Homework 4 - Do not submit.

- 1. A student newspaper conducted a survey of student opinions about the Vietnam War in May 1967. Responses were classified by <u>sex</u>, <u>year</u> in the program and <u>one of four opinions</u>. The survey was voluntary. The data may be found in the dataset uncviet.
 - (a) Compute the <u>proportion</u> favoring <u>each policy</u> within <u>each year by sex combination</u>. Plot these proportions as year varies with a different line type for each policy. Plot men and women on separate panels. Your plot should look similar to <u>Figure 7.1</u>.
 - (b) Fit a proportional odds model with policy as the <u>response</u> with sex and year as predictors (include their interaction). Use the <u>weight</u> argument to set the <u>number of respondents</u> for each case. Why is it sensible to include an interaction term?
 - (c) Now fit a model with <u>main effects only</u>, excluding the interaction. Compare this model to the previous one. Which is preferred?
 - (d) Compute the <u>predicted proportion</u> for each case and plot in the same format as (a)
 - (e) Use the type="probs" argument to predict. You will need to select only one of the four probabilities for each case. Comment on the plot and compare it to (a).
 - (f) Compute the <u>raw residuals</u> as the difference between the <u>predicted</u> and <u>observed</u> <u>proportions</u>. Use the same format to plot these <u>residuals</u> as the <u>predicted proportions</u>. Comment on the plot and suggest how the definition of the residual might be improved.
 - (g) Examine the regression summary output to find the <u>coefficients</u> that are <u>relevant</u> to the opinions of <u>women</u>. What does the <u>significance</u> (or lack thereof) say about how the opinions of <u>women</u> vary across the <u>year groups</u>?
 - (h) Fit the multinomial model with main effects and compare to (c). You can get the deviance and chisq from the function gof in the multinom_housing.R script.
 - (i) What happens if you fit the model with interaction? How does this compare to the model in (b).
- The pneumo data gives the number of coal miners classified by radiological examination into one of three categories of pneumonoconiosis and by the number of years spent working at the coal face divided into eight categories.
 - (a) Make a plot showing how the proportion on miners in the three categories at each year point varies over time. Comment on the relationship.
 - (b) Treating the pneumonoconiosis status as response variable as nominal, build a model for predicting the frequency of the three outcomes in terms of length of service. What does the model say about the similarity of the proportions falling into the mild and severe categories?
 - (c) Would it be better to use log (year) as the predictor?
 - (d) Produce a plot of the predicted probabilities in the same format as (a).
 - (e) Fit a proportional odds model to the data. Take care to order response correctly. You will need to specify the number falling in each case using the weight argument. What is the estimated value of θ_1 and how should it be interpreted?
 - (f) Repeat the comparison with using log (year)
 - (g) Extract the predicted probabilities from the model and plot in the same format as (a). Compare to the predictions from the nominal model.
 - (h) Fit a hierarchical model to the status response. First fit a binomial response—normal and not normal. Interpret the effect of year on the odds of getting the lung disease.
 - (i) Now fit a binomial model for mild vs. severe lung disease. Is the year effect signficant. What is the probability of a mild disease within the diseased subgroup?
 - (j) Compute the predicted probabilities of the three categories by combining the two binomial model predictions. Plot in the same format as (a) and comment.