BÖLÜM 2. MERKEZİ EĞİLİM ve DAĞILIM ÖLÇÜLERİ

Merkezi eğilim ölçüleri kitleye ilişkin bir değişkenin bütün farklı değerlerinin çevresinde toplandığı merkezi bir değeri gösterirler. Dağılım ölçüleri ise değişkenin aldığı değerlerin birbirinden ne kadar farklı olduğunun ölçüsüdür. En sık kullanılan merkezi eğilim ölçüleri aritmetik ortalama, tepe değer, ortanca, çeyreklikler ve geometrik ortalamadır. En sık kullanılan dağılım ölçüleri ise, değişim genişliği, çeyrek sapma, varyans, standart sapma, standart hata ve değişim katsayısıdır.

2.1. Merkezi Eğilim Ölçüleri

2.1.1. Aritmetik Ortalama

Aritmetik ortalama, en çok kullanılan merkezi eğilim ölçüsüdür. Birimlerin belirli bir değişken bakımından aldıkları değerlerin toplamının birim sayısına bölümü olarak tanımlanır. Eşit aralıklı ve oran ölçme düzeyinde ölçülen değişkenler için kullanılır. Aritmetik ortalama hem kitle hem de örneklem için hesaplanır.

μ: kitleye ilişkin aritmetik ortalama
$\bar{x}$: örnekleme ilişkin aritmetik ortalama

Aritmetik ortalama sınıflandırılmamış ve sınıflandırılmış verilerde incelenecektir.

Sınıflandırılmamış Verilerde Aritmetik Ortalama:

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

$\bar{x}$: aritmetik ortalama
x_i: örneklemdeki i. birimin değeri
n : örneklemdeki birim sayısı

Örnek 2.1: Tablo 1.1 de verilen ham verilerin aritmetik ortalamasını hesaplayınız.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{120} x_i}{120} = \frac{408.33}{120} = 3.40$$

Örnek 2.2: Ankara Üniversitesi Tıp Fakültesi Hastanesi kardiyoloji servisinde yatan hastaların hastanede kalış süreleri hakkında bilgi sahibi olunmak istenmektedir. Rasgele seçilen 15 hastanın hastanede kalış süreleri aşağıdaki gibi saptanmıştır.
$x_i(Gün)$: 20 40 10 10 26 17 17 15 22 12 12 5 5 14 15

Hastaların hastanede kalış sürelerine ilişkin aritmetik ortalamayı hesaplayınız.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{15} x_i}{15} = \frac{20 + 40 + \cdots + 15}{15} = \frac{240}{15} = 16$$

Sınıflandırılmış Verilerde Aritmetik Ortalama:

Frekans tablosu düzenlenmiş verilerde aritmetik ortalama aşağıda verilecek eşitliklerle hesaplanabilir.

I. Frekans tablosundaki sınıf değeri ve frekans sütunundan yararlanılarak hesaplanan aritmetik ortalama,

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

$\bar{x}$: aritmetik ortalama
S_i: i. sınıfın sınıf değeri
f_i: i. sınıfın frekans değeri
k : sınıf sayısı
n : örneklemdeki birim sayısı

Örnek 2.3: Tablo 1.2 ile verilen frekans tablosundan yararlanarak 120 erkek bebeğin ağırlıklarına ilişkin aritmetik ortalamayı hesaplayınız.

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i S_i}{n} = \frac{\sum_{i=1}^{10} f_i S_i}{120} = (2.9 \times 1 + 3.01 \times 2 + \cdots. + 3.89 \times 4)/120 = \frac{409,38}{120} = 3.411$$

II. Sınıf değeri ve frekanslar büyük olduğunda b_i değerinin kullanılması aritmetik ortalamanın hesaplanmasında kolaylık sağlar. b_i değeri,

$$b_i = \frac{S_i - S_a}{c}$$

ile tanımlanır. Buradan, $S_i = S_a + b_i c,\ \ i = 1,2,\dots,k$ değeri yukarıda verilen aritmetik ortalama formülünde yerine konulursa,

$$\bar{x} = S_a + c\frac{\sum_{i=1}^{k} f_i b_i}{n}$$

biçimine dönüşür.

$\bar{x}$: aritmetik ortalama
S_a: i. sınıf değeri sütununda kabul edilen değer
f_i: i. sınıfın frekans değeri
b_i : i. sınıfın sıra sayısı
c : sınıf aralığı

Örnek 2.4: Tablo 1.2 ile verilen frekans tablosundan yararlanarak 120 erkek bebeğin ağırlıklarına ilişkin aritmetik ortalamayı b_i değerinin kullanarak hesaplayınız.

Sınıf (S_i)	b_i	Frekans(f_i)
2.90	-4	1
3.01	-3	2
3.12	-2	11
3.23	-1	18
3.34	**0**	**22**
3.45	1	35
3.56	2	14
3.67	3	8
3.78	4	5
3.89	5	4

$$\bar{x} = S_a + c\frac{\sum_{i=1}^{k} f_i b_i}{n}$$

$$\sum_{i=1}^{k} f_i b_i = 1 \times (-4) + 2 \times (-3) + \cdots + 5 \times 4 = 77$$

$$\bar{x} = 3.34 + 0.11(77/120) = 3.411$$

Aritmetik ortalamanın özellikleri,

- Bir veri seti için sadece bir aritmetik ortalama vardır.
- Nicel verilere uygulanabilir.

- Birim değerlerinde meydana gelen değişim çok küçük olsa bile aritmetik ortalamayı etkiler.
- Aritmetik ortalama ile birim değerleri arasındaki farkların toplamı sıfırdır. $\sum_{i=1}^{k}(x_i - \mu) = 0$
- Aritmetik ortalama ile birim değerleri arasındaki farkların kareleri toplamı minimum bir değerdir.

Örnek 2.5: Bir ülkedeki hastanelerin yatak kapasitelerine göre dağılımları aşağıda verilmiştir.

As	Üs	Sınıf (S_i)	Frekans(f_i)	b_i
100	199	149.5	10	-3
200	299	249.5	40	-2
300	399	349.5	60	-1
400	499	449.5	100	0
500	599	549.5	50	1
600	699	649.5	30	2
700	799	749.5	10	3

Yatak kapasitesine ilişkin aritmetik ortalamayı hesaplayınız.

I. $$\bar{x} = \frac{\sum_{i=1}^{k} f_i s_i}{n} = \bar{x} = \frac{\sum_{i=1}^{7} f_i s_i}{300} = \frac{149.5x10 + \cdots + 749.5x10}{300} = \frac{131850}{300} = 439.5$$

II. $$\bar{x} = S_a + c\frac{\sum_{i=1}^{k} f_i b_i}{n} = 449.5 + 100\left(\frac{-30}{300}\right) = 439.5$$

2.1.2. Tepe Değer (Mod)

Bir veri grubunda en çok tekrarlanan değere tepe değer(mod) denir. Tepe değerin hesaplanmasında birimlerin büyüklük sırasına konulması şart değilse de, bu işlemin yapılması tepe değerin bulunmasında kolaylık sağlar.

Sınıflandırılmamış Verilerde Tepe Değer

Sınıflandırılmamış verilerde en çok tekrarlanan değer tepe değer olarak alnır.

Örnek 2.6. Bir grup öğrencinin ağırlıklarına ilişkin veriler sırasıyla şöyledir:

56, 57, 57, 58, 69, 69, 69, 80, 81, 82

Öğrencilerin ağırlıklarına ilişkin tepe değeri hesaplayınız.

Yukarıdaki veri gurubunda en çok tekrarlanan değer 69 olduğundan $TD = 69$, dur.

Örnek 2.7. 12 hastanın kan basınçları 90, 80,100, 110, 100, 120, 100, 90, 100, 110, 120, 110 olarak ölçülmüştür. Kan basınçlarına ilişkin tepe değeri hesaplayınız.

Kolaylık olması için veriler sıraya dizilirse,

80, 90, 90, 100, 100, 100, 100, 110, 110,110, 120, 120

En çok tekrarlanan değer 100 olduğundan, $TD = 100$ olur.

Sınıflandırılmış Verilerde Tepe Değer

Sınıflandırılmış verilerde tepe değeri hesaplayabilmek için öncelikle tepe değer sınıfının belirlenmesi gerekir. Frekansı en yüksek olan sınıf tepe değer sınıfıdır. Bu sınıfta yer alan tepe değeri bulmak için aşağıdaki formül kullanılır.

$$TD = A_s + c\left(\frac{F_1}{F_1 + F_2}\right)$$

Bu formülde;

TD: Tepe Değer
A_s : En büyük sınıfın bulunduğu sınıfın alt sınırı
c : Sınıf aralığı
F_1: En büyük frekans ile bir önceki sınıfın frekansı arasındaki fark
F_2: En büyük frekans ile bir sonraki sınıfın frekansı arasındaki farkdır.

Örnek 2.8. Tablo 1.2 ile verilen frekans tablosundan yararlanarak 120 erkek bebeğin ağırlıklarına ilişkin tepe değeri hesaplayınız.

Tablo 1.2: 120 Erkek Bebeğin Ağırlıklarına İlişkin Frekans Tablosu

As	Üs	Sınıf (S_i)	Frekans(f_i)	Göreli frekans(p_i)
2.85	2.95	2.90	1	0.01
2.96	3.06	3.01	2	0.02
3.07	3.17	3.12	11	0.09
3.18	3.28	3.23	18	0.15
3.29	3.39	3.34	22	0.18
3.40	3.50	3.45	**35**	0.29
3.51	3.61	3.56	14	0.12
3.62	3.72	3.67	8	0.07
3.73	3.83	3.78	5	0.04
3.84	3.94	3.89	4	0.04

$$TD = A_s + c\left(\frac{F_1}{F_1+F_2}\right) = 3.40 + 0.11\left(\frac{(35-22)}{(35-22)+(35-14)}\right) = 3.40 + 0.11\left(\frac{13}{13+21}\right) = 3.442$$

Tepe değerin özellikleri,

- Denek sayısı az olduğunda tepe değer güvenilir bir ölçü değildir.
- Bazı örneklemlerde bir tepe değer yerine iki ya da daha çok tepe değer olabilir. Bu durumda ya tepe değerini hesaplamaktan vazgeçilir ya da frekans tablosu tek tepe değerli bir dağılım olacak şekilde yeniden düzenlenir.
- Tepe değer hesaplanırken birimlerin tümü işleme katılmadığı için uç değerlerden etkilenmez.
- Nicel ve nitel verilerin her iki türü için de uygundur.
- Eğrisi J, ters J ve U şeklinde olan veriler için tepe değer kullanılmaz.

Örnek 2.9. Bir fakültede öğrencilerin boy uzunluklarına ilişkin veriler aşağıdadır. Bu öğrencilerin boyları için tepe değeri bulunuz.

As	Üs	Frekans(f_i)
140	149	5
150	159	100
160	169	**250**
170	179	60
180	189	10

$$TD = A_s + c\left(\frac{F_1}{F_1 + F_2}\right) = 160 + 10\left(\frac{(250 - 100)}{(250 - 100) + (250 - 60)}\right) = 164.412$$

2.1.3. Ortanca (Medyan)

Bir veri gurubundaki değerlerin küçükten büyüğe sıralandığında tam ortaya düşen değer ortanca değeridir. Kitledeki birimlerin sayısı çok fazla ise verilerin özetlenmesinde merkezi eğilim ölçüsü olarak ortanca kullanılabilir. Ortanca, sınıflama ölçme düzeyi ile ölçülen değişkenler için kullanılmaz. Eşit aralıklı, oran ve sıralama ölçme düzeyinde ölçülen değişkenler için kullanılır.

Sınıflandırılmamış Verilerde Ortanca

Birim sayısının tek veya çift olmasına göre medyanın bulunması değişir. İki durumda da, ilk olarak eldeki veriler büyüklük sırasına göre (küçükten büyüğe veya büyükten küçüğe) sıraya konulur. Birim sayısı n ile gösterilmek üzere,

$$ortanca(OR) = \begin{cases} x_{\mathrm{j}} & , \quad \mathrm{j} = \dfrac{\mathrm{n}+1}{2} \quad \mathrm{n\ tek} \\ \dfrac{x_{\mathrm{j}} + x_{\mathrm{j}+1}}{2} & , \quad \mathrm{j} = \dfrac{\mathrm{n}}{2} \quad n\ çift \end{cases}$$

formülü ile hesaplanır.

Örnek 2.10: Aynı hastalığa sahip 12 kişilik bir gruba yapılan ilaç tedavisi sonucu iyileşme süreleri gün olarak aşağıdaki gibi verilmiştir.

15,16,18,14,12,17,18,20,19,14,15,18

İyileşme sürelerinin ortancasını bulunuz.

Verilerin küçükten büyüğe sıralanmış hali,

12,14,14,15,15,16,17,18,18,18,19,20

biçimindedir. Veri sayısı, $n = 12$ çift olduğu için ortaya düşen iki değerin ortalaması ortanca olacaktır. $\mathrm{j} = \frac{\mathrm{n}}{2} = \frac{12}{2} = 6$ ve $\mathrm{j} + 1 = 7$ olmak üzere,

12,14,14,15,15,16,17,18,18,18,19,20

x_6 x_7

$$OR = \frac{x_6 + x_7}{2} = \frac{16 + 17}{2} = 16.5$$

olarak bulunur.

Örnek 2.11: Bir bulaşıcı hastalığın kuluçka dönemi gün olarak aşağıdaki gibi gözlenmiştir.

5,6,5,3,7,9,6,4,8

Ortancayı bulunuz.

Veri sayısı, $n = 9$ tek olduğu için veriler sıralandığında ortaya düşen değer ortanca olacaktır. $\mathrm{j} = \frac{\mathrm{n}+1}{2} = \frac{10}{2} = 5$ olmak üzere verilerin küçükten büyüğe sıralanmış hali

3,4,5,5,6,6,7,8,9

x_5

$$OR = x_5 = 6$$

Sınıflandırılmış Verilerde Ortanca

Sınıflandırılmış verilerde ortanca hesaplanırken birikimli frekans tablosunda bulunan sınıf ara değeri ve den daha az frekans sütunları kullanılır. İlk olarak, veri sayısının tek ya da çift oluşuna göre $j = n/2$ veya $j = (n+1)/2$ değeri bulunur. Daha sonra aşağıdaki eşitlik kullanılarak ortanca değeri hesaplanır.

$$\text{ortanca(OR)} = S_{\text{ad}} + \text{c}\left(\frac{\frac{\text{n}}{2} - \text{N}_1}{\text{N}_2 - \text{N}_1}\right)$$

Burada,

S_{ad}: j. değerin bulunduğu sınıfın alt sınıf ara değeri
N_1: j. değerin bulunduğu sınıfın alt den daha az frekansı
N_2: j. değerin bulunduğu sınıfın üst den daha az frekansı
c: Sınıf aralığı
n: birim sayısı

dır. Burada, $\text{N}_2 - \text{N}_1$ aslında j. değerin bulunduğu sınıfın frekansıdır.

Ortancanın özellikleri:

- Aşırı uç değerlerden etkilenmez.
- Birim değerleri ile ortanca arasındaki farkın yarısı negatif yarısı pozitiftir.
- $\sum|x_i - ortanca| = minimum$ dur.

Örnek 2.12: Tablo 1.2 ile verilen frekans tablosundan yararlanarak 120 erkek bebeğe ilişkin ortancayı hesaplayınız.

f_i	S_{ad}	Den Daha Az fi
	2.845	0
2.95	2.955	1
3.06	3.065	3
3.17	3.175	14
3.28	3.285	32
3.39	3.395	**54**
3.50	3.505	**89**
3.61	3.615	103
3.72	3.725	111
3.83	3.835	116
3.94	3.945	120

$j = 60$ değeri bu iki sınıf arasındadır.

$c = 0.11$ ve $n = 120$ dir. Burada, $j = \frac{120}{2} = 60$ olup den daha az frekans sütununda bu değer yoktur. Bu durumda bu değeri içeren aralık belirlenir. Buna göre yukarıda belirtilen ve bu sınıfı içine alan $\mathrm{N}_1: 54$ ve $\mathrm{N}_2: 89$ dur. Burada, N_1' e karşılık gelen $S_{\mathrm{ad}} = 3.395$ tir. Bu değerler ortanca formülünde yerine konulduğunda,

$$\mathrm{OR} = S_{\mathrm{ad}} + \mathrm{c}\left(\frac{\frac{\mathrm{n}}{2} - \mathrm{N}_1}{\mathrm{N}_2 - \mathrm{N}_1}\right)$$

$$\mathrm{OR} = 3.395 + 0.11\left(\frac{60 - 54}{89 - 54}\right) \cong 3.414$$

olarak hesaplanır.

Eğer, j değeri den daha az frekans sütununda bulunabilirse ortanca, den daha az frekansına karşılık gelen sınıfın sınıf ara değeri olarak alınır.

Örnek 2.13: 271 hastanın kandaki şeker miktarına ilişkin dağılımı aşağıda verilmiştir. Ortancayı hesaplayınız.

A_s	$Ü_s$	Kandaki şeker miktarı (S_i)	Hasta Sayısı (f_i)
100	119	109.5	30
120	139	129.5	80
140	159	149.5	60
160	179	169.5	50
180	199	189.5	51

İlk olarak birikimli frekans tablosunu oluşturulması gerekir.

Hasta Sayısı (f_i)	S_{ad}	Den Daha Az f_i
30	99.5	0
80	119.5	30
60	139.5	110
50	159.5	170
51	179.5	220
	199.5	271

j = 136 değerini içeren aralık

$c = 20$, $j = \frac{271+1}{2} = 136$ olup den daha az frekans sütununda bu değer yoktur. Bu durumda bu değeri içeren aralık belirlenir. Buna göre yukarıda belirtilen ve bu sınıfı içine alan $\mathrm{N}_1: 110$ ve $\mathrm{N}_2: 170$ dur. Burada, N_1' e karşılık gelen $S_{ad} = 139.5$ tir. Bu değerler ortanca formülünde yerine konulduğunda,

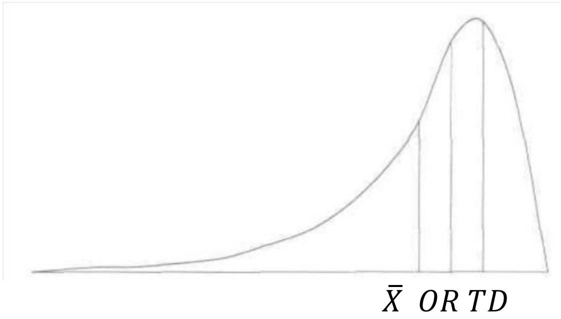
$$\mathrm{OR} = S_{ad} + \mathrm{c}\left(\frac{\frac{\mathrm{n}}{2} - \mathrm{N}_1}{\mathrm{N}_2 - \mathrm{N}_1}\right)$$

$$\mathrm{OR} = 139.5 + 20\left(\frac{135.5 - 110}{170 - 110}\right) \cong 148$$

olarak hesaplanır.

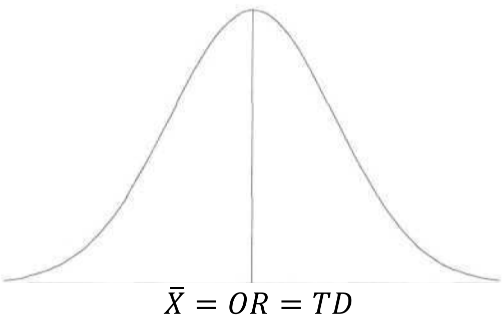
2.1.4. Aritmetik Ortalama, Tepe Değer ve Ortanca Arasındaki İlişki

Aritmetik ortalama, ortanca ve tepe değer arasındaki ilişki verilerin dağılımının çarpıklığı hakkında bilgi verir.



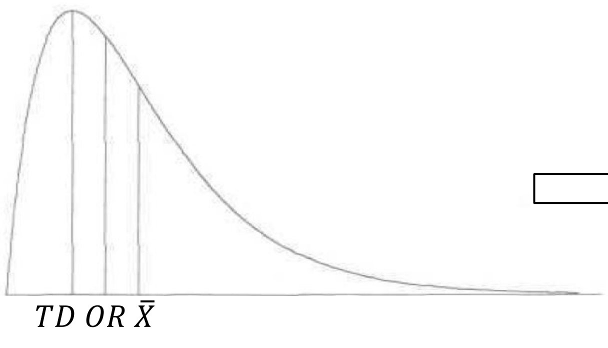
$\bar{X} < OR < TD$

dağılım sola çarpık ya da negatif yöne eğimli



$\bar{X} = OR = TD$

dağılım simetrik



$TD < OR < \bar{X}$

dağılım sağa çarpık ya da pozitif yöne eğimli

2.1.5. Çeyreklikler

Küçükten büyüğe doğru sıralanmış verileri dört eşit parçaya bölen değerlere çeyrek değerler denir. Birinci çeyreklik (Q_1), veriler küçükten büyüğe sıralandığında verilerin %25 ini sağında, %75 ini solunda bırakan değerdir. İkinci çeyreklik ortancaya denk gelmektedir. Üçüncü çeyrek değer (Q_3), veriler küçükten büyüğe sıralandığında verilerin %75 ini sağında,

%25 ini solunda bırakan değerdir. Yani sıralı verilerde, ortancadan küçük olan değerlerin ortancası birinci çeyrek değer, ortancadan büyük olan verilerin ortancası üçüncü çeyrek değerdir.

Örnek 2.14: İlaçla tedavi edilen 8 hastanın iyileşme süreleri gün olarak aşağıda verilmiştir. Çeyrek değerleri hesaplayınız.

30, 20, 24, 40, 65, 70, 10, 62

İlk olarak verilerin sıralanması gerekir,

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
10	20	24	30	40	62	65	70

$n = 8$ çift olduğu için $ortanca = \frac{x_4+x_5}{2} = \frac{30+40}{2} = 35$ dir. Ortancadan küçük olan değerlerin ortancası birinci çeyrek değere, ortancadan büyük değerlerin ortancası üçüncü çeyrek değerdir.

Ortancadan küçük olan değerler

x_1	x_2	x_3	x_4
10	20	24	30

$Q_1 = \frac{20+24}{2} = 22$

Ortancadan büyük olan değerler

x_5	x_6	x_7	x_8
40	62	65	70

$Q_3 = \frac{62+65}{2} = 63.5$

Örnek 2.15: Orta yaşlı 9 erkeğe ait sistolik kan basınçları aşağıda verilmiştir.

151, 124, 132, 170, 146, 124, 113, 111, 134

Çeyreklikleri hesaplayınız.

İlk olarak verilerin sıralanması gerekir,

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
111	113	124	124	132	134	146	151	170

Ortanca

$$Q_1 = \frac{113 + 124}{2} = 118.5 \qquad Q_3 = \frac{146 + 151}{2} = 148.5$$

2.1.6. Geometrik Ortalama

Geometrik ortalama, gözlem sonuçları bir önceki gözlem sonucuna bağlı olarak değişiyorsa ve bu değişimin hızı belirlenmek isteniyorsa kullanılan bir merkezi eğilim ölçüsüdür. Veriler (gözlem sonuçları), $x_1, x_2, \dots, x_n$ ile gösterilmek üzere geometrik ortalama,

$$GO = \sqrt[n]{x_1 . x_2 \dots x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

eşitliği ile hesaplanır. Her iki tarafın logaritması alınırsa, aşağıdaki eşitlikte kullanılabilir.

$$\log(GO) = \frac{1}{n}(logx_1 + logx_2 + \cdots + logx_n)$$
$$\log(GO) = \frac{1}{n}\sum_{i=1}^{n} logx_i$$

Geometrik Ortalamanın Özellikleri

- Herhangi bir veri sıfır veya negatif değerli ise geometrik ortalama hesaplanamaz.
- Uç değerlerden aritmetik ortalama kadar etkilenmez.
- Aritmetik ortalamadan küçüktür.
- Gözlem sonuçlarının geometrik ortalamaya oranlarının çarpımları 1 dir.

$$\frac{x_1}{GO}\frac{x_2}{GO}\dots\frac{x_n}{GO} = 1$$

Örnek 2.16. Aşağıdaki verilerin geometrik ortalamasını hesaplayınız.

4,6,5,8,7

$GO = \sqrt[5]{4.6.5.8.7} = \sqrt[5]{6720} = (6720)^{1/5}$

$GO = 5.8274$

Örnek 2.17. Bir canlı organizmanın ilk gün ağırlığı 3.2 gr. dır. Bu organizmanın ağırlığı altıncı günün sonuna kadar her gün % 25 artış gösteriyor. Bu canlı organizmanın ağırlığına ilişkin geometrik ortalamayı hesaplayınız.

günler	1. gün	2. gün	3. gün	4. gün	5. gün	6. gün
ağırlıklar	3.20	4	5	6.25	7.8125	9.765625
log(ağırlıklar)	0.50515	0.60206	0.69897	0.79588	0.89279	0.9897

$$\log(GO) = \frac{1}{6}(log3.2 + log4 + log5 + log6.25 + log7.8125 + log9.7656)$$

$$= \frac{1}{6}(0.51 + 0.6 + 0.7 + 0.8 + 0.9 + .99)$$

$$= \frac{1}{6}(4.5) = 0.75$$

$$GO = 5.6234$$

2.2. Dağılım Ölçüleri

2.2.1. Değişim Genişliği

Bir veri grubunda en büyük değer ile en küçük değer arasındaki farka değişim genişliği denir, R ile gösterilir.

$$R = En\ büyük\ değer - En\ küçük\ değer$$

Değişim genişliği, değişim aralığını gösteren bir dağılım ölçüsüdür. Değişim genişliğinin hesaplanmasında sadece iki uç değer işleme alındığından, diğer değerlerin hiçbir etkisi yoktur. Bu nedenle değişim genişliği yaygın olarak kullanılan bir dağılım ölçüsü değildir.

Örnek 2.18. Aşağıdaki veriler için değişim genişliğini hesaplayınız.

30, 12, 15, 22, 40, 55, 20, 58, 25, 60, 23, 72

$R = 72 - 12 = 60$ dır.

2.2.2. Çeyrek Sapma

Ortalama yerine ortanca kullanıldığında ya da veri setinde aşırı uç değerler bulunduğunda değişim genişliği yerine çeyrek sapma kullanılır. Çeyrek sapma Q ile gösterilir.

$$Q = \frac{Q_3 - Q_1}{2}$$

Eşitlikte,
Q : Çeyrek sapma
Q_1: Birinci çeyreklik
Q_3: Üçüncü çeyrekliktir.
Dağılımdaki bütün değerler kullanılmadığı için Q yeterli bir dağılım ölçüsü değildir.

Örnek 2.19: İlaçla tedavi edilen 8 hastanın iyileşme süreleri gün olarak aşağıda verilmiştir. Çeyrek sapma değerini hesaplayınız.

30, 20, 24, 40, 65, 70, 10, 62

Önce iyileşme sürelerini küçükten büyüğe sıraya dizelim:
$x_1,\ x_2,\ x_3, x_4, x_5, x_6,\ x_7,\ x_8$
10, 20, 24, 30, 40, 62, 65, 70

$n = 8, \qquad çift$

$$Q_1 = \frac{X_{(2)} + X_{(3)}}{2} = \frac{20 + 24}{2} = 22$$

$$Q_3 = \frac{X_{(6)} + X_{(7)}}{2} = \frac{62 + 65}{2} = 63.5$$

O halde çeyrek sapma $Q = \frac{63.5-22}{2} = 20.75$ dir.

2.2.3. Varyans ve Standart Sapma

Varyans, birim değerlerinin ortalamadan sapmalarının kareler toplamının birim sayısına bölünmesi ile elde edilir. Varyans gözlem sonuçlarının aritmetik ortalamadan ne ölçüde farklı olabileceğini ortaya koyan bir dağılım ölçüsüdür. Kitle varyansı σ^2, örneklem varyansı s^2 ile gösterilir.

Standart sapma varyansın kareköküdür. Kitle standart sapması σ, örneklem standart sapması s ile gösterilir.

Sınıflandırılmamış Verilerde Varyans ve Standart Sapma

Varyans,

$$s^2 = \frac{\sum_{j=1}^{n}(x_j - \bar{x})^2}{n-1} = \frac{\sum_{j=1}^{n} x_j^2 - \frac{\left(\sum_{j=1}^{n} x_j\right)^2}{n}}{n-1}$$

standart sapma,

$$s = \sqrt{\frac{\sum_{j=1}^{n}(x_j - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{j=1}^{n} x_j^2 - \frac{\left(\sum_{j=1}^{n} x_j\right)^2}{n}}{n-1}} = \sqrt{Varyans}$$

eşitlikleriyle bulunur. Burada,

s: standart sapma
x_j: j. denek değeri
$\bar{x}$: aritmetik ortalama
n: birim sayısıdır.

Örnek 2.20: Tablo 1.1. deki ham verileri kullanarak varyans ve standart sapmayı hesaplayınız.

$\sum_{j=1}^{120} x_j = 408.23$, $\sum_{j=1}^{120} x_j^2 = 1394.05$

$$s^2 = \frac{\sum_{j=1}^{n} x_j^2 - \frac{\left(\sum_{j=1}^{n} x_j\right)^2}{n}}{n-1} = \frac{1394.05 - \frac{408.23^2}{120}}{119} = 0.044$$

$$s = \sqrt{Varyans} = \sqrt{0.044} = 0.211$$

Örnek 2.21: 10, 15, 22, 26, 31, 40 verilerinin varyans ve standart sapması nedir?

$$\sum_{j=1}^{6} x_j = 10 + 15 + 22 + 26 + 31 + 40 = 144$$

$$\sum_{j=1}^{6} x_j^2 = 10^2 + 15^2 + 22^2 + 26^2 + 31^2 + 40^2 = 4046$$

$$s^2 = \frac{\sum_{j=1}^{n} x_j^2 - \frac{\left(\sum_{j=1}^{n} x_j\right)^2}{n}}{n-1} = \frac{4046 - \frac{144^2}{6}}{5} = \frac{4046 - 3456}{5} = 118$$

$$s = \sqrt{Varyans} = \sqrt{118} = 10.863$$

Sınıflandırılmış Verilerde Varyans ve Standart Sapma

I. Frekans tablosundaki sınıf değeri ve frekans sütunundan yaralanılarak hesaplanan varyans ve standart sapma formülü,

$$s^2 = \frac{\sum_{i=1}^{k} f_i S_i^2 - \frac{\left(\sum_{i=1}^{k} f_i S_i\right)^2}{n}}{n-1} \quad , \quad s = \sqrt{\frac{\sum_{i=1}^{k} f_i S_i^2 - \frac{\left(\sum_{i=1}^{k} f_i S_i\right)^2}{n}}{n-1}} = \sqrt{Varyans} \text{ dır.}$$

Eşitlikte,
f_i: i’inci sınıfın frekansı
S_i: i’inci sınıfın sınıf değeri
k: sınıf sayısı
n: birim sayısıdır.

Örnek 2.22: Tablo 1.2 ile verilen frekans tablosundan yararlanarak 120 erkek bebeğin ağırlıklarının varyans ve standart sapmasını bulunuz.

As	Üs	S_i	f_i
2.85	2.95	2.90	1
2.96	3.06	3.01	2
3.07	3.17	3.12	11
3.18	3.28	3.23	18
3.29	3.39	3.34	22
3.40	3.50	3.45	35
3.51	3.61	3.56	14
3.62	3.72	3.67	8
3.73	3.83	3.78	5
3.84	3.94	3.89	4

$$s^2 = \frac{\sum_{i=1}^{k} f_i S_i^2 - \frac{\left(\sum_{i=1}^{k} f_i S_i\right)^2}{n}}{n-1} = \frac{1400.564 - \frac{409.27^2}{120}}{119} = 0.0396$$

$$s = \sqrt{Varyans} = \sqrt{0.0396} = 0.199$$

II. Sınıf değerleri ve frekanslar büyük sayılar olduğunda varyans ve standart sapma hesaplanmasında b_i değerinden yararlanılarak kolaylık sağlanabilir.

$$s^2 = c^2\left(\frac{\sum_{i=1}^{k} f_i b_i^2 - \frac{\left(\sum_{i=1}^{k} f_i b_i\right)^2}{n}}{n-1}\right) \quad , \quad s = c\sqrt{\frac{\sum_{i=1}^{k} f_i b_i^2 - \frac{\left(\sum_{i=1}^{k} f_i b_i\right)^2}{n}}{n-1}} = \sqrt{Varyans} \text{ dır.}$$

Eşitlikte,
c: sınıf aralığı
f_i: i'inci sınıfın frekansı
b_i: i'inci sınıfın sıra sayısı
k: sınıf sayısı
n: birim sayısıdır.

Örnek 2.23: Tablo 1.2 ile verilen frekans tablosundan yararlanarak 120 erkek bebeğin ağırlıklarına ilişkin varyans ve standart sapmayı b_i değerini kullanarak hesaplayınız.

As	Üs	f_i	b_i
2.85	2.95	1	-4
2.96	3.06	2	-3
3.07	3.17	11	-2
3.18	3.28	18	-1
3.29	3.39	22	**0**
3.40	3.50	35	1
3.51	3.61	14	2
3.62	3.72	8	3
3.73	3.83	5	4
3.84	3.94	4	5

$$\sum_{i=1}^{10} f_i b_i = 1 \times (-4) + 2 \times (-3) + \cdots + 5 \times 4 + 4 \times 5 = 77$$

$$\sum_{i=1}^{10} f_i b_i^2 = 1 \times (-4)^2 + 2 \times (-3)^2 + \cdots + 5 \times 4^2 + 4 \times 5^2 = 439$$

$$s^2 = c^2\left(\frac{\sum_{i=1}^{10} f_i b_i^2 - \frac{(\sum_{i=1}^{10} f_i b_i)^2}{n}}{n-1}\right) = 0.11^2\left(\frac{439 - \frac{77^2}{120}}{119}\right) = 0.0396$$

$$s = \sqrt{Varyans} = \sqrt{0.0396} = 0.199$$

Örnek 2.24: İlaçla tedavi edilen hastaların iyileşme süresine göre dağılımı aşağıdadır.

As	Üs	S_i	f_i	b_i
5	14	9.5	30	-2
15	24	19.5	60	-1
25	34	29.5	100	0
35	44	39.5	80	1
45	54	49.5	30	2

İlaçla tedavi edilen hastaların iyileşme sürelerinin varyans ve standart sapmasını hesaplayınız.

I. $$s^2 = \frac{\sum_{i=1}^{k} f_i S_i^2 - \frac{\left(\sum_{i=1}^{k} f_i S_i\right)^2}{n}}{n-1} = \frac{310875-(9050^2/300)}{299} = 126.644$$

$$s = \sqrt{126.644} = 11.254$$

II. $$s^2 = c^2\left(\frac{\sum_{i=1}^{k} f_i b_i^2 - \frac{\left(\sum_{i=1}^{k} f_i b_i\right)^2}{n}}{n-1}\right) = 100\left(\frac{380-(\frac{20^2}{300})}{299}\right) = 126.644$$

$$s = \sqrt{126.644} = 11.254$$

2.2.4. Standart Hata

Örneklem ortalamalarının oluşturduğu dağılımın standart sapması örneklem ortalamalarından her birinin standart hatası sayılır. Bir örneklemin ortalamasının standart hatası,

$S_{\bar{x}} = \sqrt{\frac{S^2}{n}}$ veya $S_{\bar{x}} = \frac{S}{\sqrt{n}}$

eşitlikleri ile hesaplanır.

Örnek 2.25: Örnek 2.21 ile verilen verilerin standart hatasını bulunuz.

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{118}{6}} = 19.67$$

Örnek 2.26: Örnek 2.24 ile verilen, ilaçla tedavi edilen hastaların iyileşme sürelerinin standart hatasını bulunuz.

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{126.644}{300}} = 0.422$$

2.2.5. Değişim Katsayısı

Sadece, kitle varyansına bakılarak iki kitleden birinin diğerine göre daha homojen birimlerden oluştuğu söylenemez. Bunu söyleyebilmek için iki varyansın da aynı ölçekte olması gerekir. Örneğin birinci değişken uzunluk birimi, ikinci değişken ağırlık birimi ile ölçülmüş ise karşılaştırma yapılamayacağı açıktır. Bu gibi durumlarda değişim katsayısı olarak tanımlanan ve standart sapmanın ortalamaya bölümü olarak hesaplanan değişkenlik ölçüsü kullanılır.

$$DK = \frac{\sigma}{\mu} \qquad DK = \frac{S}{\bar{x}}$$

ile hesaplanır ve birimsizdir. Değişim katsayısının büyük çıkması, birim değerlerinin ortalama değerinden büyük olduğu, küçük çıkması birim değerlerinin ortalama değere yakın olduğu anlamına gelir. Birim değerleri ortalama değere eşit ise değişim katsayısı sıfır olacaktır.

Ör 2.27: İlaçla tedavi edilen7 hastanın ve ameliyatla tedavi edilen 9 hastanın iyileşme süreleri aşağıda gün olarak verilmiştir. İyileşme süresi bakımından ilaçla tedavi edilenler mi yoksa ameliyatla tedavi edilenler mi daha homojendir?

İlaçla (x_i):	10	20	22	34	18	22	34		
Ameliyatla(y_i):	30	40	50	52	40	52	48	34	32

İlk olarak aritmetik ortalama ve varyans hesaplanmalıdır.

İlaçla tedavi edilen hastalar için,

$\bar{x}=\frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{7}(160) = 22.86$

$S_x^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{1}{6}(446.8572) = 74.4762$

$S_x = \sqrt{74.4762} = 8.630$

$\bar{y}=\frac{\sum_{i=1}^{n} y_i}{n}=\frac{1}{9}(378)=42$

$S_y^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}=\frac{1}{8}(616)=77$

$S_y=\sqrt{77}=8.775$

Değişim katsayıları;

$$DK_1=\frac{S_x}{\bar{x}}=\frac{8.630}{22.86}=0.378$$

$$DK_2=\frac{S_y}{\bar{y}}=\frac{8.775}{42}=0.209$$

olarak hesaplanır. Buna göre, ameliyatla tedavi edilen hastaların iyileşme süresi bakımından ilaçla tedavi edilen hastalara göre daha homojen oldukları söylenebilir.

Örnek 2.28: Bir biyoloji deneyinde öğrenciler turp tohumlarını bir çantada üç gün karanlıkta bekletmişler ve bu üç günün sonunda 14 turp tohumunun uzunluklarını mm olarak aşağıdaki gibi bulmuşlardır.

15, 20, 11, 30, 33, 22, 37, 20, 29, 35, 8, 10, 13, 25

Turp tohumlarının uzunlukları için aritmetik ortalama, ortanca, tepedeğer, varyans, standart sapma ve standart hata değerlerini hesaplayınız. Dağılımın yönü için ne söylersiniz.

Verilerin sıralanmış hali;

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
8,	10,	11,	13,	15,	20,	20,	22,	25,	29,	30,	33,	35,	37

$$\bar{x}=\frac{\sum_{i=1}^{n} x_i}{n}=\frac{\sum_{i=1}^{14} x_i}{14}=\frac{8+10+\cdots+37}{14}=\frac{308}{14}=22$$

Ortanca için veri sayısının tek mi çift mi olduğuna bakılmalı. $n=14$ çift olduğundan (tam ortaya düşen iki değerin) x_7 ile x_8' in ortalaması ortancayı verecektir.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
8,	10,	11,	13,	15,	20,	20,	22,	25,	29,	30,	33,	35,	37

$$\text{OR} = \frac{x_7 + x_8}{2} = \frac{20+22}{2} = 21$$

Tepedeğer için en çok tekrarlanan değer bulunmalıdır. 20 değeri iki kez tekrarlandığı için tepedeğerdir.

TD=20, OR=21, $\bar{x}$=22 olduğundan sağa çarpık bir dağılımdır.

$$s^2 = \frac{\sum_{j=1}^{n} x_j^2 - \frac{\left(\sum_{j=1}^{n} x_j\right)^2}{n}}{n-1} = \frac{8012 - \frac{308^2}{14}}{13} = \frac{1236}{13} = 95.077$$

$$s = \sqrt{Varyans} = \sqrt{95.077} = 9.751$$

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{95.077}{14}} = 2.61$$

Örnek 2.29: 40 hasta üzerinde yapılan bir araştırmada kandaki kolestrol miktarları aşağıdaki gibi elde edilmiştir:

140	160	210	220	162	175	220	240	260	310
320	300	145	170	167	182	195	220	155	174
216	325	220	120	220	170	165	182	145	174
185	185	135	310	325	195	205	174	197	306

- **a)** Sınıf sayısını $k = 8$ alarak frekans ve birikimli frekans tablosunu oluşturunuz.
- **b)** Elde edilen frekans tablosunu kullanarak aritmetik ortalama, ortanca ve tepe değeri hesaplayınız. Dağılımın yönü hakkında ne söylersiniz?
- **c)** Elde edilen frekans tablosunu kullanarak kandaki kolestrol miktarları için çizgi grafiği ve histogram çiziniz.
- **d)** Elde edilen frekans tablosunu kullanarak varyans, standart sapma ve standart hatayı hesaplayınız.

a) $R = 325 - 120 = 205$

$$c = \frac{205 + 1}{8} = 25.75 \cong 26$$

As	Üs	S_i	f_i	S_{ad}	den daha az f_i	p_i	den daha çok f_i	p_i
				119.5	0	0.00	40	1.00
120	145	132.5	5	145.5	5	0.13	35	0.88
146	171	158.5	7	171.5	12	0.30	28	0.70
172	197	184.5	11	197.5	23	0.58	17	0.43
198	223	210.5	8	223.5	31	0.78	9	0.23
224	249	236.5	1	249.5	32	0.80	8	0.20
250	275	262.5	1	275.5	33	0.83	7	0.18
276	301	288.5	1	301.5	34	0.85	6	0.15
302	327	314.5	6	327.5	40	1.00	0	0.00

b)

A_s	$Ü_s$	S_i	f_i	b_i
120	145	132.5	5	-3
146	171	158.5	7	-2
172	197	184.5	11	-1
198	223	210.5	8	0
224	249	236.5	1	1
250	275	262.5	1	2
276	301	288.5	1	3
302	327	314.5	6	4

En çok frekansa sahip sınıf

$$\bar{x} = S_i + c\frac{\sum_{i=1}^{k} f_i b_i}{n} = 210.5 + 26 \times \frac{(-10)}{40} = 204$$

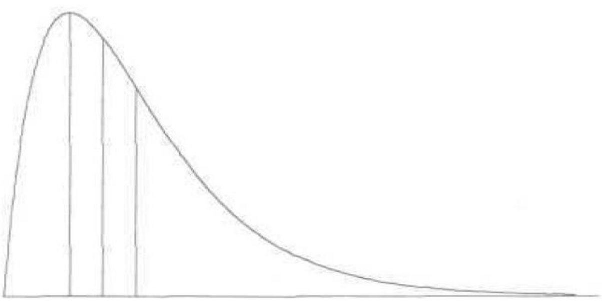
$$TD = A_s + c\left(\frac{F_1}{F_1 + F_2}\right) = 172 + 26 \times \left(\frac{4}{4+3}\right) = 186.86$$

$$j = \frac{n}{2} = \frac{40}{2} = 20$$

S_{ad}	Den daha az f_i
119.5	0
145.5	5
171.5	12
197.5	23
223.5	31
249.5	32
275.5	33
301.5	34
327.5	40

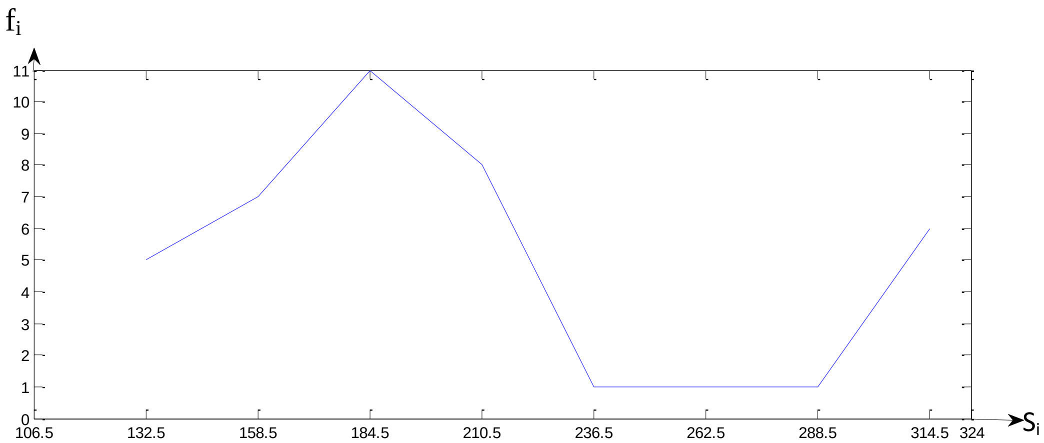
$j = 20$ değerini içeren aralık

$$OR = S_{ad} + c\left(\frac{\frac{n}{2} - N_1}{N_2 - N_1}\right) = 171.5 + 26 \times \left(\frac{20 - 12}{23 - 12}\right) = 190.41$$

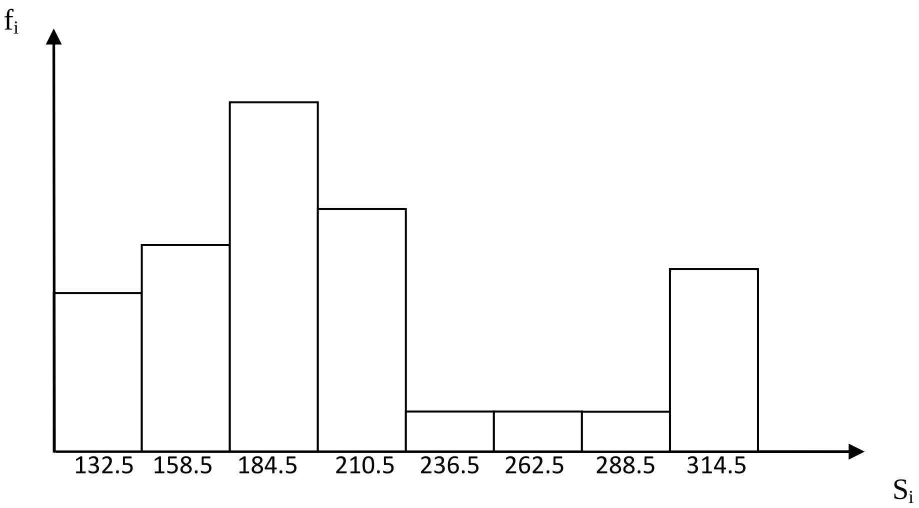


$TD < OR < \bar{X}$ olduğundan dağılım sağa çarpıktır.

c)



Kandaki kolestrol miktarı için çizgi diyagramı



Kandaki kolestrol miktarı için histogram

d) Varyansı $s^2 = c^2 \left(\dfrac{\sum_{i=1}^{8} f_i b_i^2 - \frac{\left(\sum_{i=1}^{8} f_i b_i\right)^2}{n}}{n-1} \right) = 26^2 \times \left(\dfrac{194 - \frac{(-10)^2}{40}}{39} \right) = 3319.33$ bulunur.

$s = \sqrt{3319.33} = 57.61$

$S_{\bar{x}} = \dfrac{S}{\sqrt{n}} = \dfrac{57.61}{\sqrt{40}} = 9.11$

Örnek 2.30: 40 kız öğrencinin kemiklerindeki büyüme miktarları verileri aşağıdaki gibi verilmiştir.

17.6	14.1	13.9	13.1	16.8	14.7	13.2	15.7	13.5	15.4
14.31	13.9	13.6	15.3	13.2	16.6	14.7	18.1	18.1	15.6
15.5	15.4	14.7	17.6	14.4	19.8	13.6	14.7	14.7	13.9
17.8	18.2	13.8	13.1	13.5	13.6	14.9	14.7	14.1	13.2

a) Kemiklerdeki büyüme miktarlarının frekans ve birikimli frekans tablosunu düzenleyiniz.

b) Elde edilen frekans tablosunu kullanarak aritmetik ortalama, ortanca ve tepe değeri hesaplayınız. Dağılımın yönü hakkında ne söylersiniz?

c) Elde edilen frekans tablosunu kullanarak kemiklerdeki büyüme miktarları için çizgi grafiği ve histogram çiziniz.

d) Elde edilen frekans tablosunu kullanarak varyans, standart sapma ve standart hatayı hesaplayınız.

a) $R = 19.8 - 13.1 = 6.7$

$$c = \frac{6.7 + 0.1}{8} = 0.85 \cong 0.9$$

As	Üs	S_i	f_i	SAD	den daha az f_i	den daha az p_i	den daha cok f_i	den daha cok p_i
				13.05	0	0.00	40	1
13.1	13.9	13.5	14	13.95	14	0.35	26	0.65
14.0	14.8	14.4	10	14.85	24	0.60	16	0.40
14.9	15.7	15.3	7	15.75	31	0.78	9	0.22
15.8	16.6	16.2	1	16.65	32	0.80	8	0.20
16.7	17.5	17.1	1	17.55	33	0.83	7	0.17
17.6	18.4	18	6	18.45	39	0.98	1	0.02
18.5	19.3	18.9	0	19.35	39	0.98	1	0.02
19.4	20.2	19.8	1	20.25	40	1	0	0

b)

As	Üs	S_i	f_i	b_i
13.1	13.9	13.5	14	-3
14.0	14.8	14.4	10	-2
14.9	15.7	15.3	7	-1
15.8	16.6	**16.2**	**1**	**0**
16.7	17.5	17.1	1	1
17.6	18.4	18	6	2
18.5	19.3	18.9	0	3
19.4	20.2	19.8	1	4

En çok frekansa sahip sınıf

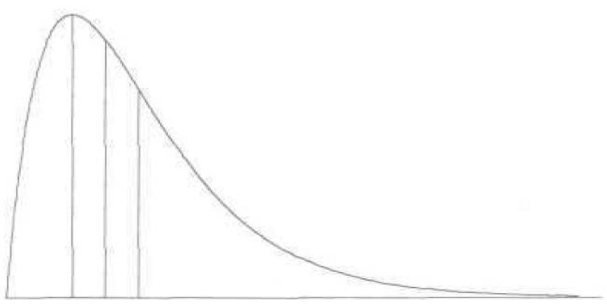
$$\bar{x} = S_i + c\frac{\sum_{i=1}^{k} f_i b_i}{n} = 16.2 + 0.9 \times \frac{-52}{40} = 15.03$$

$$TD = A_s + c\left(\frac{F_1}{F_1 + F_2}\right) = 13.1 + 0.9 \times (\frac{14}{14 + 4}) = 13.8$$

SAD	den daha az f_i
13.05	0
13.95	14
14.85	24
15.75	31
16.65	32
17.55	33
18.45	39
19.35	39
20.25	40

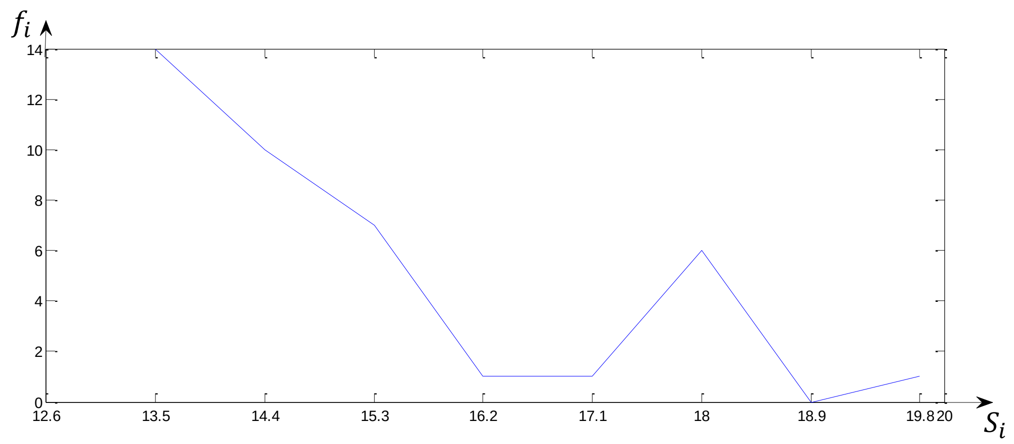
$j = \frac{n}{2} = \frac{40}{2} = 20$ değerini içeren aralık

$$\mathrm{OR} = S_{\mathrm{ad}} + \mathrm{c}\left(\frac{\frac{\mathrm{n}}{2} - \mathrm{N}_1}{\mathrm{N}_2 - \mathrm{N}_1}\right) = 13.95 + 0.9 \times \left(\frac{20 - 14}{24 - 14}\right) = 14.9$$

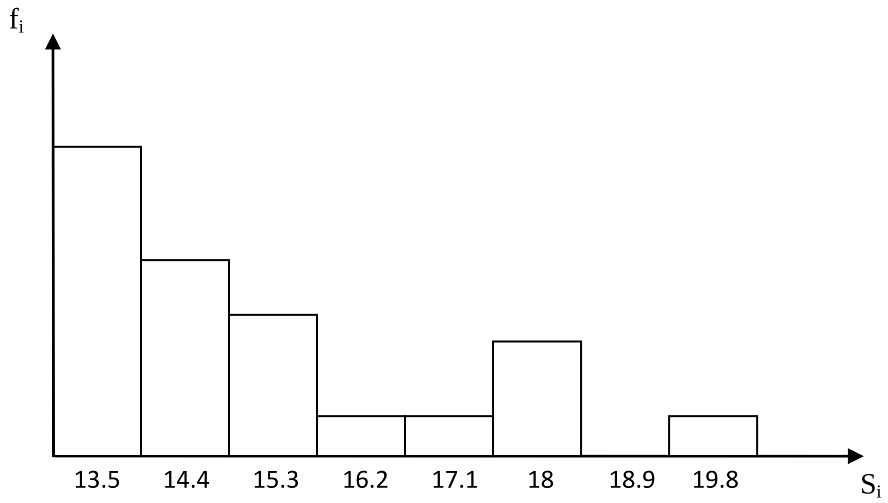


TD < OR < $\bar{x}$ olduğundan kemiklerdeki büyüme miktarlarının dağılımı sağa çarpıkdır.

c)



Kemiklerdeki büyüme miktarları için çizgi diyagramı



Kemiklerdeki büyüme miktarları için histogram

d) $s^2 = c^2 \left(\frac{\sum_{i=1}^{k} f_i b_i^2 - \frac{\left(\sum_{i=1}^{k} f_i b_i\right)^2}{n}}{n-1} \right) = 0.9^2 \left(\frac{214-(\frac{-52^2}{40})}{39} \right) = 3.041$

$S = \sqrt{3.041} = 1.744$

$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{1.744}{\sqrt{40}} = 0.276$