Instructions for Reproducing This Analysis

Mohsen Monji

2024-09-16

Introduction

This document provides an overview of the steps required to reproduce the analysis of disparities in self-rated mental health among young adults (aged 18-39) in Canada using the 2017 General Social Survey-Public Microdata File. The analysis is organized into three stages using three R scripts:

1. `Data Wrangling`: Cleaning and recoding variables.
2. `Unweighted Descriptive Statistics`: Producing frequencies and proportions of self-rated mental health by various socio-demographic variables among young adults (aged 18-39).
3. `Chi-Square and Logistic Regression`: Analyzing associations between self-rated mental health and predictors using chi-square tests and logistic regression.

Please ensure that you have downloaded the GSS-2017 dataset and placed it in the main directory of the cloned repository. You can also download the Zip file in this repository, unzip it and place the data you downloaded in the same directory.

Step 1: Clone the GitHub Repository

Clone the repository by using the following command in your terminal:

```
git clone https://github.com/Mohsnmonji/Reproducible-Research_GSS2017.git
```

This will download the repository to your local machine.

Step 2: Download the GSS 2017 Dataset

The GSS-2017 dataset is not included in the repository due to licensing restrictions. Download it from ODESI at [https://odesi.ca/] and follow the steps below:

1. Unzip the file and rename it to `gss2017.csv`.
2. Move the `gss2017.csv` file to the main directory of the cloned repository or where the R scripts are located.

Step 3: Running R Scripts

Once the repository is cloned and the downloaded dataset is in the correct directory, follow the steps below to reproduce the analysis:

3.1 Data Wrangling

First, run the `Code for Data Wrangling.R` script. It will recode the variables of interest, drop cases with missing values, and output a cleaned dataset with only the selected variables named (`gss_2017_selected.csv`), which is used in the subsequent analysis. The `gss_2017_selected.csv` will be saved in the main directory of the cloned repository where the R scripts are located.

3.2 Descriptive Statistics

Second, run the `Code for Descriptive Statistics.R` script. It will first filter the saved dataset from the previous step to include only young adults (aged 18-39) and then calculate the frequencies and proportions of self-rated mental health by various predictors such as sex, marital status, household income, etc.

3.3 Chi-Square and Logistic Regression

Third, run the `Code for Logistic Regression.R` script. It will perform chi-square tests for associations between self-rated mental health and the categorical predictors. It will also fit a logistic regression model to further explore these associations adjusting for confounders. The script will display the results of the chi-square tests and the logistic regression model, including odds ratios and confidence intervals.

Conclusion

By following the steps outlined above, you will be able to reproduce the analysis of disparities in self-rated mental health among young adults (aged 18-39) in Canada using the 2017 General Social Survey data. Ensure the dataset is placed in the correct directory, and the scripts will guide you through the process.

If you encounter any issues, feel free to reach out for assistance!