MAT7381 Solution # 6.37

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.

Le vecteur $\boldsymbol{\mu}$ des moyennes est estimé par le vecteur $\bar{\mathbf{y}}$ des moyennes échantillonnales. Si on admet les hypothèses habituelles d'un modèle linéaire (normalité et homoscédasticité), $\bar{y}$ est normal de moyenne $\boldsymbol{\mu}$ et de matrice de convariance $\sigma^2\mathbf{V}$, où $\mathbf{V}$ est une matrice diagonale dont la i^e composante est $1/n_i$. La variance σ^2 est estimée par $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^4 (n_i - 1)S_i^2}{n-4}$, $n = \Sigma n_i$. Voici donc les estimations :

$\bar{y}$ =

```
    hp    hn    fp    fn
 13.50 15.52 13.17 13.73
```

S_i

```
   hp   hn   fp   fn
 4.38 4.21 4.57 4.39
```

n_i

```
 hp hn fp fn
 38 38 79 86
```

$\hat{\sigma}$ = 4,421259

$\hat{\sigma}^2\mathbf{V}$ =

```
0.5048526 0.0000000 0.0000000 0.0000000
0.0000000 0.4664237 0.0000000 0.0000000
0.0000000 0.0000000 0.2643658 0.0000000
0.0000000 0.0000000 0.0000000 0.2240942
```

Les hypothèses énoncées en a), b), c) et d) sont toutes de la forme $\boldsymbol{\ell}'\boldsymbol{\mu} = 0$, où $\boldsymbol{\ell}' = [\ell_1\ ;\ \ell_2\ ;\ \ell_3\ ;\ \ell_4]$ est un vecteur de valeurs fixes. La statistique pour tester une telle hypothèse est $T = \dfrac{\sum_{i=1}^k \ell_i \bar{y}_i}{\hat{\sigma}\sqrt{\ell_i^2 / n_i}}$. Sous l'hypothèse, T est de loi de Student à n-4 degrés de liberté.

a) Tester l'hypothèse que les hommes et les femmes ont la même moyenne.

Ici $\boldsymbol{\ell}' = [1\ \ 1\ \ -1\ \ -1]$. $T = \dfrac{2{,}12}{1{,}226194} = 1{,}73$. La valeur p est 0,08512. On ne peut pas rejeter l'hypothèse à 5 %.

b) Tester l'hypothèse que les personnes nées sous un signe positif ont la même moyenne que ceux nés sous un signe négatif

Ici $\boldsymbol{\ell}' = [1\ \ -1\ \ 1\ \ -1]$. $T = \dfrac{2{,}58}{1{,}226194} = -2{,}1$. La valeur p est 0,03642. On peut rejeter l'hypothèse à 5 %.

c) Tester l'hypothèse que la différence entre les positifs et les négatifs est la même chez les hommes et les femmes.

Ici ℓ' = [1 -1 -1 1]. $T = \dfrac{1,46}{1,226194}$ = -1,19. La valeur p est 0,23497. On ne peut pas rejeter l'hypothèse à 5 %.

d) Tester l'hypothèse qu'il n'y a pas de différence entre les positifs et les négatifs chez les hommes.

Ici ℓ' = [1 -1 0 0]. $T = \dfrac{2,02}{1,014306}$ = -1,99. La valeur p est 0,0,04757. On ne peut pas rejeter l'hypothèse à 5 %.

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.

On teste l'hypothèse **L'μ** = **0**, où $\mathbf{L} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}$. La statistique de test est $F = \hat{\boldsymbol{\mu}}'\mathbf{L}[\mathbf{L}'\mathbf{V}\mathbf{L}]^{-1}\mathbf{L}'\hat{\boldsymbol{\mu}}/2$,

une statistique qui suit, sous l'hypothèse, une loi F à 2 et n-4 degrés de liberté.

$\mathbf{L}'\hat{\boldsymbol{\mu}}$ = -2.02
-0.56

L'VL= 1.028817 0.000000
0.000000 0.474734

F = 2.313344. Valeur p = 0.1011599. On ne peut pas rejeter l'hypothèse à 5 %.

Remarque Ce test peut aussi être basé sur la statistique $F = \dfrac{(\mathrm{SCR_o} - \mathrm{SCR})/2}{\mathrm{MCR}}$ où SCR est la somme des carrés résiduelle dans le modèle, MCR est la moyenne des carrés résiduelle (donc $\hat{\sigma}^2$) et $\mathrm{SCR_o}$ est la somme des carrés résiduelle dans le modèle restreint par l'hypothèse nulle. Selon H_o, les deux groupes d'hommes appartiennent en fait à un seul groupe de même moyenne μ_1; de même pour les femmes, dont la moyenne est μ_2. Il est possible de reconstituer les moyennes et les écarts-types pour hommes combinés et pour femmes combinées. Les voici :

	n	$\bar{y}$	s
Hommes	76	14,51	4,386558
Femmes	165	13,461	4,472179

Nous devons donc déterminer la somme des carrés résiduelle dans un modèle d'analyse de variance à un facteur à deux niveaux. $\mathrm{SCR_o} = \dfrac{(76-1)(4,386558)^2 + (165-1)(4,472179)^2}{76+165-2}$ = 4723,06. Le nombre de degrés de liberté de la différence $\mathrm{SCR_o}$-SCR est la différence entre le nombre de paramètres dans le modèle initial (4 moyennes) et le nombre de paramètres dans le modèle restreint (2 moyennes). D'où le numérateur ($\mathrm{SCR_o}$-SCR)/2. Sous H_o, la statistique F suit une loi F à 2 et n-k = 241-4 = 237 degrés de liberté. On a SCR =4632,765, MCR = 19,54753 et, finalement, F =2,313344. La valeur p est 0,1011599.