

AP Statistics Formula and Strategy Guide

Exploring data, collecting data, probability, and statistical inference — the four AP Statistics pillars, with the moment each procedure actually applies. Work a section, then try the matching questions in the practice worksheet.

SECTIONS	5	WORKED EXAMPLES	2	FORMAT	Print friendly
----------	---	-----------------	---	--------	----------------

How to use this guide

A formula you cannot place is not knowledge yet. Work the recall drill below before the questions.

- Name the procedure before running it: one-sample, two-sample, proportion, or mean — the label determines everything else.
- Always check conditions before running a test. A correct calculation with unchecked conditions earns partial credit at best.
- A conclusion always ties back to context — 'reject H_0 ' alone is not a finished answer.

1. Exploring one- and two-variable data

NAME	FORMULA	WHEN TO USE
Mean	$\text{sum} / \text{count}$	A typical value with no major outliers.
Standard deviation	typical distance of data from the mean	Describing spread.
Five-number summary	min, Q1, median, Q3, max	Building a boxplot or checking for outliers.
IQR outlier rule	below $Q1 - 1.5(IQR)$ or above $Q3 + 1.5(IQR)$	Flagging an unusually extreme value.
Correlation r	measures strength/direction of a LINEAR relationship	Describing a scatterplot's pattern.
Least-squares regression	$\hat{y} = a + bx$, minimizes sum of squared residuals	Predicting y from x with a line of best fit.

WORKED EXAMPLE

A data set has $Q1 = 20$, $Q3 = 32$. Find the IQR and the upper outlier boundary.

- $IQR = Q3 - Q1 = 32 - 20 = 12$
- Upper boundary = $Q3 + 1.5(IQR) = 32 + 18$
- $= 50$

Check: Any value above 50 would be flagged as an outlier.

2. Collecting data

NAME	FORMULA	WHEN TO USE
Simple random sample	every group of n has equal chance of selection	Reducing bias in sample selection.
Stratified sample	divide into groups, randomly sample within each	Ensuring subgroup representation.
Observational study	no treatment is imposed	Only association, not causation, can be concluded.
Randomized experiment	treatments are randomly assigned	Allows a causal conclusion.
Confounding variable	affects both the explanatory and response variable	An alternative explanation for an observed association.

3. Probability and random variables

NAME	FORMULA	WHEN TO USE
Addition rule	$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$	Combining probabilities of two events.
Multiplication rule (independent)	$P(A \text{ and } B) = P(A) \times P(B)$	Two events don't affect each other.
Conditional probability	$P(A B) = P(A \text{ and } B)/P(B)$	The probability of A given that B has occurred.
Mean of a random variable	$\sum x \cdot P(x)$	The expected value of a random variable.
Binomial probability	n choose k , times p^k , times $(1-p)^{n-k}$	A fixed number of independent yes/no trials.

WORKED EXAMPLE

A biased coin lands heads with probability 0.3. Find the probability of exactly 2 heads in 3 flips.

1. Binomial with $n=3$, $k=2$, $p=0.3$
2. $C(3,2) \times 0.3^2 \times 0.7^1$
3. $3 \times 0.09 \times 0.7 = 0.189$

Check: 0.189 is a reasonable probability, well under 1.

4. Sampling distributions and inference

NAME	FORMULA	WHEN TO USE
Sampling distribution of $\hat{p}$	mean p , standard deviation $\sqrt{p(1-p)/n}$	Describing how a sample proportion varies.
Confidence interval (general form)	estimate $\pm$ (critical value)(standard error)	Estimating a population parameter with a margin of error.

NAME	FORMULA	WHEN TO USE
Conditions for inference (proportions)	Random, 10%, Large Counts	Checking whether a procedure is valid before running it.
Significance test logic	assume H_0 true, find how surprising the data is (p-value)	Deciding whether evidence supports a claim.
Type I / Type II error	rejecting a true H_0 / failing to reject a false H_0	Describing the two ways a decision can be wrong.

5. Chi-square and regression inference

NAME	FORMULA	WHEN TO USE
Chi-square goodness of fit	compares observed counts to an expected distribution	Testing whether one categorical variable matches a claimed distribution.
Chi-square test of association	tests whether two categorical variables are related	Analyzing a two-way table.
Inference for slope	tests whether the true regression slope is nonzero	Deciding whether x is a useful predictor of y .

Notes and tutor hand-off

Bring this page to your next session so the tutor can start from what actually blocked you.

Which inference procedure (proportion or mean, one-sample or two-sample) needs review?

One condition you keep forgetting to check:

Worksheet question numbers to review with a tutor:
