CAnD3 - RR Assignment: Program

Research question: Is the tenure status of one’s current dwelling independently associated with self-rated mental health status for older adults? Does the association differ for males and females?

Program:

1. Clean dataset -Select Analytical Sample
2. Recode and Label Variables of Interest
3. Create Table of Summary Statistics
4. Conduct Logistic Regression
5. Create Table of Regression Results
6. **Clean the GSS dataset – Analytic Sample Selection (Cleaning & Selecting Sample.do)**
   1. Drop respondents aged 64 and under [agec]
   2. Drop respondents currently not living in urban population centres [luc\_rst]
      1. Few previous studies have shown that housing and health associations can differ for rural/urban areas – analysis sample only includes urban respondents
7. **Code and Label Variables of Interest (Recoding and Labelling Variables.do)**
   1. Independent Variable – tenure
      1. Create binary dummy variable for tenure [odr\_10]
         * Code ‘Owned’ with value ‘1’ and ‘rented’ with value ‘0’
         * Code alll other answer categories missing ‘ . ‘
   2. Outcome Variable – self-rated mental health
      1. Create dummy for mental health [srh\_115]
         * Code and collapse answer categories ‘excellent’, ‘very good’, and ‘good’ into ‘good’ with value ‘1’
         * Code and collapse answer categories ‘fair’ and ‘poor’ into ‘not good’ with value ‘0’
         * Code all other answer categories missing ‘ . ‘
   3. Stratification Variable - sex
      1. Create dummy for sex [sex] (name new variable sex1 to differentiate between created dummy and existing categorical variable)
         * Code ‘female’ with value ‘1’ and ‘male’ with value ‘0’
         * Code all other answer categories missing ‘ . ‘
   4. Confounder Variables
      1. Continuous variable for age already exists [agec]
      2. Create dummy for marital status [marstat]
         * Code and collapse answer categories ‘married’ and ‘living common law’ into ‘married’ with value ‘1’
         * Code and collapse answer categories ‘widowed’, ‘separated’, ‘divorced’, and ‘single, never married’ into ‘unmarried’ with value ‘0’
         * Code all other answer categories missing ‘ . ‘
      3. Create binary dummy variable for dwelling type [dwelc]
         * Code ‘Single detached house’ into ‘house’ with value ‘1’
         * Code answer categories ‘low-rise apartment’ and ‘high-rise apartment’ and ‘other’ into ‘apartment & other’
         * Code all other answer categories missing ‘ . ‘
      4. Create dummy for household size [hsdsizec]
         * Code answer categories ‘four person household’, ‘five person household’, and ‘six or more person household’ into ‘more than 4’ with value ‘1’
         * Code answer categories ‘one person household, ‘two person household’, and ‘three person household’ into ‘less than 4’ with value ‘0’
         * Code all other answer categories missing ‘ . ‘
   5. Label all recoded variables using the ‘label define’ and ‘label values’ commands according to how they were coded in the previous step
8. **Create Table of Summary Statistics (Summary Table.do)**
   1. Create a table using the ‘tabstat’ command
      1. Include variables for age, marital status, tenure, dwelling type, household size, and mental health
      2. Stratify by sex variable (gives 3 rows of table - female, male, and total)
      3. Include count, mean, and standard deviation
   2. Since the variables are binary variables, tables of frequencies and proportions might be more interesting than a table that shows the mean, but I am not sure how to make one composite table of all of the variables and their frequencies/proportions. Maybe you can try making a table showing a few relevant variables.
      1. Create a sex-stratified table of tenure and mental health using the ‘bysort’ and ‘tab’ commands – will show proportion of respondents in each tenure category with ‘good’ and ‘not good’ mental health for each sex
         * Include tenure as the first var – makes it the row var
         * Include ‘missing’ as an option
         * Include ‘row’ as an option to get percentage proportion within each tenure category
9. **Conduct Logistic Regression Analysis (Analysis & Results Table.do)**
   1. Run a logistic regression of tenure and mental health with age, marital status, dwelling type, and household size as confounders for the whole sample (use ‘logistic’ command – gives odds ratios)
   2. Run a separate logistic regression with the same variables for males and females (can use if statements)
10. **Create Table of Regression Results (Analysis & Results Table.do)**
    1. Store estimates of each regression separately using ‘estimates store’ command then use ‘etable’ command to make 1 table that shows the results for all models
       1. Option ‘column(estimates)’ names each column the name that you used to store the estimation results
       2. Include the ‘showstars’ and ‘showstarsnote’ options to include stars for significant results