CEE 260/MIE 273: Probability and Statistics in Civil Engineering

Lecture 5B: Inference for Difference of Two Proportions

Jimi Oke

UMass Amherst

College of Engineering

November 6, 2025

Outline

Today's objectives

- Normality test for difference of two proportions ($p_1 - p_2$)
- Compute CIs for $p_1 - p_2$
- Conduct hypothesis tests for $p_1 - p_2$
- Using the pooled proportion $\hat{p}_{pooled}$

Difference of two proportions

- Earlier, we considered how to perform inference for a *single* population proportion p, using sample estimates $\hat{p}$ (sample proportion) and n (sample size)
- However, there are cases whereby we want to compare proportions from *two groups/populations*
- In such cases, we perform inference on the **difference** of two population proportions p_1 and p_2
- Thus, the parameter of interest is $p_1 - p_2$, and we define the following sample statistics:
  - $\hat{p}_1$: sample proportion for group 1
  - $\hat{p}_2$: sample proportion for group 2
  - $\hat{p}_1 - \hat{p}_2$: difference two sample proportions
  - n_1: sample size of group 1
  - n_2: sample size of group 2

Normality conditions

The difference of two sample proportions $\hat{p}_1 - \hat{p}_2$ can be assumed to follow a normal distribution if:

- The data are obtained from 2 independent random samples (or from a randomized experiment) **[Independence (extended)]**
- The **success-failure condition** holds for both groups separately, i.e.

$$
\begin{aligned}
n_1 \hat{p}_1 &\geq 10 \qquad (1)\\
n_1(1-\hat{p}_1) &\geq 10 \qquad (2)\\
n_2 \hat{p}_2 &\geq 10 \qquad (3)\\
n_2(1-\hat{p}_2) &\geq 10 \qquad (4)
\end{aligned}
$$

If these conditions hold, then we can use the normal distribution to find appropriate critical values in order to compute CIs and perform hypothesis tests.

Standard error of $\hat{p}_1 - \hat{p}_2$

In order to compute CIs and perform hypothesis tests for a difference of two proportions $p_1 - p_2$, we need to first find the standard error.
If the normality conditions are satisfied, then the **standard error** of $\hat{p}_1 - \hat{p}_2$ is given by:

$$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \tag{5}$$

where:

- p_1: population proportion for group 1
- p_2: population proportion for group 2
- n_1: sample size of group 1
- n_2: sample size of group 2

In cases where we do not know the population proportions (which is typically the case), we can **approximate** the SE using the sample proportions $\hat{p}_1$ and $\hat{p}_2$. Thus:

$$SE_{\hat{p}_1 - \hat{p}_2} \approx \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \tag{6}$$

CI for difference of two proportions

The confidence interval (CI) for a difference of two proportions is given by:

$$\langle p_1 - p_2 \rangle_{(1-\alpha)} = \underbrace{\hat{p}_1 - \hat{p}_2}_{\text{point estimate}} \pm \underbrace{z^* \times SE}_{\text{margin of error}} \tag{7}$$

where:

$$SE \approx \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \tag{8}$$

$$z^* = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad \text{(critical } Z\text{-score)} \tag{9}$$

Also:

- Φ^{-1} is the inverse CDF function of the standard normal distribution
- α is defined as the **level of significance**
- $1 - \alpha$ is the **level of confidence**

Relationship between confidence and significance levels

For example, if 95% is the desired confidence level, then $1 - \alpha = .95$ and $\alpha = 0.05$

Example 1: CPR Study

Construct and interpret a 95% confidence interval for the difference between two groups of patients in a cardiopulmonary resuscitation (CPR) study. The treatment group received a blood thinner, while the control group did not. Outcome variable of interest: proportion of patients who survived for at least 24 hours.

	Survived	**Died**	**Total**
Treatment (t)	14	26	40
Control (c)	11	39	50
Total	25	65	90

Define:

- p_t: survival rate in treatment group
- p_c: survival rate in control group

Example 1: CPR Study (cont.)

First, we compute the difference of the two sample proportions:

$$\hat{p}_t - \hat{p}_c = \frac{14}{40} - \frac{11}{50} = 0.35 - 0.22 = 0.13$$

Next, we compute the SE:

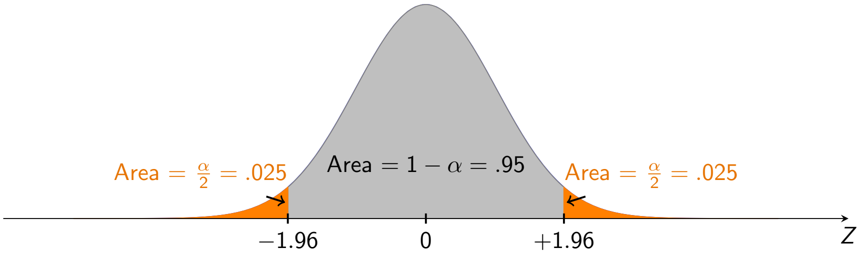
$$\begin{aligned} SE &\approx \sqrt{\frac{\hat{p}_t(1-\hat{p}_t)}{n_t} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_c}} \\ &= \sqrt{\frac{0.35(1-0.35)}{40} + \frac{0.22(1-0.22)}{50}} = 0.095 \end{aligned}$$

Example 1: CPR Study (cont.)

Then we obtain the appropriate critical Z-score z^*.
For a 95% CI, $\alpha = 0.05$. Thus,

$$\begin{aligned} z^* &= \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(1-0.05/2) = \Phi^{-1}(0.975) \\ &= \texttt{norminv}(.975) = 1.96 \end{aligned}$$

Below is the standard normal distribution showing the critical Z-scores corresponding to the desired confidence level of 95%:

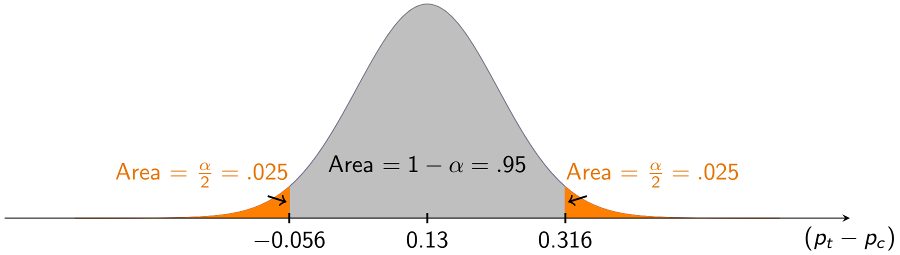


Example 1: CPR Study (cont.)

Thus, the CI is given by:

$$\begin{aligned}\langle p_t - p_c \rangle_{.95} &= 0.13 \pm (1.96 \times 0.095) = 0.13 \pm 0.186 \\ &= (-0.056, 0.316)\end{aligned}$$

Now, we show the corresponding sampling distribution of $p_t - p_c$ with the computed CIs (i.e. the standard normal distribution converted into the actual scale of the point estimate $\hat{p}_1 - \hat{p}_2$)



Example 1: CPR Study (cont.)

Interpretation

Thus, we are 95% confident (*OR* 95% of the time), the difference in survival rate between those who are treated by the blood thinner and those who are not lies between -5.6 and 31.6 percentage points.

Example 2: Political Polling Study

A polling organization surveys registered voters in two different counties to assess support for a proposed infrastructure referendum. Construct and interpret a 99% confidence interval for the difference in support between the two counties.

	Support	**Oppose**	**Total**
County A (a)	312	288	600
County B (b)	275	325	600
Total	587	613	1200

Define:

- p_a: proportion supporting the referendum in County A
- p_b: proportion supporting the referendum in County B

Example 2: Political Polling Study (cont.)

First, we compute the difference of the two sample proportions:

$$\hat{p}_a - \hat{p}_b = \frac{312}{600} - \frac{275}{600} = 0.52 - 0.458 = 0.062$$

Next, we compute the SE:

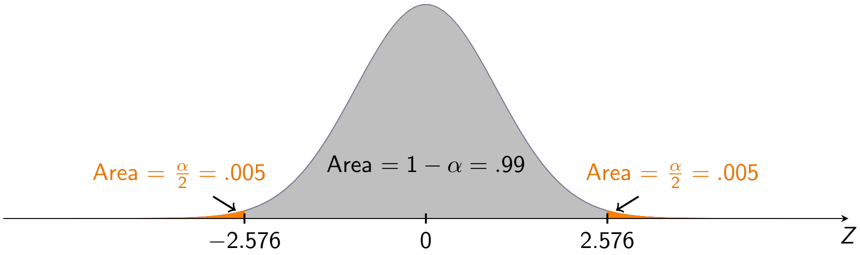
$$\begin{aligned} SE &\approx \sqrt{\frac{\hat{p}_a(1-\hat{p}_a)}{n_a} + \frac{\hat{p}_b(1-\hat{p}_b)}{n_b}} \\ &= \sqrt{\frac{0.52(1-0.52)}{600} + \frac{0.458(1-0.458)}{600}} = 0.0288 \end{aligned}$$

Example 2: Political Polling Study (cont.)

Then we obtain the appropriate critical Z-score z^*.
For a 99% CI, $\alpha = 0.01$. Thus,

$$
\begin{aligned}
z^* &= \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(1-0.01/2) = \Phi^{-1}(0.995) \\
&= \texttt{norminv}(.995) = 2.576
\end{aligned}
$$

Below is the standard normal distribution showing the critical Z-scores corresponding to the desired confidence level of 99%:

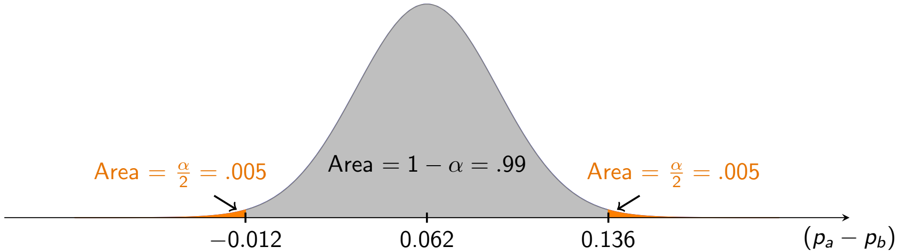


Example 2: Political Polling Study (cont.)

Thus, the CI is given by:

$$\begin{aligned}\langle p_a - p_b \rangle_{.99} &= 0.062 \pm (2.576 \times 0.0288) = 0.062 \pm 0.074 \\ &= (-0.012, 0.136)\end{aligned}$$

Now, we show the corresponding sampling distribution of $p_a - p_b$ with the computed CIs (i.e. the standard normal distribution converted into the actual scale of the point estimate $\hat{p}_1 - \hat{p}_2$)



Example 2: Political Polling Study (cont.)

Interpretation

We are 99% confident that the difference in support for the infrastructure referendum between County A and County B lies between -1.2 and 13.6 percentage points. Since the interval contains zero, we cannot conclude at the 99% confidence level that there is a significant difference in support between the two counties.

Hypothesis testing using critical value

1. State the hypotheses H_0 and H_1
2. Compute the point estimate $\hat{p}_1 - \hat{p}_2$
3. Find the standard error SE
4. Find the test statistic $z = \frac{(\hat{p}_1 - \hat{p}_2) - \Delta_0}{SE}$, where Δ_0 is the null value of the difference
5. Find the critical value(s) z^*
6. Compare the test statistic to the critical value
7. Clearly state the outcome from your hypothesis test
8. Write a final concluding statement in response to the question

Hypothesis testing using p-value

1. State the hypotheses H_0 and H_1
2. Compute the point estimate $\hat{p}_1 - \hat{p}_2$
3. Find the standard error SE
4. Find the p-value
5. Compare the p-value to the level of significance α
6. Clearly state the outcome from your hypothesis test
7. Write a final concluding statement in response to the question

Pooled proportion

If the hypothesis test is to check whether $p_1 = p_2$ or that $p_1 - p_2 = 0$ (null hypothesis), then:

$$\begin{aligned} H_0 : p_1 - p_2 &= \Delta_0 = 0 \qquad (10) \\ H_1 : p_1 - p_2 &\neq 0 \qquad (11) \end{aligned}$$

In this case, we use the **pooled proportion** $\hat{p}_{pooled}$ to verify the success-failure condition and to estimate the standard error

$$\begin{aligned} \hat{p}_{pooled} &= \frac{\text{number of successes}}{\text{number of successes} + \text{number of failures}} \qquad (12) \\ &= \frac{\text{number of successes}}{\text{total number of cases}} \qquad (13) \\ &= \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2} \qquad (14) \end{aligned}$$

Thus:

$$SE \approx \sqrt{\frac{\hat{p}_{pooled}(1 - \hat{p}_{pooled})}{n_1} + \frac{\hat{p}_{pooled}(1 - \hat{p}_{pooled})}{n_2}}$$

Example 3: Concrete quality control study

A large-scale quality control study is conducted over several years with approximately 90,000 concrete samples from two different cement suppliers to determine whether Supplier A produces concrete with a different failure rate than Supplier B.

	Failed quality test?	
	Yes	No
Supplier A (a)	500	44,425
Supplier B (b)	505	44,405

Conduct a hypothesis test to determine whether there is significant evidence to suggest that the two suppliers have different failure rates. Use a significance level $\alpha = 0.05$

Hypotheses

Thus, the null hypothesis H_0 is: $p_a - p_b = 0$ (i.e. the null difference $\Delta_0 = 0$).
And the alternative hypothesis H_1 is: $p_a - p_b \neq 0$.

Example 3 (cont.)

In this case the pooled proportion is given by

$$\begin{aligned}\hat{p}_{pooled} &= \frac{\#\text{ samples that failed quality test in the entire study}}{\#\text{ samples in the entire study}} \\ &= \frac{500 + 505}{500 + 44,425 + 505 + 44,405} \\ &= 0.0112\end{aligned}$$

Thus, the SE is given by:

$$\begin{aligned}SE &\approx \sqrt{\frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_a} + \frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_b}} \\ &= \sqrt{\frac{0.0112(1-0.0112)}{44,925} + \frac{0.0112(1-0.0112)}{44,910}} = 0.0007\end{aligned}$$

Example 3 (cont.): test statistic

Now we find the point estimate $\hat{p}_a - \hat{p}_b$:

$$\hat{p}_a - \hat{p}_b = \frac{500}{44,925} - \frac{505}{44,910} = -0.00012 \tag{15}$$

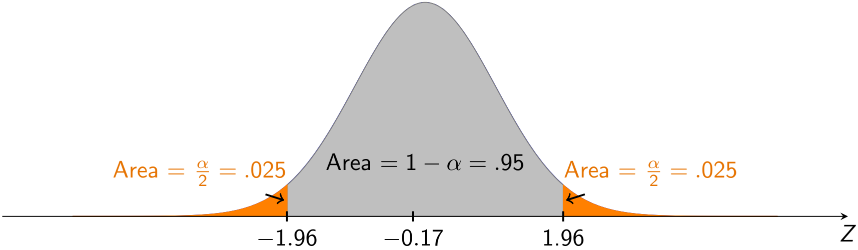
Then, the test statistic is given by:

$$\begin{aligned} z &= \frac{(\hat{p}_a - \hat{p}_b) - \Delta_0}{SE} \\ &= \frac{-0.00012 - 0}{.0007} = -0.17 \end{aligned}$$

Example 3 (cont.): critical value

The critical value z^* in this case is given by:

$$z^* = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(1-.025) = 1.96$$



Since $-0.17 > -1.96$ and $-0.17 < 1.96$ (i.e. the test statistic is contained within the interval $(-1.96, 1.96)$), then we **fail to reject** the null hypothesis H_0. **Thus, we conclude that there is no significant evidence to suggest that the two suppliers have different failure rates for concrete quality**.

Example 3 (cont.)

Next, we will conduct the hypothesis test using p-values (to be continued)

Example 4: Concrete quality control (p-value approach)

Using the same data from Example 3, we now conduct the hypothesis test using the p-value approach.

	Failed quality test?	
	Yes	No
Supplier A (a)	500	44,425
Supplier B (b)	505	44,405

Hypotheses

H_0: $p_a - p_b = 0$ (no difference in failure rates)
H_1: $p_a - p_b \neq 0$ (different failure rates)
Significance level: $\alpha = 0.05$

Example 4 (cont.): Computing the test statistic

From Example 3, we already computed:

$$\begin{aligned}
\hat{p}_{pooled} &= 0.0112 \\
SE &= 0.0007 \\
\hat{p}_a - \hat{p}_b &= -0.00012 \\
z &= \frac{(\hat{p}_a - \hat{p}_b) - \Delta_0}{SE} = \frac{-0.00012 - 0}{0.0007} = -0.17
\end{aligned}$$

Example 4 (cont.): Computing the p-value

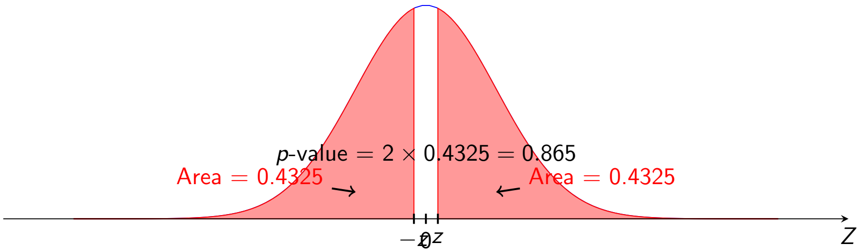
For a two-tailed test with test statistic $z = -0.17$:

$$\begin{aligned} p\text{-value} &= 2 \times P(Z \leq -0.17) \\ &= 2 \times \Phi(-0.17) \\ &= 2 \times 0.4325 \\ &= 0.865 \end{aligned}$$

We multiply by 2 because this is a two-tailed test (alternative hypothesis is $p_a - p_b \neq 0$).

Example 4 (cont.): Visualization of p-value

The p-value represents the total area in both tails beyond $z = \pm 0.17$:



Example 4 (cont.): Decision and conclusion

Decision Rule:

- If p-value $< \alpha$, reject H_0
- If p-value $\geq \alpha$, fail to reject H_0

Our Decision:

- p-value $= 0.865$
- $\alpha = 0.05$
- Since $0.865 > 0.05$, we **fail to reject** H_0

Conclusion

There is no significant evidence to suggest that the two cement suppliers have different failure rates for concrete quality. The observed difference in failure rates could easily be due to random chance.

Note: This conclusion matches our result from Example 3 using the critical value approach!

Confidence Intervals for Difference of Two Proportions

Point Estimate:

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

Standard Error (for CI):

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Critical Value:

$$z^* = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

- For 95% CI: $z^* = 1.96$
- For 99% CI: $z^* = 2.576$

Confidence Interval:

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \times SE$$

Hypothesis Testing for Difference of Two Proportions

Hypotheses:

- H_0: $p_1 - p_2 = \Delta_0$ (usually $\Delta_0 = 0$)
- H_1: $p_1 - p_2 \neq \Delta_0$ (two-tailed) *or* $p_1 - p_2 > \Delta_0$ or $p_1 - p_2 < \Delta_0$ (one-tailed)

Pooled Proportion (when $\Delta_0 = 0$):

$$\hat{p}_{pooled} = \frac{x_1 + x_2}{n_1 + n_2}$$

Standard Error (for hypothesis testing):

$$SE = \sqrt{\frac{\hat{p}_{pooled}(1 - \hat{p}_{pooled})}{n_1} + \frac{\hat{p}_{pooled}(1 - \hat{p}_{pooled})}{n_2}}$$

Test Statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - \Delta_0}{SE}$$

Decision Rules for Hypothesis Testing

Critical Value Approach:

- Two-tailed test: Reject H_0 if $|z| > z^*$ where $z^* = \Phi^{-1}(1 - \alpha/2)$
- Right-tailed test: Reject H_0 if $z > z^*$ where $z^* = \Phi^{-1}(1 - \alpha)$
- Left-tailed test: Reject H_0 if $z < -z^*$ where $z^* = \Phi^{-1}(1 - \alpha)$

p-value Approach:

- Two-tailed: p-value $= 2 \times P(Z \geq |z|) = 2 \times (1 - \Phi(|z|))$
- Right-tailed: p-value $= P(Z \geq z) = 1 - \Phi(z)$
- Left-tailed: p-value $= P(Z \leq z) = \Phi(z)$
- Decision: Reject H_0 if p-value $< \alpha$

Key Difference:

- For CI: Use individual sample proportions $\hat{p}_1$ and $\hat{p}_2$ in SE
- For hypothesis testing (when H_0: $p_1 = p_2$): Use pooled proportion in SE

Python Commands for Critical Values

Import the necessary library:

```
from scipy.stats import norm
```

Computing Critical z-scores:

- For 95% CI (two-tailed, $\alpha = 0.05$):

  ```
  z_star = norm.ppf(1 - 0.05/2)  # 1.96
  ```
- For 99% CI (two-tailed, $\alpha = 0.01$):

  ```
  z_star = norm.ppf(1 - 0.01/2)  # 2.576
  ```
- General formula: `norm.ppf(1 - alpha/2)`

Python Commands for p-values

Computing p-values:

- Two-tailed test (given test statistic z):

  ```
  p_value = 2 * (1 - norm.cdf(abs(z)))
  ```

- Right-tailed test:

  ```
  p_value = 1 - norm.cdf(z)
  ```

- Left-tailed test:

  ```
  p_value = norm.cdf(z)
  ```

Example: For $z = -0.17$ (two-tailed):

```
p_value = 2 * (1 - norm.cdf(abs(-0.17)))
# Result: 0.865
```

Temporary page!

LaTeX was unable to guess the total number of pages correctly. As the
unprocessed data that should have been added to the final page this e
has been added to receive it.
If you rerun the document (without altering it) this surplus page will g
because LaTeX now knows how many pages to expect for this documen