

R Diliyle İstatistik Uygulamaları

Ergül DEMİR

3. Baskı





Dr. Ergül DEMİR

R DİLİYLE İSTATİSTİK UYGULAMALARI

ISBN 978-605-241-885-7

Kitap içeriğinin tüm sorumluluğu yazarına aittir.

© 2023, PEGEM AKADEMI

Bu kitabın basım, yayım ve satış hakları Pegem Akademi Yay. Eğt. Dan. Hizm. Tic. AŞ'ye aittir. Anılan kuruluşun izni alınmadan kitabın tümü ya da bölümleri, kapak tasarımı; mekanik, elektronik, fotokopi, manyetik kayıt ya da başka yöntemlerle çoğaltılamaz, basılamaz ve dağıtılamaz. Bu kitap, T.C. Kültür ve Turizm Bakanlığı bandrolü ile satılmaktadır. Okuyucularımızın bandrolü olmayan kitaplar hakkında yayınevimize bilgi vermesini ve bandrolsüz yayınları satın almamasını diliyoruz.

Pegem Akademi Yayıncılık, 1998 yılından bugüne uluslararası düzeyde düzenli faaliyet yürüten **uluslararası akademik bir yayınevi**dir. Yayımladığı kitaplar; Yükseköğretim Kurulunca tanınan yükseköğretim kurumlarının kataloglarında yer almaktadır. Dünyadaki en büyük çevrimiçi kamu erişim kataloğu olan **WorldCat** ve ayrıca Türkiye'de kurulan **Turcademy.com** tarafından yayınları taranmaktadır, indekslenmektedir. Aynı alanda farklı yazarlara ait 1000'in üzerinde yayını bulunmaktadır. Pegem Akademi Yayınları ile ilgili detaylı bilgilere <http://pegem.net> adresinden ulaşılabilir.

1. Baskı: Eylül 2019, Ankara

3. Baskı: Ekim 2023, Ankara

Yayın-Proje: Ferdi Akkaya

Dizgi-Grafik Tasarım: Müge Kuyrukcu

Kapak Tasarımı: Pegem Akademi

Baskı: Sonçağ Yayıncılık Matbaacılık Reklam San Tic. Ltd. Şti.

İstanbul Cad. İstanbul Çarşısı 48/48 İskitler/Ankara

Tel: (0312) 341 36 67

Yayıncı Sertifika No: 51818

Matbaa Sertifika No: 47865

İletişim

Macun Mah. 204. Cad. No: 141/A-33 Yenimahalle/ANKARA

Yayınevi: 0312 430 67 50

Dağıtım: 0312 434 54 24

Hazırlık Kursları: 0312 419 05 60

İnternet: www.pegem.net

E-ileti: pegem@pegem.net

WhatsApp Hattı: 0538 594 92 40

İÇİNDEKİLER

Giriş ve Ön Bilgiler	v
R Dili	v
Amaç, Kapsam ve Sınırlılıklar	viii
Hedef Kitle	x
Kullanım Rehberi	x
İkinci Baskıya Birkaç Söz ve Teşekkür	xi
Yazar Hakkında ve İletişim	xi
1. R Ortamında Veri Yapıları	1
1.1.Tek Elemanlı Vektörler (Scalars)	1
1.2.Vektörler	1
1.3.Faktörler (Factors)	6
1.4.Matrisler	11
1.5.Üç Boyutlu Matrisler (Arrays)	14
1.6.Verit Setleri (Data Frames)	16
1.7.Listeler (Lists)	20
2. Veri Seti Yükleme, Düzenleme ve Çıkarma	25
2.1.R Ortamına Veri Aktarma (Import)	26
2.2.R Ortamında Veri Düzenleme	34
2.3.R Ortamından Veri Çıkarma (Export)	41
3. Betimsel İstatistikler	43
3.1.Ortalama	43
3.2.Standart Sapma ve Varyans	45
3.3.Medyan, Çeyreklikler ve Yüzdelikler	48
3.4.Medyan Mutlak Sapması, Çeyrek Açıklık ve Çeyrek Sapma	50
3.5.Diğer Betimsel İstatistikler	51
3.6.Örnek Fonksiyon Çalışması	52
3.7.R Ortamında Bir Fonksiyonun Veri Yapılarına Uygulanması	54
4. Frekans Tabloları ve Çapraz Tablolar	61
4.1.Frekans Tabloları	62
4.2.Çapraz Tablolar	65
5. Grafikler	69
5.1.R’da Grafik Fonksiyonları ve Parametreleri	70
5.2.Kategorik Verilere Uygulanan Grafikler	71
5.3.Sürekli Verilere Uygulanan Grafikler	84
5.4.Fonksiyon Grafikleri	100
6. Olasılık Yoğunluk Dağılımları	111
6.1.Normal Dağılım	112
6.2.Student’in t Dağılımı	127
6.3.F Dağılımı	135
6.4.Ki-Kare Dağılımı	139
6.5.Tek Biçimli Dağılım	143
6.6.Binom Dağılımı	145
6.7.Poisson Dağılımı	152

7. Merkezi Limit Teoremi ve Güven Aralıkları	157
7.1.Evren Ortalamasının Tahmini	158
7.2.Örneklem Ortalamaları Arasındaki Farkın Tahmini	165
7.3. Bootstrap Yöntemi ile Yeniden Örnekleme	168
8. Normallik ve Homojenlik	175
8.1.Örneklem Dağılımının Normalliği	176
8.2.Örneklem Varyanslarının Homojenliği	190
9. Hipotez Testlerinin Sınıflandırılması	201
10. Student'in t Testleri	213
10.1. Tek Örneklem t Testi	216
10.2. Eşleştirilmiş Örneklem t Testi	229
10.3. Bağımsız Örneklem t Testi	238
11. Varyans Analizi	249
11.1.Tek-Yönlü Varyans Analizi	254
11.2.İki-Yönlü Varyans Analizi	270
11.3.Çok Değişkenli Varyans Analizi	278
11.4.Tekrarlı Ölçümler Varyans Analizi	287
12. Korelasyon	299
12.1.Pearson'un Momentler Çarpımı Korelasyon Katsayısı	300
12.2.Kısmi Korelasyon	310
12.3.Nokta-Çift Serili ve Çift Serili Korelasyon Katsayıları	313
12.4.Fi (ϕ) ve Tetrakorik Korelasyon Katsayıları	320
12.5.Spearman'ın Ro (ρ_s) ve Kendall'ın Tau (τ) Katsayıları	324
12.6.Cohen'in Kappa ve Kendall'ın Konkordans Katsayıları	332
13. Regresyon	337
13.1.Basit Doğrusal Regresyon	338
13.2.Polinom Regresyon	368
13.3.Çoklu Doğrusal Regresyon	379
13.4.Doğrusal Regresyon Modeli Olarak ANOVA	394
13.5.Planlı İkili Karşılaştırmalar	398
13.6.Doğrusal Regresyon Modeli Olarak ANCOVA	406
14. χ^2 Testi	415
14.1. Varsayımlar ve Gözlenen Frekansların Büyüklüğü	421
14.2. Etki Büyüklüğü	429
14.3. McNemar Testi	432
14.4. Cochran-Mantel-Haenszel Testi	434
15. Parametrik Olmayan Testler	441
15.1. Wilcoxon Sıra Toplamları Testi	443
15.2. Wilcoxon İşaretli Sıralar Testi	454
15.3. Kruskal-Wallis H Testi	461
15.4. Friedman'ın ANOVA Testi	466
Kaynaklar	473
Ekler	477
Dizin	483

Giriş ve Ön Bilgiler

R Dili

R dili, kendi kendini fazlasıyla anlatan örnek bir uluslararası çabadır. Gerekli hemen her türlü bilgiye ve dokümana <https://www.r-project.org/> bağlantısını kullanarak bizzat R'in kendi proje sayfasından ulaşabilir ve bunları ayrıntılarıyla çalışabilirsiniz (R core team, 2020). R'in ne olduğunu anlamak ve ötesinde R'i bilgisayarınıza indirip kullanmaya başlamak için başkaca bir kaynağa ihtiyacınız olacağını düşünmüyorum. Bununla birlikte, genel bir bilgilendirme ve tanıtım amacıyla bazı önemli vurguları ve açıklamaları yapmamız beklenebilir.

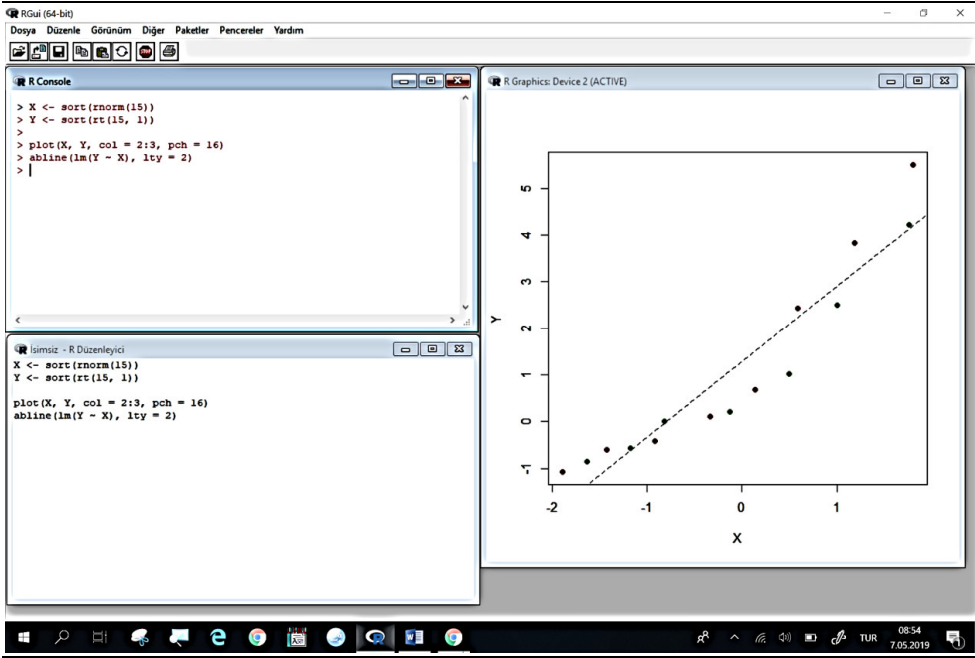
R; istatistiksel hesaplamalar ve grafiksel gösterimler için GNU sertifikalandırma sistemine göre düzenlenmiş, web ve bulut tabanlı, açık kaynak kodlu ve ücretsiz bir programlama dilidir. John Chambers tarafından 1976 yılında geliştirilmiş S dilinin genişletilmiş bir uygulaması olarak değerlendirilir. İlk geliştiricileri Ross Ihaka ve Robert Gentleman'dır. "R" isminin kaynağı anlaşılyordur sanıyorum.

R, 1992 yılında bir GNU projesi olarak tasarlanmıştır ve ilk defa 1993 yılında *s-news* e-mail listesinde duyurulmuştur. GNU lisansını alarak kullanıma açılan ilk sürümü 1995 yılındadır. R'in uluslararası düzeyde yaygınlaşmasında bu dönemdeki stratejik düzenlemelerin etkisi çoktur. Robert Ihaka başta olmak üzere Robert Gentleman ve Martin Machler, R projesi kapsamında "çekirdek takım (core team)" olarak tanımlı geniş bir uzman grubunu bir araya getirmiştir. Bu grup 1997'de resmi olarak tanıtıldığında 20 uzmandan oluşuyordu. Sonrasında 2017 yılına kadar bu çekirdek takıma 5 uzman daha eklenmiştir. Çoğunluğu profesörlerden oluşan bu uzman grup, R'in kullanıcılara kapalı temel kodlarını ve paketlerini düzenleme ve geliştirme görevini üstlenmiştir. Bunun yanı sıra 60'ın üzerinde bir uzman grubu daha "katkı sağlayanlar (contributors)" olarak sürece dâhil olmuştur. Bu grupta yer alan uzmanlar da kendi kodlarını, düzeltmelerini ve dokümanlarını paylaşarak R diline katkı sağlamışlardır.

R, bulut tabanlı bir yapıya sahiptir. İnternet üzerinde "Comprehensive R Archive Network – CRAN" olarak tanımlı bir merkez arşive sahiptir. Bu merkez arşiv Avusturya'da Viyana Üniversitesindedir ve "ayna (mirror)" olarak tanımlı diğer merkezler üzerinden erişilebilir biçimde paylaşılmaktadır. R'in, 3.6.0 sürümünün kullanıma sunulduğu 26.04.2019 tarihi itibarıyla, merkez arşivi dışında, 46 ülkede toplam 97 ayna arşivi bulunmaktadır. Böylece farklı ülkelerdeki kullanıcıların R ortamına ve araçlarına hızlı erişimleri sağlanmış, R dilinin kullanımı ve R ortamında paylaşımlar kolaylaştırılmıştır.

R; UNIX, Windows ve MacOS işletim sistemlerinde çalışabilmektedir. Kişisel bilgisayarlara indirilip yüklenmesi son derece kolaydır ve bu işlemler ayna merkezler sayesinde çok hızlı bir şekilde yapılabilir. Bunun için <https://www.r-project.org/> bağlantısındaki "indirme (download)" yönergelerini takip etmeniz yeterlidir. Bu yönergelerle göre öncelikle bulunduğunuz ülkeye ait ya da en yakın ayna arşivi seçersiniz, sonrasında bilgisayarınızın işletim sistemine uygun indirme sekmesini işaretlersiniz, gerekli kurulum dosyası bilgisayarınıza indirilir, bu dosyayı çalıştırarak kurulum işlemini tamamlarsınız.

R'in bilgisayarınıza indirip kuracağınız temel yazılımı "R GUI – R Graphical User Interface" olarak tanımlıdır. Ayrıca RGUI'ye eklemleyebileceğiniz "RStudio" olarak isimlendirilmiş bir geliştirme ortamı (IDE - integrated development environment) da bulunmaktadır. R yazılımını bilgisayarınızda açtığınızda çok sade bir görünümde olduğunu farkedersiniz. "R console" olarak tanımlı kod giriş penceresi kendiliğinden aktif hale gelir. Bu çok sade bir kullanım arayüzüdür, yaygın kullanımı olan hazır paket ve programlardan farklıdır. R ortamındaki tüm çalışmalar bu konsol penceresi üzerinde, kod ve fonksiyon yazımıyla yapılır. Betik (script) ve grafiksel gösterimler için ayrıca pencereler açılır. Aşağıda basit bir örnek ekran görüntüsü yer alıyor.



Gördüğümüz gibi R dili, kendine özgü semboller ve biçimler içermektedir. Bir "dil" olarak tanımlanması boşuna değil. Bildiğiniz gibi, paket ve programların tamamının geri planında, çoğunlukla farkında olmadığımız bir yazılım diliyle düzenlenmiş kapsamlı işlemler ve süreçler tanımlıdır. İşte R, bu geri planı ve zemini oluşturan yazılım dillerinden birisidir. Yani çalışmalarınızda R kullanırsanız, istatistiksel işlemlerinizin geri planına hâkim olursunuz, süreçlere müdahale edebilir, hatta tüm süreçleri kendiniz düzenleyebilirsiniz. Hazır paket ve programların, ancak tanımlı biçimlerini ve şablonlarını kullanma sınırlılığını büyük ölçüde aşmış olursunuz. Yani R, size özgürlük sağlar. Elbette R'in bir araç olduğunu unutmamak gerekir. Gerekli istatistik bilgi ve becerilerine sahip olmayanların bu aracı etkin bir şekilde kullanabilmeleri pek mümkün değildir. Çoğu durumda "iyi" düzeyde istatistik bilgi ve becerileri de yeterli değildir. En az bu düzeyde alan uzmanlığı ya da yeterliği de gereklidir. Aklımızda bulunsun!

R ortamının çok işlevsel bir yazılım mimarisi vardır. İç içe yuvalanmış ortamlar biçiminde yapılandırılmıştır. En içte R'ın temel içeriğini taşıyan ortam (R base) ve en dışta kullanıcının aktif olarak çalıştığı ortam (global environment) yer almaktadır. Bu iki ortam arasında kendiliğinden aktif hale geçen R temel paketleri ve sonradan indirilip aktif hale getirilen diğer paketler bulunur. R yazılımını bilgisayarınıza indirip yükledikten sonra, R konsol üzerinde `search()` komutunu çalıştırarak bu ortamları ve paketleri görebilirsiniz.

```
search()
[1] ".GlobalEnv"          "package:stats"      "package:graphics"
[4] "package:grDevices"   "package:utils"      "package:datasets"
[7] "package:methods"     "Autoloads"          "package:base"
```

Kullanıcı, en içteki temel içeriği taşıyan ortama (R base) müdahale edemez. Buraya müdahale yetkisi sadece R çekirdek ekibindedir. Aradaki kendiliğinden aktif hale geçen temel paketler, pasif hale getirilebilir. Fakat bu durumda birçok işlem yapılamaz hale gelir. Bu temel paketlerin yanı sıra, katkı sağlayanlar tarafından hazırlanan birçok paket program, R ortamında erişilebilir biçimde arşivlenmektedir. R'ın 3.6.0 sürümünün kullanıma sunulduğu 26.04.2019 tarihi itibarıyla R ortamında kullanılabilecek toplam 14.161 hazır paket bulunmaktadır. Sayısı sürekli artan bu paketlerin güncel listesine basitçe `install.packages()` komutuyla erişilebilir. Aynı komutun içerisine paket ismi tanımlanarak `install.packages("paket adı")` komutuyla, paket indirme işlemi de yapılabilir. İndirilen paketler, kullanıcının bilgisayarında tanımlı “çalışma dizinine (working directory)” kaydedilir. Bu paketler `library(paket adı)` komutuyla aktif, `detach(...)` komutuyla da pasif hale getirilir. Yüklenen ve aktif hale getirilen her paket, R'ın temel içeriğini taşıyan ortam ve aktif olarak üzerinde çalışılan ortam arasında konumlandırılır. R ortamında kullanılabilir olan her paketin ayrıntılı kılavuzları, kuramsal içeriği, kullanım bilgileri ve uygulama örnekleri arşiv dokümanları içerisinde bulunmaktadır. Tüm paketlere yönelik bu dokümanlara `help(package = "paket adı")` fonksiyonuyla kolaylıkla erişilebilmektedir.

Fark ettiğiniz gibi daha giriş kısmında bazı kod ve fonksiyon kullanımlarından bahsetmeye başladık. Bu kapsamda hemen belirtelim; R'ın güçlü yönlerinden biri, son derece kullanıcı dostu olmasıdır. Bu kitapta yüzlercesini göreceğiniz kod ve fonksiyonları, R'ın temel mantığını ve işleyişini genel olarak kavradıktan sonra, biraz da alanın temel terminolojisine hâkimseniz kolaylıkla tahmin edebilirsiniz. Ayrıca R ortamı, olası hatalı ya da eksik kod ve fonksiyon tanımlamalarında, etkili uyarılar ve olası çözümler de sunmaktadır. Bunun da ötesinde çok geniş kapsamlı bir destek mekanizmasına sahiptir. R ortamının yardım sayfaları ve kullanıcı kılavuzları çok zengin içeriktedir. Bu kapsamda `help()` fonksiyonu, `?` ve `??` ön tanımlamaları, gerekli yardım sayfalarına kolaylıkla ulaşılmasını sağlar. Her bir pakete yönelik ayrıntılı kullanım kılavuzları bulunmaktadır. Ayrıca açık kaynak kodlu olmasından dolayı R ortamında kullanılan tüm fonksiyonların kod dizimine doğrudan ulaşmak da mümkündür.

Sonuç olarak başladığımız noktaya dönmüş oluyoruz; R kendi kendisini fazlasıyla tanıtan ve anlatan bir yazılım dili. Dahası hayranlık uyandıracak düzeyde kullanıcı dostu, çok geniş kapsamlı destek ve yardım dokümanlarını da içeriyor. Dolayısıyla R dilini öğrenmek ve istatistiksel çalışmalarınızda kullanmak, sadece gerekli emek ve çabayı göstermenize kalıyor.

Peki, aşına olduğunuz ve belki de yıllardır kullandığınız hazır ve güçlü istatistiksel paketler ve programlar varken niçin R gibi bir yazılımı öğrenmeye ve kullanmaya ihtiyacınız olsun? Bu tür bir soruya Crawley'in (2013) yanıtını samimi ve vurgularıyla önemli bulurum. Bu yanıtın kapsamını biraz daha genişleterek benzer bir yanıt verelim. Sınırlı bir istatistiksel araç kullanımı size yetiyorsa, alandaki güncel gelişmeleri takip etmeye fazlaca ihtiyaç duymuyorsanız ve mevcut durumunuzdan menunsanız, haklısınız, R gibi bir dili öğrenmeye çaba harcamayabilirsiniz. Fakat bir gerçeğin farkına varalım; "yapay zekâ (artificial intelligence)", "sanal gerçeklik (virtual reality)", "artırılmış gerçeklik (augmented reality)", "makine öğrenmesi (machine learning)", "robotik (robotics)" gibi uygulamalarla kendini gösteren teknolojik sıçrama, evrendeki konumumuzun ve görelî kontrolümüzün sınırlarını zorluyor. Bu sıçramada geleneksel istatistik çoktan "veri bilimine (data science)" evrildi ya da en azından evrilme sürecine girdi. Artık verilerle uğraşmak, özel ve disiplinler arası bir bilim alanı. Bireysellik ve bireysellik biçim değişti. "İletişimli (communicative)" ve "işbirlikli (collaborative)" çalışma ve üretme modellerine akış hızlandı ve kaçınılmaz oldu. Şimdi bir de etrafınıza ve alanınıza bir bakınız, araştırmacıların ve bilim insanlarının birçoğu çoktan R dili gibi yazılımları kullanmaya başladılar. Geline nokta milyonlarca kullanıcı olan devasa bir R ailesi oluşmuş durumda. Bu ailenin üyeleri, dünyanın her köşesinden, her ülkesinden, her ırkından, her cinsiyetinden, her kesiminden ve düzeyinden. R, en azından şimdilik, geleneksel anlamda bir ayırım gözetmeksizin birlikte üretmek isteyenleri bir araya getiriyor. Bu devasa ailenin üyeleri çalışmalarını R tabanlı yürütüyor, R ortamında geliştiriyor ve güncelliyor, R ortamında raporlaştırıyor, R ortamında paylaşıyor ve hatta R ortamında yayımlıyor. Kısaca, günceli yakalamak ve alanınızda etkin olmak istiyorsanız, bu sürecin bir parçası olmayı ciddi biçimde düşünmelisiniz.

Amaç, Kapsam ve Sınırlılıklar

Amacım, istatistiği ve ötesinde veri biliminin doğasını yeniden öğrenmekti. En iyi bildiğimi düşündüğüm kavramlara, kuramlara, teoremlere ve modellere yeniden ve en baştan çalıştım. İlk kuramcılarının çalışmalarını, bu çalışmaları alıp başkaca çözümler geliştirenlerin çalışmalarını tekrar tekrar inceleme ve dahası bunlar üzerinde uygulamalı çalışma fırsatı buldum. Bu sıradışı zihinlerin, nasıl bir bütün oluşturduğunu ve bu bütünün insanlığın evrilmesini nasıl şekillendirebildiğini gördüm. Eksiklerimin, kavram yanlışlarımlarım, hatalarımlarım en azından bir kısmını fark edebildiğimi düşünüyorum. Hala birçoğunun olduğundan ise kuşku yok. Sonuçta sınırlılıklarımı görerek görece özgürleştiğimi düşünüyorum. Daha ötesinde, problemleri görmek ve bunlara sistematik çözümler üretmek için çabalayan devasa bir ailenin bir parçası olduğumu hissediyorum. Bu kitapta da sadece bu yeniden öğrenme ve görelî özgürleşme sürecimi, bu süreçte parçası olduğumu hissettiğim aileyi sizlerle paylaşıyorum. Bunun ötesinde birtakım iddialı amaçların anlamsız olduğuna inanıyorum!

Bu kitapta, istatistiğin temel yaklaşımlarını ve yaygın kullanımı olan istatistiksel modelleri, R uygulamalarıyla birlikte derinlemesine çalışma imkânı bulabileceğinizi düşünüyorum. İlk iki bölüm, temel düzeyde R kullanımı için gerekli başlangıç becerilerine yöneliktir. Bu bölümlerdeki açıklama ve örnek uygulamaların sistematigi, aynı zamanda, daha ileri düzey öğrenmelerine bir hareket noktası oluşturması planlanarak belirlenmiş ve bölüm içerikleri buna göre düzenlenmiştir. Sonraki üç bölüm, verilerin betimlenmesi üzerinedir. Betimsel istatistikler,

frekans tabloları, çapraz tablolar ve grafikler kullanılarak yürütülen betimleme çalışmaları, verilerle çalışmanın başlangıç aşamalarıdır. Verilerin yapısını ve karakteristiğini tanıyabilmek ve bunlara hâkim olabilmek için öncelikle bu tür betimsel çalışmaların yapılması, kritik öneme sahiptir. Sonraki üç bölüm, yazarken ve çalışırken çok keyif aldığım, geleneksel istatistiğin temelini oluşturan “iyi” tanımlı hipotetik modelleri, merkezi limit teoremini, normallik ve homojenlik kavramlarını tartışmaya ve açıklamaya yöneliktir. R’in bu bölümlerde sağladığı hareket alanından çok memnun olduğumu özellikle belirtmek istiyorum. Kuramsal bilgilerimi uygulamaya dönüştürme ve uygulamalarla test etme fırsatı bulmak çok keyifli oldu. Dokuzuncu bölüm, sonraki bölümlere bir hazırlık aşaması. İstatistiksel çözüm üretme sürecinin tipik temsilcisi olan hipotez testinin ne olduğunu ve her biri ayrı bir istatistikel model olan hipotez testlerinin nasıl sınıflandırılabileceğini tartıştığımız bir bölüm. Hipotez kurma ve test etme mantığının anlaşılması, verilerle çalışmada önemlidir. Bunun, alanın felsefi temeli olduğunu ifade edebiliriz. Bundan sonraki bölümlerde yaygın kullanımı olan hipotez testleri, ayrı birer istatistiksel model olarak ele alınmıştır. Bu modeller; kullanım durumları, sınırlılıkları, temel algoritmaları, çıktıları, değerlendirme ölçüt ve biçimleriyle açıklanmıştır. Bu açıklamalar uygulama örnekleriyle desteklenmiş ve zenginleştirilmiştir. Yaygın kullanımı olan modeller dikkate alınarak olabildiğince geniş bir kapsam tanımlanmıştır. Tek değişkenli ve çok değişkenli modeller, bağımsız örneklemeler ve eşleştirilmiş örneklemeler içeren modeller, kısmen deneysel modeller, parametrik ve parametrik olmayan modeller üzerinde çalışılmıştır.

İstatistik, günümüzde “veri bilimi (data science)” olarak tanımlanan bambaşka boyutlara evrilme sürecinde. Yüzlerce yıllık bilgi birikiminin yanı sıra bilgi ve iletişim teknolojilerindeki sıçramalar, bu evrilme sürecinde belirleyici rol oynuyor. Artık paket program kullanabilme becerilerinin ötesinde bunların geri planında çalışan yazılım dilinin kullanılmasına ve böylece sürece güçlü bir şekilde hâkim olunmasına yönelik beceriler önem kazanmış durumda. Bu noktada R dili ve benzer araçlar ön plana çıkıyor ve kaçınılmaz bir gerekliliği tanımlıyor. Bu kitapta R dilinin bir araç olarak kullanılması tercihi, temelde bu gerekliliğe dayalıdır. Vurguya dikkat ediniz. R, bir araçtır. Bu aracın etkin bir şekilde kullanılabilmesi, gerekli istatistiksel bilgi ve becerilere sahip olmaya, daha önemlisi gerekli alan uzmanlığına sahip olmaya bağlıdır. Bunu açıkça ifade etmiş olalım.

Bu kitabın belirtmem gereken sınırlılıklarından biri, uygulama örneklerinin tamamı olmasa da çoğunluğunun psikolojik özelliklere yönelik olmasıdır. Psikometri alanında çalışan birisi olarak, haddimi aşmamak adına başkaca disiplinlere yönelik örneklerden sınırlı bir şekilde yararlanmayı tercih ettiğimi belirtmeliyim. Fakat örneklerin diğer alan ve disiplinlere kolaylıkla uyarlanabilir biçimde olmasına özen gösterdim.

Bir diğer sınırlılık, bu kitabın ilk basımında R’in 3.6.0 sürümünün kullanılmış olmasıdır. Gözden geçirilmiş ikinci baskıda, bu kitaptaki tüm uygulamaları R’in 4.0.2 sürümü ile gözden geçirip gerekli düzeltmeleri yaptım. Fakat R, hala ve hızla güncellenerek gelişmeye devam ediyor. Bu güncellemelerin bazı farklılıklara yol açabileceği göz önünde bulundurulmalıdır. Kitap içeriğinin bu güncellemelerden olabildiğince az etkilenmesi için R’in temel içeriğinde tanımlı olan ve genel yapısı itibarıyla fazlaca değiştirilmeyen fonksiyonlar dışında başkaca fonksiyonları kullanmamaya özen gösterdim. Ayrıca istatistiksel işlemlere ve modellere yönelik temel algoritmaların uygulamalarını da örneklemeye çalıştım. R ortamında kullanılabilen hazır paketlerden, olabildiğince az yararlanmaya çalıştım. Ancak örnek olması için bazı çok yaygın paketlerin kullanımına örnekler oluşturdum.

Son bir sınırlılık; bu kitaptaki R uygulamalarının, Windows işletim sistemine yönelik olarak hazırlanmış ve düzenlenmiş olmasıdır. Uygun bir işletim sistemini belirlemem gerekiyordu. Yaygınlığını düşünerek Windows işletim sistemini tercih ettim. Bunun, başedilebilecek bir sınırlılık olduğunu düşünüyorum. Uygulamalar, bazı temel ayrımlar dışında diğer işletim sistemlerinde de benzer ya da yakın biçimde yapılabilir.

Hedef Kitle

Bu kitap, geniş kapsamda, istatistiği yeniden ve farklı bir gözle tanımak isteyenlere yöneliktir. Özelde R dilini tanımak ve R dilini istatistik çalışmalarında kullanmak için çözüm arayanlara yöneliktir. Üniversitelerin lisans ve lisansüstü programlarında öğrenim gören öğrencilerin, akademisyenlerin, bilim insanlarının, verilerle çalışan ya da çalışmak isteyen her araştırmacı ve uzmanın yararlanabileceği bir içerikte ve akışta düzenlenmeye çalışılmıştır.

Kullanım Rehberi

Kitap içeriğinin genel sistematığı, olabildiğince bağımsız bölümler oluşturulması üzerine kuruludur. Bölümleri, birbirinden olabildiğince bağımsız biçimde yapılandırmaya çalıştım. Yine de bölümlerin sıralanmasında planlı bir örüntüyü takip ettim. Bu nedenle önerim, baştan sona sıralı bir biçimde çalışılması ya da incelenmesidir.

Bölümlerin içerik yapılandırmasında ise öncelikle kuramsal anlatım ve açıklamalara yer verilmektedir. Bu kısımlar, olabildiğince ilk kaynaklara ya da alanın üzerinde uzlaşmış önemli kaynaklarına atıflar verilerek zenginleştirilmiştir. Sonrasında formül ve algoritmalar ile R fonksiyonları tanıtılmaktadır. Bu kapsamda ilgili yardım sayfalarına ve kaynaklara yönlendirmeler de yapılmıştır. Bu ön hazırlıklar sonrasında R destekli örnek uygulamalara ve çalışmalara yer verilmektedir. Örnek çalışmalar, bir araştırma durumu ya da problemi üzerine yapılandırılmıştır. Bu kapsamda gerçek veri setlerinden yararlanıldığı gibi hipotetik veri setlerinden de yararlanılmıştır. Örnek çalışmalarda öncelikle ilgili formül ve algoritmaların uygulaması, sonrasında R'ın hazır fonksiyonlarının uygulaması aşama aşama gösterilmektedir. Her uygulama sonrasında, araştırma durumu ya da problemine yönelik olarak elde edilen sonuçlara göre genel bir değerlendirme yapılmaktadır. Bu kapsamda elde edilen sonuçların nasıl raporlaştırılabileceği de örneklendirilmektedir.

R uygulamaları, kitap içerisinde, ayrı bir yazı karakteri ile ve paragraf içi verilmektedir. R konsol üzerinde girişi yapılacak kodlar, koyu verilmiştir. Koyu olarak verilen bu kod satırlarının çalıştırılması sonrasında elde edilen çıktılar ise normal tonda verilmiştir. Koyu olarak belirlenen kodları çalıştırarak uygulamaları yapabilir, çıktıları bu kitaptaki çıktılarla karşılaştırabilirsiniz.

Özellikle grafiksel gösterimlerde, renklendirmeler de kullanılmıştır. Kitabın siyah-beyaz baskılarında bunları görme olanağına sahip değilsiniz, elbette. Bu nedenle renkli grafiksel gösterimler, sınırlı tutulmuş, gölgelendirme ve farklı karakter gösterimleri tercih edilmiştir. Yine de ilgili R kodlarını kendi bilgisayarınızda çalıştırdığınızda, renklendirmeleri görebilirsiniz.

Kitapta yer alan kodlar, okuyucunun kolaylıkla kullanabileceği ve R ortamında tanımlayabileceği biçimde düzenlenmiştir. Hazır bir kaynaktan kopyalama ya da aktarmaya ihtiyaç olmamasına ya da bu tür bir ihtiyacın minimum düzeyde olmasına özen gösterdim. Bu kodları hazır bir metin içerisinden kopyalamak yerine satır satır inceleyip kendiniz girerseniz, ayrıntıları daha iyi fark edebilirsiniz. Bu biçimde çalışmanızın öğrenmelerinize olumlu katkıları olacağını düşünüyorum. Kod yazma ve kod okuma, R dilini etkili bir şekilde kullanmada önemli beceriler arasında değerlendirilir. Benzer şekilde kitap ekinde yer alan veri setlerini de kendiniz R ortamında tanımlayabilir ya da dış bir ortamda hazırlayıp R ortamına aktarabilirsiniz. Yine de zamandan kazanmak isteyen okuyucular, veri setlerini, erguldemirRbook@gmail.com e-posta adresini kullanarak talep edebilirler.

Belirtmek istediğim ve önemli gördüğüm bir ayrıntı var. Bu kitapta göreceğiniz kod satırları ve çözümleri, birçok farklı biçimde yazılabilir. Hatta her bir kullanıcının kendine özgü bir kod yazması ya da çözüm geliştirmesi mümkündür. İleriye dönük bir fikir vermek ve hareket alanı oluşturabilmek amacıyla, örnek çalışma olarak sunduğum bazı sınırlı fonksiyon yazma uygulamaları var. Bunları da birçok farklı şekilde düzenlemek ve geliştirmek mümkün. Bu hareket alanı, R dilinin araçsal olarak güçlü yönlerinden birisidir. Bu bağlamda, bu kitapta verilen R kod satırlarını uyguladıktan sonra, kendinizce farklı çözümleri üretmenizin ve denemenizin, öğrenmelerinize çok olumlu katkıları olacağını düşünüyorum.

İkinci Baskıya Birkaç Söz ve Teşekkür

Kitabımızın ilk baskısının, beklentilerimden çok daha hızlı bir sürede tükenmiş olmasından dolayı bir yandan memnun ve mutlu diğer yandan mahcubum. Ne diyebilirim, bu kitapta ortaya koymaya çalıştığım öğrenme mücadelesinde birlikte yol aldığımıza inandığım tüm okuyuculara en samimi duygularıyla teşekkür ediyorum. Sizlerden aldığım olumlu ve güçlü geribildirimlerle, gözden geçirilmiş bu ikinci baskıda elimden geldiğince iyileştirmeler ve düzeltmeler yapma imkanı buldum.

R'ın mevcut en güncel sürümü olan R 4.0.2 ile tüm işlemleri baştan sona gözden geçirdim ve gerekli uyumlaştırmaları yaptım. Daha belirgin ve anlaşılır olması için kitaptaki tüm görsellerin çözünürlük düzenlemelerini yaparak bunları güncelledim. Bazı kodlardaki ve işlemlerdeki eksik gösterimleri düzelttim. Birçok kod satırına ek argümanlar ekleyerek daha işlevsel çıktılar alınmasını sağlamaya çalıştım. Bazı kuramsal açıklama ve yorumları kısmen yeniden düzenleyerek daha akıcı ve anlaşılır hale getirmeye çalıştım. Elbette biçimsel ve dilbilgisel düzeltmelerim de oldu. Sonuç olarak, bu kitabı, birlikte, daha yararlanılabilir bir kitap haline getirebildiğimizi düşünüyorum. Bu bağlamda, bu kitabın gözden geçirilmiş ikinci baskının yayımlanmasındaki katkı ve destekleriniz için özellikle ve ayrıca teşekkür ediyorum.

Yazar Hakkında ve İletişim

Yazarın, matematik öğretmenliği ve işletme programlarında lisans, ölçme ve değerlendirme ve eğitim yönetimi alanlarında yüksek lisans mezuniyetleri bulunmaktadır. Doktorasını 2013 yılında Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme programında

tamamlamıştır. 1998-2007 yılları arasında özel okullarda ve Millî Eğitim Bakanlığının taşra teşkilatlarında, 2007-2012 yılları arasında Millî Eğitim Bakanlığı merkez teşkilatında matematik öğretmeni, alan uzmanı ve araştırmacı olarak görev yapmıştır. 2012 yılında Ankara Üniversitesi Eğitim Bilimleri Fakültesinde başladığı akademik görevine halen öğretim üyesi olarak devam etmektedir. Yazar, lisans ve lisansüstü programlarda ölçme ve değerlendirme, psikolojik ölçmeler ve test kuramları, istatistik, ileri istatistik ve araştırma yöntemleri alanlarında dersler vermektedir. Ayrıca lisansüstü tez danışmanlıkları, kamu kurumlarında akademik danışmanlık ve uzmanlık, kurumsal projelerde yürütücü, araştırmacı ve akademik danışmanlık görevleri bulunmaktadır. Yazarın, eğitim bilimleri alanında kitap ve kitap bölümleri, ulusal ve uluslararası makaleleri ve diğer birçok yayını bulunmaktadır.

Akademik nezaket ve haddini bilme gereği “yazar hakkında” üçüncül tekil şahıs açıklamalarından sonra, konunun özüne dönelim. Bu kitabı aldıysanız ve ilgilendiyseniz, muhtemelen, R diliyle istatistiği yeniden ve farklı bir gözle öğrenmeye gönüllüsünüz ya da açkırsınız demektir. Bu kitabın hipotezi budur; “istatistiği yeniden ve farklı bir gözle okumak”! Katkım olabilirse ne mutlu! Görüş, düşünce, eleştiri, çözüm, öneri ve deneyimlerinizi paylaşmak isterseniz, memnuniyet duyarım!

E-posta : erguldemirRbook@gmail.com

R Ortamında Veri Yapıları

R ortamında verileri depolamak ve kullanmak için bir dizi obje tanımlamak gerekir. Basitten karmaşığa doğru bu objeler; tek elemanlı vektörler (scalar), vektörler, vektörlerin özel bir biçimi olan faktörler, matrisler, üç boyutlu matrisler (array), veri setleri (data frame) ve listeler (list) şeklinde sıralanabilir.

1.1.Tek Elemanlı Vektörler (Scalars)

R'da obje olarak tanımlanan tek bir nümerik değer, karakter ya da mantıksal değer, tek elemanlı ve tek boyutlu bir vektör oluşturur. Bu tanımlama nümerik, karakter ya da mantıksal değer olarak yapılabilir. Nümerik değerler, herhangi bir ek sembol kullanmaksızın doğrudan belirli bir objeye atanabilir. Karakter değerler, tırnak işareti içerisinde tanımlanır. Mantıksal değerler de nümerik değerler gibi TRUE-FALSE ya da bunların kısaltmaları olan T-F olmak üzere tırnak işareti olmaksızın tanımlanır.

```
x1 <- 35
x1
[1] 35
class(x1)
[1] "numeric"
```

```
y1 <- "A"
y1
[1] "A"
class(y1)
[1] "character"
```

```
z1 <- TRUE
z1
[1] TRUE
class(z1)
[1] "logical"
```

```
t1 <- F
t1
[1] FALSE
class(t1)
[1] "logical"
```

Matematiksel ya da istatistiksel bazı işlemler yapılırken, örneğin bir sınıfın mevcudunu, deney ve kontrol gruplarındaki denek sayılarını, bir ölçekte yer alan toplam madde sayısını, bir testten alınan toplam puanların ortalamasını ya da belli bir hesaplama sonucu elde edilen bazı değerleri tek elemanlı bir vektör olarak tanımlama ihtiyacı olabilir. Obje olarak tanımlanmış bu değerler, çoğunlukla yapılacak ileri işlemlerde kullanılır ve işlem kolaylığı sağlar.

1.2.Vektörler

Vektörler, tek boyutlu veri yapılarıdır ve değişken tanımlamalarında yaygın olarak kullanılır. R'da bir vektörü nümerik, karakter ya da mantıksal olarak tanımlamak mümkündür. Bir vektörün elemanları aynı türdür.

R'da vektörler, `c()` sembolü içerisinde ve virgül ayracı kullanılarak tanımlanır. Tek elemanlı vektörlerde olduğu gibi nümerik ve mantıksal vektörlerin elemanları tırnak işareti olmadan, karakter vektör elemanları ise tırnak işareti kullanılarak tanımlanır.

```
x2 <- c(18, 20, 21, 23, 20)
x2
[1] 18 20 21 23 20

y2 <- c("A", "A", "B", "C", "A")
y2
[1] "A" "A" "B" "C" "A"

(z2 <- c(TRUE, TRUE, FALSE, FALSE, TRUE))
[1] TRUE TRUE FALSE FALSE TRUE

(t2 <- c(T, T, F, F, T))
[1] TRUE TRUE FALSE FALSE TRUE
```

Vektörler, R ortamında yapılan işlemlerde kullanılan veriler için en temel yapılardan biridir. Nümerik vektörlerle, temel dört işlem, üs alma, kök alma, sayma, sıralama, birleşim ya da kesişim alma, karşılaştırma, mantıksal karşılaştırma gibi birçok işlem yapılabilir. Benzer şekilde karakter ya da mantıksal değerler içeren vektörlerle de bu işlemlerin bir kısmı yapılabilir. Yapılan işlemlerden elde edilen sonuçlar, ayrı bir vektör olarak tanımlanıp R geri planında depolanabilir. Yapılabilecek işlemlere yönelik bazı örnekler aşağıda, nümerik iki vektör kullanılarak verilmiştir.

```
#İki nümerik vektör tanımlayalım

xv <- c(12, 13, 16, 18)
yv <- c(20, 18, 17, 22)

#Vektörlerin uzunlukları belirlenebilir.

length(xv)          length(yv)
[1] 4                [1] 4

#Küçükten büyüğe veya tersi şekilde sıralanabilir.

sort(xv)             rev(sort(yv))
[1] 12 13 16 18       [1] 22 20 18 17

#Vektör elemanlarının toplamı belirlenebilir.

sum(xv)              sum(yv)
[1] 59                [1] 77
```

#Vektör elemanlarının toplamları birbirinden çıkarılabilir.

```
sum(xv) - sum(yv)      sum(yv) - sum(xv)
[1] -18                [1] 18
```

#Vektör elemanlarının kareleri ya da karekökleri alınabilir.

```
xv^2                  sqrt(yv)
[1] 144 169 256 324   [1] 4.472136 4.242641 4.123106 4.690416
```

#Ondalık değerle tanımlı nümerik bir vektörün elemanları, belli bir ondalığa yuvarlanabilir.

```
zv <- sqrt(yv)

round(zv)              round(zv, 2)
[1] 4 4 4 5            [1] 4.47 4.24 4.12 4.69
```

#Bir vektörden belli bir değer çıkarılabilir. Bu durumda her bir elemandan bu değer çıkarılır.

```
xv - 10
[1] 2 3 6 8
```

#Bir vektörle belli bir değer çarpılabilir. Bu durumda vektörün her bir elemanı ile bu değer çarpılır.

```
yv * 2
[1] 40 36 34 44
```

#Vektörler birbiri ile toplanabilir, birbirinden çıkarılabilir, vb. Bu durumda sıralı olarak karşılıklı elemanlar, istenen işleme alınır. Sonuçta eşit uzunlukta yeni bir vektör elde edilir.

```
xv + yv              xv - yv
[1] 32 31 33 40      [1] -8 -5 -1 -4
```

#Vektörlerin farklarının kareleri alınabilir.

```
(xv - yv)^2
[1] 64 25 1 16
```

#Vektörler, mantıksal olarak, belli bir değerle veya birbirleri ile karşılaştırılabilir. R ortamında mantıksal argümanların (logics) kullanımı ayrı bir yeterlik alanıdır. Mantıksal kullanım