

AP Statistics - Sampling Distributions and the Central Limit Theorem - Lesson

Common Core AP STAT 5.1, AP STAT 5.2, AP STAT 5.3, AP STAT 5.4, AP STAT 5.5, AP STAT 5.6, AP STAT 5.7, HSS-IC.A.1, HSS-IC.A.2 · maths-pdf.fr/en/

A sampling distribution shows how a statistic varies across repeated random samples of the same size. The sample proportion has mean p and standard deviation $\sqrt{p(1-p)/n}$, and the sample mean has mean μ and standard deviation $\sigma/\sqrt{n}$. By the central limit theorem, sample means are approximately normal for large n , whatever the population shape.

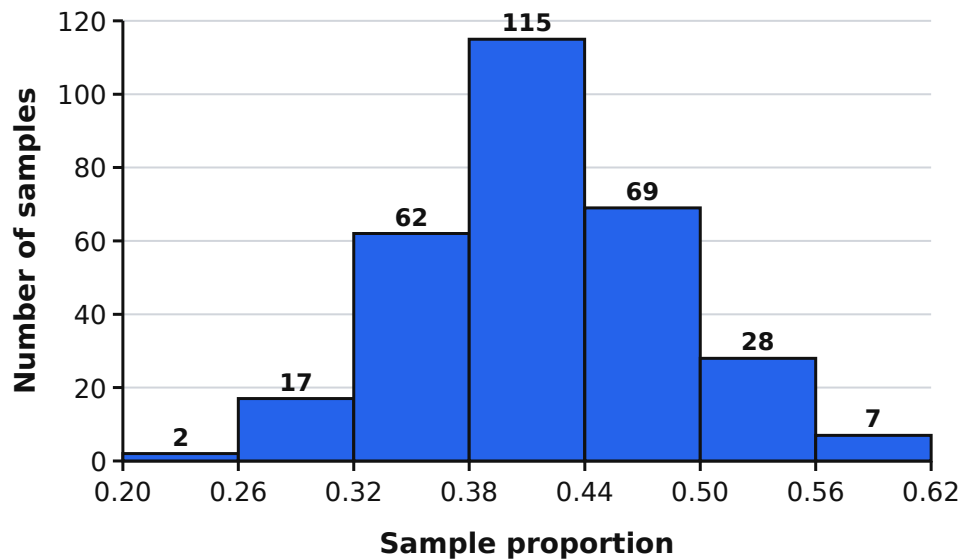
AP Statistics · Common Core State Standards: **AP STAT 5.1, AP STAT 5.2, AP STAT 5.3, AP STAT 5.4, AP STAT 5.5, AP STAT 5.6, AP STAT 5.7, HSS-IC.A.1, HSS-IC.A.2**

Key facts

- A sampling distribution describes how a statistic varies from sample to sample.
- $\hat{p}$: mean p , sd $\sqrt{p(1-p)/n}$; approximately normal if $np \geq 10$ and $n(1-p) \geq 10$.
- $\bar{x}$: mean μ , sd $\sigma/\sqrt{n}$; normal if the population is normal or n is large (CLT).
- Differences of independent statistics: means subtract, variances add.
- Larger n reduces variability but never corrects bias; the 10% condition is needed for independence.

What is a sampling distribution?

A **sampling distribution** shows all the values a statistic takes in repeated random samples of the same size from the same population, with how often each occurs. It describes *sampling variability*: the statistic changes from sample to sample.



300 simulated sample proportions: centered at 0.40, approximately normal.

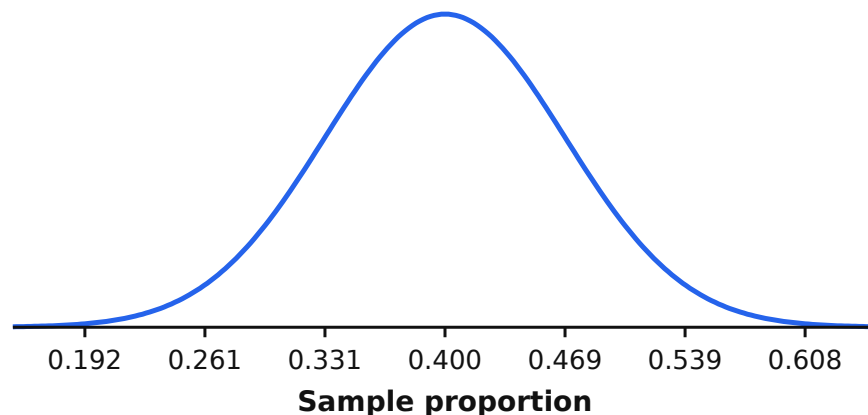
The histogram shows 300 simulated sample proportions from samples of $n=50$ with true $p=0.40$. The center is near 0.40 and the shape is roughly normal.

What is the sampling distribution of a sample proportion?

For a random sample of size n from a population with proportion p :

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

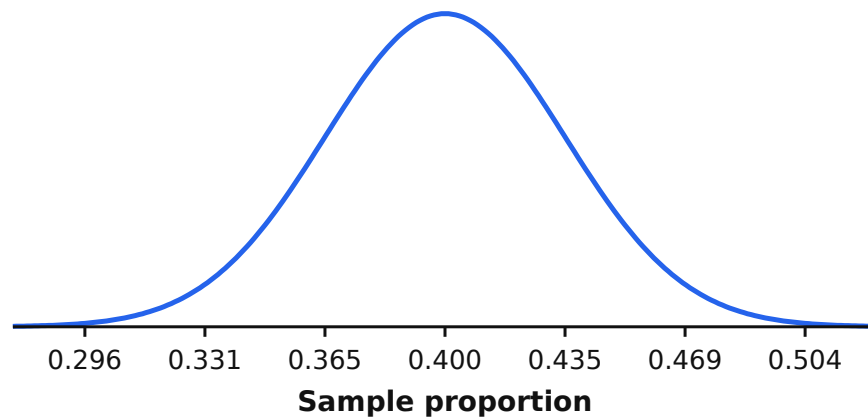
The shape is approximately normal if $np \geq 10$ and $n(1-p) \geq 10$, and the standard deviation formula needs the sample to be less than 10% of the population. Here $\sigma_{\hat{p}} = \sqrt{\frac{0.4(0.6)}{50}} = 0.0693$.



Sampling distribution of $\hat{p}$ for $n = 50$: mean 0.40, standard deviation 0.069.

The statistic $\hat{p}$ is an **unbiased estimator** of p : the center of its sampling distribution equals

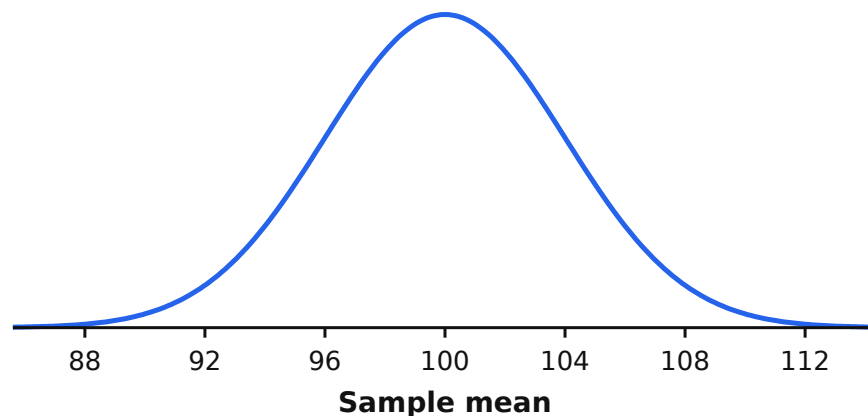
the parameter. Larger samples reduce variability, not bias.



For $n = 200$ the standard deviation is 0.035: quadrupling n halves the spread.

What is the central limit theorem?

For a sample mean $\bar{x}$ from a population with mean μ and standard deviation σ , $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. The **central limit theorem** says that when n is large (usually at least 30), the sampling distribution of $\bar{x}$ is approximately normal, whatever the shape of the population. If the population is itself normal, $\bar{x}$ is normal for any n .



Sampling distribution of $\bar{x}$ for $n = 25$: mean 100, standard deviation 4.

Example: IQ-like scores have $\mu = 100$ and $\sigma = 20$. For $n = 25$, $\sigma_{\bar{x}} = \frac{20}{5} = 4$, and $P(\bar{x} > 106) = P(Z > 1.5) = 0.0668$.

How do differences behave?

Statistic	Mean	Standard deviation	Conditions
p-hat	p	$\sqrt{p(1-p)/n}$	$np \geq 10, n(1-p) \geq 10$
x-bar	μ	$\sigma / \sqrt{n}$	population normal or $n \geq 30$
p1-hat - p2-hat	$p_1 - p_2$	$\sqrt{p_1q_1/n_1 + p_2q_2/n_2}$	both counts ≥ 10
x1-bar - x2-bar	$\mu_1 - \mu_2$	$\sqrt{s_1^2/n_1 + s_2^2/n_2}$	both normal or large n

Summary of sampling distributions.

For independent samples, the means subtract and the variances add. The standard deviation of $\hat{p}_1 - \hat{p}_2$ is $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$, never the difference of the two standard deviations.

Questions about sampling distributions and the central limit theorem

What does the standard deviation of a sampling distribution measure?

It measures how far the statistic typically falls from its mean (the parameter, for an unbiased estimator) across repeated samples. For a sample proportion, it is $\sqrt{p(1-p)/n}$.

What does the central limit theorem say?

It says that the sampling distribution of the sample mean is approximately normal when the sample size is large, even if the population is skewed. A common rule is n of at least 30.

Why does a bigger sample give a narrower sampling distribution?

The standard deviation contains $\sqrt{n}$ in the denominator. Quadrupling the sample size cuts the standard deviation in half.

What is an unbiased estimator?

A statistic is unbiased if the mean of its sampling distribution equals the parameter it estimates. The sample proportion and sample mean are unbiased.

What are the conditions for using the normal model for p-hat?

The sample must be random, the sample size must be less than 10% of the population, and both np and n(1-p) must be at least 10.