

Exercice 1

On a relevé les durées de gestation de 16 filles de 30 taureaux qui avaient été tirés au sort dans la population devant être étudiée. On voudrait savoir dans quelle mesure la durée de gestation est un caractère héréditaire - caractère se transmettant par les pères. Il s'agit de répondre, grâce à un échantillon comportant peu de taureaux, à une question concernant toute la population. Pour pouvoir étendre les résultats obtenus sur l'échantillon, il faut que celui-ci soit représentatif de toute la population et donc qu'il ait été obtenu par tirage au sort (indépendants et équiprobables). Il en découle que les taureaux de l'échantillon sont aléatoires et leurs effets sur leurs descendants sont a fortiori aléatoires. Ecrire le modèle d'analyse de variance associé à ce problème tout en précisant toutes les hypothèses associées à ce dernier.

Exercice 2

Une compagnie pharmaceutique veut tester les performances d'une méthode de spectroscopie (NIR = Near Infrared Reflectance) permettant de déterminer le contenu en substance active de comprimés. La méthode utilisée en routine (HPLC) est plus coûteuse et laborieuse. Pour cela, 10 comprimés ont été tirés au hasard et les 2 méthodes ont été utilisées sur chacun des comprimés. La question posée est "Existe-t-il une différence de performance entre les 2 méthodes testées HPLC et NIR?". Le fichier de données est le suivant :

Comprimé	HPLC	NIR	Différence HPLC - NIR
1	10.4	10.1	0.3
2	10.6	10.8	-0.2
3	10.2	10.2	0.0
4	10.1	9.9	0.2
5	10.3	11.0	-0.7
6	10.7	10.2	0.2
7	10.3	10.2	0.1
8	10.9	10.9	0.0
9	10.1	10.4	-0.3
10	9.8	9.9	-0.1
$\bar{y}_{HPLC} = 10.34$		$\bar{y}_{NIR} = 10.36$	diff = -0.05
$s_{HPLC} = 0.3239$		$s_{NIR} = 0.4033$	$s_{diff} = 0.2953$

1. Utiliser un test de Khi 2 pour répondre au problème posé.
2. (a) Définir un modèle d'analyse de variance pouvant aider à résoudre le problème posé.
(b) Faire le test ANOVA et répondre aux problèmes posés.
3. Nous pensons que l'idée du modèle mixte qui est de combiner les 2 approches, c'est à dire utiliser un modèle linéaire et y considérer certains facteurs comme aléatoires peut aussi être mise en oeuvre.
(a) Définir un modèle mixte pouvant aider à résoudre le problème posé.
(b) Utiliser le dit modèle pour conclure quand à la question posée.

Exercice 3

Supposons que l'on cherche à comparer 2 traitements A et B ; 4 élevages ont été sélectionnés pour participer à cet essai. Dans chaque élevage un échantillon d'animaux a été tiré au hasard, une moitié des animaux de l'échantillon ont reçu le traitement A et l'autre moitié le traitement B. Les données brutes ont été analysées et les analyses ont montré que le traitement B a une plus grande efficacité que le traitement A. Que peut-on conclure ?

Exercice 4

Parmi trois variétés de blé, quelle est celle qui donne le meilleur rendement ? Pour répondre à cette question, on met en culture ces trois variétés, sur trois parcelles distinctes. Chaque parcelle est divisée en trois sous-parcelles, et ainsi chaque variété est cultivée sur chaque parcelle. Les apports d'engrais azoté, de fongicide et d'insecticide sont les mêmes et sont considérés comme optimaux pour chaque variété. Les mesures ont été reportées dans le tableau ci-dessous. Le sol de chaque parcelle est censé être de même qualité mais on préfère le vérifier en prenant en compte un effet parcelle dans le modèle d'analyse de la variance. Ce type d'effet est appelé « effet bloc ». On ne s'intéresse pas à cet effet, mais on le prend en compte dans l'analyse de la variance pour mieux mettre en évidence les éventuels effets des autres effets.

Parcelle (bloc)	Variété	Rendement
1	A	61,00
2	A	70,92
3	A	55,94
1	L	70,80
2	L	66,44
3	L	68,12
1	M	55,30
2	M	57,30
3	M	54,39

1. Pourquoi la prise en compte de l'effet bloc peut-elle être utile pour conclure à un éventuel effet variété ?
2. Y a-t-il un effet variété au niveau de confiance de 5% ? On supprime l'effet bloc de l'analyse.
3. Pour quelle raison, en pratique, pourrait-on effectuer cette suppression ? Construire le tableau d'analyse de la variance à un facteur (le facteur variété).
4. Construire le test global de significativité de l'effet variété. Comment expliquez-vous que le test devienne significatif ?

Exercice 5

Pendant la cuisson, les croissants absorbent la graisse en quantité variable. Nous avons relevé la quantité de graisse absorbée lors de la cuisson de six fournées de croissantes pour quatre types de graisse. Les mesures sont présentées dans le tableau ci-dessous.

Fournée/Graisse	1	2	3	4
1	64	78	75	55
2	72	91	93	66
3	68	97	78	49
4	77	82	71	64
5	56	85	63	70
6	95	77	76	68

1. Quels sont les modèles que vous pouvez utiliser pour analyser ces données ? Pour chacun d'entre eux, détaillez les hypothèses testées.
2. Nous nous intéressons uniquement à ces quatre types de graisse. Par contre, nous cherchons à savoir si il y a un effet des fournées en général sur la quantité de graisse utilisée. Pour cela, quel est donc le modèle à choisir parmi les trois précédents ? Nous allons maintenant travailler avec le modèle que nous venons de choisir.
3. Vérifiez que les conditions d'application de l'ANOVA II sont vérifiées.
4. Construisez, avec l'aide du logiciel R, le tableau d'analyse de la variance pour ce modèle.
5. La fournée a-t-elle un effet significatif sur la quantité de graisse absorbée ? Écrivez explicitement les hypothèses nulle et alternative de votre test et justifiez votre réponse.
6. La graisse a-t-elle un effet significatif sur la quantité de graisse absorbée ? Écrivez explicitement les hypothèses nulle et alternative de votre test et justifiez votre réponse.

Exercice 6

Davies et Goldsmith ont récolté les données d'une expérience dont le but était d'étudier les différentes sources de variabilité possibles de la résistance d'un ciment fabriqué à Portland. On note Y la variable associée à la résistance du ciment. L'expérience s'est déroulée ainsi : plusieurs petits prélèvements d'un même type de ciment ont été mélangés à de l'eau et travaillés par trois personnes différentes, les « mélangeurs ». On a alors formé douze cubes à l'aide de chacune des préparations des « mélangeurs ». Puis on a donné ces 36 cubes à trois personnes chargées d'évaluer leur résistance, les « casseurs ». La répartition des 36 cubes entre ces « casseurs » a été faite de telle sorte que chaque « casseur » reçoive quatre cubes provenant de chacune des préparations des « mélangeurs » soit douze cubes au total. Tous les tests de résistance ont été faits sur la même machine. L'objectif principal de cette expérience était d'étudier et de quantifier l'importance de la variabilité dans les tests de résistance qui pouvait provenir des différences individuelles entre les « mélangeurs » et les « casseurs ». Les données ci-dessous, exprimées dans les unités d'origine c'est-à-dire en livres par pouces carrés, ont été recopiées dans le tableau ci dessous.

	« Casseur » 1		«Casseur » 2		«Casseur » 3	
« Mélangeur » 1	5280	5520	4340	44400	4160	5180
	4760	5800	5020	6200	5320	4600
« Mélangeur » 2	4420	5280	5340	4880	4180	4800
	5580	4900	4960	6200	4600	4480
« Mélangeur » 3	5360	6160	5720	4760	4460	4930
	5680	5500	5620	5560	4680	5600

Partie I :

Dans l'expérience d'origine, on ne s'intéressait qu'aux différences ne pouvant être dues qu'à ces trois « mélangeurs » et à ces trois « casseurs ».

1. Écrire le modèle d'analyse de la variance relatif à cette étude. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.
2. Les hypothèses du modèle sont-elles vérifiées ? Calculer les estimations de tous les paramètres du modèle.
3. Existe-t-il une interaction dans l'évaluation de la résistance du ciment entre les « mélangeurs » et les « casseurs » ?
3. Existe-t-il des différences dans l'évaluation de la résistance dues aux « mélangeurs » ?
4. Existe-t-il des différences dans l'évaluation de la résistance dues aux « casseurs » ?

5. Compte tenu de la nature des deux facteurs peut-on procéder à des comparaisons multiples ? Il y a-t-il un facteur pour lequel cette procédure se justifie, si oui procéder aux tests correspondants.

Partie II :

On souhaite désormais que les résultats de cette expérience ait une portée plus générale que celle de la première partie : on ne restreint plus l'étude à cette population de « casseurs » et de « mélangeurs ».

6. Écrire le modèle d'analyse de la variance relatif à cette étude. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.
7. Les hypothèses du modèle sont-elles vérifiées ? Calculer les estimations de tous les paramètres du modèle.
8. Existe-t-il une interaction dans l'évaluation de la résistance du ciment entre les « mélangeurs » et les « casseurs » ?
9. Existe-t-il des différences dans l'évaluation de la résistance dues aux « mélangeurs » ?
10. Existe-t-il des différences dans l'évaluation de la résistance dues aux « casseurs » ?

Exercice 7

Une entreprise cherche à tester quatre modèles de machines à écrire. Pour faire ce test, elle demande à cinq secrétaires professionnelles de taper un texte pendant 5 minutes. À la fin du test, on compte le nombre moyen de mots tapés en une minute. On répète l'expérience le lendemain. Les résultats (nombre moyen de mots par minute) sont présentés dans le tableau au verso.

Machines à écrire	Secrétaires				
	1	2	3	4	5
1	33	31	34	34	31
	36	31	36	33	31
2	32	37	40	33	35
	35	35	36	36	36
3	37	35	34	31	37
	39	35	37	35	40
4	29	31	33	31	33
	31	33	34	27	33

Soit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_{i,j} + \epsilon_{i,j,k}$$

avec $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4, 5$ et $k = 1, 2$ et où α correspond à l'influence de la machine à écrire, β l'influence de la secrétaire et γ l'interaction entre la machine i et la secrétaire j .

1. Décrire le modèle, et ses conditions d'utilisation, qui permet de tester les trois hypothèses nulles suivantes :

(a) Que cherche-t-on à tester avec les deux hypothèses suivantes ?

$$\mathcal{H}_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$$

contre

\mathcal{H}_1 : les valeurs des α_i ne sont pas toutes égales, $i = 1, 2, 3, 4$.

(b) Que cherche-t-on à tester avec les deux hypothèses suivantes ?

$$\mathcal{H}_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$$

contre \mathcal{H}_1 : les valeurs des β_j ne sont pas toutes égales, $j = 1, 2, 3, 4, 5$.

(c) Que cherche-t-on à tester avec les deux hypothèses suivantes ?

$$\mathcal{H}_0 : \quad \gamma_{1,1} = \gamma_{1,2} = \cdots = \gamma_{1,5} = \gamma_{2,1} = \cdots = \gamma_{4,5}$$

contre

\mathcal{H}_1 : les valeurs des $\gamma_{i,j}$ ne sont pas toutes égales, $i = 1, 2, 3, 4$ et $j = 1, 2, 3, 4, 5$.

2. Décrire le modèle, et ses conditions d'utilisation, qui permet de tester les trois hypothèses nulles suivantes :

(a) Que cherche-t-on à tester avec les deux hypothèses suivantes ?

$$\mathcal{H}_0 : \quad \sigma_A^2 = 0$$

contre

$$\mathcal{H}_1 : \quad \sigma_A^2 \neq 0.$$

(b) Que cherche-t-on à tester avec les deux hypothèses suivantes ?

$$\mathcal{H}_0 : \quad \sigma_{AB}^2 = 0$$

contre

$$\mathcal{H}_1 : \quad \sigma_{AB}^2 \neq 0.$$

3. Décrire le modèle, et ses conditions d'utilisation, qui permet de tester les trois hypothèses nulles suivantes :

(a) Que cherche-t-on à tester avec les deux hypothèses suivantes ?

$$\mathcal{H}_0 : \quad \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$$

contre

\mathcal{H}_1 : les valeurs des α_i ne sont pas toutes égales, $i = 1, 2, 3, 4$.

(b) Que cherche-t-on à tester avec les deux hypothèses suivantes ?

$$\mathcal{H}_0 : \quad \sigma_B^2 = 0$$

contre

$$\mathcal{H}_1 : \quad \sigma_B^2$$

(c) Que cherche-t-on à tester avec les deux hypothèses suivantes ?

$$\mathcal{H}_0 : \sigma_{AB}^2 = 0$$

contre

$$\mathcal{H}_1 : \sigma_{AB}^2 \neq 0.$$

4. Quelles différences doit-on faire dans l'interprétation des résultats des obtenus aux questions 1., 2. et 3 .

Exercice 8

Lors d'un test de dégustation hédonique, on s'intéresse à l'appréciation globale de trois chocolats. Pour cela, 45 juges ont participé à cette évaluation qui a eu lieu sur 2 jours (on dispose de 15 échantillons par chocolat). Les notes d'appréciation des juges, comprises en 0 et 7 , sont données dans le tableau suivant. Chaque juge n'a évalué qu'un chocolat. Comme chacun choisit son jour de dégustation et le chocolat qu'il évalue, le nombre de données et la répartition des chocolats évalués ne sont pas les mêmes d'un jour

à l'autre. On souhaite d'une part vérifier qu'il y a bien un effet chocolat, s'il y a un effet jour (les chocolats pouvant être plus ou moins appréciés lors du premier ou du deuxième jour) et un effet de l'interaction entre chocolat et jour.

	Chocolat 1			Chocolat 2			Chocolat 3		
Jour 1	5.2	6	5.4	4.2	5.2	5	4.6	5.6	5.2
	5.2	6.6		4.4	5.4	4.8	4.8	5.8	6.4
				4.4	5.6	6	5	6.2	5.2
Jour 2							5.4	6.4	
	3.2	4	3.6	3.2	3.6	4.2	3	3.8	3.4
	4.2	3.8	3.4	4	3.6	4	4.4		
	4	4.4	4.6						
		4							

1. Ecrire le modèle permettant de répondre à la problématique. Que signifie l'interaction entre les facteurs chocolat et jour ?
2. Comment (par quelle formule) sont estimés les paramètres de ce modèle ?
3. Il y a-t-il un effet jour, un effet chocolat et un effet de l'interaction ?
4. Par chocolat, calculer la moyenne des notes et la comparer avec la moyenne ajustée ($\hat{\mu} + \hat{\alpha}_i$). Qu'en pensez-vous ? Quel le chocolat préféré ?
5. Tester l'hypothèse " H_0 : l'effet du chocolat 1 est nul" (préciser les hypothèses du test, la statistique de test, la loi de la statistique de test sous H_0 , et la décision)

Exercice 9

Vous étudiez le rendement du blé en fonction des trois variables suivantes : la présence ou l'absence d'irrigation, la quantité d'engrais apportée (faible, moyenne ou importante) et la présence ou l'absence de fongicide. Pour chaque combinaison des facteurs, deux essais ont été mis en place.

1. Définir le modèle (sous la forme indiquée) que vous utilisez si vous considérez que toutes les interactions d'ordre 2 sont présentes.
Les interactions fongicide \times engrais et fongicide \times irrigation sont considérées a priori comme négligeables par l'expérimentateur.
2. Déterminez les valeurs de tous les coefficients du modèle.
3. Décrire le test de significativité de l'interaction irrigation \times engrais. Conclure.
4. Y a-t-il un effet engrais (Décrire le test et la décision) ? Pour le facteur engrais, la modalité "importante" fait-elle varier le rendement significativement par rapport au rendement moyen (Décrire le test et la décision) ?
5. Quel rendement prédirez-vous dans les conditions suivantes : avec irrigation, avec fongicide et une dose d'engrais importante.