Computer Lab Session #6

We will work again with the regressions considered in Lab #5, and refer back to what we have observed for their residuals.

**BEARS.MTW:**

Let’s consider a *quadratic regression* for y=Weight on x=Length (much more on this later). Use Calc > Calculator to produce w=Length$^2$, and then use again Stat > Regression > Regression (upload both x and w in the list of predictors). With Graphs, produce the plot of regular residuals vs fits, and again the normal probability plot for regular residuals.

- Why are we using the plot of residuals vs fits (instead of x=Length) here?
- What can you notice in the plot of residuals vs fits? Is the curvature gone? Is there evidence that the error variance may not be constant, even when considering a quadratic model?
- What can you notice in the normal probability plot? Is there evidence that the deviations from the regression parabola may not behave like Gaussian errors? Are they closer to normality than the deviations from the regression line?

To visualize the fitted quadratic regression, use Stat > Regression > Fitted line plot, and the “Quadratic” option. Is the $R^2$ for the quadratic regression higher than the one obtained with the simple linear regression? Using Stat > Basic Statistics > Correlation, compute the correlation between y=Weight and x=Length. Verify that the $R^2$ for the simple linear regression is the square of the correlation coefficient, and the $R^2$ for the quadratic regression is not.

**SALARY.MTW:**

The output for the regression of y=1991$ on x=Begin$ provides *Unusual Observations*. Identify the observation with the largest standardized residual, and use Data >Subset worksheet (“Specify row number” option) to remove this observation from the data set. Then run the regression again.

- Did the regression fit change in a substantial way? Look at $R^2$, and estimated slope and intercept.

A substantial change in the estimated regression parameters would make this outlier *influential* (much more about this later).
MASSCOLL.MTW:

For the regression of y=“%WhoGrad” on x=CSAT, besides the plot of regular residuals vs x=CSAT, you can use again Stat > Regression > Regression > Graphs to produce the plot of regular residuals vs other variables in the data set. These are variables that you may want to consider in addition to x=CSAT as predictors in a more complex regression model. The issue here is using the residuals of the regression of y=“%WhoGrad” on x=CSAT to see if any other variable may “capture something” of y that has been left out by x. If an additional variable, say w, captures something that has been left out by x, the plot of residuals vs w ought to show a “trace” of this in the form of a systematic pattern.

- Can you find any variables that might constitute useful additional predictors? How about w=SFratio?

NEWMARKET.MTW:

For the regression of y=Sales on x=Advertis, besides the plot of regular residuals vs x=Advertis, you can use again Stat > Regression > Regression > Graphs to produce the plot of regular residuals vs Index (this variable just records the order of the 10 consecutive months).

- Is there any evident mean pattern in the residuals vs the order expressed by Index?

You can think of the natural order among the observed units (months) as a special kind of “additional predictor”. The existence of a pattern in the plot of residuals vs this order indicates that the errors corresponding to various units are not independent – in other words, the order captures something about y.