Seminar in Methodology and Statistics

Fisher's Exact Test

Livi Ruffle & Maria Trofimova
21st March 2005
Outline

• Theory
  • Why use Fisher's Exact Test?
  • Justification of the formula

• Practice
  • Broca's (1 group, 2 questions)
  • Broca's and Wenicke's (2 groups, 1 question)
Why use Fisher's Exact Test?

• Chi-squared test is suitable only when all the cell frequencies are above a lower bound.

• Exact vs. approximate probability distributions.
The derivation

<table>
<thead>
<tr>
<th>Variable Y</th>
<th>Variable X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>d</td>
</tr>
</tbody>
</table>
The derivation

<table>
<thead>
<tr>
<th>Variable Y</th>
<th>Variable X</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
<td>b+d</td>
</tr>
</tbody>
</table>
If we knew only these marginal totals and the overall size of the sample involved, what would the probability be of achieving our result by chance?
The derivation

\[ P = \frac{\text{(number of favorable outcomes)}}{\text{(number of suitable outcomes)}} \]
The derivation

<table>
<thead>
<tr>
<th>Variable Y</th>
<th>Variable X</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
<td>b+d</td>
</tr>
</tbody>
</table>

Number of cases where the marginal totals match for X: \[
\begin{pmatrix}
N \\
(a + c)
\end{pmatrix}
\]

This value is the number of suitable outcomes.
The derivation

So now we have:

\[ P = \frac{\text{(number of favorable outcomes)}}{N} \]

How do we calculate the numerator?
The derivation

Number of cases where the marginal totals match for $X$: $\begin{pmatrix} N \\ a + b \end{pmatrix}$
The derivation

Number of cases where the marginal totals match for X: \( \frac{N}{a + c} \)

Number of cases where \( a \) and \( c \) correlate with Y: \( \frac{a + c}{a} \)
The derivation

Number of cases where the marginal totals match for \( X \):
\[
N = a + c
\]

Number of cases where \( a \) and \( c \) correlate with \( Y \):
\[
a + c
\]

Number of cases where \( b \) and \( d \) correlate with \( Y \):
\[
b + d
\]
The derivation

So out of all the cases where the marginal totals solve for $X$, the ones we want are where $a$, $b$, $c$ and $d$ correlate with $Y$.

Thus:

$$P = \frac{(a + c)(b + d)}{N}$$
The derivation

This value

\[
P = \frac{\begin{pmatrix} a + c \\ a \end{pmatrix} \begin{pmatrix} b + d \\ d \end{pmatrix}}{\begin{pmatrix} N \\ a + b \end{pmatrix}}
\]

is equivalent to that given in Agressi, given a 2x2 table
The derivation

It's also equivalent to:

\[ P(\text{outcome}) = \frac{(a+b)! \ (c+d)! \ (a+c)! \ (b+d)!}{N! \ a! \ b! \ c! \ d!} \]

(try it if you don't believe me)
• Example 1
  – Prepositional case-assignment by Broca’s patients

• Example 2
  – Case-assignment by Broca’s and Wernicke’s patients
Case

- A syntactic notion that relates to a dependency between the constituents in a sentence

- Is assigned to a noun phrase by case-assigners (verbs, prepositions)
Case-assignment

\[ \text{Hij . NOM. geeft een ball aan } \text{hem . ACC.} \]

\[ *\text{Hij . NOM. geeft een ball aan } \text{hij . NOM.} \]

\[ \text{Acc.case} \]

\[ \text{Hij . NOM. zie } \text{haar . ACC.} \]

\[ *\text{Hij . NOM. zie } \text{zij . NOM.} \]
Example 1

Prepositional case-assignment in the free speech of Broca’s patients

- N = 19
- Production of case-assigner (X):
  - 9 – YES, 10 – NO
- Correct case-marking (Y):
  - 9 – YES, 10 - NO
## Contingency Table

<table>
<thead>
<tr>
<th></th>
<th>NO</th>
<th>YES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>NO</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
</tbody>
</table>

\[a+c \quad b+d \quad N\]
Contingency table

<table>
<thead>
<tr>
<th>Case-assigner</th>
<th>Correct case-marking</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>YES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

X

Y
## Contingency table

<table>
<thead>
<tr>
<th>Case-assigner</th>
<th>Correct case-marking</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>YES</td>
<td>2</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>NO</td>
<td>8</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>9</td>
<td>19</td>
</tr>
</tbody>
</table>
The logic of Fisher’s Test

Ho:

There is no association between X (correct case-marking) and Y (production of case-assigner)

The question of statistical significance:

If the Ho were true how likely is it that we may end up with the result this large or larger?
The logic of Fisher’s Test

<table>
<thead>
<tr>
<th></th>
<th>Y Case-assigner</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

Correct case-marking

<table>
<thead>
<tr>
<th></th>
<th>NO</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>NO</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

10 9 19
The logic of Fisher’s Test

<table>
<thead>
<tr>
<th>(\hat{\theta}_1)</th>
<th>(\hat{\theta}_2)</th>
<th>(\hat{\theta}_3)</th>
<th>(\hat{\theta}_4)</th>
<th>(\hat{\theta}_5)</th>
<th>(\hat{\theta}_6)</th>
<th>(\hat{\theta}_7)</th>
<th>(\hat{\theta}_8)</th>
<th>(\hat{\theta}_9)</th>
<th>(\hat{\theta}_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

“this large or larger”
Relative frequency
(Probability)
The logic of Fisher’s Test

1. Figure out the exact probability of each possible outcome “this large or larger”

2. Add up the probabilities

3. Get the result!
Probability of an outcome

\[ P(\text{outcome}) = \frac{(a+b)! \ (c+d)! \ (a+c)! \ (b+d)!}{N! \ a! \ b! \ c! \ d!} \]
Probability of an outcome

\[ P(\hat{O}_{10}) = \frac{9! \ 10! \ 10!9!}{19! \ 0! \ 9! \ 10! \ 0!} = 0.000010825 \]

**NB!** x! - "x factorial

0! = 1
1! = 1
2! = 2\times1 = 2
3! = 3\times2\times1 = 6
4! = 4\times3\times2\times1 = 24
5! = 5\times4\times3\times2\times1 = 120
etc.
Probability of an outcome

\[
P(\hat{O}_9) = \frac{9! \ 10! \ 10!9!}{19! \ 1! \ 8! \ 9! \ 1!} = 0.000974258
\]

\[
P(\hat{O}_8) = \frac{9! \ 10! \ 10!9!}{19! \ 2! \ 7! \ 8! \ 2!} = 0.017536642
\]
Probability of an outcome

The probability of getting the result “this large or larger”

\[ P = P(\hat{O}10) + P(\hat{O}9) + P(\hat{O}8) \]

\[ P = 0.000010825 + 0.000974258 + 0.017536642 = 0.0185 \]
What do we get?

- $P = 0.0185$ is statistically significant
- $H_0$ can be rejected
- $X$ and $Y$ tend to be associated for this particular type of Subjects
Conclusion

The production of correct case-assigner is associated with the realization of correct case-marking in the free speech of Broca’s aphasic patients
Example 2

Syntactic prepositions by Broca’s and Wernicke’s patients

● Groups (Y)
  • Broca’s aphasia - syntactic disorder, \( N_{\text{BROCA'S}} = 5 \)
  • Wernicke’s aphasia - lexical disorder, \( N_{\text{WERNICKE'S}} = 5 \)
  • \( \sum = 10 \)

● Production of syntactic preposition (X):
  • 6 – YES, 4 – NO
## Contingency table

**Production of syntactic preposition**

<table>
<thead>
<tr>
<th></th>
<th>NO</th>
<th>YES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Groups</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wernicke’s</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Broca’s</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><strong>4</strong></td>
<td><strong>6</strong></td>
<td><strong>10</strong></td>
</tr>
</tbody>
</table>
Ho:

There is no association between a type of impairment (Broca’s vs. Wernicke’s) and production of syntactic prepositions

The question of statistical significance:
If the Ho were true how likely is it that we may end up with the result this large or larger?
## Contingency table

Production of syntactic preposition

<table>
<thead>
<tr>
<th>Y Groups</th>
<th>NO</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wernicke’s</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Broca’s</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X</th>
<th>NO</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wernicke’s</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Broca’s</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

Total: 10
The logic of Fisher’s Test

<table>
<thead>
<tr>
<th>Ô1</th>
<th>Ô2</th>
<th>Ô3</th>
<th>Ô4</th>
<th>Ô5</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>1</th>
<th>4</th>
<th>0</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

“this large”
Probability of an outcome

\[ P(\text{outcome}) = \frac{(a+b)! \ (c+d)! \ (a+c)! \ (b+d)!}{N! \ a! \ b! \ c! \ d!} \]

\[ P(\hat{5}) = \frac{5! \ 5! \ 4! \ 6!}{10! \ 0! \ 5! \ 4! \ 1!} = \frac{120 \times 120 \times 27 \times 720}{3628800 \times 1 \times 120 \times 24 \times 1} = 0.0238 \]
Results

- $P = 0.0238$ is statistically significant
- $H_0$ can be rejected
- There is certain association between a type of impairment and a type of linguistic difficulties

Conclusion

Broca’s patients as opposed to Wernicke’s have more problems with syntactic prepositions