Statistics for Data Analysis in Python Training

Duration: **2 Day**

BILGINC IT Academy
info@bilginc.com
bilginc.com

Overview

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 Python or R – if this is not currently the case we suggest completing one of our Python or R for 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 Python or R that you will need as part of this journey.

If you wish to expand your understanding of Maths and Statistics related to Data Science then this course will give you all the required pre-requisite statistical knowledge needed for our more in depth programmes.

Throughout the course you will engage with practical labs, activities, and discussions with one of our technical specialists. All modules involve the use of Python or 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

This course is intended for those who are already at ease with handling data in Python and may form part of a learning journey in Data Analytics, Data Engineering, or Data Science.

- Data Analysts
- Data Engineers
- Data Scientists
- Software Developers

Statistics for Data Analysis in Python Training

Duration: **2 Day**

BILGINÇ IT Academy
info@bilginc.com
bilginc.com

What You Will Learn

During this course you will cover:

- How to use python for statistical analysis
- A review of fundamental statistics and probability in the context of implementing these calculations in python
- How to begin using and interpreting advanced level notation for probability and statistics
- The need for recognising how data is distributed and the unexpected effects that sampling can have when calculating summary statistics
- A detailed introduction to inferential statistics and hypothesis testing which will give you a deeper understanding when interpreting the meaning of p-values
- Consideration of how linear regression methods are based on statistical techniques

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
- Discuss when the above measures are appropriate
- Apply methods for automating identification of outliers
- Discuss appropriate handling of outliers
- Practical Lab Activities with Python

Visualisations and Skew

- Using an appropriate software tool, create:
  - Histograms
  - Scatter Plots
- Use these to:
  - Identify skew and the effect this may have on modelling
  - Identify the location of the averages
  - Compare two samples (e.g.taken at different times or fromdifferent locations)
  - Determine the appropriate shape of a model and whetherthere are opportunities to linearise
- Practical Lab Activities with Python

Introduction to Probability

- Interpret P() notation and calculate simple and conditionalprobabilities
- Use Venn diagrams with set notation to calculate probabilities
- Use Tree diagrams and simple combinatorics to calculateprobabilities
- Practical Lab Activities with Python

Introduction to Distributions

- Recognise what a probability or data distribution is
- Identify when a distributionis considered to beBinomial, Poisson,or Normal
- Identify when a distribution can be treated as Normal and whatthis means for analytical methods

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com

- Practical Lab Activities with Python

Sampling

- Critique different sampling techniques
- Explain the impact a sampling or data gathering method mayhave on analytical model results
- Recognise methods for estimating summary statistics for apopulation from a sample
- Practical Lab Activities with Python

Introduction to Hypothesis Testing

- Recognise the steps required for a Hypothesis test from thesetup, 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 Python

Linear Regression

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