

|  |
| --- |
| Computational Assignment #5: Logistic Regression Computations  *MSDS 410* |

In this assignment we will be calculating the various summative statistics that are associated with logistic regression, as well as fitting logistic regression models and interpreting the results. Students are expected to show all work in their computations. A good practice is to write down the generic formula for any computation and then fill in the values need for the computation from the problem statement. Throughout this assignment keep all decimals to three places, i.e. X.xxx. Students are expected to use correct notation and terminology, and to be clear, complete and concise with all interpretations of results. This computational assignment is worth 50 points. The points associated with each problem are given with the specific question.

Any computations that involve “the log function”, denoted by log(x), ***are always meant to mean the natural log function (which will show as ln() on a calculator).*** The only time that you should ever use a log function other than the natural logarithm is if you are given a specific base.

1. For the 2x2 table, determine the odds and the probabilities of texting while driving among males and females. Then compute the odds ratio of texting while driving that compares males to females. (5 points)

|  |  |  |
| --- | --- | --- |
| **Texting While Driving** | **MALE** | **FEMALE** |
| **YES** | 30 | 34 |
| **NO** | 10 | 6 |

2. Download the data file RELIGION.CSV and import it into R. Use R and your EDA skills to gain a basic understanding of this dataset. Please note, there is a variable labeled RELSCHOL. This variable indicates if a survey respondent attends a religiously affiliated private secondary school (1) or not (0). Use this dataset to address the following questions: (10 points)

a. Compute the overall odds and probability of attending a religious school, assuming this data is from a random sample.

b. Cross-tabulate RELSCHOL with RACE (coded: 0=non-white, 1=white). What are the probabilities that non-white students and white students attend religious schools? What are the odds that white students and non-white students attend religious schools? What is the odds ratio that compares white and non-white students?

c. Plot RELSCHOL (Y) by INCOME as a scatterplot. The INCOME variable is actually an ordinal variable that is associated with income brackets. This is an old dataset, so for example, INCOME=4 🡪 $20,000-$29,999. Is there a value of INCOME that seems to separate or discriminate between those attending religious schools and those that don’t? Create a variable that dichotomizes INCOME based on this value you observed. Call this new variable D\_INCOME. Cross-tabulate RELSCHOL with D\_INCOME. What are the probabilities that low income students and higher students attend religious schools? What are the odds that lower income students and higher income students attend religious schools? What is the odds ratio that compares lower and higher income students?

d. Plot RELSCHOL (Y) by ATTEND as a scatterplot. The ATTEND variable is the number of times the survey respondent attends a service during a month. Cross-tabulate RELSCHOL with ATTEND. Are the proportion profiles the same for those attending religious school versus not, across the values of the ATTEND variable? Is there a value of ATTEND that seems to separate or discriminate between those attending religious schools and those that don’t? Save this value for later.

3. First, fit a logistic model to predict RELSCHOL (Y) using only the RACE (X) variable. Call this Model 1. Report the logistic regression model and interpret the parameter estimates for Model 1. Report the AIC and BIC values for Model 1. (3 points)

4. Next, fit a logistic model to predict RELSCHOL (Y) using only the INCOME(X) variable. Call this Model 2. For Model 2, do the following: (6 points)

a. Report the logistic regression model and interpret the parameter estimates for Model 2. Report the AIC and BIC values for Model 2. How do these compare to Model 1?

b) Use the logit predictive equation for Model 2 to compute PI for each record. Plot PI (Y) by INCOME(X). At what value of X, does the value of PI exceed 0.50? How does this value compare to your visual estimate from problem 2c)?

5. Next, fit a logistic model to predict RELSCHOL (Y) using only the ATTEND(X) variable. Call this Model 3. For Model 3, do the following: (6 points)

a. Report the logistic regression model and interpret the parameter estimates for Model 3. Report the AIC and BIC values for Model 3. How do these compare to Models 1 and 2?

b) Use the logit predictive equation for Model 3 to compute PI for each record. Plot PI (Y) by INCOME(X). At what value of X, does the value of PI exceed 0.50? How does this value compare to your visual estimate from problem 2d)?

6. Finally, fit a logistic model to predict RELSCHOL (Y) using RACE, INCOME and ATTEND as explanatory (X) variables. Please consider INCOME and ATTEND to be continuous variables. Call this Model 4. For Model 4, do the following: (9 points)

a. Report the logistic regression model and interpret the parameter estimates for Model 4. Report the AIC and BIC values for Model 4. How does this model compare to Modesl 1, 2 and 3?

b. For those who attend religious service 5 days per month (attend=5) and have a family income of $20-$29,000 (INCOME=4), what are the predicted odds of attending a religious school for white and non-white students?

c. What is the adjusted odds ratio for race? Interpret this odds ratio.

7. For Models 1, 2 and 3, use the logit models to make predictions for RELSCHOL. Note, you will have to calculate the estimated logit and then convert it into PI\_estimates for each module. The classification rule is: If PI < 0.50, predict 0; otherwise predict 1 for RELSCHOL. Obtain a cross-tabulation of RELSCHOL with the predicted values for each model. Compare the correct classification rates for each of the three models. (6 points)

8. In plain English, what do you conclude about the relationship between a student’s race/ethnicity, religious service attendance, family income and attending a religious school? (5 points)