



دانشگاه علوم پزشکی و توانبخشی

شبیه‌سازی با نرم‌افزار R



دانشگاه علوم پزشکی و توانبخشی

شبیه‌سازی با نرم‌افزار R

تألیف: دکتر اکبر بیگلریان - دکتر عنایت‌الله بخشی

مؤلفان:

دکتر اکبر بیگلریان
دکتر عنایت‌الله بخشی

Simulation with R

By: Akbar Biglarian (PhD in Biostatistics)

Enayatollah Bakhshi (PhD in Biostatistics)

R یک برنامه قوی برای تجزیه و تحلیل داده‌های آماری است که در سال‌های اخیر مورد توجه بسیاری از پژوهش‌گران قرار گرفته است. هدف، آموزش دانش‌جویان و اساتید رشته آمار و سایر علاقه‌مندان در به کارگیری این نرم‌افزار برای شبیه‌سازی است. در این کتاب سعی شده تا انتقال مفاهیم با تمرین‌های کاربردی آسان‌تر انجام شود. بدون تردید مطالعه‌ی این کتاب برای آنان که علاقه‌مند به بررسی و مطالعه‌ی پارامترها با استفاده از شبیه‌سازی هستند بسیار مفید خواهد بود.



انتشارات دانشگاه علوم پزشکی و توانبخشی

شابک: ۹۷۸-۶۰۰-۶۴۴۲-۰۷-۵

قیمت: ۱۰۰۰۰۰ ریال

نزدیک است
به نام خدایم که در

شبیه سازی با نرم افزار



دکتر اکبر بیگلریان
دکتر عنایت الله بخشی

سرشناسه	: بیگلریان ، اکبر ، ۱۳۵۴ -
عنوان و نام پدیدآور	: شبیه سازی با نرم افزار R / مولفان اکبر بیگلریان ، عنایت الله بخشی .
مشخصات نشر	: تهران: دانشگاه علوم بهزیستی و توانبخشی، ۱۳۹۳ .
مشخصات ظاهری	: ۱۰۸ ص.: مصور، جدول، نمودار.
شابک	: ۹۷۸-۶۰۰-۶۴۴۲-۰۷-۵-۱۰۰۰۰ ریال
وضعیت فهرست نویسی	: فیبا
یادداشت	: واژه نامه .
یادداشت	: کتابنامه: ص. ۹۷ .
یادداشت	: نمایه .
موضوع	: آر (زبان برنامه نویسی کامپیوتر)
موضوع	: آمار ریاضی -- داده پردازی
شناسه افزوده	: بخشی، عنایت اله، ۱۳۴۰ -
شناسه افزوده	: دانشگاه علوم بهزیستی و توانبخشی
رده بندی کنگره	: ۴۵/۲۷۶QA ۱۳۹۲ ۹ب۴/آ
رده بندی دیویی	: ۵۰۲۸۵۵۱۳۳/۵۱۹
شماره کتاب شناسی ملی	: ۳۴۲۸۸۷۷



دانشگاه علوم بهزیستی و توانبخشی



مؤلفان: دکتر اکبر بیگلریان و دکتر عنایت الله بخشی	
شمارگان: ۱۰۰۰ جلد	قیمت: ۱۰۰۰۰ ریال
نوبت چاپ: اول، ۱۳۹۳	چاپ: کاوش
نشانی: اوین، بلوار دانشجو، خیابان کودکان، دانشگاه علوم بهزیستی و توانبخشی	
تلفکس: ۲۲۱۸۰۰۸۱	شابک: ۹۷۸-۶۰۰-۶۴۴۲-۰۷-۵

با قدردانی از همکاری:

- ویراستار علمی: دکتر امید مساح • ویراستار صوری و زبانی: سجاد محمدی (مؤسسه ویراستاران حرفه ای پارس) • صفحه آرا: سیدحمید حیدری ثانی • کنترل فنی صفحه آرای: سیدمحمد موسوی خطاط • هماهنگی و نظارت: پرستو تاجیک • طراحی جلد: گروه گرافیکی ارشیا

همه حقوق نشر برای ناشر محفوظ است. هرگونه استفاده از متن به هر صورتی در قالب کتابچه، جزوه، سی دی آموزشی، نرم افزار، استخراج آزمون، ترجمه به دیگر زبان ها و استفاده از مطالب کتاب، اشکال، تصاویر و... بدون اجازه کتبی از ناشر، غیرقانونی است.



فهرست

پیشگفتار	۹
فصل اول: مروری بر پالایش داده‌ها	۱۱
۱.۱. مقدمه	۱۱
۲.۱. شیوه ورود داده‌ها	۱۱
الف. با استفاده از فرمان (scan)	۱۲
ب. با استفاده از فرمان (C)	۱۲
۳.۱. شیوه فراخوانی داده‌ها	۱۲
۴.۱. شیوه ذخیره‌سازی داده‌ها و خروجی‌ها	۱۵
۱.۴.۱. شیوه حذف یک فایل	۱۷
۲.۴.۱. شیوه ذخیره‌سازی خروجی‌ها در یک فایل متنی	۱۸
۵.۱. کیفی کردن یک متغیر	۱۸
۶.۱. ترسیم نمودار	۱۹
نمودار پراکنش	۲۳
تمرین	۲۶
فصل دوم: تابع نویسی در R	۳۱
۱.۲. مقدمه	۳۱
۲.۲. مقادیر گم‌شده	۳۱
۳.۲. عمل‌گرهای منطقی	۳۲
۴.۲. عبارات شرطی	۳۴

۳۵	۵.۲. حلقه‌ها
۳۸	۶.۲. توابع و ایجاد آن‌ها
۴۱	تمرین
۴۳	فصل سوم: شبیه‌سازی
۴۳	۱.۳. مقدمه
۴۴	۲.۳. مراحل اجرای شبیه‌سازی
۴۴	۱. تعریف مسئله
۴۵	۲. مشخص کردن هدف‌ها
۴۵	۳. بیان شرح کلی از سیستم
۴۵	۴. جمع‌آوری داده‌ها
۴۵	۵. ساخت مدل
۴۶	۶. اعتبارسنجی و اعتباربخشی
۴۶	۷. اجرای مدل
۴۶	۸. تجزیه و تحلیل نتایج
۴۷	۹. جمع‌بندی و ارائه گزارش
۴۸	۳.۳. شبیه‌سازی در آمار
۵۰	۴.۳. نمونه‌گیری تصادفی از توزیع‌های مختلف آماری
۵۰	۳.۴.۱. نمونه‌گیری از یک بردار
۵۱	۳.۴.۲. نمونه‌گیری از یک توزیع آماری مشخص
۵۳	۳.۴.۳. تنظیم دستی مقدار اولیه تولید داده تصادفی
۵۴	۵.۳. شبیه‌سازی مونت کارلویی
۵۶	۶.۳. شبیه‌سازی با نمونه‌گیری بوت استرپ
۵۸	۳.۶.۱. گام‌های اجرای بوت استرپ
۵۹	۳.۶.۲. ضعف‌ها و قوت‌های نمونه‌گیری بوت استرپ
۶۰	۷.۳. قانون قوی اعداد بزرگ
۶۰	۸.۳. روش عکس قضیه تبدیل انتگرال احتمال برای تولید اعداد تصادفی



۹۰.۳	روش استنتاج ریاضی برای تولید اعداد تصادفی	۶۴
۱۰۰.۳	روش رد و قبول برای تولید اعداد تصادفی	۶۷
۱۱۰.۳	روش نمونه‌گیری نقاط پرنفوذ در محاسبه انتگرال با استفاده از شبیه‌سازی	۷۰
۱۲۰.۳	کاربرد تابع نفوذ طبقه‌ای در محاسبه انتگرال با استفاده از شبیه‌سازی	۷۲
۱۳۰.۳	الگوریتم متروپولیس هستینگز و نمونه‌گیری گیبس	۷۹
۱۰۱۳.۳	الگوریتم متروپولیس هستینگز	۸۰
۲۰۱۳.۳	نمونه‌گیری گیبس	۸۶
۳۰۱۳.۳	روش تشخیص همگرایی	۸۹
۱۴۰.۳	سایر مثال‌ها	۹۰
۹۳	تمرین	
۹۷	کتابنامه	
۹۹	واژه‌نامه فارسی انگلیسی	
۱۰۳	واژه‌نامه انگلیسی فارسی	
۱۰۷	نمایه	



پیشگفتار

نرم افزار R از وبگاه <http://www.cran.r-project.org> به صورت رایگان قابل فراخوانی و بارگذاری بر رایانه های شخصی است. این برنامه ای قوی برای تجزیه و تحلیل داده های آماری است که در سال های اخیر، بسیاری از دانشجویان و اساتید رشته آماری (در تمام گرایش های محض، کاربردی، بیمه و زیستی) و برخی رشته های فنی، توجه بسیاری به آن نشان داده اند و علاقه مندی به استفاده از این نرم افزار، به طور چشمگیری در حال افزایش است.

همان طور که می دانید، یکی از روش های تولید اعداد تصادفی یا شبه تصادفی، طوری که رفتاری مشابه با جزء تصادفی الگوی (مدل) طراحی شده داشته باشند، شبیه سازی است. در روش های آماری بیزی و گاهی در روش های آمار کلاسیک، برای استنباط درباره پارامترهای مدل و انجام دادن پیش بینی، به انتگرال گیری از تابع چگالی احتمالی نیاز است که در بیشتر موارد چندان ساده است. در روش های بیزی با انتگرال تابع چگالی پسین پارامترها، به شرط مشاهدات و در روش های فراوانی گرا به انتگرال تابع چگالی احتمال مشاهدات به شرط پارامترها نیاز است. اغلب برای محاسبه این انتگرال ها، از روش های شبیه سازی استفاده می شود.

بر کسی پوشیده نیست که در بیشتر پایان نامه های کارشناسی ارشد و رساله های دکتری دانشجویان رشته آماری زیستی و آمار کاربردی و آمار محض، از شبیه سازی استفاده می شود و لذا، آشنایی و یادگیری شبیه سازی برای دانشجویان رشته آماری در تمامی گرایش ها، بسیار ضروری به نظر می رسد. این کتاب در سه فصل تنظیم شده است. در فصل اول، مقدمه آورده شده و

پالایش و مدیریت داده‌ها معرفی شده است. در ادامه، مقدمات برنامه‌نویسی در R، در فصل دوم آمده است که در آن، عبارات منطقی، شرطی، حلقه‌ها و توابع معرفی شده‌اند. در پایان، شبیه‌سازی در فصل سوم معرفی شده است. در این فصل، به شبیه‌سازی مونت کارلویی، بوت استرپ، روش عکس قضیه تبدیل انتگرال احتمال، روش استنتاج ریاضی، روش رد و قبول، روش نمونه‌گیری نقاط پرنفوذ، الگوریتم متروپولیس هستینگز و نمونه‌گیری گیبس اشاره و سعی شده است که با مثال‌های متنوع، مفاهیم هرکدام شرح داده شود.

نگارش این کتاب براساس تجربه تدریس درس سیستم‌های اطلاع‌رسانی، روش‌های آمار زیستی ۲ و ۳ و شبیه‌سازی در دو دانشگاه علوم بهزیستی و توانبخشی و تربیت مدرس تهران صورت گرفته است. هرچند سعی شده است که با درج مثال‌های متنوع، مفاهیم این کتاب به خوانندگان محترم انتقال یابد؛ ولی ممکن است کاستی‌هایی نیز وجود داشته باشد که از دید نگارندگان پوشیده مانده باشد؛ لذا از تمامی اساتید و دانشجویان و پژوهش‌گران خواهشمند است نظرات پیشنهادی و اصلاحی خود را به رایانامه abiglarian@gmail.com یا abiglarian@uswr.ac.ir یا ebakhshi@uswr.ac.ir ارسال کنند و ما را از راهنمایی‌های خود بهره‌مند سازند.

در پایان، از آقایان دکتر جواد فردمال و احمد رضا باغستانی برای راهنمایی‌ها و در اختیارگذاری مطالب و مثال‌های متنوع و همچنین آقایان دکتر مسعود کریملو و جواد فردمال برای ویراستاری ارزشمند این کتاب سپاس‌گزاری می‌شود.





مروری بر پالایش داده‌ها

۱.۱. مقدمه

در حال حاضر، نسخه 3.0.2 نرم‌افزار R از تارگاه <http://www.cran.r-project.org> به صورت رایگان قابل فراخوانی و بارگذاری بر رایانه‌های شخصی است. این برنامه‌ای قوی برای تجزیه و تحلیل داده‌های آماری است که در سال‌های اخیر، بسیاری از دانشجویان و اساتید رشته آمار (در تمام گرایش‌های محض، کاربردی، بیمه و زیستی) به آن توجه نشان داده‌اند و علاقه‌مندی به استفاده از این نرم‌افزار، به طور چشمگیری در حال افزایش است. در این فصل، ابتدا ورود و فراخوانی و ذخیره‌سازی داده‌ها معرفی می‌شود. در ادامه، مبتنی بر چارچوب داده ایجاد شده، شیوه ترسیم نمودار و انجام دادن برخی عملیات آماری (توصیف داده‌ها) ارائه می‌شود.

در عین حال، خوانندگان محترم برای مطالعه بیشتر، می‌توانند به چهار فصل اول کتاب کاربرد نرم‌افزار R در تجزیه و تحلیل داده‌های آماری، تألیف آقایان دکتر عنایت‌الله بخشی و محمدرضا اشراقیان، انتشارات دانشگاه علوم بهزیستی و توان بخشی مراجعه کنند.

۲.۱. شیوه ورود داده‌ها

پس از نصب و آماده‌سازی نرم‌افزار، می‌توانید داده‌ها را وارد و فرمان‌های مختلف را اجرا کنید. یادآوری می‌شود که شیوه تخصیص و مقداردهی با یکی از عمل‌گرهای < - یا > یا = صورت می‌پذیرد که در این کتاب، از عمل‌گر مساوی = استفاده می‌شود. برای وارد کردن داده‌ها به صورت مستقیم، می‌توانید به دو شیوه زیر عمل کنید: