



VERİ BİLİMİ İÇİN TEMEL İSTATİSTİK

hafta-3

CEMİLE YILDIZÇAKAR

22.12.2020





İÇİNDEKİLER

Cemile YILDIZÇAKAR

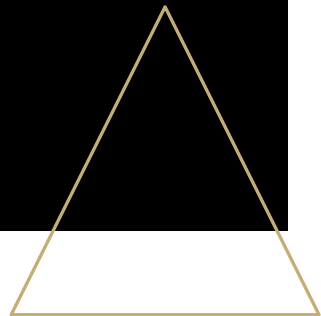


Değişim Ölçüleri

- Değişim:

Bir dağılımda ölçümler arasında gözlenen farklılık ve değişikliğe **değişim**, veriler arasındaki değişimden kaynaklanan farklılıkların istatistiksel ölçülerine ise **değişim ölçüleri** denir.

Cemile YILDIZÇAKAR



Nitel Veriler için Frekans Tablosu

Örnek;

Bir sağlık meslek yüksek okulunda ilgili programlara 400 öğrenci Hemşirelik, 300 öğrenci ameliyathane hizmetleri, 200 öğrenci Anestezi teknikeri, 100 öğrenci acil servis hizmetlerine kabul edilmiş olsun.

Sınıf	Frekans (fi)	Görelî Frekans Relative Frequency	Yüzde (%)	Kümülatif (%)
Hemşirelik	400	0,333333333	33,33333	33,33333333
Ameliyathane h.	300	0,25	25	58,33333333
Anestezi t.	200	0,166666667	16,66667	75
Acil Servis h.	100	0,083333333	8,333333	83,33333333
Diğer	200	0,166666667	16,66667	100
TOPLAM ::	1200	1	100	

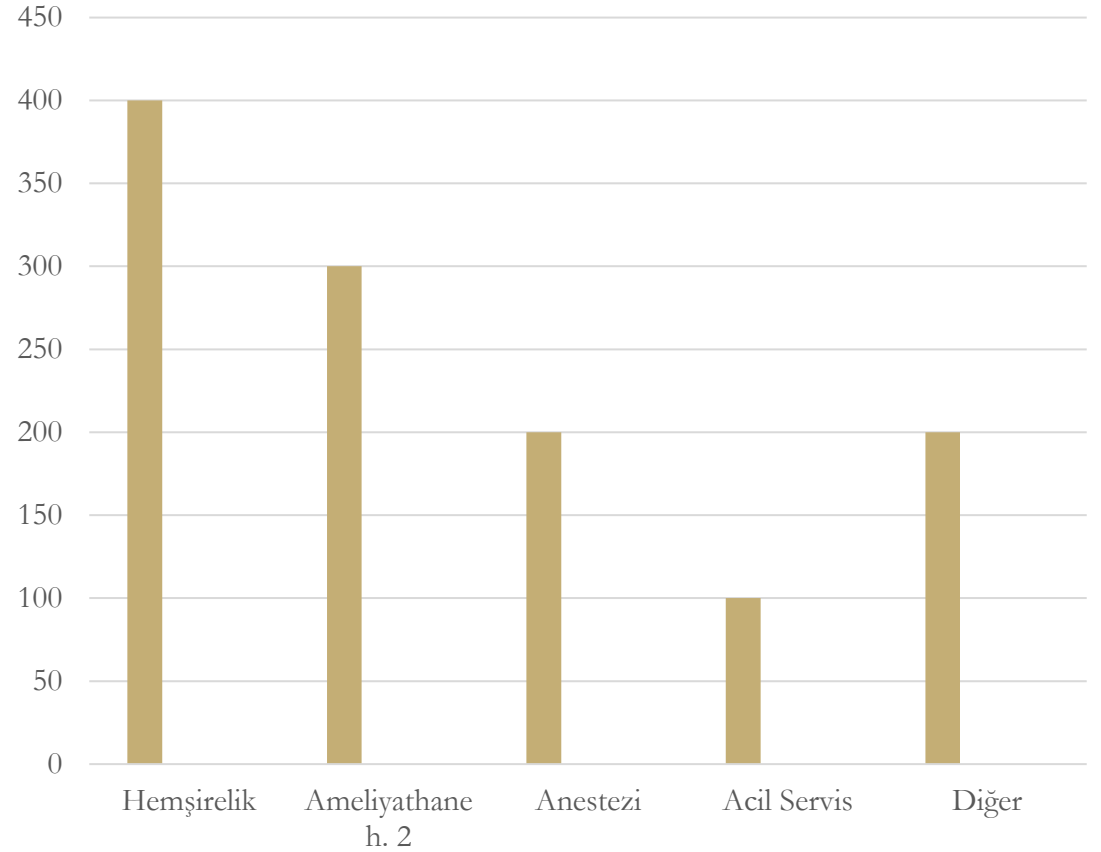
Cemile YILDIZÇAKAR

Çubuk grafiği (Bar Chart)

Dağılım çubukları grafiği, kesikli nicel verilerde ve nitel verilerde kullanılır. Çubuk grafiğinde sınıflar, tabanları eşit ve birbirine bitişik olmayan dikdörtgenlerle temsil edilir.

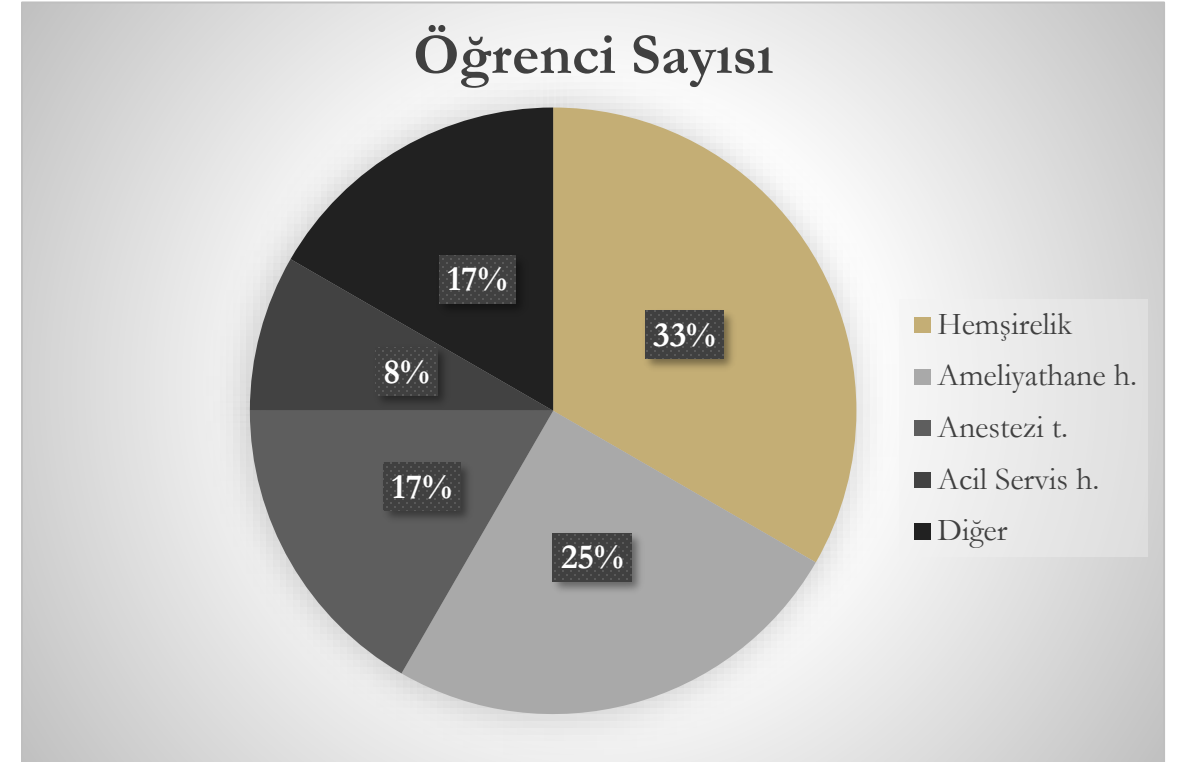
Cemile YILDIZÇAKAR

Bar Chart



Paste Grafiđi (Pie Chart)

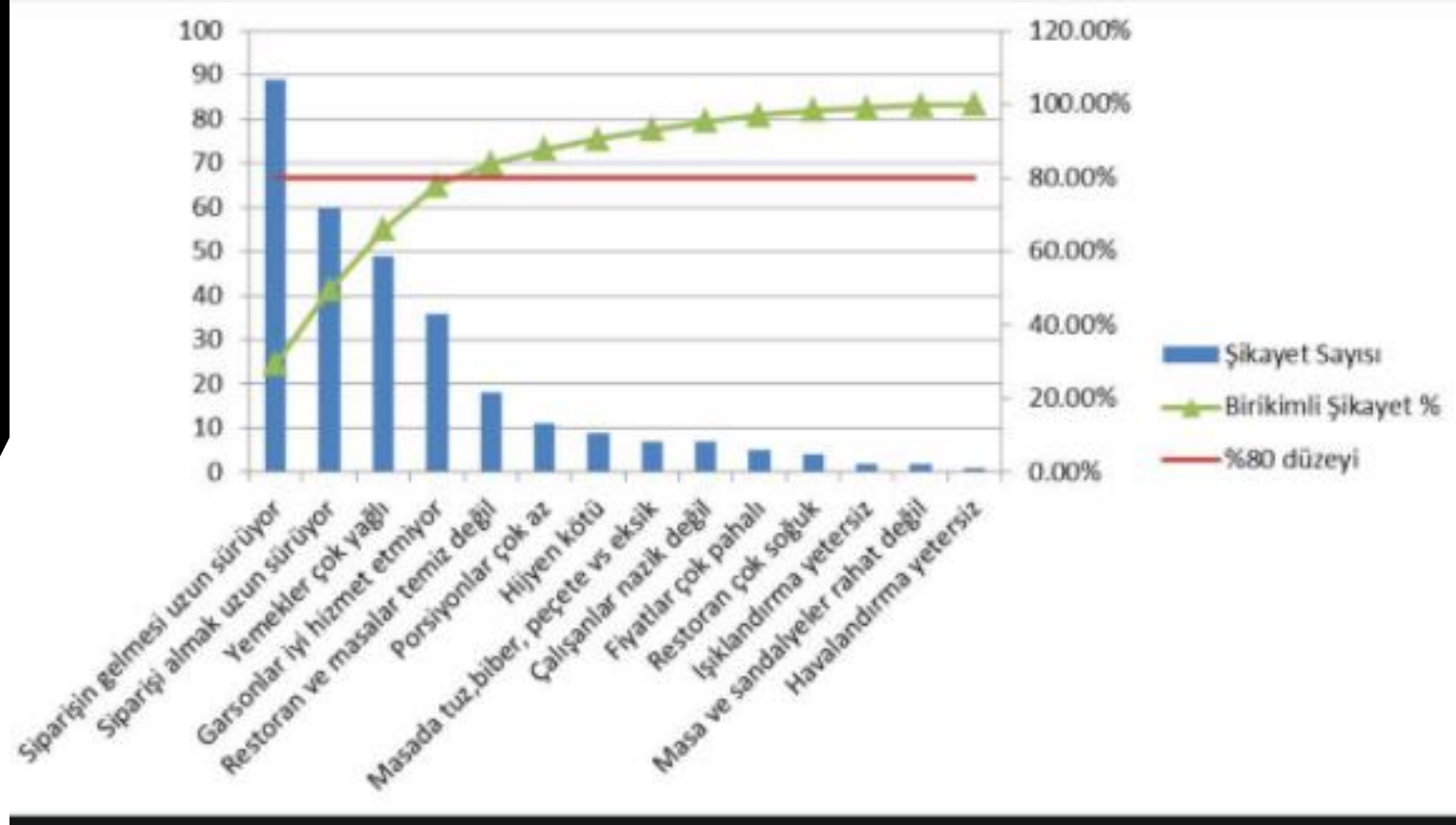
Her sınıfa düşen frekansın bir dairenin parçası ile gösterildiđi grafik türüdür. Bu grafiđi çizebilmek için görel frekanslar hesaplanır. Her sınıfa ilişkin görel frekans 3600 ile çarpılarak o sınıfa ilişkin daire dilimleri bulunur. Tüm sınıflar için yapıldığında daire tamamlanmış olur. Daha çok sınıflandırılabilen verilerde kullanılır.



Cemile YILDIZÇAKAR

Pareto Diagramı

Soldan sağa doğru azalan sırada yüksekliğe göre düzenlenmiş nitel değişken kategorilerini içeren bir çubuk grafiğidir.



Nicel Veriler İçin Tanımlayıcı Yöntemler

NOKTA GARFIĞI (DOT PLOT)

- Her bir verinin tek bir nokta ile gösterildiği grafikdir.
- Veri setindeki boşluklar, kümeler, verinin yayılımı net olarak görülebilir.

Cemile YILDIZÇAKAR

data set:

6, 3, 6, 3, 5,
7, 4, 6, 5, 3
4, 4, 5, 1, 0,
3, 2, 2, 1, 1

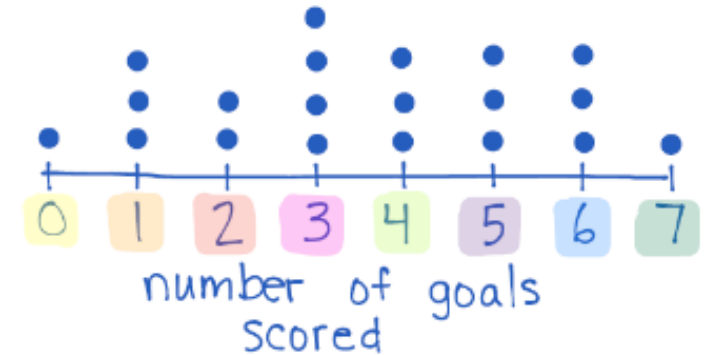
number of goals	frequency
0	1
1	3
2	2
3	4
4	3
5	3
6	3
7	1

← there is one "0" in the data set

← there are four "3"s in the data set

data set:

6, 3, 6, 3, 5,
7, 4, 6, 5, 3
4, 4, 5, 1, 0,
3, 2, 2, 1, 1



Gövde Yaprak Grafiği (Steam and Leaf display)

44, 46, 47, 49, 63, 64, 66, 68, 68, 72, 72, 75, 76, 81, 84, 88, 106

Stem	Leaf
4	4 6 7 9
5	
6	3 4 6 8 8
7	2 2 5 6
8	1 4 8
9	
10	6

- Veri setinde yer alan bir değerin bir kısmının gövde diğer kısmının yaprak olarak ayrılarak gösterilmesidir.

Cemile YILDIZÇAKAR

Gruplandırılmış Frekans Dağılımı Oluşturma

- **Veri Seti (S):** 20 kişilik bir sınıftaki öğrencilerin matematik test puanları

Ham Veri

96	90	80	67	60	51	40	30	51	60	60	67	80	90	51	60	67	60	60	51
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Büyükten Küçüğe Sıralanmış Veri

96	90	90	80	80	67	67	67	60	60	60	60	60	60	51	51	51	51	40	30
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

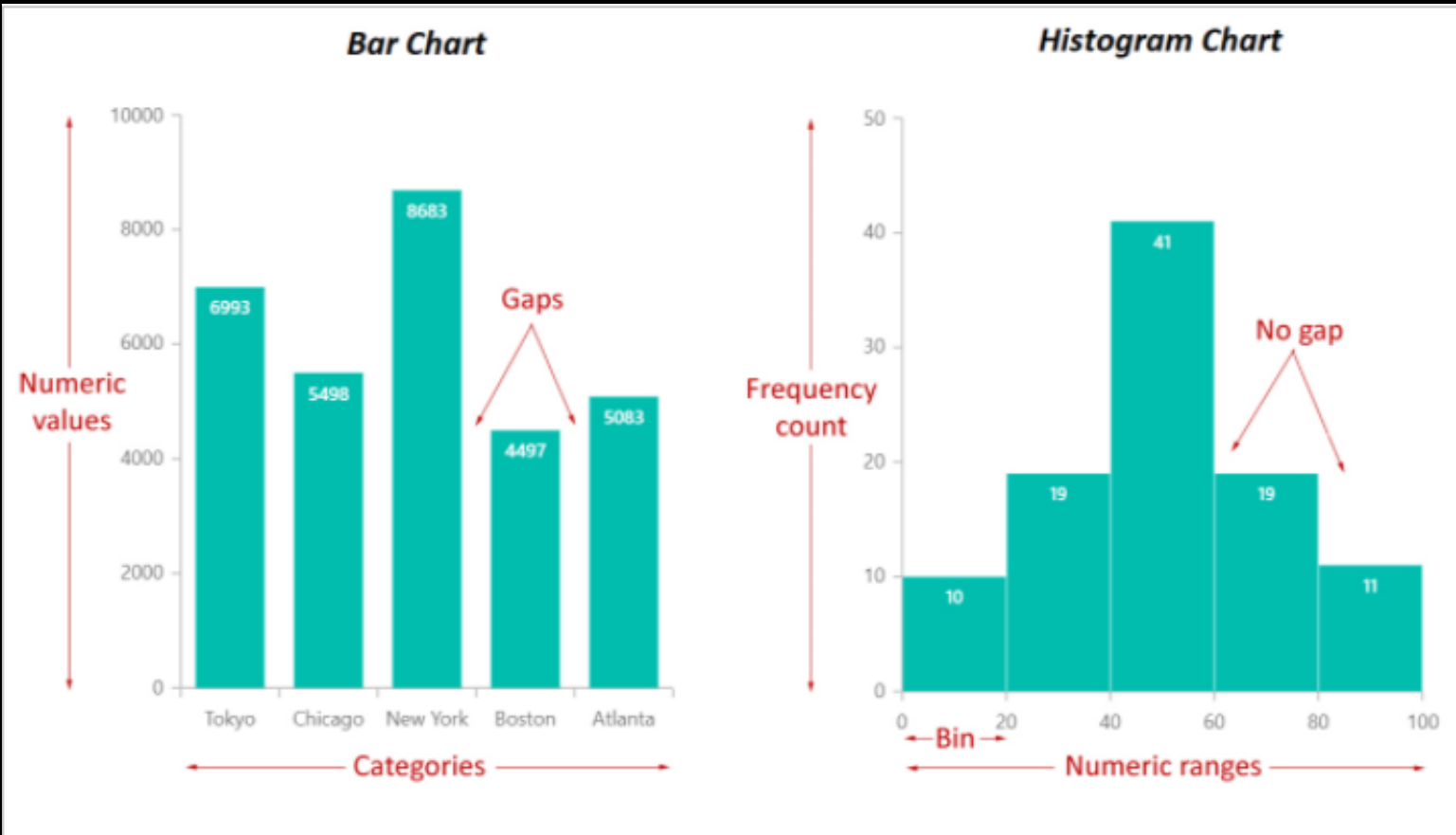
Interval	Frequency	Kümülatif Frekans	Relative Frequency	Percentage
30 - 47	2	2	0,1	10
48 - 65	10	12	0,5	50
66 - 83	5	17	0,25	25
84 - 101	3	20	0,15	15
total::	20		1	100

□ Sınıf sayısı: 4 $number\ of\ classes \cong \sqrt{number\ of\ observations}$

□ Sınıf genişliği: 17 $((96-30)/4)$ $class\ width = \frac{highest\ score - lowest\ score}{number\ of\ classes}$

**Sınıflar ayırık, sınıf genişliği sabit olmalı.

Cemile YILDIZÇAKAR



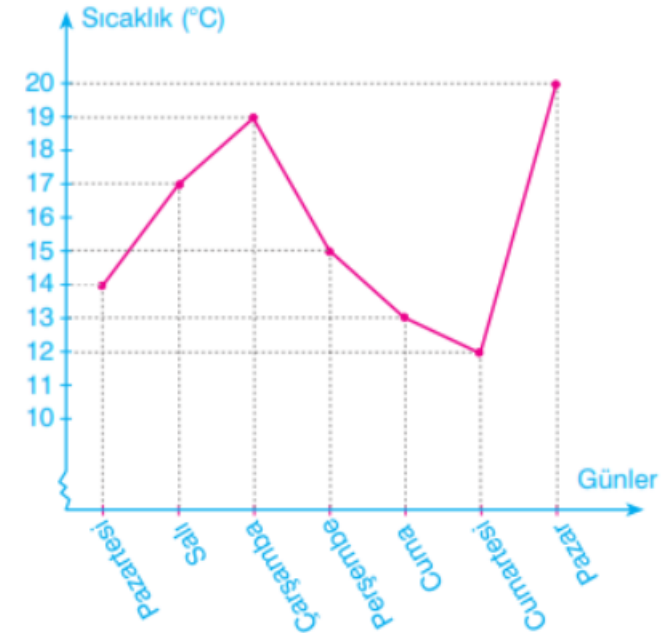
Cemile YILDIZÇAKAR

Diğer Grafikler

Cemile YILDIZÇAKAR

Çizgi Grafiği (Line Chart)

Aşağıdaki grafikte bir yerleşim biriminde bir hafta boyunca ölçülen hava sıcaklık değerleri gösterilmiştir.

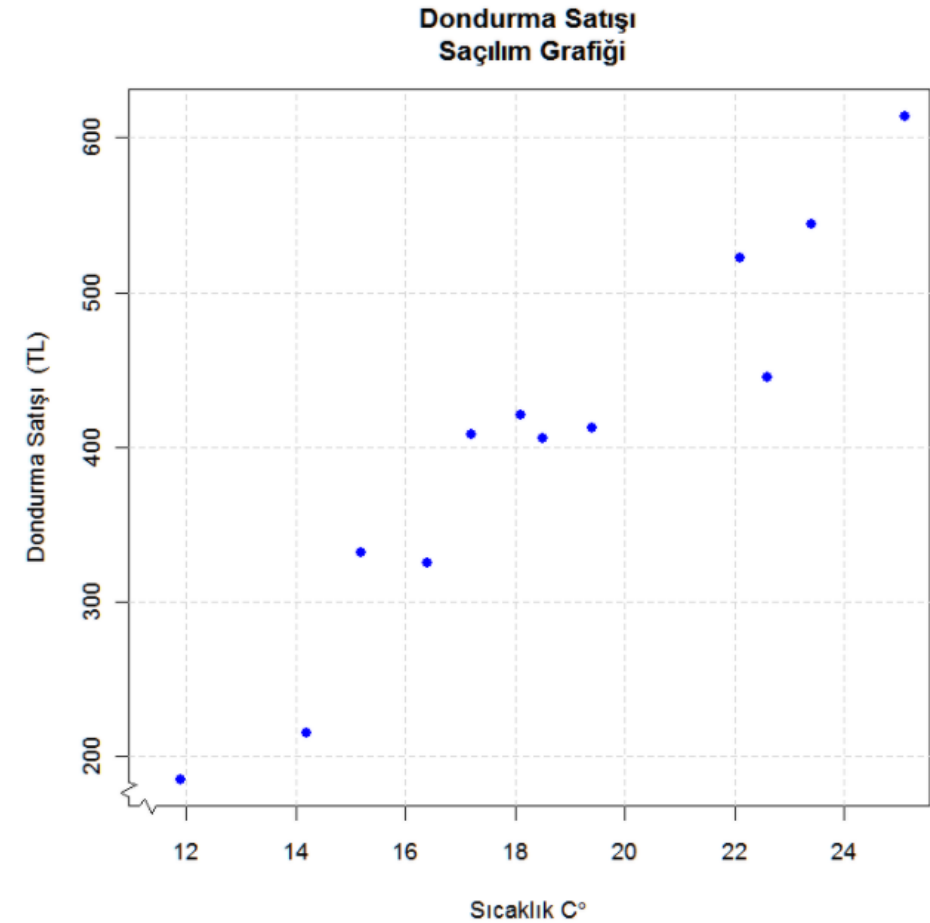


Diğer Grafikler

Sıcaklık (C°)	Dondurma Satışı (TL)
14,2	215
16,4	325
11,9	185
15,2	332
18,5	406
22,1	522
19,4	412
25,1	614
23,4	544
18,1	421
22,6	445
17,2	408

Cemile YILDIZÇAKAR

Serpme Diagramı (Scatter Diagram)



Çapraz Tablo

Türkiye AB'ye üye olmalı * Cinsiyet Crosstabulation

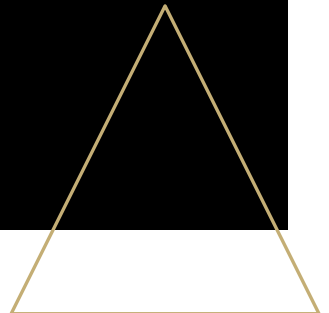
			Cinsiyet		Total
			Kadin	Erkek	
Türkiye AB'ye üye olamali	Kesinlikle Katiliyorum	Count	15	28	43
		% within Cinsiyet	11,5%	16,1%	14,1%
	Katiliyorum	Count	36	45	81
		% within Cinsiyet	27,7%	25,9%	26,6%
	Kararsizim	Count	38	40	78
		% within Cinsiyet	29,2%	23,0%	25,7%
	Katilmiyorum	Count	24	19	43
		% within Cinsiyet	18,5%	10,9%	14,1%
	Kesinlikle Katilmiyorum	Count	17	42	59
		% within Cinsiyet	13,1%	24,1%	19,4%
Total	Count	130	174	304	
	% within Cinsiyet	100,0%	100,0%	100,0%	



Merkezi Eğilim Ölçüleri (Measures of central tendency)

- Merkezi Eğilim Ölçüleri, belli bir özelliğe ya da değişkene ilişkin ölçme sonuçlarının, hangi değer etrafında toplandığını gösteren ve veri grubunu özetleyen ölçülerdir.
- Konum ölçüleri olarak da bilinir.

Cemile YILDIZÇAKAR



ORTALAMA (MEAN)

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

Sample mean

\bar{X}

X-bar

Population mean

μ

Greek letter (mu)

Cemile YILDIZÇAKAR

The **sample mean** is represented by *x bar* \bar{x} . It is given by the formula

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Örneklem
veri sayısı

Örneklemde
gözlenen veriler



The **population mean** is represented by the Greek letter *mu* (μ). It is given by the formula

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N}$$

Kitledeki veriler

Kitle veri sayısı

Veri setindeki her bir değişken hesaba katıldığı için bir verinin değişmesi tüm ortalamayı etkiler.

Cemile YILDIZÇAKAR

Gruplanmış Seriler İçin Aritmetik Ortalama

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

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

f : frekans

k: grup sayısı

i = 1, 2, 3, ..., k

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

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

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i} = \frac{51(1) + 66(3) + \dots + 94(7)}{1 + 3 + 4 + 5 + 7} \\ &= \frac{1605}{20} = 80,25\end{aligned}$$

Kaynak: Hamdi Emeç

Cemile YILDIZÇAKAR

Sınıflanmış Seriler İçin Aritmetik Ortalama

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

f : frekans

k : sınıf sayısı

i = 1, 2, 3, ..., k

m : sınıf orta noktası

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

- Sınıflanmış serilerde her bir sınıf içindeki değerlerin neler olduğu bilinmediğinden dolayı ve yalnızca her bir sınıfın frekans değerleri bilindiğinden dolayı sınıfı temsil etmek üzere sınıf orta noktaları hesaplamada kullanılır.

Kaynak: Hamdi Emeç

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

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

$$\bar{x} = \frac{\sum_{i=1}^k m_i f_i}{\sum_{i=1}^k f_i} = \frac{33(2) + 39(6) + \dots + 63(1)}{30}$$
$$= \frac{1398}{30} = 46,6 \text{ kg.}$$

Cemile YILDIZÇAKAR

ORTANCA (MEDIAN)

- Bir veri集中的 tüm değerlerin büyükten küçüğe sıralandığında orta noktasıdır.
- Ortanca veri setini iki eşit parçaya ayırır.
- Eşit aralıklı, oran ve sıralama ölçme düzeyinde ölçülen değişkenler için kullanılır.
- Ortancanın konumunu belirlemek için;

$$ortanca(OR) = \begin{cases} x_j & , j = \frac{n+1}{2} \text{ n tek} \\ \frac{x_j + x_{j+1}}{2} & , j = \frac{n}{2} \text{ n çift} \end{cases}$$

$$\frac{MED + IAN}{2}$$

Cemile YILDIZÇAKAR

5, 13, 9, 7, 1, 9, 2, 9, and 11

put in
ascending order

1, 2, 5, 7, 9, 9, 9, 11, 13

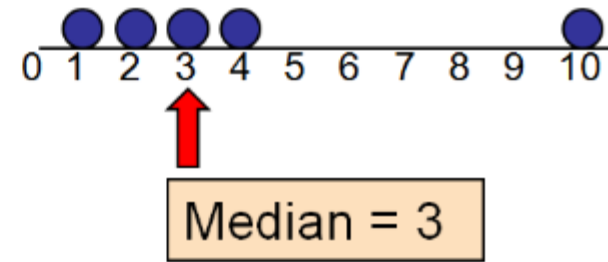
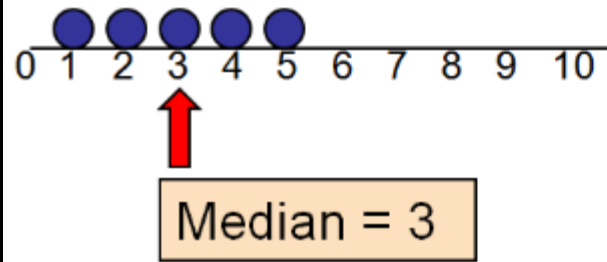
Median
(middle value)

1 2 4 7 8 9 18 17

3 3

Median

7.5



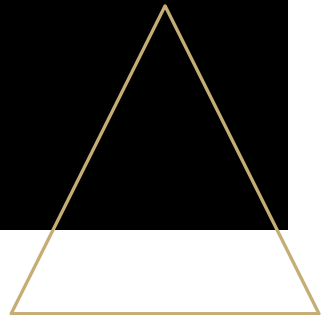
Cemile YILDIZÇAKAR

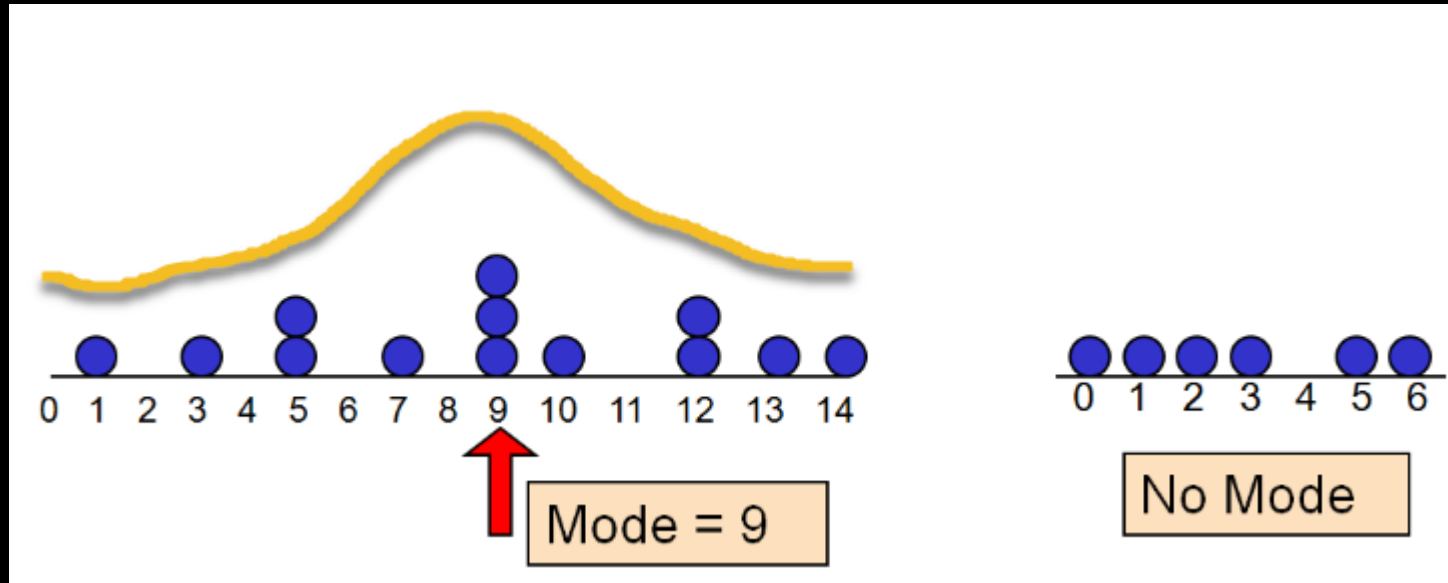


TEPE DEĞER (MODE)

- Bir veri grubunda en çok tekrar eden değere mod denir. Yani en fazla frekansa sahip değer olarak tanımlanır.
- Hiçbir aritmetik işlem gerektirmez.
- Bazı durumlarda, en yüksek frekansa sahip değer iki veya daha fazla sayıda olabilir. Bu durumda veri setinin tek tepe değeri olmaz.
- Bir veri setinde frekanslar eşit ise tepe değeri yoktur.

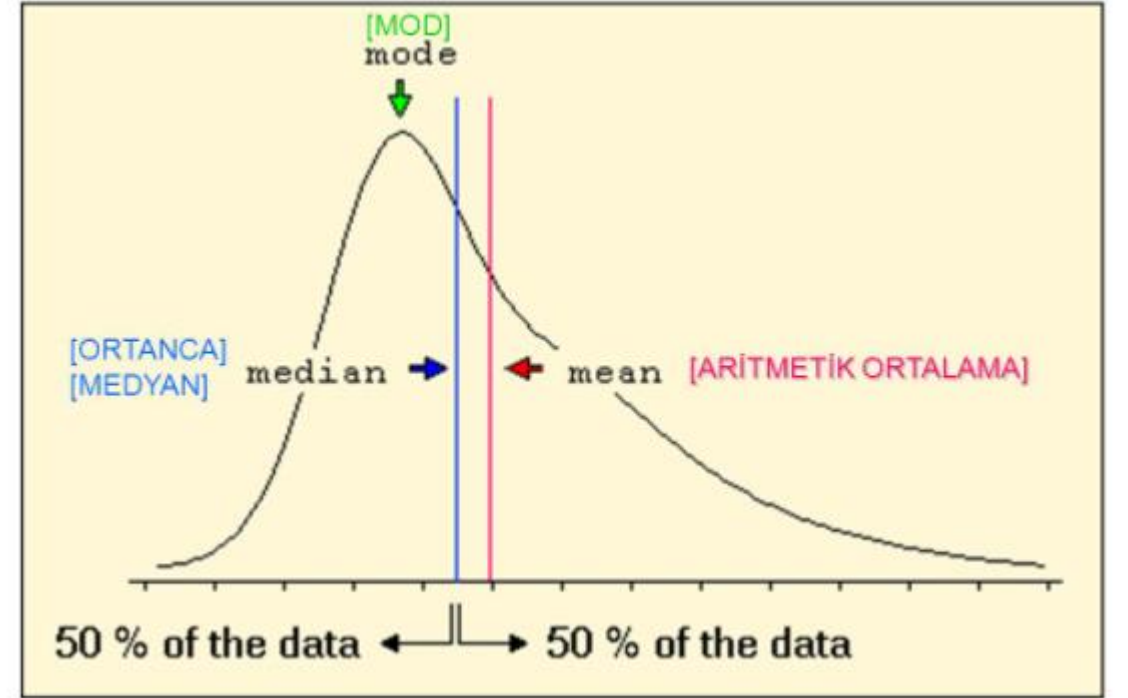
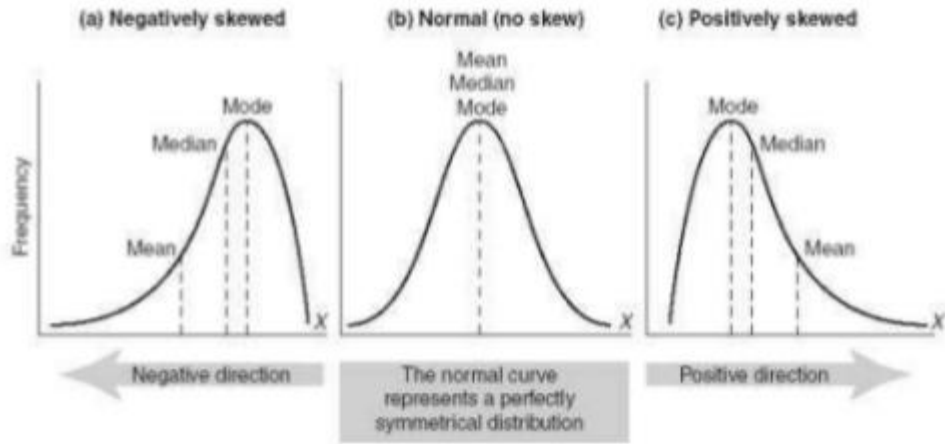
Cemile YILDIZÇAKAR





Cemile YILDIZÇAKAR

Dağılım Şekli



Cemile YILDIZÇAKAR

Teşekkür Ederim



LinkedIn

<https://www.linkedin.com/in/cemile-yildizcakar-34782248/>



Email

yildizcakar.cemile@gmail.com

Cemile YILDIZÇAKAR

A life without love
is like a year
without summer.

A SWEDISH PROVERB