ST 411/511 Homework 6

Due on February 27

Vijay Tadimeti

Winter 2019

Instructions

This assignment is due by 11:59 PM, February 27, 2019 on Canvas.

You should submit your assignment as either a PDF or Word document, which you can compile (should you choose – recommended) from the provided .Rmd (R Markdown) template. Please include your code.

Problems (25 points total)

Question 1

The table below shows a partially completed ANOVA table. (Note: if you are looking at this in RStudio it may be helpful to knit the file to properly view the table.)

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-statistic	p-value
Between Groups	35819	7	5117	3.5	0.009941808
Within Groups	35088	24	1462		
Total	70907	31			

(a) (1 point) How many groups were there?

```
I <- (31-24) + 1
I
```

```
## [1] 8
```

(b) (4 points) Fill in the rest of the table. Values to be calculated are indicated by a "."

```
SSW = 35088
SST = 70907
SSB=SST-SSW
SSB
```

```
## [1] 35819
```

```
Degrees_of_freedom_within_groups = 24
total = 31
Degrees_of_freedom_between_groups = total - Degrees_of_freedom_within_groups
Degrees_of_freedom_between_groups
```

```
## [1] 7
```

```
#n - I = 24
#n - 1 = 31
n <- 32

#n - I  = 24
#32 - I = 24
I <- 8


MSW = SSW/(n - I)
MSW
```

```
## [1] 1462
```

```
MSB = SSB/(I - 1)
MSB
```

```
## [1] 5117
```

```
F = MSB/MSW
F
```

```
## [1] 3.5
```

```
1-pf(3.5,df1 = 7,df2 = 24)
```

```
## [1] 0.009941808
```

(c) (2 points) What is your conclusion from the one-way ANOVA analysis? State the hypothesis you are testing and what your decision/strength of evidence are.

Ho : $\mu 1 = \mu 1 = \mu 1 = \mu 1 = \mu 1 = \mu 1 = \mu 1 = \mu 1 =$ HA : Atleast one of the population mean μj is not equal to the others

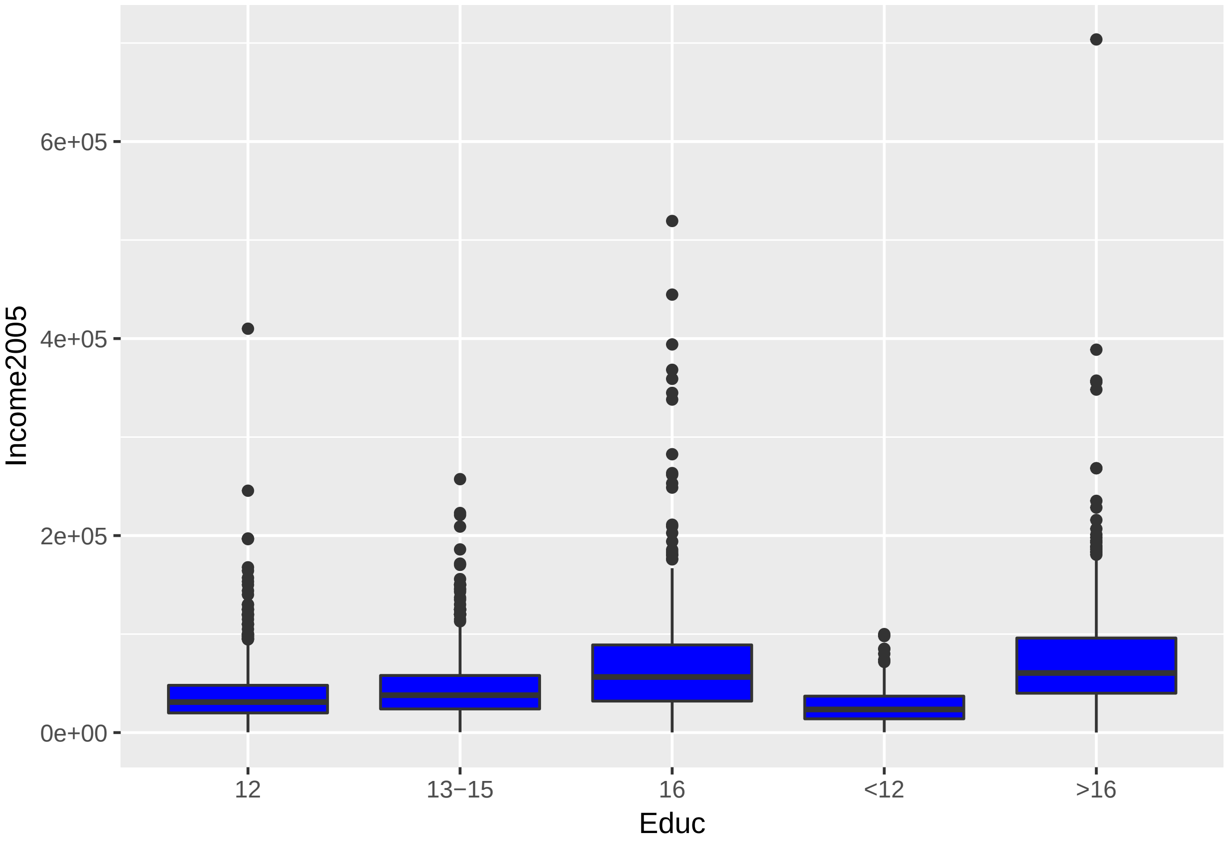
Here ,we reject the null hypothesis in favour of the (two sided) alternative hypothesis that atleast one of the population is not equal to others since p value is less than alpha = 0.1

Question 2 (Modified from *Sleuth* 5.25)

The data file `ex0525` contains annual incomes in 2005 of a random sample of 2584 Americans who were selected for the National Longitudinal Survey of Youth in 1979 and who had paying jobs in 2005. The data set also includes a code for the number of years of education that each individual had completed by 2006: <12, 12, 13-15, 16, and >16. Perform an analysis of variance *by hand* (i.e. not using the built-in anova functions like `lm()` and `anova()`) to assess whether or not the population mean 2005 incomes were the same in all five education groups. Work through the following steps:

(a) (1 point) Create a side-by-side boxplot of 2005 income grouped by education category.

```
ggplot(ex0525,aes(x=Educ,y=Income2005)) + geom_boxplot(fill = "blue")
```



(b) (2 points) Find the grand mean and the mean of each of the five education groups.

```
GM <- mean(ex0525$Income2005)
GM
```

```
## [1] 49417
```

```
ID_11 <- which(ex0525$Educ=="<12")
ID_12 <- which(ex0525$Educ=="12")
ID_13 <- which(ex0525$Educ=="13-15")
ID_14 <- which(ex0525$Educ=="16")
ID_15 <- which(ex0525$Educ==">16")
Mean_1 <- mean(ex0525$Income2005[ID_11])
Mean_2 <- mean(ex0525$Income2005[ID_12])
Mean_3 <- mean(ex0525$Income2005[ID_13])
Mean_4 <- mean(ex0525$Income2005[ID_14])
Mean_5 <- mean(ex0525$Income2005[ID_15])
Mean_1
```

```
## [1] 28301.45
```

```
Mean_2
```

```
## [1] 36864.9
```

```
Mean_3
```

```
## [1] 44875.96
```

```
Mean_4
```

```
## [1] 69996.97
```

```
Mean_5
```

```
## [1] 76855.46
```

(c) (2 points) Find the sums of squares between and within groups.

```
SSW <-
  sum((ex0525$Income2005[ID_11]-Mean_1)^2) +
  sum((ex0525$Income2005[ID_12]-Mean_2)^2) +
  sum((ex0525$Income2005[ID_13]-Mean_3)^2) +
  sum((ex0525$Income2005[ID_14]-Mean_4)^2) +
  sum((ex0525$Income2005[ID_15]-Mean_5)^2)
SSW
```

```
## [1] 4.951743e+12
```

```
SST <- sum((ex0525$Income2005 - GM)^2)
SST
```

```
## [1] 5.639978e+12
```

```
SSB <- SST-SSW
SSB
```

```
## [1] 688235137516
```

(d) (1 point) Find the mean squares between and within groups.

```
n<- nrow(ex0525)
I <- 5
dfW <- n - I
dfT <- n - 1
dfB <- I - 1

MSW <- SSW/dfW
MSB <- SSB/dfB

MSW
```

```
## [1] 1920024320
```

```
MSB
```

```
## [1] 172058784379
```

(e) (1 point) Find the F-statistic and p-value.

```
F<- MSB/MSW
F
```

```
## [1] 89.61282
```

```r
P<- 1-pf(F, df1=dfB, df2=dfW)
P
```

```
## [1] 0
```

(f) (1 point) State the conclusion of your test.

Here we reject the null hypothesis that the mean of all the five education group is zero, infavour of the (two sided)alternate hypothesis that atleast one mean is not equal to the rest of the group since our value is at: p < alpha = 0.01.

(g) (1 point) We can also state things we have calculated in the model testing framework. You should not need to calculate anything new for this part. What is the extra sum of squares? What is the pooled variance?

```r
ESS <- SST-SSW
ESS
```

```
## [1] 688235137516
```

```r
sp<- MSW
sp
```

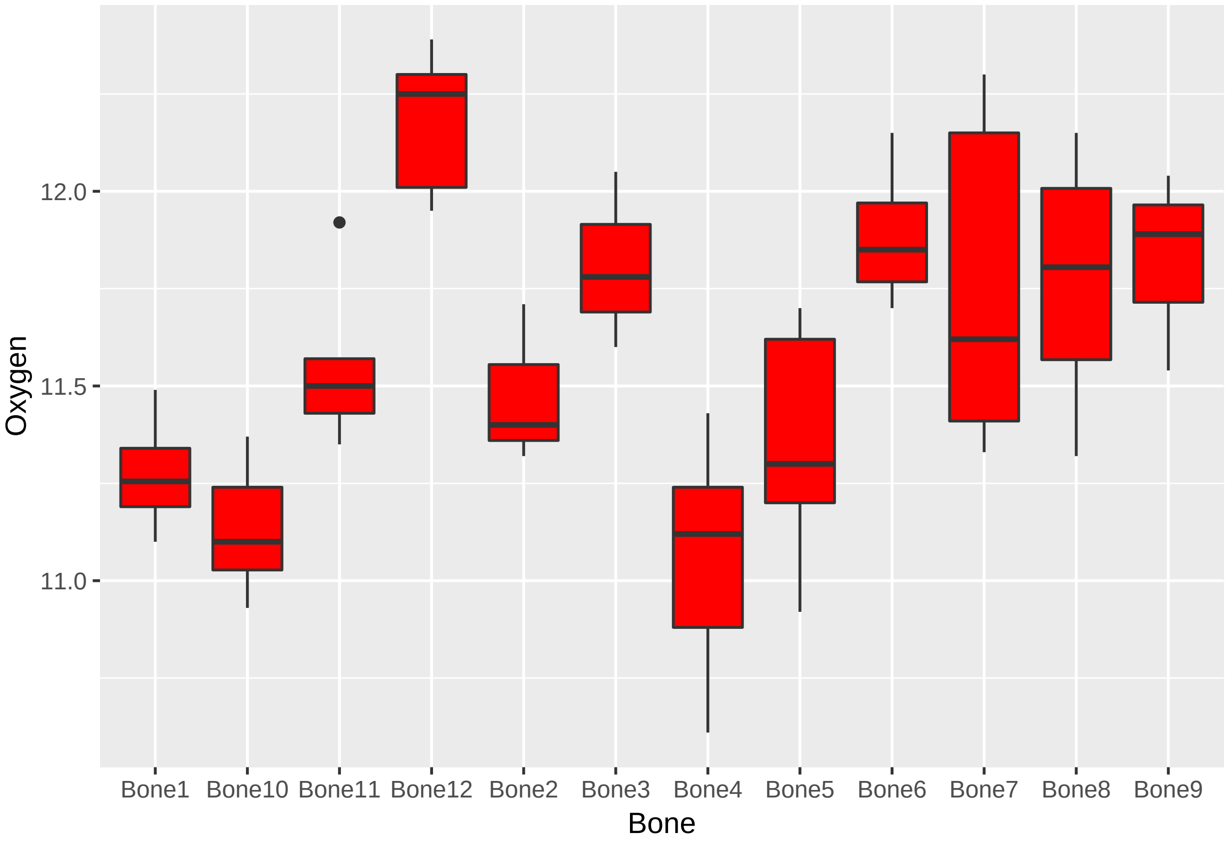
```
## [1] 1920024320
```

Question 3 (Modified from *Sleuth* 5.23)

Was Tyrannosaurus Rex warm-blooded? Several measurements of the oxygen isotopic composition of bone phosphate in each of 12 bone specimens from a single *Tyrannosaurus rex* skeleton were taken. It is known that the oxygen isotopic composition of vertebrate bone phosphate is related to the body temperature at which the bone forms. Differences in means at different bone sites would indicate nonconstant temperatures throughout the body. Minor temperature differences would be expected in warm-blooded animals. Is there evidence that the means are different for the different bones? The data are in `ex0523` in the `Sleuth3` library.

(a) (2 points) Plot the oxygen isotopic composition for each of the bones using a side-by-side boxplot. Comment on whether or not you think the population means are the same for all 12 bones based on your plot.

```r
ggplot(ex0523,aes(x=Bone,y=Oxygen)) + geom_boxplot(fill = "red")
```



The population means are not same for all the 12 bones based on the plot. We observe that there is a difference between the means of bone 12 and 4.

(b) (2 points) Perform an analysis of variance to test whether or not all the population mean oxygen isotopic compositions are the same in the 12 bone types. State your *p*-value and conclusion of the test. You may use the built-in ANOVA functions in R.

```
anova(lm(Oxygen ~ Bone, data= ex0523))
```

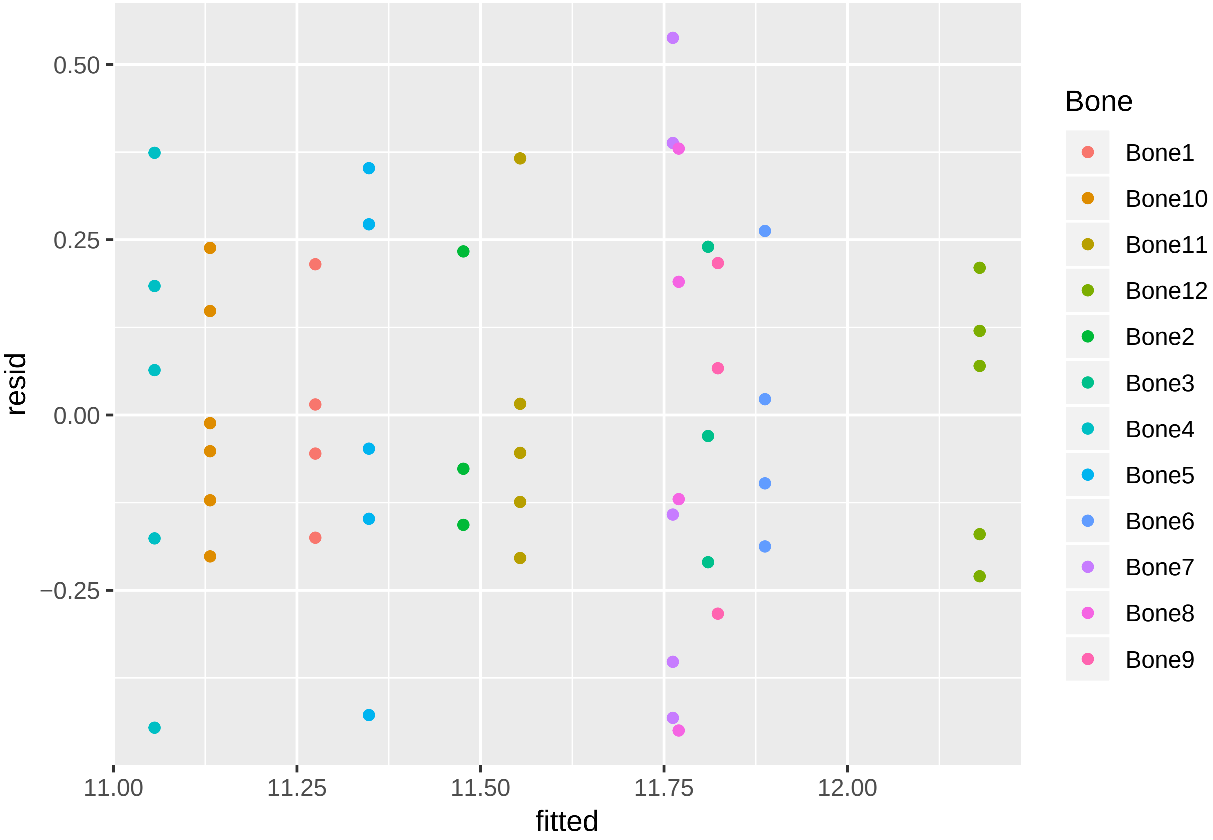
```
## Analysis of Variance Table
##
## Response: Oxygen
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Bone      11 6.0675 0.55159  7.4268 9.73e-07 ***
## Residuals 40 2.9708 0.07427
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here we observe that : P value is 9.73e-07. Hence we see that it is less that alpha = 0.1 we reject the null hypothesis in favour of the alternate hypothesis that atleast one mean oxygen isotopic composition is not same.

(c) (2 points) Assess the assumption that the population variances are the same in each group by creating a diagnostic plot using the residuals. Does this assumption appear to have been met?

```
data(ex0523)
mod = lm(Oxygen ~ Bone, data = ex0523)
ex0523$fitted = mod$fitted
ex0523$resid = mod$resid

ggplot(ex0523, aes(x=fitted, y=resid, color = Bone)) + geom_point()
```



The population variance here looks similar except for some outliers in bone 4 and 7.

(d) (3 points) Perform a Kruskal-Wallis test using the `kruskal.test()` function. What do you conclude from this test? Compare your conclusion with your result from the analysis of variance in part (b).

```
kruskal.test(Oxygen ~ Bone, data = ex0523)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Oxygen by Bone
## Kruskal-Wallis chi-squared = 34.938, df = 11, p-value = 0.0002537
```

Here we reject the null hypothesis in favour of the alternate hypothesis that atleast one mean of the oxygen isotopic composition is not same since p value is less that alpha = 0.1. We notice that Both the conclusions are the same but the p value and test statistics values are different. It is beacause Kruskal-Wallis test performs really good even when the residual plots indicate a problem with assumptions.