HACETTEPE ÜNİVERSİTESİ

İSTATİSTİK BÖLÜMÜ

İST155 İSTATİSTİĞE GİRİŞ I

DERS 5 –KONUM ÖLÇÜLERİ-I

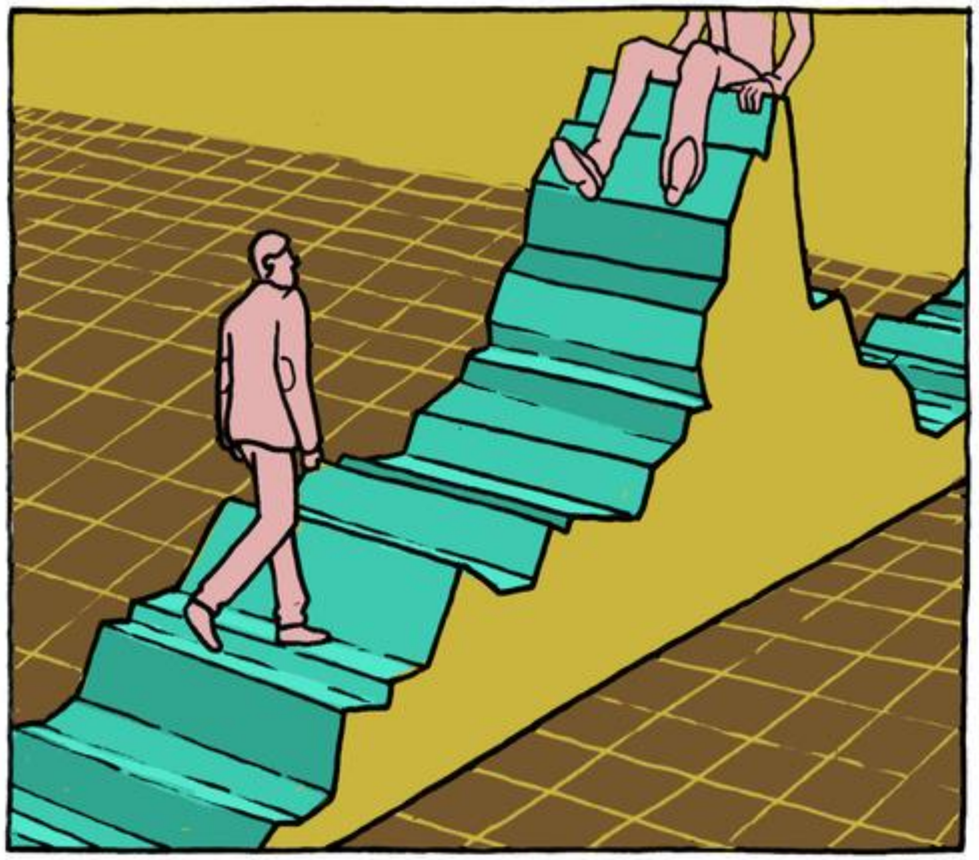
Ders sorumluları: Prof.Dr.Serpil AKTAŞ ALTUNAY (01 Şubesi)
Doç.Dr. Ayten YİĞİTER (02 Şubesi)

KONUM ÖLÇÜLERİ (MEASURES OF LOCATION –MEASURES OF CENTRAL TENDENCY)

Konum ölçüleri, verilerin dağılımlarının birbirlerine göre yerlerini, birbirlerine olan uzaklıklarını kısacası konumlarını belirlemek için kullanılan ölçülerdir. Dağılım hakkında bilgi elde etmenizi sağlarlar. Günlük hayatta en sık «Ortalama» sözcüğü kullanılır. Örneğin ortalama ücret, ortalama not, ortalama gelir vb.

Konum Ölçüleri

- Nicel veriler için
- Nitel veriler için



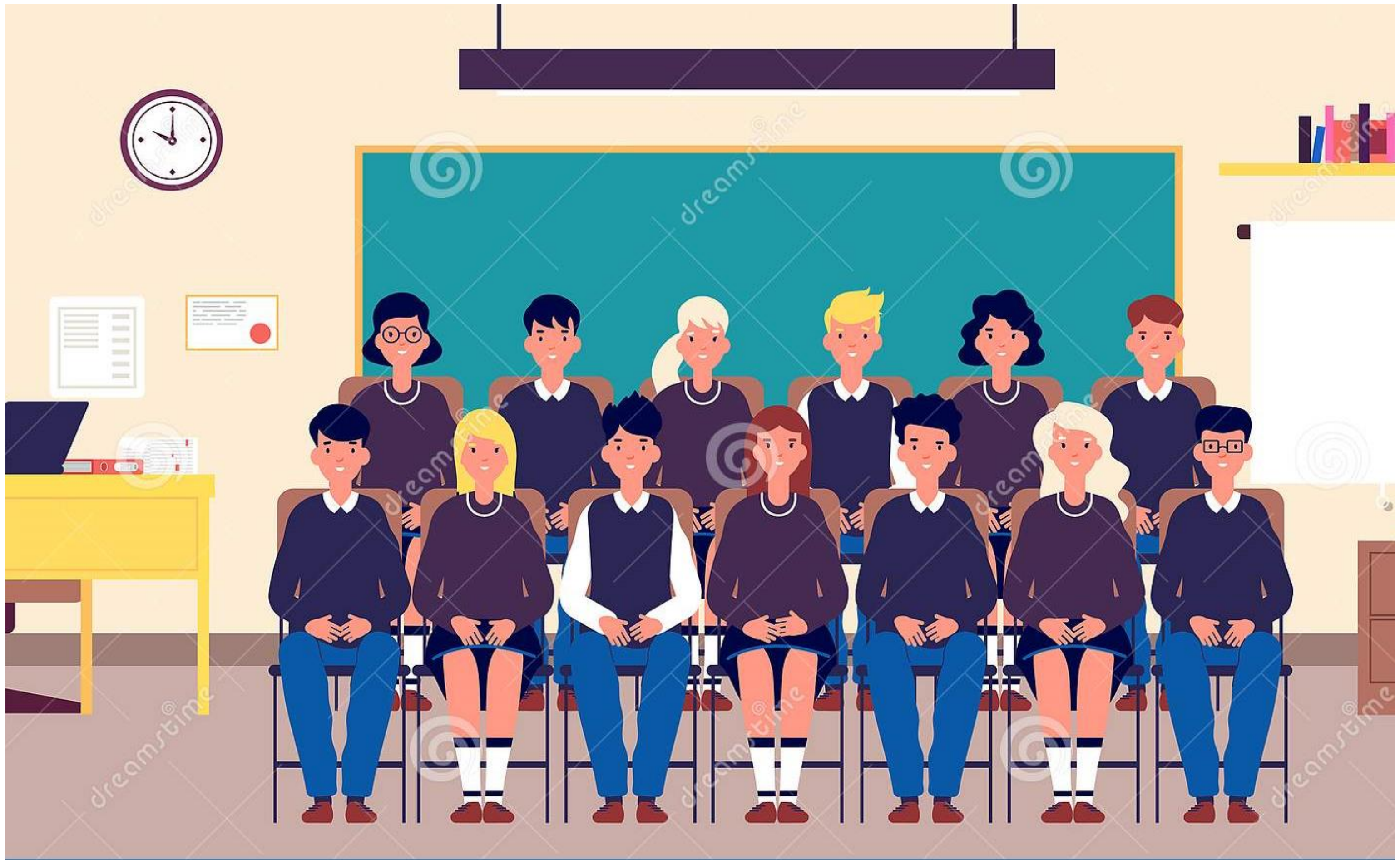
Verilerden elde edilen özetleyici bilgilerin nerede konumlandığını gösterir, onun için «Konum Ölçüsü» diyoruz.

Cep telefonundan konum göndeririz, o konum aslında bizim bir alan üzerinde nerede konumlandığımızı ve diğer birimlere olan uzaklığımızı tanımlar …

Biz İstatistiksel konum ile ilgileniyoruz !

Nicel Verilerde Konum Ölçüleri

- Aritmetik Ortalama (mean-average)
- Tepe Değeri-Mod (Mode)
- Ortanca-Medyan (Median)
- Çeyreklikler (Quartiles)
- Yüzdelikler (Percentiles)
- Ağırlıklı, Geometrik, Harmonik ortalama



Bu sınıftaki öğrencilerin yaş ortalaması kaçtır ?

Bir İstatistikçi kamu'da ortalama ne kadar maaş alır ?

Günde ortalama kaç kez telefon görüşmesi yapıyoruz?

Günlük yaşamda bu gibi soruları çok sorarız !

1-) ARİTMETİK ORTALAMA

Günlük yaşamda en sık kullanılan konum ölçüsüdür. $\bar{X}$ ile gösterilir.

Kitle

- $\mu = \frac{\sum_{i=1}^{N} X_i}{N}$

Kitle büyüklüğü: N

Örneklem büyüklüğü: n

Sınıf sayısı:k

Örneklem

- Ham veri: $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
- Sınıflandırılmış veri: $\bar{X} = \frac{\sum_{i=1}^{k} f_i \times s_i}{n}$

$$\bar{X} = \sum_{i=1}^{k} p_i \times s_i$$

Aritmetik ortalamanın özellikleri şunlardır:

- Dağılımda bulunan tüm değerler dikkate alınarak hesaplanır.

- Dağılımdaki verilerin aritmetik ortalamadan ayrılışları toplamı sıfırdır.

- Dağılım sınırları (uç değerler) aritmetik ortalama üzerinde etkilidir. Eğer dağılımın sınırlarında aşırı uç değerler varsa, aritmetik ortalama dağılımın konumu hakkında tutarlı bilgi vermez. Bu durumda dağılım sınırlarından etkilenmeyen ortanca, çeyrek değerler, tepe değeri gibi konum ölçülerinden yararlanılabilir.

SINIFLANDIRILMAMIŞ (HAM) VERİLERDE ARİTMETİK ORTALAMA

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

Örnek: 30 öğrenciye ait dönem sonu akademik ortalama puanları verilmiştir.

2.12	3.42	2.76	3.18	0.50	2.14
2.35	3.70	2.84	1.96	1.29	2.36
2.78	2.80	2.15	1.65	1.03	3.87
2.90	2.91	2.19	1.72	1.59	2.40
3.05	2.63	3.26	0.98	1.70	2.61

$\sum_{i=1}^{30} X_i$=74.84 $\qquad$ $\bar{X} = \frac{74.84}{30}$=2.3613

SINIFLANDIRILMAMIŞ (HAM) VERİLERDE ARİTMETİK ORTALAMA

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

Örnek: 30 öğrenciye ait dönem sonu akademik ortalama puanları verilmiştir.

2.12	3.42	2.76	3.18	0.50	2.14
2.35	3.70	2.84	1.96	1.29	2.36
2.78	2.80	2.15	1.65	1.03	3.87
2.90	2.91	2.19	1.72	1.59	2.40
3.05	2.63	3.26	0.98	1.70	2.61

$\sum_{i=1}^{30} X_i$=74.84 $\bar{X} = \frac{74.84}{30}$=2.3613

SINIFLANDIRILMIŞ VERİLERDE ARİTMETİK ORTALAMA

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i \times s_i}{n}$$

Örnek: 30 öğrenciye ait dönem sonu akademik ortalama puanları verilmiştir.

AS	ÜS	si	fi	fi×si
0.5	1.19	0.845	3	2.535
1.2	1.89	1.545	5	7.725
1.9	2.59	2.245	8	17.96
2.6	3.29	2.945	11	32.395
3.3	3.99	3.645	3	10.935
			30	71.55

$$\bar{X} = \frac{\sum_{i=1}^{5} fi \times s_i}{30} = \frac{71.55}{30} = 2.385$$

SINIFLANDIRILMIŞ VERİLERDE ARİTMETİK ORTALAMA

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i \times s_i}{n}$$

$$p_i$$

$$\bar{X} = \sum_{i=1}^{k} p_i \times s_i$$

Örnek: 30 öğrenciye ait dönem sonu akademik ortalama puanları verilmiştir.

AS	ÜS	si	fi	pi	pi×si
0.5	1.19	0.845	3	0.1	0.0845
1.2	1.89	1.545	5	0.167	0.2575
1.9	2.59	2.245	8	0.267	0.5987
2.6	3.29	2.945	11	0.367	1.0798
3.3	3.99	3.645	3	0.1	0.3645
			30	1	2.385

$$\bar{X} = \sum_{i=1}^{5} p_i \times s_i$$

$$\bar{X} = 2.385$$

SINIFLANDIRILMIŞ VERİLERDE ARİTMETİK ORTALAMA

$$\bar{X} = A + c\left(\frac{\sum_{i=1}^{k} f_i \times b_i}{n}\right)$$

$$b_i = \frac{s_i - A}{c} \rightarrow s_i = cb_i + A$$

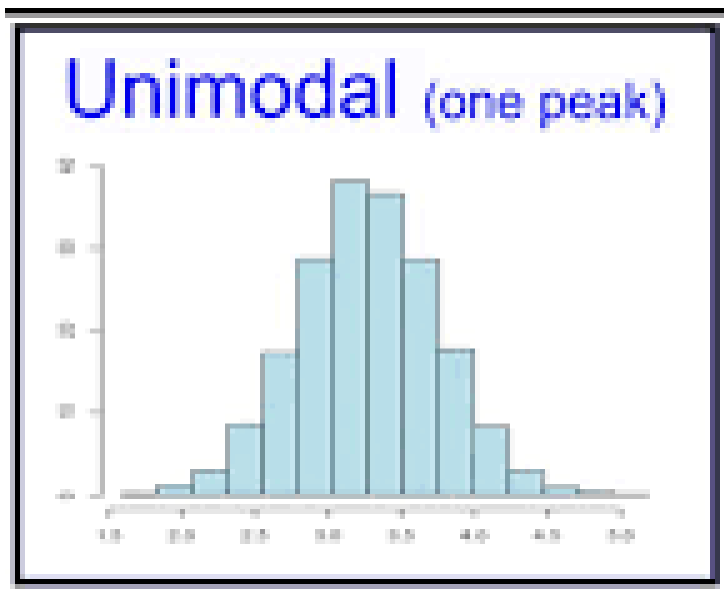
AS	ÜS	si	fi	bi	fi*bi
0.5	1.19	0.845	3	-2	-6
1.2	1.89	1.545	5	-1	-5
1.9	2.59	2.245	8	0	0
2.6	3.29	2.945	11	1	11
3.3	3.99	3.645	3	2	6
			30		Top=6

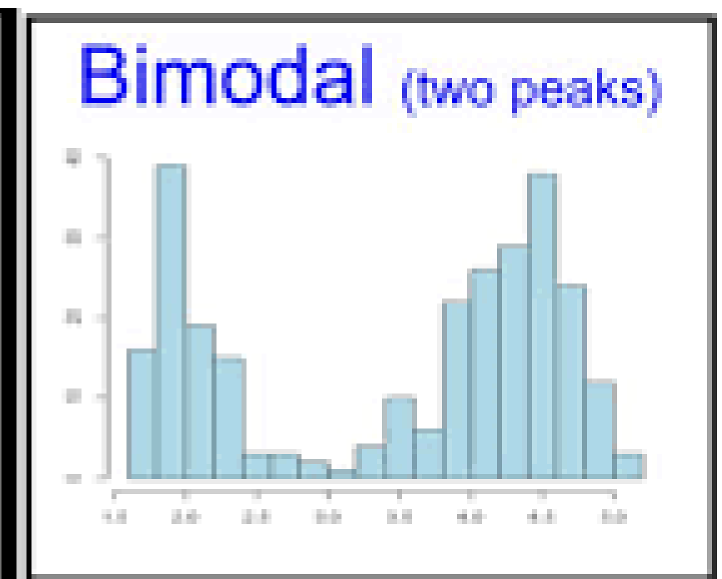
$$\bar{X} = \frac{\sum_{i=1}^{k} f_i \times s_i}{n}$$

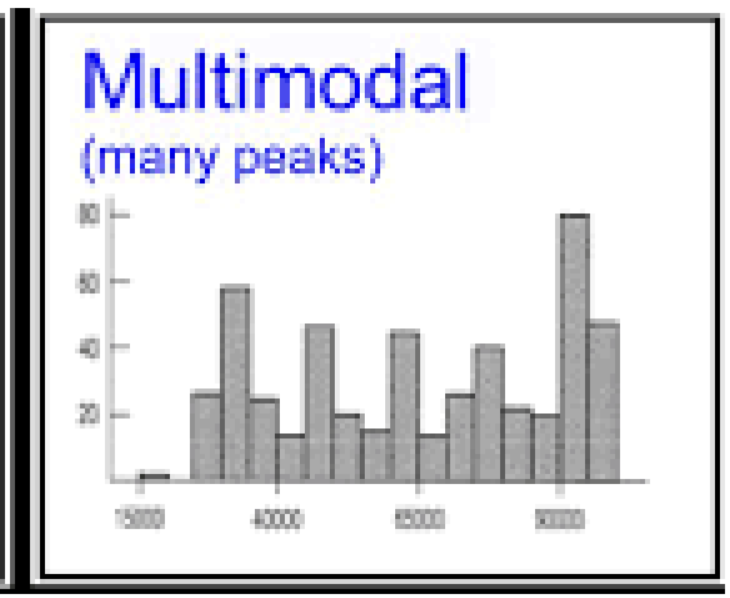
$$\bar{X} = A + c\frac{\sum_{i=1}^{k} f_i \times b_i}{n} = 2.245 + 0.7\left(\frac{6}{30}\right) = 2.245 + 0.14 = 2.385$$

2-) TEPE DEĞERİ-MOD (MODE)

Bir veri kümesinde en sık tekrarlanan değer, yani en yüksek sıklığı sahip olan değer tepe değeri olarak adlandırılır. Örneklem dağılımları her zaman için tek tepeli olmayabilir. Bazen en çok tekrarlanan değer iki ya da daha çok olabilir. Bu durumda dağılım iki tepeli ya da çok tepeli olarak adlandırılır. Dağılım tek tepeli ise «unimodal» iki tepeli «bimodal», çok tepeli ise «multimodal» adını alır.







SINIFLANDIRILMAMIŞ (HAM) VERİLERDE TEPE DEĞERİ-MOD

Örnek: Müşterilerini telefonla bilgi veren bir şirket, 20 farklı müşterinin telefondaki sürelerini kaydediyor. Bu süreler aşağıda verilmiştir. Bu verilere ilişkin tepe değerini bulunuz.

15	23	34	38	41
16	25	37	40	43
19	25	38	40	43
20	28	38	41	44

Bu veri kümesinde en sık tekrarlanan değer 38 (3 kez) olduğu için tepe değeri MOD= 38'dir.

SINIFLANDIRILMIŞ VERİLERDE TEPE DEĞERİ-MOD

$$\text{MOD}= \hat{X} = AS + c\left(\frac{D_1}{D_1+D_2}\right)$$

D_1: En büyük sıklık ile bir önceki sınıfın sıklığı arasındaki fark

AS: Sıklığı en büyük olan sınıfın alt sınırı

c: sınıf aralığı

D2: En büyük sıklık ile bir sonraki sınıfın sıklığı arasındaki fark

Örnek: Aşağıda bir sınava giren öğrencilerin yaptıkları hata sayılarının sıklık dağılımı verilmiştir. Bu veriler için tepe değerini bulunuz.

Sınıflar	fi
8-10	22
11-13	15
14-16	25
17-19	24
20-22	6
23-25	5
26-28	3

c=3

D1=25-15=10

En büyük sıklık ⇒AS=14

D2=25-24=1

$$\text{MOD} = \hat{X} = AS + c\left(\frac{D_1}{D_1+D_2}\right)$$

$$\text{MOD} = \hat{X} = 14 + 3 \times \frac{10}{10+1} = 16.73$$

Örnek: 30 öğrenciye ait dönem sonu akademik ortalama puanlarına ilişkin sıklık çizelgesi aşağıda verilmiştir.

AS	ÜS	fi
0.5	1.19	3
1.2	1.89	5
1.9	2.59	8
2.6	3.29	11
3.3	3.99	3
		30

c=0.7

D1=11-8=3

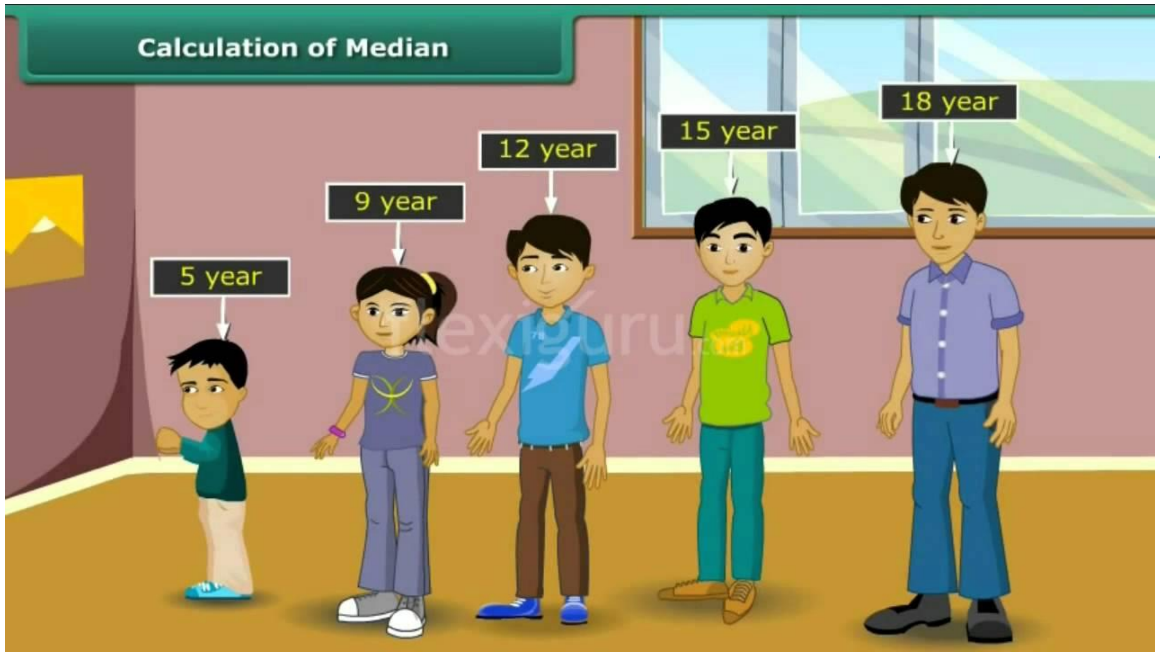
En büyük sıklık $\Rightarrow$ AS=2.6

D2=11-3=8

$$\text{MOD} = \hat{X} = AS + c\left(\frac{D_1}{D_1 + D_2}\right)$$

$$\text{MOD} = \hat{X} = 2.6 + 0.7 \times \frac{3}{3+8} = 2.6 + 0.19 = 2.79$$

3-) ORTANCA-MEDYAN (MEDIAN)



12 yaşındaki çocuk kendini nasıl tanımlar ?

Bu takımdaki oyuncuların boy ortalamasını nasıl hesaplayacağız ? En uzun boylu çocuk ortalamayı etkilemez mi?

SINIFLANDIRILMAMIŞ VERİLERDE ORTANCA-MEDYAN

Sınıflandırılmamış verilerde ortanca (medyan), küçükten büyüğe sıralanmış veriler için en ortadaki yani verileri iki eşit parçaya bölen değerdir. Gözlem sayısının tek ya da çift olmasına göre verileri sıraladıktan sonra aşağıdaki gibi hesaplanır:

n tek ise

$$\frac{n+1}{2}.\,gözlem\ değeri$$

n çift ise

$$\frac{\frac{n}{2}.\,gözlem\ değeri+\left(\frac{n}{2}+1\right).\,gözlem\ değeri}{2}$$

Örnek: Bir ilaç şirketi, üretmiş olduğu ağrı kesicilerin, ilaç alındıktan kaç dakika sonra etkili olduğu ile ilgili bir araştırmada rasgele seçilen 12 kişinin ilaç aldıktan kaç dakika sonra rahatladıklarını kayıt ediyor.

13.0	12.9	13.2	12.7	13.1	13.0	13.1	13.0	12.6	13.1	13.0	13.1
Verilenmiş sıralanmış hali											
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
12.6	12.7	12.9	13.0	13.0	13.0	13.0	13.1	13.1	13.1	13.1	13.2

Gözlem sayısı n=12 çift olduğu için 6. ve 7. Gözlemlerin ortalaması bu örneklemin ortancası olacaktır. Yani $\frac{13+13}{2} = 13\ ortancadır$.

Örnek : Zayıflama programına başvuran 15 kişinin başlangıç kiloları verilmiştir. Bu verilerin dal yaprak grafiği aşağıda verilmiştir. Ortancayı(medyanı) bulunuz.

Dal	Yaprak
5	3
6	5 8 9
7	0 2 2 9
8	3 4 5 7 9
9	0 4

MEDYAN=$\overline{X}' = \frac{(n+1)}{2}.\,gözlem\ değeri$

$n = 15 \qquad \frac{(n+1)}{2} = \frac{16}{2} = 8.\,gözlem\ değeri$

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
53	65	68	69	70	72	72	79	83	84	85	87	89	90	94

SINIFLANDIRILMIŞ VERİLERDE ORTANCA-MEDYAN

Sınıflandırılmış verilerde ortanca birikimli sıklık değerleri kullanılarak bulunur. Gözlemlerin %50'lilik değerine karşılık gelen değeri ortanca olacaktır. Ortanca aşağıda verilen formülle hesaplanır:

$\sum_{i=1}^{m-1} f_i$: ortancanın bulunduğu sınıfa kadar olan sınıfların sıklıklarının toplamı

$$\text{MEDYAN}=\bar{X}' = AS + c \times \frac{n \times \%50 - \sum_{i=1}^{m-1} f_i}{f_m}$$

f_m: Ortancanın bulunduğu sınıfın sıklığı

AS: Ortancanın bulunduğu sınıfın alt sınırı

c: sınıf aralığı

Örnek: 30 öğrenciye ait dönem sonu akademik ortalama puanlarına ilişkin sıklık çizelgesi aşağıda verilmiştir.

c=0.7

AS	ÜS	fi	Fi
0.5	1.19	3	3
1.2	1.89	5	8
1.9	2.59	8	16
2.6	3.29	11	27
3.3	3.99	3	30
		30	

$$\sum_{i=1}^{m-1} f_i = \sum_{i=1}^{2} f_i = 8$$

Ortancanın bulunduğu sınıfın sıklığı $\Rightarrow f_m = f_3 = 8$

n×%50=30×%50=15

MEDYAN=$\overline{X}' = AS + c \times \frac{n\times\%50-\sum_{i=1}^{m-1} f_i}{f_m} = 1.9 + 0.7 \times \frac{15-8}{8} = 2.5125$

Örnek: 25 öğrenciye ait İstatistik dersinden alınan notların sıklık çizelgesi aşağıda verilmiştir.

c=10

Notlar	fi	Fi
40-49 arası	2	2
50-59 arası	2	4
60-69 arası	6	10
70-79 arası	$\boxed{10}$	20
80-90 arası	5	25
	25	

$$\sum_{i=1}^{m-1} f_i = \sum_{i=1}^{3} f_i = 10$$

Ortancanın bulunduğu sınıfın sıklığı $\Rightarrow f_m = f_4 = 10$

n×%50=25×%50=12.5

MEDYAN=$\overline{X}' = AS + c \times \frac{n\times\%50-\sum_{i=1}^{m-1} f_i}{f_m} = 70 + 10 \times \frac{12.5-10}{10} = 72.5$

MEDYAN'IN ÖZELLİKLERİ

Verilerde yer alan uç değerlerden etkilenmez.

Ortalamaya karşı daha duyarlı bir konum ölçüsüdür.

Sağlam (Robust) bir konum ölçüsü olarak bilinir.

Açık uçlu sıklık çizelgeleri için en uygun konum ölçüsüdür (medyan gelir gibi)

DİĞER SAĞLAM ORTALAMALAR

- Kesilmiş Ortalama (Trimmed mean)
- Winsorized ortalama
- Huber’in M ortalaması
- Tukey’in Biweight ortalaması
- Hampel’in M ortalaması
- Andrew’ın Wave Ortalaması

En sık kullanılanlar «Ortanca-Medyan» ve «Kesilmiş ortalama» dır .

Kesilmiş (trimmed) Ortalama

Veri kümesi içinde verilerin genel yapısından ayrılış gösteren uç değerler yer alıyorsa bu uç değerlerin en büyük ve en küçük belli bir yüzdesi kesilerek geri kalan veriler üzerinden aritmetik ortalama hesaplanır. Bu ortalamaya % a'lık kesilmiş ortalama denir. «a» verilerin uçlardan yüzde kaçının kesildiğini gösterir.

1	…	g	…	…	s	…	n

Kesilen gözlemler

Kesilen gözlemler

$$\bar{X}_{Kesilmiş} = \sum_{i=g}^{s} \frac{X_i}{n - 2(\ \%a \times n)}$$

Örnek: Bir sınıftaki 20 öğrencinin notları verilsin.

38, 55, 3,2,9,45,55,99,100,76,67,36,66,75,77,80,16,45,78,80.

%10 Kesilmiş ortalama kaçtır?

20 verinin %10'u 20x0.10=2 olacağından, dağılımın başından ve sonundan 2'şer gözlem kesilir.

{2,3} 9,16,36,38,45,45,55,55,66,67,75,76,77,78,80,80 {99,100}

Geri kalan gözlemlerin ortalaması, $\bar{X} = \frac{898}{16} = 56.125$ elde edilir.

%30 Kesilmiş ortalama kaçtır?

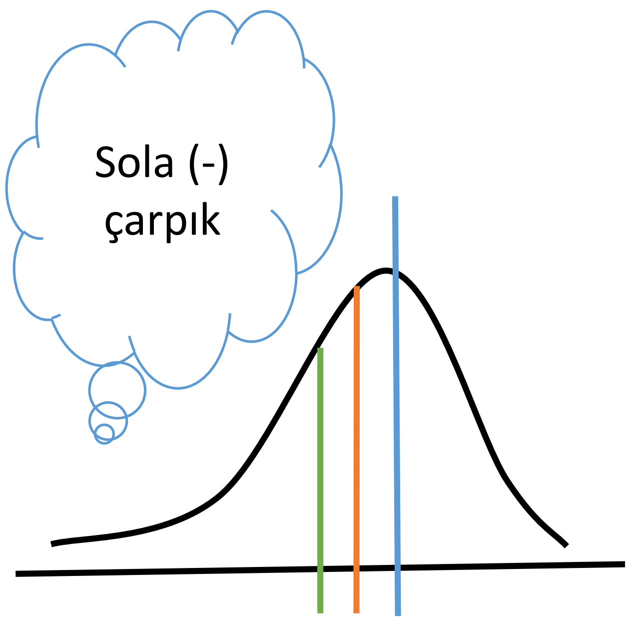
20 verinin %30'u 20x0.30=6 olacağından, dağılımın başından ve sonundan 6'şar gözlem kesilir.

{2,3,9,16,36,38} 45,45,55,55,66,67,75,76 {77,78,80,80,99,100}

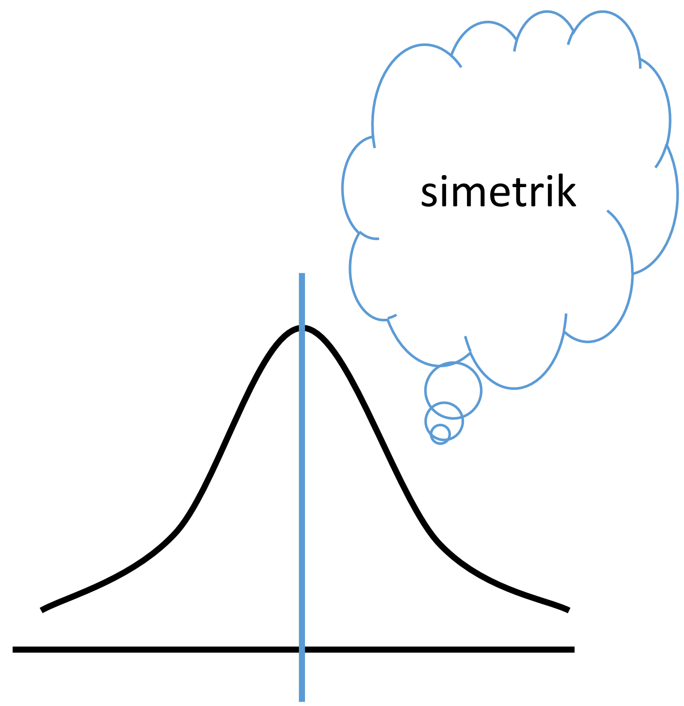
Geri kalan gözlemlerin ortalaması, $\bar{X} = \frac{484}{8}$=60.5

ORTALAMA, ORTANCA VE TEPE DEĞERI ARASINDAKI BAĞINTI

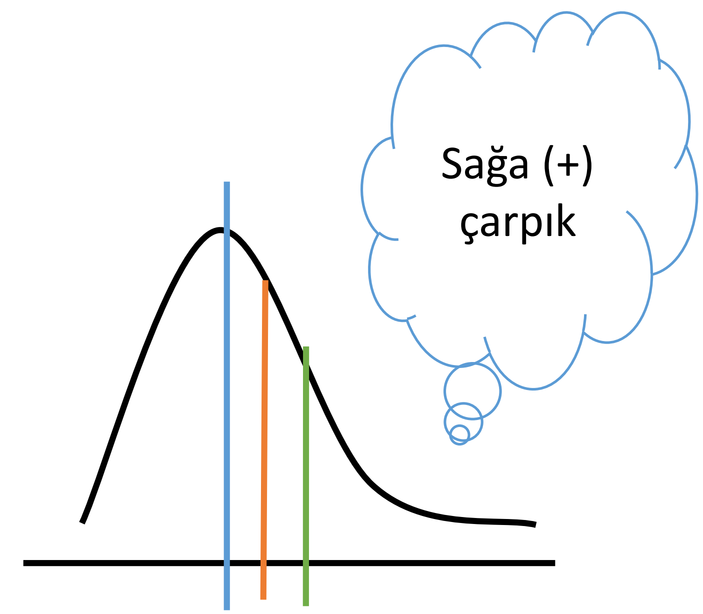
- **Ortalama < Ortanca < Tepe değeri** ise, dağılım sola çarpık (eksi - yöne çarpık) bir dağılımdır.
- **Ortalama = Ortanca = Tepe değeri** ise, dağılım simetriktir.
- **Ortalama > Ortanca > Tepe değeri** ise, dağılım sağa çarpık (artı + yöne çarpık) bir dağılımdır.



Ortalama < Ortanca < Tepe değeri

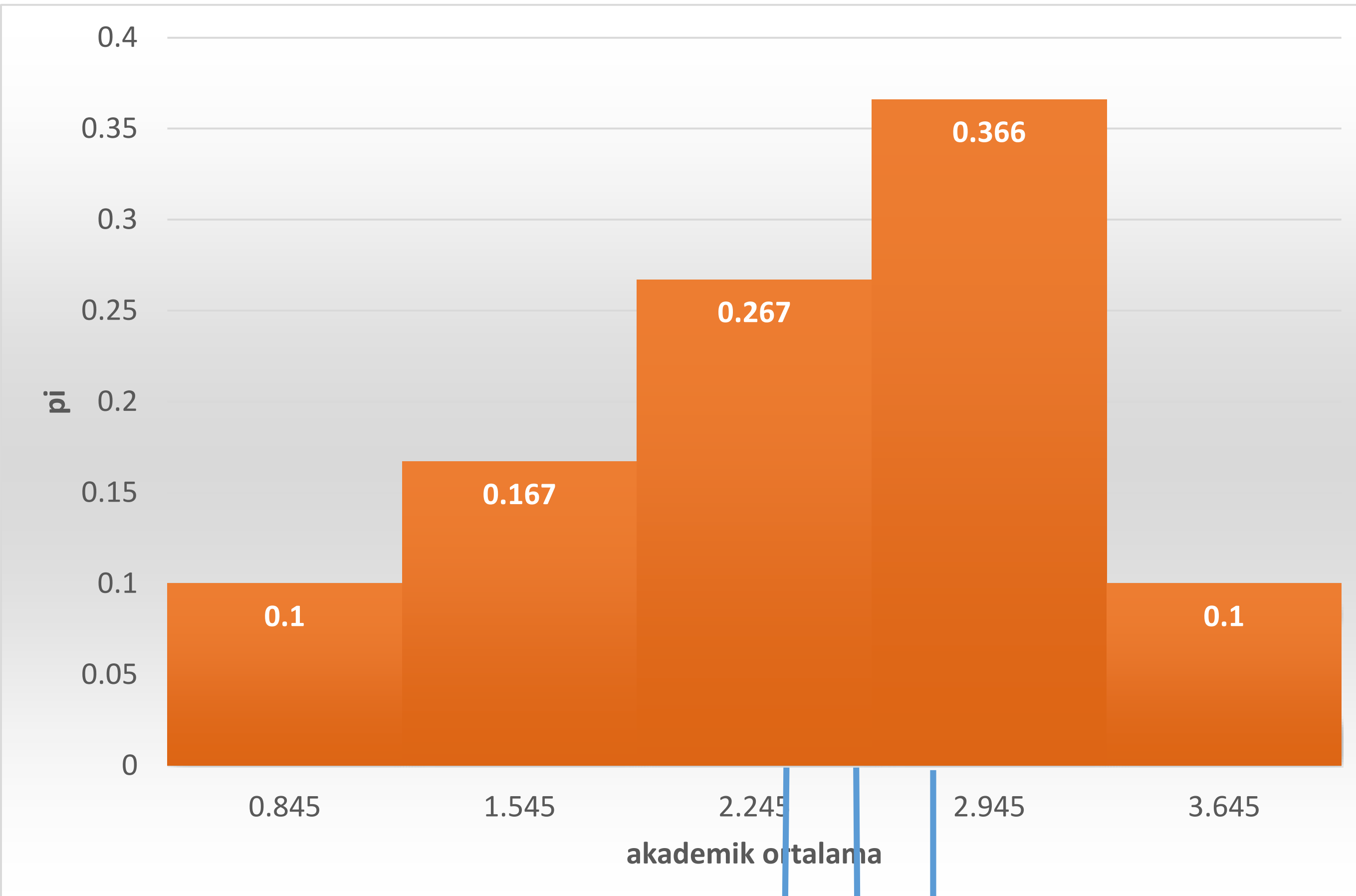


Ortalama = Ortanca = Tepe değeri



Ortalama > Ortanca > Tepe değeri

Örnek: 30 öğrenciye ait dönem sonu akademik ortalama puanlarına ilişkin ortalama, ortanca, tepe değerini kullanarak dağılımı biçimi hakkında ne söylenebilir.



Ortalama=2.385
Ortanca=2.5125
Tepe değeri=2.79

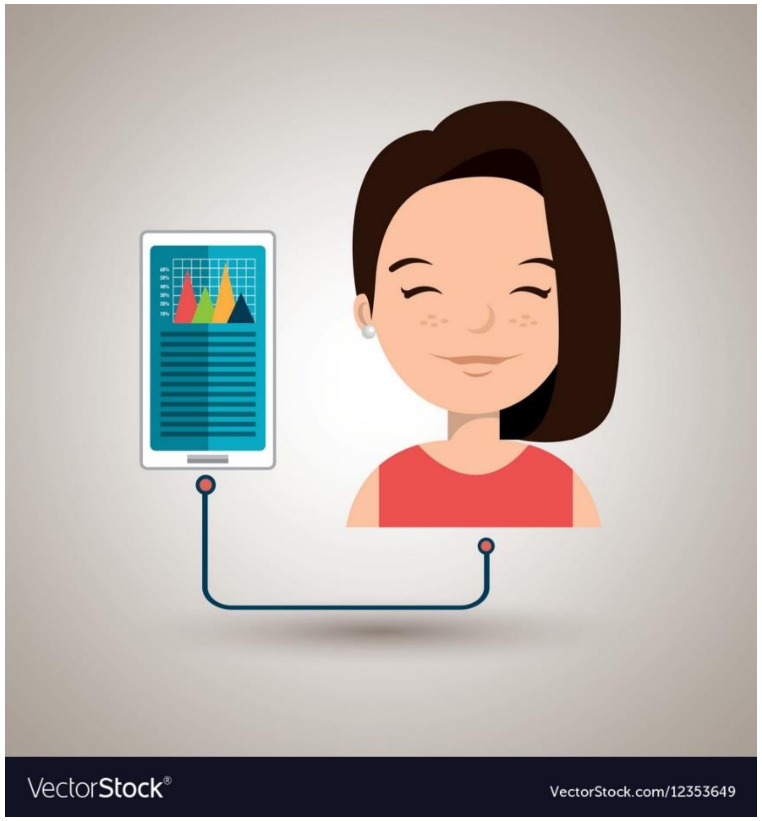
Ortalama< Ortanca < Tepe değeri

Sola (-) çarpık

Tepe değeri:2.79

Ortanca:2.5125

Ortalama:2.385



Bir sonraki derste «Konum ölçüleri» devam...

KAYNAKLAR

1-) H.Demirhan, C.Hamurkaroğlu ,“İstatistiksel Yöntemlere Giriş”, H.Ü.Yayınları, 2011.
2-) Serpil Cula, Zehra Muluk, “Temel İstatistik Yöntemler”, Başkent Üniversitesi yayınları,2006.
3-) Levent Özbek, Esin Köksal Babacan, ‘’İstatistiğe Giriş’’, TÜBİTAK e-kitap.
4-) Birdal Şenoğlu, Mehmet Yılmaz, Sibel Açık Kemaloğlu, İstatistiğe Giriş, TÜBİTAK e-kitap.