Hello Philipp, Tobias,

I apologize for the delay. Here is the information about the cases with occupation codes we were able to recode. ==Steve

Undocumented occupation codes have been revised for two variables, affecting a small number of cases. Listed below are the existing incorrect codes, the respondent id number (R00001.00) and the corrected code. These corrections will be added in the next public data release.

R72096.00 OCCALL-EMP.01 OCCUPATION (CENSUS 3 DIGIT 00 CODES (ALL), JOB #01 2002

Undocumented code 26		Undocumented code 426	
R00001.00	Revised code	R00001.00	Revised Code
1306	56	6170	625
4564	56,	6785	625
5585	385		
7576	56		
9112	874		

R7898100   OCCALL-EMP.02   OCCUPATION (CENSUS 4 DIGIT, 00 CODES) (ALL) JOB #02 2004

Undocumented code 260		Undocumented code 30	
R00001.00	Revised code	R00001.00	Revised code
1306	560	2027	3920
3933	3800	2469	3800

Dear Steve,

I hope all is well, greetings from Bonn. Tobias and I have been working together on the NLSY dataset for the last months. We just wanted to thank you very much for your guidance on navigating the information in the NLSY and all the intricacies involved.

Even though we will be sure to further expand and improve our research dataset in the future, we will now pause for a bit and take some time to document the current version. So we thought this a good moment to share our appreciation for your help.

Hello Tobias,

I'm embarrassed and apologize terribly for not getting back to you sooner. I talked to the archivist and there isn't anything we can do about the undocumented codes from the older rounds. This information was collected on paper and pencil questionnaires and not saved. However, with the information collected electronically since 1994, we can likely bring up the respondent's verbatim responses.

I also see you have an another email I did not respond to. I will get back to you on these as soon as I can. Again, I apologize for the delay.

Steve

From: Tobias Raabe [mailto:tobiasraabe@uni-bonn.de]
Sent: Monday, July 24, 2017 8:16 AM
To: Steve McClaskie <steve.mcclaskie@chrr.osu.edu>
Subject: AW: Coherent occupational categories

Hey Steve,

I just wanted to ask whether there are any updates on this issue and on the a different email titled "NLSY79 - Questions regarding occupational and educational data".

I know I had many questions to you over the past three months or so and I am very grateful for your support because otherwise I would not have been able to do all the things I wanted until now. I also understand if you delay responses to me since I am not the only person who needs help and I depleted my "budget".

Nonetheless, can you give me a short update on both issues whether they are forgotten, have stalled or a response is expected in near future?

Thank you very much.

Best regards,

Tobias

Von: Tobias Raabe
Gesendet: Donnerstag, 22. Juni 2017 13:32
An: Steve McClaskie
Betreff: AW: Coherent occupational categories

Hello Steve,

I have some examples for you:

CASEID	Undocumented value	Reference numbers	Census Code set
1906	7	R3522800	1970
4921	7	R4206100, R4587904, R5310900	1970
3245	19	R3354700	1970
1306, 4564, 5585, 7576, 9112	26	R7209600	2000 (three-digit)
6170, 6785	426	R7209600	2000 (three-digit)
1306	260	R7898100	2002 (four-digit)
2027, 2469, 3933	380	R7898100	2002 (four-digit)

I hope this sample helps to verify my problem. In addition, with my current approach I assume that there is a total of 83 unique values which cannot be mapped to 1970 Census codes, 6 unique values for the three-digit 2000er codes, and 4 unique values for the 2002 codes.

Up to know, I have not compared the 2003 codes, since I do not have access to a dataset version of the PDF.

I hope you can help me to clarify my problems. Thank you for your effort. If these are indeed incorrect values – and I hope I am not wasting your time on this – I can give you a full list of incorrect values, case ids and reference numbers.

Best regards,

Tobias

Von: Steve McClaskie
Gesendet: Mittwoch, 21. Juni 2017 18:34
An: Tobias Raabe
Betreff: RE: Coherent occupational categories

Hi Tobias,

Could you give me the reference number for the variables where you are finding undocumented codes. Thanks.

Steve

From: Tobias Raabe [mailto:tobiasraabe@uni-bonn.de]
Sent: Wednesday, June 21, 2017 10:24 AM
To: Steve McClaskie <steve.mcclaskie@chrr.osu.edu>
Subject: AW: Coherent occupational categories

Hello Steve,

thank you very much for your answer.

I decided to use the IPUMS' OCC1990 as the foundation of the coherent occupation category system and do not map each decennial Census system to the next. It was also mentioned in the essay you added to your last mail (https://usa.ipums.org/usa/chapter4/chapter4.shtml).

While mapping 1970er, and the three version of 2000er Census codes to OCC1990, I discovered that the NLSY79 data includes occupation codes which are not documented in the PDFs (http://nlsinfo.org/content/cohorts/nlsy79/other-documentation/codebook-supplement/nlsy79-attachment-3-industrial-and). Some examples:

Undocumented values for 1970 (83 counts):
        7, 19, 60, 66, 67, …
Undocumented values for 2000 three-digit (95 counts):
        24, 25, 44, 45, …
Undocumented values for 2002 four-digit codes (3 counts):
        260, 380, 9840

How do I have to think about those values? I interpreted the gaps in the occupation codes as unused values. Does seven in 1970er codes specify a subclass of six or something like this?

Best regards,

Tobias

Von: Steve McClaskie
Gesendet: Mittwoch, 14. Juni 2017 18:28
An: Tobias Raabe
Betreff: RE: Coherent occupational categories

Hello Tobias,

The question of harmonizing the NLSY79 I/O codes comes up from time to time. Below is my reply. I think this pretty much mirrors your description of how you are going about the problem.

There's no easy answer. The 1970 Census 3-digit codes are used for all jobs through 2000, then beginning in 2002 there is a change to the 2000/2002 Census industry and occupation codes are. You will need to map the 1970 codes to 1980 codes, 1980 codes to 1990 codes, and then 1990 codes to 2000 codes. Attached is a technical paper by Census that discusses converting 1970 codes to 1980 codes. Also attached is a Census memo on converting 1980 to 1990 codes. Here is the link to a Census page on revisions from 1990 to 2000, https://www.census.gov/people/io/methodology/. The University of Minnesota has a center called IPUMS. Their mission is to release all the Census variables. They have a good manual on converting I/O over time into a consistent framework. Their documentation, which goes much deeper is found here, https://usa.ipums.org/usa/chapter4/chapter4.shtml.

It has been our understanding that a trailing 0 was added to the 3-digit Census 2000 codes to create the 4-digit 2002 codes. However, I cannot point you to the Census page that documents this, other than the individual tables at https://www.census.gov/people/io/methodology/.

Our 2000 codes are in pdf format and so is the same Census table, so far as I can find. However, the 2002 Census tables are in Excel. Might you delete the trailing 0 to create your own Excel file of 2000 codes?

I hope this helps. Please let me know if you have further questions or concers.

Steve

Steve McClaskie
NLS User Services
614 442-7366

Dear NLSY support,

My name is Tobias Raabe and I am a research assistant at the Institute for Applied Microeconomics in Bonn.

I am currently working on building an occupational classification for NLSY79 respondents which is coherent for all years. As indicated in http://nlsinfo.org/content/cohorts/nlsy79/other-documentation/codebook-supplement/nlsy79-attachment-3-industrial-and, the problem is that NSLY79 uses 1970 census occupation codes from years 1979-2000. After that, it uses three different category systems of 2000er codes which are all based on SOC2000 (https://www.bls.gov/soc/).

My general question is whether there is some coherent occupation classification available for the NLSY79 or whether you know of classifications from other sources.

My current attempt to tackle the problem is to use a classification created by IPUMS called OCC1990 (https://usa.ipums.org/usa-action/variables/OCC1990#description_section for more description and https://usa.ipums.org/usa/resources/volii/documents/occ1990_xwalk.xls for the actual crosswalk). The variable builds a bridge from the 1970er census codes to the three-digit 2000er census codes. Unfortunately, the two four-digit codes are not included.

To convert the four-digit codes to the 2000er census codes, I think I can use the underlying SOC classification, but then I would need this document (https://www.nlsinfo.org/sites/nlsinfo.org/files/attachments/121217/att300.pdf) from attachment 3 as a dataset and with SOC codes for the 2000er codes. Is this document available as a dataset?

I looked at other places for a crosswalk between the different 2000er classifications but one problem is that I have not found a three-step revision process as you described in your documentation for the census system in the 2000er. I assume the change from three to four digit codes was adding a trailing zero which seems to be mostly true but I have not found something similar to a changelog for the census classification. (I have looked a other sources following information on BLS https://www.bls.gov/cps/cpsoccind.htm but I am not sure how their revisions relate to yours and vice versa.)

I would appreciate your remarks and I hope you can help me to address my general or more specific problem.

Best regards,

Tobias Raabe