



EPsy 8220—Methods for Categorical Response Data

Lab #4

16pts.

MULTINOMIAL REGRESSION

In this assignment, you will use the *Admissions.csv* data to examine the effect of financial aid on student success. To do this, you will fit several multinomial regression models to the data. To aid in the analyses you will carry out, the predictors in the data set can be grouped as follows:

- **Financial Aid:** unmet, need, loan, and scholarship
- **Demographic and Social Integration:** age19 and onCampus
- **Geographic Origin:** outstate and reciprocity
- **Academic Background:** firstgen, act, ap, and remedial
- **Academic Performance:** ratio, ccount, dcount, and wcount

Assignment Guidelines: Conduct the statistical analyses necessary to complete this assignment. Please submit a word-processed document that responds to each of the questions below. Any R output in the documentation file should be typeset and graphics should be resized so that they don't take up more room than necessary. *All tables and figures should be publication quality (i.e., follow the APA guidelines).*



Statistical Analyses

Because the sample size is very large, we are going to randomly split the sample into two samples to perform a cross-validation. The first sample will be referred to as the *training* sample and will consist of 2/3 of the observations ($n = 1710$). We will do our model building on the training sample. The second sample, consisting of 1/3 of the observations ($n = 855$), will be referred to as the *validation* sample. We will fit our chosen model to the validation sample to see how it performs on a 'new' sample.

In order to randomly split the admissions data into two samples, issue the following commands. (This assumes that the data were read into an object called `admissions`.)

```
#Make the sample split reproducible (put a different value
#in place of 32)
set.seed(32)
# Sample the ID numbers to be in the training sample
training.id = sample(1:nrow(admissions), 1710, replace = FALSE)
#Create the training sample
training = admissions[admissions$id %in% training.id, ]
#Create the validation sample
validation = admissions[!admissions$id %in% training.id, ]
```

Fit a series of multinomial models using the training sample.

1. Fit a multinomial model regressing academic outcome on the predictors in the Financial Aid grouping. Add the results of this fitted model to a regression summary table. Call this model, 'Financial Aid'. **(2pts)**
2. Fit each of the multinomial models that regresses academic outcome on the predictors in Financial Aid and another predictor grouping (e.g., the Geographic Origin model will include the predictors from Financial Aid and Geographic Origin). Add the results of these four additional fitted models to your regression summary table. Use the grouping names as the names for each model. **(3pts)**
3. Fit a multinomial model using all of the predictors. Pare this model down by removing any un-needed predictors (be careful not to remove any of the focal predictors). Report the 'final model' in your regression table. **(3pts)**



EPsy 8220—Methods for Categorical Response Data

Lab #4

16pts.

Fit the 'final' model to the validation sample.

4. Fit a multinomial model to the academic outcomes in your validation sample using the predictors you identified in the previous section for your 'final model'. Add the results to the regression table. **(3pts)**
5. Comment on any potential differences between these models for the training and validation samples. **(2pts)**
6. Based on the the results of the fitted models from your validation sample, comment on the effect of financial aid on academic outcomes. **(3pts)**