Chapter 9-10

Confidence Intervals and Hypothesis Testing

HT and CI for Two Proportions

Statistics

Mehmet Güray Güler, PhD

Last updated 06.08.2020

Tests on Two Proportions

- In many applications, we want to test that the proportion of a certain characteristic in two different populations are equal to each other:
  - H_0: $p_1 = p_2$, vs.
  - A. H_1: $p_1 > p_2$, B. H_1: $p_1 < p_2$, or C. H_1: $p_1 \neq p_2$.
- It would be desirable to have large samples (of sizes n_1 and n_2) from each population so that we can use the Z-statistic:

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

Tests on Two Proportions

- From H_0 we assume that both proportions are equal to each other.
- Therefore
  - p_1-p_2 = 0
  - Recall,we used s_p in the case of comparing two means when standard deviations are equal to each other.
    - We will do the same thing here.
- The new Z will be [corrected]:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}\,(1 - \hat{p}\,)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Tests on Two Proportions

- where $\hat{p} = \frac{X_1+X_2}{n_1+n_2}$ is the pooled estimate of the proportion of successes in the combined sample.
- X_1 = number of "success"es in the first sample, and
- X_2 = number of "success"es in the second sample.

Tests on Two Proportions

- Example 4. I want to test the coffee loyalty of the students in ÖzÜ and YTU and asked whether they favor Cafe Nero over Starbucks or not.
- The poll results are summarized as follows:
  - 120 of 200 ÖzÜ favor Cafe Nero, and
  - 240 of 500 YTÜ favor Cafe Nero.
- Do these data suggest that the proportion ÖzÜ favoring Cafe Nero is higher than the proportion of YTÜ?
- Use an $\alpha = 0.01$ level of significance.
- We will begin by writing the hypotheses of interest and the decision rule.

Tests on Two Proportions

Example 4. $H_0: p_1 = p_2$ vs. $H_1: p_1 > p_2$.

At the level of significance $\alpha = 0.01$, we reject H_0 if $z > 2.33$.

Computations:

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{120}{200} = 0.60 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{240}{500} = 0.48$$

$$\hat{p} = \frac{120 + 240}{200 + 500} = 0.51 \quad \Rightarrow \quad z = \frac{0.60 - 0.48}{\sqrt{(0.51)(0.49)(1/200 + 1/500)}} = 2.9.$$

Tests on Two Proportions

- Example 4. H_0: $p_1 = p_2$ vs. H_1: $p_1 > p_2$.
- **Decision Rule :** Reject H_0 if $Z > Z_{0.05}$=2.33.
- The standardized test statistic is computed as $Z = 2.9 > 2.33$.
- **Decision:** We reject H_0.
- **Conclusion:** The proportion of ÖzÜ favoring Cafe Nero is higher than the proportion of YTÜ.
- P-value = $P(Z > 2.9) \approx 0.002$.

The Confidence Interval

For Single Proportion

For Two Proportions

The Test Statistic

Suppose two populations with (unknown) proportions p_1 and p_2

For example,

- p_1 may be the proportion of smokers with lung cancer
- p_2 may be the proportion of nonsmokers with lung cancer,
  - we want to estimate the difference in between.
- Or, we want to make inference about $p_1 - p_2$.
- How?

The Test Statistic

- To make inference about $p_1 - p_2$,
  - a point estimator is $\hat{P}_1 - \hat{P}_2$,
  - the proportions calculated from the random sample of sizes n1&n2
- use the sampling distribution of $\hat{P}_1 - \hat{P}_2$.
- From the CLT we can assert:
- The sampling distribution of $\hat{P}_1 - \hat{P}_2$ is approximately normal with
  - mean p1 – p2 ,
  - the standard deviation

$$s_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}.$$

The Test Statistic

- Therefore, we can write (approximately)

$$P(-z_{a/2} < Z < z_{a/2}) = 1 - a,$$

$$\text{where } Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}}.$$

- If $\hat{p}_1$ *and* $\hat{p}_2$ are the proportions of “success”es in independent random samples of sizes n1 and n2 , respectively, then an approximate (1− α) 100% confidence interval for $p_1 - p_2$ is given by

$$p_1 - p_2 \in (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}."$$

Example: Comparing Defective Proportions

- Example: A change in a manufacturing process of parts is planned.
- Samples are taken under both the existing and the planned process to determine if the new process reduces defectives.
- **SAMPLE 1:**
  - 90 out of 1500 parts from existing process are defective.
- **SAMPLE 2:**
  - 80 out of 2000 parts from new (planned) process are defective.
- Calculate a 90% CI for the true difference between the proportion of defectives for the two processes.

Example: Comparing Defective Proportions

Solution: First we find a point estimate for the difference:

$$\hat{p}_1 = 90/1500 = 0.06, \hat{p}_2 = 80/2000 = 0.04,$$

$$\Rightarrow \hat{p}_1 - \hat{p}_2 = 0.02.$$

$$p_1 - p_2 \in (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$\Rightarrow p_1 - p_2 \in 0.02 \pm 1.645\sqrt{\frac{(0.06)(0.94)}{1500} + \frac{(0.04)(0.96)}{2000}}$$

$$0.008 < p_1 - p_2 < 0.032.$$

- We can conclude at the 90% confidence level that the planned process seems to be improving the proportion of defective parts.
- (WHY?)