

Cheat Sheet: Confidence Intervals for Population Means

Essential Concepts

- The mathematical formulas to find the mean and standard deviation of the sampling distribution of the sample mean for samples of size n are:
 - Mean of the sampling distribution of the sample mean = μ
 - Standard deviation of the sampling distribution of the sample mean = $\frac{\sigma}{\sqrt{n}}$

where μ and σ represent the mean and standard deviation of the original population, respectively.

- If the population distribution is normal, the distribution of the sample means will also follow a normal distribution, for any sample size.
- If the population distribution is not normal, the Central Limit Theorem states that the distribution of the sample means still follows an approximate normal distribution as long as the sample size is large (e.g., $n \geq 30$) and the population distribution is not strongly skewed.
- When we estimate σ using the sample standard deviation, s , we use the t -statistic:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \text{with a degree of freedom of } df = n - 1.$$

- The formula for a confidence interval for a population mean is:

estimate \pm margin of error

$$\bar{x} \pm (t\text{-critical value}) \frac{s}{\sqrt{n}}$$

where \bar{x} is the sample mean and the standard error used is the standard error of

the sample mean, $\frac{s}{\sqrt{n}}$. The t -critical value (t -score in the data analysis tool) in the confidence interval will depend on the sample size (degrees of freedom for the t -distribution: $df = n - 1$) and the confidence level.

- When you are interested in estimating a difference in population means, you usually start with data from samples from each of the populations of interest. There are two different strategies for selecting the two samples.
 - Independent samples: to select a sample from one population and then independently select a sample from the second population.
 - Paired samples or Dependent samples: if samples are chosen in a way that results in the observations in one sample being paired with the observations in the other sample.
- The conditions that you need to check for the two-sample t confidence interval are:
 1. The samples are independent.
 2. Each sample is a random sample from the corresponding population of interest or it is reasonable to regard the sample as if it were a random sample. It is reasonable to regard the sample as a random sample if it was selected in a way that should result in the sample being representative of the population. If the data are from an experiment, you just need to check that there was random assignment to experimental groups—this substitutes for the random sample condition and also results in independent samples.
 3. For each population, the distribution of the variable that was measured is approximately normal, or the sample size for the sample from that population is large. Usually, a sample of size 30 or more is considered to be “large.” If a sample size is less than 30, you should look at a plot of the data from that sample (a dotplot, a boxplot, or, if the sample size isn’t really small, a histogram) to make sure that the distribution looks approximately symmetric and that there are no outliers.
 - The form of the confidence intervals for a difference in sample means is:

estimate \pm margin of error, where margin of error = $(t\text{-statistic})(\text{standard error}) =$

$$(t\text{-statistic}) \left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

Key Equations

confidence interval for one sample

$$\bar{x} \pm (t - \text{critical value}) \frac{s}{\sqrt{n}}$$

confidence interval for two samples

$$(\bar{x}_1 - \bar{x}_2) \pm (t - \text{critical value}) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

degrees of freedom

$$df = n - 1$$

t-score

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

margin of error

$$(\text{t-statistic})(\text{standard error}) = (\text{t-statistic}) \left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

standard error of a sample mean

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

standardized statistic

$$z = \frac{\bar{x} - [\text{mean of } \bar{x}'s]}{\text{std. deviation of } \bar{x}'s} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Glossary

independent sample

random samples from each population

paired samples, dependent samples

samples that are chosen in a way that results in the observations in one sample being paired with the observations in the other sample

sampling distribution

the probability distribution of a sample statistic

standardized statistic

a z-score for a statistic using simulation to estimate the mean and standard deviation of the sample mean