



مؤسسه فرهنگی هنری
دیبگران تهران

به نام خدا



مؤسسه فرهنگی هنری
دیبگران تهران

راهنمای جامع برنامه نویسی

R

(مقدماتی و پیشرفته)

مؤلف:

دکتر سید مهدی صالحی

(عضو هیئت علمی دانشگاه نیشابور)

با همکاری: ریحانه عرفه



هرگونه چاپ و تکثیر از محتویات این کتاب بدون اجازه کتبی ناشر ممنوع است. متخلفان به موجب قانون حمایت حقوق مؤلفان، مصنفان و هنرمندان تحت پیگرد قانونی قرار می گیرند.

◀ عنوان کتاب: راهنمای جامع برنامه نویسی R (مقدماتی و پیشرفته)

◀ مولف: دکتر سید مهدی صالحی

◀ ویراستار: ریحانه عرفه

◀ ناشر: موسسه فرهنگی هنری دیباگران تهران

◀ صفحه آرای: فرنوش عبدالهی

◀ طراح جلد: داریوش فرسایی

◀ نوبت چاپ: اول

◀ تاریخ نشر: ۱۳۹۸

◀ چاپ و صحافی: صدف

◀ تیراژ: ۱۰۰ جلد

◀ قیمت: ۷۹۰۰۰۰ ریال

◀ شابک: ۹۷۸-۶۲۲-۲۱۸-۲۵۳-۳

◀ نشانی واحد فروش: تهران، میدان انقلاب،

خ کارگر جنوبی، روبروی پاساژ مهستان،

پلاک ۱۲۵۱

تلفن: ۲۲۰۸۵۱۱۱-۶۶۴۱۰۰۴۶

◀ فروشگاههای اینترنتی دیباگران تهران:

WWW.MFTBOOK.IR

www.dibbook.ir

www.dibagarantehran.com

◀ نشانی تلگرام: @mftbook

◀ نشانی اینستاگرام دیبا: [dibagaran_publishing](https://www.instagram.com/dibagaran_publishing)

هر کتاب دیباگران، یک فرصت جدید شغلی.

هر گوشی همراه، یک فروشگاه کتاب دیباگران تهران.

از طریق سایتها و اپ دیباگران، در هر جای ایران به کتابهای ما دسترسی دارید.

فهرست مطالب

فصل اول / معرفی نرم افزار و زبان برنامه نویسی R.....۱۹

- ۱-۱ - مقدمه..... ۲۰
- ۲-۱ - چرا R؟..... ۲۰
- ۳-۱ - چند نکته اساسی درباره‌ی زبان برنامه نویسی R..... ۲۲
- ۴-۱ - شروع کار با R..... ۲۳
- ۱-۴-۱ - نحوه‌ی اجرای برنامه‌های R..... ۲۳
- ۱-۱-۴-۱ - روش تعاملی..... ۲۳
- ۲-۱-۴-۱ - روش BATCH..... ۲۵
- ۲-۴-۱ - آشنایی اجمالی با منوهای نرم افزار..... ۲۵
- ۳-۴-۱ - روش‌های جستجو در نرم افزار..... ۲۶
- ۵-۱ - بسته‌های نرم افزاری..... ۲۶
- ۱-۵-۱ - نصب بسته‌های نرم افزاری..... ۲۸
- ۲-۵-۱ - فراخوانی بسته‌های نرم افزاری..... ۲۹
- ۳-۵-۱ - به روزرسانی بسته‌های نرم افزاری..... ۲۹
- ۶-۱ - مفاهیم پایه..... ۳۰
- ۱-۶-۱ - محاسبات ساده‌ی ریاضی..... ۳۰
- ۲-۶-۱ - متغیرها..... ۳۲
- ۱-۲-۶-۱ - ایجاد متغیرها..... ۳۲
- ۲-۲-۶-۱ - حذف کردن متغیرها..... ۳۳
- ۷-۱ - انواع داده‌ها..... ۳۳
- ۱-۷-۱ - داده‌های عددی (double یا numeric)..... ۳۴
- ۲-۷-۱ - داده‌های صحیح (integer)..... ۳۴
- ۳-۷-۱ - داده‌های موهومی (complex)..... ۳۵
- ۴-۷-۱ - داده‌های رشته‌ای (character)..... ۳۵
- ۵-۷-۱ - داده‌های منطقی (logical)..... ۳۶

فصل دوم / بردارها ۳۸

۱-۲-۱-۲- مقدمه ۳۹

۲-۲-۱-۲- ایجاد بردارها ۳۹

۱-۲-۲- تابع `c()` ۳۹

۲-۲-۲- تابع `scan()` ۴۱

۳-۲-۲- ایجاد دنباله‌های حسابی ۴۲

۴-۲-۲- دنباله‌ی اعداد تکراری ۴۳

۳-۲-۳- شناسایی و ویرایش درایه‌ها ۴۳

۴-۲-۴- استخراج عناصر مشروط ۴۵

۵-۲-۵- اعمال جبری درایه به درایه ۴۶

۶-۲-۶- معرفی برخی توابع مهم برای بردارها ۴۸

۱-۶-۲- توابع آماری ۴۹

۱-۱-۶-۲- تابع `summary()` ۴۹

۲-۱-۶-۲- تابع `sample()` ۴۹

۳-۱-۶-۲- توابع `table()` و `ftable()` ۵۰

۲-۶-۲- توابع ریاضی ۵۰

۱-۲-۶-۲- خانواده توابع `cum...` ۵۰

۲-۲-۶-۲- تابع `order()` ۵۱

۳-۲-۶-۲- توابع `round()`، `floor()` و `ceiling()` ۵۲

۴-۲-۶-۲- خانواده توابع `gamma()` ۵۲

۳-۶-۲- چند تابع و عملگر متفرقه ۵۳

۱-۳-۶-۲- تابع `which()` ۵۳

۲-۳-۶-۲- توابع `any()` و `all()` ۵۵

۳-۳-۶-۲- عملگر `'%in%'` ۵۷

۴-۳-۶-۲- تابع `names()` ۵۸

۷-۲-۷- توابع برداری شده ۵۹

۸-۲-۸- داده‌های گمشده ۶۰

۱-۸-۲- کار با داده‌های گمشده ۶۱

۹-۲-۹- معرفی بیش‌تر مقادیر `NaN`، `Inf` و `NULL` ۶۲

۱۰-۲-۱۰- توابع مهم در یک نگاه ۶۳

فصل سوم / ماتریس‌ها و آرایه‌ها ۶۶

۱-۳-۱-۳- مقدمه ۶۷

۲-۳-۲-۳- ایجاد ماتریس‌ها ۶۷

۶۷matrix() تابع	۱-۲-۳
۷۱dim() تابع	۲-۲-۳
۷۱cbind() و rbind() توابع	۳-۲-۳
۷۲diag() تابع	۴-۲-۳
۷۳فراخوانی و ویرایش درایه‌ها	۳-۳
۷۷اعمال جبری درایه به درایه	۴-۳
۷۹توابع و عملگرهای ماتریسی	۵-۳
۷۹عملگرهای ضرب	۱-۵-۳
۸۱solve() تابع	۲-۵-۳
۸۵آرایه‌ها	۶-۳
۸۷فراخوانی درایه‌ها، بردارها و ماتریس‌ها	۱-۶-۳
۸۷اعمال جبری	۲-۶-۳

فصل چهارم / عامل‌ها، داده‌های چارچوب‌دار و فهرست‌ها

۹۲مقدمه	۱-۴
۹۲عامل‌ها(FACTOR)	۲-۴
۹۲ایجاد عامل‌ها	۱-۲-۴
۹۲factor() تابع	۱-۱-۲-۴
۹۵gl() تابع	۲-۱-۲-۴
۹۵cut() تابع	۳-۱-۲-۴
۹۶داده‌های چارچوب‌دار(DATAFRAME)	۳-۴
۹۶ایجاد داده‌های چارچوب‌دار	۱-۳-۴
۹۷فراخوانی و ویرایش ستون‌ها، ردیف‌ها و درایه‌ها	۲-۳-۴
۹۷تابع '['	۱-۲-۳-۴
۹۸attach() تابع	۲-۲-۳-۴
۹۹تابع '\$'	۳-۲-۳-۴
۱۰۰subset() تابع	۴-۲-۳-۴
۱۰۰اعمال جبری درایه به درایه	۳-۳-۴
۱۰۱برخی توابع مهم برای داده‌های چارچوب‌دار	۴-۳-۴
۱۰۱str() تابع	۱-۴-۳-۴
۱۰۲merge() تابع	۲-۴-۳-۴
۱۰۳فهرست‌ها(LIST)	۴-۴
۱۰۳ایجاد فهرست‌ها	۱-۴-۴
۱۰۳list() تابع	۱-۱-۴-۴

۱۰۴ vector() تابع ۲-۱-۴-۴
۱۰۵ split() تابع ۳-۱-۴-۴
۱۰۷ فراخوانی و ویرایش آیتم‌ها ۲-۴-۴
۱۰۸ اعمال جبری ۳-۴-۴
۱۰۹ فهرست‌های بازگشتی ۴-۴-۴
۱۱۰ هموار کردن یک فهرست ۵-۴-۴
۱۱۳ شناسایی اشیاء معرفی شده ۵-۴-۴
۱۱۴ فراخوانی و ذخیره‌ی داده‌ها در خارج از محیط R ۶-۴-۴
۱۱۵ ۱-۶-۴ ذخیره‌ی موقت داده‌ها با فرمت RData
۱۱۶ ۲-۶-۴ ذخیره‌ی داده‌ها با قالب‌های معروف
۱۱۶ ۱-۲-۶-۴ قالب متنی
۱۱۷ ۲-۲-۶-۴ قالب اکسل
۱۲۱ ۳-۲-۶-۴ داده‌های نرم‌افزارهای آماری SPSS, SAS, Minitab, Stata
۱۲۲ ۳-۶-۴ ذخیره‌سازی و فراخوانی مرحله به مرحله‌ی داده‌ها
۱۲۲ ۱-۳-۶-۴ تابع cat()
۱۲۳ ۲-۳-۶-۴ readLines()
۱۲۴ ۳-۳-۶-۴ writeLines()

۱۲۶..... فصل پنجم / توابع پیشرفته

۱۲۷ ۱-۵ مقدمه
۱۲۷ ۲-۵ تابع apply()
۱۳۳ ۳-۵ تابع lapply()
۱۳۶ ۴-۵ تابع sapply()
۱۳۸ ۵-۵ تابع mapply()
۱۳۹ ۶-۵ تابع tapply()
۱۴۳ ۷-۵ تابع aggregate()
۱۴۵ ۸-۵ تابع by()
۱۴۸ ۹-۵ تابع get()
۱۵۰ ۱۰-۵ تابع replicate()

۱۵۳..... فصل ششم / اصول برنامه‌نویسی R

۱۵۴ ۱-۶ مقدمه
۱۵۴ ۲-۶ حلقه‌ها
۱۵۵ ۱-۲-۶ حلقه‌ی for
۱۵۹ ۲-۲-۶ حلقه‌ی while

۱۶۱	۳-۶- دستورات شرطی
۱۶۱	۱-۳-۶- دستور if
۱۶۲	۲-۳-۶- دستور if-else
۱۶۳	۳-۳-۶- تابع ifelse()
۱۶۴	۴-۳-۶- تابع switch()
۱۶۵	۴-۶- دستورات کنترلی
۱۶۸	۵-۶- تابع نویسی
۱۷۵	۱-۵-۶- توابع بازگشتی
۱۷۷	۲-۵-۶- تابع Vectorize()
۱۸۰	۳-۵-۶- مشخص کردن پیش فرض برای آرگومان‌ها
۱۸۲	۴-۵-۶- ساخت عملگرها و توابع جایگذاری
۱۸۵	۵-۵-۶- متغیرهای محلی و سراسری
۱۹۸	۶-۶- گریزی به برنامه‌نویسی موازی
۱۹۸	۱-۶-۶- بسته‌ی snow
۲۰۲	۲-۶-۶- بسته‌ی parallel
۲۰۴	۳-۶-۶- بسته‌ی foreach
۲۰۵	۴-۶-۶- موازی یا سری؟ کدامیک!؟
۲۰۶	فصل هفتم / کار با داده‌های متنی
۲۰۷	۱-۷- مقدمه
۲۰۷	۲-۷- ساختار داده‌های متنی
۲۰۸	۳-۷- ایجاد کاراکترها و رشته‌های متنی
۲۰۹	۱-۳-۷- تابع character()
۲۱۰	۲-۳-۷- توابع paste() و paste0()
۲۱۱	۳-۳-۷- تابع toString()
۲۱۲	۴-۷- چاپ داده‌های متنی
۲۱۲	۱-۴-۷- تابع print()
۲۱۳	۲-۴-۷- تابع noquote()
۲۱۳	۳-۴-۷- تابع cat() برای کاراکترها
۲۱۵	۴-۴-۷- تابع sprintf()
۲۱۷	۵-۷- مدیریت داده‌های متنی
۲۱۸	۱-۵-۷- تابع nchar()
۲۱۹	۲-۵-۷- توابع toupper() و tolower()
۲۱۹	۳-۵-۷- تابع casefold()

۲۱۹.....	chartr() تابع ۴-۵-۷
۲۲۰.....	grep() تابع ۵-۵-۷
۲۲۱.....	gsub() تابع ۶-۵-۷
۲۲۲.....	abbreviate() تابع ۷-۵-۷
۲۲۲.....	substring() تابع ۸-۵-۷
۲۲۳.....	strsplit() تابع ۹-۵-۷
۲۲۷.....	۶-۷- توابع مجموعه‌ای در بردارهای کاراکتری
۲۲۷.....	union() تابع ۱-۶-۷
۲۲۷.....	intersect() تابع ۲-۶-۷
۲۲۷.....	setdiff() تابع ۳-۶-۷
۲۲۸.....	setequal() تابع ۴-۶-۷
۲۲۸.....	identical() تابع ۵-۶-۷
۲۲۸.....	is.element() یا %in% تابع ۶-۶-۷
۲۲۹.....	sort() تابع ۷-۶-۷
۲۲۹.....	۷-۷- پردازش متن با استفاده از بسته‌ی نرم‌افزاری tm
۲۳۰.....	۱-۷-۷- ورود اطلاعات با استفاده از تابع corpus()
۲۳۱.....	۲-۷-۷- پیش‌پردازش متن
۲۳۲.....	۱-۲-۷-۷- تابع content_transformer()
۲۳۲.....	۲-۲-۷-۷- تابع tm_map()
۲۳۵.....	۳-۷-۷- تحلیل متن
۲۳۵.....	۱-۳-۷-۷- ماتریس کلمات (Term-Document Matrix)
۲۳۶.....	۲-۳-۷-۷- استخراج کلمات با فراوانی مورد نظر
۲۳۶.....	۳-۳-۷-۷- حذف کلمات نادر
۲۳۷.....	۴-۳-۷-۷- بررسی روابط میان کلمات
۲۳۷.....	۴-۷-۷- ابر کلمات (Word Cloud)

فصل هشتم / آشنایی با محیط گرافیکی R ۲۴۰

۲۴۱.....	۱-۸- مقدمه
۲۴۱.....	۲-۸- گرافیک پایه
۲۴۱.....	۱-۲-۸- رنگ‌ها در R
۲۴۲.....	۲-۲-۸- نمودارهای آماری (توصیفی)
۲۴۲.....	۱-۲-۲-۸- نمودار دایره‌ای
۲۴۴.....	۲-۲-۲-۸- نمودار میله‌ای
۲۴۶.....	۳-۲-۲-۸- نمودار جعبه‌ای

۲۴۷ نمودار مستطیلی (هیستوگرام) ۴-۲-۲-۸
۲۴۸ نمودار پراکنش ۵-۲-۲-۸
۲۵۱ نمودار چندک-چندک ۶-۲-۲-۸
۲۵۲ نمودار برآورد چگالی کرنل ۷-۲-۲-۸
۲۵۳ نمودارهای ریاضی ۳-۲-۸
۲۵۳ منحنی توابع ۱-۳-۲-۸
۲۵۴ رسم پاره خط ۲-۳-۲-۸
۲۵۵ چند ضلعی ۳-۳-۲-۸
۲۵۶ نمودارهای سه بُعدی ۴-۲-۸
۲۵۶ رویه (ژرفنما) ۱-۴-۲-۸
۲۵۷ نمودار خط تراز (کانتور) ۲-۴-۲-۸
۲۵۸ ویرایش و سفارشی کردن نمودارها ۵-۲-۸
۲۵۸ افزودن نقاط به نمودار ۱-۵-۲-۸
۲۵۸ افزودن خطوط به نمودار ۲-۵-۲-۸
۲۵۹ ایجاد زمینه شطرنجی ۳-۵-۲-۸
۲۵۹ تعیین مختصات نقاط نمودار ۴-۵-۲-۸
۲۶۰ شناسه‌های اختیاری ۵-۵-۲-۸
۲۶۳ افزودن جعبه راهنما به نمودار ۶-۵-۲-۸
۲۶۴ نوشتن متن درون نمودار ۷-۵-۲-۸
۲۶۸ طراحی صفحه گرافیکی ۸-۵-۲-۸
۲۶۹ ذخیره‌ی نمودارها ۶-۲-۸
۲۶۹ بسته نرم‌افزاری ggplot2 ۳-۸
۲۷۰ آشنایی با بسته گرافیکی ggplot2 ۱-۳-۸
۲۷۰ لایه‌های ضروری ۱-۱-۳-۸
۲۷۰ لایه‌های اختیاری ۲-۱-۳-۸
۲۷۱ تابع ggplot() ۲-۳-۸
۲۷۱ نمودار میله‌ای ۱-۲-۳-۸
۲۷۳ هیستوگرام ۲-۲-۳-۸
۲۷۴ نمودار پراکنش ۳-۲-۳-۸
۲۷۵ نمودار جعبه‌ای ۴-۲-۳-۸
۲۷۷ بهبود کارایی نمودارها ۳-۳-۸
۲۷۹ ذخیره‌ی نمودارها ۴-۳-۸

پیوست الف / تمرین ها ۲۸۱

۲۸۲..... تمرین های فصل اول

۲۸۲..... تمرین های فصل دوم

۲۸۴..... تمرین های فصل سوم

۲۸۶..... تمرین های فصل چهارم

۲۸۷..... تمرین های فصل پنجم

۲۸۹..... تمرین های فصل ششم

۲۹۰..... تمرین های فصل هفتم

۲۹۰..... تمرین های فصل هشتم

پیوست ب / کار با توزیع های آماری ۲۹۲

نمایه ۲۹۶

۳۰۱..... مراجع

فهرست چالش‌ها

(سطح متوسط و پیشرفته)

۴۴	«چالش ۱-۲»: استاندارد و یکنواخت کدنویسی کنیم!
۴۷	«چالش ۲-۲»: با استفاده از ویژگی بازتولید برنامه‌ی بهینه‌تری بنویسیم!
۵۲	«چالش ۳-۲»: NaN را تبدیل به عدد کنیم!
۵۴	«چالش ۴-۲»: اهمیت استفاده از متغیرهای منطقی
۵۵	«چالش ۵-۲»: کاربرد تابع any()
۵۶	«چالش ۶-۲»: آزمون برابری دو بردار: all() یا identical()؟!
۵۸	«چالش ۷-۲»: کاربرد عملگر %in در تقریب احتمال
۷۵	«چالش ۱-۳»: فراخوانی درایه‌های مشروط در ماتریس‌ها
۷۸	«چالش ۲-۳»: ویژگی بازتولید در ماتریس‌ها (۱)
۷۹	«چالش ۳-۳»: ویژگی بازتولید در ماتریس‌ها (۲)
۸۱	«چالش ۴-۳»: معکوس تعمیم‌یافته‌ی یک ماتریس یا ماتریس مور-پنروز
۸۴	«چالش ۵-۳»: ساختن ماتریس ضریب همبستگی
۸۸	«چالش ۶-۳»: ویژگی ناب بازتولید در آرایه‌ها!
۹۴	«چالش ۱-۴»: یک نکته‌ی قابل تأمل درباره‌ی عامل‌ها!
۱۰۷	«چالش ۲-۴»: تفاوت عملگرهای فراخوانی [.] و [[]] در فهرست‌ها
۱۱۴	«چالش ۳-۴»: تابع ls() را بیشتر بشناسیم!
۱۱۸	«چالش ۴-۴»: fread() در برابر read.table()
۱۱۹	«چالش ۵-۴»: یک تکنیک ساده برای فراخوانی کلان‌داده
۱۲۴	«چالش ۶-۴»: ذخیره‌سازی مرحله به مرحله‌ی یک خروجی
۱۳۰	«چالش ۱-۵»: کاربرد apply() در آرایه‌ها
۱۳۲	«چالش ۲-۵»: apply(..., 2, sum) یا colSums()؟!
۱۳۵	«چالش ۳-۵»: کاربرد تابع lapply()
۱۴۰	«چالش ۴-۵»: ترکیب توابع خانواده‌ی apply
۱۴۶	«چالش ۵-۵»: کاربرد تابع by()
۱۵۱	«چالش ۶-۵»: کاربرد تابع replicate()
۱۵۷	«چالش ۱-۶»: بایدها و نبایدها در استفاده از حلقه‌ی for

- «چالش ۶-۲»: R از ویژگی Lazy evaluation پیروی می‌کند. ۱۷۱.....
- «چالش ۶-۳»: توابع به‌عنوان اشیاء! ۱۷۲.....
- «چالش ۶-۴»: return() یا break;print()؟! ۱۷۴.....
- «چالش ۶-۵»: توابع بازگشتی به‌عنوان یک شاه کلید! ۱۷۶.....
- «چالش ۶-۶»: کاربرد Vectorize() ۱۷۸.....
- «چالش ۶-۷»: ایجاد توابع جایگذاری ۱۸۴.....
- «چالش ۶-۸»: اولویت محاسبات با متغیرهای محلی است! ۱۸۶.....
- «چالش ۶-۹»: چاپ کردن متغیرهای محلی ۱۸۸.....
- «چالش ۶-۱۰»: سلسله مراتب را در توابع تودرتو بشناسیم! ۱۸۸.....
- «چالش ۶-۱۱»: آشنایی با محیط (Environment) ۱۹۰.....
- «چالش ۶-۱۲»: تفاوت عملگرهای انتساب '=', '<-' و '<<-' ۱۹۲.....
- «چالش ۶-۱۳»: کمی بیشتر از عملگر '<<-' بدانیم! ۱۹۴.....
- «چالش ۶-۱۴»: ساختن یک شمارنده‌ی خودکار ۱۹۶.....
- «چالش ۶-۱۵»: صرفه‌جویی در زمان با موازی‌نویسی. ۲۰۰.....
- «چالش ۷-۱»: کاربرد عبارات منظم در توابع کاراکنتری ۲۲۰.....

فهرست جداول

جدول ۱-۱: عملگرهای حسابی.	۳۰
جدول ۲-۱: معرفی برخی نمادهای مهم در R.	۳۱
جدول ۳-۱: عملگرهای منطقی.	۳۶
جدول ۱-۲: توابع پرکاربرد.	۶۳
جدول ۲-۲: توابع مهم آماری.	۶۴
جدول ۳-۲: توابع مهم ریاضی.	۶۴
جدول ۱-۳: توابع ماتریسی.	۸۲
جدول ۱-۴: توابع مهم داده‌های چارچوب‌دار.	۱۰۱
جدول ۲-۴: توابع فراخوانی داده.	۱۲۱
جدول ۱-۷: توابع چاپ کاراکترها.	۲۱۲
جدول ۲-۷: انواع کاراکترهای فرمت در تابع <code>printf()</code>	۲۱۵
جدول ۳-۷: نشانه‌های کنترل خروجی در تابع <code>printf()</code>	۲۱۶
جدول ۴-۷: توابع آماده برای مدیریت داده‌های متنی.	۲۱۸
جدول ۵-۷: توابع مجموعه‌ای.	۲۲۷
جدول ۶-۷: توابع پیش‌پردازش آماده.	۲۳۲
جدول ۱-۸: انواع نمودارهای قابل ترسیم با تابع <code>plot()</code>	۲۴۸
جدول ۲-۸: شناسه‌های اختیاری در رسم نمودار.	۲۶۰
جدول ۳-۸: انواع خط در رسم نمودارها.	۲۶۱
جدول ۴-۸: کد متناظر نمادها برای فرمول‌نویسی در نمودارها.	۲۶۵
جدول ۵-۸: فرمول‌نویسی در نمودارها.	۲۶۶
جدول ب-۱: نحوه‌ی استفاده از توزیع‌های آماری در R.	۲۹۳
جدول ب-۲: نماد توزیع‌های آماری معروف در R.	۲۹۵

فهرست شکل‌ها

۲۴	شکل ۱-۱: پنجره‌ی R Console
۲۷	شکل ۲-۱: دسته‌بندی موضوعی بسته‌های نرم‌افزاری
۲۸	شکل ۳-۱ : دسته‌بندی موضوعی بسته‌های نرم‌افزاری (ادامه)
۷۷	شکل ۳-۱: پنجره‌ی Data Editor
		شکل ۷-۱: کلمات پرتکرار تست‌های زبان تخصصی کنکور کارشناسی ارشد آمار طی
۲۳۸	سال‌های ۱۳۸۸-۱۳۹۷
۲۴۲	شکل ۸-۱: طیف رنگ‌ها در محیط گرافیکی R

خط مشی کیفیت انتشارات مؤسسه فرهنگی هنری دیباگران تهران در عرصه کتاب‌های است که بتواند خواسته‌های به روز جامعه فرهنگی و علمی کشور را تا حد امکان پوشش دهد. هر کتاب دیباگران تهران، یک فرصت جدید شعلی

حمد و سپاس ایزد منان را که با الطاف بی‌کران خود این توفیق را به ما ارزانی داشت تا بتوانیم در راه ارتقای دانش عمومی و فرهنگی این مرز و بوم در زمینه چاپ و نشر کتب علمی دانشگاهی، علوم پایه و به ویژه علوم کامپیوتر و انفورماتیک گام‌هایی هرچند کوچک برداشته و در انجام رسالتی که بر عهده داریم، مؤثر واقع شویم.

گسترده‌گی علوم و توسعه روزافزون آن، شرایطی را به وجود آورده که هر روز شاهد تحولات اساسی چشمگیری در سطح جهان هستیم. این گسترش و توسعه نیاز به منابع مختلف از جمله کتاب را به عنوان قدیمی‌ترین و راحت‌ترین راه دستیابی به اطلاعات و اطلاع‌رسانی، بیش از پیش روشن می‌نماید.

در این راستا، واحد انتشارات مؤسسه فرهنگی هنری دیباگران تهران با همکاری جمعی از اساتید، مؤلفان، مترجمان، متخصصان، پژوهشگران، محققان و نیز پرسنل ورزیده و ماهر در زمینه امور نشر درصدد هستند تا با تلاش‌های مستمر خود برای رفع کمبودها و نیازهای موجود، منابعی پُر بار، معتبر و با کیفیت مناسب در اختیار علاقمندان قرار دهند.

کتابی که در دست دارید با همت "دکتر سید مهدی صالحی و با همکاری سرکار خانم ریحانه عرفه" و تلاش جمعی از همکاران انتشارات میسر گشته که شایسته است از یکایک این گرامیان تشکر و قدردانی کنیم.

کارشناسی و نظارت بر محتوا: زهره قزلباش

در خاتمه ضمن سپاسگزاری از شما دانش‌پژوه گرامی درخواست می‌نماید با مراجعه به آدرس dibagaran.mft.info (ارتباط با مشتری) فرم نظرسنجی را برای کتابی که در دست دارید تکمیل و ارسال نموده، انتشارات دیباگران تهران را که جلب رضایت و وفاداری مشتریان را هدف خود می‌داند، یاری فرمایید.

امیدواریم همواره بهتر از گذشته خدمات و محصولات خود را تقدیم حضورتان نماییم.

مدیر انتشارات

مؤسسه فرهنگی هنری دیباگران تهران
bookmarket@mft.info

❖ مقدمه مولف

نگارش کتاب حاضر در راستای شناخت و کسب درک عمیق در خصوص توانمندی‌های زبان برنامه‌نویسی R و کاربرد آن در پردازش داده‌ها تدوین شده است. اگرچه R یک زبان برنامه‌نویسی آماری قلمداد می‌شود؛ لیکن به واسطه‌ی نقش موثر و غیر قابل انکار آمار در علوم مختلف و همچنین ویژگی‌های منحصر به فرد این نرم‌افزار، می‌توان آن را یکی از معدود زبان‌های برنامه‌نویسی دانست که با سرعت زیادی در حوزه‌های مختلف علوم رو به پیشرفت است. از جمله ویژگی‌های خاص این نرم‌افزار که آن را از سایر نرم‌افزارهای هم‌رده و توانمند آن در عصر کنونی مجزا می‌سازد می‌توان به رایگان بودن، سرعت اجرای قابل توجه، قابلیت اجرا در سیستم عامل‌های مختلف، امکان فراخوانی و تبدیل ساده‌ی کدها به سایر زبان‌های برنامه‌نویسی معروف و بالعکس، سرعت به‌روزرسانی بالا و پشتیبانی توسط برنامه‌نویس‌های بزرگ سراسر دنیا اشاره نمود. همچنین به دلیل کدباز بودن، سهولت استفاده و دربرگیری بسته‌های نرم‌افزاری متنوع، کاربران زیادی را خصوصاً در حوزه‌های آمار و علوم داده به خود جذب کرده است. بنابراین، بر آن شدیم تا راهنمای مناسبی جهت استفاده از این نرم‌افزار مهیا کنیم. این کتاب در سطوح مقدماتی، متوسط و در بعضی موارد پیشرفته تنظیم شده است که حاصل تجربیات دوازده ساله‌ی نویسنده در تدریس دروس دانشگاهی مرتبط با R، برگزاری کارگاه‌های متعدد در دانشگاه‌های سراسر کشور و همچنین انجام پروژه‌های مشترک با یکی از بزرگان این حوزه در اروپا است. متن پیش‌زمینه‌ی کتاب (که با فونت نرمال تایپ شده است) در سطح مقدماتی و به‌سادگی نگارش شده است تا برای آن دسته از کاربرانی که قصد آشنایی با این زبان برنامه‌نویسی را دارند مفید باشد. برای اکثر خروجی‌ها تفسیرهایی به‌صورت مورد به مورد (با فونت ایتالیک و اندازه‌ی کوچکتر) ارائه شده است که ممکن است از حوصله‌ی کاربران سطح مقدماتی خارج باشد. این نکات برای افرادی که می‌خواهند سطح کدنویسی خود را در R افزایش دهند، مفید خواهد بود و جهت پاسخ‌گویی به نیاز این دسته از خوانندگان طراحی شده است. در سراسر کتاب (به‌جز فصل ۸) برخی نکات پیشرفته‌ی برنامه‌نویسی در قالب «چالش» ارائه شده است. در این موارد در حقیقت چالش‌های پیش‌روی کاربر که در حین کار کردن با این زبان برنامه‌نویسی با آن‌ها مواجه می‌شود، مطرح شده و در برخورد با این چالش‌ها به ذکر نکات ظریف و با اهمیتی که برای رویارویی با آن‌ها در این کتاب گنجانده شده است، می‌پردازیم. مطالعه‌ی این چالش‌ها به کاربران حرفه‌ای R توصیه می‌شود. هدف نویسنده از نوشتن این کتاب پرداختن به پیاده‌سازی مدل‌های آماری از جمله رگرسیون، سری‌های زمانی و ... به کمک نرم‌افزار R نبوده، بلکه «آشنایی و ارتقای سطح برنامه‌نویسی با R» و البته نمایش قدرت آن در «پردازش داده» بوده است. بنابراین در حین نگارش سعی بر استفاده‌ی کمتر اصطلاحات و روش‌های آماری و احتمالی شده است. اما به‌رحال برای آشنایی بیشتر با توزیع‌های آماری در R، مطالعه‌ی پیوست (ب) مفید می‌باشد. در پیوست (الف) نیز چند تمرین برای هر فصل در نظر گرفته شده است که حل آن‌ها به کاربران سطح مقدماتی توصیه می‌شود. کتاب حاضر در سطح مقدماتی برای دانشجویان محترم اکثر رشته‌ها، به‌طور ویژه رشته‌های آمار، علوم داده، داده‌کاوی، علوم کامپیوتر، مهندسی کامپیوتر (نرم‌افزار)، سایر علوم مهندسی و همچنین علوم پزشکی توصیه می‌شود و مطالب غیرمقدماتی آن برای تمامی پژوهشگران محترم حوزه‌های یاد شده مفید خواهد بود.

این کتاب در ۸ فصل تنظیم شده است که در ادامه به‌طور خلاصه توضیحاتی از هر فصل داده می‌شود: در فصل اول به معرفی مقدمات نرم‌افزار و زبان برنامه‌نویسی R می‌پردازیم. از نکات مهم مطرح شده در این فصل می‌توان به نحوه‌ی اجرای کدهای R به روش BATCH، بررسی مفاهیم پایه‌ی زبان برنامه‌نویسی R و همچنین معرفی انواع داده‌ها و تبدیل آن‌ها به یک‌دیگر بر اساس اصل coercion، اشاره کرد.

در فصل‌های دوم تا چهارم ساختارهای مختلف داده در R را معرفی می‌کنیم. اما با توجه به این‌که بردارها پایه‌ی تمام ساختارهای داده در R می‌باشند، فصل دوم را به‌طور کامل به معرفی آن‌ها اختصاص می‌دهیم. از جمله موارد ارائه شده در این فصل می‌توان به نحوه‌ی ایجاد بردارها به روش‌های گوناگون، نحوه‌ی فراخوانی و ویرایش درایه‌های یک بردار، بررسی ویژگی بازتولید در بردارها، معرفی برخی توابع مهم و پرکاربرد برداری، اهمیت استفاده از متغیرهای منطقی و نحوه‌ی مدیریت داده‌های گمشده و مبهم اشاره کرد.

فصل سوم را بر پایه‌ی معرفی ماتریس‌ها بنا نهادیم، ساختاری که قدرت R را در برنامه‌نویسی به‌خوبی نمایان می‌کند. در این فصل پس از بررسی موارد مرسوم در هر ساختار داده (نحوه‌ی ایجاد، دسترسی و ویرایش درایه‌ها و اعمال جبری)، به واکاوی ویژگی بازتولید در ماتریس‌ها، معرفی انواع عملگرهای ضرب ماتریسی و همچنین توابع مهم آماده برای این ساختار داده می‌پردازیم. انتهای این فصل به معرفی آرایه‌ها، به‌عنوان تعمیمی از ماتریس‌ها، اختصاص دارد.

در فصل چهارم عامل‌ها، داده‌های چارچوب‌دار و فهرست‌ها را به‌عنوان ساختارهای مهم داده در R معرفی می‌کنیم. یکی از نکات قابل توجهی که در داده‌های چارچوب‌دار و فهرست‌ها به آن اشاره می‌شود، بحث نحوه‌ی فراخوانی متغیرها در این‌گونه ساختارها است که بسیار حائز اهمیت است. به‌علاوه فهرست‌های بازگشتی و هموار کردن این ساختار داده نیز موضوع مهم دیگری است که در این فصل گنجانده شده است. سپس، نحوه‌ی ذخیره‌سازی ساختارهای داده در R و همچنین فراخوانی فایل‌های داده با قالب‌های گوناگون را آموزش می‌دهیم. یک تکنیک مفید برای فراخوانی مجموعه داده‌های بزرگ و همچنین ذخیره‌سازی مرحله به مرحله خروجی برنامه‌های R را نیز در انتهای این فصل ارائه می‌دهیم.

پس از معرفی ساختارهای گوناگون داده، در فصل پنجم چند تابع پیشرفته و بسیار کاربردی در برنامه‌نویسی R را معرفی خواهیم کرد. این توابع قادرند جایگزین مناسبی برای حلقه‌ها شوند و امر برنامه‌نویسی را در این محیط بسیار ساده‌تر و بعضاً سریع‌تر کنند. کاربرد تک‌تک این توابع در افزایش قدرت پردازش و دسترسی به داده‌ها در قالب چندین مثال و چالش بررسی خواهد شد.

در فصل ششم، ساختارهای برنامه‌نویسی R را با جزئیات قابل توجهی شرح می‌دهیم. در سطح مقدماتی این فصل، اصول مهم دستورنویسی، حلقه‌ها، دستورات شرطی و کنترلی و همچنین به‌صورت دقیق‌تر مبحث مهم توابع در R را معرفی می‌کنیم. اما از نکات مهمی که در سطوح بالاتر برنامه‌نویسی در این فصل مطرح می‌شوند، می‌توان به باید‌ها و نبایدها در استفاده از حلقه‌ی for، بررسی ویژگی lazy evaluation در R، نکات مهم در استفاده از تابع return()، نحوه‌ی به‌کارگیری توابع بازگشتی، آشنایی با متغیرهای محلی و سراسری، آشنایی با محیط‌ها و بحث سلسله مراتب، بررسی تفاوت عملگرهای مختلف انتساب و برنامه‌نویسی موازی در R، اشاره کرد.

فصل هفتم به داده‌های متنی (کاراکتر) اختصاص داده شده است. در این فصل، با این نوع از داده‌ها به‌طور ویژه آشنا خواهیم شد. به‌طور دقیق‌تر، با توابع مهم مختلف برای ایجاد، چاپ کردن و مدیریت داده‌های متنی آشنا خواهیم شد. از نکات برجسته‌ای که در این فصل گنجانده شده‌اند، می‌توان به بررسی ویژه‌ی تابع

`sprintf()`، نحوه‌ی مدیریت عبارات منظم (regular expressions)، حل برخی الگوریتم‌های ریاضی با کمک کاراکترها و همچنین ارائه‌ی مقدمه‌ای بر متن کاوی، اشاره کرد.

در فصل هشتم به صورت مقدماتی به بررسی محیط گرافیکی قدرتمند R می‌پردازیم. نقطه‌ی قوت این فصل، آموزش نحوه‌ی فرمول‌نویسی و همچنین ایجاد و نمایش نمودارها به کمک بسته‌ی نرم‌افزاری بی‌نظیر `ggplot2` (که توسط هادلی ویکه‌ام، دانشمند نیوزلندی R و استاد آمار دانشگاه استنفورد، معرفی شده است) می‌باشد.

در پایان بر خود لازم می‌دانم از همکاران محترمی که دست‌نوشته‌های این کتاب را با دقت، نقد و بررسی کردند و مسئولیت ویراستاری کتاب را بر عهده داشته‌اند، صمیمانه قدردانی کنم. همچنین به‌طور ویژه از سرکار خانم ریحانه عرفه، دانشجوی کارشناسی ارشد آمار-ریاضی دانشگاه فردوسی مشهد، که در امر نگارش این کتاب به ویژه در فصل‌های هفتم و هشتم و تمرین‌ها کمال همکاری را داشته‌اند و سهم زیادی در کیفیت‌بخشی به این اثر را ایفا کرده‌اند، سپاس‌گزاری می‌نمایم.

با وجود تلاش و دقتی که در تهیه این کتاب مبذول شده است، یقیناً محتویات این مجموعه خالی از اشکال نیست؛ بنابراین در راستای بهبود اثر و ارتقای سطح کیفی و علمی آن در چاپ‌های آتی، از خوانندگان محترم خواهش‌مندیم که با ارائه‌ی نظرات سازنده خود از طریق ایمیل، نگارنده را در رفع اشکالات احتمالی یاری فرمایند. امیدواریم تدوین این اثر گامی هر چند کوچک در راستای اعتلای رسالت فرهنگی و علمی به حساب آید و مورد استفاده‌ی خوانندگان ارجمند قرار گیرد.

سید مهدی صالحی

عضو هیئت علمی گروه ریاضی و آمار دانشگاه نیشابور

salehi2sms@gmail.com

زمستان ۱۳۹۸