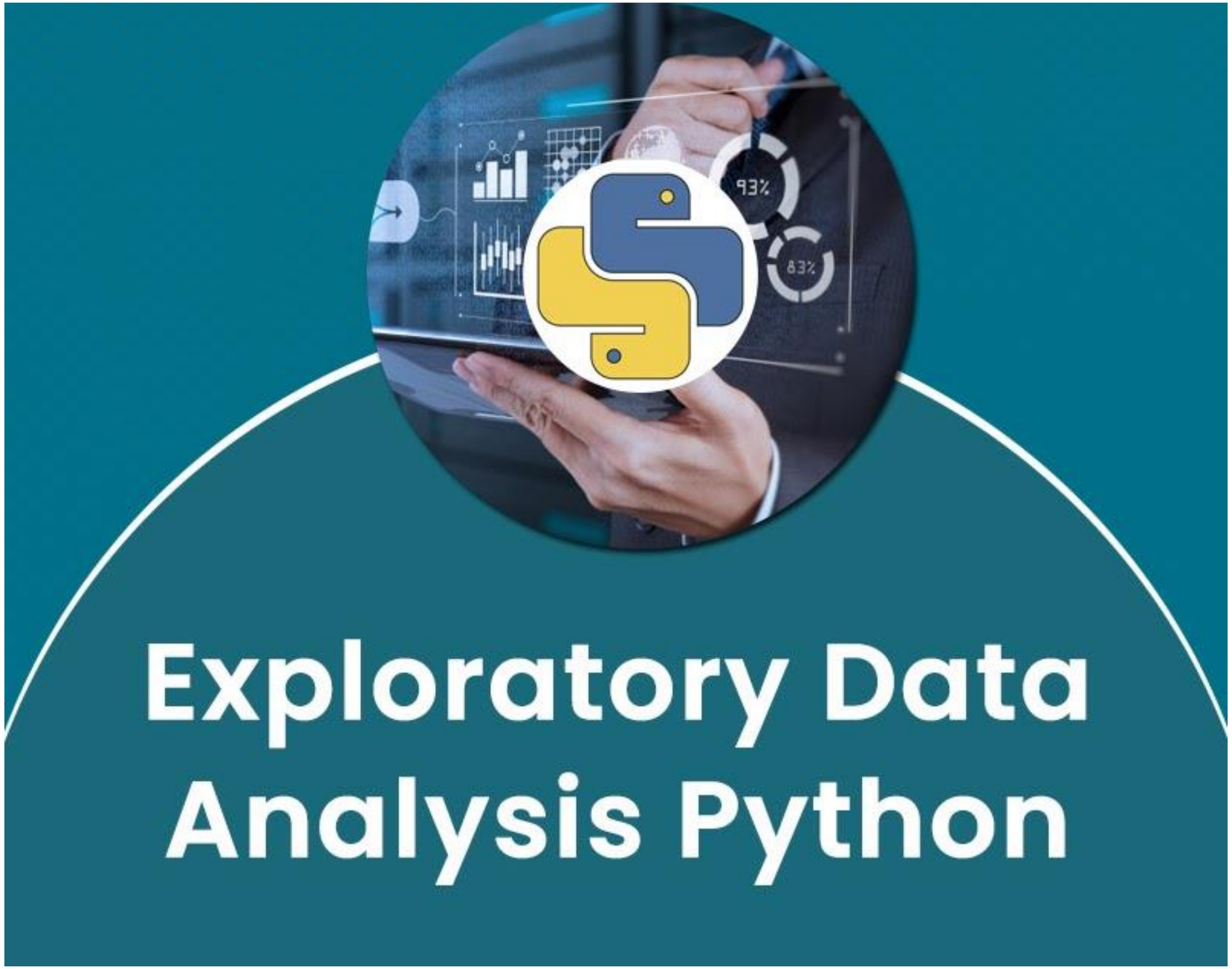
Part II:



Outline

- Measurment Functions
- Stationarity
- (Partial) Autocorrelation Function ((P)ACF):
- Random walks
- White Noise
- Time Series Decompositions and Smoothing
- Regression and Correlation
- Cross Correlation Analysis
- Multivariate Time Series Analysis
- PCA Analysis

Measurement Functions

- **Mean function**
  - The mean function is defined as

$$\mu_t = \mu_{X_t} = E[X_t] = \int_{-\infty}^{\infty} x f_t(x) dx,$$

    provided it exists, where E denotes the usual expected value operator.
- Clearly for white noise series, $\mu_{w_t} = E[w_t] = 0$ for all t.
- For random walk with drift $(\delta \neq 0)$,

$$\mu_{X_t} = E[X_t] = \delta t + \sum_{i=1}^{t} E[w_i] = \delta t$$

Autocovariance for Time Series

- Lack of independence between adjacent values in time series X_s and X_t can be numerically assessed.
- **Autocovariance Function**
  - Assuming the variance of X_t is finite, the autocovariance function is defined as the second moment product

    $$\gamma(s,t) = \gamma_X(s,t) = cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)],$$

    for all s and t.
  - Note that $\gamma(s,t) = \gamma(t,s)$ for all time points s and t.
- The autocovariance measures the linear dependence between two points on the same series observed at different times.
  - Very smooth series exhibit autocovariance functions that stay large even when the t and s are far apart, whereas choppy series tend to have autocovariance functions that are nearly zero for large separations.

Autocorrelation for Time Series

- **Autocorrelation Function (ACF)**
  - The autocorrelation function is defined as

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

  - According to Cauchy-Schwarz inequality

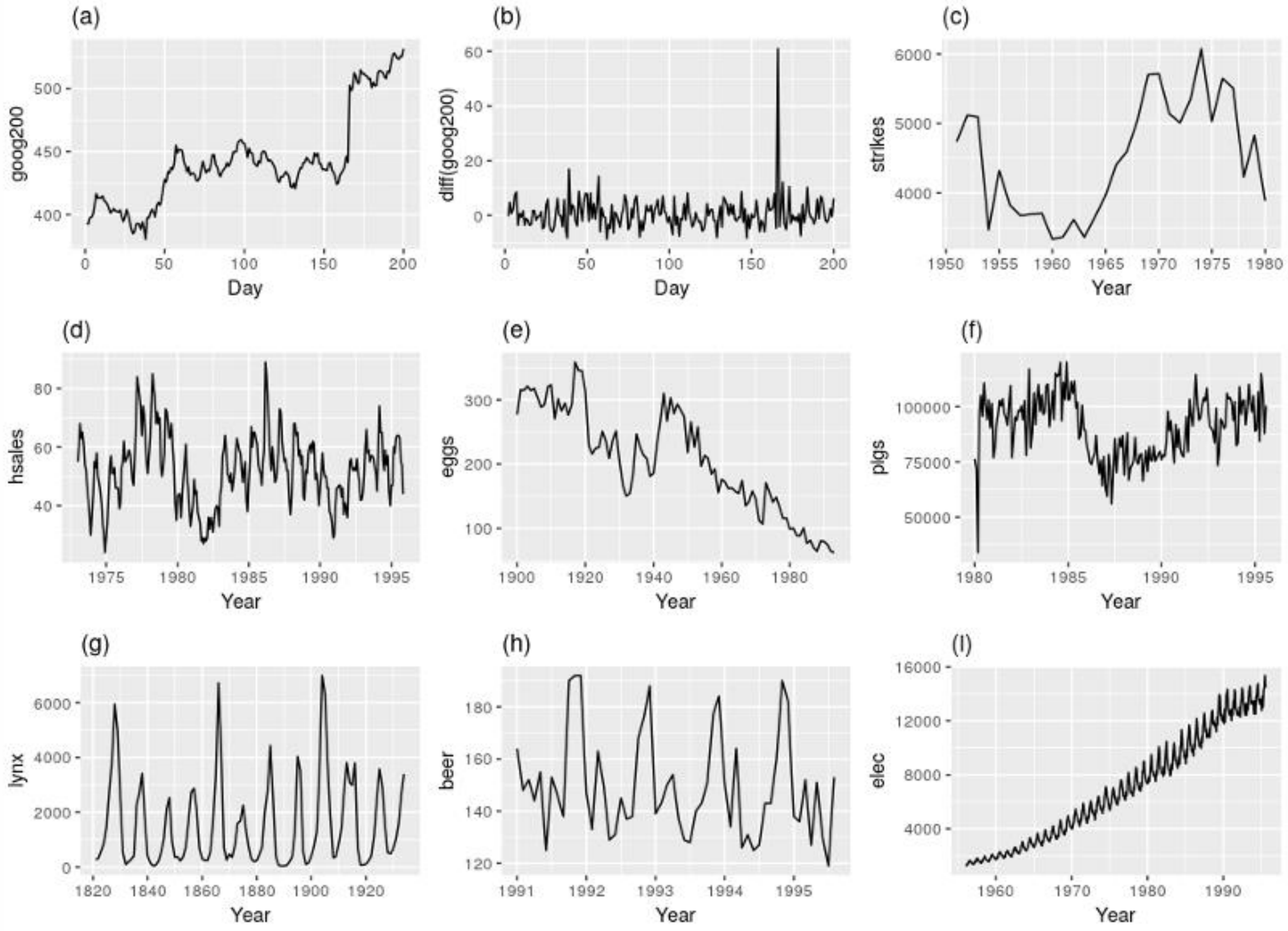
$$|\gamma(s,t)|^2 \leq \gamma(s,s)\gamma(t,t),$$

    it's easy to show that $-1 \leq \rho(s,t) \leq 1$.
- ACF measures the linear predictability of X_t using only X_s.
  - If we can predict X_t perfectly from X_s through a linear relationship, then ACF will be either $+1$ or -1.

Stationarity of Stochastic Process

- Forecasting is difficult as time series is non-deterministic in nature, i.e. we cannot predict with certainty what will occur in the future.
- But the problem could be a little bit easier if the time series is stationary: you simply predict its statistical properties will be the same in the future as they have been in the past!
  - A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time.
- Most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary after mathematical transformations.

Which of these are stationary?



Strict Stationarity

- There are two types of stationarity, i.e. strictly stationary and weakly stationary.
- **Strict Stationarity**
  - The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be strictly stationary if the joint distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ is the same as that of $(X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h})$.
  - In other words, strict stationarity means that the joint distribution only depends on the "difference" h, not the time $(t_1, t_2, \ldots, t_k)$.
- However in most applications this stationary condition is too strong.

Weak Stationarity

- **Weak Stationarity**
  - The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be weakly stationary if
    1. $E[X_t^2] < \infty, \quad \forall t \in \mathbb{Z};$
    2. $E[X_t] = \mu, \quad \forall t \in \mathbb{Z};$
    3. $\gamma_X(s,t) = \gamma_X(s+h, t+h), \quad \forall s, t, h \in \mathbb{Z}.$
  - In other words, a weakly stationary time series $\{X_t\}$ must have three features: finite variation, constant first moment, and that the second moment $\gamma_X(s,t)$ only depends on $|t-s|$ and not depends on s or t.
- Usually the term *stationary* means weakly stationary, and when people want to emphasize a process is stationary in the strict sense, they will use strictly stationary.

Remarks on Stationarity

- Strict stationarity does not assume finite variance thus strictly stationary does NOT necessarily imply weakly stationary.
  - Processes like i.i.d Cauchy is strictly stationary but not weakly stationary.
- A nonlinear function of a strictly stationary time series is still strictly stationary, but this is not true for weakly stationary.
- Weak stationarity usually does not imply strict stationarity as higher moments of the process may depend on time t.
- If time series $\{X_t\}$ is Gaussian (i.e. the distribution functions of $\{X_t\}$ are all multivariate Gaussian), then weakly stationary also implies strictly stationary. This is because a multivariate Gaussian distribution is fully characterized by its first two moments.

Autocorrelation for Stationary Time Series

University of Bremen

- Recall that the autocovariance $\gamma_X(s,t)$ of stationary time series depends on s and t only through $|s-t|$, thus we can rewrite notation $s = t + h$, where h represents the time shift.

$$\gamma_X(t+h,t) = cov(X_{t+h}, X_t) = cov(X_h, X_0) = \gamma(h,0) = \gamma(h)$$

- **Autocovariance Function of Stationary Time Series**

$$\gamma(h) = cov(X_{t+h}, X_t) = E[(X_{t+h}-\mu)(X_t-\mu)]$$

- **Autocorrelation Function of Stationary Time Series**

$$\rho(h) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t+h,t+h)\gamma(t,t)}} = \frac{\gamma(h)}{\gamma(0)}$$

Partial Autocorrelation

- Another important measure is called partial autocorrelation, which is the correlation between X_s and X_t with the linear effect of "everything in the middle" removed.
- **Partial Autocorrelation Function (PACF)**
  - For a stationary process X_t, the PACF (denoted as ϕ_{hh}), for $h = 1, 2, \ldots$ is defined as

$$\phi_{11} = \text{corr}(X_{t+1}, X_t) = \rho_1$$
$$\phi_{hh} = \text{corr}(X_{t+h} - \hat{X}_{t+h}, X_t - \hat{X}_t), \quad h \geq 2$$

  where $\hat{X}_{t+h}$ and $\hat{X}_t$ is defined as:

$$\hat{X}_{t+h} = \beta_1 X_{t+h-1} + \beta_2 X_{t+h-2} + \cdots + \beta_{h-1} X_{t+1}$$
$$\hat{X}_t = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \cdots + \beta_{h-1} X_{t+h-1}$$

  - If X_t is Gaussian, then ϕ_{hh} is actually conditional correlation

$$\phi_{hh} = \text{corr}(X_t, X_{t+h} | X_{t+1}, X_{t+2}, \ldots, X_{t+h-1})$$

White Noise

- What is white noise
- Different aspects of white noise
- Types, features, and advantages of white noise
- why it is important in time series analysis

Types
- Gaussian white noise
- Uniform white noise

Features
- Mean of Zero
- Constant Variance
- Independence
- Randomness

Advantages
- Simplicity
- Modeling Power
- Stationarity

White Noise Test (Box-Pierce and Ljung-Box Tests)

- **Box-Pierce Test (1970):**

$$Q_{BP}(m) = n\sum_{k=1}^{m} r_k^2$$

$$H_0 : \rho_1 = \cdots = \rho_m = 0 \qquad H_1 : \rho_i \neq 0 \text{ for some } i \in \{1 : m\}$$

where $1 \leq m < T$ is any given integer and T is the sample size.

Under the null hypothesis that $\{X_t\}$ is a white noise (H_0 is true), $Q_{BP}(m)$ asymptotically follows a chi-squared distribution $\chi^2(m)$ with m degrees of freedom.

- **Ljung-Box Test (1978):**

$$Q_{LB}(m) = n(n+2)\sum_{k=1}^{m} \frac{r_k^2}{n-k}$$

which still asymptotically and better follows the chi-squared distribution $\chi^2(m)$

- **P-value Interpretation:** Note that under the null hypothesis, for every integer $1 \leq m < T$, the p-value for $Q_{LB}(m)$ should be greater than 0.05 (the level of significance).

Statistical Definition of Random Walk

Definition A time series $\{X_t\}$ is called a random walk if it satisfies the following equation

$$\boxed{X_t = X_{t-1} + W_t}$$

where $\{W_t\}$ is a white noise and, for all t, W_t and X_{t-1} are uncorrelated.

we can easily obtain

$$\boxed{X_t = X_{t-1} + W_t = X_{t-2} + W_{t-1} + W_t = \cdots = X_0 + W_1 + W_2 + \cdots + W_{t-1} + W_t}$$

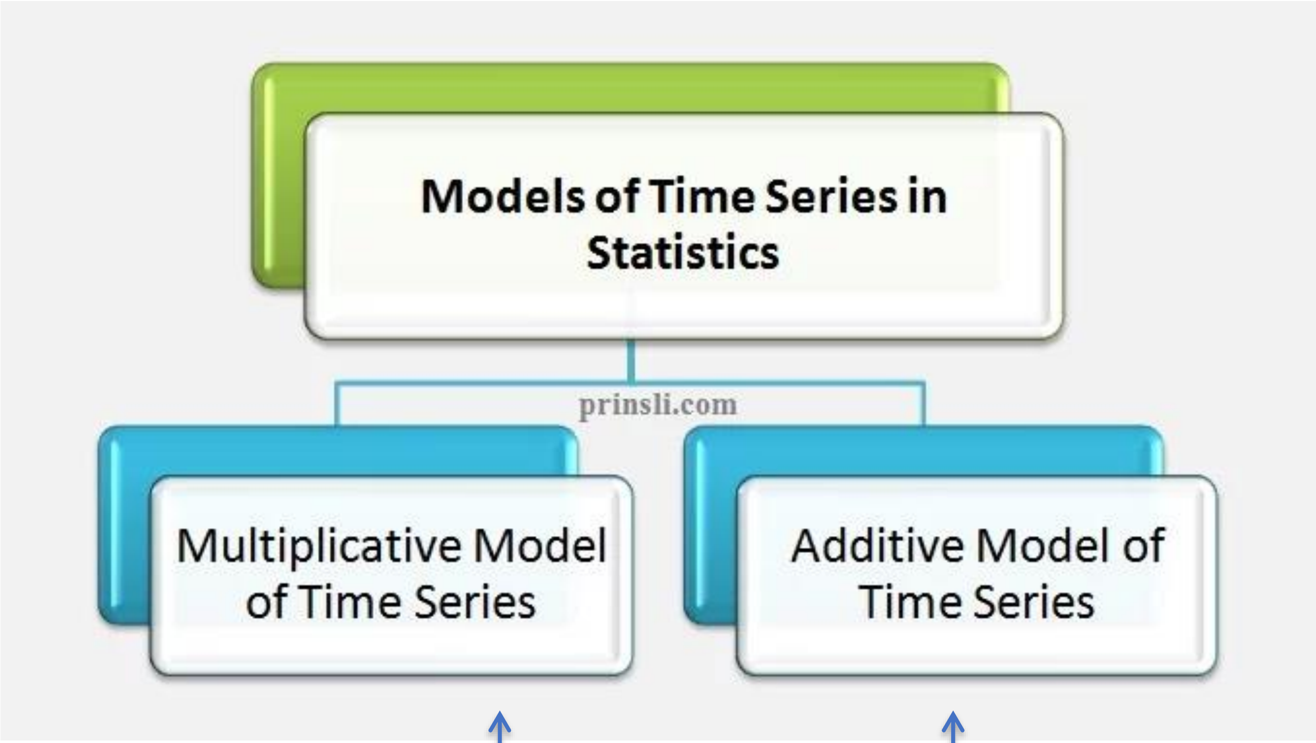
Therefore, for all t, $\mathrm{E}(X_t) = \mathrm{E}(X_0)$ is a constant. That is, the random walk is mean stationary.

On the other hand,

$$\boxed{\mathrm{Var}(X_t) = \mathrm{Var}(X_{t-1}) + \sigma_w^2 > \mathrm{Var}(X_{t-1})}$$

Thus it can be seen that the random walk is not variance stationary.

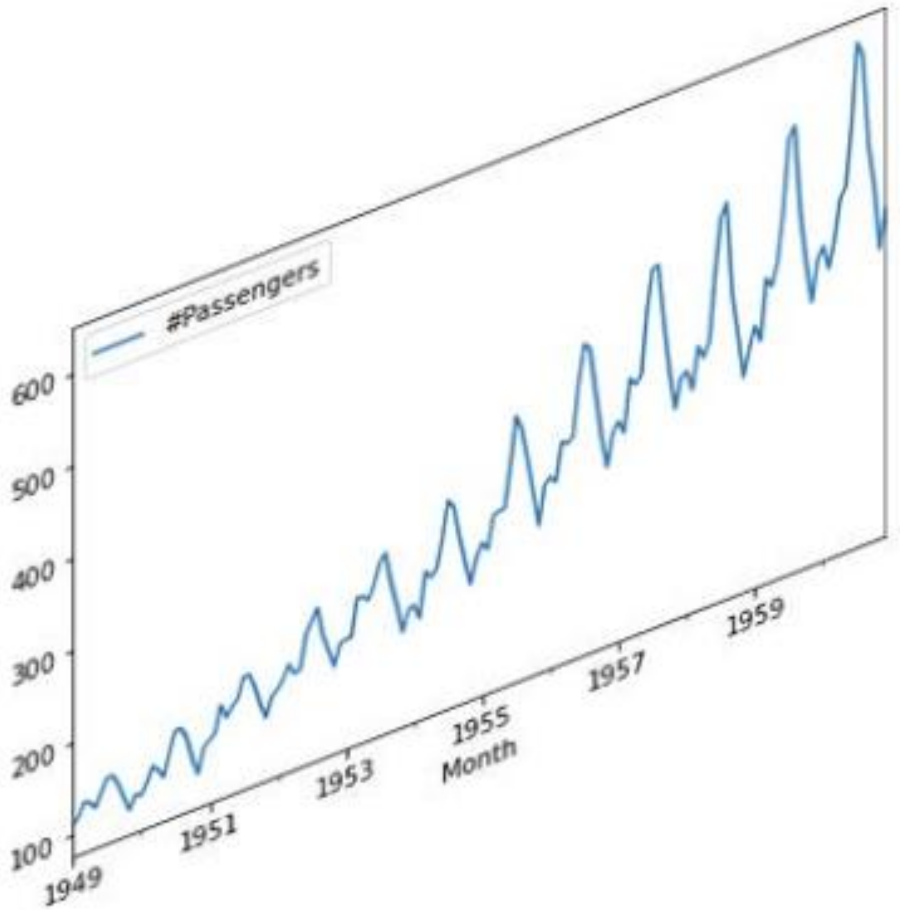
Additive vs. Multiplicative Time Series Models

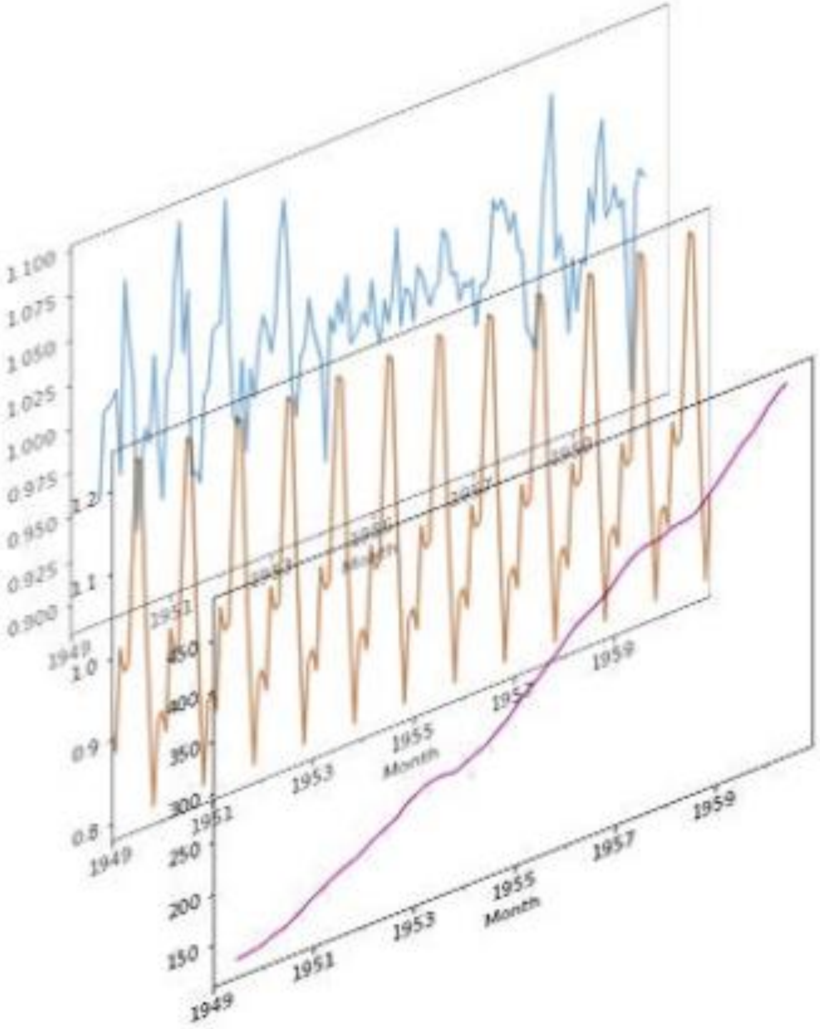


$$Y(t) = T(t) \times S(t) \times C(t) \times I(t)$$

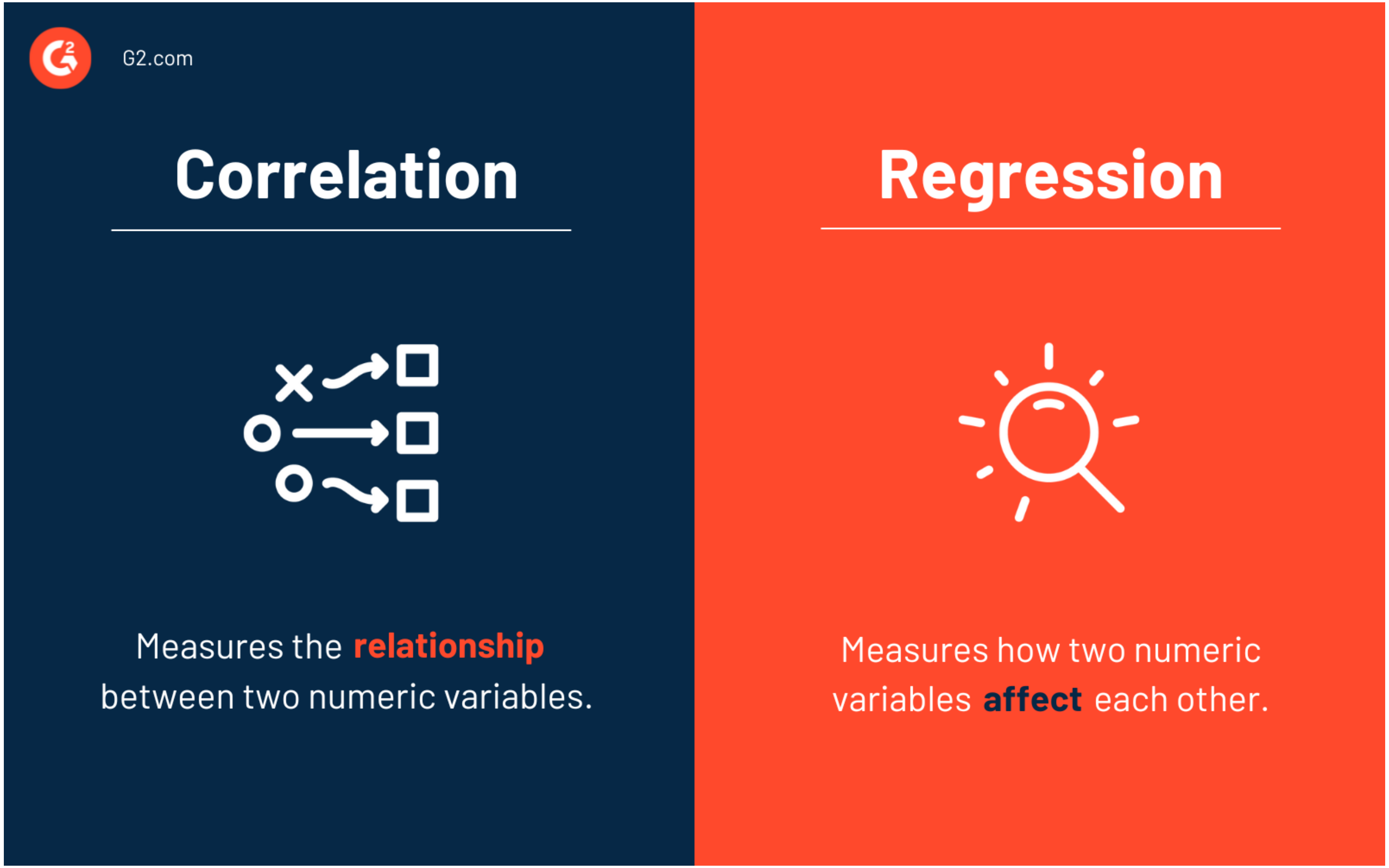
$$Y(t) = T(t) + S(t) + C(t) + I(t)$$

Time Series Decomposition





Correlation and Regression



What is Linear Correlation Coefficient?

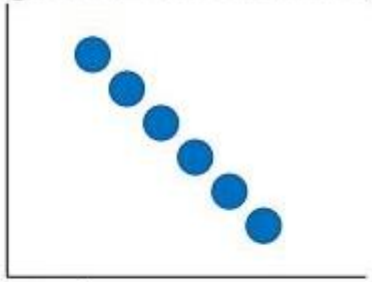
- A measure of the strength and the direction of a linear relationship between two variables.
- ***r*** represents the sample correlation coefficient.
- ***ρ*** (rho) represents the population correlation coefficient

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}}$$

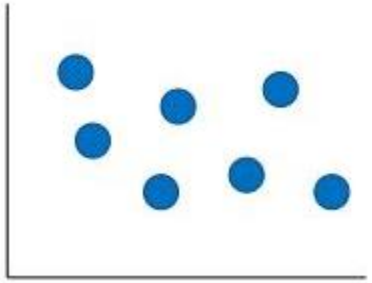
n is the number of data pairs

- The range of the correlation coefficient is -1 to 1.

If ***r* = -1** there is a perfect negative correlation

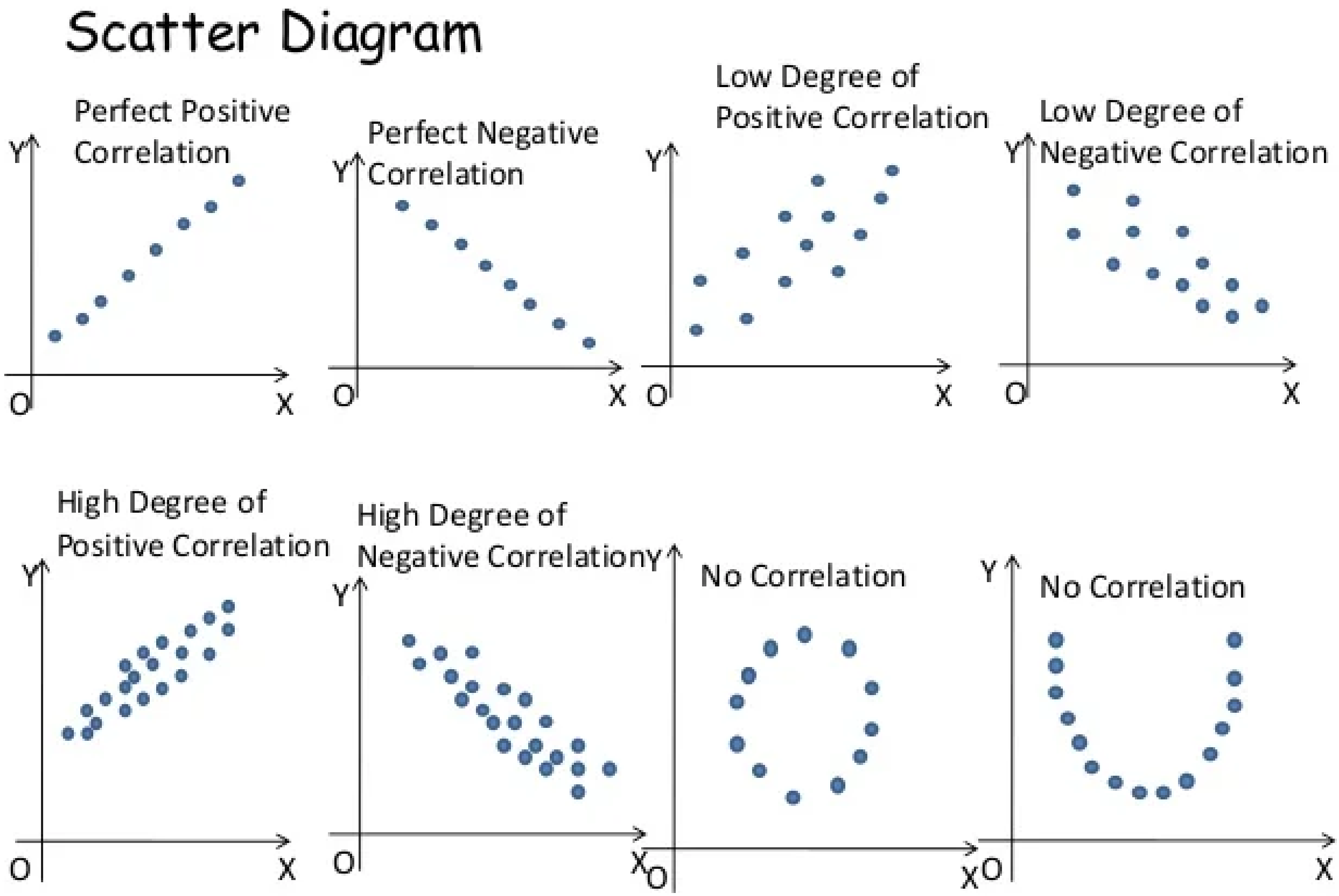


If ***r* is close to 0** there is no linear correlation



If ***r* = 1** there is a perfect positive correlation

Visualizing Linear Correlation with Scatter Plots



Introduction to Regression Line

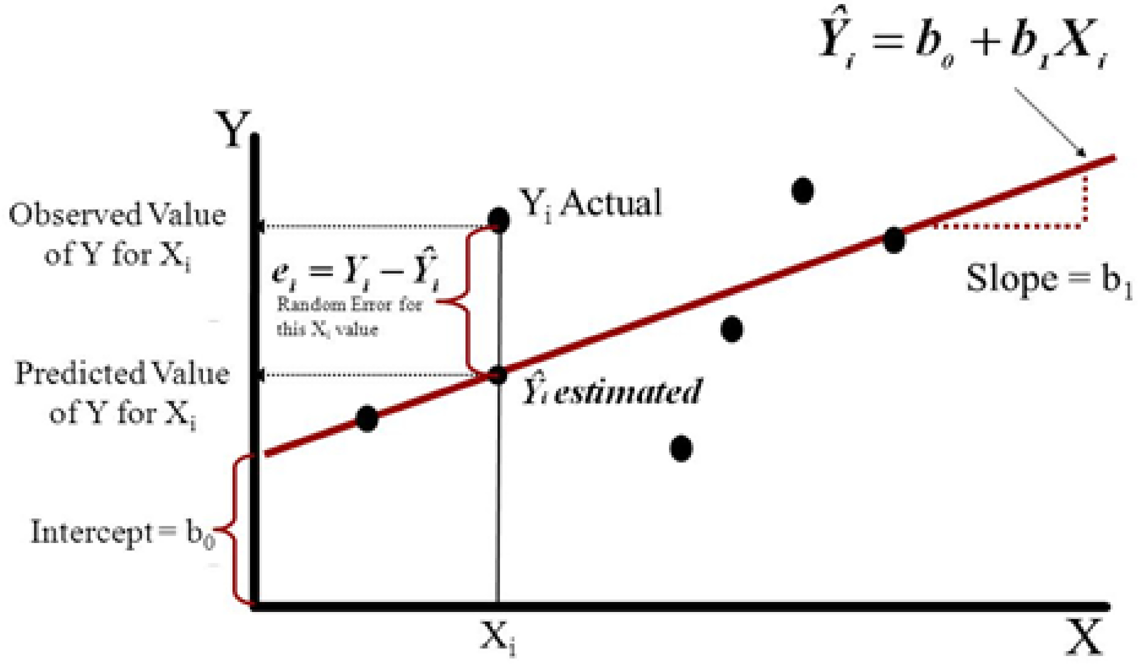
- After verifying that the linear correlation between two variables is significant, next we determine the equation of the line that best models the data (**regression line**).
- Can be used to predict the value of *y* for a given value of *x*.

Formulas for b_0 and b_1

$$b_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

Simple Linear Regression Model



What is Cross-Correlation?

- **Definition:** Cross-correlation measures the similarity between two time series as a function of the lag of one relative to the other.

- **Purpose:** Identifies how one time series is correlated with another over different time lags.

- **Applications in Environmental Science:**

➢ **Temperature and CO_2 levels:** Analyze how temperature changes are related to CO_2 concentration over time.

➢ **Rainfall and River Flow:** Measure how rainfall patterns influence river flow at different time lags.

Multivariate Time Series

Definition: A multivariate time series consists of multiple time-dependent variables observed simultaneously over time.

Purpose: Unlike univariate time series, which looks at a single variable, multivariate time series allows us to analyze the interactions between multiple variables over time.

Examples:

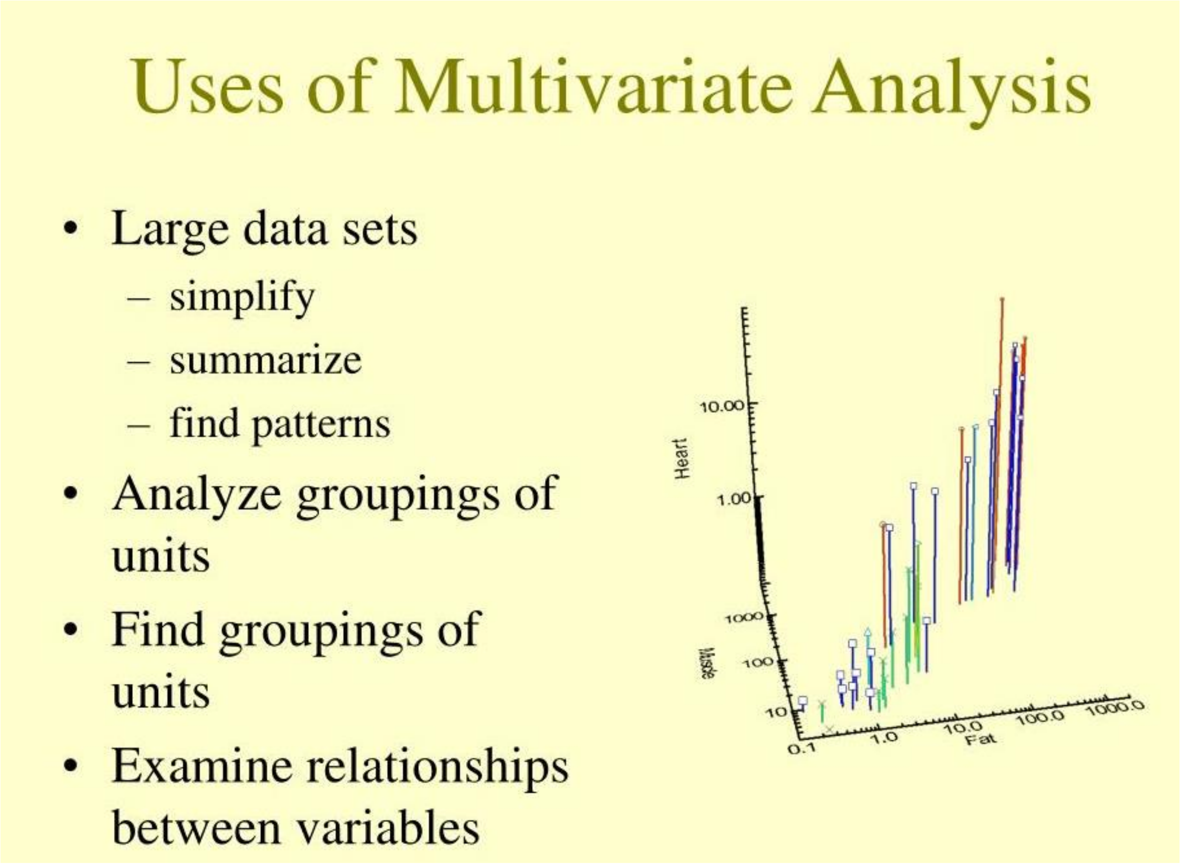
- **Environmental Science:** Studying how temperature, humidity, and wind speed change over time to predict weather conditions.
- **Economics:** Tracking GDP, inflation, and unemployment rates over time to understand economic trends.

Applications and Benefits of Multivariate Time Series

Key Benefits:

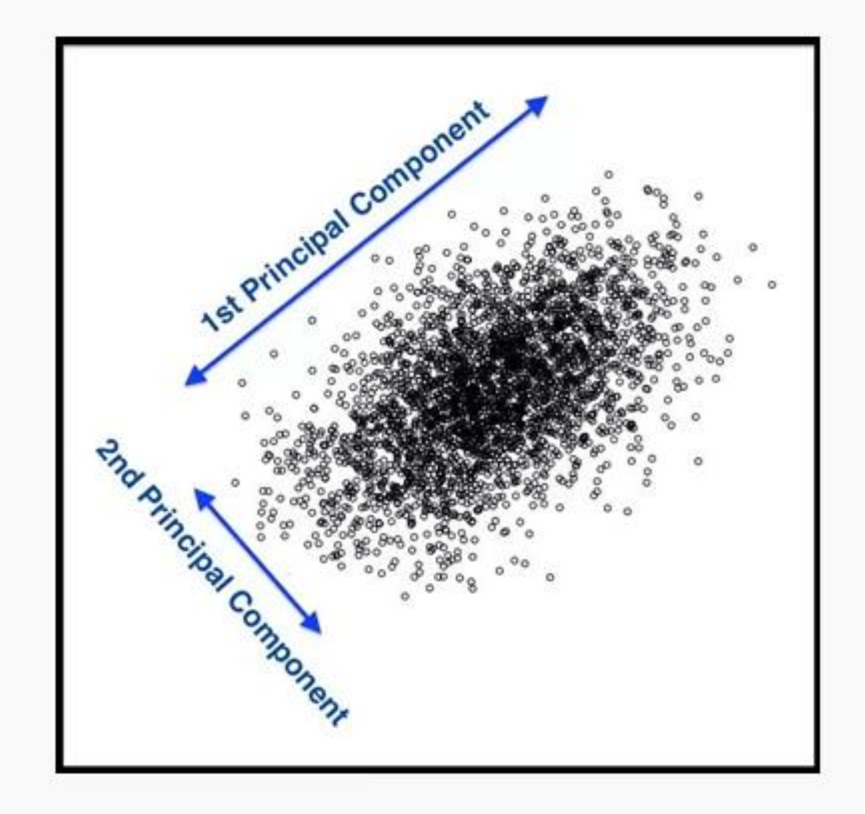
1) **Capture Interdependencies**: Analyze relationships between multiple variables (e.g., temperature affecting humidity).
2) **Improved Forecasting**: More accurate predictions due to the inclusion of multiple factors.
3) **Advanced Models**: Uses models like Vector Autoregression (VAR), which account for interactions between variables.

Example: Predicting energy consumption using temperature, time of day, and population data together.



What is Principal Component Analysis (PCA)?

1) A statistical technique used for dimensionality reduction.
2) Transforms original features into a new set of uncorrelated variables called principal components.
3) Aims to capture the maximum variance in the data with the fewest number of components.



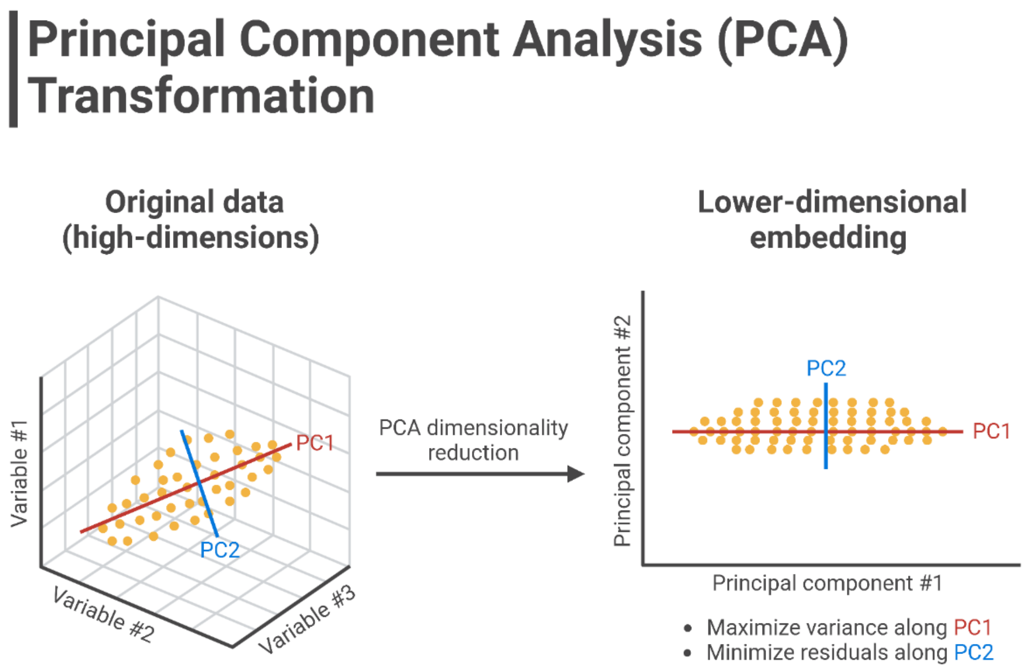
How PCA Works

1) **Standardization:** Scale the data to have a mean of zero and a standard deviation of one.

2) **Covariance Matrix:** Calculate the covariance matrix to understand how features vary together.

3) **Eigenvalues and Eigenvectors:** Compute the eigenvalues and eigenvectors of the covariance matrix.

4) **Principal Components:** Select the top k eigenvectors (principal components) based on the highest eigenvalues.



Benefits of PCA

1) **Dimensionality Reduction:** Reduces the number of features, simplifying models and improving performance.
2) **Noise Reduction:** Helps to filter out noise from the data.
3) **Data Visualization:** Enables visualization of high-dimensional data in 2D or 3D plots.
4) **Improved Model Performance:** Can lead to faster training times and better generalization.

GO VS