

1. DESCRIPTION OF THE DATA SET

In view of its relevance for social policy, it is not surprising that analysis of the closely related topics of the determinants of educational attainment and the determinants of earnings has long been a major application of econometrics. Particularly sensitive issues are those relating to differences in educational attainment and earnings attributable to ethnicity, sex, and genetic endowment, to interactions in the effects of these factors, and to changes through time. The data sets described here will allow you to explore some of these issues using a subset of a major US data-base, the National Longitudinal Survey of Youth 1979– (NLSY79).

NLSY79 is a panel survey with repeated interviews of a nationally representative sample of young males and females aged 14 to 21 in 1979. From 1979 to 1994 the interviews took place annually. Since 1994 they have been conducted at two-year intervals. The core sample originally consisted of 3,003 males and 3,108 females. In addition there are special supplementary samples (some now discontinued) of ethnic minorities, those in poverty, and those serving in the armed forces. Extensive background information was obtained in the base-year survey in 1979 and since then information has been updated each year on education, training, employment, marital status, fertility, health, child care and assets and income. In addition special sections have been added from time to time on other topics – for example, drug use. The surveys have been extremely detailed and the quality of the execution of the survey is very high. As a consequence NLSY79 is regarded as one of the most important data bases available to social scientists working with US data.

For the practical work there are 20 parallel data subsets each consisting of 540 observations, 270 drawn randomly from the male respondents in the source data set and the same number drawn randomly from the female respondents. Each subset contains data for each respondent on the following variables (C indicates a continuous variable, D a dummy variable):

Personal variables

<i>FEMALE</i>	D	Sex of respondent (0 if male, 1 if female)
<i>MALE</i>	D	Sex of respondent (1 if male, 0 if female)
<i>Ethnicity:</i>		
<i>ETHBLACK</i>	D	Black
<i>ETHHISP</i>	D	Hispanic
<i>ETHWHITE</i>	D	Non-black, non-hispanic
<i>AGE</i>	C	Age in 2002
<i>S</i>	C	Years of schooling (highest grade completed as of 2002)
<i>Highest educational qualification:</i>		
<i>EDUCPROF</i>	D	Professional degree
<i>EDUCPHD</i>	D	Doctorate
<i>EDUCMAST</i>	D	Master's degree
<i>EDUCBA</i>	D	Bachelor's degree
<i>EDUCAA</i>	D	Associate's (two-year college) degree
<i>EDUCHSD</i>	D	High school diploma or equivalent
<i>EDUCDO</i>	D	High school drop-out

Marital status

<i>SINGLE</i>	D	Single, never married
<i>MARRIED</i>	D	Married, spouse present
<i>DIVORCED</i>	D	Divorced or separated

Scaled score on a component of the ASVAB battery (see Section 6 for further details on the ASVAB variables)

<i>ASVAB2</i>	C	Arithmetic reasoning
<i>ASVAB3</i>	C	Word knowledge
<i>ASVAB4</i>	C	Paragraph comprehension
<i>ASVAB5</i>	C	Numerical operations (speed test)
<i>ASVAB6</i>	C	Coding speed (speed test)
<i>ASVABC</i>	C	Composite of <i>ASVAB2</i> (with double weight), <i>ASVAB3</i> and <i>ASVAB4</i>

Faith:

<i>FAITHN</i>	D	None
<i>FAITHC</i>	D	Catholic
<i>FAITHJ</i>	D	Jewish
<i>FAITHP</i>	D	Protestant
<i>FAITHO</i>	D	Other
<i>HEIGHT</i>	C	Height, in inches, in 1985
<i>WEIGHT85</i>	C	Weight, in pounds, in 1985
<i>WEIGHT02</i>	C	Weight, in pounds, in 2002

Family background variables

<i>SM</i>	C	Years of schooling of respondent's mother
<i>SF</i>	C	Years of schooling of respondent's father
<i>SIBLINGS</i>	C	Number of siblings
<i>Living at age 14:</i>		
<i>L14TOWN</i>	D	in a town or city
<i>L14COUN</i>	D	in the country, not on a farm
<i>L14FARM</i>	D	on a farm
<i>LIBRARY</i>	D	Member of family possessed a library card when respondent was 14
<i>POV78</i>	D	Family living in poverty in 1978

Work-related variables

<i>EARNINGS</i>	C	Current hourly earnings in \$ reported at the 2002 interview
<i>HOURS</i>	C	Usual number of hours worked per week, 2002 interview
<i>TENURE</i>	C	Tenure (years) with current employer at the 2002 interview
<i>EXP</i>	C	Total out-of-school work experience (years) as of the 2002 interview.
<i>COLLBARG</i>	D	Pay set by collective bargaining, 2002

Category of employment

<i>CATGOV</i>	D	Government
<i>CATPRI</i>	D	Private sector
<i>CATSE</i>	D	Self-employment
<i>URBAN</i>	D	Living in an urban area at 2002 interview

Living in 2002 in:

<i>REGNC</i>	D	North central census region
<i>REGNE</i>	D	North eastern
<i>REGS</i>	D	Southern
<i>REGW</i>	D	Western

Not all of the variables are used in the exercises suggested in the next section. You should feel free to experiment by trying alternative regression specifications with the extra variables. Further details of some of the variables are given in Section 6.