7.20 a. The predicted response at the point is $\hat{y} = .2689$ and a 95% confidence interval on the mean response at the point is $(.2106, .3272)$.

b. The predicted response at the point is $\hat{y} = .2512$ and a 95% confidence interval on the mean response at the point is $(.2185, .2840)$.

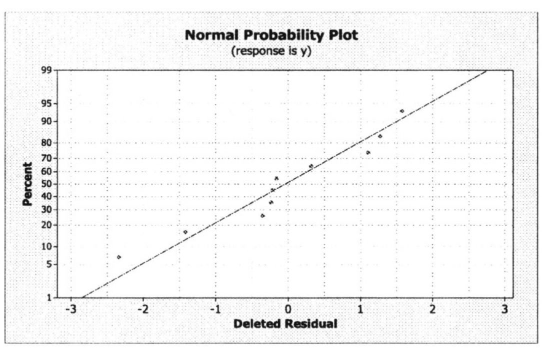
c. From the confidence intervals, it appears that the model without the pure quadratic terms might be better but the MS_{Res} are basically the same.

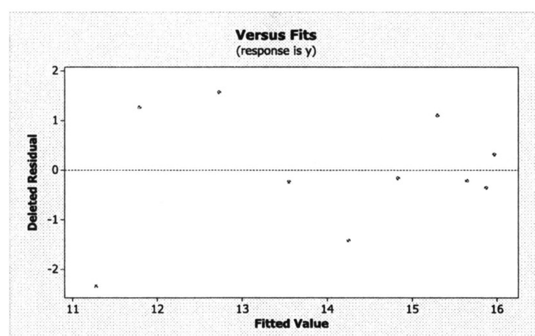
7.21 a. $\hat{y} = -1709 + 2.02x - .00059x^2$.

b. $F = 300.11$ with $p = 0.000$ which is significant.

c. $F = \dfrac{2.428}{.044} = 55.18$ which is significant. Both terms should be included in the model.

d. There is a problem with normality and a possibility of nonconstant variance.



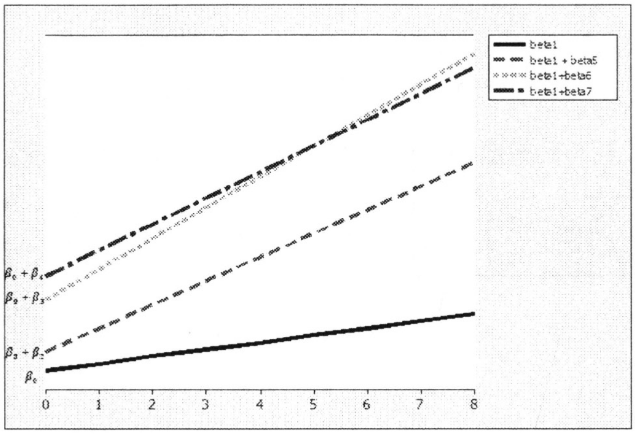


7.22 a. At $x = 1750$, the predicted response is $\widehat{y} = 14.8324$ and a 95% confidence interval on the mean response at the point is $(14.2841, 15.3808)$. At $x = 1775$, the predicted response is $\widehat{y} = 13.153$ and a 95% confidence interval on the mean response at the point is $(12.617, 13.6889)$.

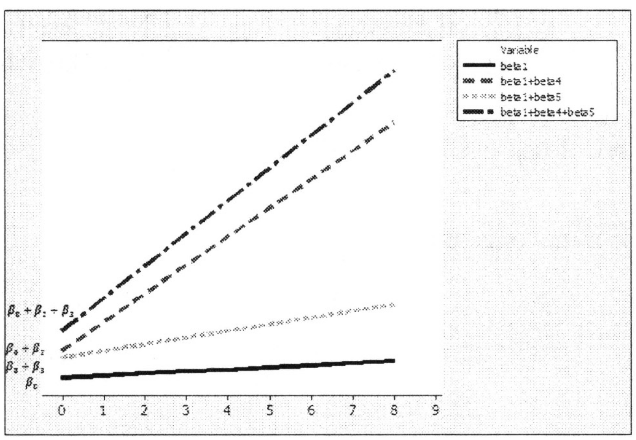
b. At $x = 1750$, the predicted response is $\widehat{y} = 14.303$ and a 95% confidence interval on the mean response at the point is $(12.888, 15.718)$. At $x = 1775$, the predicted response is $\widehat{y} = 12.996$ and a 95% confidence interval on the mean response at the point is $(11.548, 14.444)$. The predicted values are closer to the actual values using the quadratic model. Also, the prediction intervals are shorter with the quadratic model.

Chapter 8: Indicator Variables

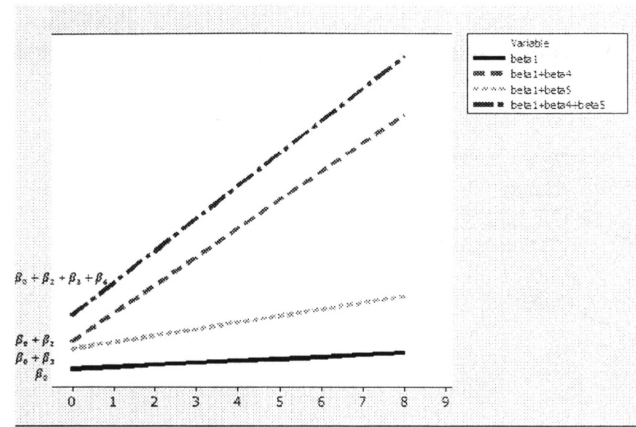
8.1 $\beta_0, \beta_2, \beta_3$, and β_4 determine the intercept while the other parameters determine the slope.



8.2 a.



b.



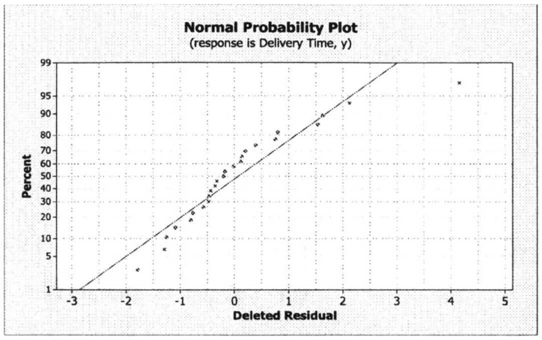
8.3 a. Let

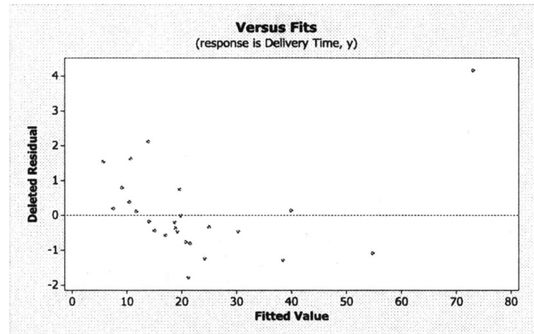
$$x_3 = \begin{cases} 1 & \text{if San Diego} \\ 0 & \text{otherwise} \end{cases} \qquad x_4 = \begin{cases} 1 & \text{if Boston} \\ 0 & \text{otherwise} \end{cases} \qquad x_5 = \begin{cases} 1 & \text{if Austin} \\ 0 & \text{otherwise} \end{cases}$$

Then $\widehat{y} = .42 + 1.77x_1 + .011x_2 + 2.29x_3 + 3.74x_4 - .45x_5$.

b. No, $F = \dfrac{64.2/3}{8.9} = 2.41$ which is not significant.

c. There is a problem with normality and a pattern to the residuals.





8.4 a. $\widehat{y} = 33.6 - .0457x_1 - .5x_{11}$. No, the $t = -.22$ with $p = 0.824$ which is not significant.

b. $\widehat{y} = 42.92 - .117x_1 - 13.46x_{11} + .082x_1x_{11}$. There is a significant interaction between engine displacement and the type of transmission. When the transmission is automatic, $\widehat{y} = (42.92 - 13.46) + (-.117 + .082)x_1 = 29.46 - .035x_1$ which indicates that on average for every increase of one cubic inch in displacement, miles per gallon decreases by .035. When the transmission is manual, $\widehat{y} = 42.92 - .117x_1$ which indicates that on average for every increase of one cubic inch in displacement, miles per gallon decreases by .117.

8.5 a. $\widehat{y} = 39.2 - .0048x_{10} - 2.7x_{11}$. No, the $t = -1.36$ with $p = 0.184$ which is not significant.

b. $\widehat{y} = 58.1 - .0125x_{10} - 26.2x_{11} + .009x_{10}x_{11}$. There is a significant interaction between vehicle weight and the type of transmission. When the transmission is automatic, $\widehat{y} = (58.1 - 26.2) + (-.0125 + .009)x_{10} = 31.9 - .0035x_{10}$ which indicates that on average for every increase of one cubic inch in displacement, miles per gallon decreases by .0035. When the transmission is manual, $\widehat{y} = 58.1 - .0125x_{10}$ which indicates that on average for every increase of one cubic inch in displacement, miles per gallon decreases by .0125.

8.6 Let

$$x_{51} = \begin{cases} 1 & \text{if } x_5 \text{ is negative} \\ 0 & \text{if } x_5 = 0 \end{cases} \qquad x_{52} = \begin{cases} 0 & \text{if } x_5 = 0 \\ 1 & \text{if } x_5 \text{ is positive} \end{cases}$$

This yields $\widehat{y} = 19.4 - .007x_7 - .006x_8 + .46x_{51} + 2.33x_{52}$. The effect of turnovers is assessed by $F = \dfrac{22.276/2}{5.462} = 2.04$ which is not significant.

8.7 $E(y) = S(x) = \beta_{00} + \beta_{01}x_1 + \beta_{11}(x_1 - t)x_2$ where $x_2 = \begin{cases} 0 & \text{if } x_2 \leq t \\ 1 & \text{if } x_2 > t \end{cases}$.

8.8 $E(y) = S(x) = \beta_{00} + \beta_{01}x_1 + \beta_{10}x_2 + \beta_{11}(x_1 - t)x_2$ where $x_2 = \begin{cases} 0 & \text{if } x_2 \leq t \\ 1 & \text{if } x_2 > t \end{cases}$.

8.9

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{41} \\ y_{42} \\ y_{43} \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

No, $\widehat{\beta}_0 = \bar{y}_{..} - \bar{y}_{1.} - \bar{y}_{2.} - \bar{y}_{3.} = \bar{y}_{4.}$, $\widehat{\beta}_1 = \bar{y}_{1.} - \bar{y}_{4.}$, $\widehat{\beta}_2 = \bar{y}_{2.} - \bar{y}_{4.}$, $\widehat{\beta}_3 = \bar{y}_{3.} - \bar{y}_{4.}$.

8.10 a. $y_{1j} = \beta_0 + \beta_1 + \varepsilon_{1j}$, $y_{2j} = \beta_0 + \beta_2 + \varepsilon_{2j}$, $y_{3j} = \beta_0 - \beta_1 - \beta_2 + \varepsilon_{3j}$ which gives

$$\begin{aligned} \mu_1 &= \beta_0 + \beta_1 \\ \mu_2 &= \beta_0 + \beta_2 \\ \mu_3 &= \beta_0 - \beta_1 - \beta_2. \end{aligned}$$

Therefore, $\mu_1 + \mu_2 + \mu_3 = 3\beta_0$ implying that $\beta_0 = \frac{\mu_1 + \mu_2 + \mu_3}{3} = \bar{\mu}$, $\beta_1 = \mu_1 - \beta_0 = \mu_1 - \bar{\mu}$, and $\beta_2 = \mu_2 - \beta_0 = \mu_2 - \bar{\mu}$.

b.

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \\ y_{31} \\ y_{32} \\ \vdots \\ y_{3n} \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ \vdots & \vdots & \vdots \\ 1 & -1 & -1 \end{pmatrix}$$

c.

$$\begin{aligned} SS_R(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \\ &= \begin{pmatrix} \bar{y}_{..} & \bar{y}_{1.} - \bar{y}_{..} & \bar{y}_{2.} - \bar{y}_{..} \end{pmatrix} \begin{pmatrix} y_{..} \\ y_{1.} - y_{3.} \\ y_{2.} - y_{3.} \end{pmatrix} \\ &= y_{..}\bar{y}_{..} + (y_{1.} - y_{3.})(\bar{y}_{1.} - \bar{y}_{..}) + (y_{2.} - y_{3.})(\bar{y}_{2.} - \bar{y}_{..}) \\ &= (y_{1.} + y_{2.} + y_{3.})\bar{y}_{..} + y_{1.}(\bar{y}_{1.} - \bar{y}_{..}) + y_{2.}(\bar{y}_{2.} - \bar{y}_{..}) - y_{3.}(\bar{y}_{1.} + \bar{y}_{2.} - 2\bar{y}_{..}) \\ &= y_{1.}\bar{y}_{1.} + y_{2.}\bar{y}_{2.} + y_{3.}(3\bar{y}_{..} - \bar{y}_{1.} - \bar{y}_{2.}) \\ &= y_{1.}\bar{y}_{1.} + y_{2.}\bar{y}_{2.} + y_{3.}\bar{y}_{3.} \end{aligned}$$

which is the same as the usual sum of squares.

8.11 a.

$$\mathbf{y} = \begin{pmatrix} 7 \\ 7 \\ 15 \\ 11 \\ 9 \\ 12 \\ 17 \\ 12 \\ 18 \\ 18 \\ 14 \\ 18 \\ 18 \\ 19 \\ 19 \\ 19 \\ 25 \\ 22 \\ 29 \\ 23 \\ 7 \\ 10 \\ 11 \\ 15 \\ 11 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

b. $\widehat{\beta}_0 = 10.8$, $\widehat{\beta}_1 = -1$, $\widehat{\beta}_2 = 4.6$, $\widehat{\beta}_3 = 6.8$, $\widehat{\beta}_4 = 10.8$.

c. $\widehat{\beta}_1 - \widehat{\beta}_3 = -1 - 6.8 = -7.8$.

d. $F = 14.76$ with $p = 0.000$ which is significant and indicates that the mean tensile strength is not the same for all five cotton percentages.

8.12 a. Since $y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk}$ for $i = 1, 2$, $j = 1, 2$ and $k = 1, 2$, we get

$$\begin{aligned}
y_{11k} &= \mu + \tau_1 + \gamma_1 + (\tau\gamma)_{11} + \varepsilon_{11k} \\
y_{12k} &= \mu + \tau_1 + \gamma_2 + (\tau\gamma)_{12} + \varepsilon_{12k} \\
y_{21k} &= \mu + \tau_2 + \gamma_1 + (\tau\gamma)_{21} + \varepsilon_{21k} \\
y_{22k} &= \mu + \tau_2 + \gamma_2 + (\tau\gamma)_{22} + \varepsilon_{22k}
\end{aligned}$$

Let

$$x_1 = \begin{cases} -1 & \text{if level 1 of treat type 1} \\ 1 & \text{if level 2 of treat type 1} \end{cases} \qquad x_2 = \begin{cases} -1 & \text{if level 1 of treat type 2} \\ 1 & \text{if level 2 of treat type 2} \end{cases}$$

Then, $y_{ijk} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon_{ijk}$.

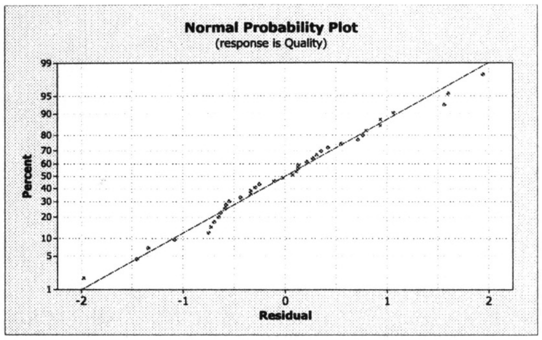
b.

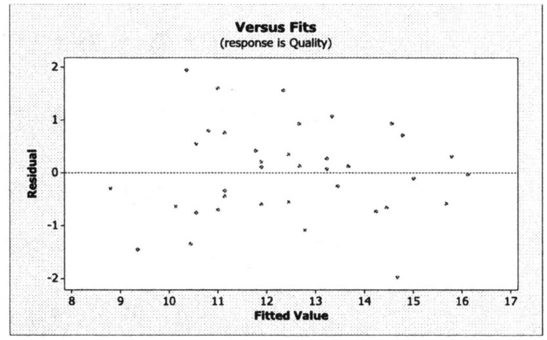
$$\mathbf{y} = \begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

c. To test $H_0 : \tau_1 = \tau_2 = 0$ obtain the sum of squares for the first treatment type and form the ratio $F = \frac{MS_A}{MS_{Res}}$. Do the same for the other treatment type and the interaction.

8.13 a. $\widehat{y} = 8.32 + 1.12x_4 - 1.22r_1 - 2.76r_2$. The region does have an impact, $F = \frac{30.961/2}{.8} = 19.35$.

b. There is a slight departure from normality.





c. There are 2 outliers: observations 12 and 25.

d. $\hat{y} = 10.1 + .796x_4 - 3.38r_1 - 6.28r_2 + .403x_4r_1 + .714x_4r_2$. No, the model is not superior to the model in part a.

8.14 The model in question 8.13 is superior.

Model	R^2	MS_{Res}	Region is Significant	Nonconstant Variance
Problem 8.13	80.9%	0.800	Yes	No
Problem 8.14	61.9%	1.584	No	Yes

8.15 Because LifeExp is the average between the male and female life expectancy, to predict average life, we can let

$$x_1 = \begin{cases} -1 & \text{if female} \\ 1 & \text{if male} \end{cases}$$

Also, recall from Problem 6.17 that a transformation was needed. If we again use the square roots of the response and the regressors, the model is

$\widehat{y^*} = 8.67 + 0.154x_1 - 0.0326sqrt(x_2) - 0.00704sqrt(x_3)$, with $F = 45.05$ and $p = 0.000$. $R^2 = 65.2\%$ and $R^2_{Adj} = 63.8\%$. $MS_{Res} = 0.0935$. Observations 8, 21, 30, 46, 59, and 68 are influential, as before, and considering this, there are no problems with the residual plots. This is very close to our model for average expected life from Problem 6.17: $\widehat{y^*} = 8.67 - 0.0323sqrt(x_2) - 0.00713sqrt(x_3)$ with $MS_{Res} = 0.0902$ but includes the adjustment for gender, so all three responses can be fit with a single model.

8.16 The response variable INHIBIT was transformed by taking the square root due to problems with nonconstant variance in the original model. Let

$$x_2 = \begin{cases} 0 & \text{if Surface} \\ 1 & \text{if Deep} \end{cases}$$

The model is $\widehat{y^*} = -0.264 + 121x_1 + 2.25x_2$. $F = 11.45$ with $p = 0.001$. $R^2 = 62.1\%$ and $R^2_{Adj} = 56.6\%$. No observations are influential, and the residual plots confirm the assumptions are not violated.

8.17 Adding the indicator variable has not improved the model. There is no evidence to support the claim that medical and surgical patients differ in their satisfaction as evident by the fact that the indicator variable is insignificant ($t = 0.48$ and $p-value = 0.633$).

The regression equation is $\hat{y} = 140 - 1.06x_{age} - 0.441x_{sev} + 1.99x_{sur-med}$.

Coefficient	test statistic	p-value
β_{age}	-6.51	0.000
β_{sev}	-2.42	0.025
$\beta_{sur*med}$	0.48	0.633

8.18 The addition of the indicator variable to the fuel consumption data does not seem to improve the analysis. In the analysis the only variable that significantly impacts fuel consumption is the initial boiling point x_5. The analysis below shows that adding the indicator variable is not a significant additional to the model. The proper analysis of these data is given in Exercise 5.20.

The regression equation is $\hat{y} = 413 - 4.25x_1 - 0.264x_5$.

Coefficient	test statistic	p-value
β_1	-1.09	0.295
β_5	-2.76	0.016

8.19 The model for the wine quality data was reduced to find significant predictors. The only significant predictor turned out to be wine color x_5. When the indicator for wine variety was added to the model, the variable was not significant at the 0.05 level with a $p - value = 0.17$. For this data we will also note that there was a strong problem with multicollinearity, so we are hesitant on the accuracy of this model.

The regression equation is $\hat{y} = 12 - 0.628x_1 + 0.850x_5$.

Coefficient	test statistic	p-value
β_1	-1.41	0.170
β_5	5.59	0.000

8.20 The regression for the methanol oxidation data was completed in Exercise 5.7. The indicator for reactor system was already included in the regression model. Exercise 5.7 concludes that the indicator variable is not significant for the transformed model.

Chapter 9: Multicollinearity

9.1 a. The correlation between x_1 and x_2 is .824.

b. The variance inflation factors are 3.1.

c. The condition number of $\mathbf{X}'\mathbf{X}$ is $\kappa = 40.68$ which indicates that multicollinearity is not a problem in these data.

9.3 The eigenvector associated with the smallest eigenvalue is

Eigenvector
−0.839
0.081
0.437
0.117
0.289

All four factors contribute to multicollinearity.

9.5 There are two large condition indices in the non-centered data. In general, it is better to center.

Number	Condition Indices	x_1	x_2	x_3	x_4
1	1.000	.00037	.00002	.00021	.00004
2	7.453	.01004	.00001	.00266	.0001
3	14.288	.00058	.00032	.00159	.00168
4	109.412	.05745	.00278	.04569	.00088
5	62,290.176	.93157	.99687	.94985	.9973

9.7 a. The correlation matrix is

$$
\begin{array}{c}
 \\ x_2 \\ x_3 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11}
\end{array}
\begin{array}{c}
\begin{array}{cccccccc}
x_1 & x_2 & x_3 & x_6 & x_7 & x_8 & x_9 & x_{10}
\end{array} \\
\begin{pmatrix}
0.945 & & & & & & & \\
0.989 & 0.964 & & & & & & \\
0.659 & 0.772 & 0.653 & & & & & \\
-0.781 & -0.643 & -0.746 & -0.301 & & & & \\
0.855 & 0.797 & 0.864 & 0.425 & -0.663 & & & \\
0.801 & 0.718 & 0.788 & 0.316 & -0.668 & 0.885 & & \\
0.946 & 0.883 & 0.943 & 0.521 & -0.718 & 0.948 & 0.902 & \\
0.835 & 0.727 & 0.801 & 0.417 & -0.855 & 0.686 & 0.651 & 0.772
\end{pmatrix}
\end{array}
$$

which indicates that there is a potential problem with multicollinearity.

b. The variance inflation factors are

Regressor	VIF
x_1	117.6
x_2	33.9
x_3	116.0
x_6	4.6
x_7	5.4
x_8	18.2
x_9	7.6
x_{10}	78.6
x_{11}	5.1

which indicates there is evidence of multicollinearity.

9.9 The condition indices are

1.00
9.65
61.93
126.11
2015.02
5453.08
44836.79
85564.32
5899200.59
8.86×10^{12}

which indicate a serious problem with multicollinearity.

9.11 The condition number is $\kappa = 24,031.36$ which indicates a problem with multicollinearity. The variance inflation factors shown below indicate evidence of multicollinearity.

Regressor	VIF
x_1	3.67
x_2	7.73
x_3	19.20
x_4	7.46
x_5	4.69
x_6	7.73
x_7	1.12

9.13 The condition number is $\kappa = 12400885.78$ which indicates a problem with multicollinearity. The variance inflation factors shown below indicate evidence of multicollinearity.

Coefficient	test statistic	p-value	VIF
β_1	-0.88	0.408	1.00
β_2	0.16	0.874	1.901
β_3	0.35	0.734	168.467
β_4	-0.19	0.854	43.104
β_5	-0.47	0.655	60.791
β_6	-0.12	0.911	275.473
β_7	-0.10	0.925	185.707
β_8	-0.23	0.822	44.363

9.15 The condition number is $\kappa = 286096.79$ which indicates a problem with multicollinearity. The variance inflation factors shown below indicate evidence of multicollinearity, especially x_2 and x_3.

Coefficient	test statistic	p-value	VIF
β_1	3.09	0.009	1.519
β_2	5.70	0.000	26.284
β_3	3.91	0.002	26.447
β_4	0.21	0.840	2.202
β_5	-0.21	0.833	1.923

9.17 a. Using $k = .008$ gives a model with $R^2 = 97.8\%$ and $\sqrt{MS_{Res}} = .041$.

b. Without the use of ridge regression it is .0196 and with ridge regression it is .0218, which is an increase of about 11%.

c. Both are good models.

9.19 a. The ridge trace leads to $k = .18$, but the resulting model is not adequate.

b. Without the use of ridge regression it is 0.00104 and with ridge regression it is 0.56265, which is an increase of about 540%.

c. Without the use of ridge regression it is 99.2% and with ridge regression it is 43.7%, which is an decrease of over 50%.

9.21 a. Principal components regression yields $R^2 = 96.5\%$ while least squares yields $R^2 = 98.2\%$. The loss is minimal at around 2%.

b. The coefficient vector is reduced to one term.

c. The principal components model has virtually the same R^2 but has a higher $SS_E = 0.0351$ compared to the $SS_E = 0.0218$ with the ridge model.

9.23 a. The variance inflation factors are given below.

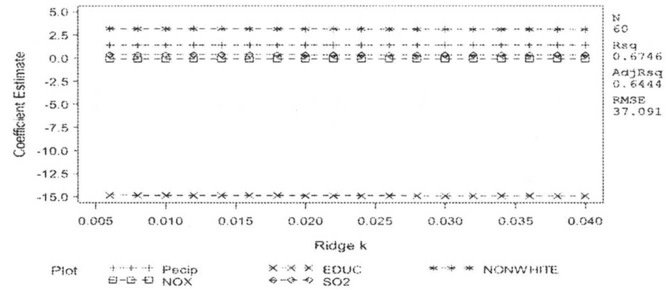
Regressor	VIF
PRECIP	2.0
EDUC	1.5
NONWHITE	1.3
NOX	1.7
SO2	1.4

The correlation matrix is

$$
\begin{array}{c|cccc}
 & PREC & EDUC & NONWHITE & NOX \\
\hline
EDUC & -0.490 & & & \\
NONWHITE & 0.403 & -0.209 & & \\
NOX & -0.486 & 0.230 & 0.025 & \\
SO2 & -0.107 & -0.234 & 0.162 & 0.412
\end{array}
$$

There is no evidence of multicollinearity.

b. The ridge trace shows flat lines.



c. The ridge trace indicates $k = 0$, therefore the estimates of the coefficients for ridge and OLS are the same.

d. Principle-component regression gives

Eigenvalue	1.9648	1.4736	0.8348	0.4062	0.3206
Proportion	0.393	0.295	0.167	0.081	0.064
Cumulative	0.393	0.688	0.855	0.936	1.000

Variable	PC1	PC2	PC3	PC4	PC5
PRECIP	-0.641	0.007	0.093	0.038	0.761
EDUC	0.490	-0.305	0.551	0.510	0.323
NONWHITE	-0.345	0.410	0.750	0.011	-0.387
NOX	0.471	0.484	0.167	-0.596	0.401
SO2	0.095	0.710	-0.312	0.619	0.080

The principal components regression accounts for 85.5% of the variation with three variables while OLS (and ridge regression since k=0) accounts for only 67.5% of the variation in the model with five variables.

9.25 The shrinkage is on the scale versus the location.

9.27 You cannot find the k that minimizes $E(L_1^2)$ because the k does not depend on j. Thus the sums will not collapse making it impossible to isolate k (see problem 9.24).

9.29 Attempting to shrink only the independent variables that are contributing to the multicollinearity instead of shrinking the entire vector of independent variables will introduce less bias. However, shrinking only a subset of the regressors can create new problems and one must be sure of the subset they are choosing to shrink. It is still better to use ordinary ridge regression.

Chapter 10: Variable Selection and Model Building

10.1 a. With $\alpha = 0.10$, the model chosen is $y = \beta_0 + \beta_2 x_2 + \beta_7 x_7 + \beta_8 x_8 + \varepsilon$.

b. With $\alpha = 0.10$, the model chosen is $y = \beta_0 + \beta_2 x_2 + \beta_7 x_7 + \beta_8 x_8 + \varepsilon$.

c. With $\alpha_{IN} = 0.05$ and $\alpha_{OUT} = 0.10$, the model chosen is $y = \beta_0 + \beta_2 x_2 + \beta_7 x_7 + \beta_8 x_8 + \varepsilon$.

d. The three procedures chose the same model.

10.3 The choice of cut-off values is to prevent the circular addition-subtraction of the variables.

10.5 a. The model involves just x_1 with $R_p^2 = 75.3\%$, $C_p = -1.8$ and $\sqrt{MS_{Res}} = 3.12$.

b. Stepwise leads to the same model involving just x_1.

10.7 When $F_{IN} = F_{OUT} = 4.0$, the model involves only x_6. However, when $F_{IN} = F_{OUT} = 2.0$, the model involves x_6 and x_7.

10.9 The model involves x_1, x_2, x_3, and x_4 with $R_p^2 = 95.3\%$, $C_p = 5$ and $\sqrt{MS_{Res}} = .002$.

10.11 The model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4$ which has a PRESS = 85.35 and $R^2_{Pred} =$ 96.86%.

10.13 a. From Section 3.7, we get $\widehat{\boldsymbol{\beta}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}^0$ with $\mathbf{W}'\mathbf{W} = \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix}$. Therefore, $(\mathbf{W}'\mathbf{W})^{-1} = \begin{pmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{12}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{pmatrix}$ which means that $Var(\widehat{\beta}_1^*) = \frac{\sigma^2}{1-r_{12}^2}$.

b. Since we are fitting a model with only one regressor, the $\mathbf{W}'\mathbf{W}$ is the scalar 1. Thus its inverse is also the scalar 1 and the $Var(\widehat{\beta}_1) = \sigma^2$.

c. We have seen from problem 3.31 earlier that in general $E(\widehat{\boldsymbol{\beta}}_1) = \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2$. For this problem, we have only 2 parameters and we are using the correlation form of the variables. Thus, $E(\widehat{\beta}_1) = \beta_1 + r_{12}\beta_2$ since the $\mathbf{W}_1$'s are the scalar 1 and $\mathbf{W}_2$ is the scalar r_{12}.

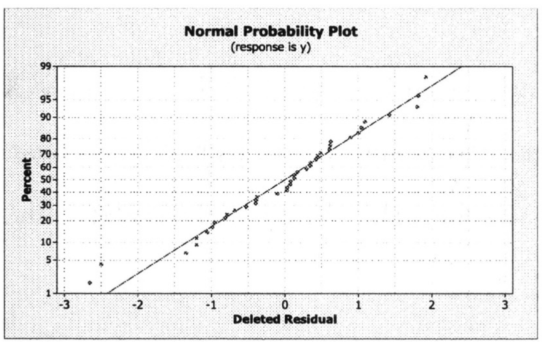
d. $MSE = Var(\widehat{\beta}_1) + [E(\widehat{\beta}_1) - \beta_1]^2 = \sigma^2 + r_{12}^2\beta_2^2$. For $\widehat{\beta}_1$ to be preferable, we need $MSE(\widehat{\beta}_1) < Var(\widehat{\beta}_1^*)$ which can be written as satisfying $\beta_2^2 < \frac{\sigma^2}{1-r_{12}^2}$.

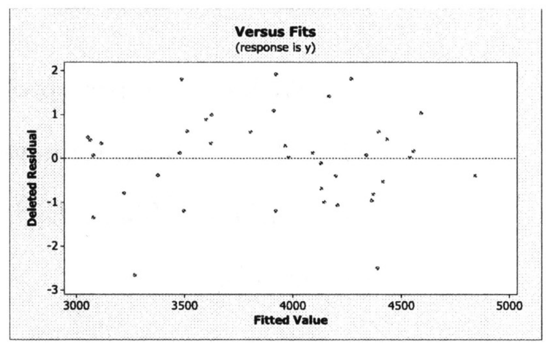
10.15 Stepwise produces the model with x_4, x_5, r_1 and r_2 which is the model with the lowest C_p from 9.14 part a.

10.17 The model with SOAKTIME and DIFFTIME is selected with $C_p = 2.8$. There is a slight departure from normality and the several outliers and influential points.

10.19 The model with DIFFTIME, x_1 and x_2 is selected with $C_p = 4.7$. There is still a slight departure from normality and the several outliers and influential points.

10.21 The model with x_1, x_3, x_5 and x_6 is selected with $C_p = 5.6$. There is a departure from normality in the tails but the residual plot show the model is adequate. There are a couple of outliers.





10.23 The confidence intervals for the model in 10.21 are narrower than those for 10.22. Also, the value for the PRESS statistic is smaller, 31685.4 compared to 34081.6.

10.25 Stepwise produces the same model as 10.24.

10.27 a. The model with x_1, x_2, x_3 and x_4 is selected ($C_p = 4.3$) which is the same model as in 10.24.

b. Stepwise produces the same model.

c. The confidence intervals for the new data set without observation 2 are narrower than the one from 10.26. The large residual from observation 2 increased MS_{Res} which in turn widened the confidence intervals in 10.26.

10.29 a. As in Section 10.4, we will use the log of the response and the log of viscosity for the model. For Run 0, performing best subsets produces the following table.

Variables	R-sq	R-sq(adj)	Mallow's Cp	S	$Log(x_1)$	x_2	x_3	x_5	x_6
1	64.7	62.2	6.7	0.12703	X				
1	20.1	14.4	30.2	0.19113		X			
2	75.2	71.4	3.1	0.11043	X	X			
2	67.6	62.6	7.1	0.12635	X		X		
3	78.3	72.8	3.5	0.10769	X	X			X
3	76.5	70.7	4.4	0.11191	X	X		X	
4	80.2	73.0	4.5	0.10737	X	X		X	X
4	79.3	71.8	4.9	0.10972	X	X	X		X
5	81.1	71.6	6.0	0.11006	X	X	X	X	X

From this, we would choose the model with 3 variables, $Log(x_1), x_2$, and x_6, which are Log(Visc), Surface, and Voids, respectively. This gives the prediction equation $\widehat{Log(y)} = -1.54 - 0.507Log(x_1) + 0.454x_2 + 0.109x_6$.

For Run = 1, we get the following table from best subsets regression.

Variables	R-sq	R-sq(adj)	Mallow's Cp	S	$Log(x_1)$	x_2	x_3	x_5	x_6
1	59.7	56.6	6.3	0.16536	X				
1	21.6	15.5	22.6	0.23060				X	
2	71.3	66.5	3.3	0.14527	X				X
2	68.4	63.1	4.6	0.15239	X			X	
3	77.2	71.0	2.8	0.13522	X		X		X
3	74.6	67.6	3.9	0.14278	X			X	X
4	78.1	69.3	4.4	0.13901	X	X	X		X
4	77.8	68.9	4.5	0.13999	X		X	X	X
5	79.0	67.4	6.0	0.14335	X	X	X	X	X

From this, we choose a 3-variable model including $Log(x_1)$, x_3, and x_6, which are Log(Visc), base, and voids, respectively. The prediction equation is $\widehat{Log(y)} = -2.06 - 0.613Log(x_1) + 0.485x_3 + 0.187x_6$.

b. When we look at the separate runs, we see different regressors are most appropriate. While Log(Visc) and Voids are significant in both models, the percentage of asphalt in the surface course (x_2) is significant only in the first run ($x_4 = 0$). Also, the percentage of asphalt in the base course (x_3) is significant in the second run ($x_4 = 1$) but not the first.

c.

Model	R^2_{Adj}	MS_{Res}	Cp
Run = 0	72.8%	0.01160	3.5
Run = 1	71.0%	0.01828	2.8
Section 9.4	95.3%	0.09150	2.9

The model in Section 10.4 has more predictive power, but greater error than the models created for the two runs. Because the indicator variable for Run (x_4) was determined to not be significant in the model in Section 10.4, we would not expect an advantage in modeling the runs separately, other than this decrease in error.

10.31 The model with age and severity is selected ($C_p = 2.0$) from the all-possible-regressions selection. This same modal is selected from stepwise regression. An analysis of this data can be found in Section 3.6.

10.33

The all-possible-regressions selection on the wine quality of young red wines produced multiple candidate models. We first chose to look at a 6 regressor model (x_1, x_2, x_3, x_4, x_5, x_8) with a $C_p = 5.7$, $R^2 = 66.6$, $R^2_{adj} = 58.5$, and $s = 1.1403$.

The VIF's still indicate a problem with multicollinearity between x_4 and x_5. Without any advice from a subject matter expert, the decision was made to remove x_4 from the model. This results in a slight increase in s, but this is preferred since the model no longer suffers from the multicollinearity problem. The residual analysis does not indicate any problems with model adequacy.

Stepwise regression suggested the simple linear regression model only containing x_5. The fit criteria for this model include $s = 1.27181$, $R^2 = 50.1\%$ and $R^2_{adj} = 48.4\%$.

Chapter 11: Validation of Regression Models

11.1 a. PRESS = 87.4612 with

$$R^2_{Pred} = 1 - \frac{\text{PRESS}}{SS_T}$$

$$= 1 - \frac{87.4612}{326.964}$$

$$= 73.25\%$$

The predictive power is not bad.

b. $\hat{y} = -8.5.004x_2 + .28x_7 - .005x_8$

y	$\hat{y}$
10	5.83
11	8.84
11	12.07
4	0.73
10	7.46
5	2.82

The model does not predict very well.

c. The model does a good job predicting these observations.

City	y	$\hat{y}$
Dallas	11	10.71
Los Angeles	10	12.25
Houston	5	5.29
San Francisco	8	8.42

11.3 PRESS = 70.82 with $R^2_{Pred} = 59.5\%$ which agrees with problem 11.2 that the model does not predict well.

11.5 a. $\hat{y}_p = 4.42 + 1.53x_1 + .012x_2$.

b. $\widehat{y}_e = 3.51 + 1.39x_1 + .016x_2$. The models are similar which indicates the overall model should be valid.

c. The model predicts fairly well and is consistent with Example 11.3.

11.7 PRESS = 337.37 with $R^2_{Pred} = 72.74\%$ which indicates that the predictive performance of the model is not bad.

11.9 The model is not predicting very well.

y	$\widehat{y}$
18.9	12.43
18.25	14.52
34.7	23.09
36.5	22.33
14.89	5.26
16.41	16.89
13.9	11.67
20.0	19.22

11.11 The standard errors are larger in the estimation set.

Problem 15.11		Problem 3.5	
Coefficient	Standard Error	Coefficient	Standard Error
$\widehat{\beta}_0$	2.409	$\widehat{\beta}_0$	1.535
$\widehat{\beta}_1$	0.009	$\widehat{\beta}_1$	0.006
$\widehat{\beta}_2$	0.936	$\widehat{\beta}_2$	0.671

11.13 The DUPLEX algorithm is probably not efficient for large sample sizes since $\binom{n}{2}$ is going to very large.

11.15 From Appendix C, we get

$$\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}}{1-h_{ii}}$$

If we postmultiply the above by $\mathbf{x}_i$ we get

$$\begin{aligned}
\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i}{1-h_{ii}} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i h_{ii}}{1-h_{ii}} \\
&= \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i[1-h_{ii}+h_{ii}]}{1-h_{ii}} \\
&= \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i}{1-h_{ii}}
\end{aligned}$$

Now, we will postmultiply the result from Appendix C by $\mathbf{X}'\mathbf{y}$

$$\begin{aligned}
\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\mathbf{X}'\mathbf{y} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}{1-h_{ii}} \\
\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\left[\mathbf{X}'_{(i)}\mathbf{y}_{(i)} \;\; \mathbf{x}_i y_i\right] &= \widehat{\boldsymbol{\beta}} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i\widehat{\boldsymbol{\beta}}}{1-h_{ii}} \\
\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)} + \left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i y_i &= \widehat{\boldsymbol{\beta}} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\widehat{y}_i}{1-h_{ii}} \\
\widehat{\boldsymbol{\beta}}_{(i)} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i y_i}{1-h_{ii}} &= \widehat{\boldsymbol{\beta}} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\widehat{y}_i}{1-h_{ii}} \\
\widehat{\boldsymbol{\beta}}_{(i)} &= \widehat{\boldsymbol{\beta}} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\widehat{y}_i}{1-h_{ii}} - \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i y_i}{1-h_{ii}} \\
\widehat{\boldsymbol{\beta}}_{(i)} &= \widehat{\boldsymbol{\beta}} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i(\widehat{y}_i - y_i)}{1-h_{ii}} \\
\widehat{\boldsymbol{\beta}}_{(i)} &= \widehat{\boldsymbol{\beta}} - \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\widehat{e}_i}{1-h_{ii}}
\end{aligned}$$

11.17 a. The model is $y = \beta_0 + \beta_1 x_1 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$ with $C_p = 5.0$.

b. The fitted model is $\widehat{y} = -302+1.11x_1+5.32x_4+1.56x_5-13.3x_6$ with $R^2_{Pred} = 99.62\%$ which indicates the model is adequate and predicts very well.

11.19 a. $R^2_{Pred} = 75.81\%$

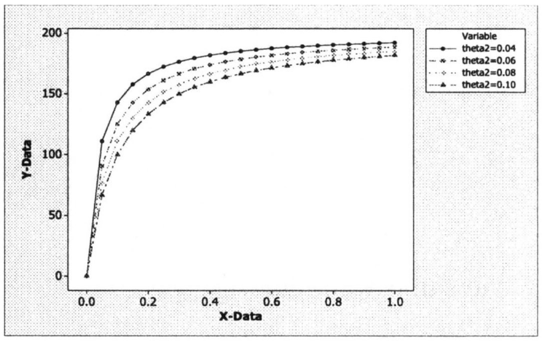
b. $\overline{R}^2_{Pred} = 77.45\%$

c. $R^2_{Pred} = 78.04\%$

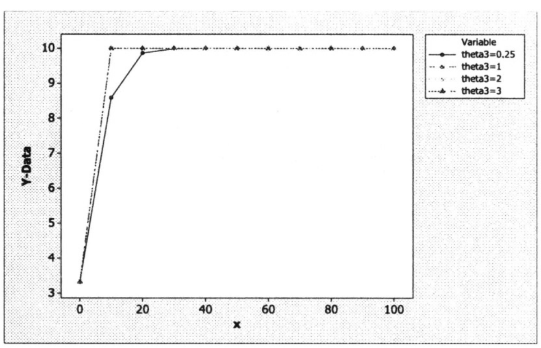
d. All three parts produce relatively the same value for R^2_{Pred}.

Chapter 12: Introduction to Nonlinear Regression

12.1 As θ_2 decreases, the curve becomes steeper.



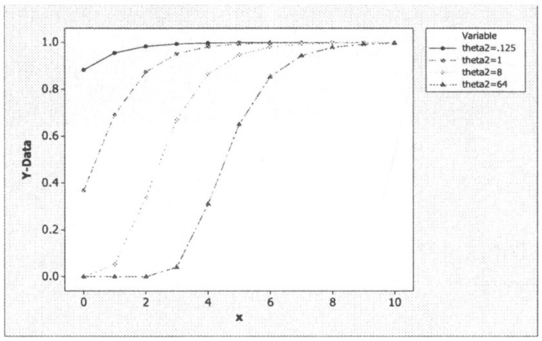
12.3 As θ_3 increases, the curve becomes steeper.



12.5 a. As θ_2 decreases, the curve becomes steeper.

b. As $x \to \infty$, $E(y) \to 1$.

c. When $x = 0$, $E(y) = \theta_1 \exp\{-\theta_2\}$.



12.7 a. This is an intrinsically linear model.

$$\begin{aligned} y &= \left[\theta_1 e^{\theta_2 + \theta_3 x}\right] \varepsilon \\ \ln(y) &= \ln(\theta_1) + \theta_2 + \theta_3 x + \ln(\varepsilon) \\ y^* &= (\theta_1^* + \theta_2) + \theta_3 x + \varepsilon^* \end{aligned}$$

b. The model is nonlinear.

c. The model is nonlinear.

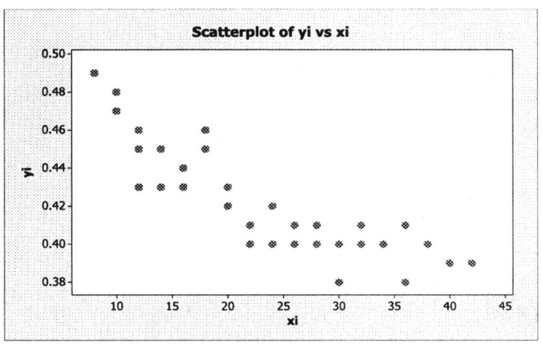
d. This is an intrinsically linear model.

$$\begin{aligned} y &= \left[\theta_1 (x_1)^{\theta_2} (x_2)^{\theta_3}\right] \varepsilon \\ \ln(y) &= \ln(\theta_1) + x_1 \ln(\theta_2) + x_2 \ln(\theta_3) + \ln(\varepsilon) \\ y^* &= \theta_1^* + x_1 \theta_2^* + x_2 \theta_3^* + \varepsilon^* \end{aligned}$$

e. The model is nonlinear.

12.9 $\widehat{y} = -.121x_2 + 1.066e^{.4928x_1}$. An approximate 95% confidence interval for θ_3 is $(-1.027, .785)$. Since this interval contains 0, we conclude there is no difference in the two days.

12.11 a.



b. $\widehat{y} = .3896 - (-.2194)e^{-.0992x}$. The starting values were obtained by plotting the expectation function.

c. $F = 141.55$ with $p =< 0.0001$ which is significant.

d. An approximate 95% confidence interval for θ_1 is (.3778, .4014). An approximate 95% confidence interval for θ_2 is $(-.2828, -.1560)$. An approximate 95% confidence interval for θ_3 is (.0626, .1357). θ_2 is not different from zero.

e. The residuals show that the model is adequate.

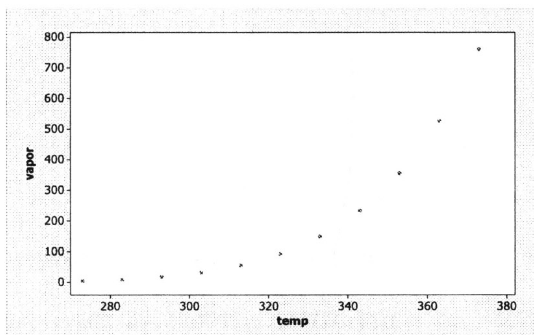
12.13 a. $\widehat{y} = .8703(x_1)^{.783}(x_2)^{.227}$.

b. $F = 422.93$ with $p = 0.000$ which is significant. Both variables appear to have important effects.

c. The residual plots are better than in 12.12. The model seems adequate.

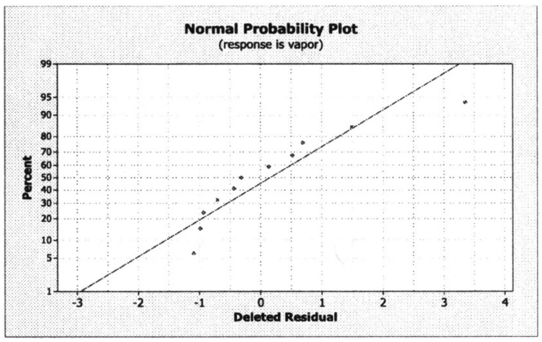
d. The nonlinear model.

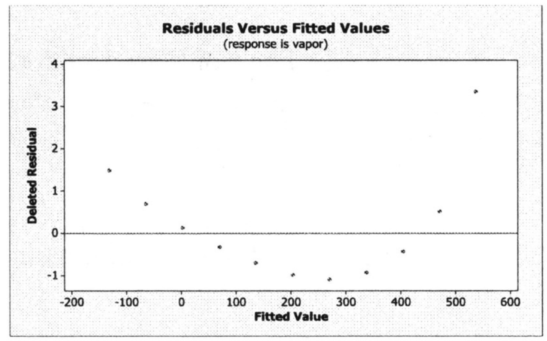
12.15 a. There is a nonlinear pattern.



b. There is a problem with normality and a nonlinear pattern in the residuals.

The regression equation is $vapor = -1956 + 6.69temp$.

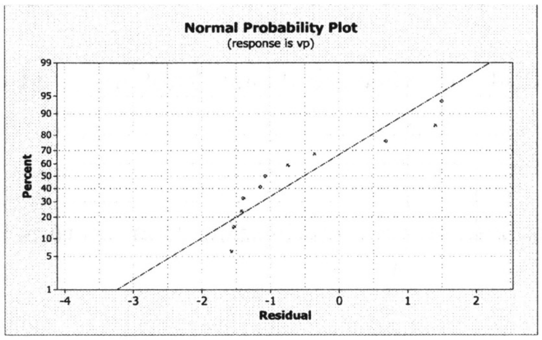


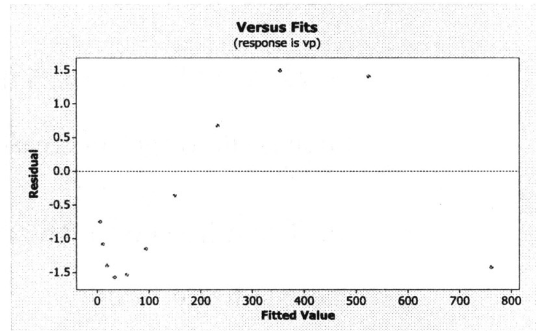


c. There is a slight improvement in the model. However, there contiues to be a problem with normality and a nonlinear pattern in the residuals.

The new regression equation is $ln(vapor) = 20.6074 - 5200.76(1/temp)$

d. The appropriate nonlinear model is $vapor = \theta_0 e^{\theta_1 (1/temp)}$. The estimated coefficients are $\widehat{\theta_0} = 576741131$ and $\widehat{\theta_1} = -5050$. We still notice a pattern in the residuals.





Note: To determine starting values, the nonlinear equation was linearized and the estimates from simple linear regression on a subset of the data were used as starting values. Another way for determining the starting value for θ_1 would be to use the chemical theory that the heat of vaporization (H_v) for water is $H_v = 9729 cal/mole$. The ideal gas constant (R) is $R = 1.9872 cal/mole°K$. Therefore,a starting value for θ_1 is $\frac{H_v}{R} = 4895.8$.

e. The simple linear regression models differ from the nonlinear model in terms of the error structures. We prefer the nonlinear model because it appears to be a better fit to the data. However, there is a still a problem with the residuals because the chemical theory assumes an idea gas and that assumption is violated with real data.

Chapter 13: Generalized Linear Models

13.1 a. $\widehat{\pi} = \dfrac{1}{1 + e^{(-6.07 + .0177x)}}$

b. Deviance $= 17.59$ with $p = 0.483$ indicating that the model is adequate.

c. $\widehat{O}_R = e^{-.0177} = .9825$ indicating that for every additional knot in speed the odds of hitting the target decrease by 1.75%.

d. The difference in the deviances is basically zero indicating that there is no need for the quadratic term.

13.3 a. $\widehat{\pi} = \dfrac{1}{1 + e^{(-5.34 + .0015x)}}$

b. Deviance $= .372$ with $p = 1.000$ indicating that the model is adequate.

c. The difference in the deviances is $\text{Dev}(x) - \text{Dev}(x, x^2) = .372 - .284 = .088$ indicating that there is no need for the quadratic term.

d. For $H_0 : \beta_1 = 0$, the Wald statistic is $Z = -.42$ which is not significant. For $H_0 : \beta_2 = 0$, the Wald statistic is $Z = -.30$ which is not significant.

e. An approximate 95% confidence interval for β_1 is $(-.0018, .0033)$ and an approximate 95% confidence interval for β_2 is $(7.15 \times 10^{-7}, 5.27 \times 10^{-7})$.

13.5 a. $\widehat{\pi} = \dfrac{1}{1 + e^{(12.35 - .0002x_1 - 1.259x_2)}}$

b. Deviance $= 14.76$ indicating that the model is adequate.

c. For $\widehat{\beta}_1$, we get $\widehat{O}_R \approx 1$ indicating that the odds are basically even. For $\widehat{\beta}_1$, we get $\widehat{O}_R = 3.52$ indicating that every one year increase in the age of the current car increases the odds of purchasing a new car by 252%.

d. $\widehat{\pi} = \dfrac{1}{1+e^{(12.35-.0002(45000)-1.259(5))}} = .76$

e. The difference in the deviances is $\text{Dev}(x_1, x_2) - \text{Dev}(x_1, x_2, x_1x_2) = 14.764 - 10.926 = 3.838$ indicating that the interaction term could be included.

f. For $H_0 : \beta_1 = 0$, the Wald statistic is $Z = -.26$ which is not significant. For $H_0 : \beta_2 = 0$, the Wald statistic is $Z = -.80$ which is not significant. For $H_0 : \beta_{12} = 0$, the Wald statistic is $Z = 1.13$ which is not significant.

g. An approximate 95% confidence interval for β_1 is $(-.0005, .0004)$, an approximate 95% confidence interval for β_2 is $(-10.827, 4.555)$ and an approximate 95% confidence interval for β_{12} is $(-.0001, .0003)$.

13.7 a. $\widehat{\pi} = e^{(-3.61-.0014x_1+.0626x_2-.0021x_3-.0289x_4)}$

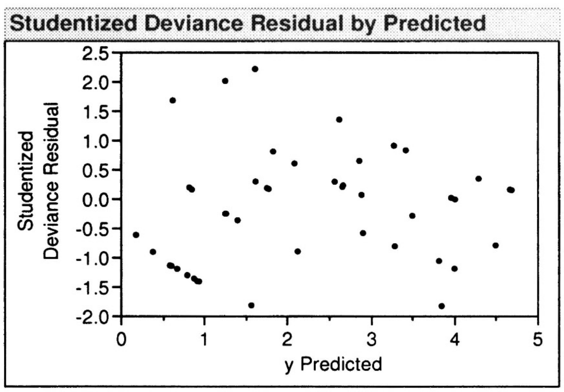
b. Deviance = 37.92 indicating that the model is adequate.

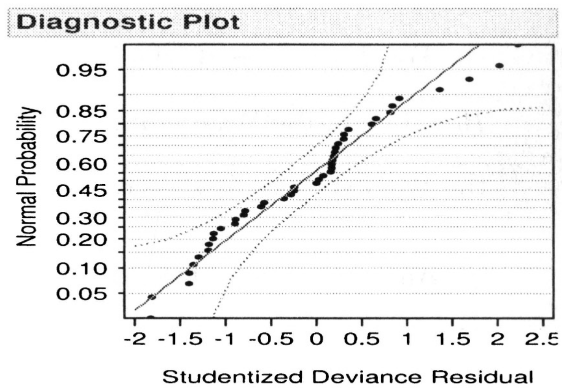
c. This indicates that x_3 should be removed.

d. Consider $\alpha = 0.05$ for all tests. For $H_0 : \beta_1 = 0$, the Wald statistic is $Z = 1.73$ which is not significant. For $H_0 : \beta_2 = 0$, the Wald statistic is $Z = 5.08$ which is significant. For $H_0 : \beta_3 = 0$, the Wald statistic is $Z = .13$ which is not significant. For $H_0 : \beta_4 = 0$, the Wald statistic is $Z = 1.87$ which is not significant.

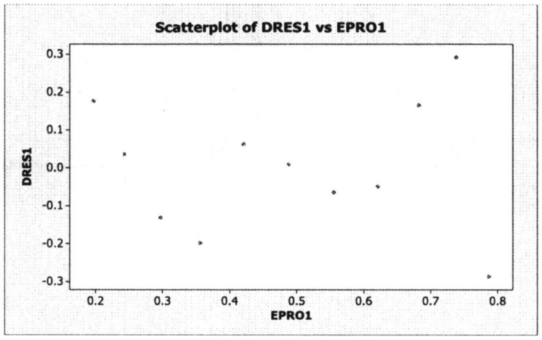
e. An approximate 95% confidence interval for β_1 is $(-.0031, .0002)$, an approximate 95% confidence interval for β_2 is $(.0384, .0867)$, an approximate 95% confidence interval for β_3 is $(-.012, .0079)$, and an approximate 95% confidence interval for β_4 is $(-.0592, .0014)$.

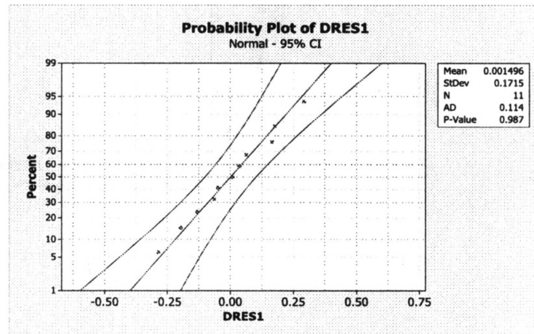
13.9 Normality seems to be satisfied but there is a pattern to the residuals.





13.11 Normality seems to be satisfied and the residual plot show that the model is satisfactory.





13.13 $f(y, r, \lambda) = a(\theta_1, \theta_2)b(y)\exp\{\sum c_j(\theta_1, \theta_2)d_j(y)$ gives $a(\theta_1, \theta_2) = \frac{\lambda^r}{\Gamma(r)}$, $b(y) = y^{-1}$, and $\sum c_j(\theta_1, \theta_2) = -\lambda y + r\ln(y)$.

13.15 Another way to write the exponential family is

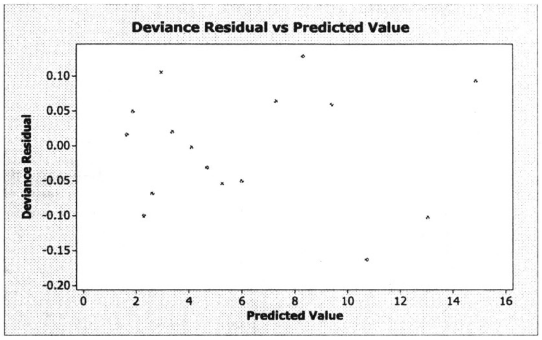
$$f(y; \theta) = B(\theta)e^{Q(\theta)R(y)}h(y)$$

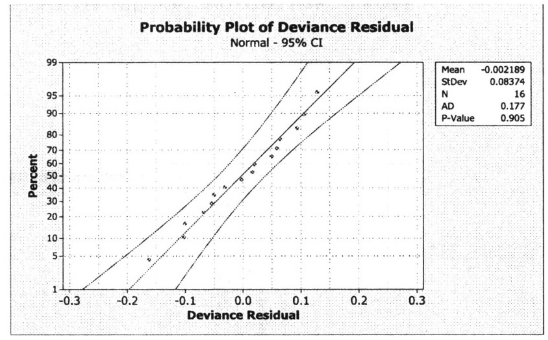
For the negative binomial, if replace $(1 - \pi)^y$ by $e^{\log(1-\pi)y}$ we get

$B(\theta)$	$Q(\theta)$	$R(y)$	$h(y)$
π^α	$\log(1 - \pi)$	y	$\binom{y+\alpha-1}{\alpha-1}$

13.17 There is no need to rework the problem since all of the regressor were important.

13.19 Both plots look good and indicate the model is adequate.





13.21 Look at

$$\begin{aligned}\widehat{\eta}(x_1+1)-\widehat{\eta}(x_1) &= \widehat{\beta}_0+\widehat{\beta}_1(x_1+1)+\widehat{\beta}_2x_2+\widehat{\beta}_{12}(x_1+1)(x_2)-\left(\widehat{\beta}_0+\widehat{\beta}_1(x_1)+\widehat{\beta}_2x_2+\widehat{\beta}_{12}x_1x_2\right)\\ &= \widehat{\beta}_1+\widehat{\beta}_{12}x_2\end{aligned}$$

Therefore, $\widehat{O}_R = e^{\widehat{\beta}_1+\widehat{\beta}_{12}x_2}$ which includes the estimated interaction coefficient and x_2 has to be fixed.

13.23 The logit model from Problem 14.5 is $\widehat{\pi} = \frac{1}{1+e^{(7.047-0.00007x_1-0.9879x_2)}}$.

G = 6.644 with p-value = 0.036, D = 18.3089 with p-value = 0.306.

The probit function is $\widehat{\pi} = \frac{1}{1+e^{(4.350-0.000046x_1-0.6099x_2}}$.

G = 6.771 with p-value = 0.034, D = 18.1819 with p-value = 0.313.

The complimentary log-log model is $\widehat{\pi} = \frac{1}{1+e^{(5.737-0.000057x_1-0.7219x_2)}}$.

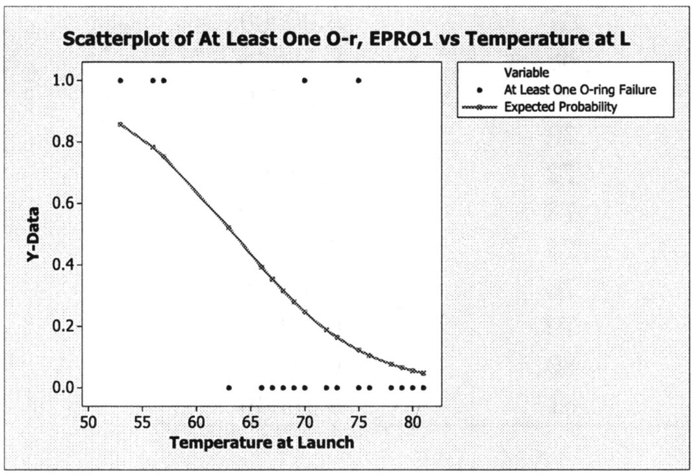
G = 6.871 with p-value = 0.032, D = 18.0827 with p-value = 0.319.

The likelihood ratio tests all show model significance for the three links. Also, the goodness of fit tests using the deviance show the models are very similar. This is to be expected since for small sample sizes, the three models do not show meaningful differences.

13.25 a. Using the logit function, $\widehat{\pi} = \frac{1}{1+e^{(-10.875+0.1713x_1)}}$.

The model fits the data well.

G=5.944 with a p-value of 0.015 and D=15.7592 with p-value=0.398.



b. $\widehat{O_R} = 0.84$ This implies that an additional degree (Fahrenheit) of temperature decreases the odds of o-ring failure by 16%.

c. $\widehat{\pi} = 0.9097$ at 50 °F.

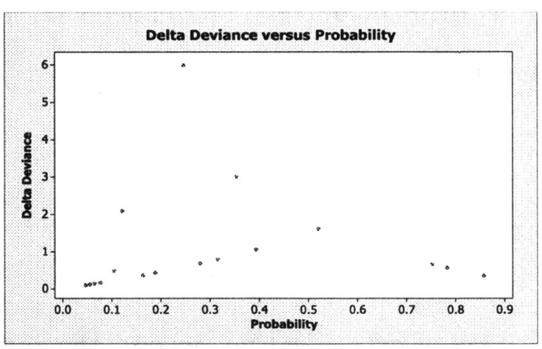
d. $\widehat{\pi} = 0.1221$ at 75 °F.

e. $\widehat{\pi} = 0.9962$ at 31 °F. There is danger in extrapolating beyond the range of temperatures used in the model, but we can see from the graph of estimated probabilities and from the calculated values in parts c and d that the probability of failure at this low temperature is very high.

f. The deviance residuals are shown below.

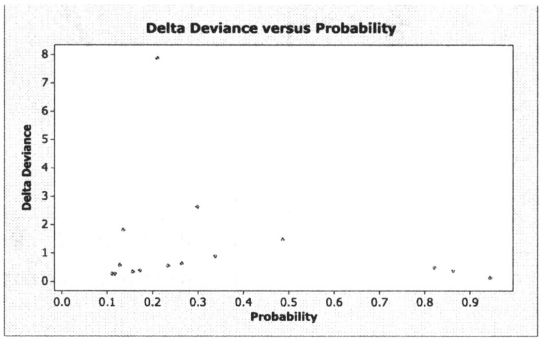
Temperature (F)	Deviance Residual
53	0.35569
56	0.56743
57	0.65666
63	1.60629
66	1.05038
67	3.00090
68	0.78648
69	0.67896
70	5.99277
72	0.43192
73	0.36997
75	2.09057
76	0.47858
79	0.14041
80	0.11883
81	0.10046

There may be some problems with the model.



g. Using the logit function,

$\widehat{\pi} = \frac{1}{1+e^{(-39.1593+1.01923x_1-0.00630x_1^2)}}$. $G = 6.386$ with $p = 0.041$. D $= 15.3177$ with $p = 0.357$. The plot of deviance residuals for this model looks better than that for the model in part a., suggesting this model may be an improvement to the original.

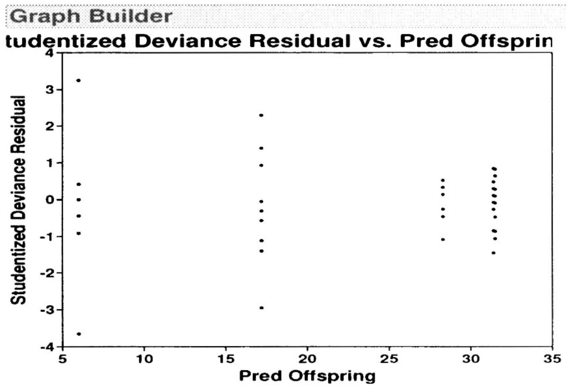


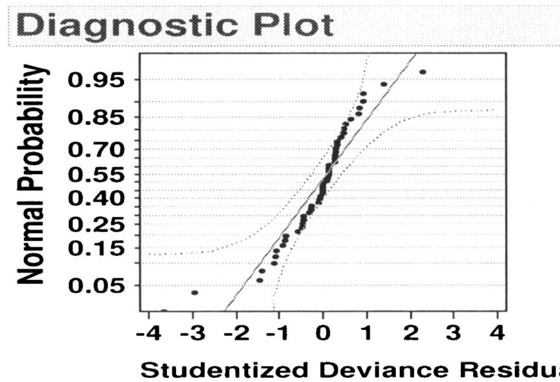
13.27

Four indicator variables were used to incorporate the five levels of dose into the analysis. A Poisson regression model with a log link was used to determine the effect of dose on the number of offspring. The model adequacy checks based on deviance ($\chi^2 = 47.44$) and the Pearson chi-square ($\chi^2 = 50.7188$) statistics are satisfactory. From the analysis, we notice when comparing to the control, dosages 235 and 310 have a significant effect on number of offspring.

Source	Test Statistic	p-value
80	0.0016	0.9682
160	1.61	0.2044
235	42.10	< 0.0001
310	189.07	< 0.0001

The residual plots show some problems with normality and model fit.



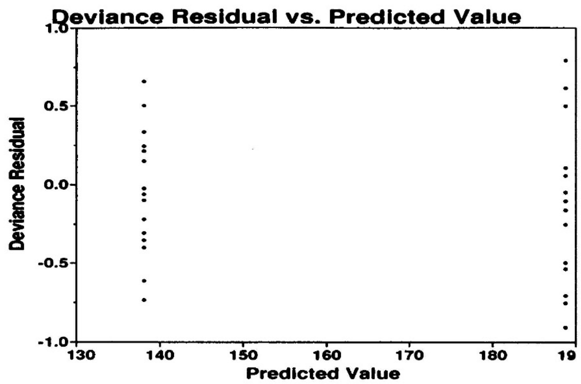


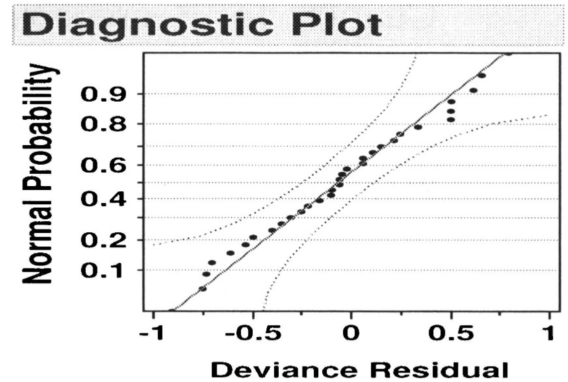
13.29

A regression with a gamma response distribution and a log link function was performed on the resistivity of a urea formaldehyde resin data. For the full model, the scaled deviance is 32.97 indicating that the model is adequate. The LR statistics for the Type III analysis indicate that some of the regressors should be removed from the model because they are not significant. Insignificant regressors were removed from the model and the resulting model only has E, the water collection time as the single predictor. The same analysis was completed using the canonical link but this had no effect on the conclusions for the analysis.

Source	Test Statistic	p-value
E	3.83	0.0503

Normality seems to be satisfied and the residual plot shows the model is satisfactory.

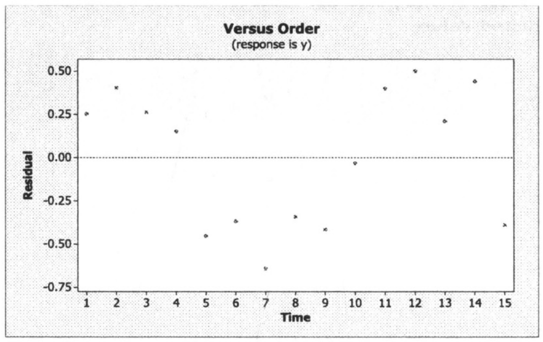

Deviance Residual vs. Predicted Value
Deviance Residual
1.0
0.5
0.0
-0.5
-1.0
130
140
150
160
170
180
19
Predicted Value


Diagnostic Plot
Normal Probability
0.9
0.8
0.6
0.4
0.2
0.1
-1
-0.5
0
0.5
1
Deviance Residual

Chapter 14: Regression Analysis of Time Series Data

14.3 a.

$\hat{y} = 24.6 - 0.0892x$. The residual plot versus time indicates there is autocorrelation.



b. $d = .81$ which rejects the null hypothesis and indicates that there is evidence of positive autocorrelation.

c. We get

$$\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_t^2}$$

$$= \frac{1.1693}{2.1610}$$

$$= .5411.$$

The new regression equation is $\widehat{y'} = 12.0854 - 0.1105x'$

The standard errors of the regression coefficients are $se(\widehat{\beta}_0') = 0.5542$ and $se(\widehat{\beta}_1') =$ 0.0.01403.

d. $d = .90$ which indicates there is still evidence of positive autocorrelation.

14.5

The regression through the origin for the first difference approach yields an estimataed slope of 0.28943 with a standard error of 0.02508. The previous estimate for $\hat{\beta}_1$ was 0.29799 with a standard error of 0.0123. As a result, the estimates are very similar, but the standard error is smaller for the Cochrane-Orcutt approach in exercise 14.4

14.7

The objective function is

$$\sum_{t=1}^{T} \frac{1}{t} \left[y_t - \hat{\beta}_0 - \hat{\beta}_1 t\right]^2.$$

Taking the derivatives with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and setting equal to 0, we obtain

$$2\sum_{t=1}^{T} \frac{1}{t} \left[y_t - \hat{\beta}_0 - \hat{\beta}_1\right](-1) = 0$$

and

$$2\sum_{t=1}^{T} \frac{1}{t} \left[y_t - \hat{\beta}_0 - \hat{\beta}_1\right](-t) = 0$$

$$2\sum_{t=1}^{T} \left[y_t - \hat{\beta}_0 - \hat{\beta}_1\right](-1) = 0$$

The resulting normal equations are

$$\hat{\beta}_0 \sum_{t=1}^{T} \frac{1}{t} + \hat{\beta}_1 = \sum_{t=1}^{T} \frac{y_t}{t}$$

and

$$T\left(\hat{\beta}_0 + \hat{\beta}_1\right) = \sum_{t=1}^{T} y_t.$$

noindent Let

$$H_t = \sum_{t=1}^{T} \frac{1}{t}.$$

We note that H_T is the *harmonic number* represented by the partial sum through T terms. The resulting solutions to the normal equations are

$$\hat{\beta}_0 = \frac{\sum_{t=1}^{T} \frac{y_t}{t} - \overline{y}}{H_T - 1}$$

and

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_0.$$

14.9

The regression equation using the Cochran-Orcutt procedure in exercise 14.3 is $\widehat{y}' = 12.0854 - 0.1105x'$ and the standard errors of the regression coefficients are $se(\widehat{\beta}_0') = 0.5542$ and $se(\widehat{\beta}_1') = 0.01403$.

The regression equation using the Cochran-Orcutt procedure in exercise 14.3 is $\widehat{y}' = 12.0854 - 0.1105x'$ and the standard errors of the regression coefficients are $se(\widehat{\beta}_0') = 0.5542$ and $se(\widehat{\beta}_1') = 0.0.01403$.

The time series regression model with autocorrelated errors produces the regression equation $\widehat{y} = 26.1875 - 0.1075x$ and the standard errors of the regression coefficients are $se(\widehat{\beta}_0) = 1.1827$ and $se(\widehat{\beta}_1) = 0.0131$.

The estimates are very similar but the standard error is smaller for the time series regression.

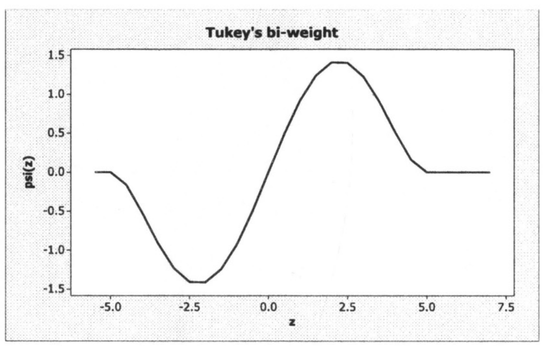
14.11

The time series regression has $\hat{\phi} = 0.600189$, which is the same as the Cochrane-Orcutt procedure from before. The previous estimate for $\hat{\beta}_1$ was 0.29799 with a standard error of 0.01230. The time series regression estimate is 0.2910 with a standard error of .009776. As a result, the estimates are very similar, but the standard error is smaller for the time series regression.

Chapter 15: Other Topics in the Use of Regression Analysis

15.1 It is possible, especially in small data sets, that a few outliers that follow the pattern of the "good" points can throw the fit off.

15.3 They are both oscillating functions that have similar shapes with Tukey's bi-weight being a faster wave. However, Tukey's bi-weight can exceed 1 while Andrew's wave function cannot.



15.5 The fitted model is $\widehat{y} = 2.34 - .288x_1 + .248x_2 + .45x_3 - .543x_4 + .005x_5$ with a couple of outliers.

15.7 a. The estimate is

$$\begin{aligned}\widehat{x}_0 &= \frac{y_0 - \widehat{\beta}_0}{\widehat{\beta}_1} \\ &= \frac{17 - 33.7}{-.0474} \\ &= 352.32\end{aligned}$$

b. First we solve the following

$$\begin{aligned}d^2\left[(-.0474)^2 - \frac{(2.042)^2(9.39)}{426101.75}\right] - 2d(-.0474)(17-20.223) \\ + \left[(17-20.223)^2 - (2.042)^2(9.39)\left(1+\frac{1}{32}\right)\right] &= 0 \\ .022d^2 - 3.055d - 29.99 &= 0\end{aligned}$$

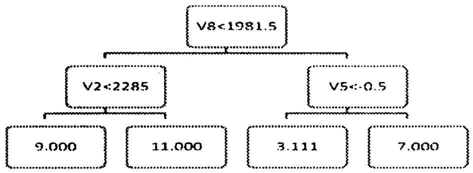
which gives $d_1 = -66.41$ and $d_2 = 205.27$. Then the confidence interval is

$$\begin{aligned}285.04 - 66.41 &< x_0 < 285.04 + 205.27 \\ 218.63 &< x_0 < 490.31\end{aligned}$$

15.9 The normal-theory confidence interval for β_2 is $.014385 \pm 1.717(.003613) = (.0082, .0206)$. The bootstrap confidence interval is $(.0073, .0240)$ which is similar to the normal-theory interval.

15.11 First, fit the model. Then, estimate the mean response at x_0. Bootstrap this m times and store all of these mean responses. Finally, find the standard deviation of these responses.

15.15 Regression tree for NFL data:



15.17 $Var(\widehat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}}{S_{XX}} \right)$ which for fixed n is minimized when $\bar{x} = 0$. If this is not possible, then the experimenter should maximize S_{XX}.

15.19 a. Let $\mathbf{D}$ be the $\mathbf{X}$-matrix without the intercept column. Then $\mathbf{D} = (\, \mathbf{d}_1 \quad \mathbf{d}_2 \quad \cdots \quad \mathbf{d}_k \,)$. Suppose the spread of the design is bounded (it has to be) then, $\mathbf{d}_i'\mathbf{d}_i \le c_i^2$ for $i = 1, 2, \ldots k$ and some constant c_i. This is equivalent to

$$d_{ii} \le c_i^2 \qquad i = 1, 2, \ldots, k$$

where d_{ii} is the i^{th} diagonal element of $\mathbf{D'D}$. It can be shown that

$$d^{ii} \ge \frac{1}{d_{ii}} \qquad i = 1, 2, \ldots, k$$

where d^{ii} is the i^{th} diagonal element of $(\mathbf{D'D})^{-1}$. There is equality in the above expression only when all the d_{ij}'s $= 0$. Therefore, if the design is orthogonal

$$
\begin{aligned}
Var(\widehat{\beta}_i) &= \sigma^2 d^{ii} \\
&\geq \frac{\sigma^2}{c_i^2} \\
&= \frac{\sigma^2}{c_i^2}
\end{aligned}
$$

since $\mathbf{d}_i'\mathbf{d}_j = 0$ for $i \neq j$ and $\mathbf{d}_i'\mathbf{d}_i = c_i^2$ when the design is orthogonal.

b. $Var(\widehat{y}) = (\sigma^2)\mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0$. Since the design is orthogonal, we have

$$
\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \begin{pmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \frac{1}{n} & 0 \\ 0 & 0 & \frac{1}{n} \end{pmatrix}
$$

Consider the center of the design as $\mathbf{0}$, then for any $\mathbf{x}_0 = \begin{pmatrix} 1 & x_i & x_j \end{pmatrix}$ it has distance from the center of $d = \sqrt{1 + x_i^2 + x_j^2}$ and

$$
\begin{aligned}
Var(\widehat{y}) &= \left(\frac{1}{n}\right)^2 + \frac{x_i^2}{n^2} + \frac{x_j^2}{n^2} \\
&= \frac{1}{n^2}\left(1 + x_i^2 + x_j^2\right) \\
&= \frac{d^2}{n^2}
\end{aligned}
$$

Thus, for any point with distance d the variance will be the same which means the design is rotatable.