

B R I G H T S C H O L A R S

AP+ EXAM PREP SERIES

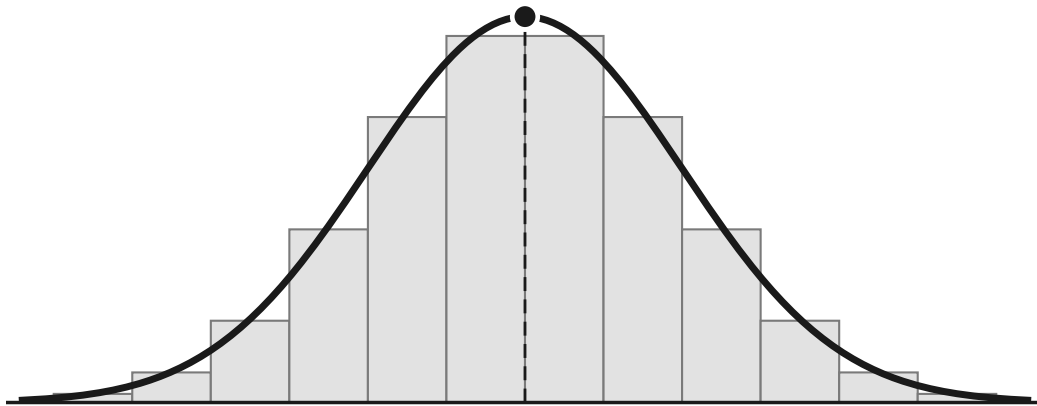
STATISTICS

REVIEW WORKBOOK

2027 EXAM

Data Analysis • Probability • Inference • FRQ Strategy

Interpret Data. Justify Conclusions. Think Statistically.



Aligned to the revised 2026-27 course and May 2027 digital exam • Diagnostic test • 5 units • 2 full-length exams

THIS BOOK BELONGS TO

Name _____

School / Teacher _____

AP exam date _____

Contents

GETTING STARTED

The 2027 AP Statistics Exam	4
How to Use This Workbook	5
8-Week Study Plan and Progress Tracker	6
Diagnostic Test	8
Diagnostic Answer Key and Score Tracker	16
Statistics Skills Overview	18

UNITS

1 Exploring One-Variable Data and Collecting Data	20
2 Probability, Random Variables, and Probability Distributions	42
3 Inference for Categorical Data: Proportions	64
4 Inference for Quantitative Data: Means	88
5 Regression Analysis	108

SKILLS LABS AND FRQ PRACTICE

Statistical Writing Lab	122
FRQ Builder	126
Calculator + Formula Practice Lab	137
Common Mistake Fixer	141
Error Log	143

PRACTICE EXAMS

Full-Length Practice Exam 1	144
Section II: Free Response	154
Full-Length Practice Exam 2	162
Section II: Free Response	172

ANSWERS AND FINAL REVIEW

Answer Key	180
Final Progress Tracker	192
Before Exam Day Checklist	193
Final Notes and Last-Minute Review	194

i KNOW IT

Short concept notes placed right next to the problems that use them.

2 FRQ EDGE

Full free-response practice with a labeled answer space for each part.

! FIX IT

A weak response next to a stronger one, so you learn the common point losses.

UNIT

1

Exploring One-Variable Data and Collecting Data

AP EXAM WEIGHT: 20–30% OF THE MULTIPLE-CHOICE SECTION

BIG IDEA

Learn to identify variables, summarize one-variable data, describe distributions, compare data displays, and evaluate how data are collected.

✓ BY THE END OF THIS UNIT

- ☐ Classify variables.
- ☐ Build and read frequency tables.
- ☐ Construct and interpret dotplots, histograms, and boxplots.
- ☐ Calculate and interpret center and spread.
- ☐ Evaluate sampling methods and study designs.

i IN THIS UNIT

- 1.1** Variables, Distributions, and Categorical Data
 - 1.2** Displaying and Summarizing Quantitative Data
 - 1.3** Investigative Questions and Sampling
 - 1.4** Experiments and Observational Studies
- Unit Review: Multiple Choice, FRQ Edge, Fix It, Exit Ticket, READY+ Check

📋 KEY TERMS

statistical question · observational unit · categorical / quantitative variable · discrete vs. continuous · relative frequency · z-score · outlier · random sampling · random assignment · bias · confounding · blocking · matched pairs

! COMMON AP TRAPS

Confusing the variable with the observational unit; calling random assignment “random sampling”; claiming an observational study proves causation; describing a graph vaguely instead of in context.

✓ READY+ GOAL

By the end of this unit, you should be able to fully describe and compare distributions in context, and correctly evaluate whether a study can support a cause-and-effect claim.

UNIT 1 | LESSON 1.1

Variables, Distributions, and Categorical Data

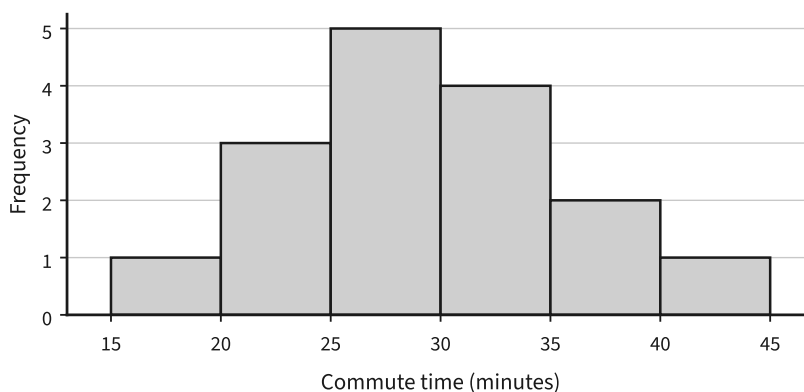
i QUICK REVIEW: VARIABLES AND DATA

An observational unit is the individual person, object, or event being studied. A variable is a characteristic that is recorded for each observational unit. Categorical variables place individuals into groups (e.g., eye color); quantitative variables take numerical values that can be measured (e.g., height). Quantitative variables are further split into discrete (countable, e.g., number of siblings) and continuous (measurable on a continuum, e.g., time).

i KNOW IT

The observational unit answers “who or what is being measured?” The variable answers “what is being recorded about them?” Mixing these up is one of the most common Unit 1 errors.

Describing a Distribution



A complete description of a quantitative distribution always includes shape (symmetric, skewed left, skewed right), center (mean or median), spread (standard deviation, range, or IQR), and any unusual features (outliers, gaps, clusters), stated in the context of the problem.

i DATA CHECK: CATEGORICAL DATA DISPLAYS

A survey asked 120 students which transportation method they use most often to get to school. Results are shown below.

Method	Frequency	Relative Frequency
Bus	54	45%
Car	36	30%
Walk	18	15%
Bike	12	10%

UNIT 1 | LESSON 1.2

Displaying and Summarizing Quantitative Data



STEM-AND-LEAF PLOTS

A stem-and-leaf plot displays quantitative data while preserving individual values. The stem contains the leading digit(s), and each leaf contains the final digit. Always include a key so the reader knows the place value.

Back-to-back stem-and-leaf

The quiz scores below compare two classes. Read Class A leaves from the stem outward to the left and Class B leaves from the stem outward to the right.

Class A leaves	Stem	Class B leaves
9 7 5 2	6	1 4 8
8 8 6 4 1	7	0 2 5 7 9
9 6 3 2	8	1 4 6 8
4	9	0 3

Key: 7 | 5 represents a score of 75.

Your turn

1

Which class has the higher typical score?

2

Which class has greater spread?

3

Identify one feature visible in the stem-and-leaf display that would be harder to see from summary statistics alone.



DATA CHECK: DOTPLOTS, HISTOGRAMS, AND BOXPLOTS

The dataset below shows the number of minutes 15 students spent on homework last night: 20, 25, 30, 30, 35, 40, 40, 45, 45, 50, 55, 60, 65, 70, 95.

Segmented Bar Charts

Segmented bar charts display conditional distributions and help compare categorical variables.

If two variables are associated, the segment percentages will differ between groups. Similar segment percentages suggest little or no association.

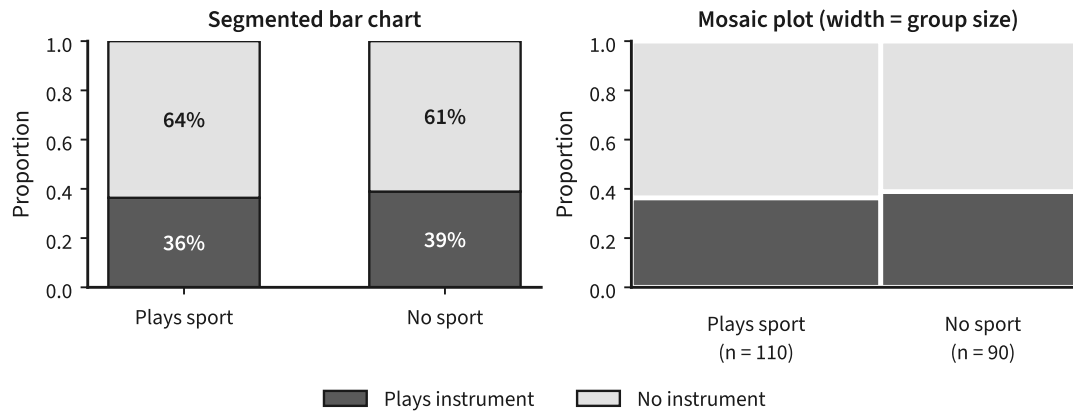
Mosaic Plots

A mosaic plot displays the relationship between two categorical variables using rectangles whose widths and heights represent proportions.

Mosaic plots help identify possible associations by comparing the sizes of different sections.



VISUAL COMPARISON: SEGMENTED BAR CHART AND MOSAIC PLOT



UNIT 1 | REVIEW

Fix It and Exit Ticket

! FIX IT: COMMON MISTAKES IN UNIT 1

WEAK RESPONSE	STRONGER RESPONSE
“The variable is the students.”	“The observational units are the students; the variable is the number of hours of homework per week.”
“The graph is skewed.”	“The distribution is skewed right, with most values between 2 and 6 hours and a few students reporting 12+ hours.”
“They randomly sampled the treatment groups.”	“Subjects were randomly assigned to treatment groups” (random assignment, not random sampling).
“This proves breakfast causes higher grades.”	“This observational study shows an association; it cannot establish causation because of possible confounding variables.”

How will you catch yourself making one of these mistakes next time?

⦿ MIXED PRACTICE

A quality-control inspector records the diameter (in mm) of 10 randomly selected bolts from a production line: 5.01, 5.02, 4.98, 5.00, 5.03, 4.99, 5.01, 5.50, 5.00, 4.97.

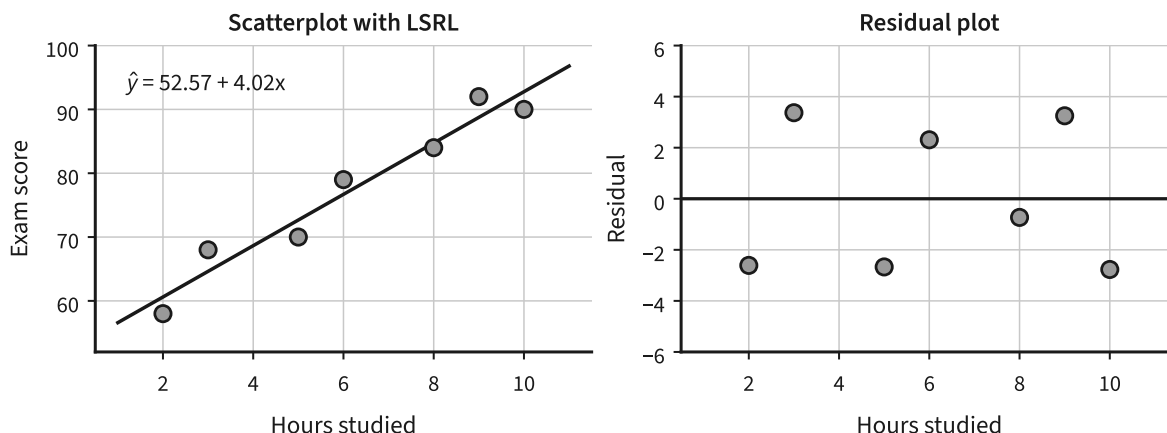
⦿ MIXED PRACTICE

Calculate the mean and median of this dataset. Identify any outliers using the $1.5 \times \text{IQR}$ rule, and explain which measure of center better represents a “typical” bolt diameter.

FRQ EDGE | UNIT 5

Regression

A study examines the relationship between hours spent studying and exam score for a random sample of 7 students. Hours studied (x): 2, 3, 5, 6, 8, 9, 10. Exam score (y): 58, 68, 70, 79, 84, 92, 90. A calculator regression output for these data gives: $\hat{y} = 52.57 + 4.02x$, $r \approx 0.973$, $r^2 \approx 0.947$. A scatterplot and residual plot are shown below.



a Identify the explanatory and response variables.

b Interpret the slope in context.

[illegible][illegible]

Full-Length Practice Exam 1

42 Multiple-Choice Questions + 4 Free-Response Questions

This full-length practice exam draws from all five units. Section I contains 42 multiple-choice questions and should be completed in 90 minutes. Section II contains 4 free-response questions and should be completed in 90 minutes. Questions 36–38 form a probability set, and Questions 40–42 form a regression set, mirroring the revised exam structure.

SECTION	QUESTIONS	TIME	SCORE WEIGHT
Section I: Multiple Choice	42	90 minutes	50%
Section II: Free Response	4	90 minutes	50%

BEFORE YOU START

Set a timer, work in one sitting if you can, and mark your answers on the answer sheet as you go. Use a graphing calculator throughout, just as you will on exam day.

Section I Answer Sheet

Fill in one bubble per question.

1 ☐ A ☐ B ☐ C ☐ D

4 ☐ A ☐ B ☐ C ☐ D

7 ☐ A ☐ B ☐ C ☐ D

10 ☐ A ☐ B ☐ C ☐ D

13 ☐ A ☐ B ☐ C ☐ D

16 ☐ A ☐ B ☐ C ☐ D

19 ☐ A ☐ B ☐ C ☐ D

22 ☐ A ☐ B ☐ C ☐ D

25 ☐ A ☐ B ☐ C ☐ D

28 ☐ A ☐ B ☐ C ☐ D

31 ☐ A ☐ B ☐ C ☐ D

34 ☐ A ☐ B ☐ C ☐ D

37 ☐ A ☐ B ☐ C ☐ D

40 ☐ A ☐ B ☐ C ☐ D

2 ☐ A ☐ B ☐ C ☐ D

5 ☐ A ☐ B ☐ C ☐ D

8 ☐ A ☐ B ☐ C ☐ D

11 ☐ A ☐ B ☐ C ☐ D

14 ☐ A ☐ B ☐ C ☐ D

17 ☐ A ☐ B ☐ C ☐ D

20 ☐ A ☐ B ☐ C ☐ D

23 ☐ A ☐ B ☐ C ☐ D

26 ☐ A ☐ B ☐ C ☐ D

29 ☐ A ☐ B ☐ C ☐ D

32 ☐ A ☐ B ☐ C ☐ D

35 ☐ A ☐ B ☐ C ☐ D

38 ☐ A ☐ B ☐ C ☐ D

41 ☐ A ☐ B ☐ C ☐ D

3 ☐ A ☐ B ☐ C ☐ D

6 ☐ A ☐ B ☐ C ☐ D

9 ☐ A ☐ B ☐ C ☐ D

12 ☐ A ☐ B ☐ C ☐ D

15 ☐ A ☐ B ☐ C ☐ D

18 ☐ A ☐ B ☐ C ☐ D

21 ☐ A ☐ B ☐ C ☐ D

24 ☐ A ☐ B ☐ C ☐ D

27 ☐ A ☐ B ☐ C ☐ D

30 ☐ A ☐ B ☐ C ☐ D

33 ☐ A ☐ B ☐ C ☐ D

36 ☐ A ☐ B ☐ C ☐ D

39 ☐ A ☐ B ☐ C ☐ D

42 ☐ A ☐ B ☐ C ☐ D

12 C Binomial probability (Unit 2)

$$P(X = 2) = C(5, 2)(0.3)^2(0.7)^3 = 10(0.09)(0.343) = 0.3087.$$

13 D Normal distribution (Unit 2)

69 inches is $(69 - 64)/2.5 = 2$ standard deviations above the mean. About 95% of values are within 2 SD, leaving about 2.5% in the upper tail (2.3% with a calculator).

14 B Expected value (Unit 2)

$$E(\text{winnings}) = 5(0.2) + 1(0.3) + 0(0.5) = 1.00 + 0.30 = \$1.30.$$

15 A Independent events (Unit 2)

$$\text{For independent spins, } P(\text{red and red}) = 0.2 \times 0.2 = 0.04.$$

16 C Binomial setting (Unit 2)

Each spin is a trial with a success (red) or failure (not red), $p = 0.2$ is constant, trials are independent, and $n = 10$ is fixed, so X is binomial.

Unit 2 • FRQ: Sport/Music Two-Way Table p. 56

Acceptable answer elements: correct conditional probability calculation; correct comparison of $P(A) \cdot P(B)$ to $P(A \text{ and } B)$ for independence; a clear, contextual distinction between independence and mutual exclusivity.

SAMPLE STRONG RESPONSE

" $P(\text{sport}) = 144/240 = 0.60$. $P(\text{music} | \text{sport}) = 48/144 \approx 0.333$. Since $P(\text{music}) = 84/240 = 0.35 \neq 0.333$, sport and music participation are not independent, because knowing a member plays a sport changes the probability that the member participates in music. Independent events do not change each other's probabilities, while mutually exclusive events cannot occur together at all. Sport and music occur together for 48 members here, so they are not mutually exclusive either."

COMMON WEAK RESPONSE

"They are independent because some students do both."
This confuses the ability to co-occur with statistical independence.

Scoring notes: Full credit requires the correct conditional probability, a numeric independence check, and a conceptually correct distinction (not just a definition recited without connecting to the data).

Unit 2 • FRQ: Vending Machine Distribution p. 58

Acceptable answer elements: mean interpreted as a long-run average, not a guarantee for any single day; a probability-based judgment (using mean and standard deviation) about selling out, rather than a bare yes/no; a correct normal probability for fewer than 30 snacks, with the z-score or calculator inputs shown.

SAMPLE STRONG RESPONSE

"The mean of $X = 40$ means that, over many days, the vending machine sells an average of about 40 snacks per day, though the actual number varies day to day. Restocking 50 snacks is $50 - 40 = 10$ above the mean, or about $(50 - 40)/6 \approx 1.67$ standard deviations above the mean. Using a normal model, $P(X \geq 50) \approx 0.048$, so there is only about a 5% chance of selling out. Selling out is possible but not likely on a typical day. For fewer than 30 snacks, $z = (30 - 40)/6 \approx -1.67$, so $P(X < 30) \approx 0.048$. About 4.8% of days would have fewer than 30 snacks sold."

COMMON WEAK RESPONSE

"50 is close to 40, so it will probably sell out." This response compares the numbers without using the standard deviation and gives no probability.

Scoring notes: Full credit requires a context-based interpretation of the mean, a probability or z-score based judgment about selling out (not just a comparison of raw numbers), and a correct normal probability for fewer than 30 snacks.

Unit 2 • Extended FRQ p. 60

- a There is a fixed number of trials (12 bulbs), each bulb is defective or not, the probability of a defect is the same for each bulb (0.04), and defects are independent.
- b $\mu = np = 12(0.04) = 0.48$ defective bulbs; $\sigma = \sqrt{(12 \times 0.04 \times 0.96)} \approx 0.68$.
- c $P(X = 0) = (0.96)^{12} \approx 0.613$.
- d $P(X = 1) = 12(0.04)(0.96)^{11} \approx 0.306$, so $P(X \geq 2) = 1 - 0.613 - 0.306 \approx 0.081$.
- e Let L = the lifetime of a working bulb. $z = (1,500 - 1,200)/150 = 2$, so $P(L > 1,500) \approx 0.023$ (about 2.3% of bulbs).

Scoring: 2 points per part. Calculations must show the formula or calculator command with its inputs, not just a final number.

UNIT 3**Unit 3 • Lessons and practice activities** p. 65

Sampling proportions for $n=50$ are approximately normal, centered at 0.40, with $SD \sqrt{(0.4 \cdot 0.6/50)} \approx 0.069$; at $n=200$, $SD \approx 0.035$. Voter interval: $\hat{p} = 84/150 = 0.56$; 90% CI $\approx (0.493, 0.627)$. We are 90% confident that the true proportion of voters in this electorate who support the measure is between 0.493 and 0.627. Customer test: $H_0: p = 0.85$ vs $H_a: p \neq 0.85$; $z \approx -1.534$, $p \approx 0.125$; fail to reject H_0 . A p-value of 0.125 means results this far from 0.85 or farther would occur about 12.5% of the time if $p = 0.85$.

Type I: conclude satisfaction differs from 85% when it does not. Type II: fail to detect a real difference. Review-program test: $H_0: p_A - p_B = 0$ vs. $H_a: p_A - p_B \neq 0$, where p_A and p_B are the true pass rates for students like these under each program. Conditions: students were randomly assigned to programs; pooled $\hat{p} = 75/155 \approx 0.484$, and the expected counts (about 38.7, 41.3, 36.3, 38.7) are all at least 10. $z \approx 2.023$, $p \approx 0.043$. Because $0.043 < 0.05$, reject H_0 : there is convincing evidence that the true pass rates differ, with Program A appearing higher. Drug-placebo: $\hat{p}_1 = 0.65$, $\hat{p}_2 = 0.48$; 95% CI for $p_1 - p_2 \approx (0.016, 0.324)$, entirely above 0, so there is evidence the improvement rates differ. Teaching methods: $z \approx 2.31$, $p \approx 0.021$. Chi-square example: expected counts 40, 40, 35, 35; $\chi^2 \approx 10.71$, $df = 1$, $p \approx 0.001$. School-start interval: $\hat{p} = 0.60$; 95% CI $\approx (0.535, 0.665)$. Cereal test: $H_0: p = 0.20$ vs $H_a: p \neq 0.20$; $z \approx 1.85$, $p \approx 0.064$; fail to reject H_0 at 0.05. Large-counts checks use p_0 for a test: 18 expected prizes and 72 expected non-prizes.

Estimators: Estimator A is unbiased because its sampling distribution is centered at the parameter, 0.40. Estimator B is less variable (SD 0.03 vs. 0.06). Less variable is not automatically better: B is centered at 0.37, so it consistently underestimates p even though its estimates are tightly grouped.

Chi-square practice: H_0 : grade and preferred study time are independent for students at this school; H_a : they are