

۸۱۲۱۰

مدل سازی خطی

با استفاده از نرم افزار R

نویسنده: جولیان فی. فاراوی

مترجمین:

دکتر عباس سقایی

مونا عیسی بخش

www.ketab.ir



سرشناسه

: فاراوی، جولیان جیمز

Faraway, Julian James

عنوان و نام پدیدآور

: مدل سازی خطی با استفاده از نرم افزار R / نویسنده جولیان فی. فاراوی؛ مترجمین عباس سقایی، مونا عیسی بخش.

مشخصات نشر

: تهران: عباس سقایی، ۱۳۹۴.

مشخصات ظاهری

: ۳۰۳ص: مصور، جدول، نمودار.

شابک

: ۲۰۰۰۰۰ ریال 7-3244-04-600-978 :

وضعیت فهرست نویسی

: فیپا

یادداشت

: عنوان اصلی: Linear models with R 2ed, 2015

یادداشت

: ازه نامه

موضوع

: تحلیل واریانس

موضوع

: برگشت (ریاضیات)

شناسه افزوده

: سقایی، عباس، ۱۳۸۶ - مترجم

شناسه افزوده

: عیسی بخش، مونا، ۱۳۸۰ - مترجم

رده بندی کنگره

: ۱۳۹۴ م ۴۲۹/۳ QA۲۷۹

رده بندی دیویی

: ۵۱۹/۵۳۸۰۲۸۵۵۱۳۳

شماره کتابشناسی ملی

: ۴۰۱۰۱۵۷

عنوان: مدل سازی خطی با استفاده از نرم افزار R

ناشر: عباس سقایی

نویسنده: جولیان فی. فاراوی

مترجمین: دکتر عباس سقایی، مونا عیسی بخش

شابک: 7-3244-04-600-978

نوبت چاپ: اول

سال چاپ: ۱۳۹۴

تیراژ: ۱۰۰۰ نسخه

صفحات: ۳۰۳

بها: ۲۰۰۰۰۰ ریال

فهرست مطالب

۱۰	مقدمه
۱۱	شروع کار با R
۱۳	فصل ۱:
۱۳	پیش درآمد
۱۳	۱.۱ در شروع کار
۱۴	۲.۱ تجزیه و تحلیل ابتدائی داده‌ها
۲۱	۳.۱ چه زمانی باید تجزیه و تحلیل رگرسیونی استفاده شود:
۲۲	۴.۱ تاریخچه
۲۵	تمرین‌ها:
۲۷	فصل ۲:
۲۷	برآورد
۲۷	۱.۲ مدل خطی
۲۸	۲.۲ نمایش ماتریسی
۲۹	۳.۲ برآورد β :
۳۰	۴.۲ برآورد حداقل مربعات:
۳۲	۵.۲ نمونه‌هایی از محاسبه β
۳۳	۶.۲ قضیه گاوس-مارکوف:
۳۵	۷.۲ نیکویی برازش
۳۷	۸.۲ مثال
۴۴	تمرین‌ها
۴۵	فصل ۳:
۴۵	استنباط
۴۶	۱.۳ آزمون‌های فرض، برای مقایسه مدل‌ها
۴۸	۲.۳ مثال‌هایی از آزمون
۵۶	۳.۳ آزمون‌های جایگشتی:

۵۹	۴.۳ فواصل اطمینان B:
۶۲	۵.۳ فواصل اطمینان پیش‌بین‌ها:
۶۷	۶.۳ آزمایش‌های طراحی شده:
۷۲	۷.۳ داده‌های مشاهداتی:
۷۸	۸.۳ مشکلات عملی:
۸۲	تمرین‌ها
۸۴	فصل ۴:
۸۴	تشخیص
۸۴	۱.۴ ب.جی و سیات خطا
۸۵	۱.۱.۴ وایانس
۹۱	۱.۲.۴ نرمال ردن:
۹۴	۳.۱.۴ خطاهای همبسته
۹۸	۲.۴ یافتن مشاهدات غیر معمول
۹۹	۱.۲.۴ داده‌های آهرمی
۱۰۱	۲.۲.۴ نقاط دورافتاده
۱۰۷	۳.۲.۴ مشاهدات اثرگذار
۱۰۹	۳.۴ بررسی ساختار مدل
۱۱۳	تمرین‌ها
۱۱۴	فصل ۵:
۱۱۴	مشکلات با پیش‌بین‌ها
۱۱۴	۱.۵ خطای پیش‌بین‌ها:
۱۲۰	۲.۵ تغییرات مقیاس
۱۲۲	۳.۵ هم‌خطی
۱۲۷	تمرین‌ها
۱۳۰	فصل ۶:
۱۳۰	مشکلات با خطاها:

۱۳۰	۱.۶. حداقل مربعات تعمیم یافته
۱۳۵	۲.۶. حداقل مربعات موزون
۱۳۹	۳.۶. آزمون نقص برازش
۱۴۴	۴.۶. رگرسیون استوار
۱۴۵	۱.۴.۶. M-برآورد
۱۴۸	۲.۴.۶. حداقل مربعات پیرواسته
۱۵۴	تمرین ها
۱۵۶	فصل ۷:
۱۵۶	تبدیل
۱۵۷	۱.۷. تبدیل متغیر پاسخ
۱۶۱	۲.۷. تبدیل پیش بین ها
۱۶۲	۱.۲.۷. رگرسیون قطعه ای
۱۶۴	۲.۲.۷. چند جمله ای ها
۱۶۷	۳.۲.۷. اسپلاین های رگرسیون
۱۶۹	۴.۲.۷. مرور
۱۷۰	تمرین ها
۱۷۳	فصل ۸:
۱۷۳	گزینش متغیر
۱۷۴	۱.۸. مدل های سلسله مراتبی
۱۷۵	۲.۸. روش های مبتنی بر آزمون
۱۷۵	۱.۲.۸. حذف پسرو
۱۷۵	۲.۲.۸. گزینش پیشرو
۱۷۶	۳.۲.۸. رگرسیون گام به گام
۱۷۶	۴.۲.۸. قضاوت در مورد روش های مبتنی بر آزمون
۱۷۹	۳.۸. روش های مبتنی بر معیار
۱۸۵	۴.۸. خلاصه

۱۸۶	تمرین‌ها
۱۸۷	فصل ۹:
۱۸۷	روش‌های انقباضی
۱۸۷	۱.۹ مولفه‌های اصلی
۱۹۷	۲.۹ حداقل مربعات جزئی
۲۰۰	۳.۹ رگرسیون تیفه‌ای
۲۰۴	تمرین‌ها
۲۰۵	فصل ۱۰:
۲۰۵	راهبرداآمار و عدم قطعیت مدل
۲۰۵	۱.۱۰ راهبرد
۲۰۶	۲.۱۰ آزمون‌ی در مورد بساطت مدل
۲۰۸	۳.۱۰ مباحثه
۲۱۱	فصل ۱۱:
۲۱۱	خط قرمز بیمه-یک مثال کامل
۲۱۲	۱.۱۱ همبستگی‌های اکولوژیکی
۲۱۴	۲.۱۱ تحلیل ابتدایی داده‌ها
۲۱۸	۳.۱۱ مدل ابتدایی و تشخیص
۲۲۰	۴.۱۱ تبدیل و گزینش متغیر
۲۲۴	۵.۱۱ مباحثه
۲۲۷	فصل ۱۲:
۲۲۷	داده‌های گمشده
۲۳۳	فصل ۱۳:
۲۳۳	تحلیل کواریانس
۲۳۵	۱.۱۳ یک مثال دو سطحی
۲۴۰	۲.۱۳ کدینگ پیش‌بین‌های کیفی

۲۴۲	۳.۱۳ مثالی از عامل چندسطحی
۲۴۸	تمرین‌ها
۲۵۱	فصل ۱۴:
۲۵۱	تحلیل واریانس یک طرفه
۲۵۱	۱.۱۴ مدل
۲۵۳	۲.۱۴ مثال
۲۵۶	۳.۱۴ تشخیص
۲۵۸	۴.۱۴ مقایسات زوجی
۲۶۱	تمرین‌ها
۲۶۳	فصل ۱۵:
۲۶۳	طرح‌های عاملی
۲۶۳	۱.۱۵ ANOVA دوطرفه
۲۶۵	۲.۱۵ ANOVA دوطرفه با یک مشاهده در هر سلول
۲۶۹	۳.۱۵ ANOVA دوطرفه با بیش از یک مشاهده در هر سلول
۲۷۳	۴.۱۵ آزمایش‌های عاملی بزرگتر
۲۷۸	تمرین‌ها
۲۷۹	فصل ۱۶:
۲۷۹	طرح‌های بلوکی
۲۸۰	۱.۱۶ طرح بلوکی تصادفی شده
۲۸۶	۲.۱۶ مربع‌های لاتین
۲۹۰	۳.۱۶ طرح بلوکی ناقص متعادل
۲۹۵	تمرین‌ها:
۲۹۶	ضمیمه A:
۲۹۶	نصب R، توابع و داده‌ها
۲۹۸	ضمیمه B:
۲۹۸	پیش‌درآمدی کوتاه بر R

۲۹۹	۱.B خواندن داده‌ها
۲۹۹	۲.B خلاصه‌های عددی
۳۰۰	۳.B خلاصه‌های نموداری
۳۰۱	۴.B انتخاب زیرمجموعه‌ای از داده‌ها
۳۰۲	۵.B یادگیری بیشتر R
۳۰۴	مراجع

www.ketab.ir

مقدمه

کتاب متعددی در حوزه رگرسیون^۱ و آنالیز واریانس^۲ موجودند. این کتاب نیازمند سطوح آمادگی متفاوتی هستند و به صورت‌های مختلفی بر محتوا تاکید دارند. این کتاب مقدماتی نیست و مطالعه آن نیازمند قدری دانش در حوزه تئوری آماری پایه و تمرین است. از خوانندگان انتظار می‌رود که با ملزومات استنباط آماری مانند برآورد^۳، آزمون‌های فرض^۴ و بازه‌های اطمینان^۵ آشنا باشند. دانشی ابتدایی در حوزه تحلیل داده، قدری اطلاعات در مورد جبرخطی و محاسبات نیز مورد نیاز است.

تاکید این نوشتار بر تمرین رگرسیون و آنالیز واریانس است. هدف این است که بدانیم چه روش‌هایی وجود دارند، از آن مهم‌تر این روش‌ها را چه زمانی باید به کار برد. برای شفاف‌سازی کاربرد این روش‌ها و نمایاندن نتایج حاصل از آن‌ها، مثال‌های متعددی ارائه شده است. بر مباحث تئوری ریاضی تاکید چندانی نشده است، چراکه انتظار می‌رود خوانندگان تا حدی در این مورد اطلاعات قبلی داشته باشند و از طرفی مناسب است که در جایی دیگر به این موضوعات پرداخته شود. مباحث تئوری حائز اهمیتند؛ چراکه رویکردهای استنباطی را هدایت می‌کنند. به تئوری آماری به طور عمیق‌تری پرداخته‌ایم. این تنها شامل قضایای مرسوم نیست. مفاهیم آماری کیفی نیز به همان میزان مهمند چراکه ما را قادر می‌سازند به جای تنها صحبت کردن در مورد موضوعی، دست به عمل بزنیم. یادگیری این قواعد کیفی سخت‌تر است، چراکه بیان دقیق آنها دشوار می‌باشد؛ اما این قواعد آمارگران موفق و با تجربه را یاری می‌نمایند.

تحلیل داده را نمی‌توان بدون تمرین آموخت؛ یعنی باید از یک بسته محاسباتی آماری استفاده شود. گزینه‌های زیادی برای انتخاب از میان این بسته‌ها وجود دارد. این بسته‌ها برای استفاده افراد مختلفی طراحی شده‌اند و دارای نقاط قوت و ضعف متفاوتی هستند. در این کتاب ما^۶ را انتخاب کرده‌ایم. دلایل متعددی برای انتخاب R وجود دارد:

تطبیق‌پذیری: R یک زبان برنامه‌نویسی نیز است. در نتیجه در مورد برایندهای از پیش برنامه‌نویسی شده در یک بسته نرم‌افزاری دچار محدودیت نمی‌شویم. برنامه‌نویسی روش‌های جدید در R نسبتاً ساده است.

¹ regression

² analysis of variance

³ estimation

⁴ hypothesis test

⁵ confidence interval

تعامل: تحلیل داده به طور ذاتی فرآیندی تعاملی است. برخی از بسته‌های آماری قدیمی زمانی طراحی شدند که محاسبات گران‌تر بوده و پردازش دسته‌ای باب بود. با وجود پیشرفت سخت‌افزارها، پارادایم قدیمی پردازش دسته‌ای در استفاده از آنها باقی مانده است. R در هر قدم یک عمل انجام می‌دهد و ما را قادر می‌سازد که بر اساس آنچه حین تحلیل می‌بینیم تغییراتی اعمال کنیم.

آزادی: R بر پایه S-که بسته تبلیغاتی S-plus از آن مشتق گشته- ایجاد شده است. R نرم افزاری متن‌باز^۱ است و به صورت رایگان در اختیار افراد قرار می‌گیرد.^۲ R و S-plus با یکدیگر سازگارند؛ بدین معنی که برای بسیاری از مثال‌های استفاده شده در این کتاب می‌توان از S-plus استفاده کرد.

محبوبیت: در کاربردهای عمومی، SAS رایج‌ترین بسته آماری است، اما R یا S میان محققان آماری محبوب‌ترند یا نگاهی به نشریات آماری این محبوبیت تایید می‌شود. R همچنین برای کاربردهای کل در امور مالی دارای شهرت است.

شروع کار با R

یادگیری R نیازمند آشنایی با روشش است. این کوشش با افزایش بهره‌وری پاسخ داده خواهد شد. شما می‌توانید طریقه دسترسی به R^۳ و همچنین نصب نرم‌افزارها و داده‌های فرعی دیگری که در این کتاب استفاده شده‌اند را در ضمیمه^۴ ببینید.

این کتاب مقدمه‌ای بر R نیست. در ضمیمه^۲ زبان برنامه‌نویسی به طور مختصر توضیح داده شده است، اما این به تنهایی کفایت نمی‌کند. تمامی فرمان‌هایی که برای تولید نتایج موجود در این کتاب استفاده شده‌اند به طور عمدی در متن قرار داده شده‌اند. این یعنی شما می‌توانید این تحلیل‌ها را دوباره تولید کرده و تغییرات را پیش از آشنایی کامل با R تجربه کنید.^۳

در انتها جا دارد از سازندگان این نرم‌افزار -که بدون آنها تهیه این کتاب امکان‌پذیر نبود- سپاس‌گزاری کنیم.^۴

^۱ open-source software

^۲ نسخه‌های لینوکس، مکینتاش، ویندوز و سایر نسخ UNIX را می‌توان از www.r-project.org تهیه کرد.

^۳ می‌توانید برای دسترسی به راهنمای مقدماتی مجانی R به وبسایت www.r-project.org مراجعه نمایید.

^۴ داده‌های ذکر شده در متن کتاب را می‌توانید در وبسایت www.stat.lsa.umich.edu/~faraway/LMR بیابید.