

Chi-Square Tests

AP Statistics ■ Unit 8

AP Statistics



Chi-Square Tests

What we will cover

- 1 The chi-square idea
- 2 Goodness-of-fit test
- 3 Two-way tables
- 4 Homogeneity vs independence
- 5 Choosing a procedure



1 Idea

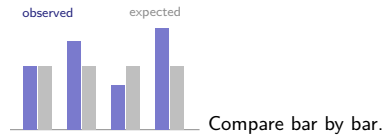
The chi-square idea

The **chi-square** 卡方 tests compare counts:

- **observed** 观测频数 – what happened
- **expected** 期望频数 – what the null predicts

Big gaps $\Rightarrow$ doubt the claim.

Goodness-of-fit 拟合优度 checks one variable; a **two-way table** 双向表 checks two.



2 GOF

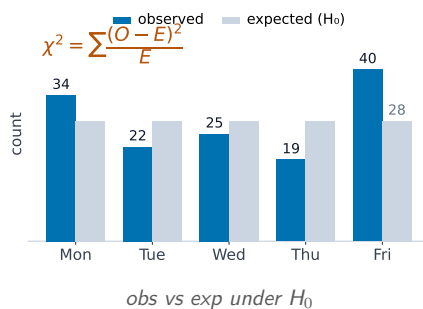
Goodness-of-fit test

Does one variable follow a claimed distribution?

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}}$$

- expected = total $\times$ claimed proportion
- condition: every expected ≥ 5
- $df = \text{categories} - 1$

Large $\chi^2 \Rightarrow$ big gaps; the p-value is the **right-tail** area.



3 Two-way

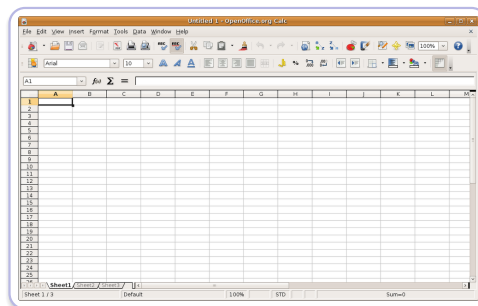
Chi-Square Tests
Two-way

Two-way tables

For a two-way table, compute each **expected count** 期望频数 from the margins (under “no relationship”):

$$\text{exp} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

- every expected ≥ 5
- $df = (\text{rows} - 1)(\text{cols} - 1)$



Organise categorical counts before the test

4 Homogeneity

Homogeneity vs independence

Same χ^2 mechanics – the *design* sets the name:

Homogeneity 齐性
several separate samples, one variable
 H_0 : same distribution across groups

Independence 独立性
one sample, two variables recorded
 H_0 : the two variables are unrelated

Separately survey men and women = homogeneity; one sample recording gender & preference = independence.

5 Choose

Choosing a procedure

Count the samples and variables:

One proportion: one-sample z

Two proportions: two-sample z

Goodness-of-fit: one variable vs a claim

Homogeneity: several samples, one variable

Independence: one sample, two variables

Number of samples 样本数 + **number of variables** 变量数 point straight to the test.

6

Recap

Chi-Square Tests

└ Recap

Worked example – chi-square goodness of fit

A die is rolled 100 times: four faces give 30, 20, 25, 25 over four categories. Test whether it is fair.

- H_0 : all four categories are equally likely $\Rightarrow$ expected = $100/4 = 25$ each.
- $\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{25}{25} + \frac{25}{25} + 0 + 0 = 2.0$
- df = categories – 1 = 3; critical value at $\alpha = 0.05$ is 7.81.
- $2.0 < 7.81 \Rightarrow$ fail to reject H_0 : no evidence the die is unfair.

Use **counts**, never percentages, and check every expected count is ≥ 5 .
“Fail to reject” is **not** “proved fair” – you just lack evidence against it.

Chi-Square Tests

└ Recap

Unit 8 – key takeaways

1 Idea

compare observed to expected counts; right-tail p-value

2 Statistic

$\chi^2 = \sum (O - E)^2 / E$; every expected ≥ 5

3 Tables

expected = row $\times$ col / grand; $df = (r - 1)(c - 1)$

4 Which test

fit, homogeneity, independence by design

Check yourself

- 1 Write the chi-square statistic formula.
- 2 What is the df for a goodness-of-fit with 6 categories?
- 3 How do you find an expected count in a two-way table?
- 4 Several separate samples, one variable: which test?



Answers 1. $\chi^2 = \sum (O - E)^2 / E$ 2. $6 - 1 = 5$ 3. $\text{row} \times \text{col} / \text{grand total}$ 4. homogeneity