**RUTGERS UNIVERSITY**

**Bloustein School of Planning and Public Policy**

**Applied Multivariate Methods**

**Fall 2020**

**Dawne Mouzon, Ph.D.**

**Problem Set #13**

**BIVARIATE AND MULTIVARIATE ORDINAL LOGISTIC REGRESSION**

**NAME: \_Hassan Khurshid\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**SUBMISSION INSTRUCTIONS: Please upload to Canvas by 11:59 pm next Monday, 12/14.**

**INTRODUCTION TO ORDINAL 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) Please use the 2017 NJ ACS dataset (file is on the Canvas site).* Please run a series of nested multivariate ordinal logistic regression models to predict income in quartiles (input variable: inctot, bottom-code negative values to 0) based on:**

**Model 1: Dummy variable for presence of one (or more than one) difficulty, using input variables diffrem diffphys diffmob diffcare diffsens\***

**Model 2: Model 1 + race/ethnicity (race\_eth\*\*, ref=NH Whites) + sex (sex, ref=women)**

**Model 3: Model 2 + education (educ5cat\*\*, ref=graduate education) + age in years (age)**

**Your analysis should include:**

1. **Univariate exploration of all analytic variable:** run the appropriate commands for a variable based on its variable-type, i.e. categorical vs. interval ratio and submit correct output for each
   1. **Be sure to include frequencies for the new dummy variable**
2. **Missing data analysis of analytic variables:** note any patterns or concerns you have; investigate abnormal results and hypothesize reasons for missing data
   1. **Simple descriptives**
   2. **mdesc**
   3. **mvpatterns**
3. **Code and output for proper analytic flag used throughout Models**
4. **An edited table from esttab command output. Be sure the table shows proportional 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**

\*HINT: This is a multi-step process you will need to recode each variable, create a scale, and then a final dummy variable.

\*\*see page 3 for code for race\_eth and educ5cat

**a)**

**sum inctot, detail**

**recode inctot (min/0 = 0)**

**sum inctot, detail**

**recode inctot (0/7299 = 1) (7300/32999 =2) (33000/94999 =3) (95000/9999999 = 4), gen(inctot\_bcq)**

**label variable inctot\_bcq "quartile range variable version of inctot"**

**label define bcq 1 "0-24th percentile" 2 "25th to 49th percentile" 3 "50th to 74th percentile" 4 "75th percentile and higher" .i "IAP, DK, NA, uncodeable"**

**label values inctot\_bcq bcq**

**tab inctot\_bcq, miss**

**tab1 diffrem diffphys, miss**

**recode diffrem (0 = .n) (1 =0) (2=1), gen(diffrem\_rec)**

**label define yngss 0 "No" 1 "Yes" .n "NA" .i "IAP" .d "DK" .c "Can't Choose"**

**label values diffrem\_rec yngss**

**numlabel yngss, add force**

**tab diffrem\_rec, miss**

**recode diffphys (0 = .n) (1 =0) (2=1), gen(diffphys\_rec)**

**label values diffphys\_re yngss**

**numlabel yngss, add force**

**tab diffphys\_rec, miss**

**\*Constructing a composite scale now\***

**egen diff\_scale = rowtotal(diffrem\_rec diffphys\_rec), miss**

**label variable diff\_scale "2 item difficulty scale"**

**sum diff\_scale diffrem\_rec diffphys\_rec**

**tab diff\_scale, miss**

**label values diff\_scale miss**

**tab diff\_scale, miss**

**recode diff\_scale (1/2 =1), gen(diff\_dum)**

**label variable diff\_dum "dummy for difficulty from two item scale"**

**label define diff 0 "No Difficulty" 1 "Yes Difficulty" .n "Not Available"**

**label values diff\_dum diff**

**numlabel diff, add force**

**tab diff\_dum, miss**

**tab1 inctot\_bcq diff\_dum race\_eth sex educ5cat, miss**

**fre age, tabulate(10)**

tab1 inctot\_bcq diff\_dum race\_eth sex educ5cat, miss

-> tabulation of inctot\_bcq

quartile range variable |

version of inctot | Freq. Percent Cum.

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

0-24th percentile | 2,199 25.00 25.00

25th to 49th percentile | 2,187 24.86 49.86

50th to 74th percentile | 2,206 25.08 74.93

75th percentile and higher | 2,205 25.07 100.00

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

Total | 8,797 100.00

-> tabulation of diff\_dum

dummy for |

difficulty from |

two item scale | Freq. Percent Cum.

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

0. No Difficulty | 7,521 85.50 85.50

1. Yes Difficulty | 779 8.86 94.35

. | 497 5.65 100.00

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

Total | 8,797 100.00

-> tabulation of race\_eth

race/ethnicity | Freq. Percent Cum.

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

1. NH White | 4,779 54.33 54.33

2. NH Black | 1,063 12.08 66.41

3. hispanic/latino, any race | 1,666 18.94 85.35

4. NH Nat Amer or Alask Nat | 13 0.15 85.50

5. NH Asian or Pac Islander | 1,053 11.97 97.47

6. NH Other Race | 41 0.47 97.93

7. NH 2 or More Race Groups | 182 2.07 100.00

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

Total | 8,797 100.00

-> tabulation of sex

sex | Freq. Percent Cum.

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

1. male | 4,374 49.72 49.72

2. female | 4,423 50.28 100.00

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

Total | 8,797 100.00

-> tabulation of educ5cat

5-category educational |

attainment | Freq. Percent Cum.

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

1. less than HS diploma | 2,147 24.41 24.41

2. HS diploma or equivalent | 2,437 27.70 52.11

3. some college or associate's | 1,418 16.12 68.23

4. Bachelor's degree | 1,705 19.38 87.61

5. graduate education | 1,090 12.39 100.00

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

Total | 8,797 100.00

. fre age, tabulate(10)

age -- age

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

| Freq. Percent Valid Cum.

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

Valid 1 1. 1 | 130 1.48 1.48 1.48

2 2. 2 | 124 1.41 1.41 2.89

3 3. 3 | 122 1.39 1.39 4.27

4 4. 4 | 121 1.38 1.38 5.65

5 5. 5 | 113 1.28 1.28 6.93

6 6. 6 | 100 1.14 1.14 8.07

7 7. 7 | 104 1.18 1.18 9.25

8 8. 8 | 91 1.03 1.03 10.29

9 9. 9 | 99 1.13 1.13 11.41

10 10. 10 | 76 0.86 0.86 12.28

: | : : : :

85 85. 85 | 24 0.27 0.27 98.33

86 86. 86 | 19 0.22 0.22 98.54

87 87. 87 | 23 0.26 0.26 98.81

88 88. 88 | 16 0.18 0.18 98.99

89 89. 89 | 17 0.19 0.19 99.18

90 90. 90 (90+ in 1980 and 1990) | 8 0.09 0.09 99.27

91 91. 91 | 11 0.13 0.13 99.40

93 93. 93 | 3 0.03 0.03 99.43

94 94. 94 | 7 0.08 0.08 99.51

95 95. 95 | 43 0.49 0.49 100.00

Total | 8797 100.00 100.00

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

b) **su inctot\_bcq diff\_dum race\_eth sex educ5cat age**

**mdesc inctot\_bcq diff\_dum race\_eth sex educ5cat age**

**mvpatterns inctot\_bcq diff\_dum race\_eth sex educ5cat age**

su inctot\_bcq diff\_dum race\_eth sex educ5cat age

Variable | Obs Mean Std. Dev. Min Max

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

inctot\_bcq | 8,797 2.502103 1.118655 1 4

diff\_dum | 8,300 .0938554 .291645 0 1

race\_eth | 8,797 2.130272 1.546266 1 7

sex | 8,797 1.502785 .5000207 1 2

educ5cat | 8,797 2.676481 1.355821 1 5

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

age | 8,797 34.76515 21.0237 1 95

. mdesc inctot\_bcq diff\_dum race\_eth sex educ5cat age

Variable | Missing Total Percent Missing

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

inctot\_bcq | 0 8,797 0.00

diff\_dum | 497 8,797 5.65

race\_eth | 0 8,797 0.00

sex | 0 8,797 0.00

educ5cat | 0 8,797 0.00

age | 0 8,797 0.00

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

. mvpatterns inctot\_bcq diff\_dum race\_eth sex educ5cat age

variables with no mv's: inctot\_bcq race\_eth sex educ5cat age

Variable | type obs mv variable label

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

diff\_dum | float 8300 497 dummy for difficulty from two item scale

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

Patterns of missing values

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

| \_pattern \_mv \_freq |

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

| + 0 8300 |

| . 1 497 |

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

Since none of the variables are missing more than 10 percent values, so we are good to make an analytical flag of 8300 values.

**c)**

**mark inctot\_bcq\_flag**

**markout inctot\_bcq\_flag inctot\_bcq diff\_dum race\_eth sex educ5cat age**

**label variable inctot\_bcq\_flag "flag for total income (NJC ACS Data (N= 8300)"**

**tab inctot\_bcq\_flag , miss**

tab inctot\_bcq\_flag , miss

flag for |

total |

income (NJC |

ACS Data |

(N= 8300) | Freq. Percent Cum.

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

0 | 497 5.65 5.65

1 | 8,300 94.35 100.00

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

Total | 8,797 100.00

**d)**

**ologit inctot\_bcq i.diff\_dum if inctot\_bcq\_flag == 1, or**

**estimates store inctotologit1**

**ologit inctot\_bcq i.race\_eth ib2.sex if inctot\_bcq\_flag== 1, or**

**estimates store inctotologit2**

**ologit inctot\_bcq i.race\_eth ib2.sex ib5.educ5cat age if inctot\_bcq\_flag == 1, or**

**estimates store inctotologit3**

**esttab inctotologit1 inctotologit2 inctotologit3, varwidth(30) eform label star(+ 0.10 \* 0.05 \*\* 0.01 \*\*\* 0.001) stats(N r2\_p p) not compress nogaps title("Multivariate Ordinal Logistic Regression Models" "Predicting income Based on Race/Ethnicity and Sociodemographic Characteristics," "2017 New Jersey American Community Survey") mtitle("Model 1" "Model 2" "Model 3")**

**e)**

Multivariate Ordinal Logistic Regression Models Predicting income Based on Race/Ethnicity and Sociodemographic Characteristics, 2017 New Jersey American Community Survey

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

Model 1 Model 2 Model 3

OR OR OR

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

Total Income Quartile Range (back coded)

**Difficulty Dummy (ref = No difficulty)**

1. Yes Difficulty **0.486\*\*\*** **0.508\*\*\* 0.519\*\*\***

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

2. NH Black **0.500\*\*\*** **0.562\*\*\***

3. hispanic/latino, any race **0.813\*\*\*** **0.812\*\***

4. NH Nat Amer or Alask Nat 0.539 0.563

5. NH Asian or Pac Islander **1.111 0.759\*\*\***

6. NH Other Race 1.564 1.276

7. NH 2 or More Race Groups 0.893 0.865

**Sex (ref =Female)**

1. male **1.549\*\*\* 1.624\*\*\***

**Education (ref = Graduate Education)**

1. less than HS diploma 1.142

2. HS diploma or equivalent  **0.154\*\*\***

3. some college or associate's  **0.223\*\*\***

4. Bachelor's degree  **0.540\*\*\***

**Age** **1.002\***

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

/

cut1 **0.330\*\*\* 0.394\*\*\* 0.117\*\*\***

cut2 **1.039+ 1.260\*\*\* 0.423\*\*\***

cut3 **3.642\*\*\* 4.485\*\*\* 1.839\*\*\***

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

N 8300 8300 8300

r2\_p 0.00549 0.0166 0.0782

p 3.38e-29 2.09e-77 0 ---------------------------------------------------------------------

Exponentiated coefficients

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

**f) Writeup:**

A nested multivariate ordinal logistic regression model series was conducted to predict income in quartiles, based on race/ethnicity, other sociodemographic characteristics, and presence of difficulty (including cognitive and ambulatory difficulty only). Model 1 included presence of difficulty as a dummy variable; Model 2 included race/ethnicity (NH White, NH Black, Hispanic, NH Asian, NH Native American, Other, 2 or more races) and sex; Model 3 included education (less than high school, high school diploma or equivalent, some college or associate’s, bachelor’s degree and graduate education) and age (range: 1-95).

In Model 1, presence of difficulty was a strong predictor of income. The respondents having a difficulty had 51.4 % lower odds of appearing in a higher quartile of income compared to those without a difficulty (*p<0.001*). Model 1 was significant (*p*<0.001) and explained 0.55% of the variation in income.

Model 2, after adding for race and sex found similar results for difficulty presence. Those with a difficulty have 49.2 % lower odds of appearing in a high quartile of income compared to those without a difficulty (*p <0.001*)*,* when controlling for race and sex. It also showed that NH Blacks had 50 percent lower (*p<0.001*) and Hispanics/Latino (*p < 0.001*) had 28.7% lower odds of appearing in a higher income quartile compared to NH Whites, when controlling the sex and difficulty presence of respondents. There was no significant difference in the odds of appearing in a higher income quartile for NH Native Americans (OR = 0.54, *p =0.217*), NH Asians (OR = 1.11, *p =0.108*), NH Other races (OR = 1.56, *p =0.143*), and NH 2 or more race groups (OR = 0.89, *p =0.45*), versus NH Whites. However, Sex was a strong predictor of income, controlling for difficulty presence and race/ethnicity. Males had 54.9% higher odds of being in a higher income quartile compared to women *(p<0.001*). Model 2 was also significant *(p<0.001*) and explained 1.66% of the variation in income.

Model 3 after adding for age and educational level that the presence of difficulty remains as a strong predictor of income controlling for rest of the variables in the model. Compared to those without a difficulty, respondents having a difficulty had 48.2 % lower odds of appearing in a higher quartile of income (*p <0.001*). NH Blacks had 43.2% lower odds of being in a higher income quartile compared to NH Whites (*p <0.001*). Hispanics had 18.8% lower odds of being in a higher income quartile compared to NH Whites (*p <0.001*). Unlike Model2, NH Asians had significantly 34% lower odds of being in a higher income quartile compared to NH Whites (*p <0.001*). There was no significant difference in odds between NH Native Americans (OR= 0.560, (*p <0.001),* Others (OR= 1.27, *p=0.441),* and those of two or more races (OR=0.865, *p=0.339*) with respect to NH Whites. Moreover, like Model 2, Sex was a strong predictor of income, controlling for difficulty presence, race/ethnicity, and age. Males had 62.3% higher odds of being in a higher income quartile compared to women (*p<0.001).* Education turned out to be a significant predictor of income controlling for other variables in the model. Those with a high school diploma or equivalent have 85% lower odds of being in a higher income quartile compared to those with a graduate education*(p<0.001*). Those with some college or associates’ degree have 78% lower odds of being in a higher income quartile compared to those with a graduate education*(p<0.001*). Those with a bachelor’s degree have 46.1% lower odds of being in a higher income quartile compared to those with a graduate education*(p<0.001).* However, those with less than a high school diploma did not have significant difference in proportional odds of being in a higher income quartile compared to those with a graduate education*(p=0.103*). Age was also a significant predictor of income; an increase in age of one year increases the proportional odds of being in a higher income quartile by 0.2 %*(p=0.037)* when controlling for race/ethnicity, sex, education, and difficulty presence. Model 3 was significant *(p<0.001)* and explained 7.8% of the variation in income.

1. **(50 points) Using any dataset of your choice, design your own nested multivariate ordinal logistic regression model series (4 models). There must be at least 7 IVs in the model series and your DV must be a variable we have not used in class or for the problem sets before. Your final analytic sample must include at least 400 cases. Remember to choose a DV that is appropriate for ordinal logistic regression (be sure the variables are ordinal -- distance between categories is unknown).**

**Your analysis should include:**

1. **Univariate exploration of all analytic variables:** run the appropriate commands for a variable based on its variable-type, i.e. categorical vs. interval ratio and submit correct output for each
   1. **Be sure to include frequencies for the new dummy variable**
2. **Missing data analysis of analytic variables:** note any patterns or concerns you have; investigate abnormal results and hypothesize reasons for missing data
3. **Code and output for proper analytic flag used throughout Models**
4. **An edited table from esttab command output. Be sure the table shows proportional 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)**

**=> USING GSS 2016 dataset:**

**recode snsmoth (2 =0) (.n =.n) (.i = .i) (.d = .d), gen(social)**

**recode absingle (2 =0) (.n =.n) (.i = .i) (.d = .d), gen(single)**

**label values social single yngss**

**numlabel yngss, add force**

**tab1 single social, miss**

**recode happy (3= 0) (2=1) (1 =2), gen(happy\_rc)**

**label variable happy\_rc "reverse coded variable of happiness level (ordinal)"**

**label define hpy 0 "Not too happy" 1 "Pretty happy" 2 "Very happy" .d "DK" .n "NA"**

**label values happy\_rc hpy**

**numlabel hpy, add**

**tab happy\_rc, miss**

**tab1 happy\_rc sex race\_eth single social, miss**

**fre age educ conrinc10k, tabulate(10)**

. tab1 happy\_rc sex race\_eth single social, miss

reverse coded |

variable of |

happiness level |

(ordinal) | Freq. Percent Cum.

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

0. Not too happy | 452 15.77 15.77

1. Pretty happy | 1,601 55.84 71.61

2. Very happy | 806 28.11 99.72

.d. DK | 3 0.10 99.83

.n. NA | 5 0.17 100.00

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

Total | 2,867 100.00

-> tabulation of sex

respondents |

sex | Freq. Percent Cum.

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

1. male | 1,276 44.51 44.51

2. female | 1,591 55.49 100.00

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

Total | 2,867 100.00

-> tabulation of race\_eth

recode of race variable | Freq. Percent Cum.

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

1. NH White | 1,893 66.03 66.03

2. NH Black | 468 16.32 82.35

3. Hispanic | 369 12.87 95.22

4. Other | 135 4.71 99.93

.u. unknown Hisp origin | 2 0.07 100.00

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

Total | 2,867 100.00

-> tabulation of single

RECODE of |

absingle (not |

married) | Freq. Percent Cum.

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

0. No | 1,051 36.66 36.66

1. Yes | 782 27.28 63.93

.d. DK | 30 1.05 64.98

.i. IAP | 977 34.08 99.06

.n. NA | 27 0.94 100.00

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

Total | 2,867 100.00

-> tabulation of social

RECODE of |

snsmoth1 (do you |

use any other |

social networks) | Freq. Percent Cum.

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

0. No | 1,209 42.17 42.17

1. Yes | 165 5.76 47.92

.i. IAP | 1,483 51.73 99.65

.n. NA | 10 0.35 100.00

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

Total | 2,867 100.00

. fre age educ conrinc10k, tabulate(10)

age -- age of respondent

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

| Freq. Percent Valid Cum.

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

Valid 18 | 7 0.24 0.25 0.25

19 | 33 1.15 1.16 1.40

20 | 26 0.91 0.91 2.31

21 | 33 1.15 1.16 3.47

22 | 44 1.53 1.54 5.01

23 | 49 1.71 1.72 6.72

24 | 35 1.22 1.23 7.95

25 | 56 1.95 1.96 9.91

26 | 42 1.46 1.47 11.38

27 | 58 2.02 2.03 13.41

: | : : : :

80 | 25 0.87 0.88 95.73

81 | 21 0.73 0.74 96.46

82 | 11 0.38 0.39 96.85

83 | 22 0.77 0.77 97.62

84 | 11 0.38 0.39 98.00

85 | 11 0.38 0.39 98.39

86 | 12 0.42 0.42 98.81

87 | 9 0.31 0.32 99.12

88 | 3 0.10 0.11 99.23

89 89. 89 or older | 22 0.77 0.77 100.00

Total | 2857 99.65 100.00

Missing .n .n. NA | 10 0.35

Total | 2867 100.00

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

educ -- highest year of school completed

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

| Freq. Percent Valid Cum.

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

Valid 0 | 2 0.07 0.07 0.07

1 | 3 0.10 0.10 0.17

2 | 3 0.10 0.10 0.28

3 | 3 0.10 0.10 0.38

4 | 2 0.07 0.07 0.45

5 | 4 0.14 0.14 0.59

6 | 31 1.08 1.08 1.68

7 | 18 0.63 0.63 2.31

8 | 48 1.67 1.68 3.99

9 | 59 2.06 2.06 6.05

10 | 90 3.14 3.15 9.20

11 | 118 4.12 4.13 13.33

12 | 824 28.74 28.83 42.16

13 | 242 8.44 8.47 50.63

14 | 359 12.52 12.56 63.19

15 | 137 4.78 4.79 67.98

16 | 485 16.92 16.97 84.95

17 | 108 3.77 3.78 88.73

18 | 149 5.20 5.21 93.95

19 | 63 2.20 2.20 96.15

20 | 110 3.84 3.85 100.00

Total | 2858 99.69 100.00

Missing .a | 9 0.31

Total | 2867 100.00

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

conrinc10k -- income in $10k

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

| Freq. Percent Valid Cum.

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

Valid .0363 | 25 0.87 1.53 1.53

.1452 | 51 1.78 3.13 4.66

.2541 | 32 1.12 1.96 6.62

.3267 | 30 1.05 1.84 8.46

.3993 | 31 1.08 1.90 10.36

.4719 | 31 1.08 1.90 12.25

.5445 | 24 0.84 1.47 13.73

.6534 | 34 1.19 2.08 15.81

.81675 | 51 1.78 3.13 18.93

.99825 | 45 1.57 2.76 21.69

: | : : : :

2.7225 | 108 3.77 6.62 56.80

3.267 | 158 5.51 9.68 66.48

3.993 | 137 4.78 8.39 74.88

4.9005 | 141 4.92 8.64 83.52

5.9895 | 79 2.76 4.84 88.36

7.26 | 68 2.37 4.17 92.52

8.712 | 36 1.26 2.21 94.73

10.164 | 21 0.73 1.29 96.02

11.616 | 14 0.49 0.86 96.88

18.92115 | 51 1.78 3.13 100.00

Total | 1632 56.92 100.00

Missing . | 1235 43.08

Total | 2867 100.00

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

**b)**

**su happy\_rc sex race\_eth single social age educ conrinc10k**

**mdesc happy\_rc sex race\_eth single social age educ conrinc10k**

**mvpatterns happy\_rc sex race\_eth single social age educ conrinc10k**

. su happy\_rc sex race\_eth single social age educ conrinc10k

Variable | Obs Mean Std. Dev. Min Max

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

happy\_rc | 2,859 1.12382 .6517909 0 2

sex | 2,867 1.554935 .4970596 1 2

race\_eth | 2,865 1.562304 .886965 1 4

single | 1,833 .426623 .4947215 0 1

social | 1,374 .1200873 .325182 0 1

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

age | 2,857 49.15576 17.69279 18 89

educ | 2,858 13.73723 2.963886 0 20

conrinc10k | 1,632 3.482252 3.625953 .0363 18.92115

. mdesc happy\_rc sex race\_eth single social age educ conrinc10k

Variable | Missing Total Percent Missing

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

happy\_rc | 8 2,867 0.28

sex | 0 2,867 0.00

race\_eth | 2 2,867 0.07

single | 1,034 2,867 36.07

social | 1,493 2,867 52.08

age | 10 2,867 0.35

educ | 9 2,867 0.31

conrinc10k | 1,235 2,867 43.08

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

. mvpatterns happy\_rc sex race\_eth single social age educ conrinc10k

variables with no mv's: sex

Variable | type obs mv variable label

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

happy\_rc | byte 2859 8 reverse coded variable of happiness level (ordinal)

race\_eth | float 2865 2 recode of race variable

single | byte 1833 1034 RECODE of absingle (not married)

social | byte 1374 1493 RECODE of snsmoth1 (do you use any other social networks)

age | byte 2857 10 age of respondent

educ | byte 2858 9 highest year of school completed

conrinc10k | float 1632 1235 income in $10k

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

Patterns of missing values

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

| \_pattern \_mv \_freq |

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

| +++.+++ 1 621 |

| +++.++. 2 531 |

| ++.++++ 1 444 |

| +++++++ 0 436 |

| ++.+++. 2 253 |

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

Three variables, single status, social media usage, and constant income had more than 10 percent missing values. Since backcoding was not possible after evaluation and the valid data set included more than 400 cases, so this combination is carried forward.

**c)**

**mark happy\_rc\_flag**

**markout happy\_rc\_flag happy\_rc sex race\_eth single social age educ conrinc10k**

**label variable happy\_rc\_flag "flag for happiness level (GSS 2018 - N= 436)"**

**tab happy\_rc\_flag , miss**

. tab happy\_rc\_flag, miss

flag for |

happiness |

level (GSS |

2018 - N= |

436) | Freq. Percent Cum.

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

0 | 2,431 84.79 84.79

1 | 436 15.21 100.00

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

Total | 2,867 100.00

**d)**

Multivariate Ordinal Logistic Regression Models Predicting happiness level Based on Race/Ethnicity and Sociodemographic Characteristics, 2016 GSS Survey

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

Model 1 Model 2 Model 3 Model 4

OR OR OR OR

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

Happiness level (ordinal form)

**Sex (ref = male)**

2. female 0.777 0.741 **0.708+** 0.864

**Age of respondent** 1.003 1.004 0.999

**Highest year of school** **1.070\* 1.079\*** 1.037

**Marital Status (ref = No)**

1. Yes 0.766 0.729

**Use of social media (ref = No)**

1. Yes 0.825 0.905

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

2. NH Black **0.504\***

3. Hispanic 0.726

4. Other 0.628

**Income in $10k 1.088\*\***

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

/

cut1 **0.151\*\*\***  0.457 0.433 **0.236\***

cut2 **1.991\*\*\***  **6.182\*\*\*** **5.929\*\*** **3.497\***

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

N 436 436 436 436

r2\_p 0.00218 0.00812 0.0111 0.0315

p 0.173 0.0751 0.0927 0.00151

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

Exponentiated coefficients

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

**e) WRITE-UP:**

A nested multivariate ordinal logistic regression model series was conducted to predict happiness level of respondents (0/ Not too happy, 1/Pretty Happy, 2/Very Happy) based on socioeconomic and demographic variables including sex, age of respondents (18 – 89 or older), highest education year in integer (0/ lowest – 20/ Highest), race (NH Whites, Hispanics, NH Blacks, NH Others), and Income in 10000 dollars unit (0.0363/lowest – 18.92115/Highest). Model 1 included sex; Model 2 had sex plus age and educational years; Model 3 in addition to Model 2 had marital status dummy variable and usage of social media dummy variable; Model 4 in addition to Model 3 had race and income of respondents.

In model 1, females did not have significantly different proportional odds of being in higher category of happiness compared to males (OR = 0.78, *p =0.174*). Model 1 itself was not significant (*p = 0.17*) and accounted for 0.22 % of variation in happiness level of respondents.

Model 2, after adding age and education found sex to remain insignificant predictor of happiness; females still do not have significantly different odds of being in a higher category of happiness than males, when controlling for age and education (OR = 0.74, *p =0.112*). Age, when controlling sex and educational years, also was not significantly associated with odds of being in a higher category of happiness (OR = 1.00, *p =0.62*). However, when sex and age of respondents was controlled, an increase in educational years of respondents increases the odds of being in a higher category of happiness by 7% (*p =0.033*). Model 3 was marginally significant (*p =0.07*) and accounted for 0.8% of variation in happiness level of respondents.

Model 3 added for marital status dummy variable and social media usage dummy variable. It found that sex of respondents changed into a significant predictor of happiness level. Compared to males, females had marginally 30 percent lower proportional odds of being in a higher happiness category (*p =0.07*), when controlling for age, educational years, marital status, and social media usage of respondents. Similarly, like Model 2, each additional educational year increased the odds of being in a higher category of happiness by 8 % (*p =0.019*), when controlling for the rest of the variables. But age remained insignificantly associated with proportional odds of being in a higher category of happiness (OR = 1, *p =0.58*). Model 3 found that there is no significant difference in proportional odds of being in a higher category of happiness between those who are married and those who are not (OR = 0.77, *p =0.169*) when controlling for the rest of the variables therein. Likewise, there is no significant difference in proportional odds of being in a higher category of happiness between those who use social media and those who do not (OR = 0.83, *p =0.537*), when controlling for the rest of the variables therein. Model 3 was only marginally significant (*p =0.09*) and accounted for 1.11% variation in happiness level of respondents.

Model 4 added for race and constant income in 10000 dollars of respondents. Converse to Model 3, sex of respondents changed into an insignificant predictor of happiness level when controlling for the rest of the variables in Model 4 (OR =0.86, *p =0.46*). Age remained insignificantly associated with proportional odds of being in a higher category of happiness (OR = 1.0, *p =0.908*), when controlling for all other variables in the model. Unlike Model 3, education became insignificantly associated with proportional odds of being in a higher category of happiness (OR = 1.03, *p =0.3*), when controlling for all other variables in the model. However, similar to Model 3, there is no significant difference in proportional odds of being in a higher category of happiness between those who are married and those who are not (OR = 0.73, *p =0.109*) when controlling for the rest of the variables therein. Likewise, there is no significant difference in proportional odds of being in a higher category of happiness between those who use social media and those who do not (OR = 0.90, *p =0.75*), when controlling for the rest of the variables therein. But race turned out to be significantly associated with the happiness level: Among the races, while controlling for all other variables, NH Blacks had 49.6 percent lower odds of being in a higher category of happiness than NH Whites (*p =0.015*). But Hispanics (OR = 0.73, *p =0.29*) and NH Others (OR = 0.63, *p =0.29*) did not have significantly different odds of being in a higher category of happiness compared to NH Whites. And lastly an addition of 10000 dollars in income of respondents increased the odds of being in a higher category of happiness by 8.8% (*p =0.004*). Model 4 was statistically significant (*p =0.0015*) and accounted for 3.15% variation in happiness of respondents.