## Assignment #2 — Due Friday, February 9, 2007, by 4:00 P.M.

Turn in homework in lecture, discussion, or your TA's mailbox. Please indicate the discussion section you expect to attend to pick up this assignment.

**311:** Tues. 1:00–2:15        **312:** Wed. 2:30–3:45        **313:** Wed. 1:00–2:15

Many problems on this assignment require using the computer. Your turned in solutions should not include all of the computer output and graphs that you will produce. Write your solutions and include only sparingly computer output or graphs when necessary to support a point you are making in response to the problem question.

If a problem asks for a graph, provide it. If the problem asks for you to comment about a graph, you do not need to include the graph in your solution.

1. The file `fevdata.txt` is on the course web page. Download it. Save it as plain text, not as a Word document.

   The first line contains the variable names which are `age`, `fev`, `ht`, `sex`, and `smoke`. For this assignment, we will consider only age (age of children, measured in years) and FEV (forced expiratory volume, a measure of lung capacity, measured in liters).

   Load the data into R using the `read.table` command. It is advisable to change your working directory to be the same place as where the data file is saved. On both Macs and Windows, a command like this opens a dialog box allowing you to browse for the file.

   ```
   > fevdata = read.table(file.choose(),header=TRUE)
   ```

   It is a good idea to examine the structure of the data you read in to make sure it is correct.

   ```
   > str(fevdata)
   ```

   This should indicate that `fevdata` is a data frame with 654 observations and 5 variables. The variable `age` is integer valued, `fev` and `ht` are numeric, and `sex` and `smoke` are factors, each with two levels.

   If you attach the data frame, you can refer to variables directly.

   ```
   > attach(fevdata)
   ```

   Use R to fit a linear model to estimate FEV from age. Summarize the fit model. Make one or more residual plots to examine the quality of the fit. State if the fit model is acceptable and briefly explain why or why not.

2. The growth of the mold *Aspergillus niger* is inhibited by the toxicant methyl 4-hydroxybenzoate. Various concentrations of the toxicant are available. The inhibitions are measured as growth rate changes (in mm) per 24 hours.

   | conc | 0.32 | 0.40 | 0.51 | 0.64 | 0.80 | 1.0 | 1.3 | 1.6 | 2.0 | 2.5 | 3.2 | 4.0 |
   |---|---|---|---|---|---|---|---|---|---|---|---|---|
   | inhibit | 2.0 | 2.3 | 2.4 | 2.8 | 3.0 | 3.3 | 3.6 | 4.3 | 4.8 | 5.1 | 5.6 | 6.6 |

   (a) Fit linear models for both `inhibit` versus `conc` and for the log transforms of both variables. Examine residual plots and decide which model is more appropriate. Briefly explain.

   The following R command is what you could use to fit the log-transformed data.

   ```
   > logfit.lm = lm( log(inhibit) ~ log(conc) )
   ```

(b) Find a 95% confidence interval for the mean inhibition at a concentration of 2.8 (in the unstated units). (If you are using the transformed data, find the end points of the log-transformed response and back-transform using the `exp` function.) Interpret the interval in the context of the problem.

(c) Find 95% prediction intervals for the inhibition when the concentration of the toxicant is 3.3 and 1.3. Explain briefly why the prediction intervals are different widths.

3. A forester is interested in relating the height of a tree (in feet), `Height`, to the stem diameter at breast height (in inches) `DBH`. Observations on height and stem diameter were recorded for 198 48-year-old trees.

The data are in the file `tree.txt` at the course web site. The first column consists of a plot id number, the second column is the tree number, the third column is height, and the fourth column is diameter at breast height (in inches).

(a) Does a simple linear regression model fit well in estimating `Height` from `DBH`? Explain why.

(b) Consider fitting a linear model for the subset of trees for which the diameter at breast height exceeds 15 inches.

```
> trees = read.table(file.choose(),header=TRUE)
> treeBig15.lm = lm(Height ~ DBH, subset = (DBH > 15), data = trees)
```

(c) Perform checks of residual plots to assess the model.

(d) Report 95% confidence intervals for the model slope and intercept.

(e) Use the regression model to predict heights of trees with DBH of measures of 5.5 and 7.5 inches with a 95% prediction intervals.

(f) Find the heights of the trees in the data with DBH of 5.5 and 7.5. Do the heights fall into the prediction intervals? Briefly summarize your observations and discuss the reasoning behind what you find.

**Work to do, but not turn in.**

• If you have not already done so, acquire the textbook.

• Work through the R examples in sections 1.2–1.4, pages 8–18.

• Revisit Chapter 5 of the textbook.