**RUTGERS UNIVERSITY**

**Bloustein School of Planning and Public Policy**

**Applied Multivariate Methods**

**Fall 2020**

**Dawne Mouzon, Ph.D.**

**Problem Set #10:**

**Multivariate Ordinary Least Squares Regression III**

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

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

**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.**

**MULTIVARIATE ORDINARY LEAST SQUARES REGRESSION III**

**(100 points)**

**Use 2018 General Social Survey for all questions.**

1. **(5 pts) Please create a new variable (*bmi2018*) based on the formula/techniques we used in the Advanced Recoding and Computing Lab. Please submit:**
   1. **Frequencies on the new variable (show only first and last ~10 rows of output)**
   2. **Extended descriptives on the new variable**
   3. **Appropriate consistency check (for two input variables and one output variable). Show only 5-10 cases. NOTE: Only check the scenario for no missing values on any variable; no need to check the other two scenarios.**

**tab1 weight height, miss**

**gen bmi2018 = (weight/((height\*height))) \* 703**

**label variable bmi2018 "2018-BMI"**

**replace bmi2018= .d if weight== .d & height == .d**

**replace bmi2018= .n if weight== .n & height == .n**

**replace bmi2018= .i if weight== .i & height == .i**

**label define bmimiss .i "IAP" .d "don’t know" .n "no answer"**

**label values bmi2018 bmimiss**

**numlabel bmimiss, add**

**tab bmi2018, miss**

**fre bmi2018, tabulate(6)**

**sum bmi2018, detail**

**list bmi2018 height weight if !missing(bmi2018) & !missing(height) & !missing(weight) in 1/10, nolabel**

. fre bmi2018, tabulate(6)

bmi2018 -- 2018-BMI

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

| Freq. Percent Valid Cum.

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

Valid 14.765805244446 | 1 0.04 0.07 0.07

15.504096031189 | 1 0.04 0.07 0.15

15.545004844666 | 1 0.04 0.07 0.22

16.871999740601 | 1 0.04 0.07 0.29

17.149559020996 | 1 0.04 0.07 0.36

17.629243850708 | 1 0.04 0.07 0.44

: | : : : :

52.680164337158 | 1 0.04 0.07 99.64

53.950313568115 | 1 0.04 0.07 99.71

54.548667907715 | 1 0.04 0.07 99.78

56.110061645508 | 1 0.04 0.07 99.85

56.670406341553 | 1 0.04 0.07 99.93

60.351196289063 | 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

. sum bmi2018, detail

2018-BMI

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

Percentiles Smallest

1% 18.45726 14.76581

5% 20.25244 15.5041

10% 21.45386 15.545 Obs 1,374

25% 23.74578 16.872 Sum of Wgt. 1,374

50% 27.36592 Mean 28.3612

Largest Std. Dev. 6.318966

75% 31.32101 54.54867

90% 36.8409 56.11006 Variance 39.92933

95% 40.34665 56.67041 Skewness 1.111694

99% 48.78378 60.3512 Kurtosis 4.878616

. list bmi2018 height weight if !missing(bmi2018) & !missing(height) & !missing(weight) in 1/10,

> nolabel

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

| bmi2018 height weight |

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

1. | 31.001125335693 73 235 |

3. | 34.207397460938 68 225 |

4. | 25.845588684082 68 170 |

8. | 21.787189483643 66 135 |

10. | 28.49800491333 69 193 |

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

1. **(5 pts) Please round both *bmi2018* and *sei10* into new variables (*bmi2018rd* and *sei10rd*). Submit:**
   1. **extended descriptives for the original and recoded variables**
   2. **your code for the whole question**

**gen bmi2018rd = round(bmi2018, 1)**

**label variable bmi2018rd "BMI rounded to whole number"**

**list bmi2018 bmi2018rd in 1/10, nolabel**

**sum bmi2018 bmi2018rd, detail**

**gen sei10rd = round(sei10, 1)**

**label variable sei10rd "SEI rounded to whole number"**

**list sei10 sei10rd in 1/10, nolabel**

**sum sei10 sei10rd, detail**

. list bmi2018 bmi2018rd in 1/10, nolabel

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

| bmi2018 bmi201~d |

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

1. | 31.001125335693 31 |

2. | .i .i |

3. | 34.207397460938 34 |

4. | 25.845588684082 26 |

5. | .i .i |

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

6. | .i .i |

7. | . . |

8. | 21.787189483643 22 |

9. | . . |

10. | 28.49800491333 28 |

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

. sum bmi2018 bmi2018rd, detail

2018-BMI

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

Percentiles Smallest

1% 18.45726 14.76581

5% 20.25244 15.5041

10% 21.45386 15.545 Obs 1,374

25% 23.74578 16.872 Sum of Wgt. 1,374

50% 27.36592 Mean 28.3612

Largest Std. Dev. 6.318966

75% 31.32101 54.54867

90% 36.8409 56.11006 Variance 39.92933

95% 40.34665 56.67041 Skewness 1.111694

99% 48.78378 60.3512 Kurtosis 4.878616

BMI rounded to whole number

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

Percentiles Smallest

1% 18 15

5% 20 16

10% 21 16 Obs 1,374

25% 24 17 Sum of Wgt. 1,374

50% 27 Mean 28.35007

Largest Std. Dev. 6.330921

75% 31 55

90% 37 56 Variance 40.08056

95% 40 57 Skewness 1.119914

99% 49 60 Kurtosis 4.869881

. list sei10 sei10rd in 1/10, nolabel

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

| sei10 sei10rd |

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

1. | 65.29999999999999716 65 |

2. | 14.80000000000000071 15 |

3. | 83.40000000000000568 83 |

4. | 69.29999999999999716 69 |

5. | 68.59999999999999432 69 |

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

. sum sei10 sei10rd, detail

r's socioeconomic index (2010)

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

Percentiles Smallest

1% 12.6 10.6

5% 14 10.6

10% 19.7 10.6 Obs 2,248

25% 25.2 12.6 Sum of Wgt. 2,248

50% 41 Mean 46.94448

Largest Std. Dev. 23.02837

75% 65.45 92.8

90% 81 92.8 Variance 530.3057

95% 84.5 92.8 Skewness .2939409

99% 91.1 92.8 Kurtosis 1.804432

SEI rounded to whole number

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

Percentiles Smallest

1% 13 11

5% 14 11

10% 20 11 Obs 2,248

25% 25 13 Sum of Wgt. 2,248

50% 41 Mean 47.00845

Largest Std. Dev. 22.9892

75% 65.5 93

90% 81 93 Variance 528.5033

95% 85 93 Skewness .2983242

99% 91 93 Kurtosis 1.805381

1. **(90 pts) Please run a series of nested multivariate ordinary least squares regression models predicting rounded and bottom-coded BMI (*bmi2018rd\_bc; see part f*). Model 1 should include age in categories (*age4cat*; ref=18-24), Model 2 should add sex (*sex*; ref=female), race/ethnicity (*race\_eth*; ref=NH White), and region (*region4cat*; ref=South), and Model 3 should include rounded and top-coded socioeconomic index (*sei10rd\_tc; see part g*).**

**Please complete parts a-j in the order they are asked. For this problem, please submit (in this order, after each part):**

**For parts a-e, please use the following variables: *bmi2018rd*, *age4cat*, *sex*, *race\_eth*, *region4cat*, *sei10rd***

1. **A small Word table that lists all variables in your model series. Include four columns for the model #, variable name, variable description, and categories/values on the variable.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model #** | **Variable Name** | **Variable Description** | **Categories/ Values** |
| 1 | Age4cat (ref =18-24) | Age in nominal categories | [(25 – 55), (45-64), (65 or older)] years |
| 2 | Sex (ref = Female) | Gender Category | Male |
| 2 | Race\_eth (ref=NH White) | Ethnicity/ Race Categories | NH Black, Hispanic, Other |
| 2 | region4cat (ref=South) | Regions of respondents | Northeast, Midwest, West |
| 3 | Sei10rd\_tc | Social economic index for 2010 | Interval/Ratio |

1. **Univariate analysis of all analytic variables:**
   1. **frequencies for the DV and the categorical IVs**
   2. **fre command for interval-ratio IVs**

**tab1 bmi2018rd age4cat sex race\_eth region4cat, miss**

**fre sei10rd**

. tab1 bmi2018rd age4cat sex race\_eth region4cat, miss

-> tabulation of bmi2018rd

BMI rounded to |

whole number | Freq. Percent Cum.

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

15 | 1 0.04 0.04

16 | 2 0.09 0.13

17 | 2 0.09 0.21

18 | 9 0.38 0.60

19 | 23 0.98 1.58

20 | 44 1.87 3.45

21 | 59 2.51 5.96

22 | 88 3.75 9.71

23 | 90 3.83 13.54

24 | 94 4.00 17.55

25 | 91 3.88 21.42

26 | 93 3.96 25.38

27 | 123 5.24 30.62

28 | 83 3.53 34.16

29 | 76 3.24 37.39

30 | 91 3.88 41.27

31 | 73 3.11 44.38

32 | 42 1.79 46.17

33 | 44 1.87 48.04

34 | 37 1.58 49.62

35 | 30 1.28 50.89

36 | 30 1.28 52.17

37 | 28 1.19 53.36

38 | 20 0.85 54.22

39 | 15 0.64 54.86

40 | 21 0.89 55.75

41 | 16 0.68 56.43

42 | 6 0.26 56.69

43 | 6 0.26 56.94

44 | 8 0.34 57.28

45 | 4 0.17 57.45

46 | 7 0.30 57.75

47 | 2 0.09 57.84

48 | 1 0.04 57.88

49 | 3 0.13 58.01

50 | 1 0.04 58.05

51 | 2 0.09 58.13

52 | 3 0.13 58.26

53 | 1 0.04 58.30

54 | 1 0.04 58.35

55 | 1 0.04 58.39

56 | 1 0.04 58.43

57 | 1 0.04 58.48

60 | 1 0.04 58.52

. | 34 1.45 59.97

.d. don’t know | 4 0.17 60.14

.i. IAP | 900 38.33 98.47

.n. no answer | 36 1.53 100.00

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

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 region4cat

categorical variable for regions | Freq. Percent Cum.

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

1. Northeast (New Engl and Mid Atl) | 356 15.16 15.16

2. Midwest (W-No Central and E-No Centr | 513 21.85 37.01

3. South (S Atl, E-S Central, W-S Centr | 953 40.59 77.60

4. West (Mountain and Pacific) | 526 22.40 100.00

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

Total | 2,348 100.00

. fre sei10rd, tabulate(6)

sei10rd -- SEI rounded to whole number

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

| Freq. Percent Valid Cum.

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

Valid 11 | 3 0.13 0.13 0.13

13 | 42 1.79 1.87 2.00

14 | 78 3.32 3.47 5.47

15 | 17 0.72 0.76 6.23

16 | 2 0.09 0.09 6.32

17 | 17 0.72 0.76 7.07

: | : : : :

88 | 19 0.81 0.85 97.86

89 | 14 0.60 0.62 98.49

90 | 1 0.04 0.04 98.53

91 | 19 0.81 0.85 99.38

92 | 9 0.38 0.40 99.78

93 | 5 0.21 0.22 100.00

Total | 2248 95.74 100.00

Missing .i | 100 4.26

Total | 2348 100.00

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

1. **An analysis of missing data patterns, including output from simple descriptives, mdesc, and mvpatterns. 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 possible, and write up your methods in a few sentences.**

**sum bmi2018rd age4cat sex race\_eth region4cat sei10rd**

**mdesc bmi2018rd age4cat sex race\_eth region4cat sei10rd**

**mvpatterns bmi2018rd age4cat sex race\_eth region4cat sei10rd**

**tab bmi2018rd age4cat, miss**

**tab bmi2018rd sex, miss**

**fre height weight, tabulate(6)**

**recode height (min/60 =1) (61/81 =2), gen(new\_height)**

**recode weight (min/150 =1) (151/395 =2), gen(new\_weight)**

**tab1 new\_height new\_weight, miss**

**tab new\_height new\_weight, miss**

. sum bmi2018rd age4cat sex race\_eth region4cat sei10rd

Variable | Obs Mean Std. Dev. Min Max

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

bmi2018rd | 1,374 28.35007 6.330921 15 60

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

region4cat | 2,348 2.7023 .9801845 1 4

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

sei10rd | 2,248 47.00845 22.9892 11 93

. mdesc bmi2018rd age4cat sex race\_eth region4cat sei10rd

Variable | Missing Total Percent Missing

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

bmi2018rd | 974 2,348 41.48

age4cat | 7 2,348 0.30

sex | 0 2,348 0.00

race\_eth | 15 2,348 0.64

region4cat | 0 2,348 0.00

sei10rd | 100 2,348 4.26

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

. mvpatterns bmi2018rd age4cat sex race\_eth region4cat sei10rd

variables with no mv's: sex region4cat

Variable | type obs mv variable label

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

bmi2018rd | float 1374 974 BMI rounded to whole number

age4cat | float 2341 7 dummy age

race\_eth | float 2333 15 recode of race variable

sei10rd | float 2248 100 SEI rounded to whole number

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

Patterns of missing values

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

| \_pattern \_mv \_freq |

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

| ++++ 0 1357 |

| .+++ 1 870 |

| .++. 2 92 |

| .+.+ 2 8 |

| +++. 1 7 |

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

| ++.+ 1 6 |

| +.++ 1 4 |

| ..++ 2 3 |

| .+.. 3 1 |

It turns out that the dependent variable (bmi2018rd) has more than 10 percent missing values (40%). This reduces are analytical flag to N = 1357. So, our effort will be towards reclaiming lost values through back-coding by running tabs for bmi2018 with categorical independent variables.

. tab bmi2018rd age4cat, miss

BMI rounded to | dummy age

whole number | 1. 18-24 2. 25-55 3. 45-64 4. 65 or .n. NA | Total

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

15 | 0 0 1 0 0 | 1

16 | 1 1 0 0 0 | 2

17 | 1 1 0 0 0 | 2

18 | 4 3 2 0 0 | 9

19 | 6 14 2 1 0 | 23

20 | 2 23 14 4 1 | 44

21 | 9 31 16 2 1 | 59

22 | 13 43 26 6 0 | 88

23 | 11 37 32 10 0 | 90

24 | 9 46 29 9 1 | 94

25 | 8 44 30 9 0 | 91

26 | 8 31 42 12 0 | 93

27 | 5 60 50 7 1 | 123

28 | 5 38 29 11 0 | 83

29 | 3 31 36 6 0 | 76

30 | 1 35 45 10 0 | 91

31 | 3 27 33 10 0 | 73

32 | 3 21 17 1 0 | 42

33 | 6 21 14 3 0 | 44

34 | 1 15 18 3 0 | 37

35 | 1 15 10 4 0 | 30

36 | 3 16 11 0 0 | 30

37 | 4 11 12 1 0 | 28

38 | 1 10 8 1 0 | 20

39 | 2 6 6 1 0 | 15

40 | 1 13 7 0 0 | 21

41 | 0 6 9 1 0 | 16

42 | 0 3 3 0 0 | 6

43 | 0 3 3 0 0 | 6

44 | 0 5 3 0 0 | 8

45 | 0 2 2 0 0 | 4

46 | 1 6 0 0 0 | 7

47 | 0 1 1 0 0 | 2

48 | 0 1 0 0 0 | 1

49 | 0 1 2 0 0 | 3

50 | 0 1 0 0 0 | 1

51 | 0 2 0 0 0 | 2

52 | 0 2 1 0 0 | 3

53 | 0 1 0 0 0 | 1

54 | 0 0 1 0 0 | 1

55 | 0 1 0 0 0 | 1

56 | 0 0 1 0 0 | 1

57 | 0 0 1 0 0 | 1

60 | 1 0 0 0 0 | 1

. | 1 18 11 3 1 | 34

.d. don’t know | 1 2 1 0 0 | 4

.i. IAP | 80 188 210 422 0 | 900

.n. no answer | 2 15 15 2 2 | 36

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

Total | 197 851 754 539 7 | 2,348

. tab bmi2018rd sex, miss

BMI rounded to | respondents sex

whole number | 1. male 2. female | Total

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

15 | 1 0 | 1

16 | 1 1 | 2

17 | 2 0 | 2

18 | 3 6 | 9

19 | 8 15 | 23

20 | 16 28 | 44

21 | 18 41 | 59

22 | 37 51 | 88

23 | 39 51 | 90

24 | 59 35 | 94

25 | 47 44 | 91

26 | 52 41 | 93

27 | 65 58 | 123

28 | 46 37 | 83

29 | 46 30 | 76

30 | 54 37 | 91

31 | 38 35 | 73

32 | 25 17 | 42

33 | 14 30 | 44

34 | 19 18 | 37

35 | 14 16 | 30

36 | 15 15 | 30

37 | 8 20 | 28

38 | 8 12 | 20

39 | 2 13 | 15

40 | 8 13 | 21

41 | 6 10 | 16

42 | 2 4 | 6

43 | 3 3 | 6

44 | 2 6 | 8

45 | 1 3 | 4

46 | 3 4 | 7

47 | 2 0 | 2

48 | 1 0 | 1

49 | 1 2 | 3

50 | 0 1 | 1

51 | 0 2 | 2

52 | 2 1 | 3

53 | 1 0 | 1

54 | 0 1 | 1

55 | 0 1 | 1

56 | 1 0 | 1

57 | 1 0 | 1

60 | 0 1 | 1

. | 7 27 | 34

.d. don’t know | 1 3 | 4

.i. IAP | 359 541 | 900

.n. no answer | 14 22 | 36

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

Total | 1,052 1,296 | 2,348

. fre height weight, tabulate(6)

height -- r is how tall

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

| Freq. Percent Valid Cum.

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

Valid 48 | 1 0.04 0.07 0.07

55 | 1 0.04 0.07 0.14

56 | 2 0.09 0.14 0.29

57 | 1 0.04 0.07 0.36

58 | 2 0.09 0.14 0.50

59 | 15 0.64 1.07 1.57

: | : : : :

76 | 13 0.55 0.93 99.00

77 | 8 0.34 0.57 99.57

78 | 2 0.09 0.14 99.71

79 | 1 0.04 0.07 99.79

80 | 2 0.09 0.14 99.93

81 | 1 0.04 0.07 100.00

Total | 1402 59.71 100.00

Missing .d .d. Don't know | 9 0.38

.i .i. IAP | 900 38.33

.n .n. No answer | 37 1.58

Total | 946 40.29

Total | 2348 100.00

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

weight -- r weighs how much

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

| Freq. Percent Valid Cum.

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

Valid 85 | 1 0.04 0.07 0.07

98 | 2 0.09 0.14 0.22

100 | 3 0.13 0.22 0.43

103 | 1 0.04 0.07 0.51

105 | 6 0.26 0.43 0.94

106 | 1 0.04 0.07 1.01

: | : : : :

342 | 1 0.04 0.07 99.35

344 | 1 0.04 0.07 99.42

350 | 2 0.09 0.14 99.57

365 | 1 0.04 0.07 99.64

380 | 4 0.17 0.29 99.93

395 | 1 0.04 0.07 100.00

Total | 1380 58.77 100.00

Missing .d .d. Don't know | 11 0.47

.i .i. IAP | 900 38.33

.n .n. No answer | 57 2.43

Total | 968 41.23

Total | 2348 100.00

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

When bmi2018 was investigated with other variables no potential was found for back-coding. So, the variables height and weight were then investigated whereupon it was hypothesized that it is possible for either height to have a missing variable while the same person is able to give the weight, or vice versa, which could result in wrongfully inputting the values as missing values. Upon running cross tabs for height and weight, after recoding them, it was found that the number of values that could be backcoded were not substantial. Thus, we will stick with our current set.

. tab new\_height new\_weight, miss

RECODE of |

height (r |

is how | RECODE of weight (r weighs how much)

tall) | 1 2 .d .i .n | Total

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

1 | 31 23 0 0 1 | 55

2 | 353 967 7 0 20 | 1,347

.d | 3 2 4 0 0 | 9

.i | 0 0 0 900 0 | 900

.n | 1 0 0 0 36 | 37

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

Total | 388 992 11 900 57 | 2,348

1. **Create an analytic flag for this model series. Paste your code and frequencies for the flag below.**

**mark bmi2018rd\_flag**

**markout bmi2018rd\_flag bmi2018rd age4cat sex race\_eth region4cat sei10rd**

**label variable bmi2018rd\_flag "Analtical flag for 2018 BMI N= 1357"**

**tab bmi2018rd\_flag, miss**

. tab bmi2018rd\_flag, miss

bmi2018rd\_f |

lag | Freq. Percent Cum.

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

0 | 991 42.21 42.21

1 | 1,357 57.79 100.00

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

Total | 2,348 100.00

1. **An assessment of outliers (interval-ratio variables only). Be sure to limit this analysis to the analytic sample only and use only the rounded variables.**
   1. **Histograms**
   2. **Frequencies**
   3. **Extended Descriptives**

**histogram bmi2018rd if bmi2018rd\_flag ==1, freq**

**histogram sei10rd if bmi2018rd\_flag ==1, freq**

**sum sei10rd bmi2018rd if bmi2018rd\_flag == 1, detail**

**fre sei10rd if bmi2018rd\_flag == 1, tabulate(10)**

**fre bmi2018rd if bmi2018rd\_flag ==1, tabulate(10)**

****

. sum sei10rd bmi2018rd if bmi2018rd\_flag == 1, detail

SEI rounded to whole number

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

Percentiles Smallest

1% 13 11

5% 14 11

10% 21 13 Obs 1,357

25% 26 13 Sum of Wgt. 1,357

50% 43 Mean 47.91157

Largest Std. Dev. 23.03296

75% 68 92

90% 83 92 Variance 530.5172

95% 85 93 Skewness .2450231

99% 91 93 Kurtosis 1.781062

BMI rounded to whole number

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

Percentiles Smallest

1% 18 15

5% 20 16

10% 21 16 Obs 1,357

25% 24 17 Sum of Wgt. 1,357

50% 27 Mean 28.3692

Largest Std. Dev. 6.346789

75% 31 55

90% 37 56 Variance 40.28173

95% 40 57 Skewness 1.114828

99% 49 60 Kurtosis 4.851942

. fre sei10rd if bmi2018rd\_flag == 1, tabulate(10)

sei10rd -- SEI rounded to whole number

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

| Freq. Percent Valid Cum.

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

Valid 11 | 2 0.15 0.15 0.15

13 | 24 1.77 1.77 1.92

14 | 47 3.46 3.46 5.38

15 | 10 0.74 0.74 6.12

17 | 8 0.59 0.59 6.71

18 | 3 0.22 0.22 6.93

19 | 12 0.88 0.88 7.81

20 | 28 2.06 2.06 9.87

21 | 37 2.73 2.73 12.60

22 | 28 2.06 2.06 14.66

: | : : : :

84 | 43 3.17 3.17 94.25

85 | 19 1.40 1.40 95.65

86 | 4 0.29 0.29 95.95

87 | 9 0.66 0.66 96.61

88 | 13 0.96 0.96 97.57

89 | 9 0.66 0.66 98.23

90 | 1 0.07 0.07 98.31

91 | 14 1.03 1.03 99.34

92 | 7 0.52 0.52 99.85

93 | 2 0.15 0.15 100.00

Total | 1357 100.00 100.00

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

. fre bmi2018rd if bmi2018rd\_flag ==1, tabulate(10)

bmi2018rd -- BMI rounded to whole number

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

| Freq. Percent Valid Cum.

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

Valid 15 | 1 0.07 0.07 0.07

16 | 2 0.15 0.15 0.22

17 | 2 0.15 0.15 0.37

18 | 9 0.66 0.66 1.03

19 | 23 1.69 1.69 2.73

20 | 43 3.17 3.17 5.90

21 | 58 4.27 4.27 10.17

22 | 87 6.41 6.41 16.58

23 | 90 6.63 6.63 23.21

24 | 92 6.78 6.78 29.99

: | : : : :

49 | 3 0.22 0.22 99.12

50 | 1 0.07 0.07 99.19

51 | 2 0.15 0.15 99.34

52 | 3 0.22 0.22 99.56

53 | 1 0.07 0.07 99.63

54 | 1 0.07 0.07 99.71

55 | 1 0.07 0.07 99.78

56 | 1 0.07 0.07 99.85

57 | 1 0.07 0.07 99.93

60 | 1 0.07 0.07 100.00

Total | 1357 100.00 100.00

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

It turns out that the bmi2018\_rd interval-ratio variable is positively skewed which means we must bottom code it to prevent outliers to the right from influencing the mean and standard deviation.

1. **Bottom-code BMI at 19 or less (*bmi2018rd\_bc*). Be sure to add value labels. Submit frequencies of the new variable (first and last ~10 rows only) and simple descriptives of the old and new variable. Limit both analyses to the analytic sample only.**

**gen bmi2018rd\_bc = bmi2018rd**

**replace bmi2018rd\_bc = 19 if bmi2018rd <= 19**

**label variable bmi2018rd\_bc "2018 BMI back-coding at 19 or less"**

**label define bmi2018rd\_bc .i "IAP" .n "no answer" .d "don't know" 19 "19 or less"**

**label values bmi2018rd\_bc bmi2018rd\_bc**

**numlabel bmi2018rd\_bc, add**

**fre bmi2018rd\_bc if bmi2018rd\_flag ==1, tabulate(10)**

**sum bmi2018rd\_bc bmi2018rd if bmi2018rd\_flag ==1, detail**

. fre bmi2018rd\_bc if bmi2018rd\_flag ==1, tabulate(10)

bmi2018rd\_bc -- 2018 BMI back-coding at 19 or less

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

| Freq. Percent Valid Cum.

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

Valid 19 19. 19 or less | 37 2.73 2.73 2.73

20 | 43 3.17 3.17 5.90

21 | 58 4.27 4.27 10.17

22 | 87 6.41 6.41 16.58

23 | 90 6.63 6.63 23.21

24 | 92 6.78 6.78 29.99

25 | 89 6.56 6.56 36.55

26 | 92 6.78 6.78 43.33

27 | 120 8.84 8.84 52.17

28 | 81 5.97 5.97 58.14

: | : : : :

49 | 3 0.22 0.22 99.12

50 | 1 0.07 0.07 99.19

51 | 2 0.15 0.15 99.34

52 | 3 0.22 0.22 99.56

53 | 1 0.07 0.07 99.63

54 | 1 0.07 0.07 99.71

55 | 1 0.07 0.07 99.78

56 | 1 0.07 0.07 99.85

57 | 1 0.07 0.07 99.93

60 | 1 0.07 0.07 100.00

Total | 1357 100.00 100.00

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

. sum bmi2018rd\_bc bmi2018rd if bmi2018rd\_flag ==1, detail

2018 BMI back-coding at 19 or less

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

Percentiles Smallest

1% 19 19

5% 20 19

10% 21 19 Obs 1,357

25% 24 19 Sum of Wgt. 1,357

50% 27 Mean 28.38615

Largest Std. Dev. 6.318703

75% 31 55

90% 37 56 Variance 39.926

95% 40 57 Skewness 1.144032

99% 49 60 Kurtosis 4.876304

BMI rounded to whole number

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

Percentiles Smallest

1% 18 15

5% 20 16

10% 21 16 Obs 1,357

25% 24 17 Sum of Wgt. 1,357

50% 27 Mean 28.3692

Largest Std. Dev. 6.346789

75% 31 55

90% 37 56 Variance 40.28173

95% 40 57 Skewness 1.114828

99% 49 60 Kurtosis 4.851942

1. **Top-code socioeconomic index and 92 or more (*sei10rd\_tc*). Be sure to add value labels. Submit frequencies of the new variable (first and last ~10 rows only) and simple descriptives of the old and new variable. Limit both analyses to the analytic sample only.**

**recode sei10rd (92/93= 92), gen(sei10rd\_tc)**

**label variable sei10rd\_tc "socio-economic index, top coded at 92 and above"**

**label define sei10rd\_tc 92 "92 or more index number"**

**label values sei10rd\_tc sei10rd\_tc**

**numlabel sei10rd\_tc, add**

**fre sei10rd\_tc if bmi2018rd\_flag == 1, tabulate(10)**

**sum sei10rd\_tc sei10rd if bmi2018rd\_flag == 1**

. fre sei10rd\_tc if bmi2018rd\_flag == 1, tabulate(10)

sei10rd\_tc -- socio-economic index, top coded at 92 and above

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

| Freq. Percent Valid Cum.

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

Valid 11 | 2 0.15 0.15 0.15

13 | 24 1.77 1.77 1.92

14 | 47 3.46 3.46 5.38

15 | 10 0.74 0.74 6.12

17 | 8 0.59 0.59 6.71

18 | 3 0.22 0.22 6.93

19 | 12 0.88 0.88 7.81

20 | 28 2.06 2.06 9.87

21 | 37 2.73 2.73 12.60

22 | 28 2.06 2.06 14.66

: | : : : :

83 | 18 1.33 1.33 91.08

84 | 43 3.17 3.17 94.25

85 | 19 1.40 1.40 95.65

86 | 4 0.29 0.29 95.95

87 | 9 0.66 0.66 96.61

88 | 13 0.96 0.96 97.57

89 | 9 0.66 0.66 98.23

90 | 1 0.07 0.07 98.31

91 | 14 1.03 1.03 99.34

92 92. 92 or more index number | 9 0.66 0.66 100.00

Total | 1357 100.00 100.00

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

. sum sei10rd\_tc sei10rd if bmi2018rd\_flag == 1

Variable | Obs Mean Std. Dev. Min Max

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

sei10rd\_tc | 1,357 47.9101 23.0301 11 92

sei10rd | 1,357 47.91157 23.03296 11 93

1. **An edited table from esttab output (*bmi2018rd\_bc, age4cat, sex, race\_eth, region4cat, sei10rd\_tc* – use reference categories specified above). The table should only show unstandardized coefficients but the write-up should include an analysis of the standardized coefficients. Please put the table on its own page when submitting.**

Multivariate OLS Models Predicting Socioeconomic Index Based on Marital Status and Sociodemograp

> hic Characteristics, 2010 General Social Survey

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

Model 1 Model 2 Model 3

B B B

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

**Age (ref = 18-24)**

2. 25-55 1.991\*\* 1.904\*\* 2.152\*\*\*

3. 45-64 2.481\*\*\* 2.461\*\*\* 2.752\*\*\*

4. 65 or older 0.958 0.988 1.229

**Sex (ref =Female)**

1. male -0.172 -0.170

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

2. NH Black 1.111\* 0.925+

3. Hispanic 0.267 0.011

4. Other -2.325\*\* -2.242\*\*

**Region (ref =South)**

1. Northeast (New Engl and Mid Atl) -1.486\*\* -1.433\*\*

2. Midwest (W-No Central and E-No Centr -0.533 -0.592

4. West (Mountain and Pacific) -1.034\* -0.999\*

**Socio-economic index**, **top coded at 92** -0.020\*\*

Constant 26.455\*\*\* 27.061\*\*\* 27.837\*\*\*

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

N 1357.000 1357.000 1357.000

r2\_a 0.010 0.027 0.032

p 0.001 0.000 0.000

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

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

1. **Code from your esttab command (that produced the table in part h)**

**regress bmi2018rd\_bc i.age4cat if bmi2018rd\_flag == 1**

**estimates store mod1**

**regress bmi2018rd\_bc i.age4cat ib2.sex i.race\_eth ib3.region4cat if bmi2018rd\_flag == 1, beta**

**estimates store mod2**

**regress bmi2018rd\_bc i.age4cat ib2.sex i.race\_eth ib3.region4cat sei10rd\_tc if bmi2018rd\_flag == 1, beta**

**estimates store mod3**

**esttab mod1 mod2 mod3, varwidth(40) label b(%6.3f) star(+ 0.10 \* 0.05 \*\* 0.01 \*\*\* 0.001) stats(N r2\_a p) compress nogaps not ///**

**title("Multivariate OLS Models Predicting Socioeconomic Index Based on Marital Status and Sociodemographic Characteristics, 2010 General Social Survey") ///**

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

1. **A standard write-up of your findings from the entire model series**

A series of ordinary least squares regression models was conducted to determine whether age categories, sex, race, region and the socioeconomic index of respondents (from 11 to 92) from 2010 survey were significant linear predictors of their calculated body mass index from 2018 (starting from 19 – 60). Model 1 comprised of age categories; Model 2 added variables sex, race, and region; Model 3 added socioeconomic index. Model 1 found initially that age was strongly associated with the body mass index of participants: Compared to people from the age group 18-24, people who were 25-55 years old had 2 points better bmi ( *p =0.002), and people who were 45-64 years old had 2.48 points better bmi (p <0.001)*, at average*.* However, there was no significant difference between the bmi index of people falling in age group 65 and older, with respect to people who are 18-24 years old, at average (*p =0.26 ,* B= 0.96). Model 1 was significant (*p<0.001)* and accounted for 1.24 percent variation in bmi index.

In Model 2, age remained strongly associated with the bmi, controlling for sex, race, and region. Like Model 1, the bmi of participants in age group 25-55 and 45-65 remained significantly greater, on average, compared to the bmi of participants in 18-24 age group (*p = 0.003, p <0.001, respectively)* (B = 1.9, B =2.46, respectively). And similarly, people from the age group 65 and above had a bmi which was not significantly different than bmi of participants who were 18-24, on average (*p =0.244,* B =0.99). Moreover, sex, upon controlling for age, race, and region was not associated with bmi. Relative to females, males’ bmi was not significantly lesser, on average (*p =0.61,* B =-0.17). Furthermore, upon controlling for age, sex, and region, the race of individuals showed strong association with bmi: Compared to NH whites, NH blacks had 1.11 points better bmi (*p =0.02)* and the rest of the races, excluding NH Hispanic had on average 2.3 points worser bmi, on average (*p =0.004).* However, NH Hispanics’ bmi was not significantly different than NH whites (*p =0.59,* B =0.27). Model 2 also found, upon controlling all variables except for the region of participants, that the region had strong link with the bmi of participants. Relative to those coming from South region, people from Northeast had a 1.5 points worser bmi (*p =0.004*), and those coming from the West had a 1.03 points worser bmi (*p= 0.028*), on average. But there was no significant difference in bmi of participants from South and Midwest (*p =0.24,* B =-0.53). All in all, it turns out that age was the strongest linear predictor of bmi of participants (β = 0.19 for 45-64 years old vs β = 0.15 for 25-55 years old). Next to age, race (β =-0.79 for others vs NH whites) and region (β =-0.084 for people from Northeast vs South) were strongest predictors. Model itself accounted for 2.74 percent of variation in bmi and was significant (*p <0.001*).

Lastly, upon adding Socio-economic index variable of participants, it was observed that age became more strongly associated with bmi as now the statistical significance of the difference in bmi of those that are aged 25-55, with respect to people aged 18-24, has increased to (*p = 0.001*); and now, on average, these people have a 2.15 points greater bmi than 18-24 years old. Like Model 2, no significant difference in bmi of 65 and older age people with respect to 18-24 years old (*p =0.15*). And similarly, upon controlling for all other variables in Model 3, the sex also remains a non-strong predictor bmi (*p =0.62,* B= -0.17 for Males w.r.t Females).

As for race, when other variables were controlled except this, the significant difference between bmi of NH blacks with respect to NHWhites dropped to a marginal significance (*p =0.061).* So now, NH blacks have a bmi that is 0.93 points greater than NH whites on average. People from Other category in race variable retained their significant difference in bmi with NH whites (*p =0.006,* B =-2.24). And Hispanics as well retained their non significant difference in bmi with NH Whites (*p = 0.98,* B =0.01). In Model 3 region variable did not change its significance as a predictor of BMI. People from Northeast and West had a significantly different bmi than people from South, on average (*p =0.006 ,* B =-1.4 for people from Northeast) (*p =0.033,* B = -1.0, for people from west). And lastly, it was ascertained that socioeconomic index of participants was negatively associated with their bmi (*p =0.01*). So, for an increase in one index number on the socio-economic index, the bmi decreased by -0.02 points. And just like Model 2, in Model 3, age was the strongest linear predictor of bmi (β = 0.2 for 45-64 years old vs β =0.17 for 25-55 years old). Next to precedence were the region of respondents (which had a β of -0.08 for people from Northeast with respect to people from South) and the socioeconomic index (β =-0.07). All in all, Model 3 accounted 3.16 percent variation in bmi of participants and was significant (p<0.001).