
Exact Tests for 2–Way Tables

Edpsy/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology



Outline

Introduction

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

- Introduction
- Fisher's Exact Test
- Various criteria
- Problems with Exact Tests
- Large tables

Introduction

Introduction

● Introduction

● Example: Imposing Views,
Imposing Shoes

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

- **Problem:** “Sparse ” tables.
- When samples are small, the distributions of X^2 , G^2 , and M^2 are not well approximated by the chi-squared distribution (so p -values for hypothesis tests are not good).
- **Solution:** Perform “exact tests” (or “estimates of exact tests”).
- **2×2 Tables:** The case of small samples and small tables.
- The basic principles are the same for exact tests for larger 2-way tables and higher-way tables (and other cases).

Example: Imposing Views, Imposing Shoes

Alper & Raymond (1995). "Imposing Views, Imposing Shoes: A Statistician as a Sole Model."

Classes were assigned randomly to one of two groups — in the control groups, professors wore ordinary shoes and in the treatment groups, professors wore Nikes. After 3 times/week for 14 weeks, checked to see if students purchased Nikes.

		Students			
		Buy Nikes?			
		Yes	No		
Professor Wore Nikes?	Yes	4	6	10	$\hat{\theta} = .857$
	No	7	9	16	
		11	15	26	

Introduction

● Introduction

● Example: Imposing Views,
Imposing Shoes

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

Fisher's Exact Test

- Fisher's test conditions on the margins of the observed 2×2 table.
- Consider the set of all tables with the exact same margins as the observed table.
- In this set of tables, once you know the value in 1 cell, you can fill in the rest of the cells.
- Nike example: If we know the row totals ($n_{1+} = 10, n_{2+} = 16$), the column totals ($n_{+1} = 10, n_{+2} = 15$), and one cell, say $n_{11} = 4$, then we can fill in the rest.

		Students		
		Buy Nikes?		
Professor	Yes	Yes	No	10
	Wore Nikes?	No	4	
		11	15	26

Introduction

Fisher's Exact Test

- Fisher's Exact Test
- Fisher's Exact Test
- Example: Fisher's Exact Test
- Hypothesis Test that $H_0 : \theta = 1$
- Left Tail Alternative
- Tables that favor H_a
- "Right tail" test, $H_A : \theta > 1$

Different Criteria

Problems with of Exact Tests

Fisher's Exact Test

Introduction

Fisher's Exact Test

- Fisher's Exact Test
- Fisher's Exact Test
- Example: Fisher's Exact Test
- Hypothesis Test that $H_O : \theta = 1$
- Left Tail Alternative
- Tables that favor H_A
- "Right tail" test, $H_A : \theta > 1$

Different Criteria

Problems with of Exact Tests

- Therefore, to find the probability of observing a table, we only need to find the probability of 1 cell in the table (rather than the probabilities of 4 cells).
- Typically, we use the (1, 1) cell, and compute the probabilities that $n_{11} = y$.
- Computing Probabilities of Tables assuming $H_O : \theta = 1$
 - ◆ When $\theta = 1$, the probability distribution of n_{11} (and therefore of the set of tables with fixed margins) is

$$P(n_{11}) = \frac{\binom{n_{1+}}{n_{11}} \binom{n_{2+}}{n_{+1} - n_{11}}}{\binom{n}{n_{+1}}}$$

where

$$\binom{a}{b} = \frac{a!}{b!(a-b)!}$$

"Binomial Coefficient".

- ◆ This probability distribution is "hypergeometric".

Example: Fisher's Exact Test

Introduction

Fisher's Exact Test

- Fisher's Exact Test
- Fisher's Exact Test
- Example: Fisher's Exact Test
- Hypothesis Test that
 $H_O : \theta = 1$
- Left Tail Alternative
- Tables that favor H_a
- "Right tail" test,
 $H_A : \theta > 1$

Different Criteria

Problems with of Exact Tests

		Students		
		Buy Nikes?		
		Yes	No	
Professor Wore Nikes?	Yes	4	6	10
	No	7	9	16
		11	15	26

For the Nike example with $n_{11} = 4$,

$$P(4) = \frac{\binom{10}{4} \binom{16}{7}}{\binom{26}{11}} = \frac{(210)(11,440)}{7,726,160} = .311$$

If $H_O : \theta = 1$ is true, then the probability of observing this particular table given the margins equals .311.

Hypothesis Test that $H_0 : \theta = 1$

- The p -value equals

$p\text{-value} = \sum$ (probabilities of tables that favor H_A , including the probability for the observed table).

- To compute the p -value, we need the alternative H_A .
- $H_A : \theta < 1$ or a “Left tail” test,
 - ◆ Find the odds ratio of the observed table,

$$\theta = n_{11}n_{22}/n_{12}n_{21}$$

- ◆ Compute the probabilities for the tables where the odds ratios are less than odds ratio from the observed table.
- For our example,

$p\text{-value} = \text{sum } P(y)$ for tables with $\theta \leq .857$

Introduction

Fisher's Exact Test

- Fisher's Exact Test
- Fisher's Exact Test
- Example: Fisher's Exact Test
- Hypothesis Test that $H_0 : \theta = 1$
- Left Tail Alternative
- Tables that favor H_A
- “Right tail” test, $H_A : \theta > 1$

Different Criteria

Problems with of Exact Tests

Left Tail Alternative

Introduction

Fisher's Exact Test

- Fisher's Exact Test
- Fisher's Exact Test
- Example: Fisher's Exact Test
- Hypothesis Test that $H_O : \theta = 1$
- Left Tail Alternative
- Tables that favor H_a
- "Right tail" test, $H_A : \theta > 1$

Different Criteria

Problems with of Exact Tests

Left Tail Test hypothesis

$$H_O : \theta = 1 \quad \text{versus} \quad H_A : \theta < 1$$

(1) Find the odds ratio of the observed table,

$$\theta = n_{11}n_{22}/n_{12}n_{21}$$

(2) Compute the probabilities for the tables where the odds ratios are less than odds ratio from the observed table.

For our example,

$$p\text{-value} = \text{sum } P(y) \text{ for tables with } \theta \leq .857$$

Tables that favor H_a

$H_O : \theta = 1$ versus $H_A : \theta < 1$

	yes	no	
yes	4	6	10
no	7	9	16
	11	15	26

$\theta = .857$

$$P(4) = \frac{\binom{10}{4} \binom{16}{7}}{\binom{26}{11}} = .31094$$

	yes	no	
yes	1	9	10
no	10	6	16
	11	15	26

$\theta = .067$

$$P(1) = \frac{\binom{10}{1} \binom{16}{10}}{\binom{26}{11}} = .01037$$

	yes	no	
yes	3	7	10
no	8	8	16
	11	15	26

$\theta = .428$

$$P(3) = \frac{\binom{10}{3} \binom{16}{8}}{\binom{26}{11}} = .19989$$

	yes	no	
yes	0	10	10
no	11	5	16
	11	15	26

$\theta = .000$

$$P(0) = \frac{\binom{10}{0} \binom{16}{11}}{\binom{26}{11}} = .00057$$

	yes	no	
yes	2	8	10
no	9	7	16
	11	15	26

$\theta = .194$

$$P(2) = \frac{\binom{10}{2} \binom{16}{9}}{\binom{26}{11}} = .06663$$

Left tail p -value equals

$$= .31094 + .19989 + .06663 + .01037 + .00057 = .588$$

Introduction

Fisher's Exact Test

- Fisher's Exact Test
- Fisher's Exact Test
- Example: Fisher's Exact Test
- Hypothesis Test that $H_O : \theta = 1$
- Left Tail Alternative
- Tables that favor H_a
- "Right tail" test, $H_A : \theta > 1$

Different Criteria

Problems with of Exact Tests

“Right tail” test, $H_A : \theta > 1$

Compute the probabilities for tables where $\hat{\theta} >$ the odds ratio from the observed table. e.g.,

$$p\text{-value} = \text{sum } P(y) \text{ for tables with } \theta \geq .857$$

θ	y	$P(n_{11} = y)$	Left tail p -value	Right tail p -value
.000	0	.000565	.000565	1.000000
.067	1	.010365	.010930	.999435
.194	2	.066631	.077561	.989070
.429	3	.199892	.277453	.922439
.857	4	.310943	.588396	.722547
1.833	5	.261193	.849589	.411604
3.300	6	.118724	.968313	.150411
7.000	7	.028268	.996581	.031687
17.333	8	.003262	.999843	.003419
63.000	9	.000156	.999999	.000157
∞	10	.000001	1.00000	.000001

Introduction

Fisher's Exact Test

● Fisher's Exact Test

● Example: Fisher's Exact Test

● Hypothesis Test that

$H_O : \theta = 1$

● Left Tail Alternative

● Tables that favor H_a

● “Right tail” test,

$H_A : \theta > 1$

Different Criteria

Problems with of Exact Tests

Different Criteria for Two-tail test

For “Two-tail” test, $H_A : \theta \neq 1$, there are 2 main ways to compute p -values for two-tailed tests:

- “Probability Criterion”
- “ X^2 ” Criterion

Probability Criterion:

p -value = sum of probabilities of tables that are no more likely than the observed table.

that is,

$$p\text{-value} = \sum_y P(y) \quad \text{where } P(y) \leq P(n_{11})$$

Introduction

Fisher's Exact Test

Different Criteria

● Different Criteria for Two-tail test

● Probability Criterion

● X^2 Criterion for $H_A : \theta \neq 1$

Problems with of Exact Tests

Probability Criterion

For our example ...

y	$P(n_{11} = y)$	Left tail	Right tail	Two tail
0	.000565	.000565	1.000000	.000722
1	.010365	.010930	.999435	.014349
2	.066631	.077561	.989070	.109248
3	.199892	.277453	.922439	.427864
4	.310943	.588396	.722547	1.000000
5	.261193	.849589	.411604	.689057
6	.118724	.968313	.150411	.227972
7	.028268	.996581	.031687	.042617
8	.003262	.999843	.003419	.003984
9	.000156	.999999	.000157	.000157
10	.000001	1.00000	.000001	.000001

So, for a two-tailed test when $n_{11} = 4$,

$$p\text{-value} = .59 + .41 = 1.00.$$

Introduction

Fisher's Exact Test

Different Criteria

● Different Criteria for Two-tail test

● Probability Criterion

● X^2 Criterion for $H_A : \theta \neq 1$

Problems with of Exact Tests

X^2 Criterion for $H_A : \theta \neq 1$

p -value equals the sum of probabilities of tables whose Pearson's X^2 is at least as large as the value for the observed table.

y	$P(n_{11} = y)$	Left tail	Right tail	Two tail	Pearson's X^2
0	.000565	.000565	1.000000	.000722	11.917
1	.010365	.010930	.999435	.014349	6.949
2	.066631	.077561	.989070	.109248	3.313
3	.199892	.277453	.922439	.427864	1.008
4	.310943	.588396	.722547	1.000000	.035
5	.261193	.849589	.411604	.689057	.394
6	.118724	.968313	.150411	.227972	2.084
7	.028268	.996581	.031687	.042617	5.105
8	.003262	.999843	.003419	.003984	9.458
9	.000156	.999999	.000157	.000157	15.143
10	.000001	1.000000	.000001	.000001	22.159

For $n_{11} = 4$, the two-tailed p -value equals 1.00.

Introduction

Fisher's Exact Test

Different Criteria

- Different Criteria for Two-tail test

- Probability Criterion

- X^2 Criterion for $H_A : \theta \neq 1$

Problems with of Exact Tests

Discreteness of Exact Tests

Introduction

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

- Discreteness of Exact Tests
- Nike Example
- Fisher's Test is Conservative
- Reduce the Conservativeness of Exact Tests
- Admission Scandal Results Revisited
- Conditioning on Both Margins
- Exact Tests for Larger Tables

p -values and Type I Errors

■ Yates Continuity Correction.

- ◆ This is an approximation of the exact p -value.
- ◆ It involves adjusting Pearson's X^2 ; however, since computers can compute exact p -values, no real need for this anymore.

■ Type I Errors.

- ◆ The smaller n , the smaller the number of possible p -values.
- ◆ Since there are only a fairly small number of possible p -values, setting an α level does not work real well.

Nike Example

If

- $H_O : \theta = 1$ is true
- $H_A : \theta > 1$ (i.e., right tail test)
- $\alpha = .05$

Then

- We can never achieve $\alpha = .05$.
- The only time that we can get $p\text{-value} < .05$ is when $n_{11} \geq 7$ (or $\theta \geq 7.00$), and $P(y \geq 7) = .032$.

y	$P(n_{11} = y)$	Left tail	Right tail	Two tail	Pearson's X^2
0	.000565	.000565	1.000000	.000722	11.917
1	.010365	.010930	.999435	.014349	6.949
2	.066631	.077561	.989070	.109248	3.313
3	.199892	.277453	.922439	.427864	1.008
4	.310943	.588396	.722547	1.000000	.035
5	.261193	.849589	.411604	.689057	.394
6	.118724	.968313	.150411	.227972	2.084
7	.028268	.996581	.031687	.042617	5.105
8	.003262	.999843	.003419	.003984	9.458
9	.000156	.999999	.000157	.000157	15.143
10	.000001	1.00000	.000001	.000001	22.159

Introduction

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

- Discreteness of Exact Tests
- Nike Example
- Fisher's Test is Conservative
- Reduce the

Conservativeness of Exact Tests

- Admission Scandal Results Revisited

- Conditioning on Both Margins
- Exact Tests for Larger Tables

Fisher's Test is Conservative

- Consider the expected value of p -values.
- Normally, when H_0 is true, the distribution of p -values is uniform on the interval $(0,1)$; that is,

$$E(p\text{-value}) = .5$$

- For Fisher's test and our the Nike example (and any table with the exact same margins), the expected p -values equals

$$\text{Left tailed test} \quad E(p\text{-value}) = .612$$

$$\text{Right tailed test} \quad E(p\text{-value}) = .612$$

$$\text{Two-tailed test} \quad E(p\text{-value}) = .612$$

- What to do?

Introduction

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

- Discreteness of Exact Tests
- Nike Example
- Fisher's Test is Conservative
- Reduce the Conservativeness of Exact Tests
- Admission Scandal Results Revisited
- Conditioning on Both Margins
- Exact Tests for Larger Tables

Reduce the Conservativeness of Exact Tests

- Use a different definition of p -value: “mid p -value”.
 - ◆ Mid p -value equal half the probability of the observed table plus the probability of more extreme tables.

- ◆ Nike example with $H_A : \theta > 1$,

$$\text{half probability of observed} = .310943/2 = .1554714$$

$$\text{probability of more extreme} = .411604$$

$$\text{mid } p\text{-value} = .155 + .412 = .567$$

which is certainly much smaller than .722 using the other definition of p -value.

- ◆ Mid p -value definition doesn't guarantee that the true Type I error rate is less than desired α .
- Report p -values and treat them as indices of how much evidence you have against H_O .

Introduction

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

● Discreteness of Exact Tests

● Nike Example

● Fisher's Test is Conservative

● Reduce the

Conservativeness of Exact

Tests

● Admission Scandal Results

Revisited

● Conditioning on Both Margins

● Exact Tests for Larger Tables

Admission Scandal Results Revisited

Introduction

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

- Discreteness of Exact Tests
- Nike Example
- Fisher's Test is Conservative
- Reduce the Conservativeness of Exact Tests
- Admission Scandal Results Revisited
- Conditioning on Both Margins
- Exact Tests for Larger Tables

	Admission		Total
	no	yes	
I list	37	123	160
general	8000	18000	26000
Total	8037	18123	26160

Fisher's Test Results:

Fisher's Exact Test

```

-----
Cell(1,1) Frequency (F)           37
Left-sided Pr <= F                0.0206
Right-sided Pr >= F               0.9869

Table Probability (P)              0.0075
Two-sided Pr <= P                  0.0389
    
```

Even the most conservative test comes out significant!

Conditioning on Both Margins

Introduction

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

- Discreteness of Exact Tests
- Nike Example
- Fisher's Test is Conservative
- Reduce the Conservativeness of Exact Tests
- Admission Scandal Results Revisited
- Conditioning on Both Margins
- Exact Tests for Larger Tables

Any other problems with the Nike or Admissions scandal examples and our use of Fisher's test?

Fisher's exact test conditions on both margins, but only 1 margin in the Nike experiment was fixed and nothing was fixed in the Admissions example (maybe total admissions).

There are other exact tests that condition on only 1 margin and on only the total.

There are other exact tests for different situations.

Exact Tests for Larger Tables

Introduction

Fisher's Exact Test

Different Criteria

Problems with of Exact Tests

- Discreteness of Exact Tests
- Nike Example
- Fisher's Test is Conservative
- Reduce the

Conservativeness of Exact Tests

- Admission Scandal Results Revisited

- Conditioning on Both Margins
- Exact Tests for Larger Tables

- SAS/FREQ: By default, Fisher's is computed for 2×2 tables whenever the "CHISQ" options is included in the "TABLES" command,

TABLES profs*student / CHISQ ;

- Exact tests conditioning on both margins can be computed on larger tables by adding the "EXACT" option to the "TABLES" command,

TABLES row*col / EXACT ;

- There is a limit to how large tables can be to use this. The test is not practical (in terms of CPU time) when

$$\frac{n}{(I-1)(J-1)} > 5$$

- An alternative to exact tests. . .

StatXact & other packages use randomization methods to compute approximations of exact p -values.