

DATE : 29.02.2024

DT/NT : NT

LESSON : STATISTICS

**SUBJECT: GRAPHICAL
REPRESENT**

BATCH : 247

 techproeducation.com

 +1 (585) 304 29 59



TECHPRO
EDUCATION

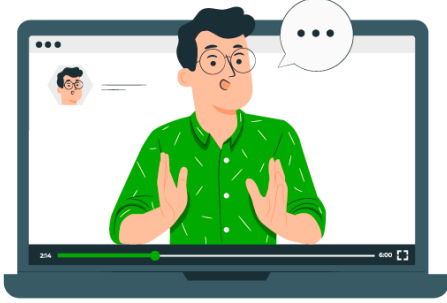


STATISTICS - 1

Data Science Program

Session - 3

Session - 2 Content



**Bu derste ne
öğreneceğiz?**

Graphical Represent

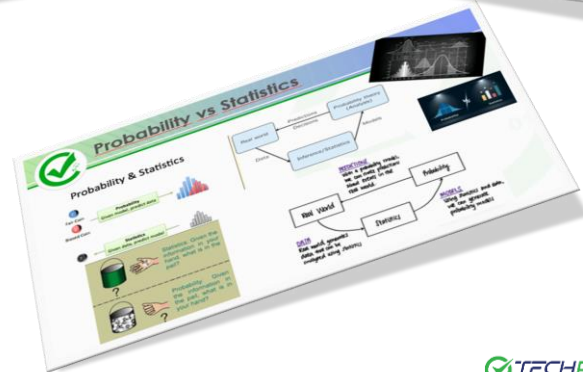
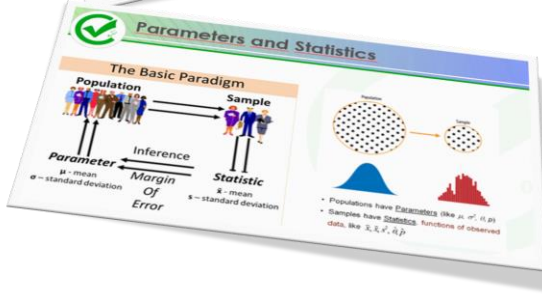
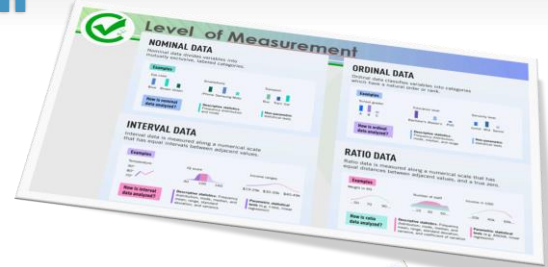
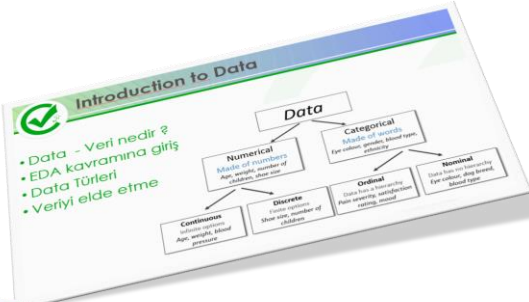
- Patterns
- Frequency Table
- Bar Chart
- Pie Chart
- Histogram



**Önceki dersten
hatırladıklarınızdan
bir cümle yazar
mısınız?**



Recap – Previous Lesson

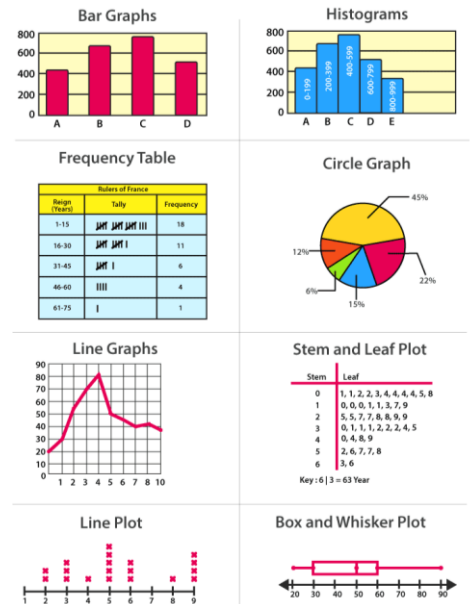


Sizi bugünkü derse
hazırlayacak **pre-class**
materyalleri ile
antrenman yaptınız
mı?

Data Visualization - Graphical Represent

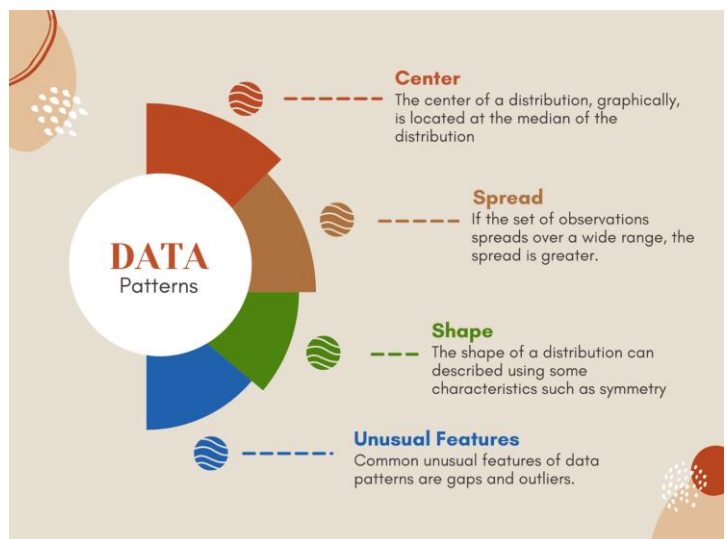
Graphical Representation of Data

- Center
- Spread
- Shape
- Unusual Features



Data Patterns

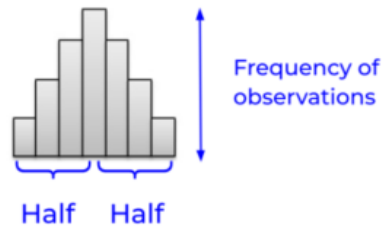
- Data Patterns
 - Center
 - Spread
 - Shape
 - Symmetric
 - Number of peaks
 - Skewness
 - Uniform
 - Unusual Features
 - Gaps
 - Outliers



Graphical Representation of Data

Center

- Dağılımın merkezi, grafiksel olarak dağılımın medyanındır olur
- Gözlemlerin yarısı her iki taraftadır
- Sütunun yüksekliği, gözlemlerin sıklığını gösterir.



Center

Spread

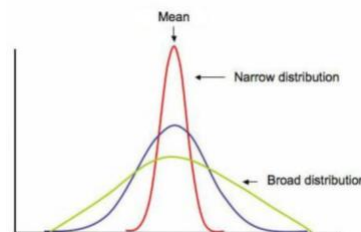
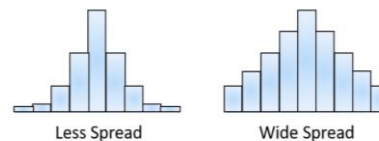
Shape

Unusual Fea.

Graphical Representation of Data

Spread

- Verilerin varyasyonu
- Gözlem kümesi geniş bir aralığa yayılıyorsa
- Gözlemler daha dar bir aralıkta tek bir değer etrafında ortalanırsa.....



Center

Spread

Shape

Unusual Fea.

Normally Distribution Videos

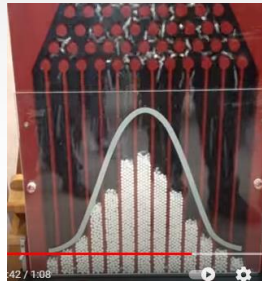
▶ Video-1

- <https://www.youtube.com/watch?v=Bampgm0HKDU>



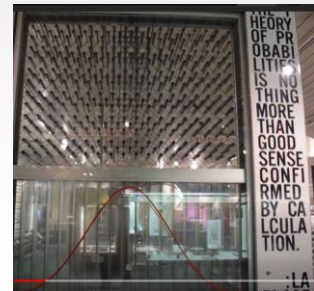
▶ Video-2

- <https://www.youtube.com/watch?v=4HpvBZhHOVI>

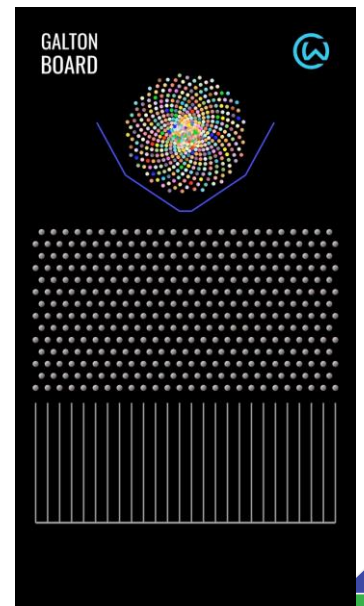
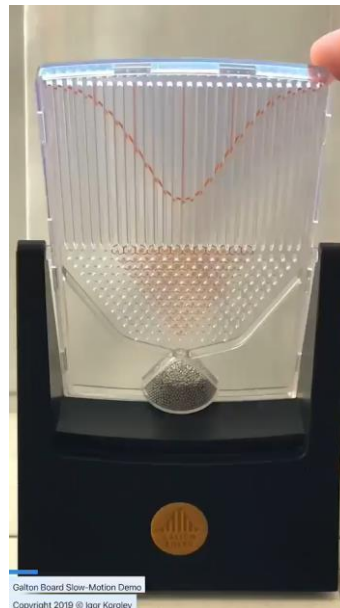


▶ Video-3

- <https://www.youtube.com/watch?v=Ph2DmwZMhGo>



Normal distribution is everywhere..

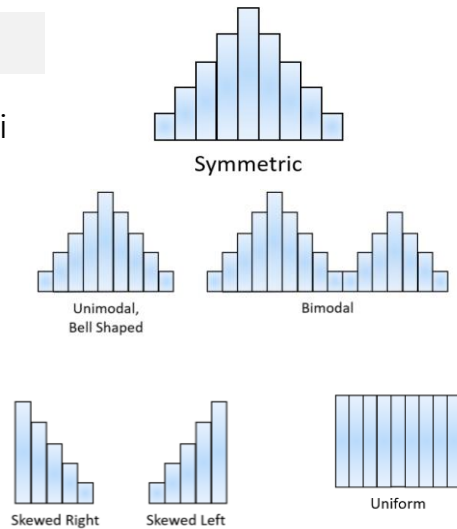


Graphical Representation of Data

Shape

Bir dağılımın şekli aşağıdaki özellikler kullanılarak tanımlanabilir.

- Symmetric
- Number of Peaks
- Skewness
- Uniform



Center

Spread

Shape

Unusual Fea.

Probability distributions

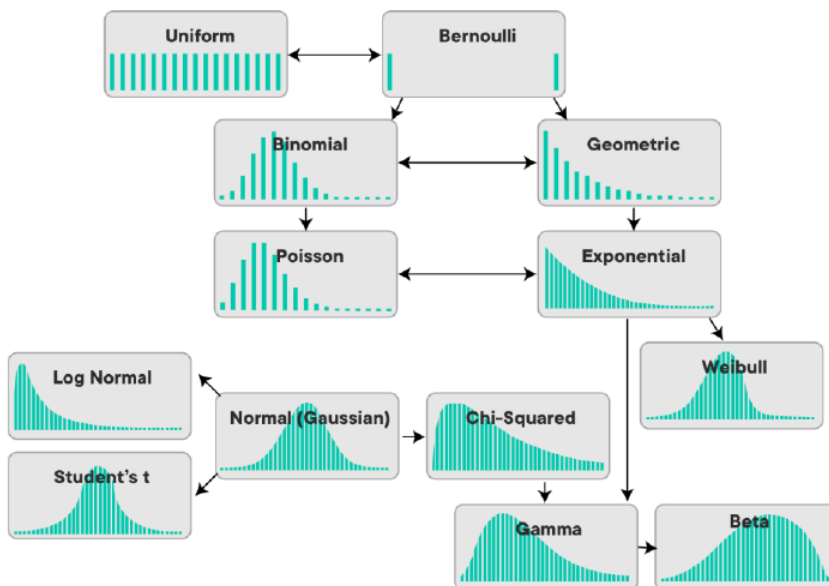
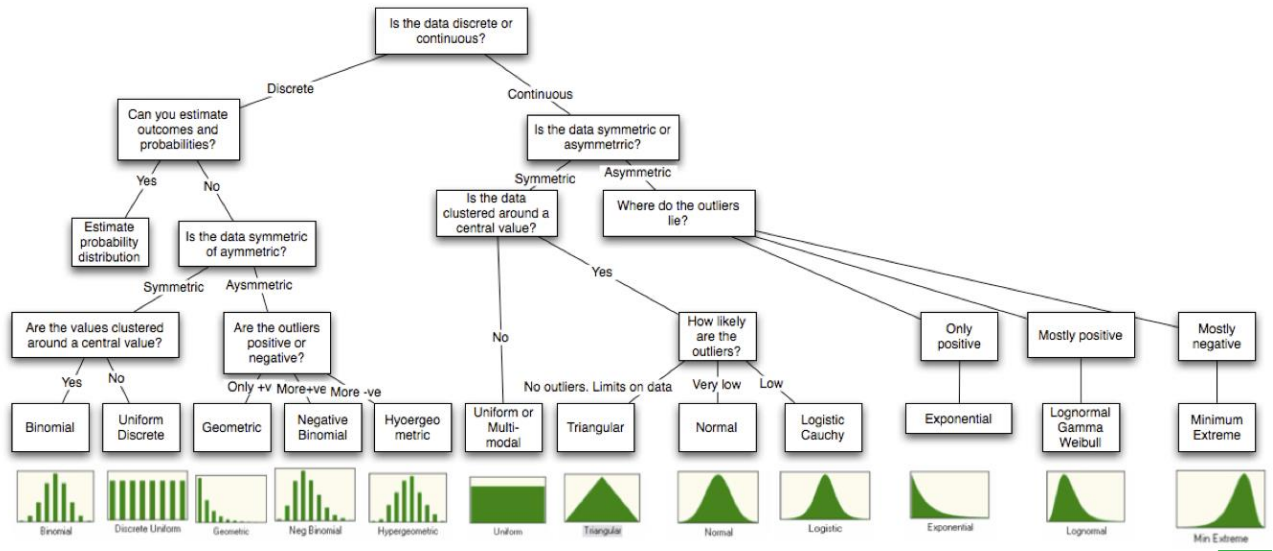


Figure 6A.15: Distributional Choices

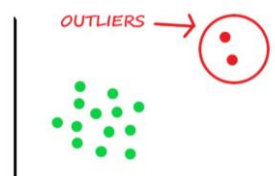
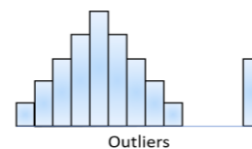
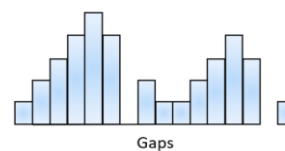


Graphical Representation of Data

Unusual Features

Veri modellerinin ortak
olağandışı özellikleri, boşluklar
ve aykırı değerlerdir

- Gaps
- Outliers



Center

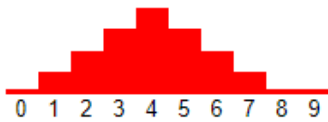
Spread

Shape

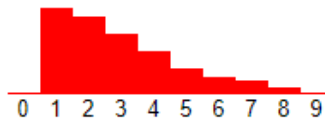
Unusual
Fea.



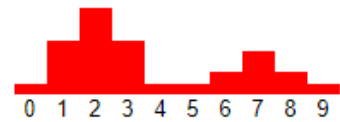
Data Patterns



Symmetric, unimodal,
bell-shaped



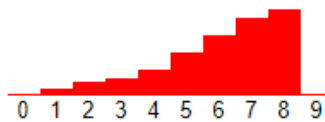
Skewed right



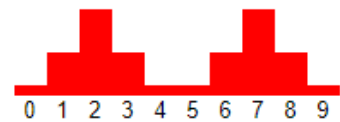
Non-symmetric, bimodal



Uniform



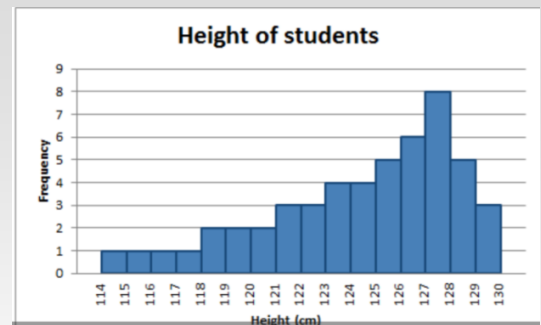
Skewed left



Symmetric, bimodal



TEST



Hangi pattern'e uygundur ?

- a. Right-skewed with no outliers
- b. Right-skewed with one outliers
- c. Left-skewed with no outliers
- d. Symmetric

Frequency

Descriptive istatistikte kullanılan yöntemler:

- Frekans Tabloları
- Şekiller ve Grafikler
- Histogram ve Frekans Poligonları
- Sütun ve Pasta Grafikleri

Developer Type	Frequency	Relative Frequency
Front-end Developer	25	0.25
Backend Developer	15	0.15
Full-stack Developer	20	0.20
Data Scientist	40	0.40

Sınıflar	Frekans, f
1 → 4	4
5 → 8	5
9 → 12	3
13 → 16	4
17 → 20	2

Üst Sınıf Limiti

Sıklıklar



Frequency

Frequency

- Bir veri değerinin meydana gelme sayısı

DATA VALUE	FREQUENCY
3	5
4	3
5	6
6	2
7	1

Relative Frequency

- bir şeyin ne sıklıkla gerçekleştiğinin tüm sonuçlara bölünmesi

DATA VALUE	FREQUENCY	RELATIVE FREQUENCY
2	3	$\frac{3}{20}$ or 0.15
3	5	$\frac{5}{20}$ or 0.25
4	3	$\frac{3}{20}$ or 0.15
5	6	$\frac{6}{20}$ or 0.30
6	2	$\frac{2}{20}$ or 0.10
7	1	$\frac{1}{20}$ or 0.05

Cumulative Frequency

- Önceki relative frekansların birikimi

DATA VALUE	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
2	3	$\frac{3}{20}$ or 0.15	0.15
3	5	$\frac{5}{20}$ or 0.25	0.15 + 0.25 = 0.40
4	3	$\frac{3}{20}$ or 0.15	0.40 + 0.15 = 0.55
5	6	$\frac{6}{20}$ or 0.30	0.55 + 0.30 = 0.85
6	2	$\frac{2}{20}$ or 0.10	0.85 + 0.10 = 0.95
7	1	$\frac{1}{20}$ or 0.05	0.95 + 0.05 = 1.00



QUESTION

**En fazla 12 yıla kadar
(at most) yaşayanların
oranı nedir ?**

Data	Frequency	Relative Frequency	Cumulative Relative Frequency
0	2	$\frac{2}{19}$	0.1053
2	3	$\frac{3}{19}$	0.2632
4	1	$\frac{1}{19}$	0.3158
5	3	$\frac{3}{19}$	0.4737
7	2	$\frac{2}{19}$	0.5789
10	2	$\frac{2}{19}$	0.6842
12	2	$\frac{2}{19}$	0.7895
15	1	$\frac{1}{19}$	0.8421
20	1	$\frac{1}{19}$	1.0000

Graphs and Charts

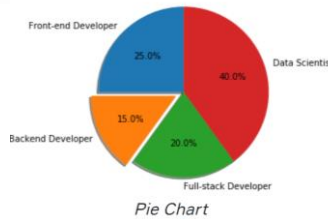
► Why Charts ?

- Anlaşılabilirlik artırılır.
- Dikkat çekilecek hususlar belirtilir.
- Dağılımın biçimi hakkında bilgi sağlanır.
- Tahmin kolaylaşır



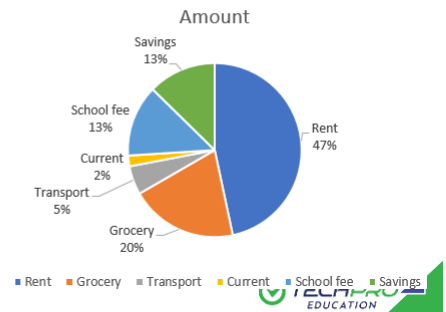
Pie Charts

- Genelde nominal ve ordinal değişkenlerle kullanılır
- Daire toplamda %100 ü tamamlayacak şekilde pasta dilimleri şeklinde kesilerek gösterilir
- Her dilim değişkenin niteliğini sunmuş olur



Pie Chart Examples

1	Expenses	Amount
2	Rent	7000
3	Grocery	3000
4	Transport	800
5	Current	300
6	School fee	2000
7	Savings	1900
8		
9		
10		
11		
12		
13		

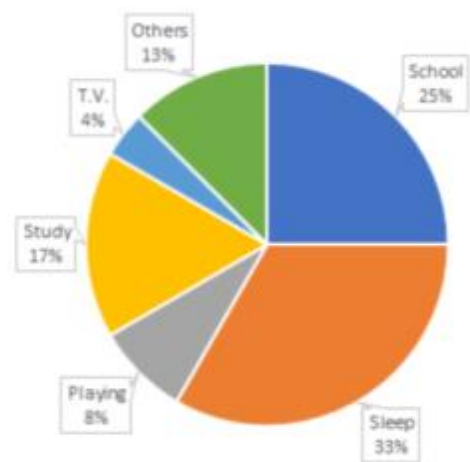


Pie Chart

Örnek

- Dilim yüzdesi hesaplama

Activity	No. of Hours	Measure of central angle
School	6	$(\frac{6}{24} \times 360)^\circ = 90^\circ$
Sleep	8	$(\frac{8}{24} \times 360)^\circ = 120^\circ$
Playing	2	$(\frac{2}{24} \times 360)^\circ = 30^\circ$
Study	4	$(\frac{4}{24} \times 360)^\circ = 60^\circ$
T. V.	1	$(\frac{1}{24} \times 360)^\circ = 15^\circ$
Others	3	$(\frac{3}{24} \times 360)^\circ = 45^\circ$



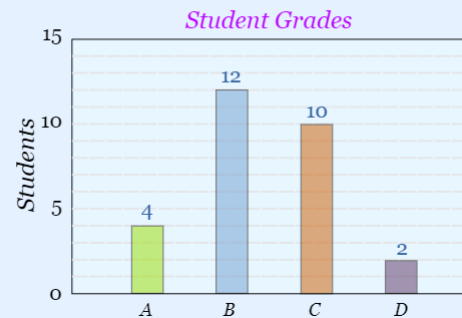
Bar Charts

- Genelde nominal ve ordinal değişkenlerle kullanılır
- Barların (sütunların) her biri bir değişkenin farklı değerlerini temsil eder
- Her bar yüksekliği her niteliğin frekansını gösterir



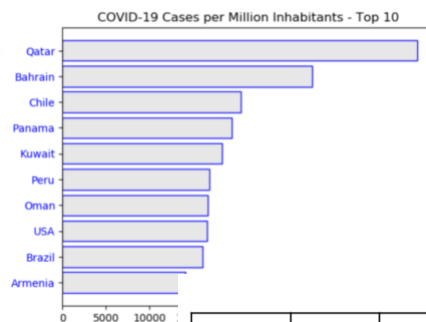
Grade:	A	B	C	D
Students:	4	12	10	2

bar graph:

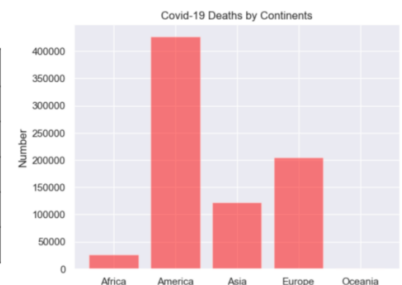


Bar Charts

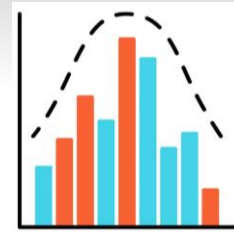
countriesAndTerritories	cases	deaths	popData2019	casesPer1M
Qatar	115661	193	2832071.0	40839.724710
Bahrain	47185	175	1641164.0	28750.935312
Chile	388855	10546	18952035.0	20517.849402
Panama	82790	1809	4246440.0	19496.331044
Kuwait	77470	505	4207077.0	18414.210151
Peru	549321	26658	32510462.0	16896.745423
Oman	83418	597	4974992.0	16767.464149
USA	5482416	171821	329064917.0	16660.591016
Brazil	3407354	109888	211049519.0	16144.808177
Armenia	41846	832	2957728.0	14148.021725



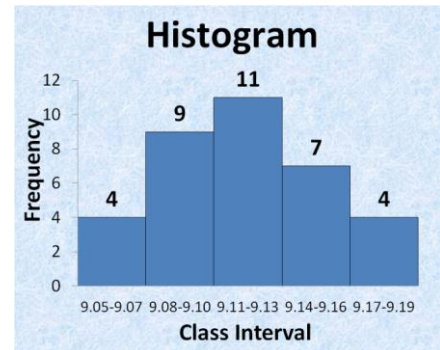
continent	cases	deaths
Africa	1119579	26260
America	11698368	427207
Asia	5606210	122034
Europe	3239237	205144
Oceania	25742	471



Histogram

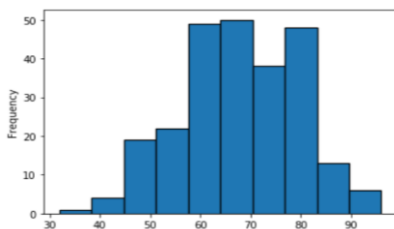


- Interval / Ratio değişkenlerle kullanılır
- Bir değişken için her bir niteliğin frekansını temsil eder
- Datanızın dağılımına iyi bir kuşbakışı bakma imkanı verir



Histogram

- Örnek bir Histogram çizim aşamaları



Histogram of scores on a Statistics test.

Interval's Lower Limit	Interval's Upper Limit	Class Frequency
32	38.4	1
38.4	44.8	4
44.8	51.2	19
51.2	57.6	22
57.6	64	49
64	70.4	50
70.4	76.8	38
76.8	83.2	48
83.2	89.6	13
89.6	96	6

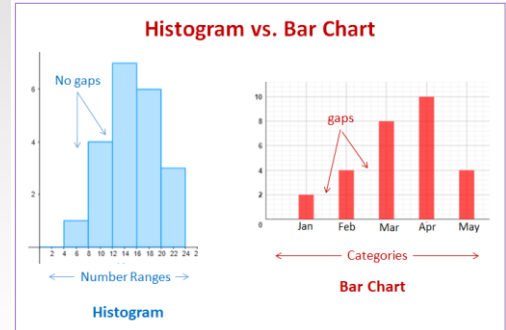
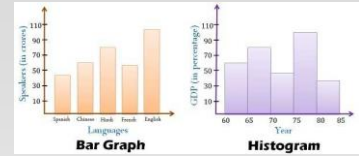
Bar Chart vs. Histogram

Bar Chart

- Kategoriler vardır
- ayırık değişkenlerin şematik bir karşılaştırması
- Kategorik veriler sunar
- Barlar arası boşlukludur

Histogram

- Grafik gösterime atıfta bulunur
- sürekli değişkenlerin frekans dağılımı
- Sayısal veriler sunar
- Barlar arası boşluk olmaz



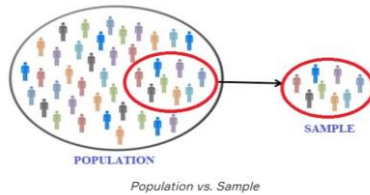
Populations & Samples

Populations & Samples

► Popülasyon

► Sample

- İstatistiki bir çalışma tamamen veri kümesi veya çözüm uzayının incelenmesine dayanır.



!! Sample'ları gözlemliyoruz ama popülasyonlarla ilgileniyoruz



Parameters & Statistics

**Population
Attributes**



Parameters

Sample Attributes



Statistics

Bir parametre, popülasyonun sayısal bir özetidir ve bir istatistik, örneklemin sayısal bir özetidir.





Task -1

EXERCISE 1. At what level are each of the following variables operationalized?

	Variable Name	Value Labels	Data Type
1	Age	1, 2, 3,	Continuous
2	Sex	male, female, other	
3	Class standing	freshman, sophomore, junior, senior	
4	Marital Status	married, single, divorced, widowed	
5	Median Household Income	annual household income in dollars	
6	How much you like the food on campus?	a lot, some, a little, other	
7	The number of toxic waste sites in your community	0, 1, 2, 3,	
8	The number of toxic waste sites in your community	0, 1–5, 6–10, 11+	
9	Your GPA	below average, average, above average	



Task -2

EXERCISE 2. The number of passengers of an airline company by years is given in the table below. Create a bar chart based on these data.

The number of passengers per year

Years	2010	2012	2013	2014	2015	2016	2017	2018	2019
Number of passengers (x1000)	5	7	13	10	20	22	17	16.5	27

EXERCISE 3. Create a frequency histogram from the data in the table below. What you can conclude about the shape of the distribution?

Income (In thousands of dollars)	Number of families
16–22	2
23–29	3
30–36	5
37–43	8
44–50	8
51–57	10



Central Tendency (Measure of Centre)

Merkezi Eğilim ve Dağılım Ölçüleri

Content

▶ Central Tendency (Measure of Centre)

Merkezi Eğilim Ölçüleri

- Mean
- Median
- Mode



▶ Dispersion (Measure of Spread)

Dağılım Ölçüleri

- Range
- IQR
- Standart Deviation
- Variation

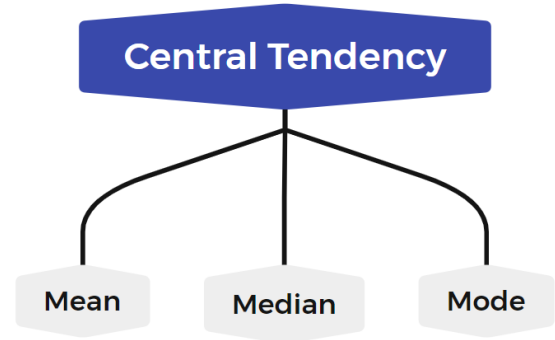
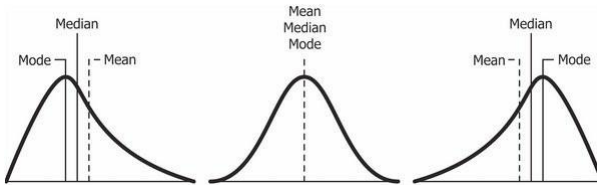


Central Tendency

Merkezi Eğilim

Tek değerle verileri en iyi tanımlama

- Ortalama
- Medyan (Ortanca)
- Mode (Tepe Değeri)



Mean (Average)

- Dataların toplamını, toplam gözlem sayısına bölmek
- Dağılımın yerinin belirlenmesinde kullanılır

Staff	Salary (thousand \$)
1	102
2	33
3	26
4	27
5	30
6	25
7	33
8	33
9	24

Population Mean	Sample Mean
$\mu = \frac{\sum_{i=1}^N x_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$
N = number of items in the population	n = number of items in the sample

Kitle Ortalaması: $\mu = \frac{\sum x}{N}$
 ↑
 "mü"

Örneklem Ortalaması: $\bar{x} = \frac{\sum x}{n}$
 ↑
 "x-bar"



Mean Example

Örnek-1

Örnek:

Aşağıdakiler küçük bir şirketin yedi çalışanın yaşlarıdır:

53 32 61 57 39 44 57

Kitle ortalamasını hesaplayın.

$$\mu = \frac{\sum x}{N} = \frac{343}{7} \quad \text{Yaşları yoplayın ve 7'ye bölün.}$$

$$= 49 \text{ years}$$

Çalışanların yaş ortalaması 49'dur.

Örnek-2

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

$$\bar{x} = \frac{10 \times 3 + 12 \times 5 + 15 \times 2 + 17 \times 6 + 20 \times 1 + 24 \times 4}{3 + 5 + 2 + 6 + 1 + 4}$$

$$\bar{x} = \frac{338}{21}$$

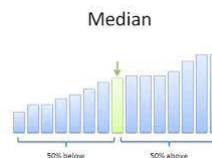
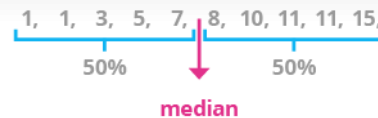
$$\bar{x} = 16.095$$

x	frequency
10	3
12	5
15	2
17	6
20	1
24	4



Median

- küçükten büyüğe sıralanmış bir veri kümesinin orta puanıdır
- Data sayısı tek ise median 1 değerdir ama çift sayı ise medianı bulurken ortadaki 2 değer ortalaması alınır
- Medyan, orta puandır. Örneklem büyüklüğü 9 ise, beşinci eleman medyandır.



Staff	Salary (thousand \$)
1	24
2	25
3	26
4	27
5	30
6	33
7	33
8	33
9	102

Median Formula $\left(\frac{n+1}{2}\right)^{\text{th}}$



Median Example

Örnek:

Yedi çalışanın ortanca yaşını hesaplayın.

53 32 61 57 39 44 57

Medyanı bulmak için verileri sıralayın.

32 39 44 53 57 57 61

Çalışanların ortanca yaşı 53'tür.

• Örnek-2

• Örnek-1



\$4000



\$15.000



\$20.000



\$33.000



\$1.800.000

Mean:

$$\mu = \frac{\sum X}{N}$$

$$\mu = \frac{\$4000 + \$15000 + \$20000 + \$33000 + \$1800000}{5}$$

$$\mu = \frac{\$1872000}{5} = \$374400$$

Median:

\$20000

Mean vs. Median

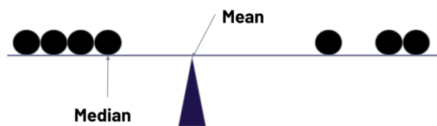


Mean



Median

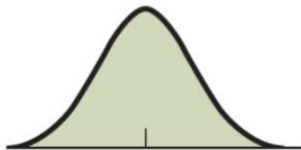
- Eğer skorların küçük bir kümesinde outlier varsa median daha iyidir.
- büyük data setlerinde outlier yoksa mean daha iyidir.
- Salary teklifinde median daha iyi olabilir



#1. Basic Definition		#4. Calculation	
Mean Mean can be referred to the simple average or arithmetic average of the given set of data or the quantities or the values.	Median Median can be defined as the middle most numerical in an ordered list (i.e. from lowest to highest or vice versa) of values.	Mean Mean can be calculated by adding up or taking up the sum of all the observations of the data set and then dividing that summation or the value obtained by the number of observations in the sample provided.	Median To calculate median, first one needs to be arranged the data set in an ascending or in a descending order and then the value which shall fall in the exact middle of the new data set or of the sample, will be the median.
#2. Actually meaning		#5. What does it represent	
Mean Mean can also be termed as arithmetic average.	Median Median can be meant as an positional average.	Mean Mean will represent the central gravity of the data set given.	Median The mid-point of the data set will be represented by the median.
#3. Type of distribution		#6. Outliers bias	
Mean For Mean, normal distribution would apply.	Median For median to be used and to be find it as more appropriate to use than mean, there should be skewed distribution.	Mean Mean will represent the central gravity of the data set given.	Median The mid-point of the data set will be represented by the median.

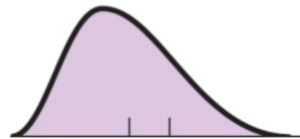
Mean vs Median

Symmetric Distribution



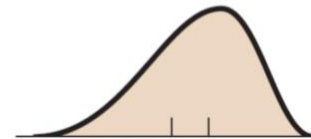
Mean = Median

Right-Skewed Distribution



Median < Mean

Left-Skewed Distribution



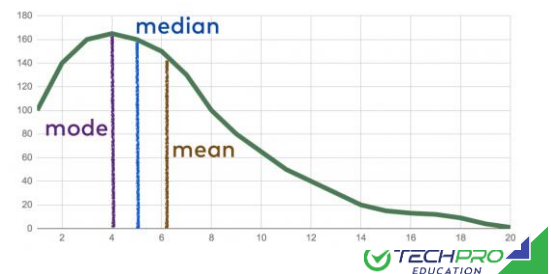
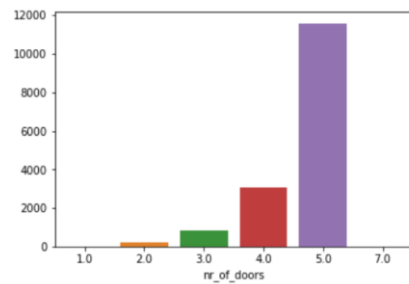
Mean < Median

Symmetric	Skewed right (positive)	Skewed left (negative)



Mode

- Mode tepe değeri diye adlandırılır
- Mode: Data setinde nın fazla karşılaşılan, en popüler değer
- hem numeric hem kategorik değişkenler için kullanılabiliyor
- Avantaj- Dezavantajları



Mode Example

• Örnek-1

Örnek:

Yedi çalışanın yaş grubunu bulun..

53 32 61 **57** 39 44 **57**

Mod 57, çünkü diğer veriler bir kez varken 57 iki kez tekrarlanıyor.

Ortalama-Mod-Medyan Karşılaştırılması

Örnek:

29 yaşında bir çalışan şirkete katılıyor ve çalışanların yaşları şimdi:

53 32 61 57 39 44 57 **29**

Ortalama, medyan ve modu yeniden hesaplayın. Bu yeni yaş eklendiğinde hangi merkezî eğilim ölçüsü etkilendi?

Mean = 46.5 Ortalama her değeri hesaba katar, ancak aykırı değerden etkilenir.

Median = 48.5 Ortanca ve mod uç değerlerden etkilenmez.

Mode = 57

• Örnek-2

Örnek 4:

Aşağıdaki verilerin modunu ve medyanını belirleyiniz.

120 100 130 100 160 130 86 100 94 90

Çözüm 3:

Verileri küçükten büyüğe sıralayalım.

1.değer	2.değer	3.değer	4.değer	5.değer	6.değer	7.değer	8.değer	9.değer	10.değer
86	90	94	100	100	100	120	130	130	160

Veri grubunda en çok tekrarlanan değer 100 olduğu için **Mod=100**

Veri sayısı n=10 → çift

$$\frac{n}{2} = \frac{10}{2} = 5. \text{değer} \rightarrow 100$$

$$\frac{n}{2} + 1 = \frac{10}{2} + 1 = 6. \text{değer} \rightarrow 100$$

$$\Rightarrow \text{Medyan} = \frac{100 + 100}{2} = 100$$

Statistic with Python

• Input

```
import numpy as np
from scipy import stats

salary = [102, 33, 26, 27, 30, 25, 33, 33, 24]

mean_salary = np.mean(salary)
print("mean:", mean_salary)

median_salary = np.median(salary)
print("median:", median_salary)

mode_salary = stats.mode(salary)
print("mode:", mode_salary)
```

• Output

```
mean: 37.0
median: 30.0
mode: ModeResult(mode=array([33]), count=array([3]))
```

Calculate Mean, Median and Mode with Python

YouTube Öneri Video



IN THIS VIDEO
WE WILL BE LOOKING AT

MODE **MEDIAN** **MEAN**
RANGE **STANDARD DEVIATION**

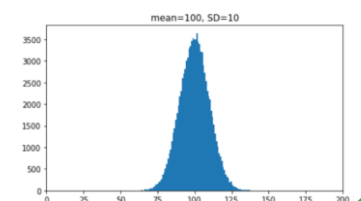
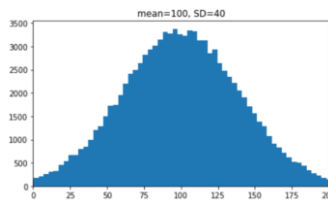
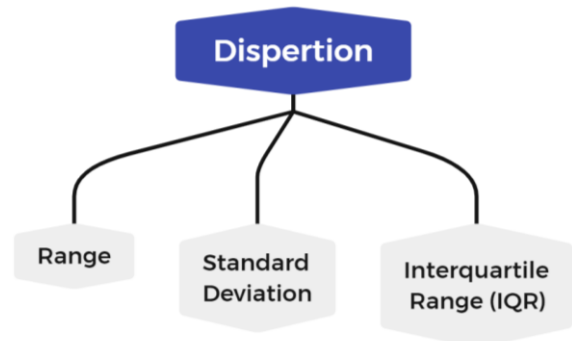
Mode, Median, Mean, Range, and Standard Deviation

<https://www.youtube.com/watch?v=mk8tOD0t8M0>

Dispersion (Measure of Spread)

► Dağılım Ölçüleri

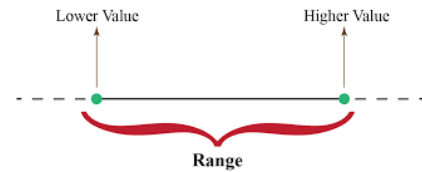
- merkezi eğilim) ölçüleri tek başına dağılımı karakterize etmez
- İki veri grubu ortalamasının eşit olması dağılımlarının aynı olmasını gerektirmez
- bir dağılım, merkezi eğilimin yaptığından daha fazlasını açıklar



Range

Aralık-Açıklık - Değişim Genişliği

- Bir veri kümesinin aralığı, kümedeki maksimum ve minimum veri girişleri arasındaki farktır
- Değişkenliğin en basit ölçüsüdür.



$$\text{Range} = \text{Largest} - \text{Smallest} = 9 - 2 = 7$$

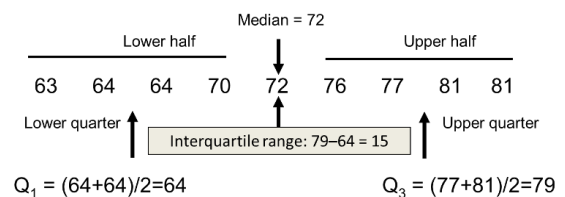
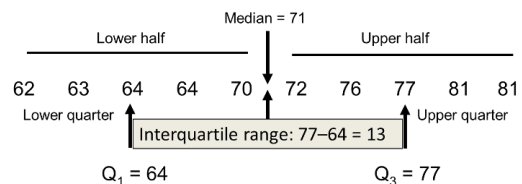
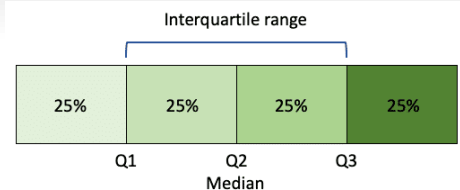


Inter Quartile Range (IQR)

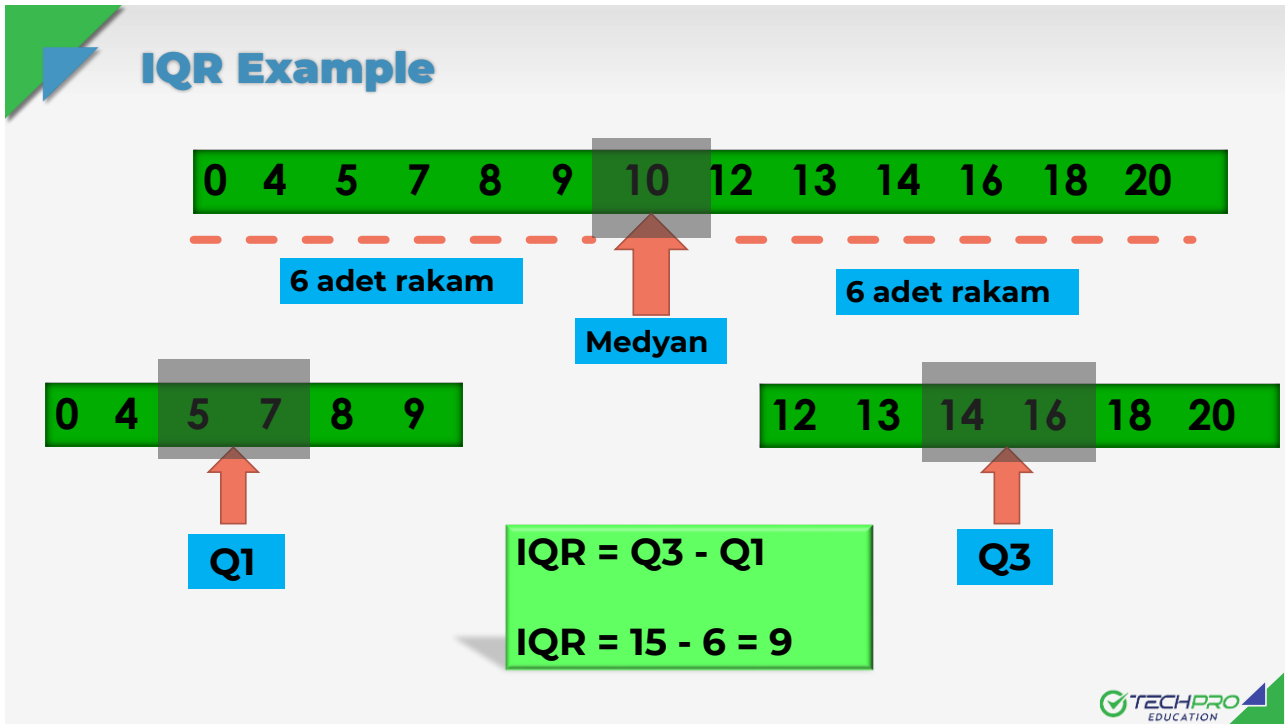
IQR

- bir sayı grubunu dörde bölen değerlerdir
- Q2 tüm datasetinin median'ıdır
- Q1, medianın altında kalan kısmın medianı'dır
- Q3, medianın üstünde kalan kısmın medianı'dır.

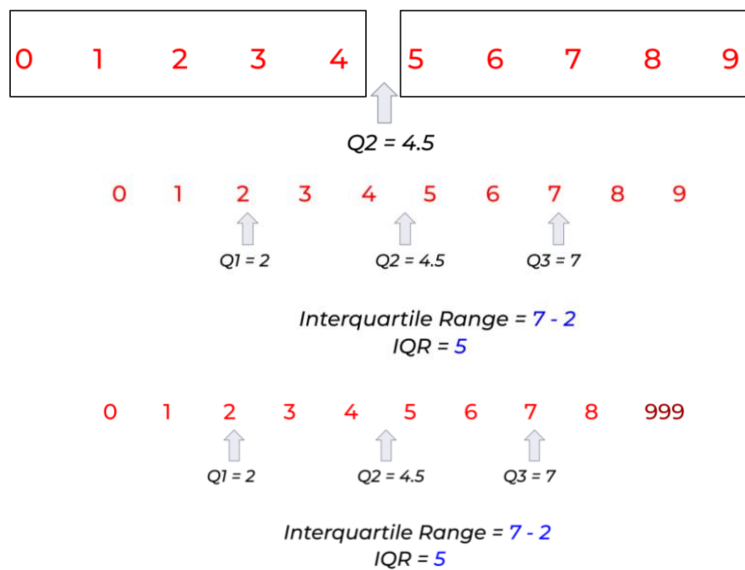
$$\text{IQR} = Q3 - Q1$$



IQR Example



IQR Example - 2



QUESTION

What is the

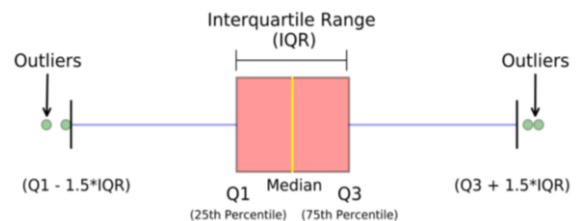
- mean
- Q1
- Q3
- Median
- IQR

27 28 30 32 34 38 41 42 43 44 46 53 56 62

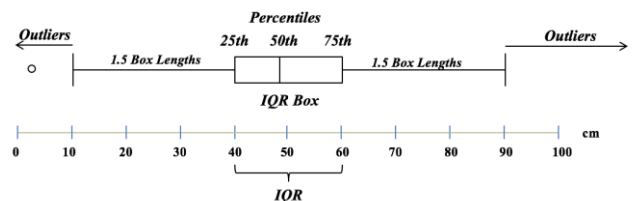
IQR

► Outlier Nasıl bulunur

- Outlier, Q1'in altında veya Q3'ün üzerinde 1.5 IQR' den fazla olan veri noktalarıdır
- list = [1, 5, 8, 10, 12, 15, 40]
- $Q1 - (1.5 * IQR) = 5 - 15 = -10$
- $Q3 + (1.5 * IQR) = 15 + 15 = 30$



Boxplot aka Box and Whiskers Plot



Variance (Population)

Varyans

- Varyans, ortalamadan farkların karelerinin ortalaması olarak tanımlanır
- Her bir skorun mean'den uzaklaştığı miktardır.

Variance

Sample variance

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

S^2 =sample variance

x_i =value of ith element

\bar{x} = sample mean

n =sample size

Population variance

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

σ^2 =population variance

x_i =value of ith element

μ =population mean

N =population size

sample variance $S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$

observation x_i mean \bar{x}

number of observations n

variance $\sigma^2 = \frac{\sum (x - \mu)^2}{N}$

element x mean μ

number of elements N



Variance Example

- Altındaki 4 değer için Varyans

0 1 5 6

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

0 1 5 6

Mean: $\mu = \frac{\sum x}{N} = \frac{0+1+5+6}{4} = \frac{12}{4} = 3$

Dev Sum of Squares: $SS = \sum (x - \mu)^2$

$$SS = (0 - 3)^2 + (1 - 3)^2 + (5 - 3)^2 + (6 - 3)^2$$

$$SS = 9 + 4 + 4 + 9 = 26$$

Variance: $\sigma^2 = \frac{\sum (x - \mu)^2}{N}$

$$\sigma^2 = \frac{26}{4} = 6.5$$

- Örnek-2

10 12 17 20 25 27 42 45

- Hem sample hem de popülasyon için bulalım.

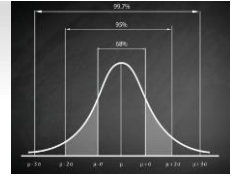


$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

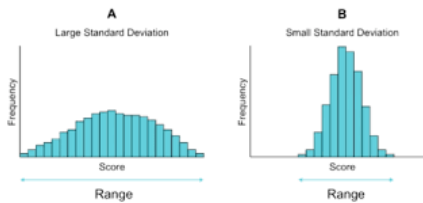


Standard Deviation



Standart Sapma

- Varyansın kareköküdür.
- Veriler ne kadar çok yayılırsa, standart sapma o kadar büyük olur.



Sample

Population

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

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

standard deviation $\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$

element x

mean μ

number of elements N



Std. Dev. Example

Staff	Salary (thousand \$)
1	24
2	25
3	26
4	27
5	30
6	33
7	33
8	33
9	102

$$\mu = \frac{24+25+26+27+30+33+33+33+102}{9}$$

$$\mu = \frac{333}{9} = 37$$

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

$$\sigma = \sqrt{\frac{(24-37)^2 + (25-37)^2 + (26-37)^2 + (27-37)^2 + (30-37)^2 + (33-37)^2 + (33-37)^2 + (33-37)^2 + (102-37)^2}{9}}$$

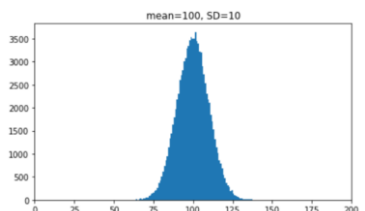
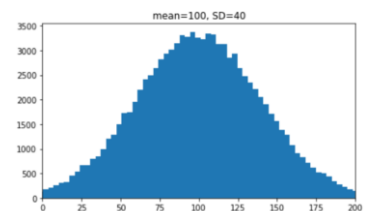
$$\sigma = \sqrt{\frac{(-13)^2 + (-12)^2 + (-11)^2 + (-10)^2 + (-7)^2 + (-4)^2 + (-4)^2 + (-4)^2 + (65)^2}{9}}$$

$$\sigma = \sqrt{\frac{169+144+121+100+49+16+16+16+4225}{9}}$$

$$\sigma = \sqrt{\frac{4856}{9}}$$

$$\sigma = \sqrt{539.55}$$

$$\sigma = 23.22833518$$

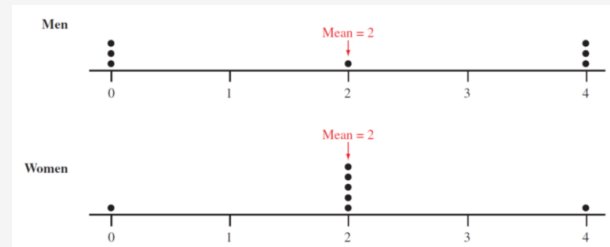


Std. Dev. Example - 2

Men: 0 0 0 2 4 4 4

Women: 0 2 2 2 2 2 4

- Bir aile için ideal çocuk sayısını cevaplayanlardan oluşan yukardaki 2 grup dağılım için (7 şer kişi),
- Varyansı nedir



Men:
$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{24}{6}} = \sqrt{4} = 2.0.$$

Women:
$$s = 1.2$$



Std. Dev with python

input :

```
import numpy as np

salary = [102, 33, 26, 27, 30, 25, 33, 33, 24]

print("Range: ", (np.max(salary)-np.min(salary)))

print("Variance: ", (np.var(salary)))

print("Std: ", (np.std(salary)))
```

output :

```
Range: 78
Variance: 539.5555555555555
Std: 23.22833518691246
```



Empirical Rule

3 Sigma Kuralı

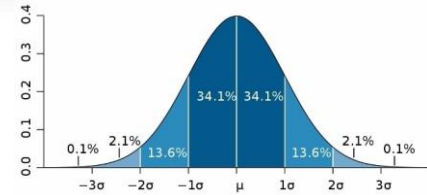
- Three Sigma Rule veya 68-95-99.7 kuralı diye de bilinir. .

- Ampirik Kural :**

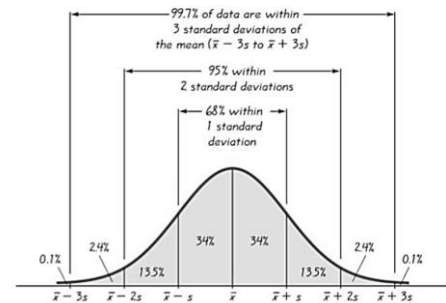
1. % 68'de kural,
= (Ortalama - standart sapma) ve (Ortalama + standart sapma)

2. % 95'de kural,
= (Ortalama - 2 × standart sapma) and (Ortalama + 2 × standart sapma)

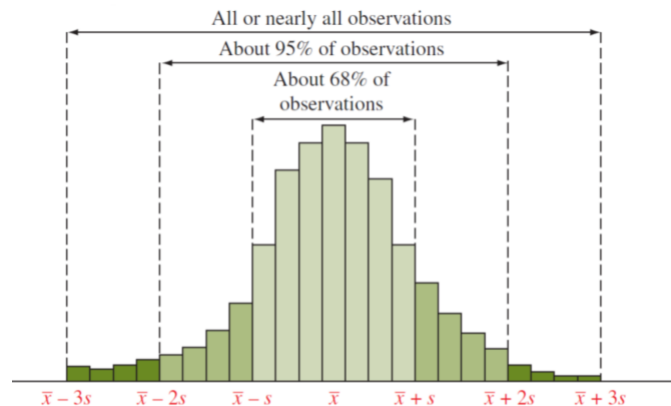
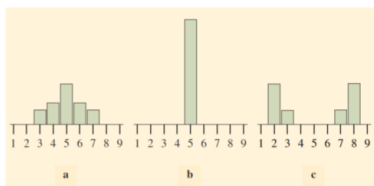
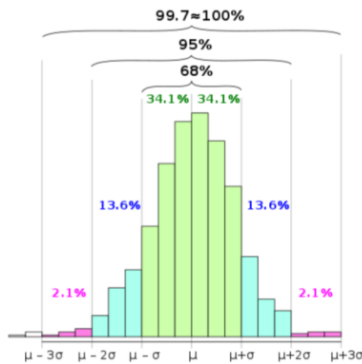
3. % 97.7'de kural,
= (Ortalama - 3 × standart sapma) and (Ortalama + 3 × standart sapma)



The Empirical Rule



Empirical Rule



Statistics Practice-1



Python Notebook zamanı