1.4. Merkezi Eğilim ve Dağılım Ölçüleri

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, çeyrekliklerdir. En sık kullanılan dağılım ölçüleri ise, değişim genişliği, çeyrek sapma, varyans, standart sapma, standart hata ve değişim katsayısıdır.

1.4.1. Merkezi Eğilim Ölçüleri

1.4.1.1. Aritmetik Ortalama

Aritmetik ortalama, en çok kullanılan merkezi eğilim ölçüsüdür. Birimlerin belirli bir değişken bakımından aldıkları değerlerin toplamının birim sayısına bölümü olarak tanımlanır. Eşit aralıklı ve oran ölçme düzeyinde ölçülen değişkenler için kullanılır. Aritmetik ortalama hem kitle hem de örneklem için hesaplanır.

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

Aritmetik ortalama sınıflandırılmamış ve sınıflandırılmış verilerde incelenecektir.

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 4.1: Tablo 2 de verilen ham verilerin aritmetik ortalamasını hesaplayınız.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{100} x_i}{100} = \frac{888.8400}{100} = 8.8884$$

Sınıflandırılmış Verilerde Aritmetik Ortalama:

Frekans tablosu düzenlenmiş verilerde aritmetik ortalama sınıf orta değeri ve frekans sütunundan yararlanılarak aşağıdaki gibi hesaplanabilir.

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

$\bar{x}$: aritmetik ortalama

S_i: i. sınıfın sınıf orta değeri

f_i: i. sınıfın frekans değeri

k : sınıf sayısı

n : örneklemdeki birim sayısı

Örnek 4.2: Tablo 2 ile verilen frekans tablosundan yararlanarak 100 kız bebeğin ağırlıklarına ilişkin aritmetik ortalamayı hesaplayınız.

S_i	f_i
6.0450	4
6.7950	5
7.5450	16
8.2950	18
9.0450	28
9.7950	10
10.5450	10
11.2950	7
12.0450	1
12.7950	1

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i S_i}{n} = \frac{\sum_{i=1}^{10} f_i S_i}{100} = (6.045 \times 4 + 6.795 \times 5 + \cdots . + 12.795 \times 1)/100 = \frac{888.75}{100}$$

$$= 8.8875$$

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.

1.4.1.2. Tepe Değer (Mod)

Bir veri grubunda en çok tekrarlanan değere tepe değer(mod) denir. Tepe değerin hesaplanmasında birimlerin büyüklük sırasına konulması şart değilse de, bu işlemin yapılması tepe değerin bulunmasında kolaylık sağlar.

Sınıflandırılmamış Verilerde Tepe Değer

Sınıflandırılmamış verilerde en çok tekrarlanan değer tepe değer olarak alınır.

Örnek 4.3. Bir iş yerinde emeklilik yaşına ilişkin veriler sırasıyla şöyledir:

65, 67, 69, 56, 57, 57, 58, 69, 69, 69, 60, 71, 72

Emeklilik yaşına ilişkin tepe değeri hesaplayınız.

Yukarıdaki veri gurubunda en çok tekrarlanan değer 69 olduğundan $TD = 69$ dur.

Sınıflandırılmış Verilerde Tepe Değer

Frekans tablosu oluşturulmuş gözlem değerleri için tepe değer hesaplanmak istendiğinde, ilk olarak tepe değer sınıfı belirlenir. Frekansı en yüksek olan sınıf, tepe değer sınıfı olarak adlandırılır ve tepe değeri bulmak için aşağıdaki eşitlik kullanılır.

$$TD = L + \frac{F_1}{F_1 + F_2} c$$

Bu eşitlikte;

TD: Tepe değer
L : En büyük sıklığa sahip sınıfın alt sınırı
c : Sınıf aralığı
F_1: En büyük sıklık ile bir önceki sınıfın sıklığı arasındaki fark
F_2: En büyük sıklık ile bir sonraki sınıfın sıklığı arasındaki fark

göstermektedir.

Örnek 4.5. Tablo.3 ile verilen frekans tablosundan yararlanarak 100 kız bebeğin ağırlıklarına ilişkin tepe değeri hesaplayınız.

Bebeklerin ağırlıkları	S_i	f_i
$5.67 \leq x < 6.42$	6.0450	4
$6.42 \leq x < 7.17$	6.7950	5
$7.17 \leq x < 7.92$	7.5450	16
$7.92 \leq x < 8.67$	8.2950	18
$8.67 \leq x < 9.42$	9.0450	28
$9.42 \leq x < 10.17$	9.7950	10
$10.17 \leq x < 10.92$	10.5450	10
$10.92 \leq x < 11.67$	11.2950	7
$11.67 \leq x < 12.42$	12.0450	1
$12.42 \leq x < 13.17$	12.7950	1

En çok frekansa sahip olan sınıf

$$TD = L + \frac{F_1}{F_1 + F_2} c$$

$$= 8.67 + 0.75\left(\frac{(28-18)}{(28-18)+(28-10)}\right) = 8.67 + 0.75\left(\frac{10}{10+18}\right) = 8.9379$$

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.
- Sınıflandırılmış verilerde ilk sınıf veya son sınıf tepe değere sahip ise kullanılmaz.
- Eğrisi J, ters J ve U şeklinde olan veriler için tepe değer kullanılmaz.

1.4.1.3 Ortanca (Medyan)

Bir veri gurubundaki değerler küçükten büyüğe sıralandığında tam ortaya düşen değer ortanca değeridir. Kitledeki birimlerin sayısı çok fazla ise verilerin özetlenmesinde merkezi eğilim ölçüsü olarak ortanca kullanılabilir. Ortanca, sınıflama ölçme düzeyi ile ölçülen değişkenler için kullanılmaz. Eşit aralıklı, oran ve sıralama ölçme düzeyinde ölçülen değişkenler için kullanılır.

Sınıflandırılmamış Verilerde Ortanca

Birim sayısının tek veya çift olmasına göre medyanın bulunması değişir. İki durumda da, ilk olarak eldeki veriler büyüklük sırasına göre (küçükten büyüğe veya büyükten küçüğe) sıraya konulur. Birim sayısı n ile gösterilmek üzere,

$$ortanca = OR = \begin{cases} x_{\mathrm{j}} & , \quad \mathrm{j} = \dfrac{\mathrm{n}+1}{2} \quad \text{n tek} \\ \dfrac{x_{\mathrm{j}} + x_{\mathrm{j}+1}}{2} & , \quad \mathrm{j} = \dfrac{\mathrm{n}}{2} \quad n\,çift \end{cases}$$

formülü ile hesaplanır.

Örnek 4.6. Aynı yaş grubunda bulunan 10 kız çocuğunun sahip olduğu Barbie bebek sayıları aşağıdaki gibi verilmiştir.

18,14,12,17,18,20,19,14,15,18

Barbie bebek sayılarının ortancasını bulunuz.

Verilerin küçükten büyüğe sıralanmış hali,

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
12	14	14	15	17	18	18	18	19	20

biçimindedir. Veri sayısı, $n = 10$ çift olduğu için ortaya düşen iki değerin ortalaması ortanca olacaktır. $\mathrm{j} = \frac{\mathrm{n}}{2} = \frac{10}{2} = 5$ ve $\mathrm{j} + 1 = 6$ olmak üzere,

$$OR = \frac{x_5 + x_6}{2} = \frac{17 + 18}{2} = 17.5$$

olarak bulunur.

Örnek 4.7. Bir ilacın etki süresi saat olarak aşağıdaki gibi gözlenmiştir.

5,6,5,3,7,9,6,4,8

Ortancayı bulunuz.

Veri sayısı, $n = 9$ tek olduğu için veriler sıralandığında ortaya düşen değer ortanca olacaktır. $\mathrm{j} = \frac{\mathrm{n}+1}{2} = \frac{10}{2} = 5$ olmak üzere verilerin küçükten büyüğe sıralanmış hali

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
3	4	5	5	6	6	7	8	9

$OR = x_5$

Sınıflandırılmış Verilerde Ortanca

Sınıflandırılmamış verilerde olduğu gibi, ilk olarak gözlem sayısının tek ya da çift olup olmadığına bakılır. Buna göre, gözlem sayısı çift ise $n/2$, gözlem sayısı tek ise $(n+1)/2$ değerleri bulunarak, ortanca değerinin bulunduğu sınıf belirlenir. Daha sonra,

$$ortanca = L + \frac{c}{f}\left(\frac{n}{2} - d\right)$$

eşitliği kullanılarak ortanca değer hesaplanır. Burada,

L : ortanca sınıfının alt sınıf değeri,

c : sınıf aralığı

f : ortanca sınıfın sıklığı

n : gözlem sayısı

d : ortanca sınıfından bir önceki sınıfın birikimli frekansı

göstermektedir.

Ortancanın özellikleri:

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

Örnek 4.8. Tablo 3 ile verilen frekans tablosundan yararlanarak 100 kız bebeğe ilişkin ortancayı hesaplayınız.

Bebeklerin ağırlıkları	S_i	f_i	Eklemeli f_i
$5.67 \leq x < 6.42$	6.0450	4	4
$6.42 \leq x < 7.17$	6.7950	5	9
$7.17 \leq x < 7.92$	7.5450	16	25
$7.92 \leq x < 8.67$	8.2950	18	43
$8.67 \leq x < 9.42$	9.0450	28	71
$9.42 \leq x < 10.17$	9.7950	10	81
$10.17 \leq x < 10.92$	10.5450	10	91
$10.92 \leq x < 11.67$	11.2950	7	98
$11.67 \leq x < 12.42$	12.0450	1	99
$12.42 \leq x < 13.17$	12.7950	1	100

Ortanca sınıfı $(100/2 = 50)$

Tablodan görülebileceği gibi, $c = 0.75$ ve $n = 100$ değerini alır. Ortanca sınıfı, $100/2 = 50$ değerini içeren sınıftır. Buna göre frekans tablosundan, $f = 28,\ d = 43$ olarak bulunur. Bulunan bu değerler ortanca eşitliğinde kullanılırsa,

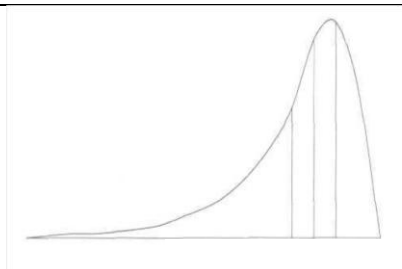
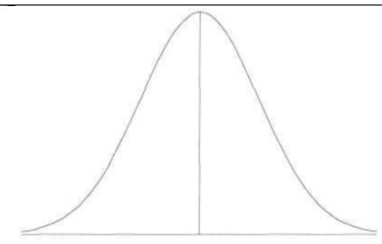
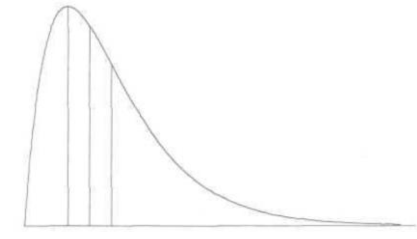
$$ortanca = L + \frac{c}{f}\left(\frac{n}{2} - d\right)$$

$$ortanca = 8.67 + \frac{0.75}{28}\left(\frac{100}{2} - 43\right) = 8.8575$$

olarak hesaplanır.

1.4.1.4 Aritmetik Ortalama, Tepe Değer ve Ortanca Arasındaki İlişki

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

		
AO<OR<TD Sola Çarpık	AO=OR=TD Simetrik	TD<OR<AO Sağa Çarpık

1.4.1.5 Çeyreklikler

Küçükten büyüğe doğru sıralanmış verileri dört eşit parçaya bölen değerlere çeyreklikler 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 çeyreklik, ortancadan büyük olan verilerin ortancası üçüncü çeyrekliktir.

Örnek 4.9. Aşağıda verilen 8 gözleme ait çeyreklikleri hesaplayınız.

40, 30, 34, 50, 75, 80, 20, 72

İlk olarak verilerin sıralanması gerekir,

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
20	30	34	40	50	72	75	80

$n = 8$ çift olduğu için $ortanca = \frac{x_4+x_5}{2} = \frac{40+50}{2} = 45$ dir. Ortancadan küçük olan değerlerin ortancası birinci çeyreklik, ortancadan büyük değerlerin ortancası üçüncü çeyrekliktir.

Ortancadan küçük olan değerler Ortancadan büyük olan değerler

x_1	x_2	x_3	x_4
20	30	34	40

$$Q_1 = \frac{30+34}{2} = 32$$

x_5	x_6	x_7	x_8
50	72	75	80

$$Q_3 = \frac{72+75}{2} = 73.5$$

Örnek 4.10. Orta yaşlı 9 kadına ait sistolik kan basınçları aşağıda verilmiştir.

141, 114, 122, 150, 136, 114, 103, 101

Çeyreklikleri hesaplayınız.

İlk olarak verilerin sıralanması gerekir,

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
101	103	114	114	122	124	136	141	160

Ortanca

$$Q_1 = \frac{103 + 114}{2} = 108.5 \qquad Q_3 = \frac{136 + 141}{2} = 138.5$$

1.4.2. Dağılım Ölçüleri

1.4.2.1 Değişim Genişliği

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

$$R = En\ büyük\ değer - En\ küçük\ değer$$

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.

Örnek 4.11. Aşağıdaki veriler için değişim genişliğini hesaplayınız.
20, 32, 15, 42, 26, 58, 27, 48, 25, 60, 23, 76

$R = 76 - 15 = 61$ dir.

1.4.2.2 Çeyrek Sapma

Ortalama yerine ortanca kullanıldığında ya da veri setinde aşırı uç değerler bulunduğunda değişim genişliği yerine çeyrek sapma kullanılır. Çeyrek sapma Q ile gösterilir.

$$Q = \frac{Q_3 - Q_1}{2}$$

Eşitlikte,

Q : Çeyrek sapma
Q_1: Birinci çeyreklik
Q_3: Üçüncü çeyrekliktir.

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

1.4.2.3 Varyans ve Standart Sapma

Varyans, gözlem değerlerinin ortalamadan sapmalarının bir ölçüsüdür. Gözlem değerlerinden elde edildiğinden, örneklem varyansı olarak adlandırılır ve s^2 ile gösterilir. 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

Varyans,

$$s^2 = \frac{\sum_{j=1}^{n}(x_j - \bar{x})^2}{n-1} = \frac{\sum_{j=1}^{n} x_j^2 - \frac{\left(\sum_{j=1}^{n} x_j\right)^2}{n}}{n-1}$$

standart sapma,

$$s = \sqrt{\frac{\sum_{j=1}^{n}(x_j - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{j=1}^{n} x_j^2 - \frac{\left(\sum_{j=1}^{n} x_j\right)^2}{n}}{n-1}} = \sqrt{Varyans}$$

eşitlikleriyle bulunur. Burada,

s: standart sapma
x_j: j. denek değeri
$\bar{x}$: aritmetik ortalama
n: birim sayısıdır.

Sınıflandırılmış Verilerde Varyans ve Standart Sapma

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

$$s^2 = \frac{\sum_{i=1}^{k} f_i S_i^2 - \frac{\left(\sum_{i=1}^{k} f_i S_i\right)^2}{n}}{n-1} \quad , \quad s = \sqrt{\frac{\sum_{i=1}^{k} f_i S_i^2 - \frac{\left(\sum_{i=1}^{k} f_i S_i\right)^2}{n}}{n-1}} = \sqrt{Varyans} \text{ dır.}$$

Eşitlikte,

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

Örnek 4.12. Tablo 3 ile verilen frekans tablosundan yararlanarak 100 kız bebeğin ağırlıklarına ilişkin varyans ve standart sapmayı bulunuz.

Bebeklerin ağırlıkları	S_i	f_i
$5.67 \leq x < 6.42$	6.0450	4
$6.42 \leq x < 7.17$	6.7950	5
$7.17 \leq x < 7.92$	7.5450	16
$7.92 \leq x < 8.67$	8.2950	18
$8.67 \leq x < 9.42$	9.0450	28
$9.42 \leq x < 10.17$	9.7950	10
$10.17 \leq x < 10.92$	10.5450	10
$10.92 \leq x < 11.67$	11.2950	7
$11.67 \leq x < 12.42$	12.0450	1
$12.42 \leq x < 13.17$	12.7950	1

$$s^2 = \frac{\sum_{i=1}^{k} f_i S_i^2 - \frac{\left(\sum_{i=1}^{k} f_i S_i\right)^2}{n}}{n-1} = \frac{8090.3 - \frac{789880}{100}}{99} = 1.9343$$
$$s = \sqrt{Varyans} = \sqrt{1.9343} = 1.3908$$

1.4.2.4.Standart Hata

Örneklem ortalamasının anlamlı olabilmesi için, standart hata ile birlikte gösterilmesi gerekmektedir. Standart hata genellikle yapılan örnekleme hatasının derecesini belirtmede kullanılan bir ölçüt olarak göz önüne alınabilir. Örneklem sayısı arttıkça standart hata azalır. Standart hata,

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

veya

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

eşitlikleri ile hesaplanır.

Örnek 4.13. Gözlem değerleri,

5, 7, 9, 12, 15, 6, 3, 8, 4, 8, 5, 2, 9, 12

için varyans 13.80769 olarak hesaplanır. Bu gözlem değerlerine ilişkin standart sapma,

$$s = \sqrt{s^2} = \sqrt{13.80769} = 3.71587$$

olarak bulunur.

1.4.2.5 Değişim Katsayısı

Farklı iki değişkene ilişkin gözlem değerleri verildiğinde, birinin diğerine göre daha homojen olup olmadığının örneklem varyansları ile söylenebilmesi için her iki gözleminde aynı ölçü birimine göre ölçülerek alınmış olması gerekir. Örneğin, birinci değişkene ilişkin gözlem değerleri ağırlık birimi ile ölçülerek alınmış, ikinci gözlem değerleri uzunluk birimi ile ölçülerek alınmış ise, bu iki gözlem değerini karşılaştırmak için sadece örneklem varyansına bakmak yanıltıcı olacaktır. Bu gibi durumlarda değişim katsayısı olarak tanımlanan ve standart sapmanın aritmetik ortalamaya bölümü ile hesaplanan ve birimsiz olan bir ölçüt kullanılır. Değişim katsayısı olarak adlandırılan bu ölçüt,

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

eşitliği kullanılarak hesaplanır. Değişkenliğin % olarak ölçüsünü verir. Değişim katsayısı büyük ise, gözlem değerlerinin aritmetik ortalamadan uzak, küçük ise gözlem değerlerinin aritmetik ortalamaya yakın olduğu şeklinde yorum yapılabilir.

Örnek 4.14. 6 yaşındaki 10 tane kız çocuğuna ait boy ve kilolar ölçülmüş ve çocuklar boylarına göre mi yoksa kilolarına göre mi daha homojen bir dağılım sergiliyorlar görülmek istemiştir. Bu iki değişkene ilişkin gözlem değerleri aşağıda verildiği gibidir.

Buna göre değişim katsayısı hesaplanarak istenilen cevap elde edilebilir.

	$Boy(x)$	$kilo(y)$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
	125	25	9.61	9
	121	18	0.81	16
	124	28	4.41	36
	118	22	15.21	0
	122	19	0.01	9
	125	21	9.61	1
	118	16	15.21	36
	119	26	8.41	16
	123	26	1.21	16
	124	19	4.41	9
Toplam	1219	220	689	148
	$\bar{x}$ =121.9	$\bar{y}$ =22	S_x^2 =7.66	S_y^2 =16.44
			S_x =2.77	S_y =4.06
			$DK_1 = \frac{S_x}{\bar{x}} = 0.023$	$DK_2 = \frac{S_y}{\bar{y}} = 0.184$

Tablo 5. Kız çocuklarının boy ve kilolarına ilişkin değişim katsayısı

Yukarıdaki tablo incelendiğinde, boy değişkenine göre değişim katsayısı $DK_1 \times 100 = 2.3$ ve kilo değişkenine ilişkin değişim katsayısı $DK_2 \times 100 = 18.4$ olarak bulunur. Buna göre, çocukların boylarına göre daha homojen bir dağılım gösterdikleri söylenebilir.