Epi5143 Winter 2021 Quiz 2

Due Tuesday March 2nd, 2021 by 11:59pm, submit via GitHub.

There are two datasets for this quiz, called visits and diagnosis.
This quiz requires you to apply your skills in linking and flat-filing.

1. Create a new dataset from the visits dataset that only includes visits with an admission date on or after January 1st 2003. Keep only the encounter ID, (HraEncWID), patient ID (HraPatWID), admission datetime (hraAdmDtm) , and patient gender (hraGenderCd) in this new dataset. This is your spine dataset.
2. Create a new dataset from the diagnosis dataset and your spine dataset that only includes diagnoses from encounters in your spine dataset.
3. In this new dataset, create a variable called diagCat from the variable hdgcd rolled up to the first 3 characters/numbers? (ie Z370 would be rolled up to Z37).
4. Use PROC FREQ to identify the 5 most frequent diagnosis codes based on the diagCat variable. (**Ans: 125, 110, Y83, O70, E87**).
5. What condition do each of these 5 codes represent? **(Ans: E87 = Endocrine, nutritional & metabolic diseases. Y83 = External causes of morbidity & mortality. 070, 110 & 125 = Infectious & parasitic diseases)**.
6. Create a new dataset with one row per encounter that creates:
    a. a flag if your most frequent diagnosis is recorded during each visit (1 if yes, 0 if no).
    b. A flag if any of your top five diagnosis are recorded during each visit (1 if yes, 0 if no).
    c. A count of the total number of diagnoses recorded during each visit.
7. Generate a frequency table for your most frequent diagnosis flag, your 'any of the top 5' flag, and your 'diagnosis count' variable.
8. How many different patients are there in your spine dataset? **(Ans=7338)**
9. Using the patient as the unit of analysis, create a new analytical dataset and create a new variable in that dataset that counts the total number of encounters for each patient.
10. Generate a PROC FREQ frequency table for your number of visits variable and the number of visits by patient gender. **(Ans=4266 Females, 3072 Males)**

Please provide your documented SAS code in addition to answers to the questions and related SAS output.