



دانشگاه سیم نور

آمار و احتمالات کاربردی

دکتر معروف خلیلی دکتر همایون کانونی

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 پر بار 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

یازده

پیشگفتار

فصل اول. مرور کلی و آمار توصیفی.....	۱
هدف‌های کلی.....	۱
هدف‌های یادگیری.....	۱
مقدمه.....	۱
۱-۱ جامعه.....	۳
۲-۱ نمونه.....	۳
۱-۲-۱ نمونه‌گیری تصادفی.....	۵
۲-۲-۱ نمونه‌گیری منظم.....	۶
۳-۲-۱ نمونه‌گیری طبقه‌ای.....	۶
۴-۲-۱ نمونه‌گیری خوشه‌ای.....	۸
۵-۲-۱ نمونه‌گیری خوشه‌ای چند مرحله‌ای.....	۸
۳-۱ آمار توصیفی.....	۹
۴-۱ آمار استنباطی.....	۱۰
۵-۱ متغیر.....	۱۱
۶-۱ مقیاس‌های اندازه‌گیری.....	۱۲
۷-۱ علامت جمع در آمار.....	۱۴
۸-۱ تبدیل‌های خطی.....	۱۶
۹-۱ معرفی نرم‌افزار SPSS.....	۱۷
۱-۹-۱ ساخت و مدیریت یک فایل در نرم‌افزار.....	۱۸
۱۰-۱ داده‌ها و طبقه‌بندی آنها.....	۲۱
۱-۱۰-۱ توزیع فراوانی.....	۲۱
۱۱-۱ شاخص‌های تمایل مرکزی.....	۲۳

۲۴.....	۱-۱۱-۱ مد یا نمای توزیع فراوانی
۲۵.....	۲-۱۱-۱ میانه
۲۷.....	۳-۱۱-۱ میانگین
۳۵.....	۴-۱۱-۱ چارک متوسط
۳۶.....	۱۲-۱ شاخص‌های پراکندگی
۳۷.....	۱-۱۲-۱ دامنه تغییرات
۳۸.....	۲-۱۲-۱ انحراف متوسط
۳۹.....	۳-۱۲-۱ انحراف معیار
۴۱.....	۴-۱۲-۱ واریانس
۴۷.....	۵-۱۲-۱ ضریب تغییرات
۴۸.....	خودآزمایی

۵۱.....	فصل دوم. نمودارهای آماری
۵۱.....	هدف‌های کلی
۵۱.....	هدف‌های یادگیری
۵۱.....	مقدمه
۵۲.....	۱-۲ جداول توزیع فراوانی
۵۳.....	۲-۲ نمودار قطاعی یا دایره‌ای
۵۴.....	۳-۲ نمودار میله‌ای
۵۶.....	۴-۲ رسم نمودار برای داده‌های کمی
۶۵.....	۵-۲ نمودار جعبه‌ای
۶۷.....	خودآزمایی

۶۹.....	فصل سوم. احتمالات
۶۹.....	هدف‌های کلی
۶۹.....	هدف‌های یادگیری
۶۹.....	مقدمه
۷۰.....	۱-۳ قانون‌های شمارش برای محاسبه احتمالات
۷۰.....	۱-۱-۳ تبدیل یا جایگشت
۷۱.....	۲-۱-۳ ترکیب
۷۲.....	۳-۱-۳ ترتیب
۷۵.....	۲-۳ مفاهیم ضروری در احتمالات
۷۹.....	۳-۳ احتمال وقوع دو یا چند واقعه
۸۰.....	۴-۳ احتمال A و B
۸۰.....	۵-۳ احتمال A یا B
۸۱.....	۶-۳ احتمال شرطی
۸۲.....	۷-۳ قانون بیز
۸۴.....	خودآزمایی

فصل چهارم. متغیرهای تصادفی گسسته و توزیع آن‌ها.....	۸۵
هدف‌های کلی.....	۸۵
هدف‌های یادگیری.....	۸۵
مقدمه.....	۸۵
۴-۱ توزیع احتمالی متغیر تصادفی گسسته.....	۸۸
۴-۲ امید ریاضی متغیر تصادفی گسسته.....	۸۹
۴-۳ واریانس متغیر تصادفی گسسته.....	۹۰
۴-۴ توزیع دوجمله‌ای.....	۹۱
۴-۵ میانگین و انحراف معیار توزیع دوجمله‌ای.....	۹۴
۴-۶ محاسبه توزیع دوجمله‌ای با SPSS.....	۹۶
۴-۷ توزیع دوجمله‌ای منفی.....	۹۸
۴-۸ توزیع چندجمله‌ای.....	۹۸
۴-۹ توزیع پواسن.....	۹۹
۴-۱۰ توزیع هندسی.....	۱۰۱
۴-۱۱ توزیع فوق هندسی.....	۱۰۱
خودآزمایی.....	۱۰۲
فصل پنجم. متغیرهای تصادفی پیوسته و توزیع آن‌ها.....	۱۰۳
هدف‌های کلی.....	۱۰۳
هدف‌های یادگیری.....	۱۰۳
مقدمه.....	۱۰۳
۵-۱ تابع چگالی احتمال $f(x)$	۱۰۴
۵-۲ امید ریاضی توزیع تصادفی پیوسته.....	۱۰۵
۵-۳ واریانس توزیع تصادفی پیوسته.....	۱۰۵
۵-۴ توزیع یکنواخت.....	۱۰۵
۵-۵ توزیع نرمال.....	۱۰۷
۵-۶ توزیع نرمال استاندارد (توزیع Z).....	۱۰۹
۵-۷ فرض نرمال بودن داده‌ها.....	۱۱۶
۵-۸ آزمون نرمال بودن داده‌ها در نرم‌افزار SPSS.....	۱۱۸
۵-۹ رابطه بین توزیع‌های گسسته و پیوسته.....	۱۱۹
۵-۱۰ قضیه حد مرکزی.....	۱۱۹
خودآزمایی.....	۱۲۲
فصل ششم. تئوری نمونه‌گیری.....	۱۲۳
هدف‌های کلی.....	۱۲۳
هدف‌های یادگیری.....	۱۲۳
مقدمه.....	۱۲۳
۶-۱ مزایا و معایب سرشماری و نمونه‌گیری.....	۱۲۴
۶-۲ روش‌های کلی نمونه‌گیری.....	۱۲۶

۱۲۶	۱-۲-۶ نمونه‌گیری غیر احتمالی
۱۲۶	۲-۲-۶ نمونه‌گیری احتمالی
۱۲۷	۳-۶ خطا در نمونه‌گیری
۱۲۹	۴-۶ روش‌های نمونه‌گیری احتمالی
۱۳۲	۵-۶ اندازه نمونه
۱۳۵	خودآزمایی

۱۳۷	فصل هفتم. استنباط آماری براساس نمونه‌ی بزرگ
۱۳۷	هدف‌های کلی
۱۳۷	هدف‌های یادگیری
۱۳۷	۱-۷ فرضیه آماری
۱۴۲	۲-۷ فواصل اطمینان
۱۴۵	۳-۷ آزمون‌های فرض یک طرفه و دوطرفه
۱۴۸	۴-۷ مفهوم درجه آزادی
۱۵۱	۵-۷ نکاتی پیرامون انتخاب آزمون آماری مناسب
۱۵۲	خودآزمایی

۱۵۳	فصل هشتم. استنباط آماری بر مبنای نمونه‌های کوچک
۱۵۳	هدف‌های کلی
۱۵۳	هدف‌های یادگیری
۱۵۳	۱-۸ توزیع فراوانی نمونه‌های کوچک
۱۵۶	۲-۸ حدود اعتماد میانگین
۱۵۷	۳-۸ نحوه‌ی آزمون t با یک عدد ثابت و خروجی نرم‌افزار SPSS
۱۵۹	۴-۸ آزمون t دو نمونه‌ی مستقل
۱۶۴	۵-۸ آزمون t با داده‌های جفت‌شده
۱۶۶	۶-۸ فواصل اطمینان تفاوت دو میانگین
۱۶۷	خودآزمایی

۱۶۹	فصل نهم. تجزیه‌ی داده‌های طبقه‌بندی شده
۱۶۹	هدف‌های کلی
۱۶۹	هدف‌های یادگیری
۱۶۹	۱-۹ مقایسه دو واریانس و توزیع F
۱۷۲	۲-۹ تجزیه واریانس
۱۷۳	۳-۹ آنالیز واریانس یک طرفه در SPSS
۱۷۷	۴-۹ تفسیر خروجی‌ها
۱۷۹	خودآزمایی

فصل دهم. تجزیه‌ی کای اسکور و آزمون‌های آماری.....	۱۸۱
هدف‌های کلی.....	۱۸۱
هدف‌های یادگیری.....	۱۸۱
مقدمه.....	۱۸۱
۱-۱۰ جدول توافقی بین فراوانی.....	۱۸۵
۲-۱۰ کای اسکور تصحیح‌شده.....	۱۸۶
۳-۱۰ آزمون فرض واریانس.....	۱۸۸
۴-۱۰ حدود اطمینان واریانس جامعه.....	۱۸۹
۵-۱۰ آزمون نسبت‌ها و فراوانی‌ها.....	۱۹۰
۶-۱۰ حدود اعتماد نسبت‌ها و اندازه نمونه.....	۱۹۲
۷-۱۰ توزیع تفاوت نسبت‌ها.....	۱۹۳
خودآزمایی.....	۱۹۴

فصل یازدهم. تبدیل داده‌ها.....	۱۹۷
هدف‌های کلی.....	۱۹۷
هدف‌های یادگیری.....	۱۹۷
مقدمه.....	۱۹۷
۱-۱۱ تبدیل جذری.....	۱۹۸
۲-۱۱ تبدیل لگاریتمی.....	۲۰۰
۳-۱۱ تبدیل زاویه‌ای.....	۲۰۱
۴-۱۱ تبدیل معکوس.....	۲۰۱

فصل دوازدهم. همبستگی و رگرسیون ساده‌ی خطی.....	۲۰۳
هدف‌های کلی.....	۲۰۳
هدف‌های یادگیری.....	۲۰۳
مقدمه.....	۲۰۳
۱-۱۲ آزمون ضریب همبستگی.....	۲۰۷
۲-۱۲ آزمون برابری ضریب همبستگی با صفر.....	۲۰۸
۳-۱۲ آزمون برابری p با یک عدد مشخص غیر صفر.....	۲۰۹
۴-۱۲ آزمون فرض برابری ضریب همبستگی در دو جامعه.....	۲۱۰
۵-۱۲ آزمون معنی‌دار بودن ضریب رگرسیون.....	۲۱۳
۶-۱۲ آزمون اختلاف ضریب رگرسیون با عددی فرضی.....	۲۱۶
۷-۱۲ آزمون اختلاف ضرایب رگرسیون در دو جامعه.....	۲۱۷
۸-۱۲ حدود اعتماد ضریب رگرسیون.....	۲۱۹
۹-۱۲ حدود اعتماد پیش‌بینی.....	۲۲۱
۱۰-۱۲ مشاهدات با تکرار برای x در رابطه‌ی رگرسیون.....	۲۲۵
۱۱-۱۲ پیش‌بینی x ها (متغیرهای مستقل) از روی y ها (متغیرهای وابسته).....	۲۲۷
۱۲-۱۲ مفروضات رابطه رگرسیون.....	۲۲۸

۲۲۹.....	۱۳-۱۲ تجزیه باقیمانده‌ها.....
۲۳۲.....	۱۴-۱۲ رگرسیون خطی و همبستگی از طریق SPSS.....
۲۴۱.....	خودآزمایی.....
۲۴۳.....	فصل سیزدهم. آمار غیرپارامتری.....
۲۴۳.....	هدف‌های کلی.....
۲۴۳.....	هدف‌های یادگیری.....
۲۴۳.....	مقدمه.....
۲۴۵.....	۱-۱۳ آزمون علامت (آزمون نشانه).....
۲۴۷.....	۲-۱۳ انجام آزمون علامت با استفاده از نرم‌افزار SPSS.....
۲۴۹.....	۳-۱۳ آزمون علامت رتبه‌ها.....
۲۵۰.....	۴-۱۳ آزمون ویلکاکسون با استفاده از نرم‌افزار SPSS.....
۲۵۱.....	۵-۱۳ آزمون u.....
۲۵۲.....	۶-۱۳ انجام آزمون U (مان-ویتنی) در نرم‌افزار SPSS.....
۲۵۴.....	۷-۱۳ ضریب همبستگی رتبه.....
۲۵۷.....	خودآزمایی.....
۲۵۹.....	پاسخنامه‌ی خودآزمایی‌ها.....
۲۷۱.....	جداول آماری.....
۲۸۹.....	منابع.....

پیشگفتار

تئوری آمار و احتمالات یک علم ضروری برای تعداد روزافزون مهندسان و کارشناسان علوم مختلف است. به دلیل روشن نبودن عوامل مؤثر بر عملکرد و پیامدهای اقتصادی و اجتماعی آن بر نظام‌های فناوری، در حال حاضر مفاهیم اساسی روش‌های تجزیه و استفاده از این نظام‌ها مبتنی بر نظریه‌ی احتمال و آمار ریاضی هستند. با این وصف، مفاهیم آمار و احتمالات هنوز به اندازه کافی درک یا تفسیر نشده‌اند. کتاب حاضر تلاشی برای ارتقاء این وضعیت به وسیله‌ی فراهم کردن توضیحات ساده‌ی قابل فهم از مفاهیم به ظاهر پیچیده است. هدف از این کتاب فراهم آوردن شرحی مختصر و شفاف از مفاهیم بسیار مهمی است که عموماً در علوم مهندسی به کار می‌روند.

این کتاب از مجموعه مطالب تدریس شده توسط مؤلفان در درس «آمار و احتمالات» رشته‌های مختلف تحصیلی طی سال‌های گذشته تدوین و به زبانی ساده و با استفاده از ابزارهای محدود ریاضی و تأکید بر جنبه‌های آموزشی نگارش شده‌است. همه‌ی مفاهیم و روش‌های نظری به‌طور نظام‌مند با مثال‌های عددی ارائه شده‌اند و از برنامه آماری SPSS به عنوان نرم‌افزار انتخابی برای حل مسائل استفاده شده‌است.

غیر از اطلاع پایه از ریاضیات عمومی، پیش نیاز خاصی برای این درس لازم نیست. این کتاب به شکل یک کتاب درسی نگارش شده، ولی می‌تواند به عنوان کتابی مرجع نیز به وسیله‌ی دانشجویان دوره های کارشناسی و کارشناسی ارشد و همه متخصصانی که در زمینه ارزیابی آماری داده‌ها کار می‌کنند، مورد استفاده قرار گیرد.

به دلیل میدان عمل محدود کتاب، برخی از مفاهیم و روش ها بدون توضیحات نظری ارائه شده اند. در این موارد، خواننده برای مطالعه ی بیشتر به مأخذ مورد نظر ارجاع داده شده است.

نظرات و راهنمایی های خوانندگان محترم در رفع نواقص و تکمیل مطالب در چاپ های بعدی ما را یاری خواهد کرد.

انما العلم عند ا...

معروف خلیلی - همایون کانونی

فصل اول

مرور کلی و آمار توصیفی

هدف‌های کلی

هدف کلی این فصل آشنایی با تعاریف و مفاهیم اولیه آمار و انواع متغیرها، مقیاس‌های مختلف اندازه‌گیری، شاخص‌های تمایل مرکز و شاخص‌های پراکندگی است.

هدف‌های یادگیری

جهت دستیابی به اهداف کلی فوق باید به هدف‌های جزئی زیر دست یابید به‌طوری که بتوانید:

۱. تعریف علم آمار و کاربرد آمار در زندگی روزمره را بیان کنید.
۲. تفاوت بین آمار توصیفی و آمار استنباطی را درک کنید.
۳. با مفاهیم کلی آمار آشنا شوید.
۴. با متغیرهای «پیوسته و گسسته»، «مستقل و وابسته»، «کیفی و کمی» آشنا شوید.
۵. تفاوت جامعه و نمونه آماری را تشخیص دهید.
۶. با انواع مقیاس‌های اندازه‌گیری آشنا شوید.
۷. مهم‌ترین شاخص‌های تمایل به مرکز و پراکندگی را توضیح دهید.

مقدمه

این فصل با بحث در مورد آمار و اهمیت آن شروع شده و در قسمت‌های بعدی عناوین متنوعی در خصوص مفاهیم اساسی علم آمار مرور می‌شوند. درک این مطالب برای تفهیم بخش‌های بعدی مهم است، زیرا در فصول آتی اصول و مفاهیم آمار توضیح داده خواهد شد.

- آمار دربرگیرنده حقایق و ارقام مختلف است. برای مثال:
- بزرگ‌ترین زمین‌لرزه جهان به میزان ۹/۲ در مقیاس ریشتر ثبت شده‌است.
- عملکرد ارقام اصلاح شده‌ی گندم حداقل سه برابر بیشتر از ارقام بومی این محصول است.
- یک‌هشتم مردم آفریقا ناقل بیماری HIV هستند.
- در سال ۲۰۲۰ به ازای هر ۱۵ نفر با سن بالای ۶۵ سال یک نوزاد متولد خواهد شد. این مطالعات عموماً مبنای آماری داشته و با محاسبات عددی به‌دست آمده‌اند، ولی این نتایج بیشتر براساس این‌که انتخاب چگونه انجام‌شده و تحلیل آماری آن‌ها چگونه بوده به‌دست آمده‌اند. مثال‌های زیر را در نظر بگیرید. در این مثال‌ها، خواننده پی خواهد برد که ممکن است اعداد درست باشند، ولی تفسیر می‌تواند غلط انجام‌شده باشد.
- پخش یک آگهی تبلیغاتی جدید برای بستنی میهن در اواخر اردیبهشت سال گذشته به مدت سه ماه، در افزایش فروش این بستنی به میزان ۳۰ درصد موثر بوده است. پس نتیجه‌گیری می‌شود، آگهی تبلیغاتی مؤثر واقع شده‌است.
- ایراد اصلی نتیجه‌گیری مزبور در این است که بدون ارتباط با آگهی، معمولاً در ماه‌های خرداد، تیر و مرداد مصرف بستنی افزایش می‌یابد. پس این نتیجه‌گیری قدری با اثرات جانبی آن اختلاط دارد.
- در شهرهایی که تعداد بیشتری سینما وجود دارد جرم و جنایت بیشتر صورت می‌گیرد. پس نتیجه گرفته می‌شود وجود تعداد بیشتر سینما در یک شهر به جنایات بیشتر منجر می‌شود.
- ایراد نتیجه‌گیری در این است که تعداد بیشتر سینماها و جرائم با بیشتربودن جمعیت یک شهر در ارتباط هستند. در شهرهای بزرگ‌تر هم سینماهای بیشتری وجود دارد و هم اتفاقات و جنایات بیشتری رخ می‌دهد. در فصل‌های بعد توضیح داده خواهد شد که در چنین مواردی ممکن است متغیر سومی هم مطرح باشد.
- در سال زراعی ۹۳-۱۳۹۲، میزان نزولات جوی تأثیری در افزایش عملکرد محصول نداشته است. این نتیجه نیز با اشکال مواجه است، زیرا در این مورد به اطلاعات دقیق‌تری نیاز است. با اطلاعات ثبت شده در یک سال نمی‌توان چنین نتیجه‌گیری کرد و لذا چنین آماری قابل اعتماد نیستند.

این مثال‌ها نشان می‌دهند، آمار فقط اعداد و ارقام نیستند، بلکه آمار چیزی بیش از این‌هاست. بنا به تعریف، آمار به دامنه‌ای از فنون و روش‌ها برای تجزیه، تفسیر و نمایش داده‌ها و تصمیم‌گیری براساس داده‌های موجود اطلاق می‌شود. در آمار اصطلاحاتی به‌طور مستمر و مکرر به کار می‌روند. از جمله می‌توان به جامعه، نمونه، پارامتر، آماره، متغیر و غیره اشاره کرد که در این فصل هرکدام از این موارد تعریف و تشریح می‌شوند.

۱-۱ جامعه^۱

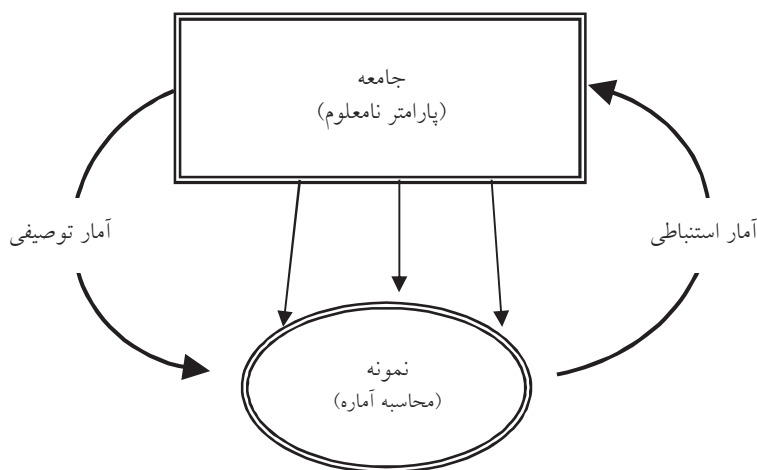
مجموعه‌ای از افراد یا عناصر با حداقل یک ویژگی مشترک یک جامعه خوانده می‌شوند. جوامع آماری به دو نوع متناهی^۲ (محدود) و غیرمتناهی^۳ (نامحدود) تقسیم می‌شوند. جامعه متناهی به جامعه‌ای گفته می‌شود که تعداد عناصر آن محدود باشد. مثلاً تعداد دانش‌آموزان دبیرستان هدف شهرستان ارومیه. در جامعه‌ی غیرمتناهی تعداد افراد (عناصر) یا نامحدود است یا به دلیل این که تعداد بسیار زیاد است آن را نامحدود در نظر می‌گیرند، مانند جامعه ماهی‌های اقیانوس اطلس.

۱-۲ نمونه^۴

نمونه در برابر مفهوم جامعه مطرح می‌شود. نمونه قسمتی از جامعه است که اطلاعات مربوط به آن برای استنتاج درباره جامعه به کار می‌رود. در واقع، هر زیرمجموعه‌ای از واحدهای جامعه را نمونه می‌نامند. برای مثال جامعه بوته‌های فستوک موجود در مراتع استان زنجان را در نظر بگیرید. هر گاه برای تجزیه و تحلیل، تعدادی از این بوته‌ها انتخاب شوند، یک نمونه از جامعه آماری انتخاب شده است. نمونه ممکن است احتمالی یا غیر احتمالی باشد. در آمار نمونه احتمالی بر نمونه غیراحتمالی ارجح است، زیرا اولاً قوانین احتمال در مورد آن صادق است و استنباط آماری براساس همین قوانین بنا می‌شود؛ ثانیاً آن که نمونه غیراحتمالی بیانگر کل جامعه نبوده و ممکن است سلیقه‌ای باشد. از اندازه‌گیری توصیفی به دست آمده از نمونه برای

1. Population.
2. Finite population.
3. Infinite population.
4. Sample.

برآورد ویژگی کل جامعه استفاده می‌شود. با این تعریف طبیعی است که از هر جامعه آماری بتوان نمونه‌های متعددی انتخاب کرد. به‌طور کلی، در بررسی‌های آماری دو هدف دنبال می‌شود. هدف اول که در آمار توصیفی به دنبال آن هستیم، شامل جمع‌آوری، گروه‌بندی، خلاصه‌کردن، رسم نمودار و تعیین شاخص‌های مرکزی و پراکندگی مانند میانگین، میانه، نما، دامنه تغییرات، انحراف متوسط، انحراف معیار و واریانس است. از این شاخص‌ها برای محاسبه شاخص‌های دیگر استفاده می‌شود. هدف دوم که در بحث آمار استنباطی گنجانده می‌شود عبارت‌است از برآورد ویژگی‌های توصیفی جامعه براساس خصوصیات نمونه مورد مطالعه با احتمالی مشخص. در این مفهوم، آمار استنباطی پلی است برای ارتباط بین جامعه و نمونه (شکل ۱-۱).



شکل ۱-۱. رابطه بین جامعه و نمونه

پس بحث عمده‌ی آمار، تصمیم‌گیری در مورد یک جامعه براساس نمونه‌ای از آن جامعه است (حکم از جزء به کل) که در علم منطق حکم از جزء به کل یا از نمونه به جامعه را استدلال استقرائی^۱ می‌گویند. استدلال قیاسی^۲ در مقابل استدلال استقرائی قرار دارد که در آن از کل به جزء یا از جامعه به نمونه حکم می‌شود.

1. Inductive reasoning.
2. Deductive reasoning.

۱-۲-۱ نمونه‌گیری تصادفی

برای استنباط رفتار افراد جامعه از روی مشاهده رفتار افراد نمونه، منطق حکم می‌کند که گروه نمونه معرف جامعه باشد. در چنین حالتی مفهوم «نمونه‌گیری تصادفی» اهمیت پیدا می‌کند. در نمونه‌گیری تصادفی احتمال انتخاب شدن برای همه اعضای جامعه یکسان و معلوم است و هیچ عاملی جز عامل شانس و تصادف در انتخاب شدن افراد نمونه از جامعه وجود ندارد. یعنی، در نمونه‌گیری تصادفی، انتخاب افراد نمونه، مستقل از یکدیگر صورت می‌گیرد. نمونه انتخاب شده به این شیوه معرف جامعه است و می‌توان نتایج حاصل از نمونه را به جامعه تعمیم داد. چهار نوع نمونه‌گیری تصادفی وجود دارد:

- نمونه‌گیری تصادفی ساده^۱
- نمونه‌گیری منظم^۲
- نمونه‌گیری طبقه‌ای^۳
- نمونه‌گیری خوشه‌ای^۴.

در نمونه‌گیری تصادفی ساده احتمال انتخاب شدن برای همه اعضای جامعه یکسان است و هر یک از اعضای جامعه شانس برابر و مستقل برای قرار گرفتن در نمونه را دارند. نمونه‌گیری تصادفی ساده را به دو شیوه می‌توان انجام داد: قرعه کشی و استفاده از جدول اعداد تصادفی. در روش نمونه‌گیری منظم از یک جامعه تعریف شده، نمونه را با نظم خاصی انتخاب می‌کنیم. در نمونه‌گیری طبقه‌ای زیر گروه‌ها با همان نسبتی که در جامعه وجود دارند، به همان نسبت نیز در نمونه به عنوان نماینده جامعه وجود دارند. هنگامی که جامعه مورد تحقیق بزرگ و گسترده باشد و فهرست کامل افراد جامعه مورد تحقیق در دسترس نباشد می‌توان از نمونه‌گیری خوشه‌ای استفاده کرد. در نمونه‌گیری خوشه‌ای واحد نمونه‌گیری فرد نیست، بلکه واحد نمونه‌گیری گروهی از افراد یا خوشه‌ای از افراد است. روش‌های مختلفی که مورد بحث قرار خواهند گرفت، بیشترین کاربرد را در تحقیقات علوم انسانی دارند.

1. Simple Random Sampling
 2. systematic Sampling
 3. stratified Sampling
 4. cluster Sampling

۱-۲-۲ نمونه‌گیری منظم

نمونه‌گیری منظم یا سیستماتیک^۲ شبیه نمونه‌گیری تصادفی ساده است و همان محدودیت‌ها را دارد ولی ساده‌تر است. برای انتخاب نمونه سیستماتیک ابتدا باید با تقسیم حجم جمعیت بر حجم نمونه نسبت نمونه‌گیری^۱ را به دست آوریم. نسبت نمونه‌گیری جمعیتی با ۵۰ نفر و نمونه‌ای با ۱۰ نفر $\frac{1}{5}$ خواهد بود یعنی از هر ۵ نفر یک نفر را انتخاب می‌کنیم. نمونه‌گیری سیستماتیک علاوه بر مسائل نمونه‌گیری تصادفی ساده با مسئله دیگری هم روبروست: تناوب^۲ در چهارچوب نمونه‌گیری. یعنی ممکن است نوع خاصی از افراد در فاصله‌های منظمی در چهارچوب نمونه‌گیری قرار گرفته باشند. اگر نسبت نمونه‌گیری با این فاصله منطبق گردد نمونه فقط شامل این نوع خاص می‌شود و به طور منظم از بقیه مستثنی می‌گردد. به عنوان مثال اگر فهرستی از زوجین تهیه کرده باشیم به گونه‌ای که نام هر زنی به دنبال شوهرش آمده باشد و نسبت نمونه‌گیری ما $\frac{1}{4}$ باشد، نمونه ما فقط شامل یک جنس خواهد بود. اگر در چهارچوب نمونه‌گیری تناوب وجود داشت باید آن را به هم بزنیم یا این که از نمونه‌گیری تصادفی ساده استفاده کنیم.

۱-۲-۳ نمونه‌گیری طبقه‌ای

نمونه‌گیری طبقه‌ای شکل اصلاح شده‌ای از نمونه‌گیری تصادفی ساده و سیستماتیک است که هدف آن رسیدن به نمونه‌های معرف‌تر و در نتیجه دقیق‌تر است که البته روال پیچیده‌تری دارد. با وجود این روی هم‌رفته از محدودیت‌های مشابه دو روش فوق برخوردار است. معرف بودن نسبت گروه‌های مختلف در نمونه مستلزم آن است که همانند نسبت آن گروه‌ها در جمعیت باشد. اما به علت تصادف (خطای نمونه‌گیری) همواره چنین امری رخ نمی‌دهد. فی‌المثل چه بسا نسبت افراد عضو طبقه متوسط یا جوانان در نمونه خیلی بیشتر باشد. گاه این عدم تناسب اهمیتی ندارد اما اگر خصیصه‌ای که نمونه در ارتباط با معرف نباشد به مسئله اصلی مطالعه مربوط باشد کار به تحریف می‌کشد. به عنوان مثال در مطالعه‌ای درباره رفتار مردم در انتخابات، نمونه‌ای که نسبت جوانان در آن کمتر از نسبت واقعی باشد ارقام کلی گمراه‌کننده‌ای درباره

1. Sampling fraction

2. periodicity

گرایش‌های مردم در انتخابات به دست خواهد آمد، چرا که آراء انتخاباتی جوانان با افراد مسن فرق می‌کند.

نمونه‌گیری طبقه‌بندی به رفع این نقیصه کمک می‌کند. در استفاده از این روش ابتدا باید متغیر (یا متغیرهای) مورد طبقه‌بندی مربوط را انتخاب کنیم. متغیر طبقه‌بندی خصیصه‌ای است که می‌خواهیم نمونه از نظر آن کاملاً معرف باشد. با تعیین این متغیر، چهارچوب نمونه‌گیری را مطابق با طبقات (یا لایه‌های) آن متغیر دسته‌بندی می‌کنیم و سپس با استفاده از نمونه‌گیری سیستماتیک افراد را به نسبت مقتضی از هر طبقه انتخاب می‌کنیم. مثلاً گیریم در پی پیمایش دانشجویان یک دانشگاه برای تعیین نگرش‌های آنها به اتحادیه ملی دانشجویان هستیم. می‌توان دانشجویان را بر حسب دانشکده طبقه‌بندی کرد تا نمونه ما شامل نسبت مناسبی از هر دانشکده باشد. بدین منظور لازم است چهارچوب نمونه‌گیری خود را به گونه‌ای تنظیم کنیم که دانشجویان هر دانشکده در یک دسته قرار بگیرند (جدول ۱-۱). سپس نمونه‌ای سیستماتیک انتخاب می‌کنیم. راز نمونه‌گیری طبقه‌بندی در طریقه دسته‌بندی افراد در چهارچوب نمونه‌گیری نهفته است.

جدول ۱-۱. مثالی از نمونه‌گیری سیستماتیک و طبقه‌ای

نمونه		نسبت نمونه = $\frac{1}{5}$	جمعیت		
درصد	تعداد		درصد	تعداد	دانشکده
۵	۲۵		۵	۵۰۰	کشاورزی
۳۰	۱۵۰		۳۰	۳۰۰۰	هنر
۲۰	۱۰۰		۲۰	۲۰۰۰	علوم
۵	۲۵		۵	۵۰۰	پزشکی
۷	۳۵		۷	۷۰۰	مهندسی
۱۶	۸۰		۱۶	۱۶۰۰	بازرگانی
۷	۳۵		۷	۷۰۰	حقوق
۱۰	۵۰		۱۰	۱۰۰۰	علوم تربیتی
	۵۰۰			۱۰۰۰۰	تعداد کل

۱-۲-۴ نمونه‌گیری خوشه‌ای

در نمونه‌گیری خوشه‌ای، واحد اندازه‌گیری فرد نیست بلکه گروهی از افراد هستند که به صورت طبیعی شکل گرفته و گروه خود را تشکیل داده‌اند. نمونه‌گیری خوشه‌ای زمانی به کار می‌رود که انتخاب گروهی از افراد امکان‌پذیر و آسان‌تر از انتخاب افراد در یک جامعه تعریف شده باشد. این موقعیت زمانی فرا می‌رسد که بتوانیم فهرست افراد یا اعضای جامعه را تهیه و تدوین کنیم. به عنوان مثال، فرض کنید جامعه مورد نظر و تعریف شده ما عبارت است از کلیه افراد یک شهر که بیشتر از ۱۸ سال سن دارند. در این جا نمونه‌گیری تصادفی ساده و نمونه‌گیری منظم زمانی میسر است که فهرست کامل تمام افراد یک شهر را با سن آنها در دست داشته باشیم. چنانچه چنین فهرستی موجود نباشد، نمونه‌گیری خوشه‌ای مناسب‌ترین روش انتخاب نمونه خواهد بود.

در اجرای روش‌های نمونه‌گیری تصادفی ساده و منظم شهر مورد نظر باید به ۱۶ ناحیه تقسیم‌بندی شود، و افراد هر ناحیه ابتدا فهرست‌بندی و سپس شماره گذاری شوند. سپس کلیه فهرست‌های شماره‌گذاری شده به یک فهرست کلی تبدیل شوند و مجدداً شماره‌گذاری گردند. به جای انجام این عملیات طاقت فرسا و انتخاب افراد به عنوان واحد نمونه‌گیری، منطقه را واحد نمونه‌گیری قرار می‌دهیم و به روش نمونه‌گیری تصادفی ساده از بین مناطق ۱۶ گانه، منطقه یا مناطق مورد نیاز را انتخاب می‌کنیم.

۱-۲-۵ نمونه‌گیری خوشه‌ای چند مرحله‌ای^۱

این روش نوع دیگری از نمونه‌گیری خوشه‌ای است. زمانی که منطقه به صورت تصادفی انتخاب شد، می‌توان نمونه‌گیری را در داخل منطقه نیز ادامه داد. به عنوان مثال، مطالعه کننده ممکن است آدرس کلیه افرادی را که در یک منطقه زندگی می‌کنند، داشته باشد. بنابراین از بین این افراد، ۱۰ نفر را به‌طور تصادفی انتخاب می‌کند. در روش نمونه‌گیری خوشه‌ای چند مرحله‌ای فهرست نمونه‌گیری دو بار تهیه می‌شود. این فهرست در بعضی مواقع ممکن است بیش از دو بار باشد.

نمونه‌گیری خوشه‌ای برخی مواقع در تحقیقات آموزشی به کار می‌رود. در این نوع تحقیقات از کلاس به عنوان واحد نمونه‌گیری استفاده می‌شود. به عنوان مثال، فرض کنید که می‌خواهیم پرسش‌نامه‌ای را برای ۳۰۰ دانش‌آموز در جامعه‌ای که مرکب از تمام دانش‌آموزان سال پنجم دبستان یک شهر است، اجرا کنیم. فرض کنید در این شهر ۱۲۵۰ دانش‌آموز به صورت ۵۰ کلاس که هر کدام به طور متوسط ۲۵ دانش‌آموز را در بر می‌گیرد، وجود دارد. یک روش، تهیه فهرستی از این ۱۲۵۰ دانش‌آموز و سپس انتخاب ۳۰۰ نفر از آنها با استفاده از روش نمونه‌گیری تصادفی ساده یا نمونه‌گیری منظم است. روش دیگر استفاده از نمونه‌گیری خوشه‌ای است، به صورت انتخاب تصادفی ۱۲ کلاس از ۵۰ کلاس موجود و سپس اجرای پرسش‌نامه مورد نظر برای دانش‌آموزان این ۱۲ کلاس. مزیت عمده نمونه‌گیری خوشه‌ای، جلوگیری از اتلاف وقت و صرفه‌جویی در منابع مالی است. استفاده از این روش نمونه‌گیری، پژوهشگر را قادر می‌سازد که پرسش‌نامه‌ای را فقط برای ۱۲ کلاس از ۵۰ کلاس اجرا کند. اگر از نمونه‌گیری تصادفی ساده استفاده می‌شد، پژوهشگر باید ترتیبی اتخاذ می‌کرد که به ۵۰ کلاس دسترسی پیدا کند. حتی اگر ناگزیر می‌شد در بعضی از این کلاس‌ها فقط یک دانش‌آموز را برای نمونه خود انتخاب کند.

• نمونه‌گیری خوشه‌ای معایبی دارد که عبارتند از:

- دقت آن از نمونه‌گیری تصادفی کمتر است، زیرا در نمونه‌گیری تصادفی ساده فقط یک اشتباه نمونه‌گیری وجود دارد در صورتی که در نمونه‌گیری خوشه‌ای در هر مرحله یک اشتباه نمونه‌گیری وجود خواهد داشت.
- برای داده‌های جمع‌آوری‌شده از این نوع نمونه‌گیری فرمول آسانی را نمی‌توان به کار برد، زیرا به‌کاربردن یک نوع ابزار آماری در جامعه‌های مختلف، دقت آن را کاهش می‌دهد. این معایب باید با صرفه‌جویی در وقت و منابع مالی که از مزایای این روش است، مورد مقایسه قرار گیرد.

۳-۱ آمار توصیفی^۱

به بخشی از آمار که در آن داده‌ها خلاصه و تشریح می‌شوند، آمار توصیفی می‌گویند.