

## Zaman Serisi Nedir?

Zaman serisi, belirli bir süre boyunca toplanan ve kronolojik olarak sıralanan veri noktalarının bir koleksiyonudur. Zaman serisi verileri, farklı veya düzenli zaman aralıklarında kaydedilen bir dizi veri noktası veya gözlemdir. Kaydedilen veri noktalarının sıklığı saatlik, günlük, haftalık, aylık, üç aylık veya yıllık gibi belirli periyotlarda olabilir. Örneğin, hava durumu, nüfus sayımı, günlük satışlar gibi verilerdir. Zaman serileri temelinde 4 bileşenden oluşur.

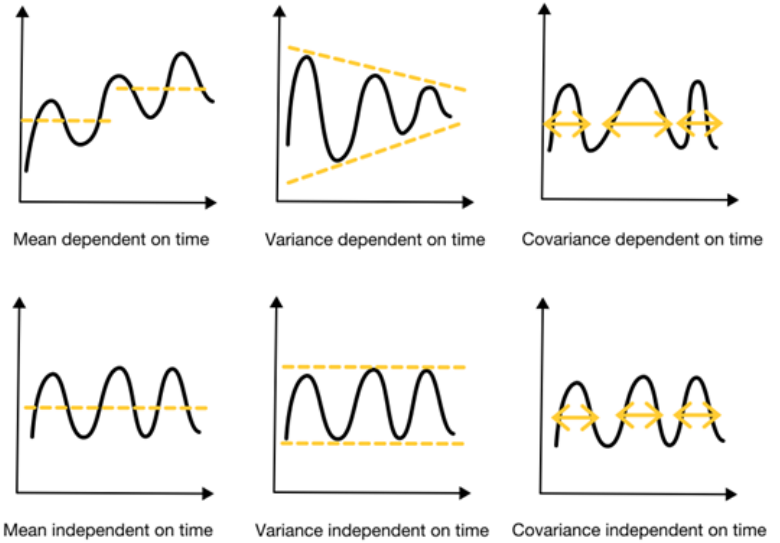
- 1- **Durağanlık (Stationary)**
- 2- **Trend**
- 3- **Mevsimsellik (Seasonality)**
- 4- **Döngü (Cycle)**

### 1-Durağanlık(Stationary)

Durağanlık bileşeni kısaca serinin istatistiksel özelliklerinin zaman içerisinde değişmemesidir. Zaman serilerinde durağanlığı **Dickey-Fuller** veya **Kwiatkowski-Phillips-Schmidt-Shin** test ile istatistiksel olarak test edebiliriz. Bir zaman serisinin ortalamasının veya varyansının zaman içinde değişip değişmediğini test etmek için kullanılır. Örneğin Dickey-Fuller bir zaman serisi için null hipotezi ( $H_0$ ) olarak serinin durağan olmadığını ve alternatif hipotezi olarak serinin durağan olduğunu kabul eder. Eğer test sonucunda null hipotezi reddedilirse, seri durağan olarak kabul edilir.

Üst resimlerde görüldüğü gibi serinin istatistiksel özellikleri zamanla değişmiş. Ama alttaki resimlere bakarsak istatistiksel özellikleri zamanla değişmemiş.

**Not :** Bir zaman serisinin varyansı ve kovaryansı zaman boyunca sabit kalıyorsa, serinin durağan olduğu kabul edilir. Ortalama varyans ya da kovaryansın biri zamana göre değişiyorsa veri kümesi durağan olmayan olarak adlandırılır.



### Otokorelasyon (ACF)

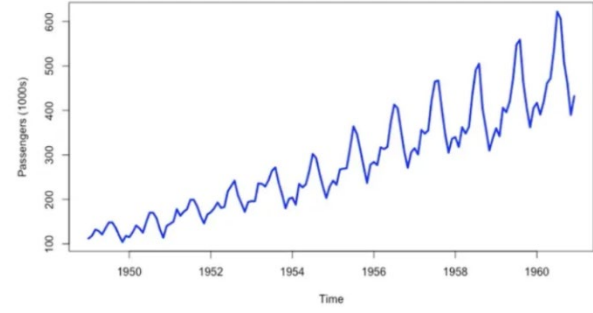
ACF grafiği, verilerin zaman grafiğine bakmanın yanı sıra, durağan olmayan zaman serilerini belirlemek için de kullanışlıdır.

**ACF** durağan olmayan bir zaman serisi için yavaşça düşer. Mesela çok yüksek bir değerden başlayıp sıfıra doğru çok yavaş düşmesi. Durağan bir zaman serisi içinse otokorelasyon fonksiyonu hızla sıfıra yaklaşır. Özetle,

Yukarıdaki durağanlık sınav yöntemlerinden sadece bir tanesinin kullanılmasından ziyade hem grafiksel hem de parametrik yöntemlerin birlikte kullanılması seri hakkında daha doğru çıkarımlar yapılmasını sağlayacaktır.

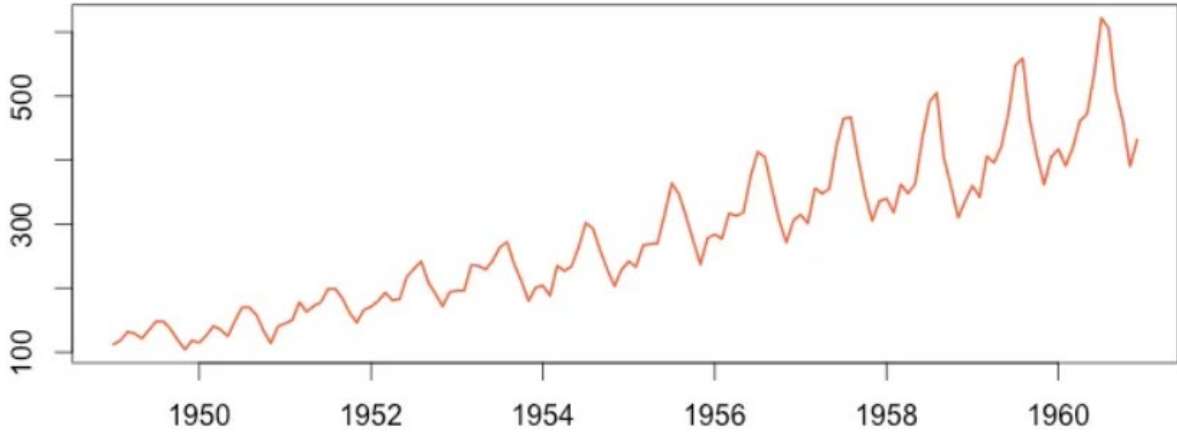
## 2-Trend

Bir zaman serisinde belli bir yöne doğru genel bir eğilimin varlığını ifade eder. Trend Negatif veya Pozitif veya Boş Trend olabilir. Örneğin, bir hisse senedinin uzun vadeli yükseliş eğilimi göstermesi gibidir. Genel olarak, bir trend gözlenen zaman serilerinde durağanlık olmadığı söylenebilir. Bunun temel sebeplerinden biri gözlem değerlerinin trend yönünde ortalamaların artması veya azalmasıdır. Durağanlık tanımı gereği bir serinin istatistikî değerlerinin zamana göre değişmemesi gerektiğinden genel olarak trend varlığı yapısı gereği durağanlıkla uyuşmayacaktır.



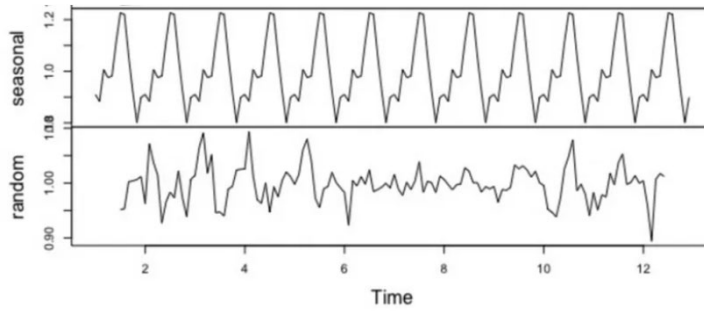
## 3-Mevsimsellik (Seasonality)

Zaman serisinde belli bir periyot için sürekli olarak tekrarlanan bir eğilimin varlığını ifade eder. Mevsimsellik yıllık, aylık, haftalık veya günlük periyotlar için görülebilir. Zaman serileri analizinde önemlidir çünkü gelecekteki değerlerin tahmin edilmesinde veya verilerin anlaşılmasında etkilidir.



## 4-Döngü (Cycle)

Döngüsellik de bir çeşit mevsimselliiktir. Fakat döngüsellik daha uzun vadeli, daha belirsiz, gün, hafta, mevsim gibi zamanlarla örtüşmeyecek şekilde gerçekleşir. Daha çok yapısal, nedensel, konjonktürel değişimlerle ortaya çıkar. İş dünyası ya da politik dünyadaki değişimlerden etkilenir.



Mevsimsellik ve Döngüsellik farkı

## Yumuşama — Düzeltme Yöntemleri (Smoothing)

Verilerdeki gürültüyü azaltmak ve trendleri veya desenleri daha iyi göstermek için kullanılan istatistiksel tekniklerdir.

### Single Exponential Smoothing (SES)

SES üssel düzeltme yaparak tahminde bulunur. Gelecek yakın geçmişle daha ilişkilidir varsayımıyla geçmişin etkilerini ağırlandırır. Geçmiş gerçek değerler ve geçmiş tahmin edilen değerlerin üssel olarak ağırlıklandırılmasıyla tahmin yapılır. Her bir gözlem için ağırlık, alpha düzgünleştirme parametresi kullanılarak hesaplanır ve 0 ile 1 arasındadır. Alpha değeri 1'e yakın olursa daha çok ağırlık verilir ,0'a yakın olursa daha eski gözlemlere verilir. Alpha yüksek olduğunda öğrenme becerisi yüksek, hatırlama becerisi daha düşük olduğundan gerçek değerlere daha yakın sonuç verir.

$$\hat{y}_t = \alpha \cdot y_{t-1} + (1 - \alpha) \cdot \hat{y}_{t-1}$$

### Double Exponential Smoothing — DES

DES, SES'ten daha gelişmiştir ve trend veya mevsimsellik içeren veriler için daha doğru sonuçlar verebilir. SES'te olduğu gibi geçmiş gözlemlerin ağırlıklı ortalamasını kullanır ancak ayrıca bir trend değeri de ekler. Trend değeri, verideki eğilimi ölçen bir değerdir ve tahminleri daha doğru hale getirir. Bu yöntemde iki farklı parametre kullanılır: alpha ve beta. Alpha verinin anlık değerini yansıtmak için kullanılır, beta ise trendi yansıtmak için kullanılır.

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$$

$$\hat{y}_{t+1} = \ell_t + b_t$$

### Triple Exponential Smoothing (TES veya Holt-Winters)

$$\ell_t = \alpha(y_t - s_{t-p}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(y_t - \ell_t) + (1 - \gamma)s_{t-p}$$

$$\hat{y}_{t+m} = \ell_t + mb_t + s_{t-p+1+(m-1)modp}$$

Hem trend hem de mevsimsellik göz önünde bulundurulur. DES'ten daha ileri bir versiyondur ve verideki uzun vadeli trendleri ve desenleri tahmin edebilir. Holt-Winters metodu, gelecekteki değerleri tahmin etmek için üç farklı bileşen kullanır: seviye bileşeni, trend bileşeni ve mevsimsel bileşen. Seviye bileşeni serinin ortalama değerini temsil eder, trend bileşeni serinin uzun vadeli trendini temsil eder ve mevsimsel bileşen serideki tekrar eden desenleri temsil eder. Üç düzgünleştirme parametresi kullanılır: alpha, beta ve gamma. Alpha düzgünleştirme parametresi seviye bileşeni için kullanılır, beta düzgünleştirme parametresi trend bileşeni için kullanılır ve gamma düzgünleştirme parametresi mevsimsel bileşen için kullanılır.

### İstatistiksel Yöntemler

### Oto Regresyon (Auto Regression — AR)

AR, bir zaman serisini kendi geçmiş değerlerinin doğrusal bir kombinasyonu olarak modelleyen bir istatistiksel yöntemdir. Bu, zaman serisi cevap değişkeni olan bir regresyon analizi türüdür ve tahmin edici değişken aynı zaman serisi belirli bir zaman adımı ile gecikmiştir.

AR modelinin ana amacı, geçmiş değerlere dayanarak zaman serisi gelecek değerlerini tahmin etmektir. Katsayıların değerlerini tahmin ederek. Bu yöntem zaman serisi verilerindeki otokorelasyonu yakalamak için kullanılır.

$$\text{AR}(1) \text{ model : } y_t = a_1 y_{t-1} + \epsilon_t$$

$$\text{AR}(2) \text{ model : } y_t = a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t$$

$$\text{AR}(p) \text{ model : } y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \epsilon_t$$

### Hareketli Ortalama (Moving Average — MA)

Bir zaman serisinin sıradaki değeri kendisinden k adet önceki değerlerin ortalamasıdır yaklaşımıyla çalışır. Hareketli ortalama pratikte genellikle bir tahmin için değil, verilerin dalgalanmasını azaltmak ve trend yakalamak için kullanılır. Fakat makine öğrenmesi kapsamında yeni özellikleri üretilirken hareketli ortalamaya yönelik özellikler de üretilir.

### Ağırlıklı Ortalama (Weighted Average)

Hareketli ortalamaya benzer bir mantıkta çalışır. Hareketli ortalamadan farklı olarak daha yakın olan gözlemlerin etkisini artırmak için ağırlıklandırır.

### ARMA (AR + MA)

ARMA zaman serisi verilerinin geçmiş değerleriyle oluşan otoregresif bir modele (AR) ve hareketli ortalama modeline (MA) dayanır. Geçmiş değerlerin katsayıları kullanılarak otoregresif bir model oluşturulur ve aynı zaman diliminde gözlenen gürültüleri tanımlamak için hareketli ortalama modeli kullanılır. Single Exponential Smoothing modelinin kardeşidir.

$$y_t = a_1 y_{t-1} + m_1 \epsilon_{t-1} + \epsilon_t$$

### ARIMA (Autoregressive Integrated Moving Average)

ARIMA (Autoregresif Entegre Hareketli Ortalama), bir zaman serisini kendi geçmiş değerleriyle ve geçmiş hata değerleriyle oluşan bir kombinasyon olarak modelleyen bir istatistiksel yöntemdir.

Bu yöntem, AR (Auto-Regression) ve MA (Moving Average) yöntemlerinin birleşimidir. Öncelikle verideki durağanlık eksikliği öncelikle düzeltilir ve daha sonra zaman serisi verileri AR ve MA modelleri ile modellenir. 3 parametre ile tanımlanır: p, d, q. p, verinin geçmiş değerlerinden etkilenme derecesini temsil eder, d ise verinin durağan hale getirilme derecesi ve q ise verinin hareketli ortalama derecesini temsil eder.

### SARIMA (Seasonal Autoregressive Integrated Moving Average)

SARIMA (Sezonal Autoregresif Entegre Hareketli Ortalama), zaman serisi verilerinde mevsimsellik etkilerini dikkate alan bir ARIMA yöntemidir. Bu yöntem, mevsimsel bir periyodikliği olan veriler için kullanılır.

SARIMA modeli, standart ARIMA modeline ek olarak mevsimsel komponentleri de içererek daha doğru tahminler yapabilir. Model, sezonun mevsimsel etkilerini tanımlayan 4 parametre ile tanımlanır:

P, D, Q, m.

P, D ve Q, sırasıyla mevsimsel otoregresif, mevsimsel entegre ve mevsimsel hareketli ortalama derecesini temsil ederken, m ise mevsimsel periyodu temsil eder.

## PYTHON KÜTÜPHANLERİ

### 1. statsmodels.tsa.stattools.adfuller

**Açıklama:**

- **Augmented Dickey-Fuller (ADF) Testi:** Bir zaman serisinin durağan olup olmadığını test etmek için kullanılır. Durağan olmayan bir seriyi durağan hale getirmek için fark alma veya dönüşümler uygulanabilir.

### 2. statsmodels.tsa.arima\_model.ARIMA

**Açıklama:**

- **ARIMA Modeli:** Otoregresif Entegre Hareketli Ortalama modelidir. Zaman serisi verilerinde trend ve mevsimsellik gibi bileşenleri modellemek için kullanılır.
- **Parametreler:** p (otoregresif kısmın derecesi), d (durağan hale getirmek için fark sayısı), q (hareketli ortalama kısmın derecesi).

### 3. statsmodels.tsa.arima\_model.ARIMAResults

**Açıklama:**

- **ARIMA Sonuçları:** ARIMA modelinin sonuçlarını ve tahminlerini içerir. Modelin uygunluğu ve performansını değerlendirmek için kullanılır.

### 4. statsmodels.tsa.seasonal.seasonal\_decompose

**Açıklama:**

- **Mevsimsel Ayrıştırma:** Bir zaman serisini trend, mevsimsellik ve rastgele bileşenlere ayırmak için kullanılır. Bu ayrıştırma, serinin farklı bileşenlerini analiz etmeyi kolaylaştırır.

### 5. statsmodels.graphics.tsaplots.plot\_acf

**Açıklama:**

- **Otokorelasyon Fonksiyonu (ACF) Grafiği:** Bir zaman serisinin otokorelasyonlarını gösterir. ACF, serinin kendi gecikmiş değerleri ile olan ilişkisini analiz etmeye yarar.

### 6. statsmodels.graphics.tsaplots.plot\_pacf

**Açıklama:**

- **Kısmi Otokorelasyon Fonksiyonu (PACF) Grafiđi:** Bir zaman serisinin gecikmeli deđerlerinin kısmi otokorelasyonlarını gösterir. PACF, serinin her bir gecikmesi arasındaki ilişkiyi doğrudan analiz eder.