Stat 445: Midterm Project #2
Multiple Sclerosis Clinical Trial

Due: Friday, April 8 *in class.* However, papers handed in by Monday, April 4 will be marked by Monday, April 11.

Background: The data for this project comes from a multi-centre, randomized, double-blind, placebo-controlled clinical trial of a new drug in patients with relapsing-remitting multiple sclerosis (MS). Several covariates (e.g., age, sex) were recorded at "Baseline" (also called "Visit 0"), the time that each patient joined the trial. The patient was then randomized to one of three treatment groups (placebo, low dose, and high dose) and treated for a period of more than two years. Response data was collected throughout this period during regularly scheduled visits every 6 weeks.

One of the centres participating in this clinical trial was the UBC MS Clinic, and the data to be analyzed in this project are from the 50 UBC patients. The data include the baseline covariates and responses collected for 17 visits occurring at 6-week intervals. The primary objective of the clinical trial was to assess the efficacy (i.e., effectiveness in treating the disease) of this new drug. Your task in this project is to assess its efficacy based on the data from these 50 patients.

MS Clinical Responses: For relapsing-remitting MS patients, there are two standard clinical responses, or measures of the status of their disease:

- *Relapses:* This is the number of MS attacks the patient suffers in a given period of time. These attacks generally last for a period of several weeks, so a patient would typically have at most one relapse in a 6-week period.
- *EDSS Score:* The current clinical standard for describing neurological impairment in MS is the patient's score on Kurtzke's Extended Disability Status Scale (EDSS), an ordinal scale ranging from 0 to 10 in steps of 0.5 points and based on a detailed neurological examination. A score of 0 indicates normal function while a score of 10 indicates death due to MS.[1]

Recently in MS clinical trials, it has become common to also carry out magnetic resonance imaging (MRI) scans of patients' brains. From these images, it is possible to identify abnormal regions, called "lesions," which are thought to be related to MS. Each successive MRI scan on a given patient is compared to the previous MRI scan on the same patient to identify lesions that have changed in the intervening period; these are referred to as *active lesions*. The total area (in mm^2) of all the lesions appearing on the MRI scan is also determined; this is referred to as the *burden of disease*. These UBC patients had MRI scans done at Baseline and every 6 weeks thereafter.

Dataset Description: The complete dataset is divided up into four files available at:

`http://unixlab.stat.ubc.ca/~stat445/project/`

One file contains covariates collected at Baseline, one row per patient. The other three files contain data on three responses collected every six weeks at Visits 1–17. Each of these three files is organized with the data for each patient in a single row: the first column contains the patient's ID and the next 17 columns contain the responses collected in the 17 scheduled visits. NAs in the files correspond to patients missing a visit (usually because the patient dropped out of the trial).

[1] See `http://www.ucihealth.com/ms/kurtzkeedss.htm` for the meanings of intermediate scores.

The data files are:

- `baseline.dat`: covariates at Baseline:

`age`	age of the patient at Baseline (in years)
`edss`	EDSS score at Baseline
`duration`	duration of disease at Baseline (in years)
`dose`	treatment group (`0` = placebo, `1` = low dose, `2` = high dose)
`origin`	patient's place of residence (`0` = Washington State, `1` = B.C.)
`sex`	(`0` = female, `1` = male)
`bod`	burden of disease on the Baseline MRI scan (in mm^2)

- `relapse.dat`: whether the patient experienced a relapse in the 6-week period immediately prior to each visit (`0` = no, `1` = yes);
- `lesions.dat`: the number of *active* lesions identified on the MRI scan for each visit;
- `bod.dat`: the "burden of disease," the total area (in mm^2) of all lesions identified on the MRI scan for each visit.

Project: The objective of this clinical trial was to investigate whether this new drug in low and high doses reduces the rate of relapses, EDSS scores, the number of active lesions, or the burden of disease, as compared to the placebo treatment. Your task in this project is to address this question using the data on these 50 UBC MS patients for *one* of the following responses:

- burden of disease;
- relapses;
- active lesions.

(Pick whichever response interests you.)

Your report should be approximately 6 pages in length (maximum 8 pages) including any essential figures and tables, and these should be fully integrated into the text of the report. Your report should start with an "Executive Summary" of no more than one page (and included in the above page limit) providing a brief overview of your report including any key statistical findings and your interpretation of them.

Additional Guidelines: As for the first project, this project is intended to:

- to provide experience in formulating problems statistically;
- to provide experience in statistical data analysis; and
- to provide experience in writing reports.

Therefore, your emphasis in these projects should be on statistical problem solving and on writing a clear, complete, and concise report.

You should write your report so that an individual without much statistical training, such as your future boss or a future client for whom you are acting as a statistical consultant, can understand your analysis and your conclusions. In particular, your report should make very clear how you have formulated the problem as a statistical problem and how the results of your analysis relate to making conclusions that would be relevant to someone interested in the original problem (and *not* someone interested in, say, linear regression theory or R programming techniques).

You will also want to discuss any reservations you have about the original data, the appropriateness of the analysis you carried out, and your conclusions.