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)*.	
	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.	
	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.	
	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.	
	1.6 Describing the Distribution of a Quantitative Variable	2.A Describe data presented numerically or graphically.	
	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.	
	1.8 Graphical Representations of Summary Statistics	2.B Construct numerical or graphical representations of distributions. 2.A Describe data presented numerically or graphically.	
	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.	

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	ESSENTIAL KNOWLEDGE
VAR-1.B Identify variables in a set of data. **[Skill 2.A]**	**VAR-1.B.1** A variable is a characteristic that changes from one individual to another.
VAR-1.C Classify types of variables. **[Skill 2.A]**	**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

UNC-1.A

Represent categorical data using frequency or relative frequency tables. **[Skill 2.B]**

UNC-1.B

Describe categorical data represented in frequency or relative tables. **[Skill 2.A]**

ESSENTIAL KNOWLEDGE

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.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

UNC-1.G

Represent quantitative data graphically. **[Skill 2.B]**

ESSENTIAL KNOWLEDGE

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	ESSENTIAL KNOWLEDGE
VAR-2.B Determine proportions and percentiles from a normal distribution. **[Skill 3.A]**	**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.
VAR-2.C Compare measures of relative position in data sets. **[Skill 2.D]**	**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.