Discussion of Student Work for Unit 2 Project

(A) Introduction

Section A: Introduction

In this project, I will be analyzing the relationship between shoe size (x) and height (y) using data collected by asking female classmates and family their measurements. This data will be recorded on a Google Spreadsheet. I will be using frequency tables to organize my data, scatter plots to graph my data, and lines of best fit to determine what kind of correlation exists between the two variables. My prediction for the relationship is that there is a positive correlation between height and shoe size. Meaning, as a female's height increase, her shoe size will also increase, and vice versa. Using a scatter plot graph, calculating the correlation coefficient, and finding a line of best fit will allow me to come to a conclusion.

Frequency Table

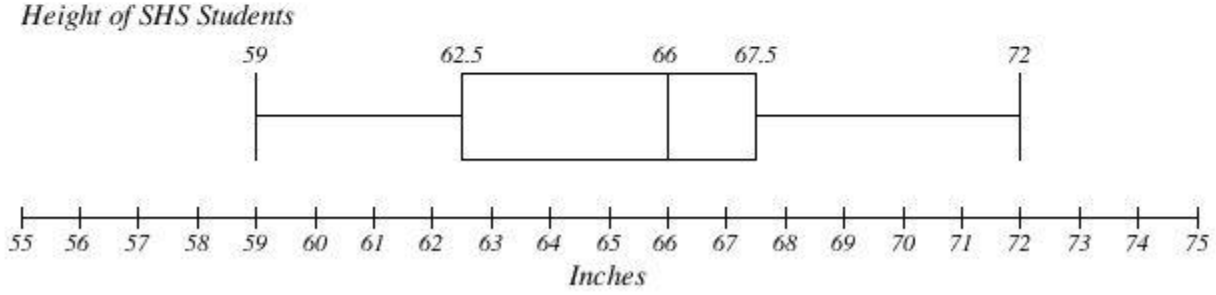
Shoe Size	Tallies	Frequency
6.5	I	1
7		0
7.5		0
8		0
8.5		0
9	I	1
9.5	I	1
10	IIIII	5
10.5	IIII	4
11	II	2
11.5	I	1
12	II	2
12.5		0
13	II	2
14	I	1
Totals	20	20

Five-Number Summary

Five Number Summary for Height:

59,60,61,61,62 || 63,63,65,65,65 || 67,67,67,67,67 || 68,68,69,69,72

Min = 59 in
Q1 = (62 + 63)/2 = 62.5 in
Median = (65 + 67)/2 = 66 in
Q3 = (67 + 68)/2 = 67.5 in
Max = 72 in



Finding Correlation Coefficient

Finding the correlation coefficient

$$r = \frac{n\times\sum xy - \sum x \sum y}{\sqrt{[n\times\sum x^2 - (\sum x)^2][n\times\sum y^2 - (\sum y)^2]}}$$

$$r = \frac{20\times 10434 - 1305\times 158.8}{\sqrt{[20\times 85379 - (1305)^2][20\times 1311.75 - (158.5)^2]}}$$

$$r = \frac{208680 - 207234}{\sqrt{[1707580 - (1703025)\,][26235 - (25122.25)]}}$$

$$r = \frac{1466}{\sqrt{[4555][1112.75]}}$$

$$r = 0.651$$

Significance Level (α)

- **α = 0.10:** You are 90% confident in the significance of your results.
- **α = 0.05:*** You are 95% confident in the significance of your results.
- **α = 0.01:** You are 99% confident in the significance of your results.

* Most Commonly Used

For Your Project

- Degrees of Freedom = n – 2 = … ?

df = *n* -2				
Level of Significance (p) for Two-Tailed Test	**.10**	**.05**	**.02**	**.01**
df				
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
15	.412	.482	.558	.606
16	.400	.468	.542	.590
17	.389	.456	.528	.575
18	.378	.444	.516	.561
19	.369	.433	.503	.549
20	.360	.423	.492	.537
21	.352	.413	.482	.526

Finding Correlation Coefficient

Finding the correlation coefficient

$$r = \frac{n\times\sum xy - \sum x \sum y}{\sqrt{[n\times\sum x^2 - (\sum x)^2][n\times\sum y^2 - (\sum y)^2]}}$$

$$r = \frac{20\times 10434 - 1305\times 158.8}{\sqrt{[20\times 85379 - (1305)^2][20\times 1311.75 - (158.5)^2]}}$$

$$r = \frac{208680 - 207234}{\sqrt{[1707580 - (1703025)\,][26235 - (25122.25)]}}$$

$$r = \frac{1466}{\sqrt{[4555][1112.75]}}$$

$$r = 0.651$$

(D) Interpretation of Results

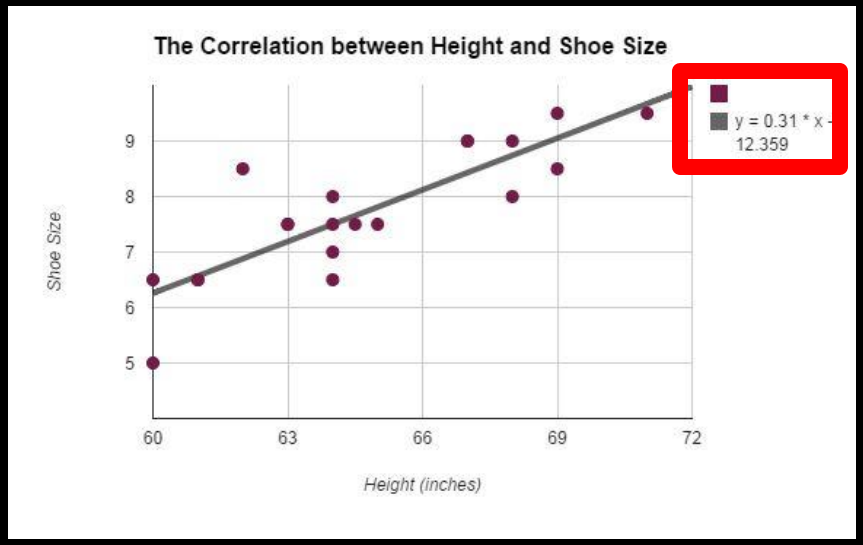
"Based on the *r* value of 0.651, using 18 degrees of freedom, my result is statistically significant at the 99% level because it's greater than the critical value of 0.561.

This means that based on my sample, there exists a positive relationship between height and shoe size among SHS's female population."

Coefficient of Determination

- The R^2 value.
- Explains the % of variation which can be explained by the regression equation.
- Calculate by squaring the correlation coefficient *r*.
- In general, the higher the R^2, the better the model fits the data.

R^2 in Action.



$$r = \frac{1214.75}{\sqrt{2093412.5625}}$$

$$r = \frac{1214.75}{1446.863}$$

$$r = 0.8396$$

- $R^2 = (0.8396)^2 = 0.7049$
- Thus, 70.49% of the variation between height and shoe size can be explained by the regression equation.
- The other 29.51% remains unexplained.
- Our regression line is a pretty good fit for the data. It can predict shoe size based on height pretty well.


J