SESSION 9

DATABASE MANAGEMENT

Moderator—David M. Sargent, SUNY Oswego

NLSDBA: The National Longitudinal surveys on compact disk

CAROL TOMASTIK SHEETS
Ohio State University, Columbus, Ohio

The National Longitudinal Survey of Labor Market Experience of Youth (NLSY) is an ongoing study of 12,686 young men and women who were aged 14 to 21 as of January 1, 1979. Over the years, the respondents have answered a wide variety of over 26,000 questions about their work experience, military experience, training, schooling, family background, marital status, income and assets, health, fertility, child care, and drug and alcohol abuse. The data from the survey are available on CD-ROM. This paper briefly describes the data set and how it may be accessed.

The National Longitudinal Survey of Labor Market Experience of Youth (NLSY) is an ongoing study of 12,686 young men and women who were aged 14 to 21 as of January 1, 1979. Over 90% of these respondents have participated in an annual personal interview, approximately 1 h in length, since 1979. In 1986, a second panel was added, composed of the children, aged 0 through 18, of the female respondents. By 1988, the number of children being interviewed was 7,346.

The study has been funded and supported by the U.S. Department of Labor, Bureau of Labor Statistics, with additional funding from the U.S. Department of Defense, the U.S. Department of Health and Human Services, including the National Institute on Child Health and Human Development, the National Institute on Alcohol and Alcohol Abuse, and the National Institute on Drug Abuse.

Over the years, the respondents have answered a wide variety of over 26,000 questions about their work experience, military experience, training, schooling, family background, marital status, income and assets, health, fertility, child care, and drug and alcohol abuse. The children were administered an extensive set of assessment instruments encompassing cognitive, socioemotional and physiological aspects of the child's development, as well as information about the quality of the home environment.

These, along with items administered to the mothers, are an additional 5,000 variables.

The Ohio State University's Center for Human Resource Research is responsible for the design, creation, and distribution of the data. The data are distributed on compact disk (CD-ROM). Designed for use in a PC/MS-DOS environment, the disk contains searching and indexing software, named NLSDBA, that allows a user to peruse documentation, select variables, create an extract file, and run descriptive statistics.

The software is menu-driven and has help screens to help the user step through the process of building an extract data file that can be used by SAS, SPSSX, or any programming language. Up to 1,024 variables can be selected for an extract. The capacity of the PC's hard disk is a limiting factor on the size of the extract. An extract of 1,024 variables from the Youth database with all 12,686 respondents would take over thirty 30 MB of disk space.

The user first selects either the Youth database or the Child database. A particularly unique and advantageous feature is available in the Child database. When creating a file of children records, any of the mothers' 26,000 variables from the Youth database can be added to the child's record. For example, if, in addition to variables from the Child database, variables about the mother's work history are desired, the software links the child to the mother and adds the selected variables from the Youth database.

The database's documentation files can be searched via numerous paths. The data are divided into general areas of interest for broad searching, as well as indexed by

Correspondence should be addressed to Carol Tomastik Sheets, Center for Human Resource Research, 921 Chatham Lane, The Ohio State University, Columbus, OH 43210.

predetermined keywords. Contextual searching can also be done by user-defined words. Additionally, those familiar with the data and the hard copy instruments may search for items by question number. As variables of interest are found, the user "selects" them by means of a function key and builds a list (an *Extract Specification File*) of variables to be extracted from the database and written to a file on the PC.

Several features assist the user in making selections. A codebook page is available for view on any item by pressing a function key. This documents the variable with question number, year, title, frequencies, and so on. The list of variables can be reviewed and altered at any time. The list is saved on the PC for use in future sessions.

Once a list of variables for extracting is complete, the user can restrict the universe of respondents. Menus offer a list of basic universe restrictors, such as race, sex, and number of cases. Additionally, a user-defined Boolean string can be constructed to further control the sample.

Along with the ASCII data file, either SAS or SPSSX control cards and an index with descriptive statistics for the data file are generated. A codebook of the selected variables is optional.

Execution times will, of course, vary by hardware, with the access time of the CD reader being the critical factor. However, typical wall-clock time for all 12,686 NLSY respondents (the number of variables being extracted has a negligible effect on processing times) takes about 2½ h on a 386-based PC. Similarly, a child-based extract of all 7,346 children takes about 45 min. A child-based extract containing mother variables takes about 3 h.

Availability

The NLSY CD-ROM and documentation are available from PLS Public User's Office, Center for Human Resource Research, 921 Chatham Lane, Suite 200, Columbus, OH, 43210-2418 (or E-mail: usersvc@OHSTHR).