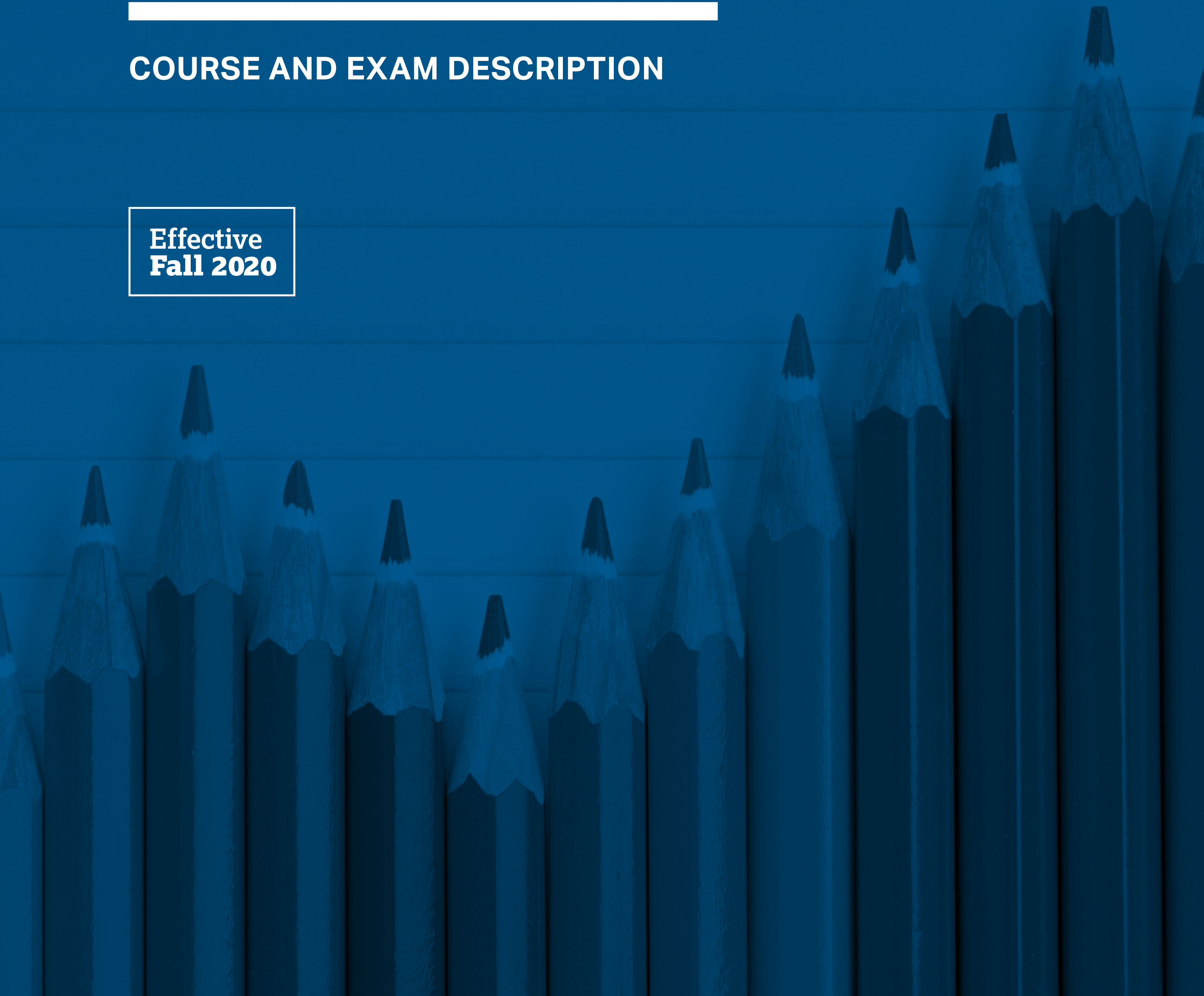
CollegeBoard AP

INCLUDES
- Course framework
- Instructional section
- Sample exam questions
- Classroom poster

AP® Statistics

COURSE AND EXAM DESCRIPTION

Effective Fall 2020



AP® CollegeBoard

AP® Statistics

COURSE AND EXAM DESCRIPTION

Effective
Fall 2020

AP COURSE AND EXAM DESCRIPTIONS ARE UPDATED PERIODICALLY

Please visit AP Central (**apcentral.collegeboard.org**) to determine whether a more recent course and exam description is available.

About College Board

College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success—including the SAT® and the Advanced Placement® Program. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools.

For further information, visit **collegeboard.org**.

AP Equity and Access Policy

College Board strongly encourages educators to make equitable access a guiding principle for their AP programs by giving all willing and academically prepared students the opportunity to participate in AP. We encourage the elimination of barriers that restrict access to AP for students from ethnic, racial, and socioeconomic groups that have been traditionally underrepresented. Schools should make every effort to ensure their AP classes reflect the diversity of their student population. College Board also believes that all students should have access to academically challenging coursework before they enroll in AP classes, which can prepare them for AP success. It is only through a commitment to equitable preparation and access that true equity and excellence can be achieved.

Designers: Sonny Mui and Bill Tully

Visit College Board on the web: **collegeboard.org**.

Contents

THIS PAGE IS INTENTIONALLY LEFT BLANK.

Acknowledgments

College Board would like to acknowledge the following committee members, consultants, and reviewers for their assistance with and commitment to the development of this course. All individuals and their affiliations were current at the time of contribution.

Ellen Breazel, *Clemson University, Clemson, SC*

Christy Brown, *Clemson University, Clemson, SC*

Paul Buckley, *Gonzaga College High School, Washington, DC*

Mine Cetinkaya-Rundel, *Duke University, Durham, NC*

Jeff Eicher, Jr., *Classical Academy High School, Escondido, CA*

Kerri Swails Freeland, *University High School, Morgantown, WV*

Kenneth Koehler, *Iowa State University, Ames, IA*

Michael Lacey, *Peters Township High School, McMurray, PA*

Laura Marshall, *Phillips Exeter Academy, Exeter, NH*

S. Leigh Nataro, *Kent Place School, Summit, NJ*

Kathleen Petko, *Palatine High School, Palatine, IL*

Paul Rodriguez, *Troy High School, Fullerton, CA*

Penny Smeltzer, *Austin Peace Academy, Austin, TX*

David Spohn, *Hudson High School, Hudson, OH*

Daren Starnes, *The Lawrenceville School, Lawrenceville, NJ*

Robert Stephenson, *Iowa State University, Ames, IA*

Jessica Utts, *University of California, Irvine, Irvine, CA*

Adam Yankay, *Western Reserve Academy, Hudson, OH*

College Board Staff

Sara Hunter, *Associate Director, AP Curricular Publications*

Tiffany Judkins, *Director, AP Instructional Design and PD Resource Development*

Claire Lorenz, *Senior Director, AP Instructional Design and PD Resource Development*

Daniel McDonough, *Senior Director, AP Content Integration*

Stephanie Ogden, *Director, AP Mathematics Content Development*

SPECIAL THANKS

John R. Williamson, Brian Robinson, and Benjamin Hedrick

THIS PAGE IS INTENTIONALLY LEFT BLANK.

AP STATISTICS

UNIT 1

Exploring One-Variable Data

15–23%
AP EXAM WEIGHTING

~14–16
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 1

Multiple-choice: ~35 questions
Free-response: 2 questions

- Exploring Data
- Exploring Data

UNIT 1

15–23% AP EXAM WEIGHTING

~14–16 CLASS PERIODS

Exploring One-Variable Data

BIG IDEA 1
Variation and Distribution **VAR**

- Is my cat old, compared to other cats?

BIG IDEA 2
Patterns and Uncertainty **UNC**

- How certain are we that what seems to be a pattern is not just a coincidence?

Developing Understanding

Unit 1 introduces students to data and the vocabulary of statistics. Students also learn to talk about data in real-world contexts. Variability in data may seem to suggest certain conclusions about the data distribution, but not all variation is meaningful. Statistics allows us to develop shared understandings of uncertainty and variation. In this unit, students will define and represent categorical and quantitative variables, describe and compare distributions of one-variable data, and interpret statistical calculations to assess claims about individual data points or samples. Students will also begin to apply the normal distribution model as an introduction to how theoretical models for populations can be used to describe some distributions of sample data. Later units will more fully develop probabilistic modeling and inference.

Building Course Skills

2.A 2.B 2.D

Having access to a world of data is meaningless without the ability to organize and analyze that information. To develop these skills, students will need multiple opportunities to interact with data presented in different formats, i.e., as a table, a graph, or even just a list of values. Students should be asked to verbally describe the patterns and characteristics they see in the data (including shape, center, variability, and unusual features for a quantitative variable) and then compare the characteristics of two different sets of data. Students should also create displays that appropriately represent the data (e.g., using a bar graph for categorical data).

Teachers can provide explicit feedback on students' verbal responses so they understand the level of detail needed. For example, when students are asked to describe a distribution of quantitative data, they often provide an acronym associated with that type of distribution (e.g., SOCS or CUSS) but then struggle to discuss all the elements the acronym stands for. In particular, students often neglect to discuss unusual features such as gaps or outliers. Teachers can reinforce that these elements must be addressed in their descriptions and that all data has context (e.g., the variable of interest, including any units of measurement).

Preparing for the AP Exam

In preparation for the AP Exam, teachers can encourage students to carefully read each question and completely answer the question asked. When interpreting representations of quantitative data, for example, students should describe shape, center, and variability, as well as unusual features, such as outliers. A response focused only on the center, for example, would be considered incomplete. Students should also provide complete explanations in context for all conclusions made from data. If asked to justify the selection of a particular conclusion over other options, students should include both a reasoning for their choice and rationales for not choosing the others.

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods ~14–16 CLASS PERIODS
VAR-1	1.1 Introducing Statistics: What Can We Learn from Data?	1.A Identify the question to be answered or problem to be solved *(not assessed)*.	
VAR-1	1.2 The Language of Variation: Variables	2.A Describe data presented numerically or graphically.	
UNC-1	1.3 Representing a Categorical Variable with Tables	2.B Construct numerical or graphical representations of distributions. 2.A Describe data presented numerically or graphically.	
UNC-1	1.4 Representing a Categorical Variable with Graphs	2.B Construct numerical or graphical representations of distributions. 2.A Describe data presented numerically or graphically. 2.D Compare distributions or relative positions of points within a distribution.	
UNC-1	1.5 Representing a Quantitative Variable with Graphs	2.A Describe data presented numerically or graphically. 2.B Construct numerical or graphical representations of distributions.	
UNC-1	1.6 Describing the Distribution of a Quantitative Variable	2.A Describe data presented numerically or graphically.	
UNC-1	1.7 Summary Statistics for a Quantitative Variable	2.C Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
UNC-1	1.8 Graphical Representations of Summary Statistics	2.B Construct numerical or graphical representations of distributions. 2.A Describe data presented numerically or graphically.	
UNC-1	1.9 Comparing Distributions of a Quantitative Variable	2.D Compare distributions or relative positions of points within a distribution.	
VAR-2	1.10 The Normal Distribution	2.D Compare distributions or relative positions of points within a distribution. 3.A Determine relative frequencies, proportions, or probabilities using simulation or calculations.	

AP

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 1. 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	1.5	**Gallery Walk** Have students work in groups of four to construct a dotplot, a stem-and-leaf plot, a histogram, or a boxplot for a set of student-generated data (e.g., time in minutes to get to school). After the gallery walk, discuss what information can be seen more easily in each graph (e.g., boxplots can easily show the IQR).
2	1.6 1.8	**FRQ Partner Quiz** Have students work in pairs to answer **2017 FRQ 4**. Have one student write and the other perform the calculations. (Although the first part of the question does not require any calculations, the second part requires calculations to justify the solution.) Discussing and crafting a solution with a partner may require more time than if students completed the FRQ individually.
3	1.9	**Notice and Wonder** Display just the graphs from **2018 FRQ 5**. Have students think individually for one minute about how the graphs compare. Then ask them, "What do you notice? What do you wonder? What questions could be answered with these graphs?" Have students share their ideas with a partner then debrief the ideas as a class.
4	1.10	**Reversing Interpretations** Give pairs of students four pictures of normal distributions with various parts shaded. Have students create the question that could have resulted in the picture shown (e.g., if a value of 15 is labeled and the distribution is shaded to the right of 15, students could write "What is the probability that a value is more than 15?").

SKILL

Selecting Statistical Methods

1.A

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

AVAILABLE RESOURCE

- Classroom Resource > **Coke® Versus Pepsi®: An Introductory Activity for Test of Significance** (may be used in Topic 1.1 to introduce the course or in Topic 6.4 to introduce inference tests)

TOPIC 1.1

Introducing Statistics: What Can We Learn from Data?

Required Course Content

ENDURING UNDERSTANDING

VAR-1

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

LEARNING OBJECTIVE

VAR-1.A

Identify questions to be answered, based on variation in one-variable data.
[Skill 1.A]

ESSENTIAL KNOWLEDGE

VAR-1.A.1

Numbers may convey meaningful information, when placed in context.

TOPIC 1.2

The Language of Variation: Variables

SKILL

Data Analysis

2.A

Describe data presented numerically or graphically.

Required Course Content

ENDURING UNDERSTANDING

VAR-1

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

LEARNING OBJECTIVE

VAR-1.B

Identify variables in a set of data. **[Skill 2.A]**

ESSENTIAL KNOWLEDGE

VAR-1.B.1

A variable is a characteristic that changes from one individual to another.

LEARNING OBJECTIVE

VAR-1.C

Classify types of variables. **[Skill 2.A]**

ESSENTIAL KNOWLEDGE

VAR-1.C.1

A categorical variable takes on values that are category names or group labels.

VAR-1.C.2

A quantitative variable is one that takes on numerical values for a measured or counted quantity.

ILLUSTRATIVE EXAMPLES

Categorical variables:

- Dominant hand
- Age group (young or old)
- Highest degree earned

Quantitative variables:

- Age of a structure
- Height of a child
- Concentration of a sample

SKILLS

Data Analysis

2.B

Construct numerical or graphical representations of distributions.

2.A

Describe data presented numerically or graphically.

TOPIC 1.3

Representing a Categorical Variable with Tables

Required Course Content

ENDURING UNDERSTANDING

UNC-1

Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-1.A Represent categorical data using frequency or relative frequency tables. **[Skill 2.B]**	UNC-1.A.1 A frequency table gives the number of cases falling into each category. A relative frequency table gives the proportion of cases falling into each category.
UNC-1.B Describe categorical data represented in frequency or relative tables. **[Skill 2.A]**	UNC-1.B.1 Percentages, relative frequencies, and rates all provide the same information as proportions. UNC-1.B.2 Counts and relative frequencies of categorical data reveal information that can be used to justify claims about the data in context.

TOPIC 1.4

Representing a Categorical Variable with Graphs

Required Course Content

ENDURING UNDERSTANDING

UNC-1

Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-1.C Represent categorical data graphically. **[Skill 2.B]**	**UNC-1.C.1** Bar charts (or bar graphs) are used to display frequencies (counts) or relative frequencies (proportions) for categorical data. **UNC-1.C.2** The height or length of each bar in a bar graph corresponds to either the number or proportion of observations falling within each category. **UNC-1.C.3** There are many additional ways to represent frequencies (counts) or relative frequencies (proportions) for categorical data.
UNC-1.D Describe categorical data represented graphically. **[Skill 2.A]**	**UNC-1.D.1** Graphical representations of a categorical variable reveal information that can be used to justify claims about the data in context.
UNC-1.E Compare multiple sets of categorical data. **[Skill 2.D]**	**UNC-1.E.1** Frequency tables, bar graphs, or other representations can be used to compare two or more data sets in terms of the same categorical variable.

SKILLS

Data Analysis

2.B Construct numerical or graphical representations of distributions.

2.A Describe data presented numerically or graphically.

2.D Compare distributions or relative positions of points within a distribution.

SKILLS

Data Analysis

2.A
Describe data presented numerically or graphically.

2.B
Construct numerical or graphical representations of distributions.

ILLUSTRATIVE EXAMPLES

A discrete variable:

- Number of students in a class

A continuous variable:

- Height of a child

TOPIC 1.5
Representing a Quantitative Variable with Graphs

Required Course Content

ENDURING UNDERSTANDING

UNC-1
Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-1.F Classify types of quantitative variables. **[Skill 2.A]**	UNC-1.F.1 A discrete variable can take on a countable number of values. The number of values may be finite or countably infinite, as with the counting numbers. UNC-1.F.2 A continuous variable can take on infinitely many values, but those values cannot be counted. No matter how small the interval between two values of a continuous variable, it is always possible to determine another value between them.
UNC-1.G Represent quantitative data graphically. **[Skill 2.B]**	UNC-1.G.1 In a histogram, the height of each bar shows the number or proportion of observations that fall within the interval corresponding to that bar. Altering the interval widths can change the appearance of the histogram. UNC-1.G.2 In a stem and leaf plot, each data value is split into a "stem" (the first digit or digits) and a "leaf" (usually the last digit).

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-1.G Represent quantitative data graphically. **[Skill 2.B]**	**UNC-1.G.3** A dotplot represents each observation by a dot, with the position on the horizontal axis corresponding to the data value of that observation, with nearly identical values stacked on top of each other. **UNC-1.G.4** A cumulative graph represents the number or proportion of a data set less than or equal to a given number. **UNC-1.G.5** There are many additional ways to graphically represent distributions of quantitative data.

SKILL

Data Analysis

2.A

Describe data presented numerically or graphically.

TOPIC 1.6

Describing the Distribution of a Quantitative Variable

Required Course Content

ENDURING UNDERSTANDING

UNC-1

Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE

UNC-1.H

Describe the characteristics of quantitative data distributions. **[Skill 2.A]**

ESSENTIAL KNOWLEDGE

UNC-1.H.1

Descriptions of the distribution of quantitative data include shape, center, and variability (spread), as well as any unusual features such as outliers, gaps, clusters, or multiple peaks.

UNC-1.H.2

Outliers for one-variable data are data points that are unusually small or large relative to the rest of the data.

UNC-1.H.3

A distribution is skewed to the right (positive skew) if the right tail is longer than the left.
A distribution is skewed to the left (negative skew) if the left tail is longer than the right.
A distribution is symmetric if the left half is the mirror image of the right half.

UNC-1.H.4

Univariate graphs with one main peak are known as unimodal. Graphs with two prominent peaks are bimodal. A graph where each bar height is approximately the same (no prominent peaks) is approximately uniform.

UNC-1.H.5

A gap is a region of a distribution between two data values where there are no observed data.

continued on next page

LEARNING OBJECTIVE

UNC-1.H

Describe the characteristics of quantitative data distributions. **[Skill 2.A]**

ESSENTIAL KNOWLEDGE

UNC-1.H.6

Clusters are concentrations of data usually separated by gaps.

UNC-1.H.7

Descriptive statistics does not attribute properties of a data set to a larger population, but may provide the basis for conjectures for subsequent testing.

SKILLS

Data Analysis

2.C

Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response.

Statistical Argumentation

4.B

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

TOPIC 1.7

Summary Statistics for a Quantitative Variable

Required Course Content

ENDURING UNDERSTANDING

UNC-1

Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE

UNC-1.I

Calculate measures of center and position for quantitative data. **[Skill 2.C]**

ESSENTIAL KNOWLEDGE

UNC-1.I.1

A statistic is a numerical summary of sample data.

UNC-1.I.2

The mean is the sum of all the data values divided by the number of values. For a sample, the mean is denoted by x-bar: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, where x_i represents the i^{th} data point in the sample and n represents the number of data values in the sample.

UNC-1.I.3

The median of a data set is the middle value when data are ordered. When the number of data points is even, the median can take on any value between the two middle values. In AP Statistics, the most commonly used value for the median of a data set with an even number of values is the average of the two middle values.

UNC-1.I.4

The first quartile, Q1, is the median of the half of the ordered data set from the minimum to the position of the median. The third quartile, Q3, is the median of the half of the ordered data set from the position of the median to the maximum. Q1 and Q3 form the boundaries for the middle 50% of values in an ordered data set.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-1.I Calculate measures of center and position for quantitative data. **[Skill 2.C]**	**UNC-1.I.5** The p^{th} percentile is interpreted as the value that has $p\%$ of the data less than or equal to it.
UNC-1.J Calculate measures of variability for quantitative data. **[Skill 2.C]**	**UNC-1.J.1** Three commonly used measures of variability (or spread) in a distribution are the range, interquartile range, and standard deviation. **UNC-1.J.2** The range is defined as the difference between the maximum data value and the minimum data value. The interquartile range (IQR) is defined as the difference between the third and first quartiles: Q3 − Q1. Both the range and the interquartile range are possible ways of measuring variability of the distribution of a quantitative variable. **UNC-1.J.3** Standard deviation is a way to measure variability of the distribution of a quantitative variable. For a sample, the standard deviation is denoted by s: $s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$. The square of the sample standard deviation, s^2, is called the sample variance. **UNC-1.J.4** Changing units of measurement affects the values of the calculated statistics.
UNC-1.K Explain the selection of a particular measure of center and/or variability for describing a set of quantitative data. **[Skill 4.B]**	**UNC-1.K.1** There are many methods for determining outliers. Two methods frequently used in this course are: i. An outlier is a value greater than $1.5 \times \text{IQR}$ above the third quartile or more than $1.5 \times \text{IQR}$ below the first quartile. ii. An outlier is a value located 2 or more standard deviations above, or below, the mean. **UNC-1.K.2** The mean, standard deviation, and range are considered nonresistant (or non-robust) because they are influenced by outliers. The median and IQR are considered resistant (or robust), because outliers do not greatly (if at all) affect their value.

SKILLS

Data Analysis

2.B
Construct numerical or graphical representations of distributions.

2.A
Describe data presented numerically or graphically.

TOPIC 1.8

Graphical Representations of Summary Statistics

Required Course Content

ENDURING UNDERSTANDING

UNC-1
Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-1.L Represent summary statistics for quantitative data graphically. **[Skill 2.B]**	**UNC-1.L.1** Taken together, the minimum data value, the first quartile (Q1), the median, the third quartile (Q3), and the maximum data value make up the five-number summary. **UNC-1.L.2** A boxplot is a graphical representation of the five-number summary (minimum, first quartile, median, third quartile, maximum). The box represents the middle 50% of data, with a line at the median and the ends of the box corresponding to the quartiles. Lines ("whiskers") extend from the quartiles to the most extreme point that is not an outlier, and outliers are indicated by their own symbol beyond this.
UNC-1.M Describe summary statistics of quantitative data represented graphically. **[Skill 2.A]**	**UNC-1.M.1** Summary statistics of quantitative data, or of sets of quantitative data, can be used to justify claims about the data in context. **UNC-1.M.2** If a distribution is relatively symmetric, then the mean and median are relatively close to one another. If a distribution is skewed right, then the mean is usually to the right of the median. If the distribution is skewed left, then the mean is usually to the left of the median.

TOPIC 1.9

Comparing Distributions of a Quantitative Variable

SKILL

Data Analysis

2.D

Compare distributions or relative positions of points within a distribution.

Required Course Content

ENDURING UNDERSTANDING

UNC-1

Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-1.N Compare graphical representations for multiple sets of quantitative data. **[Skill 2.D]**	**UNC-1.N.1** Any of the graphical representations, e.g., histograms, side-by-side boxplots, etc., can be used to compare two or more independent samples on center, variability, clusters, gaps, outliers, and other features.
UNC-1.O Compare summary statistics for multiple sets of quantitative data. **[Skill 2.D]**	**UNC-1.O.1** Any of the numerical summaries (e.g., mean, standard deviation, relative frequency, etc.) can be used to compare two or more independent samples.

SKILLS

Data Analysis

2.D
Compare distributions or relative positions of points within a distribution.

Using Probability and Simulation

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

ILLUSTRATIVE EXAMPLES

Variables that can by modeled by a normal distribution:

- Body temperature
- Weight of a loaf of bread

TOPIC 1.10

The Normal Distribution

Required Course Content

ENDURING UNDERSTANDING

VAR-2

The normal distribution can be used to represent some population distributions.

LEARNING OBJECTIVE

VAR-2.A

Compare a data distribution to the normal distribution model. **[Skill 2.D]**

ESSENTIAL KNOWLEDGE

VAR-2.A.1

A parameter is a numerical summary of a population.

VAR-2.A.2

Some sets of data may be described as approximately normally distributed. A normal curve is mound-shaped and symmetric. The parameters of a normal distribution are the population mean, μ, and the population standard deviation, σ.

VAR-2.A.3

For a normal distribution, approximately 68% of the observations are within 1 standard deviation of the mean, approximately 95% of observations are within 2 standard deviations of the mean, and approximately 99.7% of observations are within 3 standard deviations of the mean. This is called the empirical rule.

VAR-2.A.4

Many variables can be modeled by a normal distribution.

continued on next page

LEARNING OBJECTIVE

VAR-2.B

Determine proportions and percentiles from a normal distribution. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

VAR-2.B.1

A standardized score for a particular data value is calculated as (data value – mean)/(standard deviation), and measures the number of standard deviations a data value falls above or below the mean.

VAR-2.B.2

One example of a standardized score is a z-score, which is calculated as $z\text{-score} = \left(\frac{x_i - \mu}{\sigma}\right)$. A z-score measures how many standard deviations a data value is from the mean.

VAR-2.B.3

Technology, such as a calculator, a standard normal table, or computer-generated output, can be used to find the proportion of data values located on a given interval of a normally distributed random variable.

VAR-2.B.4

Given the area of a region under the graph of the normal distribution curve, it is possible to use technology, such as a calculator, a standard normal table, or computer-generated output, to estimate parameters for some populations.

LEARNING OBJECTIVE

VAR-2.C

Compare measures of relative position in data sets. **[Skill 2.D]**

ESSENTIAL KNOWLEDGE

VAR-2.C.1

Percentiles and z-scores may be used to compare relative positions of points within a data set or between data sets.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

AP STATISTICS

UNIT 2

Exploring Two-Variable Data

5–7%
AP EXAM WEIGHTING

~10–11
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 2

Multiple-choice: ~35 questions
Free-response: 2 questions

- Exploring Data
- Investigative Task

UNIT 2

5–7% AP EXAM WEIGHTING

~10–11 CLASS PERIODS

Exploring Two-Variable Data

BIG IDEA 1
Variation and Distribution **VAR**
- Does the fact that the number of shark attacks increases with ice cream sales necessarily mean that ice cream sales cause shark attacks?

BIG IDEA 2
Patterns and Uncertainty **UNC**
- How might you represent incomes of individuals with and without a college degree to help describe similarities and/or differences between the two groups?

BIG IDEA 3
Data-Based Predictions, Decisions, and Conclusions **DAT**
- How can you determine the effectiveness of a linear model that uses the number of cricket chirps per minute to predict temperature?

Developing Understanding

Building on Unit 1, students will explore relationships in two-variable categorical or quantitative data sets. They will use graphical and numerical methods to investigate an association between two categorical variables. Skills learned while working with two-way tables will transfer to calculating probabilities in Unit 4.

Students will describe form, direction, strength, and unusual features for an association between two quantitative variables. They will assess correlation and, if appropriate, use a linear model to predict values of the response variable from values of the explanatory variable. Students will interpret the least-squares regression line in context, analyze prediction errors (residuals), and explore departures from a linear pattern.

Building Course Skills

2.C 2.D 4.B

In Unit 2, students are looking at the relationship between variables. The ability to calculate and describe statistical values, such as a conditional relative frequency or the slope of a regression line, is critical for data analysis because students must be able to analyze patterns before drawing conclusions about the data. Students should be allowed to perform their calculations using technology to help them become more aware of procedural errors. Students will also need practice translating output from technology ("calculator speak") into appropriate statistical language.

As any statistician will assert, a numerical calculation is only as good as one's ability to interpret what it means in the real world. Rather than just reporting values from their calculations, students must be able to connect their numerical results to the scenario's context and formulate a verbal response that makes that connection clear. Teachers can model good communication and provide high-quality feedback to help students use accurate statistical language when comparing side-by-side bar graphs, for example, and to avoid common errors in reasoning, such as using the word "line" to explain why a relationship is linear.

Preparing for the AP Exam

Students need ongoing practice with interpretation of vocabulary and calculated values in context. It is typically not sufficient to speak generally about the direction of a relationship, for example. If the question is about a linear model for predicting the weight of a wolf based on its length, students should write that a positive relationship means that longer wolves tend to have higher weights (see **2017 FRQ 1**). Students can communicate statistical uncertainty by using words such as "tend to have" and "on average," being careful to be precise with language. For example, when explaining evidence of a linear relationship, the difference between discussing a *rate* of change, as opposed to a change, is the difference between right and wrong. For the sake of clarity, the word "correlation" should be reserved for discussions about relationships between two quantitative variables.

UNIT 2

Exploring Two-Variable Data

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods ~10–11 CLASS PERIODS
VAR-1	**2.1 Introducing Statistics: Are Variables Related?**	1.A Identify the question to be answered or problem to be solved *(not assessed)*.	
UNC-1	**2.2 Representing Two Categorical Variables**	2.D Compare distributions or relative positions of points within a distribution.	
UNC-1	**2.3 Statistics for Two Categorical Variables**	2.C Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response. 2.D Compare distributions or relative positions of points within a distribution.	
UNC-1, DAT-1	**2.4 Representing the Relationship Between Two Quantitative Variables**	2.B Construct numerical or graphical representations of distributions. 2.A Describe data presented numerically or graphically.	
DAT-1	**2.5 Correlation**	2.C Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
DAT-1	**2.6 Linear Regression Models**	2.C Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response.	
DAT-1	**2.7 Residuals**	2.B Construct numerical or graphical representations of distributions. 2.A Describe data presented numerically or graphically.	
DAT-1	**2.8 Least Squares Regression**	2.C Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
DAT-1	**2.9 Analyzing Departures from Linearity**	2.A Describe data presented numerically or graphically. 2.C Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response.	

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 2. 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	2.5 2.8	**Quickwrite** Give students a scatterplot and its associated computer output. Have them identify and describe the meaning of the following values in the context of the problem: slope, y-intercept, coefficient of determination, and standard error of the residuals. Also have them calculate the correlation and explain how they found it. Have students compare their write-ups in groups of three to four.
2	2.7	**Reversing Interpretations** Instead of asking students to interpret a residual, give them the residual and the equation of the least-squares regression line and ask them to make a prediction for a particular observation (e.g., "One wolf in the pack had a length of 1.4 m and a residual of −9.87. What does that −9.87 tell us about that particular wolf?")
3	2.8	**Build the Model Solution** Provide students with strips of paper containing portions of the model solution for **2018 FRQ 1** and have them work to assemble the phrases into a solution for the FRQ. Words can be grouped for part a, as follows: [The estimate of the intercept is] [72.95]. [It is] [estimated that] [the average time to] [finish checkout] [if there are no other customers in line] [is 72.95 seconds]. Additional numbers or phrases for part a could include [174.40], [is 174.50 seconds], and [the time to].
4	2.9	**Predict and Confirm** Have students toss a handful of M&Ms and record how many land M side up. This is trial 1. Then have them remove the ones that were M side up. For trial 2, have students toss the remaining candies (the ones left over after removing the ones that landed M side up) and record how many land M side up on the second toss. Ask students to think about the trend and make a prediction: Will it be linear? A scatterplot of many trials should show a nonlinear relationship.

SKILL

Selecting Statistical Methods

1.A

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

TOPIC 2.1

Introducing Statistics: Are Variables Related?

Required Course Content

ENDURING UNDERSTANDING

VAR-1

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

LEARNING OBJECTIVE

VAR-1.D

Identify questions to be answered about possible relationships in data.
[Skill 1.A]

ESSENTIAL KNOWLEDGE

VAR-1.D.1

Apparent patterns and associations in data may be random or not.

TOPIC 2.2

Representing Two Categorical Variables

SKILL

Data Analysis

2.D

Compare distributions or relative positions of points within a distribution.

Required Course Content

ENDURING UNDERSTANDING

UNC-1

Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE

UNC-1.P

Compare numerical and graphical representations for two categorical variables. **[Skill 2.D]**

ESSENTIAL KNOWLEDGE

UNC-1.P.1

Side-by-side bar graphs, segmented bar graphs, and mosaic plots are examples of bar graphs for one categorical variable, broken down by categories of another categorical variable.

UNC-1.P.2

Graphical representations of two categorical variables can be used to compare distributions and/or determine if variables are associated.

UNC-1.P.3

A two-way table, also called a contingency table, is used to summarize two categorical variables. The entries in the cells can be frequency counts or relative frequencies.

UNC-1.P.4

A joint relative frequency is a cell frequency divided by the total for the entire table.

SKILLS

Data Analysis

2.C
Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response.

2.D
Compare distributions or relative positions of points within a distribution.

TOPIC 2.3
Statistics for Two Categorical Variables

Required Course Content

ENDURING UNDERSTANDING

UNC-1
Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-1.Q Calculate statistics for two categorical variables. **[Skill 2.C]**	**UNC-1.Q.1** The marginal relative frequencies are the row and column totals in a two-way table divided by the total for the entire table. **UNC-1.Q.2** A conditional relative frequency is a relative frequency for a specific part of the contingency table (e.g., cell frequencies in a row divided by the total for that row).
UNC-1.R Compare statistics for two categorical variables. **[Skill 2.D]**	**UNC-1.R.1** Summary statistics for two categorical variables can be used to compare distributions and/or determine if variables are associated.

TOPIC 2.4

Representing the Relationship Between Two Quantitative Variables

SKILLS

Data Analysis

2.B
Construct numerical or graphical representations of distributions.

2.A
Describe data presented numerically or graphically.

Required Course Content

ENDURING UNDERSTANDING

UNC-1
Graphical representations and statistics allow us to identify and represent key features of data.

LEARNING OBJECTIVE

UNC-1.S
Represent bivariate quantitative data using scatterplots. **[Skill 2.B]**

ESSENTIAL KNOWLEDGE

UNC-1.S.1
A bivariate quantitative data set consists of observations of two different quantitative variables made on individuals in a sample or population.

UNC-1.S.2
A scatterplot shows two numeric values for each observation, one corresponding to the value on the x-axis and one corresponding to the value on the y-axis.

UNC-1.S.3
An explanatory variable is a variable whose values are used to explain or predict corresponding values for the response variable.

continued on next page

ENDURING UNDERSTANDING

DAT-1

Regression models may allow us to predict responses to changes in an explanatory variable.

LEARNING OBJECTIVE

DAT-1.A

Describe the characteristics of a scatter plot. **[Skill 2.A]**

ESSENTIAL KNOWLEDGE

DAT-1.A.1

A description of a scatter plot includes form, direction, strength, and unusual features.

DAT-1.A.2

The direction of the association shown in a scatterplot, if any, can be described as positive or negative.

DAT-1.A.3

A positive association means that as values of one variable increase, the values of the other variable tend to increase. A negative association means that as values of one variable increase, values of the other variable tend to decrease.

DAT-1.A.4

The form of the association shown in a scatterplot, if any, can be described as linear or non-linear to varying degrees.

DAT-1.A.5

The strength of the association is how closely the individual points follow a specific pattern, e.g., linear, and can be shown in a scatterplot. Strength can be described as strong, moderate, or weak.

DAT-1.A.6

Unusual features of a scatter plot include clusters of points or points with relatively large discrepancies between the value of the response variable and a predicted value for the response variable.

TOPIC 2.5

Correlation

Required Course Content

ENDURING UNDERSTANDING

DAT-1

Regression models may allow us to predict responses to changes in an explanatory variable.

LEARNING OBJECTIVE

DAT-1.B

Determine the correlation for a linear relationship. **[Skill 2.C]**

ESSENTIAL KNOWLEDGE

DAT-1.B.1

The correlation, r, gives the direction and quantifies the strength of the linear association between two quantitative variables.

DAT-1.B.2

The correlation coefficient can be calculated by:

$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$. However, the most common way to determine r is by using technology.

DAT-1.B.3

A correlation coefficient close to 1 or −1 does not necessarily mean that a linear model is appropriate.

continued on next page

SKILLS

Data Analysis

2.C

Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response.

Statistical Argumentation

4.B

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

LEARNING OBJECTIVE

DAT-1.C

Interpret the correlation for a linear relationship. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

DAT-1.C.1

The correlation, r, is unit-free, and always between −1 and 1, inclusive. A value of $r = 0$ indicates that there is no linear association. A value of $r = 1$ or $r = -1$ indicates that there is a perfect linear association.

DAT-1.C.2

A perceived or real relationship between two variables does not mean that changes in one variable cause changes in the other. That is, correlation does not necessarily imply causation.

TOPIC 2.6

Linear Regression Models

SKILL

Data Analysis

2.C

Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response.

Required Course Content

ENDURING UNDERSTANDING

DAT-1

Regression models may allow us to predict responses to changes in an explanatory variable.

LEARNING OBJECTIVE

DAT-1.D

Calculate a predicted response value using a linear regression model. **[Skill 2.C]**

ESSENTIAL KNOWLEDGE

DAT-1.D.1

A simple linear regression model is an equation that uses an explanatory variable, x, to predict the response variable, y.

DAT-1.D.2

The predicted response value, denoted by $\hat{y}$, is calculated as $\hat{y} = a + bx$, where a is the y-intercept and b is the slope of the regression line, and x is the value of the explanatory variable.

DAT-1.D.3

Extrapolation is predicting a response value using a value for the explanatory variable that is beyond the interval of x-values used to determine the regression line. The predicted value is less reliable as an estimate the further we extrapolate.

SKILLS

Data Analysis

2.B

Construct numerical or graphical representations of distributions.

2.A

Describe data presented numerically or graphically.

TOPIC 2.7

Residuals

Required Course Content

ENDURING UNDERSTANDING

DAT-1

Regression models may allow us to predict responses to changes in an explanatory variable.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
DAT-1.E Represent differences between measured and predicted responses using residual plots. **[Skill 2.B]**	**DAT-1.E.1** The residual is the difference between the actual value and the predicted value: residual $= y - \hat{y}$. **DAT-1.E.2** A residual plot is a plot of residuals versus explanatory variable values or predicted response values.
DAT-1.F Describe the form of association of bivariate data using residual plots. **[Skill 2.A]**	**DAT-1.F.1** Apparent randomness in a residual plot for a linear model is evidence of a linear form to the association between the variables. **DAT-1.F.2** Residual plots can be used to investigate the appropriateness of a selected model.

TOPIC 2.8

Least Squares Regression

Required Course Content

ENDURING UNDERSTANDING

DAT-1

Regression models may allow us to predict responses to changes in an explanatory variable.

LEARNING OBJECTIVE

DAT-1.G

Estimate parameters for the least-squares regression line model. **[Skill 2.C]**

ESSENTIAL KNOWLEDGE

DAT-1.G.1

The least-squares regression model minimizes the sum of the squares of the residuals and contains the point $(\bar{x}, \bar{y})$.

DAT-1.G.2

The slope, b, of the regression line can be calculated as $b = r\left(\frac{s_y}{s_x}\right)$ where r is the correlation between x and y, s_y is the sample standard deviation of the response variable, y, and s_x is the sample standard deviation of the explanatory variable, x.

DAT-1.G.3

Sometimes, the y-intercept of the line does not have a logical interpretation in context.

DAT-1.G.4

In simple linear regression, r^2 is the square of the correlation, r. It is also called the coefficient of determination. r^2 is the proportion of variation in the response variable that is explained by the explanatory variable in the model.

continued on next page

SKILLS

Data Analysis

2.C

Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response.

Statistical Argumentation

4.B

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

LEARNING OBJECTIVE

DAT-1.H

Interpret coefficients for the least-squares regression line model. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

DAT-1.H.1

The coefficients of the least-squares regression model are the estimated slope and y-intercept.

DAT-1.H.2

The slope is the amount that the predicted y-value changes for every unit increase in x.

DAT-1.H.3

The y-intercept value is the predicted value of the response variable when the explanatory variable is equal to 0. The formula for the y-intercept, a, is $a = \overline{y} - b\overline{x}$.

TOPIC 2.9

Analyzing Departures from Linearity

SKILLS

Data Analysis

2.A

Describe data presented numerically or graphically.

2.C

Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response.

Required Course Content

ENDURING UNDERSTANDING

DAT-1

Regression models may allow us to predict responses to changes in an explanatory variable.

LEARNING OBJECTIVE

DAT-1.I

Identify influential points in regression. **[Skill 2.A]**

ESSENTIAL KNOWLEDGE

DAT-1.I.1

An outlier in regression is a point that does not follow the general trend shown in the rest of the data and has a large residual when the Least Squares Regression Line (LSRL) is calculated.

DAT-1.I.2

A high-leverage point in regression has a substantially larger or smaller x-value than the other observations have.

DAT-1.I.3

An influential point in regression is any point that, if removed, changes the relationship substantially. Examples include much different slope, y-intercept, and/or correlation. Outliers and high leverage points are often influential.

continued on next page

LEARNING OBJECTIVE

DAT-1.J

Calculate a predicted response using a least-squares regression line for a transformed data set.
[Skill 2.C]

ESSENTIAL KNOWLEDGE

DAT-1.J.1

Transformations of variables, such as evaluating the natural logarithm of each value of the response variable or squaring each value of the explanatory variable, can be used to create transformed data sets, which may be more linear in form than the untransformed data.

DAT-1.J.2

Increased randomness in residual plots after transformation of data and/or movement of r^2 to a value closer to 1 offers evidence that the least-squares regression line for the transformed data is a more appropriate model to use to predict responses to the explanatory variable than the regression line for the untransformed data.

AP STATISTICS

UNIT 3

Collecting Data

12–15%
AP EXAM WEIGHTING

~9–10
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 3

Multiple-choice: ~20 questions
Free-response: 2 questions

- Exploring Data and Collecting Data
- Collecting Data

UNIT 3

12–15% AP EXAM WEIGHTING | **~9–10** CLASS PERIODS

Collecting Data

BIG IDEA 1
Variation and Distribution **VAR**
- What do our data tell us?

BIG IDEA 3
Data-Based Predictions, Decisions, and Conclusions **DAT**
- Why might the data we collected not be valid for drawing conclusions about an entire population?

Developing Understanding

Depending on how data are collected, we may or may not be able to generalize findings or establish evidence of causal relationships. For example, if random selection is not used to obtain a sample from a population, bias may result and statistics from the sample cannot be assumed to generalize to the population. For data collected using well-designed experiments, statistically significant differences between or among experimental treatment groups are evidence that the treatments caused the effect. Students learn important principles of sampling and experimental design in this unit; they will learn about statistical inference in Units 6–9.

Building Course Skills

1.B **1.C** **4.A** **4.B**

Statisticians must be adept at determining "What is this question asking?" Students should get into the habit of identifying the task in the given prompt before they begin, then checking that their response addresses that task. For example, when asked if it would be appropriate to generalize the results of a given experiment, students need to provide a clear "yes" or "no" decision in their response, along with an explanation that supports their decision.

Although students may recognize that they need to justify their reasoning, they often struggle to include explicit evidence supporting their claims. For instance, claims about non-response bias should be supported with evidence indicating whether the sample result is likely to be too high or too low compared to the population value that is being estimated. As another example, students need to clearly explain why a particular variable might lead to confounding in a given setting.

Preparing for the AP Exam

Students should continue to practice using precise language in their writing for the AP Exam. For example, some students refer to random selection when they should be writing about random assignment, or vice versa. Students should write about *random selection* when discussing generalizing results of a sample to the population. Students should write about *random assignment* when discussing experiments or their results. Because random assignment to treatments in an experiment tends to balance the effects of uncontrolled variables across groups, researchers can conclude that statistically significant differences in the response are caused by the effects of the treatments. When justifying a claim that experimental treatments cause such statistically significant differences, students should cite the random assignment of treatments and should not refer to other irrelevant elements of well-designed experiments. Students should also continue to practice connecting responses to the specific context of the question.

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods ~9–10 CLASS PERIODS
VAR-1	**3.1 Introducing Statistics: Do the Data We Collected Tell the Truth?**	1.A Identify the question to be answered or problem to be solved *(not assessed)*.	
DAT-2	**3.2 Introduction to Planning a Study**	1.C Describe an appropriate method for gathering and representing data. 4.A Make an appropriate claim or draw an appropriate conclusion.	
DAT-2	**3.3 Random Sampling and Data Collection**	1.C Describe an appropriate method for gathering and representing data.	
DAT-2	**3.4 Potential Problems with Sampling**	1.C Describe an appropriate method for gathering and representing data.	
VAR-3	**3.5 Introduction to Experimental Design**	1.C Describe an appropriate method for gathering and representing data. 1.B Identify key and relevant information to answer a question or solve a problem.	
VAR-3	**3.6 Selecting an Experimental Design**	1.C Describe an appropriate method for gathering and representing data.	
VAR-3	**3.7 Inference and Experiments**	4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 3. 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	3.2	**Graphic Organizer** Provide students with a table listing all possible combinations of whether a study involves random sampling (yes or no) and random assignment (yes or no). Ask them to fill in each cell with both the type of conclusion that is appropriate (association or causation) and the generalizability of the results (to the population or to only those similar to the study participants).
2	3.2	**Odd One Out** After modeling an odd one out example, have students form groups of four and give each of them a description of a statistical study. Explain that three of the studies are of the same type (observational or experimental) and one is different. Have students work together in their groups to determine which study is the odd one out and explain why.
3	3.2 3.3	**Password-Style Games** After completing the lessons on sampling and surveying, use the following 10 terms in a password-style game: census, simple random sample, stratified random sample, cluster sample, systematic random sample, bias, voluntary response bias, undercoverage, nonresponse bias, and response bias. The winner is the pair whose partner guesses the most terms correctly from the descriptions given.
4	3.5	**Sentence Starters** Provide students with the scenario from **2006 Form B FRQ 5**. Have them complete the following sentence starter to explain confounding: "__________ are confounded with __________ because each __________ was used in only one __________. If a difference in the draft is observed, we will not know whether the difference is due to the __________ or the __________."
5	3.5	**Think-Pair-Share** Provide students with a description of a well-designed experiment (e.g., **2010 FRQ 1**) and ask them to individually identify the type of design, the experimental units, the treatments, and how the study addresses the principles of a well-designed experiment (including random assignment, control, blinding, and replication). Then have students share their thoughts with their neighbor.

SKILL

Selecting Statistical Methods

1.A

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

TOPIC 3.1

Introducing Statistics: Do the Data We Collected Tell the Truth?

Required Course Content

ENDURING UNDERSTANDING

VAR-1

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

LEARNING OBJECTIVE

VAR-1.E

Identify questions to be answered about data collection methods. **[Skill 1.A]**

ESSENTIAL KNOWLEDGE

VAR-1.E.1

Methods for data collection that do not rely on chance result in untrustworthy conclusions.

TOPIC 3.2

Introduction to Planning a Study

SKILLS

Selecting Statistical Methods

1.C

Describe an appropriate method for gathering and representing data.

Statistical Argumentation

4.A

Make an appropriate claim or draw an appropriate conclusion.

Required Course Content

ENDURING UNDERSTANDING

DAT-2

The way we collect data influences what we can and cannot say about a population.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
DAT-2.A Identify the type of a study. **[Skill 1.C]**	**DAT-2.A.1** A population consists of all items or subjects of interest. **DAT-2.A.2** A sample selected for study is a subset of the population. **DAT-2.A.3** In an observational study, treatments are not imposed. Investigators examine data for a sample of individuals (retrospective) or follow a sample of individuals into the future collecting data (prospective) in order to investigate a topic of interest about the population. A sample survey is a type of observational study that collects data from a sample in an attempt to learn about the population from which the sample was taken. **DAT-2.A.4** In an experiment, different conditions (treatments) are assigned to experimental units (participants or subjects).
DAT-2.B Identify appropriate generalizations and determinations based on observational studies. **[Skill 4.A]**	**DAT-2.B.1** It is only appropriate to make generalizations about a population based on samples that are randomly selected or otherwise representative of that population.

continued on next page

LEARNING OBJECTIVE

DAT-2.B

Identify appropriate generalizations and determinations based on observational studies. **[Skill 4.A]**

ESSENTIAL KNOWLEDGE

DAT-2.B.2

A sample is only generalizable to the population from which the sample was selected.

DAT-2.B.3

It is not possible to determine causal relationships between variables using data collected in an observational study.

TOPIC 3.3

Random Sampling and Data Collection

SKILL

Selecting Statistical Methods

1.C

Describe an appropriate method for gathering and representing data.

Required Course Content

ENDURING UNDERSTANDING

DAT-2

The way we collect data influences what we can and cannot say about a population.

LEARNING OBJECTIVE

DAT-2.C

Identify a sampling method, given a description of a study. **[Skill 1.C]**

ESSENTIAL KNOWLEDGE

DAT-2.C.1

When an item from a population can be selected only once, this is called sampling without replacement. When an item from the population can be selected more than once, this is called sampling with replacement.

DAT-2.C.2

A simple random sample (SRS) is a sample in which every group of a given size has an equal chance of being chosen. This method is the basis for many types of sampling mechanisms. A few examples of mechanisms used to obtain SRSs include numbering individuals and using a random number generator to select which ones to include in the sample, ignoring repeats, using a table of random numbers, or drawing a card from a deck without replacement.

DAT-2.C.3

A stratified random sample involves the division of a population into separate groups, called strata, based on shared attributes or characteristics (homogeneous grouping). Within each stratum a simple random sample is selected, and the selected units are combined to form the sample.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
DAT-2.C Identify a sampling method, given a description of a study. **[Skill 1.C]**	**DAT-2.C.4** A cluster sample involves the division of a population into smaller groups, called clusters. Ideally, there is heterogeneity within each cluster, and clusters are similar to one another in their composition. A simple random sample of clusters is selected from the population to form the sample of clusters. Data are collected from all observations in the selected clusters. **DAT-2.C.5** A systematic random sample is a method in which sample members from a population are selected according to a random starting point and a fixed, periodic interval. **DAT-2.C.6** A census selects all items/subjects in a population.
DAT-2.D Explain why a particular sampling method is or is not appropriate for a given situation. **[Skill 1.C]**	**DAT-2.D.1** There are advantages and disadvantages for each sampling method depending upon the question that is to be answered and the population from which the sample will be drawn.

TOPIC 3.4

Potential Problems with Sampling

SKILL

Selecting Statistical Methods

1.C

Describe an appropriate method for gathering and representing data.

Required Course Content

ENDURING UNDERSTANDING

DAT-2

The way we collect data influences what we can and cannot say about a population.

LEARNING OBJECTIVE

DAT-2.E

Identify potential sources of bias in sampling methods. **[Skill 1.C]**

ESSENTIAL KNOWLEDGE

DAT-2.E.1

Bias occurs when certain responses are systematically favored over others.

DAT-2.E.2

When a sample is comprised entirely of volunteers or people who choose to participate, the sample will typically not be representative of the population (voluntary response bias).

DAT-2.E.3

When part of the population has a reduced chance of being included in the sample, the sample will typically not be representative of the population (undercoverage bias).

DAT-2.E.4

Individuals chosen for the sample for whom data cannot be obtained (or who refuse to respond) may differ from those for whom data can be obtained (nonresponse bias).

DAT-2.E.5

Problems in the data gathering instrument or process result in response bias. Examples include questions that are confusing or leading (question wording bias) and self-reported responses.

DAT-2.E.6

Non-random sampling methods (for example, samples chosen by convenience or voluntary response) introduce potential for bias because they do not use chance to select the individuals.

SKILLS

Selecting Statistical Methods

1.C
Describe an appropriate method for gathering and representing data.

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

TOPIC 3.5
Introduction to Experimental Design

Required Course Content

ENDURING UNDERSTANDING

VAR-3
Well-designed experiments can establish evidence of causal relationships.

LEARNING OBJECTIVE

VAR-3.A
Identify the components of an experiment. **[Skill 1.C]**

ESSENTIAL KNOWLEDGE

VAR-3.A.1
The experimental units are the individuals (which may be people or other objects of study) that are assigned treatments. When experimental units consist of people, they are sometimes referred to as participants or subjects.

VAR-3.A.2
An explanatory variable (or factor) in an experiment is a variable whose levels are manipulated intentionally. The levels or combination of levels of the explanatory variable(s) are called treatments.

VAR-3.A.3
A response variable in an experiment is an outcome from the experimental units that is measured after the treatments have been administered.

VAR-3.A.4
A confounding variable in an experiment is a variable that is related to the explanatory variable and influences the response variable and may create a false perception of association between the two.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-3.B Describe elements of a well-designed experiment. **[Skill 1.B]**	**VAR-3.B.1** A well-designed experiment should include the following: a. Comparisons of at least two treatment groups, one of which could be a control group. b. Random assignment/allocation of treatments to experimental units. c. Replication (more than one experimental unit in each treatment group). d. Control of potential confounding variables where appropriate.
VAR-3.C Compare experimental designs and methods. **[Skill 1.C]**	**VAR-3.C.1** In a completely randomized design, treatments are assigned to experimental units completely at random. Random assignment tends to balance the effects of uncontrolled (confounding) variables so that differences in responses can be attributed to the treatments. **VAR-3.C.2** Methods for randomly assigning treatments to experimental units in a completely randomized design include using a random number generator, a table of random values, drawing chips without replacement, etc. **VAR-3.C.3** In a single-blind experiment, subjects do not know which treatment they are receiving, but members of the research team do, or vice versa. **VAR-3.C.4** In a double-blind experiment neither the subjects nor the members of the research team who interact with them know which treatment a subject is receiving. **VAR-3.C.5** A control group is a collection of experimental units either not given a treatment of interest or given a treatment with an inactive substance (placebo) in order to determine if the treatment of interest has an effect. **VAR-3.C.6** The placebo effect occurs when experimental units have a response to a placebo.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-3.C Compare experimental designs and methods. **[Skill 1.C]**	**VAR-3.C.7** For randomized complete block designs, treatments are assigned completely at random within each block. **VAR-3.C.8** Blocking ensures that at the beginning of the experiment the units within each block are similar to each other with respect to at least one blocking variable. A randomized block design helps to separate natural variability from differences due to the blocking variable. **VAR-3.C.9** A matched pairs design is a special case of a randomized block design. Using a blocking variable, subjects (whether they are people or not) are arranged in pairs matched on relevant factors. Matched pairs may be formed naturally or by the experimenter. Every pair receives both treatments by randomly assigning one treatment to one member of the pair and subsequently assigning the remaining treatment to the second member of the pair. Alternately, each subject may get both treatments.

TOPIC 3.6

Selecting an Experimental Design

SKILL

Selecting Statistical Methods

1.C

Describe an appropriate method for gathering and representing data.

Required Course Content

ENDURING UNDERSTANDING

VAR-3

Well-designed experiments can establish evidence of causal relationships.

LEARNING OBJECTIVE

VAR-3.D

Explain why a particular experimental design is appropriate. **[Skill 1.C]**

ESSENTIAL KNOWLEDGE

VAR-3.D.1

There are advantages and disadvantages for each experimental design depending on the question of interest, the resources available, and the nature of the experimental units.

SKILL

Statistical Argumentation

4.B

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

TOPIC 3.7
Inference and Experiments

Required Course Content

ENDURING UNDERSTANDING

VAR-3

Well-designed experiments can establish evidence of causal relationships.

LEARNING OBJECTIVE

VAR-3.E

Interpret the results of a well-designed experiment. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

VAR-3.E.1

Statistical inference attributes conclusions based on data to the distribution from which the data were collected.

VAR-3.E.2

Random assignment of treatments to experimental units allows researchers to conclude that some observed changes are so large as to be unlikely to have occurred by chance. Such changes are said to be statistically significant.

VAR-3.E.3

Statistically significant differences between or among experimental treatment groups are evidence that the treatments caused the effect.

VAR-3.E.4

If the experimental units used in an experiment are representative of some larger group of units, the results of an experiment can be generalized to the larger group. Random selection of experimental units gives a better chance that the units will be representative.

AP STATISTICS

UNIT 4

Probability, Random Variables, and Probability Distributions

10–20%
AP EXAM WEIGHTING

~18–20
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 4

Multiple-choice: ~45 questions
Free-response: 2 questions

- Probability
- Investigative Task

UNIT 4

10–20% AP EXAM WEIGHTING

~18–20 CLASS PERIODS

Probability, Random Variables, and Probability Distributions

BIG IDEA 1
Variation and Distribution **VAR**
- How can an event be both random and predictable?

BIG IDEA 2
Patterns and Uncertainty **UNC**
- About how many rolls of a fair six-sided die would we anticipate it taking to get three 1s?

Developing Understanding

Probabilistic reasoning allows statisticians to quantify the likelihood of random events over the long run and to make statistical inferences. Simulations and concrete examples can help students to understand the abstract definitions and calculations of probability. This unit builds on understandings of simulated or empirical data distributions and fundamental principles of probability to represent, interpret, and calculate parameters for theoretical probability distributions for discrete random variables. Interpretations of probabilities and parameters associated with a probability distribution should use appropriate units and relate to the context of the situation.

Building Course Skills

2.B 3.A 3.B 4.B

Probability is a notoriously difficult topic for students to grasp because it's difficult to conceptualize future outcomes in concrete ways. Before introducing new formulas, teachers can help students get an intuitive feel for why the formulas (and related notation) make sense. For example, the probability formulas for $P(A \text{ or } B)$ and for $P(A|B)$ can be presented intuitively with two-way tables. Simulations can also help students internalize what it means to quantify random behavior. To help students understand when to apply different probability rules, teachers can use explicit strategies such as matching verbal scenarios to their corresponding probability formulas.

Students frequently misinterpret probability distributions and parameters for random variables. Teachers can reinforce that a complete interpretation will include context and units. A common misconception later in the course is that every question involving probability requires a significance test. Students should practice making predictions and decisions based on probability alone to avoid this misconception early on. They should revisit these problems in later units to practice differentiating between inference and probability problems.

Preparing for the AP Exam

To help students prepare for the AP Exam, teachers can model showing all steps in probability calculations and expect students to do the same. Calculations on the AP Exam should include presentation of an appropriate expression that communicates the structure of the formula, substitution of relevant values extracted from the problem, and an answer. In **2017 FRQ 3**, for example, a student who writes "$P(G) = P(G|J) \cdot P(J) + P(G|K) \cdot P(K) = (0.2119)(0.7) + (0.8413)(0.3) = 0.4007$" has communicated the products in the multiplication rule, the sum in the addition rule, and an understanding that the events are mutually exclusive—all components of a complete response. To avoid a common error, students who present the same work using a tree diagram should practice using probabilities in the diagram correctly. Students importing incorrect solutions from one part of a multipart question to solve another will not be penalized a second time, unless the subsequent result is not a reasonable value (like a probability less than 0 or greater than 1).

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods ~18–20 CLASS PERIODS
VAR-1	**4.1 Introducing Statistics: Random and Non-Random Patterns?**	1.A Identify the question to be answered or problem to be solved *(not assessed)*.	
UNC-2	**4.2 Estimating Probabilities Using Simulation**	3.A Determine relative frequencies, proportions, or probabilities using simulation or calculations.	
VAR-4	**4.3 Introduction to Probability**	3.A Determine relative frequencies, proportions, or probabilities using simulation or calculations. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.4 Mutually Exclusive Events	4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.5 Conditional Probability	3.A Determine relative frequencies, proportions, or probabilities using simulation or calculations.	
	4.6 Independent Events and Unions of Events	3.A Determine relative frequencies, proportions, or probabilities using simulation or calculations.	
VAR-5	**4.7 Introduction to Random Variables and Probability Distributions**	2.B Construct numerical or graphical representations of distributions. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.8 Mean and Standard Deviation of Random Variables	3.B Determine parameters for probability distributions. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.9 Combining Random Variables	3.B Determine parameters for probability distributions. 3.C Describe probability distributions.	

continued on next page

UNIT AT A GLANCE *(cont'd)*

Enduring Understanding	Topic	Skills	Class Periods ~18–20 CLASS PERIODS
UNC-3	**4.10 Introduction to the Binomial Distribution**	**3.A** Determine relative frequencies, proportions, or probabilities using simulation or calculations.	
	4.11 Parameters for a Binomial Distribution	**3.B** Determine parameters for probability distributions. **4.B** Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.12 The Geometric Distribution	**3.A** Determine relative frequencies, proportions, or probabilities using simulation or calculations. **3.B** Determine parameters for probability distributions. **4.B** Interpret statistical calculations and findings to assign meaning or assess a claim.	

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 4.
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	4.3 4.5 4.8	**Error Analysis** Using **2015 FRQ 3**, provide students with several answers containing errors for each part. Provide some responses with incorrect notation, incorrect work, missing work, work that shows calculator commands only, an incorrect formula or approach, and an incorrect final answer. Then ask students to identify the errors.
2	4.3 4.5 4.6	**Think-Pair-Share** Provide students with a set of five probability questions: one for the complement rule, the conditional probability formula, the general multiplication rule, the multiplication rule for independent events, and the general addition rule. Ask students to individually identify the formula needed to solve each problem, without doing the final calculations. Then have them share their thoughts with a partner.
3	4.5 4.6	**Create Representations** Provide students with the scenario from **2018 FRQ 3**. Ask them to create a tree diagram to organize the information in the problem. Then ask them to use the information in the problem to set up a hypothetical 100,000 table (to make the decimals easy to work with), such as the one below. Encourage students to try both representations when solving probability questions in the future.

	Multiple Birth	Single Birth	Total
Left handed	770	10,615	11,385
Right handed	2,730	85,885	88,615
Total	3,500	96,500	100,000

Activity	Topic	Sample Activity
4	4.10 4.12	**Odd One Out** After modeling an odd one out example, have students form groups of four and give each of them a description of either a binomial or a geometric random variable. Explain that three of their variables follow the same probability distribution and one is different. Have students work together in their groups to determine whose is the odd one out and explain why.
5	4.2 4.12	**Predict and Confirm** Ask students to consider couples who plan to continue having children until they have one girl and predict how many children they think these couples will have, on average. Then ask each student to perform 10 trials using a coin toss where Heads = Girl and Tails = Boy. A trial is finished once one girl is observed and the number of total children is recorded. Combine the class results and calculate the average. Confirm with the geometric mean formula once it is discussed.

TOPIC 4.1

Introducing Statistics: Random and Non-Random Patterns?

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.F

Identify questions suggested by patterns in data. **[Skill 1.A]**

ESSENTIAL KNOWLEDGE

VAR-1.F.1

Patterns in data do not necessarily mean that variation is not random.

SKILL

Using Probability and Simulation

3.A

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

AVAILABLE RESOURCES

- Classroom Resources >
  - **Graphing Calculator Simulations Simplified**
  - **Three Calculator Simulation Activities**

ILLUSTRATIVE EXAMPLES

An outcome:

- Rolling a particular value on a six-sided number cube is one of six possible outcomes.

An event:

- When rolling two six-sided number cubes, an event would be a sum of seven. The corresponding collection of outcomes would be (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1), where the ordered pairs indicate (face value on one cube, face value on the other cube).

TOPIC 4.2
Estimating Probabilities Using Simulation

Required Course Content

ENDURING UNDERSTANDING

UNC-2

Simulation allows us to anticipate patterns in data.

LEARNING OBJECTIVE

UNC-2.A

Estimate probabilities using simulation. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

UNC-2.A.1

A random process generates results that are determined by chance.

UNC-2.A.2

An outcome is the result of a trial of a random process.

UNC-2.A.3

An event is a collection of outcomes.

UNC-2.A.4

Simulation is a way to model random events, such that simulated outcomes closely match real-world outcomes. All possible outcomes are associated with a value to be determined by chance. Record the counts of simulated outcomes and the count total.

UNC-2.A.5

The relative frequency of an outcome or event in simulated or empirical data can be used to estimate the probability of that outcome or event.

UNC-2.A.6

The law of large numbers states that simulated (empirical) probabilities tend to get closer to the true probability as the number of trials increases.

TOPIC 4.3

Introduction to Probability

SKILLS

Using Probability and Simulation

3.A

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

Statistical Argumentation

4.B

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

Required Course Content

ENDURING UNDERSTANDING

VAR-4

The likelihood of a random event can be quantified.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-4.A Calculate probabilities for events and their complements. **[Skill 3.A]**	**VAR-4.A.1** The sample space of a random process is the set of all possible non-overlapping outcomes. **VAR-4.A.2** If all outcomes in the sample space are equally likely, then the probability an event E will occur is defined as the fraction: $\frac{\text{number of outcomes in event E}}{\text{total number of outcomes in sample space}}$ **VAR-4.A.3** The probability of an event is a number between 0 and 1, inclusive. **VAR-4.A.4** The probability of the complement of an event E, E' or E^C, (i.e., not E) is equal to $1 - P(E)$.
VAR-4.B Interpret probabilities for events. **[Skill 4.B]**	**VAR-4.B.1** Probabilities of events in repeatable situations can be interpreted as the relative frequency with which the event will occur in the long run.

SKILL

Statistical Argumentation

4.B

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

TOPIC 4.4

Mutually Exclusive Events

Required Course Content

ENDURING UNDERSTANDING

VAR-4

The likelihood of a random event can be quantified.

LEARNING OBJECTIVE

VAR-4.C

Explain why two events are (or are not) mutually exclusive. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

VAR-4.C.1

The probability that events A and B both will occur, sometimes called the joint probability, is the probability of the intersection of A and B, denoted $P(A \cap B)$.

VAR-4.C.2

Two events are mutually exclusive or disjoint if they cannot occur at the same time. So $P(A \cap B) = 0$.

TOPIC 4.5

Conditional Probability

SKILL

Using Probability and Simulation

3.A

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

Required Course Content

ENDURING UNDERSTANDING

VAR-4

The likelihood of a random event can be quantified.

LEARNING OBJECTIVE

VAR-4.D

Calculate conditional probabilities. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

VAR-4.D.1

The probability that event A will occur given that event B has occurred is called a conditional probability and denoted $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$.

VAR-4.D.2

The multiplication rule states that the probability that events A and B both will occur is equal to the probability that event A will occur multiplied by the probability that event B will occur, given that A has occurred. This is denoted $P(A \cap B) = P(A) \cdot P(B \mid A)$.

SKILL

Using Probability and Simulation

3.A

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

TOPIC 4.6

Independent Events and Unions of Events

Required Course Content

ENDURING UNDERSTANDING

VAR-4

The likelihood of a random event can be quantified.

LEARNING OBJECTIVE

VAR-4.E

Calculate probabilities for independent events and for the union of two events. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

VAR-4.E.1

Events A and B are independent if, and only if, knowing whether event A has occurred (or will occur) does not change the probability that event B will occur.

VAR-4.E.2

If, and only if, events A and B are independent, then $P(A \mid B) = P(A)$, $P(B \mid A) = P(B)$, and $P(A \cap B) = P(A) \cdot P(B)$.

VAR-4.E.3

The probability that event A or event B (or both) will occur is the probability of the union of A and B, denoted $P(A \cup B)$.

VAR-4.E.4

The addition rule states that the probability that event A or event B or both will occur is equal to the probability that event A will occur plus the probability that event B will occur minus the probability that both events A and B will occur. This is denoted $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

TOPIC 4.7

Introduction to Random Variables and Probability Distributions

Required Course Content

ENDURING UNDERSTANDING

VAR-5

Probability distributions may be used to model variation in populations.

LEARNING OBJECTIVE

VAR-5.A

Represent the probability distribution for a discrete random variable. **[Skill 2.B]**

ESSENTIAL KNOWLEDGE

VAR-5.A.1

The values of a random variable are the numerical outcomes of random behavior.

VAR-5.A.2

A discrete random variable is a variable that can only take a countable number of values. Each value has a probability associated with it. The sum of the probabilities over all of the possible values must be 1.

VAR-5.A.3

A probability distribution can be represented as a graph, table, or function showing the probabilities associated with values of a random variable.

VAR-5.A.4

A cumulative probability distribution can be represented as a table or function showing the probability of being less than or equal to each value of the random variable.

LEARNING OBJECTIVE

VAR-5.B

Interpret a probability distribution. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

VAR-5.B.1

An interpretation of a probability distribution provides information about the shape, center, and spread of a population and allows one to make conclusions about the population of interest.

SKILLS

Data Analysis

2.B

Construct numerical or graphical representations of distributions.

Statistical Argumentation

4.B

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

ILLUSTRATIVE EXAMPLES

Outcomes of trials of a random process:

- The sum of the outcomes for rolling two dice
- The number of puppies in a randomly selected litter for a certain breed of dog

SKILLS

Using Probability and Simulation

3.B
Determine parameters for probability distributions.

Statistical Argumentation

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

TOPIC 4.8

Mean and Standard Deviation of Random Variables

Required Course Content

ENDURING UNDERSTANDING

VAR-5
Probability distributions may be used to model variation in populations.

LEARNING OBJECTIVE

VAR-5.C
Calculate parameters for a discrete random variable.
[Skill 3.B]

ESSENTIAL KNOWLEDGE

VAR-5.C.1
A numerical value measuring a characteristic of a population or the distribution of a random variable is known as a parameter, which is a single, fixed value.

VAR-5.C.2
The mean, or expected value, for a discrete random variable X is $\mu_X = \sum x_i \cdot P(x_i)$.

VAR-5.C.3
The standard deviation for a discrete random variable X is $\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 \cdot P(x_i)}$.

LEARNING OBJECTIVE

VAR-5.D
Interpret parameters for a discrete random variable.
[Skill 4.B]

ESSENTIAL KNOWLEDGE

VAR-5.D.1
Parameters for a discrete random variable should be interpreted using appropriate units and within the context of a specific population.

TOPIC 4.9

Combining Random Variables

SKILLS

Using Probability and Simulation

3.B

Determine parameters for probability distributions.

3.C

Describe probability distributions.

Required Course Content

ENDURING UNDERSTANDING

VAR-5

Probability distributions may be used to model variation in populations.

LEARNING OBJECTIVE

VAR-5.E

Calculate parameters for linear combinations of random variables. **[Skill 3.B]**

ESSENTIAL KNOWLEDGE

VAR-5.E.1

For random variables X and Y and real numbers a and b, the mean of $aX + bY$ is $a\mu_x + b\mu_y$.

VAR-5.E.2

Two random variables are independent if knowing information about one of them does not change the probability distribution of the other.

VAR-5.E.3

For independent random variables X and Y and real numbers a and b, the mean of $aX + bY$ is $a\mu_x + b\mu_y$, and the variance of $aX + bY$ is $a^2\sigma^2_x + b^2\sigma^2_y$.

LEARNING OBJECTIVE

VAR-5.F

Describe the effects of linear transformations of parameters of random variables. **[Skill 3.C]**

ESSENTIAL KNOWLEDGE

VAR-5.F.1

For $Y = a + bX$, the probability distribution of the transformed random variable, Y, has the same shape as the probability distribution for X, so long as $a > 0$ and $b > 0$. The mean of Y is $\mu_y = a + b\mu_x$. The standard deviation of Y is $\sigma_y = |b|\sigma_x$.

SKILL

Using Probability and Simulation

3.A

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

TOPIC 4.10
Introduction to the Binomial Distribution

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE

UNC-3.A

Estimate probabilities of binomial random variables using data from a simulation. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

UNC-3.A.1

A probability distribution can be constructed using the rules of probability or estimated with a simulation using random number generators.

UNC-3.A.2

A binomial random variable, X, counts the number of successes in n repeated independent trials, each trial having two possible outcomes (success or failure), with the probability of success p and the probability of failure $1 - p$.

LEARNING OBJECTIVE

UNC-3.B

Calculate probabilities for a binomial distribution. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

UNC-3.B.1

The probability that a binomial random variable, X, has exactly x successes for n independent trials, when the probability of success is p, is calculated as $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \ldots, n.$ This is the binomial probability function.

TOPIC 4.11

Parameters for a Binomial Distribution

Required Course Content

ENDURING UNDERSTANDING

UNC-3
Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE

UNC-3.C
Calculate parameters for a binomial distribution. **[Skill 3.B]**

ESSENTIAL KNOWLEDGE

UNC-3.C.1
If a random variable is binomial, its mean, μ_x, is np and its standard deviation, σ_x, is $\sqrt{np(1-p)}$.

LEARNING OBJECTIVE

UNC-3.D
Interpret probabilities and parameters for a binomial distribution. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

UNC-3.D.1
Probabilities and parameters for a binomial distribution should be interpreted using appropriate units and within the context of a specific population or situation.

SKILLS

Using Probability and Simulation

3.B
Determine parameters for probability distributions.

Statistical Argumentation

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

SKILLS

Using Probability and Simulation

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

3.B
Determine parameters for probability distributions.

Statistical Argumentation

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

TOPIC 4.12
The Geometric Distribution

Required Course Content

ENDURING UNDERSTANDING

UNC-3
Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.E Calculate probabilities for geometric random variables. **[Skill 3.A]**	**UNC-3.E.1** For a sequence of independent trials, a geometric random variable, X, gives the number of the trial on which the first success occurs. Each trial has two possible outcomes (success or failure) with the probability of success p and the probability of failure $1-p$. **UNC-3.E.2** The probability that the first success for repeated independent trials with probability of success p occurs on trial x is calculated as $P(X=x)=(1-p)^{x-1}p, x=1,2,3,...$. This is the geometric probability function.
UNC-3.F Calculate parameters of a geometric distribution. **[Skill 3.B]**	**UNC-3.F.1** If a random variable is geometric, its mean, μ_x, is $\frac{1}{p}$ and its standard deviation, σ_x, is $\frac{\sqrt{(1-p)}}{p}$.
UNC-3.G Interpret probabilities and parameters for a geometric distribution. **[Skill 4.B]**	**UNC-3.G.1** Probabilities and parameters for a geometric distribution should be interpreted using appropriate units and within the context of a specific population or situation.

QUICK REFERENCE FOR NOTATION AND FORMULAS FOR PROBABILITY DISTRIBUTIONS

Distribution	Notes	Parameter(s)	Random Variable	Conditions	Mean for Distribution	Standard Deviation for Distribution
Probability distribution for a random variable	▪ Represent discrete random variables using frequency/ relative frequency tables or histograms ▪ Represent continuous random variables with density functions.	μ_x, σ_x	X	▪ All probabilities must be between 0 and 1. ▪ $\sum \text{Probabilities} = 1$.	$\mu_X = \sum x_i \cdot P(x_i)$ "expected value"	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 \cdot P(x_i)}$
Sum or difference of independent random variables	See Unit 5 for distributions of other linear transformations of random variables.	$\mu_X, \sigma_X, \mu_Y, \sigma_Y$	$X+Y$ or $X-Y$	To calculate the variance or standard deviation of the difference, the random variables must be independent.	$\mu_{X+Y} = \mu_X + \mu_Y$ or $\mu_{X-Y} = \mu_X - \mu_Y$	Variance, $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$ Variance, $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$
Binomial probability distribution	Binomial probability function: $P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, 2, 3, \ldots, n$.	n and p	X	▪ n is predetermined. ▪ Binary ▪ Independent ▪ p is the same for each trial.	$\mu_x = np$	$\sigma_X = \sqrt{np(1-p)}$
Geometric probability distribution	Geometric probability formula: $P(X=x) = (1-p)^{x-1} p$, $x = 1, 2, 3, \ldots$.	p	X	▪ n is not predetermined. ▪ Binary ▪ Independent ▪ p is the same for each repetition (random).	$\mu = \frac{1}{p}$ expected number of trials to get the first success	$\sigma = \frac{\sqrt{1-p}}{p}$

Note: Other notation could also be correct if properly defined. Incorrect notation will result in lost points on the AP exam.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

AP STATISTICS

UNIT 5

Sampling Distributions

7–12%
AP EXAM WEIGHTING

~10–12
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 5

Multiple-choice: ~35 questions
Free-response: 2 questions

- Probability and Sampling Distributions
- Investigative Task

UNIT 5

7–12% AP EXAM WEIGHTING

~10–12 CLASS PERIODS

Sampling Distributions

BIG IDEA 1
Variation and Distribution **VAR**

- How likely is it to get a value this large just by chance?

BIG IDEA 2
Patterns and Uncertainty **UNC**

- How can we anticipate patterns in the values of a statistic from one sample to another?

Developing Understanding

This unit applies probabilistic reasoning to sampling, introducing students to sampling distributions of statistics they will use when performing inference in Units 6 and 7. Students should understand that sample statistics can be used to estimate corresponding population parameters and that measures of center (mean) and variability (standard deviation) for these sampling distributions can be determined directly from the population parameters when certain sampling criteria are met. For large enough samples from any population, these sampling distributions can be approximated by a normal distribution. Simulating sampling distributions helps students to understand how the values of statistics vary in repeated random sampling from populations with known parameters.

Building Course Skills

3.B 3.C 4.B

The probabilities associated with the normal distribution are what statisticians use to justify claims about populations they'll never be able to measure directly. Revisiting these properties early in Unit 5 will reinforce why sampling distributions allow statisticians to approximate parameters for the population of interest. Sketching, shading, and labeling a normal distribution aids in understanding the probability being calculated. Students should practice creating graphical representations, labeling the mean, and marking off values 1, 2, and 3 standard deviations from the mean.

Students often struggle to interpret parameters of probability distributions in context, simply describing features of the graph rather than explicitly connecting those features to the situation described in the problem. Teachers can remind students that context is about a variable ("tip amounts," for example), not just the units (dollars). It's also critical that students explicitly show that the appropriate conditions have been verified, and that they avoid using nonspecific language like "it" in their interpretations. Using an error analysis strategy with sample responses can help familiarize students with these issues before they make similar mistakes.

Preparing for the AP Exam

Responses on the AP Exam often uncover gaps in understanding of sampling distributions. Students must clearly communicate whether they are talking about the distribution of a population, a sample of values (heights, for example), or a sample statistic from repeated samples (mean heights, for example). Broad generalizations, such as "larger samples have less variability," leave the exam reader unsure of whether the student is referring to variability within a sample (for which the statement would be false) or a sampling distribution. The word "it" often introduces ambiguity to a response. Students frequently show confusion about what condition to check when asserting that the sampling distribution of a given statistic is approximately normal. Students should support normal probability calculations with a sketch or a calculation of a standardized score (z-score), rather than relying on calculator syntax.

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods ~10–12 CLASS PERIODS
VAR-1	**5.1 Introducing Statistics: Why Is My Sample Not Like Yours?**	1.A Identify the question to be answered or problem to be solved *(not assessed)*.	
VAR-6	**5.2 The Normal Distribution, Revisited**	3.A Determine relative frequencies, proportions, or probabilities using simulation or calculations. 3.C Describe probability distributions.	
UNC-3	**5.3 The Central Limit Theorem**	3.C Describe probability distributions.	
	5.4 Biased and Unbiased Point Estimates	4.B Interpret statistical calculations and findings to assign meaning or assess a claim. 3.B Determine parameters for probability distributions.	
	5.5 Sampling Distributions for Sample Proportions	3.B Determine parameters for probability distributions. 3.C Describe probability distributions. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
	5.6 Sampling Distributions for Differences in Sample Proportions	3.B Determine parameters for probability distributions. 3.C Describe probability distributions. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
	5.7 Sampling Distributions for Sample Means	3.B Determine parameters for probability distributions. 3.C Describe probability distributions. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	
	5.8 Sampling Distributions for Differences in Sample Means	3.B Determine parameters for probability distributions. 3.C Describe probability distributions. 4.B Interpret statistical calculations and findings to assign meaning or assess a claim.	

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 5.
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	5.2	**Think Aloud** Group students into pairs within a larger group of four. Have each student individually read **2014 FRQ 3** and think aloud with their partner, brainstorming ways to begin each part of the question. Each student then independently completes all parts. Have the pairs compare answers within their groups, improving their individual responses as necessary. Groups can then compare their responses with other groups. Finally, have students score their responses according to the rubric.
2	5.3	**Use Manipulatives** From a large container of pennies, have each student take two random samples of size 5, two of size 10, and two of size 25, and record the dates on those pennies. Have students calculate the mean of the dates in each sample and then construct four "dotplots" on the floor: one using the pennies, one using nickels placed at the mean of the student's sample size 5, one using dimes placed at the mean of the sample size 10, and one using quarters placed at the mean of the sample size 25.
3	5.5 5.7	**Password-Style Games** Have partners sit facing opposite sides of the room. Display vocabulary terms from the unit on the classroom screen. Have the students facing the screen describe the terms to their partner who then tries to guess the terms described. After half of the terms have been used, have students switch roles. Terms to include: parameter, statistic, sampling distribution, distribution of sample data, sample distribution, unbiased estimator, sampling variability of a statistic, bias, sample proportion, sample mean, $\mu_{\hat{p}}$, $\sigma_{\hat{p}}$, $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$, and central limit theorem.

SKILL

Selecting Statistical Methods

1.A

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

TOPIC 5.1

Introducing Statistics: Why Is My Sample Not Like Yours?

Required Course Content

ENDURING UNDERSTANDING

VAR-1

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

LEARNING OBJECTIVE

VAR-1.G

Identify questions suggested by variation in statistics for samples collected from the same population. **[Skill 1.A]**

ESSENTIAL KNOWLEDGE

VAR-1.G.1

Variation in statistics for samples taken from the same population may be random or not.

TOPIC 5.2

The Normal Distribution, Revisited

SKILLS

Using Probability and Simulation

3.A

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

3.C

Describe probability distributions.

ILLUSTRATIVE EXAMPLE

Continuous random variable:

If one looks at a clock at a random time, the probability that the minute hand is between the 3 and the 6 is one fourth.

Required Course Content

ENDURING UNDERSTANDING

VAR-6

The normal distribution may be used to model variation.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-6.A Calculate the probability that a particular value lies in a given interval of a normal distribution. **[Skill 3.A]**	**VAR-6.A.1** A continuous random variable is a variable that can take on any value within a specified domain. Every interval within the domain has a probability associated with it. **VAR-6.A.2** A continuous random variable with a normal distribution is commonly used to describe populations. The distribution of a normal random variable can be described by a normal, or "bell-shaped," curve. **VAR-6.A.3** The area under a normal curve over a given interval represents the probability that a particular value lies in that interval.
VAR-6.B Determine the interval associated with a given area in a normal distribution. **[Skill 3.A]**	**VAR-6.B.1** The boundaries of an interval associated with a given area in a normal distribution can be determined using z-scores or technology, such as a calculator, a standard normal table, or computer-generated output.

continued on next page

LEARNING OBJECTIVE

VAR-6.B

Determine the interval associated with a given area in a normal distribution.
[Skill 3.A]

ESSENTIAL KNOWLEDGE

VAR-6.B.2

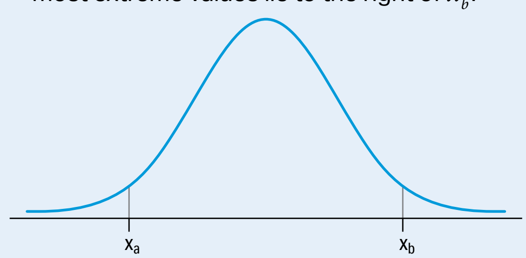
Intervals associated with a given area in a normal distribution can be determined by assigning appropriate inequalities to the boundaries of the intervals:

a. $P(X < x_a) = \frac{p}{100}$ means that the lowest $p\%$ of values lie to the left of x_a.

b. $P(x_a < X < x_b) = \frac{p}{100}$ means that $p\%$ of values lie between x_a and x_b.

c. $P(X > x_b) = \frac{p}{100}$ means that the highest $p\%$ of values lie to the right of x_b.

d. To determine the most extreme $p\%$ of values requires dividing the area associated with $p\%$ into two equal areas on either extreme of the distribution: $P(X < x_a) = \frac{1}{2}\frac{p}{100}$ and $P(X > x_b) = \frac{1}{2}\frac{p}{100}$ means that half of the $p\%$ most extreme values lie to the left of x_a and half of the $p\%$ most extreme values lie to the right of x_b.



LEARNING OBJECTIVE

VAR-6.C

Determine the appropriateness of using the normal distribution to approximate probabilities for unknown distributions.
[Skill 3.C]

ESSENTIAL KNOWLEDGE

VAR-6.C.1

Normal distributions are symmetrical and "bell-shaped." As a result, normal distributions can be used to approximate distributions with similar characteristics.

TOPIC 5.3

The Central Limit Theorem

SKILL

Using Probability and Simulation

3.C

Describe probability distributions.

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE

UNC-3.H

Estimate sampling distributions using simulation. **[Skill 3.C]**

ESSENTIAL KNOWLEDGE

UNC-3.H.1

A sampling distribution of a statistic is the distribution of values for the statistic for all possible samples of a given size from a given population.

UNC-3.H.2

The central limit theorem (CLT) states that when the sample size is sufficiently large, a sampling distribution of the mean of a random variable will be approximately normally distributed.

UNC-3.H.3

The central limit theorem requires that the sample values are independent of each other and that n is sufficiently large.

UNC-3.H.4

A randomization distribution is a collection of statistics generated by simulation assuming known values for the parameters. For a randomized experiment, this means repeatedly randomly reallocating/reassigning the response values to treatment groups.

UNC-3.H.5

The sampling distribution of a statistic can be simulated by generating repeated random samples from a population.

SKILLS

Statistical Argumentation

4.B

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

Using Probability and Simulation

3.B

Determine parameters for probability distributions.

TOPIC 5.4

Biased and Unbiased Point Estimates

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.I Explain why an estimator is or is not unbiased. **[Skill 4.B]**	UNC-3.I.1 When estimating a population parameter, an estimator is unbiased if, on average, the value of the estimator is equal to the population parameter.
UNC-3.J Calculate estimates for a population parameter. **[Skill 3.B]**	UNC-3.J.1 When estimating a population parameter, an estimator exhibits variability that can be modeled using probability. UNC-3.J.2 A sample statistic is a point estimator of the corresponding population parameter.

TOPIC 5.5

Sampling Distributions for Sample Proportions

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.K Determine parameters of a sampling distribution for sample proportions. **[Skill 3.B]**	UNC-3.K.1 For independent samples (sampling with replacement) of a categorical variable from a population with population proportion, p, the sampling distribution of the sample proportion, $\hat{p}$, has a mean, $\mu_{\hat{p}} = p$ and a standard deviation, $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. UNC-3.K.2 If sampling without replacement, the standard deviation of the sample proportion is smaller than what is given by the formula above. If the sample size is less than 10% of the population size, the difference is negligible.
UNC-3.L Determine whether a sampling distribution for a sample proportion can be described as approximately normal. **[Skill 3.C]**	UNC-3.L.1 For a categorical variable, the sampling distribution of the sample proportion, $\hat{p}$, will have an approximate normal distribution, provided the sample size is large enough: $np \geq 10$ and $n(1-p) \geq 10$
UNC-3.M Interpret probabilities and parameters for a sampling distribution for a sample proportion. **[Skill 4.B]**	UNC-3.M.1 Probabilities and parameters for a sampling distribution for a sample proportion should be interpreted using appropriate units and within the context of a specific population.

SKILLS

Using Probability and Simulation

3.B

Determine parameters for probability distributions.

3.C

Describe probability distributions.

Statistical Argumentation

4.B

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

AVAILABLE RESOURCES

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

SKILLS

Using Probability and Simulation

3.B
Determine parameters for probability distributions.

3.C
Describe probability distributions.

Statistical Argumentation

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

AVAILABLE RESOURCES

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

TOPIC 5.6
Sampling Distributions for Differences in Sample Proportions

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE

UNC-3.N

Determine parameters of a sampling distribution for a difference in sample proportions. **[Skill 3.B]**

ESSENTIAL KNOWLEDGE

UNC-3.N.1

For a categorical variable, when randomly sampling with replacement from two independent populations with population proportions p_1 and p_2, the sampling distribution of the difference in sample proportions $\hat{p}_1 - \hat{p}_2$ has mean, $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ and standard deviation, $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$.

UNC-3.N.2

If sampling without replacement, the standard deviation of the difference in sample proportions is smaller than what is given by the formula above. If the sample sizes are less than 10% of the population sizes, the difference is negligible.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.O Determine whether a sampling distribution for a difference of sample proportions can be described as approximately normal. **[Skill 3.C]**	**UNC-3.O.1** The sampling distribution of the difference in sample proportions $\hat{p}_1 - \hat{p}_2$ will have an approximate normal distribution provided the sample sizes are large enough: $n_1 p_1 \geq 10, n_1(1-p_1) \geq 10, n_2 p_2 \geq 10, n_2(1-p_2) \geq 10.$
UNC-3.P Interpret probabilities and parameters for a sampling distribution for a difference in proportions. **[Skill 4.B]**	**UNC-3.P.1** Parameters for a sampling distribution for a difference of proportions should be interpreted using appropriate units and within the context of a specific populations.

SKILLS

Using Probability and Simulation

3.B
Determine parameters for probability distributions.

3.C
Describe probability distributions.

Statistical Argumentation

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

AVAILABLE RESOURCES

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

TOPIC 5.7
Sampling Distributions for Sample Means

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.Q Determine parameters for a sampling distribution for sample means. **[Skill 3.B]**	**UNC-3.Q.1** For a numerical variable, when random sampling with replacement from a population with mean μ and standard deviation, σ, the sampling distribution of the sample mean has mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. **UNC-3.Q.2** If sampling without replacement, the standard deviation of the sample mean is smaller than what is given by the formula above. If the sample size is less than 10% of the population size, the difference is negligible.
UNC-3.R Determine whether a sampling distribution of a sample mean can be described as approximately normal. **[Skill 3.C]**	**UNC-3.R.1** For a numerical variable, if the population distribution can be modeled with a normal distribution, the sampling distribution of the sample mean, $\bar{x}$, can be modeled with a normal distribution.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.R Determine whether a sampling distribution of a sample mean can be described as approximately normal. **[Skill 3.C]**	**UNC-3.R.2** For a numerical variable, if the population distribution cannot be modeled with a normal distribution, the sampling distribution of the sample mean, $\bar{x}$, can be modeled approximately by a normal distribution, provided the sample size is large enough, e.g., greater than or equal to 30.
UNC-3.S Interpret probabilities and parameters for a sampling distribution for a sample mean. **[Skill 4.B]**	**UNC-3.S.1** Probabilities and parameters for a sampling distribution for a sample mean should be interpreted using appropriate units and within the context of a specific population.

SKILLS

Using Probability and Simulation

3.B

Determine parameters for probability distributions.

3.C

Describe probability distributions.

Statistical Argumentation

4.B

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

AVAILABLE RESOURCES

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

TOPIC 5.8

Sampling Distributions for Differences in Sample Means

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE

UNC-3.T

Determine parameters of a sampling distribution for a difference in sample means. **[Skill 3.B]**

ESSENTIAL KNOWLEDGE

UNC-3.T.1

For a numerical variable, when randomly sampling with replacement from two independent populations with population means μ_1 and μ_2 and population standard deviations σ_1 and σ_2, the sampling distribution of the difference in sample means $\bar{x}_1 - \bar{x}_2$ has mean $\mu_{(\bar{x}_1-\bar{x}_2)} = \mu_1 - \mu_2$ and standard deviation,

$$\sigma_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

UNC-3.T.2

If sampling without replacement, the standard deviation of the difference in sample means is smaller than what is given by the formula above. If the sample sizes are less than 10% of the population sizes, the difference is negligible.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.U Determine whether a sampling distribution of a difference in sample means can be described as approximately normal. **[Skill 3.C]**	**UNC-3.U.1** The sampling distribution of the difference in sample means $\bar{x}_1 - \bar{x}_2$ can be modeled with a normal distribution if the two population distributions can be modeled with a normal distribution. **UNC-3.U.2** The sampling distribution of the difference in sample means $\bar{x}_1 - \bar{x}_2$ can be modeled approximately by a normal distribution if the two population distributions cannot be modeled with a normal distribution but both sample sizes are greater than or equal to 30.
UNC-3.V Interpret probabilities and parameters for a sampling distribution for a difference in sample means. **[Skill 4.B]**	**UNC-3.V.1** Probabilities and parameters for a sampling distribution for a difference of sample means should be interpreted using appropriate units and within the context of a specific populations.

QUICK REFERENCE FOR NOTATION AND FORMULAS FOR SAMPLING DISTRIBUTIONS

Distribution	Notes	Parameter(s)	Statistic	Conditions	Mean for Distribution	Standard Deviation for Distribution
Normal distribution	A continuous random probability distribution	μ and σ			μ	σ
Sampling distribution for a sample proportion	Compare to the mean and standard deviation of a binomial random variable, X	p	$\hat{p}$	▪ Simple random sample (Random) ▪ Normal or $np \geq 10$ and $n(1-p) \geq 10$, (Large counts) ▪ For standard deviations: population $\geq 10n$ (10% rule)	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
Sampling distribution for a difference in sample proportions		$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	▪ Simple random samples (Random) ▪ Large counts ▪ 10% rule	$\mu_{(\hat{p}_1-\hat{p}_2)} = p_1 - p_2$	$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$
Sampling distribution for the sample mean		μ	$\bar{x}$	▪ SRS (Random) ▪ Normal or sample size ≥30 ▪ 10% rule	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Sampling distribution for the difference in sample means		$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	▪ SRS (Random) ▪ Normal or sample sizes ≥30 ▪ 10% rule	$\mu_{(\bar{x}_1-\bar{x}_2)} = \mu_1 - \mu_2$	$\sigma_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Standard deviation		σ	s			

Note: Other notation could also be correct if properly defined. Incorrect notation will result in lost points on the AP exam.

AP STATISTICS

UNIT 6

Inference for Categorical Data: Proportions

12–15%
AP EXAM WEIGHTING

~16–18
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 6

Multiple-choice: ~55 questions
Free-response: 2 questions

- Inference
- Investigative Task

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.	
UNC-4	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?

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.

SKILL

Selecting Statistical Methods

1.A

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

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	ESSENTIAL KNOWLEDGE
UNC-4.A Identify an appropriate confidence interval procedure for a population proportion. **[Skill 1.D]**	**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.
UNC-4.B Verify the conditions for calculating confidence intervals for a population proportion. **[Skill 4.C]**	**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

UNC-4.B

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

ESSENTIAL KNOWLEDGE

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 $n \le 10\% N$, 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

UNC-4.D

Calculate an appropriate confidence interval for a population proportion.
[Skill 3.D]

ESSENTIAL KNOWLEDGE

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.

LEARNING OBJECTIVE

UNC-4.E

Calculate an interval estimate based on a confidence interval for a population proportion.
[Skill 3.D]

ESSENTIAL KNOWLEDGE

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.

LEARNING OBJECTIVE

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]**

ESSENTIAL KNOWLEDGE

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

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**

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

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 $n \leq 10\%N$. 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

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**

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

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
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 two independent, 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

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

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**

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 two independent, 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, $\hat{p}_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$. Assuming that H_0 is true ($p_1 - p_2 = 0$ or $p_1 = p_2$), check that $n_1\hat{p}_c$, $n_1(1-\hat{p}_c)$, $n_2\hat{p}_c$, and $n_2(1-\hat{p}_c)$ 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 $\hat{p}_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$.

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.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

AP STATISTICS

UNIT 7

Inference for Quantitative Data: Means

10–18%
AP EXAM WEIGHTING

~14–16
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 7

Multiple-choice: ~50 questions
Free-response: 2 questions

- Inference and Collecting Data
- Investigative Task

UNIT 7

10–18% AP EXAM WEIGHTING | ~14–16 CLASS PERIODS

Inference for Quantitative Data: Means

BIG IDEA 1
Variation and Distribution **VAR**

- How do we know whether to use a t-test or a z-test for inference with means?

BIG IDEA 2
Patterns and Uncertainty **UNC**

- How can we make sure that samples are independent?

BIG IDEA 3
Data-Based Predictions, Decisions, and Conclusions **DAT**

- Why is it inappropriate to accept a hypothesis as true based on the results of statistical inference testing?

Developing Understanding

In this unit, students will analyze quantitative data to make inferences about population means. Students should understand that t^* and t-tests are used for inference with means when the population standard deviation, σ, is not known. Using s for σ in the formula for z gives a slightly different value, t, whose distribution, which depends on sample size, has more area in the tails than a normal distribution. The boundaries for rejecting a null hypothesis using a t-distribution tend to be further from the mean than for a normal distribution. Students should understand how and why conditions for inference with proportions and means are similar and different.

Building Course Skills

1.E 1.F 4.C 4.E

Unit 7 focuses on means, which has many similarities to the conditions and procedures for proportions. Since students sometimes confuse t-tests with z-tests, it will help to review the underlying rationales each time conditions come up. This will help students develop understanding through repeated practice in new situations. Teachers can encourage students to be mindful of notation and use the formula sheet as a reference.

Teachers can reinforce that inference testing requires careful selection of a procedure based on specific conditions for a given problem. Common errors include mislabeling conditions (e.g., incorrectly associating the large sample condition with independence), relying upon vague references to the normal distribution, or applying an inappropriate large sample condition. The null and alternative hypotheses must be clearly stated in terms of population parameters, not sample statistics. A formal decision compares the p-value to the level of significance. Students should also practice providing a numerical reference to support their claim (e.g., "Because $p < 0.05$, we reject the null hypothesis.") and interpreting findings within the context of the question.

Preparing for the AP Exam

It is critical for students to recognize that free-response questions asking whether data provide convincing evidence of some finding are asking for a significance test, not just a descriptive analysis. When using statistical inference for significance tests, students should identify the correct parameter and hypotheses, identify an appropriate test procedure and check conditions, calculate a test statistic and p-value, and provide a conclusion in context, along with a justification based on linkage between the p-value and the conclusion. For inference with means, the appropriate test will often be a t-test, but if σ is known, a z-test would be appropriate (see **2018 FRQ 6(a)**). For a t-test, conditions are (1) random sample and (2) large sample (e.g., $n > 30$). When sampling without replacement, students should also verify that the sample size is at most 10% of the population.

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods ~14–16 CLASS PERIODS
VAR-1	7.1 Introducing Statistics: Should I Worry About Error?	1.A Identify the question to be answered or problem to be solved *(not assessed)*.	
VAR-7, UNC-4	7.2 Constructing a Confidence Interval for a Population Mean	3.C Describe probability distributions. 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	7.3 Justifying a Claim About a Population Mean Based on a Confidence Interval	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-7	7.4 Setting Up a Test for a Population Mean	1.E Identify an appropriate inference method for significance tests. 1.F Identify null and alternative hypotheses. 4.C Verify that inference procedures apply in a given situation.	
VAR-7, DAT-3	7.5 Carrying Out a Test for a Population Mean	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.	

continued on next page

UNIT AT A GLANCE *(cont'd)*

Enduring Understanding	Topic	Skills	Class Periods ~14–16 CLASS PERIODS
UNC-4	**7.6 Confidence Intervals for the Difference of Two Means**	**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.	
	7.7 Justifying a Claim About the Difference of Two Means Based on a Confidence Interval	**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-7	**7.8 Setting Up a Test for the Difference of Two Population Means**	**1.E** Identify an appropriate inference method for significance tests. **1.F** Identify null and alternative hypotheses. **4.C** Verify that inference procedures apply in a given situation.	
VAR-7, DAT-3	**7.9 Carrying Out a Test for the Difference of Two Population Means**	**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.	
	7.10 Skills Focus: Selecting, Implementing, and Communicating Inference Procedures	N/A	

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 7.
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	7.2	**Predict and Confirm** After introducing the confidence interval formula for a population mean when sigma is known, $\overline{x} \pm z^* \frac{\sigma}{\sqrt{n}}$, have students discuss in small groups what will happen if we substitute the sample standard deviation s into the formula when σ is unknown. Have students use the Simulating Confidence Intervals for Population Parameter applet to test their conjectures (see link for Rossman/Chance Applets on page 210).
2	7.2 7.3 7.5	**Team Challenge** Give each team of three to four students a copy of **2004 FRQ 6**, which focuses on the connection between a one-sample t-interval, a one-sample t-test, and the unfamiliar concept of a one-sided confidence interval. Challenge teams to collaboratively produce a model solution in 30 minutes.
3	7.4	**Discussion Groups** Ask each group of three to four students to identify the conditions for performing a test about a population mean. For each condition, have them explain why the condition is required and what would go wrong with the test if the condition were violated. Have groups pair up and compare answers.
4	7.9	**Team FRQ** Give each team of four students copies of a free-response question that involves performing a two-sample t-test (e.g., **2011 FRQ 4**). Have each team member take responsibility for writing one part of the model solution (hypotheses, procedure and conditions, calculations, conclusion) with group input.
5	7.10	**Graphic Organizer** Have students work in teams of two to three to develop a flowchart for determining which inference procedure from Units 6 and 7 to use in a given setting.

TOPIC 7.1

Introducing Statistics: Why Should I Worry About Error?

Required Course Content

ENDURING UNDERSTANDING

VAR-1

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

LEARNING OBJECTIVE

VAR-1.I

Identify questions suggested by probabilities of errors in statistical inference. **[Skill 1.A]**

ESSENTIAL KNOWLEDGE

VAR-1.I.1

Random variation may result in errors in statistical inference.

SKILL

Selecting Statistical Methods

1.A

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

SKILLS

Using Probability and Simulation

3.C
Describe probability distributions.

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

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.

TOPIC 7.2
Constructing a Confidence Interval for a Population Mean

Required Course Content

ENDURING UNDERSTANDING

VAR-7
The t-distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-7.A
Describe t-distributions.
[Skill 3.C]

ESSENTIAL KNOWLEDGE

VAR-7.A.1
When s is used instead of σ to calculate a test statistic, the corresponding distribution, known as the t-distribution, varies from the normal distribution in shape, in that more of the area is allocated to the tails of the density curve than in a normal distribution.

VAR-7.A.2
As the degrees of freedom increase, the area in the tails of a t-distribution decreases.

continued on next page

ENDURING UNDERSTANDING

UNC-4

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

LEARNING OBJECTIVE

UNC-4.O

Identify an appropriate confidence interval procedure for a population mean, including the mean difference between values in matched pairs. **[Skill 1.D]**

ESSENTIAL KNOWLEDGE

UNC-4.O.1

Because σ is typically not known for distributions of quantitative variables, the appropriate confidence interval procedure for estimating the population mean of one quantitative variable for one sample is a one-sample t-interval for a mean.

UNC-4.O.2

For one quantitative variable, X, that is normally distributed, the distribution of $t = \frac{(\bar{x} - \mu)}{\frac{s}{\sqrt{n}}}$ is a t-distribution with $n-1$ degrees of freedom.

UNC-4.O.3

Matched pairs can be thought of as one sample of pairs. Once differences between pairs of values are found, inference for confidence intervals proceeds as for a population mean.

LEARNING OBJECTIVE

UNC-4.P

Verify the conditions for calculating confidence intervals for a population mean, including the mean difference between values in matched pairs. **[Skill 4.C]**

ESSENTIAL KNOWLEDGE

UNC-4.P.1

In order to calculate confidence intervals to estimate a population mean, 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 $\bar{x}$ is approximately normal (shape):
   - i. If the observed distribution is skewed, n should be greater than 30.
   - ii. If the sample size is less than 30, the distribution of the sample data should be free from strong skewness and outliers.

continued on next page

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-4.Q Determine the margin of error for a given sample size for a one-sample *t*-interval. **[Skill 3.D]**	**UNC-4.Q.1** The critical value t^* with $n-1$ degrees of freedom can be found using a table or computer-generated output. **UNC-4.Q.2** The standard error for a sample mean is given by $SE = \frac{s}{\sqrt{n}}$, where s is the sample standard deviation. **UNC-4.Q.3** For a one-sample *t*-interval for a mean, the margin of error is the critical value (t^*) times the standard error (*SE*), which equals $t^*\left(\frac{s}{\sqrt{n}}\right)$.
UNC-4.R Calculate an appropriate confidence interval for a population mean, including the mean difference between values in matched pairs. **[Skill 3.D]**	**UNC-4.R.1** The point estimate for a population mean is the sample mean, $\bar{x}$. **UNC-4.R.2** For the population mean for one sample with unknown population standard deviation, the confidence interval is $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$. **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.*

TOPIC 7.3

Justifying a Claim About a Population Mean Based on a Confidence Interval

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.S

Interpret a confidence interval for a population mean, including the mean difference between values in matched pairs. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

UNC-4.S.1

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

UNC-4.S.2

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

UNC-4.S.3

An interpretation of a confidence interval for a population mean includes a reference to the sample taken and details about the population it represents.

LEARNING OBJECTIVE

UNC-4.T

Justify a claim based on a confidence interval for a population mean, including the mean difference between values in matched pairs. **[Skill 4.D]**

ESSENTIAL KNOWLEDGE

UNC-4.T.1

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

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.

ILLUSTRATIVE EXAMPLE

UNC-4.S.3:

For interpreting a 96% confidence interval for mean foot length for all footprints found in a cave based on a particular randomly selected sample of footprints in the cave:

"We are 96% confident that the mean foot length for all footprints found in the cave falls within the confidence interval" (based on **2000 FRQ 2**).

LEARNING OBJECTIVE

UNC-4.U

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

ESSENTIAL KNOWLEDGE

UNC-4.U.1

When all other things remain the same, the width of a confidence interval for a population mean tends to decrease as the sample size increases.

UNC-4.U.2

For a single mean, the width of the interval is proportional to $\frac{1}{\sqrt{n}}$.

UNC-4.U.3

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

TOPIC 7.4

Setting Up a Test for a Population Mean

Required Course Content

ENDURING UNDERSTANDING

VAR-7

The t-distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-7.B

Identify an appropriate testing method for a population mean with unknown σ, including the mean difference between values in matched pairs.
[Skill 1.E]

ESSENTIAL KNOWLEDGE

VAR-7.B.1

The appropriate test for a population mean with unknown σ is a one-sample t-test for a population mean.

VAR-7.B.2

Matched pairs can be thought of as one sample of pairs. Once differences between pairs of values are found, inference for significance testing proceeds as for a population mean.

LEARNING OBJECTIVE

VAR-7.C

Identify the null and alternative hypotheses for a population mean with unknown σ, including the mean difference between values in matched pairs.
[Skill 1.F]

ESSENTIAL KNOWLEDGE

VAR-7.C.1

The null hypothesis for a one-sample t-test for a population mean is $H_0 : \mu = \mu_0$, where μ_0 is the hypothesized value. Depending upon the situation, the alternative hypothesis is $H_a : \mu < \mu_0$, or $H_a : \mu > \mu_0$, or $H_a : \mu \neq \mu_0$.

VAR-7.C.2

When finding the mean difference, μ_d, between values in a matched pair, it is important to define the order of subtraction.

continued on next page

SKILLS

Selecting Statistical Methods

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

1.F Identify null and alternative hypotheses.

Statistical Argumentation

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

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

LEARNING OBJECTIVE

VAR-7.D

Verify the conditions for the test for a population mean, including the mean difference between values in matched pairs. **[Skill 4.C]**

ESSENTIAL KNOWLEDGE

VAR-7.D.1

In order to make statistical inferences when testing a population mean, 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 $n \leq 10\% N$.

b. To check that the sampling distribution of $\bar{x}$ is approximately normal (shape):
   i. If the observed distribution is skewed, n should be greater than 30.
   ii. If the sample size is less than 30, the distribution of the sample data should be free from strong skewness and outliers.

TOPIC 7.5

Carrying Out a Test for a Population Mean

Required Course Content

ENDURING UNDERSTANDING

VAR-7

The t-distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-7.E

Calculate an appropriate test statistic for a population mean, including the mean difference between values in matched pairs. **[Skill 3.E]**

ESSENTIAL KNOWLEDGE

VAR-7.E.1

For a single quantitative variable when random sampling with replacement from a population that can be modeled with a normal distribution with mean μ and standard deviation σ, the sampling distribution of $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ has a t-distribution with $n - 1$ degrees of freedom.

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.

4.E

Justify a claim using a decision based on significance tests.

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

ENDURING UNDERSTANDING

DAT-3

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

LEARNING OBJECTIVE

DAT-3.E

Interpret the p-value of a significance test for a population mean, including the mean difference between values in matched pairs.
[Skill 4.B]

DAT-3.F

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

ESSENTIAL KNOWLEDGE

DAT-3.E.1

An interpretation of the p-value of a significance test for a population mean should recognize that the p-value is computed by assuming that the null hypothesis is true, i.e., by assuming that the true population mean is equal to the particular value stated in the null hypothesis.

DAT-3.F.1

A formal decision explicitly compares the p-value to the significance α. If the p-value $\le \alpha$, then reject the null hypothesis, $H_0 : \mu = \mu_0$. If the p-value $> \alpha$, then fail to reject the null hypothesis.

DAT-3.F.2

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

TOPIC 7.6

Confidence Intervals for the Difference of Two Means

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

Identify an appropriate confidence interval procedure for a difference of two population means.
[Skill 1.D]

ESSENTIAL KNOWLEDGE

UNC-4.V.1

Consider a simple random sample from population 1 of size n_1, mean μ_1, and standard deviation σ_1 and a second simple random sample from population 2 of size n_2, mean μ_2, and standard deviation σ_2. If the distributions of populations 1 and 2 are normal or if both n_1 and n_2 are greater than 30, then the sampling distribution of the difference of means, $\bar{x}_1 - \bar{x}_2$ is also normal. The mean for the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is $\mu_1 - \mu_2$. The standard deviation of $\bar{x}_1 - \bar{x}_2$ is $\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}$.

UNC-4.V.2

The appropriate confidence interval procedure for one quantitative variable for two independent samples is a two-sample t-interval for a difference between population means.

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	ESSENTIAL KNOWLEDGE
UNC-4.W Verify the conditions to calculate confidence intervals for the difference of two population means. **[Skill 4.C]**	**UNC-4.W.1** In order to calculate confidence intervals to estimate a difference of population means, we must check for independence and that the sampling distribution is approximately normal: a. To check for independence: i. Data should be collected using two independent, random samples or a randomized experiment. ii. When sampling without replacement, check that $n_1 \leq 10\% N_1$ and $n_2 \leq 10\% N_2$. b. To check that the sampling distribution of $(\bar{x}_1 - \bar{x}_2)$ should be approximately normal (shape): i. If the observed distributions are skewed, both n_1 and n_2 should be greater than 30.
UNC-4.X Determine the margin of error for the difference of two population means. **[Skill 3.D]**	**UNC-4.X.1** For the difference of two sample means, the margin of error is the critical value (t^*) times the standard error (*SE*) of the difference of two means. **UNC-4.X.2** The standard error for the difference in two sample means with sample standard deviations, s_1 and s_2, is $\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$.
UNC-4.Y Calculate an appropriate confidence interval for a difference of two population means. **[Skill 3.D]**	**UNC-4.Y.1** The point estimate for the difference of two population means is the difference in sample means, $\bar{x}_1 - \bar{x}_2$.

continued on next page

LEARNING OBJECTIVE

UNC-4.Y

Calculate an appropriate confidence interval for a difference of two population means. **[Skill 3.D]**

ESSENTIAL KNOWLEDGE

UNC-4.Y.2

For a difference of two population means where the population standard deviations are not known, the confidence interval is $\left(\bar{x}_1 - \bar{x}_2\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ where $\pm t^*$ are the critical values for the central C% of a t-distribution with appropriate degrees of freedom that can be found using technology.

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.

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.

ILLUSTRATIVE EXAMPLE

UNC-4.Z.2:

For interpreting a confidence interval for a difference between mean response times for two fire stations (northern – southern): "Based on these samples, one can be 95 percent confident that the difference in the population mean response times (northern – southern) is between −2.37 minutes and 0.37 minutes" (2009 FRQ 4).

TOPIC 7.7

Justifying a Claim About the Difference of Two Means Based on a Confidence Interval

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.Z Interpret a confidence interval for a difference of population means. **[Skill 4.B]**	**UNC-4.Z.1** In repeated random sampling with the same sample size, approximately C% of confidence intervals created will capture the difference of population means. **UNC-4.Z.2** An interpretation for a confidence interval for the difference of two population means should include a reference to the samples taken and details about the populations they represent.
UNC-4.AA Justify a claim based on a confidence interval for a difference of population means. **[Skill 4.D]**	**UNC-4.AA.1** A confidence interval for a difference of population means provides an interval of values that may provide sufficient evidence to support a particular claim in context.
UNC-4.AB Identify the effects of sample size on the width of a confidence interval for the difference of two means. **[Skill 4.A]**	**UNC-4.AB.1** When all other things remain the same, the width of the confidence interval for the difference of two means tends to decrease as the sample sizes increase.

TOPIC 7.8

Setting Up a Test for the Difference of Two Population Means

Required Course Content

ENDURING UNDERSTANDING

VAR-7

The t-distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-7.F

Identify an appropriate selection of a testing method for a difference of two population means. **[Skill 1.E]**

ESSENTIAL KNOWLEDGE

VAR-7.F.1

For a quantitative variable, the appropriate test for a difference of two population means is a two-sample t-test for a difference of two population means.

LEARNING OBJECTIVE

VAR-7.G

Identify the null and alternative hypotheses for a difference of two population means. **[Skill 1.F]**

ESSENTIAL KNOWLEDGE

VAR-7.G.1

The null hypothesis for a two-sample t-test for a difference of two population means, μ_1 and μ_2, is: $H_0: \mu_1 - \mu_2 = 0$, or $H_0: \mu_1 = \mu_2$. The alternative hypothesis is $H_a: \mu_1 - \mu_2 < 0$, or $H_a: \mu_1 - \mu_2 > 0$, or $H_a: \mu_1 - \mu_2 \neq 0$, or $H_a: \mu_1 > \mu_2$, or $H_a: \mu_1 < \mu_2$, or $H_a: \mu_1 \neq \mu_2$.

continued on next page

SKILLS

Selecting Statistical Methods

1.E

Identify an appropriate inference method for significance tests.

1.F

Identify null and alternative hypotheses.

Statistical Argumentation

4.C

Verify that inference procedures apply in a given situation.

LEARNING OBJECTIVE

VAR-7.H

Verify the conditions for the significance test for the difference of two population means. **[Skill 4.C]**

ESSENTIAL KNOWLEDGE

VAR-7.H.1

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

a. Individual observations should be independent:
   i. Data should be collected using simple 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. The sampling distribution of $\bar{x}_1 - \bar{x}_2$ should be approximately normal (shape).
   i. If the observed distribution is skewed, both n_1 and n_2 should be greater than 30.
   ii. If the sample size is less than 30, the distribution of the sample data should be free from strong skewness and outliers. This should be checked for BOTH samples.

TOPIC 7.9

Carrying Out a Test for the Difference of Two Population Means

Required Course Content

ENDURING UNDERSTANDING

VAR-7

The t-distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-7.I

Calculate an appropriate test statistic for a difference of two means. **[Skill 3.E]**

ESSENTIAL KNOWLEDGE

VAR-7.I.1

For a single quantitative variable, data collected using independent random samples or a randomized experiment from two populations, each of which can be modeled with a normal distribution, the sampling distribution of $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ is an approximate t-distribution with degrees of freedom that can be found using technology. The degrees of freedom fall between the smaller of $n_1 - 1$ and $n_2 - 1$ and $n_1 + n_2 - 2$.

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

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**

ILLUSTRATIVE EXAMPLE

VAR-7.I.1:

In a study comparing mean recovery times for two surgical procedures to repair a torn anterior cruciate ligament (ACL), the group receiving one procedure had a sample size of 110, while the other group receiving the other procedure had a sample size of 100. The degrees of freedom fall between 100 (the smaller of 110 and 100) and 208 (110 + 100 − 2). The degrees of freedom may be determined using technology. If the test statistic for this study is $t \approx 7.13$, then the p-value is the area greater than 7.13 for a t-distribution with $df = 207.18$ (**2018 FRQ 4**).

ENDURING UNDERSTANDING

DAT-3

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

LEARNING OBJECTIVE

DAT-3.G

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

ESSENTIAL KNOWLEDGE

DAT-3.G.1

An interpretation of the p-value of a significance test for a two-sample difference of population means should recognize that the p-value is computed by assuming that the null hypothesis is true, i.e., by assuming that the true population means are equal to each other.

LEARNING OBJECTIVE

DAT-3.H

Justify a claim about the population based on the results of a significance test for a difference of two population means in context. **[Skill 4.E]**

ESSENTIAL KNOWLEDGE

DAT-3.H.1

A formal decision explicitly compares the p-value to the significance α. If the $p\text{-value} \leq \alpha$, then reject the null hypothesis, $H_0: \mu_1 - \mu_2 = 0$, or $H_0: \mu_1 = \mu_2$. If the p-value $> \alpha$, then fail to reject the null hypothesis.

DAT-3.H.2

The results of a significance test for a two-sample test for a difference between two population means can serve as the statistical reasoning to support the answer to a research question about the populations that were sampled.

TOPIC 7.10

Skills Focus: Selecting, Implementing, and Communicating Inference Procedures

Required Course Content

This topic is intended to focus on the skill of selecting an appropriate inference procedure, now that students have a range of options. Students should be given opportunities to practice when and how to apply all learning objectives relating to inference involving proportions or means.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

AP STATISTICS

UNIT 8

Inference for Categorical Data: Chi-Square

2–5%
AP EXAM WEIGHTING

~10–11
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 8

Multiple-choice: ~30 questions
Free-response: 2 questions

- Inference
- Inference and Exploring Data/ Collecting Data

UNIT 8

2–5% AP EXAM WEIGHTING | **~10–11** CLASS PERIODS

Inference for Categorical Data: Chi-Square

BIG IDEA 1
Variation and Distribution **VAR**

- How does increasing the degrees of freedom influence the shape of the chi-square distribution?

BIG IDEA 3
Data-Based Predictions, Decisions, and Conclusions **DAT**

- Why is it inappropriate to use statistical inference to justify a claim that there is *no association* between variables?

Developing Understanding

Unit 6 introduced inference for proportions of categorical data. Unit 8 introduces chi-square tests, which can be used when there are two or more categories. Students need to understand how to select from the following tests: the chi-square test for goodness of fit (for a distribution of proportions of one categorical variable in a population), the chi-square test for independence (for associations between categorical variables within a single population), or the chi-square test for homogeneity (for comparing distributions of a categorical variable across populations or treatments). To integrate conceptual understanding, teachers can make connections between frequency tables, conditional probability, and calculating expected counts. The chi-square statistic is introduced to measure the distance between observed and expected counts relative to expected counts.

Building Course Skills

1.E **3.E** **4.C** **4.E**

In Unit 8, students should continue applying the same problem-solving structure to chi-square significance testing: State the hypotheses in words, explicitly identify the correct procedure, verify conditions, calculate the test statistic and the p-value, and then draw a conclusion in context that is directly linked to the p-value. Students should have opportunities to practice the distinctive elements for each type of chi-square test, such as analysis of expected counts, degrees of freedom, verbally stated hypotheses, and two-way tables.

When the p-value is large, drawing an appropriate conclusion is challenging for students. Saying there is "no association" between two variables is equivalent to incorrectly "accepting the null hypothesis." Instead, teachers can teach students to use nondeterministic language in their conclusions, that is, "The data do not provide strong enough evidence to conclude that the variables are associated." Students should have frequent opportunities to practice writing, with detailed feedback to help them improve.

Preparing for the AP Exam

When writing hypotheses, students should refer to the population, using language from the question. For example, "The null hypothesis is that the age group at diagnosis and gender are independent (i.e., they are not associated) for the population of people currently being treated for schizophrenia" (see **Scoring Guidelines for 2017 FRQ 5**). As always, students should name the test and provide evidence verifying appropriate conditions. For chi-square tests, the conditions are (1) random selection or randomized experiment and (2) large counts. Students should be sure to say that all expected counts (rather than actual counts) are at least 5. Students need to clearly present calculations and state the conclusion in context with linkage to p-values. Students should avoid tacitly accepting the null hypothesis. If the p-value is greater than conventional significance levels, the correct conclusion of a chi-square test for independence would be that there is insufficient evidence that there is an association.

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods ~10–11 CLASS PERIODS
VAR-1	**8.1 Introducing Statistics: Are My Results Unexpected?**	1.A Identify the question to be answered or problem to be solved *(not assessed)*.	
VAR-8	**8.2 Setting Up a Chi-Square Goodness of Fit Test**	3.C Describe probability distributions. 1.F Identify null and alternative hypotheses. 1.E Identify an appropriate inference method for significance tests. 3.A Determine relative frequencies, proportions, or probabilities using simulation or calculations. 4.C Verify that inference procedures apply in a given situation.	
VAR-8, DAT-3	**8.3 Carrying Out a Chi-Square Test for Goodness of Fit**	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.	
VAR-8	**8.4 Expected Counts in Two-Way Tables**	3.A Determine relative frequencies, proportions, or probabilities using simulation or calculations.	
VAR-8	**8.5 Setting Up a Chi-Square Test for Homogeneity or Independence**	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-8, DAT-3	**8.6 Carrying Out a Chi-Square Test for Homogeneity or Independence**	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.	
	8.7 Skills Focus: Selecting an Appropriate Inference Procedure for Categorical Data	N/A	

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 8. 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	8.1	**Simulation** Prepare several bags with an identical mix of at least 250 chips or beads of three colors in different proportions (e.g., red = 0.5, white = 0.3, blue = 0.2). Have each student take a random sample of 25 chips or beads from the bag, calculate $\sum \frac{(Observed\ count - Expected\ count)^2}{Expected\ count}$, and plot their value on a class dotplot. Use this graph to introduce the chi-square distribution with $df = 2$.
2	8.5	**Discussion Groups** Give each group of three to four students an example of a chi-square test involving a two-way table. Have students work together to state appropriate hypotheses, describe a Type 1 and Type 2 error in context, and give a possible consequence of each of those errors.
3	8.7	**Graphic Organizer** Have students work in teams of two to three to develop a chart that summarizes the three types of chi-square tests, including when each is appropriate, as well as the hypotheses, conditions, and degrees of freedom.

SKILL

Selecting Statistical Methods

1.A

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

TOPIC 8.1

Introducing Statistics: Are My Results Unexpected?

Required Course Content

ENDURING UNDERSTANDING

VAR-1

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

LEARNING OBJECTIVE

VAR-1.J

Identify questions suggested by variation between observed and expected counts in categorical data.
[Skill 1.A]

ESSENTIAL KNOWLEDGE

VAR-1.J.1

Variation between what we find and what we expect to find may be random or not.

TOPIC 8.2

Setting Up a Chi-Square Goodness of Fit Test

Required Course Content

ENDURING UNDERSTANDING

VAR-8

The chi-square distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-8.A

Describe chi-square distributions. **[Skill 3.C]**

ESSENTIAL KNOWLEDGE

VAR-8.A.1

Expected counts of categorical data are counts consistent with the null hypothesis. In general, an expected count is a sample size times a probability.

VAR-8.A.2

The chi-square statistic measures the distance between observed and expected counts relative to expected counts.

VAR-8.A.3

Chi-square distributions have positive values and are skewed right. Within a family of density curves, the skew becomes less pronounced with increasing degrees of freedom.

LEARNING OBJECTIVE

VAR-8.B

Identify the null and alternative hypotheses in a test for a distribution of proportions in a set of categorical data. **[Skill 1.F]**

ESSENTIAL KNOWLEDGE

VAR-8.B.1

For a chi-square goodness-of-fit test, the null hypothesis specifies null proportions for each category, and the alternative hypothesis is that at least one of these proportions is not as specified in the null hypothesis.

continued on next page

SKILLS

Using Probability and Simulation

3.C

Describe probability distributions.

Selecting Statistical Methods

1.F

Identify null and alternative hypotheses.

1.E

Identify an appropriate inference method for significance tests.

Using Probability and Simulation

3.A

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

Statistical Argumentation

4.C

Verify that inference procedures apply in a given situation.

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-8.C Identify an appropriate testing method for a distribution of proportions in a set of categorical data. **[Skill 1.E]**	**VAR-8.C.1** When considering a distribution of proportions for one categorical variable, the appropriate test is the chi-square test for goodness of fit.
VAR-8.D Calculate expected counts for the chi-square test for goodness of fit. **[Skill 3.A]**	**VAR-8.D.1** Expected counts for a chi-square goodness-of-fit test are (sample size) (null proportion).
VAR-8.E Verify the conditions for making statistical inferences when testing goodness of fit for a chi-square distribution. **[Skill 4.C]**	**VAR-8.E.1** In order to make statistical inferences for a chi-square test for goodness of fit we must check the following: a. To check for independence: i. Data should be collected using a random sample or randomized experiment. ii. When sampling without replacement, check that $n \leq 10\%N$. b. The chi-square test for goodness of fit becomes more accurate with more observations, so large counts should be used (shape). i. A conservative check for large counts is that all expected counts should be greater than 5.

TOPIC 8.3
Carrying Out a Chi-Square Test for Goodness of Fit

Required Course Content

ENDURING UNDERSTANDING

VAR-8

The chi-square distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-8.F

Calculate the appropriate statistic for the chi-square test for goodness of fit.
[Skill 3.E]

ESSENTIAL KNOWLEDGE

VAR-8.F.1

The test statistic for the chi-square test for goodness of fit is

$$\chi^2 = \sum \frac{(Observed\ count - Expected\ count)^2}{Expected\ count},$$

with

$degrees\ of\ freedom = number\ of\ categories - 1.$

VAR-8.F.2

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 (chi-square).

LEARNING OBJECTIVE

VAR-8.G

Determine the p-value for chi-square test for goodness of fit significance test.
[Skill 3.E]

ESSENTIAL KNOWLEDGE

VAR-8.G.1

The p-value for a chi-square test for goodness of fit for a number of degrees of freedom is found using the appropriate table or computer generated output.

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.

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

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

ENDURING UNDERSTANDING

DAT-3

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

LEARNING OBJECTIVE

DAT-3.I

Interpret the p-value for the chi-square test for goodness of fit. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

DAT-3.I.1

An interpretation of the p-value for the chi-square test for goodness of fit is the probability, given the null hypothesis and probability model are true, of obtaining a test statistic as, or more, extreme than the observed value.

LEARNING OBJECTIVE

DAT-3.J

Justify a claim about the population based on the results of a chi-square test for goodness of fit. **[Skill 4.E]**

ESSENTIAL KNOWLEDGE

DAT-3.J.1

A decision to either reject or fail to reject the null hypothesis is based on comparison of the p-value to the significance level, α.

DAT-3.J.2

The results of a chi-square test for goodness of fit can serve as the statistical reasoning to support the answer to a research question about the population that was sampled.

TOPIC 8.4

Expected Counts in Two-Way Tables

SKILL

Using Probability and Simulation

3.A

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

Required Course Content

ENDURING UNDERSTANDING

VAR-8

The chi-square distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-8.H

Calculate expected counts for two-way tables of categorical data. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

VAR-8.H.1

The expected count in a particular cell of a two-way table of categorical data can be calculated using the formula:

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

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 8.5
Setting Up a Chi-Square Test for Homogeneity or Independence

Required Course Content

ENDURING UNDERSTANDING

VAR-8

The chi-square distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-8.I

Identify the null and alternative hypotheses for a chi-square test for homogeneity or independence. **[Skill 1.F]**

ESSENTIAL KNOWLEDGE

VAR-8.I.1

The appropriate hypotheses for a chi-square test for homogeneity are:

H_0: There is no difference in distributions of a categorical variable across populations or treatments.

H_a: There is a difference in distributions of a categorical variable across populations or treatments.

VAR-8.I.2

The appropriate hypotheses for a chi-square test for independence are:

H_0: There is no association between two categorical variables in a given population or the two categorical variables are independent.

H_a: Two categorical variables in a population are associated or dependent.

continued on next page

LEARNING OBJECTIVE

VAR-8.J

Identify an appropriate testing method for comparing distributions in two-way tables of categorical data. **[Skill 1.E]**

ESSENTIAL KNOWLEDGE

VAR-8.J.1

When comparing distributions to determine whether proportions in each category for categorical data collected from different populations are the same, the appropriate test is the chi-square test for homogeneity.

VAR-8.J.2

To determine whether row and column variables in a two-way table of categorical data might be associated in the population from which the data were sampled, the appropriate test is the chi-square test for independence.

LEARNING OBJECTIVE

VAR-8.K

Verify the conditions for making statistical inferences when testing a chi-square distribution for independence or homogeneity. **[Skill 4.C]**

ESSENTIAL KNOWLEDGE

VAR-8.K.1

In order to make statistical inferences for a chi-square test for two-way tables (homogeneity or independence), we must verify the following:

a. To check for independence:
   i. For a test for independence: Data should be collected using a simple random sample.
   ii. For a test for homogeneity: Data should be collected using a stratified random sample or randomized experiment.
   iii. When sampling without replacement, check that $n \leq 10\% N$.

b. The chi-square tests for independence and homogeneity become more accurate with more observations, so large counts should be used (shape).
   i. A conservative check for large counts is that all expected counts should be greater than 5.

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 8.6

Carrying Out a Chi-Square Test for Homogeneity or Independence

Required Course Content

ENDURING UNDERSTANDING

VAR-8

The chi-square distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-8.L

Calculate the appropriate statistic for a chi-square test for homogeneity or independence. **[Skill 3.E]**

ESSENTIAL KNOWLEDGE

VAR-8.L.1

The appropriate test statistic for a chi-square test for homogeneity or independence is the chi-square statistic:

$$\chi^2 = \sum \frac{(Observed\ count - Expected\ count)^2}{Expected\ count},$$

with degrees of freedom equal to:

$(number\ of\ rows - 1)(number\ of\ columns - 1).$

LEARNING OBJECTIVE

VAR-8.M

Determine the p-value for a chi-square significance test for independence or homogeneity. **[Skill 3.E]**

ESSENTIAL KNOWLEDGE

VAR-8.M.1

The p-value for a chi-square test for independence or homogeneity for a number of degrees of freedom is found using the appropriate table or technology.

VAR-8.M.2

For a test of independence or homogeneity for a two-way table, the p-value is the proportion of values in a chi-square distribution with appropriate degrees of freedom that are equal to or larger than the test statistic.

continued on next page

ENDURING UNDERSTANDING

DAT-3

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

LEARNING OBJECTIVE

DAT-3.K

Interpret the p-value for the chi-square test for homogeneity or independence. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

DAT-3.K.1

An interpretation of the p-value for the chi-square test for homogeneity or independence is the probability, given the null hypothesis and probability model are true, of obtaining a test statistic as, or more, extreme than the observed value.

LEARNING OBJECTIVE

DAT-3.L

Justify a claim about the population based on the results of a chi-square test for homogeneity or independence. **[Skill 4.E]**

ESSENTIAL KNOWLEDGE

DAT-3.L.1

A decision to either reject or fail to reject the null hypothesis for a chi-square test for homogeneity or independence is based on comparison of the p-value to the significance level, α.

DAT-3.L.2

The results of a chi-square test for homogeneity or independence can serve as the statistical reasoning to support the answer to a research question about the population that was sampled (independence) or the populations that were sampled (homogeneity).

TOPIC 8.7

Skills Focus: Selecting an Appropriate Inference Procedure for Categorical Data

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

Required Course Content

This topic is intended to focus on the skill of selecting an appropriate inference procedure now that students have a range of options. Students should be given opportunities to practice when and how to apply all learning objectives relating to inference for categorical data.

AP STATISTICS

UNIT 9

Inference for Quantitative Data: Slopes

2–5%
AP EXAM WEIGHTING

~7–8
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 9

Multiple-choice: ~25 questions
Free-response: 1 question

- Inference and Exploring Data

UNIT 9

2–5% AP EXAM WEIGHTING | **~7–8** CLASS PERIODS

Inference for Quantitative Data: Slopes

BIG IDEA 1
Variation and Distribution **VAR**

- How can there be variability in slope if the slope statistic is uniquely determined for a line of best fit?

BIG IDEA 2
Patterns and Uncertainty **UNC**

- When is it appropriate to perform inference about the slope of a population regression line based on sample data?

BIG IDEA 3
Data-Based Predictions, Decisions, and Conclusions **DAT**

- Why do we not conclude that there is no correlation between two variables based on the results of a statistical inference for slopes?

Developing Understanding

Students may be surprised to learn that there is variability in slope. In their experience in previous courses, the slope of the line of best fit does not vary for a particular set of bivariate quantitative data. However, suppose that every student in a university physics course collects data on spring length for 10 different hanging masses and calculates the least-squares regression line for their sample data. The students' slopes would likely vary as part of an approximately normal sampling distribution centered at the (true) slope of the population regression line relating spring length to hanging mass. In this unit, students will learn how to construct confidence intervals for and perform significance tests about the slope of a population regression line when appropriate conditions are met.

Building Course Skills

1.A **4.B** **4.E**

In Unit 9, students should have multiple opportunities to practice interpreting the slope, y-intercept, r^2, standard deviation of the residual s, and standard error of the slope in context from computer output. They should refrain from using deterministic language such as "a 1-foot increase in X is associated with a 0.445-point increase in Y," instead framing the association in terms of potential outcomes (i.e., "a *predicted* 0.445-point increase"). Students should also practice writing "increase" or "additional" for both variables, not just the dependent variable.

Students should practice identifying what the question is asking or what needs to be solved. Without careful reading, students often provide answers that are not relevant or required, for example, conducting a significance test when the question does not call for one, or giving the expected number of successes or failures when asked to calculate a probability. Teachers can have them practice identifying the task before they begin, then checking that the response they've provided addresses the task.

Preparing for the AP Exam

Students should pay attention to timing as they work through full-length sections of past exams in order to leave enough time to complete the investigative task, which is weighted more heavily than the other free-response questions. The investigative task includes both familiar course content and questions requiring extended reasoning. As an example of a straightforward application of a topic from this unit, **2007 Form B FRQ 6 part b** asks students to find a 95% confidence interval for the slope of a regression line. This familiar task gives students an opportunity to gain confidence and earn some credit, and it serves as an entry to subsequent parts of the question. Although the investigative task will require students to transfer course skills to unfamiliar settings, students who understand course content will have everything they need to complete the task.

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods ~7–8 CLASS PERIODS
VAR-1	**9.1 Introducing Statistics: Do Those Points Align?**	**1.A** Identify the question to be answered or problem to be solved *(not assessed)*.	
UNC-4	**9.2 Confidence Intervals for the Slope of a Regression Model**	**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	**9.3 Justifying a Claim About the Slope of a Regression Model Based on a Confidence Interval**	**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-7	**9.4 Setting Up a Test for the Slope of a Regression Model**	**1.E** Identify an appropriate inference method for significance tests. **1.F** Identify null and alternative hypotheses. **4.C** Verify that inference procedures apply in a given situation.	
VAR-7, DAT-3	**9.5 Carrying Out a Test for the Slope of a Regression Model**	**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.	
	9.6 Skills Focus: Selecting an Appropriate Inference Procedure	N/A	

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 9.
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	9.2	**Note-Taking** Begin by having students use a chart to record the symbols for statistics and parameters that have been used previously to construct confidence intervals: Statistic — Parameter $\hat{p}$ — p $\bar{x}$ — μ s — σ Then, when constructing a confidence interval for the population slope parameter, have students add a new row for the symbols for the sample slope and population slope: b and β, respectively. This will reinforce that the slope of the least-squares regression line is a sample statistic and can be used to estimate the population parameter slope.
2	9.3	**Error Analysis** Give students some raw data on the distance and cost to fly from their hometown to various major cities. For example: Flying from ____ to ____ Distance — Cost 512 miles — $179 1256 miles — $257 3256 miles — $387 Then introduce some questions justifying a claim and error analysis. For example, how could you refute a claim that the average cost per mile (the population slope) is $0.50 per mile if you believe it to be false?
3	9.5	**Notation Read Aloud** Have students read AP Exam questions aloud (e.g., **2011 FRQ 5**, **2010 Form B FRQ 6**, **2005 Form B FRQ 5**, and **2001 FRQ 6**), including the given notation. Remind students that the computer output provides the two-sided p-value, and that there are two different p-values in the chart: The top p-value is for the intercept, and the bottom p-value is for the slope. Then have students discuss each of the values in the computer output and carry out a test for the slope of a regression model.

Activity 1 chart:

Statistic	Parameter
$\hat{p}$	p
$\bar{x}$	μ
s	σ

Activity 2 data:

Flying from ____ to ____	
Distance	**Cost**
512 miles	$179
1256 miles	$257
3256 miles	$387

SKILL

Selecting Statistical Methods

1.A

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

TOPIC 9.1

Introducing Statistics: Do Those Points Align?

Required Course Content

ENDURING UNDERSTANDING

VAR-1

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

LEARNING OBJECTIVE

VAR-1.K

Identify questions suggested by variation in scatter plots. **[Skill 1.A]**

ESSENTIAL KNOWLEDGE

VAR-1.K.1

Variation in points' positions relative to a theoretical line may be random or non-random.

TOPIC 9.2

Confidence Intervals for the Slope of a Regression Model

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.AC

Identify an appropriate confidence interval procedure for a slope of a regression model. **[Skill 1.D]**

ESSENTIAL KNOWLEDGE

UNC-4.AC.1

Consider a response variable, y, that is linearly related to an explanatory variable, x. For a simple random sample of n observations, the sample regression line, $\hat{y} = a + bx$, is an estimate of the population regression line $\mu_y = \alpha + \beta x$. For a particular observation, (x_i, y_i), the residual from the sample regression line, $y_i - \hat{y}_i = y_i - (a + bx_i)$, is an estimate of $y_i - (\alpha + \beta x_i)$, the deviation of the response variable from the population regression line. For all points (x, y) in the population, the standard deviation of all of the deviations of the response variable from the population regression line, σ, can be estimated by the standard deviation of the residuals from the sample regression line, $s = \sqrt{\frac{\sum\left(y_i - \hat{y}_i\right)^2}{n-2}}$. (Note: This formula uses $n-2$ in the denominator instead of $n-1$ because two parameters, α and β, must be estimated to obtain the predicted values from the least-squares regression line.)

UNC-4.AC.2

For a simple random sample of n observations, let b represent the slope of a sample regression line. Then the mean of the sampling distribution for b equals the population slope: $\mu_b = \beta$. The standard deviation of the sampling distribution for b is $\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{n}}$.

UNC-4.AC.3

The appropriate confidence interval for the slope of a regression model is a t-interval for the slope.

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.

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-4.AD Verify the conditions to calculate confidence intervals for the slope of a regression model. **[Skill 4.C]**	**UNC-4.AD.1** In order to calculate a confidence interval to estimate the slope of a regression line, we must check the following: a. The true relationship between x and y is linear. Analysis of residuals may be used to verify linearity. b. The standard deviation for y, σ_y, does not vary with x. Analysis of residuals may be used to check for approximately equal standard deviations for all x. c. To check for independence: i. Data should be collected using a random sample or a randomized experiment. ii. When sampling without replacement, check that $n \leq 10\%N$. d. For a particular value of x, the responses (y-values) are approximately normally distributed. Analysis of graphical representations of residuals may be used to check for normality. i. If the observed distribution is skewed, n should be greater than 30.
UNC-4.AE Determine the given margin of error for the slope of a regression model. **[Skill 3.D]**	**UNC-4.AE.1** For the slope of a regression line, the margin of error is the critical value (t^*) times the standard error (*SE*) of the slope. **UNC-4.AE.2** The standard error for the slope of a regression line with sample standard deviation, s, is $SE = \frac{s}{s_x\sqrt{n-1}}$, where s is the estimate of σ and s_x is the sample standard deviation of the x values.
UNC-4.AF Calculate an appropriate confidence interval for the slope of a regression model. **[Skill 3.D]**	**UNC-4.AF.1** The point estimate for the slope of a regression model is the slope of the line of best fit, b. **UNC-4.AF.2** For the slope of a regression model, the interval estimate is $b \pm t^*(SE_b)$.

TOPIC 9.3

Justifying a Claim About the Slope of a Regression Model Based on a Confidence Interval

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.AG

Interpret a confidence interval for the slope of a regression model. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

UNC-4.AG.1

In repeated random sampling with the same sample size, approximately C% of confidence intervals created will capture the slope of the regression model, i.e., the true slope of the population regression model.

UNC-4.AG.2

An interpretation for a confidence interval for the slope of a regression line should include a reference to the sample taken and details about the population it represents.

LEARNING OBJECTIVE

UNC-4.AH

Justify a claim based on a confidence interval for the slope of a regression model. **[Skill 4.D]**

ESSENTIAL KNOWLEDGE

UNC-4.AH.1

A confidence interval for the slope of a regression model provides an interval of values that may provide sufficient evidence to support a particular claim in context.

LEARNING OBJECTIVE

UNC-4.AI

Identify the effects of sample size on the width of a confidence interval for the slope of a regression model. **[Skill 4.A]**

ESSENTIAL KNOWLEDGE

UNC-4.AI.1

When all other things remain the same, the width of the confidence interval for the slope of a regression model tends to decrease as the sample size increases.

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 > **Inference**

SKILLS

Selecting Statistical Methods

1.E

Identify an appropriate inference method for significance tests.

1.F

Identify null and alternative hypotheses.

Statistical Argumentation

4.C

Verify that inference procedures apply in a given situation.

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

TOPIC 9.4

Setting Up a Test for the Slope of a Regression Model

Required Course Content

ENDURING UNDERSTANDING

VAR-7

The *t*-distribution may be used to model variation.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-7.J Identify the appropriate selection of a testing method for a slope of a regression model. **[Skill 1.E]**	**VAR-7.J.1** The appropriate test for the slope of a regression model is a *t*-test for a slope.
VAR-7.K Identify appropriate null and alternative hypotheses for a slope of a regression model. **[Skill 1.F]**	**VAR-7.K.1** The null hypothesis for a *t*-test for a slope is: $H_0 : \beta = \beta_0$, where β_0 is the hypothesized value from the null hypothesis. The alternative hypothesis is $H_0 : \beta < \beta_0$ or $H_0 : \beta > \beta_0$, or $H_0 : \beta \neq \beta_0$.

continued on next page

LEARNING OBJECTIVE

VAR-7.L

Verify the conditions for the significance test for the slope of a regression model.
[Skill 4.C]

ESSENTIAL KNOWLEDGE

VAR-7.L.1

In order to make statistical inferences when testing for the slope of a regression model, we must check the following:

a. The true relationship between x and y is linear. Analysis of residuals may be used to verify linearity.

b. The standard deviation for y, σ_y, does not vary with x. Analysis of residuals may be used to check for approximately equal standard deviations for all x.

c. To check for independence:

   i. Data should be collected using a random sample or a randomized experiment.

   ii. When sampling without replacement, check that $n \leq 10\%N$.

d. For a particular value of x, the responses (y-values) are approximately normally distributed. Analysis of graphical representations of residuals may be used to check for normality.

   i. If the observed distribution is skewed, n should be greater than 30.

   ii. If the sample size is less than 30, the distribution of the sample data should be free from strong skewness and outliers.

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 9.5
Carrying Out a Test for the Slope of a Regression Model

Required Course Content

ENDURING UNDERSTANDING

VAR-7

The t-distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-7.M

Calculate an appropriate test statistic for the slope of a regression model. **[Skill 3.E]**

ESSENTIAL KNOWLEDGE

VAR-7.M.1

The distribution of the slope of a regression model assuming all conditions are satisfied and the null hypothesis is true (null distribution) is a t-distribution.

VAR-7.M.2

For simple linear regression when random sampling from a population for the response that can be modeled with a normal distribution for each value of the explanatory variable, the sampling distribution of $t = \frac{b - \beta}{SE_b}$ has a t-distribution with degrees of freedom equal to $n - 2$. When testing the slope in a simple linear regression model with one parameter, the slope, the test for the slope has $df = n - 1$.

continued on next page

ENDURING UNDERSTANDING

DAT-3

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

LEARNING OBJECTIVE

DAT-3.M

Interpret the p-value of a significance test for the slope of a regression model.
[Skill 4.B]

ESSENTIAL KNOWLEDGE

DAT-3.M.1

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

LEARNING OBJECTIVE

DAT-3.N

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

ESSENTIAL KNOWLEDGE

DAT-3.N.1

A formal decision explicitly compares the p-value to the significance α. If the p-value $\leq \alpha$, then reject the null hypothesis, $H_0 : \beta = \beta_0$. If the p-value $> \alpha$, then fail to reject the null hypothesis.

DAT-3.N.2

The results of a significance test for the slope of a regression model can serve as the statistical reasoning to support the answer to a research question about that sample.

TOPIC 9.6

Skills Focus: Selecting an Appropriate Inference Procedure

AVAILABLE RESOURCE

- Classroom Resource > **Inference**

Required Course Content

This topic is intended to focus on the skill of selecting an appropriate inference procedure now that students have a range of options. Students should be given opportunities to practice when and how to apply all learning objectives relating to inference.