

Lecture 31

Inference for Proportions II

STAT 8010 Statistical Methods I November 8, 2019

> Whitney Huang Clemson University

Last Time: Inference for *p*



Point estimate:

$$\hat{p} = \frac{X}{n}$$

where X is the number of "successes" in the sample with sample size n, and the probability of success, p, is the parameter of interest

• $100(1-\alpha)\%$ Wald CI (when \hat{p} is not too close to 0 or 1):

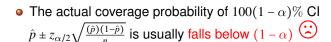
$$\hat{p}\pm z_{\alpha/2}\sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}$$

• Hypothesis Testing: $H_0: p = p_0$ vs. $H_a: p > \text{ or } \neq \text{ or } < p_0$

$$z_{obs} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Another CI for p: Wilson Score Confidence Interval





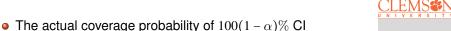
• E.B. Wilson proposed one solution in 1927 **Idea**: Solving $\frac{p-\hat{p}}{\sqrt{\frac{p(1-p)}{n}}} = \pm z_{\alpha/2}$ for p

$$\Rightarrow (p - \hat{p})^2 = z_{\alpha/2}^2 \frac{p(1-p)}{n}$$



Another CI for p: Wilson Score Confidence Interval





- $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}$ is usually falls below $(1-\alpha)$
- E.B. Wilson proposed one solution in 1927 **Idea**: Solving $\frac{p-\hat{p}}{\sqrt{\frac{p(1-p)}{n}}} = \pm Z_{\alpha/2}$ for p

$$\Rightarrow (p - \hat{p})^2 = z_{\alpha/2}^2 \frac{p(1-p)}{n}$$

 $100(1-\alpha)\%$ Wilson Score Confidence Interval:

$$\frac{X + \frac{z_{\alpha/2}^2}{2}}{n + z_{\alpha/2}^2} \pm \frac{z_{\alpha/2}}{n + z_{\alpha/2}^2} \sqrt{\frac{X(n - X)}{n} + \frac{z_{\alpha/2}^2}{4}}$$

Example



Suppose we would like to estimate p, the probability of being vegetarian (for all the CU student). We take a sample with sample size n=20 and none of them are vegetarian. Construct a 95% CI for p.





When $\hat{p} = 0$, we have

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{(\hat{p})(1-\hat{p})}{n}} = 0 \pm z_{\alpha/2} \times 0 = (0,0)$$

Similarly, when $\hat{p} = 1$, we have

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{(\hat{p})(1-\hat{p})}{n}} = 1 \pm z_{\alpha/2} \times 0 = (1,1)$$

These Wald CIs degenerate to a point , which do not reflect the estimation uncertainty. Here we could apply the rule of three to approximate 95% CI:

$$(0,3/n),$$
 if $\hat{p} = 0$
 $(1-3/n,1),$ if $\hat{p} = 1$

Comparing Two Population Proportions p_1 and p_2



- We often interested in comparing two groups, e.g., does a particular treatment increase the survival probability for cancer patients?
- We would like to infer p₁ p₂, the difference between two population proportions ⇒ point estimate, interval estimate, hypothesis testing

Notation



Parameters

- p_1, p_2 : population proportions
- $p_1 p_2$: the difference between two population proportions

Sample Statistics

- n_1, n_2 : sample sizes
- $\hat{p}_1 = \frac{X_1}{n_1}, \hat{p}_2 = \frac{X_2}{n_2}$: sample proportions

Point/Interval Estimation for $p_1 - p_2$





Point estimate:

$$\hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

• $100(1-\alpha)\%$ CI based on CLT:

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

Hypothesis Testing for $p_1 - p_2$



State the null and alternative hypotheses:

$$H_0: p_1 - p_2 = 0$$
 vs. $H_a: p_1 - p_2 > \text{ or } \neq \text{ or } < 0$

Compute the test statistic:

$$z_{obs} = \frac{\hat{p} - p_0}{\sqrt{\frac{\bar{p}(1 - \bar{p})}{n_1} + \frac{\bar{p}(1 - \bar{p})}{n_2}}},$$

where
$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

Make the decision of the test:

Rejection Region/ P-Value Methods

Oraw the conclusion of the test:

We (do/do not) have enough statistical evidence to conclude that (H_a in words) at α % significant level.

Example



A Simple Random Simple of 100 CU graduate students is taken and it is found that 79 "strongly agree" that they would recommend their current graduate program. A Simple Random Simple of 85 USC graduate students is taken and it is found that 52 "strongly agree" that they would recommend their current graduate program. At 5 % level, can we conclude that the proportion of "strongly agree" is higher at CU?