Sehba Wani

INST314

HW9

High School GPA and University GPA

Calculate the correlation coefficient. Express the results properly in words. (5 pts)

H0: The true correlation is equal to zero

Ha: The true correlation is not equal to zero.

The two variables were correlated, and the relationship was statistically significant where r(103)= 0.77918 , p = 2.2e-16. The relationship is statistically significant.

**Assumptions for Correlation Analysis:**

* Variables quantitative (variables must be quantitative)

We assume and can see that the data is quantitative as they are ratios/proportions.

* Linear Relationship

Chart

Description automatically generatedWe assume that they do have a linear relationship.

* Normality

Based on the qqplot shown above, we assume that the data is not normally distributed. There was no data transformation done to rectify that.

* Related pairs

We can assume that the data is not paired.

* Chart, box and whisker chart

  Description automatically generatedNo outliers

We can assume that there are outliers based on the boxplot shown of the residual of the University and High school GPA shown.

Create a scatter plot comparing the two. (5 pts)

Chart, scatter chart

Description automatically generatedThe scatterplot is not showing a strong correlation.

Do a simple linear regression. Include an assessment of assumptions and express your results in words and mathematically. (5 pts for assessing each assumption, 5 pts for the regression analysis and 5 pts for the correct reporting of the results)

Y’ = 1.09682 + 0.6748x

A = Baseline GPA

X = High school GPA rate

B = Coefficient that shows the trend of how gpa increases

Y = University GPA

This assumes the average expected gpa for university in terms of the high school gpa.

Assumptions of Regression:

* Linearity

There is no linear relationships between the data.

* Independence

We can assume that the variables are independent of each other due to the lack of closeness between the variables.

* Homoscedasticity

Chart, scatter chart

Description automatically generatedWe can assume that there is no homeodascity, due to the lack of variance shown in plot and a skewedness to the right.

* Normality

Chart

Description automatically generatedWe can assume that the data is normal due to the QQplot of the residuals indicating normal distribution.

Math SAT and Computer Science GPA

Graphical user interface, text, application

Description automatically generatedCalculate the correlation coefficient. Express the results properly in words.(5 pts)

H0: The true correlation is equal to zero

Ha: The true correlation is not equal to zero.

The two variables were correlated, and the relationship was statistically significant where r(103)= 0.6877209, p = 5.34e-16. The relationship is statistically significant.

Assumptions for Correlation Analysis:

* Variables quantitative

We can assume that the variables are quantitative due to the fact that the data we are using are proportions/ratios.

* Linear Relationship

Chart

Description automatically generatedThere is a linear relationship,

* Normality

We can assume that the data is not normally distributed,

Based on the qq plot provided, there isn’t normal distribution showing

And no data transformation occurred.

* Related pairs

We can assume that there are related pairs.

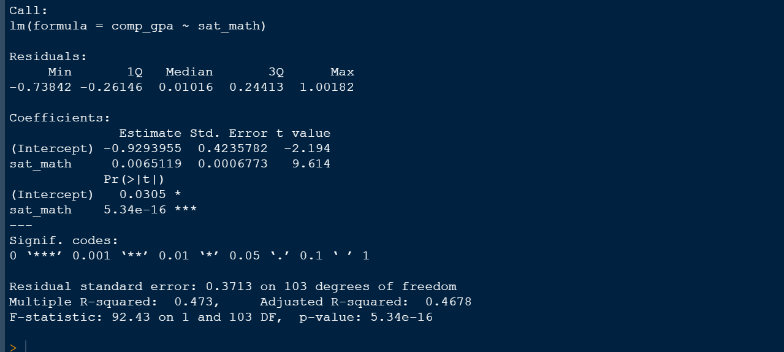
* No outliers

Chart, scatter chart

Description automatically generatedThe assumption of no outliers has been violated.

* 1. Create a scatter plot comparing the two. (5 pts)

The Scatterplot is shows a not so strong correlation.

Do a simple linear regression. Include an assessment of assumptions and express your results in words and mathematically. (5 pts for assessing each assumption, 5 pts for the regression analysis and 5 pts for the correct reporting of the results)

Y’ = 0.0305 + 0.0065119(SAT\_Math)

Y = Comp Sci GPA

A = Intercept, baseline comp sci gpa

B = Average increase of math score

X (sat math) = Specific Math SAT score

The line of best fit is attempting to establish a relationship between a computer science gpa,

* Linearity

There is no linear relationships between the math and sat seen in the scatterplot in question 2.

* Independence

We assume that there is independence between each variable, due to the lack of association among the data, there is not parametric data. The variables are separate of each other.

* Chart, scatter chart

  Description automatically generatedHomoscedasticity

We cannot assume that the data has homeocdascity.

* Normality

Chart

Description automatically generatedWe can assume that the data is not normally distributed. There was no data transformation performed.