Simfit

Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

Multiple testing is when several statistical tests are performed on the same data so it is necessary to control false results. One procedure is the Bonferroni correction where, for m tests and p values, results are considered significant at level α if $p \le \alpha/m$, rather than $p \le \alpha$ for single tests.

If at least one of the p values satisfies the Bonferroni restriction, the FDR(BH) false discovery rate technique (Benjamini and Hochberg J.R.satist.Soc. B (1995) 57,1, 289–300, and Benjamini et al Behavioural Brain Research 125 (2001) 279–284) is available to see if there other p values, not necessarily satisfying the Bonferroni restriction, that could also be regarded as possibly significant.

Example 1: FDR(BH) for a vector of p values

From the main SIMFIT menu choose [Statistics] then [Statistical calculations] and then [False discovery rates from a vector p(i)] and scrutinize the default test file `fdr_bh.tf1` provided. After selecting to calculate the false discovery rates, view the table of results for all the data arranged into order of rank which is displayed next. Here m is the number of tests and i is the rank of the sample in terms of the ordered p values.

False discovery rates for a vector of p(i) values: 1
Title: Data for BH False Discovery rate calculation
Sample size = 17
Number rejected = 0
Number analysed = 17
Significance level, α = 0.05

Rank	Sample	p p-value	$m * p/i$ p-adjusted	$\alpha * i/m$ BH-level	Result
1	12	0.000001	0.000017	0.002941	1
2	1	0.000013	0.000110	0.005882	1
3	3	0.000065	0.000368	0.008824	1
4	6	0.000630	0.002678	0.011765	1
5	5	0.000800	0.002720	0.014706	1
6	16	0.001700	0.004817	0.017647	1
7	2	0.003200	0.007771	0.020588	1
8	7	0.006500	0.013813	0.023529	1
9	11	0.014800	0.027956	0.026471	1
10	13	0.049000	0.083300	0.029412	0
11	14	0.094000	0.145273	0.032353	0
12	17	0.110000	0.155833	0.035294	0
13	9	0.150000	0.196154	0.038235	0
14	8	0.240000	0.291429	0.041176	0
15	15	0.450000	0.510000	0.044118	0
16	10	0.560000	0.595000	0.047059	0
17	4	0.870000	0.870000	0.050000	0

There are other options to view the results in sample order or to just show significant results, but the above table is the easiest to understand and follows the example given by Benjamini et al on this same data set.

In order to understand the FDR(BH) technique we shall explain the meanings of the above columns and, in particular, the interpretation of the colors and meaning of the 1's and 0's in the last column.

1. **Column 1**

   This is the rank i of the sample with respect to the p values. That is, the rows of the table are arranged so that the samples in row i are arranged in order of increasing p values.

2. **Column 2**

   This registers the actual number of the sample in the original order.

3. **Column 3**

   Here are the p values corresponding to the rank recorded in column 1 for the sample identified in column 2.

4. **Column 4**

   If this is table line for rank i then this is the p value adjusted by the rank and the sample size m. In other words, the adjusted p value is mp/i. Note that this column only depends on p, i and m, and the last adjusted p value is always the same as the uncorrected p value since $i = m$.

5. **Column 5**

   Here are listed the BH-levels, i.e., the BH threshold values $\alpha i/m$. Note that these only depend on α, i and m, and they have the following sequence. The value at row 1 is the Bonferroni corrected level for significance testing, and the value at line m is the significance level α, while between these extremes the values slowly increase as a function of the rank.

6. **Column 6**

   This column has a 1 if the sample is in the FDR(BH) set and a 0 otherwise

The systematic FDR(BH) procedure

The technique starts at row m and advances up the table until the first rank is encountered, say k, where the p value is less than or equal to the BH threshold. We then conclude that all samples from line 1 up to line k must be considered as possibly significant. So the set of possibly significant samples contains those where

$$p \leq \alpha i/m,$$

$$\text{or equivalently } mp/i \leq \alpha.$$

So now the importance of the color change will be clear and the interpretation of the table is obvious.

All samples numbered in column 2 up to level k with a 1 in column 6 are colored blue, which makes identification of the set of possibly significant samples easy to recognize.

The table can also be rearranged into sample order and can be displayed in such a way as to only identify the set of possibly significant samples. Also, for very large samples it is possible to scroll through the table to select sections or even to write the whole table to file.

Example 2: FDR(BH) for a matrix of p values

Some procedures result in matrices of p values, and this requires a more complicated approach because we have to keep track of the row and column indices. As a typical example, select the option for false discovery rate for a matrix and read in the default test file `matrix_p.tf1` which is as follows

0.00023	0.00060	0.40906	0.41318
0.00050	0.00005	0.32055	0.23282
0.00560	0.01362	0.43751	0.06327

This results from the directed correlation procedure in the multivariate statistics options using the default test files `matrix_a.tf1` for the A matrix which has dimensions 30 by 3, and `matrix_b.tf1` for the B matrix which has dimensions 30 by 4.

Proceeding with the false discovery option we obtain the following table in rank order.

False discovery rates for a matrix of p(i,j) values: 1
Title: Data from directed correlation
Number of columns = 4
Number of rows = 3
Number out of range = 0
Significance level, α = 0.05

$A(i)$	$B(j)$	p-value	p-adjusted	BH-level	Result
2	2	0.000053	0.000632	0.004167	1
1	1	0.000231	0.001387	0.008333	1
2	1	0.000500	0.002000	0.012500	1
1	2	0.000598	0.001793	0.016667	1
3	1	0.005602	0.013446	0.020833	1
3	2	0.013624	0.027247	0.025000	1
3	4	0.063269	0.108461	0.029167	0
2	4	0.232822	0.349234	0.033333	0
2	3	0.320548	0.427398	0.037500	0
1	3	0.409063	0.490875	0.041667	0
1	4	0.413176	0.450738	0.045833	0
3	3	0.437508	0.437508	0.050000	0

As before the set of possibly significant samples is easy to identify by the 1 in the last column or blue color, but columns 1 and 2 need some explanation.

In this example column 1 indicates what the row indices of p values are, because the matrix of p values had 3 rows which originated from the 3 columns of the A matrix in the directed correlation. The second column identifies column indices for the 4 columns corresponding to the 4 columns of the B matrix. This situation is valid only for this matrix, and results from the convention dictating the way that the matrix of p values was constructed.