The False Discovery Rate

- Back to this table:

	H_0 is True	H_0 is False	Total
Reject H_0	V	S	R
Do Not Reject H_0	U	W	$m - R$
Total	m_0	$m - m_0$	m

- The FWER rate focuses on controlling $\Pr(V > 1)$, i.e., the probability of falsely rejecting *any* null hypothesis.

The False Discovery Rate

- Back to this table:

	H_0 is True	H_0 is False	Total
Reject H_0	V	S	R
Do Not Reject H_0	U	W	$m - R$
Total	m_0	$m - m_0$	m

- The FWER rate focuses on controlling $\Pr(V > 1)$, i.e., the probability of falsely rejecting *any* null hypothesis.
- This is a tough ask when m is large! It will cause us to be super conservative (i.e. to very rarely reject).

The False Discovery Rate

- Back to this table:

	H_0 is True	H_0 is False	Total
Reject H_0	V	S	R
Do Not Reject H_0	U	W	$m - R$
Total	m_0	$m - m_0$	m

- The FWER rate focuses on controlling $\Pr(V > 1)$, i.e., the probability of falsely rejecting *any* null hypothesis.
- This is a tough ask when m is large! It will cause us to be super conservative (i.e. to very rarely reject).
- Instead, we can control the *false discovery rate*:

$$\text{FDR} = \text{E}(V/R).$$

Intuition Behind the False Discovery Rate

$$\mathrm{FDR} = \mathrm{E}\left(\frac{V}{R}\right) = \mathrm{E}\left(\frac{\text{number of false rejections}}{\text{total number of rejections}}\right)$$

Intuition Behind the False Discovery Rate

$$\mathrm{FDR} = \mathrm{E}\left(\frac{V}{R}\right) = \mathrm{E}\left(\frac{\text{number of false rejections}}{\text{total number of rejections}}\right)$$

- A scientist conducts a hypothesis test on each of $m = 20,000$ drug candidates.

Intuition Behind the False Discovery Rate

$$\mathrm{FDR} = \mathrm{E}\left(\frac{V}{R}\right) = \mathrm{E}\left(\frac{\text{number of false rejections}}{\text{total number of rejections}}\right)$$

- A scientist conducts a hypothesis test on each of $m = 20,000$ drug candidates.
- She wants to identify a smaller set of promising candidates to investigate further.

Intuition Behind the False Discovery Rate

$$\mathrm{FDR} = \mathrm{E}\left(\frac{V}{R}\right) = \mathrm{E}\left(\frac{\text{number of false rejections}}{\text{total number of rejections}}\right)$$

- A scientist conducts a hypothesis test on each of $m = 20,000$ drug candidates.
- She wants to identify a smaller set of promising candidates to investigate further.
- She wants reassurance that this smaller set is really "promising", i.e. not too many falsely rejected H_0's.

Intuition Behind the False Discovery Rate

$$\text{FDR} = \text{E}\left(\frac{V}{R}\right) = \text{E}\left(\frac{\text{number of false rejections}}{\text{total number of rejections}}\right)$$

- A scientist conducts a hypothesis test on each of $m = 20,000$ drug candidates.
- She wants to identify a smaller set of promising candidates to investigate further.
- She wants reassurance that this smaller set is really "promising", i.e. not too many falsely rejected H_0's.
- FWER controls Pr(at least one false rejection).

Intuition Behind the False Discovery Rate

$$\text{FDR} = \text{E}\left(\frac{V}{R}\right) = \text{E}\left(\frac{\text{number of false rejections}}{\text{total number of rejections}}\right)$$

- A scientist conducts a hypothesis test on each of $m = 20,000$ drug candidates.
- She wants to identify a smaller set of promising candidates to investigate further.
- She wants reassurance that this smaller set is really "promising", i.e. not too many falsely rejected H_0's.
- FWER controls Pr(at least one false rejection).
- FDR controls the fraction of candidates in the smaller set that are really false rejections. This is what she needs!

Benjamini-Hochberg Procedure to Control FDR

Benjamini-Hochberg Procedure to Control FDR

1. Specify q, the level at which to control the FDR.

Benjamini-Hochberg Procedure to Control FDR

1. Specify q, the level at which to control the FDR.
2. Compute p-values $p_1, \ldots, p_m$ for the null hypotheses $H_{01}, \ldots, H_{0m}$.

Benjamini-Hochberg Procedure to Control FDR

1. Specify q, the level at which to control the FDR.
2. Compute p-values $p_1, \ldots, p_m$ for the null hypotheses $H_{01}, \ldots, H_{0m}$.
3. Order the p-values so that $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$.

Benjamini-Hochberg Procedure to Control FDR

1. Specify q, the level at which to control the FDR.
2. Compute p-values $p_1, \ldots, p_m$ for the null hypotheses $H_{01}, \ldots, H_{0m}$.
3. Order the p-values so that $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$.
4. Define $L = \max\left\{j : p_{(j)} < qj/m\right\}$.

Benjamini-Hochberg Procedure to Control FDR

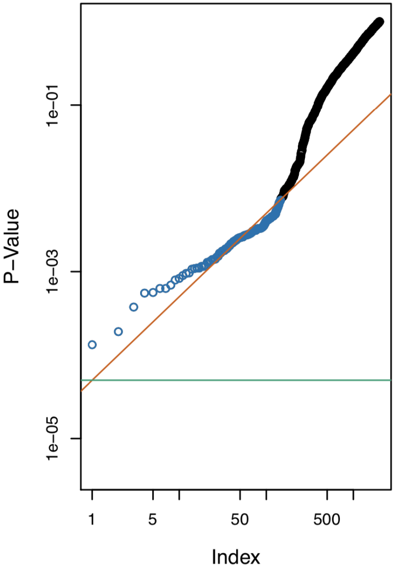
1. Specify q, the level at which to control the FDR.
2. Compute p-values $p_1, \ldots, p_m$ for the null hypotheses $H_{01}, \ldots, H_{0m}$.
3. Order the p-values so that $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$.
4. Define $L = \max\left\{j : p_{(j)} < qj/m\right\}$.
5. Reject all null hypotheses H_{0j} for which $p_j \leq p_{(L)}$.

Benjamini-Hochberg Procedure to Control FDR

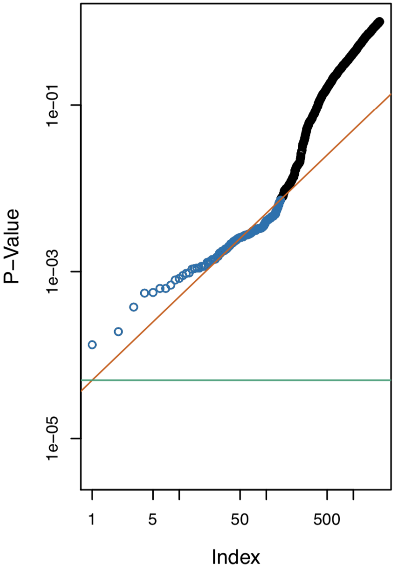
1. Specify q, the level at which to control the FDR.
2. Compute p-values $p_1, \ldots, p_m$ for the null hypotheses $H_{01}, \ldots, H_{0m}$.
3. Order the p-values so that $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$.
4. Define $L = \max\left\{j : p_{(j)} < qj/m\right\}$.
5. Reject all null hypotheses H_{0j} for which $p_j \leq p_{(L)}$.

Then, $\mathrm{FDR} \leq q$.

A Comparison of FDR Versus FWER, Part 1

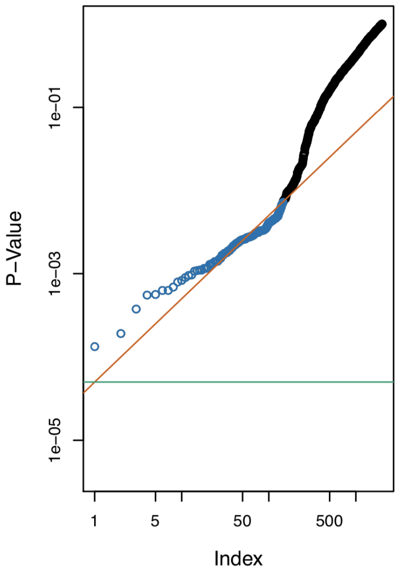


A Comparison of FDR Versus FWER, Part 1



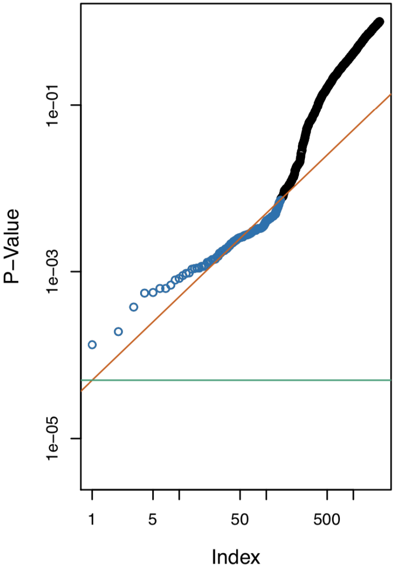
- Here, p-values for $m = 2,000$ null hypotheses are displayed.

A Comparison of FDR Versus FWER, Part 1



- Here, p-values for $m = 2,000$ null hypotheses are displayed.
- To control FWER at level $\alpha = 0.1$ with Bonferroni: reject hypotheses below green line. *(No rejections!)*

A Comparison of FDR Versus FWER, Part 1



- Here, p-values for $m = 2,000$ null hypotheses are displayed.
- To control FWER at level $\alpha = 0.1$ with Bonferroni: reject hypotheses below green line. *(No rejections!)*
- To control FDR at level $q = 0.1$ with Benjamini-Hochberg: reject hypotheses shown in blue.

A Comparison of FDR Versus FWER, Part 2

A Comparison of FDR Versus FWER, Part 2

- Consider $m = 5$ p-values from the *Fund* data:
  $p_1 = 0.006, p_2 = 0.918, p_3 = 0.012, p_4 = 0.601, p_5 = 0.756.$

A Comparison of FDR Versus FWER, Part 2

- Consider $m = 5$ p-values from the *Fund* data:
  $p_1 = 0.006, p_2 = 0.918, p_3 = 0.012, p_4 = 0.601, p_5 = 0.756.$
- Then $p_{(1)} = 0.006$, $p_{(2)} = 0.012$, $p_{(3)} = 0.601$, $p_{(4)} = 0.756$, and $p_{(5)} = 0.918$.

A Comparison of FDR Versus FWER, Part 2

- Consider $m = 5$ p-values from the *Fund* data: $p_1 = 0.006, p_2 = 0.918, p_3 = 0.012, p_4 = 0.601, p_5 = 0.756.$
- Then $p_{(1)} = 0.006$, $p_{(2)} = 0.012$, $p_{(3)} = 0.601$, $p_{(4)} = 0.756$, and $p_{(5)} = 0.918$.
- To control FDR at level $q = 0.05$ using Benjamini-Hochberg:
  - Notice that $p_{(1)} < 0.05/5$, $p_{(2)} < 2 \times 0.05/5$, $p_{(3)} > 3 \times 0.05/5$, $p_{(4)} > 4 \times 0.05/5$, and $p_{(5)} > 5 \times 0.05/5$.
  - So, we reject H_{01} and H_{03}.

A Comparison of FDR Versus FWER, Part 2

- Consider $m = 5$ p-values from the *Fund* data: $p_1 = 0.006, p_2 = 0.918, p_3 = 0.012, p_4 = 0.601, p_5 = 0.756.$
- Then $p_{(1)} = 0.006$, $p_{(2)} = 0.012$, $p_{(3)} = 0.601$, $p_{(4)} = 0.756$, and $p_{(5)} = 0.918$.
- To control FDR at level $q = 0.05$ using Benjamini-Hochberg:
  - Notice that $p_{(1)} < 0.05/5$, $p_{(2)} < 2 \times 0.05/5$, $p_{(3)} > 3 \times 0.05/5$, $p_{(4)} > 4 \times 0.05/5$, and $p_{(5)} > 5 \times 0.05/5$.
  - So, we reject H_{01} and H_{03}.
- To control FWER at level $\alpha = 0.05$ using Bonferroni:
  - We reject any null hypothesis for which the p-value is less than $0.05/5$.
  - So, we reject only H_{01}.