P8157 - Analysis of Longitudinal Data, Fall 2020
Homework - 2
Due : October 13, 2020

September 28, 2020

Question - 1

In a randomized, double-blind, parallel-group, multicenter study comparing two oral treatments (denoted A and B) for toe-nail infection (De Backer etal., 1998; also see Lesaffre and Spiessons, 2001), patients were evaluated for the degree of onycholysis (the degree of separation of the nail plate from the nail-bed) at baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48 thereafter. The onycholysis outcome variable is binary (none or mild versus moderate or severe). The binary outcome was evaluated on 294 patients comprising a total of 1908 measurements. The main objective of the analyses is to compare the effects of oral treatments A and B on changes in the probability of the binary onycholysis outcome over the duration of the study. The raw data are stored in an external file: toenail.dat Each row of the data set contains the following five variables: ID,Y,Treatment,Month,Visit. The binary onycholysis outcome variable Y is coded 0 = none or mild, 1 = moderate or severe. The categorical variable Treatment is coded 1=oral treatment A, 0=oral treatment B. The variable Month denotes the exact timing of measurements in months. The variable Visit denotes the visit number (visit numbers 1-7 correspond to scheduled visits at 0, 4, 8, 12, 24, 36, and 48 weeks).

1. Consider a marginal model for the log odds of moderate or severe onycholysis. Using GEE, set up a suitable model assuming linear trends. Use month as the time variable. Assume "exchangeable" correlation for the association among the repeated binary responses.

2. Provide Interpretations for the coefficients in your model.

3. From the results of your analysis what conclusions do you draw about the effect of treatment on changes in the severity of onycholysis over time? Provide results that support your conclusions.

4. Try Different correlation structures. Is the analysis and inference sensitive to this choice?

Question - 2

The Skin Cancer Prevention Study was a randomized, double-blind, placebo-controlled clinical trial of beta carotene to prevent non-melanoma skin cancer in high-risk subjects (Greenberg et al., 1989, 1990; also see Stukel, 1993). A total of 1805 subjects were randomized to either placebo or 50 mg of beta carotene per day for 5 years. Subjects were examined once a year and biopsied if a cancer was suspected to determine the number of new skin cancers occurring since the last exam. The outcome variable is a count of the number of new skin cancers per year. The outcome was evaluated on 1683 subjects comprising a total of 7081 measurements. The main objective of the analyses is to compare the effects of beta carotene on skin cancer rates. The raw data are stored in an external file: skin.dat Each row of the data set contains the following 9 variables: ID,Center, Age, Skin, Gender, Exposure, Y, Treatment and Year.

Note: The outcome variable Y is a count of the of the number of new skin cancers per year. The categorical variable Treatment is coded 1=beta carotene, 0 =placebo. The variable Year denotes the year of follow-up. The categorical variable Gender is coded 1 male, 0 female. The categorical variable Skin denotes skin type and is coded 1 = burns, 0 otherwise. The variable Exposure is a count of the number of previous skin cancers. The variable Age is the age (in years) of each subject at randomization.

1. Set up a suitable GEE model for rate of skin cancers with Treatment and Year as covariates.

2. Provide Interpretations for the coefficients in your model.

3. From the results of your analysis what conclusions do you draw about the effect of beta carotene on the rate of skin cancers? Provide results that support your conclusions.

4. Repeat the above analysis adjusting for skin type, age, and the count of the number of previous skin cancers. What conclusions do you draw about the effect of beta carotene on the adjusted rate of skin cancers?

5. Try Different correlation structures. Is the analysis and inference sensitive to this choice?

6. Do you need to account for overdisperion. Comment.