Midterm Part II Report

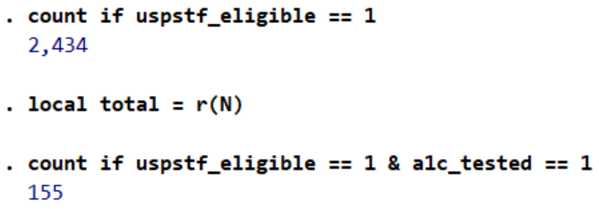
Xuange Liang, xl3493

1. (4 points) USPSTF recommends regularly testing the blood sugar (HbA1c) of patients who are:

- 35 years or older, AND
- Overweight or obese (BMI >=25), AND
- Not already diagnosed with any form of diabetes (Type 1, Type 2, or Unknown Type)

Create a binary flag for everyone who meets all of these inclusion criteria. Then, **tell us what percent of those visits had an HbA1c test completed** during the visit or within the previous 12 months (the HbA1c testing variable is called "A1C").

The percentage of those visits having an HbA1c test completed during the visit or within the previous 12 months is 6.37%.



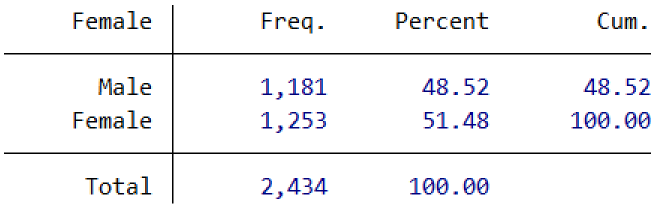
```
. count if uspstf_eligible == 1
  2,434

. local total = r(N)

. count if uspstf_eligible == 1 & a1c_tested == 1
  155
```

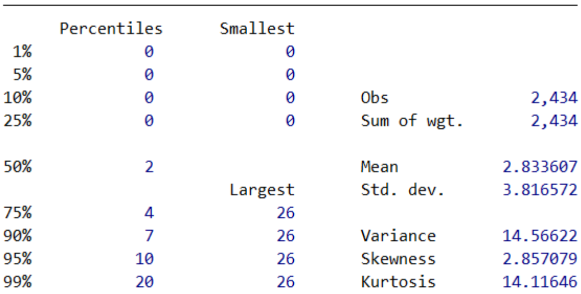
2. (3 points) We need to prepare some other variables in the data set for our analyses. Please conduct the following steps to create variables that can be used in your analyses. **Provide the frequency table or summary statistics (mean, SD, min max) for each of the variables, using the inclusion criteria in Question 1:**

- **Sex:** A 0/1 indicator for whether sex is female



```
     Female |      Freq.     Percent        Cum.
------------+-----------------------------------
       Male |      1,181       48.52       48.52
     Female |      1,253       51.48      100.00
------------+-----------------------------------
      Total |      2,434      100.00
```

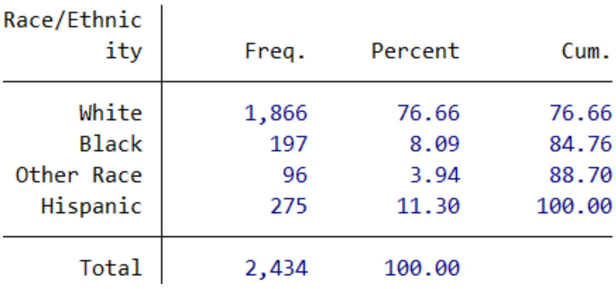
- **Past visits:** The number of past visits the patient has had with this provider in the last 12 months, top coded at 26. New patients should have 0 prior visits.



```
      Percentiles      Smallest
 1%            0              0
 5%            0              0
10%            0              0       Obs               2,434
25%            0              0       Sum of wgt.       2,434

50%            2                      Mean           2.833607
                        Largest       Std. dev.      3.816572
75%            4             26
90%            7             26       Variance       14.56622
95%           10             26       Skewness       2.857079
99%           20             26       Kurtosis       14.11646
```

- **Race/ethnicity:** No changes needed (already clean); just provide the frequency table

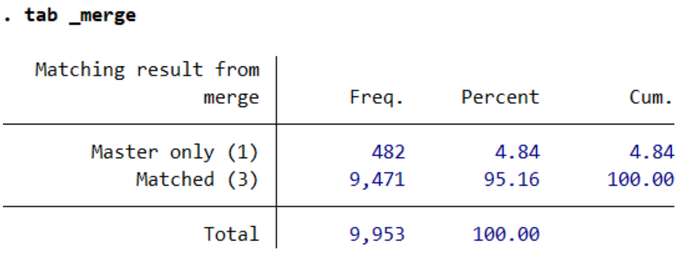


```
 Race/Ethnic |
         ity |      Freq.     Percent        Cum.
-------------+-----------------------------------
       White |      1,866       76.66       76.66
       Black |        197        8.09       84.76
  Other Race |         96        3.94       88.70
    Hispanic |        275       11.30      100.00
-------------+-----------------------------------
       Total |      2,434      100.00
```

3. (4 points) We are also going to use diagnosis information. We want to group diagnoses into CCSR "body system" groups, which are broad diagnosis groupings (e.g., all diseases of the heart and circulatory system are grouped together in a body system called "CIR"). **Using the file called ICD_CCSR_key_final.dta, merge in the CCSR body system categories to your data set, by diagnosis code.** It may help to know that the diagnosis code variable in that file is called "ICD10_4digit" and that the file has one row per diagnosis code.

Briefly **explain why you chose the type of merge** you did (1:1, m:1, 1:m, m:m) and **explain which observations you will keep** in the data set.

I will choose m:1 (many to 1 merge) because multiple visits can have the same diagnosis code". However, I will keep all observation in the master dataset (midtermLSD) because we need to keep all the important visiting records, even though some of them do not match a certain diagnosis group.



```
. tab _merge

   Matching result from |
                  merge |      Freq.     Percent        Cum.
------------------------+-----------------------------------
        Master only (1) |        482        4.84        4.84
            Matched (3) |      9,471       95.16      100.00
------------------------+-----------------------------------
                  Total |      9,953      100.00
```

4. (4 points) The USPSTF recommends that people from racial/ethnic groups with high diabetes prevalence be monitored for potential diabetes more carefully. We want to test whether there are differences in HbA1c screening by race/ethnicity. **Using the survey weights and the same inclusion criteria as question 1, run a regression to test whether the HbA1c testing varies by race/ethnicity groups**, controlling for age (continuous), number of prior visits (continuous), the female indicator, and CCSR Body System. In 1-2 sentences, **report your findings on whether there are differences by race/ethnicity.**

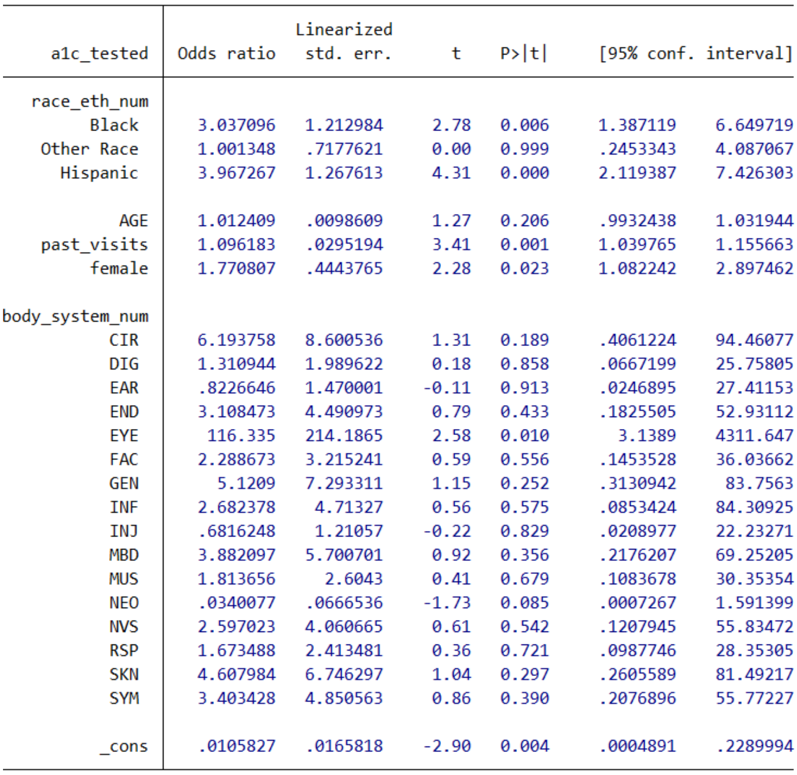
> **NOTE:** You may use a logistic regression or a linear regression for this (in the real world, either is acceptable for this type of outcome, depending on your academic field/department). Whichever you choose, though, you must appropriately interpret the output.

Null hypothesis: There are no differences in HbA1c testing between different race/ethnicity group. (2.race_eth_num = 3.race_eth_num = 4.race_eth_num = 0)

Alterative hypothesis: Not H_0.

I use logistic regression and the F test to analysis the problem. According to the Stata output, the p-value is smaller than 0.05, which means that, there are statistically significant differences in HbA1c screening rates by race/ethnicity (F-test $p < 0.05$).

Compared to White patients (reference), both Black and Hispanic patients have significantly higher odds of receiving A1C testing (Black: 3.04, Hispanic: 3.97), after controlling for age, prior visits, sex, and diagnosis."



```
                           Linearized
  a1c_tested  Odds ratio   std. err.      t    P>|t|     [95% conf. interval]

race_eth_num
       Black    3.037096    1.212984    2.78   0.006     1.387119    6.649719
  Other Race    1.001348    .7177621    0.00   0.999     .2453343    4.087067
    Hispanic    3.967267    1.267613    4.31   0.000     2.119387    7.426303

         AGE    1.012409    .0098609    1.27   0.206     .9932438    1.031944
 past_visits    1.096183    .0295194    3.41   0.001     1.039765    1.155663
      female    1.770807    .4443765    2.28   0.023     1.082242    2.897462

body_system_num
         CIR    6.193758    8.600536    1.31   0.189     .4061224    94.46077
         DIG    1.310944    1.989622    0.18   0.858     .0667199    25.75805
         EAR    .8226646    1.470001   -0.11   0.913     .0246895    27.41153
         END    3.108473    4.490973    0.79   0.433     .1825505    52.93112
         EYE     116.335    214.1865    2.58   0.010       3.1389    4311.647
         FAC    2.288673    3.215241    0.59   0.556     .1453528    36.03662
         GEN      5.1209    7.293311    1.15   0.252     .3130942     83.7563
         INF    2.682378     4.71327    0.56   0.575     .0853424    84.30925
         INJ    .6816248     1.21057   -0.22   0.829     .0208977    22.23271
         MBD    3.882097    5.700701    0.92   0.356     .2176207    69.25205
         MUS    1.813656      2.6043    0.41   0.679     .1083678    30.35354
         NEO    .0340077    .0666536   -1.73   0.085     .0007267    1.591399
         NVS    2.597023    4.060665    0.61   0.542     .1207945    55.83472
         RSP    1.673488    2.413481    0.36   0.721     .0987746    28.35305
         SKN    4.607984    6.746297    1.04   0.297     .2605589    81.49217
         SYM    3.403428    4.850563    0.86   0.390     .2076896    55.77227

       _cons    .0105827    .0165818   -2.90   0.004     .0004891    .2289994
```

```
.
. * Test joint significance of race/ethnicity
. test 2.race_eth_num 3.race_eth_num 4.race_eth_num

Adjusted Wald test

 ( 1)  [a1c_tested]2.race_eth_num = 0
 ( 2)  [a1c_tested]3.race_eth_num = 0
 ( 3)  [a1c_tested]4.race_eth_num = 0

       F(  3,  1071) =    7.19
            Prob > F =    0.0001


.
. * Interpretation
. di _newline(2) "Interpretation:"
```