# Day 57

## DIY Solution

### Q1. What is Stationary Time Series?

**Answer:** Many traditional statistical time series models rely on a time series that is stationary. Generally speaking, a stationary time series has relatively stable statistical properties over time, particularly concerning mean and variance. This seems straightforward.

Nonetheless, stationarity can be a slippery concept, mainly when applied to actual time series data. It is too intuitive and easy to fool yourself into relying on your natural intuition.

A stationary time series is one in which a time series measurement reflects a system in a steady state. Sometimes it is difficult to assert what this means, and it can be easier to rule things out as not being stationary rather than saying something is stationary.

### Q2. What is a White Noise Time Series? Why does it matter?

**Answer:** A time series is white noise if the variables are independent and identically distributed with a mean of zero.

This means that all variables have the same variance ($\sigma^2$), and each value has a zero correlation with all other values in the series.

If the variables in the series are drawn from a Gaussian distribution, the series is called Gaussian white noise.

White noise is an essential concept in time series for analysis and forecasting.

It is essential for two main reasons:

**Predictability:** If your time series is white noise, then, by definition, it is random. You cannot reasonably model it and make predictions.

**Model Diagnostics:** The series of errors from a time series forecast model should ideally be white noise.

Model Diagnostics is an essential area of time series forecasting.

Time series data are expected to contain some white noise component on top of the signal generated by the underlying process.

For example:

$y(t)$ = signal($t$) + noise($t$)

Once a time series forecast model has made predictions, they can be collected and analyzed. The series of forecast errors should ideally be white noise.

When forecast errors are white noise, the model harnesses all the signal information in the time series to make predictions. All that is left is the random fluctuations that cannot be modeled.

A sign that model predictions are not white noise is an indication that further improvements to the forecast model may be possible.

## Q3. What is Seasonality?

**Answer:** Seasonality is a characteristic of a time series in which the data experiences regular and predictable changes that recur every calendar year. Any predictable fluctuation or pattern that recurs or repeats over one year is said to be seasonal.

Seasonal effects are different from cyclical effects, as seasonal cycles are observed within one calendar year. In contrast, cyclical effects, such as boosted sales due to low unemployment rates, can span time periods shorter or longer than one calendar year.

Seasonality refers to periodic fluctuations in specific business areas and regular cycles based on a particular season. A season may refer to a calendar season such as summer or winter, or it may refer to a commercial season such as the holiday season.

## Q4. What is Non-Stationary Time Series Data? Explain the types of Non-Stationary Processes.

**Answer:** Data points are often non-stationary or have means, variances, and covariances that change over time. Non-stationary behaviors can be trends, cycles, random walks, or combinations of the three.

Non-stationary data, as a rule, are unpredictable and cannot be modeled or forecasted. The results obtained by using non-stationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist. The non-stationary data must be transformed into stationary data to receive consistent, reliable results.

**Pure Random Walk**: ($Y_t = Y_{t-1} + \varepsilon_t$ ) Random walk predicts that the value at a time "$t$" will be equal to the last period value plus a stochastic (non-systematic) component that is a white noise, which means $\varepsilon_t$ is independent and identically distributed with mean "0" and variance "$\sigma^2$." The random walk can also be named a process integrated of some order, a unit root, or a process with a stochastic trend. It is a non-mean-reverting process that can either move away from the mean positively or negatively. Another characteristic of a random walk is that the variance evolves and goes to infinity as time goes to infinity; therefore, a random walk cannot be predicted.

**Random Walk with Drift**: ($Y_t = \alpha + Y_{t-1} + \varepsilon_t$) If the random walk model predicts that the value at a time "$t$" will equal the last period's value plus a constant, or drift ($\alpha$), and a white noise term ($\varepsilon_t$), then the process is a random walk with a drift. It also does not revert to a long-run mean and has variance dependent on time.

**Deterministic Trend**: ($Y_t = \alpha + \beta_t + \varepsilon_t$) Often, a random walk with drift is confused for a deterministic trend. Both include drift and a white noise component, but the value at time "$t$" in the case of a random walk is regressed on the last period's value ($Y_{t-1}$), while in the case of a deterministic trend, it is regressed on a time trend ($\beta_t$). A non-stationary process with a deterministic trend has a mean that grows around a fixed trend, which is constant and independent of time.

**Random Walk with Drift and Deterministic Trend**: ($Y_t = \alpha + Y_{t-1} + \beta + \varepsilon_t$ ) Another example is a non-stationary process that combines a random walk with a drift component ($\alpha$) and a deterministic trend ($\beta_t$). It specifies the value at time "*t*" by the last period's value, a drift, a trend, and a stochastic component.

## Q5. What is an Augmented Dickey-Fuller Test, and why do we use it?

### Answer:

A simple definition of a stationary process is the following: a process is stationary if, for all possible lags, *k*, the distribution of $y_t$, $y_{t+1}$,..., $y_{t+k}$, does not depend on *t*.

Statistical tests for stationarity often come down to whether there is a unit root—whether 1 is a solution of the process's characteristic equation. A linear time series is nonstationary if there is a unit root, although the lack of a unit root does not prove stationarity. Addressing stationarity as a general question remains tricky and determining whether a process has a unit root remains a current area of research.

Nonetheless, a simple intuition for what a unit root is can be gleaned from the example of a random walk:

$Y_t = \varphi \times y_{t-1} + \varepsilon_t$

In this process, the value of a time series at a given time is a function of its value at the immediately preceding time and some random error. If $\varphi$ is equal to 1, the series has a unit root, will "run away", and will not be stationary. Interestingly, the series not being stationary does not mean it has to have a trend. A random walk is an excellent example of a nonstationary time series with no underlying trend.

Tests for determining whether a process is stationary are called hypothesis tests. An Augmented Dickey-Fuller (ADF) test is the most used metric for stationarity problems to assess a time series. This test posits a null hypothesis that a unit root exists in a time series. Depending on the test results, this null hypothesis can be rejected for a specified significance level, meaning the presence of a unit root test can be left at a given significance level.

Note that tests for stationarity focus on whether the mean of a series is changing. The variance is handled by transformations rather than formally tested. Whether a series is stationary is thus a test of whether a series is integrated. An integrated series of order d is a series that must be differenced d times to become stationary.

The framing of the Dickey-Fuller test is as follows:

$\Delta y_t = y_t - y_{t-1} = (\varphi - 1) \times y_{t-1} + \varepsilon_t$

Then the test of whether $\varphi = 1$ is a simple *t*-test of whether the parameter on the lagged $y_{t-1}$ is equal to 0. The difference that the ADF test makes is to account for more lags so that the underlying model takes higher-order dynamics into account, which can be written as a series of differenced lags:

$Y_{t-\varphi 1} \times y_{t-1-\varphi 2} \times y_{t-2} \ldots = \varepsilon_t$

This requires somewhat more algebra to write as a series of differenced lags. The expected distribution against which to test the null hypothesis is slightly different from the original Dickey-Fuller test. The ADF test is the most widely presented test for stationarity in the time series literature.