۱-۲- مفهوم آمار

کلمهٔ آمار در فارسی به معنای شمارش است و اصل این واژه از زبان فارسی پهلوی می‌باشد. چنین به نظر می‌رسد که کلمهٔ لاتین آمار، Statistics از کلمهٔ لاتین State به معنی دولت استخراج شده باشد. از کلمهٔ آمار دو مفهوم برداشت می‌شود: یکی به معنی اعداد و ارقام است؛ مانند آمار بیکاری، آمار تولیدات یک کارخانه و غیره. مفهوم دوم آن به معنای علم آمار است که از سوی خبرگان، معانی گوناگونی برای آن ارائه شده است که در ادامه به آن‌ها اشاره خواهیم کرد.

اینک هر روزه شاهد گسترش بیش از پیش کاربردهای آمار در زمینه‌های مختلف هستیم. مدیران ما چه در سطوح میانی و چه در سطوح عالی، قطعاً با کاربرد آمار و استفاده از قابلیت‌های فراوان این علم، چه در زمینهٔ توصیف اطلاعات و چه در زمینهٔ استنباط اطلاعات آگاه هستند و تأثیر چشمگیر آن را در موفقیت و پیشرفت کار تأیید می‌کنند. تولیدکنندگان محصولات مختلف می‌توانند با استفاده از آینهٔ آمار، جایگاه خویش را در میان رقیبان ببینند و بسنجند و با استفاده از این علم جهت صحیح را پیدا کنند. اقتصاددانان و مدیران اقتصادی در صورت استفاده از این علم، خواهند فهمید که برنامه‌ریزی‌های اقتصادی چقدر آسان‌تر، علمی‌تر و اثربخش‌تر خواهد بود. در نهایت پژوهشگران، محققان، استادان دانشگاه‌ها و دانشجویان نیز با روی آوردن بیشتر به استفاده از آمار و تحلیل‌های آماری بر ارزش و اثربخشی تحقیقات خود خواهند افزود.

۱-۳- علم آمار

صاحب نظران، برای علم آمار تعریف‌های گوناگونی ارائه داده‌اند که تقریباً همهٔ آن‌ها به تعریف زیر نزدیکتر هستند:

- علم آمار به مجموعه روش‌های علمی اطلاق می‌شود که برای جمع‌آوری اطلاعات اولیه، مرتب کردن، خلاصه کردن، طبقه‌بندی و تجزیه و تحلیل اطلاعات اولیه و تفسیر آن‌ها به کار می‌رود.

Scanned with CamScanner

که استنباط آماری از آن بشود را **آمار توصیفی** می‌نامیم.

ذیلاً به طور مختصر به معرفی بعضی از کاربردهای آمار در رشته‌های مختلف می‌پردازیم.

الف: آمار و برنامه‌ریزی

عصر حاضر، عصر برنامه‌ریزی است و آمار ضرورت آن. موفقیت دولت‌ها در برنامه‌ریزی، به خصوص آن‌ها که از نظر اقتصادی در حال توسعه‌اند، زمانی است که داده‌های آماری مربوطه را به درستی تجزیه و تحلیل کرده باشند.

ب: آمار و اقتصاد

داده‌های آماری و روش‌های تحلیل آماری در حل مسائل اقتصادی مانند دستمزد، قیمت، تحلیل سری‌های زمانی، تحلیل تقاضا و ... به مقیاس وسیع مورد استفاده است و در توسعه تئوری اقتصادی نقش مهمی دارد. آمار و ریاضیات و اقتصاد، موضوع اقتصادسنجی را به همراه آورده‌اند.

پ: آمار و بازرگانی

آمار یکی از ضرورت‌های کنترل محصول است. مدیران بازرگانی با روش‌های آماری می‌توانند به مطالعه نیاز و سلیقه خریداران بپردازند. موفقیت بازرگانان کم و بیش به صحت و دقت پیش‌بینی‌های آماری آنان بستگی دارد.

ت: آمار و صنعت و کشاورزی

در صنعت، آمار به طور وسیع در «کنترل کیفیت» محصول، بازرسی فنی کالای تولید شده و ... مورد استفاده دارد و همچنین پیشرفت‌های عظیم در تولید محصولات کشاورزی و اصلاح بذر و تأثیر کودهای شیمیایی در افزایش محصول و غیره، مدیون دانشمندانی است که در مطالعات خود، روش‌های آماری را به کار گرفته‌اند.

ث: آمار و ریاضیات

آمار به ریاضیات وابستگی بسیار دارد، به طوری که روش‌های پیشرفته آمار محصول ریاضیات کاربردی پیشرفته است و آمار ریاضی شاخه‌ای از آن است.

Scanned with CamScanner

ج: آمار و زیست‌شناسی، نجوم، علوم پزشکی، روانشناسی و آموزش و پرورش

باروری، بهداشت، بیماری و رشد یک جامعه با روش‌های آماری قابل تجزیه و تحلیل است. نظریه گاوس «قانون نرمال خطاها» برای مطالعه حرکت ستارگان با اصل کمترین توان‌های دوم بسط یافته است.

در علوم پزشکی، برای تعیین میزان اثر داروها روی انواع بیماری‌ها و زمان استفاده از آن‌ها و غیره از «آزمون‌های معنی‌دار بودن» به مقیاس وسیع استفاده می‌شود. همچنین آمار از ابزار مطالعه در روانشناسی و آموزش و پرورش است.

ج: آمار و جنگ

از شروع جنگ دوم جهانی از روش‌های آمار بسیار استفاده شده است و خود عامل پیدایش روش‌های تازه در آمار بوده است.

محدودیت‌های کاربرد آمار:

- آمار در زمینه پدیده‌های کیفی، مانند شرافت، فقر، فرهنگ و جز آن کاربرد مناسبی ندارد.
- آمار مطالعه را به صورت انفرادی انجام نمی‌دهد بلکه روش مطالعه آمار، توده‌ای است.
- قوانین آمار مانند قوانین علوم فیزیک، دقیق نبوده بلکه تقریبی و یا به عبارتی، احتمالی است.
- مهمترین محدودیت آمار آن است که حتماً بایستی به وسیله کارشناسان و مختصصان به کارگرفته شود و گرنه آمار ابزار خطرناکی است در دست غیرمتخصصان.

۳.۱ آشنایی با مفاهیم آماری

فرض کنید می‌خواهیم در مورد درآمد ساکنین شهر تهران، یک بررسی آماری انجام دهیم. در این مورد، آمارگر از هر فردی که در خیابان عبور می‌کند، نباید سؤال کند، اول باید مطمئن شود که این فرد، ساکن تهران است و سپس سؤال خود را مطرح کند. لذا ملاحظه می‌کنید، افرادی که ممکن است مورد سؤال قرار گیرند، لااقل باید در یک صفت، یعنی ساکن تهران بودن، مشترک باشند.

تعریف ۲.۱. هر مجموعه‌ای از اشیاء یا افراد که لااقل دارای یک صفت مشترک باشند را جامعه آماری می‌نامیم.

Scanned with CamScanner

۱-۴- آمار توصیفی

در آمار توصیفی سعی بر این است که به هر شکل ممکن بتوانیم ویژگی‌های اطلاعات جمع آوری شده را توصیف کنیم؛ خواه با رسم نمودار، خواه با دسته‌بندی کردن و تهیهٔ جدول توزیع فراوانی یا با محاسبهٔ شاخص‌های مرکزی یا پراکندگی. آنچه اهمیت دارد این است که بتوان ماهیت اطلاعات را برای استفاده‌کنندگان به زبان آماری و در عین حال ساده بیان کرد. اغلب آمار توصیفی را قبل از انجام آزمون آماری به کار می‌گیرند.

۱-۵- آمار استنباطی

ممکن است یک محقق به شکل نظری، الگوهایی را در یک مجموعه داده ببیند. اما آیا این الگوها به صورت تصادفی و غیرواقعی به وجود آمده‌اند و ناشی از تغییرپذیری نمونه‌گیری می‌باشند؟ یا الگوهای واقعی هستند که با هر بار تکرار پروژه، مشاهده خواهند شد؟ پاسخ به این گونه پرسش‌ها را باید در مباحث آمار استنباطی جستجو کرد. در آمار استنباطی مباحث نظری همچون احتمال و توابع توزیع، پایه‌ای برای تحلیل اطلاعات و آزمون‌های آماری هستند.

۱-۶- جامعه و نمونه

هر مجموعه از افراد یا اشیا را که حداقل دارای یک خصوصیت مشترک باشند، **جامعه** می‌گویند؛ مانند:

- دانشجویان رشته حسابداری در یک دانشگاه؛
- دارندگان حساب جاری در یک بانک؛
- ماشین حساب‌های تولید شده در یک خط تولید؛
- فاکتورهای فروش کالا در یک فروشگاه در یک ماه معین.

به هر عضو از جامعهٔ آماری، یک **واحد آماری** می‌گویند. در مثال‌های بالا یک دانشجو، یک دارنده حساب جاری، یک ماشین حساب، هر کدام یک واحد آماری هستند. تعداد کل واحدهای جامعه را **حجم جامعه** می‌گوییم و آن را با حرف N نمایش می‌دهیم.

Scanned with CamScanner

در اغلب اوقات به جامعهٔ اصلی دسترسی نداریم و معمولاً جزئی از جامعه را مورد مطالعه قرار می‌دهیم؛ به آن جزء، **نمونه** می‌گوییم. تعداد نمونه را با n نمایش می‌دهیم.

۱-۷- متغیرها و داده‌ها

در یک جامعهٔ آماری هر عضو ویژگی‌هایی دارد که با آن ویژگی‌ها می‌توانیم یک عضو را توصیف کنیم. به هر ویژگی یک **صفت آماری** می‌گویند؛ مثلاً رنگ، وزن و جنسیت یک فرد، صفت‌هایی برای آن فرد هستند. همین طور زمان تصادف، مکان تصادف، تعداد مصدومان و... صفت‌هایی برای یک تصادف هستند.

صفت‌هایی که در همهٔ اعضای جامعه مشترکند و محدوده مطالعه و تحقیق را معلوم می‌کنند **صفت مشخصه** هستند؛ مانند جامعه «دانشجویان پسر دانشگاه آزاد، متأهل که معدل بالای ۱۷ دارند». هر یک از این صفت‌ها یک صفت مشخصه است.

بر خلاف صفت‌های مشخصه، **صفت‌های متغیر** صفت‌هایی هستند که از هر عضو به عضو دیگر تغییر می‌کنند. از این رو، این گونه صفت‌ها را به اختصار **متغیر** می‌گوییم. این متغیرها هستند که عملاً در حین تحقیق مورد سؤال و اندازه‌گیری قرار می‌گیرند. در مثال بالا درآمد یک دانشجو، صفت متغیری است که پس از اندازه‌گیری آن، یک عدد در اختیار خواهیم داشت. مجموعهٔ این اعداد را **داده‌های آماری** می‌نامیم.

متغیرها نیز انواع مختلفی دارند.

۱-۸- متغیرهای کیفی

بعضی از متغیرهایی که به ظاهر غیرقابل شمارش و اندازه‌گیری هستند؛ مثل رنگ چشم، گروه خون، درجه کیفیت کالا و...را **متغیرهای کیفی** می‌گوییم. این گونه متغیرها ماهیت اندازه پذیری ندارند و تنها با حالات یا وضعیت‌هایی که دارند، نشان داده می‌شوند. متغیرهای کیفی را با پرسشنامه و مصاحبه و امثال آن، ثبت می‌کنند. متغیرهای کیفی می‌توانند **اسمی یا ترتیبی** باشند؛ مثلاً اگر به اعداد ۱ و ۲ را به جنسیت یک فرد

اختصاص دهید، با یک متغیر **کیفی اسمی** روبرو هستید و اگر میزان مهارت او را با اعداد ۱ تا ۳ رتبه‌بندی کنید، یک متغیر **کیفی ترتیبی** خواهید داشت.

۱-۹- متغیرهای کمّی

متغیرهای کمی متغیرهایی قابل شمارش و اندازه‌گیری‌اند و همیشه نتیجهٔ اندازه‌گیری آن‌ها یک عدد است؛ مثل تعداد فرزندان یک خانواده، دمای هوا و مانند آن. در این میان متغیرهایی را که قابل شمارش هستند، متغیر **کمی گسسته** می‌گویند؛ مانند تعداد افراد خانواده، تعداد چک‌های برگشتی در یک بانک و مانند آن. متغیرهایی را که اندازه پذیر هستند و مقادیر آن‌ها در فاصله‌ای از اعداد حقیقی قرار دارد، **متغیرهای کمی پیوسته** می‌گویند. هزینه خانواده، مدت زمان تولید یک قطعه و مانند این‌ها مثال‌هایی از متغیرهای پیوسته است. در فصل چهارم کتاب به مقیاس متغیرها نیز پرداخته‌ایم.

۱-۱۰- جمع‌آوری اطلاعات

جمع‌آوری اطلاعات یعنی اندازه‌گیری صفت‌های متغیر در اعضای جامعه که در موضوع تحقیق پیش بینی شده‌اند. هر نوع جمع‌آوری آمار با طرح‌های دقیق آماری که به وسیلهٔ کارشناسان آمار تهیه و تنظیم می‌شود، انجام می‌گیرد.

۱) اگر از تمامی واحدهای یک جامعه آماری کسب اطلاع کنیم، مانند هنگامی که همهٔ افراد یک کشور را مورد بررسی قرار می‌دهیم، به آن **سرشماری** می‌گویند؛ مثل سرشماری نفوس و مسکن که هر ۱۰ سال یک بار در کشور ما انجام می‌گیرد؛

۲) اگر با روش‌های خاص، بخشی از جامعه را برای مطالعه برگزینیم، به آن **نمونه‌گیری** می‌گویند.

در اغلب موارد به جای سرشماری از نمونه‌گیری استفاده می‌شود؛ زیرا نمونه‌گیری مزایایی مانند سرعت، دقت، کم شدن هزینه‌ها و هدر نرفتن واحدهای جامعه را در پی خواهد داشت.

Scanned with CamScanner

۱-۱۱- نمادگذاری مجموعه‌ی داده‌ها و عمل جمع

برای آشنایی بیشتر با داده‌ها و فرمول‌هایی که در این کتاب ارائه خواهند شد، لازم است با بعضی از نمادهای ضروری آشنا شوید. در آمار همیشه مجموعه‌ی داده‌ها را با نمادهایی نشان می‌دهیم تا بحث ما محدود به یک مجموعه‌ی معلوم از اعداد نباشد. مجموعه‌ی داده‌ها متشکل از تعدادی اندازه است که به طور نمادی به صورت $x_۱,x_۲,...,x_n$ به ترتیب نشان داده می‌شوند. آخرین زیر نویس (n) مربوط به x_n نشان دهنده تعداد داده‌ها است. $x_۱,x_۲,...,x_n$ به ترتیب نشان دهندهٔ اولین داده، دومین داده والی آخر هستند. مثلاً مجموعه داده‌هایی را که از ۵ اندازهٔ ۳ و ۳/۵ و ٤ و ٦ و ۷/۳ تشکیل شده است، با نمادهای $x_۵,x_٤,x_۳,x_۲,x_۱$ نشان می‌دهیم که در آن:

$$x_۱=۳\ ,\ x_۲=۳/۵\ ,\ x_۳=٤\ ,\ x_٤=٦\ ,\ x_۵=۷/۳$$

در مطالعهٔ آمار همه جا با عمل جمع داده‌ها سروکار داریم. برای اجتناب از نوشتن علامت (+)، نماد Σ (سیگما) را به عنوان اختصار برای عمل جمع به کار می‌بریم.

۱-۱۲- نماد جمع (Σ)

نماد $\sum_{i=۱}^{n} x_i$ نشان دهنده مجموع n عدد است و خوانده می‌شود: «مجموع تمام x_i ها که i از ۱ تا n تغییر می‌کند».

$$\sum_{i=۱}^{n} x_i = x_۱ + x_۲ + ... + x_n$$

جمله‌ای که بعد از سیگما نوشته می‌شود، نشان دهندهٔ مقادیری است که باید جمع شوند و نمادهای پایین و بالای Σ دامنهٔ زیرنویس i را معین می‌کنند.

مثال۱: در حالتی که داشته باشیم:

$$\sum_{i=۱}^{۳} x_i = x_۱ + x_۲ + x_۳$$

دامنهٔ زیر نویس i به صورت i=۱,۲,۳ است.

هم چنین برای

$$\sum_{i=۱}^{٤} (x_i - ۲) = (x_۱ - ۲) + (x_۲ - ۲) + (x_۳ - ۲) + (x_٤ - ۲)$$

دامنهٔ زیر نویس i به صورت i=۱,۲,۳,٤ است.

Scanned with CamScanner

مقیاس‌های اندازه‌گیری

چهار نوع مقیاس اندازه‌گیری وجود دارد که عبارتند از: ۱ـ مقیاس اسمی
۲ـ مقیاس ترتیبی یا رتبه‌ای، ۳ـ مقیاس فاصله‌ای، ۴ـ مقیاس نسبتی

مقیاس اسمی: در اندازه‌گیری به طریق مقیاس اسمی از اعداد به منظور جدا کردن نمودها یا عناصر طبقات مختلف، استفاده می‌شود. اعدادی که به نتایج مشاهدات نسبت داده می‌شوند، نقش «علامت» یا «نام» را برای طبقات دارند که مشاهدات در آن‌ها قرار می‌گیرند.

در این اندازه‌گیری اعدادی که به عناصر نسبت داده می‌شوند، هیچ معنای کمی ندارند و نمی‌توان بر روی آن‌ها هیچ‌گونه محاسبات ریاضی (چهار عمل اصلی از قبیل ضرب و جمع) انجام داد.

مقیاس ترتیبی یا رتبه‌ای

در اندازه‌گیری با مقیاس ترتیبی، از اعداد برای مقایسه عناصر از نظر کوچک‌تر یا بزرگ‌تر یا برابر بودن استفاده می‌شود که براساس نتایج این مقایسه‌ها نیز عناصر طبقه‌بندی می‌گردند. اعدادی که به عنوان مقدار در اندازه‌گیری به دست می‌آیند فقط به منظور قرار دادن عناصر در ترتیب از کوچک به بزرگ مورد استفاده قرار می‌گیرند. در این مقیاس، عناصر بر اساس اندازه‌های نسبی که برای آن‌ها به دست می‌آید، مرتب می‌شوند. بر روی اعداد مقیاس ترتیبی نیز هیچ‌گونه عملیات ریاضی انجام نمی‌گیرد.

مقیاس فاصله‌ای

در اندازه‌گیری به شیوه مقیاس فاصله‌ای، عناصر مورد اندازه‌گیری نه تنها از نظر ترتیب اندازه می‌توانند مقایسه شوند، بلکه بر حسب طول فاصله‌ها بین اندازه‌های عناصر نیز می‌توانند مقایسه و طبقه‌بندی گردند. مقیاس فاصله‌ای از مفهوم واحد فاصله استفاده می‌کند و به همین جهت می‌تواند فاصله را بین دو اندازه بر حسب تعداد این واحد فاصله بیان کند.

بر خلاف مقیاس اسمی و مقیاس ترتیبی که روی نتایج اندازه‌گیری، عملیات ریاضی (چهار عمل اصلی) غیرممکن است، در مقیاس فاصله‌ای انجام بعضی از عملیات ریاضی روی اعداد به دست آمده از اندازه‌گیری امکان‌پذیر می‌باشد.

Scanned with CamScanner

مقیاس نسبتی

این مقیاس هم زمانی که اعداد به دست آمده از اندازه‌گیری، از نظر ترتیب و فاصله بین دو مقدار، دارای اهمیت هستند، به کار می‌رود و هم وقتی که نسبت دو عدد به دست آمده از اندازه‌گیری‌ها، دارای اهمیت باشد. به عبارت دیگر، مقیاس نسبتی امکان می‌دهد که تعیین کنیم یک اندازه از اندازه دیگر چقدر بیشتر یا کمتر است و نیز می‌توان تعیین کرد این اندازه برای یک عنصر چند برابر اندازه عنصر دیگر است. وقتی اندازه‌گیری به منظور مقایسه خاصیت مورد مطالعه برای دو عنصر، به صورت نسبت اندازه‌های آن‌ها باشد، استفاده از اندازه‌گیری با مقیاس نسبتی اجتناب‌ناپذیر است. تنها اختلاف اساسی بین مقیاس نسبتی و مقیاس فاصله‌ای در این است که برای مقیاس نسبتی اندازه طبیعی به نام «صفر مطلق» وجود دارد، در صورتی که در اندازه‌گیری بر طبق مقیاس فاصله‌ای، اندازه صفر به طور دلخواه تعیین می‌گردد. در مقیاس نسبتی نیز مانند مقیاس فاصله‌ای «واحد فاصله» بین دو اندازه برای عناصر به طور دلخواه تعیین می‌شود. بر خلاف مقیاس فاصله‌ای که انجام عمل تقسیم روی اعداد به دست آمده از اندازه‌گیری غیرممکن می‌باشد، در مقیاس نسبتی انجام این عمل امکان‌پذیر است و می‌توان بین نسبت‌هایی از اعداد که از اندازه‌گیری‌ها به دست می‌آیند تساوی برقرار کرد. مثلاً در این مقیاس می‌توان گفت: $\frac{۶}{۱۲} = \frac{۱}{۲}$ که نتیجه وجود نقطه صفر تثبیت شده در این مقیاس می‌باشد و برای دو اندازه ۱۲ و ۶ می‌توان گفت که اندازه اولی ۲ برابر اندازه دومی است.

آن قسمت از علم آمار که درباره تخمین پارامترهای جامعه از روی پارامترهای نمونه بحث می‌کند را آمار استنباطی و آن قسمت از علم آمار که درباره خلاصه کردن و توصیف خصوصیات مهم مجموعه داده‌ها بحث می‌کند؛ بدون آن که استنباط آماری از آن بشود را آمار توصیفی می‌نامیم که مشتمل است بر خلاصه کردن داده‌ها در قالب جداول، نمایش ترسیمی آن‌ها به وسیله نمودارها و محاسبه شاخص‌های عددی گرایش به مرکز و پراکندگی.

Scanned with CamScanner

۱-۱۳- آمار توصیفی

خلاصه کردن و توصیف خصوصیات مهم داده‌ها و آنالیز اکتشافی اطلاعات را **آمار توصیفی** می‌گویند. این امر شامل مراحل زیر است:

۱-خلاصه کردن، دسته‌بندی کردن و تهیه جدول‌های توزیع فراوانی؛

۲-محاسبه شاخص‌هایی که گرایش به مرکز و پراکندگی داده‌ها را معین می‌کنند؛

۳-رسم نمودارهای مناسب برای توصیف بصری داده‌ها؛

۱-۱۴- طبقه‌بندی کردن اطلاعات

در نگاه اول ممکن است دادّه‌ها اطلاعات مفیدی به ما ندهند و برای اینکه بدانیم اعداد شامل چه اطلاعاتی هستند، باید آن‌ها را مرتب و دسته‌بندی کنیم تا قابل تفسیر و بهره‌برداری شوند.

۱-۱۵- طبقه‌بندی داده‌های کیفی

از آنجایی که متغیرهای کیفی، وضعیت و حالات واحدهای جامعه را نشان می‌دهند، در آمار توصیفی، جز با رسم نمودار و تعیین چند شاخص محدود، نمی‌توان آن‌ها را به خوبی توصیف کرد. البته می‌توانیم با دسته بندی کردن مشاهدات کیفی، با دقت بیشتری آن‌ها را بررسی کنیم.

مثال۳: در مدیریت، انسان‌ها را به لحاظ ارتباط به ٤ دستهٔ تصویری، احساسی، صوتی و ارقامی تقسیم می‌کنند. کارکنان یک سازمان را از این لحاظ مورد بررسی قرار داده و حاصل نتایج در جدول زیر ثبت شده است. این نتایج قبل از دسته‌بندی شدن به صورت دنباله‌ای از ۱۰۰ حالت بوده‌اند که پس از دسته‌بندی به صورت جدول زیر در آمده‌اند:

X_i نوع ارتباط	تصویری	احساسی	صوتی	ارقامی
f_i فراوانی	۳۵	۲۰	۳۲	۱۳

Scanned with CamScanner

به صورت زیر بیان شده است

گروه خونی	تعداد
A	۲۰
B	۱۵
O	۳۵
AB	۳۰

واضح است که داده‌های فوق از نوع **صفت کیفی اسمی** می‌باشند که طبقه‌بندی شده‌اند و طبقات را نمی‌توان براساس یک طرح، مرتب کرد زیرا ملاکی برای آن که بگوییم یکی از طبقات بهتر از دیگری و یا آن که بزرگتر یا کوچکتر از دیگری است، نداریم و این داده‌ها را نمی‌توان برای محاسبات به کار برد، زیرا طبقات دارای هیچگونه ترتیب و یا مقدار معنی‌دار نیستند، مثلاً نمی‌توان معدل آن‌ها را محاسبه نمود، حتی اگر مثلاً به گروه خونی A عدد ۱ و به B عدد ۲ و به O عدد ۳ و به AB عدد ۴ را نسبت دهیم، این اعداد هیچ معنی خاصی ندارند. ضمناً اگر متوسط را محاسبه کنیم (۱ تا ۲۰، ۲ تا ۱۵، ۳ تا ۳۵ و ۴ تا ۳۰)

$$\frac{۱}{۱۰۰}(۲۰ \times ۱ + ۱۵ \times ۲ + ۳۵ \times ۳ + ۳۰ \times ۴) = ۲/۷۵$$

و این عدد (۲/۷۵) معرف هیچ چیزی نیست و هیچ معنای خاصی ندارد.

حال اگر در همان کارخانه، مهارت کارگران را مورد تحقیق آماری قرار دهیم و آن‌ها را برحسب درجه مهارت طبقه‌بندی کنیم، فرض کنید جدول زیر به دست آید

گروه	تعداد
ضعیف	۲۰
متوسط	۱۵
خوب	۳۵
عالی	۳۰

برای طبقه‌بندی داده‌های پیوسته، به صورت زیر عمل می‌کنیم:

الف) ابتدا دامنهٔ تغییرات داده‌ها را با R نمایش داده و از رابطه زیر به دست می‌آوریم:

$$R = x_{max} - x_{min}$$

که در آن x_{max} بزرگ‌ترین مقدار و x_{min} کوچک‌ترین مقدار در بین داده‌ها است.

ب) تعداد طبقات را به دلخواه و با توجه تعداد داده‌ها در نظر می‌گیریم. معمولاً تعداد طبقات را بین ۵ تا ۲۵ انتخاب می‌کنند.

پ) از تقسیم دامنه تغییرات بر تعداد طبقات، فاصله طبقه‌ها را به دست می‌آوریم. در این حالت فاصله‌ها برای همه طبقات یکسان خواهند بود.

$$C = \frac{R}{k}$$ با تقریب اضافی

برای پوشش تمامی داده‌ها، این مقدار را با تقریب اضافه در نظر می‌گیریم.

ت) با تعیین حدود هر طبقه، جدول توزیع فراوانی را تشکیل می‌دهیم. برای تعیین حدود طبقه اول، کم‌ترین مقدار را نوشته و با توجه به فاصلهٔ طبقات، حد بالای طبقه را می‌نویسیم. برای سایر طبقات نیز به همین صورت عمل می‌کنیم.

مثال ۵: داده‌های زیر درآمد (بر حسب هزار تومان) ۲۵ روز یک فروشگاه است:

۱۰۵	۹۸	۱۰۳	۱۰۴	۱۱۰	۱۰۶	۱۰۳	۱۱۰	۱۰۲
۱۰۴	۹۷	۱۰۱	۱۰۰	۱۰۵	۱۰۴	۱۱۰	۱۰۳	۱۰۵
۱۰۹	۹۸	۹۹	۱۰۴	۱۰۳	۱۰۹	۱۰۰		

این داده‌ها را به پنج طبقه، دسته‌بندی کنید.

حل: در بین داده‌ها، کم‌ترین مبلغ ۹۷ هزار تومان و بیشترین مقدار ۱۱۰ هزار تومان است، دامنهٔ تغییرات را به دست می‌آوریم:

$$R = ۱۱۰ - ۹۷ = ۱۳$$

فاصلهٔ طبقات را نیز می‌توان به صورت زیر به دست آورد:

$$c = \frac{۱۳}{۵} = ۲/۶$$

که با تقریب اضافی باید آن را ۳ در نظر بگیریم.

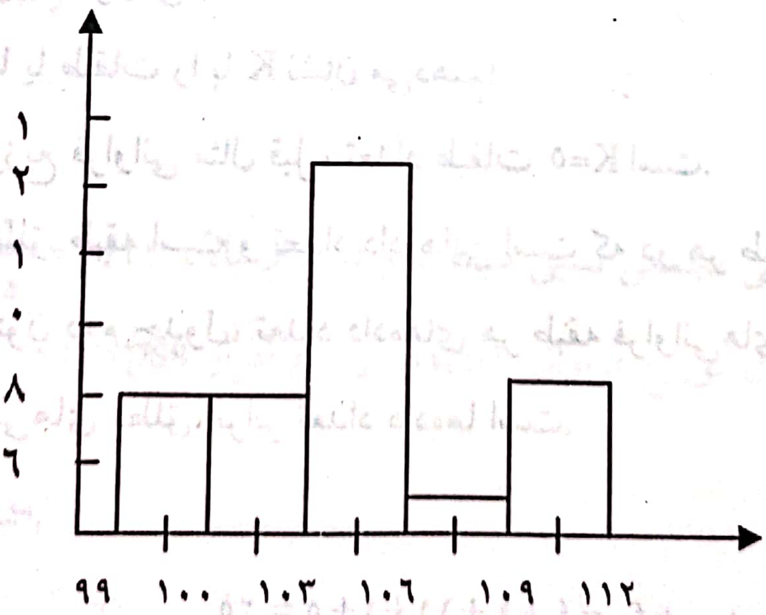
با توجه به فاصله طبقه (۳) و کم‌ترین مقدار (۹۷)، فاصله طبقه اول بین ۹۷ تا ۱۰۰ به دست می‌آید. حدود سایر طبقه‌ها را با رعایت فاصلهٔ طبقات به همین صورت به دست آورده و در یک ستون می‌نویسیم. در ستون دیگر تعداد داده‌هایی را که در هر طبقه قرار می‌گیرند، به عنوان فراوانی آن طبقه ثبت می‌کنیم.

توجه: برای اختصاص دادن داده‌ها به طبقات، حدود هر طبقه را در بازه ( ] قرار دهید.

در نهایت جدول توزیع فراوانی به صورت زیر خواهد بود:

حدود طبقات	فراوانی مطلق f_i
۹۷ — ۱۰۰	٤
۱۰۰ — ۱۰۳	٤
۱۰۳ — ۱۰٦	۱۱
۱۰٦ — ۱۰۹	۱
۱۰۹ — ۱۱۲	٥

در نمودار زیر که به هیستوگرام فراوانی معروف است، می‌توان تصویر بهتری از مجموعه داده‌ها را مشاهده کرد.



نمودار درآمد ۲۵ روز یک فروشگاه

Scanned with CamScanner

داده‌ایم.

۱-۱۸- جدول توزیع فراوانی

جدول توزیع فراوانی می‌تواند اطلاعات نهفته در داده‌ها را آشکار سازد. در یک جدول توزیع فراوانی ممکن است فراوانی‌های دیگری نیز لحاظ شوند که این موضوع به درک بهتری از مجموعه داده‌ها منجر خواهد شد.

مثال۶: به جدول توزیع فراوانی زیر که در آن فراوانی‌های دیگر مثال بالا نیز آمده است، توجه کنید. هر یک از فراوانی‌های این جدول را در ادامه، تعریف کرده‌ایم:

حدود طبقات	f_i	x'_i	r_i	F_i	R_i
۹۷ -۱۰۰	٤	۹۸/۵	۰/۱٦	٤	۰/۱٦
۱۰۰ -۱۰۳	٤	۱۰۱/۵	۰/۱٦	۸	۰/۳۲
۱۰۳ -۱۰٦	۱۱	۱۰٤/۵	۰/٤٤	۱۹	۰/۷٦
۱۰٦ -۱۰۹	۱	۱۰۷/۵	۰/۰٤	۲۰	۰/۸۰
۱۰۹ -۱۱۲	۵	۱۱۰/۵	۰/۲۰	۲۵	۱

در یک جدول توزیع فراوانی:

الف) تعداد رده‌ها یا طبقات را با K نشان می‌دهیم؛

مثلاً در جدول توزیع فراوانی مثال قبل، تعداد طبقات K=۵ است.

ب) f_i فراوانی مطلق طبقه است و تعداد داده‌هایی است که در هر طبقه قرار می‌گیرند.

در مثال بالا در ستون دوم جدول، تعداد داده‌های هر طبقه فراوانی‌های مطلق هستند.

پ) مجموع فراوانی‌های مطلق، برابر تعداد داده‌ها است.

برای این مثال داریم:

$$n = \sum_{i=۱}^{۵} f_i = f_۱ + f_۲ + \ldots + f_۵ = ٤ + ٤ + ۱۱ + ۱ + ۵ = ۲۵$$

Scanned with CamScanner

ت) حد پایین هر طبقه را با L_i و حد بالای آن را H_i نام‌گذاری می‌کنیم؛

به عنوان مثال $L_۱ = ۹۷$، کران پایین طبقه اول و $H_۳ = ۱۰۶$ کران بالای طبقه سوم است.

ث) نمایندۀ طبقه، بین حد بالا و پایین آن طبقه قرار دارد و از رابطه زیر به دست می‌آید:

$$x_i' = \frac{L_i + H_i}{۲}$$

به عنوان مثال، نمایندۀ طبقۀ اول به صورت زیر است:

$$x_۱' = \frac{۹۷ + ۱۰۰}{۲} = ۹۸/۵$$

ج) r_i فراوانی نسبی طبقه است و از تقسیم فراوانی مطلق بر تعداد داده‌ها معین می‌شود.

$$r_i = \frac{f_i}{n}$$

مثلاً فراوانی نسبی طبقه دوم در جدول فراوانی مثال بالا به صورت زیر است:

$$r_۲ = \frac{٤}{۲٥} = ۰/۱٦$$

چ) مجموع فراوانی‌های مطلق یک طبقه و طبقات قبل، فراوانی تجمعی آن طبقه است.

$$F_i = F_۱ + F_۲ + \ldots + F_i$$

در همین مثال، فراوانی تجمعی طبقۀ سوم $F_۳ = ٤+٤+۸=۱۹$ است.

ح) از تقسیم فراوانی تجمعی هر طبقه بر کل فراوانی‌ها، فراوانی تجمعی نسبی (R_i) به دست می‌آید:

$$R_i = \frac{F_i}{n}$$

به عنوان مثال فراوانی تجمعی نسبی طبقۀ دوم در مثال بالا، $R_۲ = \frac{۸}{۲٥} = ۰/۳۲$ است.

توجه: تشکیل جدول توزیع فراوانی داده‌ها، در نرم افزار spss را در فصل پنجم کتاب توضیح داده‌ایم.

Scanned with CamScanner

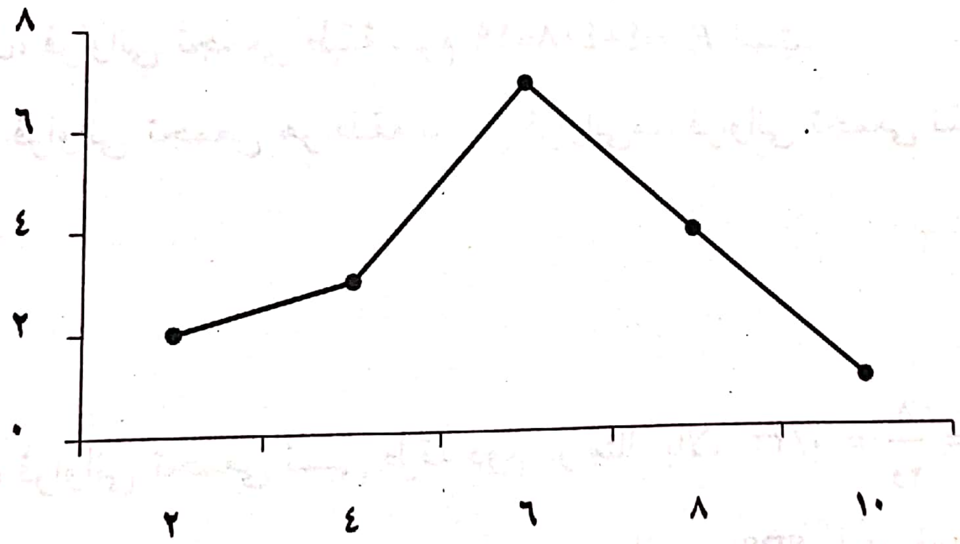
۱-۱۹- رسم نمودار چند ضلعی

در رسم این نمودار، ابتدا روی محور افقی، نمایندهٔ طبقات (x'_i) و روی محور عمودی فراوانی‌ها (f_i) را معلوم می‌کنیم. سپس نقاط (f_i , x'_i) را روی صفحه تعیین کرده و آن‌ها را به یکدیگر وصل می‌نماییم. این نمودار توجیهی از منحنی توزیع است.

توجه: ممکن است در رسم نمودار چند ضلعی، از فراوانی تجمعی نیز استفاده شود.

مثال۷: جدول فراوانی زیر را در نظر گرفته و چند ضلعی فراوانی مطلق آن را به صورت زیر رسم می‌کنیم:

حدود طبقات	f_i	F_i
۱-۳	۲	۲
۳-۵	۳	۵
۵-۷	۷	۱۲
۷-۹	٤	۱٦
۹-۱۱	۱	۱۷



نمودار چندضلعی فراوانی مطلق

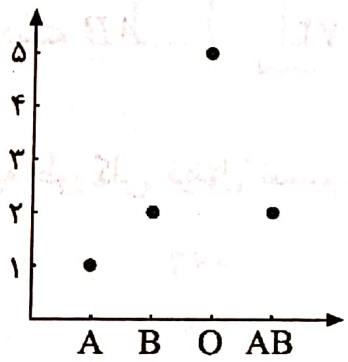
توجه: نحوه رسم اینگونه نمودارها در نرم افزار spss را در فصول بعد توضیح داده‌ایم.

Scanned with CamScanner

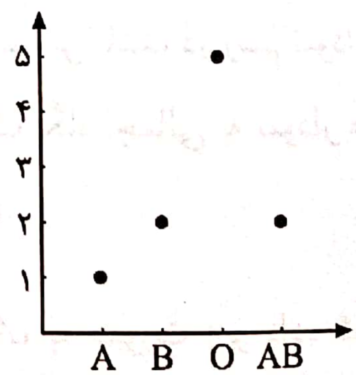
ابتدا نمایش جدولی (جدول توزیع فراوانی) را تشکیل می‌دهیم.

f_i	گروه خونی
۱	A
۲	B
۵	O
۲	AB

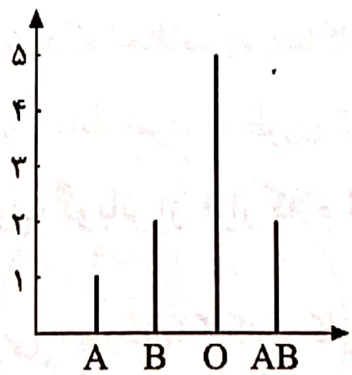
حال دو محور عمود بر هم رسم می‌کنیم، روی محور افقی چهار نقطه دلخواه برای A، B، O، AB انتخاب می‌کنیم (این محور مقیاس ندارد و ترتیب انتخاب نقاط مهم نیست زیرا داده‌ها از نوع کیفی اسمی می‌باشند.) روی محور عمودی که مربوط به فراوانی است نقطه‌ای را به دلخواه انتخاب و آن را ۵ اختیار می‌کنیم، سپس فاصله مبدأ تا ۵ را به ۵ قسمت مساوی تقسیم و به این ترتیب محور را مدرج می‌کنیم، سپس از A خطی عمود بر محور افقی رسم و از ۱ نیز خطی عمود بر محور عمودی رسم می‌کنیم تا یکدیگر را قطع کنند، محل تلاقی را پیدا کرده و همین عمل را برای نقاط دیگر هم انجام می‌دهیم، شکل روبرو به دست می‌آید.



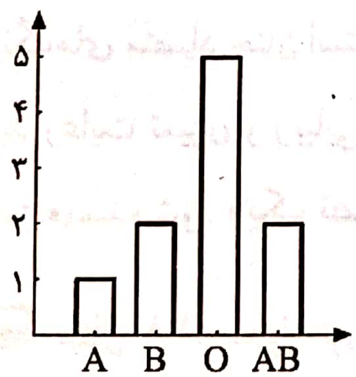
حال اگر این نقاط را پر رنگ کنیم نمودار حاصل را **نمودار سوزنی** (شکل الف) و اگر از این نقاط، خطوط پر رنگی بر محور افقی عمود کنیم، آن را **نمودار میله‌ای** (شکل ب) و اگر به جای خطوط در نمودار میله‌ای از نوارهای پهن استفاده شود آن را **نمودار ستونی** می‌نامیم (شکل پ).



(الف)



(ب)



(پ)

نمودار دیگری که برای داده‌های کیفی رسم می‌شود، **نمودار دایره‌ای** نام دارد در این نمودار، دایره‌ای

Scanned with CamScanner

رسم و آن را به چند بخش طوری تقسیم می‌کنیم که هر بخش نماینده یک طبقه باشد، نحوه تقسیم در این مثال به صورت زیر می‌باشد.

دسته A $\quad \begin{array}{ccc} 10 & & 1 \\ 360 & & x \end{array} \quad , \quad x = \frac{360}{10} \times 1 = 36$

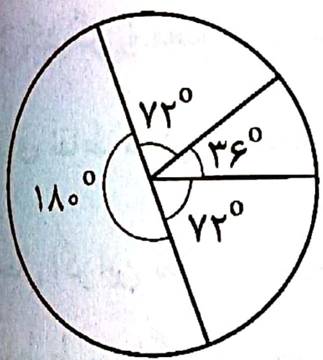
دسته B $\quad \begin{array}{ccc} 10 & & 2 \\ 360 & & x \end{array} \quad , \quad x = \frac{360}{10} \times 2 = 72$

دسته O $\quad \begin{array}{ccc} 10 & & 5 \\ 360 & & x \end{array} \quad , \quad x = \frac{360}{10} \times 5 = 180$

دسته AB $\quad \begin{array}{ccc} 10 & & 2 \\ 360 & & x \end{array} \quad , \quad x = \frac{360}{10} \times 2 = 72$

به طور کلی فرمول تقسیم‌بندی دایره به صورت زیر می‌باشد.

$$x_i = \frac{360}{\text{کل فراوانی}} \times (\text{فراوانی هر دسته})$$



توجه: در نمودار دایره‌ای معمولاً درصد فراوانی هر قسمت را روی آن می‌نویسند و در ضمن از رنگ‌های متضاد چنان استفاده می‌شود که اختلاف زاویه‌ها کاملاً محسوس باشد، در رسم نمودارها باید رعایت تمیزی و زیبایی شکل را لحاظ نمود، به طوری که با یک نگاه اجمالی به نمودار همه چیز فهمیده شود **«یک تصویر خوب گویاتر از هزار کلام است»**.

مثال ۲۸. جدول توزیع فراوانی مهارت ۲۲ کارگر کارخانه A، به صورت زیر داده شده، نمودارهای زیر را رسم کنید.

الف. نمودار سوزنی

ب. نمودار میله‌ای

مثال ۳۱. با استفاده از یک نمودار دایره‌ای نسبت کارکنان یک شرکت را که به ترتیب در جدول زیر رده‌بندی شده است را نشان دهید.

رده	نسبت
مدیریت	۰٫۰۸
دفتری	۰٫۲۲
فروشندگان	۰٫۲۸
خدمات	۰٫۴۲

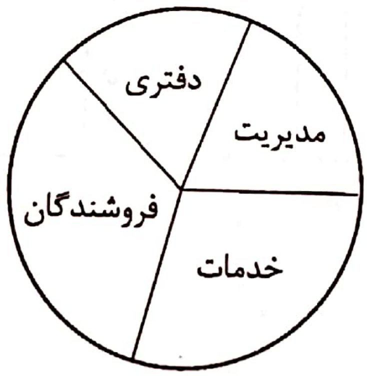
حل: ابتدا با توجه به نسبت داده شده برای هر رده زاویه قطاع مربوط به رده‌ها را تعیین می‌کنیم.

$$\text{زاویه قطاع مربوط به رده مدیریت} = \frac{۸}{۱۰۰} \times ۳۶۰ = ۲۸٫۸$$

$$\text{زاویه قطاع مربوط به رده دفتری} = \frac{۲۲}{۱۰۰} \times ۳۶۰ = ۷۹٫۲$$

$$\text{زاویه قطاع مربوط به رده فروشندگان} = \frac{۲۸}{۱۰۰} \times ۳۶۰ = ۱۰۰٫۸$$

$$\text{زاویه قطاع مربوط به رده خدمات} = \frac{۴۲}{۱۰۰} \times ۳۶۰ = ۱۵۱٫۲$$

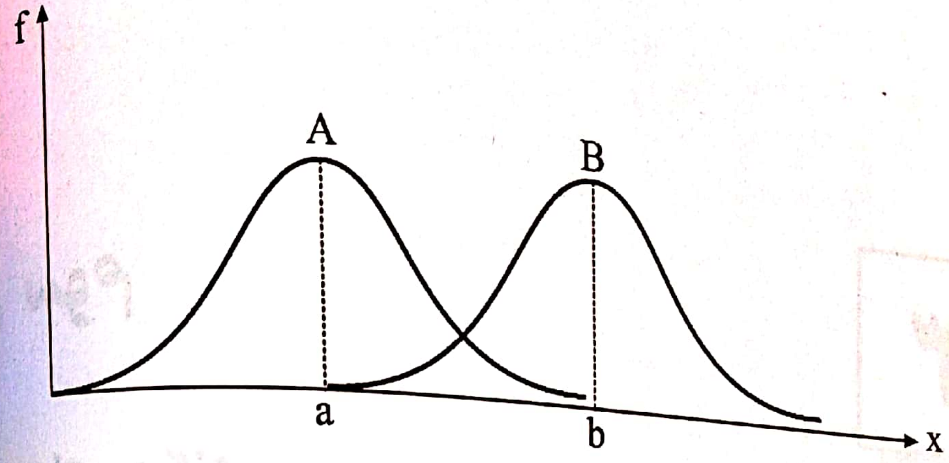


مثال ۳۲. توزیع فراوانی کارگران یک کارخانه بر حسب نوع فعالیت به صورت زیر است:

نوع فعالیت	فراوانی
مکانیک	۲۵
نجار	۱۶
لوله‌کش	۲۰
برق‌کار	۹
	۷۰

Scanned with CamScanner

برای روشن شدن مطلب، فرض کنید، متغیر X را در دو جمعیت مورد بررسی قرار داده‌ایم. ممکن است نوع توزیع X در هر دو جمعیت یکسان باشد، مثلاً X طول قد افراد باشد و توزیع X در جمعیت A متقارن و در جمعیت B نیز متقارن باشد ولی منحنی‌های فراوانی A و B متمایز باشند.



با توجه به شکل بالا، منحنی‌های فراوانی در A و B، هر دو متقارن می‌باشند ولی در جمعیت A، طول قد ۵۰٪ افراد کوچکتر از a و در جمعیت B، طول قد ۵۰٪ افراد کمتر از b می‌باشد و $a \neq b$، لذا در چنین حالتی، می‌گوییم توزیع‌های A و B متفاوت می‌باشند و یا این که می‌گوییم، توزیع‌های A و B از لحاظ تمرکز و یا اندازه مرکزی تفاوت دارند.

۱.۳ شاخص‌های مرکزی

فرض کنید طول قد دانش‌آموزان ۷ ساله در یک روستا را اندازه‌گیری کرده‌ایم و یا مزد کارگران ساده ساختمانی در یک شهرستان و یا قیمت یک کیلو پرتقال از میوه‌فروشی‌ها در یک خیابان مشخص را پرسیده‌ایم، اگر به این داده‌ها توجه کنیم، خواهیم دید که در هر مثال، داده‌ها، حول یک مقدار جمع می‌شوند.

به طورکلی داده‌های یک جامعه آماری، یک نوع تجمع و فشردگی در اطراف یک مقدار خاص از صفت متغیر مورد مطالعه را به وجود می‌آورند، می‌خواهیم این مقدار خاص را به عنوان یک شاخص مرکزی مشخص کنیم و متذکر شویم که یک شاخص مرکزی وقتی با ارزش است که دارای خواص زیر باشد.

۱. در محاسبه آن از تمام داده‌ها استفاده شود.

Scanned with CamScanner

۲. دارای خصوصیات ساده قابل محاسبه باشد.

۳. به فرم ریاضی قابل محاسبه باشد.

از شاخص‌های مهم مرکزی به بررسی میانگین، میانه و مد می‌پردازیم.

۱.۱.۳ میانگین

میانگین خود بر چند نوع است، میانگین حسابی، میانگین حسابی وزنی، میانگین هندسی، میانگین همساز، میانگین درجه دوم، میانگین پیراسته و میانگین وینزوری.

الف. میانگین حسابی که مهمترین شاخص مرکزی است و آن را به اختصار میانگین می‌نامیم، اندازه‌ای از صفت متغیر است که اگر به جای افراد جامعه قرار گیرد، مجموع صفت در جامعه تغییر نکند، میانگین جامعه را با نماد μ و میانگین نمونه را با $\bar{x}$ نمایش می‌دهیم، فرض کنید صفت متغیر دارای مقادیر $x_۱$، $x_۲$، ...، x_n باشد، اگر به جای هر یک از آن‌ها، $\bar{x}$ قرار دهیم، باید داشته باشیم

$$x_۱ + x_۲ + \cdots + x_n = \bar{x} + \cdots + \bar{x}$$

به عبارت دیگر

$$\sum_{i=۱}^{n} x_i = n\bar{x}$$

بنابراین میانگین از فرمول زیر محاسبه می‌شود.

$$\bar{x} = \frac{۱}{n}\sum_{i=۱}^{n} x_i \qquad (۱)$$

مثال ۱. میانگین داده‌ها زیر را محاسبه کنید.

$$۲, ۸, ۴, ۷, ۶, ۵, ۹, ۱, ۳, ۴$$

حل:

$$\bar{x} = \frac{۱}{n}\sum_{i=۱}^{n} x_i$$

$$= \frac{۱}{۱۰}(۲+۸+۴+۷+۶+۵+۹+۱+۳+۴) = \frac{۴۹}{۱۰} = ۴/۹$$

□

Scanned with CamScanner

توجه: اگر توزیع فراوانی صفت گسسته X را به صورت زیر داشته باشیم:

x	$x_۱$	$x_۲$	$\cdots$	x_i	$\cdots$	x_k
فراوانی	$f_۱$	$f_۲$	$\cdots$	$f_۱$	$\cdots$	f_k

با توجه به تعریف میانگین، داریم:

$$f_۱x_۱ + f_۲x_۲ + \cdots + f_kx_k = n\bar{x}$$

$$\boxed{\bar{x} = \frac{۱}{n}\sum_{i=۱}^{k} f_ix_i \quad , \quad \sum_{i=۱}^{k} f_i = n \qquad (۲)}$$

مثال ۲. جدول توزیع فراوانی نمرات ۲۵ دانشجو به صورت زیر بیان شده، مطلوب است محاسبه میانگین

x_i	f_i
۱۴	۶
۱۵	۷
۱۷	۸
۱۹	۴
	۲۵

حل: با توجه به فرمول (۲) داریم:

$$\bar{x} = \frac{۱}{۲۵}\sum_{i=۱}^{۴} f_ix_i = \frac{۱}{۲۵}(۶ \times ۱۴ + ۷ \times ۱۵ + ۸ \times ۱۷ + ۴ \times ۱۹)$$

$$= \frac{۴۰۱}{۲۵} = ۱۶/۰۴ \quad \square$$

مثال ۳. توزیع فراوانی در جدول زیر داده شده است. میانگین را محاسبه کنید.

حدود طبقات	۳ − ۵	۶ − ۸	۹ − ۱۱	۱۲ − ۱۴
f_i	۵	۵	۳	۷

Scanned with CamScanner

حل:

x_i	۴	۷	۱۰	۱۳
f_i	۵	۵	۳	۷
$f_i x_i$	۲۰	۳۵	۳۰	۹۱

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i = \frac{20+35+30+91}{5+5+3+7} = \frac{176}{20} = 8/82 \qquad \square$$

محاسبه میانگین در داده‌های گروه‌بندی شده

اگر صفت متغیر از نوع کمی گسسته با تعداد زیاد و یا از نوع پیوسته باشد، در این صورت داده‌ها را طبقه‌بندی می‌کنیم، فرض کنید در طبقه اول، فراوانی برابر f_1 و در طبقه دوم، فراوانی f_2 و ... در طبقه kام، فراوانی برابر f_k باشد و مثلاً فرض کنید حدود واقعی طبقه اول به صورت $(17/5 - 12/5]$ باشد، این بدان معنی است که f_1 نفر در فاصله $(17/5 - 12/5]$ قرار دارند ولی نمی‌دانیم چند نفر از آن‌ها مثلاً ۱۳، چند نفر ۱۴ و ... می‌باشند، لذا مجبور هستیم برای این افراد، یک نماینده انتخاب کنیم و مرسوم است که نماینده این افراد را وسط طبقه انتخاب می‌کنند یعنی

$$\boxed{\text{نماینده گروه } i\text{ام}، \; x_i = \frac{1}{2}(\text{کرانه پایین گروه } i\text{ام} + \text{کرانه بالای گروه } i\text{ام})}$$

و سپس میانگین را از فرمول (۲) محاسبه می‌نماییم که در آن f_i فراوانی طبقه iام، x_i نماینده آن طبقه است.

مثال ۴. جدول مقابل، جدول توزیع فراوانی نمرات درس فیزیک ۵۰ دانشجو را نشان می‌دهد. میانگین را محاسبه کنید.

حدود طبقات	f_i
۳ – ۵	۹
۶ – ۸	۱۰
۹ – ۱۱	۹
۱۲ – ۱۴	۱۰
۱۵ – ۱۷	۶
۱۸ – ۲۰	۶
	۵۰

Scanned with CamScanner

حل: با توجه به فرمول (۲)، به دو ستون، یکی مربوط به x_i و دیگری مربوط به $f_i x_i$ نیاز داریم لذا، ابتدا این دو ستون را تشکیل می‌دهیم

حدود طبقات	f_i	x_i	$f_i x_i$
۳ − ۵	۹	۴	۳۶
۶ − ۸	۱۰	۷	۷۰
۹ − ۱۱	۹	۱۰	۹۰
۱۲ − ۱۴	۱۰	۱۳	۱۳۰
۱۵ − ۱۷	۶	۱۶	۹۶
۱۸ − ۲۰	۶	۱۹	۱۱۴
			۵۳۶

$$\bar{x} = \frac{۱}{n}\sum_{i=۱}^{۶} f_i x_i = \frac{۵۳۶}{۵۰} = ۱۰/۷۲ \qquad \square$$

مثال ۵. جدول زیر، مربوط به توزیع فراوانی طول قد ۱۰۰ دانشجو می‌باشد. میانگین طول قد را محاسبه کنید.

حدود طبقات	f_i	x_i	$f_i x_i$
۱۵۳ − ۱۵۶/۹	۲	۱۵۴/۹۵	۳۰۹/۹
۱۵۷ − ۱۶۰/۹	۶	۱۵۸/۹۵	۹۵۳/۷
۱۶۱ − ۱۶۴/۹	۲۵	۱۶۲/۹۵	۴۰۷۳/۵
۱۶۵ − ۱۶۸/۹	۳۶	۱۶۶/۹۵	۶۰۱۰/۲
۱۶۹ − ۱۷۲/۹	۲۳	۱۷۰/۹۵	۳۹۳۱/۸۵
۱۷۳ − ۱۷۶/۹	۷	۱۷۴/۹۵	۱۲۲۴/۶۵
۱۷۷ − ۱۸۰/۹	۱	۱۷۸/۹۵	۱۷۸/۹۵
	۱۰۰		۱۶۶۸۳/۰۰

Scanned with CamScanner

ب. میانگین حسابی وزنی

بحث را با یک مثال ساده شروع می‌کنیم، فرض کنید دانشگاهی برای انتخاب دانشجو، صرفاً به معدل او توجه ندارد بلکه مثلاً ترجیح می‌دهد نمره فیزیک را با ضریب ۳، نمره ریاضی را با ضریب ۲ و نمره ادبیات را با ضریب ۱ در نظر بگیرد، می‌خواهیم میانگین نمرات دانشجویی را حساب کنیم که، در درس فیزیک نمره ۱۲، در درس ریاضی نمره ۱۵ و در درس ادبیات نمره ۱۴ گرفته است. واضح است که باید به عدد ۱۲، سنگینی ۳ و به عدد ۱۵ سنگینی ۲ و به عدد ۱۴ سنگینی ۱ نسبت داده شود.

$$\bar{x} = \frac{۳ \times ۱۲ + ۲ \times ۱۵ + ۱ \times ۱۴}{۳ + ۲ + ۱} = \frac{۸۰}{۶} = ۱۳{,}۳$$

به طور کلی اگر داده‌های $x_۱$، $x_۲$، ...، x_k به ترتیب دارای ضرایب وزنی $\omega_۱$، $\omega_۲$، ...، ω_k باشند، آن‌گاه

$$\bar{x} = \frac{\sum \omega_i x_i}{\sum \omega_i} \qquad (۸)$$

فرمول (۲) را نیز می‌توان میانگین حسابی وزنی دانست.

پ. میانگین هندسی

اگر داده‌های $x_۱$، $x_۲$، ...، x_n همگی مثبت باشند، میانگین هندسی از فرمول زیر محاسبه می‌شود.

$$G = \sqrt[n]{x_۱ \times x_۲ \times \cdots \times x_n} \qquad (۹)$$

و اگر داده‌ها گروه‌بندی شده باشند

$$G = \sqrt[n]{x_۱^{f_۱} \times \cdots \times x_k^{f_k}} \qquad (۱۰)$$

که در آن x_i، نماینده طبقه iام و f_i، فراوانی طبقه iام و k تعداد طبقات می‌باشد.

Scanned with CamScanner

مثال ۱۱. میانگین هندسی اعداد زیر را حساب کنید.

$$۲\ ,\ ۳\ ,\ ۸\ ,\ ۲۷$$

حل: با توجه به فرمول (۹) داریم

$$G = \sqrt[۴]{۲ \times ۳ \times ۸ \times ۲۷} = \sqrt[۴]{۲ \times ۳ \times ۲^{۳} \times ۳^{۳}} = \sqrt[۴]{۲^{۴} \times ۳^{۴}} = ۶ \qquad \square$$

مثال ۱۲. اگر جدول توزیع فراوانی به صورت زیر بیان شده باشد، مطلوب است محاسبه میانگین هندسی.

حدود طبقات	f_i
۱۰ − ۱۲	۳
۱۳ − ۱۵	۵
۱۶ − ۱۸	۲

حل: ابتدا ستون x_i را تشکیل داده و با توجه به فرمول (۱۰) داریم

حدود طبقات	f_i	x_i
۱۰ − ۱۲	۳	۱۱
۱۳ − ۱۵	۵	۱۴
۱۶ − ۱۸	۲	۱۷
	۱۰	

$$G = \sqrt[۱۰]{۱۱^{۳} \times ۱۴^{۵} \times ۱۷^{۲}} \qquad \square$$

برای محاسبه G، معمولاً از لگاریتم استفاده می‌کنیم، از طرفین فرمول‌های بالا لگاریتم می‌گیریم. داریم:

$$\log G = \frac{۱}{n} \sum_{i=۱}^{n} \log x_i$$

$$\log G = \frac{۱}{n} \sum_{i=۱}^{k} f_i \log x_i$$

Scanned with CamScanner

مثال ۱۴. جدول زیر، تولید کارخانه‌ای را در سال‌های مختلف نشان می‌دهد. مطلوب است محاسبه رشد متوسط.

سال	حجم تولید X_i	رشد x_i
۱۳۶۳	۱۵	
۱۳۶۴	۲۰	$\frac{۴}{۳}$
۱۳۶۵	۳۰	$\frac{۳}{۲}$
۱۳۶۶	۳۰	۱
۱۳۶۷	۳۵	$\frac{۷}{۶}$
۱۳۶۸	۴۰	$\frac{۸}{۷}$

حل: را حل اول، ابتدا رشد هر سال را نسبت به سال قبل محاسبه می‌کنیم با توجه به فرمول (۹) داریم:

$$G = \sqrt[۵]{\frac{۴}{۳} \times \frac{۳}{۲} \times ۱ \times \frac{۷}{۶} \times \frac{۸}{۷}} = \sqrt[۵]{\frac{۸}{۳}}$$

راه حل دوم، حجم تولید در سال آخر یعنی ۴۰ را بر حجم تولید در سال اول یعنی ۱۵ تقسیم می‌کنیم و سپس ریشه ۱ $- n$ام می‌گیریم

$$G = \sqrt[n-۱]{\frac{X_n}{X_۱}}$$

زیرا

$$x_۱ = \frac{X_۲}{X_۱} \quad , \quad x_۲ = \frac{X_۳}{X_۲} \, , \; \ldots \; , \; x_{n-۱} = \frac{X_n}{X_{n-۱}}$$

$$G = \sqrt[n-۱]{\frac{X_۲}{X_۱} \times \frac{X_۳}{X_۲} \times \cdots \times \frac{X_n}{X_{n-۱}}} = \sqrt[n-۱]{\frac{X_n}{X_۱}}$$

لذا

$$G = \sqrt[۵]{\frac{۴۰}{۱۵}} = \sqrt[۵]{\frac{۸}{۳}} \qquad \square$$

Scanned with CamScanner

از جذر میانگین درجه دوم، معمولاً در کاربردهای فیزیکی استفاده می‌شود مثلاً در سیستم‌های قدرت ولتاژها و جریان‌ها. یکی از کاربردهای دیگر آن که در فصل بعد مطالعه می‌شود در محاسبه انحراف معیار است ک در واقع جذر میانگین درجه دوم انحرافات داده‌ها از میانگین می‌باشد.

توجه: رابطه بین میانگین‌های بیان شده به صورت زیر می‌باشد:

$$H \leq G \leq \bar{x} \leq Q$$

چون در محاسبه میانگین‌های پیراسته و وینزوری، باید از چارک‌های اول و سوم استفاده نمود، لذا بحث این میانگین‌ها را بعد از معرفی چارک‌ها بیان می‌کنیم.

۲.۱.۳ میانه

یکی دیگر از مشخص کننده‌های مرکزی میانه می‌باشد که آن را با نماد M_d نشان می‌دهیم. اگر داده‌ها را به طور غیرنزولی مرتب کنیم، میانه اندازه‌ای از صفت متغیر است که در وسط قرار گرفته باشد.

بنابراین اگر حجم نمونه فرد باشد و داده‌ها را به طور غیرنزولی مرتب کرده باشیم، میانه، داده‌ای است که در وسط قرار می‌گیرد. ولی اگر حجم نمونه زوج باشد در این صورت، میانه برابر با میانگین دو عدد وسطی است. به عبارت دیگر میانه اندازه‌ای از متغیر است که ٪۵۰ داده‌ها که به طور غیرنزولی مرتب شده‌اند کمتر از آن و ٪۵۰ داده‌ها بیشتر از آن باشد. فرض کنید داده‌ها را به طور غیرنزولی مرتب کرده باشیم.

$$x_۱ \leq x_۲ \leq \cdots \leq x_k \leq x_{k+۱} \leq \cdots \leq x_n$$

اگر n فرد و $n = ۲k - ۱$ باشد در این صورت $k = \frac{n+۱}{۲}$ و $M_d = x_k$

و اگر $n = ۲k$ در این صورت $k = \frac{n}{۲}$ و $M_d = \frac{۱}{۲}(x_k + x_{k+۱})$

مثال ۲۴. میانه داده‌های زیر را تعیین کنید.

۹ ۶ ۳ ۷ ۵

Scanned with CamScanner

x_i	f_i	$F(x_i)$
۱۴	۳	۳
۱۵	۶	۹
۱۷	۵	۱۴
۱۸	۲	۱۶
۱۹	۴	۲۰
	۲۰	

حل: چون تعداد داده‌ها زوج است، عدد ۱۰ $= \frac{n}{۲} = k$ را با ستون فراوانی تجمعی مقایسه می‌کنیم. چون عدد ۱۰ بین اعداد ۹ و ۱۴ قرار دارد، لذا نمره نفر ۱۰ام، عدد ۱۷ می‌باشد و بنابراین میانه برابر ۱۷ است. از طرفی طبق فرمول زیر داریم

$$F(x_۲) = ۹ < k = ۱۰ < ۱۴ = f(x_۳) \Longrightarrow M_d = x_۳ = ۱۷ \qquad \square$$

محاسبه میانه در داده‌های گروه‌بندی شده

ابتدا ستون مربوط به فراوانی تجمعی را تشکیل می‌دهیم، سپس عدد $k = \frac{n}{۲}$ را با فراوانی‌های تجمعی مقایسه می‌کنیم، اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی k باشد را به عنوان طبقه میانه‌دار انتخاب می‌کنیم و میانه را از فرمول زیر محاسبه می‌کنیم.

$$\boxed{M_d = L + \frac{\frac{n}{۲} - F_c}{f_i} \times C \qquad (۱۸)}$$

که در آن:

L کران پایین طبقه میانه‌دار

F_c فراوانی تجمعی طبقه قبل از طبقه میانه‌دار

f_i فراوانی مطلق طبقه میانه‌دار

c فاصله طبقه میانه‌دار

Scanned with CamScanner

$$\text{میانه} = \text{کرانه پایین طبقه میانه‌دار} + \frac{\frac{n}{۲} - \text{فراوانی تجمعی طبقه قبل از طبقه میانه‌دار}}{\text{فراوانی مطلق طبقه میانه‌دار}} \times \text{فاصله طبقات}$$

مثال ۲۸. جدول توزیع فراوانی وزن ۱۰۰ دانشجو در جدول زیر داده شده، میانه وزن دانشجویان را محاسبه کنید.

حدود طبقات	f_i	F_c
۶۰ − ۶۲	۵	۵
۶۳ − ۶۵	۱۸	۲۳
۶۶ − ۶۸	۴۲	۶۵
۶۹ − ۷۱	۲۷	۹۲
۷۲ − ۷۴	۸	۱۰۰
	۱۰۰	

حل: (ستون مربوطه به فراوانی تجمعی را بعداً تشکیل داده‌ایم.) حال عدد $k = \frac{n}{۲} = \frac{۱۰۰}{۲} = ۵۰$ را با ستون فراوانی تجمعی مقایسه می‌کنیم، اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی ۵۰ است، طبقه سوم می‌باشد که به عنوان طبقه میانه‌دار انتخاب می‌شود. حال با توجه به فرمول (۱۸) داریم

$$L = ۶۵/۵ \;,\quad \frac{n}{۲} = ۵۰ \;,\quad F_c = ۲۳ \;,\quad f_i = ۴۲ \;,\quad c = ۳$$

$$M_d = ۶۵/۵ + \frac{۵۰ - ۲۳}{۴۲} \times ۳ = ۶۷/۴ \qquad \square$$

مثال ۲۹. در جدول توزیع فراوانی زیر، میانه را محاسبه کنید.

حدود طبقات	f_i	F_c
۵ − ۹	۸	۸
۱۰ − ۱۴	۷	۱۵
۱۵ − ۱۹	۱۰	۲۵
۲۰ − ۲۴	۱۵	۴۰
۲۵ − ۲۹	۱۰	۵۰
	۵۰	

Scanned with CamScanner

مثال ۱۰ : نما در داده‌های زیر عدد ۳ است؛ زیرا بیشتر از همه تکرار شده است.

۳ ۱ ۵ ۹ ۲ ۳ ۱ ۷ ۳ ۲ ۳ ٤ ۳ ۳ ۲

۲-۱۲- محاسبه نما در داده‌های دسته بندی شده

الف- در داده‌های دسته‌بندی شده گسسته نیز به سادگی می‌توان نما را مانند مثال زیـر، به دست آورد.

مثال ۱۱: نما در داده‌های جدول زیر ۱۵ است؛ زیرا عدد ۱۵ بیشترین فراوانی را دارد.

X_i	۱۰	۱۵	۲۰	۲۵
f_i	۸	۱۹	۱۳	۵

توجه: نما عددی منحصر به فرد نیست و ممکن است یک مجموعه داده، فاقد نما باشد یا بیشتر از یک نما داشته باشد.

ب- در داده‌های طبقه‌بندی شده پیوسته، مراحل زیر را برای محاسبه نما دنبال کنید:

۱- ابتدا طبقه‌ای را که فراوانی آن ماکزیمم است، معلوم می‌کنیم (طبقهٔ نما دار).

۲- سپس از فرمول زیر برای محاسبهٔ نما استفاده می‌کنیم:

$$Mo = l_i + \frac{D_1}{D_1 + D_2} \times C$$

در این فرمول:

L_i = کران پایین طبقه‌ای است که فراوانی مطلق آن ماکزیمم است (طبقه نما دار).

D_1 = تفاضل فراوانی مطلق طبقهٔ نما دار با طبقه قبل از آن است؛ یعنی: $D_1 = f_i - f_{i-1}$

D_2 = تفاضل فراوانی مطلق طبقه نما دار با طبقه بعد از آن است؛ یعنی: $D_2 = f_i - f_{i+1}$

C = فاصله طبقه نما دار.

مثال ۱۲: برای داده‌های مثال قبل، نما را محاسبه کنید.

حل: جدول توزیع فراوانی مثال بالا را یک بار دیگر می‌آوریم:

Scanned with CamScanner

کران طبقات	f_i
۱ – ٤	۸
٤ – ۷	۵
۷ – ۱۰	۱۳
۱۰ – ۱۳	۱۵
۱۳–۱٦	۹

طبقه نما دار

$$Mo = ۱۰ + \frac{۲}{۲+٦} \times ۳ = ۱۰ + ۰/۷۵ = ۱۰/۷۵$$

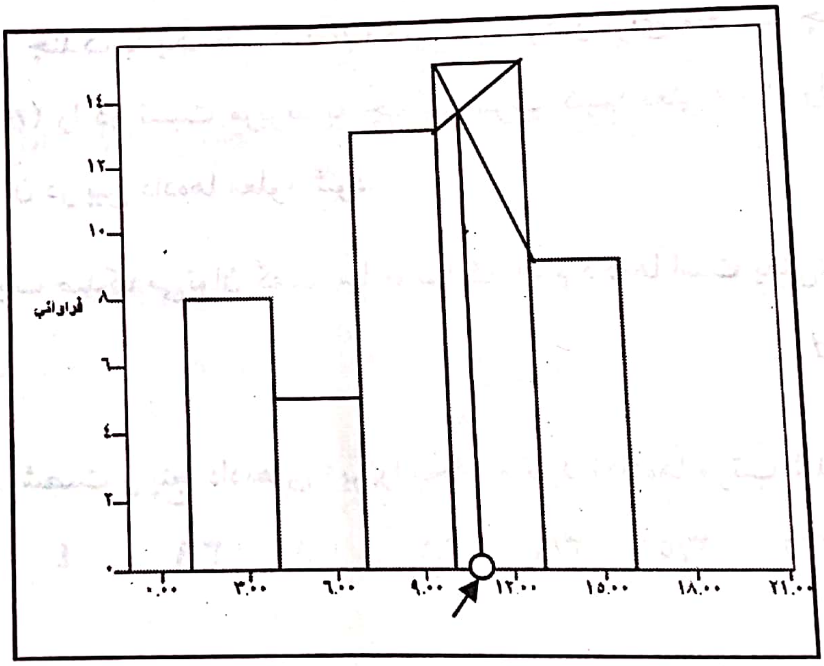
$$D_۱ = ۱۵ - ۱۳ = ۲$$

$$D_۲ = ۱۵ - ۹ = ٦$$

$$Mo = ۱۰/۷۵$$

۲-۱۳- تعیین نما از روی نمودار

در نمودار هیستوگرام فراوانی می‌توان نما را به شکل زیر توجیه کرد:



نمایش نما روی نمودار فراوانی مطلق مثال بالا

Scanned with CamScanner

۳۳) در داده‌های مرتب ...

الف. داده‌های وسط را به دست آورید.

ب. اگر میانه برابر $X_{۲۲}$ باشد مقدار n را به دست آورید.

۳۴) میانه داده‌های $X_{۲k+۱}$ و $X_{۲k+۲}$ و $X_{۶k-۱}$ برابر است با $X_{۹۶}$ مقدار k را به دست آورید.

۳۵) در جدول زیر میانه و میانگین را محاسبه کنید.

حدود طبقات	۲ – ۵	۵ – ۸	۸ – ۱۱	۱۱ – ۱۴	۱۴ – ۱۷
f_i	۴	۱۱	۳	۲	۵

۳۶) در داده‌های زیر مد را تعیین کنید.

الف. ۳ , ۴ , ۵ , ۷ , ۵ , ۶

ب. ۳ , ۴ , ۳ , ۵ , ۴ , ۷

پ. ۲ , ۳ , ۴ , ۵ , ۶

ت. ۱ , ۳ , ۱ , ۴ , ۳ , ۴

۳۷) برای جداول توزیع فراوانی تمرین‌های (۶) و (۸)، مقدار مد را محاسبه کنید.

۳۸) رابطه تجربی پیرسن، فرمول (۲۰) را برای جداول توزیع فراوانی تمرین‌های (۶) و (۸) تحقیق کنید.

۲.۳ چندک‌ها

فرض کنید داده‌ها را به طور غیرنزولی مرتب کرده‌ایم، صدک $۱۰۰p$ام عددی است که لااقل $۱۰۰p$ درصد داده‌ها کوچکتر از آن و $۱۰۰(۱-p)$ درصد داده‌ها بزرگتر از آن باشند $(۰ < p < ۱)$. در واقع میانه یعنی صدک ۵۰ام. چندک‌های معروف عبارتند از:

الف. چارک‌ها که به ازای $p = ۰/۲۵$ و $۰/۵۰$ و $۰/۷۵$ به دست می‌آیند و آن‌ها را با $Q_۱$ و $Q_۲$ و $Q_۳$ نشان می‌دهیم. ($Q_۲$ همان میانه است).

Scanned with CamScanner

ب. دهک‌ها که به ازای ۰/۹ و ... ۰/۲ و ۰/۱ $= p$ به دست می‌آیند و آن‌ها را با $D_1, D_2, \ldots, D_9$ نشان می‌دهیم.

پ. صدک‌ها که به ازای ۰/۹۹ و ... و ۰/۰۲ و ۰/۰۱ $= p$ به دست می‌آیند و آن‌ها را با $P_1, P_2, \ldots, P_{99}$ نشان می‌دهیم.

توجه: $P_{50} = D_5 = Q_2 = M_d$ و مثلاً $P_{75} = Q_3$ و $P_{40} = D_4$.

۱.۲.۳ روش محاسبه چندک‌ها در داده‌های جدا

فرض کنید n داده داریم و آن‌ها را به طور غیرنزولی مرتب کرده‌ایم، برای محاسبه صدک $100p$ام، ابتدا $(n+1)p$ را حساب می‌کنیم اگر $(n+1)p$ برابر عدد صحیح مانند k باشد x_k (عدد kام) نشان‌دهنده صدک $100p$ام می‌باشد و اگر $(n+1)p$ عدد صحیح نباشد، در این صورت

$$(n+1)p = k + r \quad , \quad 0 < r < 1$$

و صدک $100p$ام از فرمول زیر محاسبه می‌شود.

$$\text{صدک } 100p\text{ام} = (1-r)x_k + r x_{k+1} \qquad (21)$$

مثال ۳۸. داده‌های زیر مربوط به نمرات دروس یک دانش‌آموز می‌باشد

الف. چارک اول **ب.** چارک دوم **پ.** چارک سوم **ت.** دهک سوم

ث. صدک ۴۵ام را حساب کنید.

۱۷ , ۱۶ , ۱۹ , ۱۳ , ۱۴ , ۱۸ , ۱۶ , ۱۵ , ۱۷

حل: ابتدا داده‌ها را به طور غیر نزولی مرتب می‌کنیم.

۱۹ , ۱۸ , ۱۷ , ۱۷ , ۱۶ , ۱۶ , ۱۵ , ۱۴ , ۱۳

Scanned with CamScanner

الف. چارک اول برابر با صدک ۲۵ام می‌باشد ۰/۲۵

$$(n+۱)p = (۹+۱) \times ۰/۲۵ = ۲/۵$$

چون $(n+۱)p = ۲/۵$ عدد صحیح نیست، آن را به صورت $۲/۵ = ۲ + ۰/۵$ در نظر می‌گیریم

$$(n+۱)p = ۲/۵ = ۲ + ۰/۵$$

$$k = ۲ \quad , \quad r = ۰/۵ \quad , \quad x_۲ = ۱۴ \quad , \quad x_۳ = ۱۵$$

با توجه به فرمول (۲۱) داریم

$$Q_۱ = (۱ - ۰/۵)x_۲ + ۰/۵x_۳ = ۰/۵ \times ۱۴ + ۰/۵ \times ۱۵ = ۱۴/۵$$

ب. با روشی مشابه با قسمت «الف» و با انتخاب $p = ۰/۵۰$ داریم

$$(x+۱)p = (۹+۱) \times ۰/۵۰ = ۵$$

چون $(n+۱)p = ۵$ عدد صحیح می‌باشد، لذا چارک دوم برابر با $x_۵$ یعنی پنجمین عدد است

$$Q_۲ = x_۵ = ۱۶$$

پ. $Q_۳ = Q_{۰/۷۵}$

$$(n+۱)p = (۹+۱) \times ۰/۷۵ = ۷/۵ = ۷ + ۰/۵$$

$$k = ۷ \quad , \quad r = ۰/۵ \quad , \quad x_۷ = ۱۷ \quad , \quad x_۸ = ۱۸$$

$$Q_۳ = (۱ - ۰/۵)۱۷ + ۰/۵ \times ۱۸ = ۱۷/۵$$

ت. $D_۳ = Q_{۰/۳۰}$

$$(n+۱) \times p = (۹+۱) \times ۰/۳۰ = ۳$$

چون $(n+۱)p = ۳$ عدد صحیح می‌باشد، لذا $D_۳$ برابر با عدد سوم می‌باشد

$$D_۳ = x_۳ = ۱۵$$

Scanned with CamScanner

$$k = ۱۳ \quad , \quad r = ۰/۲۳ \quad , \quad x_{۱۳} = ۱۹ \quad , \quad x_{۱۴} = ۱۹$$

$$P_{۶۳} = (۱ - ۰/۲۳)x_{۱۳} + ۰/۲۳x_{۱۴} = ۰/۷۷ \times ۱۹ + ۰/۲۳ \times ۱۹ = ۱۹ \quad \square$$

۲.۲.۳ روش محاسبه چندک‌ها در داده‌های طبقه‌بندی شده

روش محاسبه چندک‌ها، کاملاً مشابه به روش محاسبه میانه است و برای محاسبه صدک $۱۰۰p$ام، عدد np را با ستون فراوانی تجمعی مقایسه می‌کنیم، اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی np باشد، طبقه صدک $۱۰۰p$ام نامیده می‌شود و صدک $۱۰۰p$ام را از فرمول زیر به دست می‌آوریم

$$\text{صدک } ۱۰۰p\text{ام} = L + \frac{np - F_c}{f} \times C \qquad (۲۲)$$

که در آن

Lـ کرانه پایین طبقه صدک $۱۰۰p$ام.

F_cـ فراوانی تجمعی طبقه قبل از طبقه صدک $۱۰۰p$ام

fـ فراوانی طبقه صدک $۱۰۰p$ام

Cـ فاصله طبقه صدک $۱۰۰p$ام

مثال ۴۰. جدول زیر، جدول توزیع فراوانی نمرات در دروس ریاضی ۵۰ دانشجو را نشان می‌دهد. مطلوب است محاسبه:

الف. چارک اول   **ب.** چارک سوم   **پ.** دهک هفتم   **ت.** صدک ۶۸ام

Scanned with CamScanner

حدود طبقات	f_i	F_c
۷ − ۵	۸	۸
۱۰ − ۸	۷	۱۵
۱۳ − ۱۱	۱۰	۲۵
۱۶ − ۱۴	۱۲	۳۷
۱۹ − ۱۷	۱۳	۵۰
	۵۰	

حل: الف. برای محاسبه چارک اول، عدد $np = ۵۰ \times ۰/۲۵ = ۱۲/۵$ را با ستون فراوانی تجمعی مقایسه می‌کنیم. اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی ۱۲/۵ باشد را به عنوان چارک اول انتخاب می‌شود، لذا طبقه دوم که فراوانی تجمعی آن ۱۵ است را به عنوان طبقه چارک اول انتخاب می‌کنیم و با توجه به فرمول (۲۲) داریم

$$Q_۱ = ۷/۵ + \frac{۱۲/۵ - ۸}{۷} \times ۳ = ۹/۴۳$$

ب. $np = ۵۰ \times ۰/۷۵ = ۳۷/۵$، اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی ۳۷/۵ می‌باشد طبقه آخر است و این طبقه به عنوان طبقه چارک سوم انتخاب می‌شود.

$$Q_۳ = ۱۶/۵ + \frac{۳۷/۵ - ۳۷}{۱۳} \times ۳ = ۱۶/۶۲$$

پ. $np = ۵۰ \times ۰/۷ = ۳۵$ ، طبقه چهارم، به عنوان طبقه دهک هفتم انتخاب می‌شود، زیرا اولین طبقه‌ای است که فراوانی تجمعی آن بزرگتر یا مساوی ۳۵ است.

$$D_۷ = ۱۳/۵ + \frac{۳۵ - ۲۵}{۱۲} \times ۳ = ۱۶$$

ت. $np = ۵۰ \times ۰/۶۸ = ۳۴$، صدک ۶۸ام در طبقه چهارم قرار دارد، با توجه به فرمول (۲۲) داریم:

$$D_{۶۸} = ۱۳/۵ + \frac{۳۴ - ۲۵}{۱۲} \times ۳ = ۱۵/۷۵ \quad \square$$

Scanned with CamScanner

$$\bar{u} = \frac{۱}{n}\sum f_i u_i = \frac{۱}{۵۰} \times ۲۵ = ۰/۵$$

$$(A \cdot D)_U = \frac{۱}{n}\sum f_i d_i = \frac{۱}{۵۰} \times ۴۰ = ۰/۸$$

$$(A \cdot D)_X = C(A \cdot D)_U = ۳ \times ۰/۸ = ۲/۴ \qquad \square$$

۴.۱.۴ واریانس و انحراف معیار

یکی از مهمترین پارامترها در علم آمار، واریانس است و آن عبارت است از میانگین مربع انحرافات از میانگین، و آن را با نماد $V(X)$ یا $S^۲$ نشان می‌دهیم[1]. واریانس جامعه را با نماد $\sigma^۲$ نشان می‌دهیم و آن را با استفاده از فرمول‌های زیر محاسبه می‌کنیم.

الف. داده‌های طبقه‌بندی شده

$$S^۲ = \frac{۱}{n}\sum_{i=۱}^{n}(x_i - \bar{x})^۲ \qquad (۶)$$

ب. داده‌های طبقه‌بندی شده

$$S^۲ = \frac{۱}{n}\sum_{i=۱}^{k} f_i(x_i - \bar{x})^۲ \qquad (۷)$$

مثال ۵. واریانس داده‌های زیر را محاسبه کنید.

$$۱۲ \,,\, ۶ \,,\, ۷ \,,\, ۳ \,,\, ۱۵ \,,\, ۱۰ \,,\, ۱۸ \,,\, ۵$$

حل: ابتدا $\bar{x}$ را محاسبه می‌کنیم

$$\bar{x} = \frac{۱}{n}\sum x_i = \frac{۱}{۸}(۱۲ + ۶ + ۷ + ۳ + ۱۵ + ۱۰ + ۱۸ + ۵) = ۹/۵$$

۱) واریانس نمونه را با $S^۲$ نشان می‌دهیم، اگر فقط منظور ما توصیف نمونه باشد می‌توانیم $S^۲ = \frac{۱}{n}\sum f_i(x_i - \bar{x})^۲$ را محاسبه نماییم ولی چون می‌خواهیم استنباط آماری درباره جامعه به عمل آوریم، بهتر است به جای n، به $n-۱$ تقسیم کنیم زیرا این عمل باعث می‌شود که واریانس نمونه، برآوردکننده نااریب برای واریانس جامعه باشد.

Scanned with CamScanner

با توجه به فرمول (۶) داریم

$$S^۲ = \frac{۱}{n}\sum_{i=۱}^{۸}(x_i - \bar{x})^۲$$

$$= \frac{۱}{۸}\Big((۱۲ - ۹{,}۵)^۲ + (۶ - ۹{,}۵)^۲ + (۷ - ۹{,}۵)^۲ + (۳ - ۹{,}۵)^۲$$

$$+(۱۵ - ۹{,}۵)^۲ + (۱۰ - ۹{,}۵)^۲ + (۱۸ - ۹{,}۵)^۲ + (۵ - ۹{,}۵)^۲\Big) = ۲۳{,}۷۵ \quad \square$$

مثال ۶. جدول توزیع فراوانی نمرات درس آمار ۱۰۰ دانشجو به صورت زیر مفروض است مطلوب است محاسبه واریانس.

حدود طبقات	f_i
۶۰ − ۶۲	۵
۶۳ − ۶۵	۱۸
۶۶ − ۶۸	۴۲
۶۹ − ۷۱	۲۷
۷۲ − ۷۴	۸
	۱۰۰

حل:

حدود طبقات	f_i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^۲$	$f_i(x_i - \bar{x})^۲$	$f_i x_i$
۶۰ − ۶۲	۵	۶۱	−۶,۴۵	۴۱,۶۰۲۵	۲۰۸,۰۱۲۵	۳۰۵
۶۳ − ۶۵	۱۸	۶۴	−۳,۴۵	۱۱,۹۰۲۵	۲۱۴,۲۴۵۰	۱۱۵۲
۶۶ − ۶۸	۴۲	۶۷	−۰,۴۵	۰,۲۰۲۵	۸,۵۰۵۰	۲۸۱۴
۶۹ − ۷۱	۲۷	۷۰	۲,۵۵	۶,۵۰۲۵	۱۷۵,۵۶۷۵	۱۸۹۰
۷۲ − ۷۴	۸	۷۳	۵,۵۵	۳۰,۸۰۲۵	۲۴۶,۴۲۰۰	۵۸۴
	۱۰۰				۸۵۲,۷۵۰۰	۶۷۴۵

$$\bar{x} = \frac{۱}{n}\sum f_i x_i = \frac{۶۷۴۵}{۱۰۰} = ۶۷{,}۴۵$$

Scanned with CamScanner

$$S^{۲} = \frac{۱}{n}\sum f_i(x_i - \bar{x})^{۲} = \frac{۸۵۲/۷۵}{۱۰۰} = ۸/۵۲۷۵ \qquad \square$$

عیب واریانس آن است که آن را حتی با خود صفت هم نمی‌توان مقایسه کرد، زیرا اگر فرض کنیم صفت بر حسب متر باشد در این صورت $x_i - \bar{x}$ نیز بر حسب متر می‌باشد ولی $(x_i - \bar{x})^{۲}$ برحسب مترمربع است، لذا قابل مقایسه نمی‌باشند و این عیب واریانس است برای رفع این عیب از جذر مثبت واریانس استفاده می‌کنیم.

انحراف معیار یا انحراف استاندارد: عبارت است از جذر مثبت واریانس و آن را با نماد σ یا S نشان می‌دهیم.

$$\sigma = \sqrt{\text{واریانس}}$$

مثال ۷. در دو مثال بالا، انحراف معیار را محاسبه کنید.

حل: در مثال (۵)

$$s = \sqrt{۲۳/۷۵} = ۴/۸۷$$

و در مثال (۹)

$$s = \sqrt{۸/۵۲۷۵} = ۲/۹۲ \qquad \square$$

خواص واریانس

الف. اگر به تمام داده‌ها، مقدار ثابت b را اضافه کنیم، واریانس داده‌های جدید برابر با واریانس داده‌های قبلی است. به عبارت دیگر دو متغیر X و Y، دارای واریانس یکسان می‌باشند اگر

$$Y = X + b$$

زیرا

$$V(Y) = V(X+b) = \frac{۱}{n}\sum\left(X_i + b - (\overline{X} + b)\right)^{۲} \qquad ۱$$

$$= \frac{۱}{n}\sum(X_i - \overline{X})^{۲} = V(X)$$

Scanned with CamScanner

همانطور که ملاحظه می‌کنید اعداد به دست آمده در قسمت‌های «الف»، «ب» و «پ» با اعداد داده شده در قسمت‌های (۱) و (۲) و (۳) مطابقت دارد. تحقیق در مورد بندهای (۴) و (۵) به عنوان تمرین به عهده دانشجویان می‌باشد.

۲.۴ پارامترهای نسبی پراکندگی

فرض کنید می‌خواهیم دو جمعیت را با هم مقایسه کنیم، برای این منظور نمی‌توان از ضریبی استفاده نمود که دارای بعد باشد. مثلاً می‌خواهیم در مورد میزان درآمد افراد یا نمرات درس دانش‌آموزان تحقیق کنیم، فرض کنید در یک کشور واحد پول، A و در کشور دیگر واحد پول، B باشد، چطور می‌توان میزان درآمد افراد این دو کشور را با هم مقایسه نمود. یا مثلاً در یک دانشگاه، نمرات دانشجویان بر مبنای ۱۰۰ و در دانشگاه دیگر نمره دانشجویان بر مبنای ۲۰ باشد. لذا، نمی‌توان دو جمعیت را با هم مقایسه نمود.

برای مقایسه دو جمعیت باید از ضریبی استفاده شود که بعد نداشته باشد مثلاً برای مقایسه دو جمعیت نمی‌توان از انحراف معیار استفاده نمود زیرا ممکن است دو صفت از لحاظ واحد اندازه‌گیری یکی نباشند، برای این منظور از پارامترهای نسبی پراکندگی استفاده می‌کنیم. پارامترهای نسبی پراکندگی از نسبت پارامترهای پراکندگی به یک پارامتر مرکزی هم بعد و یا از نسبت یک پارامتر پراکندگی به یک پارامتر پراکندگی دیگر هم بعد به دست می‌آیند. مثلاً اگر صفت بر حسب کیلوگرم باشد می‌دانیم مد بر حسب کیلوگرم و انحراف معیار نیز بر حسب کیلوگرم می‌باشد، لذا $\frac{\text{انحراف معیار}}{\text{مد}} = \frac{\sigma}{M_o}$ یک پارامتر نسبی پراکندگی است. همانطور که ملاحظه می‌کنید، پارامترهای نسبی پراکندگی به صورت یک عدد بدون بعد هستند، لذا قابل مقایسه می‌باشند. ذیلاً سه پارامتر نسبی را بررسی می‌کنیم.

۱. ضریب تغییرات
۲. ضریب چولگی
۳. ضریب کشیدگی

Scanned with CamScanner

۱.۲.۴ ضریب تغییرات یا ضریب واریانس یا پراکندگی نسبی

ضریب تغییرات که با نماد $C \cdot V$ نشان داده می‌شود عبارت است از:

$$C \cdot V = \frac{S}{\bar{x}} \times \%۱۰۰ \qquad (۱۲)$$

مثال ۱۲. فرض کنید میانگین و انحراف معیار طول قد دانشجویان دو کلاس به صورت زیر مفروض باشند.

$$\bar{x} = ۱۷۵ \quad , \quad S_X = ۵ \quad , \quad \bar{y} = ۱۶۰ \quad , \quad S_Y = ۵$$

برای مقایسه پراکندگی این دو جمعیت از ضریب تغییرات استفاده می‌کنیم

$$C \cdot V_X = \frac{S_X}{\overline{X}} \times ۱۰۰ = \frac{۵}{۱۷۵} \times ۱۰۰ = \%۲/۸۵$$

$$C \cdot V_Y = \frac{S_Y}{\overline{Y}} \times ۱۰۰ = \frac{۵}{۱۶۰} \times ۱۰۰ = \%۳/۱۲۵$$

یعنی پراکندگی صفت Y بیش از پراکندگی صفت X است. □

مثال ۱۳. یک تولیدکننده لامپ تصویر تلویزیون دو نوع لامپ تصویر تولید می‌کند. نوع A و B، عمر متوسط A برابر ۱۴۹۵ ساعت و انحراف معیار آن برابر ۲۸۰ ساعت است. عمر متوسط نوع B برابر ۱۸۷۵ ساعت و انحراف معیار آن ۳۱۰ ساعت است. کدام یک از این دو نوع لامپ تصویر دارای پراکندگی نسبی بیشتری است.

حل: ضریب تغییرات را برای هر دو لامپ تصویر محاسبه می‌کنیم.

$$C \cdot V_A = \frac{S_A}{\overline{X}_A} \times ۱۰۰ = \frac{۲۸۰}{۱۴۹۵} \times ۱۰۰ = \%۱۸/۷$$

$$C \cdot V_B = \frac{S_B}{\overline{X}_B} \times ۱۰۰ = \frac{۳۱۰}{۱۸۷۵} \times ۱۰۰ = \%۱۶/۵$$

بنابراین مشاهده می‌شود که لامپ تصویر A دارای پراکندگی بیشتری است. □

Scanned with CamScanner

۱۴-۱- اندازه گیری رابطه بین متغیرها

در بحث مربوط به تابع احتمال توأم، دیدیم که ضریب همبستگی شاخصی است که میزان رابطه بین متغیرها را نشان می‌دهد. یک ملاک مناسب برای تعیین همبستگی دو متغیر کمّی؛ ضریب همبستگی پیرسن بود که، آن را به عنوان ضریب همبستگی جامعه به صورت زیر معرفی کردیم:

$$\rho_{xy} = \frac{E(XY) - E(X)E(Y)}{\sqrt{[E(X^{2}) - E^{2}(X)][E(Y^{2}) - E^{2}(Y)]}} \quad , \quad -1 \le \rho_{xy} \le 1$$

۱۴-۲- ضریب همبستگی نمونه

از طرف دیگر وقتی شما صفت‌های متعددی از یک جمعیت را مطالعه می‌کنید؛ ممکن است بخواهید بدانید بین این صفت‌ها رابطه‌ای وجود دارد یا خیر؟ مثلاً می‌دانیم به طور معمول بین قد و وزن افراد رابطه مستقیم وجود دارد. یعنی اشخاص بلند قد وزن بیشتری نسبت به افراد کوتاه قد دارند. برای چنین بررسی باید با روش‌های نمونه گیری از جمعیت هدف، یک نمونه تصادفی انتخاب کرده و به کمک یک برآوردگر، ضریب همبستگی جامعه را برآورد کرد. چنین برآوردی را ضریب همبستگی نمونه می‌گویند و از آن برای اندازه گیری شدت رابطه بین متغیرها استفاده می‌کنند. در آمار با وجود انواع متغیرها، فرمول‌های متعددی نیز برای اندازه گیری رابطه بین آن‌ها وجود دارد. ضریب همبستگی پیرسن یکی از آن‌ها است که به صورت زیر تعریف می‌شود:

$$r_{xy} = \frac{\overline{x.y} - \bar{x}.\bar{y}}{\sqrt{s_x^{2}.s_y^{2}}}$$

هر یک از اجزاء این فرمول به صورت زیرند:

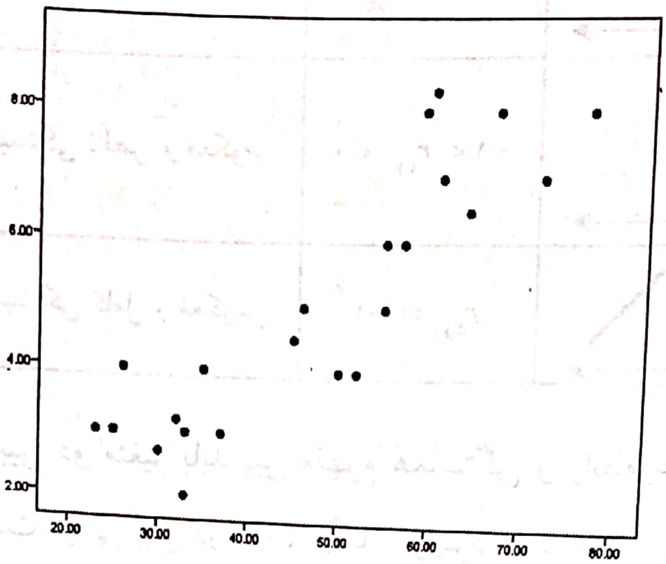
$$\bar{x} = \frac{1}{n}\sum x_i \qquad \bar{y} = \frac{1}{n}\sum y_i \qquad \overline{xy} = \frac{1}{n}\sum x_i y_i$$

$$s_y^{2} = \overline{y^{2}} - \bar{y}^{2} \qquad s_x^{2} = \overline{x^{2}} - \bar{x}^{2}$$

Scanned with CamScanner

۱۴-۳- نمودار پراکنش

یکی از راه‌های مشخص شدن رابطه بین متغیرها، رسم نمودار پراکنش (Scatter plot) آن‌ها است. مطالعه‌ای را در نظر بگیرید که در آن می‌خواهیم رابطه بین میزان بارندگی و مقدار یک محصول کشاورزی را بررسی کنیم. در نمودار زیر، متغیر میزان بارندگی در محور افقی (x) و مقدار محصول را در محور عمودی (y) قرار داده‌ایم. می‌بینید که نقاط بدست آمده روی نمودار نمایش میزان همبستگی آن‌ها را نشان می‌دهد.



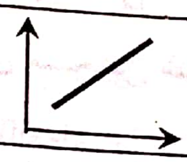
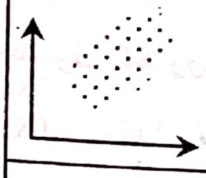
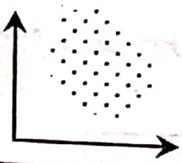
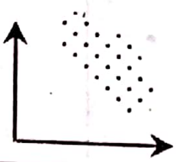
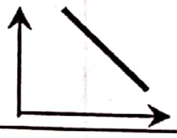
نمودار پراکنش مربوط به میزان بارندگی و مقدار محصول

هر قدر این نقاط به یکدیگر نزدیکتر و در امتداد یک خط قرار داشته باشند، دلیل وجود رابطه‌ای قوی‌تر بین دو متغیر است و بر عکس هر قدر فاصله نقاط از یکدیگر بیشتر و نقاط در امتداد یکدیگر قرار نداشته باشند، دلیل بر عدم وجود رابطه بین آن‌هاست.

۱۴-۴- تفسیر همبستگی نمونه

قبلاً در مورد ضریب همبستگی در جامعه دیدیم که ضریب همبستگی همواره مقداری بین منفی یک تا یک است. در ضریب همبستگی نمونه نیز بر حسب اینکه مقدار همبستگی؛ مثبت، منفی یا صفر باشد، می‌شود نتیجه را به صورتی که در جدول زیر آمده است، توجیه کرد:

Scanned with CamScanner

تفسیر همبستگی	مقدار	نمودار
همبستگی کامل و مستقیم	$r_{xy} = ۱$	
همبستگی ناقص و مستقیم	$۰ < r_{xy} < ۱$	
ناهمبسته (عدم همبستگی)	$r_{xy} = ۰$	
همبستگی ناقص و معکوس	$-۱ < r_{xy} < ۰$	
همبستگی کامل و معکوس	$r_{xy} = -۱$	

در مطالعه ارتباط بین دو متغیر باید بین مفهوم همبستگی و رابطه علت و معلولی بین آن‌ها تفاوت گذاشت. وقتی بین دو متغیر ارتباط وجود دارد، لازم نیست حتماً یکی از آن‌ها علت دیگری باشد. مثلاً در یک تحقیق رابطه بین تعداد خودروهای آتش نشانی در صحنه حادثه با میزان خسارت مالی رابطه ۹۰ درصدی دارد. آیا علت بالا بودن خسارت مالی آتش سوزی، وجود خودروهای آتش نشانی در صحنه حادثه است؟

مثال ۱:

در یک تحقیق، مدت خواب (بر حسب ساعت) و تعداد سیگار ۵ نفر از افراد سیگاری را به صورت جدول زیر در اختیار داریم. با توجه به این داده‌ها، بررسی کنید؛

آیا بین تعداد سیگار و ساعات خواب افراد سیگاری رابطه وجود دارد؟

(x = تعداد سیگار و y = مدت خواب است)

سیگار	۱	۲	۳	۴	۵
مدت خواب	۸	۶	۷	۶	۴

Scanned with CamScanner

حل:

ابتدا محاسبات لازم را به صورت جدول زیر انجام می‌دهیم.

x	y	xy	$x^۲$	$y^۲$
۱	۸	۸	۱	٦٤
۲	٦	۱۲	٤	۳٦
۳	۷	۲۱	۹	٤۹
٤	٦	۲٤	۱٦	۳٦
٥	٤	۲۰	۲٥	۱٦
۱٥	۳۱	۸٥	٥٥	۲۰۱

$$\bar{x}=\frac{۱٥}{٥}=۳ \qquad \bar{y}=\frac{۳۱}{٥}=٦/۲$$

$$\overline{x^۲}=\frac{٥٥}{٥}=۱۱ \qquad \overline{y^۲}=\frac{۲۰۱}{٥}=٤۰/۲ \qquad \overline{xy}=\frac{۸٥}{٥}=۱۷$$

اگر در فرمول ضریب همبستگی پیرسن قرار دهیم داریم:

$$r_{xy}=\frac{\overline{x.y}-\bar{x}.\bar{y}}{\sqrt{s_x^۲.s_y^۲}}=\frac{۱۷-۱۸/٦}{\sqrt{(۱۱-۹)(٤۰/۲-۳۸/٤٤)}}=\frac{-۱.٦}{۱/۸۸}=-۰/۸٥$$

با این نتیجه بین تعداد سیگار و ساعات خواب رابطه معکوس وجود دارد.

۱٤-٥- رگرسیون

سکه رابطه دو رو دارد، یک روی آن مقدار همبستگی بین دو متغیر است که دیدیم با ضریب همبستگی نمونه اندازه گیری می‌شود. روی دیگر این شکه، استفاده از این رابطه و یافتن معادله‌ای است که این زابطه را تبیین می‌کند. در صورتی که رابطه بین متغیرها معنی‌دار باشد، می‌توان آن را با الگوهای ریاضی بیان کرد. از آنجایی که ممکن است در یک مطالعه، متغیرهایی با مقیاس‌های متفاوت وجود داشته باشند، مجموعه‌ای از روش‌ها وجود دارند که به کمک آن‌ها، معادله ریاضی بین متغیرها را برآورد می‌کنند.

Scanned with CamScanner

۱۴-۶- هدف

در رگرسیون، معمولاً یکی از متغیرها مستقل و دیگری وابسته است. متغیر مستقل (x)، متغیری تأثیر گذار است و به متغیر تأثیر پذیر، متغیر وابسته (y) می‌گویند. در رگرسیون هدف آن است که با استفاده از معادله رگرسیون و به کمک یک نمونه تصادفی و بعضی روش‌های آماری، رفتار متغیر وابسته (y) را با آگاهی از مقادیر و مشخصات متغیرهای مستقل (x)، پیش بینی کنیم.

۱۴-۷- رگرسیون خطی

به معادله‌ای که رابطه بین دو متغیر مستقل و وابسته را نشان می‌دهد، **معادله رگرسیون** می‌گویند. معمولاً چنین رابطه‌ای ممکن است از نوع خطی یا غیر خطی باشد. اگر بتوان الگوی همبستگی را به صورت یک معادله خط نوشت، به آن **معادله خط رگرسیون** می‌گویند. در رگرسیون خطی ساده؛ در صورتی که y را متغیر وابسته و x را متغیر مستقل در نظر بگیریم، می‌توان معادله خط رگرسیون را به صورت خط راست زیر فرض کرد:

$$Y = aX + b$$

در معادله خط، a را شیب (یا ضریب زاویه) خط و b را عرض از مبداء خط می‌گوییم. در این معادله a و b را پارامترهای خط هم می‌گویند و مانند هر پارامتر دیگری در آمار، می‌توان آن‌ها را برآورد کرد.

۱۴-۸- برآورد پارامترهای خط رگرسیون

روش‌های متفاوتی برای برآورد کردن دقیق دو پارامتر a و b وجود دارد. یکی از این روش‌ها، روش «**کم‌ترین مربعات خطا**» است. در عمل، نمودار پراکنش نشان می‌دهد؛ بین نقاط روی خط راست و نقاط به دست آمده در نمونه، اختلاف وجود دارد. اساس

Scanned with CamScanner

روش کم‌ترین مربعات خطا، بر مینیمم کردن مجذور این اختلافات است. بدون هیچ توضیح اضافه‌ای، در این روش دو پارامتر a و b به صورت زیر برآورد می‌شوند:

$$\hat{a}=\frac{\overline{xy}-\bar{x}.\bar{y}}{\overline{x^{۲}}-\bar{x}^{۲}} \qquad \hat{b}=\bar{y}-\hat{a}\bar{x}$$

بنا بر این معادله برآورد شده خط رگرسیون به صورت زیر خواهد بود:

$$y=\hat{a}x+\hat{b}$$

مثال۲: در مثال قبل، با توجه به داده‌های جدول زیر، دیدیم که بین تعداد سیگار و ساعات خواب افراد سیگاری رابطه وجود دارد. مطلوب است این رابطه را به صورت یک معادله خط رگرسیون بنویسید.

(x = تعداد سیگار، متغیر مستقل و y = ساعات خواب، متغیر وابسته است)

x	y
۱	۸
۲	٦
۳	۷
٤	٦
۵	٤

حل:

با توجه به داده‌های فوق می‌توانیم پارامترهای خط رگرسیون را به صورت زیر برآورد کرده و معادله خط را بنویسیم.

$$\bar{x}=\frac{۱۵}{۵}=۳ \qquad \bar{y}=\frac{۳۱}{۵}=٦/۲$$

$$\overline{x^{۲}}=\frac{۵۵}{۵}=۱۱ \qquad \overline{xy}=\frac{۸۵}{۵}=۱۷$$

برآوردهای a و b به صورت زیر است:

Scanned with CamScanner