Ain Shams University

Faculty of Computer and Information Science

Time Series

By:

Amna Ahmed Mirghani	2021170087
Fatma Gamal Salleh	2021170386
Habiba Alaa Eldin Adel	2021170161
Habiba Bakr Ahmed	2021170159
Khira Ebe	2021170178
Rana Wahid Tammam	2021170195
Sumaya Bile	20201701273
Wessal Salah	2021170614

Course Name: Statistical Analysis

Instructors: Prof. Safia Abbas, Dr. Ayat Mohammed

Department: Mainstream

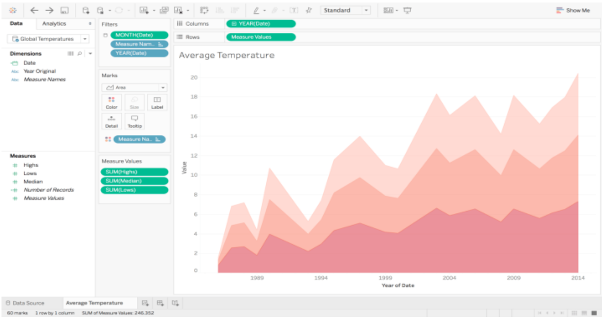
December 10, 2022

Introduction

Our chosen topic of research is time series and time series analysis. We will study it along with its types, its models and its applications in real life, and discover what is new.

We chose this topic because our goal is to understand it well, and this is due to our interest in statistics, data science and data analysis.

There are many practical situations where data might be correlated. This is particularly so where repeated observations on a given system are made sequentially in time. (Reiner, 2010).



A time series is a set of statistics gathered sequentially in time. Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. It is the systematic approach by which one goes about answering the mathematical and statistical questions posed by these time correlations.

Historically, time series methods were applied to problems in the physical and environmental sciences. This fact accounts for the basic engineering flavor permeating the language of time series analysis.

The first step in any time series investigation always involves careful scrutiny of the recorded data plotted over time.

In this research, we'll cover the following items for time series analysis:

1. Summary
2. History
3. Definition
4. Types
5. Models and techniques
6. Applications

Summary

A time series is a set of statistics, usually collected at regular intervals.This research aims to identify this important topic for those who don't have any background about it or want to know more and clearly understand it.

This research begins by discussing the history of time series: who first used it and why, the pioneer of time series analysis and when he recorded and published his efforts. All of these and more you can find in the history section.

Then comes the definition section, which presents scientific definitions for time series and time series analysis. This section also shows the importance of time series, and categorizes them into two types: **stationary** and **non-stationary**.

After definition section comes the turn of the types of time series analysis.This section shows that time series data can be classified into two main categories: stock time series data and flow time series data.Then it defines the types of time series analysis types, which include curve fitting, descriptive analysis, segmentation, forecasting, exploratory analysis, explanatory analysis, functional analysis and trend analysis. It's very important for answering business questions.

After discussing the types of time series we present their models and techniques, which are fundamentally important in various practical domains. This section shows a variety of methods to study data like decompositional models, smoothing-based models, moving-average model, exponential smoothing model, Holt-Winters method and autoregressive integrated moving average (ARIMA) model. The section includes an elaborate definition of the ARIMA model and its standard notation, and talks about Box-Jenkins ARIMA models.

Finally, the thing that should be mentioned is the applications of time series. People can make use of time series in many domains. For instance: in weather forecasting, people use time series to predict what the weather would be in the next hour, day or week. In the financial and business domain, time series and forecasting are essential processes for explaining the dynamic and influential behavior of financial markets. Other domains in which time series analysis is useful are astronomy and medicine.

Historical overview

Since the beginning of time, people have used the basics of time series analysis in order to predict and forecast events, mainly astronomical and weather phenomena. An early example is ancient Egyptians keeping records of data collected by the use of sundials and shadow clocks and using it to forecast sunrise times.

However, the first recorded and published effort that employed time series analysis was the work of J. Graunt in 1962. Although his methods were improper by mathematical standards, he is considered the pioneer of time series analysis. In his research, he compiled and analyzed the Bills of mortality (weekly mortality statistics in London) to make estimates about birth and mortality rates, and the rise and spread of certain diseases.

In the early 1800s, J. Fourier laid the mathematical foundation for time series analysis in his book "Théorie Analytique de la Chaleur", focused on infinite periodic functions. He discovered Fourier transforms, which decompose functions depending on space and time into functions depending on frequency. From then on, the field has been progressing at an accelerating rate, and many groundbreaking discoveries have been made. A notable few are mentioned below.

In a paper published in 1880, T.N. Thiele formulates a time series model that he uses to solve a problem of astronomical geodesy (the measurement of Earth's geometrical shape and size. His model represents a time series as a sum of a regression component, a Brownian motion and a white noise. He then presents a procedure to estimate the regression coefficients and predict the values of Brownian motion, that is referred to as Kalman filtering.

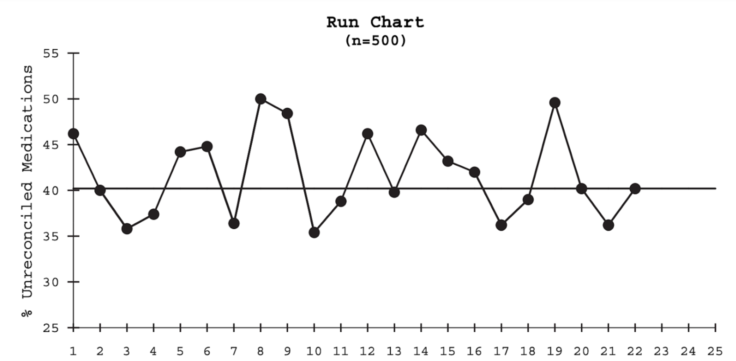
In 1999, the first successful computer tool for working with time series data, the round-robin database tool (RRDT) was developed. It facilitates storage and manipulation of a time series database, and also provides functionality to graph and visualize the data.

The Nobel Prize in Economic Sciences was awarded in 2003 to Clive W.J. Granger for his work focused on time series analysis. He presented the concept of cointegration, which solved many problems researches had been facing in their analysis of time series, including the problem of spurious regression–nonsensical apparent correlation between variables that are not actually correlated. He also discovered that the difference between a pair of integrated (nonstationary) time series could be stationary.

Definition

- Time Series

A time series is a sequence of data points (indexed or listed or graphed) collected over an interval of time. Thus, it is a sequence of discrete-time data. In time series analysis, analysts record data points at intervals over a set period of time rather than just recording the data points intermittently or randomly. A time series is very frequently plotted via a run chart, which is a temporal line chart.



Time series can show how variables change over time. In other words, time is a crucial variable because it shows how the data adjusts over the course of the data points as well as the final results. It provides an additional source of information and a set order of dependencies between the data. It typically requires a large number of data points to ensure consistency and reliability, as an extensive data set ensures you have a representative sample size and that analysis can cut through noisy data. It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance.

A time series can be of two types:

- **Stationary:** A dataset should follow the below thumb rules without having trend, seasonality, cyclical, and irregularity components of time series.
    - The MEAN value of them should be completely constant in the data during the analysis
    - The VARIANCE should be constant with respect to the time-frame
    - The COVARIANCE, which measures the relationship between two variables, should be constant between periods of identical distance
- **Non- Stationary:** This is just the opposite of a stationary time series.

- Time Series Analysis

  Time series *analysis* comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. It deals with time series data, or trend analysis. It means that data is in a series of particular time periods or intervals.

Types of Time Series Analysis

Time series analysis and forecasting models must define the types of data relevant to answering the business question. Once analysts have chosen the relevant data they want to analyze, they choose what types of analysis and techniques are the best fit. Some of these types are listed below.

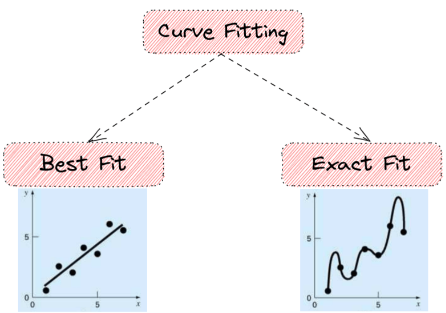
- Data classification

  This type of analysis identifies and assigns categories to the data. Time series data can be classified into two main categories:

  - **Stock time series data**, measuring attributes at a certain point in time, like a static snapshot of the information as it was.
  - **Flow time series data**, measuring the activity of the attributes over a certain period (interval), which is generally part of the total whole and makes up a portion of the results.

- Curve fitting

  This type of analysis plots the data along a curve to study the relationships of variables within the data. It is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints.

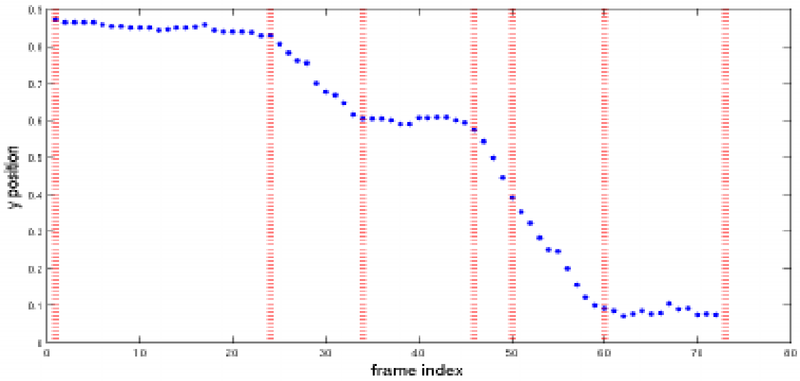
  Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are available, and to summarize the relationships among two or more variables.

- ## Descriptive analysis

This type identifies patterns in time series data, like trends, cycles, or seasonal variation. It is the process of using current and historical data to identify trends and relationships. It is sometimes called the simplest form of data analysis because it describes trends and relationships but doesn't dig deeper. Being relatively accessible, it is likely something your organization uses daily. It is especially useful for communicating change over time and uses trends as a springboard for further analysis to drive decision making.

- ## Segmentation

This type of analysis attempts to split a time series into a sequence of segments. It is often the case that a time-series can be represented as a sequence of individual segments, each with its own characteristic properties. For example, the audio signal from a conference call can be partitioned into pieces corresponding to the times during which each person was speaking. In time-series segmentation, the goal is to identify the segment boundary points in the time-series, and to characterize the dynamical properties associated with each segment.



- ## Forecasting

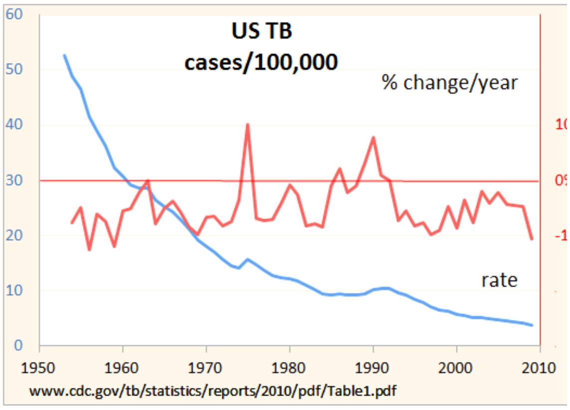
This type of analysis attempts to predict future data based on historical trends. It uses historical data as a model for future data, predicting scenarios that could happen along future plot points.

- ## Intervention analysis

This type of analysis studies how an event can change the data.

- # Exploratory analysis

Exploratory data analysis (EDA) is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. A straightforward way to examine a regular time series is manually with a line chart. A statistical model may or may not be used, but primarily, EDA is for seeing what the data can tell us beyond the formal modeling and thereby contrasts traditional hypothesis testing.



Exploratory analysis is just what it sounds like. You’re exploring the data to identify trends and outliers.

- # Explanatory analysis

This type of analysis to understand the data and the relationships within it, as well as cause and effect. Explanatory analysis is the step beyond exploratory. Instead of explaining what happened, you’re more focused on how and why it happened and what should happen next and, in most cases, communicating that to the necessary decision-makers and stakeholders.

An explanatory analysis of your Google Analytics data, for example, might include a closer look at why your top-performing pages attract so much traffic and recommendations for how they could be used to increase goal completions.

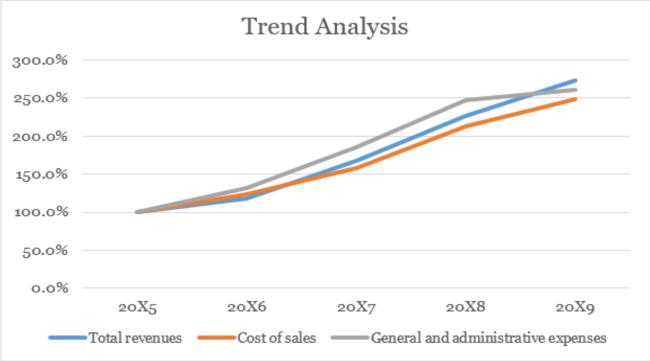
- # Functional analysis

This type can pick out the patterns and relationships within the data to identify notable events.As memory space grows, storing data becomes cheaper and cheaper, which in turn means that more and more data is stored. In case of time series, this means that data is collected more frequently.It is not clear, however, how to model time series that are recorded with a (very) high frequency, especially when (multiple) seasonality is involved. In many cases it is best to view the observations as functions of time and analyze this functional time series itself.

- Trend analysis

This type is concerned with determining consistent movement in a certain direction. There are two types of trends:

  - **Deterministic**, where we can find the underlying cause.
  - **Stochastic**, which is random and unexplainable.



Models and Techniques

Just as there are many types of analysis, there are also a variety of models and methods to study data. Here are the most common.

- Decompositional models

Time series data can exhibit a variety of patterns, so it is often helpful to split a time series into components, each representing an underlying pattern category. This is what decompositional models do.

The decomposition of time series is a statistical task that deconstructs a time series into several components, each representing one of the underlying categories of patterns. When we decompose a time series into components, we think of a time series as comprising three components: a trend component, a seasonal component, and residuals or ”noise” (containing anything else in the time series).

- Smoothing-based models

In time series forecasting, data smoothing is a statistical technique that involves removing outliers from a time series data set to make a pattern more visible. Inherent in the collection of data taken over time is some form of random variation. Smoothing data removes or reduces random variation and shows underlying trends and cyclic components.

- Exponential Smoothing model

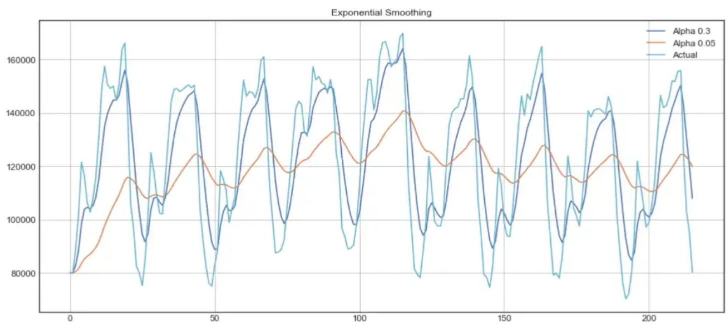
Exponential smoothing is a rule-of-thumb technique for smoothing time series data using the exponential window function. Exponential smoothing is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality. Different types of exponential smoothing include single exponential smoothing, double exponential smoothing, and triple exponential smoothing. Mathematically, exponential smoothing is expressed as

$$y = \alpha x_t + (1 - \alpha) y_{t-1}, t > 0$$

Here, *alpha* is a **smoothing factor** that takes values between 0 and 1. It determines how *fast* the weight decreases for previous observations.

There are three types of exponential smoothing methods:

- [ ] **Single Exponential Smoothing** – suitable for forecasting data with no trend or seasonal pattern, where the level of the data may change over time.
- [ ] **Double Exponential Smoothing** – for forecasting data where trends exist.
- [ ] **Triple Exponential Smoothing** – used for forecasting data with trend and/or seasonality.



From the plot above, the dark blue line represents the exponential smoothing of the time series using a smoothing factor of 0.3, while the orange line uses a smoothing factor of 0.05.

As you can see, the smaller the smoothing factor, the smoother the time series will be. This makes sense because as the smoothing factor approaches 0, we approach the moving average model.

- ## Moving-Average Model

The moving average model is probably the most naive approach to time series modeling. This model simply states that the next observation is the mean of all past observations. Although simple, this model might be surprisingly good and it represents a good starting point. Otherwise, the moving average can be used to identify interesting trends in the data. We can define a *window* to apply the moving average model to *smooth* the time series and highlight different trends.

Moving average has 3 types according to the weight of each value in the data set which are: Simple Moving average , Weighted Moving average and Exponential Moving average,

The most commonly used is Weighted Moving average as it is based on giving more weight the more recent data than that of the old data , it can be represented by the following equation:

$$\text{Weighted Moving Average} = \frac{\text{Price}_1 \times n + \text{Price}_2 \times (n-1) + \ldots\ldots\ldots\ \text{Price}_n}{[n \times (n+1)] / 2}$$

Where n : time period.

Moving average is a technical way that refers to the average of particular trading over a specific period of time either by making the weight of all values equal or making some of them less than others.

- ## Autoregressive Integrated Moving Average (ARIMA)

ARModel is another widely used forecasting technique that involves the combination of two or more time series models. This model is suitable for multivariate non-stationary data. ARIMA method is based on the concepts of autoregression, autocorrelation, and moving average. In the case of seasonal data, a variation of the model called SARIMA is applied.

SARIMA (Seasonal ARIMA) is basically an extension of ARIMA that considers the seasonal element of the time series. While ARIMA can analyze data with a trend, SARIMA supports data with both trend and seasonality. Besides the three trend factors of autoregression, difference, and moving average, SARIMA considers the three seasonal parameters for the same as well as the fourth factor for seasonal periods and seasonal period length. The advantage of this model is that it can contain many parameters and their combination.

An ARIMA model can be understood by outlining each of its components as follows:

- **Autoregression (AR)**: refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.

- **Integrated (I)**: represents the differencing of raw observations to allow for the time series to become stationary, i.e., data values are replaced by the difference between the data values and the previous values.

- **Moving average (MA)**: incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

A standard notation is used for *ARIMA(p,d,q)* where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

The parameters of the ARIMA model are defined as follows:

- First, we have the autoregressive model AR(p). This is basically a regression of the time series onto itself. Here, we assume that the current value depends on its previous values with some lag. It takes a parameter p which represents the maximum lag.
- Then, we add the moving average model MA(q). This takes a parameter q which represents the biggest lag after which other lags are not significant on the autocorrelation plot.
- After, we add the order of integration I(d). The parameter d represents the number of differences required to make the series stationary.

Real-World Applications

- Financial and Business Domain

Most financial, investment and business decisions are taken into consideration on the basis of future changes and demands forecasts in the financial domain.

Time series analysis and forecasting essential processes for explaining the dynamic and influential behavior of financial markets. Via examining financial data, an expert can predict required forecasts for important financial applications in several areas such as risk evolution, option pricing & trading, portfolio construction, etc.

- Weather Forecasting

With the time, customized weather forecasts began printed in newspapers and later on with the advancement in technology, currently forecasts are beyond the general weather conditions.

In order to conduct atmospheric measurements with computational methods for fast compilations, many governments have established thousands of weather forecasting stations around the world.

These stations are equipped with highly functional devices and are interconnected with each other to accumulate weather data at different geographical locations and forecast weather conditions at every bit of time as per requirements.

- Astronomy

One of the contemporary and modern applications where time series plays a significant role are different areas of astronomy and astrophysics,

Being specific in its domain, astronomy hugely relies on plotting objects, trajectories and accurate measurements, and due to the same, astronomical experts are proficient in time series in calibrating instruments and studying objects of their interest.

Time series data had an intrinsic impact on knowing and measuring anything about the universe, it has a long history in the astronomy domain, for example, sunspot time series were recorded in China in 800 BC, which made sunspot data collection as well-recorded natural phenomena.

- Medical Domain

  Medicine has evolved as a data-driven field and continues to contribute in time series analysis to human knowledge with enormous developments.

  It has potential research applications in the study of chronic diseases such as diabetes, hypertension, and herpes simplex. Both intervention models and multivariate models are covered, with examples illustrating the utility of time series techniques in chronic disease research. Time series modeling of a subject with diabetes before and after being placed on a regimen of chlorpropamide is used to demonstrate the potential of intervention analysis. Multivariate time series techniques are illustrated by modeling the relationship between exercise and blood glucose, and by modeling the relationship between psychosocial distress and lymphocyte subsets of the cellular immune system.

References

[1]K. Ghosh, “Feature article time series analysis: A brief history and its future challenges,” *Indian Science Cruiser*, vol. 34, no. 5, p. 22, Feb. 2021, doi: 10.24906/isc/2020/v34/i5/206994.

[2]K. D. Foote, “A Guide to Time Series Databases,” *DATAVERSITY*, Sep. 15, 2022. https://www.dataversity.net/a-guide-to-time-series-databases/ (accessed Dec. 08, 2022).

[3]I. Sutherland, “John Graunt: A Tercentenary Tribute,” *Journal of the Royal Statistical Society. Series A (General)*, vol. 126, no. 4, p. 537, 1963, doi: 10.2307/2982578.

[4]J. Korstanje, “Fourier Transform for Time Series,” *Medium*, Oct. 31, 2021. https://towardsdatascience.com/fourier-transform-for-time-series-292eb887b101 (accessed Dec. 08, 2022).

[5]S. L. Lauritzen, “Time Series Analysis in 1880: A Discussion of Contributions Made by T.N. Thiele,” *International Statistical Review / Revue Internationale de Statistique*, vol. 49, no. 3, pp. 319–331, 1981, doi: 10.2307/1402616.

[6]C. W. J. Granger, “Time Series Analysis, Cointegration, and Applications,” *The American Economic Review*, vol. 94, no. 3, pp. 421–425, 2004, Accessed: Dec. 05, 2022. [Online]. Available: https://www.jstor.org/stable/3592936

[7]N. Tyagi, “5 Applications of Time Series Analysis | Analytics Steps,” *www.analyticssteps.com*. https://www.analyticssteps.com/blogs/5-applications-time-series-analysis (accessed Dec. 08, 2022).

[8]M. Žáček, “Introduction to Time Series,” *Advances in Computational Intelligence and Robotics*, pp. 32–52, 2017, doi: 10.4018/978-1-5225-0565-5.ch002.

[9]A. Hayes, “Understanding Time Series,” *Investopedia*, Mar. 31, 2020. https://www.investopedia.com/terms/t/timeseries.asp (accessed Dec. 08, 2022).

[10]“What is Time Series Analysis?,” *TIBCO Software*. https://www.tibco.com/reference-center/what-is-time-series-analysis (accessed Dec. 08, 2022).

[11]Tableau, “Time Series Analysis: Definition, Types, Techniques, and When It’s Used,” *Tableau*, 2022. https://www.tableau.com/learn/articles/time-series-analysis (accessed Dec. 08, 2022).

[12]“(Complete Guide) Time Series Analysis: Types & Examples,” *Simplilearn*. https://www.simplilearn.com/tutorials/statistics-tutorial/what-is-time-series-analysis (accessed Dec. 08, 2022).

[13]“A Comprehensive Guide to Time Series Analysis,” *Analytics Vidhya*, Oct. 23, 2021. https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/ (accessed Dec. 08, 2022).

[14]ChannelMix, “The Difference Between Exploratory and Explanatory Analysis,” *ChannelMix*, May 20, 2021. https://channelmix.com/blog/exploratory-vs-explanatory-analysis/ (accessed Dec. 08, 2022).

[15]F. Heinrichs, “Functional Time Series,” *Medium*, Jan. 12, 2021. https://towardsdatascience.com/functional-time-series-83b717cca12 (accessed Dec. 08, 2022).

[16]K. Arora, “A Guide to Option Trading Strategies for Beginners | Analytics Steps,” *www.analyticssteps.com*. https://www.analyticssteps.com/blogs/guide-option-trading-strategies-beginners (accessed Dec. 08, 2022).