DT/NT : DATA SCIENCE
LESSON : STATISTICS -2
SUBJECT : PROBABILITY DISTRIBUTIONS

BATCH: 223

TECHPRO EDUCATION

techproeducation.com +1 (585) 304 29 59

# STATISTICS - 2

Data Science Program
***Statistics Session -6***

# Session - 6 Content

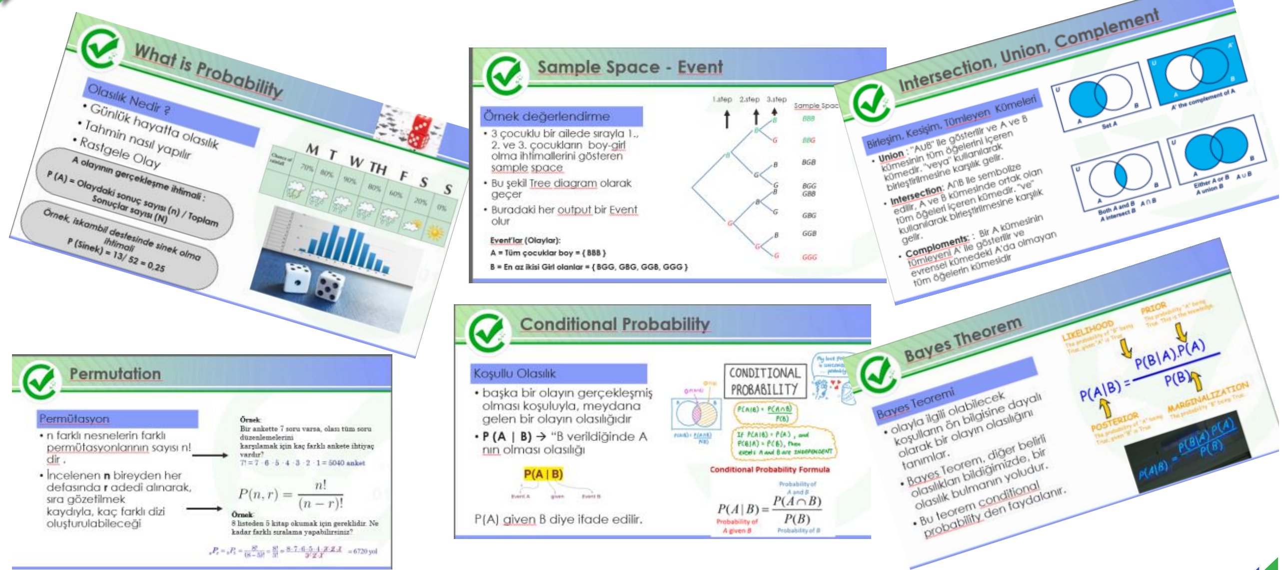
## Content

- Random Variables
- Discrete Probability Distributions
  - Binomial Distribution
  - Bernoulli Distribution
  - Poisson Distribution
- Continous Probability Distributions
  - Uniform Distribution
  - Normal Distribution
  - Standard Distribution
  - T Distribution

RECAP

Herkes önceki dersten hatırladığı 1 cümle yazabilir mi?

# Recap – Previous Lesson




TECHPRO EDUCATION

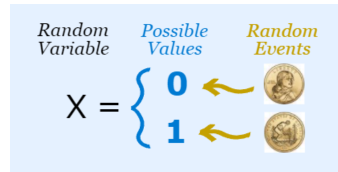

# RANDOM VARIABLES

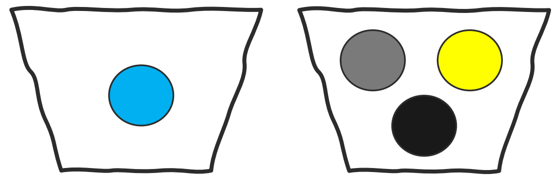
*Rasgele (Rassal) Değişken*

# Random Variable

## Rasgele Değişken

- Bir değişkenin değeri istatistiksel bir deneyin sonucuysa, bu değişken rastgele bir değişkendir.
- Şansa dayalıdır
- Bu değerleri önceden kesin olarak bilmemiz mümkün değildir



Ordinary Variable

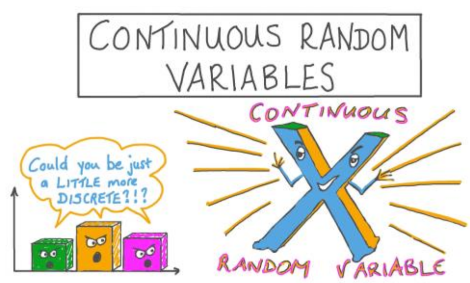
Random Variable

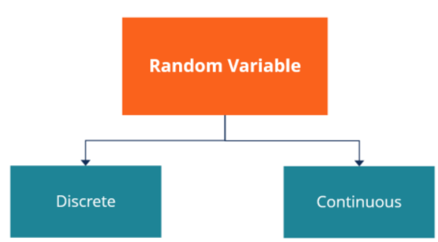
# Random Variable: Discrete - Continous

## Kesikli – Sürekli Değişkenler

- **Discrete** (Kesikli (Ayrık)): sadece belli sayılar gelebilir (zar atımında 1-2-3..... olur ama 1,5 Olmaz..

- **Continous** Sürekli: Bir aralıktaki değerlerin herhangi birini alabilir. Tamsayı veya küsurlu olabilir. Boy, kilo gibi..



# Discrete - Continous

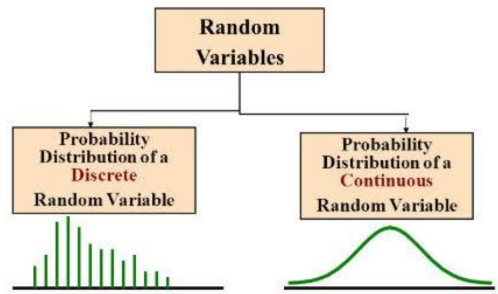
## Kesikli – Sürekli Değişkenler

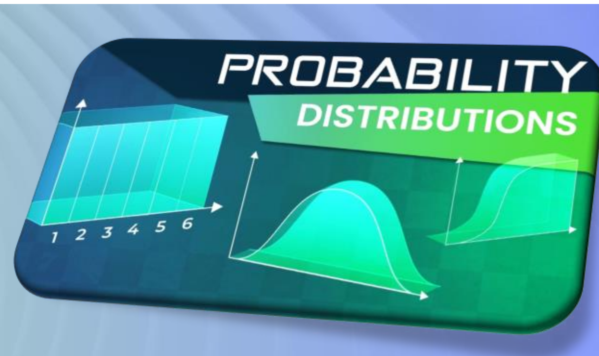
**Örnek**

- X rasgele değişkenin kesikli mi yoksa sürekli mi olduğuna karar verin.

a.) Arabanızın bir benzin deposuyla gittiği mesafe

b.) Data Science sınıfında bugünkü derse katılan öğrenci sayısı




# PROBABILITY DISTRIBUTION

*Olasılık Dağılımları*

# Probability Distribution

## Olasılık Dağılımı

- Olasılık dağılımlarını belirsizliği ortadan kaldırmak amacıyla kullanırız.
- istatistiksel bir deneyin her sonucunu, gerçekleşme olasılığıyla ilişkilendiren bir tablo veya denklemdir.
- Rastgele bir değişkenin değerlerinin dağılımı, olasılık dağılımları ile tanımlanır

**Rastgele değişkene ait matematiksel modeller (fonksiyonlar)**

**Discrete Probab. Distr.**

Discrete Random Variables için Olasılık dağılımları

**Continous Probab. Distr.**

Continous Random Variables için Olasılık dağılımları

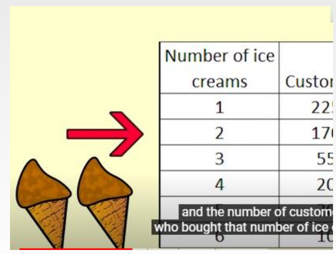
# Olasılık dağılımı neden önemlidir?

1. Olasılık tahminlerinde bulunmaya yardımcı olur.
2. İstatistiksel analizler yapılabilmesini sağlar.
3. Bazı ML modelleri olasılık dağılımı varsayımlarıyla çalışır.

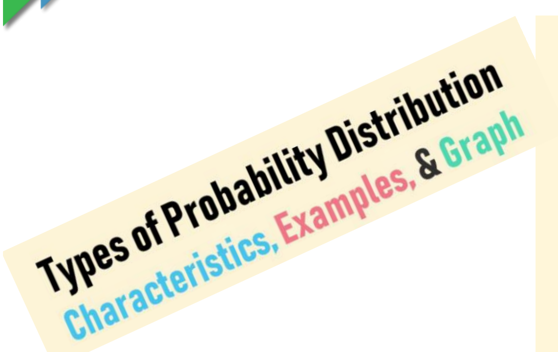
# YOUTUBE VIDEO ONERILERİ

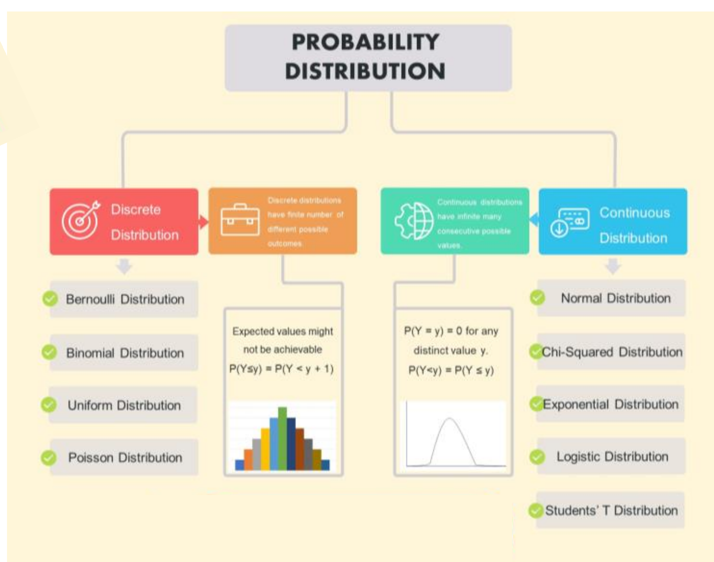
https://www.youtube.com/watch?v=lHCpYeFvTs0

- Random Variables and Probability Distributions




## Probability Distribution



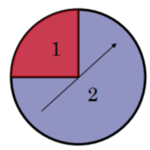
# Discrete Probability Distributions

## Kesikli Olasılık Dağılımı

- Genelde bir tablo ile anlaşılır.
- Şu koşullar sağlanmalı:
  - Kesikli rasgele değişkenin her değerinin olasılığı 0 ile 1 arasındadır.
  - Tüm olasılıklar toplamı 1'e eşittir.
- Yol haritası:
  - Olası sonuçlar için bir frekans dağılımı yapın
  - Frekansların toplamını bulun.
  - Frekansları, frekansların toplamına bölerek olası her sonucun olasılığını bulun
  - Her olasılığın 0 ile 1 arasında olduğunu ve toplamın 1 olduğunu kontrol edin.

## Örnek

- Aşağıdaki iki bölüme ayrılmış alandan 1 nolu yere inme olasılığı 0.25; 2 nolu yere inme olasılığı 0.75 olsun. X üzerine inilen yerin sayısı ise X rasgele değişkeni için bir olasılık dağılımı oluşturun




| $x$ | $P(x)$ |
|---|---|
| 1 | 0.25 |
| 2 | 0.75 |

Her olasılık 0 ile 1 arasında

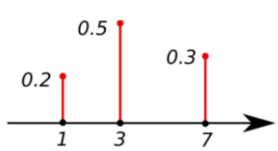
Toplamları 1 olmalı.

# Discrete Probability Distributions

## Probability Mass Function (PMF)

$$p_X(x_i) = P(X = x_i)$$

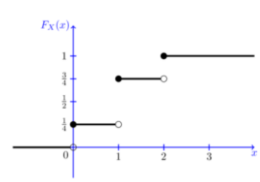
| Values of X | Probability |
|---|---|
| 1 | 0.2 |
| 3 | 0.5 |
| 7 | 0.3 |




## Cumulative Distribution Function (CDF)

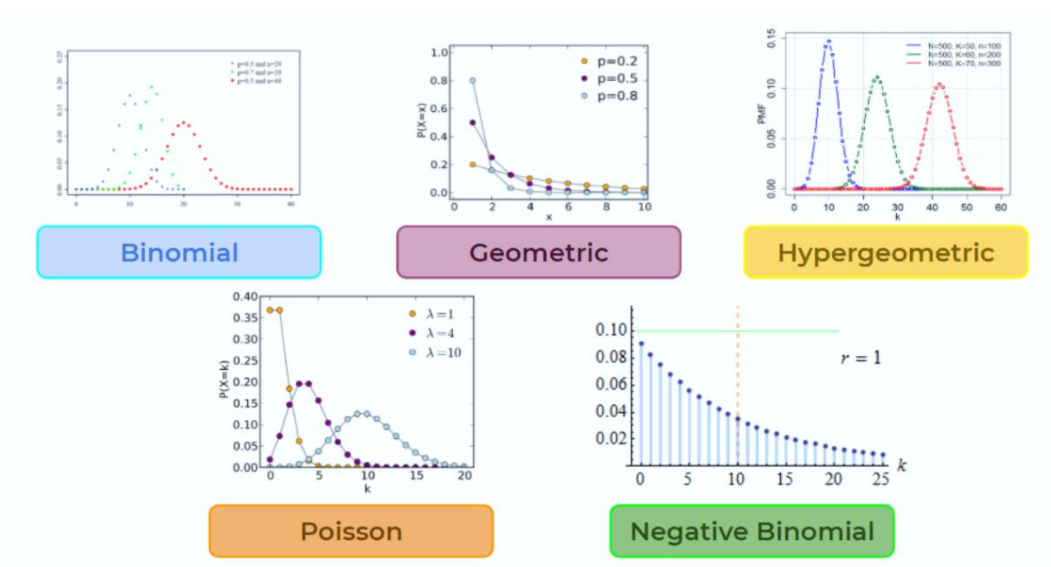
$$F_X(x) = \mathrm{P}(X \le x) \quad \textbf{(Eq.1)}$$

| Values of X | Probability |
|---|---|
| 0 | 0.25 |
| 1 | 0.5 |
| 2 | 0.25 |



# DISCRETE DISTRIBUTIONS

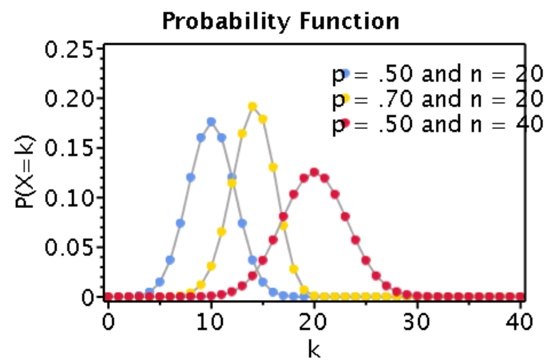
## Discrete Distributions



## Binomial Distributions

### Binom Dağılımı Özellikleri

- Tekrarlanan denemeler vardır
- İki muhtemel sonuç vardır (Success veya Failure)
- Success olasılığı sabittir
  - Başarı olasılığı (p) ve başarısızlık olasılığı q=1-p dir.
- Denemeler bağımsızdır




## Binomial Distributions

### Binom Formülü

$$P(X=x) = \frac{n!}{x!\,(n-x)!} * P^x * (1-P)^{n-x}$$

- n= deneme sayısı
- x=istenen başarı sayısı
- p=bir denemede başarı elde etme olasılığı
- *q=1- p (bir denemede başarısızlık olasılığı)*

Yazı Tura atmada n= 3 için tüm ihtimaller :

TTT, TTY, TYT, TYY, YTT, YTY, YYT, YYY)

P(X=0)=1/8
P(X=1)=3/8

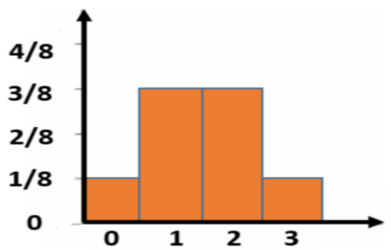
P(X=2)=3/8
P(X=3)=1/8

n= 3 için yeşil renkli alandak tüm ihtimaller ışığında bu olasılıklar bulunur. Örnek, P(x=1)=3/8 için 1 kez Y gelmesi ihtimalidir.. TTY – TYT - YTT lerde Yazı vardır. 3/8

# Binomial Distributions

## Binom Dağılımı

P(X=0)=1/8
P(X=1)=3/8
P(X=2)=3/8
P(X=3)=1/8




## Binom Dağılımnda Mean and Std. Dev.

- Mean

$$\mu = np$$

- Standard Deviation

$$\sigma = \sqrt{np(1-p)}$$

INTERVIEW QUESTION

**9. What is the Binomial Probability Formula?**

"The binomial distribution consists of the probabilities of each of the possible numbers of successes on N trials for independent events that each have a probability of π (the Greek letter pi) of occurring." *Read more*

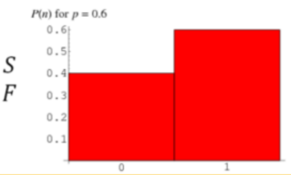
# Bernaulli Distribution

### Bernolli Dağılımı Özellikleri

- ‘Special case of Binomial’
- İki muhtemel sonuç vardır
  - Success
  - Failure
- Denemeler bağımsızdır
- Başarı olasılığı sabittir

$$p(x) = P(X = x) = \begin{cases} q = 1 - p & x = 0 \\ p & x = 1 \end{cases}$$

$$X = \begin{cases} 1 & Bernoulli\ trial = S \\ 0 & Bernoulli\ trial = F \end{cases}$$

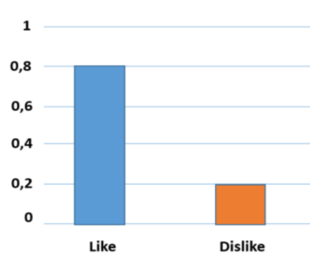



TECHPRO EDUCATION

# Bernaulli Distributions

### Bernoulli Dağılımı

- Bir deneyde başarı ve başarısızlık diye nitelendirilen iki sonuçla ilgilenildiğinde bu deneye (iki sonuçlu) Bernoulli deneyi ya da Bernoulli denemesi denir.




### Bernoulli Dağılımnda Mean and Std. Dev.

- Mean

$$\mu = p$$

- Standard Deviation

$$\sigma = \sqrt{p(1-p)}$$

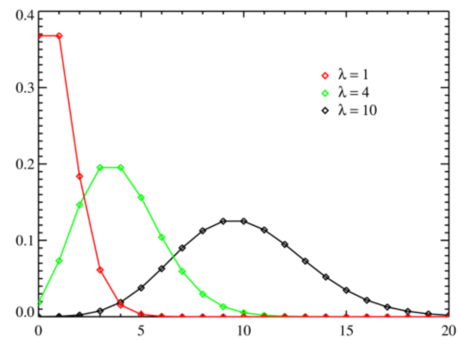
TECHPRO EDUCATION

# Poisson Distributions

## Poisson Dağılımı Özellikleri

- Poisson dağılımı, T zamanında meydana gelen X olay sayısının olasılığını verir.
- Denek sayısı olan n büyük iken p de çok küçük ise binom dağılımı poisson dağılımına yaklaşır

$$P(X) = \frac{\lambda^X . e^{-\lambda}}{X!}$$




# Poisson Distributions

## Poisson Dağılımı

- λ = meydana gelen ortalama olay sayısı
- X = aradığımız olay sayısı.
- e = 2.71828 (Euler sayısı, bir sabit)

- Olaylar bağımsızdır
- Ortalama oranı sabittir
- İki olay aynı anda gerçekleşmez

$$P(X) = \frac{\lambda^X . e^{-\lambda}}{X!}$$

$$P(X) = \frac{\lambda^X . e^{-\lambda}}{X!}$$

$$P(X) = \frac{3^4 . e^{-3}}{4!}$$

$$P(X) = \frac{81.(0,04978)}{24}$$

$$P(X) = 0,16$$

# Poisson Distributions

## Poisson Dağılımı Örnek

Bir şehirde ender rastlanan bir hastalıktan, bir hafta içinde ortalama ölen kişi sayısı 4’ dür. Belli bir hafta içinde bu hastalıktan,

a) Hiç kimsenin ölmemesi

b) En az 2 kişinin ölmesi

c) 3 kişinin ölmesi    olasılıkları nedir.

$X$: *bir haftada bu hastalıktan ölenlerin sayısı*

$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,\ldots, \lambda = 4$

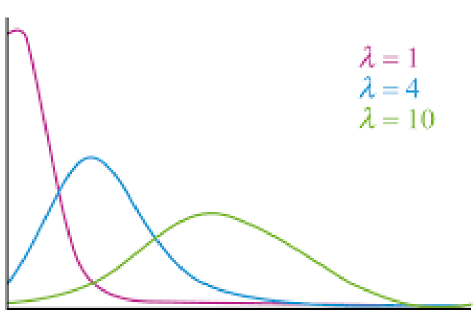
**a)** $P(X = 0) = \frac{e^{-4}4^0}{0!} = 0.0183$

**b)** $P(X \geq 2) = 1 - P(X < 2) = 1 - \left(P(X = 0) + P(X = 1)\right) = 1 - \left(\frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!}\right) =$
$= 1 - (0.0183 + 0.0733) = 1 - 0.0916 = 0.9084$

**c)** $P(X = 3) = \frac{e^{-4}4^3}{3!} = 0.195$

# Poisson Distributions

## Pisson Dağılımı




## Poisson Dağılımnda Mean and Std. Dev.

- Mean

$$\mu = \lambda$$

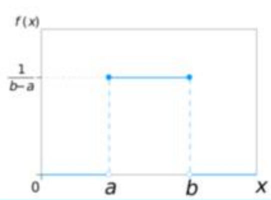
- Variance

$$\sigma^2 = \lambda$$

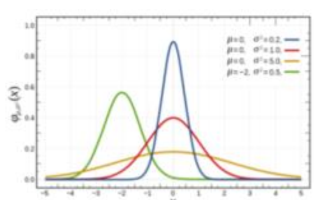
# CONTINOUS PROBABILIY DISTRIBUTIONS

Sürekli Olasılık Dağılımları

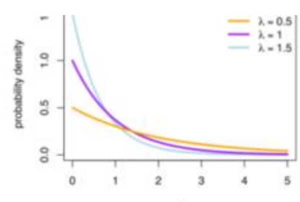
## Continous Probability Distributons



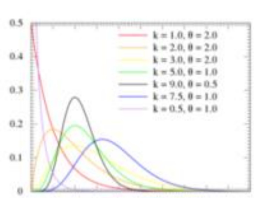

Uniform



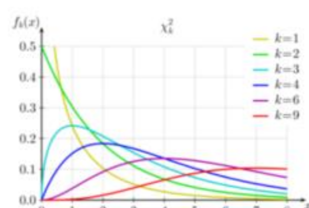
Normal




Exponential



Gamma

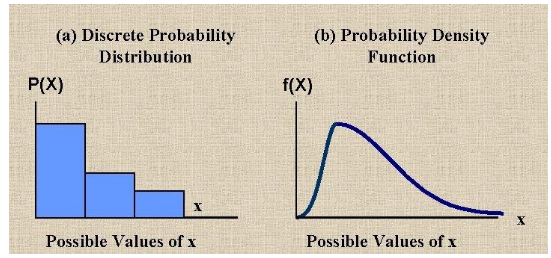


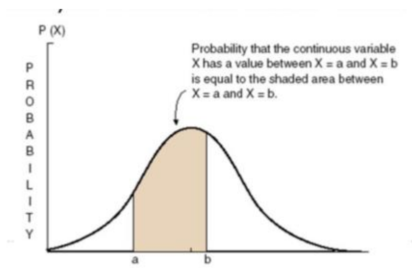

Chi-square

## Continous Probability Distributions

### Sürekli Olasılık Dağılımları

- Her aralığın olasılığı 0 ile 1 arasındadır. Bu, eğrinin altında, o aralığın üzerinde kalan alandır.
- Tüm olası değerleri içeren aralığın olasılığı 1'e eşittir, dolayısıyla eğrinin altındaki toplam alan 1'e eşittir.



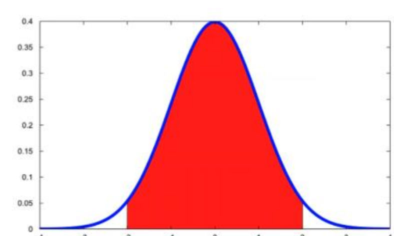

**Sürekli Anakütle Dağılımları**

Sürekli düzgün dağılım
Normal dağılım
Üstel dağılım
Lognormal dağılım
Gamma dağılımı
Ki-kare dağılımı
…………………
…………………

## Continous Probability Distributions

### Sürekli Olasılık Dağılımları

- Probability Density Function (PDF)
  - Y; X random değişkenin bir fonksiyonudur
  - Y; tüm X değerleri için 0'a eşit veya büyüktür
  - Eğri altındaki kalan alan 1 e eşittir.

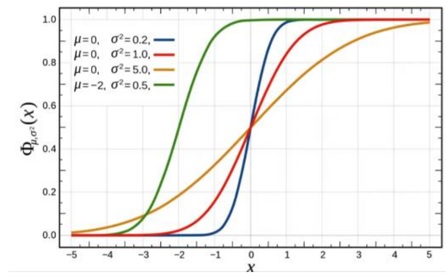



Alan = genişlik*Uzunluk = 1 * 1 = 1

# Continous Probability Distributions

## Sürekli Olasılık Dağılımları

- Cumulative Distribution Function (CDF)
- X sürekli bir rastgele değişken olsun. Bu durumda kümülatif dağılım fonksiyonu yandaki şekilde tanımlanır




$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

# Uniform Distributions

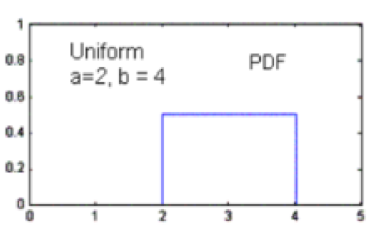
## Düzgün Üniform Dağılım

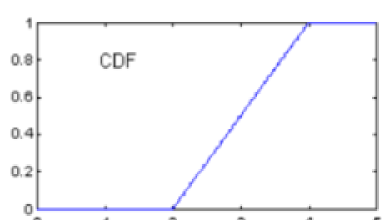
- Rasgele bir değişkenin eşit olasılıklarla meydana gelebilmesidir

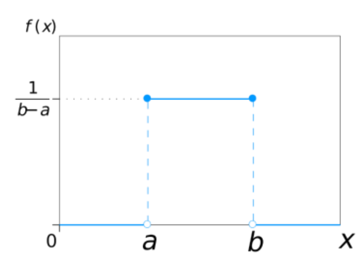
**a** = location parameter

**b** = scale parameter

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$




Mean: $\frac{1}{2}(a+b)$

Variance: $\frac{1}{12}(b-a)^2$

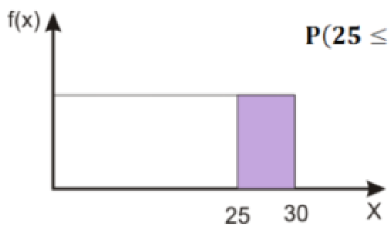
# Uniform Distributions

Süper marketteki kasaya 30 dakikalık periyotta bir müşteri gelmiştir. Bu müşterinin son 5 dakikada gelmiş olma ihtimalini hesaplayınız.

**Örnek 3 ÇÖZÜM :**

Olasılık yoğunluk fonksiyonu:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{diğer} \end{cases} \qquad f(x) = \begin{cases} \dfrac{1}{30-0} & 0 \le x \le 30 \\ 0 & \text{diğer} \end{cases}$$



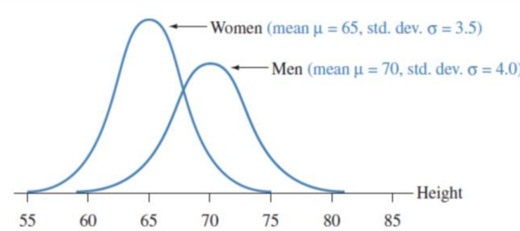

$$P(25 \le X \le 30)$$

$$= \int_{25}^{30} f(x)dx = \int_{25}^{30} \frac{1}{30} dx = \frac{30-25}{30} = 1/6$$

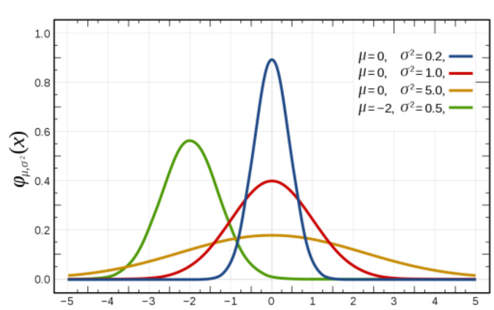
# Normal Distributions

## Normal Dağılım Özellikleri

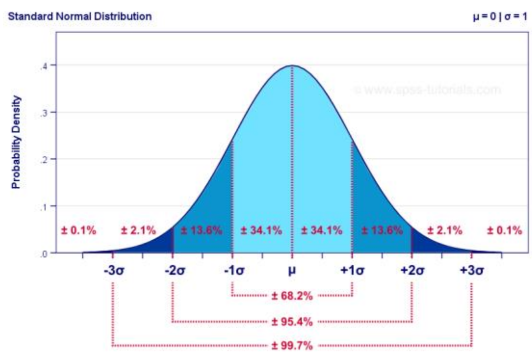
- Dikey eksene göre Simetriktir
- Çan şeklindedir, Mean-mode-medyan eşittir
- Ortalama μ ile ve Standart Sapma σ ile gösterilir.
- Değerler merkez etrafında kümelenme eğilimi gösterir.
- Dağılımın her iki ucu giderek yatay eksene yaklaşır, ancak hiçbir zaman bu eksene değmez (asimptomatik).
- Eğri altında kalan alan 1'e eşittir

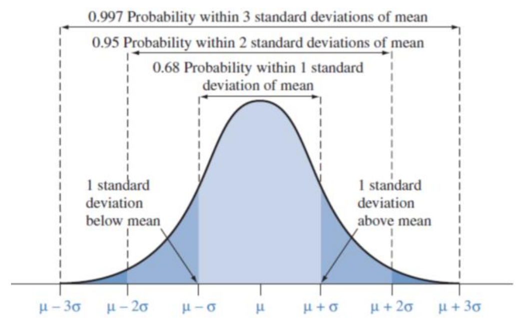



Normal Distributions for Women's Height and Men's Height.



# Normal Distributions




- **x** = normal random variable
- **μ** = mean
- **σ** = standard deviation
- **π** = 3.14159
- **e** = 2.71828

**Mean: μ**

**Variance: $\sigma^2$**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu - x)^2}{2\sigma^2}}$$

PDF

- Örnek olarak; Mean:40, SD:5 ise datanın %68 i hangi aralıkta olur ?

**Question15: What is Normal Distribution?**

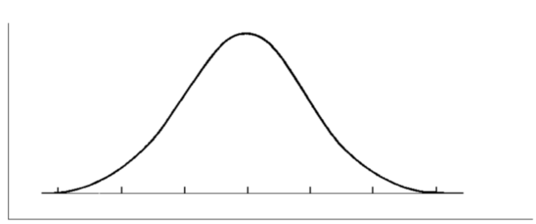
Normal Distribution is a probability distribution that is symmetric about the mean. It is also known as Gaussian Distribution. The distribution appears as a Bell-shaped curve which means the mean is the most frequent data in the given data set.

In Normal Distribution:

- Mean = Median = Mode
- Total area under the curve is 1.

**Question26: What do you understand by the term Normal Distribution?**

Normal distribution, also known as the Gaussian distribution, is a bell-shaped frequency distribution curve. Most of the data values in a normal distribution tend to cluster around the mean.



**Question27: What is the assumption of normality?**

This assumption of normality dictates that if many independent random samples are collected from a population and some value of interest (like the sample mean) is calculated, and then a histogram is created to visualize the distribution of sample means, a normal distribution should be observed.

**Question29: What are some of the properties of a normal distribution?**

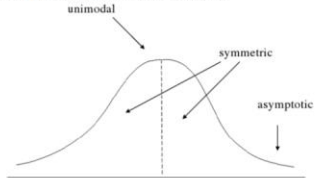
Some of the properties of a Normal Distribution are as follows:

Unimodal: normal distribution has only one peak. (i.e., one mode)

Symmetric: a normal distribution is perfectly symmetrical around its centre. (i.e., the right side of the centre is a mirror image of the left side)

The Mean, Mode, and Median are all located in the centre (i.e., are all equal)

Asymptotic: normal distributions are continuous and have tails that are asymptotic. The curve approaches the x-axis, but it never touches.




INTERVIEW QUESTION

**Question57: What is a bell-curve distribution?**

A bell-curve distribution is represented by the shape of a bell and indicates normal distribution. It occurs naturally in many situations especially while analyzing financial data. The top of the curve shows the mode, mean and median of the data and is perfectly symmetrical. The key characteristics of a bell-shaped curve are –

- The empirical rule says that approximately 68% of data lies within one standard deviation of the mean in either of the directions.
- Around 95% of data falls within two standard deviations and
- Around 99.7% of data fall within three standard deviations in either direction.

**Q6. What do you understand by the term Normal Distribution?**

Data is usually distributed in different ways with a bias to the left or to the right or it can all be jumbled up.

However, there are chances that data is distributed around a central value without any bias to the left or right and reaches normal distribution in the form of a bell-shaped curve.

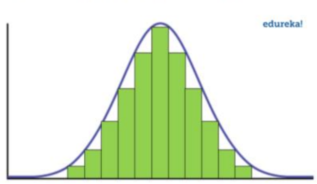


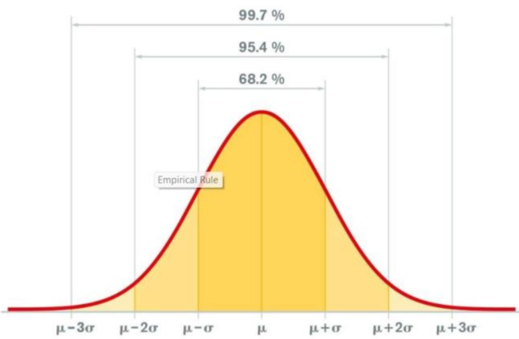

**Figure:** *Normal distribution in a bell curve*

INTERVIEW QUESTION

**Question16: What is the empirical rule?**

Empirical Rule is often called the ***68 – 95 – 99.7*** rule or **Three Sigma Rule**. It states that on a Normal Distribution:

- 68% of the data will be within one Standard Deviation of the Mean
- 95% of the data will be within two Standard Deviations of the Mean
- 99.7 of the data will be within three Standard Deviations of the Mean




INTERVIEW QUESTION

## Normal Distributions – Z Table

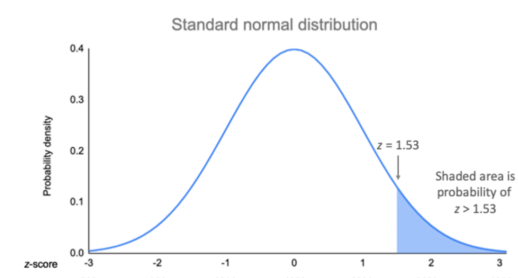
### Z tabloları ile Alan hesaplama

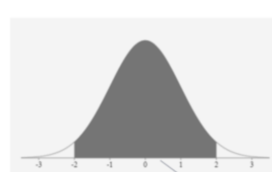
- Z tablosu olarak adlandırılan bu tablolar farklı şekillerde düzenlenmektedir
- Z puanı, puanınızın ortalamadan kaç standart sapma ötede olduğunu size söyler.
- Verilen tablo yardımıyla normal dağılıma ait her türlü olasılık hesaplanabilmektedir
- ortalamanın sağında kalan kısmı tablolarda verilmekte, diğer yarısının aynı olduğu bilinmektedir.

**Z Score**

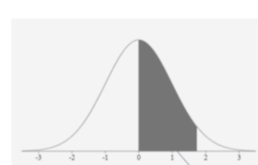
$$z = \frac{x - \mu}{\sigma}$$

Örneğin, x= 205 olduğunu varsayalım. Testin ortalaması (μ) 180 ve standart sapması (σ) 20'dir. Normal bir dağılım varsayarsak, z puanınız 1,25 olur



P(-2≤Z≤2) 0.9545



P(0≤Z≤1.73) 0.4582

# Normal Distributions – Z Table

## Z tabloları ile Alan hesaplama

- Z-tablosu, z-puanı ile bir olasılık hesaplaması yapmamıza yardımcı olur.

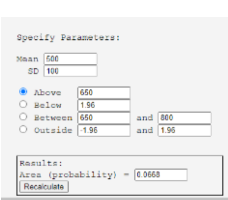
| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |

# Normal Distributions – Z Table

## Z tabloların Kullanma

- Normal dağılmış bir olayda, Mean = 500, SD= 100 ise X=650 ve yukarısı için için olasılık nedir ?

$$z = \frac{x - \mu}{\sigma} = \frac{650 - 500}{100} = 1.50$$



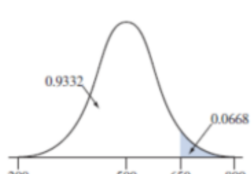
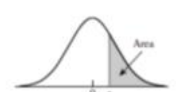




**Table 4 Normal Curve Areas**
Standard normal probability in right-hand tail
(for negative values of $z$, areas are found by symmetry)

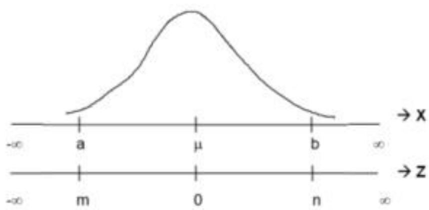



| z | Second decimal place of z: .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .4960 | .4920 | .4880 | .4840 | .4801 | .4761 | .4721 | .4681 | .4641 |
| 0.1 | .4602 | .4562 | .4522 | .4483 | .4443 | .4404 | .4364 | .4325 | .4286 | .4247 |
| 0.2 | .4207 | .4168 | .4129 | .4090 | .4052 | .4013 | .3974 | .3936 | .3897 | .3859 |
| 0.3 | .3821 | .3783 | .3745 | .3707 | .3669 | .3632 | .3594 | .3557 | .3520 | .3483 |
| 0.4 | .3446 | .3409 | .3372 | .3336 | .3300 | .3264 | .3228 | .3192 | .3156 | .3121 |
| 0.5 | .3085 | .3050 | .3015 | .2981 | .2946 | .2912 | .2877 | .2843 | .2810 | .2776 |
| 0.6 | .2743 | .2709 | .2676 | .2643 | .2611 | .2578 | .2546 | .2514 | .2483 | .2451 |
| 0.7 | .2420 | .2389 | .2358 | .2327 | .2296 | .2266 | .2236 | .2206 | .2177 | .2148 |
| 0.8 | .2119 | .2090 | .2061 | .2033 | .2005 | .1977 | .1949 | .1922 | .1894 | .1867 |
| 0.9 | .1841 | .1814 | .1788 | .1762 | .1736 | .1711 | .1685 | .1660 | .1635 | .1611 |
| 1.0 | .1587 | .1562 | .1539 | .1515 | .1492 | .1469 | .1446 | .1423 | .1401 | .1379 |
| 1.1 | .1357 | .1335 | .1314 | .1292 | .1271 | .1251 | .1230 | .1210 | .1190 | .1170 |
| 1.2 | .1151 | .1131 | .1112 | .1093 | .1075 | .1056 | .1038 | .1020 | .1003 | .0985 |
| 1.3 | .0968 | .0951 | .0934 | .0918 | .0901 | .0885 | .0869 | .0853 | .0838 | .0823 |
| 1.4 | .0808 | .0793 | .0778 | .0764 | .0749 | .0735 | .0722 | .0708 | .0694 | .0681 |
| 1.5 | .0668 | .0655 | .0643 | .0630 | .0618 | .0606 | .0594 | .0582 | .0571 | .0559 |
| 1.6 | .0548 | .0537 | .0526 | .0516 | .0505 | .0495 | .0485 | .0475 | .0465 | .0455 |
| 1.7 | .0446 | .0436 | .0427 | .0418 | .0409 | .0401 | .0392 | .0384 | .0375 | .0367 |
| 1.8 | .0359 | .0352 | .0344 | .0336 | .0329 | .0322 | .0314 | .0307 | .0301 | .0294 |
| 1.9 | .0287 | .0281 | .0274 | .0268 | .0262 | .0256 | .0250 | .0244 | .0239 | .0233 |

# Normal Distributions – Z Table

İstenen X rastgele değişkeninin belirli aralıkta değer alma olasılığını hesaplamak için izlenecek yaklaşımlar şöyle özetlenebilir:

**1.** Verilen $a < X < b$ aralığı $m < Z < n$ aralığına dönüştürülür. Yani,




Bu amaçla

$$Z = \frac{X-\mu}{\sigma}$$

dönüşümü kullanılır.

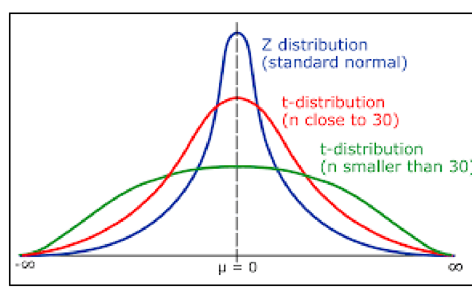
**2.** Karşı gelen P(m<Z<n) değeri tablo yardımıyla belirlenir. Öyle ise P(A<X>b):

$$P\left(\underbrace{\frac{a-\mu}{\sigma}}_{m} \le \underbrace{\frac{x-\mu}{\sigma}}_{x} \le \underbrace{\frac{b-\mu}{\sigma}}_{n}\right) \Rightarrow P(m \le Z \le n)$$ hesaplanır.

# t Distribution (aka, Student's t-distribution)

## t Dağılımı (Student Test)

- örneklem boyutu küçük olduğunda ve/veya popülasyon standart sapması (σ) bilinmediğinde popülasyon parametrelerini tahmin etmek için kullanılır
- Örneklem boyutu ne kadar büyük olursa, t dağılımı o kadar normal dağılıma yaklaşır.
- 30'dan büyük örneklem büyüklükleri için dağılım normal dağılıma çok benzer.




*df = sample size -1*

# t Distribution (aka, Student's t-distribution)

## t Dağılımı (Student Test)

- Sample büyüklüğü küçük ise ve popülasyonun standart sapmasını da bilmiyor isek, ozaman bu durumda t skoru veya t istatistiği kullanılabilir.
- Örnek: Bir Lamba üreticisi ampülün 300 gün yandığını iddia ediyor. Araştırmacı bir şirket rasgele 15 lambayı test ediyor. Bu lambalar ortalama 290 gün yanıyor ve standart sapma 50 gün.
  - Eğer şirketin iddiası doğru olsaydı, bu seçilen 15 lambanın ortalamasının 290 günden fazla az olma olasılığı ne olurdu?

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

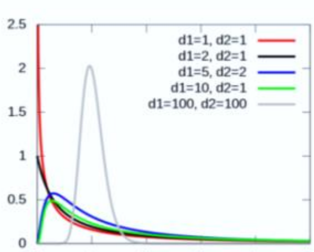
$$t = \frac{290-300}{\frac{50}{\sqrt{15}}}$$

$$= \frac{-10}{12.909945} = -0.7745966$$

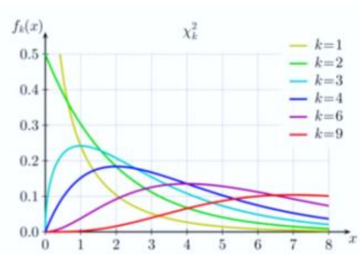
**df = 15 – 1 = 14**
**Popülasyon ort: 300**
**Örneklem ort: 290**
**Örneklem SD: 50**

**Bu t skora göre t table a bakılırsa 0,226 bulunur. İddia 0,23 olasılıkla gerçekleşir.**

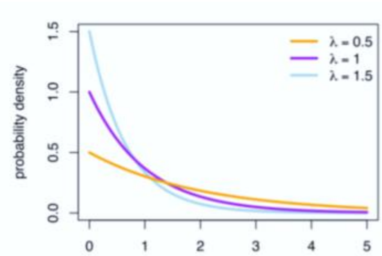
# Continous Probability Distributions



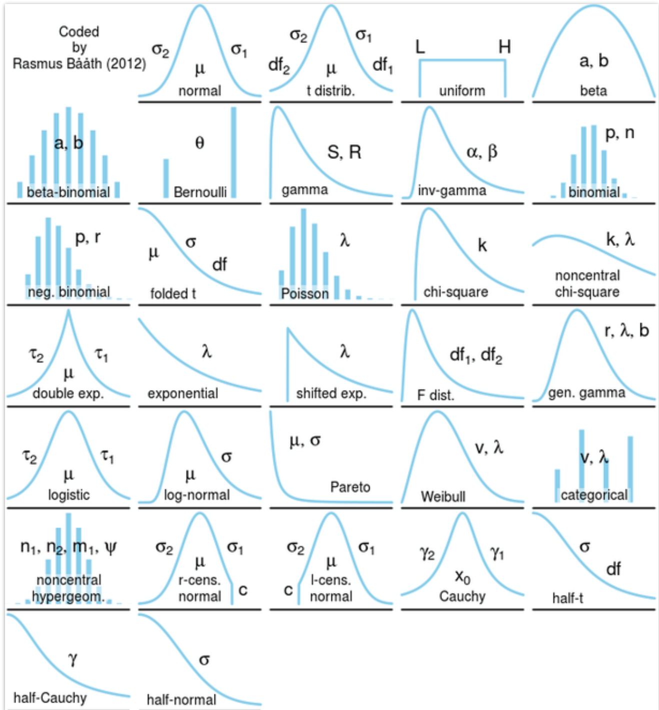

F




Chi-square




Exponential



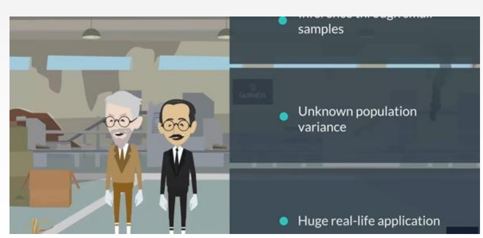

## YOUTUBE VIDEO ONERİLERİ

- The Normal Distribution
- Z-Scores, Standardization, and the Standard Normal Distribution
- Student's T Distribution

https://www.youtube.com/watch?v=mtbJbDwqWLE

https://www.youtube.com/watch?v=2tuBREK_mgE

https://www.youtube.com/watch?v=32CuxWdOlow



## Python Coding

**ProbabilityDistributions_student.ipynb dosyasına bakalım..**

- Bu notebook ta Discerete ve Continous Distribituons için hesaplamalar bulunmaktadır.

**Bugünkü ders verimli geçti**