

T. D. n° 1

Exemples classiques de dispositifs expérimentaux

- **Exercice 1, 2, 3** : ANOVA 1F
- **Exercice 4** : ANOVA 1F et transformations stabilisatrice de variance.
- **Exercice 5, 6, 7** : Plan en blocs complets
- **Exercice 8, 9, 10, 11, 12, 13** : Deux facteurs croisés
- **Exercice 9** : Deux facteurs blocs croisés
- **Exercice 14** : Deux facteurs déséquilibrés. Plan non-orthogonal.
- **Exercice 15** : Deux facteurs croisés avec un facteur bloc.
- **Exercice 16, 17** : Trois facteurs croisés
- **Exercice 18** : Deux facteurs emboîtés.
- **Exercice 19, 20, 21** : ANOVA 1F à mesures répétées.
- **Exercice 22, 23, 24, 25** : ANOVA 2F à mesures répétées (I).
- **Exercice 26, 27** : ANOVA 2F à mesures répétées (II).
- **Exercice 28** : Split plot avec répétitions.
- **Exercice 29** : Split-plot complexe.
- **Annexe 30** : Approches non paramétriques et conditionnelles.

Exercice 1. *Mangez des pommes!*

On souhaite comparer la teneur en vitamine C de cinq variétés de pommes notées V_1 , V_2 , V_3 , V_4 et V_5 . Pour chaque variété la teneur en vitamine C, exprimée en $mg/(100g)$ a été mesurée dans cinq pommes prises au hasard. On obtient les données suivantes regroupées dans le tableau ci-dessous :

V_1	V_2	V_3	V_4	V_5
93,6	95,3	94,5	98,8	94,6
95,3	96,9	97,0	98,2	97,8
96,0	95,8	97,8	97,8	98,0
93,7	97,3	97,0	97,2	95,0
96,2	97,7	98,3	97,9	98,9

Qu'est-il possible de conclure? Pour répondre à cette question, on construira le tableau de l'analyse de la variance.

Exercice 2. Un facteur. D'après Prum. *Modèle linéaire. Comparaison de groupes et régression*. Les éditions INSERM, 1996.

Nous souhaitons comparer trois traitements, notés A , B et C contre l'asthme : le traitement B est un nouveau traitement, que nous souhaitons mettre en compétition avec les traitements classiques A et C . Nous répartissons par tirage au sort les patients venant consulter dans un centre de soin, et nous leur affectons l'un des trois

traitements. Nous mesurons sur chaque patient la durée, en jours, séparant de la prochaine crise d'asthme. Les mesures sont reportées dans le tableau ci-dessous :

Traitement A	Traitement B	Traitement C
26; 27; 35; 36	29; 42; 44; 44	26; 26; 30; 30
38; 38; 41; 42	45; 48; 48; 52	33; 36; 38; 38
45; 50; 65	56; 56; 58; 58	39; 46; 47; 51
	60; 61; 63; 63	51; 56; 75
	69	

Pouvons-nous conclure que les traitements ont une efficacité différente pour le critère « temps séparant la prochaine crise ? »

1. Écrire le modèle statistique de l'analyse de la variance à un facteur à effets fixes.
2. Quelles sont les conditions d'utilisation du modèle d'analyse de la variance précédent ? Sont-elles vérifiées ?
3. Donner, à l'aide du logiciel SAS, le tableau de l'ANOVA correspondant à cette étude.
5. Réaliser le test de Fisher au seuil de significativité 5% puis de 1%. Qu'est-il possible d'en déduire ?
6. Donner une estimation de la variance σ^2 .
7. Dans le cas de cette étude, est-il possible de procéder à des comparaisons multiples ? Pourquoi ? Si oui, réaliser alors ces comparaisons.

Exercice 3. Un facteur. D'après Husson et Pagès. *Statistiques générales pour utilisateurs : Tome 2, Exercices et corrigés*. Presses Universitaires de Rennes, 2005.

Le transport d'animaux d'élevage implique une succession de manipulations et de confinements qui, inévitablement, sont responsables de stress. L'objectif de cette étude est d'examiner le comportement des animaux lors de transports de longue durée.

Pour mesurer le niveau de stress, nous avons calculé le pourcentage de temps que les animaux passent couchés. La fatigue des animaux (et donc le stress) est d'autant plus grande que le temps passé couché est important.

L'expérimentation a consisté à observer 18 veaux transportés de France en Italie, la durée du voyage étant de 29 heures. Trois traitements ont été proposés durant la pause :

- Traitement 1 : les veaux ne reçoivent ni eau ni aliment à la pause.
- Traitement 2 : les veaux reçoivent de l'eau et sont alimentés à la pause par deux abreuvoirs.
- Traitement 3 : les veaux reçoivent de l'eau et sont alimentés à la pause par cinq abreuvoirs.

Pour enregistrer les comportements des animaux, la bétailière est équipée de caméras et chaque animal est individualisé par un signe distinctif sur la peau. Les cassettes vidéo ont été dépouillées à l'aide d'un logiciel. Les données sont disponibles dans le tableau ci-dessous.

Traitement 1	17,40	20,00	26,70	31,70	35,80	47,80
Traitement 2	14,65	37,22	37,73	43,61	46,07	47,40
Traitement 3	18,76	19,49	27,19	45,42	53,20	61,27

1. Pourquoi faut-il faire ici une analyse de la variance à un facteur et non pas une analyse de la régression linéaire simple ?
2. Écrire le modèle statistique de l'analyse de la variance à un facteur à effets fixes.
3. Quelles sont les conditions d'utilisation du modèle d'analyse de la variance précédent ? Sont-elles vérifiées ?
4. Donner, à l'aide d'un logiciel, le tableau de l'ANOVA correspondant à cette étude.
5. Réaliser le test de Fisher au seuil de significativité 5% puis de 1%. Qu'est-il possible d'en déduire ?
6. Donner une estimation de la variance σ^2 .
7. Dans le cas de cette étude, est-il possible de procéder à des comparaisons multiples ? Pourquoi ? Si oui, réaliser alors ces comparaisons.

Exercice 4. Transformation. D'après Parreins. *Techniques statistiques, moyens rationnels de choix et de décision*. Dunod technique, 1974.

Nous irradiions des espèces vivantes avec des doses croissantes de rayons X, exprimées en roentgens, et nous calculons le pourcentage de décès. Les résultats expérimentaux ont été reportés dans le tableau suivant :

Dose 0	Dose 500	Dose 1 000	Dose 2 000	Dose 3 000
2,5	5,0	10,0	30,0	46,5
2,7	4,5	6,5	27,0	41,0
2,5	3,5	8,0	26,0	43,0

1. Écrire le modèle statistique de l'analyse de la variance à un facteur à effets fixes.
2. Quelles sont les conditions d'utilisation du modèle d'analyse de la variance précédent ? Sont-elles vérifiées ? En quoi cela est-il limitant ?
3. Pour stabiliser la variance, transformer les données en utilisant la fonction suivante :

$$\phi(x) = \arcsin(\sqrt{x})$$

qui convertit les pourcentages en degrés. Justifier l'utilisation de cette transformation en introduisant $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires telle que, pour tout $n \in \mathbb{N}$, X_n suit une loi binomiale $\mathcal{B}(n, p)$ avec $n \in \mathbb{N}^*$ et $p \in]0; 1[$ et en appliquant la méthode δ .

4. Déterminer alors s'il existe un effet de la dose de rayons X sur la variable calculée en 4.

Exercice 5. Plan en blocs complets. D'après G. Parreins. *Techniques Statistiques : moyens rationnels de choix et de décision*. Dunod technique, 1974.

Nous voulons tester quatre types de carburateurs : A_1 , A_2 , A_3 et A_4 . Pour chaque type de carburateur nous disposons de six pièces qui sont montées successivement en parallèle sur quatre voitures que nous supposons avoir des caractéristiques parfaitement identiques. Le tableau ci-dessous indique pour chacun des essais la valeur d'un paramètre lié à la consommation :

<i>Essai</i>	A_1	A_2	A_3	A_4
1	21	23	18	20
2	24	23	19	21
3	25	32	28	25
4	20	23	19	15
5	34	32	24	29
6	17	15	14	9

Nous décidons **de tenir compte** de la possible influence de l'ordre de réalisation des essais, c'est-à-dire du facteur *Essai*.

1. Proposer une méthode statistique permettant d'étudier conjointement l'influence du facteur *Carburateur* et du facteur *Essai* sur la consommation. Énoncer le modèle et les hypothèses nécessaires au modèle que vous projetez d'utiliser. Ce modèle comporte-t-il des répétitions ?
2. Il y a-t-il des différences entre les carburateurs ? Il y a-t-il des différences dues au facteur bloc essais ?
3. Quelles sont les estimations des paramètres du modèle ? Si nécessaire et si possible, comparer les différents niveaux du facteur *Carburateur* ainsi que les différents niveaux du facteur *Essai*.

Exercice 6. Plan en blocs complets. D'après Husson et Pagès. *Statistiques générales pour utilisateurs : Tome 2, Exercices et corrigés*. Presses Universitaires de Rennes, 2005.

Parmi trois Doses de blé, quelle est celle qui donne le meilleur rendement ? Pour répondre à cette question, on met en culture ces trois Doses, sur trois parcelles distinctes. Chaque parcelle est divisée en trois sous-parcelles, et ainsi chaque Dose 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 Dose. 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)	Dose	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 Dose ?
2. Y a-t-il un effet Dose 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 Dose).
4. Construire le test global de significativité de l'effet Dose. Comment expliquez-vous que le test devienne significatif ?

Exercice 7. Plan en blocs complets. Viennoiseries. Y. Dodge. *Analyse de régression appliquée*. Éditions Dunod, 1999.

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, détailler 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. Vérifier que les conditions d'application de l'ANOVA II sont vérifiées.

4. Construire, avec l'aide de MINITAB, le tableau d'analyse de la variance pour ce modèle.
5. La fournie a-t-elle un effet significatif sur la quantité de graisse absorbée? Écrire 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? Écrire explicitement les hypothèses nulle et alternative de votre test et justifiez votre réponse.
7. Calculer η^2 . Que concluez-vous?
8. Évaluer la puissance dans chacun des tests du tableau de l'ANOVA II.
9. Quelles sont les comparaisons multiples que vous pouvez réaliser? Réaliser-les et analyser-les.

Exercice 8. Deux facteurs croisés. D'après B. Falissard. *Comprendre et utiliser les statistiques dans les sciences de la vie*. Masson, 2005.

Nous testons l'influence de différents régimes alimentaires sur des rats de laboratoire. Le gain de poids des rats est désigné par la variable *Poids*, exprimée en grammes, les deux facteurs sont les variables *Calorie* et *Vitamine*. La variable *Calorie* vaut 1 si les rats n'ont pas suivi un régime hypercalorique et 2 s'ils ont suivi un tel régime hypercalorique. La variable *Vitamine* vaut 1 si les rats n'ont pas reçu de compléments vitaminés et 2 s'ils ont reçu de tels compléments.

Calorie	Vitamine	Poids	Calorie	Vitamine	Poids
1	1	84	1	1	66
1	2	62	1	2	59
2	1	87	2	1	89
2	2	103	2	2	90
1	1	66	1	1	56
1	2	84	1	2	74
2	1	92	2	1	101
2	2	107	2	2	116
1	1	82	1	1	79
1	2	73	1	2	74
2	1	77	2	1	95
2	2	95	2	2	112
1	1	62	1	1	89
1	2	75	1	2	74
2	1	88	2	1	91
2	2	96	2	2	92

1. Quels modèles d'analyse de la variance à deux facteurs pouvez-vous utiliser pour étudier ces données? Nous décidons de retenir, pour répondre aux questions suivantes, le modèle le plus complet parmi ceux dont il est possible de se servir. Rappeler les hypothèses associées au modèle.

2. Procéder à l'étude à l'aide de SAS.
3. Quelles sont les estimations des paramètres du modèle ?
4. Devons-nous réaliser des tests de comparaisons multiples ? Si oui, pour quel(s) facteur(s) ? Le(s) faire.

Exercice 9. Deux facteurs ou blocs croisés. Comparaison de l'évaluation de la résistance d'un même type de ciment. D'après H. Sahai et M. I. Ageel « *The Analysis of Variance* ».

Davies et Goldsmith¹ 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	4400	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 : deux facteurs fixes croisés

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.

1. Davies, O.L. et Goldsmith, P.L. (Eds.), *Statistical Methods in Research and Production*, 4th edition, Oliver and Boyd, Edinburgh, 1972.

3. Existe-t-il une interaction dans l'évaluation de la résistance du ciment entre les « mélangeurs » et les « casseurs » ?
4. Existe-t-il des différences dans l'évaluation de la résistance dues aux « mélangeurs » ?
5. Existe-t-il des différences dans l'évaluation de la résistance dues aux « casseurs » ?
6. 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 : deux facteurs blocs croisés

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 ».

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

Exercice 10. Deux facteurs croisés. Activité d'une enzyme

L'étude de l'activité d'une enzyme a été réalisée chez des sujets jeunes en fonction de l'âge et du sexe.

48 essais ont été réalisés dans des conditions similaires sur 48 individus distincts. Leurs résultats, exprimés en unités arbitraires, ont été reproduits dans le tableau ci-dessous.

Age	Sexe	Enzyme	Age	Sexe	Enzyme
moins de 12 ans	garçon	4,9	plus de 12 ans	garçon	2,1
moins de 12 ans	garçon	2,9	plus de 12 ans	garçon	2,2
moins de 12 ans	garçon	2,7	plus de 12 ans	garçon	1,1
moins de 12 ans	garçon	3,9	plus de 12 ans	garçon	2,9
moins de 12 ans	garçon	4,6	plus de 12 ans	garçon	5,0
moins de 12 ans	garçon	3,3	plus de 12 ans	garçon	3,5
moins de 12 ans	garçon	5,9	plus de 12 ans	garçon	2,4
moins de 12 ans	garçon	4,8	plus de 12 ans	garçon	4,4
moins de 12 ans	garçon	4,1	plus de 12 ans	garçon	2,1
moins de 12 ans	garçon	3,5	plus de 12 ans	garçon	3,0
moins de 12 ans	garçon	7,2	plus de 12 ans	garçon	3,9
moins de 12 ans	garçon	6,1	plus de 12 ans	garçon	5,6
moins de 12 ans	fille	4,5	plus de 12 ans	fille	2,4
moins de 12 ans	fille	6,9	plus de 12 ans	fille	3,6
moins de 12 ans	fille	4,0	plus de 12 ans	fille	4,8
moins de 12 ans	fille	5,4	plus de 12 ans	fille	3,9
moins de 12 ans	fille	1,9	plus de 12 ans	fille	5,5
moins de 12 ans	fille	3,6	plus de 12 ans	fille	5,0
moins de 12 ans	fille	4,8	plus de 12 ans	fille	6,8
moins de 12 ans	fille	3,3	plus de 12 ans	fille	2,2
moins de 12 ans	fille	7,5	plus de 12 ans	fille	3,1
moins de 12 ans	fille	5,8	plus de 12 ans	fille	5,0
moins de 12 ans	fille	4,4	plus de 12 ans	fille	4,1
moins de 12 ans	fille	6,0	plus de 12 ans	fille	4,7

Partie I

- I.1. Écrire le modèle d'analyse de la variance relatif au modèle dont les résultats sont reproduits dans la suite de la **partie I**. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.
- I.2. Les hypothèses du modèle sont-elles vérifiées? Calculer les estimations de tous les paramètres du modèle.
- I.3. Procéder à l'étude du tableau de l'analyse de la variance.
- I.4. Compte tenu de la nature des 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 et interpréter les résultats.

Partie II

- II.1. Écrire le modèle d'analyse de la variance relatif au modèle dont les résultats sont reproduits ci-dessous. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.
- II.2. Les hypothèses du modèle sont-elles vérifiées? Calculer les estimations de tous les paramètres du modèle.
- II.3. Procéder à l'étude du tableau de l'analyse de la variance.

- II.4.** Compte tenu de la nature des 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 et interpréter les résultats.
- II.5.** En comparant les résultats obtenus dans cette seconde partie avec ceux de la première partie, que constatez-vous ?

Exercice 11. Deux facteurs croisés. Quatre modèles de machines à écrire. D'après François Husson et Jérôme Pagès *Statistiques générales pour utilisateurs*, éditions PUR.

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 :

- (i) 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, $i = 1, 2, 3, 4$.

- (ii) 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, $j = 1, 2, 3, 4, 5$.

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

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

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 :

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

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

contre

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

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

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

contre

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

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

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

contre

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

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

(i) 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, $i = 1, 2, 3, 4$.

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

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

contre

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

(iii) 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 12. Deux facteurs croisés. Génotypes et masse. D'après H. Sahai et M. I. Ageel « *The Analysis of Variance* ».

Scheffé (1959, pp. 140–141) a reproduit les données d'une expérience destinée à étudier les variations dans la masse, en g , de rats femelles hybrides en fonction du génotype de la mère nourricière de la portée et du génotype de la portée. Nous reprenons ici une partie de ces données expérimentales. Il s'agit des masses moyennes des individus femelles de 48 portées qui ont été reportées dans le tableau ci-après en fonction du génotype de la mère nourricière de la portée et du génotype de la portée. Précisons que les portées proviennent de mères distinctes et n'ont pas eu les mêmes mères nourricières.

Lignée	Mère	Masse	Lignée	Mère	Masse	Lignée	Mère	Masse
A	A	65,0	F	F	64,0	I	I	55,7
A	A	68,2	F	F	61,7	J	I	45,2
A	A	64,0	I	F	56,3	J	I	57,0
F	A	48,0	I	F	69,8	J	I	61,4
F	A	49,3	I	F	67,0	A	J	42,0
F	A	51,7	J	F	59,5	A	J	48,2
I	A	36,3	J	F	52,8	A	J	54,0
I	A	37,0	J	F	56,0	F	J	51,3
I	A	40,0	A	I	52,5	F	J	40,5
J	A	54,0	A	I	52,7	F	J	45,0
J	A	54,7	A	I	49,5	I	J	50,0
J	A	59,0	F	I	53,0	I	J	43,8
A	F	55,0	F	I	56,5	I	J	54,5
A	F	42,0	F	I	59,0	J	J	44,8
A	F	60,2	I	I	46,0	J	J	51,5
F	F	62,0	I	I	55,3	J	J	53,0

1. Quels modèles d'analyse de la variance à deux facteurs pouvez-vous utiliser pour étudier ces données. On décide de retenir, pour répondre aux questions suivantes, le modèle le plus complet parmi ceux dont il est possible de se servir. Rappeler les hypothèses associées au modèle.
2. Procéder à l'étude à l'aide d'un logiciel.
3. Quels sont les estimations des coefficients du modèle ?
4. Doit-on réaliser des comparaisons multiples ? Si oui pour quel facteur ? Le faire.

Exercice 13. Deux facteurs croisés. Agents toxiques.

Box et Cox ont reproduit les résultats d'une expérience conçue pour évaluer les effets d'agents toxiques dans certaines conditions expérimentales. Des groupes de quatre animaux ont été aléatoirement attribués à ces trois poisons pour chacune des quatre situations, appelées traitements A, B, C et D, étudiées dans l'expérience.

Les temps de survie, observés sur 48 animaux différents et exprimés en dizaines d'heures, ont été reproduits dans le tableau ci-dessous.

	Traitement			
Poison	A	B	C	D
I	0,31	0,82	0,43	0,45
	0,45	1,10	0,45	0,71
	0,46	0,88	0,63	0,70
	0,43	0,72	0,76	0,62
II	0,36	0,92	0,44	0,56
	0,29	0,61	0,35	0,82
	0,40	0,49	0,31	0,71
	0,23	0,84	0,40	0,38
III	0,22	0,30	0,13	0,20
	0,21	0,37	0,25	0,36
	0,08	0,38	0,24	0,31
	0,33	0,29	0,32	0,33

1. Donner le modèle complet de l'analyse de la variance. Donner les effets principaux et interactions qu'il est possible de tester. Les conditions d'utilisation du modèle sont-elles remplies ?
2. Effectuer l'analyse de la variance en indiquant comment sont obtenus les degrés de liberté.
3. Donner, le cas échéant, les effets et les interactions qui ont un effet significatif avec le risque de mauvaise décision associé. Analyser ces effets à l'aide de graphiques et / ou, si cela est possible, des comparaisons multiples.
4. Évaluer, le cas échéant, le risque de mauvaise décision pour les effets principaux et les interactions qui ne sont pas significatifs.
5. Donner un bref commentaire de l'ensemble des résultats.

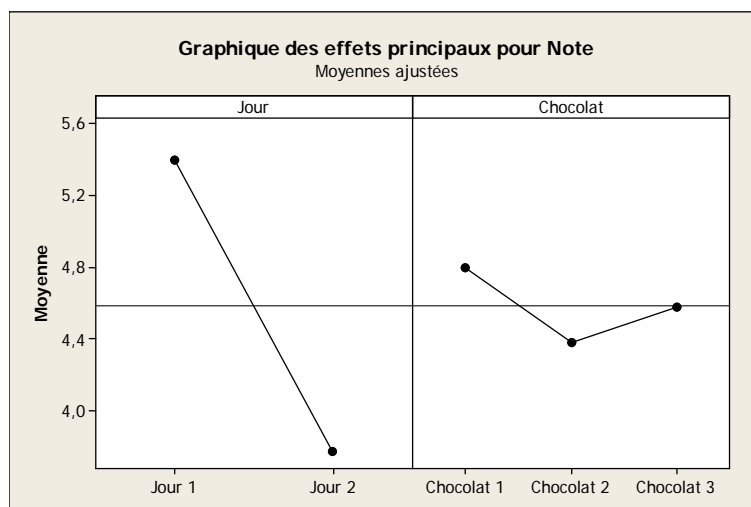
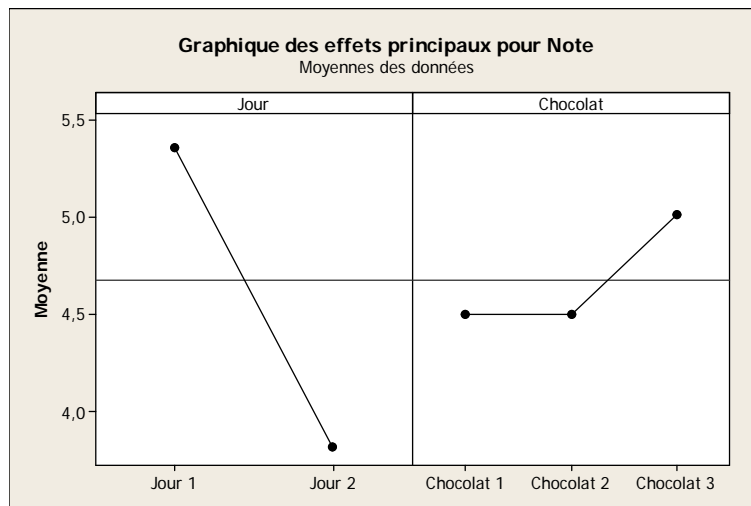
Exercice 14. Deux facteurs croisés déséquilibrés. Analyse sensorielle de trois chocolats. D'après François Husson et Jérôme Pagès *Statistiques générales pour utilisateurs*, éditions PUR.

Lors d'un test de dégustation, 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	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. Le plan utilisé pour cette expérience est-il orthogonal pour le modèle à deux facteurs complet ?
2. Il y a-t-il un effet *jour*, un effet *chocolat* et un effet de l'interaction ?
3. Par chocolat, calculer la moyenne des notes et la comparer avec la moyenne ajustée ($\hat{\mu} + \hat{\alpha}_i$). Qu'en pensez-vous ? Quel est le chocolat préféré ? Expliquez pourquoi il faut être particulièrement vigilant le plan de l'analyse de la variance est déséquilibré. Commentez les graphiques ci-après.



Exercice 15. Deux facteurs croisés avec un facteur bloc. Détermination d'une fumure optimale pour le blé. D'après Pierre Dagnélie *Statistique Théorique et Appliquée*, éditions de Boeck.

Une expérience en blocs aléatoires complets a été réalisée sur du blé au Rwanda. Trois doses d'acide phosphorique (100, 200 et 300 *kg/ha*) et trois doses de chaux (1000, 4500 et 8000 *kg/ha*) ont été testées. Les neuf combinaisons des engrais ont été affectées chacune au hasard et indépendamment à une parcelle au sein de chacun des trois blocs et les rendements obtenus reportés dans le tableau suivant.

Fumures		Bloc		
P ₂ O ₅	CaO	1	2	3
1	1	1,30	0,80	2,01
1	2	2,05	2,37	2,52
1	3	2,07	2,60	2,25
2	1	2,25	2,36	2,71
2	2	2,99	2,92	3,63
2	3	2,62	2,89	3,43
3	1	2,61	2,36	3,29
3	2	3,22	2,93	3,85
3	3	3,15	3,35	3,67

1. Quels modèles d'analyse de la variance peut-on utiliser avec ces données ? Le plan est-il équilibré ? Il y a-t-il des facteurs à effets aléatoires ?
2. On décide dans cette question de négliger l'influence du facteur *Bloc*. Procéder à l'étude de ces données en utilisant le modèle d'analyse de la variance à trois facteurs le plus complet que vous pouvez utiliser.
3. Comment modéliser le facteur *Bloc* ? L'intégrer à l'analyse.

Nous remarquons qu'ici nous constatons que SAS, qui utilise les modèles restreints, réalise ici l'approximation de Satterthwaite alors que cela n'est pas justifié avec un modèle non-restreint.

Exercice 16. Trois facteurs croisés. Résistance de panneaux de particules à l'arrachage des clous. D'après Pierre Dagnélie *Statistique Théorique et Appliquée*, éditions de Boeck.

Une des qualités essentielles des panneaux agglomérés constitués de particules ou de fibres de bois est la résistance de ces panneaux à l'arrachage des clous. Au cours d'une étude nous avons étudié simultanément l'influence de trois facteurs : la grosseur des clous, le diamètre des anneaux sur lesquels sont déposées les éprouvettes soumises aux essais et la vitesse d'arrachage.

Les essais ont été effectués sur des éprouvettes carrées de 50 *mm* de côté, les modalités des trois facteurs sont :

- a) le diamètre (*gros.*) de la tête des clous est soit de 6,5 *mm* soit de 8 *mm*,
- b) le diamètre (*ann.*) des anneaux servant de support est soit de 22 *mm* soit de 30 *mm*,

c) les vitesses (*vit.*) d'arrachage ont été de 22, 45 et 90 par minute.

De plus 5 éprouvettes ont été utilisées pour chacune des 12 combinaisons des modalités du facteur.

Nous disposons ainsi au total de 60 résultats de la mesure de la résistance (*resist.*) de panneaux de particules à l'arrachage des clous exprimée en *kg*.

gros.	ann.	vit.	essai	resist.	gros.	ann.	vit.	essai	resist.
1	1	1	1	54	2	1	1	1	67
1	1	1	2	56	2	1	1	2	71
1	1	1	3	58	2	1	1	3	72
1	1	1	4	51	2	1	1	4	70
1	1	1	5	57	2	1	1	5	81
1	1	2	1	57	2	1	2	1	79
1	1	2	2	58	2	1	2	2	80
1	1	2	3	61	2	1	2	3	81
1	1	2	4	59	2	1	2	4	80
1	1	2	5	55	2	1	2	5	85
1	1	3	1	60	2	1	3	1	78
1	1	3	2	68	2	1	3	2	78
1	1	3	3	61	2	1	3	3	77
1	1	3	4	61	2	1	3	4	83
1	1	3	5	67	2	1	3	5	79
1	2	1	1	54	2	2	1	1	67
1	2	1	2	51	2	2	1	2	66
1	2	1	3	47	2	2	1	3	62
1	2	1	4	51	2	2	1	4	71
1	2	1	5	59	2	2	1	5	69
1	2	2	1	52	2	2	2	1	72
1	2	2	2	56	2	2	2	2	67
1	2	2	3	52	2	2	2	3	75
1	2	2	4	52	2	2	2	4	70
1	2	2	5	53	2	2	2	5	71
1	2	3	1	63	2	2	3	1	78
1	2	3	2	54	2	2	3	2	81
1	2	3	3	65	2	2	3	3	67
1	2	3	4	62	2	2	3	4	76
1	2	3	5	60	2	2	3	5	75

1. Quel modèle d'analyse de la variance pouvons-nous utiliser avec ces données ? Le plan est-il équilibré ? Il y a-t-il des facteurs à effets aléatoires ?
2. Nous décidons dans cette question de négliger l'influence du facteur *essai*. Procéder à l'étude de ces données en utilisant le modèle d'analyse de la variance à trois facteurs le plus complet que nous pouvons utiliser.

Exercice 17. Trois facteurs croisés. Effets secondaires d'un médicament sur les performances de conduite

Un chercheur est chargé d'investiguer les effets sur la performance de conduite de deux nouvelles drogues contre le rhume des foins, A et B. On suspecte qu'au moins l'une de ces drogues peut avoir des effets différents sur des conducteurs alertes ou fatigués, cet effet pouvant aussi dépendre du sexe du conducteur. La firme développant ces drogues a besoin de s'assurer qu'elles n'ont pas d'effets nuisibles sur la performance de conduite.

Toutes les simulations de conduite ont été réalisées sur la même machine. Les performances de conduite de soixante sujets différents, que nous considérerons comme les réalisations d'une variable quantitative continue, ont été reproduites dans le tableau ci-dessous. Des expériences ont été réalisées pour les drogues A et B ainsi que pour un placebo P.

Vigilance	Sexe	Drogue	PerfCond	Vigilance	Sexe	Drogue	PerfCond
A	H	P	24	F	H	P	13
A	H	P	25	F	H	P	12
A	H	P	13	F	H	P	14
A	H	P	22	F	H	P	16
A	H	P	16	F	H	P	17
A	F	P	23	F	F	P	13
A	F	P	18	F	F	P	4
A	F	P	19	F	F	P	3
A	F	P	24	F	F	P	2
A	F	P	26	F	F	P	6
A	H	A	18	F	H	A	21
A	H	A	8	F	H	A	24
A	H	A	9	F	H	A	22
A	H	A	14	F	H	A	23
A	H	A	16	F	H	A	20
A	F	A	15	F	F	A	13
A	F	A	6	F	F	A	11
A	F	A	9	F	F	A	17
A	F	A	8	F	F	A	13
A	F	A	17	F	F	A	16
A	H	B	27	F	H	B	21
A	H	B	14	F	H	B	11
A	H	B	19	F	H	B	14
A	H	B	29	F	H	B	22
A	H	B	27	F	H	B	19
A	F	B	23	F	F	B	9
A	F	B	19	F	F	B	14
A	F	B	17	F	F	B	11
A	F	B	20	F	F	B	21
A	F	B	25	F	F	B	18

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. Procéder à l'étude du tableau de l'analyse de la variance.
4. Compte tenu de la nature des 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 et interpréter les résultats.

Exercice 18. Deux facteurs emboîtés. Cyclamens. D'après Pierre Dagnélie *Statistique Théorique et Appliquée*, éditions de Boeck.

Les résultats suivants sont relatifs à la vitesse moyenne de croissance, en *mm* par jour, des pédoncules floraux observés sur huit plantes de cyclamen, réparties au hasard entre deux milieux de culture.

Milieu 1				Milieu 2			
Plante 1	Plante 2	Plante 3	Plante 4	Plante 5	Plante 6	Plante 7	Plante 8
0,76	0,97	0,58	0,86	0,81	1,07	0,62	0,93
0,66	0,74	0,57	0,93	0,77	0,82	0,99	1,07
0,61		0,51			1,02		0,95
0,65		0,62					0,79
							0,66

1. De quel modèle d'analyse de la variance peut-on se servir pour étudier ces données?
2. Peut-on conclure sur cette base à l'existence d'une différence significative de vitesse de croissance entre les deux milieux?

Exercice 19. Blocs complets. Pression sanguine

Une étude de la relation entre la dose d'un médicament augmentant la pression sanguine et l'augmentation moyenne observée de la pression sanguine diastolique a été menée de la manière suivante : douze lapins ont reçu, dans un ordre aléatoire, les six différentes doses du médicament, l'intervalle entre chacune de ces prises étant suffisamment important pour que le lapin ne soit plus sous l'effet de la dose précédente.

Lapin i	Dose (j)					
	0,1	0,3	0,5	1,0	1,5	3,0
1	21	21	23	35	36	48
2	19	24	27	36	36	46
3	12	25	27	26	33	40
4	9	17	18	27	34	39
5	7	10	19	25	31	38
6	18	26	26	29	39	44
7	9	12	17	22	33	40
8	20	20	30	30	38	41
9	18	18	27	31	42	49
10	8	12	11	24	26	31
11	18	22	25	32	38	38
12	17	23	26	28	34	35

- a) Expliquer pourquoi le plan expérimental qui a été utilisé est en blocs complets.
 b) Que conclure ?

Exercice 20. Blocs complets.

Dans une expérience sur l'esthétique, on demandait à chaque participant de produire trois dessins utilisant juste un de trois matériaux différents pour chacun des dessins : Crayon, Pinceau ou Feutre. La variable dépendante était la note obtenue après l'évaluation du dessin par un panel de juges. Cette variable est considérée comme quantitative continue. La variable indépendante était le type d'instrument utilisé pour produire le dessin. Étant donné que les participants varieraient certainement par leur qualité artistique, il a été décidé de demander à chacun de produire trois dessins, un avec chaque type d'instrument. Pour essayer de neutraliser les effets d'apprentissage, l'ordre des instruments a été randomisé pour chacun des participants.

30 dessins ont été réalisés dans des conditions similaires sur 10 individus distincts. Leurs résultats, exprimés en unités arbitraires, ont été reproduits dans le tableau ci-dessous.

Participant	Ustensile		
	Crayon	Feutre	Pinceau
1	12,71	14,89	12,21
2	18,94	16,35	12,06
3	20,14	16,30	15,99
4	12,00	12,41	10,36
5	19,34	21,91	20,39
6	25,78	20,17	22,83
7	18,98	17,45	16,34
8	22,21	18,31	18,58
9	17,96	12,53	14,04
10	23,81	18,88	20,34

1. Écrire le modèle d'analyse de la variance relatif au modèle dont les résultats sont reproduits dans la suite. 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. Procéder à l'étude du tableau de l'analyse de la variance.
4. Compte tenu de la nature des 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 et interpréter les résultats.

Exercice 21. Blocs complets.

Des chercheurs ont souhaité évaluer l'effet de l'alcool sur les erreurs de conduite. Ils ont fait conduire cinq sujets, à plusieurs reprises et pour trois quantités différentes d'alcool présentes dans le sang (0,04; 0,06 et 0,08 en mg/l). Ces périodes d'essai duraient 15 minutes et se déroulaient dans un simulateur.

Les données obtenues ont été reproduites dans le tableau ci-dessous.

Sujet	Qte_alc_sang	Erreurs
1	0.04	14
2	0.04	16
3	0.04	19
4	0.04	13
5	0.04	11
1	0.06	17
2	0.06	19
3	0.06	21
4	0.06	23
5	0.06	21
1	0.08	25
2	0.08	29
3	0.08	24
4	0.08	21
5	0.08	22

1. Donner le modèle complet de l'analyse de la variance. Donner les effets principaux et interactions qu'il est possible de tester. Les conditions d'utilisation du modèle sont-elles remplies?
2. Effectuer l'analyse de la variance en indiquant comment sont obtenus les degrés de liberté.
3. Donner, le cas échéant, les effets et les interactions qui ont un effet significatif avec le risque de mauvaise décision associé. Analyser ces effets à l'aide de graphiques et / ou, si cela est possible, des comparaisons multiples.

4. Évaluer, le cas échéant, le risque de mauvaise décision pour les effets principaux et les interactions qui ne sont pas significatifs.
5. Donner un bref commentaire de l'ensemble des résultats.

Exercice 22. Split plot. Organisation des rayons

Une étude expérimentale a été réalisée afin d'étudier un éventuel effet d'une différence en matière de présentation sur un rayonnage d'un produit d'entretien ménager. Huit magasins ont été choisis au hasard et répartis, également au hasard, en deux groupes de quatre. Les ventes du produit d'entretien ménager ont été relevées simultanément dans chacun de ces huit magasins à quatre reprises.

Type de présentation	Magasin	Relevé			
		$k = 1$	$k = 2$	$k = 3$	$k = 4$
$j = 1$	$i = 1$	956	953	938	1049
	$i = 2$	1008	1032	1025	1123
	$i = 3$	350	352	338	438
	$i = 4$	412	449	385	532
$j = 2$	$i = 1$	769	766	739	859
	$i = 2$	880	875	860	915
	$i = 3$	176	185	168	280
	$i = 4$	209	223	217	301

- a) Calculer les espérances des carrés moyens pour le modèles d'ANOVA à deux facteurs à mesures répétées suivant l'un des deux facteurs et le modèle split plot.
- b) Trouver l'expression des statistiques de test fournies dans le cours.
- c) Exprimer la puissance de chacun des tests obtenus à la question précédente.
- d) Expliquer pourquoi le plan expérimental qui a été utilisé est un plan split plot (en parcelles divisées).
- e) Que conclure ?

Exercice 23. Split plot. Incentive stimulus

On cherche l'effet de deux incentive stimulus sur la capacité d'une personne à résoudre deux problèmes de nature différente. Douze personnes ont été choisies au hasard et affectées en nombre égal à deux groupes. Chacun de ces groupes est associé à l'un des deux incentive stimulus. L'ordre dans lequel nous avons fait résoudre les problèmes aux sujets a été choisi au hasard pour chacun des sujets. Les scores reportés dans le tableau ci-dessous caractérisent la capacité du sujet à résoudre le problème auquel il était confronté ; plus le score est haut, plus le sujet a montré une capacité de résolution du problème élevée.

Incentive stimulus	Sujet	Type de problème	
		Abstrait $k = 1$	Concret $k = 2$
$j = 1$	$i = 1$	10	18
	$i = 2$	14	19
	$i = 3$	17	18
	$i = 4$	8	12
	$i = 5$	12	14
	$i = 6$	15	20
$j = 2$	$i = 1$	16	25
	$i = 2$	19	22
	$i = 3$	22	27
	$i = 4$	20	23
	$i = 5$	24	29
	$i = 6$	21	22

- a) Expliquer pourquoi le plan expérimental qui a été utilisé est un plan split plot (en parcelles divisées).
- b) Que conclure ?

Exercice 24. Split plot. D'après « Effect of musical expertise on visuospatial abilities : Evidence from reaction times and mental imagery », Renaud Brochard, André Dufour and Olivier Després, *Brain and Cognition*, Volume 54 (2004) pp. 103-109.

Des chercheurs se sont intéressés à l'effet de l'apprentissage de la musique sur les capacités visuo-spatiales des sujets. Dans l'une des expériences réalisées, les chercheurs utilisent 22 sujets (11 musiciens et 11 non-musiciens). Dans une première condition, ils mesurent le temps de réaction des sujets soumis à un stimulus simple : un petit disque lumineux est affichée pendant 70 ms et les sujets doivent appuyer le plus rapidement possible sur un bouton lorsqu'ils aperçoivent le disque. Dans une deuxième condition, les sujets doivent effectuer un choix : le cercle lumineux est soit vert, soit rouge et les sujets doivent appuyer sur la flèche gauche ou la flèche droite d'un clavier selon la couleur du stimulus.

Les données (temps de réaction moyen sur 80 essais dans chaque condition) observées lors d'une reprise de cette expérience figurent dans le tableau ci-dessous.

Sujets	Expertise	Simple	Choix	Sujets	Expertise	Simple	Choix
s_1	Mus.	160	303	s_{12}	Non-Mus.	158	323
s_2	Mus.	208	230	s_{13}	Non-Mus.	202	345
s_3	Mus.	181	272	s_{14}	Non-Mus.	181	341
s_4	Mus.	133	236	s_{15}	Non-Mus.	164	340
s_5	Mus.	190	283	s_{16}	Non-Mus.	161	244
s_6	Mus.	215	282	s_{17}	Non-Mus.	204	333
s_7	Mus.	183	261	s_{18}	Non-Mus.	177	333
s_8	Mus.	205	291	s_{19}	Non-Mus.	194	287
s_9	Mus.	126	322	s_{20}	Non-Mus.	195	324
s_{10}	Mus.	190	268	s_{21}	Non-Mus.	201	311
s_{11}	Mus.	157	284	s_{22}	Non-Mus.	184	304

1. Indiquer pourquoi il faut considérer un troisième facteur « Sujets », de nature aléatoire, hiérarchisé dans le facteur « Expertise ».
2. Donner le modèle complet de l'analyse de la variance. Donner les effets principaux et interactions qu'il est possible de tester.
3. Effectuer l'analyse de la variance en indiquant comment sont obtenus les degrés de liberté.
4. Donner, le cas échéant, les effets et les interactions qui ont un effet significatif avec le risque de mauvaise décision associé. Analyser ces effets à l'aide de graphiques et / ou, si cela est possible, des comparaisons multiples.
5. Évaluer, le cas échéant, le risque de mauvaise décision pour les effets principaux et les interactions qui ne sont pas significatifs.
6. Donner un bref commentaire de l'ensemble des résultats.

Exercice 25. Split plot. Exercice et régime alimentaire.

Vous avez mesuré la pulsation cardiaque de dix-huit sujets dans trois conditions expérimentales différentes. Les premières mesures (EXERTYPE=1) ont été prises lors de l'échauffement des sujets avant l'effort. Les secondes mesures (EXERTYPE=2) ont été prises après 10 minutes de course. Les dernières mesures (EXERTYPE=3) ont été prises lors de la phase de récupération après l'effort. La variable régime alimentaire (DIET) indique les préférences alimentaires des sujets : (DIET=1) pour les carnivores et (DIET=2) pour les végétariens.

Les résultats sont présentés dans le tableau ci-après.

1. Écrire le modèle d'analyse de la variance relatif au modèle dont les résultats sont reproduits dans la suite. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.
2. Indiquer pourquoi il faut considérer dans le modèle un troisième facteur « Indiv », de nature aléatoire.
3. Les hypothèses du modèle sont-elles vérifiées? Calculer les estimations de tous les paramètres du modèle.

4. Procéder à l'étude du tableau de l'analyse de la variance.
5. Compte tenu de la nature des 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 et interpréter les résultats.

EXERTYPE	DIET	INDIV	PULSE	EXERTYPE	DIET	INDIV	PULSE	EXERTYPE	DIET	INDIV	PULSE	EXERTYPE	DIET	INDIV	PULSE
1	1	1	112	2	1	1	166	3	1	1	215				
1	1	2	111	2	1	2	166	3	1	2	225				
1	1	3	89	2	1	3	132	3	1	3	189				
1	2	4	95	2	2	4	134	3	2	4	186				
1	2	5	66	2	2	5	109	3	2	5	150				
1	2	6	69	2	2	6	119	3	2	6	177				
1	1	7	125	2	1	7	177	3	1	7	241				
1	1	8	85	2	1	8	117	3	1	8	186				
1	1	9	97	2	1	9	137	3	1	9	185				
1	2	10	93	2	2	10	151	3	2	10	217				
1	2	11	77	2	2	11	122	3	2	11	178				
1	2	12	78	2	2	12	119	3	2	12	173				
1	1	13	81	2	1	13	134	3	1	13	205				
1	1	14	88	2	1	14	133	3	1	14	180				
1	1	15	88	2	1	15	157	3	1	15	224				
1	2	16	58	2	2	16	99	3	2	16	131				
1	2	17	85	2	2	17	132	3	2	17	186				
1	2	18	78	2	2	18	110	3	2	18	164				

Exercice 26. Blocs complets croisés. Efficacité d'une calculatrice

Afin de tester l'efficacité de son nouveau modèle de calculatrice, une entreprise a choisi au hasard six ingénieurs familiers de l'utilisation à la fois de ce nouveau modèle et d'un modèle plus ancien. On leur a demandé à chacun de résoudre deux problèmes, l'un de nature statistique et l'autre d'ingénierie, à l'aide des deux calculatrices. Le temps de résolution, exprimé en minutes, a été reporté dans le tableau ci-dessous.

Sujet i	Problème statistique $j = 1$		Problème ingénierie $j = 2$	
	Nouveau modèle $k = 1$	Ancien modèle $k = 2$	Nouveau modèle $k = 1$	Ancien modèle $k = 2$
	1 Jones	3,1	7,5	2,5
2 Williams	3,8	8,1	2,8	5,3
3 Adams	3,0	7,6	2,0	4,9
4 Dixon	3,4	7,8	2,7	5,5
5 Erickson	3,3	6,9	2,5	5,4
6 Maynes	3,6	7,8	2,4	4,8

- Calculer les espérances des carrés moyens pour le modèles d'ANOVA à deux facteurs à mesures répétées suivant les des facteurs et le modèle en deux blocs complets.
- Trouver l'expression des statistiques de test fournies dans le cours.
- Exprimer la puissance de chacun des tests obtenus à la question précédente.
- Essayer d'analyser ces données avec un modèle pour un plan expérimental en blocs complets.
- Essayer d'analyser ces données avec un modèle pour mesures répétées.
- Que conclure ?

Exercice 27. Blocs complets croisés. Reconnaissance de symboles

Une expérience a été menée pour rechercher les effets de deux facteurs (variables indépendantes) sur la reconnaissance de symboles présentés brièvement sur un écran, mesurée par le nombre d'identifications correctes sur un nombre fixe d'essais. Les facteurs sont Symbole (avec les niveaux Chiffres, Minuscule, Majuscule) et la police de caractères (avec les niveaux Gothique et Roman). Chacun des participants de l'expérience est testé sous toutes les six combinaisons de deux facteurs traitements.

Les résultats sont présentés dans le tableau ci-dessous.

	Chiffres		Minuscule		Majuscule	
Sujet	Gothique	Roman	Gothique	Roman	Gothique	Roman
1	2	6	18	3	20	5
2	4	9	20	6	18	2
3	3	10	15	2	21	3
4	1	12	10	9	25	10
5	5	8	13	8	20	8
6	6	10	14	10	16	6

1. Écrire le modèle d'analyse de la variance relatif au modèle dont les résultats sont reproduits dans la suite. On précisera la nature des facteurs explicatifs ainsi que les hypothèses faites.
2. Indiquer pourquoi il faut considérer dans le modèle un troisième facteur « Sujets », de nature aléatoire.
3. Les hypothèses du modèle sont-elles vérifiées ? Calculer les estimations de tous les paramètres du modèle.
4. Procéder à l'étude du tableau de l'analyse de la variance.
5. Compte tenu de la nature des 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 et interpréter les résultats.

Exercice 28. Split plot avec répétitions. Coquillages.

La réponse expérimentale Y est la densité de peuplement, exprimée en $\sqrt{cm^{-2}}$, de larves *Semibalanus balanoides*, un petit crustacé qui se trouve en abondance sur les plages rocheuses européennes.

Ces crustacés ne peuvent se déplacer qu'au stade larvaire.

La question est de savoir si le nombre de crustacés présents sur un rocher influe sur le nombre de larves venant s'y accrocher.

Il est également vraisemblable que le taux de peuplement des rochers voisins puisse influencer ce nombre.

Plusieurs rochers, ont été nettoyés des crustacés les habitant à l'exception d'un emplacement central où deux, huit ou 32 adultes n'ont pas été touchés. Ce facteur est le facteur **Traitement**.

Pour déterminer la possible influence de la densité de peuplement des rochers voisins, qui est une caractéristique globale de la plage donc commune à tous les rochers de celle-ci, ces emplacements ont été préparés dans des conditions similaires sur différentes plages pour lesquels cette densité est élevée ou faible (Facteur **Recrute-ment**). Pour chacune de ces densités deux plages (Facteur **Plage**), ont été utilisées.

Recrutement élevé						
T	Cowes			Seaview		
2	0,386	0,397	0,432	0,279	0,411	0,260
8	0,484	0,482	0,514	0,625	0,531	0,478
32	0,484	0,520	0,569	0,738	0,570	0,620

Recrutement faible						
T	Totland			Ventnor		
2	0,190	0,177	0,300	0,304	0,302	0,278
8	0,268	0,261	0,396	0,402	0,351	0,254
32	0,384	0,319	0,334	0,244	0,401	0,324

- Expliquer pourquoi le plan expérimental qui a été utilisé est un plan split plot (en parcelles divisées) avec répétitions.
- Que conclure ?

Exercice 29. Un plan split-plot complexe. Dynamique de communautés.

Nous cherchons à étudier l'influence de la floraison d'arbres sur la dynamique des communautés d'oiseaux.

Nous disposons de huit sites répartis suivant un plan factoriel complet à deux facteurs, A et C, à deux modalités et répété deux fois.

- Le facteur A indique si le site est majoritairement peuplé d'arbres d'espèce A1 ou d'arbres d'espèce A2.
- Le facteur B désigne le site.
- Le facteur C indique si le site est situé au Nord ou au Sud d'une limite géographique naturelle.
- Le peuplement des oiseaux a été évalué une seule fois sur chaque site tous les mois pendant un an, ce qui constitue le facteur D.

1. Préciser quels sont les facteurs croisés et quels sont ceux emboîtés ?
2. Déterminer le modèle restreint complet associé à ce dispositif expérimental. Préciser les conditions d'utilisation du modèle.
3. Trouver l'expression des carrés moyens.
4. Calculer les espérances de ces carrés moyens.
5. En déduire des estimateurs sans biais des composantes de variances.
6. À l'aide des espérances des carrés moyens calculées précédemment, pour quels termes du modèle est-il possible de proposer une statistique de test ?
7. Enfin, déterminer la puissance de chacun des tests introduits précédemment.
8. Nous supposons désormais que l'expérience a duré plusieurs années et que l'effet année est négligeable, c'est-à-dire que les observations obtenues chaque an peuvent être considérées comme des répétitions. Reprendre l'étude précédente.

Exercice 30. Approches non paramétriques et conditionnelles. Rayons X strikes back.

Dans la suite il faut remplacer le chemin de travail défini par `setwd` par le chemin où est stocké le fichier de données de l'exercice.

Retour sur l'exercice 1 du TD1

Récupération des données

```
wd="/Users/fbertran/Documents/Enseignements/public_html/enseignement/ENSAI_2015"
setwd(wd)
exo1TD1 <- read.csv(file="ENSAI_TD1_EX_1.CSV", sep=";", dec=",",
                    encoding="latin1")
```

R mémorise les caractéristiques du tableau de données

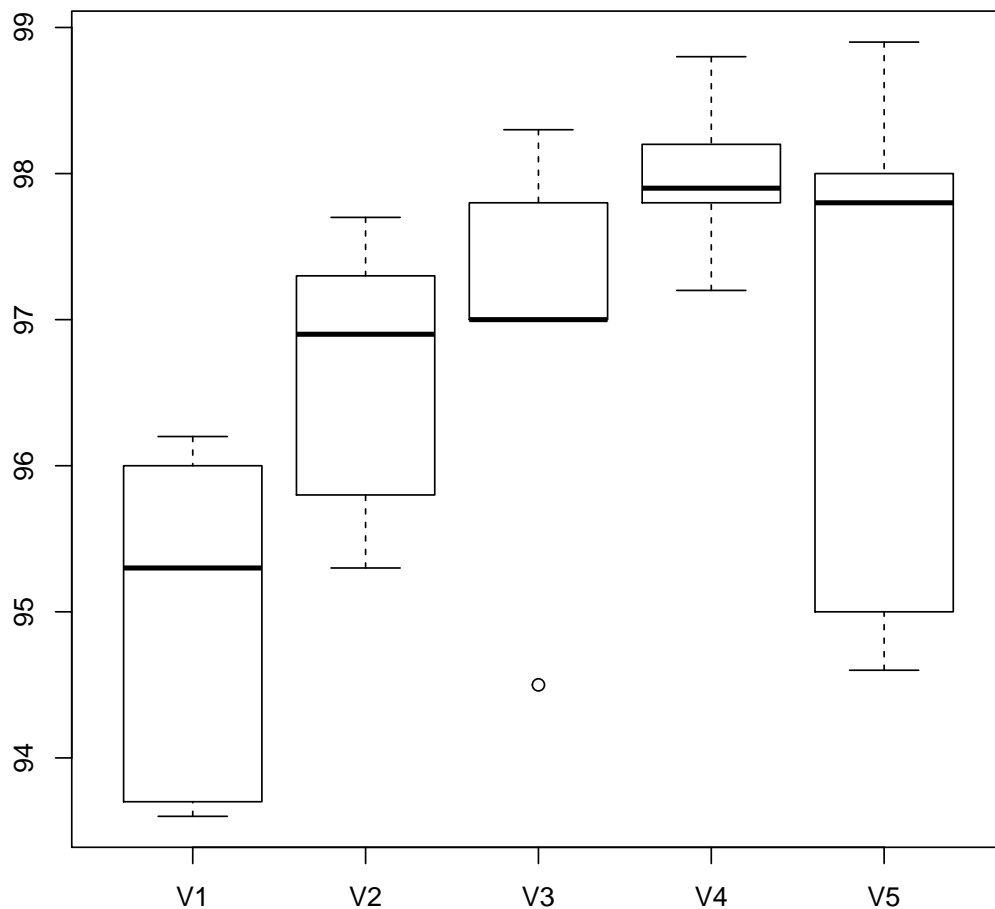
```
attach(exo1TD1)
```

Nous vérifions que les données sont bien celles de l'énoncé

```
str(exo1TD1)
## 'data.frame': 25 obs. of 2 variables:
## $ Variété : Factor w/ 5 levels "V1","V2","V3",...: 1 1 1 1 1 2 2 2 2 2
## $ Teneur.Vitamine.C: num 93.6 95.3 96 93.7 96.2 95.3 96.9 95.8 97.3 97.7 ...
head(exo1TD1)
## Variété Teneur.Vitamine.C
## 1 V1 93.6
## 2 V1 95.3
## 3 V1 96.0
## 4 V1 93.7
## 5 V1 96.2
## 6 V2 95.3
tail(exo1TD1)
## Variété Teneur.Vitamine.C
## 20 V4 97.9
## 21 V5 94.6
## 22 V5 97.8
## 23 V5 98.0
## 24 V5 95.0
## 25 V5 98.9
```

Avant toutes choses nous représentons les données

```
plot(Variété,Teneur.Vitamine.C)
```



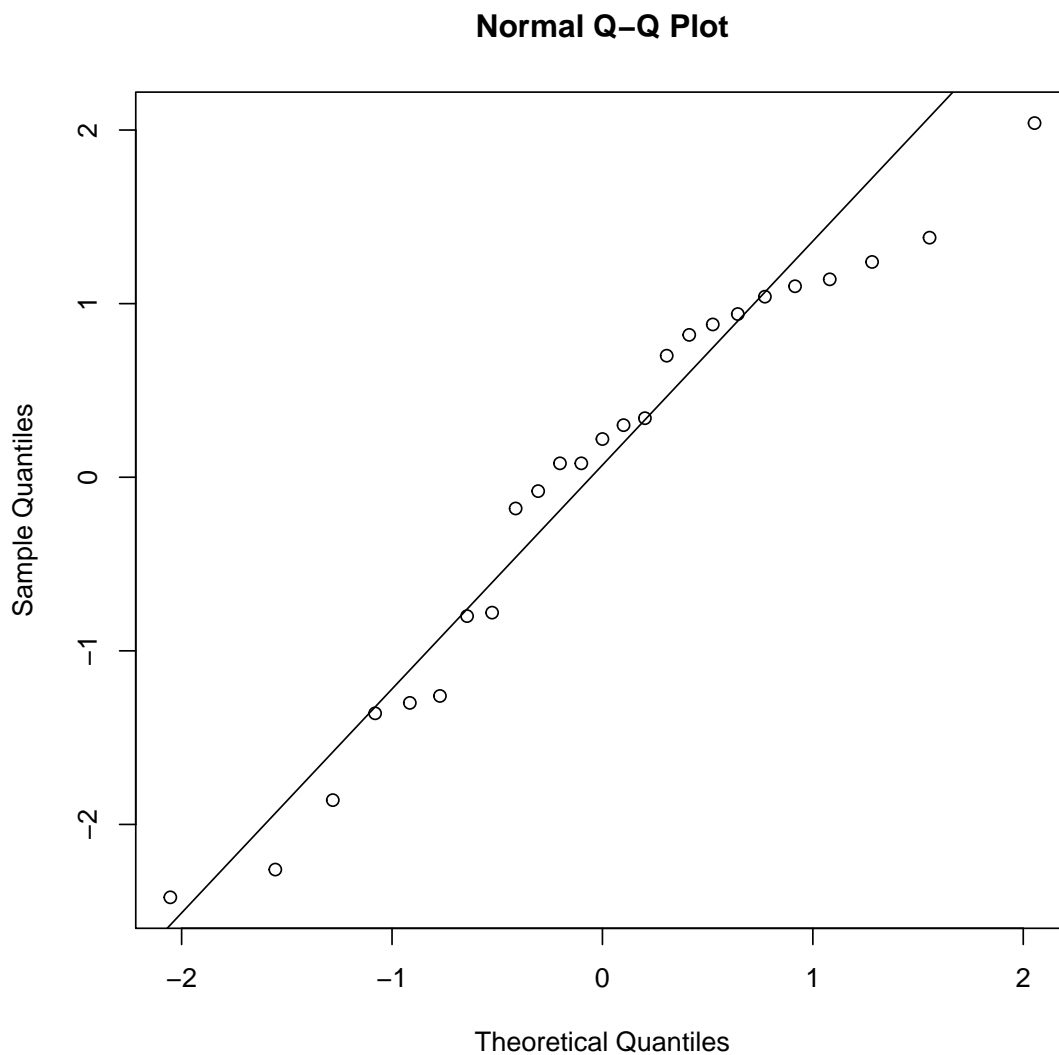
Nous procédons au calcul de l'analyse de la variance de la variable Teneur.Vitamine.C par la variable Variété la variable Bière

```
Model1 <- aov(Teneur.Vitamine.C~Variété)
```

Pour utiliser le test du tableau de l'analyse de la variance nous devons vérifier les conditions d'utilisation du modèle.

Normalité

```
qqnorm(residuals(Model1))  
qqline(residuals(Model1))
```



```
shapiro.test(residuals(Model1))  
##  
## Shapiro-Wilk normality test  
##  
## data: residuals(Model1)  
## W = 0.94597, p-value = 0.203
```

Homogénéité

```
bartlett.test(residuals(Model1)~Variété)  
##  
## Bartlett test of homogeneity of variances  
##  
## data: residuals(Model1) by Variété  
## Bartlett's K-squared = 4.9075, df = 4, p-value = 0.2969
```

La lecture de la p -valeur du test du tableau de l'analyse de la variance est légitime. Conclusion.

```
summary(Model1)
##              Df Sum Sq Mean Sq F value Pr(>F)
## Variété      4  23.72   5.929   3.377 0.0289 *
## Residuals   20  35.12   1.756
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

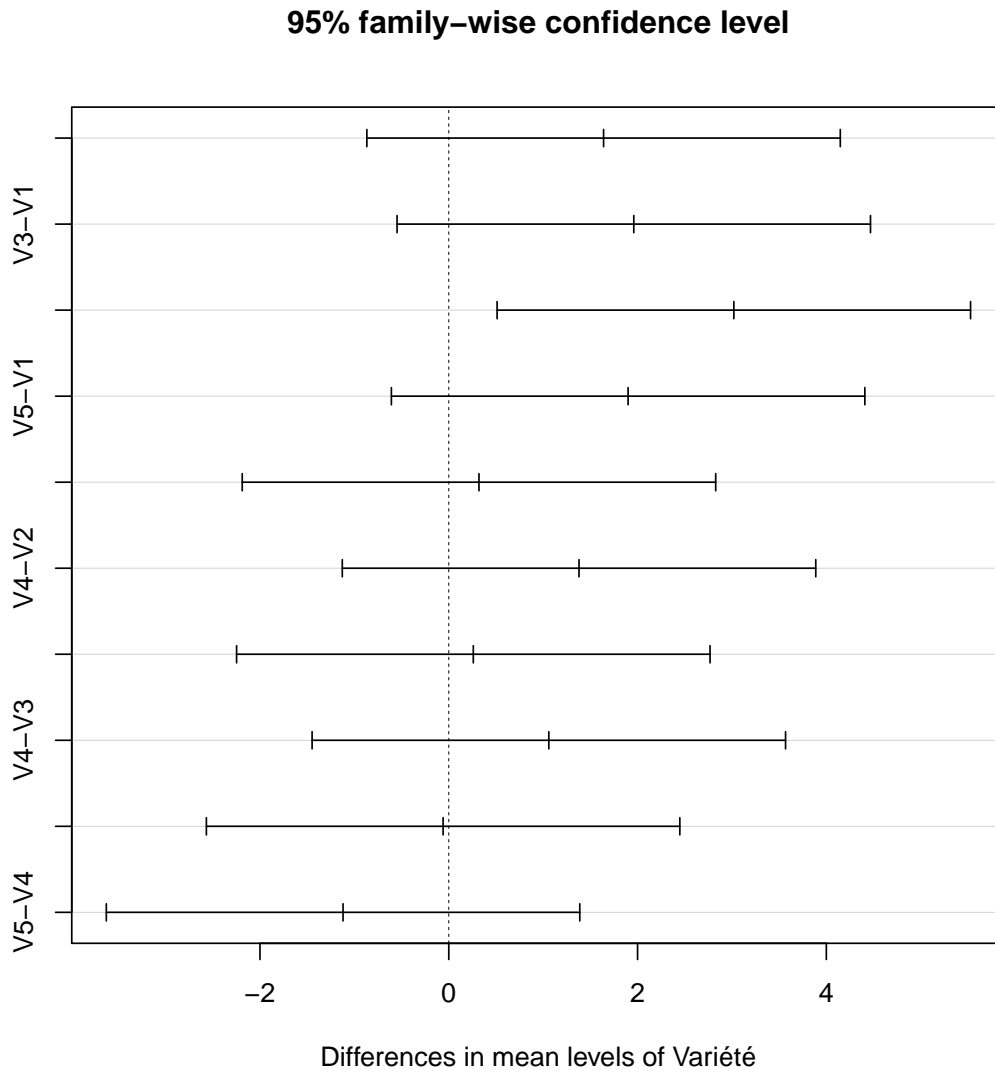
Nous aurions aussi pu utiliser

```
oneway.test(Teneur.Vitamine.C~Variété, var.equal=TRUE)
##
## One-way analysis of means
##
## data:  Teneur.Vitamine.C and Variété
## F = 3.3767, num df = 4, denom df = 20, p-value = 0.02888
```

Voici les estimations des effets fixes du facteur Variété

Nous pouvons réaliser une comparaison individuelle à l'aide de la procédure de Tukey puisque l'hypothèse nulle a été rejetée. «echo=FALSE» model.tables(Model1) @

```
TukeyHSD(Model1, "Variété", ordered = FALSE)
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Teneur.Vitamine.C ~ Variété)
##
## $Variété
##      diff      lwr      upr      p adj
## V2-V1  1.64 -0.8678884  4.147888 0.3216050
## V3-V1  1.96 -0.5478884  4.467888 0.1739336
## V4-V1  3.02  0.5121116  5.527888 0.0136044
## V5-V1  1.90 -0.6078884  4.407888 0.1966726
## V3-V2  0.32 -2.1878884  2.827888 0.9951144
## V4-V2  1.38 -1.1278884  3.887888 0.4868631
## V5-V2  0.26 -2.2478884  2.767888 0.9978120
## V4-V3  1.06 -1.4478884  3.567888 0.7147077
## V5-V3 -0.06 -2.5678884  2.447888 0.9999935
## V5-V4 -1.12 -3.6278884  1.387888 0.6727621
plot(TukeyHSD(Model1, "Variété"))
```

Inférence conditionnelle
Tests de permutations (lmPerm)

```
library(pgirmess)
Res_perm <- PermTest(Model1, B=1000)
print(Res_perm)
##
## Monte-Carlo test
##
## Call:
## PermTest.lm(obj = Model1, B = 1000)
##
## Based on 1000 replicates
## Simulated p-value:
##      p.value
## Variété    0.03
```

```
#Tests de permutations II (lmPerm)
library("lmPerm")
```

```
Model1.perm <-aovp(Teneur.Vitamine.C~Variété)
## [1] "Settings: unique SS "
summary(Model1.perm)
## Component 1 :
##           Df R Sum Sq R Mean Sq Iter Pr(Prob)
## Variété     4  23.718   5.9294 4504 0.02176 *
## Residuals  20  35.120   1.7560
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Permutational Anova III, KW et post-hocs

```
library(coin)
## Loading required package: survival
library(multcomp)
## Loading required package: mvtnorm
## Loading required package: TH.data
library(car)
Anova_oneway <- oneway_test(Teneur.Vitamine.C ~ Variété, data = exo1TD1,
                           distribution = "asymptotic")
Anova_oneway_MC <- oneway_test(Teneur.Vitamine.C ~ Variété,
                              data = exo1TD1, distribution =
                              approximate(B = 90000))
```

Anova_oneway_MC

```
print(Anova_oneway)
##
## Asymptotic K-Sample Fisher-Pitman Permutation Test
##
## data: Teneur.Vitamine.C by Variété (V1, V2, V3, V4, V5)
## chi-squared = 9.6745, df = 4, p-value = 0.04628
print(Anova_oneway_MC)
##
## Approximative K-Sample Fisher-Pitman Permutation Test
##
## data: Teneur.Vitamine.C by Variété (V1, V2, V3, V4, V5)
## chi-squared = 9.6745, p-value = 0.03124
```

IC p-value perm Anova_oneway

```
print(pvalue(Anova_oneway))
## [1] 0.04628319
```

```

print(pvalue(Anova_oneway_MC))
## [1] 0.03124444
## 99 percent confidence interval:
## 0.02976992 0.03276877
if (pvalue(Anova_oneway_MC) <= 0.05) {
Tukey_posthoc <- oneway_test(Teneur.Vitamine.C ~ Variété, data = exo1TD1,
                             ytrafo = function(data) trafo(data, numeric_trafo = rank),
                             xtrafo = function(data) trafo(data, factor_trafo = function(x)
                             model.matrix(~x - 1) %*% t(contrMat(table(x), "Tukey"))),
                             teststat = "max", distribution = approximate(B = 90000))
###Posthocs
print(pvalue(Tukey_posthoc, method = "single-step"))
}
## [1] 0.03498889
## 99 percent confidence interval:
## 0.03343025 0.03659692

```

Test de Kruskal-Wallis

```

print(leveneTest(Teneur.Vitamine.C,Variété))
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 4 0.8547 0.5076
##      20
kruskalwallis <- kruskal_test(Teneur.Vitamine.C ~ Variété, data = exo1TD1,
                             distribution = "asymptotic")
kruskalwallis_MC <- kruskal_test(Teneur.Vitamine.C ~ Variété, data = exo1TD1,
                                distribution = approximate(B = 90000))
print(kruskalwallis)
##
## Asymptotic Kruskal-Wallis Test
##
## data:  Teneur.Vitamine.C by Variété (V1, V2, V3, V4, V5)
## chi-squared = 9.5241, df = 4, p-value = 0.04925
print(kruskalwallis_MC)
##
## Approximative Kruskal-Wallis Test
##
## data:  Teneur.Vitamine.C by Variété (V1, V2, V3, V4, V5)
## chi-squared = 9.5241, p-value = 0.03552

```

p-value KW

```

print(pvalue(kruskalwallis))
## [1] 0.04925373
print(pvalue(kruskalwallis_MC))

```

```
## [1] 0.03552222
## 99 percent confidence interval:
## 0.03395201 0.03714177
if (pvalue(kruskalwallis) <= 0.05) {
NDWD <- oneway_test(Teneur.Vitamine.C ~ Variété, data = exo1TD1,
  ytrafo = function(data) trafo(data, numeric_trafo = rank),
  xtrafo = function(data) trafo(data, factor_trafo = function(x)
  model.matrix(~x - 1) %*% t(contrMat(table(x), "Tukey"))),
  teststat = "max", distribution = approximate(B = 90000))
#Variable
### global p-value
print(pvalue(NDWD))
### adjusted posthocs
print(pvalue(NDWD, method = "single-step")) }
## [1] 0.03628889
## 99 percent confidence interval:
## 0.03470222 0.03792481
##
## [1] 0.03628889
## 99 percent confidence interval:
## 0.03470222 0.03792481
```