Two sample test for proportions

In the videos, you learnt how to test if the means from two populations were different from one another. In this reading item, you will see a similar test but for comparing proportions from two different populations. A common application of two sample test for proportions is in A/B testing.

An example

Imagine you want to compare the proportion of households that own a car in Chicago (p_1) with the proportion of households that do in New York (p_2). Note that p_1 and p_2 are **population** proportions.

A possible set of hypothesis for this problem is

$$H_0 : p_1 = p_2 \text{ vs. } H_1 = p_1 \neq p_2$$

Consider for this test a significance level of 0.05

Suppose you randomly sample $n_1 = 100$ households from Chicago, 62 of which own a car, and $n_2 = 120$ households from New York, 58 of which own a car.

Defining X ="number of households that own a car in Chicago" and Y ="number of households that own a car in New York: , a good approximation for $p_1$and p_2 are

$$\hat{p}_1 = \frac{X}{100} \quad \hat{p}_2 = \frac{Y}{120}$$

Naturally, a good approximation for $\Delta = p_1 - p_2$ is

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

In order to get a good test statistic, you need to find the distribution of $\hat{\Delta}$. First, note that in $n_1$and n_2 are big enough, which in this case they are.

$$\hat{p}_1 = \frac{X}{100} \sim \mathcal{N}\left(p_1, \frac{p_1(1-p_1)}{100}\right) \quad \hat{p}_2 = \frac{Y}{120} \sim \mathcal{N}\left(p_2, \frac{p_2(1-p_2)}{120}\right)$$