Steven Shuken
03/21/22

Instruction Manual, Parameter Explanations, and Plot/Table Explanations: Simulator of P-Value Multiple Hypothesis Correction (SIMPLYCORRECT)

See Also:
[1]Shuken, S. R.; McNerney, M. W. "Visualizing the Costs and Benefits of Correcting P-Values for Multiple Hypothesis Testing in Omics Data." *BioRxiv* **2021**, DOI: https://doi.org/10.1101/2021.09.09.459558

Instructions

1. Visit https://shuken.shinyapps.io/SIMPLYCORRECT.
2. Read the Welcome tab. It is also recommended to read Shuken and McNerney, 2021 before using SIMPLYCORRECT.[1]
3. Click on tab 1, 2, or 3 at the top to simulate a quantitative omics experiment. The model title and parameters will be visible immediately whereas the plots and/or tables will take a few moments to appear. The parameters have generally been chosen to model after a real-life proteomics dataset[1,2]
4. For information about parameters, result plots, and result tables, see the following sections ("Parameter Explanations" and "How to Interpret Plots and Tables").
5. To set up an additional simulation, alter the choices in the parameter panel on the left. The plots/tables do not respond dynamically or instantaneously when you modify the parameters. When the parameters are suitable, click the "Go!" button to generate new result plots and result tables.

Parameter Explanations

I. Model #1: No Analytes Change

In this model, two distributions have been sampled that are identical. An example of this would be a proteomics experiment in which a drug was applied to cells and the drug has no true effect.

1. Show result tables: Toggles whether numerical tables are displayed beneath the plots.
2. Correction methods: All that are checked will be performed when the "Go!" button is clicked.
    a. Benjamini-Hochberg (BH): This is the standard in the field for some quantitative omics analyses. The procedure is described in layperson's terms in the reference above[1] and in more technical detail by Benjamini and Hochberg.[3] If the P-values are accurate and the analyte values are mutually independent or have positive regression dependence, the false discovery proportion (FDP) is guaranteed to be equal to or less than the corrected P-value threshold.
    b. Bonferroni: This lowers the P-value threshold by a factor of *m* where *m* is the number of analytes. Assuming that the P-values are accurate, this correction reduces the probability of at least one false discovery in the whole dataset (i.e., the family-wise error rate, FWER) to the original P-value threshold.
    c. Permutation FDR: An increasingly popular method, especially in quantitative proteomics, this method shuffles samples and artificially relabels them "test" and "control" and repeats the statistical tests with each shuffled sample configuration. The average number of statistically significant results across these random artificial experiments is then used as an estimate of the number of false discoveries in the original experiment. The P-value threshold is then corrected so that the estimated false discovery proportion (FDP) is equal to the original significance threshold. For more details, see Ref. 1.

3. Number of analytes: The number *m* of analytes, e.g., proteins, genes, transcripts, peptides, or metabolites quantified in the simulated experiment. Reasonable choices include: for mouse cerebrospinal fluid proteomics, 200–500 proteins; for cellular proteomics without offline fractionation, 3,000 proteins; for cellular proteomics with offline fractionation, 10,000 proteins; for transcriptomics, 20,000 genes.
4. Number of samples in control group: Sample size in the first of the two groups to be compared ($n_{control}$).
5. Number of samples in test group: Sample size in the second of the two groups to be compared (n_{test}).
6. Significance threshold: The original P-value threshold. For any analytes with corrected P-values lower than this number, the null hypothesis will be rejected (it will be accepted that this analyte is differential between test and control). After BH correction, this number is an upper bound for the FDR; after Bonferroni correction, this number becomes the FWER; after permutation FDR, this number is an estimate of the FDR.
7. Experimental variability within analyte (C.V.): This is the standard deviation of the distribution from which each measurement is sampled, divided by the analyte quantity; that is, the standard deviation of each analyte measurement as a fraction of analyte quantity, which is called the coefficient of variation (C.V.).
8. Show advanced settings: Toggles whether advanced settings are displayed.
    a. Number of permutations (only visible if Permutation FDR checkbox is checked): The number of random permutations of samples that will be used to estimate the FDP. See section I.2.c.
    b. Average quantity across analytes: The average analyte quantity. In the very first step of the simulation, *m* numbers are sampled from a normal distribution. This parameter is the mean μ of that normal distribution. Each number sampled, μ_i, is the underlying quantity of the i^{th} analyte. Later, when the analyte measurements are simulated, each measurement will be sampled from a normal distribution with a mean of μ_i.
        i. A reasonable choice for mass spectrometry-based proteomics is 22, which is a typical log(intensity) value. Log(intensity) values are used because they, unlike raw intensity values, are normally distributed.
        ii. Because the statistical tests in this model (Model #1) are comparative and because they assume normally distributed quantities, this parameter actually has no effect on any of the results reported for this model. However, it is included for potential future use by researchers who wish to modify and/or extend the SIMPLYCORRECT code.
    c. Standard deviation across analytes: The standard deviation of analyte quantities. This is the standard deviation σ of the normal distribution from which *m* underlying analyte quantities are sampled in the very first step of the simulation.
    d. Set seed: The seed for random number generation. If all parameters go unchanged including this one, then repeated SIMPLYCORRECT simulations will have identical results. To see additional random results with the same parameters, change this parameter only.

II. Model #2: All Analytes Change

In this simple model, two distributions have been sampled that have different means. Although this is not an accurate model of many realistic omics experiments, it can be used to illustrate the effects of p-value correction in highly differential situations. For example, if a drug is known to strongly affect a particular family of genes, a transcriptomic measurement of only those genes in cells treated with and without that drug might resemble this model. For more details, see Ref. 1.

1. Show result tables: Toggles whether numerical tables are displayed beneath the plots.

2. Correction methods: All that are checked will be performed when the "Go!" button is clicked.
    a. Benjamini-Hochberg (BH): See section I.2.a.
    b. Bonferroni: See section I.2.b.
    c. Permutation FDR: See section I.2.c.
3. Effect: For all analytes, this is the difference between the mean $\mu_{control,i}$ of the underlying true quantity of the analyte in the control group and the mean $\mu_{test,i}$ of the underlying true quantity of the analyte in the test group. Note: while $\mu_{control,i}$ and $\mu_{test,i}$ may vary from analyte to analyte (hence the index i of the i^{th} analyte), $Effect = \mu_{test,i} - \mu_{control,i}$ is the same for all analytes in this model.
4. Number of analytes: The number m of analytes. See section I.3.
5. Number of samples in control group: Sample size in the first of the two groups to be compared ($n_{control}$).
6. Number of samples in test group: Sample size in the second of the two groups to be compared (n_{test}).
7. Significance threshold: The original P-value threshold. See section I.6.
8. Experimental variability within analyte (C.V.): This is the coefficient of variation (C.V.). See section I.7.
9. Show advanced settings: Toggles whether advanced settings are displayed.
    a. Number of permutations (only visible if Permutation FDR checkbox is checked): The number of random permutations of samples that will be used to estimate the FDP. See section I.2.c.
    b. Average quantity across analytes: The average analyte quantity. See section I.8.b.
    c. Standard deviation across analytes: The standard deviation of analyte quantities. See section I.8.c.
    d. Set seed: The seed for random number generation. If all parameters go unchanged including this one, then repeated SIMPLYCORRECT simulations will have identical results. To see additional random results with the same parameters, change this parameter only.

III. Model #3: Realistic Model

In this model, each analyte has its own control and test distributions from which control and test observations are sampled. This is a much more realistic scenario than Model #2 and allows for a mixture of differential and non-differential analytes in the same experiment. For more details, see Ref. 1.

1. Show result tables: Toggles whether numerical tables are displayed beneath the plots.
2. Correction methods: All that are checked will be performed when the "Go!" button is clicked.
    a. Benjamini-Hochberg (BH): See section I.2.a.
    b. Bonferroni: See section I.2.b.
    c. Permutation FDR: See section I.2.c.
3. Typical effect: This is the typical magnitude of the underlying difference between the test and control group. This input must be positive. Some analytes will have an increased mean in the test group ($Effect_i = \mu_{test,i} - \mu_{control,i}$ is positive); some will have a decreased mean in the test group ($Effect_i = \mu_{test,i} - \mu_{control,i}$ is negative). But generally the magnitude of the underlying change will be similar to Typical Effect. This is achieved by sampling m times a normal distribution with a mean of 0 and a standard deviation equal to the Typical Effect and setting an analyte's $Effect_i$ to this sampled value each time.
    a. Note: If $Effect_i$ is lesser in magnitude than the i^{th} Biological Variability (see below), then $Effect_i$ is reassigned to 0 before measurements are simulated (see Ref. 1). Because of

this, the average $Effect_i$ magnitude across all analytes is decreased and the average $Effect_i$ magnitude across differential analytes is increased.

4. Typical biological variability (C.V.): The typical Biological Variability as a fraction of true quantity (C.V. = Coefficient of Variation = standard deviation / mean). Each analyte is assigned a Biological Variability as a C.V. before measurements are simulated. These biological variability C.V.s are determined by sampling *m* times a scaled gamma distribution whose mean is equal to Typical Biological Variability (C.V.).
5. Experimental variability (C.V.): The Experimental Variability as a fraction of true quantity (C.V.). All analytes have the same Experimental Variability C.V. When measurements are simulated, a normal distribution is sampled. The standard deviation of this normal distribution is equal to:
   $(Biological\ Variability\ C.V.)_i \times (True\ Quantity)_i +$
   $(Technical\ Variability\ C.V.)_i \times (True\ Quantity)_i$.
6. Number of analytes: The number *m* of analytes. See section I.3.
7. Number of samples in control group: Sample size in the first of the two groups to be compared ($n_{control}$).
8. Number of samples in test group: Sample size in the second of the two groups to be compared (n_{test}).
9. Significance threshold: The original P-value threshold. See section I.6.
10. Show advanced settings: Toggles whether advanced settings are displayed.
    a. Number of permutations (only visible if Permutation FDR checkbox is checked): The number of random permutations of samples that will be used to estimate the FDP. See section I.2.c.
    b. Average quantity across analytes: The average analyte quantity. See section I.8.b.
    c. Standard deviation across analytes: The standard deviation of analyte quantities. See section I.8.c.
    d. Set seed: The seed for random number generation. If all parameters go unchanged including this one, then repeated SIMPLYCORRECT simulations will have identical results. To see additional random results with the same parameters, change this parameter only.
    e. Biological variability distribution parameters: The shape and scale parameters of the scaled gamma distribution from which Biological Variability C.V.s are sampled (see section III.4 and Rice, 2007).[4] After the unscaled gamma distribution is sampled using these shape and scale parameters, the sampled values are scaled (multiplied) so that the scaled values follow a scaled gamma distribution whose mean is specified by the user. These scaled values are the biological variabilities of the analytes.[1]

How to Interpret Plots and Tables

Plots

The x-axis of each plot is labeled “Observed Effect.” For each analyte, the Observed Effect is the difference between the average measurement of that analyte in the control group and the average measurement of that analyte in the test group. If the test group has a higher average measurement, the Observed Effect is positive; if the test group has a lower average measurement, the Observed Effect is negative.

The y-axis of each plot is labeled “-log10(P.Val).” For each analyte, its position on the y-axis is $-\log_{10}$(P-value), where the P-value is calculated from a two-sided Student’s *t*-test.

Each point on each plot represents an analyte. It is color-coded according to whether it is a false positive (red), true positive (dark blue), false negative (light blue), or true negative (salmon) as

defined below in the "Tables" section. Red and salmon points represent non-differential analytes while blue points represent differential analytes as defined below.

Tables

The rows of the output tables can be interpreted as follows.

- False Discovery Proportion (FDP): This is calculated as (# false positives) / (total positives), where a "positive" is a statistically significant hit after P-value correction, that is, an analyte with $p \leq$ Significance Threshold (usually 0.05) and therefore $-\log_{10}(p) \geq -\log_{10}$(Significance Threshold) (usually ~1.3). A false positive is defined as a positive whose underlying test and control group means, μ_{test} and $\mu_{control}$, are equal. (For Model #1 this is always true.)
    - Note: FDP differs from false discovery rate (FDR) in that the FDP is a property of a single experiment/simulation (a "study-specific" property) while the FDR is a property of an analytical procedure that can be applied to many experiments (a "long-run" property).
- False Negative Proportion (FDP): This is calculated as (# false negatives) / (total negatives), where a "negative" is a statistically non-significant hit after P-value correction, that is, an analyte with $p >$ Significance Threshold (usually 0.05) and therefore $-\log_{10}(p) < -\log_{10}$(Significance Threshold) (usually ~1.3). A false negative is defined as a negative whose underlying test and control group means, μ_{test} and $\mu_{control}$, are unequal.
- Observed Sensitivity: This is calculated as (# true positives) / (total differential analytes), where a "differential analyte" is an analyte whose underlying test and control group means, μ_{test} and $\mu_{control}$, are unequal.
    - Note: "Observed Sensitivity" differs from the general term "sensitivity" in that the Observed Sensitivity is a property of a single experiment/simulation (a "study-specific" property) while sensitivity is a property of an analytical procedure that can be applied to many experiments (a "long-run" property).
- Observed Specificity: This is calculated as (# true negatives) / (total non-differential analytes), where a "non-differential analyte" is an analyte whose underlying test and control group means, μ_{test} and $\mu_{control}$, are equal.
    - Note: "Observed Specificity" differs from the general term "specificity" in that the Observed Specificity is a property of a single experiment/simulation (a "study-specific" property) while specificity is a property of an analytical procedure that can be applied to many experiments (a "long-run" property).
- Total Hits: Total positives.
- Total False Positives: Total false positives as defined above.
- Total True Positives: Total true positives. A true positive is a positive differential analyte.
- Total False Negatives: Total false negatives as defined above.
- Total True Negatives: Total true negatives. A true negative is a negative non-differential analyte.
- Hits (% of Data): Total hits as a percent of total analytes.
- False Positives (% of Data): Total false positives as a percent of total analytes.
- True Positives (% of Data): Total true positives as a percent of total analytes.
- False Negatives (% of Data): Total false negatives as a percent of total analytes.
- True Negatives (% of Data): Total true negatives as a percent of total analytes.

Steven Shuken
03/21/22

References

[1] Shuken, S. R.; McNerney, M. W. “Visualizing the Costs and Benefits of Correcting P-Values for Multiple Hypothesis Testing in Omics Data.” *BioRxiv* **2021**, DOI: https://doi.org/10.1101/2021.09.09.459558.
[2] Shuken, S. R.; Rutledge, J.; Iram, T.; Losada, P. M.; Wilson, E. N.; Andreasson, K. I.; Leib, R. D.; Wyss-Coray, T. “Limited proteolysis-mass spectrometry reveals aging-associated changes in cerebrospinal fluid protein abundances and structures.” *Nature Aging*, Accepted.
[3] Benjamini, Y.; Hochberg, Y. “Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing.” *J. R. Statist. Soc. B* **1995**, *57*, 289–300.
[4] Rice, J. A. *Mathematical Statistics and Data Analysis*. 3rd ed. Delhi: Cengage Learning India Private Limited; **2007**.