

Converting restricted-use SESTAT 1993-2013 and public SESTAT 1993-2013 from SAS to Stata format

The purpose of this project is to provide users of SESTAT data with Stata DO files that convert the data in SAS transport format to Stata data datasets having all variables (other than refid) in numeric format and with their values labeled according to each year's label definitions.

Description

The folder Do_files in the compressed folders EAGER_NCSES_SAS_to_Stata_SESTAT_Public_1993_2013 and EAGER_NCSES_SAS_to_Stata_SESTAT_Restricted_1993_2013 contain Stata do files that, for each survey year, recode text variables to numeric and labels the values of each variable according to the format definitions provided in each data year's source materials. Other than recoding text characters to numeric, the content of each variable is not modified. For example, variables are not harmonized to have the same possible values for each variable. That is, if a variable has a different number of categories in different years, these are left as is. Rather, the numeric values assigned to recoded variables were defined making sure that they would have the same meaning in all survey years that they appear.

Instructions for how to use the code

The steps for using the code that converts yearly public-use SESTAT files into Stata format are in the file **Instructions_for_SAS_to_Stata_SESTAT_Public_1993_2013.docx**. Similarly, the steps for the restricted-use SESTAT data are in **Instructions_for_SAS_to_Stata_SESTAT_Restricted_1993_2013.docx**. Note: only Stata version 10 or older is required for the code to work. It may work with versions 8 and 9, but we have not tested it with these versions.

IPF-IPEDS-FICE-DUNS crosswalks for institutions of higher education

We constructed two crosswalks intended to be used with NIH ExPORTER data and with the NSF's Higher Education Research and Development Survey (HERD). The first crosswalk is meant to be used with ExPORTER data only, and the second one is a standardized version of the first, and can be linked to other data.

IPF-IPEDS-FICE-DUNS crosswalks for institutions of higher education

The file UniversityIPF_forWeb.zip contains Stata and Excel versions of a crosswalk that assigns to each organization name their corresponding NIH identifier (IPF code), IPEDS code (UNITID), FICE code, and Dun & Bradstreet (DUNS) number. Ideally, we would have used one of the institution identifiers as matching variable. There was no public-use crosswalk that matched NIH IPF codes to IPEDS codes which are used in higher education research. As a result, we matched using institution names. Moreover, the spelling of institution names as recorded in these data has changed over time, and in many cases it is not possible to assign IPF codes to projects from previous years based on institution names. This crosswalk is only valid for IPF codes from 2014 to the present.

The crosswalk also includes columns with indicators for name changes and unknown or missing DUNS number, as well as columns for previous known names of institutions whose name has changed over time (e.g. 'Columbia College' to 'Columbia University', 'The Agricultural and Mechanical College of Texas' to 'Texas A&M University', etc.), for the year when their name changed, and for the year when institutions merged or became affiliated with another institution, whenever this applies.

Standardized IPF-IPEDS-FICE-DUNS crosswalk for institutions of higher education

The file Standardized_UniversityIPF_forWeb.zip contains a standardized version of the above crosswalk. The matching variables are the institution's IPEDS code (UNITID) and NIH identifier (IPFCODE). For each combination of these, we kept the most recently used institution name as recorded in ExPORTER (NIH_ORG_NAMES) and IPEDS (INSTNM), as well as the most recent FICE code found in IPEDS data. If a user would like a crosswalk that is unique by UNITID, they can obtain one by dropping all variables except for UNITID, INSTNM and FICE, and then drop duplicate observations.

NIH funding rank

The file NIH_Funding_Rank_1985_2017.zip contains higher education institution rankings by year, ordered according to their NIH funding as reported in ExPORTER, and aggregated at the UNITID and IPFCODE identifiers level.

Higher Education Research and Development Survey (HERD) Funding Rank

The file HERD_Funding_Rank_forWeb.zip contains yearly rankings of institutions of higher education from 1990 to 2015, constructed using the NSF's Higher Education Research and Development Survey (HERD) (<https://www.nsf.gov/statistics/srvyherd/>). The Excel file HERD_Funding_Rank_forWeb.xlsx has rankings based on total funding on sheet **All Fields**, and rankings for broad fields of degree in the rest of the sheets. The Stata file HERD_Funding_Rank_Agg_forWeb has the same information as in the **All Fields** Excel sheet, and HERD_Funding_Rank_byField_forWeb has the rankings by field for all fields.