ST4241: Design and Analysis of Clinical Trials

2009/2010: Semester I

Tutorial 9

1. The effect of five different ingredients (A,B, C, D, E) on the reaction time of a chemical process is being studied . Each batch of new material is only large enough to permit five runs to be made. Furthermore, each run requires approximately $1\frac{1}{2}$ hours, so only five runs can be made in one day. The experimenter decides to run the experiment as a Latin square so that day and batch effects may be systematically controlled. The data of the experiment is as follows.

	Day				
Batch	1	2	3	4	5
1	8 (A)	7 (B)	1 (D)	7 (C)	3 (E)
2	11 (C)	2 (E)	7 (A)	3 (D)	8 (B)
3	4 (B)	9 (A)	10 (C)	1 (E)	5 (D)
4	6 (D)	8 (C)	6 (E)	6 (B)	10 (A)
5	4 (E)	2 (D)	3 (B)	8 (A)	8 (C)

The following R code is for the computation.

```
x=c(8,11,4,6,4,
    7,2,9,8,2,
    1,7,10,6,3,
    7,3,1,6,8,
    3,8,5,10,8)
batch = factor(rep(c(1:5),5))
day = factor(kronecker(c(1:5),c(rep(1,5))))
tmt = factor(c(1,3,2,4,5,
               2,5,1,3,4,
               4,1,3,5,2,
               3,4,5,2,1,
               5,2,4,1,3))
options(contrasts=c("contr.treatment","contr.poly"))
lm.fit1 = lm(x~batch+day+tmt)
```

(i) Derive the ANOVA table for the data above.

The ANOVA table is retrieved by `anova(lm.fit1)` as

```
          Df  Sum Sq Mean Sq F value    Pr(>F)
batch      4  15.440   3.860  1.2345 0.3476182
day        4  12.240   3.060  0.9787 0.4550143
tmt        4 141.440  35.360 11.3092 0.0004877 ***
Residuals 12  37.520   3.127
```

(ii) Test whether the five ingredients have significant difference in their effect on reaction time at level 0.05. Draw your conclusion.

The F-ratio for testing the ingredient effect can be read from the ANOVA table above as $F = 11.3092$ which has a p-value 0.00049. The significance of the effect is claimed at any level greater than 0.00049.

2. The following table provides the data of the study comparing four infant formulas with a single Latin square design, where the observation X_{213} is missing.

	Week			
Infant	1	2	3	4
1	0.40(2)	1.11(3)	1.16(4)	0.88(1)
2	(3)	1.04(4)	0.57(1)	0.80(2)
3	1.14(1)	1.11(2)	1.32(3)	1.38(4)
4	1.08(4)	1.34(1)	1.73(2)	1.55(3)

(i) Define a linear model for the data above.

The linear model is as follows:

$$X = \mu_0 + \sum_{i=2}^{4} \alpha_i s_i + \sum_{j=2}^{4} \beta_j w_j + \sum_{k=2}^{4} \gamma_k t_k + \epsilon,$$

where

$$s_i = \begin{cases} 1, & \text{if infant } i, \\ 0, & \text{otherwise}, \ i = 2, 3, 4; \end{cases}$$

$$w_j = \begin{cases} 1, & \text{if week}\text{J}, \\ 0, & \text{otherwise}, \ j = 2, 3, 4; \end{cases}$$

$$t_k = \begin{cases} 1, & \text{if formula } k, \\ 0, & \text{otherwise}, \ k = 2, 3, 4. \end{cases}$$

The model is fitted by the R code below:

```
x=c(0.4,0.2,1.14,1.08,
    1.11,1.04,1.11,1.34,
    1.16,0.57,1.32,1.73,
    0.88,0.8,1.38,1.55)
infant = factor(rep(c(1:4),4))
week = factor(c(rep(1,4),rep(2,4),rep(3,4),rep(4,4)))
tmt = factor(c(2,3,1,4,3,4,2,1,4,1,3,2,1,2,4,3))
options(contrasts=c("contr.treatment","contr.poly"))
x1 =x[-2]
infant1=infant[-2]
week1=week[-2]
tmt1=tmt[-2]
lm.fit=lm(x1~infant1+week1+tmt1)
b=lm.fit$coef
v=vcov(lm.fit)
```

(ii) Using the linear model approach, test the significance of the treatment (Formula) effect at level $\alpha = 0.05$. Give the value of the F statistic and draw your conclusion.

The R code below computes the F statistic:

```
b.t=b[8:10]
v.t = v[8:10,8:10]
F.t= t(b.t)%*%solve(v.t)%*%b.t/3
```

The computed value is $F.t = 0.4940758$. The p-value `1-pf(F.t,3,5)` is 0.701933. The formula effect is not significant at level 0.05.

(iii) Using the linear model approach, test the significance of the time (Week) effect at level $\alpha = 0.05$. Give the value of the F statistic and draw your conclusion .

The R code below computes the F statistic:

```
b.w=b[5:7]
v.w= v[5:7,5:7]
F.w=t(b.w)%*%solve(v.w)%*%b.w/3
```

The computed value is $F.w = 1.835866$. The p-value `1-pf(F.w,3,5)` is 0.2577941. The time effect is not significant at level 0.05.

Hint: the F statistics can be obtained from the Wald statistics corresponding to the two tests.

3. The yield of a chemical process was measured using five batches of raw material, five acid concentrations, five standing times (A,B,C,D,E) and five catalyst concentrations ($\alpha, \beta, \gamma, \delta, \epsilon$). The Graeco-Latin square that follows was used. The observations of yield are imposed on the square.

	Acid concentration				
Batch	1	2	3	4	5
1	$A\alpha = 26$	$B\beta = 16$	$C\gamma = 19$	$D\delta = 16$	$E\epsilon = 13$
2	$B\gamma = 18$	$C\delta = 21$	$D\epsilon = 18$	$E\alpha = 11$	$A\beta = 21$
3	$C\epsilon = 20$	$D\alpha = 12$	$E\beta = 16$	$A\gamma = 25$	$B\delta = 13$
4	$D\beta = 15$	$E\gamma = 15$	$A\delta = 22$	$B\epsilon = 14$	$C\alpha = 17$
5	$E\delta = 10$	$A\epsilon = 24$	$B\alpha = 17$	$C\beta = 17$	$D\gamma = 14$

The data is fitted to a linear model by the following R code:

```
x=c(8,11,4,6,4,
    7,2,9,8,2,
    1,7,10,6,3,
    7,3,1,6,8,
    3,8,5,10,8)
batch = factor(rep(c(1:5),5))
acid = factor(kronecker(c(1:5),c(rep(1,5))))
time = factor(c(1,2,3,4,5,
                2,3,4,5,1,
                3,4,5,1,2,
                4,5,1,2,3,
                5,1,2,3,4))
catalyst = factor(c(1,3,5,2,4,
                2,4,1,3,5,
                3,5,2,4,1,
                4,1,3,5,2,
                5,2,4,1,3))
options(contrasts=c("contr.treatment","contr.poly"))
lm.fit = lm(x~batch+acid+time+catalyst)
```

(i) Derive the ANOVA table of the data above.

The ANOVA table is retrived by `anova(lm.fit)` as

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
batch	4	15.44	3.86	0.2572	0.8973
acid	4	12.24	3.06	0.2039	0.9292
time	4	21.04	5.26	0.3504	0.8369
catalyst	4	37.84	9.46	0.6302	0.6547
Residuals	8	120.08	15.01		

(ii) Analyze the data and draw conclusions on various effects.

Conclusion can be drawn on all the four factors: Batch, Acid concentration, Standing time, and Catalyst concentration. The F values given in the ANOVA table above do not support the singnificance of the effect of any of the four factors. Note that the smallest p-value is 0.6547.