

Assignment 08

Multinomial Logistic Regression

This assignment is intended to give you experience working with multinomial logistic regression models to analyze polytomous outcome data. *Do not include any R syntax or output unless it is specifically required in the question.* Please submit your responses to each of the questions below. Please submit your responses to each of the questions below in a printed document. All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption. All tables should also have an appropriate caption.

This assignment is worth 15 points. Each question is worth 1 point unless otherwise noted.

For this assignment, you will use the file, *admissions.csv*. This file contains a sample of 500 randomly selected first-time, full-time freshman from the 2002 cohort at a large, midwestern research university classified by the Carnegie Foundation as having very high research activity. The variables in this file are:

- **outcome:** Whether the student graduated from the institution they enrolled in as a freshman, transferred to another institution, or did not graduate.
- **loan:** Amount of loan aid the student accepted (in \$1000)
- **scholarship:** Amount of scholarship offered to student (in \$1000)
- **cdw_count:** Number of “C” grades (C-, C, C+), “D” grades (D-, D, D+), and “W”s (withdraws) received in student’s first semester
- **first_gen:** Student is a first generation college student? (0=No; 1=Yes)
- **on_campus:** Student lived in on-campus housing (residence hall/dormitory) their first semester? (0=No; 1=Yes)

The data are from: Jones-White, D. R., Radcliffe, P. M., Lorenz, L. M., & Soria, K. M. (2014). Priced out?: The influence of financial aid on the educational trajectories of first-year students starting college at a large research university. *Research in Higher Education*, 55(4), 329–350. <https://doi.org/10.1007/s11162-013-9313-8>.

You will use these data to focus on three research questions:

1. Does living on campus increase the likelihood that a student will graduate from college?
2. Does living on campus increase the likelihood that a student will graduate from college, after accounting for whether the student is a first-generation college student?
3. Are the effects of living on campus different for first-generation and non-first-generation students?

To answer these questions you will fit and interpret several multinomial logistic regression models. Because transfer and drop-out are also related to academic performance and financial background, you will control for several variables related to students’ first semester academic performance and financial aid.

Explore Sample Effects of Living On-Campus

1. Compute and report the counts (and proportions) of students who graduated, transferred, and did not graduate by whether or not they lived on-campus. (Hint: Filter by whether the student lived on-campus and then compute the proportion of those students with the different outcomes.)
2. Based on these summaries, describe the effect of living on-campus on student outcomes.
3. Compute and report the counts (and proportions) of students who graduated, transferred, and did not graduate by whether or not they were first-generation students. (Hint: Filter by whether the student is a first-generation student and then compute the proportion of those students with the different outcomes.)
4. Based on these summaries, describe any differences in the effects of living on-campus on student outcomes between first-generation and non-first-generation students.

Build a Control Model

Fit a model (Model 1) that regresses the student outcome variable on the two financial aid predictors and the first-semester academic performance variable (three predictors total). In this model, and all future models you fit, use the students who did not graduate as your reference group.

5. Interpret the odds ratio for both coefficients associated with the `cdw_count` effect from this model.

Examine RQ #1

Fit a second multinomial logistic regression model (Model 2) that includes all the predictors in the control model and the predictor of whether or not a student lived in on-campus housing.

6. Carry out and report the results from a likelihood ratio test (LRT) comparing Model 1 and Model 2.
7. Based on the results of the LRT and the odds-ratios from Model 2, provide an answer to the first research question.

Examine RQ #2

Fit a third multinomial logistic regression model (Model 3) that includes all the predictors in the control model, the predictor of whether the student is a first-generation college student, and the predictor of whether or not a student lived in on-campus housing.

8. Use the `texreg()`, `htmlreg()`, or `screenreg()` function to output a table of the regression results (estimated coefficients, standard errors, and statistical significance) for this fitted model.
9. Based on the estimate and statistical significance of the coefficients from Model 3, provide an answer to the second research question.

Examine RQ #3

Fit a fourth multinomial logistic regression model (Model 4) that includes all the predictors in the control model, the predictor of whether the student is a first-generation college student, the predictor of whether or not a student lived in on-campus housing, and the interaction between these two focal predictors.

10. Carry out and report the results from a likelihood ratio test (LRT) comparing Model 3 and Model 4.
11. Based on the results of the LRT, provide an answer to the third research question.

Present Regression Tables of the Fitted Models

12. Create a regression table that includes the estimated coefficients (logit scale), standard errors, and statistical significance of the coefficients for the four models you fitted. You will need to include the results for both sets of coefficients for each of the four models. (Hint: You should have eight columns in this table.)

Adopt and Plot the Results for a “Final” Model

13. Based on the results of all the analyses, which of the four fitted models should be adopted? Explain.
14. Use the model you adopted in Question #13 to create a plot of the predicted probabilities for each of the three outcomes. In this plot, put the `loan` predictor on the x -axis. Show the effects of whether the student lived in on-campus housing via two separate lines. Finally, facet on both whether the student is a first-generation student and the outcome using `facet_grid()`. Set all the other predictors to their median value. (Hint: The resulting plot should have six panels (2 first-generation values x 3 outcome values) with two lines (on-campus housing) in each panel.) **(2pts.)**