Math 3080 § 1.	**Pumpkin Example:**	Name: Example
Treibergs	**Flaws in Diagnostics: Correcting Models**	March 1, 2014

From Levine Ramsey & Smidt, *Applied Statistics for Engineers and Scientists,* Prentice Hall, Upper Saddle River, NJ, 2001. The circumference (in cms) and weight (in grams) is recorded for 22 pumpkins. Can one predict the weight from circumference?

Data Set Used in this Analysis :

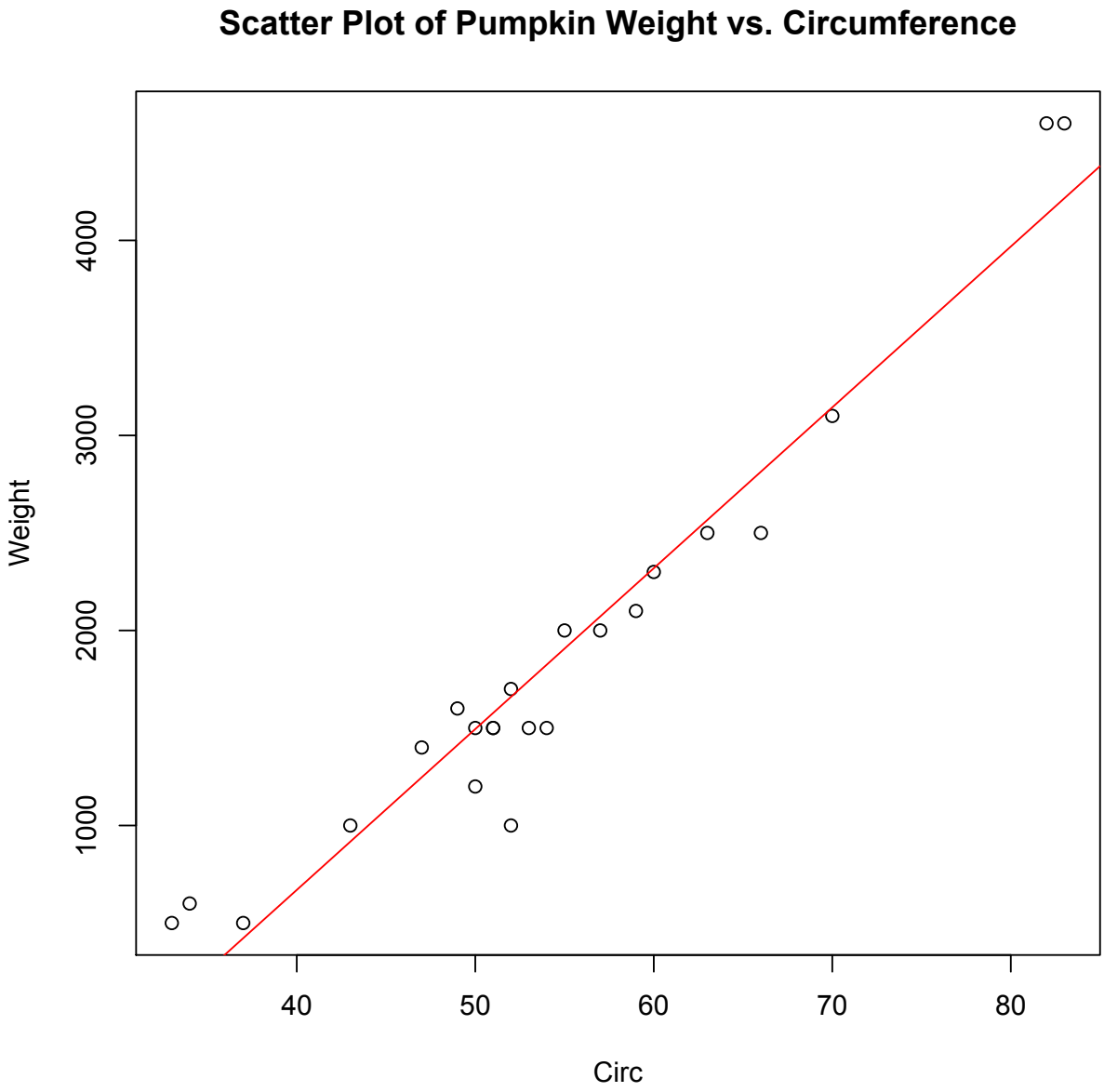
```
# Math 3080 - 1      Pumpkin Data          March 1, 2014
# Treibergs
#
# From Levine Ramsey & Smidt, Applied Statistics for Engneers and
# Scientists, Prentice Hall, Upper saddle River, NJ, 2001.
# The circumference (in cms) and weight (in grams) is recorded
# for 22 pumpkins. Can one predict the weight from circumference?
#
"Circ" "Weight"
50 1200
55 2000
54 1500
52 1700
37  500
52 1000
53 1500
47 1400
51 1500
63 2500
33  500
43 1000
57 2000
66 2500
82 4600
83 4600
70 3100
34  600
51 1500
50 1500
49 1600
60 2300
59 2100
```

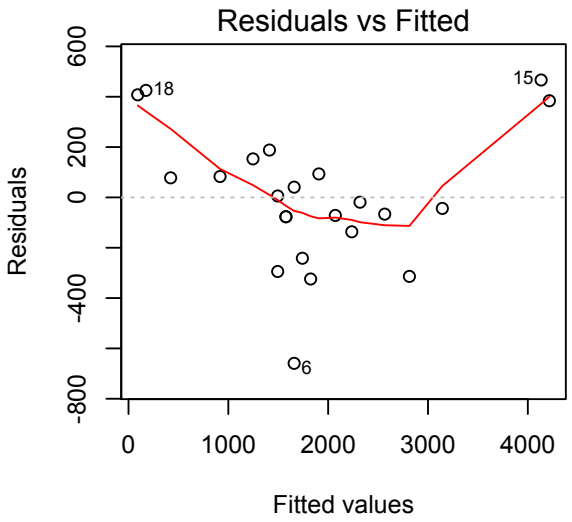
R Session:

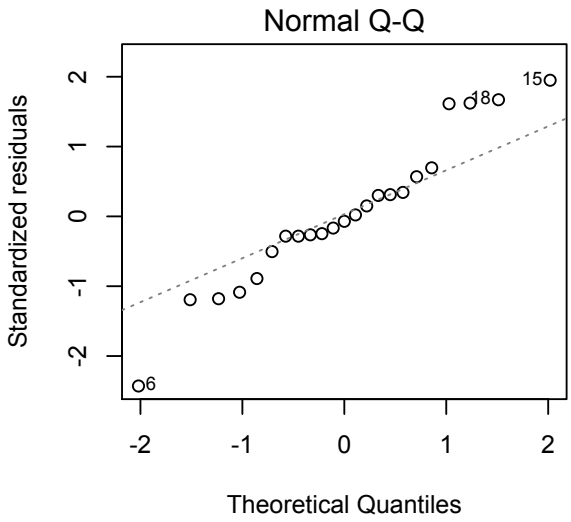
```
R version 2.13.1 (2011-07-08)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-apple-darwin9.8.0/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.41 (5874) i386-apple-darwin9.8.0]

[History restored from /Users/andrejstreibergs/.Rapp.history]

> tt=read.table("M3082DataPumpkin.txt",header=T)
> attach(tt)
> names(tt)
[1] "Circ"   "Weight"
> ############## SCATTERPLOT OF WEIGHT VS CIRC WITH REGRESSION LINE ######
> f1=lm(Weight~Circ)
> abline(f1,col=2)
> plot(Weight~Circ, main="Scatter Plot of Pumpkin Weight vs. Circumference")
> abline(f1,col=2)
> ############### DIAGNOSTIC PLOTS OF LINEAR MODEL ######################
> layout(matrix(c(1,3,2,4),nrow=2))
> plot(f1)
> ######################  SUMMARY AND ANOVA TABLE FOR LINEAR MODEL  #####
> summary(f1); anova(f1)

Call:
lm(formula = Weight ~ Circ)

Residuals:
    Min      1Q  Median      3Q     Max
-659.31 -106.72  -19.08  123.16  466.54

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2629.222    259.823  -10.12 1.57e-09 ***
Circ           82.472      4.657   17.71 4.20e-14 ***
```

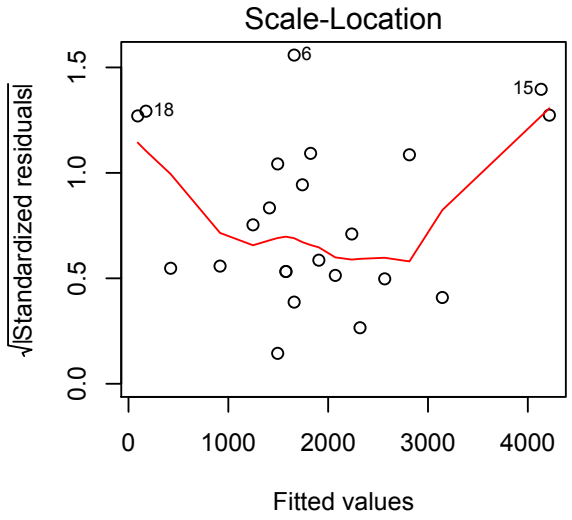
```
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 277.7 on 21 degrees of freedom
Multiple R-squared: 0.9372,Adjusted R-squared: 0.9343
F-statistic: 313.7 on 1 and 21 DF,  p-value: 4.203e-14

Analysis of Variance Table

Response: Weight
          Df   Sum Sq  Mean Sq F value    Pr(>F)
Circ       1 24196482 24196482  313.65 4.203e-14 ***
Residuals 21  1620040    77145
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1


> layout(1)
> ##############  REGRESSION LOG(WEIGHT) VS LOG(CIRC)  #################
> f2=lm(log(Weight)~log(Circ))
> summary(f2); anova(f2)

Call:
lm(formula = log(Weight) ~ log(Circ))

Residuals:
     Min       1Q   Median       3Q      Max
-0.41569 -0.04353  0.03247  0.07678  0.19928

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.3159     0.5149  -4.498 0.000198 ***
log(Circ)     2.4396     0.1295  18.842 1.23e-14 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.1422 on 21 degrees of freedom
Multiple R-squared: 0.9442,Adjusted R-squared: 0.9415
F-statistic:   355 on 1 and 21 DF,  p-value: 1.231e-14

Analysis of Variance Table

Response: log(Weight)
          Df Sum Sq Mean Sq F value    Pr(>F)
log(Circ)  1 7.1801  7.1801  355.04 1.231e-14 ***
Residuals 21 0.4247  0.0202
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> ############ SCATTERPLOT WITH FITTED POWER CURVE  ##########################
> plot(Weight~Circ,main="Scatter Plot of Pumpkin Weight vs. Circumference")
> curve(exp(f2$coefficients[1]+f2$coefficients[2]*log(x)),30,86,col=4, add=T)
```

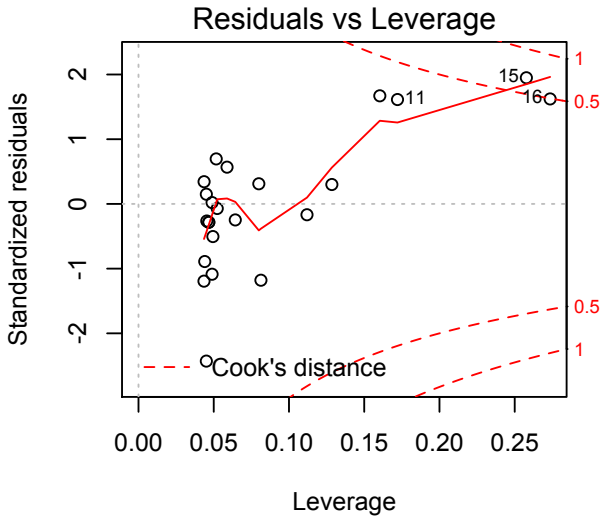
```
>
> ##############  DIAGNOSTIC PLOT OF LOG-LOG REGRESSION  ######################
> layout(matrix(c(1,3,2,4),nrow=2))
> plot(f2)
>
> #################  REMOVE OBSERVATIONS 5 AND 6 FROM DATA  ##################
> W=Weight[-6:-5]; C=Circ[-6:-5]
> f3=lm(log(W)~log(C))
> summary(f3); anova(f3)

Call:
lm(formula = log(W) ~ log(C))

Residuals:
      Min        1Q    Median        3Q       Max
-0.179337 -0.014944 -0.002274  0.043806  0.155355

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.83368    0.32880  -5.577 2.23e-05 ***
log(C)       2.32695    0.08231  28.271  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.08515 on 19 degrees of freedom
Multiple R-squared: 0.9768,Adjusted R-squared: 0.9756
F-statistic: 799.2 on 1 and 19 DF,  p-value: < 2.2e-16

Analysis of Variance Table

Response: log(W)
          Df Sum Sq Mean Sq F value    Pr(>F)
log(C)     1 5.7946  5.7946  799.23 < 2.2e-16 ***
Residuals 19 0.1378  0.0073
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
>
> ###############  SCATTERPLOT OF DATA WITHOUT OBS 5 AND 6  ###############
> layout(1)
> plot(W~C,main="Scatter Plot of Pumpkin Wt vs. Circ Omitting Obs 5,6")
> curve(exp(f3$coefficients[1]+f3$coefficients[2]*log(x)),30,86,col=4, add=T)
>
> ###############  DIAGNOSTICS FOR REGRESSION WITHOUT POINTS 5 AND 6  ####
> layout(matrix(c(1,3,2,4),nrow=2))
> plot(f3)
>
```

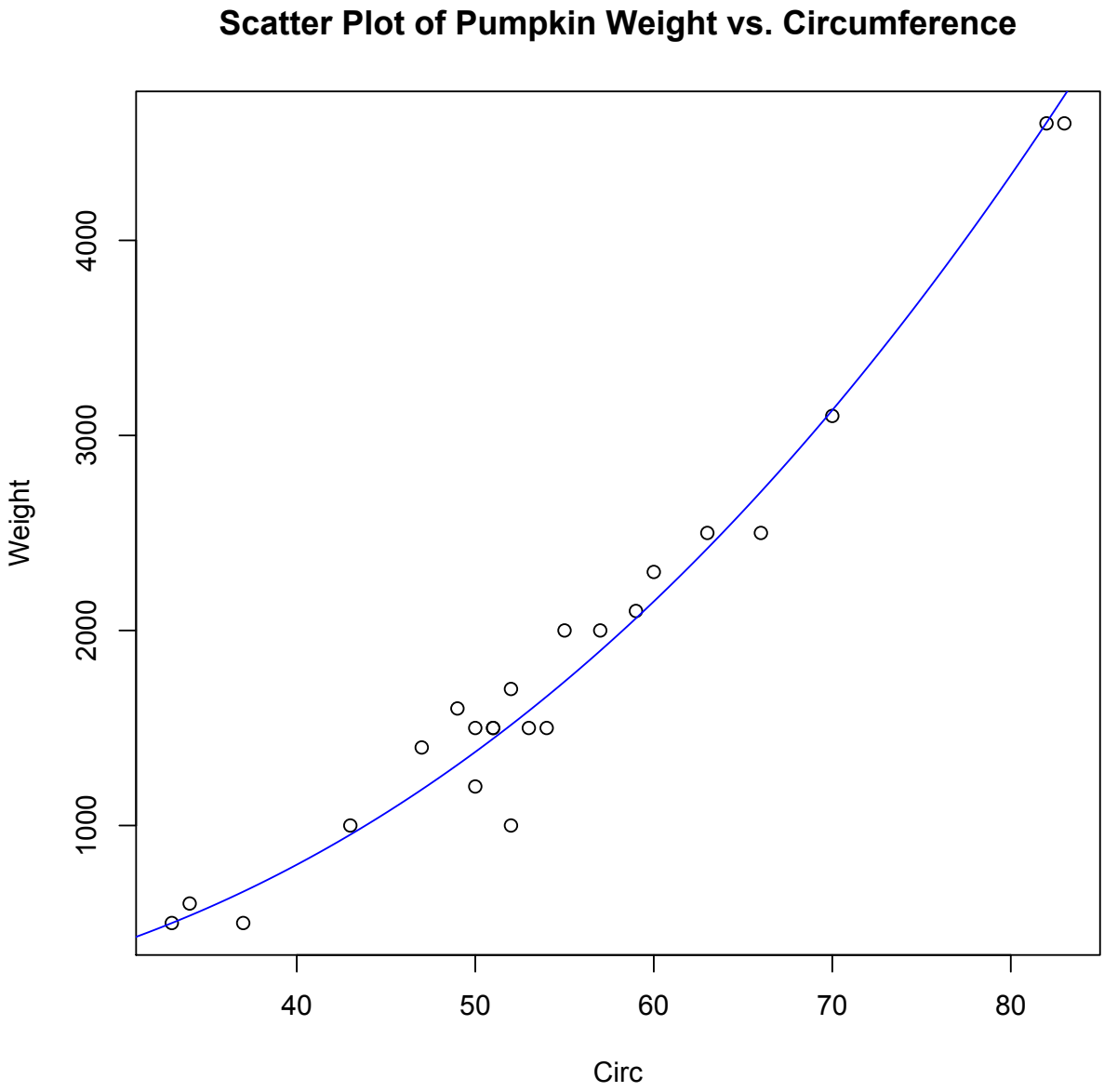

Scatter Plot of Pumpkin Weight vs. Circumference
Weight
1000
2000
3000
4000
40
50
60
70
80
Circ

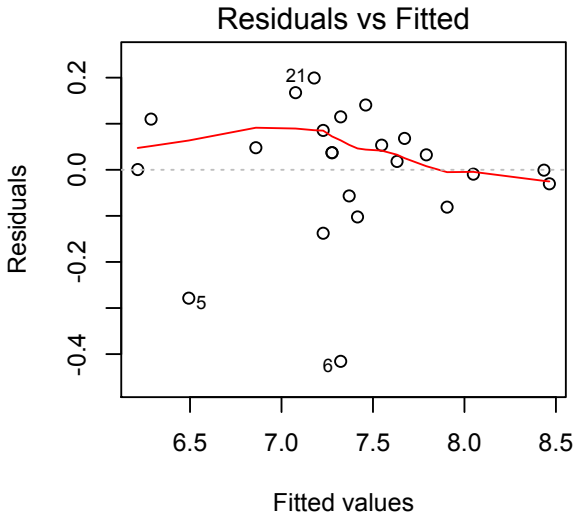

Residuals vs Fitted
Residuals
Fitted values
18
15
6
600
200
0
-400
-800
0
1000
2000
3000
4000

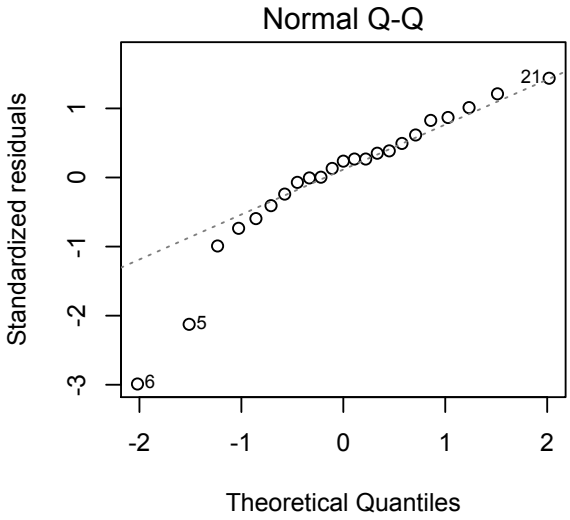

Normal Q-Q
Standardized residuals
Theoretical Quantiles
15
18
6
2
1
0
-1
-2

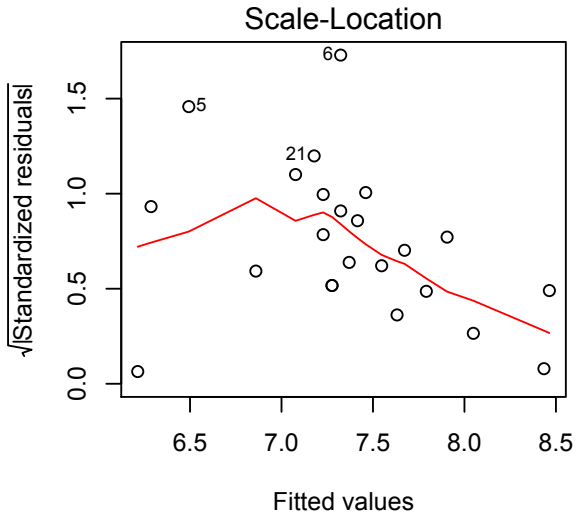

Scale-Location
√|Standardized residuals|
Fitted values
6
18
15
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000

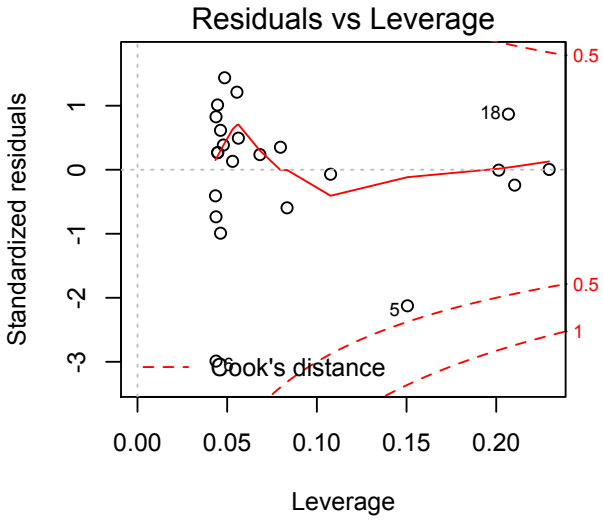

Residuals vs Leverage
Standardized residuals
Leverage
15
11
16
Cook's distance
1
0.5
0.5
1
2
1
0
-1
-2
0.00
0.05
0.10
0.15
0.20
0.25

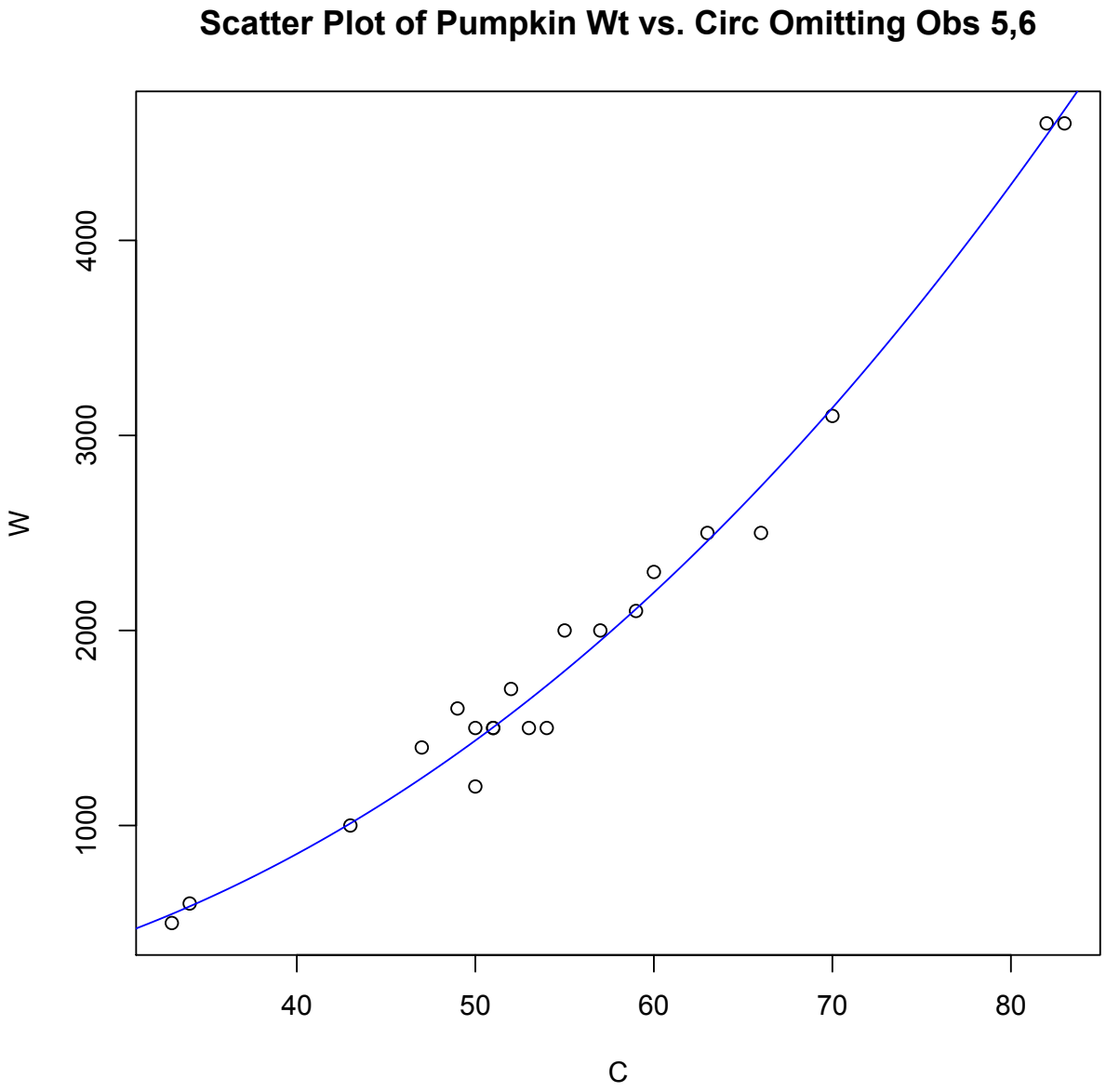

Scatter Plot of Pumpkin Weight vs. Circumference
Weight
1000
2000
3000
4000
40
50
60
70
80
Circ

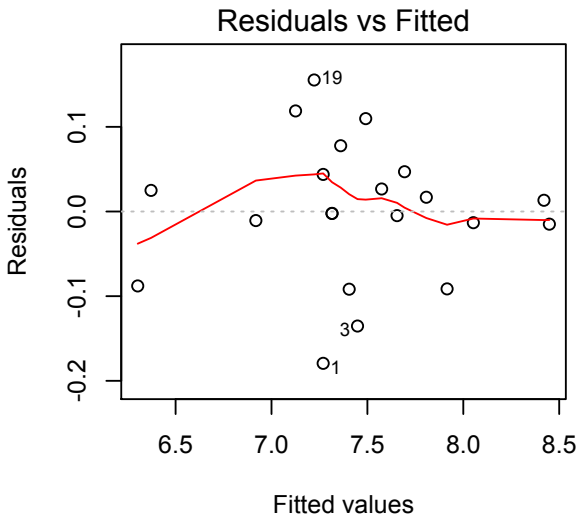

Residuals vs Fitted
Residuals
Fitted values
21
5
6
0.2
0.0
-0.2
-0.4
6.5
7.0
7.5
8.0
8.5

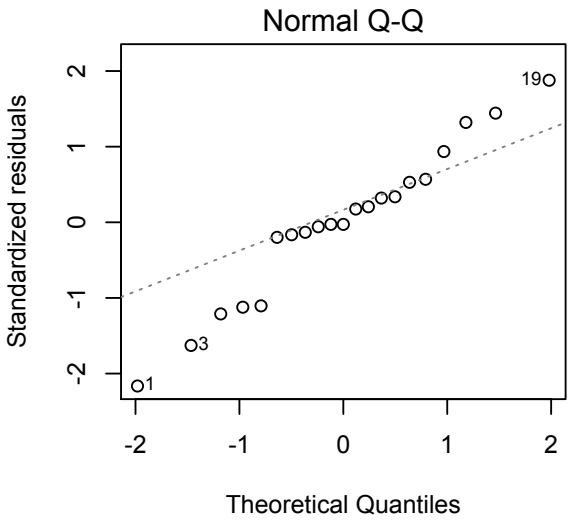

Normal Q-Q
Standardized residuals
Theoretical Quantiles
21
5
6
1
0
-1
-2
-3
-2
-1
0
1
2

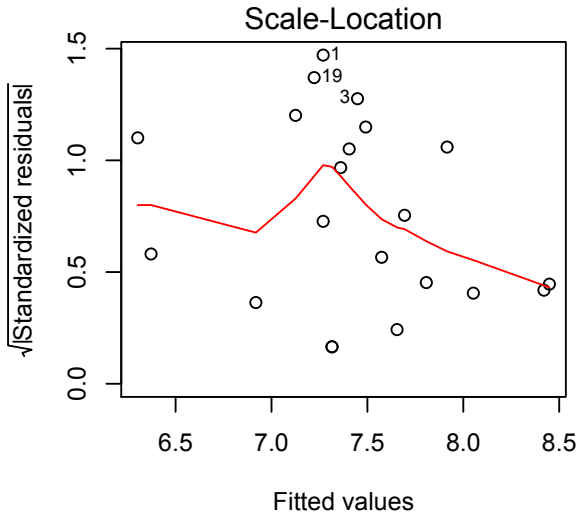

Scale-Location
√|Standardized residuals|
Fitted values
6
5
21
1.5
1.0
0.5
0.0
6.5
7.0
7.5
8.0
8.5

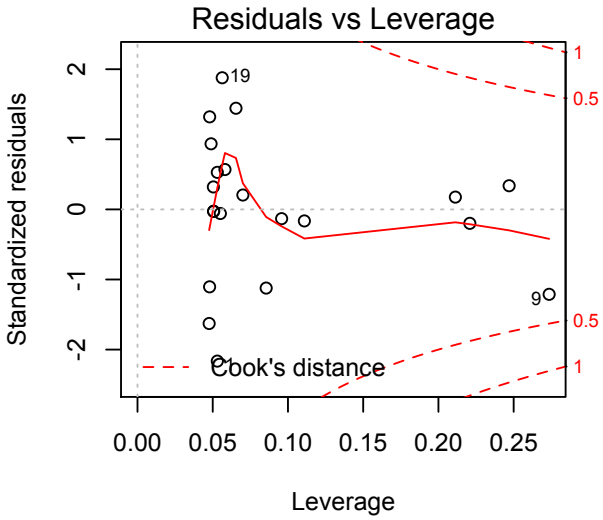

Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
18
5
6
0.5
0.5
1
1
0
-1
-2
-3
0.00
0.05
0.10
0.15
0.20


Scatter Plot of Pumpkin Wt vs. Circ Omitting Obs 5,6
W
1000
2000
3000
4000
40
50
60
70
80
C


Residuals vs Fitted
Residuals
Fitted values
19
3
1
-0.2
-0.1
0.0
0.1
6.5
7.0
7.5
8.0
8.5


Normal Q-Q
Standardized residuals
Theoretical Quantiles
19
3
1
-2
-1
0
1
2


Scale-Location
√|Standardized residuals|
Fitted values
1
19
3
0.0
0.5
1.0
1.5
6.5
7.0
7.5
8.0
8.5


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
19
9
1
0.5
0.5
1
-2
-1
0
1
2
0.00
0.05
0.10
0.15
0.20
0.25