AP STATISTICS

UNIT 9

Inference for Quantitative Data: Slopes

2–5%
AP EXAM WEIGHTING

~7–8
CLASS PERIODS

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.

<table>
<tr><th>Activity</th><th>Topic</th><th>Sample Activity</th></tr>
<tr><td>1</td><td>9.2</td><td>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:
<table><tr><th>Statistic</th><th>Parameter</th></tr><tr><td>$\hat{p}$</td><td>p</td></tr><tr><td>$\bar{x}$</td><td>μ</td></tr><tr><td>s</td><td>σ</td></tr></table>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.</td></tr>
<tr><td>2</td><td>9.3</td><td>Error Analysis
Give students some raw data on the distance and cost to fly from their hometown to various major cities. For example:
<table><tr><th colspan="2">Flying from ____ to ____</th></tr><tr><th>Distance</th><th>Cost</th></tr><tr><td>512 miles</td><td>$179</td></tr><tr><td>1256 miles</td><td>$257</td></tr><tr><td>3256 miles</td><td>$387</td></tr></table>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?</td></tr>
<tr><td>3</td><td>9.5</td><td>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.</td></tr>
</table>

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 normally distributed with standard deviation, σ. The standard deviation σ can be estimated using the standard deviation of the residuals, $s = \sqrt{\frac{\sum\left(y_i - \hat{y}_i\right)^2}{n-2}}$.

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 of b equals the population mean slope: $\mu_b = \beta$. The standard deviation of the sampling distribution for β 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**

Inference for Quantitative Data: Slopes

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	ESSENTIAL KNOWLEDGE
UNC-4.AG Interpret a confidence interval for the slope of a regression model. **[Skill 4.B]**	**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.
UNC-4.AH Justify a claim based on a confidence interval for the slope of a regression model. **[Skill 4.D]**	**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.
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]**	**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

Inference for Quantitative Data: Slopes

UNIT 9

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]**

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.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.

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.