Computer Lab Session #10

Load the file BEARS.MTW from the Minitab example data sets.

1. Fit a regression model for \( y = \text{Weight} \) as a linear form in \( x = \text{Age} \), which can differ (both in intercept and in slope) by Sex. Note: Sex is coded as 1 = Male, 2 = Female. Before you fit the model, you need to create a new column in which Sex is coded in a binary fashion (e.g., 0 for Male and 1 for Female). You can do this using Data > Code > Numeric to Numeric, or Calc > Calculator.

- Do the intercept and/or the slope for Weight as a function of Age differ significantly between Male and Female bears? Identify the appropriate tests, and interpret.
- Use Graph > Scatter Plot > With Groups to create a plot of Weight vs Age with separate regression lines and lowess smooths for males and females.

2. Suppose we are comparing two models expressing Weight as a function of various predictors and terms available from the data. If the two models are nested, i.e., if one (the smaller) can be obtained from the other (the larger) dropping some terms, we can use the General Linear Test (Chapter 2, pg 72-73) to assess whether \( H_0: \) the smaller (REDUCED) model is enough, vs \( H_a: \) it is not, and the larger (FULL) model is needed. The test is based on an F-ratio; namely

\[
F = \frac{(SSE(\text{Red}) - SSE(\text{Full}))/df(\text{Red}) - df(\text{Full}))}{SSE(\text{Full})/df(\text{Full})} 
\]

You can fit different models to obtain the sums of squares required to compute this type of F ratios, and use Calc > Probability Distribution > F to obtain p-values. Use this approach to compare:

\[
\text{Red}: \text{Weight}_i = \beta_0 + \beta_1\text{Age}_i + \varepsilon_i \\
\text{Full}: \text{Weight}_i = \beta_0 + \beta_0\text{Sex}_i + \beta_1\text{Age}_i + \beta_1\text{Sex}_i\cdot\text{Age}_i + \varepsilon_i
\]

\[
\text{Red}: \text{Weight}_i = \beta_0 + \beta_1\text{Age}_i + \varepsilon_i \\
\text{Full}: \text{Weight}_i = \beta_0 + \beta_1\text{Age}_i + \beta_2\text{Length}_i + \beta_3\text{Chest.G}_i + \varepsilon_i
\]

3. We want to select a good model for Weight as a function of the six quantitative predictors Age, Head.L, Head.W, Neck.G, Length, Chest.G, considered linearly, or possibly with various powers, other transformations and/or interaction terms in them. Form terms as needed using Calc > Calculator, and then Stat > Regression > Best Subsets to explore...
possible models. As discussed in class, once you have identified a few interesting candidate models with this automated procedure, fit them individually, and evaluate them in terms of diagnostics. Also, check interdependencies among the terms used in your models using Graph > Matrix Plot and Stat > Basic Statistics > Correlation to produce scatter plots and correlation coefficients, respectively.

Save your work here, you will use your results again in Lab session #11.

4. Take the model you like the best from 3. How parsimonious and interpretable is it? Produce confidence intervals for the mean response and predictions intervals for this model, using “Options” in Regression > Regression. Note: now you are required to put in not just ONE value (the value of x at which to build the intervals), but one value for each of the terms in your model – a string of values, one for each term, is what defines the point at which you are making the inferences. If you make inferences at several points, and want to correct the intervals, you can do so, for instance using the Bonferroni procedure, exactly as before.