



# Descriptive Data Analysis with Excel and Python in Excel

**Program Description:** This course gives an introduction to basic methods used in applied descriptive data analysis and how they can be implemented in Excel and Python in Excel. All statistical methods are explained in detail and illustrated with example calculations.

**Target Audience:** junior up to experienced investment professionals, risk managers, investment analysts, quantitative analysts, portfolio managers, IT professionals.

**Materials:** Participants will receive the slides presented, spreadsheets containing example calculations and important papers in PDF format.

**Course Delivery:** This course will be delivered online (MS Teams) over a) two days or b) four afternoon. The minimum number of participants is 4.

**Price:** 550 CHF per participant. If more than one member of the same company participates, a discount of 10% is given on the total course fee. Members of the Swiss CFA Society will receive a 10%.

The content of this program can be combined with content from different programs for customized **inhouse trainings**. Please contact [email@andreassteiner.net](mailto:email@andreassteiner.net) for details. **More information** is available on [www.andreassteiner.net/consulting](http://www.andreassteiner.net/consulting)

## Introduction

- Overview and organisation of the course
- Objectives and expected outcomes

## Session 1: Fundamentals of Descriptive Data Analysis

- What is descriptive data analysis?
- Importance in data-driven decision-making
- Types of data: qualitative vs. quantitative
- How NOT to cheat with statistics: good and bad practices

## Session 2: Data Preparation

- Importing data from various sources
- Data cleaning techniques
- Handling missing values

## Session 3: Univariate Statistics

- Creating and interpreting distributions
  - Frequency distributions (probabilities distribution functions, pdf)
  - Cumulative frequency distributions (density functions, cdf)
  - Visualisation: histograms, box plots, log pdf, cdf, inverse cdf
- Quantiles, percentiles, quartiles
- Measures of central tendency



- mean, median, mode
- weighted means
- arithmetic and geometric means
- Measures of variability: range, variance, standard deviation, interquartile range
- Moments: location and dispersion, skewness and kurtosis
- Empirical probability distribution and empirical density
- Fitting distributional models: normal, Student t and NIG distributions
- Concentration measurement: Lorenz curve, Gini and Herfindahl index, Entropy
- Indexing: single and compound indices, change of basis, basis effects, chain-linking

#### Session 4: Bivariate Statistics

- Correlation analysis
  - Covariance
  - Pearson correlation coefficients
  - Spearman correlation coefficients
  - Correlation versus causation
  - Visualization; scatter plots
- Dependency analysis: empirical copulas
- Linear regression
  - Method of least squares
  - Interpretation of regression parameters
  - Coefficient of determination
  - Graphical representation of linear regression
  - Desirable and undesirable properties of residuals
  - Data transformations
- Bivariate distributions: pdf, cdf and contours

#### Session 5: Time Series Analysis

- Decomposition of a time series into trend and seasonal components
- Dependency over time: autocorrelation
- White noise and random walks
- Basic autoregressive and moving average time series models
- Forecasting
- Structural breaks

#### Course Conclusion and Feedback

- Review key concepts
- Feedback from participants
- Additional resources for continued learning