

Reading Assignment: Diggle, Heagerty, Liang, and Zeger (DHLZ), Chapters 7, 8, 9, 10

Written Assignment: Solutions are posted on the course web page

Final Exam: Monday, December 12, 7:00-9:30 am, in Snedecor 319

1. Consider the data from the cross-over experiment in example 8.1 in DHLZ. These data were analyzed on pages 817-828 of the notes. Use the GLIMMIX macro in SAS or the nlme function in S-Plus or any other software of your choice to fit a logistic regression model with random effects to those data. Fit the model

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{1ij}X_{2ij} + \eta_i$$

where π_{ij} is the conditional probability of an abnormal electrocardiogram for the i -th patient and the j -th treatment given η_i , Y_{ij} is the result for the j -th treatment applied to the i -th patient, $Y_{ij} | \eta_i \sim \text{Binomial}(1, \pi_{ij})$, and $\eta_i \sim \text{NID}(0, \sigma^2)$.

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{1ij}X_{2ij} + \eta_i$$

- a) Give interpretations for $\beta_1, \beta_2, \beta_3$.
- b) Give a formula for the conditional variance of Y_{ij} given the value of η_i .
- c) Give formulas for the unconditional (marginal) expectation and variance of Y_{ij} averaging across all possible values of η_i .
- d) Find the approximate REML estimate of σ^2 .
- e) Do the results from fitting this model to the data disagree with the results on page 820 in the notes? Explain.
- f) Do the results from fitting this model to the data disagree with the results on page 823 in the notes? Explain.
- g) Estimate the conditional probability of an abnormal electrocardiogram for a patient treated with the active drug, when the active drug is given in the first period. Give a standard error for your estimate.
- h) What are typical values for the conditional probability in part (g) for the patients in this population? Show how you obtained your answer.

2. In a study of the effects of energy-restricted diets on preventing skin cancer, mice were randomly assigned to either an ad lib (all you can eat) or an energy restricted diet. Restricting calorie intake cause the adrenal gland to produce more a substance that is thought to help prevent skin tumors caused by exposure to ultra violet radiation. To examine this, two thirds of the mice in each diet group were randomly selected to have their adrenal gland surgically removed. Within each diet group, half of the mice that had their adrenal gland removed received injections to restore levels of the particular substance, called GCH, thought to help prevent skin cancer. Consequently, there were six treatment groups corresponding to a 2×3 factorial arrangement; two diets (ad lib and energy restricted) and three levels of the surgery factor (no surgery, adrenal gland removed, adrenal gland removed and injections of GCH). Effects of these six treatments on weight gain or loss over the first 27 weeks of the experiment were analyzed in problem 4 on assignment 4. Here you are asked to analyze the effects of these six treatments on the formation of papillomas (small tumors on the skin). The data are posted in the file mpap.dat on the assignment section of the course web page. The variables are mouse id, treatment, diet, surgery, time, and the number of papillomas found on each mouse at 10, 14, 18, 22, 26, and 30 weeks. There is a separate line on the data file for the count at each inspection time on each mouse. Code for entering these data into SAS is posted as mpap.sas, and code for entering these data into an S-PLUS data frame is posted as mpap.ssc. Both sets of code produce plots of average counts across time for the six treatment groups. Some mice did not survive the entire 30 weeks. The treatment variable is coded from 1 through 6 as follows:

- 1 adlib diet, no surgery (adlib/no surgery)
- 2 adlib diet, adrenal gland removed (adlib/ADX)
- 3 energy restricted diet, no surgery (DER/no surgery)
- 4 energy restricted diet, adrenal gland removed (DER/ADX)
- 5 adlib diet, adrenal gland removed, GCH injection (adlib\ADX-GCH)
- 6 DER diet, adrenal gland removed, GCH injection (DER\ADX-GCH)

Note that mice that are lost to follow-up contribute to average numbers of papillomas early in the experiment but do not contribute to the average numbers of papillomas at later inspection times. We will assume that the missing data are missing at random, which means that the missing results for mice lost to follow up, if we could see them, would look like a random sample of results for mice in the respective treatment group.

- (a) First fit a separate log-linear model to compare treatment effects at each of the seven inspection times. Assume independent Poisson counts at each inspection time. Since mice were randomly assigned to treatment groups, you would not expect any significant treatment effects at time zero. At what time point do significant differences among treatments appear? Which treatments are protective in the sense that they delay or help prevent the formation of papillomas?

- (b) Is there any indication that the assumption of independent Poisson distributions for the papilloma counts used in part (a) is incorrect? If so, modify your analysis treatment effects in part (a) in an appropriate way.
- (c) Fit a marginal log-linear model to the count data using both treatments and time as classification variables. Include treatment \times time interactions in the model. This is the most general model you can fit to describe the change of the expected number of papillomas across time and treatments. Look for an appropriate covariance structure for counts taken at different time points on the same mouse. Report the simplest covariance structure that is appropriate for these data. Justify your conclusion. Are there any indications of diet or time effects?
- (d) Using the covariance structure identified in part (b), find the simplest log-linear model for the effects of time and treatments on the expected number of papillomas. Report a formula for your model using the estimated coefficients and show the standard errors for the estimated coefficients. Give a brief interpretation of your estimated model.
3. Use the GLIMMIX macro in SAS or the nmls function in S-Plus or any other software of your choice to fit the following log-linear model with random effects to the data in problem 2:

$$\log(\mu_{ijk}) = \lambda + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \eta_{ik}$$

where the conditional distribution of Y_{ijk} , the number of papillomas on the k -th mouse in the i -th treatment group at the j -th inspection time is $\text{Poisson}(\mu_{ijk})$, and $\eta_{ik} \sim \text{NID}(0, \sigma^2)$ are random mouse effects.

- a) Does this model indicate any significant time or treatment effects? If so, describe the effects.
- b) Suppose you wanted to select mice that are susceptible to developing papillomas for use in breeding a race of susceptible mice. Which of the mice that survived to the end of this study would you choose? (Assume that all of the mice used in this study were males).