Cleaning Instructions

Gavin Gordon

2022-09-22

For questionnaires using an XLS Form

Phase One

Create the cleaning meta files and initial formatting and documentation

1. create the folder for cleaning that one dataset.
2. download the dataset into downloads folder, download the xls form into the meta files folder
3. use the create value labels and create meta data scripts to curate the meta data
4. import and do the initial prep on the data using the cleaning template and store data frame.

Phase Two

Select the variables to clean

use the create skeleton script to curate the code

clean those variables

store in cleaned datafiles

Phase Three

Create the codebook and review

Meta Data File

Perhaps the most important document.

name the labels with label at the end to avoid naming conflicts

Create line by line code from the data frame for changing each of the cell contents in a high level manner in a separate file to control for length

Create an initial document, from which we will copy and paste over named vectors to serve as a log for the connections with variable names with the labels, and other meta data. It is assumed that not much change will be made to this file in terms of the addition of new variables.

var_name : nothing special

var_label

brief_desc

var_type: will become important soon enough interval type: must either be continous or discrete

interval_type: discrete or continuous

possible_invalid_codes" : write all of what to check for in a string separated by commas

invalid_codes : do not touch this the code will fill in

new_page : this is to move the entire variable info over to the next page; will likely have to change to allow different tables to move

table_break : split the value table at the rows indicated in the string separated by commas

You have to handle dates in a special way

These have to be character types

which will have value labels where each valid date will have a "valid date value. Otherwise it should be the same. get all the unique valid date entries and attach valid date as the value label

For date variables you have to pass in the date format

An indicator if the time information is relevant such as time zone

Always create the instrument in kobo or something; the person will have a nice copy; you will have access to the value codes