Applied Statistics With R
J. Fox
WU Wien
May/June 2006

Part 2a: Fitting Linear Models

1. The data given in the data frame `Burt` in the `car` package, on the IQs of 27 pairs of identical twins reared apart, were reported by Sir Cyril Burt (1966). (These "data" are wholly fraudulent.) One twin in each pair was raised by his or her biological parents; the other twin was raised in a foster home. In each case, Burt recorded (i.e., made up) the "social class" to which the twins' biological parents belonged.
    a. Explore the data graphically by plotting `IQbio` (as the response variable) against `IQfoster`, using a different symbol and plotting a separate linear regression line for each social class. (You may use the `scatterplot` function in the `car` package.)
    b. Then regress the IQ of the twins reared by their biological parents (`IQbio`) on the IQ of the twins reared by foster parents (`IQfoster`), dummy variables to represent the three social classes (`class`), and regressors for the interaction between foster-twin IQ and social class. [*Suggestion*: You may want to re-order the categories of the factor `class` so that they are in their natural order rather than in the (default) alphabetic order.]
    c. Test the interaction between foster-twin IQ and social class. If the interaction proves to be non-significant, test the partial effects of foster-twin IQ and social class on biological-twin IQ. Compute the appropriate incremental F-tests using the `Anova` function in the `car` package.
    d. Based on statistical considerations alone, how can you tell from the data analysis that you performed that Burt's data are almost surely fraudulent?

2. The data in the `Adler` data frame in the `car` package are from a social-psychological experiment, reported by Adler (1973), on "experimenter effects" in social research – that is, how researchers' expectations can influence the data that they collect. Adler recruited research assistants, who showed photographs of individuals' faces to respondents; the respondents were asked by the research assistants to rate the apparent "successfulness" of the individuals in the photographs. In fact, Adler chose photographs that were average in their appearance of success, and the true subjects in the study were the research assistants.

    Adler manipulated two factors, named expectation and instruction in the data set:
    - `expectation`. Some assistants were told to expect high ratings, while others were told to expect low ratings.
    - `instruction`. In addition, the assistants were given different instructions about how to collect data: Some were instructed to try to collect "good" data; others were instructed to try to collect "scientific" data; and still others were given no special instruction of this type.

Adler randomly assigned 18 respondents to each of the six resulting experimental conditions – combinations of the two categories of the factor `expectation` (`high` or `low`) and the three categories of the factor `instruction` (`good`, `scientific`, or `none`). I deleted some of these observations at random to produce an unbalanced dataset.

a. Treating `instruction` as the row factor and `expectation` as the column factor, calculate the mean and standard deviation of scores in each cell. Graph the cell means and comment on the apparent results of the experiment. (Feel free to use whatever software you wish to draw the graph, or to draw it by hand.)
b. Construct a two-way analysis-of-variance table for Adler's data, testing hypotheses about the main effects and interactions of instruction and expectation. What conclusions would you draw from this ANOVA?