STA 102 Spring 2002
Chapter 18 Solutions to Suggested Even Problems

2,4, 6 not shown

8.
a) If the four data points corresponding to infants whose gestational age is 38 weeks or greater are removed, CBVR no longer appears to increase in magnitude as gestational age increases.
b) This information would add further credibility to the statement that there is no relationship between CBVR and gestational age.

10.
a) Scatter plot not shown.
b) The estimated least squares regression line is y_hat = -760.524+ 51.239x. Due to the influence of the point corresponding to an infant with an exceptionally small length and a large weight, the two lines probably differ.
c) After the outlier has been removed, the new least squares regression line is y_hat = -1864.27+79.889x. Removing the outlier causes the line to become steeper; the slope increases and the intercept decreases.
d) The value of R^2 increases from 41.78% to 75.79%. After removal of the outlier a greater amount of the variation in weight is explained by its linear relationship with length.
The value of $s_{y|x}$ decreases from 199.86 to 130.85. The standard deviation from regression is a measure of the variability of the data points around the line. The outlier was relatively far from the fitted line and therefore had a large residual; when it is eliminated, $s_{y|x}$ decreases.

12.
a) Scatter plot not shown. It appears that fatality rate at first decreases as year increases, but, after a number of years, begins to increase again.
b) The estimated least squares regression line is y_hat = 0.181-0.0102x. The fit of the model seems to be less than adequate. About 55.9% of the variation in fatality rate is explained by its linear relationship with calendar year. However, the plot of the residuals versus the fitted values displays a U-shaped trend that is far from random. This suggests that transformation of either response variable or the explanatory variable is necessary.
d) The estimated least squares regression line is y_hat = 0.214-0.059 ln(x). While the increase in R^2 from 55.9% to 84% indicates that a greater proportion of the variability is explained by the new model, the residual plot still displays a U-shaped curve.
e) Scatter plot not shown.
f) The estimated least squares regression line is y_hat = 0.068 + 0.181(1/x). The increase in R^2 from 84% to 94.3% again signifies an improvement in fit relative to the previous models. In addition the residual plot no longer has a U-shaped pattern.
g) The model containing the reciprocal of year appears to fit the data best. It has the largest coefficient of determination and the smallest SD. In addition the residual plot displays a more random scatter than either of the other plots. However, there are a couple of outliers.