

Herkes İçin İstatistiksel Programlama ve Analiz



Richard Cotton

Çeviri Editörü: Önder Sünbül

O'REILLY®

A PEGEM
AKADEMİ



Richard Cotton

Çeviri Editörü: Doç. Dr. Önder Sünbül

HERKES İÇİN İSTATİSTİKSEL PROGRAMLAMA VE ANALİZ: R

ISBN 978-605-318-572-7
DOI 10.14527/9786053185727

Kitabın orijinal adı: Learning R
Yazarı: Richard Cotton
Baskı sayısı: 1. Baskı
ISBN: 978-1-449-35710-8
O'Reilly Media, Inc
Authorized Turkish translation of the English edition of Learning R, ISBN 9781449357108 © 2013 Richard Cotton. This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

Kitap içeriğinin tüm sorumluluğu yazarlarına aittir.

© 2020, PEGEM AKADEMİ

Bu kitabın basım, yayım ve satış hakları Pegem Akademi Yay. Eğt. Dan. Hizm. Tic. A.Ş. ye aittir. Anılan kuruluşun izni alınmadan kitabın tümü ya da bölümleri, kapak tasarımı; mekanik, elektronik, fotokopi, manyetik kayıt ya da başka yöntemlerle çoğaltılamaz, basılamaz, 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 yayınevidir**. 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** ve **Pegemindex.net** 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: Şubat 2020, Ankara

Yayın-Proje: Özge Yüksek
Dizgi-Grafik Tasarım: Tuğba Kaplan
Kapak Tasarımı: Pegem Akademi

Baskı: Ay-bay Kırtasiye İnşaat Gıda Pazarlama ve Ticaret Ltd. Şti.
Çetin Emeç Bulvarı 1314. Cadde No: 37A-B Çankaya/ANKARA
Tel: (0312) 472 58 55

Yayıncı Sertifika No: 36306
Matbaa Sertifika No: 46661

İletişim

Karanfil 2 Sokak No: 45 Kızılay/ANKARA
Yayınevi: 0312 430 67 50 - 430 67 51
Dağıtım: 0312 434 54 24 - 434 54 08
Hazırlık Kursları: 0312 419 05 60
İnternet: www.pegem.net
E-ileti: pegem@pegem.net
WhatsApp Hattı: 0538 594 92 40

Ön Söz

R bir programlama dili ve veri analizi ve istatistik için bir yazılım ortamıdır. Bir GNU projesidir ve bu onun ücretsiz ve açık kaynak kodlu yazılım olduğu anlamına gelir. Sayısı bir milyon üzerinde olan kullanıcıların pek çok ölçüm ve değerlendirmeleri ile kademeli olarak gelişmektedir. Topluluğun katkıda bulunduğu 4000'in üzerinde eklenti paketleri vardır ve bu rakam her yıl %25 artmaktadır. Tiobe programlama dili popülerliği endeksi R'ı bu kitabın yazıldığı zamanda 24. sıraya koymuştur, ki bu da SAS ve MATLAB ile neredeyse eşit düzeydedir.

R, istatistik ya da veri analizine ihtiyaç duyulan neredeyse her alanda kullanılır. Diğer daha küçük çalışma alanlarının yanı sıra finans, pazarlama, farmasotik, genom bilimi, epidemiyoloji, sosyal bilimler ve öğretim alanlarının tümünü kapsar.

Kitap Hakkında

R öncelikli olarak istatistiksel analizler yapmanıza olanak sağlamak için tasarlandığından, R hakkında yazılan kitapların çoğu size istatistik veya model veri kümelerinin nasıl hesaplanacağını öğretmeye odaklanır. Bu da ne yazık ki veri analizinin gerçekliğinin büyük bir bölümünün gözden kaçırılmasıdır. İleri düzeyde bir araştırma yapmadığınız sürece, kullandığınız istatistiksel teknikler genellikle rutin olacaktır ve çalışmanızın modelleme kısmı çok büyük olmayabilir. Veri analizinin eksiksiz iş akışı şuna benzer:

1. Bazı verileri al .
2. Verileri temizle.
3. Verileri açım ile ve görselleştir.
4. Verileri modelle ve yordamalarda bulun.
5. Sonuçlarını sun veya yayınla.

Elbette, her aşamada sonuçlarınız sizi daha fazla veri aramaya ya da mevcut verilerinizi işlemenin farklı bir yolunu bulmanızı sağlayacak ve sizi bir adım geriye götürebilecek ilginç sorular oluşturacaktır. İş akışı tekrarlamalı olabilir fakat her bir adımın atılması gereklidir.

Bu kitabın ilk bölümü size R'ı sıfırdan öğretmek için dizayn edilmiştir – dilini anlamak için herhangi bir deneyime ihtiyacınız yoktur. Aslında hiçbir programlama deneyimine ihtiyaç yoktur, ancak bazı temel programlama bilgisine sahipseniz bu size yardımcı olacaktır. Örneğin, kitap kodunuzu nasıl yorumlayacağınızı ve bir for döngüsünü nasıl yazacağınızı açıklar, ancak bunların ne olduklarını detaylı bir şekilde açıklamaz. Nasıl programlanacağı konusunda gerçek bir tanıtım metni istiyorsanız Jason R. Briggs tarafından yazılan *Yeni başlayanlar (çocuklar) için Python* başlangıç için iyi bir seçimdir!.

Kitabın ikinci kısmı R'daki tüm veri analizi iş akışında sizi yönlendirir. Burada, bazı temel istatistiksel bilgilere sahip olunduğu varsayılmaktadır. Örneğin, ortalama ve standart sapma gibi terimleri ve sütun grafiğinin ne olduğunu anlamalısınız.

Kitap, nesne yönelimli programlama ve paket program oluşturma gibi bazı daha gelişmiş R konuları ile sonlanmaktadır. *Garret Grelomund'un R ile Veri Analizi* kitabı, veri analizi iş akışını daha detaylı bir şekilde kapsayarak bu kitabın bıraktığı yerden başlıyor.

Bir uyarı: bu bir referans kitabı değildir ve konuların çoğu detaylı bir şekilde ele alınmamıştır. Bu kitap size R ile neler yapabileceğiniz hakkında fikirler sunmak ve uygulama yapmanıza imkan vermek için öğretiler sağlamaktadır. 4000 eklenti paketinin tümünü kapsayacak kadar yer yoktur, fakat okumayı bitirdiğinizde ihtiyacınız olanları bulabilmeli ve onları kullanmaya başlamak için ihtiyacınız olan yardımı alabilmelisiniz.

Bu Kitapta Neler Var?

Bu, iki kısımdan oluşan bir kitaptır. İlk bölüm R'ı kullanmanız için sahip olmanız gereken teknik becerileri sağlamak için tasarlanmıştır; her bölüm farklı bir veri türleri setine (örneğin 4.bölüm vektörleri, matrisleri ve dizileri kapsar) ya da kavrama (örneğin 8.bölüm dallanma ve döngülemeyi kapsar) kısa bir giriş niteliğindedir.

Kitabın ikinci bölümü eğlenceyi artırıyor; gerçek veri analizini uygulamada görebiliyorsunuz. Her bölüm, verileri almaktan sonuçlarınızı yayınlamaya kadar giden standart veri analizi iş akışı ile ilgili bir bölüm kapsar.

I. Kısım "R Dili"nde bulacaklarınız:

- 1. Bölüm *Giriş*, R'in nasıl kurulacağını ve nereden yardım alacağınızı açıklar.
- 2. Bölüm *Bir Bilimsel Hesap Makinesi R'ı nasıl bilimsel hesap makinesi olarak kullanacağınızı gösterir.*
- 3. Bölüm *Değişkenlerin İncelenmesi ve Kişisel Çalışma Alanınız* değişkenleri farklı şekillerde incelemenize olanak sağlar.
- 4. Bölüm *Vektörler, Matrisler ve Diziler* vektörler, matrisler ve dizinleri kapsamaktadır.
- 5. Bölüm *Listeler ve Data Frame'ler*, listeleri ve data frameleri kapsar (elektronik tablo benzeri veri için).
- 6. Bölüm *Çevreler ve Fonksiyonlar*, çevreler ve fonksiyonları kapsar.

- 7. Bölüm *Metin ve Faktörler*, metinleri ve faktörleri kapsar (kategorik veri için).
- 8. Bölüm *Akış Kontrolü ve Döngüler*, dallanmayı (eğer varsa) ve temel döngülemeyi kapsar.
- 9. Bölüm *İleri Döngüleme*, uygulama işlevine sahip gelişmiş döngülemeyi ve onun çeşitlerini kapsar.
- 10. Bölüm *Paketler* eklenti paketlerinin nasıl kurulacağını ve kullanılacağını açıklar.
- 11. Bölüm *Tarihler ve Zamanlar* tarihleri ve zamanları kapsar.

II. Kısım *Veri Analizi İş Akışı*'nda ele alınan konular aşağıdaki gibidir:

- 12. Bölüm *Veri Temini* verilerin R'a nasıl alınacağını gösterir.
- 13. Bölüm *Temizleme ve Dönüştürme* verilerin temizliğini ve değiştirilmesini açıklar.
- 14. Bölüm *Keşfetme ve Görselleştirme* istatistikleri hesaplayarak ve grafikleyerek verileri keşfetmenizi sağlar.
- 15. Bölüm *Dağılımlar ve Modelleme* modellemeyi tanıtır.
- 16. Bölüm *Programlama* çeşitli ileri programlama tekniklerini kapsar.
- 17. Bölüm *Paket Oluşturma* işinizi başkaları için nasıl paketleyeceğinizi gösterir.

Son olarak III. Kısım yani Appendix'te faydalı referanslar bulunmaktadır:

- Ek A, *Değişkenlerin Özellikleri* farklı değişken türlerinin özelliklerini karşılaştıran tablolar içerir.
- Ek B, *R ile Yapılabilen Diğer Şeyler*, R'da yapabileceğiniz diğer şeyleri açıklar.
- Ek C, *Kısa Sınav Cevapları* bölüm sonu testlerinin cevaplarını içerir.
- Ek D, *Alıştırma Sorularının Çözümleri* bölüm sonu programlama alıştırmalarının cevaplarını içerir.

Hangi Bölümleri Okumalıyım?

Daha önce hiç R kullanmadıysanız, en baştan başlayın ve her bölümü detaylı bir şekilde inceleyin. Eğer zaten R ile ilgili bir deneyime sahipseniz ilk bölümü atlamak isteyip, R çekirdek dilindeki (temel R dilindeki) bölümleri gözden geçirebilirsiniz.

Her bölüm farklı bir konuyu ele almaktadır, bu yüzden bir bölümden diğerine az miktarda bağlılık olmasına rağmen, ilginizi çeken bölümleri seçmek mümkündür.

Geçenlerde bu konuyu “Aptallar için R” kitabının yazarı Andrie de Vries ile tartıştım. O pes etmemi, bunun yerine onun kitabını okumamı önerdi.

Kitapta Kullanılan Yazım Kuralları

Bu kitapta aşağıdaki yazı tipi kuralları kullanılmıştır:

İtalik yazılar

Yeni terimleri, URL’leri, E-posta adreslerini, dosya ve yol adlarını, ve dosya uzantılarını gösterir.

Sabit Genişlik

Paragraflar içinde olduğu gibi sözlü olarak kopyalanması gereken kod örnekleri için kullanılır; değişken veya fonksiyon adları, veri çeşitleri, çevre değişkenleri, ifadeler ve anahtar kelimeler gibi program öğelerinden söz etmek amacıyla kullanılır. Kod bloklarından alınan çıktı da sabit genişliktedir, öncesinde çift **hash** (##) vardır.

İtalik Sabit Genişlik

Kullanıcının sağladığı değerlerle veya içerik tarafından belirlenen değerlerle değiştirilmesi gereken metni gösterir.

<http://4dpicharts.com/r-codestyle-guide> adresinde bu kitapta kullanılan kod için bir stil rehberi bulunmaktadır.



Bu simge bir ipucu, öneri veya genel notu gösterir.



Bu simge bir uyarıyı gösterir.

Hedefler, Özetler, Kısa Sınavlar ve Alıştırmalar

Her bölüm, ilerleyen sayfalarda ne bulacağınızı bilmenizi sağlayacak bir hedef listesi ile başlar ve öğrendiklerinizi tekrarlayan bir özetle biter. Ayrıca, konsantre olduğunuzdan (televizyon seyrederken okuyormuş gibi yapmadığınızdan) emin olmak için bir sınava tabi tutulacaksınız. Soruların cevapları bölüm içinde bulunabilir (veya kopya çekmek istiyorsanız kitabın sonunda). Son olarak her bölüm, sizi bazı R kodlarını yazmaya dahil eden bazı alıştırmalarla son bulur. Her bir alıştırtma tanımından sonra köşeli parantez içinde, alıştırtmayı tamamlamanızın kaç dakika süreceğine ilişkin cömert bir tahmin belirten bir sayı vardır.

Kod Örneklerini Kullanma

Ek materyaller (kod örnekleri, alıştırtmalar, vb) şu adresten indirilebilir:<http://cran.r-project.org/web/packages/learningr>.

Bu kitap işinizi yapmanıza yardımcı olmak için burada. Genel olarak, eğer örnek kod bu kitapla sunulursa, onu programlarımızda ve belge işlemede kullanabilirsiniz. Kodun önemli bir miktarını çoğaltmadığınız sürece izin almak için bizimle iletişime geçmenize gerek yoktur. Örneğin, bu kitaptan birkaç kod parçası kullanan bir program yazmak izin almayı gerektirmez. O'Reilly kitaplarından örnek CD satmak veya dağıtmak izin gerektirir. Bu kitabı alıntıylaarak bir soruya cevap vermek ve örnek kod alıntılanak izin gerektirmez. Kitaptan ürününüzün belge işlemesine önemli miktarda örnek kod dahil etmek izin almayı gerektirir. Talep etmeyiz fakat atfınıza müteşekkir oluruz. Bir atf genellikle başlığı, yazarı, yayınevi-ni, Uluslararası Standart Kitap Numarasını (ISBN) içerir. Örneğin “*Learning R* by Richard Cotton (O'Reilly). Copyright 2013 Richard Cotton, 978-1-449-35710-8.”

Kod örnekleri kullanımınızın adil kullanım veya yukarıda verilen iznin dışında kaldığını düşünüyorsanız permissions@oreilly.com. adresi üzerinden bizimle iletişime geçmekten çekinmeyin.

Safari® Çevrimiçi Kitaplar



Safari Çevrimiçi Kitaplar teknoloji ve iş dünyasında dünyanın önde gelen yazarlarının uzman içeriğini hem kitap hem video biçiminde sunan (dağıtan, servis eden) talep üzerine çalışan dijital bir kütüphanedir.

Teknoloji uzmanları, yazılım geliştiriciler, web tasarımcıları, işletme ve yaratıcı profesyoneller Safari Çevrimiçi Kitaplarını araştırma, problem çözmeye, öğrenme, ve sertifikasyon eğitimi için temel (birincil) kaynak olarak kullanmaktadırlar.

Safari Books Online, kuruluşlar, devlet kurumları ve bireyler için bir dizi ürün karması ve fiyatlandırma programı sunmaktadır. Aboneler, O'Reilly Media, Prentice Hall Professional, Addison-Wesley Professional, Microsoft Press, Sams, Que, Peachpit Press, Focal Press, Cisco Press, John Wiley & Sons, Syngress, Morgan Kaufmann, IBM Redbooks, Packt, Adobe Press, FT Press, Apress, Manning, New Riders, McGraw-Hill, Jones & Bartlett, Course Technology gibi yayıncılardan tamamen araştırılabilir bir veri tabanında binlerce kitaba, eğitim videolarına ve yayım öncesi makalelere erişme imkanına sahiptirler. Safari Books Online hakkında daha fazla bilgi için lütfen bizi internet üzerinden ziyaret edin.

Bizimle Nasıl İletişime Geçilir?

Lütfen bu kitapla ilgili görüş ve sorularınızı yayıncıya iletin:

O'Reilly Media, Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

800-998-9938 (Amerika ya da Kanada'da)

707-829-0515 (uluslararası ya da yerel)

707-829-0104 (fax)

Bu kitap için yazım/dizgi hatalarını, örnekleri, ve diğer ek bilgileri listelediğimiz bir web sayfamız mevcuttur. Bu sayfaya <http://oreil.ly/learningR> adresinden erişebilirsiniz.

Bu kitap hakkında yorum yapmak ya da teknik sorular sormak için bookquestions@oreilly.com. adresine e-posta gönderebilirsiniz.

Kitaplarımız, kurslarımız, konferanslarımız ve haberlerimiz hakkında daha fazla bilgi için <http://www.oreilly.com>. adresindeki web sitemizi ziyaret ediniz.

Bizi facebook'ta bulun: <http://facebook.com/oreilly>

Bizi Twitter'da takip edin: <http://twitter.com/oreillymedia>

Bizi YouToubé'da izleyin: <http://www.youtube.com/oreillymedia>

Teşekkür

Pek çok muhteşem insan bu kitabın hazırlanmasında yardımcı oldu, bunlardan biri de akıllıca tavsiyelerle dolu harika editörüm Meghan Blanchett'dir.

Veriler birkaç harika insan tarafından bağışlandı:

AMD'den Bill Hogan Alpe d'Huez bisiklet (cycling) veri setini buldu ve temizledi, ve beni gonore veri setine yönlendirdi. Onun hastaliksız olduğunu vurgulamamı istedi, Bayanlar.

CEFAS'tan Ewan Hunter Kuzey Denizi yengeç veri setini tedarik etti.

Cambridge Üniversitesinden Corina Logan geyik kafatası verisini derledi ve tedarik etti.

Leiden Üniversitesinden Edwin Thoen Obama'ya karşı McCain veri setini tedarik etti.

Gwern Branwen aşırı miktarda Japon çizgi romanı izleyerek ve okuyarak hafu veri setini tedarik etti. Kudos. (övgü).

Diğer birçok kişi bana veri setleri gönderdi; hepsine yer yoktu, ama yine de teşekkür ederim!

Marin Yazılım'dan Daisy Vincent ve JD Long gibi Bill Hogan da kitabı gözden geçirdi. JD'nin nerede çalıştığını bilmiyorum ama Bermudada yaşıyor, bu yüzden muhtemelen üçgenleri kapsıyordu. TDX grubundan James White, Ben Hanks, Beccy Smith, ve Guy Bourne; HSL'den Alex Hogg ve Adrian Kelsey; CEFAS'tan Tom Hull, Karen Vanstaen, Rachel Beckett, Georgina Rimmer, Ruth Wortham, Bernardo Garcia-Carreras, ve Joana Silva; Tel Aviv Üniversitesinden Tal Galili; R Stüdyo'dan Garrett Grolemond ve New York Şehir Üniversitesi'nden John Verzani ayrıca yorum ve geri bildirim sağlamışlardır. CEFAS'tan David Maxwell hemen hemen CEFAS'ta çalışan herkesi kitabımı incelemek için harika bir şekilde toparlamıştır.

John Verzani de bu kitabın ortaya çıkmasına yardımcı olduğu ve kitabın yapısı konusunda tavsiyelerde bulunduğu için büyük bir övgüyü hak etmiştir.

O'Reilly'den Sanders Kleinfeld metindeki karakter kodlamaları ile uğraştığım sırada saçlarımı çekerken mükemmel bir teknik destek sağlamıştır. Yihui Xie görev alanının dışına çıkarak AsciiDoc'u üretmemde bana yardımcı oldu. Rachel Head yazıyı kontrol edip düzeltirken 4000'in üzerinde program ve yazım hatasını tek başına bulmuştur.

Garib Murshudov 2004 yılında R'ı bana ilk öğreten öğretim görevlisidi.

Son olarak, Janette Bowler, ben yazmakla meşgulken gösterdiği sonsuz sabır ve destek için teşekkürü hak ediyor.

Çeviri Editörünün Ön Sözü

Günümüz teknoloji çağı olarak adlandırılmaktadır. Teknoloji her geçen gün ivmeleyen bir gelişim süreci içerisinde ve bu gelişimin nereye kadar gideceği ve nasıl sonuçlanacağı hakkında bazıları iyimser bazıları kötümser olan çeşitli teoriler geliştirilmiştir. İster iyi olsun ister kötü olsun şunu çok açık bir şekilde iddia edebiliriz: Ciddi bir değişim süreci içerisindeyiz ve oluşan değişimlerden etkilenmemiz kaçınılmaz. Bilişim sektörünün ve veriyi yönetip işleyebilmenin önemi giderek artmaktadır. Bu gelişmelerin yarattığı dalgalanmaların etkilerini bütün bilim alanlarında görebilmemiz mümkün. Verinin önemi ve niteliği de bu değişim sürecinden payını almış durumdadır. Veri petrolden daha değerli hale gelmiştir. Akan veriyi kontrol altına alabilmek ve yönetebilmek geçmişe göre çok daha önemlidir. Özellikle istatistiksel veri işlenmesi konusunda kullanılan araçlar incelendiğinde, R ve Python olmak üzere iki önemli dilin öne çıktığı görülebilir.

Python'un daha çok programcılar ve yazılım geliştiriciler tarafından tercih edildiği, R'in ise daha çok akademisyenler tarafından tercih edildiği bilinmektedir. R'in çıkış yapılması istatistiksel ve akademik ihtiyaçların karşılanması üzerineyken, günümüzde R birçok farklı amaçlar için de kullanılabilir kapasiteye erişmiştir. R dünya çapında geniş bir kitle tarafından kullanılan açık kaynak kodlu, bilgi paylaşmak içindir felsefesine dayalı yaşayan ve güncel bir dildir. Kullanıcı ve geliştirici kitlesi de her geçen gün artmaktadır. Özellikle akademik amaçlar için bir iletişim dili olma yolunda hızla ilerlemektedir.

Yurt dışında oldukça popüler olan bu kitabı çevirmemizdeki temel amaçlardan biri, söz edilen gelişmeler ve değişimlere uyum sağlayabilmesi açısından ülkemize katkı sağlamaktır.

Kitap içeriği incelendiğinde, kitabın adında yer alan "Herkes için" ibaresinin gerçekten karşılığının bulunduğu görülebilir. İstatistiksel programlama hakkında hiçbir deneyimi olmayan okuyuculardan, yazılım geliştiricilere kadar geniş bir kitlenin temel ihtiyaçlarını karşılayacak niteliktedir.

Prof. Dr. Necmi Gürsakağın deyimiyle bu veri akarken biz bakmamalıyız! Çağımızın olmazsa olmazı veriyi yönetebilme ve analiz edebilme becerilerimizi, genel geçer bir dil kullanarak geliştirmek kendimiz için yapmanız gereken bir yatırımdır.

Her ne kadar yazarın espri anlayışı biraz farklı olsa da, çevirmekten ve ülkemize kazandırmaktan büyük keyif aldık.

Ülkemiz akademisyenlerine ve akademisyen adaylarına yararlı olması dileğiyle.

Kitabın çeviri sürecinde bizlere geri bildirimleri ile önemli katkılar sağlayan Doç. Dr. Kasım KIROĞLU'na ve süreçte vermiş oldukları desteklerden dolayı PEGEM AKADEMİ'ye teşekkür ederiz.

Çeviri Editörü

Doç. Dr. Önder SÜNBÜL

Şubat, 2020

BÖLÜMLER VE ÇEVİRENLERİ

Çeviri Editörü: Doç. Dr. Önder SÜNBÜL

1. KISIM: R DİLİ

1. BÖLÜM: GİRİŞ

Doç. Dr. Önder SÜNBÜL, Mersin Üniversitesi

2. BÖLÜM: BİR BİLİMSEL HESAP MAKİNESİ

Dr. Öğr. Üyesi Seçil ÖMÜR SÜNBÜL, Mersin Üniversitesi

3. BÖLÜM DEĞİŞKENLERİN İNCELENMESİ VE KİŞİSEL ÇALIŞMA ALANI

Dr. Öğr. Üyesi Seçil ÖMÜR SÜNBÜL, Mersin Üniversitesi

4. BÖLÜM: VEKTÖRLER, MATRİSLER VE DİZİLER

Dr. Öğr. Üyesi Seçil ÖMÜR SÜNBÜL, Mersin Üniversitesi

5. BÖLÜM LİSTELER VE DATA FRAME'LER

Dr. Öğr. Üyesi Seçil ÖMÜR SÜNBÜL, Mersin Üniversitesi

6. BÖLÜM: ÇEVRELER VE FONKSİYONLAR

Dr. Öğr. Üyesi Seçil ÖMÜR SÜNBÜL, Mersin Üniversitesi

7. BÖLÜM: METİN VE FAKTÖRLER

Dr. Öğr. Üyesi Seçil ÖMÜR SÜNBÜL, Mersin Üniversitesi

8. BÖLÜM: AKIŞ KONTROLÜ VE DÖNGÜLER

Doç. Dr. Önder SÜNBÜL, Mersin Üniversitesi

9. BÖLÜM: İLERİ DÖNGÜLEME

Doç. Dr. Önder SÜNBÜL, Mersin Üniversitesi

10. BÖLÜM: PAKETLER

Doç. Dr. Önder SÜNBÜL, Mersin Üniversitesi

11. BÖLÜM: TARİHLER VE ZAMANLAR

Doç. Dr. Önder SÜNBÜL, Mersin Üniversitesi

2. KISIM: VERİ ANALİZİ İŞ AKIŞI

12. BÖLÜM: VERİ TEMİNİ

Dr. Aytaç KARABAY, Groningen Üniversitesi

13. BÖLÜM: TEMİZLEME VE DÖNÜŞTÜRME

Dr. Aytaç KARABAY, Groningen Üniversitesi

14. BÖLÜM: KEŞFETME VE GÖRSELLEŞTİRME

Dr. Aytaç KARABAY, Groningen Üniversitesi

15. BÖLÜM: DAĞILIMLAR VE MODELLEME

Dr. Aytaç KARABAY, Groningen Üniversitesi

16. BÖLÜM: PROGRAMLAMA

Doç. Dr. Önder SÜNBÜL, Mersin Üniversitesi

17. BÖLÜM: PAKET OLUŞTURMA

Doç. Dr. Önder SÜNBÜL, Mersin Üniversitesi

3. KISIM: EKLER

EK - A: DEĞİŞKENLERİN ÖZELLİKLERİ

Dr. Aytaç KARABAY, Groningen Üniversitesi

EK - B: R İLE YAPILABİLEN DİĞER ŞEYLER

Dr. Aytaç KARABAY, Groningen Üniversitesi

EK - C: KISA SINAV CEVAPLARI

Dr. Aytaç KARABAY, Groningen Üniversitesi

EK - D: ALIŞTIRMA SORULARININ ÇÖZÜMLERİ

Dr. Aytaç KARABAY, Groningen Üniversitesi

İÇİNDEKİLER

Ön Söz..... iii

1. KISIM: R DİLİ

1. Giriş	3
Bölümün Amaçları.....	3
R Nedir?.....	3
R'ın İndirilmesi.....	4
Bir IDE Seçmek	5
Emacs + ESS	5
Eclipse/Architect.....	6
RStudio	6
Revolution – R.....	7
Live – R.....	7
Diğer IDE ve Düzenleyiciler.....	8
İlk Programınız.....	8
R'da Nasıl Yardım Alırsınız?.....	8
İlave İlgili Yazılımların İndirilmesi	11
Özet	12
Bilginizi Test Edin: Kısa Sınav	13
Bilginizi Test Edin: Alıştırmalar	13

2. Bir Bilimsel Hesap Makinesi

Bölümün Amaçları.....	15
Matematiksel Operatörler ve Vektörler.....	15
Değişkenler Atama.....	20
Özel Sayılar	22
Mantıksal Vektörler.....	23
Özet	26
Bilginizi Test Edin: Kısa Sınav	26
Bilginizi Test Edin: Alıştırmalar	26

3. Değişkenlerin İncelenmesi ve Kişisel Çalışma Alanı	29
Bölümün Amaçları.....	29
Sınıflar.....	29
Sayıların Farklı Türleri.....	30
Diğer Ortak Sınıflar	31
Sınıfları Kontrol Etme ve Değiştirme	34
Değişkenlerin İncelenmesi.....	38
Çalışma Alanı	41
Özet	42
Bilginizi Test Edin: Kısa Sınav	43
Bilginizi Test Edin: Alıştırmalar	43
4. Vektörler, Matrisler ve Diziler.....	45
Bölümün Amaçları.....	45
Vektörler	45
Sıralar.....	47
Uzunluklar	48
İsimler.....	49
Vektörleri İndeksleme	50
Vektör Geri Dönüşümü ve Tekrarı	52
Matrisler ve Diziler	53
Dizileri ve Matrisleri Oluşturma.....	54
Satırlar, Sütunlar ve Boyutlar.....	56
Satır, Sütun ve Boyut İsimleri.....	58
Dizileri İndeksleme.....	59
Matrisleri Birleştirme	60
Dizi Aritmetiği	61
Özet	63
Bilginizi Test Edin: Kısa Sınav	63
Bilginizi Test Edin: Alıştırmalar	64
5. Listeler ve Data Frame'ler	65
Bölümün Amaçları.....	65
Listeler	65
Listeleri Oluşturmak.....	66
Atomik ve Özyinelemeli Değişkenler.....	69

Liste Boyutları ve Aritmetik İşlem.....	69
Listelerin İndekslenmesi	70
Vektörler ve Listeler Arasındaki Dönüştürme	74
Listelerin Birleştirilmesi	76
NULL	77
Eşlisteler.....	82
Data Frame'ler	83
Data Frame Oluşturmak	83
Data Frame İndeksleme	87
Temel Data Frame Değişimlenmesi.....	88
Özet	90
Bilginizi Test Edin: Kısa Sınav	91
Bilginizi Test Edin: Alıştırmalar	92
6. Çevreler ve Fonksiyonlar.....	93
Bölümün Amaçları.....	93
Çevreler	93
Fonksiyonlar	96
Fonksiyonları Oluşturma ve Çağırma.....	97
Fonksiyonları Diğer Fonksiyonlara Geçirme ve Diğer Fonksiyonlardan Geçirme	101
Değişken Kapsamı	104
Özet	106
Bilginizi Test Edin: Kısa Sınav	106
Bilginizi Test Edin: Alıştırmalar	107
7. Metinler ve Faktörler	109
Bölümün Amaçları.....	109
Metinler	109
Metinleri Oluşturma ve Yazdırma	110
Sayıları Biçimlendirme.....	111
Özel Karakterler	113
Harfleri Değiştirmek.....	114
Alt Metinlerin Ayıklanması.....	115
Metinleri Ayırma.....	115
Dosya Yolu	117

Faktörler	118
Faktörleri Oluşturma	118
Faktör Düzeylerini Değiştirmek	120
Faktör Düzeylerini Düşürmek	121
Sıralı Faktörler	122
Sürekli Değişkenlerin Kategorik Değişkenlere Dönüştürülmesi	123
Kategorik Değişkenleri Sürekli Değişkenlere Dönüştürme	124
Faktör Düzeylerini Üretmek	125
Faktörleri Birleştirme	126
Özet	126
Bilginizi Test Edin: Kısa Sınav	126
Bilginizi Test Edin: Alıştırmalar	127
8. Akış Kontrolü ve Döngüler	129
Bölüm Amaçları	129
Akış Kontrolü	129
if ve else	129
Vektörelleştirilmiş if	132
Çoklu Seçimler	133
Döngüler	135
repeat Döngüleri	135
while Döngüleri	137
for Döngüleri	139
Özet	141
Bilginizi Test Edin: Kısa Sınav	141
Bilginizi Test Edin: Alıştırmalar	142
9. İleri Döngüleme	145
Bölüm Amaçları	145
Tekrarlama	145
Listeler Üzerinde Dönmek	147
Diziler Üzerinde Dönülmesi	155
Çok Girdili Apply	158
Hazır Vektörelleştirme	159
Split-Apply-Combine	159

plyr Paketi	161
Özet	165
Bilginizi Test Ediniz: Kısa Sınav	165
Bilginizi Test Edin: Alıştırmalar	165
10. Paketler	169
Bölüm Amaçları	169
Paketlerin Yüklenmesi	169
Search Yolu.....	172
Kütüphaneler ve İndirilmiş Paketler	172
Paketlerin İndirilmesi	174
Paketlerin Muhafazası	176
Özet	177
Bilginizi Test Edin: Kısa Sınav	177
Bilginizi Test Edin: Alıştırmalar	177
11. Tarihler ve Zamanlar	179
Bölüm Amaçları	179
Tarih ve Zaman Sınıfları.....	179
POSIX Tarihleri ve Zamanları.....	180
Date Sınıfı.....	182
Diğer Tarih Sınıfları.....	182
Metine veya Metinden Dönüştürme.....	182
Tarihlerin Çözümlemesi.....	183
Tarihlerin Formatlanması.....	184
Zaman Dilimleri.....	184
Tarihlerle ve Zamanlarla Aritmetik	186
Lubridate	188
Özet	193
Bilginizi Test Edin: Kısa Sınav	193
Bilginizi Test Edin: Alıştırmalar	193

2. KISIM: VERİ ANALİZİ İŞ AKIŞI

12. Veri Temini	197
Bölümün Amaçları.....	197
Gömülü Veri Setleri	197
Metin Belgelerinin Okunması	198
Virgül ya da Sekme ile Ayrılmış Dosyalar	198
Yapılandırılmamış Metin Dosyaları	203
XML ve HTML Dosyaları.....	204
JSON ve YAML Dosyaları.....	206
İkili Veri Dosyalarının (Binary Files) Okunması.....	209
Excel Dosyalarının Okunması.....	210
SAS, Stata, SPSS ve MATLAB Dosyalarının Okunması	212
Diğer Dosya Tiplerinin Okunması	212
Web Verileri	213
Uygulama Programı Arayüzlü (API) İnternet Siteleri	213
İnternet Sitelerini Ayıklama.....	215
Veri Tabanlarına Erişim	218
Özet	222
Bilginizi Test Edin: Kısa Sınav	222
Bilginizi Test Edin: Alıştırmalar	223
13. Temizleme ve Dönüştürme	225
Bölümün Amaçları.....	225
Metinleri Temizleme.....	225
Data Frame Manipülasyonu	232
Sütun Ekleme ve Değiştirme	232
Kayıp Değerlerin Üstesinden Gelmek.....	233
Geniş ve Uzun Formların Birbirlerine Dönüştürülmesi.....	235
SQL Kullanımı	236
Sıralama	237
Fonksiyonel Programlama (Functional Programming)	239
Özet	242
Bilginizi Test Edin: Kısa Sınav	242
Bilginizi Test Edin: Alıştırmalar	243

14. Keşfetme ve Görselleştirme	245
Bölümün Amaçları.....	245
İstatistiksel Özetler.....	245
Üç Grafik Çizim Sistemi.....	250
Saçılım Grafikleri	251
Deneme 1: base Grafik Sistemi	252
Deneme 2: lattice Grafik Sistemi.....	257
Deneme 3: ggplot2 Grafik Sistemi	263
Çizgi Grafikleri.....	269
Histogram.....	277
Kutu Grafiği	288
Çubuk Grafikleri	292
Diğer Çizim Ek Paketleri ve Sistemleri	299
Özet	300
Bilginizi Test Edin: Kısa Sınav	301
Bilginizi Test Edin: Alıştırmalar	301
15. Dağılımlar ve Modelleme	303
Bölüm Amaçları	303
Rastgele Sayılar	304
sample Fonksiyonu	304
Dağılımlardan Örneklemeye.....	305
Dağılımlar	307
Formül	307
İlk Model: Doğrusal Regresyon.....	309
Modelleri Karşılaştırma ve Güncelleme.....	312
Modellerin İncelenmesi ve Çizilmesi	320
Diğer Model Türleri	325
Özet	326
Bilginizi Test Edin: Kısa Sınav	327
Bilginizi Test Edin : Alıştırmalar	327
16. Programlama	329
Bölümün Amaçları.....	329
Mesajlar, Uyarılar ve Hatalar	329
Hatalarla Başa Çıkma	333

Hata Ayıklama	337
Test etme.....	340
RUnit.....	340
testthat	344
Sihir	345
Metnin Koda Dönüştürülmesi	345
Kodun Metine Dönüştürülmesi	347
Nesne Tabanlı Programlama	349
S3 Sınıfları	349
Referans Sınıfları	351
Özet	358
Bilginizi Test Edin: Kısa Sınav	358
Bilginizi Test Edin: Alıştırmalar	359
17. Paket Oluşturma	361
Bölüm Amaçları	361
Neden Paket Oluştururuz?.....	361
Ön Gereklilikler.....	362
Paket Dizin Yapısı.....	362
İlk Paketiniz	363
Paket Dokümantasyonu	366
Paketlerin Kontrolü ve Oluşturulması	370
Paketlerin Bakımı.....	371
Özet	373
Bilginizi Test Edin: Kısa Sınav	374
Bilginizi Test Edin: Alıştırmalar	374
3. KISIM: EKLER	
EK-A	377
EK-B.....	380
EK-C	382
EK-D	391
Kaynakça	424
Dizin	425

1. KISIM

R DİLİ

1. BÖLÜM

GİRİŞ

Tebrikler! Bir R programcısı olmak için yolculuğa başlıyorsunuz. Bu bölümde zihninizi zorlayacak herhangi birşey olmayıp sizi ısındırarak başlatacak bir içerik vardır. Kodlamaya başlamadan önce, R'in ne olduğu, nasıl indirileceği ve R ile nasıl çalışılacağı hakkında konuşacağız. Bundan sonra ilk programınızı yazmayı deneyebilir ve nasıl yardım alabileceğinizi öğrenebilirsiniz.

Bölümün Amaçları

Bu bölümü okuduktan sonra aşağıdakileri yapabilirsiniz:

- R kullanarak yapabileceğiniz hakkında birşeyler bilebilirsiniz.
- R'in ve R ile çalışılacak bir IDE'nin nasıl indirileceğini bilirsiniz.
- R'da basit bir program yazabilirsiniz.
- R'da nasıl yardım alınabileceğini bilirsiniz.

R Nedir?

Hemen kafanızı karıştırmak için şunu söyleyeyim: R iki şey belirtir. Bunlardan birincisi programlama dili için olan R ve diğeri de R'da yazılmış olan programları çalıştıracak yazılım olan R'dır. Neyse ki çoğu zaman hangi R'in kullanıldığı içerikten net olarak anlaşılır.

R (dil olan) 1990lar öncesinde University of Auckland'da çalışan Ross Ihaka ve Robert Gentleman tarafından yaratıldı. Temel olarak 1970'lerde Bell Laboratuvarları'nda geliştirilen John Chambers'ın başını çektiği S diline dayanır. R (yazılım olan) bir GNU projesidir. Kendini önemli, serbest ve açık kaynak kodlu bir yazılım olarak gösterir. Hem dil olan hem de yazılım olan R (şimdilerde) R Core Team olarak bilinen 20 kişi tarafından geliştirilmektedir.

R'in geçmişinin 1970'lere kadar dayanması önemlidir çünkü bu zamana yayılmış bir evrimdir. Acele ve hızlı (örneğin "yaratılmışlık"¹ hissi veren Microsoft'un .NETFramework'leri gibi) yapılmamıştır. Evrim süreci yaşam formlarında olduğu gibi bazı garipliklere ve tutarsızlıklara yol açmıştır. R'in daha serbest formdaki (freelicense) doğasının iyi yönü; eğer R'da yapılan herhangi bir şeyi beğenmezseniz, kendi istediğiniz şekilde yapabilecek bir paket yazabilmenizdir. Birçok insan bunu çoktan yaptı; artık günümüzdeki yaygın soru "Ben bunu R'da yapabilir miyim?" değil, "Bu üç uygulamadan hangisini kullanmalıyım?" şeklindedir.

R yorumlanan bir dildir (bazen betik dili olarak da adlandırılır). Bu kodunuzun çalışmadan önce derleme gerektirmediği anlamına gelir. Bu, yüksek-seviyeli bir dil olduğu için kodlarınızı çalıştırdığınız bilgisayarınızın içindeki işlemlere erişim hakkınız olmaz, herşey verinizi analiz etmenize yardım etmeye ayarlanır.

R programlama paradigmasının bir karışımını destekler. Özünde tanımlamalı bir dildir (hesaplamanın birini yaptıktan sonra diğerini yapan bir betik) ama nesne tabanlı programlamayı destekler. (Veri ve fonksiyonlar sınıflar içinde birleştirilmiştir) ve fonksiyonel programlamayı da (fonksiyonlar *birinci sınıf nesnelere*dir, bunlara diğer nesnelere davrandığınız gibi davranabilirsiniz ve bunları özyinelemeli olarak çağırabilirsiniz.) Bu programlama stilleri karışımı R kodunun diğer birçok dil ile benzerlik taşıdığı anlamına gelir. Küme parantezi C'ye benzer tanımlamalı kod yazabileceğiniz anlamına gelir. (Bölüm 2'de tartışacak olduğumuz R'in vektörel doğası gereğince daha az döngünüz olacaktır.) Eğer referans sınıflarını kullanırsanız, C# veya Java kodlarına biraz benzer nesne tabanlı kod yazabilirsiniz. Fonksiyonel programlama yapıları Lisp'ten esinlenilmiştir (değişken takip kuralları LispScheme diyalektiğindedir) ama daha az parantez vardır. Bütün bunlar R'in Perlethos'u takip ettiğini söylemenin dolambaçlı yollarıdır:

Birşeyi yapabilmenin birden fazla yolu vardır.

–Larry Wall

R'in İndirilmesi

Eğer bir Linux makinesi kullanıyorsanız büyük ihtimalle paket yöneticinizde güncel olmayan bir R vardır. Bunun dışındaki herkes R indirmek için ilk önce <http://www.r-project.org> adresine gitmelidir. Eski görünümü² web sayfası sizi aldatmasın, bu R'in kalitesini yansıtmaz. Sayfanın altında yer alan "Getting Started" bölümündeki "download R" linkinine tıklayın.

1 Akıllıca dizayn edilmiş

2 Wayback Machine İnternet arşivleri Mayıs 2004'ten bu yana değişmediğini söylemektedir.