ST4241: Design and Analysis of Clinical Trials

2009/2010: Semester I

Tutorial 5

1. The following is the results of a factorial study as given in Question 2 of Tutorial 4.

	Preparation 1		Preparation 2	
Block	Dose A	Dose B	Dose A	Dose B
1	3.0	5.5	5.0	6.0
2	2.0	4.0	4.5	5.5
3	2.5	5.0	4.0	5.0
4	3.0	4.5	4.5	6.0
5	3.0	4.0	2.5	5.5
6	3.5	4.5	4.5	5.5

Consider the four combinations of Preparation and Dose as four treatments. Conduct the Friedman's test for the significance of treatment effect at level 0.05.

The ranks within blocks and average ranks of the treatments are given in the following:

```
            r1   r2       r3       r4
      1.000000 3.00 2.000000 4.000000
      1.000000 2.00 3.000000 4.000000
      1.000000 3.50 2.000000 3.500000
      1.000000 2.50 2.500000 4.000000
      2.000000 3.00 1.000000 4.000000
      1.000000 2.50 2.500000 4.000000
Rmean 1.166667 2.75 2.166667 3.916667
```

Note that $g = 4, n. = 6$. *The Friedman's test statistic is computed as*

$$\chi^2 = \frac{12n.}{g(g+1)} \sum_{j=1}^{n.} (\bar{R}_j - \frac{g+1}{2})^2 = 14.25.$$

The p-value of the test is

$$p = P(\chi^2(3) \geq 14.25) = 0.0026.$$

The test is significant.

2. The data for the comparison of the antidepressant drug Imipramine and a placebo in a matched pair design is as follows.

Pair	Imi	Pla	d	Pair	Imi	Pla	d
1	6	4	2	16	6	8	-2
2	4	7	-3	17	10	10	0
3	6	12	-6	18	3	9	-6
4	7	10	-3	19	5	8	-3
5	5	2	3	20	4	5	-1
6	6	11	-5	21	6	8	-2
7	8	9	-1	22	7	7	0
8	7	5	2	23	5	6	-1
9	8	11	-3	24	6	9	-3
10	3	8	-5	25	3	3	0
11	9	7	2	26	10	5	5
12	4	6	-2	27	5	11	-6
13	8	8	0	28	4	7	-3
14	11	9	2	29	4	3	1
15	12	9	3	30	7	10	-3

(i) Complete the following anova table for the data above:

The anova table is completed as follows:

Source of variation	*df*	*SS*	*MS*	*F ratio*
Treatment	*1*	*24.06667*	*24.06667*	*5.6315*
Pairs	*29*	*237.7333*	*8.1977*	
Residuals	*29*	*123.93333*	*4.2736*	

The content of the table is computed using the formulae for the sum of squares. The computation is done by the following R-codes:

```
Imi = c(6,4,6,7,5,6,8,7,8,3,9,4,8,11,12,6,10,3,5,4,6,7,5,6,3,10,5,4,4,7)
Pla = c(4,7,12,10,2,11,9,5,11,8,7,6,8,9,9,8,10,9,8,5,8,7,6,9,3,5,11,7,3,10)
y=c(Imi, Pla)
y.m=mean(y)
Imi.m=mean(Imi)
Pla.m=mean(Pla)
X = cbind(Imi,Pla)
b.mean = apply(X,1,mean)
TSS = 30*((Imi.m-y.m)^2 + (Pla.m-y.m)^2)
BSS = 2*sum((b.mean-y.m)^2)
```

```
X1 = b.mean%*%t(c(1,1))
X2 = rep(1,30)%*%t(c(Imi.m,Pla.m))
X3 = rep(1,30)%*%t(c(1,1))*y.m
RSS=sum(apply((X-X1-X2+X3)^2,2,sum))

# another way to compute RSS
TTSS = sum((y-y.m)^2)
RSS = TTSS-TSS-BSS
```

Confirm that the value of the F-ratio is the squared value of the t statistic given by

$$t = \frac{\bar{d}\sqrt{n}}{s_d}.$$

The t-value is computed as -2.3731. $(-2.3731)^2 = 5.6315$.

(ii) In general, show that whenever $g = 2$ treatments are compared in a randomized blocks design study, the F ratio from the anova table is identically equal to the square of the t-statistic given above.

When $g = 2$,

$$\bar{X}_{\cdot 1} - \bar{X}_{\cdot\cdot} = \bar{X}_{\cdot 2} - \bar{X}_{\cdot\cdot} = \frac{\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2}}{2}.$$

Thus

$$\text{TSS} = \frac{n}{2}(\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2})^2 = \frac{n}{2}\bar{d}^2.$$

When $g = 2$, we can write

$$\begin{aligned} d_i - \bar{d} &= X_{i1} - X_{i2} - (\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2}) \\ &= 2(X_{i1} - \bar{X}_{i\cdot} - \bar{X}_{\cdot 1} + \bar{X}_{\cdot\cdot}) \\ &= 2(X_{i2} - \bar{X}_{i\cdot} - \bar{X}_{\cdot 2} + \bar{X}_{\cdot\cdot}). \end{aligned}$$

Thus

$$\begin{aligned} \frac{1}{2}\sum_{i=1}^{n}(d_i - \bar{d})^2 &= 2\sum_{i=1}^{n}(X_{i1} - \bar{X}_{i\cdot} - \bar{X}_{\cdot 1} + \bar{X}_{\cdot\cdot})^2 \\ &= \sum_{i=1}^{n}(X_{i1} - \bar{X}_{i\cdot} - \bar{X}_{\cdot 1} + \bar{X}_{\cdot\cdot})^2 + \sum_{i=1}^{n}(X_{i2} - \bar{X}_{i\cdot} - \bar{X}_{\cdot 2} + \bar{X}_{\cdot\cdot})^2 \\ &= 29 \text{ ISS} = 29 \text{ RSS}. \end{aligned}$$

Hence RMS $= s_d^2/2$ *and*

$$F = \frac{\text{TMS}}{\text{RMS}} = \left[\frac{n}{2}\bar{d}^2\right] / \left[\frac{s_d^2}{2}\right] = t^2.$$

3. Refer to the data below on the blood samples from 8 persons.

	Treatment				
Subject	1	2	3	4	Mean
1	8.4	9.4	9.8	12.2	9.950
2	12.8	15.2	12.9	14.4	13.825
3	9.6	9.1	11.2	9.8	9.925
4	9.8	8.8	9.9	12.0	10.125
5	8.4	8.2	8.5	8.5	8.400
6	8.6	9.9	9.8	10.9	9.800
7	8.9	9.0	9.2	10.4	9.375
8	7.9	8.1	8.2	10.0	8.550
Mean	9.300	9.713	9.938	11.025	9.994
sd	1.550	2.294	1.514	1.815	

(i) Define dummy variables for the subjects and treatments. Describe the data by a linear model using the dummy variables.

The dummy variables are defined as follows:

$$t_j = \begin{cases} 1, & \textit{if treatment } j+1, \\ 0, & \textit{otherwise,} \end{cases} \quad j = 1, 2, 3.$$

$$s_i = \begin{cases} 1, & \textit{if subject } i+1, \\ 0, & \textit{otherwise,} \end{cases} \quad i = 1, \ldots, 7.$$

The linear model for one measurement is given by

$$y = mu_0 + \sum_{i=1}^{7} \alpha_i s_i + \sum_{j=1}^{2} \beta_j t_j + \epsilon.$$

(ii) Using the R function `lm`, obtain the anova table for the data.

The R-codes are given below:

```
y1 = c(8.4,12.8,9.6,9.8,8.4,8.6,8.9,7.9)
y2 = c(9.4,15.2,9.1,8.8,8.2,9.9,9.0,8.1)
y3 = c(9.8,12.9,11.2,9.9,8.5,9.8,9.2,8.2)
y4 = c(12.2,14.4,9.8,12.0,8.5,10.9,10.4,10)
y=c(y1,y2,y3,y4)
tmt=factor(c(rep(1,8),rep(2,8),rep(3,8),rep(4,8)))
blk=factor(rep(1:8,4))
options(contrasts=c("contr.treatment","contr.poly"))
lm.fit= lm(y~tmt+blk)
anova(lm.fit)
```

The anova table is obtained as follows:

```
          Df Sum Sq Mean Sq F value    Pr(>F)
tmt        3 13.016   4.339   6.615  0.002550 **
blk        7 78.989  11.284  17.204 2.197e-07 ***
Residuals 21 13.774   0.656
```

(iii) Using the results from the object obtained by `lm`, test the significance of the following contrasts:

$$C_1 = \mu_{\cdot 1} - \mu_{\cdot 3},$$
$$C_2 = \mu_{\cdot 2} - (\mu_{\cdot 1} + \mu_{\cdot 3})/2,$$
$$C_3 = \frac{1}{3}(\mu_{\cdot 1} + \mu_{\cdot 2} + \mu_{\cdot 3}) - \mu_{\cdot 4}.$$

where $\mu_{\cdot j}$ is the mean response of treatment j. Use an appropriate criterion to control the overall error rate at level 0.05.

In terms of the parameters in the linear model, the contrasts are equivalent to

$$C_1 = \beta_2,$$
$$C_2 = \beta_1 - \beta_2/2,$$
$$C_3 = \frac{1}{3}(\beta_1 + \beta_2) - \beta_3.$$

The following R-codes are used to compute the test statistics for these contrasts.

```
vv=vcov(lm.fit)
V=vv[2:4,2:4]
bb=lm.fit$coef[2:4]
c1=c(0,1,0)
c2=c(1,-1/2,0)
c3=c(1/3,1/3,-1)
C1 =  t(c1)%*%bb
C2 =  t(c2)%*%bb
C3 =  t(c3)%*%bb
v1 = V[2,2]
v2 = t(c2)%*%V%*%c2
v3 = t(c3)%*%V%*%c3
L1 = C1/sqrt(v1)
L2 = C2/sqrt(v2)
L3 = C3/sqrt(v3)
```

It is obtained that

$$(L1, L2, L3) = (1.5743, 0.2673, -4.1587).$$

The critical value of Scheffe's and Bonferroni's criteria are

```
scheffe = sqrt(3*qf(0.95,3,21)) =  3.0360
bonferroni = qt(1-0.05/6, 21) = 2.6014
```

The appropriate critical value is the one given by Bonferroni's criterion. Compared with this critical value, only the third contrast is significant.

Note: for those who are required to submit the assignment, please submit the solutions to Question 1 and 2 only.