Overview of the National Longitudinal Survey of Youth 1997 (NLSY97)

Hsueh-Sheng Wu
CFDR Workshop Series
March 14, 2022
Outline

• Introduction to the NLSY97
• Advantages and challenges of using NLSY97
• Navigation through NLSY 97 website
  ➢ Asterisk tables
  ➢ Marital history, childcare, and fertility
  ➢ NLSY 97 sample
  ➢ Using and understanding Data
  ➢ Other documentations
• How to obtain the NLSY97 data
  ➢ NLS Investigator website
  ➢ How to obtain weight variables
• How to weight NLSY97 data
• An example of Using NLSY97 data
• Journal articles using NLSY97
• Conclusions
Introduction to the NLSY 97

• NLSY97 is one of the five cohort studies conducted by the U.S. Bureau of Labor Statistics (see https://www.bls.gov/nls/cohorts.htm).

• NLSY97 was designed to document how the youth transits from school to work and into adulthood.

• The subject areas of NLSY97 now include:

<table>
<thead>
<tr>
<th>Education, Training &amp; Achievement Scores</th>
<th>Income, Assets &amp; Program Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment</td>
<td>Health</td>
</tr>
<tr>
<td>Household, Geography &amp; Contextual Variables</td>
<td>Attitudes, Expectations, Non-Cognitive Tests, Activities</td>
</tr>
<tr>
<td>Parents, Family Process &amp; Childhood</td>
<td>Crime &amp; Substance Use</td>
</tr>
<tr>
<td>Dating, Marriage &amp; Cohabitation; Sexual Activity, Pregnancy &amp; Fertility; Children</td>
<td></td>
</tr>
</tbody>
</table>
• NLSY 97 is a cohort survey whose target population is youths who were 12 to 16 years old as of December 31, 1996.

• NLSY97 was conducted annually between 1997 and 2012 and biennially afterwards. Nineteen rounds of NLSY97 data are currently available. Not every subject topic was examined in every round of NLSY97.

• NLSY97 used a multi-stage stratified sampling method to select eligible households. Blacks and Hispanics were over-sampled. At the first phase, 96,512 households were selected from 1,748 sample segments embedded within 147 primary sampling units (PSUs). At the second phase, screening interviews were used to identify 9,907 households with eligible youths, and 8,984 of them participated in the 1997 survey.

• The NLSY97 data were mainly collected using Computer-Assisted Personal Interview (CAPI), Audio Computer-Assisted Self-Interview (ACASI), and questionnaires.
Advantages and Challenges of Using NLSY97

Advantages:
- NLSY97 covers many research topics.
- NLSY97 oversampled Black and Hispanic respondents, which allows for examining racial differences on many topics.
- NLSY97 data have been collected at a one- or two-year intervals, which provides vast information on when, why, and how youths transitioned into adulthood.
- Many constructed variables are included in NSLY97 data, and tutorials are provided for researchers to construct additional variables.
- Custom weight variables can be generated for researchers using certain rounds of NLSY97.

Challenges:
- As data collected by a cohort design, the findings of NLSY 97 may not be able to generalized to people of other birth cohorts.
- Changes to designs and questionnaire of NLSY97 over time create some problems in analyzing data longitudinally.
- Researchers should learn to use the NLS Investigator website to select and download the data.
- Special statistical techniques are needed when researchers analyzing data from multiple rounds of NLSY97.
NLSY 97 Website
(https://www.nlsinfo.org/content/cohorts/nlsy97)
### National Longitudinal Survey of Youth 1997

**V. Family Formation (age restrictions as of end of previous calendar year—12/31/96 in rd 1, 12/31/97 in rd 2, and so on)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>R7</th>
<th>R8</th>
<th>R9</th>
<th>R10</th>
<th>R11</th>
<th>R12</th>
<th>R13</th>
<th>R14</th>
<th>R15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spells Y married or lived with sexual partner</td>
<td>Y216</td>
<td>Y216</td>
<td>Y216</td>
<td>Y216</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Y216</td>
<td>Y216</td>
<td>Y216</td>
<td>Y216</td>
<td>Y216</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Y's current marital status</td>
<td>Y216</td>
<td>Y216</td>
<td>Y216</td>
<td>Y216</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Duration of Y's and partner's cohabitation</td>
<td>Y216</td>
<td>Y216</td>
<td>Y216</td>
<td>Y216</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
NLSY 97 Website (Cont.)
Marital history, childcare, and fertility
NLSY 97 Website (Cont.)

NLSY 97 sample
NLSY 97 Website (Cont.)

Using and understanding data

National Longitudinal Surveys
A Program of the U.S. Bureau of Labor Statistics

Getting Started | Cohorts | Access Data / Investigator

NLSY97
- Topical Guide to the Data
- Intro to the Sample
- Using & Understanding the Data
- Survey Instruments (Questionnaires)
- Types of Variables: Raw, Symbols, Rosters & Created
- Sample Weights & Design Effects
- Interviewer Remarks, Characteristics & Contacts
- Item Nonresponse & Interview Timings
- Interview Validation
- NLS Investigator
- NLSY97 Documentation
- Other Documentation
- Codebook Supplement
- Geocode Codebook Supplement
- Questionnaires
- Errata

National Longitudinal Survey of Youth | 1997

Using and Understanding the Data
Information on survey instruments, variable types, the interviewing process, item nonresponse, sample weights and design effects, data documentation, and how to access the data.

- Survey Instruments (Questionnaires)
- Types of Variables: Raw, Symbols, Rosters & Created
- Sample Weights & Design Effects
- Interviewer Remarks, Characteristics & Contacts
- Item Nonresponse & Interview Timings
- Interview Validation
- NLS Investigator
- NLSY97 Documentation
NLSY 97 Website (Cont.)

Other documentations

National Longitudinal Surveys
A Program of the U.S. Bureau of Labor Statistics

Getting Started | Cohorts | Access Data / Investigator

Cohorts
- NLSY97
  - Topical Guide to the Data
  - Intro to the Sample
  - Using & Understanding the Data
  - Other Documentation
    - Codebook Supplement
    - Geocode Codebook Supplement
    - Questionnaires
    - Errata
    - Tutorials
    - Technical Sampling Report
  - Get Data
    - NLSY79
    - NLSY76 ChildYA
    - NLS Mature and Young Women
    - NLS Older and Young Men

Helpful Links
Subscribe for Updates
Tutorials
Investigator User Guide

OfficeSetup.exe  |  nlsy97_2022 - HS...p0x  |  nlsy97_2022 - HSP.p0x

National Longitudinal Survey of Youth 1997

Other Documentation
Tutorials, Codebook Supplement, Geocode Supplement, Technical Sampling Report, List of Errata, and questionnaires for all rounds.

- Codebook Supplement
- Geocode Codebook Supplement
- Questionnaires
- Errata
- Tutorials
- Technical Sampling Report
How to Obtain NLSY97 Data

NLS Investigator
(https://www.nlsinfo.org/investigator/pages/login)
How to Obtain the NLSY97 Data (Cont.)
How to Obtain NLSY97 Data (Cont.)

Select the study you want to work with:
- NLSY97 (National Longitudinal Survey of Youth 1997)

Select a substudy:
- NLSY97 1997-2019 (rounds 1-19)

Additional Resources:
- Errata, Documentation (user's guide, questionnaires and other materials)
- Custom Weights

To start a new search, click here.
How to Obtain the NLSY97 Data (Cont.)
How to Obtain NLSY97 Data (Cont.)
How to Obtain NLSY97 Data (Cont.)
How to Obtain Weight Variables of NLSY 97

Sampling weights at each round, sampling panel weights, and design effect variables at 1997 are available via NLS Investigator.
How to Obtain NLSY 97 Weight Variables (Cont.)

![NLS Investigator screenshot](https://www.nlsinfo.org/investigator/pages/search#)

Select the study you want to work with:
- NLSY97 (National Longitudinal Survey of Youth 1997)

Select a sub-study:
- NLSY97 1997-2019 (rounds 1-19)

Additional Resources:
- Errata Documentation (user's guide, questionnaires and other materials)
- Custom Weights

To start a new search [click here](#)

Create search criteria below:

- **Word in Title** (pick from list)
- **is** VARIANCE (12)

**Index of Selected Variables**

<table>
<thead>
<tr>
<th>RNUM</th>
<th>QUESTION NAME</th>
<th>VARIABLE TITLE</th>
<th>YEAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R1489700</td>
<td>VSTRAT</td>
<td>1997</td>
</tr>
<tr>
<td>2</td>
<td>R1489800</td>
<td>VPSU</td>
<td>1997</td>
</tr>
<tr>
<td>3</td>
<td>R9707600</td>
<td>ASVAB GS POST VARIANCE</td>
<td>1999</td>
</tr>
<tr>
<td>4</td>
<td>R9707700</td>
<td>ASVAB AR POST VARIANCE</td>
<td>1999</td>
</tr>
<tr>
<td>5</td>
<td>R9707800</td>
<td>ASVAB WK POST VARIANCE</td>
<td>1999</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>2000</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>2001</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>2002</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>2004</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>2005</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>2006</td>
</tr>
</tbody>
</table>

[12 Variables](#) [Display Variables](#)
How to Obtain NLSY 97 Weight Variables (Cont.)

Customized Panel Weights are available at:
https://www.nlsinfo.org/weights/nlsy97
How to Obtain NLSY97 Weight Variables (Cont.)

- NLSY97 panel weight variables can be customized, depending on how many rounds of surveys are used and whether the focus is on respondents participating in all or any of these rounds.

- The customized panel weight variables were created by adjusting for differences in base weight, nonresponse to the screener, different response rates to the screener, different cross-sectional weights for Black and Hispanic respondents, nonresponse to the main interview.

- NLSY97 customized panel weight variables are post-stratification person weight variable.
How to Weight NLSY97 Data

Source: https://www.nlsinfo.org/content/cohorts/nlsy97/using-and-understanding-the-data/sample-weights-design-effects/page/0/1

use NLSY97DATAFILE.dta, clear
svyset [pweight=WTVAR], strata(vstrat) psu(vpsu) singleunit(scaled)
svy: mean ANALYSISISVAR // mean for continuous variables
svy: proportion ANALYSISISVAR // proportion for categorical variables
svy, subpop (if SUBGROUP==1): mean ANALYSISISVAR // mean within a subpopulation
svy: tabulate ANALYSISISVAR // one way table
An example of Using NLSY97 data

- Research questions: Do age and education predict the respondent’s marital status on December 2019?
- Select variables and weight variables
- Download the data
- Use the Stata to read in the data
- Examine the data to make sure the data are read in correctly.
- Code the missing values of variables
- Identify the analytic sample
- Specify survey design characteristics
- Conduct the descriptive and inferential analysis.
Journal Articles Using NLSY 97


Conclusions

- NLSY97 provides an excellent opportunity to examine how the youth transited into the adulthood. With measures of various subject areas, NLSY may shed light on why some youths made the transition while some did not.

- NLSY97 already includes many constructed variables on the timing and history of many events, so researchers do not need to start everything from scratch.

- Excelletn tutorials are provided at NLSY website and document how to link different rounds of NLSY data and constuct additional variables.

- When analyzing NLSY97 data, researchers need to correctly specify the survey design characteristics to obtain the correct estimate.

- If you have any questions about using NLSY97 data. Please feel free to contact Hsueh-Sheng Wu at wuh@bgsu.edu.