

Unit 3: Inference for Categorical Data: Proportions

Unit 3 is where statistics starts answering questions about populations. You learn how to estimate a population proportion with a confidence interval, how to test a claim about a proportion with a z-test, how to compare two proportions, and how to use chi-square tests for categorical data in two-way tables.

AP Statistics

Proportions

About 14 minutes to read

How to use this guide

Read it in order the first time. The sampling distribution is the engine underneath everything, confidence intervals use it to estimate, hypothesis tests use it to decide, and the two-sample and chi-square procedures repeat the same logic with more groups. On the exam, free-response questions follow a strict script: state the parameter in context, check the conditions, do the mechanics, then conclude in context. Practice writing all four steps every time.

After the first read, use the trap boxes to review the interpretations the exam grades hardest: what a confidence level actually means, what a p-value actually means, and how Type I and Type II errors play out in context. Finish with the practice questions and the recall check on the last page.

What this unit is worth. Inference for categorical data is one of the most heavily tested units on the AP Statistics exam, and its procedures are the template for everything in the units on means and regression later. The four-step process you learn here, state, plan, do, conclude, is the same process every inference free-response question expects.

3.1 Estimators and 3.2 Sampling Distributions for Sample Proportions

An **estimator** is a sample statistic used to estimate a population parameter. The sample proportion $\hat{p}$ is the **point estimator** of the population proportion p : a single number standing in for the unknown truth. It is an **unbiased estimator** because, over many samples, it neither systematically overestimates nor underestimates p . Unbiased does not mean always right. It means the errors balance out in the long run.

The **sampling distribution of a sample proportion** is the distribution of $\hat{p}$ values across all possible samples of size n . When the sampled values are independent, its mean is $\mu(\hat{p}) = p$ and its standard deviation is $\sigma(\hat{p}) = \sqrt{p(1-p)/n}$. The standard deviation has a name in inference: it is the **standard error**, $SE(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}$, which measures how much a statistic typically varies from the parameter.

Before using the normal model for $\hat{p}$, check three conditions. The **randomization condition** requires a random sample, or two independent random samples, or a randomized experiment. The **10% condition** says the sample must be no more than 10% of the population when sampling without replacement, so the draws are close enough to independent. The **large counts condition** says the sampling distribution is approximately normal when the expected number of successes and failures are both at least 10: $np \geq 10$ and $n(1-p) \geq 10$.

Trap. The large counts condition uses different numbers for intervals and tests. For a confidence interval you check the *observed* counts, $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$, because no null value exists yet. For a hypothesis test you check the *expected* counts under the null, $np_0 \geq 10$ and $n(1-p_0) \geq 10$. Mixing them up is a common lost point.

3.3 Constructing a Confidence Interval for a Population Proportion

The **one-sample z-interval for a population proportion** is $\hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$. A **confidence interval** is a point estimate plus or minus a **margin of error**, and the margin of error is the **critical value** z^* times the standard error. The critical value z^* is the z-score enclosing the middle C% of the standard normal curve: 1.645 for 90%, 1.96 for 95%, 2.576 for 99%.

The four-step script starts by stating the parameter in context: the true proportion of the response variable in the population, naming all three. Then check the conditions: random sample, 10% condition, and the observed counts $n\hat{p}$ and $n(1-\hat{p})$ each at least 10. Then do the mechanics with the formula above. Then interpret.

The correct interpretation is: we are C% confident that the interval captures the true population proportion. "Confident" describes the method, not this particular interval. In repeated random samples, about C% of intervals built this way will contain the parameter. This one either does or does not.

Trap. The single most-tested misconception in this unit is saying there is a C% probability that the true proportion lies in the interval. That is wrong. The parameter is a fixed number, not a random one. What varies is the interval from sample to sample, and C% of those intervals capture the parameter. Write "we are C% confident the interval captures" and you are safe.

3.4 Justifying a Claim Based on a Confidence Interval

A confidence interval is a range of plausible values for the parameter. To justify a claim, check whether the claimed value falls inside the interval. If a 95% interval for the proportion of voters supporting a measure is (0.48, 0.56), then 0.50 is a plausible value and there is not convincing evidence the true proportion differs from one half. If the claimed value falls outside the interval, the data provide evidence against it.

Three quantities move together. For a fixed sample, raising the confidence level raises z^* , which raises the margin of error and widens the interval. You pay for extra confidence with a wider interval. Raising the sample size shrinks the standard error, which shrinks the margin of error and narrows the interval. Interval width falls roughly like $1/\sqrt{n}$, so quadrupling the sample halves the width.

The margin of error formula also runs backward into sample size planning. Solving $MOE = z^* \sqrt{\hat{p}(1-\hat{p})/n}$ for n gives $n = z^{*2} \hat{p}(1-\hat{p}) / MOE^2$. When no prior estimate of $\hat{p}$ exists, use $\hat{p} = 0.5$, which gives the largest possible required n and guarantees the margin of error you want.

3.5 Setting Up a Test for a Population Proportion

A **hypothesis test** is a procedure for deciding about a parameter value. The **null hypothesis** H_0 is the status-quo claim, assumed true unless the data convince you otherwise. For a proportion it is $H_0: p = p_0$, where p_0 is the null hypothesized value. The **alternative hypothesis** H_a is the claim you are gathering evidence for. It can be one-sided, $H_a: p < p_0$ or $H_a: p > p_0$, or two-sided, $H_a: p \neq p_0$.

Choose the direction of H_a from the research question, not from the data. If the question asks whether a proportion has increased, the alternative is one-sided: $H_a: p > p_0$. If it asks whether the proportion differs from a value, with no direction given, the alternative is two-sided. Peeking at the sample first and then picking the side that matches is cheating, and the exam will penalize it.

The setup step also states the parameter in context and checks the conditions: random sample, 10% condition, and the expected counts np_0 and $n(1-p_0)$ each at least 10. Remember, tests use the null value p_0 for the counts, not the sample value $\hat{p}$.

Trap. The hypotheses are about the population parameter p , never about the sample statistic $\hat{p}$. Writing $H_0: \hat{p} = 0.5$ is wrong on its face, because $\hat{p}$ is known from the data and needs no test. If you see a hat in a hypothesis, fix it before you do anything else.

3.6 p-Values and 3.7 Carrying Out the Test

The **test statistic** for a one-sample z-test is $z = (\hat{p} - p_0) / \sqrt{(p_0(1-p_0)/n)}$. It measures how many standard errors the sample result sits from the null value. Under H_0 , this z follows the standard normal distribution, which is the **null distribution** here: the pattern of test-statistic values you would expect if the null were true.

The **p-value** is the probability, assuming H_0 is true, of getting a test statistic at least as extreme as the one observed, in the direction or directions H_a specifies. For a one-sided test it is one tail of the null distribution. For a two-sided test it is both tails. A small p-value means the observed result would be surprising under the null, which is evidence for the alternative.

The **decision rule** compares the p-value to the **significance level** α , the preset probability of rejecting a true null. If $p\text{-value} \leq \alpha$, reject H_0 : there is convincing statistical evidence for H_a . If $p\text{-value} > \alpha$, fail to reject H_0 : there is not convincing evidence for H_a . The conclusion is stated in context, in terms of the alternative, with non-definitive language. A test never proves H_0 true.

Trap. A p-value is not the probability that the null hypothesis is true, and it is not the probability that your conclusion is wrong. It is the probability of data this extreme *given* that H_0 is true. Any answer choice that treats the p-value as $P(H_0 \text{ is true})$ is the distractor. Read the conditioning direction carefully.

3.8 Potential Errors When Performing Tests

Every test can err in two ways. A **Type I error** is rejecting H_0 when it is actually true: concluding there is convincing evidence for the alternative when there is not. Its probability is the significance level α , set before data collection. A **Type II error** is failing to reject H_0 when the alternative is actually true: missing real evidence. Its probability is $1 - \text{power}$.

Power is the probability of correctly rejecting a false null, ideally at least 0.80. Power rises when the sample size increases, when the standard error decreases, when the true parameter value lies farther from the null, or when the significance level increases. Each of these makes it easier to detect a real effect.

Before the study, weigh the consequences of each error in context. The consequence of a Type I error drives the choice of α : if a false alarm is costly, pick a small α . The consequence of a Type II error drives the sample size: if missing a real effect is costly, collect more data. State both errors in context on the exam. "A Type I error would mean concluding the drug works when it does not" earns the point; "rejecting a true null" without context does not.

ERROR	WHAT HAPPENS	PROBABILITY
Type I	Reject a true H_0	α
Type II	Fail to reject a false H_0	$1 - \text{power}$
Correct rejection	Reject a false H_0	Power

3.9 through 3.11 Two-Sample Procedures for Proportions

Comparing two groups repeats the one-sample logic with a difference. The **sampling distribution of $\hat{p}_1 - \hat{p}_2$** has mean $p_1 - p_2$ and standard deviation $\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$. It is approximately normal when all four expected counts clear 10: n_1p_1 , $n_1(1-p_1)$, n_2p_2 , $n_2(1-p_2)$.

The **two-sample z-interval** is $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{(\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2)}$. Check the randomization condition for both samples, the 10% condition for both, and the observed counts for both samples. Interpret it the same way: we are C% confident the interval captures the true difference $p_1 - p_2$.

For a difference interval, zero is the decision point. If the interval contains 0, the data do not give convincing evidence of a difference between the population proportions. If the entire interval sits above 0, there is convincing evidence that p_1 exceeds p_2 . If it sits entirely below 0, the evidence points the other way.

3.12 and 3.13 The Two-Sample z-Test

The **two-sample z-test** compares two population proportions. The null hypothesis states no difference: $H_0: p_1 = p_2$, or equivalently $H_0: p_1 - p_2 = 0$. The alternative can be one-sided or two-sided, chosen from the research question before seeing the data.

Because H_0 assumes one common proportion, the test uses the **pooled proportion** $\hat{p}_c = (n_1\hat{p}_1 + n_2\hat{p}_2) / (n_1 + n_2)$, which combines the successes from both groups. The **test statistic** is $z = ((\hat{p}_1 - \hat{p}_2) - 0) / \sqrt{(\hat{p}_c(1-\hat{p}_c)(1/n_1 + 1/n_2))}$. The normality check uses the pooled value: $n_1\hat{p}_c$, $n_1(1-\hat{p}_c)$, $n_2\hat{p}_c$, and $n_2(1-\hat{p}_c)$ must all be at least 10.

Trap. The interval and the test use different standard errors, and this is the most common mechanical error in the unit. The two-sample *interval* uses the unpooled standard error with $\hat{p}_1$ and $\hat{p}_2$ kept separate. The two-sample *test* uses the pooled proportion $\hat{p}_c$ because H_0 says the two population proportions are equal. Pool for tests, do not pool for intervals.

3.14 and 3.15 Chi-Square Tests for Homogeneity and Independence

Chi-square tests handle categorical data in two-way tables. The **chi-square test for homogeneity** asks whether the distribution of one categorical variable differs across two or more populations or treatments. The **chi-square test for independence** asks whether two categorical variables are associated within a single sampled population. Same statistic, different question, and the exam expects you to name the right one from the study design.

The hypotheses match the question. For homogeneity: H_0 says there is no difference in the distribution of the categorical variable across populations, and H_a says there is a difference. For independence: H_0 says there is no association between the two variables, and H_a says there is an association. Note that chi-square alternatives never have a direction. There is no one-sided chi-square test.

The **expected counts** under the null are (row total $\times$ column total) / table total for each cell. The conditions are the randomization condition, the 10% condition, and the expected counts condition: every expected count must be at least 5. The **test statistic** is $\chi^2 = \sum((\text{Observed} - \text{Expected})^2 / \text{Expected})$, summed over all cells, with **degrees of freedom** $df = (\text{rows} - 1)(\text{columns} - 1)$. The chi-square distribution takes only positive values and is skewed right, with less skew as df grows. The p-value comes from the upper tail: large χ^2 values mean the observed counts sit far from what the null predicts.

Trap. Rejecting the chi-square null does not tell you *which* cells differ or *which* variables drive the association. The test answers only whether a difference or association exists somewhere in the table. A conclusion that names a specific cell or direction goes beyond what the test supports.

Confusions That Cost Points

PAIR	HOW TO KEEP THEM STRAIGHT
Confidence level vs probability the parameter is in the interval	The level describes the method: C% of intervals from repeated samples capture the parameter. The parameter is fixed, so this one interval either captures it or does not.
p-value vs $P(H_0 \text{ is true})$	The p-value assumes H_0 is true and measures how surprising the data are under that assumption. It never gives the probability that a hypothesis is true.
One-sided vs two-sided p-value	One-sided uses one tail. Two-sided doubles it. The choice comes from the research question, fixed before data collection.
Type I vs Type II error	Type I rejects a true null, probability α . Type II misses a false null, probability $1 - \text{power}$. State each in context.
Pooled vs unpooled standard error	Two-sample tests pool with $\hat{p}_c$ because H_0 claims equal proportions. Two-sample intervals keep the groups separate.
Interval counts vs test counts	Intervals check observed counts $n\hat{p} \geq 10$. Tests check expected counts $np_0 \geq 10$. Chi-square checks expected counts ≥ 5 .
Homogeneity vs independence	Homogeneity compares the distribution of one variable across separate populations. Independence tests association between two variables in one population.

Practice Questions

Original questions written for this guide in the style of the AP exam. Answers and explanations are on the next page, so complete the questions before checking them.

1. A 95% confidence interval for the proportion of high school students who eat breakfast daily is (0.42, 0.50). Which statement is correct?
 - A. There is a 95% probability that the true proportion is between 0.42 and 0.50.
 - B. We are 95% confident that the interval from 0.42 to 0.50 captures the true proportion of high school students who eat breakfast daily.
 - C. 95% of high school students eat breakfast daily with probability between 0.42 and 0.50.
 - D. The sample proportion is guaranteed to lie between 0.42 and 0.50.
2. A researcher tests $H_0: p = 0.30$ against $H_a: p > 0.30$ at $\alpha = 0.05$ and finds a p-value of 0.03. Which conclusion is correct?
 - A. There is a 3% chance the null hypothesis is true.
 - B. There is convincing evidence that the true proportion exceeds 0.30.
 - C. The null hypothesis is proven false with 97% certainty.
 - D. There is a 3% chance of making a Type I error on this test.
3. Two independent random samples give $\hat{p}_1 = 0.60$ ($n_1 = 200$) and $\hat{p}_2 = 0.52$ ($n_2 = 200$). For the two-sample z-test of $H_0: p_1 = p_2$, which standard error is correct?
 - A. $\sqrt{(0.60(0.40)/200 + 0.52(0.48)/200)}$, using the separate sample proportions
 - B. $\sqrt{(\hat{p}_c(1-\hat{p}_c)(1/200 + 1/200))}$, where $\hat{p}_c = 0.56$, the pooled proportion
 - C. $\sqrt{(0.5(0.5)/200 + 0.5(0.5)/200)}$, using 0.5 for both groups
 - D. $(\sqrt{(0.60(0.40)/200)} + \sqrt{(0.52(0.48)/200)}) / 2$, the average of the two standard errors
4. A chi-square test for independence of major choice and class year gives $\chi^2 = 18.4$ with $df = 6$ and a p-value of 0.005. At $\alpha = 0.05$, what is the correct conclusion?
 - A. There is convincing evidence of an association between major choice and class year in the population.
 - B. There is convincing evidence that seniors choose different majors than freshmen.
 - C. The p-value of 0.005 means there is a 0.5% chance the variables are independent.
 - D. Fail to reject H_0 because the test statistic is positive.

Answer Key

- 1. B.** This is the textbook-correct interpretation: confidence describes the method, and the sentence names the parameter, the population, and the interval. A is the most common wrong answer on the exam. It treats the parameter as random, but the true proportion is a fixed number. C confuses the interval for the parameter with a statement about individual students. D is wrong because the sample proportion is the center of the interval, not something the interval needs to capture.
- 2. B.** The p-value 0.03 is less than $\alpha = 0.05$, so reject H_0 and state the conclusion in context with non-definitive language. A treats the p-value as $P(H_0 \text{ true})$, which it is not. C claims proof and certainty, which no test provides. D confuses the p-value with α : the Type I error probability is the preset 0.05, not the observed 0.03.
- 3. B.** A hypothesis test assumes H_0 is true, and H_0 says the two population proportions are equal, so the test pools the successes into $\hat{p}_c = (120 + 104)/400 = 0.56$. A is the unpooled standard error, which belongs to the confidence interval, not the test. C invents 0.5 values with no basis. D averages two standard errors, which is not a valid formula.
- 4. A.** The p-value 0.005 is less than 0.05, so reject the null of no association. The conclusion stays general: evidence of an association somewhere in the table, without naming which cells drive it. B goes further than the test supports by naming specific groups. C misreads the p-value as the probability the null is true. D is nonsense: chi-square statistics are always non-negative, and positivity has nothing to do with the decision.

One-Page Recall Check

- Define an estimator and an unbiased estimator, and explain what unbiased does and does not promise.
- Write the mean and standard deviation of the sampling distribution of $\hat{p}$.
- State the three conditions for proportion procedures and explain what each one protects against.
- Explain why intervals check observed counts while tests check expected counts under the null.
- Write the one-sample z-interval formula and name the critical values for 90%, 95%, and 99%.
- Interpret a confidence interval correctly, and explain why "95% probability the parameter is inside" is wrong.
- Explain how confidence level, margin of error, sample size, and interval width move together.
- Write H_0 and H_a for a one-sample proportion test, and explain why the hypotheses use p and not $\hat{p}$.
- Define the p-value in a way that keeps the conditioning direction straight.
- State the decision rule and write a conclusion in context with non-definitive language.
- Define Type I and Type II errors in context, and list four ways to increase power.
- Explain when to pool the proportion and when not to, and why.
- Write the chi-square test statistic and the degrees of freedom formula.
- Distinguish the chi-square test for homogeneity from the test for independence.

Where to go next. Turn every missed item above into flashcards and drill them spaced out over several days rather than in one sitting. In Rycal, open the proportions deck under AP Statistics. The deck covers the terms in this guide, and its practice questions target the same traps named here. If you have a test date, add it in the Test Planner. You can also start your next review with a Brain Dump, then check what you missed against this guide.

Key terms for this unit

Estimator, Unbiased estimator, Point estimator, Sampling distribution of a sample proportion, Randomization condition (proportions), 10% condition (proportions), Large counts condition (sampling distribution of a sample proportion), One-sample z-interval for a population proportion, Confidence interval, Stating the parameter in context (proportion interval), Confidence level, Critical value (z^*), Standard error (of a statistic), Margin of error, Point estimate, Normality condition (one-sample z-interval for a population proportion), Confidence interval interpretation (population proportion), Relationships among confidence level, margin of error, sample size, and interval width, Justifying a claim based on a confidence interval (population proportion), One-sample z-test for a population proportion, Hypothesis test, Null hypothesis, Alternative hypothesis, Normality condition (one-sample z-test for a population proportion), p-value, Null distribution, Test statistic (one-sample z-test for a proportion), Significance level, Statistical significance, Decision rule (hypothesis test), Hypothesis test conclusion, Type I error, Type II error, Power, Factors that increase power, Consequences of Type I and Type II errors, Sampling distribution of the difference in sample proportions, Large counts condition (sampling distribution of the difference in sample proportions), Two-sample z-interval for a difference between population proportions, Standard error (difference in proportions), Normality condition (two-sample z-interval for a difference in proportions), Confidence interval for a difference in proportions (interpretation and claim), Two-sample z-test for a difference between two population proportions, Pooled proportion, Null and alternative hypotheses (two-sample z-test for proportions), Normality condition (two-sample z-test for a difference in proportions), Test statistic (two-sample z-test for proportions), Chi-square statistic, Chi-square distribution, Chi-square test for homogeneity, Chi-square test for independence, Null and alternative hypotheses (chi-square test for homogeneity), Null and alternative hypotheses (chi-square test for independence), Randomization condition (chi-square test), 10% condition (chi-square test), Expected counts condition (chi-square test), Expected counts (two-way table), Chi-square test statistic, Degrees of freedom (chi-square), Stating the parameter in context (proportion test), Stating the parameters in context (two-sample