DEĞİŞİM ÖLÇÜLERİ

Verilerin değişkenlik durumunu ve dağılış şeklini tespit etmek için kullanılan ölçülere dağılış(değişim) ölçüleri denir.

Bir anakütleyi tanıtmak, başka ana kütlelerle karşılaştırabilmek için merkezi eğilim ölçüleri yanında değişim ölçüsünün verilmesi gerekmektedir.

Bir serinin değişkenliği, gözlem değerlerinin birbirinden yada ortalamadan uzaklıkları ile ifade edilir. Dağılma veya değişkenlik ne kadar düşükse, ortalama o kadar temsil edicidir.

Ortalamaları eşit olan seriler, değişkenlikleri farklı olduğunda birbirine benzemez. Bu nedenle serileri tam olarak tanımlayabilmek için ortalamayı, değişimlerini ve bölünme şekillerini incelemek gerekir.

Örnek. Aşağıda verilenlere göre hangi serinin ortalaması daha temsil edicidir?

X_i : 49, 49, 49, 50, 51, 52 ; $\bar{x} = 50$

Y_i : 35, 41, 50, 55, 58, 61 ; $\bar{y} = 50$

Z_i : 15, 21, 33, 49, 90, 92 ; $\bar{z} = 50$

Üç serinin de ortalaması aynıdır. Değişkenliklerine bakmamız gerekir.

a) Değişim Genişliği (Ranj): En basit değişim ölçüsü olup

$$D.G. = (En\ \text{büyük}\ \text{gözlem}\ \text{değeri} - En\ \text{küçük}\ \text{gözlem}\ \text{değeri}) = X_{Max} - X_{Min}$$

İle belirlenir. Genellikle kalite kontrol çalışmalarında kullanılır.

X_i : 49, 49, 49, 50, 51, 52 ; $\bar{x} = 50$, $D.G.(X) = 52 - 49 = 3$

Y_i : 35, 41, 50, 55, 58, 61 ; $\bar{y} = 50$, $D.G.(Y) = 61 - 35 = 26$

Z_i : 15, 21, 33, 49, 90, 92 ; $\bar{z} = 50$, $D.G.(Z) = 92 - 15 = 77$

Değişim genişliklerine baktığımızda, en az değişim gösteren X_i serisinin ortalaması daha temsil edicidir.

b) Çeyrek Sapma (Kartiller arası sapma, Çeyrek ayrılış ölçüsü)

Değişim genişliğinin serinin iki ucunda yer alan aşırı uç değerlerden etkilenir. Aşırı uç değerlerden etkilenmeyi ortadan kaldırmak *çeyrek sapma* kullanılır.

Çeyrek sapma hesaplanırken, serinin her iki ucundaki %25'lik gözlem değerleri dikkate alınmaz. Çeyrek sapma Q ile gösterilir ve aşağıdaki formülle hesaplanır.

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

Örnek. 1, 3, 4, 6, 10, 12 serisinin çeyrek sapmasını bulunuz?

$Q_1 = 3$, $Q_2 = 5$, $Q_3 = 10$ olup

$$Q = \frac{10 - 3}{2} = 3.5$$

elde edilir.

Örnek. Yabancı dil sınavına hazırlanan öğrencilerden 27 tanesinin bir günde çözdükleri soru sayısı aşağıda verilmiştir. Çeyrek sapmayı(kartiller arası sapmayı) bulunuz.

28, 19 , 23, 27, 51, 38, 30, 33, 47, 35, 37, 20, 37, 42, 39, 43, 43, 47, 46, 43, 45, 17, 45, 48, 48, 44, 55

Cevap= 8

c) Ortalama Sapma

Verilerin aritmetik ortalamadan sapmalarının mutlak değerlerinin toplamının, toplam gözlem sayısına bölünmesiyle elde edilir. Seri türlerine göre ortalama sapma formülleri aşağıdaki gibidir:

Ortalama Sapma	**Basit Serilerde**	$\mathrm{O.S.} = \frac{\sum \lvert X_i - \bar{x} \rvert}{n}$
	Frekans Serilerinde	$\mathrm{O.S.} = \frac{\sum f_i \lvert X_i - \bar{x} \rvert}{n}$
	Gruplanmış Serilerde	$\mathrm{O.S.} = \frac{\sum f_i \lvert m_i - \bar{x} \rvert}{n}$

Örnek. Aşağıdaki basit serinin ortalama sapmasını bulunuz?

X_i : 1, 4, 5, 5, 7, 14

$\bar{x} = 6$ olup,

$$\mathrm{O.S.} = \frac{\sum |X_i - \bar{x}|}{n} = \frac{|1-6| + |4-6| + |5-6| + |5-6| + |7-6| + |14-6|}{6} = \frac{18}{6} = 3$$

bulunur.

Örnek. Aşağıdaki frekans serisinin ortalama sapmasını bulunuz?

$\mathbf{X_i}$	$\mathbf{f_i}$	$f_i X_i$	$\lvert X_i - \bar{x} \rvert$	$f_i \lvert X_i - \bar{x} \rvert$
1	**2**	2	5.2	10.4
5	**7**	35	1.2	8.4
7	**9**	63	0.8	7.2
12	**2**	24	5.8	11.6
Toplam	20	124		37.6

$\bar{x} = \frac{\sum f_i X_i}{n} = \frac{124}{20} = 6.2$ olup, $\mathrm{O.S.} = 1.88$ dir.

d) Varyans

En yaygın olarak kullanılan değişim ölçüsüdür. Varyansın büyük olması, değişkenliğin fazla olduğunu gösterir. Kitle ve örneklem için varyans hesaplama formülleri aşağıdaki gibidir.

$$Kitle(Anakütle)\ varyansı = \sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

$$Örneklem\ varyansı = s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n-1}$$

Formüllerde;

X_i : Değişken değerleri (gözlem değerleri)

N : Kitle hacmi

μ : Kitle ortalaması

n : Örneklem hacmi

$\bar{x}$: Örneklem ortalaması

$n-1$: Serbestlik derecesi

anlamındadır.

NOT: $s^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{x})^2}{n-1} = \frac{1}{n-1}\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right]$ dir.

Örnek: Aşağıdaki veri için varyans değerini bulunuz.

X_i : 49, 49, 49, 50, 51, 52

$\bar{x} = 50$, $n = 6$ olup

$$s^2 = \frac{(49-50)^2 + (49-50)^2 + (49-50)^2 + (50-50)^2 + (51-50)^2 + (52-50)^2}{6-1} = \frac{8}{5} = 1.6$$

bulunur.

SORU: Frekans serilerinde ve gruplandırılmış serilerde varyans formülleri nasıl olur?

$$Frekans\ serilerinde\ \ s^2 = \frac{\sum_{i=1}^{n} f_i (X_i - \bar{x})^2}{n-1}$$

$$Gruplandırılmış\ serilerde\ \ s^2 = \frac{\sum_{i=1}^{n} f_i (m_i - \bar{x})^2}{n-1}$$

olur.

Ödev. Aşağıdaki veri seti için varyansı bulunuz. (Cevap = 0.722)

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

Varyansın Özellikleri

1) c sabit bir sayı olmak üzere,

$V(X \pm c) = V(x)$ ‘dir.

Yani, sabitin varyansı sıfırdır. $V(X \pm c) = V(X) \pm V(c) = V(x)$

2) Bir serinin bütün terimlerini aynı sabit sayı ile çarptığımızda (böldüğümüzde) varyans, çarptığımız (böldüğümüz) sayının karesiyle orantılı olarak büyür (küçülür).

$$V(cX) = c^2 V(X)$$

$$V\left(\frac{X}{c}\right) = \frac{1}{c^2} V(X)$$

3) Birbirinden bağımsız X ve Y serileri için aşağıdaki eşitlik geçerlidir.

$$V(X \pm Y) = V(X) + V(Y)$$

NOT: Varyansın birimi yoktur.

e) Standart Sapma

Varyansın birimi yoktur. Bu nedenle bazı durumlarda kullanışlı olmamaktadır. Değişkenliği, verilerin elde edildiği birimle ifade etmek için varyansın karekökü olan standart sapma kullanılır.

Standart sapma gözlemlerin ortalamadan ne kadar uzaklaştığını gösteren dağılım ölçüsüdür.

$$\boldsymbol{Standart\ sapma} = \boldsymbol{s} = \sqrt{\boldsymbol{Varyans}} = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1}{n-1}\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right]}$$

formülüyle hesaplanır.

Örnek. Aşağıdaki veri için standart sapma değerini bulunuz.

X_i : 49, 49, 49, 50, 51, 52

$\bar{x} = 50$, $n = 6$ olup

$$s^2 = \frac{(49-50)^2 + (49-50)^2 + (49-50)^2 + (50-50)^2 + (51-50)^2 + (52-50)^2}{6-1} = \frac{8}{5} = 1.6$$

olup

$$\boldsymbol{s} = \sqrt{s^2} = \sqrt{1.6} = 1.2649$$

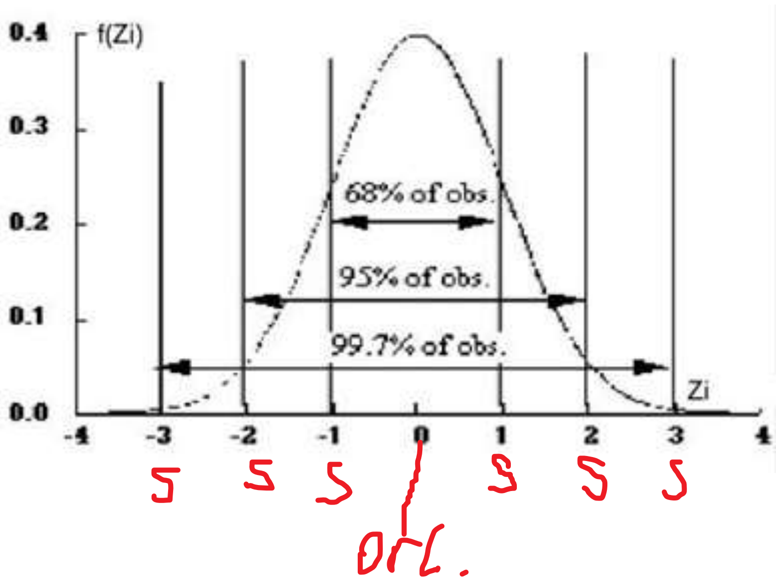
bulunur.

NOT: Veriler Normal Dağılıma sahipse verilerin,

% 68'i "ortalama ± 1 (standart sapma)",

% 95'i "ortalama ± 2 (standart sapma)",

% 99'u "ortalama ± 3 (standart sapma)" aralığındadır.



NOT: Merkezi eğilim ölçüsü olarak ortancanın(medyanın) kullanıldığı durumlarda, değişkenlik(dağılım) ölçüsü olarak çeyrek sapma kullanılır.

NOT: Merkezi eğilim ölçüsü olarak ortalamanın kullanıldığı durumlarda, değişkenlik(dağılım) ölçüsü olarak standart sapma kullanılır.

Standart Hata

Standart hata örnekleme dağılımındaki ortalamaların standart sapmasıdır. Seçilecek örneklemlerde ortalamalar arasındaki yaygınlığı gösterir. Standart hata örnek büyüklüğünün fonksiyonudur. Böylece n attıkça hata küçülür. ***Standart hata*** Aritmetik ortalamada oluşan hatanın belirlenmesi için bulunur.

Standart hata ortalamalarla ilgilidir, deneklerle ilgili değildir. Standart hata, standart sapma gibi değişkenin dağılışı hakkında bilgi vermez.

$$\boldsymbol{Standart\ hata} = \boldsymbol{s_{\bar{x}}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$$

Standart hata aynı popülasyondan seçilecek, aynı büyüklükteki örneklemlerin ortalamalarının yayılmasının ölçütüdür. Standart hata grupların ortalamalarının birbirleri ile karşılaştırılmasında kullanılır.

Ortalamanın standart hatası, ortalamanın dağılımındaki varyasyonu (değişimi) gösterir, örneklem sayısının artması ile küçülür. Standart hatanın küçük olması popülasyon parametresine ait yapılacak olan tahminler açısından ve daha dar güven aralığı sınırlar bulma açısından önemlidir.

Standart hata ne kadar küçükse, popülasyon ait tahminlerimiz de o kadar isabetli olmaktadır.

Not:Standart Sapma mı? Standart Hata mı?

Çalışma gruplarına ait veriler, sadece ilgili olduğu grubun özelliğini/özelliklerini gösteriyorsa *"ortalama ±standart sapma"* tercih edilmelidir.

Verileri birbiri ile karşılaştırarak, gruplar arasında fark olup olmadığı öğrenilmek isteniyorsa; *"ortalama ± standart hata"* kullanılması daha uygun olacaktır.

f) Değişim (Değişkenlik, Varyasyon) Katsayısı

Bir serinin standart sapmasının aritmetik ortalamasına bölünüp 100 ile çarpılması sonucu elde edilen değere değişim katsayısı adı verilir. Standart sapmanın ortalamaya göre nispi değeri olarak ifade edilir. Ölçü birimi yoktur.

$$D.K. = \frac{s}{\bar{x}} 100$$

formülü kullanılarak hesaplanır.

Standart sapma, **bir** kitle veya örneklemdeki değişkenliğin ölçülmesinde kullanılır.

Değişim katsayısı, **birden fazla** kitle veya örneklemdeki değişkenliğin karşılaştırılmasında kullanılır.

Değişim katsayısı küçük olan verilerin diğerlerine göre daha az değişim gösterdiği söylenir. Bunun anlamı, verilerin aritmetik ortalama etrafında daha homojen olarak dağıldığıdır.

NOT: Genel olarak D.K. < 20 ise grup homojen, D.K. > 25 ise grup heterojen, diğer durumlarda normal olarak nitelendirilir.

Örnek. Bir sınıftan rastgele belirlenen 10 Kız ve 10 Erkek öğrencinin F dersinden aldıkları notlar aşağıdaki gibidir.

Kız Öğrenciler : 65, 30, 45, 70, 40, 50, 85, 60, 55, 50

Erkek Öğrenciler : 90, 60, 70, 60, 50, 75, 80, 90, 70, 40

Kız öğrencilerin mi yoksa Erkek öğrencilerin mi notları daha az değişkenlik göstermektedir?

X_i : Kız öğrencilerin notları , $\bar{x} = 55$

Y_i : Erkek öğrencilerin notları , $\bar{y} = 68.5$

X_i	Y_i	$X_i - \bar{x}$	$(X_i - \bar{x})^2$	$Y_i - \bar{y}$	$(Y_i - \bar{y})^2$
65	90				
30	60				
45	70				
70	60				
40	50				
50	75				
85	80				
60	90				
55	70				
50	40				
Toplam			2250		2272.5

D.K.(X) = 28.74 D.K.(Y) = 23.85

D.K.(Y) < D.K.(X) olduğundan Erkek öğrencilerin notları daha az değişim göstermektedir.