Homework 2 - 6413A

Peter Williams

2.10) *In order to analyze possible differences between five treaments, a one-way experiment layout was carried out. Each of the treatments was tested on three machines, resulting in a total of 15 experimental runs. After fitting the one-way model (2.1) (which has no block effect) to the data, ther residuals were plotted against machine number, as shown in Figure 2.9.* What do you learn from the plot? How would you modify your model and analysis?

The plot reveals information about possible model violations. Plot 2.8 suggests that property (c), $r \sim MN(0, \sigma^2(I - H))$ is violated. If the model were true, the points should be centered about zero in a somewhat parallel band. However, in the plot, the residuals for Machine 2 are centered closer to -13, revealing the tendency of the model to over-estimate treatments on Machine 2.

Assuming the experiment can not be conducted again using a block design (where treatments are randomly assigned to units within blocks or machines), a more complex modeling approach such as a two-way layout or regression modeling approach could help the researcher understand the effect of each treatment, taking into account potential differences in measurement taken between machines.

2.11) *The bioactivity of four different drugs A, B, C, and D for treating a particular illness were compared in a study and the following ANOVA table was given for the data.*

(a) Describe a proper design of the experiment to allow valid inferences to be made from the data.

Assuming the goal of the experiment is to determine whether there are differences in bioactivity levels for the different drugs, a one-way layout with fixed effects could be suitable.

The treatment is also the main factor level of interest with k=4 levels (referring to drugs (A,B,C,D)).

A proper design would also include:

1. Effective blocking of batches of drugs if applicable
2. Randomization applied to unit selection if drug batch blocking cannot be accomplished
3. Balancing of sample sizes across treatment levels
4. Sampling size planning to ensure the number of replications employed have the needed statistical power

(b) Use an F test to test at the 0.01 level the null hypothesis that the four treaments have the same biaoactivity. Compute the p value of the observed F statistic.

Null hypothesis (no difference between 4 treatments (A,B,C,D)): $H_0 : \tau_1 = ... = \tau_4$,

F = MSTr / MSE = (64.42/3) / (62.12/26) = 21.47 / 2.39

```
FObs <- (64.42/3) / (62.12/26)
print(paste0('F statistic = ',FObs))
```

```
## [1] "F statistic = 8.98755097660442"
```

```
df1 <- 3; df2 <- 26;
pValue <- 1 - pf(FObs,df1=df1, df2 = df2)
```

```
## [1] "p-value of 0.000296730851043536, is < than 0.01, thus, by the F test,  H0 is rejected."
```

The F test suggests that there is a difference between treatment levels at $\alpha = 0.01$.

(c) The treatment average as follows: $\bar{y}_A = 66.10$ (7 samples), $\bar{y}_B = 65.75$ (8 samples), $\bar{y}_C = 62.63$ (9 samples), $\bar{y}_D = 63.85$ (6 samples). Use the Tukey method to perform multiple comparisons of the four treatments at the 0.01 level.

$$t_{ij} = \frac{\bar{y}_j - \bar{y}_i}{\hat{\sigma}\sqrt{(1/n_j + 1/n_i)}}$$

1. A vs. B

```
tstat = abs( (66.10 - 65.75) / sqrt(2.39) * sqrt(1/7 + 1/8) )
tcrit = (1/sqrt(2))*qtukey(p=0.99, nmeans=4, df = 26)
paste0('A different from B? ',tstat > tcrit)
```

```
## [1] "A different from B? FALSE"
```

2. B vs. C

```
tstat = abs( (65.75 - 62.63) / sqrt(2.39) * sqrt(1/8 + 1/9) )
tcrit = (1/sqrt(2))*qtukey(p=0.99, nmeans=4, df = 26)
paste0('B different from C? ',tstat > tcrit)
```

```
## [1] "B different from C? FALSE"
```

3. C vs. D

```
tstat = abs( (62.63 - 63.85) / sqrt(2.39) * sqrt(1/9 + 1/6) )
tcrit = (1/sqrt(2))*qtukey(p=0.99, nmeans=4, df = 26)
paste0('C different from D? ',tstat > tcrit)
```

```
## [1] "C different from D? FALSE"
```

4. A vs. C

```
tstat = abs( (66.10 - 62.62) / sqrt(2.39) * sqrt(1/7 + 1/9) )
tcrit = (1/sqrt(2))*qtukey(p=0.99, nmeans=4, df = 26)
paste0('A different from C? ',tstat > tcrit)
```

```
## [1] "A different from C? FALSE"
```

5. A vs. D

```
tstat = abs( (66.10 - 63.85) / sqrt(2.39) * sqrt(1/7 + 1/6) )
tcrit = (1/sqrt(2))*qtukey(p=0.99, nmeans=4, df = 26)
paste0('A different from D? ',tstat > tcrit)
```

```
## [1] "A different from D? FALSE"
```

6. B vs. D

```
tstat = abs( (65.75 - 63.85) / sqrt(2.39) * sqrt(1/8 + 1/6) )
tcrit = (1/sqrt(2))*qtukey(p=0.99, nmeans=4, df = 26)
paste0('A different from D? ',tstat > tcrit)
```

```
## [1] "A different from D? FALSE"
```

Under the Tukey multiple comparisons test, none of the treatments are found to be significantly different from each other at $\alpha = 0.01$.

Interestingly, the sensitivity of the ANOVA test detected lower variability within treatments, than between. But the Tukey test did not highlight differences between any pairwise means.

(d) It turns out that A and B are brand name drugs and C and D are generic drugs. to compare brand-name vs. generic drugs, the contrast $1/2(\bar{y}_A + \bar{y}_B) - 1/2(\bar{y}_C + \bar{y}_B)$ is computed. Obtain the p value of the computed contrast and test its significance at the 0.01 level. Comment on the difference between brand-name and generic drugs.

$$\hat{L} = 1/2(\bar{y}_A + \bar{y}_B) - 1/2(\bar{y}_C + \bar{y}_B) = \frac{1}{2}(66.10 + 65.75) + \frac{1}{2}(62.63 + 63.85)$$

$$\hat{T} = \frac{\hat{L}}{\hat{\sigma}\sqrt{(\frac{1}{2})^2(1/7 + 1/8 + 1/9 + 1/6)}}$$

```
options(scipen = 10)
MSE <- 2.39
contrastHat <- (1/2)*(66.10+65.75) - (1/2)*(62.63+63.85)
seContrast <- sqrt( MSE* sum( c((0.5^2)/7, (0.5^2)/8, (0.5^2)/9, (0.5^2)/6) ) ) )
tstat = contrastHat/ seContrast
pValue = 1 - pt(tstat, df = 26)
paste0('T-statistic = ', tstat)
```

```
## [1] "T-statistic = 4.70245075830375"
```

```
paste0('p-value of t-test is: ',pValue)
```

```
## [1] "p-value of t-test is: 0.0000368605649108122"
```

Based on the results of the t-test at $\alpha = 0.01$, the computed contrast of factor levels associated with brand names, vs. generic are significantly different.

2.17) *A new electronic device for measuring blood pressure is introduced in the the market. An experiment has been conducted to compare the precision (measurement variation) of measurement taken by the new device with those taken by doctors with existing devices. Three devices are randomly selected from the store and three doctors (with their ow existing devices) are randomly chosen from all available doctors. One patient's blood pressure had been monitored for 15 days. On each day, the blood pressure was read by all three devices and three doctors.*

A quick visualization of the data for both doctors and devices yields some insight into the data. Blood pressure measurments by doctor have similar scales of variation, but difference mean measurements. Blood pressure by device has similar scales of variation, and similar sample means across devices in the study.

Blood Pressure by Doctor
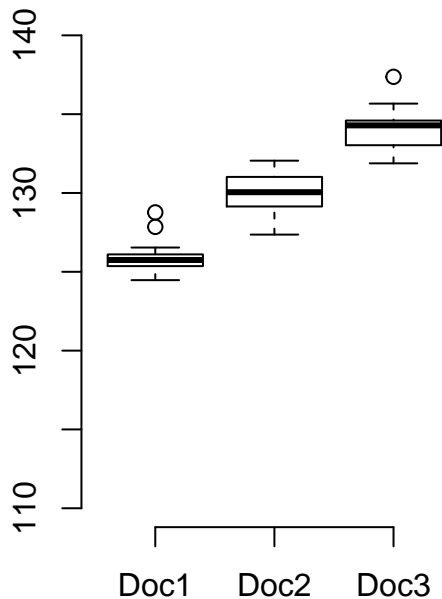
Blood Pressure by Device
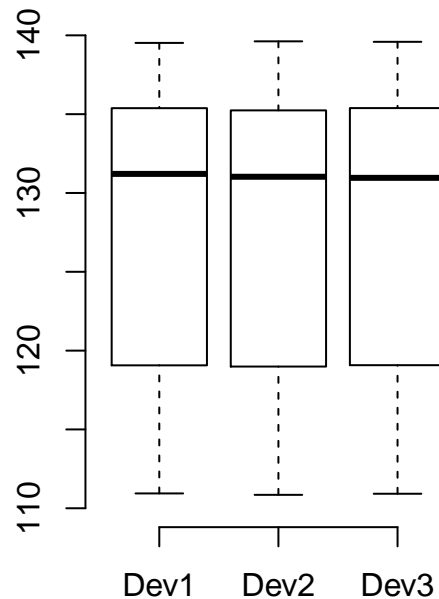
(a) Analyze the blood pressures measured by devices and by doctors separately using a one-way random effects model. Your work should include two ANOVA tables, one for devices and one for doctors. You should also include F tests and estimates of variance components.

Analysis of doctor data:

```
docMod <- summary(aov(value~Error(variable), data = docData))
print('Summary, ANOVA for doctors:')
```

```
## [1] "Summary, ANOVA for doctors:"
```

```
docMod
```

```
##
## Error: variable
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  2  496.3   248.2
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 42  74.93   1.784
```

```
#doctor data one-way random effects
MSTr <- docMod[[1]][[1]]$'Mean Sq'
MSE <- docMod[[2]][[1]]$'Mean Sq'
Fstat <- MSTr/MSE
Fcrit <- qf(0.95, df1 = 2, df2 = 42)
```

At an alpha level of 0.05, we compute an F-statistic for the treatment effect of 139.0972623, which is greater than the critical value of 3.2199423. Thus we reject the null hypothesis of no treatment effect by doctor, in favor the of the alternative that there is evidence of a treatment effect.

4

Using the results from the one-way random effects test, with n_i's being equal across groups, we compute $\sigma_\tau^2 = MSTr - MSE/15$, which equals: 16.4252328.

```r
devMod <- summary(aov(value~Error(variable), data = devData))
print('Summary, ANOVA for devices:')
```

```
## [1] "Summary, ANOVA for devices:"
```

```r
devMod
```

```
##
## Error: variable
##            Df  Sum Sq  Mean Sq F value Pr(>F)
## Residuals   2 0.01356 0.006782
##
## Error: Within
##            Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 42   3699   88.08
```

```r
#device data one-way random effects
MSTr <- devMod[[1]][[1]]$'Mean Sq'
MSE <- devMod[[2]][[1]]$'Mean Sq'
options(scipen=10)
Fvalue <- MSTr/MSE
Fcrit <- qf(0.95, df1 = 2, df2 = 42)
```

At an alpha level of 0.05, we compute an F-statistic for the treatment effect of 139.0972623, which is less than the critical value of 3.2199423. Thus we fail to reject the null hypothesis of no treatment effect by device.

Using the results from the one-way random effects test, with n_i's being equal across groups, we compute $\sigma_\tau^2 = MSTr - MSE/15$, which equals: -5.8716702. However, statistically, we don't have evidence that device to device mean measurement, or variation is different

(b) What can you conclude from (a)?

Based on the results of the one-way random effects models for doctors, (at an alpha level of 0.05), we find evidence different measurements of blood pressure from doctor to doctor. We find that doctor to doctor variance in blood pressure measurement is estimated to be more than 25% larger than day to day variance.

Based on the results of the one-way random effects models for devices, (at an alpha level of 0.05), we do not find evidence of different measurements of blood pressure from device to device. Variance in measurement across devices is similar, and the variance component of day to day measurement (residual level) is significantly larger than device to device variation.

(c) Find 95% confidence intervals for the mean blood pressure measured by devices and the mean blood pressure measured by doctors.

Here is a function to calculate the CI based on a t-interval:

```r
tinterval <- function(alpha, xbar,s, n){
  intervalEstimate <- (s/sqrt(n)) * qt(1-alpha, df = (n-1))
  paste0('[', round(xbar + intervalEstimate,digits = 3),',',
         round(xbar - intervalEstimate,digits=3),']')
}
```

For the doctor data, the computation above yields a 95% CI(t) of:

```
tinterval(alpha = 0.025, xbar = mean(docData$value),
          n = nrow(docData), s = sd(docData$value))
```

```
## [1] "[131.104,128.939]"
```

For the device data the CI computed is:

```
tinterval(alpha = 0.025, xbar = mean(devData$value),
          n = nrow(devData), s = sd(devData$value))
```

```
## [1] "[130.821,125.312]"
```

Computation of the CI of the mean for each the doctor and the device dataset yields a confidence interval that overlaps, highlighting that the mean of the pooled groups of measurements for both datasets is not different. However, visualization of the data, and the ANOVA yield insights that highlight that there are sources of variation of interest in measurement between doctors, and that measurements across devices are similar in spread and central tendency.