T. D. n° 2
Exemples classiques de dispositifs expérimentaux

- **Exercice 1 :** ANOVA 1F
- **Exercice 2 :** ANOVA 1F et transformations stabilisatrice de variance.
- **Exercice 3 :** Plan en blocs complets
- **Exercice 4 :** Deux facteurs croisés
- **Exercice 5 :** Deux facteurs blocs croisés
- **Exercice 6 :** Deux facteurs déséquilibrés. Plan non-orthogonal.
- **Exercice 7 :** Deux facteurs croisés avec un facteur bloc.
- **Exercice 8 :** Deux facteurs emboités.
- **Exercice 9 :** ANOVA 1F à mesures répétées.
- **Exercice 10 :** ANOVA 2F à mesures répétées (I).
- **Exercice 11 :** ANOVA 2F à mesures répétées (II).

Exercice 1. 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 38; 38; 41; 42 45; 50; 65	29; 42; 44; 44 45; 48; 48; 52 56; 56; 58; 58 60; 61; 63; 63 69	26; 26; 30; 30 33; 36; 38; 38 39; 46; 47; 51 51; 56; 75

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

Nous irradions 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 :

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

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 3. 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 :

Essai	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 4. 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 5. Comparaison de l'évaluation de la résistance d'un même type de ciment.

Davies et Goldsmith [1] 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

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

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.
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 facteur 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 » ?

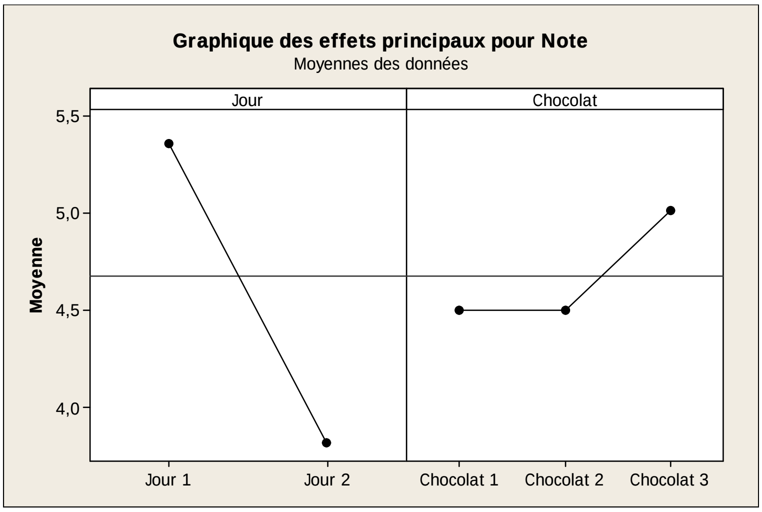
Exercice 6. Deux facteurs croisés déséquilibrés. Analyse sensorielle de trois chocolats [2]

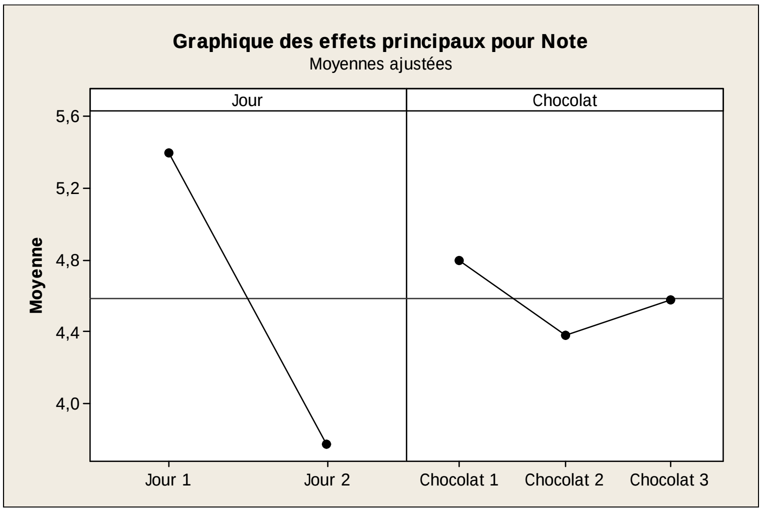
Lors d'un test de dégustation, on s'intéresse à l'appréciation globale de trois chocolats. Pour cela, 45 juges ont participé à cette évaluation qui a eu lieu sur 2 jours (on dispose de 15 échantillons par chocolat). Les notes d'appréciation des juges, comprises en 0 et 7, sont données dans le tableau 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 5.2 6.6	4.2 5.2 5 4.4 5.4 4.8 4.4 5.6 6	4.6 5.6 5.2 4.8 5.8 6.4 5 6.2 5.2 5.4 6.4
Jour 2	3.2 4 3.6 4.2 3.8 3.4 4 4.4 4.6 4	3.2 3.6 4.2 4 3.6 4	3 3.8 3.4 4.4

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 ($\widehat{\mu} + \widehat{\alpha}_i$). 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.

2. Les données de cet exercice sont tirées du livre d'exercices de François Husson et de Jérôme Pagès intitulé *Statistiques générales pour utilisateurs*, éditions PUR.





Exercice 7. Deux facteurs croisés avec un facteur bloc. Détermination d'une fumure optimale pour le blé [3]

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 neufs 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.

3. Les données de cet exercice sont inspirées du livre de Pierre Dagnélie *Statistique Théorique et Appliquée*, éditions de Boeck.

Fumures		Bloc		
P_2O_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 8. Deux facteurs emboîtés. Cyclamens [4]

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 milleux ?

4. Les données de cet exercice sont tirées du livre de Pierre Dagnélie *Statistique Théorique et Appliquée*, éditions de Boeck.

Exercice 9. 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	Dose (j)					
i	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 10. 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 11. 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.

		Problème statistique $j=1$		Problème ingénierie $j=2$	
	Sujet i	Nouveau modèle $k=1$	Ancien modèle $k=2$	Nouveau modèle $k=1$	Ancien modèle $k=2$
1	Jones	3.1	7.5	2.5	5.1
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

a) 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.
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) Essayer d'analyser ces données avec un modèle pour un plan expérimental en blocs complets.
e) Essayer d'analyser ces données avec un modèle pour mesures répétées.
f) Que conclure ?