The sampling distribution for $\hat{p}_1 - \hat{p}_2$

Assumptions: A SRS of size n_1 is selected from a population with population proportion p_1, and another SRS of size n_2 is independently selected from a large population with population proportion p_2. Both samples satisfy the 10% Condition and the Success/Failure Condition. The plan is to compare the proportions for the two populations.

- the mean of the sampling distribution of differences $\hat{p}_1 - \hat{p}_2$ is $p_1 - p_2$
- the variance of the sampling distribution of differences is the *sum* of the variances of the individual proportion variables; therefore, …
- the standard deviation of the sampling distribution of differences is

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

- since the population parameters are typically unknown, we must estimate the standard deviation by the **standard error**

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

Two-proportion z-interval

A level C confidence interval for $p_1 - p_2$ is

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

where z^* is the upper $\frac{1-C}{2}$ normal critical value.

[TI-83: `STAT TESTS 2-PropZInt…`]

- A **plus-four** method can be used here, adding one success and one failure to each of the two samples.

- **pooled proportion**
  If we have reason to believe that both samples come from the *same* population, these samples can be combined to compute a pooled proportion

$$\hat{p}_{\text{pooled}} = \frac{\text{(combined count of successes in both samples)}}{n_1 + n_2},$$

in which case

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

Pooled statistics are based on more data, so are more reliable.

The 2-Proportion z-test

- **State hypotheses:**
  Null hypothesis: H_0: $p_1 = p_2$
  Alternative hypothesis:
  H_A: $p_1 > p_2$, or $p_1 < p_2$, or $p_1 \neq p_2$

- **Choose model:**
  Two SRS are independently selected from populations satisfying the 10% Condition and the Success/ Failure Condition, so normal model applies to sampling distribution for $\hat{p}_1 - \hat{p}_2$

- **Mechanics:**
  Compute z-statistic based on H_0:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE(\hat{p}_1 - \hat{p}_2)}$$

  Probability associated with appropriate H_A:
  $P = P(Z \geq z)$, or $P = P(Z \leq z)$, or $P = 2P(Z \geq z)$

- **Conclusion**:
  Assess evidence against H_0 in favor of H_A depending on how small P is.

  [TI-83: `STAT TESTS 2-PropZTest…`]