

## Final Exam

- I generated data from the GSS (General Social Survey) website cross-tabulating Newspaper Reading Habits (1=Every Day, 2=A Few times a Week, 3=Weekly, 4=Once a month, 5=Never), Highest Degree (1=Less than High School, 2=High School, 3=Junior/Community College, 4=Bachelor, 5=Graduate) and Gender (F=Female, M=Male). We will study cumulative logit models with Newspaper Reading Habits as the response and Highest Degree as an ordinal explanatory variable.
  - Fit the proportional odds main effects model for Newspaper Reading Habits and Highest Degree. Is Highest Degree significant? Is the proportional odds assumption appropriate?
  - Fit a non-proportional odds main effects model. Interpret the slope parameter estimates as odds/odds ratios and comment on any patterns in these odds.
- Divide the counts in the GSS data set by 10 before proceeding (the working data set is so large, that every term in the following analysis would be significant otherwise). Using the GSS data, find an appropriate loglinear model for counts based on the 3-way cross-classification of Reading Habits, Highest Degree, and Gender.
- Using the same data as in the previous problem, fit an additive cumulative logit model for Newspaper Reading Habits (response) as a function of Highest Degree and Gender. Use reference parameterizations for all three variables, with the reference category being the category with the highest marginal count for each variable. Write down equations for each of the logits, and interpret terms in the logit model comparing the most common response category to the least common response category.
- Consider the simple capture/recapture example, in which we assumed that counts  $(n_{00}, n_{01}, n_{10}, n_{11})$  had a Multinomial( $n, \pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}$ ) distribution. Under the marginal independence assumption, we found MLE  $\hat{n} = \frac{n_{11} + n_{10}}{n_{11}}$ . Using the delta method, find the asymptotic variance of  $\log(n)$  under the general multinomial sampling model.
- A student had a male and female artist (1) draw photos of a jar and (2) write a simple phrase. 48 subjects were asked to decide independently whether each photo was drawn by a male or female, and whether each phrase was drawn by a male or female. Results of the study can be found in the accompanying data set. The first variable is the gender of the respondent, and the next 4 variables are the respondents' guesses at Drawing 1 sample, Drawing 2 sample, Writing 1 sample, and Writing 2 sample.
  - Generate frequency tables of Drawing 1 vs. Drawing 2, Writing 1 vs. Writing 2, and Writing 1 vs. Writing 2 while controlling for Gender of Respondent. Comment on any marginal patterns.
  - Test marginal homogeneity models for Drawing 1 and Drawing 2; label the repeated measure as Sample. Interpret output parameters.
  - Test marginal homogeneity for Writing 1 and Writing 2 (Use Sample as the repeated measure again), while accounting for the gender of the respondent by including terms for `_response_`, gender, and their interaction in your model statement. Based on patterns in the 3-way frequency table, explain any significant results.