 # stationary and **non-stationary** series have distinct characteristics, particularly in terms of statistical properties over time. Here are the key differences between the two:

1. Mean and Variance

- **Stationary**: The mean and variance of the time series remain constant over time. The data fluctuates around a constant mean.
- **Non-Stationary**: The mean and/or variance changes over time. There may be trends or varying levels of volatility.

2. Trend

- **Stationary**: There is no long-term trend in the data; the values oscillate around a constant mean level.
- **Non-Stationary**: The data often exhibits a trend, either increasing or decreasing over time.

3. Autocovariance

- **Stationary**: The autocovariance function depends only on the lag between observations, not on the actual time points. This makes it easier to model and predict.
- **Non-Stationary**: The autocovariance depends on the time at which it is measured, making it more complex to model because the relationships between values can change over time.

4. Differencing

- **Stationary**: Differencing is usually not required because the series is already stable around a constant mean.
- **Non-Stationary**: Differencing is often used to transform the data into a stationary series, particularly when the data shows a trend or varying variance.

5. Seasonality

- **Stationary**: The series may exhibit constant periodic patterns or seasonality, but these patterns do not grow or decay over time.
- **Non-Stationary**: The seasonal patterns may change in amplitude or frequency over time, making it harder to capture using simple models.

6. Predictability

- **Stationary**: Easier to predict future values because past patterns tend to persist.
- **Non-Stationary**: Harder to predict, especially without transformations, because trends and other non-stationary elements (e.g., shocks) can affect future values.

7. Modeling

- **Stationary**: Often modeled using autoregressive (AR), moving average (MA), or ARMA models.
- **Non-Stationary**: Typically requires transforming the series (e.g., through differencing or logarithmic transformation) before applying time series models like ARIMA.

To apply ARIMA models or other time series techniques that require stationarity, non-stationary time series data must first be converted into a **stationary** form. This process ensures that the statistical properties, such as the mean and variance, remain constant over time, making the data suitable for modeling. There are three main methods to achieve stationarity: **detrending**, **differencing**, and **transformation**.

Detrending

Detrending involves removing a **trend** component from the time series. A trend represents a long-term increase or decrease in the data over time, which can distort the ability to model the series properly.

1. Detrending With a Constant Model

This is the simplest method of detrending, which assumes that the trend is a straight horizontal line. The trend is removed by subtracting the **mean** of the time series from each data point.

- **Assumption**: The time series has no upward or downward trend, and the mean is constant over time.

Example:

If you have a time series $Y(t) = [5, 7, 6, 8, 7]$, the mean of this series is:

$$\mu = \frac{5+7+6+8+7}{5} = 6.6$$

Detrended series:

$$Y'(t) = Y(t) - \mu = [5-6.6, 7-6.6, 6-6.6, 8-6.6, 7-6.6] = [-1.6, 0.4, -0.6, 1.4, 0.4]$$

2. Detrending With a Linear Model

A linear trend occurs when the time series has an increasing or decreasing trend over time. In this method, we fit a straight line (linear regression) to the data and subtract it to remove the linear trend.

- **Assumption**: The time series has a linear trend, meaning it increases or decreases at a constant rate over time.

Example:

Suppose we have a time series with a linear trend:

$$Y(t) = [2, 4, 6, 8, 10]$$

Here, the trend is increasing linearly by 2 at each time step. We fit a line to this series and subtract it to get the detrended series. The line equation might look like $Y_t = 2t$.

Detrended series:

$$Y'(t) = Y(t) - (2t)$$

If you subtract this trend from the original series, you will get a flat, detrended series: $[0, 0, 0, 0, 0]$.

3. Detrending With a Quadratic Model

If the time series exhibits a non-linear trend, like a **curved (quadratic) trend**, then a quadratic model is more suitable for detrending. This method fits a second-degree polynomial (a parabola) to the data and subtracts it.

- **Assumption**: The time series has a non-linear, quadratic trend (curved).

Example:

Suppose the time series $Y(t) = [1, 4, 9, 16, 25]$ follows a quadratic trend $Y_t = t^2$. To remove this quadratic trend, we fit a quadratic model and subtract it.

Detrended series:

$$Y'(t) = Y(t) - t^2 = [1 - 1, 4 - 4, 9 - 9, 16 - 16, 25 - 25] = [0, 0, 0, 0, 0]$$

4. Detrending With a Moving Average Model

A **moving average model** smooths the data by calculating the average value over a sliding window of fixed size. The smoothed series is subtracted from the original data to remove the trend. This method is useful when the trend is not simple but varies over time.

- **Assumption**: The time series has a trend that changes over time, but you can average out local fluctuations using a moving window.

Example:

Given a time series $Y(t) = [1, 3, 5, 7, 9, 11, 13]$, you can apply a moving average with a window size of 3:

- Moving averages: $[\frac{1+3+5}{3}, \frac{3+5+7}{3}, \frac{5+7+9}{3}, \frac{7+9+11}{3}, \frac{9+11+13}{3}] = [3, 5, 7, 9, 11]$.

Detrended series:

$$Y'(t) = [1-3, 3-5, 5-7, 7-9, 9-11] = [-2, -2, -2, -2, -2]$$

Differencing is a method used in time series analysis to make a non-stationary series stationary by removing trends or seasonal components. The main idea behind differencing is to transform a time series by subtracting the current value from the previous value, thus focusing on the changes between consecutive observations rather than the absolute values.

Why Differencing is Important:

- **Stationarity**: Many time series models, such as ARIMA, require the series to be stationary. Differencing helps stabilize the mean and reduces trends or seasonality, making the data easier to model.
- **Removing Trends**: By differencing, you eliminate the systematic upward or downward trends that may be present.
- **Reducing Seasonality**: Seasonal patterns can also be addressed by differencing at a lag corresponding to the seasonal period.

Types of Differencing

1. **First-Order Differencing:**

   - The most basic form of differencing.
   - It involves subtracting the current observation from the previous one.

   $$Y_t' = Y_t - Y_{t-1}$$

   where Y_t is the value at time t, and Y_{t-1} is the value at time $t - 1$.

   This transformation eliminates linear trends and focuses on the changes between consecutive time steps.

2. **Second-Order Differencing:**

   - If the first-order differencing is not enough to achieve stationarity, you can apply second-order differencing.
   - This means taking the difference of the first differenced series.

   $$Y_t'' = (Y_t' - Y_{t-1}') = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$$

   This can remove more complex trends, such as quadratic trends, from the data.

3. **Seasonal Differencing:**

   - Seasonal differencing is useful for removing seasonal effects.
   - It involves subtracting the observation from the same season in the previous cycle.

   $$Y_t' = Y_t - Y_{t-s}$$

   where s is the seasonal period (e.g., 12 for monthly data with yearly seasonality).

AIC:

Akaike Information Criterion (AIC):

Purpose:

AIC helps evaluate the balance between a model's goodness of fit and its complexity. It's a tool for model selection that penalizes models with too many parameters to prevent overfitting, ensuring that the model generalizes well to new data.

Primary Formula:

$$AIC = -2\ln(L) + 2k$$

- L: Maximum log-likelihood, which measures how well the model fits the observed data (higher log-likelihood indicates a better fit).
- k: Number of parameters in the model.

In this formulation:

- **Fit:** The term $-2\ln(L)$ assesses how well the model captures the variance in the data. A higher likelihood means better fit, so this part aims to minimize errors.
- **Complexity:** The $2k$ term penalizes the number of parameters, discouraging overly complex models that may fit the current data well but perform poorly on new data.

Alternative Formula (from the image):

$$AIC = N \cdot \ln\left(\frac{SS_e}{N}\right) + 2K$$

- N: Number of observations in the dataset.
- SS_e: Sum of squared errors, representing the residual variance (lower SS_e means better fit).
- K: Number of parameters in the model.

This alternative form gives the same idea:

- **Model Fit:** The term $\ln\left(\frac{SS_e}{N}\right)$ reflects the error in model predictions. Lower residuals indicate better fit.
- **Penalty for Complexity:** The $2K$ term penalizes adding more parameters to prevent overfitting.

Interpretation:

- **Lower AIC is better**: The model with the smallest AIC is preferred, as it strikes the best balance between fitting the data and avoiding unnecessary complexity.
- **Penalty for additional parameters**: The $2k$ (or $2K$) term ensures that models are not overly complex. Even if a model improves in fit with more parameters, the increase in complexity must be justified.