BÖLÜM 3

MERKEZİ EĞİLİM VE DAĞILIM ÖLÇÜLERİ

Merkezi eğilim ölçüleri kitleye ilişkin bir değişkenin bütün farklı değerlerinin çevresinde toplandığı merkezi bir değeri gösterirler. Dağılım ölçüleri ise değişkenin aldığı değerlerin birbirinden ne kadar farklı olduğunun ölçüsüdür. En sık kullanılan merkezi eğilim ölçüleri aritmetik ortalama, tepe değer(), ortanca, çeyreklikler ve geometrik ortalamadır. En sık kullanılan dağılım ölçüleri ise, değişim genişliği, çeyrek sapma, varyans, standart sapma, standart hata ve değişim katsayısıdır.

Sınıflandırılmış verilerde yalnızca aritmetik ortalama ve varyans hesabı verilecektir.

3.1. Merkezi Eğilim Ölçüleri

3.1.1. Aritmetik Ortalama:

Aritmetik ortalama, en çok kullanılan merkezi eğilim ölçüsüdür. Birimlerin belirli bir değişken bakımından aldıkları değerlerin toplamının birim sayısına bölümüdür. Eşit aralıklı ve oran ölçme düzeyinde ölçülen değişkenler için kullanılır.

μ: kitleye ilişkin aritmetik ortalama
$\bar{x}$: örnekleme ilişkin aritmetik ortalama

Simetrik dağılım gösteren verileri en iyi temsil eden bir merkezi eğilim ölçüsüdür.

Sınıflandırılmamış Verilerde Aritmetik Ortalama:

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

$\bar{x}$: aritmetik ortalama
x_i: örneklemdeki i. birimin değeri
n : örneklemdeki birim sayısı

Örnek : Tablo 1.1. deki veriler için aritmetik ortalamayı hesaplayınız

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{50} x_i}{50} = \frac{962,2}{50} = 19,244$$

Sınıflandırılmış Verilerde Aritmetik Ortalama:

Frekans tablosu düzenlenmiş verilerde aritmetik ortalama

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i . S_i}{n},$$

$\bar{x}$: aritmetik ortalama

S_i: i. sınıfın sınıf orta değeri

f_i: i. sınıfın frekans değeri

k : sınıf sayısı

n : örneklemdeki birim sayısı

şeklinde hesaplanabilir.

Örnek : Tablo 1.1. deki veriler için aritmetik ortalamayı sınıflandırılmış veriler için hesaplayınız

$$\overline{x} = \frac{\sum_{i=1}^{k} f_i.S_i}{n} = \frac{965,2}{50} = 19.304$$

Aritmetik ortalamanın özellikleri,

- Bir veri seti için sadece bir aritmetik ortalama vardır.
- Nicel verilere uygulanabilir.
- Birim değerlerinde meydana gelen değişim çok küçük olsa bile aritmetik ortalamayı etkiler.
- Aritmetik ortalama ile birim değerleri arasındaki farkların toplamı sıfırdır.

$$\sum_{i=1}^{n}(x_i - \mu) = 0$$

- Aritmetik ortalama ile birim değerleri arasındaki farkların kareleri toplamı minimum bir değerdir.

3.1.2.Mod (Tepe Değer) : Bir veri setinde en çok tekrar eden değere denir.
Örnek : 7, 6, 8, 12, 7, 3, 2, 15$\Rightarrow$7

Örnek: Bir grup öğrencinin herhangi bir dersten aldıkları notlara ilişkin veriler aşağıdaki gibidir.

94,65, 50, 50, 58, 72, 72, 72, 83, 85

Öğrencilerin notlarına ilişkin tepe değeri hesaplayınız.

Yukarıdaki veri gurubunda en çok tekrarlanan değer 72 olduğunda $TD = 72$ dir.

Tepe değerin özellikleri,

- Denek sayısı az olduğunda tepe değer güvenilir bir ölçü değildir.
- Bazı örneklemlerde bir tepe değer yerine iki ya da daha çok tepe değer olabilir. Bu durumda ya tepe değerini hesaplamaktan vazgeçilir ya da frekans tablosu tek tepe değerli bir dağılım olacak şekilde yeniden düzenlenir.
- Tepe değer hesaplanırken birimlerin tümü işleme katılmadığı için uç değerlerden etkilenmez.
- Nicel ve nitel verilerin her iki türü için de uygundur.
- Eğrisi J, ters J ve U şeklinde olan veriler için tepe değer kullanılmaz.
-

3.1.3. Medyan (Ortanca) : Bir veri gurubundaki değerlerin küçükten büyüğe sıralandığında tam ortaya düşen değer ortanca denir. Örneklem hacmi tek ise ortanca tam ortaya düşen verinin değeridir. Örneklem hacmi çift ise tam ortaya düşen iki değerin aritmetik ortalamasıdır.

Ortanca, sınıflama ölçme düzeyi ile ölçülen değişkenler için kullanılmaz. Eşit aralıklı, oran ve sıralama ölçme düzeyinde ölçülen değişkenler için kullanılır.

Bir dağılımı iki eşit parçaya bölen konum ölçüsüdür. Medyan değerinin altında o veri setinin %50'si, medyan değerinin üzerinde yine veri setinin %50'si bulunur.

$$\text{Medyan} \quad Q_2 = \begin{cases} \dfrac{x_j + x_{j+1}}{2}, \; j = \dfrac{n}{2}, \; n \text{ çift ise} \\ x_j \;, \; j = \dfrac{n+1}{2}, \; n \text{ tek ise} \end{cases}$$

Ortancanın özellikleri:

- Aşırı uç değerlerden etkilenmez.
- Birim değerleri ile ortanca arasındaki farkın yarısı negatif yarısı pozitiftir.
- $\sum|x_i - ortanca| = minimum$ dur.

Örnek : 16, 12, 38, 42, 28, 50 , 17, 64,82, 25 verilerin medyanını bulunuz.

İlk olarak verilerin sıralanması gerekir,

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
12	16	17	25	28	38	42	50	64	82

$n = 10$ çift sayı olduğundan

Medyan= $\dfrac{x_5 + x_6}{2} = \dfrac{28+38}{2} = 33$

Eğer $n = 9$ tek sayı ise

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
12	16	17	25	28	38	42	50	64

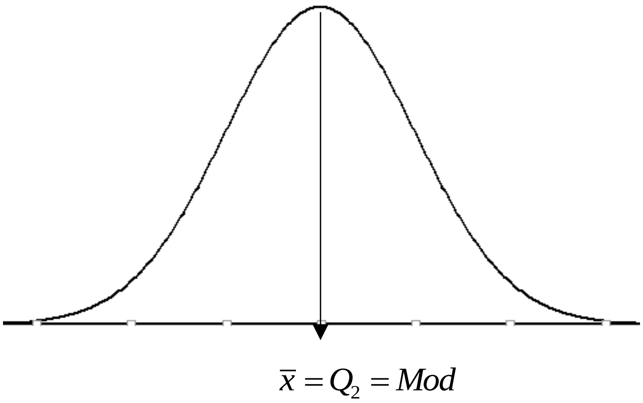
$$j = \frac{n+1}{2} = \frac{9+1}{2} = 5$$

Medyan= $Q_2 = x_5 = 28$

3.1.4. Aritmetik Ortalama, Tepe Değer ve Ortanca Arasındaki İlişki

Aritmetik ortalama, ortanca ve tepe değer arasındaki ilişki verilerin dağılımının çarpıklığı hakkında bilgi verir.

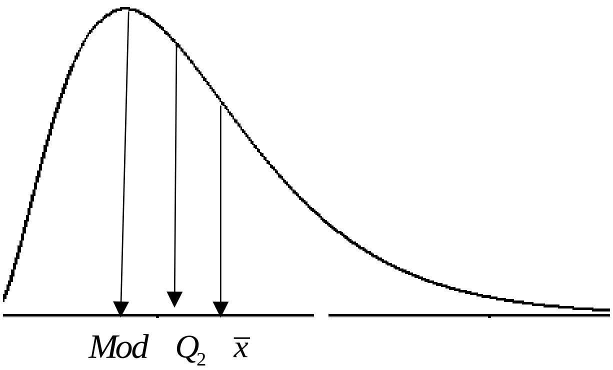
i) Bir sıklık dağılımı göze alındığında, dağılım simetrik olduğunda;



$\bar{X} = Q_2 = TD$

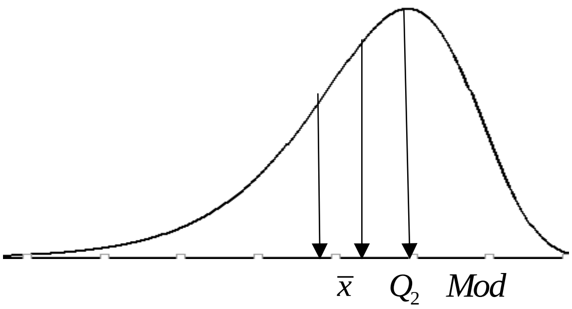
dağılım simetrik

ii) Sağa çarpık ya da pozitif yöne eğimli ise



$TD < Q_2 < \bar{X}$

dağılım sağa çarpık ya da pozitif yöne eğimli



$\bar{X} < Q_2 < TD$

dağılım sola çarpık ya da negatif yöne eğimli

2.1.5.Çeyreklikler

Küçükten büyüğe doğru sıralanmış verileri dört eşit parçaya bölen değerlere çeyrek değerler denir. Birinci çeyreklik (Q_1), veriler küçükten büyüğe sıralandığında verilerin %25 ini sağında, %75 ini solunda bırakan değerdir. İkinci çeyreklik ortancaya denk gelmektedir. Üçüncü çeyrek değer (Q_3), veriler küçükten büyüğe sıralandığında verilerin %75 ini sağında, %25 ini solunda bırakan değerdir. Yani sıralı verilerde, ortancadan küçük olan değerlerin ortancası birinci çeyrek değer, ortancadan büyük olan verilerin ortancası üçüncü çeyrek değerdir.

Örnek: 25, 50, 9, 68, 64, 74, 30, 13, 15 verilerin birinci ve üçüncü çeyrek değeri

hesaplayınız.

Küçükten büyüğe doğru sılalanmış veriler

$$\begin{array}{ccccccccc} x_1, & x_2, & x_3, & x_4, & \boxed{\begin{array}{c} x_5 \\ 30 \end{array}} & x_6, & x_7, & x_8, & x_9 \\ 9, & 13, & 15, & 25, & & 50, & 64, & 68, & 74 \end{array}$$

$$n=9\,,\quad j=\frac{n+1}{2}=\frac{9+1}{2}=5$$

$$Q_2=x_5=30\,,$$

Birinci çeyreklik

$$\begin{array}{cccc} x_1, & x_2, & x_3, & x_4 \\ 9, & 13, & 15, & 25, \end{array}$$

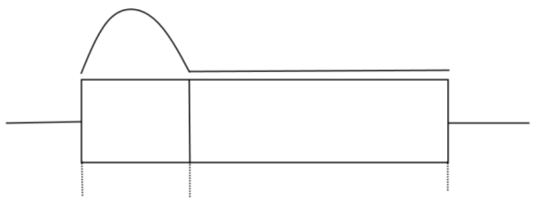
$$Q_1=\frac{x_2+x_3}{2}=\frac{13+15}{2}=14$$

Üçüncü çeyreklik

$$\begin{array}{cccc} x_6, & x_7, & x_8, & x_9 \\ 50, & 64, & 68, & 74 \end{array}$$

$$Q_3=\frac{x_6+x_7}{2}=\frac{64+68}{2}=66$$

***Kutu Diyagramı:** (Sağa çarpık)



$Q_1=14$ $Q_2=30$ $Q_3=66$

3.2. Değişim Ölçüleri:

Bir sıklık dağılımı hakkında bazı bilgiler edinmek ya da sıklık dağılımlarını karşılaştırmak için sadece konum ölçüleri yeterli değildir. Bu konum ölçüleri etrafındaki yayılma derecesini değişkenin alabileceği değerlerin birbirinden ne kadar farklı olabileceğini gösteren ölçülere de ihtiyaç vardır.

3.2.1.Değişim Genişliği:

Bir veri grubunda en büyük değer ile en küçük değer arasındaki farka “değişim genişliği” denir. “ R ” ile gösterilir.

$R = en\,büyük\,değer - en\,küçük\,değer$

Aşırı uç değerlerden etkilenir. Örneklem hacmi eşit olmayan iki veri setini karşılaştırmak anlamlı değildir. Değişim genişliği, değişim aralığını gösteren bir dağılım ölçüsüdür. Değişim genişliğinin hesaplanmasında sadece iki uç değer işleme alındığından, diğer değerlerin hiçbir etkisi yoktur. Bu nedenle değişim genişliği yaygın olarak kullanılan bir dağılım ölçüsü değildir.

3.2.2.Çeyrek Sapma:

Ortalama yerine medyan kullanıldığında ya da aşırı uç değerler bulunduğunda değişim genişliği yerine çeyrek sapma kullanılır.

$$\theta = \frac{\theta_3 - \theta_1}{2}$$

Eşitlikte,

Q : Çeyrek sapma

Q_1: Birinci çeyreklik

Q_3: Üçüncü çeyreklikti r.

Dağılımdaki bütün değerler kullanılmadığı için Q yeterli bir dağılım ölçüsü değildir.

3.2.3. Varyans & Standart Sapma:

Dağılımı tanımlayan iki önemli ölçü ortalama ve varyanstır. Standart sapma varyansın pozitif kareköküdür. Varyans ortalama etrafındaki yayılım miktarıdır. Varyans, birim değerlerinin ortalamadan sapmalarının kareler toplamının birim sayısına bölünmesi ile elde edilir. Kitle varyansı σ^2, örneklem varyansı s^2 ile gösterilir.

Standart sapma varyansın kareköküdür. Kitle standart sapması σ, örneklem standart sapması s ile gösterilir.

Sınıflandırılmamış Verilerde Varyans ve Standart Sapma

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

$S = \sqrt{S^2}$ varyans hiçbir zaman negatif çıkamaz.

eşitlikleriyle bulunur. Burada,

S: standart sapma
x_j: j. denek değeri
$\bar{x}$: aritmetik ortalama
n: birim sayısıdır.
$\sum_{i=1}^{n} x_i^2$: değerlerin kareleri toplamı

Örnek: Tablo 1.1'deki veriler için varyans

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_i)^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}{n-1} = \frac{20155,22 - 50(19,244)^2}{50-1} = 33.44$$

$$S = \sqrt{S^2} = \sqrt{33,44} = 5,78$$

Sınıflandırılmış Verilerde Varyans ve Standart Sapma

Frekans tablosundaki sınıf değeri ve frekans sütunundan yaralanılarak hesaplanan varyans ve standart sapma formülü,

$$S^2 = \frac{\sum_{i=1}^{k} f_i . S_i^2 - \frac{\left(\sum_{i=1}^{k} f_i . S_i\right)^2}{n}}{n-1}$$

$S = \sqrt{S^2}$

Eşitlikte,

f_i: i'inci sınıfın frekansı
S_i: i'inci sınıfın sınıf orta değeri
k: sınıf sayısı
n: birim sayısıdır.

Sınıf orta değeri (S_i)	**Frekans** (f_i)	$f_i \times S_i$	$f_i \times S_i^2$
7.5	2	15	112,5
10.1	4	40,4	408,04
12.7	3	38,1	483,87
15.3	6	93	1404,54
17.9	10	179	3204,1
20.5	8	164	3362
23.1	7	161,7	3735,27
25.7	5	128.5	3302,45
28.3	3	84,9	2402,67
30.9	2	61.8	1909,62
Toplam=		966,4	20325,06

$$S^2 = \frac{\sum_{i=1}^{k} f_i . s_i^2 - \frac{\left(\sum_{i=1}^{k} f_i . s_i\right)^2}{n}}{n-1} = \frac{20325,06 - \frac{(966,4)^2}{50}}{50-1} = \frac{20325,06 - 18678,58}{49} = \frac{1646,48}{49} = 33,601$$

$$S = \sqrt{S^2} = \sqrt{33,601} = 5,797$$

3.2.4. Standart Hata

Örneklem ortalamalarının oluşturduğu dağılımın standart sapması örneklem ortalamalarından her birinin standart hatası sayılır. Bir örneklemin ortalamasının standart hatası,

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}} \text{ veya } S_{\bar{x}} = \frac{s}{\sqrt{n}}$$

eşitlikleri ile hesaplanır.

Örnek : 50 ölçümden oluşan sülfür oksit miktarlarının tablo 1.1.'de verilen verilerin standart hatasını bulunuz.

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{33,44}{50}} = 0,83$$

3.2.5. Çarpıklık: Dağılım simetrik olduğunda ortalama medyan ve mod değerleri birbirine eşittir. Dağılımın simetrik olmadığı durumda dağılımlar arası çarpıklık karşılaştırmaları yapabilmek için bir katsayı tanımlanır. Bu katsayıya "çarpıklık katsayısı" denir.

$$ÇK = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3 / n}{S^3} = -0,11$$

$ÇK = 0$ ise dağılım ortalamaya göre simetriktir.

$ÇK < 0$ ise dağılım sola çarpık yada negatif yöne eğimlidir.

$ÇK > 0$ ise dağılım sağa çarpık yada pozitif yöne eğimlidir,

3.2.6. Basıklık Katsayısı:

Bir dağılımın normal dağılıma uygunluğu sadece simetrik olması ile ölçülmez. Dağılım gösterdiği yüksekliğe göre normal, dik ya da basık dağılım olarak nitelendirilebilir.

$$BK = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_i)^4}{S^4} = 2,58$$

$BK = 3$ ise dağılımın yüksekliği standart normal dağılımına uygundur.

$BK < 3$ ise dağılım standart normal dağılımından daha basıktır.

$BK > 3$ ise dağılım standart normal dağılıma göre daha sivridir(diktir).

KAYNAKLAR

1. Uygulamalı İstatistik (1994)

   Ayşen APAYDIN , Alaettin KUTSAL, Cemal ATAKAN

2. Olasılık ve İstatistik Problemler ve Çözümleri ile (2008)

   Prof. Dr. Semra ERBAŞ

3. Olasılık ve İstatistik (2006)

   Prof. Dr. Fikri Akdeniz

4. Olasılık ve İstatistiğe Giriş I-II (2011)

   Prof. Dr. Fikri Öztürk

5. Fikri Öztürk web sitesi

   http://80.251.40.59/science.ankara.edu.tr/ozturk/index.html