AP Statistics Module 3E
Chi-Square Inference

Unit 3: Inference for Categorical Data (Proportions), Sections 3.14, 3.15
Topics: Two-way tables, expected counts, chi-square test for homogeneity and independence

Learning Goals

By the end of this module, you should be able to:

- explain what a chi-square statistic measures;
- distinguish between a chi-square test for **homogeneity** and a chi-square test for **independence**;
- calculate expected counts for a two-way table;
- write appropriate hypotheses, in words and in context, for each chi-square test;
- check the chi-square conditions and determine degrees of freedom;
- interpret a chi-square test statistic and p-value, and state a conclusion in context using appropriate (non-definitive) language.

Scope note. This module covers the *chi-square tests* for two-way tables of counts: the test for **homogeneity** (comparing the distribution of one categorical variable across several independent groups) and the test for **independence** (testing whether two categorical variables recorded on a single sample are associated). These extend the two-proportion comparison of Modules 3C and 3D to categorical variables that may have more than two categories. The chi-square *goodness-of-fit* test, which compares one variable's counts to a claimed distribution, is *not* on the 2026 AP exam; it is noted later only so you recognize it. As always, a test never *proves* the null hypothesis; it measures how surprising the observed counts would be if the null model were true.

Condensed Lecture

What chi-square is really measuring

Chi-square inference is used for **categorical data recorded as counts**. Instead of comparing a sample mean or a sample proportion to a claimed value, we compare:

the counts we observed vs. *the counts we would expect if the null model were true.*

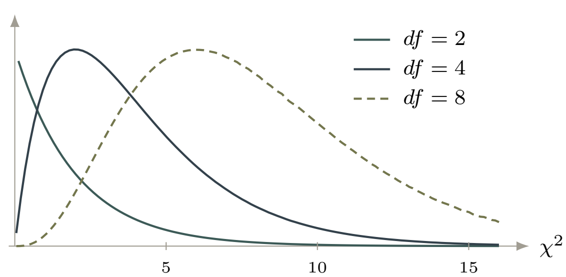
The chi-square statistic measures how far the observed counts are from the expected counts:

$$\chi^2 = \sum \frac{(O-E)^2}{E}.$$

Small differences between observed and expected counts give a small χ^2 statistic. Large differences give a large χ^2 statistic.

The shape of a chi-square distribution

Because $\chi^2 = \sum (O-E)^2/E$ adds up squared quantities, it can never be negative, and its distribution is *right-skewed* rather than symmetric. The only thing that sets its shape is the **degrees of freedom**: as the degrees of freedom grow, the distribution shifts right and spreads out (a larger table tends to produce a larger χ^2 just by chance).



How chi-square fits with earlier inference

The structure is the same as earlier hypothesis tests:

null model → what should happen → what actually happened → is the difference too large?

Test type	What we compare	Test statistic measures
z-test	sample proportion vs. claimed proportion	standardized distance
t-test	sample mean vs. claimed mean	standardized distance
χ^2-test	observed counts vs. expected counts	distance from expected counts

The two chi-square tests

Test	When to use it	Example question
Homogeneity	Two or more *separate* samples or treatment groups. Compare whether the distribution of one categorical variable is the same across the groups.	Do weekday and weekend gym members have the same distribution of primary activity?
Independence	*One* sample classified by *two* categorical variables. Ask whether the two variables are associated.	Is age group associated with whether someone uses a meal-delivery app?

The single most important distinction: count the samples.

Separate samples or treatment groups compared on one response $\Rightarrow$ **homogeneity**.

One sample measured on two variables $\Rightarrow$ **independence**.

Both use the same expected-count formula, the same χ^2 statistic, and the same degrees of freedom; only the design and the wording of the hypotheses differ.

Enrichment (not on the 2026 exam): a third procedure, the **goodness-of-fit** test, compares the counts of one categorical variable to a claimed distribution, using $E = n(\text{claimed proportion})$ and $\text{df} = k - 1$ for k categories. You may meet it in a textbook, but it will not be assessed.

Expected counts (two-way table)

$$E = \frac{(\text{row total})(\text{column total})}{\text{table total}}.$$

Conditions

For a chi-square test for homogeneity or independence, check:

- **Randomization:** data come from a random sample (independence) or a randomized experiment / separate random samples (homogeneity).
- **10% condition:** if sampling without replacement, each sample is no more than 10% of its population.
- **Expected counts:** all *expected* counts (not observed counts) are at least 5.

Degrees of freedom

For a two-way table with r rows and c columns:

$$\text{df} = (r-1)(c-1).$$

Hypotheses

Memory aid — the null matches the test's name.

- **Homogeneity** $\Rightarrow H_0$: the distribution of the response is *the same* across the groups. ("homogeneity" = same)
- **Independence** $\Rightarrow H_0$: the two variables are *independent* (not associated) in the population.

The alternative is always the opposite, and on the AP Exam both must be written *in words and in context*, naming the populations or variables from the problem.

Test	Null hypothesis	Alternative hypothesis
Homogeneity	The distribution of the categorical response is the same across the groups (or treatments).	The distribution of the categorical response is not the same across all of the groups.
Independence	The two categorical variables are independent (not associated) in the population.	The two categorical variables are associated in the population.

Full Lecture

Why we need expected counts

In earlier inference, the null hypothesis told us what value to expect. For example, if $H_0 : p = 0.40$, then a sample proportion near 0.40 is not surprising, while a sample proportion very far from 0.40 may be surprising.

For categorical counts, the same idea applies, but now there may be several categories at once. The null model tells us what counts we would expect in each cell of a two-way table. Then we compare the observed counts to the expected counts.

> A chi-square test is asking:
>
> Are the observed counts too far from the expected counts to be explained by random chance?

Why the formula looks the way it does

The test statistic is

$$\chi^2 = \sum \frac{(O - E)^2}{E}.$$

Each cell contributes

$$\frac{(O - E)^2}{E}.$$

This does three useful things:

- $O - E$ measures the difference between observed and expected.
- Squaring makes all differences positive and makes large differences count more.
- Dividing by E scales the difference relative to the expected size of the cell.

Because the differences are squared, χ^2 is always nonnegative. It is never negative.

Homogeneity vs. independence: same mechanics, different design

Both tests use a two-way table, the same expected-count formula, the same χ^2 statistic, and df $= (r - 1)(c - 1)$. What separates them is *how the data were collected*.

Test for homogeneity. We have two or more *separate* groups — either independent random samples from different populations, or the treatment groups of a randomized experiment — and we record one categorical response. We ask whether the *distribution* of that response is the same across the groups.

Test for independence. We have *one* random sample, and we measure *two* categorical variables on each individual. We ask whether the two variables are associated.

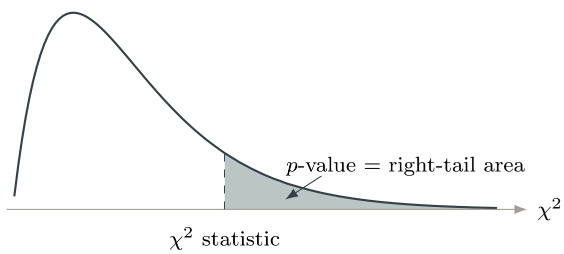
A fast diagnostic: *How many times did someone reach into a population to sample?* Once, recording two variables ⇒ **independence**. Several times, one per group, recording one response ⇒ **homogeneity**.

Interpreting the p-value

A chi-square p-value is the probability, assuming the null hypothesis is true, of getting a chi-square statistic as large as or larger than the one observed.

A large χ^2 statistic gives evidence against the null hypothesis because it means the observed counts are far from the expected counts. Chi-square tests are **right-tailed** tests.

The p-value is the area in the *right* tail of the chi-square distribution beyond your statistic: the probability of a χ^2 at least as large as the one you got, if the null model were true. A bigger statistic sits farther right and leaves a smaller tail, hence a smaller p-value.



Writing the conclusion (AP language)

The AP CED is explicit about conclusions. When you reject H_0, say there *is* convincing evidence of the relationship in the alternative. When you fail to reject H_0, do **not** say the variables are unrelated or that the distributions are equal — that tacitly "accepts" the null. Instead use non-definitive language:

Conclusion template: "Because the p-value is *(less/greater)* than $\alpha = 0.05$, we *(reject/fail to reject)* H_0. There *(is/is not)* convincing statistical evidence that *(restate H_a in context)*."

Finding the p-value on the TI-84 Plus CE

On the AP Exam you are expected to compute the chi-square statistic and the degrees of freedom yourself. The calculator's job here is just to turn that statistic into a p-value. Because a chi-square test is always right-tailed, the p-value is the area to the *right* of your statistic, which the χ^2cdf command gives directly.

The command. Press `2nd` → `VARS` to open the **DISTR** menu, then choose χ^2cdf. The syntax is

$$\chi^2\text{cdf(lower bound, upper bound, df)}.$$

For a chi-square p-value, set:

- **lower bound** = your test statistic χ^2;
- **upper bound** = a very large number for "infinity." Enter `1E99` by typing `1`, then `2nd` → `,` (the `EE` key), then `99`;
- **df** = your degrees of freedom, $(r-1)(c-1)$.

Why the cdf and not the test menu? The TI-84 also has a χ^2-Test (for two-way tables, entered through matrices) and a χ^2GOF-Test. We use χ^2cdf instead so that *you* carry out the expected counts, the statistic, and the degrees of freedom by hand, and let the calculator do only the final area. This mirrors what earns full credit on the AP Exam, where the statistic and df must be shown.

The upper bound is *not* optional. Using χ^2cdf with only two inputs, or putting 0 as the lower bound, will give the area to the *left* of your statistic — the complement of the p-value. For a right-tailed chi-square test, always put the test statistic as the lower bound and a large number such as `1E99` as the upper bound.

Model Problem 1: Chi-Square Test for Homogeneity

A fitness center manager takes *separate* random samples of 100 members who visit on weekdays and 150 members who visit on weekends, and records each member's primary activity. The results are below. Is there convincing evidence that the distribution of primary activity differs between weekday and weekend members? Use $\alpha = 0.05$.

	Cardio	Weights	Classes	Total
Weekday	50	30	20	100
Weekend	60	60	30	150
Total	110	90	50	250

Solution

Use a chi-square test for homogeneity, because two *separate* samples (weekday and weekend members) are compared on one categorical response (primary activity).

$$H_0 : \text{The distribution of primary activity is the same for weekday and weekend members.}$$
$$H_a : \text{The distribution of primary activity is not the same for weekday and weekend members.}$$

$E = \dfrac{\text{(row total)(column total)}}{250}$.

	Cardio	Weights	Classes
Weekday	$\frac{100 \cdot 110}{250} = 44$	$\frac{100 \cdot 90}{250} = 36$	$\frac{100 \cdot 50}{250} = 20$
Weekend	$\frac{150 \cdot 110}{250} = 66$	$\frac{150 \cdot 90}{250} = 54$	$\frac{150 \cdot 50}{250} = 30$

- Randomization: the two samples were taken randomly and independently.
- 10% condition: it is reasonable that each sample is less than 10% of all weekday (resp. weekend) members.
- Expected counts: all expected counts (44, 36, 20, 66, 54, 30) are at least 5.

$$\chi^2 = \frac{(50-44)^2}{44} + \frac{(30-36)^2}{36} + \frac{(20-20)^2}{20} + \frac{(60-66)^2}{66} + \frac{(60-54)^2}{54} + \frac{(30-30)^2}{30}$$
$$\approx 0.818 + 1.000 + 0 + 0.545 + 0.667 + 0 \approx 3.03.$$

$$\text{df} = (2-1)(3-1) = 2.$$

$$p\text{-value} = P(\chi^2 \geq 3.03) \approx 0.22.$$

Because the p-value of about 0.22 is greater than $\alpha = 0.05$, we fail to reject H_0. The data do not provide convincing evidence that the distribution of primary activity differs between weekday and weekend members.

Model Problem 2: Chi-Square Test for Independence

Using the meal-delivery-app table from Model Problem 1 (one random sample of 200 people classified by age group and app use), is there convincing evidence of an association between age group and app use? Use $\alpha = 0.05$.

Solution

Use a chi-square test for independence, because *one* sample is classified by *two* categorical variables.

$$H_0 : \text{Age group and app use are independent (not associated) in the population.}$$
$$H_a : \text{Age group and app use are associated in the population.}$$

$$E_{11} = \frac{120 \cdot 108}{200} = 64.8, \quad E_{12} = \frac{120 \cdot 92}{200} = 55.2$$
$$E_{21} = \frac{80 \cdot 108}{200} = 43.2, \quad E_{22} = \frac{80 \cdot 92}{200} = 36.8$$

The sample was random. The 10% condition is reasonable (200 people is less than 10% of the population). All expected counts are at least 5.

$$\chi^2 = \frac{(72-64.8)^2}{64.8} + \frac{(48-55.2)^2}{55.2} + \frac{(36-43.2)^2}{43.2} + \frac{(44-36.8)^2}{36.8} \approx 4.35.$$

$$\text{df} = (2-1)(2-1) = 1.$$

$$p\text{-value} = P(\chi^2 \geq 4.35) \approx 0.037.$$

Because the p-value of about 0.037 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence of an association between age group and meal-delivery-app use in the population.

Procedure Recognition Practice

For each situation, decide whether a chi-square test for **homogeneity** or a chi-square test for **independence** is appropriate. The key question: were there *separate* samples/groups (homogeneity) or *one* sample measured on two variables (independence)?

a. A researcher takes one random sample of voters and records both political party affiliation and whether each voter supports a new city tax.

   Solution. *Independence.* A single random sample is classified by two categorical variables (party and support), so we test whether those two variables are associated.

b. A school takes separate random samples of ninth, tenth, and eleventh graders and asks each student which type of school lunch they prefer.

   Solution. *Homogeneity.* There are three separate groups (one sample per grade level) compared on one response (lunch preference), so we test whether the distribution of preference is the same across grades.

c. A random sample of adults is asked whether they exercise regularly and whether they usually sleep at least seven hours per night.

   Solution. *Independence.* One random sample is measured on two categorical variables (exercise and sleep), so we test whether they are associated.

d. A company surveys separate random samples of employees in its marketing, finance, and operations departments and asks each whether they prefer in-office, hybrid, or remote work.

   Solution. *Homogeneity.* Separate samples (one per department) are compared on one response (work arrangement), so we test whether the distribution is the same across departments.

e. In a randomized experiment, 150 patients are randomly assigned to a new drug or a placebo, and each patient's outcome (improved, no change, worse) is recorded.

   Solution. *Homogeneity.* The two treatment groups are compared on one categorical response (outcome). Comparing distributions across treatment groups in an experiment uses the test for homogeneity.

f. A guidance office selects one random sample of seniors and records each student's intended path after graduation (work, two-year college, four-year college) and whether the student played a varsity sport.

   Solution. *Independence.* One sample is classified by two categorical variables (intended path and sport participation), so we test for an association.

Guided Problem Solving

Guided 1: Expected counts

In a random sample of 300 men and 200 women, 50 of the men and 40 of the women said that they brush their teeth from side-to-side rather than up-and-down. If there is no difference between the proportions of men and women who brush their teeth from side-to-side, how many men and women in the sample would be expected to brush their teeth from side-to-side?

(A) 45 men and 45 women

(B) 50 men and 40 women

(C) 54 men and 36 women

(D) 150 men and 100 women

(E) 250 men and 160 women

Solution. If there were no difference between men and women, then both groups would say "side-to-side" at the same overall rate. So we first find that single rate using everyone in the sample.

Altogether, $50 + 40 = 90$ of the 500 people said side-to-side. The overall proportion is

$$\frac{90}{500} = 0.18.$$

Now apply that one rate to each group separately:

$$\text{expected men} = 300(0.18) = 54, \qquad \text{expected women} = 200(0.18) = 36.$$

The answer is **(C)**.

Guided 2: Identifying the expected table

A random sample of 200 people was anonymously surveyed as to whether they sneak food into movie theaters. Of the 200 people, 120 were teenagers and 80 were adults; 150 answered yes and 50 answered no. If age group and whether someone sneaks food are independent, which table should we expect?

A	Yes	No	Total
Teenagers	75	45	120
Adults	75	5	80
Total	150	50	200

B	Yes	No	Total
Teenagers	80	40	120
Adults	70	10	80
Total	150	50	200

C	Yes	No	Total
Teenagers	90	30	120
Adults	60	20	80
Total	150	50	200

D	Yes	No	Total
Teenagers	110	10	120
Adults	40	40	80
Total	150	50	200

Solution. If age group and the response are independent, then teenagers and adults should answer "yes" at the same rate — namely the overall "yes" rate for the whole sample. That overall rate is

$$\frac{150}{200} = 0.75.$$

Apply this rate to each group to get the expected "yes" counts, then subtract from the group total to get the expected "no" counts:

$$\text{Teenagers: } 120(0.75) = 90 \text{ yes,} \qquad 120 - 90 = 30 \text{ no,}$$

$$\text{Adults: } 80(0.75) = 60 \text{ yes,} \qquad 80 - 60 = 20 \text{ no.}$$

These are exactly the counts shown in table C, so the answer is **(C)**.

Guided 3: Perfect independence

In the following table, what value of n results in a table showing perfect independence?

	Age under 30	Age 30 and older
Use Twitter	40	60
Do not use Twitter	50	n

(A) 30

(B) 50

(C) 70

(D) 75

(E) 100

Solution. "Perfect independence" means the two age groups use Twitter at the same rate, so the fraction of Twitter users should be equal in each column.

Among people under 30, the proportion who use Twitter is

$$\frac{40}{40+50} = \frac{40}{90}.$$

Among people 30 and older, that proportion is $\dfrac{60}{60+n}$. Setting the two equal,

$$\frac{60}{60+n} = \frac{40}{90}.$$

Cross-multiplying and solving for n:

$$60(90) = 40(60+n) \;\Rightarrow\; 5400 = 2400 + 40n \;\Rightarrow\; 40n = 3000 \;\Rightarrow\; n = 75.$$

The answer is **(D)**.

Guided 4: Reading a two-way table

A random sample of 200 movies were cross-classified by genre and rating.

	G	PG	PG-13	R	Total
Action	5	20	15	5	45
Comedy	10	25	10	5	50
Documentary	25	10	3	2	40
Drama	20	15	17	13	65
Total	60	70	45	25	200

Of the movies in the sample that were classified as PG or PG-13, what proportion were either action or drama?

(A) $\dfrac{45+65}{200}$

(B) $\dfrac{70+45}{200}$

(C) $\dfrac{20+15+15+17}{200}$

(D) $\dfrac{20+15+15+17}{70+45}$

Solution. The phrase "of the movies that were PG or PG-13" tells us to look only at those two columns, so they form the denominator:

$$\text{PG or PG-13 total} = 70 + 45.$$

Within those two columns, count the movies that are action or drama. Action contributes $20 + 15$ and drama contributes $15 + 17$. So the desired proportion is

$$\frac{20+15+15+17}{70+45}.$$

The answer is **(D)**.

Guided 5: Independence using a conditional proportion

For a class project, a student surveyed 125 cars in a high school parking lot, recording whether each car had a student or staff tag and whether each was an American or foreign model. Of the 50 American cars, 20 had staff tags. Given that tag type and model origin are independent, how many of the foreign cars had student tags?

(A) 20

(B) 30

(C) 45

(D) 50

(E) 75

Solution. Start with the American cars, where the breakdown is known. Of the 50 American cars, 20 had staff tags, so the rest had student tags:

$$50 - 20 = 30 \text{ American cars with student tags,} \qquad \frac{30}{50} = 0.60.$$

If tag type and model origin are independent, the foreign cars have student tags at the same rate, 0.60. The number of foreign cars is

$$125 - 50 = 75,$$

so the expected number of foreign cars with student tags is

$$75(0.60) = 45.$$

The answer is **(C)**.

Guided 6: FRQ

A researcher randomly assigns 200 volunteers to use either a new sleep app or no app (control) for one month, then classifies each person's sleep quality as poor, fair, or good. The results are below. Is there convincing evidence that the distribution of sleep quality differs between the two treatments? Use $\alpha = 0.05$.

	Poor	Fair	Good	Total
Sleep app	20	30	50	100
Control	40	35	25	100
Total	60	65	75	200

Solution. Use a chi-square test for homogeneity, because two separate treatment groups (sleep app and control) are compared on one categorical response (sleep quality).

The two groups are the sleep-app treatment and the control treatment. For each group we consider the distribution of sleep quality (poor, fair, good) among all subjects who would receive that treatment.

$$H_0 : \text{The distribution of sleep quality is the same for the sleep-app and control treatments.}$$

$$H_a : \text{The distribution of sleep quality is not the same for the two treatments.}$$

We will use $\alpha = 0.05$.

1. *Random:* the 200 volunteers were randomly assigned to the two treatments.

2. *Independence / 10% Condition:* the random assignment makes the responses independent; because this is a randomized experiment, the 10% condition does not need to be checked.

3. *Large Counts:* all expected counts are at least 5. Because each treatment group has 100 people, each expected count is $\dfrac{\text{(row total)(column total)}}{200}$:

$$E_{\text{Poor}} = \frac{100 \cdot 60}{200} = 30, \qquad E_{\text{Fair}} = \frac{100 \cdot 65}{200} = 32.5, \qquad E_{\text{Good}} = \frac{100 \cdot 75}{200} = 37.5,$$

the same three values for each treatment; all are at least 5.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(20-30)^2}{30} + \frac{(30-32.5)^2}{32.5} + \frac{(50-37.5)^2}{37.5} + \frac{(40-30)^2}{30} + \frac{(35-32.5)^2}{32.5} + \frac{(25-37.5)^2}{37.5} \approx 15.38.$$

$$\text{df} = (2-1)(3-1) = 2.$$

$$p\text{-value} = P(\chi^2 \geq 15.38) = \chi^2\text{cdf}(15.38,\ 1\text{E}99,\ 2) \approx 0.0005.$$

Because the p-value of about 0.0005 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence that the distribution of sleep quality is not the same for the sleep-app and control treatments. Because the volunteers were randomly assigned to treatments, it is reasonable to conclude that the sleep app caused the difference.

Guided 7: FRQ

A city planner takes one random sample of 240 residents and classifies each by age group and whether they support building a new park. The results are below. Is there convincing evidence of an association between age group and support? Use $\alpha = 0.05$.

	Support	Oppose	Total
Under 40	90	30	120
40 and older	70	50	120
Total	160	80	240

Solution. Use a chi-square test for independence, because one random sample of residents is classified by two categorical variables (age group and whether they support the park).

The population is all residents of this city. The two categorical variables measured on each resident are age group (under 40, 40 and older) and support for the new park (support, oppose).

H_0 : Age group and support for the park are independent for residents of this city.

H_a : Age group and support for the park are associated for residents of this city.

We will use $\alpha = 0.05$.

1. *Random:* the residents were a random sample.
2. *Independence / 10% Condition:* 240 residents is reasonably less than 10% of all residents of the city.
3. *Large Counts:* all expected counts are at least 5. Each expected count is $\frac{\text{(row total)(column total)}}{240}$, and because both rows total 120,

$$E_{\text{Support}} = \frac{120 \cdot 160}{240} = 80, \qquad E_{\text{Oppose}} = \frac{120 \cdot 80}{240} = 40,$$

the same two values in each row; all are at least 5.

$$\chi^2 = \frac{(90-80)^2}{80} + \frac{(30-40)^2}{40} + \frac{(70-80)^2}{80} + \frac{(50-40)^2}{40} = 7.5.$$

$$\text{df} = (2-1)(2-1) = 1.$$

$$p\text{-value} = P(\chi^2 \geq 7.5) = \chi^2\text{cdf}(7.5,\ 1\text{E}99,\ 1) \approx 0.006.$$

Because the p-value of about 0.006 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence that age group and support for the new park are associated among residents of this city.

Independent Problem Solving

Independent 1: Expected counts

In a two-way table, one row total is 90, one column total is 64, and the grand total is 240. What is the expected count for that cell under the null model?

Solution. $E = \frac{\text{(row total)(column total)}}{\text{table total}} = \frac{90 \cdot 64}{240} = 24.$

Independent 2: Choosing a procedure

A random sample of adults is classified by whether they own a pet and whether they live in an apartment or a house. The goal is to determine whether housing type is associated with pet ownership. Which chi-square procedure is appropriate, and why?

Solution. A chi-square test for *independence*: one random sample is classified by two categorical variables (housing type and pet ownership), so we test whether they are associated.

Independent 3: Degrees of freedom

A chi-square test on a two-way table has 4 rows and 3 columns. What are the degrees of freedom?

Solution. $\text{df} = (4 - 1)(3 - 1) = 6.$

Independent 4: Chi-square contribution

A cell has expected count 42.5 and observed count 51. What is this cell's contribution to the chi-square statistic?

Solution. $\frac{(51 - 42.5)^2}{42.5} = \frac{72.25}{42.5} \approx 1.70.$

Independent 5: Interpreting the statistic

A large chi-square statistic means the observed counts are ________________ from the expected counts. Does this tend to give stronger or weaker evidence against the null hypothesis?

Solution. *Far.* A large statistic means the observed counts are far from what the null predicts, which gives *stronger* evidence against the null hypothesis (chi-square tests are right-tailed).

Independent 6: Writing hypotheses

A random sample of voters is classified by political party and whether they support a new transit measure. Write the hypotheses for the appropriate chi-square test, in words and in context.

Solution. This is a test for independence (one sample, two variables). H_0: political party and support for the transit measure are independent (not associated) in the population of voters. H_a: political party and support for the transit measure are associated in the population of voters.

Independent 7: FRQ

A veterinarian takes separate random samples of 50 cats and 50 dogs and records which of three food textures (dry, wet, mixed) each pet prefers. Is there convincing evidence that the distribution of texture preference differs between cats and dogs? Use $\alpha = 0.05$.

	Dry	Wet	Mixed	Total
Cats	30	10	10	50
Dogs	20	20	10	50
Total	50	30	20	100

Solution. Use a chi-square test for homogeneity, because two separate samples (cats and dogs) are compared on one categorical response (food-texture preference).

The two groups are all cats and all dogs of the kind sampled. For each group we consider the distribution of preferred food texture (dry, wet, mixed).

$$H_0 : \text{The distribution of texture preference is the same for cats and dogs.}$$

$$H_a : \text{The distribution of texture preference is not the same for cats and dogs.}$$

We will use $\alpha = 0.05$.

1. *Random:* the cats and dogs were two separate random samples.
2. *Independence / 10% Condition:* 50 cats and 50 dogs are each reasonably less than 10% of all such animals.
3. *Large Counts:* all expected counts are at least 5. For each group, dry $= \frac{50 \cdot 50}{100} = 25$, wet $= \frac{50 \cdot 30}{100} = 15$, and mixed $= \frac{50 \cdot 20}{100} = 10$; all are at least 5.

$$\chi^2 = \frac{(30-25)^2}{25} + \frac{(10-15)^2}{15} + \frac{(10-10)^2}{10} + \frac{(20-25)^2}{25} + \frac{(20-15)^2}{15} + \frac{(10-10)^2}{10} \approx 5.33.$$

$$\text{df} = (2-1)(3-1) = 2.$$

$$p\text{-value} = P(\chi^2 \geq 5.33) = \chi^2\text{cdf}(5.33,\ 1\text{E}99,\ 2) \approx 0.070.$$

Because the p-value of about 0.070 is greater than $\alpha = 0.05$, we fail to reject H_0. The data do not provide convincing evidence that the distribution of texture preference is different for cats and dogs.

Independent 8: FRQ

One random sample of 200 high-school students is classified by whether they play a sport and whether they have a part-time job. Is there convincing evidence of an association? Use $\alpha = 0.05$.

	Job	No job	Total
Plays a sport	40	60	100
No sport	60	40	100
Total	100	100	200

Solution. Use a chi-square test for independence, because one random sample of students is classified by two categorical variables (playing a sport and having a part-time job).

The population is all students of this type. The two categorical variables measured on each student are whether they play a sport (yes, no) and whether they have a part-time job (job, no job).

$$H_0: \text{Playing a sport and having a part-time job are independent for these students.}$$

$$H_a: \text{Playing a sport and having a part-time job are associated for these students.}$$

We will use $\alpha = 0.05$.

1. *Random:* the 200 students were a random sample.
2. *Independence / 10% Condition:* 200 students is reasonably less than 10% of all such students.
3. *Large Counts:* every expected count is $\frac{100\cdot 100}{200} = 50$, which is at least 5.

$$\chi^2 = \frac{(40-50)^2}{50} + \frac{(60-50)^2}{50} + \frac{(60-50)^2}{50} + \frac{(40-50)^2}{50} = 2+2+2+2 = 8.$$

$$\text{df} = (2-1)(2-1) = 1.$$

$$p\text{-value} = P(\chi^2 \geq 8) = \chi^2\text{cdf}(8,\ 1\text{E}99,\ 1) \approx 0.005.$$

Because the p-value of about 0.005 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence that playing a sport and having a part-time job are associated for these students.

AP Statistics Module 3E

Chi-Square Inference Overview

Chi-square inference is used for **categorical data recorded as counts**. We compare the counts we actually *observed* with the counts we would *expect* if the null hypothesis were true. On the 2026 AP Exam there are two chi-square tests, **homogeneity** and **independence**, which share the same formulas and differ only in study design and hypothesis wording.

Key formulas

$$\chi^2 = \sum \frac{(O-E)^2}{E} \qquad \text{df} = (r-1)(c-1)$$

$$\text{Expected count for a cell} = \frac{(\text{row total})(\text{column total})}{\text{table total}}$$

Here O is an observed count, E the matching expected count, and r, c the numbers of rows and columns. The statistic is never negative and the test is **right-tailed**: a larger χ^2 means observed counts farther from expected — stronger evidence against H_0.

Choosing the right test

Count how many times a population was sampled. *Separate* samples or treatment groups compared on one categorical response ⇒ **test for homogeneity**. *One* sample with *two* categorical variables recorded per individual ⇒ **test for independence**.

Writing the hypotheses (the null matches the test's name)

Homogeneity: H_0 = the distribution of the response is the *same* across groups (H_a: not all the same). *Independence:* H_0 = the two variables are *independent*, i.e. not associated (H_a: associated). Write both in words and in context.

Checking the conditions

1. **Random:** the data come from a random sample, or a randomized experiment / separate random samples.
2. **Independence / 10% condition:** when sampling without replacement, each sample is no more than 10% of its population.
3. **Large counts:** every *expected* count (not observed count) is at least 5.

Stating the conclusion

Find the p-value as the right-tail area $P(\chi^2 \geq \text{stat}) = \chi^2\text{cdf}(\text{stat, 1E99}, df)$, then complete the template:

> "Because the p-value is *(less/greater)* than $\alpha = 0.05$, we *(reject/fail to reject)* H_0. There *(is/is not)* convincing statistical evidence that *(restate H_a in context)*."

Common traps

Put *counts* in the formula, never proportions or percents. Check *expected* counts ≥ 5, not observed. Never write "accept H_0"; say "fail to reject H_0," and don't conclude the variables are unrelated.