

کاربرد نرم افزار R در آمار

نویسنده
سارا استول

مترجمان

دکتر زهرا طهماسبی

استادیار دانشگاه ایلام

مهندس سیده زهرا طباطبایی

دانشجوی دکتری دانشگاه ایلام

مهندس زینب رستمی

دانشجوی دکتری دانشگاه ایلام

سرشناسه	: استونل، سارا Stowell, Sarah
عنوان و نام پدیدآور	: کاربرد نرم‌افزار R در آمار/ نویسنده سارا استونل؛ مترجمان زهرا طهماسبی، سیده‌زهرا طباطبایی‌پور، زینب رستمی.
مشخصات نشر	: ایلام: دانشگاه ایلام، ۱۳۹۶.
مشخصات ظاهری	: ۳۰۶ ص.: مصور، جدول.
شابک	: 978-600-6184-33-3
وضعیت فهرست نویسی	: فیپا
یادداشت	: عنوان اصلی: Using R for statistics, 2014.
موضوع	: آر (زبان برنامه‌نویسی کامپیوتر)
موضوع	: R (Computer program language)
موضوع	: آمار -- داده‌پردازی
موضوع	: Statistics -- Data processing
شناسه افزودن	: طهماسبی، زهرا، ۱۳۶۰-، مترجم
شناسه افزودن	: طباطبایی‌پور، سیده‌زهرا، ۱۳۶۵-، مترجم
شناسه افزودن	: رستمی، زینب، ۱۳۶۸-، مترجم
شناسه افزودن	: دانشگاه ایلام
رده بندی کنگره	: QA۷۶۷/۳۰۲/۱۳۹۶
رده بندی دیویی	: ۵۰۰/۵۱۹/۱۵۵۱۰۰۰۵۰
شماره کتابشناسی ملی	: ۲۹۳۳۵۶۳



انتشارات دانشگاه ایلام

کاربرد نرم‌افزار R در آمار

نویسنده	سارا استونل
مترجمان	دکتر زهرا طهماسبی مهندس زهرا طباطبایی مهندس زینب رستمی
سال چاپ اول	۱۳۹۶
شمارگان	۱۰۰۰
ناشر	انتشارات دانشگاه ایلام
صفحه آرایي	سجاد حاتمی
طراح جلد	محمد طهماسبی
قیمت	۲۰۰۰۰ تومان

۰۸۴۱۳۳۲۲۱۹۰۶

آدرس ناشر: ایلام، بلوار پژوهش، دانشگاه ایلام، انتشارات دانشگاه

پیش‌گفتار مترجمان

سیاس و ستایش خداوند بلند مرتبه را که به لطف و عنایت بی‌کران او توفیق ترجمه کتاب "کاربرد نرم افزار R در آمار" حاصل آمد. از این رو در آغاز، کمترین ادب پندگی بجای آورده و یاد می‌کنیم آموزنده بی‌آموزگار را.

امروزه داده کاوی یکی از کارهای اجتناب ناپذیر هر علمی شده است هر پژوهشی در قسمت تجزیه و تحلیل به جهت تبیین و بیان روابط و مدل‌های علی نیاز به نرم‌افزاری دارد که با اخذ اطلاعات و ترکیبات آنها، قادر به ارائه گزارش و نمودارهای لازم شود و با جمع‌بندی داده‌ها در حاتم به خوبی به تحلیل داده‌ها و تصمیم‌گیری موفق نایل آید.

R یک زبان برنامه‌نویسی محیط نرم‌افزاری برای محاسبات آماری و علم داده‌ها است، که بر اساس زبان‌های اسکیم پیاده‌سازی شده است. R اولین بار در سال ۱۹۹۶ ظهور یافت که دو پرفسور از رولند به نام‌های Robert Gentleman و Ross Ihaka یک کد برنامه نویسی را به همراه نرم‌افزار به صورت رایگان منتشر کردند. شاید دلیل نامگذاری R به خاطر حرف اول اسم کوچک این دو پرفسور آمار باشد. دلیل محبوبیت این نرم‌افزار در سالهای اخیر در دهه اول رایگان بودن این نرم‌افزار است. همچنین این نرم‌افزار بر روی انواع سیستم عامل قابل نصب است. از دیگر مزایای این نرم‌افزار فراهم کردن زمینه‌ای بی‌نظیر برای نوشتن و اجرای آماری است و قابلیت توسعه همچون دیگر نرم‌افزارهای باز را دارد. این برنامه در حال تبدیل شدن به یک زبان برنامه نویسی بین‌المللی است. امروزه شرکت‌های مختلف از جمله Google, Pfizer, Merck, Bank of America, Shell از آن به صورت روزمره استفاده می‌کنند. R این قابلیت را به کاربران می‌دهد که کدهای خود را که با زبان برنامه نویسی مانند سی++ نوشته می‌شود را برای محققان دیگر به اشتراک بگذارند، کدهای نوشته شده را اصلاح و از بسته‌های نرم‌افزاری دیگران استفاده کنند. بنابراین R هم یک زبان برنامه نویسی و هم یک بسته نرم‌افزاری برای کاربران می‌باشد که حاوی محدوده گسترده‌ای از تکنیک‌های آماری (از جمله: مدل‌سازی خطی و غیرخطی،

آزمون‌های کلاسیک آماری، تحلیل سری‌های زمانی، رده‌بندی، خوشه‌بندی و غیره) و قابلیت‌های گرافیکی است. در محیط R، کدهای سی، سی++ و فورترن قابلیت اتصال و فراخوانی هنگام اجرای برنامه را دارند و کاربران خبره می‌توانند توسط کدهای سی، مستقیماً اشیا R را تغییر دهند. گرچه نرم‌افزار R اغلب به منظور انجام محاسبات آماری به کار می‌رود، این نرم‌افزار قابل به کارگیری در محاسبات ماتریسی را دارد و در این زمینه، همپای نرم‌افزارهایی چون اکتاو و نسخه تجاری آن متلب (MATLAB) است. کتابی که پیش‌رو دارید حاصل ترجمه سلیس و روان کتاب "Using R for statistics" است که به زبانی ساده و با بیان مفاهیم مقدماتی ولیکن ارزشمند و مفید سعی در آموزش گام به گام تجزیه و تحلیل داده‌های آماری با نرم‌افزار R را دارد و لذا برای اساتید، محققان، دانشجویان کلیه رشته‌های مهندسی، پزشکی، بیولوژی و علوم اجتماعی و انسانی قابل استفاده می‌باشد. در پایان از کلیه صاحب نظران، اساتید معزز و دانشجویان گرامی درخواست می‌شود نظرات خود را در خصوص اثر حاضر با ما در میان گذارند.

"زهرا طهماسبی، سیده زهرا طباطبایی و ریند رستمی"

فهرست مطالب

۱۱	مقدمه
	فصل ۱:
۱۵	اصول کار با نرم افزار R
	فصل ۲:
۴۱	کار با فایل داده‌ها
	فصل ۳:
۶۱	آماد سازی و دستکاری داده‌ها
	فصل ۴:
۹۳	ترکیب و باز سازی دیتاها
	فصل ۵:
۱۱۵	آماره‌های اختصاری نتایج های بسته داده‌ها
	فصل ۶:
۱۴۱	داده‌های با ساختار جدول داده‌ها
	فصل ۷:
۱۶۷	توزیع‌های احتمال
	فصل ۸:
۱۸۵	ایجاد کردن نمودار
	فصل ۹:
۲۱۳	ایجاد تغییرات در ظاهر نمودارها
	فصل ۱۰:
۲۴۷	آزمون‌های فرض
	فصل ۱۱:
۲۷۷	رگرسیون و مدل‌های خطی عمومی
۳۴۳	درباره‌ی مؤلف
۳۴۴	درباره‌ی داور (بازبینی کننده) فنی
۳۴۵	سپاسگزاری‌ها

مقدمه

به کاربرد R در آمار خوش آمدید. این کتاب برای تمام کسانی که می‌خواهند از نرم‌افزار R برای تجزیه‌ی داده‌ها و ایجاد نمودارهای آماری استفاده کنند، نوشته شده است. این کتاب برای کسانی که در کار با نرم‌افزار R کم یا بی‌تجربه هستند، مناسب است و قصد دارد، بدون نیاز به یادگیری تمامی جزئیات برنامه‌نویسی، آن را فرا گرفته و به سرعت پیش بروید.

در باره‌ی R

نرم‌افزار R یک نرم‌افزار تجزیه‌ی آماری و محیط گرافیکی و نیز یک زبان برنامه‌نویسی است. R نرم‌افزاری فرمان‌بر و بسیار مشابه نرم‌افزار تجاری تولید شده $S\text{-Plus}^{\text{®}}$ می‌باشد.

R نرم‌افزاری رایگان، جهت نصب تحت مجوز عمومی کلی GNU است که توسط گروهی از افراد داوطلب مشهور تحت عنوان تیم اصلی R نوشته شده است. نسخه‌ی اولیه‌ی این نرم‌افزار توسط مارکین R در سراسر جهان که بسیاری از آن‌ها متعلق به جامعه‌ی دانشگاهی جهان هستند، با افزودن بیش از ۵۰۰۰ بسته توسعه یافته و تکمیل شده است. بسته‌های اضافه شده به نرم‌افزار توسط این افراد طیف گسترده‌ای از تکنیک‌های آماری را پوشش داده‌اند. تمامی افراد می‌توانند در بسته‌های افزودنی مشارکت کنند که قبل از افزودن آن‌ها به مجموعه پایستی از لحاظ کیفیت بررسی شوند. در زمان نوشتن این اثر، نسخه‌ی رایج نرم‌افزار R ، ۳.۱.۱ است.

آنچه که خواهید آموخت

این کتاب برای ارائه راهنمایی‌های ساده و عملی جهت اجرای روش‌های آماری متداول در R طراحی شده است. جنبه‌ی برنامه‌نویسی R ، تنها به طور مختصر بررسی شده است. بعد از خواندن این کتاب قادر خواهید بود:

۱. نصب و اجرای نرم‌افزار R ،

۲. وارد کردن داده،

۳. دست‌ورزی داده‌ها،

۴. محاسبه‌ی آماره‌های توصیفی،
۵. ایجاد نمودارهای آماری و تنظیمات ظاهری آن‌ها،
۶. انجام آزمون‌های فرض آماری از جمله آزمون t استیودنت و تجزیه‌ی واریانس،
۷. ساخت مدل‌های رگرسیونی،
۸. دسترسی به قابلیت‌های اضافی با استفاده از بسته‌های افزودنی،
۹. ایجاد انواع مختلف توابع،

اگرچه در این کتاب برخی یادآوری‌ها در مورد روش‌های آماری وجود دارد؛ اما برای آموزش آمار در نظر گرفته نشده است؛ بنابراین برای استفاده از کتاب به برخی معلومات قبلی، نیاز خواهید داشت. برای یادگیری اطلاعات موجود در کتاب بایستی قادر به انتخاب مناسب‌ترین روش آماری برای اهداف خود و تفسیر نتایجتان باشید. همچنین بایستی با مفاهیم و اصطلاحات آماری، رایج آشنا باشید؛ بنابراین اگر در مورد هر کدام از روش‌های مورد بحث در کتاب اطلاعات کافی ندارید، توصیه می‌کنیم، این کتاب را همراه با یک کتاب آمار پایه، استفاده کنید.

در این کتاب فرض شده است که شما هیچ دانش قبلی از R یا برنامه‌نویسی ندارید. در اینجا تنها مهارت‌های عمومی کامپیوتر و آشنایی با سیستم عامل کامپیوتر شما مورد نیاز است.

قراردادهای استفاده شده در این کتاب

این کتاب از قراردادهای تایپی زیر استفاده می‌کند:

۱. برای تمایز تمامی دستورات R و خروجی از متن اصلی، از خط *Fixed width* استفاده شده است.

۲. خط *Normal fixed width* برای ایجاد نام توابع R ، نام عبارت‌ها، توابع نحوی، نام مجموعه‌ی داده‌های خاص و متغیرها و هر بخش دیگری از دستورات که می‌تواند کلمه به کلمه کپی شود، استفاده می‌شود.

۳. خط *Slanted fixed width* برای مجموعه‌ی داده‌های عام و نام متغیرها و هر بخش دیگری از دستورات که بایستی با مقدارهای خود کاربر جایگزین شود، استفاده می‌شود.

۴. اغلب جا شدن یک دستور کامل تا عرض صفحه امکان پذیر نیست. در این موارد، این دستور به روی خط بعدی و با فاصله ادامه می‌یابد. در هر جایی که این حالت را می‌بینید، دستور همچنان بایستی بتواند بر روی یک خط جداگانه در دستگاه وارد شود.

۵. جعبه‌های حاوی متن که از متن اصلی جدا هستند، حاوی یادآوری‌هایی از تئوری یا روش‌های آماری می‌باشند.
۶. نمونه‌های عملی در بخش‌های شماره‌گذاری شده جداگانه ارائه شده است.

مجموعه‌ی داده‌های مورد استفاده در این کتاب

تعداد زیادی از مجموعه‌ی داده‌های مثال زده شده در R وجود دارند و این‌ها در هنگام باز نمودن نرم‌افزار جهت استفاده، قابل دسترس خواهند بود. این کتاب باعث می‌شود که از چندین مجموعه‌ی داده برای اهداف نمایشی استفاده شود. هم‌چنین تعدادی از مجموعه‌ی داده‌های اضافی مورد استفاده در سراسر این کتاب وجود دارد که در پیوست C شرح داده شده است. این داده‌ها جهت دانلود در سایت www.apress.com/9781484201404 در دسترس می‌باشند.

ارتباط با نویسنده

در صورت داشتن هرگونه پیشنهاد یا انتقاد، بسیار مایلم آن‌ها را از طریق ایمیل s.stowell@instantr.com دریافت نمایم.