

**DT/NT : DATA SCIENCE  
LESSON : STATISTICS -1  
SUBJECT: INTRODUCTION**

**BATCH: 223**

**10.11.2023**



**TECHPRO  
EDUCATION**



techproeducation.com

+1 (585) 304 29 59

## **STATISTICS - 1**

Data Science Program

***Session -1***

## Peardeck Time !

**İstatistik konularında daha önce çalıştım ve iyi durumdayım.. ?**

**Interaction**

**İstatistik benim korkulu rüyam !!**



**Part-1**

- The meaning of Statistics
  - Eisenhower Matrix
  - Characterization
    - Collection
    - Analyzing
    - Visualization
    - Inference
    - Presentation
  - Why Statistics
  - Importance of Statistics
    - Data Science vs Statistics
  - How much Statistics for us
  - Statistic Types
    - Descriptive
    - Inferential
  - Data Types
  - Parameters and Statistics
  - Probability vs Statistics
  - Level of Measurement
    - Nominal
    - Ordinal
    - Interval
    - Ordinal

**Part-2**

- Data Visualization - Graphical Representations
  - Patterns
  - Center
    - Spread
    - Shape
      - Symmetric
      - Number of peaks
      - Skewness
      - Uniform
    - Unusual Features
      - Gaps
      - Outliers
  - Frequency Table
    - Relative Frequency
    - Cumulative Frequency
  - Bar Chart
  - Pie Chart
  - Histogram
- Populations & Samples
  - Parameters & Statistics
- Central Tendency (Measure of Centre)
  - Mean
  - Median
  - Mode
- Dispersion (Measure of Spread)
  - Range
  - IQR
  - Standard Deviation
    - Empirical Rule
  - Variation

**Part-3**

- Scatter Plot
  - Linearity
  - Slope
  - Strength
  - Unusual Features
    - Clusters
    - Gaps
    - Outliers
- Box Plot
  - Min & Max Values
  - $1.5 \times \text{IQR}$  (John Tukey)
- Covariance
- Correlation
  - Pearson Correlation Coefficient
  - Correlation – Linear Relationship

**Part-4**

- Linear Regression
  - Dependent & Independent Variables
  - Regression Equation
    - Pearson's r Calculation
    - Residual term ( $e$ )
  - Coefficient of Determination –  $R^2$

**Part-5**

- Probability
  - Law of Large Estimates
  - Sample Space – Event
  - Independent – Dependent Event
  - Probability of 2 Independent Event
  - Intersection, Union, Complement
  - Permutation
  - Combination
  - Conditional Probability
    - Independence check
- Bayes Theorem

**Part-6**

- Random Variables
- Probability Distributions
  - Discrete Probability Distributions
    - Probability Mass Function (PMF)
    - Cumulative Distribution Function (CDF)
    - Binomial Distribution
    - Bernoulli Distribution
    - Poisson Distribution
    - Geometric, Hypergeometric, Negative Binomial
  - Continuous Probability Distributions
    - Uniform Distribution
    - Normal Distribution
      - Z Table
    - Standard Distribution
    - T Distribution (aka, Student's T-test, T-Distribution)
    - Exponential, Gamma, Chi-square, F

**Part-7**

- Sample Distribution
  - Simple Random Sampling (SRS)
  - Standard Error of the Mean
  - Central Limit Theorem
    - Normal Distribution Advantages
  - Confidence Interval

**Part-8**

- Hypothesis (Significance) Test
  - Hypothesis Test Steps
    - Assumptions
    - Hypotheses
      - Null Hypotheses
      - Alternative Hypotheses
    - Test Statistic
    - P – Value
    - Conclusions
  - Significance Level ( $\alpha$  – alpha)
  - Type I – II Error
  - One – Two Tail Tests
    - Left Tail Test
    - Right Tail Test
    - Two Side (Two Tail) Test
  - T Test
  - Z Test

**Part-9**

- Independent Samples T Test
- Dependent T test – (Paired T Test)
- One Way ANOVA
  - Test Statistics – ANOVA Table
    - SSR: Regression Sum of Squares
    - SSE: Error Sum of Squares
    - SST: Total Sum of Squares
  - MSR: Regression Mean Squares
  - MSE: Mean Squares Error
  - F Statistic
- Categorical Data Analysis
- Chi-Square Test

## Fundamentals Statistics

### Content and References

- Fundamentals Statistics
- Probability
- Distributions
- Central Limit Theorem
- Hypothesis Testing

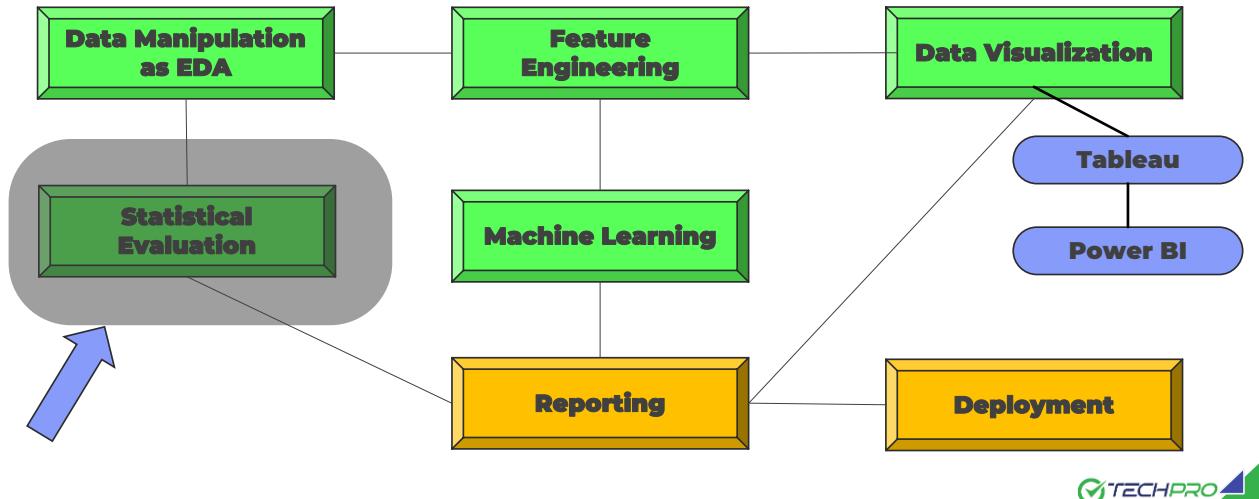
### Kaynaklar

- YouTube StatQuest
- YouTube CrashCourse
- <https://www.khanacademy.org/>
- <https://onlinestatbook.com/>
- <https://gelecegiyazanlar.turkcell.com.tr/konu/veribilimi-icin-istatistik>

### Dokumanlar

- LMS page
  - PDF Document on Internet
- Books:**
- Applied Statistics and Probability – Montgomery
  - An Introduction to Statistical Learning – James –Witten
  - Probability&Statistics - Myers

# Internship Project Steps



# Statistics in Internship Project

## Statistical Evaluation

İSTATİSTİK ANALİZLERİ		EKONOMETRİ VE ZAMAN SERİLERİ
Mod, Medyan, Ortalama, St. Sapma Analizleri		
Featurelerin Normallik Analizi (Veriler normal değilse, normalleştirme çalışmalarının yapılması)		
Aykırı Gözlemlerin belirlenmesi, gerektiğinde verilerden çıkarılmasının sağlanması		
Bağımlı Featurelerin Eğiklik ve basılık durumuna göre eğimlerinin belirlenmesi		
Farklı platformların birbirine üstünlüklerinin araştırılması Analizi (ANOVA)		
Reklam öncesi ve reklam sonrası satışların karşılaştırılması (A/B Analizleri - Student analizi)		
Korelasyon analizi	Normallik Varsayımları	
Korelasyon Analizinin Anlamlılığı testi	Duraklık Analizi	
Effect Size Analizi	Otokorelasyon Analizi	
	Çoklu Doğrusal Bağımlılık Analizi	
	Değişen Varyans analizi	
	Basit Doğrusal Regresyon Analizi	
	Çoklu Doğrusal regresyon Analizi	
	Hesaplanan denklemlerinin tahmin gücünün hesaplanması R <sup>2</sup>	
	Doğrusal Olmayan Regresyon Analizi	
	Verinin Durumuna göre ARIMA veya GARCH Modelleri ile tahmin edilmesi	
	Yapışsal Kırılma Testi (Qusum)	
	Etki ve Tepki analizi (Korelasyon analizinde belirlenmiş en yüksek ilişkili parametrelerle göre etki ve tepki analizleri yapılacaktır)	



# Statistics in Internship Project

## Statistical Evaluation

### Woocommerce Korelasyon Analizi

- Korelasyon veri kolonları arasında ilişki olup olmadığını gösteren istatistiksel bir katsayıdır.
- Katsayı 1 veya -1'e yaklaşıkça iki kolon arasındaki yerde kuvvetli ilişki vardır. Birinci yükseldikçe diğer de yükseltir veya biri azaldıkça diğer de azalır.
- Kırmızı ile işaretlenen kolonlar arasında kuvvetli/orta ilişki vardır.



### Reklamların Woocommerce Satışlarına Etkisi - İstatistiksel Test

- Woocommerce platformu için verilen reklamların 2021 yılı içinde satışa belirgin bir pozitif bir etkisi olmadığı test edilmiş ve pozitif etki olduğu sonucuna varılmıştır. 2021 Haziran'da başlayan Woocommerce reklamlarının pozitif etkisi 2022 itibarı ile başlamıştır.
- Yukarıdaki durumlar istatistiksel hipotez testi yöntemi ile test edilmiş, satışların sabit kalmasına veya artışının bilinmeyen etkenlere bağlı, tesadüfi vb olmadığı, gerçek bir artış olduğu doğrulanmıştır.



## Section-1 Content

### Content

- The meaning of Statistics**
- Why Statistics**
- Importance of Statistics**
- How much Statistics for us**
- Statistic Types**
- Level of Measurement**





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



## Eisenhower Matrix



	ACİL	ACİL DEĞİL
ÖNEMLİ	<b>YAP</b> Hemen şimdi yap.	<b>PLANLA</b> Daha sonra yapmak için bir zaman belirle.
ÖNEMLİ DEĞİL	<b>DEVRET</b> Otomasyona al ya da başka birine devret.	<b>ELE</b> Bu işi sil ve yapma.

**Prioritize your works...**





## What is Statistics

- Statistics is all about data



Collection

Characterization

Analysing

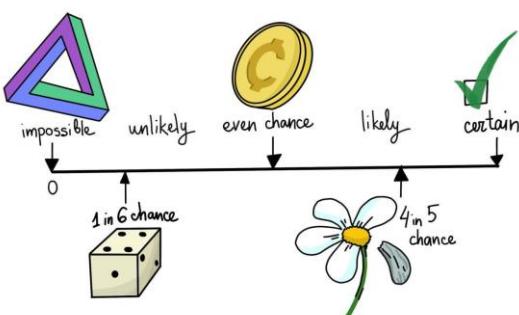
Visualisation

Inference

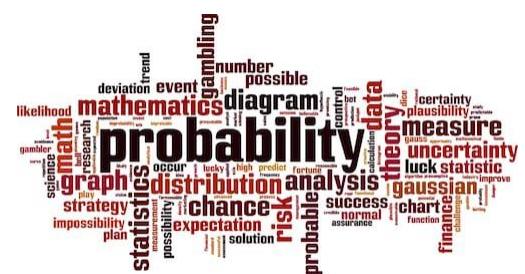
Presentation



## Science and Probability



@luminousmen.com



shutterstock.com • 1382524664





## Science and Probability



Peardeck Time

**İstatistik deyince aklıma ilk  
gelenler..**



## Statistics

- Bazı haber metinlerinden örnekler
- İstatistik tanımı
- İstatistik İngilizce kavram kullanımı
- Data Science alanında İstatistik kullanımı
- İstatistik ne kadar güvenilir bir alan



Statistics is the grammar of science.

(Karl Pearson)

### Metod Olarak İstatistik:

İstatistiğe konu olabilen olaylara ait nicel verilerin derlenmesi, işlenmesi, analizi ve yorumlanmasında kullanılan teknikler bütünüdür.



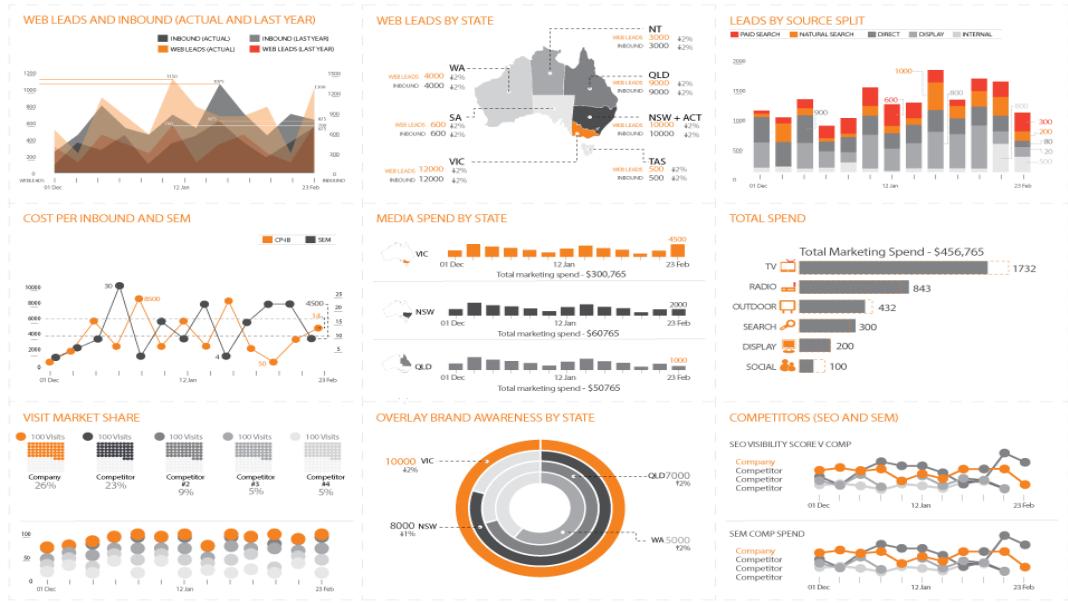
## Statistics Tanımlar

- Her türlü kuruluşça ülke geneli veya ülkenin daha küçük bir bölümü (il, ilçe, bölge vb ) hakkında toplanan verileri ifade eder. Milli eğitim istatistikleri, nüfus istatistikleri, tarım istatistikleri...
- Bilimsel araştırmalarda toplanan ve işlenen verilerin, araştırma probleminin aydınlatılması amacıyla, uygun ve etkin olarak betimlenmesi, çözümlenmesi ve yorumlanması yollarının tümüdür.
- Örneğin; belirli bir y değişkeninin x değişkenine bağlı olup olmadığı belirli testler yardımıyla incelenir. Bu testlere başvurulduğunda hesaplanan istatistik, verilen tabloda bulunan değerler ile kıyaslandıktan sonra yorum yapılır. İfadesinde olduğu gibi...
- İstatistik belli bir ana kütleden çekilen randomly (rassal) bir örneklemin “aritmetik ortalaması”, oran gibi herhangi bir özet değerine verilen addır.
- Örneğin; bir anakütleden hesaplanan aritmetik ortalama, standart sapma, oran gibi özet değerlere parametre denir. Bir parametrenin gerçek değerini hesaplamak için tam sayılmayıp olabilecegi gibi, bunu tahmin etmek için örneklemeden de yararlanılabilir. Örneklemeye başvurulduğunda örneklemden hesaplanan özet değerlere istatistik adı verilir.



## Example of Statistics as Dashboard

Geleceği dashboard u linki.



Statistics in Daily life examples?





**SCORE** 000450    **LIVES** 3    **TOP** 012000

## Statistics in Daily life examples

### ► Statistics Role In Real Life

- Government
- Emergency Preparedness
- Political Campaigns
- Sports
- Research
- Education
- Prediction
- Predicting Disease
- Insurance
- Financial Market
- Business Statistics

[https://medium.com/@john\\_marshall/statistics-role-in-real-life-a6ba727e0ad8](https://medium.com/@john_marshall/statistics-role-in-real-life-a6ba727e0ad8)

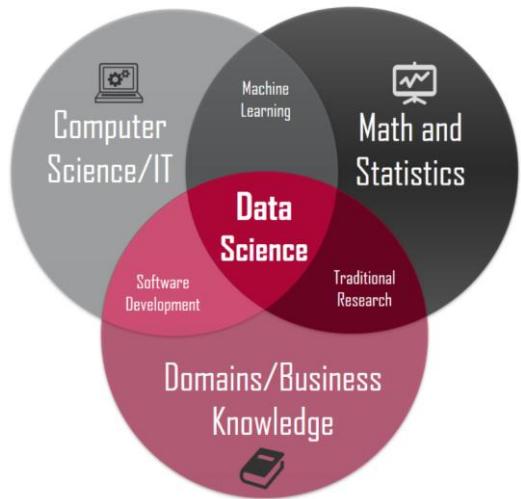
### ► Statistics In Our Day to Day Life

- Medical Study. Statistics are used behind all the medical study. ...
- Weather Forecasts.
- Quality Testing. A company makes thousands of products every day and make sure that they sold the best quality items.
- Stock Market.
- Consumer Goods.

<https://stataanalytica.com/blog/uses-of-statistics/>

## Why Statistics

- Araştırmacılara elde ettikleri bulguları açıklarken yardımcı olur.
- Araştırmacıların, verilerin ötesinde daha genel sonuçlara gitmesine yardımcı olur.
- «Bir programcıdan daha fazla istatistik, bir istatistikçiden daha fazla programlama bilen kişi» tanımı
- Ne kadar İstatistik öğreneceğiz?



## Why statistical knowledge are important in data science

Should I run for the bus?

Which stock should I buy?

Which products should I list in Website?

Should I take this medication?

Should I have my children vaccinated?

 **Josh Wills** @josh\_wills

Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

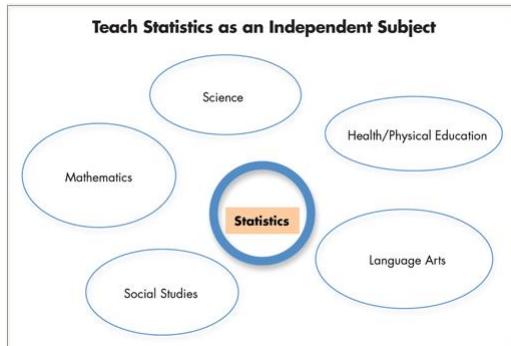
Follow 

RETWEETS 1,046 FAVORITES 532

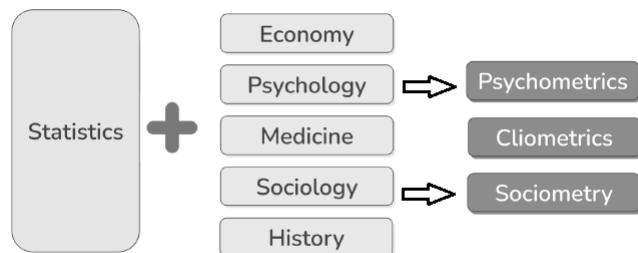
6:55 PM - 3 May 2012



# Importance of Statistics in Science



<https://www.emathzone.com/tutorials/basic-statistics/importance-of-statistics-in-different-fields.html>



## Data Science vs Statistics

### Differences

#### Similarities

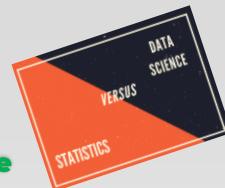
- understanding of mathematics
- investigating problems
- exploratory data analysis
- analyzing trends
- creating forecasts
- visualizations
- reporting findings to non-technical users

#### Statistics

- one-off reports
- use of SAS programming
- focus on diagnostic plots
- focus on significance testing
- focus on t-tests, ANOVA, and MANOVA, etc.
- more manual data collection (sometimes from surveys)
- usually, you will find Statisticians in the healthcare and economics field
- or more academic settings

#### Data Science

- automation
- use of SQL querying for data collection
- machine-learning libraries like sklearn and TensorFlow
- use of Python and R programming languages
- deployment of automated models (into an app)
- focus on software engineering practices



## DS vs Statistics

A data scientist makes hundreds of decisions every day.

Many of these decisions require a strong foundation in math and statistics.

Data science requires descriptive statistics and probability theory, at a minimum.

- ▶ Statistics is a mathematically-based field which seeks to collect and interpret quantitative data.
- ▶ In contrast, the problems studied by statistics are more often focused on drawing conclusions about the world at large.
- ▶ Statisticians are usually trained by math departments.

- ▶ Data science is a multidisciplinary field which uses scientific methods, processes, and systems to extract knowledge from data.
- ▶ Data science problems often relate to making predictions and optimizing search of large databases.
- ▶ Data scientists tend to come from engineering backgrounds.



## Data Flowing





## Soruları 2 cümle ile cevaplayalım

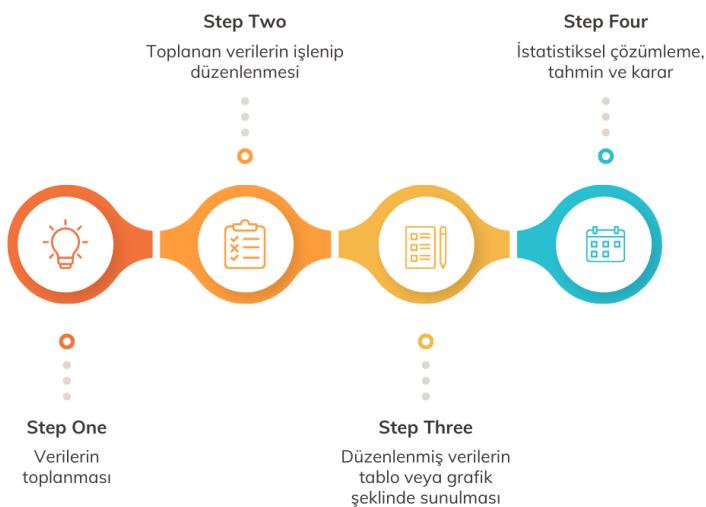
What is Data Science?

Why are we learning Statistics ?



## Statistical Order

İstatistikte izlenen sıra



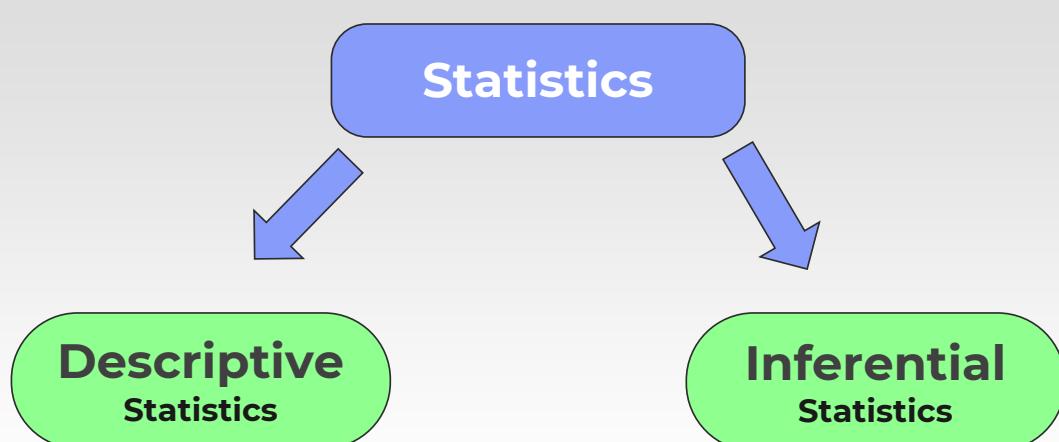
## Topics of Statistics - 1

### ▶ Tipik Olay

○ Birbirinin tam benzeri olan olaylara “**tipik olay**” denir. Olaylar birbirinin tam benzeri olunca, bunlardan sadece biri bile oluşturdukları topluluğu temsil edebilir. Bu durumda çok sayıda olayı incelemeye gerek kalmaz.

### ▶ Kolektif Olay

○ Bazı ortak yönleri olmakla birlikte, birbirine benzemeyen genelde aralarında önemli farklılıklar da bulunan olaylara ise “**kolektif (toplu) olay**” denir. Olaylar benzemediği için bunlardan sadece biri oluşturdukları topluluğu temsil edemez.



## Topics of Statistics - 2



### Descriptive Statistics

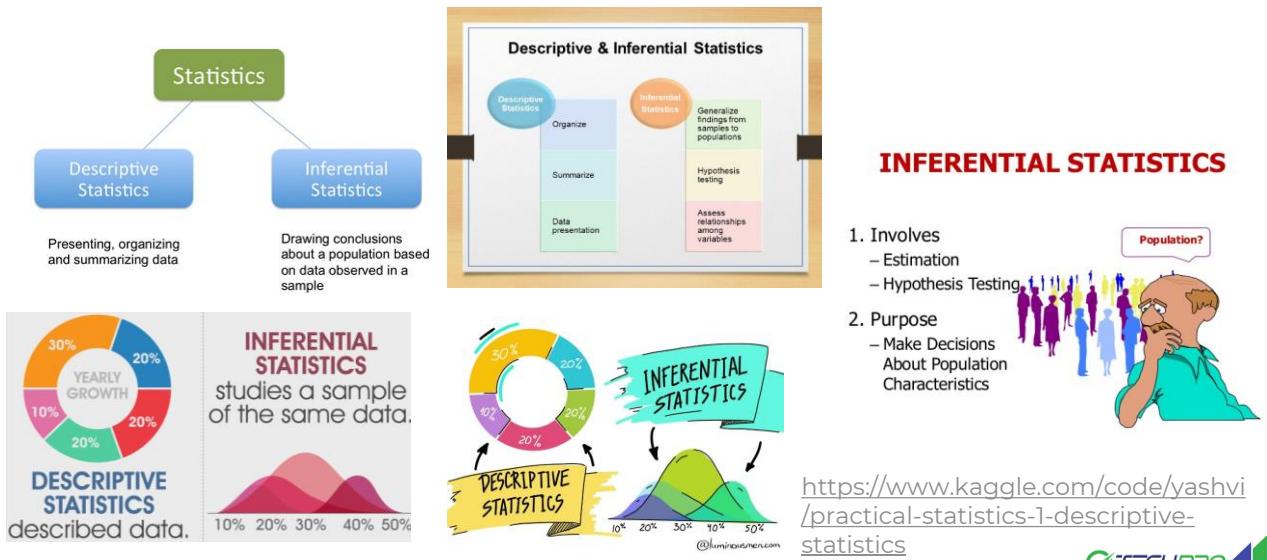
- Kolektif olaylara ilişkin verileri toplamak,
- Toplanan verileri işleyip düzenlemek,
- Düzenlenmiş verileri tablolar veya grafikler şeklinde sunmak,
- Ve nihayetinde kolektif olayların eğilimini ortaya çıkarmak...

### Inferential Statistics

- Örnekleme ve örnekleme bölünmesi kavramlarına dayanarak bir anakütleden randomly olarak Seçilen bir örneklem yardımıyla;
- Anakütle parametrelerini tahmin,
- Anakütle parametrelerine ilişkin hipotezleri test etme,
- Ve geleceğe dönük tahminler yapmak...
- Söz konusu çıkarımlar mutlak doğru olamayacağından "olasılık" kelimesi ile ifade edilir.
- Çıkarımsal istatistikte genelde tümevarım yöntemi kullanılır

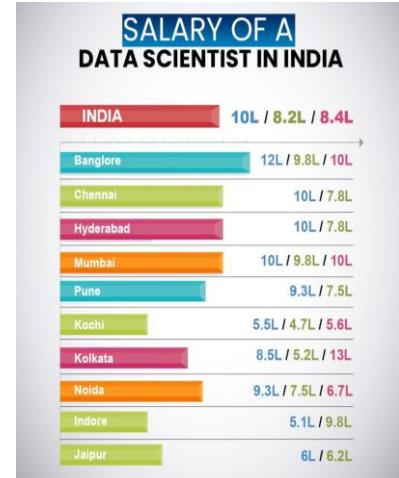
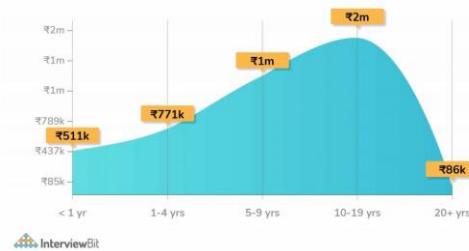


## Descriptive vs Inferential Statictics



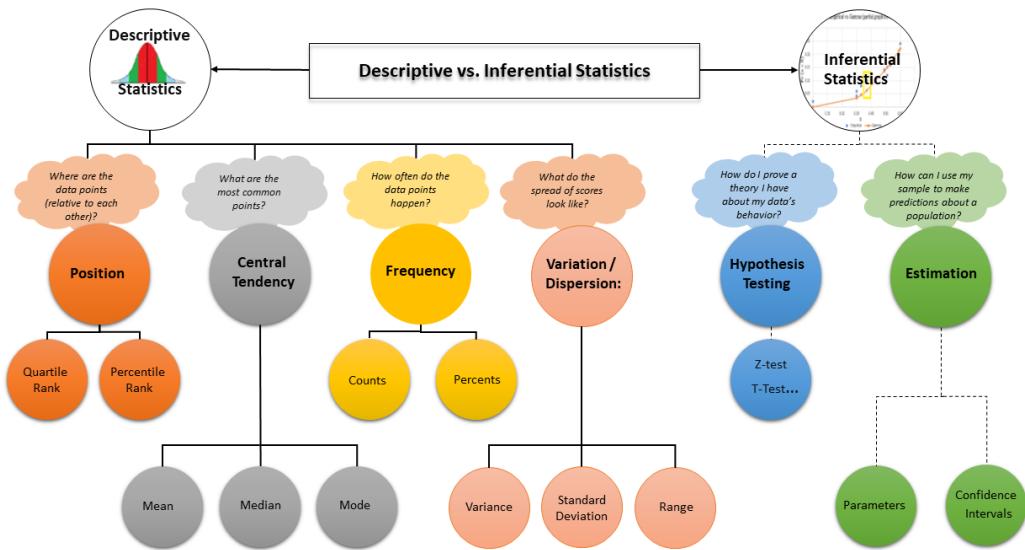
# Why do we need Inferential Statistics?

- Hindistan'da bir Data Scientist ortalama ücreti ne kadar olabilir?
- Inferential Statistics kullanımı



## Descriptive vs Inferential Statistics

<https://www.datasciencecentral.com/profiles/blogs/descriptive-vs-inferential-statistics-in-one-picture>



# Descriptive vs Inferential Statictics

## Examples of Descriptive and Inferential Statistics

### Descriptive Statistics

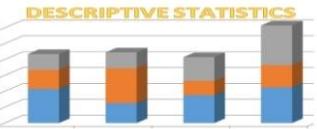
- Graphical
  - Arrange data in tables
  - Bar graphs and pie charts
- Numerical
  - Percentages
  - Averages
  - Range
- Relationships
  - Correlation coefficient
  - Regression analysis

### Inferential Statistics

- Confidence interval
- Margin of error
- Compare means of two samples
  - Pre/post scores
  - $t$  Test
- Compare means from three samples
  - Pre/post and follow-up
  - ANOVA = analysis of variance

### Descriptive Statistics

Descriptive statistics are numbers that are used to summarize and describe data. The word "data" refers to the information that has been collected from an experiment, a survey, a historical record, etc.

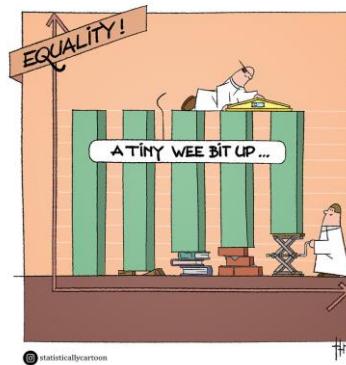
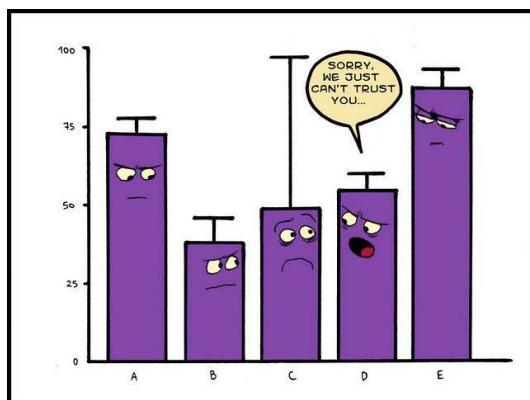


### Inferential Statistics

The sample is a set of data taken from the population to represent the population. Probability distributions, hypothesis testing, correlation testing and regression analysis all fall under the category of **inferential statistics**.



## Sometimes funny content



# Interview nedir ?



- <https://towardsdatascience.com/22-new-data-science-interview-questions-for-2022-964c687f735b>
- İlk soruların İstatistik

## **22 New Data Science Interview Questions for 2022**

- <https://towardsdatascience.com/22-new-data-science-interview-questions-for-2022-964c687f735b>
  - İlk soruların İstatistik alanında olduğuna dikkat ediniz.



## ■ Data - Veri nedir ?



Fred's Classic Movie Events									
Venue	Fred's Movie Emporium			Today			13/12/2012		
Date	Movie	Tickets Sold	% Capacity	Tickets Remaining	% Tickets to Sell	Status	Da		
Wednesday 11 May 2011	Grease	50	33%	100	67%	Promote			
Sunday 15 May 2011	Jaws	150	100%	0	0%	Last Few Seats			
Monday 23 May 2011	Citizen Kane	105	70%	45	30%	Last Few Seats			
Wednesday 01 Jun 2011	The Wizard of Oz	150	100%	0	0%	Last Few Seats			
Friday 10 Jun 2011	Singin' In The Rain	85	57%	65	43%	Promote			

## Statistical Data Analysis

- veri toplama, veri yorumlama ve veri doğrulama
- veri analizi, istatistiksel işlemleri gerçekleştiren bir prosedür
- verileri nicelleştirmeye çalışan ve istatistiksel analiz uygulayan bir nicel araştırma
- SDA da Datanın değişkenleri vardır
- Veriler tek değişkenli veya çok değişkenli
- Değişkenlerin sayısına ve türüne bağlı olarak, farklı istatistiksel teknikler
- Birden fazla veri varsa çok değişkenli veri analizi



**DT/NT : DATA SCIENCE  
LESSON : STATISTICS -1  
SUBJECT: GRAPHICAL  
REPRESENTATION**

**BATCH: 223**

**10.11.2023**



**TECHPRO  
EDUCATION**





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

## Recap – Previous Lesson

**What is Statistics**

- Statistics is all about data



Collection  
Characterization  
Analysing  
Visualisation  
Inference  
Presentation

**Why Statistics**

- Araştırmacılara elde ettikleri bulgular açıklayarak yardımcı olur.
- Araştırmacılara, verilerin ötesinde daha genel sonuçlara girmesine yardımcı olur.
- Bir programcının daha fazla istatistik, bir istatistikçinin daha fazla programlama bilen kişi tanımı
- Ne kadar İstatistik Öğreneceğiz?

Computer Science/IT  
Machine Learning  
Math and Statistics  
Software Development  
Data Science  
Traditional Research  
Domains/Business Knowledge

**Descriptive vs Inferential Statistics**

**STATISTICS**

**DESCRIPTIVE STATISTICS**

**INFERENTIAL STATISTICS**

**INFERENTIAL STATISTICS**

1. Involves:  
- Estimation  
- Hypothesis Testing

2. Purpose:  
- Make Decisions  
- About Population Characteristics

**Introduction to Data**

**Data - Veri nedir ?**

Frida's Classic Movie Events

Date	Movie	Total	% of Capacity	Tickets Remaining	Tickets Sold	Status
Wednesday, 21 May 2014	Citizen	100	100%	0	100	Sold Out
Wednesday, 21 May 2014	Shrek	100	100%	0	100	Sold Out
Wednesday, 21 May 2014	The Wizard of Oz	100	100%	0	100	Sold Out
Wednesday, 21 May 2014	Stranger In The Woods	100	100%	0	100	Sold Out
Friday, 23 May 2014	Stranger In The Woods	100	100%	0	100	Sold Out

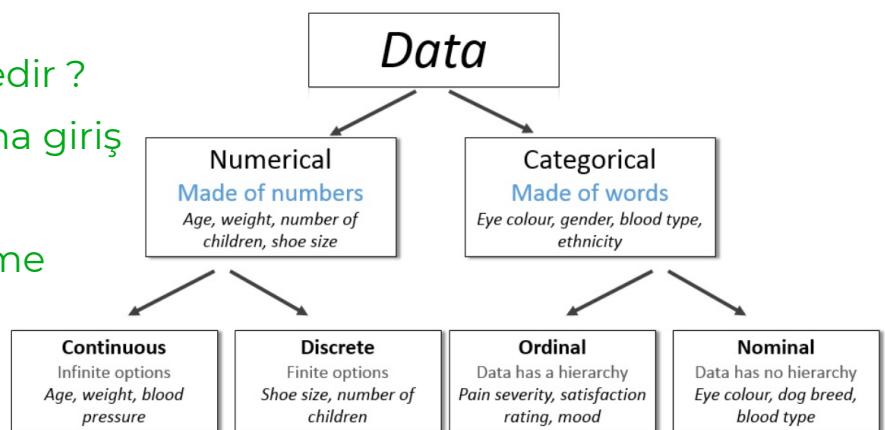
TECHPRO EDUCATION



**Sizi bugünü derse  
hazırlayacak pre-class  
materyalleri ile  
antrenman yaptınız  
mi?**

## Introduction to Data

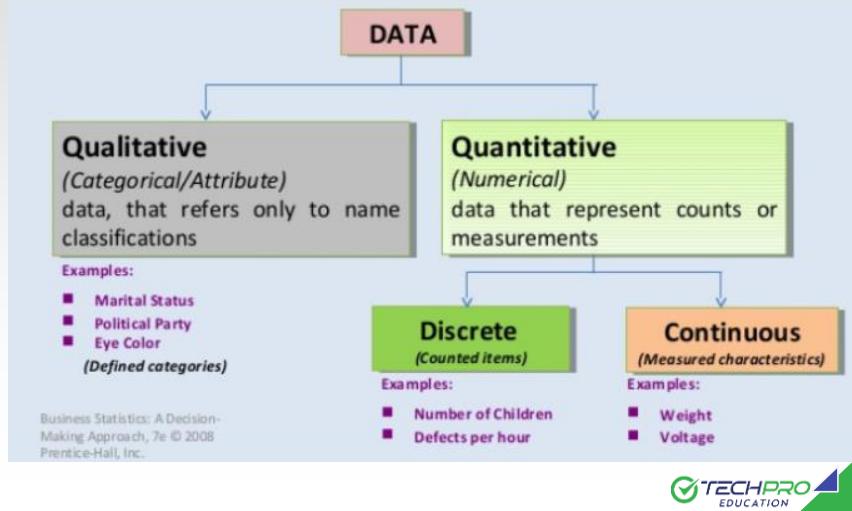
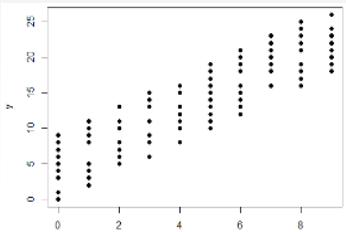
- Data - Veri nedir ?
- EDA kavramına giriş
- Data Türleri
- Veriyi elde etme



## Data Types

### Numerical Data

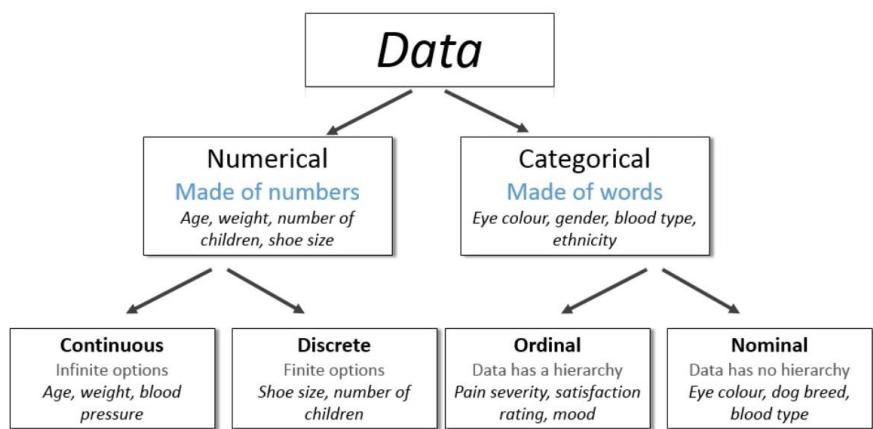
- Discrete
- Continuous



## Data Types

### Categorical Data

- Nominal
- Ordinal



# QUESTION

Race

Height

**Which is continuous data ?**

## ML Models Content as Categorical Data

### Label Encoder

#### Raw Data (Ham -Orj. Data)

- Country sütunu ML model için encoding yapılıyor

#### Ready for Model (ML Model)

- ML modele uygun hale getirildi

Encoding

ID	Country	Population
1	Japan	127185332
2	U.S	326766748
3	India	1354051854
4	China	1415045928
5	U.S	326766748
6	India	1354051854

ID	Country	Population
1	0	127185332
2	1	326766748
3	2	1354051854
4	3	1415045928
5	1	326766748
6	2	1354051854

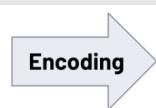
## ML Models Content as Categorical Data

### One Hot Encoder

#### Raw Data

- Country sütunu One Hot Encoding ile düzenleniyor

ID	Country
1	Japan
2	U.S
3	India
4	China
5	U.S
6	India



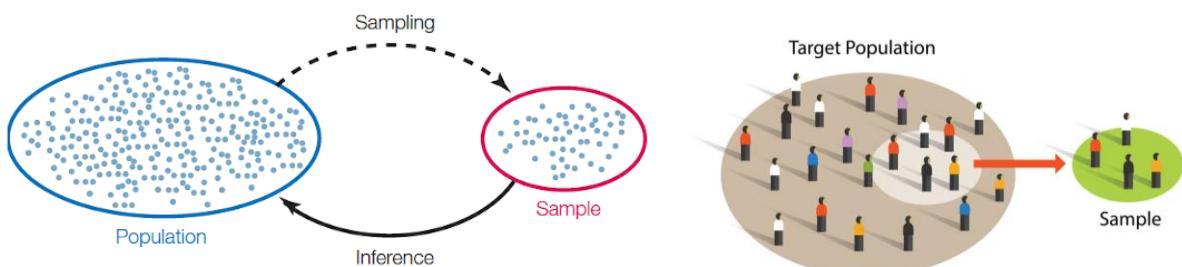
#### Ready for Model (ML Model)

- ML modele uygun hale getirildi

ID	Country_Japan	Country_U.S	Country_India	Country_China
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
5	0	1	0	0
6	0	0	1	0

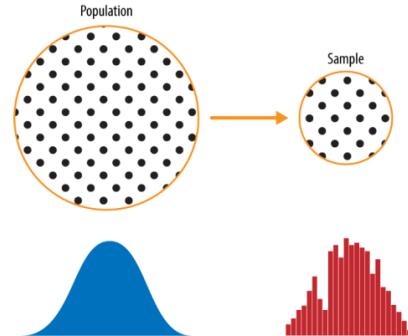
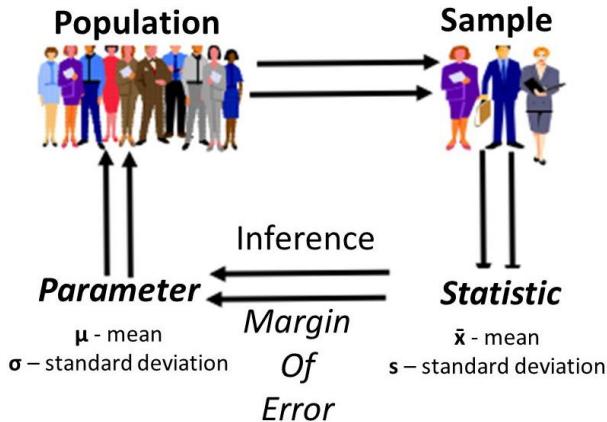


## Parameters and Statistics



## Parameters and Statistics

### The Basic Paradigm



- Populations have Parameters (like  $\mu$ ,  $\sigma^2$ ,  $\theta$ ,  $p$ )
- Samples have Statistics, functions of observed data, like  $\bar{x}$ ,  $\tilde{x}$ ,  $s^2$ ,  $\hat{\theta}$ ,  $\hat{p}$

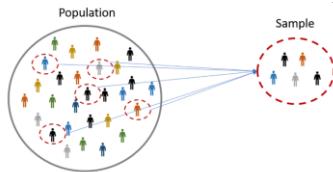


## Parameters and Statistics

- Parametre, *kitlenin* sahip olduğu özelliklerin sayısal bir açıklamasıdır.
- İstatistik, *örneklemi*n sahip olduğu özelliklerin sayısal bir açıklamasıdır.

$$\begin{array}{ccc} \text{Parametre} & \longrightarrow & \text{Kitle} \\ \hline \hline \end{array}$$

$$\begin{array}{ccc} \text{İstatistik} & \longrightarrow & \text{Örneklem} \\ \hline \hline \end{array}$$



Sample (Örneklem):

- Belirtilen değerin bir popülasyon (kitle) parametresini mi yoksa örneklem istatistiğini mi tanımladığına karar verelim.
- a.) 450 üniversite öğrencisinden oluşan bir örneklemle ilgili bir anket, öğrenciler için ortalama haftalık gelirin 325 dolar olduğunu bildirmiştir.
- 325 doların ortalaması bir örneklem dayandığından, bu bir örneklem istatistiğidir.
- b.) Tüm öğrenciler için ortalama haftalık gelir 405 dolardır.
- Ortalama 405doları bir kitleye dayandığından, bu bir kitle parametresidir.



# Probability vs Statistics

## ▶ Probability

- **Olasılıkta (İhtimaliyet hesabında),** randomization (rastgelelik) süreci varsayılar veya bilinir, ve hesaplamalar aşağıdaki gibidir:

Uygun veya hatasız bir madeni para verildiğinde, 10 para atışından 7'sinin tura gelme ihtimali-olasılığı nedir?

## ▶ Statistics

- İstatistikte, sonuç bilinir ve süreç hakkında bir çıkarımda bulunmak (Inferention) için kullanılır.

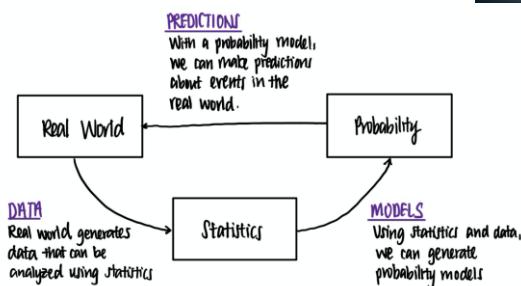
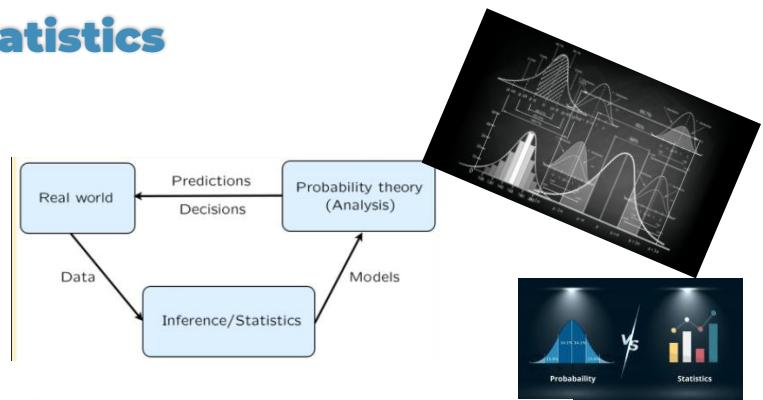
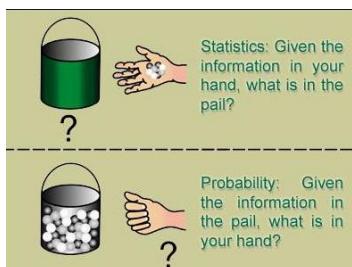
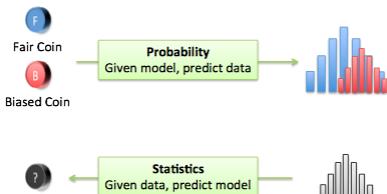
Örnek,

10 yazı turadan 7'sinin sonucu az önce gözlemlediğime göre, yazı tura sonuçları makul mü?



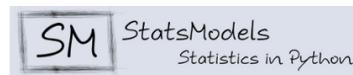
# Probability vs Statistics

## Probability & Statistics





## Statistics with Python



pandas  
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



xarray



NumPy



SciPy



TIMEOUT

Ne durumdayız ?



## Level of Measurement

- 1940'larda, Stanley Smith Stevens dört ölçüm seviyesini -ölçeği tanıttı: nominal, ordinal(sıralı), Interval (aralık) ve Ratio (oran).
- Bunlar bugün hala bir değişkenin özelliklerini tanımlamamanın bir yolu olarak yaygın olarak kullanılmaktadır. Bir değişkenin ölçüm seviyesini-ölçeğini bilmek, doğru istatistiksel analizin seçilmesinde önemli bir husustur.

## LEVEL OF MEASUREMENT

TYPE	MEASURE PROPERTY
NOMINAL	Classification, membership = , ≠
ORDINAL	Comparison Level > , <
INTERVAL	Difference , closeness + , -
RATIO	Magnitude, amount *, /

## Level of Measurement

### NOMINAL DATA

Nominal data divides variables into mutually exclusive, labeled categories.

#### Examples



#### How is nominal data analyzed?

Descriptive statistics: Frequency distribution and mode

Non-parametric statistical tests

### ORDINAL DATA

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

#### Examples



#### How is ordinal data analyzed?

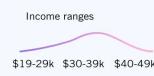
Descriptive statistics: Frequency distribution, mode, median, and range

Non-parametric statistical tests

### INTERVAL DATA

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

#### Examples



#### How is interval data analyzed?

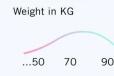
Descriptive statistics: Frequency distribution, mode, median, and mean; range, standard deviation, and variance

Parametric statistical tests (e.g. t-test, linear regression)

### RATIO DATA

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

#### Examples

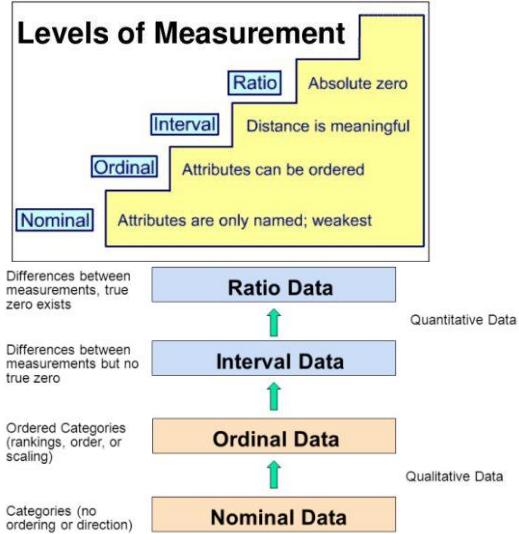


#### How is ratio data analyzed?

Descriptive statistics: Frequency distribution, mode, median, and mean; range, standard deviation, variance, and coefficient of variation

Parametric statistical tests (e.g. ANOVA, linear regression)

# Level of Measurement



## THE FOUR LEVELS OF MEASUREMENT:

Nominal	Ordinal	Interval	Ratio
Categorizes and labels variables	✓	✓	✓
Ranks categories in order	✓	✓	✓
Has known, equal intervals		✓	✓
Has a true or meaningful zero			✓

## Levels of Measurement

Nominal	Ordinal	Interval	Ratio
"Eye color"	"Level of satisfaction"	"Temperature"	"Height"
Named	Named	Named	Named
	Natural order	Natural order	Natural order
	Equal interval between variables	Equal interval between variables	Has a "true zero" value, thus ratios between values can be calculated

# Level of Measurement

## ▶ Nominal Level o Measurement

Sınıflayıcı ölçme düzeyi de denir

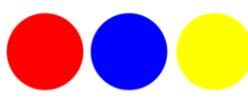
- Sıralama yok
- Kategorize etme var
- Aidiyet var



MALE



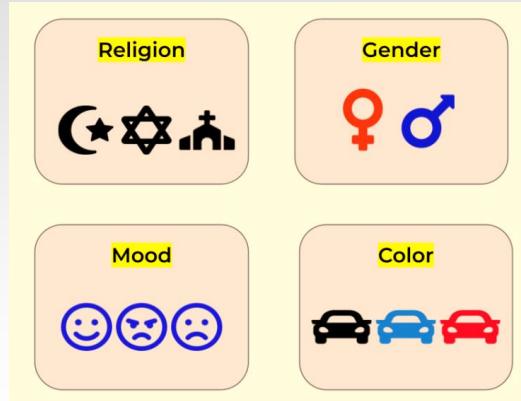
FEMALE



RED

BLUE

YELLOW



# Level of Measurement

## NOMINAL DATA

Nominal data divides variables into mutually exclusive, labeled categories.

### Examples

Eye color  
  
 Blue      Brown      Green

Smartphone  
  
 iPhone      Samsung      Moto

Transport  
  
 Bus      Train      Car

### How is nominal data analyzed?

Descriptive statistics:  
 Frequency distribution and mode

Non-parametric statistical tests



## Level of Measurement

### Ordinal Level of Measurement

Sıralayıcı ölçme düzeyi de denir

- Sıralama var
- Gerçek değer yok
- Karşılaştırma var



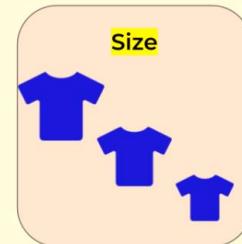
HOT



HOTTER



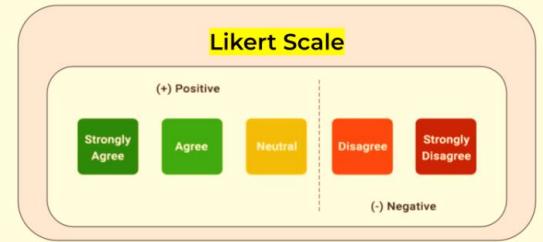
HOTTEST



Size



Quality



# Level of Measurement

## ORDINAL DATA

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

### Examples

School grades



Education level



Seniority level



### How is ordinal data analyzed?

**Descriptive statistics:**  
Frequency distribution,  
mode, median, and range

**Non-parametric**  
statistical tests



## Types of Measurement Scales

### Nominal scale

It's used to label variables in different classifications and does not imply a quantitative value or order.



### Ordinal Scale

It's used to represent non-mathematical ideas such as frequency, satisfaction, happiness, a degree of pain, etc.

### Interval Scale

It's defined as a numerical scale where the order of the variables as well as the difference between these variables is known.



### Ratio Scale

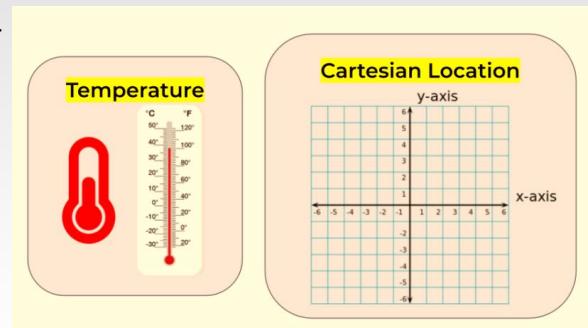
It's a variable measurement scale that not only produces the order of the variables, but also makes the difference between the known variables along with information about the value of the true zero.

## Level of Measurement

### ▶ Interval Level of Measurement

Eşit aralık ölçme düzeyi de denir

- Ölçülebilir
- True zero yoktur, arbitrary zero var
- Büyüklük ifadesi var



## Level of Measurement

### INTERVAL DATA

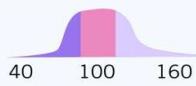
Interval data is measured along a numerical scale that has equal intervals between adjacent values.

#### Examples

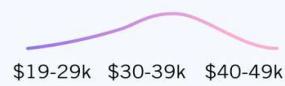
Temperature



IQ score



Income ranges



#### How is interval data analyzed?

**Descriptive statistics:** Frequency distribution; mode, median, and mean; range, standard deviation, and variance

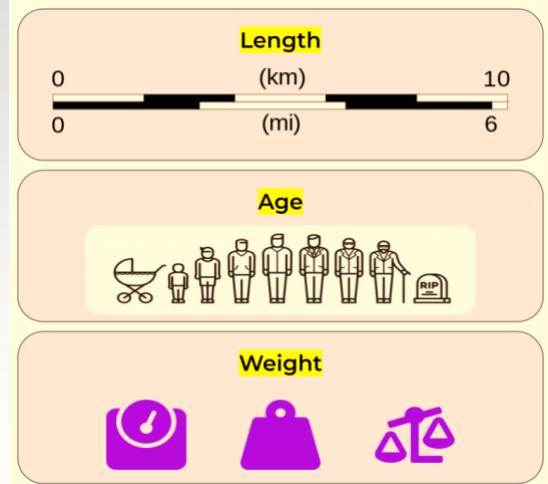
**Parametric statistical tests** (e.g. t-test, linear regression)

## Level of Measurement

### Ratio Level of Measurement

Oranlı ölçme düzeyi de denir

- Ölçülebilir
- True zero vardır
- Büyüklük ifadesi var



## Level of Measurement

### RATIO DATA

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

#### Examples

Weight in KG



Number of staff



Income in USD



#### How is ratio data analyzed?

**Descriptive statistics:** Frequency distribution; mode, median, and mean; range, standard deviation, variance, and coefficient of variation

**Parametric statistical tests** (e.g. ANOVA, linear regression)



# QUESTION

**GSMH hangi tür bir Level of Measurement ?**



<https://youtu.be/hZxnzfnt5v8>



## Matching Variable vs. Level of Measurement

**Nominal**

**School grade**

**Ordinal**

**Temperature**

**Interval**

**Weight**

**Ratio**

**Eye Color**

**Length**



**Bugünkü ders verimli  
geçti**

