

Introduction to Statistics

*Course Description*

This course provides a practical introduction to statistics and inference. Participants should leave the course understanding the difference between the population parameters we want to know, and the statistics from samples that estimate them. In addition, participants should be able to use summary statistics to describe the central value and spread in a distribution. The course will show how to design an experiment, including an estimate of sample size, to determine if the means of two populations are different. Lessons incorporate both lecture and hands-on exercises with a focus on cultivating practical skills.

*Learning outcomes*

After studying this program, you will be able to:

* Use statistical inference to estimate the mean of a population with a confidence interval from information provided in a random sample
* Have a working conceptual understanding of probability distributions and the cumulative distribution function
* Select the appropriate probability distribution for classical probability problems
* Distinguish between statistics calculated on samples, and expectations calculated on distributions
* Have a working conceptual understanding of the central limit theorem and the law of large numbers
* Distinguish between the concepts of correlation and causation
* Articulate the difference between Type I and Type II errors, and provide reasonable arguments about the type of error you would want to minimize on a particular problem
* Have a working conceptual understanding of hypothesis tests
* Perform hypothesis tests and A/B tests: i.e. test whether different populations have different parameters based on samples, while estimating the probability of making Type I and Type II errors
* Design experiments that have sufficient statistical power to calculate an effect of interesting size

*Outline*

* **Day 1: Intro to statistics and probability**
  + Statistical measures of centrality (mean, median, mode) and spread (quartiles, variance, standard deviation)
  + Discrete distributions (binomial, Poisson)
  + Continuous distributions (Exponential, Normal)
  + PDF, CDF, expected value of distributions
  + Bayes’ theorem
  + Law of Large Numbers and the Central Limit Theorem
  + Correlation coefficient
* **Day 2: Hypothesis testing and experimentation**
  + Correlation vs causation
  + Introduction to hypothesis testing: distinguishing between Type I and Type II error
  + Confidence intervals and p-values
  + Performing two-sample t-tests
  + Multiple comparisons and Bonferroni adjustments
  + False Discovery Rate
  + Power and Sample size calculations
  + Design of experiments
  + Chi-square tests for categorical data