Validation and replication results report exercise

**Data description**

**Data Sources**

* Data source: General Social Survey (2017): Family (version 2020-09). From Statistics Canada.
* Access not described but was restricted to users with a DLI License.

**Analysis Data Files**

* Analysis datafiles mentioned but not provided due to restricted access.
  + File name: gss-12M0025-E-2017-c-31\_F1.csv

**Code description**

* Output tables were provided directly with the code in pdf format.

**Stated Requirements**

* No requirements specified



**Missing Requirements**

* Software Requirements



* + R



* + - Version 4.1.1



* + - R packages:
      * haven
      * psych
      * tidyverse
      * lmtest
* Computational Requirements not specified



* Time Requirements not specified



* + Took 1-3 minutes to run

**Computing Environment of the Replicator**

* Original computing environment not specified
* Replicator computing environment:
  + XPS 8940 Dell Desktop computer
  + Processor 11th Gen Intel® Core™ i9-11900 @ 2.50GHz
  + Installed RAM 64.0 GB (63.7 GB usable)
  + System type: 64-bit operating system, x64-based processor
  + R Studio: R version 4.1.1

**Replication steps**

1. Downloaded files from OneDrive.
2. Downloaded code from URL
3. Set working directory
4. Loaded data in RStudio
5. Ran code as per program instructions.

**Findings**

* SRH\_110, RLR\_110, REE\_02, STS\_410 were all significant predictors of SLM\_01 in model 1.
* SRH\_1101 and RLR\_1101 were significant predictors of SLM\_01 in model 2.

**Data Preparation Code**

* Program 1- was readable however regression outputs differed from original output. The original output had different slopes and significant predictors than the replication.
* Looking at the original code, the researcher had created dummy variables to create binary variables that were added as new variables to their dataset. Unfortunately, they ran their models with the original variable names, not the new names which was not consistent with the program they wrote. This seems to be the cause of lack of replicability.

**Tables**

* Table 1 and 2 look completely different

**Classification**

* not able to reproduce most or all of the results (reasons see above)

