Resampling Techniques

Share on
-
-
-
-

Sampling >

Resampling techniques are a set of methods to either repeat sampling from a given sample or population, or a way to estimate the precision of a statistic. Although the method sounds daunting, the math involved is relatively simple and only requires a high school level understanding of algebra.

Informally, *resample* can mean something a little simpler: repeat *any* sampling method. For example, if you're conducting a Sequential Probability Ratio Test and don't come to a conclusion, then you resample and rerun the test. For most intents and purposes though, if you read about *resampling* (as opposed to "resample"), then the author is most likely talking about a specific resampling technique.

Specific Resampling Techniques

The main techniques are:

1. Bootstrapping and Normal resampling (sampling from a normal distribution).
2. Permutation Resampling (also called Rearrangements or Rerandomization),
3. Cross Validation.

1. Bootstrapping and Normal Resampling

Bootstrapping is a type of resampling where large numbers of smaller samples of the same size are repeatedly drawn, with replacement, from a single original sample. **Normal resampling** is very similar to bootstrapping as it is a special case of the normal shift model—one of the assumptions for bootstrapping (Westfall et al., 1993). Both bootstrapping and normal resampling both assume that samples are drawn from an actual population (either a real one or a theoretical one). Another similarity is that both techniques use sampling *with* replacement.

Ideally, you would want to draw large, non-repeated, samples from a population in order to create a sampling distribution for a statistic. However, limited resources may prevent you from getting the ideal statistic. Resampling means that you can draw small samples over and over again from the same population. As well as saving time and money, the samples can be quite good approximations for population parameters.

2. Permutation Resampling

Unlike bootstrapping, permutation resampling doesn't need any "population"; resampling is dependent only on the assignment of units to treatment groups. The fact that you're dealing with actual samples, instead of populations, is one reason why it's sometimes referred to as the Gold standard bootstrapping technique (Strawderman and Mehr, 1990). Another important difference is that permutation resampling is a without replacement sampling technique.

3. Cross Validation

Cross-validation is a way to validate a predictive model. Subsets of the data are removed to be used as a validating set; the remaining data is used to form a training set, which is used to predict the validation set.

References

Good, P. (2006). Resampling Methods: A Practical Guide to Data Analysis. Springer Science & Business Media.
Strawderman & Mehta (1990). On the validation of exact tests in software packages. Unpublished Manuscript.
Westfall, P. et al., (1993). Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment. John Wiley & Sons.

By Stephanie | December 7, 2018 | Statistics How To |

Find an article

Search
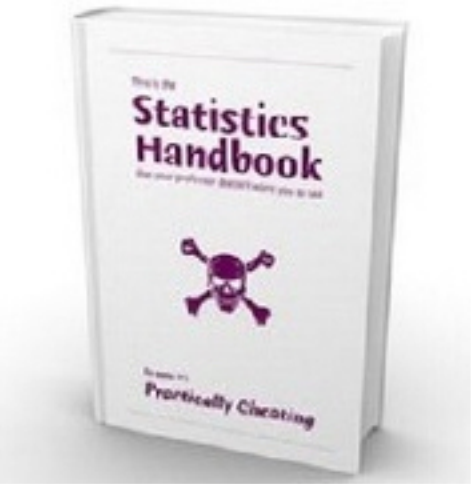
Feel like "cheating" at Statistics? Check out the grade-increasing book that's recommended reading at top universities!

NEED HELP NOW with a homework problem? CLICK HERE!

Probability and Statistics Topic Indexes

Basic Statistics.
Bayesian Statistics and Probability
Descriptive Statistics: Charts, Graphs and Plots.
Probability.
Binomial Theorem.
Definitions for Common Statistics Terms.
Critical Values.
Hypothesis Testing.
Normal Distributions.
T-Distributions.
Central Limit Theorem.
Confidence Intervals.
Chebyshev's Theorem.
Sampling and Finding Sample Sizes.
Chi Square.
Online Tables (z-table, chi-square, t-dist etc.).
Regression Analysis / Linear Regression.
Non Normal Distributions.