

EDUCATIONAL ATTAINMENT AND EARNINGS FUNCTIONS

TABLE OF CONTENTS

1. Description of the data sets
2. Exercises
3. Further details of the variables

DOWNLOADING A DATA SET

You will find data set on the research methods 2 course materials page

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.

The data is on the following variables (C indicates a continuous variable, D a dummy variable):

Personal variables

FEMALE

D Sex of respondent (0 if male, 1 if female)

MALE

D Sex of respondent (1 if male, 0 if female)

Ethnicity:

ETHBLACK

D Black

ETHHISP

D Hispanic

ETHWHITE

D Non-black, non-hispanic

AGE

C Age in 2002

S

C Years of schooling (highest grade completed as of 2002)

Highest educational qualification:

EDUCPROF

D Professional degree

EDUCPHD

D Doctorate

EDUCMAST

D Master's degree

EDUCBA

D Bachelor's degree

EDUCAA

D Associate's (two-year college) degree

EDUCHSD

D High school diploma or equivalent

EDUCDO

D High school drop-out

Marital status

SINGLE

D Single, never married

MARRIED

D Married, spouse present

DIVORCED

D Divorced or separated

Scaled score on a component of the ASVAB battery (see Section 3 for further details on the ASVAB variables)

ASVAB2

C Arithmetic reasoning

ASVAB3

C Word knowledge

ASVAB4

C Paragraph comprehension

ASVAB5

C Numerical operations (speed test)

ASVAB6

C Coding speed (speed test)

ASVABC

C Composite of ASVAB2 (with double weight), ASVAB3 and ASVAB4

Faith:

FAITHN

D None

FAITHC

D Catholic

FAITHJ

D Jewish

FAITHP

D Protestant

FAITHO

D Other

HEIGHT

C Height, in inches, in 1985

WEIGHT85

C Weight, in pounds, in 1985

WEIGHT02

C Weight, in pounds, in 2002

Family background variables

SM

C Years of schooling of respondent's mother

SF

C Years of schooling of respondent's father

SIBLINGS

C Number of siblings

Living at age 14:

L14TOWN

D in a town or city

L14COUN

D in the country, not on a farm

L14FARM

D on a farm

LIBRARY

D Member of family possessed a library card when respondent was 14

POV78

D Family living in poverty in 1978

Work-related variables

EARNINGS

C Current hourly earnings in \$ reported at the 2002 interview

HOURS

C Usual number of hours worked per week, 2002 interview

TENURE

C Tenure (years) with current employer at the 2002 interview

EXP

C Total out-of-school work experience (years) as of the 2002 interview.

COLLBARG

D Pay set by collective bargaining, 2002

Category of employment

CATGOV

D Government

CATPRI

D Private sector

CATSE

D Self-employment

URBAN

D Living in an urban area at 2002 interview

Living in 2002 in:

REGNC

D North central census region

REGNE

D North eastern

REGS

D Southern

REGW

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.

2. EXERCISES

Note: In all of the regressions you should include a constant. Most regression packages do this automatically for you.

Exercise 1 Simple regression analysis

Does educational attainment depend on intellectual ability?

In the United States, as in most countries, there is a positive correlation between educational attainment and cognitive ability.

S (highest grade completed by 2002) is the number of years of schooling of the respondent.

ASVABC is a composite measure of numerical and verbal ability with mean 50 and standard deviation 10 (both approximately; for further details of the measure, see Section 6).

Perform a regression of S on ASVABC and interpret the regression results.

Exercise 2 Simple regression analysis

Do earnings depend on education?

EARNINGS is the hourly earnings of the respondent, in dollars, in 2002. Perform a regression of EARNINGS on S and interpret the regression results.

Exercise 3 Multiple regression analysis

Does educational attainment depend on parents' education?

Regress S on ASVABC and SM, the years of schooling of the respondent's mother. Interpret the regression results and perform t tests. Repeat the regression using SF, the years of schooling of the father, instead of SM, and again including both as regressors. There is a saying that if you educate a male, you educate an individual, while if you educate a female, you educate a nation. The premise is that the education of a future mother has a future beneficial knock-on effect on the education of her children. Do your regression results support this view?

Exercise 4 Binary choice models

What are the factors influencing going to college?

Define a binary variable COLLEGE that is equal to 1 if $S > 12$, 0 otherwise.

Regress COLLEGE on ASVABC, MALE, SM, and SF

- (a) using ordinary least squares
- (b) using logit analysis
- (c) using probit analysis

Calculate the marginal effects using logit and probit analysis and compare them with those obtained using OLS.

Exercise 5 Sample selection bias

Does sample selection bias affect the OLS estimate of the return to college education?

Using your data set, investigate whether there is evidence that selection bias affects the least squares estimate of the returns to college education. Define $\text{COLLYEAR} = S - 12$ if $S > 12$, 0 otherwise, and $\text{LGEARNCL} = \text{LGEARN}$ if $\text{COLLYEAR} > 0$, missing otherwise.

Use the heckman procedure to regress

LGEARNCL on COLLYEAR, EXP, ASVABC, MALE, ETHBLACK, and ETHHISP, with ASVABC, MALE, ETHBLACK, ETHHISP, SM, SF, and SIBLINGS being used to determine whether the respondent attended college. Run the equivalent regression using least squares and tobit. Comment on your findings.

3. FURTHER DETAILS OF THE VARIABLES

The meanings of most of the variables are obvious from the definitions given in Section 1. This section provides information on those that may need some further explanation.

The FAITH variables

The dummy variables are defined according to the response to the question "In what religion were you raised?", asked during the 1979 interview.

The ASVAB variables

The Armed Services Vocational Aptitude Battery is a series of ten tests taken by potential recruits to the military. Nearly all the NLSY79 respondents took the test as part of a project sponsored by the Department of Defense to obtain updated information on the distribution of scores that could be expected and hence to allow the raw score on a test item (number of correct responses) to be mapped to a distribution with mean 50 and standard deviation 10. Eight of the tests are power tests, that is, tests where the questions start by being very easy and then progressively become more difficult, with enough time allowed for time not to be a factor. Three of these are cognitive tests (relating to basic intelligence) and five are knowledge tests. The variables ASVAB2-ASVAB4 are the scores on the cognitive tests.

ASVAB2 is arithmetic reasoning,
ASVAB3 is word knowledge, and
ASVAB4 is paragraph comprehension.

Even the most difficult test items are fairly easy, the general purpose of the ASVAB being to discriminate among those whose education is limited to high school, the most important source of recruitment to the armed forces. The ASVABC score is a composite of ASVAB2 - ASVAB4 constructed specifically for the present data sets. It combines ASVAB2 with weight 0.5, with ASVAB3 and ASVAB4, each with weight 0.25.

An adjustment has been made to preserve the standard deviation at approximately 10 without changing the mean of approximately 50. (A similar composite, known as the Armed Forces Qualification Test score, is constructed by the military but, because it is scaled in the form of percentiles, it cannot be compared directly with the scores from which it is constructed.)

The other two tests are speed tests, consisting of very easy items with no difficulty in gradient but with so little time allowed that only a small minority of respondents can complete. One is ASVAB5, numerical operations, where a typical test item is multiplying 3 by 3. The other is ASVAB6, coding speed, in which four-digit numbers are translated to words using a simple key, the key being changed periodically. Again, the difficulty of each item is very low; the score depends on concentration and short-term memory. ASVAB5 and ASVAB6 are not specified in any of the exercises but you should try experimenting with them.

The LIBRARY variable

This dummy variable is defined according to the response to the 1979 interview question "When you were about 14 years old, did you or anyone else living with you have a library card?". Try using it in the educational attainment function.

The COLLBARG variable

This is defined to be 1 if the respondent said that her or his earnings in 2002 were determined by a collective bargaining agreement. Try using it in the earnings function.