An Algorithm for Analyzing Regression Data
Ronald Christensen

To analyze regression data, I suggest you follow this algorithm. I do not claim it is perfect, but it at least covers the highlights.

When finished, you should write up a three page, typed, double spaced, summary of your procedures and conclusions. You should include an appendix, but do not expect the appendix to be read. A well organized appendix is much more likely to get read.

When analyzing regression data:

1. Plot the data, describe the data.

    (a) Scatter plot matrix.
    (b) Do any variables look like good candidates for transformations? (Check max/min.)
    (c) Do any variables look like factors?

2. Decide on initial variables and fit the full model.

    (a) In constructing the full model, account for any factor variables.
    (b) Check assumptions.
    (c) Check for outliers.
    (d) Do any variables look like good candidates for transformations? (Check Box-Cox transformations on both dependent and predictor variables.)

3. If necessary, transform the dependent variable or change the model (which may include transforming the predictors or dropping cases).

4. Repeat until full model seems adequate.

5. Check for collinearity.

    (a) If collinearity is a problem, do principal component regression.
    (b) If not, do variable selection.

6. Choose new candidate models, fit all models, evaluate assumptions, decide on final models.

7. Interpret final models in clear English.