

Module 3: Statistics & Visualization

Data Science with Python – BSc Course

Why Statistics & Visualization?

Long-Term Capital Management (LTCM) collapsed in 1998 because their models assumed normal distributions. They lost \$4.6 billion in four months. Understanding your data – its shape, its tails, its correlations – prevents billion-dollar mistakes.

Statistics is not just about calculating means and standard deviations. It is about understanding risk, testing hypotheses, and communicating findings convincingly to stakeholders.

Understanding distributions is the difference between profit and catastrophic loss

- **Risk measurement:** VaR and Expected Shortfall depend on distribution assumptions
- **Regulatory reporting:** Stress tests require statistical evidence for scenarios
- **Investment decisions:** Is this alpha statistically significant or just noise?
- **Communication:** Present data-driven arguments that convince boards and regulators

Statistics turns data into defensible decisions

By the end of this module, you will be able to:

- Calculate and interpret descriptive statistics (mean, median, variance, skewness)
- Understand probability distributions and their financial applications
- Conduct and interpret hypothesis tests (t-tests, p-values)
- Measure and visualize correlations between assets
- Create publication-quality charts with matplotlib and seaborn
- Tell compelling data-driven stories through visualization

From raw numbers to statistically rigorous insights

Lesson Roadmap

Lesson	Topic	Narrative Arc	Focus
L13	Descriptive Statistics	"The Numbers Behind the Numbers"	Central tendency, dispersion
L14	Probability Distributions	"The Shape of Randomness"	Normal, t, fat tails
L15	Hypothesis Testing	"The Evidence Standard"	t-tests, p-values, errors
L16	Correlation Analysis	"The Web of Relationships"	Pearson, Spearman, heatmaps
L17	Matplotlib Basics	Line, bar, scatter plots	(Wave 1b)
L18	Seaborn Statistical Plots	Distribution, regression	(Wave 1b)
L19	Multi-Panel Figures	Subplots, grid layouts	(Wave 1b)
L20	Data Storytelling	Narrative visualization	(Wave 1b)

Module Story: This module turns raw data into understanding. L13 measures the vital signs. L14 reveals the shape. L15 tests your claims. L16 maps the relationships. Then L17-L20 teach you to show it all visually.

Eight lessons: from numbers to insights to compelling visual stories

- **Descriptive Statistics** – Summarize data with mean, median, std, skewness, kurtosis
- **Probability Distributions** – Normal, t, chi-squared and their finance use cases
- **Hypothesis Testing** – t-tests, p-values, confidence intervals
- **Correlation** – Pearson and Spearman correlations for portfolio diversification
- **Matplotlib & Seaborn** – Professional charts for reports and presentations
- **Data Storytelling** – Turn charts into compelling narratives

Statistics and visualization together make insights actionable

Scenario: Risk Report Generator

Using the skills from this module, you can:

- Compute Value at Risk (VaR) and Expected Shortfall
- Test whether returns follow a normal distribution (Jarque-Bera test)
- Visualize portfolio risk with professional charts
- Present findings in a board-ready report with statistical evidence

This is what risk managers deliver to boards monthly – and you will build it from scratch.

Professional risk reporting requires statistics and visualization together

Who Uses This?

- **Risk Management** – Every bank computes VaR and stress tests daily
- **Research** – Morgan Stanley publishes 1,000+ research reports per year with charts
- **Regulators** – Stress test results are communicated through visualizations
- **Asset Management** – Morningstar ratings are built on statistical analysis

Statistics and visualization are core skills across finance

Statistics gave you the tools to describe data. Regression gives you the power to predict it.

Module 4 introduces linear regression, regularization, and factor models – the foundation of quantitative finance and machine learning.

Prerequisite check: Can you compute correlation, test hypotheses, and create professional charts? If yes, you are ready.

From descriptive statistics to predictive models

Let's Begin!

First up: L13 – Descriptive Statistics

Open your laptop and follow along.