

Information for Add Health Data

Aaron Gullickson

3/9/2017

The data we are using comes from the National Longitudinal Study of Adolescent to Adult Health (Add Health), conducted by the Carolina Population Center at UNC-Chapel Hill and supported by a grant from the National Institute of Child Health and Human Development. The first wave of the study surveyed adolescents between 7th and 12th grade in school in the 1994-95 school year. These respondents were then re-interviewed in three later waves, and there is currently a fifth wave of data being collected right now. This re-interviewing allows researchers to follow these respondents over time as they have grown from adolescents into full adulthood. This is what makes the study *longitudinal*.

We will only be using the publicly available Wave I data from when these respondents were adolescents in 1994-95. One of the particularly valuable features of the Add Health survey is that many respondents were in the “saturation sample” which sampled *all* students at 16 schools. In this saturation sample, students were asked about who were their friends and sexual partners, which allows researchers to construct network maps of adolescent social systems.

We will use a various basic measure of that network that estimates students’ popularity. This measure, which is called “in degree” in the network analysis literature, measures the number of times a student was nominated as a friend by other students in the school. We will treat it as a simple proxy measure of a student’s popularity. We can then look at what other student characteristics were positively or negatively associated with a student’s popularity.

From this full public dataset, I have created an extract with just a few variables for our analysis. Here is a full description of all variables in the dataset that we will use.

- **indegree**: The number of friend nominations received by other students at the same school. This is the measure of popularity that we will use.
- **race**: A six-category nominal variable indicating the race that the student best thought described them when asked to choose a single race: white, black, Latino, Asian, American Indian, other.
- **sex**: Add Health reports this as a student’s “biological” sex. Students were only reported as male or female.
- **grade**: current grade of the student as a quantitative variable.
- **psuedoGPA**: Students were asked for the most recent letter grade in four course types: math, language arts, science, and math. This variable was constructed by calculating GPA from those responses.
- **honorsociety**: A true/false variable for whether a student was in honor society or not.
- **alcoholuse**: A true/false variable that is true if the student reported drinking at least once or twice a month in the last twelve months.
- **smoker**: A true/false variable that is true if student smoked more than 5 cigarettes in the past 30 days.
- **bandchoir**: a true/false variable that is true if the student was in band or choir.
- **academicclub**: a true/false variable that was true if the student was in an academically-oriented club such as math club, book club, etc.
- **nsports**: The number of different school sports a student reported participating in. Students who reported more than six sports were top-coded at the value of six.
- **parentinc**: Parent’s household income measured in \$1000’s of dollars.

In the original data, there are missing values for some observations on some of these variables. To simplify our analysis, I have used some advanced techniques that are beyond the scope of our class to impute these missing values.