

Financial Data Analytics

Ralf Kellner

Semester 2024 / 2025

Organization of the course

- The course consists of a lectures on Wednesdays 12-14, HK 14b (SR 105) and exercises on Wednesdays 14-16, HK 14b (SR 017)
- Lectures discuss theoretical and empirical topics of the course
- Exercises repeat the content, however, with a perspective on applications and calculations
- You should participate in both, if you can't, team up with your fellow students
- The exam is paper based, it includes output from data analysis as well as questions which include calculations or open questions

Outline

- 1 Introduction
- 2 Data access
- 3 Descriptive Analysis

Introduction

Welcome to the Financial Data Analytics course.

Due to advancing digitalization, large amounts of data on companies, products, and individuals have been generated in the last decades. The increase in storage and computing capacities has fueled the evolution of data analysis methods, allowing more automated and complex analyses of data.

The goal of data analysis is to gain insights that can generate new value. This course focuses on empirical analysis of financial data, but the methods covered are broadly applicable.

Course Content

The course covers the following key topics:

- Data types, data retrieval, and data preparation
- Descriptive and graphical analysis of data
- Regression models and model evaluation
- Dimension reduction
- Clustering

We will use a "learning by doing" approach, working hands-on with data. The course materials include digital worksheets and an online book.

Python in Financial Data Analytics

Python as a Tool for Data Analysis:

Python is a popular programming language, especially useful for data analysis. In this course, you will not be learning Python programming comprehensively. Instead, Python serves as a tool to help you understand data analysis and interpret results.

Course notebooks and examples will provide practical experience without requiring prior programming knowledge.

Outline

- 1 Introduction
- 2 Data access
- 3 Descriptive Analysis

Access to Data

There are many ways to obtain data for analysis. In this section, we will discuss a few common methods of data acquisition:

- Downloading and importing smaller files, such as CSV or TXT files.
- Using databases.
- Accessing data via APIs.

This is just a brief overview; more complex methods exist, which may require further knowledge.

File Import

One of the most common ways to obtain data is by sharing files. Common formats include CSV or TXT, where each line typically corresponds to an observation, and variables are separated by commas or other delimiters.

When reading such files, it is important to specify the correct delimiter based on how the data is structured (e.g., space, comma). For example, a CSV file can be downloaded from sources such as Yahoo Finance.

Databases and APIs

Using Databases:

Data can also be accessed via databases, which store large amounts of structured data. Query languages such as SQL are often used to retrieve information from databases.

Accessing Data via APIs:

Application Programming Interfaces (APIs) allow programmatic access to data from various sources. APIs often provide structured data, such as JSON or XML, which can be processed and analyzed directly.

Outline

- 1 Introduction
- 2 Data access
- 3 Descriptive Analysis

Descriptive Analysis of Data

Descriptive analysis is an essential part of any data analysis process. It is typically done at the beginning of an analysis to gain a first impression of the data's characteristics and potential correlations.

Descriptive analysis includes:

- Interpretation of data distributions.
- Exploration of correlations between variables.
- Use of descriptive statistics and graphical representations.

Data Collection

Every data analysis starts with data collection. The goal is to identify relationships that are generally valid. For example, in analyzing rental prices of housing, we must:

- Define the relevant unit (e.g., apartments).
- Determine which factors (besides rental price) may influence the analysis.

Since collecting data on the entire population is often impossible, a representative and random sample is taken.

Descriptive Statistics

Descriptive statistics provide quantitative measures that summarize the characteristics of data. Common metrics include:

- Mean, median, and mode
- Standard deviation and variance
- Quartiles and percentiles

These statistics give a clear view of the data's central tendency, dispersion, and distribution.

Graphical Representations

Visualizing data is an essential part of descriptive analysis. Common graphical methods include:

- Histograms: Show frequency distribution.
- Boxplots: Display data spread and potential outliers.
- Scatter plots: Visualize relationships between two variables.

These visualizations help identify patterns and potential anomalies in the data.

Data Interpretation

Distribution of Characteristics:

Understanding the distribution of variables helps in drawing meaningful conclusions. For example, does a higher number of study hours correlate with better grades?

Correlation Between Variables:

Descriptive analysis also looks at how variables relate to each other. For example, analyzing the relationship between family education level and student performance.

Conclusion

Descriptive analysis serves as the foundation of any comprehensive data analysis. It provides the first look at the data and helps guide further in-depth statistical analyses.

By using a combination of descriptive statistics and graphical representations, we can summarize and interpret data effectively, identifying key trends and relationships.