Sampling Distribution of $\overline{x}$

AP Statistics · Mr. Merrick · January 16, 2026

Suppose we take a random sample of size n from a population.

- Let $X_1, X_2, \ldots, X_n$ be the sampled values.
- Let μ be the true population mean, and σ the true population standard deviation.
- For a finite population of size N with values $v_1, \ldots, v_N$,

$$\mu = \frac{1}{N}\sum_{i=1}^{N} v_i \qquad \text{and} \qquad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(v_i - \mu)^2}.$$

- The sample mean is

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Because the sample is random, $\overline{x}$ is a **random variable**. To make probability statements about $\overline{x}$, we must decide whether a **normal model** is appropriate. That depends on:

- whether observations can be treated as **independent**, and
- whether the sampling distribution of $\overline{x}$ is **(approximately) normal**.

Normality for $\overline{x}$ can come from either:

- the **population is (approximately) normal** *(then $\overline{x}$ is normal for any n)*, or
- the **Central Limit Theorem (CLT)** *(if $n \geq 30$, then $\overline{x}$ is approximately normal).*

	Independence OK ($n \leq 0.10N$)	**Independence NOT OK** ($n > 0.10N$)
Normality NOT OK (*population not normal & $n < 30$*)	**No reliable normal model** $\overline{x}$ may be skewed. Use simulation/technology or reconsider conditions. (For inference, consider a larger n.)	**No reliable normal model (finite population)** Dependence and non-normality both interfere. Exact/technology methods may be needed (simulation or enumeration for small N).
Normality OK (*population normal OR $n \geq 30$*)	**Normal model for $\overline{x}$** If normality is reasonable and independence holds, then $$\overline{x} \approx \mathcal{N}\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right).$$	**Normal model with finite population correction** When sampling without replacement and the 10% condition fails, $$\overline{x} \approx \mathcal{N}\left(\mu,\ \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right).$$ This is the **FPC** adjustment.

Important inference note (AP Stats): In practice, we usually do *not* know σ. When normality is reasonable, we estimate σ with the sample standard deviation s and use a ***t*-model for inference about μ**. (This handout focuses on the *sampling distribution* of $\overline{x}$.)

1. For each situation below, identify the **most appropriate model/approach** for the sampling distribution of $\overline{x}$. *No probability calculations are required. Justify your choice.*

   (a) A nutritionist takes a random sample of $n = 12$ teenagers from a very large city and records daily sugar intake (grams). The population distribution of sugar intake is strongly right-skewed.

   Independence is reasonable because the city is very large. But $n = 12 < 30$ and the population is strongly skewed, so the sampling distribution of $\overline{x}$ may not be close to normal. Therefore, **no reliable normal model** for $\overline{x}$ is justified; use simulation/technology or increase n.

   (b) A factory claims the mean diameter of ball bearings is $\mu = 8.00$ mm with population standard deviation $\sigma = 0.04$ mm. A random sample of $n = 36$ bearings is taken from a day's production (huge production run).

   Independence is reasonable (tiny sample from a huge production run). Also $n = 36 \geq 30$, so by the CLT, $\overline{x}$ is approximately normal. Use **Normal model for** $\overline{x}$ with mean μ and SD $\sigma/\sqrt{n}$.

   (c) A teacher randomly selects $n = 20$ tests *without replacement* from a stack of $N = 120$ tests to check the average score. The distribution of scores in the stack is approximately normal with known μ and σ.

   Sampling is without replacement, but $20 \leq 0.10(120) = 12$ is *false* (since $20 > 12$), so independence fails. Normality is reasonable because the population is approximately normal. Use the **normal model with finite population correction (FPC)** for the SD: $\sigma/\sqrt{n} \cdot \sqrt{(N-n)/(N-1)}$.

   (d) A researcher samples $n = 25$ people from a population known to be normally distributed for systolic blood pressure.

   If the population is normal, then $\overline{x}$ is **exactly normal** for any n (assuming independence is reasonable). So use the **Normal model for** $\overline{x}$.

   (e) A small company has $N = 40$ employees. A manager selects $n = 18$ employees without replacement and records commute times. The employee commute-time distribution is strongly right-skewed.

   Independence fails because $18 > 0.10(40) = 4$. Also $n = 18 < 30$ with strong skew, so normality is not justified. Therefore, **no reliable normal model**; use simulation/technology or reconsider the plan (increasing n may help normality but does not restore independence).).

2. A brand of granola bars has mean fat content $\mu = 7.5$ grams and population standard deviation $\sigma = 1.2$ grams. A random sample of $n = 64$ bars is selected from a very large shipment.

   (a) Identify the appropriate model for the sampling distribution of $\overline{x}$.

   Check conditions

   Random sampling: The problem states that a random sample of granola bars is selected.

   Independence: The bars are sampled from a very large shipment, so the sample size is less than 10% of the population. Independence is reasonable.

   Normality: The sample size is $n = 64$, which is at least 30. By the Central Limit Theorem, the sampling distribution of $\overline{x}$ is approximately normal.

$$\overline{x} \approx \mathcal{N}\left(7.5,\ \frac{1.2}{\sqrt{64}}\right).$$

   (b) Find $P(\overline{x} > 7.8)$.

$$\overline{x} \approx \mathcal{N}\left(7.5,\ \frac{1.2}{\sqrt{64}}\right), \qquad \mathrm{SD}_{\overline{x}} = 0.15.$$

   Do calculation:

$$z = \frac{7.8 - 7.5}{0.15} = 2.00.$$

$$P(\overline{x} > 7.8) \approx P(Z > 2.00) = 0.0228.$$

3. A town has $N = 500$ households. The mean household water use last month was $\mu = 12.4$ (thousand gallons) with population standard deviation $\sigma = 3.0$ (thousand gallons). A simple random sample of $n = 80$ households is taken without replacement.

(a) Explain why the usual independence condition fails.

Check conditions

Random sampling: The problem states a simple random sample of households is taken.

Independence: Sampling is done without replacement, and

$$n = 80 > 0.10(500) = 50.$$

Because the 10% condition is violated, the observations are not considered independent.

Conclusion: The usual independence condition does not hold.

(b) If the population distribution of water use is roughly normal, find the approximate probability that $\overline{x} < 11.8$. Use the finite population correction.

State distribution:

$$\overline{x} \approx \mathcal{N}\left(12.4,\ \frac{3.0}{\sqrt{80}}\sqrt{\frac{500-80}{500-1}}\right).$$

Do calculation:

$$\mathrm{SD}_{\overline{x}} = \frac{3.0}{\sqrt{80}}\sqrt{\frac{420}{499}} \approx 0.307.$$

$$z = \frac{11.8 - 12.4}{0.307} \approx -1.95.$$

$$P(\overline{x} < 11.8) \approx P(Z < -1.95) = 0.0256.$$

4. A college reports that the mean time students spend studying per week is $\mu = 14$ hours with population standard deviation $\sigma = 6$ hours. A random sample of $n = 30$ students is selected.

(a) Explain why a normal model for $\overline{x}$ is reasonable even if the population distribution is somewhat skewed.

Check conditions

Random sampling: A random sample of students is selected.

Independence: The population of college students is large, so $n = 30$ is less than 10% of the population and independence is reasonable.

Normality: Although the population may be somewhat skewed, the sample size is $n = 30$. By the Central Limit Theorem, the sampling distribution of $\overline{x}$ is approximately normal, assuming no extreme outliers.

Conclusion: A normal model for $\overline{x}$ is reasonable.

(b) Find $P(13 \le \overline{x} \le 16)$.

State distribution:

$$\overline{x} \approx \mathcal{N}\left(14,\ \frac{6}{\sqrt{30}}\right), \qquad \text{SD}_{\overline{x}} \approx 1.095.$$

Do calculation:

$$z_1 = \frac{13 - 14}{1.095} \approx -0.91, \qquad z_2 = \frac{16 - 14}{1.095} \approx 1.83.$$

$$P(13 \le \overline{x} \le 16) \approx P(-0.91 \le Z \le 1.83) = 0.9664 - 0.1814 = 0.785.$$

5. A hospital states that the mean waiting time in the ER is $\mu = 42$ minutes. Assume the population standard deviation is $\sigma = 18$ minutes. A random sample of $n = 50$ patients is selected from a very large time period.

Determine the probability that the sample mean waiting time is less than 38 minutes. Clearly justify your model choice.

Check conditions

Random sampling: A random sample of ER patients is selected.

Independence: The sample is taken from a very large time period, so $n = 50$ is less than 10% of the population and independence is reasonable.

Normality: Since $n = 50 \geq 30$, the Central Limit Theorem applies and the sampling distribution of $\overline{x}$ is approximately normal.

State distribution:

$$\overline{x} \approx \mathcal{N}\left(42,\ \frac{18}{\sqrt{50}}\right), \qquad \mathrm{SD}_{\overline{x}} \approx 2.545.$$

Do calculation:

$$z = \frac{38 - 42}{2.545} \approx -1.57.$$

$$P(\overline{x} < 38) \approx P(Z < -1.57) = 0.058.$$