

Exercise walk-through

8.7

Skills Focus: Selecting an Appropriate Inference Procedure for Categorical Data

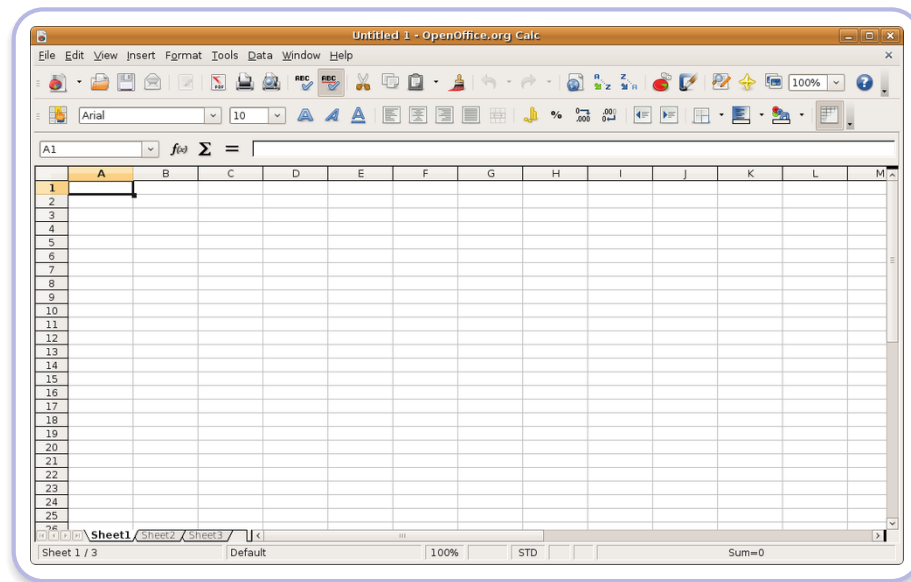
AP Statistics

Two-way tables

For a two-way table, compute each **expected count** 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

Three chi-square tests, one calculation

the shared arithmetic

$\chi^2 = \sum \frac{(O - E)^2}{E}$ in every case – only the question and the df change

goodness-of-fit

one categorical variable against a claimed distribution. $df = k - 1$

test for homogeneity

several populations, one variable – are the distributions the same? Data from separate samples

test for independence

one population, two variables – are they associated? Data from one sample

the two-way df

$(\text{rows} - 1)(\text{columns} - 1)$ for both

the condition

every expected count at least 5

Homogeneity and independence use identical arithmetic and answer different questions. Which you have is decided by **how the data were collected**, not by the table.

Exam tips

- Use $\chi^2 = \sum \frac{(O-E)^2}{E}$ for **categorical** data; always divide by the **expected** count.
- Pick the right test: goodness-of-fit (one variable), independence, or homogeneity (two-way table).
- Compute expected counts as $\frac{\text{row total} \times \text{column total}}{\text{grand total}}$ and check each is ≥ 5 .
- A large χ^2 (small p-value) means observed counts differ from expected by more than chance.
- State degrees of freedom correctly (categories -1 , or $(r-1)(c-1)$).

Exercises 2.1-2.3

2.1 State when to use a goodness-of-fit test.

[1]

2.2 State when to use a test of independence.

[1]

2.3 State the difference between a two-proportion test and a chi-square test of homogeneity.

[1]

Exercises 2.1-2.3

2.1 State when to use a goodness-of-fit test. [1]

2.1

- when checking whether one categorical variable matches a claimed distribution
-

2.2 State when to use a test of independence. [1]

2.2

- when testing whether two categorical variables are associated in one sample
-

2.3 State the difference between a two-proportion test and a chi-square test of homogeneity. [1]

2.3

- a two-proportion test compares two groups on a two-category variable; homogeneity compares a variable with more than two categories across groups

Exercises 3.1-3.2

3.1 To test whether one categorical variable's distribution matches claimed proportions, use [1]

☐

A a t-test

B a goodness-of-fit test

C regression

D a z-interval

3.2 To test the association between two categorical variables in one sample, use [1]

☐

A a goodness-of-fit test

B a test of independence

C a mean test

D a slope test

Exercises 3.1-3.2

3.1 To test whether one categorical variable's distribution matches claimed proportions, use [1]

☐

- A a t-test
- B a goodness-of-fit test
- C regression
- D a z-interval

3.1 B

- a goodness-of-fit test.

3.2 To test the association between two categorical variables in one sample, use [1]

☐

- A a goodness-of-fit test
- B a test of independence
- C a mean test
- D a slope test

3.2 B

- a test of independence.

Exercise 3.3 (a)

3.3 A researcher compares the distribution of blood types across 3 hospitals.

(a) Name the type of data.

[1]

Exercise 3.3 (a)

3.3 A researcher compares the distribution of blood types across 3 hospitals.

(a) Name the type of data.

[1]

(a)

- categorical (counts)

Exercise 3.3 (b)

3.3 A researcher compares the distribution of blood types across 3 hospitals.

(b) Name a suitable chi-square test.

[1]

Exercise 3.3 (b)

3.3 A researcher compares the distribution of blood types across 3 hospitals.

(b) Name a suitable chi-square test.

[1]

(b)

- a chi-square test of homogeneity

Exercise 3.3 (c)

3.3 A researcher compares the distribution of blood types across 3 hospitals.

(c) State the degrees-of-freedom formula for the table.

[1]

Exercise 3.3 (c)

3.3 A researcher compares the distribution of blood types across 3 hospitals.

(c) State the degrees-of-freedom formula for the table.

[1]

(c)

- $(r - 1)(c - 1)$