PUP INFLATION

A twitter handler, @dog_rates, provides ratings on cuteness of user's dog pictures. The answer to the question about grade inflation on the @dog_rates Twitter is answered by analyzing the data which consists of ratings, comments, time of comment. First, I gathered the useful data and extracted the ratings. After gathering ratings, I removed outliers (ratings that are too large). The result is plotted which is shown in the plot to the left. The blue dots show (time, ratings) plotting which is actual data and are represented in a scatter plot. The x -axis represents date (year) and y axis represents ratings. The linear fitting is applied on rating and timestamp to get the slope and intercept for best-fit line. The predicted values of ratings are calculated. The linear fit line is plotted in the same plot with red color which represents the relation between predicted ratings and timestamp values.

Then the answers to the question whether slope is zero is calculated by calculating p-value. A lower p-value claims that slope is not zero and higher p value claims the opposite. After plotting the regression fit line, its slope is calculated, and it is seen that the slope is not zero as the p value was less than 0.05. The resulting residuals (actual rating – predicted rating) were calculated and plotted on histogram as below. The x-axis represents residuals and y axis represents the count. The resulting residuals were close enough to being normal and due to positive slope, there is increase in the ratings.