STAT 6910: HW 8

David Angeles

```
## Warning: package 'emmeans' was built under R version 3.4.4

## NOTE: As of emmeans versions > 1.2.3,
##       The 'cld' function will be deprecated in favor of 'CLD'.
##       You may use 'cld' only if you have package:multcomp attached.

## Warning: package 'dae' was built under R version 3.4.4

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.4.4
```

Problem 1 (6 in text)

The purpose of the experiment run by M. Weber, R. Zielinski, J. Y. Lee, S. Xia, and Y. Guo in 2010 was to determine the best way to heat 3 cups of water (for preparation of boxed meals) to 90°F on a kitchen stove as qsuickly as possible. In this experiment, only one stove was used, and the three treatment factors were

C: diameter of pot (5.5, 6.25 and 8.625 inches; coded 1, 2, 3)
D: burner size (small, large; coded 1, 2)
E: cover (no, yes; coded 1, 2).

(b) A pilot experiment suggested that the error variance σ^2 would be no larger than 318.9 sec^2. The experimenters wanted to be able to test the hypothesis of no differences in the effects of heating time due to the 12 treatments, with a probability of 0.9 of rejecting the hypothesis if the true difference was $\Delta = 60$ secs. The test was to be done at level $\alpha = 0.05$. Calculate the number of observations that should be taken on each of the 12 treatments.

(In Part (b), either follow the instructions in Chapter 3 (ignoring blocking) or follow the instructions in Section 10.6.3, which are quite similar, assuming b = 4 blocks.)

solution

```r
v = 12; sig2 = 318.9; alpha = 0.05; pwr = 0.90; del = c(60); list_of_size <- NULL
for (i in 1) {delta <- del[i]
x<- pwr.anova.test(k = v, sig.level = alpha , power = pwr ,
f = sqrt(delta^2/(2*v*sig2)))
list_of_size <- c(list_of_size, x$n)
}
list_of_size
```

```
## [1] 4.643032
```

If we ignore blocking and use $\sigma^2 = 318.9$, $\Delta = 60$, $v = 12$, $\pi = .90$, and an α-level of $\alpha = .05$ we get that the sample size needd in 5.

(c)

The experimenters ultimately decided that they would use a randomized complete block design with $b = 4$ blocks for the experiment, where each block was defined by experimenter and day. The data are shown in Table 10.20. Using the block–treatment model (10.4.1), p. 310, for a randomized complete block design, check the assumptions on the model and test the hypothesis of no effects on the heating time due to treatments.

(In Part (c), you may detect an outlier – in practice you might run the subsequent analyses with and without the outlier and include both results, commenting on any discrepancies. For this homework, please just proceed with the outlier still in the data set – this may not be the best choice, but it will make everyone's answers uniform!)

Using the block–treatment model (10.4.1) we have:

$$\begin{gathered} Y_{hi} = \mu + \theta_h + \tau_i + \epsilon_{hi}, \\ \epsilon_{hi} \sim N(0, \sigma^2) \\ \epsilon_{hi}\text{'s mutually independent,} \\ h = 1, \dots, b;\ i = 1, \dots, v. \end{gathered}$$

where μ is a constant, θ_h is the effect of the hth block, τ_i is the effect of the ith treatment, Y_{hi} is the random variable representing the measurement on the treatment i observed in the block h, and ϵ_{hi} is the associated random error.

solution_

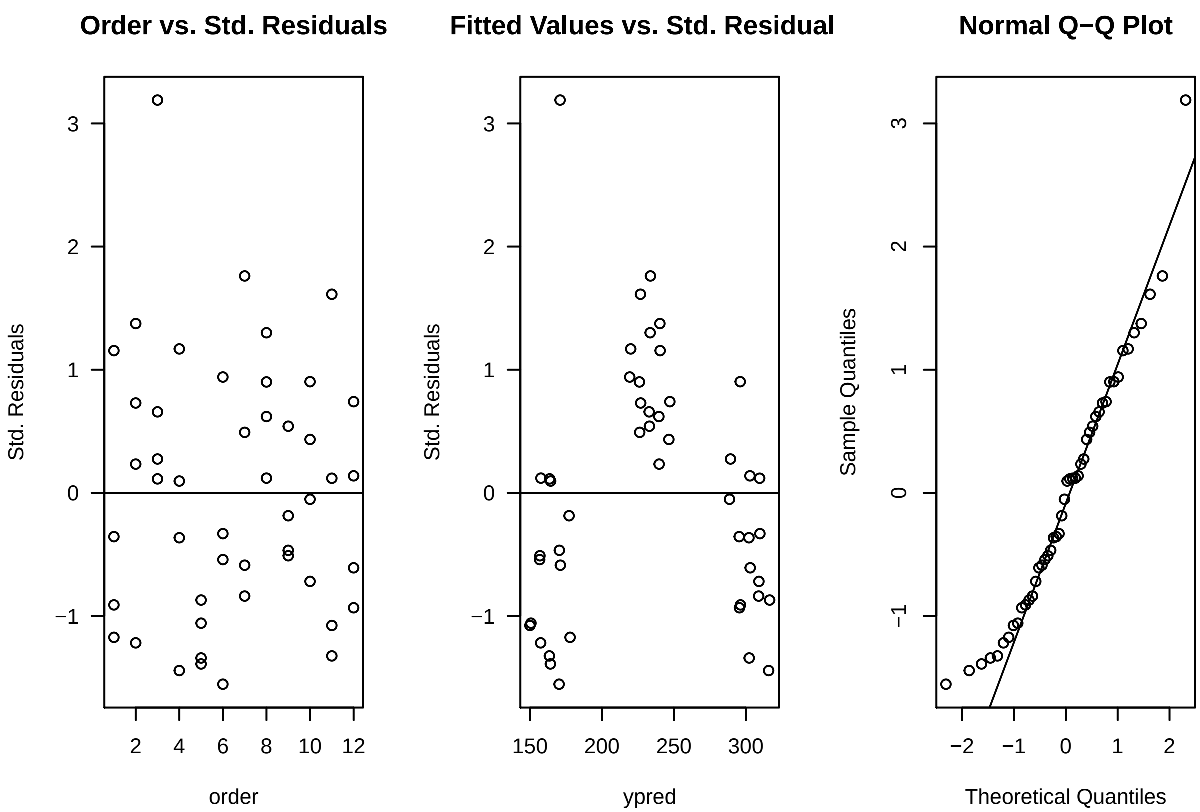
```r
two.way_model <- aov(time ~ block+trtmt, data = water.heating)

water.heating = within(water.heating, {
  # Compute predicted, residual, and standardized residual values
  ypred = fitted(two.way_model)
  e = resid(two.way_model)
  z = rstandard(two.way_model)
  new.code= rep(1:12, each =4)})
# Display first 10 lines of water.heating, 4 digits per variable
print(head(water.heating, 5), digits = 4)
```

```
##   trtmt C D E block  time order new.code       z      e ypred
## 1   111 1 1 1     1 261.0     1        1 -0.9107 -35.27 296.3
## 2   111 1 1 1     2 279.0    12        1 -0.6093 -24.03 303.0
## 3   111 1 1 1     3 296.7     6        1 -0.3318 -13.09 309.8
## 4   111 1 1 1     4 282.8     5        1 -0.8712 -33.74 316.5
## 5   112 1 1 2     1 259.4    12        2 -0.9338 -36.18 295.6
```

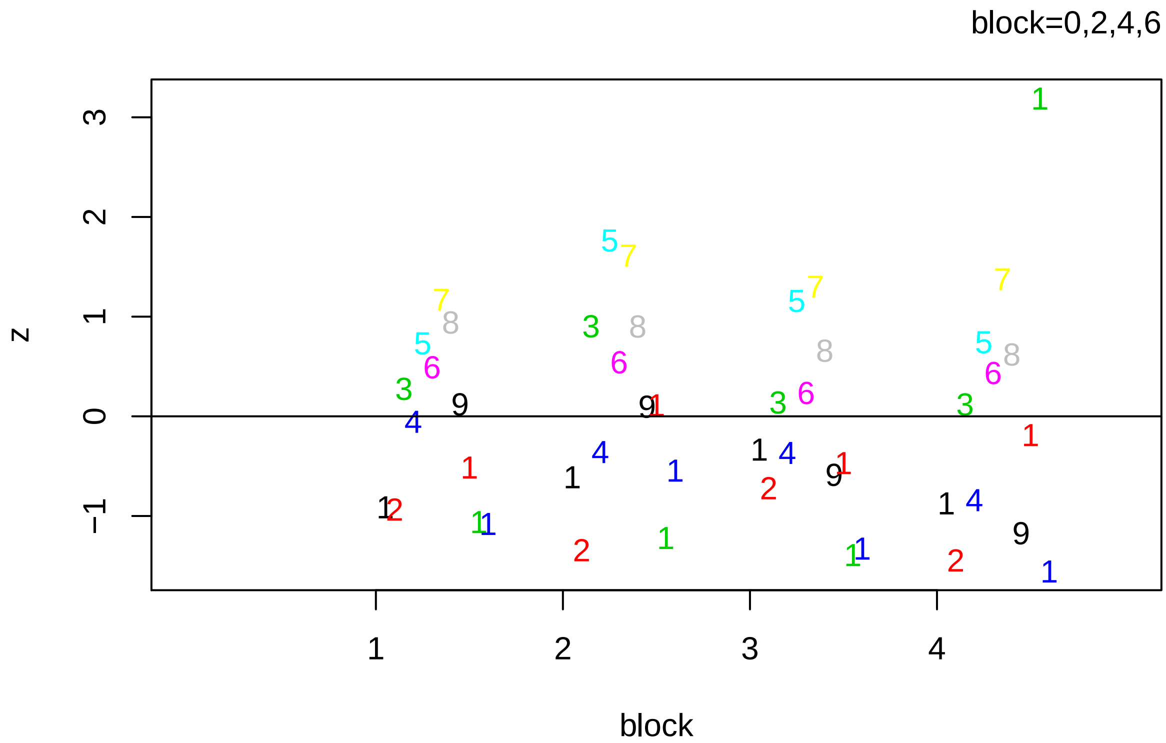
```r
# Generate residual plots
par(mfrow = c(1,3))
```

```r
plot(z ~ order, data=water.heating,
     main= "Order vs. Std. Residuals",
     ylab="Std. Residuals", las=1)
abline(h=0)
plot(z ~ ypred, data=water.heating,
      main= "Fitted Values vs. Std. Residuals",
     ylab="Std. Residuals", las=1)
abline(h=0)
qqnorm(water.heating$z)
# Line through 1st and 3rd quantile points
qqline(water.heating$z)
```



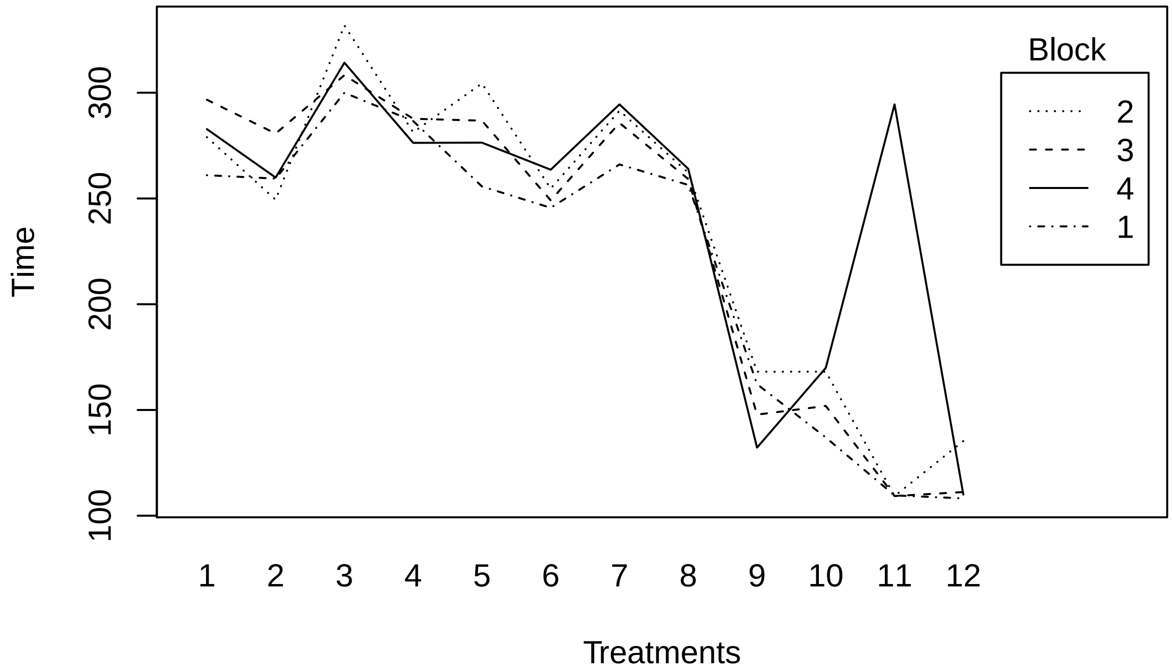
Assumption (a): The error have mean 0:

```r
attach(water.heating)
plot(z ~ block, xaxt = "n", type = "n", xlim = c(0, 5)) # Suppress x-axis, pts
axis(1, at = c(1, 2, 3,4))
points(x = as.numeric(block) + as.numeric(new.code) * 0.05, y = z,
pch = as.character(as.numeric(new.code)), col = as.numeric(new.code))
mtext("block=0,2,4,6", side = 3, adj = 1, line = 1)
abline(h = 0) # Horizontal line at zero
```




```
# Margin text, top-rt, line 1 abline(h = 0)

interaction.plot(x.factor = new.code, trace.factor = block,
                 response= time, xlab="Treatments", ylab="Time",
                 trace.label="Block", leg.bty="b", leg.bg="white")
```



Assumption (b): The error have constant variance:

By plotting the standardized residuals against the fitted values we can see that the spread of the standardized residuals, fluctuate as the fitted values increas. There also seems to be an outlier. The outlier may be of interest to inspect. Hence, we can say that the equal variance assumption is not satisfied.

Assumption (c): The error are normally distributed:

From the qq-plot above we see that the data is fairly straight. However, there is just one apparent outliers. Therefore, the normality assumption is fairly reasonable.

Assumption (d): The errors are independent:

From the plot of Order vs. Std. Residuals we can see that there isn't an apparent pattern as time increases. Therefore we feel comfortable to say that the assumption that errors are independent is approximately satisfies.

```
anova(two.way_model)
```

```
## Analysis of Variance Table
##
## Response: time
##           Df Sum Sq Mean Sq F value    Pr(>F)
## block      1   2739    2739  1.6562    0.2047
## trtmt      1 154492  154492 93.4106 1.496e-12 ***
## Residuals 45  74426    1654
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table above we can see that the p-value associated with the treatments is $<<$ .05. Therefore we can conclude that there is some effect on the heating time due to treatments.

(d)

Using the factorial form of the block–treatment model similar to (10.8.15), p. 325, but with three treatment factors, test the hypotheses of no interactions between pairs of treatment factors, each test done at level 0.01.

(In Part (d), the model they are asking for is still a block-treatment model (not a block-treatment interaction model) so please use the factorial version of the model you considered in Part (c) – so you should have a single term for block, but then all 2-factor and 3-factor interactions of Factors C,D and E. You only need to evaluate the 2-factor interactions.)

```
block.trtmt.model <- aov(time ~ block+ C*D*E, data = water.heating)
anova(block.trtmt.model)
```

```
## Analysis of Variance Table
##
## Response: time
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## block      1   2739    2739  1.7396   0.19488
## C          1 155473  155473 98.7406 3.061e-12 ***
## D          1    256     256  0.1624   0.68913
## E          1   6080    6080  3.8611   0.05657 .
## C:D        1   4211    4211  2.6746   0.11001
## C:E        1     82      82  0.0518   0.82111
## D:E        1    942     942  0.5980   0.44399
## C:D:E      1    467     467  0.2969   0.58896
## Residuals 39  61408    1575
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Notice from the table above that we have a single term for block and all 2-factor and 3-factor interactions of Factors C,D and E. Thus, at the $\alpha = .01$ level, we fail to reject test the hypotheses of no interactions between pairs of treatment factors, since the p-value for the interactions between C and D is .11. The p-value for the interactions between C and E is .82. And the p-value for the interactions between D and E is .44.

(e)

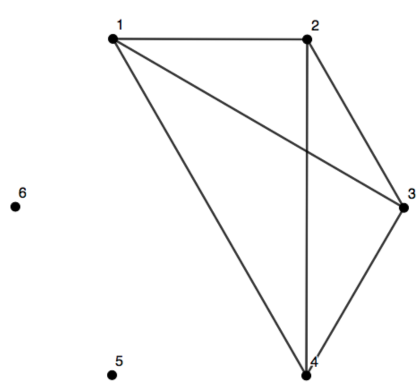
Taking into account any interactions discovered in part(d), list the contrasts that are of interest to you and, using the Scheffé method, calculate a set of 95% confidence intervals for the contrasts of interest.

(For Part (e), please provide Scheffé intervals for the two main effects contrasts in levels of C; you are certainly welcome to provide more intervals, but these are the ones that will be graded.)

Problem 2 (11.1 a-c)

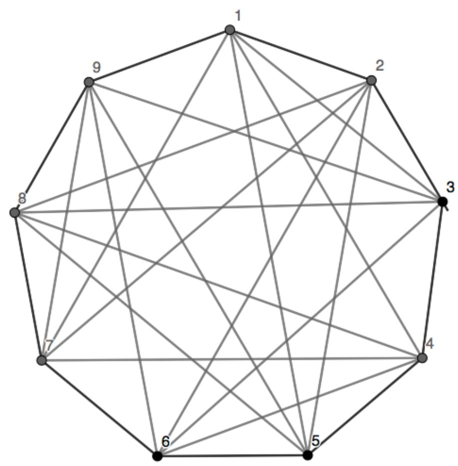
(a) For each of the three block designs in Table 11.27, draw the connectivity graph for the design, and determine whether the design is connected.

Block Design I



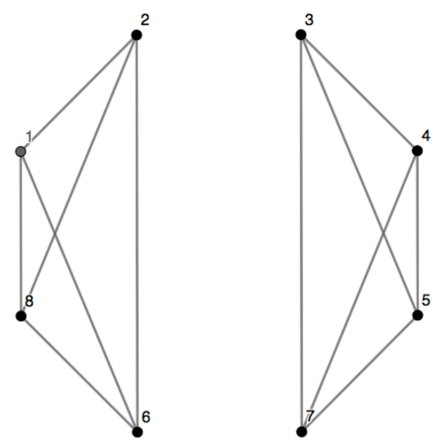
Block Design I is not connected.

Block Design II



Block Design II is connected.

Block Design III



Block Design III is not connected.

(b) If the design is connected, determine whether or not it is a balanced incomplete block design.

Design II is connected and we have that $v = 9$, $b = 9$, $k = 3$, and $r = 3$. However it is not a balanced imcomplete block design by the calculations below.

$$vr = 9(3) = bk \quad \checkmark$$
$$b = 9 = v \Longrightarrow b \geq v \quad \checkmark$$
$$\frac{r(k-1)}{v-1} = \frac{3(2)}{9-1} = \frac{6}{8} = \frac{3}{4} = \lambda \notin \mathbb{Z} \Longrightarrow r(k-1) \neq \lambda(v-1) \quad \times$$

(c) For designs II and III, determine graphically whether or not $\tau_1 - \tau_5$ and $\tau_1 - \tau_6$ are estimable.

solution

Since Block Design II is connected, then all contrasts in the treatment effects are estimable in the design. Hence $\tau_1 - \tau_5$ and $\tau_1 - \tau_6$ are estimable. Futhermore, we can see that there doesn't exist a path from 1 to 5, which means that $\tau_1 - \tau_5$ is not estimable. However there is a path from 1 to 6, so $\tau_1 - \tau_6$ is estimable.

Problem 3 (11.5)

In the following questions, consider an experiment to compare $v = 7$ treatments in blocks of size $k = 5$.

(a) Show that, for this experiment, a necessary condition for a balanced incomplete block design to exist is that r is a multiple of 5 and b is a multiple of 7.

Suppose that $b = 7n$ for some $n \in \mathbb{N}$ and $r = 5m$ for some $m \in \mathbb{N}$.

Now we must check the three conditions needed to be an incomplete block design:

$$vr = 7(5m) = 35m \text{ and } bk = (7n)5 = 35n \implies \text{ if } n = m \quad \checkmark$$
$$b = 7n \text{ and } v = 7 \implies b \geq v \text{ if } n \geq 1 \quad \checkmark$$
$$\tfrac{r(k-1)}{v-1} = \tfrac{5m(4)}{6} = \tfrac{20m}{6} = \tfrac{10m}{3} \implies \lambda \in \mathbb{Z} \text{ if } m = 3\ell \qquad \checkmark$$

So we can conclude that a necessary condition for a balanced incomplete block design for this experiment to exist is that r is a multiple of 5 and b is a multiple of 7.

(b) Show that r must be at least 15.

From part (a) we see that that m must be a multiple of 3 in order for $\lambda \in \mathbb{Z}$. So we must have that $r5(3\ell) = 15\ell$ where $\ell \in \mathbb{Z}$ So r must be a multiple of 15. Next we must show that $\ell \in \mathbb{N}$. Notice again from part (a) that it must me true that $n \geq 1$ and $m = n$. So m must be psotive and therefore $\ell \in \mathbb{N}$. Hence r must be at least 15.

(c) Taking all possible combinations of five treatments from seven gives a balanced incomplete block design with $r = 15$. Calculate the number of blocks that must be in this design.

Since we are still assuming $k = 5$, then we must have that $vr = bk$, so $5(15) = b(5)$. Then $b = 5$. So this design must have 5 blocks.