
総 説

FDR の概説とそれを制御する多重検定法の比較

Introduction of FDR and Comparisons of Multiple Testing Procedures that Control It

松田真一

Shin-ichi Matsuda

南山大学 数理情報学部 情報システム数理学科

Department of Information Systems and Mathematical Sciences,
Faculty of Mathematical Sciences and Information Engineering,
Nanzan University
e-mail:matsu@nanzan-u.ac.jp

In this paper, we introduce the definition of FDR (False Discovery Rate), which gets a lot of attention as a new concept for considering multiplicity effect, and expound properties of it.

Furthermore, we enumerate multiple testing procedures that control FDR; Benjamini-Hochberg procedure (linear step-up procedure), Adaptive Benjamini-Hochberg procedure, Benjamini-Yekutieli procedure, Storey's procedure, two-stage linear step-up procedure, and Student-Newman-Kuels procedure. In addition, we review comparisons of them in order to consider which procedure is best to use.

As a result, we show the conservativeness of Benjamini-Hochberg procedure and the availability of two-stage linear step-up procedure by Monte-Carlo simulation, in the case of dependent test statistics.

Key words: false discovery rate; FWER (familywise error rate); multiple comparisons; linear step-up procedure; two-stage procedure; Monte-Carlo simulation.

1. はじめに

多重検定(特に多重比較)は歴史的には古い方法と言ってもよいが、概念として複雑な側面もあるため理解が難しかったり誤解されたりする統計的方法である。そのため、本論文では多重検定の基礎的な話から始め、最近話題となっている FDR (false discovery rate) の定義とそれを制御する多重検定法について説明していく。さらに、多重検定法の比較についてまとめを行い、どの手法を用いるのがよいかを提案する。なお、多重比較についてさらに詳細な情報を必要とする場合は永田・吉田 (1997) を参照されたい。また、本論文でも少し厳密に述べたい部分は途中で挟むと煩雑になるため付録として分けて述べることにした。(*ⁿ のマークでその箇所を明示している。)

2. 多重検定とは

検定は統計的推測の重要な概念であるが、それは一つの帰無仮説が成り立つかどうかを調べるために用いる方法論である。帰無仮説が複数あってそのどれが成り立つかを知りたい場合は多重検定となる。この状況が最も生じやすいのが多重比較であるのでまずその説明を行う。

多重比較は R. A. Fisher によって提唱された考え方である。多群の平均の比較を行うには分散分析が基礎となるが、その検定結果ではすべての群の平均に差がないか、どこかに差があるかしかならない。差があるとなった場合にどの群とどの群に差があるのかを突き止めるのが多重比較である。したがって、解析者が当然必要とする情報を提供するために発生した統計的方法である。しかし、多重比較の性質を理解していなかったり、誤用をしたりするとその必要性が失われる可能性があるののでしっかりした理解が重要である。

今、どの群も大きさ n であり、各群の平均を \bar{x}_i 、共通の分散の推定値を $\hat{\sigma}^2$ とした場合の 2 群間の比較を考える。

$$\frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{2\hat{\sigma}^2/n}}$$

が 2 群間比較の検定統計量となるが、これらを同時に扱うため多重比較と呼ばれる。検定が複数行われるわけであるから多重比較とは多重検定の一種である。

問題となるのはこの検定統計量の棄却限界値をどうするかであるが、そのためには多重比較や多重検定における第 1 種の過誤とは何かを考える必要がある。

3. 多重比較における第 1 種の過誤とは

多重検定は Fisher の頃より多重比較を中心として研究が進んできた。そのためまず多重比較の第 1 種の過誤について考える。

多重比較の考え方は簡単である。多重比較はまず分散分析のような全体を見る視点が欠かせない。したがって、全体としての有意水準を守るように考えればよい^{*1}。Fisher はそのため最初に分散分析を行い、それで棄却された場合のみ 2 群間の比較を普通の t 検定で行えばよいと考えた^{*2}。この方法を protected LSD (least significant difference) 法と呼ぶ。この方法は一見問題がなさそうに見えるが、J. W. Tukey によって提唱され、その後整備された多重比較の第 1 種の過誤の概念である FWE (familywise error) から考えると、不都合が生じる。(FWE の概念の詳細は Hochberg and Tamhane (1987), 永田・吉田 (1997) を参照のこと)

FWE とは「真の帰無仮説のうち、誤って一つでも棄却すること」を指している^{*3}。FWE の確率のことを FWER (familywise error rate) と呼ぶ^{*4}。この FWER が多重比較における有意水準に当たる。前節での多群の比較の問題で考えてみよう。

群が 4 群であった場合、元々の分散分析の帰無仮説は $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ となる。多重比較においては仮説は 2 群間に分けられ $H_{01}: \mu_1 = \mu_2$, $H_{02}: \mu_1 = \mu_3$, $H_{03}: \mu_1 = \mu_4$, $H_{04}: \mu_2 = \mu_3$, $H_{05}: \mu_2 = \mu_4$, $H_{06}: \mu_3 = \mu_4$ となる。6 つの仮説 $H_{01} \sim H_{06}$ がすべて真の場合は H_0 も成り立つことになり、Fisher の protected LSD 法でも FWER は制御できる。しかし、 H_{01} と H_{06} だけが真で残りが偽の場合はどうであろうか？ この場合は当然 H_0 は偽となる。したがって、最初の分散分析で棄却されることは望ましいことで H_{01} と H_{06} をそれぞれ設定された有意水準で検定する

ことになる．そのため，protected LSD 法では実質的に $1 - (1 - \alpha)^2$ ($\alpha = 0.05$ のときは 0.0975) の確率でどちらかが棄却されることになり，FWER は制御されない．

4. 多重検定における第 1 種の過誤とは

多重比較は全体の検定に意味があり，それを制御する必要から FWE という概念が生じた．一方，多重検定という場合，多重比較より広く色々な場合が考えられる．たとえば，多項目検定が代表的な例であろう．この場合，全体的な制御が必要であろうか？ もし全体的な制御が必要であるような検証的な側面があるなら多重比較と同様な FWER を基準とするのが妥当であろう．もしそうでないならどうすればよいのかが大きな問題であった．特に第 2 種の過誤も考慮すべき場合，従来では個々に通常の検定を行えばよいという考えもあった．しかし，仮説の数が多い場合，第 1 種の過誤は全体的に見ると全く守られないものとなり，いくら探索的であっても非効率ではないかと思われる．この問題点についてはまだ実用レベルで解決していないが，FDR を制御する多重検定法を用いることがよいと考えられる．検定を繰り返すと FWER を制御のどちらもよくないという状況は遺伝子解析と似た探索的な問題を含む状況だからである．(Benjamini and Hochberg (1995) の例題および Benjamini et al. (2005) での解説もその可能性を示唆している.)

5. FDR とは

Benjamini and Hochberg (1995) が多重検定に対する新たな概念として提案したのが FDR である．

5.1 FDR の定義

前もって与えられた m 個の帰無仮説を同時に検定する問題を考える．そのうち m_0 個の帰無仮説が真， m_1 個の帰無仮説が偽とする．($m = m_0 + m_1$ となる．) また， R は検定によって棄却される帰無仮説の数を表す確率変数とする．表 1 は，その状況をまとめたものである． R は，観測可能な確率変数であるが， U, V, S, T は，一般的には観測不可能な確率変数である．なお，これらの確率変数に関して，FWER は $P(V \geq 1)$ と表せる．

表 1. m 個の帰無仮説の検定結果の分割表

	検定		計
	採択	棄却	
真の帰無仮説	U	V	m_0
偽の帰無仮説	T	S	$m - m_0$
計	$m - R$	R	m

R 個の棄却された帰無仮説のうち誤って棄却された過誤の割合は，確率変数 $Q = \frac{V}{R}$ によって調べることができる．ただし， $R = 0$ のときは $Q = 0$ と定義する．このとき，Benjamini and Hochberg (1995) は， Q に対してこの検定全体での期待値をとった以下の Q_e を定義した．

$$Q_e = E(Q) = E\left(\frac{V}{R}\right)$$

この Q_e が狭い意味での FDR の定義である。方法として守るべき基準は帰無仮説の真偽の様々な状況における FDR の最大値であり、それが広い意味での FDR となる*5。

5.2 FDR と FWER の関係

FWER は確率で FDR は期待値であるが、FWER と FDR の間には次のような関係がある。(Benjamini and Hochberg (1995) 参照)

- 1) 全ての帰無仮説が真であるならば、FDR は FWER に等しい
- 2) $m_0 < m$ の時、FDR は FWER より小さいか等しい

その結果、FWER を制御するどんな方法も、FDR は制御されている。しかし、FDR を制御する方法は、FWER を制御しない場合がある。そのため、検証的側面があるときに FDR を制御する方法を用いるのは全体との整合性の面で問題がある。すなわち、FDR を制御する方法は探索的問題への適用に限定すべきである。また、帰無仮説の数多くてそれなりの偽の帰無仮説がある場合により効果を発揮するため、遺伝子解析で広く使われる理由となっている。

5.3 FDR の基準値

FWER の場合の有意水準に当たる FDR を制御する基準値 (Benjamini and Hochberg (1995) に習って q^* で表すことにする) はどのぐらいにすべきであろうか。前節の結果より FDR は FWER を越えないこともあって一般的には通常の有義水準と同じ 0.05 を用いることが多いようである。有意水準自身も特別意味のある数字ではないのでこの基準値 q^* も自由にとると意味が後付になりやすくなる。その観点からしても有意水準と同じ値を用いることが定着するとよいのではないかと考えられる。(有意水準も問題によっては 0.01 や 0.1 を用いるのと同じ意味合いでこの基準値 q^* を変更することは構わないが、意味もなく変えないことが重要だということである。)

6. FDR を制御する多重検定法

FDR を制御する多重検定法は大きく分けると 3 つの種類に分けられる。(Benjamini et al. (2005) 参照。)

- (1) p 値に基づく線形上昇手順
- (2) p 値に基づく線形上昇手順の基準値に何らかの修正を施したもの
- (3) その他の方法

分類(1)に含まれるのは Benjamini and Hochberg (1995) の方法(以降 BH 法とする)であり、その変形として Benjamini and Yekutieli (2001) の方法(以降 BY 法とする)も挙げられる。分類(2)に含まれるのは BH 法の基準値を直接改良した方法とベイズ流に何らかのパラメータを導入し BH 法の基準値を推定した方法である。分類(3)には BH 法と全く系統の異なる方法や 2 段階法がある。

ここでは分類(1)の BH 法と BY 法、分類(2)に含まれる Adaptive BH 法と Storey の方法、さらに分類(3)に含まれる 2 段階法と SNK 法を述べる。

6.1 BH 法

帰無仮説 H_1, H_2, \dots, H_m は、それぞれ対応した p 値として P_1, P_2, \dots, P_m を持つと仮定する。さらに、昇順に並べ替えられた p 値を $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$ と表し、 $P_{(i)}$ に相当する帰無仮説は $H_{(i)}$ と表す。このとき、BH 法は以下のような手順である。(Benjamini and Hochberg (1995), 堀内・松田 (2006) 参照)

手順 1 $i = m$ とする。

手順 2 $P_{(i)} \leq \frac{i}{m} q^*$ を満たすならば、 $k = i$ として**手順 3**へ進む。そうでなければ $i - 1$ を改めて i とおき、**手順 2**を繰り返す。ただし、 $i = 1$ まで達したのなら、どの仮説も棄却することなく終了する。

手順 3 仮説 $H_{(i)}; i = 1, 2, \dots, k$ をすべて棄却する。

BH 法は帰無仮説が互いに独立である場合に FDR を制御することが証明されている。BH 法は p 値の大きい方から検定を行い、棄却限界値が直線的なため線形上昇手順 (linear step-up procedure) と呼ばれている。

6.2 BY 法

最初の条件は BH 法と同じである。BY 法は以下のような手順である。(Benjamini and Yekutieli (2001), 堀内・松田 (2006) 参照)

手順 1 $q^{**} = q^* / \sum_{j=1}^m \left(\frac{1}{j} \right)$ とする。

(以下、 q^* の代わりに q^{**} を用いて、BH 法を行う。)

手順 2 $i = m$ とする。

手順 3 $P_{(i)} \leq \frac{i}{m} q^{**}$ を満たすならば、 $k = i$ として**手順 4**へ進む。そうでなければ $i - 1$ を改めて i とおき、**手順 3**を繰り返す。ただし、 $i = 1$ まで達したのなら、どの仮説も棄却することなく終了する。

手順 4 仮説 $H_{(i)}; i = 1, 2, \dots, k$ をすべて棄却する。

BY 法は正の相関がある場合でも FDR を制御することが証明されている。

6.3 Adaptive BH 法

最初の条件は BH 法と同じである。Adaptive BH 法 (以降 ABH 法とする) は以下のような手順である。(Benjamini and Hochberg (2000), 堀内・松田 (2006) 参照)

手順 1 基準値 q^* で BH 法を実行する。もし 1 つも仮説を棄却することができないならば、すべての仮説を棄却することなく終了する。それ以外ならば**手順 2**へ進む。

手順 2 計算式 $S_i = \frac{1 - P_{(i)}}{m + 1 - i}$ によって、 $S_i (i = 1, 2, \dots, m)$ を計算する。

手順 3 $i = 2$ とおく。

手順 4 $S_i \geq S_{i-1}$ ならば $i+1$ を改めて i とおき、手順 4 を繰り返す。

手順 5 $S_i < S_{i-1}$ と初めてなった時の S_i を S^* とおく。最後まで $S_i \geq S_{i-1}$ ならば、 S_m を S^* とおく。

手順 6 $\hat{m}_0 = \min \left(\left\lceil \frac{1}{S^*} + 1 \right\rceil, m \right)$ とおく。ここで $\lceil \cdot \rceil$ はガウス記号を表す。

手順 7 $i = m$ とする。

手順 8 $P_{(i)} \leq \frac{i}{\hat{m}_0} q^*$ を満たすならば、 $k = i$ として次へ進む。そうでなければ $i-1$ を改めて i とおき、手順 8 を繰り返す。($\hat{m}_0 \leq m$ なので必ず次に進む k が見つかる。)

手順 9 仮説 $H_{(i)}; i = 1, 2, \dots, k$ をすべて棄却する。

ABH 法は帰無仮説が互いに独立である場合にシミュレーションにより FDR を制御することが確認されている。

6.4 Storey の方法

Storey et al. (2004) の提案した方法は Storey (2002) の方法 (以降 ST 法とする) がベースになっている。その方法は λ というパラメータを導入してベイズ流の考えで真の仮説数の割合を推定しながら検定を行うもので、記号で $t_{q^*}(\widehat{\text{FDR}}_\lambda)$ と表される p 値の棄却域を用いる方法が元になっている。この詳細な説明は後に回して重要な 2 つ補題を先に述べる。

補題 1 p 値の棄却域が $t_{q^*}(\widehat{\text{FDR}}_{\lambda=0})$ で定められる上昇手順は BH 法と同値である。

補題 2 p 値の棄却域が $t_{q^*}(\widehat{\text{FDR}}_\lambda)$ で定められる上昇手順は m を $\hat{\pi}_0(\lambda)m$ と置き換えた BH 法と同値である*⁶。

したがって、ST 法は BH 法の手順で行えることになる。(Storey et al. (2004) では、有限標本の場合の修正を施している (以降その方法を MST 法とする) が、詳細は次節に譲る。) Storey et al. (2004) によると $\lambda = 0.5$ による ST 法は BH 法より q^* を正確に守るとされているが、ある一つの設定でのシミュレーションによる結果である。また、 λ を固定した場合の ST 法は仮説が独立な場合だけでなく、漸近的にある条件を満たす弱い従属の場合でも q^* を守ることが示されている。

Storey et al. (2004) では λ を固定せずブートストラップによる探索を行う方法も示しているが、それは q^* をより正確に守るための工夫である。その詳細はここでは省くが、Storey によってプログラムが公開されている (Storey (2008) 参照*⁷) のでそれを用いるとよい。

6.4.1 Storey の記号の説明

表 1 の記号法を拡張する。 p 値 P_1, P_2, \dots, P_m とある値 t ($0 \leq t \leq 1$) に対して

$$V(t) = (P_i \leq t \text{ となる真の帰無仮説数})$$

$$S(t) = (P_i \leq t \text{ となる偽の帰無仮説数})$$

$$R(t) = V(t) + S(t)$$

とおく. さらに,

$$\text{FDR}(t) = E \left[\frac{V(t)}{\max(R(t), 1)} \right]$$

とおく. Storey (2002) は調整変数 λ を用いてその関数 $\text{FDR}(t)$ を次の関数 $\widehat{\text{FDR}}_\lambda(t)$ で推定することを提案している.

$$\widehat{\text{FDR}}_\lambda(t) = \frac{\hat{\pi}_0(\lambda)t}{\max(R(t), 1)/m}$$

ここで, $\hat{\pi}_0(\lambda)$ は $\pi_0 = m_0/m$ の推定値でベイズ流の考えにより次式で与えられるものである.

$$\hat{\pi}_0(\lambda) = \frac{m - R(\lambda)}{(1 - \lambda)m}$$

また, $[0, 1]$ 上で定義された関数 F に対する $t_{q^*}(F)$ という関数は

$$t_{q^*}(F) = \sup\{0 \leq t \leq 1 : F(t) \leq q^*\}$$

で与えられ, これは p 値の上昇手順の棄却域を定めることに対応する. これらの関数により, p 値の棄却域が $t_{q^*}(\widehat{\text{FDR}}_\lambda)$ で定められる ST 法が構成される.

一方, MST 法での修正は 2 点あり, 1 点は $\hat{\pi}_0(\lambda)$ の代わりに

$$\hat{\pi}_0^*(\lambda) = \frac{m + 1 - R(\lambda)}{(1 - \lambda)m}$$

を使うことである.

もう 1 点は $\widehat{\text{FDR}}_\lambda(t)$ の代わりに

$$\widehat{\text{FDR}}_\lambda^*(t) = \begin{cases} \frac{\hat{\pi}_0^*(\lambda)t}{\max(R(t), 1)/m} & (t \leq \lambda \text{ のとき}) \\ 1 & (t > \lambda \text{ のとき}) \end{cases}$$

を使うことである.

6.5 2 段階線形上昇手順

最初の条件は BH 法と同じである. 2 段階線形上昇手順 (Two-stage linear step-up procedure. 以降 TST 法とする) は Benjamini et al. (2005) で提案された以下のような方法である.

手順 1 基準値 $q' = q^*/(1 + q^*)$ を用いて BH 法を行う. 棄却された仮説数を r_1 とおく. $r_1 = 0$ ならばすべての仮説を棄却することなく終了する. $r_1 = m$ ならばすべての仮説を棄却して終了する. それ以外なら **手順 2** へ進む.

手順 2 $\hat{m}_0 = m - r_1$ とおく.

手順 3 基準値 q'/\hat{m}_0 を用いて BH 法を行う.

この方法は帰無仮説が互いに独立である場合に FDR を制御することが証明されている.

6.6 SNK 法

Student-Newman-Kuels の方法(以降 SNK 法とする)は多重比較法である。多重比較専用の方法として FDR を制御できる方法である(Oehlert (2000) 参照)。もちろん先に述べたように多重比較は多重検定の一種であるので多重比較の問題に対して BH 法などの多重検定法を用いることはできる。

実はこの方法は Tukey-Welsch の方法が提案されるまで多段階手順としてよく知られた方法であったが, FWER を守らないことが分かったため使われなくなった方法である。FDR の概念の登場とともに復活したと言えるであろう。本論文の主題である多重検定法ではないため, 手順は概略のみに留める。

SNK 法は全体での分散分析の後, その部分群の分散分析を行っていく多段階法である。Tukey-Welsch の方法はその際有意水準を変更していくが, SNK 法は常に α を使い続けるという方法である。そのため FWER は守られない。検定統計量としては分散分析の F 統計量以外もあるため手順の詳細は永田・吉田 (1997) に譲る。

7. 適用例

前節での多重検定法の適用例を示す。

表 2 は平成 18 年の悪性新生物による部位別死亡者数について, 中部以西の本州の 20 府県で違いがあるかどうか分析した p 値である*⁸。

表 2. 平成 18 年悪性新生物部位別死亡者数要因分析 p 値

富山県	6.18×10^{-6}	岐阜県	4.02×10^{-5}	京都府	0.0966	鳥取県	0.9030
石川県	0.0151	静岡県	4.17×10^{-6}	大阪府	0.0129	島根県	0.0047
福井県	0.4400	愛知県	1.19×10^{-13}	兵庫県	0.1171	岡山県	0.0356
山梨県	0.0018	三重県	0.0013	奈良県	0.0092	広島県	1.09×10^{-7}
長野県	6.02×10^{-8}	滋賀県	0.3284	和歌山県	0.0608	山口県	0.1202

「胃」, 「結腸・直腸」, 「肝・胆・膵」, 「気管支・肺」, 「その他」の区分で平成 18 年の悪性新生物による死亡者数を調べ, 20 府県全体での比率を所与として各府県の乖離度である p 値を自由度 4 のカイ 2 乗分布による適合度検定で求めたものである。 p 値が小さくて棄却されるならその府県独自の傾向があることになる。こういった府県がそれに当たるか FDR を制御する多重検定法で求めてみる。今回適用するのは BH 法, ABH 法, $\lambda = 0.5$ の ST 法, TST 法である。また, $q^* = 0.05$ とする。

7.1 BH 法の適用

まず, p 値を小さいものから順に並べる。

$$\begin{aligned}
 P_{(1)} &= 1.19 \times 10^{-13}, & P_{(2)} &= 6.02 \times 10^{-8}, & P_{(3)} &= 1.09 \times 10^{-7}, & P_{(4)} &= 4.17 \times 10^{-6}, \\
 P_{(5)} &= 6.18 \times 10^{-6}, & P_{(6)} &= 4.02 \times 10^{-5}, & P_{(7)} &= 0.0013, & P_{(8)} &= 0.0018, \\
 P_{(9)} &= 0.0047, & P_{(10)} &= 0.0092, & P_{(11)} &= 0.0129, & P_{(12)} &= 0.0151, \\
 P_{(13)} &= 0.0356, & P_{(14)} &= 0.0608, & P_{(15)} &= 0.0966, & P_{(16)} &= 0.1171, \\
 P_{(17)} &= 0.1202, & P_{(18)} &= 0.3284, & P_{(19)} &= 0.4400, & P_{(20)} &= 0.9030
 \end{aligned}$$

手順 1 $i = 20$ とする.

手順 2 $P_{(20)} = 0.9030 > \frac{20}{20}q^* = 0.05$ なので $i = 19$ として手順 2 を繰り返す.

$P_{(19)} = 0.4400 > \frac{19}{20}q^* = 0.0475$ なので $i = 18$ として手順 2 を繰り返す.

以下同様にして $i = 13$ までは式を満たさない.

$P_{(12)} = 0.0151 \leq \frac{12}{20}q^* = 0.0300$ なので $k = 12$ として手順 3 へ進む.

手順 3 仮説 $H_{(i)}; i = 1, 2, \dots, 12$ をすべて棄却する.

棄却された府県は、愛知県、長野県、広島県、静岡県、富山県、岐阜県、三重県、山梨県、島根県、奈良県、大阪府、石川県の 12 府県である.

7.2 ABH 法の適用

手順 1 前節の結果より BH 法で棄却される仮説が存在する.

手順 2 $S_1 = \frac{1 - P_{(1)}}{20 + 1 - 1} = 0.05$, $S_2 = \frac{1 - P_{(2)}}{20 + 1 - 2} = 0.0526$, $S_3 = 0.0556$, $S_4 = 0.0588$,
 $S_5 = 0.0625$, $S_6 = 0.0667$, $S_7 = 0.0713$, $S_8 = 0.0767$, $S_9 = 0.0829$, $S_{10} = 0.0901$,
 $S_{11} = 0.0987$, $S_{12} = 0.1094$, $S_{13} = 0.1206$, $S_{14} = 0.1342$, $S_{15} = 0.1506$,
 $S_{16} = 0.1766$, $S_{17} = 0.2200$, $S_{18} = 0.2239$, $S_{19} = 0.2800$, $S_{20} = 0.0970$

手順 3 $i = 2$ とおく.

手順 4 $S_2 \geq S_1$ なので $i = 3$ として手順 4 を繰り返す. 以下, $i = 19$ まで満たす.

手順 5 $S_{20} < S_{19}$ となるので S_{20} を S^* とおく.

手順 6 $\hat{m}_0 = \min\left(\left\lceil \frac{1}{S^*} + 1 \right\rceil, 20\right) = 11$ となる.

手順 7 $i = 20$ とする.

手順 8 $P_{(20)} = 0.9030 > \frac{20}{\hat{m}_0}q^* = 0.0909$ なので $i = 19$ として手順 8 を繰り返す.

$i = 15$ まで式を満たさない. $i = 14$ で $P_{(14)} = 0.0608 \leq 0.0636$ なので,

$k = 14$ として手順 9 へ進む.

手順 9 仮説 $H_{(i)}; i = 1, 2, \dots, 14$ をすべて棄却する.

BH 法の 12 府県の外に、岡山県、和歌山県が棄却される.

7.3 $\lambda = 0.5$ の ST 法の適用

$\hat{\pi}_0(0.5) = \frac{20 - R(0.5)}{(1 - 0.5) \times 20} = \frac{1}{10} = 0.1$ である.

したがって, m を $\hat{\pi}_0(0.5) \times 20 = 2$ と置き換えて BH 法を適用する.

手順 1 $i = 20$ とする.

手順 2 $P_{(20)} = 0.9030 > \frac{20}{2}q^* = 0.5$ なので $i = 19$ として手順 2 を繰り返す.

$$P_{(19)} = 0.4400 \leq \frac{19}{2}q^* = 0.475 \text{ なので } k = 19 \text{ として手順 3 へ進む.}$$

手順 3 仮説 $H_{(i)}; i = 1, 2, \dots, 19$ をすべて棄却する.

鳥取県以外がすべて棄却される.

ちなみに, Storey のプログラムにより, ブートストラップによる ST 法は $\hat{\pi}_0 = 0.091$ で同じ結果となり, スプラインによる ST 法は $\hat{\pi}_0 = 0.318$ で 17 府県が棄却される.

7.4 TST 法の適用

手順 1 基準値 $q' = q^*/(1 + q^*) = 0.04762$ を用いて BH 法を行う.

$$P_{(13)} = 0.0356 > \frac{13}{20}q' = 0.03095, \quad P_{(12)} = 0.0151 \leq \frac{12}{20}q' = 0.02857 \text{ なので}$$

棄却された仮説数は $r_1 = 12$ で **手順 2** へ進む.

手順 2 $\hat{m}_0 = 20 - r_1 = 8$ とおく.

手順 3 基準値 $q' \frac{20}{\hat{m}_0} = 0.11905$ を用いて BH 法を行う.

$$P_{(15)} = 0.0966 > \frac{15}{20} \times 0.11905 = 0.08929, \quad P_{(14)} = 0.0608 \leq \frac{14}{20} \times 0.11905 = 0.08333$$

なので仮説 $H_{(i)}; i = 1, 2, \dots, 14$ をすべて棄却する.

ABH 法と同じ結果で 14 府県が棄却される.

7.5 考 察

この要因分析自身の話は付録^{*9}に譲るとしてここでは多重検定法の適用特性の考察にのみ話を絞る.

BH 法はその特性から棄却される p 値の最大値は q^* 以下であり, 個別に検定した場合を越えないという意味で整合性が取れている. ABH 法と TST 法はともに 0.0608 の p 値が棄却されている. これは FDR の特徴とも言えるものであるが, この不自然さは棄却される仮説数が全体の仮説数に対して多くなると見受けられる. 真に棄却すべき仮説が多いのなら割合である FDR の制御で大きな p 値の棄却に寛容になるのは仕方がない. そもそもこのデータは Bonferroni の方法(限界値 $0.05/20 = 0.0025$)や Holm の方法(限界値 $0.05/(m - i + 1)$ を用いる下降手順)でも 8 府県が棄却される違いの大きなデータであり, ある程度の仮説の棄却という点では FDR を制御する多重検定法に頼るまでもないものであった.

一方, $\lambda = 0.5$ の ST 法は棄却されすぎに感じる. λ を推定する ST 法の方が実用的であろう.

8. 多重検定法の比較

Benjamini et al. (2005) では様々な多重検定法をシミュレーションで比較している. そこでは独立標本の場合に λ を固定した MST 法がよい成績を挙げている. 一方, 正の相関がある場合は λ を固定した MST 法と ABH 法には FDR を制御できない場合が見られ, TST 法が全体的によりとされている.

堀内・松田 (2006) では負の相関も含めたより広い設定でシミュレーション比較を行っている. その目的は, 様々な設定において BH 法, ABH 法, BY 法が FDR を基準値 $q^* = 0.05$ 以下に保

てているのかを確認することである。以下のようなパターンで仮説数や相関を変えた合計 116 通りの設定について各々 10000 回のシミュレーションを行った。(乱数の発生方法等は堀内・松田 (2006) を参照せよ。)

1 〈仮説間が独立な場合〉

- 正の相関

2 〈仮説間がすべて同じ正の相関である場合〉

3 〈移動平均で偶数番号の変数に 1 を加え、それを偽の仮説と設定したもの〉

4 〈移動平均で後半の番号の変数に 1 を加え、それを偽の仮説と設定したもの〉

- 負の相関のある場合

5 〈移動平均で偶数番号の変数を反転し 1 を加えたものを偽の仮説と設定したもの〉

6 〈移動平均で偶数番号の変数を反転させ、後半の番号の変数に 1 を加え、それを偽の仮説と設定したもの〉

7 〈対比較の設定で p 値を計算したもの[対比較の場合]〉

- 仮説数を拡張したもの

8 〈パターン 3 を偶数番号ではなく、もっと広げた間隔で行ったもの[偽の仮説数は 5 に固定]〉

9 〈パターン 4 を偽の仮説数 5 に固定して、仮説数を増やしたもの〉

10 〈パターン 6 を偽の仮説数 5 に固定して、仮説数を増やしたもの〉

結果のまとめを表 3 に掲げる。これによると BH 法はあらゆる設定で FDR を制御することができている。Benjamini and Hochberg (1995) の証明は無相関のみであるが、どんな相関があろうとも FDR を制御できると考えられる。一方、ABH 法はパターン 2 で正の等相関が 0.5 以上と強い場合や負の相関が混じるパターン 5 の場合に FDR を制御できない。また、BY 法は FDR を常に制御できるもののたいていの場合 FWER さえも守れるほど保守的であることが分かった。

表 3. 様々な設定におけるシミュレーションでの FDR 値

	BH	ABH	BY
最小値	0.00495	0.00542	0.00137
最大値	0.04353	0.11152	0.01088

8.1 TST 法のシミュレーション

いくつかのパターンに対して TST 法のシミュレーションを行ってみて堀内・松田 (2006) の結果と比較してみた。表 4 がその結果である。

選んだパターンは主に ABH 法が FDR 値を制御できず苦手とするものである。(パターン 3 だけは遺伝子解析を意識したものであるが、このパターンは全体的にどの方法でも保守的な結果が出るのが分かっている。) TST 法はそのどの場合でも FDR 値を制御しており、とても有効な方法であることが分かる。

8.2 ま と め

特殊なケースで提案されている FDR を制御する多重検定法を除いて一般的に推奨される方法とはどれであろうか。可能性のあるものを挙げておく。

表 4. TST 法とその他の方法の FDR 値の比較

	BH	ABH	BY	TST
パターン 2: 相関 0.5, 両側, 仮説数 20 (真は 15)	0.03692	0.05463	0.01001	0.04393
パターン 2: 相関 0.9, 両側, 仮説数 20 (真は 15)	0.02669	0.07553	0.00780	0.03500
パターン 2: 相関 0.5, 片側, 仮説数 20 (真は 15)	0.03424	0.09708	0.00899	0.04611
パターン 2: 相関 0.9, 片側, 仮説数 20 (真は 15)	0.02811	0.11152	0.00819	0.03552
パターン 3: 仮説数 20 (真は 10)	0.01473	0.01455	0.00416	0.01570
パターン 5: 仮説数 20 (真は 10)	0.02284	0.08829	0.00629	0.04770

1. BH 法
2. ABH 法
3. $\lambda = 0.5$ の ST 法
4. λ を推定する ST 法
5. TST 法

BH 法のメリットはその万能性である。どのような相関でも耐えうる頑健さがある。しかし、偽の帰無仮説数が多い場合などに少し保守的になる欠点がある。

ABH 法は BH 法の欠点を補う方法であるが、強い相関があると q^* を守れない可能性があり、証明もできていない。

$\lambda = 0.5$ とした ST 法は上の 2 つの方法とほぼ同じ程度の手順で行えるが、弱い相関まで q^* を守ることが証明されている。しかし、強い相関がある場合は守れない場合がある。

λ を推定する ST 法は直感的に理解するのは難しいが探索的に定めているので多くの場合有効であると思われる。(強い相関がある場合もある程度の数の仮説があるならば BH 法に近づいて制御することが期待される。)

最後に TST 法であるが, Benjamini et al. (2005) で推奨されており, 今回のシミュレーションでも一部とはいえとても良好な結果が得られたので有用な方法であると言える。

以上の方法のうちどれを推奨するかは、どのように使いたいかにも依存する。第二種の過誤も重要であるが、全体の仮説数が多いならそれほどたくさんの棄却された仮説は要らないことも多いだろう。私ならまず BH 法で求めてそれで十分と思われる仮説数が棄却されたらそれで終わりとし、足りないようなら λ を推定する ST 法、または TST 法で探索するというのが合理的に思われる。FDR を制御する多重検定法は探索的な場面で使うものであるのだからこのように順序をつけても構わないだろう。

9. おわりに

本論文では FDR を制御する多重検定法を説明するために初歩的な概念から順序立てて説明することを試みた。これにより FDR と従来の多重比較との違いや FDR を制御する方法の種類と性質について理解してもらえるように工夫した。また、FDR を制御する方法の比較も行い、それなりの指針を示せたと思う。結果として「FDR とは何か」ということをつかんでいただけたら幸い

である。

謝 辞

本論文は 2008 年度日本計量生物学会年会の予稿を改稿したものである。予稿の作成にあたり特別セッションのオーガナイザーとして重要な議論をしていただいた京都大学の松井茂之准教授、久留米大学の森川敏彦教授を初めとする関係各位に感謝いたします。

また、丁寧に査読していただいた査読者にも感謝します。

参考文献

- Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Statist. Soc. ser.B*, **57**(1), 289-300.
- Benjamini, Y. and Hochberg, Y. (2000). On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. *J. Edu. Behavioral Statistics*, **25**(1), 60-83.
- Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2005). Adaptive Linear Step-up Procedures that control the False Discovery Rate. Unpublished paper,
<http://www.math.tau.ac.il/~ybenja/MyPapers/bkymarch9.pdf>
- Benjamini, Y. and Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *Annals of Statistics*, **29**(4), 1165-1188.
- Hochberg, Y. and Tamhane, A.C. (1987). *Multiple Comparison Procedures*, John Wiley and Sons.
- 堀内賢太郎・松田眞一 (2006). FDR を制御する多重比較法の性能比較. 南山大学紀要アカデミア数理情報編, **6**, 17-30.
(URL: <http://www.seto.nanzan-u.ac.jp/msie/nas/academia/vol.006pdf/06-017-030.pdf>)
- 永田靖・吉田道弘 (1997). 統計的多重比較法の基礎. サイエンティスト社.
- Oehlert, G. W. (2000). Student-Newman-Kuels controls the false discovery rate. *Statistics and Probability Letters*, **46**, 381-383.
- Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Statist. Soc. ser.B*, **64**(3), 479-498.
- Storey, J. D. (2008). Q-Value. <http://genomics.princeton.edu/storeylab/qvalue/>
- Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong Control, Conservative Point Estimation, and Simultaneous Conservative Consistency of False Discovery Rates: a Unified Approach. *J. R. Statist. Soc. ser.B*, **66**(1), 187-205.

付 録

本文中で記した記号に基づき、厳密な意見や補足をここで述べる。

*¹ これは簡便に述べているだけで多重比較を行う場合に必ずしも分散分析のような全体の検定は必要ではない。全体の検定と親和性のある多重比較法もあるが、併用して使うと有意水準を損なうものの方が多い。したがって、多重比較を行う場合は全体の検定を行わない方がよい。逆にどうしても全体の検定を行いたい場合は親和性のある多重比較法を用いるとよい。たとえば、 F 検定に基づく Tukey-Welsch の方法などである。詳細は永田・吉田 (1997) を参照のこと。

*² 正確には単なる 2 群間の t 検定の繰り返しではなく、全群が等分散であると考えて共通分散の推定量を用い、2 節に示した統計量を用いるのが普通である。もし、等分散性がないと考えられるなら個別分散による 2 群間の比較 (Welch の方法) を用いることも可能である。

*³ FWE の定義を「真の仮説のうち、誤って一つでも棄却すること」とし、帰無仮説以外も考えることがある。仮説が帰無仮説のとき、FWE は第 1 種の過誤を指していることになるが、この定義の下では帰無仮説に対する FWE を厳密に type I FWE と呼ぶことがある。永田・吉田 (1997) でもそのように呼んでいる。

なお、「真の帰無仮説」という表現そのものに違和感を持つ人もいるかもしれない。多重検定では帰無仮説は複数あって、神のみぞ知る本当の状態ではそのうちどれかは真でどれかは偽と定まっている。「真の帰無仮説」とは「帰無仮説全体のうち、真であるものの集まり」を指している。それは固定したものであるが、人には未知なのでどの仮説が真であるかは事前には分からない。それに関わらず真である仮説の過誤のみを制御しようとするところに多重検定の難しさがあるのである。(通常の検定では帰無仮説は 1 つなのでそれが真のときのみを考えればよかったのであるが、多重検定では複数の帰無仮説の真偽のパターンすべてを相手にしなければならぬのである。) そのため、真偽の状態が未知のままであっても過誤の確率をパターンに応じてうまく制御できる方法が優れた方法であると言える。

*⁴ 過誤とその確率を区別せず一律に FWE と呼ぶ場合もあることに注意する。(永田・吉田 (1997) では FWE を用いている。FWER という呼び方は FDR と対比させた呼び方である。) また、FWER は仮説の状態ごとに定義できるが、通常どれが真の仮説であるかは分からないため、広い意味で制御するのはその最大値となる。永田・吉田 (1997) ではそれに名前をつけて最大タイプ I FWE (generalized type I FWE) と呼んでいる。海外を含めた多くの多重比較の書籍では区別を明確にしている名前をつけていないことが多い。たとえば、Hochberg and Tamhane (1987) では最大値が制御できいることを「FWE を強く制御する」という表現で説明している。

*⁵ FDR に関しても上記と同じことが言えるのであるが、今のところ広い意味での FDR に別の名称をつけているのは見たことがない。原点である Benjamini and Hochberg (1995) でも明確に区別していないため、5.1 節の定義にも関わらず FDR の制御といえば広い意味での FDR を指すことが多いと思われる。

⁶ BH 法の手順に適用するにはこれでは分かりにくいので、実際に手順を用いる際には「 q^ を $q^*/\hat{\pi}_0(\lambda)$ で置き換える」と考えた方がよい。

*⁷ 配布されているプログラムは ST 法のみのものである。しかし、MST 法とは仮説数が 100 ぐらい以上であればほとんど変わらない。プログラムでは Storey が研究した別の推定法であるスプラインによる λ の推定法も実装しているが、ブートストラップ法とその方法との優劣は明確ではない。例で見たように仮説数が多くない場合はスプラインによる方法の方が安定して優れているかもしれない。

*⁸ 悪性新生物の部位別死亡率に都道府県ごとの違いがあることはよく知られている。これは特に食生活の影響であると思われるので地域ごとに類似した傾向が見られる。したがって、47 都道府県全部で傾向を見ると p 値は小さなものばかりとなる。今回は多重検定法の適用で違いが見られる程度の地域の大きさを考えた結果、この 20 府県となった。

*⁹ BH 法で棄却された 12 府県は次のような特徴がある。

「結腸・直腸」が多く、「肝・胆・膵」が少ない：愛知県
「気管支・肺」が少ない：長野県
「肝・胆・膵」が多い：広島県
「その他」が多い：静岡県
「胃」が多い：富山県、石川県
「胃」、「結腸・直腸」が少なく、「肝・胆・膵」が多い：岐阜県
「気管支・肺」が多い：三重県
「肝・胆・膵」が多く、「気管支・肺」が少ない：山梨県
「胃」が少ない：島根県
「胃」が若干多く、「結腸・直腸」が若干少ない：奈良県
「肝・胆・膵」が若干多い：大阪府

奈良県や大阪府のように特徴が顕著に出ていない府県もあり、7.5 節で述べたように BH 法で棄却されたものでも十分すぎる棄却数で p 値の大きい府県は棄却されたと言ってもすでに明瞭な違いはないのかもしれない。三重県は四日市のコンビナートのせいであろうし、その逆が内陸の長野県、山梨県であろうかと思われるが、その他の府県の原因ははっきりしない。（主に食物の影響であると思われる。酒類や塩分の摂取などが考えられる。）近隣である富山県と石川県が同じ傾向を示していることから地域性はあると思われるが、上記のように各府県の特徴はばらついたものとなり地域性の影響は強いとは言えないことが分かる。

なお、この 20 府県の部位別比率を用いて 47 都道府県で p 値を調べると地域性の大まかな特徴はつかめる。

「胃」は東日本で多く、九州で少ない。
「結腸・直腸」は東日本で多い。
「肝・胆・膵」は西日本で多い。
「気管支・肺」は東海・近畿が多く、コンビナートのある県でも多い。