Confidence Intervals for Two Groups

David Armstrong

UCI

Specific Formulas for Confidence Intervals

Parameter Description	Confidence Interval	Parameter Estimated
Difference in 2 population proportions	$\hat{p}_1 - \hat{p}_2 \pm Z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$	$p_1 - p_2$
Population Mean of paired differences (σ_d unknown)	$\bar{X}_d \pm t_{n_d-1}\frac{s_d}{\sqrt{n_d}}$	μ_d
Difference in 2 population means for independent samples (σ_1 and σ_2 unknown) ($\sigma_1 \neq \sigma_2$)	$\bar{X}_1 - \bar{X}_2 \pm t_{min(n_1-1,n_2-1)}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$\mu_1 - \mu_2$
Difference in 2 population means for independent samples (σ_1 and σ_2 unknown) ($\sigma_1 = \sigma_2$) Pooled	$\bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$	$\mu_1 - \mu_2$

Assumptions for 2 Groups

Difference in 2 population proportions

- $n_1\hat{p}_1 \geq 10$ and $n_1\hat{q}_1 \geq 10$
- $n_2\hat{p}_2 \geq 10$ and $n_2\hat{q}_2 \geq 10$
- Sample size is less than 10% of the population size; if we are sampling is w/out replacement
- The sample can be regarded as a simple random sample from the population of interest.
- The data values are assumed to be independent of each other.

Population mean of paired differences (2 dependent samples)

- We need to have a large enough sample size of pairs ($n > 30$). For $n < 30$ with extreme skewness or outliers, you cannot use this method.
- Sample size is less than 10% of the population size; if we are sampling is w/out replacement
- The sample can be regarded as a simple random sample from the population of interest.
- The differences between the pairs are assumed to be independent of each other.
- Differences within pairs are dependent.

Difference in 2 population means for independent samples

- We need to have a large enough sample size for each group ($n_1 > 30$ and $n_2 > 30$). For $n_1 < 30$ or $n_2 < 30$ with extreme skewness or outliers, you cannot use this method.
- Sample size is less than 10% of the population size; if we are sampling is w/out replacement
- The sample can be regarded as a simple random sample from the population of interest.
- The data values are assumed to be independent of each other.
- The groups are independent as well.

EX: A local restaurant keeps records of reservations and no-shows. In a random sample of 150 Saturday reservations, it is found that 70 of them were no-shows. In a random sample of 125 Friday reservations, it is found that 45 of them were no-shows.

a. Find the difference between the sample proportion of Saturday no-shows versus Friday no-shows.

b. Calculate the standard error of the difference between the sample proportion of Saturday no-shows and the sample proportion of Friday no-shows.

c. Find a 95% confidence interval for the true difference between the proportion of Saturday no-shows versus the proportion of Friday no-shows at this restaurant.

d. Which of the following statements is a correct interpretation of the confidence interval obtained?

   We can be 95% confident that the difference between the proportion of Saturday no-shows versus Friday no-shows in the sample is within the interval obtained.

   If this study were to be repeated with a sample of the same size, there is a .95 probability that the difference between the sample proportion of Saturday no-shows versus Friday no-shows would be in the interval obtained.

   We can be 95% confident that the difference between the true proportion of Saturday no-shows versus Friday no-shows is within the interval obtained.

   There is a 95% probability that the the difference between the population proportion of Saturday no-shows versus Friday no-shows is within the interval obtained.

EX: Suppose we have 10 football players that are selected to participate in a study. Although speed and strength are a necessity, flexibility and grace can also help their game. These players flexibility was measured in a sit and reach before taking Ballet classes for a month and then measured again at the end of the class. The Institute of Ballet claims that the average difference should be 4 inches.

Before	12	6	7	8	9	10	11	15	3	5
After	13	12	10	9	10	8	10	15	9	8
Difference										

a. Create a 99% Confidence interval for the true average increase in flexibility of the pairs if we know the sample standard deviation for the differences is 2.7 inches.

b. Based on the interval, do you think that the Institute's claim is correct?

EX: Suppose we have two groups of people that we would like to compare. The first group received a new weight loss drug. The second group thought they were receiving the drug, but instead were given a "sugar pill". The participants were weighed at the beginning of the study. After 4 weeks, the participants were weighed again. There weight loss was measured by subtracting their weight at the end of the study from their weight at the beginning of the study.

Group 1		Group 2	
n_1	12	n_2	20
s_1	4	s_2	3
Observed average weight loss	10	Observed average weight loss	4

a. If we assume that the new drug does not help people lose weight, what value do you think would be in our confidence interval?

b. What is the sample average difference between group 1 and group 2?

c. What is the standard error of the sample average difference between group 1 and group 2?

d. Create a 95% confidence interval for the true average difference in weight loss for group 1 versus group 2.

e. Interpret your interval in context of the problem

f. Do you think the weight loss drug works?

EX: A 95% confidence interval for the true difference between two population proportions is (0.2850, 0.3700).

a. Find the point estimate for the difference between the two population proportions.

b. Find the margin of error.

c. Find the standard error.

EX: Below is a 90% confidence interval for the true average amount of weight women lose after 1 month of taking a diet pill:
(7.1398, 12.86).

a. What is the point estimate?

b. What is the margin of error?