**RUTGERS UNIVERSITY**

**Bloustein School of Planning and Public Policy**

**Applied Multivariate Methods**

**Fall 2020**

**Dawne Mouzon, Ph.D.**

**Problem Set #7**

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

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

**BIVARIATE ORDINARY LEAST SQUARES REGRESSION**

**(100 points total)**

**Please be sure to follow all instructions for the problems and submit the specified code and output. The requested output can be found after each question. Please copy and paste your code and output after each question/subquestion.**

**Your output must be copied/pasted. Do not use screenshots or pictures. Use the Courier New font, size 10 to make your output line up. Do not use bold font for output. Minimize the font if necessary.**

**Copy and paste your code from your do-file only (not the log); copy and paste your output from the results window.**

**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 2014 General Social Survey for the following questions.***

1. **(20 pts) Create an interval-ratio variable (bmi2014) that combines 2014 weight (weight) and 2014 height (height) using the formula we learned in the Advanced Recoding and Computing lab. Carry over the extended missing values to bmi2014 (.i/IAP, .d/don’t know, .n=no answer) only if each value is the same across weight and height. Otherwise, allow those cases to default to system-missing on the new variable. Please note that these variables are only available in the 2014 wave. Please submit:**
   1. Frequencies (copy and paste only the first and last 10 values of the variable to avoid too long of a document)
   2. Extended descriptive (i.e. sum, detail)
   3. Your code

**tab1 weight height, miss**

**gen bmi2014 = (weight/((height\*height))) \* 703**

**label variable bmi2014 "2014-BMI"**

**replace bmi2014= .d if weight== .d & height == .d**

**replace bmi2014= .n if weight== .n & height == .n**

**replace bmi2014= .i if weight== .i & height == .i**

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

**label values bmi2014 bmimiss**

**numlabel bmimiss, add**

**tab bmi2014, miss**

**sum bmi2014, detail**

tab bmi2014, miss

2014 BMI | Freq. Percent Cum.

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

16.03749 | 1 0.04 0.04

16.94559 | 1 0.04 0.08

17.21633 | 1 0.04 0.12

17.35802 | 1 0.04 0.16

17.6374 | 1 0.04 0.20

17.67798 | 1 0.04 0.24

17.80916 | 1 0.04 0.28

17.93367 | 1 0.04 0.32

17.96556 | 1 0.04 0.35

18.30296 | 1 0.04 0.39

[DATA OMITTED TO FIT ON PAGE]

50.48837 | 1 0.04 45.94

50.84131 | 1 0.04 45.98

51.48232 | 1 0.04 46.02

54.08707 | 1 0.04 46.06

58.42105 | 1 0.04 46.10

67.25826 | 1 0.04 46.14

. | 66 2.60 48.74

.d. don’t know | 1 0.04 48.78

.i. IAP | 1,288 50.75 99.53

.n. no answer | 12 0.47 100.00

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

Total | 2,538 100.00

. sum bmi2014, detail

2014-BMI

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

Percentiles Smallest

1% 18.45726 16.03749

5% 20.50417 16.94559

10% 21.78719 17.21633 Obs 1,171

25% 23.91156 17.35802 Sum of Wgt. 1,171

50% 27.09977 Mean 27.99723

Largest Std. Dev. 5.855369

75% 30.68042 51.48232

90% 35.42454 54.08707 Variance 34.28535

95% 39.15126 58.42105 Skewness 1.347993

99% 48.41598 67.25826 Kurtosis 6.518682

1. **Run separate bivariate ordinary least squares regression models to predict BMI (bmi2014) based on the variables below. For each test, please submit your code for running the test and output from the test, as well as a standard write-up.**
   1. (10 pts) Age in years (**age**)

**tab1 bmi2014 age, miss**

**regress bmi2014 age**

regress bmi2014 age

Source | SS df MS Number of obs = 1,169

-------------+---------------------------------- F(1, 1167) = 3.71

Model | 127.259991 1 127.259991 Prob > F = 0.0542

Residual | 39986.1715 1,167 34.2640715 R-squared = 0.0032

-------------+---------------------------------- Adj R-squared = 0.0023

Total | 40113.4315 1,168 34.3436913 Root MSE = 5.8536

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

bmi2014 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

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

age | .0245094 .0127176 1.93 0.054 -.0004426 .0494614

\_cons | 26.91787 .5857833 45.95 0.000 25.76857 28.06718

A bivariate ordinary least squares regression was conducted to determine whether age (in years) was a significant linear predictor of body mass index in 2014. The overall model was marginally significant (*p*= 0.0542). Age was marginally positively associated to the body mass index; for each additional year of age the BMI goes up by 0.025 (*p*= 0.054). The model accounted for 0.23% of variation in the BMI index.

* 1. (10 pts) Number of days of poor physical health in the past month (**physhlth**)

**tab1 bmi2014 physhlth, miss**

**regress bmi2014 physhlth**

regress bmi2014 physhlth

Source | SS df MS Number of obs = 1,170

-------------+---------------------------------- F(1, 1168) = 6.19

Model | 211.328872 1 211.328872 Prob > F = 0.0130

Residual | 39896.0716 1,168 34.1575955 R-squared = 0.0053

-------------+---------------------------------- Adj R-squared = 0.0044

Total | 40107.4005 1,169 34.3091535 Root MSE = 5.8444

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

bmi2014 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

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

physhlth | .0682836 .0274524 2.49 0.013 .014422 .1221451

\_cons | 27.82867 .1834923 151.66 0.000 27.46866 28.18868

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

A bivariate ordinary least squares regression was conducted to determine whether the number of days of poor physical health is a significant linear predictor of body mass index in 2014. The overall model was significant (p= 0.0130). The number of days of poor physical health is positively associated to the body mass index; for each additional day of poor health, the BMI goes up by 0.068 (p= 0.013). The model accounted for 0.53% of variation in the BMI index.

* 1. (10 pts) Number of hours per week on the internet (**wwwhr**)

**tab1 bmi2014 wwwhr, miss**

**regress bmi2014 wwwhr**

. regress bmi2014 wwwhr

Source | SS df MS Number of obs = 728

-------------+---------------------------------- F(1, 726) = 2.63

Model | 86.4259539 1 86.4259539 Prob > F = 0.1055

Residual | 23884.6388 726 32.8989516 R-squared = 0.0036

-------------+---------------------------------- Adj R-squared = 0.0022

Total | 23971.0648 727 32.9725788 Root MSE = 5.7358

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

bmi2014 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

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

wwwhr | .0215311 .0132842 1.62 0.105 -.0045489 .0476112

\_cons | 27.62992 .2691431 102.66 0.000 27.10153 28.15831

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

A bivariate ordinary least squares regression was conducted to determine whether the number of hours per week spent on the internet is a significant linear predictor of body mass index in 2014. The overall model was not significant (p= 0.1055). The number of hours per week on the internet is not a significant predictor of the BMI index (B=0.022, p= 0.105). The model accounted for approximately 0 percent of the variation in the BMI index.

* 1. (10 pts) Sex (**sex** – use women as the reference category/base level)

**tab1 bmi2014 sex, miss**

**regress bmi2014 ib2.sex**

regress bmi2014 ib2.sex

Source | SS df MS Number of obs = 1,171

-------------+---------------------------------- F(1, 1169) = 0.12

Model | 4.04728871 1 4.04728871 Prob > F = 0.7313

Residual | 40109.8141 1,169 34.3112182 R-squared = 0.0001

-------------+---------------------------------- Adj R-squared = -0.0008

Total | 40113.8614 1,170 34.2853516 Root MSE = 5.8576

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

bmi2014 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

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

sex |

1. male | -.1175955 .3423945 -0.34 0.731 -.789372 .5541809

2. female | 0 (base)

|

\_cons | 28.05698 .2440657 114.96 0.000 27.57813 28.53584

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

A bivariate ordinary least squares regression model was made to determine whether sex is a significant predictor of body mass index in 2014. The overall model was not significant (p= 0.7313). Sex was not a predictor of the body mass index (B= -0.118 for males v/s females, p= 0.731). The model accounted for less than 0% of the variation in the body mass index.

* 1. (10 pts) Nativity status (**born–** use those born in the U.S. as the reference category)

**tab1 bmi2014 born, miss**

**regress bmi2014 i.born**

regress bmi2014 i.born

Source | SS df MS Number of obs = 1,171

-------------+---------------------------------- F(1, 1169) = 4.64

Model | 158.486236 1 158.486236 Prob > F = 0.0315

Residual | 39955.3751 1,169 34.1791062 R-squared = 0.0040

-------------+---------------------------------- Adj R-squared = 0.0031

Total | 40113.8614 1,170 34.2853516 Root MSE = 5.8463

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

bmi2014 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

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

born |

1. yes | 0 (base)

2. no | -1.079706 .5014065 -2.15 0.031 -2.063463 -.0959485

|

\_cons | 28.14199 .1835952 153.28 0.000 27.78178 28.5022

A bivariate ordinary least squares regression was conducted to determine whether nativity status is a significant predictor of body mass index in 2014. The overall model was significant (p= 0.0315). Nativity status is a significant predictor of the body mass index. Those not born in the US scored 1.08 points lower than those born in the US (p=0.031) on the BMI index. The model accounted for 0.31% of the variation in the body mass index.

* 1. (15 pts) Self-rated health (**health** – use “excellent” as the reference category)

**tab1 bmi2014 health, miss**

**regress bmi2014 i.health**

. regress bmi2014 i.health

Source | SS df MS Number of obs = 770

-------------+---------------------------------- F(3, 766) = 16.17

Model | 1578.08834 3 526.029446 Prob > F = 0.0000

Residual | 24921.9835 766 32.5352265 R-squared = 0.0596

-------------+---------------------------------- Adj R-squared = 0.0559

Total | 26500.0718 769 34.4604315 Root MSE = 5.704

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

bmi2014 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

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

health |

1. excellent | 0 (base)

2. good | 1.99914 .4658103 4.29 0.000 1.084724 2.913556

3. fair | 3.586124 .6257453 5.73 0.000 2.357745 4.814503

4. poor | 7.653664 1.758149 4.35 0.000 4.202301 11.10503

|

\_cons | 26.12219 .365159 71.54 0.000 25.40536 26.83902

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

A bivariate ordinary least squares regression model was conducted to determine whether self rated health (excellent, good, fair, poor) was a significant predictor of the body mass index in 2014. The overall model was significant (p < 0.001). Self-reported health was a significant predictor of the body mass index. On average, the worser the self rated health gets, the higher is the value of BMI index. For example, compared to those with excellent health, those with good health scored 2 points higher (p < 0.001), those with fair health scored 3.59 points higher (p < 0.001), and those with poor health scored 7.65 points higher (p <0.001) on the BMI index. The model accounted for 5.6% of the variation in the body mass index in 2014.

* 1. (15 pts) Race/ethnicity (**race\_eth**) - use “Hispanic” as the reference category

**tab1 bmi2014 race\_eth, miss**

**regress bmi2014 ib3.race\_eth**

. regress bmi2014 ib3.race\_eth

Source | SS df MS Number of obs = 1,168

-------------+---------------------------------- F(3, 1164) = 3.10

Model | 317.288131 3 105.76271 Prob > F = 0.0261

Residual | 39752.666 1,164 34.151775 R-squared = 0.0079

-------------+---------------------------------- Adj R-squared = 0.0054

Total | 40069.9542 1,167 34.3358648 Root MSE = 5.844

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

bmi2014 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

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

race\_eth |

1. NH White | .3822827 .4721535 0.81 0.418 -.5440844 1.30865

2. NH Black | 1.20345 .6164052 1.95 0.051 -.0059396 2.412839

3. Hispanic | 0 (base)

4. Other | -1.512478 .9353335 -1.62 0.106 -3.347606 .3226507

|

\_cons | 27.63867 .4217509 65.53 0.000 26.8112 28.46615

A bivariate ordinary least squares regression model was conducted to determine whether race/ethnicity (NH White, NH Black, Hispanic, Other) was a significant predictor of the body mass index in 2014. The overall model was significant (p =0.0261). Compared to Hispanics, NH Blacks scored on average, 1.2 points marginally higher on the BMI index (p =0.051). Those who were NH White (B= 0.382, p = 0.42) and Others (B = -1.51, p =0.11) did not have significantly different mean scores on the BMI index than Hispanics. The model accounted for 0.54% of the variation in the body mass index in 2014.