Tanımlayıcı İstatistikler

Yrd. Doç. Dr. Emre ATILGAN

Tanımlayıcı İstatistikler

- **Yer Gösteren Ölçüler**
  - **Merkezi Eğilim Ölçüleri**
  - **Konum Ölçüleri**
- **Yaygınlık Ölçüleri**

Tanımlayıcı İstatistikler

Yer Ölçüleri

1)Aritmetik ort.
2)Geometrik ort.
3)Harmonik ort.
4)Mod
5)Medyan
6)Kartiller

Değişkenlik Ölçüleri

1) Range (Değişim Aralığı)
2) Ort. Mutlak sapma
3) Varyans
4) Standart Sapma
5) Değişkenlik(Varyasyon) Katsayısı

Çarpıklık Ölçüleri

1)Pearson Asimetri Ölçüsü
2)Bowley Asimetri Ölçüsü

Basıklık Ölçüleri

Aritmetik Ortalama

- Aritmetik ortalama, veri setindeki tüm değerlerin toplanması ve bu toplamın veri sayısına bölünmesiyle elde edilir.

Örnek 2: 9 kişinin yaşları 12, 13, 11, 12, 14, 29, 12, 13, 11 olsun. Buna göre yaş ortalaması

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{12+13+11+12+14+29+12+13+11}{9} = 14.11$$

- **Aritmetik ortalama dağılımdaki tüm değerleri dikkate alır. Ancak dağılımdaki aşırı değerlerden etkilenir. Bu dağılımda 29 yaş aşırı bir değerdir ve ortalamayı etkiler ve aritmetik ortalamanın yüksek çıkmasına neden olur.**

Gruplanmış Serilerde Aritmetik Ortalama

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

$$\sum_{i=1}^{k} f_i = n$$

f : frekans

k: grup sayısı

i = 1,2,3,..........,k

Grup	Frekans	$x_i f_i$
51	1	**51**
66	3	**198**
72	4	**288**
82	5	**410**
94	7	**658**
	$\sum f_i$ =**20**	**1605**

Örnek: Yandaki tabloda bir Samsung bayisindeki LCD televizyonların ekran boyutlarına göre satış miktarları verilmiştir. Frekans dağılımının aritmetik ortalamasını hesaplayınız.

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{51(1) + 66(3) + ... + 94(7)}{1+3+4+5+7}$$

$$= \frac{1605}{20} = 80,25$$

Örnek:

Bir ana kütlede 20 adet birim bulunmaktadır. Bu birimlerin bir F değişkeni için ölçüm sonuçları izleyen basit seride verilmiştir. Ana kütle aritmetik ortalamasını basit seri ve frekans serisi eşitlikleri yardımıyla hesaplayınız.

1	1	1	2	2
2	2	3	3	3
3	3	3	3	4
4	4	5	5	6

a) Ana kütle aritmetik ortalaması (Basit Seri):

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{N} = \frac{1 + \dots + 5 + 6}{20} = \frac{60}{20} = 3$$

b) Gruplanmış Frekans Serisi Aritmetik Ortalaması

Gözlem Değeri	Frekans	Gözlem Değeri x Frekans
1	3	1 x 3 = 3
2	4	2 x 4 = 8
3	7	3 x 7 = 21
4	3	4 x 3 = 12
5	2	5 x 2 = 10
6	1	6 x 1 = 6
Toplam	20	$\sum f_i x_i = 60$

$$\mu = \frac{\sum_{i=1}^{N} f_i x_i}{N} = \frac{60}{20} = 3$$

Sınıflanmış Seriler İçin Aritmetik Ortalama

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

f : frekans

k : sınıf sayısı

i = 1,2,3,.........,k

m : sınıf orta noktası

$$\sum_{i=1}^{k} f_i = n$$

- Sınıflanmış serilerde her bir sınıf içindeki değerlerin neler olduğu bilinmediğinden dolayı ve yalnızca her bir sınıfın frekans değerleri bilindiğinden dolayı sınıfı temsil etmek üzere sınıf orta noktaları hesaplamada kullanılır.
- Kullanılan formül gruplanmış seriler için kullanılan formüle benzerdir.

Örnek: Aşağıdaki tabloda 30 günlük süre içinde bir restoranın kullandığı et miktarının dağılımı verilmiştir. Günlük kullanılan et miktarının aritmetik ortalamasını hesaplayınız.

Sınıflar	f_i	m_i	$m_i f_i$
30-36'dan az	2	33	66
36-42'den az	6	39	234
42-48'den az	10	45	450
48-54'dan az	7	51	357
54-60'den az	4	57	228
60-66'den az	1	63	63
Toplam	30		1398

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{\sum_{i=1}^{k} f_i} = \frac{33(2) + 39(6) + \ldots + 63(1)}{30}$$

$$= \frac{1398}{30} = 46{,}6\ kg.$$

Örnek:

İzleyen tabloda verilen gruplandırılmış frekans serisi için ana kütle aritmetik ortalaması μ'yü hesaplayınız.

Ağırlık	Frekans
0 - 50	3
50 - 100	5
100 - 150	9
150 - 200	2
200 - 250	1
Toplam	20

Gruplandırılmış frekans serilerinde öncelikle her sınıfın orta noktası belirlenerek bir sütunda gösterilir. Daha sonra bu orta nokta değerleri ile ilgili sınıf frekansları çarpılarak yeni bir sütun oluşturulur. İzleyen tabloda bu işlemler gösterilmiştir.

Ağırlık	Frekans	Orta Değer	Orta değer x Frekans	
0 - 50	3	25	25 x 3 = 75	
50 - 100	5	75	75 x 5 = 375	
100 - 150	9	125	125 x 9 = 1.125	
150 - 200	2	175	175 x 2 = 350	
200 - 250	1	225	225 x 1 = 225	
Toplam	20		$\Sigma f_i x_i$ = 2.150	

$$\mu = \frac{\sum_{i=1}^{N} f_i x_i}{N} = \frac{2.150}{20} = 107{,}50$$

GEOMETRİK ORTALAMA

- ✓ Bir veri setinde bulunan n adet elemanın çarpımının n'nci dereceden kökünün alınmasıyla elde edilen yer ölçüsüdür.

$$G \cdot O = \sqrt[n]{(x_1) \cdot (x_2) \ldots \cdot (x_n)}$$

- ✓ Geometrik ortalamanın formülüne bakıldığında hesaplama zorluğu olduğundan dolayı logaritma ifadesi kullanılır. Genellikle basit seriler için kullanışlı olup negatif sayılar için kullanışlı değildir.

$$\mathrm{LogG} = \frac{\sum_{i=i}^{n} \log x_i}{n}$$

Geometrik Ortalama'nın Kullanım Alanları

Ortalama oranları,

- Değişim Oranları,
- Logaritmik dağılış gösteren veri setleri,

için kullanışlıdır.

Örnek: fiyat indeksleri, faiz formülleri.

Örnek:

Bir bakterinin 5 farklı zamanda çoğalma miktarları (yüzde olarak) 10, 17, 42, 21 ve 19 olarak hesaplanmıştır. Bakteri çoğalma miktarları için geometrik ortalamayı hesaplayınız.

İlk olarak geometrik ortalamanın standart formülünü kullanalım. Bu problem için geometrik ortalama,

$$G.O = \sqrt[5]{(10)(17)(42)(21)(19)} = 19{,}54$$

olur. Bu çalışmada bakterilerin ortalama çoğalma miktarları yaklaşık %20'dir. Aynı geometrik ortalama değerine logaritma kullanılarak da ulaşılabilir. Eşitliğe göre,

$$\log(G.O) = \frac{1}{5}\left(\log(10) + \log(17) + \log(42) + \log(21) + \log(19)\right) = 1{,}290$$

olur. Geometrik ortalamamız bu değerin anti-logaritması alınarak yine 19,54 (yaklaşık %20) bulunur.

Medyan (Ortanca)

Bir veri grubu küçükten büyüğe sıralandığında, terim sayısı tek ise ortadaki sayı, çift ise ortadaki iki sayının toplamının yarısıdır.

Örnek 3: 9 kişinin yaşları küçükten büyüğe doğru sıralandığında

11, 11, 12, 12, 12 , 13, 13, 14, 29

Gözlem sayısı tektir. **Ortanca =(9+1)/2=5. değer**

Denek sayısı çift olduğunda Ortanca

Ör: Denek sayısı 10 ve yaşlar aşağıdaki gibi olsaydı

12, 13, 11, 12, 14, 29, 12, 13, 15 11

Yaşlar sıraya dizildiğinde

11 11 12 12 12 13 13 14 15 29

Denek sayısı çift olduğundan

Ortanca (n/2)=5 ve (n+2)/2=6 değerlerinin ortalamasıdır.

$$\text{Ortanca} = \frac{12 + 13}{2} = 12.5$$

-

- ✓ **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.

Medyan (Ortanca)

$$Medyan = L + \frac{c}{f_m}\left(\frac{n}{2} - f_k\right)$$

Eşitlikte,

- L; medyanı içerdiği düşünülen medyan sınıfının alt limitini,
- c; medyan sınıfının sınıf aralığını,
- f_m; medyan sınıfının frekansını,
- n; toplam terim sayısını,
- f_k; medyan sınıfından önce yer alan kümülatif frekansı

Örnek: Bir test sonucunda elde edilen gözlem sonuçları 80, 84, 89, 90, 68, 75, 78, 79, 94 olarak verilmiştir. Bu serinin medyanı kaçtır?

Sıralı Seri: 68, 75, 78, 79, 80, 84, 89, 90, 94

n=9 (tek Sayı)

Medyan=(n+1)/2 = (9+1)/2 = 5. gözlem

Cevap: 80

Örnek: Bir test sonucunda elde edilen gözlem sonuçları 80, 84, 89, 90, 68, 75, 78, 79, 94, **92** olarak verilmiştir. Bu serinin medyanı kaçtır?

Sıralı Seri: 68, 75, 78, 79, 80, 84, 89, 90, 92, 94

n=10 (çift sayı)

Medyan=(n)/2 ve (n+2)/2 değerlerinin ortalaması

Medyan=10/2 ve (10+2)/2 değerlerinin ortalaması
= 5. ve 6. değerlerin ortalaması

Cevap: (80 + 84)/2 = 82

Örnek: *Aşağıda verilen gruplandırılmış frekans serisi için medyan değerini hesaplayınız.*

Hemoglobin (g/dl)	Hasta Sayısı	Kümülatif Frekans
4 - 6	2	2
6 - 8	7	9
8 - 10	14	23
10 - 12	6	29
12 - 14	5	34
Toplam	34	

$$Medyan = L + \frac{c}{f_m}\left(\frac{n}{2} - f_k\right)$$

Eşitlikte,

- L; medyanı içerdiği düşünülen medyan sınıfının alt limitini,
- c; medyan sınıfının sınıf aralığını,
- f_m; medyan sınıfının frekansını,
- n; toplam terim sayısını,
- f_k; medyan sınıfından önce yer alan kümülatif frekansı

Öngörülen Medyan Sınıfı
= 34/2 = 17. gözlem

17. Gözlem 8-10 grubunda yer alır.

L=8

$$Medyan = 8 + \frac{2}{14}\left(\frac{34}{2} - 9\right) = 9{,}143$$

MOD

- ✓ Bir seride en çok tekrarlanan terimin değerine **mod** denir.
- ✓ Veri setinin modu olmayacağı gibi, birden fazla da modu olabilir.
- ✓ Mod genellikle kesikli şans değişkenli için oluşturulan gruplanmış serilerde aritmetik ortalama yerine kullanılabilir

Basit Seriler İçin Mod

Örnek: Bir fabrikada çalışan 5 endüstri mühendisinin bildiği yabancı dil sayıları aşağıda verilmiştir. Buna göre bu mühendislerin bildiği yabancı dil sayısının modunu hesaplayınız.

x_i : 2,0,1,2,0,1,0 0,0,0,1,1,2,2.

Veri setinde en çok tekrar eden eleman 0 olduğundan (3 kez) mod değeri 0 'dır.

- Eğer veri seti 1,0,1,2,0,1,0 şeklinde olsaydı veri seti iki modlu olacaktı. (0 ve 1)
- Eğer veri seti 2,0,1,2,0,1 şeklinde olsaydı veri setinin modunun olmadığı ifade edilecekti.

Gruplanmış Seriler İçin Mod

Örnek: Aşağıdaki tabloda bir Samsung bayisindeki LCD televizyonların ekran boyutlarına göre satış miktarları verilmiştir. Frekans dağılımının aritmetik ortalamasını hesaplayınız.

Ekran	Satış Adedi
51	1
66	3
72	4
82	5
94	7

- ✓ **Frekans dağılımına bakıldığında en fazla satış miktarı 94 ekran LCD televizyonda olduğundan dolayı (7 adet) dağılımın modunun 94 olduğu söylenir.**

- ✓ Eğer 82 ekran LCD televizyonlarından da 7 adet satılsaydı dağılımın iki modu olduğu ifade edilirdi. (82 ve 94)

Sınıflanmış Seriler İçin Mod

- ✓ Sınıflanmış serilerde mod değeri hesaplanırken ilk olarak mod sınıfı belirlenir.
- ✓ Mod sınıfı frekansı en yüksek olan sınıftır.
- ✓ Mod sınıfı belirlendikten sonra bu sınıf içerisinde yer alan modun tam değeri sınıf frekansı ve kendine komşu olan sınıf frekansları dikkate alınarak hesaplanır.

$$\textbf{Mod} = L_{\text{mod}} + \frac{\Delta_1}{\Delta_1 + \Delta_2} . i$$

L_{Mod} = Mod Sınıfı Aralığının Alt Sınırı

Δ_1 = Mod Sınıfı Frekansı - Kendinden Bir Önceki Sınıf Frekansı

Δ_2 = Mod Sınıfı Frekansı – Kendinden Bir Sonraki Sınıf Frekansı

i = Mod Sınıfının Sınıf Aralığı

Örnek: Aşağıdaki tabloda 30 günlük süre içinde bir restoranın kullandığı et miktarının dağılımı verilmiştir. Günlük kullanılan et miktarının modunu hesaplayınız.

Sınıflar	f_i
30-36'dan az	2
36-42'den az	6
42-48'den az	10
48-54'dan az	7
54-60'den az	4
60-66'den az	1
Toplam	30

Mod sınıfı

$$Mod = L_{mod} + \frac{\Delta_1}{\Delta_1 + \Delta_2}.i$$

$$= 42 + \frac{(10-6)}{(10-6)+(10-7)}.6 = 45,4\ kg.$$

Örnek:

Bir araştırmada ilgilenilen değişkenin aldığı 10 değer 7, 9, 11, 11, 11, 14, 14, 16, 16, 17 olarak ortaya çıkmıştır. Bu basit seride en çok tekrar sayısına 3 tekrar ile 11 değeri sahip olduğundan bu basit serinin modu 11'dir. Bu araştırmada daha fazla gözlemin elde edildiği ve aşağıdaki gruplandırılmış frekans serisinin oluşturulduğu varsayılsın. Mod değerini tekrar hesaplayalım. Yeni veriler ile hazırlanan gruplandırılmış frekans serisi aşağıda verilmiştir.

Gözlem	Frekans
0 - 4	21
4 - 8	34
8 - 12	45
12 - 16	49
16 - 20	42
Toplam	191

En yüksek frekans 12 -16 sınıfı içindir. Mod sınıfı bu sınıf olarak alınmalıdır. İstenen mod değeri eşitlik yardımıyla,

$$Mod = 12 + \frac{4}{4+7}4 = 13,45$$

DEĞİŞKENLİK (YAYGINLIK) ÖLÇÜLERİ

- **Bir dağılımdaki değerlerin, birbirlerine ya da kendi ortalamalarına göre farklılıklarını gösterir.**

- Bu farklılıkların derecesi dağılımın yaygınlığı kavramını oluşturur. İki dağılım aynı ortalama ya da ortanca değerine sahipken yaygınlıkları farklı olabilir.

Yaygınlık Ölçüleri

- **Bir dağılımdaki değerlerin, birbirlerine ya da kendi ortalamalarına göre farklılıklarını gösterir.**

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

Dağılım I		**Dağılım II**	
6		3	
1		7	
6	$\overline{X} = 6$ Ortanca $= 6$	6	$\overline{X} = 6$ Ortanca $= 6$
15		5	
6		6	
2		9	

Dağılım I'deki değerlerin aritmetik ortalamaya olan uzaklığı dağılım II'ye göre daha fazladır.

Dağılım I, dağılım II'ye göre daha yaygındır.

Dağılımların yaygınlığı hakkında bilgi veren ve en çok kullanılan ölçüler ;

- **Dağılım (Değişim) Aralığı**
- **Standart Sapma**
- **Varyans**
- **Çeyreklikler Arası Genişlik**
- **Çeyrek Sapma**
- **Değişim Katsayısı**

Dağılım Aralığı

- ✓ Dağılım aralığı en basit yaygınlık ölçüsüdür.

  Dağılımdaki en büyük değerden en küçük değerin çıkartılması ile bulunur.
  R ile gösterilir

  R= En Büyük Değer-En Küçük Değer

Dağılım Aralığı

- 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.

Varyans ve Standart Sapma

Veri setindeki her bir gözlem değerinin aritmetik ortalamadan farklarının karelerinin toplamının örnek hacminin bir eksiğine bölünmesinden elde edilen değişkenlik ölçüsüne örnek varyansı adı verilir

- Varyans ve Standart Sapma bir dağılımın yaygınlığını gösteren en önemli yaygınlık ölçülerinden biridir.
- Dağılımdaki tüm değerlerin aritmetik ortalamaya olan uzaklıklarının ortalamasıdır.
- Dağılımın yaygınlığı arttıkça standart sapma büyür.
- Dağılımdaki değerler aynı ise yaygınlık yoktur ve standart sapma sıfırdır.

Varyans ve Standart Sapma

$$Varyans = (standart\ sapma)^2$$

Ya da

$$Standart\ Sapma = \sqrt{Varyans}$$

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

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

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

Frekans serilerinde standart sapma

Frekans serilerinde ana kütle standart sapması:

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

Frekans serilerinde ana kütle standart sapması:

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

Örnek I: Aşağıdaki Dağılım için Standart Sapma hesaplayınız.

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

Eşitliğine göre standart sapma hesaplanması

x_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$
6	0	0
1	-5	25
6	0	0
15	9	81
6	0	0
2	-4	16
		122

$$S = \sqrt{\frac{122}{6-1}} = 4.94$$

Örnek II: Aşağıdaki Dağılım için Standart Sapma hesaplayınız.

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

Eşitliğine göre standart sapma hesaplanması

x_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$
3	-3	9
7	1	1
6	0	0
5	-1	1
6	0	0
9	3	9
		20

$$S=\sqrt{\frac{20}{6-1}}=2$$

Örnek: İstatistik I dersini alan 10 öğrencinin vize notları aşağıdaki gibi sıralanmıştır. Buna göre vize notları için varyans ve standart sapmayı hesaplayınız.

30,41,53,61,68,79,82,88,90,98

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{30+41+....+98}{10} = 69$$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{(30-69)^2 + (41-69)^2 + ... + (98-69)^2}{9}$$

$$= \frac{4538}{9} \approx 504{,}22$$

$$s^2 \approx 504{,}22 \quad \rightarrow s = \sqrt{s^2} = \sqrt{504{,}22} \approx 22{,}45$$

İstatistik I vizesinden alınan notların ortalama etrafında yaklaşık olarak 22 puan değiştiği görülmektedir.

Örnek: Aşağıda verilen frekans serisi için varyansı ve standart sapmayı hesaplayınız.

Kalori	Ürün Sayısı
100	17
200	34
300	30
400	19
Toplam	100

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{n} = \frac{25.100}{100} = 251$$

Kalori	Ürün Sayısı	$(x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x}$
100	17	100 - 251 = -151	22.801	22.801 x 17 = 387.617
200	34	200 - 251 = - 51	2.601	2.601 x 34 = 88.434
300	30	300 - 251 = 49	2.401	2.401 x 30 = 72.030
400	19	400 - 251 = 149	22.201	22.201 x 19 = 421.819
Toplam	100			969.900

$$s = \sqrt{\frac{\sum_{i}^{n} \left(x_i - \bar{x}\right)^2 f_i}{n-1}} = \sqrt{\frac{969.900}{100-1}} = \sqrt{9.796,9697} = 98,979$$

Varyans:

$$s^2 = (98,979)^2 = 9.796,9697$$

FREKANS DAĞILIMLARINDA SİMETRİ VE ASİMETRİ

- ✓ **Simetrik serilerde:**

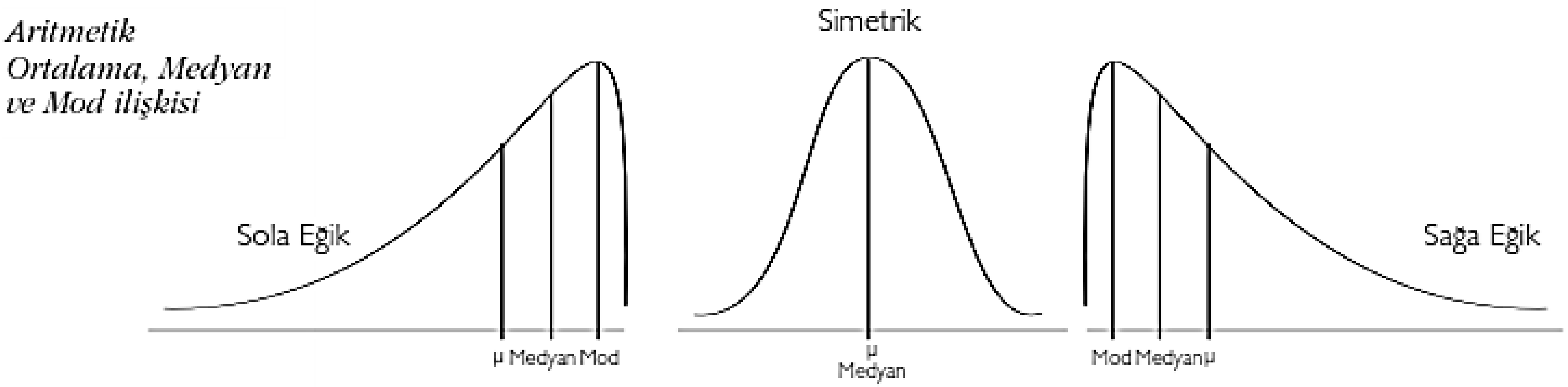
Aritmetik Ortalama = Medyan = Mod

- ✓ **Sola eğik frekans dağılımlarında:**

Aritmetik Ortalama < Medyan < Mod

- ✓ **Sağa eğik frekans dağılımlarında:**

Aritmetik Ortalama > Medyan > Mod



Eğiklik Katsayısı

- İlgilenilen değişkenin frekans dağılımının simetri derecesini tespit etmek amacı ile eğiklik katsayısı hesaplanır.
- Eğikliğin hesaplanması için bir kaç teknik bulunmakla birlikte bunların en basit olanı Karl Pearson (1837-1936) tarafından geliştirilmiştir.
- Aritmetik ortalaması, medyanı ve standart sapması bilinen bir veri setinde Pearson eğiklik katsayısı izleyen eşitlik yardımıyla hesaplanır.

$$Sk_p = \frac{3(\bar{x} - med)}{s}$$

Sk_P < 0 →Negatif çarpık(Sola)

Sk_P > 0 → Pozitif Çarpık(Sağa)

Sk_P = 0 ise dağılış simetrik

Örnek: Aşağıdaki tabloda 30 günlük süre içinde bir restoranın kullandığı et miktarının dağılımından elde edilen bazı tanımlayıcı istatistikler verilmiştir. Buna göre pearson asimetri ölçüsünü hesaplayıp yorumlayınız

Aritmetik Ort.	Medyan	Varyans
46,6	46,2	54,46

$$Sk_p = \frac{3(\bar{x} - med)}{s} = \frac{3(46,6 - 46,2}{\sqrt{54,46}} \approx 0,16 > 0$$

Sağa Çarpık
Pozitif Asimetri

CHEBYSHEV TEOREMİ

- ✓ Küçük standart sapma değerinin serideki terimlerin aritmetik ortalama etrafında yoğunlaştığını ve büyük bir standart sapma değerinin de terimlerin ortalamadan uzak bir yayılıma sahip olduğunun bir göstergesi olabileceği belirtilmişti.
- ✓ Herhangi bir veri kümesinde, ilgilenilen değişkenin frekans dağılımının şekline bakmaksızın
- ✓ (simetrik, sağa eğik veya sola eğik gibi) aritmetik ortalama değerinden belirli bir standart sapma uzaklıkta yer alan birimlerin en küçük oranı **P.L. Chebyshev** tarafından bir teorem yardımıyla gösterilmiştir.

CHEBYSHEV TEOREMİ

- ✓Herhangi bir veri seti için (örneklem veya ana kütle) aritmetik ortalamadan k standart sapma uzaklıkta, k>1 olmak üzere, yer alacak terimlerin en düşük oranı $1-(1/k^2)$ olur.

- ✓Chebyshev teoremine göre birimlerin %75’i aritmetik ortalamadan artı eksi 2 standart sapma aralığında değerler almaktadır.

Örnek: Bir süt ürünleri fabrikasında üretilen yoğurtlar 500 gr.’lık kutularda piyasaya sürülmektedir. Fabrika kalite kontrol sorumlusu, son 45 dakika içinde üretilen 500 gr.’lık yoğurtlardan 35 tanesini alarak tartılmalarını istemiştir. Tartılan yoğurtların ortalaması 502 gr. ve standart sapması da 1 gr. olarak hesaplanmıştır. Üretilen 500 gr.’lık yoğurtların en az yüzde kaçı artı eksi 3,5 standart sapma aralığında yer almaktadır?

Çözüm: Burada yoğurt ağırlığı değişkenin frekans dağılımı hakkında hiç bir bilgi verilmemiştir. Dağılımın simetrik ya da asimetrik olduğu konusunda bir bilgi yoktur. Dolayısıyla çözüm için Chebyshev teoremi kullanılabilir. İstenen aralık 3,5 standart sapma aralığı olduğu için teoreme göre yoğurtların en az:

$$1-\left(\frac{1}{k^2}\right)=1-\left(\frac{1}{3,5^2}\right)=0,92$$

%92’si 500 ± 3,5(1) aralığında yer almaktadır.