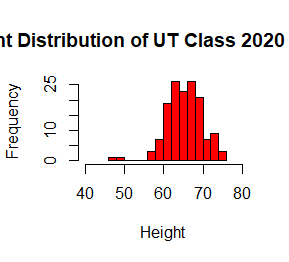
Preliminary Analysis Report

Intro: The goal of this study was to find a correlation between nutrition, specifically milk, and height and how strong the relationship is for UT students of class of 2020. I assumed that the more cups of milk a person drank weekly during the ages of 14–18 years old, the taller the person will be. However, drinking less cups of milk does not shrink a person’s height. Additionally, I expected that males will overall have a greater height than the heights of females due to biological differences.

Data Collection: I collected my data via a survey form created on Google Docs Forms. I posted the link to this survey in the Facebook group Class of 2020, The University of Texas. This group is not official but does contain 10,000 members whose majority of members does fit my population of interest. My sample subjects were specifically those in the Facebook group Class of 2020 who clicked and filled out my survey. My final sample size after removing response biases and outliers was 146. I had to remove the responses that said that they were not class of 2020 and responses that claimed to not be male or female. There was also one that claimed to be an extremely large height and an extremely large number of cups of milk drank. The two numbers were inhuman and impossible to have so I removed that response.

Descriptive Analysis of Response Variable:

The distribution of heights appears to be normally distributed with the exception of a potential outlier. The outlier can be justified to be removed because it is very low, but we would want to keep it to analyze the lower height extreme’s relationship with nutrition. The mean of these heights is 65.87116 inches with the standard deviation 4.524983.



> stats <- read.csv("Heights of UT Students clean.csv")

> hist(stats$Height..inches., main = "Height Distribution of UT Class 2020 Students",

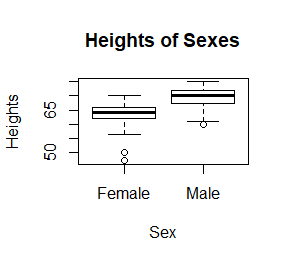
+ xlab = 'Height', col = 'red', xlim=c(40, 80), breaks = 10)

> mean(stats$Height..inches.)

[1] 65.87116

> sd(stats$Height..inches.)

[1] 4.524983



Investigation of Explanatory Variable #1:

> boxplot(stats$Height..inches. ~ stats$Sex, main = "Heights of Sexes",

+ xlab = "Sex", ylab = "Heights")

> males <- stats$Height..inches.[stats$Sex=='Male']

> females <- stats$Height..inches.[stats$Sex=='Female']

> summary(males)

Min. 1st Qu. Median Mean 3rd Qu. Max.

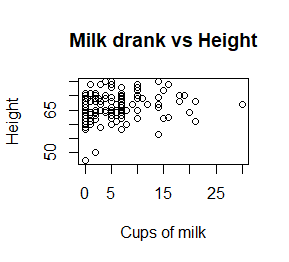
60.00 67.50 70.00 69.47 72.00 75.00

> summary(females)

Min. 1st Qu. Median Mean 3rd Qu. Max.

47.00 62.00 64.00 63.94 66.00 70.00

The males’ IQR of heights is greater than that of the females. The median, minimum, maximum, Q2, Q4 of the males’ distribution are all greater than those of the females. The relationship appears to be that being male increases your height which corresponds to my assumption at the beginning of this study. The males’ IQR is 14.5, max is 75, min is 60, median is 70, Q1 is 67.5, Q3 is 72. The females’ IQR is 4, max is 70, min is 60, median is 64, Q1 is 62, Q3 is 66.

Investigation of Explanatory Variable #2:

> plot(stats$Cups.of.milk.drank.weekly, stats$Height..inches., xlab =

+ 'Cups of milk', ylab = 'Height', main = 'Milk drank vs Height')

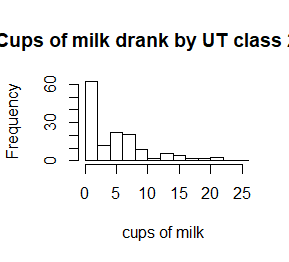
> hist(stats$Cups.of.milk.drank.weekly, xlim=c(0,25), breaks=11)

> summary(stats$Cups.of.milk.drank.weekly)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 1.000 4.000 5.295 7.000 30.000

The distribution of cups of milk drank is right skewed as shown

by the histogram. The scatter plot shows a nonlinear, weak, positive correlation between height and cups of milk drank. This corresponds to what I expected but the relationship appears weaker than I expected. This may be due to nutrition other than milk that factors more heavily into an individual’s height or that genetics plays a larger role than amount of milk drank. The cups of milk min is 0, Q1 is 1, median is 4, Q3 is 7, max is 30 and the IQR is 6.