MERKEZİ EĞİLİM ÖLÇÜLERİ

Betimsel istatistikte temel amacın verileri en iyi şekilde kullanıma sunmak olduğunu söylemiştik. Bunu bir ölçüde çeşitli tablolar ve grafikler kullanarak gösterdik. Ancak bazen verilerin sadece tablo ya da grafik ile gösterilmesi yeterli olmayabilir. Özellikle veri gruplarının mukayesesinde, bu verilerin ortalama gibi tek bir değer ile temsil edilmesi gerekir. Sözgelimi, Fransa ile Türkiye'de yaşayan insanların gelir düzeylerini mukayese edecek olursak, ancak bu iki ülkedeki ortalama gelir düzeyi hesaplanarak bu mukayese yapılabilir.

İstatistikte, basit aritmetik ortalama dışında kullanılan başka ortalama türleri de mevcuttur. Ortalamalara ve oranlara genel olarak Merkezi Eğilim Ölçüleri ya da Konum ölçüleri denilmektedir. Çünkü, ortalama veya oranlar, seri değerlerinin sayı ekseninde hangi merkez nokta etrafında yığılma eğilimine sahip olduğunu veya sayı eksenindeki konumunu temsil eder. Bir serinin merkezi eğilim noktası, fizikte bir cismin ağırlık merkezi kavramına tekabül etmektedir. Bu bölümde aritmetik ortalama, tartılı aritmetik ortalama, geometrik ortalama, harmonik ortalama mod ve medya gibi merkezi eğilim ölçülerine değinilecektir

Analitik Ortalamalar

Serinin tüm değerini ortalama hesabına dahil eden ortalama türlerine Analitik ortalamalar adı verilir. *N* toplam terim sayısı, *X* seri değerleri, *f* her bir seri değerine karşı gelen frekans, m_j, *j* inci grubun ortası olmak üzere, basit, sınıflanmış ve gruplanmış serilerde analitik ortalamalar için genel formüller sırasıyla,

$$M_r = \left(\frac{\sum X^r}{N} \right)^{\frac{1}{r}} \qquad M_r = \left(\frac{\sum f_j X_j^r}{\sum f_j} \right)^{\frac{1}{r}} \qquad M_r = \left(\frac{\sum f_j m_j^r}{\sum f_j} \right)^{\frac{1}{r}}$$

olup; *r*'ye ortalamanın mertebesi adı verilir. Teorik olarak *r* herhangi bir değer alabilirse de, uygulamada en çok kullanılan değerler ve ortaya çıkan ortalama türleri şunlardır:

$r = -1$	için	M_{-1}	Harmonik Ortalama
$r = 0$	için	M_0	Geometrik Ortalama
$r = 1$	için	M_1	Aritmetik Ortalama
$r = 2$	için	M_2	Karesel Ortalama

Şimdi bunları ayrı ayrı görelim.

Aritmetik Ortalama

Genel ortalama formülünde ortalama mertebesinin $r=1$ alınmasıyla elde edilir. en yaygın kullanım alanına sahip olan ortalama türü aritmetik ortalamadır. Basitçe, terimlerin toplamının terim sayısına bölünmesi diye tanımlanabilir. Bir $X_1, X_2, \ldots, X_N$ serisi verildiğinde bunun aritmetik ortalaması özel olarak μ ile gösterilir ve;

$$M_1 = \left[\frac{\sum X_i}{N}\right]^1$$

veya

$$A = \frac{\sum X_i}{N}$$

biçiminde bulunur.

Örnek: Bir öğrencinin dört ayrı dersten aldığı puanlar sırasıyla 72, 68, 87 ve 53 olsun. Bu öğrencinin bu dört dersten aldığı puanlar itibariyle başarı ortalaması;

$$A = \frac{72+68+87+53}{4} = 70 \text{ puan}$$

olarak bulunur.

Aritmetik Ortalamanın Özellikleri

1. Bir seride terimlerin genel topluluğuna nazaran aşırı küçük veya aşırı büyük olan terimlere aşırı uç değerler veya kısaca uç değerler diyoruz. Aritmetik ortalama hesabına giren bütün terimlerin ağırlığı aynıdır. Bu da aritmetik ortalamayı terim değerlerine karşı hassas kılmaktadır. Eğer bir seride aşırı uç değerler varsa arimetik ortalama bu uç değerlerden hemen etkileneceğinden ortalamanın temsili olma özelliği ortadan kalkar.

2. Aritmetik ortalama ile terim sayısı çarpımı terim toplamını verir.

$$NA = \sum X$$

3. Bütün terimlere bir k sayısı eklendiğinde (çıkarıldığında) aritmetik ortalama da k sayısı kadar büyür (küçülür).

$$\frac{\sum(X+k)}{N}=\frac{\sum X+Nk}{N}=\frac{\sum X}{N}+k=A+k$$

4. Bütün terimler bir k sayısı ile çarpıldığında (bölündüğünde) aritmetik ortalama da k katı kadar büyür (küçülür).

$$\frac{\sum kX}{N}=\frac{k\sum X}{N}=k\frac{\sum X}{N}=kA$$

5. Terimlerin aritmetik ortalamadan farkları toplamı daima sıfırdır.

$$\sum(X_i-A)=\sum X_i-NA=0$$

6. Bir X serisinin aritmetik ortalaması A olmak üzere;

$$KT=\sum(X-A)^2$$

ile gösterilen aritmetik ortalamadan sapmaların kareleri toplamının bir minimum olduğunu gösterilebilir.

Tartılı Aritmetik Ortalama

Aritmetik ortalama hesaplanırken bazan ortalamaya giren bütün terimler aynı önleme haiz olmayabilir. Terimler arasında böyle önem farkları olduğunda, her bir terime önemi ile orantılı bir ağırlık (tartı) verilerek hesaplanan ortalamaya tartılı aritmetik ortalama denir. Bu ortalama, simgesel olarak A_t ile gösterilir ve $A_t=\frac{\sum t_j X_j}{\sum t_j}$ ile hesaplanır. Burada t_j 'ler her bir X_j 'ye karşı gelen tartıları göstermektedir. Açıktır ki, $\sum t_j$ de tartıların toplamını göstermektedir.

Örnek: Tartılı aritmetik ortalamanın en tipik uygulama alanı öğrenci başarı notu hesaplamasıdır. Bir öğrencinin aldığı derslerin önemi farklı farklıdır ve bu fark, derslerin kredileri ile orantılıdır. Bir öğrenci dört dersten aşağıdaki puan ve kredilere sahip olsun.

	Puanlar	Krediler
İstatistik	67	4
Muhasebe	85	3
İktisat	63	3
Matematik	52	2

Bu öğrencinin ortalama başarısını tartılı aritmetik ortalama ile hesaplayalım.

$$A_t = \frac{67(4)+85(3)+63(3)+52(2)}{4+3+3+2} = \frac{816}{12} = 68 \text{ Puan}$$

olarak ortalama başarı düzeyi bulunur.

Geometrik Ortalama

G harfi ile gösterilen geometrik ortalama, terimlerin çarpımının terim sayısı mertebesinden kökünün alınması ile elde edilir. Genel ortalama formülünde $r = 0$ alınarak geometrik ortalamanın elde edilişinde, matematiksel olarak önce bir belirsizlik ortaya çıkar. Belirsizliği gidermek için bilinen kurallar uygulanarak aşağıdaki formül elde edilir[1].

$$G = \sqrt[N]{X_1.X_2...X_N} = \left(X_1.X_2...X_N\right)^{\frac{1}{N}}$$

Ancak uygulamada rakamlar büyüdükçe çarpma işlemi, terim sayısı arttıkça da kök alma işlemi zorlaşacaktır. Bu işlemleri, eşitliğin her iki yanının logaritmasını alarak kolaylaştırıyoruz. Böylelikle rakamlar küçülecek, çarpma işlemi toplama işlemine, kök alma işlemi de çarpma işlemine dönüşecektir. Şimdi, yukarıdaki eşitlikte bu işlemleri yapalım:

$$\log G = \log\left(X_1.X_2...X_N\right)^{\frac{1}{N}}$$

$$= \frac{1}{N}\log\left(X_1.X_2...X_N\right)$$

$$= \frac{1}{N}\left(\log X_1 + \log X_2 + \cdots + \log X_N\right)$$

$$= \frac{1}{N}\sum \log X_i$$

Bu şekilde elde edilen log G'den G'nin elde edilmesi için $10^{\log G} = G$ biçiminde antilog almak gerekir.

Örnek: Bir ülkenin nüfusu her yirmi beş yılda 2, 4, 8, 16, 32 şeklinde geometrik bir seri ile artmaktadır. Bu ülkenin yirmi beş yılda ortalama nüfus artış hızını hesaplayalım.

$$G = (2.4.8.16.32)^{\frac{1}{5}}$$

$$= \left(2.2^2.2^3.2^4.2^5\right)^{\frac{1}{5}}$$

$$= \left(2^{15}\right)^{\frac{1}{5}} = 2^3 = 8$$

olarak bulunur. Bu örnekte rakamların uygunluğu işlemleri kolaylaştırdığından logaritmaya gerek kalmamıştır.

Örnek: Son beş yılda portakal fiyatları nisbi olarak 100, 98, 110, 112, 121 şeklinde değişmiştir. Portakal fiyatlarındaki son beş yıla ait nisbi değişmenin ortalamasını bulunuz.

$$G = (100.98.110.112.121)^{\frac{1}{5}}$$

$$\log G = \frac{1}{5}(\log 100 + \log 98 + \log 110 + \log 112 + \log 121)$$

$$= \frac{1}{5}(2.00 + 1.99 + 2.0 + 2.05 + 2.08)$$

$$= \frac{1}{5}(10.16) = 2.03$$

$$10^{\log G} = G = 10^{2.03} = 107.6$$

olarak geometrik ortalama bulunur.

Geometrik Ortalamanın Özellikleri

1. Geometrik ortalama, terimler arasında 0 veya negatif değerli olan varsa hesaplanamayacağı açıktır.
2. Geometrik ortalama daima aritmetik ortalamadan küçük çıkar.
3. Geometrik ortalama uç değerlere karşı aritmetik ortalama kadar hassas değildir.
4. Terimlerin geometrik ortalamaya bölümlerinin çarpımı 1 eder.

$$\frac{X_1}{G}.\frac{X_2}{G}\cdots\frac{X_N}{G}=1$$

5. Geometrik ortalamanın N inci kuvveti terimlerin çarpımına eşittir.

$$G^N = X_1.X_2 \ldots X_N$$

6. Terimlerin k'ıncı kuvveti alındığında geometrik ortalama da k ıncı kuvvete yükseltilmiş olur.

$$\sqrt[N]{X_1^k.X_2^k \ldots X_N^k} = \left(X_1.X_2 \ldots X_N\right)^{\frac{k}{N}} = G^k$$

Örnek : Geometrik ortalamanın özelliklerinden ikincisini basit olarak *X*: 1, 2, 3 serisi ile gösterebiliriz. Bu seride;

$$A=\left(\frac{1}{3}\right)(1+2+3)=\frac{6}{3}=2$$

ve

$$G=(1.2.3)^{\frac{1}{3}}=6^{\frac{1}{3}}=1.82$$

olduğundan $G < A$ bulunur.

Harmonik Ortalama

Genel ortalama formülünde $r=-1$ ile elde edilir. Harmonik ortalama, terimlerin çarpmaya göre terslerinin aritmetik ortalamasının tersi olup, *H* ile gösterilir. Formül olarak *N* elemanlı bir seride harmonik ortalama;

$$H=\left[\frac{\sum X_i^{-1}}{N}\right]^{-1}$$

veya

$$H=\frac{N}{\sum \frac{1}{X_i}}$$

Biçiminde ifade edilir. şu halde, bir serinin harmonik ortalamasının hesabı için önce bir terimin çarpmaya göre tersi alınmalıdır.

Örnek: Bir X_i: 2, 4, 3, 5 serisi verilmiş olsun. Bu serinin harmonik ortalaması için önce terimlerin terslerini alalım.

$$\frac{1}{X_i} : 0.5,\ 0.25,\ 0.33,\ 0.20$$

ve terslerin toplamı;

$$\sum\left(\frac{1}{X_i}\right) = 1.28$$

Terim sayısı $N = 4$ olduğundan harmonik ortalama;

$$H = \frac{4}{1.28} \text{ veya } H = 3.125$$

bulunur.

Harmonik Ortalamanın Özellikleri

1. Terimlerden biri 0 ise harmonik ortalama sonsuz (belirsiz) çıkar. Negatif terimlerin varlığı halinde ise sonuç anlamsız olur.

2. Harmonik ortalama daima Geometrik ortalamadan küçük çıkar.

Örnek: *X*: 1, 2, 3 basit serisi için $A = 2,\ G = 1.82$ bulunmuştu. Bu seri için harmonik ortalama;

$$H = \left(\frac{\frac{1}{1} + \frac{1}{2} + \frac{1}{3}}{3}\right)^{-1}$$

$$= \left(\frac{1.833}{3}\right)^{-1} = (0.611)^{-1} = 1.64$$

değerine sahip olduğundan $H < G$ olduğu görülür.

Karesel Ortalama

Karesel ortalama terimlerin önce karelerinin alınması ve daha sonra karelerin aritmetik ortalamasının karekökünün alınması ile elde edilir. Karesel ortalama *K* harfi ile gösterilir. Formül olarak karesel ortalama, *N* elemanlı bir *X* serisi için;

$$K = \sqrt{\frac{\sum X_i^2}{N}}$$

şeklinde yazılır.

Örnek 12: Bir X serisi 2, 5, 4, 3, 6 olarak verilmiştir. Bu serinin karesel ortalaması için ara değerler şöyledir:

$$\begin{array}{llllllll} X: & 2 & 5 & 4 & 3 & 6 & \\ X^2: & 4 & 25 & 16 & 9 & 36 & \text{Toplam} = 90 \end{array}$$

Buna göre serinin karesel ortalaması:

$$K = \sqrt{\frac{\sum X_i^2}{N}} = \sqrt{\frac{90}{5}} = \sqrt{18} = 4.24$$

olarak bulunur.

Karesel Ortalamanın Özellikleri

1. Karesel ortalamanın en önemli özelliği, terimleri arasında 0 veya negatif değerler olsa da hesaplanmasının mümkün olmasıdır. Bu özellik karesel ortalamanın diğer ortalamalara üstünlüğüdür.
2. Karesel ortalama sayısal olarak diğer ortalamalara nazaran daha büyük çıkar. Bu yüzden de uç değerlerden en fazla etkilenen ortalama türüdür.

Örnek: Daha önce X: 1, 2, 3 serisi için $H = 1.64$, $G = 1.82$ ve $A = 2$ bulunmuştu. Aynı seri için karesel ortalama;

$$K = \sqrt{\frac{1^2 + 2^2 + 3^2}{3}}$$

$$= \sqrt{\frac{14}{3}} = \sqrt{4.66} = 2.2$$

Bulunur. Geometrik ortalamanın aynı seri için en büyük değeri verdiği görülüyor.

Analitik ortalamalar arasında $H \le G \le A \le K$ ilişkisi vardır. eşitlik durumunun sadece değişken terimlerinin tamamının eşit olması halinde, yani $X_i = a,\ a$ bir sabit sayı ve $i = 1,\ 2, ..., N$ durumunda sağlandığı gösterilebilir.

Örnek : Serimiz X: 2, 2 olarak verilmiş olsun. Bu seri için dört tip ortalamayı hesaplayalım.

X	$\frac{1}{X}$	X^2	$\log_2 X$
2	0.5	4	1
2	0.5	4	1
4	1.0	8	2

$$H = \frac{2}{1.0} = 2 \qquad \log G = \frac{2}{2} = 1 \qquad G = 2^1 = 2$$

$$A = \frac{4}{2} = 2 \qquad K = \left(\frac{8}{2}\right)^{\frac{1}{2}} = 2$$

Bütün ortalamalar eşit çıkmıştır.

Analitik Olmayan Ortalamalar

Sadece seride bulunan bazı terimlerin hesaplamaya katıldığı ortalama türleridir. Burada bunlardan mod ve medyan türlerini göreceğiz.

Mod

Bir serinin modu, o seride en çok tekrar edilen değerdir.

Örnek : Bir öğrencinin dokuz ayrı dersten ara sınav puanları şöyle olsun: 50, 66, 25, 80, 50, 75, 66, 50, 90. Bu öğrencinin mod cinsinden başarı ortalaması 50 (en çok tekrar eden değer)dir.

Modun Özellikleri

1. Mod aşırı uç değerlerden etkilenmez, bu yüzden temsil kabiliyeti iyi bir ortalama türüdür.
2. Serideki bütün terimler eşit sayıda tekrar ediyorsa bu durumda serinin modu mevcut değildir.
3. Seride en çok tekrar sayısına sahip iki veya daha fazla değer, dolayısıyla iki veya daha fazla mod söz konusu ise bu değerlerden hangisinin mod olarak alınacağı problem olur.

Medyan

Seri terimleri büyüklük sırasına göre dizildiğinde serinin tam orta yerindeki değere medyan (ortanca) adı verilir ve Me ile gösterilir. Genel olarak, toplam terim sayısı *N* tek sayı ise, medyanın sıra numarası $\frac{(N+1)}{2}$'dir. Yani;

$$Me = X_{\frac{(N+1)}{2}}$$

Toplam terim sayısı *n,* çift sayı ise medyan değeri tam ortada yer alan iki değerin aritmetik ortalamasıdır. Yani;

$$Me = \frac{X_{\frac{N}{2}} + X_{\left(\frac{N}{2}\right)+1}}{2}$$

Örnek: Örnek 14'teki öğrencinin başarı ortalamasını Medyan cinsinden bulmak için önce terimlerin büyüklük sırasına koyulması gerekir:

25, 50, 50, 50, 66, 66, 75, 80, 90

Burada medyan sırası;

$$\frac{(9+1)}{2} = 5$$

olup $Me = 66$ bulunur.

Örnek: Yukarıdaki örnekte terim sayısı bir eksik olsun. Yani;

25, 50, 50, 50, 66, 66, 75, 80

Bu durumda terim sayısı $N = 8$ çift sayıdır. Tam ortaya gelen iki değer $X_4 = 50 \; ve \; X_5 = 66$'dir. Buna göre Medyan;

$$Me = \frac{X_{\frac{8}{2}} + X_{\frac{8}{2}+1}}{2} = \frac{X_4 + X_5}{2} = \frac{50+66}{2} = 58$$

olarak bulunur.

Medyanın Özellikleri

1. Medyan da mod gibi uç değerlerden etkilenmez. Dolayısıyla temsil kabiliyeti yüksektir.

2. Terimlerin Medyandan mutlak sapmalarının toplamı bir minimumdur. Bu minimum terimlerin aritmetik ortalamadan mutlak sapmalarının toplamından daha küçüktür.

$$\sum |X_i - Me| = \min$$

$$\sum |X_i - Me| < \sum |X_i - \mu|$$

3. Alttan ya da üstten açık uçlu serilerde medyan özellikle uygun bir ortalama ölçüsü olmaktadır.

Sınıflanmış Seride Ortalamalar

Buraya kadar gördüğümüz ortalama tanımları ve verilen örnekler basit bir seri içindi. Şimdi sınıflanmış bir seride yukarıda verilen ortalamaların nasıl hesaplanacağına bakalım. Bilindiği gibi sınıflanmış seride iki sütun vardır. birinci sütunda serinin sınıf değerleri (X_j), ikinci sütunda her bir sınıftaki tekrar sayısını veren frekans değerleri f_j yer almaktadır. İşte böyle bir seride aritmetik ortalama;

$$\mu = \frac{\sum f_j X_j}{\sum f_j}$$

Burada $j = 1, 2, ..., m$ sınıf sayısı ve $\sum f_j = N$ toplam frekanstır.

Örnek: Bir atölyede çalışan işçilerin aldıkları yevmiyeye göre dağılımları şöyledir:

Yevmiye (TL)	**İşçi Sayısı**
55	5
64	4
72	7
80	3
90	1

Bu işyerindeki ortalama ücreti bulalım. Toplam işçi sayısı (toplam frekans):

$$\sum f_j = 5+4+7+3+1=20$$

$$A=\frac{1}{20}\left[55(5)+64(4)+72(7)+80(3)+90(1)\right]$$

$$A=\frac{1}{20}1365=68.25\ TL$$

olarak ortalama ücret bulunur.

Sınıflanmış seride medyanın bulunması için kümültatif frekanslardan faydalanıyoruz. Burada medyan hesabına geçmeden önce, verilen serinin büyüklük sırasına göre verilmiş olduğuna dikkat edilmelidir.

Örnek: Yukarıdaki örnek için kümültatif frekanslar:

X_j	f_j	F_j
55	5	5
64	4	9
72	7	16
80	3	19
90	1	20

Bu örnekte terim sayısı $N=20$ olduğundan medyan değeri $\frac{20}{2}=10.$ ile $\frac{20}{2}+1=11.$ sıradaki X değerlerinin aritmetik ortalamasıdır. Yani;

$$Me=\frac{X_{10}+X_{11}}{2}$$

10. ve 11. X değerlerini kümültatif frekans sütunundan arıyoruz. Bu iki sıra (10. ve 11.), ilk 5 de yok, ilk 9 da yok, fakat ilk 16 içerisinde var, 16 kümülatif değerine karşı gelen X değeri medyandır. O halde, $Me=72$ olur.

Gruplanmış Seride Ortalamalar

Gruplanmış bir seride aritmetik ortalama hesabı için önce her bir grubun ortası (m_j) bulunur. Bunu basitçe her grubun üst sınırı ile alt sınırını toplayıp ikiye bölmek suretiyle elde ederiz. Daha sonra serinin aritmetik ortalamasını;

$$A = \frac{\sum f_j m_j}{\sum f_j}$$

ile hesaplanır.

Örnek 20: 1993 yılında İktisat Bölümünden mezun olan öğrencilerin mezuniyet derecelerine göre dağılımı şöyledir.

Mezuniyet Derecesi	**Öğrenci Sayısı**
$50 \leq X < 60$	14
$60 \leq X < 70$	21
$70 \leq X < 80$	10
$80 \leq X < 90$	5
$90 \leq X < 100$	2

Bütün öğrenciler için ortalama mezuniyet derecesini bulalım. Bunun için önce her bir grubun orta değeri bulunmalıdır. Daha sonra bu değerler karşı gelen frekanslarla çarpılır. Her bir grup için bu işlemler yapılarak aşağıdaki bilgileri elde edilir.

m_j	f_j	$f_j m_j$
55	14	770
65	21	1365
75	10	750
85	5	425
95	2	190
Toplam	52	3500

Bu bilgilerden $\sum f_j m_j = 3500$ ve $\sum f_j = 52$ olduğundan, ortalama mezuniyet derecesi;

$$A = \frac{3500}{52} \qquad A = 67.3 \text{ puan}$$

olarak bulunur.

Gruplanmış seride mod değerinin hesabı için önce mod grubu belirlenir. Mod grubu, frekansı en büyük olan gruptur. Mod grubu belirlendikten sonra aşağıdaki formül kullanılır.

$$Mod = L + \frac{f_s}{f_ö + f_s}.c$$

Burada;

L : Mod grubunun alt sınırını

$f_ö$: Mod gurubundan önceki grubun frekansını

f_s : Mod grubundan sonraki grubun frekansını

c : Grup genişliğini

temsil etmektedir.

Örnek 22: Örnek 20'deki gruplanmış serinin mod değerini hesaplayalım. Burada mod grubu, frekansı en yüksek olan ikinci sıradaki gruptur. Buna göre,

$$L = 60, \quad f_ö = 14, \quad f_s = 10, \quad c = (70 - 60) = 10$$

ile

$$Mod = 60 + \frac{10}{14+10}.10 = 60 + \frac{100}{24} = 64.16 \text{ puan}$$

olarak bulunur.

Gruplanmış seride medyan hesabı için, yine önce medyan grubu belirlenmelidir. Medyan grubunun belirlenmesi için kümülatif frekans sütunundan yararlanıyoruz. Medyan değeri $\frac{N}{2}$ inci sıradaki değer olduğuna göre, bu değer kümülatif frekans sütununda ilk sağlandığı yer medyan grubunu belirler. Bundan sonra;

$$Me = L + \frac{c}{f}\left(\frac{N}{2} - d\right)$$

ile medyan bulunur. Burada;

L : Medyan grubunun alt sınırını

c : Grup genişliğini

f : Medyan grubunun frekansını

d : Medyan grubundan bir önceki grubun kümülatif frekansını göstermektedir.

Örnek 23: Yukarıdaki örnekte verilen seri için mezuniyet ortalamasını medyan cinsinden hesaplayalım. Bunun için önce kümülatif frekanslar bulunur.

X	f	F
$50 \leq X < 60$	14	14
$60 \leq X < 70$	21	35
$70 \leq X < 80$	10	45
$80 \leq X < 90$	5	50
$90 \leq X < 100$	2	52

$N = 52$ ve $\frac{N}{2} = 26$ medyanın sıra numarasıdır. Kümülatif frekans sütununda bu sıra numarasının ilk sağlandığı (içerildiği) satır, ikinci satırdır. Şu halde ikinci satırda yer alan $60 \leq X < 70$ grubu medyan grubudur. Buna göre medyan değeri şöylece hesaplanır:

$$L = 60, \quad c = 10, \quad f = 21, \quad d = 14$$

ile

$$Me = 60 + \frac{10}{21}(26 - 14) = 60 + \frac{120}{21} = 65.7 \text{ puan}$$

Ortalama Türünün Seçimi

Bu görmüş olduğumuz ortalama türleri karşılaşılan her veri kümesine rastgele uygulanamaz. Şimdi bunların her birinin kullanım özelliklerini sıralayalım.

Aritmetik ortalama en hassas ortalamadır. Yani, ortalama seri değerlerinden birinde meydana gelen bir değişiklik hemen ortalamayı etkiler. Bu nedenle eğer seride aşırı uç değerler (çok büyük ya da çok küçük değerler) mevcut ise bu ortalama türü uygun olmaz. Eğer ortalama hesabının amacı hassas mukayeseler yapmak ise aritmetik ortalama tercih edilir. Öte yandan aritmetik ortalama hesabında, terimler arasında önem farkı varsa, tartılı aritmetik ortalama kullanılmalıdır. Veriler mutlak değerler olarak değil de bunların nispi değişimleri olarak verilmiş ise, bu durumda geometrik ortalama kullanılır. Yalnız burada rakamlardan birinin sıfır veya negatif olması halinde geometrik ortalama hesaplanamaz. Geometrik ortalama, aritmetik ortalama kadar hassas değildir. Birim başına ortalama değer arandığı durumlarda harmonik ortalama kullanılır. Örneğin, saat başına hız, dönüm başına tarımsal verim, işçi başına üretilen miktarlar türünden verilerde harmonik ortalama hesaplanır. Harmonik ortalama da terimlerden biri sıfır olduğunda ortalama hesaplanamaz.

Mod ve medyan, aşırı uç değer içeren verilerde temsil kabiliyeti en iyi olan ortalama türleridir. Verilerimiz sayısal olmayan bir değişkene ait ise (örneğin göz rengi, meslek, cinsiyet vb) bu durumda ortalama türü olarak sadece mod kullanılabilir. Bir seride bazen birden fazla mod olabilir. Bu durumda mod kullanmak uygun olmaz. Fakat özellikle mod gerekiyorsa sınıflanmış serileri gruplanmış seriye dönüştürerek çift moddan kurtulabiliriz. Gruplanmış seride çift mod grubu varsa burada da grupları birleştirmek suretiyle çift mod durumundan kurtulabiliriz.

DEĞİŞKENLİK ÖLÇÜLERİ

Önceki bölümde bir seriye ait rakamlar topluluğunun hangi merkez (ortalama) etrafında toplanma eğiliminde olduğunu araştırdık. Bunun için muhtelif merkezi eğilim ölçülerini izah ettik. Ancak merkezi eğilim ölçüleri seriler hakkında yeterli bilgi vermeyebilir. Aşağıdaki iki basit seriyi göz önüne alınız:

X	*Y*
58	12
59	70
63	98

Bu serilerin her ikisi de aynı ortalama değerine sahiptir $\left(\bar{X}=60 \text{ ve } \bar{Y}=60\right)$. Oysa bu iki seri arasında önemli bir mahiyet farkı vardır. *X* serisi *Y* serisine nazaran daha homojen, terim değerleri birbirine daha yakındır. Sözgelimi, bu iki seri iki ayrı öğrencinin üç dersten aldıkları puanlarını gösteriyorsa, birinci öğrenci her üç dersten vasat bir başarı gösterirken, ikinci öğrencinin bazı derslerden başarısız bazı derslerden çok başarılı olduğu görülüyor. Halbuki her iki öğrencinin ortalama başarı düzeyi aynıdır. Şu halde, bir serinin aldığı değerlerin ne kadar değişken olduğunun bilinmesi gerekir.

İstatistikte bir serinin değişkenliği kavramı, fizikte cisimlerin yoğunluk kavramına benzetilebilir. Serilerdeki değişkenliğin derecesini ölçen çeşitli ölçüler geliştirilmiştir. Bunlardan değişim aralığı, kartiller arası fark, ortalama sapma, varyans, standart sapma ve değiş

Değişim Aralığı

Değişim aralığına açıklık veya range adı da verilmektedir. Bir *X* serisinin (değişkeninin) değişim aralığı, serideki en büyük değer $\left(X_{\max}\right)$ ile en küçük değer $\left(X_{\min}\right)$ arasındaki farktır. Yani, $DA = X_{\max} - X_{\min}$. Yukarıdaki örneğimizde $DA_x = 63-58=5$ ve $DA_y = 98-12=86$ bulunur. Görüldüğü gibi, *Y*'nin *X*'e nazaran değişkenliği büyüktür.

Ortalama Sapma

Değişim aralığı, sadece serideki iki değeri (maksimum ve minimum değerleri) dikkate almakta fakat bu ikisi arasında kalan değerlerin değişkenliği hakkında bir fikir vermemektedir. Kartiller arası fark da aynı şekilde sadece iki değeri dikkate almaktadır. Bu yetersizliği ortadan kaldıracak bir değişkenlik ölçüsü ortalama sapmadır. Ortalama sapma seride yer alan bütün değerlerin değişkenliğini dikkate alan bir değişkenlik ölçüsüdür. Ortalama Sapma, serideki bütün değerlerin aritmetik ortalamadan mutlak sapmalarının aritmetik ortalamasıdır. Formül olarak;

$$OS = \frac{\sum |X_i - \mu|}{N}$$

ile ifade edilir. Burada aritmetik ortalama μ ile gösterilmiştir.

Örnek : X_i : 3, 4, 2, 6, 8, 7 serisi verilmiştir. Bu serinin ortalama sapmasını bulalım. Önce seri aritmetik ortalaması μ bulunmalıdır.

$$\mu = \frac{\sum X_i}{N} = \frac{1}{6}(3+4+2+6+8+7) = 5$$

olarak bulunur. Terimlerin ortalamadan sapmaları ve bunların mutlak değerleri aşağıdaki tablo ile verilmiştir.

X_i	$X_i - \mu$	$\lvert X_i - \mu \rvert$
3	$3-5=-2$	2
4	$4-5=-1$	1
2	$2-5=-3$	3
6	$6-5=1$	1
8	$8-5=3$	3
7	$7-5=2$	2

Tabloda son sütunun toplamından $\sum |X_i - \mu| = 12$ bulunur. Buna göre bu serinin ortalama sapması;

$$OS = \frac{\sum |X_i - \mu|}{N} = \frac{12}{6}$$

$$OS = 2$$

olarak hesaplanır.

Varyans ve Standart Sapma

Varyans σ^2 notasyonu ile gösterilir. Standart sapma ise varyansın pozitif kareköküdür. Yani standart sapma da σ ile gösterilir. Bir serinin varyansı, o serideki tüm terimlerin aritmetik ortalamadan farklarının karelerinin aritmetik ortalamasıdır. Buna göre varyans ve standart sapma formülü sırasıyla;

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

şeklinde ifade edebiliriz.

Örnek : X_i : 213, 219, 252, 243, 222, 267, 215 serisi bir gazete bayisinde bir hafta boyunca günlük gazete satışına ilişkin rakamlardır. Gazete bayisinin sattığı günlük gazete sayısındaki değişkenliğin ölçüsünü varyans ve standart sapma cinsinden bulalı. Serinin ortalaması;

$$\mu = \frac{1}{7}.(213 + 219 + 252 + 243 + 222 + 267 + 215)$$

$$\mu = 233$$

olarak bulunur. Varyans için gerekil değerler tablo halinde verilmiştir.

X_i	$X_i - \mu$	$(X_i - \mu)^2$
213	-20	400
219	-14	196
252	19	361
243	10	100
222	-11	121
267	34	1156
215	-18	324

Son sütunun toplamı $\sum\left(X_i-\mu\right)^2=2658$ olmaktadır. Buna göre günlük gazete satışlarındaki varyans,

$$\sigma^2=\frac{\sum\left(X_i-\mu\right)^2}{N}=\frac{2658}{7}$$

$$\sigma^2=379.7$$

olarak bulunur. Serinin standart sapması ise;

$$\sigma=\sqrt{379.7}=19.5$$

olur.

Değişim Katsayısı

Bir serideki değişkenlik mutlak olarak ölçüldüğünde bir anlam ifade etmeyebilir. Bu bakımdan, serideki değişkenliğin serideki rakamların büyüklüğüne nispetle ölçülmesi daha anlamlı olabilir. Değişim katsayısı, bir serideki değişkenliği nispi olarak ölçer. Bir serinin standart sapmasının serinin aritmetik ortalamasına oranı değişim katsayısını verir. Yani;

$$DK=\frac{\sigma}{\mu}$$

Burada, standart sapma aritmetik ortalamasının bir yüzdesi olarak ifade edilmiş olmaktadır. Çıkan sonuç 100 ile çarpılırsa daha rahat yorumlanabilir.

Örnek : Yukarıda verdiğimiz X_i: 213, 219, 252, 243, 222, 267, 215 (gazete satışları) serisinde aritmetik ortalama ve standart sapma sırasıyla, $\mu=233$, $\sigma=19.5$ olarak bulunmuştu. Bu seri için değişim katsayısı;

$$DK=\frac{\sigma}{\mu}=\frac{19.5}{233} \qquad DK=0.08$$

olarak elde edilir. Değişim katsayısını standart sapmanın aritmetik ortalama içindeki yüzdesi olarak da okuyabiliriz. Bunun için değişim katsayısı 100 ile çarpılır. Buna göre yukarıdaki örneğin değişim katsayısı % 8 olarak bulunur.

Sınıflanmış Seride Değişkenlik

Sınıflanmış serilerde kartil sınıfının bulunması ile kartiller bulunmuş olur. bunu da medyanda olduğu gibi, kartillerin sıra numarasını kümülatif frekans sütunundan takip ederek buluyoruz. Kümülatif frekans sütununda,

$\frac{(N+2)}{4}$ sıra numarasının içerildiği ilk sınıf 1. kartil sınıfı

$\frac{(2N+2)}{4}$ sıra numarasının içerildiği ilk sınıf 2. kartil sınıfı

$\frac{(3N+2)}{4}$ sıra numarasının içerildiği ilk sınıf 3. kartil sınıfı

olur. Bu sınıflara karşı gelen seri değerleri ilgili kartil değerini verir.

Örnek 6: Sınıflanmış bir seri kümülatif frekansları ile birlikte aşağıdaki tablo ile verilmiştir.

X	f	F
150	7	7
160	9	16
170	13	29
180	11	40
190	6	46
200	4	50

Toplam terim sayısı $N=50$'dir. Buna göre;

$\frac{(50+2)}{4}=13$ sıra numarası 1. kartilin sınıfını

$\frac{(2\times50+2)}{4}=25.5$ sıra numarası 2. kartilin sınıfını

$\frac{(3\times50+2)}{4}=38$ sıra numarası 3. Kartilin sınıfını

Belirler. Kümülatif sütunda 13 sıra numarası ikinci sınıfta 25.5 sıra numarası, üçüncü sınıfta 38 sıra numarası dördüncü sınıfta, sağlanmaktadır. Buradan,

$$Q_1 = 160, \quad Q_2 = 170, \quad Q_3 = 180$$

değerleri belirlenir. Bu seri için kartillerarası fark;

$$KAF = 180 - 160 = 20$$

ve çeyrek sapma;

$$ÇS = \frac{(180-160)}{2} = 10$$

bulunur. Bu değerlerin her ikisi de serinin değişkenliği hakkında birer ölçüdür.

Sınıflanmış bir seride ortalama sapma ve varyans ölçüleri sırasıyla şu şekilde yazılabilir.

$$OS = \frac{\sum f_j \left| X_j - \mu \right|}{\sum f_j}$$

$$\sigma^2 = \frac{\sum f_j \left(\bar{X}_j - \mu \right)^2}{\sum f_j}$$

Burada f_j lerin frekansları gösterdiğini hatırlayınız.

Örnek 7: Aşağıda verilen sınıflanmış seri için Ortalama Sapma ve Varyansı hesaplayalım.

X_j	f_j
10	4
11	2
16	1
14	3

Bu sınıflanmış seride $m = 4$ sınıf mevcuttur. Frekansların toplamı ise $\sum f_j = 10$ dur. Serinin aritmetik ortalaması;

$$\mu = \frac{\sum f_j.X_j}{\sum f_j} = \frac{(4)10+(2)11+16+(3)14}{10} = 12$$

olarak bulunur. Buna göre ara sonuçlar;

$\lvert X_i-\mu \rvert$	$f_j \lvert X_i-\mu \rvert$	$(X_i-\mu)^2$	$f_j(X_i-\mu)^2$
2	8	4	16
1	2	1	2
4	4	16	16
2	6	4	12
Toplam	20	-	46

olarak elde edilir. Buna göre, bu serinin ortalama sapması,

$$OS = \frac{\sum f_j \lvert X_i-\mu \rvert}{\sum f_j} = \frac{20}{10} = 2$$

bulunur. Serinin varyansı ve standart sapması ise;

$$\sigma^2 = \frac{\sum f_j (X_j-\mu)^2}{\sum f_j} = \frac{46}{10} = 4.6$$

$$\sigma = 2.1$$

olarak bulunur.

Gruplanmış Seride Değişkenlik

$$OS = \frac{\sum f_j \lvert m_j-\mu \rvert}{\sum f_j} \qquad \sigma^2 = \frac{\sum f_j (m_j-\mu)^2}{\sum f_j}$$

Burada $\overline{X}_j$ nin *j* inci grubun ortalaması olduğunu hatırlayınız.

Örnek : Bir meslek edindirme kursunun sonunda kursiyerlere 10 üzerinden puanlar verilmiştir. Puanların dağılımı şöyledir:

Puanlar	Kursiyer Sayısı
0-2	3
2-4	7
4-6	13
6-8	11
8-10	6

Kursiyerlerin başarı derecelerindeki değişkenliği ortalama sapma ve varyans cinsinden hesaplayalım. OS ve varyans hesabı için gerekli ara sonuçlar arka sayfadaki tabloda özetlenmiştir. Bu tablodaki ara değerler yardımı ile gruplanmış seride ortalama bulan formülden;

$$\mu = \frac{\sum f_j m_j}{\sum f_j} \qquad \mu = \frac{220}{40} = 5.5$$

olarak hesaplanır.

m_j	f_j	$f_j m_j$	$\lvert m_j - \mu \rvert$	$f_j \lvert m_j - \mu \rvert$	$(m_j - \mu)^2$	$f_j (m_j - \mu)^2$
1	3	3	4.5	13.5	20.25	60.75
3	7	21	2.5	17.5	6.25	43.75
5	13	65	0.5	6.5	0.25	3.25
7	11	77	1.5	16.5	2.25	24.75
9	6	54	3.5	21.0	12.25	73.50
-	40	220	-	75.0	-	206.00

Tablodaki diğer ara değerlerden Ortalama Sapma,

$$OS = \frac{\sum f_j \lvert m_j - \mu \rvert}{\sum f_j} = \frac{75}{40} = 1.875$$

ve varyansı;

$$\sigma^2 = \frac{\sum f_j \left(m_j - \mu\right)^2}{\sum f_j} = \frac{206}{40} = 5.15$$

olur. Standart sapma ise;

$$\sigma = \sqrt{5.15} = 2.27$$

olarak bulunur.