

Chi-Square Tests for Two-Way Tables

AP Statistics | Unit 8 | expected counts, χ^2 , independence vs homogeneity

Running example. A random sample of 300 students: phone use in bed and sleep quality. Observed counts, with the expected counts in parentheses:

	Poor	OK	Good	total
Phone in bed	48 (37.5)	62 (60)	40 (52.5)	150
No phone	27 (37.5)	58 (60)	65 (52.5)	150
total	75	120	105	300

1. Expected counts

What the table would look like with the same totals and *no association*:

Expected count of a cell

$$\text{expected} = \frac{\text{row total} \times \text{column total}}{\text{table total}} \quad \text{e.g.} \quad \frac{150 \times 75}{300} = 37.5$$

Expected counts can be decimals; do not round them.

2. The chi-square statistic

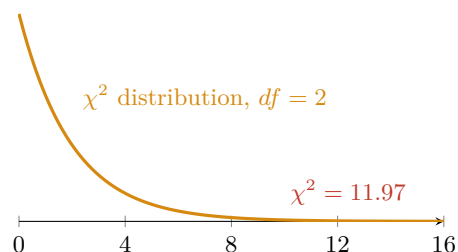
$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}, \quad df = (r-1)(c-1)$$

Cell by cell: $\frac{(48 - 37.5)^2}{37.5} = 2.94$, and so on:

$$\chi^2 = 2.94 + 0.07 + 2.98 + 2.94 + 0.07 + 2.98 \approx 11.97.$$

$df = (2 - 1)(3 - 1) = 2$ (do not count the totals).

The p-value is always the **right tail**: $P(\chi^2 \geq 11.97) \approx 0.0025$.



3. Independence or homogeneity?

Same arithmetic; the design decides the name and the hypotheses (written in words, in context).

Test for independence	Test for homogeneity
One random sample; each individual classified by <i>two</i> variables.	Several samples or treatment groups, compared on <i>one</i> variable.
H_0 : there is no association between the two variables in the population.	H_0 : the distribution of the variable is the same in every population.
H_a : there is an association.	H_a : the distributions are not all the same.

Our data are one sample of 300 students classified two ways: a test for **independence**.

4. Conditions and conclusion

- **Random:** a random sample, independent random samples, or a randomized experiment.
- **10%:** when sampling without replacement, n is at most 10% of the population.
- **Large Counts:** all *expected* counts are at least 5 (here the smallest is 37.5). Show the expected counts.

Conclusion. Because the p-value $0.0025 < 0.05$, we reject H_0 : there is convincing evidence of an association between phone use in bed and sleep quality in the population of students. This is an association, not a cause: the study was observational.

Where students lose points

1. Checking large counts on *observed* counts, or not showing expected counts.
2. df = number of cells $- 1$ instead of $(r - 1)(c - 1)$.
3. Naming the wrong test: independence vs homogeneity.
4. Putting percentages into the formula instead of counts.
5. Concluding “causes”, or hypotheses with no context.

Now practise it

Theory sticks once you use it. On **sad-academy.com** you'll find practice questions for this exact subtopic in the AP format, and an AI that grades your free-response answers against the scoring guidelines, the way an AP reader would.

sad-academy.com

Data are hypothetical. AP[®] is a registered trademark of College Board, which is not affiliated with and does not endorse this material.