

R Programlama Dili ile

Temel İstatistikler ve Raporlama

Mehmet Taha ESER

Gökhan AKSU

Cem Oktay GÜZELLER

2. Baskı





Dr. Mehmet Taha ESER - Dr. Gökhan AKSU - Prof. Dr. Cem Oktay GÜZELLER

**R Programlama Dili ile
TEMEL İSTATİSTİKLER VE RAPORLAMA**

ISBN 978-605-241-652-5

DOI 10.14527/9786052416525

Kitap içeriğinin tüm sorumluluğu yazarlarına aittir.

© 2019, 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 yayineimize bilgi vermesini ve bandrolsüz yayınları satın almamasını diliyoruz.

Pegem Akademi Yayıncılık, 1998 yılından bugüne uluslararası düzeyde düzenli faaliyet yürüten **uluslararası akademik bir yayinevidir**. 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 ulaşabilmektedir.

1. Baskı: Mayıs 2019, Ankara

2. Baskı: Aralık 2019, Ankara

Yayın-Proje: Şehriban Türüldür

Dizgi-Grafik Tasarım: Tuğba Kaplan

Kapak Tasarımı: Pegem Akademi

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

İstanbul Cad. İstanbul Çarşısı 48/48 İskitler - Ankara

(0312 341 36 67)

Yayıncı Sertifika No: 36306

Matbaa Sertifika No: 25931

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

Dr. Mehmet Taha ESER

2007 yılında Gazi Üniversitesi Eğitim Fakültesi Sınıf Öğretmenliği Bölümünden mezun olmuş, 2011 yılında Hacettepe Üniversitesi Eğitimde Ölçme ve Değerlendirme Alanında yüksek lisansını tamamlamıştır. 2014 yılında Hacettepe Üniversitesinde Eğitimde Ölçme ve Değerlendirme Alanında doktora eğitimine başlayan yazar, 2019 yılında doktora eğitimini tamamlamıştır. Çalışma konuları Ölçme ve Değerlendirme, Veri Madenciliği, Makine Öğrenimi, Çok Değişkenli İstatistik ve Programlama olan araştırmacı Akdeniz Üniversitesi İstatistik Danışma Merkezi'nde Öğretim Görevlisi olarak çalışmaktadır.

İletişim Adresi: Akdeniz Üniversitesi, Konyaaltı Mevkii, Antalya / Eposta: ta-haeser@akdeniz.edu.tr

Dr. Gökhan AKSU

2007 yılında Dokuz Eylül Üniversitesi Ortaöğretim Matematik Öğretmenliği Bölümünden mezun olmuş, 2012 yılında Adnan Menderes Üniversitesi Eğitim Programları ve Öğretim alanında yüksek lisansını tamamlamıştır. 2013 yılında Hacettepe Üniversitesinde Eğitimde Ölçme ve Değerlendirme alanında doktora eğitimine başlayan yazar, 2018 yılında doktora eğitimini tamamlamıştır. Çalışma konuları Ölçme ve Değerlendirme, PISA, Büyük veri, Veri Madenciliği, Çok Değişkenli İstatistik ve Programlama olan araştırmacı Adnan Menderes Üniversitesinde İstatistik ve Olasılık, İstatistik-1, İstatistik-2, Psikolojide Ölçme ve Ölçek Geliştirme gibi derslere girmektedir.

İletişim Adresi: Adnan Menderes Üniversitesi, Aytepe Mevkii, Aydın / Eposta: gokhanaksu@adu.edu.tr

Prof. Dr. Cem Oktay GÜZELLER

1993 yılında Gazi Üniversitesi Eğitim Fakültesi Bilgisayar Öğretmenliğinden, 1999 yılında Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi İşletme Bölümünden mezun olmuştur. 2005 yılında Hacettepe Üniversitesi Eğitimde Ölçme ve Değerlendirme Alanında doktora eğitimini tamamlamış, 2012 yılında Ölçme ve Değerlendirme alanında doçentlik, 2018 yılında profesör ünvanını almıştır. Çalışma konuları istatistiksel analiz ve psikometri olan araştırmacı Akdeniz Üniversitesi'nde Öğretim Üyesi olarak çalışmaktadır.

İletişim Adresi: Akdeniz Üniversitesi, Konyaaltı Mevkii, Antalya / Eposta: cemguzeller@akdeniz.edu.tr

ÖN SÖZ

R istatistiksel hesaplamalarda, analizlerde, grafik oluşturmada ve raporlamada son yıllarda oldukça popüler olan bir programlama dilidir. R başta UNIX, Windows ve Mac gibi birçok işletim sistemi ile uyumlu haldedir. R programlama dili Ross Ihaka ve Robert Gentleman tarafından 1990'lı yıllarda geliştirilmiştir. Yazılım ismini yazarların isimlerinin baş harflerinden almaktadır. R ile analiz yapmanın en büyük avantajlarından biri açık kaynak kodlu ve ücretsiz olmasıdır. Önceleri R Çekirdek Grup (R CoreGroup) ve daha sonraları R Vakfı (R Foundation) olarak birçok yazılımcının bir araya gelmesiyle R topluluğu zamanla büyük bir aile haline gelmiştir.

R programla diliyle analizlerin gerçekleştirilebileceği en ideal ortamlardan biri RStudio'dur. Araştırmacılar daha önce farklı yazılımcılar ve istatistikçiler tarafından oluşturulmuş olan paketler yardımıyla analizlerini kolaylıkla gerçekleştirebilmektedir. Programlama konusunda bilgi sahibi olan araştırmacılar ise ilgili analize ilişkin fonksiyon R Vakfına göndererek ilgili önerinin kabul edilmesi durumunda dünya çapında R kullanan tüm araştırmacılar ile önerdikleri paket ya da fonksiyonu paylaşmış olacaktırlar. Bu kitap kapsamında R programlama dilinin ne olduğu, R Studio'nun nasıl kullanılacağı, RStudio ile betimsel ve anlam çıkarıcı istatistiklerin nasıl gerçekleştirileceği anlatılmaktadır. Özellikle lisans ve lisansüstü düzeyinde eğitim alan araştırmacıların bilgisayar karşısına geçerek tüm analizleri kendilerinin gerçekleştirebilmesi amacıyla analiz aşamasına ilişkin ekran görüntüleri ile analiz sonuçları ayrı ayrı rapor edilmiştir. Bu sayede eğitim bilimleri ile sosyal bilimlerde araştırma yapmak isteyenlere açık kaynak kodlu, ücretsiz ve kullanıcı dostu bir yazılım olan RStudio'nun nasıl kullanılacağı gösterilmeye çalışılmıştır. Özellikle genç bilim insanlarının temel düzeydeki istatistik bilgileri ile R Studio programlama dilinin özelliklerini birleştirmeleri sonucunda; zaman, emek ve para açısından hem ülkemizin kaynakları bakımından hem de kişisel kaynakların etkin ve verimli kullanımı açısından önemli olduğunu düşünmekteyiz.

Kitabın yenilenen ikinci baskısına Tekrarlı Ölçümler için ANOVA ve Friedman testleri eklenerek en az üç veya daha fazla zaman diliminde elde edilmiş ölçme sonuçları arasındaki farklılıkların da öğrenciler tarafından R ortamında test edilmesi amaçlanmıştır. Tüm öğrenci ve genç akademisyenlere faydalı olması ümidiyle...

Dr. Mehmet Taha ESER

Dr. Gökhan AKSU

Prof. Dr. Cem Oktay GÜZELLER

İÇİNDEKİLER

Ön Söz.....	v
-------------	---

1. BÖLÜM

R İSTATİSTİKSEL YAZILIM GELİŞTİRME ORTAMINA GİRİŞ

R İstatistiksel Yazılım Geliştirme Ortamı ve Özellikleri	1
R İstatistiksel Yazılım Geliştirme Ortamının Kişisel Bilgisayarlara Yüklenmesi	4
R İstatistiksel Yazılım Geliştirme Ortamı Sürümlerindeki Değişiklikler	4
R İstatistiksel Yazılım Geliştirme Ortamının Bilgisayara Kurulumu	4
RStudio	11
RStudio'nun Kurulumu.....	12
Paketler	19
RStudio Kapsamında Paket Yükleme ve Çağırma	20
Paketler Sekmesi Kullanılarak Paket Yükleme ve Çağırma	20
Komut Kullanılarak Paket Yükleme ve Çağırma.....	23

2. BÖLÜM

RSTUDIO İLE BETİMSSEL İSTATİSTİKLER VE KOMUT DİZİLERİNİN KULLANIMI

RSTUDIO ile Betimsel İstatistikler	27
RStudio Ortamında Veri Seti Oluşturma.....	28
Excel Dosyasının RStudio'ya Okutulması/Çağırılması	30
SPSS Programı Dosyasının RStudio'ya Okutulması/Çağırılması	33
RStudio'da ggplot2 Paketi İle Veri Görselleştirme	41
RSTUDIO ile Komut Dizilerinin (Scripts) Kullanımı.....	44

3. BÖLÜM

RSTUDIO İLE ANLAM ÇIKARICI İSTATİSTİKLER

RStudio ile Korelasyon Analizi	49
Pearson Momentler Çarpımı Korelasyon Katsayısı.....	49
Nokta Çift Serili Korelasyon Katsayısı.....	57
Spearman Sıra Farkları Korelasyon Katsayısı	62
Kendall'ın Tau Sıralama Korelasyon Katsayısı	66
Basit Doğrusal Regresyon Analizi.....	70

Tek Örneklem T-Testi	81
Bağımsız Örneklem T-Testi.....	86
Bağımlı Örneklem T-Testi.....	92
Tek Yönlü Varyans Analizi (ANOVA)	98
Tekrarlı Ölçümlerde Tek Yönlü Varyans Analizi (Tekrarlı Ölçümlerde ANOVA) ...	110
Tek Yönlü ANCOVA	119
Mann-Whitney U Testi (Wilcoxon Sıra-Toplam Testi).....	130
Kruskal-Wallis H Testi.....	136
Wilcoxon İşaretli Sıralar Testi.....	141
Friedman Testi	146
Ki-Kare Bağımsızlık Testi.....	153
KAYNAKLAR	161

ŞEKİLLER LİSTESİ

Şekil 1. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu I	5
Şekil 2. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu II.....	6
Şekil 3. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu III	6
Şekil 4. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu IV	7
Şekil 5. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu V.....	8
Şekil 6. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu VI.....	8
Şekil 7. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu VII.....	9
Şekil 8. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu VIII	10
Şekil 9. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu IX	10
Şekil 10. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu X.....	11
Şekil 11. RStudio'nun İndirilmesi I	12
Şekil 12. RStudio'nun İndirilmesi II.....	13
Şekil 13. RStudio'nun Çalıştırılması I	14
Şekil 14. RStudio'nun Çalıştırılması II.....	14
Şekil 15. RStudio'nun Çalıştırılması III	15
Şekil 16. RStudio'nun Çalıştırılması IV	15
Şekil 17. RStudio'nun Çalıştırılması V.....	16
Şekil 18. RStudio'nun Çalıştırılması VI	16
Şekil 19. RStudio Ana Ekranı.....	17
Şekil 20. RStudio Ana Ekrandaki Menüler	19
Şekil 21. RStudio Paketler Menüsü	21
Şekil 22. RStudio'da Paketlerin Yüklenmesi I.....	21
Şekil 23. RStudio'da Paketlerin Yüklenmesi II	22
Şekil 24. RStudio'da Pakete İlişkin Bilgiler	22
Şekil 25. RStudio'da "psych" Paketinin Çağrılması	23
Şekil 26. RStudio'da Komut Kullanarak Paket Yükleme I.....	23
Şekil 27. RStudio'da Komut Kullanarak Paket Yükleme II	24
Şekil 28. RStudio'da Komut Kullanarak Paket Çağırma	24
Şekil 29. Komut Kullanarak Paket Yükleme	25
Şekil 30. RStudio'da Kullanıcı Kütüphanesi	25
Şekil 31. RStudio'da Sistem Kütüphanesi.....	26
Şekil 32. Veri Düzenleyici Ekranı I	28
Şekil 33. Değişken Düzenleyici Penceresi	29
Şekil 34. Veri Düzenleyici Ekranı II.....	29
Şekil 35. Değişkenlere İlişkin Objelerin Görüntülendiği RStudio Ortamı.....	30
Şekil 36. Excel Veri Setini RStudio'ya Yükleme	31
Şekil 37. Veri Setinin RStudio'ya Okutulması I	31
Şekil 38. Veri Setinin RStudio'ya Okutulması II.....	32
Şekil 39. Veri Setinin RStudio'ya Okutulması III	32

Şekil 40. SPSS Programına İlişkin Veri Setini RStudio'ya Yükleme.....	33
Şekil 41. Veri Setinin RStudio'ya Okutulması I	34
Şekil 42. Veri Setinin RStudio'ya Okutulması II.....	34
Şekil 43. Veri Setinin RStudio'ya Okutulması III	35
Şekil 44. Veri Setinin İlk 10 Satırı.....	35
Şekil 45. Veri Setinin Boyutuna İlişkin Bilgiler	36
Şekil 46. Veri Setinde Yer Alan Değişkenler	37
Şekil 47. Veri Setinde Yer Alan Değişkenlere İlişkin Tanımlayıcı Bilgiler	37
Şekil 48. Veri Setinde Yer Alan Değişkenlere İlişkin Standart Sapma Değerleri	38
Şekil 49. Veri Setinde Yer Alan Değişkenlere İlişkin Çarpıklık Değerleri.....	38
Şekil 50. Veri Setinde Yer Alan Değişkenlere İlişkin Basıklık Değerleri.....	38
Şekil 51. Veri Setinde Yer Alan Değişkenlere İlişkin Varyans Değerleri.....	39
Şekil 52. Veri Setinde Yer Alan Değişkenlere İlişkin Kayıp Veriler	39
Şekil 53. Veri Setinde Yer Alan Değişkenlere İlişkin Kayıp Verilerin Yüzdesi.....	39
Şekil 54. Veri Setinde Yer Alan Değişkenlere İlişkin Kayıp Verilerin Durumu	40
Şekil 55. Veri Setinin RStudio'ya Okutulması ve Ggplot2 Paketinin Çağrılması.....	42
Şekil 56. Akademik Başarı Değişkeninin Frekansına İlişkin Görsel	43
Şekil 57. Akademik Başarı ve Çalışma Süresi Değişkenlerinin Birbirlerine Göre Dağılımı	43
Şekil 58. Veri Setinin RStudio'ya Okutulması ve İliştirilmesi.....	44
Şekil 59. Komut Dizini Penceresi Açma Adımları	45
Şekil 60. Komut Dizini Penceresinin Çalıştırılması.....	46
Şekil 61. Komut Dizini Penceresi İle Elde Edilen Analiz Sonuçları.....	47
Şekil 62. Pearson Veri Setinin RStudio'ya Okutulması.....	50
Şekil 63. Graphics Paketinin RStudio'ya Çağrılması.....	51
Şekil 64. Sınav Kaygısı ve Motivasyon Ölçeği Puanlarına İlişkin Saçılım Grafiği	51
Şekil 66. Sınav Kaygısı Puanlarına İlişkin Kutu Grafiği	52
Şekil 66. Motivasyon Puanlarına İlişkin Kutu Grafiği.....	52
Şekil 67. Sınav Kaygısı Puanlarına İlişkin Q-Q Grafiği.....	53
Şekil 68. Motivasyon Puanlarına İlişkin Q-Q Grafiği	54
Şekil 69. Sınav Kaygısı Puanlarına İlişkin Shapiro-Wilk Normallik Testi Sonuçları.....	54
Şekil 70. Motivasyon Puanlarına İlişkin Shapiro-Wilk Normallik Testi Sonuçları	55
Şekil 71. Kolmogorov-Smirnov Normallik Testi Sonuçları	55
Şekil 72. Sınav Kaygısı ve Motivasyon Puanları Arasındaki Korelasyon Analizi Sonuçları.....	56
Şekil 73. Sınav Kaygısı ve Motivasyon Puanları Arasındaki İlişkiler	57
Şekil 74. Noktaçift Veri Setinin RStudio'ya Okutulması	58
Şekil 75. Cinsiyet Değişkenine Göre Elde Edilen Kutu Grafikleri.....	59
Şekil 76. Kolmogorov-Smirnov Normallik Testi Sonuçları	60
Şekil 77. Varyansların Homojenliğine İlişkin Analiz Sonuçları.....	60

Şekil 78. Nokta Çift Serili Korelasyon Analizi Sonuçları	61
Şekil 79. Nokta Çift Serili Korelasyon Analizi Manidarlık Düzeyi.....	61
Şekil 80. Spearman Veri Setinin RStudio'ya Okutulması	63
Şekil 81. Hakemlerin Puanlarına İlişkin Saçılım Grafikleri	64
Şekil 82. Hakem 1 ve Hakem 2'nin Sıralamaları Arasındaki Korelasyon Analizi Sonuçları.....	65
Şekil 83. Hakem 1 ve Hakem 2'nin Sıralamalarına İlişkin Analiz Sonuçları.....	66
Şekil 84. Spearman Veri Setinin RStudio'ya Okutulması	67
Şekil 85. Hakem 2 İçin Elde Edilen Saçılım Grafiği.....	68
Şekil 86. Hakem 1 İçin Elde Edilen Saçılım Grafiği.....	68
Şekil 87. Hakem 1 ve Hakem 2'nin Sıralamaları Arasındaki Korelasyon Analizi Sonuçları.....	69
Şekil 88. Hakem 1 ve Hakem 2 Sıralamalarına İlişkin Korelasyon Analizi Sonuçları.....	70
Şekil 89. PISA Veri Setinin RStudio'ya Okutulması I	72
Şekil 90. Regresyon Veri Setinin RStudio'ya Okutulması II	72
Şekil 91. Regresyon Veri Setinin Görünümü.....	73
Şekil 92. Objeler Olarak Kaydedilen Regresyon Veri Setinin Görünümü.....	73
Şekil 93. Regresyon Veri Setinin RStudio'ya Eklenmesi.....	73
Şekil 94. Bağımlı ve Bağımsız Değişkenler İçin Doğrusallık Varsayımının İncelenmesi	74
Şekil 95. Bağımlı ve Bağımsız Değişkenler İçin Saçılım Grafiği	75
Şekil 96. RStudio Ortamına “reg” isimli Objenin Aktarılması.....	76
Şekil 97. Basit Doğrusal Regresyon Analizi Sonuçları	76
Şekil 98. Regresyon Doğrusunun Elde Edilmesi.....	77
Şekil 99. Durbin-Watson Test Sonuçları	77
Şekil 100. Doğrusallık Varsayımına İlişkin Görsel	78
Şekil 101. Hataların/Artıkların Normal Dağılımına İlişkin Görsel.....	79
Şekil 102. Eş-Varyanslılık Varsayımına İlişkin Görsel.....	79
Şekil 103. Uç Değer Varsayımına İlişkin Görsel.....	80
Şekil 104. “lm” Fonksiyonuna İlişkin Ayrıntılı Bilgiler	80
Şekil 105. Tek Örneklem Veri Setinin RStudio'ya Okutulması I.....	82
Şekil 106. Tek Örneklem Veri Setinin RStudio'ya Okutulması II	82
Şekil 107. Tek Örneklem Veri Setinin RStudio'ya Okutulması III.....	83
Şekil 108. Tek Örneklem Veri Setinin RStudio'ya Okutulması IV	83
Şekil 109. Özyeterlik Değişkenine İlişkin Kutu Grafiği.....	84
Şekil 110. Özyeterlik Değişkenine İlişkin Shapiro-Wilk Testi Sonuçları	85
Şekil 111. Tek Örneklem t-testine İlişkin Analiz Sonuçları	85
Şekil 112. Bağımsız Örneklem Veri Setinin RStudio'ya Okutulması I	87
Şekil 113. Bağımsız Örneklem Veri Setinin RStudio'ya Okutulması II.....	88
Şekil 114. Bağımsız Örneklem Veri Setinin RStudio'ya Okutulması III	88

Şekil 115. Bağımsız Örneklem Veri Setinin RStudio'ya Okutulması IV	88
Şekil 116. Başarı Değişkenine ilişkin Kutu Grafiği	89
Şekil 117. Kız Öğrenciler için Normallik Testi Sonuçları	90
Şekil 118. Erkek Öğrenciler için Normallik Testi Sonuçları	90
Şekil 119. Varyansların Homojenliği Testi Analiz Sonuçları	91
Şekil 120. Bağımsız Örneklem t-testi Sonuçları	92
Şekil 121. Bağımlı Örneklem Veri Setinin RStudio'ya Okutulması I.....	94
Şekil 122. Bağımlı Örneklem Veri Setinin RStudio'ya Okutulması II	94
Şekil 123. Bağımlı Örneklem Veri Setinin RStudio'ya Okutulması III.....	95
Şekil 124. Bağımlı Örneklem Veri Setinin RStudio'ya Okutulması IV	95
Şekil 125. Ön Test Puanlarında Uç Değerlerin Belirlenmesine Yönelik Rosner Testi Sonuçları.....	96
Şekil 126. Son Test Puanlarında Uç Değerlerin Belirlenmesine Yönelik Rosner Testi Sonuçları	96
Şekil 127. Ön Test ve Son Test Puanları Arasındaki Farkların Dağılımına İlişkin Analiz Sonuçları.....	97
Şekil 128. Ön Test ve Son Test Puanları Arasındaki Farklara ilişkin Analiz Sonuçları.....	98
Şekil 129. ANOVA Veri Setinin RStudio'ya Okutulması I	100
Şekil 130. ANOVA Veri Setinin RStudio'ya Okutulması II.....	101
Şekil 131. ANOVA Veri Setinin RStudio'ya Okutulması III	101
Şekil 132. ANOVA Veri Setinin RStudio'ya Okutulması IV	101
Şekil 133. Değişkenlere İlişkin Kutu Grafiği.....	102
Şekil 134. Mod Adlı Objenin RStudio Ortamına Kaydedilmesi	103
Şekil 135. Normal Dağılıma İlişkin Shapiro-Wilk Testi Sonuçları	104
Şekil 136. Hataların Dağılımına İlişkin Histogram Grafiği.....	105
Şekil 137. Cerceve Adlı Objenin RStudio Ortamına Kaydedilmesi.....	105
Şekil 138. Barlett Testi Sonuçları	106
Şekil 139. Artıkların Dağılımına İlişkin Q-Q Grafiği.....	107
Şekil 140. Varyans Analizi Sonuçları	107
Şekil 141. Varyans Analizi Sonuçlarına İlişkin Etki Büyüklüğü.....	108
Şekil 142. Games Howell İkili Karşılaştırma Testi Sonuçları.....	108
Şekil 143. Çoklu Karşılaştırma Analizi Sonuçları	109
Şekil 144. Repeated Veri Setinin RStudio'ya Okutulması I.....	112
Şekil 145. Repeated Veri Setinin RStudio'ya Okutulması II	112
Şekil 146. Repeated Veri Setinin RStudio'ya Okutulması III.....	113
Şekil 147. Repeated Veri Setinin RStudio'ya Okutulması IV	113
Şekil 148. Gruplara İlişkin Kutu Grafiği.....	114
Şekil 149. Puanlara İlişkin Shapiro-Wilk Normallik Testi Sonuçları	115
Şekil 150. Veri Setinin Son Görüntüsüne İlişkin Görsel.....	116
Şekil 151. Tekrarlı Ölçümlerde Tek Yönlü Varyans Analizi Sonuçları.....	117

Şekil 152. Bonferroni İkili Karşılaştırma Testi Sonuçları.....	118
Şekil 153. ANCOVA Veri Setinin RStudio'ya Okutulması I	121
Şekil 154. ANCOVA Veri Setinin RStudio'ya Okutulması II.....	121
Şekil 155. ANCOVA Veri Setinin RStudio'ya Okutulması III	121
Şekil 156. ANCOVA Veri Setinin RStudio'ya Okutulması IV	122
Şekil 157. Son Test Puanlarına İlişkin Kutu Grafiği.....	123
Şekil 158. Model İsimli Objenin RStudio Ortamına Kaydedilmesi.....	124
Şekil 159. “res” İsimli Objenin RStudio Ortamına Kaydedilmesi.....	124
Şekil 160. “zres” İsimli Objenin RStudio Ortamına Kaydedilmesi.....	125
Şekil 161. Artıkların/Hataların Normal Dağılım Testi Sonuçları	125
Şekil 162. Varyansların Homojenliği Testi Sonuçları	126
Şekil 163. “pred” İsimli Objenin RStudio Ortamına Kaydedilmesi.....	127
Şekil 164. Hataların Doğrusallık Testi Sonuçları	127
Şekil 165. ANCOVA Sonuçları	128
Şekil 166. ANCOVA'ya İlişkin Etki Büyüklükleri.....	130
Şekil 167. MannWhitney Veri Setinin RStudio'ya Okutulması I	132
Şekil 168. Mann Whitney Veri Setinin RStudio'ya Okutulması II.....	132
Şekil 169. Mann Whitney Veri Setinin RStudio'ya Okutulması III	132
Şekil 170. Mann Whitney Veri Setinin RStudio'ya Okutulması IV	133
Şekil 171. Depresyon Değişkenine İlişkin Kutu Grafiği.....	134
Şekil 172. Kız Öğrenciler İçin Depresyon Puanlarının Dağılımı.....	134
Şekil 173. Erkek Öğrenciler İçin Depresyon Puanlarının Dağılımı.....	135
Şekil 174. Mann Whitney U-Testi Analiz Sonuçları.....	136
Şekil 175. Kruskal Wallis Veri Setinin RStudio'ya Okutulması I.....	138
Şekil 176. Kruskal Wallis Veri Setinin RStudio'ya Okutulması II	139
Şekil 177. Kruskal Wallis Veri Setinin RStudio'ya Okutulması III.....	139
Şekil 178. Kruskal Wallis Veri Setinin RStudio'ya Okutulması IV.....	139
Şekil 179. Kruskal Wallis Veri Seti İçin Normallik Testi Sonuçları.....	140
Şekil 180. Çoklu Karşılaştırma Analiz Sonuçları	141
Şekil 181. Wilcoxon Veri Setinin RStudio'ya Okutulması I.....	143
Şekil 182. Wilcoxon Veri Setinin RStudio'ya Okutulması II.....	143
Şekil 183. Wilcoxon Veri Setinin RStudio'ya Okutulması III	143
Şekil 184. Wilcoxon Veri Setinin RStudio'ya Okutulması IV	144
Şekil 185. Ocak ve Şubat Aylarına İlişkin Kutu Grafiği.....	145
Şekil 186. Wilcoxon İşaretili Sıralar Testi Analiz Sonuçları.....	146
Şekil 187. Friedman Veri Setinin RStudio'ya Okutulması I.....	148
Şekil 188. Friedman Veri Setinin RStudio'ya Okutulması II.....	148
Şekil 189. Friedman Veri Setinin RStudio'ya Okutulması III	149
Şekil 190. Friedman Veri Setinin RStudio'ya Okutulması IV	149
Şekil 191. Farklı Zaman Dilimlerindeki Ölçmelere İlişkin Kutu Grafiği.....	150

Şekil 192. Örnek Adlı Objenin RStudio Ortamına Kaydedilmesi	150
Şekil 193. Friedman Testine İlişkin Analiz Sonuçları.....	151
Şekil 194. Nemenyi Testine İlişkin İkili Karşılaştırma Sonuçları.....	151
Şekil 195. Ölçme Sonuçlarına İlişkin Betimsel İstatistikler	152
Şekil 196. Ki-kare Veri Setinin RStudio'ya Okutulması I.....	154
Şekil 197. Ki-kare Veri Setinin RStudio'ya Okutulması II.....	155
Şekil 198. Ki-kare Veri Setinin RStudio'ya Okutulması III.....	155
Şekil 199. Ki-kare Veri Setinin RStudio'ya Okutulması IV	156
Şekil 200. Cinsiyet ve Sigara İçme Durumu Değişkenleri İçin Kontenjans Tablosu	156
Şekil 201. Cinsiyet ve Sigara İçme Durumu Değişkenleri İçin Sütun Grafikleri	157
Şekil 202. Cinsiyet ve Sigara İçme Durumu Değişkenleri İçin Ki-kare Testi Sonucu	158
Şekil 203. Cinsiyet ve Sigara İçme Durumu Değişkenleri için Beklenen Frekans Değerleri.....	158
Şekil 204. Cinsiyet ve Sigara İçme Durumu Değişkenleri için Fisher Testi Sonucu	159

1. BÖLÜM

R İSTATİSTİKSEL YAZILIM GELİŞTİRME ORTAMINA GİRİŞ

R programlama dili farklı amaçlar için geliştirilmiş pek çok yazılımı/uygulamayı içerisinde barındırmaktadır. Bu kapsamda, istatistik, meta analiz, veri madenciliği, yapay zeka vb. uygulamalara yönelik olarak geliştirilen ve kullanılan pek çok yazılımı (SPSS programı, SAS, CMA, WEKA, MATLAB vb.) kendi içerisinde barındırdığını belirtmek doğru olacaktır. R, web tabanlı büyük veri ve veri yönetimi süreçlerinde kullanılmaktadır. Bu kadar geniş bir alanda kullanımı bulunan R programlama diline; üniversiteler, bankalar, fabrikalar, uluslararası kalkınma ve yardım kuruluşları, hastaneler gibi çok kuruluş tarafından da değişik alanlarda başvurulmaktadır. R genellikle, yazılım dillerini bilen veya yazılım dillerine aşina olan programcılar tarafından kullanılmaktadır. Bu kitap kapsamında temel istatistikler R programlama dili ile gerçekleştirilmiş ve elde edilen analiz sonuçlarına dair raporlanma biçimi ele alınmıştır.

R İstatistiksel Yazılım Geliştirme Ortamı ve Özellikleri

Araştırmacıların büyük bir çoğunluğu veri analizi için SPSS, Minitab, Statistica, SAS, Excel vb. araçları kullanmaktadır fakat R İstatistiksel Yazılım Geliştirme Ortamını kullanmak diğer araçlara kıyasla kullanıcılarına birçok avantaj sunmaktadır. Bu avantajlardan bazıları aşağıda listelenmiştir.

- Açık kaynak kodlu ve ücretsizdir.
- Genel anlamda birden çok amaca hizmet etmektedir.
- İşletim sistemi farketmeksizin çalışabilmektedir.
- Tekrarlanabilir çalışmalar oluşturulabilmektedir.
- Güçlü ve dinamik bir topluluğa sahip olma ve çok sayıda yerel R kullanıcı grubu ile farklı platformlarda yaşanan sorunlara hızlı bir şekilde çözüm bulunabilmektedir.
- Hazır kodlar kullanılabilir.

- Profesyonel yazılımcılar ve meraklılar için kurulmuş olan Stackoverflow adlı sitede kayıtlı olan uzman istatistikçi ve yazılımcılar ile program ile ilgili her konuda iletişime geçilerek destek alınabilmektedir.
- Kod ve çıktılar tek bir doküman içerisinde yer almaktadır.
- Çok büyük veri setleri ile çalışan araştırmacıların hızlı bir şekilde sonuca gitmesine yardımcı olmaktadır.
- Aynı amaca hizmet eden programlara göre miktar ve kalite anlamında öğrenme kaynakları daha fazladır.
- Tanıtımına ve kullanımına ilişkin ücretsiz çevrimiçi kitaplar mevcuttur.
- Yoğun hesaplamalar için C, C++ ve Fortran gibi programlama dilleri ile çalışabilmektedir.
- Komut mantığı ile çalışabilmektedir.
- Her anlamda gelişmiş ve ayrıntılı grafikler oluşturulabilmektedir.
- LaTeX benzeri belgelendirme imkanına sahiptir.
- Excel, SPSS, SAS, STATA, SYSTAT uzantılı farklı veri dosyalarını okuyabilmektedir.
- Birçok veri işleme, veri madenciliği yönteminin yanı sıra meta analiz ve sosyal ağ analizi gibi analizler de gerçekleştirilebilmektedir.

R, son yıllarda popülerite anlamında diğer istatistiksel yazılımlara göre çok büyük bir yol almıştır. Bu popülerite göz önünde bulundurularak, R'in yeni ve gelişmekte olan bir dil olduğunu söyleyebiliriz. Şaşırtıcı bir şekilde, R, 1992'de Robert Gentleman ve Ross Ihaka tarafından başlatılan bir araştırma projesi kapsamında geliştirilmeye başlanmıştır ve 1995'ten bu yana açık kaynak kodlu bir yazılım olarak piyasada yer almaktadır. R programlama dilinin S programlama diline rakip olarak çıktığı söylenebilir. Yazılımın popüleritesinin aniden ve çok hızlı bir şekilde artış göstermesi, veri biliminin bir kariyer ve çalışma alanı olarak ortaya çıkması ile doğru orantılıdır. Ancak, veri biliminin temelleri uzun yıllara dayanmaktadır. İstatistik, doğrusal cebir, yöneylem araştırması, yapay zeka ve makine öğrenimi, modern bir veri bilimcisinin kullandığı araçlara katkıda bulunmaktadır. R istatistiksel yazılım geliştirme ortamı, bu araçlar ile gerçekleştirilenlerin neredeyse hepsini tek bir ara yüz ile gerçekleştirebilme yeteneğine sahiptir. R istatistiksel yazılım geliştirme ortamının tarihsel olarak nasıl bir gelişim gösterdiği aşağıda sırasıyla açıklanmıştır.

1992: R, Yeni Zelanda'da bir araştırma projesi kapsamında Robert Gentleman ve Ross Ihaka* tarafından geliştirilmeye başlandı.

1993: R'in iki kodlu ilk sürümü Statlib'de yayınlandı.

1995: R, GPL2 lisansı altında açık kaynak kodlu bir yazılım olarak piyasaya sürüldü.

1997: R geliştiricileri "The R Core Group" adı altında bir araya geldiler.

1997: CRAN (The Comprehensive R Archive Network; Kapsamlı R Arşiv Ağı) oluşturuldu.

1999: R web sitesi, www.r-project.org kuruldu

1999: The R Core Group Viyana'da İstatistik konulu konferansta (Inaugural Directions in Statistical Computing Conference) bir araya geldi.

2000: R'in 1.0.0 sürümü yayınlandı (Şubat 29).

2000: 1998'de ACM Yazılım Sistemleri Ödülünü S yazılımı ile alan John Chambers The R Core Group'a katıldı.

2001: R News (daha sonra The R Journal; R Dergisi oldu) kuruldu.

2003: R Vakfı (The R Foundation) kuruldu.

2004: Viyana'da ilk UseR! Konferansı gerçekleştirildi.

2004: R'in 2.0.0 sürümü yayınlandı.

2009: R dergisinin ilk sayısı yayınlandı.

2013: R'in 3.0.0 sürümü yayınlandı.

2015: R Vakfının katılımıyla "R Birliği" kuruldu.

2016: Yeni logo kullanılmaya başlandı.

2017: R 3.3.3, R 3.4.0, R 3.4.1, R 3.4.2 ve R 3.4.3 yayınlandı.

2018: R 3.4.4, R 3.5.0, R 3.5.1 ve R 3.5.2 yayınlandı.

2019: R 3.5.3, R 3.6.0, R 3.6.1 yayınlandı. R 3.6.1 hali hazırda en güncel R sürümüdür.

R istatistiksel yazılım geliştirme ortamının kullanımına başlamadan programa ilişkin indirme ve yükleme işlemlerinin gerçekleştirilmesi gerekmektedir.

* Ihaka, R. and R. Gentleman (1996). "R: A language for data analysis and graphics," Journal of Computational and Graphical Statistics, 5, 299–314.

R İstatistiksel Yazılım Geliştirme Ortamının Kişisel Bilgisayarlara Yüklenmesi

R'a ilişkin yükleme dosyasına, R'ın geliştiricisi olan Kapsamlı R Arşiv Ağının (CRAN) sitesinden (<https://cran.r-project.org/>) ulaşılabilir. Siteye giriş yapıldığında, sayfanın üst kısmında Windows, Mac OS ve Linux işletim sistemleri için programa ilişkin indirme bağlantıları yer almaktadır. Windows ve Mac OS işletim sistemleri için önceden kurulmuş dosyalar mevcutken, Linux'a ilişkin kurulum dosyaları genellikle ilgili kaynaklardan derlenmiş bir haldedir. Program, bu platformlar kullanılarak çok basit bir şekilde bilgisayara yüklenebilmektedir.

Windows kullanıcıları programın kurulum dosyasını bilgisayarlarına indirmek için, önce Download R for Windows daha sonra base ve son olarak Download R 3.x.x sekmelerine tıklamalıdır. Buradaki x, mevcut R istatistiksel yazılım geliştirme ortamı sürümüne işaret etmektedir. Programa ilişkin güncellemeler oldukça x periyodik olarak değişmektedir.

Benzer şekilde, Mac kullanıcıları önce Download R for (Mac) OS X sonra R-3.X.X.PKG sekmelerine tıklamalıdır. Yine buradaki x harfi, mevcut R istatistiksel yazılım geliştirme ortamı sürümüne işaret etmektedir. Bu link programın hem 32 hem de 64 bit sürümlerini bilgisayara kurmaktadır.

Linux kullanıcıları, apt-get (Ubuntu ve Debian) / yum (Red Hat) / zypper (SUSE) veya başka bir kaynaktan sağlanan standart dağılım mekanizmalarını kullanarak programı yükleyebilirler. Bu mekanizmalar hem programı indirecek hem de yükleyecektir.

R İstatistiksel Yazılım Geliştirme Ortamı Sürümlerindeki Değişiklikler

Bu kitabın yazıldığı dönemde programa ilişkin mevcut sürüm R 3.6.1'dir. CRAN, programın her sene, program sürümünü belirten 3 rakamlı kodlama şablonunun ortasında yer alan rakamı majör değişiklikleri temsilen bir sayı arttırarak yeni sürümünü piyasaya sürmektedir. 2015 yılında R 3.2.0, 2016 yılında 3.3.0, 2017 yılında ise 3.4.0 piyasaya sürülmüştür. Programın mevcut sürüm kodu ise R 3.6.1'dir. Sürüm kodunda yer alan son rakam ise küçük (minör) güncellemeleri temsil etmektedir. Çoğu R işlevi, genellikle önceki sürümlerle geriye dönük bir şekilde uyum içerisindedir.

R İstatistiksel Yazılım Geliştirme Ortamının Bilgisayara Kurulumu

R istatistiksel yazılım geliştirme ortamının bilgisayara kurulumundan önce sistemin 32 ve 64 byte kapsamındaki sürümleri ile ilgili bilgi sahibi olmada fayda

vardır. 32 byte ve 64 byte kullanımı arasındaki seçim, bilgisayarın zemininde kurulu işletim sistemine bağlı olarak değişiklik göstermektedir.

64 byte'ı destekleyen R 3.0.0 ve üzeri sürümler R nesnelerinde çok fazla veri depolayabilmektedir. Geçmişte, bazı paketler R istatistiksel yazılım geliştirme ortamının 32 byte kapsamındaki sürümünü gerektiriyordu ancak son yıllarda bu durumla oldukça az karşılaşılmaktadır. Programın 32 byte olan sürümünü kurmanın tek nedeni, bu sürümün bazı eski analiz türlerini desteklemesidir.

R istatistiksel yazılım geliştirme ortamının yüklenmesi için öncelikle daha önce de bahsedildiği gibi <https://cran.r-project.org/bin/windows/base/> adlı siteye giriş yapılır. Kurulum (set up) dosyasının bilgisayara indirilmesi için ana sayfanın üst tarafında **Download R 3.6.1 for Windows** adlı link yer almaktadır (Kurulum linkine ilişkin site uzantısı R'nin yeni sürümleri çıktıkça değişebilmektedir. Arama motoruna "R Download" yazıldığı takdirde R istatistiksel yazılım geliştirme ortamının yüklenmesine ilişkin web sitesi ilk seçenekler içerisinde görüntülenmektedir). Bu aşamada yapılması gerekenler Şekil 1'de gösterilmiştir.

R-3.6.1 for Windows (32/64 bit)

Download R 3.6.1 for Windows (81 megabyte)

[Installation and other instructions](#)
[New features in this version](#)

Download R 3.6.1 for Windows tıklanır.

If you want to double-check that the package you have downloaded is the same as the one computed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the [CRAN website](#). Both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

Note to webmasters: A stable link which will redirect to the current Windows binary release is [<CRAN MIRROR>/bin/windows/base/release.htm](#).

Şekil 1. R İstatistiksel Yazılım Geliştirme Ortamının Kurulumu I

Download R 3.6.1 for Windows adlı linke tıkladıktan sonra yaklaşık 81 megabyte yer kaplayan program kurulum dosyasının bilgisayara indirilme süreci başlamaktadır. İlk basamak programın dilinin belirlenmesine ilişkindir. Şekil 2'de programın dilinin belirlenmesine ilişkin pencereyi göstermektedir. Programın dili pencerede yer alan sekme vasıtasıyla seçilir ve OK düğmesine tıklanır.