

EXERCICES SUR LE CHAPITRE 3 – 2^{ème} PARTIE

La comparaison de plusieurs moyennes – ANOVA I

FORMULES

- $F_{obs} = \frac{\sum_{k=1}^K \frac{n_k(\bar{y}_k - \bar{y})^2}{K-1}}{s_p^2} = \frac{MC(Facteur)}{MC(Erreur)} < F_{K-1, n-K}(0.95)$ avec $s_p^2 = \frac{(n_1-1)s_1^2 + \dots + (n_k-1)s_k^2}{n-K}$
- $(\mu_k - \mu_j | données) \sim t_{n-K}(\bar{y}_k - \bar{y}_j, s_p^2 \left(\frac{1}{n_k} + \frac{1}{n_j}\right))$
- $(\bar{y}_k - \bar{y}_j) \pm t_{n-K}(0.975) \sqrt{s_p^2 \left(\frac{1}{n_k} + \frac{1}{n_j}\right)}$

EXERCICES À FAIRE À LA SÉANCE DE TP

1. Afin de vérifier si les hôtels à New-York sont plus chers que les hôtels d'autres grandes villes, 8 hôtels ont été sélectionnés aléatoirement dans 3 grandes villes d'après le guide du voyageur de 2005 (comparaison sur base du prix en dollars d'une chambre double classique). Ces villes sont Washington DC (DC), Los Angeles (LA) et New-York (NY). Voici les résultats obtenus :

Ville	n	\bar{y}_k	s_k^2
DC	8	132,875	42,52
LA	8	130,125	32,43
NY	8	195,125	45,64

- a. Évaluez la plausibilité que les prix moyens d'une chambre double classique soient identiques dans les trois grandes villes.
 - b. Peut-on conclure que les hôtels à New-York sont plus chers que les hôtels d'autres grandes villes?
2. Une étude a été menée afin de vérifier l'hypothèse selon laquelle il est plus difficile de s'adapter au décalage horaire lorsque l'on voyage vers l'est. Dix-huit personnes se sont vues offrir un voyage en avion et ont été réparties au hasard en 3 groupes de 6 personnes. Six personnes ont voyagé vers l'est, 6 vers l'ouest et 6 dans le même fuseau horaire. Les trois groupes ont volé pendant la même durée. Voici le nombre de jours nécessaires aux voyageurs pour s'habituer au décalage horaire:

Ouest	2	1	3	3	2	4
Est	6	4	6	8	5	7
Même Fuseau	1	0	1	1	0	0

- a. Évaluez la plausibilité que le nombre moyen de jours nécessaires aux voyageurs pour s'habituer au décalage horaire soit le même quelle que soit la direction vers laquelle on voyage.
- b. Comparez, si nécessaire, les difficultés d'adaptation au décalage horaire selon la direction vers laquelle on voyage.

3. Une étude a été menée en vue de comparer les loyers mensuels d'appartements récents et meublés dans quatre régions américaines (Nord-est, Midwest, Sud et Ouest). Voici la table d'analyse de variance produite à partir du résultat de quatre échantillons:

Tests Univariés de Significativité de Var2 (Feuille de données1) Modèle sur-paramétré Décomposition de Type III					
Effet	SC	Degré de Liberté	MC	F	p
Ord.Orig.	3585139	1	3585139	229,0229	0,000000
"Var1"	61537	3	20512	1,3103	0,305516
Erreur	250465	16	15654		

- Que pouvez-vous en déduire quant au lien entre le loyer mensuel d'un appartement récent et meublé et la région américaine de laquelle il provient?
 - Serait-il intéressant de comparer les régions 2 à 2 par rapport au loyer mensuel d'un appartement récent et meublé?
4. Une étude sur l'efficacité d'un projet de contrôle des infections nosocomiales (infection attrapée à l'hôpital) a été menée dans le but de vérifier si les programmes de surveillance et de contrôle de l'infection ont permis de réduire la proportion d'infections nosocomiales aux Etats-Unis. Les données consistent en un échantillon aléatoire de 113 hôpitaux provenant de quatre régions américaines (28 hôpitaux du Nord-est, 32 hôpitaux du Midwest, 37 hôpitaux du Sud et 16 hôpitaux de l'Ouest). Intéressons-nous plus particulièrement à la variable traduisant la probabilité d'attraper une infection nosocomiale dans chaque hôpital (exprimée en %). Voici la table d'analyse de variance étudiant cette variable en fonction des régions:

Tests Univariés de Significativité de Var4 (APC1) Modèle sur-paramétré Décomposition de Type III					
Effet	SC	Degré de Liberté	MC	F	p
Ord.Orig.	1971,035	1	1971,035		0,000000
"Var9"	13,997	3	4,666		0,048386
Erreur	187,383	109	1,719		

- Cette table est incomplète. Déterminez les valeurs manquantes. Déduisez-en si la probabilité moyenne d'acquérir une infection nosocomiale varie selon la région.
- S'il y a lieu, utilisez l'information suivante pour déterminer les régions qui diffèrent.

"Var9"; Moy. Pondérées (APC1) Effet courant : F(3, 109)=2,7140, p=,04839 Décomposition de Type III						
N°Cellu.	Var9	Var4 Moy.	Var4 Err-Type	Var4 -95,00%	Var4 +95,00%	N
1	1	4,860714	0,240224	4,367816	5,353612	28
2	2	4,393750	0,236743	3,910910	4,876590	32
3	3	3,927027	0,239859	3,440571	4,413483	37
4	4	4,381250	0,219131	3,914184	4,848316	16

5. Un économiste décide d'étudier l'augmentation de production d'équipement électronique pour ordinateur sur un échantillon de firmes produisant ce type d'équipement. L'amélioration de productivité est mesurée sur une échelle de 0 à 100. De plus, les firmes sont classées selon les moyens financiers qu'elles ont mis à disposition de leur secteur « recherche et développement ». Ces moyens financiers ayant été catégorisés pour former K groupes. Voici la table d'analyse de variance étudiant l'augmentation moyenne de productivité selon le groupe :

Tests Univariés de Significativité de Var1 (CH16PR10) Modèle sur-paramétré Décomposition de Type III					
Effet	SC	Degré de Liberté	MC	F	p
Ord.Orig.	1623,262	1			
"Var2"	20,125	2			
Erreur	15,362	24			

- À partir des informations disponibles dans cette table, déterminez le nombre de groupes traduisant les différents moyens financiers que les firmes ont mis à disposition de leur secteur « recherche et développement ».
 - Complétez cette table d'analyse de variance. Déduisez-en si l'augmentation moyenne de productivité varie selon le groupe (selon les moyens financiers attribués à la recherche et au développement).
 - Serait-il intéressant de comparer les groupes 2 à 2 par rapport l'augmentation moyenne de productivité?
6. Voici une table d'analyse de variance étudiant le nombre de jours nécessaires à un rétablissement complet après une opération du genou en fonction de la condition physique avant l'opération (catégorisée en trois groupes : en dessous de la moyenne, moyenne, au-dessus de la moyenne).

Tests Univariés de Significativité de Var1 (CH16PR12) Modèle sur-paramétré Décomposition de Type III					
Effet	SC	Degré de Liberté	MC	F	p
Ord.Orig.	22560,00	1			0,000000
"Var2"		2	336,00		0,000041
Erreur	416,00	21			

- Complétez cette table d'analyse de variance. Déduisez-en si le nombre moyen de jours nécessaires à un rétablissement complet après une opération du genou varie en fonction de la condition physique avant l'opération.
- Serait-il intéressant de comparer 2 à 2 les moyennes du nombre de jours nécessaires à un rétablissement complet après une opération du genou en fonction de la condition physique avant l'opération?

7. Reprenons l'étude de l'exercice 4 sur l'efficacité d'un projet de contrôle des infections nosocomiales (infection attrapée à l'hôpital). Les données consistaient en un échantillon aléatoire de 113 hôpitaux. Intéressons-nous plus particulièrement à l'effet de l'âge moyen des patients sur la probabilité d'acquérir une infection nosocomiale dans chaque hôpital (exprimée en %). Dans ce but, l'âge moyen des patients a été catégorisé en 4 groupes: <50 ans, ≥50 mais <55, ≥55 mais <60 et ≥60. Voici la table d'analyse de variance étudiant la probabilité d'acquérir une infection nosocomiale en fonction des catégories d'âge moyen:

Tests Univariés de Significativité de Var4 (APC1)					
Modèle sur-paramétré					
Décomposition de Type III					
Effet	SC	Degré de Liberté	MC	F	p
Ord.Orig.	.		.	650,1771	0,000000
"Var13"			.	0,5620	0,641197
Erreur			1,819		

- Cette table est incomplète. Déterminez les valeurs manquantes. Déduisez-en si la probabilité moyenne d'acquérir une infection nosocomiale varie selon la catégorie d'âge.
- Serait-il intéressant de comparer 2 à 2 les probabilités moyennes d'acquérir une infection nosocomiale selon les catégories d'âge?

EXERCICES COMPLÉMENTAIRES

1. On a mesuré la pression artérielle systolique au repos dans 3 groupes d'âge. On dispose d'échantillons indépendants de respectivement 10, 8 et 12 individus pour les groupes d'âge: 30-35 ans, 50-55 ans et 70-75 ans. Voici les résultats obtenus :

30-35 ans	11	11	9	14	12	11	12	11	11	15		
50-55 ans	15	10	14	10	17	10	19	14				
70-75 ans	16	16	18	15	17	14	15	16	15	15	13	12

- Ces données permettent-elles d'évaluer si la pression artérielle moyenne change avec l'âge ?
 - Comparez, si nécessaire, les groupes d'âge au niveau de leur pression artérielle moyenne.
2. On a mesuré le temps de survie de 15 patients atteints du SIDA. Parmi ces 15 patients, 5 ont reçu un médicament A, 5 un médicament B et 5 un médicament C contre le sida. Voici le temps moyen (mois) de survie pour chacun des trois groupes de patients (selon le médicament reçu): 41,8 mois, 39,8 mois et 68,6 mois respectivement. Voici les variances du temps de survie: 546,7 mois-carré, 549,7 mois-carré et 1829,3 mois-carré respectivement.
- Peut-on mettre en évidence une différence entre les temps de survie moyens sous les 3 traitements ? Quelles sont les hypothèses de travail ?
 - Comparez, si nécessaire, les temps moyens de survie selon les groupes.

3. Dans le cadre d'une étude écotoxicologique, la concentration en DDT et en ses dérivés a été mesurée chez des brochets de différents âges. Voici la table d'analyse de variance obtenue à partir de ces données:

Tests Univariés de Significativité de Var2 (Feuille de données4) Modèle sur-paramétré Décomposition de Type III					
Effet	SC	Degré de Liberté	MC	F	p
Ord.Orig.	2,589894	1	2,589894	702,6912	0,000000
"Var1"	0,181205	2	0,090602	24,5823	0,000003
Erreur	0,077399	21	0,003686		

- Déterminez le nombre de groupes de brochets.
 - Déterminez le nombre total d'observations.
 - La concentration moyenne en DDT et ses dérivés varie-t-elle avec l'âge des brochets?
 - Serait-il intéressant de comparer la concentration moyenne en DDT et ses dérivés selon les âges des brochets ?
4. On désire déterminer si la quantité moyenne de nitrate varie d'une station à une autre le long d'une rivière. Pour cela, on prélève en 10 points une certaine quantité d'eau et cela dans trois stations différentes. Voici la table d'analyse de variance étudiant cette variable en fonction des stations:

Tests Univariés de Significativité de Var2 (Feuille de données1) Modèle sur-paramétré Décomposition de Type III					
Effet	SC	Degré de Liberté	MC	F	p
Ord.Orig.			944832,5	151,5909	0,000000
"Var1"	1732,5				
Erreur					

- Cette table est incomplète. Déterminez les valeurs manquantes. Déduisez-en si la quantité moyenne de nitrate varie d'une station à une autre.
 - Serait-il intéressant de comparer les niveaux moyens de nitrate des différentes stations?
5. On désire étudier l'effet de la région géographique sur le taux de criminalité des villes aux Etats-Unis en 1977. Les données consistent en un échantillon aléatoire de 141 villes provenant de quatre régions américaines (27 villes du Nord-est, 35 villes du Midwest, 51 villes du Sud et 28 villes de l'Ouest). Voici la table d'analyse de variance:

Tests Univariés de Significativité de Var13 (APC2) Modèle sur-paramétré Décomposition de Type III					
Effet	SC	Degré de Liberté	MC	F	p
Ord.Orig.	409315,6				0,000000
"Var12"				28,289	0,000000
Erreur	18009,6				

- a. Complétez cette table d'analyse de variance. Déduisez-en si le taux moyen de criminalité varie selon la région.
- b. S'il y a lieu, utilisez l'information suivante pour déterminer les régions qui diffèrent.

"Var12"; Moy. Pondérées (APC2) Effet courant : F(3, 137)=28,289, p=,00000 Décomposition de Type III						
N°Cellu.	Var12	Var13 Moy.	Var13 Err-Type	Var13 -95,00%	Var13 +95,00%	N
1	1	42,20579	2,631874	36,79590	47,61569	27
2	2	52,77845	1,482272	49,76611	55,79079	35
3	3	57,41214	1,689356	54,01896	60,80531	51
4	4	70,08682	2,074974	65,82932	74,34431	28