Comparing Two Population Proportions

March 4, 2010

In this worksheet, we will test whether or not the proportions of blue M&M's and red M&M's in a random package are the same.

Before beginning, do you think they will be the same?

1. You have been given 10 bags of M&M's. Separate these into two piles of 5 bags each. One of these piles will be your sample for red M&M's, and the other will be your sample for blue M&M's.

2. Open the five bags in one of your piles (it doesn't matter which one). Count the total number of M&M's, this is n_1. Count the total number of red M&M's and compute the proportion. This is $\widehat{p}_1$.

$$n_1 =$$

$$\widehat{p}_1 =$$

3. Open the five bags in the other pile. Count the total number of M&M's, this is n_2. Count the total number of blue M&M's and compute the proportion. This is $\widehat{p}_2$.

$$n_2 =$$

$$\widehat{p}_2 =$$

4. Formulate (mathematically) the null hypothesis that these proportions are the same.

$$H_0 :$$

5. Formulate the alternative hypotheses. Is it one or two sided?

$$H_a :$$

6. In order to proceed, we compute a new statistic, called $\widehat{p}_C$. This is the total proportion of successes:

$$\widehat{p}_C = \frac{\text{total number of successes in both samples}}{\text{combined size of both samples}}.$$

Compute this statistic for your data.

$$\widehat{p}_C =$$

7. We need to check that the conditions for inference are satisfied. Confirm as follows:

   (a) **SRS**: Does our data come from simple random samples?

   (b) **Normality**: Compute $n_1\widehat{p}_C$, $n_2\widehat{p}_C$, $n_1(1-\widehat{p}_C)$, and $n_2(1-\widehat{p}_C)$. Make sure these are each greater than or equal to ten.

   (c) **Independence**: Is the population of M&M's greater than 10 times n_1 and n_2?

8. The standard error in this case is:

$$\mathrm{SE}_{\widehat{p}_1-\widehat{p}_2} = \sqrt{\widehat{p}_C(1-\widehat{p}_C)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}.$$

   Compute the standard error for your data.

$$\mathrm{SE}_{\widehat{p}_1-\widehat{p}_2} =$$

9. The test statistic is

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

   As usual, this is our estimate of the parameter of interest divided by the standard error. Compute the test statistic for your data:

$$z =$$

10. Our test statistic has the standard normal distribution, so we can use the z-score. Find the P-value for our test statistic.

$$P =$$

11. Is our result significant at the $\alpha = 0.05$ level? How about the $\alpha = 0.01$ level? If we use the standard $\alpha = 0.05$ level, what do you conclude?