ماهان

آمادگی آزمون دکتری

به نام خداوند بخشنده مهربان

آمار و احتمالات مهندسی

مجموعه:

مهندسی منابع آب

مؤلفان:

شهاب کاوه کار

المیرا ولی‌پور

ماهان
مشاوران صعود

کاوه کار ، شهاب (۱۳۶۶)

آمار و احتمالات مهندسی مجموعه مهندسی منابع آب / شهاب کاوه‌کار ـ المیرا ولی‌پور (۱۳۷۰)

تهران ـ مشاوران صعود ماهان: ۱۳۹۹

۲۵۲ص: جدول، نمودار، (آمادگی آزمون دکتری)

ISB/N: 978-600-458-683-2

وضعیت فهرستنویسی: فیپا مختصر

فارسی ـ چاپ اول

۱-آمار و احتمالات مهندسی　　۲- آزمونها و تمرینها

۲- آزمون دوره‌های تحصیلات تکمیلی　　۴-دانشگاه‌ها و مدارس عالی ـ ایران ـ آزمونها

شهاب کاوه‌کار ـ المیرا ولیپور

ج ـ عنوان

شماره کتابشناسی ملی ۳۸۴۳۵۶۱

انتشارات مشاوران صعود ماهان

موسسه آموزش عالی آزاد ماهان

www.mahan.ac.ir

❑ **نام کتاب:** .... آمار و احتمالات مهندسی

❑ **مدیران مسئول:** .... مجید و هادی ستّاری

❑ **مولفین:** .... شهاب کاوه‌کار ـ المیرا ولی‌پور

❑ **مسئول تولید:** .... سمیه بیگی

❑ **ناشر:** .... مشاوران صعود ماهان

❑ **نوبت و تاریخ چاپ:** .... اول /۱۳۹۹

❑ **تیراژ:** .... ۱۰۰۰ نسخه

❑ **قیمت:** .... ۶۶۰/۰۰۰ ریال

❑ **شابک:** ISBN۹۷۸-۶۰۰-۴۵۸-۶۸۳-۲

انتشارات مشاوران صعود ماهان: تهران ـ خیابان ولیعصر، بالاتر از تقاطع ولیعصر مطهری، پلاک ۲۰۵۰

تلفن: ۸۸۱۰۰۱۱۳ و ۸۸۴۰۱۳۱۳

مقدمه ناشر

بنام خدا

ایمان داریم که هر تغییر و تحول بزرگی در مسیر زندگی بدون تحول معرفت و نگرش میسر نخواهد بود. پس بیایید با اندیشه توکل، تفکر، تلاش و تحمل در توسعه دنیای فکریمان برای نیل به آرامش و آسایش توأمان اولین گام را برداریم. چون همگی یقین داریم دانایی، توانایی می‌آورد.

شاد باشید و دلی را شاد کنید

برادران سیاری

۹۴

پیشگفتار

کتابی که پیش روی شماست یک مجموعه کامل از سرفصل‌های موجود در درس تئوری احتمالات و کاربردهای آن می‌باشد. تمام هدف و تمرکز این کتاب بر ارائه روشی نوین در بیان مطالب و کمک به دانشجویان در ایجاد درک عمیقی از موضوعات مورد بحث می‌باشد. برای رسیدن به این هدف، هر فصل از کتاب شامل دو بخش اصلی می‌باشد. در بخش اول، متن هر فصل به صورتی ساده و شیوا بیان شده است. ارائه مثال‌های فراوان در هر قسمت از متن، موجب افزایش سرعت یادگیری و تفهیم مطالب می‌گردد. از آنجاییکه مؤلفین اعتقاد دارند برای کسب مهارت بالا و تسلط کامل بر موضوعات، نیاز به حل مسائل متنوع است، از این رو بخش دوم هر فصل از کتاب شامل تعدادی مسئله در غالب سوالات چهارگزینه‌ای می‌باشد. جهت اثر بخشی بیشتر، اکثر سوالات دارای پاسخ‌های تشریحی می‌باشد. بنابراین به نظر می‌رسد دانشجویان با حل سوالات چهارگزینه‌ای و تکرار آنها می‌توانند به سطح بالایی از مهارت در حل مسائل مختلف تئوری احتمالات دست یابند.

بخشی از سوالات چهارگزینه ای، شامل مسائلی است که در کنکور کارشناسی ارشد و دکتری رشته‌ی مهندسی منابع آب وجود دارد و بخش دیگری از سوالات نیز به وسیله مؤلفین طراحی گردیده است. در طراحی و انتخاب سوالات سعی شده است تا مسائلی انتخاب شوند که از نظر کمک به فراگیری مطالب، به اندازه کافی ارزشمند و قابل تأمل باشند. در هر فصل خلاصه‌ای از مباحث مهم، نکات کلیدی مربوطه، مجموعه سوالات کنکور ارشد سال‌های ۱۳۸۳ تا ۱۳۹۴ و مجموعه سوالات آزمون دکتری منابع آب سال‌های ۱۳۹۱ تا ۱۳۹۴ آورده شده است. در خاتمه از خوانندگان عزیز و دانش‌پژوهان گرامی خواهشمندیم ما را با ارائه پیشنهادها و انتقادهای خود در بهبود کمی و کیفی کارهای انجام شده راهنمایی نمایند تا بتوانیم در آینده کتاب‌هایی با کیفیت بهتر تقدیم حضورشان کنیم.

شهاب کاوه‌کار

المیرا ولیپور

فهرست مطالب

فصل اول

آمار توصیفی

هدف:

انسان درزمینه‌ی همه علوم تجربی، اجتماعی و اقتصادی و در سایه اتخاذ روش صحیح و علمی برای تحقیق به پیشرفت‌های شگرفی دست یافته است. تحقیق به مجموعه فعالیت هایی اتلاق می‌گردد که بر اساس یک نظم فکری و طرح و اندیشه، به منظور کسب آگاهی از واقعیت‌ها تدوین شده اند. روش علمی با مشاهده شروع می‌شود و به دنبال آن مسئله‌ای مطرح می‌گرددو فرض هائی بنا نهاده می‌شوند. در روش علمی مشاهدات در معرض آزمون قرار می‌گیرند و نتایج مورد قضاوت آماری قرار می‌گیرند؛ بنابراین در روش علمی پدیده هائی مطرح هستند که قابل مشاهده و تکرار شدنی باشند. هدف از بررسی‌های آماری عبارت است از: فراهم نمودن مبانی علمی برای تجزیه و تحلیل پدیده‌ها و مساولی که در آنها مشاهدات از قانون علیت انحراف دارند. استدلال استقرائی که یک روش استدلال منطقی است در مورد این نوع مشاهدات دارای کاربرد است و امروزه به طور گسترده‌ای در تحقیقات علمی به کار گرفته می‌شود. برای اینکه مفهوم روش منطقی استدلال روشن گردد، به برخی از مفاهیم مقدماتی اشاره می‌گردد. آمار به عنوان علم و هنر جمع آوری، گروه بندی، توصیف، تجزیه و تفسیر داده‌های حاصل از سر شماری و آزمایش تعریف می‌شود. برای هرگونه محاسبه آماری لازم است که داده‌های آزمایش معمولاً به صورت توده‌ای از اعداد خام هستند به شکل خاصی منظم گردند. در ادامه به اصول و قواعد تنظیم و مرتب کردن داده‌های خام پرداخته می‌شود. در بسیاری از تحقیقات لازم می‌شود که کمیت یا شاخصی را به عنوان معرف یا نماینده کلی داده‌ها و مشاهدات ارئه نمود و از آن در بررسی نتایج تحقیق استفاده کرد. مقایسه توزیع فراوانی یا نمودار فراوانی کافی نیست و برای مقایسه دقیق تر بایستی مقدار یا شاخصی را به عنوان بهترین نماینده ارائه داد. به همین دلیل از شاخص‌های تمایل مرکزی استفاده می‌نماییم. مهم ترین شاخص هایی که بر مرکزیت یا نقطه ثقل داده‌ها دلالت می‌نماید، شامل نما، میانه، پارک ها، میانگین‌ها می‌باشند. با اینکه جداول توزیع فراوانی و شاخص‌های تمایل مرکزی اطلاعات بسیار مفیدی را در خصوص داده‌های تحقیقاتی بیان می‌کنند، ولی تصویر کاملی از ماهیت داده‌ها را در اختیار قرار نمی دهند. از این رو به معرفی شاخص‌های پراکندگی می‌پردازیم. شاخص‌هایی که بر تنوع و پراکندگی دلالت دارند، وضعیت و توزیع مشاهدات را در اطراف شاخص‌های تمایل مرکزی نشان می‌دهند. مهم‌ترین شاخص‌های پراکندگی عبارتند از: دامنه کلی تغییرات، چارک متوسط و انحراف معیار که شاخص اخیر از نظر اصول و مبانی آمار و کاربرد مهمترین آنها می‌باشد.

۱ – ۱ روش‌های استدلال

اصولاً کلیه مسائل استدلالی دو دسته استقرائی و قیاسی تقسیم می‌شوند. استدلال استقرائی حکم از جزء به کل بوده و دارای دو مشخصه است. (۱) نتیجه گیری بر مبنای اطلاعاتی است که به وسیله قسمتی از جامعه کسب می‌گردد و به کل تعمیم داده می‌شود و (۲) نتیجه‌ای که ازآن این استدلال‌ها حاصل می‌گردد صد در صد صحیح نبوده و احتمالی است و وقتی می‌توانیم آن را مورد استفاده قرار دهیم که احتمال صحت نتیجه زیاد باشد. لازمه انجام چنین استدلال هائی پیاده نمودن یک سرس آزمایش و کسب تعدادی مشاهده یا اطلاعات می‌باشد که این همتن بررسی جزء است. سپس بر مبنای نتایج حاصل از تجزیه و تحلیل این اطلاعات، استدلال استقرائی انجام می‌گیرد تا یک کلیت نسبی حاصل شود. در استدلال قیاسی از کل به جزء و یا از اصل به نتیجه و از قانون به موارد اطلاق آن حکم می‌شود. تئوری آمار کاملا قیاسی می‌باشد و روش‌های مختلف را برای استدلال استقرائی فراهم می‌سازد. در استدلال قیاسی یک یا مجموعه‌ای از اصول در دست می‌باشد و هدف تعیین واقعه‌ای است که در تحت شرایط معینی به وقوع می‌پیوندد.

۱ – ۲ تعریف علم و روش تحقیق علمی

علم به مجموعه دانش‌های سازمان یافته‌ای گفته می‌شود که به روش عینی درباره واقعیت‌ها و پدیده‌های طبیعی کسب شده‌اند. روش علمی روشی است که برای پی بردن به روابط علت و معلول بین پدیده‌های طبیعی مورد استفاده قرار می‌گیرد. تحقیق علمی یک تلاش و جدیت فکری است برای استدلال، استنباط و کشف حقیقت نسبی، درباره‌ی موضوع و مسئله‌ای است که مورد بررسی قرار می‌گیرد. مراحل تحقیق علمی عبارتند از:

۱) تعریف موضوع و تعیین حدود آن
۲) تعیین فرضیه ها
۳) تجزیه یا انجام آزمایش
۴) استقراء
۵) تفسیر نتایج و تعمیم آنها

آمار به عنوان علم و هنر جمع آوری، گروه بندی، توصیف، تجزیه و تفسیر داده‌های حاصل از سر شماری و آزمایش تعریف می‌شود. در حال حاضر روش‌های آماری به طور قابل موثر در زمینه‌های علوم زیستی، مهندسی و علوم انسانی به کار می‌روند و در ارتباط با پژوهش‌های کشاورزی نیز جایگاه ویژه‌ای برای خود باز کرده اند. روش‌های آماری را می‌توان به دو بخش آمار توصیفی و آمار استنباطی تقسیم کرد. آمار توصیفی شاخه‌ای از علم آمار است که به ساماندهی، تجزیه و تحلیل، خلاصه سازی و نمایش داده‌ها می‌پردازد. از مهم ترین روش‌های آماری می‌توان روش‌های رسم انواع نمودار، تجزیه‌ی غیر استنباطی داده ها، محاسبه میانگین ها، واریانس‌ها و انحراف استانداردها را نام برد. در مقابل مفهوم آمار توصیفی، شاخه‌ای دیگر از علم آمار که به آمار استنتاجی یا آمار استنباطی مشهور است. اینشاخه از دانش آمار به استنباط، نتیجه گیری و قضاوت درباره یک جامعه آماری از طریق مطالعه و بررسی یک نمونه کوچکتر از آن می‌پردازد. مهم ترین تفاوت میان این دو شاخه از دانش آمار در این نکته است که در آمار استنتاجی، هدف تعمیم نتایج بدست آمده از نمونه به جامعه و قضاوت در مورد آن است، در حالی که در آمار توصیفی نتایج غالباً به جامعه تعمیم داده نمی شوند، مگر آنکه تمام مشاهدات جامعه در تحلیل وجود داشته باشند.

۱ – ۳ جمعیت و نمونه

در آمار استنباطی، جامعه، نمونه و برآورد از جمله مفاهیم اساسی می‌باشند. به مجموعه بزرگی از افراد که دست کم دارای یک صفت مشترک باشند، جمعیت گفته می‌شود. به طور مثال به افرادی که در یک کشور زندگی می‌کنند جمعیت آن کشور و به

مجموعه ژنوتیپ‌های متعلق به یک گونه گیاهی یا جانوری، جمعیت آن گونه گیاهی یا جانوری گفته می‌شود. اگر شمار افراد یک جمعیت مشخص باشد آن جمعیت را محدود می‌نامند؛ مانند جمعیت درختان یک گونه که توسط آفت بخصوصی مورد حمله قرار می‌گیرد. در صورتی که تعداد افراد بی شمار باشد آن جمعیت را نا محدود می‌گویند. به منظور بررسی خصوصیات جمعیت تعدادی از افراد را به عنوان نمونه انتخاب و ارزیابی می‌کنند. به مجموعه‌ای از افراد یا موادی که از یک جمعیت انتخاب می‌شوند نمونه می‌گویند. برا رسیدن به یک نتیجه معتبر، نمونه باید نمایانگر جمعیت مورد نظر باشد. معمولاً یکی از روش‌های مناسب انتخاب نمونه، انتخاب تصادفی است تا کلیه افراد جمعیت شانس مساوی در تشکیل آن داشته باشند. علاوه بر این، هر چه تعداد افراد در نمونه بیشتر باشد، برآورد ویژگی‌های جمعیت دقیق تر خواهد بود. در آمار اگر اندازه نمونه بیشتر از ۳۰ باشد، آن را نمونه بزرگ و اگر کمتر از ۳۰ باشد نمونه کوچک می‌گویند.

۱ – ۴ معیار جمعیت و معیار نمونه

کمیتی که ویژگی‌های جمعیت را نشان می‌دهد معیار جمعیت یا پارامتر نامیده می‌شود. معیار نمونه نیز ویژگی‌های نمونه را نمایان می‌سازد. پارامتر جمعیت کمیتی ثابت ولی ناشناخته است و آنرا از طریق نمونه برداری و محاسبه‌ی معیار نمونه برآورد می‌کنند. مقدار عددی معیار نمونه قابل نوسان است. معیار نمونه را به طور معمول با حروف کوچک لاتین و پارامتر جمعیت را با حروف کوچک یونانی نشان می‌دهند.

مطالعه شاخص‌های آماری در جامعه به علت زیادی تعداد افراد جامعه و حتی فرضی بودن برخی از جوامع امکان پذیر نیست و به طریقی که بحث خواهد شد با توجه به نمونه برآورد می‌گردد. تمام سعی آمار استنباطی در این است که معلوم نماید که تا چه حد این برآورد خوب و مناسب است و می‌توان با توان با توجه به آن اطلاعاتی را در خصوص پارامتر کسب نمود.

۱ – ۵ داده و متغیر

داده‌ها (مشاهدات) اعداد و ارقامی هستند که از اندازه گیری صفت یا متغیر مورد مطالعه در افراد حاصل می‌شوند متغیر عبارت است از عامل یا صفتی است که ارزش آن از فردی به فرد دیگر تغییر می‌کند. اگر مقادیر به طور تصادفی از جمعیت برداشت شوند. ان را متغیر تصادفی می‌نامند، در حالی که اگر یک متغیر به طور غیر تصادفی و توسط پژوهشگر تعیین شود، به آن متغیر ثابت می‌گویند. هنگامی که ارزش متغیر با اعداد و ارقام قابل بیان نباشد و افراد درون جمعیت به چند نوع مجزا تقسیم شوند، در این حالت به آن متغیر کیفی می‌گویند؛ مانند رنگ گل ها. برعکس در متغیر کمی ارزش افراد توسط اعداد به دست آمده مشخص می‌شود. مثل ارتفاع گیاه، تعداد دانه در سنبله گندم. از طرف دیگر متغیرها به دو گروه پیوسته و ناپیوسته تقسیم می‌شوند. به طور معمول متغیر پیوسته مودر اندازه گیری قرار گرفته و افراد می‌توانند هر مقدار ممکن را در دامنه تغییرات (اختلاف دو حد بالا و پایین متغیر) موجود را به خود اختصاص می‌دهند. در این نوع متغیر ها، اختلاف ارزش‌های نزدیک به هم کوچک و غالباً قابل تشخیص نیست؛ به عبارت دیگر ارزش افراد در یک مقیاس پیوسته نوسان می‌کند. صفاتی مانند عملکرد علوفه، درصد چربی شیر در زمره‌ی متغیرهای پیوسته به حساب می‌آیند. در متغیرهای ناپیوسته، ارزش افراد از طریق شمارش به دست می‌آید و قابل اندازه گیری نیست. نظر به اینکه مقادیر حاصل جدا از هم بوده و در دسته‌های مجزا گروه بندی می‌شوند، تغییرات در این نوع متغیرها در مقیاس نا پیوسته به چشم می‌خورد. مثل تعداد گلبرگ‌های یک گل. متغیرهای کمّی می‌توانند به دو صورت پیوسته و ناپیوسته تظاهر یابند. متغیرها را می‌توان به دو گروه وابسته و مستقل نیز متمایز کرد. مقادیر متغیر مستقل معمولاً توسط آزمایش کننده تعیین می‌شود؛ بنابراین متغیر مستقل غالباً متغیر ثابت نیز استو از طرف دیگر مقادیر متغیر وابسته بستگی به سطح انتخابی متغیر مستقل دارد و نظر به اینکه مقدار متغیر وابسته قبل از آزمایش قابل تشخیص نیست، متغیر وابسته همواره یک متغیر تصادفی به حساب می‌آید.

۱ – ۶ انواع مقیاس اندازه گیری

مقیاس‌های اندازه گیری عبارتند از اسمی، رتبه ای، فاصله‌ای و نسبتی. در مقیاس اسمی اعداد را به عنوان اسم به کار می‌برند. این اعداد مقدارکمّی واقعی نیستند و فقط اختلافات را نشان می‌دهند. مثال کلاسیک در این مورد، شماره بازیکنان فوتبال است که هیچ مفهومی به غیر از متمایز کردن بازیکنان در زمین بازی ندارد. از مقیاس اسمی برای گروه‌بندی افراد یا اشیاء استفاده می‌کنند. داده‌های کیفی بر اساس مقیاس اسمی اندازه گیری می‌شوند. در حالی که داده‌های کمّی بر مبنای ۳ مقیاس دیگر قابل اندازه گیری هستند.

مقیاس رتبه‌ای افزون بر اینکه مانند مقیاس اسمی، تفاوت‌ها را نمایان می‌سازد، افرد واشیاء را در طول یک مقیاس کمّی مرتب می‌کند. از این نوع مقیاس می‌توان رتبه‌ی دانشجویان در یک دوره تحصیلی را نام برد. در مقیاس فاصله‌ای علاوه بر اینکه ویژگی‌های دو مقیاس قبلی وجود دارد، فاصله بین داده‌های اندازه گیری یکسان است. به عنوان مثال فاصله بین ۱۰ تا ۲۰ درجه سانتی گراد مانند فاصله بین ۲۰ تا ۳۰ درجه سانتیگراد است. در صورتی که در مقیاس رتبه‌ای ممکن است فاصله دانشجویان رتبه‌ی اول از دوم ۰/۵ نمره ولی فاصله دانشجوی رتبه‌ی دوم از سوم ۱/۵ نمره باشد. در این مقیاس نمی توان صحبتی از نسبت به میان آورد. مقیاس نسبتی کلیه ویژگی‌های ۳ مقیاس قبلی را دارا است. علاوه بر این، امکان گزارش داده‌ها به صورت نسبت وجود دارد.

اطلاع از نوع مقیاس اندازه گیری نقش مهمی در انتخاب روش‌های آماری مناسب برای تجزیه داده‌ها ایفا می‌کند. داده هایی که از طریق مقیاس رتبه‌ای به دست می‌آیند، فاقد توزیع نرمال هستند بنابراین تجزیه آنها باید توسط روش‌های آماری غیر پارامتری انجام شود. داده‌های برخوردار از مقیاس فاصله‌ای و نسبتی می‌توانند دارای توزیع نرمال باشند؛ بنابراین برای تجزیه دادها روش‌های پارامتری مبتنی بر توزیع نرمال به کار می‌روند.

۱ – ۷ قوانین جمع در آمار

مثال ۱ – ۱- رشته اعداد زیر را در نظر بگیرید:

x_1	x_2	x_3	x_4
1	2	3	4

$$\sum_{i=1}^{n} x_i = 1+3+5+7=16$$

حالت‌های دیگر جمع در آمار :

$$\sum_{i=1}^{n} c = nc \leftrightarrow \sum_{i=1}^{5} 4 = 20$$

(۱- ۱)

$$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i \qquad (۲ - ۱)$$

$$\sum_{i=1}^{n}(x_i+y_i)=\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}y_i \qquad (۳ - ۱)$$

مثال ۱ – ۲ - چنانچه $\sum_{i=1}^{10} x_i = 10$ در آن صورت حاصل عبارت $\sum_{i=1}^{10}(2x_i+3)$ عبارت است از :

$$\sum_{i=1}^{10}(2x_i+3) = 2\sum_{i=1}^{10}x_i + \sum_{i=1}^{10}3 = 2\times10+30 = 50$$

بعضی از مشاهدات آماری دارای فراوانی بیشتری هستند. از طرف دیگر منظور از تهیه جداول فراوانی تنها منظم و مرتب کردن داده‌های آماری نیست، بلکه کسب اطلاعاتی در مورد مشاهدات جمع‌آوری‌شده نیز موردنظر می‌باشد. برای دستیابی به این‌گونه اطلاعات از دو نوع فراوانی دیگر به نام‌های فراوانی‌های تراکمی و نسبی استفاده می‌شود.

۱ – ۸ فراوانی تراکمی

فراوانی تراکمی یا تجمعی به فراوانی هر داده و یا هر طبقه به‌اضافه فراوانی کلیه داده‌ها یا طبقه‌های ماقبل یا مابعد گفته می‌شود. بدین منظور در جدول توزیع فراوانی از دومین طبقه یا داده شروع کرده، فراوانی هر یک را با فراوانی طبقه یا داده اول جمع می‌کنیم و مجموع دو فراوانی را در مقابل آن می‌نویسیم و سپس در مقابل طبقه سوم، مجموع فراوانی ساده آن طبقه و فراوانی تراکمی طبقه دوّم را ثبت می‌کنیم و این عمل را تا آخرین طبقه ادامه می‌دهیم. توجه داشته باشید که فراوانی تراکمی آخرین طبقه برابر با مجموع کل فراوانی‌های مطلق یا تعداد کل داده‌های آزمایش یا تحقیق است. بدین ترتیب فراوانی تراکمی صعودی حاصل می‌شود. محاسبه‌ی تراکمی را می‌توان از پایین‌ترین طبقه یا داده شروع نمود. در این صورت فراوانی تراکمی نزولی حاصل می‌شوند. هر فراوانی تراکمی نزولی نشان می‌دهد که چه تعداد از مشاهدات مساوی یا بیشتر از یک مشاهده خاص می‌باشند. به‌عبارت‌دیگر مقدار چند مشاهده حداقل برابر با یک مقدار خاص است. هر فراوانی تراکمی صعودی نشان می‌دهد که چه تعداد از مشاهدات مساوی یا کمتر از یک داده یا رقم موردنظر می‌باشند. به‌عبارت‌دیگر مقدار چند مشاهده حداکثر برابر با یک مقدار خاص بوده است. این مهم‌ترین مفهوم و کاربرد فراوانی تراکمی است.

فراوانی تراکمی را نیز می‌توان به دستگاه مختصات منتقل و نمودارهای چندضلعی، خطی و ستونی آن‌ها را رسم نمود.

۱ – ۹ فراوانی نسبی

فراوانی نسبی یک داده یا طبقه خارج‌قسمت فراوانی مطلق آن داده یا طبقه به تعداد کل مشاهدات یا داده‌ها هست.

$$P_i = \frac{f_i}{N} \qquad (۸ – ۱)$$

چون $\sum f_i = N$ و $\sum P_i = \sum \frac{f_i}{N}$ می‌باشد، مجموع فراوانی نسبی برابر با یک است.

فراوانی نسبی را نیز می‌توان به‌صورت نمودارهای مختلف نشان داد. توجه داشته باشید که عرض و طول هر یک از مستطیل‌ها متناسب با دامنه طبقه‌ها و فراوانی هر طبقه می‌باشند. در تهیه نمودارهای ستونی می‌توان به‌جای ارتفاع مستطیل‌ها مساحت آن‌ها را برابر با فراوانی نسبی قرارداد. مزیت این روش این است که چنان چه دامنه طبقه و بنابراین تعداد مستطیل‌ها را تغییر دهیم، شکل کلی نمودار تغییر چندانی نمی‌کند. خواص نمودار ستونی که بدین روش تهیه می‌شود به شرح زیر است :

۱) احتمال وقوع یک پیشامد برابر با تعداد دفعاتی که در یک آزمایش آن پیشامد به وقوع می‌پیوندد، تقسیم‌بر تعداد کل پیشامدها است. بنابراین فراوانی نسبی طبقه‌ها برابر است با احتمال وقوع پیشامدها در آن طبقه‌ها، که مساوی با مساحت مستطیل‌ها می‌باشد. لذا احتمال اینکه یک یاز مشاهدات را با قرعه انتخاب کنیم و در یک یا چندطبقه معین قرار گیرد برابر با مساحت مستطیل آن طبقه یا مستطیل‌های آن طبقه‌ها می‌باشد.

۲) چون مجموع فراوانی‌های نسبی برابر با یک است، مجموع مساحت کلیه مستطیل‌ها نیز مساوی با یک می‌باشد.

۱ – ۱۰ درصد

شاخص آماری درصد، مفهوم دیگری از فراوانی نسبی با احتمال وقوع پیشامدها را دارد. درصد عبارت است از فراوانی یا تعداد معینی از یک مشاهده که نسبت به تعداد کل مشاهدات سنجیده می‌شود. به‌عبارت‌دیگر در بیان تعداد مشاهدات به‌صورت درصد، عدد ۱۰۰ مبنای مقایسه قرار می‌گیرد. بنابراین

$$P = \frac{f}{N} \times 100 \qquad (۹ – ۱)$$

درصد یکی از ساده‌ترین و مرسوم‌ترین شاخص‌های آماری است. بااین‌وجود خصوصاً در مواردی که تعداد کل مشاهدات کم است، درصد شاخص معتبری نیست. هرچه N کوچک‌تر باشد، درصد شاخص نامطمئن‌تر می‌گردد.

۱ – ۱۱توصیف عددی داده‌ها

در این رویکرد، توابعی روی مشاهدات موجود در نمونه تعریف گردیده و به کمک آن به توصیف خواص نمونه پرداخته می‌شود. این توابع در حقیقت معیارهایی برای بررسی خواص مشاهدات هستند که عمدتاً می‌توان آن‌ها را به سه دسته اصلی تقسیم نمود. این سه دسته به ترتیب معیارهای مکانی (تمایل مرکزی)، معیارهای مقیاسی (پراکندگی) و معیارهای شکل تابع نام دارند.

۱ – ۱۱ – ۱ معیارهای مکانی (تمایل مرکزی)[1]

دسته‌ای از معیارها که بیان‌کننده مکان تمرکز داده‌ها می‌باشند به معیارهای مکانی معروف‌اند. مد، میانه، میانگین حسابی، میانگین هندسی و میانگین هارمونیک برخی از معیارهای مکانی شناخته‌شده می‌باشند که در ادامه به معرفی برخی از آن‌ها پرداخته می‌شود.

نما[2]

مد یا نمای نمونه مقداری است که بیشترین فراوانی یا تکرار را در یک نمونه داشته باشد. معمولاً در دامنه میانی قرار دارد.

در مواقعی که نمی‌توان از میانگین یا میانه استفاده کرد، مد معیار مناسبی است. مثلاً وقتی مشاهدات از نوع اسمی باشند، نمی‌توان میانگین یا میانه محاسبه نمود، اما می‌توان مشاهده‌ای که دارای بیشترین فراوانی است را مشخص کرد.

یک جامعه یا یک نمونه می‌تواند بیش از یک مد داشته باشد. چنان چه یک توزیع آماری صرفاً یک مد داشته باشد ، توزیع تک نمایی[3] نامیده می‌شود. اگر توزیع دو یا چند مد داشته باشد، به ترتیب دو نمایی[4] و چند نمایی[5] نام دارد. اگر جامعه‌ای دو یا چند مد داشته باشد، مد بهترین معیار برای تمرکز داده‌ها است. اگر فراوانی همه مشاهدات با یکدیگر برابر باشند، آنگاه نمونه یا جامعه مد ندارد.

برای محاسبه نما در یک جدول توزیع فراوانی با حدود دسته باید به روش زیر عمل نمود :

---

[1] - Measures Of Location (Central Tendency)
[2] - Mod
[3] -Uni-modal
[4] - bi-modal
[5] -Multi-modal

۱) چنان چه حدود دسته کوچک باشد، نمره متوسط طبقه‌ای که دارای بیشترین فراوانی است به‌عنوان نما در نظر گرفته می‌شود.
۲) در حالتی که حدود دسته بزرگ‌تر باشند برای محاسبه نما باید از فرمول زیر استفاده کرد.

$$mo = L + \left(\frac{d_1}{d_1 + d_2}\right) J \qquad (۱۰-۱)$$

L: حد پایین دسته نما (دسته نما دسته‌ای است که دارای بیشترین فراوانی مطلق می‌باشد.)

d_1 : تفاضل فراوانی دسته نما از دسته قبل

d_2 : تفاضل فراوانی دسته نما از دسته مابعد

J : حدود دسته‌ها

مثال ۱ – ۳:

طبقات	فراوانی
۱۹ – ۲۱	۱
۲۱ – ۲۳	۲
۲۳ – ۲۵	۴
۲۵ – ۲۷	۶
۲۷ – ۲۹	۶
۲۹ – ۳۱	۸
۳۱ – ۳۳	۷
۳۳ – ۳۵	۶
۳۵ – ۳۷	۴
۳۷ – ۳۹	۴
۳۹ – ۴۱	۲

$$mo = 29 + \left(\frac{(8-6)}{(8-6)+(8-7)}\right) \times 2 = 30 / 33$$

میانه

میانه نمونه کمیتی است که نمونه را به دونیمه مساوی تقسیم می‌کند به‌طوری‌که حداقل نیمی از داده‌ها بزرگ‌تر یا مساوی آن و نیمی دیگر کوچک‌تر یا مساوی آن باشند. در حالتی که تعداد مشاهدات فرد باشد،میانه عبارت است از داده وسط و زمانی که تعداد مشاهدات زوج باشد میانه عبارت است از میانگین دو داده وسطی. در هریک از حالات فوق رشته اعداد حتماً باید به صورت صعودی یا نزولی مرتب‌شده باشد. اگر مشاهدات به‌صورت بازه‌هایی در جدول فراوانی مشخص‌شده باشد، از روش دیگری برای محاسبه میانه استفاده می‌شود. گام‌های زیر را برای محاسبه میانه مشاهداتی که در بازه‌های پیوسته هستند ارائه‌شده است :

۱) اگر تعداد کل مشاهدات n در نظر گرفته شود، ابتدا مقدار عددی $0/5n$ محاسبه می‌گردد.

۲) رده یا طبقه میانه تعیین می‌گردد. این رده یا طبقه در جدول فراوانی، کوچک‌ترین رده‌ای است که فراوانی تجمعی مشاهدات آن از $0/5n$ بزرگ‌تر است.

۳) با استفاده از رابطه زیر میانه مشاهدات را به دست می‌آید.

$$me = L + \left(\frac{\frac{N}{2} - F_C}{f_i} \right) J \qquad (۱-۱۱)$$

L: حد پایین دسته میانه

F_C : فراوانی تراکمی دسته ماقبل دسته میانه

f_i : فراوانی مطلق دسته میانه

J : حدود دسته‌ها

مثال ۱- ۴: در مثال ۱–۳ مجموع کل مشاهدات $N = 50$ است درنتیجه $0.5N = 0.5 \times 50 = 25$ می‌باشد. کوچک‌ترین رده‌ای که فراوانی تجمعی آن از ۲۵ بزرگ‌تر است رده $[29-31)$ می‌باشد که مقدار فراوانی تجمعی آن برابر با :

$$1+2+4+6+6+8=27$$

$$Me = 29 + \left(\frac{25-19}{8} \right) \times 2 = 30.5$$

مزیت اصلی میانه به میانگین این است که میانه نسبت به مشاهدات دورافتاده حساس نیست و برای بسیاری از داده‌ها که به‌طور ذاتی شامل نقاط دورافتاده هستند، میانه یک معیار تمرکز مکانی مناسب است.

اگر مشاهدات $x_1, x_2 x_3, \ldots, x_n$ مربوط به یک نمونه باشند، تابع مجموع فواصل مختصاتی به‌صورت $\sum_{i=1}^{n} |x_i - a|$ وقتی مینیمم می‌شود که a برابر با میانه مشاهدات باشد.

چندک[6]

چندک k ام کمیتی است که در اصطلاح k چندم از داده‌ها کوچک‌تر یا مساوی آن هستند. برای مثال صدک هفتم مقداری است که $\frac{7}{100}$ مشاهدات کوچک‌تر یا مساوی آن هستند. برخی از چندک‌ها به دلیل کاربرد فراوان اسامی خاصی دارند. چارک i ام چندکی است که $\frac{i}{4}$ مشاهدات کوچک‌تر یا مساوی آن هستند. صدک یا صده i ام نقطه‌ای است که $\frac{i}{100}$ یا $\%i$ مشاهدات کوچک‌تر یا مساوی آن می‌باشند.

نکته‌ای که باید به آن دقت کنیم این است که چندک‌ها لزوماً بیان‌کننده مکان تمرکز مشاهدات نیستند بلکه مکانی از توزیع را نشان می‌دهند که کسری از مشاهدات کوچک‌تر یا مساوی آن هستند. برای محاسبه هر یک از انواع پارک‌ها در یک جدول توزیع فراوانی با حدود دسته باید به ترتیب زیر عمل نمود :

گام ۱) ابتدا مشاهدات را به‌صورت صعودی مرتب کرده و مشاهدات مرتبه iام را به صورت x_i نشان می‌دهیم.

گام ۲) اگر p نسبت چندک‌ها یا کسر متناظر با آن و n اندازه نمونه باشد، آنگاه کمیت c را از رابطه زیر محاسبه می‌کنیم.

---

[6] Quantile

$$c=(n+1)\,p \qquad (۱-۱۲)$$

گام ۳) فرض کنید قسمت صحیح کمیت c را با r و قسمت اعشاری آن را با w نشان دهیم. در این صورت چندک به صورت زیر به دست می‌آید:

$$(1-w)\,x_r+wx_{r+1}=x_r+w\left[x_{r+1}-x_r\right] \qquad (۱-۱۳)$$

مثال ۱ -۵– سن یازده نفر از مراجعین به کلینک اطفال به صورت 18،17،16،17،15،13،12،14،13،10،11 گزارش شده است. پنجمین هفتک این داده‌ها را حساب کنید.

حل: ۱- مشاهدات را به صورت صعودی مرتب می‌کنیم 10،11،12،13،13،14،15،16،17،17،18

۲- کسر متناظر با پنجمین هفتک را به صورت $p=\frac{5}{7}$ می‌نویسیم و c را به صورت زیر محاسبه می‌کنیم:

$$c=(11+1)\times\frac{5}{7}=\frac{60}{7}=8+\frac{4}{7}$$

۳- با توجه به گام ۲، مقادیر $r=8$ و $w=\frac{4}{7}$ به دست می‌آید. بنابراین داریم:

$$x_8+w\left[x_9-x_8\right]=16+\frac{4}{7}\times[17-16]=16.571$$

برای مشاهداتی که به صورت جدولی نمایش داده شده اند، چندک به صورت زیر محاسبه می‌شود:

گام ۱ – اگر p نسبت چندک یا کسر متناظر با آن و n اندازه نمونه باشد، آنگاه کمیت $c=np$ را به دست آورید. برخی از چارک‌های مهم در زیر آورده شده است.

$$\text{چارک اول}=\tfrac{1}{4}\,\mathrm{n} \qquad \text{چارک سوم}=\tfrac{3}{4}\,\mathrm{n}$$

$$\text{دهک چهارم}=\frac{4}{10}\,\mathrm{n} \qquad \text{صدک بیستم}=\frac{20}{100}\,\mathrm{n}$$

۱) دسته‌ای را پارک مورد نظر در آن قرار می‌گیرد تعیین می‌کنیم. (این دسته عبارت است از اولین دسته‌ای که فراوانی تراکمی آن بیشتر از c باشد).

۲) با توجه به فرمول زیر مقدار پارک مورد نظر را حساب می‌کنیم که در این فرمول داریم :

$$\text{مقدار پارک مورد نظر}=\mathrm{L}+\left(\frac{\mathrm{c}-\mathrm{F_c}}{\mathrm{f_i}}\right)\mathrm{J} \qquad (۱-۱۴)$$

L: حد پایین دستهای که پارک مورد نظر در آن قرار دارد.

F_c: فراوانی تراکمی دسته ماقبل دسته پارک مورد نظر

f_i : فراوانی مطلق دسته پارک مورد نظر

J : حدود دسته‌ها

مثال ۱ – ۶: چارک دوم را حساب کنید؟

دسته	f_i	F_c
2 – 4	3	3
4 – 6	5	8
6 - 8	12	20
8 – 10	10	30
	N=30	

$$\mathrm{p_n}=\frac{2}{4}\mathrm{N}=\frac{2}{4}\times 30=15$$

$$Q_2 = L + \left(\frac{p_n - F_c}{f_i} \right) J$$

$$Q_2 = 6 + \left(\frac{15-8}{12} \right) \times 2 = 7.1$$

میانگین

میانگین مهم ترین شاخص تمایل مرکزی می‌باشد که دارای بیشترین استفاده است. میانگین‌ها دارای انواع مختلفی هستند که هر کدام کاربرد خاص خود را داراست. اما به طور کلی مهم ترین انواع آن‌ها را می‌توان به صورت زیر بیان نمود :

۱) میانگین حسابی
۲) میانگین هندسی یا لگاریتمی
۳) میانگین همساز یا هارمونیک
۴) میانگین وزنی یا هم وزن شده
۵) میانگین متحرک

میانگین حسابی : اگر $x_1, x_2, \ldots, x_n$ یک نمونه n تایی به دست آمده از یک جامعه باشد، آنگاه میانگین حسابی نمونه به صورت زیر تعریف می‌شود.

$$\bar{x} = \frac{\sum_{i=1}^{n} \sum x_i}{n} \qquad (۱۵ - ۱)$$

چنانچه بجای یک نمونه، کل مشاهدات موجود در یک جامعه متناهی با جمعیت N را بررسی کنیم، میانگین حسابی جامعه به صورت زیر به دست می‌آید:

$$\mu = \frac{\sum_{i=1}^{N} \sum x_i}{N} \qquad (۱۶ - ۱)$$

توجه داشته باشید که همیشه با میانگین یک سری داده آزمایش واحد یا مقیاس اندازه گیری نیز بیان می‌شود. وقتی داده‌ها در جدول فراوانی بدون طبقه بندی تنظیم شده باشند از فرمول زیر برای محاسبه میانگین استفاده می‌گردد.

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum fx}{N} \qquad (۱۷ - ۱)$$

در مواقعی که داده‌ها در جدول فراوانی طبقه بندی شده تنظیم شده باشند از فرمول زیر برای محاسبه میانگین استفاده می‌گردد.

$$\bar{x} = \frac{\sum f_i x_c}{\sum f_i} = \frac{\sum f x_c}{N} \qquad (۱۸ - ۱)$$

در این فرمول x_c مقدار متوسط هر طبقه است. چنانچه مشاهدات یک نمونه با x_i نشان داده شده باشد، میانگین نمونه را با $\bar{x}$ نشان می‌دهند.

اگر مقادیر تمامی مشاهدات نمونه به اندازه کمیت ثابتی مانند α افزوده یا کم شوند، در اینصورت مقدار $\bar{x}$ نیز به اندازه α افزوده و یا کم خواهد شد.

$$x_i' = x_i \pm \alpha \quad \Rightarrow \bar{x}' = \bar{x} \pm \alpha$$

اگر مقادیر تمامی مشاهدات نمونه در کمیت ثابتی به اندازه α ضرب شوند، در اینصورت مقدار $\bar{x}$ نیز α برابر خواهد شد.

$$x_i' = \alpha x_i = \quad \Rightarrow \bar{x}' = \alpha \bar{x}$$

وزن هر یک از n مشاهده در میانگین حسابی برابر با $\frac{1}{n}$ است.

ضعف اصلی میانگین حسابی این است که در برابر مشاهدات دور افتاده حساس است. مشاهدات دور افتاده، داده هایی از یک نمونه هستند که به طرز معنی داری از سایر مشاهدات موجود در نمونه دور هستند. این دورافتادگی اگرچه می‌تواند به صورت تصادفی رخ داده باشد، اما غالباً دلیل مشخصی دارد.

مثال ۱ – ۷: میانگین حسابی داده‌های هر یک از مجموعه‌های A و B را بیابید.

$$A = \{1, 2, 5, 5, 8, 10\} \qquad B = \{1, 2, 5, 5, 8, 100\}$$

$$\bar{x}_A = \frac{1+2+5+5+8+10}{6} = 5.166 \qquad \bar{x}_B = \frac{1+2+5+5+8+100}{6} = 20.166$$

این مثال ساده نشان می‌دهد که اگر چه هر دو مجموعه فقط در یک مشاهده تفاوت دارند، اما میانگین‌های انها نزدیک به هم نیستند. این به دلیل وجود یک نقطه دور افتاده با مقدار ۱۰۰ در مشاهدات مجموعه B می‌باشد که میانگین را تحت تأثیر قرار داده است.

میانگین هندسی : چنانچه مشاهدات حاصل از یک بررسی بر حسب درصد، نسبت و امثال اینها باشد، معمولاً میانگین دیگری بنام میانگین هندسی محاسبه می‌شود. میانگین هندسی n داده آزمایشی برابر است با ریشه‌ی n ام حاصل ضرب این داده‌ها و آن را با m_g نشان می‌دهند.

$$\bar{x}_g = \sqrt[n]{x_1.x_2.x_3....} \qquad (۱۹ - ۱)$$

در عمل برای محاسبه میانگین هندسی، ابتدا لگاریتم $\bar{x}_g$ را بدست آورده و سپس آنتی لگاریتم آن را محاسبه می‌نمایند، و به همین جهت به آن میانگین لگاریتمی نیز گفته می‌شود.لذا به جای اینکه میانگین xها گرفته شود، آنتی لگاریتم xها محاسبه می‌گردد.

$$\log \bar{x}_g = \frac{1}{N}(\log x_1 + \log x_2 + ... + \log x_n) = \frac{\sum \log x_i}{N} \qquad (۲۰ - ۱)$$

و در یک جدول توزیع فراوانی با حدود دسته خواهیم داشت:

$$\log \bar{x}_g = \frac{\sum f_i \log x_i}{\sum f_i} \qquad (۲۱ - ۱)$$

برای مثال فرض نمائید که در چهار مرتبه نمونه برداری از مزرعه‌ای میزان خسارت ناشی از یک آفت ۲۵، ۳۰، ۲۴ و ۴۵ درصد بوده است. مقدار متوسط خسارت عبارت خواهد بود از:

$$\bar{x}_g = \sqrt[4]{25\times30\times24\times45} = 30$$

$$\log \bar{x}_g = \frac{1}{4}(\log 25 + \log 30 + \log 24 + \log 45) = 1.4771$$

آنتی لگاریتم $1/4771$ برابر با ۳۰ می‌باشد. یعنی مقدار متوسط خسارت ۳۰ درصد خواهد بود.

یکی از فرض‌های لازم در محاسبات آماری این است که داده‌ها دارای توزیع فراوانی نرمال باشند. با این توزیع در فصل‌های بعد آشنا خواهید شد. چنانچه توزیع داده‌ها نرمال نباشد، امکان استفاده از اکثر فرمول‌های آماری از بین می‌رود، زیرا اصول آمار بر مبنای توزیع نرمال بنا نهاده شده است. توزیع فراوانی نسبت‌ها و درصدها نرمال نیستند. از طرفی ثابت می‌شود که توزیع لگاریتم این داده‌ها و توزیع سینوس معکوس آنها نرمال است. لذا با توضیح مختصر فوق روشن می‌شود که چون درصدها دارای توزیع نرمال نمی باشند،نمی توان میانگین حسابی آنها را محاسبه کرد. بلکه ابتدا با گرفتن لگاریتم، توزیع آنها به نرمال یا تقریباً نرمال تبدیل می‌گردد وسپس میانگین حسابی آنها محاسبه می‌شود.

میانگین هندسی عموماً برای مشاهداتی استفاده می‌شود که هر مشاهده نسبت به مشاهده قبل با یک نرخ تصاعدی از نوع هندسی تغییر کرده باشد. به عبارت دیگر اگر مشاهده x_i نسبت به مشاهده x_{i+1}، d_i برابر شده باشد (d_i قدر نسبت مشاهدات است)، برای محاسبه متوسط نرخ تغییر (میانگین d_i ها) از میانگین هندسی استفاده می‌شود. به عنوان مثال برای محاسبه متوسط بسیاری از مدل‌های رشد، مانن نرخ رشد جمعیت، نرخ رشد تورم و نرخ رشد شاخص‌های مالی از میانگین هندسی استفاده می‌شود.

مثال ۱ – ۸: قیمت کالایی در شش ماه متوالی به صورت جدول زیر تغییر کرده است. متوسط نرخ رشد قیمت این شش ماه چقدر است؟

ماه	۱	۲	۳	۴	۵	۶
قیمت	۱۰۰	۱۲۰	۱۴۵	۱۸۰	۲۲۰	۲۶۰

حل: این مشاهدات در طول زمان به صورت افزایشی زیاد شده اند، بنابراین برای محاسبه متوسط نرخ رشد قیمت از میانگین هندسی استفاده می‌شود. همچنین نرخ رشد در هر ماه، از نسبت قیمت در آن ماه به قیمت در ماه قبل به دست می‌آید(دقت کنید که نرخ رشد در ماه اول را نداریم).

$$\bar{x}_g = (1.20\times1.21\times1.24\times1.22\times1.18)^{\frac{1}{5}} = 1.21$$

میانگین همساز یا هارمونیک : چنانچه مقادیر xها را معکوس نمائیم و میانگین آنها را محاسبه کرده و سپس این میانگین را معکوس کنیم، میانگین همساز یا هامونیک به دست می‌آید. کاربرد این میانگین در مواردی است که شدت و ارزش داده‌ها متفاوت باشد.

$$\bar{x}_h = \frac{1}{\sum \frac{\frac{1}{x_i}}{N}} = \frac{N}{\sum \frac{1}{x_i}} \qquad (۱ - ۲۲)$$

در یک جدول توزیع فراوانی شامل حدود دسته میتوان میانگین هارمونیک را به صورت زیر محاسبه کرد.

$$\bar{x}_h = \frac{1}{\sum \frac{\frac{f_i}{x_i}}{N}} = \frac{\sum f_i}{\sum f_i \frac{1}{x_i}} \qquad (۱ - ۲۳)$$

برای روشن شدن کاربرد میانگین همساز مثالی ذکر می‌گردد. فرض کنید کارگری یک کار معین را در ۳ روز، دومی در ۴ روز و سومی در ۶ روز تمام کند. میانگین تعداد روزهای لازم برای اتمام این کار در این کارگاه چند روز است؟

$$\bar{x}_h = \frac{3}{\frac{1}{3} + \frac{1}{4} + \frac{1}{6}} = 4$$

به طور کلی برای یک گروه از اعداد مثبت رابطه زیر بین سه میانگین وجود دارد.

$$\bar{x}_h \le \bar{x}_g \le \bar{x} \qquad (۱ - ۲۴)$$

میانگین هموزن شده: در برخی از مطالعات ارزش مشاهدات متفاوت است و هر مشاهده دارای ارزش و وزنه خاص خود می‌باشد. در این صورت فرمول محاسبه‌ی میانگین به شرح زیر است :

$$\bar{x}_w = \frac{\sum f_i x_i}{\sum f_i} \qquad (۱ - ۲۵)$$

مثال ۱ – ۹: دانشجویی در یک نیمسال در درس ۳ واحدی آمار نمره ۱۵، در درس ۲ واحدی انگلیسی نمره ۱۴ و در درس یک واحدی تربیت بدنی نمره ۱۷ کسب نموده است. معدل یا میانگین نمرات وی برابر است با :

$$\bar{x}_w = \frac{3(15) + 2(24) + 1(17)}{3 + 2 + 1} = 15$$

وقتی مشاهدات از نسبت دو کمیت مختلف بوجود آمده باشند (مانند سرعت که نسبت مسافت به زمان است) و مقادیر مخرج کسر به ازای مشاهدات مختلف در نمونه متغیر باشند، از میانگین هارمونیک برای محاسبه مکان تمرکز داده‌ها استفاده می‌شود.

مثال ۱ – ۱۰- یک اتومبیل به مدت t ثانیه با سرعت یکنواخت ۹۰ متر بر ثانیه و به مدت t ثانیه دیگر با سرعت یکنواخت ۳۰ متر بر ثانیه در حرکت است. میانگین سرعت این خودرو طی این دو مدت چقدر بوده است؟

حل: اتومبیل در t ثانیه اول مسافتی برابر با x_1 و در t ثانیه بعد، مسافتی برابر با x_2 طی کرده است. چون سرعت در t ثانیه اول با سرعت در t ثانیه دوم برابر نیست، بنابراین مسافت طی شده در این زمان‌ها یکسان نخواهد بود. از طرفی زمان طی دو مسافت مذکور با هم برابر است. در نتیجه می‌توان نوشت:

$$\overline{V}=\frac{x_1+x_2}{t_1+t_2}\Rightarrow\frac{x_1+x_2}{2t}=\frac{1}{2}\left(\frac{x_1}{t}+\frac{x_2}{t}\right)=\frac{1}{2}(V_1+V_2)=\frac{1}{2}(90+30)=60\,{}^{m}\!/_{s}$$

مثال ۱ – ۱۱: در مثال قبلی اگر اتومبیل از مکان مشخص A با سرعت ۹۰ متر بر ثانیه به مکان B حرکت کند، سپس با سرعت ۳۰ متر بر ثانیه از مکان B به مکان A بازگردد، میانگین سرعت این اتومبیل چقدر است؟

حل: در این مثال برخلاف مثال قبل، مسافت ثابت بوده و فاصله‌ای که اتومبیل با سرعت یکنواخت ۹۰ متر بر ثانیه طی می‌کند برابر با فاصله‌ای است که این اتومبیل با سرعت ۳۰ متر بر ثانیه طی می‌کند.در این شرایط سرعت متوسط از رابطه زیر به دست می‌آید.

$$\overline{V}=\frac{x_1+x_2}{t_1+t_2}\quad\Rightarrow\frac{x+x}{t_1+t_2}=\frac{2x}{\frac{x}{V_1}+\frac{x}{V_2}}=\frac{2}{\frac{1}{V_1}+\frac{1}{V_2}}=\left(\frac{1}{2}\left(\frac{1}{V_1}+\frac{1}{V_2}\right)\right)^{-1}$$

$$\left(\frac{1}{2}\left(\frac{1}{90}+\frac{1}{30}\right)\right)^{-1}=45\,{}^{m}\!/_{s}$$

چنانچه دریافتید در مثال ۱- ۱۰ زمانها برابر بودند اما مسافت طی شده متفاوت بود (مخرج‌ها مساوی و صورت‌ها متفاوت) و از میانگین حسابی استفاده شد. اما در مثال ۱ – ۱۱ مسافت طی شده یکسان و زمانها متفاوت بودند که میانگین هامونیک برای محاسبه سرعت متوسط استفاده شد (صورت‌ها مساوی و مخرج‌ها متفاوت).

میانگین متحرک : در مواردی داده‌ها در سری زمانی جمع آوری می‌گردند. به عنوان مثال می‌توان از درجه حرارت یا میزان بارندگی در دوره‌های زمانی معین نام برد. بنابر این میانگین‌های متحرک با رتبه یا ترتیب n برای $x_n,\ldots,x_3x_2,x_1$ برابر است با :

$$\frac{x_1+x_2+\ldots+x_n}{n},\frac{x_2+x_3+\ldots+x_{n-1}}{n},\frac{x_3+x_4+\ldots+x_{n-2}}{n},\ldots \qquad (۲۶-۱)$$

مثال ۱ – ۱۲: جدول زیر میزان بارندگی در روز اول مهر ماه را نشان می‌دهد.

سال	۱۳۶۰	۱۳۶۱	۱۳۶۲	۱۳۶۳	۱۳۶۴	۱۳۶۵	۱۳۶۶	۱۳۶۷	۱۳۶۸
میزان بارندگی (میلی متر)	۱۰	۸	۲۰	۷	۵	۶	۴	۹	۳

میانگین متحرک سه ساله به شرح زیر هستند.

$$\frac{10+8+20}{3},\frac{8+20+7}{3},\frac{20+7+5}{3},\frac{7+5+6}{3},\frac{5+6+4}{3},\frac{6+4+9}{3},\frac{4+9+3}{3}$$

خصوصیات شاخص‌های تمایل مرکزی

۱) در بین شاخص‌های تمایل مرکزی، اهمیت میانگین و آن هم میانگین حسابی در بررسی‌های آماری بیشتر است. تفاوت اساس بین میانگین، میانه و نما این است که میانگین بر حسب مقیاس داده‌ها است و در محاسبه آن فراوانی و کمیت دادها منظور می‌شود. اما میانه و نما شاخص‌ها و داده هائی در بین مشاهدات هستند که تابع ترتیب و فراوانی داده‌ها می‌باشند. لذا میانگین از ترکیب داده‌ها حاصل شده و هر افزایش و یا کاهش در مقدار آن را تغییر خواهد داد. در نقطه مقابل چنانچه افزایش‌ها و یا کاهش‌ها ترتیب داده‌ها تغییر ندهند، در مقدار میانه و مد تغییری حاصل نمی شود.

۲) در یک توزیع فراوانی میانگین مرکز ثقل مشاهدات است و جمع جبری تفاوت داده‌ها نسبت به آن مساوی صفر است. فاصله هر داده آماری از میانگین آن دادها، انحراف آن داده از میانگین و به عبارت ساده تر در ریاضی و آمار انحراف نامیده می‌شود. در یک توزیع فراوانی، میانگین تنها شاخصی است که مجموع انحراف‌ها نسبت به آن مساوی با صفر است. و جزء آن هیچ شاخص دیگری این خصوصیت را دارا نیست.

۳) در یک توزیع فراوانی، چنانچه میانگین یا شاخص‌های دیگری در بین داده‌های آزمایشی در نظر گرفته شوند و انحراف داده‌ها نسبت به هریک از آنها محاسبه گردد، مجموع توان‌های دوّم انحراف‌ها از میانگین کمترین مقدار را خواهد داشت. در آمار به مجموع توان‌های دوّم انحراف داده‌ها از میانگین به اختصار مجموع مربعات گفته می‌شود.

۴) بین میانگین، میانه و مد روابط ساده‌ای برقرار است که در مقایسه با توزیع‌های فراوانی و تبدیل یک شاخص به شاخص دیگر مورد استفاده قرار می‌گیرند. در نمودارها و جداول متقارن هر سه شاخص مقدار متوسط با هم مساوی هستند.

۱ – ۱۱ – ۲ معیارهای مقیاسی (پراکندگی)

معیارهای پراکندگی یا مقیاسی معیارهایی می‌باشند که میزان پراکندگی حول پارامترهای مکانی یک توزیع را اندازه گیری می‌کنند. انحراف معیار، واریانس، میانگین قدر مطلق انحرافات، دامنه، دامنه بین چارک و ضریب تغییرات از معیارهای پراکندگی معروف می‌باشند.

دامنه کلی تغییرات داده‌ها

این شاخص از تفاضل بزرگ ترین و کوچک ترین مشاهده به دست می‌آید و درباره پراکندگی مشاهدات یک دید کلی به ما می‌دهد.در واقع دامنه تغییرات فقط اختلاف کمترین و بیشترین داده آزمایشی را نشان می‌دهد، در صورتی که مفهوم واقعی پراکندگی به فاصله داده‌ها از شاخص‌های تمایل مرکزی و فراوانی آنها بستگی دارد.

$$R = X_{Max} - X_{Min} \qquad (۲۷-۱)$$

دامنه به وجود نقاط دور افتاده به شدت حساس است و در صورت وجود نقطه دور افتاده مقدار دامنه به صورت غیر طبیعی بزرگ می‌شود. همچنین دامنه معیاری است که از اطلاعات موجود در تمام مشاهدات استفاده نمی کند و فقط به اطلاعات کوچکترین و بزرگترین مشاهده نیازمند است. بنابراین ممکن است نماینده خوبی از پراکندگی مشاهدات نباشد. اما مزیت آن استفاده از آن می‌باشد.

چارک متوسط

همانگونه که گفته شد دامنه کلی تغییرات در سلسله داده‌ها تنها بر مبنای دو مشاهده اول و آخر محاسبه می‌گردد، تنوع و پراکندگی داده‌های دیگر در آن تأثیری ندارد. بنابراین می‌توان شاخص پراکندگی را نسبت به شاخص‌های تمایل مرکزی از

جمله میانه ، میانگین محاسبه کرد. به طور کلی هر چه مشاهدات به شاخص‌های تمایل مرکزی نزدیک تر باشند، پراکندگی کمتری دارند. یکی از شاخص‌های پراکندگی که پراکندگی را نسبت به میانه نشان می‌دهد و از دامنه کلی تغییرات، اعتبار آماری بیشتری دارد، فاصله چارکی است. فاصله چارکی تفاضل چارک اول از چارک سوم می‌باشد. معمولاً مقدار فاصله چارکی را به دو تقسیم کرده و آن را چارک متوسط می‌نامند. چارک متوسط شاخص پراکندگی خاص میانه است و با آن مورد استفاده و تفسیر قرار می‌گیرد.

فاصله چارکی : $Q_3 - Q_1$ و چارک متوسط : $\frac{Q_3-Q_1}{2}$ (۲۸ – ۱)

انحراف متوسط

در یک توزیع فراوانی هر مشاهده یا مساوی با میانگین است و یا از آن بیشتر و یا کمتر می‌باشد. در بند ۲ خصوصیات شاخص‌های تمایل مرکزی با مفهوم انحراف داده‌ها از میانگین آشنا شدید. همانگونه که ذکر شد جمع جبری این انحراف‌ها مساوی صفر است. هر چه داده‌ها نسبت به میانگین پراکنده تر باشند، انحراف‌ها نیز بیشتر خواهند بود. چنانچه قدر مطلق انحراف‌ها جمع گردند و بر تعدادشان تقسیم گردند، انحراف متوسط و یا به عبارت دیگر میانگین قدر مطلق انحراف‌ها به دست خواهد آمد. لذا انحراف متوسط، شاخص پراکندگی نسبت به میانگین است. از آنجائی که میانگین با اعتبارترین شاخص تمایل مرکزی است، انحراف متوسط نیز از دامنه کلی تغییرات و فاصله چارکی معتبرتر است. فرمول انحراف متوسط به شرح زیر است.

$$\mathrm{AD} = \frac{\sum |x_i - \mu|}{N} \qquad (۲۹ – ۱)$$

در یک توزیع متقارن حدود ۸۰ ٪ از مشاهدات در فاصله یک انحراف متوسط بیشتر و کمتر از میانگین قرار می‌گیرند درمحاسبه انحراف متوسط باید توجه داشت که قدر مطلق انحرافات مثبت از میانگین بررب است با قدر مطلق انحرافات منفی.

مثال ۱ – ۱۳- انحراف متوسط داده‌های آماری زیر را محاسبه نمائید :

6	5	4	3	2	x_i
1	2	2	2	1	f_i

$$\mu = \frac{2+3(2)+4(2)+5(2)+6}{8} = 4$$

$$\mathrm{AD} = \frac{|2-4|+2|3-4|+2|4-4|+2|5-4|+|6-4|}{8} = 1$$

انحراف معیار

انحراف معیار مهم ترین و معتبرترین شاخص آماری برای نشان دادن پراکندگی مشاهدات نسبت به میانگین است. این شاخص علاوه بر اینکه پراکندگی مشاهدات را به طور دقیق اندازه گیرب می‌کند، در محاسبه و تفسیر بسیاری از شاخص‌های دیگر آماری نیز مورد استفاده قرار می‌گیرد. انحراف متوسط در آمار توصیفی و آمار استنباطی موارد استفاده فراوان دارد. انحراف معیار همانند انحراف متوسط بر مبنای انحراف مشاهدات از میانگین حاصل می‌گردد و متوسط پراکندگی مشاهدات در اطراف میانگین را نشان می‌دهد.در بین انحراف داده‌ها از میانگین، بعضی مثبت و برخی منفی می‌باشند و جمع جبری آنها مساوی

صفر می‌گردد. برای حصول شاخص پراکندگی، محاسبه میانگین قدر مطلق این انحراف‌ها پیشنهاد شده که به آن انحراف متوسط اطلاق گردید. روش دیگر برای حذف علامت جبری انحراف ها، محاسبه میانگین توان‌های دوم انحراف‌ها است که به آن واریانس گفته می‌شود. با این روش نیز مقیاس اندازه گیری به توان دو می‌رسد، ولی می‌توان از میانگین حاصل جذر گرفت و آن را به مقیاس اصلی برگردانین. این شاخص آماری را انحراف معیار گویند. بنابراین انحراف معیار عبارت است از جذر واریانس که خود میانگین مجموع مربع انحراف‌ها می‌باشد.

$\frac{2}{3}$ از داده‌ها در فاصله یک انحراف معیار، بیشتر و کمتر از میانگین قرار می‌گیرد.

$$\mu \pm \sigma \quad \to \frac{2}{3}\ \mathrm{N}$$

$$\mu \pm 1/64\sigma \quad \to 90\%\ \mathrm{N}$$

$$\mu \pm 1/96\sigma \quad \to 95\%\ \mathrm{N}$$

$$\mu \pm 2/58\sigma \quad \to 99\%\ \mathrm{N}$$

$$\mu \pm 3\sigma \quad \to 99/7 = 100\%\ \mathrm{N}$$

در توزیع‌های متقارن و نرمال دامنه تغییرا در حدود ۶ انحراف معیار است.

$$r = 6\sigma \qquad (۳۰-۱)$$

واریانس

واریانس یکی از معروف ترین معیارهای اندازه گیری پراکندگی مجموعه‌ای از داده‌ها می‌باشد.اگر $x_n, \ldots, x_2, x_1$ مقادیر به دست آمده از یک نمونه یا جامعه باشند، آنگاه واریانس را می‌توان از رابطه زیر به دست آورد:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \qquad \text{واریانس نمونه} \qquad (۳۱-۱)$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} \qquad \text{واریانس جامعه} \qquad (۳۲-۱)$$

وقتی مشاهدات در جدول فراوانی نشان داده شده باشند، واریانس را می‌توان از رابطه زیر به دست آورد:

$$S^2 = \frac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{n-1} \qquad \text{واریانس نمونه} \qquad (۳۳-۱)$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} f_i(x_i-\mu)^2}{N} \qquad \text{واریانس جامعه} \qquad (۳۴-۱)$$

که در رابطه فوق μ میانگین جامعه،$\bar{x}$ میانگین نمونه و f_i نیز فراوانی رده i ام در جدول است. رابطه واریانی را می‌توان به صورت‌های دیگری نیز نوشت که برخی ازآنها به صورت زیر می‌باشند:

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n-1} = \frac{\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}{n-1} = \frac{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2}{n-1} \qquad (۳۵-۱)$$

یکی از اشکالات واریانس این است که واحد آن با واحد مشاهدات یکسان نیست. به عنوان مثال اگر مشاهدات دارای واحد متر باشند، واحد واریانس مجذور متر خواهد بود. این مسأله گاهی باعث می‌شود این معیار تعبیر عملی ملموسی نداشته باشد. به عنوان مثال اگر واحد مشاهدات مجذور اینچ $(inch)^2$ باشد، واحد واریانس $(inch)^4$ خواهد بود کهدر حال حاضر در دنیای واقعی معنی ندارد. اما مزیت عمده واریانس نمونه‌ای این است که خواص اماری بسیار مناسبی دارد.

به مجموع مربع انحراف‌ها از میانگین، یعنی $\sum f x^2$، به اختصار مجموع مربعات (SS) گفته می‌شود. که خود نیز از شاخص‌های مهم آماری است. مجموع مربعات، مقدار کلی پراکندگی داده‌ها را در مقیاس توان دوم بیان می‌کند و واریانس، متوسط این مقدار است. به همین جهت واریانس را میانگین مربعات (MS) نیز می‌گویند. هر کدام از این سه شاخص (واریانس، مجموع مربعات و انحراف معیار) یک شاخص پراکندگی محسوب می‌شوند. مزیت اصلی انحراف معیار در آن است که پراکندگی را به مقیاس اصلی داده‌ها نشان می‌دهد و مستقیماً با میانگین و شاخص‌های دیگر قابل مقایسه است.

چنانچه برای سهولت محاسبات، داده‌های حاصل از آزمایش از طریق تفریق کوچک شوند، واریانس داده‌های جدید با واریانس داده‌های اصلی تفاوتی نخواهد داشت. چون مبنای محاسبه واریانس انحراف داده‌ها از میانگین است، لذا چنانچه داده‌ها از طریق تفریق یا جمع تغییر داده شوند، میانگین آنها نیز به همان نسبت تغییر داده می‌کند و انحراف‌ها ثابت باقی می‌ماند.

چنانچه کلیه مشاهدات در ضریب ثابت c ضرب شوند و یا بر آن تقسیم شوند، مقدار واریانس به اندازه c^2 تغییر می‌کند.

واریانس تفاوت‌ها و مجموع‌ها

توزیع فراوانی تفاوت‌ها یکی از توزیع‌های مرسوم در آمار است. از این توزیع در مقایسه تفاوت بین جوامع، که موضوع متداول در تحقیقات است، استفاده می‌شود.

s_i = مجموع دو متغیر y_i و x_i $\qquad {\sigma_{s_i}}^2 = {\sigma_{x_i}}^2 + {\sigma_{y_i}}^2 + 2\,Cov_{x,y}$ (۱ – ۳۶)

D_i = تفاضل دو متغیر y_i و x_i $\qquad {\sigma_{D_i}}^2 = {\sigma_{x_i}}^2 + {\sigma_{y_i}}^2 - 2\,Cov_{x,y}$ (۱ – ۳۷)

Cov : کواریانس بیانگر ماهیت و شدت ارتباط بین آنهاست. چنانچه مقدار Cov مثبت باشد بیانگر وجود رابطه مستقیم و در صورتی که مقدار آن منفی باشد بیانگر وجود رابطه معکوس بین x و y می‌باشد.

$$\mathrm{cov}_{x,y} = \frac{\sum (x-\bar{x})(y-\bar{y})}{n-1} \qquad (۱ – ۳۸)$$

$$\mathrm{cov}_{x,y} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{n-1} \qquad (۱ – ۳۹)$$

ضریب پراکندگی یا تنوع

در مواردی مقایسه پراکندگی دو یا چند توزیع فراوانی مورد نظر است. برای این مقایسه در صورتی می‌توان از واریانس یا انحراف معیار استفاده نمود که دامنه و واحد اندازه گیری داده‌های مختلف با هم مساوی باشند. دلیل این امر این است که انحراف معیار، متناسب با دامنه تغییرات و واحد مشاهدات است. وقتی مقیاس سنجش یا اندازه گیری متفاوت باشد باید از شاخص دیگری به نام ضریب پراکندگی استفاده کرد.

اگر $x_1, x_2, \ldots, x_n$ مقادیر به دست آمده از یک نمونه باشند، آنگاه ضریب تغییرات نمونه از رابطه زیر به دست می‌آید:

$$C.V = (\frac{S}{\bar{x}}) \qquad (۴۱-۱)$$

ضریب تغییرات جامعه نیز به سهولت با جایگزینی σ به جای S و μ به جای $\bar{x}$ به دست می‌آید.

ضریب پراکندگی نشان می‌دهد که میزان پراکندگی چند درصد میانگین است. چنانچه از رابطه ضریب تغییرات پیداست، این معیار بدون واحد است. در نتیجه برای مقایسه پراکندگی دو مشخصه آماری که واحدهای یکسانی ندارند، می‌توان از این معیار استفاده کرد. همچنین برای مقایسه پراکندگی نسبی دو مشخصه که میانگین‌های متفاوتی دارند نیز از این معیار استفاده می‌شود.

مثال ۱ – ۱۴: میانگین و انحراف معیار وزن و قد یک گروه از افراد به صورت جدول زیر است. به نظر شما کدامیک از این دو مشخصه (قد و وزن) پراکندگی بیشتری دارد؟

	قد (اینچ)	وزن (پوند)
$\bar{X}$	۶۹	۱۴۵
S	۱۲	۲۵

حل: چون دو مشخصه قد و وزن واحدهای متفاوتی دارند، بهترین شاخص برای اندازه گیری پراکندگی ضریب تغییرات است.

$$C.V_1 = \frac{12}{69} = 0.1739 \qquad C.V_2 = \frac{25}{145} = 0.1724$$

بنابراین پراکندگی نسبی قد و وزن در این گروه تقریباً یکسان است (پراکندگی نسبی قد کمی بیشتر از وزن است).

نمره استاندارد یا نمره سیگمائی

وقتی مقیاس داده‌ها و دامنه تغییرات انها در دو یا چند توزیع فراوانی مشابه نباشد، داده‌ها و مشاهدات را نیز می‌توان با هم مقایسه نمود. برای این منظور نیز باید مشاهدات را به اعداد نسبی و فاقد واحد اندازه گیری تبدیل کرد. این عدد نسبی را نمره معیار یا داده سیگمائی یا نمره Z می‌نامند. بنابراین :

$$Z = \frac{X-\mu}{\sigma} \qquad (۴۲-۱)$$

بررسی نمره Z را پس از آشنا شدن با منحنی نرمال به صورت آموزنده تر و دقیق تر انجام خواهیم داد. توجه داشته باشید که در مورد یک سری مشاهده نمره سیگمائی ارقام بیشتر از میانگین مثبت و نمره سیگمائی ارقام کمتر از میانگین منفی است و نمره سیگمائی ارقام مساوی با میانگین مساوی صفر است.

۱ – ۱۱ – ۳ معیارهای شکل تابع

گاهی اوقات دو توزیع مختلف معیارهای مکانی و پراکندگی یکسانی دارند، اما شکل آنها کاملاً با هم متفاوت است. ب عبارت دیگر گشتاور اول و دوم این توزیع‌ها یکسان بوده اما گشتاورهای بالاتر انها با هم فرق می‌کند. وجود معیارهایی که بتواند گشتاورهای بالاتر را اندازه گیری کند می‌تواند در بسیاری از تحلیل‌ها مفید باشد. در این بخش دو معیار چولگی و کشیدگی معرفی می‌شوند که به عنوان بهترین معیارهای شکل تابع شناخته می‌شوند.

ضریب چولگی

اگر $x_1, x_2, \ldots, x_n$ یک نمونه n تایی از یک جامعه باشند، آنگاه چولگی نمونه به صورت زیر تعریف می‌شود.

$$b_1 = \frac{m_3}{(m_3)^{\frac{3}{2}}} \qquad (۴۰-۱)$$

در رابطه بالا m_k یا گشتاور $\mathbf{k}$ ام نمونه به صورت زیر تعریف می‌شود.

$$m_k = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^k}{n} \qquad (۴۱-۱)$$

چولگی معیاری مناسب برای بررسی قرینگی توزیع می‌باشد. به این ترتیب که هر گاه توزیع مشاهدات قرینه باشد، مقدار چولگی صفر است. وقتی توزیع مشاهدات دمی در طرف راست خود داشته باشد، چولگی مقداری مثبت خواهد داشت. در این حالت توزیع اصطلاحاً چوله به راست یا چوله مثبت نامیده می‌شود. همچنین وقتی توزیع مشاهدات دمی در طرف چپ خود داشته باشد چولگی مقداری منفی خواهد داشت و در این حالت توزیع اصطلاحاً چوله به چپ یا چوله منفی نامیده می‌شود.

- وقتی یک توزیع تک نمایی قرینه باشد، آنگاه رابطه $\mu = me = mo$ برقرار است.
- وقتی توزیع چوله به راست باشد، آنگاه رابطه $\mu > me > mo$ برقرار است.
- وقتی توزیع چوله به چپ باشد، آنگاه رابطه $\mu < me < mo$ برقرار است.

برای محاسبه ضریب چولگی روابط زیر نیز ارائه شده است که عبارتند از:

$$\mathrm{SK} = \frac{\mu - mo}{\sigma} \qquad (۴۲-۱)$$

$$SK = \frac{3(\mu - me)}{\sigma} \qquad (۴۳-۱)$$

$$SK = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} \qquad (۴۴-۱)$$

$$SK = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}} \qquad (۴۵-۱)$$

ضریب کشیدگی

اگر $x_1, x_2, \ldots, x_n$ یک نمونه $\mathbf{n}$ تایی از یک جامعه باشند، آنگاه کشیدگی نمونه به صورت زیر تعریف می‌شود:

$$b_2 = \frac{m_4}{(m_2)^2} - 3 \qquad (۴۶-۱)$$

علت کم کردن عدد ۳ در رابطه بالا این است که نسبت گشتاور چهارم توزیع نرمال به مجذور گشتاور دوم آن برابر ۳ است و در واقع مقدار کشیدگی داده‌ها نسبت به توزیع نرمال سنجیده می‌شود. همچنین مقدار کشیدگی بر حسب میانگین و میانه و فاصله چارکی نیز محاسبه می‌شود که به شرح زیر است:

$$K = \frac{\frac{1}{2}(Q_3 - Q_1)}{P_{90} - P_{10}} \qquad (۴۵ - ۱)$$

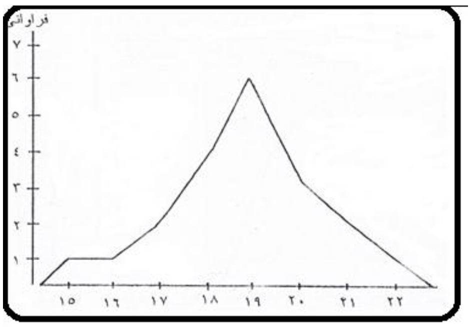
در توزیع نرمال که توزیع قرینه‌ای است و میانگین، مد و میانه در آن بر هم منطبق می‌باشند، ضریب کشیدگی برابر با ۰/۲۶۳ می‌باشد.

۱ – ۱۲ نمایش گرافیکی داده ها

در این رویکرد، داده‌ها به کمک نمودارها، اشکال و گراف‌های شناخته شده نمایش داده می‌شوند. بطوریکه این اشکال بتوانند به طرز مفیدی نتایج و تحلیل‌های مربوط به نمونه را ارئه دهند. از بین ابزارهای نمایش گرافیکی داده‌ها نمودار شاخه و برگ و هیستوگرام معروف تر و پرکاربرد تر هستند که در این بخش به شناسایی آنها می‌پردازیم.

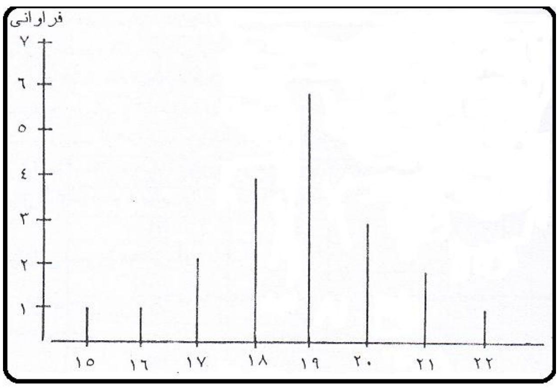
نمایش هندسی داده‌ها در جدول توزیع فراوانی را نمودار فراوانی گویند. بدین منظور در محورهای مختصات، داده‌های آزمایش را در طول محور (x) و فراوانی‌ها در عرض محور (y) نشان می‌دهیم. نمودارهای فراوانی چند نوع هستند که در اینجا مهم ترین آنها مورد بحث قرار می‌گیرند.

برای رسم نمودار چند ضلعی یا منحنی فراوانی، هر مشاهده و یا چنانچه مشاهدات طبقه بندی شده باشند، نمره‌ی متوسط هر طبقه و فراوانی آن طبقه را در دستگاه مختصات منتقل نموده و سپس نقاط حاصل را بهم متصل می‌سازیم. ضمناً نمره‌ی متوسط طبقه‌های قبل و بعد از طبقه‌های انتهائی را با فراوانی صفر مشخص می‌نمائیم.



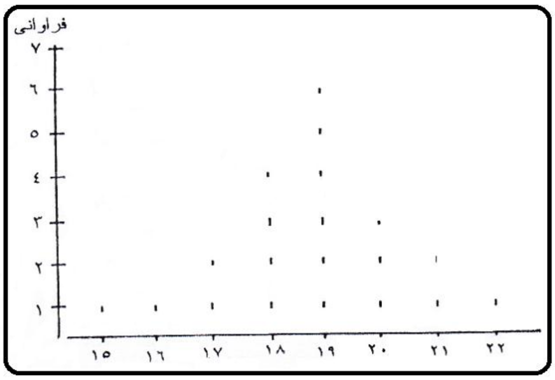
شکل ۱ – ۱ نمودار چند ضلعی

چنانچه از یک خط عمود بر محور xها در نقطه مربوط به آن داده آماری استفاده شود، نمودار خطی به دست می‌آید. طول خط رسم شده در بالای هر x_i مبین فراوانی آن می‌باشد.



شکل ۱ – ۲ نمودار خطی

روش دیگر مورد استفاده این است که بجای هر یک از خطوط از یک یا چند نقطه که تعداد آنها برابر با تعداد دفعاتی است که یک داده آماری مشاهده می‌شود، استفاده گردد. در این صورت تعداد دفعاتی که هر مشاهده اتفاق می‌افتد به سهولت قابل تشخیص است.



شکل ۱ – ۳ نمودار نقطه‌ای

نموادار شاخه و برگ

نمودار شاخه و برگ روشی برای نمایش داده‌ها است که در آن شاخه‌ها بیانگر دامنه مقادیر داده‌ها و برگ‌ها بیانگر رقم بعدی مشاهدات هستند. معمولاً وقتی اندازه نمونه داده‌ها خیلی کم یا خیلی زیاد نباشند از نمودار شاخه و برگ بیشتر استفاده می‌شود.

نحوه ترسیم نمودار شاخه و برگ

گام ۱- هر x_i را به دو بخش تقسیم می‌کنیم. شاخه، که شامل یک یا چند رقم اول است و برگ که شامل سایر رقم‌های موجود در داده است.

گام ۲- شاخه‌ها را در یک ستون به ترتیب می‌نویسیم.

گام ۳- مقادیر برگ‌های متناظر با هر شاخه را در جلوی شاخه مورد نظر قرار می‌دهیم.

مثال ۱ – ۱۵- داده‌های زیر نشان دهنده مقاومت کششی یک نوع آلیاژ مرکب از آلومینیوم و لیتیم است. نمودار شاخه و برگ مربوط به این مشاهدات را رسم کنید.

۷۱	۹۳	۶۶	۸۱	۶۱
۷۳	۸۴	۹۵	۸۸	۷۸
۸۳	۸۷	۷۰	۷۸	۸۴
۶۴	۶۵	۶۳	۷۵	۹۲
۸۸	۷۷	۷۱	۶۵	۷۹

دو نمودار شاخه و برگ مختلف مربوط به مشاهدات مقاومت کششی آلیاژ در زیر نشان داده شده است. در شکل سمت راست، رقم دهگان به عنوان شاخه و رقم یکان به عنوان برگ در نظر گرفته شده است. در شکل سمت چپ، هر رقم دهگان به دو بخش تبدیل شده است. به عنوان مثال $6l$ به معنی مشاهداتی است که رقم دهگان آنها ۶ و رقم یکان آن کمتر یا مساوی ۵ است. از جدول داده‌ها می‌توان دریافت که اعداد ۶۱ ، ۶۳ و ۶۴ اعداد شاخه ۶ هستند که رقم یکان انها کمتر از ۵ است و در شاخه $6l$ مشخص شده اند.

شاخه	برگ
$6L$	1 3 4
$6U$	5 5 6
$7L$	0 1 1 3
$7U$	5 7 8 8 9
$8L$	1 3 4 4
$8U$	7 8 8
$9L$	2 3
$9U$	5

شاخه	برگ
۶	1 3 4 5 5 6
۷	0 1 1 3 5 7 8 8 9
۸	1 3 4 4 7 8 8
۹	2 3 5

در انتخاب شاخه‌های مورد نظر باید دقت کرد. اگر تعداد شاخه‌ها زیاد باشد، برگ‌های کمی در هر شاخه وجود خواهد داشت. در اینصورت نمودار اثر بخشی لازم را ندارد. همچنین اگر تعداد شاخه‌ها بیش از اندازه کم باشد، برگ‌ها بسیار زیاد خواهد شد و تفاوت بین شاخه‌ها آشکار نمی شود. بهتر است از یک سطح با تعداد شاخه کم شروع و تا جایی ادامه داد که نمودار شکل مناسبی به خود بگیرد.به عنوان مثال شکل نمودار سمت راست شکل واضحی ندارد زیرا تعداد شاخه آن کم است، اما نمودار سمت چپ به اندازه کافی شاخه دارد.

نمودار هیستوگرام

هیستوگرام ابزری برای نمایش فراوانی داده‌ها می‌باشد که کاربرد فراوانی در تعیین خصوصیات مربوط به مشاهدات، شکل توزیع و نوع آنها دارد. در هر هیستوگرام دامنه مشاهدات به فواصلی تقسیم می‌شود که معمولاً رده، کلاس یا بازه نامیده می‌شوند. رده‌ها یا کلاس‌ها معمولاً فواصل مساوی دارند، که در این صورت تفسیر آنها راحت تر خواهد بود.

نحوه ترسیم نمودار هیستوگرام

گام ۱- دامنه مشاهدات را به دست آورید و با تقسیم آن بر تعداد رده یا کلاس مورد نظر، طول هر کلاس را محاسبه می‌کنیم.

گام ۲- حد بالا و پایین هر کلاس را مشخص نموده، فراوانی نسبی مربوط به هر کلاس را به دست می‌آوریم.

گام ۳- مستطیل هایی به طول فراوانی نسبی و به عرض فاصله هر کلاس ترسیم می‌نمائیم.

از هیستوگرام معمولاً به عنوان یک روش ابتدائی برای تعیین توزیع مجموعه‌ای از مشاهدات استفاده می‌شود. همه معیارهای مکانی، پراکندگی و شکل توزیع در هیستو گرام مشخص است و از این رو کاربرد فراوانی در خلاصه سازی داده‌ها دارد.