Tanımlayıcı İstatistikler

- Bir veya birden fazla dağılışı karşılaştırmak için kullanılan ve ayrıca örnek verilerinden hareket ile frekans dağılışlarını sayısal olarak özetleyen değerlere **tanımlayıcı istatistikler** denir**.**
- Analizlerde kullanılan seri tiplerine (*basit, gruplanmış, sınıflanmış*) göre hesaplamalarda kullanılacak formüller değişmektedir.

Tanımlayıcı İstatistikler

Yer Ölçüleri

1)Aritmetik ort.
2)Geometrik ort.
3)Harmonik ort.
4)Mod
5)Medyan
6)Kartiller

Değişkenlik Ölçüleri

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

Çarpıklık Ölçüleri

1)Pearson Asimetri Ölçüsü
2)Bowley Asimetri Ölçüsü

Basıklık Ölçüleri

Yer Ölçüleri

- Veri setini tanımlamak üzere kullanılan ve genellikle tüm elemanları dikkate alarak veri setini özetlemek için kullanılan ifadelerdir.

- Veri setindeki tüm elemanları temsil edebilecek merkez noktasına yakın bir değerdir.

- Merkezi eğilim ölçüleri olarak da adlandırılırlar.

1) 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

Basit Seriler İçin Aritmetik Ortalama

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

n: örnek hacmi

i = 1,2,3,……….,n

Örnek: Bir fabrikada çalışan 5 endüstri mühendisinin bildiği yabancı dil sayıları aşağıda verilmiştir. Buna göre bu mühendislerin bildiği yabancı dil sayısının aritmetik ortalamasını hesaplayınız.

2,0,1,2,0 X_i = 0,0,1,2,2. n = 5 i = 1,2,…,5

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2+0+1+2+0}{5} = 1$$

Gruplanmış Seriler İçin Aritmetik Ortalama

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

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

f : frekans

k: grup sayısı

i = 1,2,3,……….,k

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

Grup	Frekans	$x_i f_i$
51	1	**51**
66	3	**198**
72	4	**288**
82	5	**410**
94	7	**658**
	$\sum f_i =$ **20**	**1605**

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{51(1) + 66(3) + ... + 94(7)}{1+3+4+5+7}$$

$$= \frac{1605}{20} = 80,25$$

Sınıflanmış Seriler İçin Aritmetik Ortalama

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

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 <u>değerlerin</u> neler olduğu bilinmediğinden dolayı ve yalnızca her bir sınıfın <u>frekans değerleri</u> bilindiğinden dolayı sınıfı temsil etmek üzere sınıf orta noktaları hesaplamada kullanılır.

- Kullanılan formül gruplanmış seriler için kullanılan formüle benzerdir.

Örnek: Aşağıdaki tabloda 30 günlük süre içinde bir restoranın kullandığı et miktarının dağılımı verilmiştir. Günlük kullanılan et miktarının aritmetik ortalamasını hesaplayınız.

Sınıflar	f_i	m_i	$m_i f_i$
30-36'dan az	2	*33*	*66*
36-42'den az	6	*39*	*234*
42-48'den az	10	*45*	*450*
48-54'dan az	7	*51*	*357*
54-60'den az	4	*57*	*228*
60-66'den az	1	*63*	*63*
Toplam	30		*1398*

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{\sum_{i=1}^{k} f_i} = \frac{33(2) + 39(6) + \ldots + 63(1)}{30}$$

$$= \frac{1398}{30} = 46{,}6\ kg.$$

2) Geometrik Ortalama

- Bir veri setinde bulunan n adet elemanın çarpımının n nci dereceden kökünün alınmasıyla elde edilen yer ölçüsüdür.

$$G = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$$

- Geometrik ortalamanın formülüne bakıldığında hesaplama zorluğu olduğundan dolayı logaritma ifadesi kullanılır. Genellikle basit seriler için kullanışlı olup negatif sayılar için kullanışlı değildir.

$$Log\ G = \frac{\sum_{i=1}^{n} \log x_i}{n} \qquad G = anti\log \frac{1}{n} \sum_{i=1}^{n} \log x_i$$

Geometrik Ortalama'nın Kullanım Alanları

- Ortalama oranları,
- Değişim Oranları,
- Logaritmik dağılış gösteren veri setleri,

için kullanışlıdır.

Örnek: fiyat indeksleri, faiz formülleri.

Örnek: Bir alışveriş merkezindeki 5 farklı meyvenin satış fiyatı aşağıdaki gibidir. Buna göre meyvelerin satış fiyatlarının geometrik ortalamasını hesaplayınız.

Elma: 1,5 YTL. Üzüm: 2,5 YTL Erik: 1 YTL

Muz : 3 YTL. Armut : 2 YTL.

$$G = \sqrt[n]{x_1 \times x_2 \times .... \times x_n} = \sqrt[5]{1(1{,}5)(2)(2{,}5)(3)}$$

$$= \sqrt[5]{22{,}5} \approx 1{,}86\,YTL.$$

$$Log\ G = \frac{\sum_{i=1}^{n} \log x_i}{n} = \frac{0 + 0{,}17609 + 0{,}30103 + 0{,}39794 + 0{,}47712}{5}$$

$$Log\,G = \frac{1{,}35218}{5} \approx 0{,}27045$$

G = anti log 0,27045 = 10^{0,27045} ≈ 1,86 YTL.

3) Harmonik Ortalama

- Bir veri setinde bulunan n adet elemanın çarpma işlemine göre terslerinin ortalamasının tersinin alınmasıyla elde edilen yer ölçüsüdür. Genellikle basit seriler için kullanışlıdır.

$$H = \frac{1}{\left(\dfrac{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \ldots + \dfrac{1}{x_n}}{n} \right)} = \frac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \ldots + \dfrac{1}{x_n}}$$

$$\frac{1}{H} = \frac{\sum_{i=1}^{n} \dfrac{1}{x_i}}{n}$$

Harmonik Ortalama’nın Kullanım Alanları

- Belirli fiyat tipleri,
- Zaman serileri,

için kullanışlıdır.

Örnek: Zaman birimi başına hız, para birimi başına satın alınan birim sayısı.

NOT: ARİTMETİK ORT. > GEOMETRİK ORT. > HARMONİK ORT.

Örnek: Bir tekstil fabrikasında çalışan dört kişinin bir pantolonu ütüleme süreleri aşağıda verilmiştir. Buna göre bu fabrikada bir pantolon ortalama kaç dakikada ütülenir?

İşçi 1: 10 dk. İşçi 2: 6 dk. İşçi 3: 4 dk. İşçi 4 : 5 dk.

$$\frac{1}{H}=\frac{\sum_{i=1}^{n}\frac{1}{x_i}}{n}=\frac{\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{10}}{4}=\frac{43}{240}$$

$$H=\frac{240}{43}\approx 5{,}58\,dk.$$

4) 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 5 endüstri mühendisinin bildiği yabancı dil sayıları aşağıda verilmiştir. Buna göre bu mühendislerin bildiği yabancı dil sayısının modunu hesaplayınız.

$\boldsymbol{x_i}$ **:** 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 Samsung 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.

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

L_{Mod} = Mod Sınıfı Aralığının Alt Sınırı

Δ_1 = Mod Sınıfı Frekansı - Kendinden Bir Önceki Sınıf Frekansı

Δ_2 = Mod Sınıfı Frekansı – Kendinden Bir Sonraki Sınıf Frekansı

i = Mod Sınıfının Sınıf Aralığı

Örnek: Aşağıdaki tabloda 30 günlük süre içinde bir restoranın kullandığı et miktarının dağılımı verilmiştir. Günlük kullanılan et miktarının modunu hesaplayınız.

Sınıflar	f_i
30-36'dan az	2
36-42'den az	6
42-48'den az	10
48-54'dan az	7
54-60'den az	4
60-66'den az	1
Toplam	30

Mod sınıfı → 42-48'den az

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

$$= 42 + \frac{(10-6)}{(10-6)+(10-7)}.6 = 45,4\ kg.$$

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 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.

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

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

- 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.

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

L_{med} : Medyan sınıfının alt sınırı

f_l : Medyan sınıfından bir önceki sınıfın kümülatif frekansı

f_{med} : Medyan sınıfının frekansı

i : Medyan sınıfının sınıf aralığı

Örnek: Aşağıdaki tabloda 30 günlük süre içinde bir restoranın kullandığı et miktarının dağılımı verilmiştir. Günlük kullanılan et miktarının medyanını hesaplayınız.

Sınıflar	f_i	$\sum f_i$
30-36'dan az	2	2
36-42'den az	6	8
Medyan sınıfı → 42-48'den az	10	18
48-54'dan az	7	25
54-60'den az	4	29
60-66'den az	1	30
Toplam	30	

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

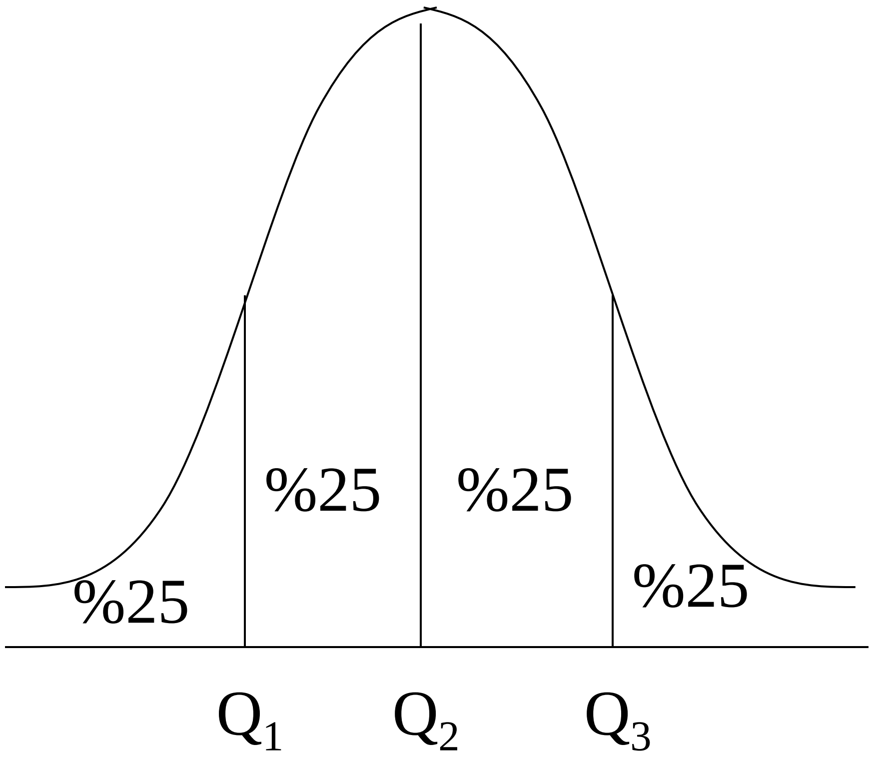
$$= 42 + \frac{15-8}{10}.6 = 46,2\,kg.$$

6) Kartiller

•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.*



Basit Seriler İçin Kartiller

- **1.Kartil Q_1**

Veri Setinin Hacmi Tek Sayı İse;

$$\frac{n+1}{4}$$ *nci gözlem değeri,*

- Veri Setinin Hacmi Çift Sayı İse;

$$\frac{n}{4} \quad ve \quad \frac{n}{4}+1$$ *nci gözlem*

değerlerinin aritmetik ortalaması 1.Kartili verir.

- **3.Kartil Q_3**

Veri Setinin Hacmi Tek Sayı İse;

$$\frac{3n+1}{4}$$ *nci gözlem değeri,*

Veri Setinin Hacmi Çift Sayı İse;

$$\frac{3n}{4} \quad ve \quad \frac{3n}{4}+1$$ *nci gözlem*

değerlerinin aritmetik ortalaması 3.Kartili verir.

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/4 ve (n/4)+1 nci elemanların aritmetik ortalaması 56 olduğundan Q_1=56, 3n/4 ve (3n/4)+1 nci elemanların aritmetik ortalaması 88 olduğundan Q_3=88 ‘dir	Veri Seti, 30,42,56,61,68,79,82,88,90 şeklinde 9 hacimli olsaydı, (n+1)/4 ncü eleman olan 49 1.Kartil, (3n+1)/4 ncü eleman olan 82 3.Kartil olarak ifade edilirdi.

Gruplanmış Seriler İçin Kartiller

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

- Gruplanmış serilerde 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.

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

Örnek: Yandaki tabloda bir Samsung bayisindeki LCD televizyonların ekran boyutlarına göre satış miktarları verilmiştir. Frekans dağılımının 1. ve 3 ncü Kartillerini hesaplayınız.

- n/4 değerine karşılık gelen sıra grup değeri 82 olduğundan 1.Kartil, ve 3n/4 değerine karşılık gelen grup değeri 94 olduğundan 3.Kartil olarak ifade edilir.

- Frekans dağılımı yandaki gibi verilmiş olsaydı

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

$Q_1 = 66$ ve $Q_3 = 82$ olacak idi.

Sınıflanmış Seriler İçin Kartiller

- Sınıflanmış serilerde kartiller hesaplanırken ilk olarak kümülatif frekans sütunu oluşturularak kartil sınıfları belirlenir.

- Kartil sınıfları belirlenirken gruplanmış serilerde 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 kümülatif 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: Aşağıdaki tabloda 30 günlük süre içinde bir restoranın kullandığı et miktarının dağılımı verilmiştir. Günlük kullanılan et miktarının 1 nci ve 3 ncü kartillerini hesaplayınız.

	Sınıflar	f_i	$\sum f_i$
	30-36'dan az	2	2
Q_1 sınıfı →	36-42'den az	6	8
	42-48'den az	10	18
Q_3 sınıfı →	48-54'dan az	7	25
	54-60'den az	4	29
	60-66'den az	1	30
	Toplam	30	

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

$$= 36 + \frac{7{,}5 - 2}{6}.6 = 41{,}5\,kg. \qquad = 48 + \frac{22{,}5 - 18}{7}.6 = 51{,}9\,kg.$$

Uygulama

1. Beş kişinin yaşları: 24, 26, 20, 18, 24 olduğuna göre aritmetik ortalamasını hesaplayınız.

2. 10 bayanın ayakkabı numaraları:

35, 38, 36, 36, 37, 36, 38, 35, 39, 37 olduğuna göre bu seriyi gruplandırılmış seri yaparak aritmetik ortalamasını bulunuz.

3. Aşağıdaki sınıflandırılmış serinin aritmetik ortalamasını, modunu ve medyanını bularak grafik üzerinde gösteriniz.

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

4. Aşağıdaki sınıflandırılmış serinin aritmetik ortalamasını, medyanını ve modunu bulunuz.

Sınıflar	f_i	m_i	$m_i f_i$	Σf_i
40-49	3	45	135	3
50-59	5	55	275	8
60-69	11	65	715	19
70-79	22	75	1650	41
80-89	15	85	1275	56
90-100	6	95	570	62
toplam	62		4620	

5. Aşağıdaki sınıflandırılmış serinin 1.,2. ve 3. kartillerini elde ediniz.

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