



Lab Exercise #3: Linking ASEC to Basic Monthly (Mechanical Linking)

Learning Goals

- Use a long file to identify individuals using CPSIDP in both the ASEC and Basic Monthly
- Understand the implications of ASEC oversample for linking
- Explore pros and cons of long versus wide linked files

Summary

In this exercise, you will link data from the 2014 Food Security supplement (FSS) to the 2015 ASEC. Make an extract including these two samples along with the variables CPSIDP, ASECOVERP, MISH, FSSTATUS, FSHWTSKALE, CTCCRD, AGE, SEX, and RACE.

Future exercises will build on this work. Save any do files to your IPUMS flash drive. You may also choose to save extracted or modified data files locally.

Exercises

1. Why is CPSIDP preferable over MARBASECID for linking the ASEC to the December FSS?
-
-

Create a file that contains both the 2014 FSS and 2015 ASEC in which records are stacked one on top of the other (long file).

Hint: A default IPUMS extract will include a new row for each person-month set of observations. See the Stata syntax guide for information on ordering or sorting your data.

2. Check for duplicate values of CPSIDP. Does it look like you expect? What does it mean when a CPSIDP value appears only once? Twice? Explain.
-
-

3. What values of CPSIDP appear multiple times? Tabulate the variable you created to identify duplicate observations with ASECOVERP. Who are the individuals that have a CPSIDP value of 0?
-

4. What are the MISH values for December records that link? ASEC records?
-

5. How many records link successfully between the two time points? How many records from the December FSS that were eligible to link to the ASEC did not link successfully? What are reasons they might not link?
-
-

Create the same linked file with data from the two time points with every observation for a person on a single record (wide file).

6. Do you have the same number of links? _____
 7. In what situation might a wide file be preferable to a long one?
-
-

Answers

1. Why is CPSIDP preferable over MARBASECID for linking the ASEC to the December FSS?

CPSIDP is suitable for linking across any CPS samples, while MARBASECID can only link ASEC samples to March basic monthly samples. Using CPSIDP is cleaner, faster, and safer.

2. Check for duplicate values of CPSIDP. Does it look like you expect? What does it mean when a CPSIDP value appears only once? Twice? Explain.

There are 69,213 records that have a non-unique CPSIDP value of 0. This is expected, as ASEC samples contain an oversample that cannot be linked to basic monthly samples. The number of times a given CPSIDP value is observed indicates how many times that person appeared in the cps during the months included in your dataset.

3. What values of CPSIDP appear multiple times? Tabulate the variable you created to identify duplicate observations with ASECOVERP. Who are the individuals that have a CPSIDP value of 0?

CPSIDP values of 0 appear multiple times. All of these observations come from the ASEC oversample.

4. What are the MISH values for December records that link? ASEC records?

Records that link have MISH values of one and five in December 2014 and four and eight in the 2015 ASEC.

5. How many records link successfully between the two time points? How many records from the December FSS that were eligible to link to the ASEC did not link successfully? What are reasons they might not link?

28,174 link successfully, 31,729 are eligible. There are 3,555 eligible records in the Food Security supplement that do not link to the ASEC. This maybe because of migration or death.

6. Do you have the same number of links? Yes.

7. In what situation might a wide file be preferable to a long one?

In instances such as this where one or more variables only appears in one of the months being linked, a wide file may be preferable as there is no need to deal with missing values