SI-06 SI-06: Statistics, Error Bars & Evidence Claims

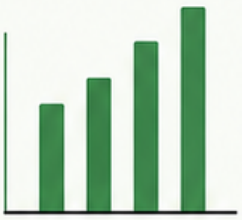
From Data to Defensible Claim — Writing Statistical Conclusions That Earn Full Points

AP Bio FRQs constantly ask students to support, refute, or evaluate claims using data. The bottleneck isn't math — most students can compute a chi-square value or read an error bar. The bottleneck is **writing the conclusion sentence**: which words are scientifically defensible, which words ('prove', 'definitely') are landmines, and what the rubric is

Q1, Q2, Q3, Q6 | Skills 5 & 6 | Statistical reasoning | Argumentation

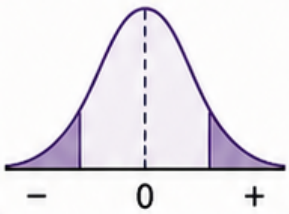
1 The Data -> Claim Pipeline

1 DATA (raw)



2 TEST/COMPARE (error bars, chi-square, CI, etc.)

3 STATISTICAL JUDGMENT (overlap? sig? reject null?)



4 WRITTEN CLAIM (sentence with defensible language)

The pipeline AP graders are checking:

1. Did you read the data correctly?
2. Did you apply the correct statistical comparison?
3. Did you reach a defensible conclusion (not 'this proves...')?
4. Did you tie the conclusion back to the biological claim?

If any link breaks, you lose points even if the math is right.

Page focus: follow the full path from raw data to defensible claim.

SI-06 SI-06: Statistics, Error Bars & Evidence Claims

2. SD vs. SE vs. 95% CI

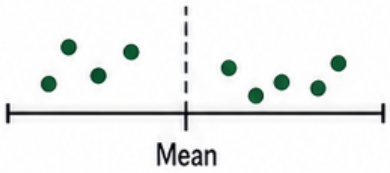
Know what each statistic measures before you interpret the graph.

AP Biology commonly tests three different error/variability measures. Each answers a different question about your data—make sure you know which one your graph is showing.

1 Standard Deviation (SD)

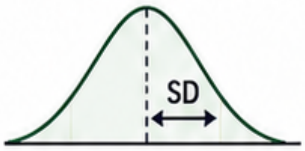
What it measures

The spread of individual data points around the mean.



When you'll see it on AP

When graphs show how much individual data values vary ("dispersion," "variability," or "spread" of the data).



2 Standard Error (SE)

What it measures

The uncertainty in the mean itself.
SE = SD / square root of *n*.

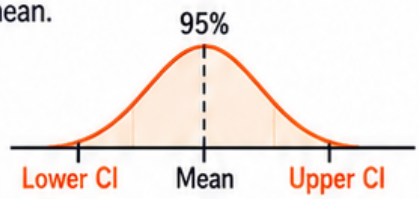
When you'll see it on AP

Most often as error bars on the mean (bar graphs or dot plots) to show how precisely the mean has been estimated.

3 95% Confidence Interval (CI)

What it measures

The estimated range for the population mean. If many samples were taken and a CI calculated for each, about 95% of those intervals would contain the true mean.



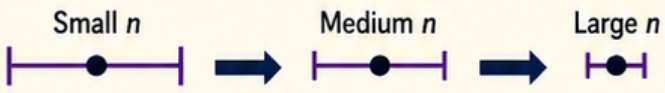
When you'll see it on AP

When graphs report a 95% CI around the mean (whiskers or brackets that show the estimated range of the population mean).

Statistic	What it measures	When you'll see it on AP
SD	Spread of individual data points around the mean.	When graphs show how much individual data values vary ("dispersion," "variability," or "spread" of the data).
SE	Uncertainty in the mean itself. SE = SD / square root of *n*.	Most often as error bars on the mean (bar graphs or dot plots) to show how precisely the mean has been estimated.
95% CI	Estimated range for the population mean. If many samples were taken and a CI calculated for each, about 95% of those intervals would contain the true mean.	When graphs report a 95% CI around the mean (whiskers or brackets that show the estimated range of the population mean).

Key Relationship

Larger sample size *n*
→ smaller SE → tighter error bars
→ more confident estimate of the mean.



On the AP equations sheet:

$$SE = \frac{SD}{\sqrt{n}}$$

SE = SD / square root of *n*.
It is provided.
You do not need to memorize it, but you do need to use it correctly.

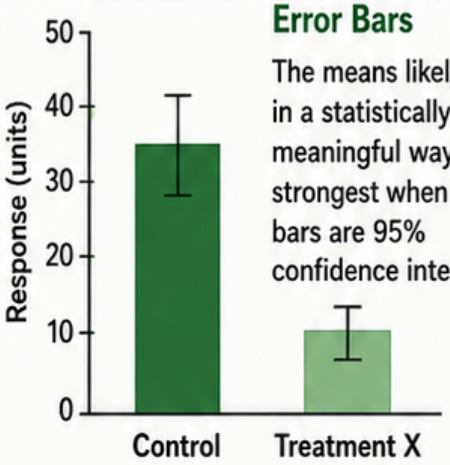
Page focus: distinguish variability in data from uncertainty in the mean

SI-06 SI-06: Statistics, Error Bars & Evidence Claims

One of the most-tested visual interpretation skills in AP Bio FRQs.

3 Error Bar Overlap: The 3 Cases

CASE 1 Non-Overlapping Error Bars



The means likely differ in a statistically meaningful way, strongest when the bars are 95% confidence intervals.

What to write

"Because the 95% confidence intervals of the two groups do not overlap, the difference between the means is likely statistically significant. The data support the claim that [treatment X has an effect on Y]."

Important note

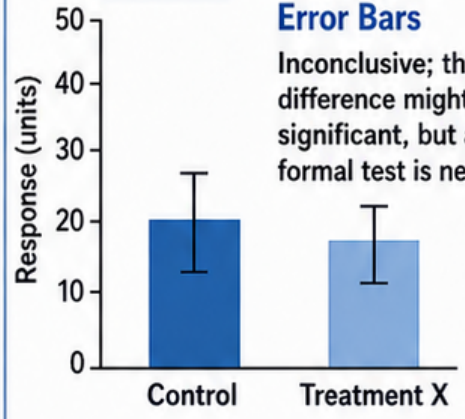
If the graph shows SE error bars (not CIs), do not treat SE bars as 95% CIs directly. AP allows the approximation 95% CI ≈ mean ± 2 SE for normally distributed data; only then use CI-overlap logic.

What NOT to write

- ✗ Do not say "This proves X causes Y."
- ✗ Do not say "X definitely causes Y."
- ✗ Do not treat SE bars as CIs without the ± 2 SE approximation.

Safer wording:
"These data are consistent with X causing Y" or "support the claim that…"

CASE 2 Partially Overlapping Error Bars



Inconclusive; the difference might be significant, but a formal test is needed.

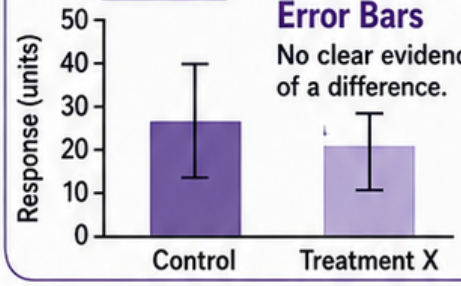
Cautious writing

"The error bars partially overlap, so a statistical test (such as a t-test) would be needed to determine whether the difference is significant. The data are consistent with but do not strongly support the claim that [X affects Y]."

AP context note

Unless a formal test is provided, describe partial overlap cautiously, especially if the bars are SE rather than CIs.

CASE 3 Fully Overlapping Error Bars



No clear evidence of a difference.

What to write

"Because the error bars overlap substantially, the data do not provide clear evidence of a statistically significant difference between the groups. The claim that [X affects Y] is not supported by these data."

SUMMARY: Match the overlap to your conclusion.

- **Non-overlap = stronger support** — Likely significant (especially with 95% CIs).
- **Partial overlap = cautious / need test** — Inconclusive without a formal statistical test.
- **Full overlap = not supported** — No clear evidence of a difference.

Page focus: match your wording to the strength of the overlap evidence

SI-06 SI-06: Statistics, Error Bars & Evidence Claims

Observed vs. expected: use the full AP workflow and write the correct conclusion.

chi-square (chi-squared, symbol chi with superscript 2) tests whether observed data fit an expected distribution such as a Mendelian ratio.

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

4 Chi-Square: The 6-Step Protocol

1 STATE THE NULL HYPOTHESIS

There is no significant difference between observed and expected values; the data fit the expected ratio; any differences are due to chance alone.

All are AP-acceptable phrasings.

2 CALCULATE EXPECTED VALUES

Calculate expected values from predicted ratio times total observed.

3 COMPUTE CHI-SQUARED

Compute chi-squared for each category: $(O - E)^2 / E$, then sum.

4 DETERMINE DEGREES OF FREEDOM

df = number of categories – 1.

5 FIND THE CRITICAL VALUE

Find the critical value at $p = 0.05$ using the AP equations sheet.

6 COMPARE AND CONCLUDE

Compare calculated chi-square to the critical value and **write a conclusion** that answers the biological question.

RESULT	IF...	THEN CONCLUDE...
chi-squared > critical value	→	**Reject the null hypothesis;** observed values differ significantly from expected; data do not fit the predicted ratio.
chi-squared < critical value	→	**Fail to reject the null hypothesis;** observed values are consistent with expected; data fit the predicted ratio.

Worked Example — Mendelian Cross

A monohybrid cross is expected to produce a 3:1 ratio of tall:short plants. Observed: 78 tall and 22 short (total = 100).

Calculation Table

Category	Observed (O)	Expected (E) (3:1 ratio)	$(O - E)^2$	$(O - E)^2 / E$
Tall	78	75 (0.75 × 100)	$(78 - 75)^2 = 9$	9 / 75 = 0.12
Short	22	25 (0.25 × 100)	$(22 - 25)^2 = 9$	9 / 25 = 0.36
TOTAL	100	—	—	TOTAL $\chi^2 = 0.48$

Steps 4–6

Degrees of freedom (df):
df = number of categories – 1 = 2 – 1 = 1

Critical value at $p = 0.05$ (df = 1):
3.84

Comparison:
0.48 < 3.84 → fail to reject the null hypothesis.

Because chi-squared (0.48) is less than the critical value (3.84) at df = 1 and $p = 0.05$, we fail to reject the null hypothesis. The observed data are consistent with a 3:1 Mendelian ratio.

Page focus: AP scoring loves the full chain — null, calculation, df, critical value, conclusion

SI-06 SI-06: Statistics, Error Bars & Evidence Claims

Use scientifically defensible language and know what changes when n increases.

5 p-Value Language Traps

✗ Don't say	✓ Say this
"The data **prove** the hypothesis."	"The data **support** the hypothesis" or "are **consistent** with the hypothesis."
"We **accept** the alternative hypothesis."	"We **reject** the null hypothesis."
"$p < 0.05$ means the result is **true**."	$p < 0.05$ means there is less than a 5% probability of observing results this extreme or more extreme, assuming the null hypothesis is true.
"There is **definitely** an effect."	"The data provide **statistically significant** evidence of an effect."
"$p = 0$ means **certainty**."	p never equals zero in practice; **smaller** p means stronger evidence against the null.

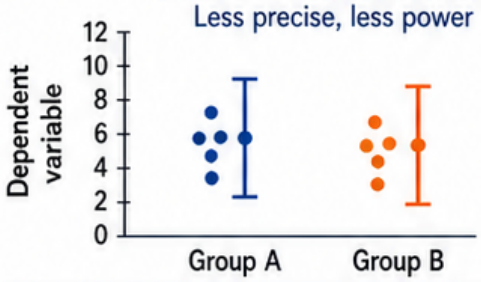
Two unbreakable rules

1. Never use the word "**prove**" in a statistical conclusion. Science supports, refutes, or fails to refute.
2. Safest AP wording: "**reject the null**" or "**fail to reject the null**"—not prove, not accept the alternative.

6 Sample Size & Standard Error

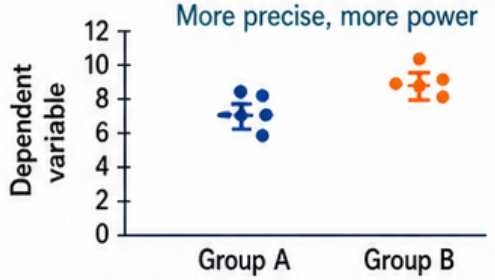
Sample Size and Standard Error: The Inverse Relationship

Small *n* → Wide Error Bars
Less precise, less power



Wide error bars overlap → harder to detect real differences.

Large *n* → Narrow Error Bars
More precise, more power



Narrow error bars overlap less → easier to detect real differences.

Increasing sample size **decreases standard error** (**SE = SD / square root of *n***), which produces narrower error bars. With narrower error bars, smaller true differences between groups become detectable, and statistical conclusions can be drawn with higher confidence—though SD itself reflects population variability and may not change with larger *n*.

1. **SD** measures spread of individual data points and **does not** decrease just because *n* increases.
2. **SE** measures uncertainty in the mean and **does** decrease with larger *n*.
3. **CI** is built from SE and also **narrows** with larger *n*.

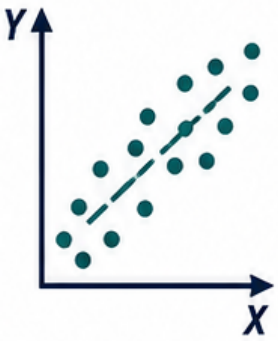
Page focus: stronger evidence needs stronger language;

SI-06 SI-06: Statistics, Error Bars & Evidence Claims

Correlation alone is not causation; claim, evidence, and reasoning must match the data.

7 Correlation ≠ Causation

Correlation ≠ Causation



The data show a strong positive correlation between [X] and [Y]. However, correlation **does not establish causation**—a confounding variable could affect both, or the direction of causation could be reversed. A controlled experiment manipulating [X] would be required to support a causal claim.

When can you claim causation?

1. The experiment manipulated the independent variable directly.
2. Confounding variables were controlled.
3. Replicates support reproducibility.
4. A plausible biological mechanism exists.

8 CER Templates for Data-Based Claims

A Template A Supporting a Claim with Data

Claim:

The data support the claim that enzyme X is sensitive to high temperature.

Evidence:

At 60 °C, enzyme X activity averaged 22.4 units, while at 40 °C it averaged 68.7 units—a difference of 46.3 units. The error bars do not overlap.

Reasoning:

Proteins can denature at temperatures above their optimal range, reducing active sites and decreasing enzyme activity.

B Template B Refuting a Claim with Data

Claim:

The data refute the claim that enzyme X is sensitive to high temperature.

Evidence:

There is no significant difference between 40 °C and 60 °C, or the results show the opposite direction, or the error bars overlap.

Reasoning:

If the claim were correct, we would expect a large decrease in activity at high temperature. The data do not match this pattern, so the claim is not supported.

C Template C Inconclusive Data

Claim:

The data are inconclusive about whether enzyme X is sensitive to high temperature.

Evidence:

The data show a partial trend, but the error bars overlap, and/or the sample size is small.

Reasoning:

A statistical test or larger sample size would be needed to determine if the difference is significant. With the current data, we cannot strongly support or refute the claim.

Strong AP answers connect biological reasoning to the exact data pattern.

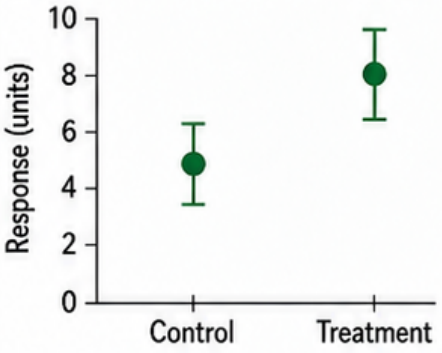
Page focus: cite the data, then explain what the data can and cannot justify

SI-06: Statistics, Error Bars & Evidence Claims

Practice the wording AP graders want to see.

9 Practice: Write the Defensible Conclusion

1 Error Bar Interpretation



Control:
mean = 5.0
SE = 0.5
(error bar 4.5 to 5.5)

Treatment:
mean = 8.0
SE = 0.6
(error bar 7.4 to 8.6)

Q: Does the evidence support the claim that the treatment has an effect?

Defensible Answer:

The data are reported with SE error bars. Using the AP approximation that 95% CI is about mean plus or minus 2 SE, the control CI is about 4.0 to 6.0 and the treatment CI is about 6.8 to 9.2. These intervals do not overlap, so the data support the claim that the treatment has a statistically significant effect. The treatment mean is about 3.0 units higher than the control mean.

2 Chi-Square

Genetic cross: observed counts and expected 3:1 ratio.

Phenotype	Observed (O)	Expected (E) 3:1 ratio	$(O-E)^2 / E$
Dominant phenotype	84	82.5	0.027
Recessive phenotype	26	27.5	0.082
Total	**110**	**110**	**0.109**

- Chi-squared = 0.027 + 0.082 = 0.109
- df = 1
- Critical value (0.05 level) = 3.84

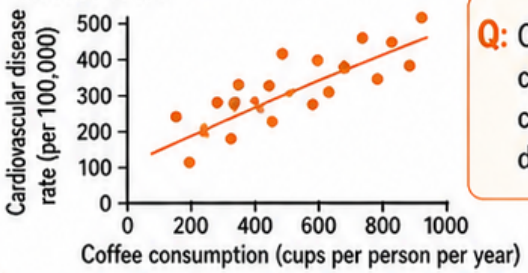
Q: Do the observed data fit the expected 3:1 ratio?

Defensible Answer:

Chi-squared = 0.109 with df = 1. The critical value at $\alpha = 0.05$ is 3.84. Because $0.109 < 3.84$, we fail to reject the null hypothesis. The observed data are consistent with the expected 3:1 Mendelian ratio.

3 Correlation vs. Causation

Data: Countries with higher per-capita coffee consumption have higher cardiovascular disease rates.



Q: Can we conclude coffee causes cardiovascular disease?

Defensible Answer:

The data show a positive correlation at the population level. However, correlation does not establish causation. Confounding variables such as smoking rates, diet, age distribution, or healthcare access could explain the pattern. A controlled longitudinal study would be needed to evaluate causality.

Defensible conclusions are specific, cautious, and tied to the correct statistical logic.

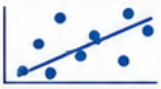
Page focus: practice the exact sentence forms that earn points

SI-06 SI-06: Statistics, Error Bars & Evidence Claims

Ultra-compact review: phrase your conclusion to match the strength of the evidence.

10 One-Page Cheat Card

The 6 must-know phrasings

1		Non-overlapping 95% CIs	→	**"likely statistically significant"** (treat SE bars as CIs only after using plus/minus 2 SE approximation).
2		Substantially overlapping error bars	→	**"no clear evidence of difference"** Partial overlap requires cautious language or a formal test.
3	$\chi^2 < \text{crit}$	Chi-squared < critical	→	**"fail to reject the null"**
4	$\chi^2 > \text{crit}$	Chi-squared > critical	→	**"reject the null"**
5		Correlation only	→	**"does not establish causation"**
6		Larger n	→	**"decreases SE, narrower error bars, more confident conclusion"**

Words to avoid (every time)

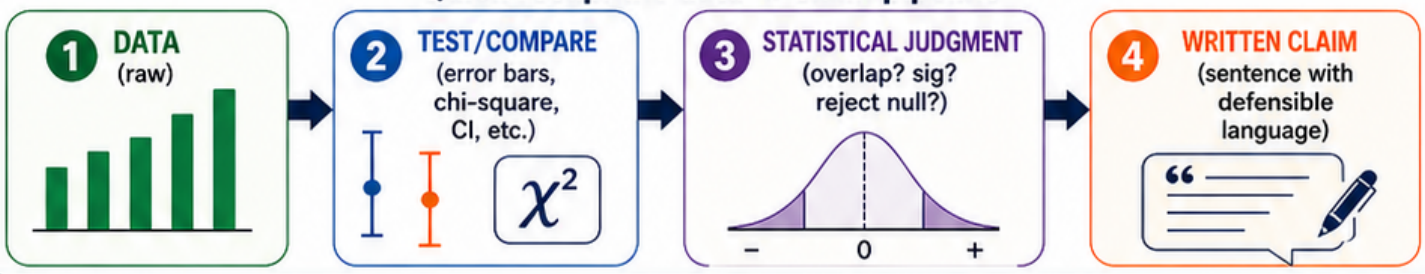
- ✗ prove
- ✗ definitely
- ✗ accept the alternative
- ✗ "the data show that X causes Y" without manipulation

Words to use

- ✓ support
- ✓ consistent with
- ✓ fail to reject
- ✓ reject the null
- ✓ statistically significant

Match your conclusion language to the strength of the evidence.
Strong, non-overlapping data -> support.
Weak or overlapping data -> consistent with or *do not support*.
Correlation only -> does not establish causation.

Quick recap: the data -> claim pipeline



End of SI-06 content — review complete.

Page focus: careful wording turns correct analysis into earned points.