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

Required Course Content

ENDURING UNDERSTANDING

VAR-6

The normal distribution may be used to model variation.

LEARNING OBJECTIVE

VAR-6.A

Calculate the probability that a particular value lies in a given interval of a normal distribution. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

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.

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.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

SKILLS

Using Probability and Simulation

3.A

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

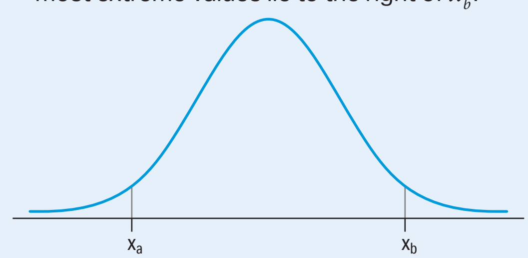
3.C

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.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-6.B Determine the interval associated with a given area in a normal distribution. **[Skill 3.A]**	**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. 
VAR-6.C Determine the appropriateness of using the normal distribution to approximate probabilities for unknown distributions. **[Skill 3.C]**	**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

UNC-3.K

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

ESSENTIAL KNOWLEDGE

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.

LEARNING OBJECTIVE

UNC-3.L

Determine whether a sampling distribution for a sample proportion can be described as approximately normal. **[Skill 3.C]**

ESSENTIAL KNOWLEDGE

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 $n(1-p) \geq 10$

LEARNING OBJECTIVE

UNC-3.M

Interpret probabilities and parameters for a sampling distribution for a sample proportion. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

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.