Cheatsheets / **Learn R**

codecademy

Learn R: Hypothesis Testing

Sample Vs. Population Mean

In statistics, we often use the mean of a sample to estimate or infer the mean of the broader population from which the *sample* was taken. In other words, the *sample mean* is an estimation of the *population mean*.

Hypothesis Test P-value

Statistical *hypothesis tests* return a *p-value*, which indicates the probability that the *null hypothesis* of a test is true. If the p-value is less than or equal to the *significance level*, then the null hypothesis is rejected in favor of the alternative hypothesis. And, if the p-value is greater than the significance level, then the null hypothesis is not rejected.

Hypothesis Test Errors

Type I errors, also known as *false positives*, is the error of rejecting a null hypothesis when it is actually true. This can be viewed as a miss being registered as a hit. The acceptable rate of this type of error is called *significance level* and is usually set to be `0.05` (5%) or `0.01` (1%).
Type II errors, also known as *false negatives*, is the error of not rejecting a null hypothesis when the alternative hypothesis is the true. This can be viewed as a hit being registered as a miss.
Depending on the purpose of testing, testers decide which type of error to be concerned. But, usually `type I` error is more important than `type II`.

Central Limit Theorem

The *central limit theorem* states that as samples of larger size are collected from a population, the distribution of sample means approaches a normal distribution with the same mean as the population. No matter the distribution of the population (uniform, binomial, etc), the sampling distribution of the mean will approximate a normal distribution and its mean is the same as the population mean.

The central limit theorem allows us to perform tests, make inferences, and solve problems using the normal distribution, even when the population is not normally distributed.

Print Share