AP Statistics Exam Prep

Complete Review Bundle

Grades 11 - 12

- - -

>> 9 Unit Concept Summaries <<

>> 150 Multiple-Choice Practice Questions + Answer Key <<

>> 10 Free-Response Questions + Scoring Guides <<

>> Formula Sheet & Graphing Calculator Tips <<

>> Exam Pacing Strategy & Score Interpretation Chart <<

UM Printables

Premium Teaching Resources

Table of Contents

PART ONE

Unit Concept Summaries

Unit 1: Exploring One-Variable Data

1.1 Types of Variables

- **Categorical (qualitative)**: names or labels with no numerical meaning (e.g., eye color, zip code).
- **Quantitative**: numerical values that represent measurable quantities (e.g., height, temperature).
- Quantitative variables can be **discrete** (countable) or **continuous** (measurable on an interval).

1.2 Displaying Categorical Data

- **Bar chart**: one bar per category; bars do not touch; y-axis shows frequency or relative frequency.
- **Pie chart**: shows parts of a whole; only appropriate when all categories represent 100% of the data.

1.3 Displaying Quantitative Data

- **Dot plot**: each data value is a dot above a number line; best for small datasets.
- **Stem plot**: preserves actual data values; shows shape; useful for small to moderate datasets.
- **Histogram**: bars represent frequency of values in intervals (bins); bars touch; shows shape for large datasets.
- **Box plot**: displays five-number summary; good for comparing distributions; highlights outliers.

> **TIP:** *When describing a distribution from a graph, always use SOCS: Shape, Outliers, Center, Spread - in that order.*

1.4 Describing Distributions - SOCS

- **Shape**: Symmetric, Skewed right (positive), Skewed left (negative), Uniform, or Bimodal.
- **Outliers**: Data points that fall far outside the overall pattern. Use the 1.5 * IQR rule.
- **Center**: Typical value: mean (x-bar) or median. Mean is pulled toward the tail in skewed distributions.
- **Spread**: How variable the data are: range, IQR, standard deviation.

1.5 Measures of Center

Formulas:

Sample Mean: $\bar{x} = (\textbf{Sum}\ x_i) / n$

Population Mean: $\mu = (\textbf{Sum}\ x_i) / N$

- The **mean** is the arithmetic average. It is **not resistant** - outliers pull the mean toward them.
- The **median** is the middle value when data are ordered. It **is resistant** to outliers.

1.6 Measures of Spread

Formulas:	Range = Max - Min IQR = Q_3 - Q_1 Sample Variance: s^2 = **Sum**(x_i - x-bar)^2 / (n - 1) Sample Std. Dev.: s = sqrt[**Sum**(x_i - x-bar)^2 / (n - 1)]

- **IQR** (Interquartile Range) is the range of the middle 50% of the data. It is **resistant** to outliers.
- **Standard deviation** measures typical distance of data values from the mean. It is **not resistant**.

1.7 The 1.5 x IQR Rule for Outliers

Outlier Boundaries:	Low outlier: below Q_1 - 1.5(IQR) High outlier: above Q_3 + 1.5(IQR)

1.8 Five-Number Summary & Box Plot

- Five-number summary: **Min, Q_1, Median, Q_3, Max**.

1.9 Percentiles & Z-Scores

Key Formulas:	Z-score: z = (x - *mu*) / *sigma*

- A **z-score** tells how many standard deviations a value is from the mean.

1.10 The Normal Distribution

Key Rule:	**Empirical Rule (68-95-99.7 Rule):** ~ 68% of data within *mu* +/- 1 *sigma* ~ 95% of data within *mu* +/- 2 *sigma* ~ 99.7% of data within *mu* +/- 3 *sigma*

1.11 Transforming Data

- Linear transformation: x_{new} = a + bx -> mu_{new} = a + b*mu*, $sigma_{new}$ = |b|*sigma*.

Key Terms: Categorical | **Quantitative** | **Mean** | **Median** | **Standard Deviation** | **IQR** | **Outlier** | **Z-score** | **Normal Distribution**

Unit 2: Exploring Two-Variable Data

2.1 Scatterplots

- A **scatterplot** displays the relationship between two quantitative variables.
- Explanatory (independent) variable on the x-axis; Response (dependent) variable on the y-axis.
- Describe the relationship using: **Direction** (positive/negative), **Form** (linear/curved), **Strength** (strong/moderate/weak), and **Outliers**.

2.2 Correlation (r)

Formula: $r = (1/(n-1))\ \textbf{Sum}\ [(x_i\text{-x-bar})/s_x][(y_i\text{-y-bar})/s_y]$

$-1 <= r <= 1$

- **r** measures the strength and direction of the **linear** relationship between two quantitative variables.
- r > 0 -> positive association; r < 0 -> negative association.
- |r| close to 1 -> strong linear relationship; |r| close to 0 -> weak linear relationship.
- r does **NOT** distinguish between explanatory and response variables (r is symmetric).
- r is **not resistant** to outliers.
- **Correlation does NOT imply causation!**

TIP: *r measures only linear association. A strong curved relationship can have r approx 0. Always look at the scatterplot!*

2.3 Least-Squares Regression Line (LSRL)

Formulas:

Predicted value: $\hat{y} = b_0 + b_1 x$

Slope: $b_1 = r(s_y / s_x)$

Intercept: $b_0 = \bar{y} - b_1 \bar{x}$

- The LSRL minimizes the sum of the squared residuals (SSE = **Sum** e_i^2).
- **Slope (b_1)**: for each unit increase in x, y-hat changes by b_1 units on average.
- **Intercept (b_0)**: predicted value of y when x = 0. Only meaningful if x = 0 is within the data range.
- The LSRL always passes through the point (x-bar, y-bar).

2.4 Residuals

Definition: Residual = y - y-hat = observed - predicted

- A **positive residual** means the model underestimates the actual value.
- A **negative residual** means the model overestimates the actual value.
- **Residual plot**: residuals vs. x-value. A good linear model shows random scatter around 0.
- A pattern in the residual plot suggests the linear model is not appropriate.

2.5 Coefficient of Determination (r^2)

- **r^2** = the proportion of variation in y that is explained by the LSRL of y on x.
- If r = 0.8, then r^2 = 0.64, meaning 64% of variation in y is explained by the linear relationship with x.
- 1 - r^2 = proportion of variation in y that is NOT explained by the model.

2.6 Influential Points

- An **influential point** is a point that, if removed, significantly changes the position of the LSRL.
- High-leverage points have extreme x-values. They may or may not be influential.
- Not all outliers are influential, and not all influential points are outliers.

2.7 Interpolation vs. Extrapolation

- **Interpolation**: predicting within the range of observed x-values - generally reliable.
- **Extrapolation**: predicting outside the range of observed x-values - risky and often unreliable.

2.8 Categorical Data: Two-Way Tables

- A **two-way table** displays counts for two categorical variables.
- **Marginal distribution**: distribution of one variable alone (row or column totals).
- **Conditional distribution**: distribution of one variable given a specific value of the other.
- **Simpson's Paradox**: a reversal of association direction when data are aggregated vs. separated by groups.

Key Terms: Scatterplot | Correlation (r) | LSRL | Residual | r^2 | Extrapolation | Influential Point | Two-Way Table | Conditional Distribution | Simpson's Paradox

Unit 3: Collecting Data

3.1 Population vs. Sample

- **Population**: the entire group of individuals we want to learn about.
- **Sample**: a subset of the population that is actually observed.
- **Census**: collecting data from every member of the population.
- **Sampling frame**: the list of individuals from which the sample is drawn.

3.2 Bad Sampling Methods

- **Convenience sampling**: selecting individuals who are easiest to reach - almost always biased.
- **Voluntary response sampling**: individuals choose to participate - biased toward strong opinions.

> **TIP:** *Convenience and voluntary response samples almost always produce biased results. Avoid them on the AP exam!*

3.3 Good Sampling Methods

- **Simple Random Sample (SRS)**: every possible sample of size n has an equal chance of being selected. Every individual has equal chance of selection.
- **Stratified Random Sample**: divide population into strata (homogeneous groups), then take an SRS within each stratum. Reduces variability; ensures representation.
- **Cluster Sample**: divide population into clusters (heterogeneous groups, ideally reflecting the population), randomly select entire clusters, then survey all members of chosen clusters. Practical for large, spread-out populations.
- **Systematic Random Sample**: select every kth individual from a list, starting from a randomly chosen position. Not the same as SRS, but often similar in practice.

> **TIP:** *Know the difference: Stratified = SRS within EACH group; Cluster = random selection of WHOLE groups. Stratified reduces variability; Cluster saves time and money.*

3.4 Bias in Sampling

- **Undercoverage**: some groups in the population are excluded from the sampling frame.
- **Nonresponse**: selected individuals fail to participate or respond.
- **Response bias**: respondents give inaccurate answers (social desirability, leading questions, interviewer influence).
- **Wording bias**: poorly worded or leading questions influence responses.
- **Voluntary response bias**: people with strong opinions are more likely to respond.

3.5 Observational Study vs. Experiment

- **Observational study**: observes individuals and measures variables without imposing treatments. Cannot establish causation; can suggest association.

PART TWO

Multiple-Choice Practice

Answer the following 150 questions to test your knowledge of all 9 AP Statistics units.

Unit 1: Exploring One-Variable Data

1 Which of the following variables is categorical?

(A) Zip code
(B) Annual income
(C) Number of siblings
(D) Temperature in C

2 A distribution is skewed to the right. Which statement is true?

(A) Mean < Median
(B) Mean = Median
(C) Mean > Median
(D) IQR = Range

3 According to the Empirical Rule, approximately what percent of data falls within 2 standard deviations of the mean?

(A) 68%
(B) 95%
(C) 99.7%
(D) 100%

4 The 1.5 * IQR rule is used to identify:

(A) The standard deviation
(B) The median
(C) Outliers
(D) The mean

5 Which statistic is resistant to outliers?

(A) Mean
(B) Standard deviation
(C) Range
(D) Median

6 Standard deviation measures:

(A) Center
(B) Shape
(C) Spread
(D) Frequency

7 The formula for a z-score is:

(A) (x - mu) / sigma
(B) (mu - x) / sigma
(C) x - mu
(D) (x - x-bar) / s

8 If 5 points are added to every score in a dataset, which measure does NOT change?

(A) Mean
(B) Standard deviation
(C) IQR
(D) Both B and C

9 If every value is multiplied by 3, what happens to the standard deviation?

(A) Unchanged.
(B) Multiplied by 3.
(C) Multiplied by 9.
(D) Increases by 3.

10 A percentile measures the:

(A) Percentage of data above a value
(B) Percentage of data below a value
(C) Value at the mean
(D) Range of the middle 50%

11 Which display is best for a categorical variable?

(A) Histogram
(B) Boxplot
(C) Bar chart
(D) Scatterplot

12 A box plot displays the:

(A) Mean and standard deviation
(B) Five-number summary
(C) Frequencies of bins
(D) Individual data points

Answer Key

1. A	**26.** B	**51.** A	**76.** B	**101.** B	**126.** B
2. C	**27.** B	**52.** B	**77.** B	**102.** B	**127.** B
3. B	**28.** C	**53.** B	**78.** A	**103.** C	**128.** B
4. C	**29.** B	**54.** A	**79.** C	**104.** B	**129.** B
5. D	**30.** B	**55.** B	**80.** B	**105.** B	**130.** A
6. C	**31.** B	**56.** B	**81.** A	**106.** B	**131.** B
7. A	**32.** C	**57.** B	**82.** B	**107.** B	**132.** B
8. D	**33.** A	**58.** B	**83.** B	**108.** B	**133.** B
9. B	**34.** B	**59.** B	**84.** D	**109.** A	**134.** B
10. B	**35.** A	**60.** C	**85.** B	**110.** B	**135.** B
11. C	**36.** B	**61.** C	**86.** B	**111.** B	**136.** B
12. B	**37.** B	**62.** C	**87.** B	**112.** B	**137.** B
13. C	**38.** B	**63.** A	**88.** B	**113.** B	**138.** B
14. C	**39.** B	**64.** B	**89.** B	**114.** B	**139.** B
15. B	**40.** C	**65.** C	**90.** B	**115.** B	**140.** B
16. A	**41.** B	**66.** B	**91.** B	**116.** B	**141.** B
17. B	**42.** B	**67.** B	**92.** B	**117.** B	**142.** B
18. B	**43.** B	**68.** B	**93.** D	**118.** B	**143.** B
19. C	**44.** B	**69.** B	**94.** B	**119.** B	**144.** B
20. B	**45.** B	**70.** A	**95.** C	**120.** B	**145.** B
21. B	**46.** B	**71.** B	**96.** B	**121.** B	**146.** B
22. B	**47.** C	**72.** B	**97.** A	**122.** B	**147.** B
23. B	**48.** B	**73.** B	**98.** B	**123.** B	**148.** B
24. B	**49.** A	**74.** B	**99.** B	**124.** B	**149.** B
25. B	**50.** C	**75.** B	**100.** B	**125.** B	**150.** B

PART THREE

Free-Response Practice

Answer the following 10 free-response questions. Allocate approximately 12-15 minutes per question for questions 1-9, and 25 minutes for question 10.

FRQ 1: Exploring One-Variable Data

Prompt:

A study was conducted to investigate the lengths, in minutes, of movies released in 2023. The distribution is roughly symmetric with a mean of 115 minutes and a standard deviation of 18 minutes.

(a) Describe the distribution of movie lengths.
(b) Using the Empirical Rule, approximate the interval that contains the middle 95% of movie lengths.
(c) Would a movie that is 160 minutes long be considered an outlier based on the Empirical Rule? Justify your answer.

Scoring Guide

• **Part (a)**: States shape (approximately symmetric), center (mean = 115 min), and spread (SD = 18 min) in context.
• **Part (b)**: Correctly calculates the interval using the Empirical Rule: 115 +/- 2(18) = (79, 151).
• **Part (c)**: States that 160 min is outside the 95% interval, making it unusual (potential outlier) because 160 > 151.

Sample Response:

(a) The distribution of movie lengths is approximately symmetric, with a mean of 115 minutes and a standard deviation of 18 minutes. There are no obvious outliers.

(b) According to the Empirical Rule, approximately 95% of the data falls within 2 standard deviations of the mean. This gives the interval: 115 +/- 2(18) = 115 +/- 36 = (79, 151) minutes.

(c) A movie length of 160 minutes would be considered unusual. Since 160 minutes is greater than the upper bound of the 95% interval (151 minutes), it falls in the upper 2.5% of the distribution.

FRQ 2: Collecting Data

Prompt:

A pharmaceutical company wants to test a new drug designed to reduce blood pressure. They have 200 volunteers with high blood pressure. They randomly assign 100 to take the new drug and 100 to take a placebo. Blood pressure is measured before and after a 6-week period.

(a) Is this an observational study or an experiment? Explain.
(b) Explain the purpose of the placebo.
(c) Can the results be generalized to all individuals with high blood pressure? Why or why not?

Scoring Guide

- **Part (a)**: Identifies as an experiment because treatments (drug/placebo) are deliberately imposed.
- **Part (b)**: Explains placebo acts as a control for the psychological expectation of treatment (placebo effect).
- **Part (c)**: States results CANNOT be generalized because volunteers are not a random sample from the population.

Sample Response:

(a) This is an experiment because the researchers actively imposed a treatment (the new drug vs. placebo) on the volunteers.

(b) The purpose of the placebo is to act as a control group. It allows researchers to separate the actual physiological effect of the drug from the psychological effect of simply taking a pill (the placebo effect).

(c) No, the results cannot be generalized to all individuals with high blood pressure. The participants were volunteers, not randomly selected from the population. Volunteers may have different characteristics than the general population.

PART FOUR

Exam Resources

Unit : AP Statistics Formula Sheet

I. Descriptive Statistics

Key Formulas:	Sample Mean: x-bar = (**Sum** x_i) / n Sample Standard Deviation: s = sqrt[**Sum**$(x_i$ - x-bar$)^2$ / (n - 1)] Correlation: r = (1/(n-1)) **Sum**$[(x_i$-x-bar$)/s_x][(y_i$-y-bar$)/s_y]$ Linear Regression: y-hat = b0 + b1*x Slope: b1 = r(s_y/s_x) Residual: e = y - y-hat Z-Score: z = (x - *mu*) / *sigma*

II. Probability

Key Formulas:	Complement: P(A^c) = 1 - P(A) Addition: P(A or B) = P(A) + P(B) - P(A and B) Multiplication: P(A and B) = P(A) * P(B|A) Conditional: P(B|A) = P(A and B) / P(A) Binomial: P(X = k) = (n choose k) $p^k(1-p)^{n-k}$ Mean of Binomial: *mu* = np Std Dev of Binomial: *sigma* = sqrt[np(1-p)] Geometric: P(X = k) = $(1-p)^{k-1}$p Mean of Geometric: *mu* = 1 / p Variance of Sum/Diff (Ind): $sigma_{X+/-Y}^2 = sigma_X^2 + sigma_Y^2$

III. Sampling Distributions

Key Formulas:	Sample Proportion Mean: $mu_{p\text{-}hat}$ = p Sample Proportion SD: $sigma_{p\text{-}hat}$ = sqrt[p(1-p)/n] Sample Mean Mean: $mu_{x\text{-}bar}$ = *mu* Sample Mean SD: $sigma_{x\text{-}bar}$ = *sigma* / sqrt(n)

IV. Inferential Statistics

Key Formulas:	CI: statistic +/- (critical value) * (standard error) 1-Sample z-test for prop: z = (p-hat - p0) / sqrt[p0(1-p0)/n] 2-Sample z-test for props: z = (p-hat1 - p-hat2) / sqrt[p-pool(1-p-pool)/n1 + p-pool(1-p-pool)/n2] Pooled p-hat: p-pool = (x1 + x2) / (n1 + n2) 1-Sample t-test for mean: t = (x-bar - *mu*0) / (s/sqrt(n)) Paired t-test: t = (d-bar - 0) / (s_d/sqrt(n)) 2-Sample t-test: t = (x-bar1 - x-bar2) / sqrt[$s1^2/n1 + s2^2/n2$] *chi*-Square Test: chi^2 = **Sum** $(O - E)^2 / E$ Expected Count: E = (row total * col total) / grand total t-Test for Slope: t = (b1 - 0) / SE_{b1}

Unit : Graphing Calculator Tips (TI-83/84)

Knowing calculator shortcuts saves valuable time on the AP Exam. All commands are accessed via the calculator's menu system.

1. Descriptive Statistics

• **STAT > CALC > 1-Var Stats**: Finds mean (x-bar), sum, sample SD (Sx), population SD, median, Q1, Q3, min, max. Specify your list (e.g., L1).

• **STAT > CALC > 2-Var Stats**: Finds means, SDs, min/max for both x and y, plus sample size (n).

2. Linear Regression

• **STAT > CALC > LinReg(a+bx)**: Enter x-list and y-list. IMPORTANT: Turn diagnostics on via **2nd > 0 (CATALOG) > DiagnosticOn** to get r and r-squared.

• **STAT > TESTS > LinRegTTest**: Gets slope, intercept, r, r-squared, s, t-statistic, and P-value for slope.

3. Normal Distribution

• **2nd > VARS (DISTR) > normalcdf**: Finds proportion/probability between two values. Use -1E99 for negative infinity and 1E99 for positive infinity.

• **2nd > VARS (DISTR) > invNorm**: Finds the value that corresponds to a given cumulative area (left-tail).

4. Binomial & Geometric Distributions

• **2nd > VARS > binompdf(n, p, x)**: P(X = x) for exactly x successes.

• **2nd > VARS > binomcdf(n, p, x)**: P(X <= x) for at most x successes. For 'at least k', use 1 - binomcdf(n, p, k-1).

• **2nd > VARS > geometpdf(p, x)** and **geometcdf(p, x)**: Similar logic for first success on trial x.

5. Hypothesis Tests & Confidence Intervals

• Access all via **STAT > TESTS**. Choose 'Data' if you have raw data in a list, or 'Stats' if you have calculated values.

• **1-PropZTest / 1-PropZInt**: Enter p0, x, and n for tests. (x is the count, not p-hat!).

• **2-SampTTest / 2-SampTInt**: Select 'No' for Pooled (AP exam does not pool variances unless instructed).

6. Chi-Square Tests

• **STAT > TESTS > Chi-Square Test**: You MUST enter observed counts into a matrix first. Press **2nd > x^-1 (MATRIX) > EDIT > [A]** to enter dimensions and values.

Unit : Exam Pacing Strategy

The AP Statistics Exam is 3 hours long. Time management is critical.

Section I: Multiple Choice

- **Format:** 40 Questions | **Time:** 90 Minutes | **Pace:** approx 2 mins 15 secs per question.
- **Strategy:** First pass (60 mins): Answer easiest questions first. Second pass (25 mins): Return to circled questions, eliminate choices, guess. Final Check (5 mins): Ensure bubbles are filled correctly. No penalty for guessing!

Section II: Free Response

- **Format:** 5 Short Answer + 1 Investigative Task | **Time:** 90 Minutes
- **Pace:** Approx 12 mins per short-answer question; 25 mins for the Investigative Task.
- **Strategy:** Read all 6 prompts quickly. Start with the question you feel most confident about. Write clearly, show work/calculator inputs, use proper vocabulary. Never leave a FRQ blank; partial credit is available.

Unit : Score Interpretation Chart

AP Scores are translated from a weighted composite score. The ranges below are approximate.

5	70 - 100	Extremely Well Qualified
4	55 - 69	Well Qualified
3	40 - 54	Qualified
2	25 - 39	Possibly Qualified
1	0 - 24	No Recommendation

How it's calculated: Multiple-Choice accounts for 50% of your composite score, and Free-Response accounts for the other 50%. Question 6 (Investigative Task) counts for roughly 25% of the FRQ section weight.

Thank You for Your Purchase!

Your support means the world to **UM Printables**. We put hundreds of hours into crafting premium, rigorous, and beautifully designed educational resources.

How you can help:

If you found this AP Statistics Exam Prep resource helpful, please take a moment to leave a review on our Teachers Pay Teachers store. Your honest feedback not only helps us improve but also helps other educators find high-quality materials for their classrooms.

Just click the **"Leave a Review"** button on your "My Purchases" page on TPT. Every review earns you TPT credits toward future purchases!

Thank you for choosing UM Printables. We wish your students the best of luck on the AP Exam!

UM Printables

Premium Teaching Resources for Educators