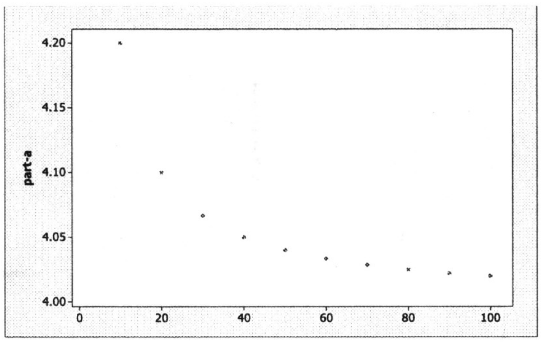
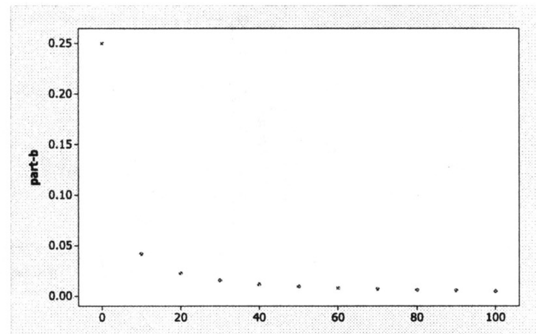
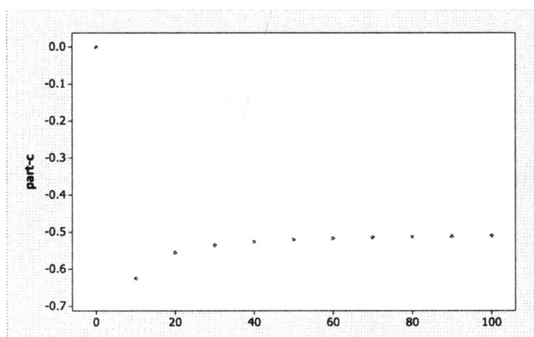
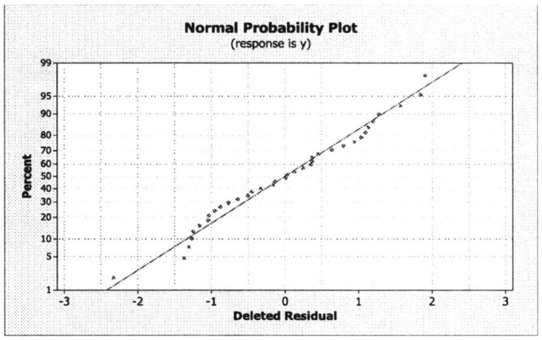
5.8 The models were sketched with $\beta_0 = 4$, $\beta_1 = 2$ and for $0 \leq x \leq 100$ by tens. The pattern is more consistent with a.

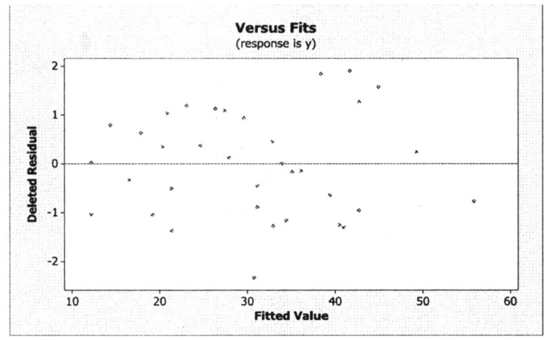




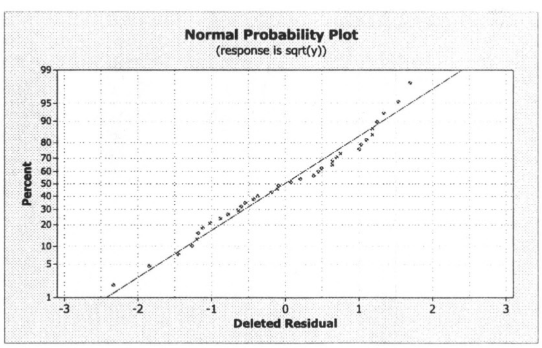


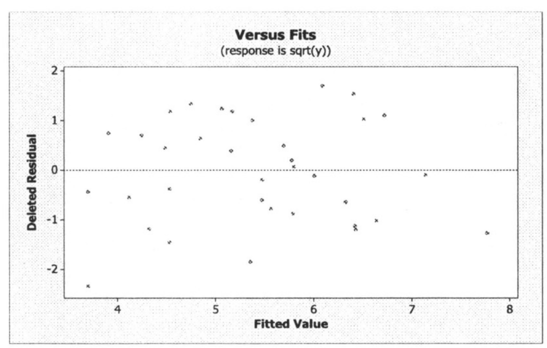
5.9 a. There is a problem with normality and a drifting in the residuals. There is an outlier at observation 28. x_2 has a nonlinear pattern.



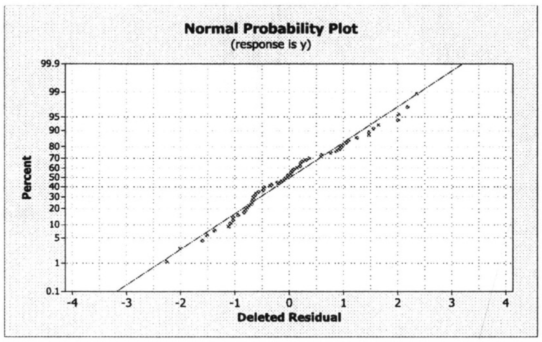


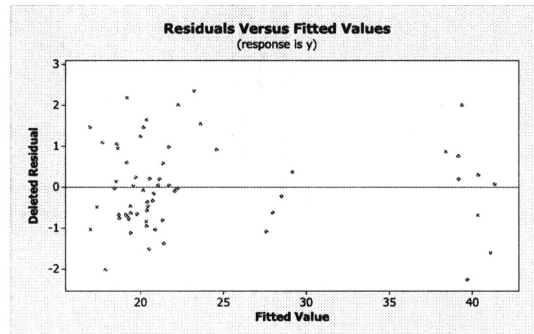
b. A square root transformation on y was used.



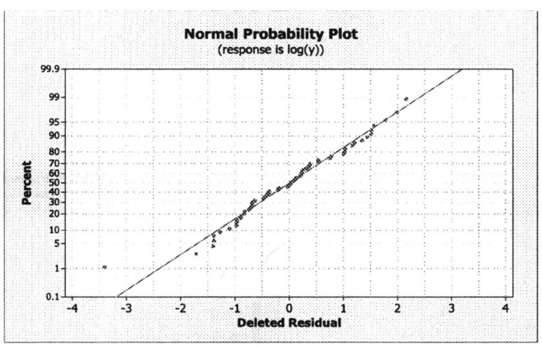


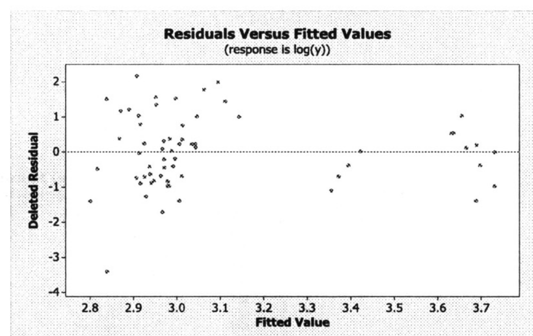
5.10 a. There is no problem with normality but a drifting in the residuals. There are outliers. x_4 has a nonlinear pattern.



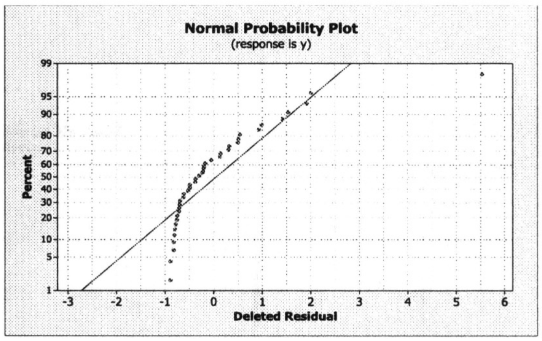


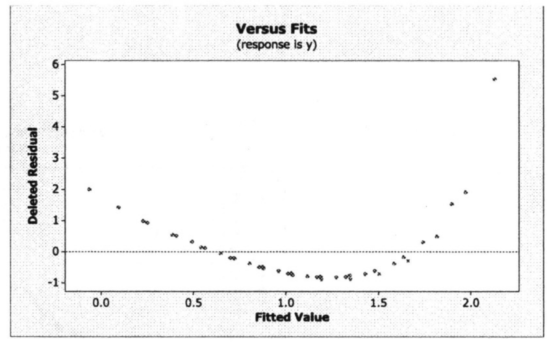
b. A natural log transformation on y was used.



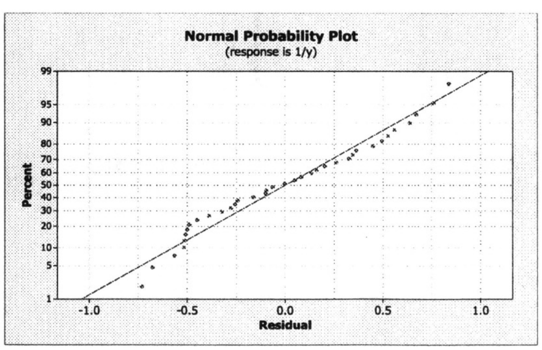


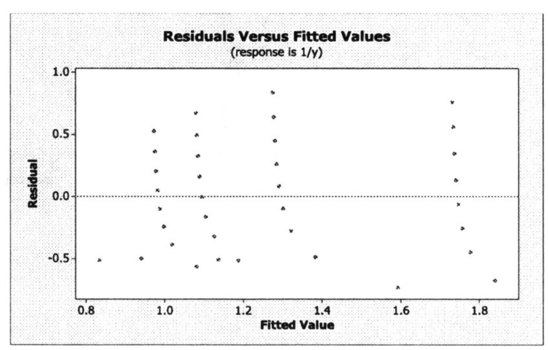
5.11 a. There is a problem with normality and a nonlinear pattern in the residuals. x_2 has a nonlinear pattern.



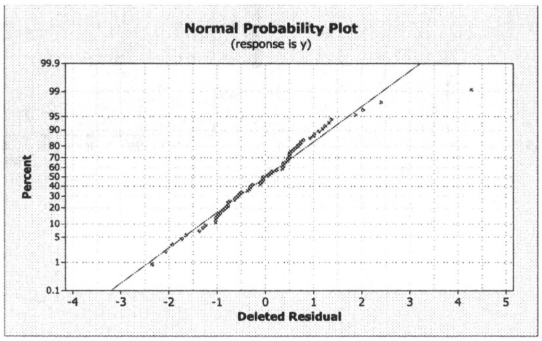


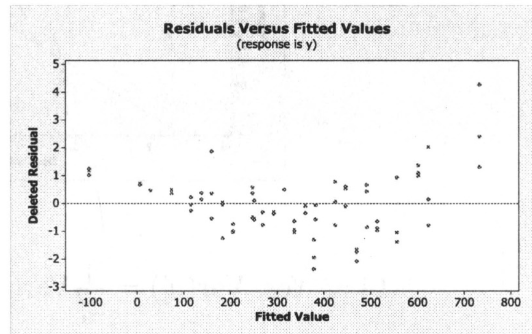
b. A transformation of $1/y$ was used along with inverting both of the independent variables.





5.12 a. There is a departure from normality in the tail. There is a nonlinear pattern to the residuals. There is nonconstant variance. There are many potential outliers.

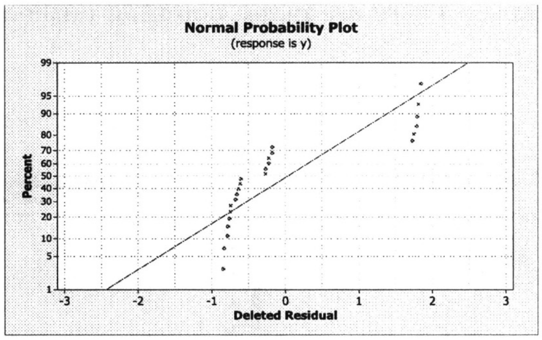


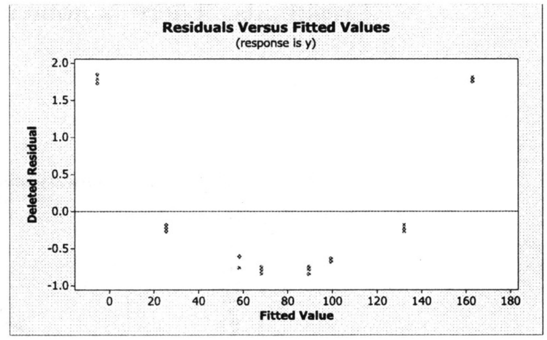


b. This corrects the nonconstant variance.

c. Use a square root transformation of the sample variance and model the sample standard deviation.

5.13 a. There is a departure from normality and a nonlinear pattern to the residuals.





5.14 a. Yes, $\mathrm{Var}(y_i') = \frac{1}{x_i^2}\mathrm{Var}(y_i) = \sigma^2$.

b. Their roles are reversed.

c. The values of the parameters are the same but, by (b), their roles are reversed.

5.15 a. $S(\beta) = \sum_{i=1}^{n} w_i(y_i - \beta x_i)^2$. Taking the derivative with respect to β and setting it equal to zero gives $\sum_{i=1}^{n} w_i(Y_i - \widehat{\beta}x_i)(-x_i) = 0$. Solving for $\widehat{\beta}$ yields $\widehat{\beta} = \frac{\sum_{i=1}^{n} w_i x_i y_i}{\sum_{i=1}^{n} w_i x_i^2}$.

b.

$$
\begin{aligned}
\mathrm{Var}(\widehat{\beta}) &= \left(\frac{1}{\sum_{i=1}^{n} w_i x_i^2}\right)^2 \sum_{i=1}^{n} w_i^2 x_i^2 \mathrm{Var}(y_i) \\
&= \left(\frac{1}{\sum_{i=1}^{n} w_i x_i^2}\right)^2 \sum_{i=1}^{n} w_i^2 x_i^2 (\sigma^2/w_i) \\
&= \frac{\sigma^2}{\sum_{i=1}^{n} w_i x_i^2}
\end{aligned}
$$

c. Here, we have $w_i = 1/x_i$. Therefore,

$$\begin{aligned}\widehat{\beta} &= \frac{\sum\limits_{i=1}^{n}(1/x_i)x_iy_i}{\sum\limits_{i=1}^{n}(1/x_i)x_i^2} \\ &= \frac{\sum\limits_{i=1}^{n} y_i}{\sum\limits_{i=1}^{n} x_i}\end{aligned}$$

with $\mathrm{Var}(\widehat{\beta}) = \frac{\sigma^2}{\sum\limits_{i=1}^{n} x_i}$.

d. Here, we have $w_i = 1/x_i^2$. Therefore,

$$\widehat{\beta} = \frac{\sum\limits_{i=1}^{n}(1/x_i^2)x_iy_i}{\sum\limits_{i=1}^{n}(1/x_i^2)x_i^2} = (1/n)\sum_{i=1}^{n}\frac{y_i}{x_i}$$

with $\mathrm{Var}(\widehat{\beta}) = \frac{\sigma^2}{n}$.

5.16 Let $\boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{pmatrix}$, p_2 be the number of parameters in $\boldsymbol{\beta}_2$, $\mathbf{K}' = (\,\mathbf{0} \quad \mathbf{I}\,)$, $\mathbf{m} = \mathbf{0}$, and the rank of $\mathbf{K}' = p_2$. Note this gives $\mathbf{K}'\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_2$. Then the appropriate test statistic is

$$F_0 \frac{\left(\mathbf{K}'\widehat{\boldsymbol{\beta}} - \mathbf{m}\right)' \left[\mathbf{K}' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{K}\right]^{-1} \left(\mathbf{K}'\widehat{\boldsymbol{\beta}} - \mathbf{m}\right)}{p_2 MSE}$$

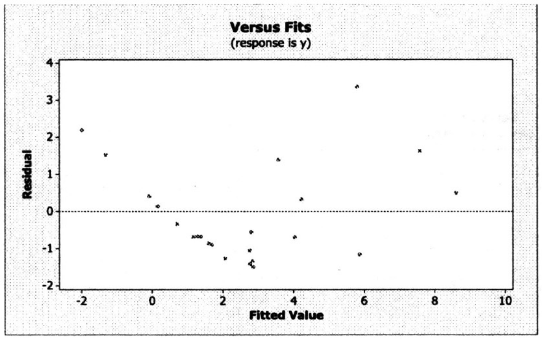
Now under H_0, F_0 above has a central F distribution and under H_1 it has a noncentral F distribution.

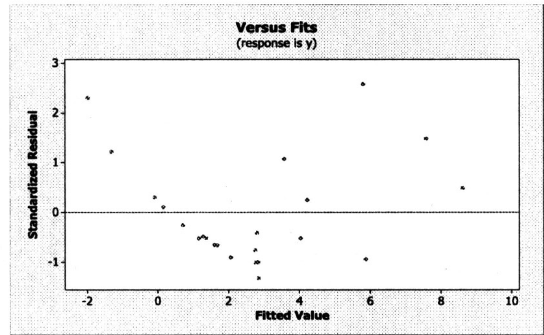
5.17 Notice that we can write the top as the quadratic form,

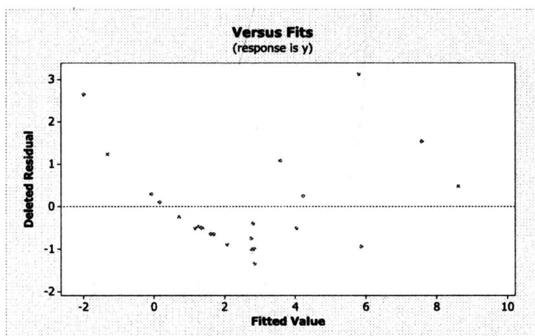
$$\mathbf{y}'\left[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]\mathbf{y}.$$

Call the matrix in the brackets $\mathbf{A}$. Then from Appendix C, we get $E(\mathbf{y}'\mathbf{A}\mathbf{y}) = \text{trace}[(\mathbf{A})(\sigma^2\mathbf{V})] + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$ where for us, $\boldsymbol{\mu} = E(\mathbf{y}) = \mathbf{0}$. It is easy to show that $[\mathbf{AV}]$ is idempotent, so it's trace is equal to its rank, which is $n - p$. Thus, in this case, $E(\mathbf{y}'\mathbf{A}\mathbf{y}) = \text{trace}[(\mathbf{A})(\sigma^2\mathbf{V})] + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} = (n-p)\sigma^2$.

5.18 a. There is a nonlinear pattern to the residuals.

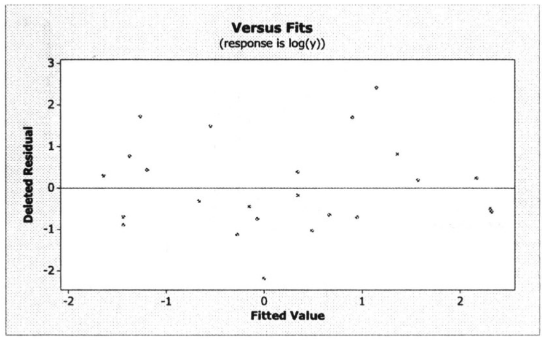






b. Use a natural log transformation on y. It does not improve the model.

c. Use a natural log transformation on each of the regressors in addition to the transformation in part b.



5.19 a.

$$\begin{aligned} Var(\mathbf{y}) &= Var(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\delta + \epsilon) \\ &= \mathbf{Z}Var(\delta)\mathbf{Z}' + Var(\epsilon) \\ &= \mathbf{Z}(\sigma_\delta^2\mathbf{I}_m)\mathbf{Z}' + \sigma^2\mathbf{I} \\ &= \sigma_\delta^2\mathbf{Z}\mathbf{Z}' + \sigma^2\mathbf{I}. \end{aligned}$$

b.

From part a, we have $Var(\mathbf{y}) = \sigma^2\mathbf{I} + \sigma_\delta^2\mathbf{Z}\mathbf{Z}' = \boldsymbol{\Sigma}$.

Then

$$\begin{aligned}\mathbf{I} &= \mathbf{\Sigma}\mathbf{\Sigma}^{-1} \\ &= [\sigma^2\mathbf{I} + \sigma_\delta^2\mathbf{Z}\mathbf{Z}'][\frac{1}{\sigma^2}\mathbf{I} - k\mathbf{Z}\mathbf{Z}'].\end{aligned}$$

In order to solve for $\mathbf{\Sigma}^{-1}$, we must solve for k. Multiplying $\mathbf{\Sigma}\mathbf{\Sigma}^{-1}$ leads us to setting the following quantity equal to 0.

$$\begin{aligned}0 &= -k\sigma^2\mathbf{Z}\mathbf{Z}' + \frac{\sigma_\delta^2}{\sigma^2}\mathbf{Z}\mathbf{Z}' - k\sigma_\delta^2\mathbf{Z}\mathbf{Z}'\mathbf{Z}\mathbf{Z}' \\ &= \mathbf{Z}[-k\sigma^2\mathbf{I} + \frac{\sigma_\delta^2}{\sigma^2}\mathbf{I} - k\sigma_\delta^2\mathbf{Z}'\mathbf{Z}]\mathbf{Z}'.\end{aligned}$$

Therefore,

$$\begin{aligned}0 &= -k\sigma^2\mathbf{I} + \frac{\sigma_\delta^2}{\sigma^2}\mathbf{I} - k\sigma_\delta^2\mathbf{Z}'\mathbf{Z} \\ &= -k\sigma^2\mathbf{I} + \frac{\sigma_\delta^2}{\sigma^2}\mathbf{I} - kn\sigma_\delta^2\mathbf{I}.\end{aligned}$$

We solve for k

$$k = \frac{\sigma_\delta^2}{\sigma^2(\sigma^2 + n\sigma_\delta^2)}.$$

Then

$$\mathbf{\Sigma}^{-1} = \frac{1}{\sigma^2}\mathbf{I} - \frac{\sigma_\delta^2}{\sigma^2(\sigma^2 + n\sigma_\delta^2)}\mathbf{Z}\mathbf{Z}'.$$

Now we must show $(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. First let's solve $\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}$

$$\begin{aligned}
\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X} &= \mathbf{X}'[\frac{1}{\sigma^2}\mathbf{I} - \frac{\sigma_\delta^2}{\sigma^2(\sigma^2+n\sigma_\delta^2)}\mathbf{Z}\mathbf{Z}']\mathbf{X} \\
&= \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} - \frac{\sigma_\delta^2}{\sigma^2(\sigma^2+n\sigma_\delta^2)}\mathbf{X}'\mathbf{Z}\mathbf{Z}'\mathbf{X} \\
&= \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} - \frac{n\sigma_\delta^2}{\sigma^2(\sigma^2+n\sigma_\delta^2)}\mathbf{X}'\mathbf{X} \\
&= \frac{1}{\sigma^2+n\sigma_\delta^2}\mathbf{X}'\mathbf{X}.
\end{aligned}$$

Now let's solve $\mathbf{X}'\mathbf{\Sigma}^{-1}$

$$\begin{aligned}
\mathbf{X}'\mathbf{\Sigma}^{-1} &= \mathbf{X}'[\frac{1}{\sigma^2}\mathbf{I} - \frac{\sigma_\delta^2}{\sigma^2(\sigma^2+n\sigma_\delta^2)}\mathbf{Z}\mathbf{Z}'] \\
&= \frac{1}{\sigma^2}\mathbf{X}' - \frac{\sigma_\delta^2}{\sigma^2(\sigma^2+n\sigma_\delta^2)}\mathbf{X}'\mathbf{Z}\mathbf{Z}' \\
&= \frac{1}{\sigma^2}\mathbf{X}' - \frac{n\sigma_\delta^2}{\sigma^2(\sigma^2+n\sigma_\delta^2)}\mathbf{X}' \\
&= \frac{1}{\sigma^2+n\sigma_\delta^2}\mathbf{X}'.
\end{aligned}$$

As a result , $(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y}$ becomes

$$\begin{aligned}
(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y} &= (\sigma^2+n\sigma_\delta^2)(\mathbf{X}'\mathbf{X})^{-1}\frac{1}{(\sigma^2+n\sigma_\delta^2)}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.
\end{aligned}$$

This proves that the ordinary least squares estimates for $\boldsymbol{\beta}$ are the same as the generalized least squares estimates.

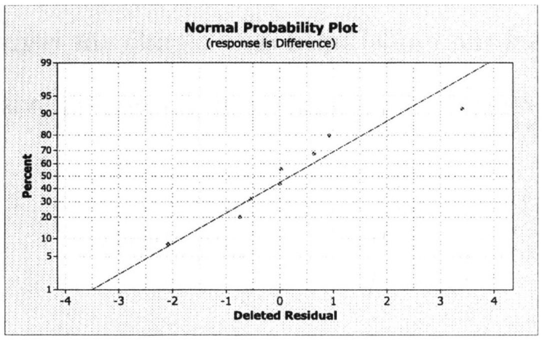
5.20 The proper analysis for the fuel consumption data is a regression analysis on the difference in fuel consumption (y) based on the batch. Because the batches of oil were divided into two with one batch going to the bus and the other batch going to the truck, a regression analysis on the difference to overcome the effect of batch. Also, for the regression analysis, we reduced the model until only significant regressors were present in the model. This leaves regressors x_4 and x_5 in the model, viscosity and initial boiling point.

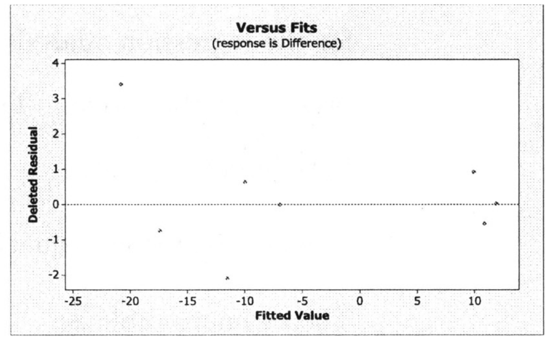
The new regression equation is $\widehat{y}_{Difference} = -106 - 13.0x_4 + 0.651x_5$.

The estimate table is:

Coefficient	test statistic	p-value
β_4	-3.09	0.027
β_5	8.99	0.000

The residual plots for this reduced model are seen below. The analysis on the difference in fuel consumption has alleviated the problems identified in problem 4.27.





5.21 A regression analysis using 3 indicator variables for mix rate was carried out for the tensile strength data. (Note: An ANOVA anlaysis could be performed on this data. The results from the ANOVA are equivalent with the results from the regression analysis.)

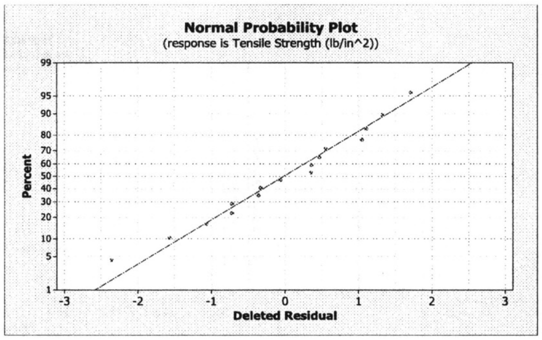
The regression equation is $\widehat{y} = 2666 + 430x_{150} + 490x_{175} + 267x_{200}$.

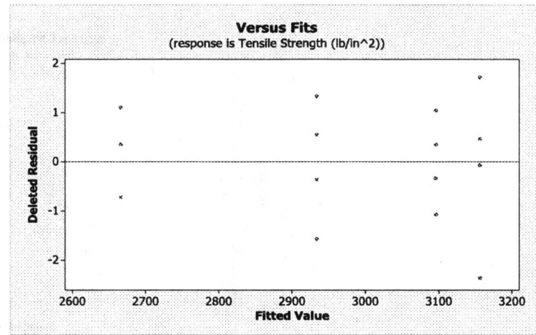
The estimate table is:

Coefficient	test statistic	p-value
β_{150}	5.83	0.000
β_{175}	6.65	0.000
β_{200}	3.63	0.003

The regression indicates that mix rate (rpm) has an effect on tensile strength. The p-values from the estimate table are computed from comparisons of average tensile strength from mix rates 150, 175, and 200 with mix rate 225. The average tensile strengths for mix rates 150, 175, and 200 are significantly higher compared to the tensile strength for a mix rate of 225.

The residual plots for this model seen below indicate no problems.





5.22 A regression analysis using 3 indicator variables for temperature was carried out for the density data. (Note: An ANOVA analysis could be performed on this data. The results from the ANOVA are equivalent with the results from the regression analysis.)

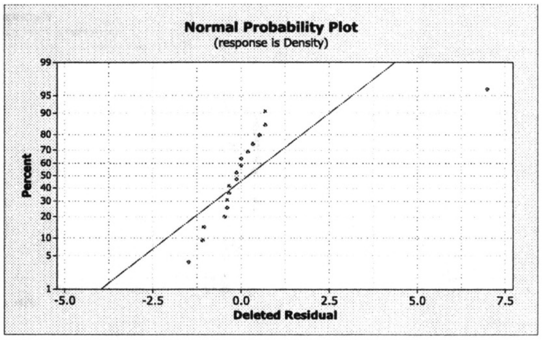
The regression equation is $\hat{y} = 23.2 - 1.46x_{900} - 0.700x_{910} - 0.280x_{920}$.

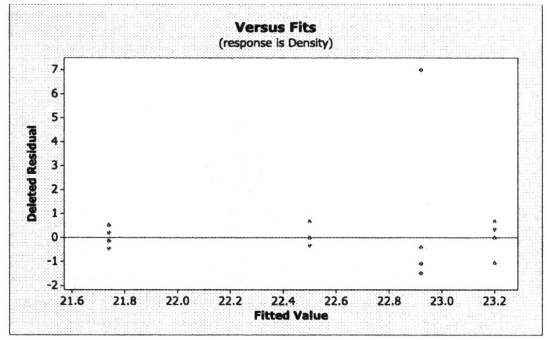
The estimate table is:

Coefficient	test statistic	p-value
β_{900}	-6.61	0.000
β_{910}	-3.00	0.009
β_{920}	-1.27	0.226

The regression indicates that peak kiln temperature has an effect on density of bricks. The p-values from the estimate table are computed from comparisons of average density from temperatures 900, 910, and 920 with temperature 930. The average density for temperatures 900 and 910 are significantly lower compared to the average density at 930. The average density is not significantly different for temperatures at 920 and 930.

The residual plots indicate a potential outlier in observation 10 (Temp. 920, Density 23.9).





5.23 This Fixed Effects tests for the subsampling analysis indicates that the three vat pressures do not have a significant effect on strength ($F = 2.3984$, $p-value = 0.1716$). The variance component for batch is 0.743. A high percentage (73%) of the total variability is due to the batch-to-batch variability.

Effect Tests

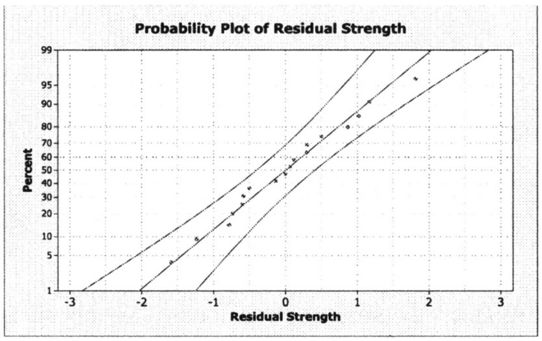
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Pressure	2	2	4.2238889	2.3984	0.1716

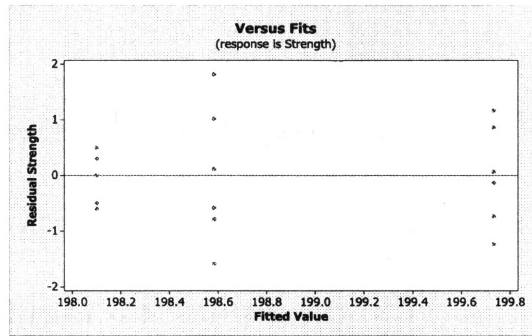
REML Variance Component Estimates

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
Batch[Pressure]	2.7020202	0.7430556	0.5125044	-0.261435	1.7475457	72.988
Residual		0.275	0.1296362	0.1301072	0.9165344	27.012
Total		1.0180556				100.000

-2 LogLikelihood 41.917470887

The plot of the residuals versus fits shows that the model is reasonable and the normal probability plot does not show a problem with the normality assumption.





Chapter 6: Diagnostics for Leverage and Influence

6.1 Observation 1 is identified as influential. It affects the coefficients for x_3 and x_5.

6.2 No observations show up as influential.

6.3 Observation 14 is identified as influential. It seriously affects the coefficients for x_5 and x_6.

6.4 No observations show up as influential.

6.5 No observations show up as influential.

6.6 No observations show up as influential.

6.7 No observations show up as influential.

6.8 Observations 50-53 show up as influential.

6.9 Observation 31 shows up as influential.

6.10 Appendix C establishes that $\widehat{\boldsymbol{\beta}}_{(i)} - \widehat{\boldsymbol{\beta}} = \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i e_i}{1-h_{ii}}$. Therefore,

$$
\begin{aligned}
D_i &= \frac{\left(\widehat{\boldsymbol{\beta}}_{(i)} - \widehat{\boldsymbol{\beta}}\right)' \mathbf{X}'\mathbf{X} \left(\widehat{\boldsymbol{\beta}}_{(i)} - \widehat{\boldsymbol{\beta}}\right)}{pMS_{Res}} \\
&= \frac{\mathbf{x}_i (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i e_i^2}{(1-h_{ii})^2 pMS_{Res}} \\
&= \left(\frac{e_i}{1-h_{ii}}\right)^2 \left(\frac{h_{ii}}{pMS_{Res}}\right) \\
&= \left(\frac{e_i^2}{MS_{Res}(1-h_{ii})}\right) \left(\frac{1}{p}\right) \left(\frac{h_{ii}}{1-h_{ii}}\right) \\
&= \frac{r_i^2}{p} \left(\frac{h_{ii}}{1-h_{ii}}\right)
\end{aligned}
$$

6.11 Appendix C establishes

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

Therefore,

$$
\begin{aligned}
COVRATIO_i &= \frac{\left|\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1} S^2_{(i)}\right|}{\left|(\mathbf{X}'\mathbf{X})^{-1} MS_{Res}\right|} \\
&= \frac{\left(S^2_{(i)}\right)^p}{MS^p_{Res}} \frac{\left|(\mathbf{X}'\mathbf{X})^{-1} + \dfrac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}}{1-h_{ii}}\right|}{\left|(\mathbf{X}'\mathbf{X})^{-1}\right|} \\
&= \left[\frac{S^2_{(i)}}{MS_{Res}}\right]^p \left(\frac{\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i + \dfrac{\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i}{1-h_{ii}}}{\mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i}\right)
\end{aligned}
$$

(note the determinants have been dropped because they are scalars)

$$
\begin{aligned}
&= \left[\frac{S^2_{(i)}}{MS_{Res}}\right]^p \left(\frac{h_{ii} + \dfrac{h^2_{ii}}{1-h_{ii}}}{h_{ii}}\right) \\
&= \left[\frac{S^2_{(i)}}{MS_{Res}}\right]^p \left(\frac{1}{1-h_{ii}}\right)
\end{aligned}
$$

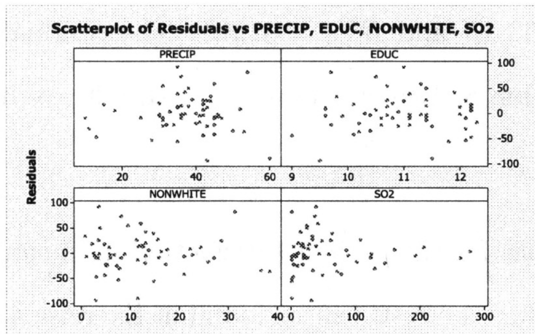
6.12 No observations show up as influential.

6.13 The last observation shows up as influential.

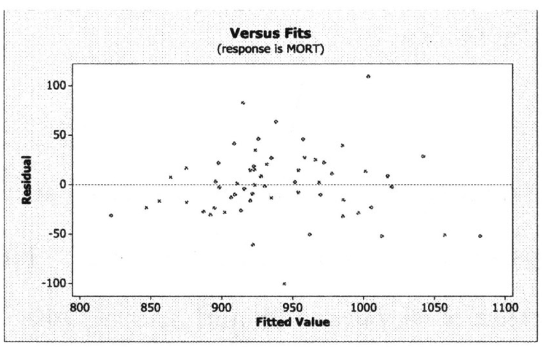
6.14 Observation 20 shows up as influential.

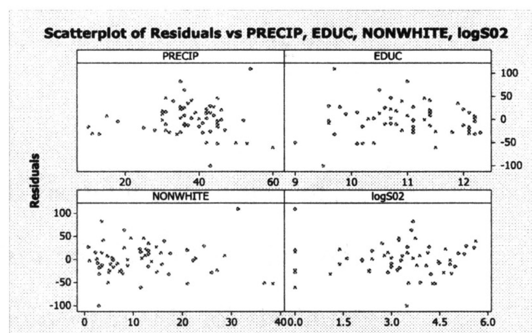
6.15 Observations 2 and 4 show up as influential.

6.16 In looking at the plots of the residuals vs. the predictors, we can see a pattern with SO2.



We take the log of SO2 to obtain the model $\hat{y} = 942 - 13.8EDUC + 3.34NONWHITE + 1.67PRECIP + 34.3logSO2$ (Recall that NOX was not significant in our previous analyses.) The model is significant with $F = 30.14$ and $p = 0.000$ with an $R^2 = 68.7\%$ and $R^2_{Adj} = 66.4\%$. The residuals look fine plotted against the fitted values and the individual regressors. None of the observations are influential.





6.17 For all three models, we transform the data using square roots of both the response and the regressors. For Life Expectancy, this gives the model, $\widehat{y^*} = 8.67 - 0.0323sqrt(x_1) - 0.00713sqrt(x_2)$. $F = 30.25$ with $p = 0.000$, so the model is significant. $R^2 = 63.4\%$ and $R^2_{Adj} = 61.3\%$. The residuals look fine, except for the outlier from observation 8. Observations 8, 21, and 30 are influential for each model.

6.18 The regression analysis for the patient satisfaction data can be found in section 3.6 of the text and the residual analysis can be found in Exercise 4.26. The influence analysis for this regression indicates that observations 9 and 17 are highly influential.

6.19 From Exercise 5.20 we recognized that the analysis for the fuel consumption data requires an analysis on the difference in fuel consumption for buses versus trucks. See Exercise 5.20 for the regression analysis of these data. The residuals indicate observation 5 as a possible outlier. The influence analysis for this regression indicates that observation 5 is influential for the model.

6.20 The regression analysis for the wine quality of young red wines data can be found in Exercise 3.19 and the residual analysis can be found in Exercise 4.28. The influence analysis for this regression indicates that observations 28 and 32 are highly influential.

6.21 The regression and residual analysis for the methanol oxidation data can be found in Exercise 5.7. To improve the model we took a log transformation of the response and reduced the model to only contain the significant predictor x_3. The influence analysis for this regression indicates that observation 1 is highly influential.

Chapter 7: Polynomial Regression Models

7.1 Yes there are potential problems since the correlation $(x,\, x^2) = .995$

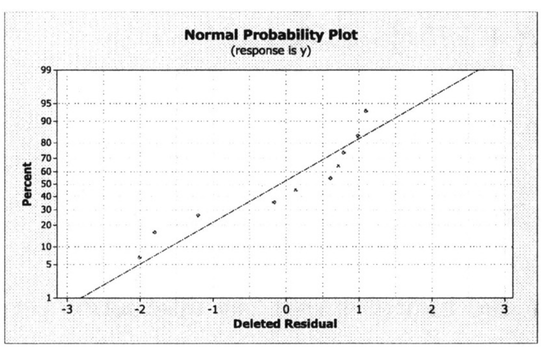
7.2 a. $\widehat{y} = 1.63 - 1.23x + 1.49x^2$.

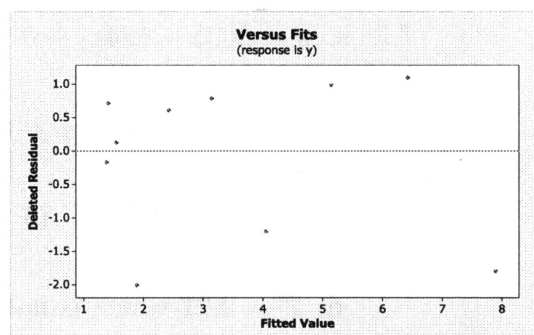
b. $F = 1.86 \times 10^6$ with $p = 0.000$ which is significant.

c. $F = \frac{4.607}{0.000} \approx \infty$ which is significant.

d. Since it is a quadratic model, there can be potential hazards in extrapolating.

7.3 There is a problem with normality. The residuals seem to show that the model is adequate.





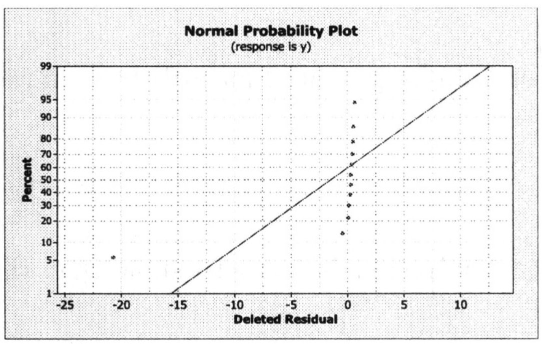
7.4 a. $\widehat{y} = -4.5 + 1.38x + 1.47x^2$.

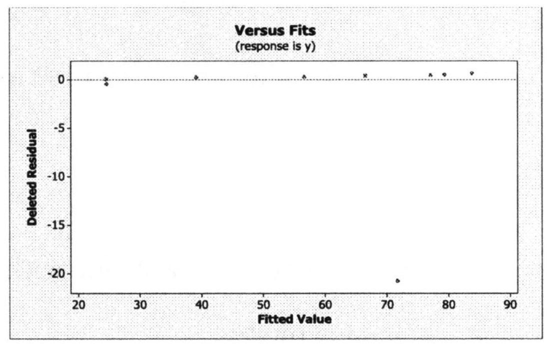
b. $F = 1044.99$ with $p = 0.000$ which is significant.

c. $F = 48.7$ with $p = 0.001$ which is indicates lack of fit.

d. $F = \frac{24.3}{2.7} = 9$ which is significant and indicates the term cannot be deleted.

7.5 There is an outlier which affects the normality and the residual plot which shows the model is not adequate.





7.6 a. $\widehat{y} = 3025 - 194x_1 - 6.1x_2 + 3.63x_1^2 + 1.15x_2^2 - 1.33x_1x_2$.

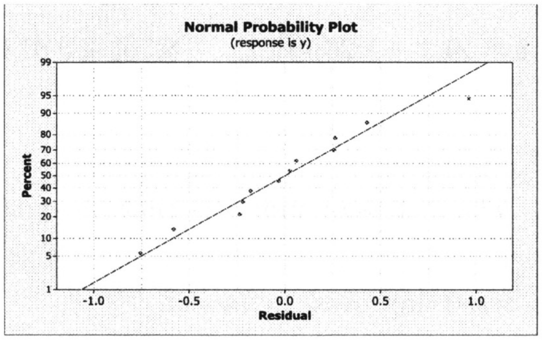
b. $F = 177.17$ with $p = 0.000$ which is significant.

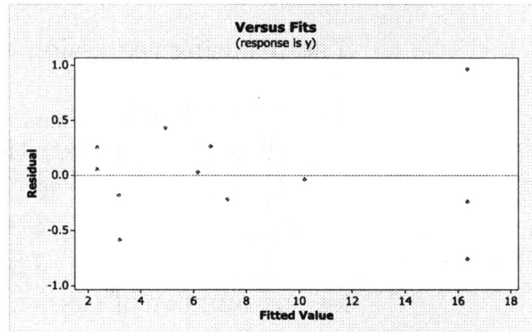
c. $F = .46$ with $p = .73$ which indicates there is no lack of fit.

d. $F = 2.21$ which is not significant and indicates that the interaction term does not contribute significantly to the model.

e. The quadratic term for x_2 contributes significantly to the model while the quadratic term for x_1 does not.

7.7 Observation 7 is influential which affects the plots. Normality looks pretty good and the residual plot is ok.





7.8 a. $\widehat{y} = 3.535 + .360P_1(x) + .187P_2(x)$.

b. $SS_R(\alpha_1, \alpha_2) = .360(118.71) + .187(24.66) = 47.31$. The linear and quadratic terms account for all of the variation in the data. Thus, the cubic term is not necessary.

7.9 a. To test $H_0 : \beta_{10} = \beta_{11} = \beta_{12} = 0$ use $F = \dfrac{SS_R(\beta_{10}, \beta_{11}, \beta_{12}|\beta_{00}, \beta_{01}, \beta_{02})/3}{MS_E}$.

b. Delete the term $\beta_{10}(x - t)^0$.

c. Also, delete the term $\beta_{11}(x - t)^1$.

7.10 A complete second-order model was fit to the delivery time data in Example 3.1. The analysis was done on centered data. Insignificant regressors were removed from the model.

The resulting regression equation is $\hat{y} = 21.1 + 1.26 * (x_{num} - 8.76) + 0.0136 * (x_{dist} - 409.28) + 0.0306 * (x_{num} - 8.76)^2$.

Coefficient	test statistic	p-value
β_{num}	6.70	0.000
β_{dist}	4.36	0.000
β^2_{num}	2.98	0.007

The regression indicates that the quadratic term for the number of cases of product stocked improves the model.

7.11 A complete second order model was fit to the patient satisfaction data where the data have been centered.

The regression equation is $\hat{y} = 69.1 - 1.029 * (x_{age} - 50.84) - 0.422 * (x_{sev} - 45.92) + 0.0031 * (x_{age} - 50.84) * (x_{sev} - 45.92) - 0.0065 * (x_{age} - 50.84)^2 - 0.0082 * (x_{sev} - 45.92)$.

Coefficient	test statistic	p-value
β_{age}	-5.54	0.000
β_{sev}	-1.95	0.067
$\beta_{age*sev}$	0.14	0.892
β^2_{age}	-0.56	0.584
β^2_{sev}	-0.44	0.663

There is no indication that it is necessary to add these second-order terms to the model.

7.12 a. Change the ranges to $x \leq t_1$, $t_1 < x \leq t_2$, and $x > t_2$.

b. Delete the terms $\beta_{10}(x - t_1)^0$ and $\beta_{20}(x - t_2)^0$.

c. Also, delete the terms $\beta_{11}(x - t_1)^1$ and $\beta_{21}(x - t_2)^1$.

7.13 $\widehat{y} = 15.1 - .0502x + .0389(x - 200)^1$. Test $H_0 : \beta_{11} = 0$, which gives a $t = 6.53$ and $p = 0.000$. The data do support the fit of this model.

7.14 $\widehat{y} = 15.298 - .0516x + .325(x - 200)^0 + .0373(x - 200)^1$. Test $H_0 : \beta_{10} = 0$, which gives a $t = 0.79$ and $p = 0.456$. There is no change in the intercept but a change in the slope.

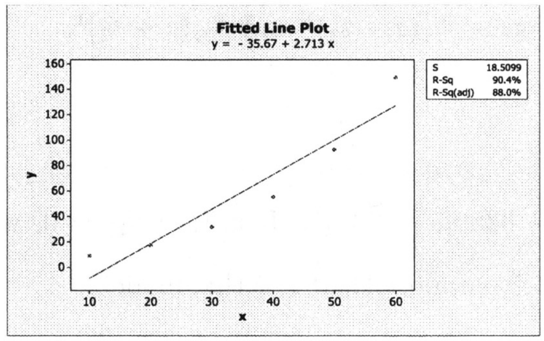
7.15 The variance inflation factors are 4.9 which do not indicate a multicollinearity problem.

7.16 a. The variance inflation factors are 19.9 which indicates there is a multicollinearity problem.

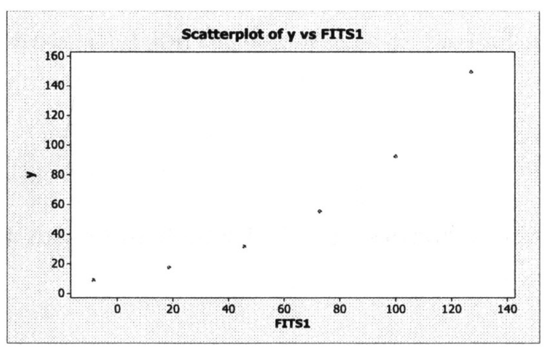
b. The variance inflation factors are 1.0 which indicates there is not a multicollinearity problem.

c. Many times centering can remove the multicollinearity problem.

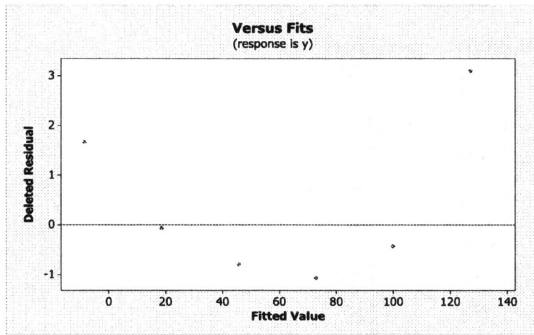
7.17 a. The data are nonlinear.



b. This also shows the data is nonlinear.

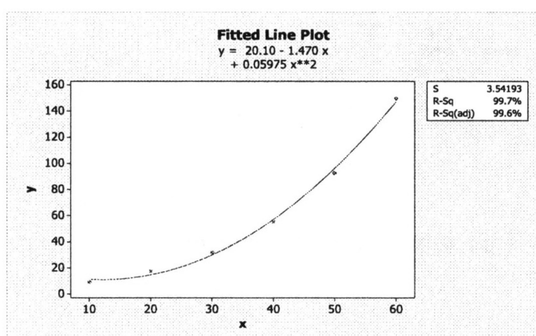


c. There is a quadratic pattern.



d. $\widehat{y} = 20.1 - 1.47x + .059x^2$. The test on the quadratic term is $F = \frac{1332.8}{12.5} = 106.62$ which is significant.

e. Yes, the second order model fits better.

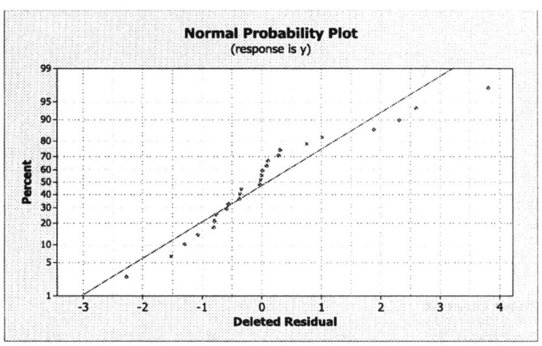


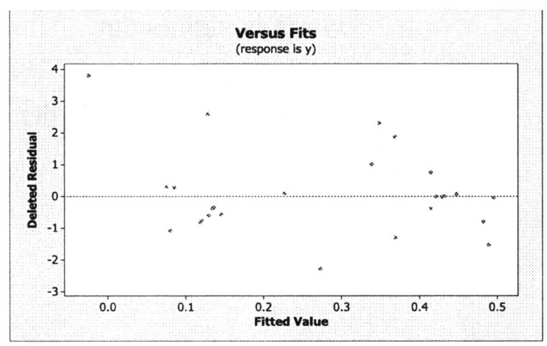
7.18 a. $\widehat{y} = -1.77 + .421x_1 + .222x_2 - .128x_3 - .0193x_1^2 + .007x_2^2 + .0008x_3^2 - .019x_1x_2 + .009x_1x_3 + .003x_2x_3$.

b. $F = 19.63$ with $p = 0.000$ which is significant. All are non-significant.

Coefficient	test statistic	p-value
β_1	1.43	0.172
β_2	1.70	0.108
β_3	−1.82	0.087
β_{11}	−1.15	.267
β_{22}	−.62	.545
β_{33}	.57	.575
β_{12}	−1.63	.118
β_{13}	1.20	.247
β_{23}	.37	.719

c. There are several outliers which affect normality and the residual plot.





d. $F = \dfrac{.035908/6}{.003712} = 1.61$ which is not significant.

7.19 The variance inflation factors are all very large indicating there is a serious problem with multicollinearity.