1. **Class Project**

   Indicate your intended class project. What will you present? You may talk to me before this homework due date; duplicate proposals will be settled on a first-come basis, or failing that in reverse alphabetical order.

2. **Berkeley Admissions Data and Log Linear Models**

   Demonstrate how to re-analyze the Berkeley Admissions Data as multinomial log linear models. Perform this analysis using the `glm` function. Is one approach preferred over the other?

3. **Poisson Overdispersion**

   A cohort of subjects, some non-smokers and others smokers, was observed for several years. The number of cases of cancer of the lung diagnosed among the different categories was recorded. Data regarding the number of years of smoking were also obtained from each individual. For each category the person-years of observation were calculated. The investigators wish to address the question of the relative risks of smoking.

   These data are available from the class website. The variables are CigsperDay, Years, Person-Years, and NumCases.

   Analyze these data to model the rate of lung cancer cases (per year) in terms of CigsperDay and Years. Is there evidence of lack of fit? Is there evidence of overdispersion?

   Compare fitting this model as a negative binomial model to using a quasi-likelihood approach with overdispersion. How does this choice affect the sample covariance of the estimated parameters?

   Are these data really best thought of as Poisson? If not, can the use of a Poisson model be justified?

4. **Some Light Poisson Model Theory**

   (a) Show that when an intercept is included in the Poisson model, the deviance simplifies to
   \[ D(y, \mu) = 2 \sum_{i=1}^{n} y_i \log \left( \frac{y_i}{\mu_i} \right) \]

   Is it legitimate to change deviance residuals accordingly?

   (b) M+N 6.3.

5. **Polytomous Data In Action**

   M+N Problem. 5.5; include an exploration of various types of models for the probabilities and an examination of overdispersion. Give details on how the latter was carried out. These data are available on the class website.