BIOST 509: Homework 5

Instructor: Amy Willis, Biostatistics, UW

Due date: 1:00pm on November 8, 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. Please also submit your plots as **pdfs** with your submission to Canvas if you use to submit a ".R" file.*

OPTIONAL: I encourage you to use R Markdown to create the pdf file with your results. See the In-Class Exercise for a brief introduction. The R Markdown file (".Rmd") used to produce the Homework 2 Solutions is available on Canvas under `Pages/Module 5 materials` to guide you.

Description of the dataset

All questions for this homework relate to the data is "seizure.txt", which is available from Canvas in `Pages/Module 5 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

1. How many individuals are in each of the two treatment groups? What is the mean and standard deviation of the baseline seizure count for each of the two treatment groups?
2. Create a variable which represents the total number of seizures in the 16 weeks after randomization. Provide the mean and standard deviation of total seizure count for the entire sample, as well as for each treatment group separately.
3. Create a boxplot of the total number of seizures in the 16 weeks after randomization for each treatment group separately.
4. Conduct a t-test for differences in the mean total seizure count in the 16 weeks after randomization for the two treatment groups, assuming equal variances for the two treatment groups.

5. Make a `ggplot` of total seizure count in the 16 weeks after randomization versus baseline seizure count, where the color (or plotting characters) of the points for the active-treatment group are different from the placebo group.
6. Perform a linear regression analysis with total seizure count in the 16 weeks after randomization as the response and treatment group and baseline seizure count (variable `BL`) and as the predictors. Interpret the coefficients.
7. Is there significant evidence of an association between total seizure count after randomization and baseline seizure count? Provide evidence to support your conclusion.
8. (Optional) Which analysis do you prefer and why?