Chapter 8: Linear Regression

A linear model is often used to describe the relationship between two variables.

Residuals are the differences between data values and the corresponding values predicted by the regression model.

Residual = observed value - predicted value = y - y

The line of best fit is the line for which the sum of the squared residuals is smallest.

The units of slope for this line are always the units of y per unit of x.

A linear model's equation has the form $y = b_0 + b_1x$.

b_0 is the y-intercept and b_1 is the slope

b_1 can be calculated using the equation $b_1 = rSy / Sx$

b_0 can be calculated by solving the equation $b_0 = y - b_1x$, plugging in the mean-mean point for x and y.

The regression equation should be written using the actual names of the y and x variables.

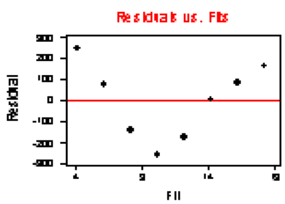
Linear models apply only when certain conditions are true.

Quantitative Variables Condition: Correlation applies only to quantitative variables, check the variables' units and what they measure.

Straight Enough Condition: The linear model assumes that the relationship between variables is linear, check to see if the scatter plot looks reasonably straight.

Outlier Condition: Outliers could change the regression as they may have large residuals, check for them to make sure no point should be singled out for special attention.

The plot of the residuals should be checked for any patterns. If there is a pattern, as in the residual plot below, the linear model is not appropriate.



R^2 is the square of the correlation between y and x. It gives the fraction of the *variability of y accounted for by the least squares linear regression on x, the model.* It measures how successful the regression is in linearly relating y to x.

Sample Problem:

Given that x =10, Sx= 2, y = 20, Sy= 3, and r = 0.5, determine the linear model equation.

$b_1 = rSy / Sx = (0.5)(3) / (2) = 0.75$ $b_0 = y - b_1x = 20 - [(.75)(10)] = 12.5$

$y = 12.5 + 0.75x$