

## **Data Set: Vocabulary Data from the U.S. General Social Surveys**

*Data File:* Vocabulary-2.txt

*Source:* GSS Cumulative Datafile 1972-2004, downloaded from <http://sda.berkeley.edu/archive.htm>.

*Variables:*

id: Respondent ID number.

year: Year of survey.

education: Level of education, <HS, HS, JC (junior college degree), Bach (bachelor's degree), Grad (graduate degree).

sex: Male or Female.

nativeBorn: Yes or No.

vocabulary: Vocabulary-test score, number of questions right out of 10.

age: in years.

agegroup: 2 (20s) through 6 (60 or more)

*Note:* This data file, with more variables but fewer observations than Vocabulary.txt, is used in Section 8.3.2 of the text.