Homework 2

BIOST 544

Autumn 2024

Please submit your homework 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)

1) Suppose we are comparing a new treatment to standard-of-care. We run a randomized clinical trial (with basic simple randomization), and would like to assess if treatment is more effective than control (using a binary outcome — response vs non-response). Suppose efficacy is assessed by the difference in response proportion between new-treatment and standard-of-care.

   Write a function that takes in data from a clinical trial run as above (a matrix with two columns: `tx` for treatment assignment, and `outcome`, with `outcome=1` indicating response), and runs a permutation/re-randomization test to evaluate if the data are consistent with the hypothesis that standard-of-care is at least as effective as new-treatment.

2) Consider the following *adaptive randomization* scenario:

   Suppose we have pre-clinical evidence that a new treatment is much *more effective than standard-of-care. Given this, it may not be ethical to randomize a ton of patients to standard-of-care. However, preclinical evidence is not always to be believed. We decide to run a trial in which patients are enrolled one-at-a-time (sequentially). Each time a patient is enrolled, they are randomized to new treatment/control and then outcome is observed (all before the next patient is enrolled). Furthermore, we change our randomization probabilities after each outcome is observed, using the following scheme*

$$p_{new} \leftarrow \frac{1 + 3\,(success_{new} + failures_{old})}{2 + 3 \cdot totalPatients}$$

   where p_{new} is our randomization probability for the new-treatment arm; with $success_{new}$ the number of successes on the new-treatment arm (up to that point in the trial), $failures_{old}$ the number of failures on the control-arm, and $totalPatients$ the total number of patients enrolled (up to that point in the trial). As patients are enrolled, randomized, and outcomes are recorded, p_{new} will change. This adaptation strategy was proposed and discussed (in more generality) in Wei and Durham (1978). It is known as a **randomized play the winner** *rule. At the end of the trial, we assess effectiveness of treatment by looking at*

$$\hat{\pi}_{new} - \hat{\pi}_{old}$$

   the difference between proportion of successes on the new treatment, and successes on the control, calculated on all the patients from the trial.

2a) Write a function that will take in data from a trial run as above (ie. a matrix with one column `tx` of treatment assignments, one column `outcome` of binary outcomes, and one column `order` of enrollment order) and run a rerandomization test. The test should assess if the observed difference in response proportions is consistent with the hypothesis that standard-of-care is at least as effective as the new-treatment.

2b) Read in the data "HW2-adaptive-trial.txt" from the website; and, using your function from (2a), evaluate if the data are consistent with the hypothesis that standard-of-care is at least as effective as the new treatment.

3) Now suppose a collaborator provided the data in "HW2-adaptive-trial.txt", but forgot to say that they were generated using an adaptive trial. In this case we would likely accidentally analyze our data using the function written in (1). How does the sampling distribution of $\hat{\pi}_{new} - \hat{\pi}_{old}$ generated from the simple permutations in (1) compare to the sampling distribution generated using our adaptive rerandomization from the function in (2a)?

References

Wei, LJ, and S Durham. 1978. "The Randomized Play-the-Winner Rule in Medical Trials." *Journal of the American Statistical Association* 73 (364): 840–43.