Statistics for Data Analysis in R Training

Duration: **2 Day**

BILGINC IT Academy
info@bilginc.com
bilginc.com

This two-day course is designed for those who analyse data or who are creating machine learning models, but who wish to firm their understanding in core concepts as well as expanding into types of data distributions, inferential statistics (hypothesis tests), statistical significance, and a deeper understanding of how linear regression works. It is expected that you will have experience with a programming language used for data analysis such as R or Python – if this is not currently the case we suggest completing one of our R or Pythonfor Data Handling courses.

As well as providing a business context to using core concepts such as averages, spread, and interpreting analyst visualisations, you will take this knowledge further and learn how distributions, sampling, and hypothesis testing can be used to analyse data in an organisation and in automatically highlighting significant results or anomalies.

If you are on a learning journey with Machine Learning and AI this course will give you a strong starting point in the statistical methods that underpin a large number of algorithms without overloading, you with too many mathematical formulae or notations that are otherwise commonly used to communicate advanced mathematics. Your focus will be on business problems and applying tools such as R that you will need as part of this journey.

Throughout the course you will engage with practical labs, activities, and discussions with one of our technical specialists. All modules involve the use of R to practice the techniques taught – setting you up to succeed in analysing, interpreting, and getting value from your data.

Prerequisites

- Minimum of GCSE Maths or equivalent
- Experience with Python or R for Data Handling

Target Audience

Anyone wishing to expand their understanding of Maths and Statistics related to Data Science. This course will provide all the required pre-requisite statistical knowledge needed for our more in-depth programmes.

Training Outline

Central Tendency, Variation, and Outliers - Using an appropriate software tool, calculate:

- Mean, Mode, Median, Mid-range
- Population and Sample Standard Deviation & Variance
- Inter-Quartile Range
- Apply methods for automating identification of outliers
- Discuss appropriate handling of outliers
- Practical Lab Activities with R

Visualisations and Skew - Using an appropriate software tool, create:

BILGINC IT Academy

info@bilginc.com

bilginc.com

- Histograms
- Scatter Plots
- Use these to:
  - Identify skew and the effect this may have on modelling
  - Identify the location of the averages
  - Compare twsamples (e.g. taken at different times or from different locations)
  - Determine the appropriate shape of a model and whether there are opportunities t linearise
- Practical Lab Activities with R

Introduction to Probability

- Interpret P() notation and calculate simple and conditional probabilities
- Use Venn diagrams with set notation tcalculate probabilities
- Use Tree diagrams and simple combinatorics tcalculate probabilities
- Practical Lab Activities with R

Introduction to Distributions

- Recognise what a probability or data distribution is
- Identify when a distribution is considered tbe Binomial, Poisson, or Normal
- Identify when a distribution can be treated as Normal and what this means for analytical methods
- Practical Lab Activities with R Sampling
- Critique different sampling techniques
- Explain the impact a sampling or data gathering method may have on analytical model results
- Recognise methods for estimating summary statistics for a population from a sample
- Practical Lab Activities with R

Introduction to Hypothesis Testing

- Recognise the steps required for a Hypothesis test from the set- up, assumptions, testing, and interpretation of p-values
- Identify a variety of tests and when they are used
- Evaluate the output of tests from an appropriate software tool
- Practical Lab Activities with R

Linear Regression

- Recognise when a linear regression is an appropriate method tuse
- Interpreting y = mx + c
- Evaluate linear models
- Practical Lab Activities with R