STOR 155 Introductory Statistics

Lecture 19: Comparing two proportions

Section 8.2
Two populations: an extension of Lecture 18

- Population 1: with proportion $p_1$
- Population 2: with proportion $p_2$
- Interested in the difference $p_1 - p_2$

- Sample 1: size $n_1$ count $X_1$ proportion $\hat{p}_1 = X_1/n_1$
- Sample 2: size $n_2$ count $X_2$ proportion $\hat{p}_2 = X_2/n_2$

- Consider the difference $D = \hat{p}_1 - \hat{p}_2$

- Assume the two samples are independent, and both $n_1$ and $n_2$ are large.
Useful probability facts

The random variable $D$ has approximately a normal distribution with mean $p_1 - p_2$ and standard deviation

$$\text{SD}(D) = \sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}$$

An estimate of $\text{SD}(D)$:

$$SE_D = \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2}$$
Confidence Interval for $p_1 - p_2$

- Expression: $[D - m, D + m]$

  where the margin of error $m = z^* SE_D$

  – Confidence level $C$ determines $z^*$
Hypothesis testing for $p_1 - p_2$

- We want to test $H_0 : p_1 = p_2$ versus some (1-sided or 2-sided) alternative. Recall the 4 steps …

- Step 1: need to specify the alternative $H_a$
Hypothesis testing for $p_1 - p_2$ (continued)

- **Step 2**: Test statistic $z = \frac{D}{SE_{D_p}}$ where

$$SE_{D_p} = \sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}$$

and $\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$

- **Note**: The pooled standard error $SE_{D_p}$ is different from $SE_D$ on page 3
Hypothesis testing (continued)

• Step 3: The $P$-value will be equal to
  $P(Z > z)$ for 1-sided (upper tail) $H_a : p_1 > p_2$
  $P(Z < z)$ for 1-sided (lower tail) $H_a : p_1 < p_2$
  $2 \ P(Z > |z|)$ for 2-sided $H_a : p_1 \neq p_2$

• Step 4: Compare the $P$-value with the significance level $\alpha$ and draw your conclusion.
Gender difference in frequent binge drinking?

Proportion of frequent binge drinkers

- Population 1: (male college students) $p_1$
- Population 2: (female college students) $p_2$

- Sample 1: $n_1 = 7180$, $X_1 = 1630$, $\hat{p}_1 = 0.227$
- Sample 2: $n_2 = 9916$, $X_2 = 1684$, $\hat{p}_2 = 0.170$

- Total: $n_1 + n_2 = 17096$, $X_1 + X_2 = 3314$, $\hat{p} = 0.194$
Gender difference? (continued)

- Test $H_0 : p_1 = p_2$ vs $H_a : p_1 > p_2$

- Test statistic:

  $z = (0.227 - 0.170) \div \sqrt{(0.194)(0.806)(1/7180 + 1/9916)} = 9.34$

- $P$-value = $P(Z > 9.34) = 0.00$
- Reject $H_0$

- 95% CI for $p_1 - p_2$ is (0.045, 0.069), where

  $SE_D = \sqrt{(0.227)(0.773)/7180 + (0.170)(0.830)/9916} = 0.00622; 
  m = z^*SE_D = (1.96)(0.00622) = 0.012$
Take Home Message

- CI for the difference $p_1 - p_2$
- Hypothesis testing for comparing $p_1$ and $p_2$ … 4 steps
- Note: different standard errors are used --- $SE_D$ in CI

$SE_{Dp}$ in testing