Epi5143 Winter 2024 Quiz 2

Nora Szabo

Due Monday February 19th, 2024 by 11:59pm, submit via GitHub.

There is a dataset provided called q2_diag.sas7bdat.
Please provide your documented SAS code, log file and output as well as your answers to the questions.

1. **The quiz data includes diagnoses for several years of hospital admissions. Create a new dataset from the quiz dataset that only includes data for admissions from 2002.**

2. **In this new dataset, create a variable called diag_cat from the variable hdgcd rolled up to the first 3 characters/numbers. (ie Z370 would be truncated as Z37).**

3. **Use PROC FREQ to create a frequency table for diag_cat in order of most to least frequent diagnosis, and report what is the most frequent diagnosis from this table.**

   The most frequent diagnosis is I10

4. **Create a new variable in your dataset from Q2. that has the value 1 when the most common diagnosis determined in Q3. is present, and 0 otherwise.**

5. **Using one of the methods we discussed in class, create a flat file with respect to hdgHraEncWID that includes a new variable that identifies whether each admission included the most common diagnosis or not. This dataset should have 1 row for each unique encounter ID, and a value of 0 or 1 for the presence of the most common diagnosis.**

6. **Generate a frequency table for your most frequent diagnosis flag. What number and proportion of admissions have the most common diagnosis?**

262 (8.66%) of admissions have the most frequent diagnosis