Exam 2021

2AMS10: Longitudinal Data Analysis

A Multi-center Randomized Controlled Trial on Cognition Training in the Elderly

November 5, 2021

Edwin van den Heuvel

Data Description

A multi-site randomized controlled trial was conducted to investigate the effect of three different cognition trainings on the cognitive performance of independent-living elderly (i.e., 65 years and older). The study included 2802 participants collected from six different area's of the United States. They were randomly allocated to one of four different trainings (memory, reasoning, speed, and a control). The study investigated performances on memory, reasoning, and speed. The participants were first tested on cognition before they undergone a cognition training (the so-called baseline period) and they were tested four times after training. The first time just after the training session and then approximately annually. The goal is to understand if treatment has an influence on the cognitive trajectories of the participants.

1. Trial design variables:
   (a) `SITE:` indicating an area of the USA
   (b) `ID:` indicating a participant number
   (c) `DAYS:` the number of days measured from the time of the training
   (d) `PERIOD:` a categorical variable indicating the period of the observation
   (e) `ORDER:` a numerical support variable for the period
   (f) `TRT:` A treatment indicator with four levels (1 = memory; 2 = reasoning; 3 = speed; 4 = control)
2. Participant baseline characteristics:
   (a) `SEX:` a binary variable for sex of the participant (0 = male; 1 = female)
   (b) `AGEC:` a categorical variable for age group (0 = [65, 70); 1 = [70, 75); 2 = $\geq$ 75)
3. Response variables:
   (a) `HVLT:` Hopkins verbal learning test
   (b) `ALVT:` Auditory verbal learning test
   (c) `WORD:` A speed score measuring the number of correct words out of 30 tested words

Calculation Exam Grade

The maximum number of credits for each of the questions of the exam are provided in the following table. You will receive 4 credits for free (if you have spelled your name correct). The total number of obtained credits, divided by four will be the grade of the exam.

Table 1: Maximum number of credits for each of the questions

Question 1							Question 2			Question 3		
1a	1b	1c	1d	1e	1f	1g	2a	2b	2c	3a	3b	3c
3	4	2	2	3	3	3	3	2	3	3	2	3

Questions

1. In this exercise we will focuse on the response `HVLT` in the sixth site (`SITE`=6), observed at the first and second moment after training (i.e., post-training and first annual moment). We will consider the factors subject (`ID`), period (`RERIOD`), and training (`TRT`). Make sure that you create an appropriate data set or otherwise use an appropriate selection statement for the analysis questions.

   (a) Write down the **full** ANOVA model in mathematical terms with all relevant assumptions (assuming a balanced design).

   (b) Provide the expected mean squares of this ANOVA model in mathematical terms, give the variance component estimators, and provide the F-test for all terms in the model (assuming a balanced design again).

   (c) Determine the quadratic term mathematically for the effect of training in the expected mean square for training. Thus determine the first term of $\mathbb{E}[MS_{\text{TRT}}]$.

   (d) Since the design is not entirely balanced, what ANOVA estimation technique would you recommend for your ANOVA model? Explain your answer.

   (e) Fit your ANOVA model and provide the complete ANOVA table. Based on this table, do you believe there is evidence for a training effect? Explain your answer.

   (f) Based on your ANOVA model, calculate the correlation between the two scores of participants with their 95% confidence intervals? Explain your opinion on the strength of this correlation.

   (g) To investigate any effect of training, conduct a likelihood ratio test. Formulate the null and alternative hypotheses in mathematical terms, report the value of the test statistic, its degrees of freedom, and its P-value. Do you believe there is a training effect? Explain your answer.

2. In this exercise we will analyze the response on reasoning (`WORD`) at baseline and at post training using several marginal mixed models. We will concentrate on the variables subject (`ID`), location (`SITE`), period (`PERIOD`), training (`TRT`), sex (`SEX`), and age group (`AGEC`), although these variables will not be used in all questions.

   (a) Fit a marginal model with `PERIOD` and `TRT` and their two-factor interaction as fixed effects, assuming that the participants have an unstructured variance covariance matrix. Provide the SAS code and report the output of the overall test statistics and the residual variance-covariance matrix. Explain why the interaction effect `TRT*PERIOD` could not be tested.

   (b) Based on the model in 2(a), formulate the null hypothesis and the alternative hypothesis for an effect of training from baseline to post training. Determine the effect sizes with their 95% confidence intervals for each of the four trainings. Report whether the cognition training has increased the response.

   (c) Fit a marginal model with `SITE`, `PERIOD`, `TRT`, `SEX`, and `AGEC` and with all two-factor interactions with `TRT` (except of course for interaction `TRT*PERIOD`) using

the unstructured variance-covariance stucture. Determine whether you can reduce the complexity of the variance-covariance matrix and the fixed effects. Provide your model selection approach, your final marginal model, and the details of how you came to this model

3. In this exercise we will analyze the trial data using several subject-specific mixed models for the response on memory (`AVLT`). We will built polynomial time profiles and use the different trainings (`TRT`) to understand how training affect the the time profile. Here we will use only the observations after training (no baseline data). Create an appropriate data set and make a new time variable: `TIME=DAYS/365`.

   (a) Fit a linear time profile for each patient separately and assume that the residuals are independent. The coefficients of the linear time profile are assumed dependent. Make sure that you have an average profile for each training (`TRT`). Provide the SAS code, the output of this model, and report the conclusions about the effect of training. [**Tip:** Be aware that the model can take some time to fit]

   (b) For the model in (a) determine the time point for which variability in `AVLT` is minimal when the reasoning training (`TRT=2`) is applied. Report this time point and provide an estimate of the minimal variation.

   (c) Calculate the 95% upper finite sample prediction limit for the control training (`TRT=4`). Report the formula as function of years.

Good luck with the examination!