

Sampling Distributions

AP Statistics ■ Unit 5

AP Statistics



What we will cover

- 1 Statistic vs parameter
- 2 The Central Limit Theorem
- 3 Bias and variability
- 4 Sampling distribution of a proportion
- 5 Sampling distribution of a mean
- 6 Comparing two groups





Statistic vs parameter

A **statistic** ($\bar{x}$, $\hat{p}$) varies sample to sample – **sampling variability**. A **parameter** (μ , p) is fixed.

- **parameter** → **population**; **statistic** → **sample**

The **sampling distribution** of a statistic is its distribution over **all** samples of size n .

$\hat{p}$ each differs

0.60

0.52

0.68

three samples of $n = 25$

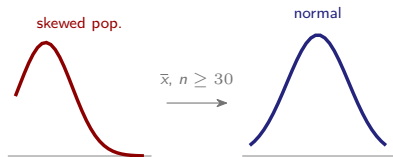


The Central Limit Theorem

The **CLT** : for large n , the sampling distribution of $\bar{x}$ is **approximately normal** – *whatever* the population shape.

- AP guideline: $n \geq 30$
- larger $n \Rightarrow$ more normal & narrower

Averaging cancels the extremes, smoothing a skewed population into a bell.





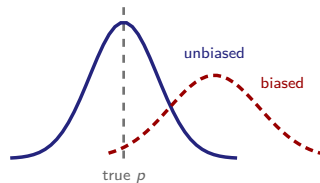
Bias and variability

A **point estimate** uses one sample.

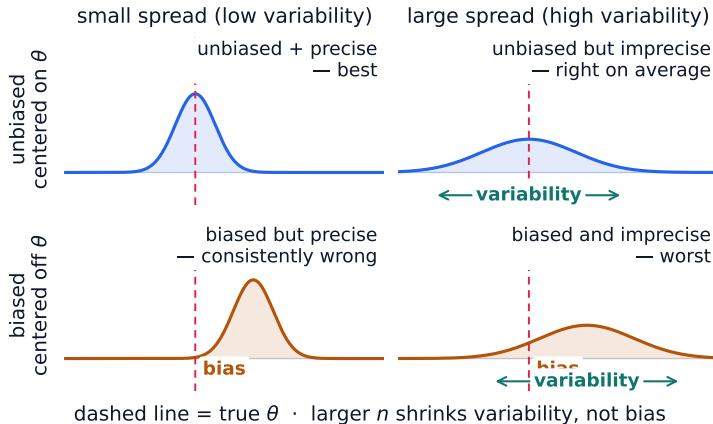
An **estimator** is **unbiased** when its sampling distribution centers **on** the parameter.

- bias is about *aim*
- **variability** shrinks as n grows

Ideal: unbiased *and* low variability
– a bigger sample buys the second.



Good Guesses and Bad Guesses: Bias



Bias and variability are separate faults. Only the top-left estimator is both centered on θ and tight; the bottom-left one is precise but



Sampling distribution of a proportion

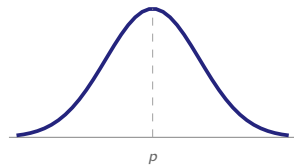
For a sample proportion $\hat{p}$:

$$\mu_{\hat{p}} = p \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Normal when two conditions hold:

- **large counts** : $np \geq 10$ and $n(1-p) \geq 10$
- **10% condition** 10%: $n \leq 0.10 N$

$\hat{p}$ is unbiased; bigger n under the root
 $\Rightarrow$ less bounce.



Centered on p ,

spread $\sigma_{\hat{p}}$.

Worked example – the standard error of a proportion

40% of voters favour a measure ($p = 0.4$) and you sample $n = 100$. Find the standard error, and the chance of seeing $\hat{p} \geq 0.5$.

- $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.4(0.6)}{100}} = 0.049$
- check the conditions: $np = 40$ and $n(1-p) = 60$, both ≥ 10 , so the distribution is approximately normal
- $z = \frac{0.5 - 0.4}{0.049} = 2.04$, so $P \approx 0.021$

The standard error shrinks as $\sqrt{n}$: to **halve** it you need **four times** the sample. That is why large surveys cost what they do.

Worked example – check before you normal

12% of a town's 40,000 adults cycle to work. A survey takes $n = 150$. Is $\hat{p}$ approximately normal, and how much does it bounce?

large counts: $np = 150(0.12) = 18$

$$n(1 - p) = 150(0.88) = 132$$

both ≥ 10 ✓

10%: $150 \leq 0.10(40,000) = 4000$ ✓

So normal, with $\sigma_{\hat{p}} = \sqrt{\frac{0.12(0.88)}{150}} \approx 0.027$.

Check **both** counts – the smaller one fails first, and here $np = 18$ is what nearly bit. State the conditions **with their numbers**: “ $np \geq 10$ ” alone earns nothing.



Sampling distribution of a mean

For a sample mean $\bar{x}$:

$$\mu_{\bar{x}} = \mu \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The spread $\sigma/\sqrt{n}$ is the **standard error**.

- normal population $\Rightarrow \bar{x}$ normal for any n
- else lean on the CLT ($n \geq 30$)

Quadruple n to **halve** the spread – n sits under a square root.

Scores

$\mu = 500, \sigma = 100, n = 25$:

$$\sigma_{\bar{x}} = \frac{100}{\sqrt{25}} = 20$$

$P(\bar{x} > 530)$: $z = \frac{530-500}{20} = 1.5$,
area ≈ 0.067 .



Comparing two groups

For a difference of independent statistics, means combine and **variances add**:

Two proportions

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

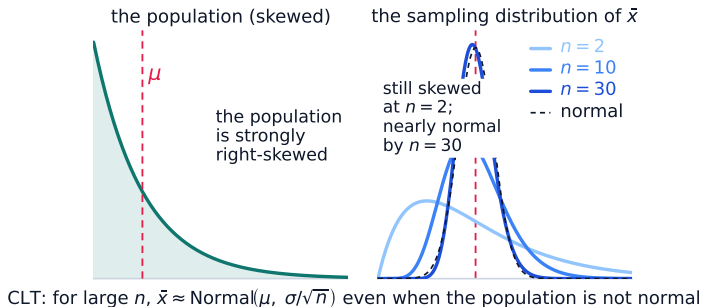
Two means

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Add the two variances under the root – for *both* a sum and a difference – then take the square root.

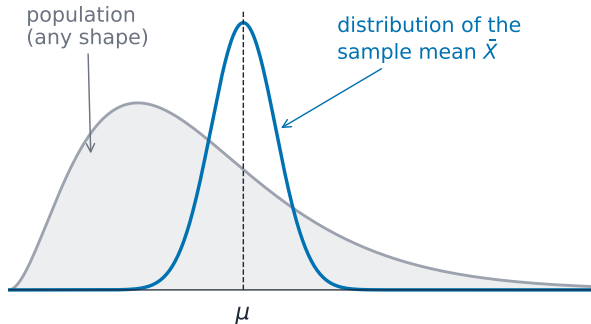
The Sampling Distribution of a Sample Mean

The population on the left is strongly skewed, yet every sampling distribution of $\bar{x}$ is centered at μ . A larger n shrinks the standard error $\sigma/\sqrt{n}$, so the curve gets taller and narrower – and it also straightens: still clearly skewed at $n = 2$, almost exactly normal (dashed) by $n = 30$.



The population on the left is strongly skewed, yet every sampling distribution of $\bar{x}$ is centered at μ .

The Central Limit Theorem, continued



The sample mean is nearly normal whatever the shape of the population

Exam tips

- A **sampling distribution** is the distribution of a statistic over many samples, **centered on the true parameter**.
- The **Central Limit Theorem**: for a large enough sample the sample mean is approximately normal, even if the population is not.
- Larger samples give **less** variability (a smaller standard error).
- Check the conditions (random, independent/10%, large enough) before using a normal model.
- Keep straight what varies – the statistic – versus the fixed parameter.



Unit 5 – key takeaways

1

Terms

parameter fixed; statistic varies;
sampling distribution

2

CLT

$n \geq 30 \Rightarrow \bar{x}$ normal for any population

3

Proportion

$\mu_{\hat{p}} = p$, $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$; large counts

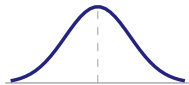
4

Mean

$\mu_{\bar{x}} = \mu$, standard error $\sigma/\sqrt{n}$;
variances add

Check yourself

- 1 Is μ a parameter or a statistic?
- 2 What does the CLT promise for large n ?
- 3 Write the standard error of a sample mean.
- 4 When combining independent variables, what adds?



Answers 1. parameter 2. $\bar{x}$ is approximately normal 3. $\sigma/\sqrt{n}$ 4. the variances