ST4241: Design and Analysis of Clinical Trials

2009/2010: Semester I

Tutorial 4

1. In the table below, each block represents a different subject, the units within blocks are four blood samples from each subject, and four treatments were randomly assigned to the blood samples within each set. The values are the clotting times of plasm, in minutes.

	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) Suppose we are only interested in investigating the following three differences: (a) $\mu_1 - \mu_4$, (b) $\mu_2 - \mu_4$ and (c) $\mu_3 - \mu_4$, where μ_j is the mean effect of treatment j. Test the significance of these differences at level 0.05 by using an appropriate multiple comparison criterion.

The ANOVA table which is given in Lecture notes 4 is as follows:

Source	*df*	*SS*	*MS*	*F ratio*
Treatments	*3*	*13.0163*	*4.3388*	*6.62*
Subjects	*7*	*78.9888*	*11.2841*	
Residuals	*21*	*13.7737*	*0.6559*	
Total	*31*	*105.7788*		

The three test statistics are computed below:

$$
\begin{aligned}
L_1 &= \frac{\hat{\mu}_1 - \hat{\mu}_4}{\text{RMS}}\sqrt{\frac{n}{2}} = \frac{9.3 - 11.025}{0.6559}\sqrt{4} = -5.26;\\
L_2 &= \frac{\hat{\mu}_2 - \hat{\mu}_4}{\text{RMS}}\sqrt{\frac{n}{2}} = \frac{9.713 - 11.025}{0.6559}\sqrt{4} = -4.00;\\
L_3 &= \frac{\hat{\mu}_3 - \hat{\mu}_4}{\text{RMS}}\sqrt{\frac{n}{2}} = \frac{9.938 - 11.025}{0.6559}\sqrt{4} = -3.31.
\end{aligned}
$$

The Dunnett's criterion is appropriate for these comparisons. The critical value of the Dunnett's criterion at 0.05 level for two sided test is $d_{3,21,0.05} \approx 2.54$. Comparing the absolute values of the L_j's with this critical value shows that all the differences are significant.

(ii) Give the critical values controlling the overall error rate at 0.05 of Scheff's, Tukey's, Dunnett's and Bonferroni's criterion respectively.

Scheffe's critical value:

$$\sqrt{3F_{3,21,0.05}} = 3.036.$$

Tukey's critical value:

$$q_{4,21,0.05}/\sqrt{2} \approx 3.958/\sqrt{2} = 2.799.$$

Bonferroni's critical value:

$$t_{21,0.05/6} = 2.60.$$

Dunnett's critical value is 2.54 as given in part (i).

(iii) Test the three differences in part (i) using each of the criteria in part (ii). Comment on the results.

All three differences are still significant by using any of the criteria.

Since all the criteria control the overall error rate at 0.05 level, the ones with smaller critical values are more powerful in detecting significant differences.

2. The following is the results of a factorial study:

	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

(i) Complete the following table:

Source of variation	df	SS
Preparations		
Doses		
Interaction		
Blocks		
Residual		

The completed table is as follows:

Source of variation	*df*	*SS*
Preparations	*1*	*8.1667*
Doses	*1*	*15.0417*
Interaction	*1*	*0.1667*
Blocks	*5*	*3.3333*
Residual	*15*	*4.25*

The content of the table can be obtained by using the R function `lm` *as follows:*

```
x = c(3,2,2.5,3,3,3.5,
      5.5,4,5,4.5,4,4.5,
      5,4.5,4,4.5,2.5,4.5,
      6,5.5,5,6,5.5,5.5)
p = factor(c(rep(1,12), rep(2,12)))
d = factor(rep( c(rep(1,6), rep(2,6)), 2))
b = factor( rep(1:6,4) )
options(contrasts=c("contr.treatment","contr.poly"))
lm.o = lm(x~p+b+d+p*d)
anova(lm.o)
```

It produces the following anova table:

```
           Df  Sum Sq Mean Sq F value    Pr(>F)
p           1  8.1667  8.1667 28.8235 7.817e-05 ***
b           5  3.3333  0.6667  2.3529   0.09133 .
d           1 15.0417 15.0417 53.0882 2.669e-06 ***
p:d         1  0.1667  0.1667  0.5882   0.45501
Residuals 15  4.2500  0.2833
```

The sum of squares can also be computed manually by using the formulae for various sum of squares. For the computation of the sum of squares for a particular factor, other factors can be ignored. For instance, the SS for doses can be computed as

$$\begin{aligned} \mathrm{SS}_d &= \sum_{j=1}^{2} 12(\bar{X}_{..1} - \bar{X}_{...})^2 \\ &= 12[(3.5 - 4.292667)^2 + (5.083333 - 4.292667)^2] = 15.0417. \end{aligned}$$

(ii) Test whether the interaction effect is significant at level 0.05. If not, test the significance of the effects of the two factors: preparation and dose.

The interaction is tested by the ratio

$$F = \frac{0.1667}{0.2833} = 0.4550,$$

which is not significant compared with $F_{1,15,0.05} = 4.5430$.

The p-values for testing preparation and dose can be found from the anova table. They are respectively 7.817e-05 and 2.669e-06. Therefore, both pareparation and dose are significant.

3. The following concerns the nonparametric tests for CRBD data.

(i) Prove that the numerator of the Signed rank test statistic Z is equal to

$$\frac{n_+ n_-}{n.}(\bar{R}_+ - \bar{R}_-) + \frac{(n. + 1)(n_+ - n_-)}{4}.$$

The statistic Z is given by

$$Z = \frac{R_+ - \frac{n.(n.+1)}{4}}{\sqrt{\frac{n.(n.+1)(2n.+1)f}{24}}},$$

where f is the factor for adjusting ties.

$$
\begin{aligned}
& R_+ - \frac{n.(n.+1)}{4} \\
= \;& R_+ - \frac{2n_+(n.+1)}{4} + \frac{(n_+ - n_-)(n.+1)}{4} \\
= \;& R_+ - \frac{n_+}{n.}\frac{n.(n.+1)}{2} + \frac{(n_+ - n_-)(n.+1)}{4} \\
= \;& n_+\bar{R}_+ - \frac{n_+}{n.}(n_+\bar{R}_+ + n_-\bar{R}_-) + \frac{(n_+ - n_-)(n.+1)}{4} \\
= \;& \frac{n_+}{n.}(n_-\bar{R}_+ - n_-\bar{R}_-) + \frac{(n_+ - n_-)(n.+1)}{4} \\
= \;& \frac{n_+n_-}{n.}(\bar{R}_+ - \bar{R}_-) + \frac{(n_+ - n_-)(n.+1)}{4}.
\end{aligned}
$$

(ii) Prove that, when $g = 2$, the Friedman's test statistic is equal to

$$\chi^2 = \frac{(n_+ - n_-)^2}{n.}.$$

When $g = 2$, the Friedman's test statistic is given by

$$\chi^2 = 2n.[(\bar{R}_1 - \frac{3}{2})^2 + (\bar{R}_2 - \frac{3}{2})^2].$$

Note that

$$\bar{R}_1 - \frac{3}{2} = \frac{1}{n.}\sum_{i=1}^{n.}(R_{i1} - \frac{3}{2}).$$

Since $R_{i1} = 1$ or 2, $R_{i1} - \frac{3}{2} = -\frac{1}{2}$ or $\frac{1}{2}$. There are n_+ of the differences which are $\frac{1}{2}$ and n_- of them which are $-\frac{1}{2}$. Therefore

$$\bar{R}_1 - \frac{3}{2} = \frac{(n_+ - n_-)}{2n.}.$$

A similar argument yields

$$\bar{R}_2 - \frac{3}{2} = \frac{(n_- - n_+)}{2n.}.$$

Thus

$$\chi^2 = 4n. \left[\frac{(n_- - n_+)}{2n.}\right]^2 = \frac{(n_- - n_+)^2}{n.}.$$

4. The following table is the layout of the data from a complete random blocks design:

	Treatment					
Block	1	$\cdots$	j	$\cdots$	g	Mean
1	X_{11}	$\cdots$	X_{1j}	$\cdots$	X_{1g}	$\bar{X}_{1\cdot}$
$\vdots$						
i	X_{i1}	$\cdots$	X_{ij}	$\cdots$	X_{ig}	$\bar{X}_{i\cdot}$
$\vdots$						
n	X_{n1}	$\cdots$	X_{nj}	$\cdots$	X_{ng}	$\bar{X}_{n\cdot}$
Mean	$X_{\cdot 1}$	$\cdots$	$X_{\cdot j}$	$\cdots$	$X_{\cdot g}$	$X_{\cdot\cdot}$
sd	s_1	$\cdots$	s_j	$\cdots$	s_g	

A linear model for the data is as follows:

$$X_{ij} = \mu_j + b_i + \epsilon_{ij},$$

where μ_j is the underlying mean response to Treatment j, b_i is the effect on the response due to the particular characteristics of the experimental units constituting Block i, and ϵ_{ij} is a random variable representing chance measurement errors and other random perturbations in the data. Assume that b_i's are random variables with mean zero and variance σ_b^2 and are independent of ϵ_{ij}'s.

(i) Prove that the expected value of the mean square for blocks is $\sigma_\epsilon^2 + g\sigma_b^2$.

(ii) Prove that the expected value of the residual mean squares is σ_ϵ^2.

First,

$$\text{BMS} = \frac{g}{n-1}\sum_{i=1}^{n}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2 = \frac{g}{n-1}[\sum_{i=1}^{n}\bar{X}_{i\cdot}^2 - n\bar{X}_{\cdot\cdot}^2].$$

$$\begin{aligned}\text{RMS} &= \frac{1}{(g-1)(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{g}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{\cdot\cdot})^2 \\ &= \frac{1}{(g-1)(n-1)}[\sum_{i=1}^{n}\sum_{j=1}^{g}X_{ij}^2 - g\sum_{i=1}^{n}\bar{X}_{i\cdot}^2 - n\sum_{j=1}^{g}\bar{X}_{\cdot j}^2 + ng\bar{X}_{\cdot\cdot}^2].\end{aligned}$$

Next,

$$
\begin{aligned}
& Var(X_{ij}) = \sigma_b^2 + \sigma_e^2,\ E(X_{ij}) = \mu_j \\
\Rightarrow\ & E(X_{ij}^2) = \sigma_b^2 + \sigma_e^2 + \mu_j^2.
\end{aligned}
$$

$$
\begin{aligned}
& Var(\bar{X}_{i\cdot}) = \sigma_b^2 + \sigma_e^2/g,\ E\bar{X}_{i\cdot} = \bar{\mu}. \\
\Rightarrow\ & E(\bar{X}_{i\cdot}^2) = \sigma_b^2 + \sigma_e^2/g + \bar{\mu}_{\cdot}^2.
\end{aligned}
$$

$$
\begin{aligned}
& Var(\bar{X}_{\cdot j}) = \sigma_b^2/n + \sigma_e^2/n,\ E\bar{X}_{\cdot j} = \mu_j \\
\Rightarrow\ & E(\bar{X}_{\cdot j}^2) = \sigma_b^2/n + \sigma_e^2/n + \mu_j^2.
\end{aligned}
$$

$$
\begin{aligned}
& Var(\bar{X}_{\cdot\cdot}) = \sigma_b^2/n + \sigma_e^2/(ng),\ E\bar{X}_{\cdot\cdot} = \bar{\mu}. \\
\Rightarrow\ & E(\bar{X}_{\cdot\cdot}^2) = \sigma_b^2/n + \sigma_e^2/(ng) + \bar{\mu}_{\cdot}^2.
\end{aligned}
$$

Substituting the above expression into the mean sum of squares yields

$$
\begin{aligned}
E(\text{BMS}) &= \frac{g}{n-1}[n(\sigma_b^2 + \sigma_e^2/g + \bar{\mu}_{\cdot}^2) - (\sigma_b^2 + \sigma_e^2/g + n\bar{\mu}_{\cdot}^2) \\
&= \frac{g}{n-1}[(n-1)\sigma_b^2 + (n-1)\sigma_e^2/g] = g\sigma_b^2 + \sigma_e^2.
\end{aligned}
$$

$$
\begin{aligned}
E(\text{RMS}) &= \frac{1}{(g-1)(n-1)}[ng(\sigma_b^2 + \sigma_e^2) + n\sum_{j=1}^{g}\mu_j^2 - ng(\sigma_b^2 + \sigma_e^2/g + \bar{\mu}_{\cdot}^2) \\
&\quad -n(g\sigma_b^2/n + g\sigma_e^2/n + \sum_{j=1}^{g}\mu_j^2) + ng(\sigma_b^2/n + \sigma_e^2/(ng) + \bar{\mu}_{\cdot}^2)]. \\
&= \frac{1}{(g-1)(n-1)}(ng - n - g + 1)\sigma_e^2 = \sigma_e^2.
\end{aligned}
$$