

Lucy Lin  
Professor Finch  
AMS 315  
10 November 2020

## AMS 315 Project 1 Part B

### Introduction

The task for Part B is to find a transformation for the dependent or independent variables or both. There was no missing data in the 882 data values. A lack of fit test should be performed after the near-repeated data is binned. Then, the null hypothesis that there is no relationship between the variables will be tested. All these tasks were performed using R.

### Methodology

Various transformations of log, inverse, square root, and exponential were tested. The highest  $R^2$  value that was achieved was from 0.5761 in the original data to 0.5886 with the transformation of  $IV = \sqrt{IV}$  and  $DV = \ln(DV)$ . The transformed data was binned with an interval of 0.03 into 39 groups. The lack of fit test was performed using the alr3 package and there was no significant lack of fit.

### Results

Prior to the transformation, the  $R^2$  value was 0.5761 and the original fitted linear function was  $y = 96.683 + 223.031x$ . The confidence intervals of the slopes have changed due to the transformations. The  $R^2$  value after the transformation was 0.5886 (the fraction of variation in dependent variable explained) and the newly transformed fitted function was estimated to be  $\ln(y) = 1.18833\sqrt{x} + 4.59264$ . Additionally, the cor() function returned 0.7672019 for the transformed x and y values. The confidence intervals for the coefficient of  $\sqrt{x}$ , or the parameter estimate, for the fitted function was (1.122601, 1.254061) for 95% confidence and (1.101878, 1.274784) for 99% confidence. Since the p value is 0.7326 and the F value is 0.9021, the null hypothesis can be rejected. Thus there is no significant lack of fit. The ANOVA tables, summaries, confidence intervals, and plots are shown below.

### **Conclusion/Discussion**

The transformation applied to the original model improves the data slightly by increasing the  $R^2$  value. There was no significant lack of fit as tested. Although the improvement may not be extreme, the final model  $\ln(y) = 1.18833\sqrt{x} + 4.59264$  is a reasonable representation of the original data.

## Before Transformation

Table: ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
IV	1	10557042	10557042.479	1195.887	0
Residuals	880	7768457	8827.792	NA	NA

## Summary()

Call:

lm(formula = DV ~ IV, data = PartB)

Residuals:

	Min	1Q	Median	3Q	Max
	-304.61	-64.27	-7.68	61.36	394.72

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	96.683	13.292	7.274	7.73e-13 ***
IV	223.031	6.449	34.582	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 93.96 on 880 degrees of freedom

Multiple R-squared: 0.5761, Adjusted R-squared: 0.5756

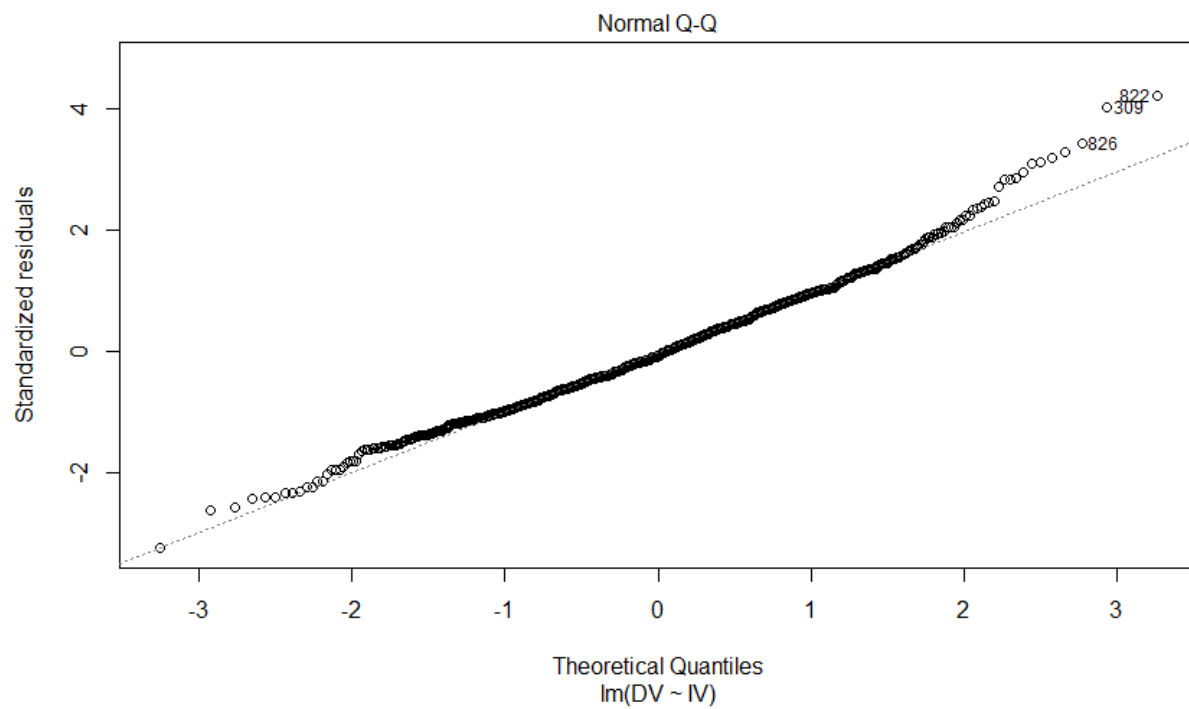
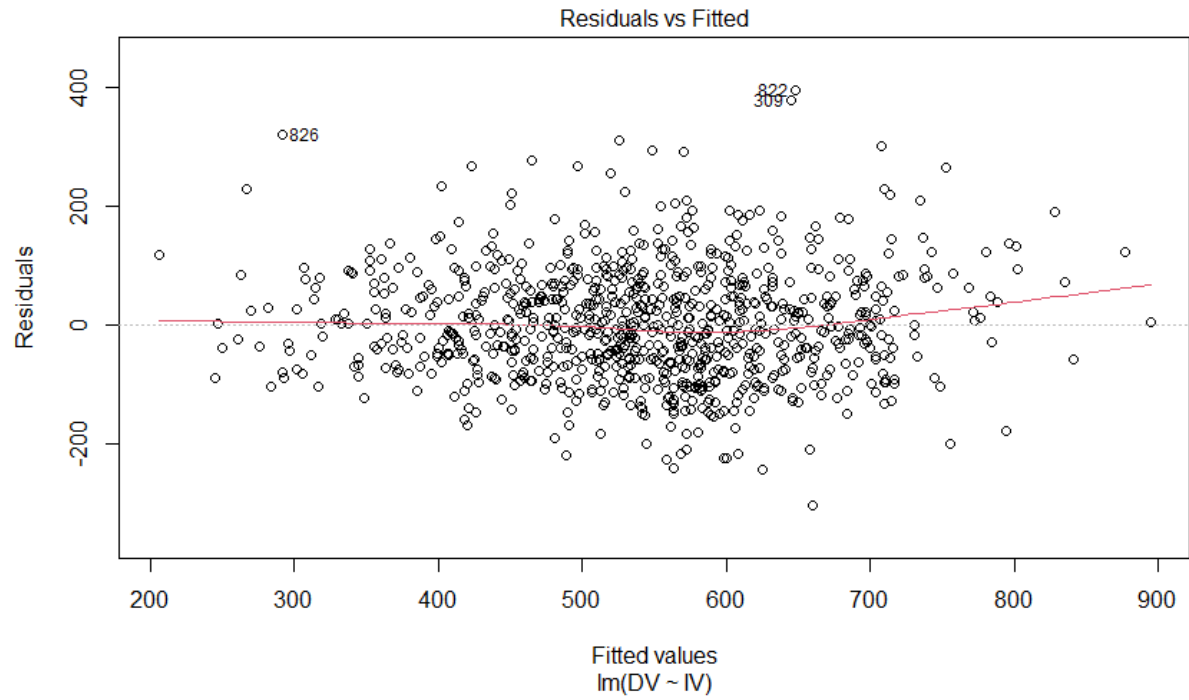
F-statistic: 1196 on 1 and 880 DF, p-value: < 2.2e-16

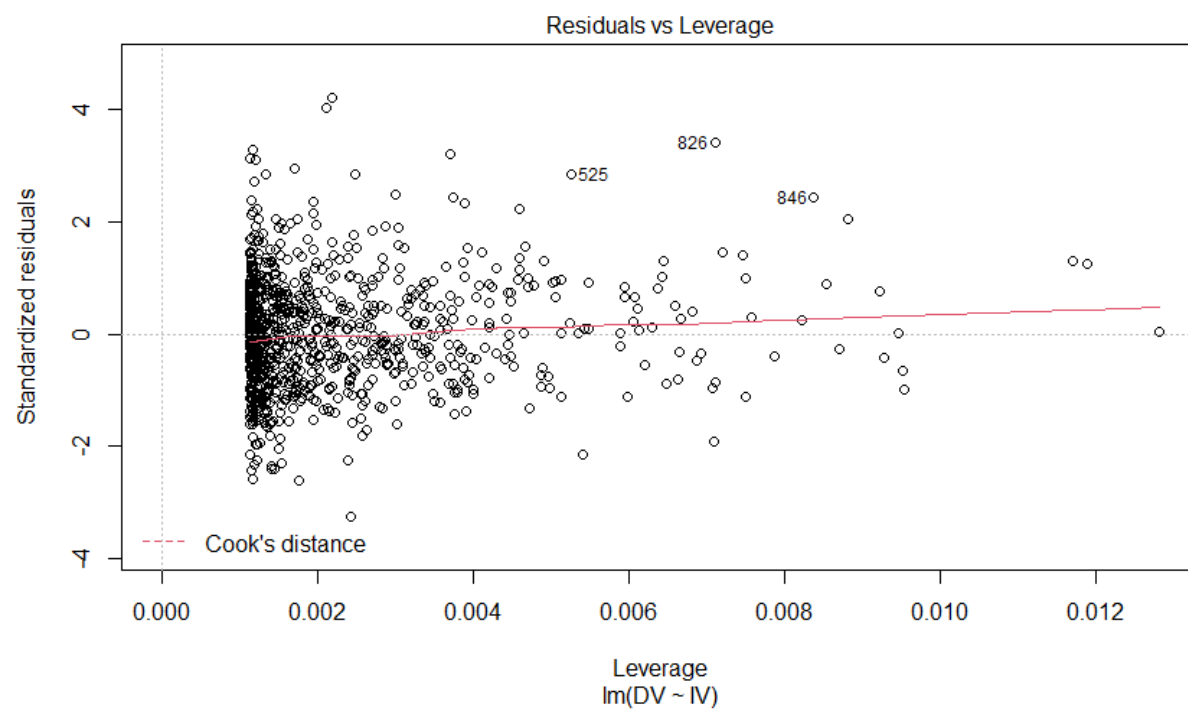
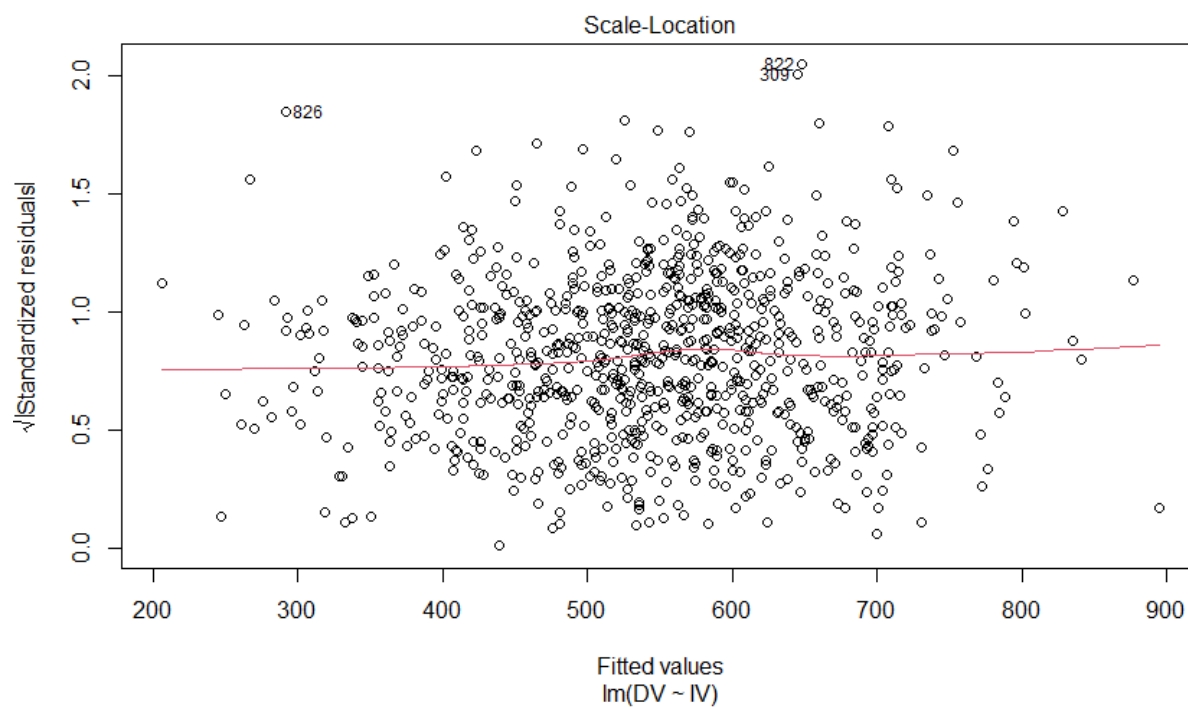
## Confidence Interval 95%

	2.5 %	97.5 %
(Intercept)	70.59556	122.7701
IV	210.37293	235.6890

## Confidence Interval 99%

	0.5 %	99.5 %
(Intercept)	62.37108	130.9946
IV	206.38226	239.6796





## After Transformation

### Analysis of Variance Table (PureErrorAnova())

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x	1	40.547	40.547	1246.9843	<2e-16 ***
Residuals	880	28.312	0.032		
Lack of fit	95	2.787	0.029	0.9021	0.7326
Pure Error	785	25.525	0.033		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Call:

lm(formula = DVtrans ~ IVtrans, data = data\_trans)

Residuals:

	Min	1Q	Median	3Q	Max
	-0.6082	-0.1150	0.0024	0.1201	0.7122

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.59264	0.04738	96.93	<2e-16 ***
IVtrans	1.18833	0.03349	35.48	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1794 on 880 degrees of freedom

Multiple R-squared: 0.5886, Adjusted R-squared: 0.5881

F-statistic: 1259 on 1 and 880 DF, p-value: < 2.2e-16

### Confidence Interval 95%

	2.5 %	97.5 %
(Intercept)	4.499641	4.685633
IVtrans	1.122601	1.254061

### Confidence Interval 99%

	0.5 %	99.5 %
(Intercept)	4.470322	4.714951
IVtrans	1.101878	1.274784

