**Epi5143 Winter 2020 Quiz 5**

Due by Tuesday March 24th, 2019 by 11:59pm. Submit quiz via Github (link will be provided by email)

From the class data:

Determine the proportion of admissions which recorded a diagnosis of diabetes for admissions between January 1st 2003 and December 31st, 2004. Generate a frequency table of frequency of diabetes diagnoses, with the denominator being the total number of admissions between January 1st 2003 and December 31st, 2004 .

This exercise requires sorting, flat-filing, and linking (merging) tables.

Hints:

1. From the NhrAbstracts dataset, you will have to create a new dataset which contains only unique admissions (hraEncWID) with admit dates (hraAdmDtm) between January 1st, 2003 and December 31st, 2004: this is your spine dataset.

"/folders/myshortcuts/MYFOLDERS";\*creating folders for data\*;

libname ex "/folders/myshortcuts/MYFOLDERS/Large database/epi5143 work folder/data";

data ex.abstracts; \*using the abstracts data\*;

set classdat.nhrabstracts; \*it's located in classdata\*;

run;

data ex.spine; \*creating newdataset called spine\*;

set ex.abstracts; \*extracting the information from the original set\*;

where datepart(hraAdmDtm) between '01Jan2003'd and '31Dec2004'd;

keep hraAdmDtm and hraEncWID;

run;

\*sql alternative to creating spine dataset\*;

proc sql;

create table ex.sqlspine as

select hraEncWID, hraAdmDtm

from ex.abstracts

where datepart(hraAdmDtm) between '01Jan2003'd and '31Dec2004'd;

run;

1. From the NhrDiagnosis table you will need to determine if one or more diagnosis codes (hdgcd) for diabetes (ICD9 starting with ‘250’ or ICD10 starting with ‘E11’ or ‘E10’) was present for each encounter in the diagnosis table and create an indicator variable called DM (=0 for no diabetes codes, =1 for one or more diabetes codes).

data ex.diagnosis; \*we are making a dataset called diagnosis and we are using info from nhrdiagnosis\*;

set classdat.nhrdiagnosis;

run;

data ex.diabetes;

set ex.diagnosis;

by hdghraencwid;

if hdgcd in: ('250','E10','E11') then DM=1;

else DM=0;

run;

proc sort data=ex.diabetes;

by hdghraencwid;

run;

1. You will need to flatten your diabetes diagnoses dataset with respect to encounter ID (hdgHraEncWID).

proc transpose data=ex.diabetes out=ex.flat; \*flatten dataset\*;

by hdgHraEncWID;

var DM;

run;

data ex.flat2;\*here, we organize DM\*;

set ex.flat;

if col1=1 or col2=1 or col3=1 or col4=1 or col5=1

or col6=1 or col7=1 or col8=1 or col9=1 or col10=1

or col11=1 or col12=1 or col13=1 or col14=1 or col15=1

or col16=1 or col17=1 or col18=1 or col19=1 or col20=1

or col21=1 or col22=1 or col23=1 or col24=1 then DM=1;

else DM=0;

run;

1. You will need to link the spine dataset you generated from NhrAbstracts and the flattened diabetes diagnoses dataset you generated based on the NhrDiagnosis table using the encounter id’s from each database (renaming as required).

proc sql;

create table ex.linked as

select s.hraencwid as ID, f.dm

from spine as s

left join ex.flat2 as f

on s.hraencwid = f.hdghraencwid;

quit;

data ex.finallink; \*This adds value to the missing data of denominator\*;

set ex.linked;

if dm=. then dm=0;

run;

1. Your final dataset should have the same # of observations (and include all encounter IDs) found in your the spine dataset and have an indicator variable, DM which is 1 if any diabetes code was present, and 0 otherwise.



proc freq data=ex.finallink; \*generating frequency table for DM\*;

table DM;

run;

**The final dataset had the same number of observations 2230 as the spine dataset. The final dataset indicates that 3.72% of the cases had a diagnosis of diabetes.**

* Please provide your final SAS code, and resulting frequency table for the indicator variable you created (plain text so it can be uploaded to Github as a plain text .sas file). You are encouraged to include inline comments to explain the purpose of each step in your program (which may earn you part marks if your code doesn’t actually do what it is supposed to).