










**BATCH** : **B150** Data Science  
**LESSON** : **STATISTICS-2**  
**DATE** : 17.06.2023  
**SUBJECT** : **Linear Regression**  
 **$R^2$**

 techproeducation  
 techproeducation  
 techproeducation  
 techproeducation  
 techproedu

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



# STATISTICS - 2

Data Science Program  
**Statistics Sessions -4**



## Interaction



**Statistics-1 bölümünde  
anlamadığım konu yoktur**



## RECAP

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





## Session - 4 Content

### Content

- Linear Regression
- Regression Equation
- Coefficient of Determination



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

# Recap – Previous Lesson

## Scatter Plot

- İki değişkenli bir scatter plot, Y eksenindeki bir değişkenin değerlerini ve X eksenindeki diğer değişkenin değerlerini gösterir.
- değişkenler arasındaki ilişkinin yönünü ve büyüklüğünü gösterir.

## Pattern of Data in Scatterplot

• Meli

described in terms of direction, outliers, linearity, and strength (DOLS)

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

• *linear, fairly strong, no outliers* Direction is *neutral*

## Box Plot

- Ortadaki quartiles denem kısmının tamamının %50'si
- Q3+1.5IQR sonrası outlier, Q3+3IQR extreme outlier.

## Covariance

**Cov (x,y) > 0**

- ilişki pozitifdir.
- X artarken Y de artar



**Cov (x,y) < 0**

- ilişki negatiftir.
- X artarken Y azalır

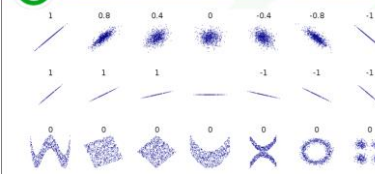


**Cov (x,y) = 0**

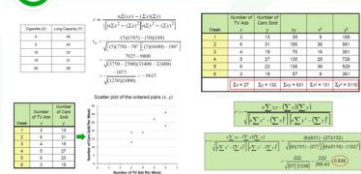
- İki değişkenin arasında ilişki yoktur, birbirinden bağımsızdır.



## Correlation



## Correlation - r Calculation

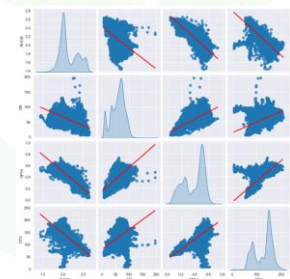
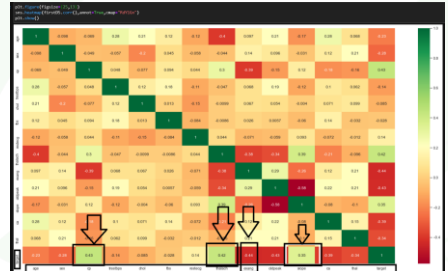
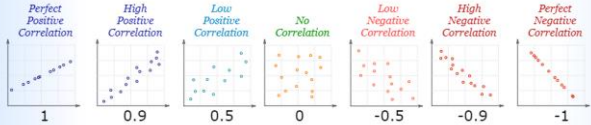


# Review

## Konular

- Correlation
- Pearson katsayısı
- Sample ve Population corr.
- R hesaplanması

A correlation is assumed to be **linear** (following a line).





## Linear Regression

### Lineer Regresyon

- Amaç: İki değişken arasındaki ilişkiye dayanarak ileri dönük tahmin yapmak
- Sebep-sonuç ilişkisi içinde, Independent variable (bağımsız değişken) sebep, bağımlı değişken ise sonuçtur.

X Değişkeni (Sebep - Input)  
(Independent Variable)



Y Değişkeni (Sonuç - Output)  
(Dependent Variable)



## Linear Regression

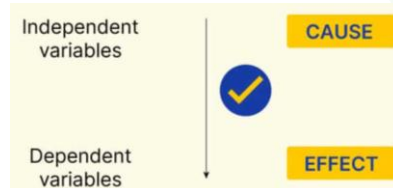
### Independent Variable

- Gelir
- Araç sahipliliği
- IQ değeri
- ???



### Dependent Variable

- Yaşam konforu
- Trafik hacmi
- İş performansı
- ???





## RECAP

Aklınıza gelen Linear regreesyon örneklerini yazar mısınız?



## Matching on Peardeck



Daily temperature

Annual Salary

Number of exams passed

Life Expectancy

Amount of time spent studying

Electricity Consumption

GDP per Capita

Number of vacations taken



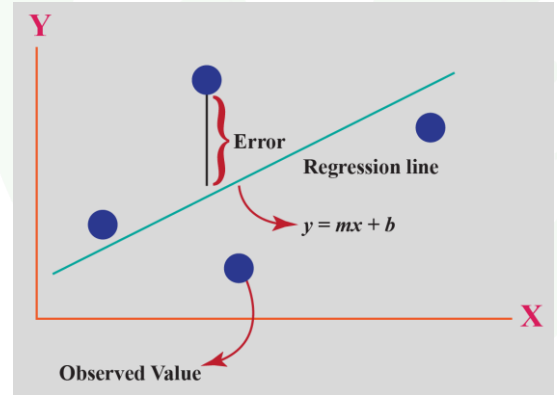


## Linear Regression and Equation

### En küçük Kareler Yöntemi

- The least squares (en küçük kareler) yöntemi
- X bağımsız değişkenin değerine bağlı olarak, Y bağımlı değişkenin değerini tahmin etmek için kullanılan bir yöntem

$$Y = aX + b$$



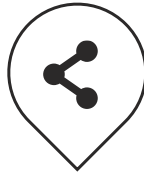
## Linear Regression Requirement

### Değişken Sayısı



1 Bağımlı değişken  
1 Bağımsız değişken

### Lineerlik

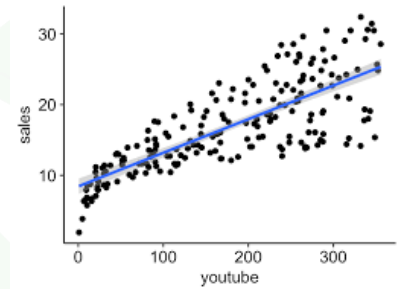


Lineer ilişki olmalı,  
nonlineer vb. değil

### Ölçülebilirlik



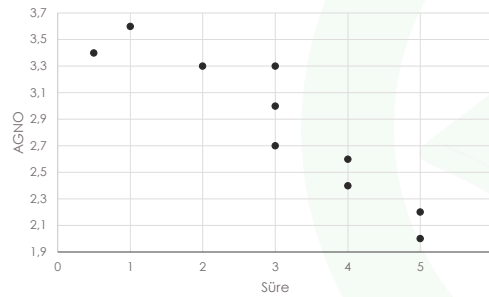
Interval veya  
ratio scale





## Linear Regression Sample

Herhangi Ekranda geçirilen süre -X	Ağırlıklı Genel Not Ortalaması -Y
3	2,7
5	2,2
2	3,3
0,5	3,4
5	2
3	3
1	3,6
4	2,4
3	3,3
4	2,6
3,05 (ort)	2,85 (ort)



intercept

$$Y = a + bX$$

slope



## Linear Regression Sample

$X - X_{ort}$	$(X - X_{ort})^2$	$Y - Y_{ort}$	$(Y - Y_{ort})^2$	$(X - X_{ort}) * (Y - Y_{ort})$
-0,05	0,0025	-0,15	0,0225	0,0075
1,95	3,8025	-0,65	0,4225	-1,2675
-1,05	1,1025	0,45	0,2025	-0,4725
-2,55	6,5025	0,55	0,3025	-1,4025
1,95	3,8025	-0,85	0,7225	-1,6575
-0,05	0,0025	0,15	0,0225	-0,0075
-2,05	4,2025	0,75	0,5625	-1,5375
0,95	0,9025	-0,45	0,2025	-0,4275
-0,05	0,0025	0,45	0,2025	-0,0225
0,95	0,9025	-0,25	0,0625	-0,2375
	21,225		2,725	-7,025 Toplam
	SSx		SSy	SP

$$b = SP / SSx = -7,025 / 21,225 = -0,3310$$

$$a = y_{ort} - b * x_{ort} = 2,85 - (-0,3310) * 3,05 = 3,85$$

$$Y = a + bX$$

$$Y = 3,85 - 0,331X$$

intercept

slope

**SSx:** Sum of Square for independent variable  
**SSy:** Sum of Square for dependent variable  
**SP :** Sum of products



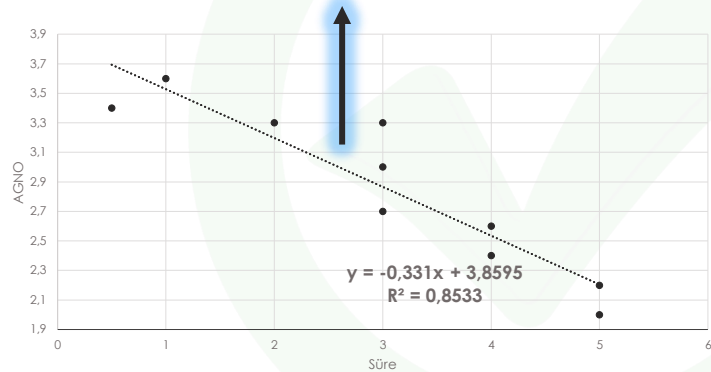


## Linear Regression Sample

En iyi eğilim çizgisi (Line of Best Fit)

$$Y = 3,85 - 0,331X$$

intercept      slope



Regresyon çizgisi, gerçek değerler ve tahmin edilen değerler arasındaki 'sum of square farklarını' minimize eder



## Linear Regression Sample - Python

# Linear Regression Sample



```
In [2]: import numpy as np
        from scipy import stats

In [3]: Ekran_sure = np.array([3,5,2,0.5,5,3,1,4,3,4])

In [4]: AGNO = np.array([2.7,2.2,3.3,3.4,2,3,3.6,2.4,3.3,2.6])

In [5]: reg = stats.linregress(Ekran_sure, AGNO)

In [7]: print("a: ", reg.intercept)
        print("b: ", reg.slope)

a: 3.859481743227327
b: -0.330977620730271
```

$$Y = a + bX$$

$$Y = 3,85 - 0,331X$$

intercept

slope



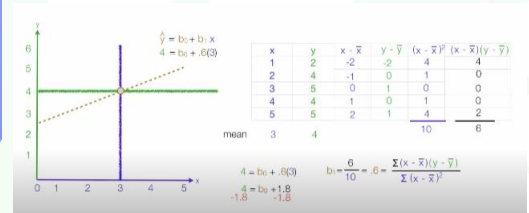
Google  
Sheets





## YOUTUBE VIDEO ONERI

- How to calculate linear regression using least square method



<https://www.youtube.com/watch?v=JvS2triCgOY>



## Pearson's r Calculation

$X - X_{ort}$	$(X - X_{ort})^2$	$Y - Y_{ort}$	$(Y - Y_{ort})^2$	$(X - X_{ort}) * (Y - Y_{ort})$	
-0,05	0,0025	-0,15	0,0225	0,0075	
1,95	3,8025	-0,65	0,4225	-1,2675	
-1,05	1,1025	0,45	0,2025	-0,4725	
-2,55	6,5025	0,55	0,3025	-1,4025	
1,95	3,8025	-0,85	0,7225	-1,6575	
-0,05	0,0025	0,15	0,0225	-0,0075	
-2,05	4,2025	0,75	0,5625	-1,5375	
0,95	0,9025	-0,45	0,2025	-0,4275	
-0,05	0,0025	0,45	0,2025	-0,0225	
0,95	0,9025	-0,25	0,0625	-0,2375	
	<b>21,225</b>		<b>2,725</b>	<b>-7,025</b>	Toplam
	<b>SSx</b>		<b>SSy</b>	<b>SP</b>	

Formula of Pearson's Correlation Coefficient

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

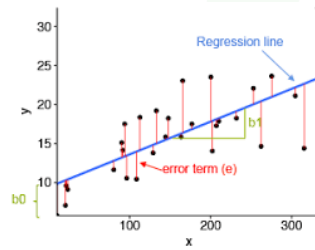
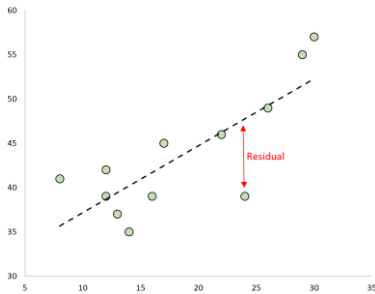
$$r = \frac{SP}{\sqrt{SS_x SS_y}}$$

$$r = -0,92$$



## Residual term (e)

**Residual = Observed value – Predicted value**



### Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Dependent Variable  $Y_i$  = Population Y intercept  $\beta_0$  + Population Slope Coefficient  $\beta_1$  × Independent Variable  $X_i$  + Random Error term  $\epsilon_i$   
 Linear component      Random Error component



## Coefficient of Determination – $R^2$

*Determinasyon – Belirlilik Katsayısı*



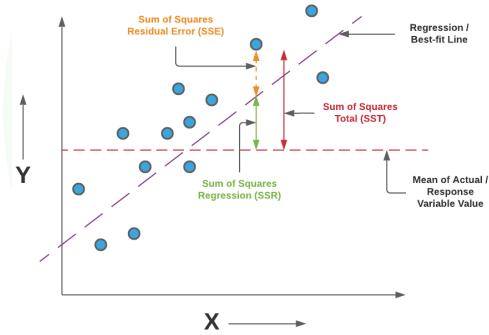
## Coefficient of Determination – $R^2$

### $R^2$ R-square

- Analizimizde iki değişken arasındaki ilişki hakkında fikir sunar
- $R^2$  değeri bize bağımlı değişkendeki toplam varyansın yüzde kaçının bağımsız değişken tarafından açıklandığını söyler.
- $R^2$  0 -1 arasında değişir

$$r = -0,92$$

$$R^2 = 0,85$$



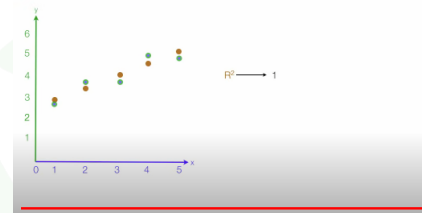
$$R^2 = \frac{SSR}{SST} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$



### YOUTUBE VIDEO ONERI

<https://www.youtube.com/watch?v=w2FKXOa0HGA>

- How to Calculate R Squared Using Regression Analysis





## Kahoot Uygulaması



## Python Calculation

- It is time to code by Python...



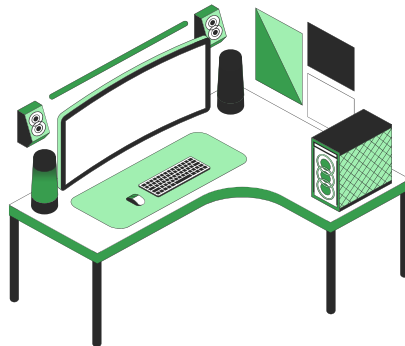


## Final Question

**Bu dersi tamamen anladım..**



**FINISH**



Do you  
have any  
questions?

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