≡ ⌂ AimyCroner / RRWM

<> Code ⊙ Issues ⑂ Pull requests ▷ Actions ⊞ Projects ⚠ Security ⪤ Insights

RRWM / **Program Template.md** ⧉

👤 **AimyCroner** Create Program Template.md

018b3e8 · 2 weeks ago

111 lines (77 loc) · 5.1 KB

Preview | Code | Blame

Raw ⧉ ⭳ ✎ ▾ ☰

Program Template

1. Cleaning the GSS 2017: "CAnD3 RRWM 1.do"

a. Import the GSS.csv 2017 data set

b. Create a dummy variable for physical health of the respondent (srh_110)

```
• The categories should be good or poor health

• Recode the missing values (7,8,9) as "."

• For simpler interpretation, label the categories as "Good health" and
"Poor health"
```

c. Create a categorical variable for age (agec)

```
• Drop respondents aged below 20 and those older than 65. The sample should
include respondents aged 20 to 65.
• Recode missing values as "." Note: According to the codebook, there
should be no missing values.
• Create a new variable named 'age group' and create & label the following
categories:
  ○ 20-29, 30-39, 40-49, 50-59, 60+
```

d. Create a dummy variable for marital status (marstat).

- Recode the missing values (97, 98) as ".".
- Create a new binary variable for marital status and create the following categories
  - Single
  - Married / Living Common Law
  - For simpler interpretation, label the categories

e. Create categorical variables for main activity of the spouse (map_110)

- Recode missing values (96,97,98,99) as ".".
- Create the following categories
  - Working, Unemployed / Job Seeking, Studying, Homemaking/Caregiving, Not in the labour force, Other
  - For simpler interpretation, label the categories

f. Create a dummy variable for spouse's work status (map_130)

- Recode missing values (6,7,8,9) as ".".
- Create the following categories:
  - Paid job or self-employed in last 12 mos
  - Not employed in the last 12 mos
  - For simpler interpretation, label the categories

g. Create a categorical variable for spouse's work schedule (map_155)

- Recode missing (96,97,98,99) as ".".
- Create new categories and label them as follows:
  - Day shift/Standard, Evening/Night shift, Rotating/Irregular shift, Split shift or other

h. Create a dummy variable for respondent's work status (worklyr)

- Recode missing values as ".". Note that, according to the codebook, there should be no missing values.
- Create and label the following categories
  - Worked last yr
  - Did not work last yr

i. Create a dummy variable for sex (sex)

- Recode missing values as "." Note that, according to the codebook, there should be no missing values.
- Create/label the following categories
  - Female
  - Male

j. Create a categorical variable for the respondent's level of education (ehg3_01b)

- Recode the missing values (97,98,99) as "."
- Create and label the following categories
  - Less than a high school degree
  - High school degree or equivalent
  - College or trade certificate
  - University degree or equivalent certificate

k. Create a categorical variable for family income (famincg2)

- Recode the missing values as "." Note that, according to the codebook, there should be no missing values.
- Create and label the following categories
  - Less than $25k
  - $25k-$49,999k
  - $50k-74,999k
  - $75k-99,999k
  - $100k +

l. Create a categorical variable for average hours worked per week (uhw_16gr)

- Recode missing values (7,8,9) as "."
- Note: valid skips (6) here are our unemployed category - worklyr shows us that those who said they did not work in the past year are unemployed (7087) - this same number is present in our uhw_16gr variable, where Valid skips are 7087 respondents, which are our unemployed category. Labelling them as unemployed rather than missing
- Create and label the following categories
  - 0 hrs
  - Less than 30 hrs
  - 30-40 hrs
  - 40.1-50 hrs
  - "50.1+ hrs
  - Unemployed

2. Create the following tables:

a. Summary Descriptives Table: Using the summarize function, summarize all variables that were cleaned up in stage 1

b. Table 1: Cross-tab respondents' self-rated physical health by spouse's main activity (row percentages)

c. Table 2: Cross–tab respondents'self-rated physical health by spouse's employment status (row percentages)

d. Table 3: Cross–tab respondents'self-rated physical health by spouse's work schedule (row percentages)

4. Create the following regression on the binary of self-reported physical health using logistic regression

a. Respondents' self-rated Physical health & spouse's main activity (make sure to control for all variables that were cleaned up, except for spouse employment status in the last 12 mos, spouse's work schedule)

b. Physical health and spouse employment status in the last 12 mos (make sure to control for all variables that were cleaned up, except for the spouse's main activity, spouse's work schedule)

c. Physical health and spouse work schedule (make sure to control for all variables that were cleaned up, except for the spouse's main activity, spouse's employment status in the last 12 mos)