**RUTGERS UNIVERSITY**

**Bloustein School of Planning and Public Policy**

**Applied Multivariate Methods**

**Fall 2020**

**Dawne Mouzon, Ph.D.**

**Problem Set #8:**

**Multivariate Ordinary Least Squares Regression I**

**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.**

**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.**

**MULTIVARIATE ORDINARY LEAST SQUARES REGRESSION:**

**(100 points total, 50 points each)**

**#1. (2010 GSS) We are interested in running a series of nested multivariate ordinary least squares regression models predicting respondent income (*conrinc*) based on some of the following categories of variables below. There must be at least 5-6 IVs in the model series. Be sure to use an analytic flag. Your final analytic sample must include at least 400 cases.**

* **Model 1 – Education**
  + **Potential variables: educational attainment (*educ*), maternal education in years (*maeduc*), and/or paternal education in years (*paeduc*)**
* **Model 2 -- Demographics** 
  + **Potential variables: race/ethnicity (*race\_eth;* ref=NH white), *sex* (ref=male), age (*age*)**
* **Model 3 – Labor Participation** 
  + **Potential variables: hours worked last week (*hrs1*), weeks worked last year (*weekswrk*), work part-time or full-time (*partfull*), usual work schedule (*wrksched*), supervise others at work (*suprvsjb*), and/or work arrangement at main job (*wrktype*)**

**For this problem, please submit:**

1. **A completed multivariate table in Word, based on the output from your models (template below, please do not make major changes to formatting). Please put the table on its own page and make sure it fits on one page**
2. **A standard write-up**
3. **Your code for the entire problem**

**tab1 conrinc race\_eth sex suprvsjb, miss**

**fre educ age weekswrk, tabulate(4)**

**mark conrinc\_flag**

**markout conrinc\_flag conrinc educ race\_eth sex weekswrk suprvsjb**

**label variable conrinc\_flag "analytic flag for income in constant dollars - conrinc(N =963)"**

**tab conrinc\_flag, miss**

**regress conrinc educ if conrinc\_flag == 1**

**regress conrinc educ i.race\_eth age i.sex if conrinc\_flag == 1, beta**

**regress conrinc educ i.race\_eth age i.sex weekswrk i.suprvsjb if conrinc\_flag == 1, beta**

**Table I. Nested Multivariate Ordinary Least Squares Regression Models Predicting Income Based on Education, Demographics and Labour Participation, 2010 General Social Survey (N = 963)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Model 1** | **Model 2** | | **Model 3** | |
|  | **B** | **B** | **β** | **B** | **β** |
| **Educational Attainment (in years)** | 4008.27\*\*\* | 4079.64\*\*\* | 0.42\*\*\* | 3662.5\*\*\* | 0.38\*\*\* |
| **Race/Ethnicity (ref= NH White)** |  |  | |  | |
| NH Black | --- | -6108.4\* | -0.068\* | -4428.6+ | -0.05+ |
| Hispanic | --- | -1542.8 | -0.017 | 745.04 | 0.008 |
| NH Other | --- | 2672.1 | 0.018 | 3214.8 | 0.021 |
| **Age** | --- | 284.5\*\*\* | 0.13\*\*\* | 293.6\*\*\* | 0.14\*\*\* |
| **Sex (ref= male)** |  |  | |  | |
| Female | --- | -13803\*\*\* | -0.23\*\*\* | -12295.2\*\*\* | -0.21\*\*\* |
| **Weeks worked last year** | --- | --- | --- | 727.4\*\*\* | 0.25\*\*\* |
| **Supervise others at work (ref = Yes)** |  |  | |  | |
| No | --- | --- | --- | -9406.6\*\*\* | -0.16\*\*\* |
| **Model Significance (*p*)** | <0.001 | <0.001 | | <0.001 | |
| **Adjusted R2** | 0.1678 | 0.2475 | | 0.3325 | |
| **N** | 963 | 963 | | 963 | |

\*\*\* *p* < 0.001; \*\* *p* < 0.01; \* *p <* 0.05; + *p* < 0.10.

**Standard Write-up:**

A series of ordinary least squares regression models was conducted to determine whether education, demographics (race/ethnicity, sex, age) and labor participation (no. of weeks worked last year, Supervise others at work) were significant linear predictors of income in constant dollars of respondents. Model 1 included education attainment in years, Model 2, in addition, comprised of age in years, sex, and race/ethnicity, and Model 3 added number of weeks worked in last year and whether the respondent supervised someone or not. In Model 1, education attained in years was positively associated with income such that each one-year increase in education was associated with a $4008.27 increase in income (p < 0.001). Model 1 was significant (p < 0.001) and accounted for 16.78% of the variation in income.

In Model 2, education remained positively associated with income (B = 4079.64, p<0.001), when variables Race, Age, and Sex were controlled. Moreover, Age, and Sex also turned out to be significantly associated with the variation in income. Upon an addition in age of one year, the income increases by $284.5 (p<0.001). On the contrary sex turned out to be negatively associated with income: Females made on average $13803 lesser than males(p<0.001). Those who identified as NH Blacks earned $6108.4 less on average than those who identified as NH Whites (p =0.017), while there was no significant difference in the earnings of those who identified as NH Other and those who identified as Hispanics compared to those who identified as NH Whites (B= 2672.1 for NH Other, -1542.8 for Hispanic). In Model 2, Education was the strongest predictors of income (β= 0.42 vs β= -0.23 for sex vs β= -0.068 for NH Black income with respect to NH Whites). Model 2 was significant (p < 0.001) and explained 24.75% of the variation in income.

In Model 3, after controlling for additional variables related to labor participation (Weeks worked and Supervision of others), the educational attainment remained positively associated with income (B= 3662.5, p < 0.001). Age and sex were also still significant predictors of income: Age remained positively associated, with one extra year of age adding $293.6 in income (p < 0.001), controlling for sex, race/ethnicity, education and labor participation. Females earned $12295.2 less than males (p < 0.001), on average, when all other variables were controlled. However, those who identified as NH Blacks dropped to marginal significance (p = 0.068) and now on average, NH Blacks earn, on average, $4428.6 less than NH Whites. Other Ethnicities and Hispanics still have no significant difference in their earnings with respect to NH Whites.

Model 3 also found that weeks worked last year was positively associated with income (p<0.001), with an increase in one week resulting in $727.4 increase in income, while controlling for education, age, sex, race, and supervision of others variables. As for the supervision of others, it was found that there was a significant difference (p<0.001)in the income of those who supervise others vs those who do not: Relative to those who supervise others, those who do not ,earn on average, $9406.6 lesser, while controlling for education, age, sex, race, and weeks worked last year. In Model 3, educational attainment was the strongest predictor of income (β= 0.38 vs β= 0.25 for weeks worked last year vs β= -0.21 for sex vs β= -0.16 for supervising others or not vs β= 0.14 for age vs β= -0.05 for NH Black’s average income with respect to NH Whites). Model 3 was significant (p<0.001) and explained 33.25% of the variation in income.

. regress conrinc educ if conrinc\_flag == 1

Source | SS df MS Number of obs = 963

-------------+---------------------------------- F(1, 961) = 194.92

Model | 1.4026e+11 1 1.4026e+11 Prob > F = 0.0000

Residual | 6.9150e+11 961 719559774 R-squared = 0.1686

-------------+---------------------------------- Adj R-squared = 0.1678

Total | 8.3175e+11 962 864608399 Root MSE = 26825

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

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

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

educ | 4008.27 287.0973 13.96 0.000 3444.86 4571.679

\_cons | -22121 4142.746 -5.34 0.000 -30250.88 -13991.13

. regress conrinc educ i.race\_eth age i.sex if conrinc\_flag == 1, beta

Source | SS df MS Number of obs = 963

-------------+---------------------------------- F(6, 956) = 53.73

Model | 2.0975e+11 6 3.4958e+10 Prob > F = 0.0000

Residual | 6.2200e+11 956 650632564 R-squared = 0.2522

-------------+---------------------------------- Adj R-squared = 0.2475

Total | 8.3175e+11 962 864608399 Root MSE = 25508

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

conrinc | Coef. Std. Err. t P>|t| Beta

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

educ | 4079.674 285.6581 14.28 0.000 .4179579

|

race\_eth |

1. NH White | 0 (base) 0

2. NH Black | -6108.402 2563.65 -2.38 0.017 -.0681532

3. Hispanic | -1542.779 2722.647 -0.57 0.571 -.0170232

4. Other | 2672.136 4208.403 0.63 0.526 .0179232

|

age | 284.5434 61.59393 4.62 0.000 .1321525

|

sex |

1. male | 0 (base) 0

2. female | -13803.17 1664.371 -8.29 0.000 -.2343006

|

\_cons | -27387.81 5109.35 -5.36 0.000 .

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

.

regress conrinc educ i.race\_eth age i.sex weekswrk i.suprvsjb if conrinc\_flag == 1, beta

Source | SS df MS Number of obs = 963

-------------+---------------------------------- F(8, 954) = 60.89

Model | 2.8115e+11 8 3.5144e+10 Prob > F = 0.0000

Residual | 5.5060e+11 954 577153515 R-squared = 0.3380

-------------+---------------------------------- Adj R-squared = 0.3325

Total | 8.3175e+11 962 864608399 Root MSE = 24024

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

conrinc | Coef. Std. Err. t P>|t| Beta

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

educ | 3662.527 272.5089 13.44 0.000 .3752217

|

race\_eth |

1. NH White | 0 (base) 0

2. NH Black | -4428.613 2421.652 -1.83 0.068 -.0494113

3. Hispanic | 745.0434 2572.792 0.29 0.772 .0082209

4. Other | 3214.809 3970.013 0.81 0.418 .0215632

|

age | 293.6001 58.01985 5.06 0.000 .1363588

|

sex |

1. male | 0 (base) 0

2. female | -12295.22 1580.152 -7.78 0.000 -.208704

|

weekswrk | 727.3877 78.34246 9.28 0.000 .2465879

|

suprvsjb |

1. yes | 0 (base) 0

2. no | -9406.633 1633.206 -5.76 0.000 -.1558436

|

\_cons | -51666.55 6118.261 -8.44 0.000 .

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

**#2. (2016 GSS) We are interested in predicting institutional confidence (*inst\_conf\_scale*) based on 5-6 control variables of your choice. Please use at least three models. There must be at least 5-6 IVs in the model series. Be sure to use an analytic flag. Your final analytic sample must include at least 400 cases.**

**For this problem, please submit:**

1. **Output from your univariate (Garbage In, Garbage Out) analysis**
2. **A completed multivariate table in Word, based on the output from your models (template below, please do not make major changes to formatting). Please put the table on its own page and make sure it fits on one page.**
3. **A standard write-up**
4. **Your code for the entire problem**

**Table II. Nested Multivariate Ordinary Least Squares Regression Models Predicting Confidence in Institutions1 Based on Based on Race, Depression Symptom’s scale (0 – 12), and Socioeconomic status, 2016 General Social Survey (N = 554)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Model 1** | **Model 2** | | **Model 3** | |
|  | **B** | **B** | **β** | **B** | **β** |
| **Race/Ethnicity (ref= NH White)** |  |  |  |  |  |
| NH Black | 1.00\* | 1.1\* | 0.097\* | 1.2\* | 0.105\* |
| Hispanic | 0.92+ | 0.91+ | 0.07+ | 1.11+ | 0.089+ |
| NH Other | 1.76\* | 1.87\* | 0.096\* | 1.82\* | 0.09\* |
| **Depression Symptom's Scale [ref = low (<25th percentile)]** |  |  | |  | |
| medium (25th-74th percentile) | --- | -1.05\* | -0.12\* | -1.1\* | -0.12\* |
| high (75th percentile or higher) | --- | -1.68\*\* | -0.16\*\* | -1.67\*\* | -0.16\*\* |
| **Age** | --- | --- | --- | 0.006 | 0.021 |
| **Education** | --- | --- | --- | 0.06 | 0.044 |
| **Income (Dollars)** | --- | --- | --- | 0.000001 | 0.01 |
| **Model Significance (*p*)** | 0.0288 | 0.0049 | | 0.0204 | |
| **Adjusted R2** | 0.0109 | 0.0212 | | 0.0183 | |
| **N** | 554 | 554 | | 554 | |

**Standard Write-up:**

A series of ordinary least squares regression models was conducted to determine whether Race, 4 item depression symptom’s scale (0/lowest – 12/highest) in categorical form, Age, Education, and Income were significant linear predictors of confidence in institutions (13-item scale where 0/lowest confidence in institutions -26/highest confidence in institutions). Model 1 included race, Model 2 added Depression Symptom’s Categorical Scale, and Model 3 included variables Age, Education, and Income additionally.

In Model 1, race had a positive association with confidence in institution’s scale. Compared to NH Whites, those who were NH Blacks scored 1 point higher, on average (p <0.05), on the confidence scale; NH Others and Hispanics also scored higher than NH Whites (NH others scored 1.76 points higher – p<0.05 – and Hispanics scored marginally 0.92 points higher – p = 0.09 – on the confidence scale). Model 1 was statistically significant (p<0.029) and it explained 1.09 percent variation in the confidence in institution scale.

In Model 2 as well, NH blacks, Hispanics, and Others retained their higher average scores with respect to NH Whites on the confidence scales, while controlling for the depression symptom’s variable. There was no appreciative change in their respective significance as well. The Depression symptom’s categories turned out to be strongly associated with the confidence in institutions scale. Relative to those that lie in the lower category of Depression symptoms (<25 percentile), those in the medium category (25 – 74 percentile) had 1.05 points lower confidence (p <0.05), and those in the high depression category (>74 percentile) had 1.68 points lower confidence in institutions (p<0.01). In Model 2, Depression Symptoms Categorical scale was the strongest linear predictor of confidence in institution scale (β = -0.16 for high percentile, β = -0.12 for medium percentile). The Model had a strong significance overall (p =0.0049) and explained 2.12 percent of variation in confidence in institution scale.

In Model 3, race remained strongly associated with confidence in institution scale, controlling Depression Symptoms categorical variable, Age, Education, and Income. As before, in Model 3, those that identified as NH Blacks (p <0.05)or Others (p<0.05) had higher confidence in institutions than NH Whites whereas Hispanics remain marginally more confident (p=0.051) than NH Whites on institutions. Depression Symptoms also remains strongly associated with confidence in institution scale when the rest of the variables are controlled. There is no noteworthy change in the coefficients and their respective statistical significance for Depression Symptoms despite controlling for Age, Income, and Education. Like Model 1 and Model 2, neither of these (Age, Income, and Education) were significant predictors of confidence in institution scale. In Model 3, just like Model 2, Depression Symptoms was the strongest predictor of confidence in institution scale (β = -0.16 for high percentile, β = -0.12 for medium percentile). Race also happens to be a strong predictor after Depression symptoms as NH Blacks have a β = 0.105 which is comparable to β of medium percentile in Depression Symptom’s categories. The adjusted R square value decreased to 0.0183 for Model 3 denoting that it can explain 1.83 percent of variation in confidence in institution scale. However, the overall model has an appreciable statistical significance of p =0.0204 i.e. p<0.05.

**tab1 inst\_conf\_scale race\_eth cesd\_scale\_cat, miss**

**fre educ conrinc age , tabulate(4)**

**mark inst\_conf\_scale\_flag**

**markout inst\_conf\_scale\_flag race\_eth cesd\_scale\_cat educ conrinc age**

**label variable inst\_conf\_scale\_flag "analytic flag for confidence in 7 item institution scale (0-26) - (N =554)"**

**tab inst\_conf\_scale\_flag, miss**

**regress inst\_conf\_scale i.race\_eth if inst\_conf\_scale\_flag ==1**

**regress inst\_conf\_scale i.race\_eth i.cesd\_scale\_cat if inst\_conf\_scale\_flag == 1, beta**

**regress inst\_conf\_scale i.race\_eth i.cesd\_scale\_cat age educ conrinc if inst\_conf\_scale\_flag == 1, beta**

. regress inst\_conf\_scale i.race\_eth if inst\_conf\_scale\_flag ==1

Source | SS df MS Number of obs = 554

-------------+---------------------------------- F(3, 550) = 3.03

Model | 163.346306 3 54.4487688 Prob > F = 0.0288

Residual | 9871.63023 550 17.9484186 R-squared = 0.0163

-------------+---------------------------------- Adj R-squared = 0.0109

Total | 10034.9765 553 18.1464313 Root MSE = 4.2366

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

inst\_conf\_~e | Coef. Std. Err. t P>|t| [95% Conf. Interval]

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

race\_eth |

1. NH White | 0 (base)

2. NH Black | 1.003495 .4927986 2.04 0.042 .0354969 1.971492

3. Hispanic | .9245434 .5438065 1.70 0.090 -.1436483 1.992735

4. Other | 1.763095 .8311869 2.12 0.034 .130406 3.395784

|

\_cons | 12.30833 .2232862 55.12 0.000 11.86974 12.74693

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

. regress inst\_conf\_scale i.race\_eth i.cesd\_scale\_cat if inst\_conf\_scale\_flag == 1, beta

Source | SS df MS Number of obs = 554

-------------+---------------------------------- F(5, 548) = 3.40

Model | 301.858845 5 60.371769 Prob > F = 0.0049

Residual | 9733.11769 548 17.7611637 R-squared = 0.0301

-------------+---------------------------------- Adj R-squared = 0.0212

Total | 10034.9765 553 18.1464313 Root MSE = 4.2144

inst\_conf\_scale | Coef. Std. Err. t P>|t| Beta

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

race\_eth |

1. NH White | 0 (base) 0

2. NH Black | 1.103591 .4922224 2.24 0.025 .0969142

3. Hispanic | .9101731 .5412328 1.68 0.093 .0723344

4. Other | 1.865137 .827807 2.25 0.025 .0959996

|

cesd\_scale\_cat |

1. low (<25th percentile) | 0 (base) 0

2. medium (25th-74th percentile) | -1.047492 .5177547 -2.02 0.044 -.118949

3. high (75th percentile or higher) | -1.680291 .6021394 -2.79 0.005 -.164557

|

\_cons | 13.32235 .4821022 27.63 0.000 .

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

. regress inst\_conf\_scale i.race\_eth i.cesd\_scale\_cat age educ conrinc if inst\_conf\_scale\_flag == 1, beta

Source | SS df MS Number of obs = 554

-------------+---------------------------------- F(8, 545) = 2.29

Model | 326.281941 8 40.7852426 Prob > F = 0.0204

Residual | 9708.69459 545 17.8141185 R-squared = 0.0325

-------------+---------------------------------- Adj R-squared = 0.0183

Total | 10034.9765 553 18.1464313 Root MSE = 4.2207

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

inst\_conf\_scale | Coef. Std. Err. t P>|t| Beta

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

race\_eth |

1. NH White | 0 (base) 0

2. NH Black | 1.201049 .5033861 2.39 0.017 .1054726

3. Hispanic | 1.115964 .5704422 1.96 0.051 .0886892

4. Other | 1.819492 .8346138 2.18 0.030 .0936503

|

cesd\_scale\_cat |

1. low (<25th percentile) | 0 (base) 0

2. medium (25th-74th percentile) | -1.094601 .5202177 -2.10 0.036 -.1242986

3. high (75th percentile or higher) | -1.673744 .6038325 -2.77 0.006 -.1639159

|

age | .0062721 .0129467 0.48 0.628 .0206424

educ | .0599786 .0638734 0.94 0.348 .0441148

conrinc | 1.21e-06 5.68e-06 0.21 0.832 .0096203

\_cons | 12.13736 1.167191 10.40 0.000 .

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

**Multivariate Regression Table Template**

**Please use the following table as a guide for what your table should look like. Make sure to update all relevant information in the table (like the title) to be relevant to each model series. The order of your final submission should be table, followed by write-up for each series.**

**Table X. Nested Multivariate Ordinary Least Squares Regression Models Predicting Support of Abortion Rightsa Based on Marital Status and Sociodemographic Characteristics, 2016 General Social Survey (N = 1,844)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Model 1** | | **Model 2** | | **Model 3** | |
|  | **B** | **β** | **B** | **β** | **B** | **β** |
| **Marital Status**  **(ref=Married)** |  |  |  |  |  |  |
| Separated/divorced/ widowed | 0.17 | 0.03 | 0.22 | 0.04 | 0.15 | 0.03 |
| Never married | 0.67\*\*\* | 0.12\*\*\* | 0.62\*\*\* | 0.11\*\*\* | 0.46\*\* | 0.08\*\* |
|  |  |  |  |  |  |  |
| **Age in Years** | --- | --- | -0.01 | -0.02 | -0.01 | -0.02 |
|  |  |  |  |  |  |  |
| **Male (ref=female)** | --- | --- | 0.22+ | 0.04+ | 0.31\*\* | 0.06\*\* |
|  |  |  |  |  |  |  |
| **Political Party Affiliation**  **(ref=Republican)** |  |  |  |  |  |  |
| Democrat | --- | --- | --- | --- | 1.34\*\*\* | 0.25\*\*\* |
| Independent | --- | --- | --- | --- | 0.85\*\*\* | 0.17\*\*\* |
| Other |  |  |  |  | 1.49\*\*\* | 0.09\*\*\* |
|  |  | |  | |  | |
| **Model Significance (*p*)** | < 0.001 | | < 0.001 | | <0.001 | |
| **Adjusted R2** | 0.010 | | 0.012 | | 0.050 | |
| **N** | 1,844 | | 1,844 | | 1,844 | |

\*\*\* *p* < 0.001; \*\* *p* < 0.01; \* *p <* 0.05; + *p* < 0.10.

a The summed 7-item “Support of Abortion Rights” scale ranges from 0/lowest support– 7/highest support.