Homework #3

Dustin Leatherman

September 29, 2019

Contents

1 2.23

1.1 d

> What is the absolute magnitude of the reduction in the variation of Y when X is introduced into the regression model? What is the relative reduction? What is the name of the latter measure?

The absolute magnitude of the reduction in the variation of Y when X is introduced is $SSR = 3.588$.

The relative reduction in variance is $R^2 = 0.0726$.

1.2 e

Obtain r and attach appropriate sign

$r = \sqrt{R^2} = \sqrt{0.0726} = 0.2695$

1.3 f

Which measure, R^2 or r has the more clear-cut operational interpretation?

R^2 dictates the variance in Y explained by X and r measures the linear association between Y and X. From an operational perspective, R^2 is more clear-cut as there is less interpretation required than the correlation coefficient in which the sign is important.

2 2.26

2.1 c

> Plot the deviations $Y_i - \hat{Y}_i$ against X_i on a graph. Plot the deviations $\hat{Y}_i - \bar{Y}$ on another graph. Does SSE or SSR appear to be the larger component of SSTo? What does this imply about the magniture of R^2?

According to the two graphs, SSR appears to be a larger component of SSO. This would mean that the magnitude of R^2 is high.

2.2 d

Calculate R^2 and r

```
summary(plastic.model1)

#Call:
#lm(formula = Hardness ~ Hours, data = plastic)

#Residuals:
#    Min      1Q  Median      3Q     Max
#-5.1500 -2.2188  0.1625  2.6875  5.5750

#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
```

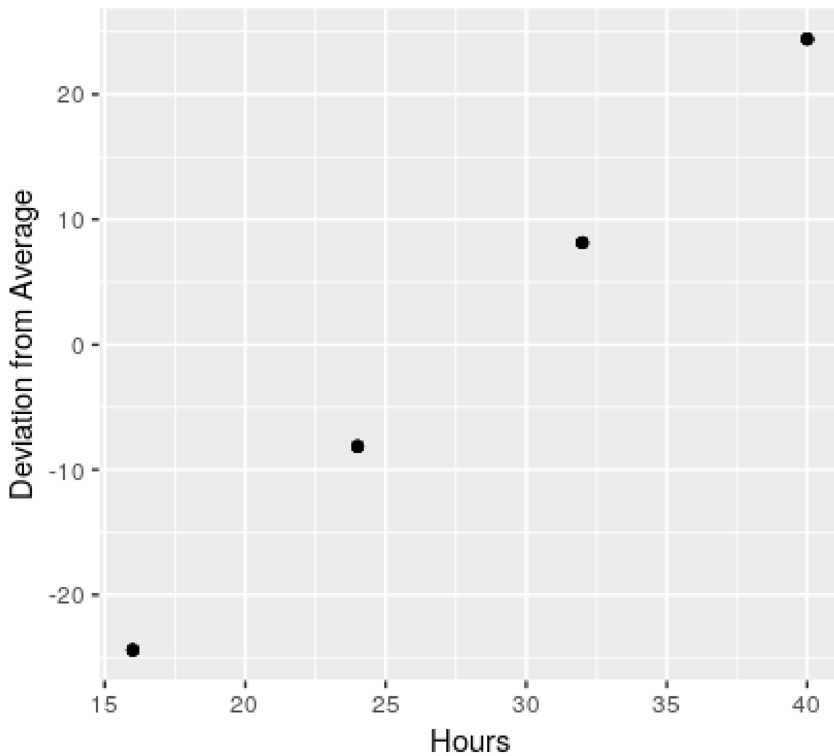

Figure 1: $Y_i - \bar{Y}$

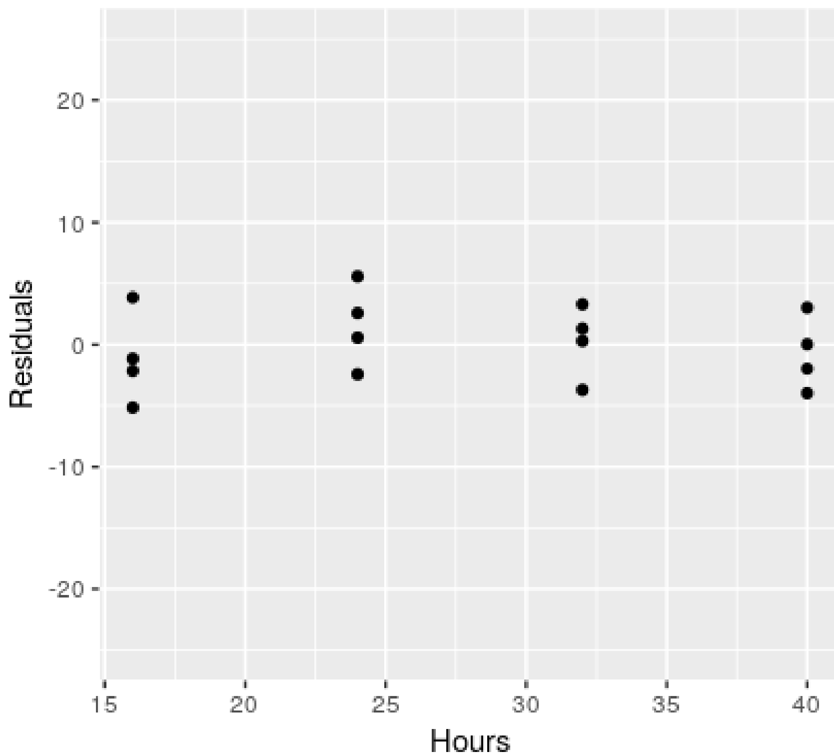

Figure 2: Residuals

```
#(Intercept) 168.60000    2.65702   63.45  < 2e-16 ***
#Hours         2.03438    0.09039   22.51 2.16e-12 ***
#---
#Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 3.234 on 14 degrees of freedom
#Multiple R-squared:  0.9731,Adjusted R-squared:  0.9712
#F-statistic: 506.5 on 1 and 14 DF,  p-value: 2.159e-12
```

$R^2 = 0.9731$
$r = \sqrt{R^2} = 0.9865$

3 3.2

Prepare a prototype residual plot for each of the following cases:

1. error variance decreases with X
2. true regression function is $\cup$ shaped but a linear regression function is fitted

4 3.6

4.1 a

Obtain the residuals and prepare a boxplot. What information is provided?

The boxplot shows the median, quartiles, and min/max of the residual values from the SLR Model: $\hat{Y_H} = 168.6 + 2.0344 \times \text{Hours}$. This shows that over 50% of Residuals lie between -3 and 3 with the 50th percentile laying near 0.

4.2 b

Plot the residuals against the fitted values to ascertain whether any departures from the regression model are evident. State your findings

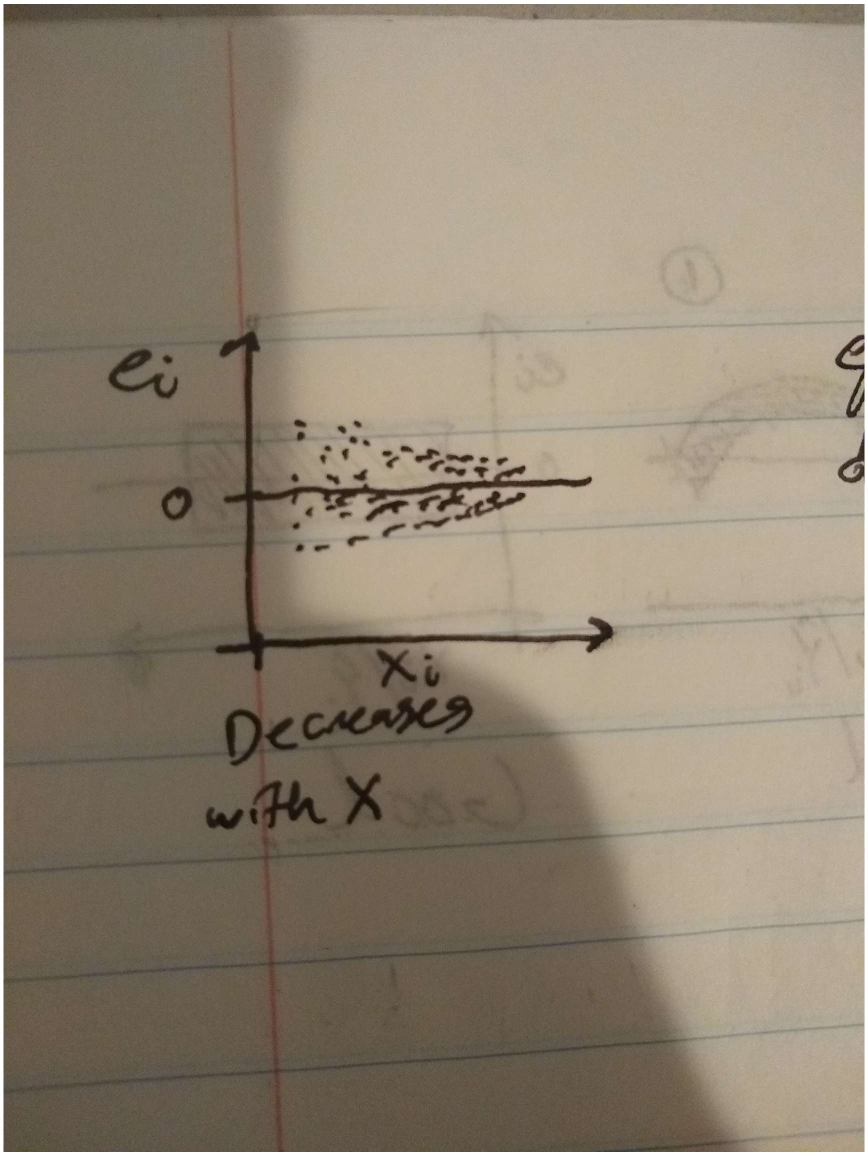

Figure 3: Error Variance decreases with X

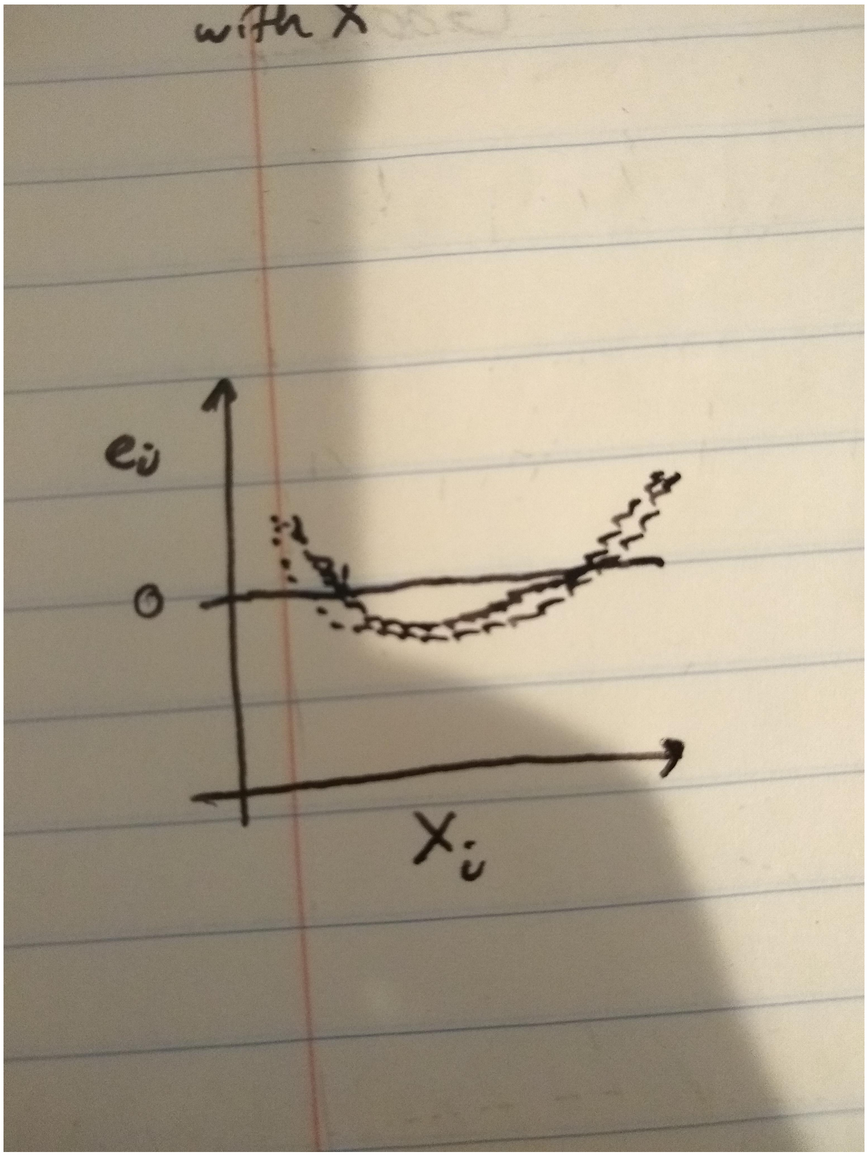

Figure 4: True Regression is Parabolic

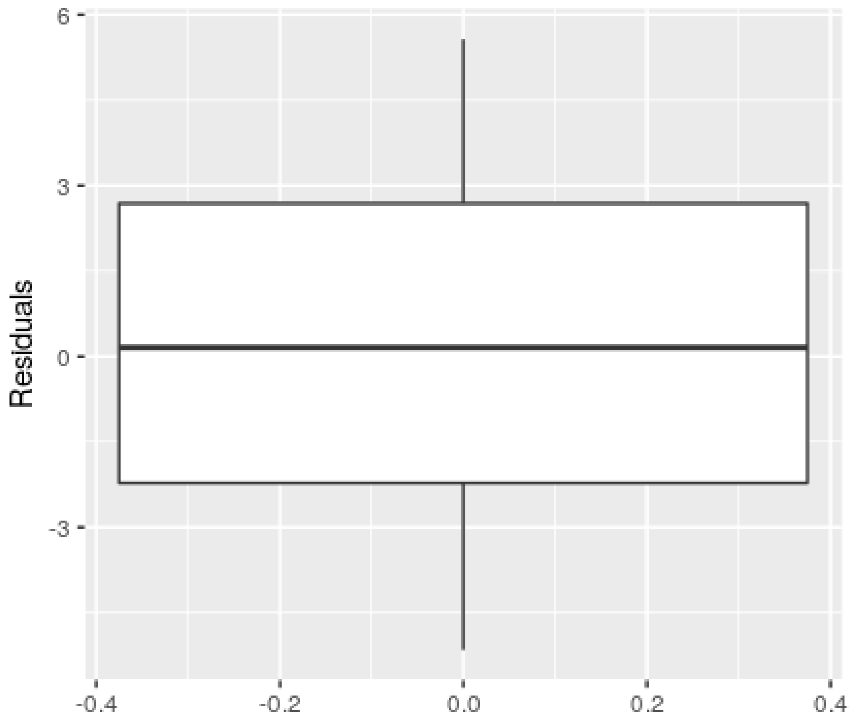

Figure 5: Residual Boxplot

There is a noticable wave-like pattern with the residuals that starts below $Y = 0$ and $X = 201.15$, rises above $Y = 0$ at $X = 217.425$, and dips below again at the X values afterwards. This indicates that the regression function is **not** linear, and that the error terms may not be independent.

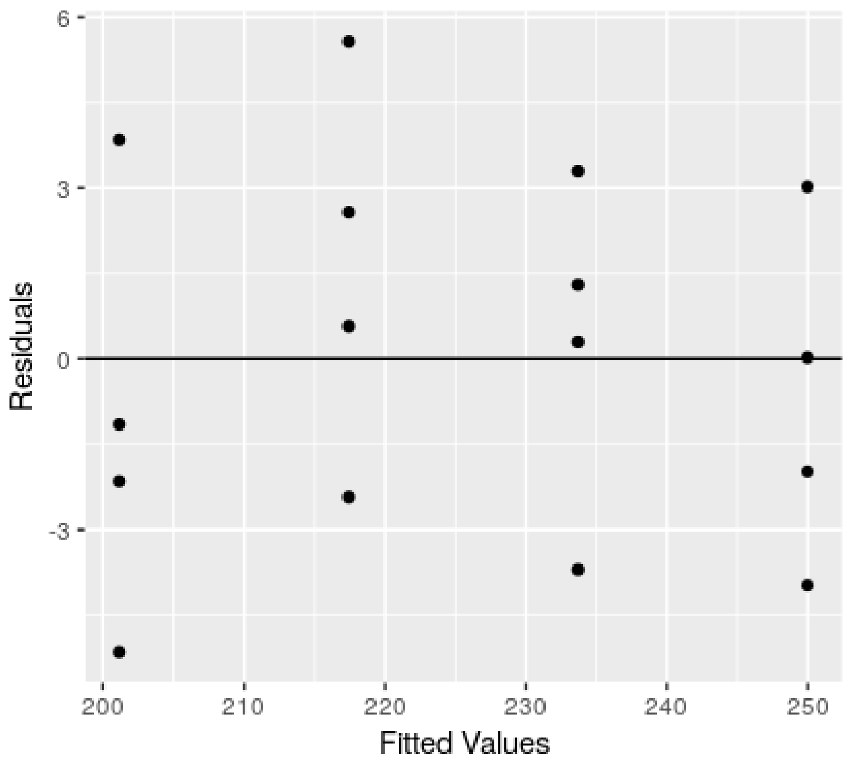

Figure 6: Residuals vs Fitted Values

4.3 c

> Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

There is not enough evidence to suggest that the residuals are not normal (Shapiro-Wilk. p-value = 0.8914). The Normal Probability plot supports this conclusion as well. Thus the normality assumption appears to be reasonable.

```
shapiro.test(x = plastic.model1$residuals)
# Shapiro-Wilk normality test

#data:  plastic.model1$residuals
#W = 0.97348, p-value = 0.8914
```

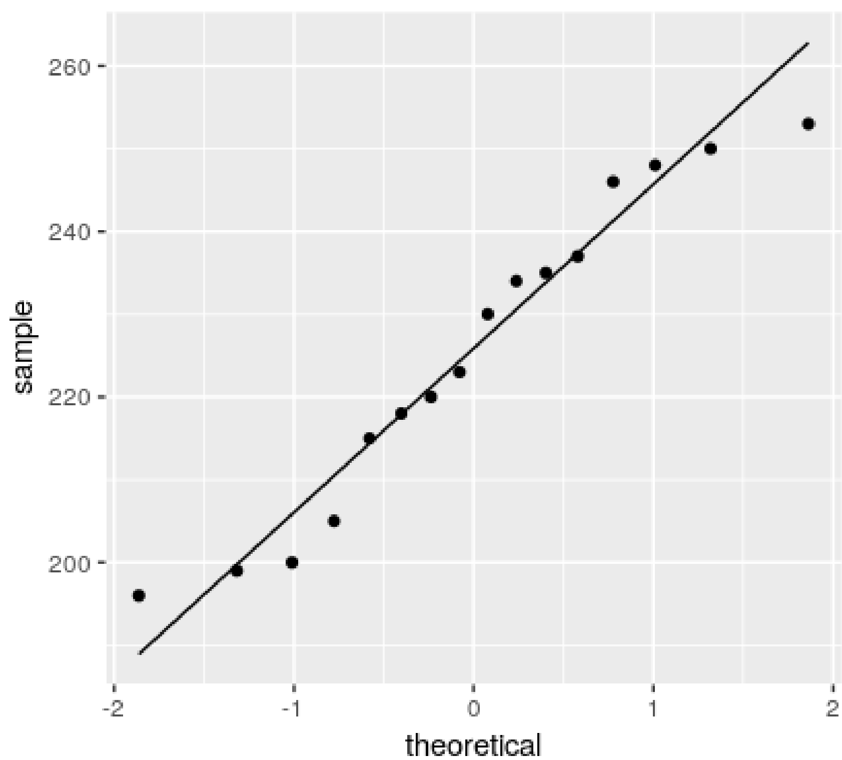

Figure 7: Normal Probability Plot

4.4 e

Use the Brown-Forsythe test to determine whether or not the error variance varies with the level of X. Divide the data into 2 groups, X $<=$ 24 and X$>$ 24. State the decision rule and conclusion. Does your conclusion support your preliminary findings in b?

There is not enough evidence to suggest that data does not have constant variance (Brown-Forsyth Levene's Test. p-value = 0.04065). Examining the residual plot in B showed that variance was more or less constant so this confirms the initial conclusion.

```
plastic$Group <- plastic$Hours <= 24
levene.test(y = plastic.model1$residuals, group = plastic$Group)

# Modified robust Brown-Forsythe Levene-type test based on the absolute deviations fro

#data:  plastic.model1$residuals
#Test Statistic = 0.73237, p-value = 0.4065
```