Assignment 04: Portfolio 1 – Partial replication of Pescosolido et al. 2010

Quantitative Methods for the Social Sciences | Winter 2022/23

Complete the following tasks and generate a do-file. You have to upload the **do-file** to ILIAS. Please include your solution to **all** tasks in this do-file.

This assignment is **graded**. You can work on your own or in teams of two. In case of teamwork, both students will be awarded the same number of points.

Submitted do-files will be cross-checked for plagiarism.

ILIAS **submission deadline:** Nov 13, 2022, 10:00 CET.

Do not miss the deadline. Start your submission well ahead of the deadline. No points will be awarded for late submissions.

General instructions

In this assignment, you turn your do-file prepared in the previous week into a partial replication of the results reported in Pescosolido et al. 2010. Your will extend this do-file and revise it so that it meets the standards for data work introduced in this week's lecture. Your do-file from the previous week might already meet some of these standards.

Work in the same folder that you have used for Assignment 3

Copy your do-file from last week and rename it into **portfolio1_YOURLASTNAME**. In case of teamwork, enter both last names separated by an underscore (_).

Students under the **new PO** complete **all tasks**, for a total of 20 points.

Students under the **old PO** complete **all tasks except T3 and T6**, for a total of 15 points. If you study under the old PO, add one line stating **"* OLD PO"** at the top of your do-file.

T1 Write a header at the top of your do-file. In this header, give a brief step-by-step summary of what your do-file does. Write these instructions so that another researcher could understand (a) what the purpose of your do-file is, (b) how your do-file is organized, and (c) how to run it (e.g., how to customize the working directory). Add a table of contents to structure your do-file. Ensure that all globals are defined correctly after customizing the path to the working directory. **1 point.**

T2 In a brief comment, describe most important findings of the study by Pescosolido et al. 2010. What has changed regarding the neurobiological understanding? What has changed regarding public stigma? **2 points.**

T3 In the previous week's assignment, almost all variables were prepared for analysis. One variable that is still missing is Neurobiological Conception. `generate` a new variable called NEURO for Neurobiological Conception. To do this, follow the note below Table 1 in the article. `tabulate` NEURO showing the missings. Note: If you do not manage to create the variable NEURO, continue without it. **2 points.**

T4 Revise your code on all data transformations and selections done so far to meet the standards of data work introduced in this week's lecture. Note: If your existing code already meets the standards, you can leave it as it is. The following instructions provide guidance for your revision. **6 points.**

- Annotate every step so that a reader can understand what you are doing and why. Start every step with the annotation and report the original coding (as a comment) before your code.
- Separate data transformation (recoding and defining new variables) from data selection (keeping or dropping cases) in your do-file. Do not switch between the two. Do the data transformation (i.e., changing and creating variables) before the data selection (i.e., keeping and dropping cases). Use comments to track how many cases are lost due to each selection (i.e., sample exclusions).
- Remember to keep all original variables in their original scaling and to recode them into new variables.
- Minimize the code needed for recoding your variables (e.g., combine identical recodes, use loops, etc.). If necessary, improve the layout/formatting of your do-file so that errors are easy to spot.
- Ensure that your do-file uses consistent formatting, labeling, and naming.
- Label all variables and values, keep only the variables needed for the analysis, and finish the data preparation by posting the prepared data in the folder *2_posted*.

T5 Add a new section called "Analysis" to your do-file. Start by loading the file that you saved in the folder *2_posted*. Then replicate the %1996 and %2006 columns of Table 1 in Pescosolido et al. 2010. Draw on the prepared data and export a text-format table (.doc or .rtf) that creates the table you need. The best-practice standard is exporting a table into your subfolder *3_tables* that requires no further adjustments. If you do not manage to do this, try to get as close as you can. The following instructions provide guidance. Note: Students under the old PO (9 ECTS) do not have to replicate statistics on the variable "Neurobiological Conception." **6 points.**

- To solve this task fully, you likely need to access knowledge from the web, in particular resources such as Statalist. You will do this regularly as you learn to work with Stata independently.

- Use the "estout" package (an "ado" file) written by Ben Jann. To install it, type `ssc install estout` in your Stata command window. Then check `help estout` to get an overview of what this package allows you to do. This ado includes the `esttab` command that does what you need.
- Look at a Statalist post by clicking on the link below. The post is about a different analysis but you can adapt the piece under "I have code like this:" to solve your task. Try to understand the logic of the code, adapt it to your needs, and solve your task: https://www.statalist.org/forums/forum/general-stata-discussion/general/1395253-descriptive-statistics-table-generation
- You are not required to use weights in order to gain full points. But if you want to replicate the exact same results, you need to use the weights (the weights are in the WTSSALL variable) (use `help weights`). You can add analytical weights using [aw=WTSSALL]. For an example, see `help weights`.

T6 Conclude your do-file with a reflection comment in which you give answers to the following questions.

(1) The authors use samples from the General Social Survey collected in 1996 and in 2006. These samples contain only a few thousand respondents and some questions about mental illnesses were answered only by a few hundred respondents. Based on these samples, is it possible to say something about the hundreds of millions of people comprising the US adult population in 1996 and in 2006? Explain why (or why not). **1 point.**

(2) Pescosolido et al. 2010 did not take all the measures in their original form but applied various recodes before analyzing the data. These recodes are decisions taken by the authors that may affect the results. Reflect critically on the recoding schemes used by Pescosolido et al. 2010. Do you agree with their decisions? Is there something you would do differently? Explain why you agree and/or why you disagree with the authors' decisions. Note that you can also explain why you agree with some decisions and disagree with others. **2 points.**