Analyse de la variance à un facteur (ANOVA) : comparaison de k moyennes sur des groupes indépendants

Exercice 11

Un éditeur veut choisir entre trois couvertures possibles pour une revue. A cet effet, il a fait noter chaque couverture par un groupe de 5 sujets. Les trois groupes ainsi constitués sont indépendants. Les notes obtenues sont les suivantes:

Couv. 1	Couv. 2	Couv. 3
14	16	14
6	14	16
12	8	14
10	8	14
8	14	12

Le test indique-t-il une différence significative entre les trois couvertures?

Réponses : Exercice traité en CM. Rappel des résultats.

Calcul des sommes de carrés :

	C1	*C2*	*C3*	*Total*
$T_{\cdot j}$	*50*	*60*	*70*	*180*
$T_{\cdot j}^2$	*2500*	*3600*	*4900*	
n_j	*5*	*5*	*5*	*15*
$\frac{T_{\cdot j}^2}{n_j}$	*500*	*720*	*980*	*2200*
$\sum x_{ij}^2$	*540*	*776*	*988*	*2304*

$$SC_1 = 2200 - \frac{180^2}{15} = 40 \;;\; SC_2 = 2304 - 2200 = 104 \;;\; SC_T = 144$$

Source	*SC*	*ddl*	*CM*	*F*
$\mathcal{A}$	40	2	20	$F_{obs} = 2.31$
Résiduelle	104	12	8.67	
Total	144	14		

Au seuil de 5%, $F_{crit}(2,12) = 5.10$*. La différence entre les groupes n'est donc pas significative.*

Complément : Modèle de score. Chaque observation x_{ij} *peut s'interpréter comme la somme de trois termes :*

$$x_{ij} = \mu + a_j + e_{ij}$$

avec les règles suivantes :

- μ *est la moyenne de la variable* X *étudiée (la même, quel que soit l'individu ou le groupe) ;*
- a_j *est un effet dû au groupe (le même pour tous les individus d'un groupe), nul en moyenne ;*
- e_{ij} *est une variation due au hasard, spécifique à chaque observation, de moyenne nulle dans chaque groupe.*

Sur l'exemple traité, cette décomposition s'écrit :

$$\begin{pmatrix} 14 & 16 & 14 \\ 6 & 14 & 16 \\ 12 & 8 & 14 \\ 10 & 8 & 14 \\ 8 & 14 & 14 \end{pmatrix} = \begin{pmatrix} 12 & 12 & 12 \\ 12 & 12 & 12 \\ 12 & 12 & 12 \\ 12 & 12 & 12 \\ 12 & 12 & 12 \end{pmatrix} + \begin{pmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \\ -2 & 0 & 2 \\ -2 & 0 & 2 \\ -2 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 4 & 4 & 0 \\ -4 & 2 & 2 \\ 2 & -4 & 0 \\ 0 & -4 & 0 \\ -2 & 2 & -2 \end{pmatrix}$$

Les sommes des carrés inter-groupes et intra-groupes se retrouvent alors comme sommes des carrés des éléments des deux dernières matrices (par exemple : $40 = 5\times(-2)^2+5\times 2^2$*).*

Exercice 12

Dans un établissement scolaire, on a réparti les élèves en trois classes de troisième; les notes ci-dessous sont celles obtenues par les élèves en mathématiques au Brevet des Collèges. Peut-on dire que ces trois classes sont équivalentes? Si oui, quelles seraient les caractéristiques de la population résultant de la fusion des trois groupes?

G1	G2	G3
14	8	7
15	18	8
20	3	11
7	12	11
8	15	20
13	8	14
10	7	13
1	11	13
12	8	10
16	14	12
17	14	12
17	9	13
11	9	12
6	9	14
16	10	8

G1	G2	G3
8	14	13
10	15	12
11	14	8
11	13	8
7	10	11
10	12	15
11	10	8
12	12	14
11	12	16
8	11	13
	10	12
	10	15
	10	
	12	

Vérifier l'exactitude des tableaux ci-dessous et conclure.

	G1	G2	G3	Totaux	
n_j	25	29	27	81	
T_j	282	320	323	925	10563,27
Σx_{ij}^2	3600	3694	4091	11385	
T_j^2/n_j	3180,96	3531,03	3864,04	10576,03	
Inter	12,76				
Total	821,73				

Sources de variations	Sommes des carrés	DDL	Carrés moyens	F
Inter	12,76	2	6,38	0,62
Intra	808,97	78	10,37	
Total	821,73	80		

Réponses: Au seuil de 5%, $F_{crit}(2,78) = 3.1$. La différence entre les groupes n'est donc pas significative. De plus, l'obtention d'un F_{obs} inférieur à 1 semblerait indiquer (sans pour autant le montrer) que les classes n'ont pas été constituées au hasard, mais qu'elles ont, au contraire, été rendues artificiellement homogènes : on a composé les trois classes de façon qu'elles soient de niveau équivalent.

Exercice 13

Reprendre les données de l'exercice 8 (dossier pedago). Comparer les moyennes des deux groupes à l'aide d'une analyse de variance. Comparer les résultats avec ceux obtenus au premier semestre, à l'aide de la statistique T.

Réponses:

Les calculs intermédiaires sont résumés dans le tableau suivant :

	Péda1	Péda2	Totaux	
n_j	10	10	20	
T_j	32.5	42.5	75	281.25
Σx_{ij}^2	124.25	204.75	329	
T_j^2/n_j	105.625	180.625	286.25	
Inter	5.00			
Total	47.75			

Le tableau d'analyse de variance est donc :

Sources de variation	Sommes des carrés	DDL	Carrés moyens	F
Inter	5,0	1	5,0	2,11
Intra	42,75	18	2,375	
Total	47,75	19		

Au seuil de 5%, $F_{crit}(1,18) = 4.41$. Hypothèse H_1 rejetée.

Comparaison possible avec l'exercice vu au premier semestre : $t_{obs}^2 = (-1.45)^2 = 2.10$, c'est-à-dire la valeur de F.

Enoncé 14 *Données Bransfor*

On reprend une expérience de Bransford et al. (1972), dans laquelle on demande à des sujets d'écouter le texte suivant:

> "Si les ballons éclatent, le son ne portera pas puisque tout sera bien trop loin du bon étage. Une fenêtre fermée empêchera également le son de porter, surtout depuis que les immeubles récents sont correctement isolés. Comme l'essentiel de l'opération dépend d'une arrivée correcte d'électricité, un fil cassé causerait bien des problèmes. Evidemment, le type peut hurler. Mais la voix humaine n'est pas assez puissante pour porter bien loin. Un problème supplémentaire serait qu'une corde casse sur l'instrument. Alors il serait impossible d'accompagner le message. C'est clair que la meilleure situation impliquerait la plus petite distance. Alors, il y aurait bien moins de problèmes potentiels. Avec un contact en face à face, un bien petit nombre de choses pourrait gêner."

Le but visé par Bransford *et al.* est de montrer l'importance du contexte dans la compréhension et la mémorisation d'un texte. Pour ce faire, ils utilisent quatre groupes expérimentaux:

1. Un groupe "sans contexte" entend simplement le texte.
2. Le groupe "avec contexte avant" regarde une figure suggérant un contexte approprié pendant qu'il entend le texte.
3. Le groupe "avec contexte après" entend le texte puis regarde la figure précédente.
4. Le groupe "avec contexte partiel" regarde une figure suggérant un contexte inapproprié pendant qu'il entend le texte.

A proprement parler cette étude comprend un groupe expérimental (le groupe 2: contexte pendant) et trois groupes contrôles (les groupes 1, 3 et 4). Les groupes contrôles doivent permettre d'éliminer des explications concurrentes (en particulier, effet facilitateur sur la mémoire de l'imagerie, de l'aspect concret du matériel, etc.). L'expérimentateur s'attend, donc, à observer une performance pour le groupe 2 supérieure aux trois autres groupes.
Il choisit de mesurer le comportement des sujets par deux Variables Dépendantes: une note de compréhension donnée par les sujets (de 0 à 7, avec 0 indiquant l'incompréhension totale), et le nombre d'idées correctement rappelées (Bransford découpe le texte en 14 idées, essayez de les retrouver !). Quoique cette dernière Variable Dépendante soulève de délicats problèmes de codage (e.g., à partir de quel moment une idée est présente...), elle reflète clairement l'intérêt des auteurs de cette expérimentation.
Dans cette expérience, on utilise vingt sujets répartis en quatre groupes. Les résultats, pour la Variable Dépendante "nombre d'idées rappelées" (maximum 14) se trouvent ci-dessous (mais avant, faites ce que doit faire un bon expérimentateur: prenez une feuille et détaillez les cinq premières étapes du test statistique avant de partir à la pêche aux résultats):

Résultats de l'expérience				
	G.1	G.2	G.3	G.4
	3	5	2	5
	3	9	4	4
	2	8	5	3
	4	4	4	5
	3	9	1	4
$T_{.j}$	15	35	16	21
n_j	5	5	5	5
$\frac{T_{.j}}{n_j}$	3	7	3.2	4.2
Σx_{ij}^2	47	267	62	91

Justifiez les calculs et le tableau d'ANOVA suivants :
Table d'ANOVA:

Source	ddl	SC	CM	F_{cal}	$Pr(F > F_{cal})$
$\mathcal{A}$	3	50.95	16.98	7.22 **	.00288
$\mathcal{S}(\mathcal{A})$	16	37.60	2.35		
Total	19	88.55			

Si on utilise la procédure des valeurs critiques:
** $F_{critique} = 5.29$, au seuil $\alpha = .01$; $F_{cal} > F_{critique}$. On rejette H_0.
Les cinq étapes du test sont évidemment :

1. *Formulation des hypothèses statistiques H_0 et H_1. Ici :*
   H_0 : dans les 4 conditions, les moyennes dans la population parente sont égales
   H_1 : les 4 moyennes ne sont pas toutes égales.

2. *Choix du test : ici, une analyse de variance à un facteur. Statistique : F.*

3. *Distribution de la statistique de test : ici, le F de Fisher Snedecor avec $ddl_1 = 3$ (nombre de groupes – 1) et $ddl_2 = 16$ (nombre d'observations – nombre de groupes).*

4. *Seuil de signification choisi : ici, $\alpha = 1\%$.*

5. *Règle de décision : détermination des zones d'acceptation et de rejet de H_0. Ici, :*
   – Si $F_{cal} \leq 5.29$, on accepte H_0 (égalité des moyennes)
   – Si $F_{cal} > 5.29$, on refuse H_0 et on accepte H_1.

L'étude pourrait être poursuivie à l'aide de la méthode des contrastes orthogonaux (que nous ne détaillerons pas).
La première étape consiste opposer le groupe 2 aux trois autres groupes en testant l'hypothèse nulle : $3\mu_2 = \mu_1 + \mu_3 + \mu_4$. On calcule : $L_1 = 3\overline{x_2} - \overline{x_1} - \overline{x_3} - \overline{x_4} = 10.6$; $\sum a_j^2 = 3^2 + (-1)^2 + (-1)^2 + (-1)^2 = 12$; $SC_{contraste1} = \dfrac{nL^2}{\sum a_j^2} = 46.81$
Dans la formule précédente, n est le nombre d'observations par groupe. Ici, $n = 5$. Le F de Fisher associé à ce contraste est obtenu en divisant $SC_{contraste1}$ par le carré moyen résiduel 2.35 ; il vaut 19.92. Les degrés de liberté sont 1 et 16. Le résultat est donc significatif d'un comportement du groupe 2 différent de celui des autres groupes.
La méthode peut être poursuivie en opposant le groupe 4 aux groupes 1 et 3 (coefficients appliqués aux quatre moyennes : 1, 0, 1, -2) puis en opposant les groupes 1 et 3 (coefficients appliqués : 1, 0, -1, 0).
Pourquoi s'agit-il de contrastes orthogonaux *?*
Réponse : Les "vecteurs" associés aux coefficients des trois contrastes, à savoir $V_1 = (-1, 3, -1, -1)$, $V_2 = (1, 0, 1, -2)$, $V_3 = (1, 0, -1, 0)$ sont deux à deux orthogonaux (par exemple, $V_1 \cdot V_2 = -1 \times 1 + 3 \times 0 + (-1) \times 1 + (-1) \times (-2) = 0$), ce qui garantit l'indépendance des résultats des trois tests.
Une autre grandeur intéressante est le coefficient (souvent noté η^2) d'estimation de l'intensité de l'effet de la variable indépendante. Dans le cas d'une analyse de variance à un facteur, il est défini par :

$$\eta^2 = \frac{SC_{inter}}{SC_{total}}$$

Il vaut donc ici : $\eta^2 = 0.58 = 58\%$.
Signification : 58% de la variance de la Variable Dépendante est expliquée par la Variable Indépendante (les différentes conditions expérimentales).
η^2 est aussi le carré d'un coefficient de corrélation. η peut en effet être obtenu comme coefficient de la corrélation entre l'ensemble des données observées d'une part, et la série de données obtenue en remplaçant chaque observation par la moyenne de son groupe

d'autre part. Sur notre exemple, soit U la série des données observées et V la série des données du modèle ainsi obtenu.

u_i	3	3	2	4	3	5	9	8	4	9	2	4	5	4	1	5	4	3	5	4
v_i	3	3	3	3	3	7	7	7	7	7	3.2	3.2	3.2	3.2	3.2	4.2	4.2	4.2	4.2	4.2

On obtient : $r(U, V) = 0.7585$ *et* $r^2(U, V) = 0.575$.

Enoncé 15 *Données Loftus*
Elisabeth Loftus (Loftus et Palmer 1974) – dans une série d'expérimentations sur le thème du témoignage – désire mettre en évidence l'influence de la tournure d'une question sur la réponse de témoins. Pour ce faire, elle montre à ses sujets, un film décrivant un accident de voiture. Elle pose, ensuite, une série de questions aux sujets. Parmi celles-ci se trouve une des cinq versions d'une question relative à la vitesse des véhicules. Voici ces versions:

1) HIT: About how fast were the cars going when they *hit* each other? (A environ quelle vitesse allaient les voitures quand elles se sont "percutées").

2) SMASH: About how fast were the cars going when they *smashed* each other? (To smash : écraser, heurter avec violence).

3) COLLIDE: About how fast were the cars going when they *collided* each other? (To collide: entrer en collision, s'emboutir).

4) BUMP: About how fast were the cars going when they *bumped* each other? (To bump: cogner, frapper).

5) CONTACT: About how fast were the cars going when they *contacted* each other? (To contact: entrer en contact).

Les sujets répondaient en indiquant une vitesse exprimée en miles (nous sommes aux U.S.A). Voici les résultats obtenus (lors d'une réplique de l'expérience):

HIT	SMASH	COLLIDE	BUMP	CONTACT
22	38	43	47	27
29	40	39	29	24
33	50	32	58	46
50	45	44	34	37
19	48	29	36	31
37	56	44	43	37
33	52	45	25	34
43	47	33	58	18
40	39	48	24	28
34	40	37	31	26

Après avoir identifié les variables dépendante(s) et indépendante(s), vous tirerez les conclusions de cette expérimentation.
Pour vous aider voici quelques statistiques pour chaque groupe:

	$T_{.j}$	$T_{.j}/n_j$	$T_{.j}^2/n_j$	$\Sigma_j x_{ij}^2$
Gr. 1	340	34.0	11560	12338
Gr. 2	455	45.5	20702.5	21043
Gr. 3	394	39.4	15523.6	15894
Gr. 4	385	38.5	14822.5	16241
Gr. 5	308	30.8	9486.4	10060
Total	1882		72095	75576

La Variable Dépendante est évidemment la vitesse exprimée en miles. La Variable Indépendante est le type de verbe utilisé pour poser la question sur la vitesse des voitures. Manifestement, E. Loftus veut montrer que les "sous-entendus" des verbes sont pris en compte par les sujets dans leur décision sur la vitesse (e.g., les sujets utilisent la signification implicite des verbes comme une source d'information). Le point d'importance dans cette expérience est de remarquer que E. Loftus désire généraliser ses résultats à l'ensemble des verbes signifiant quelque chose comme "entrer en contact". Quoiqu'elle n'ait pas, à proprement parler, sélectionné ses verbes au hasard, elle les juge représentatifs de l'ensemble des verbes de mouvement. Le problème ici est de décider si le facteur expérimental est fixé ou aléatoire. Si l'on admet que les verbes choisis par Loftus représentent un échantillon représentatif, on décidera que le facteur est aléatoire (cf. La polémique initiée par Clark 1973). Si l'on juge que les modalités sont choisies en fait arbitrairement. on décidera que le facteur est fixé, et les conclusions de l'étude se restreignent aux modalités effectivement présentes dans l'expérimentation. Quelle que soit la décision prise, elle sera criticable.

Ici, le distinguo entre facteur fixé et aléatoire peut paraître sans importance car la décision (rejet ou non de l'hypothèse nulle) sera identique dans les deux cas. *Ce ne sera plus le cas dans des plans d'expérience plus complexes.* En fait, l'essentiel de l'argument de Clark (1973) est de montrer qu'une partie des recherches utilisant du matériel linguistique aboutit à des conclusions SCIENTIFIQUES erronées du fait de la confusion entre facteurs fixés et aléatoires (cf. aussi les réponses de Wike et Church 1976). Clark défend l'idée qu'une partie des conclusions de la psychologie du langage est invalide pour avoir cru que des facteurs aléatoires étaient fixes. A cette attaque répond Chastaing (1986) qui démontre méthodologiquement qu'une autre partie de la psychologie du langage est invalide d'avoir cru que des facteurs fixes étaient aléatoires !

Dans le cas présent, le choix entre les deux modèles n'a pas d'influence sur les résultats de l'analyse statistique: on aboutit à des conclusions statistiques identiques (mais pas à des interprétations psychologiques identiques !). L'analyse de variance permet de conclure en tout cas à un effet sur la vitesse estimée, du type le verbe utilisé pour poser la question. On obtient le tableau d'analyse de variance suivant :

Source	ddl	SC	CM	F_{cal}	$Pr(F > F_{cal})$
Expérimentale	4	1256.52	314.13	4.06 **	.0069
Erreur	45	3481.00	77.36		
Total	49	4737.52			

Ainsi, le type de verbe employé pour interroger les sujets sur la vitesse des véhicules, influence l'estimation qu'ils donnent ($F_{cal}(4, 45) = 4.06,\ p < .05$). On remarque la vitesse élevée induite par *to smash*. Nous pourrions poursuivre cet exemple en essayant d'apprécier les différences entre ces différents verbes les uns par rapport aux autres).

Enoncé 16 *Données Besancon*

On fait subir à 30 élèves d'une école de Besançon une épreuve de "précision perceptive" qui consiste à évaluer un nombre de points sur une diapositive projetée pendant un temps relativement court (une demi-seconde). Les auteurs de cette expérience pensent que la présence d'un témoin peut influencer la performance des sujets dans cette tâche perceptive. Pour vérifier cette idée, les expérimentateurs divisent leur échantillon en trois groupes – chaque enfant étant affecté à un groupe en utilisant une "table de nombres au hasard". Dans le premier groupe (A1) l'expérience est effectuée sans témoin; dans le second groupe (A2) l'enfant accomplit sa tâche en compagnie d'un témoin présenté par l'expérimentateur comme un spécialiste; dans le troisième groupe (A3), le témoin est présenté comme un simple curieux. On répète – pour chaque sujet – vingt-cinq fois l'expérience. Et l'on retient pour chaque sujet la moyenne des écarts absolus (i.e. en ignorant le signe) entre l'estimation fournie et le nombre exact de points.

Les expérimentateurs s'attendent à trouver des différences entre les trois conditions expérimentales; mais, plus précisément, entre la condition "sans témoin" et la condition "témoin simple curieux" (cette différence leur permettrait de contredire un de leurs collègues qui avançait dans une expérience voisine que le témoin n'agissait que parce que les enfants le jugeait spécialiste). Les auteurs veulent, également vérifier l'existence d'un effet spécifique à la condition "témoin spécialiste".

Questions:

Pourquoi les expérimentateurs décident-ils de prendre l'écart absolu et non pas – par exemple – l'écart relatif. Tout de même, pourquoi retiennent-ils la moyenne des vingt-cinq essais, plutôt qu'un seul essai?

Quelle est la (les) variable(s) indépendante(s), la (les) variable(s) dépendante(s)?

Après avoir traduit en termes statistiques les objectifs des expérimentateurs, peut-on penser que ces objectifs sont atteints? Voici les résultats obtenus:

Condition A1	140	124	118	115	110	110	108	104	102	90
Condition A2	170	164	161	158	156	148	143	140	130	126
Condition A3	136	120	112	104	102	96	92	84	81	75

Eléments de réponses. Calculs intermédiaires:

	A1	*A2*	*A3*	*Totaux*	
n_j	*10*	*10*	*10*	*30*	
T_j	*1121*	*1496*	*1002*	*3619*	*436572.03*
Σx_{ij}^2	*127303*	*225746*	*103582*	*456637*	
T_j^2/n_j	*125664.1*	*223801.6*	*100400.4*	*449866.1*	
Inter	*13294.07*				
Total	*20064.97*				

Le tableau d'analyse de variance est donné par:

Source	*ddl*	*SC*	*CM*	F_{cal}
Inter-groupes	*2*	*13294.1*	*6647.03*	*26.51 ***
Intra-groupes	*27*	*6770.9*	*250.77*	
Total	*29*	*20065*		

Les trois groupes ne sont donc pas équivalents. La méthode peut être poursuivie en décomposant la variation intra-groupes selon les deux contrastes orthogonaux suggérés par l'énoncé :
$L_1 = \overline{x_1} - \overline{x_3} = 11.9$
$L_2 = 2\overline{x_2} - \overline{x_1} - \overline{x_3} = 86.9$

On obtient alors :
$SC_{contraste1} = 708.05; F = 2.82; Pr(F) = 0.10$
$SC_{contraste2} = 12586.02; F = 50.19; Pr(F) = 1.3 \times 10^{-7}$
L'expérience ne met donc pas de différence en évidence entre les conditons "sans témoins" et "témoin simple curieux" mais par contre, montre un comportement différent dans la condition "témoin spécialiste".

Enoncé 17
Un chercheur a soumis quatre groupes de cinq élèves à un apprentissage de "résolutions de problèmes mathématiques". Chaque groupe apprend avec une méthode pédagogique propre: le premier avec une méthode uniquement verbale, le second avec une méthode écrite, le troisième avec un schéma annoté, le quatrième avec une série de schémas annotés. L'apprentissage dure une heure pour chaque groupe, et le même contenu est présent. Deux jours après l'apprentissage, les sujets sont soumis à un test de raisonnement mathématique. Ce test provient des travaux d'autres chercheurs qui ont étalonné ce test sur une population comparable à celle dont provient l'échantillon d'enfants utilisé ici; le résultat de ce test est une note (de 0 à 35: plus la note est élevée, meilleur est le résultat).
Quelle est la Variable Indépendante, la Variable Dépendante? Comment l'expérimentateur traitera-t-il les résultats de son expérience (souvenez-vous qu'il faut pouvoir répondre à cette question avant de recueillir les résultats!)?
En outre, l'auteur a mis au point cette expérience pour vérifier certaines hypothèses précises:

1. La méthode verbale diffère-t-elle de l'ensemble des autres méthodes
2. La méthode écrite diffère-t-elle des méthodes avec schémas (un ou plusieurs)'?
3. Le nombre de schémas a-t-il une influence décelable sur la performance?

L'auteur peut-il répondre simultanément à ces différentes questions, et quelles seront les réponses? Interprétez — en vous justifiant — les résultats obtenus et concluez.
Voici les résultats :

GROUPE EXPÉRIMENTAL			
A1	A2	A3	A4
6	14	22	23
13	10	11	19
16	14	19	25
14	19	19	24
14	25	23	25

Elements de réponses. Calculs intermédiaires :

	A1	*A2*	*A3*	*A4*	*Totaux*	
n_j	*5*	*5*	*5*	*5*	*20*	
T_j	*63*	*82*	*94*	*116*	*355*	*6301.25*
Σx_{ij}^2	*853*	*1478*	*1856*	*2716*	*6903*	
T_j^2/n_j	*793.8*	*1344.8*	*1767.2*	*2691.2*	*6597*	
Inter	*295.75*					
Total	*601.75*					

Le tableau d'analyse de variance est donné par :

Source	*ddl*	*SC*	*CM*	F_{cal}	$Pr(F > F_{cal})$
Inter-groupes	*3*	*295.75*	*98.58*	*5.15*	*0.011*
Intra-groupes	*16*	*306*	*19.125*		
Total	*19*	*601.75*			

La méthode peut être poursuivie en décomposant la variation inter-groupes selon les trois contrastes orthogonaux suggérés par l'énoncé :
$L_1 = 3\overline{x_1} - \overline{x_2} - \overline{x_3} - \overline{x_4} = -20.6$
$L_2 = 2\overline{x_2} - \overline{x_3} - \overline{x_4} = -9.2$
$L_3 = \overline{x_3} - \overline{x_4} = -4.4$

On obtient alors :
$SC_{contraste1} = 176.82; F = 9.24; Pr(F) = 0.0067$
$SC_{contraste2} = 70.53; F = 3.69; Pr(F) = 0.07$
$SC_{contraste3} = 48.40; F = 2.53; Pr(F) = 0.13$

Exercice 18

Ci-dessous figure un extrait d'un ouvrage de statistiques relatif à un test statistique qui n'a pas été étudié en cours, le test H de Kruskal et Wallis.

Liaison entre un caractère quantitatif et un caractère qualitatif à k classes ($k > 2$).

C'est le problème appelé, dans les chapitres précédents, “comparaison de plusieurs moyennes” et traité par analyse de variance. Le test non paramétrique correspondant le plus usuel est le test H de Kruskal et Wallis.
On classe les observations de l'ensemble des k séries, comme on le faisait pour les deux séries dans les tests précédents, puis on calcule les rangs moyens $\overline{W}_1, \overline{W}_2, \ldots, \overline{W}_k$ et le rang moyen $\overline{W}$, ce dernier valant $\dfrac{N+1}{2}$ si N représente le nombre total d'observations.
Dans l'hypothèse nulle, $\overline{W}_1, \overline{W}_2, \ldots, \overline{W}_k$ ne doivent pas trop s'écarter de $\overline{W}$, de sorte que les quantités $\left(\overline{W}_i - \overline{W}\right)^2$ ne doivent pas être trop grandes. On montre que, sous l'hypothèse nulle, la statistique :

$$H = \frac{1}{N} \frac{\sum n_i \left(\overline{W}_i - \overline{W}\right)^2}{(N+1)/12}$$

suit approximativement une loi du χ^2 à $k-1$ degrés de liberté.
Dans cette expression, les n_i désignent les effectifs des diverses séries. L'approximation n'est valable que s'ils atteignent tous la dizaine, à la rigueur 5.

1) Dans quelles situations ce test doit-il être préféré à une analyse de variance?
2) *Première application.* Une variable $\mathcal{A}$ comporte trois modalités a_1, a_2, a_3. Pour chaque modalité, on dispose de 2 ou 3 observations d'une variable numérique. Ces observations sont rassemblées dans le tableau ci-dessous.

a_1	a_2	a_3
14	12	14
16	15	18
13		14

Construire sur cet exemple le protocole des rangs (W_i) et calculer les rangs moyens et la statistique H. (Vu le faible nombre d'observations, on s'abstiendra ici d'effectuer le test).
3) *Deuxième application.* Afin de constituer un groupe suffisamment important en vue d'une recherche, un chercheur teste 3 groupes de 10 sujets. Les rangs moyens observés sur les trois groupes sont les suivants :

$$\overline{W}_1 = 9,8 \ ; \ \overline{W}_2 = 24,05 \ ; \ \overline{W}_3 = 12,65.$$

Peut-on considérer que les trois groupes testés sont issus d'une même population ?
Réponses : 1) Un test non paramétrique doit être préféré lorsque la variable est ordinale, ou lorsque l'on ne peut pas faire d'hypothèse concernant la normalité des distributions dans les populations parentes.
2) Le protocole des rangs est donné par :

a_1	a_2	a_3
4	*1*	*4*
7	*6*	*8*
2		*4*
4.33	*3.5*	*5.33*

Dans ce cas, $\overline{W} = \frac{36}{8}$; $H = \frac{1}{8}\frac{3(4.33-4.5)^2+2(3.5-4.5)^2+3(5.33-4.5)^2}{9/12} = 0.6944$
3) Dans ce cas, $H_{obs} = 14.67$*,* $ddl = 2$ *et, au seuil de 1%,* $\chi^2_c = 9.21$*. On conclut donc à l'hétérogénéïté des groupes.*

Plans d'expériences, interactions

Enoncé 19

Déterminez VI et VD dans les hypothèses suivantes :
a) Un individu est d'autant plus attaché à une opinion qu'il s'est davantage engagé à la défendre.
b) Le degré de violence d'un événement modifie sa mémorisation.
c) Je l'aime plus qu'hier et bien moins que demain.
d) Les individus agressifs assurent plus souvent le leadership dans un groupe, mais ils en satisfont moins les membres que les leaders non agressifs.
e) On retient plus facilement un matériel significatif qu'un matériel dépourvu de sens.
f) C'est dans les vieux pots que l'on fait les meilleures confitures.
g) On restitue d'autant mieux une information que celle-ci est rappelée dans le même contexte que celui où elle a été apprise.

h) Les aînés sont plus anxieux que les puînés.
i) L'influence d'un discours est d'autant plus importante que l'orateur possède du prestige.
j) Le nombre de conversations au cours d'un repas dépend étroitement de la disposition des individus autour de la table.

Indications de réponses : a) VI = intensité de l'engagement, VD = une mesure numérique liée à l'attachement à l'opinion.
b) VI = degré de violence d'un événement, VD = mesure numérique liée à la mémorisation de l'événement.
c) VI = Date, VD = mesure du degré d'amour.
d) Ici, deux hypothèses commbinées :
VI = agressivité, VD liée à la prise de leadership
VI = agressivité, VD = mesure de la satisfaction des membres du groupe.
e) VI = significativité du matériel, VD = évaluation numérique de la performance mnésique.
f) VI = âge du pot, VD = mesure de la qualité de la confiture.
g) VI1 = contexte d'apprentissage, VI2 = contexte de rappel (ou lien (même / différent) entre contexte de rappel et contexte d'apprentissage), VD = évaluation numérique de la performance mnésique.
h) VI = rang de naissance, VD = mesure de l'anxiété.
i) VI = prestige de l'orateur, VD = mesure du degré d'influence.
j) VI = disposition autour de la table, VD = nombre de conversations.

Enoncé 20 *Données Schizo*

Dans une série d'expériences destinées à éclaircir la notion de "maladie mentale" on soumet des sujets diagnostiqués comme schizophrènes et des sujets normaux à une épreuve de "formation de concept". Tous les sujets retenus pour participer à l'expérience doivent posséder un Q.I. compris entre 100 et 105. Pourquoi ?
On compte pour chaque sujet le nombre d'essais nécessaires pour arriver à former un nouveau concept.
Dans cette expérience on utilise deux ensembles de stimuli: Le premier contient des images illustrant l'approbation sociale, le second des images illustrant la désapprobation sociale.
L'auteur de cette expérience émet les prédictions suivantes (qui découlent de certaines théories de la personnalité et des performances intellectuelles):
a) Les sujets normaux devront arriver plus rapidement que les schizophrènes à accomplir l'épreuve et ce, indépendamment de la nature des images ;
b) Les sujets normaux ne seront pas influencés par la nature des stimuli;
c) Les schizophrènes devront réussir moins facilement les épreuves comportant des images exprimant la désapprobation sociale que les épreuves décrivant l'approbation sociale.
Quel est le plan d'expérience utilisé ? Les prédictions du chercheur se traduisent par des prédictions sur les hypothèses statistiques, lesquelles ?

Eléments de réponses : Facteurs : sujet $\mathcal{S}$, maladie $\mathcal{M}_2$, nature des images $\mathcal{I}_2$. L'effet du facteur QI a été éliminé par le choix initial des sujets. Plan : $\mathcal{S} < \mathcal{M}_2 > *\mathcal{I}_2$. La combinaison des hypothèses a) et c) se traduit par une interaction : les sujets réussiront moins facilement que les sujets normaux, mais cette différence sera plus marquée dans la condition i_2 que dans la condition i_1.

Enoncé 21

Une expérience a été menée en utilisant un plan de la forme $\mathcal{S}_8 < \mathcal{A}_3 * \mathcal{B}_2 > *\mathcal{C}_4$.
Combien de groupes indépendants de sujets a-t-on constitué ?

Combien de sujets différents ont été utilisés pour cette expérience ?
Combien d'observations différentes d'un même sujet a-t-on effectué ?

Réponses : Les sujets sont emboîtés dans $< \mathcal{A}_3 * \mathcal{B}_2 >$*. On a 6 combinaisons de modalités et donc 6 groupes de sujets.*
L'indice dans $\mathcal{S}_8 < \mathcal{A} * \mathcal{B} >$ *montre qu'il y a 8 sujets dans chaque groupe, soit en tout :* $6 \times 8 = 48$ *sujets.*
Les sujets sont croisés avec $\mathcal{C}_4$*, on a donc fait 4 observations de chaque sujet.*

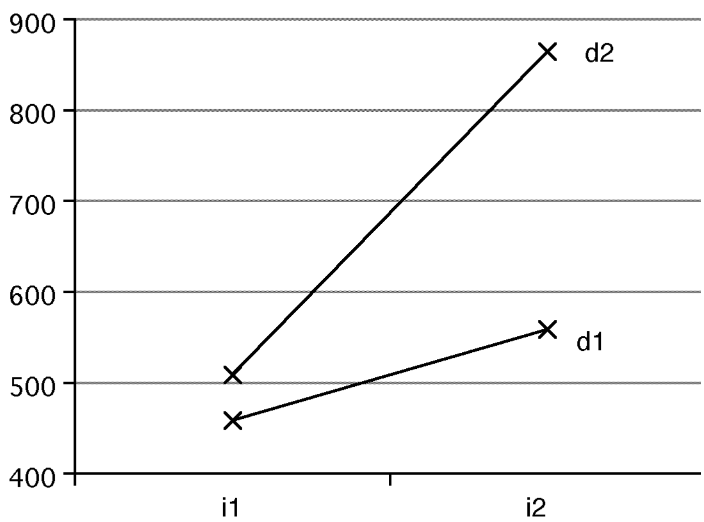
Enoncé 22

Dans une tâche de dénomination de figures géométriques, l'auteur étudie l'évolution du temps de réaction verbale en fonction de la discriminalité des figures.
Dans un premier temps, on présente aux sujets une série de figures. Pour la moitié d'entre eux, la série est constituée de 2 figures, pour l'autre moitié, de 4 figures. Dans chacun des cas, la série est constituée soit de figures facilement discriminables (triangle, carré,...) soit de figures plus complexes (octogone, décagone...).
Dans un deuxième temps, on demande à chaque sujet de nommer une figure tirée au hasard dans la série précédente et on mesure le temps de réaction verbale du sujet.
48 sujets répartis en 4 groupes de 12 ont participé à l'expérience.
Les moyennes des temps de réaction mesurés en millisecondes observés sur chacun des quatre groupes sont indiqués dans le tableau suivant :

	Discriminalité	
Incertitude	Forte	Faible
2 figures	460	510
4 figures	559	864

1) Définir la variable dépendante et les variables indépendantes prises en compte. Quel est le plan d'expérience utilisé ?
2) Au vu du tableau précédent, indiquer s'il semble y avoir une interaction entre les deux facteurs étudiés. Construire un graphe d'interaction. Commenter ce graphe en rédigeant une phrase exprimant comment se traduit l'effet d'interaction.

Eléments de réponses : 1) Le plan utilisé est ici $S_{12} < I_2 * D_2 >$*.*
2) Le temps de réaction augmente lorsque la discriminalité est plus faible. Mais cet effet est d'autant plus important que l'incertitude est élevé.



Enoncé 23 *Données Conrad*

Dans une reprise d'une expérience de Conrad (1971), on veut mettre en évidence l'hypothèse de recherche suivante: "les enfants jeunes n'utilisent pas un codage phonologique en mémoire à court terme". Pour ce faire, on sélectionne cinq enfants de 5 ans et 5 enfants de 12 ans (Variable $\mathcal{A}$, avec deux modalités). On montre à chaque enfant un certain nombre de paires d'images représentant des objets dont on s'est assuré auparavant qu'ils sont nommés d'une seule manière par les enfants. On montre les images aux enfants. Puis on retourne les images (les enfants ne voient plus que le dos des images). Ensuite, on donne aux enfants une paire d'images identiques à celles retournées. Enfin, on leur demande de placer ces nouvelles images comme les images retournées sur la table. Pour la moitié des paires d'images les noms des objets se ressemblent (e.g., noix et doigt). Pour l'autre moitié, les noms des objets ne se ressemblent pas (e.g., maison et cheval). Conrad prédit que les enfants les plus vieux réussiront dans l'ensemble mieux que les enfants les plus jeunes, mais également que les enfants les plus vieux utiliseront un codage phonologique comme mnémonique (i.e., "la parole intérieure"). De ce fait, les enfants les plus vieux devront commettre plus d'erreurs lorsque les noms se ressemblent acoustiquement que lorsque les noms diffèrent. On présente à chaque enfant cinquante paires d'images correspondant à la modalité b_1 (dissemblance acoustique), et cinquante paires d'images correspondant à la modalité b_2 (ressemblance acoustique); la Variable Dépendante choisie est le nombre de paires d'images correctement reconstituées. L'ordre de présentation est "aléatorisé" pour chaque passation (Pourquoi cette précaution?).
Essayer de traduire l'hypothèse de recherche en prédiction sur les sources de variation de l'analyse de variance.
Vous avez dû conclure que, d'une part, on s'attend à un effet principal de l'âge (qui est trivial), et, d'autre part, à un effet d'interaction: c'est le point d'importance, ou si vous préférez, le point crucial de la théorie. On retrouve, ici, le rôle essentiel de l'interaction "comme test de théorie ".

Enoncé 24 *Données Cochran*

Les données suivantes, adaptées d'une expérience de Cochran et Cox, illustrent un paradigme expérimental extrêmement courant: la comparaison de deux conditions avec contrebalancement des ordres.
Il s'agissait de comparer l'efficacité de deux types de machines à calculer ml et m2: on supposera ici que 10 sujets, s1 à sl0, ont exécuté la même séquence de calculs, successivement sur chacune des deux machines m1 et m2. Les sujets s1 à s5 ont travaillé d'abord (essai e1) avec la machine m1, puis (essai e2) avec la machine m2; les sujets s6 à s10 ont travaillé dans l'ordre inverse (m2 à l'essai e1, puis m1 à l'essai e2). Les résultats (temps d'exécution du calcul, en unités conventionnelles) sont les suivants:

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
m1	30	22	29	12	23	21	22	18	16	23
m2	14	5	17	14	8	21	13	13	7	24

Comme facteurs décrivant le protocole, nous prendrons d'abord: S (sujets: dix modalités, s1 à s10); M ou M2 (Machines: deux modalités m1 et m2); E ou E2 (essais: deux modalités e1 et e2). A ce facteur nous adjoindrons, pour des raisons qui apparaîtront plus loin, le facteur ordre O ou O2 avec:

o1: machine m1 passée à l'essai e1 et machine m2 passée à l'essai e2;

o2: machine m2 passée à l'essai e1 et machine m1 passée à l'essai e2;
N.B.: le tableau précédent correspond à la description des données selon le plan S*M2
Mais on pourrait également présenter ces données selon le tableau suivant:

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
e1	30	22	29	12	23	21	13	13	7	24
e2	14	5	17	14	8	21	22	18	16	23

Ce nouveau tableau correspondrait à la description selon le plan S*E2.
En introduisant le facteur ordre, ces mêmes données pourront encore être décrites selon l'un des plans S<O2>*M2 et S<O2 >*E2
Question principale: y a-t-il une différence d'efficacité entre les machines? Question secondaire: y a-t-il une différence entre les deux essais?
Du point de vue des objectifs de la recherche, le facteur Machine sera donc considéré comme principal, et le facteur Essai comme secondaire (ce qui n'implique nullement que, lors de la planification de l'expérience, on s'attendait à ce que l'effet du facteur Essai soit peu important; les deux ordres ont été contrebalancés précisément afin de parer à l'éventualité d'un effet même important du facteur Essai).
Réponses : Il faut ici bien comprendre que le facteur "essai" représente l'interaction entre les facteurs "machine" et "ordre" ; de même, le facteur "machine" représente l'interaction entre les facteurs "ordre" et "essai".
L'analyse, au niveau descriptif, de l'interaction entre les facteurs "ordre" et "essai" (c'est-à-dire l'analyse de l'effet "machine") pourra être faite à partir du tableau suivant obtenu à partir des moyennes calculées dans chacune des conditions e_1o_1, e_1o_2, e_2o_1, e_2o_2 :

	o_1	o_2	*Moy.*	*Diff.*
e_1	*23.2*	*15.6*	*19.4*	*7.6*
e_2	*11.6*	*20.0*	*15.8*	*-8.4*
Moy.	*17.4*	*17.8*	*17.6*	*-0.4*
Diff.	*11.6*	*-4.4*	*3.6*	*16*

Dans ce tableau, 3.6 représente deux fois l'effet "essai", 16 représente deux fois l'effet "machine". L'interaction apparaît clairement sur un graphe d'interaction.
On pourra répondre aux deux questions posées à l'aide de comparaisons de moyennes sur des groupes appareillés, en ignorant le troisème facteur. La comparaison des moyennes obtenues pour $M = m_1$ et $M = m_2$ aboutit à $T_{obs} = 3.52$, valeur significative d'une différence entre machines au seuil de 1%. La comparaison des moyennes obtenues pour $E = e_1$ et $E = e_2$ aboutit à $T_{obs} = 1.09$. La différence n'est donc pas significative. L'analyse de variance permet ici une étude plus fine. Mais, le tableau d'analyse de variance est assez complexe, car il s'agit d'un plan à mesures partiellement répétées (sujets emboîtés dans un facteur et croisés avec l'autre facteur). On obtient par exemple :

Source	*ddl*	*SC*	*CM*	F_{cal}	*Pr*
O	*1*	*0.8*	*0.8*	*0.018*	*0.89*
$S < O >$	*8*	*357*	*44.6*		
E	*1*	*64.8*	*64.8*	*3.08*	*0.11*
Interaction	*1*	*320*	*320*	*15.2*	*0.0045*
Résidu	*8*	*168.2*	*21.02*		
Total	*19*	*910.8*			