AUTHOR BY:

Sana Ur Rehman Arain

LEARNING JOURNAL 2

University of the People

2024

INSTRUCTOR: ANSON XUAN

MATH 1281-01 Statistical Inference – AY2025-T3

(a) Hypothesis Test for Whether Less Than 50% Cannot Afford College

Context:

A random sample of 441 adults (none with a college degree and not currently enrolled in school) shows that 38% (sample proportion $\hat{p} = 0.38$) said they did not attend college because they could not afford it. We test whether the true proportion (p) is less than 50% (0.50).

1. **State the Hypotheses:**
    - Null Hypothesis (H_0): $p = 0.50$
    - Alternative Hypothesis (H_a): $p < 0.50$
2. **Validate Conditions:**
    - *Independence:*

      • The sample is randomly selected.

      • The sample size (441) is less than 10% of the adult population in question.
    - *Success-Failure Condition (Under H_0):*

      • Expected successes: $n \times p_0 = 441 \times 0.50 = 220.5$

      • Expected failures: $n \times (1 - p_0) = 441 \times 0.50 = 220.5$

      (Both values are greater than 10, so the condition is met.)
3. **Compute the Test Statistic:**

   Use the formula:

   $z = (\hat{p} - p_0) / \text{sqrt}[\, p_0 (1 - p_0) / n]$

Where:

• $\hat{p} = 0.38$

• $p_0 = 0.50$

• $n = 441$

Calculation:

- Compute the standard error (SE):

  SE = sqrt [(0.50 × 0.50) / 441]

  = sqrt (0.25 / 441)

  = sqrt (0.000567)

  ≈ 0.0238

- Calculate the z-value:

  z = (0.38 − 0.50) / 0.0238

  = (−0.12) / 0.0238

  ≈ −5.04

4. **Find the P-value and Conclusion:**

Since the test is one-tailed ($p < 0.50$), the p-value corresponds to the left-tail probability for z = −5.04. This p-value is extremely small (approximately 0), much less than any standard significance level (e.g., $\alpha = 0.05$).

Conclusion:

Reject H_0. There is strong evidence that the true proportion of adults who did not attend college due to cost is less than 50%.

(b) Determining the Required Sample Size for a 90% Confidence Level with a 1.5% Margin of Error

We want a margin of error (ME) of 0.015 using a 90% confidence level.

1. **Margin of Error Formula for a Proportion:**

   $ME = z_{(\alpha/2)} \times \text{sqrt}\,[\,p\,(1 - p) / n]$

   For maximum variability (worst-case scenario), use $p = 0.50$.

2. **Determine the Critical Value:**

   For a 90% confidence level, the critical value $z_{(0.05)} \approx 1.645$.

3. **Set Up the Equation and Solve for n:**

   $0.015 = 1.645 \times \text{sqrt}\,[\,(0.50 \times 0.50) / n]$

   $0.015 = 1.645 \times \text{sqrt}\,(0.25 / n)$

   Isolate the square root:

   $\text{sqrt}\,(0.25 / n) = 0.015 / 1.645$

   ≈ 0.00912

   Square both sides:

   $0.25 / n = (0.00912)^2$

   ≈ 0.0000832

Solve for n:

$n = 0.25 / 0.0000832$

≈ 3005

Recommendation:

A survey should include approximately 3,005 respondents to achieve a 1.5% margin of error at the 90% confidence level.

Part 2 – Comparing Two Proportions

(a) Constructing a 95% Confidence Interval for the Difference in Proportions

Context:

Two groups are compared:

• Texas residents: $n_1 = 13{,}270$ with $\hat{p}_1 = 0.07$

• Dallas residents: $n_2 = 4{,}681$ with $\hat{p}_2 = 0.068$

We want the 95% confidence interval for the difference $(\hat{p}_1 - \hat{p}_2)$.

1. **Validate Conditions:**
    - *Independence:*

      • The samples are random and assumed independent.

      • Each sample size is less than 10% of its respective population.
    - *Success-Failure Condition:*

      For Texas:

• Successes = $n_1 \times \hat{p}_1 = 13{,}270 \times 0.07 \approx 929$

• Failures = $n_1 \times (1 - \hat{p}_1) = 13{,}270 \times 0.93 \approx 12{,}341$

For Dallas:

• Successes = $n_2 \times \hat{p}_2 = 4{,}681 \times 0.068 \approx 318$

• Failures = $n_2 \times (1 - \hat{p}_2) = 4{,}681 \times 0.932 \approx 4{,}363$

All values exceed 10, so the condition is met.

2. **Confidence Interval Formula:**

$(\hat{p}_1 - \hat{p}_2) \pm z_{(\alpha/2)} \times SE$, where

$SE = \text{sqrt} [(\hat{p}_1 (1 - \hat{p}_1) / n_1) + (\hat{p}_2 (1 - \hat{p}_2) / n_2)]$

For 95% confidence, $z_{(0.025)} \approx 1.96$.

3. **Calculations:**

   - **Point Estimate:**

     Difference = $\hat{p}_1 - \hat{p}_2 = 0.07 - 0.068 = 0.002$

   - **Compute Standard Error (SE):**

     For Texas:

     Variance component = $(0.07 \times 0.93) / 13{,}270$

     $\approx 0.0651 / 13{,}270$

     $\approx 4.91 \times 10^{-6}$

For Dallas:

Variance component = $(0.068 \times 0.932) / 4{,}681$

$\approx 0.06338 / 4{,}681$

$\approx 1.35 \times 10^{-5}$

Sum these components:

Total variance $\approx 4.91 \times 10^{-6} + 1.35 \times 10^{-5} \approx 1.84 \times 10^{-5}$

Then,

SE $\approx$ sqrt $(1.84 \times 10^{-5}) \approx 0.00429$

- **Margin of Error (ME):**

  ME = $1.96 \times 0.00429 \approx 0.00841$

- **Confidence Interval:**

  Lower Limit = $0.002 - 0.00841 \approx -0.00641$

  Upper Limit = $0.002 + 0.00841 \approx 0.01041$

  Thus, the 95% confidence interval is approximately $(-0.00641, 0.01041)$.

Interpretation:

We are 95% confident that the true difference in the proportion of sleep-deprived individuals (Texas minus Dallas) lies between –0.64% and 1.04%. Since the interval includes 0, there is no significant difference between the two groups.

Objective:

Test whether there is a statistically significant difference in the proportion of insufficient sleep between Texas and Dallas residents at the $\alpha = 0.05$ level.

1. **State the Hypotheses:**
   - Null Hypothesis (H_0): $p_1 = p_2$ (no difference)
   - Alternative Hypothesis (H_a): $p_1 \neq p_2$ (the proportions differ)
2. **Pooled Proportion Calculation:**

   Since H_0 assumes equal proportions, pool the successes:

   Let $x_1 = n_1 \times \hat{p}_1$ and $x_2 = n_2 \times \hat{p}_2$.

   • $x_1 \approx 13{,}270 \times 0.07 \approx 929$

   • $x_2 \approx 4{,}681 \times 0.068 \approx 318$

   Total successes = 929 + 318 = 1247

   Total sample size = $n_1 + n_2 = 13{,}270 + 4{,}681 = 17{,}951$

   Pooled proportion, $\hat{p}\_pool = 1247 / 17{,}951 \approx 0.06947$

3. **Computing Standard Error Using the Pooled Proportion:**

   $SE\_pool = \text{sqrt}\,[\,\hat{p}\_pool\,(1 - \hat{p}\_pool)\,(1/n_1 + 1/n_2)]$

First, calculate: • $1/n_1 \approx 1 / 13{,}270 \approx 7.53 \times 10^{-5}$

• $1/n_2 \approx 1 / 4{,}681 \approx 0.0002136$

• Sum $\approx 7.53 \times 10^{-5} + 0.0002136 \approx 0.0002889$

Then, SE_pool = sqrt [0.06944 × 0.93056 × 0.0002889]

≈ sqrt (0.06461 × 0.0002889)

≈ sqrt(0.00001867)

≈ 0.00432

4. **Compute the Test Statistic:**

$z = (\hat{p}_1 - \hat{p}_2) / SE\_pool$

$= (0.07 - 0.068) / 0.00432$

$= 0.002 / 0.00432$

≈ 0.46

5. **Determine the P-value and Conclusion:**

This is a two-tailed test. For $z \approx 0.46$, the area in the upper tail is approximately 0.3228.

Therefore, the p-value = $2 \times 0.3228 \approx 0.6456$.

Since $p \approx 0.646 > \alpha$ (0.05), we fail to reject H_0.

Conclusion:

There is insufficient evidence to conclude that the proportion of sleep-deprived residents is different between Texas and Dallas.