**RUTGERS UNIVERSITY**

**Bloustein School of Planning and Public Policy**

**Applied Multivariate Methods**

**Fall 2020**

**Dawne Mouzon, Ph.D.**

**Problem Set #12**

**BIVARIATE AND MULTIVARIATE BINARY LOGISTIC REGRESSION**

**NAME: \_\_HASSAN KHURSHID\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**SUBMISSION INSTRUCTIONS: Please upload to Canvas by 11:59 pm next Thursday night.**

**INTRODUCTION TO BINARY LOGISTIC REGRESSION**

**(100 points total)**

**Please be sure to follow all instructions for the problems and submit the specified code and output.**

**Reminder: the output we ask for does not include every consistency check you should be running to check your own work. Generally, what we need to grade you is less than what you should do to check your code throughout the assignment.**

**Please use the Courier New font, size 10 to make your output line up.**

1. **(50 points; 2010 General Social Survey) Please run a series of nested multivariate binary logistic regression models to predict whether or not someone is prochoice (use input variable: *abany*) based on:**

**Model 1: race/ethnicity (race\_eth, ref=NH White)**

**Model 2: Model 1 + sex (sex, ref=women) + age in categories (age4cat, ref=65+)**

**Model 3: Model 2 + education in categories (educ4cat, ref=bachelor’s degree or more)**

**Model 4: Model 3 + political views (polviews, keep as ordinal, treat as interval-ratio when running and interpreting regression) + region (region4cat, ref=Northeast (New Engl and Mid Atl))**

**Your analysis should include:**

1. **Univariate exploration of all analytic variables (GIGO)**
   1. **frequencies for the DV and the categorical IVs (show only first and last ~10 rows of output)**
   2. **fre command for interval-ratio IVs**
2. **Missing data analysis of analytic variables. Do you notice any major fluctuations in your sample (e.g., 10% or more)? If so, how do you plan to investigate further? What hypotheses do you have about your missing data? Please back-code if necessary and write up your findings in a few sentences.**
   1. **Simple descriptives**
   2. **mdesc**
   3. **mvpatterns**
3. **Code and output for proper analytic flag**
4. **An edited table from esttab output. Be sure the table shows odds ratios (“exponentiated coefficients”). Please put the table on its own page when submitting.**
5. **Code from your esttab command (that produced the table in part d)**
6. **A standard write-up of your findings from the entire model series**

**recode abany (2=0), gen(prochoice)**

**label variable prochoice "Dummy for whether or not someone is prochoice"**

**label values prochoice yngss**

**numlabel yngss, add**

**tab prochoice, miss**

**tab1 prochoice race\_eth sex age4cat educ4cat, miss**

**fre polviews, tabulate(10)**

tab1 prochoice race\_eth sex age4cat educ4cat, miss

-> tabulation of prochoice

Dummy for |

whether or not |

someone is |

prochoice | Freq. Percent Cum.

-----------------+-----------------------------------

0. No | 693 33.90 33.90

1. Yes | 537 26.27 60.18

.d. DK | 34 1.66 61.84

.i. IAP | 763 37.33 99.17

.n. NA | 17 0.83 100.00

-----------------+-----------------------------------

Total | 2,044 100.00

-> tabulation of race\_eth

recode of race variable | Freq. Percent Cum.

------------------------+-----------------------------------

1. NH White | 1,422 69.57 69.57

2. NH Black | 298 14.58 84.15

3. Hispanic | 235 11.50 95.65

4. Other | 83 4.06 99.71

.u. unknown Hisp origin | 6 0.29 100.00

------------------------+-----------------------------------

Total | 2,044 100.00

-> tabulation of sex

respondents |

sex | Freq. Percent Cum.

------------+-----------------------------------

1. male | 891 43.59 43.59

2. female | 1,153 56.41 100.00

------------+-----------------------------------

Total | 2,044 100.00

-> tabulation of age4cat

dummy age | Freq. Percent Cum.

---------------+-----------------------------------

1. 18-24 | 181 8.86 8.86

2. 25-55 | 746 36.50 45.35

3. 45-64 | 721 35.27 80.63

4. 65 or older | 393 19.23 99.85

.n. NA | 3 0.15 100.00

---------------+-----------------------------------

Total | 2,044 100.00

-> tabulation of educ4cat

highest level of school |

completed | Freq. Percent Cum.

----------------------------+-----------------------------------

1. less than high school | 346 16.93 16.93

2. high school diploma | 558 27.30 44.23

3. some college | 524 25.64 69.86

4. bachelors degree or more | 611 29.89 99.76

. | 5 0.24 100.00

----------------------------+-----------------------------------

Total | 2,044 100.00

. fre polviews, tabulate(10)

polviews -- think of self as liberal or conservative

--------------------------------------------------------------------------------

| Freq. Percent Valid Cum.

-----------------------------------+--------------------------------------------

Valid 1 1. extremely liberal | 76 3.72 3.85 3.85

2 2. liberal | 259 12.67 13.13 16.98

3 3. slightly liberal | 232 11.35 11.76 28.74

4 4. moderate | 746 36.50 37.81 66.55

5 5. slghtly conservative | 265 12.96 13.43 79.98

6 6. conservative | 315 15.41 15.97 95.95

7 7. extrmly conservative | 80 3.91 4.05 100.00

Total | 1973 96.53 100.00

Missing .d .d. DK | 61 2.98

.n .n. NA | 10 0.49

Total | 71 3.47

Total | 2044 100.00

--------------------------------------------------------------------------------

b)

**. sum prochoice race\_eth sex age4cat educ4cat polviews**

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

prochoice | 1,230 .4365854 .496164 0 1

race\_eth | 2,038 1.499019 .8516633 1 4

sex | 2,044 1.56409 .4959968 1 2

age4cat | 2,041 2.649682 .8890737 1 4

educ4cat | 2,039 2.686611 1.074448 1 4

-------------+---------------------------------------------------------

polviews | 1,973 4.079574 1.456695 1 7

**. mdesc prochoice race\_eth sex age4cat educ4cat region4cat polviews**

Variable | Missing Total Percent Missing

----------------+-----------------------------------------------

prochoice | 814 2,044 39.82

race\_eth | 6 2,044 0.29

sex | 0 2,044 0.00

age4cat | 3 2,044 0.15

educ4cat | 5 2,044 0.24

region4cat | 0 2,044 0.00

polviews | 71 2,044 3.47

----------------+-----------------------------------------------

**. mvpatterns prochoice race\_eth sex age4cat educ4cat region4cat polviews**

variables with no mv's: sex region4cat

Variable | type obs mv variable label

-------------+------------------------------------------------------------------

prochoice | byte 1230 814 Dummy for whether or not someone is prochoice

race\_eth | float 2038 6 recode of race variable

age4cat | float 2041 3 dummy age

educ4cat | byte 2039 5 highest level of school completed

polviews | byte 1973 71 think of self as liberal or conservative

--------------------------------------------------------------------------------

Patterns of missing values

+------------------------+

| \_pattern \_mv \_freq |

|------------------------|

| +++++ 0 1186 |

| .++++ 1 777 |

| ++++. 1 36 |

| .+++. 2 31 |

| +.+++ 1 4 |

|------------------------|

| .++.+ 2 3 |

| ++.++ 1 2 |

| +++.. 2 1 |

| +.++. 2 1 |

| ..+++ 2 1 |

|------------------------|

| .++.. 3 1 |

| .+.+. 3 1 |

The dependent variable abany has approximately 40 percent missing values so I will explore it to see any potential for back-coding.

.

**tab abany, miss**

abortion if |

woman wants |

for any |

reason | Freq. Percent Cum.

------------+-----------------------------------

1. yes | 537 26.27 26.27

2. no | 693 33.90 60.18

.d. DK | 34 1.66 61.84

.i. IAP | 763 37.33 99.17

.n. NA | 17 0.83 100.00

------------+-----------------------------------

Total | 2,044 100.00

It turns out that around 37 percent people have been categorized as those upon whom this question is inapplicable. Since this is a case of missing data which cannot be recovered, so I will continue with the current dataset.

c)

**mark prochoice\_flag**

**markout prochoice\_flag prochoice i.race\_eth ib2.sex ib4.age4cat ib4.educ4cat polviews i.region4cat**

**label variable prochoice\_flag "analytical flag for abortion opinion (n=1186)"**

**tab prochoice\_flag, miss**

analytical |

flag for |

abortion |

opinion |

(n=1186) | Freq. Percent Cum.

------------+-----------------------------------

0 | 858 41.98 41.98

1 | 1,186 58.02 100.00

------------+-----------------------------------

Total | 2,044 100.00

d)

Multivariate Binary Logistic Regression Models Predicting Prochoice for abortions based on Race/Ethnicity and Sociodemographic Characteristics and political-views, 2010 General Social Survey

--------------------------------------------------------------------------------------------

Model 1 Model 2 Model 3 Model 4

OR OR OR OR

--------------------------------------------------------------------------------------------

Dummy for whether or not someone is prochoice

**Race/ethnicity (ref= NH White)**

2. NH Black 0.899 0.843 0.970 0.770

3. Hispanic **0.398\*\*\*** **0.356\*\*\*** **0.481\*\*** **0.397\*\*\***

4. Other 1.296 1.171 1.105 0.783

**Sex (ref = female)**

1. male 1.153 1.172 **1.255+**

**Age (ref = 65 or older)**

1. 18-24 **2.392\*\***  **2.340\*\* 2.157\*\***

2. 25-55 **2.028\*\*\*** **1.800\*\* 1.875\*\*\***

3. 45-64 **1.773\*\***  **1.618\*\* 1.693\*\***

**Education (ref =Bachelors degree or more)**

1. less than high school **0.364\*\*\* 0.353\*\*\***

2. high school diploma **0.543\*\*\* 0.611\*\***

3. some college **0.672\* 0.703\***

**Political-views(think of self as liberal or conservative) 0.661\*\*\***

**Region (ref = Northeast (New Engl and Mid Atl**)

2. Midwest (W-No Central and E-No Centr **0.661\***

3. South (S Atl, E-S Central, W-S Centr 0.876

4. West (Mountain and Pacific) 1.341

--------------------------------------------------------------------------------------------

N 1186.000 1186.000 1186.000 1186.000

r2\_p 0.014 0.027 0.045 0.106

p 0.000 0.000 0.000 0.000

--------------------------------------------------------------------------------------------

Exponentiated coefficients

+ p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001

e)

**logistic prochoice i.race\_eth if prochoice\_flag == 1**

**estimates store prochoice1**

**logistic prochoice i.race\_eth ib2.sex ib4.age4cat if prochoice\_flag == 1**

**estimates store prochoice2**

**logistic prochoice i.race\_eth ib2.sex ib4.age4cat ib4.educ4cat if prochoice\_flag == 1**

**estimates store prochoice3**

**logistic prochoice i.race\_eth ib2.sex ib4.age4cat ib4.educ4cat polviews i.region4cat if prochoice\_flag == 1**

**estimates store prochoice4**

**esttab prochoice1 prochoice2 prochoice3 prochoice4, eform varwidth(40) label b(%6.3f) star(+ 0.10 \* 0.05 \*\* 0.01 \*\*\* 0.001) stats(N r2\_p p) compress nogaps not ///**

**title(Multivariate Binary Logistic Regression Models Predicting Prochoice for abortions based on Race/Ethnicity and Sociodemographic Characteristics and political-views, 2010 General Social Survey) ///**

**mtitle("Model 1" "Model 2" "Model 3" "Model 4")**

f)

Write-up:

A nested multivariate binary logistic regression model series was conducted to determine the predictors of being prochoice about women’s right to abortion for any reason. Model 1 included race/ethnicity (NH White, NH Black, Hispanic, Other), Model 2 added sex and age (18-24, 25-44, 45-64, 65 or older), Model 3 added education (less than high school, high school diploma, some college, and bachelor’s degree or more), and Model 4 added region (Northeast, Midwest, South, West).

In Model1, NH Hispanics had significantly lower odds for being prochoice (*p<0.001,* OR = 0.4) compared to NH Whites. There was no significant difference in odds for NH Blacks (OR= 0.9, *p =0.54*) and NH Other (OR=1.3, *p= 0.378*) compared to NH Whites. Model 1 was significant (*p<0.001*) and accounted for 1.4% of the variation in the support for being prochoice about abortion.

Model2, after adding for sex and age, denoted the same odds difference among ethnicities. Again, Hispanic had significantly lower odds (64 percent lesser) of being prochoice compared to NH Whites (OR =0.36, *p <0.001*), controlling for age and sex. Still there was no significant difference in odds of being prochoice between NH Blacks and Others, with respect to NH whites. Model 2 showed that sex was not a significant predictor of supporting abortion (OR= 1.15, *p = 0.234*). However, there were significant differences in the odds of being prochoice among people of different age groups when controlling for sex and race. Those who were 18-24 years old had 2.39 times higher odds (*p =0.001*), those who were 25 -55 had 2.03 times higher odds (*p <0.001*) , and those who were 45 – 64 years old had 77 percent higher odds (*p =0.001*) of being prochoice, with respect to people aged 65 and above. The overall model was significant (*p <0.001*) and accounted for 2.7 percent of the variation in being prochoice for abortion.

In Model 3, education was added which did not change the trends of odd’s difference for being prochoice among variables: race, age, and sex. Compared to NH Whites, Hispanics had 52 percent lower odds of being prochoice (*p = 0.001*), while NH Blacks (OR=0.97, *p-value= 0.87*) and Others (OR=1.1, *p= 0.744*) did not have odds significantly different, when controlling for sex, age, and education.

Males also did not have significantly greater odds for being prochoice (OR=0.91, *p = 0.19*), with respect to females, when controlled for age, race, and education.

Age again was a significant predictor for supporting abortion after controlling for race/ethnicity, sex and education:18-24 years old had 2.34 times higher odds (*p =0.002*), 25-55 years old had 80% higher odds (*p =0.001*), and 45-64 years old had 61.8% higher odds (*p =0.008*) of being prochoice with respect to those aged 65 and above. Model 3 also found that education was a significant predictor of whether someone was a prochoice for abortion: those with less than high school education had 63 percent lower odds (*p <0.001*)*,* those with high school diploma had 45 % lower odds (*p <0.001*), and those with some college education had 33 percent lower odds (*p =0.011*) with respect to those having a Bachelors degree or more while controlling for age, sex, and race. Overall Model 3 was significant (*p <0.001*) and accounted for 4.45 % variation in whether someone agreed with women’s right to undergo abortion for any reason.

Model 4 added for political views and region of the respondents and found that when controlling for all the variables, Hispanics still had a significantly lower odd of being prochoice compared to NH Whites (OR = 0.4, *p <0.001*). Again, NH Blacks and others didn’t have significantly lower odds with respect to NH Whites. The sex of respondents now has become marginally significant predictor of whether someone is prochoice (*p = 0.076*). Males happen to have 25 percent greater odds of being prochoice compared to women, when controlling for age, race, education, political views, and region of respondents. Age remains a significantly strong predictor of whether someone is prochoice. 18-24 years old had 2.15 times higher odds (*p =0.008*), 25-55 years old had 87.5% higher odds (*p =0.001*), and 45-64 years old had 69% higher odds (*p =0.006*) of being prochoice with respect to those aged 65 and above when controlling for race, sex, education, political views, and region of respondents. Like Model 3 education was a significant predictor of whether someone was a prochoice for abortion: those with less than high school education had 65 percent lower odds (*p <0.001*)*,* those with high school diploma had 39 % lower odds (*p =0.003*), and those with some college education had 30 percent lower odds (*p =0.033*) with respect to those having a Bachelors degree or more while controlling for age, sex, political views, region and race of respondents. Additionally, it was found that for each unit decrease in liberality of respondents, the odds of being prochoice decrease by 33 % (*p <0.001*), while controlling for all other variables in the model. And lastly, respondents coming from Midwest had significantly lower odds (OR = 0.66, *p =0.043*) of being prochoice with respect to those coming from Northeast, while controlling for all other variables in the model. And there was no significant difference in odds for being prochoice for respondents from the South (OR=0.87, *p-value= 0.484*) and West (OR= 1.34, *p-value= 0.14*), with respect to those in the Northeast. Model 4 was also significant (*p<0.001*) and accounted for 10.63% of the variation in the support for abortion for any reason.

1. **(50 points; 2018 General Social Survey) Please run a series of nested multivariate binary logistic regression models (at least three models, at least 6-8 variables) to predict whether or not someone is married based on any combination of variables in the dataset. Your final analytic sample must include at least 400 cases.**

**Your analysis should include:**

1. **Univariate exploration of all analytic variables (GIGO)**
   1. **frequencies for the DV and the categorical IVs (show only first and last ~10 rows of output)**
   2. **fre command for interval-ratio IVs**
2. **Missing data analysis of analytic variables. Do you notice any major fluctuations in your sample (e.g., 10% or more)? If so, how do you plan to investigate further? What hypotheses do you have about your missing data? Please back-code if necessary and write up your findings in a few sentences.**
   1. **Simple descriptives**
   2. **mdesc**
   3. **mvpatterns**
3. **Code and output for proper analytic flag**
4. **An edited table from esttab output. Be sure the table shows odds ratios (“exponentiated coefficients”). Please put the table on its own page when submitting.**
5. **Code from your esttab command (that produced the table in part d)**
6. **A standard write-up of your findings from the entire model series**

**a)**

**recode absingle (2=0), gen(married)**

**label variable married "Dummy for whether or not someone is married"**

**label values married yngss**

**numlabel yngss, add**

**tab married, miss**

**tab1 married age4cat sex race\_eth relig5cat, miss**

**fre conrinc10k bmi2018rd, tabulate(10)**

. tab1 married age4cat sex race\_eth relig5cat, miss

-> tabulation of married

Dummy for |

whether or not |

someone is |

married | Freq. Percent Cum.

-----------------+-----------------------------------

0. No | 816 34.75 34.75

1. Yes | 700 29.81 64.57

.d. DK | 38 1.62 66.18

.i. IAP | 774 32.96 99.15

.n. NA | 20 0.85 100.00

-----------------+-----------------------------------

Total | 2,348 100.00

-> tabulation of age4cat

dummy age | Freq. Percent Cum.

---------------+-----------------------------------

1. 18-24 | 197 8.39 8.39

2. 25-55 | 851 36.24 44.63

3. 45-64 | 754 32.11 76.75

4. 65 or older | 539 22.96 99.70

.n. NA | 7 0.30 100.00

---------------+-----------------------------------

Total | 2,348 100.00

-> tabulation of sex

respondents |

sex | Freq. Percent Cum.

------------+-----------------------------------

1. male | 1,052 44.80 44.80

2. female | 1,296 55.20 100.00

------------+-----------------------------------

Total | 2,348 100.00

-> tabulation of race\_eth

recode of race variable | Freq. Percent Cum.

------------------------+-----------------------------------

1. NH White | 1,504 64.05 64.05

2. NH Black | 371 15.80 79.86

3. Hispanic | 349 14.86 94.72

4. Other | 109 4.64 99.36

.u. unknown Hisp origin | 15 0.64 100.00

------------------------+-----------------------------------

Total | 2,348 100.00

-> tabulation of relig5cat

recoded relig affiliation | Freq. Percent Cum.

----------------------------------------+-----------------------------------

1. Protestant | 1,175 50.04 50.04

2. Catholic | 493 21.00 71.04

3. Buddh/Hindu/Moslem or Islam/Other Ea | 44 1.87 72.91

4. none | 542 23.08 96.00

5. other | 73 3.11 99.11

.d. DK | 5 0.21 99.32

.n. NA | 16 0.68 100.00

----------------------------------------+-----------------------------------

Total | 2,348 100.00

. fre conrinc10k bmi2018rd, tabulate(10)

conrinc10k -- income in $10k

--------------------------------------------------------------

| Freq. Percent Valid Cum.

-----------------+--------------------------------------------

Valid .03505 | 33 1.41 2.42 2.42

.1402 | 32 1.36 2.35 4.77

.24535 | 32 1.36 2.35 7.12

.31545 | 21 0.89 1.54 8.66

.38555 | 21 0.89 1.54 10.20

.45565 | 12 0.51 0.88 11.08

.52575 | 18 0.77 1.32 12.40

.6309 | 33 1.41 2.42 14.82

.788625 | 48 2.04 3.52 18.34

.963875 | 46 1.96 3.37 21.72

: | : : : :

2.62875 | 81 3.45 5.94 52.38

3.1545 | 136 5.79 9.98 62.36

3.8555 | 114 4.86 8.36 70.73

4.73175 | 114 4.86 8.36 79.09

5.78325 | 81 3.45 5.94 85.03

7.01 | 67 2.85 4.92 89.95

8.412 | 43 1.83 3.15 93.10

9.814 | 32 1.36 2.35 95.45

11.216 | 14 0.60 1.03 96.48

18.26959 | 48 2.04 3.52 100.00

Total | 1363 58.05 100.00

Missing . | 985 41.95

Total | 2348 100.00

--------------------------------------------------------------

bmi2018rd -- BMI rounded to whole number

-----------------------------------------------------------------------

| Freq. Percent Valid Cum.

--------------------------+--------------------------------------------

Valid 15 | 1 0.04 0.07 0.07

16 | 2 0.09 0.15 0.22

17 | 2 0.09 0.15 0.36

18 | 9 0.38 0.66 1.02

19 | 23 0.98 1.67 2.69

20 | 44 1.87 3.20 5.90

21 | 59 2.51 4.29 10.19

22 | 88 3.75 6.40 16.59

23 | 90 3.83 6.55 23.14

24 | 94 4.00 6.84 29.99

: | : : : :

49 | 3 0.13 0.22 99.13

50 | 1 0.04 0.07 99.20

51 | 2 0.09 0.15 99.34

52 | 3 0.13 0.22 99.56

53 | 1 0.04 0.07 99.64

54 | 1 0.04 0.07 99.71

55 | 1 0.04 0.07 99.78

56 | 1 0.04 0.07 99.85

57 | 1 0.04 0.07 99.93

60 | 1 0.04 0.07 100.00

Total | 1374 58.52 100.00

Missing . | 34 1.45

.d .d. don’t know | 4 0.17

.i .i. IAP | 900 38.33

.n .n. no answer | 36 1.53

Total | 974 41.48

Total | 2348 100.00

**b)**

**sum married age4cat sex race\_eth relig5cat conrinc10k bmi2018rd**

**mdesc married age4cat sex race\_eth relig5cat conrinc10k bmi2018rd**

**mvpatterns married age4cat sex race\_eth relig5cat conrinc10k bmi2018rd**

sum married age4cat sex race\_eth relig5cat conrinc10k bmi2018rd

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

married | 1,516 .4617414 .4986986 0 1

age4cat | 2,341 2.698419 .9163949 1 4

sex | 2,348 1.551959 .4973989 1 2

race\_eth | 2,333 1.598371 .9056412 1 4

relig5cat | 2,327 2.073915 1.316488 1 5

-------------+---------------------------------------------------------

conrinc10k | 1,363 3.681271 3.71783 .03505 18.26959

bmi2018rd | 1,374 28.35007 6.330921 15 60

. mdesc married age4cat sex race\_eth relig5cat conrinc10k bmi2018rd

Variable | Missing Total Percent Missing

----------------+-----------------------------------------------

married | 832 2,348 35.43

age4cat | 7 2,348 0.30

sex | 0 2,348 0.00

race\_eth | 15 2,348 0.64

relig5cat | 21 2,348 0.89

conrinc10k | 985 2,348 41.95

bmi2018rd | 974 2,348 41.48

----------------+-----------------------------------------------

. mvpatterns married age4cat sex race\_eth relig5cat conrinc10k bmi2018rd

variables with no mv's: sex

Variable | type obs mv variable label

-------------+----------------------------------------------------------------

married | byte 1516 832 Dummy for whether or not someone is married

age4cat | float 2341 7 dummy age

race\_eth | float 2333 15 recode of race variable

relig5cat | float 2327 21 recoded relig affiliation

conrinc10k | float 1363 985 income in $10k

bmi2018rd | float 1374 974 BMI rounded to whole number

------------------------------------------------------------------------------

Patterns of missing values

+------------------------+

| \_pattern \_mv \_freq |

|------------------------|

| ++++++ 0 738 |

| ++++.. 2 493 |

| .+++++ 1 417 |

| .+++.. 3 277 |

| +++++. 1 133 |

|------------------------|

| ++++.+ 1 133 |

| .+++.+ 2 65 |

| .++++. 2 53 |

| +++.++ 1 6 |

| ++.+++ 1 5 |

|------------------------|

| .+.+.. 4 5 |

| .++.++ 2 4 |

| .++... 4 4 |

| ..++++ 2 3 |

| +.++++ 1 1 |

|------------------------|

| +++..+ 2 1 |

| ++.++. 2 1 |

| ++.+.+ 2 1 |

| +++... 3 1 |

| ++.+.. 3 1 |

|------------------------|

| .++.+. 3 1 |

| ..+++. 3 1 |

| ++.... 4 1 |

| +.+... 4 1 |

| .+.... 5 1 |

|------------------------|

| ..+... 5 1 |

+------------------------+

It turned out that that married, conrinc10k, and bmi2018 were having more than 10 percent missing values. Recovering missing values from conrinc10k and bim2018 is not possible as they are interval-ratio and there is no path for them that can be tracked. As for married dummy variable, the missing values are because of people categorized as inapplicable which cannot be further explored. So, I will use dataset as with 738 values.

**c)**

**mark married\_flag**

**markout married\_flag married age4cat sex race\_eth relig5cat conrinc10k bmi2018rd**

**label variable married\_flag "analytic flag for those married (n=738)"**

**tab married\_flag, miss**

. tab married\_flag, miss

analytic |

flag for |

those |

married |

(n=738) | Freq. Percent Cum.

------------+-----------------------------------

0 | 1,610 68.57 68.57

1 | 738 31.43 100.00

------------+-----------------------------------

Total | 2,348 100.00

Multivariate Binary Logistic Regression Models Predicting marital status based on Race/Ethnicity and Sociodemographic Characteristics, 2018 General Social Survey

----------------------------------------------------------------------------

Model 1 Model 2 Model 3

OR OR OR

----------------------------------------------------------------------------

Dummy for whether someone is married

**Age (ref = 65 or older)**

1. 18-24 1.130 1.163 0.975

2. 25-55 1.342 1.487 1.079

3. 45-64 0.974 1.034 0.787

**Sex (ref =female)**

1. male 1.125 0.847

**Race/ Ethnicity (ref =NH White)**

2. NH Black **0.675+** 0.840

3. Hispanic **0.628\*** 0.701

4. Other 1.195 1.268

**BMI rounded to whole number** **0.966\*\*** **0.973\*\***

**Religion (ref = Protestant)**

2. Catholic 1.37

3. Buddh/Hindu/Moslem or Islam/Other Eas 2.24

4. none  **3.362\*\*\***

5. other **14.03\*\*\***

**Income in $10k**  **1.069\*\***

-------------------------------------------------------------------------------

N 738.000 738.000 738.000

r2\_p 0.004 0.022 0.0826

p 0.233 0.004 0.000

-------------------------------------------------------------------------------

Exponentiated coefficients

+ p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**d)**

**logistic married ib4.age4cat if married\_flag == 1**

**estimates store married1**

**logistic married ib4.age4cat ib2.sex i.race\_eth bmi2018rd if married\_flag == 1**

**estimates store married2**

**logistic married ib4.age4cat ib2.sex i.race\_eth i.relig5cat conrinc10k if married\_flag == 1**

**estimates store married3**

**esttab married1 married2 married3, eform varwidth(40) label b(%6.3f) star(+ 0.10 \* 0.05 \*\* 0.01 \*\*\* 0.001) stats(N r2\_p p) compress nogaps not ///**

**title(Multivariate Binary Logistic Regression Models Predicting marital status based on Race/Ethnicity and Sociodemographic Characteristics, 2018 General Social Survey) ///**

**mtitle("Model 1" "Model 2" "Model 3")**

**e)**

**Write-up:**

A nested multivariate binary logistic regression model series was conducted to determine the predictors of being married. Model 1 included age in groups (18-24, 25-44, 45-64, 65 or older), Model 2 added sex, race/ethnicity (NH White, NH Black, Hispanic, Other), and BMI (15/low – 60/max) and Model 3 added religion (Catholics, Protestant, Hindu/Muslim/Buddhist, none, and Others) and income in 10k increments.

In Model 1, it was found that age was not a significant predictor about whether someone is married or not. With respect to those aged 65 and above, people who were 18-24 (OR =1.1, *p =0.75*), people who were 25 -55 (OR =1.3, *p =0.32*), and people who were 45 -64 years old (OR =0.97, *p =0.93*) did not have significantly different odds of being married. The overall model was not significant (*p =0.23*) and accounted for 0.42 percent of variation in whether someone is married or not.

Model 2, after adding sex, race, and bmi of respondents observed that age still was not a significant predictor about whether someone is married. With respect to those who are 65 and above, all other age groups don’t have significantly different odds of being married. Sex turned out to be insignificant predictor as well when controlling age, race, and bmi of respondents: males don’t have significantly greater odds of getting married than females (OR =1.1, *p =0.43*). In terms of race, when all other variables were controlled, those who were NH Blacks had 33 percent marginally lower odds of getting married (*p =0.075*), and those who were Hispanics had 37 percent significantly lower odds of getting married (*p=0.028*), compared to NH Whites. Apart from these, people from ‘Other’ races did not have significantly greater or lesser odds of getting married compared to NH Whites (OR =1.2, *p =0.63*). Moreover, bmi of participants showed that with an addition of one unit in bmi index of participants, the odds of getting married reduced significantly (*p =0.005*) by 3.5 %, when all other variables in Model2 were controlled. Overall Model 2 was significant (*p =0.004*)and accounted for 2.2% variation in whether someone from the respondents is married or not.

Model 3 after adding religion and income of respondents found that age persisted to be insignificant predictor of being married. Compared to those who are 65 and above, all other age groups don’t have significantly different odds of being married, while controlling for the rest of the variables in Model 3. Similarly, like Model 2, gender remained insignificant predictor as males did not have higher odds of getting married than females (OR = 0.84, *p=0.3*). Race dropped its significance in model 3 as a predictor of getting married: Now, compared to NH Whites, none of the ethnicity’s groups have significantly greater or lesser odds of getting married. It was found once again that BMI of participants upon increasing by one unit diminishes significantly the odds of getting married by 2.8 percent (*p=0.037*). Furthermore, upon controlling for the rest of the variables, it was found that compared to Protestants, those who are Catholics (OR = 1.38, *p=0.13*) and those who come under category ‘Hindus/Muslims/Buddhists’ (OR = 2.24, *p=0.29*) did not have significantly greater odds of getting married. However, those who don’t have a religion have 3.37 times higher odds of getting married (*p<0.001*) and those who have some other religion have 14 times higher odds of getting married (*p =0.001*), compared to Protestants. It was also observed that an increase in income of 10,000 dollars increases the odds of getting married by 6.3 percent (*p =0.016*). Overall the model was significant (*p <0.001*) and accounted for 8.26 percent of variation in whether the respondent is married or not.