BIOST 509: In-Class Exercise 7

Instructor: Amy Willis, Biostatistics, UW

Due date: 6:30pm on November 15, 2019 via Canvas

Instructions

Submit your answers to the below questions in a R Script (.R), Word (.doc or .docx) or pdf file to Canvas. Provide the code that you used to get the results, the output of the code as comments, and your answers to the questions as comments.

The style of this week's homework is choose-your-own-adventure!

- If you are looking for more challenge, there is one question with no subparts. You will need to figure out the necessary steps by yourself.
- If you are looking for less challenge, the question is broken down into subparts that guide you through the steps.

Optional but encouraged: Use R Markdown to create the pdf file with your results.

Description of the dataset

All questions for this homework relate to the data is "seizure.txt", which is available from Canvas in `Pages/Module 7 materials` or `Files/datasets`.

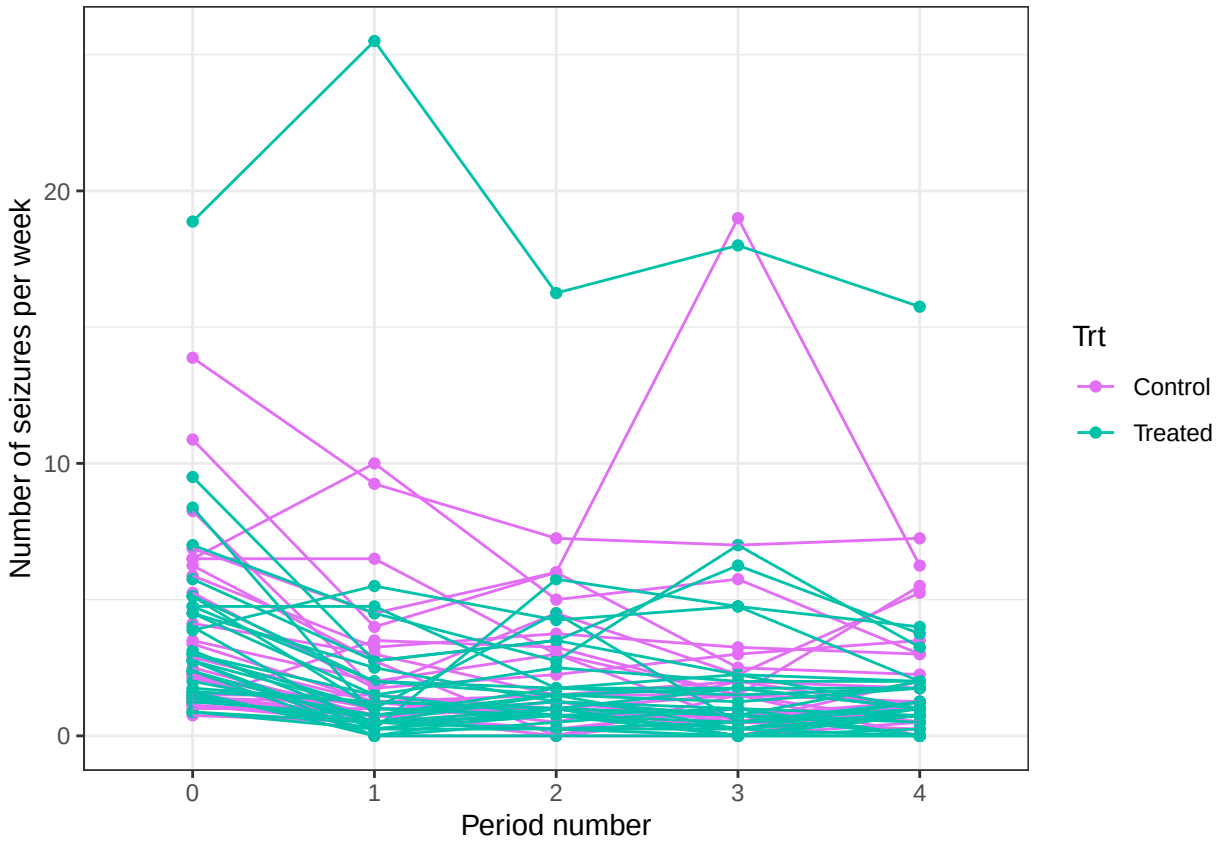
The data are from a randomized clinical trial with 59 individuals on the effectiveness of an anti-seizure medication. Prior to randomization, the 59 individuals were observed for 8 weeks and the number of seizures each patient had during this time was recorded. After the initial 8-week observation period, patients were then randomized and assigned to either the active treatment arm (i.e., these individuals received the anti-seizure medication) or the placebo arm of the trial. Patients were followed for 16 weeks, and the number of seizures within each 4-week period after randomization was recorded.

Variables names and description:

- `ID`: patient id number
- `Y1`: Number of seizures in first 4 week period after randomization
- `Y2`: Number of seizures in second 4 week period after randomization
- `Y3`: Number of seizures in third 4 week period after randomization
- `Y4`: Number of seizures in fourth 4 week period after randomization
- `Trt`: Indicator of treatment (1=active treatment, 0=placebo)
- `BL`: Number of seizures in baseline 8-week period prior to randomization
- `Age`: Patient age at randomization (years)

Questions: unguided version

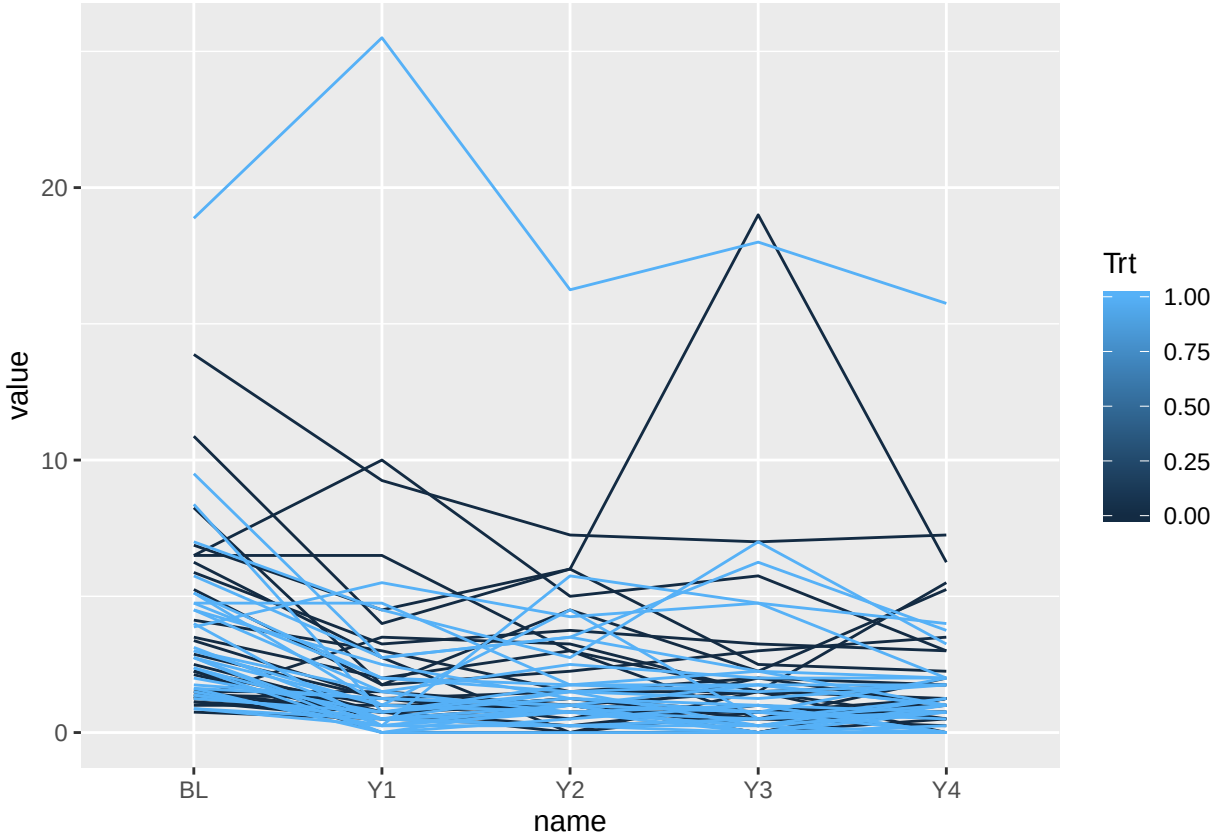
Make a line plot showing the number of seizures per week over time for each patient. Each line represents the trajectory of a patient, and the line should be coloured by treatment group. Your plot might look something like the following:


Number of seizures per week
20
10
0
0
1
2
3
4
Period number
Trt
Control
Treated

Questions: guided version

1. Order the rows in decreasing order by the number of seizures at baseline. Which three patients had the highest number of seizures in the baseline 8-week period?
2. Convert the baseline and post-randomisation number of seizures into "seizures per week".
3. The seizures dataset and the dataset that you created in Question 2 are in wide format: each row is a patient and we have the time points in each column. Restructure the seizures-per-week data so that each row contains a single observation at one time point.
4. *(Optional)* Does your response to Question 2 scale easily? For example, if there were 500 4-week periods of follow-up, would you have to modify each column individually? If your solution does not scale, provide an alternative approach that does scale to, e.g., 500 columns.
5. *(Optional)* **Reshaping your dataset is often necessary for constructing plots in `R`.** Make a line plot showing the number of seizures per week over time for each patient. Each line represents the trajectory of a patient, and the line should be coloured by treatment group.

Your plot could be as simple as the following:



It could also be more professional, like the one on Page 2.