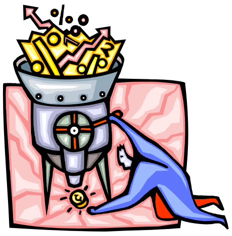
TANIMLAYICI İSTATİSTİKLER

MERKEZİ EĞİLİM ÖLÇÜLERİ



Dr. Mehmet AKSARAYLI
D.E.Ü. İ.İ.B.F.
EKONOMETRİ BÖLÜMÜ
mehmet.aksarayli@deu.edu.tr

Tanımlayıcı İstatistikler

Yer Ölçüleri (Merkezi Eğilim Ölçüleri)

- Duyarlı Ortalamalar
  - Aritmetik ort.
  - Tartılı Aritmetik
  - Geometrik ort.
  - Kareli ort.
  - Harmonik ort.
- Duyarlı Olmayan Ort.
  - Mod
  - Medyan
  - Kartiller

Değişkenlik Ölçüleri

- Değişim Aralığı (Range)
- Standart sapma
- Varyans
- Mutlak sapma
- Değişkenlik katsayısı
- Kartil sapma katsayısı
- Ortalama sapma katsayısı

Çarpıklık Ölçüleri

- Bowley asimetri ölçüsü
- Pearson asimetri ölçüsü

Basıklık Ölçüleri

Aritmetik Ortalama

- Üzerinde inceleme yapılan veri setindeki elemanların toplanıp incelenen eleman sayısına bölünmesiyle elde edilen yer ölçüsüne aritmetik ortalama denir.
- Halk dilinde ortalama ifadesi kullanıldığında ilk akla gelen kavram aritmetik ortalamadır.

- Örnek:
  - Sınav notlarının ortalaması,
  - Yaz aylarında m^2'ye düşen ortalama yağış miktarı

Yer Ölçüleri: Aritmetik – Ortalama - Mod - Medyan

Basit Veriler İçin

Yaş: 24, 26, 20, 18, 24

Aritmetik Ortalama = $\overline{X} = \dfrac{\sum X_i}{n}$ =(24+26+20+18+24)/5 =22.4

18, 20, 24, 24, 26

Medyan=24

Mod= 24

Yer Ölçüleri: Aritmetik – Ortalama - Mod - Medyan

Gruplandırılmış Veriler İçin

10 bayanın ayakkabı numaraları:

35, 38, 36, 36, 37, 36, 38, 35, 39, 37,

35, 35, 36, 36, 36, 37, 37, 38, 38, 39,

f : frekans

k: grup sayısı

i = 1,2,3,……….,k

Grup	Frekans	$X_i f_i$
35	2	70
36	3	108
37	2	74
38	2	76
39	1	39
Σ	10	367

$$\overline{X} = \frac{\sum X_i f_i}{\sum f_i} = 367/10 = 36.7$$

Medyan = (10+1)/2=5,5. Veri = 36.5

Mod=en çok tekrar eden veri=36

Sınıflanmış Seriler İçin Aritmetik Ortalama

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

f : frekans

k : sınıf sayısı

i = 1,2,3,……….,k

m : sınıf orta noktası

$$\sum_{i=1}^{k} f_i = n$$

- Sınıflanmış serilerde her bir sınıf içindeki değerlerin neler olduğu bilinmediğinden ve yalnızca her bir sınıfın frekans değerleri bilindiğinden dolayı sınıfı temsil etmek üzere sınıf orta noktaları hesaplamada kullanılır.
- Kullanılan formül gruplanmış seriler için kullanılan formüle benzerdir.

Örnek: Sınıflandırılmış Veriler İçin Aritmetik ortalama

Sınıflar	f_i	m_i	$f_i.m_i$	Birikimli$f_i = \Sigma f_i$
0 – 10'dan az	12	5	60	12
10 - 20'dan az	6	15	90	18
20 - 30'dan az	4	25	100	22
30 – 40''dan az	2	35	70	24
40 – 50'dan az	1	45	45	25
Σ	25		365	

$$\overline{X} = \frac{\sum m_i f_i}{\sum f_i} = 365/25 \quad =14.6$$

Aritmetik ortalamanın özellikleri:

- Örnek elemanları ortalama etrafında toplanma eğilimdedir yani örneği en iyi temsil eden tek bir simetrik değerdir.
-

$$\left.\begin{aligned} \bar{x} &= \frac{\sum x}{n} \\ n\bar{x} &= \sum x \\ \sum x - n\bar{x} &= 0 \\ \sum x - n\frac{\sum x}{n} &= 0 \\ \sum x - \sum x &= 0 \\ \sum (x-\bar{x}) &= 0 \end{aligned}\right\} \text{Aritmetik ortalamadan sapmaların toplamı sıfırdır.}$$

AĞIRLIKLI ARİTMETİK ORTALAMA:

- İndex sayıların hesaplanmasında, yüzdelerin ortalamasında çarpımların ortalamasının alınmasında kullanılır .

Basit seriler için $\bar{x}_w \frac{\sum wx}{\sum w}$

Frekans verileri için $\bar{x}_w = \frac{\sum wfx}{\sum wf}$ $\bar{x}_w = \Sigma f_i w_i / \Sigma w_i = \frac{f_1\bar{x}_1 + f_2\bar{x}_2 + ...}{f_1 + f_2 + ...}$

$$\bar{X}_w = \frac{x_1 w_1 ....}{w_1} = \frac{\hat{\Sigma} x_i w_i}{\sum w_i} = \frac{f_1\bar{x}_1 + f_2\bar{x}_2 + ...}{f_1 + f_2 + ...}$$

Örnek:

İstatistiki Bölge Birimleri Sınıflandırmasına göre Türkiye toplam 12 bölgeye ayrılmaktadır. Aşağıda bu bölgelere ilişkin 2000 yılı nüfus ve kişi başına düşen GSYİH (YTL) miktarları verilmektedir. Bu verilerden yola çıkarak Türkiye geneline ilişkin ortalama kişi başına düşen GSYİH miktarını bulunuz.

		x	w	wx
	BÖLGE ADI	**GSYİH (1000YTL)**	**Toplam nüfus(1.000.000)**	
1	Kuzeydoğu Anadolu	1.1	2.5	2.75
2	Ortadoğu Anadolu	1.3	3.7	4.81
3	Güneydoğu Anadolu	1.4	6.6	9.24
4	İstanbul	3.7	10.0	37
5	Batı Marmara	2.9	2.9	8.41
6	Ege	3.1	8.9	27.59
7	Doğu Marmara	3.9	5.7	22.23
8	Batı Anadolu	2.8	6.4	17.92
9	Akdeniz	2.5	8.7	21.75
10	Orta Anadolu	1.9	4.2	7.98
11	Batı Karadeniz	2.1	4.8	10.08
12	Doğu Karadeniz	1.7	3.1	5.27
	TOPLAM	**28.4**	**67.5**	**175.03**

Aritmetik ortalama ile hesaplanırsa:

$$\bar{x} = \frac{1.1+1.3+1.4+3.7+2.9+3.1+3.9+2.8+2.5+1.9+2.1+1.7}{12} = \frac{28.4}{12} = 2.367$$

Sakınca nedir?

Ağırlıklı aritmetik ortalama ile hesaplanırsa:

$$\bar{x}_w = \frac{(1.1\times 2.5)+(1.3\times 3.7)+...+(1.7\times 3.1)}{2.5+3.7+...+3.1} = 2.59$$

GEOMETRİK ORTALAMA:

- Örnek veri değerleri çarpımının, örnek hacmi derecesinden köküne eşittir .

$$G = \sqrt[n]{x_1.x_2.x_3......x_n}$$

Özellikleri:

1. $x_i > 0$ olmalıdır.

2.Serideki değerlerin her birinin yerine geometrik ortalama konulduğunda serinin çarpım sonucu değişmez.

2.4.8.16.32 = 32768 = 8.8.8.8.8

3.Geometrik ortalamanın orijinal gözlemlerinin logaritmik sapmaları eşittir. Bu özellikten dolayı ortalama oranlara, değişme oranlarına, logaritmik dağılmış şekiller uygulanır. Örneğin; fiyat indekslerinde geometrik ortalama anlamlı sonuçlar verir.

4.Aritmetik ortalama gerçekte nispi olan değerler yerine mutlak değerlenmiş gibi bir işleme bağlı tutularak çok artan nispi değerleri olduğundan fazla gösterir. Bu yüzden yukarı eğilimlidir.

5.Logaritmik bir dağılımda geometrik ortalamanın tercih nedeni böyle bir dağılımda mutlak sapmaların değil ancak merkezi eğilim etrafında nispi sapmaların simetrik olma eğilimidir.

6.G birimleri değerleri arasındaki orana göre değer alır.

7.Uç değerlerden kadar etkilenmez.

8.log x_1-log G

9. G < $\overline{X}$

10. $\frac{x_1}{G} * \frac{x_2}{G} * .... \frac{x_n}{G} = 1$

11. $x_1.x_2..........x_n = n.G$

- **Basit seri :**

$$G = \left(\prod_{i=1}^{n} x_i \right)^{1/n} = (x_1.x_2.x_3........x_n)^{1/n} \qquad G = \sqrt[n]{x_1.x_2.x_3......x_n}$$

$$\log G = \frac{1}{n}\left[\log x_1 + \log x_2 + ...\log x_n\right] = \frac{1}{n}\left[\sum_{i=1}^{n} \log x_i\right] \qquad G = anti\log \frac{1}{n}\sum_{i=!}^{n} \log x_i$$

Frekans serilerinde ise geometrik serinin hesaplanması:

$$G = \sqrt[\Sigma f]{x_1^{f_1}.x_2^{f_2}........x_n^{f_n}} = \left(x_1^{f_1} x_2^{f_2} ...x_n^{f_n} \right)^{1/\sum f}$$

$$G = \sqrt[N]{\underbrace{x_1 x_1.....x_1}_{f_1 \text{tane}} \underbrace{x_2 x_2.....x_2}_{f_2 \text{tane}} .... \underbrace{x_n x_n.....x_n}_{f_n \text{tane}}} = \sqrt[N]{x_1^{f_1} x_2^{f_2} .... x_n^{f_n}} = \sqrt{x_1^{f_1} x_2^{f_2} ...x_n^{f_n}}$$

$$\log G = \frac{1}{\Sigma f}\left[f_1 \log x_1 + f_2 \log x_2 + ... + f_n \log x_n \right] = \frac{1}{\Sigma f} \sum_{i=1}^{n} f_i \log x_i$$

$$G = anti\log \frac{1}{\Sigma f} \sum_{i=1}^{n} f_i \log x_i$$

Bileşik Faiz Formülü

P_0=başlayış miktarı
r= faiz
n=yıl
P_n=n yıl sonraki meblağ

$$P_n = P_0(1+r)^n \longrightarrow (1+r)^n = \frac{P_n}{P_0} \longrightarrow (1+r) = \sqrt[n]{\frac{P_n}{P_0}}$$

$$\longrightarrow r = \sqrt[n]{\frac{P_n}{P_o}} - 1 \qquad \longrightarrow \log(1+r) = \frac{\log P_n - \log P_o}{n}$$

Örnek: 3 yılda 1000$, 5000$ a artmıştır. Yıllık ortalama artış yüzdesi nedir?

$\frac{\%500}{3}$ gibi gözükse de bu ortalama % artışı doğru değildir.
“r ortalama artış yüzdesini göstermektedir

Başlangıç 1000
1 yıl sonra 1000(1+r)
2 yıl sonra $1000(1+r)^2$
3 yıl sonra $1000(1+r)^3=5000$

$$(1+r)^3 = 5$$

$$r = \sqrt[3]{5} - 1$$

HARMONİK ORTALAMA

Gözlemlerin terslerinin aritmetik ortalamasının tersidir.

H.O belli koşullar altında ve belli fiyat tipleri altında zaman serilerini ortalamak için kullanılır.

$$H = \frac{1}{\frac{\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots \frac{1}{x_n}}{n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \quad \text{veya} \quad \frac{1}{H} = \frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n}$$

Harmonik ortalama aşağı eğimlidir. H.O'da $x_i > 0$ olmalıdır.

Eğer $x_1 = x_2 = \ldots\ldots x_n$ ise $\bar{x} = G = H$ olur.

$$\bar{x} > G > H$$

Uygulamada sabit ve değişken birimler vardır. **Örneğin**; 1 birimlik mal A kişisi tarafından 30 dk'da ve yine 1 birimlik mal B kişisi tarafından 20 dk'da üretiliyorsa mal miktarı sabit, zaman değişkendir. Ortalaması alınan değişkendir yani zamandır.

$H = \frac{2}{\frac{1}{20} + \frac{1}{30}} = 24$ dk'da 1kg mal (ort.) üretilmektedir

Uçakla 400 km, trenle 60 km(570km)

$$H.O = \frac{2}{\frac{1}{400} + \frac{1}{60}} = \frac{2}{\frac{46}{2400}} = \frac{4800}{46} \cong 104 km/h$$

<u>Harmonik Ortalama uygulama yerleri</u>

Zaman birimi başına hız

Para birimi başına satın alınan birim sayısı

Örnek: A ve B gibi iki şehir arasında 100km lik bir yol vardır. Bir otomobilli yolun ilk yarısını 30 km/saat hızla gidiyor. Diğer yarısını 40 km/saat hızla gidiyor. Hız ortalaması nedir?

v = ortalama hız ; t = geçen zaman ; d = alınan yol

t1: Yolun ilk yarısında geçen zaman

t2: Yolun ikinci yarısında geçen zaman

$$\frac{d}{2} = v_1.t_1 = v_2.t_2$$

$$t_1 = \frac{d}{2v_1} \quad ve \quad t_2 = \frac{d}{2v_2}$$

$$t = t_1 + t_2 = \frac{d}{2}.\left(\frac{1}{v_1} + \frac{1}{v_2}\right) = \frac{vt}{2}\left(\frac{1}{v_1} + \frac{1}{v_2}\right)$$

$$v = \frac{d}{t} = \frac{vt}{\frac{vt}{2}\left(\frac{1}{v_1} + \frac{1}{v_2}\right)} = \frac{2}{\left(\frac{1}{v_1} + \frac{1}{v_2}\right)} = \frac{2}{\left(\frac{1}{30} + \frac{1}{40}\right)} = 34.28 = \frac{n}{\sum \frac{1}{x_i}}$$

Harmonik ortalama

KUADRATİK ORTALAMA :

Gözlemlerin karelerinin aritmetik ortalamasının köküdür. Standart sapmanın hesaplanmasında kullanılır. Ortalama değerlerinin ortalamasında kullanılmaz.

$$K = \sqrt{\frac{\Sigma x^2}{n}}$$

$$K > \overline{x} > G > H$$

Mod

- Bir veri setinde en çok gözlenen (en çok tekrar eden) değere veya frekansı en fazla olan şans değişkeni değerine **mod** adı verilir.

- Veri setinin *modu* olmayacağı gibi birden fazla da *modu* olabilir.

- Mod genellikle kesikli şans değişkenli için oluşturulan gruplanmış serilerde aritmetik ortalama yerine kullanılabilir.

Basit Seriler İçin Mod

Örnek: Bir fabrikada çalışan 7 endüstri mühendisinin bildiği yabancı dil sayıları aşağıda verilmiştir. Buna göre bu mühendislerin bildiği yabancı dil sayısının modunu hesaplayınız.

x_i : 2,0,1,2,0,1,0 0,0,0,1,1,2,2.

Veri setinde en çok tekrar eden eleman 0 olduğundan (3 kez) mod değeri 0 'dır.

• Eğer veri seti 1,0,1,2,0,1,0 şeklinde olsaydı veri seti iki modlu olacaktı. (0 ve 1)

• Eğer veri seti 2,0,1,2,0,1 şeklinde olsaydı veri setinin modunun olmadığı ifade edilecekti.

Gruplanmış Seriler İçin Mod

Örnek: Aşağıdaki tabloda bir TV bayisindeki LCD televizyonların ekran boyutlarına göre satış miktarları verilmiştir. Frekans dağılımının aritmetik ortalamasını hesaplayınız.

Ekran	Satış Adedi
51	1
66	3
72	4
82	5
94	7

- **Frekans dağılımına bakıldığında en fazla satış miktarı 94 ekran LCD televizyonda olduğundan dolayı (7 adet) dağılımın modunun 94 olduğu söylenir.**
- *Eğer 82 ekran LCD televizyonlarından da 7 adet satılsaydı dağılımın iki modu olduğu ifade edilirdi. (82 ve 94)*

Sınıflanmış Seriler İçin Mod

- Sınıflanmış serilerde mod değeri hesaplanırken ilk olarak mod sınıfı belirlenir.
- Mod sınıfı frekansı en yüksek olan sınıftır.
- Mod sınıfı belirlendikten sonra bu sınıf içerisinde yer alan modun tam değeri sınıf frekansı ve kendine komşu olan sınıf frekansları dikkate alınarak hesaplanır.

Sınıflandırılmış Veriler İçin

Sınıflar	f_i	m_i	$f_i.m_i$	Σf_i
0 – 10'dan az	12	5	60	12
10 - 20'dan az	6	15	90	18
20 - 30'dan az	4	25	100	22
30 – 40''dan az	2	35	70	24
40 – 50'dan az	1	45	45	25
Σ	25		365	

Mod Sınıfı

$$\text{Mod} = L_{\text{mod}} + \frac{\Delta_1}{\Delta_1 + \Delta_2}.i$$

Sınıf aralığı

Mod sınıfının alt sınırı

Mod sınıfıyla bir önceki sınıf fr. arasındaki fark

Mod sınıfıyla bir sonraki sınıf fr. arasındaki fark

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 Seriler İçin Medyan

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

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

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

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

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

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

n/2 ve (n/2)+1 nci elemanlar 68 ve 79 olup bunların ortalaması 73,5 medyan değeridir.

Veri Seti 30,42,56,61,68,79,82,88,90 şeklinde 9 adet veriden oluşsaydı (n+1)/2 nci eleman olan 68 veri setinin medyanı olacaktı.

Gruplanmış Seriler İçin Medyan

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

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

Grup	Frekans	$\sum f_i$
51	1	**1**
66	3	**4**
72	4	**8**
82	5	**13**
94	7	**20**

Örnek: Yandaki tabloda bir TV bayisindeki LCD televizyonların ekran boyutlarına göre satış miktarları verilmiştir. Frekans dağılımının medyanını hesaplayınız.

- n/2 ve (n/2)+1 nci gözlem değerlerine karşılık gelen değerler (10 ve 11 nci sıra) 82 olduğundan dolayı medyan değeri 82'dir.

Grup	Frekans	$\sum f_i$
51	1	**1**
66	3	**4**
72	4	**8**
82	5	**13**
94	2	**15**

- Frekans dağılımı yandaki gibi olsaydı (n+1)/2 nci elemana (8 nci elemana) karşılık gelen sayı 72 olduğunda dolayı veri setinin medyanı 72 olacak idi.

Sınıflanmış Seriler İçin Medyan

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

Sınıflandırılmış Veriler İçin

Sınıflar	f_i	m_i	$f_i.m_i$	Σf_i
0 – 10'dan az	12	5	60	12
10 - 20'dan az	6	15	90	18
20 - 30'dan az	4	25	100	22
30 – 40"dan az	2	35	70	24
40 – 50'dan az	1	45	45	25
Σ	25		365	

Medyan Sınıfı

$$\text{Med} = L_{med} + \frac{\frac{\sum f_i}{2} - \Sigma f_1}{f_{med}}.i$$

Medyan sınıfından bir önceki sınıfın Yığmalı frekansı

Medyan sınıfının alt sınırı

Medyan sınıfının frekansı

Sınıf aralığı

Sınıflandırılmış Veriler İçin

Sınıflar	f_i	m_i	$f_i.m_i$	Σf_i
0 – 10'dan az	12	5	60	12
10 - 20'dan az	6	15	90	18
20 - 30'dan az	4	25	100	22
30 – 40''dan az	2	35	70	24
40 – 50'dan az	1	45	45	25
Σ	25		365	

Mod Sınıfı

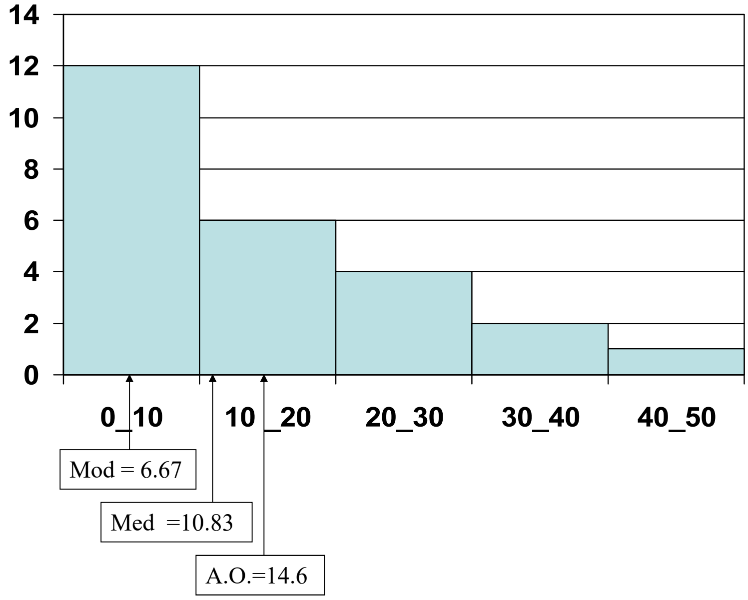
$$\text{Mod} = L_{\text{mod}} + \frac{\Delta_1}{\Delta_1 + \Delta_2}.i \quad = 0 + \frac{(12-0)}{(12-0)+(12-6)}.10 = 6.67$$

Sınıflandırılmış Veriler İçin

Sınıflar	f_i	m_i	$f_i.m_i$	Σf_i
0 – 10'dan az	12	5	60	12
10 - 20'dan az	6	15	90	18
20 - 30'dan az	4	25	100	22
30 – 40''dan az	2	35	70	24
40 – 50'dan az	1	45	45	25
Σ	25		365	

Medyan Sınıfı

$$\text{Med} = L_{\text{med}} + \frac{\frac{\sum f_i}{2} - \Sigma f_{\text{med}}}{f_{\text{med}}}.i = 10 + \frac{\frac{25}{2} - 12}{6}.10 = 10.83$$

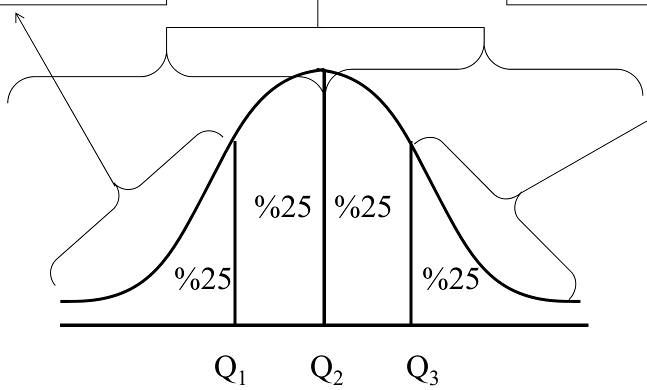


KARTİLLER

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

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

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



Sınıflar	f_i	Σf_i
0-20'den az	8	8
20-40'dan az	12	20
40-60'dan az	25	45
60-80'den az	15	60
80-100	10	70

Q_1 sınıfı

$$Q_1 = L_{Q_1} + \frac{\frac{\sum f_i}{4} - f_l}{f_{Q_1}}.i = 20 + \frac{\frac{70}{4} - 8}{12}.20 = 35.83$$

Sınıflar	f_i	Σf_i
0-20'den az	8	8
20-40'dan az	12	20
40-60'dan az	25	45
60-80'den az	15	60
80-100	10	70

Q_2 sınıfı

$$Q_2 = Med = L_{Q_2} + \frac{\frac{\sum f_i}{2} - f_l}{f_{Q_2}}.i = 40 + \frac{\frac{70}{2} - 20}{25}.20 = 52$$

Sınıflar	f_i	Σf_i
0-20’den az	8	8
20-40’dan az	12	20
40-60’dan az	25	45
60-80’den az	15	60
80-100	10	70

Q_3 sınıfı

$$Q_3 = L_{Q_3} + \frac{\frac{3\sum f_i}{4} - f_l}{f_{Q_{31}}}.i \; = 60 + \frac{\frac{3\times 70}{4} - 45}{15}.20 = 70$$