**Epi5143 Winter 2022 Quiz 2 – Luke Terrett**

Due Friday March 4th, 2022 by 11:59pm, submit via GitHub.

There are two datasets required for this quiz, a dataset called visits is provided,

download from: <https://www.dropbox.com/s/5c2bi92997uhsub/quiz2_visits.sas7bdat?dl=0>

and you will use the class nhrdiagnosis table.

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 between January 1st 2003 and Jan 31st 2004 inclusive. 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.

See code /output

1. Create a new dataset from the diagnosis dataset and your spine dataset that only includes diagnoses from encounters in your spine dataset.

See code/output

1. 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 ).

See code/output

1. Use PROC FREQ to identify the 5 most frequent diagnosis codes based on the diagCat variable you just created.

Z37, Z38, I25, I10, Y83

See code/output

1. What condition do each of these 5 codes represent?

Z37=outcome of delivery

Z38=liveborn infant

I25=ischemia heart disease

I10=hypertension

Y83=surgical operation

1. Create a new dataset with one row per encounter (make sure this dataset has the same number of rows as your spine dataset):
   1. a flag if your most frequent diagnosis is recorded during each visit (1 if yes, 0 if no).
   2. A count of the total number of diagnoses recorded during each visit.

See code/output below

1. Generate a frequency table for your most frequent diagnosis flag, and your ‘diagnosis count’ variable.

See code/output below

Table

Description automatically generated

1. How many unique patients are there in your spine dataset?

See code/output below

2419 unique patients in the spine dataset (after using nodupkey to remove duplicate encounters)

1. 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 unique patient.

See code/output below

1. Generate a PROC FREQ frequency table for your number of visits variable and the number of visits by patient gender.

See code/output below

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

/\*Epi 5143 - Quiz 2 - Luke Terrett - Mar 4th, 2022\*/

/\*variable names for visits dataset

HraEncWID = encWID

HraPatWID = encPatWID

hraAdmDtm = encStartDtm

hraGenderCd = encPatGenderCd\*/

/\*1. Create a new dataset from the visits dataset that only includes visits with an admission date

between January 1st 2003 and Jan 31st 2004 inclusive. 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.\*/

**data** quiz2\_visits;

set epi5143.quiz2\_visits;

**run**;

**proc** **contents** data=quiz2\_visits;

**run**;

**proc** **sort** data=quiz2\_visits;

by encWID encpatWID;

**run**;

**data** visits\_new;

set quiz2\_visits;

date = datepart(encstartdtm);

format date date9.;

**run**;

**proc** **sort** data=visits\_new;

by date;

**run**;

**data** visits\_new\_1;

set visits\_new;

where **'01jan2003'd** <= date <= **'31jan2004'd**;

**run**;

**proc** **contents** data=visits\_new\_1;

**run**;

**data** spine;

set visits\_new\_1 (keep=encwid encpatwid encpatgendercd encstartdtm);

**run**;

**proc** **contents** data=spine varnum;

**run**;

\*above is my spine dataset;

Table

Description automatically generated

/\*2. Create a new dataset from the diagnosis dataset and your spine dataset that

only includes diagnoses from encounters in your spine dataset.\*/

**data** diagnosis;

set epi5143.nhrdiagnosis;

encwid = hdghraencwid;

**run**;

**proc** **sort** data = diagnosis;

by encwid;

**run**;

**proc** **sort** data = spine;

by encwid;

**run**;

**proc** **contents** data=diagnosis;

**run**;

**proc** **contents** data=spine;

**run**;

**Data** new\_dataset;

merge spine (in=a) diagnosis;

by encwid;

if a=**1** then output;

**run**;

**proc** **contents** data=new\_dataset;

**run**;

Table

Description automatically generated

/\*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 ).\*/

**data** new\_dataset\_1;

length diagcat $ **3**;

set new\_dataset;

diagCat = hdgcd;

**run**;

/\*4. Use PROC FREQ to identify the 5 most frequent diagnosis codes based on the

diagCat variable you just created.\*/

**proc** **freq** data=new\_dataset\_1 order=freq;

table diagcat;

**run**;

Table

Description automatically generated

/\*5. What condition do each of these 5 codes represent?

Z37 143 5.04 143 5.04 Z37 = outcome of delivery

Z38 121 4.26 264 9.30 Z38 = liveborn infant

I25 99 3.49 363 12.79 I25 = ischemic heart disease

I10 98 3.45 461 16.24 I10 = hypertension

Y83 71 2.50 532 18.75 Y83 = surgical operation

\*/

/\*6. Create a new dataset with one row per encounter (make sure this dataset has the same

number of rows as your spine dataset):

a. a flag if your most frequent diagnosis is recorded during each visit (1 if yes, 0 if no).

b. A count of the total number of diagnoses recorded during each visit. \*/

**proc** **sort** data=new\_dataset\_1;

by encwid;

**run**;

**data** new\_dataset\_2;

set new\_dataset\_1;

mostfreq=**0**;

if diagcat ='Z37' then mostfreq=**1**;

else mostfreq=**0**;

**run**;

**proc** **means** data=new\_dataset\_2 noprint;

class encwid;

types encwid;

var mostfreq;

output out=onerow max(mostfreq)=mostfreq n(mostfreq)=count;

**run**;

**proc** **contents** data=onerow varnum;

**run**;

**proc** **freq** data=onerow;

tables mostfreq count;

**run**;

Table

Description automatically generated

/\*7.Generate a frequency table for your most frequent diagnosis flag, and your ‘diagnosis count’ variable\*/

**proc** **freq** data=onerow;

tables mostfreq count;

**run**;

Table

Description automatically generated

/\* 8. How many unique patients are there in your spine dataset?\*/

\*Answer = there are 2419 unique patients in the spine dataset;

**proc** **sort** data=spine out=spine\_ptnodup nodupkey;

by encpatwid;

**run**;

**proc** **contents** data=spine\_ptnodup;

**run**;

Graphical user interface, table, Word

Description automatically generated

/\*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 unique patient.\*/

\*nhrdiagnosis dataset does not have patientID. the only variable in common is the EncounterID;

**data** finaldataset;

set new\_dataset;

**run**;

**proc** **sort** data=finaldataset;

by encpatwid;

**run**;

**data** finaldataset\_1;

set finaldataset;

by encpatwid;

if first.encpatwid

then visits=**0**;

visits=visits+**1**;

if last.encpatwid then output;

retain visits;

**run**;

**proc** **contents** data=finaldataset\_1 varnum;

**run**;

Table

Description automatically generated

/\*10. Generate a PROC FREQ frequency table for your number of visits variable and the number of visits by patient gender.\*/

**proc** **freq** data=finaldataset\_1;

tables visits visits\*encpatgendercd;

**run**;

Table

Description automatically generated

Calendar

Description automatically generated