

## Başlık: Veri Analizi ve Korelasyon: Python ile İki Değişken Arasındaki İlişkiyi Keşfetme

Günümüzde, veri analizi ve istatistiksel yöntemler, veri bilimcilerin, araştırmacıların ve analistlerin veri setlerindeki desenleri anlamalarına yardımcı olan önemli araçlardır. Bu yazıda, Python programlama dilini kullanarak korelasyon analizi yapma sürecini ele alacağız.

Korelasyon, iki değişken arasındaki ilişkiyi, bu ilişkinin yönünü ve şiddetini ölçen istatistiksel bir yöntemdir. İki ya da daha çok değişken arasındaki ilişkinin matematiksel bağıntısı “Regresyon Analizi” ile ilişkinin yönü ve derecesi ise “Korelasyon Analizi” ile incelenir.

Korelasyon katsayısı (r), bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin gücünü gösteren bir katsayıdır. Örneğin; öğrencinin ders çalışma süresi ile aldığı istatistik notu arasında ilişki olup olmadığı korelasyon katsayısı ile incelenebilir. Korelasyon katsayısı değişkenin yönü ve etkileşimlerin nasıl olduğu hakkında bilgi verir.

Korelasyon katsayısı, iki değişken arasındaki doğrusal ilişkinin ölçüsü olup incelenen değişkenlerin birimlerinden bağımsızdır ve  $-1 \leq r \leq 1$  arasındadır. Korelasyon katsayısının 0’a yaklaşması değişkenler arasında zayıf ilişkinin varlığını gösterir. Değişkenler birlikte artıyor veya azalıyorsa pozitif yönde, değişkenlerden biri artarken diğeri azalıyorsa negatif yönde bir ilişki vardır.

Korelasyon Analizi ilişkisini Python gibi programlama dilleri ile hesaplamak oldukça yaygındır.

### Python ile Korelasyon Analizi

Öncelikle, rastgele veri seti oluşturarak işe başlayalım ve bu veri setlerini ‘matplotlib’ kütüphanesi ile grafiğe dökelim. İlk veri setimizde, x ve y değerleri arasında nasıl bir korelasyon olduğunu gözlemleyelim.

```
import random
import matplotlib.pyplot as plt

def calculate_r(x, y):
    # Korelasyon katsayısını hesaplamak için kullanılan fonksiyon
    mean_x = sum(x) / len(x)
    a = [(x_i - mean_x) ** 2 for x_i in x]
    b = sum(a) / len(a)
    sd_x = b ** 0.5

    mean_y = sum(y) / len(y)
    a = [(y_i - mean_y) ** 2 for y_i in y]
    b = sum(a) / len(a)
    sd_y = b ** 0.5

    t = 0
    for i in range(len(x)):
        x_z_or_standard_value = (x[i] - mean_x) / sd_x
        y_z_or_standard_value = (y[i] - mean_y) / sd_y
        t = t + x_z_or_standard_value * y_z_or_standard_value

    return t / len(x)

# İlk örnek veri seti
```

```
x = [random.randint(20, 60) for i in range(20)]  
y = [random.randint(50, 100) for i in range(20)]
```

```
# Scatter plot  
plt.scatter(x, y, marker='x')  
correlation_coefficient = calculate_r(x, y)  
print(f'Korelasyon Katsayısı: {correlation_coefficient}')
```

Bu adımda, oluşturduğumuz veri seti üzerindeki korelasyon katsayısını hesaplıyoruz ve grafiğini çiziyoruz. Daha sonra ise bu ilişkinin yorumunu yaparak açıklamalar ekleyebiliriz.

## Korelasyon Sonuçlarını Anlama

Kodun çıktısı olan korelasyon katsayısı, iki değişkenin arasındaki ilişkinin yönünü ve gücünü gösterir. Pozitif bir korelasyon katsayısı, bir değişkenin arttığı durumda diğer değişkenin de arttığını belirtir. Negatif bir korelasyon ise bir değişken artarken diğerinin azaldığını ifade eder. Eğer korelasyon katsayısı sıfıra yakınsa, iki değişken arasında belirgin bir ilişki yoktur.

Bu bilgileri kullanarak, örnek veri setlerimizdeki grafiği incelediğimizde, iki değişken arasında nasıl bir ilişki olduğunu anlamamız daha kolay olacaktır. Örneğin, pozitif bir korelasyon, veri noktalarının yukarı doğru bir eğilim gösterdiğini ve negatif bir korelasyon ise aşağı doğru bir eğilim gösterdiğini gözlemleyebiliriz.

## Diğer Veri Setleri İle Analiz

Aynı adımları, farklı özelliklere sahip veri setleri için de uygulayabiliriz. Örneğin, değişkenler arasında doğrusal bir ilişki olduğu durumları ve farklı dağılımları inceleyebiliriz. Bu, korelasyon analizi ile veri setlerimizdeki desenleri keşfetmemize ve bilgi çıkarmamıza yardımcı olur.

## Sonuç

Python programlama dilini kullanarak korelasyon analizi yapmak, veri analizindeki temel becerilerden biridir. Bu yazıda, basit bir örnek veri seti üzerinde adım adım korelasyon analizi yapma sürecini inceledik. Veri analizi, gerçek dünya veri setlerini daha iyi anlamamıza ve bilinçli kararlar almamıza yardımcı olan önemli bir araçtır. Korelasyon analizi de bu sürecin önemli bir parçasıdır ve iki değişken arasındaki ilişkiyi anlamamıza yardımcı olan değerli bir istatistiksel yöntemdir.