Lab 7 exercise

library(mosaic)  
library(tidyverse)

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

nhanes <- read.csv('nhanes.csv' , header = T)  
### only adults  
nhanes <- filter(nhanes,age>=18)

# Exercise:

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

Here are datasets of people with PreHypertension and Stage 1 Hypertension, respectively:

PreHyp <- filter(nhanes,sbpClass=='Prehypertension')  
  
stage1 <- filter(nhanes,sbpClass=='Stage 1 Hypertension')  
  
w1 <- stage1$weight; #Group1: weights for normal SBP;  
w2 <- PreHyp$weight; #Group2: weights for prehypertension;

* Formulate a null and alternative hypothesis for a two-sided hypothesis test for the difference in mean weight between adults who are Prehypertensive and those with Stage 1 hypertension (you may use words instead of math, if you prefer):

The population mean weight of people with sbpClass= Normal is equal to the population mean weight of people with sbpClass= Prehypertension

The population mean weight of people with sbpClass= Normal is not equal to the population mean weight of people with sbpClass= Prehypertension

We will test at level .

* What is the difference in sample means between the two populations?

y1 <- mean(w1); #sample mean of group1.   
y2 <- mean(w2); #sample mean of group2.  
y1

## [1] 82.9

y2

## [1] 87.4

* What is the standard error of the difference in sample means?

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] 1.11

* What is the T-statistic (ratio test statistic) to test versus ?

## put your answer here  
T=(y1-y2)/SE  
T

## [1] -4.13

* Use the same test statistic to estimate a p-value testing

## put your answer here  
2\*pt(-abs(T),df=min(n1,n2)-1)

## [1] 4.19e-05

* What is the correct multiplier (i.e. critical value) for a 95% confidence interval for ?

## put your answer here  
alpha <- 1-0.95  
mult <- qt(1-alpha/2,df=min(n1,n2)-1)  
mult

## [1] 1.96

* Use the difference in sample means, standard error, and critical value you calculated to estimate a 95% confidence interval for

## put your answer here  
mean(w1)-mean(w2)+c(-1,1)\*mult\*SE

## [1] -6.73 -2.39

* Use function t.test to estimate a p-value testing and a 95% confidence interval Does it (roughly) agree?

#put your answer here  
t.test(w1, w2,conf.level = 0.95); #Check the result "95 percent

##   
## Welch Two Sample t-test  
##   
## data: w1 and w2  
## t = -4, df = 1178, p-value = 4e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -6.73 -2.39  
## sample estimates:  
## mean of x mean of y   
## 82.9 87.4

* What is your decision?

we will reject the null hypothesis because p< alpha

* What can you conclude about the mean weight of people with pre-hypertension verus stage-1 hypertension?

There is no evidence that the mean weight of people with pre-hypertension us equal to the mean weigth of people with stage-1 hypertension