Inferences Concerning the Difference Between Proportions (Independent Samples)

Recall that we can measure elements of a population in terms of a qualitative property by looking at the proportion of the population that has the specific quality.

We will use the following methods to draw conclusions about the proportions of two populations.

Recall, for one population $p' = \frac{x}{n}$ is the **sample proportion.**

x is the number of successes (number of individuals with the specific quality) in n trials.

For two populations we have two sample proportions

and

and we can create a new random variable $p_1' - p_2'$, the difference of sample proportions. From this we get a new population

$P_{p_1' - p_2'}$ the population of all possible differences $p_1' - p_2'$

$P_{p_1{}'-p_2{}'}$ is approximately normal if $P_{p_1{}'}$ and $P_{p_2{}'}$ are. One way to ensure this is to check that

Formulas for $P_{p_1{}'-p_2{}'}$

$$p_1{}' = \frac{x_1}{n} \qquad p_2{}' = \frac{x_2}{n}$$

$$\mu_{p_1{}'-p_1{}'} = p_1 - p_2$$

(mean of differences of sample proportions) (actual difference in population proportions)

This means that

$$\sigma_{p_1{}'-p_1{}'} = \sqrt{\frac{p_1{}'(1-p_1{}')}{n_1} + \frac{p_1{}'(1-p_1{}')}{n_1}}$$

Which means that, for a confidence interval estimate

Example: An experiment was conducted to test effects of a new drug. 100 mice were spilt into two groups of 50 mice and infected with a virus. The results are listed below.

Control Group (no treatment):	Group treated with new drug:
18 mice survive	30 mice survive

Construct a 95% C.I. to estimate the actual difference in the survival rates between the two groups of mice.

Hypothesis testing

P_1 and P_2 are parent populations. Random samples of size n_1 and n_2

$x_1 =$ # of successes in n_1 trials from P_1

$x_2 =$ # of successes in n_2 trials from P_2

$$p_1{}' = \frac{x_1}{n} \qquad\qquad p_2{}' = \frac{x_2}{n}$$

Because we don't know the population proportions, to calculate the test statistic we will need the **pooled proportion of success**

And so the test statistic is

$$z = \frac{p_1{}' - p_2{}'}{\sqrt{p_p{}'(1-p_p{}')\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

(Notice that we use the z-table!)

Note: This is because, when using the pooled proportion of success we get

$$\sigma_{p_1{}'-p_2{}'} \approx$$

The test statistic is normally distributed provided

$$n_1 \cdot p_p{}' > 5 \qquad n_2 \cdot p_p{}' > 5$$

$$n_1 \cdot (1 - p_p{}') > 5 \qquad n_2 \cdot (1 - p_p{}') > 5$$

Example: An experiment was conducted to test effects of a new drug. 100 mice were spilt into two groups of 50 mice and infected with a virus. The results are listed below.

Control Group (no treatment):	Group treated with new drug:
18 mice survive	30 mice survive

Test at 5% significance whether the drug was effective.