

BIOST 509: Homework 6

Instructor: Amy Willis, Biostatistics, UW

Due date: 1:00pm 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.

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 **Files/datasets**. You analysed this dataset in Homework 5. Recall that 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. Begin by creating a variable which represents the total number of seizures in the 16 weeks after randomization. (You did this in Homework 5.) Then, create an indicator variable which represents whether or not a patient had more than 40 seizures in the 16 weeks following treatment. Let the indicator variable equal 1 if the patient had more than 40 seizures.
2. Perform a logistic regression analysis with your indicator variable for 41+ seizures as the response, using treatment and number of seizures during baseline period as covariates in your model.
3. According to the fitted model, does treatment increase or decrease the log-odds of having more than 40 seizures? Explain your answer.

4. Give a 95% confidence interval for the odds ratio of having more than 40 seizures when treated compared to baseline.
5. Perform a logistic regression analysis with your indicator variable for 41+ seizures as the response. Fit a model using treatment and number of seizures during baseline period as covariates in your model, and allow for an interaction between treatment and number of seizures during baseline.
6. (*Optional but encouraged, especially for students who have previous exposure to logistic regression*) Interpret the coefficients on treatment and number of seizures during baseline period in your model from Question 2.