STA 210: HW 2

Jackson Hubbard

September 20th, 2018

Question 1: Ex. #5.14

a

```
avg <- c(14.62,34.12,33.61,29.1,27,26.97,26.8)
sd <- c(5.039,11.942,6.582,4.593,3.818,9.010,5.969)
n <- c(9,5,6,9,2,6,9)
judges <- bind_cols(avg=avg,sd=sd,n=n)

judges <- judges %>% mutate(var2 = sd^2)
judges <- judges %>% mutate(numerator = (n-1)*var2)
judges <- judges %>% mutate(denominator = (n-1))


pooled_sd <- sqrt((sum(judges$numerator))/(sum(judges$denominator)))
df <- sum(judges$denominator)

pooled_sd
```

```
## [1] 6.914247
```

```
df
```

```
## [1] 39
```

b

t test for hypothesis that the Spock's mean is equal to mean for judge A $H_0 : \mu_S - \mu_A = 0$ $H_A : \mu_S - \mu_A \neq 0$

```
est <- 14.62- 34.12
se <- pooled_sd * sqrt((1/9)+(1/5))
t_stat <- est/se
t_stat
```

```
## [1] -5.05629
```

```r
pt(t_stat,39 ) #n-k = 39
```

```
## [1] 5.242132e-06
```

Looking at the output of the t-test, we see that the p-value is very small(.00000524). This means that there is sufficient evidence to reject the null hypothesis of Spock judge's mean being equal to judge A's mean.

Question 2: Ex. #5.17

```r
#values provided by book
dfw <- 24 #DF within
dft <- 31 #DF total
ssw <- 35088 #Sum of squares within
sst <- 70907 #Sum of squares total
```

```r
#fill in equations to calculate remaning values
dfb <-  dft- dfw #DF between
ssb <-  sst -ssw #Sum of squares between
msb <-  ssb / dfb #Mean square between
msw <-  ssw/ dfw #Mean square within
f_stat <- msb /msw #F -statistic
p_val <- 1- pf(f_stat,7, 24) # p-value
```

```r
# create each column of the ANOVA table
#Each column must have same number of rows,
#Use NA to hold a space for the blank parts of the ANOVA table
source <- c("Between Groups", "Within Groups", "Total")
df <- c(dfb, dfw,dft)
ss <- c(ssb, ssw, sst)
ms <- c(msb, msw,NA)
f.statistic <- c(f_stat, NA, NA)
p.value <- c(p_val,NA,NA)

# combine the columns to make a table called "anova"
anova <- bind_cols("Source"=source,"df"=df,"Sum of squares"=ss,
            "Mean square"=ms,"F-statistic"=f.statistic,"p-value"=p.value)

# print the table
kable(anova)
```

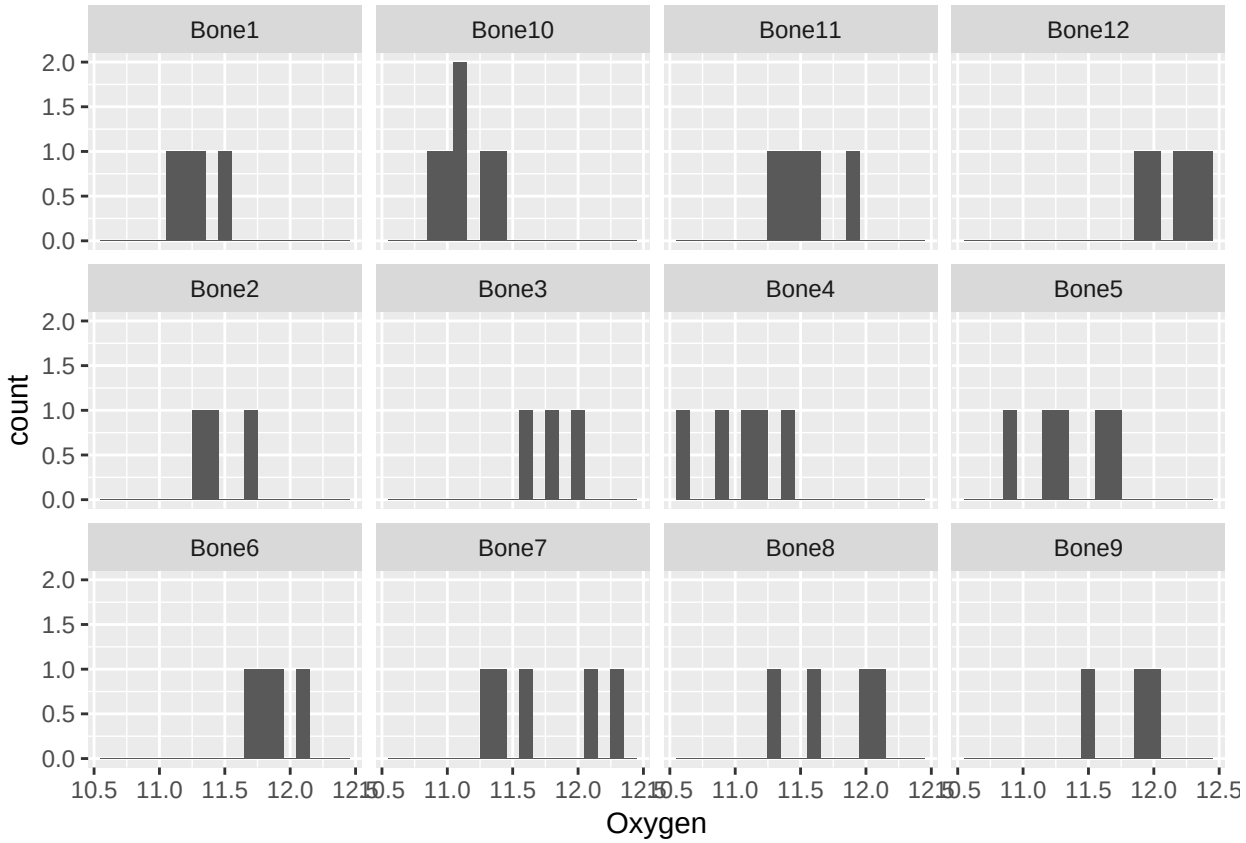
Source	df	Sum of squares	Mean square	F-statistic	p-value
Between Groups	7	35819	5117	3.5	0.0099418
Within Groups	24	35088	1462		

Source	df	Sum of squares	Mean square	F-statistic	p-value
Total	31	70907			

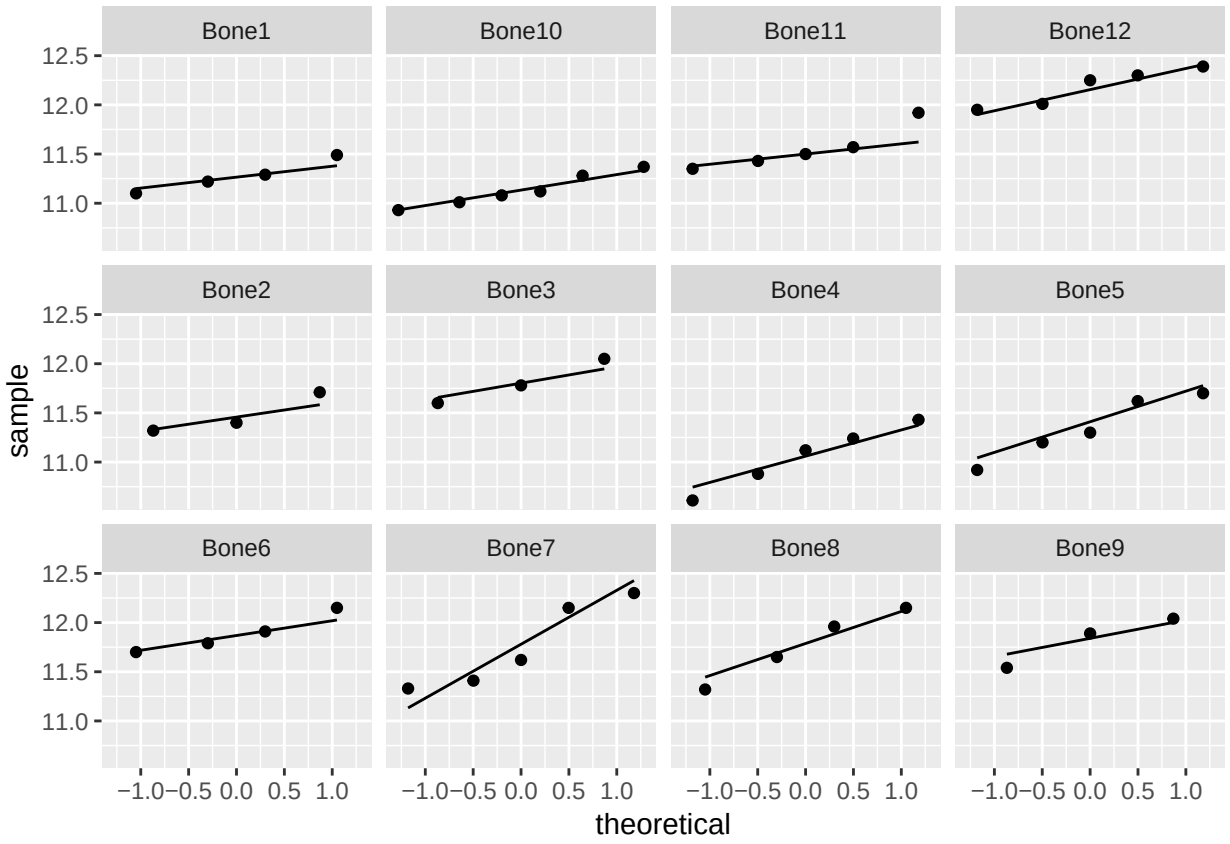
Question 3: Ex. #5.23

Check for assumpitons of ANOVA:

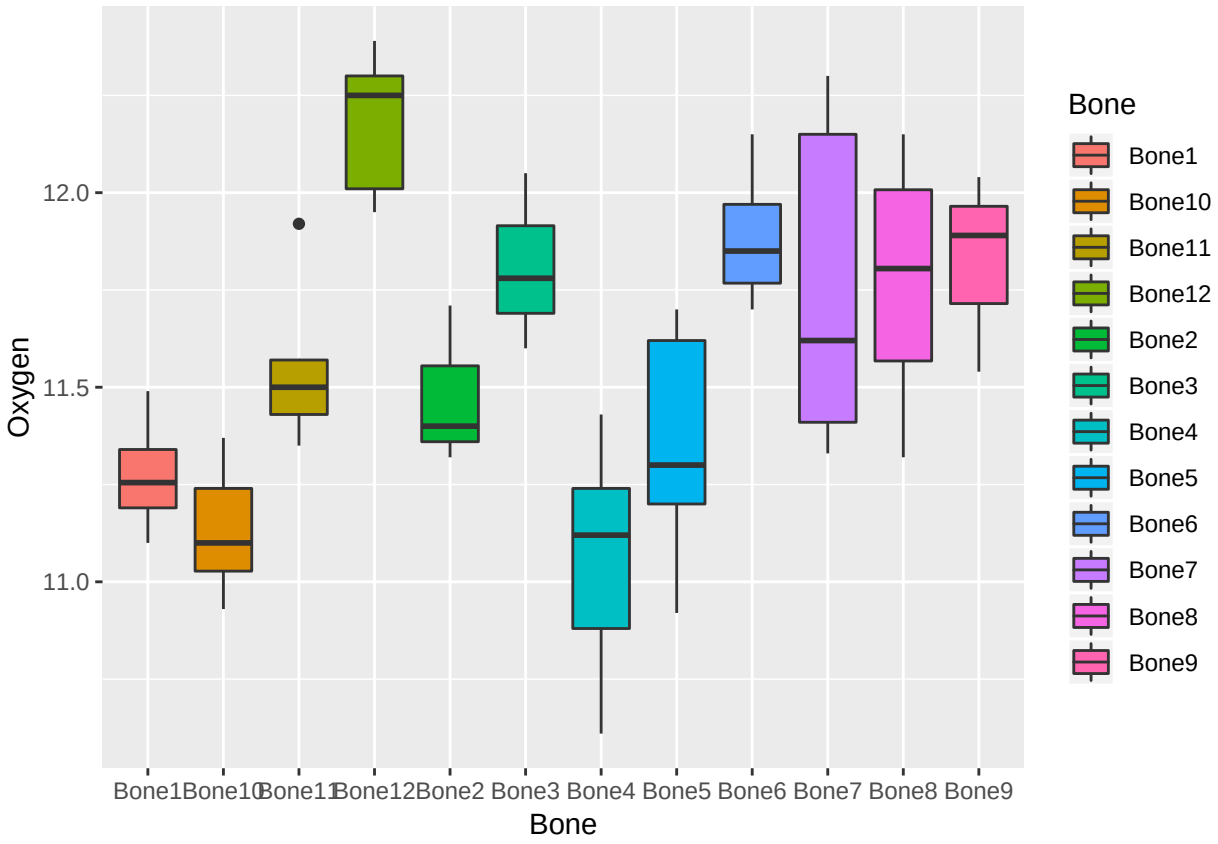
```
q3 <- ex0523
ggplot(q3, aes(x=Oxygen)) + geom_histogram(binwidth = 0.1)+ facet_wrap(Bone~.)
```




```
ggplot(q3 ,aes(sample = Oxygen)) +
  stat_qq() +
  stat_qq_line() +
  facet_wrap(.~Bone)
```




```
ggplot(q3, aes(Bone, Oxygen, fill= Bone)) + geom_boxplot()
```



First, check for normality. Looking at histograms, we see that the sample size is very small for each type of bone. Thus, it is hard to conclude much from these graphs. Looking at the qq plots for each bone type, it appears that there are no major departures from the line, but it is hard to conclude anything since the sample size is so small. Also, after examining the boxplots, we see that Bone 11 has an outlier on its upper bound.

```
q3 %>% group_by(Bone) %>% summarise(n=n(), mean=mean(Oxygen), var=var(Oxygen))
```

```
## # A tibble: 12 x 4
##    Bone       n  mean    var
##    <fct>  <int> <dbl>  <dbl>
##  1 Bone1      4  11.3 0.0267
##  2 Bone10     6  11.1 0.0274
##  3 Bone11     5  11.6 0.0485
##  4 Bone12     5  12.2 0.0363
##  5 Bone2      3  11.5 0.0424
##  6 Bone3      3  11.8 0.0513
##  7 Bone4      5  11.1 0.102
##  8 Bone5      5  11.3 0.101
##  9 Bone6      4  11.9 0.0380
## 10 Bone7      5  11.8 0.193
## 11 Bone8      4  11.8 0.132
## 12 Bone9      3  11.8 0.0658
```

Next check for constant variance. The majority of the variances of each bone group are in the range 0.03-0.05. However, there are a couple of outliers- Bone4 and 5 with variances of about 0.10 and Bone7 with a variance of 0.19. This means that the assumption of constant variance is broken. To check independence, we can think about how the data was collected. The oxygen isotopic composition of each bone was measued for only that bone so it is unlikely that one measurement would have an effect on another. However, since the bones came from the same organism, the oxygen levels may not actually be independent.

Despite hesitations about the vailidty of ANOVA's assumptions, I will continue with my analysis.

mu_i is the true mean of the oxygen isotropic composition in the ith bone ($1 \leq$ i ≤ 12)

$H_0 : mu_1 = mu_2 = mu_3 = ... = mu_12$ H_A : at least 1 μ is different

```
anova <- aov(Oxygen ~ Bone, data = q3)
kable(tidy(anova))
```

term	df	sumsq	meansq	statistic	p.value
Bone	11	6.067491	0.5515901	7.426843	1e-06
Residuals	40	2.970792	0.0742698		

Looking at the results of ANOVA, the p value (0.000001) is very small. Therefore, there is sufficient evidence to reject the null hypothesis of the means of al the different bone types are the same. There is sufficient evidence that at least one of the 12 bone specimens has a different mean oxygen isotopic composition than the other bone specimens. However, we must remember that some of the assumptions of ANOVA were broken, so this interpretation must not be taken literally.