



دانشگاه ساری
پیت

روش‌های پیشرفته آماری

دکتر پرویز نصیری دکتر محمدعلی ابراهیمی دکتر محمد جعفر آقایی

امروزه کتاب‌خوانی و علم‌آموزی، نه تنها یک وظیفه‌ی ملی، که یک واجب دینی است.

مقام معظم رهبری

در عصر حاضر یکی از شاخصه‌های ارزیابی رشد، توسعه و پیشرفت فرهنگی هر کشوری میزان تولید کتاب، مطالعه و کتاب‌خوانی مردم آن مرز و بوم است. ایران اسلامی نیز از دیرباز تاکنون با داشتن تمدنی چندهزارساله و مراکز متعدد علمی، فرهنگی، کتابخانه‌های معتبر، علما و دانشمندان بزرگ با آثار ارزشمند تاریخی، سرآمد دولت‌ها و ملت‌های دیگر بوده و در عرصه‌ی فرهنگ و تمدن جهانی به‌سان خورشیدی تابناک همچنان می‌درخشد و با فرزندان نیک‌نهاد خویش هنرنمایی می‌کند. چه کسی است که در دنیا با دانشمندان فرزانه و نام‌آور ایرانی همچون ابوعلی سینا، ابوریحان بیرونی، فارابی، خوارزمی و ... همچنین شاعران برجسته‌ای نظیر فردوسی، سعدی، مولوی، حافظ و ... آشنا نباشد و در مقابل عظمت آنها سر تعظیم فرود نیاورد. تمامی این افتخارات ارزشمند، برگرفته از میزان عشق و علاقه فراوان ملت ما به فراگیری علم و دانش از طریق خواندن و مطالعه منابع و کتاب‌های گوناگون است. به شکرانه‌ی الهی، تاریخ و گذشته ما، همیشه درخشان و پربار است. ولی اکنون در این زمینه در چه جایگاهی قرار داریم؟ آمار و ارقام ارائه‌شده از سوی مجامع و سازمان‌های فرهنگی در مورد سرانه‌ی مطالعه‌ی هر ایرانی، برایمان چندان امیدوارکننده نمی‌باشد و رهبر معظم انقلاب اسلامی نیز از این وضعیت بارها اظهار گله و ناخشنودی نموده‌اند.

کتاب، دروازه‌ای به سوی گستره‌ی دانش و معرفت است و کتاب خوب، یکی از بهترین ابزارهای کمال بشری است. همه‌ی دستاوردهای بشر در سراسر عمر جهان تا آنجا که قابل کتابت بوده است، در میان دست‌نوشته‌هایی است که انسان‌ها پدید آورده و می‌آورند. در این مجموعه‌ی بی‌نظیر، تعالیم الهی، درس‌های پیامبران به بشر، و همچنین علوم مختلفی است که سعادت بشر بدون آگاهی از آنها امکان‌پذیر نیست. کسی که با دنیای زیبا و زندگی‌بخش کتاب ارتباط ندارد بی‌شک از مهم‌ترین دستاوردهای انسانی و نیز از بیشترین معارف الهی و بشری محروم است. با این دیدگاه، به‌روشنی می‌توان ارزش و مفهوم رمزی عمیق در این حقیقت تاریخی را دریافت که اولین خطاب خداوند متعال به پیامبر گرامی اسلام(ص) این است که «بخوان!» و در اولین

سوره‌ای که بر آن فرستاده‌ی عظیم‌الشأن خداوند، فرود آمده، نام «قلم» به تجلیل یاد شده‌است: «إِقْرَأْ وَرَبُّكَ الْأَكْرَمُ. الَّذِي عَلَّمَ بِالْقَلَمِ» در اهمیت عنصر کتاب برای تکامل جامعه‌ی انسانی، همین بس که تمامی ادیان آسمانی و رجال بزرگ تاریخ بشری، از طریق کتاب جاودانه مانده‌اند.

دانشگاه پیام‌نور با گستره‌ی جغرافیایی ایران‌شمول خود با هدف آموزش برای همه، همه‌جا و همه‌وقت، به‌عنوان دانشگاهی کتاب‌محور در نظام آموزش عالی کشورمان، افتخار دارد جایگاه اندیشه‌سازی و خردورزی بخش عظیمی از جوانان جویای علم این مرز و بوم باشد. تلاش فراوانی در ایام طولانی فعالیت این دانشگاه انجام پذیرفته تا با بهره‌گیری از تجربه‌های گرانقدر استادان و صاحب‌نظران برجسته کشورمان، کتاب‌ها و منابع آموزشی درسی شاخص و خودآموز تولید شود. در آینده هم، این مهم با هدف ارتقای سطح علمی، روزآمدی و توجه بیشتر به نیازهای مخاطبان دانشگاه پیام‌نور با جدیت ادامه خواهد داشت. به‌طور قطع استفاده از نظرات استادان، صاحب‌نظران و دانشجویان محترم، ما را در انجام این وظیفه‌ی مهم و خطیر یاری‌رسان خواهد بود. پیشاپیش از تمامی عزیزانی که با نقد، تصحیح و پیشنهادهای خود ما را در انجام این وظیفه‌ی خطیر یاری می‌رسانند، سپاسگزاری می‌نماییم. لازم است از تمامی اندیشمندانی که تاکنون دانشگاه پیام‌نور را منزلگاه اندیشه‌سازی خود دانسته و ما را در تولید کتاب و محتوای آموزشی درسی یاری نموده‌اند، صمیمانه قدردانی گردد. موفقیت و بهروزی تمامی دانشجویان و دانش‌پژوهان عزیز آرزوی همیشگی ما است.

دانشگاه پیام‌نور

فهرست مطالب

پیشگفتار	یازده
فصل اول. مدل‌های خطی و تجزیه و تحلیل چندمتغیره	۱
هدف کلی	۱
اهداف یادگیری	۱
۱-۱ معرفی روش‌های یک متغیره و چند متغیره	۱
۲-۱ احتمالات	۲
۳-۱ آزمون فرض	۲
۴-۱ تجزیه‌ی واریانس	۳
۵-۱ همبستگی و رگرسیون	۴
۶-۱ روش‌های چند متغیره	۶
۷-۱ توصیف جوامع چند متغیره	۹
۸-۱ همبستگی جزئی	۹
۹-۱ رگرسیون چندمتغیره	۱۰
۱۰-۱ تجزیه به مؤلفه‌های اصلی	۱۰
۱۱-۱ تجزیه به عامل‌ها	۱۱
۱۲-۱ تابع تشخیص	۱۱
۱۳-۱ تجزیه کلاستر	۱۲
فصل دوم. جبر ماتریسی	۱۳
هدف کلی	۱۳
اهداف یادگیری	۱۳
مقدمه	۱۳
۱-۲ ماتریس	۱۴
۲-۲ نمایش ماتریس و بردار	۱۴

۱۵	۲-۳ تساوی دو ماتریس
۱۶	۲-۴ جمع و تفریق ماتریس‌ها
۱۷	۲-۵ قطر اصلی و قطر فرعی ماتریس
۱۷	۲-۶ ماتریس صفر و ماتریس قطری
۱۸	۲-۷ ماتریس واحد
۱۸	۲-۸ ماتریس یک
۱۹	۲-۹ اثر ماتریس
۱۹	۲-۱۰ برگردان یک ماتریس
۲۰	۲-۱۱ ماتریس متقارن
۲۰	۲-۱۲ ضرب دو ماتریس
۲۲	۲-۱۳ قواعد ضرب
۲۲	۲-۱۴ نرم یک بردار
۲۳	۲-۱۵ بردار یکه
۲۴	۲-۱۶ بردارهای متعامد
۲۵	۲-۱۷ صورت‌های درجه دوم
۲۵	۲-۱۸ ماتریس متمم
۲۶	۲-۱۹ دترمینان ماتریس
۲۸	۲-۲۰ ماتریس ویژه
۲۹	۲-۲۱ ماتریس مزدوج
۲۹	۲-۲۲ معکوس ماتریس
۳۲	۲-۲۳ معکوس ماتریس قطری
۳۳	۲-۲۴ مقادیر و بردارهای ویژه
۳۵	۲-۲۵ حل معادلات با استفاده از روش جبر ماتریسی
۳۶	۲-۲۶ بردار و ماتریس تصادفی
۳۷	۲-۲۷ امید ریاضی یک بردار یا یک ماتریس
۳۸	خودآزمایی فصل دوم
۴۱	فصل سوم. آزمون فرض و استنباط آماری
۴۱	هدف کلی
۴۱	اهداف یادگیری
۴۱	مقدمه
۴۲	۳-۱ فرض صفر در مقابل فرض یک
۴۴	۳-۲ خطای نوع اول و خطای نوع دوم در آزمون فرض
۴۵	۳-۳ یک دامنه و دو دامنه‌بودن آزمون
۴۶	۳-۴ برآورد احتمال
۴۷	۳-۵ توزیع نرمال
۵۱	۳-۶ تبدیل توزیع نرمال به توزیع نرمال استاندارد
۵۶	۳-۷ آزمون فرض درباره‌ی میانگین یک جامعه نرمال

۶۱	۸-۳ آزمون فرض در باره‌ی نسبت جامعه (P)
۶۳	۹-۳ آزمون فرض درباره‌ی برابری میانگین‌های دو جامعه‌ی نرمال مستقل
۶۷	۱۰-۳ آزمون فرض درباره‌ی برابری میانگین‌های دو جامعه نرمال همبسته
۶۹	۱۱-۳ آزمون فرض درباره‌ی برابری واریانس‌های دو جامعه نرمال
۷۱	خودآزمایی فصل سوم

۷۵	فصل چهارم. توصیف و نمایش داده‌های چند متغیره
۷۵	اهداف کلی
۷۵	اهداف یادگیری
۷۶	مقدمه
۷۶	۱-۴ میانگین و واریانس یک متغیره تصادفی
۷۷	۲-۴ کوواریانس و همبستگی متغیرهای تصادفی توأم
۸۰	۳-۴ همبستگی
۸۳	۴-۴ نمایش گرافیکی نمونه‌های دو یا چندمتغیره
۸۵	۵-۴ بردار میانگین
۸۷	۶-۴ ماتریس واریانس-کوواریانس
۹۰	۷-۴ ماتریس واریانس-کوواریانس نمونه‌ای
۹۱	۸-۴ ماتریس ضریب همبستگی جامعه و نمونه
۹۴	۹-۴ رابطه‌ی ماتریس واریانس-کوواریانس و ماتریس همبستگی
۹۴	۱۰-۴ توزیع نرمال چند متغیره
۹۵	۱۱-۴ برآورد تغییرات کل
۹۷	۱۲-۴ تعیین فواصل میان بردارها
۹۸	۱۳-۴ انواع همبستگی
۹۸	۱-۱۳-۴ همبستگی ساده
۹۸	۲-۱۳-۴ همبستگی چندگانه
۹۹	۳-۱۳-۴ همبستگی جزئی
۹۹	۴-۱۴ ضریب همبستگی جزئی
۱۰۴	۵-۱۵ ضریب همبستگی جزئی مرتبه دوم
۱۰۶	۶-۱۶ آزمون فرض درباره‌ی ضریب همبستگی جزئی جامعه
۱۰۸	خودآزمایی فصل چهارم

۱۱۳	فصل پنجم. رگرسیون
۱۱۳	هدف کلی
۱۱۳	اهداف یادگیری
۱۱۳	۱-۵ رگرسیون خطی دو متغیره
۱۱۴	۲-۵ برآوردکننده‌های کمترین توان‌های دوم (حداقل مجموع مربعات)
۱۱۹	۳-۵ فرم ماتریسی رگرسیون ساده

۱۲۱	۴-۵ آزمون فرض درباره β_1
۱۲۲	۵-۵ تحلیل مانده‌ها
۱۲۶	۶-۵ تجزیه‌ی واریانس رگرسیون ساده
۱۳۰	۷-۵ جدول تجزیه‌ی واریانس
۱۳۲	۸-۵ ضریب تعیین
۱۳۶	۹-۵ ضریب تعیین
۱۳۷	۱۰-۵ تجزیه و تحلیل رگرسیون چندگانه با استفاده از جبر ماتریس‌ها
۱۳۷	۱۱-۵ خط رگرسیون با k متغیر مستقل
۱۳۸	۱۲-۵ فرم ماتریسی رگرسیون با k متغیر مستقل
۱۴۰	۱۳-۵ واریانس برآوردگر b
۱۴۱	۱۴-۵ واریانس برآورد مقدار پیش‌بینی
۱۴۱	۱۵-۵ برآورد واریانس خطا
۱۴۲	۱۶-۵ تحلیل واریانس رگرسیون چند متغیره
۱۴۲	۱۷-۵ مجموع مربعات رگرسیونی
۱۴۳	۱۸-۵ مجموع مربعات خطا
۱۴۳	۱۹-۵ افراز مجموع مربعات کل
۱۴۳	۲۰-۵ آزمون F برای وجود رابطه‌ی خطی
۱۴۷	۲۱-۵ آزمون درباره ضرایب رگرسیون (β_j)
۱۴۹	۲۲-۵ ضریب تعیین چندگانه
۱۵۰	۲۳-۵ ضریب تعدیل چندگانه
۱۵۲	۲۴-۵ متغیرهای مستقل طبقه‌بندی‌شده
۱۵۷	۲۵-۵ محاسبه‌ی ضریب همبستگی جزئی با استفاده از نتایج جدول تجزیه‌ی واریانس رگرسیون
۱۵۹	خودآزمایی فصل پنجم

۱۶۹	فصل ششم. تجزیه‌ی مؤلفه‌های اصلی
۱۶۹	هدف کلی
۱۶۹	اهداف یادگیری
۱۷۰	مقدمه
۱۷۱	۱-۶ حوه‌ی استفاده از مؤلفه‌های اصلی
۱۷۲	۲-۶ ماتریس واریانس-کوواریانس
۱۷۳	۲-۶ ماتریس ضرایب همبستگی
۱۷۳	۴-۶ ماتریس واریانس-کوواریانس نمونه
۱۷۵	۵-۶ ماتریس ضریب همبستگی نمونه‌ای
۱۷۷	۶-۶ مقادیر و بردارهای ویژه ماتریس Σ
۱۷۹	۷-۶ مؤلفه‌ی اصلی i -ام
۱۸۰	۸-۶ واریانس و کوواریانس مؤلفه‌های اصلی
۱۸۳	۹-۶ مقادیر و بردارهای ویژه ماتریس همبستگی ρ
۱۸۶	۱۰-۶ سهم مؤلفه‌ی اصلی i -ام

۱۸۷	۶-۱۱ تعداد مؤلفه‌های اصلی
۱۸۹	۶-۱۲ ضریب همبستگی بین مؤلفه‌ی اصلی و متغیر
۱۹۰	۳-۱۳ مقادیر و بردارهای ویژه‌ی S یا R
۱۹۴	۶-۱۴ تحلیل مؤلفه‌های اصلی
۱۹۹	۶-۱۵ متغیرهای استاندارد شده
۲۰۴	۶-۱۶ نمودار مؤلفه اصلی t - ام در برابر مؤلفه اصلی j - ام
۲۰۷	۶-۱۷ نمودار پراکنش مشاهدات بر اساس مؤلفه‌های اصلی
۲۱۰	خودآزمایی فصل ششم

۲۱۷	فصل هفتم. تحلیل عاملی (تجزیه به عامل‌ها)
۲۱۷	هدف کلی
۲۱۷	اهداف یادگیری
۲۱۷	مقدمه
۲۱۸	۷-۱ الگوی عاملی
۲۲۲	۷-۲ تجزیه Σ
۲۲۵	۷-۳ محاسبه‌ی ماتریس ψ
۲۲۹	۷-۴ تجزیه S
۲۳۴	۷-۵ انتخاب تعداد فاکتورها
۲۳۸	۷-۶ آزمون مناسب بودن داده‌ها برای تحلیل عاملی
۲۴۲	۷-۷ چرخش عاملی
۲۴۴	۷-۸ دوران و ریماکس
۲۵۴	خودآزمایی فصل هفتم

۲۵۹	فصل هشتم. تجزیه تابع ممیزی
۲۵۹	هدف کلی
۲۵۹	اهداف یادگیری
۲۵۹	مقدمه
۲۶۰	۸-۱ ساختار داده در گروه‌ها
۲۶۳	۸-۲ تخصیص با استفاده از فاصله‌های مایلانویس
۲۶۶	۸-۳ تابع تشخیص کانونی
۲۷۱	۸-۴ فرضیات تابع تشخیص کانونی
۲۷۴	۸-۵ آزمون معنی‌داری
۲۷۶	۸-۶ نمودار پراکنش
۲۸۰	خودآزمایی فصل هشتم

۲۸۵	فصل نهم. تجزیه‌ی خوشه‌ای
۲۸۵	هدف کلی

۲۸۵	اهداف یادگیری
۲۸۵	مقدمه
۲۸۶	۱-۹ معیارهای برآورد فاصله میان مشاهدات
۲۸۹	۲-۹ استاندارد کردن داده‌ها
۲۹۲	۳-۹ معیار فاصله برای داده‌های نسبتی
۲۹۳	۴-۹ معیار فاصله برای داده‌های دودویی
۲۹۵	۵-۹ خوشه‌بندی نمونه‌ها
۲۹۵	۶-۹ خوشه‌بندی سلسله مراتبی
۲۹۷	۱-۶-۹ روش خوشه‌بندی پیوند تکی
۳۰۱	۲-۶-۹ روش خوشه‌بندی پیوند کامل
۳۰۵	۳-۶-۹ روش خوشه‌بندی درون گروه‌ها
۳۰۶	۴-۶-۹ روش خوشه‌بندی مرکزی
۳۰۶	۵-۶-۹ روش خوشه‌بندی میانه
۳۰۸	۶-۶-۹ Ward
۳۱۱	۷-۹ انتخاب تعداد خوشه‌ها
۳۱۵	۸-۹ خوشه‌بندی به روش تجزیه
۳۱۶	۹-۹ خوشه‌بندی k - میانگین
۳۱۹	خودآزمایی فصل نهم
۳۲۳	پیوست‌ها
۳۳۱	واژگان توصیفی
۳۵۳	منابع و مآخذ
۳۵۵	فهرست نمایه

"به نام پروردگار بی‌همتایی که کتاب، معجزه آخرین فرستاده اوست"

پیشگفتار

در حال حاضر علم آمار یک رشته دانشگاهی و محض محسوب نمی‌شود، بلکه بسیاری از رشته‌های مهم دانش بشری از جمله علوم زیستی، کشاورزی، اقتصاد، هوا و اقلیم‌شناسی، روانشناسی، علوم تربیتی، علوم پزشکی و بسیاری از علوم دیگر بدون بهره‌گیری از علم آمار، علمی غیرکاربردی خواهند بود. البته باید تاکید نمود آشنایی دقیق و استفاده درست، از روش‌های آماری بخصوص روش‌های پیشرفته آماری، باعث شده تا علوم دیگر در امر تحقیق، برنامه‌ریزی و رفع مشکلات بشر موثر و مفید باشند.

با توجه به نیاز دانشجویان و دانش‌پژوهان رشته‌های مختلف بخصوص علوم زیستی و کشاورزی با کاربرد روش‌های پیشرفته آماری کتاب حاضر در قالب ۹ فصل تدوین شد. در این اثر علمی روش‌های پیشرفته آماری چون مدل‌های خطی و تجزیه و تحلیل چندمتغیره، جبر ماتریسی، آزمون فرض و استنباط آماری، توصیف و نمایش داده‌های چند متغیره، رگرسیون، تجزیه مؤلفه‌های اصلی، تحلیل عاملی (تجزیه به عامل‌ها)، تجزیه تابع ممیزی و تجزیه خوشه‌ای در قالب یک منبع آموزشی و پژوهشی با ویژگی خودآموز ارائه شده‌است. خداوند بزرگ را سپاسگزاریم که این توفیق را به ما عطا فرمود.

لازم است از استاد ارجمند جناب آقای دکتر محمدرضا بی‌همتا که به عنوان ویراستار علمی ما را در ارتقای علمی این کتاب یاری کرده‌اند، سپاسگزاری نمائیم. همچنین از نظرات علمی جناب آقای دکتر محمدرضا رمضانی مقدم در تدوین این اثر بهره گرفته و

مراتب تقدیر خود را اعلام می‌داریم. همچنین از همکاری و نظرات جناب آقای مهندس امین مخدومی سپاسگزاری می‌شود. سرکار خانم اکرم مبارکی در ویراستاری ادبی این اثر نقش‌آفرین بودند، که از زحمات ایشان قدردانی می‌شود. پیگیری امور، صفحه‌آرایی، تنظیم ساختار کتاب و طراحی جلد با همکاری ارزشمند کارشناسان مدیریت دفتر تدوین و تولید کتب دانشگاه پیام نور، سرکار خانم حاجی‌انزهائی، خانم عسگری‌فر و خانم رحیمی انجام شده که از زحمات این همکاران محترم، صمیمانه قدردانی می‌گردد.

مفتخریم به داشتن خانواده‌ای محترم، که در نهایت شکیبایی و درک عمیق از مشغولیت‌های تدوین این اثر، ما را در ارائه آن یاری کرده و به ادامه این کار علمی تشویق نمودند. وظیفه خود می‌دانیم از شکیبایی و حمایت‌های آنها صمیمانه سپاسگزاری نمائیم. امید است با نظرات دانشجویان، استادان و محققان محترم در آینده شاهد ارتقای علمی این اثر باشیم.

پرویز نصیری (دانشیار گروه آمار دانشگاه پیام نور)
محمدعلی ابراهیمی (دانشیار گروه بیوتکنولوژی کشاورزی دانشگاه پیام نور)
محمد جعفرآقایی (دانشیار مؤسسه اصلاح و تهیه نهال و بذر)
پائیز ۱۳۹۴

فصل اول

مدل‌های خطی و تجزیه و تحلیل چندمتغیره

هدف کلی

هدف کلی این فصل، آشنایی دانشجویان با مدل‌های خطی و تجزیه و تحلیل داده‌های چندمتغیره است.

اهداف یادگیری

- برای رسیدن به این هدف کلی، باید به هدف‌های یادگیری زیر دست یابید:
- با روش‌های آماری برای تجزیه و تحلیل داده‌های چند متغیره آشنا شوید.
 - جامعه‌ی آماری را تعریف کنید و فرضیه‌های آماری را ارائه دهید.
 - ضریب همبستگی و مدل‌های خطی را بیان کنید.
 - با روش‌های چند متغیره، از جمله تجزیه و تحلیل مؤلفه‌های اصلی و تابع تشخیص آشنا شوید.

۱-۱ معرفی روش‌های یک متغیره و چند متغیره

علم آمار مسئله‌ی تصمیم‌گیری را بر اساس داده‌های مشاهده شده و در شرایط عدم قطعیت مورد بررسی قرار می‌دهد، ولی برداشت عام از این علم چنین است که آمار با توده‌ی عظیم داده‌ها سرکار دارد و به ارائه‌ی درصدها، متوسط‌ها یا به نمایش داده‌ها در قالب جدول‌ها و نمودارها می‌پردازد. در واقع امروزه، این موارد تنها بخش کوچکی از این علم را تشکیل می‌دهند. سایر جنبه‌های آماری که بسیار مورد استفاده‌ی پژوهشگران قرار می‌گیرند، عبارتند از: نمونه‌گیری، آزمون فرض، تجزیه‌ی واریانس، همبستگی، رگرسیون و روش‌های چند متغیره.

در بررسی‌های آماری نمونه‌گیری از آن جهت اهمیت دارد که به دلیل نامحدود بودن جوامع آماری و عدم امکان دسترسی به همه‌ی افراد جامعه، اغلب ارزیابی‌ها بر روی نمونه‌ای از افراد انجام می‌گیرد که به‌طور تصادفی از میان افراد جامعه‌ی اصلی انتخاب شده‌اند. انتخاب نمونه‌ی مناسب برای رسیدن به نتایج نااریب بسیار با اهمیت است.

۱-۲ احتمالات

تجزیه و تحلیل‌های آماری براساس علم احتمالات انجام می‌شود. احتمالات علمی به نسبت جدید محسوب می‌شود، این علم وقوع حوادث را از حدس و گمان‌های ذهنی به یک روش دقیق ریاضی ارتقا داده است. علم احتمالات برای برآورد احتمال وقوع نتایج مطلوب در هر حادثه ساده‌ای، مانند احتمال آمدن یک جفت شیر در پرتاب همزمان دو سکه، یا احتمال تولید تصادفی تعداد معینی قطعه‌ی ناسالم در یک کارخانه، یا بروز شدت معینی از بیماری زنگ زرد در مزارع گندم، روش‌های علمی و دقیقی را ارائه می‌دهد، اما حتی دقیق‌ترین برآوردها نیز فقط برآورد احتمال است و هیچ قطعیتی در کار نیست. در جهان واقعی ممکن است یک حادثه با احتمال بزرگ اتفاق نیفتد و یا یک واقعه با احتمال وقوع بسیار کوچک بروز کند. احتمال تنها نشان می‌دهد که اگر یک حادثه به دفعات بسیار زیاد تکرار شود، چه نسبتی از آنها به نتیجه مورد نظر پژوهشگر منجر خواهد شد. در واقع تأیید یا رد یک فرضیه‌ی آماری با اثبات یا عدم اثبات یک گزاره‌ی ریاضی متفاوت است. گاهی یک گزاره‌ی ریاضی را اثبات می‌کنند و یا با ارائه‌ی مثال نقض، آن را رد می‌کنند. در هر حال نتیجه‌ای که به دست می‌آید، بدون هیچ شکی برقرار است؛ اما در مقابل، نتیجه‌ی حاصل از آزمون فرضیه‌ی آماری از طریق تحلیل داده‌های تجربی و برآورد احتمال وقوع انجام می‌شود. بنابراین، هیچ نتیجه‌ای حتمی و قطعی نیست.

۱-۳ آزمون فرض

یکی روش‌های آماری که در تجزیه و تحلیل نتایج آزمایش‌های تجربی بسیار کاربرد دارد، آزمون فرض است. این روش برای بررسی صحت احکامی در مورد پارامترهای جامعه، بر اساس نتایج حاصل از بررسی نمونه‌ها انجام می‌شود و اجازه می‌دهد نتایج

به‌دست آمده از آزمایش‌های مختلف با یکدیگر و نتایج مورد انتظار مقایسه شوند. به طور کلی هدف از آزمون فرضیه‌های آماری، رسیدن به این موضوع است که آیا حدس و گمانی که پژوهشگر درباره‌ی پارامتر یا پارامترهایی از جامعه دارد، مورد تأیید قرار می‌گیرند یا خیر. هر حدس یا گمانی یک فرضیه‌ی آماری قابل آزمون نیست. مثلاً این حدس که فردا هوا بارانی است یا محصول سال آینده زیاد خواهد بود، یک فرض آماری نیست. فرضیه‌ای را که مبین وضعیت خاصی درباره‌ی پارامتر یا توزیع یک جامعه‌ی معین باشد، یک فرضیه آماری گویند. این حدس که "میانگین بارندگی در شهر الف از میانگین بارندگی در شهر ب بیشتر است" یک فرضیه‌ی آماری است که بیانگر مقایسه‌ی پارامتر میانگین میان دو جامعه است و می‌تواند مورد آزمون قرار گیرد. آزمون فرض قاعده‌ای است که به وسیله‌ی آن مشخص می‌شود آیا نمونه‌ی مورد مطالعه از نظر منطقی با فرض مورد نظر مطابقت دارد یا ندارد. آزمون فرض در دامنه‌ی وسیعی از تحقیقات آماری برای مقایسه آماره‌های به‌دست آمده از نمونه با آماره‌های جامعه یا آماره‌های به‌دست آمده از نمونه‌های مختلف با یکدیگر یا با مقادیر مورد انتظار به‌کار برده می‌شود. بخش سوم این کتاب به توضیح این نوع از آزمون‌ها و نحوه‌ی نتیجه‌گیری از آنها اختصاص دارد.

۴-۱ تجزیه‌ی واریانس

در برخی از آزمایش‌ها هدف تحقیق، بررسی اثر سطوح مختلفی از تیمارهای متعدد بر شدت یا کیفیت یک صفت معین است. در علم آمار برای تحلیل نتایج این نوع از آزمایش‌ها و بررسی تفاوت اثر تیمارها از روش تجزیه‌ی واریانس استفاده می‌شود. در تجزیه‌ی واریانس، تغییرات کل مشاهده شده در مشاهدات آزمایش، به تغییرات ناشی از اثر تیمارها، که به آن میانگین مربعات تیمار گفته می‌شود، و تغییرات تصادفی ناشی از عوامل محیطی خارج از کنترل پژوهشگر تقسیم می‌شود. به این جزء اخیر در اصطلاح میانگین مربعات خطا گفته می‌شود. بنابراین،

$$MST = MSt + MSe \quad (۱)$$

اثر خطا بر روی همه‌ی تیمارها و تکرارهای آزمایش یکسان فرض می‌شود، بنابراین،

تفاوت مشاهده شده در میان تیمارهای مختلف از اثر تیمار و اثر محیط ناشی می‌شود.

$$MST = MSt + MSe \quad MSt = r\sigma_i + \sigma_e \quad MSe = \sigma_e \quad (2)$$

برای آزمون تأثیر تیمار در نتیجه‌ی آزمایش، میانگین مربعات تیمار بر میانگین مربعات خطا تقسیم می‌شود.

$$F = \frac{MSt}{MSe} = \frac{r\sigma_i + \sigma_e}{\sigma_e} \quad (3)$$

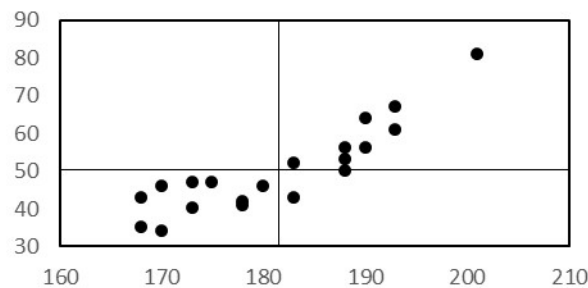
اگر حاصل این نسبت نزدیک به یک باشد به این معنی است که تفاوت‌های مشاهده شده میان تیمارها تنها از عوامل محیطی تصادفی ناشی شده‌است، اما اگر حاصل نسب خیلی بزرگ‌تر از یک باشد، تیمارهای مختلف دارای اثر متفاوتی بر صفت مورد بررسی بوده‌اند. معیار بزرگ‌بودن حاصل این نسبت متغیر تصادفی F است که از توزیع فیشر تبعیت می‌کند.

محدودیت‌های عملی در اجرای آزمایش‌ها مختلف علوم زیستی، برای بررسی تأثیر تیمارها، متعدد و متنوع هستند. این محدودیت‌ها یا نیازهای اجرایی موجب شده‌است تا اشکال متنوعی از روش‌ها برای تعریف عوامل دخیل در آزمایش و بررسی اثر آنها، بر اساس روش تجزیه واریانس تعریف شود. مجموعه‌ی این روش‌ها و انواع مختلف آن به‌ویژه در علوم کشاورزی آن‌قدر وسیع هستند که یک واحد درسی خاص تحت عنوان طرح آزمایش‌ها برای آن در نظر گرفته شده‌است.

۵-۱ همبستگی و رگرسیون

تجزیه و تحلیل آماری فقط شامل بررسی اثر تیمارها و عوامل مختلف بر صفات مورد نظر پژوهشگر نیست. در بسیاری از موارد روابط میان صفات مختلف اهمیت دارند. در این نوع از آزمایش‌ها پژوهشگر تعداد زیادی صفت را در یک جامعه‌ی زیستی اندازه‌گیری می‌کند و اکنون می‌خواهد بداند تغییرات هرکدام از صفات چه تأثیری بر سایر صفات دارد. ساده‌ترین وضعیت بررسی روابط دو به دو میان متغیرها است که به آن جامعه‌ی دو متغیره توأم گفته می‌شود. برای مثال تعداد بوته (x) و مجموع سطح برگ (y) بوته‌های گندم در واحد سطح مزرعه اندازه‌گیری شده‌است. سؤال پژوهشگر

آن است که آیا میان تعداد بوته و سطح برگ بوته‌ها رابطه‌ای وجود دارد؟ اگر رابطه‌ای میان این دو متغیر وجود دارد، کیفیت این رابطه و شدت آن چگونه است؟ بررسی نمودار پراکنش یکی از معمول‌ترین روش‌های بررسی روابط میان دو متغیر توأم است. برای این منظور اندازه‌های یکی متغیرها را روی محور افقی و دیگری را روی محور عمودی مشخص می‌کنند و نمونه‌ها را به صورت نقاطی بر روی صفحه نمایش می‌دهند.



شکل ۱-۱. نحوه‌ی نمایش نمودار پراکنش نمونه‌ها براساس اندازه‌های دو متغیر توأم

بررسی نحوه‌ی قرار گرفتن نقاط نمونه بر روی نمودار پراکنش می‌تواند ایده‌ای از وجود رابطه‌ی میان دو متغیر را نشان بدهد. اما بررسی دقیق‌تر رابطه، نیازمند محاسبه همبستگی میان متغیرها است که بر نسبت کوواریانس میان دو متغیر بر واریانس درون هر دو متغیر استوار است و مقداری بین ۱- تا ۱ را به خود می‌گیرد.

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}} \quad (۴)$$

مقادیر نزدیک به صفر بیانگر عدم وجود رابطه میان دو متغیر است در حالی که مقادیر نزدیک به ۱ بیانگر رابطه همسو و مقادیر نزدیک به ۱- بیانگر رابطه غیرهمسو میان دو متغیر است.

همبستگی یک همسویی دوطرفه میان اندازه‌های دو متغیر است و دلیلی بر رابطه‌ی علت و معلولی میان متغیرها نیست، و حتی لازم نیست که یکی از متغیرها به دیگری وابسته باشد، اما اگر از قبل مشخص باشد که تغییرات یکی از متغیرها (مستقل)

بر روی اندازه‌های متغیر دوم (وابسته) تأثیر می‌گذارد، شدت این تأثیرپذیری را می‌توان با استفاده از روش تجزیه‌ی رگرسیون برآورد کرد و حتی با تشکیل یک مدل ریاضی اندازه‌های متغیر وابسته را از روی اندازه‌های متغیر مستقل برآورد کرد.

اگرچه همه‌ی مباحث احتمالات، نمونه‌گیری، آزمون‌های فرض، تجزیه‌ی واریانس، همبستگی و رگرسیون از مهم‌ترین مباحث مربوط به علم آمار هستند، اما از آنجا که این مباحث به‌طور معمول در کتاب‌های آمار و احتمالات مقدماتی و طرح آزمایش‌ها به‌طور مفصل مطرح می‌شوند، در کتاب حاضر از ورود مفصل به این مباحث خوداری و تنها در حدی که برای توضیح مباحث خاص این کتاب لازم است مورد اشاره قرار خواهند گرفت.

۱-۶ روش‌های چند متغیره

بحث اصلی کتاب حاضر روش‌های چندمتغیره آماری هستند، که هدف کلی آن تجزیه و تحلیل مشاهدات برای چندین متغیر به‌صورت همزمان است. روش‌های چند متغیره از نظر تاریخی سابقه‌ی طولانی در علوم رفتاری و علوم زیستی دارند. این روش‌ها از گذشته شناخته شده بودند و روش‌هایی بر حل معادلات آنها پیشنهاد شده بود. اما حل مسائل واقعی با استفاده از این روش‌ها، نیازمند انجام حجم وسیعی از محاسبات بود که بسیار وقت‌گیر بوده و با بروز اشتباهات مکرر در محاسبات مواجه بود. با توسعه‌ی رایانه‌ها و فراهم‌آمدن نرم‌افزارهای محاسباتی قدرتمند، امکان انجام پیچیده‌ترین محاسبات چند متغیره در زمان‌های بسیار کوتاه و عاری از اشتباه و خطا فراهم آمد، بنابراین، توجه به استفاده از روش‌های چند متغیره آماری برای تجزیه و تحلیل روابط میان متغیرهای مورد بررسی و استفاده از آنها برای مطالعه‌ی روابط میان افراد جامعه آماری مورد بررسی بیش از پیش مورد توجه قرار گرفت. امروزه از روش‌های چند متغیره به‌طور وسیعی در تحقیقات رشته‌های شیمی، فیزیک، زمین‌شناسی، مهندسی، حقوق، تجارت، زیست‌شناسی، روان‌شناسی، کشاورزی، علوم اجتماعی، مردم‌شناسی، باستان‌شناسی و غیره مورد استفاده قرار می‌گیرد.

اولین قدم در روش‌های چندمتغیره، سازماندهی داده‌های اندازه‌گیری یا مشاهده‌شده‌ی متغیرها است. فرض کنید در یک پژوهش علمی P متغیر

۷ مدل‌های خطی و تجزیه و تحلیل چندمتغیره

$X_1, X_2, \dots, X_p$ مدنظر باشد، نمایش متغیر i -ام از j -ام را با علامت X_{ij} نشان می‌دهند، برای n مشاهده از هر متغیر را می‌توان به صورت زیر سازماندهی کرد.

مشابه متغیر	۱	۲	...	i	...	n
X_1	X_{11}	X_{12}	...	X_{1i}	...	X_{1n}
X_2	X_{21}	X_{22}	...	X_{2i}	...	X_{2n}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
X_j	X_{j1}	X_{j2}	...	X_{ji}	...	X_{jn}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
X_p	X_{p1}	X_{p2}	...	X_{pi}	...	X_{pn}

فرض کنید در بررسی وضعیت کارکنان سازمانی متغیرهای سن (X_1) ، تحصیلات (X_2) ، جنسیت (X_3) و در آمد (X_4) مدنظر باشد. برای ۵ نفر از آنها که به تصادف انتخاب شده‌اند، نتایج به صورت زیر رده بندی می‌شود.

مشابه متغیر	۱	۲	۳	۴	۵
X_1	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_2	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}
X_3	X_{31}	X_{32}	X_{33}	X_{34}	X_{35}
X_4	X_{41}	X_{42}	X_{43}	X_{44}	X_{45}

که در آن X_{ij} ، معرف مشاهده انجام شده برای متغیر i -ام از فرد j -ام است. در تجزیه‌ی چندمتغیره مشاهدات برای چندین متغیر مرتبط به طور همزمان در نظر گرفته می‌شوند که در شروع آزمایش همه آنها برای پژوهشگر مهم هستند. این متغیرها بسته به اهداف پژوهشگر و نوع جامعه مورد بررسی ممکن است بسیار متنوع بوده و دارای مقیاس‌های متفاوتی باشند. برای مثال جدول زیر را در نظر بگیرید.

جدول ۱-۱. مثالی از چند جمعیت و صفات مورد بررسی در آنها

ردیف	جامعه مورد بررسی	متغیرهای مورد بررسی
۱	دانشجویان	نمرات چند دوره مختلف در یک درس
۲	دانشجویان	نمره در ریاضی، تاریخ، موسیقی، هنر و غیره
۳	جمعیت	قد، وزن، درصد چربی، ضربان قلب
۴	جمع‌ه‌های باستانی	طول، عرض، گنجایش حجمه
۵	مقاضیان وام بانکی	درآمد، اشتغال، سابقه سکونت، پس‌انداز، بدهی وام
۶	اسکلت پرنندگان	طول استخوان‌های قسمت‌های مختلف بدن

در جدول شماره ۱-۱ متغیرهای ردیف اول و دوم از مقیاس مشابهی برخوردارند، درحالی‌که صفات اندازه‌گیری‌شده در ردیف سوم دارای مقیاس اندازه‌گیری متفاوت هستند. همه‌ی روش‌های آماری مورد استفاده برای تحلیل داده‌های یک متغیره و یا دو متغیره را می‌توان برای تحلیل داده‌های چند متغیره نیز به‌کار برد. علاوه بر آن روش‌های تحلیل چند متغیره وجود دارد که می‌توانند با در نظر گرفتن همزمان داده‌ها روابط بیشتری در میان داده‌ها آشکار نمایند. در ادامه‌ی این کتاب برخی از این روش‌ها معرفی می‌شوند، اما نکته‌ای که وجود دارد وقتی تعداد متغیرهای مورد بررسی در آزمایش افزایش می‌یابند، حجم داده‌ها و محاسبات مورد نیاز برای تحلیل آنها به‌صورت نمایی افزایش می‌یابند و ممکن است روش تحلیل بسیار پیچیده شود. بنابراین، برای معرفی یک روش ثابت و قابل درک برای انجام محاسبات چند متغیره از روش معینی در ریاضیات تحت عنوان جبر ماتریسی استفاده می‌شود. در جبر ماتریسی به‌جای انجام محاسبات جداگانه بر روی هرکدام از متغیرها و اندازه‌ها، همه‌ی داده‌ها به‌صورت یک کل در محاسبات به‌کار برده می‌شوند. بنابراین، اگرچه انجام محاسبات حتی با استفاده از جبر ماتریسی نیز همچنان وقت‌گیر و پیچیده هستند، اما درک روش‌ها را ساده‌تر می‌کند. از آنجا که برخی از این روش‌ها با روش‌ها معمول جبری متفاوت هستند و ممکن است خوانندگان با آنها آشنا نباشند، مفاهیم جبر ماتریسی و برخی از روش‌های محاسبات ماتریسی به‌طور خلاصه در فصل دوم توضیح داده می‌شود.

۷-۱ توصیف جوامع چند متغیره

همان‌طور که در ابتدای این فصل اشاره شد، روش‌های توصیفی در آمار اولین مفاهیمی هستند که به ذهن متبادر می‌شوند. این روش‌ها شامل محاسبه و نشان دادن فراوانی‌ها، برآورد پارامترهای تمایل به مرکز، پارامترهای پراکندگی و نمایش گرافیکی داده‌ها است. برای داده‌های چند متغیره هم نیاز به چنین توصیفاتی، مانند معرفی وضعیت جامعه در یک یا چند پارامتر ساده و حتی نمایش گرافیکی خصوصیات جامعه وجود دارد. اگرچه تعداد زیاد متغیرها و حجم زیاد داده‌ها، نمایش و توصیف ساده آنها را غیرممکن می‌سازد، اما روش‌های خاصی برای توصیف این نوع از داده‌ها نیز وجود دارد که در فصل چهارم مورد بحث و بررسی قرار خواهند گرفت.

۸-۱ همبستگی جزئی

ادامه فصل چهارم به توضیح انواع مختلف همبستگی اختصاص داده شده است. همبستگی ساده که در بخش اول این فصل به آن اشاره شد و در دروس آمار مقدماتی مورد بحث قرار می‌گیرند، تغییرات همزمان میان دو متغیر را بدون توجه به تغییرات سایر متغیرها مورد نظر قرار می‌دهد و به همین دلیل وجود همبستگی ساده میان دو متغیر هیچگاه به معنی وجود رابطه میان آنها تعبیر نمی‌شود. مشاهده شده است که دو متغیر، که خود تحت تأثیر متغیر دیگری بوده‌اند، با همدیگر همبستگی نشان می‌دهند، بدون آنکه هیچ رابطه‌ی علّی میان آنها وجود داشته باشد. به‌طور مثال مشاهده شده است که گسترش دامنه‌ی یک بیماری قارچی و عملکرد محصول دیم با همدیگر حتی همبستگی مثبت نشان می‌دهند، درحالی‌که به‌طور منطقی اگر همبستگی هم وجود داشته باشد، باید یک همبستگی منفی میان آنها دیده شود. با بررسی میانگین بارندگی در فصل رشد، ممکن است متوجه شویم که افزایش رطوبت، همزمان که موجب بهبود شرایط رشد و افزایش عملکرد می‌شود، به گسترش بیماری قارچی نیز کمک می‌کند. اگر شدت گسترش بیماری آنقدر زیاد نباشد که به کاهش چشمگیر در عملکرد منجر شود، افزایش میزان بارندگی، که با گسترش همزمان بیماری و افزایش عملکرد همراه است، موجب بروز همبستگی مثبت میان دو متغیر اخیر خواهد شد. همبستگی جزئی، روشی است برای بررسی همبستگی میان متغیرها به شرط ثابت ماندن سایر متغیرها. بنابراین،