CAnD3 Validation and Replication Report
Allison Laing
October 5, 2021

Author: Matthew Stackhouse
Project: The impact of parental bereavement in childhood on self-rated mental health
Validation and Replication results

1. Data description

a. Data Sources

From: Statistics Canada, Social and Aboriginal Division.
Data: General Social Survey, Cycle 31, 2017 [Canada]: Family
To download full dataset and codebook:
https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4501

b. Analysis Data Files

Replication package included Stata dta files containing data subset and recoded dataset:

- 2017_GSS_Recode.dta - 95.2 KB [download]
- 2017_GSS_Data.dta - 60.7 MB [download]

2. Code description

- There are two provided Stata do files, including a "Recode.do" and “Analysis.do”.
  - Program lacked specific category cut points for outcome and predictor variables, but these were apparent in Recode.do file
  - Program lacked clear instruction for sampling and variable weights (*or more likely, this reflects my lack of familiarity*), but again, this was apparent in Recode.do file
  - Table 1. Program did not specify the reporting parameters for Table one (I assumed percentages, 2 decimal places for SE and 95% CI), Recode.do applied default settings
  - Table 2. Program was unclear that two nested logistic regressions and two tables were required (i.e. logistic regression for each predictor variable with bivariate and multivariable regressions as Model 1 and 2)

NOTE: my quant and Stata skills are rusty, which informed my difficulties with this exercise and may not accurately reflect the clarity of Matt’s program.

3. **Stated Requirements**

   Software Requirements specified as follows:

   - Stata 15

4. **Missing Requirements**

   Computational Requirements not specified:

   - Cluster size, disk size, memory size, etc.

   Time Requirements not specified:

   - Length of necessary computation (hours, weeks, etc.)

5. **Computing Environment of the Replicator**

   Mac Laptop, MacOS Big Sur Version 11.5.2, 8 GB of memory
   Stata/BE171

6. **Replication steps**

1. Downloaded program and code through Canvas link provided.
2. Downloaded data through Canvas link provided.
3. Following the Program without looking at the do.files, I attempted to replicate Matt's summary and regression tables.
4. Recoded continuous Predictor Variables (age father died, age mother died) into categorical variables (4 categories: 1 "Childhood 0-18" 2 "Young adult 19-29" 3 "Adulthood 30-49" 4 "Older adult ≥50") using generate command, dropped missing
5. Recoded ordinal Outcome Variable into dummy (poor /fair health vs excellent /v. good /good health) using generate command, dropped missing
6. Recoded covariate marital status `marstat' into 1 "Married/Common-law" 2 "Previously married" and 3 "Single/Never Married" using generate command, dropped missing
7. Labelled covariates, dropped missing from education.
8. Generated a summary statistics table (Table 1) using proportion command, showed percentages, 2 decimal places
1. Renamed variables b/c could not apply variable labels to the table (?)
9. Conducted bivariate logistic regressions for each predictor (age father died, age mother died)
10. Conducted multivariable logistic regression using all covariates (Table 2)
    a. Of course, this didn't make sense, to use a second predictor as a covariate/ control
    b. After looking at the tables Matt generated, I conducted separate multivariable logistic regressions for each predictor (age father died, age mother died), generating nested models: bivariate regression as Model 1, adding in all covariates for Model 2

7. Findings

After adjusting for sex, age, marital status, education and income, parental bereavement was not associated with self-reported poor mental health.

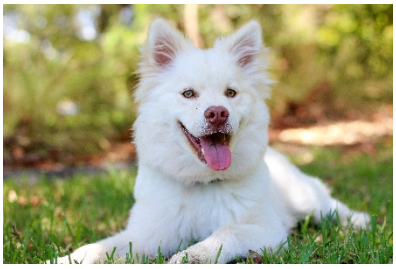
a. Data Preparation Code

- 2017_GSS_Data.dta ran without error, output expected data
- 2017_GSS_Recode.dta ran without error, output expected data

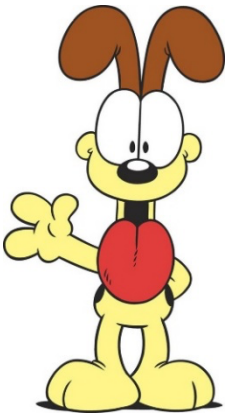
b. Tables

- Table 1:
  - due to sampling weight issue (and differences in coding?), my observation totals were off
  - cut points for categories for predictors (age father died, age mother died) were different
- Table 2:
  - Due to survey weight and above issues, my results were substantively similar (no significant associations), but the tables were as different as … these dogs.

Matt's table:



My table:



8. Classification

Failure to reproduce: only a small number of programs ran successfully, or only a small number of numbers were successfully generated (<25%)