|  |  |  |  |
| --- | --- | --- | --- |
| **PM592: Regression Analysis for Health Data Science** |  |  |  |
| **Lab 9 – Logistic Regression Assumptions & Diagnostics**  **Data Needed:** *vote\_mhealth.csv* | | | |

**This lab is devoted entirely to the exercise.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Lab 9 Exercises** |  |  |  |  |  |  |

|  |  |
| --- | --- |
| Objective(s): | Assess the linearity assumption for logistic regression using 3 techniques, translate the concepts of confounding and effect modification to logistic regression, assess logistic regression model diagnostics and goodness of fit. |
| Datasets Required: | vote\_mhealth |

Research has shown that participation in voting is higher for those with greater resources, such as time, money, and social status. It was largely unknown whether mental health status had an effect on the likelihood of voting. A working hypothesis is that individuals who experience more depression also experienced more feelings of hopelessness and decreased efficacy. This is compounded by physical correlates of depression, such as lethargy and physical aches, that must also be dealt with.

Use this data set to explore whether mental health is related to the likelihood of voting. Examine age, education, and gender as possible confounders and effect modifiers.

vote96

1 if the respondent voted in the 1996 presidential election, 0 otherwise

age

Age of respondent

educ

Number of years of formal education completed by the respondent

female

1 if respondent is female, 0 if male

mhealth

Index variable which assesses the respondent's mental health, ranging from 0 (an individual with no depressed mood) and 9 (an individual with the most severe depressed mood).

1. Before you begin, determine whether you would like to center any of the variables.
2. Examine the form of the relationship between mental health on voting.
   1. Use the grouped smooth method to assess the linearity of mental health and voting. Provide the likelihood ratio test statistic and p-value for the categorical model vs. the ordinal/linear model.
   2. Use the LOESS method to assess the linearity of mental health and voting. Provide a graph showing the relationship between mental health and the logit of the outcome, based on the LOESS smoother.
   3. Use the FP method to assess the linearity of mental health and voting.
3. Examine the form of the relationship between all covariates (age, educ, female) and voting using whichever measure you’d like.
   1. Determine how age is related to the logit.
   2. Determine how education is related to the logit.
   3. Determine how gender is related to the logit.
4. Determine your preliminary final model.
   1. Re-assess the linearity of mental health with education and age in the model.
   2. Assess confounding of each covariate on the effect of mental health on voting. You can choose how to do this:
      1. One-by-one (good when you’re doing exploratory analyses for confounders)
      2. All-at-once (good when you are certain about the set of confounders you want to examine)
   3. Write your preliminary final model.
5. Assess your preliminary final model.
   1. What is the pseudo R2 for this model?
   2. How many covariate patterns are there? Based on this, would you trust the Pearson’s or Hosmer-Lemeshow test for goodness of fit? Compute the test statistic and p-value for GOF.
   3. List a few covariate patterns that might concern you. Why do they not fit well?
   4. Are you confident in your model? Or do you need to re-assess?
6. Present your final model.
   1. Present the results of your model in a professionally formatted table. Include the unadjusted and adjusted models.
   2. Write a conclusion that briefly describes your modeling approach and explains the effect of mental health on voting. Include relevant odds ratios, confidence intervals, and p-values.