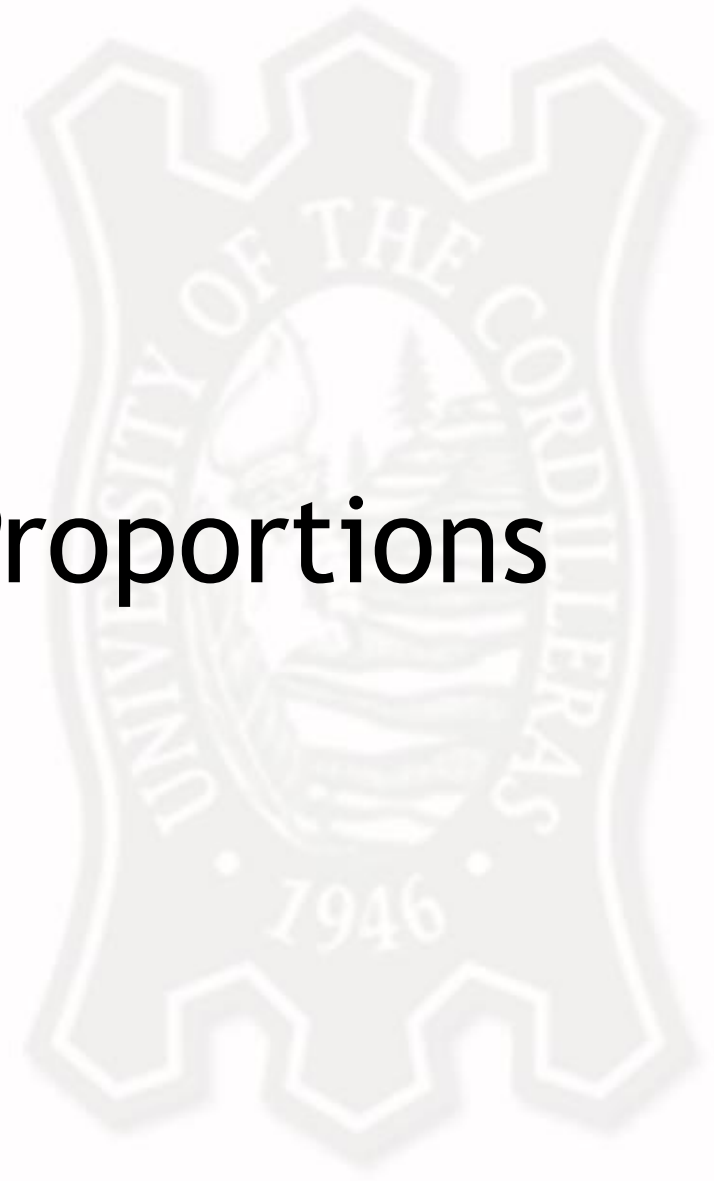


Comparing Two Proportions



Categorical Response Variable

- For a *categorical response variable*, inferences compare groups in terms of their population proportions in a particular category.
- Let p_1 represent the population proportion for the first group and p_2 the population proportion for the second group.
- We can compare the groups by their difference, $(p_1 - p_2)$. This is estimated by the difference of the sample proportions, $(\hat{p}_1 - \hat{p}_2)$.



Step 1: Assumptions

- Check assumptions
 - Population proportions are defined for each of the two groups
 - n_1 and n_2 are large enough, $(n_1 + n_2 > 30)$



Example:

Table 10.1 Whether Subject Died of Cancer, for Placebo and Aspirin Treatment Groups

Group	Death from Cancer		Total
	Yes	No	
Placebo	347	11,188	11,535
Aspirin	327	13,708	14,035

Is there a significant difference between the two groups?
Use a 95% confidence level.

Step 1: Assumptions

- Both sample sizes are large enough



Step 2: State the Hypotheses

Null Hypothesis

$$H_0: p_1 = p_2 \text{ (} p_1 - p_2 = 0 \text{)}$$

Alternative Hypothesis

$$H_a: p_1 \neq p_2$$



Step 3: Compute the Test Statistic

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{se_0}$$

where $se_0 = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ and

$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$, the pooled estimate



Step 3: Compute the Test Statistic

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{347 + 327}{11535 + 14035} = 0.0264$$

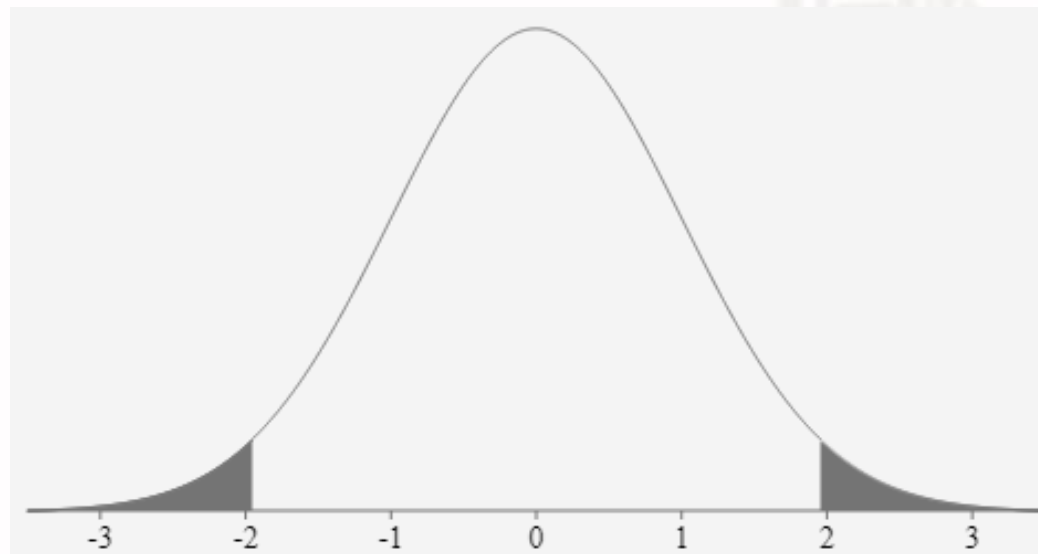
$$\begin{aligned} se_0 &= \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \\ &= \sqrt{0.0264(1 - 0.0264)\left(\frac{1}{11535} + \frac{1}{14035}\right)} = 0.002013 \end{aligned}$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{se_0} = \frac{(347/11535 - 327/14035) - 0}{0.002013} = 3.37$$



Step 4: Interpret the Test Statistic (Using Rejection Region)

- 95% confidence level, $\alpha = 0.05$
- $z_c = \pm 1.96$



Step 4: Interpret the Test Statistic (Using p-value)

- 95% confidence level, $\alpha = 0.05$
- $z = 3.37$
- $p\text{-value} = 2(0.5 - .4996) = 2(0.0004) = 0.0008$



Step 5: Make a Conclusion

- Since the test statistic lies in the RR or
- since the p-value is less than α ,
- then we reject the null hypothesis.
- Therefore, there is a difference between the proportions of the two groups.

