Chapter 3 HW

Ahmad M. Osman

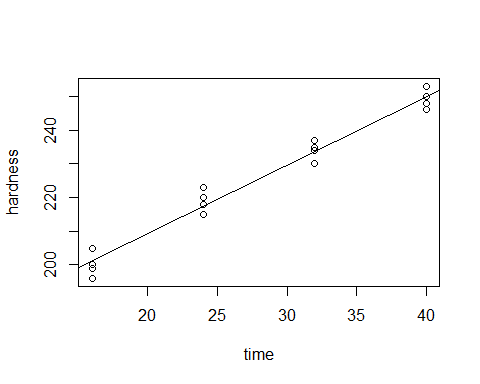
9/25/2017

Chapter 3 - Problem 3.6

**Includes answers to:**

* a, b, and c

##   
## Call:  
## lm(formula = hardness ~ time)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.1500 -2.2188 0.1625 2.6875 5.5750   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 168.60000 2.65702 63.45 < 2e-16 \*\*\*  
## time 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

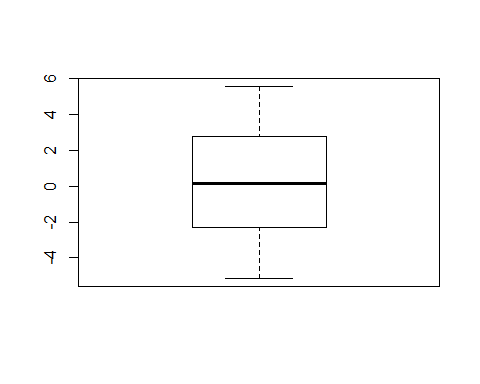
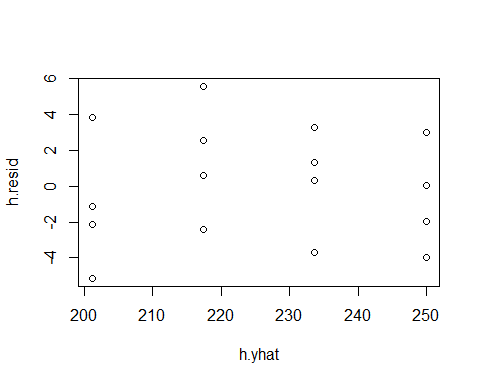


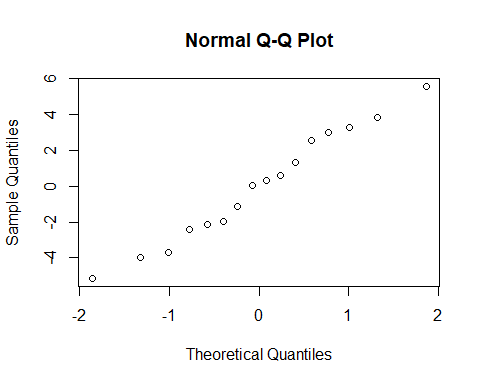
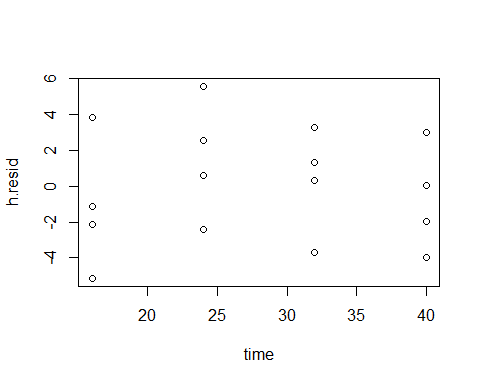
##   
## R SQUARED:

## [1] 0.9731031

##   
## CORRELATION COEFFICIENT:

## [1] 0.9864599





**Answers Explained:**

**3.6**

**a)** The residuals are

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| -2.150 | 3.850 | -5.150 | -1.150 | 0.575 | 2.575 | -2.425 | 5.575 | 3.300 | 0.300 | 1.300 | -3.700 | 0.025 | -1.975 | 3.025 | -3.975 |

Refer to the knitted section for the boxplot created. The boxplot shows how the residuals spread; which show constant variance and normal spread.

**b)** Refer to the knitted section for the graph of the residuals against the fitted values (Ŷ). From the graph, there doesn’t seem to be any outliers. Because the residual values are, relatively speaking, evenly spread out between negative 4 and positive 4, the maximum and minimum values are not extreme enough to be considered outliers. However, if the plot was of the semistudentized residuals, then the rule would have been that if the absolute value of any residual is greater than 4, then it can be considered an outlier. And since this plot is just of the simple residuals, that does not apply.

**c)** Refer to the knitted section for the normal probability plot of the residuals (Normal Q-Q Plot). From the graph, the normality assumption appears to be reasonable. Comparing the plot from this question to the plots on page 112 of the book, this question’s plot does not have any signs of abnormality.

Chapter 3 - Problem 3.11

**Includes answers to:**

* a and b

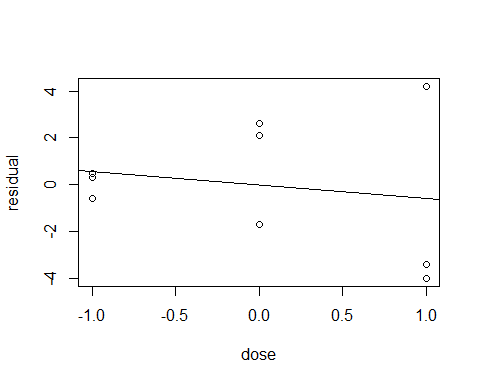
##   
## Call:  
## lm(formula = residual ~ dose)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4333 -1.7000 -0.2667 2.1000 4.7667   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.0000 0.9598 0.000 1.000  
## dose -0.5667 1.1755 -0.482 0.644  
##   
## Residual standard error: 2.879 on 7 degrees of freedom  
## Multiple R-squared: 0.03213, Adjusted R-squared: -0.1061   
## F-statistic: 0.2324 on 1 and 7 DF, p-value: 0.6445

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

##   
## studentized Breusch-Pagan test  
##   
## data: hfit  
## BP = 6.0215, df = 1, p-value = 0.01413



**Answers Explained:**

**3.11**

**a)** Refer to the knitted section above to see the plot of *ei* against X*i*. As the dose values go from -1 to 1, the absolute value of the residuals gets bigger, lead us to think of it as a non-constant variance.

**b)** The Constancy of Error Variance corresponds to having y1 = 0. The alternatives in our problem is *H0*: *y1* = 0 and *Ha*: *y1* ≠ 0. The rule that we should follow (page 119 from the books) says the following:

If X2BP ≤ x2(1 – α, *df*), then conclude *H0,* and if X2BP > x2(1 – α, *df*), then conclude *Ha.*

In this case, x2(1 – α, *df*) = 3.84. The BP value from the BP test function, is 6.0215. And since 6. 0215 > 3.84, we conclude *Ha, and* that the error variance is not constant. The P-value of the test is 0.01413, which is relatively small and is an evidence of the lack of constant variance. From this test, we can confirm the fidings in part (a) of this question suggesting non-constant variance.

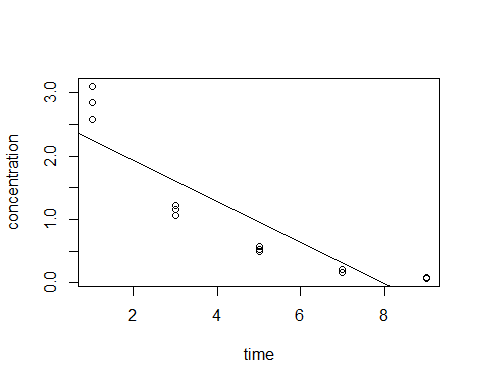
Chapter 3 - Problem 3.15

**Includes answers to:**

* a, b, and c

##   
## Call:  
## lm(formula = concentration ~ time)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.5333 -0.4043 -0.1373 0.4157 0.8487   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.5753 0.2487 10.354 1.20e-07 \*\*\*  
## time -0.3240 0.0433 -7.483 4.61e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4743 on 13 degrees of freedom  
## Multiple R-squared: 0.8116, Adjusted R-squared: 0.7971   
## F-statistic: 55.99 on 1 and 13 DF, p-value: 4.611e-06

## (Intercept) time   
## 2.575333 -0.324000



## Analysis of Variance Table  
##   
## Model 1: concentration ~ time  
## Model 2: concentration ~ 0 + as.factor(time)  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 13 2.9247   
## 2 10 0.1574 3 2.7673 58.603 1.194e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 38.50633

**Answers Explained:**

**3.15**

**a)** Refer to the knitted section above to see the plot. The function is   
Ŷ = -0.324000(X*i* ) + 2.575333

**b)** The F test alternatives, due to the lack of fit, are *H0*: *E{Y} = β0 + β1X* and *Ha*: *E{Y} ≠ β0 + β1X*. The rule we follow is

if F\* ≤ F(1 – α; c -2, n -c), conclude *H0* and if F\* > F(1 – α; c -2, n -c), conclude *Ha*.

In our case, c = 3 for the 3 subgroups, n = 5 for the number of observations in each subgroup, and α = 0.025. Through R, F() = F(0.975, 1, 2) = 38.50633. The F\* value from the ANOVA table is 58.603. Since 58.603 > 38.50633, we conclude *Ha, and* that the regression function is not a linear one. The P-value for this test is equal to 1.194e-06.

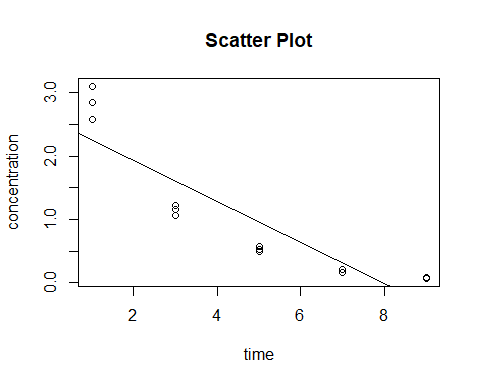
**c)** On page 127 from the book, comment number 5 states that “The alternative *Ha* includes all regression functions other than a linear one. For instance, it includes a quadratic function or a logarithmic one. If *Ha* is concluded, a study of residuals can be helpful in identifying an appropriate function”. The F\* test for lack of fit only points out that the appropriate regression function is not linear, but does not suggest a more appropriate function or model to follow/use.

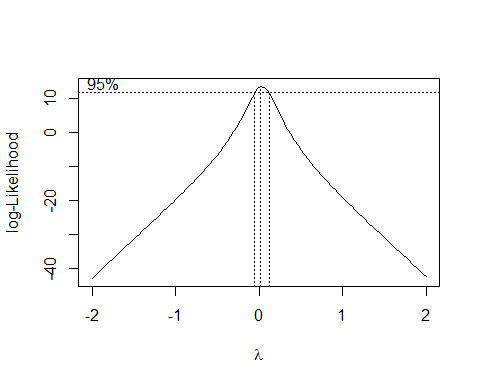
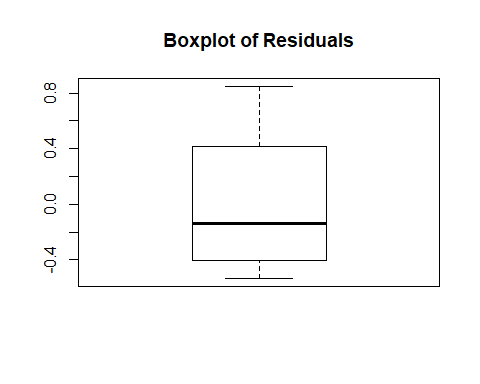
Chapter 3 - Problem 3.16

**Includes answers to:**

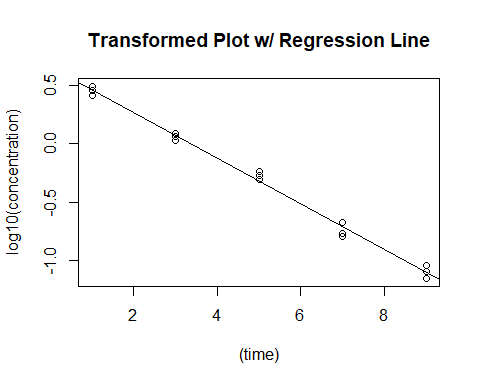
* a, b, c, d, e, and f

##   
## Call:  
## lm(formula = concentration ~ time)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.5333 -0.4043 -0.1373 0.4157 0.8487   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.5753 0.2487 10.354 1.20e-07 \*\*\*  
## time -0.3240 0.0433 -7.483 4.61e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4743 on 13 degrees of freedom  
## Multiple R-squared: 0.8116, Adjusted R-squared: 0.7971   
## F-statistic: 55.99 on 1 and 13 DF, p-value: 4.611e-06

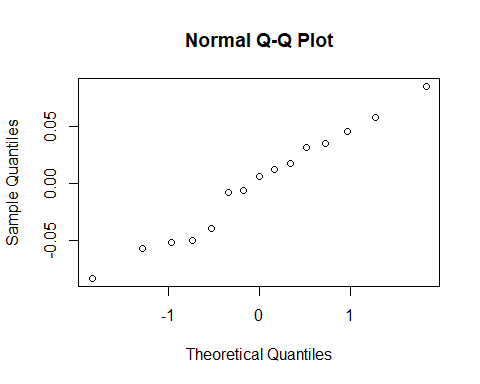
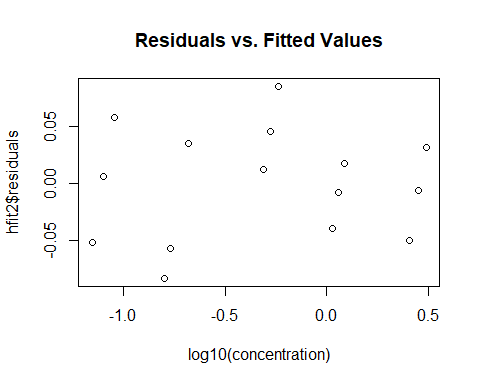
## (Intercept) time   
## 2.575333 -0.324000



## (Intercept) time   
## 0.6548798 -0.1954003



## 1 2 3 4 5   
## -0.051178946 0.057965523 0.006813001 -0.082957620 -0.056628681   
## 6 7 8 9 10   
## 0.035141692 0.012317861 0.085549775 0.046397651 0.017680995   
## 11 12 13 14 15   
## -0.007980995 -0.039295058 -0.006161112 -0.049546328 0.031882242



**Answers Explained:**

**3.16**

**a)** Refer to the knitted section above to see the scatter plot of the data. From Figure 3.15, we can learn that the best transformation on Y would be Y’ = log10 Y; in order to achieve linearity and constant variance.

**b)** Referring to the knitted section above, from the Box-Cox plot, it is clear that the best choice for λ is 0 which corresponds to Y’ = log10 Y.

**c)** By using the transformation Y’ = log10 Y, the estimated linear regression function for the transformed data is equal to Ŷ ‘ = -0.1954003 X + 0.6548798.

**d)** Refer to the knitted section above to see the plot of the estimated regression line and the transformed data. The regression line seems to be a good linear fit for the transformed data.

**e)** The residuals of the transformed data:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| -0.051178946 | 057965523 | 0.006813001 | -0.082957620 | -0.056628681 | 0.035141692 | 0.012317861 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 0.085549775 | 0.046397651 | 0.017680995 | -0.007980995 | -0.039295058 | -0.006161112 | -0.049546328 | 0.031882242 |

Refer to the knitted section above to see the plot of the residuals against the transformed fitted values. Also refer to the knitted section above to see the normal probability plot.

The plot of the residuals against the fitted values shows the variance is much more constant and much closer to 0, which means that the transformation was successful in reaching and achieving a constant variance. The normal probability plot shows a much more linear plot, which shows the transformation was successful in reaching and achieving linearity.

**f)** The estimated regression function in the original units is equal to log10 (Ŷ) = -0.1954003 X + 0.6548798.