

UP SFO- S4 - UE : Statistique  
Examen N°3 (34%) partie 7- 15 juin 2011  
Mme Bottollier Lemallaz

---

**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 les résultats finaux avec 3 chiffres significatifs.

Nom : .....	Prénom : .....	Note sur 20 : .....
-------------	----------------	---------------------

**Evaluation du cours (6 points) (15 min):**

Soit un modèle de régression linéaire multiple du 1<sup>er</sup> degré à k facteurs, expliciter la matrice d'information associée à ce modèle : dimension et termes qui la constituent.

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

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

Soit le plan expérimental associé au modèle :  $\hat{y}_l = \beta_0 + \beta_1 U_{l1} + \beta_2 U_{l2} + \beta_3 U_{l3}$

essai	U1 : t (min)	U2 : pH	U3 : T° (°C)
1	10	2	220
2	12	2	218
3	10	4	218
4	12	4	220

Vous allez retranscrire ce plan en une matrice expérimentale en remplaçant les valeurs d' $U_{ij}$  par celles de  $X_{ij}$  tel que chaque  $X_{ij}$  correspond à la variable centrée réduite de chaque  $U_{ij}$

essai	X1	X2	X3
1			
2			
3			
4			

Construisez la matrice de dispersion utilisée pour calculer les coefficients du modèle de la forme  
 $\hat{y}_i = \beta'_0 + \beta'_1 X_{i1} + \beta'_2 X_{i2} + \beta'_3 X_{i3}$

**Exercice d'application (10 points) (30 min) :**

Pour optimiser les quantités de carpes pêchées pendant la période estivale dans 9 étangs de la Dombes, un auteur étudie trois facteurs pouvant influencer les rendements des pêches. Le premier facteur est l'orientation (sinus d'un angle) de l'entrée de la zone de piégeage. Le deuxième facteur est un rapport entre la hauteur d'eau en sortie dans la zone de piégeage et celle de l'entrée. Pour finir, le dernier facteur est la charge initiale de carpes rapportées à l'hectare.

La réponse mesurée  $Y$  représente la quantité de carpes pêchées lors de la première pêche estivale, on admettra qu'elle suit une loi normale. Le plan d'expérience utilisé ainsi que les résultats observés sont donnés dans les tableaux ci-dessous.

N°	U1	U2	U3	Y
Etang	Orientation	Sortie/entrée	Charge	Poids pêché (kg)
1	-0,96	0,4	187,5	528
2	0,65	0,1	180	298
3	-0,87	0,21	250	447
4	0	0,62	242,86	1287
5	-0,13	0,43	130,77	760
6	-0,3	0,48	225	501
7	0,58	0,07	133,33	17
8	-0,87	0,33	193,55	350
9	0,89	0,5	666,67	806
<b>Somme</b>	<b>-1,01</b>	<b>3,14</b>	<b>2209,68</b>	<b>4994,00</b>
<b>SC</b>	<b>4,0933</b>	<b>1,3776</b>	<b>756450,4028</b>	<b>3824792,00</b>

- 1) Les 4 matrices calculées pour déterminer les coefficients du modèle sont dans l'annexe 1 ; A vous de les compléter et de reconnaître ces matrices pour écrire l'équation du modèle retenu :
- 

- 2) Sous forme d'un tableau, valider la pertinence du modèle en complétant et en vous aidant des résultats de l'annexe 2.

Quelles sont vos hypothèses et conclusions ?

					p- value
					0,0375

- 3) Discutez la significativité du coefficient du facteur U2. Justifiez.

- 4) Sachant que l'analyse entière a montré que les facteurs U1 et U3 n'ont pas montré des effets significatifs, quelle réponse apporteriez- vous à cette étude ?

# Annexe 1

0,65020714	0,081195025	-0,889011597	-0,000895316
0,08119502	0,351047959	0,474902839	-0,000845096
-0,88901160	0,474902839	4,824997036	-0,003018416
-0,00089532	-0,000845096	-0,003018416	0,000007550

.....
.....
2219,780
1396406,150

9,0000	-1,0100	3,1400	2209,6800
-1,0100	4,0933	-0,5031	137,2791
3,1400	-0,5031	1,3776	866,8439
2209,6800	137,2791	866,8439	756450,4028

.....
94,052
1804,801
-0,183

# Annexe 2

	Y	Y^	e
	528	.....	.....
	298	.....	.....
	447	232,1335	214,8665
	1287	1055,2308	231,7692
	760	720,5551	39,4449
	501	777,6036	-276,6036 *
	17	137,1365	-120,1365
	350	459,0152	-109,0152
	806	844,9901	-38,9901
<b>somme</b>	.....	.....	.....
<b>SC</b>	<b>3824792,00</b>	<b>3604699,4810</b>	.....

UP SFO- S4 - UE : Statistique  
Epreuve N°2 (33%) partie 6 - 19 Mai 2011  
Mme Bottollier Lemallaz

**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			
QUESTION 2			
QUESTION 3			
QUESTION 4			
QUESTION 5			
QUESTION 6			
QUESTION 7			
QUESTION 8			
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
<b>1:</b> Je ne suis pas sûr-e de ma réponse	+3	-1
<b>2:</b> J'ai confiance en ma réponse	+4	-2
<b>3:</b> 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 :  $[(\text{nombre de points récoltés}) / (8 \times 4)] \times 8$

**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  $x_{ij}$  ( $i = 1 \text{ à } 6$  ;  $j = 1 \text{ à } 4$ ) avec par exemple :

$x_{11} = 5.02$  ;  $x_{12} = 5.37$  ;  $x_{13} = 5.41...$  etc

Somme de toutes les valeurs  $x_{ij} = 120.8900$

Somme des carrés de toutes les valeurs  $x_{ij} = 610.5341$

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  $y_{ij} = 100 \times (x_{ij} - 4)$  ( $y_{11} = 102$  ;  $y_{12} = 137$  ;  $y_{13} = 141...$  etc) ? Conséquences sur les F calculés de l'ANOVA?

**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 .....facteurs étudiés avec.....répétitions. Il y a .....traitements à tester à répartir sur un total de ..... unités expérimentales.

- 2) Les résultats des tests sur les coefficients de Pearson sont résumés dans le tableau suivant :

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

- 3) Les résultats des tests du  $\chi^2$  de l'étude sont résumés dans le tableau suivant :

	$\chi^2$ calculé	Quelles sont les hypothèses et la conclusion par rapport au résultat de O*P pour $\alpha = 0,05$ ? Justifier.
O	1,917	
P	1,723	
O*P	2,8	

- 4) 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 :

- .....  
- .....  
- .....

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

- 5) 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.



## 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) doit(vent) 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 : $x_{ijr} = \mu + \alpha_i + \beta_j + \alpha_i \beta_j + \epsilon_{ijr}$
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 $e_{ijk}$ est égal à la différence entre la réponse mesurée $x_{ijk}$ et la moyenne des répétitions du traitement $A_i \cdot B_j$
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 $\hat{x}_{ir} = \mu + \alpha_i + \beta_j$ et $e_{ir} = x_{ir} - \bar{x}_{ij}$
2	$SCE_A = \sum_{i=1}^k X_{i\cdot} \bar{x}_{i\cdot} - X\mu$
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  $F_{\text{calculé}} > F_{\text{théorique}}$

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 $\alpha$
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  $\beta_1$  est inférieure à 0,05

1	On peut admettre que la distribution des résidus est symétrique et $\beta_1 = 0$ 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 $\beta_1$ 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 $\beta_1$ 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  $X^2_{\text{calculé}} < X^2_{\text{théorique}}$

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 $\sigma^2_\epsilon$ indépendante des variances des facteurs.
4	L'interaction du 2ème ordre $\alpha_i\beta_j\gamma_t$ 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	$x_{ir} = \mu + \alpha_i + \epsilon_{ir}$	$e_{ir} = x_{ir} - \bar{X}_{i.}$
2	$x_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$	$e_{ij} = x_{ij} + \mu - \bar{x}_{i.} - \bar{x}_{.j}$
3	$x_{ijr} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \epsilon_{ijr}$	$e_{ijr} = x_{ijr} - \bar{X}_{ij.}$
4	$x_{ijtr} = \mu + \alpha_i + \beta_j + \gamma_t + \alpha_i\beta_j + \alpha_i\gamma_t + \beta_j\gamma_t + \alpha_i\beta_j\gamma_t + \epsilon_{ijtr}$	$e_{ijtr} = x_{ijtr} - \bar{X}_{ijt.}$
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	