ISyE 2028, Fall 2015
Professor X

Aditya Sehgal
902905469

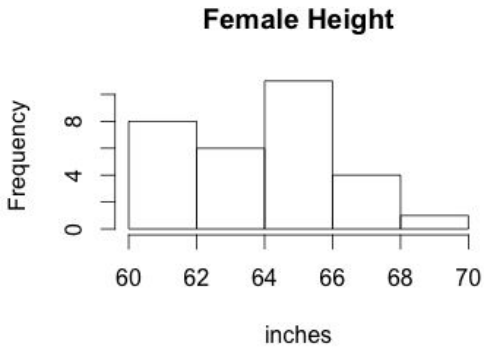
Male and Female Height and Shoe Size

For my ISyE 2028 Bonus Project, I decided to perform a study on Male and Female height and their shoe size. To collect data for this project, I sent out anonymous forms using Google Forms, through social networking sites to friends and acquaintances. I requested the persons (1) Gender (M/F), (2) Height (inches) as well as (3) Shoe Size (US). Most people that I sent out the form to are students at Georgia Tech or are in the same age group as me. Using this method of data collection, I will assume that (1) the sample is random even though the data may not truly reflect a random sample of the population of all males and females, and (2) that the people who submitted the data did not submit false data, as this is a non-sampling error. Further, I chose to select n=30 as a sample size for both male and female samples.

Here is a summary of the collected data:

	Male Height (inches)	Male Shoe Size (US)	Female Height (inches)	Female Shoe Size (US)
mean (x)=	70.2	10.6	64.3	7.2
median =	70.4	10.4	64.6	6.9
std deviation (s) =	2.17	1.66	2.54	1.31
n =	30	30	30	30

It is interesting to note that the standard deviation for Male and Female Height is more than the standard deviation for the respective Shoe Size.
The distribution of the data can be viewed below:



Now, consider the following hypothesis test (at a 0.05 significance level):

H0: “Mean height of a male is the same as the mean height of a female”
H1: “Mean height of a male is greater than the mean height of a female”

The following calculations were done with R:

```
> t = ((x1_bar - x2_bar) - 0)/sqrt((s1*s1/n1)+(s2*s2/n2))
> t
[1] 9.673214
```

From the t-table using n= 30 and a significance level of 0.05, the value from the t-table is 1.697. Since 9.673214 > 1.697, we reject the null hypothesis H0.

So there is enough evidence to conclude that the mean male height is greater than the mean female height.

Now, let’s perform the same hypothesis test on male and female shoe sizes (at a 0.05 significance level).

H0: “Mean shoe size of a male is the same as the mean shoe size of a female”
H1: “Mean shoe size of a male is greater than the mean shoe size of a female”

The following calculations for the t-test were done using R

```
> t = ((x1_bar - x2_bar) - 0)/sqrt((s1*s1/n1)+(s2*s2/n2))
> t
[1] 8.806497
```

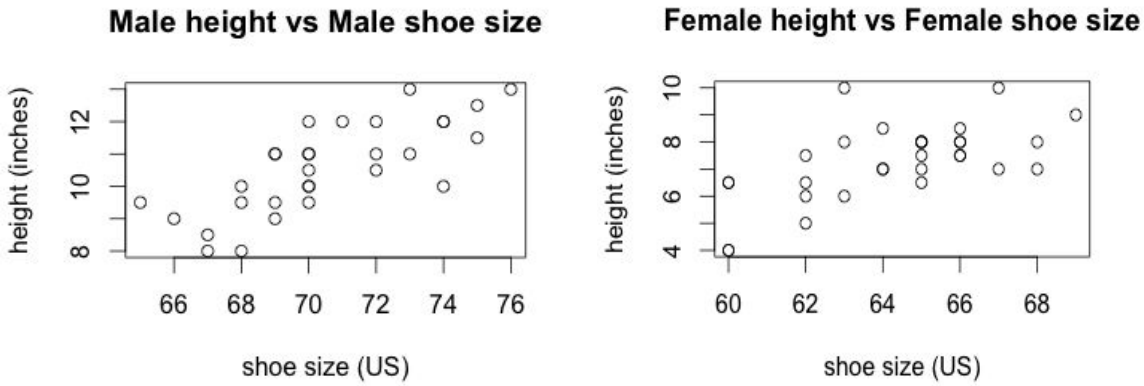
Once again, we can reject H0 in favor of H1 since the computed t-value 8.806497 > 1.697, by looking up the t-table for n=30 at a significance level of 0.05. So we can conclude that we

have sufficient evidence to believe that the mean male shoe size is greater than the mean female shoes size.

From the above tests, it is clear that both mean height and shoe size for a male is greater than the same for a female. Let's find out if there is a linear correlation between the height and shoe size for both males and females.

The correlation coefficient was found to be 0.7653334 for male height and shoes size and 0.6252704 for female height and shoe size. This suggests that height and shoe size are strongly correlated, since both the computed values are greater than the critical value of 0.37 at a 0.05 significance level.

So there is enough evidence to support that there is a correlation between male height and male shoe size as well as between female height and female shoes size. Additionally, both the coefficients are greater than 0.5, which again suggests strong correlation between the quantities.



I used the data in R to find a fitting line for male height and shoe size, and the line can be described to have a slope 1.555 of and an intercept of 54.072.
Similarly, a fitting line for female height and shoe size was found to have a slope 1.113 of and an intercept of 56.177.

Based on the tests through this project, I have found that on average, males are taller than females, males have bigger feet than females, there is a correlation between a person's height and shoe size. These conclusions seem match up with what I observe in day to day life. This project was a great way for me to verify and confirm observations that we make in our day to day life, but take for granted as we do not perform these tests on a daily basis. However, this sample was not completely random, as it focussed mainly on my classmates and friends at Tech, as well as other friends and acquaintances, so we would need a completely unbiased random sample to know for sure!

I referred to the following websites while working on this project:

http://www.cyclismo.org/tutorial/R/plotting.html
https://www.google.com/forms/about/
http://yuppal.people.ysu.edu/econ_3790/t-table.pdf
http://ww2.coastal.edu/kingw/statistics/R-tutorials/simplelinear.html
http://www.chem.utoronto.ca/coursenotes/analsci/stats/ttest.html
http://blog.datacamp.com/r-tutorial-read-excel-into-r/