Chapter 21

COMPARING TWO PROPORTIONS

1

THE STANDARD DEVIATION OF THE DIFFERENCE BETWEEN TWO PROPORTIONS

- The standard deviation of the difference between two sample proportions is

$$SD(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

- Thus, the standard error is

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

2

ASSUMPTIONS AND CONDITIONS

- Independence Assumptions:
  - Randomization Condition: The data in each group should be drawn independently and at random from a homogeneous population or generated by a randomized comparative experiment.
  - The 10% Condition: If the data are sampled without replacement, the sample should not exceed 10% of the population.
  - Independent Groups Assumption: The two groups we're comparing must be independent *of each other*.

3

ASSUMPTIONS AND CONDITIONS (CONT.)

- Sample Size Condition:
  - *Each* of the groups must be big enough…
  - Success/Failure Condition: Both groups are big enough that at least 10 successes and at least 10 failures have been observed in each.

4

THE SAMPLING DISTRIBUTION

- Provided that the sampled values are independent, the samples are independent, and the samples sizes are large enough, the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is modeled by a Normal model with
  - Mean: $\mu = p_1 - p_2$
  - Standard deviation:

$$SD(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

5

TWO-PROPORTION *Z*-INTERVAL

- When the conditions are met, we are ready to find the confidence interval for the difference of two proportions:
- The confidence interval is

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

where

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

- The critical value z^* depends on the particular confidence level, *C*, that you specify.

6

TWO-PROPORTION Z-TEST

- The conditions for the two-proportion *z*-test are the same as for the two-proportion *z*-interval.
- We are testing the hypothesis H_0: $p_1 = p_2$.
- Because we hypothesize that the proportions are equal, we pool them to find

$$\hat{p}_{pooled} = \frac{Success_1 + Success_2}{n_1 + n_2}$$

7

TWO-PROPORTION Z-TEST

- We use the pooled value to estimate the standard error:

$$SE_{pooled}\left(\hat{p}_1 - \hat{p}_2\right) = \sqrt{\frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_1} + \frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_2}}$$

- Now we find the test statistic:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE_{pooled}\left(\hat{p}_1 - \hat{p}_2\right)}$$

- When the conditions are met and the null hypothesis is true, this statistic follows the standard Normal model, so we can use that model to obtain a P-value.

8