**VERİLER TEKRAR KULLANILABİLİR AMA TEK BAŞLARINA ANLAM İÇERMEZ.**

**VERİ ÜRETİMİ TELEFON-BİLGİSAYAR-AKILLI CİHAZ-BULUT-CİHAZLAR İLE OLABİLİR**

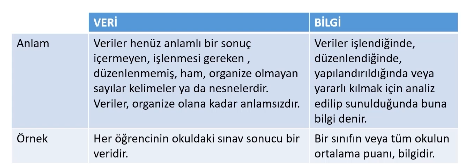
**VERİDEN BİLGİYE GEÇİŞ ÖNEMLİDİR**

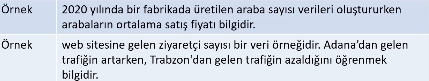
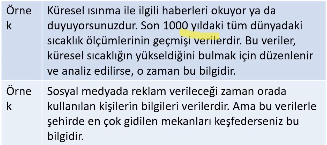
**İŞ DÜNYASINDA ÇOK FAZLA SEKTÖRDE VERİ ÜRETİLİR VE KULLANILIR.**

* **FİRMALAR KARAR ALIRKEN GEÇMİŞ VERİLERİNE GÜVENİR.**

**VERİ TOPLANDIKTAN SONRA GRUPLANARAK, SIRALANARAK VE ÖZETLENEREK KİŞİLER VEYA BİLGİSAYARLARLA İŞLENİP BİLGİYE DÖNÜŞTÜRMELERİNDE ANLAM KAZANIRLAR.**

**VERİ-BİLGİ FARKI**

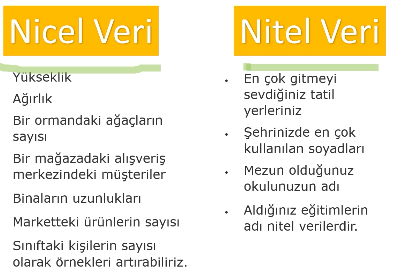
****

****

**TERİMLER VE KAVLAMLAR**

* **BÜYÜK VERİ ANALİTİĞİ=**
* **VERİ MERKEZİ=SUNUCU VE AĞ SİSTEMLERİNİN BARINDIRMAK İÇİN İŞLETMELER TARAFINDAN KULLANILAN FİZİKSEL VEYA SANAL ALTYAPI**
* **VERİ MADENCİLİĞİ=GELECEKTEKİ DAVRANIŞLARI TAHMİN ETMEK İÇİN KULLANILABİLECEK BİR VERİ GRUBUNDA GİZLİ KALIPLAR ARAYAN VERİTABANI UYGULAMALARI SINIFIDIR.**
* **VERİ TABANI=**
* **HAM VERİLER=TOPLANMIŞ ANCAK BİÇİMLENDİRİLMEMİŞ VERİ**
* **YAPILANDIRILMIŞ VERİ=SABİT BİR ALANDA BULUNAN HERHANGİ BİR VERİYİDİR. İLİŞKİSEL VERİTABANINDAVE ELEKTRONİK TABLOLARDA BULUNAN VERİLERİ İÇERİR.**
* **YAPILANDIRILMAMIŞ VERİ=GELENEKSEL BİR SATIR-SÜTUN VERİTABANINDA BULUNAN BİLGİDİR.  **

**VERİ TÜRLERİ**

1. **NİCEL(SAYILSAL)**
2. **NİTEL(SAYISAL OLMAYAN GRUPLAR)(CİNSİYET-YER-İSİM)**

**NİCEL VERİ**

* **KESİKLİ=BELİRLİ BİR ARALIKLA ALINABİLİR, TAM SAYILARDIR**
* **SÜREKLİ=ÖLÇÜM İLE BELİRLENİR, TAM DA OLUR KESİRLİ DE**

**NİTEL VERİ**

* **SIRASIZ=SIRALANMASI MÜMKÜN DEĞİLDİR, GRUPLANIR.**
* **SIRALI=KARŞILAŞTIRMA YAPARAK SIRALANABİLİR.**

**VERİ TOPLAMA YÖNTEMLERİ**

1. **İŞLEM VERİLERİ=ANLAMLI BİLGİLER ÜRETMEK İÇİN VERİ ÖGELERİN TOPLANMASI VE SONUÇ ORTAYA ÇIKARILMASI. ŞİRKETLER İÇİN ÖNEMLİDİR**
2. **WEB VERİLERİ =İNTERNETTEN ÇEKEBİLECEĞİNİZ HER TÜRLÜ VERİ İFADE EDER.**
3. **SENSÖR VERİLERİ=NESNELER TARAFINDAN ÜRETİLİR. OOP OLARAK ADLANDIRILIR.**

**NİTEL ARAŞTIRMA VERİ ANALİZİ=VERİLERİN KALİTE BİLGİSİ KELİMELERDEN, NESNELERDEN, RESİMLERDEN VE GÖRÜNTÜLERDEN OLUŞUR.**

**NİCEL ARAŞTIRMA VERİ ANALİZİ= İSTATİSTİKSEL ARAŞTIRMACILAR RAKAMLARIN GÜCÜNDEN YARARLANIR.**

**VERİ MADENCİLİĞİ**

****

**VERİ MADENCİLİĞİ UYGULAMA ALANLARI**

1. **METİN MADENCİLİĞİ**
2. **GÖRÜŞ MADENCİLİĞİ=ANKETLERDEN VERİ**
3. **GÖRSEL MADENCİLİĞİ=2 VEYA 3 BOYUTLU RESİMLERDE**
4. **GRAFİK MADENCİLİĞİ=GRAFİK MODEL MADENCİLİĞİ**

**ARAŞTIRMA KONUSU VERİ BİLİMİ VE İSTATİSTİK FARKLARI**

**VERİ BİLİMİ**

* **YAPILANDIRILMIŞ VEYA YAPILANDIRILMAMIŞ VERİLERDEN BİLGİ, ANLAM, SONUÇ VE ÖNGÖRÜ ELDE ETMEK İÇİN BİLİMSEL YÖNTEMLERİ, ANALİZ YÖNTEMLERİ, ALGORİTMALAR KULLANAN ÇOK DİSİPLİNLİ BİR ALANDIR.**
* **VERİ BİLİMİ, VERİ AKIŞININ ARTMASI NEDENİYLE DE BİRLİKTE GELİŞEN TEKNOLOJİLERİN GÜCÜNDEN FAYDALANIR. VERİDEN ANLAMLI SONUÇLAR ÜRETİLMESİ İÇİN KULLANILIR.**
* **VERİ BİLİMCİ, GEÇMİŞTEKİ VEYA ANLIK VERİLERİ KULLANARAK GELECEK İLE İLGİLİ TAHMİNLER KURAN YAZILIM PROGRAMLARI ARACILIĞIYLA BU VERİLERİ YORUMLAYAN, SONUÇLAR ÇIKARTAN, VERİLERİ GÖRSELLEŞTİREN, HİKAYELEŞTİREREK ANLATAN, BİLGİ VE ÇÖZÜM ODAKLI KİŞİDİR.**

**VERİ BİLİMCİ ÇALIŞMA SÜRECİ**

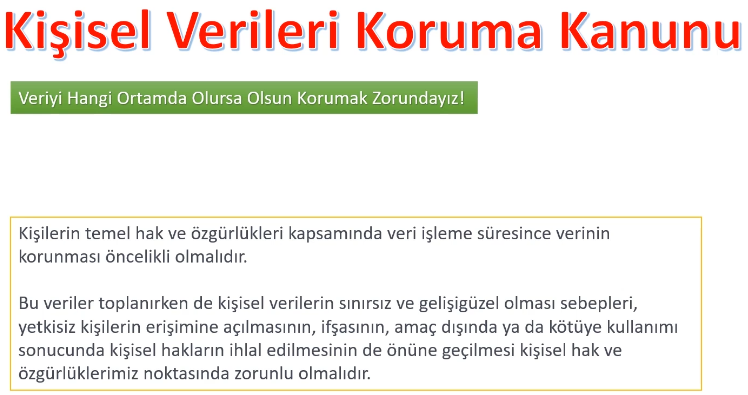
* **PROJE PLANLAMASI YAPMAK**
* **KULLANILACAK ARAÇ BELİRLEME**
* **EKİP İLE TÜM DETAYLARI PAYLAŞ**
* **GEREKLİ VERİLERİ ÇALIŞMA ORTAMINA YÜKLEMEK**
* **İSTATİSTİK VE MATEMATİK KAYNAKLI MODELLER OLUŞTURMA**
* **VERİLERİ GÖRSELLEŞTİRME**
* **DÖKÜMANLARI İYİ SAKLAMAK**
* **TEST ETMEK**
* **GERİ BİLDİRİM VE GÖRÜNTÜLEME YAPMA**
* **ANLAM YARATABİLME**

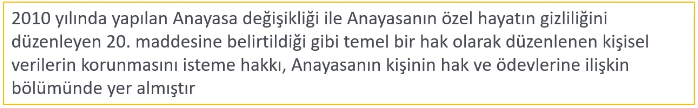
**ARAŞTIRMA KONUSU VERİ BİLİMİ VE İSTATİSTİK FARKLARININ CEVABI**

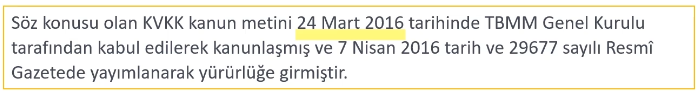
1. **İKİ FARKLI TİP VERİYİ ANALİZ EDİP BİRLEŞTİRMEK VE ANLAMLI HALE GETİRMEK. MİLYONLARCA VERİNİN KEŞFEDİLMESİ İÇİN GELENEKSEL VERİTABANI SİSTEMİ YÖNETİMİ YETERSİZ KALIR**

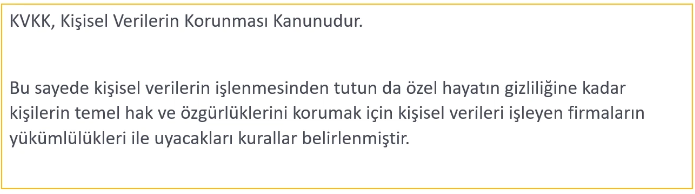
**VERİ BİLİMİNİN DİĞER BİLİMLER İLE İLİŞKİSİ**

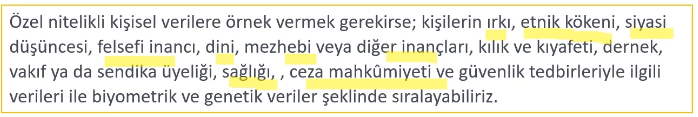
* **İŞLETME YÖNETİMİ**
* **BİLGİSAYAR BİLİMİ**
* **İSTATİSTİK-MATEMATİK**

****

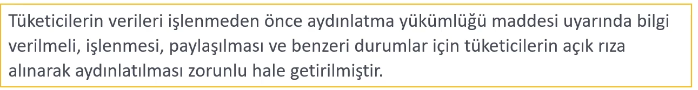
****

****

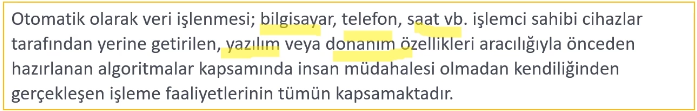
****

****

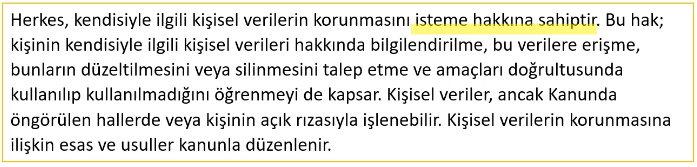
**AYDINLATMA METNİ NEDİR?**

****

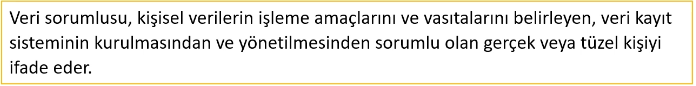
**KİŞİSEL VERİLERİN OTOMATİK YOLLARLA İŞLENMESİ NEDİR?**

****

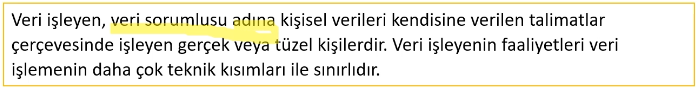
**KVK HAKKININ DAYANAĞI NEDİR VE BU HAK SINIRSIZ MI?**

****

**VERİ SORUMLUSU**

****

**VERİ İŞLEYEN**

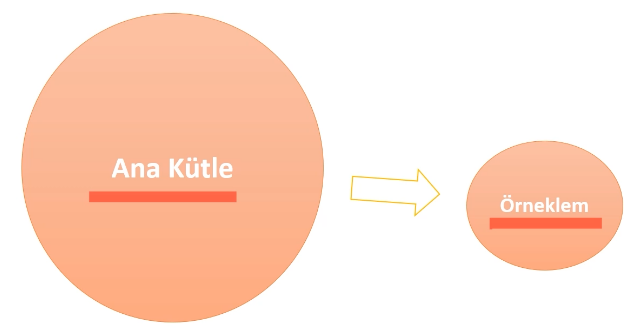
****

**İŞ DÜNYASINDA ÇOK FAZLA SEKTÖRDE VERİ ÜRETİLİR VE KULLANILIR.**

* **TIP**
* **HAVAYOLU ŞİRKETLERİ**
* **İLETİŞİM**
* **OKULLAR**
* **FİNANS**
* **MEDYA**
* **SATIŞ PAZARLAMA**

**BASİT İSTATİSTİKLER**

**Düşünme yöntemleri-**5 adımdanoluşur. Proje içeriği belirleme- verilerin belirlenmesi – verilerin gruplandırılması -verilerin analiz edilmesi -verilerin bilgiye dökülmesi.

**İstatistik**

* Tanımsal
* Çıkarımsal

Örneklem nedir?

Bir araştırmada bütününden seçilen gruba denir.

**Not= ana kütlenin tümünden veri toplanırsa buna tamsayım denir**

**Örneklem denen kullanılır? Bütünü incelemenin zorluğu-zaman-maliyet**

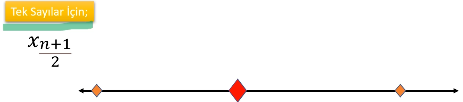
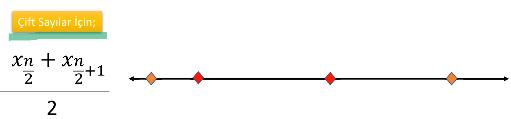
**Örneklemin özellikleri**

* Eşit ağırlıklı ve raslantısal seçilmesi gerekir
* Ana kümeyi net olarak yansıtmaz
* Örneklem büyüklüğü

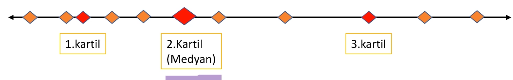
**Aritmetik ortalama**

**Medyan=**bir sayı grubu sıralandığında tam ortada bulunan sayıya denir.

* --10—20—40—45—50—60--65—70--90--100—=>(50+60)/2=55

****

**Mod=**bir sayı setinde en çok tekrar eden mod(tepe değer) denir

**Kartil=** bir veri setinde bulunan değerler sıralandığında 4 parçaya bölen 3 değere denir.

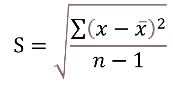
10-15-20-30-40-50-55-70-80-90-100

1.kartil=20

2.kartil(Medyan)=50

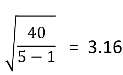
3.kartil=80

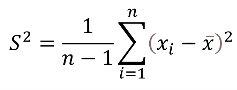
**Değişim aralığı=X()max)-x(min)**

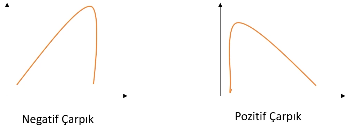
3-5-5-12-7-9-18=> 18-3=15

**Standart sapma(S)=**bir veri setinde kaç kadar sayının ortalamya yakın olduğunu verir. Bu sayı Küçükse ortaya yakın veri sayısı fazladır, büyükse az.

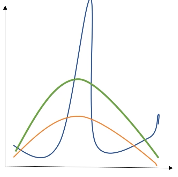


* 2-3-4-6-10=>(2+3+4+6+10)/5=>25/5=5>
  + 5-2=3=>32=>9
  + 5-3=2=>22=>4
  + 5-4=1=>12=>1
  + 6-5=1=>12=>1
  + 10-5=5=>52=>25
* A=(5-15-20-5-5)||b=(10-15-15-5-5) bu iki dizi arasında hangisinin standart sapması küçüktür.
  + 50/5=10|50/5=10
  + 52+52+102+52+5|52+52+52+52
  + √(200/4)|√(125/4)
  + 7.071|5
  + B’nin standart sapması daha küçük

**Varyans(S2)=**standart sapmanın karesidir. Bir serideki verilerin aritmetik ortalamadan farklarının karesi değer sayılarına oranı denir.

* 2-3-4-6-10=>(2+3+4+6+10)/5=>25/5=5>
  + 9+4+1+1+25=40
  + √(40/4)2
  + 40/4=10=>S2=10

**Çarpıklık=**Bir dağılımın normal olup olmadığını test etmek için çarpıklık ve yığılma testleri yapılır. Çarpıklık simetrik olup olmadığını analiz eder. Sağa çarpık genellikle ortalama>ortanca>mod olarak yer alır. Sola çarpık ise mod>ortanca>ortalama

**Basıklık**=grafik gösteriminde yola çıkarak dağılımın basıklığını ölçme sürecidir. Basık olma durumunda standart sapma yüksektir.

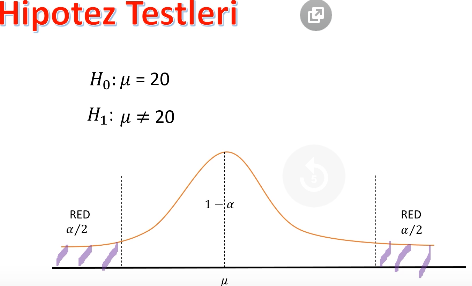
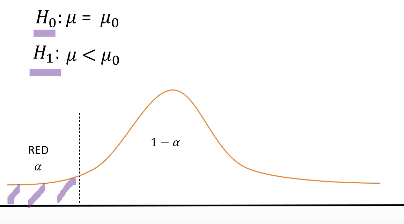
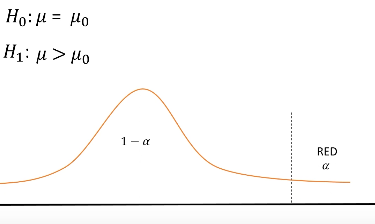
En basık grafik çizgisi turuncudur.

(standart sapması büyüktür)

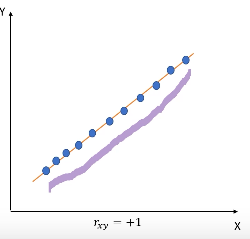
**Güven aralığı= örnek=>**

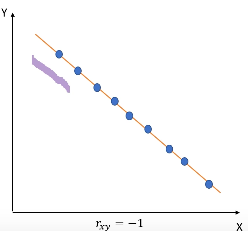
**Hipotez testleri=** güven aralığı saptana bilmesi için oluşturulan bir testtir. Hepotez testi sürecinde, örnek istatistiğine karşılık gelen değeri bulunmaya çalışan ana kütle parametsine uygun olup olmadığını saptanmasına çalışır.

Bir örneklem aracalığı ile çeşitli tahminler yapılır. Bu süreçte belirsizlikler ortaya çıkacaktır ve hata olasılıkları doğacaktır. Hipotez kurulduktan sonra bu tahmine ne derece güvenle bakıldığının bilinmesi gerekir.

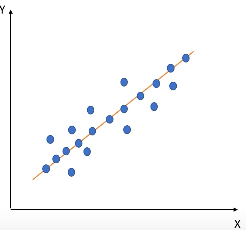


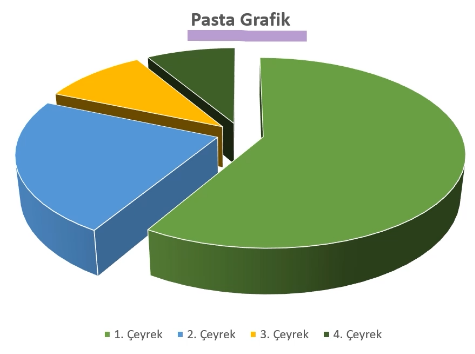
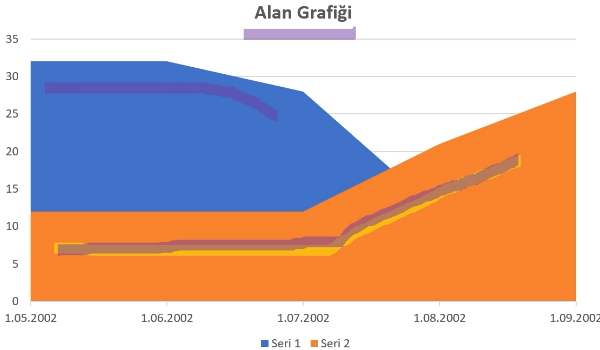
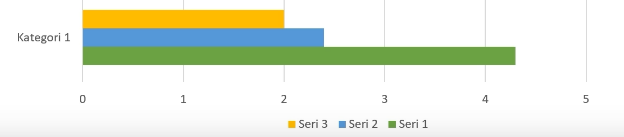
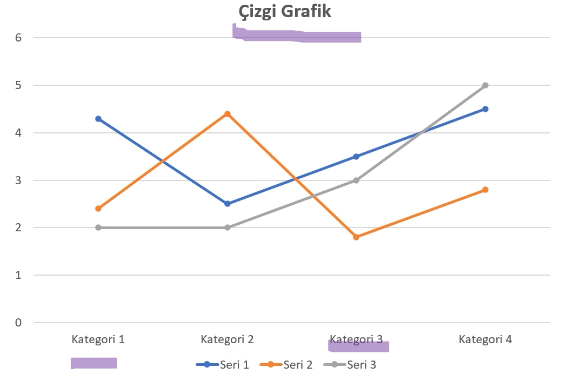
**Korelasyon(r)=** iki veya daha fazla değişkenin arasındaki doğrusal ilişkinin yönünü ve gücünü ifade eder. Bir değişkenin 2 ya da daha çok değişken ile olan ilişkisini **çoklu korelasyon** adı alır. Değişkenlerden birinin sabitlenip diğer değişkenlerle olan ilişkisi ise **kısmi korelasyon** denir.

****

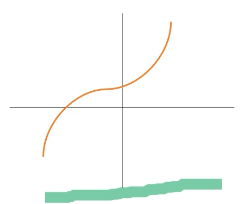
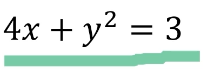
****

**Regresyon=** bir bağımlı değişken ile bağımsız değişkenler arasındaki ilişkisi belirler. Regreston testi bağımsız değişkenler değiştiğinde bağımlı değişkenlenin nasıl bir etki alacağını hesaplar. Değişkenler arasında neden sonuç ilişkisini test eder.

 Doğrusal regresyon=bir doğruda en uygun düzlemi bulmak için kullanılır



LİNEER CEBİR

**cebir**

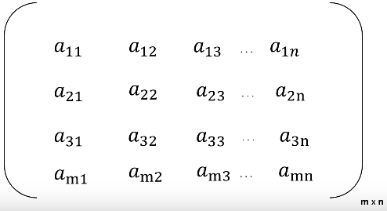
**lineer cebir=** Lineer olabilmek için

* 1.dereceden denklem olmak zorundadır.
* 2 bilinmeyen değişken birbiri

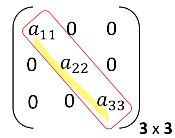
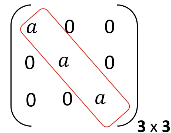
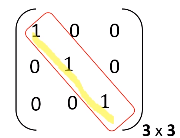
**Lineer cebir özellikleri**

* A1x+c1y=b1, a2x+c2y=b2 eğer paralel ise a1 ve a2 katsayıları işittir
* A1x+c1y=b1, a2x+c2y=b2 eğer kesişiyosa ise iki denklemi birbirine işetle
* A1x+c1y=b1, a2x+c2y=b2 eğer çakışıyosa a1/a2=c1/c2=b1/b2

**Matrisler**

****





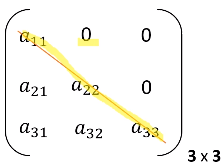
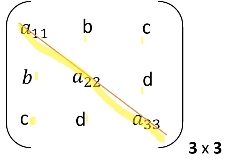
Satır matris=> 1 adet satırı vadır(m=1)

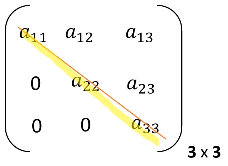
Sütun matris=>1 adet sütunu vardır(n=1)

Kare matris=>satır ve sütun sayısı aynı(m=n)

Sıfır matris=> içindeki tüm sayılar 0 ise

Köşegen matris=>köşegen dışındaki bulunan sayılar 0 ise

Skaler matris=>köşegen aynı sayı ve diğer sayılar\* 0 ise

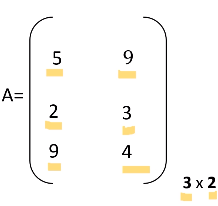
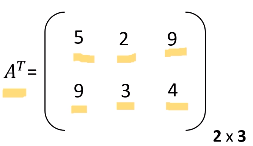
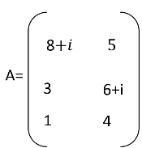
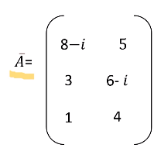
Birim matris=>köşegen 1, diğerleri 0 ise

Üst üçgen=>köşegenin üstü 0 harici ve alt tarafı 0 ise

Alt üçgen =>köşegenin altı 0 harici ve üst kısmı 0 ise

Simetrik matris=>karşılıklı gelen sütun ve sayıların eşit ise

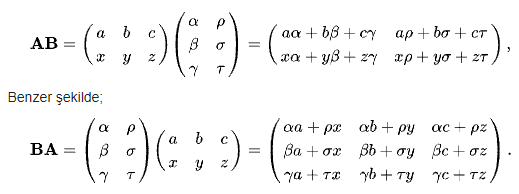
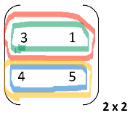
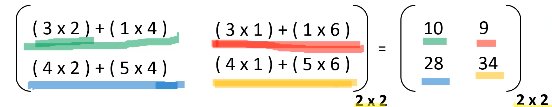
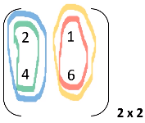
**Özellikleri**

* **İki matris eşitliği=>**iki matris eşit ise hem satır ve sütunları sayıları aynıdır hem o satır ve sütunlardaki değerler
* **Matrisin devriği(xT)=>**bir matrisin satırları sütun, sütunların satır olduğu durumdur.
* **matrisin eşleyeni( )=>**bilinmeyenlerin işaretlerini değiştir.

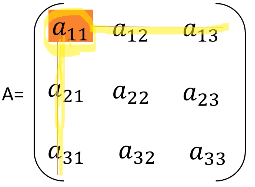
**Matriste işlemler**

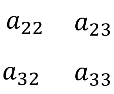
* Toplama ve çıkarmada Boyutları aynı olmak zorundadır
* Çarpma işleminde a matrisin sütunu b matrisin satır sayısına eşit olmalı
* Bir matris kat sayı ile çarpılmak istenirse matrisin tüm elemanları o katsayı ile çarpılır

Matrislerin çarpımı



Matrislerin eş çarpımı



Bir eleman seçilir. Seçilen elemanın bulunduğu satır ve sütun kapanır geriye kalan satılar ile yeni bir matris oluşturulur.

Minör nasıl hesaplanır =>içler çarpımı – dışlar çarpımı |M11|= =>

|1 3 5| |m11|=elemanının minörü

|4 2 8| | 2 8 | |m11|=(2\*9)-(8\*7)=-38 Aij=(-1)i+jx|Mij|

|6 7 9| | 7 9 |

Kofaktör hesaplama

Aij=(-1)i+jx|Mij|

Örk=

|1 3 5| |M11|=-38

|4 2 8| => A11=(-1)1+1x-38

|6 7 9| A11=-38

Matris indirgenmesi

1-satırlar

* Hi(x)=>i numaralı satırın x sayısıyla çarpmak

|4 3 5| |8 6 10|

A= |1 2 8| H1(2)= |1 2 8 |

|6 7 9| |4 7 9 |

* her hangi 2 satırı yer değiştirebiliriz=>Hxy

|4 3 5| |1 2 8|

A= |1 2 8| H12 = |4 3 5|

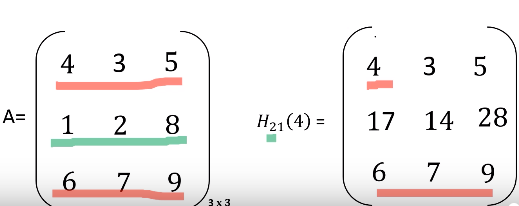
|6 7 9| |6 7 9|

* j numaralı satırı bir x sayısıyla çarpıp herhangi i satırını eklemek=>hij(x)

|4 3 5| |1 2 8 | not=1.satırı 2le çarp sonra 3.satırdakilere ekle

A= |1 2 8| H31(2) = |4 3 5 |

|6 7 9| |10 13 19|



2- sütun

J numaralı sütun elemanlarını bir x sayısıyla çarpmak=>Kj(x)

|4 3 5| |8 2 8|

A= |1 2 8| k1(2) = |2 3 5|

|6 7 9| |12 7 9|

j numaralı sürunu bir x sayısıyla çarpıp herhangi i sürununa eklemek=>Kij(x)

|4 3 5| |1 2 8 | not=1.sütunu 2le çarp sonra 3.satırdakilere ekle

A= |1 2 8| H31(2) = |4 3 5 |

|6 7 9| |6 7 21|

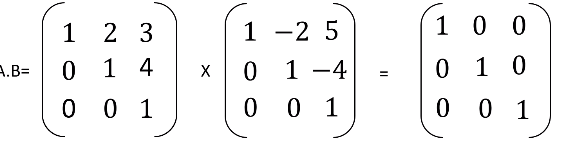
Determinant

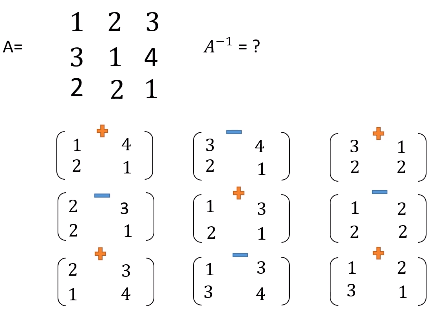
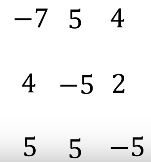
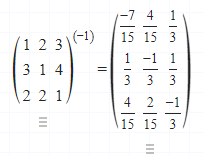
* 1x1 matrixsin determinantı içindeki sayıdır.
* 2x2 matrixsin determinantı köşegenlerinin çarğımının eksiğidir=>A=|a b|=>|A|= (ad)-(bc)

|c d|

Matrisin tersi

2 farklı matris çarpıldığında sonuç olarak birim matris çıkarsa bu iki matris birbirlerinin tersidir.



Bir matriksin tesini bulmak için:

|A|=15

**Vektörler**

**Skaler çarpım=>a.b=** |a|.|b|.cosX

**Vektörel çarpım=>a^b=** |a|.|b|.sinX

**R programlama**

R programı kurulum web site linki •https://www.r-project.org/ 2-5.Bölüm 2.Derste yer alan R

r studio •https://rstudio.com/products/rstudio/download

help(mean)-?mean

sqrt(x)=>karekök bulma-sqrt(2)-4

print(x)=>ekrana yazdırma

matrix(x(c(değerler),satırS,sütunS))>matriks oluşturma- print(matrix(c(1,0,2,4),2,2))

ls() => nesleleri sırala

rm(x)=>nesne siler

mode() => nesnenin veri türünü

mean(c(x,y,z))=>ortalamayı bulur

sd(c(x,y,z))=>standart sapma

median(c(x,y,z))=>median hesaplar

var(c(x,y,z))=>varyant hesaplar

başlangıçS:bitişS=>liste oluşturma

seq(from= başlangıçS, to=bitişS, by=artışS)

rep(sayı, kaç\_tane)=> sayıyı tekrar etme

“==,!=,<=,>=” =>karşılaştırma yapar doğruysa true yazar

1. Listeleri kaşılaştırsak tüm elemanları karşılaştırır

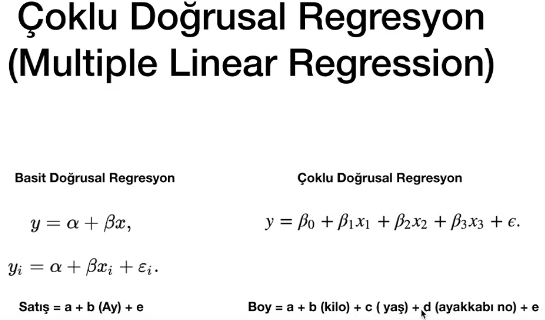
Any(liste==sayı)=>listede aranan sayı var mı

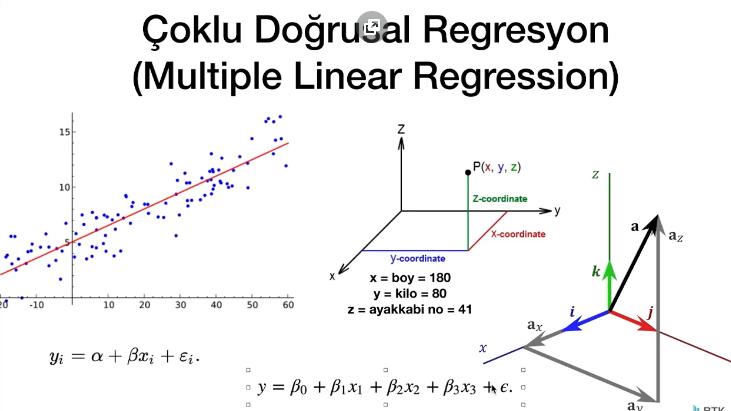
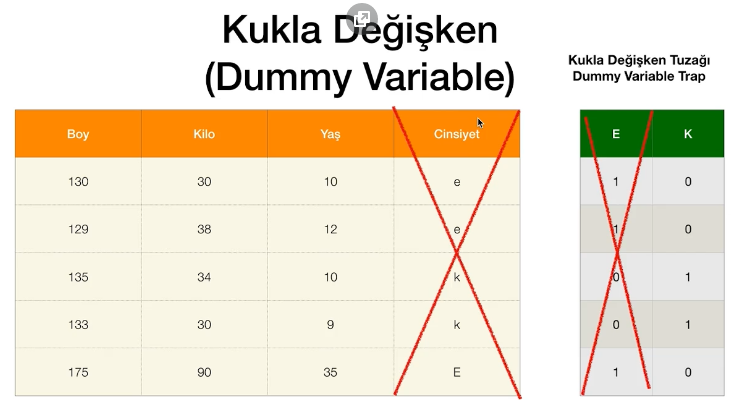
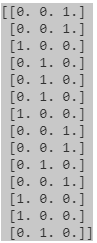
All(liste==sayı)=>listedeki tüm elemanlar aranan sayı mı

**Proje geliştirme**

* Proje belirlenmesi
* Verilerin elde edilmesi
* Veri hazırlama ve etiketleme:
* Modelleme
* Hiperparametre ayarlama
* Test

1. Roboflow
2. Colabeler
3. Label studio



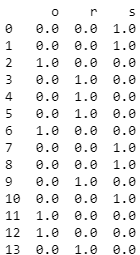
from sklearn.preprocessing import OneHotEncoder

c = veriler2.iloc[:,:1]

ohe = OneHotEncoder()

c=ohe.fit\_transform(c).toarray()

print(c)



havadurumu = pd.DataFrame(data = c, index = range(14), columns=['o','r','s'])

**p-value**

* H0: boş hipotezi, sıfır hipotezi
* H1: alternatif hipotez
* P- değeri: olasılık değeri(genelde 0.05)
  + P değeri küçükse h0 hatalı
  + P değeri büyükse h1 hatalı

**Farklı yaklaşımlar**

* Bütün değişkenleri dahil etme
* Geriye doğru eleme (backward elimination)
* İleri seçim (forward elimination)
* İki yönlü eleme (bidirection elimination)
* Skor karşılaşması (score comparison)