

TANIMLAYICI İSTATİSTİKLER (VERİLERİN ÖZETLENMESİ)

TANIMLAYICI İSTATİSTİKLER

Bundan önceki sunuda verilerin özetlenmesi için nicel ve nitel verilerde frekans tablolarının elde edilmesi ve grafiklenmesi üzerinde duruldu.

Bu derste yine istatistiğin “Tanımlama” görev kapsamı içinde yapılabilecek istatistik hesaplamalar üzerinde durulacaktır.

- Merkezi Eğilim Ölçüleri
- Dağılım (Değişim) Ölçüleri

MERKEZİ EĞİLİM VE YER ÖLÇÜLERİ

Sayı yığınlarının kolayca anlaşılması için sayı yığınlarının en fazla yığıldığı bölgeyi tarif eden tipik değerlerin verilmesi gerekir. Bu değerler dağılışın merkezini gösterdikleri için merkezi eğilim ölçüleri olarak da bilinir.

Dağılışın merkezini gösteren değişik ölçüler vardır. Bunlar:

- Aritmetik ortalama
- Ortanca(medyan)
- Tepe değer(mod)
- Geometrik ortalama
- Harmonik ortalama
- Tartılı Ortalama vb ölçülerdir.

Sınıflandırılmamış Verilerde Aritmetik Ortalama

Örnek: 20 pnömoni hastası için hastalık süreleri (gün) aşağıdaki şekilde bulunmuştur.

X_i : { 6,7,8,8,10,11,11,11,8,10,10,10,12,12,14, 14,12,7,10,11} , aritmetik ortalama:

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

$n=20$, Toplam(X_i)=202, buna göre;
Aritmetik Ortalama= $202/20=10.1$

FREKANS(SIKLIK) TABLOSU ÖRNEĞİNİ HATIRLAYALIM

Yüz yeni doğan çocuğun doğum ağırlıkları (kg) aşağıdaki gibi bulunmuştur.

2,6	2,6	3,0	2,7	3,8	2,7	3,1	2,7	4,8	4,9
3,6	4,0	4,8	2,7	3,9	2,4	2,0	3,4	2,0	3,9
3,6	4,7	2,3	2,4	3,7	2,7	1,7	2,9	2,4	2,1
5,1	3,2	3,1	4,1	3,0	4,3	3,8	3,3	2,4	3,2
1,7	5,0	4,4	2,6	3,0	3,6	4,7	2,6	2,8	2,7
3,1	2,0	2,4	3,5	2,2	3,9	2,9	3,6	2,6	4,0
2,4	3,3	4,1	2,4	2,2	3,1	4,2	3,9	2,4	3,3
3,8	3,5	5,2	2,5	3,4	3,2	3,8	4,1	2,7	3,7
3,3	4,3	3,4	2,1	2,9	3,8	2,2	3,2	2,8	2,9
2,7	5,0	3,1	2,9	2,7	3,5	2,0	2,5	4,7	2,5

Sınıf genişliği **0.5** kg olacak şekilde doğum ağırlıklarını sıklık(frekans) tablosu halinde özetlemek için **8** sınıf yapılabilir.

Sınıf aralığı	Sınıf sınırları	Sıklık(frekans)
1.5 - 1.9	1.45 - 1.94	2
2.0 - 2.4	1.95 - 2,44	18
2.5 - 2.9	2,45 – 2,94	24
3.0 - 3.4	2,95 – 3,44	19
3.5 - 3.9	3,45 – 3,94	18
4.0 - 4.4	3,95 – 4,44	9
4.5 - 4.9	4,45 – 4,94	6
5.0 - 5.4	4,95 – 5,44	4
Toplam		100

SINIFLANDIRILMIŞ VERİLERDE ARİTMETİK ORTALAMA

Sınıf sınırları	Sınıf Değeri (X_i)	Sıklık (frekans= f_i)	$f_i \cdot X_i$
1.45 - 1.94	1.7	2	3,4
1.95 - 2,44	2,2	18	39,6
2,45 – 2,94	2,7	24	64,8
2,95 – 3,44	3,2	19	60,8
3,45 – 3,94	3,7	18	66,6
3,95 – 4,44	4,2	9	37,8
4,45 – 4,94	4,7	6	28,2
4,95 – 5,44	5,2	4	20,8
	TOPLAM	100	322

- Frekans tablosunda sınıf orta değerleri (X_i) bulunur. Frekanslarla (f_i) birlikte bu değerler kullanılarak ortalama hesaplanır.
- Şöyle ki: k = sınıf sayısı, yani örnek için $k=8$ olur.

$$\bar{X} = \frac{\sum_{i=1}^k f_i \cdot X_i}{\sum_{i=1}^k f_i} = \frac{f_1 \cdot X_1 + f_2 \cdot X_2 + \dots + f_k \cdot X_k}{f_1 + f_2 + \dots + f_k}$$

Aritmetik Ortalama= $322/100= 3.22$

Olur. Aslında grafik çizilmeyecekse frekans tablosu elde etmeye gerek yoktur.

ORTANCA(MEDYAN)

Sayılar kümesini iki eşit yarıya bölen değer ortanca değerdir.

Gözlem değerleri büyüklik sırasına dizildiğinde tam ortada kalan değerdir.

Örnek: $X_i: \{ 3, 1, 13, 27, 6, 8, 6 \}$ gözlem değerlerinin ortancası nedir?

Sayılar büyüklik sırasına dizilirse,

$\{ \underline{1}, \underline{3}, \underline{6}, 6, \underline{8}, \underline{13}, \underline{27} \}$

Ortada kalan sayı 6 olduğundan

Ortanca=6 olur.

Gözlem adedi çift ise ortada iki sayı kalacağından ortanca değer bu iki değer ortalaması olur.

Örnek: $X_i: \{ 21, 9, 8, 3, 7, 9 \}$

Gözlemler büyüklük sırasına dizilirse,

$X_i : \{ \underline{3}, \underline{7}, 8, 9, \underline{9}, \underline{21} \}$

Ortada kalan iki değer ortalaması ortancadır:

$\text{Ortanca} = (8+9)/2 = 8.5$

olur.

Diğer bir ifade ile $(N+1)/2$ nci değer ortanca değerdir. Gözlem değeri çift ise sonuç şöyle bulunur.

$(N+1)/2 = (6+1)/2 = 3.5$, yani 3 üncü ve 4 üncü gözlemlerin ortalaması ortancadır.

Sınıflandırılmış verilerden ortanca hesaplamak için önce medyan sınıfının bulunması gerekir. Bunun içinden az eklemeli frekans (... den az F_i) bulunur ve bu kullanılarak $(N+1)/2$ nci gözlemin düştüğü sınıf ortanca sınıfı olarak tanımlanır.

Örnek: Çocuk doğum ağırlığı ile ilgili frekans tablosunda $N=100$ olduğu için

$$(N+1)/2=(100+1)/2=50.5,$$

yani 50 nci ve 51 nci gözlemin düştüğü sınıf ortanca sınıfıdır.

- N çift ise aşağıdaki şekilde de medyanı oluşturan değerler bulunabilir:

$$X_{\text{Enküçük}} = X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} = X_{\text{Enbüyük}}$$

$$\text{MEDYAN(Ortanca)} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{Eğer } n \text{ tek ise} \\ \frac{1}{2} \left[x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)} \right] & \text{Eğer } n \text{ çift ise} \end{cases}$$

- ★ N çift ise $X_{(n/2)}$ nci değer ile $X_{(n+2)/2}$ nci değer ortalaması alınarak bulunabilir:

SINIFLANMIŞ VERİLERDEN ORTANCA HESABI

Sınıf sınırları	Sıklık (frekans= f_i)den Az (F_i)
1.45 - 1.94	2	2
1.95 - 2,44	18	20
2,45 – 2,94	24	44
2,95 – 3,44	19	63
3,45 – 3,94	18	81
3,95 – 4,44	9	90
4,45 – 4,94	6	96
4,95 – 5,44	4	100
17.12.2019	100 BİYOİSTATİSTİK VE TIP BİLİŞİMİ A.D.	13

Sınıf Sınırı	2.95	3.45
Fi	45..... 50	51 63
	

Medyan Sınıfı
Başlangıcı

Medyan Sınıfı
Sonu

$$\text{Ortanca} = L + \frac{\frac{N}{2} - F_{(i-1)}}{f_i} \cdot C$$

Bu formül aslında basit orantıdan bulunabilir.



$$\frac{A}{a} = \frac{B}{b} \Leftrightarrow \frac{c}{\text{Med} - L} = \frac{f_i}{\frac{N}{2} - F_{(i-1)}}$$

$$\text{Or } \tan ca = L + \frac{\frac{N}{2} - F_{(i-1)}}{f_i} \cdot C$$

Burada:

L:Medyan sınıfının alt sınır değeri

N= Toplam gözlem sayısı

$F_{(i-1)}$ = Medyan sınıfından önceki sınıfların
frekans toplamı

f_i = Medyan sınıfının kendi sınıf frekansı

C= sınıf genişliği

Örnek için: $L= 2.95$, $N=100$, $F_{(i-1)}=44$, $f_i=19$ olur.

Ortanca= $2.95 + [(100/2) - 44] * 0.5 / 19 = 3.108$

Bulunur.

MEDYANIN VARYANSI

$$\text{Varyans(Medyan)} = 1.57 \cdot \frac{\text{Var}(X)}{n} = \frac{1.57 \cdot S^2_x}{n}$$

Medyanın varyansı ortalamanınkinden daha büyüktür.

Dolayısı ile medyanın güvenilirliği daha azdır.

Bu nedenle mecbur kalmadıkça aritmetik ortalama yerine medyan kullanılmaz.

$$N=50, f_i=17, F_{(i-1)}=24, C=5$$

Sınıf Sınırları	Frekans(f_i)	Eklemeli Frekans (F_i)
29.5-33.4	3	3
33.5-37.4	7	10
37.5-41.4	14	24
41.5-45.4	17	41
45.5-49.4	7	48
49.5-53.4	2	50
Ortanca= $41.5 + (25 - 24) * 4/17 = 41.74$		
	50	

Medyan sınıfı

ÇEYREKLİK VEYA YÜZDELİK DEĞERLER

Bir veri setinin p 'nci yüzdeliği bulunurken,

- $n \cdot p / 100$ sayısı tam sayı değilse büyüklük sırasına dizilmiş sette $(m+1)$ nci gözlem ilk yüzdelik olur. m ise $n \cdot p / 100$ den hemen üstündeki tamsayıdır.
- Eğer $n \cdot p / 100$ tam sayı ise $[n \cdot p / 100]$ ve $[(n \cdot p / 100) + 1]$ nci en büyük gözlemlerin ortalaması ilk yüzdelik olur. $n=8$ ise $p=0.25$ ise $8 \cdot 0.25 / 100 = 2$, 2 nci ve 3ncü gözlemlerin ortalaması ($Q_{0.25} = 5.5$) ilk %25 lik olur. $8 \cdot 0.50 / 100 = 4$ ncü ve 5 nci gözlemlerin ortalaması ($Q_{0.5} = 19$) %50 lik gözlem olur, $8 \cdot 0.75 / 100 = 6$ nci ve 7 nci gözlemler ortalaması ($Q_{0.75} = 43$) %75 lik gözlem olur.
- 2, 4, 7, 12, 26, 32, 54, 56

- Sıralanmış veriyi %25, %50 ve %75 lik parçalara ayıran değerlere 1., 2. ve 3. çeyreklik (kartil) adı verilir. Bunlar Q1, Q2 ve Q3 diye sembolize edilir. Veriyi tam ortadan bölen değer Q2 özel bir isimle isimlendirilir, buna ortanca (medyan) denir.
- Çeyreklikler arası uzaklık $(IQR)=(Q3-Q1)$ olur. IQR nin yarısı da yarı mesafe $(SIQR=[Q3-Q1]/2)$ olarak adlandırılır.

ÖRNEK:

Veri {0, 1, 2, 3, 4, 5, 6}

$$X_i = \{X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}, X_{(5)}, X_{(6)}, X_{(7)}\}$$

$$\text{Ortalama} = (0+1+2+3+4+5+6)/7 = 21/7 = 3$$

$$\text{min} = 0, \quad \text{max} = 6$$

$$Q1 = X_{(2)} = 1 = \text{Birinci Çeyreklik}$$

$$Q2 = X_{(4)} = 3 = \text{Ortanca (ikinci çeyreklik)}$$

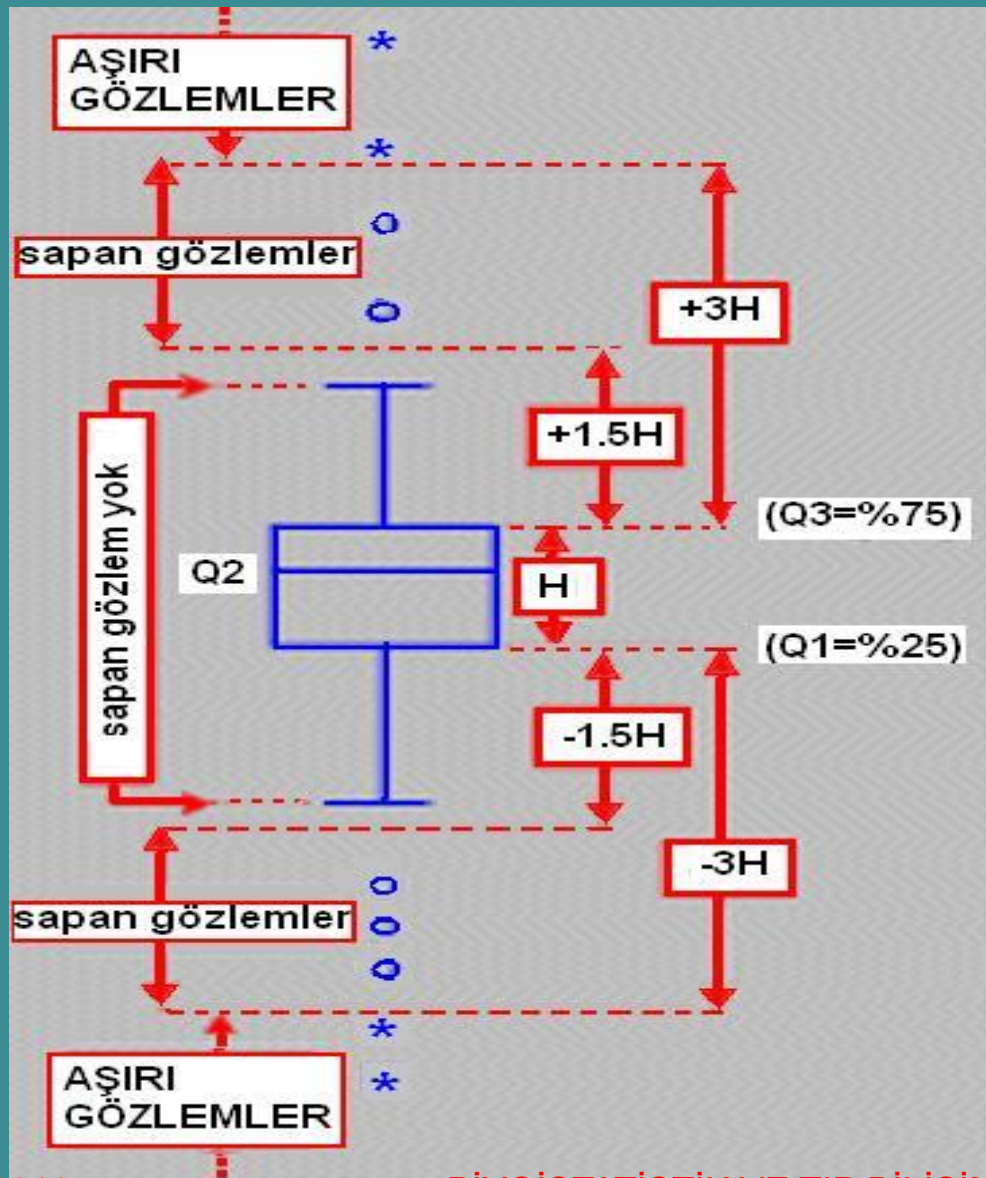
$$Q3 = X_{(6)} = 5 = \text{Üçüncü çeyreklik}$$

$$\text{Range} = \text{max} - \text{min} = 6 - 0 = 6$$

$$\text{IQR} = Q3 - Q1 = 5 - 1 = 4$$

$$s^2 = [(0^2+1^2+2^2+3^2+4^2+5^2+6^2) - 7(3^2)]/(7-1) = [91-63]/6 = 4.67$$

$$s = \text{sqrt}(4.67) = 2.16$$



Kutu grafikte bazı önemli noktalar.

TEPE DEĞER(MOD)

Bir sayı kümesi içinde en fazla tekrarlanan değer o kümenin tepe değerini oluşturur.

● $X_i: \{ 1,2, 6, 3, 7, 3,5, 6,6, 8,9 \}$

Burada en fazla tekrarlanan değer 6 olduğu için **Mod= 6** olur.

● Sınıflandırılmış Verilerde Mod Hesabı İçin:

$$\text{Tepe Değer} = L + \frac{d_1}{d_1 + d_2} \cdot C$$

Örnek: Bir taramada 50 kadının kanındaki (gr/lt) serum albümin değerleri aşağıdaki gibi bulunmuştur.

41 41 42 44 44 36 38 41 42 44
42 39 49 40 45 32 34 43 37 39
41 39 48 42 43 33 43 35 32 34
39 35 43 44 47 40 39 42 41 46
37 49 41 39 43 42 47 48 51 52

Bu verilere ait sıklık tablosu 6 sınıf olacak şekilde yapılsın.

EnK.=32, EnB.= 52, sınıf aralığı= 5 olsun.

- Önce mod sınıfı bulunur, bunun için en yüksek frekanslı sınıf seçilir. Serum albumin düzeyi Örneğindeki değerlerle formül sembollerini açıklarsak:

- En yüksek Frekans=17 olduğundan, 41.5-45.5 sınıfı mod sınıfıdır.

- Bu sınıfın alt sınır değeri $L=41.5$ dur.

- $d1 = (\text{medyan sınıfı frekansı} - \text{bir önceki sınıf frekansı})$

- $d2 = (\text{medyan sınıfı frekansı} - \text{bir sonraki sınıf frekansı})$

- $C = \text{sınıf aralığı} = 5$

Örnek: $N=50$, $\text{Max}(f_m)=17$, $C=5$, $d_1=(17-14)=3$,
 $d_2=(17-7)=10$, $L=41.5$

29.5-33.4	3	
33.5-37.4	7	
37.5-41.4	14	
41.5-45.4	17	Mod Sınıfı
45.5-49.4	7	
49.5-53.4	2	
Toplam	50	

En yüksek frekans

$$\text{Mod} = 41.5 + (17 - 14) * 5 / [(17 - 14) + (17 - 7)] = 42.42$$

- Tepe değerinin (mod) kullanışlı olabilmesi için gözlem sayısının çok fazla olması gerekir.

- Bazı durumlarda dağılışın birden fazla modlu olabilir, yani dağılış çok modlu olabilir.

Modların aynı yükseklikte (frekansta) olması da gerekmez. Ancak bu modların sınıf aralarının küçük değişikliği ile modlar kaybolmayacak ayrıklıkta olması gerekir.

Bu durumlarda örneklemin farklı grupların bir araya gelmesinden oluştuğu anlamı çıkar.

TARTILI ORTALAMA

- Bir ildeki 5 hastanede acile gelen hastaların ortalama yaşları aşağıdaki şekilde bulunmuştur.

Hastane	Hasta Sayısı (t_i)	Ortalama Yaş(X_i)
1	10	25
2	15	30
3	20	40
4	5	20
5	30	15

Bu ildeki hastanelerin acili servislerine gelen hastaların ortalama yaşı nedir ? Sorusunun cevabını bulmak için tartılı ortalama bulunur.

t_i = i nci grubun tartısı (i nci hastanedeki hasta sayısı)

X_i = i nci grubun değeri (i nci hastanedekilerin ortalama yaşı)

$$\text{Tartılı Ortalama} = \frac{\sum_{i=1}^k t_i \cdot X_i}{\sum_{i=1}^k t_i}$$

t_i	X_i
10	25
15	30
20	40
5	20
30	15

$$\sum t_i \cdot X_i = 2050$$

$$\sum t_i = 80$$

Bu ilde acile gelen hastaların ortalama yaşı hesaplanırsa:

$$\text{Tartılı Ortalama} = 2050/80 = 25,625$$

KESİKLİ VERİLERİN FREKANS TABLOSUNDAN HESAPLAMALAR

Sayılararak elde edilen verilerin çoğunda sıklık tablosu hazırlanırken sınıf değerleri doğrudan alındığı için yeniden bu tip tablolarda sınıf değeri hesaplamak gerekmez.

Örnek: Hastalık nedeniyle işe gelmeyen işçilerin gelmedikleri gün sayısını gösteren frekans tablosu aşağıdaki şekilde olsun.

Sınıflar	Gün (Xi)	İşçi Sayısı (fi)	fi*Xi	Fi
1	0	5	0	5
2	1	8	8	13
3 Mod sınıfı	2	10	20	23
4 Medyan sınıfı	3	9	27	32
5	4	6	24	38
6	5	5	25	43
7	6	4	24	47
8	7	2	14	49
9	8	1	8	50
Toplam		50	150	

- Aritmetik Ortalama = $150/50 = 3$

- Ortanca = 3, doğrudan tablodan okunur.

Ortanca sınıfının değeri (X_i) doğrudan ortanca olarak alınır.

- En yüksek sıklığa sahip sınıf mod sınıfı olduğundan bu sınıfa ait değer doğrudan mod değeri olarak alınır.

Mod = 2

ARİTMETİK ORTALAMA, ORTANCA, TEPE DEĞERİN KULLANIMI

Ortanca aritmetik ortalama kadar aşırı düşük veya aşırı yüksek değerlerden etkilenmez. Herhangi bir hastalık geçirmiş kişinin kaç yıl yaşayacağını hesaplamak için ortanca kullanılır.

$$X_i: \{4, 6, 7, 6, 6, 4, \underline{30}\}$$

Art. Ortalama $= 63/7 = 9$, son değer (30) Aritmetik Ortalamayı etkilemiştir ve yukarı çekmiştir.

$$\text{Ortanca ise } \{4, 4, 6, 6, 6, 7, 30\}$$

$$\text{Ortanca} = 6,$$

bu sayı sayı kümesini daha iyi temsil ediyor.

- Sıklık (frekans) tabloların alt ve üst sınırları açık uçlu ise bu sınıfların sınıf değerleri hesaplanamayacağı için Aritmetik ortalama hesaplanamaz ama ortanca hesaplanabilir.
- Açık Uçlu tablo-alt ve üst uçlar açık. (X_i) ler bu sınıflar için hesaplanamaz.

Sınıf Sınırları	Frekans (fi)
+33.5	3
33.5-37.5	7
37.5-41.5	14
41.5-45.5	17
45.5-49.5	7
49.5+	2

- Sıralamalı ölçümlü özelliklerde merkezi eğilim ölçüsü olarak ortancanın kullanılması uygundur.
- Bütün değerlerin elde edilmesinin uzun zaman aldığı bazı durumlarda ortanca kullanımı uygundur.

Örneğin öğrenme davranışının incelendiği bir araştırmada bazı bireyler çok geç öğrenebilir, ortalama için bunu beklemek gerekir, ortanca için bunu beklemeye gerek kalmaz.

ORTANCA KULLANMANIN SAKINICALARI NEDİR?

- Ortancanın standart hatası aritmetik ortalamadan daha büyüktür.
- Ortanca üzerinde cebirsal işlemler yapılamaz. Farklı alt grupların ortancaları biliniyorsa bu gruplar birleştğinde elde edilen birleşik grubun ortancası nedir sorusu hesaplama ile bulunamaz.

TEPE DEĞER KULLANIMI FAYDA VE SAKINICALARI

Faydaları:

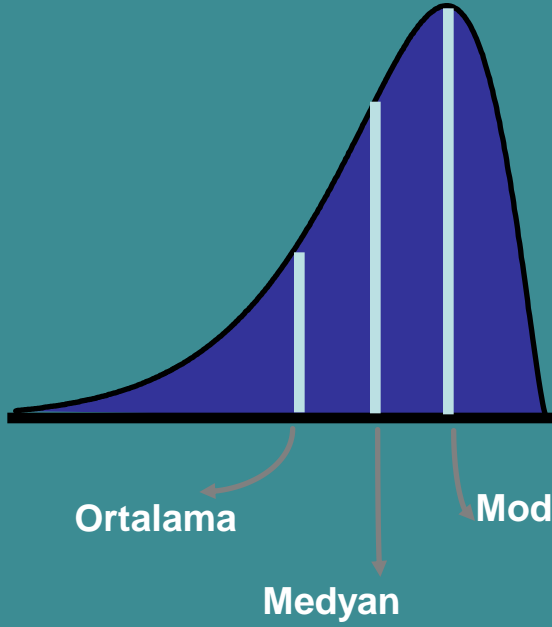
- Aşırı uç değerlerde etkilenmez.
- Örnekte alt gruplar olup olmadığını görmek için frekans eğrisinin modlarının incelenmesi gerekir.
- Ortalamayı hesaplanmasının zor olduğu durumlarda mod kullanılabilir.
- **Adlandırma** (nominal) ölçekli değişkenlerde mod kullanımı uygundur.

Sakıncalarına gelince:

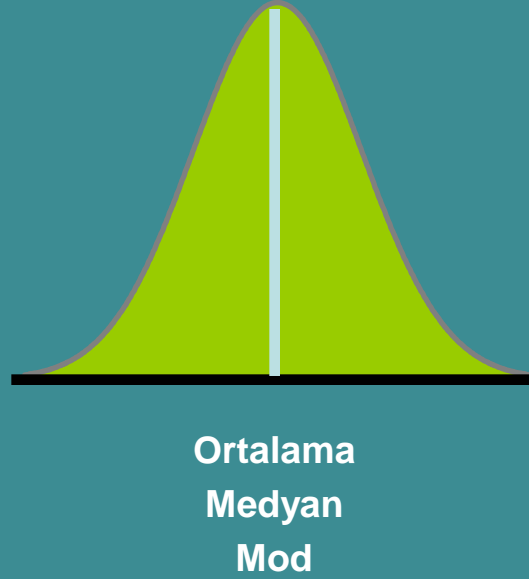
- Modun güvenirliliği azdır. Yani örnekten elde edilen mod popülasyon modundan çok farklı olabilir.
- Ortancada olduğu gibi mod üzerinde de cebirsel işlemler yapılamaz.
- Bazen verilerin ortalaması, ortancası olduğu halde modu olmayabilir. Bütün değerler farklı ise mod yoktur.

Çarpık dağılışlarda genelde: Ortalama, ortanca ve mod arasında aşağıdaki genel ilişki vardır.

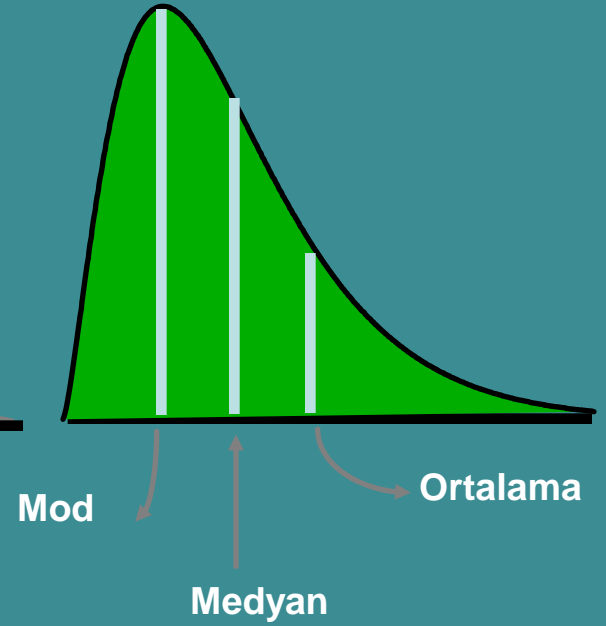
$$(Ortalama - Mod) = 3 * (Ortalama - Ortanca)$$



Negatif Çarpık



Simetrik



Pozitif Çarpık

Pearson'ın en basit çarpıklık katsayısı formülü:

$$\text{Çarpıklık} = \frac{(\text{Ortalama-medyan})}{\text{Std. Sapma}}$$

Çarpıklık $> +0.2$ veya
Çarpıklık < -0.2 ise

Çarpıklık ciddiye alınmalıdır.

GEOMETRİ ORTALAMA

● X_i : $\{X_1, X_2, \dots, X_n\}$ değerlerinin geometrik ortalaması:

$$\text{Geometrik Ortalama} = \text{GO} = \sqrt[n]{X_1 \cdot X_2 \cdot \dots \cdot X_n}$$

Şeklinde hesaplanır.

● Örnek: X_i : $\{8, 12, 25, 6\}$

$$\text{Geometrik Ortalama} = \text{GO} = \sqrt[4]{8 \cdot 12 \cdot 25 \cdot 6}$$

$$= 10.95$$

- Geometrik ortalama aşırı değerlerden aritmetik ortalamadan daha az etkilenir.
- $AO \geq GO$ ilişkisi vardır. Bütün X_i ler eşitse $AO=GO$ olur.
- Bazı değerler sıfır veya negatifse GO hesaplanamaz.
- Çoğu kişi GO'nın ne anlama geldiğini kavrayamaz, dolayısıyla kullanımı azdır.
- Nüfus çoğalması, bakteri üremesi gibi geometrik dizilerde birim zamandaki artışı bulmak için GO kullanılır.

- Başlangıçta A kadar birey varsa, bu bireyler birim zamanda r kadar bir hızla artıyorsa, n birim zaman sonra sayıları B kadar olmuş ise
- $B = A \cdot (1+r)^n$ olur.
- Ortalama artış (r) buradan hesaplanır.

$$1 + r = \sqrt[n]{\frac{B}{A}}, \quad \text{buradan}$$
$$r \text{ çekilirse}$$
$$r = \sqrt[n]{\frac{B}{A}} - 1 \quad \text{olur.}$$

- Bir bakteri kültürü 3 günde 1000 den 4000 e çıkmış ise ortalama günlük artış hızı(r)

- Burada r hesaplanırsa,

$$r = \sqrt[3]{\frac{4000}{1000}} - 1 = \sqrt[3]{4} - 1 = 0.587$$

- Yani ortalama artış hızı= $r = \%58.7$ dir.
- Örnek: Bir diyet uygulamasında deney hayvanının 4 günlük ağırlık artışı aşağıdaki gibidir. Buna göre ortalama günlük ağırlık artışı nedir?
- $X_i: \{ 70, 90, 120, 180 \}$
- Geometrik ortalamayı hesaplamadan önce 3 günlük artış oranları bulunursa:

İlk gün $(90)/70=1.29$, ikinci gün $120/90=1.33$,
Üçüncü gün $180/120=1.50$ olur. Bunların
 ortalaması (GO)

$GO = (1.29 \cdot 1.33 \cdot 1.50)^{(1/3)} = 2,57355)^{(1/3)} =$
 1.366 olur, yani Ortalama günlük artış %36.6
 olmuştur. Logaritmaların ortalaması GO olur.

Örnek: (ister e ister 10 tabanına göre logaritma
 alınabilir.)

X	Ln(X)
1	0.000000
2	0.693147
3	1.098612
10	2.302585
Geometric mean = 2.78	Arithmetic mean= 1.024 $EXP[1.024] =$ 2.78

HARMONİK ORTALAMA

- $X_i: \{ X_1, X_2, \dots, X_n \}$ değerlerinin harmonik ortalaması:

$$\text{Harmonik Ortalama} = \frac{1}{\frac{1}{n} \cdot \sum_{i=1}^n \frac{1}{X_i}}$$

Örnek: $X_i: \{ 6, 8, 3, 5, 4 \}$ veri setinin harmonik ortalaması:

$$\text{Harmonik Ortalama} = \frac{1}{\frac{1}{5} \cdot \left(\frac{1}{6} + \frac{1}{8} + \frac{1}{3} + \frac{1}{5} + \frac{1}{4} \right)} = 4.65$$

HARMONİK ORTALAMA

Örnek 1: 6 öğrenci 100 TL ile farklı eczanelerden aspirin alıyorlar. Birinci öğrenci 9 adet, ikincisi 6 adet, üçüncüsü 7 adet, dördüncüsü 8 adet, beşinci 6 adet ve altıncısı 8 adet aspirin alıyor. 100 TL ile alınabilecek ortalama aspirin sayısı ne kadardır?

- Fiyat=para/mal olduğundan ve para sabit ise harmonik ortalama alınır.

$$HO = \frac{6}{\frac{1}{9} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{6} + \frac{1}{8}}$$

- = 7.166, Yani 100 TL ile ortalama 7.166 aspirin alınabilir.

- AO, GO ve HO arasında şöyle bir ilişki vardır:
- $AO \geq GO \geq HO$, yani en küçük Harmonik ortalamadır.

Örnek:

Bir sağlık ocağının serum alımı için her yıl sabit 6 milyon liralık bütçe konuyor. 4 yıllık harcama döneminde 1. yıl 1 şişe serumun fiyatı 500 bin, ikinci yıl 1200 lira, 3. Yıl 1500 lira ve 4. Yıl için 2000 liradır. 1 şişe serum için ödenen yıllık ortalama fiyat nedir?

- Her yıl ilaç fiyatı sabit değil toplam harcanan miktar sabittir bu nedenle HO uygundur.

● Fiat= Para/mal olduğundan ödenen yıllık para sabit mal değişiyorsa HO alınır.

$$HO = 4 / (1/500 + 1/1200 + 1/1500 + 1/2000)$$

= $4 / 0,004 = 1000$ TL dir. Yani 1 şişe serum için yıllık ortalama 1000 TL harcama yapılmıştır.

● Bu basit olarak şöyle de hesaplanabilirdi.

1. Yıl $6000 * 10^3 / 500 = 12 * 10^3$ şişe,

2. Yıl $6000 * 10^3 / 1200 = 5 * 10^3$

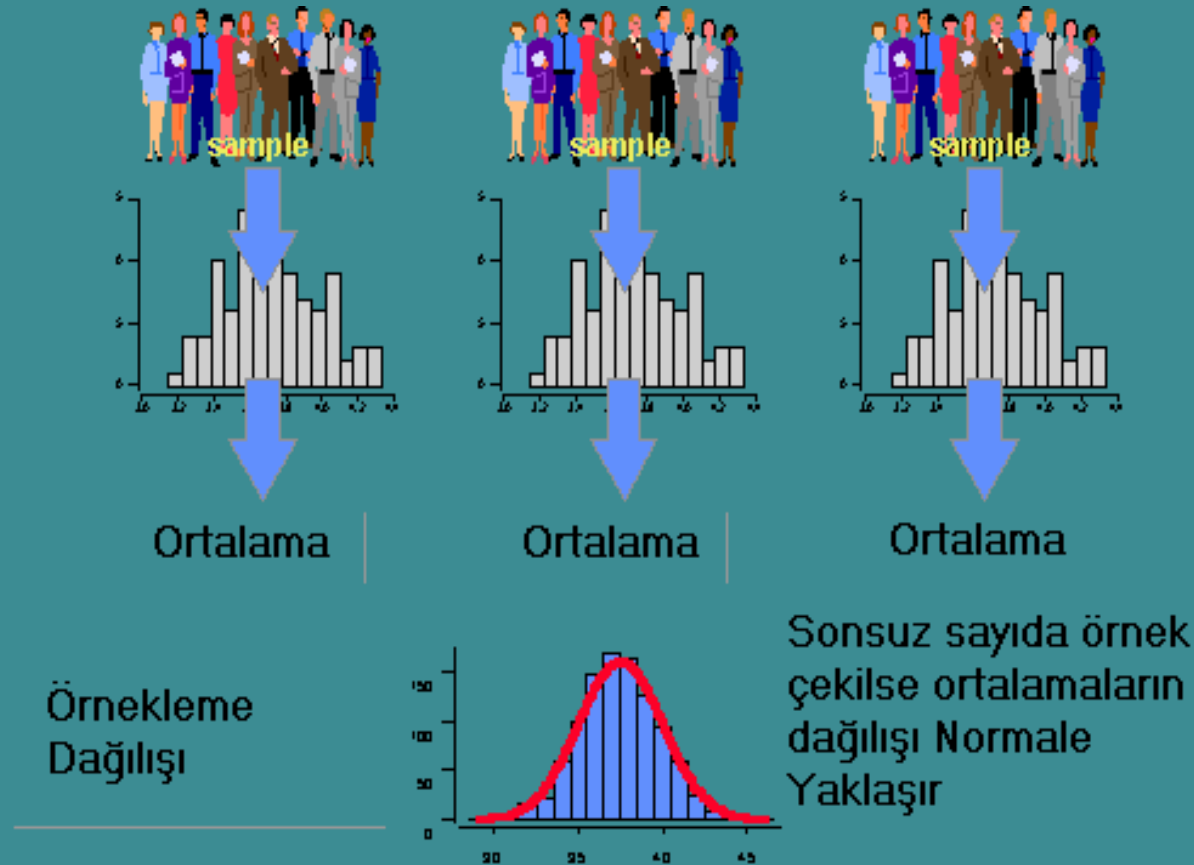
3. Yıl $6000 * 10^3 / 1500 = 4 * 10^3$

4. Yıl $6000 * 10^3 / 2000 = 3 * 10^3$

Toplam $(12+5+4+3) * 10^3 = 24 * 10^3$ şişe serum kullanılmış ve 24 milyon para harcanmıştır. Buna göre $24 * 10^3 / 24 = 1000$ TL/şişe bulunur.

ORTALAMALARIN ORTALAMASI

Ortalamaların ortalamasının örneklem dağılışı çok önemli özelliklere sahiptir.



TANIMLAYICI İSTATİSTİKLER DEĞİŞİM VEYA DAĞILIŞ ÖLÇÜLERİ

DEĞİŞKENLİK NEDİR?

● Bir ana kütleyi (popülasyonu) tanıtmak için, veya başka ana kütlelerle karşılaştırabilmek için merkezi ölçülerin yanında dağılışın genişliğini, değişkenliğin büyüklüğünü gösteren bir başka tipik değerin verilmesi gerekmektedir.

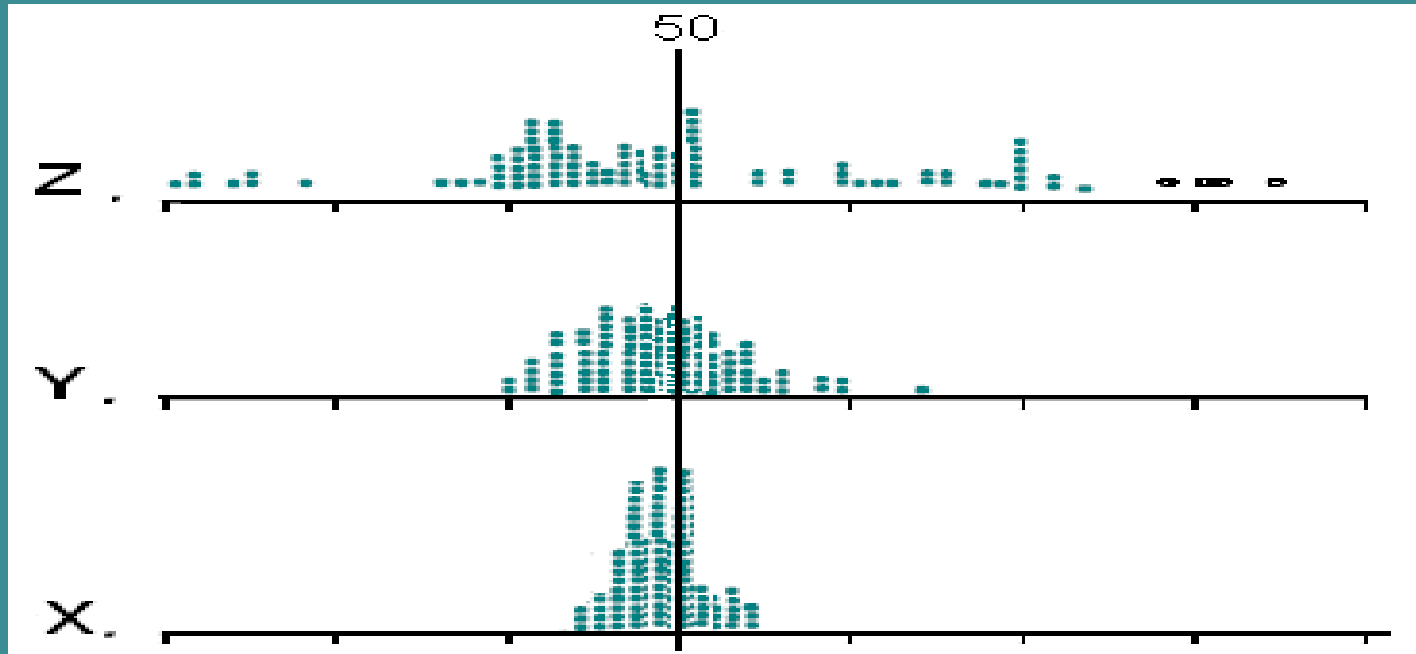
Bu tipik değer değişkenlik ölçüsüdür.

Örnek:

- ▶ $X_i: \{ 49, 49, 49, 50, 51, 52 \} ; \bar{X} = AO(X) = 50$
- ▶ $Y_i: \{ 35, 41, 50, 55, 58, 61 \} ; \bar{Y} = AO(Y) = 50$
- ▶ $Z_i: \{ 15, 21, 33, 49, 90, 92 \} ; \bar{Z} = AO(Z) = 50$

- Bu üç değişkenin ortalaması aynı olduğu halde X, Y ve Z değişkenlerinin aldığı değerlerin en küçük ve en büyük değerlerine bakıldığında birbirlerinden çok farklıdır.
- Bir popülasyonun gözlem değerlerinde görülen bu değişimin bir istatistikle ifade edilmesi gerekir.

Dağılışı veya değişimleri ifade için kullanılan bu ölçümlere dağılışı veya değişim ölçüleri denir.



X değerleri 50 nin etrafında çok yakın kümелendiği halde, Y değerleri biraz daha dağınık, Z değerleri 50 den çok uzakta yer almakta, yani daha büyük değişim göstermektedir.

Dolayısıyla değişkenin ortalama değeri 50 dir demek değişkeni tanımlamak için yeterli olmuyor. Bunun yanında başka ölçütlerde vermek gerekmektedir. Bu ölçüt değişkenlik ölçütü olabilir.

BAŞLICA DEĞİŞİM ÖLÇÜTLERİ

- Değişim Aralığı veya Genişliği (Range)
- Çeyrek Ayrılış (IQR)
- Yarı Çeyrek Ayrılış (SIQR)
- Ortalama Mutlak Sapma
- Variyans
- Standart Sapma
- Standart hata
- Değişim (Varyasyon) katsayısı

DEĞİŞİM GENİŞLİĞİ(RANGE)

En basit dağılış ölçüsü Değişim Genişliğidir.

Değişim Genişliği=DG=R=

= (En Büyük Gözlem – En Küçük Gözlem)

$$\blacktriangleright DG(X) = [X_{(enbuy)} - X_{(en kuc)}] = 52 - 49 = 3$$

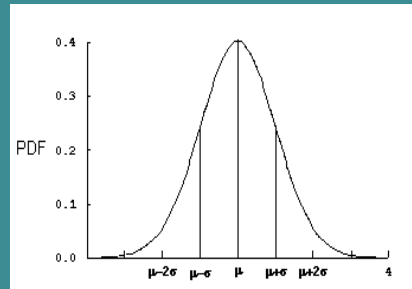
$$\blacktriangleright DG(Y) = [Y_{(enbuy)} - Y_{(en kuc)}] = 61 - 35 = 26$$

$$\blacktriangleright DG(Z) = [Z_{(enbuy)} - Z_{(en kuc)}] = 92 - 15 = 77$$

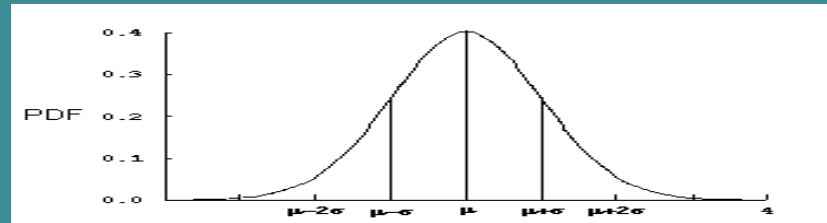
Olur.

Bu tip değişkenliğe sahip dağılışların grafiklerine bakıldığında aşağıdaki şekilde bir görünüş söz konusudur.

X



Y



Z



Değişim genişliği veya buna dayanan ölçütler (IQR, SIQR) aslında çok kullanılan ölçülerdir. Örneğin şeker hastasının en küçük ve en yüksek değerleri nedir, ateşli bir hastanın en düşük ve en yüksek ateş derecesi nedir, kişiler bu tip değerlerle sürekli ilgilenilir. Özellikle sıralı ölçekteki verilerin tanıtıcı istatistikleri verilirken bunlardan yararlanılır.

- Ancak değişim genişliğinde sadece iki aşırı uç değer kullanıldığı için istikrarlı bir ölçü değildir, bilgi kaybı çoktur, çünkü verinin büyük bir kısmı kullanılmamaktadır.

- Diğer bir sakıncası değişim genişliği örnek büyüklüğüne çok bağlıdır. Örnek büyüdükçe aşırı değerleri kapsama olasılığı artar.

ÇEYREK AYRILIŞ(IQR), YARI ÇEYREK AYRILIŞ (SIQR)

- Üçüncü ve birinci çeyrekliklerin farkı çeyrek ayrılıştır. $IQR = (Q_3 - Q_1)$

- Yarı Çeyrek Ayrılış ($SIQR = (Q_3 - Q_1)/2$)

- Bu istatistikler genelde sıralama ölçeği ile ölçülen özelliklerin tanımlayıcı istatistiklerinde değişkenliği tanımlarken kullanılır.

Sıralama ölçeğine sahip değişkenlerde merkezi bölgeyi tanımlarken medyan kullanılmasının doğru olacağı daha önce belirtilmişti. Değişkenlik ölçüsü olarak da R, IQR veya SIQR kullanılabilir.

ORTALAMA MUTLAK SAPMA

- Ortalama Mutlak Sapma veya Ayrılış

$$OA = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}|$$

- Ortalama mutlak sapma tüm veriyi kullandığı için değişim genişliğinden daha güçlü bir istatistiktir.

- Ancak bunun sakıncası da, üzerinde cebirsel işlemlerin yapılmasının zorluğudur. Yani iki grubun (n) leri ve ortalama mutlak sapmaları verilse, bunlar birleştirildiğinde birleşik örneğin ortalama mutlak sapmasını hesaplayın dendiğinde bu bulunamaz.

Bunların yerine verinin tamamını kullanan, üzerinde cebirsel işlemlerin daha kolay yapılabildiği bir değişkenlik ölçüsünün kullanılması daha faydalı olacaktır.

Bu ölçü variyans ve buna dayanan ölçülerdir.

VARIYANS

Variyans değişkenlerin aldığı değerlerin kendi ortalamasından ayrılışlarının karelerinin toplamının $(n-1)$ bölünmesi ile elde edilir. Eğer popülasyon ortalaması (μ) biliniyorsa n bölünerek te elde edilebilir. Ancak genelde μ bilinmez bu nedenle de yansız tahmin için $(n-1)$ e bölünür.

$X_i: \{X_1, X_2, \dots, X_n\}$ değerler kümesinin variyansı iki farklı şekilde hesaplanabilir:

$$\text{Variyans} = S^2 = \underbrace{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}}_{\text{ilk hesaplama şekli}} = \frac{1}{n-1} \underbrace{\left(\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right)}_{\text{ikinci hesaplama şekli}}$$

Eşitliğin sağ tarafındaki sonraki formül ile bir önceki tamamen aynıdır, biri diğerinin açılmış halidir.

Örnek:

$X_i: \{49, 49, 49, 50, 51, 52\}$

değişkeninin variyansı nedir?

İLK FORMÜLÜN KULLANILIŞI

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
49	-1,0	1,0
49	-1,0	1,0
49	-1,0	1,0
50	0,0	0,0
51	1,0	1,0
52	2,0	4,0

$$n = 6$$

$$\text{Toplam} = \sum X_i = 300 \quad \sum_{i=1}^6 (x_i - \bar{X})^2 = 8,0$$

$$\text{Ortalama} = \sum X_i / n = 50 \quad S^2 = 8/5 = 1,6$$

İKİNCİ FORMÜLÜN KULLANILIŞI

X_i	(X_i^2)
49	2401,0
49	2401,0
49	2401,0
50	2500,0
51	2601,0
52	2704,0

n= 6

Toplam= 300 15008,0

Variyans =

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

Variyans =

$$S^2 = \frac{1}{6-1} \left(15008 - \frac{300^2}{6} \right) = \frac{8}{5} = 1,6$$

Her iki formülden de aynı sonuç elde edilmektedir.

$$S^2 = [(49-50)^2 + (49-50)^2 + (49-50)^2 + (50-50)^2 + (51-50)^2 + (52-50)^2] / (6-1) = (1+1+1+0+1+4) / 5 = 8/5 = 1.6$$

Veya ikinci formül kullanılırsa;

$$S^2 = [(49^2 + 49^2 + 49^2 + 50^2 + 51^2 + 52^2) - 300^2 / 6] / 5 = (15008 - 15000) / 5 = 8/5 = 1.6$$

X'in variyansı olduğunu ifade için $S_x^2 = 1.6$ yazılabilir. Sonuç aynıdır.

Aynı şekilde Y nin ve Z nin variyansları da hesaplanabilir.

$$S_y^2 = 103.2, S_z^2 = 1144 \text{ bulunur.}$$

FREKANS(SIKLIK) TABLOSUNDAN VARIYANS HESAPLANMASI

Daha önce merkezi ölçülerin hesaplanmasında verilen örnek kullanılırsa, “yüz yeni doğan çocuğun doğum ağırlıkları (kg) ile ilgili sıklık tablosundaki gruplandırılmış olan veri setine” ait varyans nasıl hesaplanacaktır ?

ÇOCUK DOĞUM AĞIRLIĞI SIKLIK TABLOSU

Sınıf sınırları	Sınıf Değeri (X_i)	Sıklık (frekans= f_i)
1.45 - 1.94	1.7	2
1.95 - 2,44	2,2	18
2,45 – 2,94	2,7	24
2,95 – 3,44	3,2	19
3,45 – 3,94	3,7	18
3,95 – 4,44	4,2	9
4,45 – 4,94	4,7	6
4,95 – 5,44	5,2	4
		100

SINIFLANDIRILMIŞ VERİLERDE VARIYANS HESABI

Sınıf sınırları	(X_i)	(f_i)	$f_i \cdot X_i$	$f_i \cdot (X_i)^2$
1.45 - 1.95	1.7	2	3,4	5,78
1.95 - 2,45	2,2	18	39,6	87,12
2,45 – 2,95	2,7	24	64,8	174,96
2,95 – 3,45	3,2	19	60,8	194,56
3,45 – 3,95	3,7	18	66,6	246,42
3,95 – 4,45	4,2	9	37,8	158,76
4,45 – 4,95	4,7	6	28,2	132,54
4,95 – 5,45	5,2	4	20,8	108,16
	Σ	100	322	1108 , 3

$$\text{Variyans} = \frac{1}{\sum_{i=1}^k f_i - 1} \left(\sum_{i=1}^k f_i \cdot X_i^2 - \frac{(\sum_{i=1}^k f_i X_i)^2}{\sum_{i=1}^k f_i} \right)$$

$$S^2 = \frac{1}{100 - 1} \left(1108.3 - \frac{322^2}{100} \right) = 0.722$$

Variyanstan hesaplanan diğer değişkenlik ölçütleri **Standart Sapma**, **Standart Hata** ve **Varyasyon(Değişim) Katsayısı** çok sık kullanılan değişkenlik istatistikleri olarak karşımıza çıkar.

Bunların nasıl hesaplandığı üzerinde duralım.

STANDART SAPMA

Standart sapma varyansın kare köküdür. Varyans birimsiz ifade edildiği halde standart sapma ölçülen özelliğin birimi ile ifade edilir. Ortalama etrafındaki saçılma fazla ise varyans büyük olacak, dolayısıyla standart sapma da büyük çıkacaktır. Varyans hesaplanırken ortalamadan ayrılışların kareleri üzerinden hesap yapıldığı için aşırı değerlerin varyans üzerindeki etkisi ortalamaya yakın değerlerden daha fazla olur.

- Veri seti içinde aşırı değer fazla ise bunların varyansı etkilemesi olağan karşılanabilir, ancak birkaç aşırı değer varyansı büyütmesi olağan karşılanmaz.

- Aşırı değerlerin az olduğu durumlarda ortalama yerine ortanca, standart sapma yerine de çeyrek ayrılışın kullanılması uygun olur.

Ancak çeyrek sapma hesaplanırken tüm veriler yerine sadece dörtte birlik sınırlar (Q1 ve Q3) kullanıldığından standart sapmadan daha zayıf bir dağılış ölçüsüdür. Ortanca kullanmanın gerekli olduğu hallerde çeyrek sapmada değişim ölçüsü olarak kullanılır.

ÖRNEK:

$X_i: \{49, 49, 49, 50, 51, 52\}$ veri seti için varyans daha önce hesaplandı. Bunun standart sapması(S) nasıl bulunur? Varyansın karekökü standart sapmadır.

$$S = \sqrt{\text{Varyans}} = \sqrt{S^2}$$

$$\text{Standart sapma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right)}$$

Standart sapma **S** veya **S_x** şeklinde değişken ismi ile birlikte yazılabilir.

$$S = \sqrt{\text{Variyans}} = \sqrt{S^2} = \sqrt{1.6} = \mathbf{1.26},$$

Aynı şekilde Y ve Z değişkenlerinin standart sapmaları hesaplanırsa:

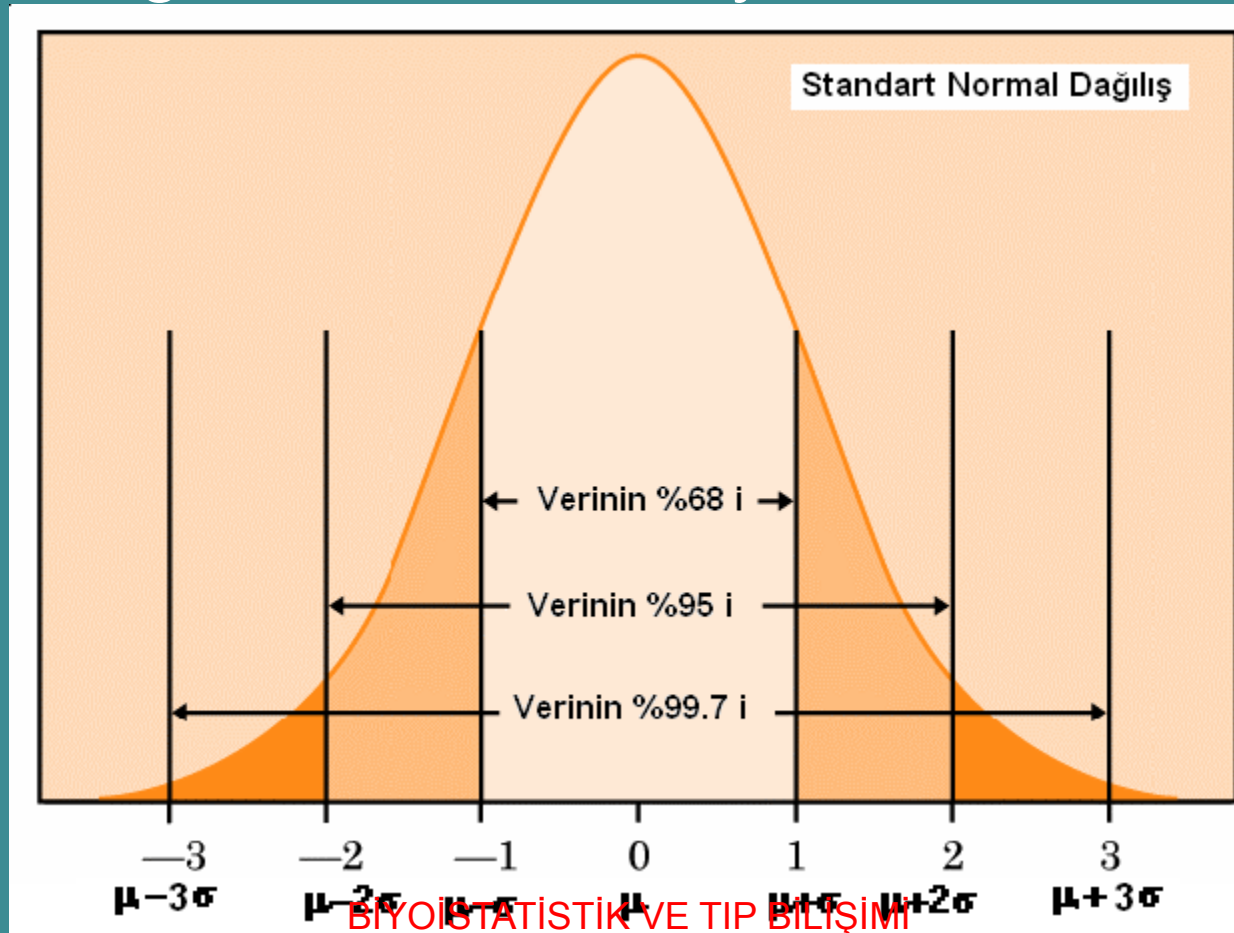
$$S_y = \sqrt{103.2} = \mathbf{10.16},$$
$$S_z = \sqrt{1144} = \mathbf{33.82} \text{ bulunur.}$$

$$\bar{X} \pm S_x \quad 50 \pm 1.26 \text{ ifadesi}$$

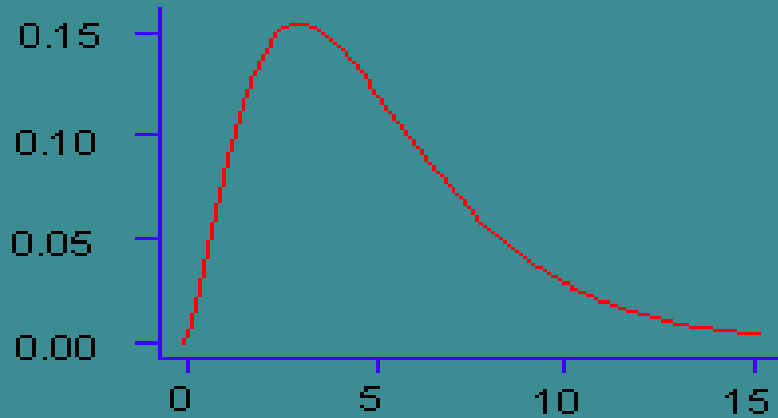
kullanılabilir

Standart sapma gözlemlerin değişkenliği ile ilgili bir göstergedir.

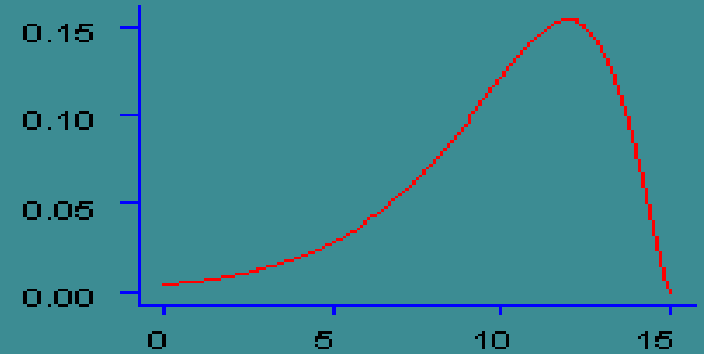
Normal dağılıшта gözlemlerin %68.27 si ortalamanın ± 1 standart sapma, gözlemlerin %95.45 i ± 2 standart sapma, %99.7 si ± 3 standart sapma sağında ve solunda yer alır.



ÇARPIK DAĞILIŞLAR



Pozitif(sağa) Çarpık



Negatif(sola) çarpık

Çarpık dağılışlarda bu tip bir ilişkiden bahsetmek söz konusu değildir. Bu nedenle çarpık dağılışlarda güven sınırından bahsetmek anlamsız olur.

STANDART HATA

$$\text{Standart Hata} = \sqrt{\frac{\text{Varyans}}{n}} \Rightarrow S_{\bar{x}} = \sqrt{\frac{S^2}{n}}$$

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{1.6}{6}}$$

$$= 0.52$$

Standart sapma ve standart hata Normal dağılış gösteren özellikler için oldukça kullanışlıdır.

GÜVEN SINIRI

Ortalamanın dağılışı:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ ise}$$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Şeklinde
standardize edilir.

Dağılışı simetrik ise ana kütle ortalamasına (μ) ait güven sınırı hesaplanabilir.

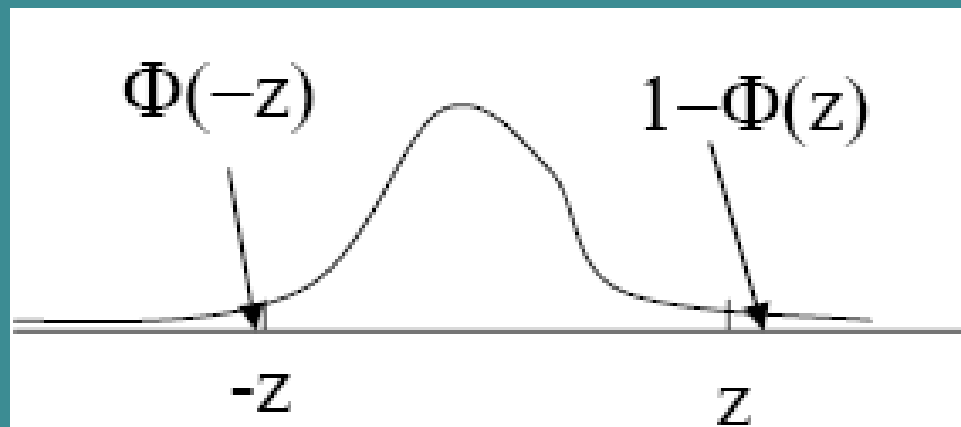
$$P\left[\bar{X} - Z_{(\alpha/2)} \cdot S_{\bar{X}} \leq \mu \leq \bar{X} + Z_{(\alpha/2)} \cdot S_{\bar{X}}\right] = \%(1 - \alpha)$$

Burada Z standart normal dağılışın (alfa düzeyindeki) tablo değeridir.

GÜVEN SINIRI

$$P\left[-z_{\alpha/2} \leq Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right] = 1 - \alpha$$

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



X DEĞİŞKENİ İÇİN %95 LİK GÜVEN SINIRLARI

$$\bar{X} = 50$$

$$S_{\bar{X}} = 0,52$$

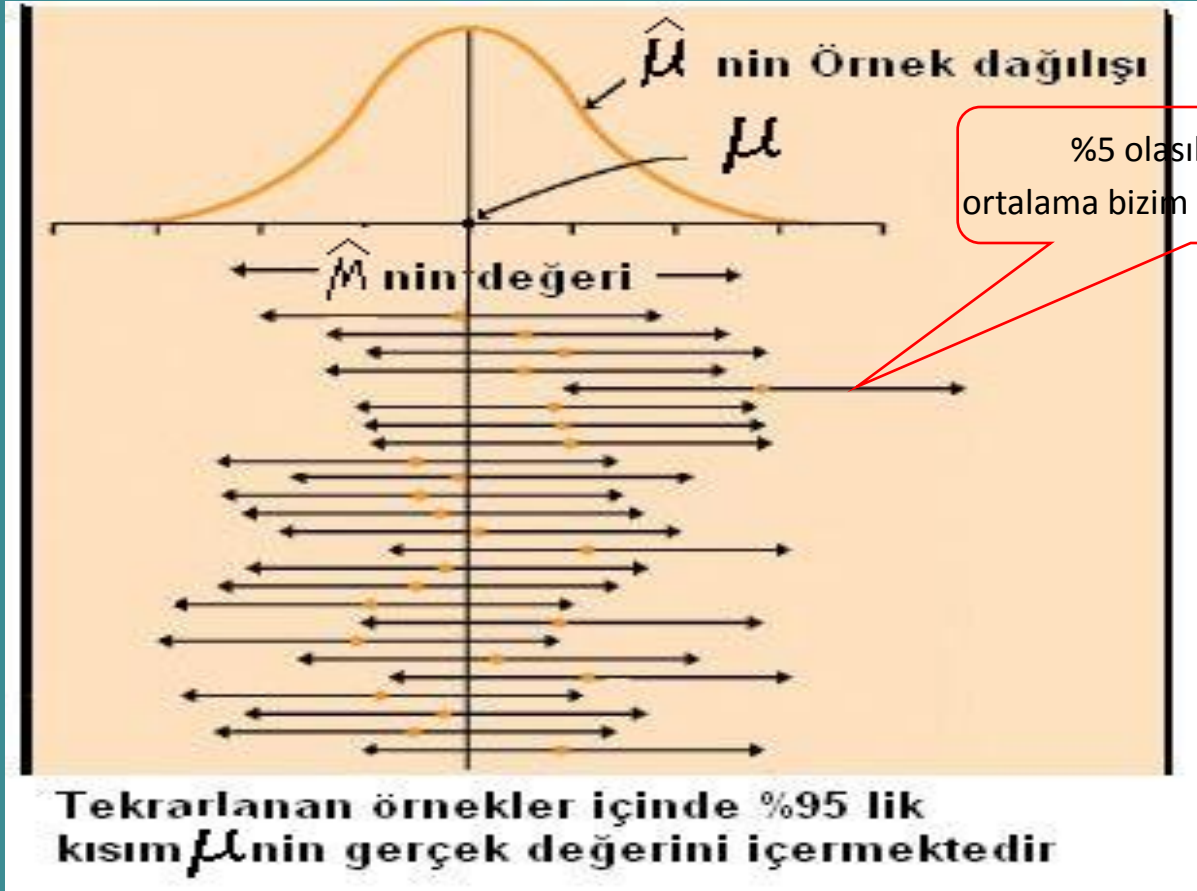
$$Z_{(0.025)} = 1,96$$

$$L = \bar{X} - Z_{(\alpha/2)} \cdot S_{\bar{X}} = 50 - 1,96 \times 0,52 = 48,98$$

$$U = \bar{X} + Z_{(\alpha/2)} \cdot S_{\bar{X}} = 50 + 1,96 \times 0,52 = 51,02$$

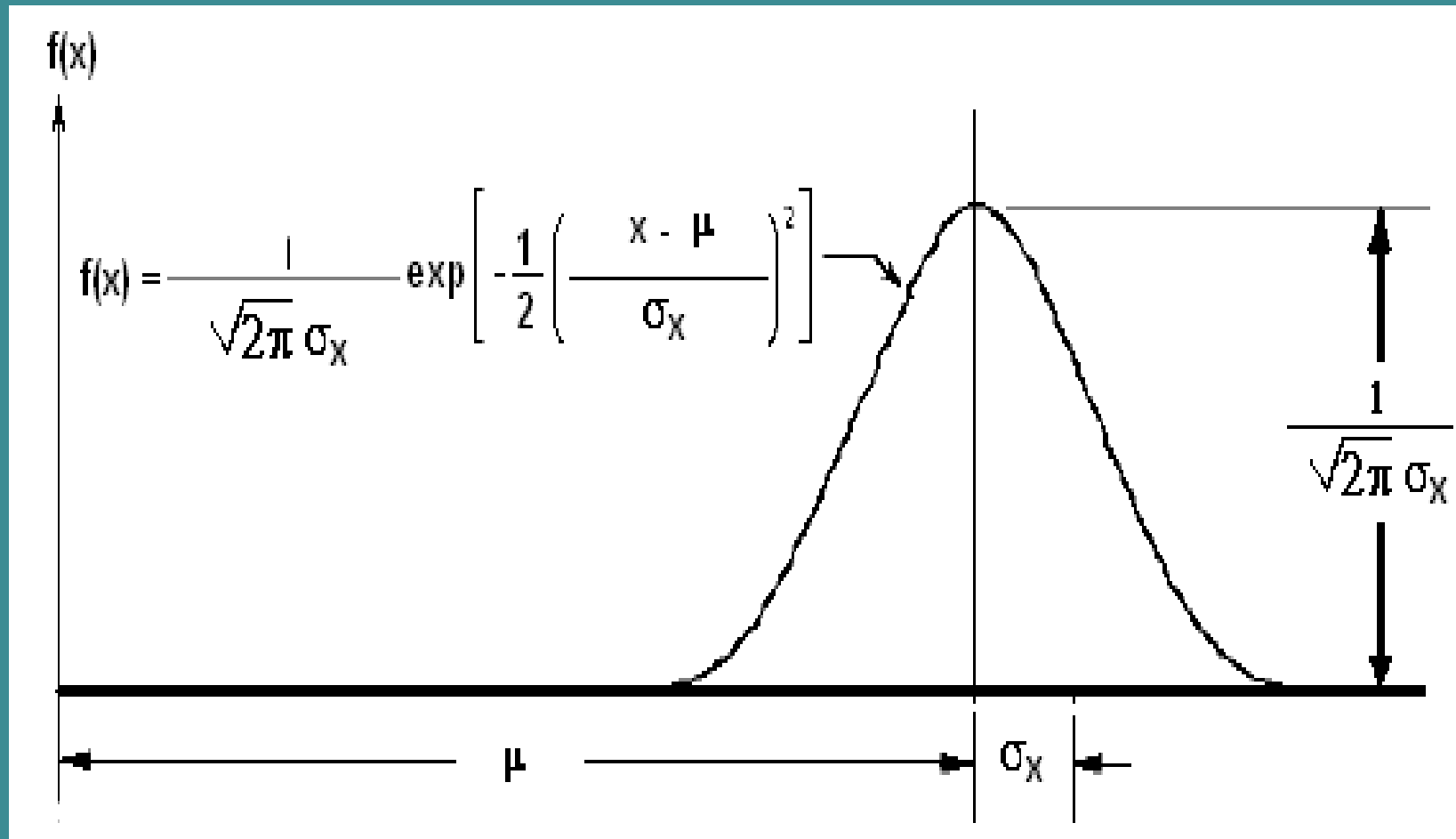
Alt güven sınırı=48,98, üst güven sınırı=51,02
Olur.

GÜVEN SINIRININ ANLAMI



Bulduğumuz güven sınırının gerçek ortalamayı (μ) kapsadığını %95 güvenle söyleyebiliriz.

Normal Dağılış Olasılık Yoğunluk Fonksiyonu



X, Y ve Z değişkenlerine ait İstatistikler

	N	DG	Ortalama	s^2 veya s_x^2 Variyans	s_x Standart Sapma	$s_{\bar{x}}$ Standart Hata
X	6	3	50	1,6	1,26	0,516
Y	6	26	50	103,2	10,16	4,147
Z	6	77	50	1144,0	33,82	13,808

Bir serinin ortalaması verilirken standart hatası ile birlikte verilir, şöyle ki:

Aritmetik Ortalama \pm Standart Hata

$$\bar{X} \pm S_{\bar{X}}$$

X_i serisi için bunun ifadesi: 50 ± 0.52 .

Y_i serisi için 50 ± 4.15 ,

Z_i serisi için 50 ± 13.8

şeklindedir olur.

Böylece ortalamaya ait değişkenlik te yanında ifade edilmiş olur.

Standart hata ortalamanın güvenirliliği ile ilgili bir ölçüdür. Standart hata ne kadar küçükse ortalama o kadar güvenilir.

DEĞİŞİM VEYA VARYASYON KATSAYISI(VK)

Bir serinin standart sapması kendi aritmetik ortalamasının yüzdesi olarak ifade edilirse değişim katsayısı bulunmuş olur.

- $V.K = \text{Standart Sapma} * 100 / \text{Ortalama}$

$$\text{Varyasyon Katsayısı} = \frac{S}{\bar{X}} * 100$$

- Xi serisi için: $V.K = 1.26 * 100 / 50 = \%2.52$
- Yi için; $V.K = 10.16 * 100 / 50 = \%20.32$
- Zi için; $V.K = 33.82 * 100 / 50 = \%67.64$ olur.

VARYASYON KATSAYISININ (VK) KULLANIMI

- A ve B gibi iki farklı popülasyondaki değişkenliği karşılaştırmak istersek doğrudan standart sapma veya standart hatalarına bakmak yanıltıcı olabilir.

- Ana kütlelerin ortalamaları büyüklük olarak birbirinden çok farklı ise (fillerin ağırlığı daha değişkendir, farelerin ağırlığı ?) standart sapma değerlerini ölçü biriminden bağımsız hale getirmek gerekir. Bunun için Değişim(varyasyon) kullanılır.

Örneğin fillerin ağırlığı daha değişkendir, farelerin ağırlığı sorusunun cevabı için V.K. kullanılır.

- İki özellik farklı birimlerle ölçülüyorsa, örneğin üzerlerinde deneme yapılan farelerin kan şekerimi daha değişkendir yoksa vücut ağırlıkları mı sorusu ile karşılaşılsa bunun için varyasyon katsayılarına bakmak gerekir.
- Ortalamanın sıfıra yakın olduğu durumlarda varyasyon katsayısının güvenilirliği azalır. Bu gibi durumlarda dikkatli olmak gerekir.
- %20 nin üzerinde varyasyon katsayısı fazla değişkenlik olduğunun göstergesi olarak görülür.

Örnek:Aynı şahıs üzerinde iki farklı yöntemle (X:autoanalyser, Y:Microenzymatic) ölçülen kolesterol miktarları aşağıda verilmiştir.

●X: { 177,193,195,209, 226} %mg/ml

●Y: {192,197,200,202,209}

Her ikisinin aritmetik ortalaması eşittir.

$$AO(X)= 1000/5=200, AO(Y)=1000/5=200$$

$$V(X)=(201360-200000)/4=340, S=18.44$$

$$V(Y)=(200158-200000)/4=39.5, S=6.28$$

$$VK(X)=18.44*100/200=\%9.22$$

$$VK(Y)=6.28*100/200=\%3.14 \text{ olur.}$$

VARIYANSLA İLGİLİ BAZI ÖNEMLİ HESAPLAMALAR

Veri setini sabit bir sayı ile toplama ve çıkarmanın ortalama ve variyans üzerine etkisi:

C sabit bir sayı ise $Y_i = X_i + C$ şeklinde verilerin tamamına **C** sayısı eklense yeni elde edilen sayının ortalama ve variyansı değişir mi?

● $E(X_i) = \mu$, $E(Y_i) = E(C + X_i) = E(C) + E(X_i) = C + \mu$
ortalama C kadar artar.

● $V(X_i) = \sigma^2$ ise, $V(Y_i) = V(C + X_i) = V(C) + V(X_i) = 0 + \sigma^2$
varyans aynı kalır.

ÖRNEK:

- $X_i: \{4, 5, 6, 9\}$ ise

Ortalama(X)= $24/4=6$, variyans(X)=4.67

Tüm değerlere $c=3$ sabiti eklense

- $Y_i = \{7, 8, 9, 12\}$ elde edilir.

Y_i nin ortalaması ve variyansı ne olur?

Ortalama(Y)= $36/4=9$, variyans(Y)=4.67

Olur.

C sabit bir sayı ise verilerin tamamı **C** sayısı ile çarpılsa yeni elde edilen sayının ortalama ve varyansı değişir mi?

- **$Y_i = C \cdot X_i$** olursa,

- $E(X_i) = \mu$, $E(Y_i) = E(C \cdot X_i) = C \cdot E(X_i) = C\mu$,
ortalama C kat artar.

- $V(X_i) = \sigma^2$ ise, $V(Y_i) = V(C \cdot X_i) = C^2 \cdot V(X_i) = C^2 \cdot \sigma^2$,
varyans sabit sayının karesi ile çarpımı kadar büyür.

- $X_i: \{4, 5, 6, 9\}$ ise
- $\text{Ortalama}(X) = 24/4 = 6$, $\text{variyans}(X) = 4.67$
- Tüm değerlere $c=3$ sabiti ile çarpılsa
- $Y_i = \{12, 15, 18, 27\}$ elde edilir.
- Y_i nin ortalaması ve varyansı ne olur?
- $\text{Ortalama}(Y) = 72/4 = 18$, $\text{variyans}(Y) = 42,0$