2AMS10: Longitudinal Data Analysis 2022-2023 Assignment

October 4, 2022

Edwin van den Heuvel
e.r.v.d.heuvel@tue.nl
Professor in Statistics

Mathematics & Computer Science

Repeated CD4 Counts

A randomized controlled trial

Acquired Immunodeficiency Syndrome (AIDS)

- CD4 cells are white blood cells that fight infection in the body
  - The more you have the better it is
  - A normal range is 500 to 1500 cells/mm^3
- CD4 cells kill the human immunodeficiency virus (HIV)
  - When HIV progresses the number of CD4 decline
  - A CD4 count less than 200 cells/mm^3 is a diagnosis for HIV
- If HIV is untreated it may lead to AIDS
  - If HIV is controlled the CD4 cells typically increase
  - People with controlled HIV can have healthy lives

Repeated CD4 Counts

A randomized controlled trial

Randomized Controlled Trial (RCT)

- Four HIV treatments
  - [TRT=1]: 600mg of zidovudine alternating monthly with 400mg didanosine
  - [TRT=2]: 600mg of zidovudine plus 2.25mg of zalcitabine
  - [TRT=3]: 600mg of zidovudine plus 400mg of didanosine
  - [TRT=4]: 600mg of zidovudine plus 400mg of didanosine plus 400mg of nevirapine
- Treatments are randomly assigned to 182 AIDS patients
- Three-time points with CD4 counts
  - Baseline: just before treatment was allocated
  - Eight-weeks after treatment
  - Sixteen-weeks after treatment

Repeated CD4 Counts

A randomized controlled trial

Observed Data Set:

- Type of variables involved
  - ID: Indicator for patient
  - TRT: Indicator for treatment {1,2,3,4}
  - AGE: The age in years at baseline
  - SEX: The sex of the patient (Female = 0; Male = 1)
  - TIME: Time point (1 = baseline; 2 = eight-weeks; 3 = sixteen-weeks)
  - CD4: Number of CD4 cells per mm^3

- **The medical scientist want to understand the effects of treatment**
  - **They ask you to analyze the data**
  - **Report back to them so that they can write one or more research papers**

Repeated CD4 Counts

A randomized controlled trial

Potential research questions:

- Is there a treatment effect?
  - Is there a difference between treatments?
  - Is there an effect over time and is this different for treatments?
- Does age affect the outcome?
  - Is an age effect mediated by treatment?
  - Is an age effect mediated by time?
- Does sex effect the outcome?
  - Is a sex effect mediated by treatment?
  - Is a sex effect mediated by time?
  - Is an age effect mediated by age?

Repeated CD4 Counts

Randomized controlled trial

Assignment:

- Develop statistical analysis plan (SAP)
  - Think carefully about the different treatment effects
  - Think carefully about the dependence structure in the data and how to capture it
- Analyze the data
  - Execute the SAP
  - Evaluate the analysis assumptions
- Report the results
  - Include also descriptives [including randomization performance]
- Report the conclusions

Assignment expectations:

- Create an analysis group: 4-6 students
  - You are responsible for task allocation
  - Maintaining the group size at full strength
- Write a two-page document
  - Maximum number of words: 1200
  - Tables, figures and references not included
  - **Goal is to write for the medical scientists, so that they can use your work in a paper**
- Document all the SAS codes
  - Report in an appendix
  - If we run the codes, we should be obtaining all your results [reproducibility]

Repeated CD4 Counts

Randomized controlled trial

Important elements of assignment:

- Conduct a **sound** data analysis:
  - Research questions are translated into testable or estimable data analytic features or statistical model parameters
  - The data analysis addresses the data structure and correlations properly
  - You provide arguments for the choice of the analysis, and you report on the strengths and weaknesses of your analysis
  - You conduct and report alternative approaches to investigate analysis or model sensitivity
- There is no “best” data analysis plan
  - Different ways of analyzing the data exists
  - Importance is that analysis fits with answering the research questions and that it can be defended data scientifically
- Built a logic to the analyses with goals
  - Descriptive analyses [who are in the study]
  - Primary analyses [main hypotheses]
  - Secondary analyses [subgroup analysis]
  - Goodness-of-fit [verifying your choices]
- Make sure you report relevant details
  - Models, assumptions, estimation, etc.
  - Understandable results and graphics

Repeated CD4 Counts

Randomized controlled trial

Structure of the report:

- Statistical analysis plan
  - Small intro to understand the experiment and research questions
  - Description of statistical analyses/models: making sure you are precise, complete and relate them to the research questions
  - Interpretation of statistical parameters and hypotheses and how they relate to the research questions
  - Investigation of the (model) assumptions to help you formulate the strength and weaknesses of the analyses: possibly sensitivity analysis
- Results
  - Report on all the analyses (following the logic of the analysis plan)
  - Do not repeat the description of the analysis plan (just refer)
  - Explanation of how to read or interpret the results
  - Do not report conclusions
- Conclusions & Discussion
  - Summarize the general conclusions
  - Describe strength and weaknesses and how they may affect the conclusions
  - Be honest in what you have found (do not over state)