HW 13 [PG] : Due Friday December 5, 2003

Chapter 18, Pages 443-447: Take a look at and answer Exercises 1 to 7. But you do not need turn in these problems.

Turn in Exercises 8, 9 and 12

Ch 18, 8: a. If the 4 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 CBRV and gestational age.

Ch 18, 9: a. Scatter plot.

b. The estimated regression line is

$$\hat{y} = 10.552 + 1.264x.$$

The slope of the line implies that each one-week increase in gestational age causes an infant's systolic blood pressure to increase by 1.264 mm Hg on average. Although it does not make sense in this case, the y-intercept of 10.552 is the predicted systolic blood pressure for an infant whose gestational age is 0.

c. To test the null hypothesis

$$H_0 : \beta = 0,$$

we note that $t = 2.898$ and $p = 0.005$. We reject H_0 at the 0.05 level of significance. We conclude that for low birth weight infants, systolic blood pressure increases as gestational age increases.

d. For the population of infants whose gestational age is 37 weeks, the estimated mean systolic blood pressure is

$$\hat{y} = 10.552 + 1.264 \times 31 = 49.748 mmHg.$$

e. To construct a 95% confidence interval, we first find the standard error of the predicted mean $\hat{y}$. In R, we find that $se(\hat{y}) = 1.434$. Thus a 95% confidence interval for the mean value of systolic blood pressure of those whose gestational age is 31 is

$$(49.748 - 1.98 \times 1.434, 49.748 + 1.98 \times 1.434) = (46.91, 52.59).$$

f. For a randomly selected new infant whose gestational age is 31 weeks, the predicted systolic blood pressure is

$$\tilde{y} = 49.748 mmHg.$$

g. To construct a prediction interval, we first find the standard error of $\hat{y}$. It is $se(\tilde{y}) =$ 11.093 mm Hg. Therefore, a 95% prediction interval for the individual value of systolic blood pressure is

$$(49.748 - 1.98 \times 11.093, 49.748 + 1.98 \times 11.093) = (27.71,\ 71.71).$$

Note that this interval is much wider than the confidence interval for the mean of systolic blood pressure.

h. The regression model does not provide a very good fit to the data. Although gestational age does help in predicting systolic blood pressure, only $R^2 = 7.9\%$ of the variability in blood pressure is explained by the linear relationship.

The plot of residuals versus the fitted values of systolic blood pressure shows no evidence that the assumption of homoscedasticity has been violated or that a transformation of either variable is necessary. However, there is one data point that seems to be an outlier.

Ch 18, 12: a. The scatter plot of fatality rate vs. year suggests that fatality rate at first decreases as year increases, but after a number of years began to increase again.

b. The estimated regression line is

$$\hat{y} = 0.181 - 0.010x.$$

The fit of the model appears to be no adequate. Te plot of residuals vs. the fitted values iF rate below displays a Uchaped trend that is far from random. This suggests that a transformation of either the response variable or the explanatory variable is necessary.

c. This is a scatter plot of fatality rate vs. natural logrithm of year.

d. The estimated regression line is

$$\hat{y} = 0.214 - 0.058ln(x).$$

While the R^2 increases from 55.9% to 84.0%, the residual plot still displays a U-shaped trend.

e. The scatter plot of fatality rate versus the reciprocal of year.

f. The estimated regression line is

$$\hat{y} = 0.068 + 0.181(1/x).$$

The increase in R^2 from 84.0% to 94.3% signifies an improvement in fit relative to the previous model. In addition, the residual plot no longer has a U-shaped pattern.

[Note: I have also gone over this problem in class. Please see

`www.stat.uiowa.edu/~jian/S101/ComputerLab/reg.txt`

]

g. The model containing the reciprocal of year appears to fit the data best. It has the largest R^2 and the smallest standard deviation from the regression. In addition, its residual plot displays a more random scatter than either of the other plots.