MAT7381 Exercices
Chapitre 6 - Analyse de variance

6.1 Montrez que le test t pour comparer deux moyennes est équivalent à une analyse de variance. [Utilisez le fait que si T est de loi t_ν, alors T^2 est de loi $F_{1;\nu}$.]

6.2 Démontrez la décomposition SCT = SCE + SCR, soit

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k} n_i(\bar{y}_i-\bar{y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$$

6.3 Démontrez que SCR/σ^2 ~ χ^2_{n-k} dans l'analyse de variance à un facteur.

6.4 Montrez que pour chaque i, $\bar{y}_i$ et $Q_i = \Sigma_j(y_{ij} - \bar{y}_i)^2$ sont indépendantes [et donc que toute fonction des $\bar{y}_i$ est indépendante de toute fonction de Q_i, en particulier de SCR.]

6.5 Montrez que SCE/σ^2 ~ $\chi^2_{k-1}(\lambda)$, $\lambda = \dfrac{\sum_i n_i(\mu_i-\bar{\mu})^2}{\sigma^2}$. [Soit $\mathbf{z}$ le vecteur des moyennes échantillonnales $\bar{y}_i$. Alors $\mathbf{z} \sim \mathcal{N}_k(\boldsymbol{\mu}, \sigma^2\mathbf{D}^{-1})$, où $\boldsymbol{\mu}$ est le vecteur des moyennes μ_i et $\mathbf{D}$ est la matrice diagonale dont les éléments sont les n_i. Vérifiez que SCE = $\Sigma_i n_i\bar{y}_i^{\,2} - n\bar{y}^2 = \mathbf{z}'\mathbf{A}\mathbf{z}$, où $\mathbf{A} = \mathbf{D} - \mathbf{D}\boldsymbol{e}(\boldsymbol{e}'\mathbf{D}\boldsymbol{e})^{-1}\boldsymbol{e}'\mathbf{D}$].

6.6 [*Pour un même i, les y_{ij} correspondent à une même valeur x_i*]

a) Montrez que si SCR = $\Sigma_i\Sigma_j(y_{ij} - \bar{y}_i)^2$ et SCR$_o$ = $\sum_i\sum_j(y_{ij}-\hat{\beta}_o-\hat{\beta}_1x_i)^2$ alors SCR$_o$ - SCR = $\sum_i n_i(\bar{y}_i-\hat{\beta}_o-\hat{\beta}_1x_i)^2$.

b) Montrez que SCR$_o$ - SCR et SCR sont indépendantes. [Il suffit de montrer que SCR$_o$ - SCR est fonction des moyennes $\bar{y}_i$ seulement (et non directement des y_{ij})].

c) Utilisez le fait que sous H$_o$, SCR$_o$/σ^2 et SCR/σ^2 sont des χ^2 ainsi que le fait que SCR$_o$ - SCR et SCR sont indépendantes pour montrer que [SCR$_o$ - SCR]/σ^2 est une χ^2.

6.7 Démontrez la formule de calcul

$$\sum_{i=1}^{k} n_i(\bar{y}_{i.}-\bar{y})^2 = \sum_{i=1}^{k} n_i\bar{y}_i^2 - n\bar{y}^2$$

6.8 Considérons le cas où $n_1 = n_2 = \ldots = n_k = m$. Montrez que SCE/$\sigma^2$: χ^2_{k-1} si H$_o$ est vraie. Sans supposer que H$_o$ est vraie, déterminez l'espérance de SCE. Montrez comment votre réponse justifie une région critique de la forme $F > \mathcal{F}_{k-1;k(m-1);\alpha}$.

6.9 Dans un modèle d'analyse de variance ($y_{ij} \sim \mathcal{N}(\mu_i\ ;\sigma^2)$, la variance échantillonnale $S_i^2 = \sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2/(n_i-1)$ est un estimateur sans biais de σ^2 pour chaque $i = 1, \ldots, k$.

a) Montrer que la moyenne de ces variances échantillonnales $\overline{S^2} = \sum_{i=1}^{k} S_i^2 \big/ k$ est un estimateur sans biais de σ^2.

b) Montrer que la variance de $\overline{S^2}$ est $\dfrac{2\sigma^4}{k^2}\sum_{i=1}^{k}\dfrac{1}{n_i-1}$

c) L'estimateur $\hat{\sigma}^2$ = MCR = $\sum_{i=1}^{k}(n_i-1)S_i^2/(n-k)$ est aussi un estimateur sans biais. Montrer que sa variance est $2\sigma^4/(n-k)$.

d) Montrer que parmi tous les estimateurs sans biais de la forme $\sum_{i=1}^{k} c_iS_i^2$, MCR est celui qui minimise la variance [*Suggestion* : Utiliser l'inégalité de Cauchy-Schwartz]

6.10 Dans un modèle d'analyse de variance à un facteur il est possible de tester l'hypothèse $H_o : \Sigma c_i\mu_i = a$ sans faire l'hypothèse d'homoscédasticité, à conditions que les effectifs n_i soient assez importants : la statistique $Z = \frac{\sum c_i \bar{y}_i - a}{\sqrt{\sum c_i^2 \sigma_i^2 / n_i}} \sim \mathcal{N}(0\,;\,1)$, et si les n_i sont importants on peut remplacer les σ_i par les S_i. On a donc approximativement $T = \frac{\sum c_i \bar{y}_i - a}{\sqrt{\sum c_i^2 S_i^2 / n_i}} \sim \mathcal{N}(0\,;\,1)$. Il est probable qu'on puisse mieux approximer la distribution de T par une variable de loi de Student. On a que $T = \frac{\left(\sum c_i \bar{y}_i - a\right) \Big/ \sqrt{\sum c_i^2 \sigma_i^2 / n_i}}{\sqrt{V}}$ où $V = \nu \sum c_i^2 S_i^2 / n_i \Big/ \sum c_i^2 \sigma_i^2 / n_i$ serait de loi t_ν si $V \sim \chi^2_\nu$. Certains proposent qu'on fasse la supposition que $T \sim t_\nu$, et qu'on choisisse ν de telle sorte que $E(V) = \nu$ et $Var(V) = 2\nu$. Montrez qu'alors $\nu = \frac{\left(\sum c_i^2 \sigma_i^2 / n_i\right)^2}{\sum c_i^4 \sigma_i^4 / n_i^2 (n_i - 1)}$. On remplace ν par $\hat{\nu} = \frac{\left(\sum c_i^2 S_i^2 / n_i\right)^2}{\sum c_i^4 S_i^4 / n_i^2 (n_i - 1)}$ et on suppose que $T \sim t_{\hat{\nu}}$.

6.11 Supposons que M. Martin peut se rendre chez lui le soir par trois routes différentes. Il essaye chacune d'elles 5 fois en prenant note du temps à chaque fois. Voici les résultats, en minutes :

Route 1: 22, 26, 25, 25, 3l
Route 2: 25, 27, 28, 26, 29
Route 3: 26, 29, 33, 30, 33.

Testez à 5% l'hypothèse que les trois routes sont comparables.

6.12 [*Exemple tiré de Snedecor et Cochran*] Une expérience est conduite pour comparer 4 traitements sur la culture de la betterave à sucre. Chaque traitement a été appliqué à 5 champs, et la récolte moyenne par arpent a été notée. Voici les résultats en centaines de livres.

		Engrais appliqué		
	Pas d'engrais	En janvier par labourage	En janvier à la volée	En avril à la volée
Moyenne	38,7	48,7	48,8	45,0

Les calculs ont donné : $\hat{\sigma}^2 = 7{,}443$. Testez chacune des hypothèses suivantes :

a) L'engrais n'a aucun effet.
b) En moyenne, l'engrais appliqué en janvier n'a ni plus ni moins d'effet que lorsqu'il est appliqué en avril.
c) L'engrais appliqué à la volée en janvier a le même effet que lorsqu'il est labouré.

Dans les cas b) et c), déterminez un intervalle de confiance pour l'effet étudié.

6.13 Les données suivantes présentent les pertes en humidité dans le sol des forêts exploitées. Les trois traitements diffèrent selon la quantité de bois laissée sur le sol après la coupe. *Traitement* 1: pas de bois; *Traitement* 2: quantité modérée de bois; *Traitement* 3: grande quantité de bois.

Traitement		
1	2	3
1,52	1,63	2,56
1,38	1,82	3,32
1,29	1,35	2,76
1,48	1,03	2,63
1,63	2,30	2,12
1,45	2,78	

a) Déterminez une table d'analyse de variance, et exprimez la conclusion dans le langage du contexte.

b) On suppose que la présence du bois aura un certain effet, comparé a l'absence de bois, indépendamment de la quantité. Testez donc l'hypothèse que les traitements 2 et 3 ont le même effet l'un que l'autre, et que l'effet moyen de 2 et 3 est le même que celui de 1.

6.14 On demande à 30 sujets, répartis en 3 groupes de 10 chacun, de taper des doigts. Le groupe 1 a bu du café décaféiné avant l'expérience, le groupe 2 a consommé 100 *mg* de caféine et le groupe 3, 200 *mg* de caféine. Le nombre de tapotements par minute a été noté. Voici les données :

***Groupe* 1**:	242	245	244	248	247	248	242	244	246	242
***Groupe* 2**:	248	246	245	247	248	250	247	246	243	244
***Groupe* 3**:	246	248	250	252	248	250	246	248	245	250

a) Déterminez une table d'analyse de variance ;
b) Testez l'hypothèse que boire 100 *mg* a le même effet que boire 200 *mg*.
c) Testez l'hypothèse que $\mu_1 = (\mu_2 + \mu_3)/2$.

6.15 Dans une grande classe de statistique, les élèves proviennent de 4 groupes distincts, définis comme suit:

Groupe 1: Les élèves n'ayant suivi aucun cours de mathématiques au Cégep
Groupe 2: Les élèves ayant suivi des cours de mathématiques et de statistique
Groupe 3: Les élèves ayant suivi des cours de mathématiques mais pas de statistique
Groupe 4: Les élèves ayant terminé un programme de science au Cégep.

Les résultats au cours de statistique sont présentés dans le tableau ci-dessous :

a) Dressez une table d'analyse de variance

b) Testez l'hypothèse que la moyenne du premier groupe est égale celle des trois autres.

Groupe 1		Groupe 2		Groupe 3		Groupe 4	
66	69	63	58	36	48	45	83
35	87	51	47	72	78	91	88
61	82	29	38	32	45	56	67
74	80	53	50	91	47	82	67
47	84	63	68	75	50	131	84
72	54	45	66	63	53	88	83
57	66	33	37	85	21	68	54
48	35	84	62	54		51	81
55	40	83	68			84	57
		59	74			60	56
		83	60			83	44
						75	

6.16 Tester l'hypothèse de linéarité pour les données suivantes :

x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
1,3	2,3	3,3	1,8	4,7	5,4	5,3	2,1
1,3	1,8	3,7	3,7	4,7	3,2	5,7	3,4
2,0	2,8	3,7	1,7	4,7	1,9	6,0	3,2
2,0	1,5	4,0	2,8	5,0	1,8	6,0	3,0
2,7	2,2	4,0	2,8	5,3	3,5	6,3	3,0
3,3	3,8	4,0	2,2	5,3	2,8	6,7	5,9

6.17 Les données suivantes portent sur un échantillon de 34 ménages. Tester l'hypothèse de linéarité dans la relation entre le nombre de personnes et le montant de la facture d'électricité.

Facture d'électricité	*Nombre de personnes*	*Facture d'électricité*	*Nombre de personnes*	*Facture d'électricité*	*Nombre de personnes*
228	2	420	3	756	2
156	1	876	1	636	2
648	2	840	7	708	4
528	1	876	2	960	2
552	3	276	2	1080	3
636	4	1236	3	480	2
444	1	372	2	96	3
144	1	276	1	1272	5
744	2	540	1	1056	2
1104	5	1044	6	156	2
204	1	552	2	396	4
				768	3

6.18 Lors d'une étude sur le « concept de soi » des adolescents nigérians, un chercheur [Jegede, R. Olukayode, *The Journal of Psychology* 110, 249-261 (1982)] a administré le test *Piers-Harris Self-Concept Scale* à 1300 élèves de niveau secondaire à Ibadan.

a) Le score moyen a été de 58,19 avec un écart-type de 10,06. Dans une étude faite antérieurement auprès de 1183 adolescents américains on avait trouvé une moyenne de 51,84 avec un écart-type de 13,87. La différence entre le Nigérians et les Américains est-elle significative ? Supposez que la variance σ^2 des deux groupes est la même et estimez σ^2.

b) L'échantillon de Jegede était composé de 552 filles et 828 garçons. Les filles avaient une moyenne de 56,82 avec un écart-type de 9,96 ; et les garçons une moyenne de 59,11 avec un écart-type de 10,01. La différence entre les garçons et les filles est-elle significative ? Supposez que la variance σ^2 des deux groupes est la même et estimez σ^2.

c) Maintenant traitez simultanément les trois groupes : filles nigérianes, garçons nigérians, et Américains. Testez l'hypothèse que les moyennes de trois groupes (μ_1, μ_2 et μ_3, respectivement) sont égales. Supposez que la variance σ^2 des trois groupes est la même et estimez σ^2.

d) Dans le cadre du modèle en c) testez l'hypothèse que les filles et les garçons nigérians ont la même moyenne.

e) Dans le cadre du modèle en c) testez l'hypothèse que la moyenne des Américains est égale à la moyenne des Nigérians en supposant qu'il y a autant de filles que de garçons au Nigéria (en d'autres termes, vous devez tester l'hypothèse que $\mu_3 = (\mu_1+\mu_2)/2$.)

6.19 Dans une étude sur la sexualité des jeunes en Australie, un chercheur [Hong, Sung-Mook, The *Journal of Psychology* 115, 17-22 (1983)] a fait remplir un questionnaire à 560 étudiants d'université. Le questionnaire rempli permet de calculer un score qui indique dans quelle mesure l'attitude du répondant est permissive (un score élevé dénote une attitude permissive). L'objectif est de déterminer si le niveau de pratique religieuse affecte l'attitude concernant les comportements sexuels. Voici les moyennes, les écarts-types et les effectifs de trois sous groupes.

Vont à l'église	$\bar{y}$	S	n	μ
Régulièrement	3,31	1,54	128	μ_1
De temps en temps	4,73	1,10	230	μ_2
Jamais	5,24	0,79	202	μ_3

a) Dresser une table d'analyse de variance. Expliquez votre conclusion. Estimer la variance σ^2.

b) Tester, dans le cadre du modèle en a) l'hypothèse que ceux qui ne vont jamais à l'église sont différents de ceux qui y vont, régulièrement ou de temps en temps, c'est-à-dire, tester l'hypothèse $\mu_3 = (\mu_1+\mu_2)/2$.

c) Tester l'hypothèse qu'il n'y a pas de différence entre ceux qui vont régulièrement à l'église et ceux qui y vont de temps en temps.

d) Tester l'hypothèse qu'il n'y a pas de différence entre ceux qui vont à l'église de temps en temps et ceux qui n'y vont jamais.

e) Amalgamez les deux derniers groupes et testez l'hypothèse que ceux qui vont régulièrement à l'Église ont la même moyenne que les autres. Supposez une même variance σ^2 et estimez σ^2.

6.20 Les données suivantes donnent le prix de vente (Prix, en milliers de dollars) et le nombre de salles de bains (Bains) pour un ensemble de maisons.

a) Déterminer si la relation entre le prix et le nombre de salles de bains est linéaire.
b) Tester l'hypothèse d'homoscédasticité

Prix	*Bains*	*Prix*	*Bains*	*Prix*	*Bains*	*Prix*	*Bains*
179	2,5	110	2,5	136	1	133	2,5
126,5	1,5	154	2,5	158	2,5	180	2,5
134,5	1	151,5	2,5	211,9	2,5	139,5	1,5
125	1	141	1,5	165	2,5	210	2,5
142	1,5	119	1	183	2,5	126,5	1
164	2,5	108,5	2,5	85	1	195	2,5
146	2,5	126,5	2,5	126,5	1	97	1
129	1	130	1,5	162	2,5	117	1
141,9	3	140	2,5	169	3	150	2,5
135	2,5	123,5	2,5	175	3	180,5	2,5
118,5	1	153,5	3	150	2,5	160	2,5
160	2,5	194,9	2,5	115	1	181,5	2,5
89,9	1	165	2,5	126,5	1,5	124	1,5
169,9	3	179,9	2,5	215	2,5	125,9	1,5
127,5	1	194,5	2,5	190	2,5	165	3
162,5	2,5	127,5	2,5	190	2,5	122	1,5
152	1,5	170	3	113,5	1,5	132	2,5
122,5	1	160	2,5	116,3	1,5	145,9	1,5
220	3	135	2,5	190	1,5	156	2,5
141	1	117	1	145	2,5	136	2,5
80,5	1	235	3	135,5	1,5	142	1
152	1,5	163,5	2,5	190	2,5	140	2,5
231,8	3	78	1	98	1	144,9	1
180	2,5	187	2,5	137,9	1,5	133	2,5
185	2,5	133	1	108	1	196,8	2,5
135	1,5	125	1	120,5	2,5	121,9	1
203	3	116	1	128,5	2,5	126	2,5
141	2,5	135	2,5	142,5	1,5	164,9	2,5
159	1,5	194,5	2,5	72	1	172	2,5
182	2,5	99,5	1	124,9	1	100	1
208	2,5	152,5	2,5	134	1,5	129,9	1
96	1	141,9	2,5	205,4	2,5	110	1
156	3	139,9	1,5	217	3	131	1
185,5	1,5	117,5	1	94	1	107	1
144,9	2,5	150	2,5	189,9	2,5	165,9	2,5
155	1,5	177	1,5	168,5	2,5		

6.21 Considérez une analyse de variance à deux facteurs, chaque facteur à deux niveaux. Les moyennes sont

$$\mu_{11} \qquad \mu_{12} \qquad\qquad \mu_{21} \qquad \mu_{22}$$

Toute fonction linéaire non singulière

$$T: (\mu_{11}\ ;\ \mu_{12}\ ;\ \mu_{21}\ ;\ \mu_{22}) \to (\mu\ ;\ \alpha\ ;\ \beta\ ;\gamma)$$

qui transforme ces quatre moyennes en quatre autres paramètres μ, α, β,γ est une paramétrisation également valable. Un choix commode de nouveaux paramètres (μ, α, β,γ) est celui pour lequel les hypothèses usuelles H_A, H_B, et H_{AB} se traduisent par $\alpha = 0$, $\beta = 0$, et $\gamma = 0$, respectivement.

a) La paramétrisation qui donne les moyennes suivantes satisfait cette condition :

$\mu - \alpha - \beta + \gamma$	$\mu - \alpha + \beta - \gamma$
$\mu + \alpha - \beta - \gamma$	$\mu + \alpha + \beta + \gamma$

Montrez qu'en effet H_A *est vraie* $\Leftrightarrow \alpha = 0$, H_B *est vraie* $\Leftrightarrow \beta = 0$, et H_{AB} *est vraie* $\Leftrightarrow \gamma = 0$ [*Suggestion*: Commencez par exprimer les moyennes des marges en fonction des paramètres μ, α, β et γ].

Si $T^{-1}(\mu\ ;\ \alpha\ ;\ \beta\ ;\ \gamma) = (\mu\text{-}a\text{-}\beta+\gamma\ ;\ \mu-\alpha+\beta-\gamma\ ;\ \mu+\alpha-\beta-\gamma\ ;\ \mu+\alpha+\beta+\gamma) = (\mu_{11}\ ;\ \mu_{12}\ ;\ \mu_{21}\ ;\ \mu_{22})$, montrez que la fonction T est orthogonale.

b) Considérez la paramétrisation suivante :

μ	$\mu + \beta$
$\mu + \alpha$	$\mu + \alpha + \beta + \gamma$

Montrez H_{AB} se traduit par $\gamma = 0$, mais que H_A ne se traduit pas par $\alpha = 0$, et que H_B ne se traduit pas par $\beta = 0$. Montrez néanmoins que si $\gamma = 0$, alors H_A est vraie $\Leftrightarrow \alpha = 0$ et H_B est vraie $\Leftrightarrow \beta = 0$. [Il suffit d'exprimer les hypothèses usuelles H_A, H_B, et H_{AB} en termes des paramètres].

6.22 Soit les moyennes échantillonnales suivantes dans une analyse de variance à deux facteurs (chacun à deux niveaux) et r observations par cellule :

$\bar{y}_{11}$	$\bar{y}_{12}$	$\bar{y}_{1\cdot}$
$\bar{y}_{21}$	$\bar{y}_{22}$	$\bar{y}_{2\cdot}$
$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	$\bar{y}$

Montrez que les vecteurs $\begin{bmatrix}\bar{y}_{1\cdot}\\ \bar{y}_{2\cdot}\end{bmatrix}$ et $\begin{bmatrix}\bar{y}_{\cdot 1}\\ \bar{y}_{\cdot 2}\end{bmatrix}$ ne sont pas indépendants, mais que la variable $\bar{y}_{1\cdot} - \bar{y}_{2\cdot}$ est indépendante du vecteur $\begin{bmatrix}\bar{y}_{\cdot 1}\\ \bar{y}_{\cdot 2}\end{bmatrix}$. Déduisez que SCA et SCB sont indépendantes.[*Suggestion*: Exprimez toutes les variables définies ici comme combinaisons linéaires du vecteur $\bar{\boldsymbol{y}} = (\bar{y}_{11}, \bar{y}_{12}, \bar{y}_{21}, \bar{y}_{22})'$, après en avoir déterminé la distribution].

6.23 Soit les moyennes échantillonnales suivantes dans une analyse de variance à deux facteurs où les effectifs n_{ij} des classes ne sont pas égaux :

$\bar{y}_{11}$	$\bar{y}_{12}$	$\bar{y}_{1\cdot}$
$\bar{y}_{21}$	$\bar{y}_{22}$	$\bar{y}_{2\cdot}$
$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	$\bar{y}$

Les sommes de carrés SCA, SCB et SCAB devraient dépendre, respectivement, de $\bar{y}_{1\cdot} - \bar{y}_{2\cdot}$, de $\bar{y}_{\cdot 1} - \bar{y}_{\cdot 2}$, et de $\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} + \bar{y}_{22}$, où les moyennes marginales sont des moyennes non pondérées des moyennes des cellules. Par exemple, $\bar{y}_{1\cdot} = (½)(\bar{y}_{11} + \bar{y}_{12})$. Montrez que ces sommes de carrés ne sont pas indépendantes. Dites dans quelles conditions

a) $\bar{y}_{1\cdot} - \bar{y}_{2\cdot}$ et $\bar{y}_{\cdot 1} - \bar{y}_{\cdot 2}$ sont indépendantes.
b) $\bar{y}_{1\cdot} - \bar{y}_{2\cdot}$ et $\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} + \bar{y}_{22}$ sont indépendantes.
c) $\bar{y}_{\cdot 1} - \bar{y}_{\cdot 2}$ et $\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} + \bar{y}_{22}$ sont indépendantes.

6.24 [Battacharryya, Gouri K. , Johnson, Richard A. (1977) *Statistical concepts and methods*, Wiley, New York, *p*.502].On étudie l'effet de lésions sur le lobe occipital du cerveau, normalement associé au sens de la vue. L'expérience est effectuée sur 24 singes rhésus, à 12 desquels on inflige une lésion sur le lobe occipital, les 12 autres servant de groupe témoin. On note les scores obtenus par tous les singes à un test d'acuité visuelle. Le deuxième facteur est l'âge du singe au moment du test, qui, pour le groupe expérimental, suit immédiatement l'opération. Voici les données :

	Groupe	
Âge	Contrôle	Expérimental
100	93, 90, 91, 88	86, 71, 68, 70
300	79, 94, 50, 70	76, 91, 51, 96
730	89, 91, 86, 84	60, 72, 93, 68

a) Faites une analyse de variance à l'aide d'une procédure d'analyse de variance à deux facteurs.
b) En utilisant les sommes de carrés calculées en a), tester l'hypothèse que les lésions n'ont pas d'effet sur l'acuité visuelle à l'âge de 100 jours ; non plus à 300 jours ; ni à 730 jours (un seul test des trois hypothèses).
c) Considérez un modèle d'analyse de variance à un facteur (6 niveaux) et testez les hypothèses usuelles H_A, H_B et H_{AB}. Vérifiez que les statistiques et conclusions sont les mêmes qu'en a).

6.25 On veut comparer l'effet sur l'apprentissage du français de trois types d'école : écoles françaises (type I), écoles bilingues (type II), et écoles d'immersion française pour anglophones (type III). Il y a deux écoles de chaque type dans la ville. On fait passer un test de français à 9 finissants choisis au hasard dans chaque école. Voici les résultats :

Type I		Type II		Type III	
École 1	École 2	École 1	École 2	École 1	École 2
58	63	82	98	43	57
76	78	77	65	56	49
56	64	99	54	83	59
65	53	33	91	46	51
74	40	58	88	61	58
67	53	80	52	69	89
76	64	58	86	67	42
83	38	84	78	58	46
75	43	54	74	92	52

Soit A le facteur « type d'école » et B le facteur « école ». Contrairement aux exemples et exercices traités dans ce chapitre, les deux facteurs ne sont pas « croisés », puisqu'il n'y a aucune correspondance entre les trois « école 1 » ou les trois « école 2 ». Néanmoins, définissez les sommes de carrés comme auparavant.

a) Expliquez pourquoi la statistique MCB/MCR ne teste pas une hypothèse raisonnable dans ce contexte.
b) Montrez (algébriquement, pas avec les chiffres actuels) que SCB + SCAB = $r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij}-\bar{y}_{i}.)^2$, avec r =9, a = 3, b = 2, et que la statistique $\dfrac{[\text{SCB+SCAB}]/[a(b-1)]}{\text{MCR}}$ teste une hypothèse raisonnable. Exprimez cette hypothèse en termes des paramètres, et testez-la.

6.26 Au numéros 6.25-b), on teste une hypothèse $\mu_{i1} = \mu_{i2} = \ldots = \mu_{ib}$ pour $i = 1, \ldots, a$ — in test basé sur la somme SCA + SCAB. Le but de cet exercice est de montrer que cette procédure n'exige pas que les effectifs des cellules soient égaux. On suppose donc que l'effectif de la cellule (i,j) est $n_{ij} \geq 1$. Soit, par définition, SCB + SCAB = $\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}(\bar{y}_{ij}-\bar{y}_{i}.)^2$, SCA = $\sum_{i=1}^{a}n_{i}.(\bar{y}_{i}.-\bar{y})^2$, et SCE = $\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}(\bar{y}_{ij}-\bar{y})^2$, où $\bar{y}_{i}. = \dfrac{1}{n_{i}.}\sum_{j=1}^{b}n_{ij}\bar{y}_{ij}$.

a) Montrez que SCA + SCB + SCAB = SCE.
b) Montrez que (SCB + SCAB)/σ^2 ~ $\chi^2_{a(b-1)}(\lambda)$, où $\lambda = \sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}(\mu_{ij}-\bar{\mu}_{i}.)^2/\sigma^2$, $\bar{\mu}_{i}. = \dfrac{1}{n_{i}.}\sum_{j=1}^{b}n_{ij}\mu_{ij}$.

c) Montrez que SCB + SCAB et SCA sont indépendantes. [*Suggestion* : Il suffit de montrer que pour chaque i, $\bar{y}_{i}.$ est indépendante de $\sum_{j=1}^{b} n_{ij}(\bar{y}_{ij} - \bar{y}_{i}.)^2$].

6.27 Le tableau suivant classe des données sur la récolte d'une céréale, classée selon deux facteurs : le facteur A est la variété et le facteur B l'espacement entre les plantes.

Variété	4 pouces	8 pouces	12 pouces	Moyenne
I	56, 45, 43, 46	60, 50, 45, 48	66, 57, 50, 50	616/12
II	65, 61, 60, 63	60, 58, 56, 60	53, 53, 48, 55	692/12
III	60, 61, 50, 53	62, 68, 67, 60	73, 77, 77, 65	773/12
Moyennes	663/12	694/12	724/12	2081/36

a) Y a-t-il une interaction entre les variétés et les espacements ?
b) Y a-t-il un effet d'espacement ?
c) Y a-t-il des différences entre les variétés ?
d) Tester l'hypothèse que, quelle que soit la variété, l'espacement n'a pas d'effet.

6.28 Considérons un modèle d'analyse de variance à deux facteurs, avec $r > 1$ observations par cellule. L'estimation de σ^2 et les statistiques F_A *et* F_B pour tester H_A et H_B sont bien établies. Supposons, cependant, que nous ajoutions, comme supposition supplémentaire du modèle, l'hypothèse que les interactions sont nulles. Ceci permet une meilleure estimation de σ^2 et modifie les statistiques de test.

a) Proposez un meilleur estimateur de σ^2.
b) Proposez de nouveaux tests pour H_A et H_B.
c) Appliquez vos résultats en a) et b) à l'exemple du numéro précédent.

6.29 Dans le modèle d'analyse de variance à deux facteurs, avec r observations par cellule,

a) calculez Cov[$(\bar{y}_{i}. - \bar{y})(\bar{y}._{j} - \bar{y})$] et déduisez de SCA et SCB sont indépendantes;
b) démontrez de la même façon que SCA et SCAB sont indépendantes.

6.30 On compare 5 variétés de maïs, chacune plantée dans 4 terrains (des "blocs"). Les blocs sont divisés en 5 parcelles chacun, une parcelle étant consacrée à chaque variété. Une mesure de la quantité récoltée est prise dans chaque parcelle. Voici les données :

	Bloc				
Variété	1	2	3	4	Moyenne
A	21	25	28	35	27,25
B	23	29	33	34	29,75
C	15	21	22	28	21,50
D	16	19	25	25	21,25
E	12	18	21	25	19,00
Moyenne	17,4	22,4	25,8	29,4	23,75

a) Testez à 5% l'hypothèse qu'il n'y a pas de différence entre les variétés, en supposant qu'il n'y a pas d'interaction entre les variétés et les blocs.
b) Testez l'hypothèse qu'en moyenne les variétés A et B produisent autant que les trois autres.

6.31 Utilisez le modèle de régression pour faire une analyse de variance à deux facteurs sans interactions. Les données suivantes représentent le nombre de graines de soya parmi 100 *qui n'ont pas germé*, selon le traitement et le bloc.

	Bloc				
Traitement	1	2	3	4	5
Aucun	8	10	12	13	11
Arasun	2	6	7	11	5
Spergon	4	10	9	8	10
Semeson Jr	3	5	9	10	6
Permate	9	7	5	5	3

a) Montrez que la paramétrisation suivante tient compte de l'hypothèse d'additivité (non interaction):

	Bloc				
Traitement	1	2	3	4	5
Aucun	μ	$\mu+\beta_2$	$\mu+\beta_3$	$\mu+\beta_4$	$\mu+\beta_5$
Arasun	$\mu+\alpha_2$	$\mu+\alpha_2+\beta_2$	$\mu+\alpha_2+\beta_3$	$\mu+\alpha_2+\beta_4$	$\mu+\alpha_2+\beta_5$
Spergon	$\mu+\alpha_3$	$\mu+\alpha_3+\beta_2$	$\mu+\alpha_3+\beta_3$	$\mu+\alpha_3+\beta_4$	$\mu+\alpha_3+\beta_5$
Semeson Jr	$\mu+\alpha_4$	$\mu+\alpha_4+\beta_2$	$\mu+\alpha_4+\beta_3$	$\mu+\alpha_4+\beta_4$	$\mu+\alpha_4+\beta_5$
Permate	$\mu+\alpha_5$	$\mu+\alpha_5+\beta_2$	$\mu+\alpha_5+\beta_3$	$\mu+\alpha_5+\beta_4$	$\mu+\alpha_5+\beta_5$

b) Exprimez les hypothèses usuelles en fonction de ces nouveaux paramètres.

6.32 Déterminez les espérance des moyennes de carrés SCA, SCB et SCAB, et montrez qu'elles sont indépendantes les unes de autres. [Il suffira de calculer les covariances entre les fonctions linéaires dont dépendent les moyennes des carrés.]

6.33 Considérons une analyse de variance à deux facteurs, chacun à 2 niveaux, avec une observation par cellule et pas d'interaction. On dénote les 4 observations par y_{11}, y_{12}, y_{21}, y_{22} et les 4 moyennes correspondantes par μ_{11}, μ_{12}, μ_{21}, μ_{22}.

a) Montrer que l'absence d'interaction entraîne que $\mu_{22} = \mu_{12} + \mu_{21} - \mu_{11}$.

b) Les estimateurs des moindres carrés $\hat{\mu}_{ij}$ sont les valeurs des μ_{ij} qui minimisent la somme $\sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij}-\mu_{ij})^2$. Montrez que $\hat{\mu}_{11}$ = (¼)($3y_{11}+y_{12}+y_{21}-y_{22}$); $\hat{\mu}_{12}$ = (¼)($y_{11}+3y_{12}-y_{21}+y_{22}$); $\hat{\mu}_{21}$ = (¼)($y_{11}-y_{12}+3y_{21}-y_{22}$); et, bien sûr, $\hat{\mu}_{22}$ = (¼)($\hat{\mu}_{12} + \hat{\mu}_{21} - \hat{\mu}_{11}$).

6.34 Montrez que le test t pour comparer deux moyennes est équivalent à une analyse de variance. [Utilisez le fait que si T est de loi t_ν, alors T^2 est de loi $F_{1;\nu}$.]

6.35 [*Exemple tiré de Snedecor et Cochran*] Une expérience est faite pour comparer 4 traitements sur la culture de la betterave à sucre. Chaque traitement a été appliqué à 5 champs, et la récolte moyenne par arpent a été notée. Voici les résultats en centaines de livres.

		Engrais appliqué		
	Pas d'engrais	En janvier par labourage	En janvier à la volée	En avril à la volée
Moyenne	38,7	48,7	48,8	45,0

Les calculs ont donné : $\hat{\sigma}^2$ = 7,443. Testez chacune des hypothèses suivantes :

a) L'engrais n'a aucun effet.

b) En moyenne, l'engrais appliqué en janvier n'a ni plus ni moins d'effet que lorsqu'il est appliqué en avril.

c) L'engrais appliqué à la volée en janvier a le même effet que lorsqu'il est labouré.

Dans les cas b) et c), déterminez un intervalle de confiance pour l'effet étudié.

6.36 Dans une grande classe de statistique, les élèves proviennent de 4 groupes distincts, définis comme suit :

Groupe 1: Les élèves n'ayant suivi aucun cours de mathématiques au Cégep
Groupe 2: Les élèves ayant suivi des cours de mathématiques et de statistique

Groupe 3: Les élèves ayant suivi des cours de mathématiques mais pas de statistique
Groupe 4: Les élèves ayant terminé un programme de science au Cégep.

Les résultats au cours de statistique sont présentés dans le tableau ci-dessous :

a) Dressez une table d'analyse de variance
b) Testez l'hypothèse que la moyenne du premier groupe est différente de celle des trois autres.

Groupe 1		Groupe 2		Groupe 3		Groupe 4	
66	69	63	58	36	48	45	83
35	87	51	47	72	78	91	88
61	82	29	38	32	45	56	67
74	80	53	50	91	47	82	67
47	84	63	68	75	50	131	84
72	54	45	66	63	53	88	83
57	66	33	37	85	21	68	54
48	35	84	62	54		51	81
55	40	83	68			84	57
		59	74			60	56
		83	60			83	44
						75	

6.37 Dans une étude sur la relation entre certains traits de personnalité et des facteurs astrologiques, des chercheurs [Sakofske, Kelly et McKerracher, *The Journal of Psychology* 110, 275-80, 1982] ont fait compléter un questionnaire (le *Eysenck Personality Questionnaire*] à 241 étudiants néo-zélandais. L'hypothèse (avancée antérieurement par des astrologues) que ces chercheurs se proposer de vérifier est que les personnes nées sous un signe positif (Bélier, Balance, Gémeaux, Lion, Verseau, Sagittaire) sont moins introverties que les personnes nées sous un signe négatif (Cancer, Capricorne, Poisson, Scorpion, Taureau, Vierge). Sur l'échelle introversion-extraversion du test, les extravertis ont un score élevé. L'échantillon était composé d'hommes et de femmes. Voici les moyennes, les écarts-types et le tailles des 4 groupes ainsi que la désignation des moyennes des populations :

Signe du zodiac	Hommes				Femmes			
	$\bar{y}$	S	n	μ	$\bar{y}$	S	n	μ
Positif	13,50	4,38	38	μ_1	13,17	4,57	79	μ_3
Négatif	15,52	4,21	38	μ_2	13,73	4,39	86	μ_4

Dans ce qui suit, considérer qu'il s'agit de quatre groupes issus de quatre populations de même variance σ^2.

a) Tester l'hypothèse que les hommes et les femmes ont la même moyenne.
b) Tester l'hypothèse que les personnes nées sous un signe positif ont la même moyenne que ceux nés sous un signe négatif
c) Tester l'hypothèse que la différence entre les positifs et les négatifs est la même chez les hommes et les femmes.
d) Tester l'hypothèse qu'il n'y a pas de différence entre les positifs et les négatifs chez les hommes.
e) Tester l'hypothèse qu'il n'y a pas de différence entre les positifs et les négatifs chez les hommes *et* qu'il n'y a pas de différence entre positifs et négatifs chez les femmes.

6.38 Lors d'un projet d'étude des problèmes de racisme dans une force policière, on prélève un échantillon de 32 policiers, dont 16 ont 11 ans de scolarité ou moins et 11 on plus de 11 ans de scolarité. Chacun des deux groupes est divisé en deux sous-groupes de 8. L'un des deux sous-groupes suit un cours de sensibilisation aux cultures ethniques, l'autre pas. Le tableau suivant donne les résultats à un test d'hostilité aux groupes ethniques.

		Scolarité	
		≤ 11 ans	> 11 ans
Cours de sensibilisation	Suivi	60 58 56 54 52 50 48 46	36 36 33 32 30 29 26 26
	Pas suivi	50 48 46 44 42 40 38 36	30 28 26 24 22 20 28 26

a) Dresser une table d'analyse de variance. Estimer la variance σ^2

b) Tester l'hypothèse que la moyenne de ceux qui ont suivi le cours est la même que celle de ceux qui ne l'on pas suivi.

c) Testez l'hypothèse que le cours a le même effet chez ceux de 11 ans de scolarité et ceux de plus de 11 ans de scolarité (l'« effet » du cours est mesuré par la différence des deux moyennes).

d) Supposez qu'on n'a pas tenu compte des niveaux de scolarité. Vous avez donc deux groupes. Tester l'hypothèse que la moyenne de ceux qui ont suivi le cours est la même que celle de ceux qui ne l'on pas suivi. Estimez σ^2.

6.39 Deux groupes de 11 enfants de troisième année du cycle primaire ont complété le test psychologique IAR (*Intelligence Achievement Responsibility*) avant et après une période de quatre mois et demi d'expérimentation avec l'un ou l'autre de deux langages informatiques : LOGO et Delta Drawing. Contrairement au LOGO, le langage Delta Drawing n'attache pas une grande importance à la décomposition d'un problème complexe ou à l'apprentissage par la correction des erreurs. Le test IAR mesure la propension du sujet à se sentir maître de ses apprentissages et de son succès intellectuel. Les chercheurs ont voulu montrer que l'exercice du langage LOGO augmente cette propension. Voici les résultats obtenus :

LOGO			Delta		
	Score			Score	
Sexe	Avant	Après	Sexe	Avant	Après
F	16	29	F	15	21
F	20	24	M	18	22
M	21	23	F	21	21
M	22	21	F	21	19
M	22	26	F	22	20
F	23	30	F	22	20
F	24	26	F	23	23
F	24	23	F	23	30
F	25	32	M	26	21
M	27	34	M	27	25
M	28	29	M	30	27

a) Soit Y la différence Après-Avant. Dresser une table d'analyse de variance considérant que l'échantillon est composé de quatre groupes : Filles*LOGO (μ_1) ; Filles*Delta (μ_2) ; Garçons*LOGO (μ_3) ; et Garçons*Delta (μ_3).

b) Tester l'hypothèse qu'il n'y a pas de différence entre LOGO et Delta.

c) Tester l'hypothèse qu'il n'y a pas de différence entre filles et garçons.

d) Tester l'hypothèse que la différence entre les deux sexes est la même pour LOGO que pour Delta.

e) Tester l'hypothèse que la différence entre LOGO et Delta est la même pour les filles que pour les garçons.

f) Supposons que la distinction entre filles et garçons n'a simplement pas été retenue pour l'expérience. Dans ce modèle à deux groupes, tester l'hypothèse qu'il n'y a pas de différence entre LOGO et Delta. Estimer la variance σ^2.

6.40 Démontrez la décomposition SCT = SCE + SCR, soit

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2=\sum_{i=1}^{k}n_i(\bar{y}_i-\bar{y})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$$

6.41 Démontrez la formule de calcul

$$\sum_{i=1}^{k}n_i(\bar{y}_{i.}-\bar{y})^2=\sum_{i=1}^{k}n_i\bar{y}_i^2-n\bar{y}^2$$

6.42 Considérons le cas où $n_1 = n_2 = \ldots = n_k = m$. Montrez que SCE/$\sigma^2$: χ^2_{k-1} si H_o est vraie. Sans supposer que H_0 est vraie, déterminez l'espérance de SCE. Montrez comment votre réponse justifie une région critique de la forme $F > \mathcal{F}_{k-1;k(m-1);\alpha}$.

6.43 Lors d'une étude sur le « concept de soi » des adolescents nigérians, un chercheur [Jegede, R. Olukayode, *The Journal of Psychology* 110, 249-261 (1982)] a administré le test *Piers-Harris Self-Concept Scale* à 1300 élèves de niveau secondaire à Ibadan.

a) Le score moyen a été de 58,19 avec un écart-type de 10,06. Dans une étude faite antérieurement auprès de 1183 adolescents américains on avait trouvé une moyenne de 51,84 avec un écart-type de 13,87. La différence entre le Nigérians et les Américains est-elle significative ? Supposez que la variance σ^2 des deux groupes est la même et estimez σ^2.

b) L'échantillon de Jegede était composé de 552 filles et 828 garçons. Les filles avaient une moyenne de 56,82 avec un écart-type de 9,96 ; et les garçons une moyenne de 59,11 avec un écart-type de 10,01. La différence entre les garçons et les filles est-elle significative ? Supposez que la variance σ^2 des deux groupes est la même et estimez σ^2.

c) Maintenant traitez simultanément les trois groupes : filles nigérianes, garçons nigérians, et Américains. Testez l'hypothèse que les moyennes de trois groupes (μ_1, μ_2 et μ_3, respectivement) sont égales. Supposez que la variance σ^2 des trois groupes est la même et estimez σ^2.

d) Dans le cadre du modèle en c) testez l'hypothèse que les filles et les garçons nigérians ont la même moyenne.

e) Dans le cadre du modèle en c) testez l'hypothèse que la moyenne des Américains est égale à la moyenne des Nigérians en supposant qu'il y a autant de filles que de garçons au Nigéria (en d'autres termes, vous devez tester l'hypothèse que $\mu_3 = (\mu_1+\mu_2)/2$.)

6.44 Les données suivantes résultent d'une expérience visant à déterminer quels facteurs influencent la progression de la pourriture dans les pommes de terre. La variable mesurée est le diamètre maximal du tissu pourri cinq jours après une injection de bactéries. Les facteurs considérés sont la température et la quantité d'oxygène. Voici les données

		Pourcentage d'oxygène	
		2%	10%
Température	0°	13, 11, 3, 20, 10, 13, 10, 11, 5, 15	15, 2, 7, 6, 5, 10, 8, 6, 10, 16
	10°	26, 19, 24, 26, 27, 17, 21, 17, 18, 17	20, 24, 8, 18, 11, 18, 15, 9, 8, 18

a) Effectuer une analyse de variance en utilisant une procédure d'analyse de variance à deux facteurs.

b) Effectuer une analyse de variance en utilisant une procédure d'analyse de variance à 1 facteur (4 niveaux) et testez les hypothèses usuelles HA, HB et HAB.