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

		Variable X	
		No	Yes
Variable Y	Yes	**a**	**b**
	No	**c**	**d**

The derivation

		Variable X		
		No	Yes	
Variable Y	Yes	**a**	**b**	**a+b**
	No	**c**	**d**	**c+d**
		a+c	**b+d**	**N**

The derivation

		Variable X		
		No	Yes	
Variable Y	Yes			**a+b**
	No			**c+d**
		a+c	**b+d**	**N**

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

$$P = \frac{\textit{(number of favorable outcomes)}}{\textit{(number of suitable outcomes)}}$$

The derivation

		Variable X		
		No	Yes	
Variable Y	Yes	**a**	**b**	**a+b**
	No	**c**	**d**	**c+d**
		a+c	**b+d**	**N**

Number of cases where the marginal totals match for X: $\binom{N}{a+c}$

This value is the number of suitable outcomes.

The derivation

So now we have:

$$P = \frac{(number\ of\ favorable\ outcomes)}{\binom{N}{a+b}}$$

How do we calculate the numerator?

The derivation

		Variable X		
		No	Yes	
Variable Y	Yes	**a**	**b**	**a+b**
	No	**c**	**d**	**c+d**
		a+c	**b+d**	**N**

Number of cases where the marginal totals match for X: $\binom{N}{a+b}$

The derivation

Variable Y		Variable X: No	Variable X: Yes	
	Yes	**a**	**b**	**a+b**
	No	**c**	**d**	**c+d**
		a+c	**b+d**	**N**

Number of cases where the marginal totals match for X: $\binom{N}{a+c}$

Number of cases where **a** and **c** correlate with Y: $\binom{a+c}{a}$

The derivation

Variable Y		Variable X: No	Variable X: Yes	
	Yes	**a**	**b**	**a+b**
	No	**c**	**d**	**c+d**
		a+c	**b+d**	**N**

Number of cases where the marginal totals match for X: $\binom{N}{a+c}$

Number of cases where **a** and **c** correlate with Y: $\binom{a+c}{a}$

Number of cases where **b** and **d** correlate with Y: $\binom{b+d}{d}$

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:

$$P = \frac{\binom{a+c}{a}\binom{b+d}{d}}{\binom{N}{a+b}}$$

The derivation

This value

$$P = \frac{\binom{a+c}{a}\binom{b+d}{d}}{\binom{N}{a+b}}$$

is equevalent 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

Acc.case

Hij .NOM. *geeft een ball* **aan** *hem* .ACC.

Hij* .NOM. *geeft een ball* **aan *hij* .NOM.

Acc.case

Hij .NOM. **zie** *haar* .ACC.

Hij* .NOM. **zie *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

X

		NO	YES	
Y	YES	a	b	a+b
	NO	c	d	c+d
		a+c	b+d	**N**

Contingency table

		X Correct case-marking		
		NO	YES	
Y Case-assigner	YES			9
	NO			10
		10	9	**19**

Contingency table

X

Correct case-marking

Y Case-assigner		NO	YES	
	YES	2	7	9
	NO	8	2	10
		10	9	**19**

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

X

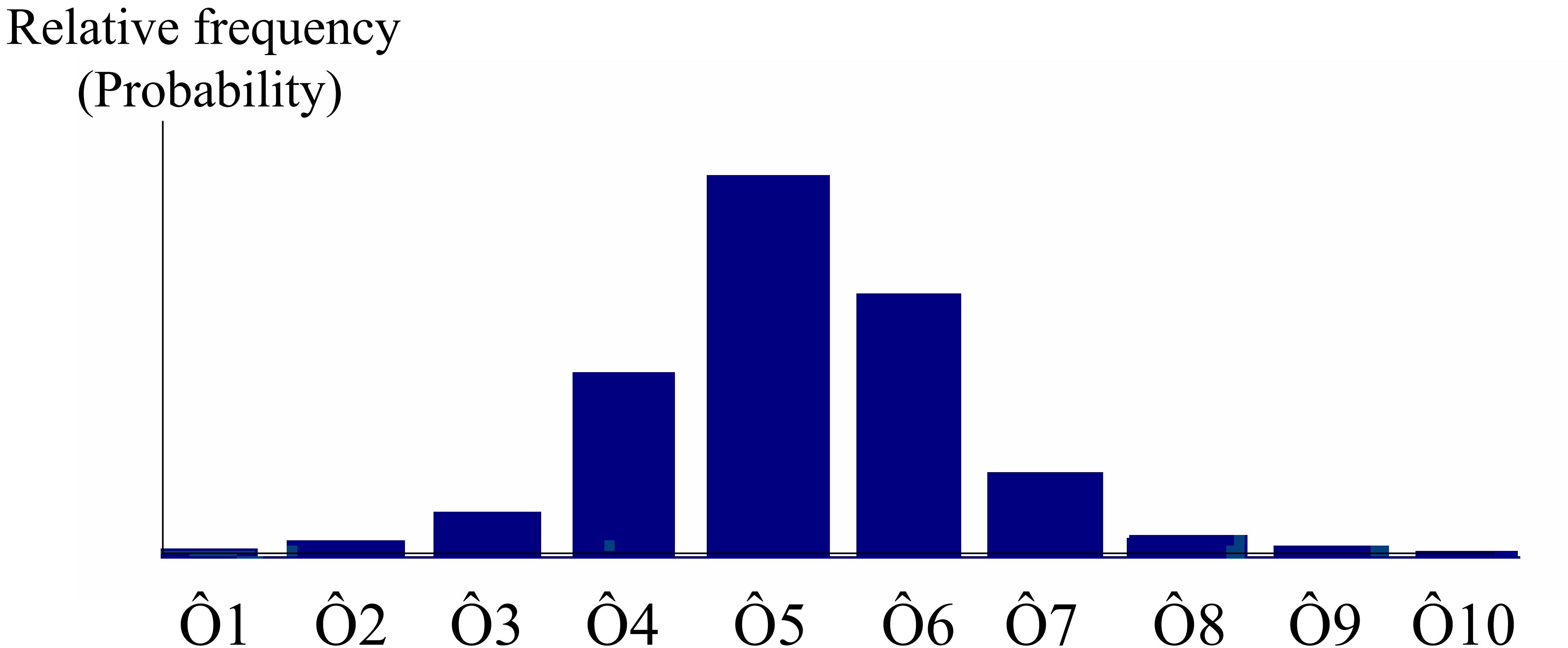
Correct case-marking

		NO	YES	
Y Case-assigner	YES			9
	NO			10
		10	9	**19**

The logic of Fisher’s Test

Ô1	Ô2	Ô3	Ô4	Ô5	Ô6	Ô7	Ô8	Ô9	Ô10
9 0	8 1	7 2	6 3	5 4	4 5	3 6	2 7	1 8	0 9
1 9	2 8	3 7	4 6	5 5	6 4	7 3	8 2	9 1	10 0

“this large or larger”


Relative frequency
(Probability)
Ô1
Ô2
Ô3
Ô4
Ô5
Ô6
Ô7
Ô8
Ô9
Ô10

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

		X		
		NO	YES	
Y	YES	a	b	a+b
	NO	c	d	c+d
		a+c	b+d	N

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

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! = 2x1 = 2

3! = 3x2x1 = 6

4! = 4x3x2x1 = 24

5! = 5x4x3x2x1 = 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(Ô10) + P(Ô9) + P(Ô8)

P = 0.000010825 + 0.000974258 + 0.017536642 = **0.0185**

What do we get?

- P = 0.0185 is statistically significant
- Ho 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, $N_{BROCA'S} = 5$
  - Wernicke's aphasia - lexical disorder, $N_{WERNICKE'S} = 5$
  - $\sum = 10$

- Production of syntactic preposition (X) :
  - 6 – YES, 4 – NO

Contingency table

X

Production of syntactic preposition

Y Groups

	NO	YES	
Wernicke's	0	5	5
Broca's	4	1	5
	4	6	**10**

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

X

Production of syntactic preposition

		NO	YES	
Y Groups	Wernicke's			5
	Broca's			5
		4	6	**10**

The logic of Fisher's Test

Ô1		Ô2		Ô3		Ô4		Ô5	
4	1	3	2	2	3	1	4	0	5
0	5	1	4	2	3	3	2	4	1

"this large"

Probability of an outcome

$$P_{(\text{outcome})} = \frac{(a+b)!\ (c+d)!\ (a+c)!\ (b+d)!}{N!\ a!\ b!\ c!\ d!}$$

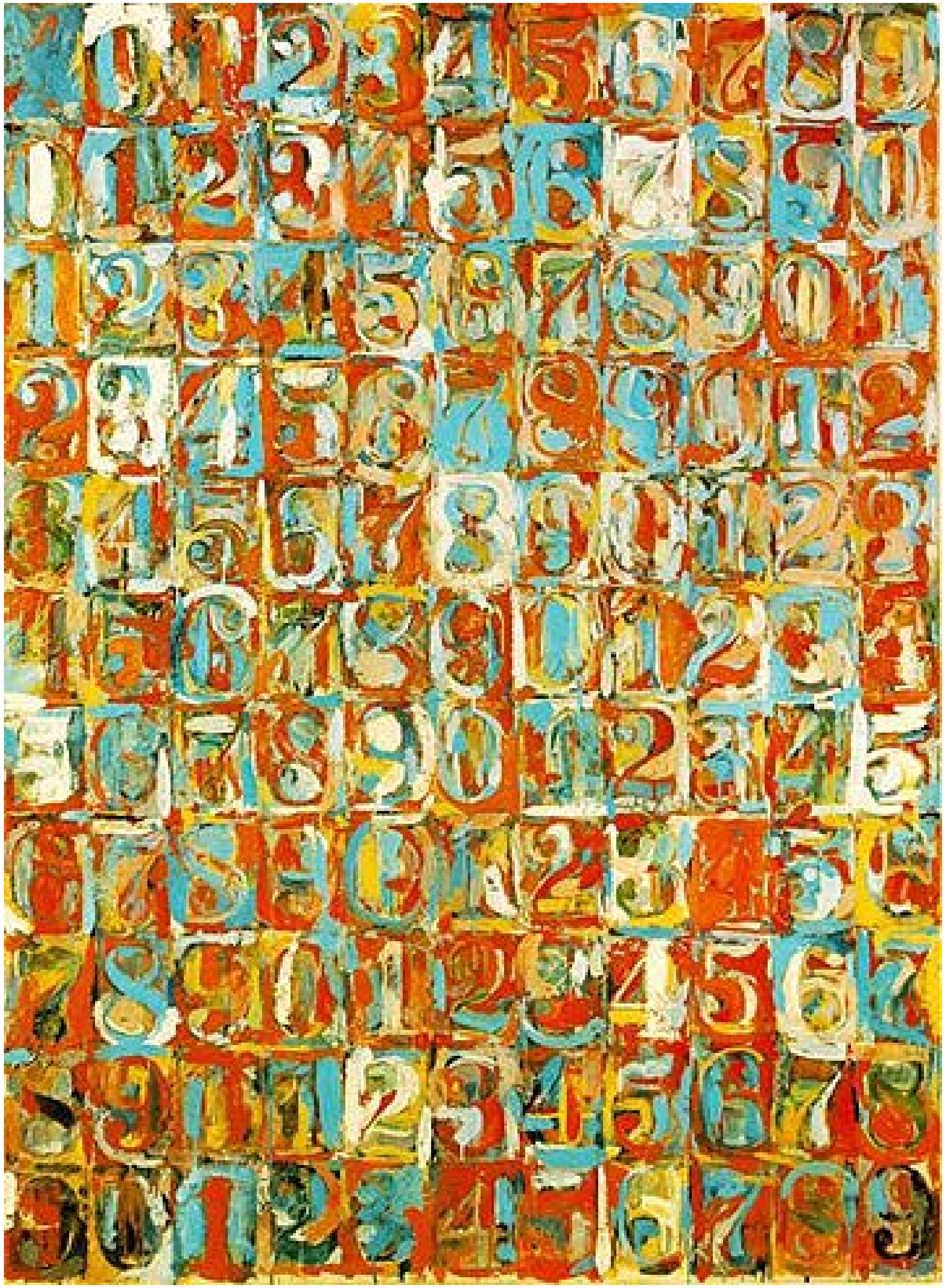
$$\mathbf{P(\hat{O}_5) =} \frac{5!\ 5!\ 4!\ 6!}{10!\ 0!\ 5!\ 4!\ 1!} = \frac{120 * 120 * 27 * 720}{3628800 * 1 * 120 * 24 * 1} = \mathbf{0.0238}$$

Results

- P = 0.0238 is statistically significant
- Ho 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



‘Numbers’ by Jasper Johns