6. Statistical hypothesis testing

“I can prove anything by statistics except the truth”

George Canning
Biology and statistics wishful thinking

Experiment

Statistics

\[
E[p_{\text{FDR}}(\gamma)] - p_{\text{FDR}}(\gamma) \geq E\left[ \frac{W(\lambda)/(1 - \lambda)\gamma - V(\gamma)}{R(\gamma) \lor 1} \right] \Pr\{R(\gamma) > 0\},
\]

where \( W(\lambda)/(1 - \lambda)\gamma - V(\gamma) \) is a linear non-increasing function of \( R(\gamma) \), and \( E[\gamma]/R(\gamma) \) is a function of \( R(\gamma) \). Thus, by Jensen’s inequality on \( R(\gamma) \) it follows that

\[
\frac{W(\lambda)/(1 - \lambda)\gamma - V(\gamma)}{R(\gamma) \Pr\{R(\gamma) > 0\}} \geq \frac{E[W(\lambda)/(1 - \lambda)\gamma - V(\gamma)|R(\gamma) > 0]}{E[R(\gamma)|R(\gamma) > 0] \Pr\{R(\gamma) > 0\}}
\]

\[
= E[R(\gamma)|R(\gamma) > 0] \Pr\{R(\gamma) > 0\}, \text{it follows that}
\]

\[
\frac{W(\lambda)/(1 - \lambda)\gamma - V(\gamma)|R(\gamma) > 0}{E[R(\gamma)|R(\gamma) > 0] \Pr\{R(\gamma) > 0\}} \geq E[\frac{W(\lambda)/(1 - \lambda)\gamma - V(\gamma)|R(\gamma) > 0}{E[R(\gamma)|R(\gamma) > 0] \Pr\{R(\gamma) > 0\}}]
\]

\[
p < 0.05
\]
P-Values: Misunderstood and Misused

Bertie Vidgen and Taha Yasseri*

The fickle P value generates irreproducible results

Lewis G Halsey, Douglas Curran-Everett, Sarah L Vowler & Gordon B Drummond

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis
Null hypothesis
Null hypothesis

Default position
$H_0$: there is no effect

Evidence against $H_0$

Strong evidence?

no  yes

Position unchanged  Reject $H_0$

Default position
Defendant is innocent

Evidence against

Strong evidence?

no  yes

Innocent  Guilty
Evidence against $H_0$

- Two samples of mice
  - 12 English mice
  - 9 Scottish mice

- Body mass difference:
  \[ \Delta M = M_S - M_E = 5.0 \text{ g} \]

- Two possibilities
  - real difference
  - fluke

- What are the chances of the fluke?

\[ M_E = 19.0 \text{ g} \quad M_S = 24.0 \text{ g} \]
**Gedankenexperiment** under the null hypothesis

One population of British mice

\[ \mu = 20 \text{ g}, \sigma = 5 \]

Select two samples size 12 and 9

\[ \Delta M = M_E - M_S \]

Build distribution of \( \Delta M \)
Gedankenexperiment: result under null hypothesis

Note: in real life we use a statistic with known distribution
Gedankenexperiment: p-value

P-value: probability of getting the observed, or more extreme result, by chance, under $H_0$.

$n = 10^6$

Observation: $n = 12,453$, $p = 0.012$
Null hypothesis and p-value

If both samples were taken from the same population,
then
the probability of observing the difference in mean body mass of 5 g, or more,
by chance (due to random sampling)
would be 1.2%

We observe an effect, but it will occur by chance in 1.2% of repeated experiments (1 in 80)
You have 1.2% chance of making a fool of yourself (if you publish this result)
P-value is the probability of making a fool of yourself
Why “more extreme”?

\[ P(X = 5) = 0 \]

Zero probability
Null hypothesis: reject or what?

Default position: $H_0$  
All other assumptions must be true!

- absorption of evidence is not evidence of absence!
- evidence too weak?

Strong evidence?  

Reductio ad absurdum

- data are incompatible with $H_0$...
- ...or any of the other assumptions
- reject $H_0$ at your own risk
You cannot confirm the null hypothesis

Differential gene expression between WT and a mutant

Genes that are “not different” from 2 replicates...

...are “significantly different” when using 16 replicates

\[ p \geq \alpha \]

\( \times \) No effect

\( \checkmark \) Insufficient evidence

Schurch et al. 2016
You cannot prove the null hypothesis
Statistical model

Null hypothesis

$H_0$: no effect

All other assumptions

Significance level

$\alpha = 0.05$

Data

$p < \alpha$?

yes

Reject $H_0$

(at your own risk)

Effect is real

no

Insufficient evidence
Fisher’s exact test
Ronald Fisher

Sir Ronald Aylmer Fisher (1890-1962)

Rothamsted Experimental Station (Hertfordshire)
The appreciation of tea

Milk first

Tea first
Null hypothesis:
Ms Bristol has no clue
Let’s draw some balls

Urn with 23 balls
13 are white
10 are black

What is the probability of finding exactly 4 white balls?

Draw 5 balls
Binomial coefficient

- “n chose k”

\[
{n \choose k} = \frac{n!}{k!(n-k)!}
\]

- In combinatorics it is the number of possible \( k \)-element subsets of an \( n \)-element set

- From a 5-element set there are 10 possible 3-element subsets

\[
{5 \choose 3} = \frac{5!}{3!2!} = \frac{120}{6 \times 2} = 10
\]
Smaller example

\[
\binom{5}{3} = 10 \quad \text{all combinations}
\]

\[
\begin{array}{c}
\text{1 2 3 4 5} \rightarrow \text{0 0 1}
\end{array}
\]

\[
\binom{3}{2} = 3 \quad \binom{2}{1} = 2 \quad \text{favourable combinations}
\]

\[
\begin{array}{c}
\text{1 2} \times \text{4} = \text{1 2 4}
\end{array}
\]

\[
\begin{array}{c}
\text{1 3} \times \text{5} = \text{1 3 4}
\end{array}
\]

\[
\begin{array}{c}
\text{2 3}
\end{array}
\]

\[
P = \frac{\text{favourable combinations}}{\text{all combinations}} =
\]

\[
= \frac{\binom{2}{1} \times \binom{3}{2}}{\binom{5}{3}} = \frac{6}{10} = 0.6
\]

Draw 3 balls. What is the probability of finding exactly 2 whites among them?
Hypergeometric probability

- \( N = 23 \) balls
- \( m = 13 \) are white
- \( n = 5 \) balls drawn

What is the probability of finding exactly \( k = 4 \) white balls in the draw?

\[
P(X = 4) = \frac{{\binom{13}{4} \cdot \binom{10}{1}}}{{\binom{23}{5}}} \]

Combinations with 3 whites

All combinations

\[
= \frac{715 \times 10}{33,649} = \frac{7,150}{33,649} \approx 0.21
\]
Hypergeometric distribution

- If sums are fixed (blue fields), the cells in the table follow hypergeometric distribution

\[ P_{0}^5 \binom{13}{5} = 0.0075 \]
\[ P_{1}^4 \binom{12}{6} = 0.081 \]
\[ P_{2}^3 \binom{11}{7} = 0.28 \]
\[ P_{3}^2 \binom{10}{8} = 0.38 \]
\[ P_{4}^1 \binom{9}{9} = 0.21 \]
\[ P_{5}^0 \binom{8}{10} = 0.038 \]
One-sided test

- What is the probability of drawing 4 or more white balls?
  \[ P(X \geq 4) = 0.021 + 0.038 = 0.25 \]

- **Enrichment**: do we have more than random? (right-sided test)

- **Depletion**: do we have fewer than random? (left-sided test)

\[ P(X = 4) \]
\[ > \ dhyper(4, 13, 10, 5) \]
\[ [1] \ 0.2124877 \]

\[ P(X > 3) \]
\[ > 1 - \ phyper(3, 13, 10, 5) \]
\[ [1] \ 0.2507355 \]
Tea tasting by Muriel Bristol

Milk first

✅ ✅ ✅ ✗

Tea first

✅ ✅ ✅ ✗
Tea tasting test

- Null hypothesis: Ms Bristol has no ability to tell the difference

- One-sided probability of getting this or more extreme result by chance is

  \[ P(X \geq 3) = 0.229 + 0.014 \approx 0.24 \]

- The null hypothesis cannot be rejected

- Insufficient data!
### 2x2 Contingency Table

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1</th>
<th>Treatment 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>$a$</td>
<td>$b$</td>
<td>$a + b$</td>
</tr>
<tr>
<td>Failure</td>
<td>$c$</td>
<td>$d$</td>
<td>$c + d$</td>
</tr>
<tr>
<td>Total</td>
<td>$a + c$</td>
<td>$b + d$</td>
<td>$a + b + c + d$</td>
</tr>
</tbody>
</table>

Contingency = association
Interpretation 1: test of independence

- $H_0$: variables are independent
- Ms. Bristol’s answers do not depend on whether she got milk or tea first

<table>
<thead>
<tr>
<th>Tea served</th>
<th>T</th>
<th>T</th>
<th>M</th>
<th>M</th>
<th>T</th>
<th>M</th>
<th>T</th>
<th>M</th>
<th>T</th>
<th>M</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms. Bristol</td>
<td>T</td>
<td>M</td>
<td>M</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>M</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

\[
\begin{array}{c|c}
4 & 5 \\
2 & 1 \\
\end{array}
\]

\[ p = 0.58 \]

<table>
<thead>
<tr>
<th>Tea served</th>
<th>T</th>
<th>T</th>
<th>M</th>
<th>M</th>
<th>T</th>
<th>M</th>
<th>M</th>
<th>M</th>
<th>T</th>
<th>T</th>
<th>M</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms. Bristol</td>
<td>T</td>
<td>T</td>
<td>M</td>
<td>T</td>
<td>M</td>
<td>T</td>
<td>M</td>
<td>T</td>
<td>M</td>
<td>T</td>
<td>M</td>
<td>M</td>
</tr>
</tbody>
</table>

\[
\begin{array}{c|c}
5 & 2 \\
0 & 5 \\
\end{array}
\]

\[ p = 0.03 \]
Interpretation 2: test of proportion

- $H_0$: proportions are the same in each column
- $H_0$: proportions are the same in each row

| Tea served | T | T | M | M | T | M | T | T | M | M |
| Ms. Bristol | T | M | M | T | T | T | T | T | M | T | T |
| 4 | 5 | 4:2 | $p = 0.58$ |
| 2 | 1 | 5:1 |

| Tea served | T | T | M | M | T | M | M | T | T | M | M |
| Ms. Bristol | T | T | M | T | M | T | M | T | M | M |
| 5 | 2 | 5:0 | $p = 0.03$ |
| 0 | 5 | 2:5 |
Enrichment analysis

All genes

Gene with GO-term GO:00301174 - regulation of DNA replication initiation

Any other gene

Differentially expressed between conditions

Is our GO-term more frequent in the selection than random?

Is GO-term enriched?
Enrichment example

- There are 668 genes in an experiment
- 7 of them are annotated with GO:00301174
- 44 genes are differentially expressed
- 6 of them have this GO term

- Is it significantly enriched?

$$P(X \geq 6) \approx 4 \times 10^{-7}$$

<table>
<thead>
<tr>
<th></th>
<th>DE</th>
<th>Not DE</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>With GO-term</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Without GO-term</td>
<td>38</td>
<td>623</td>
<td>661</td>
</tr>
<tr>
<td>Total</td>
<td>44</td>
<td>624</td>
<td>668</td>
</tr>
</tbody>
</table>

$$P(X > 5)$$

$$> 1 - \text{phyper}(5, 7, 661, 44)$$

[1] 3.893907e-07
Absolute numbers are important

- A newspaper reports clinical tests on a new cancer drug
- 15% of patients treated with drug A survived
- 30% of patients treated with drug B survived
- So, drug B is 100% better than drug A!
Absolute numbers are important

- A newspaper reports clinical tests on a new cancer drug
- 15% of patients treated with drug A survived
- 30% of patients treated with drug B survived
- So, drug B is 100% better than drug A!
- Actual numbers: 20 and 10 patients
- \( p = 0.37 \) (two-sided test)
Absolute numbers are important

- A newspaper reports clinical tests on a new cancer drug
- 15% of patients treated with drug A survived
- 30% of patients treated with drug B survived
- So, drug B is 100% better than drug A!
- Actual numbers: 20 and 10 patients
- $p = 0.37$

- If we had 80 and 100 patients and the same proportions
- $p = 0.02$

- Moral 1: don’t trust newspapers
- Moral 2: estimate the required size of your sample before you do your experiment
Never, ever use percentages in Fisher’s test!

<table>
<thead>
<tr>
<th></th>
<th>Alive</th>
<th>Dead</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug A</td>
<td>15%</td>
<td>85%</td>
<td></td>
</tr>
<tr>
<td>Drug B</td>
<td>30%</td>
<td>70%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Fisher’s exact test: summary**

<table>
<thead>
<tr>
<th>Input</th>
<th>2×2 contingency table (larger tables possible) typically columns = treatments, rows = outcomes table contains counts counts of subjects falling into categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>Examine if there is an association (contingency) between two variables; whether the proportions in one variable depend on the proportions in the other variable; if there is enrichment</td>
</tr>
<tr>
<td>Null hypothesis</td>
<td>The proportions in one variable do not depend on the proportions in the other variable</td>
</tr>
<tr>
<td>Comments</td>
<td>Exact test – count all possible combinations Use when you have small numbers For large numbers (hundreds) use chi-square test Carefully chose between one- and two-sided test</td>
</tr>
</tbody>
</table>
# Tea tasting

```r
greater = fisher.test(rbind(c(3, 1), c(1, 3)), alternative="greater")
```

Fisher's Exact Test for Count Data

data:  rbind(c(3, 1), c(1, 3))
p-value = 0.2429
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval: 0.3135693 1
sample estimates: odds ratio 6.408309

# GO enrichment

```r
greater = fisher.test(rbind(c(6, 1), c(38, 623)), alternative="greater")
```

Fisher's Exact Test for Count Data

data:  rbind(c(6, 1), c(38, 623))
p-value = 3.894e-07
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval: 14.29724 1
sample estimates: odds ratio 96.29591
Slides available at
https://dag.compbio.dundee.ac.uk/training/Statistics_lectures.html