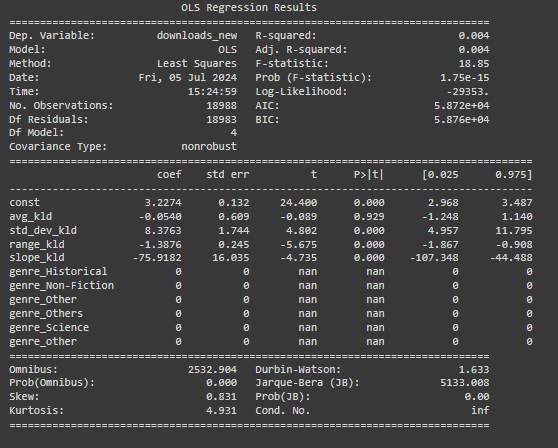
**Data Analysis**

Analysis was carried out to investigate the effect of information revelation and genre of the popularity of books. The linear regression model consisting of average, standard deviation, range and the slope of a linear regression across the course of the narrative did not explain much of the variation in the popularity (adj r-square = 0.04). The addition of the genre variable to the regression did not significantly improve this. The standard deviation range as and slope were significant predictors in the model with the higher slopes resulting in lower popularity levels.

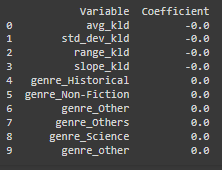
The Lasso regression further supported the findings from the linear regression as the coefficients were all zero showing that they were less significant predictors in the model. A negative R squared value obtained in the lasso regression also point to the poor fit of the model.

**Linear Regression Output without the Genre**



Variables: The variables include the average, standard deviation, range and slope.

**Lasso Regression Coefficient Output**



Variables

1. Genre: The genre was obtained using the NLP tf-idf procedure. The subjects were processed to make it easy to identify common words, which were then ranked.   
   Keywords that are common to genres were developed from online sources as shown below. , and thus were the used in identifying the groups to which the various books are likely to belong to.

