

**Part A:** Use the Alzheimer's Data (RM\_xsect.csv) to complete the work below:

- (1) Read data into R for analysis.
- (2) Generate descriptive statistics for "global\_lv," including the minimum, first quartile, median, third quartile, and maximum.
- (3) Based on results in part (2), create a categorical variable for quartiles of "global\_lv" as described below.
  - a. Create 1 categorical variable coded as:
    - i. 1: [min to Q1),
    - ii. 2: [Q1 to median),
    - iii. 3: [median to Q3),
    - iv. 4: [Q3 to max).
  - b. Use the *factor* function to label the 4 levels.
  - c. Check your work from part (3a) using the *tapply* function.
  - d. Make a frequency table for the variable created in (3a). Be sure to include both frequency and relative frequency.
  - e. Create a bar graph for the variable you created in (3a). Be sure to label your axes and make your bar graph publishable.

**Part B:** We will continue to use the Alzheimer's dataset (RM\_xsect.csv) to complete the work below:

- (1) Read data into R for analysis.
- (2) Create a new binary variable for greater than high school education. The variable "educ" is the years of education. Your binary variable should have the following levels:
  - a. 0 if education level is: High School (HS) or less education (education  $\leq$  12 years),
  - b. 1 if education level is: University and more education (education  $>$  12 years).
- (3) Separately plot the distribution of years of education for each education category.
  - a. Calculate the summary statistics for each group.
  - b. Create side-by-side boxplots of the distribution of years of education, stratified by education group.
- (4) Test whether the proportion of those with university or more education is different than 75%.