

TANITICI İSTATİSTİKLER

Üzerinden çalışılan örnekten veriler doğru olarak toplanıp özetlendikten sonra araştırmacı örneğini tanıtmak, tanımlamak veya daha iyi ifade etmek için bazı değerler hesaplar. Bu değerlere **tanıtıcı veya tanımlayıcı (descriptive) istatistikler** denir. Hesaplanan tanıtıcı istatistikler, üzerinde çalışılan örnek hakkında detaylı bilgiler verir. Tanıtıcı istatistikler Tablo 2.1’de görüldüğü gibi 2 gruba ayrılır.

Tablo 2.1. Tanıtıcı istatistikler

TANITICI İSTATİSTİKLER	
YERLEŞİM (LOCATION) ÖLÇÜLERİ (ORTALAMALAR)	DEĞİŞİM (VARIATION) ÖLÇÜLERİ
1. Aritmetik Ortalama	1. Değişim Genişliği
2. Ortanca Değer (Median)	2. Ortalama sapma
3. Tepe Değeri (Mode)	3. Varyans
4. Geometrik Ortalama	4. Standart Sapma
5. Harmonik Ortalama	5. Varyasyon Katsayısı

2.1. Yerleşim Ölçüleri (Ortalamalar)

Yerleşim ölçüleri; **ortalamalar** olarak da adlandırılır. Örnekteki bireylerden elde edilen gözlem değerlerinin etrafında toplanma eğilimi gösterdikleri değeri ifade etmek amacıyla hesaplanırlar. Yerleşim ölçüleri örneği en iyi temsil eden değerlerdir.

2.1.1. Aritmetik Ortalama

Aritmetik ortalama, gözlem değerlerinin toplamının gözlem sayısına bölünmesiyle elde edilen değerdir ve;

$$\begin{array}{ccc} \text{Populasyonda} & \text{ve} & \text{Örnekte;} \\ \mu_x = \frac{\sum_{i=1}^N X_i}{N} & & \bar{X} = \frac{\sum_{i=1}^n X_i}{n} \end{array} \quad \dots(2.1)$$

şeklinde hesaplanır. Eşitlikte, X_i , i. gözlem değeri, N, populasyondaki birey sayısını ve n, örnekteki birey sayısını göstermektedir.

Sadece **ortalama** denildiği zaman kastedilen “**aritmetik ortalama**”dır. Ortalama, verilerin yapısına bağlı olarak örneği en iyi temsil eden değerdir.

Bölüm I’de, Tablo 1.2’de verilen 80 mısır bitkisinin uzunluğuna ait ortalama (2.1) numaralı eşitlik kullanılarak aşağıdaki şekilde hesaplanır:

$$\bar{X} = \frac{88+107+70+99+\dots+81+78+88+73}{80} = \frac{6633}{80} = 82.91\text{cm}$$

yani, 80 mısır bitkisinin ortalama uzunluğu 82.91 cm’dir ve bitkilerin uzunlukları bu değer etrafında toplanma eğilimindedir.

(2.1) numaralı eşitlik kullanılarak Bölüm I’de, Tablo 1.10’da verilen 40 tütün bitkisinin ortalama yaprak sayısı;

$$\bar{X} = \frac{7+6+6+6+7+\dots+7+6+7+8+6}{40} = \frac{302}{40} = 7.55 \approx 8$$

olarak bulunur, yani 40 tütün bitkisinin yaprak sayıları 8 etrafında toplanma eğilimindedir.

Bir örnek için hesaplanan aritmetik ortalamanın, örnekteki en küçük gözlem değerinden daha küçük ve en büyük gözlem değerinden daha büyük olması mümkün değildir.

2.1.1.1. Frekans Dağılım Tablosundan Ortalamanın Hesaplanması

Araştırmacılar üzerinde çalıştıkları örnekten elde ettikleri verileri frekans dağılım tablosu oluşturarak özetleyebilirler. Eğer frekans dağılım tablosu oluşturulmuşsa ortalama, frekans dağılım tablosundan (2.2) numaralı eşitlik kullanılarak hesaplanabilir.

$$\bar{X} = \frac{\sum_{i=1}^k f_i X_i}{\sum_{i=1}^k f_i} \quad \dots(2.2)$$

eşitlikte f_i , i. sınıfın frekansını, k, frekans dağılım tablosundaki sınıf sayısını, X_i , i. sınıfın sınıf değerini göstermektedir.

(2.2) numaralı eşitlikte görüldüğü gibi, frekans dağılım tablosundan ortalama hesaplanırken, her bir sınıf değerinin aritmetik ortalamaya katkısı, o sınıfın frekansı kadar olup sınıf değeri söz konusu sınıfın frekansı ile tartılarak hesaplamaya katılır.

80 mısır bitkisinin uzunluğu için Tablo 1.3'te düzenlenen frekans dağılım tablosundan ortalama, (2.2) numaralı eşitlik kullanılarak Tablo 2.2'de görüldüğü şekilde hesaplanır.

Tablo 2.2. 80 mısır bitkisi için ortalamanın frekans dağılım tablosundan hesaplanması

Sınıf değeri	Mutlak frekans	$f_i X_i$
52.5	2	105.0
58.5	3	175.5
64.5	7	451.5
70.5	10	705.0
76.5	10	765.0
82.5	15	1237.5
88.5	10	885.0
94.5	8	756.0
100.5	10	1005.0
106.5	3	319.5
112.5	2	225.0
Toplam	$\sum f_i = 80$	$\sum f_i X_i = 6630.0$
Ortalama bitki uzunluğu, $\bar{X} = \frac{6630}{80} = 82.875$ cm olarak bulunur.		

Gözlem değerleri tek tek toplanarak ortalama 82.91 cm olarak hesaplanmış olmasına karşın, frekans dağılım tablosundan hesaplandığı zaman 82.87 cm olarak bulunmuştur. Hesaplanan ortalamalar

arasındaki bu küçük fark, frekans dağılım tablosundan ortalamanın hesaplanmasında gerçek değerler yerine her sınıfı temsil eden sınıf değerlerinin kullanılmasından kaynaklanmaktadır.

40 tütün bitkisinin yaprak sayısı için Tablo 1.11’de düzenlenen frekans dağılım tablosu kullanılarak ortalama yaprak sayısı ise (2.2) numaralı eşitlik kullanılarak Tablo 2.3’de görüldüğü şekilde hesaplanır.

Tablo 2.3. 40 tütün bitkisinin yaprak sayısı için ortalama yaprak sayısının frekans dağılım tablosundan hesaplanması

Yaprak sayıları (X_i)	Bitki sayısı (f_i)	$f_i X_i$
6	11	66
7	11	77
8	10	80
9	4	36
10	2	20
11	1	11
12	1	12
Toplam	$\sum f_i = 40$	$\sum f_i X_i = 302$
Ortalama yaprak sayısı, $\bar{X} = \frac{302}{40} = 7.55 \cong 8$ olarak bulunur. Hesaplanan bu ortalama tartılı ortalamadır çünkü yaprak sayıları bitki sayısı ile tartılmıştır.		

2.1.1.2. Aritmetik Ortalamanın Özellikleri

1. Bir örnekteki gözlem değerlerinin örnek ortalamasından farkları alınır ve toplanırsa bu farkların (sapmaların) toplamı sıfır olarak bulunur. Yani bir örneğin ağırlık merkezi veya denge noktası ortalamadır.

$$\sum_{i=1}^n (X_i - \bar{X}) = 0 \text{ 'dır.}$$

2. Üzerinde çalışılan örnekteki gözlem değerlerinin hepsine **A** gibi bir sabit sayı eklenirse, elde edilen değerlerin ortalaması, sabit sayı eklenmeden önceki ortalamadan **A** kadar fazla olur.

$$y_i = x_i + A \text{ ise}$$

$$\bar{y} = \bar{x} + A \text{ olarak bulunur.}$$

3. Üzerinde çalışılan örnekteki gözlem değerlerinin hepsinden **A** gibi bir sabit sayı çıkarılırsa, elde edilen değerlerin ortalaması, sabit sayı eklenmeden önceki ortalamadan **A** kadar az olur.

$$y_i = x_i - A \text{ ise}$$

$$\bar{y} = \bar{x} - A \text{ olarak bulunur.}$$

4. Üzerinde çalışılan örnekteki gözlem değerlerinin hepsi **c** gibi bir sabit ile çarpılırsa, elde edilen değerlerin ortalaması sabit sayı ile çarpılmadan önceki ortalamanın **c** katı kadardır:

$$y_i = c x_i$$

$$\bar{y} = c \bar{x}$$

5. Üzerinde çalışılan örnekteki gözlem değerlerinin hepsi c gibi bir sabite bölünürse, elde edilen değerlerin ortalaması sabit sayıya bölünmeden önceki ortalamanın c ile bölümüne eşit olacaktır:

$$y_i = \frac{x_i}{c}$$

$$\bar{y} = \frac{\bar{x}}{c}$$

ÖRNEK:

Aşağıdaki örneklerden A örneğinin ortalaması 3 olduğuna göre, diğer örneklerin (B, C; D; E) ortalamalarını, ortalamanın özelliklerinden yararlanarak bulunuz.

A	B	C	D	E
1	11	0.5	10	4
2	12	1.0	20	14
3	13	1.5	30	24
4	14	2.0	40	34
5	15	2.5	50	44

1. Her zaman bir örnekteki gözlem değerlerinin kendi ortalamasından sapmalarının toplamı sıfırdır.

Örneğin, $\bar{A} = 3$ 'tür. Bu durumda;

$$(1-3)+(2-3)+(3-3)+(4-3)+(5-3)=0 \text{ olacaktır.}$$

2. B örneğindeki gözlem değerleri, $B=A+10$ şeklinde elde edilmiştir. Bu sebeple B örneğinin ortalaması,

$$\bar{B} = \bar{A} + 10 = 13 \text{ olacaktır.}$$

3. C örneğindeki gözlem değerleri, $C = \frac{A}{2}$ şeklinde elde edilmiştir. Bu sebeple C örneğinin ortalaması,

$$\bar{C} = \frac{\bar{A}}{2} = \frac{3}{2} = 1.5 \text{ olacaktır. C örneğindeki gözlem değerleri } C=(B-10)/2 \text{ şeklinde de elde edilebilir.}$$

Bu durumda ise C örneğinin ortalaması, $\bar{C} = \frac{(\bar{B} - 10)}{2} = \frac{(13 - 10)}{2} = 1.5$ olarak hesaplanır.

4. D örneğindeki gözlem değerleri, $D=A \times 10$ şeklinde elde edilmiştir. Bu sebeple B örneğinin ortalaması, $\bar{D} = \bar{A} \times 10 = 30$ olacaktır.

5. E örneğindeki gözlem değerleri, $E=D-6$ şeklinde elde edilmiştir. Bu sebeple B örneğinin ortalaması, $\bar{E} = \bar{D} - 6 = 24$ olacaktır.

2.1.1.3. Kısa Yoldan Aritmetik Ortalamanın Hesaplanması

Tablo 2.2'de, ortalamanın frekans dağılım tablosundan nasıl hesaplandığı gösterilmiştir. Bu hesaplamaların tamamlanması rakamların duyarlılığına bağlı olarak hem uzun zaman almakta hem de hata yapma olasılığını yükseltmektedir. Bu sebeple aritmetik ortalamanın özelliklerinden yararlanarak aritmetik ortalama, daha az zamanda ve daha az hata ile kısa yoldan hesaplanabilir.

Kısa yoldan aritmetik ortalama hesaplanmasında temel, işlem yapması zor olan sınıf değerlerini (2.3) numaralı eşitliği kullanarak, daha kolay işlem yapılabilecek tamsayı değerlere dönüştürmektir.

$$b_i = \frac{X_i - A}{c} \quad \dots(2.3)$$

(2.3) numaralı eşitlikte, X_i , i. sınıfın sınıf değeri, A, frekans dağılım tablosunda ortaya en yakın veya ortadaki sınıfın sınıf değeridir. A değeri, araştırmacının seçtiği bir değerdir ve bunu seçerken dikkat edilmesi gereken bir kural yoktur. Sadece, A değeri seçilirken, ne kadar ortadaki bir sınıfın sınıf değeri seçilirse, hesaplanan b değerleri o kadar küçük ve işlem yapması da o kadar kolay olur. c, frekans dağılım tablosundaki sınıf aralığıdır.

(2.3) numaralı eşitlikten i. sınıfın sınıf değeri, $X_i = A + b_i c$ şeklinde yazılabilir. Dolayısı ile ortalamanın özellikleri kullanılarak $\bar{X} = A + c\bar{b}$ olarak yazılır. Bu eşitlik kullanılarak \bar{X} 'nin hesaplanabilmesi için, önce (2.4) numaralı eşitlik kullanılarak her sınıf için hesaplanan “b” değerlerine ait aritmetik ortalamanın hesaplanması gerekir.

$$\bar{b} = \frac{\sum_{i=1}^k f_i b_i}{\sum_{i=1}^k f_i} \quad \dots(2.4)$$

“b” değerlerinin ortalamasının hesaplanması için $f_i b_i$ çarpımlarının toplamı bulunurken cebirsel toplam yapılması, yani “b” değerlerinin işaretinin de dikkate alınması gerekir.

80 mısır bitkisinin uzunluğu için Tablo 1.3'te düzenlenen frekans dağılım tablosu kullanılarak kısa yoldan aritmetik ortalama Tablo 2.4'de görüldüğü şekilde hesaplanır.

Tablo 2.4. 80 mısır bitkisi için ortalamanın frekans dağılım tablosundan hesaplanması

Sınıf değeri	Mutlak frekans	b_i	$f_i b_i$
52.5	2	-5	-10
58.5	3	-4	-12
64.5	7	-3	-21
70.5	10	-2	-20
76.5	10	-1	-10
82.5=A	15	0	0
88.5	10	1	10
94.5	8	2	16
100.5	10	3	30
106.5	3	4	12
112.5	2	5	10
Toplam	$\sum f_i = 80$		$\sum f_i b_i = 5$

$$\sum f_i b_i = (-10) + (-12) + (-21) + (-20) + (-10) + (0) + (10) + (16) + (30) + (12) + (10) = 5$$

$$\text{ve } \bar{b} = \frac{5}{80} = 0.0625 \text{ olarak bulunur.}$$

$\bar{X} = A + c\bar{b}$ olduğu için, 80 mısır bitkisinin ortalama uzunluğu $\bar{X} = 82.5 + (6) \times (0.0625) = 82.875 \text{ cm}$ olarak hesaplanır. Bu ortalama, Tablo 2.2'de frekans dağılım tablosundan hesaplanan ortalama ile aynıdır. Frekans dağılım tablosundan ne şekilde hesaplanırsa hesaplanırsa ortalama değişmez.

(2.3) numaralı eşitlik kullanılarak her sınıf için “b” değerinin hesaplanabilmesi için ilk olarak A değerinin seçilmesi gerekir. Bu örnekte en ortadaki ve frekansı en yüksek olan sınıfın sınıf değeri (82.5) A değeri olarak seçilmiştir. Farklı bir sınıfın sınıf değerinin A değeri olarak seçilmesi, hesaplanacak olan ortalamayı değiştirmez. Sadece “b” değerlerinin daha büyük ve işlemlerin daha uzun zaman almasına sebep olur. Bu örnek için Sınıf aralığı, c, ise 6’dır. A değeri olarak seçilen sınıfın b değeri (2.3) numaralı eşitlikten $b=(82.5-82.5)/6=0$ olarak bulunur. Bir sonraki sınıf için b değeri, $b=(88.5-82.5)/6=1$, bir önceki sınıf için b değeri ise $b=76.5-82.5=-1$ dir. Tablo 2.4’ten de kolayca görülebileceği gibi her frekans dağılım tablosu için geçerli olmak üzere, A değeri olarak seçilen sınıfın b değerinden 1 birim çıkararak önceki sınıfların, 1 birim ekleyerek de sonraki sınıfların b değerleri oluşturulabilir. Her sınıf için “b” değerleri ve ortalamanın kısa yoldan hesaplanması Tablo 2.4’de gösterilmiştir.

2.1.2. Ortanca Değer (Median)

Ortanca değer, büyüklüklerine göre sıralanmış verilerin en ortasındaki değerdir. Örnekteki gözlem değerlerinin yarısı bu değere eşit veya küçük diğer yarısı ise bu değere eşit veya büyüktür. Üzerinde çalışılan örnekte tek sayıda gözlem değeri varsa ortanca değer $\frac{(n+1)}{2}$. gözlemin değeridir. Üzerinde çalışılan örnekte çift sayıda gözlem değeri varsa bu durumda ortanca değer, $\frac{n}{2}$. gözlem değeri ile $(\frac{n}{2} + 1)$. gözlem değerinin ortalamasıdır.

ÖRNEK:

Örnek genişliği 7 olan bir örnekten elde edilen gözlem değerleri; 9, 11, 5, 6, 8, 7, 6 şeklinde ise ortanca değerin bulunması için gözlem değerlerinin büyüklüklerine göre sıralanması gerekir. Gözlem değerleri küçükten büyüğe veya büyükten küçüğe doğru sıralanabilir. Büyüklüklerine göre sıralanması denince genellikle izlenen yol küçükten büyüğe doğru sıralamaktır. Gözlem değerleri; 5, 6, 6, 7, 8, 9, 11 şeklinde sıralandıktan sonra en ortadaki, yani baştan itibaren 4. değer ortanca değerdir ve 7’ye eşittir. Gözlem değerleri büyüklüklerine göre sıralandığı zaman, örnekte tek sayıda gözlem değeri olduğu için, $\frac{(n+1)}{2} = \frac{(7+1)}{2} = 4$. gözlem değerinin, yani 7’nin ortanca değer olduğu kolayca görülebilir.

Eğer örnek genişliği 8 ve gözlem değerleri; 9, 11, 5, 6, 8, 7, 6, 12 şeklinde ise yine gözlem değerleri ilk olarak büyüklüklerine göre 5, 6, 6, 7, 8, 9, 11, 12 şeklinde sıralanır. Örnekte çift sayıda gözlem değeri olduğu için ortanca değer, $\frac{n}{2} = \frac{8}{2} = 4$. gözlem değeri ile $(\frac{n}{2} + 1) = \frac{8}{2} + 1 = 5$. gözlem değerinin ortalamasıdır. Bu durumda ortanca değer, $OD = \frac{(7+8)}{2} = 7.5$ olarak hesaplanır.

2.1.2.1. Frekans Dağılım Tablosundan Ortanca Değerin Hesaplanması

Üzerinde durulan özellik ile ilgili toplanan gözlem değerleri için frekans dağılım tablosu düzenlenmiş ise ortanca değer de frekans dağılım tablosundan hesaplanabilir. Frekans dağılım tablosunda ortanca değer $(\frac{n}{2})$. gözlem değerinin bulunduğu sınıftadır. Örneğin, 80 mısır bitkisinin

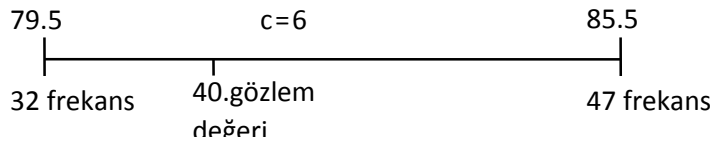
ZZT201 BİYOMETRİ DERSİ 3. HAFTA DERS NOTLARI

BÖLÜM 2

Kaynak Kitap: TEMEL BİYOMETRİ (2013) Kocabaş Z., Özkan, MM ve Başpınar E.
Ankara Üniversitesi Ziraat Fakültesi Yayın No:1606 Ders Kitabı 558

uzunlukları toplandığı zaman ortanca değer, büyüklüklerine göre sıralanmış gözlem değerlerinin en ortasındaki değerdir. 80 gözlem değerini büyüklüklerine göre sıralayarak ortanca değeri hesaplamak yerine düzenlenmiş olan frekans dağılım tablosu kullanılarak ortanca değer “...den daha az” eklemeli frekanslarından yararlanılarak hesaplanabilir. Çünkü “...den daha az” eklemeli frekanslar düzenlendiği zaman, gözlem değerleri büyüklüklerine göre sıralandırılmıştır. 80 mısır bitkisinin uzunluğu için Tablo 1.5’de düzenlenen “...den daha az” eklemeli frekansları kullanılarak ortanca değer hesaplanması Tablo 2.5’de verilmiştir.

Tablo 2.5’de verilmiş olan “...den daha az” eklemeli frekanslarından 40. gözlem değerinin 6. ve 7. sınıfın alt sınırları arasında olduğu görülmektedir. Bu durumda ortanca değer aşağıdaki şekilde hesaplanır:



Tablo 2.5. 80 mısır bitkisinin uzunluğu için “...den daha az” eklemeli frekansları kullanılarak ortanca değer hesaplanması

Alt gerçek sınır	Mutlak frekans	...den daha az eklemeli frekans	
49.5	2	0	
55.5	3	2	
61.5	7	5	
67.5	10	12	
73.5	10	22	
79.5	15	32	
85.5	10	47	
91.5	8	57	
97.5	10	65	
103.5	3	75	
109.5	2	78	
115.5		80	

Uzunluklar sıralandığı zaman 40. uzunluk ortanca değerdir. “...den daha az” eklemeli frekansları 40. gözlem değerinin 79.5 ile 85.5 alt gerçek sınırları arasında olduğunu göstermektedir.

Burada 6 birimlik sınıf aralığında 15 gözlem değerinin eşit aralıklarla dağıldığı varsayılır. Bu varsayım doğrultusunda, ortanca değer yani 40. gözlem değeri, 32. gözlem değerinden yani 79.5 cm’den 8 frekansın kapladığı sınıf aralığı kadar fazladır. 8 gözlem değerinin kapladığı sınıf aralığı aşağıdaki orantı kurularak hesaplanır:

$$\begin{array}{lcl} 6 \text{ sınıf aralığında} & 15 \text{ frekans} & \\ x & 8 \text{ frekans} & \end{array}$$

$$X = \frac{(6)(8)}{15} = 3.2$$

Hesaplanan bu sınıf aralığı 79.5'e eklenir ve Ortanca değer=79.5+3.2=82.7 cm olarak bulunur.

(2.5) numaralı eşitlik, yukarıda açıklanan çözümün formül şeklinde ifadesidir ve frekans dağılım tablosundan ortanca değer (2.5) numaralı eşitlik kullanılarak da hesaplanabilir.

$$\text{Ortanca Değer} = L_1 + \frac{(n:2) - (\sum f)_{OD}}{f_{OD}} c \quad \dots(2.5)$$

(2.5) numaralı eşitlikte, L_1 , ortanca değer bulunduğ sınıfın alt gerçek sınırını, $(\sum f)_{OD}$, ortanca değer bulunduğ sınıftan önceki sınıfların frekansları toplamını, f_{OD} , ortanca değer bulunduğ sınıfın frekansını, n , örnek genişliğini ve c , sınıf aralığını göstermektedir. (2.5) numaralı eşitlik kullanılarak Tablo 2.5'deki frekans dağılım tablosundan ortanca değer;

$$\text{Ortanca değer} = 79.5 + \frac{80/2 - 32}{15} (6) = 82.7 \text{ cm dir.}$$

Eğer 80 mısır bitkisinin uzunluğu küçükten büyüğe doğru sıralanırsa 40. ve 41. gözlem değerlerinin orta noktası, yani ortanca değer $\frac{82+83}{2} = 82.5 \text{ cm}$ olarak bulunur. Frekans dağılım tablosundan hesaplanan ortanca değer ile bu değer arasındaki önemli olmayan bu fark, frekans dağılım tablosunda verilerin özetlenmesinden yani verilerin gerçek değerlerinin kullanılmamasından ileri gelir.

Ortanca değer hesaplandıktan sonra birinci çeyrek (first quartile), yani en kısa uzunluğa sahip mısır bitkilerinin ilk %25'nin hangi uzunluktan daha kısa olduğu hesaplanabilir. En kısa uzunluğa sahip mısır bitkilerinin %25'i örneğimizde 80 adet mısır bitkisine ait uzunluklar olduğundan (80×0.25) 20. mısır bitkisinin uzunluğundan daha kısadır ve (2.5) numaralı eşitlikten;

$$\text{İlk çeyrek değer} = 73.5 + \frac{80/4 - 19}{22} (6) \cong 73.77 \text{ cm}$$

olarak hesaplanır. Eğer gözlem değerleri sıralanarak ilk çeyrek hesaplanırsa 73 cm olarak bulunur. Bu fark frekans dağılım tablosu yapılırken gözlem değerlerinin özetlenmesinden ileri gelir.

İlk çeyrek değere benzer şekilde üçüncü çeyrek değer (third quartile), yani en uzun mısır bitkilerinin %25'nin hangi uzunluktan daha uzun olduğu hesaplanabilir. En uzun bitkilerin %25'i 60. mısır bitkisinin uzunluğundan daha uzundur ve (2.5) numaralı eşitlikten;

$$\text{Üçüncü çeyrek değer} = 91.5 + \frac{80(3/4) - 57}{8} (6) = 93.75 \text{ cm}$$

olarak hesaplanır. Eğer gözlem değerleri sıralanarak üçüncü çeyrek hesaplanırsa 93 cm olarak bulunur. Bu fark frekans dağılım tablosu yapılırken gözlem değerlerinin özetlenmesinden ileri gelir.

2.1.2.2. Ortanca Değerin Kullanıldığı Durumlar

Üzerinde çalışılan her örnekten elde edilen gözlem değerleri için aritmetik ortalama her zaman iyi bir tanıtıcı değer olmayabilir. Bazı durumlarda örnekten elde edilen gözlem değerleri için ortanca değer hesaplanması gerekir. Ortanca değer hesaplanması gereken durumlar aşağıdaki gibi sıralanabilir:

1. Toplanan gözlem değerleri arasında büyük çarpıklıklar varsa, yani gözlem değerlerinden bazıları diğerlerine nazaran çok küçük veya çok büyük ise ortanca değerin hesaplanması gerekir. Çünkü aritmetik ortalama, gözlem değerlerinin toplamının gözlem sayısına bölümü olduğundan sapan değerlerden etkilenir.

Örneğin, 9 öğrencinin haftalık masrafları **65, 75, 59, 60, 65, 55, 70, 75, 250** TL olarak saptamış ise bu örnek için aritmetik ortalama;

$$\bar{X} = \frac{65 + 75 + 59 + 60 + 65 + 55 + 70 + 75 + 250}{9} = \frac{746}{9} \approx 82.89 \text{ olarak hesaplanır. Bu iyi bir}$$

tanıtıcı değer değildir. Çünkü gözlem değerleri içindeki 250 TL aritmetik ortalamanın büyük hesaplanmasına sebep olmuştur ve gözlem değerlerinin çoğunluğu bu değer etrafında toplanma eğiliminde değildir. Dolayısıyla bu örnek için ortalamanın yerine ortanca değerin hesaplanması gerekir. Bu örnek için ortanca değer, gözlem değerleri 55, 59, 60, 65, 65, 70, 75, 75, 250 şeklinde sıralandıktan sonra 65 TL olarak bulunur, yani öğrencilerin yarısı 65 TL'den daha az, diğer yarısı ise 65 TL'den daha çok masraf yapmaktadır. Bu değer, gözlem değerlerine bakıldığında gerçekten örneği daha iyi temsil eden bir değerdir.

2. Yürütülen bir çalışmada örnekteki bütün bireylerden veri toplanması çok uzun zaman gerektiriyorsa ve araştırmacının bu kadar uzun zaman araştırmaya devam etmesi mümkün değilse ortanca değer hesaplanır. Örneğin, tavuklarda cinsi olgunluk yaşı tespit edilirken tüm tavukların ilk yumurtalarını yapmalarını beklemek uzun zaman alacağından hatta bazı tavukların kısırılık sebebiyle hiç yumurtlamamasından üzerinde çalışılan tavukların yarısı yumurtaya gelince son yumurtlayan hayvanın yaşı ile ondan sonra yumurtlayacak hayvanın yaşının ortalaması ortanca değer olarak alınır. Tüm tavukların yumurtaya gelmesi beklenmeden örnek için cinsi olgunluk yaşı bulunmuş olur.

3. Toplanan gözlem değerleri için açık uçlu frekans dağılım tablosu düzenlenmiş ise aritmetik ortalama hesaplanamaz. Çünkü frekans dağılım tablosundan aritmetik ortalamanın hesaplanabilmesi için her sınıfın sınıf değerinin bilinmesi gerekir. Halbuki açık uçlu frekans dağılım tablolarında ucu açık olan sınıfların sınıf değeri bilinmez ve bu sebepten aritmetik ortalamanın hesaplanması mümkün olmaz. Dolayısıyla açık uçlu frekans dağılım tablolarında da ortalamanın yerine ortanca değerin hesaplanması gerekir.

Örneğin, 250 çiftçinin yaşları için Tablo 1.17'de düzenlenen açık uçlu frekans dağılım tablosunda araştırmacının ortanca değeri Tablo 2.6'da görüldüğü şekilde hesaplaması gerekir.

TABLO 2.6. 250 çiftçinin yaşlarına ait açık uçlu frekans dağılım tablosunda ortanca değerin hesaplanması

Yaş Grupları	Birey sayısı (f)	250 çiftçinin yaşlarına ait frekans dağılım tablosunda büyüklüklerine göre sıralandığı zaman 125. çiftçinin yaşı ortanca değerdir. Sınıflardaki frekanslar toplandığı zaman 125. çiftçi frekansı 59 olan sınıftadır ve bu sınıfın alt gerçek sınırı 34.5'tur. (2.5) numaralı eşitlikten;
<25	9	
25-29	47	
30-34	55	
35-39	59	
40-44	22	
45-49	19	
50-54	17	

55-59	15	$OD = 34.5 + \frac{250/2 - 111}{59}(5) \cong 35.69$ bulunur.
>59	7	

2.1.3. Tepe Değeri (Mode)

Örnekten toplanan gözlem değerleri arasında en çok tekrarlanan, yani frekansı en fazla olan gözlem değeri, **tepe değeri (mode)** olarak adlandırılır. Örneğin, üzerinde çalışılan gözlem değerleri **9, 11, 5, 6, 8, 7, 6, 12, 6, 9, 6** şeklinde ise bu örneğin tepe değeri 6'dır. Çünkü 6 değeri en fazla (3 kez) tekrarlanmıştır.

Eğer gözlem değerleri için frekans dağılım tablosu oluşturulmuş ise tepe değeri frekans dağılım tablosundan hesaplanır. Frekans dağılım tablosunda, tepe değeri frekansı en fazla olan sınıftır ve (2.6) numaralı eşitlik kullanılarak hesaplanır.

$$\text{Tepe Değeri} = L_1 + \frac{d_1}{d_1 + d_2} c \quad \dots(2.6)$$

eşitlikte, L_1 , tepe değerinin bulunduğu sınıfın alt gerçek sınırı, d_1 , tepe değerinin bulunduğu sınıfın frekansı ile bir önceki sınıfın frekansı arasındaki fark, d_2 , tepe değerinin bulunduğu sınıfın frekansı ile bir sonraki sınıfın frekansı arasındaki fark, c ise sınıf aralığıdır.

80 mısır bitkisinin uzunluğu için tepe değerinin hesaplanması Tablo 2.7'de verilmiştir.

Tablo 2.7. 80 mısır bitkisinin uzunluğu için tepe değerinin hesaplanması

Alt gerçek sınır	Mutlak frekans	
49.5	2	<p>80 mısır bitkisinin uzunluğuna ait tepe değeri frekansı 15 olan sınıftır. Çünkü mutlak frekanslar içinde en fazla frekans 15 dir ve (2.6) numaralı eşitlikten;</p> <p>Tepe Değeri = $79.5 + \frac{(15-10)}{(15-10) + (15-10)}(6) = 82.5$ cm</p> <p>olarak hesaplanır.</p>
55.5	3	
61.5	7	
67.5	10	
73.5	10	
79.5	15	
85.5	10	
91.5	8	
97.5	10	
103.5	3	
109.5	2	

Düzenlenen bir frekans dağılım tablosunda en yüksek frekansa sahip birden fazla sınıf olabilir. Eğer en yüksek frekansa sahip sınıflar ardışık sınıflarda ise bir tane tepe değeri hesaplanır. Eğer en yüksek frekansa sahip sınıflar ardışık sınıflar değilse en yüksek frekansa sahip sınıf sayısı kadar tepe değeri hesaplanır. Bu iki durum için tepe değerinin hesaplanması Tablo 2.8 ve 2.9'da gösterilmiştir.

Tablo 2.8. En yüksek frekansa sahip sınıfların ardışık olması durumunda tepe değerinin hesaplanması

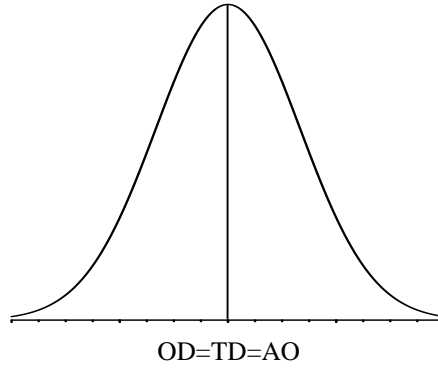
Alt gerçek sınır	Frekans	<p>Frekans dağılım tablosunda ardışık iki sınıfın frekansı 10 olup en yüksek frekanstır. Bu durumda bu iki sınıf 47.45’den başlayıp 51.45’te biten tek bir sınıf olarak dikkate alınır ve (2.6) numaralı eşitlikten, alt gerçek sınır en fazla frekansa sahip ilk sınıfın alt gerçek sınırı olmak ve sınıf aralığının da 2 katı alınmak üzere;</p> $\text{Tepe Değeri} = 47.45 + \frac{(10-6)}{(10-6)+(10-7)} (2 \times 2) \cong 49.74$ <p>olarak hesaplanır.</p>
41.45	3	
43.45	5	
45.45	6	
47.45	10	
49.45	10	
51.45	7	
53.45	6	
55.45	3	

Tablo 2.9. En yüksek frekansa sahip sınıfların ardışık olmaması durumunda tepe değerinin hesaplanması

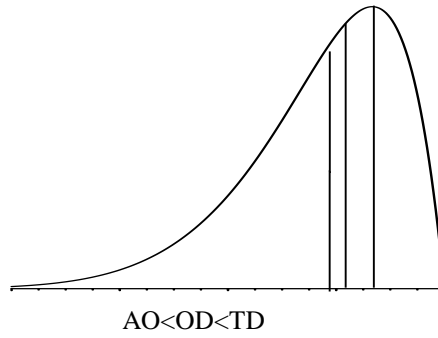
Alt gerçek sınır	Frekans	<p>Frekans dağılım tablosunda ardışık olmayan iki sınıfın frekansı 10 olup en yüksek frekanstır. Bu durumda (2.6) numaralı eşitlik kullanılarak ayrı ayrı iki tepe değeri hesaplanır:</p> $\text{Tepe Değeri}_1 = 45.45 + \frac{(10-5)}{(10-5)+(10-6)} (2) \cong 46.56$ $\text{Tepe Değeri}_2 = 49.45 + \frac{(10-6)}{(10-6)+(10-7)} (4) \cong 50.59$
41.45	3	
43.45	5	
45.45	10	
47.45	6	
49.45	10	
51.45	7	
53.45	6	
55.45	3	

2.1.4. Aritmetik Ortalama, Ortanca ve Tepe Değeri Arasındaki İlişkiler

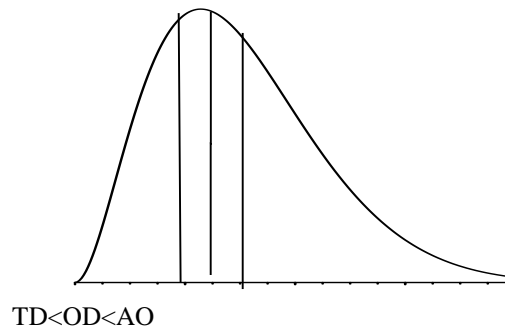
Bir araştırma yapılarak gözlem değerleri toplandığı zaman elde edilen sonuçların hem güvenilir hem de yorumlanabilir olması en uygun istatistik yöntemin seçilmesine bağlıdır. Söz konusu yöntemin uygulanmasında gözlem değerlerinin dağılımını bilmek gerekir. Aritmetik ortalama, ortanca değer ve tepe değerinin hesaplanabildiği örneklerde araştırmacı, gözlem değerlerinin dağılımı hakkında bir bilgiye sahip olur. Bu merkezi eğilim ölçülerinin sıralanışı, örnekten elde edilen gözlem değerlerinin dağılımına bağlıdır. Eğer gözlem değerleri normal dağılım gösteriyorsa dağılım simetriktir ve Şekil 2.1’de görüldüğü gibi bu üç merkezi eğilim ölçüsü birbirine eşittir. Şekil 2.2’de gösterildiği gibi aritmetik ortalama en küçük ve tepe değeri en büyük ise gözlem değerlerinin dağılımı sola yatıktır. Aritmetik ortalama en büyük ve tepe değeri en küçük ise bu durumda dağılım Şekil 2.3’de görüldüğü gibi sağa yatıktır.



ŞEKİL 2.1. Normal (Simetrik) dağılım



ŞEKİL 2.2. Sola yatık (çarpık) dağılım



ŞEKİL 2.3. Sağa yatık (çarpık) dağılım

Şekil 2.1 ve Şekil 2.2’de görüldüğü gibi gözlem değerlerinin dağılım şekline en çok etkilenen aritmetik ortalamadır. Dağılımın şekli nasıl olursa olsun ortanca değer, aritmetik ortalama ve tepe değerinin arasındadır, yani ortanca değer dağılımın şeklinden etkilenmez. Örneğin, 80 mısır bitkisinin uzunluğuna ait aritmetik ortalama 82.875 cm, ortanca değer 82.7 cm ve tepe değeri 82.5 cm olarak hesaplanmıştır. Eğer bu değerler tamsayı olarak yuvarlaklaştırılır ise 83 cm’dir ve üç merkezi eğilim ölçüsü birbirine eşittir. Bu durumda 80 mısır bitkisinin uzunluklarının simetrik bir dağılım gösterdiği kabul edilebilir.

2.1.5. Geometrik Ortalama

Geometrik ortalama, n tane pozitif değerin çarpımlarının n. dereceden köküdür ve (2.7) numaralı eşitlik kullanılarak hesaplanır.

$$G.O = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \dots X_n} = (X_1 \cdot X_2 \cdot X_3 \dots X_n)^{1/n} = \left(\prod_{i=1}^n X_i \right)^{1/n} \quad (2.7)$$

Örnek genişliği büyük ise geometrik ortalama, gözlem değerlerinin logaritmalarının ortalamasının anti logaritması olarak (2.8) numaralı eşitlik kullanılarak da hesaplanabilir.

$$\log G.O = \frac{1}{n} [(\log(x_1) + \log(x_2) + \dots + \log(x_n))] \quad \dots(2.8)$$

$$G.O = \text{antilog}(\log G.O)$$

Geometrik ortalama özellikle eşit zaman aralıklarıyla değişen oranlar için kullanılır. Örneğin, bir işçinin ücreti birbirini izleyen 4 yılda sırasıyla %5, %8, %10 ve %14 artmışsa ve bu artış oranları bir yıl önceki ücret üzerinden belirtilmişse ortalama % ücret artışı;

$$G.O = \sqrt[4]{1.05 \times 1.08 \times 1.10 \times 1.14} = 1.0920$$

% 9.20 olarak bulunur.

Genel olarak belirli bir başlangıç değeri (B), herhangi bir zaman biriminde r gibi sabit bir hızla artıyorsa, n zaman birimi sonunda varacağı değer (S);

$$S = B (1 + r)^n \quad \dots (2.9)$$

formülüyle hesaplanır.

Örneğin, belirli bir bakteri kültüründe birim alana düşen bakteri sayısı 1. günde 100'den (B) 5. günün sonunda (r), 2000'e ulaşmışsa (S) bakteri sayısı bakımından % olarak ortalama günlük artış (2.9) numaralı eşitlikten;

$2000 = 100 (1 + r)^5$ formülünden yapılacak hesaplamayla ortalama günlük artış $r = \% 82$ olarak bulunur.

2.1.6. Harmonik Ortalama

Harmonik ortalama verilerin terslerinin ortalamasının tersidir ve (2.10) numaralı eşitlik kullanılarak hesaplanır.

$$H.O = \frac{1}{\frac{1}{n} \sum \frac{1}{x_i}} = \frac{n}{\sum \frac{1}{x_i}} \quad \dots(2.10)$$

Harmonik ortalama, d/t şeklinde ifade edilen gözlem değerlerinin değişmekte olan t değerleri için kullanılır. Eğer gözlemler 1 saat de yapılan km (km (d)/saat (t)) yerine 1 km için geçen saat, 1 kg için harcanacak para (TL (d)/kg (t)) yerine 1 TL ile satın alınabilecek malzeme miktarı veya bir birimlik zamanda elde edilen adet (adet (d)/zaman (t)) yerine 1 adet için gerekli zaman şeklinde ifade edilmişse bu ve benzeri değişen değerlerin (t) hesaplanmasında harmonik ortalama kullanılır.

Örneğin, birinci işçi 5 adet fidanı 20 dakikada, ikinci işçi 24 dakikada ve üçüncü işçi 26 dakikada budamış olsa bu işçilerin ortalama iş süratini hesaplamak için kullanılacak en uygun ortalama, adet için gerekli zaman hesaplanacağından yani d (adet)/ t (zaman)' da t değiştiğinden, harmonik ortalama değildir. Bu örnek için harmonik ortalama;

$$H.O = \frac{3}{\frac{1}{20} + \frac{1}{24} + \frac{1}{26}} = 23.05 \text{ dakikadır.}$$

2.2. Değişim Ölçüleri

Üzerinde çalışılan örnekten toplanan gözlem değerleri arasında bir farklılık gözlenir. Mesela, bir sınıftaki öğrencilerin boy uzunluğu, bir ağılda yeni doğmuş kuzuların ağırlıkları, bir bahçedeki elma ağaçlarının verimleri bakımından yapılan ölçümlerin hepsinin aynı olduğuna rastlamak pek mümkün olmadığından, bu veriler arasında bir farklılığın gözlenmesi kaçınılmazdır. Bu farklılıklar gözlem değerleri arasındaki değişim (varyasyon)'dır. Bu değişimi sayısal olarak ölçebilen bir alet veya cihaz olmadığından, bir örnekten elde edilen gözlem değerleri arasındaki değişimin (varyasyonun) sayısal olarak ifadesi için kullanılan değerler DEĞİŞİM ÖLÇÜLERİ olarak adlandırılırlar.

2.2.1. Değişim Genişliği

Değişim genişliği (DG), bir örnekteki gözlem değerlerinin en büyüğü ile en küçüğü arasındaki farktır ve (2.11) numaralı eşitlik kullanılarak hesaplanır.

$$D.G = \text{En büyük gözlem değeri} - \text{En küçük gözlem değeri} \quad \dots(2.11)$$

ÖRNEK:

Örneğin, X ve Y gibi iki örnekte gözlem değerleri ve bu örnekler için hesaplanan değişim genişlikleri aşağıdaki gibi bulunmuştur.

$$X \text{ örneği: } 7, 11, 10, 12, 15 \quad DG_X = 15 - 5 = 10$$

$$Y \text{ örneği: } 14, 7, 8, 7, 15 \quad DG_Y = 15 - 5 = 10$$

Yukarıdaki örnekte görüldüğü gibi değişim genişliği iyi bir değişim ölçüsü değildir. Çünkü değişim genişliği bir örnekteki bireylerden sadece en büyük ve en küçük değerli iki gözlemi dikkate alır ve diğer gözlemleri göz ardı eder. Dolayısıyla bütün bireyler arasındaki farklılığı tam olarak yansıtmaz. Ancak gözlemlerin tamamının Değişim Genişliği kadar bir aralıkta yer aldıkları bilgisini verir. Y örneğinde bireyler arasındaki değişim daha fazla olmasına rağmen Değişim Genişliği bunu yansıtmamış, iki örneğin de aynı değişimi gösterdiği sonucuna götürmüştür.

2.2.2. Ortalama Sapma

Ortalama sapma (OS), gözlem değerlerinin kendi ortalamasından sapmalarının mutlak değerlerinin ortalamasıdır ve (2.12) numaralı eşitlik kullanılarak hesaplanır.

$$O.S = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n} \quad \dots(2.12)$$

ÖRNEK:

Yukarıda değişim genişliği hesaplanan X ve Y örnekleri için ortalama sapmalar (2.12) numaralı eşitlik kullanılarak aşağıdaki şekilde hesaplanır:

$$X \text{ örneği: } 7, 11, 10, 12, 15 \quad \bar{X} = \frac{7+11+10+12+15}{5} = 11$$

$$OS_x = \frac{|7-11|+|11-11|+|10-11|+|12-11|+|15-11|}{5} = 2.00$$

$$Y \text{ örneği: } 14, 7, 8, 7, 15 \quad \bar{Y} = \frac{14+7+8+7+15}{5} = 10.2$$

$$OS_y = \frac{|14-10.2|+|7-10.2|+|8-10.2|+|7-10.2|+|15-10.2|}{5} = 3.44$$

Ortalama sapma yukarıdaki örnekte görüldüğü gibi bireyler arasındaki değişim hakkında daha iyi bilgi verir. Bu sebeple bireyler arasındaki değişimin daha fazla olduğu Y örneği için ortalama sapma daha büyük hesaplanmıştır. Buna rağmen ortalama sapma da yaygın olarak kullanılan bir değişim ölçüsü değildir. Çünkü ortalama sapma hesaplandıktan sonra bunu, değişimi meydana getiren kaynaklara göre analiz etmek mümkün değildir. Başka bir deyişle istatistik testlere uygun değildir. Örneğin, bir sınıfta istatistik dersi notlarına ait ortalama sapma hesaplanırsa, hesaplanan bu ortalama sapma notlardaki değişimin ne kadarının öğrencilerin cinsiyet farklılığından ileri geldiğini göstermek için analiz edilemez, yani hesaplanan ortalama sapmanın ne kadarının cinsiyet farklılığından olduğu belirlenemez.

2.2.3. Varyans

İstatistik testlere uygun olan, değişimi meydana getiren kaynaklara göre analiz edilebilen ve en yaygın olarak kullanılan değişim ölçüsü varyanstır. Eğer popülasyondan hesaplanırsa parametredir, σ^2 ile gösterilir ve (2.13) numaralı eşitlik kullanılarak hesaplanır. Varyans, örnekten hesaplandığı zaman bir istatistiktir, S^2 ile gösterilir ve (2.14) numaralı eşitlik kullanılarak hesaplanır. Üzerinde çalışılan özelliğe ait gözlem değerleri için varyans hesaplandığı zaman birimi belirtilmez, yani varyansın anlamlı bir birimi yoktur.

Populasyonda:

$$\sigma_x^2 = \frac{\sum_{i=1}^N (X_i - \mu_x)^2}{N} \quad \dots(2.13)$$

Örnekte:

$$S_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)} \quad \dots(2.14)$$

(2.13) ve (2.14) numaralı eşitliklerde $\sum_{i=1}^N (X_i - \mu_x)^2$ ve $\sum_{i=1}^n (X_i - \bar{X})^2$ ifadeleri KARELER TOPLAMI olarak adlandırılır ve kısaca $\sum d_x^2$ olarak gösterilir, N popülasyondaki birey sayısı ve (n-1) örnek için serbestlik derecesidir.

Kareler toplamı ($\sum d_x^2$), gözlem değerlerinin kendi ortalamasından farklarının (sapmalarının) karelerinin toplamıdır. Örnekten varyans hesaplanırken, kareler toplamı serbestlik derecesine yani örnek

genişliğinin 1 eksiğine bölünür. Çünkü örnek üzerinde çalışılırken populasyon ortalaması (μ_x) bilinmez ve (2.13) numaralı eşitlikte μ_x yerine örnek ortalaması, yani \bar{X} kullanılır. Serbestlik derecesi bulunurken 1 çıkarılması kullanılması gereken bir parametre yerine bir istatistik kullanılmış olduğunu gösterir. Dolayısıyla SERBESTLİK DERECEŚİ, her hangi bir istatistiğin hesaplanmasında parametre yerine kullanılan her istatistik için örnek genişliğinden 1 çıkarılarak elde edilen değerdir şeklinde tanımlanır. Bu tanımlamadan anlaşılacağı üzere serbestlik derecesinin hesaplanmasında, örnek genişliğinden bazen 1 bazen 2 ve bazen de 3 çıkarılması gerekebilir.

ÖRNEK:

Değişim genişliği ve ortalama sapmaları hesaplanan X ve Y örnekleri için önce kareler toplamaları aşağıdaki şekilde hesaplanır.

X örneği: 7, 11, 10, 12, 15 $\bar{X}=11$

Y örneği: 14, 7, 8, 7, 15 $\bar{Y}=10.2$

$$\sum d_x^2 = (7-11)^2 + (11-11)^2 + (10-11)^2 + (12-11)^2 + (15-11)^2 = 34.0$$

$$\sum d_y^2 = (14-10.2)^2 + (7-10.2)^2 + (8-10.2)^2 + (7-10.2)^2 + (15-10.2)^2 = 62.8$$

Kareler toplamaları hesaplandıktan sonra X ve Y örneklerine ait varyanslar (2.14) numaralı eşitlik kullanılarak:

$$S_x^2 = \frac{\sum d_x^2}{(n-1)} = \frac{34.0}{(5-1)} = 8.5 \text{ ve } S_y^2 = \frac{\sum d_y^2}{(n-1)} = \frac{62.8}{(5-1)} = 15.7$$

olarak hesaplanır.

Örnek ortalaması tamsayı olmadığı ve örnekteki gözlem sayısı arttığı zaman kareler toplamının $\sum_{i=1}^n (X_i - \bar{X})^2$ şeklinde hesaplanması hem uzun zaman alır hem de hata yapma olasılığı fazladır. Kareler toplamının daha kolay hesaplanması için pratikte (2.15) numaralı eşitlik kullanılır.

$$\sum d_x^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \quad \dots(2.15)$$

(2.15) numaralı eşitlik kullanılarak X ve Y örnekleri için kareler toplamı aşağıdaki şekilde hesaplanır.

$$\sum d_x^2 = 7^2 + 11^2 + 10^2 + 12^2 + 15^2 - \frac{(7+11+10+12+15)^2}{5} = 34.0$$

$$\sum d_y^2 = 14^2 + 7^2 + 8^2 + 7^2 + 15^2 - \frac{(14+7+8+7+15)^2}{5} = 62.8$$

2.2.3.1. Frekans Dağılım Tablosundan Varyansın Hesaplanması

Örnekten elde edilen gözlem değerleri için frekans dağılım tablosu oluşturulmuşsa varyans (2.16) numaralı eşitlik kullanılarak frekans dağılım tablosundan hesaplanabilir.

$$S_x^2 = \frac{\sum_{i=1}^k f_i X_i^2 - \frac{(\sum_{i=1}^k f_i X_i)^2}{\sum_{i=1}^k f_i}}{\sum_{i=1}^k f_i - 1} \quad \dots(2.16)$$

(2.16) numaralı eşitlikte, k frekans dağılım tablosundaki sınıf sayısı, f_i i. sınıfın frekansı, X_i , i. sınıfın sınıf değeridir.

80 mısır bitkisinin uzunluğu için varyansın frekans dağılım tablosundan hesaplanması Tablo 2.10'da verilmiştir.

Tablo 2.10. 80 mısır bitkisinin uzunluğu için varyansın frekans dağılım tablosundan hesaplanması

Sınıf değeri	Mutlak frekans	$f_i X_i$	$f_i X_i^2$
52.5	2	105.0	5512.5
58.5	3	175.5	10266.8
64.5	7	451.5	29121.8
70.5	10	705.0	49702.5
76.5	10	765.0	58522.5
82.5	15	1237.5	102093.8
88.5	10	885.0	78322.5
94.5	8	756.0	71442.0
100.5	10	1005.0	101002.5
106.5	3	319.5	34026.8
112.5	2	225.0	25312.5
Toplam	$\sum f_i = 80$	$\sum f_i X_i = 6630.0$	$\sum f_i X_i^2 = 565326.2$
$\sum f_i$, $\sum f_i X_i$ ve $\sum f_i X_i^2$ hesaplandıktan sonra (2.15) numaralı eşitlik kullanılarak 80 mısır bitkisinin uzunluğuna ait varyans; $S_x^2 = \frac{565326.2 - \frac{(6630.0)^2}{80}}{(80-1)} \cong 200.82 \text{ olarak hesaplanır.}$			

2.2.3.2. Varyansın Özellikleri

1. Bir örnekteki gözlem değerlerin kendi ortalamalarından sapmalarının karelerinin toplamı her zaman minimumdur, yani $\sum (X_i - a)^2$ eğer $a = \bar{X}$ ise minimumdur. Ortalamadan farklı bir değerden sapmaların karelerinin toplamı her zaman ortalamadan olan sapmaların karelerinin toplamından büyüktür.

2. Örnekten elde edilen gözlem değerlerine A gibi bir sabitin eklenmesi varyansı değiştirmez. Yani;

$$y_i = x_i + A \text{ ise } \sum d_y^2 = \sum d_x^2 \text{ dolayısıyla}$$

$$S_y^2 = S_x^2$$

3. Örnekten elde edilen gözlem değerlerinden A gibi bir sabitin çıkarılması varyansı değiştirmez. Yani;

$$y_i = x_i - A \text{ ise } \sum d_y^2 = \sum d_x^2 \text{ dolayısıyla}$$

$$S_y^2 = S_x^2$$

4. Örnekten elde edilen gözlem değerleri c gibi bir sabit ile çarpılırsa, yeni değerlerin varyansı, eski değerlerin varyansı ile sabit sayının karesinin çarpımına eşit olur. Yani;

$$y_i = cx_i \text{ ise } \sum d_y^2 = c^2 \sum d_x^2 \text{ dolayısı ile}$$

$$S_y^2 = c^2 S_x^2$$

5. Örnekten elde edilen gözlem değerleri c gibi bir sabite bölünürse, yeni değerlerin varyansı, eski değerlerin varyansının sabit sayının karesine bölümüne eşittir. Yani (c sabit bir sayı olmak üzere);

$$y_i = \frac{x_i}{c} \text{ ise } \sum d_y^2 = \frac{\sum d_x^2}{c^2} \text{ dolayısı ile}$$

$$S_y^2 = \frac{S_x^2}{c^2}$$

ÖRNEK:

Aşağıdaki örneklerden A örneğinin varyansı 2.5 olduğuna göre, diğer örneklerin (B,C;D;E) varyanslarını, varyansın özelliklerinden yararlanarak bulunuz.

A	B	C	D	E
1	11	0.5	10	4
2	12	1.0	20	14
3	13	1.5	30	24
4	14	2.0	40	34
5	15	2.5	50	44

1. A örneğinin ortalaması $\bar{A} = 3$ 'tür. Bu durumda;

$\sum (X_i - 3)^2 = (1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 = 10$ olacaktır. Eğer, ortalamadan farklı bir değerden, örneğin 1'den sapmaların karelerinin toplamı hesaplanırsa;

$\sum (X_i - 1)^2 = (1-1)^2 + (2-1)^2 + (3-1)^2 + (4-1)^2 + (5-1)^2 = 30$ olacaktır. Bu da ortalamadan sapma karelerinin toplamından büyüktür.

2. B örneğindeki gözlem değerleri, $B=A+10$ şeklinde elde edilmiştir. Bu varyansı değiştirmez. Bu sebeple B örneğinin varyansı, $S_b^2 = S_a^2 = 2.5$ olacaktır.

3. C örneğindeki gözlem değerleri, $C = \frac{A}{2}$ şeklinde elde edilmiştir. Bu sebeple C örneğinin varyansı,

$$S_c^2 = \frac{S_a^2}{2^2} = \frac{2.5}{4} = 0.625 \text{ olacaktır. C örneğindeki gözlem değerleri } C=(B-10)/2 \text{ şeklinde de elde}$$

edilebilir. Bu durumda, B örneğindeki gözlem değerlerinden 10 çıkarılması varyansı etkilemez ve

$$C \text{ örneğinin varyansı, } S_c^2 = \frac{S_b^2}{2^2} = \frac{2.5}{4} = 0.625 \text{ olarak hesaplanır.}$$

4. D örneğindeki gözlem değerleri, $D=A \times 10$ şeklinde elde edilmiştir. Bu sebeple D örneğinin varyansı;

$$S_d^2 = S_a^2 \times 10^2 = (2.5) \times (10)^2 = 250 \text{ olacaktır.}$$

5. E örneğindeki gözlem değerleri, $E=D-6$ şeklinde elde edilmiştir. Bu sebeple E örneğinin varyansı,

$$S_e^2 = S_d^2 = 250 \text{ olacaktır.}$$

2.2.3.3. Frekans Dağılım Tablosundan Kısa Yoldan Varyansın Hesaplanması

Varyans da aritmetik ortalamada olduğu gibi frekans dağılım tablosundan kısa yoldan hesaplanabilir. Bunun için (2.3) numaralı eşitlik kullanılarak her sınıf için “b” değerleri;

$$b_i = \frac{X_i - A}{c}$$

şeklinde hesaplanır. Bu eşitlikten, $X_i = A + b_i c$ yazılabilir. Varyansın özelliklerinde açıklandığı gibi $b_i c$ değerlerine A sabitinin eklenmesi varyansı değiştirmez. Bu sebeple, geriye $X_i = b_i c$ kalır. Bu durumda X değerlerine ait kareler toplamı ve varyans;

$$\sum d_x^2 = c^2 \sum d_b^2 \text{ ve } S_x^2 = c^2 S_b^2$$

şeklinde hesaplanır. Frekans dağılım tablosundan “b” değerlerine ait varyans (2.16) numaralı eşitlik kullanılarak;

$$S_b^2 = \frac{\sum_{i=1}^k f_i b_i^2 - \frac{(\sum_{i=1}^k f_i b_i)^2}{\sum_{i=1}^k f_i}}{\sum_{i=1}^k f_i - 1}$$

şeklinde hesaplanır. 80 mısır bitkisinin uzunluğu için varyansın, frekans dağılım tablosundan kısa yoldan hesaplanması Tablo 2.11’de verilmiştir.

Tablo 2.11. 80 mısır bitkisinin uzunluğu için varyansın frekans dağılım tablosundan kısa yoldan hesaplanması

Sınıf değeri	Mutlak frekans	b_i	$f_i b_i$	$f_i b_i^2$
52.5	2	-5	-10	50
58.5	3	-4	-12	48
64.5	7	-3	-21	63
70.5	10	-2	-20	40
76.5	10	-1	-10	10
82.5=A	15	0	0	0
88.5	10	1	10	10
94.5	8	2	16	32
100.5	10	3	30	90
106.5	3	4	12	48
112.5	2	5	10	50
Toplam	$\sum f_i = 80$		$\sum f_i b_i = 5$	$\sum f_i b_i^2 = 441$
$\sum f_i$, $\sum f_i b_i$ ve $\sum f_i b_i^2$ değerleri hesaplandıktan sonra (2.16) numaralı eşitlik kullanılarak “b” değerlerinin varyansı; $S_b^2 = \frac{441 - \frac{(5)^2}{80}}{(80 - 1)} \cong 5.57832$ olarak bulunur.				

(2.3) numaralı eşitlik ve varyansın özelliklerinden, X değerlerinin varyansının $S_x^2 = c^2 S_b^2$ şeklinde hesaplanabileceği açıklanmıştı. Bu durumda, 80 mısır bitkisinin uzunluğuna ait varyans, $S_x^2 = (6)^2 (5.57832) \cong 200.82$ olarak bulunur.

2.2.4. Standart Sapma

Standart sapma varyansın kareköküdür ve birimi gözlem değerlerinin ölçü birimidir. Populasyondan (2.17) numaralı eşitlik kullanılarak hesaplanır. Bu durumda bir parametredir ve σ ile gösterilir. Örnekten hesaplanırsa bir istatistiktir, S ile gösterilir ve (2.18) numaralı eşitlik kullanılarak hesaplanır.

Populasyonda:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu_x)^2}{N}} \quad \dots(2.17)$$

Örnekte:

$$S_x = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}} \quad \dots(2.18)$$

80 mısır bitkisinin uzunluğu için standart sapma, $S_x = \sqrt{S_x^2} = \sqrt{200.82} = 14.17$ cm olarak bulunur.

2.2.5. Varyasyon Katsayısı

Üzerinde çalışılan örnekten toplanan gözlem değerleri büyüdükçe, yani örneğin ortalaması büyüdükçe, standart sapmaları da genellikle büyür. Eğer örnekler, gözlem değerleri arasındaki değişim bakımından karşılaştırılmak istenirse her bir örnek için varyasyon katsayısının hesaplanması gerekir. Varyasyon katsayısı (CV, Coefficient of variation), standart sapmanın ortalamaya bölünüp 100 ile çarpımı şeklinde (2.19) numaralı eşitlik kullanılarak hesaplanır. Böylece standart sapma üzerinde ortalamanın etkisi ortadan kaldırılmış olur. Bu katsayı, standart sapmayı ortalamanın yüzdesi, yani gözlem değerleri arasındaki değişimi (varyasyonu) % olarak ifade eder.

$$CV = \frac{S_x}{\bar{X}} 100 \quad \dots(2.19)$$

ÖRNEK:

Yapılan bir çalışmada salatalık turşularında, ortalama tuz oranının %4 ve tuz oranına ait standart sapmanın %0.3, ortalama su oranının ise %75 ve su oranına ait standart sapmanın da %5 olduğu saptanmıştır. Eğer hangi özellik bakımından değişimin daha fazla olduğu araştırılırsa standart sapma dikkate alınarak su oranı bakımından değişimin daha fazla olduğu söylenemez. Çünkü su oranı bakımından standart sapma %5 olarak hesaplanmıştır fakat ortalama su oranı da yüksektir. Bu sebeple önce tuz ve su oranlarına ait varyasyon katsayılarının (2.19) numaralı eşitlik kullanılarak hesaplanması gerekir.

$$CV_{\text{tuz}} = \frac{0.3}{4} 100 = \%7.5$$

$$CV_{\text{su}} = \frac{5}{75} 100 = \%6.67$$

Tuz ve su oranları için hesaplanan varyasyon katsayıları, tuz bakımından gözlem değerleri arasındaki değişimin daha fazla olduğunu göstermiştir. Su oranı için hesaplanan standart sapmanın daha büyük olması su oranına ait ortalamanın da büyük olmasından kaynaklanmıştır. Bu sebeple ortalamaları birbirinden farklı olan örnekler, gözlem değerleri arasındaki değişim bakımından karşılaştırılmak istendiği zaman mutlaka önce örnekler için varyasyon katsayılarının hesaplanması gerekir.

Örnekten elde edilen gözlem değerleri arasında sıfır ve negatif değerler varsa, bu örnek için varyasyon katsayısının hesaplanması bir anlam taşımaz. Çünkü bu gibi durumlarda hesaplanan varyasyon katsayısı %100'ü geçer.

ÖRNEK:

Mart ayının ilk haftasında sıcaklıklar -1, 0, -1, 2, 4, 3, 7 olarak ölçülmüş olsun. Bu örnek için;

$$\bar{X} = \frac{14}{7} = 2.0$$

$$\sum d_x^2 = (-1)^2 + 0 + (-1)^2 + (2)^2 + (4)^2 + (3)^2 + (7)^2 - \frac{(14)^2}{7} = 52$$

$$S_x^2 = \frac{52}{(7-1)} \cong 8.67$$

$$S_x = \sqrt{8.67} = 2.94^\circ\text{C}$$

olarak bulunur. Bu örnek için varyasyon katsayısı da (2.19) numaralı eşitlik kullanılarak;

$$CV = \frac{2.94}{2.0} 100 = \%147 \text{ olarak hesaplanır. Hesaplanan CV \%100'ü geçtiği için bir anlam}$$

taşımaz. Bu sebeple de sıfır ve negatif değerler olan örnekler için varyasyon katsayısı hesaplanmaz.

Daha önce açıklandığı gibi üzerinde çalışılan örnekteki bireylerin, çalışılan özelliği etkileyen faktörler bakımından homojen olması gerekir. Eğer örnekteki bireyler homojen değilse, gözlem değerleri arasındaki değişim, yani hata artar. Bu da elde edilecek sonuçların güvenilirliğini düşürür. Yürütülen bir araştırmadan elde edilen gözlem değerleri arasındaki değişimin kabul edilir sınırlar içinde olup olmadığı için de varyasyon katsayısı hesaplanır. Varılacak sonuçların güvenilirliği bakımından, gözlem değerleri için hesaplanacak varyasyon katsayısının %15-20'den daha büyük olması istenmez.

2.3. Sorular

1. Tanıtıcı istatistikler kavramı ne ifade etmektedir.
2. Tanıtıcı istatistikler kaç gruba ayrılır. Bu gruplardaki istatistikler nelerdir?
3. Merkezi eğilim (yerleşim) ölçüleri hangi amaçla hesaplanır? Açıklayınız.
4. Aritmetik ortalama nedir? Özellikleri nelerdir? Örnek vererek açıklayınız.
5. Ortanca değer nedir? Örnek vererek açıklayınız.
6. Ortanca değer hangi durumlarda hesaplanmalıdır? Açıklayınız.
7. Tepe değeri nedir? Örnek vererek açıklayınız.

8. Aritmetik ortalama, ortanca değer ve tepe değeri arasında nasıl bir ilişki vardır? Bu ilişkiden nasıl yararlanılır? Açıklayınız.
9. Değişim ölçüleri hangi amaçla hesaplanır? Açıklayınız.
10. Değişim genişliği nedir? Yaygın olarak kullanılan bir değişim ölçüsü müdür? Açıklayınız.
11. Ortalama sapma nedir? Neden yaygın olarak kullanılan bir değişim ölçüsü değildir? Açıklayınız.
12. Kareler toplamı nedir? Nasıl hesaplanır?
13. Serbestlik derecesi nedir? Açıklayınız.
14. Varyans nedir? Nasıl hesaplanır ve özellikleri nelerdir? Açıklayınız.
15. Varyasyon katsayısı ne amaçla hesaplanır? Açıklayınız.
16. Aşağıda A örneğine ait ortalama ve standart sapma sırasıyla 4.4 ve 2.3 olarak verilmiştir. Ortalama ve varyansın özelliklerinden yararlanarak B, C ve D örneklerine ait ortalama ve varyansı bulunuz.

A	B	C	D
2	1	11	12
4	2	13	24
8	4	17	48
5	2.5	14	30
3	1.5	12	8

17. Aşağıda A örneğine ait ortalama ve standart sapma sırasıyla 43.8 ve 12.83 olarak bulunmuştur. B, C, D örneğine ait ortalama ve varyansları hesaplayınız.

A	B	C	D
27	9	57	351
36	12	75	468
45	15	93	585
51	17	105	663
60	20	123	780

18. $n=6$, $\sum X = 45$, $\sum (X_i - \bar{X})^2 = 15$ ve $Y=5X+13$ olduğuna göre Y'nin ortalama ve varyansını hesaplayınız.
19. Mart ayının ilk haftasında sıcaklıklar -1, 0, -1, 2, 4, 3, 7 olarak ölçülmüştür. Bu sıcaklıklar için aritmetik ortalamayı, standart sapmayı ve varyasyon katsayısını hesaplayarak sonuçlarını tartışınız.
20. Aşağıdaki örneklerden A örneğinin ortalaması 3 olduğuna göre, diğer örneklerin (B,C,D;E) ortalamalarını herhangi bir hesaplama yapmadan sadece ortalamanın özelliklerinden yararlanarak bulunuz.

A	B	C	D	E
1	11	0.5	10	100
2	12	1.0	20	101
3	13	1.5	30	102
4	14	2.0	40	103
5	15	2.5	50	104

21. 30 adet üzüm tanesindeki çekirdek sayıları aşağıdaki gibi bulunduğu göre;

0	1	1	0	2	2	3	4	1	2
0	1	2	4	3	3	2	2	2	1
2	3	4	0	3	2	2	3	2	2

Frekans dağılım tablosunu yaparak aritmetik ortalamayı bulunuz.

22. Hıyar turşularında, ortalama tuz oranının %4 ve tuz oranına ait standart sapmanın da %0.2, yine hıyar turşularında ortalama su oranının %75 ve su oranına ait standart sapmanın da %5 olduğu bilinmektedir. Bu bilgilere göre, hıyar turşularında tuz ve su oranlarından hangisinde varyasyon (değişim) daha fazladır?

23. X ve Y gibi iki özellik arasında, $Y = \frac{1}{3}X + 5$ şeklinde fonksiyonel bir ilişki olduğu bilinmektedir. X özelliğinin ortalama ve standart sapması sırasıyla 12.0 ve 18.0 olarak bulunduğu göre, Y özelliğinin ortalama ve varyansı kaçtır?

24. Bir örnekten edilen gözlem değerleri 3, 1, 3, 0, 4, -1, 2 şeklinde ise bu örnekte gözlem değerleri arasındaki değişimin ölçüsü olarak hesaplanabilecek olan değişim ölçüsü nedir? Hesaplayarak açıklayınız.

25. Yapılan bir araştırmada tesadüfen seçilen 6 öğrencinin haftalık masrafları TL olarak 45, 47, 52, 265, 65 ve 58 olarak saptanmıştır. Bu örnek için hesaplanabilecek en uygun merkezi eğilim (yerleşim) ölçüsü nedir? Nedenini açıklayarak hesaplayınız.

26. Açık uçlu frekans dağılım tabloları ile çalışılırken hesaplanacak merkezi eğilim (yerleşim) ölçüsü hangisidir? Nedenini belirterek açıklayınız.

27. 67 bireyin oluşturduğu bir örnekte, ortalama 170.392, ortanca değer 173.435 ve tepe değeri de 176.087 olarak bulunduğu göre, dağılım şekli için ne söylenebilir?

28. Bir örnekten edilen gözlem değerleri -3, 0, 3, -4, 1, 2 şeklinde ise bu örnek için hesaplanamayacak olan değişim ölçüsü nedir?

29. 75 öğrencinin İstatistik dersi vize notuna ilişkin aritmetik ortalama 42.75 ve Tepe değeri de 43.12 bulunduğu göre ortanca değer kaç olabilir? Tartışınız.

30. Bir frekans dağılım tablosunda en yüksek frekansa sahip olan sınıfların 5. ve 7. sınıflar olduğu saptanmıştır. Böyle bir frekans dağılım tablosunda tepe değeri için ne söylenebilir? Tartışınız.

31. 8, 10, 12, 14, 16 şeklinde elde edilen gözlem değerlerinin varyansı 10 olarak bulunduğu göre, -8, -10, -12, -14, -16 gözlem değerlerinin varyansı ne olur? Tartışınız.

32. 25 kişilik bir örnekte X özelliğine ait ortalama $\bar{X} = 25$, varyasyon katsayısı da C.V. = %20 olarak bulunduğu göre, bu özelliğin varyansı kaçtır?

33. Aşağıdaki frekans dağılım tablosunda aritmetik ortalama, ortanca değer, tepe değeri, varyans, standart sapma ve varyasyon katsayısını hesaplayınız.

Sınıflar	Frekans
15.25 -	1
15.45 -	4
.... -	7
.... -	14
.... -	10
.... -	8
.... -	4
.... -	2

34. 125 bireyin oluşturduğu bir örnekte, ortanca değer 65.281, tepe değeri de 64.976 olarak bulunduğuna göre, aritmetik ortalama için ne söylenebilir?
35. 5, 4, 3, 54, 2, 4, 7, 3 sayısal değerlerinin elde edildiği bir örnekte, merkezi eğilim (yerleşim) ölçülerinden hangisini hesaplamak uygundur? Nedenini belirterek hesaplayınız.
36. 48 adet öğrencinin ağırlığına ait varyans 55 olarak hesaplanmıştır. Bu öğrencilerin hepsine 3'er kg. ağırlığında palto giydirildikten sonra, öğrencilerin ağırlıklarına ait varyans ne olur?
37. Ş ve Ç gibi iki özellik arasında, $\bar{X} = 2\bar{Y} - 2.5$ şeklinde fonksiyonel bir ilişki olduğu bilinmektedir. Ç özelliğinin ortalama ve standart sapması sırasıyla 4.0 ve 3.0 olarak bulunduğuna göre, Ş özelliğinin ortalama ve varyansı nedir? Hesaplayınız.
38. 8 adet analjezik ilaç kutusunun etken madde miktarlarına ait ortanca değer, tepe değeri, ortalama, varyans ve varyasyon katsayısını hesaplayınız.

225	240	235	245	250	247	235	230
-----	-----	-----	-----	-----	-----	-----	-----

39. 5 bireyde X özelliğine ait varyans 5 olarak bulunmuştur. $Y=2X+1005$ eşitliği bilindiğine göre Y'nin varyansı kaç olur?

40. $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n d_x^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}$ olduğunu gösteriniz.