Regression Analysis

We begin by hypothesizing a relationship:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \, .$$

Using relevant sample data, we estimate the coefficients:

$$Y_{pred} = a + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k \, (+ 0) \, .$$

The six "steps" to interpreting the results of a regression analysis are:

1. Predict the dependent variable (using the appropriate model) for individuals, and determine the margin of error for a prediction by referring to the standard error of the prediction.

2. Estimate the mean of the dependent variable across a large pool of similar individuals (using the appropriate model), and determine the margin of error in the estimated mean by referring to the standard error of the estimated mean.

3. Estimate the pure "effect" of an explanatory variable on the dependent variable (using the most complete model available), and determine the margin of error in that estimate by referring to the standard error of the coefficient of that variable.

4. Measure the potential "explanatory power" of a model (i.e., the extent to which variation in the dependent variable across the population might be explained by the fact that the explanatory variables vary across the population) by referring to the adjusted coefficient of determination.

5. Compare the absolute values of the beta-weights of the explanatory variables in order to rank them in order of "relative explanatory importance" (i.e., the extent to which variation in each helps to explain observed variation in the dependent variable) within the current model.

6. Look at the significance levels for the explanatory variables to see how strongly the evidence supports including each of the explanatory variables in the model (i.e., to see whether knowledge of this variable is useful when making individual predictions [because the variable has a non-zero coefficient]).