

Exercise walk-through

8.1

Introducing Statistics: Are My Results Unexpected?

AP Statistics

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.

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.

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. row $\times$ col / grand total 4. homogeneity

Exercises 2.1-2.3

2.1 State what a chi-square test compares.

[1]

2.2 State what “observed counts” are.

[1]

2.3 State what “expected counts” are.

[1]

Exercises 2.1-2.3

2.1 State what a chi-square test compares. [1]

2.1

- observed counts with the counts expected under the null hypothesis
-

2.2 State what “observed counts” are. [1]

2.2

- the counts that actually occur in each category of the data
-

2.3 State what “expected counts” are. [1]

2.3

- the counts predicted in each category if the null hypothesis were true

Exercises 3.1-3.2

3.1 A chi-square test compares observed counts to

[1]

☐

A means

B expected counts

C slopes

D proportions directly

3.2 Chi-square tests are used for _____ data.

[1]

☐

A quantitative

B categorical

C continuous

D paired

Exercises 3.1-3.2

3.1 A chi-square test compares observed counts to

[1]

☐

A means

B expected counts

C slopes

D proportions directly

3.1 B

- expected counts.

3.2 Chi-square tests are used for _____ data.

[1]

☐

A quantitative

B categorical

C continuous

D paired

3.2 B

- categorical.

Exercise 3.3 (a)

3.3 A die is rolled 60 times to test whether it is fair.

(a) Name the kind of data collected.

[1]

Exercise 3.3 (a)

3.3 A die is rolled 60 times to test whether it is fair.

(a) Name the kind of data collected.

[1]

(a)

- categorical (counts of each face)

Exercise 3.3 (b)

3.3 A die is rolled 60 times to test whether it is fair.

(b) State the expected count for each face.

[1]

Exercise 3.3 (b)

3.3 A die is rolled 60 times to test whether it is fair.

(b) State the expected count for each face.

[1]

(b)

- $60 \div 6 = 10$

Exercise 3.3 (c)

3.3 A die is rolled 60 times to test whether it is fair.

(c) Name the test comparing observed to expected counts.

[1]

Exercise 3.3 (c)

3.3 A die is rolled 60 times to test whether it is fair.

(c) Name the test comparing observed to expected counts.

[1]

(c)

- a chi-square (goodness-of-fit) test