Homework #11

Robert Campbell

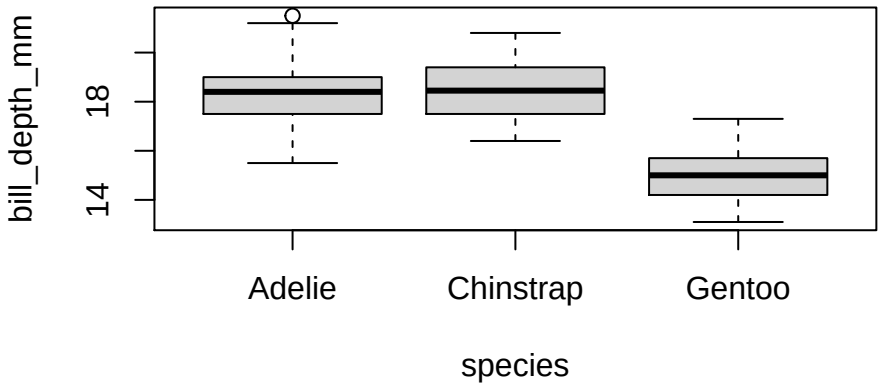
10 May 2021

Chapter 12

Problem 04

a. Looks like the data satisfies ANOVA assumptions

```
pen<-palmerpenguins::penguins
boxplot(bill_depth_mm ~ species, data=pen)
```



b. p-value < 2.2*(10^-16), showing significance at the alpha = 0.01 level. Reject the null, species does significantly predict bill depth.

```
depth_mod<-lm(bill_depth_mm ~ species, data=pen)
anova(depth_mod)
```

```
## Analysis of Variance Table
##
## Response: bill_depth_mm
##            Df Sum Sq Mean Sq F value    Pr(>F)
## species     2 903.97  451.98  359.79 < 2.2e-16 ***
## Residuals 339 425.87    1.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 05

P-value: 0.00362, reject the null. There are significant differences in the predictions of 300g objects by age group.

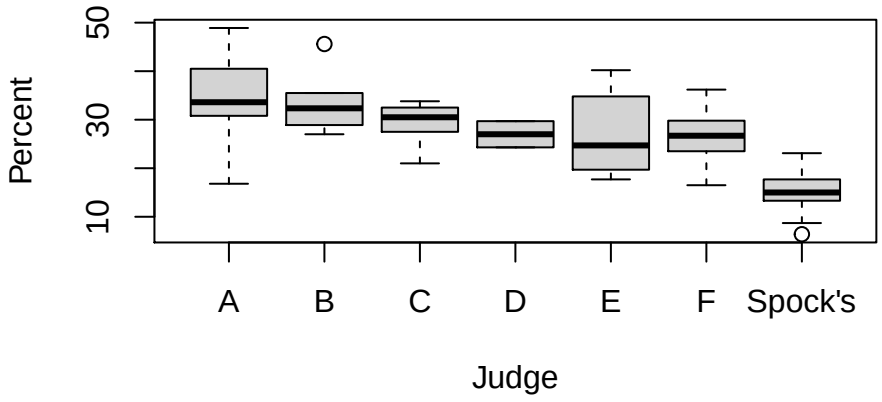
```
anova(lm(mean300 ~ age, data=fosdata::weight_estimate))
```

```
## Analysis of Variance Table
##
## Response: mean300
##           Df Sum Sq Mean Sq F value  Pr(>F)
## age        3  60786 20262.1  4.9018 0.00362 **
## Residuals 76 314155  4133.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 06

a. Spock's Lawyers may have a point, as Spock's mean percentage is visibly less than the means of the other 7 lawyers.

```
judge_perc <- Sleuth3::case0502
boxplot(Percent ~ Judge, data=judge_perc)
```



b. P-value: 0.3239, do not reject the null. The means of the percentage of women jurors in the 6 other judges courts are significantly similar.

```
judge_perc %>% filter(Judge != "Spock's") %>% lm(Percent ~ Judge, data=.) %>% anova()
```

```
## Analysis of Variance Table
##
## Response: Percent
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Judge      5  326.46  65.292  1.2183 0.3239
## Residuals 31 1661.33  53.591
```

c. The significantly similar means of the other judges justifies comparing Spock's judge to the other 6. P-value: 1.303 * (10^-6), reject the null. There is significant difference in the expected percentage of women jurors in Spock's judge's court when compared to the other judges.

```
judge_perc %>% mutate(Spock = (Judge=="Spock's")) %>%
  t.test(Percent ~ Spock, data=.)
```

```
##
##  Welch Two Sample t-test
##
## data:  Percent by Spock
## t = 7.1597, df = 17.608, p-value = 1.303e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  10.49935 19.23999
## sample estimates:
## mean in group FALSE  mean in group TRUE
##            29.49189            14.62222
```

Problem 08

The p-value: 5.93*(10^-10) looks to show significant difference in weight based on feed, however the blox-plot shows inconsistent variance in the different groups with the most extreme being when you compare "meatmeal" with "sunflower".

```
anova(lm(weight ~ feed, data=chickwts))
```

```
## Analysis of Variance Table
##
## Response: weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## feed       5 231129   46226  15.365 5.936e-10 ***
## Residuals 65 195556    3009
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
boxplot(weight ~ feed, data=chickwts)
```

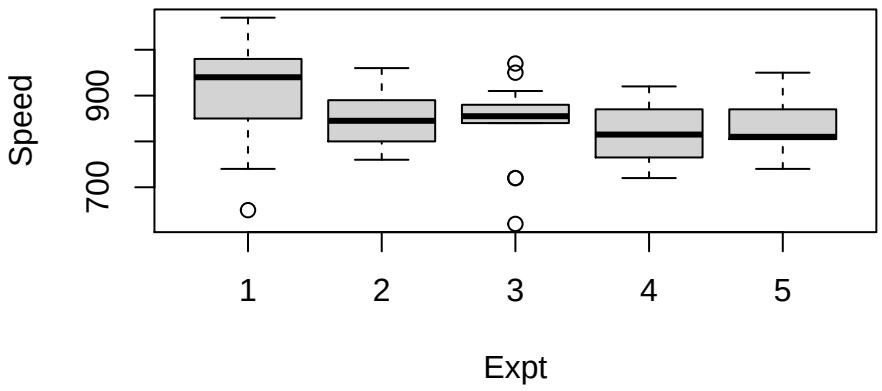
Problem 09

With P-value: 0.0004827, it looks like there is significant difference in the mean based on experiment number. However, looking at the boxplots the data does not satisfy the anova assumptions, especially experiments 3 and 5.

```
anova(lm(Speed ~ Expt, data=morley))
```

```
## Analysis of Variance Table
##
## Response: Speed
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Expt       1  72581   72581  13.041 0.0004827 ***
## Residuals 98 545444    5566
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

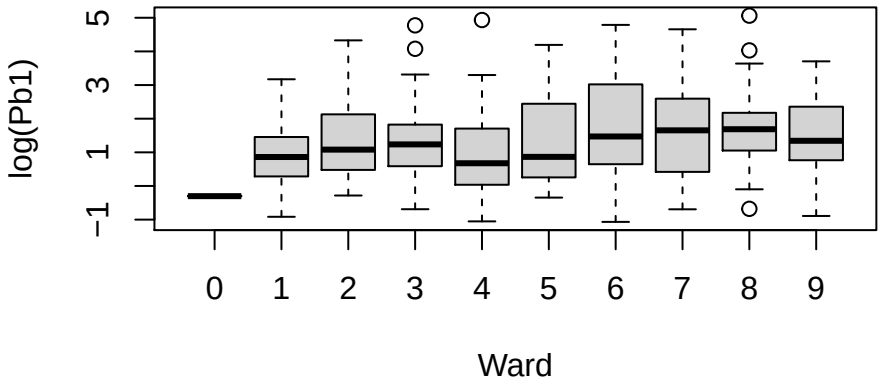
```
boxplot(Speed ~ Expt, data=morley)
```



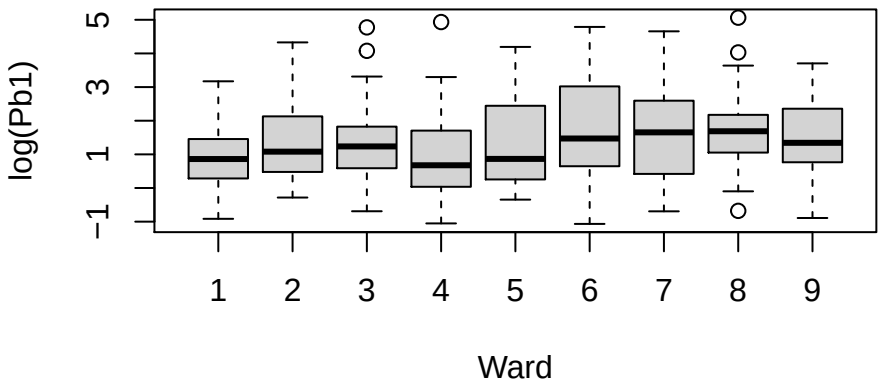
Problem 15

a.

```
flint_pb <- fosdata::flint
boxplot(log(Pb1) ~ Ward, data=flint_pb)
```




```
flint_pb<- flint_pb %>% filter(Ward != 0)
boxplot(log(Pb1) ~ Ward, data=flint_pb)
```



b. The log levels look normally distributed and the variances a similar enough to justify anova.
c. P-value: 0.03137, there is a significant difference in the mean of the log of the amount of lead in the water based on ward in Flint, MI.

```
anova(lm(log(Pb1) ~ Ward, data=flint_pb))
```

```
## Analysis of Variance Table
##
## Response: log(Pb1)
##            Df Sum Sq Mean Sq F value  Pr(>F)
## Ward        1   7.76  7.7563  4.6815 0.03137 *
## Residuals 268 444.03  1.6568
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```