ST4241: Design and Analysis of Clinical Trials

2009/2010: Semester I

Tutorial 7

1. In a stratified study, the patients are stratified into three strata and three treatments are randomly assigned to patients in each stratum such that each treatment was administered to the same number of patients. The data is given below:

	Treatment		
Stratum	1	2	3
1	130	34	20
	155	40	70
	74	80	82
	180	75	58
2	150	126	25
	188	122	70
	159	106	58
	126	115	45
3	138	174	96
	110	120	104
	168	150	82
	160	139	60

The following R-codes can be used for the computation of the desired results:

```
x1 = c(130,155,74,180,34,40,80,75,20,70,82,58)
x2 = c(150,188,159,126,126,122,106,115,25,70,58,45)
x3 = c(138,110,168,160,174,120,150,139,96,104,82,60)
x=c(x1,x2,x3)
S = factor(c(rep(1,12),rep(2,12),rep(3,12)))
T = factor(rep(c(rep(1,4),rep(2,4),rep(3,4)),3))
options(contrasts=c("contr.treatment","contr.poly"))
lm.fit=lm(x~S*T)
```

Extract the relevant results from the fitted object `lm.fit` to do the following:

(i) Give the linear model behind the R code in terms of dummy variables.

The model is as follows:

$$X = \mu_0 + \sum_{i=2}^{3} \beta_i t_i + \sum_{j=2}^{3} \gamma_j s_j + \sum_{i=2}^{3} \sum_{j=2}^{3} \xi_{ij} t_i s_j + \epsilon,$$

where

$$t_i = \begin{cases} 1, & \textit{if treatment } i, \\ 0, & \textit{otherwise}; \end{cases} \qquad s_j = \begin{cases} 1, & \textit{if stratum } j, \\ 0, & \textit{otherwise}. \end{cases}$$

(ii) Test the significance of stratum by treatment interaction at level $\alpha = 0.05$.

The anova table is as follows:

```
S           2  10633    5317  7.9834  0.001888 **
T           2  39083   19542 29.3438 1.694e-07 ***
S:T         4   9438    2359  3.5429  0.018973 *
Residuals 27  17981     666
```

The F ratio for the interaction effect is 3.5429. With a F-distribution with df 4 and 27, the p-value is 0.018973. At level $\alpha = 0.05$, the test concludes the significance of the interaction effect.

(iii) Suppose the interaction effect is significant, make pairwise comparisons within each of the three strata to find out how the treatment effects differ from stratum to stratum.

In terms of the parameters of the model, the three pairwise contrasts within stratum 1 are given by

$$\beta_2, \beta_3, \beta_2 - \beta_3;$$

the three pairwise contrasts within stratum 2 are given by

$$\beta_2 + \xi_{22}, \beta_3 + \xi_{32}, (\beta_2 + \xi_{22}) - (\beta_3 + \xi_{32});$$

and the three pairwise contrasts within stratum 3 are given by

$$\beta_2 + \xi_{23}, \beta_3 + \xi_{33}, (\beta_2 + \xi_{23}) - (\beta_3 + \xi_{33}).$$

The following R code is used for the computation of the test statistics The test statistics for the above contrasts are

$$L_{ij} = \frac{\boldsymbol{c}_{ij}^{'}\boldsymbol{b}}{\sqrt{\boldsymbol{c}_{ij}^{'}\Sigma\boldsymbol{c}_{ij}}},$$

where $\boldsymbol{b} = (\hat{\beta}_2, \hat{\beta}_3, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\xi}_{22}, \hat{\xi}_{32}, \hat{\xi}_{23}, \hat{\xi}_{33})$*,* Σ *is the estimated variance-covariance matrix of* $\boldsymbol{b}$*, and* $\boldsymbol{c}_{ij}$*'s are the contrast vector corresponding the contrasts.*

The following R-code is used for the computation:

```
V=vcov(lm.fit)[-1,-1]
b = lm.fit$coef[-1]
c11 = c(1,0,0,0,0,0,0,0)
c12 = c(0,1,0,0,0,0,0,0)
c13 = c11-c12
c21 = c(1,0,0,0,1,0,0,0)
c22 = c(0,1,0,0,0,1,0,0)
c23 = c21-c22
c31 = c(1,0,0,0,0,0,1,0)
c32 = c(0,1,0,0,0,0,0,1)
c33=c31-c32
C = rbind(c11,c12,c13,c21,c22,c23,c31,c32,c33)
l = C%*%b
s = sqrt(diag(C%*%V%*%t(C)))
L = l/s
```

The computed statistics are as follows:

```
   c11    c12    c13   c21    c22   c23   c31    c32   c33
-4.247 -4.233 -0.014 -2.11 -5.823 3.713 0.096 -3.206 3.302
```

Sheffe's criterion is appropriate for the comparisons. The critical value at level 0.05 is

$$\sqrt{6F_{6,27,0.05}} = 3.841.$$

Conclusion: In stratum 1, treatment 2 and 3 are significantly different from treatment 1, but there is no significant difference between treatment 2 and 3. In stratum 2, only the difference between treatment 1 and treatment 3 is significant. In stratum 3, none of the pairs has significant difference.

(iv) Regardless of the result in (ii), assume that there is no stratum by treatment interaction. Test the significance of the following contrasts: a) treatment 2 versus treatment 1, b) treatment 3 versus treatment 1, c) the average effect of treatment 2 and 3 versus treatment 1. (You need to refit the model without the interaction terms).

The model under the assumption is the one in (i) but without the interaction terms. In terms of the parameters of the model, the contrasts are:

$$\beta_2, \beta_3, \frac{1}{2}(\beta_2 + \beta_3).$$

The following R-code is used for the computation of the three test statistics:

```
lm.fit2=lm(x~T+S)
b=lm.fit2$coef[-c(1,4,5)]
V = vcov(lm.fit2)[-c(1,4,5),-c(1,4,5)]
c1 = c(1,0)
c2 = c(0,1)
c3 = c(1/2,1/2)
C=rbind(c1,c2,c3)
l = C%*%b
s = sqrt(diag(C%*%V%*%t(C)))
L = l/s
```

The computed statistics are as follows:

```
    c1     c2     c3
-3.137 -6.644 -5.647
```

Sheffe's criterion with df 2 and 31 is appropriate for the comparisons. The critical value at level 0.05 is

$$\sqrt{2F_{2,31,0.05}} = 2.571.$$

All the three contrasts are significant.

2. Consider the following model:

$$X_{ij} = \mu_i + \beta_i Z_{ij} + \epsilon_{ij},\ i = 1, \ldots, g; j = 1, \ldots, n_i.$$

(i) Derive the formula for the estimator of $\hat{\beta}_i$ and $\hat{\mu}_i$.

$$\begin{aligned}\hat{\beta}_i &= \frac{\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})(Z_{ij} - \bar{Z}_{i\cdot})}{\sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i\cdot})^2}. \\ \hat{\mu}_i &= \bar{X}_{i\cdot} - \hat{\beta}_i \bar{Z}_{i\cdot}.\end{aligned}$$

(ii) Show that the covariance of $\hat{\beta}_i$ and $\bar{X}_{i\cdot}$ equals zero.

$$\text{Cov}(\bar{X}_{i\cdot}, \hat{\beta}_i) = \frac{1}{\sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i\cdot})^2} \sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i\cdot})\text{Cov}(\bar{X}_{i\cdot}, X_{ij} - \bar{X}_{i\cdot}).$$

Note that

$$\text{Cov}(\bar{X}_{i\cdot}, X_{ij} - \bar{X}_{i\cdot}) = \text{Cov}(\bar{X}_{i\cdot}, X_{ij}) - \text{Var}(\bar{X}_{i\cdot}) = \frac{\sigma^2}{n_i} - \frac{\sigma^2}{n_i} = 0.$$

The result follows.

(iii) Derive the formulas for the variances of $\hat{\beta}_i$ and $\hat{\mu}_i$.

$$\begin{aligned}\text{Var}(\hat{\beta}_i) &= \frac{1}{[\sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i\cdot})^2]^2} \sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i\cdot})^2 \text{Var}(X_{ij}) = \frac{\sigma^2}{\sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i\cdot})^2} \\ \text{Var}(\hat{\mu}_i) &= \text{Var}(\bar{X}_{i\cdot}) + \text{Var}(\hat{\beta}_i \bar{Z}_{i\cdot}) \qquad \text{by (ii)} \\ &= \frac{\sigma^2}{n_i} + \frac{\sigma^2 \bar{Z}_{i\cdot}^2}{\sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i\cdot})^2}\end{aligned}$$

(iv) Suppose $\beta_i = \beta$ for all i. Show that the least square estimate of β has the form

$$\hat{\beta} = \frac{\sum_i w_i \hat{\beta}_i}{\sum_i w_i},$$

for certain w_i's.

$$
\begin{aligned}
\hat{\beta} &= \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i\cdot})(Z_{ij}-\bar{Z}_{i\cdot})}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(Z_{ij}-\bar{Z}_{i\cdot})^2} \\
&= \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(Z_{ij}-\bar{Z}_{i\cdot})^2\hat{\beta}_i}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(Z_{ij}-\bar{Z}_{i\cdot})^2}. \\
w_i &= \sum_{j=1}^{n_i}(Z_{ij}-\bar{Z}_{i\cdot})^2.
\end{aligned}
$$

3. Consider the following summary data from Example 2 of Lecture notes 6:

Quantity	Treatment 1	Treatment 2	Treatment 3
n_i	19	20	20
$\bar{X}_{i\cdot}$	5.3158	8.3000	8.5500
$\sum(X_{ij}-\bar{X}_{i\cdot})^2$	67.1053	76.2000	139.9500
$\bar{Z}_{i\cdot}$	2.3158	2.4500	3.1500
$\sum(Z_{ij}-\bar{Z}_{i\cdot})^2$	23.1050	15.9500	22.5500
$\sum(X_{ij}-\bar{X}_{i\cdot})(Z_{ij}-\bar{Z}_{i\cdot})$	27.1053	18.3000	34.3500

Assume the model of non-parallel regression lines as described in Problem 2 above for the data.

(i) Compute the estimates $\hat{\mu}_i, \hat{\beta}_i$ for $i = 1, 2, 3$.

Using the formulas in Problem 2, the estimates are computed by the following R code:

```
n=c(19, 20, 20)
x.m=c(5.3158, 8.3000, 8.5500)
s.x=c(67.1053, 76.2000, 139.9500)
z.m=c(2.3158, 2.4500, 3.1500)
s.z =c(23.1050, 15.9500, 22.5500)
s.xz=c(27.1053, 18.3000, 34.3500)
b = s.xz/s.z
mu = x.m-b*z.m
```

The computed values of the estimates are:

```
> b
[1] 1.173136 1.147335 1.523282
> mu
[1] 2.599052 5.489028 3.751663
```

(ii) Test the following null hypothesis at level $\alpha = 0.05$.

$$H_0 : \beta_1 = \beta_2 = \beta_3.$$

The hypothesis is expressed in two contrasts about the β_i's. A Wald statistic is constructed and computed by the following R code:

```
s2 = (s.x - s.xz^2/s.z)/(n-2)
ss2 = sum(s2*(n-2))/sum(n-2)
V.b = ss2*diag(1/s.z)
c1 = c(1,-1,0)
c2 = c(0,1,-1)
C = rbind(c1,c2)
W = t(C%*%b)%*%solve(C%*%V.b%*%t(C))%*% C%*%b
```

The value of the Wald statistic is 0.555, which cannot be significant at any reasonable level.

(iii) Regardless of the result of (ii), assume that $\beta_1 = \beta_2 = \beta_3 = \beta$. Compute the estimate of $\hat{\beta}$ and the estimates of μ_i, $i = 1, 2, 3$, under this assumption.

The estimates are computed by

```
bb = sum(b*s.z)/sum(s.z)
mmu = x.m-bb*z.m
```

The computed values are:

```
> bb
[1] 1.294624
> mmu
[1] 2.317710 5.128172 4.471935
```

(iv) Under assumption of (iii), test the following hypotheses simultaneously:

$$\mu_1 = \mu_2, \mu_1 = \mu_3, \mu_1 = \frac{1}{2}(\mu_2 + \mu_3).$$

The test statistics for the three contrasts are computed by:

```
nn=sum(n)
ss.x=sum(s.x)
ss.z=sum(s.z)
ss.xz =sum(s.xz)
ss2=(ss.x-ss.xz^2/ss.z)/(nn-3-1)
S = ss2*(z.m%*%t(z.m)/ss.z + diag(1/n) )
c1 = c(1,-1,0)
c2=c(1,0,-1)
c3 =c(1,-1/2,-1/2)
C=rbind(c1,c2,c3)
l=C%*%mmu
s = sqrt(diag(C%*%S%*%t(C)))
L=l/s
```

The computed values are:

```
    c1     c2     c3
-4.842 -3.528 -4.808
```

The values are compared with the Scheffe critical value

$$\sqrt{2F_{2,55,0.05}} = 2.516.$$

All the contrasts are significant at level 0.05.