

R ile Programlama

Burak ÖZKAN

Dr. Yalçın ÖZKAN



© Papatya Yayıncılık Eğitim

Ankara Caddesi, Prof. Fahreddin Kerim Gökay Vakfı İşhanı Girişi
No: 11/6, Çağaloğlu (Fatih) / İstanbul

Tel : (+90 212) 527 52 96 (+90 532) 311 31 10
Faks : (+90 212) 527 52 97
e-mail : admin@papatyabilim.com.tr
Web : www.papatyabilim.com.tr
Dağıtım : TDK Bilim – www.tdk.com.tr

R ile Programlama – Burak ÖZKAN ve Dr. Yalçın ÖZKAN

1. Basım Nisan 2017

Yayın Danışmanı : Dr. Toros Rifat ÇÖLKESEN (Post-Edu Institute)
Üretim : Necdet AVCI
Pazarlama : Mustafa DEMİR
Satış : TDK Bilim www.tdk.com.tr
Sayfa Düzenleme : Müge URCAN, Papatya ve Kelebek Tasarım
Basım ve Ciltleme : Özkaracan Matbaacılık (Sertifika No: 12228)
Evren Mah. Gülbahar Cad. No:62 Güneşli/İstanbul

©Bu kitabın her türlü yayın hakkı Papatya Yayıncılık Eğitim AŞ'ye aittir.
Yayınevinden yazılı izin alınmaksızın alıntı yapılamaz, kısmen veya tamamen hiçbir şekil ve teknikle ÇOĞALTILAMAZ, BASILAMAZ, YAYIMLANAMAZ. Kitabın, tamamı veya bir kısmının fotokopi makinesi, ofset gibi teknikle çoğaltılması, hem çoğaltan hem de bulunduranlar için yasadışı bir davranıştır.

Özkan, Burak ve Özkan, Yalçın

R ile Programlamada / Burak Özkan ve Yalçın Özkan - İstanbul: Papatya Yayıncılık Eğitim, 2017.

xiv, 336 s.; 24 cm.

Kaynakça ve Dizin var.

Sertifika No: 11218

ISBN 978-605-9594-20-2

1. Programlama Dilleri. 2. R ile İstatistik 3. İstatistiksel Programlama 4. R Dili.

I. Title

Anneme, sonsuz saygılarımla...

Burak ÖZKAN

Eşime, sonsuz sevgilerimle...

Yalçın ÖZKAN

*Bu kitabın yayınlanması
konusunda desteklerinden ötürü
Sayın Toros Rifat ÇÖLKESEN'e ve
emeđi geçen tüm
Papatya Bilim Yayınevi
çalışanlarına çok teşekkür ediyoruz.*

İçindekiler

Önsöz	xv
Bölüm 1. R Programlama Diline Giriş	15
1.1 Bu Bölümün Amacı	15
1.2 Konuyla ilgili Temel Sorular	15
1.3 R Programlama Dili	16
1.4 R'nin Kurulumu	16
1.5 R Arayüzü	17
1.5.1 Dosya	17
1.5.2 Düzenle	18
1.5.3 Görünüm	19
1.5.4 Diğer	19
1.5.5 Paketler	19
1.5.6 Pencereleler	19
1.5.7 Yardım	20
1.5.8 Komut Satırı	20
1.6 Atama Operatörü	20
1.7 Aritmetik İşlemler	21
1.8 R Nesne Değişkenleri	22
1.9 Veri Türleri	23
1.9.1 Kesirli Veri	23
1.9.2 Tamsayı Veri	24
1.9.3 Karakter Dizisi	26
1.9.4 Mantıksal Veri	27
1.9.5 Karmaşık Veri Türü	30
1.10 Hazır R Veri Kümleri	31
1.10.1 Veri Kümelerinin Görüntülenmesi	32
1.10.2 En Üstteki Satırların Görüntülenmesi	33
1.10.3 Öznitelikler	34
1.10.4 Satır ve Sütün Sayısı	36
1.11 R Paketleri	37
1.11.1 Paketlerin Kurulumu	37
1.11.2 Paketlerin Listesi	39
1.11.3 Paketlerin Oturuma Tanıtılması	40
1.12 Yardım Araçları	40
1.12.1 Yardım Sayfaları	40
1.12.2 Kısa Kitapçıklar	42
1.12.3 Demolara Erişim	44

1.13	R Çalışma Alanı	44
1.13.1	Nesne Listesiyle İlgili İşlemler	44
1.13.2	Çalışma Dizini	45
1.13.3	Seçenekler	45
1.13.4	Çalışma Alanı Görüntüsünün Kaydedilmesi	46
1.14	Özet	46
1.5	Gözden Geçirme Soruları	47

Bölüm 2. Vektör ve Faktörler **49**

2.1	Bu Bölümün Amacı	49
2.2	Konuyla İlgili Temel Sorular	49
2.3	Vektör	49
2.3.1	Vektör İndeksi	51
2.3.2	Vektör Aritmetiği	52
2.3.3	Mantıksal Vektörler	53
2.3.4	Karakter Vektörler	53
2.3.5	Vektör ve Elemanlarının Uzunluğu	54
2.3.6	Vektörlerin Birleştirilmesi	54
2.3.7	Vektör Üyelerini İsimlendirme	55
2.3.8	Vektörlerde Sorgulama	55
2.3.9	Vektör Elemanlarının Güncellenmesi	56
2.3.10	Vektörün Tüm Elemanlarının Yok Edilmesi	57
2.3.11	Vektör Elemanlarının Toplanması	57
2.3.12	Vektörün En Büyük ve En Küçük Elemanı	57
2.3.13	Vektörün Ortalaması ve Medyanı	58
2.3.14	Vektörün Sıralanması	58
2.4	Faktör	59
2.4.1	Faktörlerin Yaratılması	59
2.4.2	Vektörü Faktöre Dönüştürme	60
2.4.3	Faktör Elemanlarına Erişilmesi ve Değişiklik Yapılması	60
2.4.4	Faktöre Yeni Bir Eleman Ekleme ve Silme	61
2.4.5	Düzeylerde Değişiklik Yapılması	61
2.4.6	Düzy Frekansları	62
2.4.7	Düzy Sırası	63
2.4.8	Aritmetik İşlemler	63
2.5	Özet	64
2.6	Gözden Geçirme Soruları	64

Bölüm 3. Dizi ve Matrisler **67**

3.1	Bu Bölümün Amacı	67
3.2	Konuyla İlgili Temel Sorular	67

3.3	Diziler	68
3.4	Matrisler	69
3.4.1	Matris Oluşturma	69
3.4.2	Vektörleri Birleştirerek Matris Elde Etme	71
3.4.3	Vektörleri Boyutlandırarak Matris Elde Etme	71
3.4.4	Satır ve Sütunları İsimlendirme	72
3.4.5	Matris Elemanlarına Erişim	73
3.4.6	Bir Koşula Uyan Matris Elemanlarının Bulunması	74
3.4.7	İsmlendirilmiş Matris Elemanlarına Erişim	74
3.4.8	Matris Elemanlarının Değiştirilmesi	75
3.4.9	Matris Elemanlarının Silinmesi	76
3.4.10	Mevcut bir Matrise Yeni bir Vektör Eklenmesi	77
3.4.11	İki Matrisin Birleştirilmesi	77
3.4.12	Matrisin En Büyük ve En Küçük Elemanları	78
3.4.13	Satır ve Sütun Toplamları	79
3.4.14	Satır ve Sütun Ortalamaları	80
3.4.15	Satır ve Sütunlara Fonksiyonun Uygulanması	
3.5	Matrislerle Cebirsel İşlemler	81
3.5.1	Köşegen İşlemleri	82
3.5.2	Birim Matris	83
3.5.3	Matris Transpozesi	83
3.5.4	Matrislerin Toplanması	84
3.5.5	Matrisin Skaler Değerle Çarpımı	84
3.5.6	Matrislerin Çarpımı	85
3.5.7	Matris Determinantı	86
3.5.8	Matrisin Tersi	87
3.5.9	Matrislerle Regresyon Analizi	88
3.5.10	Doğrusal Denklem Sistemlerinin Çözümü	89
3.6	Özet	91
3.7	Gözden Geçirme Soruları	91

Bölüm 4. Listeler **93**

4.1	Bu Bölümün Amacı	93
4.2	Konuyla İlgili Temel Sorular	93
4.3	Listelerin Yaratılması	94
4.4	Vektörün Liste Olmaya Zorlanması	97
4.5	Boş Liste	98
4.6	Liste Üyelerine Erişim	98
4.7	İsmlendirilmiş Liste Üyeleri	98
4.7.1	Liste Dilimleme	99
4.7.2	Üyelere Erişim	100
4.8	Liste Üyelerinin Doğrudan Kullanımı	101

4.9	Liste Üzerinde İşlemler	101
4.9.1	Listeye Üye Ekleme	102
4.9.2	Listedeki Bir Üyeyi Silme	102
4.9.3	Listenin Bir Üyesini Güncelleme	103
4.10	Liste Üyelerine Fonksiyonların Uygulanması	104
4.11	Özet	105
4.12	Gözden Geçirme Soruları	105
Bölüm 5. Veri Çerçevesi		107
5.1	Bu Bölümün Amacı	107
5.2	Konuyla İlgili Temel Sorular	107
5.3	Veri Çerçevesi	108
5.3.1	Satır ve Sütun İsimlerinin Değiştirilmesi	108
5.3.2	Veri Görüntüleme	109
5.3.3	Veri Çerçevesi Özellikleri	110
5.3.4	Veri Çerçevesi Elemanlarına Erişim	111
5.3.5	Koşullu Erişim	112
5.3.6	Veri Çerçevesi Boyutları	113
5.4	Veri Çerçevesi İşlemleri	113
5.4.1	Sütun Ekleme	113
5.4.2	Veri Çerçevelerinin Birleştirilmesi	114
5.4.3	Araya Sütun Ekleme	115
5.4.4	Satır Ekleme	116
5.4.5	Satır veya Sütun Silme	116
5.4.6	Eleman Değerini Değiştirme	118
5.4.7	Sütunlar Üzerinde Aritmetik İşlemler	118
5.4.8	Veri Çerçevesi Sütunlarının Doğrudan Kullanılması	119
5.4.9	Fonksiyonların Uygulanması	120
5.4.10	Matris, Liste ve Veri Çerçevesi Dönüşümleri	121
5.5	Özet	122
5.6	Gözden Geçirme Soruları	122
Bölüm 6. Karakter Dizileri		125
6.1	Bu Bölümün Amacı	125
6.2	Konuyla İlgili Temel Sorular	125
6.3	Karakter Dizileri	126
6.3.1	Boş Karakter Dizisi	126
6.3.2	Karakter Sayısını Öğrenme	127
6.3.3	Karakter Dizisi Olup Olmadığını Öğrenme	127
6.3.4	Karakter Dizilerini Görüntülemek	128

6.4	Karakter Dizileri ile İşlemler	128
6.4.1	Bir Nesneyi Karakter Dizisine Dönüştürmek	128
6.4.2	Karakter Dizilerinin Birleştirilmesi	130
6.4.3	Küçük ve Büyük Harflere Dönüştürme	131
6.4.4	Kısaltma	132
6.4.5	Karakterlerin Değiştirilmesi	133
6.4.6	Alt Dizilerle İşlemler	134
6.5	Konsola Yazdırma İşlemleri	136
6.5.1	Basit Yazdırma İşlemi- Print() Fonksiyonu	136
6.5.2	Birleştirerek Yazdırma- Cat() Fonksiyonu	137
6.5.3	Formatlayarak Yazdırma İşlemi- Format() Fonksiyonu	138
6.5.4	Karakter Dizisine Dönüştürerek Yazdırma- ToString() Fonk.	140
6.5.5	C Stili Formatlama - Sprintf() Fonksiyonu	141
6.6	Örüntü Arama	143
6.6.1	Konumu Bulma	143
6.6.2	Karakter Dizisini Bulma	144
6.6.3	Bulunamayan Karakter Dizileri	144
6.6.4	Mantıksal Sonuçlar	144
6.6.5	Örüntünün Tam Adresi	145
6.6.6	Örüntüyü Değiştirme	145
6.7	Özet	146
6.8	Gözden Geçirme Soruları	147

Bölüm 7. Veri Giriş Çıkış İşlemleri **149**

7.1	Bu Bölümün Amacı	149
7.2	Konuyla İlgili Temel Sorular	149
7.3	Veri Girişleri	150
7.4	Hazır Veri Kümeleri	151
7.5	Terminalden Veri Girişleri	151
7.5.1	Terminalden Veri Girişleri ve Çıkışları	151
7.5.2	Terminalden Bir Satır Girilmesi	153
7.6	Dosyaya Yazma ve Okuma İşlemleri	153
7.6.1	Dosya ve Dizin Hakkında Bilgi Edinme	154
7.6.2	Dosyaya Yazdırma	155
7.6.2.1	Veriyi Dosyaya Yönlendirme	155
7.6.2.2	Doğrudan Bir Dosyaya Kayıt	157
7.6.2.3	CSV Dosyaya Kayıt	158
7.6.2.4	Tabular Verinin Dosyaya Kaydedilmesi	159
7.6.2.5	R Dosyalarının Yaratılması	160
7.7	Dosyadan Okuma	161
7.7.1.1	scan() Fonksiyonu ile Okuma	161
7.7.1.2	CSV Dosyalarının Okutulması	163

7.7.1.3	Tabular Verinin Okutulması	164
7.7.1.4	Panodan Veri Okuma	165
7.8	Özet	166
7.9	Gözden Geçirme Soruları	166
Bölüm 8. Program Denetimi		169
8.1	Bu Bölümün Amacı	169
8.2	Konuyla İlgili Temel Sorular	169
8.3	Denetim Yapıları	170
8.3.1	Karşılaştırma ve Mantıksal Operatörler	170
8.3.2	if Deyimi ile Karşılaştırma	171
8.3.3	Aksi Takdirde - if... else Deyimi	173
8.3.4	İççe if Deyimleri	175
8.3.5	if else() Fonksiyonu	176
8.4	Döngüler	177
8.4.1	for Döngüsü	177
8.4.2	while Döngüsü	180
8.4.3	Döngüden Çıkış	182
8.4.4	Döngüye Devam - Next Deyimi	183
8.4.5	repeat Döngüsü	185
8.4.6	İç İçe Döngüler	187
8.4.7	Döngü Performansı ve Vektörleştirme	188
8.4.8	Gizli Döngüler -Apply() Fonksiyon Ailesi	190
8.5	Listeden Seçim İşlemleri - switch() Fonksiyonu	192
8.6	Özet	193
8.7	Gözden Geçirme Soruları	194
Bölüm 9. Fonksiyonlar		195
9.1	Bu Bölümün Amacı	195
9.2	Konuyla İlgili Temel Sorular	195
9.3	Kullanıcı Tanımlı Fonksiyonlar	196
9.3.1	Fonksiyonların Yaratılması	196
9.3.2	Fonksiyonların Çalıştırılması	197
9.3.3	Bir Fonksiyondan Bir Başka Fonksiyonun Çağrılması	197
9.3.4	Fonksiyon Çalıştıktan Sonra İşlem Sırası	198
9.3.5	Fonksiyonun Değer Döndürmesi	199
9.3.6	Liste Dönüşler	200
9.3.7	Fonksiyonlara Parametre Geçirme	201
9.3.8	Kendini Çağırın Fonksiyonlar	202

9.4	Yerleşik Hazır Fonksiyonlar	203
9.4.1	Matematik Fonksiyonları	203
9.4.1.1	Trigonometrik İşlemler	204
9.4.1.2	Logaritma İşlemleri	205
9.4.1.e	Sayısının Kuvvetleri	205
9.4.1.4	Karekök İşlemi	206
9.4.1.4	Mutlak Değer İşlemi	207
9.4.1.5	Yuvarlatma İşlemi	207
9.4.1.6	Kesme İşlemi	208
9.4.2	İstatistik Fonksiyonları	209
9.4.2.1	Ortalama Hesabı	209
9.4.2.2	Standart Sapma ve Varyans Hesabı	210
9.4.2.3	Medyanın Bulunması	211
9.4.2.4	Değişim Aralığının Bulunması	212
9.5	Özet	212
9.6	Gözden Geçirme Soruları	212

Bölüm 10. Grafikler **215**

10.1	Bu Bölümün Amacı	215
10.2	Konuyla İlgili Temel Sorular	215
10.3	Grafiklerin Yaratılması	216
10.4	Grafiğin Kaydedilmesi	218
10.5	Dağılım Grafikleri	220
10.6	Nokta Grafik	223
10.7	Çizgi Grafikler	224
10.7.1	Diğer Çizgi Grafiklerin Eklenmesi	225
10.7.2	Yatay Eksen Etiketleri	226
10.8	Histogram ve Yoğunluk Grafikleri	228
10.8.1	Histogram Çizdirme	229
10.8.2	Yoğunluk Grafikleri	231
10.8.3	Ortak Histogram ve Yoğunluk Grafiklerinin Çizdirilmesi	223
10.9	Çubuk Grafikler	234
10.9.1	Basit Çubuk Grafikler	235
10.9.2	Frekansları Kullanan Çubuk Grafikler	237
10.9.3	Yatay Çubuk Grafikler	239
10.9.4	Yığılı Çubuk Grafikler	239
10.9.5	Gruplandırılmış Çubuk Grafikler	241
10.10	Pasta Grafikler	242
10.10.1	Basit Pasta Grafiklerinin Çizdirilmesi	243
10.10.2	Yüzelere Göre Pasta Grafikler	244
10.11	Kutu Grafikleri	245
10.11.1	Bir Değişken İçin Kutu Grafik	246

10.11.2 Bir Grup Değişken İçin Kutu Grafik	248
10.12 Matematiksel Fonksiyonların Grafikleri	249
10.12.1 Parabol Grafiği	249
10.12.2 Logaritmik Eksenler	251
10.12.3 Trigonometrik Fonksiyonların Grafiği	252
10.13 Grafikleri Bir Arada Çizdirme	254
10.14 Özet	256
10.15 Gözden Geçirme Soruları	256

Bölüm 11. R ile İstatistiğe Giriş **259**

11.1 Bu Bölümün Amacı	259
11.2 Konuyla İlgili Temel Sorular	259
11.3 Merkezi Eğilim Ölçüleri	259
11.3.1 Aritmetik Ortalama	260
11.3.2 Geometrik Ortalama	262
11.3.3 Harmonik Ortalama	263
11.3.4 Kareli Ortalama	263
11.3.5 Medyan	264
11.3.6 Mod	265
11.4 Dağılım Ölçüleri	266
11.4.1 Ortalama Sapma	267
11.4.2 Varyans	267
11.4.3 Standart Sapma	268
11.4.4 Değişkenlik Katsayısı	269
11.5 Olasılık Dağılımları	270
11.5.1 Normal Dağılım	270
11.5.1.1 Yoğunluk FonksiyonU	270
11.5.1.2 Kümülatif Dağılım Fonksiyonu	272
11.5.1.3 Normal Dağılıma Sahip Veri Üretme	273
11.5.2 Binom Dağılımı	274
11.5.2.1 Yoğunluk Fonksiyonu	274
11.5.2.2 Kümülatif Olasılık	276
11.5.2.3 Binom Dağılıma Sahip Veri Üretme	277
11.6 Hipotez Testleri	278
11.6.1 Hipotezin Belirlenmesi	278
11.6.2 Hipotezin Kritik Bölgesi	279
11.6.3 Test Değeri	280
11.6.4 z-test Değerinin Yorumlanması	280
11.7 Regresyon Analizi	282
11.7.1 Verinin Hazırlanması	283
11.7.2 Regresyon Modelinin Elde Edilmesi	285
11.7.3 Model İle İlgili Bilgiler	285

11.7.4	Determinasyon Katsayısı	286
11.7.5	Öngörüler	287
11.7.6	Standart Artıklar	288
11.8	Özet	289
11.9	Gözden Geçirme Soruları	200
Bölüm 12. R ile Veri Madenciliği		293
12.1	Bu Bölümün Amacı	293
12.2	Konuyla İlgili Temel Sorular	293
12.3	Veri Madenciliği	294
12.4	Veri Dönüştürme	294
12.4.1	Min-Max Normalleştirilmesi	294
12.4.2	Z-Score Normalleştirilmesi	296
12.5	Sınıflandırma Algoritmaları	297
12.5.1	Modelin Doğruluk Oranı	297
12.5.2	C4.5 Sınıflandırma Algoritması	299
12.5.3	Yapay Sinir Ağları	305
12.5.4	Bayes Sınıflandırıcılar	310
12.6	Kümeleme Algoritmaları	314
12.6.1	Uzaklık Ölçütleri	314
12.6.2	Hiyerarşik Kümeleme	315
12.6.3	k-Ortalamalar Algoritması	318
12.7	Birliktelik Kuralları	320
12.8	Özet	325
12.9	Gözden Geçirme Soruları	326
Kaynakça		329
Dizin		333

Önsöz

R programlama dilinin günümüzde özellikle akademik çevreler başta olmak üzere geniş bir uygulama alanına sahip olduğunu söyleyebiliriz. Bu dilin güçlü veri işleme araçlarına sahip olması ve açık kaynak kodlu olması nedeniyle geniş bir kullanıcı kesimi oluşmuştur. Özellikle ilişkiel veri tabanları ile entegre olabilmesi nedeniyle, sadece akademik çevrelerde kullanılmamakta, kurumsal uygulamalara da dahil olduğu görülmektedir.

Elinizdeki bu kitap R programlama dili için bir başlangıç kitabıdır. Bu nedenle konuları oldukça basit düzeyde ele alarak anlaşılır olmasına özen gösterdik. Her konuyu çok sayıda örnekle desteklemeye çalıştık. Kitabın birinci konusu R programlama diline bir giriş yapmaktadır. R arayüzünün ve yardımcı geliştirme araçlarının nasıl kullanılacağı konusu üzerinde durulmaktadır.

Kitabın beşinci bölüme kadar olan kısımlarında R'nin temel veri yapıları üzerinde durulmaktadır. Vektör, faktör, dizi, matris, liste ve veri çerçevesi gibi kavramlar bu bölümlerde ele alınmaktadır. Altıncı bölümde ise karakter dizilerinin R'de nasıl kullanılacağı anlatılmaktadır.

Yedinci bölümde veri giriş çıkışları ele alınarak özellikle dosyalarla çalışma konusu üzerinde durulmaktadır. Sekizinci bölümde programlarda yer alan denetim ve döngü işlemleri açıklanmaktadır. Onuncu bölümde ise grafiklere yer verilerek örneklerle çeşitli grafiklerin nasıl çizdirilebileceği üzerinde durulmaktadır.

Son iki bölümde ise R'nin istatistikte ve veri madenciliğinde kullanılması anlatılmaktadır. Bu iki alanda R'nin yaygın biçimde kullanıldığı bilinmektedir. Her iki konunun kavramsal temellerine inmeden basit uygulamalarla R'nin kullanımı üzerinde durulmaktadır.

Bir programlama dilinin yeterince kod yazarak öğrenilebileceği düşüncesindeyiz. Okuyucunun R programını bilgisayarına kurduktan sonra kitaptaki tüm örnekleri sırasıyla uygulamasını öneriyoruz. Kitap içinde gözden kaçmış hatalar olabilir. Bu konuda okuyucunun eleştirilerini ve uyarılarını bekliyoruz.

Burak ÖZKAN

Yalçın ÖZKAN