Chi-Squared Tests

χ^2 Goodness of Fit and Independence Tests

Yolymatics Tutorials

www.yolymaticstutorials.com

November 3, 2025

Professional Mathematics Education

Overview

The Chi-Squared Distribution

Definition

The chi-squared statistic is used to test hypotheses about categorical data by comparing observed frequencies with expected frequencies.

Formula:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i = observed frequency, E_i = expected frequency

Properties:

- Always non-negative ($\chi^2 \geq 0$)
- Depends on degrees of freedom (ν)
- Right-skewed distribution
- As ν increases, distribution approaches normal

Hypotheses and Decision Rules

Hypothesis Testing Framework:

- H_0: Null hypothesis (no association/good fit)
- H_1: Alternative hypothesis (association exists/poor fit)

Decision Rule:

- Calculate test statistic: χ^2_{calc}
- Find critical value from tables: $\chi^2_{crit}(\alpha, \nu)$
- If $\chi^2_{calc} > \chi^2_{crit}$: Reject H_0 (significant result)
- If $\chi^2_{calc} \leq \chi^2_{crit}$: Do not reject H_0 (insufficient evidence)

Conditions for Validity:

- All expected frequencies $E_i \geq 5$
- Independent observations
- Random sampling

Goodness of Fit Test

Purpose

Tests whether observed data follows a specified theoretical distribution or ratio.

Degrees of Freedom:

$$\nu = n - 1 - p$$

where n = number of categories, p = number of parameters estimated

Hypotheses:

- H_0: The data fits the specified distribution
- H_1: The data does not fit the specified distribution

Applications:

- Testing fairness of dice/coins
- Genetic ratio verification

Practice Problem 1: Fair Die Test

A die is rolled 300 times with the following results:

Face	1	2	3	4	5	6
Observed	43	49	56	45	52	55

Test whether the die is fair at the 5% significance level.

Problem 1: Working Space

Practice Problem 2: Blood Type Distribution

A medical study claims blood types are distributed as: O(45%), A(40%), B(11%), AB(4%). In a sample of 500 people:

Blood Type	O	A	B	AB
Observed	234	197	52	17

Test this claim at the 1% significance level.

Problem 2: Working Space

Practice Problem 3: Uniform Distribution

Test whether the following data follows a uniform distribution over 5 categories. Use $\alpha = 0.05$.

Category	1	2	3	4	5
Frequency	78	92	84	76	90

Problem 3: Working Space

Practice Problem 4: Binomial Distribution

A coin is tossed 200 times. Test whether it is fair at the 5% level.

Outcome	Heads	Tails
Observed	118	82

Problem 4: Working Space

Practice Problem 5: Day of Week Births

Hospital records show births on different days. Test if births are equally likely on any day ($\alpha = 0.05$).

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Births	84	78	92	88	90	65	63

Problem 5: Working Space

Test of Independence (Contingency Tables)

Purpose

Tests whether two categorical variables are independent or associated.

Degrees of Freedom:

$$\nu = (r-1)(c-1)$$

where r = number of rows, c = number of columns

Expected Frequency Formula:

$$E_{ij} = \frac{(\text{Row}_i \text{ total}) \times (\text{Column}_j \text{ total})}{\text{Grand total}}$$

Hypotheses:

- H_0: The two variables are independent
- H_1: The two variables are not independent (associated)

Practice Problem 6: Gender and Subject Preference

Test for association between gender and subject preference at $\alpha = 0.05$.

	Mathematics	Science	Arts	**Total**
Male	65	55	40	160
Female	45	65	70	180
Total	110	120	110	340

Problem 6: Working Space

Practice Problem 7: Smoking and Disease

Investigate the relationship between smoking status and lung disease. Test at $\alpha = 0.01$.

	Smoker	Non-Smoker	**Total**
Disease Present	78	32	110
No Disease	122	268	390
Total	200	300	500

Problem 7: Working Space

Practice Problem 8: Education Level and Income

Test whether education level is independent of income bracket ($\alpha = 0.05$).

	Low	Medium	High	**Total**
No Degree	85	62	23	170
Bachelor's	45	78	57	180
Postgraduate	20	60	70	150
Total	150	200	150	500

Problem 8: Working Space

Practice Problem 9: Age and Technology Adoption

Test for association between age group and smartphone adoption at $\alpha = 0.05$.

	Smartphone User	Non-User	**Total**
18-35 years	184	16	200
36-55 years	138	62	200
56+ years	78	122	200
Total	400	200	600

Problem 9: Working Space

Practice Problem 10: Treatment Effectiveness

A clinical trial tests two treatments. Determine if treatment type affects outcome ($\alpha = 0.01$).

	Improved	No Change	Worsened	**Total**
Treatment A	82	48	20	150
Treatment B	95	32	23	150
Total	177	80	43	300

Problem 10: Working Space

Chi-Squared Distribution Table

Critical values for common significance levels:

df (ν)	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188

Note: Reject H_0 if $\chi^2_{calc} > \chi^2_{crit}(\alpha, \nu)$

Key Concepts Summary

Chi-Squared Test Statistic

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Goodness of Fit

- Tests fit to distribution
- $\nu = n - 1 - p$
- Single variable
- Compares observed vs theoretical

Test of Independence

- Tests association
- $\nu = (r - 1)(c - 1)$
- Two variables
- Uses contingency tables

Essential Conditions:

- All expected frequencies ≥ 5
- Independent observations
- Random sampling

Thank You

Questions?

Yolymatics Tutorials
www.yolymaticstutorials.com

Professional Mathematics Education