

Understanding Society User Support - Support #1840

errors in 4-digits ISCO88 variable

01/16/2023 09:35 AM - Matteo Pinna Pintor

Status:	Resolved	Start date:	01/16/2023
Priority:	High	% Done:	100%
Assignee:	Understanding Society User Support Team		
Category:	Data inconsistency		
Description Dear survey help team, Last spring I submitted a request for special license version in order to use the four-digits ISCO88 codes. I received the data a couple of months ago, after quite a bit of waiting time. Now that time has come to make good use of it, however, I realize to my dismay that there are it seems to me debilitating problems with the variable. While using the code to link the survey with other, publicly available information, many non-matches on the survey side emerged. Upon inspection and cross-checking against ILO manuals, it appears to me that many values of the variable jbisco88 simply do not exist in the ISCO88 classification (and, at a glance, not even in its European variation). The codes in question all end with a zero at the fourth, last digit, while a positive integer always occupies this place in the official list. I attach a list exported from Stata to Excel. If the fourth-digit zero means, as I'm guessing, that the survey somehow could not elicit more precise information, then I would like to have this confirmed, and I would strongly advise to flag this up in the code, not to add an arbitrary zero. The zero sometimes does exist in the fourth digit, so it's available to signal missing information and its use in this sense is bound to create confusion. I looked for information on the documentation, but found no indication about this point. Second, if this is indeed the case, it amounts to having a variable that contains much less information that it could - and should, considering that by default it is reasonable for a user to expect a complete variable. Waiting many months for an incomplete variable, with 36 out of 390 codes no richer than the condensed version, is not acceptable. For most serious research purposes, such a gap makes the variable useless. This can lead to serious time waste on the user side. If instead the reason is different, and some remedy (other than selectively stepping back to use the 3-digit values) available, I'd like to know about it. Best regards, Matteo Pinna Pintor			

History

#1 - 01/16/2023 09:38 AM - Matteo Pinna Pintor

erratum (somehow edit doesn't work): "...The zero sometimes does exist in the fourth digit, so it's NOT available to signal missing information"

#2 - 01/18/2023 12:17 PM - Understanding Society User Support Team

- Status changed from New to In Progress

- % Done changed from 0 to 10

Many thanks for your enquiry. The Understanding Society team is looking into it and we will get back to you as soon as we can.

We aim to respond to simple queries within 48 hours and more complex issues within 7 working days. While we will aim to keep to this response times due to the current coronavirus (COVID-19) related situation it may take us longer to respond.

Best wishes,
Understanding Society User Support Team

#3 - 01/25/2023 01:37 PM - Understanding Society User Support Team

- Private changed from Yes to No

Dear Matteo,

First of all, I am sorry for the inconvenience, I understand that sometimes the waiting time to get access to UKHLS data can be frustrating and full explanation of all aspect of the data is not always readily available. We are currently investigating why these codes were assigned in some cases, I will get back to you when I know more.

Regards,

Piotr
UKHLS User Support Team

#4 - 01/26/2023 04:17 PM - Matteo Pinna Pintor

- File *fake4digitsisco88_2.JPG* added

Dear Piotr,

thanks for your reply and appreciated efforts - and apologies for the initial disgruntled tone. Looking at the labeling of values, I can now guess what is perhaps the case (and perhaps also documented somewhere) - that is, that the ISCO88com version has been used. The class headings of those non-matching classes track the European version much better than the international version. In three cases, the code actually exists in ISCO88com (2470, 9320, 9330). However, the problem remains for all other codes. I attach the completed version of the list provided above.

Two other apparent anomalies. In all observations coded 2470 (again, existing in the ISCO88com version), the limited 3-digit version is always coded as missing value (-9). Moreover, the code for "Armed forces" seems to be mistaken in its own way: the ISCO88 is 0110, the ISCO88com is 0100, while the survey code for jbisco88 is 100 (and its limited version 10) - in other words, the problem here is not a trailing zero in place of a positive integer, but a missing leading zero (vis-à-vis ISCO88com) or an altogether different number (vis-à-vis ISCO88).

I wait for your updates.

Best,
Matteo

#5 - 01/26/2023 05:43 PM - Matteo Pinna Pintor

I also realize now the issue is less debilitating than at first sight, because the codes with the fourth-digit zeroes seem to apply only to some observations, and there are other observations for those 3-digit classes in which the last digit is a positive number.

Best,
M

#6 - 01/27/2023 04:05 PM - Understanding Society User Support Team

- Status changed from *In Progress* to *Feedback*

- % Done changed from 10 to 50

Dear Matteo,

Thank you for your additional clarifying comments.

The coding into ISCO-88 in UKHLS was done by translating the SOC 2000 (the British Standard Occupational Classification) codes using a correspondence table, in other words, the coding into ISCO-88 was not done from the raw text fields. Wherever it was impossible to find a match for a given SOC 2000 code, one of the codes you identified was assigned. A few examples:

1140 "SENIOR OFFICIALS OF SPECIAL-INTEREST ORGANISATIONS" (non-existing 4digit ISCO-88, but present in the UKHLS variable) - the SOC 2000 code in this case was Senior officials of special interest organisations, so it's more general that the 3 ISCO-88 unit group titles (1141,1142,1143), hence a direct match is impossible

2110 "PHYSICISTS, CHEMISTS AND RELATED PROFESSIONALS" (non-existing 4digit ISCO-88, but present in UKHLS variable) - the SOC 2000 code in this case was 2113 "Physicists, geologists and meteorologists", again, a direct match is not possible, because the ISCO-88 unit group titles combine individual occupations differently, that is, meteorologists are a separate unit group and so are geologists

Thank you very much for your feedback, we will review the coding procedure to see if we can make it more accurate. However, please note that in case any amendments were to be done, this would be available - at the earliest - in the next data release, i.e. November 2023.

Best wishes,
Piotr
UKHLS User Support Team

#7 - 02/02/2023 08:49 AM - Matteo Pinna Pintor

Dear Piotr,

thanks for the clarification. Let me also point out, both to ask your confirmation and for future readers, that for at least some of the codes involved, the match up to the fourth digit was indeed sometimes possible and sometimes not. I found this out while trying to merge my data based on a modified version of the jbisco variable in which I *replaced* all codes listed above from my other data, and included instead averaged observations (in my case it was reasonable) within the respective 3-digit groups, using zero as the fourth digit number. That is, a substituted the real codes with the survey pseudo-codes. When I did this and merged the datasets, it appeared that they survey contained observations which did not match because they did indeed have ISCO codes with those three digits, AND also a positive integer as a fourth digit. Everything matched when I instead added the pseudo-codes to my other data, instead of replacing the real ones.

For example, while there are - as mentioned - survey observations with the pseudo-code ISCO 2110 ("Physicists, chemists and related professionals", the label comes from the 211 3-digit ISCO group <https://www.ilo.org/public/english/bureau/stat/isco/isco88/211.htm>), because their

SOC code is either 2113 ("Physicists, geologists and meteorologists") or 2321 ("Scientific researchers"), there are also survey observations with the real ISCO value 2113 ("Chemists"), because their SOC code is 2111 ("Chemists").

So, to further specify the diagnosis of the problem, and Piotr please either confirm or correct me: for some 3-digit ISCO groups in the survey, an unambiguous mapping with the SOC 4-digit code could be made only some of the times, hence the 4-digit ISCO code variable jbisco sometimes has a real value, which is always 1 to 9, and sometimes has a zero, denoting inability to map with SOC at the fourth-digit level. Hence, for those 3-digit groups, the jbisco variable values ending with zeroes are pseudo-values, which cannot be immediately used as link variables to merge with other databases, because they do not really exist in the ISCO classification system. A way out, if reasonable in the specific context, is to create in addition pseudo-codes ending with zeroes for those group in the other dataset, averaging values of other variables within those groups.

Piotr, please let me know if all of this sounds correct on your side.

Best
Matteo

#8 - 02/02/2023 09:16 AM - Matteo Pinna Pintor

Finally, let me signal that one of the jbisco pseudo-codes 2470 ("Public service administrative professionals") does not even have a real 3-digit group in the ISCO88 classification. I was only able to detect both the 3-digit and 4-digit codes in the European version of the ISCO88 classification, ISCO88 (COM) - see here: <https://ec.europa.eu/eurostat/documents/1978984/6037342/ISCO-88-COM.pdf>. The label comes either from ISCO88 (COM) or from the SOC classification, group 2441. In this case, for merging purposes, one might re-code the jbisco value to some other value in the ISCO88 system that has a similar description such as 3431 ("Administrative secretaries and related associate professionals") or 3439 ("Administrative associate professionals not elsewhere classified") - noting that the latter is commented in the ISCO88-08 crosswalk manual specifying "executive secretaries to government department heads and official committees" (so relevant for "public service").

#9 - 10/26/2023 03:14 PM - Piotr Marzec

- Category changed from *Special license* to *Data inconsistency*
- % Done changed from 50 to 90

Hi Matteo,

I'm really sorry we missed your follow-up. Yes, your description of our coding procedure sounds about right. Also, the merging procedure you're proposing is sensible. Thank you for highlighting the issue with the code 2470, we will have a look at it.

Best wishes,
Piotr Marzec,
UKHLS User Support

#10 - 11/30/2023 12:46 PM - Understanding Society User Support Team

- Status changed from *Feedback* to *Resolved*
- % Done changed from 90 to 100

Files

fake4digitsisco88.JPG	151 KB	01/16/2023	Matteo Pinna Pintor
fake4digitsisco88_2.JPG	189 KB	01/26/2023	Matteo Pinna Pintor