Chapter 21: Comparing Two Proportions: Key Ideas

- **Sampling distribution of the difference in sample proportions $\hat{p}_1 - \hat{p}_2$**

  Let $\hat{p}_1$ and $\hat{p}_2$ be the *sample proportions* of successes from two populations. Then the sampling distribution of $\hat{p}_1 - \hat{p}_2$:

  - Is <u>approximately normal</u>,
  - Has <u>mean</u> $p_1 - p_2$ (the difference in the population proportions of successes)
  - Has <u>standard deviation</u>

$$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

  - **Assumptions**:
    - The data are from an SRSs from the corresponding populations.
    - The populations are at least 10 times as large as the corresponding samples.
    - The sample sizes are large.

➔ N.B.: Specific guidelines for significance tests and confidence intervals are given below.

Testing the equality of proportions from two samples: *z* test for $H_0: p_1 = p_2$

1. **State** the problem in context

2. **Formulate**

   a. Define the parameters of interest

   b. State H_0 and H_a.

3. **Solve**

   a. Check that the conditions are satisfied (see below).

   b. Compute two-sample *z* statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\hat{p}$ is the *pooled sample proportion* defined as

$$\hat{p} = \frac{\text{count of successes in both samples combined}}{\text{count of observations in both samples combined}}.$$

4. **Conclude** in the context of the **stated** problem.

Note: This test can be carried out by hand calculation using the Rossman-Chance applet for critical normal values. Be sure to specific the alternative hypothesis correctly.

Conditions for use of the *z* significance test for comparing two proportions:

- The data are from two independent SRSs.
- The population is at least 10 times as large as the sample.
- The counts of success and failures are each 5 or more in both samples.

Plus four *z* confidence interval for $p_1 - p_2$

1. **State** the parameter estimation problem in context
2. **Formulate**
   a. Define the parameter of interest
   b. State the level of confidence to be used

3. **Solve**

   a. Check that the conditions are satisfied (see guidelines below).

   b. Calculate the level C confidence interval for $p_1 - p_2$ using the plus-four interval:

$$(\tilde{p}_1 - \tilde{p}_2) \pm z^* \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}}$$

   where

   - $\tilde{n}_1 = n_1 + 2$ and $\tilde{n}_2 = n_2 + 2$
   - $\tilde{p}_1$ and $\tilde{p}_2$ represent the **plus four estimate** of p_1 and p_2 respectively:

$$\tilde{p}_1 = \frac{\text{count of successes in first sample } + 1}{\tilde{n}_1},$$

$$\tilde{p}_2 = \frac{\text{count of successes in second sample } + 1}{\tilde{n}_2}.$$

   - z^* is the upper (1-C)/2 critical value for the Z distribution (using technology).

4. **Conclude** in the context of the **stated** problem.

Note: This interval can be computed easily by hand (using a Normal applet to find z^*)

Conditions for use of the *z* plus-four confidence interval for the difference in two proportions:

- The data are from two independent SRSs.
- The population is at least 10 times as large as the sample.
- $n_1 \geq 5$ and $n_2 \geq 5$.