AP STATISTICS

UNIT 6

Inference for Categorical Data: Proportions

12–15%
AP EXAM WEIGHTING

~16–18
CLASS PERIODS

UNIT 6

12–15% AP EXAM WEIGHTING | **~16–18** CLASS PERIODS

Inference for Categorical Data: Proportions

BIG IDEA 1
Variation and Distribution **VAR**
- When can we use a normal distribution to perform inference calculations involving population proportions?

BIG IDEA 2
Patterns and Uncertainty **UNC**
- How can we narrow the width of a confidence interval?

BIG IDEA 3
Data-Based Predictions, Decisions, and Conclusions **DAT**
- If the proportion of subjects who experience serious side effects when taking a new drug is smaller than the proportion of subjects who experience serious side effects when taking a placebo, how can we determine if the difference is statistically significant?

Developing Understanding

This unit introduces statistical inference, which will continue through the end of the course. Students will analyze categorical data to make inferences about binomial population proportions. Provided conditions are met, students will use statistical inference to construct and interpret confidence intervals to estimate population proportions and perform significance tests to evaluate claims about population proportions. Students begin by learning inference procedures for one proportion and then examine inference methods for a difference between two proportions. They will also interpret the two types of errors that can be made in a significance test, their probabilities, and possible consequences in context.

Building Course Skills

1.D 3.D 4.D

Unit 6 is a critical transition point in the course, as students begin learning skills that will be applied repeatedly in subsequent units. Students need to familiarize themselves with these procedures so they can build proficiency over time. Applying different inference methods requires fluency with verifying conditions. Students often check conditions superficially (e.g., just listing "SRS") without explicitly connecting them to the problem. Teachers can make sure students practice verifying conditions in context by providing numerical calculations and explaining how each condition is met.

Precision of language is key. Students often interpret *confidence intervals* and *confidence levels* incorrectly. Providing students with sentence starters or templates can help them learn to generate appropriate responses (e.g., Confidence interval: "We are 95% confident that the interval from ___ to ___ captures the [parameter in context]."). For decisions based on a hypothesis test, students may incorrectly claim that "we can accept" or "have proven" the null. Teachers can reinforce early and often that statistical tests do not provide evidence for what can be accepted or proved; they only provide evidence for "rejecting" or "failing to reject" the null.

Preparing for the AP Exam

When using statistical inference to construct confidence intervals or perform significance tests, students should identify the appropriate inference method by name or formula. For inference with population proportions, students should verify that the following conditions are met: (1) random sample and (2) large sample (e.g., $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$). When sampling without replacement, students should also verify that the sample size is at most 10% of the population. Verification should be simple and specific.

Next, students should present calculations and then interpret results in the context of the problem. Students often find it beneficial to use language provided in the question. In **2017 FRQ 2**, for example, the response might read "We can be 95% confident that the proportion of *all customers who, having asked for a cup of water when placing an order, will fill the cup with a soft drink* is between 0.1883 and 0.3867."

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods ~16–18 CLASS PERIODS
VAR-1	6.1 Introducing Statistics: Why Be Normal?	1.A Identify the question to be answered or problem to be solved *(not assessed)*.	
UNC-4	6.2 Constructing a Confidence Interval for a Population Proportion	1.D Identify an appropriate inference method for confidence intervals. 4.C Verify that inference procedures apply in a given situation. 3.D Construct a confidence interval, provided conditions for inference are met.	
	6.3 Justifying a Claim Based on a Confidence Interval for a Population Proportion	4.B Interpret statistical calculations and findings to assign meaning or assess a claim. 4.D Justify a claim based on a confidence interval. 4.A Make an appropriate claim or draw an appropriate conclusion.	
VAR-6	6.4 Setting Up a Test for a Population Proportion	1.F Identify null and alternative hypotheses. 1.E Identify an appropriate inference method for significance tests. 4.C Verify that inference procedures apply in a given situation.	
VAR-6, DAT-3	6.5 Interpreting p-Values	3.E Calculate a test statistic and find a p-value, provided conditions for inference are met. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
DAT-3	6.6 Concluding a Test for a Population Proportion	4.E Justify a claim using a decision based on significance tests.	

continued on next page

UNIT AT A GLANCE *(cont'd)*

Enduring Understanding	Topic	Skills	Class Periods ~16–18 CLASS PERIODS
UNC-5	**6.7 Potential Errors When Performing Tests**	1.B Identify key and relevant information to answer a question or solve a problem. 3.A Determine relative frequencies, proportions, or probabilities using simulation or calculations. 4.A Make an appropriate claim or draw an appropriate conclusion. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
UNC-4	**6.8 Confidence Intervals for the Difference of Two Proportions**	1.D Identify an appropriate inference method for confidence intervals. 4.C Verify that inference procedures apply in a given situation. 3.D Construct a confidence interval, provided conditions for inference are met.	
UNC-4	**6.9 Justifying a Claim Based on a Confidence Interval for a Difference of Population Proportions**	4.B Interpret statistical calculations and findings to assign meaning or assess a claim. 4.D Justify a claim based on a confidence interval.	
VAR-6	**6.10 Setting Up a Test for the Difference of Two Population Proportions**	1.F Identify null and alternative hypotheses. 1.E Identify an appropriate inference method for significance tests. 4.C Verify that inference procedures apply in a given situation.	
VAR-6, DAT-3	**6.11 Carrying Out a Test for the Difference of Two Population Proportions**	3.E Calculate a test statistic and find a p-value, provided conditions for inference are met. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim. 4.E Justify a claim using a decision based on significance tests.	

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 6.
Review the results in class to identify and address any student misunderstandings.

SAMPLE INSTRUCTIONAL ACTIVITIES

The sample activities on this page are optional and are offered to provide possible ways to incorporate various instructional approaches into the classroom. They were developed in partnership with teachers from the AP community to share ways that they approach teaching some of the topics in this unit. Please refer to the Instructional Approaches section beginning on p. 207 for more examples of activities and strategies.

Activity	Topic	Sample Activity
1	6.4 6.7 6.8	**Error Analysis** Give student pairs a worksheet with 20 sets of hypotheses (including hypotheses for a population proportion and for the difference of two proportions), each with a common student mistake. Have students circle the incorrect part, write why the circled component is incorrect, and then write the correct hypotheses. Include errors such as using statistics instead of parameters, and interchanging the = and > in the two hypotheses.
2	6.5 6.6 6.11	**Sentence Starters** For a given question, provide students with a set of hypotheses, p-value, significance level, and context. Have them compare the p-value to the significance level to determine whether or not to reject the null hypothesis. Using a given sentence starter with blanks to fill in, have students write a sentence in context explaining if they have enough evidence to "reject H_0", or if they will "fail to reject H_0." Make sure students avoid the common mistake of implying that evidence supports an "accept H_0" conclusion or a "reject H_a" conclusion.
3	6.2 6.8	**The Scribe and the Calculator** Have students work with a partner to construct and interpret a confidence interval for a population proportion. Only one partner is allowed to use the calculator, and only the other partner is allowed to write. When a calculation needs to be made, the scribe can only describe to the calculator operator which buttons to push; when writing needs to be done, the calculator operator can only describe to the scribe what needs to be written. Have students switch roles when constructing and interpreting a confidence interval for the difference of two population proportions.

TOPIC 6.1

Introducing Statistics: Why Be Normal?

SKILL

Selecting Statistical Methods

1.A

Identify the question to be answered or problem to be solved.

Required Course Content

ENDURING UNDERSTANDING

VAR-1

Given that variation may be random or not, conclusions are uncertain.

LEARNING OBJECTIVE

VAR-1.H

Identify questions suggested by variation in the shapes of distributions of samples taken from the same population. **[Skill 1.A]**

ESSENTIAL KNOWLEDGE

VAR-1.H.1

Variation in shapes of data distributions may be random or not.

SKILLS

Selecting Statistical Methods

1.D

Identify an appropriate inference method for confidence intervals.

Statistical Argumentation

4.C

Verify that inference procedures apply in a given situation.

Using Probability and Simulation

3.D

Construct a confidence interval, provided conditions for inference are met.

TOPIC 6.2

Constructing a Confidence Interval for a Population Proportion

Required Course Content

ENDURING UNDERSTANDING

UNC-4

An interval of values should be used to estimate parameters, in order to account for uncertainty.

LEARNING OBJECTIVE

UNC-4.A

Identify an appropriate confidence interval procedure for a population proportion. **[Skill 1.D]**

ESSENTIAL KNOWLEDGE

UNC-4.A.1

The appropriate confidence interval procedure for a one-sample proportion for one categorical variable is a one sample z-interval for a proportion.

LEARNING OBJECTIVE

UNC-4.B

Verify the conditions for calculating confidence intervals for a population proportion. **[Skill 4.C]**

ESSENTIAL KNOWLEDGE

UNC-4.B.1

In order to make assumptions necessary for inference on population proportions, means, and slopes, we must check for independence in data collection methods and for selection of the appropriate sampling distribution.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-4.B Verify the conditions for calculating confidence intervals for a population proportion. **[Skill 4.C]**	**UNC-4.B.2** In order to calculate a confidence interval to estimate a population proportion, p, we must check for independence and that the sampling distribution is approximately normal. a. To check for independence: i. Data should be collected using a random sample or a randomized experiment. ii. When sampling without replacement, check that $n \leq 10\%N$, where N is the size of the population. b. To check that the sampling distribution of $\hat{p}$ is approximately normal (shape): i. For categorical variables, check that both the number of successes, $n\hat{p}$, and the number of failures, $n(1-\hat{p})$ are at least 10 so that the sample size is large enough to support an assumption of normality.
UNC-4.C Determine the margin of error for a given sample size and an estimate for the sample size that will result in a given margin of error for a population proportion. **[Skill 3.D]**	**UNC-4.C.1** Based on sample data, the standard error of a statistic is an estimate for the standard deviation for the statistic. The standard error of $\hat{p}$ is $SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. **UNC-4.C.2** A margin of error gives how much a value of a sample statistic is likely to vary from the value of the corresponding population parameter. **UNC-4.C.3** For categorical variables, the margin of error is the critical value (z^*) times the standard error (SE) of the relevant statistic, which equals $z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ for a one sample proportion. **UNC-4.C.4** The formula for margin of error can be rearranged to solve for n, the minimum sample size needed to achieve a given margin of error. For this purpose, use a guess for $\hat{p}$ or use $\hat{p} = 0.5$ in order to find an upper bound for the sample size that will result in a given margin of error.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-4.D Calculate an appropriate confidence interval for a population proportion. **[Skill 3.D]**	**UNC-4.D.1** In general, an interval estimate can be constructed as point estimate ± (margin of error). For a one-sample proportion, the interval estimate is $\hat{p} \pm z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. **CLARIFYING STATEMENT:** *Formulas for interval estimates do not appear explicitly on the AP Statistics Formula Sheet provided with the AP Statistics Exam. However, these formulas do not need to be memorized, as they can be constructed based on the general test statistic formula and the relevant standard error formulas that are provided on the formula sheet.* **UNC-4.D.2** Critical values represent the boundaries encompassing the middle C% of the standard normal distribution, where C% is an approximate confidence level for a proportion.
UNC-4.E Calculate an interval estimate based on a confidence interval for a population proportion. **[Skill 3.D]**	**UNC-4.E.1** Confidence intervals for population proportions can be used to calculate interval estimates with specified units.

TOPIC 6.3

Justifying a Claim Based on a Confidence Interval for a Population Proportion

Required Course Content

ENDURING UNDERSTANDING

UNC-4

An interval of values should be used to estimate parameters, in order to account for uncertainty.

LEARNING OBJECTIVE

UNC-4.F

Interpret a confidence interval for a population proportion. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

UNC-4.F.1

A confidence interval for a population proportion either contains the population proportion or it does not, because each interval is based on random sample data, which varies from sample to sample.

UNC-4.F.2

We are C% confident that the confidence interval for a population proportion captures the population proportion.

UNC-4.F.3

In repeated random sampling with the same sample size, approximately C% of confidence intervals created will capture the population proportion.

UNC-4.F.4

Interpreting a confidence interval for a one-sample proportion should include a reference to the sample taken and details about the population it represents.

continued on next page

SKILLS

Statistical Argumentation

4.B

Interpret statistical calculations and findings to assign meaning or assess a claim.

4.D

Justify a claim based on a confidence interval.

4.A

Make an appropriate claim or draw an appropriate conclusion.

AVAILABLE RESOURCE

- Classroom Resource > **Calculations Aren't Enough! The Importance of Communication in AP Statistics**

ILLUSTRATIVE EXAMPLE

UNC-4.F.4:

For interpreting a 99% confidence interval of (0.268, 0.292), based on the proportion of a nationally representative sample of twelfth-grade students who answered a particular multiple choice question correctly:

"We are 99 percent confident that the interval from 0.268 to 0.292 contains the population proportion of all United States twelfth-grade students who would answer this question correctly" (**2011 FRQ 6(a)**).

LEARNING OBJECTIVE

UNC-4.G

Justify a claim based on a confidence interval for a population proportion. **[Skill 4.D]**

ESSENTIAL KNOWLEDGE

UNC-4.G.1

A confidence interval for a population proportion provides an interval of values that may provide sufficient evidence to support a particular claim in context.

UNC-4.H

Identify the relationships between sample size, width of a confidence interval, confidence level, and margin of error for a population proportion. **[Skill 4.A]**

UNC-4.H.1

When all other things remain the same, the width of the confidence interval for a population proportion tends to decrease as the sample size increases. For a population proportion, the width of the interval is proportional to $\frac{1}{\sqrt{n}}$.

UNC-4.H.2

For a given sample, the width of the confidence interval for a population proportion increases as the confidence level increases.

UNC-4.H.3

The width of a confidence interval for a population proportion is exactly twice the margin of error.

TOPIC 6.4

Setting Up a Test for a Population Proportion

Required Course Content

ENDURING UNDERSTANDING

VAR-6

The normal distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-6.D

Identify the null and alternative hypotheses for a population proportion.
[Skill 1.F]

ESSENTIAL KNOWLEDGE

VAR-6.D.1

The null hypothesis is the situation that is assumed to be correct unless evidence suggests otherwise, and the alternative hypothesis is the situation for which evidence is being collected.

VAR-6.D.2

For hypotheses about parameters, the null hypothesis contains an equality reference (=, ≥, or ≤), while the alternative hypothesis contains a strict inequality (<, >, or ≠). The type of inequality in the alternative hypothesis is based on the question of interest. Alternative hypotheses with < or > are called one-sided, and alternative hypotheses with ≠ are called two-sided. Although the null hypothesis for a one-sided test may include an inequality symbol, it is still tested at the boundary of equality.

VAR-6.D.3

The null hypothesis for a population proportion is: $H_0: p = p_0$, where p_0 is the null hypothesized value for the population proportion.

VAR-6.D.4

A one-sided alternative hypothesis for a proportion is either $H_a: p < p_0$ or $H_a: p > p_0$. A two-sided alternate hypothesis is $H_a: p_1 \neq p_2$.

continued on next page

SKILLS

Selecting Statistical Methods

1.F

Identify null and alternative hypotheses.

1.E

Identify an appropriate inference method for significance tests.

Statistical Argumentation

4.C

Verify that inference procedures apply in a given situation.

AVAILABLE RESOURCES

- Classroom Resources >
  - **Inference**
  - **Coke® Versus Pepsi®: An Introductory Activity for Test of Significance**

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-6.D Identify the null and alternative hypotheses for a population proportion. **[Skill 1.F]**	**VAR-6.D.5** For a one-sample z-test for a population proportion, the null hypothesis specifies a value for the population proportion, usually one indicating no difference or effect.
VAR-6.E Identify an appropriate testing method for a population proportion. **[Skill 1.E]**	**VAR-6.E.1** For a single categorical variable, the appropriate testing method for a population proportion is a one-sample z-test for a population proportion.
VAR-6.F Verify the conditions for making statistical inferences when testing a population proportion. **[Skill 4.C]**	**VAR-6.F.1** In order to make statistical inferences when testing a population proportion, we must check for independence and that the sampling distribution is approximately normal: a. To check for independence: i. Data should be collected using a random sample or a randomized experiment. ii. When sampling without replacement, check that $n \le 10\%N$. b. To check that the sampling distribution of $\hat{p}$ is approximately normal (shape): i. Assuming that H_0 is true $(p = p_0)$, verify that both the number of successes, np_0, and the number of failures, $n(1 - p_0)$ are at least 10 so that that the sample size is large enough to support an assumption of normality.

TOPIC 6.5

Interpreting p-Values

Required Course Content

ENDURING UNDERSTANDING

VAR-6

The normal distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-6.G

Calculate an appropriate test statistic and p-value for a population proportion.
[Skill 3.E]

ESSENTIAL KNOWLEDGE

VAR-6.G.1

The distribution of the test statistic assuming the null hypothesis is true (null distribution) can be either a randomization distribution or when a probability model is assumed to be true, a theoretical distribution (z).

VAR-6.G.2

When using a z-test, the standardized test statistic can be written:

$$\text{test statistic} = \frac{\text{sample statistic-null value of the parameter}}{\text{standard deviation of the statistic}}.$$

This is called a z-statistic for proportions.

VAR-6.G.3

The test statistic for a population proportion is:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

CLARIFYING STATEMENT:
The formulas for test statistics do not appear explicitly on the AP Statistics Formula Sheet provided with the AP Statistics Exam. However, these formulas do not need to be memorized, as they can be constructed based on the general test statistic formula and the relevant standard error formulas that are provided on the formula sheet.

continued on next page

SKILLS

Using Probability and Simulation

3.E
Calculate a test statistic and find a p-value, provided conditions for inference are met.

Statistical Argumentation

4.B
Interpret statistical calculations and findings to assign meaning or assess a claim.

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

LEARNING OBJECTIVE

VAR-6.G

Calculate an appropriate test statistic and p-value for a population proportion.
[Skill 3.E]

ESSENTIAL KNOWLEDGE

VAR-6.G.4

A p-value is the probability of obtaining a test statistic as extreme or more extreme than the observed test statistic when the null hypothesis and probability model are assumed to be true. The significance level may be given or determined by the researcher.

ENDURING UNDERSTANDING

DAT-3

Significance testing allows us to make decisions about hypotheses within a particular context.

LEARNING OBJECTIVE

DAT-3.A

Interpret the p-value of a significance test for a population proportion.
[Skill 4.B]

ESSENTIAL KNOWLEDGE

DAT-3.A.1

The p-value is the proportion of values for the null distribution that are as extreme or more extreme than the observed value of the test statistic. This is:

a. The proportion at or above the observed value of the test statistic, if the alternative is >.

b. The proportion at or below the observed value of the test statistic, if the alternative is <.

c. The proportion less than or equal to the negative of the absolute value of the test statistic plus the proportion greater than or equal to the absolute value of the test statistic, if the alternative is ≠.

DAT-3.A.2

An interpretation of the p-value of a significance test for a one-sample proportion should recognize that the p-value is computed by assuming that the probability model and null hypothesis are true, i.e., by assuming that the true population proportion is equal to the particular value stated in the null hypothesis.

TOPIC 6.6

Concluding a Test for a Population Proportion

Required Course Content

ENDURING UNDERSTANDING

DAT-3

Significance testing allows us to make decisions about hypotheses within a particular context.

LEARNING OBJECTIVE

DAT-3.B

Justify a claim about the population based on the results of a significance test for a population proportion. **[Skill 4.E]**

ESSENTIAL KNOWLEDGE

DAT-3.B.1

The significance level, α, is the predetermined probability of rejecting the null hypothesis given that it is true.

DAT-3.B.2

A formal decision explicitly compares the p-value to the significance level, α. If the p-value $\leq \alpha$, reject the null hypothesis. If the p-value $> \alpha$, fail to reject the null hypothesis.

DAT-3.B.3

Rejecting the null hypothesis means there is sufficient statistical evidence to support the alternative hypothesis. Failing to reject the null means there is insufficient statistical evidence to support the alternative hypothesis.

DAT-3.B.4

The conclusion about the alternative hypothesis must be stated in context.

DAT-3.B.5

A significance test can lead to rejecting or not rejecting the null hypothesis, but can never lead to concluding or proving that the null hypothesis is true. Lack of statistical evidence for the alternative hypothesis is not the same as evidence for the null hypothesis.

continued on next page

SKILL

Statistical Argumentation

4.E

Justify a claim using a decision based on significance tests.

AVAILABLE RESOURCES

- Classroom Resource >
  - **Inference**
  - **Calculations Aren't Enough! The Importance of Communication in AP Statistics**

LEARNING OBJECTIVE

DAT-3.B

Justify a claim about the population based on the results of a significance test for a population proportion. **[Skill 4.E]**

ESSENTIAL KNOWLEDGE

DAT-3.B.6

Small p-values indicate that the observed value of the test statistic would be unusual if the null hypothesis and probability model were true, and so provide evidence for the alternative. The lower the p-value, the more convincing the statistical evidence for the alternative hypothesis.

DAT-3.B.7

p-values that are not small indicate that the observed value of the test statistic would not be unusual if the null hypothesis and probability model were true, so do not provide convincing statistical evidence for the alternative hypothesis nor do they provide evidence that the null hypothesis is true.

DAT-3.B.8

A formal decision explicitly compares the p-value to the significance α. If the p-value $\leq \alpha$, then reject the null hypothesis, $H_0: p = p_0$. If the p-value $> \alpha$, then fail to reject the null hypothesis.

DAT-3.B.9

The results of a significance test for a population proportion can serve as the statistical reasoning to support the answer to a research question about the population that was sampled.

TOPIC 6.7

Potential Errors When Performing Tests

Required Course Content

ENDURING UNDERSTANDING

UNC-5

Probabilities of Type I and Type II errors influence inference.

LEARNING OBJECTIVE

UNC-5.A

Identify Type I and Type II errors. **[Skill 1.B]**

ESSENTIAL KNOWLEDGE

UNC-5.A.1

A Type I error occurs when the null hypothesis is true and is rejected (false positive).

UNC-5.A.2

A Type II error occurs when the null hypothesis is false and is not rejected (false negative).

Table of Errors

		Actual Population Value	
		H_0 true	**H_a true**
Decision	Reject H_0	Type I Error	Correct Decision
	Fail to Reject H_0	Correct Decision	Type II Error

continued on next page

SKILLS

Selecting Statistical Methods

1.B

Identify key and relevant information to answer a question or solve a problem.

Using Probability and Simulation

3.A

Determine relative frequencies, proportions, or probabilities using simulation or calculations.

Statistical Argumentation

4.A

Make an appropriate claim or draw an appropriate conclusion.

4.B

Interpret statistical calculations and findings to assign meaning or assess a claim.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-5.B Calculate the probability of a Type I and Type II errors. **[Skill 3.A]**	**UNC-5.B.1** The significance level, α, is the probability of making a Type I error, if the null hypothesis is true. **UNC-5.B.2** The power of a test is the probability that a test will correctly reject a false null hypothesis. **UNC-5.B.3** The probability of making a Type II error $= 1 - power$.
UNC-5.C Identify factors that affect the probability of errors in significance testing. **[Skill 4.A]**	**UNC-5.C.1** The probability of a Type II error decreases when any of the following occurs, provided the others do not change: i. Sample size(s) increases. ii. Significance level (α) of a test increases. iii. Standard error decreases. iv. True parameter value is farther from the null.
UNC-5.D Interpret Type I and Type II errors. **[Skill 4.B]**	**UNC-5.D.1** Whether a Type I or a Type II error is more consequential depends upon the situation. **UNC-5.D.2** Since the significance level, α, is the probability of a Type I error, the consequences of a Type I error influence decisions about a significance level.

TOPIC 6.8

Confidence Intervals for the Difference of Two Proportions

Required Course Content

ENDURING UNDERSTANDING

UNC-4

An interval of values should be used to estimate parameters, in order to account for uncertainty.

LEARNING OBJECTIVE

UNC-4.I

Identify an appropriate confidence interval procedure for a comparison of population proportions. **[Skill 1.D]**

ESSENTIAL KNOWLEDGE

UNC-4.I.1

The appropriate confidence interval procedure for a two-sample comparison of proportions for one categorical variable is a two-sample z-interval for a difference between population proportions.

LEARNING OBJECTIVE

UNC-4.J

Verify the conditions for calculating confidence intervals for a difference between population proportions. **[Skill 4.C]**

ESSENTIAL KNOWLEDGE

UNC-4.J.1

In order to calculate confidence intervals to estimate a difference between proportions, we must check for independence and that the sampling distribution is approximately normal:

a. To check for independence:
   i. Data should be collected using random samples or a randomized experiment.
   ii. When sampling without replacement, check that $n_1 \le 10\% N_1$ and $n_2 \le 10\% N_2$.

b. To check that sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal (shape).
   i. For categorical variables, check that $n_1\hat{p}_1$, $n_1(1-\hat{p}_1)$, $n_2\hat{p}_2$, and $n_2\left(1-\hat{p}_2\right)$ are all greater than or equal to some predetermined value, typically either 5 or 10.

continued on next page

SKILLS

Selecting Statistical Methods

1.D

Identify an appropriate inference method for confidence intervals.

Statistical Argumentation

4.C

Verify that inference procedures apply in a given situation.

Using Probability and Simulation

3.D

Construct a confidence interval, provided conditions for inference are met.

LEARNING OBJECTIVE

UNC-4.K

Calculate an appropriate confidence interval for a comparison of population proportions. **[Skill 3.D]**

ESSENTIAL KNOWLEDGE

UNC-4.K.1

For a comparison of proportions, the interval estimate is

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

CLARIFYING STATEMENT:

Formulas for interval estimates do not appear explicitly on the AP Statistics Formula Sheet provided with the AP Statistics Exam. However, these formulas do not need to be memorized, as they can be constructed based on the general test statistic formula and the relevant standard error formulas that are provided on the formula sheet.

UNC-4.L

Calculate an interval estimate based on a confidence interval for a difference of proportions. **[Skill 3.D]**

UNC-4.L.1

Confidence intervals for a difference in proportions can be used to calculate interval estimates with specified units.

TOPIC 6.9

Justifying a Claim Based on a Confidence Interval for a Difference of Population Proportions

SKILLS

Statistical Argumentation

4.B

Interpret statistical calculations and findings to assign meaning or assess a claim.

4.D

Justify a claim based on a confidence interval.

AVAILABLE RESOURCE

- Classroom Resource > **Calculations Aren't Enough! The Importance of Communication in AP Statistics**

Required Course Content

ENDURING UNDERSTANDING

UNC-4

An interval of values should be used to estimate parameters, in order to account for uncertainty.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-4.M Interpret a confidence interval for a difference of proportions. **[Skill 4.B]**	UNC-4.M.1 In repeated random sampling with the same sample size, approximately C% of confidence intervals created will capture the difference in population proportions. UNC-4.M.2 Interpreting a confidence interval for difference between population proportions should include a reference to the sample taken and details about the population it represents.
UNC-4.N Justify a claim based on a confidence interval for a difference of proportions. **[Skill 4.D]**	UNC-4.N.1 A confidence interval for difference in population proportions provides an interval of values that may provide sufficient evidence to support a particular claim in context.

SKILLS

Selecting Statistical Methods

1.F
Identify null and alternative hypotheses.

1.E
Identify an appropriate inference method for significance tests.

Statistical Argumentation

4.C
Verify that inference procedures apply in a given situation.

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

TOPIC 6.10
Setting Up a Test for the Difference of Two Population Proportion

Required Course Content

ENDURING UNDERSTANDING

VAR-6
The normal distribution may be used to model variation.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-6.H Identify the null and alternative hypotheses for a difference of two population proportions. **[Skill 1.F]**	**VAR-6.H.1** For a two-sample test for a difference of two proportions, the null hypothesis specifies a value of 0 for the difference in population proportions, indicating no difference or effect. **VAR-6.H.2** The null hypothesis for a difference in proportions is: $H_0 : p_1 = p_2$, or $H_0 : p_1 - p_2 = 0$. **VAR-6.H.3** A one-sided alternative hypothesis for a difference in proportions is $H_a : p_1 < p_2$, or, $H_a : p_1 > p_2$. A two-sided alternative hypothesis for a difference of proportions is $H_a : p_1 \neq p_2$.
VAR-6.I Identify an appropriate testing method for the difference of two population proportions. **[Skill 1.E]**	**VAR-6.I.1** For a single categorical variable, the appropriate testing method for the difference of two population proportions is a two-sample z-test for a difference between two population proportions.

continued on next page

LEARNING OBJECTIVE

VAR-6.J

Verify the conditions for making statistical inferences when testing a difference of two population proportions. **[Skill 4.C]**

ESSENTIAL KNOWLEDGE

VAR-6.J.1

In order to make statistical inferences when testing a difference between population proportions, we must check for independence and that the sampling distribution is approximately normal:

a. To check for independence:
   i. Data should be collected using random samples or a randomized experiment.
   ii. When sampling without replacement, check that $n_1 \le 10\% N_1$ and $n_2 \le 10\% N_2$.

b. To check that the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal (shape):
   i. For the combined sample, define the combined (or pooled) proportion, $\hat{p}_c = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$. Assuming that H_0 is true ($p_1 - p_2 = 0$ or $p_1 = p_2$), check that $n_1\hat{p}_c$, $n_1\left(1-\hat{p}_c\right)$, $n_2\hat{p}_c$, and $n_2\left(1-\hat{p}_c\right)$ are all greater than or equal to some predetermined value, typically either 5 or 10.

SKILLS

Using Probability and Simulation

3.E
Calculate a test statistic and find a p-value, provided conditions for inference are met.

Statistical Argumentation

4.B
Interpret statistical calculations and findings to assign meaning or assess a claim.

4.E
Justify a claim using a decision based on significance tests.

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

TOPIC 6.11

Carrying Out a Test for the Difference of Two Population Proportions

Required Course Content

ENDURING UNDERSTANDING

VAR-6
The normal distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-6.K
Calculate an appropriate test statistic for the difference of two population proportions. **[Skill 3.E]**

ESSENTIAL KNOWLEDGE

VAR-6.K.1
The test statistic for a difference in proportions is:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}_c(1-\hat{p}_c)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\hat{p}_c = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$.

CLARIFYING STATEMENT:
The formulas for test statistics do not appear explicitly on the AP Statistics Formula Sheet provided with the AP Statistics Exam. However, these formulas do not need to be memorized, as they can be constructed based on the general test statistic formula and the standard error formulas for each of the relevant test statistics that are provided on the formula sheet.

continued on next page

ENDURING UNDERSTANDING

DAT-3

Significance testing allows us to make decisions about hypotheses within a particular context.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
DAT-3.C Interpret the p-value of a significance test for a difference of population proportions. **[Skill 4.B]**	**DAT-3.C.1** An interpretation of the p-value of a significance test for a difference of two population proportions should recognize that the p-value is computed by assuming that the null hypothesis is true, i.e., by assuming that the true population proportions are equal to each other.
DAT-3.D Justify a claim about the population based on the results of a significance test for a difference of population proportions. **[Skill 4.E]**	**DAT-3.D.1** A formal decision explicitly compares the p-value to the significance α. If the p-value $\leq \alpha$, then reject the null hypothesis, $H_0: p_1 = p_2$, or $H_0: p_1 - p_2 = 0$. If the p-value $> \alpha$, then fail to reject the null hypothesis. **DAT-3.D.2** The results of a significance test for a difference of two population proportions can serve as the statistical reasoning to support the answer to a research question about the two populations that were sampled.