Warm Up (10. 1 Day 3)

1. **Think about the conditions that need to be met in order to create confidence intervals for proportions. Match the condition on the left with its purpose on the right.**

10% Condition	So we can generalize to both populations or, in an experiment, we can show causation.
Large Counts	So sampling without replacement is OK. If the condition is met, we can use the formulas for the standard devtation.
Random Condition	So that the sampling distribution of $\hat{p}_1 - \hat{p}_2$ will be approximately Normal and we can uthen use z^* to do calculations

2. Now read the following for background for today's lesson about significance tests

Large Counts Condition for Testing a Difference in Proportions

For use in the "Stating Hypotheses and Checking Conditions" ~~subsection on page 630 (6e)~~.

The Random and 10% conditions for performing a significance test about $p_1 - p_2$ are the same as for constructing a confidence interval. However, we check the Large Counts condition differently when performing a test of $H_0: p_1 - p_2 = 0$.

A significance test begins by assuming that the null hypothesis is true. In that case, $p_1 - p_2$. We call the common value of these two parameters p. Unfortunately, we don't know this common value. To estimate p, we combine (or "pool") the data from the two samples as if they came from one larger sample. This combined sample proportion is

$$\hat{p}_C = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}} = \frac{X_1 + X_2}{n_1 + n_2}$$

In other words, $\hat{p}_C$ gives the overall proportion of successes in the combined samples.

Let's return to the Hungry Children example. The two-way table summarizes the sample data on whether or not children missed breakfast today at the two schools.

	School		
Breakfast?	1	2	Total
No	19	26	45
Yes	61	124	185
Total	80	150	230

In this setting, the null hypothesis is $H_0: p_1 - p_2 = 0$ (or equivalently, $H_0: p_1 = p_2$). So we can estimate the common proportion p of all students at School 1 and all students at School 2 who missed breakfast today using

$$\hat{p}_C = \frac{X_1 + X_2}{n_1 + n_2} = \frac{19 + 26}{80 + 150} = \frac{45}{230} = 0.196$$

The rightmost column in the two-way table makes it easy to see that the overall proportion of successes in the two samples is $\hat{p}_C = 45/230$.

We use the combined (pooled) proportion of successes $\hat{p}_C$ when checking the Large Counts condition for a test of $H_0: p_1 - p_2 = 0$. In that case, $n_1\hat{p}_C$, $n_1(1-\hat{p}_C)$, $n_2\hat{p}_C$, and $n_2(1-\hat{p}_C)$ give us the expected number of successes and failures in the two samples.

CONDITIONS FOR PERFORMING A SIGNIFICANCE TEST ABOUT A DIFFERENCE BETWEEN TWO PROPORTIONS

- **Random:** The data come from two independent random samples or from two groups in a randomized experiment.
  - **10%:** When sampling without replacement, $n_1 < 0.10N_1$ and $n_2 < 0.10N_2$.
- **Large Counts:** The expected numbers of successes and failures in each sample or group—$n_1\hat{p}_C$, $n_1(1-\hat{p}_C)$, $n_2\hat{p}_C$, $n_2(1-\hat{p}_C)$—are all at least 10.

EXAMPLE **Hungry children** **Checking conditions**

PROBLEM: Researchers designed a study to compare the proportion of children who come to school without eating breakfast in two low-income elementary schools. An SRS of 80 students from School 1 found that 19 skipped breakfast today. At School 2, an SRS of 150 students included 26 who skipped breakfast today. More than 1500 students attend each school. Do these data give convincing evidence of a difference in the population proportions at the $\alpha = 0.05$ significance level? Check if the conditions for performing the test are met.

SOLUTION:

- Random? Independent random samples of 80 students from School 1 and 150 students from School 2. ✓
  - 10%: 80 < 10% of students at School 1; 150 < 10% of students at School 2. ✓
- Large Counts? $\hat{p}_C = \dfrac{19+26}{80+150} = \dfrac{45}{230} = 0.196$

  $n_1\hat{p}_C = 80(0.196) = 15.68$, $n_1(1-\hat{p}_C) = 80(0.804) = 64.32$, $n_2\hat{p}_C = 150(0.196) = 29.40$, $n_2(1-\hat{p}_C) = 150(0.804) = 120.6$ are all ≥ 10. ✓