Math 3080 § 1. **Polycarbonate Example:** Name: Example
Treibergs **Confidence Intervals in Polynomial Regression** March 9, 2014

This **R**© program explores polynomial regression. We investigate confidence intervals (and thus hypothesis tests) for the regression coefficients $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_3}$, for the mean response Y^* and prediction for next observation $\hat{Y}$.

This data was taken from the article "Measurements of the Thermal Conductivity and Thermal Diffusivity of Polymer Melts with the Short Hot-Wire Method," Zhang & Hendro *et al, International Journal of Thermophysics* (2002) as quoted by Navidi, *Statistics for Engineers and Scientists,* 2nd ed., McGraw Hill, 2008. Measurements of the thermal conductivity (in $Wm^{-1}K^{-1}$) and diffusivity of polycarbonate at several temperatures (in 1000° C) were reported. We seek a polynomial regression to predict the conductivity.

Data Set Used in this Analysis :

```
# Math 3080 - 1      Polycarbonate Data          March 9, 2014
# Treibergs
#
# In the article "Measurements of the Thermal Conductivity and Thermal
# Diffusivity of Polymer Melts with the Short Hot-Wire Method," Zhang &
# Hendro et al, International Journal of Thermophysics (2002) as
# quoted by Navidi, Statistics for Engineers and Scientists, 2nd ed.,
# McGraw Hill, 2008, measurements of the thermal conductivity (in W/mK) and
# diffusivity of polycarbonate at several temperatures (in 1000 degrees C)
# was reported. Use a polynomial regression to predict the conductivity.
#
"Cond" "Temp"
.236 .028
.241 .038
.244 .061
.251 .083
.259 .107
.257 .119
.257 .130
.261 .146
.254 .159
.256 .169
.251 .181
.249 .204
.249 .215
.230 .225
.230 .237
.228 .248
```

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("M3082DataPolycarbonate.txt",header=T)
> attach(tt)
> tt
    Cond  Temp
1  0.236 0.028
2  0.241 0.038
3  0.244 0.061
4  0.251 0.083
5  0.259 0.107
6  0.257 0.119
7  0.257 0.130
8  0.261 0.146
9  0.254 0.159
10 0.256 0.169
11 0.251 0.181
12 0.249 0.204
13 0.249 0.215
14 0.230 0.225
15 0.230 0.237
16 0.228 0.248
```

```
> ########### PLOT OBSERVED POINTS AND FITTED POLYNOMIALS  ########
> plot(Cond~Temp, xlab = "Thermal Conductivity (W/mK)",
  ylab = "Temperature (1000 deg. C)", main="Polycarbonate Data")
> xsquare=Temp*Temp
> xcube=Temp*xsquare
> xfourth=xsquare*xsquare
> ##### POLYNOMIAL REGRESSIONS OF FIRST TO FOURTH DEGREE  #########
> f1=lm(Cond~Temp)
> f2=lm(Cond~Temp+xsquare)
> f3=lm(Cond~Temp+xsquare+xcube)
> f4=lm(Cond~Temp+xsquare+xcube+xfourth)
> d=seq(0,.26,length.out=200)
> plot(Cond~Temp, ylab = "Thermal Conductivity (W/mK)",
  xlab = "Temperature (1000 deg. C)", main="Polycarbonate Data")
> lines(d, predict(f4, data.frame(Temp=d, xsquare=d^2, xcube=d^3,
  xfourth=d^4)), col = 2)
> lines(d, predict(f3, data.frame(Temp=d, xsquare=d^2, xcube=d^3)),
  col = 3)
> lines(d, predict(f2, data.frame(Temp=d, xsquare=d^2)), col = 4)
> lines(d, predict(f1, data.frame(Temp=d)), col = 5)
> legend(.1,.240, c("Linear","Quadratic","Cubic","Quartic"),
  fill=c(5,4,3,2), title="Polynomial Regression")
>
> ######  LOOK AT FOURTH DEGREE FIT  ##########################
> summary(f4);anova(f4)

Call:
lm(formula = Cond ~ Temp + xsquare + xcube + xfourth)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0074537 -0.0016785 -0.0005074  0.0021481  0.0071585

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.2315     0.0135  17.152 2.76e-09 ***
Temp          0.1091     0.5834   0.187    0.855
xsquare       3.4544     7.7602   0.445    0.665
xcube       -26.0224    40.4496  -0.643    0.533
xfourth      40.1571    72.2925   0.555    0.590
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.003829 on 11 degrees of freedom
Multiple R-squared: 0.9111,Adjusted R-squared: 0.8788
F-statistic:  28.2 on 4 and 11 DF,  p-value: 9.924e-06
```

```
Analysis of Variance Table
Response: Cond
          Df     Sum Sq    Mean Sq  F value    Pr(>F)
Temp       1 0.00012825 0.00012825   8.7482   0.01303 *
xsquare    1 0.00150950 0.00150950 102.9654 6.387e-07 ***
xcube      1 0.00001140 0.00001140   0.7777   0.39671
xfourth    1 0.00000452 0.00000452   0.3086   0.58968
Residuals 11 0.00016126 0.00001466
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
>
> #### We may reasonably neglect fourth power terms
>
> ############  LOOK AT THIRD DEGREE FIT #####################
> summary(f3);anova(f3)
Call:
lm(formula = Cond ~ Temp + xsquare + xcube)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0079975 -0.0018768 -0.0001387  0.0024047  0.0064944

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.225140   0.006896  32.648 4.31e-13 ***
Temp         0.411046   0.205762   1.998   0.0689 .
xsquare     -0.746513   1.688725  -0.442   0.6663
xcube       -3.672813   4.043033  -0.908   0.3815
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.003717 on 12 degrees of freedom
Multiple R-squared: 0.9087,Adjusted R-squared: 0.8858
F-statistic: 39.79 on 3 and 12 DF,  p-value: 1.636e-06

Analysis of Variance Table

Response: Cond
          Df     Sum Sq    Mean Sq  F value    Pr(>F)
Temp       1 0.00012825 0.00012825   9.2831   0.01014 *
xsquare    1 0.00150950 0.00150950 109.2610 2.215e-07 ***
xcube      1 0.00001140 0.00001140   0.8252   0.38153
Residuals 12 0.00016579 0.00001382
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
>
> ### Note that the Sum Sq are same for first three. This means that
> ### the fourth power ANOVA fits each model in turn and the Sum Sq
> ### indicates the improvement over the over fitting the previous
> ### variables. We may reasonably neglect third power terms.
```

```
> ##############  LOOK AT QUADRATIC FIT  ####################
> summary(f2);anova(f2)

Call:
lm(formula = Cond ~ Temp + xsquare)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0077387 -0.0022391  0.0001283  0.0019330  0.0071757

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.219955   0.003843  57.230  < 2e-16 ***
Temp         0.589315   0.061460   9.589 2.92e-07 ***
xsquare     -2.267888   0.215501 -10.524 9.92e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.003692 on 13 degrees of freedom
Multiple R-squared: 0.9024,Adjusted R-squared: 0.8874
F-statistic: 60.08 on 2 and 13 DF,  p-value: 2.705e-07

Analysis of Variance Table

Response: Cond
          Df     Sum Sq    Mean Sq  F value    Pr(>F)
Temp       1 0.00012825 0.00012825   9.4096  0.008992 **
xsquare    1 0.00150950 0.00150950 110.7498 9.919e-08 ***
Residuals 13 0.00017719 0.00001363
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
> ### All coefficients are significant. Tis model has the lowest
> ### adjusted R2.

> ############  LINEAR MODEL  ##################################
> summary(f1); anova(f1)
Call:
lm(formula = Cond ~ Temp)

Residuals:
      Min        1Q    Median        3Q       Max
-0.016003 -0.011270  0.004540  0.008893  0.013901
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.253167   0.006522  38.819 1.18e-15 ***
Temp        -0.041561   0.040281  -1.032     0.32
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.01098 on 14 degrees of freedom
Multiple R-squared: 0.07066,Adjusted R-squared: 0.004283
F-statistic: 1.065 on 1 and 14 DF,  p-value: 0.3197
```

```
Analysis of Variance Table

Response: Cond
          Df     Sum Sq    Mean Sq F value Pr(>F)
Temp       1 0.00012825 0.00012825  1.0645 0.3197
Residuals 14 0.00168669 0.00012048


> ### Note that the diagnostics indicate that there is nonlinear dependence
> ### and the residuals are not normal. In the linear model, the first
> ### order term may be reasonably neglected. However, adjusted R2 = .004
> ### Using adjusted R2, the best model is the quadratic Model.
>

> ###########  DIAGNOSTIC PLOTS FOR THE FOUR MODELS  ###########
> opar <- par(mfrow = c(2, 2), oma = c(0, 0, 1.1, 0),
+             mar = c(4.1, 4.1, 2.1, 1.1))
> plot(f4)
> plot(f3)
> plot(f2)
> plot(f1)

>
> #### FOR THE REST OF THIS DISCUSSION WE FOCUS ON THE QUADRATIC MODEL
> ####  REPEAT SUMMARY f2
> summary(f2)

Call:
lm(formula = Cond ~ Temp + xsquare)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0077387 -0.0022391  0.0001283  0.0019330  0.0071757

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.219955   0.003843  57.230  < 2e-16 ***
Temp         0.589315   0.061460   9.589 2.92e-07 ***
xsquare     -2.267888   0.215501 -10.524 9.92e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.003692 on 13 degrees of freedom
Multiple R-squared: 0.9024,Adjusted R-squared: 0.8874
F-statistic: 60.08 on 2 and 13 DF,  p-value: 2.705e-07

> SSE = sum((Cond-fitted(f2))^2)
> SSE
[1] 0.0001771875
> xbar = mean(Temp)
> xbar
```

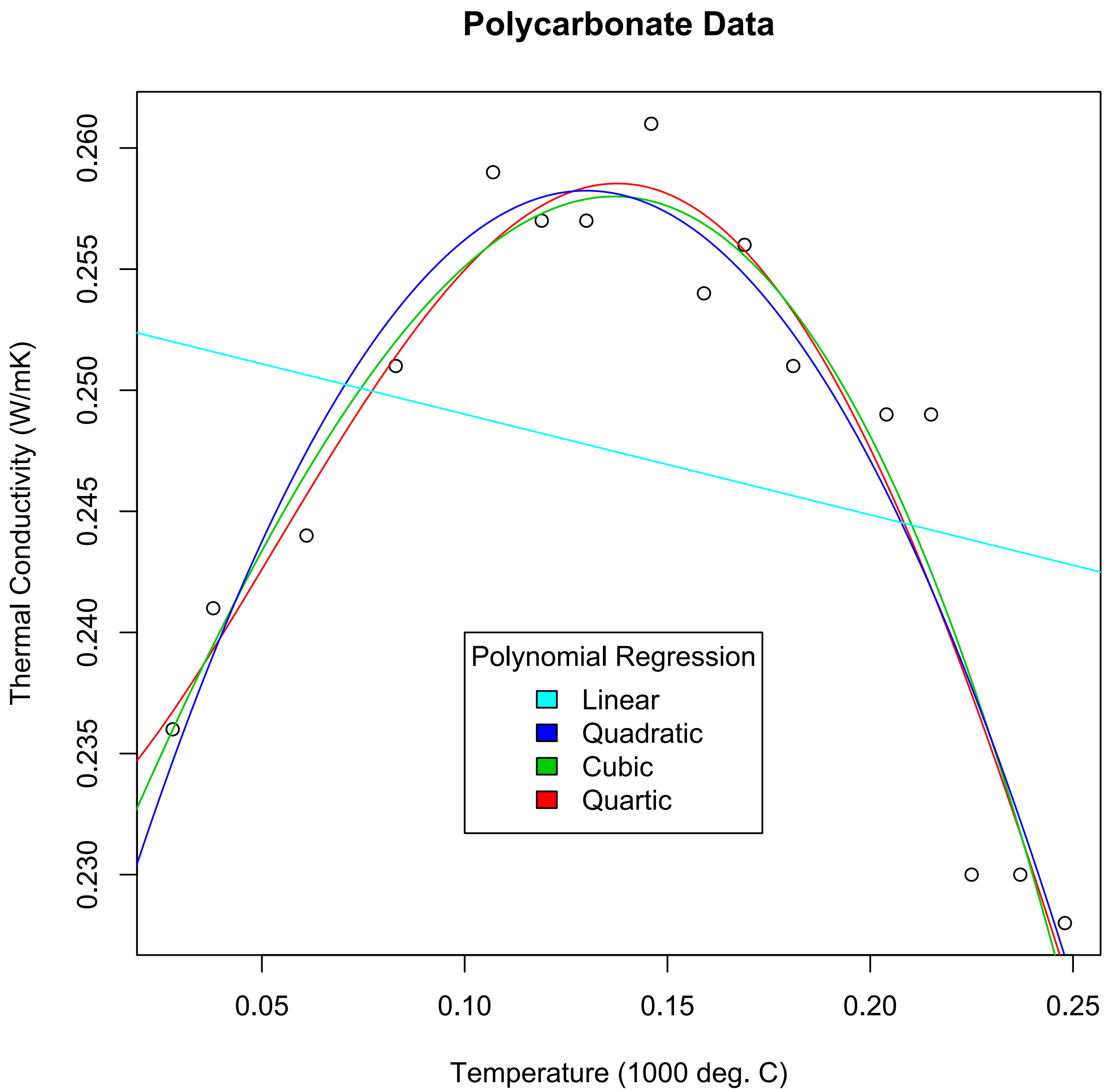
```
[1] 0.146875
> ybar=mean(Cond); ybar
[1] 0.2470625
> SST= sum((Cond-ybar)^2); SST
[1] 0.001814938
> R2 = 1-SSE/SST; R2
[1] 0.9023726
> n=length(Cond); n
[1] 16
> k=2
> AdjR2= 1-(n-1)*SSE/((n-1-k)*SST); AdjR2
[1] 0.8873531
> bhat=coefficients(f2); bhat
(Intercept)        Temp     xsquare
  0.2199548   0.5893146  -2.2678875

> ##############  5% 2-SIDED CI'S ON PARAMETERS  ###########
> confint(f2)
                 2.5 %     97.5 %
(Intercept)  0.2116517  0.2282579
Temp         0.4565391  0.7220901
xsquare     -2.7334500 -1.8023251
>
> #### Confidence intervals for the model parameters (betas)
> #### by hand.
> tcrit=qt(.025,n-1-k,lower.tail=F);tcrit
[1] 2.160369
> c(bhat[1]-tcrit*0.003843,bhat[1],bhat[1]+tcrit*0.003843)
(Intercept) (Intercept) (Intercept)
  0.2116525   0.2199548   0.2282571
> c(bhat[2]-tcrit*0.061460,bhat[2],bhat[2]+tcrit*0.061460)
     Temp      Temp      Temp
0.4565383 0.5893146 0.7220909
> c(bhat[3]-tcrit*0.215501,bhat[3],bhat[3]+tcrit*0.215501)
  xsquare   xsquare   xsquare
-2.733449 -2.267888 -1.802326
>
> ### t-Test H0: beta2=-2    vs.  Ha: beta2 < -2  ###############
> t=(bhat[3]-(-2))/0.215501;t
  xsquare
-1.243092
> qt(.05,n-1-k)
[1] -1.770933
> #### Cannot reject the null hypothesis. t not in lower tail.
> pvalue=pt(t,n-1-k);pvalue
  xsquare
0.1178936
```

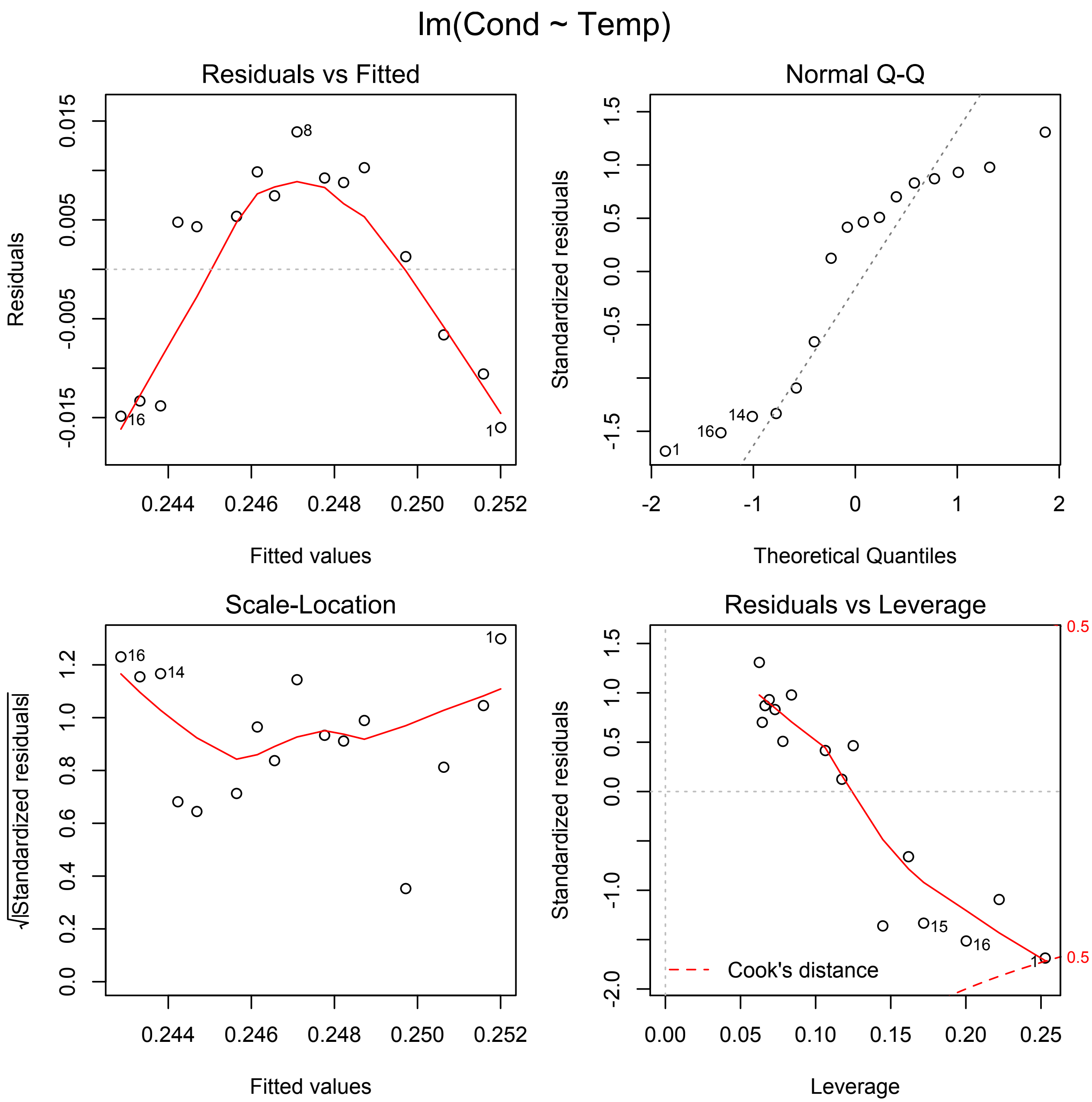
```
> ###  PRINT TABLE OF X's, FITTED, ST.DEV.FIT, RES., STANDARDIZED RES. ###

> m=matrix(c(Temp, Cond, fitted(f2), predict(f2,se=T)$se.fit, f2$residuals,
  rstandard(f2)), nrow=16)
> colnames(m)=c("Temp","Cond","Fitted","Stdev.fit","Residual","St.Resid")
> rownames(m)=1:16
> m
    Temp  Cond    Fitted   Stdev.fit      Residual   St.Resid
1  0.028 0.236 0.2346776 0.002481182  0.0013224440  0.4837421
2  0.038 0.241 0.2390739 0.002107538  0.0019261039  0.6354290
3  0.061 0.244 0.2474642 0.001515311 -0.0034641519 -1.0289923
4  0.083 0.251 0.2532444 0.001308876 -0.0022444053 -0.6501658
5  0.107 0.259 0.2570464 0.001330328  0.0019536115  0.5672770
6  0.119 0.257 0.2579677 0.001360978 -0.0009676526 -0.2819629
7  0.130 0.257 0.2582384 0.001376518 -0.0012383692 -0.3615001
8  0.146 0.261 0.2576524 0.001362964  0.0033475886  0.9756731
9  0.159 0.254 0.2563213 0.001318882 -0.0023213271 -0.6731918
10 0.169 0.256 0.2547758 0.001270686  0.0012241980  0.3531724
11 0.181 0.251 0.2523224 0.001211745 -0.0013224495 -0.3792153
12 0.204 0.249 0.2457945 0.001209107  0.0032054588  0.9189310
13 0.215 0.249 0.2418243 0.001325182  0.0071756919  2.0824313
14 0.225 0.230 0.2377387 0.001518454 -0.0077387490 -2.2996870
15 0.237 0.230 0.2322374 0.001857764 -0.0022373557 -0.7012819
16 0.248 0.228 0.2266206 0.002260568  0.0013793637  0.4725721
>
> #################  CI FOR MEAN RESPONSE  ####################
> ### for Temp=c(.10,.15,.20)

> TempStar=c(.1,.15,.2)
> predict(f2, data.frame(Temp=TempStar,xsquare=TempStar^2),
  se.fit=T, interval="confidence")
$fit
        fit       lwr       upr
1 0.2562074 0.2533717 0.2590430
2 0.2573245 0.2544030 0.2602460
3 0.2471022 0.2445318 0.2496725

$se.fit
          1           2           3
0.001312578 0.001352311 0.001189770

$df
[1] 13

$residual.scale
[1] 0.003691857
```

```
> ########  CI FOR PREDICTED NEXT OBSERVATION  #########
> predict(f2, data.frame(Temp=TempStar, xsquare=TempStar^2), se.fit=T,
  interval="prediction")
$fit
        fit       lwr       upr
1 0.2562074 0.2477425 0.2646722
2 0.2573245 0.2488305 0.2658185
3 0.2471022 0.2387225 0.2554819

$se.fit
          1           2           3
0.001312578 0.001352311 0.001189770

$df
[1] 13

$residual.scale
[1] 0.003691857


> #####  Centered Variables  ####################################
>
> xc=Temp-xbar
> xc2=xc*xc
> fc=lm(Cond~xc+xc2); summary(fc); anova(fc)

Call:
lm(formula = Cond ~ xc + xc2)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0077387 -0.0022391  0.0001283  0.0019330  0.0071757

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.257587   0.001361 189.280  < 2e-16 ***
xc          -0.076877   0.013958  -5.508 0.000101 ***
xc2         -2.267888   0.215501 -10.524 9.92e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.003692 on 13 degrees of freedom
Multiple R-squared: 0.9024,Adjusted R-squared: 0.8874
F-statistic: 60.08 on 2 and 13 DF,  p-value: 2.705e-07
```

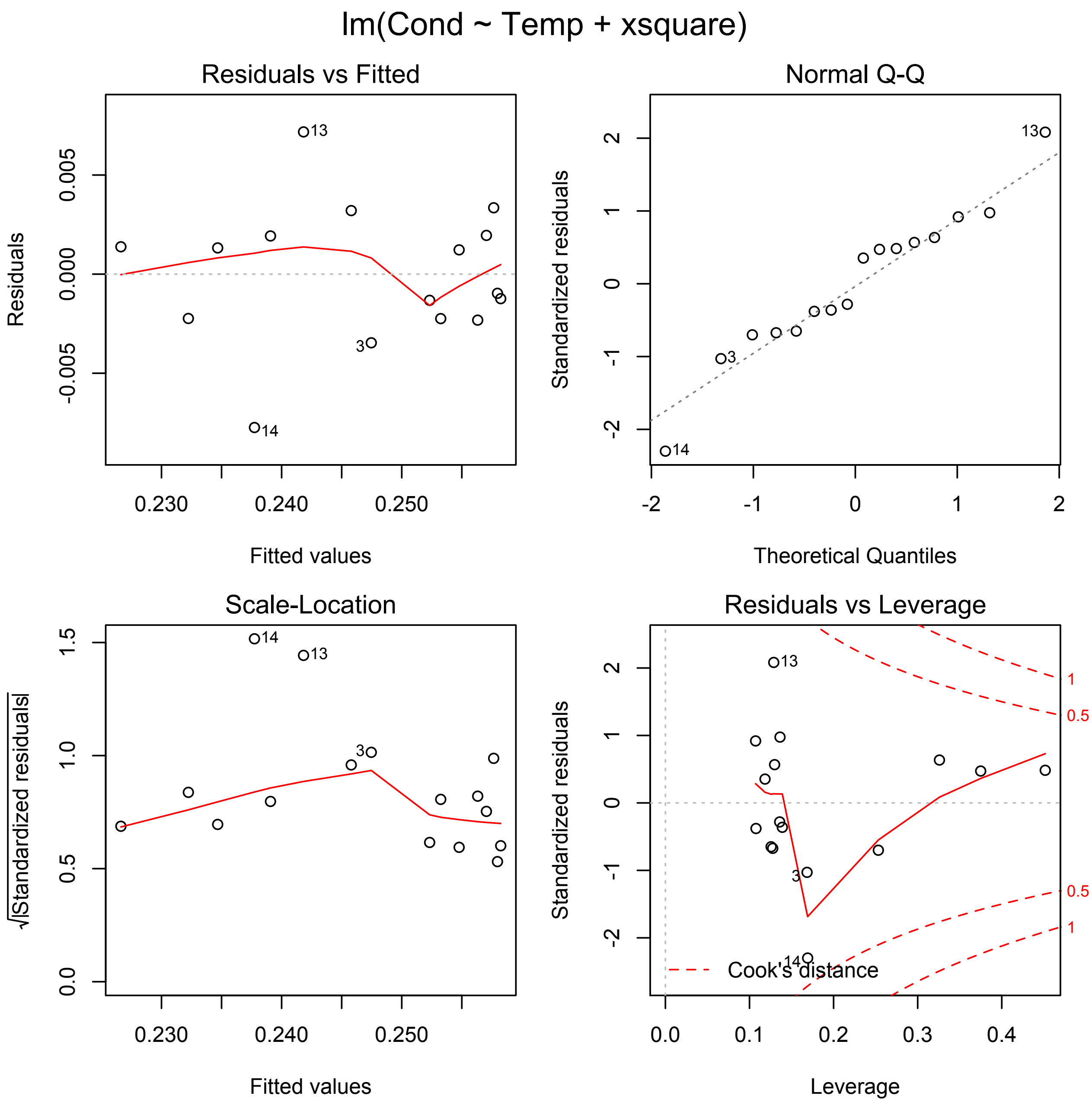
```
Analysis of Variance Table

Response: Cond
          Df     Sum Sq    Mean Sq  F value    Pr(>F)
xc         1 0.00012825 0.00012825   9.4096  0.008992 **
xc2        1 0.00150950 0.00150950 110.7498 9.919e-08 ***
Residuals 13 0.00017719 0.00001363
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> #### RELATE CENTERED COEFFICIENTS TO ORIGINAL COEFFICIENTS #####
> bc=coefficients(fc);bc
(Intercept)          xc         xc2
 0.25758688 -0.07687737 -2.26788752
> bhat
(Intercept)        Temp     xsquare
  0.2199548   0.5893146  -2.2678875
> c(bc[1]-bc[2]*xbar+bc[3]*xbar^2,bc[2]-2*bc[3]*xbar,bc[3])
(Intercept)          xc         xc2
  0.2199548   0.5893146  -2.2678875
>
> #### BEST METHOD: ORTHONORMAL POLYNOMIALS #####################
> #### The variable Temp is centered and divided by root mean square
> #### in column 1. The second column is a quadratic in Temp chosen
> #### so it is orthogonal to column 1 and also normalized.
> poly(Temp,2)
                1           2
 [1,] -0.43625763  0.44593222
 [2,] -0.39955877  0.32207979
 [3,] -0.31515141  0.08152361
 [4,] -0.23441393 -0.09078525
 [5,] -0.14633668 -0.21431577
 [6,] -0.10229806 -0.25086430
 [7,] -0.06192932 -0.26959898
 [8,] -0.00321115 -0.27163269
 [9,]  0.04449736 -0.25127880
[10,]  0.08119621 -0.22219639
[11,]  0.12523484 -0.17188727
[12,]  0.20964220 -0.02847192
[13,]  0.25001094  0.06194918
[14,]  0.28670980  0.15640832
[15,]  0.33074842  0.28516951
[16,]  0.37111716  0.41796874
attr(,"degree")
[1] 1 2
attr(,"coefs")
attr(,"coefs")$alpha
[1] 0.1468750 0.1313025

attr(,"coefs")$norm2
[1] 1.000000e+00 1.600000e+01 7.424975e-02 2.934876e-04
```

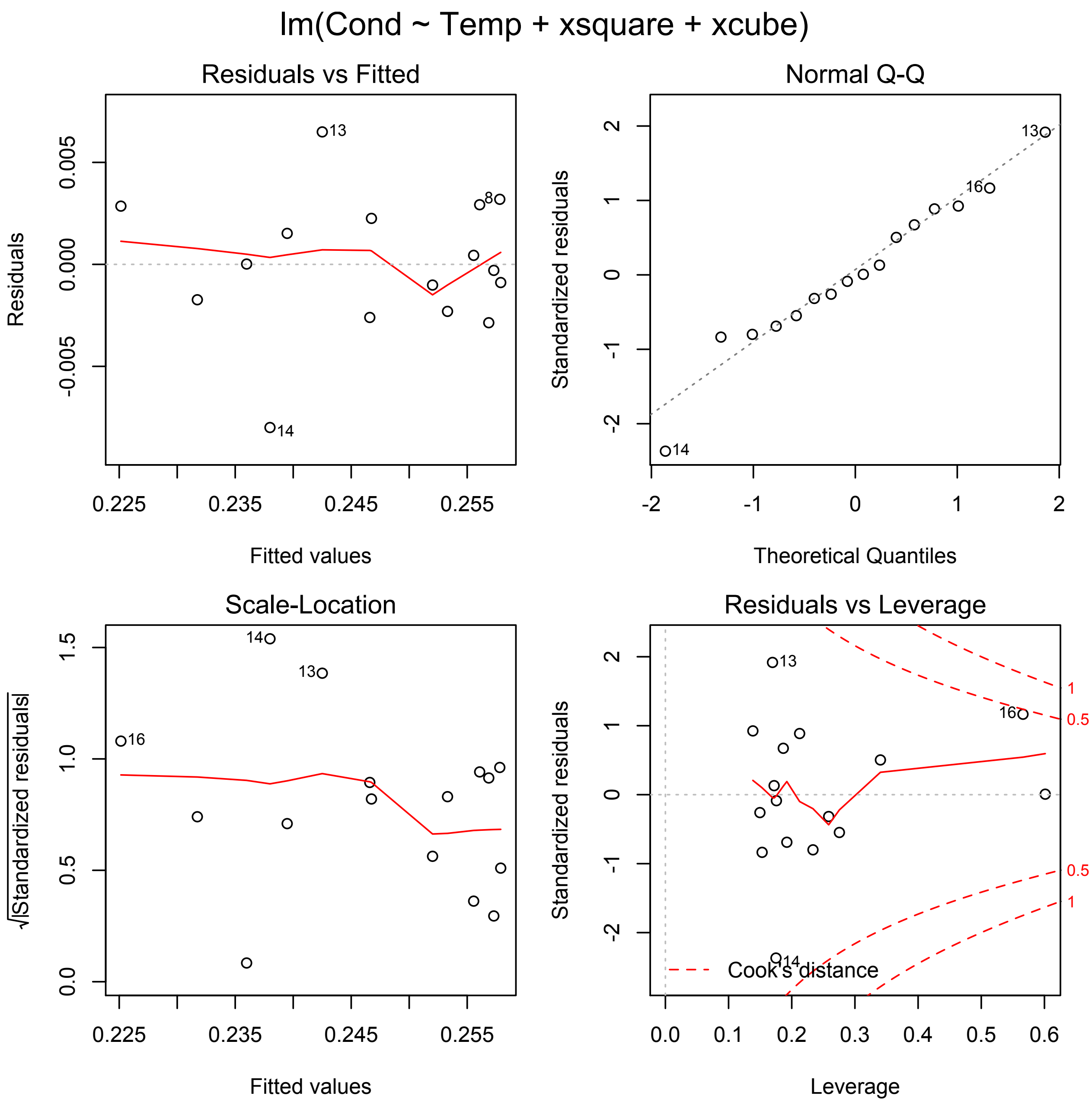
```
> q=poly(Temp,2)
> sum(q[,1]*q[,1]);sum(q[,1]*q[,2]);sum(q[,2]*q[,2])
[1] 1
[1] 7.155734e-17
[1] 1

> #####  POLY. REGRESSION WITH ORTHONORMAL QUADRATIC POLY ####

> fg=lm(Cond~q); summary(fg); anova(fg)

Call:
lm(formula = Cond ~ q)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0077387 -0.0022391  0.0001283  0.0019330  0.0071757

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.247062   0.000923 267.684  < 2e-16 ***
q1          -0.011325   0.003692  -3.068  0.00899 **
q2          -0.038852   0.003692 -10.524 9.92e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.003692 on 13 degrees of freedom
Multiple R-squared: 0.9024,Adjusted R-squared: 0.8874
F-statistic: 60.08 on 2 and 13 DF,  p-value: 2.705e-07

Analysis of Variance Table

Response: Cond
          Df     Sum Sq    Mean Sq F value    Pr(>F)
q          2 0.00163775 0.00081887   60.08 2.705e-07 ***
Residuals 13 0.00017719 0.00001363
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
> #### See canned R dataset    help(cars)
```


Polycarbonate Data
Thermal Conductivity (W/mK)
Temperature (1000 deg. C)
0.230
0.235
0.240
0.245
0.250
0.255
0.260
0.05
0.10
0.15
0.20
0.25
Polynomial Regression
Linear
Quadratic
Cubic
Quartic

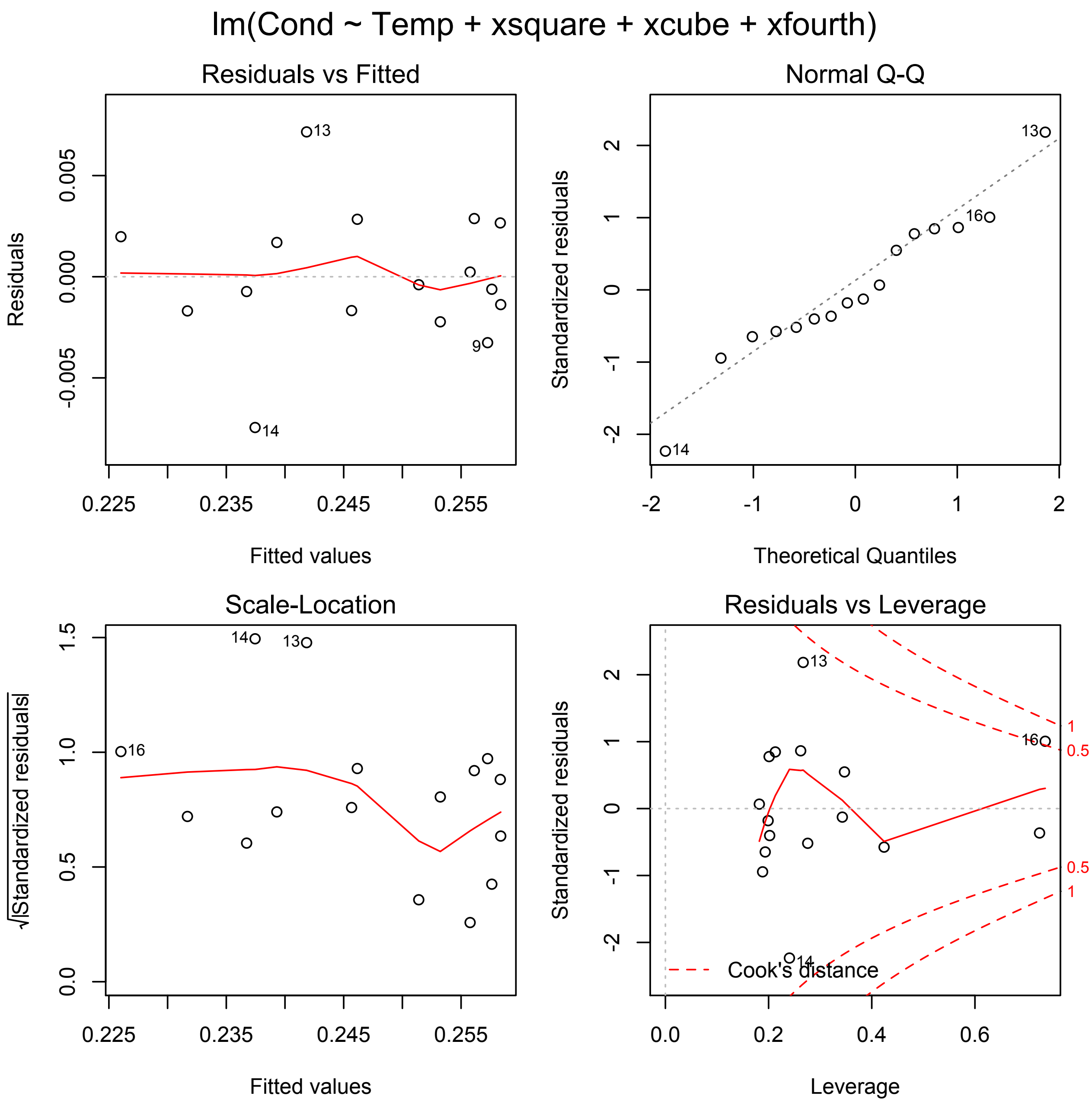


lm(Cond ~ Temp)
Residuals vs Fitted
Residuals
Fitted values
0.015
0.005
-0.005
-0.015
0.244
0.246
0.248
0.250
0.252
8
16
1
Normal Q-Q
Standardized residuals
Theoretical Quantiles
1.5
1.0
0.5
0.0
-0.5
-1.5
-2
-1
0
1
2
14
16
1
Scale-Location
√|Standardized residuals|
Fitted values
1.2
1.0
0.8
0.6
0.4
0.2
0.0
16
14
1
Residuals vs Leverage
Standardized residuals
Leverage
0.5
1.5
1.0
0.5
0.0
-1.0
-2.0
0.00
0.05
0.10
0.15
0.20
0.25
15
16
1
Cook's distance


lm(Cond ~ Temp + xsquare)
Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage



lm(Cond ~ Temp + xsquare + xcube)
Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage


lm(Cond ~ Temp + xsquare + xcube + xfourth)
Residuals vs Fitted
Residuals
Fitted values
0.225
0.235
0.245
0.255
-0.005
0.000
0.005
13
14
9
Normal Q-Q
Standardized residuals
Theoretical Quantiles
-2
-1
0
1
2
13
16
14
Scale-Location
√|Standardized residuals|
Fitted values
0.0
0.5
1.0
1.5
14
13
16
Residuals vs Leverage
Standardized residuals
Leverage
0.0
0.2
0.4
0.6
Cook's distance
13
16
14
1
0.5
0.5
1