

مبانی آمار ریاضی
(به همراه دستوره‌های R)

www.ketab.ir
عبدالرضا سیاره



شماره ۵۴۳

سرشناسه: سیاره، عبدالرضا، ۱۳۴۴ - Sayyareh, Abdolreza

عنوان و نام پدیدآور: مبانی آمار ریاضی (به همراه دستورهای R) / عبدالرضا سیاره.
 مشخصات نشر: تهران: دانشگاه صنعتی خواجه نصیرالدین طوسی، انتشارات، ۱۴۰۲.
 مشخصات ظاهری: ۵۲۶ص: مصور.

شابک: ۹۷۸-۶۲۲-۵۲۳۴-۲۶-۰

ISBN: 978-622-5234-26-0

وضعیت فهرست نویسی: فیپا.

یادداشت: واژه‌نامه.

یادداشت: کتابنامه.

یادداشت: نمایه.

موضوع: آمار ریاضی -- راهنمای آموزشی (عالی) / Mathematical statistics--Study and teaching (Higher)

آمار -- راهنمای آموزشی (عالی) / Statistics -- Study and teaching (Higher)

رده‌بندی کنگره: QA۲۷۶/۱۸

رده‌بندی دیویی: ۵۱۹/۵۰۷۶

شماره کتابشناسی ملی: ۹۴۲۳۹۹۱

press.kntu.ac.ir



ناشر: دانشگاه صنعتی خواجه نصیرالدین طوسی

عنوان: مبانی آمار ریاضی (به همراه دستورهای R)

مؤلف: دکتر عبدالرضا سیاره

نوبت چاپ: اول

تاریخ انتشار: آذر ۱۴۰۲

شمارگان: ۲۰۰ جلد

چاپ و صحافی: آرمانسا

قیمت: ۳۸۷,۰۰۰ تومان

تمام حقوق برای ناشر محفوظ است

خیابان میرداماد غربی - شماره ۴۷۰ - انتشارات دانشگاه صنعتی خواجه نصیرالدین طوسی - تلفن: ۸۸۸۸۱۰۵۲

میدان ونک - خیابان ولی‌عصر (ع) - بالاتر از چهارراه میرداماد - شماره ۲۶۲۶ - مرکز پخش و فروش انتشارات

تلفن: ۸۸۷۷۲۲۷۷ رایانامه: press@kntu.ac.ir - تارنما (فروش برخط): press.kntu.ac.ir

فهرست مطالب

فهرست مطالب ز

اول متغیرهای تصادفی

۱	توزیع‌های متغیرهای تصادفی	۱۰۱
۷	۱.۱.۱ تابع توزیع تجمعی	۱۷
۲۰	۲.۱.۱ بردارهای تصادفی	۲۰
۲۳	۲.۱ تابع احتمال متغیر تصادفی گسسته	۲۳
۳۲	۳.۱ تابع چگالی احتمال متغیر تصادفی پیوسته	۳۲
۴۲	۴.۱ توابعی از متغیرهای تصادفی	۴۲
۴۴	۵.۱ امید ریاضی	۴۴
۴۴	۱.۵.۱ مقدمه	۴۴
۴۵	۶.۱ میانگین و واریانس	۴۵
۵۱	۱.۶.۱ امید ریاضی و نامساوی‌ها	۵۱
۵۵	۷.۱ گشتاورهای توزیع	۵۵
۵۹	۸.۱ تابع مولد گشتاورها	۵۹
۶۷	۹.۱ تابع‌های مشخص و گشتاورهای فاکتوریل	۶۷
۷۷	۱۰.۱ تمرین‌ها	۷۷

دوم توزیع‌های احتمال

۸۵	۱.۲ مقدمه	۸۵
۸۶	۲.۲ تابع احتمال	۸۶
۸۹	۳.۲ برخی توزیع‌های گسسته	۸۹
۸۹	۱.۳.۲ توزیع یکنواخت گسسته	۸۹

۹۳	توزیع‌های برنولی و دوجمله‌ای	۲۰۳.۲
۱۱۳	توزیع دوجمله‌ای منفی	۳۰۳.۲
۱۱۷	توزیع هندسی	۴۰۳.۲
۱۲۱	توزیع پواسون	۵۰۳.۲
۱۲۸	فرایند پواسون	۶۰۳.۲
۱۳۱	توزیع فوق‌هندسی	۷۰۳.۲
۱۳۵	برخی توزیع‌های پیوسته	۴.۲
۱۳۵	توزیع یکنواخت پیوسته	۱۰۴.۲
۱۳۹	توزیع نرمال	۲۰۴.۲
۱۶۱	توزیع بتا	۳۰۴.۲
۱۶۳	توزیع لگ-نرمال	۴۰۴.۲
۱۶۵	توزیع گاما	۵۰۴.۲
۱۶۸	توزیع خی دو	۶۰۴.۲
۱۶۹	توزیع وایبل	۷۰۴.۲
۱۷۰	توزیع نمایی	۸۰۴.۲
۱۷۵	توزیع لجستیک	۹۰۴.۲
۱۷۷	توزیع کوشی	۱۰۰۴.۲
۱۷۹	توزیع نمایی دوگانه	۱۱۰۴.۲
۱۸۳	توزیع مثلثی	۱۲۰۴.۲
۱۸۵	تمرین‌ها	۵.۲
۱۹۱	سوم توزیع‌های چندمتغیره	
۱۹۲	تابع احتمال دو متغیره گسسته	۱.۳
۲۰۴	تابع چگالی احتمال دو متغیره پیوسته	۲.۳
۲۰۹	توزیع‌های شرطی	۳.۳
۲۱۰	متغیرهای تصادفی گسسته	۱۰۳.۳
۲۱۲	متغیرهای تصادفی پیوسته	۲۰۳.۳
۲۱۸	متغیرهای تصادفی پیوسته-گسسته	۳۰۳.۳
۲۲۰	استقلال متغیرهای تصادفی	۴.۳
۲۲۱	امید ریاضی توأم و شرطی	۱۰۴.۳
۲۲۸	ویژگی‌های امید ریاضی توأم و شرطی	۵.۳

۶.۳	کوواریانس بین دو متغیر تصادفی	۲۲۹
۱.۶.۳	ضریب همبستگی بین دو متغیر تصادفی	۲۳۳
۲.۶.۳	توزیع نرمال دو متغیره	۲۴۴
۷.۳	توزیع‌های چندمتغیره	۲۴۷
۱.۷.۳	توزیع‌های حاشیه‌ای و شرطی در توزیع چندجمله‌ای	۲۵۳
۸.۳	نمونه تصادفی ساده	۲۶۵
۹.۳	مدل‌های سلسله مراتبی	۲۶۸
۱.۹.۳	امید ریاضی امید متغیر تصادفی و واریانس شرطی	۲۶۹
۲.۹.۳	معادله والد	۲۷۶
۱۰.۳	تمرین‌ها	۲۷۸

چهارم توزیع توابعی از متغیرهای تصادفی

۱.۴	مقدمه	۲۸۳
۲.۴	مقدمه‌ای بر نظریه نمونه‌گیری	۲۸۶
۳.۴	روش تعویض متغیر	۲۸۹
۴.۴	تعویض متغیر برای تبدیلات کلی	۳۱۳
۵.۴	روش تابع توزیع تجمعی	۳۱۶
۱.۵.۴	آماره‌های ترتیبی	۳۲۰
۶.۴	چگالی آماره ترتیبی دلخواه	۳۲۴
۷.۴	روش تابع مولد گشتاورها	۳۳۱
۸.۴	توزیع میانگین و واریانس نمونه‌ای و توابعی از آن‌ها	۳۳۳
۱.۸.۴	اتحاد تجزیه واریانس	۳۳۷
۹.۴	تمرین‌ها	۳۴۰

پنجم آمار بزرگ نمونه‌ای و قضایای مجانبی

۱.۵	مقدمه	۳۴۳
۲.۵	انواع همگرایی‌ها	۳۴۴
۳.۵	روابط بین همگرایی‌ها	۳۴۸
۴.۵	توزیع مجانبی آماره‌ها	۳۵۶
۱.۴.۵	قانون اعداد بزرگ	۳۵۹
۲.۴.۵	قضیه حد مرکزی	۳۶۴
۵.۵	روش دلتا	۳۶۸

۳۷۳	تمرین‌ها	۶.۵
۳۷۷	ششم مباحثی در توزیع‌ها	
۳۷۷	مقدمه	۱.۶
۳۷۸	خانواده توزیع‌های مکان و مقیاس	۲.۶
۳۸۱	خانواده توزیع‌های نمایی	۳.۶
۳۸۷	چند توزیع ویژه	۴.۶
۳۸۷	توزیع زتای ریمان	۱.۴.۶
۳۸۷	توزیع مقادیر غایی	۲.۴.۶
۳۸۸	توزیع بُر	۳.۴.۶
۳۹۰	برخی ویژگی‌های خانواده‌های توزیع‌ها	۵.۶
۳۹۱	تمرین‌ها	۶.۶
۳۹۳	هفتم مباحث ویژه	
۳۹۳	توزیع‌های نمونه‌ای	۱.۷
۳۹۵	نمونه‌گیری تصادفی ساده	۱.۱.۷
۳۹۷	برآورد نقطه‌ای پارامترها	۲.۷
۳۹۹	برآورد گشتاوری	۱.۲.۷
۴۰۲	برآورد بیشینه درست‌نمایی	۲.۲.۷
۴۱۳	اطلاع فیشر	۳.۷
۴۱۳	اندازه اطلاع	۱.۳.۷
۴۱۵	اطلاع فیشر در حالت یک پارامتری	۲.۳.۷
۴۱۹	اطلاع فیشر در حالت چندپارامتری	۳.۳.۷
۴۲۲	تولید اعداد تصادفی	۴.۷
۴۲۲	تولید اعداد تصادفی با دنباله فان درکوپوت	۱.۴.۷
۴۲۶	تولید اعداد تصادفی با روش تبدیل وارون	۲.۴.۷
۴۳۳	آنتروپی	۵.۷
۴۳۷	آنتروپی توأم، آنتروپی شرطی و آنتروپی شرطی مورد انتظار	۱.۵.۷
۴۴۰	ویژگی‌های آنتروپی	۲.۵.۷
۴۴۱	آنتروپی و رمزگذاری	۳.۵.۷
۴۴۲	آنتروپی توأم و اطلاع متقابل	۴.۵.۷
۴۴۴	مخاطره کولبک-لیبلر برای انتخاب مدل	۶.۷

۴۴۴	مقدمه	۱۰۶۰۷
۴۴۴	معیار واگرایی کولبک-لیبلر	۷۰۷
۴۵۱	برآوردگر بیشینه درست‌نمایی تحت تابع چگالی نامناسب	۸۰۷
۴۵۳	تحلیل بقاء و مدل‌هایی برای داده‌های سانسور شده	۹۰۷
۴۶۱	تمرین‌ها	۱۰۰۷

۴۶۵

کتاب‌نامه

۴۶۷

اول پیوست آ

۴۶۹	آشنایی با R	۱.آ
۴۷۱	محاسبات در R	۲.آ
۴۷۲	بردار و ماتریس	۳.آ
۴۷۵	محاسبات برداری	۴.آ
۴۸۷	چارچوب داده‌ای	۵.آ
۴۸۸	رسم نمودار	۶.آ
۴۹۲	برنامه و تابع نویسی	۷.آ
۴۹۵	ساختار شرطی	۸.آ
۴۹۸	ساختار تکرار	۹.آ

۵۰۱

نمایه

۵۰۶

واژه‌نامه

پیش‌گفتار مؤلف

با توجه به کاربرد و اهمیت روزافزون روش‌های آماری در کمک به درک و توجیه مباحث علمی، تصمیم به تألیف کتاب حاضر گرفته شد. کتاب در هفت فصل تنظیم شده است. به جز فصل ۷، سطح این کتاب به گونه‌ای است که برای دانشجویان دوره‌های کارشناسی همه رشته‌هایی که با دروس مقدماتی آمار و احتمال آشنا هستند؛ قابل استفاده است. کتاب حاضر می‌تواند مورد استفاده دانشجویان رشته‌های آمار، ریاضی، علوم کامپیوتر و سایر علاقه‌مندان در رشته‌های دیگر و در کلیه مقاطع و به خصوص در دوره کارشناسی قرار بگیرد.

سعی شده است؛ سرفصل‌های درس احتمال یک در دوره کارشناسی در کتاب گنجانده شود. به علاوه، تلاش بر این بوده است؛ تا مباحث به گونه‌ای آورده شوند که خواننده برای درک بهتر سایر دروس و پروژه‌های خود مرجعی در زمینه مقدمات استنباط آماری، در اختیار داشته باشد. با توجه به بیان مثال‌هایی از زمینه‌های مختلف و ارائه کدهای R در مثال‌های متنوع، کتاب به صورت خودخوان نیز قابل استفاده است.

این کتاب، به‌ویژه بر پایه دروس احتمال یک در رشته‌های آمار، ریاضی و علوم کامپیوتر در دوره کارشناسی نگارش شده است و پیش‌بینی می‌شود؛ مخاطبینی در رشته‌های علوم پایه، مهندسی، علوم داده، علوم کامپیوتر داشته باشد. در پایان کتاب، فصل ضمیمه در مورد مقدمات برنامه‌نویسی به زبان R آورده شده است.

در فصل یکم به بررسی مقدمات احتمال و بیش‌تر به متغیر تصادفی و مطالب وابسته به آن پرداخته شده است. تعاریفی از تابع توزیع تجمعی، توابع چگالی، امید ریاضی، گشتاورها و توابع مولد گشتاورها به همراه برنامه‌های R مرتبط با موضوع در هر بخش، این فصل را تشکیل می‌دهد.

نظریه توزیع‌ها برای متغیرهای تصادفی گسسته و پیوسته با توجه به خانواده مدل‌های آماری به همراه توابع مولد گشتاورها برای هر توزیع به تفکیک و هم‌چنین برنامه‌های مناسب R در فصل دوم آورده شده‌اند. توزیع‌های مهم آماری، به تفکیک توزیع‌های گسسته و پیوسته

معرفی و برخی ویژگی‌های آن‌ها بیان شده است.

توزیع‌های چندمتغیره و به‌ویژه توزیع‌های دو متغیره در فصل سوم آورده شده است. این فصل از مهم‌ترین بخش‌های کتاب است. مفاهیمی هم‌چون توزیع‌های شرطی، توزیع‌های حاشیه‌ای، امید ریاضی توأم و امید ریاضی شرطی در این فصل معرفی شده‌اند. در ارتباط با توزیع‌های دو متغیره، مفاهیمی پرکاربرد مانند کواریانس و ضریب همبستگی بین دو متغیر تصادفی به همراه ویژگی‌های آن‌ها به تفکیک متغیرهای تصادفی گسسته و پیوسته بیان شده است. دو توزیع مهم آماری، توزیع نرمال دو متغیره و توزیع چندجمله‌ای در این فصل آورده شده‌اند. آخرین بخش این فصل در بردارنده مدل‌های سلسله مراتبی، بیان امید ریاضی و واریانس متغیرهای تصادفی به کمک امید ریاضی امید و واریانس شرطی که کاربردهای بسیار گسترده‌ای دارند؛ بیان و مقدمه‌ای بر نمونه تصادفی در بخش‌های انتهایی فصل آورده شده است. بیان نمونه تصادفی در این فصل به این دلیل است که توزیع نمونه تصادفی ارتباط نزدیکی با عنوان فصل دارد و محاسبه امید ریاضی و واریانس آماره‌ای مانند میانگین نمونه‌ای، می‌تواند کاربرد محاسبات مربوط به توابعی از متغیرهای تصادفی را که به‌طور غیررسمی در فصول اول تا سوم آورده شده است؛ تکمیل نماید. معادله والد آخرین بخش این فصل را تشکیل می‌دهد.

فصل چهارم با مقدمه‌ای بر نظریه نمونه‌گیری آغاز شده است تا در سرتاسر فصل، محاسبات روی متغیرهای تصادفی و توابعی از آن‌ها بهتر درک شود. به‌طور کلی، در این فصل به بررسی روش‌های یافتن توزیع توابعی از متغیرهای تصادفی پرداخته می‌شود. آماره‌های ترتیبی که در آنالیز بقاء و بررسی داده‌های سانسور شده کاربردهای فراوانی دارند؛ در این فصل معرفی شده است. توزیع‌های t -ستودنت و F -فیشر، به کمک روش تعویض متغیر معرفی و توزیع χ^2 دو به‌طور گسترده‌ای مطالعه شده است. در هر بخشی که امکان‌پذیر بوده است؛ از برنامه‌های R برای درک بهتر مفاهیم کمک گرفته شده است.

آمار بزرگ نمونه‌ای و قضایای مجانبی در فصل پنجم آورده شده است. این فصل از نتایج فصول پیشین کمک می‌گیرد؛ تا برای حجم نمونه بزرگ به بررسی رفتار متغیرهای تصادفی بپردازد. وقتی مقادیر نمونه‌ای محاسبه می‌شوند؛ باید حداقل برای حجم نمونه بزرگ، به مقدار واقعی جامعه نزدیک شوند؛ تا دلیلی برای مطالعه آن‌ها وجود داشته باشد. از سوی دیگر توزیع توابعی از متغیرهای تصادفی در حجم نمونه بزرگ حائز اهمیت است. در این فصل قوانین اعداد بزرگ و قضیه حد مرکزی در اشکال مختلف بیان شده است. این قضیه کمک می‌کند؛ تا در حالت‌های گسسته و پیوسته از همگرایی در توزیع صورت استاندارد شده آماره‌های ویژه‌ای، به توزیع نرمال استاندارد اطمینان حاصل کنیم.

فصل ششم، فصل کوتاهی است در رابطه با برخی خانواده‌ها از توزیع‌های احتمال، از

جمله، خانواده‌های مکانی-مقیاسی و خانواده توزیع‌های نمایی که کاربردهای گسترده‌ای در استنباط آماری دارند و بسیاری از ویژگی‌های آماری در این خانواده‌ها ساختار ساده‌ای می‌یابند.

در فصل هفتم که تحت عنوان مباحث ویژه آورده شده است؛ به برخی مفاهیم مقدماتی در استنباط آماری پرداخته شده است که پس از مطالعه شش فصل اول کتاب، برای خواننده به راحتی قابل درک است. اگر چه به دلیل سطح این کتاب هیچ‌کدام از مباحث مطرح شده بسط و گسترش داده نشده‌اند؛ اما امید می‌رود؛ زمینه مناسبی برای علاقه‌مندان فراهم کند. مباحثی روی توزیع‌های نمونه‌ای آورده شده است که در فصول قبل به آن اشاره شده است. یافتن برآوردهای پارامترها در مدل‌های آماری با دو روش برآورد گشتاوری و روش بیشینه درست‌نمایی ارائه شده است. این روش‌ها به طور گسترده‌ای در نظریه برآورد و انتخاب مدل به کار می‌روند. این روش‌ها، به کمک برنامه R نیز به طور مختصر آموزش داده شده است. نقش اطلاع فیشر در همه علوم به ویژه آمار و علوم کامپیوتر باعث شد؛ تا بخش ویژه‌ای به این موضوع اختصاص داده شود. اطلاع فیشر در مباحث یادگیری ماشین و علوم داده به منظور کاهش واریانس، توسط محققین به دفعات مورد استفاده قرار گرفته است. تولید اعداد تصادفی به عنوان زمینه‌ای برای شبیه‌سازی آورده شده و به دنبال آن آنتروپی معرفی شده است. ارتباط آنتروپی با معیار شناخته شده و اگرایی کولبک-لیبلر که ارتباط نزدیکی با برآوردگر بیشینه درست‌نمایی دارد به خواننده کمک می‌کند؛ تا زمینه‌ای در یادگیری ماشین به دست آورد. آخرین بخش این فصل، به طور مختصر به معرفی انواع داده‌های مانسوریده و تحلیل بقاء پرداخته است.

در ضمیمه این کتاب، اشاره‌ای به برخی دستورهای R شده است؛ تا مطالعه برنامه‌های موجود در کتاب را تسهیل کند.

من مرهون اساتید بزرگوام هستم که سهم بسزایی در نوشتار حاضر دارند. امیدوارم این کتاب مورد قبول این بزرگان واقع گردد.

سرانجام، اگر چه از سوی نویسنده کوشش شده است؛ تا این کار بدون کاستی انجام شود؛ ولی نمیداند که تا چه اندازه به این هدف نائل شده است. یادآوری کاستی‌ها و نارسایی‌های کتاب موجب سپاسگزاری است.

عبدالرضا سیاره

پاییز ۱۴۰۲