Weekend Homework Part2

## Setup

setwd("C:/Users/22700/Desktop")

## Packages

library(GGally)

## Loading required package: ggplot2

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(ggplot2)  
library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(data.table)

## Load Data

load("dt\_wages.RData")  
dt.wages2 <- data.table(dt.wages)  
rm(dt.wages)

## How to calculate the 95% confidence intervals for the population wage?

dt.wages2[, list(avg\_wage=mean(wage))]

## avg\_wage  
## 1: 5.896103

## By location  
dt.wages2[, list(avg\_wage=mean(wage)), by=northcen]

## northcen avg\_wage  
## 1: 0 5.958300  
## 2: 1 5.710455

dt.wages2[, list(avg\_wage=mean(wage)), by=south]

## south avg\_wage  
## 1: 0 6.176991  
## 2: 1 5.386898

## Confidence intervals

## Let’s create a function to calculate the 95% confidence interval

dt.wages2[, list(avg\_wage=mean(wage), sd\_wage=sd(wage))]

## avg\_wage sd\_wage  
## 1: 5.896103 3.693086

# Creating a function to calculate the 95% confidence interval  
conf.int <- function(X){  
 n <- length (X)  
 error <- qt(0.975, df=n-1) \* sd(X) / sqrt (n)  
 mean.X <- mean(X)  
 return (list( lower = mean.X- error, upper = mean.X + error ))}  
  
#Finally, apply the function to the wage variable  
dt.wages2[, conf.int(wage)]

## lower upper  
## 1: 5.579768 6.212437

# Apply the conf.int function to the wage variable  
dt.wages2[, conf.int(wage)]

## lower upper  
## 1: 5.579768 6.212437

## Calculating confidence intervals by groups

dt.wages2[, conf.int(wage), by=northcen]

## northcen lower upper  
## 1: 0 5.580891 6.335708  
## 2: 1 5.137948 6.282961

dt.wages2[, conf.int(wage), by=south]

## south lower upper  
## 1: 0 5.753979 6.600004  
## 2: 1 4.939862 5.833935

## t-test

dt.wages2[, t.test (wage ~ northcen)]

##   
## Welch Two Sample t-test  
##   
## data: wage by northcen  
## t = 0.71367, df = 255.17, p-value = 0.4761  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.4360585 0.9317484  
## sample estimates:  
## mean in group 0 mean in group 1   
## 5.958300 5.710455