

Assignment 2: Testing for heterogeneity // ABS (BI 610 Fall 2023)

Assignment: Your task is to use Rmarkdown to write a short report, readable by a technically literate person. The code you used should not be visible in the final report (unless you have a good reason to show it).

Please submit **both the .Rmd file and the resulting .html or .pdf (your choice) file.** to Canvas, by the end of day (midnight) on Thursday Oct. 26th. For tips on how to write Rmarkdown reports, and an example, please see this page. You can work with other members of class, but I expect each of you to construct and run all of the code yourself.

The Problem

You are working with a group studying immune response to different types of seasonal flu. From a large number of blood samples from people with the flu but otherwise healthy, the group has determined for each person (a) the strain of flu, (b) the strength of the immune response (in units of pg/mL of interleukin-1 β), and (c) the person's genotype at an immune-regulating gene.

The group is interested in a fine-scale classification of flu, so they have divided the flu samples up into a fairly large number of distinct strains. The main question they are interested in is whether the mean strength of immune response (measured using interleukin-1 β concentration) differs by flu strains; next they are interested in which strains (if any) tend to induce stronger, or weaker, responses. Of course, they would like to know how any differences compare to natural variation in immune response. In previous studies, mean immune response differed by genotype of the immune-regulating gene, so the analysis should take this into account. You should (a) look at the data, then answer these questions with a permutation test. More detailed instructions are below, but your report should explain, in your own words, how each analysis works what the conclusions are, and what to conclude after looking at both sets of results.

The data file is `flu.tsv`, which you can download from Canvas. Note: this is fake data; for real information on flu strains visit nextflu.org

(a) Make an *interaction plot* for this question: the x-axis should show the various flu strains; the y-axis should show immune response, and there should be one line for each genotype, connecting the mean response to each flu strain for people of that genotype. The plot should show not just the means, but also the range of variation in the data.

(b) Use as the test statistic for your permutation test the *standard deviation of strain means, averaged across genotypes*. In other words, compute (separately for each genotype) the mean response for each strain, then compute the standard deviation of those mean responses. Then average the resulting standard deviation values you get across genotypes. Since we want to look at variation across strains *while controlling for genotype*, you should shuffle immune responses *separately* for each genotype. Note that this standard deviation quantifies variation in mean immune response among strains (accounting for genotype); the permutation test gives a measure of significance.

The purposes of this homework are to practice plotting interactions; doing randomization tests, and communicating differences in variation in plain language.