ISARA Lyon Durée 1h

UP SFO- S4 – UE : Statistique

Epreuve N°2 (33%) partie 6– Mai 2011

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 7 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 du cours (8 points) (15 min):** Veuillez compléter le tableau ci-dessous à partir des questions de l’annexe .

|  |  |  |  |
| --- | --- | --- | --- |
|  | Proposition retenue 1 ou 2 ou 3…..ou 8 | Degré de certitudes 1 ou 2 ou 3 | Réservée au correcteur |
| QUESTION 1 | 8 |  |  |
| QUESTION 2 | 8 |  |  |
| QUESTION 3 | 2 |  |  |
| QUESTION 4 | 7 |  |  |
| QUESTION 5 | 4 |  |  |
| QUESTION 6 | 2 |  |  |
| QUESTION 7 | 5 |  |  |
| QUESTION 8 | 5 |  |  |
| Total des points | | |  |
| Note sur 8 | | |  |

Pour chaque question vous ne devez donner qu'une seule réponse : 1 ou 2 ou 3 ou...ou 8  
Pour chaque question vous devrez associer un coefficient de certitudes 1 ou 2 ou 3 à votre réponse.

**Barème 4 points par réponse juste, équilibrée avec 3 degrés de certitudes**

Nombre de points attribués en cas **d'absence de réponse et /ou de degré de certitude: 0**Nombre de points attribués en cas de présence d'une réponse:

|  |  |  |
| --- | --- | --- |
| Certitudes | Réponse correcte | Réponse incorrecte |
| **1**: Je ne suis pas sûr-e de ma réponse | +3 | -1 |
| 2: J'ai confiance en ma réponse | +4 | -2 |
| **3**: Je suis certain-e de ma réponse | +5 | -5 |

Explication du barème : Ce barème est calibré pour que vous ayez le nombre de points suffisant pour obtenir la note maximale lorsque vous avez la réponse correcte et que vous êtes confiant-e en votre réponse (certitude 2 qui vous rapporte alors 4 points).

Avec la certitude 3 (Je suis certain-e de ma réponse) vous pouvez donc gagner plus de points qu'il ne vous en faut pour réussir (5 au lieu de 4). Cela vous permet de rattraper des points perdus par ailleurs. Mais attention, si vous n'avez pas la bonne réponse, l'addition est salée !

Avec la certitude 1 (Je ne suis pas sûr-e de ma réponse), en cas de bonne réponse vous ne gagnez pas assez de points pour vous permettre de réussir, mais au pire vous n'en perdez qu'un si vous vous êtes trompé-e.

Au final, la note sur 8 sera calculée selon la formule : [(nombre de points récoltés)/ (8x4)] x8

**Exercice de réflexion (4points) (10 min) :**

*Toutes traces de recherches, même infructueuses, seront prises en compte dans la correction de l’exercice.*

Soit une série de réponse de la forme xij (i = 1 à 6 ; j = 1 à 4) avec par exemple :

x11 = 5.02 ; x12 = 5.37 ; x13 = 5.41… etc

Somme de toutes les valeurs xij = 120.8900

Somme des carrés de toutes les valeurs xij = 610.5341

On dispose de calculs utiles pour réaliser une ANOVA à 2 facteurs (en 4 colonnes et 6 lignes) sans répétitions :

SCE colonnes = 609.1507 - 608.9330 = 0,2177

SCE lignes = 609.7642 - 608.9330 = 0,8312

Que vaudraient ces SCE si on effectuait un changement de repère yij = 100\*(xij – 4) (y11 = 102 ; y12 = 137 ; y13 = 141… etc )? Conséquences sur les F calculés de l’ANOVA?

Somme de toutes les valeurs yij = (120.8900- 24\*4)\*100 = 2489,0000

Somme des carrés de toutes les valeurs yij = 100² ( 610.5341 – 8\*120,89 + 24\*4²) = 274141

SCE tot des xij = 610.5341 – 24\*(120,89/24)² = 1,601096

SCE tot des yij = 274141 – 24\*(2489,00/24)² = 16010,96

On remarque que les SCE sont sensibles au changement d’échelle mais pas au changement d’origine.

Donc dans la nouvelle unité:

SCE colonnes = 100² \*(609.1507 - 608.9330) = 2177

SCE lignes = 100²\*(609.7642 - 608.9330) = 8312

Les F calculés sont des rapports de : SCE/ddl ; donc les rapports vont se simplifier par 100², les F calculés sont donc inchangés.

**Exercice d’application (8points) (30 min) :**

On vous propose d’apporter une réponse à la question posée : « Quelle orientation des ruches et quel type de pollen faut-il mettre à disposition des abeilles pour obtenir une quantité maximale de miel ? ».

Pour cela on dispose de 8 serres et de 24 ruches. On choisit 2 types d’orientation (O 1 et O 2) et 4 types de pollen (P 1 à P 4).Trois ruches sont disposées, selon une orientation définie, dans 1 serre à l’intérieur de laquelle on a disposé les fleurs artificielles en tissu remplies d’un type de pollen. On a utilisé des pelotes d’un pollen que l’on a enrichi avec une substance à tester pour former les 4 types de pollen. A la fin de l’expérimentation on mesure la quantité de miel produite dans chaque ruche.

1. Compléter les propositions suivantes à l’aide de nombres :

Cette étude est une étude à …2…facteurs étudiés avec…3…répétitions. Il y a …8….traitements à tester à répartir sur un total de …24…… unités expérimentales.

1. Les résultats des tests sur les coefficients de Pearson de l’étude sont résumés dans le tableau suivant :

|  |  |  |
| --- | --- | --- |
| coefficient | p-value | Formuler les hypothèses et la conclusion relative au résultat de béta 2 uniquement  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. |
| Beta 1 | 0,806 |
| Beta 2 | 0,45 |

1. Les résultats des tests du X² de l’étude sont résumés dans le tableau suivant :

|  |  |  |
| --- | --- | --- |
|  | X² calculé | Quelles sont les hypothèses et la conclusion par rapport au résultat de O\*P pour = 0,05? Justifier.  H0 : les 8 variances résiduelles entre les traitements sont homogènes  H1 : au moins une des 8 variances résiduelles entre les traitements est supérieure à une autre  ddl = 7 p- value > 90% (table) ou X²0.95(7) = 14.1  conclusion : Le risque d’erreur est trop grand (> 5%) si on décide de rejeter H0, donc on conserve l’hypothèse d’homoscédasticité des variances résiduelles entre les traitements |
| O | 1,917 |
| P | 1,723 |
| O\*P | 2,8 |

1. Etablissez le tableau de l’ANOVA. Hypothèses et conclusions pour alpha = 5%.

Citez les 3 hypothèses qui ont du être validées pour pouvoir réaliser cette ANOVA :

* Normalité des résidus………………………………………………………………………………………………………………
* Indépendance entre résidus et traitements…………………………………………………………………………
* Homoscédasticité des variances résiduelles…………………………………………………………………………………

On donne :

|  |  |  |
| --- | --- | --- |
| somme des répétitions | O 1 | O2 |
| P 1 | 67 | 97 |
| P 2 | 37 | 106 |
| P 3 | 73 | 106 |
| P 4 | 69 | 101 |

Somme des carrés de toutes les valeurs = 19492

Somme de toutes les valeurs = 656

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

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

H0: CMO\*P / CMe = 1 H1: CMO\*P / CMe > 1

SCE tot = 19492 – 24 (656/24)² = 1561,333

SCE (O) = 19051,333 – 24 (656/24)² =1120,667

SCE (P) = 18047,667 – 24 (656/24)² = 117,000

SCE (O\*P) = 19343,333 – 24 (656/24)² - 1120,667 - 117,000 = 175,000

SCE e = 1561,333 - 1120,667 - 117,000 - 175,000 = 148,667

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | S.C.E | ddl | C.M. | F calculé | F 0,95 |
| VarTOTALE | 1561,333 | 23 | 67,884 |  |  |
| Var.FACTEUR O | 1120,667 | 1 | 1120,667 | 120,610 | 4.49 |
| Var.FACTEUR P | 117,000 | 3 | 39,000 | 4,197 | 3.24 |
| Var.INTER O\*P | 175,000 | 3 | 58,333 | 6,278 | 3.24 |
| VAR.RESIDUELLE | 148,667 | 16 | 9,292 |  |  |

Conclusions : Tous les F calculés > F théoriques. Un effet des facteurs étudiés et de l’effet d’interaction ont pu être mis en évidence avec moins de 5% de risque d’erreur. On doit donc uniquement comparer les moyennes entre les 8 traitements.

1. Mettez en œuvre une méthode statistique qui permette de répondre à la question posée au début de l’étude pour un risque d’erreur de 5%. Conclusion.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sigma² d(m) | 9.292\*2/3=6.195 |  |  | t 0,975 (16) | 2,120 |  |  |
| Sigma d(m) | 2.489 |  |  | ppds | 5.276 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| traitement | Moyenne | 3Groupes | | |
| O2 P2 | 35,333 | A |  |  | |
| O2 P3 | 35,333 | A |  |  | |
| O2 P4 | 33,667 | A |  |  | |
| O2 P1 | 32,333 | A |  |  | |
| O1 P3 | 24,333 |  | B |  | |
| O1 P4 | 23,000 |  | B |  | |
| O1 P1 | 22,333 |  | B |  | |
| O1 P2 | 12,333 |  |  | C | |

Il est préférable de choisir l’orientation 2 des ruches, quelque soit le type de pollen, pour obtenir une quantité maximale de miel.

Annexe

**Question 1:**

Les conditions d'application de l'ANOVA sont :

|  |  |
| --- | --- |
| 1 | L’indépendance des observations (doit être contrôlée a priori, au moment de l'échantillonnage aléatoire des observations) |
| 2 | Le(s) facteur(s) étudié(s) doi(ven)t être de nature quantitative |
| 3 | L’homoscédasticité des variances résiduelles |
| 4 | La normalité de la réponse mesurée et des résidus |
| 5 | Toutes les propositions 1, 2, 3, 4 sont correctes |
| 6 | Aucune des propositions 1, 2, 3, 4 n'est correcte |
| 7 | Uniquement 2 propositions sur les 4 sont correctes |
| 8 | Uniquement 3 propositions sur les 4 sont correctes |

**Question 2:**

Dans une analyse de variance à 2 facteurs A (k modalités, i = 1 à k) et B (l modalités, j =1 à l) avec r répétitions (k = 1 à r)

|  |  |
| --- | --- |
| 1 | On préconise : **xijr = µ + i + j + i j + ijr** |
| 2 | Il faut réaliser au total 3 tests de Fisher |
| 3 | Il faut réaliser au total 2 tests d’homoscédasticité : 1 pour le facteur A et 1 pour le facteur B |
| 4 | Chaque résidu eijk est égal à la différence entre la réponse mesurée xijk et la moyenne des répétitions du traitement Ai\*Bj |
| 5 | Toutes les propositions 1, 2, 3, 4 sont correctes |
| 6 | Aucune des propositions 1, 2, 3, 4 n'est correcte |
| 7 | Uniquement 2 propositions sur les 4 sont correctes |
| 8 | Uniquement 3 propositions sur les 4 sont correctes |

**Question 3:**

Dans une analyse de variance à 1 facteur étudié A (k modalités, i = 1 à k) et le même nombre r de répétitions (r = 1 à n)

|  |  |
| --- | --- |
| 1 | On préconise **=** µ + i + jete ir = x ir - |
| 2 |  |
| 3 | S'il y a hétérogénéité des variances résiduelles alors on peut réaliser une comparaison entre la variabilité expliquée par le facteur et la variabilité non expliquée. |
| 4 | CM e = SCE e /( n – k) |
| 5 | Toutes les propositions 1, 2, 3 et 4 sont correctes |
| 6 | Aucune des propositions 1, 2, 3 et 4 n'est correcte |
| 7 | Uniquement 2 propositions sur les 4 sont correctes |
| 8 | Uniquement 3 propositions sur les 4 sont correctes |

**Question 4:**

Dans une analyse de variance à 1 facteur étudié A (k modalités, i = 1 à k) et le même nombre r de répétitions (r = 1 à n), si pour le facteur A on obtient F calculé > F théorique

|  |  |
| --- | --- |
| 1 | On peut affirmer que la variance des moyennes est supérieure à la variance résiduelle |
| 2 | On a réussi à prouver sans se tromper qu’au moins 1 moyenne était différente d’une autre |
| 3 | on a pu mettre en évidence un effet du facteur testé avec un risque de se tromper inférieur à  |
| 4 | On va pouvoir ranger les moyennes par ordre croissant et les comparer 2 à 2 (ppds) |
| 5 | Toutes les propositions 1, 2, 3 et 4 sont correctes |
| 6 | Aucune des propositions 1, 2, 3 et 4 n'est correcte |
| 7 | Uniquement 2 propositions sur les 4 sont correctes |
| 8 | Uniquement 3 propositions sur les 4 sont correctes |

**Question 5:**

Si la probabilité appelée p-value associée au test sur le coefficient 1 est inférieure à 0,05

|  |  |
| --- | --- |
| 1 | On peut admettre que la distribution des résidus est symétrique et 1 = 0 avec moins de 5% de risque d’erreur |
| 2 | On peut conclure que la distribution des résidus n’est pas mésocurtique avec moins de 5% de risque d’erreur |
| 3 | On a réussi à prouver que la distribution des résidus n’est pas symétrique avec moins de 5% de risque d’erreur et que 1 est supérieur à 0 |
| 4 | On a réussi à prouver que la distribution des résidus n’est pas symétrique avec moins de 5% de risque d’erreur et que 1 est différent de 0 |
| 5 | Il manque des données pour pouvoir répondre |
| 6 | Aucune des propositions 1, 2, 3 et 4 n'est correcte |
| 7 | Uniquement 2 propositions sur les 4 sont correctes |
| 8 | Uniquement 3 propositions sur les 4 sont correctes |

**Question 6:**

Dans une analyse de variance à 1 facteur étudié A (k modalités, i = 1 à k) et le même nombre r de répétitions (r = 1 à n), si pour les k variances des résidus du facteur A on obtient X² calculé < X² théorique

|  |  |
| --- | --- |
| 1 | Les k moyennes appartiennent à une même distribution, donc je ne peux pas comparer les variances inter et intragroupes. |
| 2 | Il va être possible de calculer une estimation commune de l’erreur expérimentale. |
| 3 | Les k variances sont jugées hétérogènes, donc je ne peux pas comparer les k moyennes. |
| 4 | Les k moyennes sont jugées hétérogènes, donc je ne peux pas comparer les variances inter et intragroupes.. |
| 5 | Toutes les propositions 1, 2, 3 et 4 sont correctes |
| 6 | Aucune des propositions 1, 2, 3 et 4 n'est correcte |
| 7 | Uniquement 2 propositions sur les 4 sont correctes |
| 8 | Uniquement 3 propositions sur les 4 sont correctes |

**Question 7:**

Dans le cas d’une expérimentation à 3 facteurs A, B et C sans répétition on suppose que:

|  |  |
| --- | --- |
| 1 | Les facteurs A et B et C agissent seulement sur la moyenne |
| 2 | Les facteurs étudiés agissent de façon additive. Les effets de facteurs sont certains. |
| 3 | La loi de la variable X est une loi Normale de variance σ²ε indépendante des variances des facteurs. |
| 4 | l'interaction du 2ème ordre ij t est négligeable |
| 5 | Toutes les propositions 1, 2, 3 et 4 sont correctes |
| 6 | Aucune des propositions 1, 2, 3 et 4 n'est correcte |
| 7 | Uniquement 2 propositions sur les 4 sont correctes |
| 8 | Uniquement 3 propositions sur les 4 sont correctes |

**Question 8:**

On propose 4 modèles théoriques associés à 4 formules de calcul des résidus observés

|  |  |
| --- | --- |
| 1 | xir = µ + i + ir  eir = xir - |
| 2 | xij = µ + i + j + ij  eij = xij + µ -  - |
| 3 | xijr = µ + i + j + i j + ijr  eijr = x ijr - |
| 4 | xijtr = µ + i + j + t + ij + it + jt + ij t + ijtr eijtr = x ijtr - |
| 5 | Toutes les propositions 1, 2, 3 et 4 sont correctes |
| 6 | Aucune des propositions 1, 2, 3 et 4 n'est correcte |
| 7 | Uniquement 2 propositions sur les 4 sont correctes |
| 8 | Uniquement 3 propositions sur les 4 sont correctes |