STOR 455 Homework #1

20 points - Due Tuesday 09/12 at 12:30pm

Directions: This first assignment is meant to be a brief introduction to working with R in RStudio. You may (and should) collaborate with other students. However, you must complete the assignment by yourself. You should complete this assignment in an R Notebook, including all calculations, plots, and explanations. Make use of the white space outside of the R chunks for your explanations rather than using comments inside of the chunks. For your submission, you should knit the notebook to PDF (it is usually smoother first knit to Word then save the file as pdf) and submit the file to Gradescope. The submitted PDF should not be longer than 20 pages.

Eastern Box Turtles: The Box Turtle Connection is a long-term study anticipating at least 100 years of data collection on box turtles. Their purpose is to learn more about the status and trends in box turtle populations, identify threats, and develop strategies for long-term conservation of the species. Eastern Box Turtle populations are in decline in North Carolina and while they are recognized as a threatened species by the International Union for Conservation of Nature, the turtles have no protection in North Carolina. There are currently more than 30 active research study sites across the state of North Carolina. Turtles are weighed, measured, photographed, and permanently marked. These data, along with voucher photos (photos that document sightings), are then entered into centralized database managed by the NC Wildlife Resources Commission. The *Turtles* dataset (found under "Resources" on Sakai) contains data collected at The Piedmont Wildlife Center in Durham.

Questions

Q1

Write code that imports *Turtles.csv* into R and save that data into a data frame named **Turtles**. Print the first 6 rows of **Turtles**.

```r
library(readr)
#
setwd("C:/Users/Jabbir Ahmed/Desktop/Data")
Turtles <- read.csv("Turtles.csv")
head(Turtles)
```

```
##   LifeStage     Sex Annuli Mass StraightlineCL  MaxCW PL_AnteriortoHinge
## 1     Adult    Male     13  410         127.00 102.00              48.00
## 2     Adult    Male     19  340         113.62  93.96              44.87
## 3  Juvenile  Female      7  160          89.49  73.51              39.60
## 4     Adult    Male     16  175         127.70 101.16              54.76
## 5  Juvenile  Female      7  100          81.00  69.00              35.00
## 6     Adult Unknown     17  410         127.32 101.21              56.70
##   PL_HingetoPosterior ShellHeightatHinge
## 1               68.00              61.00
## 2               67.61              55.88
## 3               53.65              43.48
```

```
## 4                84.72                61.97
## 5                44.00                39.00
## 6                81.42                64.24
```

Q2

Create a data frame named **Adult_Turtles** that contains only the adult turtles. Print the first 6 rows of **Adult_Turtles**.

```
#
Adult_Turtles <- (subset(Turtles, LifeStage == "Adult"))
head(Adult_Turtles)
```

```
##   LifeStage     Sex Annuli Mass StraightlineCL  MaxCW PL_AnteriortoHinge
## 1     Adult    Male     13  410         127.00 102.00              48.00
## 2     Adult    Male     19  340         113.62  93.96              44.87
## 4     Adult    Male     16  175         127.70 101.16              54.76
## 6     Adult Unknown     17  410         127.32 101.21              56.70
## 7     Adult  Female     18  472         131.00 104.00              49.00
## 8     Adult  Female     20  155         122.85  99.38              51.68
##   PL_HingetoPosterior ShellHeightatHinge
## 1               68.00              61.00
## 2               67.61              55.88
## 4               84.72              61.97
## 6               81.42              64.24
## 7               80.00              59.00
## 8               74.73              64.60
```

Q3

For the adult turtles, calculate the sample average **Mass** for males and the sample average **Mass** for females. Perform a two-sample test to check if **Sex** is useful in predicting **Mass**.

```
#
mean(Adult_Turtles$Mass[Adult_Turtles$Sex == "Male"], na.rm = TRUE)
```

```
## [1] 344.5429
```

```
mean(Adult_Turtles$Mass[Adult_Turtles$Sex == "Female"], na.rm = TRUE)
```

```
## [1] 345.5612
```

```
Adult_twogender_Turtles <- (subset(Adult_Turtles, Sex %in% c("Male", "Female")))
t.test(Mass~Sex, data = Adult_twogender_Turtles)
```

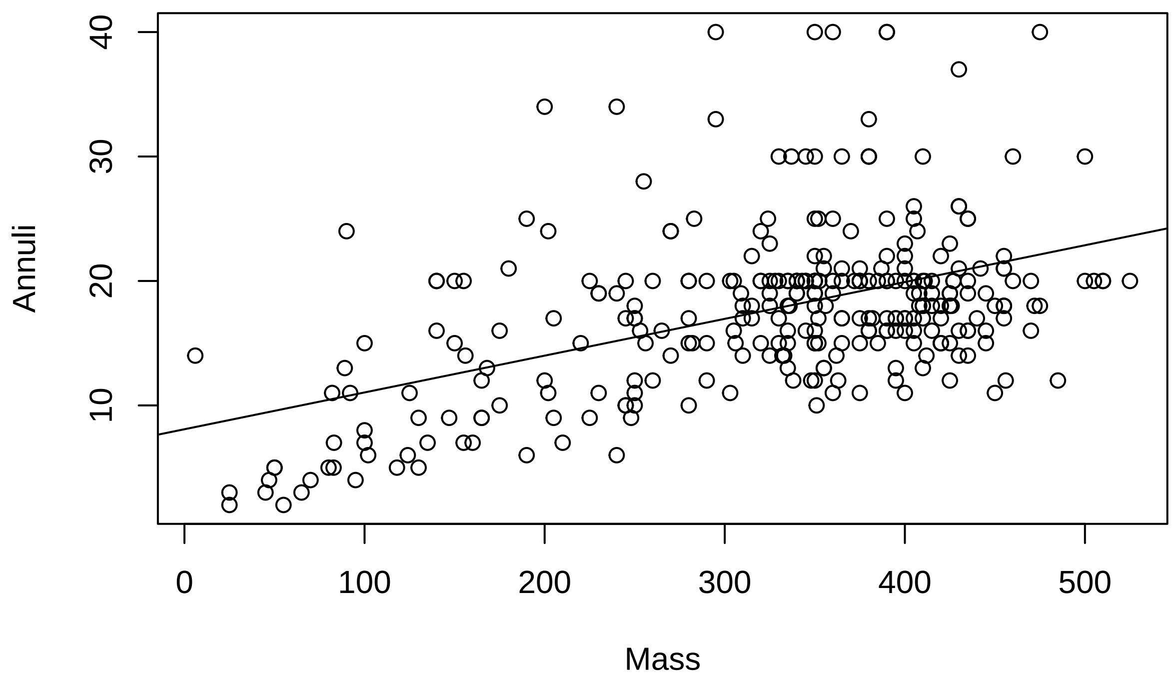
```
##
##  Welch Two Sample t-test
##
## data:  Mass by Sex
## t = 0.083763, df = 165.28, p-value = 0.9333
```

```
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -22.98420  25.02076
## sample estimates:
## mean in group Female   mean in group Male
##             345.5612             344.5429
```

Q4

The **Annuli** rings on a turtle represent growth on the scutes of the carapace and plastron. In the past, it was thought that annuli corresponded to age, but recent findings suggest that this is not the case. However, the annuli are still counted since it may yield important life history information. Construct a least squares regression line that predicts turtles' **Annuli** by their **Mass**. Produce a scatterplot of this relationship (and include the least squares line on the plot). Use the full dataset for this question and the questions after.

```
#
mod1 = lm(Annuli~Mass, data = Turtles)
plot(Annuli~Mass, data = Turtles)
abline(mod1)
```



Q5

The turtle in the first row of the **Turtles** dataset has a mass of 410 grams. What does your model predict for this turtle's number of **Annuli**? What is the residual for this case?

The model predicts that for the mass of 410 the Annuli is 20.20905. This means that there is -7.209046 residuals because of 13-20.

```
#
0.029571*410 +  8.084936
```

```
## [1] 20.20905
```

```
13-(0.029571*410 +  8.084936)
```

```
## [1] -7.209046
```

Q6

Which turtle (by row number in the original dataset) has the largest positive residual? What is the value of that residual? Which turtle (by row number in the original dataset) has the most negative residual? What is the value of that residual?

The dataset with the largest positive residuals is 185. The value of the residual is 23.19151. The dataset with the most negative residuals is 93. The value of the negative resiudal is -10.42705.

```
#
Turtles$residuals <- resid(mod1)
row_with_max_residuals <- which(Turtles$residuals == max(Turtles$residuals))
row_with_max_residuals
```

```
## [1] 185
```

```
max(Turtles$residuals)
```

```
## [1] 23.19151
```

```
row_with_min_residuals <- which(Turtles$residuals == min(Turtles$residuals))
row_with_min_residuals
```

```
## [1] 93
```

```
min(Turtles$residuals)
```

```
## [1] -10.42705
```

Q7

Find the means and standard deviations of **Mass** and **Annuli**. Also find the correlation between these two variables. Write R code and use the formulas to find the intercept and slope for regressing **Annuli** on **Mass**. Your answers should agree with those in Question 4.

The slope between the two variables is 0.029571 and the intercept is 8.084936.

```r
#
mean(Turtles$Mass)
```

```
## [1] 323.6166
```

```r
mean(Turtles$Annuli)
```

```
## [1] 17.65472
```

```r
sd(Turtles$Mass)
```

```
## [1] 111.4214
```

```r
sd(Turtles$Annuli)
```

```
## [1] 6.799029
```

```r
cor(Turtles$Annuli, Turtles$Mass)
```

```
## [1] 0.484611
```

```r
coef(mod1)[1]
```

```
## (Intercept)
##    8.084936
```

```r
coef(mod1)[2]
```

```
##       Mass
## 0.02957137
```