Introduction to the NLSY

The **National Longitudinal Surveys (NLS)** are a set of surveys designed to gather information at multiple points in time on the labor market and other significant life events of several groups of men and women.

Topics covered in NLS Surveys

- **Education, Training and Achievement Scores:** information about high school and college experience, types of training, and cognitive test scores

- **Employment:** extensive history of employment over time, including wages, hours, occupation, industry, and job search

- **Health:** Health status, weight, and health conditions

- **Attitudes, Expectations, and Non-Cognitive Tests:** personality traits, expectations about the future, locus of control

- And many more...

Survey	Birth Years	Interview Years
Older Men	1906-1921	1966-1990
Mature Women	1922-1937	1967-2003
Younger Men	1941-1952	1966-1981
Younger Women	1943-1953	1968-2003
NLS Youth – 1979 cohort	1957-1964	1979-present
NLS Youth – 1997 cohort	1980-1984	1997-present
NLSY Child and Young Adult	1975-present	1984-present

Table: List of NLS cohorts

NLS Data

- Based on birth cohorts
- US nationally representative
- Oversamples (military, poor, minorities)
- Interview-based data set focused on labor market experience
- Periodic interviews

National Longitudinal Survey of Youth, 1979 Cohort (NLSY79)

- Original sample size of 12,686 individuals born between 1957 and 1964

- First interviews in 1979 with respondents aged 14 to 22

- Annual interviewing from 1979 to 1994, biennial interviewing since 1996

- Most recent interviews in 2014 with respondents aged 49 to 58 and a sample size of 7,071

Types of information gathered in NLSY79

- ▶ Labor market behavior
- ▶ Educational experiences
- ▶ Family background
- ▶ Armed Services Vocational Aptitude Battery (measures knowledge and skills including reading and mathematics)
- ▶ High school information (from respondents' schools and school transcripts)
- ▶ Government program participation
- ▶ Family life
- ▶ Health issues
- ▶ Assets and income

Subsample Design

Three subsamples

- **Cross-Sectional Sample:** 6,111 individuals that represent the noninstitutionalized civilian segment of people living in the US in 1979

- **Supplemental Sample:** 5,295 civilian Hispanic or Latino, black, and economically disadvantaged non-black/non-Hispanic individuals living in the US in 1979

- **Military Sample:** 1,280 individuals that represent the population serving in the U.S. military as of September 30, 1978

Changes in the subsamples

The subsamples have been subject to significant changes over time:

- After the 1984 interview, 1,079 members of the military sample were no longer eligible for interview; 201 respondents randomly selected from the military sample remained in the survey

- After the 1990 interview, none of the 1,643 members of the economically disadvantaged, non-black/non-Hispanic sample were eligible for interview

Pros & Cons

Pros:

- ▶ Covers a long time period
- ▶ Large part of the data set is freely available
- ▶ Lots of helpful internet resources (use them!)

Cons:

- ▶ Significant reduction of the sample size over the years
- ▶ Not the whole data set is freely available (might be a problem for replication)

Longitudinal availability of variables

- ▶ Note that only a small share of variables is available for all points in time.

- ▶ Go to `https://www.nlsinfo.org/content/cohorts/nlsy79/topical-guide/asterisk/i-labor-market-experience-variables` to check for which timepoints a certain variable is available.

- ▶ Helpful to do this quite early in order to avoid data availability problems at later stages.

NLSY79 Child and Young Adult

- ▶ Co-resident children of women in the NLSY79 have been followed with their mothers since 1986
- ▶ Not based on a certain birth cohort
- ▶ Psychosocial assessments through age 10
- ▶ Children graduate to "Young Adults" at age 15
- ▶ Significant overlap of the children sample with NLSY97

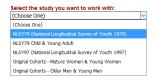# National Longitudinal Survey of Youth, 1997 Cohort

- Original sample size of 8,984, including oversamples of Blacks and Hispanics
- Annual interviewing, beginning in 1997
- Good for comparison with earlier cohorts
- Disadvantage: less (non-)cognitive info and more attrition

Access NLSY data

- To access NLSY data, visit `https://www.nlsinfo.org/investigator/pages/login.jsp`
- You can either create an account for free or search for data as a guest.
- Setting up an account only requires an email address and a username, not necessarily your real name.

Selecting a data source

Select the data source that you want to use.

For this example, we will work with NLSY79.

Tagsets tab

Tagsets tab

After selecting the study you will see a series of tabs. The first one is to choose a *tagset*. Tagsets are basically saved searches.

Note that by default six variables will be added to your data:

- ▶ id (can't remove this)
- ▶ the following optional demographics:
    - ▶ gender
    - ▶ age
    - ▶ race/ethnicity
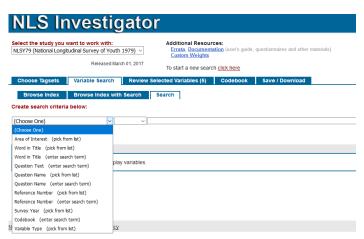   - ▶ birthday (month and year)

"Variable search" tab

"Variable search" tab

"Variable Search" offers three search modes:

- by topics (index)
- by searching the index
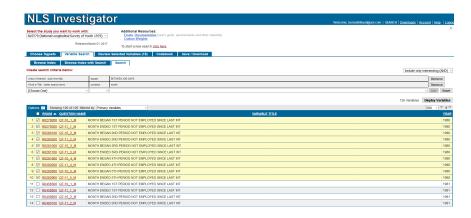- a general search.

Note that the site does not offer access to the full data sets.

Searching for data (1)

The general search offers a variety of options. From the dropdown menu select the type of search you want.
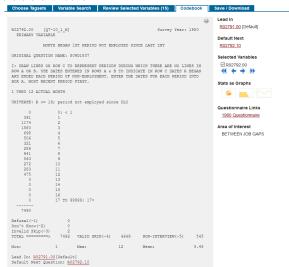
Searching for data (2)

Searching for data (2)

This is a nested search in which we are looking for "month" in "between job gaps". You can add more terms if needed.

Results will be shown in the lower box. Sometimes you will see a "+" next to RNUM which means that there are more variables available (as a subset of responses). Check the square next to RNUM to select all variables or check the ones you need.
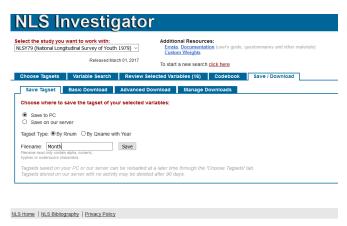
If you hover the cursor over the question name, a window will pop up with detailed information about that variable.
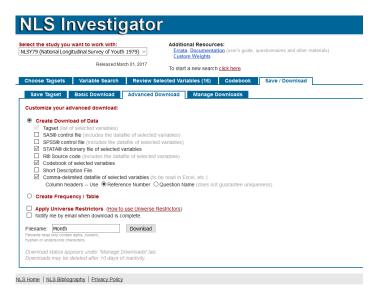
Codebook tab

Here you can get some additional information about your variables along with some basic stats

Downloading data (1)

Once you are satisfied with your search, you can save it as a "tagset". You can keep the defaults, select a name (in this case we choose "Month") and click on "Save".

Downloading data (2)

Downloading data (2)

1. Select the data format you want. For this example we will select Stata and comma-delimited along with the codebook.
2. Select a name for the file and click "Download"

Downloading data (3)

Downloading data (3)

If you set your browser to ask you where to save, a pop-up window will prompt you to select a folder location for the zip file, click "OK" and select the folder.

If it gets downloaded automatically, then the file should be in the "Downloads" folder or the default place for downloads.

Further information

- For detailed information on all NLS cohorts:
  https://www.bls.gov/nls/handbook/nlshndbk.htm

- For an extensive user guide to the NLSY79:
  https://www.bls.gov/nls/79guide/2001/nls79g0.pdf

- For an extensive user guide to the NLSY97:
  https://www.bls.gov/nls/97guide/rd5/nls97ugall.pdf

Further information

- For information on how to use NLSY in STATA;
  http://dss.princeton.edu/training/NLS101.pdf

- For a tutorial on how to construct comparable sample across NLSY79 and NLSY97:
  https://www.nlsinfo.org/content/getting-started/
  intro-to-the-nls/tutorials/
  tutorial-constructing-comparable-samples-across

- For information on sample weights/clustering adjustments:
  https://www.nlsinfo.org/content/cohorts/nlsy79/
  using-and-understanding-the-data/
  sample-weights-clustering-adjustments