Chapter 20: Inference for Proportions:
Key Ideas

- **Sampling distribution of a sample proportion $\hat{p}$**
  - Let $\hat{p}$ be the *sample proportion* of successes,

    $$\hat{p} = \frac{\text{count of success in the sample}}{n}$$

    Then, the sampling distribution of $\hat{p}$:
    - Is approximately normal,
    - Has mean p (the population proportion of successes)
    - Has standard deviation $\sqrt{\frac{p(1-p)}{n}}$
  - Assumptions:
    - The data are from an SRS from the population.
    - The population is at least 10 times as large as the sample.
    - The sample size is large enough so that both np and $n(1-p)$ are at least 10.

z test for $H_0 : p = p_0$:

1. **State** the problem in context
2. **Formulate**
   a. Define the parameter of interest
   b. State H_0 and H_a.
3. **Solve**
   a. Check that the conditions are satisfied (see below).
   b. Calculate the one-sample z statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

   c. Find the p-value using the Z distribution (using technology).
      - $H_a : p > p_0 \Rightarrow p-\text{value} = P(Z \geq z)$
      - $H_a : p < p_0 \Rightarrow p-\text{value} = P(Z \leq z)$
      - $H_a : p \neq p_0 \Rightarrow p-\text{value} = 2P(Z \geq |z|)$
4. **Conclude** in the context of the **stated** problem.

- **Conditions for use of the *z* significance test for a proportion:**
  - The data are from an SRS.
  - The population is at least 10 times as large as the sample.
  - The sample size n is large enough so that both np_0 and $n(1-p_0)$ are at least 10.

Note: This test can be carried out by hand calculation using the Rossman-Chance applet for critical normal values.

Plus Four z confidence interval for p:

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 using the plus-four interval:

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

where

- $\tilde{n} = n + 4$ and
- $\tilde{p}$ represents the **plus four estimate** of p:

$$\tilde{p} = \frac{\text{count of successes in the sample } + 2}{\tilde{n}}$$

- 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 carried out by hand calculation using the Rossman-Chance applet for critical normal values.

- **Conditions for use of the *plus four confidence interval* for a proportion**
  - The data are from an SRS.
  - The population is at least 10 times as large as the sample.
  - C is at least 90%
  - The sample size $n \geq 10$.

- **Sample size for a desired margin of error**
  - The level *C* confidence interval for a population proportion *p* will have margin of error approximately equal to a specified value *m* when the sample size is

$$n = \left(\frac{z^*}{m} \right)^2 p^*(1 - p^*)$$

where p^* is either

    - A guessed value for the sample proportion (if one is available) OR
    - $p^* = 0.5$, if no other guess is available (Note: Setting $p^* = 0.5$ will ensure that *n* is large enough no matter what *p* turns out to be.)