

Shell File: CAnD3 Data Activity

Tyler Bruefach

9/22/2021

Reproducing the Analyses

Cleaning Data

First, run the script named “**Cleaning Data.R**” to complete the following tasks:

1. Load required packages for analyses.
2. Converting data from the GSS into a tibble.
3. Cleaning and coding variables to be used in the analyses.
4. Dropping all variables but the ones used in analyses.
5. Labeling those variables.

Handling Missing Data

Next, you should run the script “**Handing Missing Data.R**”. This script completes the following tasks to handle missing data using listwise deletion:

1. Recodes missing values of self-rated health and self-rated mental health (other measures were assigned missing values during the cleaning phase).
2. Creates an index called “samppmiss” that is a count of how many variables that each respondent has missing values.
3. Creates a dataset called “sample” that only contains cases with no missing values.

Producing the Descriptive Table (Table 1)

The sample is now ready for producing descriptive and inferential statistics. Run “**Table 1.R**” to create a descriptive table. This table provides descriptive statistics across respondents’ sex.

TABLE 1. DESCRIPTIVE STATISTICS (N = 19755)

Variables	Male (8991) ¹	Female(10764) ¹
Self-Rated Health		
1	0.21	0.22

¹ Mean and standard deviations provided for continuous variables. Proportions provided for categorical variables

Variables	Male (8991)¹	Female(10764)¹
2	0.34	0.35
3	0.30	0.30
4	0.10	0.10
5	0.4	0.4
Self-Rated Mental Health		
1	0.32	0.28
2	0.33	0.35
3	0.28	0.29
4	0.6	0.7
5	0.2	0.2
Educational Attainment		
1	0.41	0.37
2	0.32	0.35
3	0.27	0.28
Household Income	3.72 (1.73)	3.46 (1.75)
Age	50.82 (17.67)	52.54 (17.75)
Marital Status	0.60	0.53
Number of Children in Household	0.45 (0.86)	0.51 (0.90)
Racial Minority Status	0.13	0.12

¹ Mean and standard deviations provided for continuous variables. Proportions provided for categorical variables

We find that there are some slight differences in the distribution of mental health and education across people identifying as male and female. Male and female respondents had nearly identical distributions of self-rated health. Although a greater proportion of female respondents had more than a secondary degree than male respondents, they also had lower household income.

Producing the Regression Table (Table 2)

Finally, you should run “**Table 2.R**” to create a table featuring two multivariable regressions. This table provides descriptive statistics across respondents’ sex. Findings suggest that education is associated with better health, but that female respondents gain a greater health-benefit from earning a postsecondary degree than male respondents:

TABLE 2. SEX MODERATES THE HEALTH-BENEFITS OF POSTSECONDARY EDUCATIONAL ATTAINMENT, NET OF CONTROLS (N = 19755)

Variables	Self-Rated Health		Self-Rated Mental Health	
	<i>b</i> ¹	<i>se</i> ²	<i>b</i> ¹	<i>se</i> ²
Educational Attainment				
1	—	—	—	—
2	-0.12***	0.024	-0.10***	0.023
3	-0.32***	0.025	-0.20***	0.025
Female	0.00	0.022	0.12***	0.022
Age	0.01***	0.000	0.00	0.000
Marital Status	-0.07***	0.017	-0.11***	0.017
Household Income	-0.07***	0.004	-0.04***	0.004
Number of Children in Household	-0.01	0.008	0.03***	0.008
Racial Minority Status	0.15***	0.018	-0.05*	0.018
Educational Attainment * Female				
2 * Female	-0.07*	0.033	-0.08*	0.032
3 * Female	0.02	0.035	-0.01	0.034

¹ *p<0.05; **p<0.01; ***p<0.001

² SE = Standard Error