Sampling Distribution of $\bar{x}_1 - \bar{x}_2$

AP Statistics · Mr. Merrick · January 22, 2026

Suppose we take two random samples from two populations.

- Sample 1 has size n_1, population mean μ_1, population standard deviation σ_1, and sample mean $\bar{x}_1$.
- Sample 2 has size n_2, population mean μ_2, population standard deviation σ_2, and sample mean $\bar{x}_2$.

We are interested in the sampling distribution of the difference

$$\bar{x}_1 - \bar{x}_2,$$

which is a random variable because both samples are **random**.
To determine whether $\bar{x}_1 - \bar{x}_2$ can be modeled using a normal distribution, we must check conditions related to **independence** and **normality / large sample size**.

Independence Condition	**Normality / Large Sample Condition**
• The two samples are independent of each other • Each sample is taken randomly • If sampling without replacement: $n_1 \leq 0.10N_1 \quad \text{and} \quad n_2 \leq 0.10N_2$	• Both populations are normal, **or** • $n_1 \geq 30$ and $n_2 \geq 30$

If all of the above conditions are satisfied, then

$$\bar{x}_1 - \bar{x}_2 \approx \mathcal{N}\left(\mu_1 - \mu_2,\ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

1. For each situation below, determine whether the sampling distribution of $\bar{x}_1 - \bar{x}_2$ can be modeled using the **normal model**:

$$\bar{x}_1 - \bar{x}_2 \approx \mathcal{N}\left(\mu_1 - \mu_2,\ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

Justify your answer by referring to the appropriate conditions.

(a) A researcher compares average commute times in two large cities. A random sample of $n_1 = 45$ commuters is selected from City A and $n_2 = 50$ commuters from City B. The distributions of commute times in both cities are strongly right-skewed.

Solution. The samples are random and independent. Although the population distributions are skewed, both sample sizes are at least 30. Therefore, by the large sample condition, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ can be approximated by a normal distribution.

(b) A school compares test scores between two small classes. A random sample of $n_1 = 12$ students is taken from Class A and $n_2 = 10$ students from Class B. Both populations are approximately normal.

Solution. The samples are random and independent (independence is assumed although population sizes are not given). Even though the sample sizes are small, both population distributions are approximately normal. Therefore, the normal model for $\bar{x}_1 - \bar{x}_2$ is appropriate.

(c) A company compares assembly times for two machines. A random sample of $n_1 = 18$ items from Machine A and $n_2 = 22$ items from Machine B is selected. Both distributions are strongly skewed.

Solution. Although the samples are random and independent, the sample sizes are less than 30 and the population distributions are skewed. Thus, the normal model for $\bar{x}_1 - \bar{x}_2$ is not appropriate.

(d) A teacher compares quiz scores between two classes by sampling $n_1 = 14$ students from a class of $N_1 = 25$ and $n_2 = 16$ students from a class of $N_2 = 30$.

Solution. Because both samples exceed 10% of their respective populations, the independence condition is violated. Therefore, the normal model for $\bar{x}_1 - \bar{x}_2$ is not appropriate.

2. A manufacturer compares the average lifetime of batteries produced by two factories.

- Factory A: $\mu_1 = 1200$ hours, $\sigma_1 = 150$ hours, $n_1 = 64$
- Factory B: $\mu_2 = 1150$ hours, $\sigma_2 = 160$ hours, $n_2 = 81$

Let $\bar{x}_1 - \bar{x}_2$ represent the difference in sample mean lifetimes (Factory A minus Factory B).

(a) Explain why a normal model for $\bar{x}_1 - \bar{x}_2$ is appropriate.

Solution. The samples are random and independent. Both sample sizes are at least 30, so the large sample condition is satisfied. Therefore, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ can be approximated by a normal distribution.

(b) Find the probability that the difference in sample means is greater than 75 hours.

Solution. Using the normal model,

$$\mu = \mu_1 - \mu_2 = 1200 - 1150 = 50,$$

$$\sigma = \sqrt{\frac{150^2}{64} + \frac{160^2}{81}} = \sqrt{351.6 + 316.0} \approx \sqrt{667.6} \approx 25.83.$$

$$z = \frac{75 - 50}{25.83} \approx 0.97.$$

$$P(\bar{x}_1 - \bar{x}_2 > 75) \approx P(Z > 0.97) \approx 0.166.$$

3. A college compares the average number of hours students study per week between freshmen and seniors.

- Freshmen: $\mu_1 = 11$ hours, $\sigma_1 = 3.5$ hours, $n_1 = 100$
- Seniors: $\mu_2 = 13$ hours, $\sigma_2 = 4.0$ hours, $n_2 = 90$

Let $\bar{x}_1 - \bar{x}_2$ represent the difference in sample means (freshmen minus seniors).

(a) Explain why a normal model for $\bar{x}_1 - \bar{x}_2$ is appropriate.

Solution. The samples are random and independent. Both sample sizes exceed 30, so the large sample condition is satisfied. Therefore, a normal model for $\bar{x}_1 - \bar{x}_2$ is appropriate.

(b) Find the probability that the freshmen sample mean is less than or equal to the senior sample mean. That is, find $P(\bar{x}_1 - \bar{x}_2 \leq 0)$.

Solution.

$$\mu = \mu_1 - \mu_2 = 11 - 13 = -2,$$

$$\sigma = \sqrt{\frac{3.5^2}{100} + \frac{4.0^2}{90}} = \sqrt{0.1225 + 0.1778} = \sqrt{0.3003} \approx 0.548.$$

$$z = \frac{0 - (-2)}{0.548} \approx 3.65.$$

$$P(\bar{x}_1 - \bar{x}_2 \leq 0) \approx P(Z \leq 3.65) \approx 0.9999.$$