 ménages par catégories. Les résultats en euro sont :

C1	C2	C3	C4	C5
56	47	55	61	69
66	50	51	62	71
54	55	59	54	55
61	46	54	54	62
56	56	59	62	53

On suppose que, pour tout $i \in \{1, \ldots, 5\}$, la consommation annuelle d'un ménage en euro de catégorie Ci peut être modélisée par une *var* X_i suivant la loi normale $\mathcal{N}(\mu_i, \sigma^2)$, avec μ_i et σ inconnus.

1. On donne $\text{sce}_T = 908.64$ et $\text{sce}_R = 556.40$. Dresser le tableau ANOVA.
2. Effectuer, au risque 5%, le test ANOVA :

$$H_0 : \mu_1 = \ldots = \mu_5 \qquad \text{contre} \qquad H_1 : \text{"il existe au moins 2 moyennes différentes"}$$

   Interpréter le résultat.
3. Peut-on affirmer, au risque 5%, que $\mu_3 \neq \mu_4$?

Exercice 2. Un organisme de certification souhaite établir des coefficients de fidélité pour la quantité de *glucosinolate* exprimée en micromoles par gramme contenue dans le Colza. Il fait appel à 5 laboratoires différents : $L1$, $L2$, $L3$, $L4$ et $L5$.

Les résultats sont :

$L1$	$L2$	$L3$	$L4$	$L5$
25.3	28.8	24.2	25.6	26.0
25.0	27.8	26.8	25.4	25.2
25.0	26.4	26.3	24.8	25.6
25.9		25.3	25.0	
25.2				

On suppose que, pour tout $i \in \{1, \ldots, 5\}$, la quantité de *glucosinolate* en micromoles par gramme contenue dans le Colza pour le laboratoire Li peut être modélisée par une *var* X_i suivant la loi normale $\mathcal{N}(\mu_i, \sigma^2)$, avec μ_i et σ inconnus. Le tableau ANOVA, incomplet, est reproduit ci-dessous :

	sce	ddl	cm	f_{obs}
Total	21.5811			
Factoriel				
Résiduel	8.1447			

1. Reproduire et compléter les cases nécessaires de ce tableau.
2. Est-ce que les laboratoires trouvent, en moyenne, des résultats significativement différents ?
3. Peut-on affirmer que, en moyenne, les résultats de $L2$ et $L4$ diffèrent significativement ?

Exercice 3. Un industriel s'intéresse à la tendreté des haricots qu'il mettra en boites. Pour mesurer cette tendreté, on met les haricots en bottes (chaque botte ayant le même nombre de haricots) et on mesure le travail qu'il faut à une trancheuse pour couper la botte. L'indice de tendreté est proportionnel à ce travail. Les haricots peuvent provenir de fournisseurs différents F1, F2, F3 et F4. Les résultats sont :

$F1$	$F2$	$F3$	$F4$
8	9	10	6
9	10	9	8
9	9	12	8
6	10	11	7

On suppose que, pour tout $i \in \{1, \ldots, 4\}$, l'indice de tendreté pour le fournisseur Fi peut être modélisé par une *var* X_i suivant la loi normale $\mathcal{N}(\mu_i, \sigma_i^2)$, avec μ_i et σ_i inconnus.

1. Peut-on affirmer, au risque 5%, que la dispersion de l'indice de tendreté diffère selon les producteurs ?
2. Le tableau ANOVA, incomplet, est reproduit ci-dessous :

	sce	ddl	cm	f_{obs}
Total	40.4375			
Factoriel	25.6875			
Résiduel				

   Reproduire et compléter les cases nécessaires de ce tableau.

3. Peut-on dire, au risque 5%, que l'indice de tendreté moyen différe suivant la provenance des haricots ?