

2026 AP Statistics Unit 1 - Descriptive Statistics and Collecting Data

Content Area: **Math**
Course(s):
Time Period: **MP1**
Length: **45**
Status: **Published**

Unit Overview

Unit Summary	Unit Rationale
Unit 1 AP Statistics gives a comprehensive introduction to the statistical problem-solving cycle by synthesizing three foundational pillars: data production, graphical representation, and statistical modeling. Rather than presenting numbers in a vacuum, this 45-day curriculum uses the Experience First, Formalize Later (EFFL) model to ground each mathematical concept in a real-world context. By introducing data collection at the beginning of the year, students develop a key understanding that the integrity of any subsequent visual display, calculated statistic, or regression model is bound by how safely and representatively the underlying data was gathered. In the first phase of the unit, students move from managing raw categorical and quantitative variables to constructing and interpreting data displays such as stem-and-leaf plots, histograms, box plots, and mosaic plots. When evaluating quantitative data, students must move past isolated descriptions and address Shape, Outliers, Center, and Variability (CUSS) while weaving the problem's context into their descriptions. This phase also establishes methods for measuring relative standing through percentiles and z-scores, culminating in density curves and the Normal distribution. Students master the Empirical Rule (68-95-99.7) and use graphing calculators to execute standard proportion calculations and inverse boundary lookups. The second phase shifts the focus to the mechanics of sound research design, teaching students how to collect high-quality data via surveys and regulated experiments. Students learn to implement and identify sampling methods, including Simple Random Sampling (SRS), stratified sampling, cluster sampling, and systematic sampling, while weighing the operational advantages of each. Crucially, students learn to detect major sources of bias, such as undercoverage, nonresponse, and response bias, and	Starting the course with this unit gives students a firm foundation for AP Statistics right from the beginning. In most math classes, data is often just a set of numbers for students to work with. This unit takes a different approach by showing statistics as a full scientific process. Students begin with real-world questions, collect data systematically, and then use that data to describe and predict outcomes. By combining the College Board's Unit 1 (One-Variable Descriptive Statistics and Collecting Data) and Unit 5 (Two-Variable Relationships) with Math Medic's Unit 10 (Two-Variable Relationships), this unit highlights a key rule: the way data is collected determines what conclusions can be made. Students quickly learn that even the most elaborate calculations or graphs are useless if the data comes from poor sampling or biased research. Organizing the course in this way helps students understand more during each 47-minute class. Starting with data collection lets students see how statistics work in real situations before they deal with complex notation or formulas. When they later create displays such as mosaic plots, histograms, and boxplots, or calculate z-scores, they already comprehend the context and the variables involved. Introducing bivariate linear regression early also helps students apply the same principles—such as describing patterns, evaluating models, calculating residuals, and considering outliers—to both single- and two-variable data before moving on to probability and inference. This unit is also important for teaching students how to communicate clearly, which is important for the AP Statistics Exam. The College Board penalizes students for using certain language or for listing numbers without context. This unit addresses those issues by teaching clear frameworks for interpretation, such as CUSS & BS for distributions, DUFS & BS for scatterplots, and templates for regression slope, r^2, and s. Spreading this unit over

explicitly predict the direction of the bias as a systematic over- or underestimation of the true population parameter. By studying the four core principles of experimental design—Comparison, Randomization, Replication, and Control—across completely randomized, randomized block, and matched pairs layouts, students learn to determine a study's true scope of inference. They discover that random selection allows for generalization to a larger population, whereas random assignment is the unique mechanism required to establish a causal relationship.

Finally, the unit concludes by examining bivariate relationships between two quantitative variables. Students begin by constructing scatterplots to analyze patterns observed from Direction, Form, Strength, and Outliers (DUFS) before building Least-Squares Regression Lines (LSRL) in the statistical format $\hat{y} = a + bx$. To satisfy rigorous grading standards, students must master precise template scripts to interpret the real-world meaning of the slope (b), y-intercept (a), standard deviation of the residuals (s), and the coefficient of determination (r^2). Students are taught to avoid deterministic language, using probabilistic phrasing such as "on average" or "predicted value" to reflect the uncertainty inherent in statistical models. To ensure the highest level of mastery before their benchmark assessment, students learn to validate their linear models by auditing residual plots for random scatter and evaluating how high-leverage outliers or influential points change the trajectory and predictive strength of their regression lines.

45 days allows time for the Experience First, Formalize Later (EFFL) model, so students can discover concepts through collaboration before learning the formal concepts. This approach gives students a firm foundation in data collection, visual understanding, and explicit communication, helping them avoid common mistakes and preparing them to make strong statistical arguments throughout the year.

NJSLS

Standards for Mathematical Practice

- | | |
|---|--|
| 1 | Make sense of problems and persevere in solving them |
| 2 | Reason abstractly and quantitatively |

3	Construct viable arguments and critique the reasoning of others
4	Model with mathematics
5	Use appropriate tools strategically
6	Attend to precision
7	Look for and make use of structure
8	Look for and express regularity in repeated reasoning

Unit Focus

Enduring Understandings	Essential Questions
1. Well-designed data collection processes (surveys and experiments) are the foundation of any valid statistical claim; the methodology used to produce data dictates whether conclusions can be generalized to a population or can establish a causal relationship. 2. Graphical displays and numerical summaries reveal underlying patterns, shapes, central trends, and departures from patterns in a dataset that the naked eye or a raw list of numbers cannot capture. 3. Mathematical structures, such as the Normal distribution curve or the Least-Squares Regression Line, serve as probabilistic models to describe reality, simplify complex distributions, and make calculated predictions under uncertainty.	1. How do we design studies and gather data in a way that allows us to confidently make claims about a broader population or prove cause-and-effect? 2. What story does this data display tell us, and how do measures of center and variability help us precisely describe a distribution's behavior? 3. How can we use mathematical models to predict future outcomes or evaluate an individual's relative standing within a dataset, and what are the limitations of those models?

Instructional Focus

Learning Targets
<i>Part 1: Exploring One-Variable Data:</i> • I can identify the population, sample, individual cases, and the exact investigative question being asked in a statistical study. • I can classify a variable as either categorical or quantitative (discrete vs. continuous). • I can represent categorical data using frequency tables, bar charts, and mosaic plots, and use them to calculate marginal and conditional relative frequencies. • I can construct and interpret quantitative data displays, including stem-and-leaf plots, histograms, and box plots.

- I can write a cohesive, contextual description of a quantitative distribution by addressing its Shape, Outliers, Center, and Variability (CUSS & BS) using comparative language when necessary.
- I can identify outliers mathematically using the $1.5 \times \text{IQR}$ rule and the mean-plus-standard-deviation boundary.
- I can calculate and interpret a data point's percentile and z-score in context, and analyze the effects of linear transformations (addition, subtraction, multiplication, division) on summary statistics.
- I can model a data distribution using density curves and the Empirical Rule (68-85-99.7), and calculate areas/proportions and boundaries using a graphing calculator (normalcdf and invNorm).

Part 2: Collecting Data

- I can distinguish between an observational study and a controlled experiment, and identify why only an experiment can establish a cause-and-effect relationship.
- I can identify and execute various sampling methodologies, including Simple Random Sampling (SRS), Stratified Sampling, Cluster Sampling, and Systematic Sampling, and describe the operational advantages of each.
- I can identify potential sources of sampling bias (undercoverage, nonresponse, response bias, and convenience sampling) and explain how each bias can alter or systematically misrepresent the true population parameter.
- I can identify the core components of an experimental design: experimental units, explanatory variables (factors), treatments, and response variables.
- I can apply the four principles of experimental design: Comparison, Randomization, Replication, and Control.
- I can design and explain randomized experimental structures, including Completely Randomized Designs, Randomized Block Designs, and Matched Pairs Designs.
- I can evaluate a study's design to state its proper Scope of Inference (determining whether results can be generalized to a broader population, and whether a causal relationship can be claimed).

Part 3: Exploring Two-Variable Data

- I can construct a scatterplot to analyze the relationship between two quantitative variables and describe its Direction, Form, Strength, and Outliers (DUFS) in context.
- I can calculate and interpret the correlation coefficient (r) and explain its mathematical properties (e.g., its vulnerability to outliers and its unitless nature).
- I can construct a Least-Squares Regression Line (LSRL) model, use it to make predictions, and explain the dangers of extrapolation.
- I can calculate and interpret the slope and y-intercept of an LSRL using strict, contextual AP Statistics template phrasing.
- I can calculate and interpret a residual, construct a residual plot, and evaluate whether a linear model is appropriate based on the plot's randomness.
- I can interpret standard computer output for linear regression to find the regression equation, the standard deviation of the residuals (s), and the coefficient of determination (r^2).
- I can explain what s and r^2 mean in the context of the model's predictive accuracy.
- I can identify influential points and high-leverage outliers on a scatterplot and analyze their specific effects on the slope, y-intercept, correlation coefficient r , and standard deviation of the residuals s .

Prerequisite Skills

Part 1: Exploring One-Variable Data

- Students must seamlessly convert between fractions, decimals, and percentages.
- Students need to confidently substitute values into multi-step formulas.
- Familiarity with reading basic number lines and early-grade displays like standard bar charts, dot plots, and line graphs.
- Prior knowledge of what a "mean" and "median" represent conceptually, even if they haven't yet studied how outliers affect them.
- Understanding intervals on a number line is critical to understanding standard deviation and the boundaries of the normal curve.

Part 2: Collecting Data

- Students must be able to read a dense 5-sentence paragraph and extract the operational core: Who is being studied? What is being measured? What was the goal?
- A baseline intuitive understanding that just because two things happen together does not mean one caused the other.
- An intuitive grasp of chance and fair selection.
- The ability to follow "if-then" scenarios to identify flaws in a process.

Part 3: Exploring Two-Variable Data

- Unhesitating ability to plot (x, y) ordered pairs and read coordinates from a standard four-quadrant scatterplot.
- Students must know what a line is, how to identify the slope (m) and the y-intercept (b), and how to use an x-value to calculate a predicted y-value.
- Understanding that the independent variable (x) is the input/predictor and the dependent variable (y) is the output/outcome.
- Remembering that slope represents a rate.

Common Misconceptions

Part 1: Exploring One-Variable Data

- When looking at a histogram or bar chart, students often confuse the value on the x-axis (the variable itself) with the value on the y-axis (the count/frequency).
 - If a histogram shows a high bar at a score of 10, students might say "the data is highly variable" because the bar is tall, rather than looking at how spread out the values are along the x-axis.
- Students intuitively believe that a wider section of a boxplot contains more data points.
 - Every single section of a boxplot (each whisker and each half of the box) contains exactly 25% of the data. A wider section simply means that 25% of the data is more spread out (variable), not that there are more individuals in it.
- Students use the words "Symmetric" and "Normal" interchangeably.
 - All Normal distributions are symmetric, but not all symmetric distributions are Normal. A uniform distribution (flat line) or a bimodal distribution can be perfectly symmetric but looks nothing like a bell curve. College Board graders will penalize students who call a distribution "Normal" without verifying the bell shape or empirical rule.
- Students write isolated, abstract descriptions like, "The shape is skewed right, there is an outlier at 50, the median is 12, and the IQR is 4."
 - This will score a "Partially Correct" on the AP Exam. They must weave the problem's context

into the sentence (e.g., "The distribution of high school student shoe sizes is skewed right...").

Part 2: Collecting Data

- Students mix up these two methods because both involve splitting a population into groups.
 - In Stratified, you sample some individuals from all groups (Homogeneous groups: "Some from all").

In Cluster, you sample all individuals from some groups (Heterogeneous groups: "All from some").
- Students think bias means the researcher had bad intentions or made a sloppy mistake. Alternatively, they define bias as just "bad sampling."
 - On the AP Exam, defining bias requires explaining the direction of the error. Students must state whether the sampling method will systematically overestimate or underestimate the true population parameter. Simply saying "it's biased because it's a convenience sample" earns no credit.
- Students think any use of randomness accomplishes both goals.
 - Random Selection (surveys) allows you to generalize your findings to the population.

Random Assignment (experiments) allows you to establish causation between treatments.
- When asked to describe a random sampling process, students will write, "I will pick 20 students randomly from the cafeteria."
 - Human eyes naturally drift toward certain people, which introduces bias. Students must explicitly state a mechanical randomization technique (e.g., "Assign every student a number from 1 to N, use a random number generator to pick 20 unique numbers, and sample those students").

Part 3: Exploring Two-Variable Data

- When interpreting a slope or making a prediction, students write absolute statements like, "For every 1-inch increase in height, a person's weight will increase by 4 pounds."
 - Regression lines are probabilistic models, not algebraic lines. Students must use "wobble" words: "On average," "the predicted weight increases by," or "we expect." Omitting these words is an automatic point deduction.
- Students assume that if the correlation coefficient $r = 0.95$, the relationship is guaranteed to be linear.
 - A strongly curved relationship (such as a distinct quadratic curve) can still yield a very high linear correlation. The only way to check if a linear model is appropriate is to look at the residual plot and ensure there is no leftover curved pattern.
- Students frequently mix up the interpretations of r (the strength and direction of a linear relationship) and r^2 (the percentage of variation in y explained by the linear model with x).
 - Force them to stick to Math Medic's rigid script templates during Phase 3 so they don't accidentally blur these two concepts together.

Spiraling For Mastery

Current Unit Content/Skills	Spiral Focus	Activity
<i>Part 1: Exploring One-Variable Statistics</i> - Constructing/interpreting graphs - Histograms, Boxplots, Mosaic plots. - Describing Quantitative Distributions (CUSS & BS) - Calculating percentiles, z-scores, and Normal curve proportions.	<i>Part 1: Exploring One-Variable Statistics</i> - Distinguishing between population vs. sample - Properly identifying variable types within context. - categorical vs. quantitative	<i>Part 1: Exploring One-Variable Statistics</i> "What's the Variable?" Snapshot - Provide a brief 3-sentence summary of a study at the start of class. Before analyzing the graphs, give students 2 minutes to identify the population and sample and classify every variable involved. This prevents them from confusing categorical counts with quantitative values during analysis.
<i>Part 2: Collecting Data</i> - Identifying sampling methods - SRS, Stratified Cluster, Systematic - Explaining sources of bias - Undercoverage, nonresponse, response bias - Designing randomized experiments vs. observational studies.	<i>Part 2: Collecting Data</i> - Using CUSS & BS to justify study design. - Reading existing data distributions to identify confounding variables.	<i>Part 2: Collecting Data</i> "The Treatment Comparison" Drill - When teaching experimental design, give students summary data (like side-by-side boxplots) from a flawed observational study. Have them write a quick 3-minute paragraph comparing the distributions using CUSS & BS, and use that

		descriptive analysis to explain why a randomized experiment is necessary to prove causation.
<i>Part 3: Exploring Two-Variable Data</i> • Interpreting scatterplots ◦ DUFS & BS • Calculating and Interpreting LSRL slope, y-intercept, r, r^2, and s. • Analyzing residuals and computer output.	<i>Part 3: Exploring Two-Variable Data</i> • Assessing how bad sampling impacts a regression model. • Analyzing the distribution of residuals using 1-variable tools.	<i>Part 3: Exploring Two-Variable Data</i> • Residual DUFS & BS and Biased Slopes ◦ Give students a histogram or box plot of the <i>residuals</i> from a linear model. Have them describe its shape and check for outliers using the $1.5 \times \text{IQR}$ rule. ◦ Present a scenario where a convenience sample was used to gather two-variable data. Ask students to predict whether the resulting slope will over- or underestimate the true population relationship.

Assessment

Formative Assessment	Summative Assessment
• Homework • Lesson Checks • Quizzes • Exit Tickets	• Topic Tests • Unit 1 Benchmark (Link-It)

• Lesson Reflections • Performance Tasks • AP Classroom Progress Checks	
---	--

Resources

Key Resources	Supplemental Resources
AP Classroom AP Statistics CED Pacing Guide	iXL Delta Math Desmos Khan Academy Math Medic Skew the Script Teacher Made worksheets Textbook - Starnes, D., & Tabor, J. (2024). <i>The Practice of Statistics for the AP Classroom</i> (7th ed.). Beford, Freeman & Worth.

Career Readiness, Life Literacies, and Key Skills

9.4.12.CI.1	Demonstrate the ability to reflect, analyze, and use creative skills and ideas (e.g., 1.1.12prof.CR3a).
9.2.12.CAP.4	Evaluate different careers and develop various plans (e.g., costs of public, private, training schools) and timetables for achieving them, including educational/training requirements, costs, loans, and debt repayment.
9.4.12.CT.2	Explain the potential benefits of collaborating to enhance critical thinking and problem solving (e.g., 1.3E.12profCR3.a).
9.4.12.IML.3	Analyze data using tools and models to make valid and reliable claims, or to determine optimal design solutions (e.g., S-ID.B.6a., 8.1.12.DA.5, 7.1.IH.IPRET.8).
9.4.12.IML.7	Develop an argument to support a claim regarding a current workplace or societal/ethical issue such as climate change (e.g., NJSLSA.W1, 7.1.AL.PRSNT.4).
9.4.12.TL.4	Collaborate in online learning communities or social networks or virtual worlds to analyze and propose a resolution to a real-world problem (e.g., 7.1.AL.IPERS.6).

Interdisciplinary Connections

RI.CR.11–12.1	Accurately cite a range of thorough textual evidence and make relevant connections to strongly support a comprehensive analysis of multiple aspects of what an informational text says explicitly and inferentially, as well as interpretations of the text.
8.1.12.DA.5	Create data visualizations from large data sets to summarize, communicate, and support different interpretations of real-world phenomena.
8.1.12.DA.6	Create and refine computational models to better represent the relationships among different elements of data collected from a phenomenon or process.
W.AW.11–12.1.B	Develop claim(s) and counterclaims avoiding common logical fallacies and using sound reasoning and thoroughly, supplying the most relevant evidence for each while pointing out the strengths and limitations of both in a manner that anticipates the audience’s knowledge level, concerns, values, and possible biases.
SL.PE.11–12.1.A	Come to discussions prepared, having read and researched material under study; explicitly draw on that preparation by referring to evidence from texts and other research on the topic or issue to stimulate a thoughtful, well-reasoned exchange of ideas. Global economic activities involve decisions based on national interests, the exchange of different units of exchange, decisions of public and private institutions, and the ability to distribute goods and services safely.
SCI.HS-LS2-6	Evaluate the claims, evidence, and reasoning that the complex interactions in ecosystems maintain relatively consistent numbers and types of organisms in stable conditions, but changing conditions may result in a new ecosystem.