

Homework 3

Biost 540

General Instructions

- Students may discuss with each other but each of you will be required to submit the work in your own writing.
- Grading will be based on completion (3 pts), accuracy (3 pts), work shown (3 pts), and neatness (1 pt).
- Be sure to show work for all problems. R code should not appear in the main body of the homework; however, the code should appear at the end of the assignment as an Appendix. It should be possible for someone to use the code to reproduce any figures or numeric results.

Problem 1: UK Understanding Society Study

The Dataset

In this problem, we will still focus on the UK Understanding Society Study which we have explored in Homework 2. Below are the description of the variables.

- pidp: subject identifier
- wave: data collection wave (1-9), annual follow up
- age_dv: age at data collection time
- sex_dv: sex (1: female)
- scghq1_dv: a well-being measure at each follow up

Questions

- a) In Homework 2, we have studied how age and sex factors are related to well-being without considering interactions. Please provide a statistical evidence on whether the association between age and well-being changes for different sex. You can focus on LMM with subject-specific random intercept.
- b) Perform the similar analysis as in a) by fitting GEE models. Consider the models with independent, exchangeable and AR1 working correlation matrix.
- c) Compare the estimates and hypothesis tests between the LMM and the GEE models for similarities and differences. To test whether the association between age and well-being differs at different sex category, which model would you choose and why? Interpret the findings of your chosen model.

Problem 2: Six City data

The Dataset

Consider the Six City data set which describes Mother's smoking behavior and childhood respiratory disease. Below are the description of the variables.

- id: Child's ID
- resp: binary indicator of respiratory disease
- age: standardized age (ages 6-9 minus 8)
- smok: mother's smoking
- aXs: interaction of age and smok

Questions

Please fit a GLMM, GEE, and transition model regressing indicator of respiratory disease on age, maternal smoking status, and interactions. Describe the output of each model (you can focus on the smok variable when printing output). Also provide a brief description of how the assumptions made by each model differ from one another. Hint: see Lecture 3 pg 49, 56, 64 for help with interpretation and assumptions.

Problem 3: Seizure data

Consider the seizure dataset. The study consisted of 59 patients randomized to the anti-epileptic drug progabide, or to placebo in addition to standard chemotherapy. Over an 8-week period prior to randomization, the “baseline” number of seizures was recorded for each participant. Over four subsequent follow-up time periods the number of seizures in each 2-week period was recorded. Listed below are the variables in the dataset

- id: patient id
- age: age of the patient
- tx: treatment (placebo or progabide)
- y0, y1, y2, y3, y4: number of seizures by visit times. y0 refers to the baseline number of seizures.

We are interested in evaluating whether the drug progabide is effective at reducing the rate of epileptic seizures.

- Produce a graphical/tabular summary of the rates of seizure incidence at baseline and over the subsequent study visits by treatment arm. Feel free to summarize the individual rates and/or the average rates within each arm. Also pay particular attention to the length of the “at risk” period at each time point and use informative labels for study visits.
- Describe and fit a Poisson generalized estimating equation (GEE) to the data to evaluate whether treatment has an effect on the rate of seizures. Consider relevant covariate adjustments. Be sure to provide a relevant interpretation of your model: describe the parameter estimate and 95% CI and the conclusion you draw from the hypothesis test. Also, discuss the implications of working correlation specification on the validity of your inference. Hint 1: include an log observation time as an offset in your model to ensure the rate is being modelled. See Lecture 3 Slide 45 for details. Hint 2: the poisson GEE should return coefficients on the log incidence rate ratio scale. You may want to report the exponentiated coefficients, which correspond to the fold changes in incidence rates associated with predictors. Hint 3: Also you may need to use the deltamethod function from the msm package if your model uses an interaction.