بسم الله الرحمن الرحیم

درس آماروااحتمالات

استاد: اکرم سلطان پور

فصل اول: کلیات

آمار: مجموعه ای از روش ها برای جمع آوری، تنظیم، خلاصه کردن داده ها، طبقه بندی آن ها و روش های تحلیلی برای انجام استنباط، پیش بینی، برآورد، نتیجه گیری و تصمیم گیری در شرایط مختلف ارائه می دهد.

تعریف آمار

روش علمی است که برای جمع آوری ، تلخیص ، تجزیه و تحلیل ، تفسیر و بطور کلی برای مطالعه و بررسی مشاهدات بکار گرفته می شود

از فنون آماری در مدیریت برای چه مقاصدی استفاده می شود ؟

۱ – برای تبدیل داده ها به اطلاعات

۲– برای بررسی صحت و سقم فرضیات

۳– برای تعیین اعتبار و پایایی تحقیقات

- **تعریف جامعه:** جامعه بزرگترین مجموعه از موجودات است که در یک زمان معین ، مطلوب ما قرار می گیرند. مثل جامعه فرهنگیان ایران و . . .

لازم به ذکر است جامعه متناسب با هدف شما تغییر می کند. به طور مثال اگر هدف شما بررسی رضایتمندی کارکنان بانک مرکزی باشد، جامعه مورد نظر شما تمام کارکنان بانک مرکزی می باشد.

اگر هدف شما بررسی رضایتمندی معلمان باشد جامعه شما به صورت تمام معلمان تعریف می شود و..

- **جامعه آماری:**

هر مجموعه از اشیا یا افراد که لااقل دارای یک صفت مشترک باشند را جامعه آماری گویند. جامعه آماری دانشجویان دانشگاه آزاد ارومیه، جامعه اتومبیل های سواری، جامعه ستارگان

- **انواع جامعه آماری**

۱- **محدود :** یعنی جامـعه مقادیر از تعـداد محدود و ثابتی تشکیل شده و پایان پذیر باشد. (کارکنان بانک مرکزی)

۲- **نا محدود :** یعنی جامعه از یک ردیف بی انتهایی از مقادیر تشکیل شده باشد. (تمام برگ های درختان)

- **صفت مشخصه:** صفتی است که بین همه عناصر جامعه اماری مشترک و متمایز کننده جامعه اماری از سایر جوامع باشد.

صفت متغیر: خاصیتی است که افراد جامعه را از یکدیگر جدا و مشخص می کند برای جامعه دانشجویان دانشگاه آزاد اسلامی واحد ارومیه صفت متغیر می تواند: سن، قد، وزن، گروه خونی و... باشد.

صفات متغیر دو نوع است: ۱- صفات کمی ۲- صفات کیفی

صفت کمی: صفتی که قابل اندازه گیری و شمارش است مانند سن، قد، وزن و...
صفت کمی خود ۲ نوع است: کمی پیوسته، کمی ناپیوسته

صفت کمی پیوسته: صفتی است که مجموعه مقادیر خود را از مجموعه اعداد حقیقی اختیار می کند مانند وزن که هر عددی مثلا بین بازه [50, 70] اختیار کند.

صفت کمی ناپیوسته: صفتی است که مجموعه مقادیر خود را از مجموع اعداد طبیعی یا معادله هم عرض آن اختیار می کند مانند تعداد فرزندان خانواده ها.

صفت کیفی: صفتی است که قابل اندازه گیری و شمارش نیست مانند جنسیت، رشته تحصیلی، گروه خونی

صفت کیفی دو نوع است. ۱- کیفی اسمی ۲- کیفی ترتیبی

۱- **کیفی اسمی:** این صفت ها قابل شمارش نیست و صرفا جهت شناسایی می باشد مانند جنس، نژاد، رشته تحصیلی، شماره شناسنامه، گروه خونی
۲- **کیفی ترتیبی:** این صفت ها قابل شمارش نیست و صرفا جهت شناسایی و بیان برتری به کار می رود مانند مقاطع تحصیلی، درجه نظامی، مهارت دانشجویان

➢ **تعریف نمونه:** نمونه عبارتست از تعداد محدودی از آحاد جامعه آماری که بیان کننده ویژگی های اصلی جامعه باشد . (نمونه انتخاب شده باشد تا حد ممکن شبیه جامعه مورد نظر باشد. برای بررسی یک روش آموزش درست نیست فقط دانشجویان با معدل بالا را انتخاب کرد بلکه انتخاب دانشجویان از هر طیف معدل الزامی است.) (نمونه: زیرمجموعه یا بخشی از جامعه آماری مورد مطالعه است که بتوان نتایجی را در مورد جامعه استخراج نمود)

اندازه گیری جامعه برای به دست آوردن برخي از شاخص هاست

➢ **انواع شاخص های آماری**

۱- **پارامتر** : شاخص هایی که از طریق سرشماری (اندازه گیری تمامی عناصر جامعه آماری) بدست می آیند.. (محاسبه متوسط درآمد کارکنان بانک مرکزی با استفاده از اندازه گیری درآمد تمام کارکنان دولت.)

۲- **آماره** : شاخص هایی که از طریق نمونه گیری (اندازه گیری بخشی از جامعه) بدست می آیند. (محاسبه متوسط درآمد کارکنان بانک مرکزی با استفاده از اندازه گیری درآمد نمونه ای از کارکنان دولت.)

سیر تحول علم آمار از نظر موضوعی عبارت است از:

۱- آمار توصیفی: آمار توصیفی شامل روش هایی است که برای خلاصه کردن و رده بندی و دسته بندی داده های موجود در مجموعه ای از داده ها، محاسبه مشخصات عددی این مجموعه و نمایش داده ها در قالب نمودارها، جداول و شکل های مختلف به کار می رود.

۲- آمار استنباطی: شامل روش هایی است که با استفاده از آن ها، اطلاعات موجود در نمونه را به کل جامعه تعمیم می دهیم.

آمار استنباطی شامل برآورد نقطه ای، برآورد فاصله ای و آزمون فرض ها می باشد.

۳- آمار ناپارمتری: این آمار در مقابل آمار پارامتریک بیان می شود. فرض اساسی در آمار پارامتریک برخوردار بودن مشاهدات از توزیع نرمال است. در حالی که در فنون ناپارامتریک این فرض ضرورتی ندارد. و برای مشاهدات فاقد توزیع آماری کاربرد دارد.

مراحل پژوهش علمی در آمار

۱- مشخص کردن هدف

۲- جمع آوری داده ها

۳- تجزیه و تحلیل داده ها

۴- بیان یافته ها

روش های ناپارامتریک

آزمون هایی که مشروط به مفروضات آمار کلاسیک نیستند و کاربرد اصلی آنها در بررسی جوامع آماری غیر نرمال ، جوامع با داده های کیفی و نمونه های کوچک آماری می باشد

دو عنصر اصلی تحقیقات رفتاری و مدیریتی

۱- فرضیه های تحقیق

۲- متغیرهایی که برای آزمودن آنها بکار گرفته می شوند

نقش متغیرها در فرضیات

متغیرها ، فرضیه ها را بصورتی نشان می دهند که محققان رفتاری و مدیریتی بتوانند آنها (فرضیه ها) را مشاهده و اندازه گیری نمایند

➢ انواع متغیرها

۱- متغیر خصیصه

۲- متغیر مستقل

۳- متغیر وابسته

۴- متغیر تعدیل کننده (واسطه ای)

۵- متغیر کنترل

- **متغیر خصیصه**

متغیری که مقدار آن از یک فرد به فرد دیگر و یا از یک عضو به عضو دیگر جامعه آماری ممکن است تغییر کند . مثل

اندازه سازمان، قد افراد، رنگ چشم افراد ، نظر آنها در مورد یک موضوع خاص و ..

- **متغیر مستقل**

به علت احتمالی یا فرضی متغیر وابسته ، متغیر مستقل یا متغیر درونداد و به عبارتی محرک گفته می شود .

مثال: فرض کنید می خواهیم تاثیر مصرف شیر را روی افزایش قد افراد بررسی کنیم در این حالت مصرف شیر روی افزایش قد تاثیر دارد ولی افزایش قد روی مصرف شیر تاثیر نداردبنابراین متغیر مستقل مصرف شیر است.

- **متغیر وابسته**

به متغیری که به تبع تغییر متغیر مستقل ، مقدارش کم و زیاد می شود متغیر وابسته ، متغیر پاسخ و یا برونداد اطلاق می شود. مثال: در مثال قبل افزایش قد از آنجایی که وابسته به مصرف شیر است به عنوان متغیر وابسته شناخته می شود.

- **متغیر تعدیل کننده (واسطه ای)**

متغر ثانوی است که پژوهشگر می خواهد تاثیر آن را در متغیر مستقل اولیه و متغیر وابسته ملاحظه کند. این متغر بدین منظور انتخاب می شود که روشن شود آیا این متغیر ، رابطه بین متغیر مستقل و متغیر وابسته را تحت تأثیر قرار می دهد یا نه.

- **متغیر کنترل**

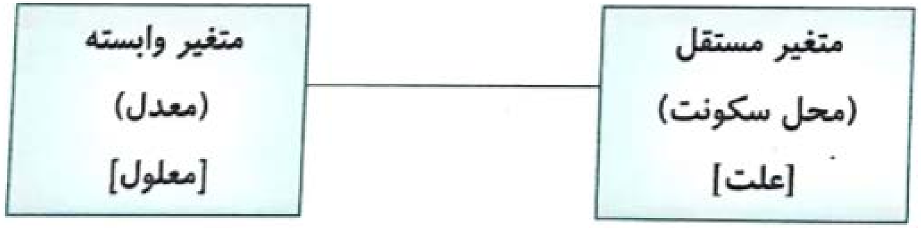
به متغیرهایی که در موقع انجام پژوهش ، لازم است تأثیر آنها خنثی شده و یا از بین برود، متغیرهای کنترل می گویند.

فرق متغیر تعدیل کننده با متغیر کنترل

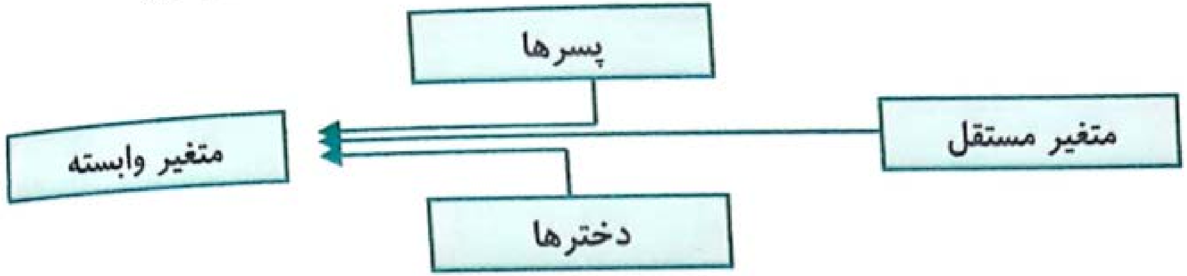
موقع انجام تحقیق ، پژوهشگر سعی می کند تأثیرات متغیر کنترل را از بین ببرد ولی تأثیرات متغیر تعدیل کننده را مورد بررسی قرار می دهد.

مثال مفهومی: برای درک بیشتر متغیرها بهتر است به مثال زیر توجه کنید:

فرض کنید می‌خواهیم تأثیر محل سکونت را بر معدل دانشجویان بررسی کنیم. برای این‌کار معدل دانشجویان روزانه رشته مدیریت را می‌یابیم. می‌دانیم که عده‌ای از دانشجویان در خوابگاه و عده‌ای خارج از خوابگاه زندگی می‌کنند و همچنین عده‌ای از دانشجویان دختر و عده‌ای پسر هستند. در این تحقیق متغیر **محل سکونت** دانشجویان (خوابگاهی، غیرخوابگاهی)، **متغیر مستقل** است و متغیری که مستقیماً تحت تأثیر متغیر مستقل است یعنی **«معدل دانشجویان»**، **متغیر وابسته** است.



متغیری که بنا بر فرض، اثر متغیر مستقل بر متغیر وابسته را تغییر می‌دهد، یعنی **جنسیت** (دختر و پسر) متغیر **تعدیل‌کننده** است؛ مثلاً ممکن است محل سکونت در معدل دختران تأثیر بیشتری داشته باشد تا در معدل پسران.



اما متغیری که در طول تحقیق ثابت نگه داشته می‌شود و تأثیرش در طول تحقیق خنثی است، یعنی **دانشجوی روزانه رشته مدیریت** بودن، متغیر **کنترل** است.

توجه کنید که در این تحقیق، **معدل** دانشجویان متغیر **خصیصه** نیز است.

توزیع‌های فراوانی

در این بخش توزیع‌های فراوانی داده‌های کیفی و کمی را به طور جداگانه مورد بررسی قرار می‌دهیم.

الف) توزیع فراوانی داده‌های کیفی: فرض کنید X یک صفت کیفی با حالات E_i $(i=1,\ldots,k)$ باشد. مثلاً X صفت کیفی مربوط به سنجش دقت دانشجویان در مطالعه باشد و حالات آن ضعیف، متوسط، خوب و عالی است. همچنین فرض کنیم f_1 نشان‌دهنده تعداد افرادی باشد که دارای حالت E_1 می‌باشند ,... , f_k نشان دهنده تعداد افرادی باشد که دارای حالت E_k می‌باشند و اگر تعداد کل افراد برابر n باشد، باید داشته باشیم:

$$\sum_{i=1}^{k} f_i = f_1 + f_2 + \ldots + f_k = n$$

حال اگر جدولی به صورت زیر، متشکل از دو ستون که یکی نشان‌دهنده حالات ودیگر نشان‌دهنده تعداد افرادی که دارای آن حالت هستند را تشکیل دهیم، چنین جدولی را جدول توزیع فراوانی می‌نامیم.

حالات	f_i
E_1	f_1
E_2	f_2
$\vdots$	$\vdots$
E_k	f_k

مثال: پزشکی ۲۰ بیمار قلبی دارد که گروه خونی آن‌ها عبارتند از: B , B , AB , O , AB , O , A , A , O , O , A , A , A , O , AB , O , A , B , O , AB می‌خواهیم جدولی فراوانی داده‌های فوق را تشکیل دهیم.

E_i (گروه خونی)	f_i (تعداد بیماران)
A	۶
B	۳
AB	۴
O	۷
	۲۰

ب) توزیع فراوانی داده‌های کمی: داده‌های کمی بر دو نوع می‌باشند، داده‌های کمی گسسته و داده‌های کمی پیوسته.

جدول توزیع فراوانی داده‌های گسسته: مانند جدول توزیع فراوانی داده‌های کیفی است.

۱–۵. مثال: فرض کنید می‌خواهیم ۲۰ کارگر یک کارخانه را از نظر نوع فعالیت که یک صفت کیفی و بار دیگر از نظر تعداد خانوار که یک صفت کمی گسسته است، طبقه بندی کنیم. جدول توزیع فراوانی را در هر دو حالت به صورت زیر تشکیل می‌دهیم.

حل: فرض کنید ۵ نفر مکانیک، ۶ نفر نجار، ۷ نفر برق کار و ۲ نفر لوله کش مشغول فعالیت باشند.

E_i	f_i
مکانیک	۵
نجار	۶
برق کار	۷
لوله کش	۲
	۲۰

حال فرض کنید تعداد خانوار ۲ نفره ۳ و تعداد خانوار ۳ نفره ۵ و تعداد خانوار ۴ نفره ۶ و تعداد خانوار ۵ نفره ۵ و تعداد خانوار ۶ نفره، ۱ خانوار باشد. جدول توزیع فراوانی به صورت زیر بیان می‌شود.

x_i	f_i
۲	۳
۳	۵
۴	۶
۵	۵
۶	۱
	۲۰

که در آن x_i نشان‌دهنده تعداد افراد خانوار است.

جدول توزیع فراوانی داده‌های پیوسته: این داده‌ها نتیجه اندازه‌گیری با مقیاس نسبی هستند. مانند وزن افراد، طول قد و یا طول عمر لامپ و غیره که به آن‌ها داده‌های پیوسته گفته می‌شود و از آن جا که مقادیر ممکن، به طور پیوسته تغییر می‌کنند، دیگر درباره فراوانی یک مقدار خاص صحبت نمی‌کنیم برای مثال فرض کنیم نمونه‌ای مرکب از ۲۰۰ دانشجو را در نظر بگیریم و قد هر یک را بر حسب سانتی متر ثبت کنیم. X می‌تواند مقادیر ۱۷۵ , ۱۷۶ , ۱۷۷ سانتی‌متر و غیره و یا هر مقدار بین آن‌ها از قبیل ۱۷۵/۲۴ باشد. هرگز دوباره کسی را مشاهده نخواهیم کرد که قدش ۱۷۵/۲۴ باشد. بنابراین، برای مثال ، نمی‌توانیم بگوییم که چهار نفر دارای قد ۱۷۵ هستند. زیرا قد این افراد در فاصله (۱۷۵/۵ — ۱۷۴/۵) قرار دارد. لذا در مورد داده‌های پیوسته مجبور هستیم برای طبقات، فاصله‌ای را منظور کنیم و مراحل زیر را انجام دهیم.

۱) انتخاب طبقات
۲) جای دادن اطلاعات آماری در طبقات انتخاب شده
۳) شمارش عناصر قرار گرفته در هر طبقه

باید توجه داشت که مسأله اصلی در ساختن هر توزیع فراوانی، مرحله اول، یعنی انتخاب طبقه بندی مناسب می‌باشد که با مثالی مراحل فوق را توضیح می‌دهیم.

۱–۶. مثال: داده‌های زیر اندازه‌های قد ۱۰۰ جوان بیست ساله در یکی از شهرهای ایران می‌باشند که بر حسب سانتی‌متر تا نزدیکترین واحد سر راست شده‌اند.

۱۵۱	۱۷۷	۱۷۲	۱۷۸	۱۷۱	۱۶۴	۱۵۰	۱۷۰	۱۸۴	۱۷۲
۱۷۰	۱۵۳	۱۵۹	۱۶۹	۱۶۴	۱۶۶	۱۷۲	۱۶۹	۱۶۵	۱۵۸
۱۵۶	۱۷۲	۱۶۵	۱۵۱	۱۶۲	۱۷۴	۱۸۲	۱۵۹	۱۵۴	۱۶۲
۱۶۵	۱۵۰	۱۵۶	۱۶۴	۱۷۱	۱۷۵	۱۸۴	۱۷۷	۱۷۰	۱۶۳
۱۵۴	۱۶۶	۱۶۹	۱۷۴	۱۷۵	۱۷۶	۱۷۲	۱۶۲	۱۶۹	۱۷۲
۱۵۶	۱۷۱	۱۶۴	۱۶۱	۱۵۹	۱۵۲	۱۶۹	۱۸۰	۱۶۷	۱۶۲
۱۷۰	۱۷۱	۱۶۴	۱۵۰	۱۷۰	۱۸۴	۱۷۲	۱۶۵	۱۷۰	۱۶۳
۱۷۲	۱۶۹	۱۷۷	۱۷۳	۱۵۴	۱۶۲	۱۷۰	۱۵۳	۱۷۷	۱۸۴
۱۶۵	۱۶۵	۱۷۳	۱۷۹	۱۷۷	۱۷۹	۱۶۹	۱۷۴	۱۶۴	۱۶۶
۱۸۳	۱۶۰	۱۶۱	۱۵۹	۱۶۲	۱۵۱	۱۵۸	۱۵۳	۱۶۳	۱۵۷

در این داده‌ها عدد ۱۵۰ کوچک‌ترین و ۱۸۴ بزرگترین داده هستند. چون داده‌ها تا نزدیکترین واحد سرراست شده‌اند، می‌توان گفت که اندازه واقعی قدها در فاصله [۱۸۴/۵ , ۱۴۹/۵] قرار دارند.
طول این فاصله، یعنی ۳۵ = ۱۴۹/۵ − ۱۸۴/۵ را دامنه تغییرات داده‌ها می‌نامیم و به چند فاصله مساوی مثلاً ۵ فاصله هر یک به طول ۷ سانتی‌متر تقسیم می‌کنیم. هر کدام از فاصله‌های کوچک، مثلاً [۱۵۶/۵ , ۱۴۹/۵] را یک طبقه (دسته) با طول واقعی ۷ می‌نامیم. عدد ۱۴۹/۵ را حد یا مرز پایین و عدد ۱۵۶/۵ را حد (مرز) بالای این طبقه می‌خوانیم. تعداد طبقه‌ها باید طوری انتخاب شوند که در هر طبقه یک یا چند داده داشته باشیم و معمولاً تعداد طبقه‌ها نباید از ۵ کمتر یا از ۲۰ بیشتر باشد. گاهی به جای مثلاً [۱۵۶/۵ , ۱۴۹/۵] می‌نویسیم [۱۵۷ , ۱۵۰] و در این حالت ۱۵۰ را کرانه پایین و ۱۵۷ را کرانه بالای این طبقه می‌خوانیم که در این حالت طبقه بعدی به صورت [۱۶۴ , ۱۵۷] خواهد بود و باید توجه کرد که کرانه بالای هر طبقه به آن تعلق ندارد، به استثنای طبقه آخر. گام دوم، جای دادن اطلاعات آماری در طبقات انتخاب شده است. بدین منظور، برای هر عدد در مقابل طبقه‌ای که در آن قرار می‌گیرد، یک خط – نشان (چوب خط) می‌زنیم. معمولاً پنجمین خط – نشان را به صورت مورب منظور می‌کنیم تا امر شمارش را ساده‌تر نماید.

فراوانی (مطلق): فراوانی هر طبقه، یعنی تعداد دفعات تکرار داده‌ها در آن طبقه که آن را با نماد f نشان می‌دهیم.

فراوانی f_i	خط نشان	حدود طبقات	حدود واقعی
۱۵	卌 卌 卌	۱۵۰–۱۵۷	۱۴۹/۵–۱۵۶/۵
۲۰	卌 卌 卌 卌	۱۵۷–۱۶۴	۱۵۶/۵–۱۶۳/۵
۳۰	卌 卌 卌 卌 卌	۱۶۴–۱۷۱	۱۶۳/۵–۱۷۰/۵
۲۵	卌 卌 卌 卌 卌	۱۷۱–۱۷۸	۱۷۰/۵–۱۷۷/۵
۱۰	卌 卌	۱۷۸–۱۸۵	۱۷۷/۵–۱۸۴/۵
۱۰۰			جمع

۱–۷. تعریف: اگر فراوانی هر طبقه را بر فراوانی کل تقسیم کنیم، فراوانی نسبی به دست می‌آید که با نماد r_i نشان می‌دهیم. مثلاً فراوانی نسبی طبقه دوم برابر $r_2 = \frac{20}{100} = \frac{1}{5}$ است. اگر فراوانی نسبی هر طبقه را در عدد ۱۰۰ ضرب کنیم، درصد فراوانی نسبی آن طبقه به دست می‌آید.

فراوانی تجمعی: مجموع فراوانی هر طبقه با طبقات بالا را فراوانی تجمعی نامیده و آن‌را با نماد F_i نشان می‌دهیم. برای مثال فراوانی تجمعی طبقه سوم «F_3» برابر است با: $F_3 = f_1 + f_2 + f_3 = 30 + 20 + 15 = 65$. لذا بنابه تعریف فراوانی تجمعی $F_5 = 100$، طبقه آخر با فراوانی کل، برابر خواهد بود.

نماینده طبقه (نشان دسته): میانگین کران بالا و کران پایین هر طبقه را نماینده آن طبقه نامیده و با نماد x_i نشان می‌دهیم.(کران بالا طبقه i ام + کران پایین طبقه i ام) $x_i = \frac{1}{2}$

تعاریف بیان شده در بالا را در مثال ذکر شده، به صورت زیر ملاحظه می‌کنید.

طبقه	x_i	f_i	r_i	F_i
۱۴۹/۵–۱۵۶/۵	۱۵۳	۱۵	۰/۱۵	۱۵
۱۵۶/۵–۱۶۳/۵	۱۶۰	۲۰	۰/۲۰	۳۵
۱۶۳/۵–۱۷۰/۵	۱۶۷	۳۰	۰/۳۰	۶۵
۱۷۰/۵–۱۷۷/۵	۱۷۴	۲۵	۰/۲۵	۹۰
۱۷۷/۵–۱۸۴/۵	۱۸۱	۱۰	۰/۱۰	۱۰۰
		۱۰۰	۱/۰۰	

مثال ۲ـ۳ (مقیاس ترتیبی) رضایت شغلی ۲۴ نفر را جویا شدیم که نتایج به این صورت است.

راضی، نسبتاً راضی، ناراضی، راضی، راضی، بی‌نظر، ناراضی، راضی، نسبتاً ناراضی، راضی، راضی، ناراضی، نسبتاً ناراضی، نسبتاً راضی، بی‌نظر، بی‌نظر، نسبتاً ناراضی، راضی، بی‌نظر، ناراضی، راضی، راضی، راضی، ناراضی

طبقه‌بندی داده‌ها به این صورت خواهد بود.

سطح رضایت	راضی	نسبتاً راضی	بی‌نظر	نسبتاً ناراضی	ناراضی	جمع
تعداد	۱۰	۲	۴	۳	۵	۲۴

توجه: تعداد طبقات (K) را مشخص می‌سازیم. علی‌رغم اینکه برای تعیین تعداد طبقات هیچ‌گونه روش دقیق ریاضی وجود ندارد، برای محاسبه آن می‌توان از این دو فرمول تجربی استفاده کرد. فرمول اول که به وسیله «استورجس» ارائه شده عبارت است از:

(۳ـ۳)
$$\boxed{K = 1 + 3/32\,\mathrm{Log}N} \quad \text{تعداد طبقات}$$

در این فرمول N تعداد داده‌های جمع‌آوری شده از جامعه است. فرمول دوم که باز یک روش تقریبی است، عبارت است از:

(۳ـ۴)
$$\boxed{K = \sqrt{N}} \quad \text{تعداد طبقات}$$

زمانی که تعداد داده‌ها زیاد باشد فرمول ۳ـ۳ جواب بهتری ارائه می‌کند. در هر صورت تعداد طبقات به تجربه پژوهشگر بستگی دارد و این فرمولها پیشنهادات ابتدایی به دست می‌دهند.

مرحله سوم. فاصله طبقات را محاسبه می‌کنیم.

$$\boxed{I = \frac{R}{K}} \quad \text{فاصله طبقات}$$

مثال ۵ـ۳ داده‌های مرتب شده ذیل مربوط به سود سالانه ۵۰ شرکت برحسب ۱۰ میلیون ریال است:

۱۰	۲۱	۳۲	۳۸	۵۰	۵۳	۶۴	۷۳	۸۵	۹۸
۱۰	۲۴	۳۳	۴۹	۵۱	۵۴	۶۴	۷۳	۸۷	۱۰۰
۱۴	۲۵	۳۴	۴۹	۵۱	۵۷	۶۵	۷۵	۸۹	۱۰۶
۱۵	۲۷	۳۵	۵۰	۵۱	۶۰	۶۸	۷۷	۹۲	۱۱۰
۱۷	۳۰	۳۷	۵۰	۵۱	۶۳	۷۰	۸۰	۹۵	۱۲۰

مرحله اول. دامنه تغییر را محاسبه می‌کنیم:

$$R = ۱۲۰ - ۱۰ = ۱۱۰$$

مرحله دوم. تعداد طبقات را با استفاده از رابطه ۴ـ۳ محاسبه می‌کنیم:

$$K = \sqrt{N} \Rightarrow K = \sqrt{۵۰} = ۷/۰۷۱ \Rightarrow K = ۸$$

مرحله سوم. فاصله طبقات را محاسبه می‌کنیم:

$$I = \frac{۱۱۰}{۸} = ۱۳/۷۵ \Rightarrow I = ۱۴$$

مرحله چهارم. براساس $K = ۸$ و $I = ۱۴$ جدول طبقه‌بندی ۳ـ۳ را تشکیل می‌دهیم و داده‌ها را سازماندهی می‌کنیم. در این مثال عرض طبقات یک واحد کمتر از طول طبقات است. از آنجا که طبقات نسبت به همدیگر ناسازگارند، برای شمارش، حدود طبقات بسته در نظر گرفته می‌شود.

طبقه i	سود سالانه (C-L)	خط و نشان	فراوانی مطلق f_i	فراوانی تجمعی
۱	۱۰ـ۲۳	卌 I	۶	6
۲	۲۴ـ۳۷	卌 IIII	۹	15
۳	۳۸ـ۵۱	卌 卌	۱۰	25
۴	۵۲ـ۶۵	卌 III	۸	33
۵	۶۶ـ۷۹	卌 I	۶	39
۶	۸۰ـ۹۳	卌	۵	44
۷	۹۴ـ۱۰۷	IIII	۴	48
۸	۱۰۸ـ۱۲۱	II	۲	50
		$N = ۵۰$	$\Sigma F_i = ۵۰$	

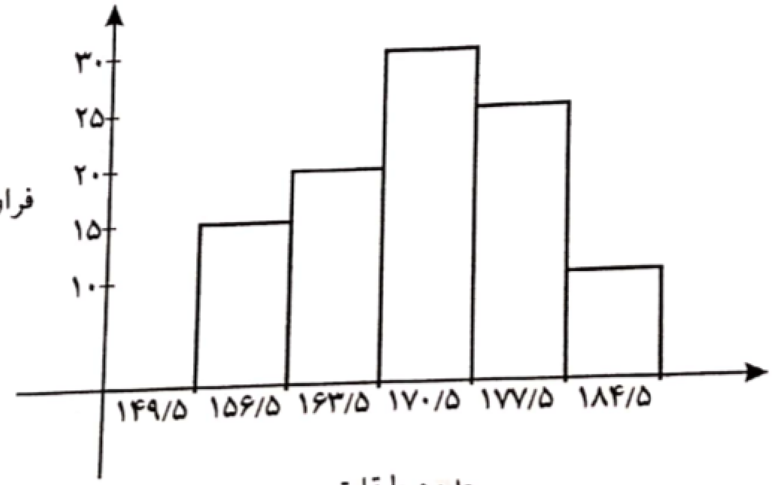
نمودارهای توزیع فراوانی

اطلاعات خلاصه شده در یک جدول توزیع فراوانی را می‌توان توسط نمودارهای هندسی با سرعت و سهولت بیشتری به بیننده انتقال داد. هر نمودار آماری باید دارای شماره، عنوان و در صورت لزوم زیر نویس و مأخذ باشد. مقیاس‌های اندازه‌گیری روی محورهای افقی و عمودی باید مشخص باشند.

«برای نمایش داده‌های پیوسته و طبقه بندی شده از دو نمودار زیر استفاده می‌کنیم.»

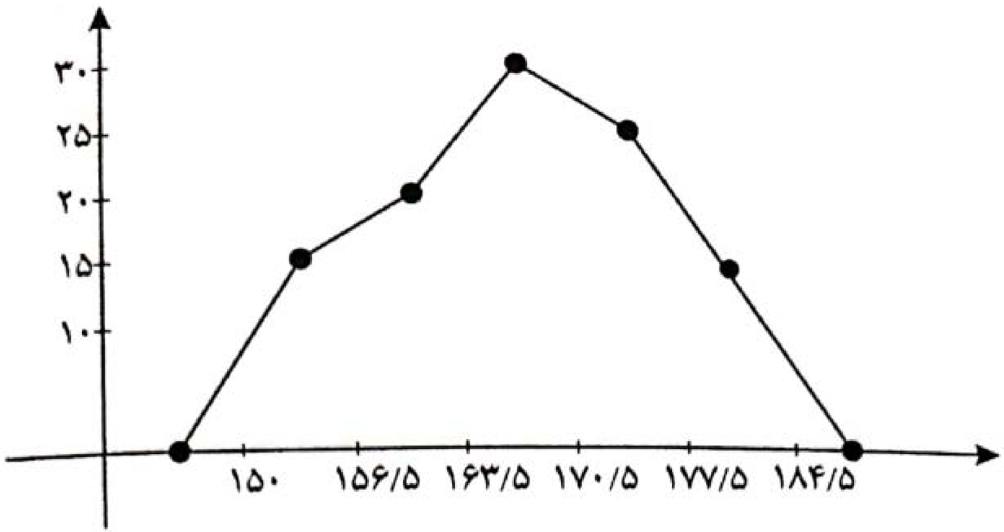
هیستوگرام: نموداری است مرکب از چند مستطیل که از روی جدول فراوانی داده‌های پیوسته ساخته می‌شود. تعداد مستطیل‌ها برابر است با تعداد طبقات قاعده. هر مستطیل روی محور xها جا دارد و طولش برابر است با طول واقعی طبقه. ارتفاع مستطیل برابر است با فراوانی نسبی یا مطلق طبقه مربوطه.

شکل زیر مربوط به نمودار هیستوگرام جدول صفحه قبل است و نشان می‌دهد که بیشترین تراکم در فاصله [۱۷۰/۵ , ۱۶۳/۵] است.



نمودار چند ضلعی فراوانی (پلی گن): نمودار دیگری که در مورد صفات پیوسته به کار می‌رود، نمودار چند ضلعی فراوانی است، در این نمودار دو نقطه دیگر با فراوانی صفر، یکی قبل از نماینده طبقه اول و دیگری بعد از طبقه آخر به همراه نقاط مشخص شده در نظر می‌گیریم. نقاطی را تعیین می‌کنیم که طول آن‌ها نماینده طبقه و عرض آن‌ها فراوانی مربوط به آن طبقه باشد، از اتصال متوالی این نقاط به یک دیگر به وسیله خطوط مستقیم، نموداری حاصل می‌شود که آن‌را نمودار چند ضلعی می‌نامند.

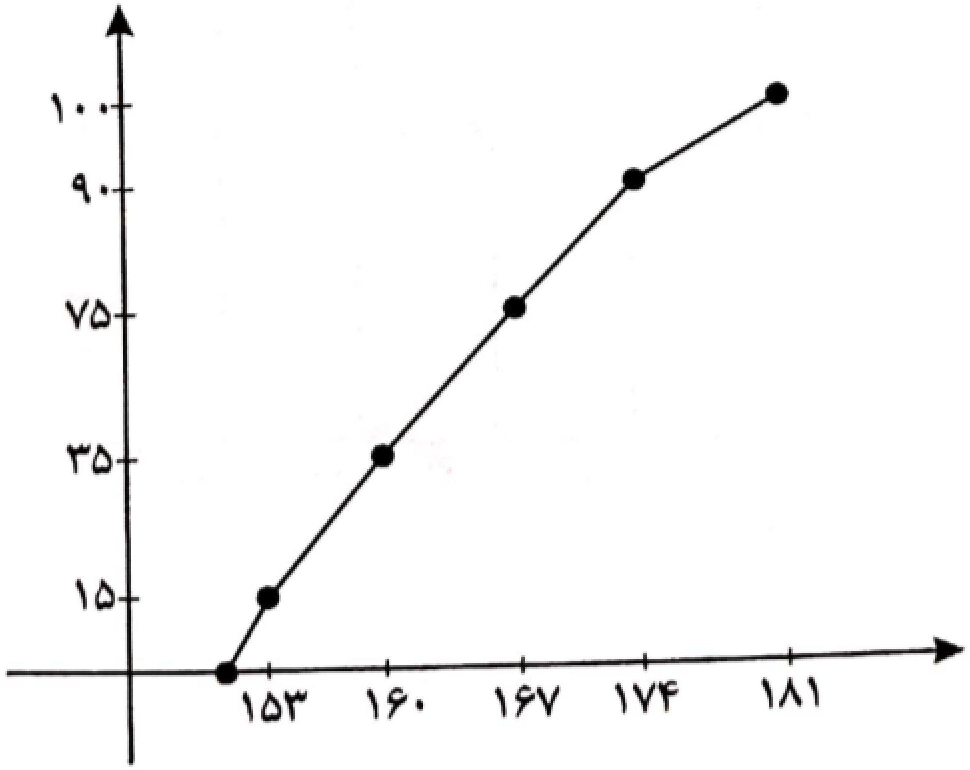
$$\text{نماینده هر طبقه} = x_i = \frac{۱}{۲}(\text{کران پایین طبقه} + \text{کران بالای طبقه})$$



نمودار چند ضلعی فراوانی مطلق اندازه قد ۱۰۰ جوان در جدول صفحه ۱۶

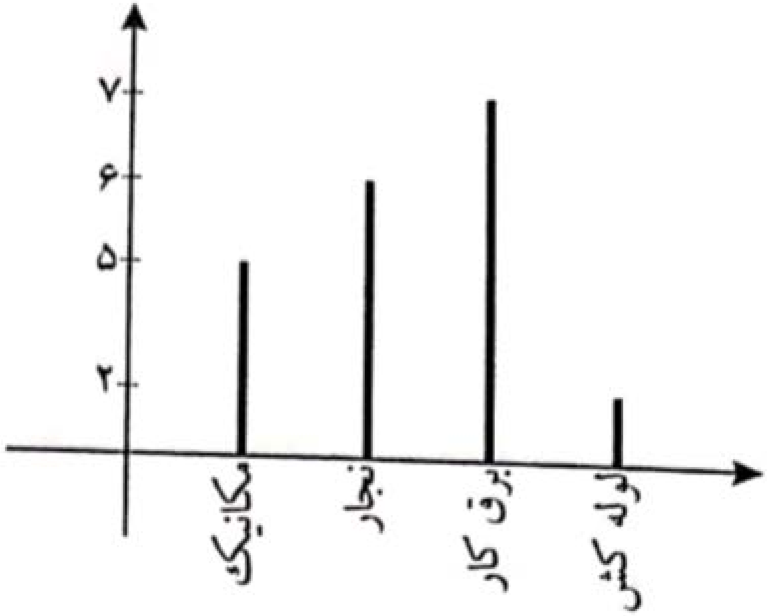
تبصره: معمولاً نمودار هیستوگرام فراوانی و چند ضلعی فراوانی را در یک دستگاه رسم می‌کنیم.

رسم نمودار فراوانی تجمعی: به منظور انتقال سریع‌تر مفهوم فراوانی تجمعی، می‌توان نموداری بـرای آن ترسیم کرد. برای رسم نمودار فراوانی تجمعی، روی محور افقی نماینده طبقات و روی محور عمـودی فراوانی‌های تجمعی را مشخص می‌کنیم. به این ترتیب نقاط متناظر به هـر طبقـه را مشـخص و سـپس بـا اتصال متوالی آن‌ها به یکدیگر نمودار فراوانی تجمعی حاصل می‌گردد.



نمودار فراوانی

نمودار میله‌ای: نمودارهای میله‌ای وقتی به کار گرفته می‌شوند که مسئله تحت بررسی دارای طبیعت کیفی باشد. در نمودار میله‌ای ارتفاع میله، فراوانی یا تعداد و محور x ها کیفیت و یا فعالیت را نشان می‌دهد. نمودار میله‌ای مربوط به جدول مثال ۱-۵ را می‌توان به صورت زیر رسم کرد.



نمودار دایره‌ای: در نمودار دایره‌ای، معمولاً دایره‌ای به چند قطاع تقسیم می‌گردد که اندازه هر قطاع متناسب با فراوانی نسبی هر یک از طبقات است و از رابطه زیر به دست می‌آید.

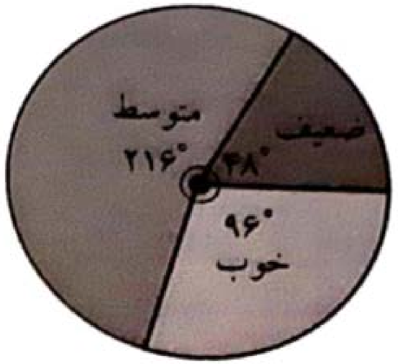
$$d_i = \frac{f_i}{n} \times ۳۶۰^\circ$$

که در آن d_i درجه، f_i فراوانی مربوطه به طبقه i ام و n فراوانی کل داده‌های آماری است.

۱-۸. مثال: جدول زیر، نظر ۳۰ دانشجو را که به صورت تصادفی انتخاب شده‌اند و در خوابگاه زندگی می‌کنند در خصوص کیفیت خوابگاه نشان می‌دهد.

کیفیت	تعداد فراوانی
ضعیف	۴
متوسط	۱۸
خوب	۸

اکنون با در اختیار داشتن اندازه کمان‌های هر قطاع، با توجه به جدول فوق، نمودار دایره‌ای را به صورت شکل زیر ترسیم می‌کنیم.



$$d_۱ = \frac{۴}{۳۰} \times ۳۶۰ = ۴۸^\circ$$

$$d_۲ = \frac{۱۸}{۳۰} \times ۳۶۰ = ۲۱۶^\circ$$

$$d_۳ = \frac{۸}{۳۰} \times ۳۶۰ = ۹۶^\circ$$

مثال ۱۰: جدول توزیع فراوانی مهارت ۲۲ کارگر کارخانه A، به صورت زیر داده شده، نمودارهای زیر را رسم کنید.

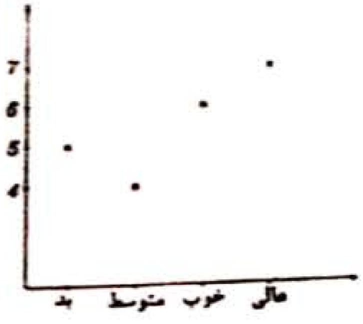
E_i	فراوانی
بد	۵
متوسط	۴
خوب	۶
عالی	۷
	۲۲

الف. نمودار سوزنی

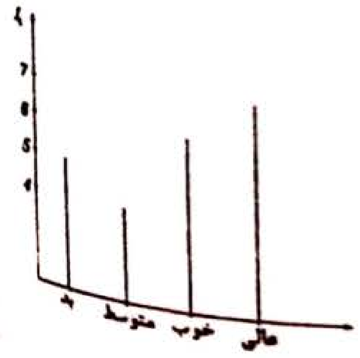
ب. نمودار میله‌ای

پ. نمودار دایره‌ای

حل:



الف



ب

۲۲   ۵

۳۶۰   x

$$x = \frac{۳۶۰}{۲۲} \times ۵ = ۸۱/۸$$

۲۲   ۴

۳۶۰   x

$$x = \frac{۳۶۰}{۲۲} \times ۴ = ۶۵/۵$$

۲۲   ۶

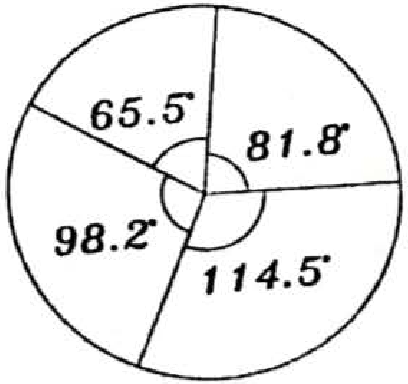
۳۶۰   x

$$x = \frac{۳۶۰}{۲۲} \times ۶ = ۹۸/۲$$

۲۲   ۷

۳۶۰   x

$$x = \frac{۳۶۰}{۲۲} \times ۷ = ۱۱۴/۵ \quad \square$$



تمرین

1– برای داده‌های آماری جدول زیر:

الف) جدول توزیع فراوانی را به نحوی تشکیل دهید که حد پایین طبقه اول ۱۳/۵ و طول طبقه معادل ۱ باشد.

ب) نمودار مستطیلی (هیستوگرام) مربوطه را رسم کنید.

ج) نمودار چند ضلعی فراوانی را رسم کنید

۱۶/۲	۱۷/۶	۱۸/۶	۱۶/۰	۱۷/۰	۱۴/۷
۱۶/۵	۱۴/۹	۱۵/۷	۱۸/۱	۱۷/۱	۱۷/۷
۱۴/۱	۱۵/۶	۱۷/۱	۱۷/۷	۱۸/۲	۱۶/۷
۱۶/۹	۱۶/۱	۱۶/۹	۱۷/۹	۱۸/۳	۱۷/۷
۱۶/۹	۱۶/۶	۱۷/۵	۱۶/۸	۱۶/۹	۱۷/۹

اگر بخواهیم دو جامعه آماری را با هم مقایسه کنیم، باید این مقایسه به صورت کمی انجام گیرد و در این خصوص از اعدادی استفاده می‌شود که آن‌ها را شاخص‌های عددی می‌گوییم و این اعداد خصوصیات کلی جامعه آماری را بیان می‌کنند و به دو دسته شاخص‌های مرکزی و شاخص‌های پراکندگی تقسیم می‌شوند که دسته اول تجمع داده و دسته دوم پراکندگی داده را نسبت به یک مبدأ دلخواه نشان می‌دهد. از شاخص‌های مهم مرکزی به بررسی مفاهیم میانگین، میانه و مد می‌پردازیم.

۱–۱۰. میانگین: میانگین خود بر چند نوع است، میانگین حسابی، میانگین حسابی موزون، میانگین هندسی، میانگین هارمونیک.

میانگین حسابی: هرگاه داده‌های حاصل از مشاهدات آماری را با هم جمع نموده و بر تعداد کل داده تقسیم کنیم، معدل یا میانگین حسابی داده‌ها به دست می‌آید و آن‌را با نماد $\overline{x}$ یا μ نشان می‌دهیم. بنابراین، اگر $x_1, x_2, \dots, x_n$ مشاهده آماری وجود داشته باشد، میانگین حسابی آن‌ها به صورت زیر است.

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

بنابراین $\sum_{i=1}^{N} x_i = N.\overline{x}$. در بعضی مواقع ممکن است μ را اندیس متغیری مانند μ_x بنویسیم. این به معنای میانگین مقادیر x است و μ_y نیز میانگین مقادیر y را نشان می‌دهد.

۱–۱۱. مثال: داده‌های زیر نمرات دروس یک دانش آموز را نشان می‌دهد. میانگین نمرات را محاسبه کنید.

۱۷, ۱۴, ۱۷, ۱۱, ۱۹, ۱۷, ۲۰

حل:

$$\overline{x} = \frac{17+14+17+11+19+17+20}{7} = \frac{115}{7} = 16/42$$

خواص میانگین حسابی:

۱) جمع جبری اختلاف مجموعه‌ای از اعداد از میانگینشان، برابر صفر است یعنی

$$\sum_{i=1}^{N}(x_i - \overline{x}) = (x_1 - \overline{x}) + \dots + (x_N - \overline{x})$$
$$= (x_1 + x_2 + \dots + x_N) - \underbrace{(\overline{x} + \dots + \overline{x})}_{N \text{ مرتبه}} = N\,\overline{x} - N\,\overline{x} = 0$$

۲) هرگاه عدد ثابتی مانند a به هر یک از اعداد داده‌های آماری اضافه شود، میانگین داده‌های جدید برابر میانگین داده‌های قبلی، به علاوه‌ی a خواهد بود. به عبارت دیگر، اگر $y_i = x_i + a$، آن‌گاه،

$$\sum_{i=1}^{N} y_i = \sum_{i=1}^{N} (x_i + a) = \sum_{i=1}^{N} x_i + Na$$

بنابراین،

$$\frac{\sum_{i=1}^{N} y_i}{N} = \frac{\sum_{i=1}^{N} x_i}{N} + \frac{Na}{N}$$

یا،

$$\mu_y = \mu_x + a$$

۳) هرگاه هر یک از داده‌های آماری در عدد ثابت b ضرب شود، میانگین اعداد حاصل، b برابر میانگین داده‌های قبلی خواهد بود.

به عبارت دیگر، اگر $y_i = bx_i$، آنگاه،

$$\sum_{i=1}^{N} y_i = \sum_{i=1}^{N} bx_i = b\sum_{i=1}^{N} x_i$$

بنابراین

$$\frac{\sum_{i=1}^{N} y_i}{N} = \frac{b\sum_{i=1}^{N} x_i}{N} \Rightarrow \bar{y} = b\bar{x} = b\mu_x$$

میانگین حسابی موزون (وزن‌دار): گاهی اوقات هر یک از مشاهدات x_i دارای وزنی در حدود w_i است. در این صورت، میانگین حسابی را میانگین حسابی موزون می‌نامیم و به صورت زیر محاسبه می‌کنیم.

x_i	x_1	x_2	$\cdots$	x_k	
w_i	w_1	w_2	$\cdots$	w_k	$\sum w_i = N$

$$\mu_w = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i} = \frac{w_1 x_1 + w_2 x_2 + \ldots + w_k x_k}{w_1 + w_2 + \ldots + w_k}$$

۱–۱۲. مثال: فرض کنید دانشجویی نمرات زیر را در دروس ترم قبل خویش گرفته باشد.

نام درس	تعداد واحد	نمره
آمار و کاربرد آن	۴	۱۹
جغرافیای اقتصادی	۲	۱۶
اقتصاد کلان ۱	۳	۱۸
اقتصاد خرد ۲	۴	۱۲
اقتصاد کشاورزی	۳	۱۴

در این مثال، تعداد واحدها، وزن‌های مربوطه را نشان می‌دهد. بنابراین، میانگین وزنی را به صورت زیر می‌توان محاسبه کرد.

$$\mu_w = \frac{(۴\times۱۹)+(۲\times۱۶)+(۳\times۱۸)\times(۴\times۱۲)+(۳\times۱۴)}{۱۶} = \frac{۲۵۲}{۱۶} = ۱۵/۷۵$$

محاسبه میانگین در داده‌های طبقه بندی شده: اگر داده‌ها از نوع کمی گسسته با تعداد زیاد و یا از نوع پیوسته باشد، در این صورت میانگین را از روی جدول فراوانی به صورت زیر محاسبه می‌کنیم.

ابتدا نماینده (نشان) دسته هر طبقه را مشخص می‌کنیم، سپس در فرمول میانگین به جای x_i نماینده طبقه iام و به جای f_i فراوانی مطلق iام را جای‌گذاری می‌کنیم و میانگین را به صورت زیر محاسبه می‌کنیم.

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$$

که در آن n تعداد کل داده‌ها است.

۱-۱۳. مثال: با توجه به مثال، میانگین قد ۱۰۰ جوان بیست ساله را با استفاده از جدول توزیع فراوانی مربوطه آن، محاسبه کنید.

طبقه	x_i	f_i	$f_i x_i$
۱۴۹/۵–۱۵۶/۵	۱۵۳	۱۵	۱۵×۱۵۳ = ۲۲۹۵
۱۵۶/۵–۱۶۳/۵	۱۶۰	۲	۲۰×۱۶۰ = ۳۲۰۰
۱۶۳/۵–۱۷۰/۵	۱۶۷	۳۰	۳۰×۱۶۷ = ۵۰۱۰
۱۷۰/۵–۱۷۷/۵	۱۷۴	۲۵	۲۵×۱۷۴ = ۴۳۵۰
۱۷۷/۵–۱۷۸/۵	۱۸۱	۱۰	۱۰×۱۸۱ = ۱۸۱۰
		۱۰۰	۱۶۶۶۵

$$\bar{x} = \frac{۱۶۶۶۵}{۱۰۰} = ۱۶۶/۶۵$$

میانگین هندسی: میانگین هندسی n داده مثبت $x_1, x_2, \ldots, x_n$ را با نماد G یا μ_G نشان داده و به صورت $G = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}$ محاسبه می‌کنیم.

اگر داده‌ها در جدول توزیع فراوانی، طبقه بندی شده و دارای فراوانی باشند از رابطه $G = \sqrt[n]{x_1^{f_1} \times \ldots \times x_k^{f_k}}$ که در آن x_i نماینده طبقه iام و f_i فراوانی طبقه iام و k تعداد طبقات است، استفاده می‌کنیم.

از این میانگین برای محاسبه متوسط آهنگ‌های تغییر (مثل جمعیت)، متوسط نسبت تغییرات (مثل افزایش قیمت‌ها) و درصدها استفاده می‌کنیم. به طور کلی، هرگاه میزان افزایش یا کاهش متناسب با موجودی اولیه باشد، مانند رشد جمعیت، و همچنین در محاسبه رشد متوسط نیز از این میانگین استفاده می‌کنیم، زیرا در این حالت، کمتر تحت تأثیر داده‌های خیلی بزرگ قرار می‌گیرد. هرگاه سال پایه ثابت نباشد و تولید هر سال نسبت به سال قبل در نظر گرفته شود، از میانگین هندسی استفاده می‌کنیم ولی اگر سال پایه ثابت باشد، از میانگین حسابی استفاده می‌کنیم.

۱–۱۴. مثال: میزان تولید یک کارخانه در چهار سال متوالی ۲، ۴، ۶ و ۲۷ برابر، نسبت به سال قبل می‌باشد. میزان افزایش متوسط تولید را به دست آورید.

حل: اگر از میانگین حسابی استفاده کنیم، آنگاه داریم:

$$\overline{x} = \frac{۲+۴+۶+۲۷}{۴} = ۹/۷۵$$

که میزان درستی نخواهد بود. زیرا در این صورت باید در سال چهارم $(۹/۷۵)^۴ = ۹۰۳۶/۸۷$ برابر سال اول تولید کرده باشد، در حالی که در سال چهارم $۲\times۴\times۶\times۲۷ = ۱۲۹۶$ برابر سال اول تولید کرده است. حال، چون در این مثال سال پایه ثابت نیست، لذا از میانگین هندسی استفاده می‌کنیم.

$$G = \sqrt[۴]{۲\times۴\times۶\times۲۷} = ۶$$

تذکر: چون محاسبه G از تعریف میانگین هندسی کمی دشوار می‌باشد لذا در محاسبه آن به صورت زیر عمل می‌کنیم.

$$G = \sqrt[n]{x_۱^{f_۱}\times\ldots\times x_n^{f_n}}$$

از طرفین معادله فوق، ln می‌گیریم.

$$\ln G = \frac{۱}{n}(f_۱ \ln x_۱ + \ldots + f_n \ln x_n) = S$$

حال تابع نمایی را برطرفین اثر می‌دهیم،

$$G = e^S$$

۱–۱۵. مثال: جدول زیر مصرف گاز مایع از سال ۱۳۶۳ تا ۱۳۶۷ را در کشور نشان می‌دهد. رشد متوسط مصرف را طی این دوره به دست آورید.

سال	x_i مصرف (هزار تن)	$\frac{x_{i+1}}{x_i}$
۱۳۶۳	۸۹۵	$\frac{۱۰۵۴}{۸۹۵} = ۱/۱۸$
۱۳۶۴	۱۰۵۴	$\frac{۱۰۳}{۱۰۵۴} = ۱/۰۳$
۱۳۶۵	۱۰۸۳	$\frac{۱۰۷۶}{۱۰۸۳} = ۰/۹۹$
۱۳۶۶	۱۰۷۶	$\frac{۱۱۹۷}{۱۰۷۶} = ۱/۱۱$
۱۳۶۷	۱۱۹۷	

حل: ابتدا نسبت هر سال به سال قبل (رشد سالانه) را به دست می‌آوریم. (به نمودار فوق توجه کنید.)

$$G = \sqrt[۴]{(۱/۱۸)(۱/۰۳)(۰/۹۹)(۱/۱۱)} \quad \text{(رشد متوسط پنج سال)}$$

$$\Rightarrow \ln G = \frac{۱}{۴}\left(\ln(۱/۱۸) + \ln(۱/۰۳) + \ln(۰/۹۹) + \ln(۱/۱۱)\right)$$

$$\Rightarrow \ln G = ۰/۰۷ \quad \Rightarrow \quad G = e^{۰/۰۷} = ۱/۰۸$$

میانگین هارمونیک(همساز): از این میانگین معمولاً برای محاسبه حد متوسط سرعت‌ها، مطالعه در شبکه‌های برق و عینک شناسی استفاده می‌شود و داده‌ها به صورت نسبت‌هایی هستند که صورت و مخرج آن‌ها دارای مقیاس یکسان نیستند و فرمول آن به صورت زیر می‌باشد.

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

اگر داده‌ها در جدول توزیع فراوانی طبقه‌بندی شده و دارای فراوانی باشند از رابطه زیر میانگین هارمونیک را محاسبه می‌کنیم.

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

که در آن x_i نشان‌دسته (متوسط دسته) طبقه i ام و f_i فراوانی طبقه i ام است.

۱–۱۸. مثال: جدول توزیع فراوانی مصرف برق ۵۰ خانوار به صورت زیر داده شده است، میانگین هارمونیک را حساب کنید.

حدود طبقات	f_i	x_i
۵۰–۵۲	۴	۵۱
۵۳–۵۵	۹	۵۴
۵۶–۵۸	۲۱	۵۷
۵۹–۶۱	۱۴	۶۰
۶۲–۶۴	۲	۶۳

$$H = \frac{n}{\sum_{i=۱}^{n} \frac{f_i}{x_i}} = \frac{۵۰}{(\frac{۴}{۵۱} + \frac{۹}{۵۴} + \frac{۲۱}{۵۷} + \frac{۱۴}{۶۰} + \frac{۲}{۶۳})} \cong ۵۶/۹۵$$

مثال ۱۰_۲ سرعت ۳۰ ماشین برحسب دور در ثانیه به این شرح است:

x_i (سرعت)	۵	۶	۱۰	۱۵	
W_i (تعداد)	۶	۱۰	۸	۶	$\Sigma W_i = ۳۰$

میانگین سرعت ماشنها با استفاده از رابطه ۷_۲ عبارت است از:

$$H = \frac{۶+۱۰+۸+۶}{\frac{۶}{۵}+\frac{۱۰}{۶}+\frac{۸}{۱۰}+\frac{۶}{۱۵}} = \frac{۳۰}{۴/۰۶۷} = ۷/۳۷۶$$

۱–۱۹. میانه: عدد m را میانه می‌نامند، هرگاه تقریباً تعداد نصف داده‌ها از m کوچکتر باشند.

محاسبه میانه برای داده‌های گسسته: فرض کنیم n داده گسسته داشته باشیم که به طور غیر نزولی، یعنی به صورت $x_۱ \le x_۲ \le \dots \le x_n$ تنظیم شده باشند. اگر n فرد باشد، داده‌ای را که در وسط قرار دارد، میانه می‌گویند. اگر n زوج باشد، نصف مجموع دو داده‌ای را که در وسط قرار دارند، میانه می‌گویند. میانه را با m یا M_d نشان می‌دهیم. بنابراین داریم:

برای n فرد، $m = x_{\frac{n+۱}{۲}}$

برای n زوج، $m = \frac{(x_{\frac{n}{۲}} + x_{\frac{n+۲}{۲}})}{۲}$

۱–۲۰. مثال: برای پیدا کردن میانه داده‌های ۱۲، ۸، ۱۷، ۳۰، ۲۰، ۲۱، ۴ و ۳۱، ابتدا آن‌ها را به طور غیر نزولی یعنی به صورت زیر می‌نویسیم.

۴ ,۵ , ۸ , ۱۲ , ۱۷ , ۲۰ , ۲۱ , ۳۰ , ۳۱

چون تعداد داده‌ها، عدد فرد ۹ = n است، میانه برابر است با $m = ۱۷$.

۱–۲۱. مثال: برای پیدا کردن میانه داده‌های ۲، ۳، ۵، ۴، ۱، ۲، ۳ و ۲ آن‌ها را به طور غیر نزولی، یعنی به صورت

۵ و ۴ و ۳ و ۳ و ۲ و ۲ و ۲ و ۱

می‌نویسیم. چون تعداد داده‌ها، عدد زوج ۸ = n است، میانه برابر است با:

$$m = \frac{۲+۳}{۲} = ۲/۵$$

محاسبه میانه از روی جدول توزیع فراوانی: ابتدا طبقه میانه دار را مشخص می‌کنیم، طبقه میانه دار طبقه‌ای است که فراوانی تجمعی آن برای اولین بار بزرگتر یا مساوی $\frac{n}{۲}$ باشد. (n فراوانی کل است) سپس با استفاده از فرمول زیر مقدار تقریبی میانه را محاسبه می‌کنیم.

$$m = l_i + \frac{(\frac{n}{۲} - F_{i-۱})}{f_i} \times c_i$$

که در آن l_i حد پایین طبقه میانه دار، c_i طول طبقه و f_i فراوانی مطلق طبقه میانه دار و $F_{i-۱}$ فراوانی تجمعی طبقه قبل از طبقه میانه‌دار است.

۱–۲۲. مثال: جدول توزیع فراوانی زیر وزن ۱۰۰ دانشجو را نشان می‌دهد. میانه وزن را محاسبه کنید.

حدود طبقات	f_i	F_i
۵۸–۶۲	۵	۵
۶۲–۶۶	۲۶	۳۱
۶۶–۷۰	۴۰	۷۱
۷۰–۷۴	۲۰	۹۱
۷۴–۷۸	۹	۱۰۰

حل: $۵۰ = \frac{۱۰۰}{۲} = \frac{n}{۲}$. اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی ۵۰ است، طبقه سوم می‌باشد که به عنوان طبقه میانه‌دار انتخاب می‌کنیم. حال با استفاده از فرمول فوق داریم:

$l_۳ = ۶۶ \;,\; f_۳ = ۴۰ \;,\; F_۲ = ۳۱$

$c_۳ = ۷۰ - ۶۶ = ۴$

$$m = ۶۶ + (\frac{۵۰ - ۳۱}{۴۰}) \times ۴ = ۶۷/۹$$

مد (نما): در یک مجموعه از داده‌های آماری، داده‌ای که بیشترین تعداد تکرار یا فراوانی را داشته باشد، مد نامیده می‌شود و آن را با Mod نمایش می‌دهیم. برای مثال در داده‌های زیر مد برابر ۲ است، چون ۲ بیشترین تکرار را دارد.

۵ ۳ ۵ ۲ ۴ ۲ ۳ ۲

بعضی اوقات مجموعه آماری به صورت دو مده ظاهر می‌شود. این اغلب به این علت است که دو توزیع مختلف با هم ادغام شده‌اند. در مثال فوق ممکن است داده ۵ نیز سه بار ظاهر شود که در این حالت ۲ و ۵ مد خواهند بود.

محاسبه مد در جدول توزیع فراوانی: در این حالت به طبقه‌ای که بیشترین تعداد فراوانی را دارد، طبقه مد دار می‌گوییم، و با استفاده از رابطه زیر مقدار تقریبی مد را محاسبه می‌کنیم.

$$M_\circ = Mod \simeq l_i + (\frac{d_1}{d_1 + d_2}) \times c_i$$

که در آن l_i حد پایین طبقه مد دار و d_1 تفاضل فراوانی طبقه ما قبل از فراوانی طبقه مددار، $(d_1 = f_i - f_{i-1})$ و d_2 تفاضل فراوانی طبقه ما بعد از فراوانی طبقه مددار، $(d_2 = f_i - f_{i+1})$ و c_i طول طبقه است.

۱–۲۳. مثال: در جدول فراوانی زیر، مد را محاسبه می‌کنیم.

حدود طبقات	f_i
۵۸–۶۲	۵
۶۲–۶۶	۲۶
۶۶–۷۰	۴۰
۷۰–۷۴	۲۰
۷۴–۷۸	۹

حل: چون طبقه سوم بیشترین فراوانی را دارد، لذا طبقه مددار است و بنابراین داریم:

$$l_3 = ۶۶ \quad , d_1 = ۴۰ - ۲۶ = ۱۴ \ , \ d_2 = ۴۰ - ۲۰ = ۲۰ \ , \ c_3 = ۷۰ - ۶۶ = ۴$$

بنابراین مد به صورت زیر محاسبه می‌شود.

$$Mod = ۶۶ + (\frac{۱۴}{۱۴ + ۲۰}) \times ۴ \cong ۶۷/۶۴$$

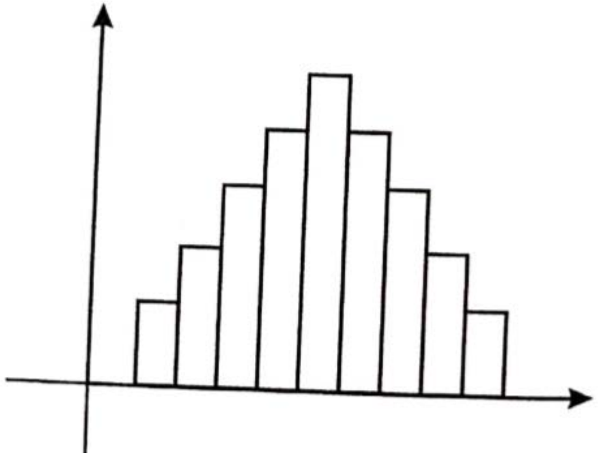
مقایسه میانگین، میانه، مد: از مزیت‌های میانگین آن است که بر کلیه داده‌ها متکی می‌باشد و با تغییر یک داده، مقدار آن عوض می‌شود، مثلاً برای داده‌های

۲ ۳ ۴ ۵ ۵ ۶ ۷

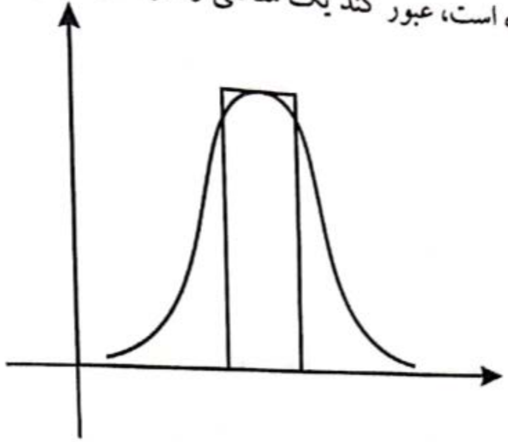
اگر میانگین، میانه و مد را حساب کنیم، داریم: $Mod = ۵$ و $M = ۵$ و $\overline{x} = ۴/۵$. حال اگر عدد ۷ را با ۷۰ عوض کنیم، میانه و مد تغییر نمی‌یابد ولی میانگین عوض می‌شود.

از دیگر محاسن میانگین آن است که نسبت به میانه و مد پایدارتر است. به عبارت دیگر، در نمونه‌گیری از یک جامعه با تغییر حجم نمونه، نمونه‌های مختلفی با میانگین و میانه و مد متفاوت به دست می‌آید که اگر پارامترها را با هم مقایسه کنیم، تغییرات میانگین کمتر از دیگر پارامترها می‌باشد.

یک توزیع را متقارن می‌نامیم، هرگاه هیستوگرام فراوانی به صورت زیر باشد. (یعنی نسبت به مستطیل وسطی متقارن باشد.)

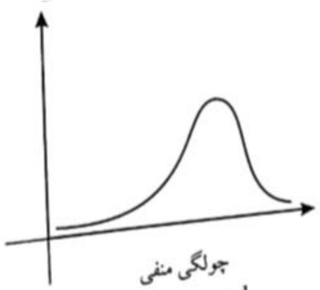


اگر برای هیستوگرام فوق یک منحنی رسم کنیم که از نقاط میانی بالای مستطیل‌ها به گونه‌ای که در شکل زیر نشان داده‌شده است، عبور کند یک منحنی زنگوله‌ای شکل کاملاً قرینه به دست خواهد آمد.



در توزیع متقارن، میانگین، میانه و مد بر هم منطبق هستند. توزیع‌هایی که متقارن نباشند را توزیع‌های چوله می‌نامیم که شاخص چولگی در مباحث بعدی بحث خواهد شد.

در توزیع‌های چوله شاخص‌های مرکزی بر هم منطبق نیستند ولی همیشه میانه بین مد و میانگین است. اگر میانه از مد بزرگتر باشد، توزیع چوله به راست است. (چولگی مثبت) و اگر میانه از مد کوچکتر باشد توزیع چوله به چپ است. (چولگی منفی)



چولگی منفی

$\mu < m < \text{mod}$

چولگی مثبت

$\text{mod} < m < \mu$

تمرین: در جدول توزیع زیر با محاسبه میانگین، میانه و مد، نشان دهید که توزیع متقارن نمی‌باشد.

حدود طبقات	f_i	F_i	x_i
۰-۴	۱۰	۱۰	۲
۴-۸	۱۵	۲۵	۶
۸-۱۲	۲۵	۵۰	۱۰
۱۲-۱۶	۳۰	۸۰	۱۴
۱۶-۲۰	۱۵	۹۵	۱۸
۲۰-۲۴	۵	۱۰۰	۲۲