Exercises

Exercise 1:

```
library(mosaic)
library(tidyverse)
```

Read in "NHANES" dataset and take the adult (age 18+ subset: (change the code below so that it points to where you have nhanes.csv saved)

```
nhanes <- read_csv('nhanes.csv')

## Rows: 5826 Columns: 10
## ── Column specification
──────────────────────────────────────────────────────────
## Delimiter: ","
## chr (4): sex, raceEthnicity, sbpClass, dbpClass
## dbl (6): age, systolic, diastolic, weight, height, bmi
##
## ℹ Use `spec()` to retrieve the full column specification for this
data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
this message.

### only adults
nhanes <- filter(nhanes,age>=18)
```

Consider weights for people with "PreHypertension," or systolic blood pressure between 120 and 140.

```
PreHyp <- filter(nhanes,sbpClass=='Prehypertension')
```

- Use the formula (as above) to calculate the 90% C.I. for the mean weight for all PreHypertensive people in the population represented by the NHANES survey.

```
NormalSBP <- filter(nhanes, sbpClass == 'Normal')
# put your answer here
weight <- NormalSBP$weight
sampleMean <- mean(weight)
sampleSD <- sd(weight)
sampleSize <- length(weight)
SE <- sampleSD/sqrt(sampleSize)
alpha <- 1- 0.90 ## proportion of time CI is wrong
critValue <- -qt(alpha/2,df=sampleSize-1)
```

```
CI.lower <- sampleMean - critValue*SE #Lower bound of the C.I.
CI.lower

## [1] 77.59759
```

- Use function t.test to calculate the 90% C.I. for the mean weight for all PreHypertensive people in the population represented by the NHANES survey.

```
t.test(weight,conf.level = 0.9)

##
##  One Sample t-test
##
## data:  weight
## t = 177.54, df = 2083, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  77.59759 79.04953
## sample estimates:
## mean of x
##  78.32356

# put your answer here
```

- What is the probability that the true population mean is inside the confidence interval you estimated? Does this question even make sense? Why or why not?

The probability is 90%. Yes it makes sense becuase that is the confidence inerval we are taking, it is likely that the mean will be inside.

- If you were to gather a large number of surveys from the same population, using the same methodology of NHANES, and estimate a confidence interval using each survey in the same way, what proportion of those confidence intervals should include the true population mean?

9/10 should include the true population mean

Exercise 2:

Now study the difference of the mean weight between adults with prehypertension and adults with SPB Stage 1 Hypertension.

We already have a dataset of prehypertensive people. Here's a dataset of people with Stage 1 Hypertension:

```
stage1 <- filter(nhanes,sbpClass=='Stage 1 Hypertension')
```

Use the formula to calculate the 90% C.I. for the mean weight difference:

- What is the standard error of the difference?

```
w1 <- PreHyp$weight
w2 <- stage1$weight
w1 <- NormalSBP$weight; #Group1: weights for normal SBP;
PreHyp <- filter(nhanes,sbpClass=='Prehypertension')
w2 <- PreHyp$weight; #Group2: weights for prehypertension;

### Calculation by formula: calculate the 95% confidence interval
#   for the difference of population means
y1 <- mean(w1); #sample mean of group1.
y2 <- mean(w2); #sample mean of group2.

s1 <- sd(w1); #sample SD
s2 <- sd(w2);

n1 <- length(w1); #sample size
n2 <- length(w2); #sample size

V1 <- s1^2/n1; #sampling variance of sample mean of group 1.
V2 <- s2^2/n2; #sampling variance of sample mean of group 2.

Vdiff <- V1+V2 # sampling variance of difference of two sample means

SE <- sqrt(Vdiff); #Standard error for the difference of sample means.
SE

## [1] 0.7426056
```

- What is a 90% CI of the difference?

```
## put your answer here
alpha <- 1- 0.90 ## proportion of time CI is wrong
critValue <- -qt(alpha/2,df=sampleSize-1)
CI.lower <- sampleMean - critValue*SE #Lower bound of the C.I.
CI.lower
```

```
## [1] 77.10154
```

- Use function t.test to calculate the 90% C.I. for the mean weight difference.

```
## put your answer here
t.test(w1, w2,conf.level = 0.90);

##
##  Welch Two Sample t-test
##
## data:  w1 and w2
## t = -12.243, df = 3148.4, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  -10.313292  -7.869618
## sample estimates:
## mean of x mean of y
##  78.32356  87.41502
```