Permutation Analysis

Benjamin Neale, Michael Neale, Manuel Ferreira
Who came up with permutation?

• Hint: it’s a statistical tool
• R. A. Fisher
• Proposed as validation for Student’s t-test in 1935 in Fisher’s The Design of Experiments
Basic Principle

1. Under the null, all data comes from the same distribution
2. We calculate our statistic, such as mean difference
3. We then shuffle the data with respect to group and recalculate the statistic (mean difference)
4. Repeat step 3 multiple times
5. Find out where our statistic lies in comparison to the null distribution
Real Example

• Case-Control data, and we want to find out if there is a mean difference

<table>
<thead>
<tr>
<th></th>
<th>case</th>
<th>control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.49274</td>
<td>1.471227</td>
</tr>
<tr>
<td>2</td>
<td>-0.30228</td>
<td>0.612679</td>
</tr>
<tr>
<td>3</td>
<td>0.093007</td>
<td>-0.47886</td>
</tr>
<tr>
<td>4</td>
<td>0.715722</td>
<td>0.746045</td>
</tr>
<tr>
<td>5</td>
<td>1.272872</td>
<td>0.871994</td>
</tr>
<tr>
<td>6</td>
<td>-1.37599</td>
<td>0.985237</td>
</tr>
<tr>
<td>7</td>
<td>-0.14798</td>
<td>-0.44421</td>
</tr>
<tr>
<td>8</td>
<td>-1.22195</td>
<td>0.246393</td>
</tr>
</tbody>
</table>

Mean difference: 0.541

Mean: case: -0.01979, control: 0.52144
## Permutation One

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1.2812</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>0.093007</td>
<td>18</td>
</tr>
<tr>
<td>17</td>
<td>0.246393</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>0.985237</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>-0.44421</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>-0.49274</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>-0.14798</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>1.471227</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>0.746045</td>
<td>8</td>
</tr>
</tbody>
</table>

Mean: 0.415354

Mean difference = .329
Simulation example

• I simulated 70 data points from a single distribution—35 cases and 35 controls
• Mean difference of -.21
• I then permuted randomly assigning case or control status
• Empirical significance=#{hits}/#{permutations}
Distribution of mean differences from permutations

![Histogram showing frequency distribution with a peak at -.21]
Distribution of mean differences from permutations
Empirical Significance

- \#hits is any permuted dataset that had a mean difference >.21 or <-.21
- \#permutations is the trials permuted datasets we generate
- Result(\#hits/\#permutations) = 2024/5000 = .4048
- T test results = .3672
General principles of permutation

• Disrupt the relationship being tested
  – Mean difference between group: switch groups
  – Test for linkage in siblings: randomly reassign the ibd sharing
  – If matched control then within pair permute
  – Shaun will describe tests for association
General advantages

• Does not rely on distributional assumptions
• Corrects for hidden selection
• Corrects for hidden correlation
How would we do QTL permutation in Mx?

1. We analyze our real data and record $\chi^2$
2. For sibpairs we shuffle the ibd probabilities for each sibpair
3. We reanalyze the data and record the new $\chi^2$
4. We generate a distribution of $\chi^2$ for the permuted sets
5. Place our statistic on the distribution
6. Repeat for all locations in genome
Some caveats

- Computational time can be a challenge
- Determining what to maintain and what to permute
- Variable pedigrees also pose some difficulties
- Need sufficient data for combinations
- Unnecessary when no bias, but no cost to doing it
Some Exercises

• What would we permute to determine if the MZ correlation is equal to the DZ correlation in a univariate analysis?
• What would we permute to determine if QTL dominance is significant in linkage analysis