Comments on Variable Naming and Coding
Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised February 15, 2015

1.   Do not confuse the name of a variable with the possible values that the variable can assume. Remember, in order for something to be a variable, it must be able to assume at least two different values. For example, a variable called FEMALE might be coded 1 if the respondent is female, 0 if the respondent is male.

2.   Sometimes dichotomous variables have names like FEMALE, BLACK, or HEAD START. Typically, these variables are coded 1 if the respondent has the trait implied by the variable's name, 0 otherwise. Hence, for FEMALE, 1 = female, 0 = male; for BLACK, 1 = black, 0 = nonblack; for HEAD START, 1 = participates in head start, 0 = does not participate.

3.   Sometimes dichotomous variables have names like GENDER or RACE. In such instances, you need to be sure to check how the variable is coded. For example, GENDER might be coded 1 = male, 0 = female. Another common coding is 1 = male, 2 = female.

4.   In the absence of information to the contrary, you can usually assume that continuous variables are coded from low to high. Hence, for INCOME, a low value on the variable means low income, a high value means high income. For POVERTY, a high value means very poor, a low value means very rich. For IQ, a high value means very smart, a low value means very dumb.

5.   While the above conventions are fairly common, any particular researcher might do things a bit differently. When reading other people's research, you should always make sure you understand how variables are coded.

6.   Suppose RACE is coded 1 = white, 0 = nonwhite. A positive correlation between RACE and INCOME would mean that whites tend to have higher incomes than nonwhites do.