Tanımlayıcı İstatistikler

Değişkenlik Ölçüleri

Hazırlayan: Dr. Gökhan AKSU

Tanımlayıcı İstatistikler

- Bir veri setini tanımak veya birden fazla veri setini karşılaştırmak için kullanılan ve ayrıca örnek verilerinden hareket ile frekans dağılışlarını sayısal olarak özetleyen değerlere **tanımlayıcı istatistikler** denir**.**

- Analizlerde kullanılan veri tiplerine (*basit*, *gruplanmış, sınıflanmış*) göre hesaplamalarda kullanılacak formüller değişmektedir.

Tanımlayıcı İstatistikler

Merkezi Eğilim
1)Aritmetik ort.
2)Geometrik ort.
3)Harmonik ort.
4)Mod
5)Medyan

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

Çarpıklık Ölçüleri

Basıklık Ölçüleri

8) Ortalama Mutlak Sapma(OMS)

- Veri setindeki her bir gözlem değerinin mutlak değerce aritmetik ortalamadan farklarının toplamının, örnek hacmine bölünmesiyle elde edilir.
- Gözlem değerlerinin aritmetik ortalamadan faklarının toplamı 0 olacağından bu problemi ortadan kaldırmak için mutlak değer ifadesi kullanılır.

Basit veriler için:

$$OMS = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n}$$

Gruplanmış veriler için:

$$OMS = \frac{\sum_{i=1}^{k} f_i |x_i - \overline{x}|}{\sum_{i=1}^{k} f_i}$$

Sınıflanmış veriler için :

$$OMS = \frac{\sum_{i=1}^{k} f_i |m_i - \overline{x}|}{\sum_{i=1}^{k} f_i}$$

Örnek: İstatistik I dersini alan 10 öğrencinin vize notları aşağıdaki gibi sıralanmıştır. Buna göre vize notları için ortalama mutlak sapma değerini hesaplayınız.

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

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

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

$$= \frac{189}{10} = 18{,}4$$

Soru:

- 25, 62, 45, 23, 32, 10, 17, 36, 40, 60

Yukarıda verilen serinin ortalama mutlak sapması kaça eşittir?

	x	x - ort	mutlak
1	25	-10	10
2	62	27	27
3	45	10	10
4	23	-12	12
5	32	-3	3
6	10	-25	25
7	17	-18	18
8	36	1	1
9	40	5	5
10	60	25	25
mean	35	total	136
	35	OMS	13,6

Sınıflanmış Veriler İçin Ortalama Mutlak Sapma Örneği

Sınıflar	f_i
150-157'den az	5
157-164'den az	7
164-171'den az	14
171-178'den az	9
178-185'den az	8
185-192'den az	4
192-199'dan az	3

$$OMS = \frac{\sum_{i=1}^{k} f_i \left| m_i - \bar{x} \right|}{\sum_{i=1}^{k} f_i}$$

Sınıflanmış Veriler İçin Ortalama Mutlak Sapma Örneği

Sınıflar	f_i	m_i	$\lvert f_i(m_i-\bar{x})\rvert$
150-157'den az	5	***153,5***	**92,4**
157-164'den az	7	***160,5***	**80,36**
164-171'den az	14	***167,5***	**62,72**
171-178'den az	9	***174,5***	**22,68**
178-185'den az	8	***181,5***	**76,16**
185-192'den az	4	***188,5***	**66,08**
192-199'dan az	3	***195,5***	**70,56**
Toplam	50		***470,96***

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{\sum_{i=1}^{k} f_i} = 171,98\,kg.$$

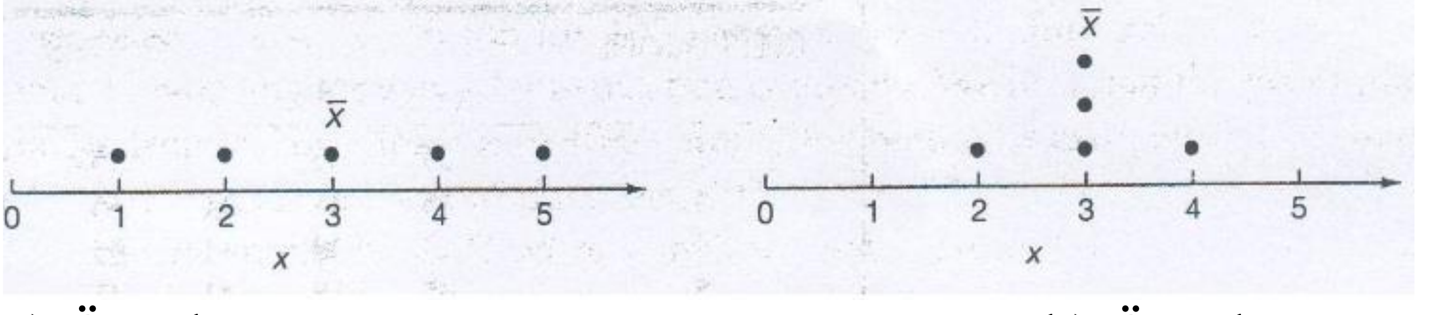
$$OMS = \frac{\sum_{i=1}^{k} f_i \left| m_i - \bar{x} \right|}{\sum_{i=1}^{k} f_i} = \frac{470,96}{50} = 9.42$$

=9,4192

Yayılma Ölçülerinin Gerekliliği

	Örnek 1	**Örnek 2**
Ölçümler	1,2,3,4,5	2,3,3,3,4
Ortalama	$\bar{x} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$	$\bar{x} = \frac{2+3+3+3+4}{5} = \frac{15}{5} = 3$
$\bar{x}$ dan Uzaklıklar	1-3, 2-3, 3-3, 4-3, 5-3 veya -2, -1, 0, 1, 2	2-3, 3-3, 3-3, 3-3, 4-3 veya -1, 0, 0, 0, 1

İki veri seti için uzaklıklar



a) Örnek 1

b) Örnek 2

9) Varyans

- Ortalama mutlak sapmada kullanılan mutlak değerli ifadeler ile işlem yapmanın zor hatta bazı durumlarda imkansız olması sebebiyle yeni değişkenlik ölçüsüne ihtiyaç bulunmaktadır.

- Mutlak değer ifadesindeki zorluk aritmetik ortalamadan farkların karelerinin alınmasıyla ortadan kalkmaktadır.

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

Basit veriler İçin:

Anakütle Varyansı:

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

μ : Anakütle Ortalaması N : Anakütle Hacmi

Örnek Varyansı :

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

Gruplanmış veriler için:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \overline{x})^2}{\sum_{i=1}^{k} f_i - 1}$$

Sınıflanmış veriler için :

$$s^2 = \frac{\sum_{i=1}^{k} f_i (m_i - \overline{x})^2}{\sum_{i=1}^{k} f_i - 1}$$

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$

ifadesi istatistikte bir çok formülde kullanılır ve ***kareler toplamı*** olarak adlandırılır.

- Matematiksel olarak hesaplama kolaylığı sağlaması açısından formüllerde kareler toplamının açılımı olan aşağıdaki eşitlik kullanılabilir.

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

Basit Veriler İçin:

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

Gruplanmış Veriler İçin:

$$s^2 = \frac{\sum_{i=}^{k} f_i x_i^2 - \frac{\left(\sum_{i=}^{k} f_i x_i\right)^2}{\sum_{i=1}^{k} f_i}}{\sum_{i=1}^{k} f_i - 1}$$

Sınıflanmış Veriler İçin :

$$s^2 = \frac{\sum_{i=}^{k} f_i m_i^2 - \frac{\left(\sum_{i=}^{k} f_i m_i\right)^2}{\sum_{i=1}^{k} f_i}}{\sum_{i=1}^{k} f_i - 1}$$

Örnek: Bir un fabrikasının satış mağazasında bir gün içinde satılan un paketlerinin gramajlarına göre göre satış adetleri aşağıda verilmiştir. Buna göre veri seti için varyans değerlerini hesaplayınız.

Kg	Satış adedi
1	5
2	12
3	30
4	10
5	12
6	6

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} f_i - 1}$$

Kg	Satış adedi	x*f		X -Xort	(X-Xort)^2	f*(x-xort)^2
1	5	5		-2,4	5,76	28,8
2	12	24		-1,4	1,96	23,52
3	30	90		-0,4	0,16	4,8
4	10	40		0,6	0,36	3,6
5	12	60		1,6	2,56	30,72
6	6	36		2,6	6,76	40,56
toplam	75	255				132
aritmetik ortalama		**3,4**			**varyans**	**1,76**

Sınıftaki öğrencilerin yaşları ve yaşlara ilişkin öğrenci sayıları aşağıda verilmiştir.

Buna göre varyans kaça eşittir?

Yaş	*Kişi sayısı*
19	4
20	6
21	12
22	16
23	8
24	2

Cevap: 1,58

yaş	öğrenci sayısı	x*f		X -Xort	(X-Xort)^2	f*(x-xort)^2
19	4	76		-2,5	6,25	25
20	6	120		-1,5	2,25	13,5
21	12	252		-0,5	0,25	3
22	16	352		0,5	0,25	4
23	8	184		1,5	2,25	18
24	2	48		2,5	6,25	12,5
toplam	48	1032				76
aritmetik ortalama		**21,5**			**varyans**	**1,5833333**

Örnek: Bir un fabrikasının satış mağazasında bir gün içinde satılan un paketlerinin gramajlarına göre göre satış adetleri aşağıda verilmiştir. Buna göre veri seti için varyans değerlerini hesaplayınız.

Kg	Satış adedi	$x_i.f_i$	$x^2_i.f_i$
1	5	***5***	***5***
2	12	***24***	***48***
3	35	***105***	***315***
4	14	***56***	***224***
5	8	***40***	***200***
6	6	***36***	***216***
toplam	80	***266***	***1008***

$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2 - \frac{\left(\sum_{i=1}^{k} f_i x_i\right)^2}{\sum_{i=1}^{k} f_i}}{\sum_{i=1}^{k} f_i - 1} = \frac{1008 - \frac{(266)^2}{80}}{79} \cong 1.56$$

Sınıflanmış Veriler İçin Ortalama Varyans Örneği

Sınıflar	f_i	m_i	$f_i(m_i-\bar{x})^2$
150-157’den az	5	***153,5***	1707,552
157-164’den az	7	***160,5***	922,5328
164-171’den az	14	***167,5***	280,9856
171-178’den az	9	***174,5***	57,1536
178-185’den az	8	***181,5***	725,0432
185-192’den az	4	***188,5***	1091,642
192-199’dan az	3	***195,5***	1659,571
Toplam	50		6444,48

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{\sum_{i=1}^{k} f_i} = 171{,}98\,kg. \quad s^2 = \frac{\sum_{i=1}^{k} f_i (m_i - \bar{x})^2}{\sum_{i=1}^{k} f_i - 1} = \frac{6444{,}48}{50-1} \approx 131{,}52$$

10) Standart Sapma

- Varyans hesaplanırken kullanılan verilerin kareleri alındığında mevcut ölçü biriminin de karesi alınmış olur.

- **Örnek: kg^2, cm^2 gibi.**

- Bu nitelendirme veriler açısından bir anlam taşımayacağından varyans yerine ortalama etrafındaki değişimin bir ölçüsü olarak onun pozitif karekökü olan ***standart sapma*** kullanılır.

Basit Veriler İçin:

Populasyon Standart Sapması:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

σ : Populasyon Standart Sapması N : Populasyon Hacmi

Örnek Standart Sapması :

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

Gruplanmış Veriler İçin:

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

Sınıflanmış Veriler İçin :

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i (m_i - \bar{x})^2}{\sum_{i=1}^{k} f_i - 1}}$$

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

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

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

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

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

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

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

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

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

Evren deseydi std.sapma 21,30 olacaktı.

Aynı soru kareler ortalamasının açılımı kullanılarak çözüldüğünde aynı sonuçları verecektir.

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

x	x^2
30	900
41	1681
53	2809
61	3721
68	4624
79	6241
82	6724
88	7744
90	8100

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x\right)^2}{n}}{n-1} = \frac{52148 - \frac{(690)^2}{10}}{9}$$

$$s^2 \approx 504,22$$

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

$$\sum_{i=1}^{n} x_i = 690 \quad \sum_{i=1}^{n} x_i^2 = 52148$$

Standart Sapmanın Yorumlanması

- Chebyshev teoreminden, frekans dağılımının şekline bakılmaksızın, ölçümlerin herhangi bir örneğine uygulanan kural:

a- Ölçümlerden hiçbirinin $\bar{x} \pm s \; yada \; (\bar{x} - s, \bar{x} + s)$ aralığına düşmemesi mümkündür.

b- Ölçümlerin en az ¾'ü $(\bar{x} - 2s, \bar{x} + 2s)$ aralığına düşer.

c- Ölçümlerin en az 8/9'u $(\bar{x} - 3s, \bar{x} + 3s)$ aralığına düşer.

d- Genellikle, ölçümlerin en az $(1\text{-}1/k^2)$'ı $(\bar{x} - ks, \bar{x} + ks)$ aralığına düşer. (k>1)

- Simekrik dağılışlarda standart sapmanın yorumu:

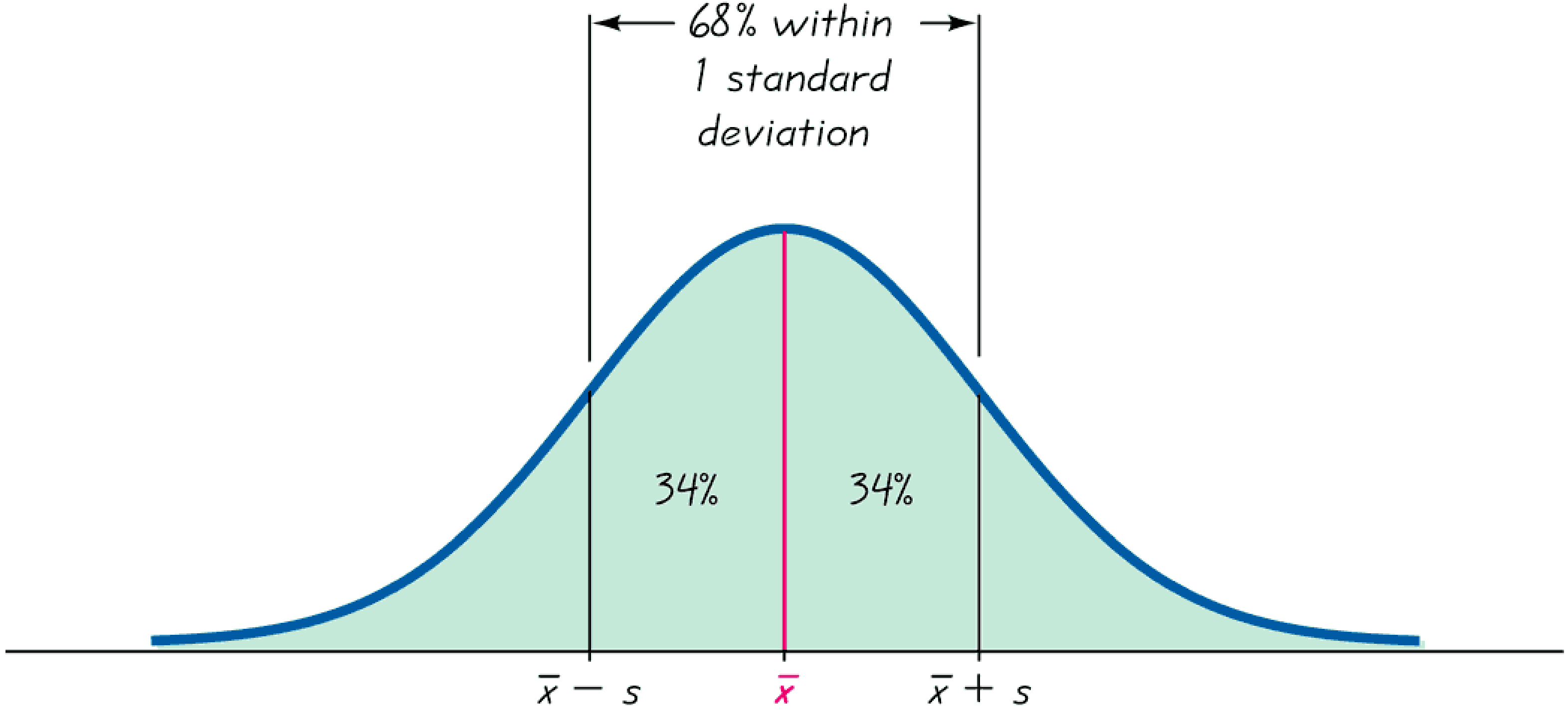
a- Ölçümlerin yaklaşık %68'i

$\bar{x} \pm s \; yada \; (\bar{x} - s, \bar{x} + s)$ aralığına düşer.- ortalamanın 1 standart sapması için

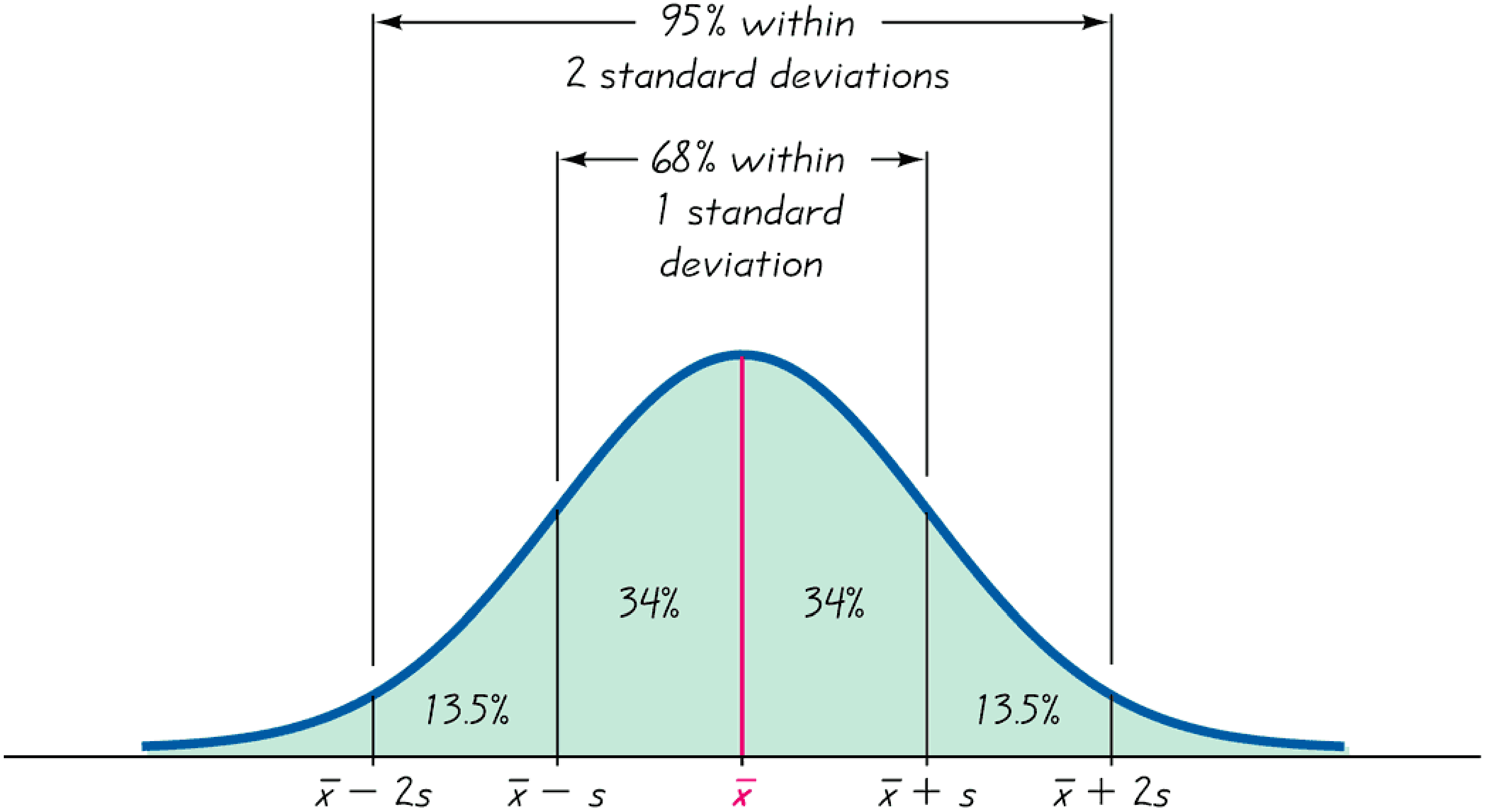
b- Ölçümlerin yaklaşık %95'i $(\bar{x} - 2s, \bar{x} + 2s)$ aralığına düşer.- ortalamanın 2 standart sapması için

c- Temelde, tüm ölçümler $(\bar{x} - 3s, \bar{x} + 3s)$ aralığına düşer. -ortalamanın 3 standart sapması için

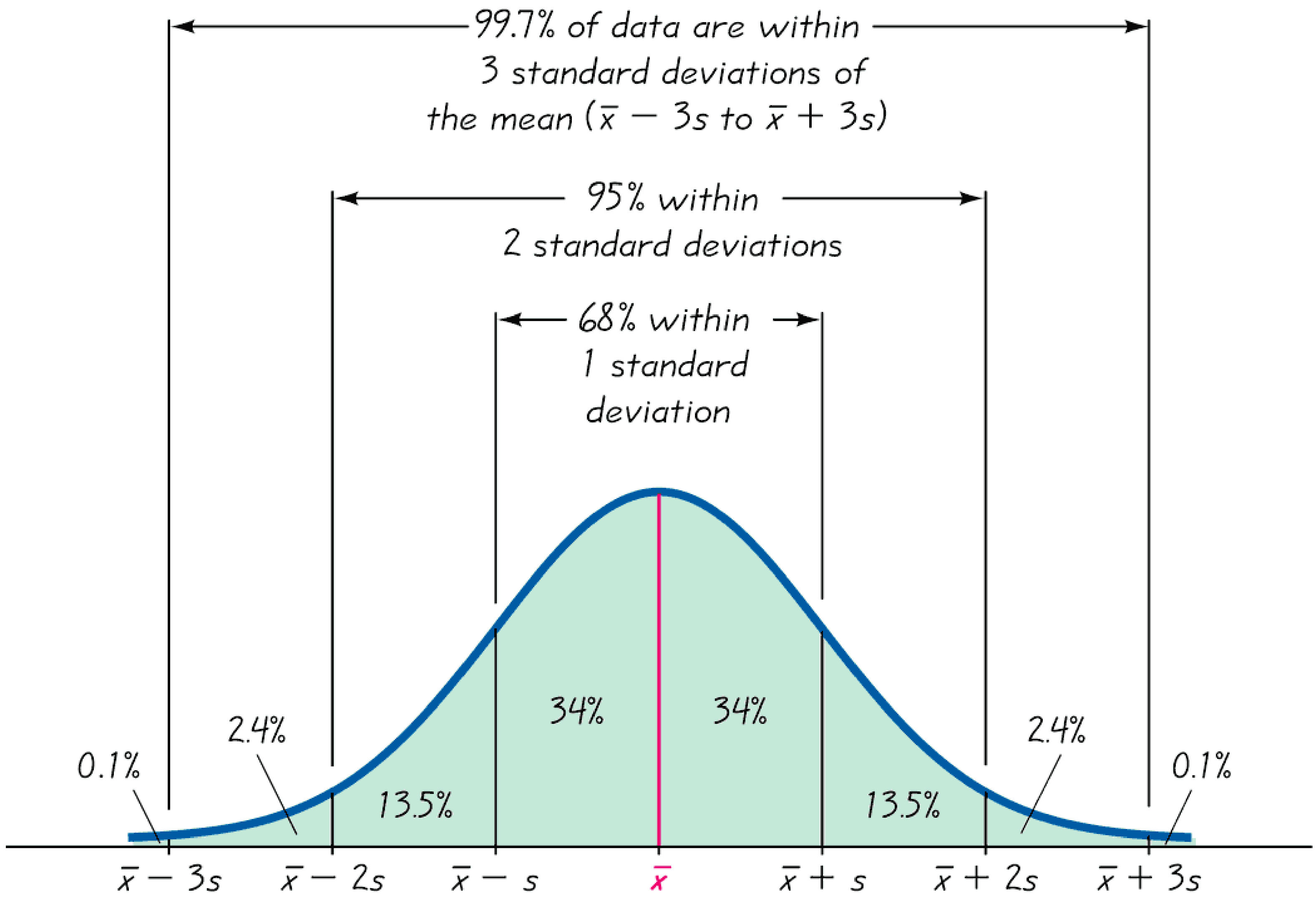
Ampirik Kural



Ampirik Kural

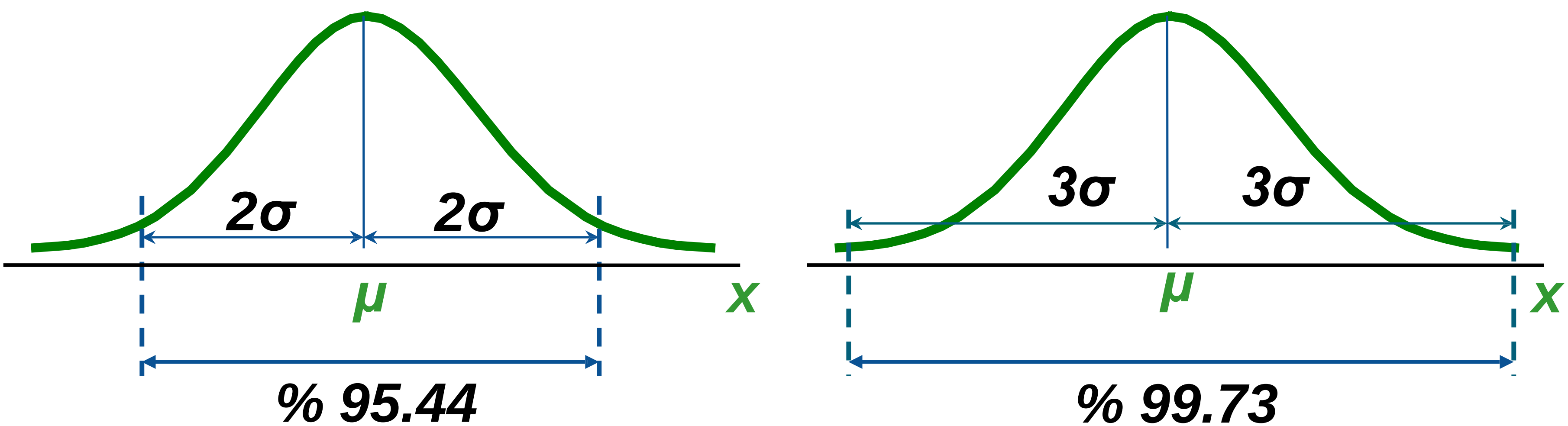


Ampirik Kural



Bazı kurallar…

- *$\mu \pm 2\sigma$ X'lerin yaklaşık % 95 'ini içerir*
- *$\mu \pm 3\sigma$ X'lerin yaklaşık % 99.7 'ini içerir*



11) z Skoru

Verilen bir gözlem değerinin ortalamanın kaç standart sapma uzağında olduğunu ölçer.

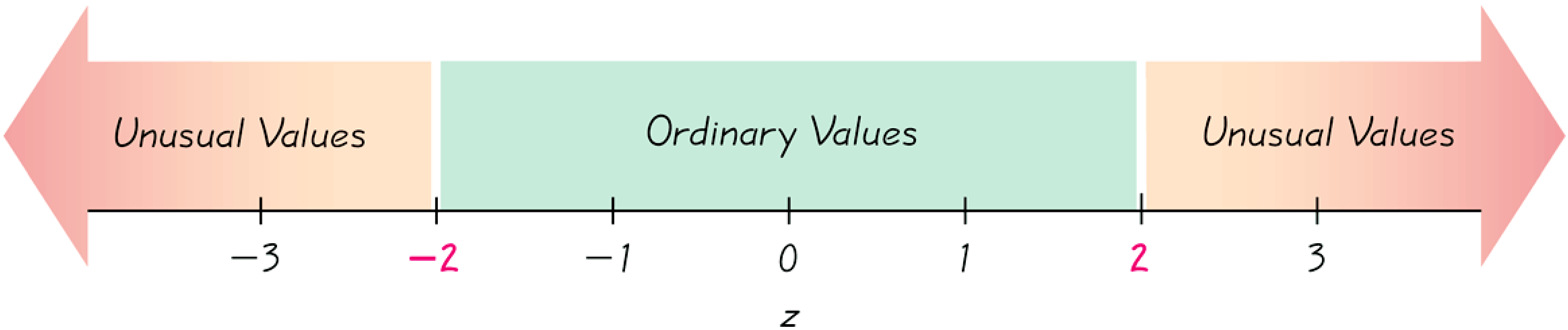
Örneklem **Anakütle**

$$z = \frac{x - \bar{x}}{s}$$

$$z = \frac{x - \mu}{\sigma}$$

2 ondalık basamağa yuvarlanır.

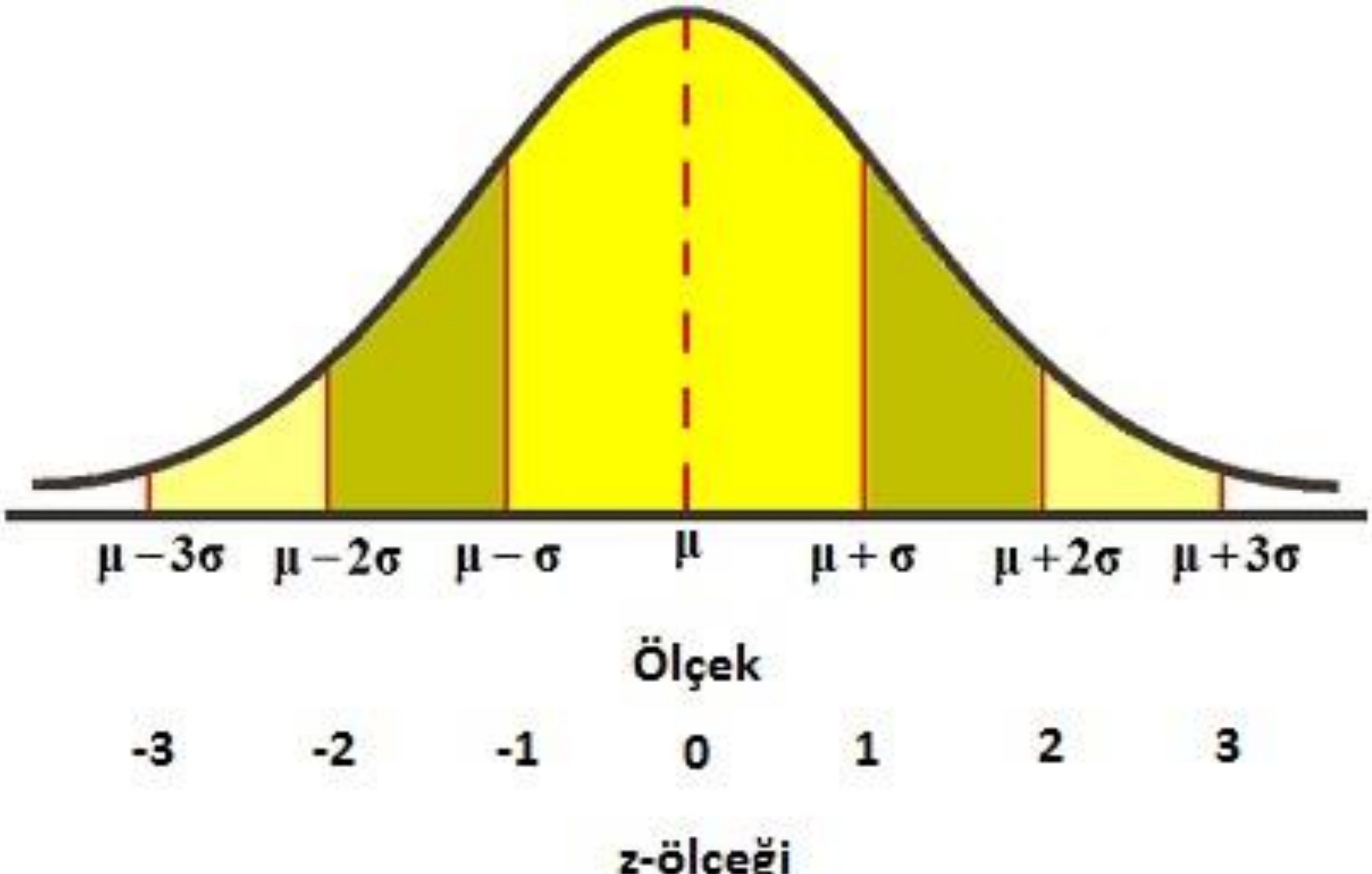
z- skorunun Yorumlanması



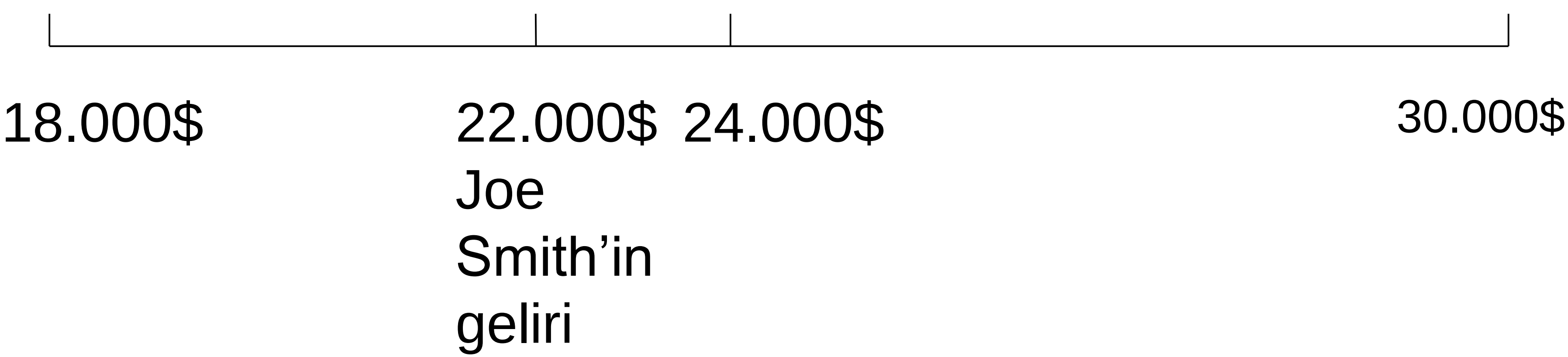
Bir veri ortalamadan küçük olursa z-skoru değeri negatif olur.

Olağan Veriler : *z* skoru –2 ve 2 s.s arasında

Olağandışı Veriler: *z* skoru < -2 veya *z* skoru > 2 s.s

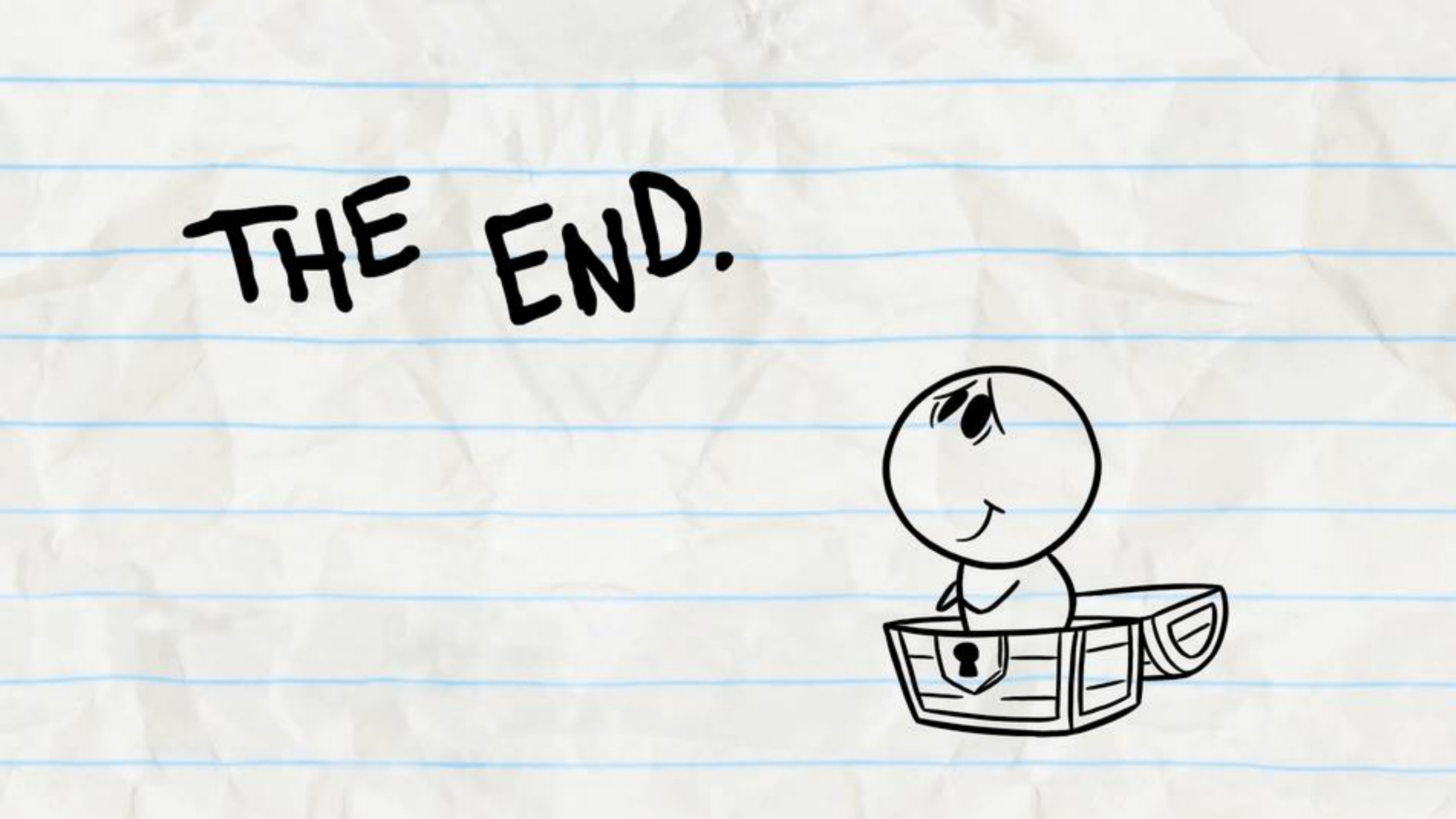

μ − 3σ
μ − 2σ
μ − σ
μ
μ + σ
μ + 2σ
μ + 3σ
Ölçek
-3
-2
-1
0
1
2
3
z-ölçeği

- **Örnek:** 200 çelik işçisinin yıllık gelirleri incelenmiş ve ortalaması = 24.000$ ve standart sapması s= 2.000$ olarak bulunmuştur. Yıllık geliri 22.000$ olan Joe Smith’in z-skoru kaçtır?



$z = \frac{x - \bar{x}}{s} = \frac{22.000\$ - 24.000\$}{2.000\$} = -1.0$ bulunur. Burada ki -1.0 ın anlamı Joe Smith'in yıllık geliri ortalamanın 1 standart sapma altındadır.

z-skorunun sayısal değeri göreli durumlar için ölçümü yansıtmaktadır. Bir x değeri için bulunan en büyük pozitif z-skoru değeri, bu x değerinin diğer bütün ölçümlerden daha büyük olduğunu gösterir ve mutlak değerce en büyük negatif z-skoru değeri de bu ölçümün diğer tüm ölçümlerden daha küçük olduğunu gösterir. Eğer z skoru 0 veya 0'a yakın ise ölçüm ortalamaya eşit veya ortalamaya çok yakındır.



Dinlediğiniz İçin Teşekkür Ederim…

Dr. Gökhan AKSU