- **Confidence intervals for 2 proportions**

  Go to the Hyperstat site to learn about confidence intervals for 2 proportions. Be sure to only work through the 4 pages on this topic.

The confidence interval on the difference between proportions is based on the same general formula as are other confidence intervals. A confidence interval on the difference between two proportions is computed in the following situation: There are two populations and the members of each population can be classified as falling into one of two categories. For example, the categories might be such things as whether or not one has a high-school degree or whether or not one has ever been arrested.

Consider a researcher interested in whether people who majored in psychology are more or less likely than physics majors to solve a problem that involves a certain type of statistical reasoning. The researcher is interested in estimating the difference in the proportions of people in the two populations that can solve the problem and in computing a 99% confidence interval on the difference. Random samples of 100 psychology majors and 110 physics majors are taken and each person is given a chance to solve the problem. Of the 100 psychology majors, 65 solve the problem; of the 110 physics majors only 45 solve it.

Therefore the proportions who solve the problem are 0.65 for the psychology majors (p_1) and 0.41 for the physics majors (p_2). The goal of the experiment is to estimate the difference between the proportion of psychology majors in thepopulation that can solve the problem (π_1) and the proportion of physics majors in the population that can solve it (π_2). The statistic $p_1 - p_2$ is used as an estimate of $\pi_1 - \pi_2$. In this experiment, the estimated value of $\pi_1 - \pi_2$ is 0.65 - 0.41 = 0.24.

To compute the confidence interval, the standard error of $p_1 - p_2$ is needed. The standard error is:

$$\sigma_{p_1-p_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

The estimated standard error is:

$$s_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{0.65(1-0.65)}{100} + \frac{0.41(1-0.41)}{110}}$$

= 0.067

Finally, the value of z for the 99% confidence interval is computed using a z table; it is 2.58. The lower limit of the confidence interval is simply:

$p_1 - p_2 - (z)$ (estimated $\sigma_{p_1-p_2}$)

$= 0.65 - 0.41 - (2.58)(0.067) = 0.07$

The upper limit is:
$= 0.65 - 0.41 + (2.58)(0.067) = 0.41$

The 99% confidence interval is therefore:

$0.07 \leq \pi_1 - \pi_2 \leq 0.41$.

This indicates that the proportion of psychology majors that can solve the problem is from 0.07 to 0.41 higher than the proportion of physics majors that can solve it.

Summary of Computations

1. Compute $p_1 - p_2$.
2. Find z for confidence interval using a z table.
3. $\sigma_{p_1 - p_2}$ with the formula:

$$s_{p_1 - p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{0.65(1-0.65)}{100} + \frac{0.41(1-0.41)}{110}}$$

4. lower limit = $p_1 - p_2 - (z)$ (estimated $\sigma_{p_1 - p_2}$)
5. upper limit = $p_1 - p_2 + (z)$ (estimated $\sigma_{p_1 - p_2}$)
6. Confidence interval: lower limit $\leq \pi_1 - \pi_2 \leq$ upper limit.

Assumptions

1. The two proportions are independent.
2. The adequacy of the normal approximation depends on the sample size (N) and π. Although there are no hard and fast rules, the following is a guide to needed sample size:

   If π is between 0.4 and 0.6 then an N of 10 is adequate. If π is as low as 0.2 or as high as 0.8 then N should be at least 25. For π as low as 0.1 or as high as 0.9, N should be at least 30.

A more conservative rule of thumb that is often recommended is that $N\pi$ and $N(1 - \pi)$ should both be at least 10.

Click here for an interactive demonstration of the normal approximation to the binomial to explore the validity of these rules of thumb.