Carter Adamson

FISH 558

Lab 1 HW

1/24/2024

**Part A** (proficient variation)

1. **Describe how you spend your 2 hours working on R skills.**

I worked mostly on data manipulation skills. The goal was to knock out two useful goals at once by using the coding opportunity from this lab to work on skills relevant to my thesis work. I worked with a new package to me (“tidyverse”) which was recommended by a friend for working with data. In working on a QA/QC problem, I made some code to compare two versions of my data and identify discrepancies between them. The goal is to help identify data entry errors. I also spent a little bit of time learning RMarkDown, and organized this tidyverse code with it.

1. **What did you find most interesting?**

RMarkDown was the most interesting part for me. I’ve been given documents and assignments generated with RMarkDown before, but never made my own document using it before. I like the way you can present and describe chunks of code. Using it in the future might help make my code more digestible and easier to return to and jump back in.

1. **What do you think will be the biggest challenge for you in using R?**

Expanding my knowledge base will be a challengeing. There is always a temptation when you are beginner-level at something to stick to what you know and get “stuck in your ways” without doing any exploration of new techniques. This class should prove helpful in forcing me out of that comfort zone to learn something new.

1. **Compare and contrast a vector, a matrix, and a dataframe.**

A vector is a linear sequence of like elements: e.g. a>b>c

A matrix is a 2D array of like elements, like a vector of vectors: e.g. a>b>c

d>e>f

g>h>i

A data frame is more like a “table,” having rows and columns that can have unique data types: e.g. letter > integer > double

a > 1 > 1.0

b > 2 > 1.1

c > 3 > 1.2

**Part B**

**# 1. Create a dataframe named "data" that has a 4 columns (named: ID, X, Y, Z) and 5 rows. ID should go from 1 to 5, X should be the numbers [13,3,2,8,5], Y should go from 10 to 30 by fives, and Z should be the letters "a" through "e".**

data = data.frame(ID = 1:5, X=c(13,3,2,8,5), Y=seq(10, 30, by=5), Z=letters[1:5])

**# 2. Create an object called P that is set equal to the element in the 3rd row, 2nd column of your dataframe.**

p<-data[3,2]

**# 3. Calculate the sum of all elements in column Y using the sum() function.**

sum.Y<-sum(data$Y)

> sum.Y

[1] 100

**# 4. What is the average value of the elements in rows 1:3 for column 3? (Use a function that references your dataset to find the answer.)**

avg.3<-mean(data[1:3,3])

> avg.3

[1] 15

**# 5. Add a fifth column (named Xsquared), with elements that are the square of the values in column X.**

mutate(data, Xsquared=X\*X)

**Part C**

**# 1. Calculate and report the mean and standard deviation for TL and FL. (1 pt)**

avg.TL<-mean(sandeel[,1])

> avg.TL

[1] 156.5133

sd.TL<-sd(sandeel[,1])

> sd.TL

[1] 13.30874

avg.FL<-mean(sandeel[,2])

> avg.FL

[1] 152.4221

sd.FL<-sd(sandeel[,2])

> sd.FL

[1] 13.17417

**# 2. Generate a plot of the observed TL vs. FL (this means that TL is on the Y-axis,**

**# and FL is on the X-axis). Change the points in the plot from open circles to solid circles.**

ggplot(sandeel, aes(x=FL, y=TL))+geom\_point()

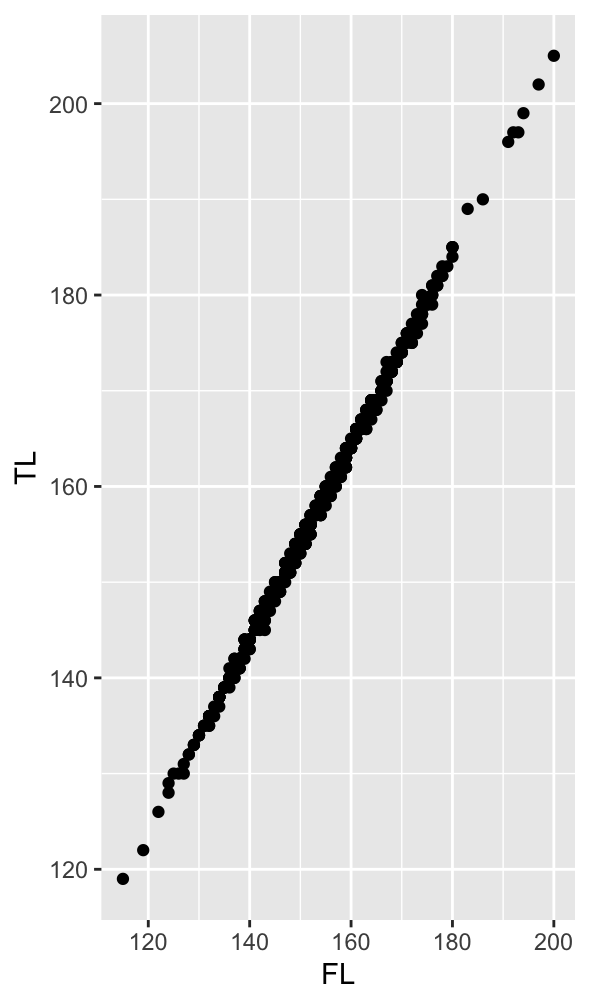


Figure 1: Observed total length (TL) vs fork length (FL) for sand eel.

**3. How many Total Hours did you spend on this homework assignment? (1 pt)**

3 hours

**4. Working with other classmates on your HW will be a tremendous help as the semester progresses.**

I worked alone on this assignment.

**Part D**

data.2<-data.frame(X=rep(1:20, each=5))

data.2<-mutate(data.2, Y=2.5\*X+10)

noise<-rnorm(100, 0, 2)

for(i in 0:100)

{

data.2[i, 2]<-data.2[i,2]+noise[i]

}

mod=lm(data.2$Y~data.2$X)

summary(mod)

ggplot(data.2, aes(x=X, y=Y))+geom\_point()+ggtitle("Y = 2.52X + 9.65")+theme(plot.title = element\_text(hjust = 0.5))

A graph showing a line graph

Description automatically generated

Figure 2: Y vs X plot for simulated data. X are integer values between 1 and 20, each repeated 5 times. Y were calculated as 2.5X + 10, then random noise was added. R-squared = 0.98.