

Tanımlayıcı İstatistikler

Tanımlayıcı İstatistikler bir değerler dizisinin istatistiksel olarak genel özelliklerini tanımlayan ölçülerdir

Tanımlayıcı İstatistikler

```
graph TD; A[Tanımlayıcı İstatistikler] --> B[Yer Gösteren Ölçüler]; A --> C[Yaygınlık Ölçüleri];
```

Yer Gösteren Ölçüler

Yaygınlık Ölçüleri

Yer Gösteren Ölçüler

- Bir dağılımı tanımlayabilmek için çeşitli yer gösteren ölçüler vardır.
- Bu ölçülere merkez ölçüleri ya da ortalama ölçüleri de denir.
- Bunlar yardımıyla dağılımdaki tüm değerleri temsil eden tek bir değer elde edilir.

Merkezi Eğilim (Ortalama) Ölçüleri

Aritmetik
Ortalama

Ortanca

Tepe
Değeri

Oran

Geometrik
Ortalama

Harmonik
Ortalama

Konum Ölçüleri

Çeyrekler

Yüzdelikler

Aritmetik Ortalama

Çoğunlukla sayısal verilerde kullanılan bir **merkezi eğilim** ölçüsüdür. Her bir gözleme ilişkin değerlerin toplamının denek sayısına bölünmesi ile elde edilir.

N : Kitledeki **n** : Örneklemden denek sayısını göstermek üzere

Kitle
A.Ortalaması

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

Örneklem
A. Ortalaması

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

Aritmetik ortalama dağılımdaki tüm değerleri dikkate alır. Ancak dağılımdaki aşırı değerlerden etkilenir.

Ortanca

Sıraya dizilmiş veri dizisinin ortasındaki değerdir.

Ortancayı bulmak için: Veriler küçükten büyüğe sıraya dizilir.

Denek sayısı tek ise
en ortadaki değer,
 $\text{Ortanca} = (n+1)/2$ 'inci değerdir.

Denek sayısı çift ise
 $(n/2)$ ve $(n+2)/2$ 'nci denek değerlerinin
ortalaması dağılımın ortancasını verir.

**Ortanca dağılımın orta noktası hakkında bilgi verir.
ve aşırı değerlerden etkilenmez. Bu nedenle dağılımda aşırı
gözlemlerin bulunduğu durumlarda, ortalama ölçüsü
olarak ortancanın kullanılması daha doğrudur.**

Tepe Değeri

Tepe değeri dağılımda en fazla tekrar edilen değerdir. Tepe değerini hesaplamak için kullanılan bir formül yoktur.

Oran

Nitelik veriler aritmetik ortalama, ortanca, tepe değeri gibi ortalama ölçüleri ile özetlenmez. Nitelik veriler çoğunlukla yüzde (oran) ile özetlenirler.

Oran (yüzde) Kullanmanın Önemi

- Yüzde kullanma verinin daha kolay anlaşılmasını sağlar.
- İki yada daha fazla sayıda grubun özellikleri karşılaştırılırken ham sayılar tek başına bir anlam ifade etmez.
- Gruplar özelliklerine göre yüzdelerle ifade edilmelidirler.

Geometrik Ortalama

- Veri geometrik artış gösteriyorsa
- Veriye logaritmik dönüşüm uygulanmışsa kullanılır.

$$GO = \sqrt[n]{x_1 \times x_2 \times x_3 \dots \times x_n} \quad \text{ya da}$$

$$GO = 10^{\frac{\sum_{i=1}^n \log_{10} x_i}{n}}$$

Harmonik Ortalama

- Veri setindeki değerler bir zaman serisi ise (birim zamanda farklı değerler)
- Hız, fiyat ortalamalarında kullanılır

$$HO = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Konum Ölçüleri

Çeyrekler: dağılımı 4 eşit parçaya bölen değerlerdir. Bunlar,

1. Çeyrek (Ç1)

Değerlerin %25'i Ç1'e eşit ya da ondan küçüktür.

2. Çeyrek (Ç2)

Değerlerin %50'si Ç2'ye eşit ya da ondan küçüktür. Bu değer aynı zamanda *ortanca*dır.

3. Çeyrek (Ç3)

Değerlerin %75'i Ç3'e eşit ya da ondan küçüktür.

Yüzdelikler

Yüzdelikler sıraya dizilmiş verilerde yığılımlı sıklıkları gösterirler.

Örneğin verilerin ilk %30'u 30. Yüzdeliğe (**Y30**) eşit ya da ondan küçüktür.

Yaygınlık Ölçüleri

- Bir dağılımdaki değerlerin farklılıklarını gösterir.
- Bu farklılıkların derecesi dağılımın yaygınlığı kavramını oluşturur. İki dağılım aynı ortalama, ortanca ya da tepe değerine sahipken yaygınlıkları farklı olabilir.

Dağılımların yaygınlığı hakkında bilgi veren ve en çok kullanılan ölçüler

- Dağılım (değişim) Aralığı
- Standart Sapma
- Varyans
- Çeyreklikler Arası Genişlik
- Çeyrek Sapma

Dağılım Aralığı

Dağılım aralığı en basit yaygınlık ölçüsüdür. Dağılımdaki en büyük değerden en küçük değerın çıkartılması ile bulunur. R ile gösterilir.

$$R = \text{En Büyük Değer} - \text{En Küçük Değer}$$

- Dağılım aralığı dağılımdaki diğer değerlerden oldukça farklı değerler alan aşırı değer(ler)den etkilenir.
- Dağılımda yalnızca 2 gözleme ilişkin değer dikkate alındığı için kaba bir yaygınlık ölçüsüdür.
- Gözlemlerin çoğunun en büyük yada en küçük değere yakın olduğu durumlarda da gerçek değişkenlik hakkında bilgi vermez.

Standart Sapma

- Bir dağılımın yaygınlığını gösteren en önemli yaygınlık ölçülerinden biridir.
- Dağılımdaki tüm değerlerin aritmetik ortalamaya olan uzaklıklarının ortalamasıdır.
- Standart sapma büyüdükçe dağılımın yaygınlığı artar.
- Dağılımdaki değerler aynı ise yaygınlık yoktur ve standart sapma sıfırdır.
- Standart sapma hesaplanırken dağılımdaki tüm değerler dikkate alınır.
- Standart sapma, aritmetik ortalama kullanıldığında bir yaygınlık ölçüsü olarak kullanılır.
- Çarpık dağılımlarda kullanılması önerilmez!

Standart Sapma

N : Kitledeki n : Örneklemdeki denek sayısını göstermek üzere

Kitle
S. Sapması

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{N}}$$

Örneklem
S. Sapması

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$$S = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1}}$$



Varyans

Standart sapmanın karesine varyans denir (σ^2). Varyansın birimi karesel olduğu için yaygınlık ölçüsü olarak veriyi tanımlamakta pek kullanılmaz.

Çeyreklikler Arası Genişlik

Dağılımdaki verilerin ortadaki 0.50 'sinin yer aldığı aralığı belirlemek için kullanılır.

$$\text{ÇAG} = \text{Ç3} - \text{Ç1}$$

- Çeyreklikler arası genişlik aşırı uç değerlerden etkilenmez.
- Çünkü çeyreklikler arası genişlik dağılımdaki değerlerin merkezdeki %50'si ile ilgilenir.
- Özellikle uçtaki değerlerden çok ortadaki değerlerle ilgilenildiği durumlarda kullanılır.
- Eğer incelenen dağılım simetrikse 25. ve 75. Yüzdelikler ortancadan eşit uzaklıktadır.

Çeyrek Sapma

Bu değer yüzdeliklerle ortanca arasındaki uzaklığın ortalama bir ölçüsüdür.

$$\zeta S = \frac{\zeta 3 - \zeta 1}{2}$$

- Çeyrek sapma, ortalama ölçüsü olarak ortancanın kullanıldığı durumlarda kullanılan yaygınlık ölçülerinden biridir.
- Özellikle aşırı değerlerin dağılımının sadece bir tarafında olduğu durumlarda kullanılması gerekir.

Değişim Katsayısı

- Standart sapma bir dağılımın yaygınlığını gösteren ölçülerden birisidir.
- Aritmetik ortalama büyüdükçe standart sapmanın büyüme eğilimi vardır.
- Standart sapmanın büyüklüğüne bakarak bir dağılımın yaygınlığı konusunda yargıya varmak her zaman doğru değildir.

İki ya da daha fazla dağılımın yaygınlığını karşılaştırmak istediğimizde standart sapmayı doğrudan kullanamayız.

Dağılımın yaygın olup olmadığına karar verebilmek için değişim katsayısını hesaplamalıyız. Değişim katsayısı dağılımdaki değerlerin ortalamaya göre yüzde kaçlık bir değişim gösterdiğini belirtir.

$$DK = \frac{s}{\bar{x}} \times 100$$