

fdr.r: Explanation of code

Goals of code:

- False discovery rate, starting with p-values.
- Subsetting a data frame, determining its size

**False discovery rate adjustment:** `p.adjust()`

The `p.adjust()` function does various adjustments to p-values. It starts with a vector of p-values and returns a vector of adjusted p-values. These can be stored, stored in a data frame, plotted, or printed just like any other column of information.

My code stores the adjusted p-values in the original data frame of tests.

**Subsetting a data frame:** `subset(tests, fdr < 0.2)`

There are (at least) two ways to extract a subset of rows that satisfy a specified condition. This is useful here when you only want to print (or save) rows with small adjusted p-values.

The simplest way to subset the rows of a data frame is to use the `subset()` function. The `subset()` function returns the subset of rows that meet the specified condition. The first argument is the name of the data frame to work with; the second argument is the condition that defines the desired subset. Variables in this condition are “looked up” in the data frame. So here, `fdr<0.2` looks at `tests$fdr`, i.e., the `fdr` variable in the `tests` data frame.

R provides many logical operators: `<` and `>` should be obvious. Less than or equal to is `<=`, and `>=` is analogous. Equals is `==`; that syntax was chosen by the R developers because it was not the same as `=` for passing an argument to a function. If R fusses with an obscure error message, check that you didn’t accidentally use only one `=` when you needed two. Not is `!()`, where the thing you are complementing is inside the `()`. So, `!(fdr > 0.2)` is the same as `fdr <= 0.2`.

The result of `subset()` is saved as a new data frame. If you don’t save it, it is printed.

There are other ways to subset either rows or columns of a data frame or to subset a vector. I will introduce those when / if we need them.

**Determining the size of a data frame:** `dim(signif)`

`dim()` returns the number of rows and number of columns in a data frame. Useful here because the number of rows in the subset is the number of tests with `fdr < 0.2`.