ISARA Lyon Durée 1h

UP SFO- S4 – UE : Statistique

Epreuve N°2 - partie 6 (25%) – Mai 2012

Mme Bottollier Lemallaz CORRIGE EN BLEU

Conditions d’examen : **Documents**: Tables statistiques autorisées

**Calculatrice** : non programmable

Remarques importantes :

1. A la fin de l’épreuve vous devez rendre les 4 pages du sujet complétées de façon lisible et propre.
2. Sauf indication contraire, vous présenterez vos résultats avec 2 chiffres significatifs.

**Nom :……………………………………………………………….Prénom :……………………………………………….Note sur 20 :……….**

**Evaluation des connaissances de base (8 points) (20 min):**

1° ) On donne 4 modèles associés à une Analyse de la variance et 4 formules de calcul des résidus.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | |  | Modèle | | I | xir = µ + i + ir | | II | xijr = µ + i + j + i j + ijr | | III | xij = µ + i + j + ij | | IV | xijtr = µ + i + j + t + ij + it + jt + ij t + ijtr | |  | |  |  | | --- | --- | |  | Formule | | a | x ijr - | | b | x ijtr - | | c | xir - | | d | xij + µ -  - | |

Pour chacun de ces modèles, précisez le nombre de facteurs étudiés (k = 1, 2, 3 ……), avec ou sans répétitions (r= oui ou non), la formule du calcul des résidus ( = a, b, c ou d) , le nombre de tests d’homoscédasticité (h = 1, 2 , 3, ….), le nombre de critère statistique calculés dans le tableau de l’ANOVA (f = 1, 2 , 3, ….), en complétant le tableau ci-dessous .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Modèle | k | r |  | h | f |
| I | 1 | oui | c | 1 | 1 |
| II | 2 | oui | a | 3 | 3 |
| III | 2 | non | d | 2 | 2 |
| IV | 3 | oui | b | 7 | 7 |

2°) Que doit on conclure lorsque l’on obtient le type de résultat ci dessous ?

|  |  |  |
| --- | --- | --- |
| Coefficient | PROB | Formuler les hypothèses et conclusion pour chaque coefficient |
| Beta 1 | 0,045 | Pour le coefficient d’asymétrie :  H0 1 = 0 H1  1  0 (distribution non symétrique)  On a 4,5% de risque de se tromper en rejetant l’hypothèse de symétrie des résidus. Ce risque est faible, on a mis en évidence que la distribution des résidus n’était pas symétrique. |
| Beta 2 | 0,45 | Pour le coefficient d’aplatissement :  H0 2 = 3 H1  2  3 (distribution non mésocurtique)  On a 45% de risque de se tromper en rejetant l’hypothèse de mésocurtie des résidus. Ce risque est trop fort, on n’a pas mis en évidence que la distribution des résidus n’était pas mésocurtique. |
| Conclusion générale | | On ne retient pas l’hypothèse de normalité des résidus. |

3°) Pour vérifier l’homoscédasticité des résidus dans une étude à 2 facteurs étudiés A et B respectivement à 2 et 4 modalités sans répétition on doit déterminer les ddl et le terme C suivants (compléter le tableau) :

|  |  |  |  |
| --- | --- | --- | --- |
| Facteur A ou B ? | Pour le calcul du X² | Pour le X² théorique | Pour le F théorique |
| Facteur …**B**…  **8 unités au total donc nj= 2 par modalité, 4 modalités** | j **= 2-1 = 1** |  = **4-1=3** |  |
|  =**4x1 = 4** |  |
| C (\*)= **1,417……………au 1/1000ème** |  |
| Facteur …**A**…  **8 unités au total donc ni = 4 par modalité, 2 modalités** |  |  | 1 **= 4-1=3** |
|  |  | 2 = **4-1 =3** |

(\*) **= 1 + (1/(3x3))(4x1 – 1/4) = 1,417**

4°) Donner les valeurs des critères statistiques théoriques pour **un risque d’erreur égal à 10%.** Poser les hypothèses de chaque test et formuler les conclusions possibles.

|  |  |  |
| --- | --- | --- |
| Facteur | Critère théorique | hypothèses et conclusions |
| A | F0,95 (3,3)  = 9,28 | H0 : les 2 variances résiduelles selon le facteur A sont égales  H1 : les 2 variances résiduelles selon le facteur A sont différentes conclusions : Si F²calc < F² théo on conserve l’hypothèse d’homoscédasticité des variances résiduelles selon le facteur A.  Si F²calc > F² théo on rejette l’hypothèse d’homoscédasticité des variances résiduelles selon le facteur A avec moins de 10% d’erreur. |
| B | X²0,90 (3)  = 6,25 | H0 : les 4 variances résiduelles selon le facteur B sont homogènes  H1 : au moins une des 4 variances résiduelles est supérieure à une autre  conclusions : Si X²calc < X² théo on conserve l’hypothèse d’homoscédasticité des variances résiduelles selon le facteur B.  Si X²calc > X² théo on rejette l’hypothèse d’homoscédasticité des variances résiduelles selon le facteur B avec moins de 10% d’erreur. |

**Exercice d’application 1 (7points) (25 min) :**

Soit une étude à 2 facteurs étudiés A et B respectivement à 3 et 2 modalités et 5 répétitions.

On donne : xijr = 340,0 et xijr² = 3974,0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Somme des 5 répétitions par traitement AiBj | A1 | A2 | A3 | Total |
| *B1* | 46 | 60 | 61 | 167 |
| *B2* | 69 | 50 | 54 | 173 |
| *Total* | 115 | 110 | 115 | 340 |

Construisez le tableau de l’analyse de la variance, calculs au 1/100ème, hypothèses et conclusions pour un risque d’erreur égal à 5%.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  | B | ni | Somme | Moyenne | Produit |
| *B1* | A1 | A2 | A3 |  |  |  | *B1* | 15 | 167,00 | 11,13 | 1859,27 |
| ni | 5 | 5 | 5 |  |  |  | *B2* | 15 | 173,00 | 11,53 | 1995,27 |
| Somme | 46 | 60 | 61 |  |  |  |  |  |  | *total fact B =* | **3854,53** |
| Moyenne | 9,2 | 12 | 12,2 |  |  |  |  |  |  |  |  |
| Produit | 423,2 | 720 | 744,2 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | A | ni | Somme | Moyenne | Produit |
| *B2* | A1 | A2 | A3 |  | total A\*B |  | A1 | 10 | 115,00 | 11,50 | 1322,50 |
| ni | 5 | 5 | 5 |  | **3922,8** |  | A2 | 10 | 110,00 | 11,00 | 1210,00 |
| Somme | 69 | 50 | 54 |  |  |  | A3 | 10 | 115,00 | 11,50 | 1322,50 |
| Moyenne | 13,8 | 10 | 10,8 |  |  |  |  |  |  |  |  |
| Produit | 952,2 | 500 | 583,2 |  |  |  |  |  |  | *total fact A =* | **3855,00** |

|  |  |
| --- | --- |
| N = | 30 |
| C = | 30 (340/30)² = 3853 |

H0: CMA / CMe = 1 H1: CMA / CMe > 1

H0: CMB / CMe = 1 H1: CMB/ CMe > 1

H0: CMA\*B/ CMe = 1 H1: CMA\*B/ CMe > 1

SCE tot = 3974 – 3853= 121

SCE (A) = 3855 – 3853= 1, 67

SCE (B) = 3854, 53 – 3853= 1, 20

SCE (A\*B) = 3922, 8 – 3853- 1, 67 – 1, 20 = 66, 60

SCE e = 121 – 1,67 – 1,20 - 66,60 = 51,20

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *variations* | *SCE* | *ddl* | *CM* | *F calculé* | *F0,95* |  |
| Facteur B | 1,20 | 1 | 1,2 |  |  | NS |
| Facteur A | 1,67 | 2 | 0,83 |  |  | NS |
| Interaction | 66,60 | 2 | 33,30 | 15,61 | 3,40 | \* |
| résiduelle | 51,20 | 24 | 2,13 |  |  |  |
| Totale | 121 | 29 |  |  |  |  |

F calculé > F théorique

Conclusions : Seul l’effet d’interaction a pu être mis en évidence avec moins de 5% de risque d’erreur. On doit donc uniquement comparer les moyennes entre les 6 traitements.

**Exercice d’application 2 (5 points) (15 min) :**

On veut étudier les effets de 2 « type d’alimentation » (origines de protéines OP1 et OP2) et de 4 « types de ration » (R1 à R4) sur la croissance pondérale d’animaux d’élevage. Chaque traitement a été répété 5 fois. Le tableau de l’ANOVA a donné les indications suivantes :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *variations* | *ddl ?* | *CM* | *F* |  |  |
| OP | 2-1 = 1 | 8,1 | 3,68 | **NS** |  |
| R | 4-1=3 | 3,77 | 1,71 | **NS** |  |
| OP\*R | 1x3= 3 | 24,70 | 11,23 | **\*\*\*** |  |
| résiduelle | 32 | 2,2 |  |  |  |
| Totale | 40-1=39 |  |  |  |  |

On donne les moyennes pour chaque traitement :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Moyenne | R1 | R2 | R3 | R4 |
| OP1 | 9,20 | 12,00 | 12,20 | 9,00 |
| OP2 | 13,80 | 10,00 | 10,80 | 11,40 |

Calculer la valeur de la ppds au 1/1000ème pour un risque égal à 0,05 et réaliser le groupement des moyennes adéquat et concluez cette étude en proposant les conditions d’alimentation qui permettent de maximiser la croissance pondérale des animaux.

L’analyse de la variance a montré que seule l’interaction est significative donc on va regrouper les 8 moyennes des 8 traitements OPxR. Pour chaque combinaison OPxR on a

nij. = 5

L’ANOVA donne CMe = 2,2 avec un ddl = 39 – (1+3+3) = 32

|  |
| --- |
| t0,975(32) = t0,975 =1,960 |
| d(m) = Racine (2,2\*( 1/5+ 1/5) )= 0,938 |
| ***PPDS=*** 0,938 \* 1,96 = 1,839 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Traitement | moyenne | Moy +ppds | Groupement (4 groupes) | | | |
| R4OP1 | 9 | 10,839 | A |  |  |  |
| R1OP1 | 9,2 | 11,039 | A |  |  |  |
| R2OP2 | 10 | 11,839 | A | B |  |  |
| R3OP2 | 10,8 | 12,639 | A | B | C |  |
| R4OP2 | 11,4 | 13,239 |  | B | C |  |
| R2OP1 | 12 | 13,839 |  |  | C | **D** |
| R3OP1 | 12,2 | 14,039 |  |  | C | **D** |
| R1OP2 | 13,8 | 15,639 |  |  |  | **D** |

Pour maximiser la croissance pondérale des animaux il faut privilégier l’OP 2 avec R1 ; ou alors si on utilise OP1 il faut soit R2 soit R3.