









BATCH : **B150** Data Science
LESSON : **STATISTICS-1**
DATE : 17.04.2023
SUBJECT : **Graphical Represent**

 techproeducation
 techproeducation
 techproeducation
 techproeducation
 techproedu

 techproeducation.com  info@techproeducation.com  +1 (917) 768-7466



STATISTICS - 1

Data Science Program
Session -2



Session - 2 Content

Content

• Graphical Represent

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



RECAP

**Herkes önceki dersten hatırladığı
1 cümle yazabilir mi?**

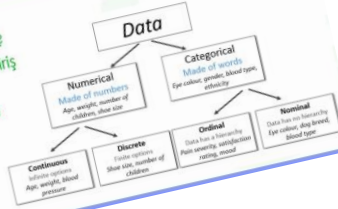




Recap – Previous Lesson

Introduction to Data

- Data - Veri nedir ?
- EDA kavramına giriş
- Data Türleri
- Veriyi elde etme



Level of Measurement

NOMINAL DATA

Nominal data divides variables into mutually exclusive, ordered categories.



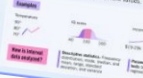
ORDINAL DATA

Ordinal data divides variables into categories which have a natural order or rank.



INTERVAL DATA

Interval data is measured along a numerical scale that has equal intervals between adjacent values.



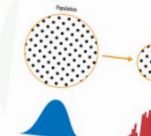
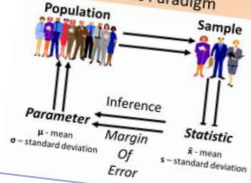
RATIO DATA

Ratio data is measured along a numerical scale that has equal distances between adjacent values, and a true zero.



Parameters and Statistics

The Basic Paradigm



- Populations have **Parameters** (like μ , σ^2 , ρ)
- Samples have **Statistics** (functions of observed data, like \bar{x} , s^2 , r)

Probability vs Statistics

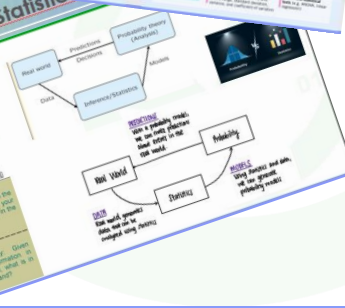
Probability & Statistics

Probability: Given the model, what is the chance of this event occurring?

Statistics: Given the data, what is the best model to fit the data?

Probability: Given the model, what is the chance of this event occurring?

Statistics: Given the data, what is the best model to fit the data?



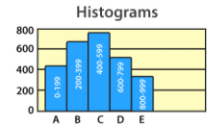
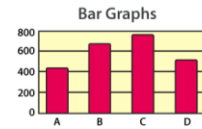
LMS Pre-Class'ta bu dersle ilgili kısma çalıştım



Data Visualization - Graphical Represent

• Graphical Representation of Data

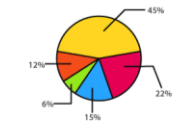
- Center
- Spread
- Shape
- Unusual Features



Frequency Table

Range (Years)	Tally	Frequency
1-15		18
16-30		11
31-45		8
46-60		4
61-75		1

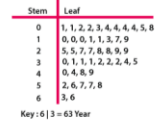
Circle Graph



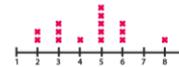
Line Graphs



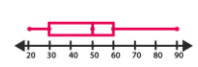
Stem and Leaf Plot



Line Plot



Box and Whisker Plot



Data Patterns

• Data Patterns

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

DATA Patterns

Center

The center of a distribution, graphically, is located at the median of the distribution

Spread

If the set of observations spreads over a wide range, the spread is greater.

Shape

The shape of a distribution can be described using some characteristics such as symmetry

Unusual Features

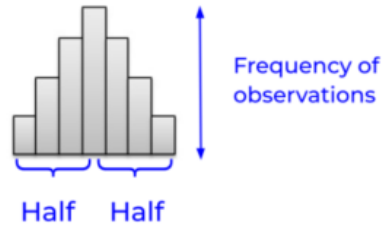
Common unusual features of data patterns are gaps and outliers.



Graphical Representation of Data

Center

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



Center

Spread

Shape

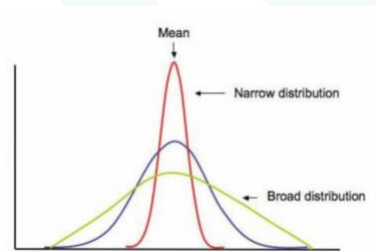
Unusual Fea.



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>



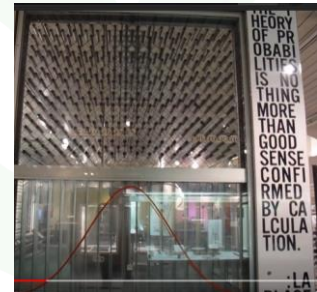
Video-2

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

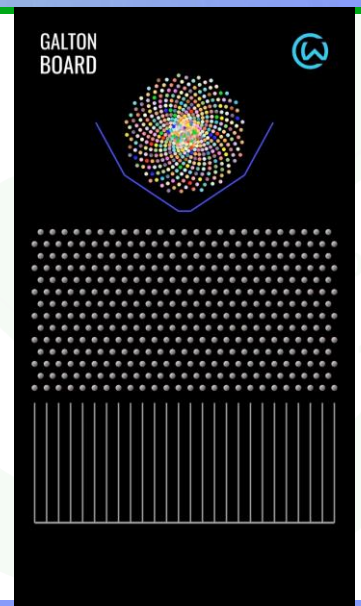
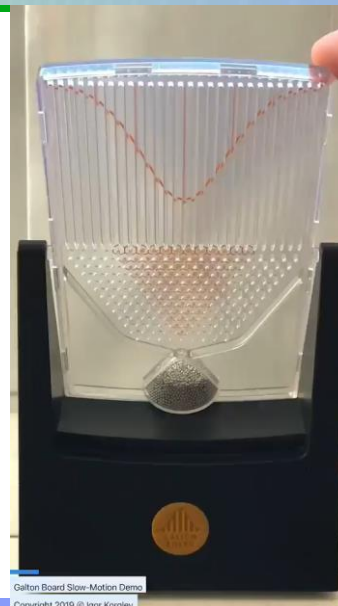
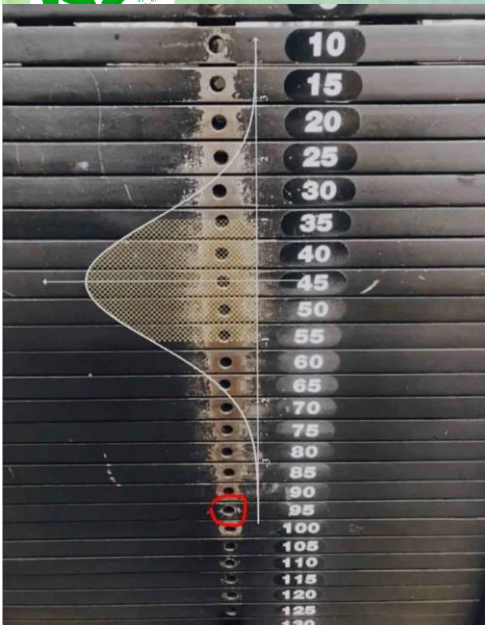


Video-3

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



Normal distribution is everywhere..



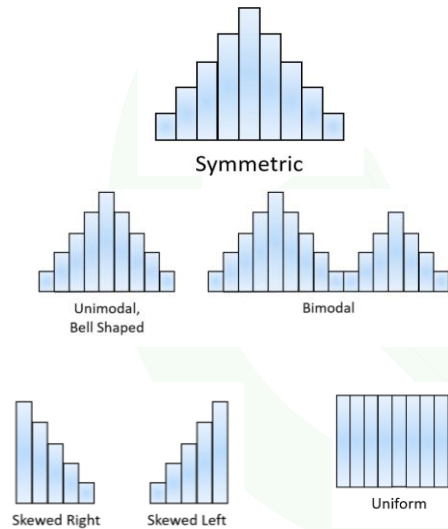


Graphical Representation of Data

Shape

Bir dağılımın şekli aşağıdaki özellikler kullanılarak tanımlanabilir.

- Symmetric
- Number of Peaks
- Skewness
- Uniform



Center

Spread

Shape

Unusual Fea.



Probability distributions

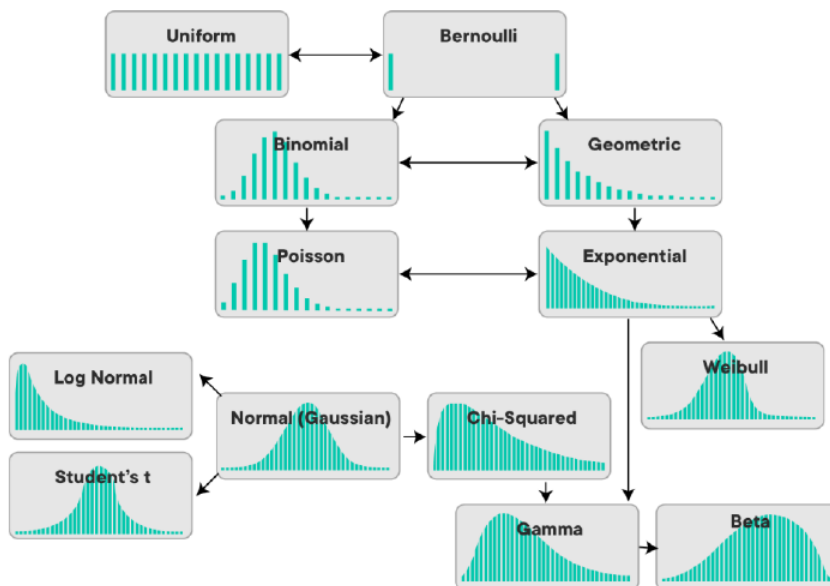
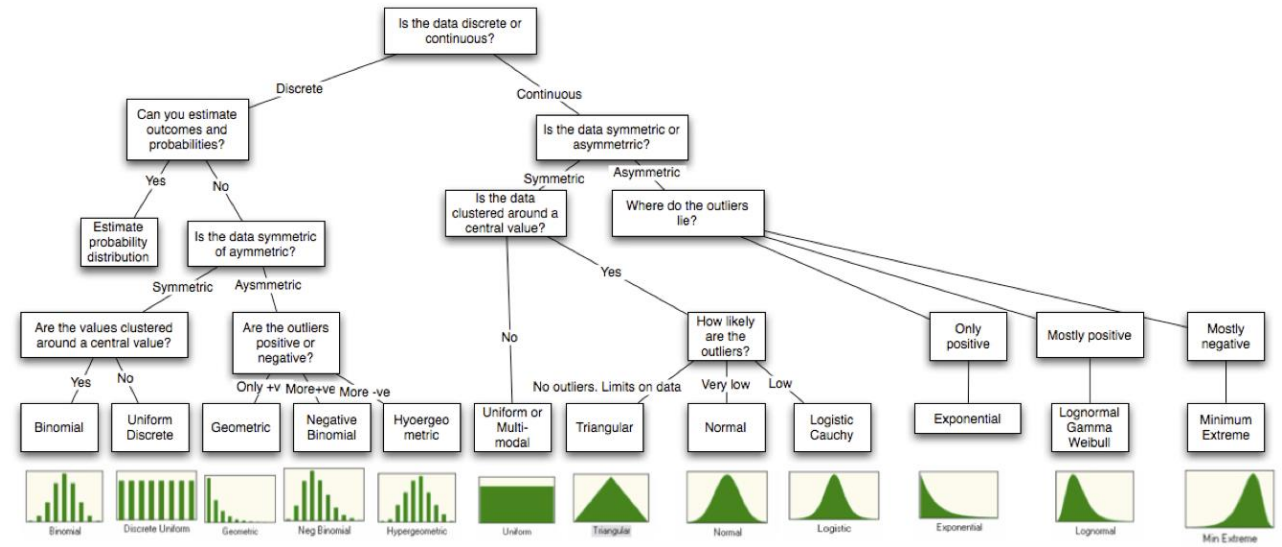




Figure 6A.15: Distributional Choices

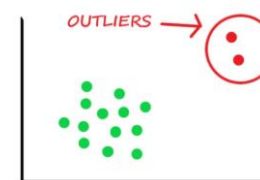
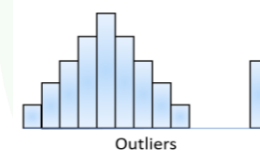
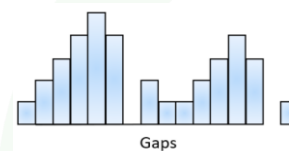


Graphical Representation of Data

Unusual Features

Veri modellerinin ortak
olağandışı özellikleri,
boşluklar ve aykırı değerlerdir

- Gaps
- Outliers



Center

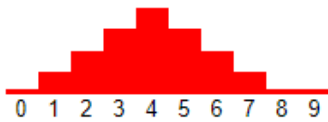
Spread

Shape

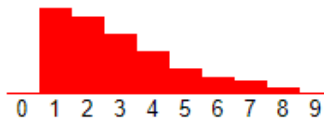
Unusual
Fea.



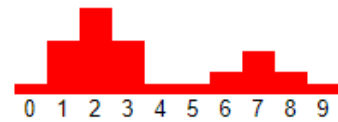
Data Patterns



Symmetric, unimodal,
bell-shaped



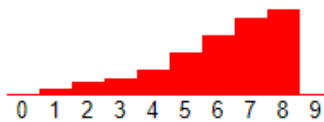
Skewed right



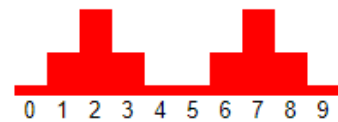
Non-symmetric, bimodal



Uniform



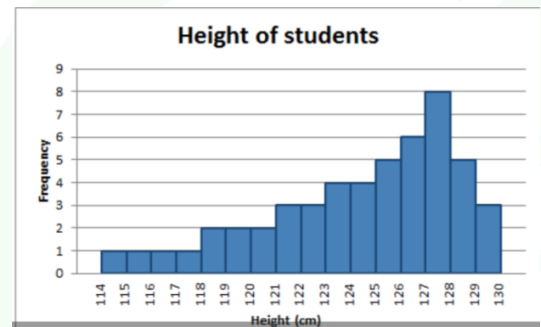
Skewed left



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 ?

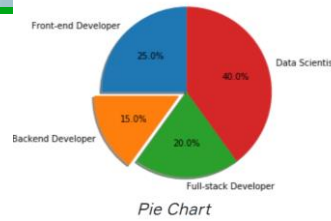
- Anlaşılabilirlik artırılır.
- Dikkat çekilecek hususlar belirtilir.
- Dağılımın biçimi hakkında bilgi sağlanır.
- Tahmin kolaylaşır





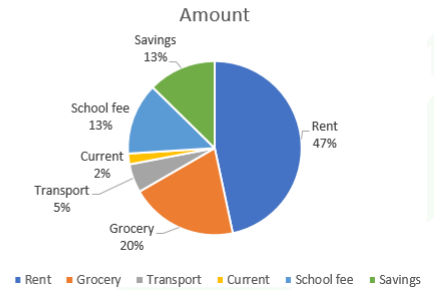
Pie Charts

- 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		

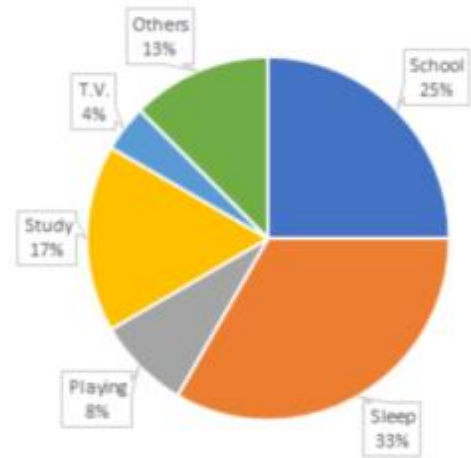


Pie Chart

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	$(\frac{3}{24} \times 360)^\circ = 45^\circ$





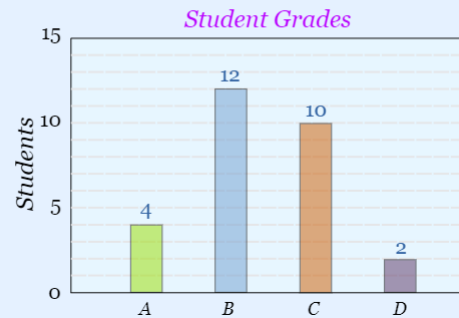
Bar Charts



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

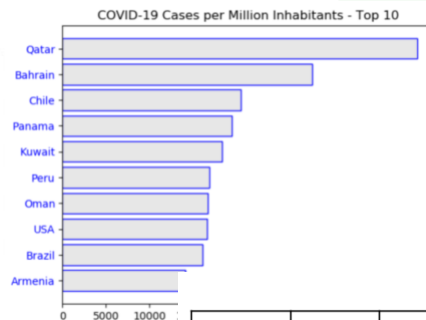
Grade:	A	B	C	D
Students:	4	12	10	2

bar graph:

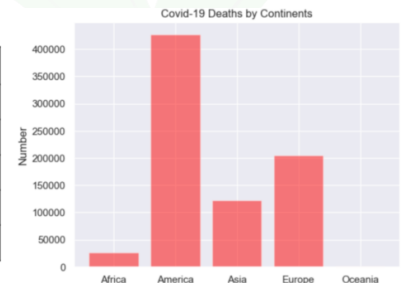


Bar Charts

countriesAndTerritories	cases	deaths	popData2019	casesPer1M
Qatar	115661	193	2832071.0	40839.724710
Bahrain	47185	175	1641164.0	28750.935312
Chile	388655	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

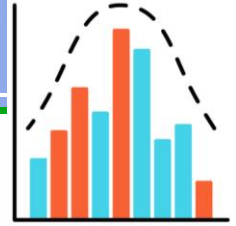


continent	cases	deaths
Africa	1119579	26260
America	11698368	427207
Asia	5606210	122034
Europe	3239237	205144
Oceania	25742	471

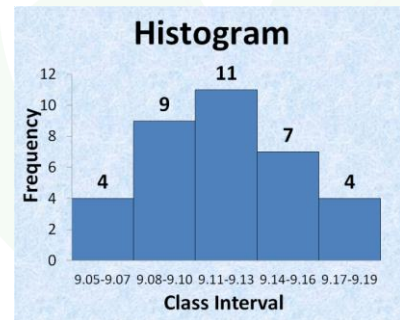




Histogram

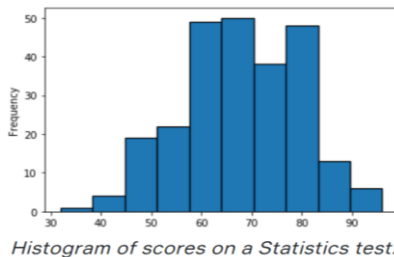


- Interval / Ratio değişkenlerle kullanılır
- Bir değişken için her bir niteliğin frekansını temsil eder
- Datanızın dağılımına iyi bir kuşbakışı bakma imkanı verir



Histogram

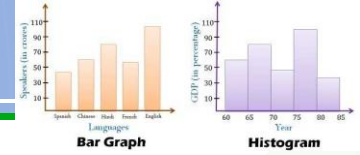
- Örnek bir Histogram çizim aşamaları



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

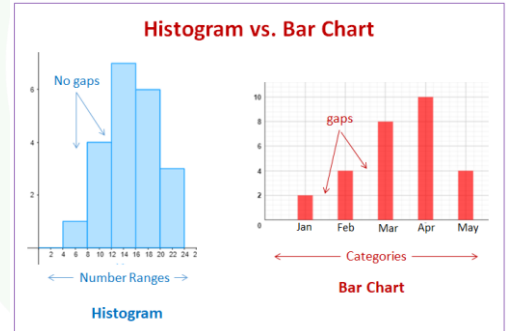


Bar Chart

- 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



BATCH : **B150** Data Science
LESSON : **STATISTICS-1**
DATE : 17.04.2023
SUBJECT : **Central Tendency**





RECAP

**Herkes önceki dersten hatırladığı
1 cümle yazabilir mi?**



**LMS Pre-Class'ta bu dersle
ilgili kısma çalıştım**



Populations & Samples

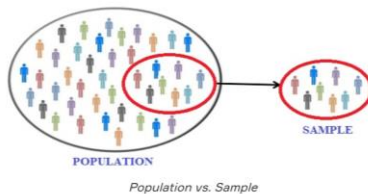


Populations & Samples

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

• Popölasyon

All people living in the
USA



• Sample

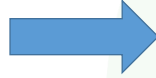


!! Sample'ları gözlemliyoruz ama popölasyonlarla ilgileniyoruz



Parameters & Statistics

Population Attributes



Parameters

Sample Attributes



Statistics

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



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

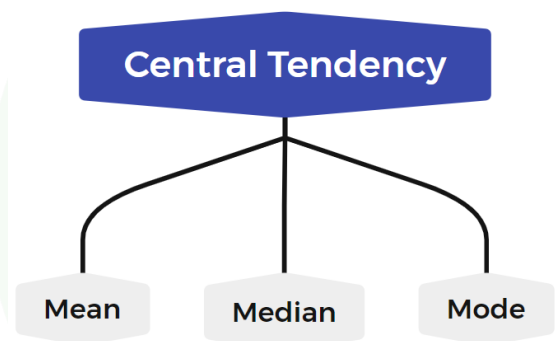
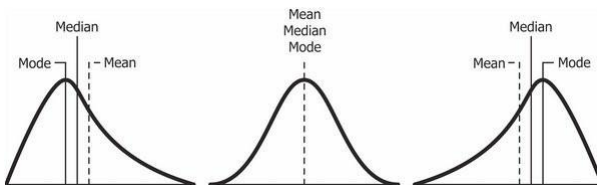


Central Tendency

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: $\mu = \frac{\sum x}{N}$ Örneklem Ortalaması: $\bar{x} = \frac{\sum x}{n}$

↑
"mü"
↑
"x-bar"



Mean Example

Örnek:

Aşağıdakiler küçük bir şirketin yedi çalışanının 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 years

Çalışanların yaş ortalaması 49'dur.

• Örnek-1

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

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

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

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

$$\bar{x} = 16.095$$

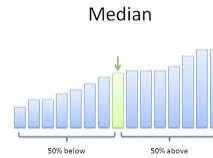
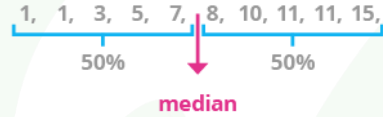
• Örnek-2



Median

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

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



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



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-1

• Örnek-2



\$4000



\$15.000



\$20.000



\$33.000



\$1.800.000

Mean:

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

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

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

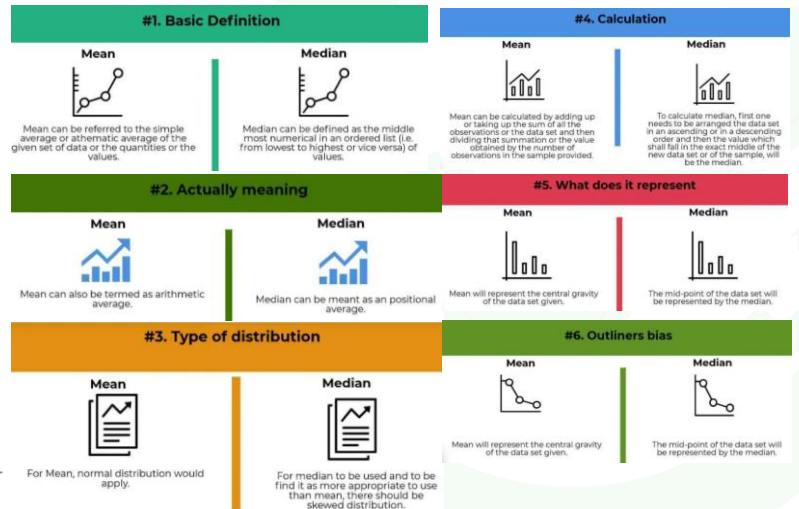
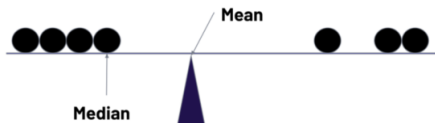
Median:

\$20000



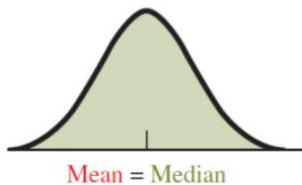
Mean vs. Median

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

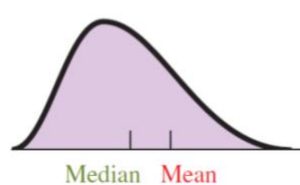


Mean vs Median

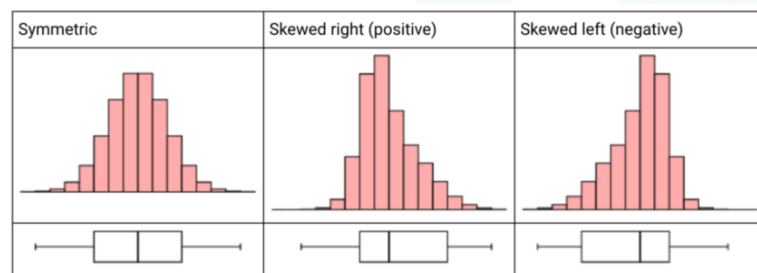
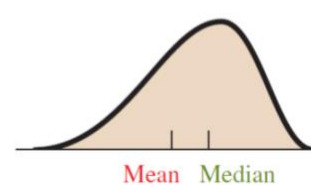
Symmetric Distribution



Right-Skewed Distribution



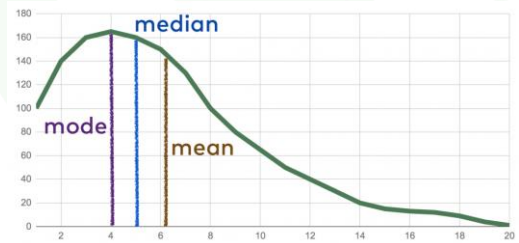
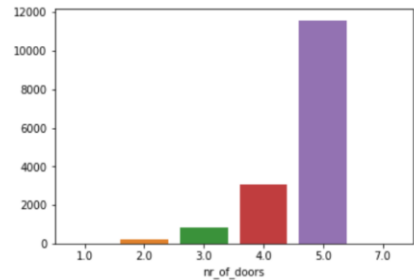
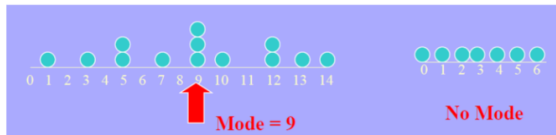
Left-Skewed Distribution





Mode

- Mode tepe değeri diye adlandırılır
- Mode: Data setinde nn fazla karşılaşılan, en popüler değer
- hem numeric hem kategorik değişkenler için kullanılabilir
- 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.

• Örnek-2

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 merkezi 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 4:

Aşağıdaki verilerin modunu ve medianını belirleyiniz.

120 100 130 100 160 130 86 100 94 90

Cözüm 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

**Mode, Median, Mean, Range,
and Standard Deviation**

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



IN THIS VIDEO
WE WILL BE LOOKING AT

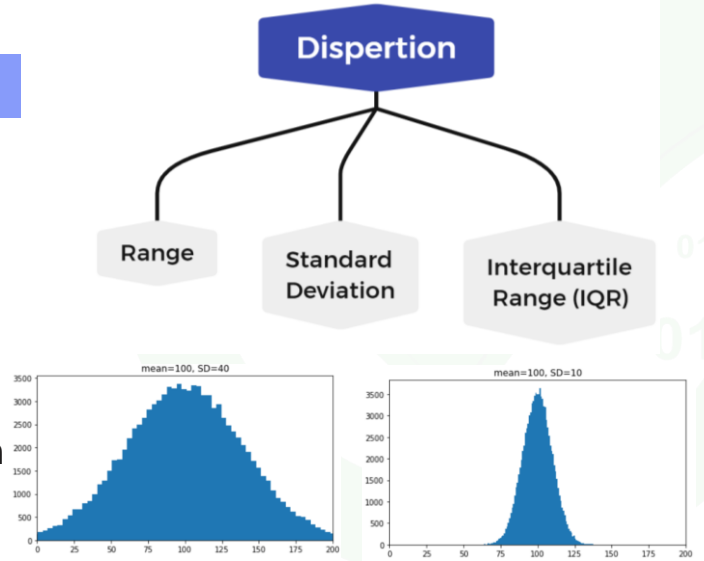
MODE	MEDIAN	MEAN
RANGE	STANDARD DEVIATION	



Dispersion (Measure of Spread)

Dağılım Ölçüleri

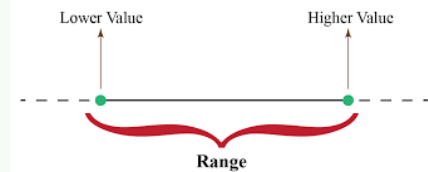
- merkezi eğilim) ölçüleri tek başına dağılımı karakterize etmez
- İki veri grubu ortalamasının eşit olması dağılımlarının aynı olmasını gerektirmez
- bir dağılım, merkezi eğilimin yaptığından daha fazlasını açıklar



Range

Aralık-Açıklık – Değişim Genişliği

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



2 4 9 5 7 3

Range = Largest - Smallest = 9 - 2 = 7

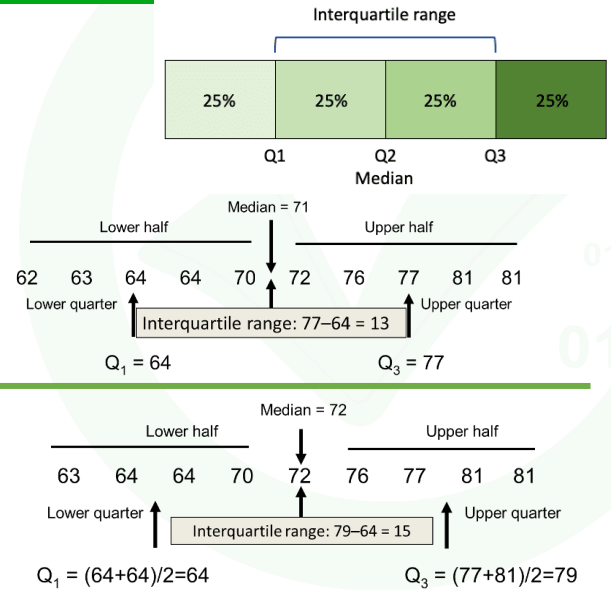


Inter Quartile Range (IQR)

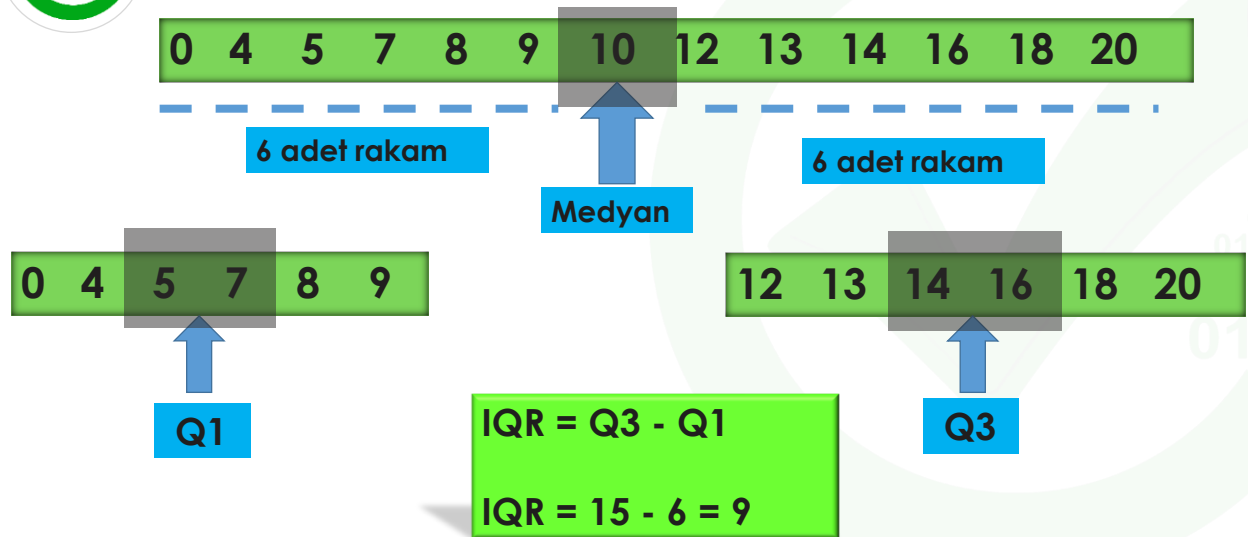
IQR

- 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.

$$IQR = Q3 - Q1$$



IQR Example





IQR Example - 2



$Q2 = 4.5$



$Q1 = 2$

$Q2 = 4.5$

$Q3 = 7$

Interquartile Range = $7 - 2$
 $IQR = 5$



$Q1 = 2$

$Q2 = 4.5$

$Q3 = 7$

Interquartile Range = $7 - 2$
 $IQR = 5$



QUESTION

What is the

- mean
- Q1
- Q3
- Median
- IQR

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



IQR

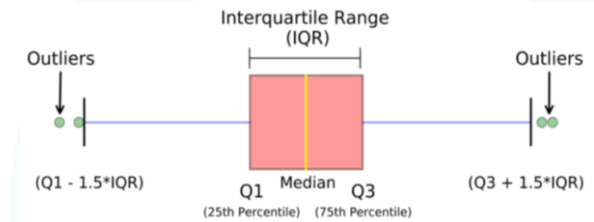
Outlier Nasıl bulunur

- Outlier, Q1'in altında veya Q3'ün üzerinde 1.5 IQR' den fazla olan veri noktalarıdır

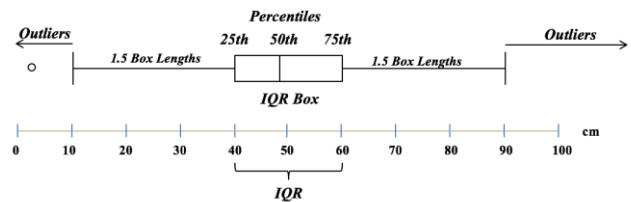
- list = [1, 5, 8, 10, 12, 15, 40]

- $Q1 - (1.5 * IQR) = 5 - 15 = -10$

- $Q3 + (1.5 * IQR) = 15 + 15 = 30$



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 i th 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 i th element
 μ = population mean
 N = population size

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

observation mean

number of observations

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

element mean

number of elements



Variance Example

- Altıdaki 4 değeri 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.

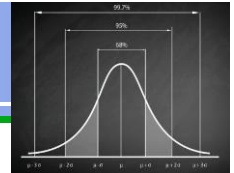


$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

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

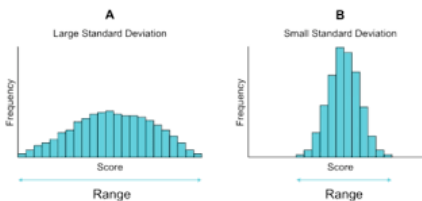


Standard Deviation



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

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

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

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

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

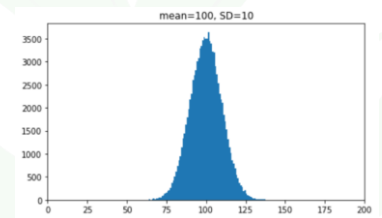
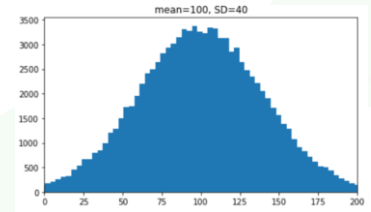
$$\sigma = \sqrt{\frac{(-13)^2 + (-12)^2 + (-11)^2 + (-10)^2 + (-7)^2 + (-4)^2 + (-4)^2 + (-4)^2 + (65)^2}{9}}$$

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

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

$$\sigma = \sqrt{539,55}$$

$$\sigma = 23,22833518$$

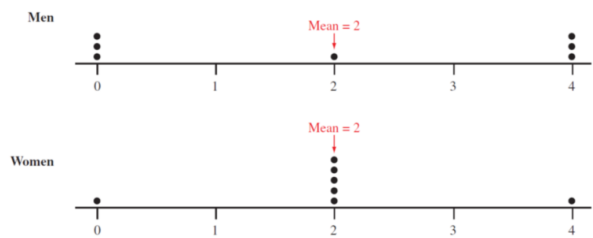


Std. Dev. Example - 2

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
```



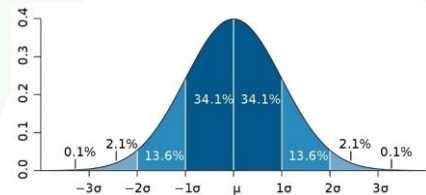
Empirical Rule

3 Sigma Kuralı

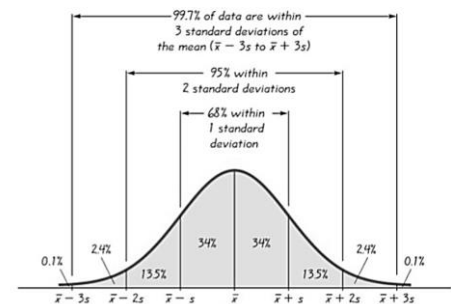
- Three Sigma Rule veya 68-95-99.7 kuralı diye de bilinir. .

- **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)

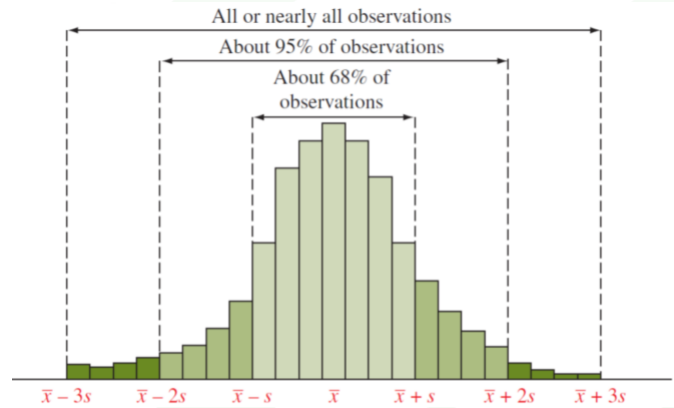
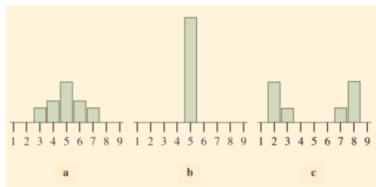
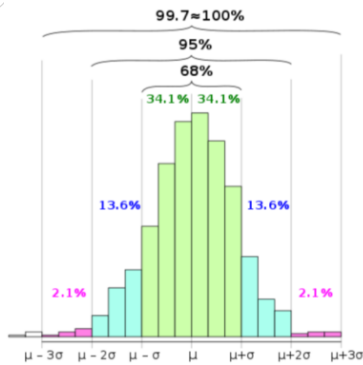


The Empirical Rule





Empirical Rule



Peardeck Interaction



Bugünkü dersi tamamen anladım

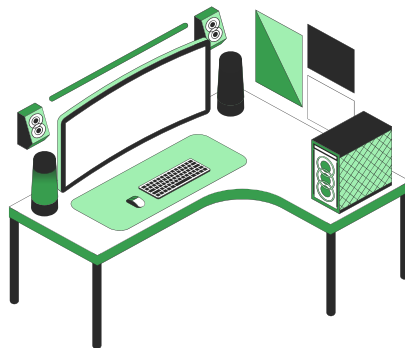
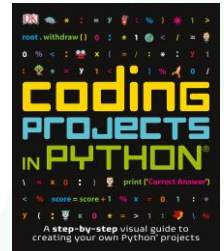




Statistics Practice-1

Python Notebook zamanı

• It is time to CODING



Do you
have any
questions?

Send it to us! We hope you learned something new.