lab2b\_estes

Andrew Estes

9/5/2021

#### Installing Packages and Data Set

library(tidyr)

## Warning: package 'tidyr' was built under R version 4.1.1

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v dplyr 1.0.7  
## v tibble 3.1.3 v stringr 1.4.0  
## v readr 2.0.0 v forcats 0.5.1  
## v purrr 0.3.4

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

kim <- read.csv("C:\\Users\\andre\\OneDrive\\Desktop\\PDAT 610\\Module2b\\Clean-KimData.csv")  
head(kim)

## Semester Gender Siblings Birth.Order Shoe.Size Height Weight dog.vs.cat  
## 1 6 F 5 Middle 11.0 71 195 cat  
## 2 4 F 0 Only 10.0 64 187 cat  
## 3 6 F 1 Last 9.5 69 150 dog  
## 4 7 F 3 First 9.5 64 193 dog  
## 5 6 M 2 Middle 13.0 73 181 dog  
## 6 6 M 2 First 10.0 68 167 dog  
## Handed On.Off.Campus Calories.per.day Servings.of.Fruit Cups.of.Water  
## 1 Right off 2200 1 4  
## 2 Right on 2100 1 3  
## 3 Right off 1500 3 4  
## 4 Right off 1290 3 10  
## 5 Right off 5000 7 10  
## 6 Right off 2500 2 5  
## Cups.of.Coffee Hours.of.Sleep Hours.spent.studying.per.week  
## 1 1 7 6  
## 2 1 7 15  
## 3 NA 9 10  
## 4 NA 8 6  
## 5 NA 7 12  
## 6 NA 8 8  
## Hours.spent.working.per.week Hours.spent.workingout.per.wee  
## 1 10 3.0  
## 2 4 4.5  
## 3 6 20.0  
## 4 16 4.0  
## 5 12 9.0  
## 6 4 5.0  
## Hours.socializing.per.w Politically.Liberal Religiously.C.or.L  
## 1 4 2 0  
## 2 8 -1 0  
## 3 20 -2 -1  
## 4 24 0 1  
## 5 8 -2 -1  
## 6 7 -2 0  
## Socially.C.or.L Phone Hrs.per.day.on.phone Hrs.day.on.phone.not.talking  
## 1 2 NA NA  
## 2 -1 NA NA  
## 3 -1 NA NA  
## 4 1 NA NA  
## 5 0 NA NA  
## 6 -1 NA NA

## Question 1: Select “male” OR “first year” and compare to “male AND first year”

#### Answer: there were 229 observations of “male” OR “first year” compared to 43 observations of “male AND first year”

maleORfirst <- filter(kim, Semester < 2 | Gender == "M" )  
count(maleORfirst)

## n  
## 1 229

## Question 2: Calculate BMI

#adding numerical BMI to the dataframe  
kimBMI <- mutate(kim, BMI = 703 \*(Weight/(h=Height^2)))  
  
#adding categorical TRUE/FALSE determining status of obese to dataframe  
kimData <- mutate(kimBMI, Obese = BMI > 30)  
head(kimData)

## Semester Gender Siblings Birth.Order Shoe.Size Height Weight dog.vs.cat  
## 1 6 F 5 Middle 11.0 71 195 cat  
## 2 4 F 0 Only 10.0 64 187 cat  
## 3 6 F 1 Last 9.5 69 150 dog  
## 4 7 F 3 First 9.5 64 193 dog  
## 5 6 M 2 Middle 13.0 73 181 dog  
## 6 6 M 2 First 10.0 68 167 dog  
## Handed On.Off.Campus Calories.per.day Servings.of.Fruit Cups.of.Water  
## 1 Right off 2200 1 4  
## 2 Right on 2100 1 3  
## 3 Right off 1500 3 4  
## 4 Right off 1290 3 10  
## 5 Right off 5000 7 10  
## 6 Right off 2500 2 5  
## Cups.of.Coffee Hours.of.Sleep Hours.spent.studying.per.week  
## 1 1 7 6  
## 2 1 7 15  
## 3 NA 9 10  
## 4 NA 8 6  
## 5 NA 7 12  
## 6 NA 8 8  
## Hours.spent.working.per.week Hours.spent.workingout.per.wee  
## 1 10 3.0  
## 2 4 4.5  
## 3 6 20.0  
## 4 16 4.0  
## 5 12 9.0  
## 6 4 5.0  
## Hours.socializing.per.w Politically.Liberal Religiously.C.or.L  
## 1 4 2 0  
## 2 8 -1 0  
## 3 20 -2 -1  
## 4 24 0 1  
## 5 8 -2 -1  
## 6 7 -2 0  
## Socially.C.or.L Phone Hrs.per.day.on.phone Hrs.day.on.phone.not.talking  
## 1 2 NA NA  
## 2 -1 NA NA  
## 3 -1 NA NA  
## 4 1 NA NA  
## 5 0 NA NA  
## 6 -1 NA NA  
## BMI Obese  
## 1 27.19401 FALSE  
## 2 32.09497 TRUE  
## 3 22.14871 FALSE  
## 4 33.12476 TRUE  
## 5 23.87746 FALSE  
## 6 25.38949 FALSE

## Question 3: Calculate BMI for each year

#dividing the semeesters into 2 to calculate the year  
#adding 1 to the calculation so Freshman are Year 1 instead of Year 0  
kimData2 <- kimData %>%   
 mutate(Year = round(Semester/2)+1)  
head(kimData2)

## Semester Gender Siblings Birth.Order Shoe.Size Height Weight dog.vs.cat  
## 1 6 F 5 Middle 11.0 71 195 cat  
## 2 4 F 0 Only 10.0 64 187 cat  
## 3 6 F 1 Last 9.5 69 150 dog  
## 4 7 F 3 First 9.5 64 193 dog  
## 5 6 M 2 Middle 13.0 73 181 dog  
## 6 6 M 2 First 10.0 68 167 dog  
## Handed On.Off.Campus Calories.per.day Servings.of.Fruit Cups.of.Water  
## 1 Right off 2200 1 4  
## 2 Right on 2100 1 3  
## 3 Right off 1500 3 4  
## 4 Right off 1290 3 10  
## 5 Right off 5000 7 10  
## 6 Right off 2500 2 5  
## Cups.of.Coffee Hours.of.Sleep Hours.spent.studying.per.week  
## 1 1 7 6  
## 2 1 7 15  
## 3 NA 9 10  
## 4 NA 8 6  
## 5 NA 7 12  
## 6 NA 8 8  
## Hours.spent.working.per.week Hours.spent.workingout.per.wee  
## 1 10 3.0  
## 2 4 4.5  
## 3 6 20.0  
## 4 16 4.0  
## 5 12 9.0  
## 6 4 5.0  
## Hours.socializing.per.w Politically.Liberal Religiously.C.or.L  
## 1 4 2 0  
## 2 8 -1 0  
## 3 20 -2 -1  
## 4 24 0 1  
## 5 8 -2 -1  
## 6 7 -2 0  
## Socially.C.or.L Phone Hrs.per.day.on.phone Hrs.day.on.phone.not.talking  
## 1 2 NA NA  
## 2 -1 NA NA  
## 3 -1 NA NA  
## 4 1 NA NA  
## 5 0 NA NA  
## 6 -1 NA NA  
## BMI Obese Year  
## 1 27.19401 FALSE 4  
## 2 32.09497 TRUE 3  
## 3 22.14871 FALSE 4  
## 4 33.12476 TRUE 5  
## 5 23.87746 FALSE 4  
## 6 25.38949 FALSE 4

#grouping the obesity rates by Year  
df <- kimData2 %>% group\_by(Year)  
df %>% summarise (  
 avg = mean(BMI, na.rm = TRUE)  
)

## # A tibble: 6 x 2  
## Year avg  
## <dbl> <dbl>  
## 1 1 25.1  
## 2 2 23.1  
## 3 3 24.0  
## 4 4 23.5  
## 5 5 24.6  
## 6 6 28.5

## Question 4: Calculate the obesity percentage for each year

#filtering obesity by year  
firstYearObese <- filter(df, Year == 1 & BMI > 30)  
secondYearObese <- filter(df, Year == 2 & BMI > 30)  
thirdYearObese <- filter(df, Year == 3 & BMI > 30)  
fourthYearObese <- filter(df, Year == 4 & BMI > 30)  
fifthYearObese <- filter(df, Year == 5 & BMI > 30)  
sixthYearObese <- filter(df, Year == 6 & BMI > 30)  
  
  
 #filtering data by year  
firstYear <- filter(df, Year == 1)  
secondYear <- filter(df, Year == 2)  
thirdYear <- filter(df, Year == 3)  
fourthYear <- filter(df, Year == 4)  
fifthYear <- filter(df, Year == 5)  
sixthYear <- filter(df, Year == 6)  
  
 #dividing total number of obesity/yr by total respondents/yr  
obeseFirstYr <- count(firstYearObese)/count(firstYear)  
obeseSecondYr <- count(secondYearObese)/count(secondYear)  
obeseThirdYr <- count(thirdYearObese)/count(thirdYear)  
obeseFourthYr <- count(fourthYearObese)/count(fourthYear)  
obeseFifthYr <- count(fifthYearObese)/count(fifthYear)  
obeseSixthYr <- count(sixthYearObese)%/%count(sixthYear)

## Error in Ops.data.frame(count(sixthYearObese), count(sixthYear)): '%/%' only defined for equally-sized data frames

#output of the above division  
obeseFirstYr

## Year n  
## 1 1 0.1727273

obeseSecondYr

## Year n  
## 1 1 0.07207207

obeseThirdYr

## Year n  
## 1 1 0.08928571

obeseFourthYr

## Year n  
## 1 1 0.05

obeseFifthYr

## Year n  
## 1 1 0.173913

obeseSixthYr

## Error in eval(expr, envir, enclos): object 'obeseSixthYr' not found

## Question 5: Analyze feet size by gender

#### Answer: I removed all observations over 2 standard deviations away from the mean

#### The male average drops from an initial value of 11.51 to 10.81

#### The male sd drops from an initial value 8.19 to 1.31

#### The female average drops from an initial value of 7.99 to 7.96

#### The female SD drops from an initial value 1.33 to 1.17

#male analysis  
maleDF <- filter(df, Gender == "M")  
mean(maleDF$Shoe.Size, na.rm = TRUE)

## [1] 11.51398

sd(maleDF$Shoe.Size, na.rm = TRUE)

## [1] 8.187464

maleDF2 <- filter(maleDF, Shoe.Size < 20)  
mean(maleDF2$Shoe.Size, na.rm = TRUE)

## [1] 10.87969

sd(maleDF2$Shoe.Size, na.rm = TRUE)

## [1] 1.508355

maleDF3 <- filter(maleDF2, Shoe.Size > 7.86 & Shoe.Size < 13.9)  
mean(maleDF3$Shoe.Size, na.rm = TRUE)

## [1] 10.80556

sd(maleDF3$Shoe.Size, na.rm = TRUE)

## [1] 1.312961

#female analysis  
femaleDF <- filter(df, Gender == "F")  
mean(femaleDF$Shoe.Size, na.rm = TRUE)

## [1] 7.992991

sd(femaleDF$Shoe.Size, na.rm = TRUE)

## [1] 1.329053

femaleDF2 <- filter(femaleDF, Shoe.Size > 5.33 & Shoe.Size < 10.65)  
mean(femaleDF2$Shoe.Size, na.rm = TRUE)

## [1] 7.958333

sd(femaleDF2$Shoe.Size, na.rm = TRUE)

## [1] 1.172122