Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

AP Statistics · Mr. Merrick · January 22, 2026

Suppose we take two random samples from two populations.

- Sample 1 has size n_1 with true proportion p_1 and sample proportion

$$\hat{p}_1 = \frac{X_1}{n_1}.$$

- Sample 2 has size n_2 with true proportion p_2 and sample proportion

$$\hat{p}_2 = \frac{X_2}{n_2}.$$

We are interested in the sampling distribution of the difference

$$\hat{p}_1 - \hat{p}_2,$$

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

Independence Condition	**Normality (Success–Failure) Condition**
• The two samples are independent of each other • Each sample is taken randomly • If sampling without replacement: $n_1 \leq 0.10N_1 \quad \text{and} \quad n_2 \leq 0.10N_2$	• For Sample 1: $n_1p_1 \geq 10 \quad \text{and} \quad n_1(1-p_1) \geq 10$ • For Sample 2: $n_2p_2 \geq 10 \quad \text{and} \quad n_2(1-p_2) \geq 10$

If all of the above conditions are satisfied, then

$$\hat{p}_1 - \hat{p}_2 \approx \mathcal{N}\left(p_1 - p_2,\ \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right).$$

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

$$\hat{p}_1 - \hat{p}_2 \approx \mathcal{N}\left(p_1 - p_2,\ \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right).$$

Justify your answer by referring to the appropriate conditions.

(a) A school compares the proportion of students who play a sport between freshmen and seniors. A random sample of $n_1 = 40$ freshmen and $n_2 = 35$ seniors is taken. The true participation rates are $p_1 = 0.10$ and $p_2 = 0.20$.

Solution. The samples are independent.

However, the normality (success–failure) condition fails.

For freshmen:

$$n_1 p_1 = 40(0.10) = 4 < 10 \quad \text{and} \quad n_1(1-p_1) = 40(0.90) = 36 \geq 10.$$

For seniors:

$$n_2 p_2 = 35(0.20) = 7 < 10 \quad \text{and} \quad n_2(1-p_2) = 35(0.80) = 28 \geq 10.$$

Since the success–failure condition is not met, the normal model for $\hat{p}_1 - \hat{p}_2$ is not appropriate.

(b) A quality-control inspector compares defect rates from two machines. Sample 1 includes $n_1 = 120$ items and Sample 2 includes $n_2 = 150$ items. The true defect rates are $p_1 = 0.18$ and $p_2 = 0.22$.

Solution. The samples are independent and random. The success–failure condition holds for both samples:

$$n_1 p_1 = 21.6, \quad n_1(1-p_1) = 98.4,$$

$$n_2 p_2 = 33, \quad n_2(1-p_2) = 117.$$

Thus, the normal model for $\hat{p}_1 - \hat{p}_2$ is appropriate.

(c) A teacher compares homework completion rates between two classes of sizes $N_1 = 28$ and $N_2 = 30$ by sampling $n_1 = 15$ and $n_2 = 18$ students without replacement.

Solution. Because $n_1 > 0.10N_1$ and $n_2 > 0.10N_2$, the independence condition is violated. Therefore, the normal model for $\hat{p}_1 - \hat{p}_2$ is not appropriate.

2. A news organization compares support for a proposed law in two different states.

- In State A, the true proportion of adults who support the law is $p_1 = 0.52$. A random sample of $n_1 = 200$ adults is selected.
- In State B, the true proportion of adults who support the law is $p_2 = 0.45$. A random sample of $n_2 = 180$ adults is selected.

Let $\hat{p}_1$ and $\hat{p}_2$ be the sample proportions from State A and State B, respectively.

(a) Explain why the sampling distribution of $\hat{p}_1 - \hat{p}_2$ can be approximated by a normal distribution.

Solution. **Random:** The problem states that both samples are random.

Independence: The samples are taken from two different states, so they are independent of each other. Additionally, both sample sizes are far less than 10% of their respective state populations.

Success–failure: For State A,

$$n_1 p_1 = 200(0.52) = 104 \geq 10 \quad \text{and} \quad n_1(1 - p_1) = 200(0.48) = 96 \geq 10.$$

For State B,

$$n_2 p_2 = 180(0.45) = 81 \geq 10 \quad \text{and} \quad n_2(1 - p_2) = 180(0.55) = 99 \geq 10.$$

Since all conditions are satisfied, a normal model is appropriate.

(b) Find the probability that the difference in sample proportions is *greater than* 0.12. That is, find $P(\hat{p}_1 - \hat{p}_2 > 0.12)$.

Solution. Using the normal model,

$$\mu = p_1 - p_2 = 0.52 - 0.45 = 0.07,$$

$$\sigma = \sqrt{\frac{0.52(0.48)}{200} + \frac{0.45(0.55)}{180}} \approx \sqrt{0.002623} \approx 0.0512.$$

$$z = \frac{0.12 - 0.07}{0.0512} \approx 0.98.$$

$$P(\hat{p}_1 - \hat{p}_2 > 0.12) \approx P(Z > 0.98) \approx 0.164.$$

3. A school district compares sleep habits between middle school students and high school students.

- Among middle school students, the true proportion who get at least 8 hours of sleep on school nights is $p_1 = 0.30$. A random sample of $n_1 = 150$ students is selected.
- Among high school students, the true proportion is $p_2 = 0.25$. A random sample of $n_2 = 120$ students is selected.

Let $\hat{p}_1 - \hat{p}_2$ represent the difference in sample proportions (middle school minus high school).

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

Solution. **Random:** Both samples are random.

Independence: The samples come from two different populations, so they are independent. Each sample size is less than 10% of its respective population.

Success–failure:

$$n_1 p_1 = 150(0.30) = 45, \quad n_1(1 - p_1) = 105,$$

$$n_2 p_2 = 120(0.25) = 30, \quad n_2(1 - p_2) = 90.$$

All values are at least 10, so the success–failure condition is met.

(b) Find the probability that the middle school sample proportion is *less than or equal to* the high school sample proportion. That is, find $P(\hat{p}_1 - \hat{p}_2 \leq 0)$.

Solution.

$$\mu = p_1 - p_2 = 0.30 - 0.25 = 0.05,$$

$$\sigma = \sqrt{\frac{0.30(0.70)}{150} + \frac{0.25(0.75)}{120}} \approx \sqrt{0.002963} \approx 0.0545.$$

$$z = \frac{0 - 0.05}{0.0545} \approx -0.92.$$

$$P(\hat{p}_1 - \hat{p}_2 \leq 0) \approx P(Z \leq -0.92) \approx 0.179.$$