



روش‌های تجزیه آماری چند متغیره

(کارشناسی ارشد علوم کشاورزی)

دکتر عزت‌اله فرشادفر دکتر محسن فرشادفر

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 دستاورد 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 عظیم‌الشان خداوند، فرود آمده، نام «قلم» به تجلیل یاد شده‌است: «اقْرَأْ وَ رَبُّكَ 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. کلیات آمار چند متغیره و جبر ماتریس.....	۱
هدف کلی	۱
هدف‌های یادگیری	۱
مقدمه	۱
۱-۱ هدف‌های کاربردی روش‌های آماری چند متغیره	۳
۲-۱ سازمان‌دهی داده‌ها	۴
۳-۱ مقدمه‌ای بر جبر ماتریس	۶
۱-۳-۱ تعریف ماتریس	۶
۲-۳-۱ عملیات ماتریس	۸
۳-۳-۱ دترمینان	۱۳
۴-۳-۱ وابستگی خطی	۱۴
۵-۳-۱ عکس ماتریس	۱۵
۶-۳-۱ خصوصیات دترمینان	۱۹
۷-۳-۱ مقادیر ویژه و بردارهای ویژه	۲۴
۸-۳-۱ نحوه پیدا کردن مقدار ویژه و بردار ویژه	۲۴
خودآزمایی چهارگزینه‌ای فصل اول	۳۲
فصل دوم. استنباط آماری چندمتغیره.....	۳۵
هدف کلی	۳۵
هدف‌های یادگیری	۳۵

۳۵	 مقدمه
۳۶	 ۱-۲ توزیع نرمال چند متغیره
۳۹	 ۲-۲ ارزیابی احتمالات
۴۰	 ۳-۲ برآورد نمونه
۴۳	 ۴-۲ آزمون‌های چند متغیره در مقابل تک متغیره
۴۵	 ۱-۴-۲ آزمون μ با Σ معلوم
۴۸	 ۲-۴-۲ آزمون μ برای وقتی که Σ نامعلوم است
۵۰	 ۵-۲ خصوصیات کلیدی T^2 هتلینگ
۵۵	 ۶-۲ مقایسه میانگین‌های دو بردار
۵۵	 ۱-۶-۲ آزمون مقایسه دو نمونه با آزمون t در آمار تک متغیره
۵۶	 ۲-۶-۲ آزمون T^2 دو نمونه‌ای چند متغیره
۵۷	 ۷-۲ خصوصیات کلیدی آزمون T^2
۶۰	 ۸-۲ آزمون مشاهدات جفت شده
۶۰	 ۱-۸-۲ حالت یک متغیره
۶۲	 ۲-۸-۲ حالت چند متغیره
۶۵	 خودآزمایی تشریحی فصل دوم
۶۷	 فصل سوم. تجزیه واریانس چند متغیره
۶۷	 هدف کلی
۶۷	 هدف‌های یادگیری
۶۷	 مقدمه
۶۹	 ۱-۳ مفروضات MANOVA
۷۰	 ۱-۱-۳ آزمون نسبت درست نمایی ویلک
۷۴	 ۲-۳ طرح بلوک‌های کامل تصادفی
۷۶	 ۳-۳ طرح مربع لاتین
۷۷	 خودآزمایی چهارگزینه‌ای فصل سوم
۷۹	 فصل چهارم. تجزیه مؤلفه‌های اصلی
۷۹	 هدف کلی
۷۹	 هدف‌های یادگیری
۷۹	 مقدمه
۸۲	 ۲-۴ استخراج مؤلفه‌های اصلی
۸۵	 ۳-۴ روش محاسبه مؤلفه‌های اصلی
۸۹	 ۴-۴ سهم هر صفت در مؤلفه اصلی
۹۱	 ۱-۴-۴ اسکری گراف
۹۲	 ۵-۴ تعداد مؤلفه‌های اصلی

۹۳	۶-۴ نمودار دوبعدی
۹۷	خودآزمایی تشریحی فصل چهارم
۹۷	خودآزمایی چهارگزینه‌ای فصل چهارم
۱۰۱	فصل پنجم. تحلیل عاملی
۱۰۱	هدف کلی
۱۰۱	هدف‌های یادگیری
۱۰۱	مقدمه
۱۰۳	۱-۵ خلاصه هدف‌های اصلی تجزیه عاملی
۱۰۵	۲-۵ مدل آماری تجزیه عاملی
۱۰۷	۱-۲-۵ مفروضات مدل ریاضی تحلیل عاملی
۱۱۱	۳-۵ تعیین تعداد عامل
۱۱۳	۴-۵ برآورد ماتریس Ψ, L
۱۱۶	۱-۴-۵ روش مؤلفه‌های اصلی (PC)
۱۲۱	۵-۵ تعاریف در تجزیه به عامل‌ها
۱۲۱	۱-۵-۵ بار عامل‌ها
۱۲۱	۲-۵-۵ کسینوس عامل‌ها
۱۲۲	۳-۵-۵ وزن عامل‌ها
۱۲۲	۴-۵-۵ نمره عامل‌ها
۱۲۲	۵-۵-۵ تفاوت عامل‌ها و نمره عامل‌ها
۱۲۳	۶-۵ دوران عامل‌ها
۱۲۳	۱-۶-۵ توضیح هندسی هدف دوران
۱۲۵	۲-۶-۵ روش‌های دوران
۱۲۹	۷-۵ تعیین تعداد عامل
۱۳۱	۸-۵ فرق بین مؤلفه‌ها و عامل‌ها
۱۳۴	خودآزمایی چهارگزینه‌ای فصل پنجم
۱۳۹	فصل ششم. تابع تشخیص
۱۳۹	هدف کلی
۱۳۹	هدف‌های یادگیری
۱۳۹	مقدمه
۱۴۰	۱-۶ تجزیه تشخیص دو گروه
۱۴۲	۲-۶ تحلیل رگرسیون چندگانه
۱۴۵	۳-۶ ضرایب ساختمانی
۱۵۹	۴-۶ توان ممیزی (تشخیصی)
۱۶۰	۵-۶ آزمون معنی

۱۶۱	۶-۶ توابع تشخیص
۱۶۳	۷-۶ مراکز ثقل گروهی
۱۶۵	۸-۶ ضریب ساختمانی
۱۶۸	خودآزمایی چهارگزینه‌ای فصل ششم

۱۷۱	فصل هفتم. تحلیل همبستگی متعارف
۱۷۱	هدف کلی
۱۷۱	هدف‌های یادگیری
۱۷۱	مقدمه
۱۷۲	۱-۷ مفهوم و کاربرد تحلیل همبستگی متعارف
۱۷۳	۲-۷ ماتریس داده‌های همبستگی متعارف
۱۷۵	۳-۷ تحلیل متعارف در متغیرهای پیوسته
۱۷۵	۴-۷ همبستگی‌های متعارف
۱۷۷	۵-۷ وزن‌های متعارف
۱۸۲	۶-۷ ضرایب ساختمانی
۱۹۴	خودآزمایی چهارگزینه‌ای فصل هفتم

۱۹۷	فصل هشتم. تجزیه خوشه‌ای
۱۹۷	هدف کلی
۱۹۷	هدف‌های یادگیری
۱۹۷	مقدمه
۱۹۸	۱-۸ تعریف تجزیه خوشه‌ای
۱۹۹	۲-۸ روش تجزیه خوشه‌ای
۱۹۹	۳-۸ هدف از تجزیه خوشه‌ای
۲۰۱	۴-۸ انواع تجزیه خوشه‌ای
۲۰۱	۱-۴-۸ تکنیک‌های مراتبی (شجره‌ای)
۲۰۱	۲-۴-۸ تکنیک‌های تجزیه‌ای
۲۰۷	۵-۸ محاسبه ضریب شباهت در صفات کمی
۲۰۷	۱-۵-۸ ضریب متوسط فاصله اقلیدسی
۲۰۹	۲-۵-۸ ضریب تفاوت شکل
۲۱۰	۳-۵-۸ ضریب کسینوس
۲۱۱	۴-۵-۸ ضریب همبستگی
۲۱۱	۵-۵-۸ ضریب متریک کانبرا
۲۱۲	۶-۵-۸ ضریب بری-کرتیس
۲۱۴	۷-۵-۸ فاصله مینسکاووسکی
۲۱۴	۵-۸ روش‌های دسته‌بندی

۲۱۵	 ۱-۵-۸ تکنیک‌های مراتبی
۲۱۵	 ۲-۵-۸ تکنیک‌های اپتی‌مم‌کردن
۲۱۵	 ۳-۵-۸ تکنیک‌های دسته‌ای
۲۱۵	 ۴-۵-۸ تکنیک‌های تراکمی
۲۱۵	 ۵-۵-۸ روش‌های گرافیکی
۲۶۵	 ۸-۸ معایب تجزیه خوشه‌ای
۲۶۵	 ۹-۸ مثال کاربردی از گروه‌بندی تعداد ۱۵ ژنوتیپ گیاه رازیانه براساس روش تجزیه کلاستر با نرم‌افزار SPSS22
۲۸۸	 خودآزمایی چهارگزینه‌ای فصل هشتم
۲۹۳	 پیوست
۳۰۷	 پاسخنامه سوالات چهارگزینه‌ای
۳۰۹	 منابع

پیشگفتار

بنام آنکه جان را فکرت آموخت چراغ دل به نور جان برافروخت

«قال رسوا الله (ص): لَيْسَ الْعِلْمُ بِكُنْهٍ التَّعْلِيمِ وَ إِنَّمَا هُوَ نُورٌ يَقْذِفُهُ اللهُ فِي قَلْبِ مَنْ يَشَاءُ»

امروز بر کسی پوشیده نیست که برای دفاع از ارزش‌های انقلاب اسلامی، گسترش فرهنگ اسلام ناب محمدی در پرتو اندیشه‌های امام راحل (ره) و رهبر معظم انقلاب اسلامی حضرت آیت‌الله خامنه‌ای (مدظله) و زمینه‌سازی برای حاکمیت مطلق قرآن و برقراری حکومت جهانی اسلام به رهبری مهدی آل‌محمد (عج)، تأمین استقلال علمی جامعه از اهمیت ویژه‌ای برخوردار است. به همین دلیل بر عهده کارشناسان، متخصصان و اساتید دانشگاه‌ها است که برای ادای دین خود به نظام مقدس جمهوری اسلامی ایران که یادگار گران‌قدر بزرگ پرچمدار اسلام ناب محمدی حضرت امام‌خمینی (قدس ره) و شهیدان انقلاب اسلامی، در سنگر علم و دانشگاه به پاسداری از دستاوردهای انقلاب اسلامی بپردازند و علوم و فنون روز را با استفاده از منبع معتبر به‌رشته تحریر درآورند تا مورد استفاده دانش‌پژوهان، دانشجویان، محققان و اساتید فن قرار گیرد.

با پیشرفت علم و فناوری در ایران اسلامی، پژوهش‌های علمی روزبه‌روز ارزش بیشتری پیدا کرده است. از آنجائی که استفاده از روش‌های آماری اجازه ارزیابی صحیح‌تر نتایج مطالعات و آزمایش‌های انجام‌شده را می‌دهد، اهمیت دانستن علم آمار روزافزون می‌شود. کتاب‌های آماری متعددی به زبان فارسی چه به‌صورت تألیف و چه به‌صورت ترجمه نوشته‌شده است که هر یک به‌نوبه خود ارزشمند است. کتاب *اصول روش‌های آماری* چند متغیره که هم‌اکنون در پیش روی دارید از جمله کتاب‌هایی است

که براساس قوانین و مقررات آموزشی دانشگاه پیام نور جهت دانشجویان کارشناسی ارشد رشته‌های کشاورزی به رشته تحریر درآمده‌است. اصول و روش‌های بیان‌شده در این کتاب برای رشته‌های مختلف علوم پایه، علوم انسانی، علوم مهندسی، علوم پزشکی و علوم مدیریت کاربرد دارد. با توجه به اینکه کتاب حاضر عاری از نقص نیست، از کلیه کسانی که با پیشنهادها سازنده خود ما را در رفع این نواقص یاری نمایند، سپاس‌گزاری می‌نماییم. در پایان از کلیه کسانی که در تهیه، تدوین، حروف‌چینی، چاپ و انتشارات دانشگاه پیام نور که مؤلفین را یاری نمودند صمیمانه تشکر و قدردانی می‌شود. از خداوند متعال مسئلت می‌نماییم که به ما توفیق خدمت‌گزاری به اسلام و مسلمین را در پرتو رهنمودهای امام راحل (ره) و مقام معظم رهبری حضرت آیت‌الله خامنه‌ای عنایت فرماید.

والعلم عندالله

عزت‌اله فرشادفر: استاد دانشگاه رازی کرمانشاه

محسن فرشادفر: دانشیار دانشگاه پیام نور

فصل اول

کلیات آمار چند متغیره و جبر ماتریس

هدف کلی

در این فصل کلیات و هدف‌های آمار چند متغیره و اصول موردنیاز جهت تجزیه و تحلیل آمار چند متغیره شرح داده خواهد شد.

هدف‌های یادگیری

هدف‌های یادگیری این فصل به شرح زیر است:

۱. تعریف آمار چند متغیره بیان می‌گردد.
۲. کاربرد آمار چند متغیره توضیح داده خواهد شد.
۳. جبر ماتریس در حد نیاز آشنا خواهید شد.

مقدمه

روش‌های آماری چند متغیره آن دسته از روش‌های آماری هستند که به‌طور هم‌زمان داده‌ها را روی چندین متغیر تجزیه می‌کنند. در تجزیه چند متغیره روابط هم‌زمان متغیرها باهم موردبحث قرار می‌گیرد. در این تکنیک‌ها پیرامون میانگین یا واریانس یک متغیر یا رابطه دو متغیر بحث نمی‌شود، بلکه کوواریانس و همبستگی‌های بین سه یا چند متغیر موردبحث قرار می‌گیرد. به عبارت دیگر روش‌های آماری چند متغیره متفاوت از یک متغیره یا دو متغیره هستند. در روش‌های آماری مقدماتی تغییرات موجود در یک متغیر تصادفی موردبحث قرار می‌گرفت و این امر حتی در رگرسیون چندگانه نیز صادق بود، زیرا در این تکنیک نیز تلاش بر آن است که دلیل تغییرات موجود در یک

متغیر وابسته بیان شود؛ اما در روش‌های آماری چند متغیره چندین متغیر تصادفی وابسته به‌طور هم‌زمان مورد ارزیابی قرار می‌گیرد، به‌طوری‌که هر یک از این متغیرها در آغاز شروع تجزیه از اهمیت یکسانی برخوردار هستند.

فرض کنید یک شرکت می‌خواهد دو محصول مسواک و خمیردندان را به بازار عرضه کند و مدیر کارخانه در نظر دارد که وضعیت مصرف‌کننده‌ها را از نظر این دو محصول ارزیابی کند. نوع تجزیه بستگی به این دارد که مدیر کارخانه چه چیزی را در نظر دارد. ممکن است مردم هر دو محصول را باهم موردتوجه قرار دهند. اگر دو محصول به هم وابسته باشند، در آن صورت دریافت دو محصول دارای همبستگی خواهد بود. واردشدن چنین تجزیه‌ای در محاسبات، محاسبات را مشکل‌تر خواهد کرد. در تجزیه رگرسیون و کورولاسیون نیز ممکن است چندین متغیر موردبحث قرار گیرند، اما در اینجا معمولاً فرض بر آن است که متغیرهای مستقل تصادفی نبوده بلکه ثابت هستند.

در تجزیه چند متغیره داده‌ها را بر روی چندین متغیر مثلاً مجموعه‌ای از X ها از قبیل X_1, X_2, X_3 و X_4 غیره به‌عنوان یک عنصر واحد در تجزیه موردتوجه قرار می‌گیرند. مثلاً در تجزیه وضعیت مصرف‌کننده‌ها، مصرف‌کننده‌ها از نظر دو محصول نمره می‌گیرند، یعنی برای هر مصرف‌کننده یک جفت پاسخ وجود دارد به نام‌های X_1 و X_2 بنابراین دو نمره دارد یکی X_1 که وضعیت مصرف‌کننده از نظر مسواک است و دیگری X_2 که وضعیت مصرف‌کننده از نظر خمیردندان است.

اگر k متغیر در تجزیه داشته باشیم در آن صورت k عدد $(X_1, X_2, \dots, X_k)$ یک عنصر خواهد بود. چنین ترتیبی از داده‌ها را یک بردار گویند؛ بنابراین عنصر اصلی تجزیه در اینجا بردار است. توزیع نرمال دارای اهمیت زیادی در تجزیه آماری است، در تجزیه چند متغیره نیز توزیع نرمال از اهمیت ویژه‌ای برخوردار است. توزیع نرمال در تجزیه چند متغیره عبارت از توزیع متغیرهای تصادفی در یک بردار به‌صورت $X=(X_1, X_2, X_k)$ است. این توزیع را توزیع نرمال چند متغیره گویند. اگر $k=2$ باشد آن را حالت دومتغیره گویند و به این ترتیب با توزیع نرمال دوبعدی مواجه هستیم.

در اینجا به‌جای منحنی زنگوله‌ای، منحنی سه‌بعدی پشته‌ای مانند یا تپه مانند وجود دارد. اگر k بزرگ‌تر از ۲ باشد در آن صورت تابع احتمالی دارای سطحی است که بیشتر از سه بعد دارد و نمی‌توان آن را رسم کرد. توزیع نرمال دومتغیره همراه با ابعاد بیشتر مبنای روش آماری چند متغیره است که از نتایج کارهای فرانسیس

گالتون^۱ در نیمه دوم سده نوزدهم است. گالتون نتایج حاصل از دو نسل نخود را مطالعه کرد و اندازه همبستگی بین آنها را در دو نسل به دست آورد. سپس کارهای بعدی کارل پیرسن^۲ و دیگران انواع متفاوت کورولاسیون را در بین متغیرها ارزیابی کردند. این تئوری‌ها مبنای روش‌های آماری چند متغیره قرار گرفت (ارقامی و همکاران ۱۳۷۰، Richard & Haris 1975).

۱-۱ هدف‌های کاربردی روش‌های آماری چند متغیره

- **کاهش داده‌ها و ساده کردن ساختار مورد مطالعه:** یعنی پدیده مورد مطالعه تا آنجا که ممکن است ساده می‌شود بدون آنکه اطلاعات ارزشمندی از دست برود. با این کار امید آن است که تفسیر نتایج آسان‌تر صورت گیرد.
- **گروه‌بندی:** براساس خصوصیات اندازه‌گیری شده گروه‌هایی از متغیرها یا مواد به وجود می‌آید، بنابراین لازم است که از قوانینی استفاده شود که متغیرها را به گروه‌های کاملاً مشخصی تبدیل کند.
- **رابطه بین متغیرها:** معمولاً در تحقیقات علمی ماهیت رابطه بین متغیرها مورد علاقه پژوهشگر است، بدین صورت که آیا متغیرها مستقل از یکدیگرند؟ آیا یک یا چند متغیر به متغیرهای دیگر وابستگی دارند؟ اگر چنین است چگونه؟
- **پیش‌بینی:** هدف از پی بردن به روابط بین متغیرها آن است که با استفاده از اطلاعات به دست آمده از یک گروه به ارزش متغیرهای دیگر پی ببریم.
- **ساختن فرضیه و آزمون:** فرضیه‌های آماری به خصوص با عنوان پارامترهای چند متغیره جمعیت‌ها فرموله شده و سپس مورد آزمون قرار می‌گیرند (فرشادفر، ۱۳۸۴).
- **کاربرد:** اصولاً تکنیک‌های آماری جزء تفکیک‌ناپذیر پژوهش‌های علمی هستند و از این رو، کاربرد آنها در علوم مختلف بسیار وسیع است. از روش‌های آماری چند متغیره نیز برای حل مشکلات مربوط به علوم پایه، علوم فنی و مهندسی، علوم کشاورزی، علوم پزشکی و علوم انسانی استفاده می‌شود.

1. Francis Galton

2. Karl Pearson

مثال. در اصلاح نباتات لازم است که پس از هر نسل، گیاهانی انتخاب کرد که به‌عنوان والدین نسل آینده مورد استفاده قرار گیرند. این انتخاب باید به‌گونه‌ای صورت گیرد که نسل‌های بعد از نظر پاره‌ای خصوصیات بهتر از نسل قبل باشد. برای رسیدن به این هدف خصوصیات زیادی اندازه‌گیری می‌شود و مورد ارزیابی قرار می‌گیرد. هدف اصلاح‌کننده آن است که در کمترین زمان بیشترین سود ژنتیکی را به‌دست آورد. از تکنیک‌های چند متغیره در اصلاح گیاهان زراعی برای تبدیل اندازه‌گیری‌های حاصل از چندین متغیر مربوط به عملکرد یا میزان پروتئین به شاخص انتخاب استفاده می‌شود. در اینجا هدف کاهش داده‌ها و تشکیل شاخصی است که بتواند جایگزین اندازه‌گیری چندین متغیر شود. امروزه برای تعیین فاصله ژنتیکی گونه‌ها از یکدیگر و پی‌بردن به منشأ آن‌ها صفات به‌خصوص را اندازه‌گیری می‌کنند؛ و سپس با استفاده از تکنیک‌های آماری چند متغیره ژنوتیپ‌ها را گروه‌بندی می‌کنند به‌طوری‌که افراد داخل یک گروه دارای شباهت بیشتری نسبت به افراد بین گروه‌ها هستند (فرشادفر، ۱۳۸۴؛ Singh & Chaudhary 1979).

۱-۲ سازمان‌دهی داده‌ها

پس از اندازه‌گیری چندین متغیر باید داده‌ها را تنظیم و به‌صورت‌های مختلفی ارائه نمود. برای نمونه می‌توان آن‌ها را به‌صورت نمودار یا جدول نشان داد، گاهی اوقات خلاصه‌ای از داده‌ها را نشان می‌دهند و بدین‌صورت بعضی از خواص کمی متغیرها را ارائه می‌کنند. به‌وجود آمدن داده‌های چند متغیره بدین‌صورت است که تعداد $P \geq 1$ متغیر یا صفت برای چندین فرد یا ماده آزمایشی یا آزمودنی اندازه‌گیری می‌شود. برای نشان‌دادن متغیر i ام از فرد j ام از علامت X_{ij} استفاده می‌شود، بنابراین اگر P متغیر و n اندازه‌گیری داشته باشیم می‌توان آن‌ها را به‌صورت ماتریس X با P ردیف و n ستون تنظیم کرد.

فرد n	...	فرد j	...	فرد دو	فرد یک	
X_{1n}	...	X_{1j}	...	X_{12}	X_{11}	متغیر اول
X_{2n}	...	X_{2j}	...	X_{22}	X_{21}	متغیر دوم
				$\vdots$	$\vdots$	$\vdots$
X_{in}	...	X_{ij}	...	X_{i2}	X_{i1}	متغیر i ام
				$\vdots$	$\vdots$	$\vdots$
X_{pn}	...	X_{pj}	...	X_{p2}	X_{p1}	متغیر P ام

۵ کلیات آمار چند متغیره و جبر ماتریس

همین وضعیت را می‌توان به صورت ماتریس X با P ردیف و n ستون نشان داد.

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & & & & \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & & & \\ x_{p1} & x_{p2} & \cdots & x_{pj} & \cdots & x_{pn} \end{bmatrix}$$

مثال. از کتاب‌فروشی چهار قبض به منظور پی بردن به وضعیت فروش کتاب دریافت گردید. در هر قبض فروش تعداد کتاب‌های فروخته شده و مقدار کل فروش مشخص شده است؛ بنابراین، متغیر اول مقدار فروش و متغیر دوم تعداد کتاب فروخته شده خواهد بود. برای مطالعه این دو متغیر چهار قبض فروش که نشان‌دهنده تعداد اندازه‌گیری‌ها است مورد بررسی قرار گرفت. هم‌اکنون می‌توان داده‌های حاصل را به صورت جدول زیر تنظیم کرد.

افراد (قبض فروش):

	۱	۲	۳	۴
متغیر اول (فروش)	۴۲	۵۲	۴۸	۵۸
متغیر دوم (تعداد کتاب)	۴	۵	۴	۳

اگر از علامت x_{ij} استفاده کنیم، خواهیم داشت:

$$\begin{array}{llll} x_{11}=42 & x_{12}=52 & x_{13}=48 & x_{14}=58 \\ x_{21}=4 & x_{22}=5 & x_{23}=4 & x_{24}=3 \end{array}$$

مقادیر بالا را می‌توان در ماتریس X خلاصه کرد:

$$X = \begin{bmatrix} 42 & 52 & 48 & 58 \\ 4 & 5 & 4 & 3 \end{bmatrix}$$

از مزایای ارائه داده‌ها به صورت ردیف و ستون یا ماتریس آن است که می‌توان بر روی آن‌ها محاسبات عددی انجام داد (فرشادفر ۱۳۸۴).

۳-۱ مقدمه‌ای بر جبر ماتریس

یکی از رشته‌های مفید ریاضیات جبر ماتریس است و از آن برای درک داده‌های مربوط به پژوهش‌های روان‌شناسی، جامعه‌شناسی، تعلیم و تربیت، کشاورزی، مدیریت و غیره استفاده می‌شود. هر قدر پژوهش بیشتر چند متغیره شود، نیاز آن به جبر ماتریس بیشتر خواهد بود. در بسیاری از موارد جبر ماتریس سبب ساده‌شدن و توضیح بیشتر مفاهیم ریاضی و آماری می‌شود. در این فصل آن بخش‌هایی از جبر ماتریس مورد بحث قرار می‌گیرد که در آمار کاربرد بیشتری دارد (Shayler, 1982).

۱-۳-۱ تعریف ماتریس

ماتریس آرایشی مستطیلی است که از تعدادی ردیف و ستون تشکیل شده‌است. معمولاً سطرها یا ردیف‌های ماتریس r را اول می‌نویسند و ستون‌های C ماتریس را دوم می‌نویسند. در متن این مستطیل یک سری عدد یا علامت است که در هر صورت آن‌ها را اعضاء یا عناصر ماتریس گویند. برای نمونه ماتریس 2×3 زیر را در نظر بگیرید:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 5 \\ 2 & 6 & 3 \end{bmatrix}$$

عناصر ماتریس را با شماره سطر و ستون می‌شناسند مثلاً a_{11} یعنی عنصر ردیف اول و ستون اول که در مثال بالا عدد ۴ است. به همین صورت a_{22} یعنی عنصر ردیف دوم و ستون سوم A که در ماتریس بالا عدد ۳ است. به‌طور کلی a_{ij} یعنی عنصر ردیف i ام و ستون j ام؛ بنابراین ماتریس را می‌توان به‌صورت زیر نوشت:

$$A = [a_{ij}] \quad i = 1, \dots, r, j = 1, \dots, c$$

برگردان^۱ ماتریس: اگر جای ردیف‌ها و ستون‌های یک ماتریس را با یکدیگر عوض کنیم برگردان آن به‌دست می‌آید. برای نمونه برگردان ماتریس A می‌شود A' .

$$A' = \begin{bmatrix} 4 & 7 & 2 \\ 2 & 6 & 3 \\ 3 & 5 & 1 \end{bmatrix}$$

۷ کلیات آمار چند متغیره و جبر ماتریس

اگر $r=c$ باشد آن را ماتریس مربع گویند. ماتریس مربع می‌تواند متقارن^۱ یا نامتقارن باشد. ماتریس متقارن آن است که عناصر بالای قطر اصلی آن مشابه عناصر زیر قطر اصلی آن باشد. به جز اینکه این عناصر زیر قطر برگردان شده‌اند. منظور از قطر اصلی ماتریس مجموعه‌ای از عناصر است که از بالاترین گوشه چپ ماتریس شروع می‌شوند و به پایین‌ترین گوشه راست ماتریس ختم می‌شوند. معمولاً در تجزیه رگرسیون چندگانه با ماتریس‌های متقارن مواجه می‌شویم، در تحلیل چند متغیره نیز این ماتریس‌ها وجود دارند. یکی از ماتریس‌های متقارن ماتریس همبستگی است، خواهیم داشت:

$$R = \begin{bmatrix} 1 & 0/7 & 0/3 \\ 0/7 & 1 & 0/4 \\ 0/3 & 0/4 & 1 \end{bmatrix}$$

عناصر روی قطر همبستگی متغیرها با خودشان است، به همین دلیل است که یک شده‌است. عناصر خارج از قطر همبستگی بین دو متغیر است که توسط اعداد مربوط به سطرها و ستون‌ها شناسایی می‌شوند.

مثلاً $r_{12}=r_{21}=0/7$ و $r_{23}=r_{32}=0/4$ ، به همین صورت است بقیه عناصر.

بردار ستونی^۲ یک ستون $1 \times n$ از اعداد است. برای نمونه:

$$b = \begin{bmatrix} 8 \\ 1/3 \\ -2 \end{bmatrix}$$

یک بردار ردیفی^۳ عبارت است از یک ردیف $1 \times c$ از اعداد است.

b' برگردان b است. بردارها را با حروف کوچک انگلیسی نشان می‌دهیم و بردار

ردیفی را با پریم نشان می‌دهیم.

$$b' = 8 \quad 1/3 \quad -0/2$$

ماتریس قطر^۱ - در کارهای آماری معمولاً با ماتریس قطر مواجه می‌شویم. ماتریس قطر آن است که مقادیر روی قطر اصلی آن صفر نیستند و عناصر خارج از قطر آن نیز صفر هستند. برای نمونه ماتریس زیر را در نظر بگیرید.

1. Symmetric matrix
2. Column vector
3. Row vector

$$B = \begin{bmatrix} ۲/۷۵۹ & ۰ & ۰ \\ ۰ & ۱/۶۴۳ & ۰ \\ ۰ & ۰ & ۰/۸۷۹ \end{bmatrix}$$

ماتریس واحد- یکی از فرم‌های ویژه ماتریس قطر ماتریس واحد است که با i نشان داده می‌شود. در روی قطر اصلی ماتریس واحد یک وجود دارد.

$$I = \begin{bmatrix} ۱ & ۰ & ۰ \\ ۰ & ۱ & ۰ \\ ۰ & ۰ & ۱ \end{bmatrix}$$

۱-۳-۲ عملیات ماتریس

مهم‌ترین عملیات دو ماتریس جمع، تفریق، ضرب و عکس است. عملیات آماری متعددی را می‌توان با اطلاع از قوانین جبری انجام داد. این عملیات عبارت‌اند از:

- **جمع و تفریق:** دو یا چند بردار را وقتی می‌توان باهم جمع یا از هم کم کرد که دارای ابعاد یکسان باشند، یعنی تعداد عناصر یکسانی داشته باشند.

مثال:

$$\begin{bmatrix} ۴ \\ ۳ \\ ۵ \end{bmatrix} + \begin{bmatrix} ۷ \\ ۷ \\ ۴ \end{bmatrix} = \begin{bmatrix} ۱۱ \\ ۱۰ \\ ۹ \end{bmatrix}$$

به همین صورت می‌توان ماتریس‌های دارای ابعاد یکسان را جمع و تفریق کرد. دو ماتریس ۳×۲ زیر را در نظر بگیرید.

$$\begin{bmatrix} ۶ & ۴ \\ ۵ & ۶ \\ ۹ & ۵ \end{bmatrix} + \begin{bmatrix} ۷ & ۴ \\ ۷ & ۴ \\ ۱ & ۳ \end{bmatrix} = \begin{bmatrix} ۱۳ & ۸ \\ ۱۲ & ۱۰ \\ ۱۰ & ۸ \end{bmatrix}$$

هم‌اکنون B را از A کم می‌کنیم:

$$\begin{bmatrix} ۶ & ۴ \\ ۵ & ۶ \\ ۹ & ۵ \end{bmatrix} - \begin{bmatrix} ۷ & ۴ \\ ۷ & ۴ \\ ۱ & ۳ \end{bmatrix} = \begin{bmatrix} -۱ & ۰ \\ -۲ & ۲ \\ ۸ & ۲ \end{bmatrix}$$

- **ضرب:** برای به‌دست آوردن حاصل‌ضرب یک بردار ردیفی در یک بردار ستونی، عناصر هر یک را در دیگری ضرب می‌کنیم و سپس باهم جمع می‌کنیم. برای نمونه، حاصل‌ضرب a' و b به‌صورت زیر است:

$$a_1 \quad a_2 \quad a_3 = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3$$

گفتنی است که حاصل ضرب یک بردار ردیفی در یک بردار ستونی عددی است که آن را اسکالر^۱ گویند. به همین دلیل است که حاصل ضرب یک ردیف در یک ستون را حاصل ضرب اسکالر بردارها گویند. مثال عددی زیر را در نظر بگیرید.

$$4 \quad 1 \quad 3 = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} = (4)(1) + (1)(2) + (3)(5) = 21$$

از حاصل ضرب اسکالر بردارها مفصلاً در تحلیل آماری استفاده می‌شود. برای مثال، برای به دست آوردن مجموع عناصر یک بردار ستونی آن‌ها در یک بردار ردیفی واحد باهمان ابعاد پیش ضرب می‌کنیم (یعنی بردار واحد را در بردار ستونی ضرب می‌کنیم) خواهیم داشت:

$$\sum X = 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad \begin{bmatrix} 1 \\ 4 \\ 1 \\ 3 \\ 7 \end{bmatrix} = 16$$

مجموع مربعات یک بردار ستونی از طریق پیش ضرب برگردان آن در آن بردار به دست می‌آید:

$$\sum X^2 = 1 \quad 4 \quad 1 \quad 3 \quad 7 \quad \begin{bmatrix} 1 \\ 4 \\ 1 \\ 3 \\ 7 \end{bmatrix} = 76$$

به همین صورت، مجموع حاصل ضرب‌های X و Y از طریق ضرب ردیف X در ستون Y یا ردیف Y در ستون به دست می‌آید.