Assignment #3 — Due Friday, February 16, 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. Problems 1 and 2 reconsider the data in the file `fevdata.txt` on the course web page. The first line contains the variable names which are `age`, `fev`, `ht`, `sex`, and `smoke`. In this assignment, we will consider models that use `age` (age of children, measured in years), `ht` (height of children in inches), and `fev` (forced expiratory volume, a measure of lung capacity, measured in liters).

   (a) Fit a linear model to predict FEV from age. Use the R function `boxcox` to find a simple power transformation (-1 = reciprocal, 0 = log, 0.5 = square root) close to the Box-Cox maximum likelihood estimate. Which simple transformation seems best?

   (b) Fit a linear model with the best simple transformed response predicted by `age` and examine the residual plot of the fit. Has the transformation improved adherence to the constant variance assumption? Is this linear model acceptable? Briefly explain why or why not.

   (c) Repeat parts (a) and (b) using `log(age)` as the explanatory variable. Which set of transformations of `fev` and `age` seems to be best to match linear model assumptions?

2. Add a second explanatory variable, `ht`, or a transformation of it, to the model found in Problem 1c. Write an equation to express the model. Find 95% confidence intervals for each model parameter (one parameter and two slopes) in the (possibly) transformed scale. As best as you can, interpret these confidence intervals in the scale of the initial measurements. (Note: There is no single set of transformations that is unambiguously best. Use your judgment!)

3. In a study of genetic variation in sugar maple, seeds were collected from native trees in the eastern United States and Canada and planted in a nursery in Wooster, Ohio. The time of leafing out of these seedlings can be related to the latitude and mean July temperature of the place of origin of the seed. The variables are X_1 = latitude, X_2 = July mean temperature, and Y = weighted mean index of leafing out time. (Y is a measure of the degree to which the leafing out process has occurred. A high value is indicative that the leafing out process is well advanced.) The data is below and in the file `maple.txt` on the course web page.

   For the following problem, check assumptions and use transformations if warranted.

   (a) Find the regression of `LeafIndex` on `Latitude`. Is latitude a useful predictor of leaf index?

   (b) Repeat part (a) for the regression of `LeafIndex` on `JulyTemp`.

   (c) Find the regression of `LeafIndex` on `Latitude` and `JulyTemp`. Compare the results of this analysis with your results from (a) and (b). How different are the slope coefficients in each case. What best explains the differences in their values?

   (d) Find ANOVA tables for the model in part (a) (`LeafIndex ~ Latitude`) and the model in part (c) (`LeafIndex ~ Latitude + JulyTemp`) using the function `anova` on the object created using `lm`. What parts of the row of the ANOVA table corresponding to `Latitude` are the same and what parts are different? To what formal hypothesis test does the p-value in the `Latitude` row of each ANOVA table correspond? Why are the p-values different?

Location	X_1	X_2	Y	Location	X_1	X_2	Y
Quebec	46.95	66.7	20.5	Ohio	40.33	73.0	16.6
Nova Scotia	45.54	64.3	21.0	Ohio	41.58	71.4	18.2
Nova Scotia	45.00	61.8	18.3	Penn.	41.92	69.7	18.3
Maine	45.42	65.0	18.6	Penn.	40.12	72.6	13.3
Vermont	44.50	63.5	18.6	Mass.	42.75	69.2	20.0
Michigan	46.37	65.5	19.8	Conn.	42.00	68.0	18.1
Michigan	45.03	67.7	20.6	Mass.	42.55	70.0	17.0
Ontario	43.00	67.9	19.2	Conn.	41.50	72.5	15.6
New York	42.80	70.2	19.8	Indiana	39.25	76.7	15.7
Michigan	42.70	71.1	15.8	Illinois	37.45	79.5	15.4
Iowa	42.42	74.4	14.5	Virginia	38.33	69.5	15.4
Iowa	40.75	76.4	18.5	Tennessee	36.25	74.3	15.8
Illinois	41.38	75.9	16.7	No. Car.	35.80	68.9	15.0
Illinois	40.35	76.4	13.9	Miss.	31.38	80.8	15.9
Ohio	40.80	71.9	14.3	Georgia	32.83	81.2	14.0
Ohio	40.92	74.2	15.2	Florida	30.75	81.0	10.8

Work to do, but not turn in.

- Work through the R examples in sections 1.5, pages 18–24.
- Read Chapter 6 of the textbook.