Mathew Witek 05/02/2019

Statistics II Final Project

Shoe Size and Height

Introduction

We are first presented with a data set containing both the shoe sizes and heights of both females and males. Our goal in this project is to split the data based on sexes, and then analyze the two data sets to see if shoe size is a useful predictor for height, along with other calculations and predictions to show that we can properly analyze and interpret the data.

Procedure

The tools used to complete this project were SAS statistical software and excel in order to compute and calculate the required values needed to adequately analyze and interpret the given data set(s). Although all code, output, and results will be attached to this writeup, screen shots of the tools and technologies used to find the answers will be provided throughout this writeup in order for the reader to better grasp where the answers came from.

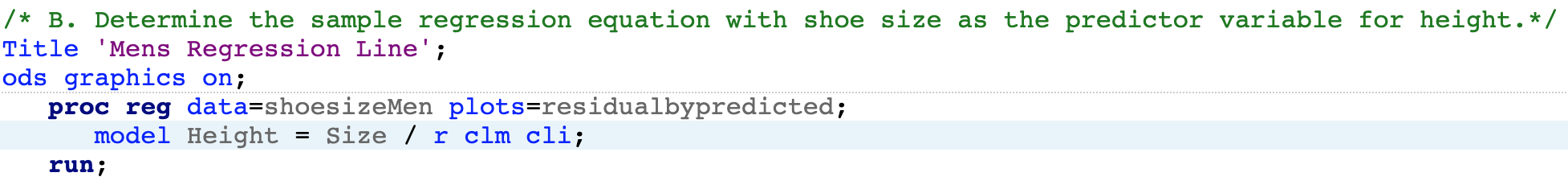
Analysis

1. Here we are asked to simply separate the data into two data sets, splitting the original data set based on sexes; male and female. This was done with a two simple ‘if’ statements in the SAS statistical software, resulting in…



The following B. through J. are all analysis of the male data set.

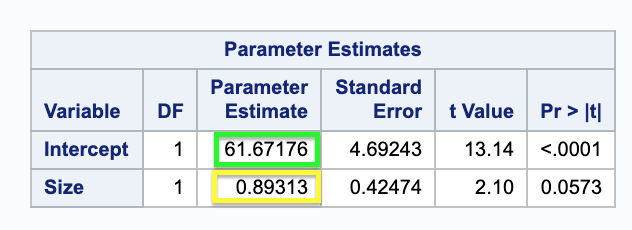
1. Moving on, we are asked to determine the sample regression equation with shoe size as the predictor variable for height. In SAS, there is a simple way to output a diagram and model a regression line based on a data set by using the proc reg data method and then using model Height = Size because we know that size is being used as the predictor variable for height. This is what I did in SAS below…



This then output a lot of results of which one must carefully read through to interpret and

find the necessary values to construct the sample regression equation. The output is

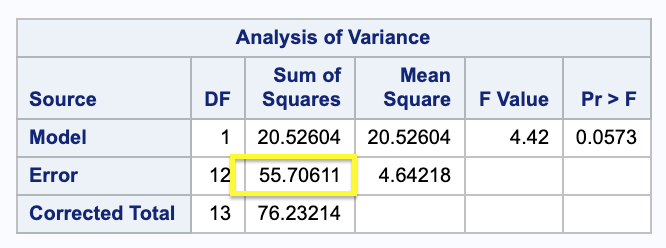
shown below and highlighted to show the y-intercept (in green) and the slope (in yellow).



With the help of SAS statistical software, we get that the sample regression equation for

the male data set is, ŷ = 61.67176 + 0.89313x.

1. In order to find the standard error (), we must first find the error sum of squares (*SSE*). The SSE is given in the output from the provided SAS code above.



Now that we have SSE = 55.70611, we can use this value in the equation .

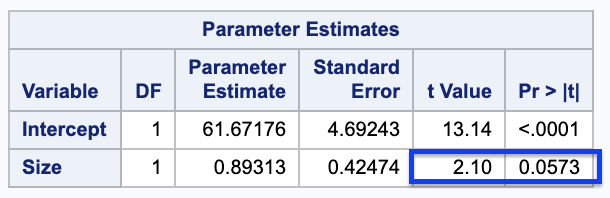
We know that n = 14 and plugging that all in gives us . This means that the

predicted height of a male in the sample differs, on average, from the observed height by

2.155 inches.

1. To determine whether shoe size is useful for predicting height, we must complete a hypothesis test, where Ho: , and Ha: . The previously shown SAS codes output already does the work for us and provides us with the needed test statistic and

P-value to interpret.



Here we see that the test statistic has a value of 2.10, and the P-value is 0.0573 when

= 0.05, which means that the P-value is greater than . That means that we cannot

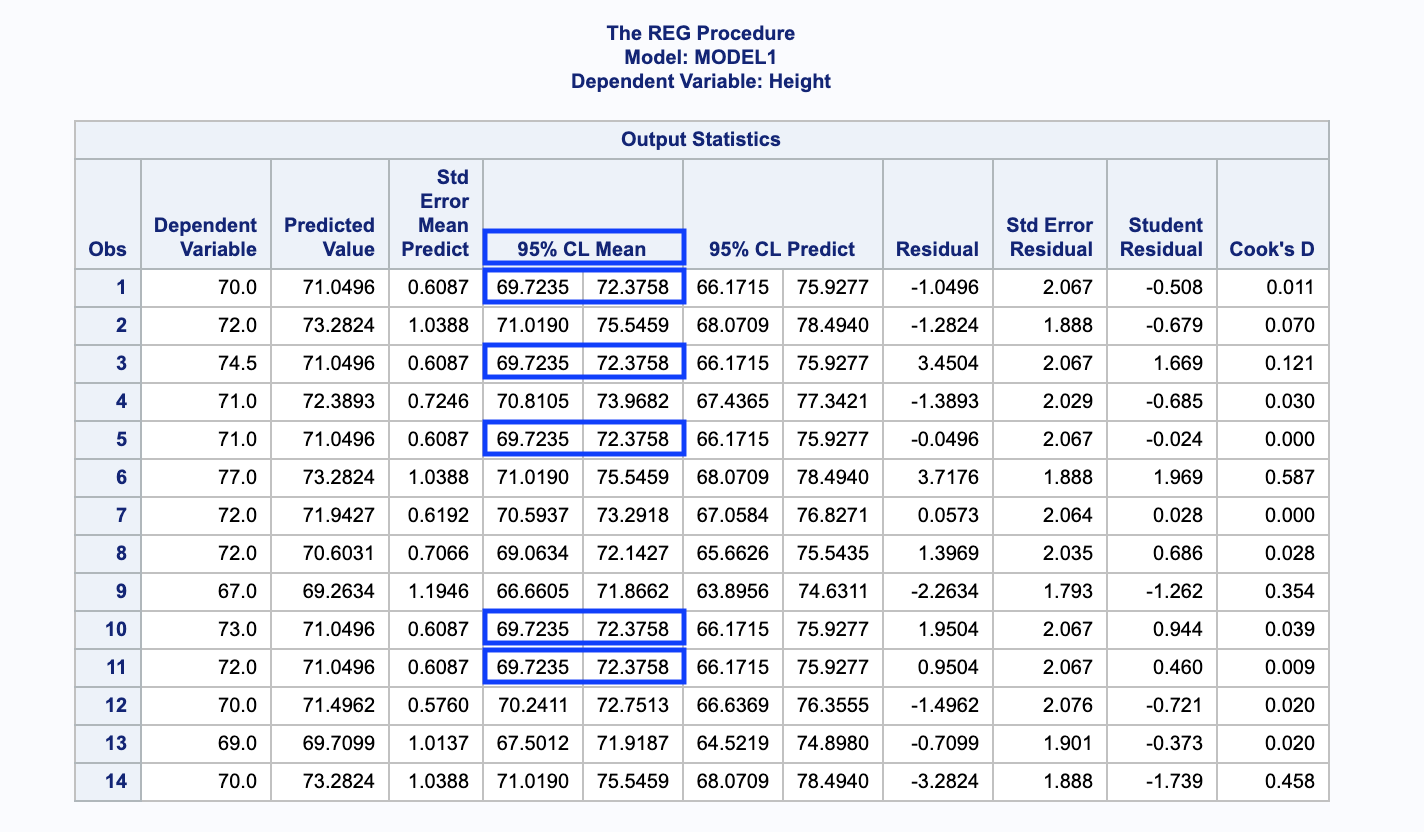
reject the null hypothesis of Ho: . This means that the data does not provide sufficient evidence to conclude that shoe size is useful for predicting height.

1. A point estimate for the mean height of all males who wear a size 10.5 shoe is given by simply plugging in 10.5 into the sample regression equation. This gives us

ŷ = 61.67176 + 0.89313(10.5)

ŷ = 71.05 inches.

1. Here we are asked to obtain a 95% confidence interval for the mean height of all males who wear a size 10.5 shoe. The SAS code shown in part B once again provides the output that we need in order to answer this question. The useful output that the code provides us with is…



Looking at the table above, we can see that the observations which correspond to a shoe

size of 10.5 are observations 1, 3, 5, 10, and 11. All of which have a 95% confidence

interval of (69.72, 72.38). We can be 95% confident that the height of men who wear a

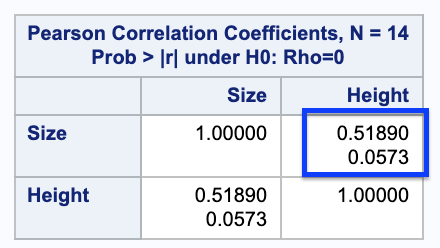
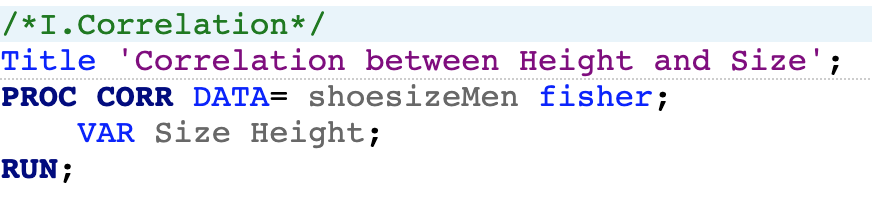
size 10.5 shoe lies between 69.72 and 72.38 inches.

1. The predicted height of a male who wears a size 10.5 shoe can be found by plugging in 10.5 into the sample regression equation.

ŷ = 61.67176 + 0.89313(10.5)

ŷ = 71.05 inches.

1. Looking at the same output used to find the 95% confidence interval in part f. we can see that next to the 95% confidence interval column, there is 95% confidence interval predict column where we will find our answer. The 95% prediction interval for the height of a male who wears a size 10.5 shoe is (66.17, 75.93). We can be 95% confident that the height of males who wear a size 10.5 shoe lies between 66.17 and 75.93 inches.
2. To find if the data provides sufficient evidence to conclude that shoe size and height are positively linearly correlated, we used the proc corr method in SAS to calculate the correlation value between Size and Height, and the P-value. The code is provided below along with the output it generated…



As we can see in the boxed in region, we are provided with the correlation value of 0.519,

and a P-value of 0.0573. The P-value of the right-tailed test is 0.0285. This gives us…

P-value (0.0285) < (0.05)

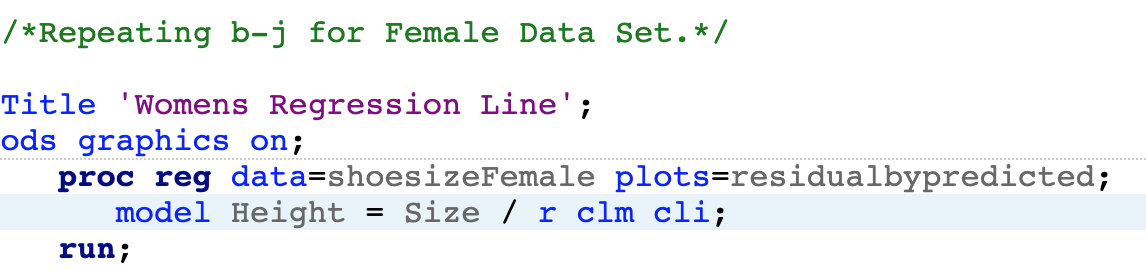
This means that we reject the Ho: p = 0, and we can conclude that at the 5% significance level, the data does provide sufficient evidence to conclude that shoe size and height are positively linearly correlated.

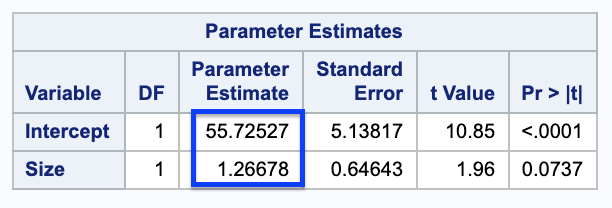
\*\*Upon completing I, J. states to repeat B-I for the female data set of shoe size and height.\*\*

B.) We are asked to determine the sample regression equation with shoe size as the

predictor variable for height. The following code is used which provides a lot of

useful output.



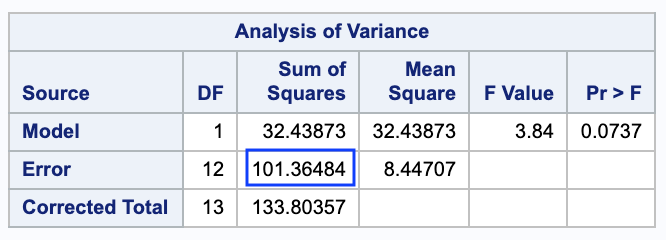


The boxed in data gives us our sample regression equation for shoe size and height

which is ŷ = 55.725 + 1.267x.

C.) In order to find the standard error (), we must first find the error sum of squares

(*SSE*). The SSE is given in the output from the provided SAS code in part B.



Now that we have SSE = 101.365, we can use this value in the equation .

We know that n = 14 and plugging that all in gives us . This means that

the predicted height of a female in the sample differs, on average, from the observed

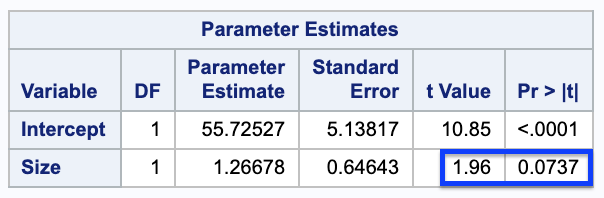
height by 2.906 inches.

D.) In order to determine whether shoe size is useful for predicting height, we must

complete a hypothesis test, where Ho: , and Ha: . The previously shown

SAS codes output already does the work for us and provides us with the needed test

statistic and P-value to interpret.



Here we see that the test statistic has a value of 1.96, and the P-value is 0.0737 when

= 0.05, which means that the P-value is greater than . That means that we cannot

reject the null hypothesis of Ho: . This means that the data does not provide

sufficient evidence to conclude that shoe size is useful for predicting height.

E.) A point estimate for the mean height of all females who wear a size 8 shoe is given

by simply plugging in 8 into the sample regression equation. This gives us

ŷ = 55.725 + 1.267(8)

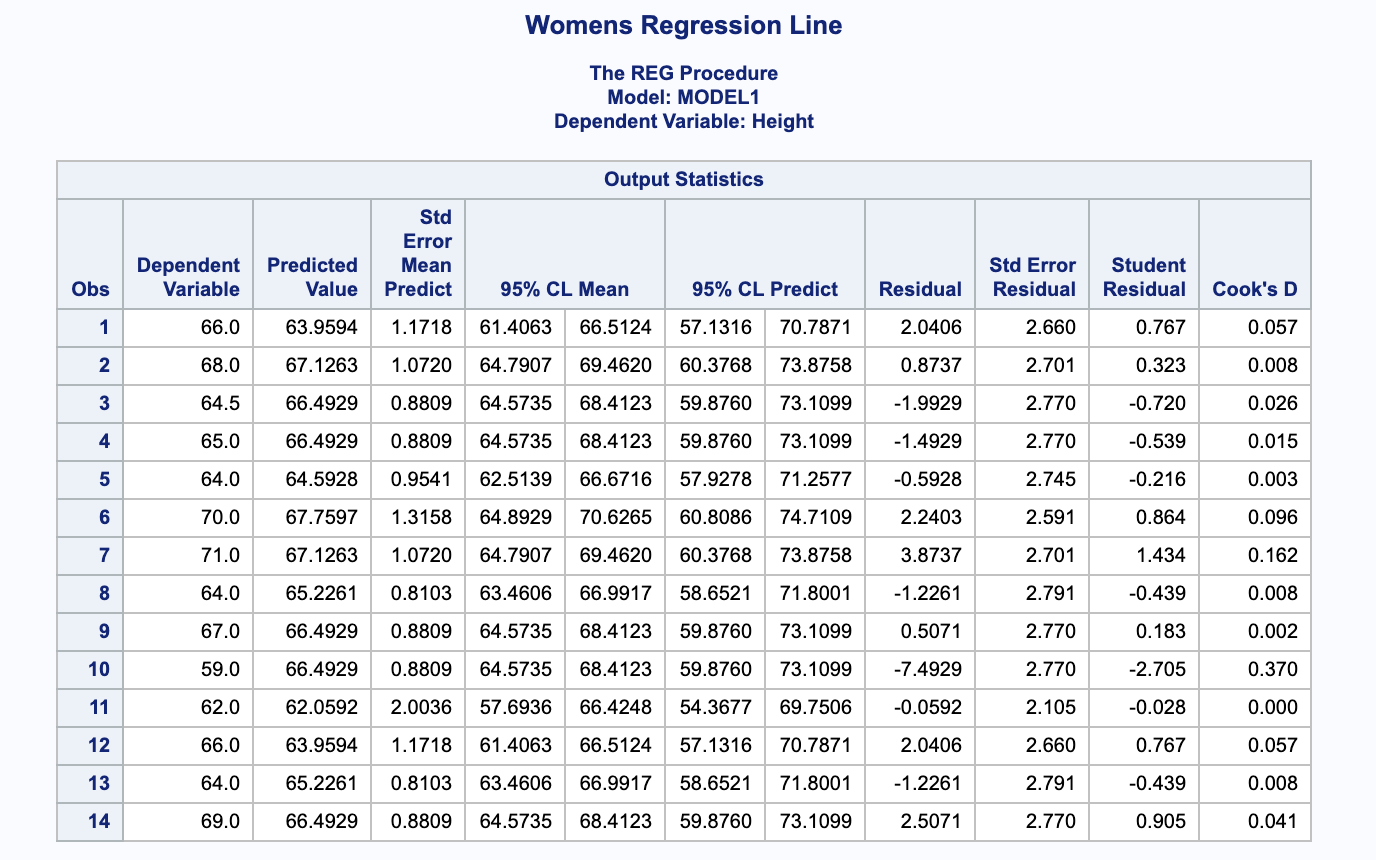
ŷ = 65.861 inches.

F.) Here we are asked to obtain a 95% confidence interval for the mean height of all

females who wear a size 8 shoe. The SAS code shown in part B once again provides

the output that we need in order to answer this question. The useful output that the

code provides us with is… (next page due to sizing)



where we can see that the 95% confidence interval for the mean height of all females

who wear a size 8 shoe is (64.16, 67.56). This means that we can be 95% confident

that the height of all females who wear size 8 shoes lies between 64.16 and 67.56

inches.

G.) The predicted height of a female who wears a size 8 shoe can be found by

plugging in 8 into the sample regression equation.

ŷ = 55.725 + 1.267(8)

ŷ = 65.861 inches.

H.) Looking at the same output used to find the 95% confidence interval in part f. we can

see that 95% prediction interval for the height of a female who wears a size 8 shoe is

(59.3, 72.42). We can be 95% confident that the height of females who wear a size 8

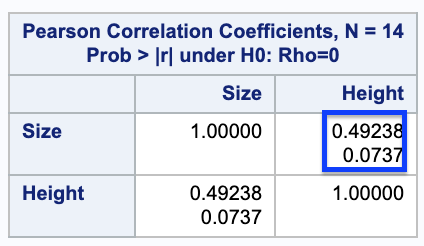
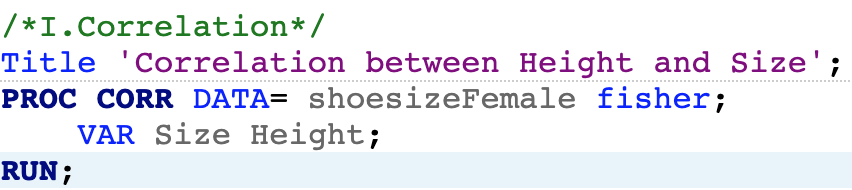
shoe lies between 59.3 and 72.42 inches.

I.) To find if the data provides sufficient evidence to conclude that shoe size and height

are positively linearly correlated, we used the proc corr method in SAS to calculate the

correlation value between Size and Height, and the P-value. The code is provided

below along with the output it generated…



As we can see in the boxed in region, we are provided with the correlation value of

0.49238, and a P-value of 0.0737. The P-value of the right-tailed test is 0.03685. This

gives us…

P-value (0.03685) < (0.05)

This means that we reject the Ho: p = 0, and we can conclude that at the 5%

significance level, the data does provide sufficient evidence to conclude that shoe size

and height are positively linearly correlated.

Conclusion

We have found that in both data sets, shoe size and height are positively linearly correlated, and that shoe size is not a good predictor variable for height. Both sample regression equations were simply computed, as were the confidence intervals. All code, data, and spreadsheets are attached to this write up. This project has helped me gain a better understanding of chapters 14 and 15, as well as gave me the chance to brush up on my SAS programming.