Computational Statistics

26.04: Multiple testing
Multiple testing correction

• See https://xkcd.com/882/

• With many tests at significance level $\alpha$, one is bound to find some significant results, even if all null hypotheses are true (since we allow a false positive probability of $\alpha$ for each test)

• Possible approaches include controlling:
  • Family wise error rate (FWER) (e.g., Bonferroni)
  • False discovery rate (FDR) (e.g., Benjamini Hochberg)
Terminology and notation

For one test:

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is true</th>
<th>$H_a$ is true</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ is not rejected</td>
<td>True negative</td>
<td>False negative or Type II error</td>
</tr>
<tr>
<td>$H_0$ is rejected</td>
<td>False positive or Type I error</td>
<td>True positive or true discovery</td>
</tr>
</tbody>
</table>

$P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ is true}) = \alpha \quad (\text{or } \leq \alpha)$

$P(\text{type II error}) = P(\text{not reject } H_0 | H_a \text{ is true}) = \beta$

Power = $1 - \beta$
Terminology and notation

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is true</th>
<th>$H_\alpha$ is true</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ is not rejected</td>
<td>$U$</td>
<td>$T$</td>
<td>$m - R$</td>
</tr>
<tr>
<td>$H_0$ is rejected</td>
<td>$V$</td>
<td>$S$</td>
<td>$R$</td>
</tr>
<tr>
<td>Total</td>
<td>$m_0$</td>
<td>$m - m_0$</td>
<td>$m$</td>
</tr>
</tbody>
</table>

- **Note**: $m$ is a fixed known number, $m_0$ is a fixed but unknown number. The case when $m = m_0$ is also called the “complete null” or “global null”. All capital letters represent random variables. Only $R$ is observable.

- $Q = V/R$ is the false discovery proportion (FDP)
  (Convention: $V/R = 0$ if $V = R = 0$)

- $E(Q)$ is the false discovery rate (FDR)

- $P(V \geq 1)$ is the family wise error rate (FWER)
Relationships

• One can show that (see board):
  • \( FWER \geq FDR \)
  • \( FWER = FDR \) under the global null
  • \( \alpha \leq FWER \leq \alpha m \)

• The last inequality leads to the Bonferroni correction:
  Conducting each test at significance level \( \alpha/m \) controls the FWER at \( \alpha \).

• Under the global null and if all tests are independent and exactly at level \( \alpha \), then
  \( FWER = 1 - (1 - \alpha)^m \). A first order Taylor expansion around \( \alpha = 0 \) shows that
  this is approximately equal to \( \alpha m \). Hence, in this situation the Bonferroni
  correction is sensible.

• In case of dependent tests, the Bonferroni correction can be much too
  strict/conservative.

• See board.
Westfall Young permutation procedure

- Data matrix of size \( n \times (m + 1) \)
  - Column for y-variable contains 1’s and 0’s to indicate treatment and control
  - Other columns are the x-variables \( X_1, \ldots, X_m \)
- Note: if \( m = m_0 \) (i.e., under the global null), one can permute the y-values
- Procedure:
  - Repeat many times: permute the y-column and do a two sample test (e.g., Wilcoxon), for each \( x_j \)-column (comparing \( x_j[y == 1] \) and \( x_j[y == 0] \)). Let \( p_j, j = 1, \ldots, m \) be the corresponding p-values. Store \( \min(p_1, \ldots, p_m) \).
  - Set \( \delta = \) empirical \( \alpha \)-quantile of the permutation distribution of \( \min(p_1, \ldots, p_m) \).
  - Reject any null hypothesis where the two-sample test on the original data has p-value \( \leq \delta \).
- This procedure provides weak control of the FWER (i.e., under the global null). One can also show strong control (under any configuration of null and alternative hypotheses) under some assumptions.