FDR for infinitely many comparisons in linear models

Peter Westfall, Texas Tech University
Setup

$X_i, i = 1, \ldots, g$, are within-group averages; $X = (X_1, \ldots, X_g)'$.

$s^2$ is the pooled-sample squared standard error from a balanced ANOVA.

Contrasts $c_j'X$ of means are of interest.
Standard ANOVA Hypothesis

\[ H_0 : \mu_i \equiv \mu \]

Decision rule: Reject \( H_0 \) when \( F \geq F_{g-1,\nu,1-\alpha} \), where \( F = \frac{s^2_X}{s^2} \), and where \( s^2_X = \sum_{i=1}^{g} (X_i - \overline{X})^2 / (g - 1) \).
“Follow-up” hypotheses

\[ H_{0j} : \mathbf{c}_j' \mathbf{\mu} = 0 \text{ for contrast vectors } \mathbf{c}_j, j = 1, \ldots, k. \]

Decision Rule: Reject \( H_{0j} \) when \( |t(\mathbf{c}_j)| \geq T_{\nu,1-\alpha/2}, \)

where \( t(\mathbf{c}_j) = \mathbf{c}_j' \mathbf{X} / (s^2 \mathbf{c}_j' \mathbf{c}_j)^{1/2} \)

(if no correction for multiplicity)

Note: My main results assume the data are fixed; i.e., no assumptions.
Multiple Comparisons Procedures

Scheffé’s method: reject $H_{0j}$ when $|t(c_j)| \geq \{(g - 1)F_{g-1,\nu,1-\alpha}\}^{1/2}$.

Bonferroni’s method: Reject $H_{0j}$ when $|t(c_j)| \geq T_{\nu,1-\alpha/(2k)}$.

Exact method: Reject $H_{0j}$ when $|t(c_j)| \geq T_{\alpha}$, where $T_{\alpha}$ is the $1 - \alpha$ quantile of the null distribution of $\max_j |T(c_j)|$.

All control the Familywise Type I error rate (FWE).
Notes on critical values

Let $q_{\text{Method}}$ denote the critical value of a given method.

- For small $k$, $q_{\text{Unadjusted}} \leq q_{\text{Exact}} \leq q_{\text{Bonferroni}} \leq q_{\text{Scheffé}}$

- For large $k$, $q_{\text{Unadjusted}} \leq q_{\text{Exact}} \leq q_{\text{Scheffé}} \leq q_{\text{Bonferroni}}$

- As $k \to \infty$, $q_{\text{Exact}} \to q_{\text{Scheffé}}$ (if the $c_j$ are space-filling)

- Confidence intervals are $c_j'X \pm q_{\text{Method}}(s^2 c_j' c_j)^{1/2}$
Fisher’s Protected LSD

If the global $F$ test rejects $H_0$, then reject $H_{0j}$ when $|t(c_j)| \geq T_{\nu,1-\alpha/2}$.

Notes:

• LSD fails to control the FWE. A counterexample is $(\mu_1, \ldots, \mu_{g-1}, \mu_g) = (0, \ldots, 0, 100)$.

• But gee whiz, it just seemed like such a nice method!
Duncan’s Bayesian (k-ratio) Method

Reject $H_{0j}$ when

$$|t(c_j)| \geq \left( \frac{F}{F-1} \right)^{1/2} t_{\infty},$$

where

$$\frac{\phi(t_{\infty}) + t_{\infty} \Phi(t_{\infty})}{\phi(-t_{\infty}) - t_{\infty} \Phi(-t_{\infty})} = k_1/k_2,$$

where $k_1/k_2$ measures severity of Type I error to Type II error.

When $k_1/k_2 = 50$, $t_{\infty} = 1.485$.

Minimizes Loss, variance component prior.
Three-Decision Loss Functions
Duncan’s Method; k-ratio = 50

Figure 1: Loss Functions
False Discovery Rate

FDR = \( E(V/R) \), where

\[ V = \text{Number of erroneous rejections} \]

\[ R = \text{Number of total rejections (0/0 defined as 0)} \]

FDR-controlling procedures ensure \( \text{FDR} \leq \alpha \).

It’s a frequentist method.
Benjamini-Hochberg (BH) Method

\(p\)-values are

\[ p(c_j) = 1 - 2P(T_\nu > |t(c_j)|). \]

Let \( p(1) \leq \cdots \leq p(k) \) denote the ordered \( p\)-values corresponding to hypotheses \( H_0(1), \ldots, H_0(k) \), and let

\[ j_0 = \begin{cases} 
\max\{j : p(j) \leq (j/k)\alpha\} & \text{if } p(j) \leq (j/k)\alpha \text{ for some } j, \\
0 & \text{otherwise.}
\end{cases} \]

Then BH will reject all \( H_0(j) \) where \( j \leq j_0 \).

BH controls FDR when \( p\)-values are independent or positively dependent.
BH as Fixed Significance Level Method

Let $r = \frac{j_0}{k}$. Genovese and Wasserman (2002) show $r \to \rho \in [0,1)$ as $k \to \infty$ under independence.

$\Rightarrow$ BH is approximately a testing procedure with a fixed critical significance level $\rho \alpha$.

Here $\rho$ depends only on the underlying data generating process, not on $k$.

However, contrasts are not independent!

My main result: For testing all contrasts, the BH critical values do converge, but to data-dependent values that are similar to the Duncan Bayesian critical values.
CDF Representations

Let $t(1) \geq \cdots \geq t(k)$ denote the ordered values of the $|t(c_j)|$, then BH rejects all $H_{0(j)}$ where $t(j) \geq T_{\nu,1-r\alpha/2}$.

Equivalent formulation: Let

$$G_k(t) = \frac{1}{k} \sum_{j=1}^{k} I(|t(c_j)| \leq t), \quad \text{and}$$

$$H_k(t) = \frac{1}{k} \sum_{j=1}^{k} I(T_{\nu,1-\frac{r\alpha}{k}} \leq t).$$

Then the BH critical value is

$$T_{\nu,1-r\alpha/2} = \begin{cases} \min \{ t : G_k(t) < H_k(t) \}, & \text{if } G_k(t) < H_k(t) \text{ for some } t \\ \infty & \text{otherwise.} \end{cases}$$
Convergence Results

Under independence or weak dependence of the test statistics,

\[ G_k(t) \to_{a.s} G(t) \text{ as } k \to \infty \] by the Glivenko-Cantelli theorem.

Further, typically \( H_k(t) \to H(t) \) (pointwise, non-stochastic) as \( k \to \infty \).

Thus, under independence, or weak dependence, the BH asymptotic critical value is

\[
T_{1-\rho \alpha/2} = \begin{cases} 
\inf \{ t : G(t) < H(t) \}, & \text{if } G(t) < H(t) \text{ for some } t, \\
\infty, & \text{otherwise.}
\end{cases}
\]
Figure 2: (a) $H_k(t)$ and $H(t)$ when $k = 200$, (b) $G(t|F = 17.49)$ and $H(t)$ when $g = 10$, $\nu = 40$, and $\alpha = 0.05$, (c) $G_{200}(t|\mathbf{X},s^2)$ and $H_{200}(t)$ for 200 randomly sampled contrasts, (d) $G_{45}(t|\mathbf{X},s^2)$ and $H_{45}(t)$ for all pairwise contrasts.
Figure 3: $g=4$, $df=16$, $a=.05$
Figure 4: (a) FDRC $\alpha$-thresholds for any significance, compared to $\alpha = 0.05$ for Sheffé’s method. (b) Critical values for FDRC (×) and Sheffé (○).
Randomly Sampled Contrasts

Let $a_1, \ldots, a_{g-1}, a_i \in \mathbb{R}^g$, denote an arbitrary orthonormal basis for $C^\perp(1_g)$, the vector space of contrasts for which $1'_g a = 0$, and let $A = [a_1 | \ldots | a_{g-1}]$. I define a “randomly selected contrast vector” by $c \propto A b$, where $b$ is randomly sampled from the unit sphere in $\mathbb{R}^{g-1}$; such a $b$ may be represented almost surely as $b = z/\|z\|$ where $z \sim N_{g-1}(0, I)$ (e.g., Watson, 1983, p.68). It will be seen that the distribution of $t(c)$ is invariant to choice of o.n. basis matrix $A$, lending credence to my definition of a “randomly sampled contrast vector.”
Main Theorem

For randomly sampled contrasts, $G_k(t|X,s^2) \rightarrow a.s. G(t|F)$

$$= 1 - 2P \left\{ T_{g-2} > \left( \frac{(g-2)t^2}{(g-1)F-t^2} \right)^{1/2} \right\}, 0 < t < \{(g - 1)F\}^{1/2}.$$

Further, $\lim_{k \to \infty} H_k(t) = H(t)$

$$= 1 - \frac{2}{\alpha} P(T_\nu > t), \text{if } t \geq T_{\nu,1-\alpha/2}.$$

Thus, for fixed $F$, the critical value for BH when testing all contrasts is

$$T_{g,\nu,\alpha,\infty}(F) = \begin{cases} \inf\{t : G(t|F) < H(t)\}, & \text{if } G(t|F) < H(t) \text{ for some } t, \\ \infty, & \text{otherwise.} \end{cases}$$
Some Details of the Proof

For fixed \((X,s^2)\), \(t(c) = c'X/(s^2c'c)^{1/2}\) is distributed identically as \(z'A'X/(s^2z'z)^{1/2}\). Let \(\mathcal{P}(M)\) denote the unique projection matrix for the column space of a matrix \(M\), specifically \(\mathcal{P}(M) = I - M(M'M)^{-1}M\), and \(\mathcal{P}(A) = I - 1_g1'_g/g\). Note that, conditional on \(X\),

\[
\frac{z'\mathcal{P}(A'X)z}{\{z'z - z'\mathcal{P}(A'X)z\}z/(g - 2)} \sim F_{1,g-2},
\]

and that

\[
\mathcal{P}(A'X) = \frac{A'XX'A}{X'AA'X} = \frac{A'XX'A}{(g - 1)s_X^2}.
\]

Using algebra and the relationship between \(t\) and \(F\) distributions, the result follows.
Connection with Duncan’s Method

For large $F$, $g$, $\nu$, and small $\alpha$, $T_{g,\nu,\alpha,\infty}(F) \approx$

\[
\left( \frac{F}{F-1} \right)^{1/2} \left\{ \frac{Z_{\alpha/2}^2 + \ln(Z_{\alpha/2}^2)}{\ln(\pi/2) - \ln(F)} \right\}^{1/2}.
\]

Recall that Duncan’s critical value is $\left( \frac{F}{F-1} \right)^{1/2} t_{\infty}$. 
Figure 5: Comparison of FDR and Duncan.
Details of Approximation

\[ H(t) = 1 - (2/\alpha)P(T_\nu \geq t) \approx 1 - (2/\alpha)P(Z \geq t) \text{ for large } \nu \]

\[ G(t|X,s^2) \approx 1 - 2P(Z \geq t/F^{1/2}) \text{ for large } g \]

\[ P(Z \geq t) \approx \exp(-t^2/2)/(t\sqrt{2\pi}) \text{ for large } t. \]
Extension to General Linear Models

General case: $X \sim N_g(\mu, \sigma^2 V)$ for known positive definite $V$. Here, $F = X' A (A' VA)^{-1} A' X / \{ s^2 (g - 1) \}$ is used for testing $H_0 : \mu_i = \mu$.

The main theorem $G_k(t|X, s^2) \to_{a.s} G(t|F)$ still holds but requires a “randomly sampled contrast” to be defined as $c \propto A b$, with $b = z / ||z||$, but with $z \sim N_{g-1}(0, V_{\perp}^{-1})$, where $V_{\perp} = A' VA$. 
Generalized Randomly Sampled Contrasts

Suppose

\[ V = \begin{bmatrix} 1/30 & 0 & 0 & 0 \\ 0 & 1/10 & 0 & 0 \\ 0 & 0 & 1/10 & 0 \\ 0 & 0 & 0 & 1/10 \end{bmatrix} \]

The table shows average cosine of the angle of a randomly selected contrast with the corresponding pairwise contrast.

<table>
<thead>
<tr>
<th></th>
<th>( \mu_1 )</th>
<th>( \mu_2 )</th>
<th>( \mu_3 )</th>
<th>( \mu_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_1 )</td>
<td>-</td>
<td>.55</td>
<td>.55</td>
<td>.55</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td>.55</td>
<td>-</td>
<td>.45</td>
<td>.45</td>
</tr>
<tr>
<td>( \mu_3 )</td>
<td>.55</td>
<td>.45</td>
<td>-</td>
<td>.45</td>
</tr>
<tr>
<td>( \mu_4 )</td>
<td>.55</td>
<td>.45</td>
<td>.45</td>
<td>-</td>
</tr>
</tbody>
</table>
Conclusion

A new multiple comparisons procedure has been developed for testing contrasts in linear models.

The method is derived from BH FDR.

Like Duncan’s method, it depends on the $F$ statistic. The method has a Bayesian correspondence, but is derived from purely frequentist principles.

For large $F$, the method reduces to unadjusted $\alpha$-level comparisons.