FISHER’S EXACT TEST

HOT NOTES FOR STATISTICS

Abstract. Compare two proportions in a two-way table, even when the sample sizes are quite small. Too small for the chi-square test.

1. Explanation and Formulas

Fisher’s exact test for two-way tables uses the hypergeometric distribution to compute an exact p-value for a one- or two-sided test.

In a two-way layout with the form

<table>
<thead>
<tr>
<th>group</th>
<th>no</th>
<th>yes</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$n_{1,1}$</td>
<td>$n_{1,2}$</td>
<td>$n_{1,+}$</td>
</tr>
<tr>
<td>B</td>
<td>$n_{2,1}$</td>
<td>$n_{2,2}$</td>
<td>$n_{2,+}$</td>
</tr>
<tr>
<td>total</td>
<td>$n_{+,1}$</td>
<td>$n_{+,2}$</td>
<td>$n_{+,+}$</td>
</tr>
</tbody>
</table>

we want to compare the two proportions $p_A$ and $p_B$ where

$$\hat{p}_A = \frac{n_{1,2}}{n_{1,+}} \quad \hat{p}_B = \frac{n_{2,2}}{n_{2,+}}$$

Fisher proposed looking at the sum of hypergeometric probabilities of all possible tables where the $n_{+,2}$ values are distributed between groups A and B, and calculating a p-value. So

$$p\text{-value} = \sum_{i=0}^{n_{1,2}} \frac{n_{1,+} \cdot (n_{2,2} - i)}{(n_{+,+} - i)}$$
2. Example

This example comes from J. H. McDonald’s *Handbook of Biological Statistics*:\(^1\)

McDonald and Kreitman (1991) sequenced the alcohol dehydrogenase gene in several individuals of three species of *Drosophila*. Varying sites were classified as synonymous (the nucleotide variation does not change an amino acid) or amino acid replacements, and they were also classified as polymorphic (varying within a species) or fixed differences between species. The two nominal variables are thus synonymicity ("synonymous" or "replacement") and fixity ("polymorphic" or "fixed"). In the absence of natural selection, the ratio of synonymous to replacement sites should be the same for polymorphisms and fixed differences.

<table>
<thead>
<tr>
<th></th>
<th>synonymous</th>
<th>replacement</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>polymorphisms</td>
<td>43</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>fixed</td>
<td>17</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>total</td>
<td>60</td>
<td>9</td>
<td>69</td>
</tr>
</tbody>
</table>

For the research hypothesis comparing \( p_P \), the proportion of synonymous pairs in polymorphic genes, to \( p_F \), the proportion in fixed genes, \( H_a : p_P < p_F \), the p-value is calculated as

\[
\frac{\binom{45}{2}}{\binom{69}{7}} \cdot \frac{\binom{24}{1}}{\binom{69}{8}} + \frac{\binom{45}{1}}{\binom{69}{8}} \cdot \frac{\binom{24}{1}}{\binom{69}{9}} = 0.00605 + 0.00584 = 0.000023 = 0.006653
\]

\(^1\)http://udel.edu/mcdonald/statfishers.html