

BIO 226 Homework Assignment 8
Due in class on Tuesday May 5, 2015

Purpose:

To introduce the fitting of a generalized marginal models.

Instructions:

- 1) For each question requiring data analysis, support your conclusions by including only the relevant SAS output in your answer.
- 2) Include your SAS program as an appendix to your solutions.
- 3) Late homework will not be graded unless you make prior arrangements with the Instructor.

Background:

The Skin Cancer Prevention Study, a randomized, double-blind, placebo-controlled clinical trial, was designed to test the effectiveness of beta-carotene in the prevention of non-melanoma skin cancer in high-risk subjects. A total of 1,683 subjects were randomized to either placebo or 50 mg of beta-carotene per day and were followed for up to 5 years. Subjects were examined once per year and biopsied if a cancer was suspected to determine the number of new cancers per year. The outcome variable, Y , is a count of the number of new skin cancers per year.

Selected data from the study are in the dataset called “skin.txt” on the course web site. Each row of the dataset contains the following 9 variables: ID, Center, Age, Skin, Gender, Exposure, Y , Treatment, Year. These variables take values as follows:

ID:	Subject identifier number
Center:	Identifier number for center of enrollment.
Age:	Subject's age in years at randomization
Skin:	Skin type (1=burns; 0 otherwise) [evaluated at randomization and doesn't change with time]
Gender:	1=male; 0=female
Exposure:	Count of number of previous skin cancers [prior to randomization]
Y :	Count of number of new skin cancers in the Year of follow-up
Treatment:	1=beta-carotene; 0=placebo
Year:	Year of follow-up after starting randomized treatment

You may assume that the counts of new skin cancers, Y , are from exact one-year periods (so that no offset term is needed).

Questions:

1. Your collaborator is interested in assessing the effect of treatment on the incidence of new skin cancers over time. As the statistician on the project, provide an analysis of the data that addresses this question. Please,
 - i) Provide a short table providing a descriptive summary of the mean count of new skin cancers by Treatment for each Year. Briefly comment on changes in incidence of new skin cancers by randomized treatment over time.
 - ii) Provide an algebraic definition for a generalized linear marginal Poisson regression model in which the only effects are for the intercept, Year (as a continuous variable) and $\text{trt} \times \text{Year}$. Make sure you write out the full model. Provide a table of results that you obtain from fitting the model in SAS that includes the regression coefficient estimates, empirical standard errors 95% confidence interval, and p-value for testing the hypothesis that the true parameter is zero. What do you conclude from this model about the effect of treatment?
 - iii) Properly interpret each of the parameter estimates from your analysis in (ii) on the untransformed scale.
 - iv) In the form of the results section of a scientific abstract (3 or 4 sentences), state with justification what you conclude about the effect of beta-carotene versus placebo based on your analysis.
2. A secondary aim of the study was to evaluate risk factors for increased incidence of new skin cancers. To accomplish this goal,
 - i) Fit an appropriate generalized linear marginal model in SAS which includes Year (a linear effect), Treatment, Center, Age, Skin, Gender and Exposure as covariates. Make sure you fully write out your chosen model.
 - ii) Provide a table from your SAS output that includes parameter estimates, empirical standard errors and p-values.
 - iii) In a short paragraph (<150 words), summarize the key findings from this model.
 - iv) Fit the model from 2(i) assuming there is no overdispersion (that is, the overdispersion parameter is equal to 1). Briefly describe the effect of this assumption on the results of the analysis.