SJSU SAN JOSÉ STATE UNIVERSITY

Time Series Data

1

SJSU SAN JOSÉ STATE UNIVERSITY

Agenda

- What is Time Series Data?
- Deterministic vs Stochastic Trends
- Stationary vs Non-stationary Time Series Data
- Transformation Methods

2

SJSU SAN JOSÉ STATE UNIVERSITY

What is Time Series?

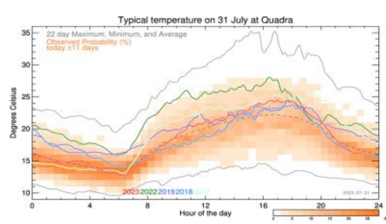
- A time series is a collection of **data points recorded at sequential time intervals**.
- The sequence captures **changes over time**, making it essential for understanding **temporal trends**, **patterns**, and **dynamics**.

3

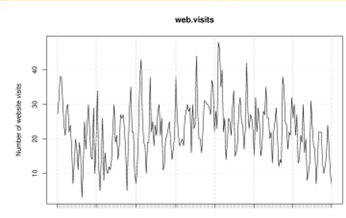
SJSU SAN JOSÉ STATE UNIVERSITY

Examples of Time Series

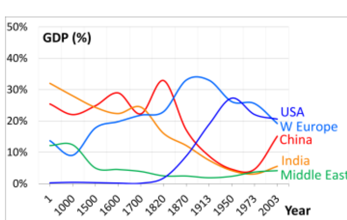
Stock Prices: Daily or minute-by-minute price movements of stocks.



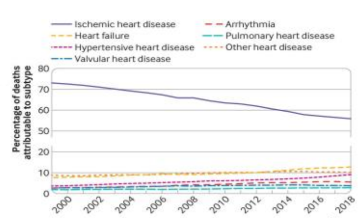
Weather Data: Daily temperature, rainfall, or humidity levels.



Website Traffic: Hourly or daily counts of website visits



Economic Indicators: Quarterly GDP, monthly unemployment rates, or annual inflation rates.



Medical Data: Heart rate or glucose levels monitored over time.

This Photo by Unknown Author is licensed under CC BY-NC-ND

4

SJSU SAN JOSÉ STATE UNIVERSITY

Characteristics of Time Series Data

Time Dependency
Each data point is dependent on the time it was recorded.

Chronological Order
Observations are always arranged in time order, which is critical for analysis.

Frequency
Data can be recorded at different intervals, such as hourly, daily, monthly, or yearly.

5

SJSU SAN JOSÉ STATE UNIVERSITY

Purpose of Time Series Analysis

Understand Historical Trends

Identify patterns like growth, decline, seasonality, or cyclic behavior in historical data.

Example: Analyzing seasonal sales trends to determine peak months for revenue.

Detect Anomalies and Patterns

Spot irregularities or unexpected events in the data.

Example: Detecting fraud in financial transactions or identifying spikes in website traffic.

Forecast Future Values

Use past data to predict future behavior.

Example: Forecasting stock prices, energy demand, or weather conditions.

Support Decision-Making

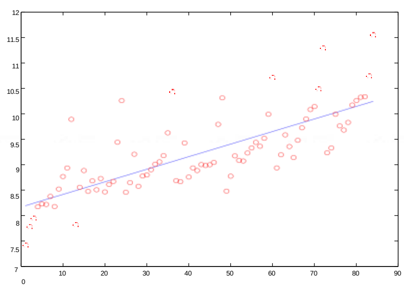
Provide actionable insights for strategic planning.
Example: Using sales forecasts to plan inventory levels or marketing strategies.

6

Key Components of Time Series Data

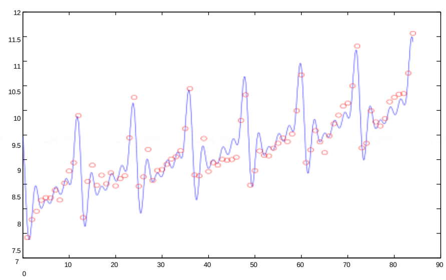
Trends

The trend represents the direction in the data, whether it's increasing, decreasing, or remaining constant.



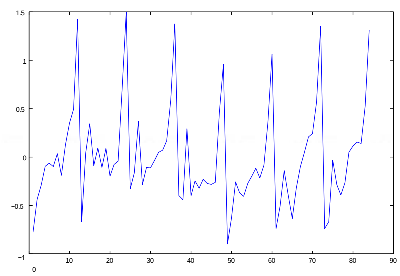
Seasonality

Seasonality refers to repeated patterns that occur at regular intervals due to factors like weather, holidays, or cyclical behavior.



Residuals

Residuals are the unexplained, random fluctuations in the data after accounting for trends and seasonality.



7

Deterministic vs. Stochastic Trends

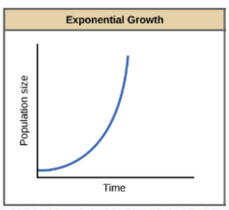
Understanding the underlying trends in time series data for accurate modeling and forecasting

8

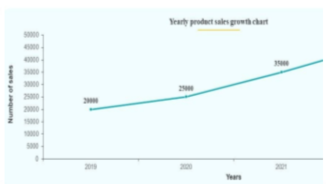
Deterministic vs. Stochastic Trends

Deterministic

- Trends follow a fixed mathematical function (e.g., linear or exponential)
- Predictable and smooth.
- Can be modeled using functions like straight lines or curves



Population growth with a clear upward exponential pattern



Sales growth following a linear increase over time.

Stochastic

- Trends involve randomness and uncertainty, influenced by external factors
- Noisy, irregular patterns
- Unpredictable in the long term, but can be analyzed probabilistically

Random Walk (e.g. daily stock prices)

9

Stationary vs. Non-Stationary Data

10

SJSU SAN JOSÉ STATE UNIVERSITY

Statistical characteristics of Time Series Data

	Mean	Variance	Autocovariance
Formula Given a time series x_t	$\mu_t = E[xt]$	$Var(x_t) = E[(x_t - \mu_t)^2]$	$\gamma_x(s,t) = \mathrm{Cov}[x_s, x_t] = E[(x_s - \mu_s)(x_t - \mu_t)]$
What it measures	The average value of a dataset	How much the data values vary (or spread out) from the mean	How the values at two different times s and t are related to each other.
Purpose	Gives a central value that summarizes the data.	Quantifies the overall variability in the data.	Helps identify time-dependent patterns or dependencies in a time series
Key Insight	Where the data is centered	How spread out the data is	How data points at different times relate to each other If $\gamma_x(s,t) > 0$: The values at s and t tend to move in the same direction (positive relationship). If $\gamma_x(s,t) < 0$: The values at s and t tend to move in opposite directions (negative relationship). If $\gamma_x(s,t) = 0$: There is no relationship between x_s, x_t
Example	Average of temperature over three days	How much the temperatures differ from the average. A higher variance means the temperatures are more spread out.	If you're looking at daily temperatures, autocovariance with lag of 1 day tells you how today's temperature is related to yesterday's.

11

SJSU SAN JOSÉ STATE UNIVERSITY

Stationarity

Stationarity refers to a property of a time series where its statistical characteristics—such as mean, variance, and autocorrelation—remain constant over time. In other words, the behavior of the series does not depend on when you observe it.

A series x_t is said to be stationary if it satisfies the following properties:

- The mean $E[x_t]$ is the same for all t.
- The variance $\mathrm{Var}(x_t)$ is the same for all t. And $\sigma(x_t) = \sqrt{\mathrm{Var}(x_t)}$ is the standard deviation
- The covariance $\gamma_x\,(t+h,t)$ between x_t and x_{t+h} is the same for all t at each time lag h = 1, 2, 3, etc or independent of t. We can write the covariance as

$$\gamma_x(h) = \gamma_x(h,0)$$

12

SJSU SAN JOSÉ STATE UNIVERSITY

Types of Stationarity

Strict/Strong Stationarity

A time series is strictly stationary if its **distribution** remains exactly the same at all points in time. This includes **mean, variance, autocovariance, skewness, and kurtosis** must be **constant over time**.

Example: Fair Coin Tosses

The process of tossing a fair coin is strictly stationary because the distribution of heads and tails does not change regardless of when or where you observe the tosses.

[1,0,1,0,1,0,1,0,1,0]

The mean is 0.5, and this would remain constant at any point in time.
The variance is 0.25, would also remain constant.
The distribution of heads and tails would not change as we shift our viewpoint in time.

This is a theoretical concept and rarely observed in real-world data.

This Photo by Unknown Author is licensed under CC BY-NC

13

SJSU SAN JOSÉ STATE UNIVERSITY

Types of Stationarity

Weak Stationarity

Weak stationarity does not require that the entire distribution remains the same. **The distribution can change, as long as the mean, variance, and autocovariance do not.**

Example: Stock Returns

Financial time series like daily stock returns are often considered weakly stationary, because the **mean** return and **variance** are often roughly constant over time (even though extreme events like crashes can change the distribution shape). The **autocovariance** (i.e., the correlation between returns over time) depends only on the lag, not on the time period itself.

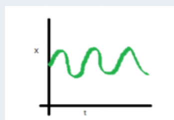
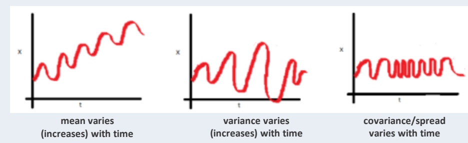
Most time series analysis methods rely on weak stationarity.

This Photo by Unknown Author is licensed under CC BY-NC

14

SJSU SAN JOSÉ STATE UNIVERSITY

Types of Time Series Data

	Stationary	Non-Stationary
Mean	Constant	Changes over time
Variance	Constant	Changes over time
Autocovariance	Constant for any given lag	Changes with time or lag
Trend	No trend (constant or cyclic patterns) 	How spread out the data is 
Examples	**Daily Temperature Deviations** Deviations from the average temperature of the month (+/- 5) **{2, -1, 1, -2, 0, 3, -3, 2}** These deviations fluctuate around zero and show no long-term trend or changing variance.	**Daily Temperatures** The actual daily temperatures often show trends or seasonality **{30°C, 32°C, 35°C, ..., 10°C, 12°C, 15°C}** Over a year, temperatures might increase in summer and decrease in winter, leading to a clear pattern or trend.

15

SJSU SAN JOSÉ STATE UNIVERSITY

Types of Time Series Data

	Stationary	Non-Stationary
Pros	• Enhances the precision of statistical model • Simpler to model and forecast • More reliable in terms of consistency over time	• Better representation of real-world data • Captures long-term trends, growth, and structural shifts. • Offers richer insights into underlying dynamics of the data.
Cons	• May not represent real-world data that shows trends • Oversimplifies complex data behaviors • May miss long-term growth or structural changes	• Harder to model and forecast without adjustments • Requires transformations (e.g., differencing, detrending)

16

SJSU SAN JOSÉ STATE UNIVERSITY

When to Convert Non-Stationary Data to Stationary

1. **You need to apply classical models** (like ARIMA) that assume stationarity
2. **The data exhibits trends** that make it unsuitable for stationary models without transformation.
3. **You want to identify relationships between variables** that are clearer after removing long-term trends or seasonality.

17

SJSU SAN JOSÉ STATE UNIVERSITY

Transformation Methods

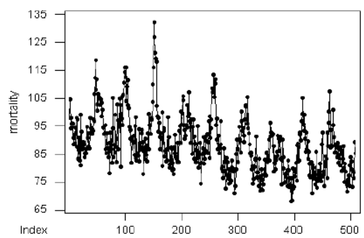
18

SJSU SAN JOSÉ STATE UNIVERSITY

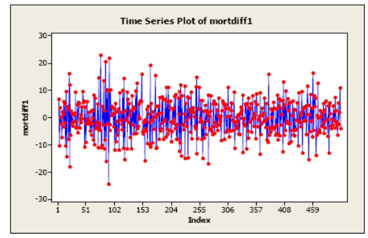
Differencing

Removes trends by computing the difference between consecutive data points

First differencing removes linear trends. (ARIMA with d>0)



$$y_t = x_t - x_{t-h}$$



Second/Seasonal differencing removes quadratic trends (SARIMA)
removes seasonal effects (e.g., subtracting values from the same season in the previous year).

$$y_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$$

Stop once the series becomes stationary (confirmed using tests like Augmented Dickey-Fuller test)

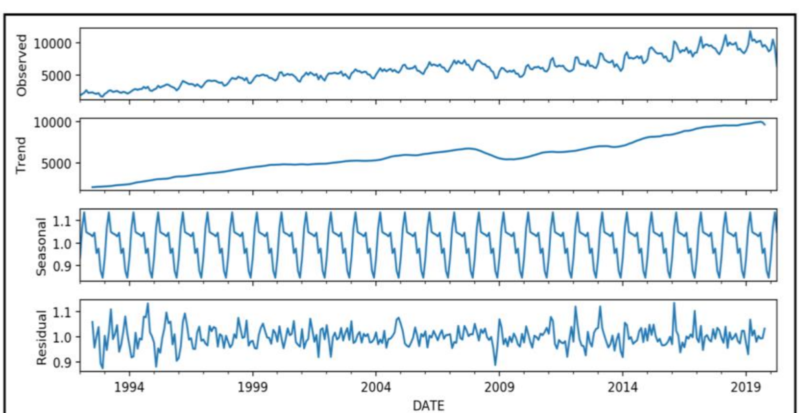
19

SJSU SAN JOSÉ STATE UNIVERSITY

Decomposition

Process of separating a time series into its key components

Trend: Long-term direction.
Seasonality: Repeated patterns over time.
Residuals: Random noise or irregular fluctuations.



Why Decompose?

- To better understand underlying patterns.
- Improve forecasting and analysis by isolating noise.
- Make accurate predictions

https://medium.com/@roshmitadey/time-series-decomposition-62cbf31ab65e

20

SJSU SAN JOSÉ STATE UNIVERSITY

Differencing OR Decomposition

	Differencing	Decomposition
Purpose	When you need to make data stationary for modeling (e.g., ARIMA).	When you need to understand the components of the data or accurate forecasting.
Complexity	The trend or seasonality is relatively simple and can be handled with basic differencing	More complex, especially for large datasets.
Example	Forecasting sales data where trends and seasonal effects are stable	Analyzing retail data with clear seasonal patterns to understand the trend, seasonality, and irregular components for better forecasting

21

SJSU SAN JOSÉ STATE UNIVERSITY

Types of Decomposition

Additive

In **additive decomposition**, the components of the time series (Trend, Seasonality, Noise) are assumed to **add up** to produce the observed values.

$$Y_t = T_t + S_t + N_t$$

Used when the seasonal fluctuations do **not change** as the overall level of the time series increases or decreases. In other words, the size of the seasonal effects remains constant over time.

Example: monthly sales data for a small shop
Trend: gradual increase in sales over time
Seasonality: may be consistent peak every December (e.g., +100 sales each December, regardless of the general trend)

Not suitable if the seasonal effects grow larger as the trend increases (e.g., in very large datasets where the magnitude of seasonality increases over time).

Multiplicative

In **multiplicative decomposition**, the components of the time series (Trend, Seasonality, Noise) are assumed to **multiply** to produce the observed values.

$$Y_t = T_t \ X \ S_t \ X \ N_t$$

Choose this when the seasonal effect **scales with the trend**. That is, as the trend increases (e.g., as sales grow), the seasonal effect also grows (e.g., the seasonal effect could be 10% of total sales rather than a fixed amount).

Example: monthly sales data for a large department store
Trend: shows a consistent growth in sales over time
Seasonality: might show that every December, sales are typically 20% higher than the average monthly sales.

More complex than additive decomposition
If seasonal changes are not proportional to the trend, the model might not fit well

22