

Lecture 23: Tests for Independence in Two-Way Tables

Chapter 6.4

1 / 26

Quiz 9

Question: While the results of the controlled experiment suggesting that women are at a disadvantage in science hiring may come as no surprise, what argument is made that this discrimination is not entirely due to overt misogyny? Answer in one sentence.

Answer: Women rated women candidates lower as well, suggesting not so much explicit misogyny, but rather manifestation of subtler prejudices internalized from societal stereotypes.

2 / 26

Quiz 9

Question: Knowing nothing else about the problem (sample sizes, SE, etc), what can we conclude about the difference in means between men and women?

Answer: No, refer to HW8 Question 7. We had two overlapping CIs, but the CI on the difference did not include 0.

3 / 26

Conditions for Chi-Square Test for Goodness-of-Fit

1. **Independence:** Each case is independent of the each other
2. **Sample size/distribution:** We need at least 5 cases in each scenario i.e. each cell in the table
3. **Degrees of freedom:** We need at least $df = 2$, i.e. $k \geq 3$

4 / 26

Today's Example

Google is always tinkering with its search ranking **algorithm**. Say we want to compare the following 3 algorithms:

1. the current version
2. test algorithm 1
3. test algorithm 2

Today's Example

They measure user satisfaction with the results for a particular search with the **new search** variable:

- ▶ no new search: User clicked on a result. Suggests user is satisfied with result.
- ▶ new search: User **did not** click on a result and tried a new related search. Suggests user is dissatisfied with result.

Today's Example

So we have two categorical variables:

- ▶ `algorithm`: current, test 1, or test 2
- ▶ `new search`: yes or no

Are they independent? i.e. independent of which algorithm is used, do we have the same levels of new search?

7 / 26

Today's Example

Say we observed the following results:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	4000	2000	2000	8000
New search	1000	500	500	2000
Total	5000	2500	2500	10000

For all 3 algorithms, there is a new search $\frac{1}{5}$ of the time.

`algorithm` and `new search` are **independent**: regardless of which algorithm used, the proportion of new searches stays the same.

8 / 26

Today's Example

Now say instead we observed the following results:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	4000	2500	1500	8000
New search	1000	0	1000	2000
Total	5000	2500	2500	10000

In this case, algorithm and new search are **not independent**: depending on which algorithm used, the proportion of new searches is different.

9 / 26

Hypothesis Test

We test at the $\alpha = 0.05$ significance level:

H_0 : the algorithms each perform equally well
vs H_A : the algorithms do not perform equally well

i.e. are the categorical variables algorithm and new search independent?

10 / 26

Different Names

The following all refer to the same test: χ^2 test for

- ▶ two-way tables
- ▶ i.e. contingency tables
- ▶ independence of two categorical variables
- ▶ homogeneity: are the algorithms homogeneous in their performance?

11 / 26

Example from Textbook

Let's make the values match the example from the textbook on page 284:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	3511	1749	1818	7078
New search	1489	751	682	2922
Total	5000	2500	2500	10000

12 / 26

Example from Textbook

Before we start, let's make each column reflect a proportion and not a count.

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	0.7022	0.6996	0.7272	0.7078
New search	0.2978	0.3004	0.2728	0.2922
Total	1	1	1	1

If all algorithms performed the same, we'd **expect**

- ▶ **0.7078** for all 3 values in the top row
- ▶ **0.2922** for all 3 values in the bottom row

Are we observing what we expect? i.e. What is the degree of this deviation?

13 / 26

What's Expected

We expect:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search				$7078 = 0.7078 \times 10000$
New search				$2922 = 0.2922 \times 10000$
Total	5000	2500	2500	10000

14 / 26

What's Expected

We expect:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search		$1769.5 = 0.7078 \times 2500$		7078
New search		$730.5 = 0.2922 \times 2500$		2922
Total	5000	2500	2500	10000

15 / 26

What's Expected

We expect:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search		$1769.5 = 0.7078 \times 2500$	1769.5	7078
New search		$730.5 = 0.2922 \times 2500$	730.5	2922
Total	5000	2500	2500	10000

16 / 26

What's Expected

We expect:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	$3539 = 0.7078 \times 5000$	1769.5	1769.5	7078
New search	$1461 = 0.2922 \times 5000$	730.5	730.5	2922
Total	5000	2500	2500	10000

17 / 26

Observed vs. Expected

Expected Counts:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	3539	1769.5	1769.5	7078
New search	1461	730.5	730.5	2922
Total	5000	2500	2500	10000

Observed Counts:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	3511	1749	1818	7078
New search	1489	751	682	2922
Total	5000	2500	2500	10000

18 / 26

Chi-Square Statistic

We compute χ^2 test statistic: for all $i = 1, \dots, 6$ cells

$$\frac{(\text{observed count}_i - \text{expected count}_i)^2}{\text{expected count}_i}$$

$$\text{Row 1, Col 1} = \frac{(3511 - 3539)^2}{3539} = 0.222$$

\vdots

$$\text{Row 2, Col 3} = \frac{(682 - 730.5)^2}{730.5} = 3.220$$

So

$$\begin{aligned}\chi^2 &= 0.222 + 0.237 + \dots + 3.220 \\ &= 6.120\end{aligned}$$

19 / 26

Chi-Square Distribution

We compare this to a χ^2 distribution to get the p-value. What are the degrees of freedom?

$$\begin{aligned}df &= (\# \text{ of rows} - 1) \times (\# \text{ of columns} - 1) \\ &= (R - 1) \times (C - 1) \\ &= (2 - 1) \times (3 - 1) = 2 \text{ in our case}\end{aligned}$$

20 / 26

Chi-Square Distribution

Looking up 6.120 in the χ^2 table on page 412 on the $df = 2$ row, it would be between 0.05 and 0.01. Since our $\alpha = 0.05$, we reject the null hypothesis and accept the alternative that the algorithms do not perform equally well.

i.e. the algorithm and new search categorical variables are independent.

21 / 26

Conditions/Assumptions

Nearly identical to conditions/assumptions for χ^2 tests for goodness-of-fit:

1. **Independence:** Each case is independent of the other
2. **Sample size/distribution:** We need at least 5 cases in each scenario i.e. each cell in the table
3. **Degrees of freedom:** (Different than before) We need $df = (R - 1) \times (C - 1) \geq 2$.

22 / 26

Why Are They Called Degrees of Freedom?

In the case of χ^2 tests, the degrees of freedom is the number of values needed before you specify **all** values in the cells of the table.

23 / 26

Why Are They Called Degrees of Freedom? Rows

Each row has $df = 2$ because if we specify 2 values, all values in the row are specified.

Example:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	X	Y		7078
New search				2922
Total	5000	2500	2500	10000

then the missing value is $7078 - X - Y$.

i.e. the **wiggle room** we have is $C - 1$ two cells

24 / 26

Why Are They Called Degrees of Freedom? Columns

Each column has $df = 1$ because if we specify 1 value, all values in the column are specified.

Example:

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	X			7078
New search				2922
Total	5000	2500	2500	10000

then the missing value is $5000 - X$.

i.e. the [wobble room](#) we have is $R - 1$ one cell

25 / 26

Why Are They Called Degrees of Freedom? Columns

So the overall df is $(C - 1) \times (R - 1)$, in our case $df = 2$.

new search	algorithm			Total
	Current	Test 1	Test 2	
No new search	X	Y		7078
New search				2922
Total	5000	2500	2500	10000

i.e. if we know these two values, we can fill the rest of the table.

26 / 26