10.1 COMPARING TWO PROPORTIONS

TWO POPULATION PROPORTIONS

M&Ms plain chocolate candies and M&Ms peanut chocolate candies are delicious. Because blue is his favorite color, Mr. Tyson likes to eat the blue candies first. Through extensive experience, Mr. Tyson thinks that there is a different proportion of blue candies among M&Ms plain than M&Ms peanut. For the sake of statistics and scientific inquiry, let's estimate the difference in the proportion of blue candies in these populations.

1. Is color a categorical or quantitative variable?

2. Mr. Tyson will provide a large bag of M&Ms plain and a large bag of M&Ms peanut. As a class, divide up the bag record the count of each color in the table below.

Color	**Plain M&Ms**	**Peanut M&Ms**
Blue		
Orange		
Green		
Yellow		
Red		
Brown		
Total		

3. Find the proportion of blue in each sample. Then compute the difference in the proportion of blue candies (plain – peanut). Subscripts are often easier to write as numbers, so we'll let 1 represent plain and 2 represent peanut.

$\hat{p}_1 =$ $\hat{p}_2 =$ $\hat{p}_1 - \hat{p}_2 =$

4. Do you expect this difference to be the exact difference in the population proportions? Why or why not?

5. Use the following formula to find the standard error of the difference in proportions. Interpret this value in context. Notice the difference between this formula and the one on the formula sheet.

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

TWO-SAMPLE Z INTERVAL FOR $P_1 - P_2$

6. Create and interpret a 95% confidence interval for the true difference in the population parameters.

When the appropriate conditions have been met, a level C confidence interval for $p_1 - p_2$ is given by

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where z^* is the value in a standard Normal distribution such that an area C lies between $-z^*$ and z^*.

Conditions:

1. **Random**. The data come from **two independent random samples** of the respective populations or from **two randomly assigned treatment groups** in an experiment.
    - **10%**: When sampling without replacement, the sample sizes are less than 10% of their respective population sizes.
2. **Normal**. Both sampling distributions approximately Normal. This occurs when $n_1\hat{p}_1 \geq 10$, $n_1(1-\hat{p}_1) \geq 10$, $n_2\hat{p}_2 \geq 10$, and $n_2(1-\hat{p}_2) \geq 10$.

TWO-SAMPLE Z TEST FOR $P_1 - P_2$

As you might expect, we can also test a hypothesis for the difference in population proportions, but the test statistic for a test of significance has one small difference from what you might expect. Since the null hypothesis is that the two population proportions are equal, we combine (or pool) the two samples to get a better estimate of what the standard error for the difference in proportions might be.

7. Compute the pooled (combined) proportion of blue candies for our sample using this formula: $\hat{p}_C = \dfrac{\text{count of successes in both samples}}{\text{count of observations in both samples}} = \dfrac{X_1 + X_2}{n_1 + n_2}$

8. Compute the pooled (combined) standard error: $SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\dfrac{\hat{p}_C(1-\hat{p}_C)}{n_1} + \dfrac{\hat{p}_C(1-\hat{p}_C)}{n_2}}$.

9. Conduct a hypothesis test to see if there is *convincing* evidence that there is a difference in the proportion of blue candies between these two populations. The formula for the test statistic is $z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (0)}{\sqrt{\dfrac{\hat{p}_C(1-\hat{p}_C)}{n_1} + \dfrac{\hat{p}_C(1-\hat{p}_C)}{n_2}}}$.