Mock Exam 2021

2AMS10: Longitudinal Data Analysis

A Multi-center Randomized Clinical Trial on Blood Process Devices for Open Hart Surgery

October 17, 2021

Edwin van den Heuvel

Data Description

A multi-center randomized clinical trial was conducted to investigate how different blood process devices helps the clinical outcome of patients who undergo open hart surgery. One of the clinical outcomes or responses of the trial was the hemoglobine values (Hb) of patients measured at six sequential days starting at the first day after the day of the operation. For this outcome the trial is considered a longitudinal trial. The trial data (presented in long form) included the following variables:

1. Trial design variables

   (a) CENTER: indicating hospital 1, 2, 3, 4, 5, and 7

   (b) ID: indicating a patient number (which is numbered uniquely)

   (c) TIME: the day at which Hb is measured

   (d) TRT: Treatment indicator with four levels (0 = control; 1 = device; 2 = filter; 3 = device + filter)

2. Patient baseline characteristics

   (a) AGE: the age of the patient at the time of the surgery (in years)

   (b) SEX: the gender of the patients (0 = female; 1 = male)

   (c) HEIGHT: the height of the patients (in cm)

   (d) WEIGHT: the weight of the patients at the time of the surgery (in kg)

   (e) ANEMIA: a variable that indicates whether the patient has a reduced number of red blood cells at the moment of the surgery (0 = No; 1 = Yes). It is defined on the basis of the hemoglobin value. Less than 13 g/dL for males and 12 g/dL for females indicates anemia.

3. Patient operation characteristics

   (a) TYPE: indicating the type of surgery (0 = valve replacement; 1 = bypass operation; 2 = bypass and valve replacement)

   (b) DURATION: the duration of the surgical operation

   (c) BLOOD: Indicating if the patient received external blood products (0 = No; 1 = Yes)

   (d) RESP: the Hb value (in g/dL)

Questions

1. In this exercise we will assume that the response (Hb) that is observed at six different days can be modeled with a mixed effects ANOVA model. We will only consider the variables "Center", "ID", "Time", "Treatment", and the "Response".

   (a) Write down the **full** ANOVA model in mathematical terms when we assume that center is considered random.

   (b) Provide the expected mean squares of this ANOVA model and give the variance component estimators if you assume that each center contains the same number of patients.

   (c) Use Proc HPMIXED to fit the ANOVA model and provide the solutions of the fixed effects and the variance component estimates (HPMIXED works similar as MIXED).

   (d) Why would the variance component estimates described in (c) be different from the proposed estimates in (b)?

   (e) Formulate the traditional null and alternative hypotheses for treatment in mathematical terms.

   (f) Assume that the only random terms in the model are center and subject within center. Use Proc MIXED to calculate a likelihood ratio test to make a decision on the hypothesis in (e), provide the degrees of freedom, the P-value, and formulate your conclusion.

   (g) The trial statistician decides to apply a linear normalization for the full ANOVA model using the center average. What are the consequences for the variance components?

2. In this exercise we will analyze the trial data using several marginal mixed models. Again we will concentrate on the variables "Center", "ID", "Time", "Treatment", and the "Response".

   (a) Use Proc MIXED to fit a marginal model with time, treatment, and its interaction as categorical variables as fixed effects and an unstructured six dimensional distribution for the residuals. Provide the SAS code and the output of this model.

   (b) Based on the model in (a) what is the correlation between the first and third Hb value of one patient? Explain both theoretically and provide the estimate.

   (c) Use Proc MIXED to fit a mixed model that includes time, treatment, and its interaction as categorical variables as fixed effects and which uses center as random effect and a Toeplitz structure for the residual error of patients. Provide the SAS code and the output of this model.

   (d) Based on the model in (c) what is the correlation between the first and third Hb value of one patient? Explain both theoretically and provide the estimate.

   (e) Between model (a) and (c) which model do you prefer? Explain

(f) Fit the same model as in (c) but now exclude the random effect of center. Perform a likelihood ratio test between model (c) and formulate your conclusion based on an appropriate critical chi-square value.

3. In this exercise we will analyze the trial data using several subject-specific mixed models. We will built polynomial time profiles and use the patient characteristics to understand which variables affect the time profile. Here we will only use the first five time points.

   (a) Use Proc MIXED to fit a quadratic time profile for each patient separately and an auto-regressive one model for the residuals. The coefficients of the quadratic profile are assumed independent. Provide the SAS code and the output of this model. Based on the output, what model would you recommend?

   (b) Investigate if the average quadratic time profile for patients who receive blood products and who do not receive blood products are different. Assume that the intercept and slope of the quadratic profiles are subject specific and independent (no subject-specific quadratic term). Use a likelihood ratio test to test your hypothesis. Report the null hypothesis, the value of the likelihood ratio test, the degrees of freedom, and your conclusion.

   (c) Based on the two average quadratic time profiles for patients who receive and do not receive blood pressure, estimate the difference at day three and day 6, together with a 95% confidence interval.

   (d) Use the model in (b) and assume that the $\boldsymbol{G}$ matrix depend on the variable blood product. Investigate if there are outliers. If you believe there are outliers, report the number of outliers and the subject numbers and time points. If you do not think there are outliers, then explain why.

   (e) Which single variable from the set of variables {TRT, AGE, SEX, WEIGHT, ANEMIA, TYPE, DURATION, BLOOD} describes the best subgroups of quadratic time profiles (using the model in (b))? Report the mathematical model and explain why this model would fit the best.

Good luck with the examination!