Homework3

BIOS6643 Fall 2021

8/20/2021

Question 1 Models for Beta Carotene data

For the Beta Carotene data (see the description of the data and the data itself in another link in the Data module). For parts **a** and **b**, model *time* and *group* as class variables, and include $group \times time$. In order to account for repeated measures over *time*, specify the UN error covariance structure.

a. Conduct a test to compare the 30 and 60mg BASF trends over *time* to see if they differ, i.e., an interaction test, but only involving these 2 *groups*.

b. Conduct a test to compare to see if the 12 week - baseline value differs between the 4 *groups*.

c. Consider the model that uses *time* as continuous, with up to cubic effects, plus interactions between group and time (up to cubic). How does this model compare with the one that uses *time* as class (plus interactions)? Discuss in a paragraph.

d. Modeling the data using $Time0$ as a covariate value, with the remaining *times* as repeated measures on the outcome (6, 8, 10, 12 weeks). What are pros and cons of this approach, relative to using all measures as outcome values in a longitudinal model? In particular, focuses on the modeling of the repeated measures, how fixed effects need to be specified, and impact of modeling of *time* as class versus continuous.

e. For the model in part **d**, estimate the linear, quadratic and cubic trends for the model that uses *time* as a class variable.

Question 2

Consider a study where *subjects* in 3 *groups* (e.g., race or treatment) are observed over 3 equally spaced *times* and some health outcome, y, is measured. Unless otherwise mentioned, include a random intercept for subjects to account for the repeated measures. For simplicity, use 2 *subjects* per *group*.

a. Consider modeling *group* and *time* as class variables, plus interaction. Write statistical models and the $\boldsymbol{X}$ matrix for the following cases.

   i. No restriction placed on the model. i.e., write the less-than-full-rank statistical model.

   ii. A set-to-0 restriction is placed on the parameters associated with highest levels.

b. Show that the linear trend for one *group* compared to another (say *GroupA* versus *GroupB*) is estimable by showing that $\boldsymbol{L} = \boldsymbol{LH}$, where the Moore-Penrose inverse is used in calculating $\boldsymbol{H}$. First you need to construct $\boldsymbol{L}$. (As a check, you can repeat using SAS's g-inverse in calculating $\boldsymbol{H}$, but you don't need to turn that in.)

c. How would answers in a change in part **a** if an AR(1) structure for $\boldsymbol{R}$ is included? (You do not need to rewrite entire models, just mention what changes).

d. Say that $Time$ is treated as continuous (i.e., not included in the CLASS statement in SAS or factor argument in R). Rewrite either the full-rank or less-than-full-rank model (clearly specify which one) and $\boldsymbol{X}$ matrices in **a**. Say the linear term for $Time$ is sufficient.

e. Say that the times of observation were at 0, 1 and 6 months rather than equally spaced.

   i. Would it be appropriate to treat $Time$ as a class variable in this case? Explain.

   ii. Suggest a structure for $\boldsymbol{R}_1$ and write it out.