

New Occupation and Industry Variables

Updated Variables: This release updates the following variables for all years 1972-2010: OCC10, INDUS10, SPOCC10, SPIND10, PAOCC10, PAIND10, MAOCC10 and MAIND10. These variables were previously only available for 2012 and 2014. This release also adds KDOCC10, KDIND10, SBOCC10, SBIND10, SPHEDOCC10 and SPHEDIND10. The verbatims for these variables were originally collected in 1994, and previously had only been coded according to the 1980 occupation and industry codes.

This release also updates occupational prestige and SEI variables to include all years 1972-2010: PRESTG10, PRESTG105PLUS, SPPRES10, SPPRES105PLUS, PAPRES, PAPRES105PLUS, MAPRES10, and MAPRES105PLUS. These variables were previously only available for 2012 and 2014. Values for these variables are assigned based on the new 2010 occupation codes. Similarly, this release adds KDPRES10, KDPRES105PLUS, SBPRES10, SBPRES105PLUS, SPHEDPRES10, and SPHEDPRES105PLUS. Please see Analysis below for instructions for working with the updated variables.

Terms for definition: When discussing the new variables below, references are occasionally made to categories of variables. Occupation and industry data were collected on seven different categories of people, over the lifetime of the General Social Survey. Starting in 1972, respondents were asked about their own occupation and industry, as well as that of their father/father substitute, and their spouse, if they were currently married. In 1994, questions about mother's occupation and industry were added, and continue to be asked going forwards. Also in 1994, respondents were asked about a random child of theirs (if applicable), a random sibling of theirs (if applicable), and their spouse's father or father substitute (if applicable). These seven types of people (Respondent, Spouse, Father, Mother, Kid, Sibling, Spouse's Head of Household) form the seven categories of occupation and industry codes collected. Each of these categories has a V code (e.g. OCCINDV) that applies to both occupation and industry, two Status codes (e.g. SPOCCSTATUS, SPINDSTATUS), and two Tag codes (e.g. MAOCCTAG, MAINDTAG). These codes are described below.

FOUND: Found is a case-level variable that indicates that at least one verbatim was recovered for the case, across the seven possible categories of verbatims. If none of these categories contained any verbatims, the case was designated as not found, and coded as 0. If any of these categories contained any verbatims, the case was designated as found, and the verbatims were assessed and used to generate the new codings using the 2010 Census occupation codes and 2007 North American Industrial Classification System codes. Starting in 2000, the GSS maintains digital files containing all verbatims collected from cases, so all cases from 2000 onwards are coded as found. Prior to 2000 original hardcopy questionnaires needed to be located and verbatims data entered.

OCCINDV, SPOCCINDV, PAOCCINDV, MAOCCINDV, KDOCCINDV, SBOCCINDV, SPHEDOCCINDV:

Indicates that a verbatim exists for this particular instance. There is a separate V code for each category of occupation and industry verbatim, designating the presence or absence of verbatims for respondent, father, mother etc. for a particular case. If the case is not found, all of the V codes are assigned -1. Otherwise, 0 indicates the absence of any verbatim for that category, and 1 indicates the presence of a verbatim (e.g. if OCCINDV=0 & PAOCCINDV=1 for a case, there are no verbatims available for the respondent, but there are for the father). Please note that the absence of a verbatim does not indicate that a case should be problematic - IAP values rarely have verbatims because they were never asked the occupation and industry questions. From 2000 onwards, all cases are considered found, and thus will only have 0 or 1 V codes. The V code is used to determine STATUS variables (see below).

OCCSTATUS, INDSTATUS, SPOCCSTATUS, SPINDSTATUS, PAOCCSTATUS, PAINDSTATUS, MAOCCSTATUS, MAINDSTATUS, KDOCCSTATUS, KDINDSTATUS, SBOCCSTATUS, SBINDSTATUS, SPHEDOCSTATUS, SPHEDINDSTATUS: STATUS variables have seven values, depending on the original occupation or industry coding, the V code for the case, and the 2007/2010 coding (i.e. 2007 industry and 2010 occupation codes). The first three codes are for found cases, while the last four are for cases that were not found. Please note that from 2000 onwards, STATUS variables codes 4-7 are not used, because all cases are considered found. The STATUS variables are used to determine TAG variables.

- 1) Found, IAP: Both the original occupation and the original industry were coded as inapplicable (IAP). Examples, would be for respondents who never worked or when respondents did not have a current spouse.
- 2) Found, No Answer: Either both the old code and the 2007/2010 update are coded as No Answer, or the 2007/2010 update is coded as No Answer. This typically occurs when the recovered verbatim was uncodeable or made no sense.
- 3) Found, Coded: The verbatim was recovered, and the case has a valid code (i.e. non-missing code) for the 2007/2010 update, and either a valid code or a no answer code for the original occupation or industry.
- 4) Not found, IAP: The original verbatim was not recovered for any variable (respondent, father, spouse, or mother) and the original code was IAP for the variable. The new code is assigned IAP.
- 5) Not found, no answer: The original verbatim was not recovered for any variable, and the original code was No Answer for the variable. The new code is assigned No answer.
- 6) Not found, coded: The original verbatim was not recovered for any variable, and the original code was valid. These cases were determined to have a strong match to a particular 2007/2010 code, and were directly assigned that code.
- 7) Not found, no code assigned: The original verbatim was not recovered for any variable, and the original code was valid. These cases did not have a strong match to a particular 2007/2010 code, and were thus assigned a 9997 code, indicating that no 2007/2010 code could be assigned. 9997 is declared to be a missing value.

OCCTAG, INDTAG, SPOCCTAG, SPINDTAG, PAOCCTAG, PAINDTAG, MAOCCTAG, MAINDTAG, KDOCCTAG, KDINDTAG, SBOCCTAG, SBINDTAG, SPHEDOCCTAG, SPHEDINDTAG: TAG variables have three values, depending on the 2007/2010 variable coding of the occupation or industry. All found cases are coded as 1, while the not found cases are divided into those that were recoded, and those that could not be recoded. If a case was not found and was assigned an IAP, no answer, or valid code under the 2010 coding, TAG variable=2. For the cases there were not found and that could not be assigned a valid 2007/2010 code, TAG variables=3.

Analysis

The default use of the new occupation and industry variables (i.e. OCC10, INDUS10, SPOCC10, SPIND10, PAOCC10, PAIND10, MAOCC10, MAIND10, KDOCC10, KDIND10, SBOCC10, SBIND10, SPHEDOC10, SPHEDIND10) includes all cases with values directly coded from verbatim and those for which codes were assigned (i.e. the TAG variables were coded 1 or 2). For example, if one used the default for the new occupational codes for respondents (OCC10), one would have valid codes based both on found cases directly coded from verbatims and on not found cases assigned by recodes of the original valid

codes. The missing values for all occupation and industry variables consist of IAP (0), Not Assignable (9997), Don't Know (9998), and No Answer (9999). If one used the default for the new occupation or industry codes, Not Assignable (9997) would not appear in any responses, as that missing value is excluded when TAG variable=3 is not used.

If one wishes to exclude the valid 2007/2010 codes that were assigned and not coded based on verbatims, one needs to exclude cases coded 6 on the appropriate STATUS variable. For example, to look at only the new occupational codes for respondents based on verbatims, one would use OCC10 and exclude cases in which OCCSTATUS equals 6.