Recent Developments in Multiple Comparison Procedures
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 47 (2004) 100–118
© Institute of Mathematical Statistics, 2004

FDR step-down and step-up procedures for the correlated case

Paul N. Somerville[1]

University of Central Florida

Abstract: Controlling the false discovery rate has been increasingly utilized in high dimensional screening studies where multiplicity is a problem. Most methods do not explicitly take the correlation between the data or the test statistics into account, with consequent loss of power. In this paper, we use least favorable configurations to obtain critical values for both step-down and step-up procedures, valid for both dependent and independent hypotheses. The concept of a "minimum critical value" (MCV) is introduced. For the step-down case with MCV = 0, our step-down procedure is the same as that of Troendle (2000). It is conjectured that, for a given MCV, there is no uniformly more powerful step-down FDR procedure. Empirical results suggest that, for maximizing power, the "optimum" MCV is a decreasing function of the number of false hypotheses. Various tables are given, with a special "condensed" table valid for numbers of hypotheses from 30 to 10,000 and $\rho = .5$ specifically designed for the case where few false hypotheses are anticipated or where a satisfactory outcome is the discovery of a few false hypotheses. Intermediate values for the latter table may be obtained by interpolation. An application to high dimensional genomic data is given.

1. Introduction

Suppose we wish to simultaneously test m hypotheses. A traditional method has been to control the familywise error (FWE) rate. The familywise error rate is defined as the probability of committing a type I error for at least one of the m hypotheses. An alternative, recently proposed for a number of situations, is to control the false discovery rate (FDR). The false discovery rate is defined to be the expected value of the proportion of rejected hypotheses which are true, with the understanding that if no hypotheses are rejected, the proportion is zero.

Benjamini and Hochberg (1995) introduced a step-up FDR (SU) valid when the hypotheses are independent. Benjamini and Liu (1999) introduced a step-down FDR (SD), valid also for independent hypotheses. Benjamini and Yekutieli (2001) proved that the approach of Benjamini and Hochberg (1995) controls the FDR if the joint distribution of the test statistics is positive regression dependent on each one. In addition, the authors introduced a slight deviation of the original procedure, which was shown to control the FDR under any correlation structure. However, these methods lack in power in that they do not take the possible correlation among the test statistics into account. Yekutieli and Benjamini (1999) applied resampling techniques to include the underlying distributional characteristics. Benjamini and Yekutieli (1999) gave a distribution-free FDR-control multiple test procedure. Benjamini and Liu (2001) presented another distribution-free FDR-control multiple test procedure. Under the assumption of normality, powerful test statistics both for

[1]Department of Statistics, University of Central Florida, 4000 Central Florida Avenue, Orlando, FL 32816-011, USA. e-mail: somervil@pegasus.cc.ucf.edu

Keywords and phrases: FDR, multiplicity, multiple comparison procedures, step-down and step-up procedures, minimum critical value, high dimensional data.

AMS 2000 subject classifications: 62F03.

the step-up and the step-down cases were derived by Troendle (2000). His method was shown to control the FDR asymptotically. A different approach to that of Benjamini and Hochberg making use of positive regression dependency was introduced by Kwong, Holland and Cheung (2002). Sarkar (2002) extended some of the above results. Horn and Dunnett (2004) conducted a study comparing the power of several FWE and FDR controlling methods. Korn, Troendle, McShane and Simon (2003) proposed two step-wise permutation based procedures to control, with specific confidence, the actual number of false discoveries, and approximately, the actual proportion of false discoveries.

In this paper we develop step-down and step-up FDR procedures which are valid for dependent or independent hypotheses. We conjecture that there are no step-down FDR procedures which, for a given MCV, are uniformly more powerful. We also introduce a "condensed" FDR step-down procedure especially designed for the case where few false hypotheses are anticipated.

2. Problem

Suppose we have m hypotheses $H_1, H_2, \ldots, H_m$ to be tested. Let the corresponding test statistics be $T_1, T_2, \ldots, T_m$. Denote the hypotheses as $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$ corresponding to the ordered test statistics $T_{(1)} \leq T_{(2)} \cdots \leq T_{(m)}$. Denote the m critical values as $d_1 \leq d_2 \leq \cdots \leq d_m$. For the step-down procedure, beginning with $i = m$, then $m-1$, etc., compare $T_{(i)}$ with d_i stopping when $T_{(i)} < d_i$, and rejecting the $m - i$ hypotheses $H_{(i+1)}, H_{(i+2)}, \ldots, H_{(m)}$. If $T_{(m)} < d_m$, reject no hypotheses. If $T_{(i)} \geq d_i$ for all m test statistics, reject all the m hypotheses.

For the step-up procedure, beginning with $i = 1$, then 2, etc., compare $T_{(i)}$ with d_i, stopping when $T_{(i)} \geq d_i$, and rejecting the $m - i + 1$ hypotheses $H_{(i)}, H_{(i+1)}$, $\ldots, H_{(m)}$. If $T_{(i)} < d_i$ for all m test statistics, reject no hypotheses.

We require that the false discovery rate (FDR), that is, the expected value of the proportion of number of rejected hypotheses which are true, to be $\leq q$. Let $Q = V/R$ where V is the number of true hypotheses which are rejected and R is the number of hypotheses which are rejected. When $R = 0$, we define $Q = 0$. $E(Q) = E(V/R)$ is the false discovery rate FDR.

In what follows, we will use n_T to be the number of hypotheses which are true (T) and n_F the number which are false (F), except in graphs where nT and nF will be used.

3. Calculation of the critical values for the step-down case

We first define A_i to be the probability that exactly i hypotheses are rejected.

$$
\begin{aligned}
A_0 &= P\big[T_{(m)} < d_m\big] \\
A_1 &= P\big[T_{(m)} \geq d_m,\ T_{(m-1)} < d_{m-1}\big] \\
&\ldots \\
A_{m-1} &= P\big[T_{(m)} \geq d_m, \ldots, T_{(2)} \geq d_2,\ T_{(1)} < d_1\big] \\
A_m &= P\big[T_{(m)} \geq d_m, \ldots, T_{(2)} \geq d_2,\ T_{(1)} \geq d_1\big].
\end{aligned}
$$

To obtain the critical values, we use m least favorable configurations (see Section 9.) of the location parameters of the test statistics. Define LFC_i as the configuration where i of the location parameters are zero and the remainder are infinite. The case where all m hypotheses are T corresponds to LFC_m.

Under LFC_m, set $A_0 = 1 - q$, and solve for d_m.

$$\begin{aligned} A_0 &= P\big[T_{(m)} < d_m\big] \\ &= P\big[\text{all } T_i < d_m\big] \\ &= 1 - q. \end{aligned}$$

To obtain d_1, use LFC_1. Then, with probability 1, $T_{(m)}, T_{(m-1)}, \ldots, T_{(2)}$ are infinite and $A_0, A_1, \ldots, A_{m-2}$ are zero. We have

$$E(Q) = A_{m-1} * (0/(m-1)) + A_m * (1/m) \leq q$$

or $P[T_{(1)} \geq d_1] \leq mq$.

To maximize power considerations, we choose the smallest value for d_1 which satisfies the equation. When we calculate critical values (or $E(Q)$) in this paper, we assume the test statistics have a joint multivariate t distribution and a common correlation coefficient ρ. For convenience, we discuss only the one-sided case.

To obtain d_i $(1 < i < m)$, given the values for $d_1, \ldots, d_{i-1}$, use LFC_i. With probability 1, $T_{(m)}, T_{(m-1)}, \ldots, T_{(i+1)}$ are infinite and $A_0, A_1, \ldots, A_{m-i-1}$ are zero. We may then write, using the basic expectation algorithm,

$$E(Q) = A_{m-i} * \big(0/(m-i)\big) + A_{m-i+1} * \big(1/(m-i+1)\big) + \cdots + A_m * i/m$$

where

$$\begin{aligned} &A_{m-i+1} = P\big[T_{(i)} \geq d_i, \ldots, T_{(m-i+1)} \geq d_{m-i+1}, T_{(m-i)} < d_{m-i}\big] \\ &\ldots \\ &A_{m-1} = P\big[T_{(i)} \geq d_i, \ldots, T_{(2)} \geq d_2, T_{(1)} < d_1\big] \\ &A_m = P\big[T_{(i)} \geq d_i, \ldots, T_{(2)} \geq d_2, T_{(1)} \geq d_1\big]. \end{aligned}$$

Since $A_{m-i+1}, \ldots, A_m$ are each decreasing functions of d_i, we choose d_i as the smallest value such that

$$E(Q) \leq q.$$

We note that larger values will satisfy the FDR requirement, but of course at the expense of "power".

On close examination, $A_0 = 1-q$, and the $m-1$ equations $E(Q) = q$ for obtaining the m critical values, are, in spite of completely different notations, equivalent to the corresponding k $(= m)$ equations (4) of Troendle (2000). Troendle's equations are based on multivariate normal assumptions and "asymptotic FDR control". Troendle uses the fact that $T_1, \ldots, T_k$ are consistent for testing $H_1, \ldots, H_k$. Thus, as the sample size for each of the test statistics increases without limit, "under any parameter configuration" the values of the test statistics corresponding to the true null hypotheses are less than values of the test statistics corresponding to the false hypotheses. No true hypothesis can then be rejected unless all of the false hypotheses are rejected, and equations (4) follow from (3).

Our equations are based on least favorable configurations of the population means of the test statistics which are assumed to have a multivariate-t distribution.

Theorem 3.1. *If $mq \geq j$, $d_1, d_2, \ldots, d_j \to -\infty$.*

Proof. Suppose the theorem is true for $j = i - 1$. Then $d_1, d_2, \ldots, d_j \to -\infty$ implies $A_{m-1}, \ldots, A_{m-i+1} \to 0$. Using LFC_i to solve for d_i, we have in addition, $A_{m-i-1}, \ldots, A_0 \to 0$ and $A_m = P[T_{(i)} \geq d_i]$.

Then $E(Q) = i * A_m/m$, and $E(Q) \leq q$ implies $P[T_{(i)} \geq d_i] \geq mq/i$. If $mq \geq i, d_j \to -\infty$.

If $j = 1$, using LFC_1, we have $T_{(m)}, \ldots, T_{(2)} \to \infty$, $A_0, \ldots, A_{m-2} \to 0$ and $A_m = P[T_{(1)} \geq d_1]$. Then $E(Q) \leq q$ implies $P[T_{(1)} \geq d_1] \leq mq$ and $d_1 \to -\infty$ if $mq > 1$, and the theorem is proved. □

When $mq \geq 1$, Troendle (2000) arbitrarily chooses $d_1 = 0$.

4. Calculation of the critical values for the step-up case

As before, denote the m critical values as $d_1 \leq d_2 \leq \cdots \leq d_m$, noting that the critical values are in general different from those for the step-down case. Define B_i to be the probability that exactly i hypotheses are rejected.

$$
\begin{aligned}
B_m &= P\big[T_{(1)} \geq d_1\big] \\
B_{m-1} &= P\big[T_{(1)} < d_1, T_{(2)} \geq d_2\big] \\
&\ldots \\
B_1 &= P\big[T_{(1)} < d_1, \ldots, T_{(m-1)} < d_{m-1}, T_{(m)} \geq d_m\big] \\
B_0 &= P\big[T_{(1)} < d_1, \ldots, T_{(m-1)} < d_{m-1}, T_{(m)} < d_m\big].
\end{aligned}
$$

We should like to follow a regime similar to that for the step-down case. That is, to obtain d_1, assume the least favorable configuration LFC_1 and find the smallest value of d_1 which solves the equation $E(Q) \leq q$. To obtain d_i, given $d_1, \ldots, d_{i-1}$, assume LFC_i, and solve for the smallest value d_i for which $E(Q) \leq q$.

Assuming LFC_1, with probability 1, $T_{(2)}, \ldots, T_{(m)}$ are infinite, and $B_0, \ldots, B_{m-2}$ equal 0. Then

$$E(Q) = B_m(1/m) \leq q$$

or

$$P[T_{(1)} \geq d_1] \leq mq.$$

To obtain d_i, given the values for $d_1, \ldots, d_{i-1}$, assume the least favorable configuration LFC_i. Then with probability 1, $T_{(i+1)}, \ldots, T_{(m)}$ are infinite, and $B_0, \ldots, B_{m-i-1}$ are 0. Then

$$E(Q) = B_{m-i+1}\big(1/(m-i+1)\big) + \cdots + B_m(i/m).$$

B_{m-i+1} is a decreasing function of d_i, and no other term depends on d_i. Choose d_i as the smallest value such that $E(Q) \leq q$.

When $mq \geq 1$, the smallest value of d_1 which is a solution of $P[T_{(1)} \geq d_1] \leq mq$ is $-\infty$. Use of that value would result in rejection of all hypotheses. When $mq < 1$, it is not difficult to find examples where, having obtained the smallest value of d_1, there are values of i for which it is not possible to obtain a sufficiently large value d_i such that $E(Q) \leq q$. For these cases we may choose the smallest value c such that if $d_1 = c$, the regime is such that $E(Q) \leq q$ under all LFC_i. That such a value may always be obtained may be seen by setting $d_1 = d_2 = \cdots d_m = \theta$. (As $\theta \to \infty$, $E(Q) \to 0$.)

As an example, for $m = 14$, $q = .05$, $\nu = \infty$, solving $P[T_i \geq d_1] = mq$ for d_1 gives $d_1 = -.5244$. However, the smallest values of d_1 for which the higher critical values can be obtained such that $E(Q) \leq .05$, are approximately $-.1$, .6 and 1.8 for values of $\rho = .1$, .5 and .9 respectively.

Table 5.1: Critical values for FDR step-down one-sided tests $m = 10$, $q = .05$, $\rho = .5$, $\nu = \infty$

MCV	1	2	3	4	5	6	7	8	9	10
0.0	0.0	0.921	1.222	1.432	1.597	1.748	1.891	2.039	2.211	2.448
0.5	0.5	0.799	1.223	1.431	1.597	1.747	1.891	2.039	2.211	2.448
1.0	1.000	1.000	1.090	1.420	1.592	1.746	1.890	2.041	2.211	2.448
1.5	1.5	1.5	1.5	1.5	1.5	1.720	1.884	2.037	2.212	2.448
1.645	1.645	1.645	1.645	1.645	1.645	1.647	1.876	2.035	2.210	2.448
2.0	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.201	2.448
2.448	2.448	2.448	2.448	2.448	2.448	2.448	2.448	2.448	2.448	2.448

Table 5.2: Critical values for FDR step-up one-sided tests $m = 10$, $q = .05$, $\rho = .5$, $\nu = \infty$

MCV	1	2	3	4	5	6	7	8	9	10
0.28	0.280	1.419	2.396	2.489	2.499	2.580	2.635	2.717	2.830	3.020
0.5	0.500	.872	1.699	1.787	1.888	2.007	2.122	2.250	2.406	2.634
1.0	1.000	1.000	1.097	1.637	1.753	1.882	2.011	2.150	2.313	2.553
1.5	1.5	1.5	1.5	1.5	1.5	1.795	1.961	2.110	2.285	2.529
1.645	1.645	1.645	1.645	1.645	1.645	1.647	1.959	2.108	2.282	2.526
2.0	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.246	2.519
2.448	2.448	2.448	2.448	2.448	2.448	2.448	2.448	2.448	2.448	2.448

5. Minimum critical values (MCV's)

For step-down FDR we obtained d_1 as the smallest value satisfying $P[T_{(1)} \geq d_1] \leq mq$. For step-up FDR, a larger value for d_1 was sometimes required. For either step-down or step-up FDR , the resulting value of d_1 may not be looked on with favor by all potential users. We doubt that many users would look with favor on an FDR procedure with the critical value $d_1 = -\infty$. Similarly, we suspect that some of those users might look with suspicion on any procedure which would allow an hypothesis to be rejected when the test statistic was less than $t_{q,\nu}$ (the critical value for rejection of a 1-sided hypothesis using level q when the standard deviation is estimated using ν degrees of freedom). In that case they might prefer to set $d_1 = c = t_{q,\nu}$.

The value we set for c, we define to be the Minimum Critical Value (MCV). Table 5.1 shows the resulting critical values for 7 different MCV's for step-down FDR when $m = 10$, $q = .05$ and $\rho = .5$, $\nu = \infty$. Table 5.2 shows the resulting critical values for 7 different MCV's for step-up FDR for the same parameter values.

6. "Optimum" critical values

For each of per pair, all pairs and any pair power, for SD and SU FDR, computations were made for $m = 14$ and 100, $q = .05$, $\rho = .5$, $\nu = \infty$, all values of n_F, various MCV's and $\Delta = 1.732$ and 3.464. (We define Δ to be the common location parameter of the n_F F test statistics. The value $\Delta = 1.732$ is equivalent to a standardized mean difference of 1 between the population means of the F treatments and the control in the comparisons with the control problem when the sample means are estimated with a sample size of 6. (See Horn and Dunnett (2004))). The graphs for $m = 100$, $\Delta = 1.732$ for the three different powers for both step-down

Table 6.1: MCVs for which per pair power is largest (Step-down FDR)

n_F	MCV
1 to 11	All equal (3 decimals)
12 to 35	2.0
36 to 60	1.645
61 to 85	1.0
86 to 95	0.0
95 to 100	$-\infty$

Table 6.2: Probability that all n_F hypotheses are rejected ($E(Q)$ on second line) FDR (Step-down) $m = 10$, $q = .05$, $\rho = .5$, $\nu = \infty$

	n_F									
MCV	10	9	8	7	6	5	4	3	2	1
0.28	**.926**	.866	.808	.759	.720	.689	.666	.654	.659	**.710**
	(.000)	(.049)	(.049)	(.050)	(.048)	(.049)	(.049)	(.050)	(.050)	(.050)
0.5	.918	**.878**	.809	.759	.720	.688	.666	.654	.660	**.710**
	(.000)	(.031)	(.049)	(.049)	(.050)	(.049)	(.049)	(.050)	(.050)	(.050)
1.	.859	.858	**.831**	.763	.722	.689	.666	.654	.660	**.710**
	(.000)	(.016)	(.034)	(.050)	(.049)	(.050)	(.050)	(.050)	(.050)	(.050)
1.5	.695	.710	.725	**.740**	**.747**	.698	.668	.655	.660	**.710**
	(.000)	(.007)	(.014)	(.024)	(.035)	(.048)	(.050)	(.050)	(.050)	(.050)
1.645	.631	.647	.666	.685	.706	**.718**	**.672**	.656	.660	**.710**
	(.000)	(.005)	(.011)	(.018)	(.026)	(.037)	(.049)	(.050)	(.050)	(.050)
2.	.461	.479	.500	.524	.553	.586	.625	**.667**	**.663**	**.710**
	(.000)	(.002)	(.005)	(.008)	(.012)	(.018)	(.024)	(.034)	(.047)	(.050)
2.448	.258	.274	.293	.316	.344	.378	.422	.482	.569	**.710**
	(.000)	(.001)	(.002)	(.003)	(.004)	(.006)	(.008)	(.012)	(.017)	(.026)

and step-up are shown in Appendix I. Graphs and tables for all cases are given in Somerville (2003). The MCV for which the calculated power was largest, was always a decreasing function of nF. This suggests that a "small" value for MCV should be used when most hypotheses are anticipated to be false and a "large" value for MCV should be used when few are expected to be false. An additional advantage of selecting an MCV related to an expected number of false hypotheses is that there may be a significant reduction in E(Q) or the "false discovery rate". If $n_F = 1$, then an FDR procedure is not needed, and Dunnett's (1955) procedure is indicated. For the step-down procedure, Table 6.1 shows the MCVs resulting in the largest per pair power when $m = 100$, $q = .05$, $\rho = .5$, $\nu = \infty$, $\Delta = 1.732$. MCVs used were $-\infty$, 0, 1, 1.645, 2.

Tables 6.2 and 6.3 give the probability that the step-down and step-up procedures will result in all n_F false hypotheses being rejected for one-sided tests when $m = 10$, $q = .05$, $\rho = .5$, $\Delta = 5.196$, and $\nu = \infty$. The corresponding values of $E(Q)$ are given in parentheses. The value in **bold** in each n_F column is the maximum probability over all MCV values. (Values in the tables are given to three decimal places. There may be error in the 3rd place).

Table 6.3: Probability that all n_F hypotheses are rejected ($E(Q)$ on second line) FDR (Step-up) $m = 10$, $q = .05$, $\rho = .5$, $\nu = \infty$

	n_F									
MCV	10	9	8	7	6	5	4	3	2	1
0.28	**.975**	.771	.401	.341	.343	.335	.349	.370	.408	.496
	(.000)	(.039)	(.050)	(.051)	(.050)	(.050)	(.050)	(.050)	(.050)	(.050)
0.5	.956	**.906**	.670	.632	.609	.586	.575	.573	.588	.644
	(.000)	(.032)	(.050)	(.050)	(.050)	(.051)	(.050)	(.051)	(.048)	(.052)
1.	.867	.875	**.861**	.704	.669	.642	.624	.617	.627	.674
	(.000)	(.016)	(.034)	(.051)	(.048)	(.050)	(.050)	(.050)	(.049)	(.052)
1.5	.695	.710	.727	**.745**	**.765**	.682	.647	.634	.639	.682
	(.000)	(.007)	(.014)	(.024)	(.036)	(.049)	(.049)	(.050)	(.048)	(.050)
1.645	.631	.648	.666	.687	.710	**.736**	**.650**	.635	.640	.683
	(.000)	(.005)	(.011)	(.018)	(.027)	(.037)	(.049)	(.049)	(.052)	(.048)
2.	.461	.479	.500	.524	.553	.586	.627	**.678**	**.655**	.686
	(.000)	(.002)	(.005)	(.008)	(.012)	(.018)	(.024)	(.032)	(.046)	(.048)
2.448	.258	.274	.293	.316	.344	.378	.422	.482	.569	**.710**
	(.000)	(.001)	(.002)	(.003)	(.004)	(.006)	(.008)	(.012)	(.017)	(.024)

7. Tables

Tables 7.1 and 7.2 are for step-down and step-up FDR respectively. They give critical values for $m = 20$ when $MCV = t_{q,\nu}$, $\rho = .1(.2).9$ and $\nu = 15$, 30 and ∞. The tables suggest that the critical values (particularly the larger ones) are decreasing functions of both ρ and ν. The tables, and intuition, suggest that underestimating ρ (the assumed common correlation) leads to a conservative rejection (discovery) procedure.

Table 7.3 gives FDR step-down critical values for values of m ranging from 30 to 10,000 for $\rho = .5$, $\nu = \infty$ and the specified MCV. The MCV is chosen as the smallest value c for which $d_1 = d_2 = \ldots = d_{m-7} = c$ results in $E(Q) \leq .05$. It is designed for situations where relatively few false hypotheses are expected, or situations where a user is satisfied to find a few false hypotheses (make a few "discoveries"). Critical values for m not included in the table can be obtained by interpolation. Each critical value d_i is approximately linear in $\ln(m)$. A more accurate value for arbitrary values of m can be obtained by using the equation obtained by regressing d_{m-i} on $\ln(m)$, $[\ln(m)]^2$ and $1/\ln(m)$ (see Section 9.)

8. Some observations on the use of fewer than m critical values in step-down FDR

A considerable amount of simulation suggests that using a reduced number of critical values in step-down FDR results in a reduced FDR with a small subsequent reduction of power. Table 8.1 gives an example for the case $m = 14$, $q = .05$, $\Delta = 3.464$ and $\rho = .5$. The notation r/s is used to denote that at least r out of s false hypotheses were rejected. The five sets of rows gives powers and FDR (in parentheses) for 14, 4, 3, 2 or 1 critical values used. The powers when all 14 critical values are used are in bold and are underlined. The powers when the number of critical values used is equal to the number of false hypotheses are also underlined and in bold. These suggest that use of more critical values than the number of false

Table 7.1: Step-down FDR critical values for $m = 20$ and $q = .05$

	d_{20}	d_{19}	d_{18}	d_{17}	d_{16}	d_{15}	d_{14}	d_{13}	d_{12}	d_{11}	d_{10} to d_1
$\underline{\rho}$						$\underline{\nu = \infty}$					
0.1	2.79	2.55	2.39	2.26	2.15	2.06	1.97	1.88	1.81	1.64	1.64
0.3	2.74	2.52	2.36	2.24	2.15	2.06	1.96	1.88	1.79	1.64	1.64
0.5	2.65	2.45	2.31	2.20	2.11	2.02	1.94	1.85	1.72	1.64	1.64
0.7	2.48	2.32	2.21	2.12	2.04	1.96	1.88	1.77	1.64	1.64	1.64
0.9	2.18	2.08	2.01	1.94	1.88	1.81	1.67	1.64	1.64	1.64	1.64
$\underline{\rho}$						$\underline{\nu = 30}$					
0.1	2.98	2.70	2.52	2.38	2.26	2.15	2.05	1.95	1.86	1.72	1.70
0.3	2.91	2.65	2.48	2.35	2.24	2.13	2.04	1.95	1.84	1.70	1.70
0.5	2.79	2.57	2.42	2.30	2.20	2.10	2.01	1.92	1.77	1.70	1.70
0.7	2.60	2.43	2.31	2.21	2.12	2.03	1.95	1.82	1.70	1.70	1.70
0.9	2.26	2.16	2.08	2.01	1.94	1.87	1.71	1.70	1.70	1.70	1.70
$\underline{\rho}$						$\underline{\nu = 15}$					
0.1	3.20	2.87	2.66	2.50	2.37	2.25	2.14	2.03	1.93	1.77	1.75
0.3	3.10	2.81	2.62	2.47	2.34	2.23	2.12	2.02	1.90	1.75	1.75
0.5	2.95	2.71	2.54	2.41	2.29	2.19	2.09	1.98	1.82	1.75	1.75
0.7	2.74	2.54	2.41	2.30	2.20	2.11	2.02	1.88	1.75	1.75	1.75
0.9	2.36	2.25	2.16	2.09	2.01	1.93	1.76	1.75	1.75	1.75	1.75

Table 7.2: Step-up FDR critical values for $m = 20$ and $q = .05$

	d_{20}	d_{19}	d_{18}	d_{17}	d_{16}	d_{15}	d_{14}	d_{13}	d_{12}	d_{11}	d_{10} to d_1
$\underline{\rho}$						$\underline{\nu = \infty}$					
0.1	2.81	2.56	2.40	2.27	2.16	2.07	1.98	1.89	1.85	1.64	1.64
0.3	2.78	2.54	2.39	2.27	2.17	2.07	1.99	1.90	1.82	1.64	1.64
0.5	2.71	2.50	2.36	2.27	2.17	2.02	1.98	1.90	1.73	1.64	1.64
0.7	2.59	2.43	2.31	2.22	2.13	2.06	1.97	1.83	1.64	1.64	1.64
0.9	2.44	2.28	2.06	1.87	1.87	1.87	1.87	1.87	1.87	1.87	1.87
$\underline{\rho}$						$\underline{\nu = 30}$					
0.1	3.01	2.72	2.53	2.39	2.27	2.16	2.06	1.97	1.89	1.72	1.70
0.3	2.96	2.69	2.52	2.38	2.26	2.16	2.07	1.98	1.87	1.70	1.70
0.5	2.87	2.63	2.47	2.35	2.25	2.15	2.06	1.97	1.78	1.70	1.70
0.7	2.74	2.55	2.42	2.32	2.23	2.14	2.05	1.87	1.70	1.70	1.70
0.9	2.58	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10
$\underline{\rho}$						$\underline{\nu = 15}$					
0.1	3.24	2.90	2.68	2.52	2.38	2.26	2.15	2.05	1.96	1.77	1.75
0.3	3.17	2.86	2.66	2.50	2.37	2.26	2.15	2.05	1.93	1.75	1.75
0.5	3.05	2.79	2.61	2.47	2.35	2.25	2.15	2.05	1.83	1.75	1.75
0.7	2.92	2.70	2.55	2.44	2.33	2.23	2.13	1.92	1.75	1.75	1.75
0.9	2.64	2.31	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10

Table 7.3: Step-down FDR critical values for m for $q = .05$, $\rho = .5$, $\nu = \infty$

m	d_m	d_{m-1}	d_{m-2}	d_{m-3}	d_{m-4}	d_{m-5}	d_{m-6}	MCV(d_{m-7} to d_1)	ln(m)
30	2.752	2.567	2.459	2.339	2.280	2.191	2.114	1.983	3.401
40	2.830	2.649	2.536	2.439	2.379	2.300	2.231	2.110	3.689
50	2.883	2.715	2.602	2.510	2.450	2.376	2.316	2.195	3.912
60	2.930	2.763	2.656	2.571	2.499	2.438	2.375	2.265	4.094
70	2.964	2.798	2.702	2.616	2.546	2.488	2.425	2.320	4.248
80	2.995	2.838	2.736	2.644	2.598	2.528	2.470	2.365	4.382
90	3.021	2.868	2.758	2.718	2.643	2.564	2.506	2.403	4.500
100	3.047	2.897	2.766	2.731	2.650	2.595	2.538	2.437	4.605
200	3.200	3.069	2.956	2.900	2.839	2.788	2.736	2.641	5.298
300	3.291	3.154	3.056	3.011	2.936	2.892	2.842	2.750	5.704
400	3.350	3.213	3.119	3.081	3.011	2.962	2.914	2.825	5.991
500	3.395	3.260	3.192	3.107	3.063	3.015	2.967	2.881	6.215
1000	3.526	3.410	3.334	3.261	3.206	3.173	3.126	3.044	6.908
5000	3.816	3.700	3.622	3.592	3.561	3.535	3.472	3.388	8.517
10000	3.940	3.834	3.781	3.696	3.682	3.641	3.552	3.546	9.210

Table 8.1: Effects on power and FDR of using only the larger critical values (**cv** s) Step-down FDR

cv s used	$n_F = 1$	$n_F = \mathbf{2}$		$n_F = 3$			$n_F = 4$			
	1/1	**1/2**	**2/2**	**1/3**	**2/3**	**3/3**	**1/4**	**2/4**	**3/4**	**4/4**
14	**.821** (.050)	**.925** (.050)	**.785**	**.958** (.050)	**.904**	**.778**	**.973** (.050)	**.945**	**.898**	**.781**
4	.821 (.049)	.925 (.046)	.785	.958 (.042)	.904	.778	**.973** (.032)	**.945**	**.898**	**.781**
3	.821 (.048)	.925 (.043)	.785	**.958** (.032)	**.904**	**.777**	.973 (.024)	.945	.898	.777
2	.820 (.045)	**.925** (.018)	**.785**	.958 (.022)	.859	.729	.973 (.022)	.945	.868	.683
1	**.820** (.027)	.925 (.018)	.715	.958 (.013)	.859	.643	.973 (.010)	.915	.804	.590

hypotheses is counter productive since power increase, if any, is negligible and FDR is increased.

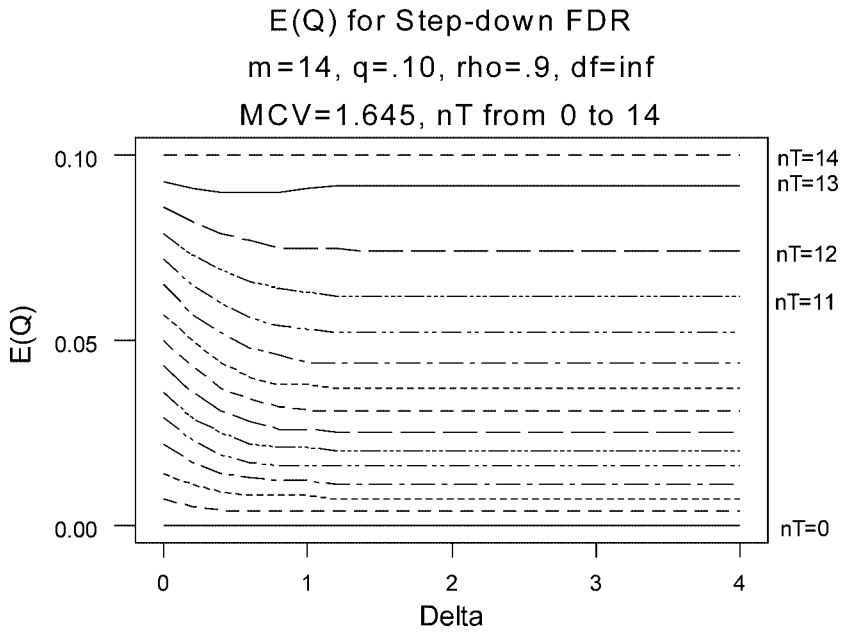
This suggests the utility of tables which limit the number of "unique" critical values such as Table 7.3. One convenient way of accomplishing this is by the use of sufficiently large MCVs.

9. Calculation of critical values and powers

Fortran 90 programs SEQDN and SEQUP can sequentially calculate the critical values d_2 to d_m for step-down and step-up FDR, respectively, for arbitrary values of m, q, ρ, and ν. N random normal multivariate vectors of size m are used to obtain each critical value.

Fortran 90 programs FDRPWRDN and FDRPWRUP calculate powers, $E(Q)$, $P[u \leq 1, 2, \ldots, 7]$, and $P[\gamma \leq .05, .10, .15]$ where u and γ are the number and proportion of false discoveries respectively. Inputs are m, ρ, ν, Δ, n_F and a set of m critical values. N random normal multivariate vectors of size m are also used. Three kinds of power are always calculated: per pair, all pairs and any pair (see Horn and Dunnett (2004)), and also $E(Q)$. The probability of rejecting at least one of the false hypotheses is called any pairs power. The probability of rejecting all false hypotheses is called all pairs power. Considering a specific hypothesis, the probability of its rejection is called the per pairs power. Since our calculations assume all the test statistics corresponding to the F hypotheses have the same location parameter, the per pair power is identical to the average power.

The Fortran 90 program FDRBIG calculates d_m for step-down FDR for arbitrary values of m, q, ρ and ν. The value d_m is also the critical value for Dunnett's (1955) comparisons with a control.



Graph 9.1:

For both step-down and step-up FDR, critical values were calculated using $m = 14$, $q = .01$, $.05$ and $.10$, values of ρ equal to 0, .05, .1, .5 and .9, various values

for MCV, and degrees of freedom for the test statistics equal to ∞. Additional computations were done using various values of m from 3 to 20, and for some special cases with $m = 9000$. The value of N was 10^7. Because, in many cases, the partial derivative of $E(Q)$ with respect to d_i is small, moderately large changes in d_i are required to effect a small change in $E(Q)$. This, combined with the dependence of the calculated values on $d_1, \ldots, d_{i-1}$, make accurate estimates, and assessments of their standard error of estimate, difficult to obtain. Error in the third decimal place should be expected. FDRPWRDN and FDRPWRUP were used to obtain not only the various powers, but to obtain $E(Q)$, to assure that the FDR requirement that $E(Q) \leq q$ was met.

FDRPWRDN and FDRPWRUP were used to calculate $E(Q)$ for $0 < n_T < m-1$ for $\Delta = 0(.2)2, 3, 4$ and sometimes 8. The values for N were 10^6 or 10^7. Calculation of critical values, $E(Q)$ and various powers for $m = 14$ and 100 , $\rho = .5$ were previously given in Somerville (2003).

When ρ was equal to .1, .5 or .9, and q equaled .01, .05 or .10, computations for step-down FDR suggested $E(Q)$ to be an increasing function of n_T for a given Δ. An example is given in Graph 9.1. Every computations did not show $E(Q) \leq q$, but the maximum value of $E(Q) - q$ was .0037, which occurred for $m = 14$, $q = .10$, $\rho = .9$, $n_T = 4$, MCV $= -1000$ and Δ large. For a given n_T, $E(Q)$ was typically, but not always an increasing function of Δ. Given the standard error of estimates for the critical values, and the programs which make extensive use of simulation, the calculations are not inconsistent with:

i) Using the m "least favorable configurations" for the calculation of critical values result in $E(Q) \leq q$.

ii) For a given Δ, $E(Q)$ is an increasing function of n_T for step-down FDR.

We again note that when MCV=0, the FDR step-down equations for the critical values are equivalent to those of Troendle (2000). We conjecture that, for a given MCV, no step-down FDR procedure is uniformly more powerful.

For Table 7.3, the values for d_m were obtained using Fortran 90 program FDRBIG. The other critical values were obtained using SEQDN with $N = 10^7$ for $m \leq 500$, and $N = 10^6$ for $m =$ 1,000, 5,000 and 10,000. The values were then "smoothed" using the regression of d_{m-i} on $\ln(m)$, $[\ln(m)]^2$ and $1/\ln(m)$. "Large" residuals (as defined by MINITAB regression calculations) were, except for one value, restricted to $m = 1,000$, 5,000 and 10,000. Except for the residuals .0070, .0041 and .0049, none exceeded .0025. The value of R^2 for each of the regression equations was 100.0%.

Values in Table 7.3 should have small standard errors, restricted to the third decimal. The critical values d_m have known standard error and $N = 10^8$ was used in FDRBIG. The values d_{m-7} were also calculated using $N = 10^8$. The small residuals resulting from the MINITAB regression on the calculated values gives further evidence of small error.

10. Comparisons with other procedures

Per pair and all pairs powers were calculated for $m = 100$ for the step-up procedure of Benjamini and Hochberg (1995), the stepdown procedure of Benjamini and Liu (1999) and the sequentially rejective procedure of Holm (1979). Graphs comparing the procedures with step-down procedures using 6 different MCV values are given in APPENDIX II. Also included, for the same procedures, is a graph comparing $E(Q)$.

11. Example

Korn, Troendle, McShane and Simon (2003) recently proposed two new procedures which control, with specified confidence, the actual number of false discoveries, and the actual proportion of false discoveries, respectively. They applied their procedures to analyze a microarray dataset consisting of measurements on approximately 9000 genes in paired tumor specimens, collected both before and after chemotherapy on 20 breast cancer patients. Their study, after elimination of cases of missing data included 8029 genes for analysis. Their Table 3 showed the genes with the 28 smallest "unadjusted paired t-test p-values for testing the null hypotheses that the mean pre and post chemotherapy expression of genes is the same".

Their Procedure A identified 28 genes where u, the number of false discoveries, was $\leq$2, with confidence .05. Procedure B identified the same 28 genes where γ, the false discovery proportion, was $\leq .10$.

We have used the Fortran program SEQDN to obtain stepdown FDR constants for $m = 8029$, $q = .05$, $\nu = \infty$, $\rho = 0$ and .1, and MCVs which result in 8 and 31 "unique" critical values. Table 11.1 gives the 34 smallest unadjusted p-values , and the corresponding critical p-values for the four cases. Table 11.2 gives the number of genes identified, and the minimum probabilities for $P[u \leq 2]$ over all possible values of n_F for the four cases. The values for $P[u \leq 2]$ were obtained using the Fortran program FDRPWRDN.

Note that as expected, decreasing MCV (increasing the number of "unique" critical values) increased the number of genes identifies. The price paid was an increased probability of false discoveries.

12. Summary and conclusions

Step-up and step-down FDR procedures are developed which are valid for dependent or independent hypotheses. The concept of a "Minimum Critical Value" (MCV) is introduced. The methodology (but not the computation) is independent of population distributions and uses least favorable configurations. The formulas developed produce a vector $\mathbf{d}$ of critical values, each element of which is the smallest possible, given the previously calculated values. We conjecture that there are no uniformly more powerful step-down FDR procedures. For the step-down case, the FDR procedure is the same as that of Troendle (2000) when the MCV = 0. For the step-up case, use only of least favorable configurations is sometimes insufficient, and for those cases, MCV's are also utilized.

Tables giving the critical values for step-up and step-down FDR when $m = 20$, $q = .05$ for several values of ρ and ν are given. A table of critical values when m is between 30 and 10,000 and $\rho = .5$ is given, designed for the case when the number of false hypotheses is small or when a satisfactory result is the selection of a small number of false hypotheses (discoveries).

An example of the use of the methodology, with data taken from the literature, is given.

Use of large values of MCV makes feasible the use of FDR for very large values of m, since fewer critical values are needed. Using fewer critical values reduces the expected number of discoveries, but is compensated by smaller expected numbers and proportions of false discoveries. Numerous calculations suggest that the "optimum" MCV (the one which results in the highest powers), is one for which the number of "unique" critical values is approximately equal to n_F, the number of false hypotheses.

Table 11.1: Unadjusted and critical p-values for $\rho = 0$ and .1 (8 and 31 "unique" critical values) Multiply all p-values by 10^{-6}

Rank	Unadjusted p-values	$\rho = 0$ MCV = 3.879	$\rho = .1$ MCV = 3.850	$\rho = 0$ MCV = 3.535	$\rho = .1$ MCV = 3.506
1	0.2	6.23	7.13	6.40	7.13
2	0.5	12.45	13.71	12.53	13.79
3	0.6	18.77	20.09	18.89	20.51
4	1.1	25.15	27.49	25.23	26.48
5	1.1	31.52	33.52	31.56	32.52
6	1.2	37.87	39.97	37.91	39.71
7	2.1	43.61	47.81	44.24	46.03
8	2.7	52.53	58.94	50.67	52.91
9	5.2	.	.	56.67	58.04
10	5.7	.	.	63.37	65.95
11	5.7	.	.	69.32	71.60
12	6.7			76.16	80.01
13	6.7			81.41	84.34
14	7.7			88.83	88.50
15	9.7			94.80	96.33
16	11.0			100.10	106.04
17	11.9			107.28	111.77
18	22.1			113.48	115.37
19	29.3			119.39	120.32
20	32.5			125.64	130.70
21	58.2			132.32	131.73
22	60.0			138.05	146.20
23	65.8			143.35	149.97
24	97.4			151.53	156.43
25	105.0			156.52	164.59
26	115.0			161.82	164.59
27	120.9			169.09	179.02
28	130.9			178.92	187.75
29	151.9			178.92	189.17
30	184.7			178.92	189.17
31	193.4	.	.	203.83	227.05
32	222.1	.	.	.	.
33	222.8	.	.	.	.
34	255.9	52.53	58.94	203.83	227.05

Table 11.2: Table of $P[u \leq 2]$ and # of genes identified

	8 "unique" critical values		31 "unique" critical values	
	$\rho = 0$	$\rho = .1$	$\rho = 0$	$\rho = .1$
# genes identified	20	21	29	33
$P[u \leq 2]$	.99	.96	.78	.78

A forthcoming paper will include tables of FDR step-down critical values for m between 50 and 10000, assuming common correlations of .0, .1 and .5, and for MCVs which result in 8 or 31 "unique" critical values. Also included will be a study of the effects of assuming a correlation of 0 when the data has higher correlations (robustness with respect to correlation).

Acknowledgments

The author is indebted to Frank Bretz for suggesting research on FDR procedures and for many contributions, comments, criticisms and suggestions. The author is also indebted to the referees for suggestions which materially improved the presentation. Manfred Horn read an earlier version of the paper and made many valuable suggestions. James Troendle supplied the unadjusted p-values for ranks 29–34 for Table 11.1. All random normal numbers were obtained using the "Monty Python" normal generator of Marsaglia and Tsang (1998). Graphs and regressions were obtained using MINITAB 13. All Fortran 90 programs were compiled and executed using Lahey Fortran 95. Almost all the computations were accomplished using an AMD Athlon 800 CPU.

References

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 289–300. MR1325392

[2] Benjamini, Y. and Liu, W. (1999). A step-down multiple hypotheses procedure that controls the false discovery rate under independence. *J. Statis. Plann. Inference* **82**, 163–170. MR1736441

[3] Benjamini, Y. and Yekutieli, D. (1999). The control of false discovery rate in multiple testing under dependency. Manuscript, Department of Mathematical Science, University of Tel Aviv.

[4] Benjamini, Y. and Liu, W. (2001). A distribution-free multiple-test procedure that controls the false discovery rate. Technical Report, Department of Statistics and Operations Research, Tel Aviv University.

[5] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165–1188. MR1869245

[6] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50**, 1096–1121.

[7] Horn, M. and Dunnett, C. W. (2004). Power and sample size comparisons of stepwise FWE and FDR controlling test procedures in the many-one case. *IMS Lecture Notes Monogr. Ser.* Inst. Math. Statist. Beechwood, OH.

[8] Korn, E. L., Troendle, James, McShane, Lisa M. and Simon, Richard (2003). Controlling the number of false discoveries: Application to high-dimensional genomic data. *J. Statis. Plann. Inference* To appear.

[9] Kwong, K. S., Holland, B. and Cheung, S. H. (2002). A modified Benjamini-Hochberg multiple comparisons procedure for controlling the false discovery rate. *J. Statis. Plann. Inference* **104**, (2002) 351–362. MR1906017

[10] Marsaglia, G. and Tsang, W. W. (1998). The Monty Python method for generating random variables. *ACM Trans. Math. Software* **24**(3), 341–350. MR1673744

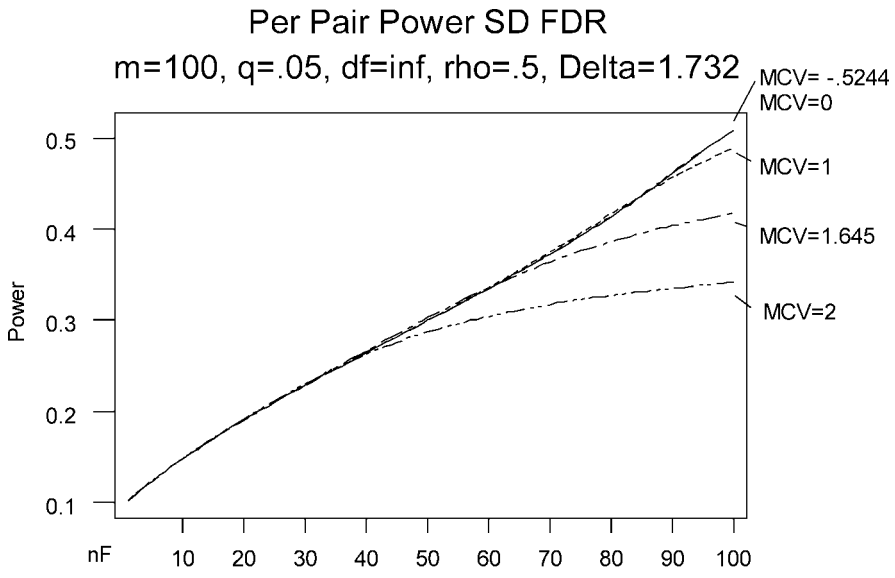
[11] Sarkar, Sanat K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.* **30**, 239–257. MR1892663

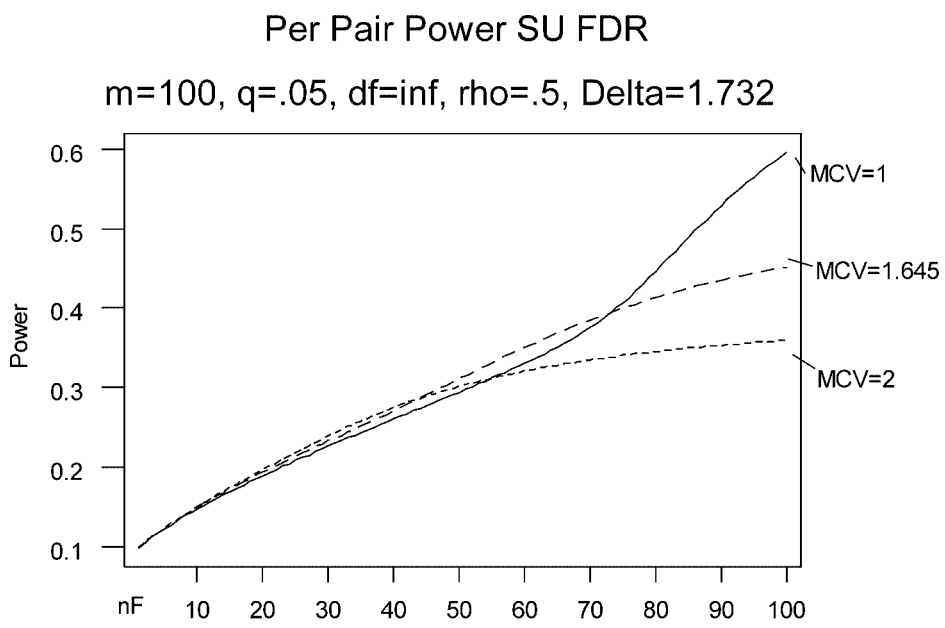
[12] Somerville, Paul N. (2003). "Optimum" FDR procedures and MCV values. Technical Report TR-03-01, Department of Statistics, University of Central Florida.

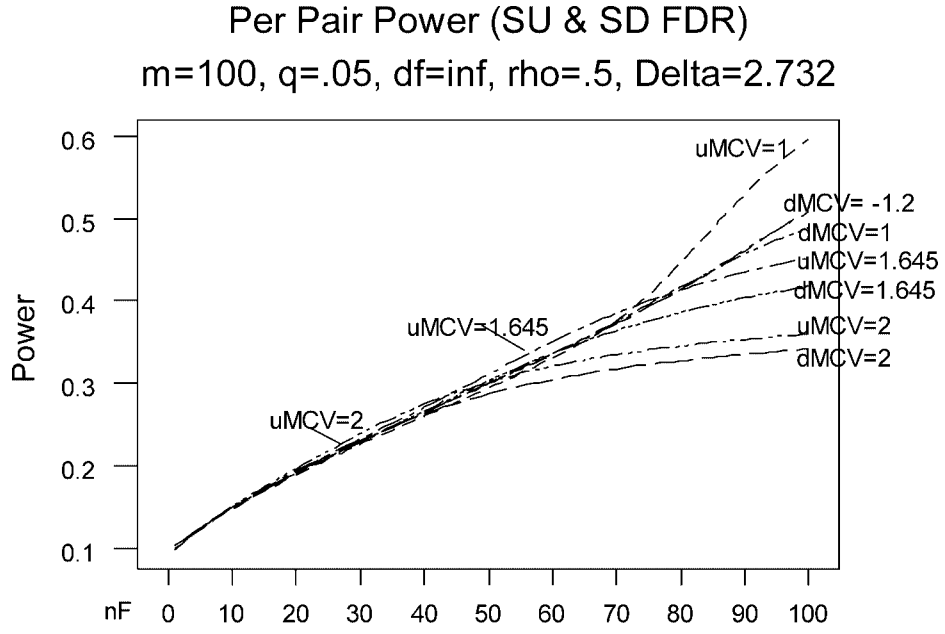
[13] Troendle, James F. (2000). Stepwise normal theory multiple test procedures controlling the false discovery rate. *J. Statis. Plann. Inference* **84**, 139–158. MR1747501

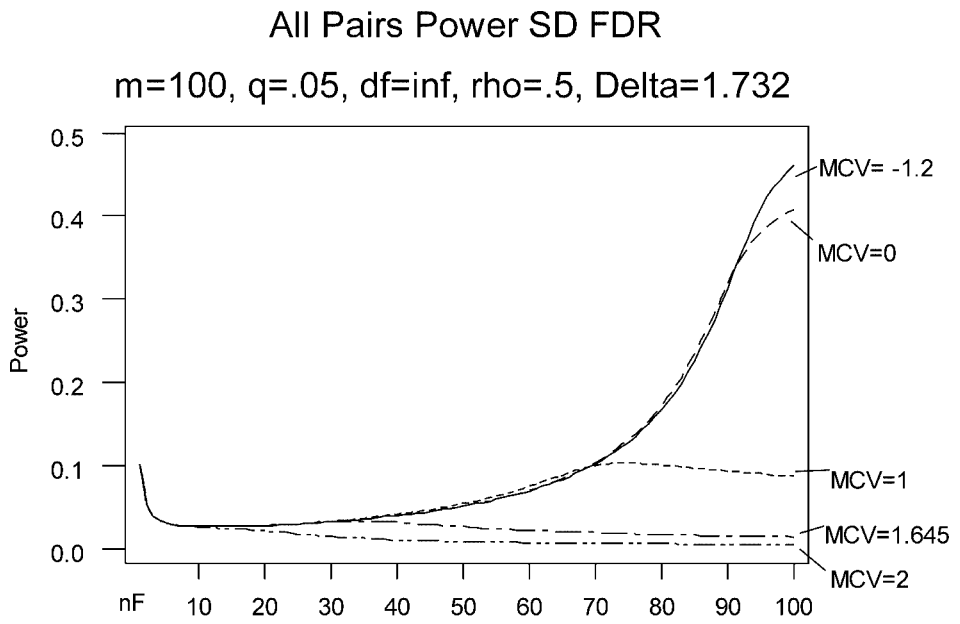
[14] Yekutieli, D. and Benjamini, Y., (1999). Resampling based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Statis. Plann. Inference* **82**, 171–196. MR1736442

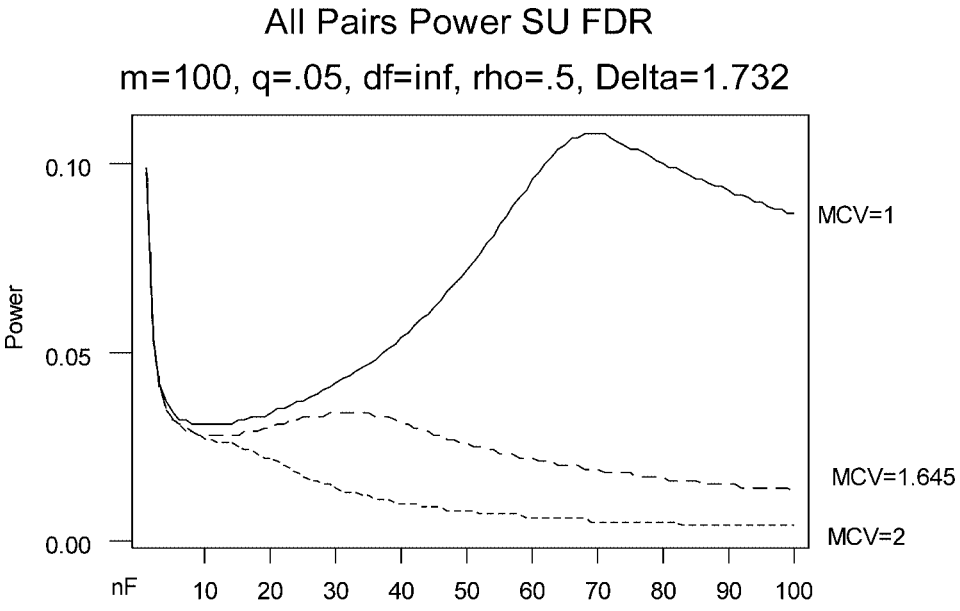
Appendix I

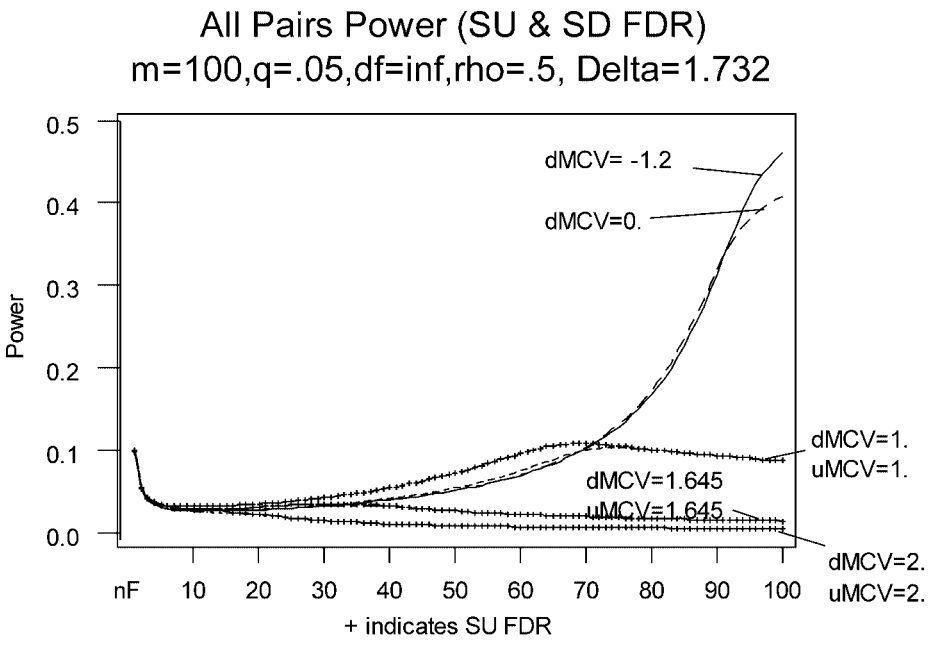


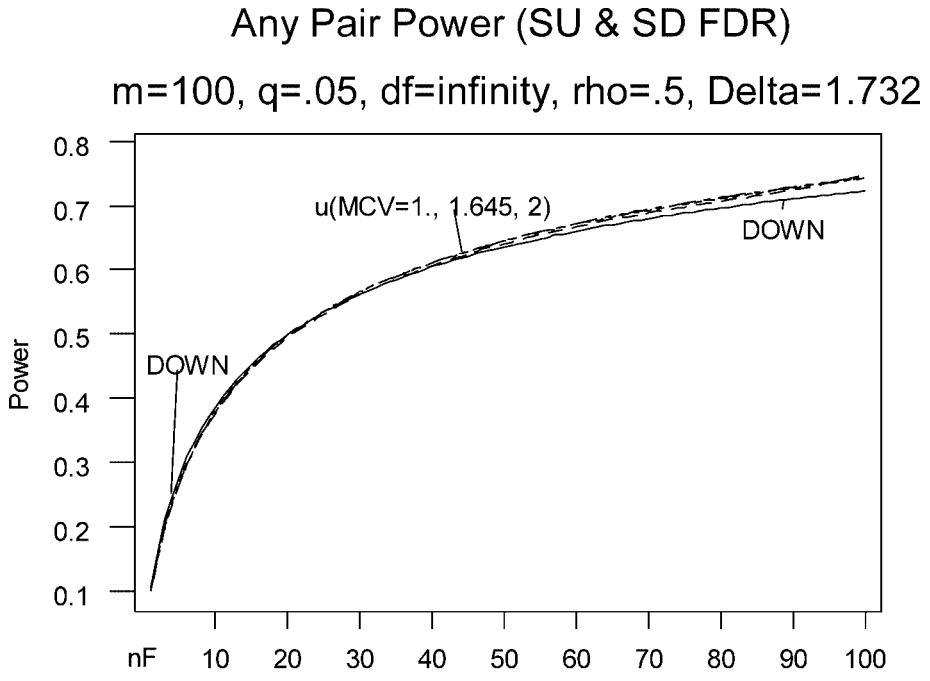

Per Pair Power SU FDR
m=100, q=.05, df=inf, rho=.5, Delta=1.732
Power
0.6
0.5
0.4
0.3
0.2
0.1
nF
10
20
30
40
50
60
70
80
90
100
MCV=1
MCV=1.645
MCV=2


Per Pair Power (SU & SD FDR)
m=100, q=.05, df=inf, rho=.5, Delta=2.732
Power
0.6
0.5
0.4
0.3
0.2
0.1
nF
0
10
20
30
40
50
60
70
80
90
100
uMCV=1
dMCV= -1.2
dMCV=1
uMCV=1.645
dMCV=1.645
uMCV=2
dMCV=2
uMCV=1.645
uMCV=2

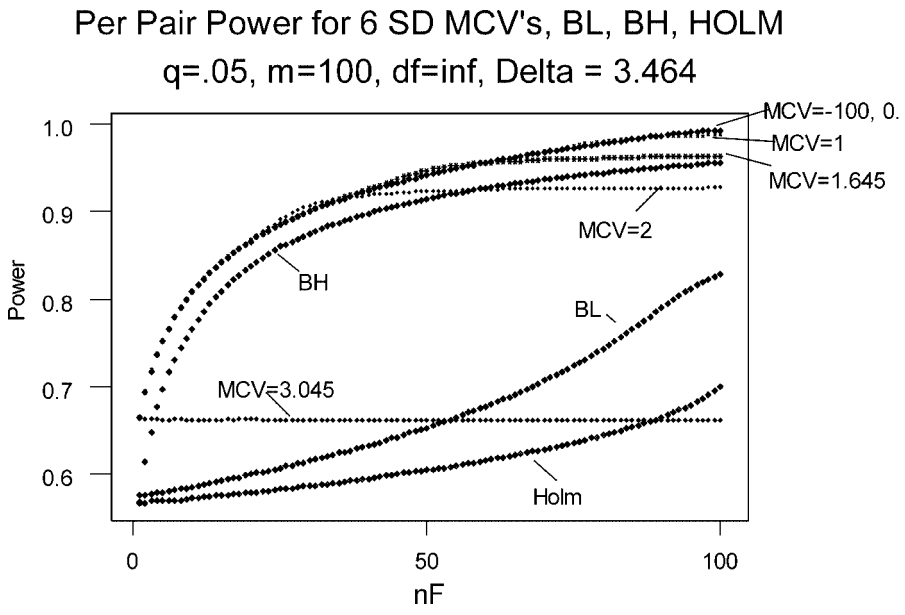

All Pairs Power SD FDR
m=100, q=.05, df=inf, rho=.5, Delta=1.732
Power
0.5
0.4
0.3
0.2
0.1
0.0
nF
10
20
30
40
50
60
70
80
90
100
MCV= -1.2
MCV=0
MCV=1
MCV=1.645
MCV=2

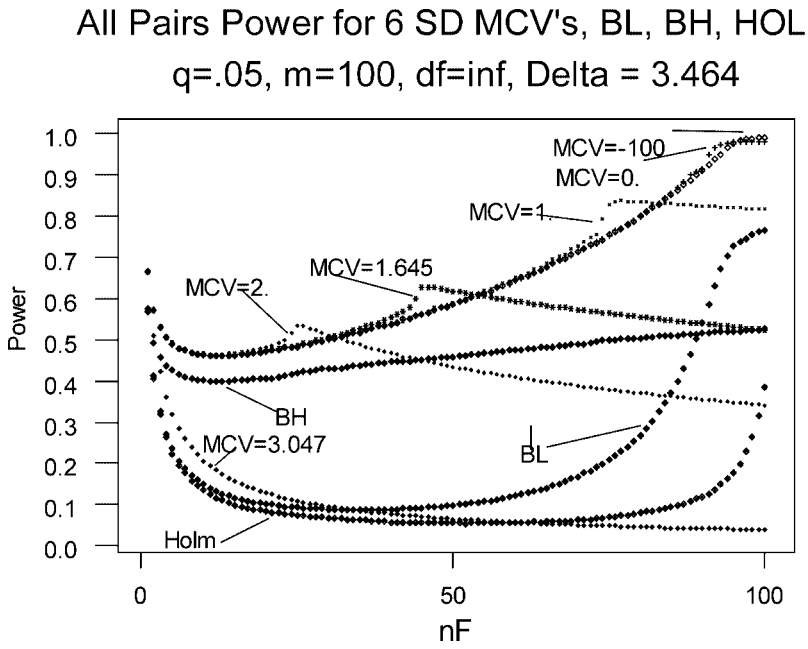

All Pairs Power SU FDR
m=100, q=.05, df=inf, rho=.5, Delta=1.732
Power
0.10
0.05
0.00
nF
10
20
30
40
50
60
70
80
90
100
MCV=1
MCV=1.645
MCV=2


All Pairs Power (SU & SD FDR)
m=100,q=.05,df=inf,rho=.5, Delta=1.732
Power
0.5
0.4
0.3
0.2
0.1
0.0
dMCV= -1.2
dMCV=0.
dMCV=1.
uMCV=1.
dMCV=1.645
uMCV=1.645
dMCV=2.
uMCV=2.
nF 10 20 30 40 50 60 70 80 90 100
+ indicates SU FDR


Any Pair Power (SU & SD FDR)
m=100, q=.05, df=infinity, rho=.5, Delta=1.732
Power
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
u(MCV=1., 1.645, 2)
DOWN
DOWN
nF 10 20 30 40 50 60 70 80 90 100

Appendix II


Per Pair Power for 6 SD MCV's, BL, BH, HOLM
q=.05, m=100, df=inf, Delta = 3.464
Power
1.0
0.9
0.8
0.7
0.6
0
50
100
nF
MCV=-100, 0.
MCV=1
MCV=1.645
MCV=2
BH
BL
MCV=3.045
Holm


All Pairs Power for 6 SD MCV's, BL, BH, HOLM
q=.05, m=100, df=inf, Delta = 3.464
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
50
100
nF
MCV=-100
MCV=0.
MCV=1
MCV=1.645
MCV=2.
BH
MCV=3.047
BL
Holm