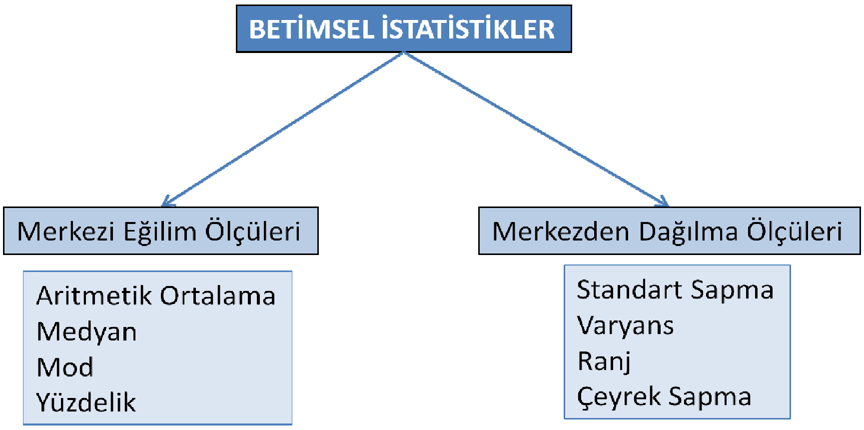
BÖLÜM 6

MERKEZDEN DAĞILMA ÖLÇÜLERİ

Gözlenen belli bir özelliği, bu özelliğe ilişkin ölçme sonuçlarını yani verileri kullanarak betimleme, istatistiksel işlemlerin bir boyutunu oluşturmaktadır. Temel sayma ve sınıflama işlemleri ile elde edilen frekans tabloları ve grafik gösterimleri, betimlemenin yol ve yöntemlerinden biridir. Bunun yanı sıra 'betimsel istatistikler (descriptive statistics)' olarak adlandırılan bazı sayısal değerler kullanılarak da betimleme yapılabilmektedir.

Betimsel istatistikler, (1) merkezi eğilim ölçüleri ve (2) merkezden dağılma ölçüleri olmak üzere iki grupta sınıflandırılmaktadır:



Şekil 1. Betimsel İstatistiklerin Sınıflandırılması

Merkezi eğilim ölçüleri, 'merkeze yığılma ölçüleri' olarak da ifade edilebilmektedir. Aritmetik ortalama[1], mod, medyan, yüzdelik gibi nokta değerler, merkezi eğilim ölçüleridir. Diğer bir deyişle bu ölçüler, tek bir nokta belirtir. Bu nokta değerler verilerin yığılma noktaları olarak ilgilenilen özelliğe dönük betimlemelerin yapılmasında dikkate alınabilir.

Merkezi dağılım ölçüleri, 'merkezden yayılma ölçüleri' olarak da ifade edilebilmektedir. Standart sapma, varyans, ranj, çeyrek sapma gibi değerler, merkezden dağılma ölçüleridir. Bu ölçüler, merkez ya da ölçüt olarak belirlenen noktalara göre verilerin yayılması, çeşitlenmesi ya da farklılaşması hakkında bilgi verir.

---

[1]Aritmetik ortalama dışında, geometrik ortalama, harmonik ortalama gibi başkaca ortalama değerler de bulunmaktadır. Fakat bunlar içerisinde en bilindik olan ve en sık kullanılan aritmetik ortalamadır. Bu nedenle bundan sonraki bölümlerde aritmetik ortalama yerine 'ortalama' kullanımı tercih edilecektir. 'Ortalama' kullanımı, bundan sonraki bölümlerde, aritmetik ortalamayı ifade etmektedir.

Veri setinin karakteristiğine ve verilerin dağılımına göre uygun betimsel istatistiklerin kestirilmesi ve dikkate alınması gerekir. Her betimsel istatistik her veri setinde anlamlı olmayabilir. Bu nedenle her bir betimsel istatistiğin hesaplanmasının yanı sıra hangi durumlarda kullanılabilir olduğunun da bilinmesi önemlidir. Aksi durumda elde edilen sayılar, yanıltıcı olabilir, yanlış ya da eksik yorumların yapılmasına yol açabilir.

Betimsel istatistikler tek başına, ilgilenilen özellik hakkında fazlaca bilgi sağlamaz. Birden fazla betimsel istatistik bir arada değerlendirilerek ya da birden fazla gruba/örnekleme yönelik betimsel istatistikler bir arada değerlendirilerek anlamlı betimsel yorumlar yapmak mümkündür.

Bu bölümde betimsel istatistiklerden merkezden dağılma ölçüleri grubunda yer alan 4 istatistik hakkında bilgi verilmektedir.

6.1. STANDART SAPMA

'Standart kayma' olarak da ifade edilen standart sapma (standard deviation); bir veri setinde her bir verinin ortalamadan uzaklıklarının standartlaştırılmış bir ölçüsüdür.

Evren standart sapması ve örneklem standart sapması, farklı sembollerle ifade edilmenin yanı sıra farklı formüllerle hesaplanır.

Evren Standart Sapması

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

Örneklem Standart sapması

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

Formüllerde yer alan;

- X_i, her bir gözlem birimine yönelik ölçme sonucunu
- μ, evren ortalamasını,
- $\bar{X}$, örneklem ortalamasını,
- n, örneklem büyüklüğünü yani örneklemde yer alan nesne yada kişi sayısını,
- N, evren büyüklüğünü yani evrende yer alan gözlem birimi sayısını

  ifade etmektedir.

Dikkat edilirse örneklem standart sapmasının belirlenmesinde (n-1)'e bölme işlemi yapılmaktadır. Bu değer 'serbestlik derecesi' olarak isimlendirilir. Daha az yanlı yani daha titiz bir kestirim sağlamaktadır.

Standart sapma, ortalamaya bağlı olarak hesaplanan bir istatistiktir. Ortalamada olduğu gibi standart sapma da sürekli değişkenlerde ve özellikle normal dağılım gösteren değişkenlerde kullanışlıdır. Değişkenin sürekli yani en az eşit aralık ölçeğinde olduğu ve değişkenin gözlenen değerlerinin normal dağılım gösterdiği durumlarda standart sapma en 'sağlam' merkezden dağılma ölçüsü olarak bilinir.

Cinsiyet, eğitim düzeyi, doğum yeri, sınıf, şube, okul türü gibi kategorik değişkenlerde yani sınıflama ve sıralama ölçeği düzeyindeki değişkenlerde, ortalama anlamlı olmadığı gibi standart sapma da anlamlı değildir.

Verilerin normal dağılımdan sapması ya da uç noktaların etkisinin söz konusu olduğu durumlarda, ortalamada olduğu gibi standart sapma da doğrudan yorumlanabilir değildir. Bu durumlarda standart sapma, yanıltıcı olabilir, eksik ve yanlış betimsel yorumlara yol açabilir.

ÖRNEK 1.

Bir basketbol takımında oynayan 10 çocuğun ağırlıkları ölçülmüştür:

Ağırlık: 45kg; 55kg; 40kg; 62kg; 55kg; 54 kg; 58kg; 48kg; 50kg; 43kg

Buna göre bu çocukların ağırlıklarının standart sapmasını hesaplayalım. Standart sapma formülüne dikkat edilirse, standart sapma, ortalamaya bağlı olarak hesaplanan bir istatistiktir. O halde öncelikle ağırlıkların ortalamasının hesaplanması gerekir. Standart sapmanın hesaplanmasında işlem adımları aşağıda verilmiştir:

1. Ortalamanın hesaplanması
2. Her bir gözlem değerinden ortalamanın çıkarılması ve ortalamadan farkların bulunması.
3. Her bir ortalamadan farkın karesinin alınması.
4. Ortalamadan farkların karelerinin toplanması.
5. Elde edilen toplamın N ya da (n-1)'e bölünmesi.
6. Elde edilen sayının karekökünün alınması.

Verilen örnekte öncelikle ortalamayı hesaplayalım:

$$\overline{X} = \frac{45 + 55 + 40 + 62 + 55 + 54 + 58 + 48 + 50 + 43}{10}$$

$$= \frac{510}{10}$$

$$= 51$$

Ortalama hesaplandıktan sonra standart sapmayı daha kolay hesaplayabilmek için üç sütunlu bir tablo kullanılabilir:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
45	45-51 = -6	$(-6)^2 = 36$
55	-4	16
40	-11	121
62	11	121
55	4	16
54	3	9
58	7	49
48	-3	9
50	-1	1
43	-8	64
	Toplam	**442**

Ortalamadan farkların kareleri toplamı 442 olarak hesaplandı. Şimdi formülde bu değeri yerine yazarak standart sapmayı belirleyelim:

$$s = \sqrt{\frac{442}{10 - 1}}$$

$$= \sqrt{49,11}$$

$$\approx 7,01$$

Çocukların ağırlıklarının standart sapması 7,01 olarak hesaplanmıştır. Yukarıda açıklandığı gibi bu değer tek başına fazlaca bir bilgi vermez. Sadece ağırlıkların orta noktaya göre sapma gösterdiği yani çocukların farklı ağırlıklarda olduğu söylenebilir.

6.2. VARYANS

Varyans, standart sapmanın karesidir. Standart sapma gibi varyans da verilerin ortalamadan uzaklaşma düzeyleri yani verilerin çeşitliliği hakkında bilgi verir.

Evren varyansı ve örneklem varyansı, farklı sembollerle gösterilmenin yanı sıra hesaplama formüller açısından da farklıdır:

<u>Evren Varyansı</u>

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

<u>Örneklem Varyansı</u>

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Standart sapma için yukarıda yapılan açıklama ve uyarılar aynen varyans için de geçerlidir. Yani varyans da sürekli ve normal dağılım gösteren değişkenlerde 'sağlam' bir kestirimdir. Bu varsayımları karşılamayan değişkenlerde varyans, ya anlamsızdır ya da yanıltıcı olabilir.

Varyansın hesaplanmasında takip edilmesi gereken işlem adımları, standart sapma için verilenlerle aynıdır. Sadece varyansı hesaplarken, standart sapma hesaplamasının son adımı olan karekök alma işlemi yapılmaz. Buna göre varyans kestiriminde takip edilecek işlem adımları şu şekildedir:

1. Ortalamanın hesaplanması
2. Her bir gözlem değerinden ortalamanın çıkarılması ve ortalamadan farkların bulunması.
3. Her bir ortalamadan farkın karesinin alınması.
4. Ortalamadan farkların karelerinin toplanması.
5. Elde edilen toplamın N ya da (n-1)'e bölünmesi.

Buna göre Örnek 1'de verilen değerlerin varyansının 49,11 olduğu açıktır.

ÖRNEK 2.

Bir sınıftaki 20 öğrencinin matematik dersine karşı tutumları bir tutum ölçeği ile ölçülmüştür. Ölçek toplam puanlarının varyansını hesaplanması aşağıdaki üç sütunlu tabloda gösterilmektedir.

n	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	40	13,40	179,56
2	35	8,40	70,56
3	35	8,40	70,56
4	30	3,40	11,56
5	26	-0,60	0,36
6	24	-2,60	6,76
7	23	-3,60	12,96
8	20	-6,60	43,56
9	15	-11,60	134,56
10	36	9,40	88,36
11	32	5,40	29,16
12	28	1,40	1,96
13	26	-0,60	0,36
14	25	-1,60	2,56
15	26	-0,60	0,36
16	17	-9,60	92,16
17	15	-11,60	134,56
18	33	6,40	40,96
19	26	-0,60	0,36
20	20	-6,60	43,56
	$\bar{x}$=26,60		$\sum$= **964,80**

Tabloda görüldüğü gibi toplam tutum puanlarının ortalaması 26,6 ve puanların ortalamadan farklarının kareleri toplamı 964,80 olarak hesaplanmıştır. Buna göre varyans bu değerin 20-1=19'a bölümünden elde edilen değer olacaktır.

$$s^2 = \frac{964,80}{20\text{-}1}$$

$$\approx \; 50,78$$

Öğrencilerin derse karşı tutum puanlarının varyansı 50,78 olarak hesaplanmıştır. Yukarıda açıklandığı gibi bu değer tek başına fazlaca bir bilgi vermez. Sadece tutum puanlarının orta noktaya göre sapma gösterdiği yani öğrencilerin farklı tutumlara sahip oldukları söylenebilir.

6.3. RANJ

Ranj (range); ölçme sonuçlarını gösteren bir veri setindeki alt ve üst değerlerin yani maksimum ve minimum değerlerin farkıdır. Yani bir değişkene yönelik gözlenen değerlerin aralığıdır.

Ranj = Maksimum Gözlenen Değer – Minimum Gözlenen Değer

Ranj, sürekli yani en az eşit aralıklı ölçek düzeyindeki değişkenlerde anlamlı bir istatistiktir. Kesikli yani sınıflama ve sıralama ölçekleri düzeyindeki değişkenlerde anlamlı değildir. Sıralama ölçeği düzeyindeki değişkenlerde, değişkenin gözlenen değerlerinin alt ve üst sınırlarının farkını ifade eden ranj yerine 'sıra farkları (rank)' kullanılabilmektedir.

Ranj, standart sapma ve varyans gibi verilerin çeşitliliği ve yayılması hakkında bilgi veren bir merkezden dağılma ölçüsüdür. Bir değişkenin gözlenen değerlerinin ranjının yanı sıra bu değerlerin alt ya da üst sınırı biliniyorsa diğer uç değer de hesaplanabilir. Örneğin, bir sınıfta öğrencilerin akademik başarı notlarının ranjı 40 ve bu sınıfta en yüksek akademik başarı notu 95 ise en düşük akademik başarı notunun 95 - 40 = 55 olduğu belirlenebilir.

ÖRNEK 3.

a) Örnek 1'de verilen, çocukların ağırlıklarının ranjını hesaplayalım:

Maksimum gözlenen ağırlık = 62

Minimum gözlenen ağırlık = 40

Ranj = 62 - 40 = 22

b) Örnek 2'de verilen öğrencilerin tutum puanlarının ranjını hesaplayalım:

Maksimum tutum puanı = 40

Minimum tutum puanı = 15

Ranj = 40 - 15 = 25

6.4. ÇEYREK SAPMA

Çeyrek sapma (quartile deviation); birinci çeyreklik ile üçüncü çeyreklik değerlerinin farkının yarısı alınarak hesaplanan bir merkezi dağılım ölçüsüdür.

Çeyreklik değerler, sıralanmış verilerin %25'lik bölümlerine karşılık gelen değerlerdir. Yani birinci çeyreklik (Q_1) 25. yüzdeliğe (Y_{25}), ikinci çeyreklik (Q_2) 50. yüzdeliğe (Y_{50}), üçüncü çeyreklik (Q_2) 75. yüzdeliğe (Y_{75}) ve dördüncü çeyreklik (Q_4) maksimum değer olarak 100. yüzdeliğe (Y_{100}) karşılık gelir.

Buna göre çeyrek sapma, 25. ve 75. yüzdeliklerin farkının yarısıdır:

$$\text{Çeyrek Sapma} = \frac{Y_{75} - Y_{25}}{2}$$

Ortalama, standart sapma ve varyansın, özellikle uç noktalardan aşırı etkilendiği yukarıda açıklanmıştı. Çeyrek sapma ilk çeyreklik olan %25'lik düşük ölçme sonuçlarını ve son çeyreklik olan %25'lik yüksek ölçme sonuçlarını dışarıda bırakan bir istatistiktir. BU nedenle çeyrek sapma, uç noktalardan etkilenmez. Bu nedenle uç noktaların etkisinin manidar olduğu veri setlerinde çeyrek sapma daha 'iyi' ve 'yansız' bir kestiricidir.

ÖRNEK 4.

a) Örnek 1'de verilen çocukların ağırlıklarını sıralayarak bu gözlenen değerlerin çeyrek sapmasını belirleyelim:

*Sıralı ağırlıklar: 40kg; 43kg; **45kg**; 48kg; 50kg; 54kg; 55kg; **55kg**; 58kg; 62kg*

$\boxed{Y_{25}}$ $\boxed{Y_{75}}$

Çeyrek sapma = (55 - 45) / 2

= 5

b) Örnek 2'de verilen öğrenci tutum puanlarını sıralayarak, bu puanların çeyrek sapmasını hesaplayalım:

Sıralı tutum puanları: 15; 15; 17; 20; **20**; 23; 24; 25; 26; 26; 26; 26; 28; 30; 32; **33**; 35; 35; 36; 40

$\boxed{Y_{25}}$ $\boxed{Y_{75}}$

Çeyrek sapma = (33 - 20) / 2

= 6,5