

Sampling Distributions

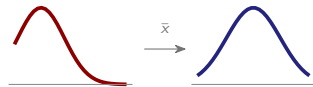
AP Statistics ■ Unit 5

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** 统计量 ($\bar{x}$, $\hat{p}$) varies sample to sample – **sampling variability** 抽样变异性. A **parameter** 参数 (μ , p) is fixed.

- parameter → population; statistic → sample

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

$\hat{p}$ each differs

0.60

0.52

0.68

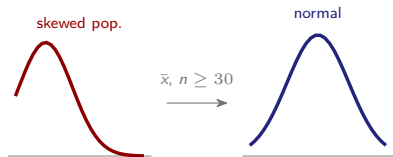
three samples of $n = 25$

The Central Limit Theorem

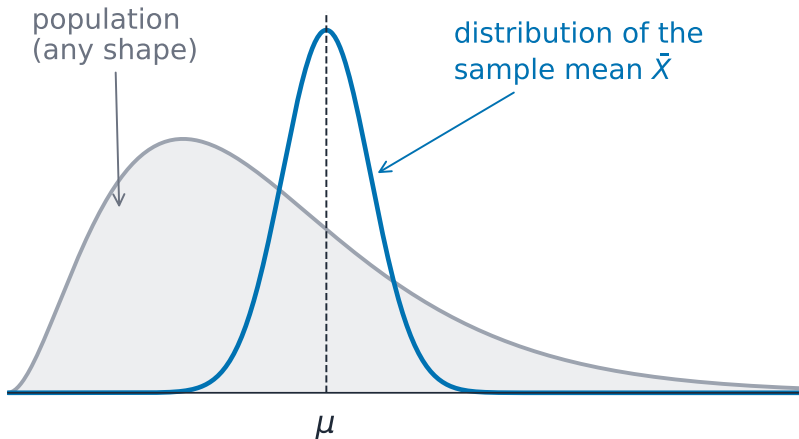
The **CLT** 中心极限定理: for large n , the sampling distribution of $\bar{x}$ is **approximately normal** – *whatever* the population shape.

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

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



The Central Limit Theorem

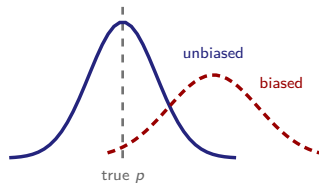


Bias and variability

A **point estimate** 点估计 uses one sample. An **estimator** 估计量 is **unbiased** 无偏 when its sampling distribution centers **on** the parameter.

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

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



Sampling distribution of a proportion

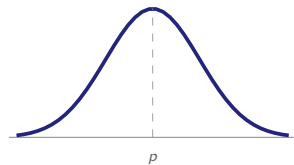
For a sample proportion $\hat{p}$:

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

Normal when two conditions hold:

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

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



Centered on p ,

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

Worked example – 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** 标准误.

- 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

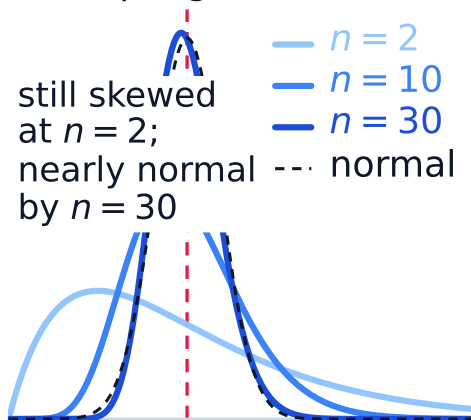
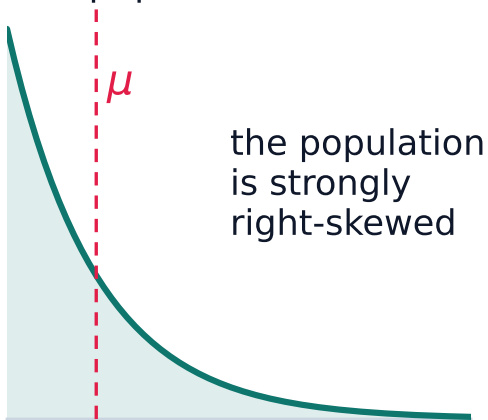
$\mu = 500$, $\sigma = 100$, $n = 25$:

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

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

Sampling distribution of a mean

the population (skewed) the sampling distribution of $\bar{x}$



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.

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