Bölüm 3

Tanımlayıcı İstatistikler

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/Çeyrekler/Yüzdelikler

Çarpıklık Ölçüleri

Basıklık Ölçüleri

5) Medyan

- Bir veri setini büyükten küçüğe veya küçükten büyüğe sıraladığımızda tam orta noktadan veri setini iki eşit parçaya ayıran değere **medyan** adı verilir.

- Veri setinde aşırı uçlu elemanlar olduğunda aritmetik ortalamaya göre daha güvenilirdir.

- Medyan, veri setindeki tüm elemanlardan etkilenmez.

Basit Veriler İçin Medyan

- **Veri Setinin Hacmi Tek Sayı İse;**

$$\frac{n+1}{2}$$ ***nci gözlem değeri medyandır.***

- **Veri Setinin Hacmi Çift Sayı İse;**

$$\frac{n}{2} \quad \textit{ve} \quad \frac{n}{2}+1$$ ***nci gözlem değerinin aritmetik ortalaması medyandır.***

5.40	1.10	0.42	0.73	0.48	1.10
0.42	0.48	0.73	1.10	1.10	5.40

Medyan bu iki noktanın arasına düşmektedir

$$\frac{0.73 + 1.10}{2}$$

MEDYAN 0.915

5.40	1.10	0.42	0.73	0.48	1.10	0.66
0.42	0.48	0.66	0.73	1.10	1.10	5.40

Tam ortadaki değer medyandır.

MEDYAN 0.73

Gruplanmış Veriler İçin Medyan

- Gruplanmış verilerde medyan değeri hesaplanırken veri setinin tam orta noktasının hangi gruba ait olduğunu belirlemek için birikimli frekans sütunu oluşturulur.

- Sıra numarası belirlendikten sonra o sıra numarasına ait grup medyan değeri olarak ifade edilir.

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

Kg	Satış adedi	***Birikimli Frekans (∑f)***
1	5	***5***
2	12	***17***
3	35	***52***
4	14	***66***
5	8	***74***
6	6	***80***

- n/2 ve (n/2)+1 nci gözlem değerlerine karşılık gelen değerler (40 ve 41 nci sıra) 3 olduğundan dolayı medyan değeri 3’tür.

•Frekans dağılımı aşağıdaki gibi olsaydı (n+1)/2 nci elemana (**45** nci elemana) karşılık gelen değer **4** olacağından dolayı veri setinin medyanı **4** olarak hesaplanacaktır.

Kg (x)	Satış adedi (f)	***Birikimli Frekans (∑f)***
1	5	***5***
2	12	***17***
3	22	***39***
4	32	***71***
5	14	***85***
6	4	***89***

Sınıflanmış Veriler İçin Medyan

- Sınıflanmış verilerde medyan değeri hesaplanırken ilk olarak medyan sınıfı belirlenir.
- Medyan sınıfı birikimli frekanslar dikkate alındığında toplam frekansın yarısını içinde bulunduran sınıftır.
- Medyan sınıfı belirlendikten sonra medyan sınıfından bir önceki sınıfın birikimli frekansı ve medyan sınıfı frekansı dikkate alınarak hesaplanır.

$$Medyan = L_{med} + \frac{\frac{\sum f_i}{2} - f_l}{f_{med}} .i$$

L_{med} : Medyan sınıfının alt sınırı

f_l : Medyan sınıfından bir önceki sınıfın birikimli frekansı

f_{med} : Medyan sınıfının frekansı

i: Sınıf Aralığı

Örnek: Bir sınıftaki öğrencilerin boyları hakkında bir araştırma yapılmaktadır. Bu amaçla 50 öğrencinin boyları ölçülerek kaydedilmiştir.Öğrencilerin boylarının mod değerini hesaplayınız.

$$Medyan = L_{med} + \frac{\frac{\sum f_i}{2} - f_l}{f_{med}} .i$$

Sınıflar	f_i	$\sum f_i$
150-157'den az	5	5
157-164'den az	7	12
164-171'den az	14	26
171-178'den az	9	35
178-185'den az	8	43
185-192'den az	4	47
192-199'dan az	3	50
Toplam	50	

Lmed=164; Tf/2=25; fl=12; fmed=14; i=7

Toplam 50 adet gözlem olduğundan dolayı, birikimli frekans sütununda 50/2 =25 nci gözlemin bulunduğu sınıf medyan sınıfı olarak belirlenir.

$$Medyan = L_{med} + \frac{\frac{\sum f_i}{2} - f_l}{f_{med}}.i$$

$$= 164 + \frac{25-12}{14}.7 = 170,5cm$$

Merkezi Ölçüm	Tanım	Nasıl Kullanılıyor	Varlığı	Her değer Dikkate Alınırmı?	Uç Değerlerden Etkilenirmi?	Avantajları ve Dezavantajları
Ortalama	$\bar{x} = \frac{\sum x}{n}$	En Bilinen 'ortalama'	Her zaman vardır.	Evet	Evet	Birçok istatistiksel metodla iyi çalışır.
Medyan	Orta değer	Sıklıkla Kullanılır	Her zaman vardır.	Hayır	Hayır	Birkaç uç değer varsa genellikle iyi bir tercihtir
Mod	En sık tekrar eden veri değeri	Ara sıra kullanılır	Olmayabilir ya da birden fazla olabilir.	Hayır	Hayır	Nominal düzeyde veriler için uygundur

Veriler mod etrafında simetrik oldukları zaman, mod, medyan ve artimetik ortalama birbirlerine eşit olur.

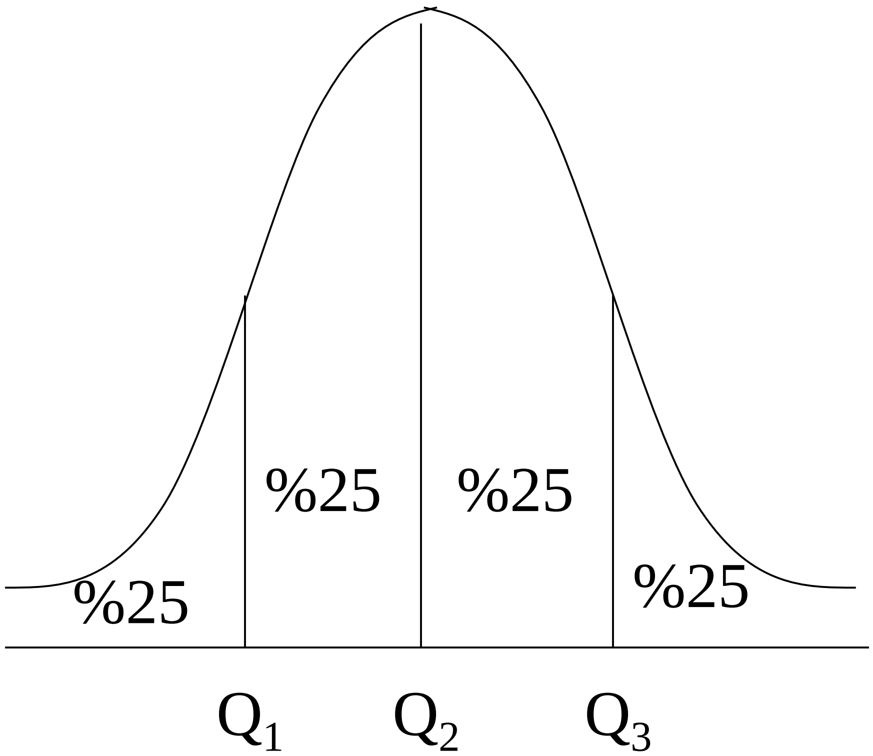
Eğer örneklem aynı anakütleden çekilmişse, aritmetik ortalama diğer ölçülere göre daha güvenilirdir

Kartiller (Yüzdelikler)

•Bir veri setini büyükten küçüğe veya küçükten büyüğe sıraladığımızda dört eşit parçaya ayıran üç değere **kartiller** adı verilir.

•İlk % 25'lik kısmı içinde bulunduran 1. Kartil (Q_1), % 50'lik kısmı içinde bulunduran 2. Kartil (Q_2), % 75'lik kısmı içinde bulunduran 3. Kartil (Q_2), olarak adlandırılır.

•*%50'lik kısmı içinde bulunduran 2. Kartil (Q_2) aynı zamanda veri setinin medyanıdır.*



Kartiller (Çeyrekler) Arası Fark

- Diğer değişkenlik ölçülerinden biri olan çeyrekler arası fark ortalamadan ne kadar sapma olduğunu göstermektedir.

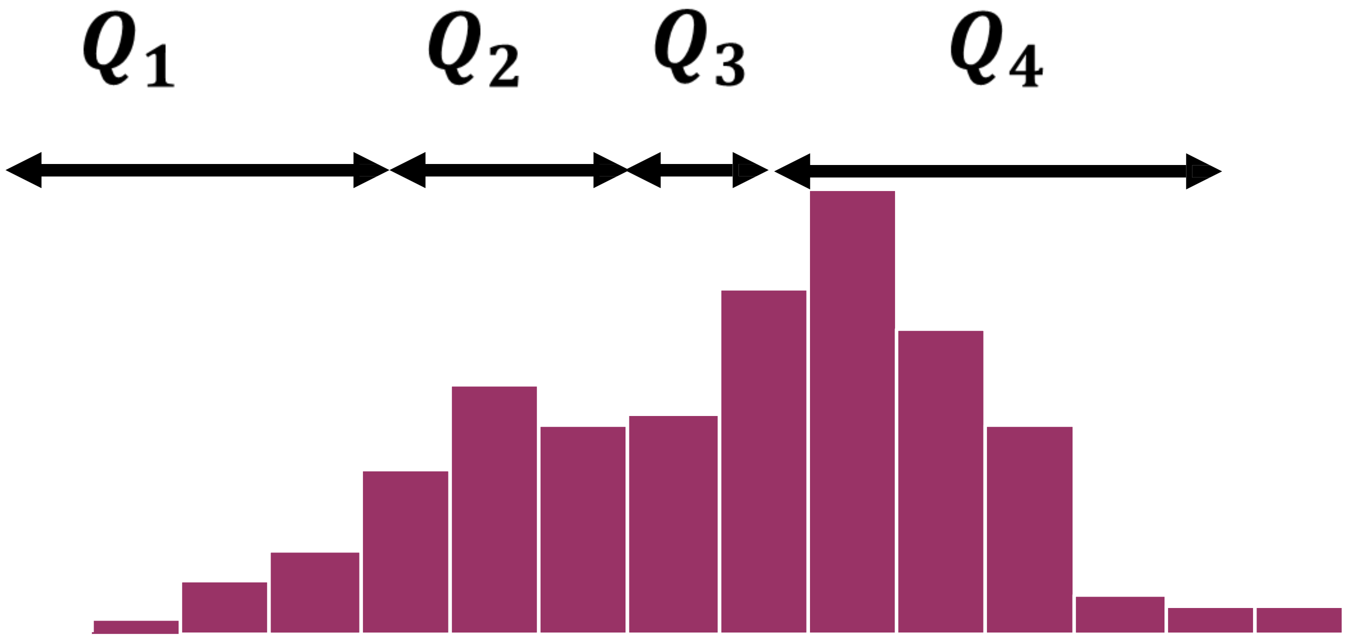
- 1. ve 3. kartiller arasındaki farka dikkat çeker.

- Çeyrek aralık (çeyrekler arası açıklık) olarak adlandırılan bu fark, Q_3-Q_1, veri setinin yarısını içeren genişliği verir.

Çeyrek sapma…

$$Ç.A.A = Q_3 - Q_1$$

$$Ç.K = \frac{Q_3 - Q_1}{2}$$



Çeyrekler

Dağılımı 4 eşit parçaya bölen değerlerdir. Bunlar,

1. Çeyrek (Ç1)	2. Çeyrek (Ç2)	3. Çeyrek (Ç3)
Değerlerin %25'i Ç1'e eşit ya da ondan küçüktür.	***Değerlerin %50'si Ç2'ye eşit ya da ondan küçüktür. Bu değer aynı zamanda ortancadır.***	***Değerlerin %75'i Ç3'e eşit ya da ondan küçüktür.***

Basit Veriler İçin Kartiller

- **1.Kartil Q_1**

$$\frac{n+1}{4}$$

nci gözlem değeri,

- **3.Kartil Q_3**

$$\frac{3(n+1)}{4}$$

nci gözlem değeri,

$$\varphi(\boldsymbol{fi}) = \frac{\boldsymbol{Y_{75}} - \boldsymbol{Y_{25}}}{\boldsymbol{2}}$$

Örnek: İstatistik I dersini alan 10 öğrencinin vize notları aşağıdaki gibi sıralanmıştır. Buna göre vize notları için Q_1 ve Q_3 değerlerini hesaplayınız.

30,42,56,61,68,79,82,88,90,98

N=10

Q1=(10+1)/4=2,75. gözlem değeri

Q3=3*(10+1)/4=8,25. gözlem değeri

Q1: 4 parça olması için; (56-42)/4=3,50

42,	45.5,	49,	52.5,	56
2.gözlem	2,25. gözlem	2.5. gözlem	2.75. gözlem	3.gözlem
			Q1	

Veri seti aşağıdaki gibi verilseydi,

30,42,56,61,**68**,79,82,88,90

(n+1)/4 ‘ncü verinin sıra numarası (9+1)/4 = 2,5’dir.

Q_1= 42 + 0, 5 .(56 - 42) = 49 ,

3(n+1)/4 ‘ncü verinin sıra numarası 3(9+1)/4 = 7,5’dir.

Q_3= 82 + 0, 5.(88 - 82) = 85 ,

olarak hesaplanacaktı.

Aşağıdaki veri, her çocuğun beslenme çantasındaki balık krakerlerin sayısını temsil ediyor.

Veriyi küçükten büyüğe sıralayınız.

~~4~~	~~4~~	~~10~~	~~11~~	~~15~~	~~7~~	~~14~~	~~12~~	~~6~~

Veri için çeyrekler açıklığını hesaplayınız.

Çeyrekler açıklığı = $13 - 5$

$= \boxed{8}$

5 13

$$4, \boxed{4, 6}, 7, \boxed{10}, 11, \boxed{12, 14}, 15$$

$$\frac{4+6}{2} = \frac{10}{2} = \boxed{5}$$

$$\frac{12+14}{2} = \frac{26}{2} = 13$$

Sıra sizde…

Aşağıdaki nokta grafiği için çeyrekler açıklığını hesaplayınız.

[ ] şarkılar

Sadık'ın koleksiyonundaki albümlerin şarkı sayısı

Sıralı: **7, 9, 9, 10, 10, 10, 11, 12, 12, 14** **n=10**

Q1: 11/4=2.75. gözlem değeri =9 olur.

Q3: 3*11/4=8.25. gözlem değeri =12 olur.

Gruplanmış Veriler İçin Kartiller (Çeyrekler=Quartil=Dördebölenler)

- Gruplanmış verilerde kartiller hesaplanırken veri setinin ilk çeyrek ve son çeyrek kısmını tam olarak ifade etmek amacıyla birikimli frekans sütünü oluşturulur.

- Gruplanmış verilerde örnek hacminin tek veya çift olduğuna bakılmaksızın

n/4 ncü eleman 1.Kartil (Q_1),

3n/4 ncü eleman ise 3.Kartil (Q_3),

olarak ifade edilir.

Örnek: Bir un fabrikasının satış mağazasında bir gün içinde satılan un paketlerinin gramajlarına göre göre satış adetleri aşağıda verilmiştir. Buna göre veri seti için Q_1 ve Q_3 nedir?

Kg	Satış adedi	***Birikimli Frekans (∑f)***
1	5	***5***
2	12	***17***
3	35	***52***
4	14	***66***
5	8	***74***
6	6	***80***

- n/4 ncü (20 nci) sıra numarasına karşılık gelen gözlem 3 olduğundan; 1.kartil 3, 3n/4 ncü (60 ncı) sıra numarasına karşılık gelen gözlem 4 olduğundan; 3.kartil 4’tür.

Sınıflanmış Veriler İçin Kartiller

- Sınıflanmış verilerde kartiller hesaplanırken ilk olarak birikimli frekans sütunu oluşturularak kartil sınıfları belirlenir.

- Kartil sınıfları belirlenirken gruplanmış verilerde olduğu gibi n/4 ve (3n)/4 ncü sıralardaki elemanların hangi sınıflara ait iseler o sınıflar kartil sınıfları olur.

- Kartil sınıfları belirlendikten sonra bu sınıflardan bir önceki sınıfın birikimli frekansı ve mevcut sınıf frekansı dikkate alınarak kartil değerleri hesaplanır.

1. Kartil

$$Q_1 = L_{Q_1} + \frac{\frac{\sum f_i}{4} - f_l}{f_{Q_1}}.i$$

2. Kartil

$$Q_2 = Medyan = L_{Q_2} + \frac{\frac{\sum f_i}{2} - f_l}{f_{Q_2}}.i$$

3. Kartil

$$Q_3 = L_{Q_3} + \frac{\frac{3\sum f_i}{4} - f_l}{f_{Q_3}}.i$$

Örnek: Bir sınıftaki öğrencilerin boyları hakkında bir araştırma yapılmaktadır. Bu amaçla 50 öğrencinin boyları ölçülerek kaydedilmiştir.Öğrencilerin boylarının birinci ve üçüncü kartillerini hesaplayınız.

Q_1 sınıfı

Q_3 sınıfı

Sınıflar	f_i	$\sum f_i$
150-157'den az	5	5
157-164'den az	7	12
164-171'den az	14	26
171-178'den az	9	35
178-185'den az	8	43
185-192'den az	4	47
192-199'dan az	3	50
Toplam	50	

$$L_{Q_1} + \frac{\frac{\sum f_i}{4} - f_l}{f_{Q_1}}.i \qquad Q_3 = L_{Q_{31}} + \frac{\frac{3\sum f_i}{4} - f_l}{f_{Q_3}}.i$$

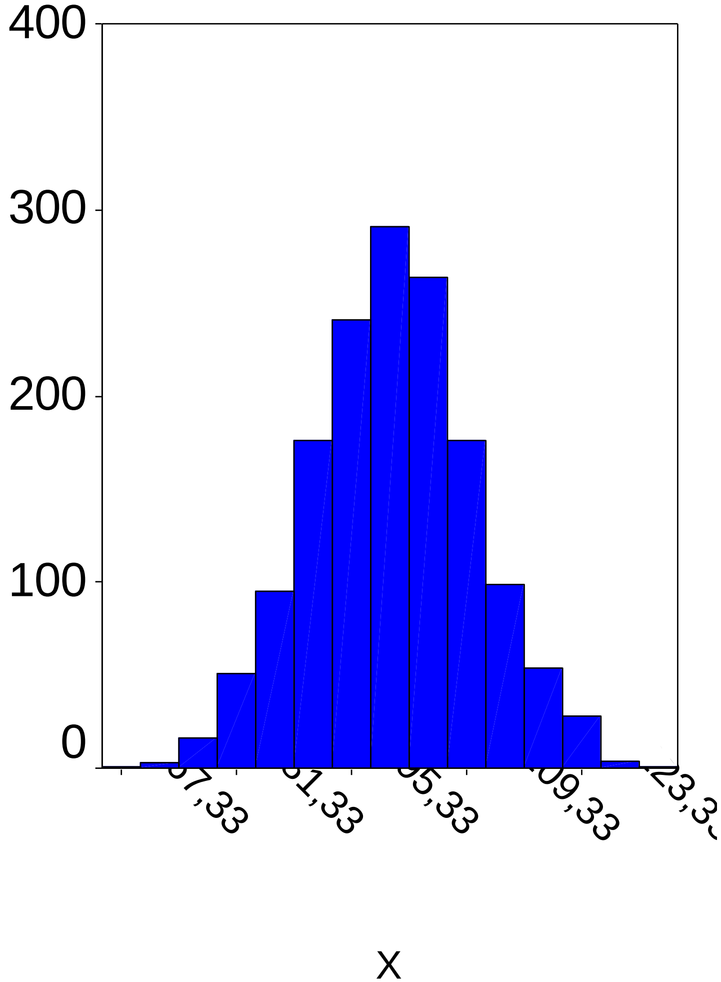
$$164 + \frac{12{,}5 - 12}{14}.7 = 164.25 \qquad 178 + \frac{37{,}5 - 35}{8}.7 = 180.19$$

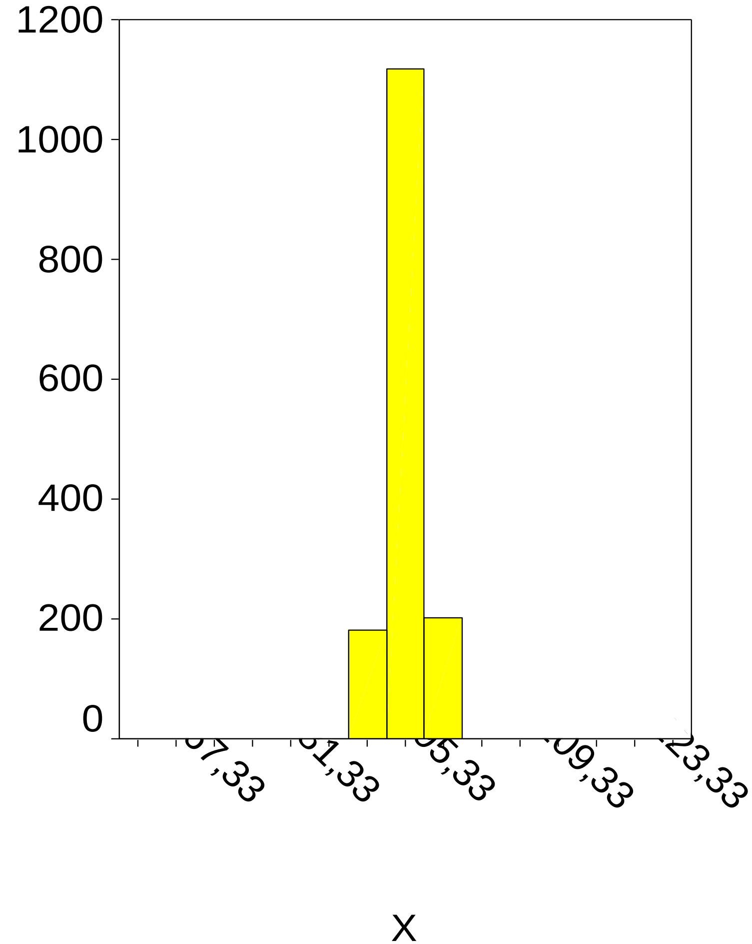
Yayılma (Değişkenlik) Ölçüleri

- Bir veri setini tanımak yada iki farklı veri setini birbirinden ayırt etmek için her zaman yalnızca yer ölçüleri yeterli olmayabilir.

- Dağılımları birbirinden ayırt etmede kullanılan ve genellikle aritmetik ortalama etrafındaki değişimi dikkate alarak hesaplanan istatistiklere yayılma (değişkenlik) ölçüleri adı verilir.

Aşağıdaki iki grafik n = 1500 hacimli iki farklı örnek doğrultusunda oluşturulan histogramlardır. Her iki örnek ortalaması yaklaşık olarak 100 olduğuna göre iki örneğin aynı anakütleden alındığı söylenebilir mi?





- Dağılımları birbirinden ayırt etmede kullanılan **yayılım ölçüleri** aritmetik ortalama etrafındaki değişimleri dikkate alan tanımlayıcı istatistiklerdir.

- **Bir veri setinde aritmetik ortalamalardan her bir gözlemin farkı alınıp bu değerlerin tümü toplandığında sonucun 0 olduğu görülür**.

- **Örnek: 4,8,9,13,16** şeklinde verilen bir basit veri için;

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4+8+9+13+16}{5} = 10$$

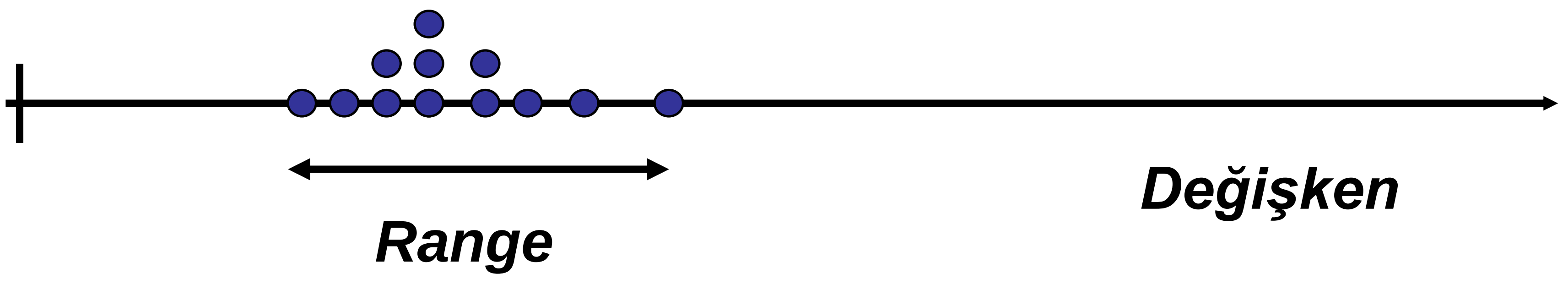
$$\sum_{i=1}^{n} \left(x_i - \bar{x}\right) = (4-10)+(8-10)+(9-10)$$

$$= (13-10)+(16-10) = 0$$

- ***Bu örnekten görüleceği üzere gözlemlerin aritmetik ortalamadan uzaklığı alıp toplandığında 0 elde edildiğinden dolayı bu problem mutlak değer kullanarak veya karesel uzaklık alınarak ortadan kaldırılır.***

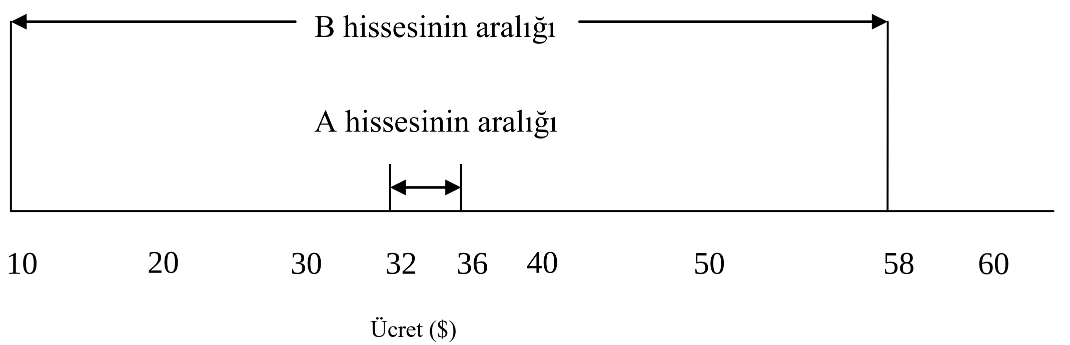
7) Range (Değişim Aralığı)

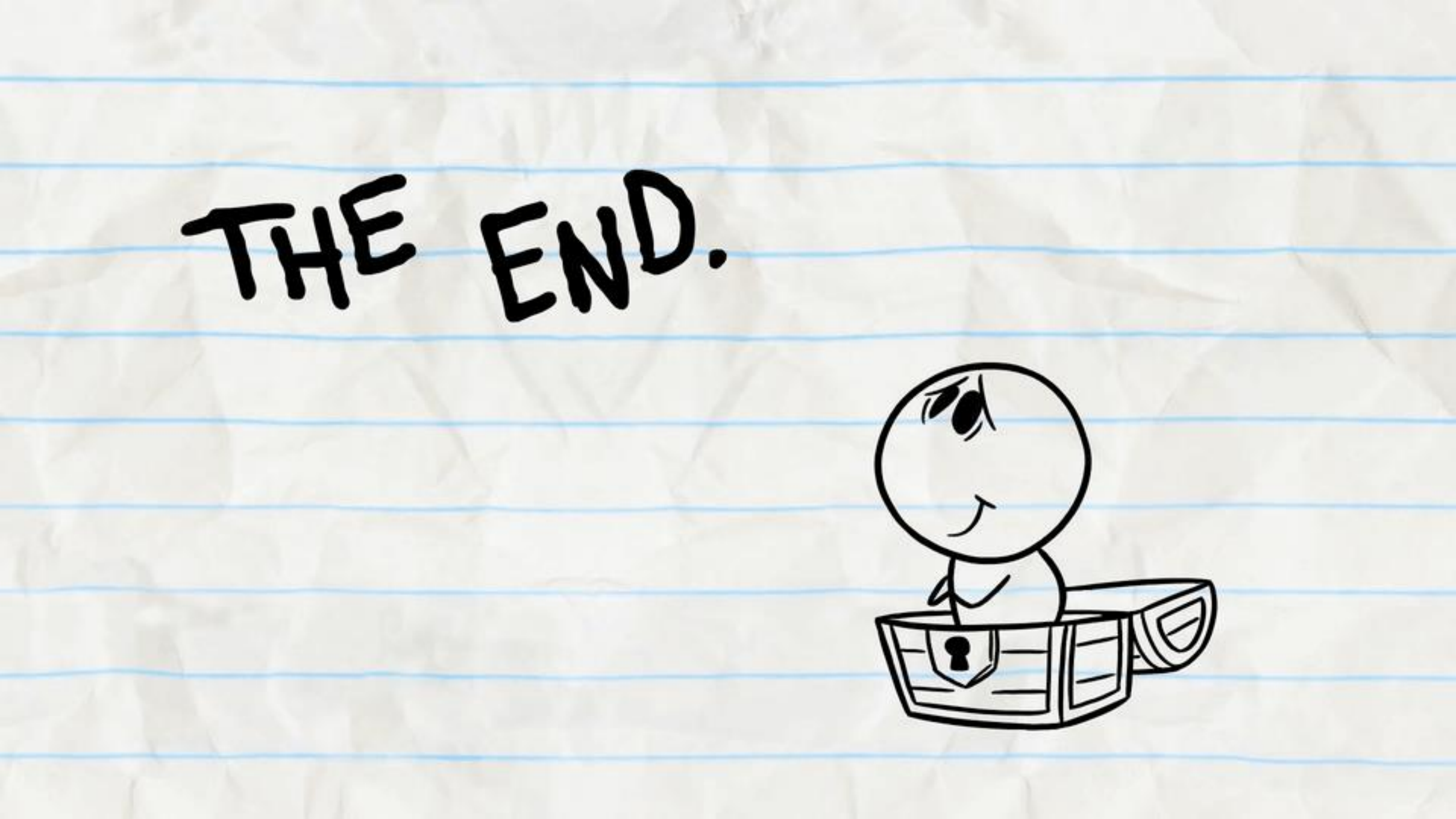
- Veri setindeki yayılımı ifade etmede kullanılan en basit ölçü, değişim aralığıdır. Genel olarak az sayıda veri için kullanılır.
- En büyük gözlem değeri ile en küçük gözlem değeri arasındaki fark değişim aralığını verir.
- Range, veri setindeki tek bir gözlemin aşırı derecede küçük veya büyük olmasından etkilendiği için bir başka ifadeyle örnekte yer alan sadece iki veri kullanılarak hesaplanmasından dolayı tüm veri setinin değişkenliğini açıklamak için yetersiz kalmaktadır.


Range
Değişken

Değişim Aralığı

Örnek:
Aralık, veri seti içindeki en büyük değerle en küçük değer arasındaki uzaklığı ölçerek verinin yayılımını ortaya koyar. Örneğin aşağıdaki şekilde gösterildiği üzere A hisse senedi belirli bir yılda 36$ ila 32$ arasında çeşitlilik gösterirken, B hisse senedi 10$ ila 58$ arasında gösterdi. Hisse senedinin fiyatındaki aralık A için 36-32 = 4$ dır; B için 58-10=48$.Aralıkları kıyasladığımızda B hisse senedinin fiyat aralığının A ya göre daha çok değişkenlik gösterdiğini söyleyebiliriz.





Dinlediğiniz İçin Teşekkür Ederim…

Dr. Gökhan AKSU