DSC 423: Data Analysis and Regression Assignment 02

Saransh Thakur

Id: 2020070

**Problem 1**

1. R-squared or Coefficient of determination defines the variability in the dependent variable as explained by the independent variable. R-square lies between 0 and 1. The better model is closed to 1 and the poor model is closed to 0. R-squared is 0.69 which means 69 % of the variability in the dependent variable is explained by the independent variable. There is only a 69% chance that the data is close to the regression line. This is not at all a good performance score for a regression model.
2. The regression (or regressive) fallacy is an informal fallacy. It assumes that something has returned to normal because of corrective actions taken while it was abnormal.

For example:

1. When the patience pain got worse, he went to a doctor, after which the pain subsided a little. Therefore, he benefited from the doctor's treatment.
2. The accident on a road reduces after a speed camera was installed. Therefore, the speed camera has improved road safety.

**Problem 2**

Table

Description automatically generated

**Output: -**

**Model 1**

A screenshot of a computer

Description automatically generated with medium confidence

**Model 2**

**A screenshot of a computer

Description automatically generated with medium confidence**

**Model 3**

**A screenshot of a computer

Description automatically generated with medium confidence**

**Model 4**

**A screenshot of a computer

Description automatically generated with medium confidence**

**Model 5**

**A screenshot of a computer

Description automatically generated with medium confidence**

**For, model 1**

**A screenshot of a computer

Description automatically generated with medium confidence**

H0: B1= 0 (Null hypothesis)

H1: B1=/= 0 (Alternative Hypothesis)

For model 1, the B0 =112.115, B1 is -7.013, and the p-value is 0.735, greater than 0.05. So, we fail to reject the null hypothesis means that we don’t know if an independent variable should be used to predict the dependent variable.

R-squared is 0.005073 which means 0.5073% of the variability in the dependent variable is explained by the independent variable.

**For model 2**

**Text, letter

Description automatically generated**

H0: B1= 0 (Null hypothesis)

H1: B1=/= 0 (Alternative Hypothesis)

For model 2, the B0 = 665.77, B1 is 41.83, and the p-value is 0.316, greater than 0.05. So, we fail to reject the null hypothesis means that we don’t know if an independent variable should be used to predict the dependent variable. R-squared is 0.04365 which means 4.365 % of the variability in the dependent variable is explained by the independent variable.

**For model 3**

**A screenshot of a computer

Description automatically generated with medium confidence**

H0: B1= 0 (Null hypothesis)

H1: B1=/= 0 (Alternative Hypothesis)

For model 3, the B0 = -1978.21, B1 is 45.78, and the p-value is 0.363 which is greater than 0.05. So, we fail to reject the null hypothesis means that we don’t know if an independent variable should be used to predict the dependent variable.

R-squared is 0.03611 which means 3.611 % of the variability in the dependent variable is explained by the independent variable.

**For model 4**

A screenshot of a computer

Description automatically generated with medium confidence

H0: B1= 0 (Null hypothesis)

H1: B1=/= 0 (Alternative Hypothesis)

For model 4, the B0 = -667.31, B1 is 38.31, and the p-value is 0.0044 which is smaller than 0.05.

As the p-value is 0.44 % that means if the null hypothesis is true (B1 = 0) then there are 0.44 % chances that we observe this outcome So, we reject the null hypothesis and accept the alternative hypothesis which means an independent variable x has an impact on the dependent variable Y.

R-squared is 0.3024 which means 30.24 % of the variability in the dependent variable is explained by the independent variable.

**For model 5**

**A screenshot of a computer

Description automatically generated with medium confidence**

H0: B1= 0 (Null hypothesis)

H1: B1=/= 0 (Alternative Hypothesis)

For model 5, the B0 = 1263.64, B1 is 44.63, and the p-value is 0.001197 which is smaller than 0.05.

As the p-value is 0.1197 %, that means if the null hypothesis is true (B1 = 0), there are 0.1197 % chances that we observe this outcome. So, we reject the null hypothesis and accept the alternative hypothesis.

R-squared is 0.3724 which means 37.24 % of the variability in the dependent variable is explained by the independent variable.

1. The best model is model 5 then model 4, model 2, model 3, and model 1. Model 5 is the best because the regression model explains 37.24% rest frame Equivalent Width.

But, in general, it’s not considered a good model because there is only a 37.24% chance that the data is close to the regression line which means that the regression line is not fitted well. Hence, it’s not a good fit.

"I have completed this work independently. The solutions given are entirely my work."