Hello Yulric and Wenshan,

Can I ask you a few questions the code in this quiz?
There were 2 approaches to do the flat filing from the lecture, and one of them was:

```
 * practice: creating a flatfiling - practicel - second approach;
data flatfile;
set clasdat.quiz5_diag;
dm = 0;
if hdgcd in:('250' 'E10' 'E11') then dm = 1;
run;

proc means data = flatfile;
class hdgHraEncWID;
types hdgHraEncWID;
var dm;
output out=flatfile2 max(dm) = diabetes n(dm) = count sum(dm) = dm_count;
run;

proc freq data=flatfile2;
tables diabetes count dm_count;
run;
```

From my understanding, I thought dm flags if there is hdgcd with any appearance of 250 or E10 or E11 (i.e. E11.1 as well since it has E11 in it) in the hdgcd. So if there is 1 of the 3 codes or 2 of the 3, it would be still dm = 1?

I think I understood the max(dm) outputs into a new dataset called 'flatfile2' as a variable name 'diabetes' and it flags if there is a diabetes diagnosis or not for hdgHraEncWID.

And sum(dm) outputs as a variable name 'diabetes' and I think it flags the number of diabetes diagnosis for the hdgHraEncWID.

And when I did proc freq tables for dm_count (dm_count), I got 10 observations that had 2 (I have attached the tables below). Does that mean that these 10 observations with 2 dm_count is the hdgHraEncWID that has duplicates? Because diabetes, which flags the diabetes, has 1724 for value 1. Also, I accidentally did 'proc sort' the nhrdiagnosis table with nodupkey, and it deleted 10 duplicates, and I am wondering if these 10 observations in 2 for dm_count is the duplicate? And when I linked this diagnosis table to the spine table, there were only 0 or 1 for the count -

'Diabetes' frequency table:

diabetes	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	31120	94.75	31120	94.75
1	1724	5.25	32844	100.00

'Dm_count' Before merging:

dm_count	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	31120	94.75	31120	94.75
1	1714	5.22	32834	99.97
2	10	0.03	32844	100.00

*10 obs with 2 dm_count;

'Dm_count' After merging:

The SAS System

The FREQ Procedure

dm_count	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	2147	96.28	2147	96.28
1	83	3.72	2230	100.00

There is no observations with 2 dm_count

2. And I am not sure if I am understanding the n(dm) = dm_count correctly. I though n(variable) counts the number of non-missing values for the dm. But how can there be 1-24 values for count, when there is largest value for it was 1? Or am I not understanding what dm variable is counting? Is dm counting the number of appearance of 250 or E10 or E11 for hdgHraEncWID? I have attached the frequency table of count variable (n(dm)).

Thank you very much for your time and help!

Arum.

count	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	7701	23.45	7701	23.45
2	8115	24.71	15816	48.15
3	5241	15.96	21057	64.11
4	3819	11.63	24876	75.74
5	2644	8.05	27520	83.79
6	1573	4.79	29093	88.58
7	1098	3.34	30191	91.92
8	769	2.34	30960	94.26
9	539	1.64	31499	95.90
10	370	1.13	31869	97.03
11	267	0.81	32136	97.84
12	202	0.62	32338	98.46
13	139	0.42	32477	98.88
14	114	0.35	32591	99.23
15	63	0.19	32654	99.42
16	53	0.16	32707	99.58
17	44	0.13	32751	99.72
18	29	0.09	32780	99.81
19	29	0.09	32809	99.89
20	18	0.05	32827	99.95
21	10	0.03	32837	99.98
22	4	0.01	32841	99.99
23	2	0.01	32843	100.00
24	1	0.00	32844	100.00