Puromycin Practice

Haley Reed

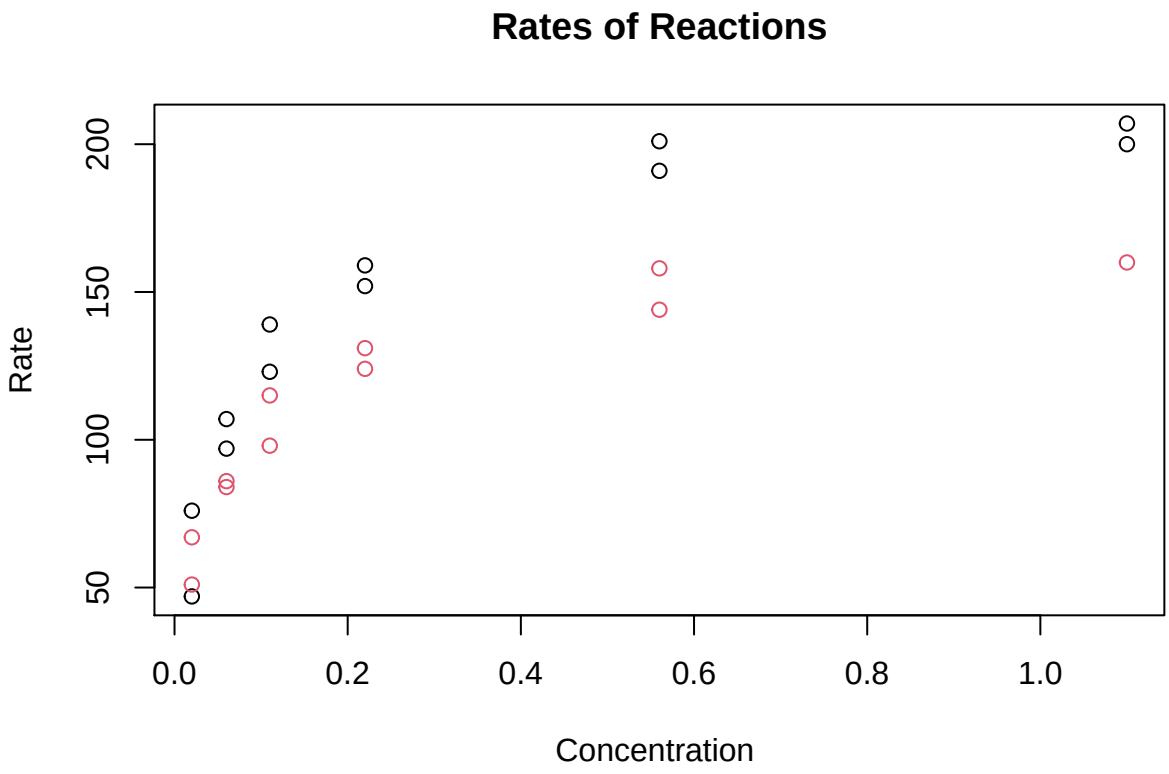
5/27/2020

Introduction

It is relevant to explore the effects and reaction times of Puromycin for its' application in biology. This report will use the data collected by Margret Treloar in 1974 in extermination with rat livers to explore these reaction times. Using this data obtained from R, this report will explore how substrate concentration (measured in parts per million) and the presence (or lack thereof) of the antibiotic Puromycin can act as explanatory variables and model the reaction rate of enzymes. The rate, our dependent variable, is measured by the number of counts per minute of radioactive product.

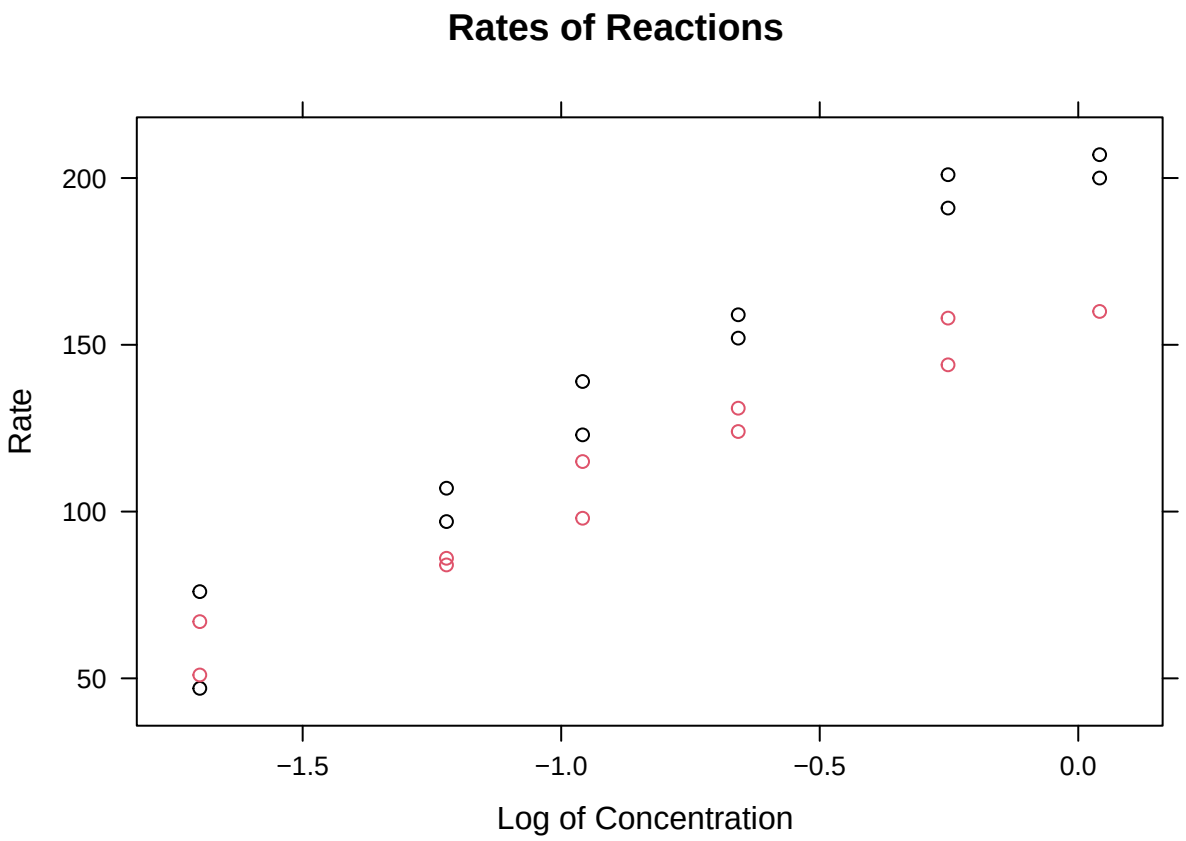
Analysis

To begin our exploration of the data, we start with a scatterplot of the rate represented by the explanatory variable concentration and differentiate treated with black pointed and untreated with red.



The data appears to be skewed to the right. The untreated data points appear to have reactions that may have statistically significantly slower reaction rates. The patterns present in this scatterplot demonstrate that

it would be helpful to build a model based on both concentration and state (treated or untreated) to model reaction rates. There appears to be a moderately strong positive relationship between the two variables of concentration, however the relationship appears to not be particularly linear. To address concerns of linearity, we will apply a logarithmic transformation to the numeric explanatory variable of concentration.



A logarithmic transformation of concentration results in a graph that looks more appropriate for a linear model.

From here, there are two major models worth discussing:

1. A Simple Linear Regression Model
2. An ANCOVA model

Even though the ANOVA model is slightly more complicated because of the added categorical value of state (treated versus untreated) we ultimately prefer it for the following reasons:

It has an adjusted multiple R-squared value of 0.9629 versus the Simple Linear Regression model's 0.9435. The ANCOVA model is able to explain 96.2902% of the variance in rate, making it a good option for model selection.

The ANCOVA model has an AIC value of 172.7171, a value 9.6658 units lower that the Simple Linear Regression model, meeting the statistical heuristic of significance for AIC because 9.6658>2.

The F-Test produces a P-Value of 2.41 × 10^(-6), a value well below the statistically significant 5%. This means we can reject the null hypothesis that the simpler model is better.

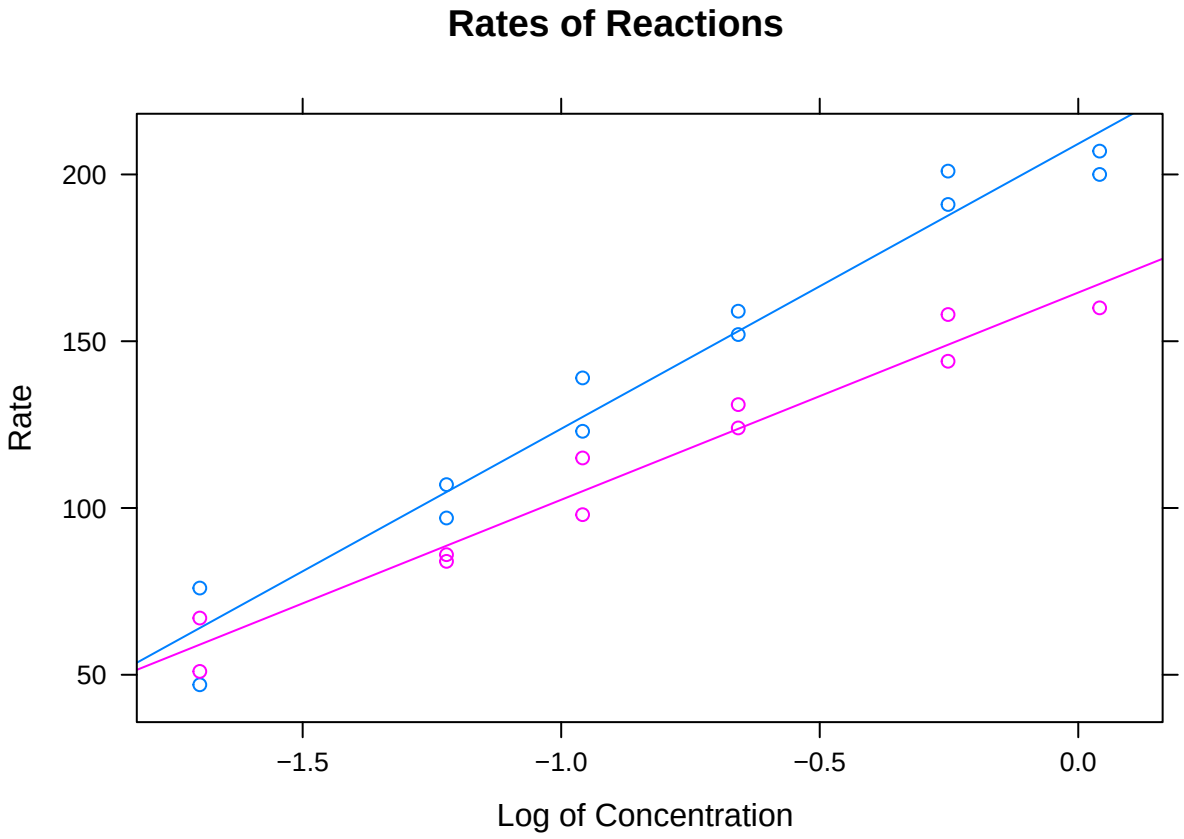
Therefore, we will go forward with the fitted model for ANCOVA given by:

$\mu[rate\ conc, state] = 209.1945 + 85.4499(logconc) + -23.3214(logconc) \times I(state = untreated) + -44.6061 \times I(state = untreated$

Since the primary difference between our two models is the differentiation between treated and untreated data, it is worth exploring the following confidence intervals:

	2.5 %	97.5 %
(Intercept)	199.87	218.516
logconc	75.97	94.934
stateuntreated	-58.86	-30.351
logconc:stateuntreated	-37.48	-9.167

The 95% confidence interval for the intercept of treated enzymes is (199.8734, 218.5156), meaning we are 95% confident the true mean of untreated enzymes with a concentration of one part per million is between 58.86 and 30.35 counts per minute slower than the true mean of treated enzyme with a concentration of one part per million.



Contextually, this means for a treated enzyme (shown in blue):

- When the concentration is 1, the model predicts a rate of 209.1945 counts per minute.
- When the concentration goes up by a factor of 10, the rate will increase by 85.4499 counts per minute.

And for an untreated Enzyme (shown in pink):

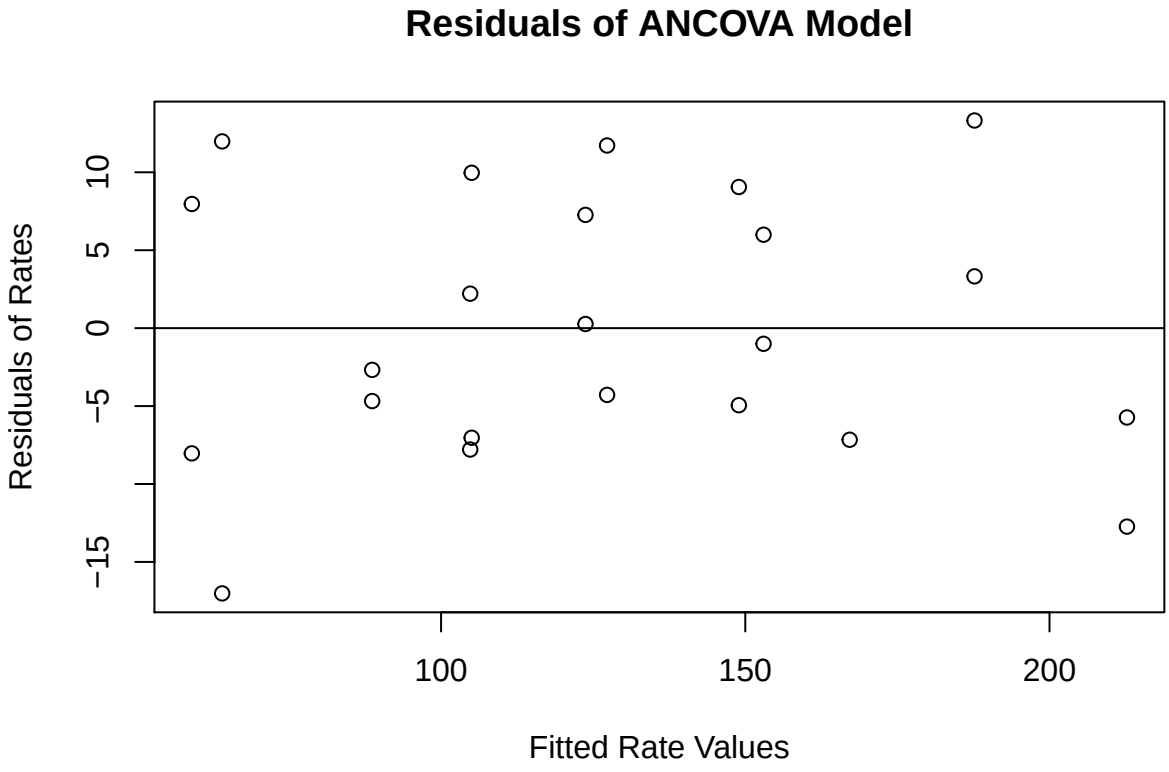
- When the concentration is 1, the model predicts a rate of 185.8731 counts per minute.
- When the concentration goes up by a factor of 10, the rate will increase by 40.8438 counts per minute.

For example, for a treated enzyme with a concentration of .4 parts per million, our model predicts a rate of between 175.2 and 195.5 counts per minute.

Sampling Variability Assumptions

We will check the following Sampling Variability Assumptions of our ANCOVA model:

1. The relationship between explanatory and dependent numerical variables is linear.
2. There is homogeneity of regression slopes.



- Residual plot appears linear, there is no obvious pattern and the scatter appears random on the residual plot. The residuals appear to be zero centered and have constant variance.
- This fulfills the first two Sampling Variability Assumptions.

3. There is no multicollinearity.

- There are no concerns over multicollinearity because we have only one numeric variable and one categorical variable. Third assumption is met.

Conclusion

Using Treloar's data, the ANCOVA model with the variable of concentration logarithmically transformed is an effective linear model for predicting the rate of reaction using the variables of concentration of the substrate and whether or not the enzyme was treated. With this we also found that untreated enzymes have statistically slower rates overall, and the rates increase in pace slower with increases in concentration than their treated

counterparts. We can use this model because it meets all sampling variability assumptions. Since this model is able to explain 96.2902% of the variation in reaction rates with the logarithmic transformation of concentration and the state (treated or untreated), it is an effective model with relatively high predicting power for exploring these reaction times. However; since these samples were pulled from enzymes within rat livers, inference to the effect of puromycin of rate of reactions for all cells is speculative. Without further biological evidence about how well the enzymes used in this experiment are representative of all cells, we should be careful to not broaden our findings and inferences drawn by the model beyond the enzymes used in Treloar's experimentation.

Works Cited

Silver, Nate. 2020. "Raw Polls." GitHub. https://github.com/fivethirtyeight/data/blob/master/pollster-ratings/raw-polls.csv.