Shabnoor Nabi

September 20th, 2024

Week 1 – RRWM, CAnD3 Replication Exercise

**RESEARCHER PROGRAM FOR REPLICATION**

**Study Overview:**

This study employs multivariate linear regression in R to examine the association between place of birth and individuals' self-rated mental health, based on data from the Canadian General Social Survey (GSS 2017). Factors like individuals’ age, gender, and marital status are treated as covariates in the multivariate analysis.

**Replication/ Analysis Design Plan (short):**

1. Import GSS 2017 data
2. Clean and recode the GSS 2017 data file for analysis
3. Subset data frame with main variables of interest
4. Create table of summary statistics
5. Run a series of bivariate and multivariate regression analysis and chart the associations

**Replication Program (extended): “Nabi\_RRWM.Rmd”**

1. Import GSS data
   1. On your computer, **UNZIP** the folder with data (GSS 2017 – Family, Cycle 31.zip)
   2. Set working directory in your R Markdown file to the path where the unzipped data folder is saved
   3. Load necessary R packages
      1. Researcher recommended: tidyverse, rio, dplyr, gtsummary, and webshot2.
   4. Import the GSS data file (gss-12M0025-E-2017-c-31\_F1.csv)
2. Clean and recode data for analysis
   1. To prepare the data for analyses, first turn the values from 6-9 where applicable (where, 6 = Valid skip, 7 = Don't know, 8 = Refusal, and 9 = Not stated) within all variables of interest into NAs and rename the variables for analytical ease.
      1. List of variables of interest: SRH\_115 (self-rated mental health), BRTHCAN (place of birth: Canada vs. outside of Canada), MARSTAT (marital status), AGEC (age at the time of interview), SEX (sex of respondent: male/female)
   2. Also, turn the following categorical variables into binary/dummy and classify them as factor.
      1. Set BRTHCAN as dummy for birth outside Canada and rename it as “immigrant”
      2. Set MARSTAT as dummy for “marriage” and/or “cohabitation” and rename it as “marr\_cohab”
      3. Set SEX as dummy for females and rename it as “female”
   3. Create a new variable called “SR\_MH\_R” by reverse coding SRH\_115 so that lower values represent poor levels of self-reported mental health (e.g., 1 = poor mental health) and higher values represent higher levels of self-reported mental health (e.g., 5 = excellent).
3. **Subset data frame with main variables of interest to have the main analytic sample**
   1. Sub-set the data frame to variable of interest with missing values (NAs) omitted across all
4. **Create table of summary statistics**
   1. Calculate the mean, standard deviation, and sample counts and percentages for each variable of interest: *AGEC, female, SR\_MH\_R, marr\_cohab, and immigrant.*
   2. Stratify the table by outcome (i.e., *SR\_MH\_R*)
   3. Import and save results as output table in .png/jpeg format
5. **Run a series of bivariate and multivariate regression analysis and chart the associations**
   1. Run the following series of linear bivariate regression model without any controls (naïve analysis) with *SR\_MH\_R* as outcome and I.V. as:
      1. *immigrant*
      2. *female*
      3. *marr\_cohab*
      4. *AGEC*
   2. For each of the four explanatory variables, *compute proportions table grouped by outcome (SR\_MH\_R), plot the results with proportions, and save the output as .png*
      1. *immigrant [bar chart]*
      2. *female [bar chart]*
      3. *marr\_cohab [bar chart]*
      4. *AGEC [box plot]*
   3. Now, run a full multivariate linear regression model that regresses main outcome self-rated mental health (SR\_MH\_R) on main explanatory variable (immigrant) while holding constant: marr\_cohab, female, and AGEC
   4. Format and import results of the multivariate analysis in an output table

**Data availability and accessibility statement:**

Data used for the present project were obtained through ODESI, a service provided by the Ontario Council of University Libraries ([https://search1.odesi.ca/#/Links to an external site.](https://search1.odesi.ca/#/)).

Access is restricted to those users who have a DLI License and can be used for statistical and research purposes. The terms of the license can be viewed here ([https://www.statcan.gc.ca/eng/dli/licenceLinks to an external site.](https://www.statcan.gc.ca/eng/dli/licence)).

As part of McGill University, the CAnD3 initiative has a license to use the data for the purposes of training. Those outside of McGill university **should not use** the data provided through CAnD3's training activities for purposes not related to the CAnD3 training.

GSS data citation:

Statistics Canada. 2020. General Social Survey, Cycle 31, 2017 [Canada]: Family (version 2020-09). Statistics Canada [producer and distributor], accessed September 10, 2021. ID: gss-12M0025-E-2017-c-31