**RUTGERS UNIVERSITY**

**Bloustein School of Planning and Public Policy**

**Applied Multivariate Methods**

**Fall 2020**

**Dawne Mouzon, Ph.D.**

**Problem Set #11**

**BIVARIATE BINARY LOGISTIC REGRESSION**

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

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

**BIVARIATE BINARY LOGISTIC REGRESSION**

**(100 points total)**

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

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

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

**Use allmidus.dta for all questions.**

1. **(10 points) Using *C1SC1*, please create a dummy variable for whether or not respondents are insured (*w3insured*). (Reminder: a dummy variable must be coded as 0/no, 1/yes.) Any missing or quirky values should be coded into separate extended missing values on the new variable. Be sure to follow all steps for creating a basic recode and show the relevant consistency check.**

**tab C1SC1, miss**

**recode C1SC1 (2 =0) (-1 = .i) (8 = .r), gen(w3insured)**

**label variable w3insured "Dummy variable for respondent's insurance coverage"**

**label define w3insured 0 "No" 1 "Yes" .r "Refused" .i "No SAQ Data"**

**label values w3insured w3insured**

**numlabel w3insured, add**

**tab1 C1SC1 w3insured, miss**

**tab C1SC1 w3insured, miss**

. tab1 C1SC1 w3insured, miss

-> tabulation of C1SC1

Covered by healthcare insurance |

currently | Freq. Percent Cum.

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

-1. RESPONDENT DOES NOT HAVE SAQ DATA | 563 7.65 7.65

1. YES | 2,590 35.18 42.83

2. NO | 113 1.53 44.36

8. REFUSED | 28 0.38 44.74

. | 4,068 55.26 100.00

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

Total | 7,362 100.00

-> tabulation of w3insured

Dummy variable |

for |

respondent's |

insurance |

coverage | Freq. Percent Cum.

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

0. No | 113 1.53 1.53

1. Yes | 2,590 35.18 36.72

. | 4,068 55.26 91.97

.i. No SAQ Data | 563 7.65 99.62

.r. Refused | 28 0.38 100.00

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

Total | 7,362 100.00

. tab C1SC1 w3insured, miss

Covered by healthcare | Dummy variable for respondent's insurance coverage

insurance currently | 0. No 1. Yes . .i. No SA .r. Refus | Total

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

-1. RESPONDENT DOES N | 0 0 0 563 0 | 563

1. YES | 0 2,590 0 0 0 | 2,590

2. NO | 113 0 0 0 0 | 113

8. REFUSED | 0 0 0 0 28 | 28

. | 0 0 4,068 0 0 | 4,068

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

Total | 113 2,590 4,068 563 28 | 7,362

1. **In the models below, please predict the odds of being covered by health insurance (*w3insured*) based on the predictors below. Please include a standard write-up of your findings for each test.**
2. (10 pts) Age (***w3age***)

**. logistic w3insured w3age**

Logistic regression Number of obs = 2,703

LR chi2(1) = 38.88

Prob > chi2 = 0.0000

Log likelihood = -449.90857 Pseudo R2 = 0.0414

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

w3insured | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

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

w3age | 1.058905 .0102181 5.93 0.000 1.039066 1.079122

\_cons | .6822835 .390966 -0.67 0.505 .2219241 2.097613

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

Note: \_cons estimates baseline odds.

A bivariate logistic regression was conducted to determine whether the age of respondents (ranging from 39 (low) to 93 (high)) was a significant predictor about whether respondents have health insurance or not. It seems that with an increase in 1 year in age of respondents, the odds of being health insured enhance by 5.9 percent (p <0.001). Overall, this model was significant (p <0.001) and accounted for 4.14 percent of variation in whether respondents have health insurance or not.

1. (10 pts) Autonomy scale (***auton\_scale***)

**. logistic w3insured auton\_scale**

Logistic regression Number of obs = 2,702

LR chi2(1) = 2.00

Prob > chi2 = 0.1570

Log likelihood = -468.30486 Pseudo R2 = 0.0021

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

w3insured | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

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

auton\_scale | 1.020005 .0141921 1.42 0.155 .9925653 1.048204

\_cons | 12.70716 5.328159 6.06 0.000 5.586458 28.90417

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

Note: \_cons estimates baseline odds.

A bivariate logistic regression was done to ascertain whether 7-item Autonomy scale of respondents was a predictor of whether they had insurance coverage or not. It turns out that Autonomy Scale was not a significant predictor of the possibility of insurance coverage of respondents (OR = 1.02, *p = 0.16*). The model was not significant (*p =0.16*) and it accounts for 0.21 % variation in whether the respondents have health insurance coverage or not.

1. (15 pts) Gender (***woman***; ref=man)

**. logistic w3insured i.woman**

Logistic regression Number of obs = 2,690

LR chi2(1) = 0.01

Prob > chi2 = 0.9219

Log likelihood = -468.78751 Pseudo R2 = 0.0000

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

w3insured | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

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

woman |

0. Male | 1 (base)

1. Female | 1.019114 .1968419 0.10 0.922 .6979319 1.488101

|

\_cons | 22.56863 3.229495 21.78 0.000 17.04908 29.87509

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

Note: \_cons estimates baseline odds.

A bivariate logistic regression was performed to determine whether gender was a predictor of the outcome that respondents are insured or not. Using man as a reference category, it was observed that woman are not at significantly greater odds for having health insurance coverage (OR = 1.02, *p = 0.922*). The overall model is not significant (*p = 0.92*) and accounts for approximately 0 % of the variation in whether respondents have health insurance or not.

1. (15 pts) Dummy for Wave 1 excellent/very good self-rated health (***evgsrhw1;*** ref=”excellent or very good”)

**. logistic w3insured ib1.evgsrhw1**

Logistic regression Number of obs = 2,701

LR chi2(1) = 23.13

Prob > chi2 = 0.0000

Log likelihood = -457.69932 Pseudo R2 = 0.0246

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

w3insured | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

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

evgsrhw1 |

0. Poor, Fair, or Good SRH | .3923386 .0779814 -4.71 0.000 .2657516 .5792235

1. Excellent or Very Good .. | 1 (base)

|

\_cons | 37.04762 5.79321 23.10 0.000 27.26818 50.33434

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

Note: \_cons estimates baseline odds.

A bivariate logistic regression was conducted to determine if respondent’s self rate health (0/ Poor, Fair, or Good – 1/Excellent or Very Good) is a predictor of respondent’s health insurance coverage. It turns out that relative to those who come under Excellent or Very Good for self rated health, those who come under Poor, Fair, or Good SRH have 60.8 percent lower odds of having a health insurance coverage (*p <0.001*). Overall, this model turned out to be statistically significant (*p < 0.001*) and accounted for 2.5 percent of variation in whether respondents had health insurance coverage or not.

1. (20 pts) Race/ethnicity (***raceeth;*** ref=NH White). Please disregard the warning you will receive for the Native Hawaiian/Other Pacific Islander category. Do not interpret Native Hawaiians/Other Pacific Islanders in your write-up.

**. logistic w3insured i.raceeth**

note: 6.raceeth != 0 predicts success perfectly

6.raceeth dropped and 1 obs not used

Logistic regression Number of obs = 2,690

LR chi2(5) = 6.55

Prob > chi2 = 0.2559

Log likelihood = -462.3833 Pseudo R2 = 0.0070

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

w3insured | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

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

raceeth |

1. NH White | 1 (base)

2. NH Black/ Afr Amer | .6590525 .3113243 -0.88 0.377 .2611127 1.663459

3. Hispanic/Latino (any k..) | .3903795 .1602518 -2.29 0.022 .174609 .8727853

4. NH Native Amer/Alask Nat | .428652 .262681 -1.38 0.167 .1289681 1.424713

5. NH Asian | .4420474 .4640147 -0.78 0.437 .0564894 3.459158

6. NH Nat Hawaiian or Paci.. | 1 (empty)

7. NH Other (Specify) | .8439086 .8682702 -0.16 0.869 .1123367 6.339705

|

\_cons | 24.88421 2.60386 30.72 0.000 20.27006 30.5487

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

Note: \_cons estimates baseline odds.

A bivariate binary logistic regression was performed to determine whether race of respondents (7-category variable ranging from Non-Hispanic White, Non-Hispanic Black, Hispanic/Latino, NH Native American/Alaskan Native, NH Asian, NH Native Hawaiian/Pacific Islander, and NH Others) was a predictor about whether respondents had health insurance coverage or not. Except for Hispanic/ Latino, all other race categories including (NH Black – OR =0.66, *p=0.38*, NH Native Amer – OR =0.43*, p =0.17*, NH Asian – OR = 0.44, *p =0.44*, and NH Other - OR =0.84, *p =0.11*) did not have significantly greater or lesser odds of having health insurance coverage, relative to NH Whites. Moreover, Hispanic/ Latino, had 61 percent lower odds of having health insurance coverage than NH Whites *(p =0.022*).

The overall model is not significant (*p =0.26*) and accounts for 0.7 percent variation in whether respondents have health insurance or not.

1. (20 pts) Wave 3 educational attainment (***educw3;*** ref=bachelor’s degree)

**. logistic w3insured ib4.educw3**

Logistic regression Number of obs = 2,696

LR chi2(4) = 23.09

Prob > chi2 = 0.0001

Log likelihood = -457.50687 Pseudo R2 = 0.0246

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

w3insured | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

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

educw3 |

1. less than high school d.. | .3516548 .1332163 -2.76 0.006 .1673627 .7388811

2. high school grad or GED | .6748334 .1951989 -1.36 0.174 .3828092 1.189627

3. some college/2 year col.. | .6507092 .1818909 -1.54 0.124 .3762296 1.125436

4. bachelor's degree | 1 (base)

5. more than bachelors | 2.384752 1.007912 2.06 0.040 1.041559 5.460121

|

\_cons | 28.2 6.416541 14.68 0.000 18.05383 44.04827

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

Note: \_cons estimates baseline odds.

A bivariate logistic regression model was conducted to determine if respondents’ level of educational attainment (ref = Bachelor’s Degree) was a significant predictor about whether respondents have health insurance coverage or not. Relative to those with a Bachelor’s degree, respondents with High school education or GED and some college/2 year col were not at significant odds of having health insurance coverage,(OR = 0.67, *p = 0.17*), (OR = 0.65, *p =0.12*), respectively.

AS for those with less than high school education, they had 64.8 percent lesser odds of having health insurance coverage with respect to those with a bachelor’s degree. (*p = 0.006*). The overall model is significant (*p <0.001*) and accounts for 2.46% of the variation in whether respondents are insured or not.