



---

# PROJE

---

Haftalık Rapor – 20.03.2021

20 MART 2021

KIRIKKALE ÜNİVERSİTESİ – BİLGİSAYAR MÜHENDİSLİĞİ İÖ

AHMET MUNGAN – 160255081

## İÇİNDEKİLER

ÖZET.....	2
İSTATİSTİĞİN VERİ BİLİMİ İLE YORUMLANMASI .....	3
Veri Biliminde Cebir Terimi.....	4
RSTUDIO İLE R DİLİNE GİRİŞ .....	5
ANAHTAR KELİMELER.....	8
REFERANS VE KAYNAKÇA .....	9

## ÖZET

İstatistiğin bileşenleri veri bilimi ile teknik olarak öğrenilmiş ve yorumlanmıştır. Bilgisayar bilimleri eğitimlerinde gösterilen derslerin veri bilimi ile alakası ve çözümlenmesi değerlendirilip öğrenilmiştir. Bunlara ek R dili için giriş yapıp temel bazı fonksiyonlar öğrenilmiştir. R diline hakim olmaya çalışılmıştır.

## İSTATİSTİĞİN VERİ BİLİMİ İLE YORUMLANMASI

Örneklemlerin kullanılması ana kütlenin bir kısmından seçilen veri setlerinin teknik olarak yorumlanması gerekir. Örneklem üzerinden veri bilimi uygulamalarının gerçekleştirilmesinin bazı nedenleri:

1. Ana kütleyi incelemenin zorluğu,
2. Amaca yönelik araştırmalarda kısıtlı bir zaman ile bulgular edinilmesi ve dolayısıyla veri üzerinde ciddi maliyet tasarruflarının meydana gelmesi.

İstatistiğin veri bilimi ile incelenmesi için gözlem birimi seçimi ve parametreler yer yer değişkenlik gösterebilir. Fakat günlük hayat istatistik tutumundan farklı ve sanal roller veri bilimi için atanabilir. Örneğin bir sesi dinleyen belirli özelliklerdeki bir insanın bu sese vereceği tepkiler ve yapacağı yorumlar sayısal parametreler genellikle taşımaz. Fakat sesi frekans sinyallerine göre ayrıştıran bir bilgisayar sayısal parametreler taşıyabilir. Bu sebepten ötürü aynı örneklemin farklı gözlem birimleri ile veri bilimi için farklı anlamlar ve parametreler taşıdığı söylenebilir.

Veri setlerinde değişim aralıkları ve dolayısıyla güven aralığı belirleme işlemleri sinyaller üzerinde mevcuttur. Örneğin; bir ses sinyalinin beklendik desibel aralığında<sup>1</sup> uyarak oluşturulduğu varsayılırsa, en yüksek ve en düşük desibel bilindiğine göre değişim aralığı hesaplanabilir. Dolayısıyla kendi örneklemimiz üzerinde bu değişim aralığına bağlı olarak güven aralığı/aralıkları yaratılabilir.

Bire bir sinyal eşlemede veriler gerçek halleri ile fonksiyon haline getirildiği için tüm bu istatistik bileşenlerini kullanmak avantajdır.<sup>2</sup> Örneğin; bir enstrüman ses kaydının oluşturduğu sinyallerde değişim aralığı matematiksel olarak küçük değerlerdir. Fakat akoru bozulmuş veya yanlış bir teknikle çalınan enstrüman genellikle kayıt zamanı tespit edilememektedir. Sinyallerinin işlenmesi sırasında

---

<sup>1</sup> Güven aralığı kastedilmiştir.

<sup>2</sup> Bire bir sinyal işleme ile ilgili yazılar için bkz. [Referans ve Kaynakça](#)

basıklık<sup>3</sup> adı verilen istatistik bileşeni bu bozuklukların tespitinde kolaylık sağlamaktadır. Bu sayede kulak ile duyulmayan fakat ses sinyallerinde oluşan bozukluklar güven aralığı dışına çıkan anlık sinyallerin grafikler üzerinden tespiti kolaylaşır.

Tüm bu istatistik bileşenleri olasılık hesabında da yardımcı olacağı gibi, daha hesaplanabilir veriler üretmeye de yardımcı olacaktır. Veriyi sadece arka planda veritabanlarında tutmak yerine daha sık kullanıp istatistiksel olarak incelenmesi gerekir.

### **Veri Biliminde Cebir Terimi**

Veri; istatistiğin bileşenleriyle birlikte yorumlanırsa ve ortaya bir fonksiyon koyulursa bu fonksiyonun sıfırı (denklemin kökü/kökleri) bulunması gerekebilir. Lineer veya lineer olmayan cebirsel ifadelerin çözümüne göre:

1. Lineer ise: Lineer cebir yöntemleri kullanılabilir,
2. Lineer değil ise: Nümerik analiz yöntemleri kullanılabilir.

Ayrıca veritabanından çekilen verilerin elde tutulması<sup>4</sup> için array tipleri kullanılabilir. Eğer 2 boyutlu array olarak tutulursa veriler lineer cebirin matris işlemleri kullanılabilir. Eğer 3 boyutlu array yapısı kullanılırsa burada x, y ve z eksenlerinde vektörler ile ifade edilebilir. Bu sebeple lineer yöntemler büyük veriler ile alakalı pratik ve hızlı çözümler sunabilmektedir. Bire bir bilgisayar bilimleriyle alakalı çözümler veri setleri özelinde çözüm üretmese bile bilinen yöntemler ile türetilen çözümler yararlı olabilmektedir.

---

<sup>3</sup> Wikipedia basıklık tanımı. Bkz. [Referans ve Kaynakça](#)

<sup>4</sup> Elde tutulması ile verilerin RAM'e geçici olarak yüklenmesi kastedilmiştir.

## RSTUDIO İLE R DİLİNE GİRİŞ

Ses filigranlamasına yönelik olarak veri biliminde mühim bir dil olan R dilini kavrayıp özelliklerine hakim olmak avantajdır. Syntax, işlev, çalışma ortamı, vb. gibi açılardan R dilinin incelenmesi gerekir.

Açık kaynak kodlu olan R dili syntax olarak Python diline benzemektedir. İlk kullanım için RStudio yardımıyla basit birkaç konum vasıtasıyla test edilmiştir.

*RS 1*

```
> 5+5
[1] 10
> 5+10*22
[1] 225
```

RS 1’de ilk uygulamalar ile basit matematiksel işlemler uygulanarak console çıkışları görülmüştür. Ayrıca sık kullanılan fonksiyonların nasıl kullanıldığına dair yardım dökümanlarına console satırından erişmek mümkündür.

*RS 2*

```
> help("mean")
```

*RS 3*

```
mean {base} R Documentation
Arithmetic Mean
Description
Generic function for the (trimmed) arithmetic mean.

Usage
mean(x, ...)

## Default S3 method:
mean(x, trim = 0, na.rm = FALSE, ...)

Arguments
x
An R object. Currently there are methods for numeric/logical vectors
and date, date-time and time interval objects. Complex vectors are
allowed for trim = 0, only.

trim
the fraction (0 to 0.5) of observations to be trimmed from each end
of x before the mean is computed. Values of trim outside that range
are taken as the nearest endpoint.

na.rm
a logical value indicating whether NA values should be stripped
before the computation proceeds.

...
further arguments passed to or from other methods.
```

RS 2'deki komut console için yazılırsa; RS 3'teki çıktı elde edilir. Bu komut aritmetik ortalama hesaplayıcı fonksiyonu içerir. RS 3'teki çıktı özel bir help alanında gösterilir ve bu gösterimde faydalı bilgilerin tamamı mevcuttur. Help kısmı kullanılarak R dili öğrenileceği için resmi kaynaklar kullanılmış olacaktır. Ayrıca help kısmından dökümanlara erişildiği gibi dökümanına erişilmek istenen fonksiyonların başına "?" konarak aynı şekilde help kısmına ulaşılabilir.

RS 4

```
> ?assign  
> ?pnorm  
> ?linearizeMlist
```

RS 4'teki üç temel komut sırasıyla:

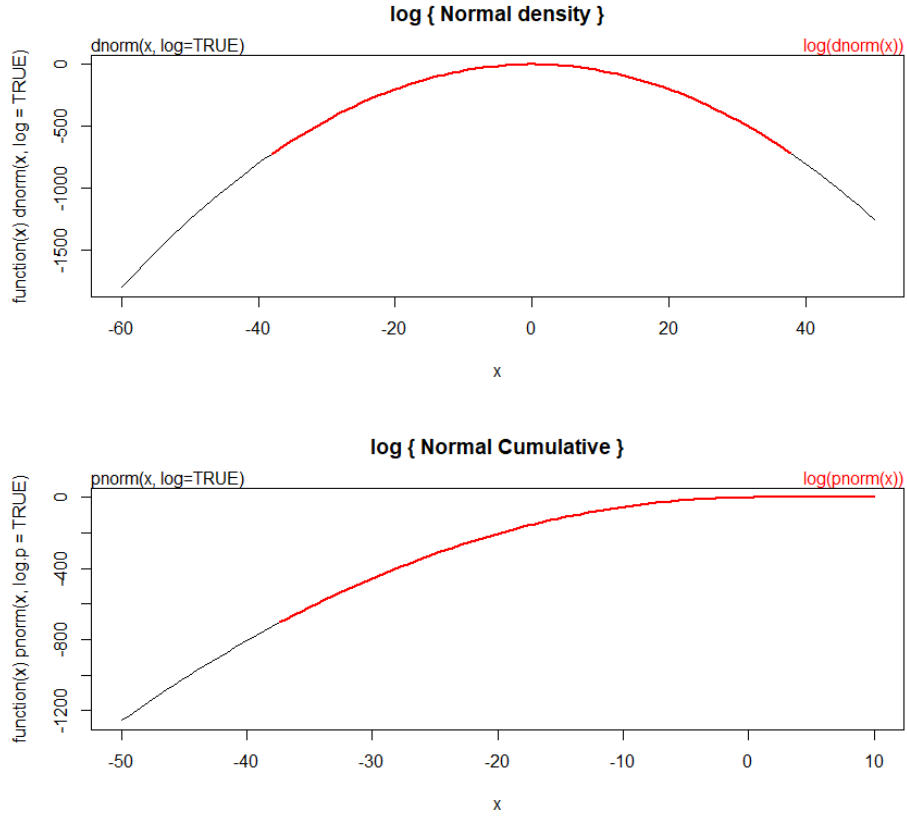
1. Nasıl atamalar yapılacağını ve syntax'ını gösteren komuttur.
2. Olasılık hesaplamalarında sıkça kullanılan normal dağılım fonksiyonunun nasıl kullanılacağını ve hangi parametreleri aldığını gösteren komuttur.
3. Lineer metot listesinin tutulduğu bir fonksiyonun nasıl kullanıldığını gösteren komuttur.

RStudio; dökümantasyon sağlamada yardımcı olduğu gibi, örnekler üzerinden kullanımın test edilmesine de imkan sağlar. Bu sebeple kendi sayılarımızı vermeden evvel doğruluğu konusunda R çekirdeğini deneme şansımız olabilmektedir. Örnek olması açısından normal dağılım aşağıda incelenmiştir.

RS 5

```
> example("pnorm")  
pnorm> require(graphics)  
pnorm> dnorm(0) == 1/sqrt(2*pi)  
[1] TRUE  
pnorm> dnorm(1) == exp(-1/2)/sqrt(2*pi)  
[1] TRUE  
pnorm> dnorm(1) == 1/sqrt(2*pi*exp(1))  
[1] TRUE  
pnorm> ## Using "log = TRUE" for an extended range :  
pnorm> par(mfrow = c(2,1))  
pnorm> plot(function(x) dnorm(x, log = TRUE), -60, 50,  
pnorm+      main = "log { Normal density }")  
Sonraki çizimi görmek için <Enter>'a basın:
```

RS 5'te görüldüğü üzere verilen komut ile normal dağılım fonksiyonunun örneği R dili tarafından oluşturulmuştur. Burada yabancı olunan birkaç fonksiyon ile alakalı helper vasıtasıyla bilgi alınmıştır. Gerekli olan kısımlar gösterilmeye gayret edilmiştir. RS 5'teki fonksiyon örneği tam anlamıyla anlaşıldıktan sonra enter tuşu ile birlikte bir çizim geliyor.



Plots 1

Plots 1'de RS 5'teki normal dağılım örneğinde verilmiş değerler ile grafiksel çıktısı görülmektedir. Olasılık hesaplamalarına göre tamamı doğru olan bu çizimler yine RStudio'nun sağladığı bir hizmettir. Plots sekmesinde kurulan fonksiyonun grafiksel çıktısı görüntülenebilir, yakınlaştırılıp uzaklaştırılabilir, export<sup>5</sup> edilebilir.

---

<sup>5</sup> RStudio'nun bazı sürümlerinde plots sekmesi export edilemeyebiliyor. Fakat bahsi geçen RStudio en güncel sürümüdür. (1.4.1106 sürümü)



## ANAHTAR KELİMELELER

- ***RS X*** : X. RStudio kodu/çıktısı.
- ***Plots X*** : X. RStudio Plots sekmesi çıktısı.
- **Normal Dağılım**: Olasılık ve İstatistik hesaplamalarında Gauss dağılımı olarak da bilinen birçok alanda pratik uygulaması olan önemli bir sürekli olasılık dağılım ailesidir.

## REFERANS VE KAYNAKÇA

- Bingöl Üniversitesi Veri Çözümleme ve Yorumlama Üzerine Araştırma Dökümanı – Link için [tıklayınız](#).
- Wikipedia Ses Tanımı – Link için [tıklayınız](#).
- Veri Bilimi İçin İstatistik Üzerine Makale – Link için [tıklayınız](#).
- Wikipedia Basıklık Tanımı – Link için [tıklayınız](#).
- Sinyal İşleme Üzerine Proje İncelemesi – Link için [tıklayınız](#).
- R Çekirdek – Link için [tıklayınız](#).
- RStudio – Link için [tıklayınız](#).
- Normal Dağılım – Link için [tıklayınız](#).