HW3\_xl2836

Xinyi Lin

10/4/2018

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.0.0 ✔ purrr 0.2.5  
## ✔ tibble 1.4.2 ✔ dplyr 0.7.6  
## ✔ tidyr 0.8.1 ✔ stringr 1.3.1  
## ✔ readr 1.1.1 ✔ forcats 0.3.0

## ── Conflicts ────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(readxl)

# Problem 1

# Problem 2

First, we need to import data “HeavySmoke.csv”.

heavysmoke\_df = read\_csv(file = "./data/HeavySmoke.csv")

## Parsed with column specification:  
## cols(  
## ID = col\_integer(),  
## BMI\_base = col\_double(),  
## BMI\_6yrs = col\_double()  
## )

## Question 1

In order to test wether BMI has changed 6 years after quitting smoking, we need to test the means of BMI\_base and BMI\_6yrs are different or not. As we don’t know the variance of two samples, we use t test.

Assuming the mean of BMI\_base is and the mean of BMI\_6yrs is , the difference between BMI\_base and BMI\_6yrs is . The samples sizes is 10.

The null hypothesis : , the alternative hypothesis : .

d = heavysmoke\_df$BMI\_6yrs - heavysmoke\_df$BMI\_base

The test statustuc :

4.314474.