وزارت علوم، تحقیقات و فناوری

دانشگاه هرمزگان

مجتمع آموزش عالی میناب

گروه کشاورزی

جزوه آمار و احتمالات

تدوین: دکتر جمشید احسانی نیا

استادیار گروه کشاورزی

فصل اول

مقدمه

پیشگفتار

علم آمار یکی از رشته هایی است که به سرعت جای خود را در بین سایر علوم باز کرده است و تقریبا اکثر رشته ها از کشاورزی گرفته تا روانشناسی و علوم مهندسی نیازمند علم آمار می باشند. در واقع اهمیت این علم زمانی مشخص می گردد که در هنگام استفادهٔ عملی از مطالب تخصصی هر رشته به علوم آماری نیازمند می گردند و برای صحت مطالب از قوانین علم آمار استفاده می کنند. البته این علم تنها در اختیار متخصصان رشته های علمی نیست بلکه دولت ها و برنامه ریزان کلان هر کشور نیز برای برنامه ریزی کوتاه مدت و بلند مدت در تمامی زمینه ها از قوانین این علم استفاده میکنند که نشان از اهمیت علم آمار دارد.

مقدمه

امروزه در همه علوم از جمله علوم پایه، علوم انسانی، هنر، علوم پزشکی، علوم کشاورزی و فنی و مهندسی پیشرفتهای زیادی حاصل شده است. این پیشرفت ها در نتیجه بکارگیری روشهای صحیح و علمی حاصل شده اند. تحقیق در لغت به معنای جستجو برای یافتن حقیقت است اما از نظر علمی به مجموعه فعالیت هایی اطلاق می شود که بر اساس یک نظام فکری درست در راستای آگاهی از مجهولات پایه ریزی شده باشد. بشر از دیرباز به طرق مختلف از جمله بدیهی پنداری، اتکا به نظریات دیگران، پیروی محض و روش علمی درصدد تغییر وقایع اطراف خود بوده است. روش علمی موجه‌ترین راه حصول آگاهی می‌باشد که با مشاهده شروع می شود و به دنبال آن مسئله ای مطرح می گردد و فرض هائی بنا نهاده می شود. بنابراین در روش علمی مشاهدات در معرض آزمون قرار می گیرند و نتایج مورد قضاوت منطقی واقع می شوند و به صورت فرضیه و تئوری بیان می گردند. لذا در روش علمی پدیده های مطرح هستند که قابل مشاهده و تکرار شدنی باشند و هدف آن دسترسی به حقیقت نسبی بوده و دستورالعمل آن روش پژوهش است. پژوهش در پدیده های طبیعی نوعی استدلال و تحقیق استقرائی است که نیازمند آمار و احتمالات است.

اصطلاحات و مفاهیم آماری

برای هر تحقیق علمی نیاز به آمار داریم و سپس نتایج آماری را بعد از آنالیز تفسیر و استدلال کنیم و چندین روش استدلال وجود دارد:

روش های استدلال در آمار

استدلال استقرائی (Induction) : حکم از جزء به کل بوده و دو مشخصه دارد:

۱- نتیجه گیری بر مبنای اطلاعات حاصل از قسمتی از جامعه (نمونه) و تعمیم آن به کل جامعه.

۲- نتیجه حاصل از این نوع استدلال ها صدرصد صحیح نبوده و احتمالی است و وقتی می توانیم آن را مورد استفاده قرار دهیم که احتمال صحت نتیجه زیاد باشد. تعیین میزان این احتمال جز موضوعات علم آمار و احتمالات است و به وسیله آن قضاوت می شود که تا چه اندازه به نتیجه تحقیق بدست آمده اعتماد داشته باشیم.

استدلال قیاسی: حکم از کل به جزء (اصل به نتیجه). به عنوان مثال فرمول کلی مساحت دایره πR^2 است و می خواهیم مساحت دایره ای به شعاع یک متر را بدست آوریم. در اینجا چون اصل و قانون وجود دارد، اقدام به قیاس می کنیم.

معمولا در آمار از روش استقرایی استفاده می کنیم یعنی از یک آزمایش کوچک نتایج را به کل جامعه تعمیم می دهیم. بدیهی است استدلا ما ممکن است ۱۰۰٪ صحیح نباشد و لذا نتیجه را با حتمال بیان می کنیم. روش آماری استقرایی است اما احتمالا قیاسی است.. استدلال ممکن است ترکیبی از قیاسی و استقرایی باشد.

انواع قضاوت های آماری:

قضاوت عینی : استدلال و تحقیقی عینی است که بدون پیش داوری و بر اساس آزمایش و تجربه انجام شود. به عبارت دیگر استدلال و نتیجه آزمایش وابسته به عقاید شخصی نباشد.

قضاوت ذهنی: هرگاه بدون توجه به مشاهدات و تجربه و تنها به استناد هوش و سلیقه قضاوت گردد استدلال ذهنی خواهد بود.و مزیت قضاوت عینی بر قضاوت ذهنی علاوه بر صحت و اعتباری که دارد، در این است که نسبت به میزان اعتماد به صحت قضاوت پی برد.

علم و روش تحقیق علمی

علم (Science): به مجموعه دانش های سازمان یافته ای گفته می شود که به روش عینی درباره واقعیت ها و پدیده های طبیعی کسب شده اند. علم عبارت از کشف حقایق است. برای کشف یک حقیقت فرضی گرفته و آزمایشی انجام می گیرد اگر آزمایش صحت فرض را ثابت نمود فرض موردنظر قبول واقع می شود و در غیر

اینصورت فرض رد می گردد. اگر صحت فرضی چندین بار به اثبات رسد فرض به صورت تئوری در می آید. تئوری معمولا حقیقتی را بیان می کند ولی ممکن است که صددرصد صحیح نبوده و برای همیشه دوام نداشته باشد و در طول زمان قسمتی و یا حتی تمام تئوری تغییر نماید.

تحقیق علمی (Scientific research): یک تلاش و جدیت فکری است برای استدلال، استنباط و کشف حقیقت نسبی درباره موضوع و مسئله ای که مورد بررسی قرار می گیرد.

روش تحقیق (Research Method): در هر شاخه علمی عبارت است از مجموع روش ها و قواعدی که برای انجام تحقیق در آن علم بکار برده می شوند. برای تبیین ضرورت کاربرد آمار در پژوهش های علمی، لازم است که به مراحل اجرای یک پژوهش علمی نگاهی بیاندازیم. به طور خلاصه در اجرای یک تحقیق، شش مرحله کلی به شرح زیر وجود دارد:

۱- تعیین موضوع تحقیق و بیان هدف
۲- تعیین فرضیه های تحقیق بر اساس تجربیات و مطالعات محقق
۳- طرح تحقیق و انجام آزمایش
۴- جمع آوری داده ها
۵- تجزیه و تحلیل مشاهدات
۶- تفسیر و تعمیم نتایج

همانطوریکه دیده می شود آمار در بیشتر مراحل اجرای یک تحقیق (۳ الی ۶) به کار گرفته می شود.

میدان عمل علم آمار

وقوع هر پدیده معلول علل مختلفی است که می توان تعدادی از آنها را شناخت و بر حسب امکانات و وسایل موجود مقدار آنها را تا حدودی اندازه گرفت. ولی تعدادی از این علل را به دلیل اینکه وسایل موجود توانایی شناخت آنها را ندارند، نمی شناسیم و ناچاراً نمی توانیم اندازه گیری کنیم. بنابراین با کمک مطالعات و تجربیات علمی می توان وجود تعدادی از این روابط و عدم و جود تعداد دیگری را تا حدودی از طریق تحقیق کشف نمود. تحقیق موقعی دارای ارزش است که بتوان از موضوع تحت بررسی اطلاع کافی حاصل نمود و نسبت به آنها قضاوت

های صحیح و مبتنی بر مشاهده، تجربه و تکرار انجام داد. در چنین مواردی است که علم آمار به روش های علمی تحقیق کمک می کند. بنابراین آگاهی از اصول آماری ضروری است. علم آمار به روش های مختلفی دسته بندی می شود :

آمار تئوری یا ریاضی یک علم نظری است و بی نیاز از توجیه تجربی. آمار کاربردی: برای هر رشته علمی متداول است و با موضوع آن علم و روش های تحقیق آن بستگی و تناسب دارد برای مثال آمار کاربردی در علوم وابسته به موجودات زنده، بیومتری (Biometry)، علوم وابسته به اقتصاد، (Econometry) و علوم وابسته به جامعه شناسی، (Sociometry) را بوجود آورده است.

فصل دوم

مفاهیم و اصطلاحات آماری

تعریف آمار (Statistic)

علم آمار عبارت است از مجموعه روشها و تکنیک هایی که برای جمع آوری ، سازماندهی و تنظیم، تجزیه و تحلیل و تفسیر و تعمیم داده ها به کار می رود. به عنوان مثال می خواهیم در مورد قد یا وزن و استعداد دانشجویان اطلاعاتی کسب کنیم.

الف) جمع آوری داده ها (Collection of data):

مهمترین قسمت در یک روش آماری جمع آوری مشاهدات است. معمولا اگر جامعه کوچک باشد به کل جامعه دسترسی داریم و داده ها را ازاین جامعه متناهی جمع آوری می کنیم. اما عموما جوامع نامتناهی بوده و یا از لحاظ وقت و بودجه به ما اجازه نمی دهد که کل جامعه را دیتابرداری کنیم لذا از این جوامع یک نمونه برداری منطقی انجام داده و با این داده ها کار می کنیم.

روش های جمع آوری داده ها عبارتند از:

- استفاده از داده های از پیش تهیه شده.
- از طریق پرسش به صورت شفاهی، مصاحبه و یا کتبی.
- از طری مشاهده و ثبت وقایع.
- از طریق انجام آزمایش.

ب) سازماندهی و تنظیم (Organization of data)

داده های جمع آوری شده را در قالب های مناسب مثل جدول، نمودار و ... نمایش می دهیم که بعدا در مورد آنها بحث خواهد شد.

ج) تجزیه و تحلیل داده ها (Analysis of data)

از این داده های جمع آوری شده اطلاعات قابل اطمینان مثل میانگین، میانه، واریانس، دامنه تغییرات و یا شاخص های شکل توزیع مثل چولگی و کشیدگی را از داده ها استخراج می کنیم.

د) تفسیر و تعمیم داده ها. (Interpretation of data)

بعد از تجزیه و تحلیل داده ها بایستی آنها را تفسیر و متعاقب آن این داده ههای حاصل از نمونه گیری را به جامعه آماری تعمیم بدهیم.

شاخص های آماری

عبارتنداز مقادیر، کمیت ها و معیارهایی که برای نشان دادن نتیجه یک بررسی آماری به کار برده می شوند.

در بررسی های آماری، دو هدف اصلی و مشخص دنبال می شود:

هدف اول که در آمار توصیفی منظور می شود.

هدف دوم که در آمار استنباطی گنجانده می شود.

آمار توصیفی (Descriptive Statistics)

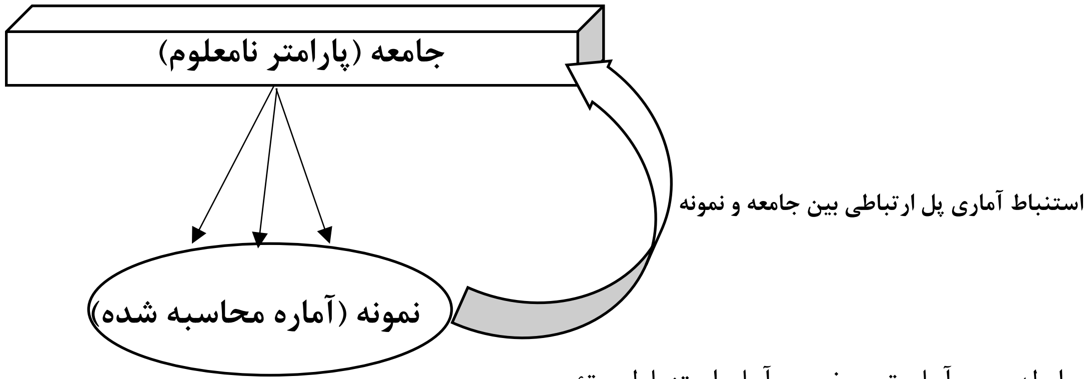
قسمتی از روش های آماری که تنها به توصیف و تجزیه و تحلیل گروه معینی، بدون تعمیم نتایج حاصله به گروه بزرگتر از آن محدود می گردد و هدف آن محاسبه پارامترهای جامعه است. توصیف داده ها ممکن است آخرین مرحله یک تجزیه و تحلیل ساده آماری باشد. برای توصیف داده ها از ابزارهایی مانند تهیه جدول توزیع فراوانی، رسم نمودار، محاسبه شاخص های مرکزی و پراکندگی (مانند میانگین، میانه، نما، دامنه تغییرات، انحراف متوسط از میانگین، انحراف معیار و واریانس) استفاده می کنند.

خصوصیات آمار توصیفی

۱. مناسب برای جوامع کوچک
۲. به توصیف خصوصیاتی از یک جامعه می پردازد بدون اینکه نتیجه گیری کند.
۳. تمامی اعضای جامعه مورد ارزیابی قرار می گیرد.
۴. نتایج آن فقط به همان جامعه محدود می شود و قابل تعمیم به جوامع بزرگ تر نیست.
۵. نتایج آن در قالب جدول و انواع نمودار ارائه می شود.

آمار استنباطی (Inferential Statistic)

به قسمتی از آمار که می تواند نتایج حاصل از تجزیه و تحلیل نمونه را به جامعه تعمیم دهد آمار استنباطی گویند. استنباط آماری پلی برای ارتباط بین جامعه و نمونه است (شکل۱).



رابطه بین آمار توصیفی و آمار استنباطی تئوری احتمالات (Probability theaory) است زیرا ما نتایج حاصل از نمونه گیری را با یک احتمال (و نه به طور قاطع) به جامعه تعمیم می دهیم.

جامعه آماری یا جمعیت:Statistical population

از نظر آماری به مجموعه افراد یا اشیائی اطلاق می شود که حداقل دارای یک خصوصیت یا ویژگی مشترک باشند. به همین علت آنها را با هم تحت مطالعه و بررسی قرار می دهیم. اعضای جامعه را عضو آماری گویند و حجم جامعه را با N نشان می دهند. اگر خصوصیتی از کل جامعه را بررسی کنیم مثل میانگین، به اینها پارامتر (Parameter) گویند. پارامتر جمعیت را معمولاً با حروف کوچک یونانی نشان می دهند. به عنوان مثال در یک جمعیت نرمال، میانگین و واریانس جمعیت را به ترتیب با μ و σ^2 نشان می دهند.

سوال: آیا امکان مطالعه روی یک جمعیت وجود دارد؟

اندازه گیری کلیه افراد جامعه به دلیل هزینه ها (costs) و زمان بر بودن (time) امکانپذیر نیست. بنابراین قسمت محدودی از جامعه به عنوان نمونه در نظر گرفته شده و تحقیق بر اساس آن انجام می شود.

انواع جامعه آماری:

متناهی (Finite): یعنی تک تک اعضای جامعه برای ما قابل دسترس و بررسی است.

نامتناهی (Infinite): که در این صورت ما به همه جامعه دسترسی نداریم و قطعا نمونه برداری یا sampling ضروری است.

نمونه (sample)

به هر بخش از جامعه آماری که مورد بررسی قرار می گیرد یک نمونه گفته می شود. مثال : کارکنان حوزه ریاست می توانند به عنوان یک نمونه از جامعه آماری مجتمع آموزش عالی میناب در نظر گرفته شوند. برای رسیدن به یک نتیجه معتبر و با اطمینان بالا نمونه باید نمایانگر جامعه موردنظر باشد یا به عبارتی نمونه طوری استخراج شود که تمامی صفات کلی جامعه را دربر داشته باشد. نمونه گیری ممکن است تصادفی (Random) و یا اریب (Biased) باشد (در اریب بودن سلیفه آمارگر یا محقق دخالت می کند ولی در نمونه های تصادفی همه عناصر احتمال مشابهی دارند تا رخ دهند). ما در آمار با نمونه های تصادفی سروکار داریم. مرز بزرگی و کوچکی نمونه عدد ۳۰ است. اگر ۳۰ $\leq$n باشد نمونه کوچک و اگر ۳۰ $<$n باشد نمونه بزرگ تلقی می شود.

کمیتی که از مطالعه نمونه حاصل می شود را آماره (Statistics) نامند که میزان آن از نمونه ای به نمونه دیگر متغیر است. معیار نمونه(آماره)را معمولاً با حروف کوچک لاتین نشان می دهند. به عنوان مثال در یک جمعیت نرمال، میانگین و واریانس نمونه را به ترتیب با $\bar{x}$و s^2نشان می دهند.

خصوصیات یک نمونه مناسب:

۱- نمونه باید به صورت تصادفی انتحاب شود تا تمامی افراد جامعه شانس مساوی در تشکیل آن داشته باشند.

۲- نمونه انتخابی معرف خوبی از جامعه باشد به عبارتی تمامی خصوصیات و ویژگی های جامعه را در بر داشته باشد.

۳- هر چه تعداد افراد در نمونه موردنظر بیشتر باشد برآورد مشخصات جامعه دقیقتر خواهد بود.

انواع روش های نمونه گیری:

نمونه‌گیری تصادفی (Random sampling)

نوعی نمونه گیری است که در آن هر یک از اعضای جامعه مورد مطالعه، برای انتخاب شدن، شانس مساوی دارند. اگر فهرستی از افراد جامعه در دسترس باشد از این روش می توان استفاده کرد. در این حالت افراد یا اشیا به طور تصادفی و به دو روش قرعه کشی و جدول اعداد تصادفی از میان عناصر جامعه انتخاب می‌شوند و هر یک از آنها از شانس مساوی برای انتخاب شدن برخوردارند. مثلاً انتخاب ۱۰ خودرو از بین ۱۰۰ خودرو تولید شده توسط کارخانه.

نمونه گیری منظم (Regular sampling)

در این روش یک نقطه از فهرست اشیا یا افراد انتخاب شده را به طور تصادفی انتخاب می کنیم و بعد از آن نمونه های موردنظر را به صورت منظم و پشت سرهم انتخاب می کنیم (انتخاب ۵ دانشجو پشت سرهم براساس شماره دانشجویی). این روش برای آن دسته از جوامع آماری که کد از پیش تعیین شده و مرتبی دارند، کاربرد فراوان دارد (شماره کارمندی-شماره دانشجویی). با مشخص شدن اولین عضو بقیه اعضا نیز مشخص می شوند.

نمونه گیری خوشه ای (Cluster sampling)

هر گاه جامعه موردنظر خیلی وسیع و گسترده باشد و فهرست افراد آن در اختیار نباشد، استفاده از روش های نمونه گیری قبلی دشوار بوده و مستلزم وقت زیاد، نیروی انسانی و هزینه های بیشتر است که به منظور صرفه جویی در این منابع از نمونه گیری خوشه ای استفاده می شود. در این روش بر خلاف سایر روش ها، واحد نمونه گیری فرد یا عضو نیست بلکه گروهی از افراد است که به صورت طبیعی در یک مجموعه یا خوشه شکل گرفته اند. به عنوان مثال می خواهیم دانشجویان دانشگاه های شهر بندرعباس را برای مطالعه انتخاب کنیم، ابتدا دانشگاههای شهر بندرعباس را در لیستی قرار می دهیم و به شیوه های تصادفی یکی از آنها را انتخاب می کنیم و دانشجویان نمونه را به یکی از شیوه های نمونه گیری (تصادفی یا منظم) از آن دانشگاه انتخاب می کنیم. اگر خوشه های انتخاب شده بزرگ باشند، آن خوشه را نیز به خوشه های کوچک تر تقسیم کرده و به صورت تصادفی یکی از آنها را انتخاب می کنیم و سپس افراد موردنظر را از آن خوشه بر می گزینیم. اگر دانشگاه هرمزگان انتخاب

شده باشد، دانشکده های دانشگاه هرمزگان را به عنوان خوشه های زیرمجموعه این دانشگاه درنظر گرفته و یکی را به تصادفی انتخاب و سپس افراد نمونه را از آن دانشکده به صورت تصادفی انتخاب می کنیم. روش فوق را نمونه گیری تصادفی خوشه ای چند مرحله ای گویند.

نمونه گیری مرحله ای (Stage sampling)

شکل گسترش یافته نمونه گیری خوشه ای است. در این حالت نمونه گیری از جامعه طی چندین مرحله صورت می گیرد. یعنی انتخاب نمونه از نمونه دیگر. به طور مثال چند سازمان یا ارگان از شهر انتخاب می کنیم سپس از بین هر سازمان چند واحد را معین می کنیم و عناصر نمونه را به صورت تصادفی بدست می آوریم.

سرشماری (Census)

اگر تمام افراد جامعه مورد مطالعه قرار گیرد به آن سرشماری گفته می شود. معمولا در سرشماری ها با مشکلاتی مواجه هستیم که مهمترین آنها عبارتند از:

- در دسترس نبودن تمام اعضای جامعه
- وقت گیر بودن دسترسی به تمام اعضای جامعه
- گران تمام شدن بررسی تمام اعضای جامعه

در مطالعهٔ جامعهٔ آماری با دو دسته صفت مواجه هستیم یک دسته صفاتی هستند که برای تمام افراد جامعه مشترک است این صفت را صفات مشخصهٔ جامعه می نامیم و دیگری صفات متغیر هستند که از یک فرد به فرد دیگر جامعه تغییر می کند. مثلاً وقتی می گوییم دانشجویان دانشگاه هرمزگان، صفات مشخصه آن عبارت از دانشجو بودن و متعلق بودن به دانشگاه هرمزگان آن را از سایر جامعه آماری جدا می سازد.

متغیر (Variable)

به ویژگی های مشترک یک جامعه آماری که از عضوی به عضو دیگر تغییر می کند و مورد پرسش یا اندازه گیری قرار می گیرد متغیر گفته می شود. به عبارت دیگر متغیر عامل یا صفتی است که ارزش آن از فردی با فرد دیگر تغییر می کند. مثل قد دانشجویان این کلاس. بنابراین داده ها مقادیر اندازه گیری شده متغیرها هستند. مثلا در جامعهٔ دانشجویان دانشگاه هرمزگان صفت مشترک دانشجوی دانشگاه هرمزگان بودن است. در حالیکه صفاتی

مانند جنس، وزن، قد، اهل کدام شهر بودن، رنگ چشم، تعداد افراد خانوار و غیره باهم متفاوت هستند و این صفات، صفات متغیر در جامعهٔ دانشجویان دانشگاه هرمزگان است.

انواع متغیر:

۱- **متغیر کمی** (quantitative variable) : متغیرهایی که قابل شمارش و یا اندازه گیری هستند و خود به دو دسته تقسیم می شوند:

الف) **کمی پیوسته** (Continous v.): هرگاه متغیری بتواند تمامی مقادیر موجود بین دو حد معین را شامل شود آن را متیغر پیوسته گویند. به عبارت دیگر متغیرهایی که می توانند اعداد اعشاری را نیز اختیار کنند. مانند: قد و وزن دانشجویان مجتمع آموزش عالی میناب.

ب-) **کمی گسسته یا ناپیوسته** (Discrete v.): اگر متغیری بتواند مقادیر شمارش پذیر(صحیح) را اختیار کند آن را متغیر گسسته گویند (بین دو عد دلخواه نتوانیم عددی را انتخاب کنیم). مانند: تعداد دانشجویان دانشگاه هرمزگان، تعداد فرزندان یک خانواده، جمعیت شهرها، تعداد کلاس ها و ...

نکته : متغیر پیوسته از طریق اندازه گیری و متغیر گسسته از طریق شمارش بدست می آید بنابراین متغیرها دارای واحد اندازه گیری و شمارش هستند.

۲- **متغیر کیفی** (Qualitative variable): متغیرهایی که قابل اندازه گیری و یا شمارش نیستند اما می توان آنها را طبقه بندی کرد که به دو گروه تقسیم می شود:

الف) اسمی (Nominal) : متغیرهایی که در آنها ترتیبی وجود ندارد مانند رنگ پوست، گروه خوانی افراد، مذهب، رنگ چشم.

ب) **رتبه ای یا ترتیبی** (Ordinal) : متغیرهایی هستند که در آنها نوعی ترتیب طبیعی وجود دارد مثل رضایت مندی از درس آمار، میزان هوش و مراحل زندگی.

داده های پیوسته، بعد داده های گسسته و سپس داده های رتبه ای و بعد داده های اسمی از لحاظ آنالیز کردن ارزششان کم و کمتر می شود و هر کدام از این ۴ نوع داده ها تجزیه و تحلیل خاص خود را می طلبد تا تفسیر ما صحیح باشد.

لازم به ذکر است برخی از محققین سعی می کنند متغیرهای کیفی و کمی را به یکدیگر تبدیل کنند که پیشنهاد می شود اگر مجبور نیستیم این کار را نکنیم مثلا قبلا رنگ سیب به رنگ های زرد، صورتی و قرمز تقسیم بندی می شد که یک متغیر اسمی است. امروزه دستگایی وجود دارد که شدت رنگ را به صورت کمی به ما می دهد که یک متغیر پیوسته است و ارزش آن از متغیر کیفی بیشتر است. یا در گیاهپزشکی میزان بیماری در ارقام مختلف گندم می تواند به صورت مقاوم، نیمه مقاوم، نیمه حساس یا حساس طبقه بندی شود که حالت کیفی رتبه ای دارد اما محقق می تواند به جای آن درصد بیماری روی گیاه را ثبت کند که در این حالت کمی خواهد بود.

داده یا مشاهدات (Data or observation)

نتایج خام یک آزمایش یا تحقیق را داده گویند و منابع داده ها عبارتند از: موارد ثبت شده روزانه، پرسش‌نامه و آزمایش و غیره. اگر داده ها را فرایند پردازش روی آنها انجام دهیم تبدیل به اطلاعات می شوند. داده های خام انبوهی از داده های عددی جمع آوری شده که هیچ نوع پردازشی روی آنها انجام نشده است و اگر روی آنها پردازش صورت گیرد تبدیل به اطلاعات (information) می شوند.

نماد جمع در آمار :

یکی از علائمی که زیاد در آمار استفاده می شود علامت زیگما یا سیگما ($\sum$) است و برای نشان دادن جمع چند عدد از این علامت استفاده می شود. متغیری که باید روی آن جمع صورت گیرد در طرف راست سیگما نوشته می شود و محدوده ای که لازم است جمع بسته شود به کمک زیر اندیس i و بالانویس n مشخص می شود.

$$\sum_{i=۱}^{n} X_i$$

اگر مقادیر $X_۱, X_۲,X_۳,X_۴,X_۵,X_۶, \dots,X_n$ در اختیار باشد حاصل جمع آنان با استفاده از زیگما می تواند به صورت زیر نمایش داده شود و فرمول آن به این صورت خوانده می شود: زیگمای $\boldsymbol{X_i}$ از ۱ تا n برابر است با ...

$$\sum_{i=۱}^{n} X_i = X_۱ + X_۲ + X_۳ + X_۴ + X_۵ + X_۶ + \cdots + X_n$$

$$\sum_{i=۱}^{n} X_i^۲ = X_۱^۲ + X_۲^۲ + X_۳^۲ + \cdots + X_n^۲$$

$$\sum_{i=۱}^{n} X_iY_i = X_۱Y_۱ + X_۲Y_۲ + \cdots + X_nY_n$$

مثال ۱ : با استفاده از مجموعه مقادیر $\{2.3.5.7.13\}$ =x، مطلوب است مقدار $\sum_1^5 X_i$ و $\sum_2^4 X_i$.

$$\sum_{۱}^{۵} X_i = ۲ + ۳ + ۵ + ۷ + ۱۳ = ۳۰$$

$$\sum_{۲}^{۴} X_i = ۳ + ۵ + ۷ = ۱۵$$

نکته ۱: در هنگام حل مسائل به حدود جمع که در بالا و پایین سیگما نوشته می شود بایستی توجه نمود.

مثال ۲ : اگر $X_۱=۳$, $X_۲=۹$, $X_۳=۱۱$ باشد مقادیر زیر را بدست آورید.

$$\sum_{۱}^{۳} X_i = X_۱ + X_۲ + X_۳ = ۳ + ۹ + ۱۱ = ۲۳$$

$$\sum_{۲}^{۳} X_i = ۹ + ۱۱ = ۲۰$$

نکته ۲: اگر حدود جمع برای سیگما مشخص نشده باشد بایستی تمامی اعداد با هم جمع شوند.

$$\sum X_i = ۳ + ۹ + ۱۱ = ۲۳$$

$$\sum X_i^۲ = ۳^۲ + ۹^۲ + ۱۱^۲ = ۱۱۵$$

قوانین جمع در علم آمار :

قانون ۱ : اگر C یک عدد ثابت باشد جمع حاصل ضرب یک عدد ثابت در متغیر برابر است با حاصل ضرب عدد ثابت در جمع آن متغیر.

$$\sum_{i}^{n} CX_i = C\sum_{i=۱}^{n} X_i$$

مثال ۳ : اگر $X_۱=۳$, $X_۲=۴$, $X_۳=۶$ و $C=۲$ باشد، حاصل عبارت زیر را بدست آورید.

$$\sum_{i=۱}^{۳} CX_i = ۲ \times ۳ + ۲ \times ۴ + ۲ \times ۶ = ۲۶ \qquad C\sum_{i=۱}^{۳} X_i = ۲(۳ + ۴ + ۶) = ۲۶$$

قانون ۲ : جمع یک عدد ثابت برابر است با حاصل ضرب n در عدد ثابت.

$$\sum_{i=۱}^{n} C = nc$$

نکته ۳ : n تعداد اعدادی است که در جمع شراکت دارند.

مثال ۴ : مقادیر عبارات زیر را بدست آورید.

$$\sum_{i=۴}^{۵} ۳ = ۲ \times ۳ = ۶ \qquad \sum_{i=۴}^{۷} ۵ = ۴ \times ۵ = ۲۰$$

$$\sum_{i=۳}^{۶} ۴ = ۴ \times ۴ = ۱۶ \qquad \sum_{i=۱}^{۴} ۳ = ۱۲$$

قانون ۳ : جمع یک جمله که خود شامل ۲ یا چند عبارت باشد برابر است با مجموع جمع تک تک عبارات.

$$\sum_{i=۱}^{n}(X_i + Y_i + Z_i) = \sum_{i=۱}^{n} X_i + \sum_{i=۱}^{n} Y_i + \sum_{i=۱}^{n} Z_i$$

مثال ۵ : اگر $X_۴=۸$, $X_۱=۹$, $X_۲=۶$, $X_۳=۵$ باشد حاصل عبارات زیر را پیدا کنید.

$$\sum_{i=۱}^{n}(X_i - ۵)^۲ = (X_۱ - ۵)^۲ + (X_۲ - ۵)^۲ + (X_۳ - ۵)^۲ + (X_۴ - ۵)^۲ + (X_۵ - ۵)^۲$$
$$= (۹ - ۵)^۲ + (۶ - ۵)^۲ + (۵ - ۵)^۲ + (۸ - ۵)^۲ + (۱۲ - ۵)^۲ = ۷۵$$

$$\sum_{i=۳}^{۵}(X_i - ۵)^۲ = (X_۳ - ۵)^۲ + (X_۴ - ۵)^۲ + (X_۵ - ۵)^۲$$
$$= (۵ - ۵)^۲ + (۸ - ۵)^۲ + (۱۲ - ۵)^۲ = ۵۸$$

$$(\sum_{i=۳}^{۴} X_i)^۲ = (۵ + ۸)^۲ = ۱۳^۲ = ۱۶۹$$

$$\sum_{i=۱}^{n}(X_i^۲ - ۲X_i) = (۹^۲ - ۲ \times ۹) + (۶^۲ - ۲ \times ۶) + (۵^۲ - ۲ \times ۵) = ۱۰۲$$

تمرین۱: حاصل عبارات زیر را بدست آورید؟

$$\sum_{i=۱}^{۵}(X_i+۲)$$

$$\sum_{i=۱}^{۳}(X_i+۳)^۲$$

تمرین ۲: چنانچه ۵ = n و ،$\sum_{i=1}^{5}X_i=20$ ،$\sum_{i=1}^{5}Y_i=17$ ،$\sum_{i=1}^{5}X_i^2=90$ ،$\sum_{i=1}^{5}X_iY_i=66$ باشد حاصل عبارات زیر را بدست آورید؟

$$\sum_{i=۱}^{۵}(۲X_i+۳)$$

$$\sum_{i=۱}^{۵}X_i(X_i-۱)$$

$$\sum_{i=۱}^{۵}(۲X_i+۱)(X_i-۳)$$

$$\sum_{i=۱}^{۵}(X_i-۵)Y_i$$

$$\sum_{i=۱}^{۵}(X_i+۵Y_i)$$

$$\sum_{i=۱}^{۵}X_i^۲-\frac{(\sum_{i=۱}^{۵}X_i)^۲}{۵}$$

تمرین ۳: اگر $X_1=4$ ،$X_2=9$ و $X_3=5$ باشد مقادیر زیر را محاسبه کنید.

الف- $\sum X^2$

ب- $\sqrt{\sum 2X}$

فصل سوم

توزیع فراوانی و طبقه بندی داده ها

تنظیم و خلاصه کردن داده ها و جدول فراوانی

در مطالعات علمی و یا برای انجام محاسبات آماری لازم است داده های حاصل از آزمایش یا تحقیق علمی که اغلب به صورت توده ای از اعداد خام هستند با روش صحیحی خلاصه و منظم شوند. بعد از جمع آوری داده ها که شکل نامنظمی دارند و تفسیر آنها مشکل خواهد بود، بایستی داده ها را طبقه بندی نموده و در جدولی بنام جدول توزیع فراوانی قرار دهیم. تنظیم داده های عددی در قالب جدول های فراوانی و ترسیم نمودار آنها از اولین مراحل تجزیه آماری می باشد.

شیوه های تنظیم و خلاصه کردن داده ها

- جداول آماری
- خلاصه کردن و توضیح داده ها به وسیله رسم نمودارها.

بعد از جمع آوری و ویرایش داده ها، طبقه بندی اولین گامی است که باید برداشته شود. مرتب کردن داده ها بر اساس ویژگی های مشترک را طبقه بندی گویند. با توجه به نوع داده ها (کیفی و کمی) می توان طبقه بندی مناسبی را انجام داد.

طبقه بندی عبارت است از تهیه جدولی که مقادیر متغیر مورد نظر را به تعداد افرادی که این کمیت ها یا کیفیت ها را معرفی کردند، مرتبط می کند. هر کدام از این تعداد را، فراوانی می نامند. جدول به دست آمده را، جدول توزیع آماری یا جدول توزیع فراوانی می گویند. بسته به نوع متغیرها(گسسته یا پیوسته) و اهداف تحقیق، به دو روش زیر می توان جدول توزیع فراوانی آنها را ترسیم کرد.

جدول آماری

در جدول آماری علاوه بر داده ها، تعداد و درصد تکرار آنها نمایش داده می شود. به طوری که با یک نگاه به جدول می توان اطلاعات مفیدی را در مورد پراکندگی و توزیع داده ها بدست آورد. معمولا داده های حاصل از یک آزمایش در قالب یک جدول ثبت می شوند. این داده ها به علت داشتن ظاهری نامنظم نمی توانند گویای

مطلبی در جامعه باشند. برای آنکه بتوان به آنها نظم بهتری داد در جدولهای مناسبی تنظیم میشوند. یکی از متداولترین جداول آماری جدول فراوانی یا جدول توزیع فراوانی است.

جدول توزیع فراوانی:

اگر داده ها و تعداد تکرارشان در یک جدول مرتب شوند، جدول حاصل جدول توزیع فراوانی نام دارد. لذا جدول توزیع فراوانی جدولی است که مقادیر متغییر مورد نظر را به تعداد افرادی که این کمیت ها یا کیفیت ها را دارا هستند، مرتبط می کند. دو نوع جدول توزیع فراوانی براساس نوع داده ها وجود دارد:

۱- جدول توزیع فراوانی برای داده های دسته بندی نشده

معمولا داده های گسسته و کیفی و فراوانی آنها در این نوع جدول ارائه می شود. یعنی در یک ستون خود داده (حروف) و در ستون روبروی آن فراوانی آنها ذکر می شود. عیب این جدول حجم بزرگ آن است زیرا ممکن است فاصله بین کوچکترین و بزرگترین اعداد جدول عددی بزرگ باشد. برای رفع این عیب داده ها را به کلاس هایی تقسیم می کنند که در قسمت ۲ آمده است.

۲- جدول توزیع فراوانی برای داده های دسته بندی شده

برای داده های پیوسته و نیز داده های گسسته ای که تنوع آنها زیاد باشد این جدول تشکیل می شود. اگر هدف این باشد تا اندازه های یک متغیر را در K طبقه، نمایش دهند، مقادیر آنها را در ستونی با عنوان X_i (اندازه ویژگی مورد مطالعه) مشخص کرده و در ستون دیگر فراوانی مطلق آنها را با F_i که (F_i تعداد X_i ها در بین کل داده هاست) و فراونی نسبی آنها با F_r نشان می دهند. به طور کلی در جدول توزیع فراوانی همواره موارد زیر را خواهیم داشت:

فراوانی یا فراوانی مطلق (Frequency/ Absolute Frequency):

تعداد دفعات تکرار هر داده در میان تمامی داده ها را فراوانی مطلق آن داده گویند که آن را با F_i نشان می دهند.

فراوانی نسبی (Relative Frequency):

نسبت فراوانی مطلق هر داده به حجم جامعه(کل داده ها) را فراوانی نسبی گویندکه آن را با f_i نشان می دهند و جمع آن برابر با ۱ خواهد بود.

$f_{i1}= \frac{F_1}{n}, \quad f_{i2}= \frac{F_2}{n}, \ldots, F_{ik}= \frac{F_k}{n}$

در اینصورت F_{i1} و F_{i2} و ... F_{ik} را فراوانی نسبی می نامیم.

درصد فراوانی نسبی ($F_r\%$):

از ضرب فراوانی نسبی هر طبقه در عدد۱۰۰ درصد فراوانی نسبی بدست می آید.

$F_r\% = F_r \times 100$

فراوانی تجمعی یا تجمعی مطلق (Cumulative Frequency):

اگر فراوانی هر داده (رده) را با فراوانی داده ها (رده های) قبل از آن جمع کنیم فراوانی تجمعی مطلق آن داده یا رده بدست می آیدکه آن را با F_{C_i} نشان می دهند.

نکته: فراوانی تجمعی طبقهٔ آخر برابر با تعداد کل داده ها می باشد.

فراوانی تجمعی نسبی: (Cumulative Relative Frequency)

فراوانی تجمعی نسبی هر داده یا طبقه (رده) از جمع فراوانی آن رده و فراوانی نسبی داده یا رده های قبل از آن حاصل می شود که آن را با f_{c_i} نشان می دهند.

درصد فراوانی تجمعی نسبی ($f_{c_i\%}$):

حاصلضرب فراوانی تجمعی نسبی هر طبقه در عدد ۱۰۰، درصد فراوانی تجمعی نسبی آن طبقه نام دارد. واضح است که درصد فراوانی نسبی تجمعی طبقهٔ آخر برابر ۱۰۰ است.

خواص جداول فراوانی:

اگر n نوع داده از نوع F_1 و F_2 و ... F_k داشته باشیم:

۱- همواره جمع فراوانی های مطلق برابر با تعداد کل داده ها است.

$$F_1 + F_2 + \ldots + F_k = \sum_{i=1}^{k} F_i = N$$

۲ - همواره جمع فراوانی های نسبی برابر با ۱ است.

$$\frac{F_1}{n} + \frac{F_2}{n} + \ldots + \frac{F_k}{n} = \frac{F_1+F_2+\cdots+F_k}{n} = \frac{\sum_{i=1}^{k} F_i}{n} = \frac{n}{n} = 1$$

۳ - فراوانی تجمعی نسبی طبقه آخر همواره برابر با ۱ است.

۴ - فراوانی تجمعی مطلق طبقه آخر برابر با تعداد کل داده ها است.

۵-فراوانی تجمعی مطلق و فراوانی مطلق طبقه اول باهم برابر است.

جدول فراوانی برای داده های گسسته:

مثال ۱ : گروههای خونی زیر را در یک جمعیت داریم. جدول توزیع فراوانی را تشکیل دهید.

B,A,AB,O,O,A,A,B,AB,B,O,A,B,O,AB,B,O,

ابتدا آنها را مرتب می کنیم:

A,A,A,A,A,A,A,B,B,B,B,B, AB,AB,AB,O,O,O,O,O

جدول فراوانی مربوط به گروه های خونی جامعه فوق چنین است:

Xi	F_i	f_i	$f_i\%$	F_{C_i}	f_{C_i}	$f_{C_i}\%$
A	۷	۰/۳۵	٪۳۵	۷	۰/۳۵	٪۳۵
B	۵	۰/۲۵	۲۵٪	۱۲	۰/۶۰	٪۶۰
AB	۳	۰/۱۵	٪۱۵	۱۵	۰/۷۵	٪۷۵
O	۵	۰/۲۵	٪۲۵	۲۰	۱	٪۱۰۰
جمع	۲۰	۱	۱۰۰			

خواص ۱ خواص ۲ خواص ٤ خواص ۳

مثال ۲ : در یک جدول فراوانی با ۴۰ داده اگر درصد فراوانی نسبی طبقه ای ۳۰ باشد فراوانی نسبی و فراوانی مطلق آن طبقه را بیابید.

$$f_c\% = f_c \times ۱۰۰ \Longrightarrow ۳۰ = f_c \times ۱۰۰ \Longrightarrow f_c = \frac{۳۰}{۱۰۰} = ۰.۳$$

$$f_c = \frac{F_i}{n} \Longrightarrow ۰.۳ = \frac{F_i}{۱۰} \Longrightarrow F_i = ۰.۳ \times ۴۰ = ۱۲$$

مثال ۳: جدول فراونی برای تعدادی از داده ها به صورت زیر می باشد. جاهای خالی را با عدد مناسب پر کنید.

X_i	F_i	f_i	$f_i\%$	F_{C_i}
۲	؟			۳
۵	؟		٪۳۵	
۷	؟	۰/۲۰		
۹	؟		٪۲۵	
۱۲	؟			۲۰

حل: چون فراوانی تجمعی طبقه آخر برابر ۲۰ است پس تعداد داده ها برابر با ۲۰ است. در طبقۀ اول فراوانی تجمعی و فراوانی مطلق با هم برابر است پس داریم: F۱=۳ و $f_۱ = ۳/۲۰ = ۰.۱۵$ و $\%f_۱ = ۰.۱۵ \times ۱۰۰ = ٪۱۵$ خواهد بود. در طبقه دوم چون $\%f_۲ = ٪۳۵$ بنابراین $f_۲ = ۰.۳۵$ و $F_۲ = ۰.۳۵ \times ۲۰ = ۷$ خواهد بود. در طبقه سوم $f_۳ = ۰.۲۰$ را داریم پس $F_۳ = ۰.۲۰ \times ۲۰ = ۴$ و $\%f_۳ = ۰.۲۰ \times ۱۰۰ = ٪۲۰$. در طبقه چهارم چون $\%f_۴ = ٪۲۵$ بنابراین $f_۴ = ۰.۲۵$ و $F_۴ = ۰.۲۵ \times ۲۰ = ۵$ خواهد بود. در طبقه پنجم چون فراوانی مطلق برابر با کل داده ها می باشد لذا $F_۵ = N - ۱۹ = ۱$ و $f_i = ۱/۲۰ = ۰.۰۵$ و $\%f_۵ = ۰.۰۵ \times ۱۰۰ = ٪۵$ خواهد بود.

X_i	F_i	f_i	$f_i\%$	F_{C_i}
۲	۳	۰/۱۵	٪۱۵	۳
۵	۷	۰/۳۵	٪۳۵	۱۰
۷	۴	۰/۲۰	٪۲۰	۱۴
۹	۵	۰/۲۵	٪۲۵	۱۹
۱۲	۱	۰/۰۵	٪۵	۲۰
جمع کل	۲۰	۱	۱۰۰	

جدول فراوانی برای داده های پیوسته

زمانیکه با داده های پیوسته سروکار داریم ابتدا داده ها را به رده های با طول مساوی تقسیم می کنیم و فراوانی داده ها را در هر رده بدست می آوریم. این جدول نیز همانند جدول دادههای گسسته است یعنی شامل فراوانی مطلق، فراوانی نسبی و فراوانی تجمعی می باشد ولی چون اعداد خام در این حالات می توانند اعداد اعشاری نیز داشته باشند ستونی به نام حدود طبقه به این جدول اضافه می شود. برای رسم این ستون به شرح زیر اقدام می کنیم:

۱) **تعیین تعداد رده ها یا طبقه ها (K):**

تعیین تعداد طبقات مناسب در مورد هر مجموعه از داده ها بستگی به تشخیص و ابتکار و تجربه فرد دارد این تعداد با درنظر گرفتن نوع داده ها، تعداد اقلام نمونه و پراکندگی آنها و سلیقه محق انتخاب میگردد. در تعیین

تعداد طبقات باید دقت کافی مبذول داشت زیرا اگر تعداد طبقات فوق العاده کم انتخاب گردد اندازه طبقات بزرگ شده و مقدار زیادی اطلاعات از دست می رود برعکس هر قدر تعداد طبقات توزیع زیادتر باشد جزئیات اقلام بهتر مشخص میگردد ولی گسترش طبقات ضمن اینکه محاسبه و نتیجه گیری را مشکل می سازد موجب می شود مقادیر در نمودارها و دیاگرام ها به صورت نامنظم و نابهنجار درآیند و تنظیم جدول مشکل شود. از این رو بهتر است تعداد طبقات با تعداد داده ها متناسب باشد. به طورکلی پیشنهاد می شود تعداد طبقات بین ۵ تا ۲۵ باشد. دو روش برای تعیین تعداد طبقات هست که در این جا به آنها اشاره می کنیم.

الف) یکی از روش های تعیین تعداد طبقات استفاده از فرمولی به نام قاعدهٔ استورجس (Sturges) به شرح زیر می باشد.

$$K = ۱ + ۳{,}۳۲۲ log_{۱۰}^{n}$$

که n تعداد کل داده ها می باشد. عدد محاسبه شده برای k لزوما صحیح نیست ولی نزدیکترین عدد صحیح به k را انتخاب می کنیم. معمولا عدد اعشاری بدست می آید که باید با تقریب اضافی آن را گرد کنیم.

مثال۴: اگر در آزمایشی ۱۲۶ داده جمع آوری کرده باشیم براساس فرمول استورجس چند طبقه خواهیم داشت؟

$$K = ۱ + ۳{,}۳۲۲ \log_{۱۰}^{۱۲۶} = ۱ + ۳.۳۲۲ \times ۲.۱ = ۶.۴۳ \cong ۷$$

ب- روش دیگر برای تعیین تعداد طبقات و تعداد داده ها برابر جدول زیر است:

تعداد طبقات	تعداد داده ها	تعداد طبقات	تعداد داده ها
۲	۵	۱۰	۱۰۰
۴	۱۰	۱۲	۲۰۰
۶	۲۵	۱۵	۵۰۰
۸	۵۰	۳۰	۱۰۰۰

۲) تعیین میزان تغییر پذیری داده ها

میزان تغییرپذیری داده ها با استفاده از رابطه زیر بدست می آوریم:

$$\text{میزان تغییرپذیری داده ها} = \frac{\text{واحد گرد شده داده ها}}{۲}$$

۳) تعیین حد بالا و پایین هر رده یا طبقه

هر طبقه در جدول توزیع فراوانی با یک داده شروع می شود و به یک داده دیگر ختم می شود که داده شروع را حد یا کران پایین طبقه یا رده و داده بالا در همان طبقه را حد یا کران بالای طبقه گویند. برای نوشتن دیگر طبقات کافی است c یا فاصله طبقات را به حد پایین یا بالای هر طبقه اضافه کنیم تا حدود طبقه بعدی بدست آید.

نکته: در جدول توزیع داده های گسسته حدود طبقات اعداد صحیح هستند و با حد طبقه بعدی خود گسستگی دارند. در حالیکه اگر جدول توزیع داده ها به صورت پیوسته باشد حدود طبقات اعداد حقیقی بوده یعنی طبقات به صورت پیوسته خواهند بود. مثال جدول زیر:

پیوسته	گسسته
۲-۶	۲-۶
۶-۱۰	۷-۱۱

مقدار واقعی کوچکترین و بزرگترین داده یعنی min و max را به صورت زیر بدست می آوریم:

min = کوچکترین داده - میزان تغییرپذیری داده ها

max = بزرگترین داده + میزان تغییرپذیری داده ها

۴) تعیین دامنه تغییرات داده ها (R)

به تفاضل بزرگترین دادهٔ آماری از کوچکترین دادهٔ آماری دامنهٔ تغییرات می گوییم.

$$R = \max - \min$$

۵) تعیین طول یا فاصله رده ها یا طبقات (C)

$$C = \frac{R}{K}$$

مثال ۵: اگر کوچکترین داده برابر ۵ و بزرگترین دادهٔ آماری برابر ۱۴۵ و تعداد داده ها ۸۰ باشد فاصلهٔ طبقات را بیابید.

$K = ۱+ ۳,۳۲۲log_{۱۰}^{۷۰} = ۱ + ۳.۲۲ \times ۱.۸۵ = ۱ + ۵.۹۴ = ۱ + ۶ = ۷$

$R = \max - \min = ۱۴۵-۵=۱۴۰$

$$C = \frac{R}{K} \Longrightarrow C = \frac{۱۴۰}{۷} = ۲۰$$

۶) تعیین نماینده طبقات

نماینده هر طبقه میانگین حد پایین و حد بالا همان طبقه است.

مثال ۶: داده های زیر طول عمر ۴۰ عدد لامپ را نشان می دهند که به نزدیکترین عدد صحیح گرد شده اند. جدول توزیع فراوانی برای این دادها را بدست آورید.

۱۵،۱۳،۷،۸،۲۰،۱۷،۱۵،۱۶،۱۹،۱۳،۱۴،۱۰،۱۱،۱۲،۲۱،۱۸،۱۱،۲۱،۱۷،۱۶،۸،۲۲،۲۳،۱۷،۱۴،۱۳،۲۰،۱۵،۱۲،۹،۱۱

۱۹،۱۸،۹,۱۱،۱۲،۲۲،۱۴،۱۶،۱۷,

$K = ۱+ ۳,۳۲۲log_{۱۰}^{۴۰} = ۶.۳۲۲ \cong ۷$

$$\text{میزان تغییرپذیری داده ها} = \frac{\text{واحد گرد شده داده ها}}{۲} = \frac{۱}{۲} = ۰.۵$$

min = کوچکترین داده - میزان تغییرپذیری داده ها = ۷-۰,۵ = ۶,۵

max = بزرگترین داده + میزان تغییرپذیری داده ها = ۲۳+۰,۵ = ۲۳.۵

$R = \max - \min = ۲۳.۵-۶,۵ = ۱۷$

$$C = \frac{R}{K} = \frac{۱۷}{۷} = ۲.۴ \cong ۳$$

حدود رده ها	X_i مرکز دسته	F_i فراوانی مطلق	f_i فراوانی نسبی	F_{C_i} فراوانی تجمعی مطلق	f_{c_i} فراوانی تجمعی نسبی
۶,۵-۹,۵	۸	۵	۰,۱۲۵	۵	۰,۱۲۵
۹,۵-۱۲,۵	۱۱	۸	۰,۲۰	۱۳	۰,۳۲۵
۱۲,۵-۱۵,۵	۱۴	۹	۰,۲۲۵	۲۲	۰,۵۵۰
۱۵,۵-۱۸,۵	۱۷	۹	۰,۲۲۵	۳۱	۰,۷۷۵
۱۸,۵-۲۱,۵	۲۰	۶	۰,۱۵	۳۷	۰,۹۲۵
۲۱,۵-۲۴,۵	۲۳	۳	۰,۰۷۵	۴۰	۱
جمع		۴۰	۱		

مثال ۷: در یک امتحان تستی که شامل۳۰ سوال بوده است یک نمونه ۲۰ نفری از دانش آموزان نمرات زیر را کسب کرده اند جدول فراوانی داده ها را در ۷ دسته طبقه بندی کنید.

۱, ۵, ۲۷, ۱۱, ۱۶, ۱۴, ۲۵, ۲۴, ۱۴, ۶, ۱۰, ۲۹, ۱۳, ۱۱, ۱۷, ۱۸, ۱۹, ۱۲, ۹, ۳.

حدود طبقات	X_i مرکز دسته	F_i فراوانی مطلق	f_i فراوانی نسبی	F_{C_i} فراوانی تجمعی مطلق	f_{c_i} فراوانی تجمعی نسبی
۱-۴,۵	۲,۵	۲	۰,۱۰	۲	۰,۱۰
۴,۵-۸,۵	۶,۵	۲	۰.۱۰	۴	۰.۲۰
۸,۵-۱۲,۵	۱۰,۵	۵	۰,۲۵	۹	۰.۴۵
۱۲,۵-۱۶,۵	۱۴,۵	۴	۰,۲۰	۱۳	۰.۶۵
۱۶,۵-۲۰.۵	۱۸,۵	۳	۰,۱۵	۱۶	۰.۸۰
۲۰.۵-۲۴,۵	۲۴,۵	۱	۰,۰۵	۱۷	۰,۸۵
۲۴,۵-۲۸.۵	۲۶,۵	۲	۰,۱۰	۱۹	۰,۹۵
۲۸,۵-۳۲.۵	۳۰,۵	۱	۰,۰۵	۲۰	۱
جمع		۲۰	۱		

$$\text{میزان تغییرپذیری داده ها} = \frac{\text{واحد گرد شده داده ها}}{۲} = \frac{۱}{۲} = ۰.۵$$

min = کوچکترین داده - میزان تغییرپذیری داده ها = ۱-۰,۵ = ۰,۵

max = بزرگترین داده + میزان تغییرپذیری داده ها = ۲۹+۰,۵ = ۲۹,۵

R = max – min = ۲۹,۵-۰.۵ = ۲۹

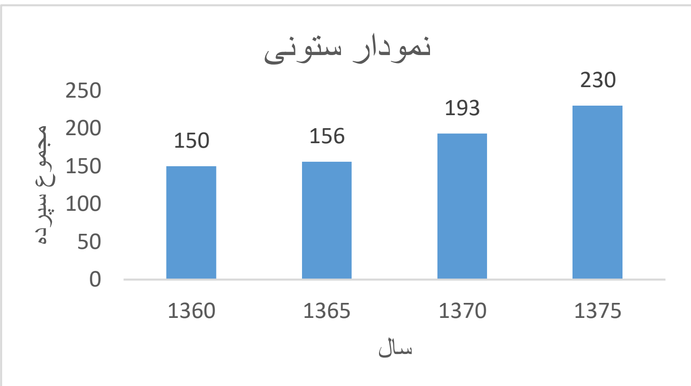
$$C = \frac{R}{K} = \frac{۲۹}{۷} = ۴.۱۴ \cong ۴$$

نمایش داده ها به صورت نمودار:

نمایش تصویری اعداد به منظور مطالعه تغییرات یک متغیر نمودار یا گراف نام دارد. برای نمایش توزیع های فراوانی، اغلب از نمودار استفاده می شود. نمودارها کمک می کنند تا تصویر توزیع به سهولت دیده شود. به کارگیری این روش ها در گزارش نویسی، یکی از راه های مفید برای انتقال موضوع گزارش به خواننده در کمترین زمان ممکن و با ساد ه ترین بیان است. نمودارها اطلاعات بیشتری نسبت به مشاهدات مرتب شده در جدول ارائه نمی نمایند اما حقایق مهم را به صورت واضح تر و قاطع تر نمایش می دهند.

مثال: مجموع سپرده های کلیه بانک های ایران از سال ۱۳۶۰ تا ۱۳۷۵ در فواصل ۵ ساله براساس جدول چنین است:

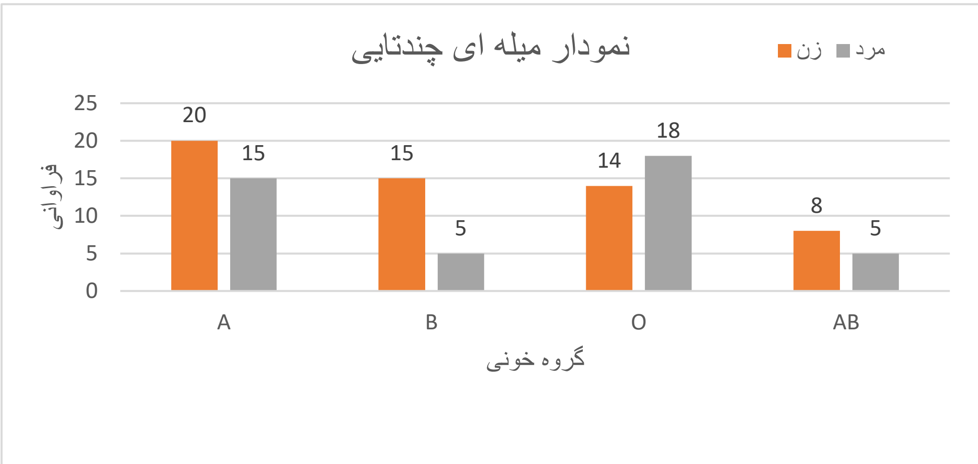
سال	مجموع سپرده ها(میلیون دلار)
۱۳۶۰	۱۵۰
۱۳۶۵	۱۵۶
۱۳۷۰	۱۹۳
۱۳۷۵	۲۳۰



نمودارهای آماری برای داده های گسسته:

۱- **نمودار میله ای یا ستونی** (Bar chart)

برای رسم نمودار میله ای بر روی محور افقی (طول ها) صفت متغیر مربوطه و بر روی محور عمودی (عرض ها) فراونی ها قرای می گیرند. اگر بیش از یک صفت کیفی مدنظر باشد از نمودار میله ای چندتایی استفاده می شود. از نمودارهای Bar Graph به منظور ترسیم داده های دسته ای یا کیفی و همچنین برای مقایسه متغیرها استفاده می شود.



۲- نمودار دایره ای Circle graph

این نوع نمودار برای نشان دادن طرز تقسیم یک مقدار کل به اجزای تشکیل دهنده اش بکار می رود. اجزا ممکن است بر حسب درصد و یا به صورت جزهایی از دایره که زاویه مرکزی آنها درصدی از ۳۶۰ درجه است نشان داده شود. به طور کلی از نمودار دایره ای برای نشان دادن متغیرهای کیفی مثل جنسیت، رشته تحصیلی و غیره استفاده می شود.

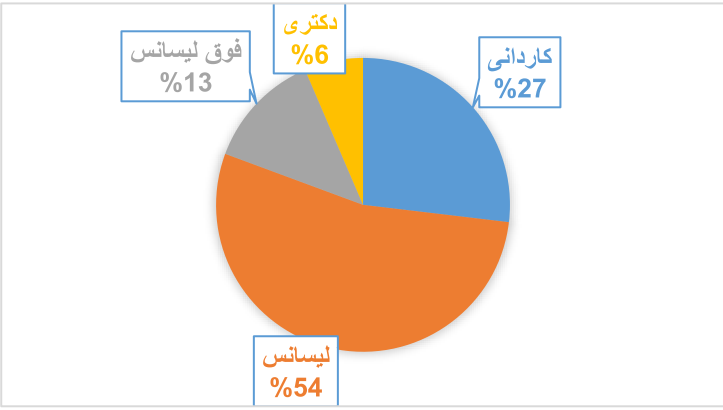
$$\text{اندازه قطاع یا زاویه} = \frac{F_i}{n} \times ۳۶۰^\circ$$

مثال ۴: تعداد کارمندان یک وزارتخانه براساس مقاطع تحصیلی به صورت زیر بوده است. نمودار دایره ای را با رسم کنید؟

۶۰۰: دکتری ۱۲۰۰: فوق لیسانس ۲۵۰۰: درصد کاردانی ۵۰۰۰: لیسانس

$$\frac{۵۰۰۰}{۹۳۰۰} \times ۳۶۰^\circ = ۱۹۴^\circ \qquad \frac{۲۵۰۰}{۹۳۰۰} \times ۳۶۰^\circ = ۹۴^\circ$$

$$\frac{۶۰۰}{۹۳۰۰} \times ۳۶۰^\circ = ۲۵^\circ \qquad \frac{۱۲۰۰}{۹۳۰۰} \times ۳۶۰^\circ = ۴۷^\circ$$

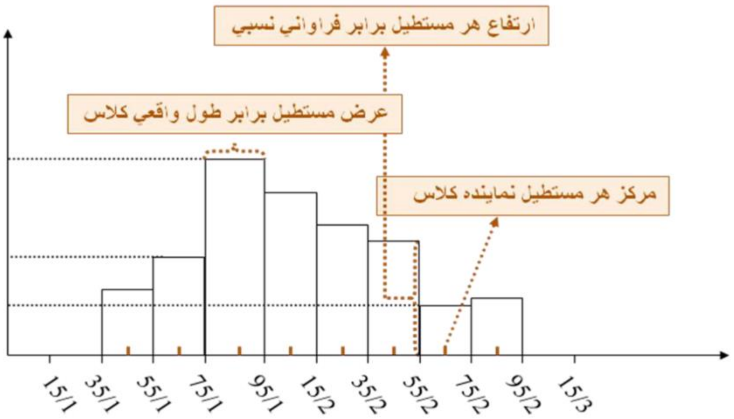


نمودارهای آماری برای داده های پیوسته

۱) **نمودار ستونی یا هیستوگرام :**

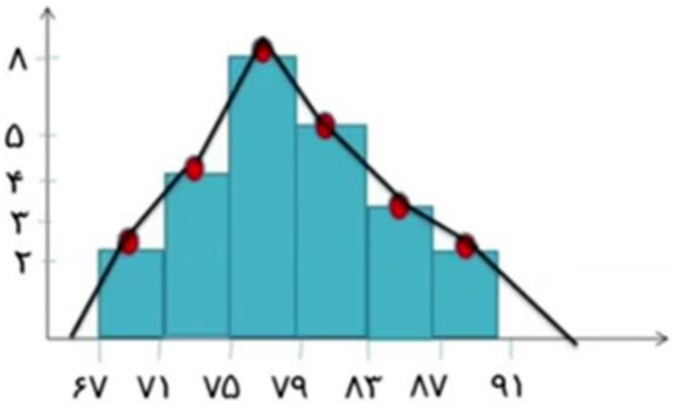
نمودار هیستوگرام به منظور نمایش توزیع متغیرهای کمی مورد استفاده قرار می گیرد. در این نوع نمودار محور افقی یا Xها نمایش دهنده فاصله رده ای (کرانه های طبقات) و محور عمودی نشان دهنده فراوانی است و به ازای هر رده مستطیلی که عرض آن برابر با طول واقعی طبقه و ارتفاع آن معادل فراوانی نسبی آن رده است رسم می

شود. مرکز مستطیل نماینده طبقه یا رده است. نمودار هیستوگرام همانند نمودار ستونی است و تنها اختلافی که بین این دو وجود دارد، نمایش ستون هاست. در هیستوگرام ستون ها به یکدیگر چسبیده اند. اتصال ستون ها در هیستوگرام موجب می شود تا این نمودار وسیله مناسبی برای نمایش داده های ناشی از اجرای متغیرهای پیوسته باشد.



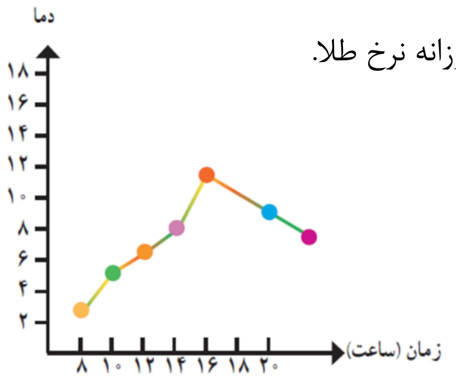
۲) نمودار چندبر یا چندضلعی (Polygon):

برای داده های پیوسته است. اگر مرکز طبقات را که وسط مستطیل های هیستوگرام است را به یکدیگر وصل کنیم، نمودار چندبر بدست می آید. اما ابتدا و انتهای آن حتما باید صفر باشد یعنی به طبقات ماقبل و بعد از داده ها حتما باید وصل شود. مساحت زیر منحنی polygon برابر مساحت هیستوگرام می باشد.



۳) نمودار خط شکسته:

از نمودار خط شکسته معمولا برای نمایش موضوعاتی استفاده می‌شود که در آن ها تغییرات داده ها اهمیت دارد مثلا:تغییرات طول یک گیاه در هر دو روز یک بار یا تغییر روزانه نرخ طلا.



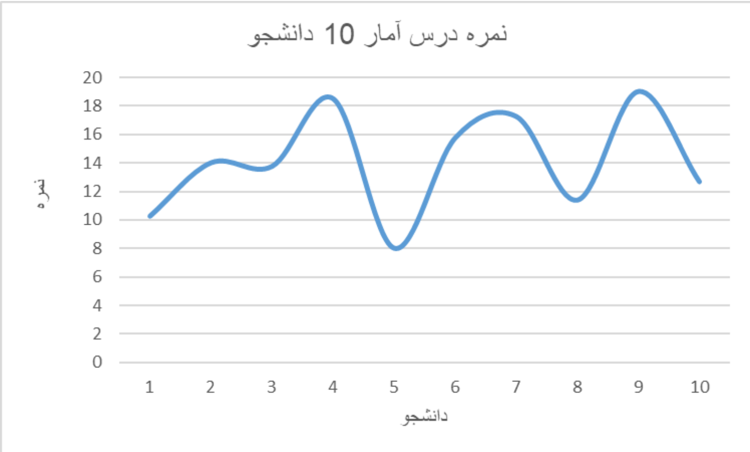
تغییرات دما در یک روز

۴) نمودار خط منحنی

فرم تغییر یافته نمودار خط شکسته است که بجای یکسری خط شکسته یک منحنی صاف نقاط را به هم وصل می نماید و بیشتر برای نشان دادن تغییرات متغیرهای پیوسته بکار می رود. در صورتیکه تعداد داده ها زیاد باشد و طول رده ها کوچک، تعداد رده ها زیاد می شود و در نتیجه اضلاع چندبر فراوانی افزایش یافته و در نهایت یک منحنی ایجاد می شود.

مثال ۵ : نمرات آمار ۱۰ دانشجو به شرح زیر است. نمودار خط منحنی نمرات این دانشجویان را رسم کنید.

نمره	دانشجو
۱۰,۲۵	۱
۱٤	۲
۱۳,۷۵	۳
۱۸,٥	٤
۸	٥
۱٥,۷٥	٦
۱۷,۲٥	۷
۱۱,۳۸	۸
۱۹	۹
۱۲,٦۷	۱۰



۵) نمودار ساقه و برگ (Stem-leaf plot)

روش دیگری که برای توصیف داده های کمی به کار برده می شود استفاده از نمودار ساقه و برگ می باشد. رسم نمودار ساقه و برگ به دلیل از دست ندادن اطلاعات به نمودارهای فراوانی ترجیح داده می شود. برای رسم نمودار ساقه و برگ، هر مقدار کمی بدست آمده را به دو بخش تقسیم نموده، یک بخش را ساقه و بخش دیگر را برگ می نامیم. آنگاه برای هر ساقه برگهای مربوطه را به طور جداگانه در مقابل آن می نویسیم. شکل حاصل را نمودار ساقه و برگ داده ها می نامند.

یک یا چند رقم از ارقام سمت چپ هر داده به عنوان ساقه در نظر گرفته می شود. رقمهای باقی مانده به عنوان برگ در نظر گرفته می شود. اعداد مربوط به قسمت ساقه بدون تکرار از کم به زیاد و از بالا به پایین نوشته می شود در جلو هر رقم ساقه قسمت های باقی مانده به صورت صعودی و البته با تکرار و با فاصله یکنواخت نوشته می شود.

مثال ۶ : فرض کنید که اعداد زیر را داشته باشیم. نمودار ساقه و برگ این مجموعه داده ها را رسم کنید.

٦,٩,١٢,١٢,١٤,١٤,١٤,١٥,١٦,١٨,١٨,١٨,١٨,١٩,١٩,٢٠,٢٠,٢١,٢١,٢١,,٢٢,٢٢,٢٢,٢٣,٢٨,
٢٨,٢٩,٣٢,٣٣,٣٣,٣٧

شاخه	برگ	فراوانی
۰	۶۹	۲
۱	۲۲۴۴۴۵۶۸۸۸۸۹۹	۱۳
۲	۰۰۱۱۱۲۲۲۲۳۸۸۹	۱۲
۳	۲۳۳۷	۴

تمرین ۱: صنعت گری چهار نوع قطعه A, B, C, D تولید می کند اگر در یک روز ۲۰ قطعه از این قطعات را به صورت زیر تولید کرده باشد جدول فراوانی را رسم کنید و مشخص کنید چند درصد قطعات تولیدی از نوع A و B است؟

B,C,C,A,D,C,C,B,D,C,A,C,D,C,B,C,C,B,D,D

تمرین ۲: اگر از ۱۰۰ کارمند یک اداره، ۵۰ نفر دارای حقوق کم، ۳۰ نفر دارای حقوق متوسط و ۲۰ نفر دارای حقوق زیاد باشند، فراوانی نسبی، فراوانی تجمی مطلق و فراوانی تجمعی نسبی را بدست آورید.

تمرین۳: در یک اداره ۱۲۰۰ کارمند براساس رضایت مندی از مدیرعامل آماربرداری شدند که نتیجه آن در جدول زیر آمده است. نمودار دایره ای را ترسیم کنید.

میزان رضایت	تعداد کارمند
ناراضی	۲۹۶
بی اعتنا	۳۶۰
راضی	۵۴۴

تمرین ۴: در جدول فراوانی ۸۰ داده آماری، فراوانی تجمعی در طبقات سوم و چهارم به ترتیب ۲۵ و ۳۴ است، درصد فراوانی نسبی طبقه چهارم را محاسبه کنید؟

طبقه	فراوانی مطلق	فراوانی تجمعی مطلق
اول		
دوم		
سوم		۲۵
چهارم		۳۴

تمرین ۵: وزن های ۴۰ قالب کره به صورت زیر می باشند. یک جدول فراوانی برای این داده ها تشکیل دهید و نمودار هیستوگرام و چندبر آن را رسم کنید.

۵۲،۴۷،۲۶،۳۱،۳۵،۳۶،۲۹،۳۰،۲۴،۳۸،۳۰،۲۶،۴۷،۵۰،۳۲،۳۵،۳۶،۴۷،۳۰،۴۵،۵۱،۳۴،۳۵،۲۳،۳۴،۴۱،۳۷،۴۳،۳۸،۴۰،۳۷،۳۱،۴۶،۴۲،

۴۱،۳۴،۳۳،۴۰،۲۱،۴۳

تمرین ۶: نمودار دایره ای زیر سهم وزنی ترکیبات تشکیل دهنده ی یک بسته غذایی کنسرو شده را نشان می دهد:

الف- چه کسری از این ترکیبات چربی است ؟ چه کسری کربو هیدرات است ؟

ب- چند گرم پروتئین در یک بسته ۶۶ گرمی از این محصول وجود دارد؟

تمرین ۷. در یک مرکز تلفن زمان انتظار مردان برای تلفن زدن (بر حسب دقیقه) به صورت زیر است. نمودار شاخه و برگ را ترسیم کنید.

.	۰	۵	۱۵	۰	۰	۱۳	۱۳
۰	۱۸	۲۵	۰	۱۹	۱۰	۱۹	۵
۱۵	۹	۱۸	۱۷	۱۷	۱۷	۷	۱۴
۳۵	۰	۰	۲۱	۰	۲۰	۱۸	۸
۰	۰	۱۰	۰	۴	۰	۴	۲۵
۰	۰	۱۰	۰	۴	۰	۱۴	۳۵
۰	۰	۴	۲۳	۲۷			

فصل چهارم:

شاخص های مرکزی

شاخص های آماری

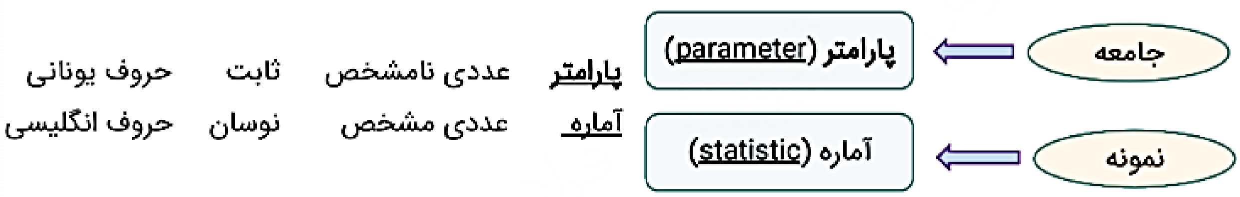
مقادیر، معیارها و کمیت هایی که برای نشان دادن نتیجه یک بررسی آماری مربوط به جامعه یا نمونه مورد استفاده قرار می گیرند. مانند میانگین، درصد، انحراف معیار و ... .

به شاخص های آماری در یک جامعه **پارامتر** گویند. این شاخص ها (پارامترها) از طریق نمونه گیری از جامعه تخمین زده می شود که در اینصورت به آنها آماره گویند.

پارامتر: عددی که خصوصیتی از جامعه را بیان می کند. مانند میانیگن جامعه ($\boldsymbol{\mu}$) و واریانس جامعه ($\boldsymbol{\delta}^2$).

آماره: در مورد نمونه استفاده می شود و خصوصیتی از آن را بررسی می کند. مانند میانگین نمونه ($\bar{\mathbf{X}}$) و واریانس نمونه ($\boldsymbol{s}^2$).

به طور کلی:



	Parameter	Statistic
Mean	μ	$\bar{x}$
Variance	σ^2	S^2
Standard deviation	σ	s
Regression coefficient	β	b
Correlation coefficient	ρ	r

شاخصهای آماری به دو دسته اصلی شاخصهای تمایل به مرکز و شاخص های پراکندگی تقسیم می شوند. با استفاده از این شاخص ها می توان ویژگی های یک جمعیت را توصیف نمود و از مقدار متوسط داده ها اطلاع یافته و چگونگی پراکندگی داده ها در اطراف مقدار متوسط جمعیت را نشان داد.

الف) شاخص های تمایل به مرکز

هر اندازه یا کمیتی که مرکزی از توزیع داده ها را نشان دهد، شاخص تمایل به مرکز گویند. مهمترین شاخص های تمایل به مرکز عبارتند از :نما یا مد، میانه، چارکها، میانگین(حسابی، هندسی، همساز یا هارمونیک، هموزن شده)

۱-نما یا مد (The mode)

- **تعیین مد برای داده های دسته بندی نشده**

در بین مشاهدات حاصل از یک بررسی آماری، نما مشاهده ای است که دارای بیشترین فراوانی یا تکرار است و آن را با m نشان می دهیم.

اگر در یک توزیع آماری تنها یک عدد وجود داشته باشد که دارای بیشترین فراوانی باشد، توزیع را یک نمایی گویند.

مثال ۱:در مجموعه داده های {۲،۴،۳،۲،۲} نما را تعیین کنید. عدد ۲ مد است.

مثال ۲: در مجموعه داده های {۲،۸،۳،۷،۲،۳،۷،۹،۳،۷،۳} نما عدد ۳ می باشد. یعنی: m=۳

در سری داده های مرتب شده اگر دو داده غیر مجاور (مابین آنها یک عدد دیگر باشد) دارای فراوانی یکسان باشد دو نمایی می گوییم.

مثال ۳: در مجموعه داده های {۵،۶،۳،۷،۹،۶،۵،۷}مطلوب است مقدار نما؟

ابتدا داده را مرتب می کنیم : ۳،۵۵۵،۶۶،۷۷۷،۹ که در نتیجه m={۵،۷}

در صورتیکه در مجموعه ای از مشاهدات، همه بطور یکسان تکرار شده باشند، نما وجود نخواهد داشت.

۱,۱,۱,۲,۲,۲,۳,۳,۳ مد ندارد ۱,۲,۳,۴,۵,۶,۷ مد ندارد

نکته : نما یا مد شاخص آماری معتبری نیست زیرا تنها با توجه به فراوانی مشاهدات محاسبه میشود و کمیت داده ها در مقدار آن دخالت ندارد.

اگر داده ها در جدول توزیع فراوانی داده شده باشند، نما یا مد دسته ای است که بیشترین فراوانی را داشته باشد.

مثال ۴: در جدول فراوانی زیر مطلوب است مقدار نما یا مد؟

X_i	۳	۵	۴	۲
F_i	۶	۹	۱	۴

نما یا مد عدد ۵ است که بالاترین فراوانی مطلق را دارد.

- **تعیین مد در جدول توزیع فراوانی های دسته بندی شده**

ابتدا طبقه ای که بیشترین فراوانی را دارد به عنوان طبقه نمادار مشخص می کنیم و سپس مد را از رابطه زیر بدست می آوریم:

$$m = L + \frac{d_۱}{d_۱+d_۲}\times C$$

که در آن m، نما و $d_۱$ تفاضل فراوانی طبقه نمادار با فراوانی طبقه ماقبل و $d_۲$ تفاضل فراوانی طبقه نمادار با فراوانی طبقه مابعد، C فاصله طبقات و L کرانه پایین طبقه نمادار می باشد.

مثال ۵: طبقه مددار و مقدار مد را در جدول توزیع فراوانی زیر محاسبه کنید.

طبقه ها	۱۹-۱۰	۲۹-۲۰	۳۹-۳۰	۴۹-۴۰	۵۹-۵۰
فراوانی مطلق	۱۲	۱۷	۲۰	۴۰	۱۰

$$m = ۳۹.۵ + \frac{۲۰}{۲۰+۳۰}\times ۱۰ = ٤۳.٥$$

$$\text{کران پایین همان طبقه} = \frac{\text{حد بالا طبقه قبل} + \text{حد پایین همان طبقه}}{۲} = \frac{۴۰+۳۹}{۲} = ۳۹.۵$$

۲- میانه (Median)

در یک سری داده مرتب شده (صعودی یا نزولی) میانه عددی است که اعداد را از نظر تعداد به دو قسمت مساوی تقسیم می نماید به طوری که نیمی از اعداد کمتر و نیم دیگر بیشتر از آن می باشند.

- **محاسبه میانه برای داده های معمولی:**

اگر تعداد داده ها فرد باشد داده وسط بعنوان میانه در نظر گرفته می شود و فرمول آن چنین است:

$$M_d = X_{(\frac{n+۱}{۲})}$$

مثال ۶: مقدار میانه را در مجموعه داده های {۳و۱و۳۰و۱۸و۲و۴۵و۱۸و۴۸و۲۰و۸و۱۲} تعیین کنید.

تعداد اعداد فرد می باشد(۱۱) لذا طبق فرمول $M_d = X_{\left(\frac{n+۱}{۲}\right)} = X(۶)$. لذا در مجموعه مرتب شده داده ها عدد ششم که ۱۸ است میانه خواهد بود.

۱و۲و۳و۸و۱۲و۱۸و۱۸و۲۰و۳۰و۴۵و۴۸

در صورتی که تعداد داده ها زوج باشد میانگین دو داده وسط میانه خواهد بود و فرمول آن چنین است:

$$M_d = \frac{X_{\left(\frac{n}{۲}\right)} + X_{\left(\frac{n}{۲}+۱\right)}}{۲}$$

مثال ۷: مقدار میانه را در مجموعه داده های {۵و۱۶و۷و۹و۲۱و۲۹و۶و۴و۴۰و۳} تعیین کنید.

تعداد اعداد زوج می باشد(۱۰) لذا طبق فرمول $M_d = \frac{X_{\left(\frac{n}{۲}\right)} + X_{\left(\frac{n}{۲}+۱\right)}}{۲}$ میانه به شرح زیر محاسبه می شود:

$$M_d = \frac{X_{\left(\frac{n}{۲}\right)} + X_{\left(\frac{n}{۲}+۱\right)}}{۲} = \frac{X_{\left(\frac{۱۰}{۲}\right)} + X_{\left(\frac{۱۰}{۲}+۱\right)}}{۲} = \frac{X_{(۵)} + X_{(۶)}}{۲} = \frac{۷+۹}{۲} = ۸$$

۳و۴و۵و۶و۷و۹و۱۶و۲۱و۲۹و۴۰   ۸= ۲/ (۹+۷)

نکته: برای داده هایی که دارای مقادیر پرت هستند استفاده از شاخص میانه بهتر از میانگین است.

- **محاسبه میانه برای داده های طبقه بندی شده:**

ابتدا طبقه میانه دار را پیدا می کنیم. طبقه میانه دار اولین رده ای است که فراوانی تجمعی مطلق آن مساوی یا بیشتر از $\frac{N}{۲}$ است. میانه عددی است که در این رده قرار دارد. حال برای محاسبه آن از فرمول زیر استفاده می کنیم:

$$M_d = L + \frac{\frac{N}{۲} - FC_{i-۱}}{F_i} \times C$$

که در آن L کران پایین رده میانه دار، N تعداد داده ها، $FC_{i-۱}$ فراوانی تجمعی مطلق قبل از رده میانه دار، F_i فراوانی مطلق رده میانه دار و C طول رده می باشد.

$$\text{کران پایین همان طبقه} = \frac{\text{حد بالای طبقه قبل} + \text{حد پایین همان طبقه}}{۲}$$

مثال ۸: میانه داده های زیر را بدست آورید.

طبقه ها	۱۰۰-۹۵	۱۰۵-۱۰۰	۱۱۰-۱۰۵	۱۱۵-۱۱۰	۱۲۰-۱۱۵
F_i	۷	۲۳	۲۲	۱۷	۴
F_{C_i}	۷	۳۰	۵۲	۶۹	۷۳

$$\sum F_i = ۷۳ = N \Longrightarrow \frac{N}{۲} = \frac{۷۳}{۲} = ۳۶.۵$$

دسته ۱۱۰-۱۰۵ جایی است که فراوانی تجمعی آن از $\frac{N}{۲}$ بیشتر است و ما آن را به عنوان دسته میانه انتخاب می کنیم.

$$M_d = L + \frac{\frac{N}{۲} - FC_{i-۱}}{F_i} \times C = ۱۰۵ + \frac{\frac{۷۳}{۲} - ۳۰}{۲۲} \times ۵ = ۱۰۵ + \frac{۶.۵}{۲۲} \times ۵ = ۱۰۶.۴۸$$

- **محاسبه میانه برای داده های با طبقات گسسته:**

در صورتیکه جدول دارای طبقات گسسته باشد ابتدا طبقات را به صورت پیوسته در می آوریم برای این کار کافی است $\frac{\text{حد پایین طبقه بعدی} + \text{حد بالا طبقه}}{۲}$ شده و از حد پایین طبقه قبل ۰/۵ واحد کم و به حد بالای طبقه آخر ۰/۵ اضافه کنیم.

مثال۹: با استفاده جدول زیر median را محاسبه کنید.

طبقات	طبقات پیوسته	فراوانی مطلق	فراوانی تجمعی مطلق
۵-۱	۰,۵-۵,۵	۴	۴
۱۰-۶	۵,۵-۱۰,۵	۱	۵
۱۵-۱۱	۱۰,۵-۱۵,۵	۱	۶
۲۰-۱۶	۱۵,۵-۲۰.۵	۱۰	۱۶
۲۵-۲۱	۲۰,۵-۲۵,۵	۳	۱۹
۳۰-۲۶	۲۵,۵-۳۰,۵	۵	۲۴
۳۵-۳۱	۳۰,۵-۳۵,۵	۳	۲۷
۴۰-۳۶	۳۵,۵-۴۰,۵	۰	۲۷
۴۵-۴۱	۴۰,۵-۴۵,۵	۱	۲۸

اولین رده ای که فراوانی تجمعی آن بیشتر از $\frac{N}{۲}$ هست

دسته ۱۵,۵-۲۰,۵ جایی است که فراوانی تجمعی آن از $\frac{N}{۲}$ بیشتر است و ما آن را به عنوان دسته میانه انتخاب کردیم.

$$\sum F_i = N = ۲۱ \qquad \frac{N}{۲} = \frac{۲۸}{۲} = ۱۴$$

$$L = \frac{\text{کران بالای طبقه ماقبل} + \text{کران پایین همان طبقه}}{۲} = \frac{۱۶ + ۱۵}{۲} = ۱۵.۵$$

$$M_d = ۱۵.۵ + \frac{\frac{۲۸}{۲} - ۶}{۱۰} \times ۴ = ۱۸.۵$$

نکته : در جدول توزیع فراوانی نما یا مد عبارت است از عددی که فراوانی مطلق آن بزرگتر باشد.

۳-چندک ها (**Quantiles**): یا پارک ها (**Partiles**) :

مقادیری از مشاهدات هستندکه جامعهٔ آماری را به فواصل مساوی با نسبت های $\frac{1}{4}$(چارک)، $\frac{1}{10}$ (دهک) و$\frac{1}{100}$ (صدک) تقسیم می کند، به طوری که فراوانی در هر یک از این فواصل درصد معینی از فراوانی کل را تشکیل می دهد.

الف- چارک (**Quartiles**):

اگر روی رشته اعدادی که آماربرداری کرده ایم ۳ نقطه با فواصل مساوی انتخاب کنیم، ۴ دامنه به صورت ۲۵ درصدی ایجاد خواهد شد که ۲۵ درصد اول جزء ضعیف ترین داده ها و ۲۵ درصد آخر جزء قوی ترین داده ها خواهند بود. به اینها چارک گویند که با Q۱، Q۲، Q۳ و Q۴ نشان می دهند. بنابراین اگر جامعه آماری را به چهار قسمت تقسیم کنیم، چارک های اول، دوم و سوم (Q۱ تا Q۳) به وجود می آیند. بعد از مرتب کردن داده ها به صورت صعودی داریم. در اینجا میانه مترادف با چارک دوم ($Q_۲$) است.

$$\begin{array}{ccccccccc} & ۱/٤=\%۲٥ & & ۱/٤=\%۲٥ & & ۱/٤=\%۲٥ & & ۱/٤=\%۲٥ & \\ X_{min} & \longleftrightarrow & Q_۱ & \longleftrightarrow & Q_۲ & \longleftrightarrow & Q_۳ & \longleftrightarrow & X_{max} \\ & & & & \text{میانه} & & & & \end{array}$$

ب- دهک (Deciles) :

اگر این رشته اعداد یا جامعۀ آماری را با ۹ نقطه مساوی جدا کنیم، ده قسمت خواهیم داشت و دهک های اول تا نهم($D_۱$ تا $D_۹$) بوجود می آیند. بعد از مرتب کردن داده ها به صورت صعودی داریم. میانه مترادف با $D_۵$ است.

$$X_{min} \overset{۱/۱۰}{\longleftrightarrow} D_۱ \overset{۱/۱۰}{\longleftrightarrow} D_۲ \ldots \underset{\text{میانه}}{\underline{D_۵}} \ldots \overset{۱/۱۰}{\longleftrightarrow} Q_۹ \overset{۱/۱۰}{\longleftrightarrow} X_{max}$$

ج- صدک (Percentiles) :

اگر جامعۀ آماری را به ۹۹ نقطه مساوی تقسیم کنیم ۱۰۰ قسمت خواهیم داشت، صدک های اول تا۹۹($P_۱$ تا $P_{۹۹}$) بوجود می آید. در اینجا میانه صدک پنجاهم ($P_{۵۰}$) می باشد. بعد از مرتب کردن داده ها به صورت صعودی داریم :

$$X_{min} \underset{۱/۱۰۰}{\longleftrightarrow} P_۱ \underset{۱/۱۰۰}{\longleftrightarrow} P_۲ \ldots \underset{\text{میانه}}{\underline{P_{۵۰}}} \ldots \underset{۱/۱۰۰}{\longleftrightarrow} P_{۹۹} \underset{۱/۱۰۰}{\longleftrightarrow} X_{max}$$

نکته ←مشاهده می شودکه بیشترین تطبیق درصد بین چندک ها (quantiles) رخ می دهد.

محاسبه پارک ها برای داده های دسته بندی نشده:

اگر N داده های ما باشند که به صورت صعودی مرتب شده باشند برای بدست آوردن پارک موردنظر (P_p) که تقریبی می باشد از فرمول زیر استفاده می کنیم:

$$Pp = (۱ - w)X_{(r)} + w\,X_{(r+۱)}$$

با استفاده از فرمول P(N+۱) مقادیر r و W را محاسبه و در رابطه فوق جایگذاری می کنیم. که r عدد صحیح و w عدد اعشاری می باشد.

مثال ۱۰: در مجموعه داده های ۷۳،۱۰،۱۵،۳۰،۸۴،۴۵،۷۳،۲۰،۱۰۰،۲۰ چارک سوم را بدست آورید.

مرتب کردن داده ها از کوچک به بزرگ:

۱۰،۱۵،۲۰،۳۰،۴۵،۷۳،۷۳،۸۴،۹۰،۱۰۰

(N+۱)P = (۱۰+۱)×۰,۷۵ = ۸,۲۵

$$Q_{۳} = (۱ - \mathrm{w})\mathrm{X}_{(\mathrm{r})} + \mathrm{w}\,\mathrm{X}_{(\mathrm{r}+۱)} = (۱ - ۰.۲۵)\,X_{۸} + ۰.۲۵\,X_{۹} = ۰.۷۵ \times ۸۴ + ۰.۲۵ \times ۹۰$$
$$= ۶۳ + ۲۲.۵ = ۸۵.۵$$

مثال ۱۱: دهک ششم مجموعۀ داده های زیر را محاسبه کنید.

۲۰،۱۰۰،۲۰،۷۳،۴۵،۸۴،۳۰،۱۵،۱۰،۷۳

مرتب کردن داده ها از کوچک به بزرگ:

۱۰,۱۵,۲۰,۳۰,۴۵,۷۳,۷۳,۸۴,۹۰,۱۰۰

(N+۱)P = (۱۰+۱)×۰,۶ = ۶,۶

$$D_{۶} = (۱ - \mathrm{w})\mathrm{X}_{(\mathrm{r})} + \mathrm{w}\,\mathrm{X}_{(\mathrm{r}+۱)} = (۱ - ۰.۶)\,X_{۶} + ۰.۶\,X_{۷} = ۰.۴ \times ۷۳ + ۰.۶ \times ۷۳$$
$$= ۲۹.۲ + ۴۳.۸ = ۷۳$$

مثال ۱۲: داده های زیر سهام ۲۵ شرکت در بورس می باشد صدک بیستم این داده ها را بدست آورید.

۲۰.۵	۱۹.۵	۱۵.۶	۲۴.۱	۹.۹
۱۵.۴	۱۲.۷	۵.۴	۱۷.۰	۲۸.۶
۱۶.۹	۷.۸	۲۳.۳	۱۱.۸	۱۸.۴
۱۳.۴	۱۴.۳	۱۹.۲	۹.۲	۱۶.۸
۸.۸	۲۲.۱	۲۰.۸	۱۲.۶	۱۵.۹

- ✓ مرتب کردن داده ها از کوچک به بزرگ :

۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲	۲۳	۲۴	۲۵
۵/۴	۷/۸	۸/۸	۹/۲	۹/۹	۱۱/۸	۱۲/۶	۱۲/۷	۱۳/۴	۱۴/۳	۱۵/۴	۱۵/۶	۱۵/۹	۱۶/۸	۱۶/۹	۱۷	۱۸/۴	۱۹/۲	۱۹/۵	۲۰/۵	۲۰/۸	۲۲/۱	۲۳/۳	۲۴/۱	۲۸/۶

(N+۱)P = (۲۵+۱)×۰,۲ = ۵,۲

$$P_{۲۰} = (۱ - \mathrm{w})\mathrm{X}_{(\mathrm{r})} + \mathrm{w}\,\mathrm{X}_{(\mathrm{r}+۱)} = (۱ - ۰.۲)\,X_{۵} + ۰.۲ \times X_{۶} = ۰.۸ \times ۹.۹ + ۰.۲ \times ۱۱.۸$$
$$= ۷.۹۲ + ۲.۳۶ = ۱۰.۲۸$$

۲۰ ٪ داده ها از ۱۰/۲۸ کمتر و ۸۰ ٪ داده ها از ۱۰/۲۸ بیشتر است.

۴- میانگین (Mean):

میانگین ها مهمترین شاخص های تمایل به مرکز بوده و دارای بیشترین استفاده می باشند که در این شاخص ها علاوه بر فراوانی مشاهدات مقدار آنها را نیز در نظر می گیرند. داده پرت (Outlier) میانگین را به شدت به سمت خود می کشد.

الف- میانگین حسابی (Arithmetic mean): در حقیقت همان معدل یا متوسط داده هاست که از تقسیم مجموع داده ها بر تعداد آنها بدست می آید.

- **میانگین داده های طبقه بندی نشده:**

میانگین نمونه:
$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

میانگین جامعه:
$$\mu = \frac{X_1 + X_2 + X_3 + \cdots X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N}$$

مثال۱۳: میانگین حسابی اعداد ۴،۸،۵،۳ در یک نمونه به صورت زیر محاسبه می شود:

$$\mu = \frac{۴ + ۸ + ۵ + ۳}{۴} = ۵$$

مثال ۱۴ : میانگین ۱۰داده آماری ۳۲/۵ است. اگر دو داده ۳۵و۴۰را از آنها کنار بگذاریم، میانگین ۸ داده حاصل را محاسبه کنید؟

$$\overline{X} = \frac{\sum_{i=۱}^{n} X_i}{n} \Rightarrow ۳۲.۵ = \frac{\sum X_i}{۱۰} \Rightarrow \sum X_i = ۳۲.۵ \times ۱۰ = ۳۲۵$$
$$\overline{X} = \frac{۳۲۵ - ۳۵ - ۴۰}{۱۰ - ۲} = \frac{۲۵۰}{۸} = ۳۱.۲۵$$

مثال ۱۵: میانگین داده های جدول توزیع فراوانی زیر چند است؟

X_i	۳	۱	۲
F_i	۲	۴	۵

$$\bar{X} = \frac{\sum_{i=۱}^{n} F_i X_i}{n} = \frac{۲ \times ۳ + ۴ \times ۱ + ۵ \times ۲}{۲ + ۴ + ۵} = ۱/۸۲$$

میانگین داده های طبقه بندی شده:

اگر داده ها به صورت طبقه بندی شده باشند ابتدا مرکز طبقات ($\tilde{X}_i$) را محاسبه و سپس میانگین طبق فرمول زیر تعیین می شود:

$$\bar{X} = \frac{\sum_{i=۱}^{n} F_i \tilde{X}_i}{N} \quad \text{یا} \quad \bar{X} = \sum_{i=۱}^{n} \mathrm{f_i}\, \tilde{X}_i \quad \text{که} \quad \mathrm{f}_i = \frac{F_i}{N}$$

که n تعداد داده ها، $\mathrm{F_i}$ فراوانی مطلق، $\tilde{\mathrm{X}}_\mathrm{i}$ نماینده طبقه (میانگین حد بالا و حد پایین هر طبقه) می باشد.

مثال ۱۶: مقدار میانگین حسابی داده های طبقه بندی شده زیر را محاسبه کنید.

طبقات	نماینده طبقه $\tilde{X}_i$	فراوانی مطلق F_i	فراوانی نسبی f_i	نماینده طبقه× فراوانی مطلق $F_i \times \tilde{X}_i$	نماینده طبقه× فراوانی نسبی $f_i \times \tilde{X}_i$
۱۱-۱۲	۱۱,٥	۲	۰,۰۲۰	۲۳	۰,۲۳
۱۲-۱۳	۱۲,٥	۱٦	۰,۱٦۰	۲۰۰	۲
۱۳-۱٤	۱۳,٥	۲۹	۰,۲۹۰	۳۹۱,٥	۳,۹۲
۱٤-۱٥	۱٤,٥	۲۷	۰,۲۷۰	۳۹۱,٥	۳,۹۲
۱٥-۱٦	۱٥,٥	۱۱	۰,۱۱۰	۱۷۰,٥	۱,۷۱
۱٦-۱۷	۱٦,٥	٦	۰,۰٦۰	۹۹	۰,۹۹
۱۷-۱۸	۱۷,٥	٤	۰,۰٤۰	۷۰	۰,۷
۱۸-۱۹	۱۸,٥	۲	۰,۰۲۰	۳۷	۰,۳۷
۱۹-۲۰	۱۹,٥	۲	۰,۰۲۰	۳۹	۰,۳۹
۲۰-۲۱	۲۰,٥	۱	۰,۰۱۰	۲۰,٥	۰,۲۰٥
		۱۰۰		۱٤٤۲	۱٤,٤۲

$$\bar{X} = \frac{\sum_{i=۱}^{n} F_i \tilde{X}_i}{n} = \frac{۱۴۴۲}{۱۰۰} = ۱۴.۴۲ \quad \text{یا} \quad \bar{X} = \sum_{i=۱}^{n} P_i\, \tilde{X}_i = ۱۴.۴۲$$

ویژگی های میانگین :

۱) همواره مجموع انحراف داده ها از میانگین صفر می باشد.

$$\sum_{i=۱}^{n} (\mathrm{X_i} - \bar{\mathrm{X}}) = (۳-۵) + (۵-۵) + (۸-۵) + (۴-۵) = ۳ - ۳ = ۰$$

و در یک جدول توزیع فراوانی فرمول به صورت زیر می باشد که F_i فراوانی هر طبقه است:

$$\sum_{\mathrm{i}=۱}^{\mathrm{n}} \mathrm{F_i}\,(\mathbf{X_i} - \bar{\mathbf{X}})$$

۲) هر گاه به هر یک از داده ها عدد ثابتی مانند a اضافه یا کم کنیم میانگین داده های حاصل برابر است با میانگین داده های قبلی به اضافه یا منهای عدد ثابت خواهد بود.

$$Y_i = x_i \pm a \Longrightarrow \boldsymbol{y_i} = \boldsymbol{\overline{x_\iota}} \pm \boldsymbol{a}$$

مثال ۱۷: اگر به اعداد ۴،۸،۵،۳ عدد ثابت ۳ را اضافه کنیم، میانگین جدید عبارتست از ۵ به اضافه ۳ که عدد ۸ خواهد شد:

اعداد جدید ۷،۱۱،۸،۶

$$\mu = \frac{۶+۸+۱۱+۷}{۴} = ۸$$

$$\overline{y_\iota} = \overline{x_\iota} + a = ۵ + ۳ = ۸$$

۳) اگر تک تک داده ها را در عدد ثابتی مانند b ضرب کنیم میانگین نیز در b ضرب خواهد شد.

$$\mathrm{y_i} = \mathrm{bX_i} \Rightarrow \bar{\mathrm{y}} = \mathrm{b}\overline{\mathrm{X}}$$

مثال ۱۸: اگر اعداد ۳، ۵، ۸، ۴ را در عدد ۳ ضرب کنیم اعداد جدید ۹، ۱۵،۲۴ و ۱۲ خواهند بود که میانگین این اعداد عبارت خواهد بود از :

$$\mu = \frac{\mathbf{9+24+15+12}}{\mathbf{4}} = ۱۵$$

میانگین اعداد جدید برابر است با میانگین قبلی ضربدر عدد ثابت:

$$\bar{\mathrm{y}} = \mathrm{b}\overline{\mathrm{X}} \Longrightarrow \bar{\mathrm{y}} = ۳ \times ۵ = ۱۵$$

۴) اگر تک تک داده ها را بر عدد ثابتی مانند b تقسیم کنیم میانگین اعداد جدید برابر با میانگین قبلی تقسیم بر b خواهد بود.

$$y = \frac{X_i}{b} \Longrightarrow \bar{\mathbf{y}} = \frac{\overline{\mathbf{X}}}{b}$$

مثال ۱۹: اگر اعداد ۱۲، ۱۸ و ۱۵ بر عدد ثابت ۳ تقسیم شوند میانگین اعداد جدید را بدست آورید.

$$\mu = \frac{۹ + ۱۸ + ۱۲}{۳} = ۱۵$$

اعداد جدید ۴، ۶ و ۵ خواهند بود که میانگین آنها عبارتست از :

$$\mu = \frac{۵ + ۶ + ۴}{۳} = ۵$$

بجای محاسبه بالا می توان نوشت که میانگین اعداد جدید برابر با میانگین قبلی تقسیم بر ۳خواهد بود.

$$\bar{y} = \frac{\bar{X}}{b} = \frac{۱۵}{۳} = ۵$$

۵) اگر دو مجموعه اعداد داشته باشیم و دسته سوم اعدادی را از روی آنها بسازیم آنگاه میانگین اعداد جدید، جمع این دو میانگین خواهد بود. این خاصیت زمانی رخ می دهد که n۱=n۲ باشد یعنی تعداد داده ها برابر باشد.

X	Y	Z=X+Y
۴	۳	۷
۳	۲	۵
۷	۶	۱۳
۶	۴	۱۰
۵	۵	۱۰
۲۵	۲۰	۴۵

$$\mu x = \frac{۴ + \cdots + ۵}{۵} = ۵ \quad \mu Y = \frac{۳ + \cdots + ۵}{۵} = ۴ \quad \mu Z = \frac{۷ + \cdots + ۱۰}{۵} = ۹$$

$$\mu Z = \mu x + \mu Y$$

حال اگر $n۱ \neq n۲$ باشد میانگین وزنی مطرح خواهد بود.

ب- میانگین وزنی (**Weighted mean**) : برای حالتی که هر کدام از مشاهدات دارای وزن های متفاوت باشند استفاده می شود که با $\bar{X}_W$ نشان داده می شود. مانند محاسبه معدل دانشجویان.

$$\bar{x} = \frac{W_۱X_۱ + W_۲X_۲ + W_۳X_۳ + \cdots + W_nX_n}{W_۱ + W_۲ + W_۳ + \cdots + W_n} = \frac{\sum_{i=۱}^{n} W_iX_i}{\sum_{i=۱}^{n} W_i}$$

مثال ۲۰: یک جیره غذایی به وزن ۵۰۰ گرم مخلوطی از ۴ ماده A,B,C,D است که در جدول زیر مقادیر هر یک و درصد پروتئین آنها آمده است متوسط پروتئین این جیره غذایی را محاسبه کنید.

ماده غذایی	وزن گرم Wi	پروتئین Xi	WiXi
A	۱۵۰	۸	۱۲۰۰
B	۵۰	۱۲	۶۰۰
C	۲۰۰	۹	۱۸۰۰
D	۱۰۰	۱۰	۱۰۰۰

$$\mu_w = \frac{\sum Wi \times Xi}{\sum Wi} = \frac{۱۵۰ \times ۸ + ۵۰ \times ۱۲ + ۲۰۰ \times ۹ + ۱۰۰ \times ۱۰}{۵۰۰} = \frac{۴۶۰۰}{۵۰۰} = ۹.۲$$

اگر میانگین وزنی چهار ماده غذایی A,B,C,D در جدول فوق برابر ۹/۲ باشد میزان (مقدار) پروتئین ماده غذایی C را محاسبه کنید .

$$۹.۲ = \frac{۱۵۰ \times ۸ + ۵۰ \times ۱۲ + ۲۰۰ \times X + ۱۰۰ \times ۱۰}{۵۰۰} = \frac{۴۶۰۰}{۵۰۰} = ۹.۲$$

$$۹.۲ = \frac{۱۲۰۰ + ۶۰۰ + ۲۰۰ \times X + ۱۰۰۰}{۵۰۰} \Longrightarrow ۹.۲ = \frac{۲۸۰۰ + ۲۰۰ \times X}{۵۰۰} \Longrightarrow ۴۶۰۰$$

$$= ۲۸۰۰ + ۲۰۰ \times X \Longrightarrow ۴۶۰۰ - ۲۸۰۰ = ۲۰۰ \times X \Longrightarrow ۱۸۰۰ = ۲۰۰ \times X$$

$$X = \frac{۱۸۰۰}{۲۰۰} = \mathbf{9}$$

فرمول دیگر میانگین وزنی:

$$\bar{X}_W = \frac{n_۱\bar{X}_۱ + n_۲\bar{X}_۲ + \cdots + n_K\bar{X}_K}{n_۱ + n_۲ + \cdots + n_K} = \frac{\sum_{i=۱}^{k} n_i\bar{X}_i}{\sum_{i=۱}^{k} n_i}$$

مثال ۲۱ : از ۵۰۰ دانشجو، ۱۵۰ دختر با میانگین قد۱۵۷/۵ سانتی متر و بقیه دانشجوها پسر با میانگین قدی ۱۷۸/۵ سانتی متر می باشد. میانگین قدی ان دانشجویان را حساب کنید.

$$\bar{X}_W = \frac{۱۵۰ \times ۱۵۷.۵ + ۳۵۰ \times ۱۷۸.۵}{۱۵۰ + ۳۵۰} = ۱۷۲.۷۰$$

مثال ۲۲ : در درس آمار و احتمالات سال تحصیلی ۱۴۰۲-۱۴۰۱، ۲۰ دانشجوی سال اول نمره ۶۸، ۱۸ دانشجوی سال دوم نمره ۷۵ و ۱۲ نفر دانشجوی سال سوم نمره ۸۶ گرفته اند. میانگین نمره دانشجویان را محاسبه کنید.

$$\bar{X}_W = \frac{۲۰ \times ۶۸ + ۱۸ \times ۷۵ + ۱۲ \times ۸۶}{۲۰ + ۱۸ + ۱۲} = ۷۴.۸$$

مثال ۲۳: اگر میانگین n داده آماری برابر ۱۵ باشد و میانگین ۱۵ داده برابر ۱۰ باشد، به طوریکه میانگین تمام داده ها ۱۲ باشد، مقدار n را بدست آورید.

$$\bar{X} = \frac{n\bar{x}_۱ + m\bar{x}_۲}{n+m} \Rightarrow ۱۲ = \frac{۱۵ \times n + ۱۰ \times ۱۵}{n + ۱۵} \Rightarrow ۱۵ \times n + ۱۵۰ = ۱۲ \times n + ۱۸۰$$
$$\Rightarrow ۳ \times n = ۳۰ \Rightarrow n = ۱۰$$

میانگین × تعداد = مجموع داده ها

ج- **میانگین هارمونیک یا همساز یا معکوس (Harmonic mean)**: اگر سرعت(وقتی فواصل طی شده برابر باشد)، شدت و یا ارزش انجام کار معینی فرق کند از این میانگین استفاده می کنیم.

$$m_h = \frac{n}{\frac{۱}{X_۱} + \frac{۱}{X_۲} + \cdots + \frac{۱}{X_n}} = \frac{n}{\sum_{i=۱}^{n} {}^{۱}/_{X_i}} = \frac{۱}{\frac{۱}{n}\sum_{i=۱}^{n} \frac{۱}{X_i}}$$

مثال ۲۴ : کارگری یک کار معین را در سه روز دومی در چهار روز و سومی در شش روز تمام می کند میانگین روزهای لازم برای تمام شدن کار در این کارگاه چند روز است.

$$m_h = \frac{۱+۱+۱}{\frac{۱}{۳} + \frac{۱}{۴} + \frac{۱}{۶}} = \frac{\frac{۳}{۱}}{\frac{۴+۳+۲}{۱۲}} = \frac{\frac{۳}{۱}}{\frac{۹}{۱۲}} = \frac{۳۶}{۹} = ۴$$

مثال۲۵: اتومبیلی فاصله ی بندرعباس تا میناب را با سرعت ۱۰۰ کیلومتر در ساعت رفته و با سرعت ۱۲۰ کیلومتر در ساعت برگشته است. میانگین سرعت رفت و برگشت این اتومبیل چه قدر است؟

$$m_h = \frac{۲}{\frac{۱}{۱۰۰} + \frac{۱}{۱۲۰}} = \frac{\frac{۲}{۱}}{\frac{۱۲+۱۰}{۱۲۰۰}} = \frac{\frac{۲}{۱}}{\frac{۲۲}{۱۲۰۰}} = \frac{۲۴۰۰}{۲۲} = ۱۰۹.۰۹$$

د- **میانگین هندسی (Geometric mean)** برای حالتی که مشاهدات حاصل از یک بررسی آماری بر حسب درصد، نسبت و امثال اینها باشد یا نسبت دو عدد متوالی ثابت باشد از میانگین هندسی استفاده می شود.

$$m_g = \sqrt[n]{x_1 \times x_2 \times x_3 \times \cdots \times x_n} \quad \rightarrow \quad \log m_g = \frac{\sum_{i=1}^{n} \log x_i}{n}$$

میانگین هندسی n داده آزمایشی برابر است با ریشه nام حاصل ضرب این داده ها

مثال ۲۶ : میانگین هندسی اعداد ۲و۴و۸ و۱۶ و ۳۲ را بدست آورید.

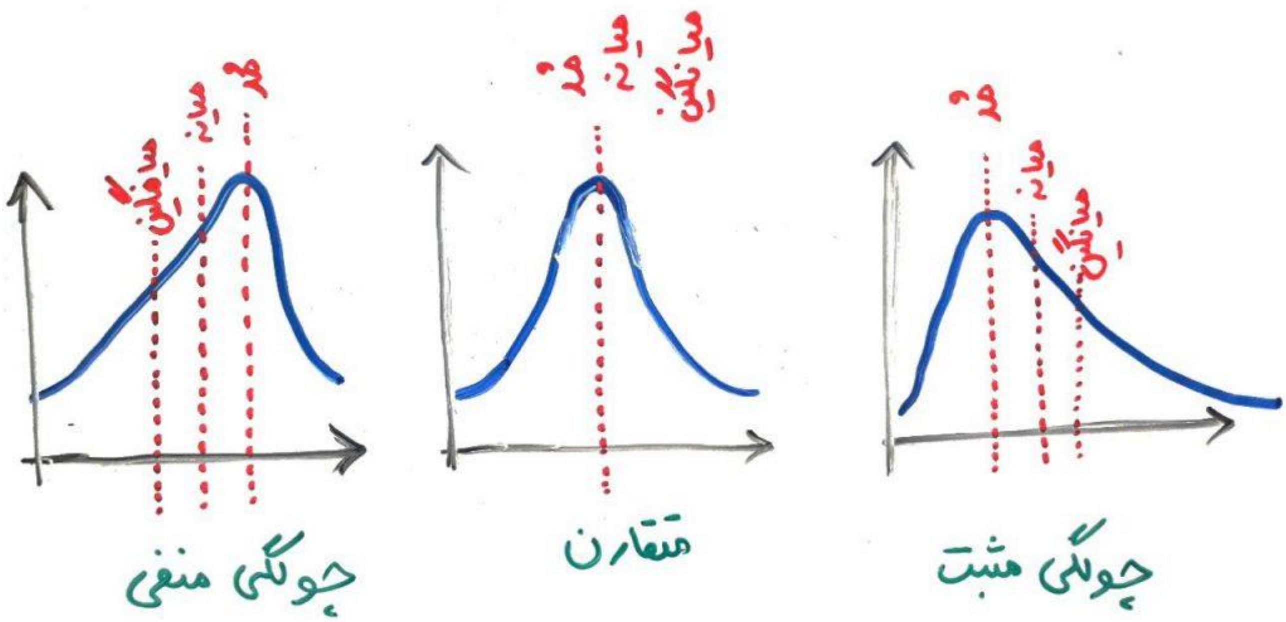
$$\mu g = \sqrt[5]{2 \times 4 \times 8 \times 16 \times 32} = \sqrt[5]{32768} = \sqrt[5]{8^5} = 8$$

رابطه بین میانگین های حسابی، هندسی و هارمونیک:

همواره مقدار میانگین حسابی از همه بیشتر، هندسی حدواسط و میانگین هارمونیک از همه کوچک تر است.

$$m_h \le m_g \le \bar{x}$$

اگر داده ها به صورت نرمال باشند میانگین، میانه و نما روی هم می افتند (باهم مساوی اند). اما اگر چولگی به چپ یا راست داشته باشد نمودارها به شرح زیر خواهند بود :



تفاوت اساسی بین میانگین، میانه و نما این است که میانگین بر حسب مقیاس داده هاست و در محاسبه آن فراوانی و کمیت داده ها منظور می شود و بستگی به ارزش عددی هر یک از مشاهدات در توزیع فراوانی دارد. اما میانه و نما شاخص ها یا داده هایی در بین مشاهدات است که تابع ترتیب و فراوانی داده ها می باشند. لذا میانگین از ترکیب داده ها حاصل شده و افزایش یا کاهش در آنها مقدار آن را تغییر می دهد. در نقطه مقابل چنانچه افزایش یا کاهش ها ترتیب داده ها را تغییر ندهند، در مقدار میانه و نما تغییری حاصل نمی شود.

تمرین ۱ : با استفاده از جدول فراوانی زیر مقدار نما را برآورد کنید.

طبقه ها	۱۰۰-۹۵	۱۰۵-۱۰۰	۱۱۰-۱۰۵	۱۱۵-۱۱۰	۱۲۰-۱۱۵
فراوانی مطلق	۷	۲۳	۲۲	۱۷	۴

تمرین ۲: میانه اعداد ۵۳،۶۷،۳۱،۳۶،۵۳ را محاسبه کنید.

تمرین ۳: میانه اعداد ۱۸،۱۳،۱۵،۱۲،۸،۲۱،۷،۱۱،۱۶،۳ را بدست آورید.

تمرین ۴ : داده های طبقه بندی شده زیر را درنظر گرفته و مقادیر میانه، میانگین و نما را محاسبه کنید:

طبقه ها	۲۰-۱۰	۳۰-۲۰	۴۰-۳۰	۵۰-۴۰
F_i	۱۰	۲۰	۳۰	۴۰

الف- میانه

ب- میانگین

ج- نما یا مد

تمرین ۵: میانه و مد داده های طبقه بندی شده زیر را محاسبه کنید:

طبقه ها	۱۰۰-۹۵	۱۰۵-۱۰۰	۱۱۰-۱۰۵	۱۱۵-۱۱۰	۱۲۰-۱۱۵
فراوانی مطلق	۷	۲۳	۲۲	۱۷	۴

الف- میانه

ب- مد

تمرین۶- فرض کنید نمرات درس آمار در یک کلاس با ۱۹ دانشجو به ترتیب صعودی به صورت زیر است:

مطلوب است :الف) چارک سوم، ب) دهک ششم، ج) صدک هفتاد و هشتم.

۵۸، ۶۰، ۶۵، ۶۵، ۶۸، ۶۹، ۷۰، ۷۲، ۷۵، ۷۵، ۷۶، ۷۷، ۸۱، ۸۲، ۸۴، ۸۷، ۹۰، ۹۵، ۹۸

الف- چارک سوم

ب- دهک ششم

ج) صدک هفتادوهشتم

تمرین ۷ : در ۱۰داده آماری میانگین ۱۸است. دو داده ۱۱و ۱۳را حذف و داده ۲۷را اضافه می کنیم. میانگین تقریبا چقدر زیاد میشود؟

تمرین۸ : میانگین ۶ داده آماری برابر ۱۳ است اگر دو عدد ۹و $4X$ را کنار آنها قرار دهیم، میانگین ۸ داده حاصل برابر ۱۵ می شود. مقدار X چقدر است؟

تمرین ۹ : جدول توزیع فراوانی زیر را درنظر بگیرید اگر $\mu = 2$ و $N = 28$ باشد، مطلوب است مقادیر a و b؟

X_i	۰	۱	۲	۳	۴
F_i	۳	a	۱۰	b	۳

تمرین ۱۰: مطلوب است میانگین داده های جدول زیر:

طبقه	۵-۰	۵-۱۰	۱۰-۱۵
F_i	۲	۳	۴

تمرین ۱۱: کارنامه دو داوطلب شرکت کننده در یک آزمون استخدامی به صورت زیر است. اگر قرار باشد از میان این دو نفر یکی استخدام شود؛ کدام یک استخدام خواهند شد؟

	زبان (ضریب ۳)	اطلاعات عمومی (ضریب ۱)	پرسش های تخصصی (ضریب ۳)
داوطلب اول	۱۹	۱۴	۱۷
داوطلب دوم	۱۷	۲۰	۱۶

تمرین ۱۲ : اگر میانگین نمرات ۷دانش آموز برابر ۱۶و میانگین نمرات ۱۳دانش آموز دیگر برابر ۱۲باشد، میانگین نمرات کل دانش آموزان را محاسبه کنید.

تمرین ۱۳: اگر ۳ خودرو مسیر ۱۲۰ کیلومتری بین دو شهر را به ترتیب با سرعت ۱۱۰، ۷۰ و ۹۰ کیلومتر در ساعت طی نمایند. میانگین سرعت این ۳ خودرو با هم چند کیلومتر بر ساعت است؟

تمرین ۱۴: میانگین حسابی، هندسی و هارمونیک اعداد ۲، ۳ و ۴ را به دست آورید.

م

تمرین ۱۵: فرض کنید در یک جشن از ۱۰ نفر از میهمان خواسته می شود که هرکدام مبلغی را برای کمک به یک موسسه خیریه پرداخت کنند. اعداد زیر مربوط به مبالغ پرداختی توسط این ۱۰ نفر بر حسب هزار تومان است. از چه پارامتری برای نشان دادن میزان شاخص پرداختی این ۱۰ نفر استفاده می کنید؟

۳۵ ،۳۷ ،۳۹ ،۴۰ ،۴۳ ،۴۴ ،۴۴ ،۹۹۰۰

تمرین ۱۶: اتومبیلی مسیری را با سرعت ۳۰ کیلومتر بر ساعت رفته و هنگام برگشت $\frac{1}{3}$ مسیر را با سرعت ۶۰ کیلومتر بر ساعت و باقی مانده را با سرعت۴۰ کیلومتر بر ساعت برگشته است، سرعت متوسط اتومبیل چقدر بوده است؟

تمرین ۱۷: در یک کارگاه ۵ ماشین با سرعت ۴ دور در ثانیه و ۳ ماشین با سرعت ۶ دور در ثانیه کار می کنند.سرعت متوسط این ماشین ها چند دور در ثانیه است؟

فصل پنجم:

شاخص های پراکندگی

شاخص های پراکندگی

شاخص های مرکزی فقط بخشی از اطلاعات لازم را ارائه می دهند و به تنهایی تصویر روشنی از متغیر ها را نمی توانند نشان دهند. فرض کنید نمرات دانشجویان دو کلاس از ۱۰۰ در یک آزمون آماری به شرح زیر می باشد.

۷۷	۷۷	۷۷	۷۴	۷۳	۷۱	۶۸	۶۷	۶۶	۶۵	کلاس A
۱۰۰	۹۳	۸۵	۷۷	۷۷	۶۷	۶۲	۵۸	۵۴	۴۲	کلاس B

شاخص های متمایل به مرکز ۲ مجموعه داده عبارت است از:

میان دامنه	نما	میانه	میانگین	کلاس
۷۱	۷۷	۷۲	۷۱/۵	A
۷۱	۷۷	۷۲	۷۱/۵	B

شاخص های متمایل به مرکز این ۲ سری داده مشابه است اما وقتی به پراکندگی داده ها نگاه می کنیم داده های گروه B پراکندگی بیشتری دارند نتیجه اینکه معیار های مرکزی به تنهایی نمی توانند معرف کاملی از داده ها باشند و برای تفسیر بهتر داده ها به شاخص های پراکندگی نیز نیاز داریم. مهمترین شاخص هایی که میزان پراکندگی را اندازه گیری میکنند، عبارتند از: دامنه تغییرات، چارک متوسط(نیم دامنه میان چارکی)، میانگین انحرافات، واریانس، انحراف معیار و ضریب تغییرات.

۱-دامنه کلی تغییرات (Range)

دامنه تغییرات از تفاضل بین بزرگترین و کوچکترین داده بدست می آید.

$R = Max(x_i) - Min(x_i)$

مثال ۱ : دامنه تغییرات مربوط به مشاهدات ۶ ,۷ ,۳ ,۲ , ۴ کدام است؟

R=۷-۵=۲

مثال ۲: در جدول طبقه بندی شده زیر دامنه تغییرات کدام است؟

حدود طبقات	۱-۴	۴-۷	۷-۱۰	۱۰-۱۳
فراوانی	۷	۱۳	۵	۶

R=۱۳-۱=۱۲

مثال ۳ : با توجه به جدول زیر دامنه تغییرات کدام است؟

مرکز طبقه	-۲	۰	۶	۷
فراوانی	۱۱	۷	۳	۴

R=۷-(-۲)=۹

۲- انحراف متوسط (Average Deviation)

هیچ یک از شاخص هایی که تا به حال در مورد آ نها صحبت شد، قادر به بیان تمام تغییرات نیستند. تغییر زمانی مفهوم پیدا می کند که تمام داده ها نسبت به یک مبدأ مقایسه شوند و بهترین مبدأ یا مرکز داده ها، میانگین است. اما از آنجاکه مجموع انحرافات از میانگین همیشه صفر است، می توانیم از متوسط قدرمطلق انحرافات حول میانگین استفاده کنیم تا به پارامتر پراکندگی جدیدی به نام انحراف متوسط از میانگین برسیم. این شاخص پراکندگی مشاهدات را نسبت به میانگین نشان می دهد و با AD یا MD نشان داده می شود.

- AD برای داده های طبقه بندی نشده

$$AD = \frac{\sum_{i=۱}^{n} |X_i - \bar{X}|}{N}$$

- AD برای داده های طبقه بندی شده

$$AD = \frac{\sum_{i=۱}^{n} \mathrm{F_i}|X_i - \bar{X}|}{N}$$

مثال ۴ : میانگین انحرافات دسته اعداد ۸، ۶، ۳، ۲ و ۱۱ را محاسبه کنید.

$$\overline{\mathrm{X}} = \frac{\sum_{\mathrm{i}=۱}^{\mathrm{n}} \mathrm{X}}{\mathrm{n}} = \overline{\mathrm{X}} = \frac{۳۰}{۵} = ۶$$

$$AD = \frac{\sum_{i=۱}^{n}|X_i - \overline{X}|}{N} = \frac{|۸-۶| + |۶-۶| + |۳-۶| + |۲-۶| + |۱۱-۶|}{۵}$$

$$= \frac{۲+۰+۳+۴+۵}{۵} = \frac{۱۴}{۵} = ۲.۸$$

مثال ۵ : انحراف متوسط از میانگین در داده های جدول فراوانی زیر کدام است؟

X_i	۳	۰	۵
F_i	۲	۷	**۳**

$$\bar{X} = \frac{\sum_{i=۱}^{n} F_i X_i}{N} = \frac{۲\times۳+۷\times۰+۳\times۵}{۲+۷+۳} = \frac{۱۵}{۱۲} = ۱.۲۵$$

$$AD = \frac{\sum_{i=۱}^{n} F_i|X_i - \bar{X}|}{N} = \frac{۲|۳-۱.۲۵| + ۷|۰-۱.۲۵| + ۳|۵-۱.۲۵|}{۱۲} = \frac{۳.۵+۱۱.۲۵}{۱۲}$$

$$= \frac{۱۴.۷۵}{۱۲} = ۱.۲۳$$

۳- واریانس (**Variance**)

یکی از مهم ترین شاخص های اندازه گیری پراکندگی مشاهدات حول میانگین، واریانس است. عبارت از میانگین مجموع مربعات انحراف از میانگین یا به اختصار میانگین مربعات می باشد که با V، $\sigma^۲$ و یا MS نشان می دهند. واریانس یکی از شاخصهای مهم پراکندگی است که در تجزیه های آماری زیاد بکار می رود. ولی باید توجه داشت که در این روش مقیاس اندازه گیری دادهها به توان دو میرسد.

الف) واریانس داده های طبقه بندی نشده

- **واریانس جامعه**

فرمول کاربردی $\sigma^۲ = \dfrac{\sum_{i=۱}^{n} X_i^۲ - \dfrac{(\sum_i^n X_i)^۲}{N}}{N}$ فرمول تئوری $\sigma^۲ = \dfrac{\sum_{i=۱}^{n}(X_i - \mu)^۲}{N}$

- **واریانس نمونه**

فرمول کاربردی $S^۲ = \dfrac{\sum_{i=۱}^{n} X_i^۲ - \dfrac{(\sum_i^n X_i)^۲}{n}}{n-۱}$ فرمول تئوری $S^۲ = \dfrac{\sum_{i=۱}^{n}(X_i - \overline{X})^۲}{n-۱}$

$\sigma^۲$ = واریانس جامعه	$S^۲$ = واریانس نمونه و N،n = به ترتیب تعداد کل مشاهدات در نمونه و جامعه

$$SS = \sum_{i=۱}^{n}(X_i - \overline{X})^۲ = \sum_{i=۱}^{n} X_i^۲ - \frac{(\sum_i^n X_i)^۲}{n}$$

$$CF = \frac{(\sum_i^n X_i)^۲}{n} \qquad USS = \sum_{i=۱}^{n} X_i^۲$$

یادآوری می شود در آمار اگر تعداد افراد بیش از ۳۰ تا باشد به آن جامعه بزرگ گویند و لذا برای این موارد باید از فرمول اولی واریانس استفاده می کنیم.

مجموع مربعات تصحیح نشده: USS= Uncorrected Sum Of Square

مجموع مربعات : Sum of Square=SS

میانگین مربعات : Mean of Square =MS= $S^۲$

$$MS = \frac{SS}{df} = \frac{USS - CF}{df}$$

درجه آزادی: $df = n - ۱$

مثال ۶: واریانس مجموعۀ داده های زیر که معرف یک جامعه می باشد را محاسبه کنید.

$X_i = ۳،۲،۵،۶،۹$

$$\mu = \frac{۳ + ۲ + ۵ + ۶ + ۹}{۵} = \frac{۲۵}{۵} = ۵$$

تئوری ←

$$\sigma^۲ = \frac{(۳ - ۵)^۲ + (۲ - ۵)^۲ + (۵ - ۵)^۲ + (۶ - ۵)^۲ + (۹ - ۵)^۲}{۵} = \frac{۳۰}{۵} = ۶$$

کاربردی ←

$$\sigma^۲ = \frac{۳^۲ + ۲^۲ + ۵^۲ + ۶^۲ + ۹^۲ - \frac{(۳ + ۲ + ۵ + ۶ + ۹)^۲}{۵}}{۵} = \frac{۱۵۵ - \frac{۶۲۵}{۵}}{۵} = \frac{۳۰}{۵} = ۶$$

ب) واریانس داده های طبقه بندی شده

$$\sigma^2 = \frac{\sum_{i=1}^{n} F_i (X_i - \mu)^2}{N} \qquad \text{فرمول تئوری} \qquad \sigma^2$$

$$= \frac{\sum_{i=1}^{n} F_i X_i^2 - \frac{\left(\sum_i^n F_i X_i\right)^2}{N}}{N} \qquad \text{کاربردی}$$

مثال ۷: واریانس داده ها با جدول فراوانی زیر کدام است؟

X_i^2	۴	۱	۰	۱
X_i	۲-	۱-	۰	۱
F_i	۲	۳	۱	۴

$$\sigma^2 = \frac{\sum_{i=1}^{n} F_i X_i^2 - \frac{\left(\sum_i^n F_i X_i\right)^2}{N}}{N} = \frac{2 \times 4 + 3 \times 1 + 1 \times 0 + 4 \times 1 - \frac{(-4-3+4)^2}{10}}{10}$$

$$= \frac{15 - \frac{9}{10}}{10} = \frac{\frac{150-9}{10}}{10} = \frac{\frac{141}{10}}{10} = \frac{141}{100} = 1.41$$

خواص واریانس:

۱) با اضافه یک عدد ثابت به مجموعه اعداد یا کم کردن عدد ثابت از مجموعه اعداد، واریانس ثابت می ماند.

$X_i:\{3, 2, 5, 6, 9\} = \mu_{Xi} = 5 \quad \sigma^2 = 6$

مجموعه اعداد بعلاوه ۲:

$$X_d : \{5, 4, 7, 8, 11\} \Longrightarrow \sigma^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} = \frac{275 - \frac{1225}{5}}{5} = \frac{275-245}{5} = \frac{30}{5} = 6$$

$\mu_{Xd} = 5 + 2 = 7$

مجموعه اعداد منهای ۲:

$$X_c : \{1, 0, 3, 4, 7\} \Longrightarrow \sigma^2 = \frac{75 - \frac{225}{5}}{5} - \frac{75-45}{5} = \frac{30}{5} = 6$$

$\mu_{Xc} = 5 - 2 = 3$

$$\begin{cases} D_i = X_{i-}C & \rightarrow \mu_D = \mu_X - C \\ D_{i=}X_i + c & \rightarrow \mu_D = \mu_X + C \end{cases}$$

۲) اگر سری اعداد را در عدد ثابت a ضرب کنیم در این صورت واریانس سری داده های جدید عبارت خواهد بود از حاصلضرب واریانس داده های اولیه در مجذور عدد ثابت ($a^۲$).

مثال ۸ : اگر سری داده های $x_i =$ ۶،۵،۴،۲،۷ را در عدد ثابت ۳ ضرب کنیم. واریانس داده های جدید را محاسبه کنید.

واریانس داده های اصلی:

$$\sigma_X^۲ = \frac{\sum X^۲ - \frac{(\sum X)^۲}{N}}{N} = \frac{۶^۲ + ۵^۲ + ۴^۲ + ۷^۲ - \frac{(۲۲)^۲}{۴}}{۴} = \frac{۱۲۶ - ۱۲۱}{۴} = ۱/۲۵$$

حال واریانس داده های جدید $u_i =$ ۱۸،۱۵،۱۲،۲۱ را محاسبه می کنیم:

واریانس داده های جدید طبق خاصیت دوم واریانس:

$$\sigma^۲ u = a^۲ \sigma^۲ x = ۳^۲ \times ۱.۲۵ = ۹ \times ۱.۲۵ = ۱۱.۲۵$$

واریانس داده های جدید طبق فرمول واریانس:

$$\sigma_u^۲ = \frac{\sum u^۲ - \frac{(\sum u)^۲}{N}}{N} = \frac{۱۱۳۴ - \frac{۴۳۵۶}{۴}}{۴} = \frac{۱۱۳۴ - ۱۰۸۹}{۴} = \frac{۴۵}{۴} = ۱۱/۲۵$$

۳) اگر سری اعداد را بر عدد ثابت a تقسیم کنیم در این صورت واریانس سری داده های جدید عبارت خواهد بود از حاصلضرب واریانس داده های اولیه در معکوس مجذور عدد ثابت ($\frac{۱}{a^۲}$).

مثال ۹: در صورتی که سری اعداد $X_i = 18،30،42،6$ بر عدد ثابت ۶ تقسیم شوند واریانس داده های جدید و قدیم را بدست آورده و چه رابطه ای بین واریانس ها برقرار است؟

واریانس داده های اصلی:

$$\sigma_X^۲ = \frac{\sum X^۲ - \frac{(\sum X)^۲}{N}}{N} = \frac{۱۸^۲ + ۳۰^۲ + ۴۲^۲ + ۶^۲ - \frac{(۹۶)^۲}{۴}}{۴} = \frac{۳۰۲۴ - ۲۳۰۴}{۴} = ۱۸۰$$

داده های تقسیم شده بر عدد ثابت ۶ عبارتنداز : $u_i =$ ۳،۵،۷،۱ و واریانس این اعداد:

طبق خاصیت سوم واریانس :

$$\sigma^۲ u = \frac{۱}{a^۲} \sigma^۲ X = \frac{۱}{٦^۲} \times ۱۸۰ = ۵$$

طبق فرمول واریانس:

$$\sigma_u^۲ = \frac{\sum u^۲ - \frac{(\sum u)^۲}{N}}{N} = \frac{۸۴ - \frac{۲۵۶}{۴}}{۴} = \frac{۸۴ - ۶۴}{۴} = \frac{۲۰}{۴} = ۵$$

۴- انحراف معیار یا انحراف استاندارد (Standard Deviation)

همانطور که گفته شد عیب عمده واریانس این است که واحد آن توان دوم واحد داده ها است. برای مثال، اگر داده‌ها برحسب گرم باشند، واریانس برحسب گرم مربع (گرم به توان ۲) خواهد بود. این موضوع باعث می‌شود که نتوان واریانس را با خود داده‌ها مقایسه کرد. برای رفع این عیب از واریانس جذر میگیریم تا واحد اندازه‌گیری این شاخص پراکندگی با واحد اندازه‌گیری داده‌ها یکی شود که به جذر واریانس انحراف معیار گویند. بنابراین انحراف معیار پراکندگی را با مقیاس داده ها(نه توان دوم) نشان میدهد و مستقیماً با میانگین و شاخص های دیگر قابل مقایسه بوده و باثبات ترین شاخص پراکندگی می باشد. انحراف معیار در آمار توصیفی و استنباطی مورد استفاده فراوانی دارد.

کاربردی $\sigma = \sqrt{\frac{\sum X_i{}^۲ - \frac{(\sum X_i)^۲}{N}}{N}}$ تئوری $\sigma = \sqrt{\frac{\sum (X_i - \mu)^۲}{N}}$

جامعه :

نمونه : کاربردی $S = \sqrt{\frac{\sum X_i{}^۲ - \frac{(\sum X_i)^۲}{n}}{n - ۱}}$ تئوری $S = \sqrt{\frac{\sum (X_i - \bar{X})^۲}{n}}$

مثال ۱۰: اگر مجموعه اعداد $X_i = 3،6،12$ مربوط به جامعه باشد، انحراف معیار را بدست آورید.

$$\mu = \frac{۳ + ۶ + ۱۲}{۳} = \frac{۲۱}{۳} = ۷$$

$$\sigma = \sqrt{\frac{(۱۲ - ۷)^۲ + (۶ - ۷)^۲ + (۳ - ۷)^۲}{۳}} = \sqrt{\frac{۴۲}{۳}} = \sqrt{۱۴} \sim ۳/۷۴$$

$$\sigma = \sqrt{\frac{۱۲^۲ + ۶^۲ + ۳^۲ - \frac{(۱۲ + ۶ + ۳)^۲}{۳}}{۳}} = \sqrt{\frac{۱۸۹ - ۱۴۷}{۳}} = \sqrt{\frac{۴۲}{۳}} = \sqrt{۱۴} \sim ۳/۷۴$$

۵- ضریب تغییرات (Coefficient of Variation)

در مواردی که مقایسه پراکندگی دو یا چند توزیع فراوانی مد نظر است برای این منظور باید دامنه و واحد اندازه گیری داده های مختلف با هم مساوی باشند، بنابراین بایستی واحد اندازه گیری دو انحراف معیار را حذف کرد که برای این کار انحراف معیار هر توزیع را بر میانگین آن تقسیم میکنیم تا واحد اندازه گیری حذف شود.

$$\mathrm{CV} = \frac{S}{\overline{X}} \times ۱۰۰ \quad . \quad \mathrm{CV} = \frac{\sigma}{\mu} \times ۱۰۰$$

مثال ۱۱: در جدول زیر میانگین و انحراف معیار و زن یک گروه از جوانان ۲۰ ساله و یک گروه از نوجوانان ۱۴ ساله گردآوری شده است. میزان تغییرات وزن دو گروه را مقایسه نمایید.

سن	۲۰ساله	۱۴ساله
میانگین وزن	۱۴۵پوند	۳۲پوند
انحراف استاندارد	۱۰پوند	۴پوند

$$\mathrm{CV} = \frac{۴}{۳۲} \times ۱۰۰ = \%۱۲/۵ \qquad \mathrm{CV} = \frac{۱۰}{۱۴۵} \times ۱۰۰ = \%۶.۹$$

بنابراین میزان تغییرات وزن در نوجوانان نسبت به جوانان بیشتر است.

مثال ۱۲: با فرض اینکه داشته باشیم $\sum_{i=۱}^{۳} X_i = ۳$ و $\sum_{i=۱}^{۳} X_i^۲ = ۶$**، ضریب تغییرات را محاسبه کنید.**

$$\mu = \frac{\sum xi}{N} = \frac{۳}{۳} = ۱ \qquad \mathrm{CV} = \frac{\sigma}{\mu} \times ۱۰۰ = \frac{۱}{۱} \times ۱۰۰ = ۱۰۰\%$$

$$\sigma_X^۲ = \frac{\sum X^۲ - \frac{(\sum X)^۲}{N}}{N} = \frac{۶ - \frac{۹}{۳}}{۳} = \frac{۶ - ۳}{۳} = ۱ \implies \sigma = ۱$$

تمرین ۱: در یک مطالعه اطلاعات زیر بدست آمده است ضریب پراکندگی را محاسبه کنید ؟

CV=?   n=۴ $\sum_{i=۱}^{۴} X_i^{۲} = ۴۲۷$ $\mu = ۱۰$

تمرین ۲: انحراف معیار، انحراف متوسط، میانگین حسابی و واریانس مجموعهٔ داده های زیر رامحاسبه کنید.

$X_i =$ ۲۱،۳۶،۱۸،۲۷،۱۴،۱۲

تمرین ۳: در ۴۰ داده آماری مجموع داد هها برابر ۱۰۰ و مجموع مجذورات آ نها ۳۴۰ است .ضریب پراکندگی کدام است؟

تمرین ۴: یک کارخانه تولید لاستیک دو نوع محصول A و B تولید می کند لاستیک نوع A دارای میانگین طول عمر ۲۰۰۰۰ کیلومتر و انحراف استاندارد ۲۰۰۰ کیلومتر می باشد و لاستیک نوع B دارای میانگین طول عمر ۱۸۰۰۰ کیلومتر و انحراف استاندارد۲۰۰کیلومتر است کدام لاستیک برای خرید مناسب است ؟

تمرین ۵: در چامعه ای به حجم N=۵۰ برای صفت متغیر X کمیت های زیر بدست آمده است:

$\sum X = 250 \qquad \sum X^2 = 2500$