

BIOS 665: Problem Set 5
Assigned: November 07, 2019
Due: November 21, 2019

Notes: For all hypothesis tests, please state the statistical method you are using, the null hypothesis, the test statistic, the p-value, and the interpretation of the test using significance level of 0.05, unless otherwise stated. Please round all estimates to **three decimal** places.

Helpful hints: For estimates and tests, simply copying and pasting SAS output without any commentary will not earn full credit, especially on exams. Highlighting is not considered commentary. However, commentary can be as simple as: The 95% CI for the odds ratio is (____, ____).

1. A researcher wishes to describe the relationship of time to occurrence of dental caries with treatment and follow-up period. The contingency table shown below displays data from a randomized study to compare an existing dental treatment to an experimental dental treatment for the prevention of dental caries (cavities). An event was defined as the first occurrence of at least one cavity at a follow-up visit (such that a participant could only have experienced the event once during the study). Following treatment, there were biannual follow-ups lasting for a period of 2 years.

Treatment	No dental caries by 2 years	Follow-up period for dental caries				Follow-up period for withdrawal				Total
		0-6 months	6-12 months	12-18 months	18-24 months	0-6 months	6-12 months	12-18 months	18-24 months	
Existing	37	3	5	14	35	3	2	7	12	118
Experimental	58	2	3	9	17	2	1	9	17	118

- a) Construct the life table format for the data separately for each treatment group.
 - b) Provide life table estimates for the cumulative probabilities (and corresponding standard errors) for no occurrence of dental caries by the end of each of the four periods for each treatment group. State the assumptions for these estimates, and assume withdrawals may be treated as not having dental caries at the time of withdrawal.
2. From the data in the contingency table in question 1; the researcher would like for you to fit a piecewise exponential regression model (including interaction terms) to the data presented in the table below. Use '0-6 months' as the reference group for follow-up period, and 'Existing treatment' as the reference group for treatment.
 - a) Mathematically specify the structure of the model that includes main effects for treatment and time, as well as their interaction. Mathematically define all variables used in the model.
 - b) Fit the model you specified in (a) and assess the effect of interaction between treatment and time.

- c) Regardless of your conclusion from part (b), fit the main effects model (i.e., the model without interaction terms) for treatment and follow-up period.
- i). Interpret the model parameters.
- ii). Provide a 95% confidence interval for the model parameter corresponding to the treatment variable.
- d) For each treatment group, provide model-predicted values using the model in part (c) for cumulative probabilities of no occurrence of dental caries by 6 months, 12 months, 18 months, and 24 months, respectively.
3. You are asked to analyze a dataset from a randomized, controlled trial for women of childbearing age to evaluate the effects of an educational program intervention for health. One response variable of interest is the participants' self-rating of health as either "good" or "poor". The researchers would like to assess the effect of the intervention on self-rated health across the follow-up period. The REPEATED.sas7bdat data set contains $n=80$ women enrolled in this trial. These data were measured at time of randomization, as well as 3 follow-up points in time: 3 months, 6 months, and 12 months post-randomization. Each observation in the dataset contains values for the following variables:
- ID: unique participant identification code
 - TIME: the visit number for this observation of this participant
 - 2 corresponds to the 3 month post-randomization visit
 - 3 corresponds to the 6 month post-randomization visit
 - 4 corresponds to the 12 month post-randomization visit
 - RX: the group to which the participant has been randomized
 - control
 - intervention
 - HEALTH: participant's self-rated level of health for this visit
 - Good
 - Poor
 - AGE_GROUP: participant's age group at time of randomization
 - 15-24 (years old)
 - 25 to 34 (years old)
 - 35+ (years old)
 - BASE: participant's self-rated level of health at randomization
 - Good
 - Poor
- a) Fit a GEE repeated measures logistic regression model across all study follow-up visits (not including at the time of randomization) to describe the marginal relationship of participants' self-rating of good health to the main effects of randomized group, visit (as a class variable), health self-rating at time of randomization, and age group as explanatory variables together

with group by visit interaction. Use 15-24 year old women in the control group with a good health assessment at randomization as your reference group, and use 3 months post-randomization as your reference visit. Specify an exchangeable working correlation structure.

- i. Assess goodness of fit through consideration of other pairwise interactions between explanatory variables in the model. Also assess whether group by visit needs to remain in the model.
 - ii. Do the results of i) affect your choice for the final model? Explain in 1-2 sentences.
 - iii. For the final model based on ii), present the table of parameter estimates, along with a standard error, test statistic, and p-value for each.
- b) Using your final recommended model from part (a), separately for each follow up visit, provide the odds ratio and corresponding 95% confidence interval that pertains to the intervention effect as estimated by the model for that visit.
- c) Briefly summarize your findings related to the intervention effect(s) (adjusted for baseline status and age group, and addressing time in a way that is consistent with your final model) in 2-3 sentences.