Homework 5

Noah Simon

Feb 9, 2017

Please turn this in by Thursday, 2/9/17 by 5:00 pm.

Please submit on Canvas, in a compiled R-markdown file (to pdf or html).

All code in this assignment should be cleanly written and well commented, with appropriate use of functions/arguments. Imagine you need to give this code to someone else and they need to understand it (which you may need to do!)

Rerandomization (and programming... and confounding)

Suppose we are comparing a new treatment to standard-of-care. We run a randomized clinical trial, and would like to assess if treatment is more effective than control (using a binary outcome — response vs non-response). Suppose we are interested in the average treatment benefit (the difference in the proportion of patients who would respond on treatment vs standard-of-care, if treatment vs standard-of-care were given to all patients in the population with the disease). However, this new treatment is potentially quite hazardous and, while we want to evaluate the treatment on all patients, we are more inclined to give the new treatment to those with more serious disease.

To evaluate disease prognosis, when patients enter our study they are given a series of tests (clinical and bio-molecular). Based on these tests, each patient is given a disease prognosis score (S) on a continuous scale from (very poor prognosis) 0 to (very good prognosis) 1. These scores are used in determining which arm to randomize patients to. Patients with poorer prognoses are more likely to be randomized to the new treatment (which is more aggressive but also more dangerous).

Assume that the set of people enrolled in the trial represents a simple random sample from the population with the disease (why might this be a bad assumption?)

1) Suppose the following randomization scheme is used (where p_new is the randomization prob for the new treatment arm):

   For patients with $S \leq 0.5$, $p_{new} \leftarrow 0.75$.

   For patients with $S > 0.5$, $p_{new} \leftarrow 0.25$.

   Write a function that takes in a matrix with one row for each patient; where the columns are named `score` (indicating S-value), `outcome` (indicating response/non-response), and `tx` (indicating treatment arm); and outputs an estimate of treatment effect. (You may use the code for Q2, with an appropriate choice of `calc.prob` to solve Q1).

2) Rather than writing out the randomization rule (as I have done above), suppose instead I give you a (potentially complicated) function called `calc.prob` which takes in a `score` (S-value) and outputs a probability p_{new} (which is a function of that score) of randomizing to the new treatment.

   Please write a function that takes in two arguments: the matrix as in Q1, but now also a function `calc.prob` (indicating how the randomization probability is determined from S); and outputs an estimate of treatment effect.

3) Suppose we now want to evaluate whether this estimated difference was actually just due to chance (and in fact new treatment is actually no more effective than control). Please write a function that uses re-randomization to assess this.

4) Consider the data in **Example_Trial.txt**. Here patients were randomized with $p_{new} \leftarrow 1 - S$. Using your code from Q2 and Q3 please evaluate the average treatment effect; and assess how consistent your estimate is with a hypothesis of new treatment being no more effective than control.

5) Suppose we accidentally assumed that **Example_Trial.txt** used simple randomization (ignoring S); estimate the average treatment effect here, and evaluate the degree of evidence we have for saying new treatment is more effective than control (do we even need to run a re-randomization test here?)

6) Suppose we knew that S was used in determining the randomization ratio, but did not know exactly how. What might we do to estimate the average treatment effect here? How might we evaluate the degree of evidence we have for saying that new treatment is more effective than control? Please implement your solution and compare it to your findings from Q4.