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

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.	
	3.3 Random Sampling and Data Collection	1.C Describe an appropriate method for gathering and representing data.	
	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.	
	3.6 Selecting an Experimental Design	1.C Describe an appropriate method for gathering and representing data.	
	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

Collecting Data

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

Collecting Data

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

VAR-3.C

Compare experimental designs and methods. **[Skill 1.C]**

ESSENTIAL KNOWLEDGE

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.