

The χ^2 Tests

Contingency Tables

Math 122

Goodness of Fit

- Goodness of Fit tests test claims that observations either do or do not match some claimed distribution.
- H_0 : The observed frequencies match the claimed distribution.
- H_1 : The observed frequencies do not match the claimed distribution.

World Series

- Below is a table listing the number of games needed to win the World Series in recent span of 91 years.
- Test the claim that the actual frequencies match the distribution expected if the teams are equally matched.

These numbers are taken from baseball-almanac.com and only looks at seasons after 1921 (which was the last best-of-9 year).

Number of Games	Actual Count	Theoretical Probability
4	18	$2/16$
5	19	$4/16$
6	20	$5/16$
7	34	$5/16$

Observed	Expected
18	$91 \times 2/16$
19	$91 \times 4/16$
20	$91 \times 5/16$
34	$91 \times 5/16$

H0: The observed counts match the expected frequencies. ← Claim

H1: The observed counts do not match the expected frequencies

Df=

$$4 - 1 = 3$$

P-value: .0446

Formal Conclusion:

Reject H0 / Support H1

Final conclusion:

There is enough evidence to reject

The claim that the # of games required

to win the series matches the theoretical dist.

Does the number of game required to win the series match the distribution expected if the teams are evenly matched?

NO

Does the distribution match the distribution expected if one team is better than the other?

NO

Lines

On the first day of class, I asked students in four classes to draw a line in a square. I had these results:

	Horizontal	Vertical	North East	South East
Geometry	4	1	1	1
Statistics	1	9	17	4
Logic	5	2	6	1
Survey	4	3	5	1



Chantix

Chantix is a drug used as an aid for those who want to stop smoking. The adverse reaction of nausea has been studied in clinical trials. A contingency table summarizes the study below.

	Placebo	Chantix
Nausea	10	30
No Nausea	795	791

Car Size

Below is a table showing the numbers of small, medium, and large cars and cars from America, Europe, and Japan among a sample of cars.

	Small	Medium	Large
America	26	53	36
Europe	19	17	4
Japan	92	54	2

Contingency Tables

- Tables such as these are called **contingency tables**.
- We can use data in contingency tables to test whether or not the rows and columns of the table are dependent.
- We will use probabilities to calculate the expected count in each cell if the rows and columns are independent.
- We then do a Chi-squared test.

Expected Values

If the rows and columns are independent, what is the probability that a randomly selected car is small and made in America?

	Small	Medium	Large
America	26	53	36
Europe	19	17	4
Japan	92	54	2

We need row and column totals.

Expected Values

	Small	Med	Large	Totals
America	26	53	36	115
Europe	19	17	4	40
Japan	92	54	2	148
Totals	137	124	42	303

Expected Values

If the rows and columns are independent what is the probability that a randomly selected car is small and made in America?

	Small	Med	Large	Totals
America				115
Europe				40
Japan				148
Totals	137	124	42	303

$$P(\text{small and America}) = P(\text{small}) \times P(\text{America}) = \frac{137}{303} \times \frac{115}{303}$$

Expected Values

What is the expected number of small American cars in a sample of 303 if the rows and columns are independent?

	Small	Med	Large	Totals
America				115
Europe				40
Japan				148
Totals	137	124	42	303

$$\begin{aligned}\text{Expected small American cars} &= \frac{137}{303} \times \frac{115}{303} \times 303 \\ &= 51.9\end{aligned}$$

Expected Values

	Small	Med	Large	Totals
America	51.9	47.1	16	115
Europe	18.1	16.4	5.5	40
Japan	67	60.5	20.5	148
Totals	137	124	42	303

$$\text{Small/Amer} = \frac{137}{303} \times \frac{115}{303} \times 303 = 51.9$$

$$\text{Small/Europe} = \frac{137}{303} \times \frac{40}{303} \times 303 = 18.1$$

$$\text{Med/Amer} = 47.1$$

$$\text{Med/Eur} = 16.4$$

$$DF = 4$$

Car Sizes

Test the claim that a car's size is dependent on where it was made.

Observed	Small	Medium	Large
America	26	53	36
Europe	19	17	4
Japan	92	54	2

Expected	Small	Medium	Large
America	52.0	47.1	15.9
Europe	18.1	16.4	5.5
Japan	66.9	60.5	20.6

Contingency Table Test

- H_0 : The rows and columns are independent.
- H_1 : The rows and columns are dependent.
- DF is $(\text{number of rows}-1) \times (\text{number of cols}-1)$

Observed	Small	Medium	Large
America	26	53	36
Europe	19	17	4
Japan	92	54	2

Claim: A car's size is dependent on where it was made.

Expected	Small	Medium	Large
America	52.0	47.1	15.9
Europe	18.1	16.4	5.5
Japan	66.9	60.5	20.6

H0: Rows and columns ⁱⁿ dependent.

H1: Rows and columns ~~in~~ dependent

Df= 4

P-value: 0

Formal Conclusion:

Reject H_0 / Support H_1

Final conclusion: Support claim that car size is dependent on where it was made.

Observed	Placebo	Chantix	
Nausea	10	30	40
No Nausea	795	791	1586
	805	821	1626

Claim: Whether or not a patient experiences nausea is dependent on whether or not that patient took Chantix.

Expected	Placebo	Chantix
Nausea	19.8	20.2
No Nausea	785.2	800.8

$$\frac{805}{1626} \times \frac{40}{1626} \times 1626 = 19.8$$

H0: Rows and columns ⁱⁿ dependent.

H1: Rows and columns ~~in~~ dependent.

Df= 1

P-value: .002

Formal Conclusion:

Reject H₀ / Support H₁

Final conclusion: Support Claim

Observed	H	V	NE	SE
Geo	4	1	1	1
Statistics	1	9	17	4
Logic	5	2	6	1
Survey	4	3	5	1

Claim: The type of line that a student drew depends on that student's class.

Expected	H	V	NE	SE
Geo				
Statistics				
Logic				
Survey				

H₀: Rows and columns ⁱⁿ dependent.

H₁: Rows and columns ~~in~~ independent.

Df= 9

P-value: 0.1

Formal Conclusion: Do not reject H₀ / Do not support H₁

Final conclusion: Do not support the claim.