

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

مفاهیم و روشهای آماری (۱)

رشته حسابداری بازرگانی

گروه تحصیلی اداری مالی

زمینه خدمات

شاخه آموزش فنی و حرفه‌ای

شماره درس ۳۹۵۳

مرآت، ابوالقاسم	۰۰۱
مفاهیم و روشهای آماری (۱) / مؤلف: ابوالقاسم مرآت. - تهران: شرکت چاپ و نشر کتابهای	/ ۴۲۲
درسی ایران، ۱۳۹۴.	م ۴۳۲ م
۱۱۳ ص. - (آموزش فنی و حرفه‌ای؛ شماره درس ۳۹۵۳)	۱۳۹۴
متون درسی رشته حسابداری بازرگانی گروه تحصیلی اداری مالی، زمینه خدمات.	
برنامه‌ریزی و نظارت، بررسی و تصویب محتوا: کمیسیون برنامه‌ریزی و تألیف کتابهای درسی	
رشته حسابداری بازرگانی دفتر تألیف کتابهای درسی فنی و حرفه‌ای و کاردانش وزارت آموزش و پرورش.	
۱. آمار. الف. ایران. وزارت آموزش و پرورش. دفتر تألیف کتابهای درسی فنی و حرفه‌ای	
و کاردانش. ب. عنوان. ج. فروست.	

همکاران محترم و دانش آموزان عزیز :

پیشنهادات و نظرات خود را درباره محتوای این کتاب به نشانی تهران-
صندوق پستی شماره ۴۸۷۴/۱۵ دفتر تألیف کتابهای درسی فنی و حرفه ای و
کار دانش، ارسال فرمایند.

info@tvoccd.sch.ir

پیام نگار (ایمیل)

www.tvoccd.sch.ir

وبگاه (وب سایت)

محتوای این کتاب براساس نظرهای هنرآموزان توسط مؤلف زیر نظر کمیسیون تخصصی
برنامه ریزی و تألیف کتابهای درسی رشته حسابداری بازرگانی در سال ۸۷-۱۳۸۶ مورد بازبینی
قرار گرفت و فصل ششم اضافه گردید.

وزارت آموزش و پرورش سازمان پژوهش و برنامه ریزی آموزشی

برنامه ریزی محتوا و نظارت بر تألیف : دفتر تألیف کتابهای درسی فنی و حرفه ای و کار دانش
نام کتاب : مفاهیم و روشهای آماری (۱) - ۳۵۹/۴
مؤلف : ابوالقاسم مرآت

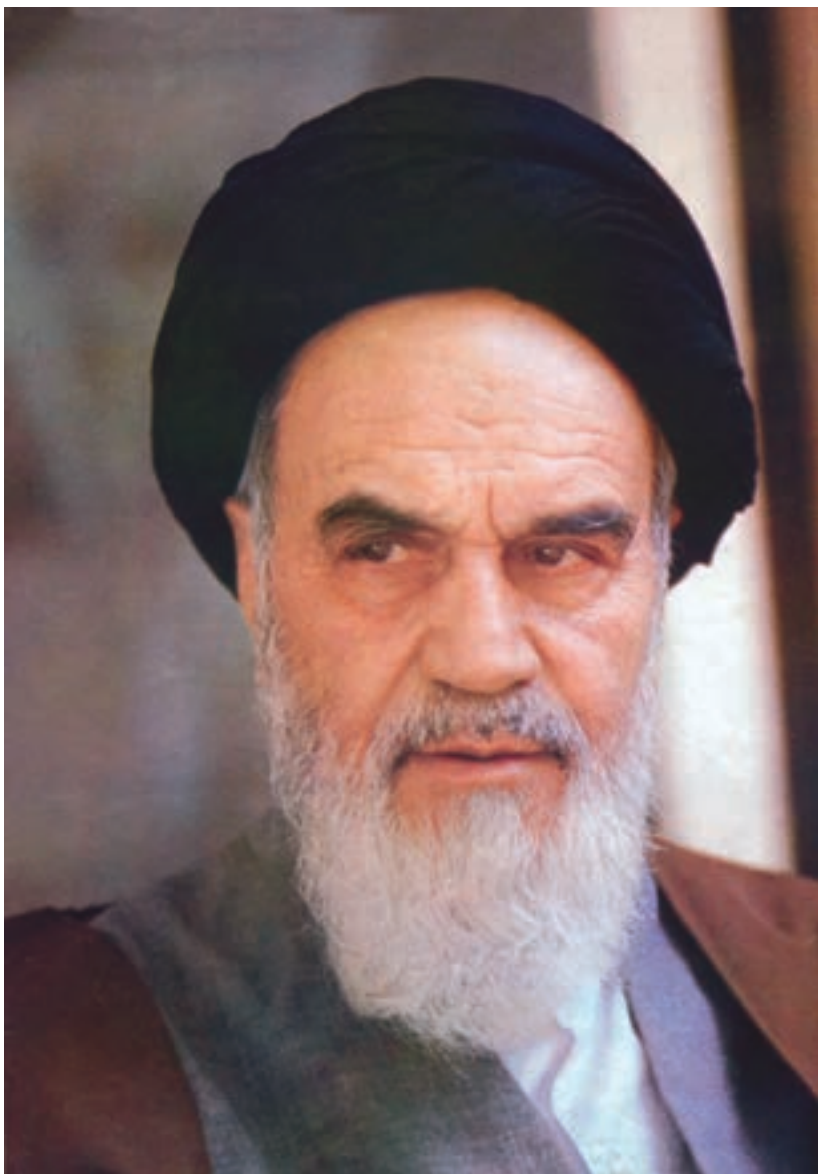
آماده سازی و نظارت بر چاپ و توزیع : اداره کل نظارت بر نشر و توزیع مواد آموزشی
تهران : خیابان ایرانشهر شمالی - ساختمان شماره ۴ آموزش و پرورش (شهید موسوی)
تلفن : ۹-۸۸۸۳۱۱۶۱، دورنگار : ۰۹۲۶۶۰۸۸۳، کدپستی : ۱۵۸۴۷۴۷۳۵۹،
وبسایت : www.chap.sch.ir

رسم : محمد پریسای

صفحه آرا : صغری عابدی

طراح جلد : مریم کیوان

ناشر : شرکت چاپ و نشر کتابهای درسی ایران - تهران - کیلومتر ۱۷ جاده مخصوص کرج - خیابان ۶۱ (داروبخش)
تلفن : ۵-۴۴۹۸۵۱۶۱، دورنگار : ۰۴۴۹۸۵۱۶۰، صندوق پستی : ۱۳۹-۳۷۵۱۵
چاپخانه : شرکت چاپ و نشر کتابهای درسی ایران «سهامی خاص»
سال انتشار و نوبت چاپ : چاپ شانزدهم ۱۳۹۴
حق چاپ محفوظ است.



اول باید اخلاصتان را قوی بکنید، ایمانتان را قوی بکنید،... و این
اخلاص و ایمان، شما را تقویت می کند و روحیه شما را بالا می برد و نیروی
شما جوری می شود که هیچ قدرتی نمی تواند (با شما) مقابله کند.
امام خمینی (ره)

فهرست

مقدمه

۱- مختصری دربارهٔ پیدایش، رشد و توسعهٔ آمار

۲- تعریف آمار به عنوان یک علم

۱	فصل اول - مفاهیم اساسی
۱	مفاهیم اولیه
۱	جامعهٔ آماری
۳	صفت متغیّر
۴	مراحل مختلف انجام یک تحقیق آماری
۴	تعیین هدف تحقیق
۴	مشاهدهٔ آماری
۵	پردازش داده‌های آماری
۵	مرحلهٔ تجزیه و تحلیل نتایج مشاهدات (داده‌های آماری)
۵	اندازه‌گیری و انواع اندازه‌گیری (مقیاسهای اندازه‌گیری)
۶	مقیاس اسمی
۶	مقیاس ترتیبی یا مقیاس رتبه‌ای
۷	مقیاس فاصله‌ای
۸	مقیاس نسبی
۹	علامت زیگما () و کاربرد آن در محاسبات
۱۰	سؤالات و تمرینها
۱۲	خودآزمونهای چهارگزینه‌ای
۱۴	فصل دوم - گروه‌بندی و پردازش داده‌های آماری
۱۴	پردازش داده‌های آماری - توزیع صفت متغیّر

- ۱۴ پردازش داده‌های آماری
- ۱۵ توزیع صفت متغیر
- ۱۸ بیان هندسی توزیع صفت متغیر (نمودار میله‌ای - نمودار چندگوش)
- ۱۹ فراوانیهای تجمعی (انباشته)
- ۲۳ طریقه عملی گروه‌بندی نتایج مشاهدات (داده‌های آماری)
- ۲۵ مراحل تشکیل جدول توزیع فراوانیهای فاصله‌ای
- ۳۰ بیان هندسی توزیع فراوانیهای فاصله‌ای (صفت متغیر پیوسته)
- ۳۴ سؤالها و تمرینها
- ۳۷ خودآزمونیهای چهارگزینه‌ای
- ۳۸ فصل سوم - مشخص‌کننده‌های مرکزی (تمایل به مرکز)
- ۳۸ مشخص‌کننده‌های آماری
- ۳۹ میانگین حسابی
- ۴۵ بعضی خواص ریاضی میانگین حسابی
- مقایسه توزیعهای فراوانی به وسیله میانگین و نتیجه‌گیری
- ۴۸ نسبت به یکسان بودن آنها
- ۵۰ میانه
- ۵۴ خواص میانه
- ۵۶ نما (مُد)
- ۶۰ سؤالها و تمرینها
- ۶۴ خودآزمونیهای چهارگزینه‌ای
- ۶۶ فصل چهارم - مشخصه‌های پراکندگی
- ۶۶ مشخصه‌های پراکندگی
- ۶۶ طول دامنه تغییرات
- ۶۷ متوسط قدر مطلق انحرافات (انحراف متوسط)
- ۶۹ متوسط مجذور انحرافات یا واریانس (پراش)

۷۲	انحراف معیار
۷۳	انحراف چارکی
۷۷	مشخصه‌های نسبی پراکندگی
۷۷	ضریب تغییرات
۷۸	انحراف خطی نسبی
۷۸	انحراف چارکی نسبی
۷۹	مشخصه چولگی (انحراف از قرینگی)
۷۹	ضریب چولگی پیرسن
۸۲	ضریب چولگی بر پایه میانگین توان سوم انحرافات
۸۴	سؤاها و تمرینها
۸۶	خودآزمونهای چهارگزینه‌ای
۸۹	فصل پنجم — توزیع صفت متغیر کیفی
۸۹	توزیع صفت کیفی
۹۲	محاسبه مشخصه‌های عددی برای صفات کیفی
۹۵	مطالعه جامعه از نظر دو صفت کیفی A و B
۹۹	بیان هندسی صفات کیفی
۱۰۱	سؤاها و تمرینها
۱۰۳	فصل ششم — توزیع مشترک دو صفت متغیر
۱۰۳	توزیع مشترک دو صفت متغیر X و Y
۱۰۵	توزیعهای حاشیه‌ای
۱۰۷	توزیعهای شرطی
۱۰۸	مشخصه‌های عددی برای توزیع مشترک دو صفت X و Y
۱۱۱	سؤاها و تمرینها
۱۱۳	منابع و مأخذ

مقدمه

واژه آمار معادل لاتین Statistics است که از کلمه Status به معنای حالت، وضع یا موقعیت گرفته شده. کلمات Stato (دولت)، Statista (دولت‌شناس) نیز از همین خانواده هستند.

۱- مختصری دربارهٔ پیدایش، رشد و توسعهٔ آمار: علم آمار مانند علوم دیگر در نتیجهٔ نیازهای بشر به وجود آمده است و تاریخی طولانی دارد و از دورانهای گذشته تاکنون رشد و توسعه آن ادامه یافته است.

با پیدایش نخستین دولتها در تاریخ، آمار، هستی یافته است. در گذشته‌های دور هر دولت شکل گرفته، دارای دستگاههای اداره‌کننده و هم‌چنین نیروی ارتش بوده است. به همین جهت به آگاهی از تعداد جمعیت، میزان دارایی و ثروت قلمرو حکومت خود نیاز داشته است تا امکانات و نیازمندیهای ضروری خود را فراهم سازد. در این باره از زمانهای قدیم، سرشماریهای مختلف از جمعیت، تعداد دامها، محصولات کشاورزی و ... انجام می‌گرفته است که می‌توان از کشورهای مصر، چین، هند و بابل نام برد. در آغاز، عملیات آماری بسیار ساده بود و فقط بعضی از نمودهای اجتماعی را دربر می‌گرفت، رفته‌رفته با رشد تولید، ظهور روابط کالا و پول و در نتیجه با پیچیده‌تر شدن زندگی اقتصادی و امور دولتی، ثبت و نگهداری حساب نفوس، ثبت مشاغل، محدودهٔ اراضی، دامها و نظایر آن منظم‌تر و کاملتر شد، در نتیجه کارهای آماری نیز توسعه یافت. ظهور و توسعهٔ تجارت بین‌المللی باعث شد که نیاز به کسب اطلاعات دربارهٔ حکومتهای خارجی نیز بیشتر گردد، به طوری که سازمانهای دولتی و مؤسسات سرمایه‌داری مجبور به جمع‌آوری اطلاعات وسیع و مختلف دربارهٔ بازار کار و فروش کالا، منابع مواد خام و نظایر آنها شدند. این اطلاعات که بر پایهٔ نیازهای یادشدهٔ آن زمان به عنوان آمار رشد یافت، تنها یکی از پایه‌های آمار امروزی را تشکیل می‌داد ولی در طول زمان نیازها فزونی می‌گرفت و

آمار همچنان تکامل می‌یافت زیرا هنوز دانستنیها کافی نبودند. بدین جهت از قرن هفدهم به بعد گروهی از دانشمندان مکتبی به نام «حسابدانه‌های سیاسی» تشکیل دادند که هدف آن بیشتر کشف نظم روابط اجتماعی و مسابلی در ارتباط با رشد و توسعه سرمایه‌داری بود. هم‌زمان با این گروه، دانشمندان دیگری در آلمان مکتب آمار توصیفی یا دولت‌شناسی را تشکیل دادند، اینان سعی کردند که به‌طور همه‌جانبه با استفاده از اعداد، دولتها و کشورها را تفسیر نمایند. در این سده‌ها رشد و توسعه آمار، مراحل مختلفی را گذرانده، به یکی از تواناترین ابزارهای شناسایی روابط اجتماعی تبدیل شده بود. اواخر قرن نوزدهم و اوایل قرن بیستم دوره رشد و تکامل سریع آمار است. در این زمان، روشهای آماری و استفاده از علوم ریاضی به‌طور وسیع توسعه یافته بود و در کنار آن به موازات آمار جمعیت، رشته‌های دیگر آمار اجتماعی و اقتصادی نیز رشد و توسعه می‌یافت.

۲- تعریف آمار به عنوان یک علم: تعریف موضوع آمار، مانند هر علم دیگر اهمیت بسیار دارد. بحث درباره موضوع و روش آمار به‌طور متفاوت در کشورهای مختلف بیش از ۱۰۰ سال است که انجام می‌گیرد.

تعیین و تعریف موضوع یک علم، از تعیین کردن محتوای آن، جایگاه آن در بین علوم دیگر و چگونگی ارتباط آن با سایر علوم شکل می‌گیرد.

با توجه به تاریخ توسعه و تکامل آمار، اکثر دانشمندان، آمار را به عنوان یک علم به‌صورت زیر تعریف نموده‌اند:

«آمار یک علم مستقل است که جنبه‌های کمی نمودهای اجتماعی را در ارتباط با کیفیت آنها، با هم مطالعه می‌کند.»

این تعریف، آمار را به عنوان یک علم مشخص کرده، اختلاف آن را از موضوعهای علوم دیگر بیان می‌کند. مطالعه جنبه‌های کمی نمودها در ارتباط با جنبه‌های کیفی آنها، نقشی اساسی در فرآیند تحلیل داده‌های آماری پیدا می‌کند. باید گفت که در روشهای مطالعه نمودهای اجتماعی، تنها از مفاهیم و روشهای آماری استفاده نمی‌شود بلکه روشها و طریقه‌های دیگری نیز وجود دارد. از این رو، آمار کاربردی را به عنوان یک علم مستقل به‌صورت زیر تعریف می‌کنند:

«آمار کاربردی علم مستقلی است که موضوع آن را تهیه و تنظیم مفاهیم، روشها و الگوهای ریاضی موردنیاز تشکیل می‌دهد. همچنین طرح مسئله، جمع‌آوری داده‌های آماری، پردازش و استخراج آنها، به‌منظور تعبیر و استنتاج علمی و عملی نیز موضوع آمار کاربردی می‌باشد.»

روشهای آماری را می‌توان به دو بخش آمار توصیفی و آمار استنباطی تقسیم نمود:

آمار توصیفی، مجموعه‌ای از روشهاست که برای گردآوری، توصیف و بیان مقداری اطلاعات

مربوط به جامعه‌های مورد مطالعه، به کار می‌رود.

آمار استنباطی برای استنباط خصوصیات یک جامعه آماری از روی نتایج یک نمونه که از آن جامعه انتخاب می‌شود، به کار می‌رود. آمار استنباطی امکان می‌دهد تا جنبه‌های مورد نظر جامعه‌های بزرگ آماری، از روی نمونه‌های نسبتاً کوچکی که نماینده آن جامعه‌ها هستند، توصیف شوند.

هدف کلی

آشنا کردن دانش‌آموزان با چگونگی توصیف و تعبیر مجموعه داده‌های آماری
(نتایج مشاهدات نسبت به یک خاصیت مورد مطالعه در یک جامعه از اشیاء یا پدیده‌ها)
با استفاده از مفاهیم و روشهای آماری.

فصل اول

مفاهیم اساسی

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند:

- ۱- مفاهیم اساسی آمار (جامعه آماری، صفت مشخصه، صفت متغیر و انواع آن) را شرح دهید.
- ۲- مفاهیم اساسی آمار (جامعه آماری، صفت مشخصه، صفت متغیر و انواع آن) را با رشته فعالیت خود تطبیق دهد.
- ۳- مشاهده آماری را به عنوان یکی از روشهای تحقیق آماری توضیح دهد.
- ۴- یک صفت متغیر را در یک جامعه مشاهده کند و نتایج آن را ارائه دهد.
- ۵- مراحل مطالعه آماری را بیان کند.
- ۶- مفهوم اندازه گیری و انواع آن را شرح دهد و در مشاهده صفتهای متغیر به کار برد.

مفاهیم اولیه

آمار نیز مانند علوم دیگر دارای مفاهیم اساسی است. این گونه مفاهیم را «مفاهیم اولیه» نیز می نامند که بدون تعریف اند و فقط توصیف می شوند.

جامعه آماری

یکی از مفاهیم اولیه در آمار، مفهوم «جامعه آماری» است که به صورت زیر بیان می شود:

«به همه افراد یا اشیايي که حداقل در یک خاصیت باهم مشترک باشند «جامعه آماری» گویند.»

به این ترتیب جامعه آماری تنها به انسانها گفته نمی‌شود بلکه، جامعه اشیا را نیز جامعه آماری می‌نامند.

«خاصیت یا خواصی که بین اعضای جامعه مشترک هستند، «صفت مشخصه» نامیده می‌شوند.»

مثلاً:

— جامعه ایرانیان که صفت مشخصه آن «ایرانی بودن» است.
— جامعه دانش‌آموزان ایرانی که صفات مشخصه آن، «دانش آموز» و «ایرانی بودن» است.
— جامعه روستاهای ایران که صفات مشخصه آن «روستا بودن» و «متعلق به ایران بودن» می‌باشد.

— جامعه حسابداران که صفت مشخصه آن «حسابدار بودن» است.
— جامعه مهره‌های ساخته شده در یک دستگاه ماشینی که صفات مشخصه آن یکی «مهره بودن» و دیگری «ساخته شدن به وسیله یک دستگاه ماشینی» است.

هریک از اعضای جامعه را «واحد جامعه» یا «فرد جامعه» می‌نامند.

مثلاً در جامعه ایرانیان، یک فرد ایرانی، عضو یا واحد یا فرد آن جامعه می‌باشد یا در جامعه دانش‌آموزان ایران، هر دانش‌آموز ایرانی یک واحد یا فرد آن جامعه به حساب می‌آید.

تعداد واحدهای تشکیل دهنده جامعه را، «حجم جامعه» گویند و آن را با حرف N نشان می‌دهند.

مثلاً اگر جمعیت ایران (در سال ۱۳۷۵) ۶۰ میلیون نفر باشد: $N = 60,000,000$.
اگر تعداد مهره‌های ساخته شده در یک دستگاه ماشینی ۱۰۰,۰۰۰ عدد باشد: $N = 100,000$.
اعضای جامعه آماری، می‌توانند محدود یا نامحدود باشند.

— **جامعه محدود (متناهی):** دارای حجم معین می‌باشد، به عبارت دیگر به تمامی اعضای آن می‌توان دسترسی داشت مانند جامعه ایرانیان که در زمان معین (مثلاً سال ۱۳۷۵) دارای حجم $N = 60,000,000$ است.

— **جامعه نامحدود (نامتناهی):** تعداد اعضای آن نامحدود است، به عبارت دیگر به تمام

اعضای آن نمی‌توان دسترسی داشت. مانند جامعه ستارگان، جامعه پرندگان و یا جامعه ماهیان دریاها.

صفت متغیر

عناصر جامعه بجز صفت مشخصه که در همه اعضای آن مشترک است، خواص دیگری نیز دارند که باهم متفاوتند. این گونه خاصیتها یا صفات را که بین عناصر جامعه متفاوت هستند «صفات متغیر» گویند. به عبارت دیگر،

صفتی که از یک عضو به عضو دیگر جامعه می‌تواند تغییر نماید، «صفت متغیر» نامیده می‌شود.

مثلاً: در جامعهٔ ایرانیان، خصوصیات چگونگی سن، طول قد، وزن، محل تولد، درآمد و ... صفات متغیر می‌باشند. در ادامهٔ بحث، صفات متغیر را مانند متغیرها در ریاضی با حروف Z, Y, X و ... نشان خواهیم داد. صفات متغیر بر دو نوعند: صفت متغیر کیفی، صفت متغیر کمی.

— صفت متغیر کیفی: صفتی است که آن را برای اعضای جامعه (در سطح علوم امروزی) یا هیچ عدد یا کمیتی نتوان بیان کرد. مانند: جنس، نوع شغل، نوع بیماری، مرغوبیت کالا و ...
— صفت متغیر کمی: صفتی است که آن را برای اعضای جامعه با عدد یا کمیت بتوان بیان کرد. مانند: طول قد افراد، درآمد افراد، تعداد اعضاء خانوار و ...

از صفات کمی: دو نوع مهم آن را مورد بحث قرار خواهیم داد:

— صفت کمی گسسته (نایبسته)

— صفت کمی پیوسته

صفت کمی گسسته: صفتی است که دامنه تعریف آن مجموعه شمارا^۱ (مجموعه نامحدود شمارش پذیر) یا هر زیر مجموعه‌ای از آن مجموعه می‌باشد.

مانند تعداد اعضای خانوار در جامعه خانوارهای یک منطقه، تعداد دانش‌آموزان یک دبیرستان، تعداد تصادفات اتومبیل در روز در یک شهر.

صفت کمی پیوسته: صفتی است که دامنه تعریف آن را مجموعه اعداد حقیقی یا هر فاصله‌ای از آن مجموعه تشکیل می‌دهد.

مانند طول قد انسان، قطر داخلی مهره‌های ساخته شده توسط یک دستگاه که با یک عدد حقیقی بیان شوند.

۱- هر مجموعه نامحدود هم‌ارز با مجموعه اعداد طبیعی را مجموعه نامحدود شمارش پذیر یا مجموعه شمارا گویند.

مراحل مختلف انجام یک تحقیق آماری

تعیین هدف تحقیق: وقتی یک جامعه آماری را مورد مطالعه قرار می‌دهیم، ابتدا باید هدف این بررسی را مشخص کنیم. پس از تعیین هدف، جامعه آماری که مطالعه در آن انجام می‌گیرد باید کاملاً تعریف شود. همچنین صفت مشخصه و صفات متغیر مورد بررسی نیز باید دقیقاً روشن گردد.

برای مثال اگر بخواهیم درآمد خانوارها را در یک جامعه بررسی کنیم، قبل از هر چیز باید معلوم کنیم که هدف از مطالعه درآمد خانوارها چیست؟ ممکن است برای آگاهی از سطح درآمد خانوارها و مقایسه با هزینه‌های زندگی، این کار انجام پذیرد تا معلوم شود آیا بین درآمد و هزینه خانوار در جامعه مورد بررسی تعادلی وجود دارد یا خیر؟ بنابراین هدف از این بررسی، مقایسه سطح درآمد خانوارها و هزینه خانوارها به منظور آگاهی از وجود تعادل بین آنهاست. سپس جامعه مورد بررسی را تعریف می‌کنیم. ممکن است جامعه مورد بررسی، خانوارهای شهر اراک باشند. در نتیجه کلیه خانوارهایی که در شهر اراک ساکن هستند، جامعه مورد بررسی را تشکیل می‌دهند. صفات متغیر این بررسی می‌تواند میزان درآمد ماهانه خانوارها، تعداد نفرات خانوارها، هزینه ماهانه خانوارها و ... باشد.

پس از تعیین هدف و جامعه مورد بررسی، برای انجام یک تحقیق آماری، سه مرحله اساسی زیر باید رعایت شود:

— مشاهده آماری: قبل از هر چیز باید صفت متغیر مورد مطالعه در یک جامعه برای اعضای آن جامعه مشخص گردد.

«مشخص کردن صفت متغیر برای هر یک از اعضای جامعه را مشاهده آماری می‌نامند».

نتایج هر مشاهده آماری اعضای جامعه در یک فرم آماری به نام پرسشنامه ثبت می‌گردد. مثلاً با مراجعه به خانوارهای یک منطقه، ثبت اطلاعاتی نظیر درآمد، هزینه، تعداد اعضای خانوار و ... در واقع مشاهده صفت متغیر «درآمد خانوار»، «هزینه خانوار»، «تعداد اعضای خانوار» و ... می‌باشد. بسته به اینکه مشاهده آماری، تمامی اعضای جامعه را یا قسمتی از جامعه آماری را (زیرمجموعه‌ای از آن) دربرگیرد، دو نوع مشاهده آماری را از هم مجزا می‌کنند.

— مشاهده سرتاسری (تمام شماری)

— مشاهده غیر سرتاسری (یا مشاهده نمونه‌ای)

در مشاهده سرتاسری، تمامی اعضای جامعه، مورد مطالعه قرار می‌گیرند. لیکن به علل زیر امکان مشاهده سرتاسری در جامعه فراهم نمی‌باشد.

— هزینه زیاد مشاهده

– دگرگونی اعضای جامعه

– نامحدود بودن حجم جامعه

مثلاً، اگر بخواهیم تمامی اعضای جامعه انسانی را در یک کشور از نظر «سلامت بودن» مشاهده کنیم، آن چنان هزینه زیادی به بودجه کشور تحمیل می‌گردد، که هیچ کشوری تن به چنین مشاهده‌ای نمی‌دهد. یا اگر بخواهیم طول عمر لامپهای تولید شده در یک کارخانه تولیدی را مطالعه کنیم، تمامی لامپهای ساخته شده باید به شبکه برق متصل گردد و پس از سوختن تمام لامپها، اطلاع راجع به طول عمر تمامی لامپها به دست خواهد آمد که چنین جامعه‌ای دیگر وجود ندارد و نابود شده است. چنین اطلاعاتی فاقد هرگونه ارزشی است یا مطالعه شدت حرارت ستارگان (درجه حرارت)، به علت نامحدود بودن حجم جامعه غیرممکن می‌باشد، در چنین شرایطی محقق ناچار به انتخاب جزئی از اعضای جامعه به عنوان نمونه می‌باشد که مشاهده غیرسرتاسری نامیده می‌شود.

– پردازش داده‌های آماری: اطلاعات به دست آمده در مرحله مشاهده (داده‌های آماری) را منظم کرده، آنها را برحسب مقادیر یا حالات صفت متغیر گروه‌بندی می‌کنند و برای بیان اجمالی آنها از جداول و نمودارهای آماری کمک می‌گیرند. این عملیات را «پردازش داده‌های آماری» گویند.

– مرحله تجزیه و تحلیل نتایج مشاهدات (داده‌های آماری): پس از گروه‌بندی نتایج مشاهدات برحسب صفت متغیر، مشخصه‌های عددی توصیف‌کننده صفت متغیر در جامعه محاسبه می‌شود و بر پایه آنها نسبت به چگونگی صفت متغیر در جامعه، قضاوت و یا نتیجه‌گیری صورت می‌گیرد. این مرحله از مطالعه جامعه آماری را «مرحله تجزیه و تحلیل آماری» می‌نامند.

اندازه‌گیری^۱ و انواع اندازه‌گیری (مقیاسهای اندازه‌گیری)

گفته شد که برای مطالعه تغییرات صفت در جامعه، اکثر اوقات، نتیجه مشاهده با اعدادی بیان می‌شوند که از اندازه‌گیری صفت متغیر کمی و یا از نسبت دادن اعداد به حالت‌های صفت کیفی به دست می‌آیند. از این‌رو، روشن ساختن مفهوم اندازه‌گیری ضرورت پیدا می‌کند.

نسبت دادن اعداد بر طبق قواعد معین به هر یک از اعضای جامعه را، اندازه‌گیری صفت متغیر در آن جامعه می‌نامند.

برای بیان تغییرپذیری صفت متغیر در جامعه، اکثر اوقات نتیجه مشاهده را با اعدادی بیان می‌کنند. هر یک از اعداد نسبت داده شده به اعضای جامعه را اندازه آن صفت گویند.

در اینجا چهار نوع مهم اندازه‌گیری را (مقیاس اندازه‌گیری^۱) مورد بحث قرار می‌دهیم:

- مقیاس اسمی
- مقیاس ترتیبی
- مقیاس فاصله‌ای
- مقیاس نسبتی

- مقیاس اسمی^۲: در اندازه‌گیری به طریق «مقیاس اسمی» از اعداد به منظور جدا کردن عناصر طبقات مختلف، استفاده می‌شود. اعدادی که به نتایج مشاهدات نسبت داده می‌شوند، نقش «علامت» یا «نام» را برای طبقات دارند که مشاهدات در آنها قرار می‌گیرند. انتخاب نام «مقیاس اسمی» نیز از همین جا حاصل شده است. در این اندازه‌گیری، اعدادی که به عناصر نسبت داده می‌شوند، هیچ معنای کمی ندارند و نمی‌توان بر روی آنها هیچ‌گونه عملیات جبری (چهار عمل اصلی) انجام داد. فرض کنید می‌خواهیم افراد یک جامعه را برحسب جنس گروه‌بندی کنیم در این صورت مثلاً به هریک از مردها عدد ۱ و به هریک از زن‌ها عدد ۲ را نسبت داده، آنها را به دو طبقه تقسیم می‌کنیم. نسبت دادن اعداد ۱ و ۲ به افراد جامعه، به منظور طبقه‌بندی آنها، اندازه‌گیری با مقیاس اسمی می‌باشد.

- مقیاس ترتیبی یا مقیاس رتبه‌ای^۳: در اندازه‌گیری با مقیاس ترتیبی، از اعداد برای مقایسه عناصر از نظر کوچکتر یا بزرگتر یا برابر بودن استفاده می‌شود که براساس نتایج این مقایسه‌ها نیز عناصر جامعه طبقه‌بندی می‌گردند. در این مقیاس، عناصر براساس اندازه‌های نسبی که برای آنها به دست می‌آید، مرتب می‌شوند. بر روی اعداد مقیاس ترتیبی نیز هیچ‌گونه عملیات جبری انجام نمی‌گیرد. مثال: در یک مؤسسه تولیدی که ۱۵ نفر کارگر مشغول انجام یک نوع کار هستند، به منظور طبقه‌بندی آنها برحسب مهارت (صفت متغیر)، می‌توان آنها را در چهار گروه: کارگر ساده، کارگر نیمه‌ماهر، کارگر ماهر و استادکار طبقه‌بندی نمود.

بدین منظور، به هریک از کارگران بسته به مهارت آنها، یکی از اعداد ۱، ۲، ۳ یا ۴ را نسبت می‌دهیم. نتایج به دست آمده به همان ترتیبی که مشاهده شده در زیر آورده شده است:

۲، ۴، ۱، ۳، ۳، ۲، ۲، ۳، ۱، ۴، ۱، ۳، ۳، ۲، ۳

براساس اعدادی که به آنها نسبت داده شده است، آنها را به طبقات زیر تقسیم‌بندی می‌کنیم:

(۴، ۴) و (۳، ۳، ۳، ۳، ۳، ۳) و (۲، ۲، ۲، ۲) و (۱، ۱، ۱)

۱- Scale of measurement

۲- Nominal scale

۳- Ordinal scale

نتیجه به دست آمده، طبقه‌بندی موردنظر است. هر طبقه، با رتبه داده شده به اعضای آن، از طبقه دیگر متمایز می‌گردد.

در بعضی موارد مقایسه عناصر و یا مقایسه چند گروه براساس امتیازهای داده شده به آنها با عدد برابر مشخص می‌گردند. مثلاً فرض کنیم به منظور رتبه‌بندی به صورت شاگرد اول، شاگرد دوم و شاگرد سوم در یک کلاس، پس از آزمونهای لازم، برای ۱۵ دانش‌آموز، نمرات زیر به دست آمده باشد:

$$۲۰, ۱۹, ۱۹, ۱۹, ۱۹, ۱۶, ۱۵, ۱۴, ۱۳, ۱۱, ۱۰, ۹, ۷, ۶, ۴$$

اگر در مشاهده، چند عضو اندازه‌های برابر داشته باشند، می‌گویند گروه هم‌رتبه‌ها به وجود آمده است. بنابراین طبق تعریف فوق، بین دانش‌آموزان، یک گروه (۱۹, ۱۹, ۱۹) هم‌رتبه وجود دارد. در بعضی مواقع در تحلیل آماری صفت متغیر در جامعه، نیاز است که عناصر در ترتیب معین با رتبه‌های یگانه مشخص گردند. در چنین مواردی، از رتبه‌های یکسان مُعدّل گرفته، آنها را ادغام می‌کنند و به جای رتبه هریک از آنها، مُعدّل رتبه‌ها را قرار می‌دهند. با یک مثال، این مطلب را روشن می‌سازیم.

در مثال بالا، عدد ۱ را به عنوان رتبه، به دانش‌آموزی که دارای بالاترین نمره است، نسبت می‌دهیم. عدد ۲ را به نمره پایین‌تر از رتبه اول به عنوان رتبه دوم نسبت می‌دهیم همچنین تا آخر. به دانش‌آموز با نمره ۲۰ رتبه اول را می‌دهیم ولی به دانش‌آموز با نمره ۱۹ رتبه دوم را می‌دهیم. در اینجا ملاحظه می‌شود که ۴ دانش‌آموز، نمره ۱۹ گرفته‌اند، بنابراین باید به آنها رتبه دوم تا رتبه پنجم را اختصاص دهیم. اما از آنجا که ارزش نمره آنها با هم یکسان است از رتبه ۲ تا ۵ مُعدّل گرفته و عدد به دست آمده را برای نفرات دوم تا پنجم قرار می‌دهیم.

$$\frac{۲+۳+۴+۵}{۴} = ۳/۵$$

درواقع برای نفرات دوم تا پنجم رتبه ۳/۵ را در نظر می‌گیریم. در مورد سایر نمرات که تکرار شده‌اند نیز به همین ترتیب عمل می‌کنیم. در نتیجه دانش‌آموزان موردنظر، برحسب رتبه به صورت زیر قرار خواهند گرفت:

$$۱, ۳/۵, ۳/۵, ۳/۵, ۳/۵, ۶, ۷, ۸, ۹, ۱۰, ۱۱, ۱۲, ۱۳, ۱۴, ۱۵$$

— مقیاس فاصله‌ای^۱: در اندازه‌گیری به شیوه «مقیاس فاصله‌ای» عناصر مورد اندازه‌گیری نه تنها از نظر ترتیب اندازه می‌توانند مقایسه شوند، بلکه برحسب طول فاصله‌ها بین اندازه‌های عناصر نیز می‌توانند مقایسه و طبقه‌بندی گردند. «مقیاس فاصله‌ای» از مفهوم «واحد فاصله» استفاده می‌کند و

۱- Interval scale

بدین جهت می‌تواند فاصله را بین دو اندازه برحسب تعداد این «واحد فاصله»ها بیان کند. مثال روشن برای اندازه‌گیری به‌طریق مقیاس فاصله‌ای، مقیاسی است که با آن شدت حرارت را نشان می‌دهیم. «واحد فاصله» در اینجا برای اندازه‌گیری «درجه» است. مقدار عددی شدت حرارت به‌طور ساده، نتیجه مقایسه با یک نقطه دلخواه است که آن را «درجه صفر» می‌نامند. در اندازه‌گیری با مقیاس فاصله‌ای، به وجود «نقطه صفر» و «واحد فاصله» نیاز می‌باشد. ولی اهمیت ندارد که این نقطه با چه نوع اندازه‌گیری مشخص شده است و با چه فاصله‌ای به‌عنوان «واحد فاصله» تعریف شده است. شدت حرارت گاهی با مقیاس سانتیگراد و گاهی با مقیاس فارنهایت اندازه‌گیری می‌شود که دارای نقطه صفر و واحد فاصله متفاوت هستند.

برخلاف مقیاس اسمی و مقیاس ترتیبی که روی نتایج اندازه‌گیری، عملیات جبری (چهار عمل اصلی) غیرممکن است، در مقیاس فاصله‌ای انجام بعضی از عملیات جبری روی اعداد به‌دست آمده از اندازه‌گیری، امکان‌پذیر است. مثلاً در این مقیاس می‌توان گفت که $5 - 9 = 14 - 18$ و برای هر دو طرف تساوی می‌توان گفت، اندازه اولی چهار واحد بیشتر از دومی است. ولی در این مقیاس نمی‌توان گفت که اندازه یک عنصر چند برابر اندازه عنصر دیگر است. مثلاً در اندازه‌گیری شدت حرارت با مقیاس سانتیگراد نمی‌توان گفت که درجه حرارت 5° دو برابر درجه حرارت 25° می‌باشد. مثال دیگر مقیاس تقویمی است که در سال شمسی مبدأ آن، هجرت حضرت محمد (ص) از مکه به مدینه می‌باشد و در سال میلادی مبدأ آن، میلاد حضرت مسیح (ع) است.

— مقیاس نسبتی^۱: این مقیاس هم زمانی که اعداد به‌دست آمده از اندازه‌گیری، از نظر ترتیب و فاصله بین دو مقدار، دارای اهمیت هستند، به‌کار می‌رود و هم وقتی که نسبت دو عدد به‌دست آمده از اندازه‌گیریها، دارای اهمیت باشد. به عبارت دیگر، مقیاس نسبتی امکان می‌دهد که تعیین کنیم یک اندازه از اندازه دیگر چقدر بیشتر یا کمتر است و نیز می‌توان تعیین کرد این اندازه برای یک عنصر چند برابر اندازه عنصر دیگر است. وقتی اندازه‌گیری به‌منظور مقایسه دو عنصر، به‌صورت نسبت اندازه‌های آنها باشد، استفاده از اندازه‌گیری با مقیاس نسبتی اجتناب‌ناپذیر است. تنها اختلاف اساسی بین مقیاس نسبتی و مقیاس فاصله‌ای در این است که برای مقیاس نسبتی اندازه طبیعی به نام «صفر مطلق» وجود دارد، در صورتی که در اندازه‌گیری بر طبق مقیاس فاصله‌ای، اندازه صفر به‌طور دلخواه تعیین می‌گردد. در مقیاس نسبتی نیز مانند مقیاس فاصله‌ای، «واحد فاصله» بین دو اندازه برای عناصر به‌طور دلخواه تعیین می‌شود. برخلاف مقیاس فاصله‌ای که انجام عمل تقسیم روی اعداد

به دست آمده از اندازه گیری غیرممکن می باشد، در مقیاس نسبتی انجام این عمل امکان پذیر است و می توان بین نسبت‌هایی از اعداد که از اندازه گیریها به دست می آیند، تساوی برقرار کرد. مثلاً در این مقیاس می توان گفت: $\frac{1}{4} \cdot \frac{6}{12}$ که نتیجه وجود نقطه صفر تثبیت شده در این مقیاس می باشد و برای دو اندازه ۱۲ و ۶ می توان گفت که اندازه اولی ۲ برابر اندازه دومی است.

صفت‌های متغیر کمی در مقیاس فاصله ای و نسبتی اندازه گیری می شوند ولی صفت‌های متغیر کیفی در سطح مقیاس اسمی و ترتیبی اندازه گیری می شوند.

علامت زیگما (Σ) و کاربرد آن در محاسبات

از آنجا که در بسیاری از فرمول‌های آماری از علامت زیگما (Σ) استفاده می شود، به منظور کاربرد صحیح آن، ضروری است با خصوصیات این علامت آشنا شویم.

قاعده ۱: برای نشان دادن عمل جمع چند عدد، از حرف یونانی (زیگما) استفاده می شود. اگر n مقدار $x_1, x_2, x_3, \dots, x_n$ در اختیار باشد، حاصل جمع آنها با استفاده از حرف زیگما می تواند به صورت $\sum_{i=1}^n x_i$ (بخوانید زیگمای x_i از ۱ تا n) نشان داده می شود. بنابراین داریم:

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_n$$

وقتی دامنه مقادیر از محتوای مسأله آشکار باشد می توان بدون نوشتن اعداد برای اندیس i حاصل جمع n عدد را به طور ساده تر به صورت x_i نوشت.

قاعده ۲: ضرب مقادیر x_i در عدد ثابت a

اگر همه اندازه‌های صفت را در عدد ثابت a ضرب کرده، سپس آنها را با هم جمع کنیم، نتیجه را با استفاده از علامت Σ می توان به صورت زیر ساده کرد:

$$ax_1 + ax_2 + ax_3 + \dots + ax_n = a(x_1 + x_2 + \dots + x_n) = a \sum_{i=1}^n x_i$$

مثال: می خواهیم مقادیر متغیر x را که ۱۴، ۱۲، ۱۰، ۸، ۶، ۴، ۲ می باشند، در عدد ۵ ضرب کرده، حاصل ضربها را جمع کنیم.

با استفاده از قاعده ۲ خواهیم داشت:

$$\begin{aligned} 5x_i &= 5 \cdot 2 + 5 \cdot 4 + 5 \cdot 6 + 5 \cdot 8 + 5 \cdot 10 + 5 \cdot 12 + 5 \cdot 14 \\ &= 5(2 + 4 + 6 + 8 + 10 + 12 + 14) = 5 \cdot 56 = 280 \end{aligned}$$

۱- حرف زیگما، معادل حرف S انگلیسی در زبان یونانی می باشد که حرف اول واژه Sum به معنای مجموع است.

قاعدهٔ ۳: حاصل جمع مجموع چند متغیر برابر است با مجموع حاصل جمع آن چند متغیر:

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$$

$$\begin{aligned} \sum_{i=1}^n (x_i + y_i) &= (x_1 + y_1) + (x_2 + y_2) + \dots + (x_n + y_n) \\ &= (x_1 + x_2 + \dots + x_n) + (y_1 + y_2 + \dots + y_n) \\ &= \sum_{i=1}^n x_i + \sum_{i=1}^n y_i \end{aligned}$$

قاعدهٔ ۴: اگر a مقدار ثابتی باشد، خواهیم داشت:

$$\begin{aligned} \sum_{i=1}^n a &= na \\ \sum_{i=1}^n a &= a + a + \dots + a = na \end{aligned}$$

سؤالها و تمرینها

- ۱- آمار را به عنوان یک علم تعریف کنید.
- ۲- جامعه آماری چیست؟
- ۳- صفت مشخصه جامعه آماری چیست؟
- ۴- فرد جامعه چیست؟
- ۵- حجم جامعه آماری چیست؟
- ۶- جامعه محدود و جامعه نامحدود را تعریف کرده، برای هر کدام مثالی بیاورید.
- ۷- صفت متغیر چیست؟
- ۸- صفت متغیر کمی چیست؟ تعریف کنید و یک مثال بیاورید.
- ۹- صفت متغیر کیفی چیست؟ تعریف کنید و یک مثال بیاورید.
- ۱۰- انواع صفت کمی را نام برده، برای هر یک مثالی بیاورید.
- ۱۱- برای به دست آوردن داده‌های آماری، چه اقدامی باید انجام گیرد؟
- ۱۲- مشاهده آماری چیست و به چند طریق انجام می‌گیرد؟
- ۱۳- مشاهده سرتاسری چه نوع مشاهده‌ای است؟ مشکلات آن را نام ببرید.

۱۴- مشاهده غیر سرتاسری چه نوع مشاهده‌ای است و در چه مواقعی الزاماً به کار

می‌رود؟

۱۵- در انجام یک تحقیق آماری چه مراحل را باید رعایت نمود؟

۱۶- پردازش آماری شامل چه اقداماتی است؟

۱۷- تحلیل آماری به چه منظور انجام می‌گیرد؟

۱۸- اندازه‌گیری صفت متغیر چیست؟

۱۹- مقیاس اندازه‌گیری چیست؟

۲۰- چند نوع مقیاس اندازه‌گیری مهم مورد استفاده قرار می‌گیرد؟

۲۱- در اندازه‌گیری به شیوه «مقیاس اسمی» اعداد نسبت داده شده به عناصر، چه معنایی

دارند؟

۲۲- آیا روی اعداد نسبت داده شده در مقیاس اسمی، می‌توان عملیات جبری را انجام داد؟

چرا؟

۲۳- در مقیاس ترتیبی، چگونه اعداد به عناصر مورد اندازه‌گیری نسبت داده می‌شوند؟ آیا

روی این اعداد، می‌توان عملیات جبری را انجام داد؟ چرا؟

۲۴- تفاضل دو اندازه ۷ و ۴ را در اندازه‌گیری به طریق مقیاس فاصله‌ای چگونه تعبیر می‌کنید؟

۲۵- اندازه‌گیری به شیوه مقیاس نسبتی، چه خصوصیت‌هایی از اعداد به دست آمده در این نوع

اندازه‌گیری را در نظر می‌گیرد؟ در چه مواردی استفاده از مقیاس نسبتی در اندازه‌گیریها، اجتناب‌ناپذیر

است؟

۲۶- اختلاف اساسی بین اندازه‌گیری به شیوه مقیاس نسبتی و مقیاس فاصله‌ای در چیست؟

۲۷- نماد (!)، به چه منظور در آمار وارد شده است؟

۲۸- چه قواعدی در محاسبات برای علامت ! وجود دارد؟

۲۹- اگر x_1, x_2, x_3, x_4, x_5 به ترتیب اعداد زوج متوالی ۲، ۴، ۶، ۸، ۱۰ باشند، مقادیر

$$\sum_{i=1}^n x_i, \sum_{i=2}^4 x_i, \text{ و } (\sum_{i=1}^n x_i)^2 \text{ را محاسبه نمایید.}$$

۳۰- در تمرین ۲۹ اگر $a=10$ باشد، مقادیر: $\sum_{i=1}^n ax_i$ و $\sum_{i=1}^n a^2 x_i^2$ و $\sum_{i=1}^n (x_i + a)$ و

$$\sum_{i=1}^n (x_i - a)^2 \text{ را محاسبه نمایید.}$$

خودآزمونهای چهارگزینه‌ای ✓

- ۱- کدام یک از گزینه‌های زیر معرف جامعه آماری است؟
 - الف - اعضای آن دارای خواص مختلفی باشند.
 - ب - همه اعضای آن حداقل یک خاصیت مشترک داشته باشند.
 - ج - اعضای آن دارای چند صفت متغیر باشند.
 - د - اعضای آن متعلق به یک گروه همگن باشند.
- ۲- کدام تعریف برای صفت مشخصه جامعه صحیح است؟
 - الف - صفتی است که از فردی به فرد دیگر تغییر می‌کند.
 - ب - صفتی است که برای اعضای با عدد یا کمیت قابل بیان باشد.
 - ج - صفت مشترک برای افراد جامعه است.
 - د - صفتی است که قابل شمارش است.
- ۳- کدام یک از گزینه‌های زیر تعریف صفت متغیر کمی است؟
 - الف - صفتی است که با مجموعه اعداد گویا بیان شود.
 - ب - صفتی است که با عدد یا کمیت قابل بیان باشد.
 - ج - با عدد یا کمیت نتوان آن را بیان کرد.
 - د - با اعداد مثبت بیان شود.
- ۴- کدام یک از گزینه‌های زیر تعریف صفت متغیر کیفی است؟
 - الف - با مجموعه اعداد صحیح بیان شود.
 - ب - با هیچ عدد یا کمیتی بیان نشود.
 - ج - با مجموعه اعداد گویا بیان نشود.
 - د - با عدد یا کمیت بیان شود.
- ۵- کدام یک از گزینه‌های زیر، تعریف صفت کمی پیوسته است؟
 - الف - صفتی است که دامنه تعریف آن مجموعه اعداد شمارش پذیر باشد.
 - ب - صفتی است که دامنه تعریف آن مجموعه اعداد حقیقی یا هر فاصله‌ای از آن باشد.
 - ج - صفتی است که دامنه تعریف آن مجموعه اعداد گویا باشد.
 - د - صفتی است که با مجموعه اعداد اصم بیان شود.

۶- مشاهده آماری یعنی :

الف - مشخص کردن صفت متغیر برای هر یک از اعضای جامعه

ب - ثبت حالات صفات کیفی در جامعه

ج - ثبت مقدار صفت متغیر کمی برای اعضای جامعه

د - اندازه گیری صفت متغیر کمی برای اعضای جامعه

۷- مهمترین مشکلات مشاهده سرتاسری کدامند؟

الف - عدم دسترسی به تمامی اعضای جامعه

ب - نداشتن پرسنل مجرب برای آمارگیری

ج - هزینه زیاد، نامحدود بودن جامعه و بعضاً انهدام جامعه

د - طولانی شدن زمان پردازش داده‌های آماری

۸- کدام یک از گزینه‌های زیر، مفهوم اندازه گیری برای صفت متغیر جامعه را بیان می کند.

الف - وسیله ای که با آن صفات را اندازه گیری می کنند.

ب - نسبت دادن اعداد برطبق قواعد معین به اعضای جامعه برای مشخص شدن

صفت متغیر آن

ج - اندازه گیری طول صفات متغیر در جامعه

د - نسبت اندازه صفات متغیر در جامعه

۹- روی کدام یک از مقیاسهای اندازه گیری، می توان عملیات جبری (چهار عمل اصلی) را

انجام داد؟

الف - مقیاس اسمی ب - مقیاس فاصله ای

ج - مقیاس رتبه ای د - مقیاس نسبی

۱۰- ! در ضرب اعداد چه مفهومی دارد؟

الف - فاکتورگیری از ضرایب ؛ Xها (متغیرها)

ب - حاصل جمع، حاصلضربهای عدد ثابت در متغیرها

ج - تفکیک ضرایب از مقادیر صفت

د - تفکیک مقادیر صفت از ضرایب

گروه‌بندی و پردازش داده‌های آماری

- هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند:
- ۱- توضیح دهد که پردازش آماری به چه منظوری انجام می‌گیرد.
 - ۲- مفهوم گروه‌بندی و نقش آن را برای توصیف آماری توضیح دهد.
 - ۳- نتایج مشاهدات را در یک جامعه آماری گروه‌بندی نماید.
 - ۴- مفهوم فراوانی را توضیح دهد.
 - ۵- توزیع فراوانیهای صفت متغیر را در جامعه بیان کند.
 - ۶- جدول «توزیع فراوانیهای صفت متغیر» را با استفاده از نتایج مشاهدات، تشکیل دهد.
 - ۷- نمودار جدول «توزیع فراوانیهای صفت متغیر» را رسم کند.
 - ۸- مفهوم «فراوانی انباشته» را تعریف نماید.
 - ۹- براساس نتایج مشاهدات، فراوانیهای انباشته را محاسبه کرده، توزیع فراوانیهای انباشته را تشکیل دهد و نقش آن را به‌عنوان مشخصه صفت متغیر توضیح دهد.
 - ۱۰- لزوم واردکردن فاصله‌ها را برای مقادیر یک صفت متغیر نشان دهد.
 - ۱۱- لزوم واردکردن «مفهوم چگالی فراوانیها» را بیان کند.
 - ۱۲- جدول توزیع چگالی فراوانی را تشکیل داده، نمودار آن را نیز رسم کند.

پردازش داده‌های آماری — توزیع صفت متغیر

پردازش داده‌های آماری^۱

داده‌های آماری که در مرحله مشاهده به‌دست آمده است، اعداد یا ارقامی هستند که هیچ‌گونه

مفهوم خاصی نداشته و نیاز به پردازش دارند تا آنها را به صورت یک مجموعه ادغام شده درآورد. پردازش نتایج مشاهدات، شامل منظم کردن، طبقه بندی یا گروه بندی داده ها، تشکیل جداول، محاسبه سرجمعها و مشخصه های عددی است که آن را «مرحله پردازش» می نامند. بنابراین هدف از پردازش، ادغام نتایج مشاهدات، یا تراکم کردن اطلاعات، به دست آوردن مشخصه های کلی به منظور توصیف اجمالی خصوصیات جامعه ها می باشد.

توزیع صفت متغیر

مطالعه صفت در جامعه، بیشتر به منظور آشکار ساختن تغییر پذیری صفت متغیر صورت می گیرد. برای این کار لازم است عناصر جامعه در گروه های یکسان دسته بندی شوند. گروه بندی نتایج مشاهدات، یکی از اساسی ترین روشهای پردازش و تحلیل اولیه اطلاعات آماری است.

اگر صفت متغیر مورد بررسی کمی گسسته باشد، مقادیر مشاهده شده یکسان را که با یک عدد بیان می شوند به عنوان یک گروه در نظر گرفته، تعداد آنها را مشخص می کنیم. سپس تمامی مقادیر مختلف صفت متغیر را به ترتیب صعودی مرتب نموده در مقابل هریک از این مقادیر تعداد آنها (فراوانی) را می نویسیم. نتیجه گروه بندی به صورت زیر به دست می آید:

$$X: x_1 \quad x_2 \quad \dots \quad x_k$$

$$F_i: F_1 \quad F_2 \quad \dots \quad F_k$$

دو مجموعه مقادیر صفت متغیر X و فراوانیها را با هم «توزیع فراوانیهای صفت متغیر» یا به طور ساده «توزیع صفت متغیر» می نامند.

تعداد عناصر را در هر گروه، «فراوانی مطلق^۱» آن گروه می نامند و بر طبق معمول آن را با F_i نشان می دهند.

واضح است که برای فراوانیهای مطلق در یک توزیع صفت، همواره تساوی زیر برقرار است:

$$F_1 + F_2 + F_3 + \dots + F_k = N \quad (1)$$

به عبارت دیگر، مجموع فراوانیهای مطلق صفت متغیر در جامعه، همواره برابر N یعنی حجم جامعه می باشد.

۱- Frequency

۲- Absolute Frequency

توزیع صفت متغیر، چگونگی تغییرات فراوانیها را برحسب تغییرپذیری صفت متغیر منعکس می‌نماید. توزیع صفت متغیر را با جدول نیز به صورت زیر می‌توان بیان نمود:

جدول ۱

X	x_1	x_2	x_3	x_k	-
F_i	F_1	F_2	F_3	F_k	N

چنین جدولی را «جدول توزیع فراوانیهای مطلق» می‌نامند.

مثال ۱: فرض کنیم جامعه مورد مطالعه $N=20$ خانوار باشد که صفت متغیر X در آن تعداد اعضای هر خانوار است. برای هر خانوار، تعداد اعضای آن با پرسش تعیین شده و نتایج مشاهدات به صورت زیر به دست آمده است:

$X: 2, 4, 3, 4, 4, 5, 2, 2, 4, 5, 5, 3, 4, 4, 3, 3, 2, 4, 3, 5$

اگر مقادیر مختلف صفت متغیر X را که به ترتیب ۲، ۳، ۴ و ۵ می‌باشند، در ۴ گروه طبقه‌بندی نموده، تعداد تکرار آنها (فراوانیها) را در برابر آنها قرار دهیم، گروههایی به صورت زیر به دست می‌آیند (جدول ۲):

جدول ۲

X تعداد اعضای خانوار	۲	۳	۴	۵	-
F_i تعداد خانوار (فراوانی)	۴	۵	۷	۴	$F_i = 20$

اگر فراوانیهای مطلق (F_i) را برحجم جامعه (N) تقسیم کنیم، توزیع فراوانیهای نسبی^۱ را به دست خواهیم آورد. فراوانیهای نسبی را با نماد f_i نشان خواهیم داد که از رابطه زیر به دست می‌آید:

$$f_i = \frac{F_i}{N} \quad (2)$$

فراوانی نسبی گروه نام در جدول توزیع فراوانیها، سهم هریک از مقادیر صفت متغیر را در کل جامعه نشان می‌دهد.

۱- Relative Frequency

حاصل جمع فراوانیهای نسبی در جدول توزیع فراوانی، همواره برابر ۱ می‌باشد.

$$f_1 + f_2 + f_3 + \dots + f_k = \sum_{i=1}^k f_i = 1 \quad (3)$$

توزیع فراوانیهای نسبی صفت متغیر را با جدول نیز می‌توان نشان داد :

جدول ۳- توزیع فراوانیهای نسبی صفت متغیر X

X	x_1	x_2	x_3	x_k	-
f_i	f_1	f_2	f_3	f_k	$f_i = 1$

برای مثال ۱، توزیع فراوانیهای نسبی را محاسبه می‌کنیم (جدول ۴) :

جدول ۴

X تعداد اعضای خانوار	۲	۳	۴	۵	-
f_i نسبت خانوارها در جامعه	$\frac{4}{20}$	$\frac{5}{20}$	$\frac{7}{20}$	$\frac{4}{20}$	۱

مثال ۲: سن کارگران یک کارخانه نساجی مورد پرسش واقع شده، نتایج به‌دست آمده در جدول زیر، گروه‌بندی شده است. می‌خواهیم جدول توزیع فراوانی نسبی سن کارگران را به‌دست آوریم :

جدول ۵

X سن	۱۹	۲۰	۲۲	۲۵	۲۶	۳۱	۳۴	۳۷	-
F_i فراوانی مطلق	۷	۶	۹	۱۰	۱۵	۱۲	۸	۱۳	۸۰

کافی است هریک از فراوانیهای مطلق (F_i) را به N (در اینجا $N=80$) تقسیم نماییم تا فراوانی نسبی هر گروه از سن کارگران به‌دست آید. نتیجه محاسبات در جدول ۶ آورده شده است :

جدول ۶

X سن	۱۹	۲۰	۲۲	۲۵	۲۶	۳۱	۳۴	۳۷	-
f_i فراوانی نسبی	$\frac{7}{80}$	$\frac{6}{80}$	$\frac{9}{80}$	$\frac{10}{80}$	$\frac{15}{80}$	$\frac{12}{80}$	$\frac{8}{80}$	$\frac{13}{80}$	$\frac{80}{80} = 1$

ملاحظه می‌شود جمع فراوانیهای نسبی برابر ۱ خواهد شد.

مفهوم فراوانی نسبی بهتر از فراوانی مطلق می‌تواند وضعیت صفت را در جامعه مشخص سازد، برای مثال اگر بگوییم $\frac{1}{8}$ کارگران سنی برابر ۲۵ سال دارند، بهتر است از اینکه بگوییم در این کارخانه ۱۰ کارگر ۲۵ ساله وجود دارد. چون فراوانی مطلق معلوم نمی‌کند از چند کارگر موجود در این کارخانه، ۱۰ نفر ۲۵ سال سن دارند، ولی فراوانی نسبی دقیقاً آن را مشخص می‌کند.

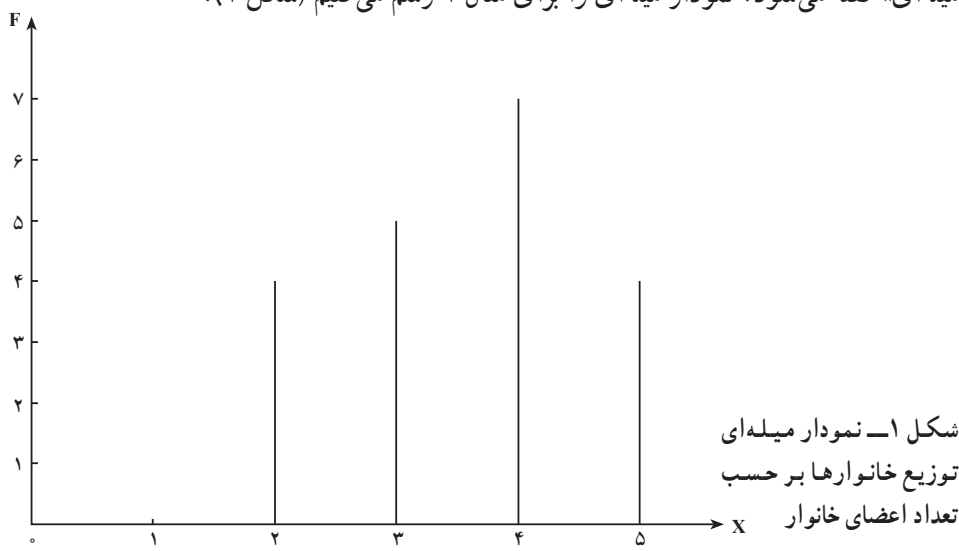
بیان هندسی توزیع صفت متغیر (نمودار میله‌ای^۱ – نمودار چندگوش^۲)

نمایش تغییرپذیری صفت متغیر با جداول توزیع فراوانیها و پی بردن به خصوصیات آن به‌ویژه وقتی این‌گونه جداول مفصل باشند، ساده نیست به طوری که از روی آنها به‌آسانی نمی‌توان تغییرات فراوانیها را از یک مقدار به مقدار دیگر صفت فهمید.

برای آشکار کردن آنها در آمار، غالباً از نمودار میله‌ای و نمودار چندگوش، استفاده می‌شود.

(البته این کار بیشتر اوقات زمانی که صفت متغیر کمی گسسته باشد صورت می‌گیرد).

برای ساختن نمودار میله‌ای در دستگاه مختصات (قائم)، بر روی محور طولها، مقادیر متغیر X و بر روی محور عرضها، فراوانیها، قرار داده می‌شوند. هر زوج F_i و X_i از توزیع فراوانیهای صفت متغیر، در دستگاه مختصات به صورت یک نقطه با طول X_i و عرض F_i نشان داده می‌شود. پس از رسم تمامی نقاط بر روی دستگاه مختصات، از نقاط به دست آمده، عمودهایی به محور X ها رسم می‌کنیم. به این ترتیب تصویری به دست می‌آید که از چند میله تشکیل شده است و به آن «نمودار میله‌ای» گفته می‌شود. نمودار میله‌ای را برای مثال ۱ رسم می‌کنیم (شکل ۱).

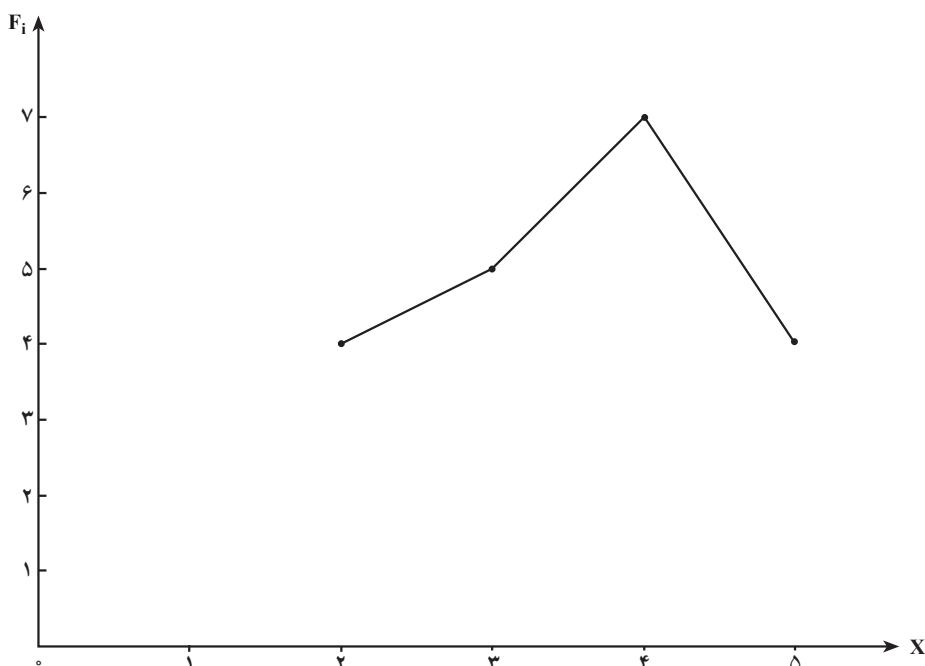


شکل ۱- نمودار میله‌ای
توزیع خانوارها بر حسب
تعداد اعضای خانوار

۱- Bar Diagram

۲- Polygon

با نمودار میله‌ای به سهولت می‌توان دریافت که خانوارهای ۴ نفره تعدادشان بیش از همه است و بعد از آن خانوارهای ۳ نفره و سپس خانوارهای ۲ و ۵ نفره به تعداد یکسان در این جامعه وجود دارند. برای رسم نمودار چندگوش، کافی است نقاط به دست آمده در دستگاه مختصات را با خطوط مستقیم به ترتیب به هم وصل نمود. در نتیجه شکلی حاصل می‌شود که از چند گوشه (زاویه) تشکیل شده است و به همین دلیل به آن «نمودار چندگوش توزیع فراوانیها» گفته می‌شود. این نوع نمودار بیشتر برای مقایسه توزیع فراوانیها در دو یا چند جامعه باهم مورد استفاده قرار می‌گیرد. چندگوش توزیع فراوانیها را برای مثال ۱ رسم می‌کنیم (شکل ۲):



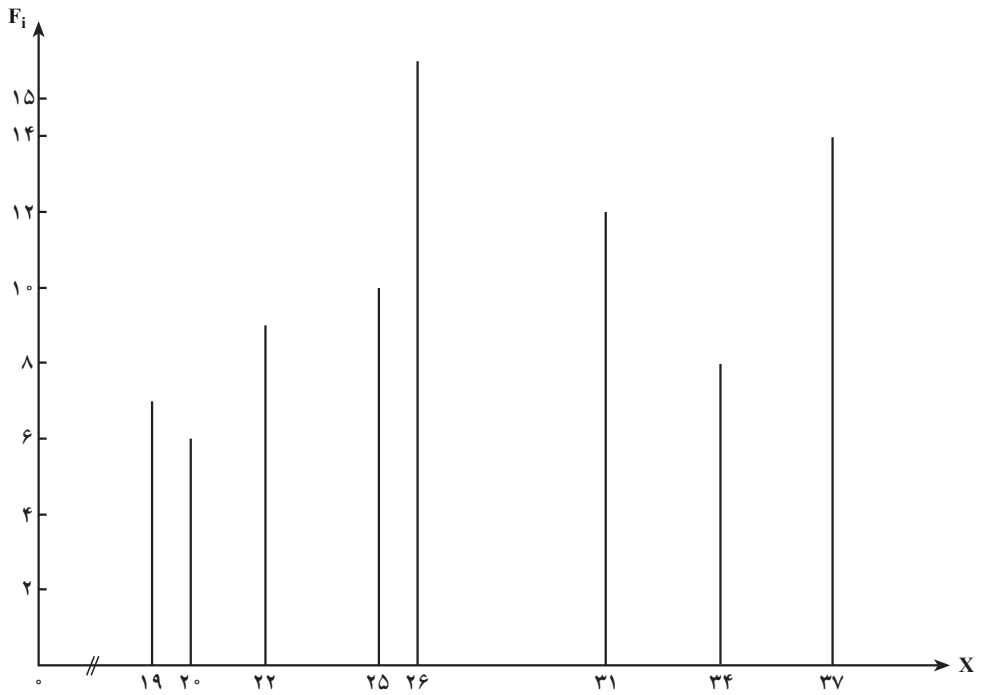
شکل ۲- نمودار چندگوش توزیع فراوانیهای تعداد اعضای خانوار

نمودار میله‌ای و چندگوش توزیع فراوانیها را برای مثال ۲ (سن کارگران کارخانه نساجی) رسم می‌کنیم (شکل‌های ۳ و ۴):

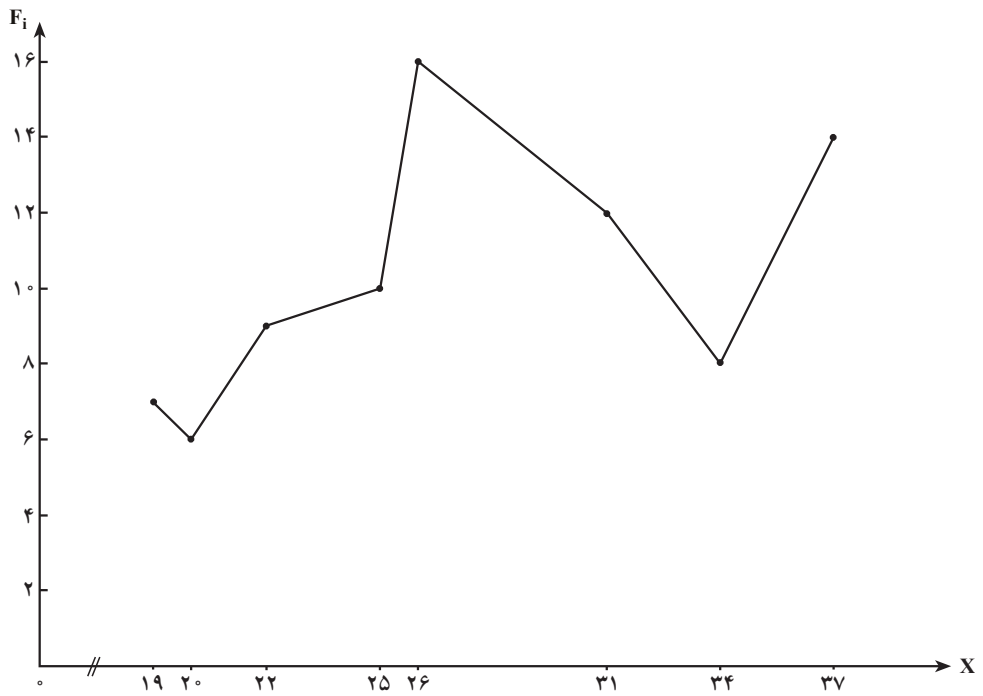
فراوانیهای تجمعی^۱ (انباشته)

فراوانیهای تجمعی نیز مانند فراوانیهای مطلق و نسبی، حجم گروهی از عناصر جامعه را بیان

^۱ - Cumulative Frequency



شکل ۳- نمودار میله‌ای سن کارگران کارخانه نساجی



شکل ۴- نمودار چندگوش سن کارگران کارخانه نساجی

می‌کند که به‌عنوان مشخصهٔ صفت متغیر، مورد استفاده قرار می‌گیرد و به‌صورت زیر تعریف می‌شود:

فراوانی تجمعی هر گروه از صفت متغیر در جدول توزیع فراوانی، برابر است با فراوانی مطلق همان گروه به‌علاوهٔ فراوانیهای مطلق گروههای ماقبل آن. فراوانی تجمعی (انباشته) مطلق را با نماد FC_i یا CF_i نشان می‌دهند.

اگر فراوانیهای مطلق را از ابتدا در جدول برای گروههای متوالی جمع کنیم، نهایتاً آخرین فراوانی انباشته (تجمعی) برابر با N یعنی حجم جامعه خواهد شد.
 فرم کلی محاسبات فراوانی انباشته به‌شکل زیر است (جدول ۷):

جدول ۷

X_i	F_i	FC_i
x_1	F_1	$FC_1 = F_1$
x_2	F_2	$FC_2 = F_1 + F_2$
x_3	F_3	$FC_3 = F_1 + F_2 + F_3$
x_k	F_k	$FC_k = F_1 + F_2 + F_3 + \dots + F_k = N$
—	N	—

فراوانیهای انباشته را برای مثال ۱ محاسبه می‌کنیم (جدول ۸):

جدول ۸

X_i	F_i	FC_i
۲	۴	$FC_1 = ۴$
۳	۵	$FC_2 = ۴ + ۵ = ۹$
۴	۷	$FC_3 = ۴ + ۵ + ۷ = ۱۶$
۵	۴	$FC_4 = ۴ + ۵ + ۷ + ۴ = ۲۰$
—	۲۰	—

ملاحظه می‌شود که آخرین فراوانی انباشته در جدول برابر 20° یعنی حجم جامعه می‌باشد. همین مفهوم به صورت نسبت، به نام فراوانیهای انباشته نسبی نیز بیان می‌شود که به صورت زیر تعریف می‌گردد:

نسبت هریک از فراوانیهای تجمعی (انباشته) در جدول توزیع فراوانی را به N (حجم جامعه) فراوانی انباشته نسبی همان گروه صفت متغیر می‌نامند و آنرا با نماد fc_i یا fc_i نشان می‌دهند.

$$fc_i = \frac{FC_i}{N} \quad (4)$$

از حاصل جمع تجمعی فراوانیهای نسبی در جدول توزیع فراوانی، به همان ترتیبی که فراوانیهای انباشته مطلق را محاسبه نمودیم نیز می‌توان فراوانی انباشته نسبی هریک از گروههای صفت متغیر را به دست آورد. جدول توزیع فراوانیهای انباشته مطلق و نسبی را برای مثال ۲ محاسبه می‌کنیم:

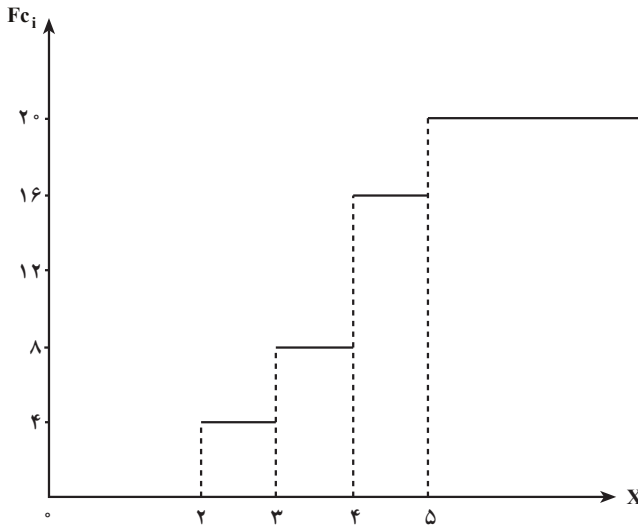
جدول ۹

X_i	F_i	f_i	FC_i	fc_i
۱۹	۷	$\frac{7}{80}$	۷	$\frac{7}{80}$
۲۰	۶	$\frac{6}{80}$	$7+6=13$	$\frac{7}{80} + \frac{6}{80} = \frac{13}{80}$
۲۲	۹	$\frac{9}{80}$	$13+9=22$	$\frac{13}{80} + \frac{9}{80} = \frac{22}{80}$
۲۵	۱۰	$\frac{10}{80}$	$22+10=32$	$\frac{22}{80} + \frac{10}{80} = \frac{32}{80}$
۲۶	۱۵	$\frac{15}{80}$	$32+15=47$	$\frac{32}{80} + \frac{15}{80} = \frac{47}{80}$
۳۱	۱۲	$\frac{12}{80}$	$47+12=59$	$\frac{47}{80} + \frac{12}{80} = \frac{59}{80}$
۳۴	۸	$\frac{8}{80}$	$59+8=67$	$\frac{59}{80} + \frac{8}{80} = \frac{67}{80}$
۳۷	۱۳	$\frac{13}{80}$	$67+13=80$	$\frac{67}{80} + \frac{13}{80} = \frac{80}{80} = 1$
—	80°	۱	—	$\frac{80}{80} = 1$

فراوانی انباشته برای بیان تعداد مقادیر صفت x معین که از یک مقدار مشخص تجاوز نکنند، به کار می‌رود. در مثال ۲ اگر بخواهیم بدانیم مجموع کارگرانی که سن آنها از ۲۶ سال تجاوز نمی‌کند،

چه تعدادی است، کافی است فراوانی انباشته ردیف ۲۶ سالگی را مشاهده کرد و آن را به دست آورد، که در اینجا این رقم برابر ۴۷ می باشد. همچنین، اگر بخواهیم نسبت کارگرانی را که سن آنها از ۲۶ سالگی تجاوز نمی کند، بدانیم، کافی است، فراوانی انباشته نسبی سن ۲۶ سالگی را در جدول قرائت نماییم. این نسبت در جدول برابر $\frac{47}{80}$ می باشد.

توزیع فراوانیهای انباشته را نیز می توان به صورت نمودار بیان نمود. برای این کار روی محور طولها، مقادیر X و روی محور عرضها، فراوانیهای انباشته مطلق یا نسبی را قرار می دهیم و برای هر زوج X_i و FC_i یا fc_i یک نقطه در صفحه مختصات به دست می آوریم و سپس با وصل کردن این نقاط به شکلی دست می یابیم که به آن «نمودار توزیع فراوانیهای تجمعی» (کومولات^۱) گویند. نمودار فراوانیهای انباشته را برای مثال ۱ رسم می کنیم (شکل ۵):



شکل ۵- نمودار فراوانیهای تجمعی تعداد اعضای خانوار

نمودار فراوانیهای انباشته را برای مثال ۲ رسم می کنیم (شکل ۶):

طریقه عملی گروه بندی نتایج مشاهدات (داده های آماری)

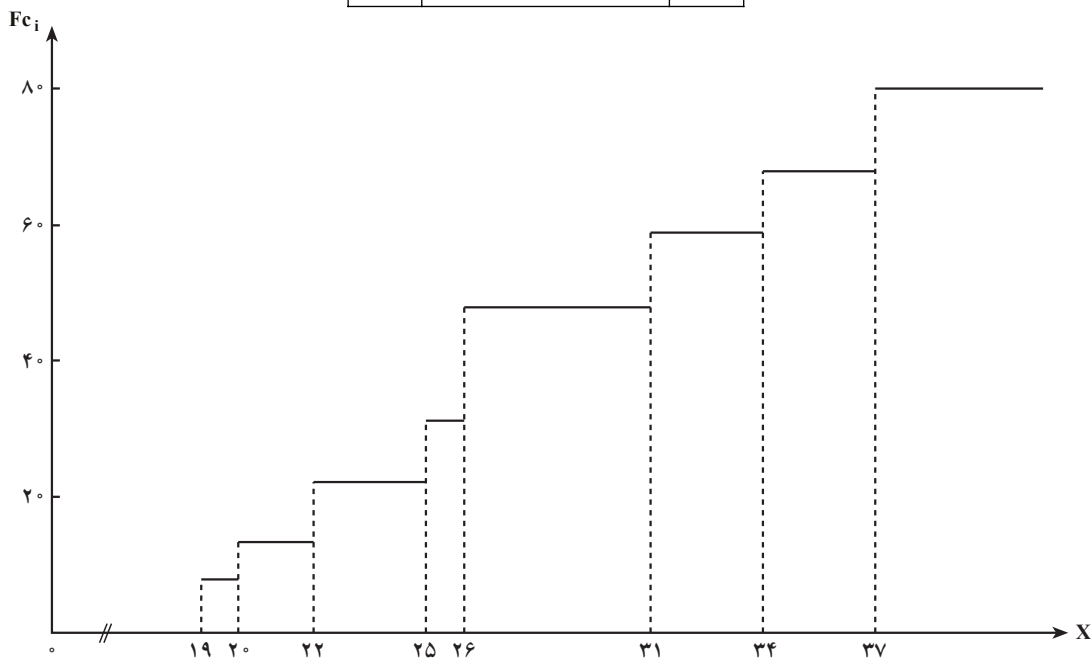
در مطالعه جامعه های با حجم کوچک، پردازش داده ها به صورت دستی انجام می گیرد. ولی در مطالعه جامعه های با حجم بزرگ، از پردازش ماشینی یا رایانه استفاده می شود. ما در اینجا به طور خلاصه فقط درباره پردازش دستی در مرحله گروه بندی بحث می کنیم.

۱- Cumulat

— روش چوب خط: در روش چوب خط، جدولها به صورت زیر تشکیل می گردد:

جدول ۱۰

X	ستون چوب خط	F_i
x_1	//	F_1
x_2	///	F_2
x_k	/	F_k
—	—	N



شکل ۶— نمودار فراوانیهای انباشته سن کارگران کارخانه نساجی

در ستون اول جدول، مقادیر صفت متغیر را به طور صعودی از کوچک به بزرگ قرار می دهیم. در ستون دوم، برای هر یک از مشاهدات در گروه X مربوط، یک چوب خط رسم می کنیم. این کار تا انتقال همه مشاهدات به صورت چوب خط ادامه می یابد. سپس تعداد چوب خطها را در هر گروه، شمارش کرده، نتیجه را به صورت فراوانی آن گروه در ستون سوم قرار می دهیم. در نهایت جمع فراوانیها برابر کل مشاهدات یعنی N خواهد شد.

اگر بخواهیم جدول فراوانیهای فاصله‌ای^۱ داشته باشیم، نحوه عملیات، مشابه جدول ۱۰ خواهد بود با این تفاوت که در ستون اول، مقادیر صفت متغیر را به صورت فاصله‌ها قرار می‌دهیم و سپس تک تک مشاهدات را در برابر فاصله مربوط با چوب خط رسم نموده، از شمارش آنها، فراوانیهای مطلق هر گروه را به دست خواهیم آورد که در ستون سوم نوشته می‌شود. به این ترتیب، جدول توزیع فراوانیهای فاصله‌ای برای صفت متغیر به دست می‌آید (جدول ۱۱):

جدول ۱۱

فاصله‌های صفت متغیر	ستون چوب خط	فراوانیها
$X_1 - X_2$	///	F_1
$X_2 - X_3$	////-	F_2
$X_3 - X_4$	//////	F_3
$X_k - X_{k+1}$	//	F_k
-	-	N

— مراحل تشکیل جدول توزیع فراوانیهای فاصله‌ای: مراحل تنظیم جدول توزیع فراوانیهای فاصله‌ای را به صورت زیر بیان می‌کنیم:

۱- پس از انجام مرحله مشاهده بزرگترین اندازه صفت متغیر (X_{\max}) و کوچکترین اندازه آن را (X_{\min}) مشخص کرده، تفاضل آنها را محاسبه می‌کنیم. یعنی

$$X_{\max} - X_{\min} = R \quad (5)$$

این تفاضل را که دامنه تغییرات^۲ نامیده می‌شود با R نشان می‌دهند.

۲- تعیین تعداد طبقات جدول - با توجه به حجم داده‌ها (N) تعداد طبقات را به دلخواه بین ۵ تا ۲۵ طبقه در نظر می‌گیرند. یکی از روشهای کلاسیک برای تعیین تعداد طبقات استفاده از فرمول استروجنس است که به صورت زیر می‌باشد:

۱- Interval Frequency

۲- Range = دامنه

$$K = 1 + 3 / 32 \log N \quad (6)$$

۳- از تقسیم دامنه تغییرات بر تعداد طبقات، فاصله طبقات به دست می آید که آن را با نماد I نشان خواهیم داد.

$$I = \frac{R}{K} \quad (7)$$

۴- تعیین فراوانیهای مطلق - در این مرحله با کمک روش چوب خط، تعداد اعضای را که اندازه آنها مربوط به هر طبقه جدول است، مشخص می کنیم.

با یک مثال، هم طریقه ساختن جدول توزیع فراوانیهای فاصله ای و هم طریقه عملی گروه بندی نتایج مشاهدات را نشان می دهیم:

مثال ۳: تعداد کارگران یک مؤسسه تولیدی $N = 100$ نفر می باشند. به منظور مطالعه تغییر پذیری سن آنها، تمامی کارگران را مورد مشاهده قرار داده ایم. نتایج مشاهدات برای سن آنها، به قرار زیر به دست آمده است:

۱۸	۲۱	۲۱	۱۸	۴۷	۵۳	۳۹	۲۸	۲۳	۵۰
۳۲	۳۳	۳۶	۴۴	۳۳	۲۲	۲۵	۲۲	۱۸	۱۹
۳۷	۲۴	۳۵	۳۵	۱۸	۳۹	۳۷	۳۲	۳۰	۴۴
۳۵	۳۵	۲۴	۲۷	۲۶	۳۲	۳۶	۳۵	۲۸	۳۲
۳۰	۴۶	۴۸	۴۴	۴۹	۵۰	۱۹	۲۰	۱۹	۴۹
۴۷	۴۳	۳۰	۳۸	۳۷	۲۰	۱۸	۳۰	۳۹	۴۵
۴۴	۲۴	۵۳	۲۸	۵۰	۲۸	۲۸	۵۲	۲۶	۱۸
۲۶	۲۷	۳۶	۳۸	۳۸	۴۰	۴۱	۴۱	۴۸	۴۳
۴۳	۳۲	۳۳	۳۲	۴۷	۲۵	۴۳	۳۲	۳۰	۳۵
۴۷	۳۴	۳۶	۳۸	۴۰	۴۱	۲۹	۴۲	۵۰	۵۴

اگر نتایج مشاهدات را بر حسب مقادیر مختلف سن طبقه بندی نمایم، ۳۶ گروه سنی تشکیل خواهد شد، که به علت زیادی گروههای صفت X ، امکان نتیجه گیری نسبت به تغییر پذیری سن مشکل است. از این رو، در اینجا مناسب است از روش گروه بندی فاصله ای استفاده کنیم، یعنی با ادغام کردن

چند گروه سنی باهم، فاصله‌های سنی بسازیم. با این کار، تعداد گروه‌های صفت متغیر کمتر و از طولانی شدن جدول کاسته می‌شود.

برای این منظور، ابتدا دامنه تغییرات صفت را محاسبه می‌کنیم.

در اینجا بزرگترین اندازه صفت (سن) ۵۴ سالگی و کوچکترین اندازه آن، ۱۸ سالگی است،

بنابراین داریم :

$$54 - 18 = 36 \quad \text{دامنه تغییرات سن}$$

اگر بخواهیم نتایج مشاهدات را در ۸ گروه طبقه‌بندی کنیم ($k=8$)، طول فاصله خواهد شد :

$$\frac{36}{8} = 4.5 \approx 5 \quad \text{طول فاصله}$$

(در چنین مواردی بهتر است طول فاصله را به عدد صحیح تبدیل کنیم).

بنابراین طول فاصله‌هایی را که در جدول توزیع فراوانی تشکیل می‌دهیم، برابر ۵ سال در نظر

می‌گیریم. برای تشکیل اولین طبقه در جدول توزیع فراوانی، کوچکترین اندازه صفت را (X_{\min})

برابر کرانه پایین طبقه اول در نظر گرفته، با توجه به مقدار I (فاصله طبقات)، سایر طبقات جدول را

تنظیم می‌کنیم. توضیح اینکه کوچکترین اندازه هر طبقه را، کرانه پایین آن طبقه و بزرگترین اندازه آن

طبقه را، کرانه بالای آن طبقه نامند.

نتایج این گروه‌بندی در جدول زیر نشان داده شده است (جدول ۱۲) :

جدول ۱۲

طبقات سنی X	چوب خط	فراوانی F_i
۱۸-۲۲	//// //	۱۵
۲۳-۲۷	//// //	۱۱
۲۸-۳۲	//// //	۱۸
۳۳-۳۷	//// //	۱۷
۳۸-۴۲	//// //	۱۳
۴۳-۴۷	//// //	۱۴
۴۸-۵۲	////	۹
۵۳-۵۷	///	۳
-	-	۱۰۰

باید توجه داشت که هریک از فاصله‌ها از ۵ سال تشکیل شده‌اند. چون سن، یک صفت ناپیوسته است، بنابراین کرانه بالای هر گروه با کرانه پایین گروه بعدی یک واحد تفاوت دارد، به عبارت دیگر، سن ۲۲ سالگی در گروه اول قرار می‌گیرد و سن ۲۳ سالگی که بلافاصله بعد از ۲۲ سالگی قرار دارد، در گروه دوم منظور می‌شود. به این ترتیب، با قرائت تک‌تک اعداد که در مرحله مشاهده به دست آمده است و انتقال آنها به فاصله مربوط در جدول، به وسیله چوب‌خط، کلیه اعداد را در ستون دوم وارد می‌کنیم و از شمارش آنها، فراوانیهای هر گروه (هر فاصله) را به دست می‌آوریم که حاصل جمع آنها نیز برابر N یعنی ۱۰۰ خواهد بود.

اهمیت فراوانیهای فاصله‌ای و مناسب بودن آن، زمانی آشکار می‌گردد که صفت متغیر مورد مطالعه «پیوسته» باشد.

در مطالعه صفت متغیر پیوسته، فاصله‌ها را باید چنان انتخاب کرد که صفت در هریک از آنها تقریباً به طور یکنواخت توزیع شده باشد. معمولاً انجام این کار وقتی امکان‌پذیر است که فاصله‌های نامساوی انتخاب شوند. لیکن انتخاب فاصله‌های نامساوی، انجام عملیات بعدی را مشکل‌تر می‌سازد. از این رو، در عمل اکثر اوقات، نتایج مشاهدات را به وسیله توزیع فراوانیهای با فاصله‌های یکسان بیان می‌کنند.

در استفاده از فاصله‌های با طولهای متفاوت، تحلیل توزیع فراوانیهای فاصله‌ای مشکل‌تر می‌گردد، زیرا فراوانیهای گروهها نه تنها از خصوصیات صفت متغیر تبعیت می‌کند، بلکه از طول فاصله انتخاب شده نیز تأثیر خواهد پذیرفت.

در اینجا به این نتیجه می‌رسیم که در مطالعه صفت متغیر که به صورت توزیع فاصله‌ای بیان شده باشد، باید از مفهوم جدیدی به نام «چگالی فراوانی» استفاده کنیم.

چگالی فراوانی، فراوانی را در واحد اندازه صفت نشان می‌دهد.

یعنی:

$$dF_i = \frac{\text{فراوانی گروه } i}{\text{فاصله همان گروه}} = \frac{F_i}{\Delta x_i} \quad (8)$$

$$(9) \quad df_i = \frac{\text{فراوانی نسبی گروه } i}{\text{فاصله همان گروه}} = \frac{f_i}{\Delta x_i}$$

مثال ۴: فرض کنیم نتایج مشاهدات برای صفت متغیر، موجودی حسابهای پس انداز در یکی از شعب بانک باشد که برحسب فاصله‌ها بیان شده است (جدول ۱۳):

جدول ۱۳

(موجودی برحسب هزار تومان) X	۷-۱۰	۱۰-۱۵	۱۵-۲۵	۲۵-۴۰	۴۰-۶۰	-
F _i فراوانیها	۳۰	۴۰	۶۰	۶۰	۱۰	۲۰۰

اگرچه در توزیع فوق با فاصله ۷-۱۰، (به طول ۳ واحد)، فراوانی ۳۰ و با فاصله ۱۵-۱۰ (به طول ۵ واحد) فراوانی ۴۰ در تناظر است، اما به این معنا نیست که با تغییر صفت از فاصله ۷-۱۰ به فاصله ۱۵-۱۰ فراوانیها افزایش یافته، بلکه برعکس در فاصله اول به سهم هر واحد طول فاصله $\frac{۳۰}{۳} = ۱۰$ عنصر جامعه تعلق می‌گیرد، در صورتی که در فاصله دوم، به هر واحد طول فاصله فقط $\frac{۴۰}{۵} = ۸$ عنصر جامعه تعلق می‌گیرد.

چگالی فراوانیها را برای فراوانیهای فاصله‌ای محاسبه کرده، توزیع چگالی فراوانیها را به دست می‌آوریم.

جدول ۱۴

فاصله‌های صفت	۷-۱۰	۱۰-۱۵	۱۵-۲۵	۲۵-۴۰	۴۰-۶۰
چگالی فراوانی $df_i = \frac{F_i}{\Delta x_i}$	$\frac{۳۰}{۳} = ۱۰$	$\frac{۴۰}{۵} = ۸$	$\frac{۶۰}{۱۰} = ۶$	$\frac{۶۰}{۱۵} = ۴$	$\frac{۱۰}{۲۰} = ۰/۵$

ملاحظه می‌شود که چگالی فراوانیها با تغییر صفت متغیر از فاصله‌ای به فاصله دیگر تغییر کرده، نزول می‌یابد. این نزولی بودن تصادفی نیست. برای اطمینان، کافی است فاصله‌ها را عوض کنیم، باز هم خواهیم دید که تغییرات چگالی فراوانیها برای این صفت متغیر، نزولی است. اگر در توزیع چگالی فراوانیها، فاصله‌ها را تغییر دهیم، شکل کلی توزیع تغییر نخواهد کرد. بدین جهت، در مطالعه توزیع فراوانیهای فاصله‌ای، با فاصله‌های نامساوی، مناسب است، این توزیع فراوانیها، به توزیع چگالی فراوانیها تبدیل گردد.

مثال ۵: فرض کنیم نتایج مشاهدات برای یک صفت متغیر کمی پیوسته X ، به وسیله جدول توزیع فراوانیهای نسبی فاصله‌ای بیان شده باشد:

جدول ۱۵

فاصله‌های صفت X	۱۱-۱۵	۱۵-۲۵	۲۵-۲۸	۲۸-۳۵	-
فراوانیهای نسبی f_i	۰/۰۸	۰/۴۰	۰/۲۴	۰/۲۸	$f_i = 1$

ظاهراً در این توزیع، بین تغییرات فراوانیها با فاصله‌های صفت متغیر، نظم خاصی دیده نمی‌شود، و بزرگترین فراوانی نسبی متعلق به فاصله ۱۵-۲۵ است. حال اگر توزیع فراوانیها را به صورت توزیع چگالی فراوانیها، درآوریم، نتیجه به شکل جدول زیر به دست می‌آید:

جدول ۱۶

فاصله‌های صفت X	۱۱-۱۵	۱۵-۲۵	۲۵-۲۸	۲۸-۳۵	-
فراوانیهای نسبی f_i	۰/۰۸	۰/۴۰	۰/۲۴	۰/۲۸	$f_i = 1$
چگالی فراوانیهای نسبی df_i	۰/۰۲	۰/۰۴	۰/۰۸	۰/۰۴	-

از توزیع چگالی فراوانیها نتیجه می‌گیریم که حداکثر آن در فاصله ۱۵-۲۵ نیست، بلکه به فاصله ۲۵-۲۸ تعلق دارد. یعنی اکثر عناصر جامعه در این فاصله گروه‌بندی شده‌اند. بنابراین، هرگاه جدول توزیع فراوانی فاصله‌ای، با فاصله‌های نامساوی داشته باشیم، برای درک این مطلب که تراکم فراوانیها در چه فاصله‌ای از صفت متغیر قرار دارند، باید، ابتدا چگالی فراوانیهای صفت متغیر را محاسبه نموده، با استفاده از آنها، نسبت به تراکم فراوانیها نتیجه‌گیری کنیم.

بیان هندسی توزیع فراوانیهای فاصله‌ای (صفت متغیر پیوسته)

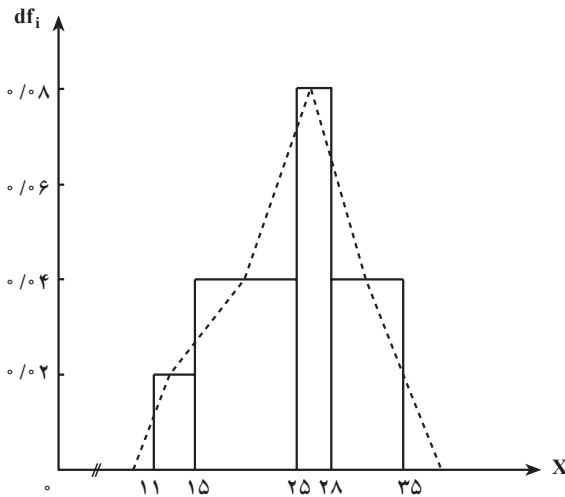
نمودار مستطیلی یا هیستوگرام^۱ (بافت‌نگار): برای بیان هندسی توزیع صفت پیوسته که با فاصله بیان شده است، از دو نوع نمودار استفاده می‌شود:

– نمودار مستطیلی یا هیستوگرام برای توزیع چگالی فراوانیها

– نمودار تجمعی یا اجایو برای توزیع فراوانیهای انباشته

در ساختن هر دو نوع نمودار، فرض بر این است که چگالی (تراکم) در هر یک از فاصله‌ها ثابت باشد.

برای ساختن هیستوگرام، بر روی محور طولها (در دستگاه مختصات قائم) فاصله‌های صفت و بر روی محور عرضها، چگالی فراوانیها را قرار می‌دهند. به عبارت دیگر بر روی فاصله‌های صفت متغیر، مستطیلهایی به ارتفاع مساوی با چگالی متناظر با آن فاصله‌ها می‌سازند. برای جدول ۱۶ در مثال ۵، هیستوگرام توزیع صفت متغیر X را رسم می‌کنیم.



شکل ۷- هیستوگرام جدول ۱۶

برای بیان هندسی توزیع صفتی که با فاصله‌ها بیان شده است، می‌توان از نمودار چندگوش نیز استفاده کرد. برای رسم چندگوش توزیع صفت، کافی است روی شکل هیستوگرام، نقاط وسط اضلاع فوقانی مستطیلهای را با خطوط مستقیم به هم وصل کنیم. از اتصال تمام نقاط وسط مستطیلهای به هم، نمودار چندگوش به دست می‌آید که مساحت زیر آن با محور X ها، برابر با مجموع مساحت مستطیلهای، می‌باشد.

در شکل ۷ چندگوش توزیع صفت به وسیله خط چین رسم شده است.

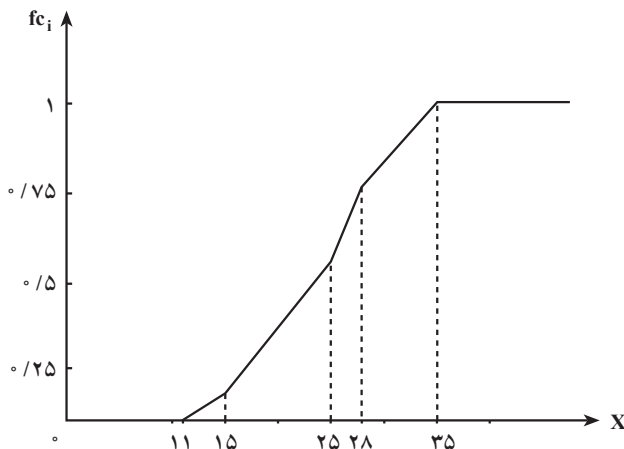
در ساختن نمودار تجمعی یا اجابو با فراوانیهای انباشته، از همان فرض ثابت بودن چگالی در فاصله‌های صفت استفاده می‌شود. در صفحه مختصات قائم، روی محور طولها، کرانه‌های هر فاصله را قرار داده، روی محور عرضها، فراوانیهای انباشته را درجه بندی می‌کنیم. به همان ترتیبی که برای صفت کمی گسسته، چندگوش فراوانیهای انباشته رسم گردید، عمل می‌کنیم و از به دست آوردن نقاط

(باطول X_i و عرض FC_i متناظر)، آنها را به وسیله خطوط مستقیم، به هم وصل می‌کنیم. در این صورت، شکلی ظاهر می‌شود که نمودار فراوانیهای انباشته نام دارد.

در زیر، جدول توزیع فراوانیهای انباشته نسبی را برای مثال ۵ تشکیل می‌دهیم و با استفاده از آن نمودار تجمعی توزیع صفت را رسم می‌کنیم (جدول ۱۷ و شکل ۸):

جدول ۱۷

X	۱۱-۱۵	۱۵-۲۵	۲۵-۲۸	۲۸-۳۵	-
f_i	۰/۰۸	۰/۴	۰/۲۴	۰/۲۸	۱
fc_i	۰/۰۸	۰/۴۸	۰/۷۲	۱	-



شکل ۸ - نمودار تجمعی (جدول ۱۷) توزیع صفت متغیر X

می‌توان نشان داد که مساحت هیستوگرام (جمع مساحت مستطیلهای) و همین‌طور مساحت زیرخط شکسته چندگوش، وقتی که برای رسم آنها از چگالی فراوانیهای مطلق استفاده شده باشد، همواره مساوی با N است و وقتی که برای رسم آنها از چگالی فراوانیهای نسبی استفاده شده باشد، مساحت هیستوگرام و همین‌طور مساحت زیرخط شکسته چندگوش، مساوی با ۱ است.

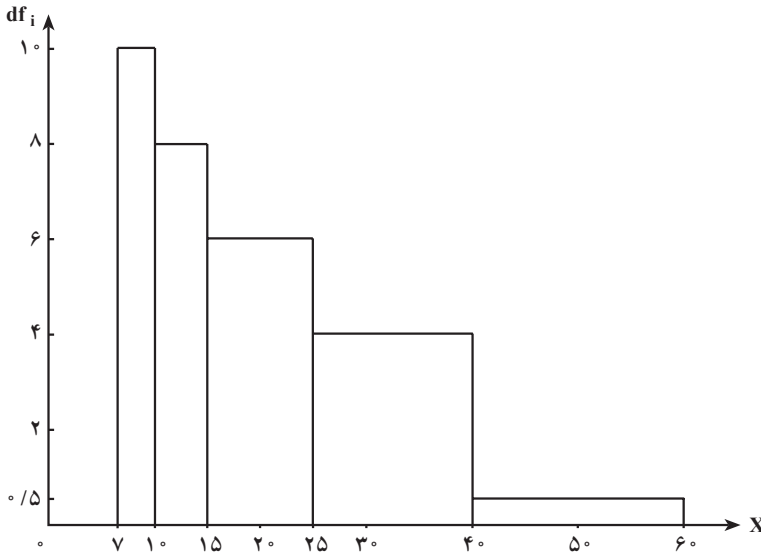
نمودارهای هیستوگرام و تجمعی توزیع صفت را برای مثال ۴ رسم می‌کنیم تا دانش‌آموزان با چگونگی رسم آنها آشنایی بیشتری پیدا کنند.

ابتدا، چگالی فراوانیهای مطلق و فراوانیهای انباشته توزیع صفت را محاسبه می‌کنیم.

جدول ۱۸

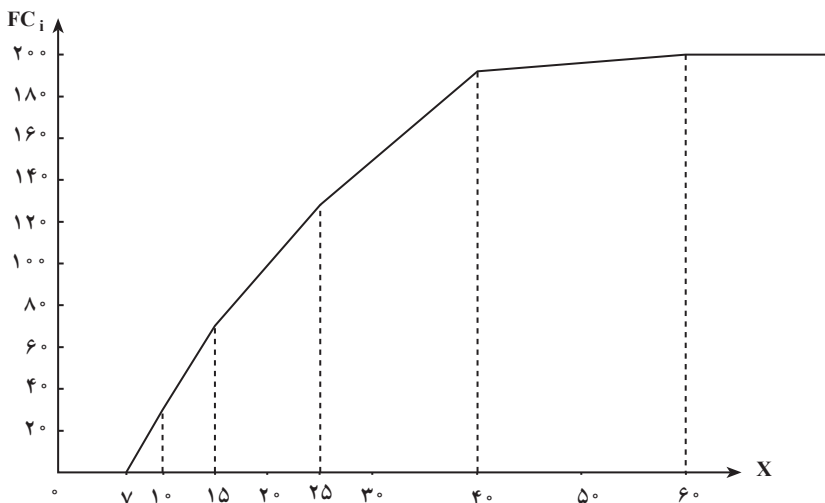
X	۷-۱۰	۱۰-۱۵	۱۵-۲۵	۲۵-۴۰	۴۰-۶۰	-
F_i	۳۰	۴۰	۶۰	۶۰	۱۰	۲۰۰
dF_i	۱۰	۸	۶	۴	۰/۵	-
FC_i	۳۰	۷۰	۱۳۰	۱۹۰	۲۰۰	-

برای رسم هیستوگرام، ابتدا، فاصله‌های صفت را روی محور طولها، در دستگاه مختصات قائم وارد کرده، روی محور عرضها، چگالی فراوانیها را قرار می‌دهیم، سپس روی هر فاصله صفت متغیر، مستطیلی با طول چگالی فراوانی همان فاصله رسم می‌کنیم.



شکل ۹ - هیستوگرام توزیع صفت متغیر X (جدول ۱۸)

برای رسم نمودار تجمعی، اندازه‌های صفت را روی محور طولها و فراوانیهای انباشته را روی محور عرضها در دستگاه مختصات قرار می‌دهیم، سپس با هر مقدار صفت و فراوانی انباشته متناظر آن، یک نقطه روی صفحه مختصات به دست می‌آوریم. با وصل نمودن نقاط به دست آمده، به وسیله خطوط مستقیم، شکل نمودار تجمعی توزیع صفت متغیر به دست می‌آید.



شکل ۱۰- نمودار تجمعی توزیع صفت متغیر X (مربوط به مثال ۴)

سؤالها و تمرینها

- ۱- لزوم گروه‌بندی نتایج مشاهدات در چیست؟
- ۲- فراوانیهای گروه، چه مفهومی دارد؟
- ۳- توزیع صفت متغیر، چیست؟
- ۴- فراوانی مطلق، چه مفهومی دارد؟
- ۵- فراوانی نسبی، چه مفهومی دارد؟
- ۶- بین فراوانیهای مطلق چه رابطه‌ای برقرار است؟
- ۷- بین فراوانیهای نسبی چه رابطه‌ای وجود دارد؟
- ۸- توزیع فراوانیهای صفت متغیر چیست؟ تعریف کنید.
- ۹- فرض کنیم صفت متغیر کمی ناپیوسته در جامعه فقط مقادیر ۵، ۱۰، ۱۵، ۲۰ و ۲۵ را به ترتیب با فراوانیهای ۵، ۱۰۰، ۲۰۰، ۱۰۰ و ۵۰ اختیار کرده باشد.
اولاً: توزیع فراوانیهای صفت متغیر فوق را بیان کنید.
ثانیاً: جدول توزیع فراوانیهای فوق را تشکیل دهید.
- ۱۰- توزیع فراوانیهای نسبی صفت متغیر ناپیوسته، چگونه به دست می‌آید؟
- ۱۱- برای صفت متغیر در تمرین ۹، فراوانیهای نسبی را محاسبه کنید.

- ۱۲- بین فراوانیهای مطلق و فراوانیهای نسبی چه رابطه‌ای برقرار است؟
- ۱۳- دلیل استفاده از نمودارها برای بیان توزیع صفت متغیر چیست؟
- ۱۴- نمودار میله‌ای، چگونه ساخته می‌شود؟
- ۱۵- برای توزیع صفت متغیر در تمرین ۹، نمودار میله‌ای رسم کنید.
- ۱۶- چندگوش توزیع فراوانیها چیست و چگونه ساخته می‌شود؟
- ۱۷- چندگوش توزیع فراوانیها را برای توزیع صفت متغیر که در جدول زیر داده شده است، رسم کنید.

X	۲	۴	۸	۱۰	۱۲	۱۴	-
F _i	۱۲	۲۴	۳۶	۴۸	۲۶	۴	۱۵۰

۱۸- توزیع صفت متغیر X با جدول زیر بیان شده است :

X	۳	۶	۹	۱۲	۱۵	-
f _i	۰/۱	۰/۳	۰/۳	۰/۲	۰/۱	۱

نمودار میله‌ای و نمودار چندگوش را برای توزیع فوق رسم کنید.

۱۹- فراوانی انباشته (تجمعی) چیست؟ تعریف کنید.

۲۰- توزیع فراوانیهای انباشته چیست؟

۲۱- توزیع فراوانی صفت متغیر، با جدول زیر بیان شده است :

X	۲	۴	۶	۸	۱۰	۱۲	-
F _i	۲۰	۴۰	۷۰	۳۵	۲۰	۱۵	۲۰۰

جدول توزیع فراوانیهای انباشته را برای صفت متغیر تشکیل دهید.

۲۲- فراوانی انباشته نسبی چیست و چگونه محاسبه می‌شود؟

۲۳- جدول توزیع فراوانیهای انباشته نسبی را برای تمرین ۲۱ تشکیل دهید.

۲۴- توزیع فراوانیهای انباشته به صورت هندسی چگونه بیان می‌شود؟

۲۵- نمودار تجمعی یا اجایو، برای صفت متغیر چگونه رسم می‌شود؟

۲۶- نمودار تجمعی را برای توزیع صفت در تمرین ۲۱ رسم کنید.

۲۷- چرا در گروه بندی نتایج مشاهدات صفت متغیر، از فاصله مقادیر صفت استفاده می شود؟

۲۸- لزوم تشکیل چنین فاصله هایی در چیست؟

۲۹- توزیع فراوانیهای فاصله ای، چیست؟

۳۰- وزن دانش آموزان کلاس چهارم دبستان A به شرح زیر اندازه گیری شده است.

۲۸/۵	۳۲/۵	۳۱/۷	۳۱/۸	۲۷/۹
۳۱/۴	۳۰/۸	۳۲/۵	۲۹/۶	۳۱/۷
۳۰/۶	۳۰/۴	۲۹/۳	۲۸/۷	۲۸/۹
۲۹/۳	۲۹/۷	۲۸/۸	۲۸/۶	۳۰/۵
۳۲/۸	۳۳/۲	۳۳/۵	۳۱/۴	۳۱/۶
۳۰/۲	۲۷/۸	۲۸/۲	۲۹/۳	۲۸/۲
۲۹/۸	۳۰/۸	۳۱/۲	۳۲/۸	۳۳/۴
۳۰/۲	۲۸/۶	۲۹/۹	۲۹/۳	۳۱/۸
۳۴/۱	۳۴/۳	۳۲/۴	۳۱/۲	۳۴/۴
۲۸/۷	۲۹/۴	۳۱/۴	۲۷/۸	۲۸/۲

الف - جدول توزیع فراوانیهای گروه بندی شده وزن دانش آموزان را در ۸ گروه تشکیل دهید.

ب - جدول توزیع فراوانیهای تجمعی وزن دانش آموزان را تشکیل دهید.

ج - چند درصد دانش آموزان وزنی بیش از ۳۰ کیلوگرم دارند؟

د - هیستوگرام وزن دانش آموزان را رسم کنید.

۳۱- در چه مواردی انتخاب فاصله های با طول نامساوی را به فاصله های با طول یکسان

ترجیح می دهید؟

۳۲- چگالی فراوانیها چیست و لزوم وارد کردن آن در چیست؟

۳۳- توزیع فراوانیهای کارگران بر حسب دستمزد روزانه در یک کارگاه ساختمانی به صورت

جدول زیر بیان شده است :

X دستمزد روزانه (صد تومان)	۱۶. ۲۵	۲۵. ۳۰	۳۰. ۳۵	۳۵. ۴۰	۴۰. ۵۰	.
F _i فراوانی	۲۲	۹۸	۶۰	۱۸	۲	۲۰۰

توزیع چگالی فراوانیها را برای دستمزد روزانه کارگران تشکیل دهید.

۳۴- دلیل ثبات چگالی فراوانیها در چیست؟

۳۵- هیستوگرام یا نمودار مستطیلی چیست؟ در چه مواردی از هیستوگرام استفاده می شود؟

- ۳۶- فرضی که در ساختن هیستوگرام باید در نظر گرفت، چیست؟
 ۳۷- برای تمرین ۳۳ هیستوگرام را رسم کنید.
 ۳۸- برای تمرین ۳۳ نمودار تجمعی یا اجایو را رسم کنید.

خودآزمونهای چهارگزینه‌ای ✓

- ۱- کدام رابطه بین فراوانیهای مطلق صحیح است؟
 الف - $F_i = 1$.
 ب - $F_i = \frac{N}{4}$.
 ج - $F_i = 0$.
 د - $F_i = N$.
- ۲- کدام رابطه بین فراوانیهای نسبی صحیح است؟
 الف - $f_i = 0$.
 ب - $f_i \neq 0/5$.
 ج - $f_i = N$.
 د - $f_i = 1$.
- ۳- فراوانی انباشته هر گروه از صفت متغیر، در جدول توزیع فراوانی برابر است با:
 الف - فراوانی مطلق گروه مربوط
 ب - فراوانی نسبی گروه مربوط
 ج - فراوانی گروه مربوط به علاوه فراوانیهای ماقبل آن
 د - N
- ۴- برای بیان هندسی صفت متغیر گسسته از کدام نمودارها، استفاده می‌شود؟
 الف - نمودار فراوانی تجمعی
 ب - نمودار میله‌ای و چندگوش
 ج - هیستوگرام و نمودار مستطیلی
 د - نمودار نقطه‌ای
- ۵- برای رسم هیستوگرام روی محورهای مختصات کدام مؤلفه‌ها قرار داده می‌شود؟
 الف - فاصله‌های صفت متغیر و چگالی فراوانی
 ب - اندازه صفت و فراوانی نسبی
 ج - فاصله‌های صفت و فراوانیهای انباشته
 د - فاصله‌های صفت و فراوانیهای مطلق
- ۶- چگالی فراوانی صفت متغیر یعنی:
 الف - فراوانی هر طبقه تقسیم بر N
 ب - فراوانی هر طبقه تقسیم بر فراوانی نسبی همان طبقه
 ج - فراوانی هر طبقه تقسیم بر فاصله آن طبقه
 د - فراوانی انباشته تقسیم بر N
- ۷- اگر هیستوگرام توزیع صفت با استفاده از چگالی فراوانیهای مطلق، رسم گردد، مساحت مستطیل‌های هیستوگرام برابر خواهد شد با:
 الف - ۱
 ب - $0/5$
 ج - N
 د - $\frac{N}{4}$

مشخص‌کننده‌های مرکزی (تمایل به مرکز)

- هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند:
- ۱- دلایل به کار بردن مشخصه‌های آماری را توضیح دهد.
 - ۲- کاربرد و دلایل استفاده از مشخص‌کننده‌های مرکزی را بیان کند.
 - ۳- مشخصه‌های مرکزی مهم در توزیع فراوانیهای صفت را به صورت جدول، محاسبه نموده، آنها را تعبیر نماید.
 - ۴- با استفاده از مشخصه‌های مرکزی، جامعه‌های مختلف را مقایسه کند.

مشخص‌کننده‌های آماری

در فصل گذشته درباره‌ی توزیع صفت متغیر در جامعه‌ها بحث شد و نحوه‌ی آرایه نتایج مشاهدات، هم با جدول و هم با نمودار، در مورد صفات کمی گسسته و پیوسته بیان گردید. ولی برای یک محقق، آگاهی از توزیع صفت در جامعه برحسب تمامی اندازه‌های صفت متغیر، چندان مفید نیست. اکثر اوقات داشتن اطلاعات اجمالی از خصوصیات صفت در جامعه، برای محقق مفیدتر خواهد بود.

کمیت‌هایی که خصوصیات جامعه‌ها را به طور اجمال بیان می‌کنند، «مشخصه‌های عددی» یا «مشخصه‌های آماری»^۱ توزیع صفت، نامیده می‌شوند.

بسته به اینکه، چه ویژگی‌هایی از توزیع صفت، مورد نظر باشد، گروه‌های مختلف مشخصه‌های آماری، تعریف می‌شوند و مورد استفاده قرار می‌گیرند.

در این فصل درباره‌ی مشخصه‌های مرکزی (مشخص‌کننده‌های مرکزی)، که نقطه‌ی تمرکز توزیع صفت را منعکس می‌کنند، بحث خواهد شد.

مشخصه‌های مرکزی، مشخصه‌هایی هستند که مرکزیت اندازه‌ی صفت متغیر را در جامعه نشان

^۱ Statistical Characteristic

می دهند. مانند: میانگین حسابی (متوسط حسابی)، میانه و نما که جزو این مشخصه‌ها، به حساب می آیند.
میانگین حسابی^۱

یکی از مهمترین مشخصه‌های مرکزی در آمار، میانگین حسابی یا متوسط حسابی است که به صورت زیر تعریف می شود:

میانگین حسابی N متغیر $X_1, X_2, X_3, \dots, X_N$ برابر است با:

حاصل جمع متغیرهای $X_1, X_2, X_3, \dots, X_N$ تقسیم بر تعدادشان، یعنی N که آن را با نماد \bar{X} (بخوانید ایکس بار) نشان می دهند. (توضیح اینکه غالباً میانگین کل جامعه را با علامت 1 و میانگین مقادیر نمونه را با \bar{X} نشان می دهند. در این کتاب برای میانگین، به طور کلی از نماد \bar{X} استفاده شده است.)

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} = \frac{\sum X_i}{N} \quad (1)$$

مثال ۱: فرض کنید ۱۰ نفر دانش آموز را وزن نموده ایم و نتیجه به صورت اعداد زیر به دست آمده است:

۴۱، ۴۴/۲، ۴۲/۸، ۴۵، ۴۷، ۴۳/۵، ۴۵، ۴۳، ۴۲/۵، ۴۳

می خواهیم متوسط وزن این ۱۰ دانش آموز را محاسبه کنیم.

اعداد بالا را باهم جمع کرده، نتیجه را بر ۱۰ (تعداد آنها) تقسیم می کنیم تا متوسط وزن آنها به دست آید:

$$\bar{X} = \frac{۴۳ + ۴۲/۵ + ۴۳ + ۴۵ + ۴۳/۵ + ۴۷ + ۴۵ + ۴۲/۸ + ۴۴/۲ + ۴۱}{۱۰} = \frac{۴۳۷}{۱۰} = ۴۳/۷$$

بنابراین متوسط وزن دانش آموزان، ۴۳/۷ کیلوگرم می باشد.

در صورتی که حجم جامعه زیاد باشد و نتایج مشاهدات صفت متغیر را در جدول توزیع فراوانی بیان کرده باشیم، چون از هر کدام از مقادیر صفت، ممکن است بیش از یک فراوانی داشته باشیم، ابتدا مقادیر صفت را در فراوانی متناظرشان ضرب کرده، نتیجه را در یک ستون در ادامه جدول می نویسیم. سپس مقادیر این ستون را باهم جمع کرده، حاصل را بر N تقسیم می کنیم. بدین ترتیب، میانگین صفت به دست می آید، (جدول ۱)، در نتیجه فرمول محاسبه میانگین به صورت زیر در می آید:

$$\bar{X} = \frac{F_1 X_1 + F_2 X_2 + \dots + F_k X_k}{\sum F_i} = \frac{\sum F_i X_i}{N} \quad (2)$$

^۱ - Arithmetic Mean

جدول ۱

X_i	F_i	$F_i X_i$
X_1	F_1	$F_1 X_1$
X_2	F_2	$F_2 X_2$
X_3	F_3	$F_3 X_3$
X_k	F_k	$F_k X_k$
-	N	$\sum F_i X_i$

تمامی محاسبات در جدول ۱ نشان داده شده است.

مثال ۲: توزیع سنی ۵۰ نفر در جدول زیر بیان شده است (جدول ۲). مطلوب است میانگین سن

برای جامعه فوق:

جدول ۲

سن X	۲۰	۲۲	۲۴	۲۶	۲۸	-
فراوانی F_i	۴	۶	۳۰	۷	۳	$\sum F_i = 50$

برای محاسبه میانگین، جدول زیر را تشکیل می‌دهیم (جدول ۳):

جدول ۳

X_i	F_i	$F_i X_i$
۲۰	۴	۸۰
۲۲	۶	۱۳۲
۲۴	۳۰	۷۲۰
۲۶	۷	۱۸۲
۲۸	۳	۸۴
-	۵۰	۱۱۹۸

$$\bar{X} = \frac{\cdot F_i X_i}{N} = \frac{1198}{50} = 23/96$$

بنابراین متوسط سن برای هریک از اعضای این گروه، ۲۳/۹۶ سال می‌باشد. گفتنی است که هرچند سن، یک صفت کمی گسسته است^۱، ولی متوسط آن می‌تواند یک عدد غیر صحیح باشد، بنابراین نباید هیچ‌گونه تغییری در آن داده شود.

همانطور که از فرمول میانگین حسابی، دانسته می‌شود، فراوانیهای صفت در آن، مانند وزنه‌هایی هستند که به مقادیر صفت اثر می‌کنند، بدین معنی که اگر فراوانی یک مقدار صفت را بزرگ کنیم، متوسط حسابی، به آن مقدار، نزدیک خواهد شد. از این رو، میانگین حسابی که طبق فرمول ۲ محاسبه می‌شود، «میانگین حسابی وزنی» نامیده می‌شود.

اگر توزیع صفت متغیر، با جدول توزیع فراوانیهای نسبی بیان شده باشد، میانگین حسابی از رابطه زیر محاسبه می‌شود:

$$\bar{X} = \sum_{i=1}^k f_i X_i \quad (3)$$

خودآزمایی: نشان دهید که دو فرمول زیر معادل هستند:

$$\frac{\cdot F_i X_i}{N} = \cdot f_i X_i$$

نحوه محاسبه میانگین با فراوانی نسبی در جدول ۴ نشان داده شده است:

جدول ۴

X_i	f_i	$f_i X_i$
x_1	f_1	$f_1 x_1$
x_2	f_2	$f_2 x_2$
x_k	f_k	$f_k x_k$
—	۱	$\cdot f_i X_i$

مثال ۳: مجموعه‌ای از ۲۰۰ سند حسابداری مربوط به خریدهای کوچک، بر حسب تعداد

۱- سن در تعریف آمار دموگرافیک، عبارت از سالهای تمام شده عمر انسان؛ بنابراین فقط می‌تواند اعداد صحیح را اختیار کند و یک صفت کمی گسسته است.

نقصها در این اسناد، گروه‌بندی و نتایج مشاهدات به وسیلهٔ جدول توزیع فراوانیهای نسبی زیر، بیان شده است:

X تعداد نقص	۰	۱	۲	۳	۴	–
f _i فراوانی نسبی	۰/۵	۰/۲	۰/۲	۰/۰۵	۰/۰۵	۱

مطلوب است: متوسط تعداد نقصها، برای یک سند.

برای محاسبهٔ متوسط تعداد نقصها، جدول زیر را تشکیل می‌دهیم (جدول ۵):

جدول ۵

X	f _i	f _i X _i
۰	۰/۵	۰
۱	۰/۲	۰/۲
۲	۰/۲	۰/۴
۳	۰/۰۵	۰/۱۵
۴	۰/۰۵	۰/۲۰
–	۱	۰/۹۵

$$\bar{X} = \sum f_i X_i = 0.95$$

بنابراین متوسط تعداد نقصها برای یک سند، ۰/۹۵ می‌باشد. یعنی به‌طور متوسط در هر سند

۰/۹۵ نقص یا تقریباً یک نقص وجود دارد.

گاهی ممکن است حجم نتایج مشاهدات کوچک باشد. در چنین شرایطی، گروه‌بندی برای

صفت انجام نمی‌شود و در این صورت میانگین حسابی به‌صورت ساده بیان می‌گردد و همانطور که قبلاً

گفته شد، آن را با استفاده از فرمول ۱ محاسبه می‌کنند.

مثال ۴: طول سابقهٔ خدمت ۱۰ نفر مراجعه‌کننده برای اشتغال در یک مؤسسه (برحسب سال)

به‌قرار زیر است:

۸, ۴, ۱۰, ۷, ۸, ۶, ۵, ۱۰, ۷, ۱۱

متوسط حسابی سابقهٔ خدمت را برای یک مراجعه‌کننده، محاسبه کنید.

در این حالت به‌علت اینکه نتایج مشاهدات، گروه‌بندی نشده است، میانگین را طبق فرمول ۱

محاسبه می کنیم :

$$\bar{X} = \frac{8+4+10+7+8+6+5+10+7+11}{10} = \frac{76}{10} = 7.6$$

بنابراین، متوسط سابقه خدمت برای یک مراجعه کننده ۷/۶ سال می باشد.
در بسیاری موارد، فراوانیها در توزیع صفت، برحسب فاصله های مقادیر صفت بیان می شوند.
در این حالت توزیع صفت را «توزیع فاصله ای» می نامند.
در محاسبه میانگین برای توزیعهای فاصله ای، ابتدا وسط فاصله ها را به دست آورده، فرض می کنند که این اعداد، خود مقادیر مشاهده شده صفت هستند. (یعنی صفت متغیر را به صورت گسسته در نظر می گیرند) و سپس میانگین این توزیع جدید را طبق فرمول ۲ یا ۳ محاسبه می کنند. روش محاسبه میانگین در چنین حالتی در جدول ۶ نشان داده شده است.

جدول ۶

فاصله گروه ها	F_i	X_i	$F_i X_i$
$X_1 - X_2$	F_1	X_2	$F_1 X_2$
$X_2 - X_3$	F_2	X_3	$F_2 X_3$
$X_k - X_{k+1}$	F_k	X_k	$F_k X_k$
-	N	-	$\sum F_i X_i$

که در اینجا X_i ها، نقاط میانی فاصله گروه هاند و از میانگین کرانه پایین و کرانه بالای هر گروه به دست می آیند، مثلاً X_2 از $X_1 + X_2$ به دست می آید.

مثال ۵: مزد و حقوق ماهانه کارکنان یک مؤسسه خدماتی به صورت جدول صفحه بعد گروه بندی

شده است (جدول ۷).

مطلوب است : متوسط مزد و حقوق ماهانه کارکنان این مؤسسه :

جدول ۷

فاصله طبقات X مزد و حقوق ماهانه بر حسب هزار تومان	فراوانی F_i
۴۰-۵۰	۴
۵۰-۶۰	۷
۶۰-۷۰	۱۲
۷۰-۸۰	۱۵
۸۰-۹۰	۸
۹۰-۱۰۰	۴
-	۵۰

برای محاسبه میانگین مزد و حقوق ماهانه، ابتدا وسط فاصله‌ها یعنی X_i هر طبقه را محاسبه کرده، در ستون سوم جدول ۸ می‌نویسیم و سپس محاسبات را طبق فرمول ۲ انجام می‌دهیم:

جدول ۸

X فاصله طبقات	F_i	X_i وسط فاصله‌ها	$F_i X_i$
۴۰-۵۰	۴	۴۵	۱۸۰
۵۰-۶۰	۷	۵۵	۳۸۵
۶۰-۷۰	۱۲	۶۵	۷۸۰
۷۰-۸۰	۱۵	۷۵	۱۱۲۵
۸۰-۹۰	۸	۸۵	۶۸۰
۹۰-۱۰۰	۴	۹۵	۳۸۰
-	۵۰	-	۳۵۳۰

$$\bar{X} = \frac{\sum F_i X_i}{N} = \frac{۳۵۳۰}{۵۰} = ۷۰/۶$$

بنابراین، متوسط مزد و حقوق ماهانه یک نفر در مؤسسه فوق، ۷۰۶۰۰ تومان است.

مثال ۶: روزهای غیبت کارگران در یک کارگاه ساختمانی به صورت جدول زیر است :

X روزهای غیبت	۰-۴	۵-۹	۱۰-۱۴	۱۵-۱۹	۲۰-۲۴	۲۵-۲۹	-
F _i فراوانی	۵	۸	۱۰	۶	۷	۴	۴۰

متوسط روزهای غیبت کارگران را محاسبه کنید.

برای محاسبه میانگین روزهای غیبت کارگران جدول زیر را تشکیل می دهیم (جدول ۹):

جدول ۹

X فاصله طبقات	F _i	X. وسط فاصله‌ها	F _i X _i
۰-۴	۵	۲	۱۰
۵-۹	۸	۷	۵۶
۱۰-۱۴	۱۰	۱۲	۱۲۰
۱۵-۱۹	۶	۱۷	۱۰۲
۲۰-۲۴	۷	۲۲	۱۵۴
۲۵-۲۹	۴	۲۷	۱۰۸
-	۴۰	-	۵۵۰

$$\bar{X} = \frac{\sum F_i X_i}{N} = \frac{550}{40} = 13.75 \text{ روز}$$

میانگین حسابی روزهای غیبت کارگران ۱۳/۷۵ روز می باشد.

— بعضی خواص ریاضی میانگین حسابی

۱- میانگین حسابی N کمیت ثابت برابر است با خود آن کمیت :

فرض کنید $X = a$ باشد، در این صورت

$$\bar{X} = \frac{\sum F_i a}{\sum F_i} = \frac{a \cdot \sum F_i}{\sum F_i} = a$$

۲- اگر فراوانیهای مقادیر صفت را چند برابر بزرگ یا کوچک کنیم، میانگین حسابی تغییری

نمی کند. برای مثال اگر تمام فراوانیها را K برابر کوچک کنیم، خواهیم داشت :

$$\bar{X} = \frac{\sum \frac{F_i}{k} X_i}{\sum \frac{F_i}{k}} = \frac{\sum F_i X_i}{\sum F_i}$$

۳- حاصل جمع انحرافهای (تفاضلهای) مقادیر صفت از میانگین خود، همواره مساوی صفر

است، یعنی :

$$\sum (X_i - \bar{X}) = 0 \Rightarrow \sum X_i - \sum \bar{X} = \sum X_i - N\bar{X} = \sum X_i - \sum X_i = 0$$

$$\sum F_i (X_i - \bar{X}) = 0 \quad \text{و به همین ترتیب}$$

۴- اگر از تمامی مقادیر صفت، یک عدد ثابت کم شود و یا یک عدد ثابت به آنها اضافه گردد، از میانگین این کمیتهای جدید نیز، همان مقدار ثابت کاسته، یا به آن افزوده خواهد شد.

$$\begin{aligned} (\overline{X+a}) &= \bar{X} + a \\ (\overline{X-a}) &= \bar{X} - a \end{aligned}$$

۵- اگر تمامی مقادیر صفت را، b برابر بزرگ یا کوچک کنیم، متوسط حسابی آنها نیز b برابر

بزرگ یا کوچک خواهد شد :

$$\begin{aligned} (\overline{bX}) &= b\bar{X} \\ \left(\frac{\bar{X}}{b}\right) &= \frac{1}{b}\bar{X} \end{aligned} \quad b \neq 0$$

محاسبه میانگین به روش کوتاه: خواص بالا می توانند برای ساده کردن محاسبه میانگین به کار

روند که «روش کوتاه» نامیده می شوند. در کاربرد این خواص، فرمول محاسبه میانگین را به صورت

زیر می توان بیان کرد :

$$\bar{X} = b\bar{Y} + a \quad (۴)$$

$$\bar{Y} = \frac{\sum F_i Y_i}{N} \quad \text{که در آن}$$

$$Y_i = \frac{X_i - a}{b} \quad \text{و}$$

a و b اعداد دلخواه هستند (در جدول توزیع فراوانیها با فاصله‌های مساوی، برای سهولت کار، a را حدّ وسط یکی از طبقات میانی جدول و b را فاصله طبقات در نظر می‌گیرند). چون از تمامی X_i ها عدد a کم شده است، بنابراین متوسط کمیّت جدید به اندازه عدد a کوچک خواهد شد (خاصیت ۴) و چون اعداد $(X_i - a)$ بر عدد b تقسیم شده‌اند، متوسط این اعداد نیز، b برابر کوچکتر خواهد شد (خاصیت ۵). از این رو، متوسط به دست آمده برای مقادیر Y_i ها (یعنی \bar{Y}) باید در عدد b ضرب و به حاصل، عدد a اضافه شود.

مثال ۷: قد دانش‌آموزان یک کلاس دبیرستانی را اندازه‌گیری کرده‌ایم و نتایج مشاهدات به صورت جدول توزیع فراوانی زیر به دست آمده است (جدول ۱۰):

جدول ۱۰

X فاصله طبقات	F_i
۱۴۰-۱۴۵	۳
۱۴۵-۱۵۰	۴
۱۵۰-۱۵۵	۱۰
۱۵۵-۱۶۰	۱۶
۱۶۰-۱۶۵	۸
۱۶۵-۱۷۰	۶
۱۷۰-۱۷۵	۳
-	۵۰

متوسط قد دانش‌آموزان این کلاس را با استفاده از روش کوتاه محاسبه می‌کنیم. برای این منظور، ابتدا وسط فاصله‌ها را به دست می‌آوریم (این اعداد در ستون سوم جدول ۱۱ قرار داده شده‌اند).

جدول ۱۱

X_i	F_i	X_i	$X_i - ۱۵۷/۵$	$Y_i = \frac{X_i - ۱۵۷/۵}{۵}$	$F_i Y_i$
۱۴۰-۱۴۵	۳	۱۴۲/۵	-۱۵	-۳	-۹
۱۴۵-۱۵۰	۴	۱۴۷/۵	-۱۰	-۲	-۸
۱۵۰-۱۵۵	۱۰	۱۵۲/۵	-۵	-۱	-۱۰
۱۵۵-۱۶۰	۱۶	۱۵۷/۵	۰	۰	۰
۱۶۰-۱۶۵	۸	۱۶۲/۵	۵	۱	۸
۱۶۵-۱۷۰	۶	۱۶۷/۵	۱۰	۲	۱۲
۱۷۰-۱۷۵	۳	۱۷۲/۵	۱۵	۳	۹
-	۵۰	-	-	-	۲

به منظور ساده کردن محاسبات، عدد دلخواه a را، مساوی با $۱۵۷/۵$ در نظر می‌گیریم. از تمامی مقادیر وسطی X_i ها، عدد دلخواه a را کم کرده، تفاضلها را در ستون ۴ جدول فوق قرار داده‌ایم. سپس تفاضلهای ستون ۴ را بر عدد $b = ۵$ تقسیم کرده، حاصل را در ستون ۵ جدول می‌نویسیم که این اعداد را با Y_i نشان داده‌ایم. حال آنها را در فراوانیهای متناظر ضرب کرده (که حاصل در ستون ۶ جدول نوشته شده است). حاصل جمع را برای محاسبه میانگین Y_i ها به دست می‌آوریم. بنابراین خواهیم داشت:

$$\bar{Y} = \frac{\sum F_i Y_i}{N} = \frac{۲}{۵۰} = ۰/۰۴$$

حال، متوسط صفت X را طبق فرمول ۴ محاسبه می‌کنیم:

$$\bar{X} = b\bar{Y} + a$$

$$= ۵ \times ۰/۰۴ + ۱۵۷/۵ = ۱۵۷/۷$$

— مقایسه توزیهای فراوانی به وسیله میانگین و نتیجه‌گیری نسبت به یکسان بودن آنها: از تعریف میانگین ملاحظه می‌شود که میانگین حساسی صفت متغیر، تابعی از توزیع صفت است، یعنی هم از مقادیر صفت و هم از فراوانیهای آن مقادیر، تبعیت می‌کند. بنابراین، اگر دو توزیع (دو جامعه)، دو میانگین حساسی متفاوت داشته باشند، آنگاه دو توزیع متفاوت خواهند بود. ولی اگر میانگینهای دو توزیع یکسان باشند، نمی‌توان نتیجه‌گیری کرد که دو توزیع یکسان هستند، زیرا ممکن است دو توزیع

یکسان باشند و نیز ممکن است یکسان نباشند. به عبارت دیگر، دو توزیع یکسان، میانگینهای حسابی یکسانی را مشخص می نمایند، ولی دو میانگین حسابی یکسان، دو توزیع یکسان را مشخص نمی کنند. به عنوان مثال، دو توزیع زیر را می توان در نظر گرفت (جداول ۱۲ و ۱۳):

جدول ۱۲

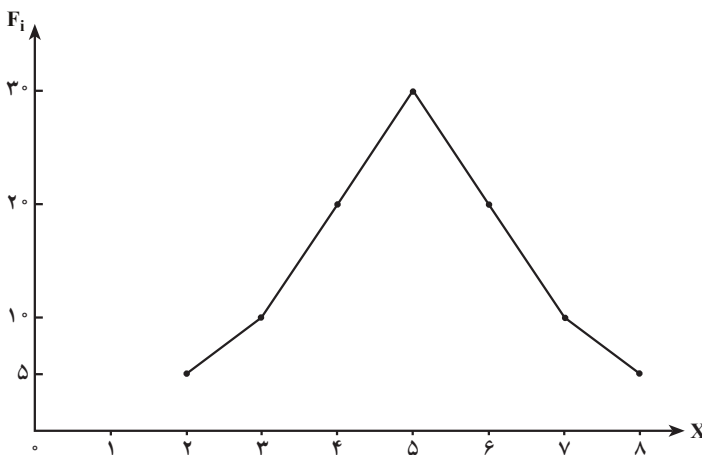
X_i	۲	۳	۴	۵	۶	۷	۸	-
F_i	۵	۱۰	۲۰	۳۰	۲۰	۱۰	۵	۱۰۰

جدول ۱۳

X	۳	۴	۵	۶	۷	-
F_i	۲	۳	۹۰	۳	۲	۱۰۰

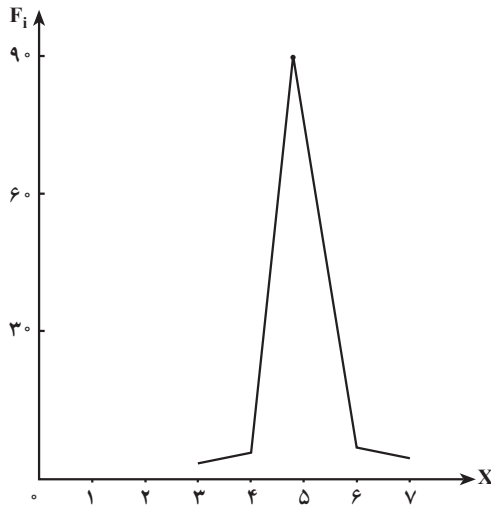
متوسط حسابی برای هر دو توزیع یکسان می باشد. $\bar{X}_1 = \bar{X}_2 = 5$

ولی به وضوح دیده می شود که این دو جدول، دو توزیع متفاوت را بیان می کنند. برای روشن شدن این تفاوت، می توان دو توزیع را با چندگوش نیز، نمایش داد (شکلهای ۱ و ۲):



شکل ۱- چندگوش توزیع صفت متغیر جدول ۱۲

بنابراین، همان گونه که از مثال فوق برمی آید، برای پی بردن به یکسان بودن توزیعهای صفت متغیر، همواره نمی توان تنها به مشخصه های مرکزی، مانند میانگین حسابی، اکتفا نمود.



شکل ۲- چند گوش صفت متغیر (جدول ۱۳)

میانۀ^۱

هم‌ردیف با میانگین، به عنوان مشخصه آماری توزیع صفت، کمیت‌های دیگری وجود دارند که در توزیع صفت متغیر به عنوان مشخصه مرکزی مورد استفاده قرار می‌گیرند. یکی از چنین مشخصه‌هایی، میانۀ می‌باشد که آن را به صورت زیر تعریف می‌کنیم:

میانۀ، مقدار صفتی است که جامعه را به ۲ گروه با حجم یکسان تقسیم می‌کند.

بر طبق معمول، میانۀ را با نماد Me نشان می‌دهیم.

از تعریف دقیق میانۀ برای متغیرهای گسسته در اینجا صرف‌نظر می‌کنیم.^۲

اگر مقادیر صفت متغیر گسسته را به طور صعودی مرتب کنیم، در صورتی که تعداد اعضای جامعه، عدد فرد باشد، یعنی اگر $(N = 2m + 1)$ باشد، آنگاه میانۀ، مقداری است که با شماره X_{m+1} مشخص می‌شود.

و در صورتی که تعداد اعضای جامعه عدد زوج باشد یعنی $(N = 2m)$ باشد، آنگاه میانۀ، متوسط حسابی عضو X_m و X_{m+1} خواهد بود، یعنی:

^۱ - Median

^۲ - به کتابهای آمار ریاضی مراجعه شود.

$$Me = \frac{X_m + X_{m+1}}{2}$$

مثلاً اگر مقادیر مشاهدات به صورت زیر به دست آمده باشد :

۵, ۲, ۳, ۳, ۴, ۳, ۲, ۶, ۴, ۲, ۳

برای تعیین میانه، ابتدا این مقادیر را به طور صعودی مرتب می‌کنیم.

۲, ۲, ۲, ۳, ۳, ۳, ۳, ۴, ۴, ۵, ۶

چون در اینجا $N = 2m + 1 = 11$ می‌باشد در نتیجه $m = 5$ خواهد بود در این صورت عدد

وسطی یعنی عضو $X_{m+1} = X_{5+1} = X_6 = 3$ میانه می‌باشد.

مثال دیگر - اگر مجموعه مشاهدات به صورت زیر باشد :

۳, ۴, ۴, ۶, ۸, ۹

چون در اینجا $N = 2m = 6$ و $m = 3$ ، در نتیجه میانه مقدار صفتی است که با میانگین

حسابی X_m و X_{m+1} مشخص می‌شود. یعنی :

$$Me = \frac{X_m + X_{m+1}}{2} = \frac{X_3 + X_4}{2} = \frac{4 + 6}{2} = 5$$

خواهد بود.

اگر حجم مشاهدات (N)، بزرگ باشد و مقادیر صفت گروه‌بندی شده باشند، دیگر نوشتن

تمامی مشاهدات، به صورت دنباله، مشکل و گاهی غیرممکن می‌شود. در چنین حالتی، عضو وسطی

را برای توزیع، به وسیله فراوانیهای انباشته، جستجو می‌کنیم.

مثال ۸: فرض کنید، توزیع صفت متغیر به وسیله جدول فراوانیهای زیر بیان شده باشد :

جدول ۱۴

X_i	۲	۳	۴	۵	۶	۷	-
F_i	۲	۲	۳	۵	۴	۴	۲۰

برای به دست آوردن میانه صفت، ابتدا فراوانیهای انباشته را، محاسبه می‌کنیم :

جدول ۱۵

X_i	۲	۳	۴	۵	۶	۷	-
F_i	۲	۲	۳	۵	۴	۴	۲۰
FC_i	۲	۴	۷	۱۲	۱۶	۲۰	-

حال چنین استدلال می‌کنیم :

حجم مشاهدات $N = 20$ است. باز هم ردیف میانه را مشخص می‌کنیم :

$$Me = \frac{X_m + X_{m+1}}{2} = \frac{X_{10} + X_{11}}{2}$$

یعنی در دنباله مقادیر صفت، عضو دهم و عضو یازدهم، اعضای وسطی هستند. می‌خواهیم ببینیم این اعضای وسطی (شماره ۱۰ و ۱۱) در کدام گروه قرار دارند. با توجه به فراوانیهای انباشته، مشاهده می‌شود، عضو یکم و دوم در گروه $X = 2$ هستند.

عضو سوم و چهارم در گروه $X = 3$ هستند.

عضو پنجم، ششم و هفتم در گروه $X = 4$ هستند.

عضو هشتم، نهم، دهم، یازدهم و دوازدهم در گروه $X = 5$ قرار دارند.

بنابراین اعضای وسطی، یعنی عضو دهم و یازدهم، در گروه $X = 5$ ، یعنی از صفت متغیر،

مقدار $X = 5$ را دارند و طبق تعریف بالا، خواهیم داشت :

$$Me = \frac{X_{10} + X_{11}}{2} = \frac{5 + 5}{2} = 5$$

اگر توزیع صفت، برحسب فاصله‌ها بیان شده باشد (اغلب در حالت صفت متغیر پیوسته)، پس از محاسبه فراوانیهای انباشته در جدول، فاصله‌ای که $\frac{N}{2}$ در آن فراوانی انباشته قرار گرفته باشد، آن فاصله را «فاصله میانه» یا طبقه میانه‌دار می‌نامند و توسط فرمول زیر مقدار میانه را به دست می‌آورند :

$$Me = X_i + \frac{\frac{N}{2} - FC_i}{F_i} \times I \quad (5)$$

که در آن :

X_i - حد پایین «فاصله میانه» یا طبقه میانه‌دار

FC_i - فراوانی انباشته طبقه قبل از طبقه میانه‌دار

F_i - فراوانی مطلق طبقه میانه‌دار

I - فاصله طبقه میانه‌دار

می‌باشند.

مثال ۹: فرض کنید می‌خواهیم میانۀ حجم تولید مؤسسات تولیدکننده مواد شیمیایی (X) را به‌دست آوریم (جدول ۱۶)، ابتدا ستون فراوانیهای انباشته را تشکیل می‌دهیم:

جدول ۱۶

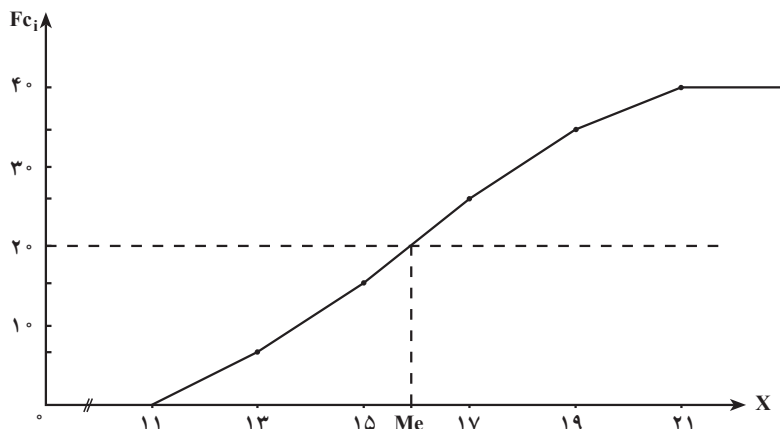
مقدار محصول X	تعداد مؤسسات F _i	فراوانی انباشته FC _i
۱۱-۱۳	۷	۷
۱۳-۱۵	۹	۱۶
۱۵-۱۷	۱۰	۲۶
۱۷-۱۹	۸	۳۴
۱۹-۲۱	۶	۴۰
-	۴۰	-

$$\frac{N}{2} = \frac{40}{2} = 20 \quad \text{سپس نصف حجم مشاهدات را تعیین می‌کنیم:}$$

چون نصف حجم جامعه، در فاصله ۱۵-۱۷ قرار گرفته، بنابراین، طبقه میانه‌دار، طبقه ۱۵-۱۷ می‌باشد، طبق فرمول ۵، میانه را به‌صورت زیر محاسبه می‌کنیم:

$$\begin{aligned} Me &= X_i + \frac{\frac{N}{2} - FC_i}{F_i} \times I \\ &= 15 + \frac{20 - 16}{10} \times 2 \\ &= 15 + \frac{8}{10} = 15.8 \end{aligned}$$

میانه را می‌توان به‌طور هندسی نیز با استفاده از چندگوش فراوانیهای انباشته، به‌دست آورد (شکل ۳).



شکل ۳- تعیین مقدار میانه برای مقدار محصول در مثال (۹)

— خواص میانه

۱- محاسبه میانه از محاسبه میانگین آسان تر است.

۲- برای توزیعهای صفت که مقادیر آن در دامنه‌های توزیع، نامعین باشد، میانگین به‌عنوان مقدار متوسط، نمی‌تواند محاسبه گردد، در صورتی که، میانه را برای چنین توزیعهایی، می‌توان محاسبه کرد.

مثلاً، اگر نمایه (لیست) حقوق کارگران یک مؤسسه به‌صورت جدول زیر در دست باشد :

جدول ۱۷

حقوق کارگران (برحسب هزار تومان) X	تعداد کارگران F _i
کمتر از ۳۰ هزار تومان	۱۰
۳۰-۴۰	۲۰
۴۰-۵۰	۴۰
۵۰-۶۰	۱۵
۶۰-۷۰	۱۰
۷۰ هزار تومان به‌بالا	۵
—	۱۰۰

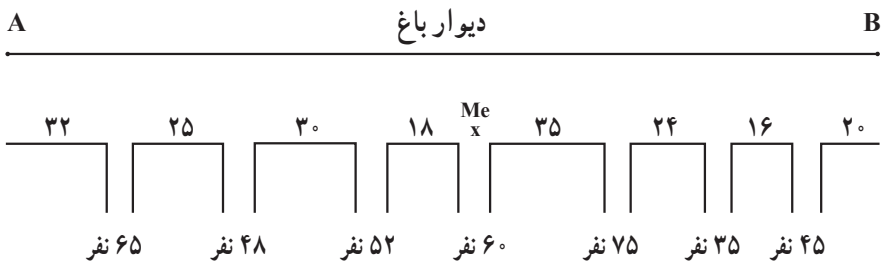
برای چنین توزیعی، محاسبه میانگین حسابی، امکان‌پذیر نیست، زیرا باید وسط فاصله‌ها را برای هرگروه محاسبه کرد، که چون گروه اول و گروه آخر، حدود معینی ندارند، بنابراین وسط فاصله آنها را نمی‌توان به‌دست آورد. ولی برای این توزیع، میانه را می‌توان محاسبه نمود، چون میانه با طبقات میانی جدول محاسبه می‌شود و به طبقات اول و آخر جدول کاری ندارد.

۳- میانه، حاصل جمع قدرمطلق انحرافات مقادیر صفت را از خودش به حداقل می‌رساند:

$. X_i - Me = \min \quad (۶)$ <p style="text-align: right; margin-right: 20px;">یا</p> $. F_i X_i - Me = \min \quad (۷)$
--

از این خاصیت برای نصب یا ایجاد وسایلی که استفاده همگانی دارند، استفاده می‌کنند. مثلاً برای نصب باجه تلفن عمومی در یک معبر و یا تعیین محل مناسب برای ایستگاه اتوبوس در یک خیابان و مواردی از این قبیل، از نقطه میانه این مکانها استفاده می‌شود.

مثال ۱۰: شرکت مخابرات در نظر دارد یک باجه تلفن عمومی در خیابان معینی نصب کند. موقعیت خیابان و فاصله مکانها و تعداد افراد ساکن در این مکانها در شکل زیر نشان داده شده است:



برای تعیین جای مناسب نصب باجه تلفن، جدول توزیع فراوانی فاصله‌ها را از نقطه A به‌دست آورده، سپس با تشکیل فراوانیهای انباشته، میانه توزیع را تعیین می‌کنیم:

جدول ۱۸

فاصله مکانها تا نقطه A X	تعداد نفرات مکانها F _i	فراوانی انباشته FC _i
۳۲	۶۵	۶۵
۵۷	۴۸	۱۱۳
۸۷	۵۲	۱۶۵
۱۰۵	۶۰	۲۲۵
۱۴۰	۷۵	۳۰۰
۱۶۴	۳۵	۳۳۵
۱۸۰	۴۵	۳۸۰
—	۳۸۰	—

نصف حجم جامعه $\left(\frac{N}{2}\right)$ را محاسبه می‌کنیم:

$$\frac{N}{2} = \frac{380}{2} = 190$$

با مقایسه $\left(\frac{N}{2}\right)$ ، ۱۹۰ با ستون فراوانیهای انباشته، معلوم می‌شود که در فراوانی انباشته ۲۲۵ قرار دارد و گروه X متناظر آن فاصله ۱۰۵ می‌باشد، بنابراین میانه این خیابان نقطه ۱۰۵ متری از نقطه A می‌باشد که در شکل خیابان نشان داده شده است.

چنانچه میانه این خیابان را از نقطه B محاسبه کنیم، میانه، فاصله ۹۵ متری از نقطه B به دست خواهد آمد که دقیقاً روی نقطه قبلی خواهد افتاد.

نما (مد)

یکی دیگر از مشخصه‌های مرکزی «نما» است که به صورت زیر تعریف می‌شود:

«مقدار صفتی که بزرگترین فراوانی یا بزرگترین چگالی فراوانی را در توزیع فراوانی، دارا باشد، «نما» نامیده می‌شود.»

نمای صفت را بعد از این با Mo نشان خواهیم داد.

مثال ۱۱: فرض کنید توزیع صفت متغیر، به وسیله جدول زیر بیان شده باشد:

جدول ۱۹

X	۲	۳	۴	۵	۶	-
F _i	۱۰	۲۵	۴۰	۲۰	۵	۱۰۰

چون بزرگترین فراوانی در جدول، ۴۰ است و این فراوانی به مقدار ۴ تعلق دارد، بنابراین، نما در این توزیع ۴ خواهد بود، یعنی:

$$Mo = 4$$

در توزیعهای فاصله‌ای، ابتدا می‌توان، طبقه یا فاصله‌ای را که شامل نما می‌شود، تعیین کرد. این فاصله را «فاصله نما» یا «طبقه نما» می‌نامند. در توزیعهای با فاصله‌های مساوی، فاصله نما، از روی بزرگترین فراوانی تعیین می‌گردد. ولی در توزیعهای با فاصله‌های نامساوی، فاصله نما، به وسیله بزرگترین چگالی فراوانی، تعیین می‌شود. برای به دست آوردن نما در توزیعهای با فاصله‌های مساوی، از فرمول زیر استفاده می‌شود:

$$Mo = X_i + \frac{F_i - F_{(i-1)}}{F_i - F_{(i-1)} - (F_i - F_{(i+1)})} \times I \quad (8)$$

که در آن

X_i - حد پایین فاصله نما

I - طول فاصله نما

F_i - فراوانی مطلق متعلق به فاصله نما (بزرگترین فراوانی در جدول)

$F_{(i-1)}$ - فراوانی متعلق به فاصله مجاور ماقبل نما

$F_{(i+1)}$ - فراوانی متعلق به فاصله مجاور بعد از نما

می‌باشند. برای ساده نمودن فرمول ۷ می‌توان از تبدیلهای زیر استفاده نمود:

$$F_i - F_{(i-1)} = d_1$$

$$F_i - F_{(i+1)} = d_2$$

در این صورت فرمول ۸ به صورت زیر نوشته می‌شود:

$$Mo = X_i + \frac{d_1}{d_1 + d_2} \times I \quad (9)$$

مثال ۱۲: توزیع قد دانش‌آموزان یک کلاس دبیرستانی در جدول ۲۰ آورده شده است. نمای قد دانش‌آموزان را محاسبه نمایید.

جدول ۲۰

X قد دانش‌آموزان	F _i
۱۴۰-۱۴۵	۳
۱۴۵-۱۵۰	۶
۱۵۰-۱۵۵	۱۵
۱۵۵-۱۶۰	۱۸
۱۶۰-۱۶۵	۱۰
۱۶۵-۱۷۰	۸
-	۶۰

چون بزرگترین فراوانی ۱۸ است، گروه متناظر آن، فاصله ۱۶۰-۱۵۵ است. بنابراین مطابق فرمول ۸، نما را محاسبه می‌کنیم:

$$\begin{aligned}
 Mo &= 155 + \frac{18-15}{(18-15)+(18-10)} \times 5 \\
 &= 155 + \frac{15}{11} = 155 + 1/35 = 156/35
 \end{aligned}$$

نکته: در هر توزیع، به بیش از یک میانگین و یک میانه، نمی‌توان دست یافت ولی می‌توان بیش از یک نما به دست آورد.

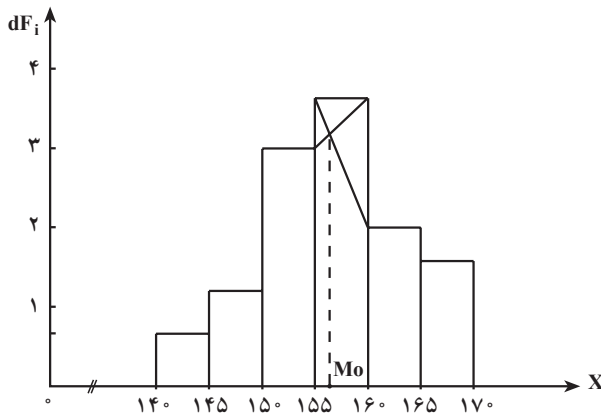
نما را نیز می‌توان با استفاده از شکل توزیع صفت در جامعه، تعیین نمود. نما را برای صفت متغیر (قد دانش‌آموزان) در مثال ۱۲ با شکل (هیستوگرام) توزیع به دست می‌آوریم.

جدول ۲۱

X	F_i	dF_i
۱۴۰-۱۴۵	۳	۰/۶
۱۴۵-۱۵۰	۶	۱/۲
۱۵۰-۱۵۵	۱۵	۳
۱۵۵-۱۶۰	۱۸	۳/۶
۱۶۰-۱۶۵	۱۰	۲
۱۶۵-۱۷۰	۸	۱/۶
-	۶۰	-

ابتدا چگالی فراوانیها را برای گروههای X در جدول توزیع فراوانی محاسبه می کنیم (جدول ۲۱).

سپس هیستوگرام توزیع را رسم می کنیم :



شکل ۴- تعیین نما با استفاده از هیستوگرام توزیع

در هیستوگرام توزیع، در داخل بلندترین ستون، از ستونهای مجاور، دو خط را مطابق شکل ۴ رسم می کنیم، سپس از نقطه تقاطع این دو خط، عمودی بر محور طولها، ترسیم می کنیم. مقدار صفتی که با طول نقطه عمود بر روی محور طولها، متناظر است، نمای صفت خواهد بود. در بعضی موارد، نما برای توزیع صفت، یک مشخصه مرکزی برتر است که بیشترین مقادیر

صفت در حول آن متمرکز می‌گردد. از این رو، در توصیف جامعه، در چنین حالاتی به عنوان مشخصه مرکزی، نما نسبت به میانگین برتری دارد.

مثلاً در مورد تقاضا نسبت به یک اندازه معین کفش یا لباس، «نما» نسبت به میانگین یک مشخصه بهتر است. به طور کلی، میانگین در اینجا ارزشی ندارد.

باید متذکر شد که برای توزیعهای کاملاً قرینه، مقادیر میانگین حسابی (\bar{X})، میانه (Me) و نما (Mo) برهم منطبق می‌شوند، یعنی:

$$\bar{X} = Me = Mo$$

ولی اگر توزیع قرینه نباشد (البته اگر از حالت قرینگی انحراف بسیار نداشته باشد)، بین این سه مشخصه مرکزی، می‌توان رابطه تقریبی زیر را نوشت:

$$\bar{X} - Mo = 3(\bar{X} - Me)$$

یا

$$Mo = 3Me - 2\bar{X}$$

از تساوی تقریبی فوق، می‌توان برای تعیین مقدار یکی از ۳ مشخصه مرکزی، وقتی که ۲ تای دیگر معلوم باشد، استفاده کرد.

مثلاً برای یک توزیع تقریباً غیرقرینه، اگر میانگین برابر ۴۲ و میانه برابر ۴۰ باشد، با استفاده از رابطه بالا، مقدار نما را محاسبه می‌کنیم:

$$Mo = \bar{X} - 3(\bar{X} - Me) = 42 - 3(42 - 40) = 36$$

یا

$$Mo = 3Me - 2\bar{X} = 3 \times 40 - 2 \times 42 = 36$$



سوآلها و تمرینها

۱- مشخصه‌های آماری چه مشخصه‌هایی هستند؟ لزوم وارد کردن آنها در آمار

چيست؟

۲- مشخصه‌های مرکزی، چه ویژگی از صفت متغیر را بیان می‌کنند؟

۳- میانگین حسابی را تعریف کنید.

۴- توزیع میزان فروش یک شرکت بازرگانی در ماه گذشته در جدول زیر بیان شده است، میانگین حسابی فروش روزانه این شرکت را محاسبه کنید.

X (میزان فروش بر حسب هزار تومان)	۱۰۰-۳۰۰	۳۰۰-۵۰۰	۵۰۰-۷۰۰	۷۰۰-۹۰۰	۹۰۰-۱۱۰۰	۱۱۰۰-۱۳۰۰	-
F _i تعداد روز	۳	۷	۱۰	۳	۲	۱	۲۶

- ۵- میانگین حسابی ساده چیست؟ در چه مواردی از آن استفاده می‌کنند؟
 ۶- میانگین حسابی برای توزیعهای فاصله‌ای، چگونه محاسبه می‌شود؟
 ۷- توزیع سن کارگران یک واحد تولیدی، در جدول زیر بیان شده است:

X سن کارگران	۱۸-۲۲	۲۳-۲۷	۲۸-۳۲	۳۳-۳۷	۳۸-۴۲	۴۳-۴۷	-
F _i فراوانی	۴	۱۴	۲۰	۱۶	۱۶	۱۰	۸۰

متوسط حسابی سن کارگران را محاسبه کنید.

۸- جدول زیر، توزیع خانوارهای یک منطقه مسکونی را بر حسب تعداد افراد خانوار نشان

می‌دهد:

X تعداد افراد خانوار	۳	۴	۵	۶	۷	۸	-
f _i فراوانی نسبی	۰/۰۵	۰/۱۴	۰/۳۰	۰/۲۵	۰/۱۶	۰/۱۰	۱

متوسط تعداد افراد خانوار را در این منطقه محاسبه کنید.

۹- خواص میانگین حسابی را بیان کنید. این خواص چه اهمیتی دارند و کاربردشان چیست؟

۱۰- اگر در توزیعهای فاصله‌ای، فاصله طبقات نامساوی باشد، می‌توان از روش کوتاه،

میانگین را محاسبه نمود؟ چرا؟

۱۱- میانگین تمرین ۷ را با استفاده از روش کوتاه، محاسبه نمایید.

۱۲- میانه را تعریف کنید.

- ۱۳- اگر حجم مشاهدات در جامعه، زوج باشد، میانه را چگونه تعیین می‌کنند؟
- ۱۴- یکی از خواص میانه که لزوم آن را در آمار، اجباری می‌سازد، کدام است؟
- ۱۵- اگر توزیع صفت، به وسیله جدول توزیع فراوانیها، بیان شده باشد، میانه را چگونه محاسبه می‌کنند؟
- ۱۶- تعداد تصادفات اتومبیل در روز، در یک شهر، طی دو ماه گذشته به صورت جدول زیر گروه‌بندی شده است :

X تعداد تصادف	۱۰	۱۲	۱۵	۱۶	۱۸	۲۰	-
F_i روز	۳	۱۵	۱۸	۱۴	۷	۳	۶۰

- میانه را برای تعداد تصادفات اتومبیل به دست آورید.
- ۱۷- مبلغ اضافه کار پرداختی در یک شرکت تولیدی در ماه گذشته به صورت جدول زیر گروه‌بندی شده است :

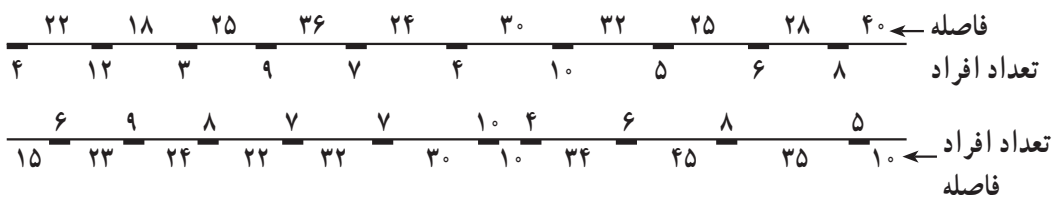
X مبلغ اضافه کار پرداختی بر حسب هزار تومان	۲۰-۳۰	۳۰-۳۵	۳۵-۴۰	۴۰-۴۵	۴۵-۵۰	-
F_i فراوانی	۸	۱۸	۲۵	۱۲	۷	۷۰

- میانه، اضافه کار پرداختی در این شرکت، چقدر است؟
- ۱۸- میزان پارچه‌های تولید شده در کارخانه‌های بافندگی کشور در سال گذشته به صورت جدول صفحه بعد گروه‌بندی شده است : (واحد میلیون متر پارچه)

تعداد کارخانه	پارچه تولیدشده برحسب میلیون متر
F_i	X
۵	۱۰ و کمتر
۱۷	۱۰-۲۰
۲۸	۲۰-۴۰
۳۵	۴۰-۶۰
۲۵	۶۰-۱۰۰
۱۰	۱۰۰ و بیشتر
۱۲۰	-

میانه پارچه تولید شده در کارخانه‌های بافندگی کشور را محاسبه کنید.

- ۱۹- یکی از خواص مهم میانه، که در برنامه‌ریزیها مورد استفاده قرار می‌گیرد، کدام است؟
- ۲۰- فرض کنید شرکت اتوبوسرانی تصمیم دارد در یک خیابان به طول ۲۸۰ متر، یک ایستگاه اتوبوس دایر کند، با توجه به مشخصات شکل زیر و با استفاده از خاصیت میانه، نقطه مناسب محل ایستگاه را مشخص کنید.



- (اعداد خارج از معبر فاصله‌ها را و اعداد داخل معبر تعداد افراد ساکن در خانه‌ها را نشان می‌دهد.)
- ۲۱- نما را تعریف کنید.

- ۲۲- نمرات درس ریاضی برای دانش‌آموزان یک کلاس به صورت زیر گروه‌بندی شده است:

X نمره ریاضی	۶	۷	۸	۱۰	۱۱	۱۲	۱۴	۱۶	-
F_i فراوانی	۳	۲	۵	۱۲	۶	۱۵	۷	۴	۵۴

نما برای نمرات درس ریاضی، چه نمره‌ای است؟

۲۳- نما را برای تمرین ۸، به دست آورید.

۲۴- نما را برای تمرین ۷ (سن کارگران یک واحد تولیدی) محاسبه کنید.

۲۵- نما را برای تمرین ۱۷ (اضافه کار پرداختی) محاسبه کنید.

۲۶- نما را برای تمرین ۱۶ (تعداد تصادفات اتومبیل) به دست آورید.

۲۷- نما را برای تمرین ۱۷ با استفاده از شکل هیستوگرام به دست آورید.

۲۸- در توزیعهای قرینه، چه رابطه‌ای بین مشخصه‌های مرکزی (میانگین، میانه و نما)، وجود

دارد؟

۲۹- در توزیعهای غیر قرینه (که انحراف آنها کم است)، چه رابطه‌ای بین میانگین، میانه و نما

وجود دارد؟ آن را بنویسید.

خودآزمونهای چهارگزینه‌ای

۱- کدام یک از گزینه‌های زیر، یکی از خواص میانگین را نشان می‌دهد؟

الف - $(X_i - \bar{X}) = 0$. ب - $(X_i - \bar{X}) = \max$.

ج - $(X_i - \bar{X}) \neq 0/5$. د - $(X_i - \bar{X}) = \min$.

۲- میانگین ۱۰ عدد برابر ۸ می‌باشد، X_i . کدام است؟

الف - ۴۰ ب - ۱۶۰ ج - ۸۰ د - ۱۶

۳- اگر به تمامی مقادیر صفت متغیر، یک عدد ثابت مانند $a = 5$ اضافه کنیم،

الف - میانگین به اندازه ۵ واحد کم می‌شود.

ب - میانگین تغییری نمی‌کند.

ج - میانگین به اندازه ۵ واحد زیاد می‌شود.

د - میانگین دو برابر می‌شود.

۴- میانه ۳۲ عضو برابر ۱۶ می‌باشد، اگر فراوانی مطلق طبقه میانه ۶ و فاصله طبقه میانه ۴ و

فراوانی انباشته طبقه ماقبل $\frac{N}{4}$ ، ۱۰ باشد، کرانه پایین طبقه میانه، کدام است؟

الف - ۱۰ ب - ۸ ج - ۱۲ د - ۶

۵- کدام یک از گزینه‌های زیر یکی از خواص میانه می‌باشد؟

الف - $(X_i - Me) = 0$. ب - $|X_i - Me| = 0$.

ج - $|X_i - Me| = \max$. د - $|X_i - Me| = \min$.

۶- کدام یک از گزینه‌های زیر تعریف میانه است؟

الف - مقدار صفتی که بیشترین فراوانی را دارد.

ب - مقداری از صفت که در وسط توزیع مرتب شده قرار دارد.

ج - حاصل جمع مقادیر صفت تقسیم بر تعدادشان

د - بزرگترین اندازه صفت در جامعه

۷- هر توزیع صفت متغیر

الف - فقط یک میانگین و یک نما دارد.

ب - فقط یک میانه و یک نما دارد.

ج - فقط یک میانگین و یک میانه دارد.

د - بیش از یک میانه دارد.

۸- نهضت سوادآموزی، زمان پخش برنامه تلویزیونی را از زنان روستایی، نظرخواهی نموده

است، کدام مشخصه مرکزی برای تعیین زمان پخش برنامه نهضت مناسب است؟

الف - میانگین ب - میانه ج - نما د - هیچکدام

۹- در توزیعهای قرینه، کدام یک از روابط زیر صحیح است؟

الف - $\bar{X} \# Me \# Mo$ ب - $\bar{X} < Me < Mo$

ج - $\bar{X} = Me = Mo$ د - $\bar{X} = 3Me - 2Mo$

۱۰- در توزیعهای تقریباً غیرقرینه کدام رابطه بین سه مشخصه مرکزی وجود دارد؟

الف - $\bar{X} = 3Me - 2Mo$ ب - $Mo = 3Me - 2\bar{X}$

ج - $Me = 3\bar{X} - 2Mo$ د - $Mo = 3\bar{X} - 2Me$

مشخصه‌های پراکندگی

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند:

- ۱- دلایل به کار بردن مشخصه‌های پراکندگی را بیان کند.
- ۲- مشخصه‌های مهم پراکندگی را معرفی کند.
- ۳- نحوه محاسبه واریانس را برای توزیعهای گسسته و پیوسته، به کار برد.
- ۴- خواص مشخصه‌های پراکندگی و کاربرد آنها را بیان کند.
- ۵- برای توزیعهای با دامنه نامعین، پراکندگی را با استفاده از چارکها محاسبه کند.
- ۶- مشخصه‌های نسبی پراکندگی را معرفی کند.
- ۷- مشخصه‌های چولگی را برای توزیعهای با دامنه معلوم و نامعین محاسبه کند.
- ۸- با مشخصه‌های پراکندگی بتواند جامعه‌ها را مقایسه کند.

در مطالعه تغییرپذیری صفت متغیر برای اعضای جامعه، تنها نمی‌توان به متوسط اندازه صفت، اکتفا نمود، زیرا قبلاً دیدیم که یک میانگین همواره به یک جامعه تعلق ندارد، بلکه دو یا چند توزیع متفاوت، ممکن است میانگینهای یکسان داشته باشند، ولی پراکندگی آنها متفاوت باشد. برای مشخص نمودن پراکندگی صفت متغیر، می‌توان از انحرافات مقادیر صفت از همدیگر یا انحرافات مقادیر صفت از مشخصه‌های دیگر مانند مشخصه‌های مرکزی استفاده کرد.

مشخصه‌های پراکندگی

طول دامنه تغییرات

در آمار برای اندازه‌گیری پراکندگی، مشخصه‌های مختلفی مورد استفاده قرار می‌گیرند. یکی از

ساده‌ترین آنها، که به آسانی با توزیع صفت متغیر، محاسبه می‌شود «طول دامنه تغییرات» صفت می‌باشد. طول دامنه تغییرات از تفاضل بزرگترین اندازه صفت و کوچکترین مقدار آن به دست می‌آید.

$$R = X_{\max} - X_{\min} \quad (1)$$

فرض کنیم مقادیر مشاهده شده صفت متغیر در جامعه به قرار زیر به دست آمده باشد:

$$X: 3, 5, 7, 9, 7, 4, 5, 8, 11, 10$$

دامنه تغییرات صفت، مطابق فرمول ۱ به صورت زیر محاسبه می‌شود:

چون بزرگترین اندازه صفت، ۱۱ و کوچکترین اندازه آن، ۳ می‌باشد، بنابراین:

$$R = 11 - 3 = 8$$

خواهد بود. ولی دامنه تغییرات، به عنوان مشخصه پراکندگی، تصویر صحیحی درباره چگونگی تراکم مقادیر، حول مقدار متوسط صفت، ارائه نمی‌دهد و فقط از تغییر کوچکترین و بزرگترین مقدار صفت، متأثر می‌شود. از این رو، برای ارزیابی نوسانات مقدار صفت نسبت به مقدار متوسط، مشخصه‌های پراکندگی دیگری مورد استفاده قرار می‌گیرند که از نظر کلی به دامنه تغییرات ارجحیت دارند.

متوسط قدر مطلق انحرافات (انحراف متوسط^۱)

اگر بخواهیم نوسانات مقادیر صفت را حول میانگین، با میانگین انحرافات ارزیابی کنیم، یعنی مشخصه پراکندگی را به صورت:

$$\frac{\sum F_i (X_i - \bar{X})}{N}$$

بیان کنیم، چون صورت کسر، مطابق خاصیت ۳ میانگین، همواره مساوی صفر است، درباره پراکندگی واقعی صفت، چیزی به دست نخواهد آمد.

برای برطرف کردن این مشکل، از قدر مطلق انحرافات و یا از مجذور انحرافات استفاده می‌شود. محاسبه میانگین قدر مطلق انحرافات که آن را با نماد AD نشان می‌دهیم، به قرار زیر خواهد بود:

$$AD = \frac{\sum F_i |X_i - \bar{X}|}{N} \quad (2)$$

و اگر فراوانیها برای مقادیر صفت، برابر ۱ باشد، رابطه بالا به صورت صفحه بعد در خواهد آمد:

۱- Average Deviation = AD

$$AD = \frac{\sum |X_i - \bar{X}|}{N} \quad (3)$$

روش محاسبه آن را با ذکر یک مثال بیان می‌کنیم.

مثال ۱: برای توزیع صفت متغیر که با جدول زیر بیان شده است، متوسط قدرمطلق انحرافات را

محاسبه می‌کنیم:

X	۲	۳	۴	۵	۶	-
F _i	۱۰	۲۰	۳۰	۱۶	۴	۸۰

ابتدا میانگین حسابی را برای این توزیع محاسبه می‌کنیم، مقدار آن برابر $\bar{X} = 3/8$ است.

بنابراین، طبق رابطه ۲ خواهیم داشت:

$$AD = \frac{\sum F_i |X_i - \bar{X}|}{N}$$

$$= \frac{10|2 - 3/8| + 20|3 - 3/8| + 30|4 - 3/8| + 16|5 - 3/8| + 4|6 - 3/8|}{80}$$

$$= \frac{18 + 16 + 6 + 19/2 + 8/8}{80} = \frac{68}{80} = 0/85$$

تمام محاسبات می‌تواند در جدول زیر انجام پذیرد (جدول ۱):

جدول ۱

X	F _i	F _i X _i	X _i - \bar{X}	X - \bar{X}	F _i X _i - \bar{X}
۲	۱۰	۲۰	۲ - ۳/۸	۱/۸	۱۸
۳	۲۰	۶۰	۳ - ۳/۸	۰/۸	۱۶
۴	۳۰	۱۲۰	۴ - ۳/۸	۰/۲	۶
۵	۱۶	۸۰	۵ - ۳/۸	۱/۲	۱۹/۲
۶	۴	۲۴	۶ - ۳/۸	۲/۲	۸/۸
-	۸۰	۳۰۴	-	-	۶۸

$$\bar{X} = \frac{\sum F_i X_i}{N} = \frac{304}{80} = 3/8$$

$$AD = \frac{\sum F_i |X_i - \bar{X}|}{N} = \frac{68}{80} = 0/85$$

یعنی به طور متوسط، هریک از مقادیر صفت، به اندازه $0/85$ واحد از متوسط خود انحراف دارند.

اگر صفت متغیر با توزیع فاصله‌ای، بیان شده باشد، همانطور که در محاسبه متوسط حسابی، گفته شد، ابتدا وسط فاصله‌ها را تعیین کرده، سپس مانند توزیع گسسته، عمل می‌کنیم.
 مثال ۲: متوسط قدرمطلق انحرافات را برای مؤسّسات تولیدکننده مواد شیمیایی در مثال ۹ از فصل سوم، محاسبه می‌کنیم (جدول ۲):

جدول ۲

X	F _i	X _i	F _i X _i	X _i - \bar{X}	F _i X _i - \bar{X}
۱۱-۱۳	۷	۱۲	۸۴	۳/۸۵	۲۶/۹۵
۱۳-۱۵	۹	۱۴	۱۲۶	۱/۸۵	۱۶/۶۵
۱۵-۱۷	۱۰	۱۶	۱۶۰	۰/۱۵	۱/۵۰
۱۷-۱۹	۸	۱۸	۱۴۴	۲/۱۵	۱۷/۲۰
۱۹-۲۱	۶	۲۰	۱۲۰	۴/۱۵	۲۴/۹۰
-	۴۰	-	۶۳۴	-	۸۷/۲۰

$$\bar{X} = \frac{634}{40} = 15/85$$

$$AD = \frac{\sum F_i |X_i - \bar{X}|}{N} = \frac{87/20}{40} = 2/18$$

متوسط قدرمطلق انحرافات، یک اندازه طبیعی برای پراکندگی توزیع می‌باشد، لیکن به علت فاقد بودن بعضی ویژگیهای ریاضی، از این مشخصه، کمتر استفاده می‌شود.

متوسط مجذور انحرافات یا واریانس^۱ (پراش)

در عمل به جای استفاده از علامت قدرمطلق، از مجذور انحرافات استفاده می‌کنند. بنابراین یکی دیگر از معیارهای پراکندگی، متوسط مجذور انحرافات است که آن را واریانس می‌نامند و با نماد $V(X)$ نشان داده و با فرمول زیر تعریف می‌شود:

۱- Variance

$$V(X) = \frac{F_1(X_1 - \bar{X})^2 + F_2(X_2 - \bar{X})^2 + \dots + F_k(X_k - \bar{X})^2}{F_1 + F_2 + \dots + F_k}$$

$$= \frac{\sum F_i(X_i - \bar{X})^2}{N} \quad (4)$$

و اگر مقادیر صفت بدون فراوانی باشند (فراوانی هر کدام یک باشد):

$$V(X) = \frac{\sum (X_i - \bar{X})^2}{N} \quad (5)$$

محاسبه واریانس را با مثال زیر نشان می‌دهیم:

مثال ۳: برای توزیع صفت متغیر که با جدول زیر بیان شده است، واریانس را محاسبه می‌کنیم:

X	۲	۳	۴	۵	-
F _i	۱۰	۴۰	۳۰	۲۰	۱۰۰

متوسط حسابی برای این توزیع $\bar{X} = 3/6$ است، بنابراین با استفاده از فرمول ۴ خواهیم داشت:

$$V(X) = \frac{\sum F_i(X_i - \bar{X})^2}{N}$$

$$= \frac{10(2 - 3/6)^2 + 40(3 - 3/6)^2 + 30(4 - 3/6)^2 + 20(5 - 3/6)^2}{10 + 40 + 30 + 20}$$

$$= 0/84$$

تمامی محاسبات می‌تواند در جدول زیر انجام پذیرد (جدول ۳):

جدول ۳

X _i	F _i	F _i X _i	X _i - \bar{X}	(X _i - \bar{X}) ²	F _i (X _i - \bar{X}) ²
۲	۱۰	۲۰	-۱/۶	۲/۵۶	۲۵/۶
۳	۴۰	۱۲۰	-۰/۶	۰/۳۶	۱۴/۴
۴	۳۰	۱۲۰	۰/۴	۰/۱۶	۴/۸
۵	۲۰	۱۰۰	۱/۴	۱/۹۶	۳۹/۲
-	۱۰۰	۳۶۰	-	-	۸۴

$$\bar{X} = \frac{360}{100} = 3/6$$

$$V(X) = \frac{\sum F_i(X_i - \bar{X})^2}{N} = \frac{184}{100} = 1.84$$

در محاسبه واریانس، اغلب به علت اینکه متوسط حسابی، با رقمهای اعشاری باید از مقادیر صفت کم شود، به توان رساندن این اعداد اعشاری، محاسبه را دشوارتر می‌سازد. از این رو، فرمول واریانس را به صورت زیر تبدیل می‌کنند:

$$V(X) = \frac{\sum F_i(X_i - \bar{X})^2}{N} = \frac{\sum F_i X_i^2}{N} - \bar{X}^2 \quad (6)$$

$$V(X) = \frac{\sum F_i(X_i - \bar{X})^2}{N} = \frac{\sum F_i X_i^2 - \frac{(\sum F_i X_i)^2}{N}}{N} \quad (7)$$

و اگر فراوانی هر کدام از اندازه‌ها مساوی یک باشد، فرمولهای ۶ و ۷ به صورت زیر نوشته خواهد شد:

$$V(X) = \frac{\sum (X_i - \bar{X})^2}{N} = \frac{\sum X_i^2}{N} - \bar{X}^2 \quad (8)$$

$$V(X) = \frac{\sum (X_i - \bar{X})^2}{N} = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{N}}{N} \quad (9)$$

فرمولهای ۶ تا ۹ را فرمولهای کاربردی واریانس نامند.
مثال ۴: واریانس را برای توزیع صفت متغیر که در مثال ۳ بیان شده است، با استفاده از فرمول ۶ محاسبه می‌کنیم (جدول ۴):

جدول ۴

X_i	F_i	$F_i X_i$	$F_i X_i^2$
۲	۱۰	۲۰	۴۰
۳	۴۰	۱۲۰	۳۶۰
۴	۳۰	۱۲۰	۴۸۰
۵	۲۰	۱۰۰	۵۰۰
—	۱۰۰	۳۶۰	۱۳۸۰

$$\bar{X} = \frac{360}{100} = 3.6$$

$$V(X) = \frac{\sum F_i X_i^2}{N} - \bar{X}^2$$

$$= \frac{1380}{40} - (3/6)^2 = 13/8 - 12/96 = 0/84$$

اگر صفت متغیر، به صورت توزیع فاصله‌ای بیان شده باشد، مانند محاسبه متوسط حسابی، ابتدا، وسط فاصله‌ها را تعیین کرده، سپس مانند توزیع گسسته، عمل می‌کنند.

مثال ۵: واریانس را برای توزیع صفت متغیر که با جدول ۲ در مثال ۲ بیان شده است، محاسبه می‌کنیم. تمامی محاسبات در جدول ۵ آورده شده است:

جدول ۵

X	F _i	X _i	F _i X _i	F _i X _i ^۲
۱۱-۱۳	۷	۱۲	۸۴	۱۰۰۸
۱۳-۱۵	۹	۱۴	۱۲۶	۱۷۶۴
۱۵-۱۷	۱۰	۱۶	۱۶۰	۲۵۶۰
۱۷-۱۹	۸	۱۸	۱۴۴	۲۵۹۲
۱۹-۲۱	۶	۲۰	۱۲۰	۲۴۰۰
-	۴۰	-	۶۳۴	۱۰۳۲۴

$$V(X) = \frac{\sum F_i X_i^2 - \frac{(\sum F_i X_i)^2}{N}}{N}$$

$$= \frac{10324 - \frac{(634)^2}{40}}{40} = \frac{275/1}{40} = 6/8775$$

انحراف معیار^۱

واریانس، دارای ویژگیهای مهم ریاضی است که در اینجا درباره آنها بحث نمی‌کنیم لیکن به علت اینکه، واریانس، پراکندگی را به صورت مجذور انحرافات بیان می‌کند، برای اینکه متوسط انحراف برای یک عضو مشخص گردد، از واریانس جذر گرفته و آن را انحراف معیار صفت متغیر می‌نامند و با حرف یونانی σ نشان می‌دهند.

^۱ - Standard Deviation

$$\begin{aligned}
 &= .+ \sqrt{V(X)} \\
 &= +\sqrt{\frac{\sum F_i (X_i - \bar{X})^2}{N}} \quad (۱۰)
 \end{aligned}$$

با وجود اینکه در آمار نظری، به عنوان اندازه پراکندگی، به طور اساسی، واریانس مورد استفاده قرار می‌گیرد، ولی انحراف معیار، مناسب‌ترین و متداولترین مشخصه پراکندگی است که در عمل از آن استفاده می‌شود. از این رو، در بحثهای بعدی، پراکندگی را با انحراف معیار، ارزیابی خواهیم کرد.

مثال ۶: انحراف معیار را برای مثال ۵ محاسبه می‌کنیم:

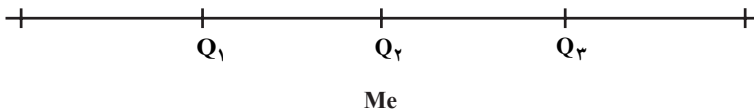
$$= .+ \sqrt{V(X)} = \sqrt{6/8775} = 2/622$$

انحراف معیار، نیز مانند متوسط قدر مطلق انحرافات، متوسط انحرافات اندازه‌های صفت را از میانگین صفت نشان می‌دهد.

انحراف چارکی^۱

به همان دلایلی که میانه به عنوان مشخصه مرکزی تعریف شده است، برای تغییرپذیری جامعه‌هایی که دامنه توزیع صفت در آنها نامشخص باشد، از مشخصه‌های آماری که از دامنه توزیع متأثر نمی‌باشند، استفاده می‌کنند. چنین مشخصه‌های آماری، چارکها^۲ هستند که میانه یکی از آنها می‌باشد. برای تعیین پراکندگی توزیع صفت در چنین حالتی از انحراف چارکی استفاده می‌شود که به صورت زیر تعریف می‌گردد:

اگر طول دامنه تغییرات صفت را به چهار قسمت مساوی از نظر حجم تقسیم کنیم، سه عددی که این دامنه را به چهار قسمت افراز می‌کند به ترتیب چارک اول، چارک دوم و چارک سوم نامند و با Q_1 ، Q_2 و Q_3 نشان داده می‌شوند.



متوسط تفاضل چارک دوم و چارک اول با تفاضل چارک سوم و چارک دوم را انحراف چارکی نامند و آن را با نماد QD نشان می‌دهند.

^۱ - Quartile Deviation

^۲ - Quartiles

فرمول آن به صورت زیر بیان می‌شود :

$$QD = \frac{(Q_2 - Q_1) + (Q_3 - Q_2)}{2} = \frac{Q_3 - Q_1}{2} \quad (11)$$

چارکها، مقدار صفت متغیر در مجموعه مرتب شده مقادیر صفت هستند که ۲۵ درصد اعضای جامعه، کوچکتر از Q_1 و ۲۵ درصد اعضای جامعه، بین Q_1 و Q_2 و ۲۵ درصد اعضای جامعه بین Q_2 و Q_3 و ۲۵ درصد بقیه اعضا، بالاتر از Q_3 قرار می‌گیرند.

چارکهای اول و سوم، یعنی Q_1 و Q_3 ، مشابه میانه، طبق فرمولهای زیر محاسبه می‌شوند :

$$Q_1 = X_i + \frac{\frac{N}{4} - FC_i}{F_i} \times I \quad (12)$$

که در آن

X_i - نقطه شروع فاصله گروهی از صفت متغیر است که چارک اول در آن قرار دارد.

I - طول فاصله‌ای است که چارک اول در آن قرار دارد.

FC_i - فراوانی انباشته گروه قبل از گروه چارک اول می‌باشد.

F_i - فراوانی مطلق گروه چارک اول است.

$$Q_3 = X_i + \frac{\frac{3N}{4} - FC_i}{F_i} \times I \quad (13)$$

که در آن

X_i - نقطه شروع فاصله گروهی از صفت متغیر است که چارک سوم در آن قرار دارد.

I - طول فاصله‌ای است که چارک سوم در آن قرار دارد.

FC_i - فراوانی انباشته گروه قبل از گروه چارک سوم می‌باشد.

F_i - فراوانی مطلق گروه چارک سوم است.

بدیهی است، برای محاسبات چارک اول و چارک سوم، باید ستون فراوانیهای تجمعی را در

جدول توزیع صفت، محاسبه کنیم.

مثال ۷: فرض کنید توزیع فراوانی کارگران برحسب مزد و حقوق ماهانه آنها در یک مؤسسه با

جدول زیر بیان شده باشد. می‌خواهیم تغییرپذیری (پراکندگی) مزد و حقوق ماهانه را برای این جامعه

محاسبه کنیم :

جدول ۶

مزد و حقوق ماهانه کارگران X	تعداد کارگران F _i	فراوانی انباشته FC _i
تا ۴۰۰۰۰	۱۲	۱۲
۴۰۰۰۰-۴۵۰۰۰	۳۸	۵۰
۴۵۰۰۰-۵۰۰۰۰	۴۴	۹۴
۵۰۰۰۰-۵۵۰۰۰	۵۰	۱۴۴
۵۵۰۰۰-۶۰۰۰۰	۶۲	۲۰۶
۶۰۰۰۰-۷۰۰۰۰	۴۲	۲۴۸
۷۰۰۰۰-۸۰۰۰۰	۲۵	۲۷۳
۸۰۰۰۰-۱۰۰,۰۰۰	۱۴	۲۸۷
۱۰۰,۰۰۰ تومان به بالا	۱۳	۳۰۰
-	۳۰۰	-

ملاحظه می‌شود که مشخصه‌هایی مانند «طول دامنه تغییرات»، «متوسط قدر مطلق انحرافات»، واریانس و انحراف معیار به‌عنوان مشخصه پراکندگی برای این توزیع نمی‌تواند محاسبه گردد. زیرا حدود بالا و پایین جامعه نامشخص است. از این رو «انحراف چارکی» را برای این توزیع فراوانیها، محاسبه می‌کنیم. ابتدا ستون فراوانیهای انباشته را محاسبه می‌کنیم که نتیجه آن در ستون سوم جدول ۶ آمده است.

محاسبه چارک اول (Q_1): برای این منظور $\frac{N}{4}$ را تعیین می‌کنیم:

$$\frac{N}{4} = \frac{300}{4} = 75$$

با مقایسه مقدار چارک اول (۷۵) با ستون فراوانیهای انباشته جدول، مشاهده می‌شود که در فراوانی انباشته (۹۴) متعلق به فاصله ۴۵۰۰۰-۵۰۰۰۰ قرار دارد، بنابراین چارک اول در فاصله ۴۵۰۰۰-۵۰۰۰۰ می‌باشد. با استفاده از فرمول ۱۲ مقدار Q_1 را محاسبه می‌کنیم.

$$X_i = 45000 \quad \text{در اینجا:}$$

$$I = 50000 - 45000 = 5000$$

$$FC_i = 50$$

$$F_i = 44$$

$$\begin{aligned}
 Q_1 &= X_i + \frac{\frac{N}{4} - FC_i}{F_i} \times I \\
 &= 450000 + \frac{75 - 50}{44} \times 50000 \\
 &= 450000 + \frac{1250000}{44} = 450000 + 2841 = 47841
 \end{aligned}$$

مشابه روش بالا، چارک سوم (Q_3) را محاسبه می‌کنیم. برای این منظور $\frac{3N}{4}$ را تعیین می‌کنیم:

$$\frac{3N}{4} = \frac{3 \times 300}{4} = 225$$

با $\frac{3N}{4}$ ، فاصله‌ای را که شامل چارک سوم است، جستجو می‌کنیم. چون ۲۲۵ در فراوانی انباشته (۲۴۸) که متعلق به فاصله ۷۰۰۰۰-۶۰۰۰۰ است، قرار دارد، از این رو، چارک سوم در فاصله ۷۰۰۰۰-۶۰۰۰۰ می‌باشد. با استفاده از فرمول ۱۳، چارک سوم را محاسبه می‌کنیم.

$$X_i = 60000 \quad \text{در اینجا:}$$

$$I = 70000 - 60000 = 10000$$

$$FC_i = 206$$

$$F_i = 42$$

$$\begin{aligned}
 Q_3 &= X_i + \frac{\frac{3N}{4} - FC_i}{F_i} \times I \\
 &= 60000 + \frac{225 - 206}{42} \times 10000 \\
 &= 60000 + \frac{190000}{42} = 60000 + 4524 = 64524
 \end{aligned}$$

حال با استفاده از فرمول ۱۱، انحراف چارکی (QD) را محاسبه می‌کنیم:

$$\begin{aligned}
 QD &= \frac{Q_3 - Q_1}{2} \\
 &= \frac{64524 - 47841}{2} = 8341/5
 \end{aligned}$$

چون تمامی مقادیر صفت متغیر، بر روی انحراف چارکی، تأثیر نمی‌کنند (حداکثر ۵۰ درصد از مقادیر صفت در آن تأثیر دارند)، از این مشخصه در مواردی که محاسبه انحراف معیار، دشوار یا غیرممکن است، استفاده می‌شود.

مشخصه‌های نسبی پراکندگی

برای مقایسه تغییرات صفت در دو یا چند جامعه با میانگینهای مختلف، از مشخصه‌های نسبی پراکندگی، استفاده می‌شود. این مشخصه‌ها از نسبت مشخصه‌های پراکندگی به مشخصه‌های مرکزی صفت، محاسبه می‌شوند.

یکی از مهمترین مشخصه‌های نسبی پراکندگی، ضریب تغییرات است که به صورت زیر تعریف می‌شود:

ضریب تغییرات^۱

نسبت انحراف معیار صفت متغیر بر متوسط حسابی صفت که برحسب درصد بیان شده باشد، ضریب تغییرات، نامیده می‌شود و آن را با C.V نشان می‌دهند.

$$C.V = \frac{s}{\bar{X}} \times 100 \quad (14)$$

که در آن:

– انحراف معیار صفت متغیر است.

\bar{X} – متوسط حسابی صفت متغیر، می‌باشد.

مثال ۸: ضریب تغییرات را برای صفت متغیر که در مثال ۵ آورده شده است، و برای آن

$$\bar{X} = 15/85 \text{ و } s = 2/622 \text{ می‌باشند، محاسبه می‌کنیم:}$$

$$C.V = \frac{s}{\bar{X}} \times 100 = \frac{2/622}{15/85} \times 100 = 16/54\%$$

یعنی تغییرات صفت، ۱۶/۵ درصد اندازه متوسط آن را تشکیل می‌دهد.

ضریب تغییرات نه تنها برای بیان تغییرپذیری نسبی به کار می‌رود، بلکه به عنوان «مشخصه همگنی» جامعه نیز مورد استفاده قرار می‌گیرد. هرگاه ضریب تغییرات صفت در جامعه از ۳۳ درصد تجاوز نکند، آن جامعه را از نظر صفت متغیر، «همگن» در نظر می‌گیرند.

مثلاً جامعه‌ای که در مثال بالا آورده شد، نسبت به صفت متغیر، همگن می‌باشد.

وقتی به عنوان اندازه تغییرپذیری، متوسط قدر مطلق انحرافات (A.D)، مورد استفاده قرار

می‌گیرد، مشخصه نسبی پراکندگی به صورت انحراف خطی نسبی، محاسبه می‌شود:

^۱ – Coefficient of variation

انحراف خطی نسبی

نسبت متوسط قدر مطلق انحرافات بر متوسط حسابی صفت متغیر که بر حسب درصد بیان شده باشد، «انحراف خطی نسبی» نامیده می‌شود. اگر آن را با $C.V_{AD}$ نشان دهیم، به صورت زیر نوشته می‌شود:

$$C.V_{AD} = \frac{AD}{\bar{X}} \times 100 \quad (15)$$

که در آن:

AD - متوسط قدر مطلق انحرافات صفت متغیر است.

\bar{X} - متوسط حسابی صفت متغیر، می‌باشد.

مثال ۹: انحراف خطی نسبی را برای صفت متغیر در مثال ۲، محاسبه می‌کنیم که برای آن

$\bar{X} = 15/85$ و $AD = 2/18$ قبلاً محاسبه شده است.

$$\begin{aligned} C.V_{AD} &= \frac{AD}{\bar{X}} \times 100 \\ &= \frac{2/18}{15/85} \times 100 = 13/75- \end{aligned}$$

درحالاتی که استفاده از مشخصه‌های تغییر پذیری (پراکندگی) مذکور، امکان پذیر نباشد، مناسب

است از مشخصه نسبی پراکندگی دیگری به نام «انحراف چارکی نسبی» استفاده گردد.

انحراف چارکی نسبی

نسبت انحراف چارکی بر میانه صفت متغیر «انحراف چارکی نسبی» نامیده می‌شود که آن را با q نشان می‌دهیم:

$$q = \frac{QD}{Me} \times 100 \quad (16)$$

مثال ۱۰: انحراف چارکی نسبی را برای صفت متغیر در مثال ۷ محاسبه می‌کنیم

که برای آن: $QD = 8341/5$ و $Me = 55484$ می‌باشد.

$$q = \frac{QD}{Me} \times 100 = \frac{8341/5}{55484} \times 100 = 15-$$

مشخصه چولگی (انحراف از قرینگی)

برای پی بردن به خصوصیات عمومی توزیع، اغلب، مشخصه‌های انحراف از قرینگی یا چولگی محاسبه می‌شود.

قبل از اینکه مشخصه‌های چولگی را معرفی کنیم، باید بدانیم توزیع قرینه چیست، تا براساس آن، غیرقرینه بودن، مشخص گردد.

هرگاه فراوانیهای متعلق به دو مقدار صفت که از مرکز توزیع، در فاصله‌های مساوی قرار دارند، یکسان باشند، آن توزیع را «قرینه» می‌خوانند. در غیر این صورت، توزیع را «غیرقرینه» یا «چوله» می‌نامند.

مثلاً در توزیع فراوانی‌هایی که با جدول زیر بیان شده است، فراوانیها، نسبت به وسط توزیع، دوه‌دو باهم برابرند:

X	۵	۱۰	۱۵	۲۰	۲۵
F _i	۱۰	۴۰	۸۰	۴۰	۱۰

در صورتی که در توزیع فراوانی‌هایی که در جدول زیر بیان شده است، فراوانیها برای مقادیر صفت که نسبت به وسط توزیع در فاصله‌های مساوی قرار گرفته‌اند، یکسان نیستند.

X _i	۲	۴	۶	۸	۱۰	۱۲
F _i	۱۰	۲۰	۴۰	۴۵	۲۵	۱۰

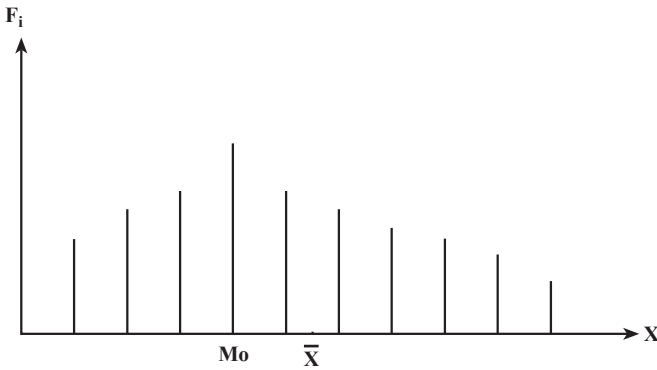
برای مقایسه درجه‌دوری توزیعها از حالت قرینه بودن، از مشخصه‌نسبی که «ضریب چولگی» نام دارد، استفاده می‌شود.

— ضریب چولگی پیرسن: در فصل گذشته، ملاحظه شد که برای توزیعهای قرینه، مقدار میانگین، میانه و نما باهم برابرند. در ارتباط با آن، ساده‌ترین مشخصه انحراف از قرینگی یا چولگی که ضریب چولگی پیرسن نام دارد، بر مناسبات بین مشخصه‌های مرکزی استوار گردیده است.

هرچه تفاوت بین مشخصه‌های مرکزی $(\bar{X} - Mo)$ ، بیشتر باشد، چولگی نیز بیشتر است. ضریب چولگی پیرسن به صورت زیر تعریف می‌شود:

$$P_{AS} = \frac{\bar{X} - Mo}{.} \quad (17)$$

کمیت P_{AS} ، می تواند مثبت، منفی یا صفر باشد. مقدار مثبت این ضریب، وجود چولگی راست را نشان می دهد یعنی دامنه راست توزیع نسبت به نما، بیشتر کشیده است تا دامنه چپ آن (مانند نمودار شکل ۱):



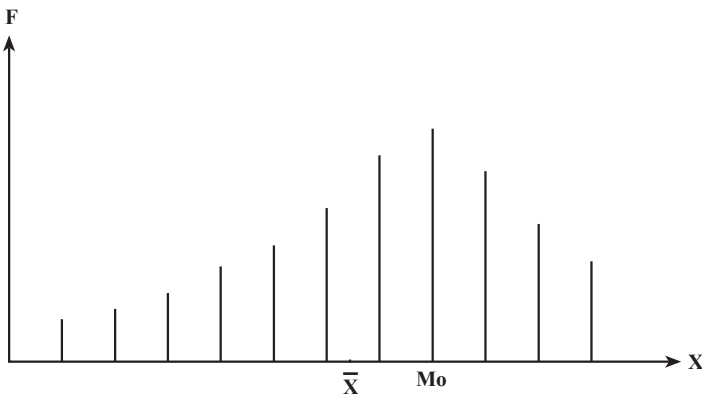
شکل ۱- وجود چولگی راست

در چولگی راست، بین مشخصه های مرکزی رابطه زیر برقرار است:

$$Mo < Me < \bar{X} \quad (18)$$

علامت منفی در این ضریب، گواه وجود چولگی چپ است که در شکل ۲ نشان داده شده

است:



شکل ۲- وجود چولگی چپ

در این حالت، بین مشخصه‌های مرکزی، رابطه زیر برقرار است :

$Mo \# Me \# \bar{X}$	(۱۹)
-----------------------	------

معمولاً وقتی ضریب چولگی کوچکتر از ۱/۵ باشد، توزیع را تقریباً قرینه و وقتی که ضریب چولگی بزرگتر از ۵/۵ باشد آن را خیلی چوله در نظر می‌گیرند. طبیعی است که در توزیعهای قرینه، ضریب چولگی مساوی با صفر خواهد بود.

مثال ۱۱: ضریب چولگی پیرسن (P_{AS}) را برای توزیع زیر محاسبه می‌کنیم :

X	۱۰-۱۵	۱۵-۲۰	۲۰-۲۵	۲۵-۳۰	۳۰-۳۵	-
F_i	۴	۱۰	۲۰	۵	۱	۴۰

برای محاسبه ضریب چولگی پیرسن، میانگین حسابی، نما و انحراف معیار توزیع را باید به دست آورد. تمامی محاسبات در جدول ۷ آورده شده است.

جدول ۷

X	F_i	X_i	$F_i X_i$	$F_i X_i^2$
۱۰-۱۵	۴	۱۲/۵	۵۰	۶۲۵
۱۵-۲۰	۱۰	۱۷/۵	۱۷۵	۳۰۶۲/۵
۲۰-۲۵	۲۰	۲۲/۵	۴۵۰	۱۰۱۲۵
۲۵-۳۰	۵	۲۷/۵	۱۳۷/۵	۳۷۸۱/۲۵
۳۰-۳۵	۱	۳۲/۵	۳۲/۵	۱۰۵۶/۲۵
-	۴۰	-	۸۴۵	۱۸۶۵۰

$$\bar{X} = \frac{\sum F_i X_i}{N} = \frac{845}{40} = 21.125$$

$$V(X) = \frac{\sum F_i X_i^2 - \frac{(\sum F_i X_i)^2}{N}}{N}$$

$$= \frac{18650 - \frac{(845)^2}{40}}{40} = \frac{799.375}{40} = 19.98$$

$$= .+ \sqrt{V(X)} = \sqrt{19/98} = 4/47$$

$$\begin{aligned} Mo &= X_i + \frac{F_i - F_{i-1}}{(F_i - F_{i-1}) + (F_i - F_{i+1})} \times I \\ &= 20 + \frac{20 - 10}{(20 - 10) + (20 - 5)} \times 5 \\ &= 20 + \frac{50}{25} = 22 \end{aligned}$$

در نتیجه، ضریب چولگی پیرسن طبق فرمول ۱۷، به صورت زیر محاسبه می‌شود:

$$P_{AS} = \frac{\bar{X} - Mo}{.} = \frac{21/125 - 22}{4/47} = -0/19$$

ولی باید متذکر شد که ضریب چولگی پیرسن همیشه نمی‌تواند مورد استفاده قرار گیرد. مثلاً

اگر توزیع دو نما داشته باشد، این ضریب نمی‌تواند محاسبه گردد.

— ضریب چولگی بر پایه میانگین توان سوم انحرافات:

یکی از متداول‌ترین ضرایب چولگی که نقص ضریب چولگی پیرسن را ندارد و از تمامی

عناصر توزیع (یعنی از مقادیر صفت و فراوانیها) در محاسبه آن استفاده می‌شود ضریب چولگی بر پایه

میانگین توان سوم انحرافات می‌باشد که با M_{AS} نشان داده می‌شود، به صورت زیر بیان می‌شود:

$$M_{AS} = \frac{\sum F_i (X_i - \bar{X})^3}{N^3} = \frac{\sum F_i (X_i - \bar{X})^3}{N^3} \quad (20)$$

مثال ۱۲: ضریب چولگی بر پایه میانگین توان سوم انحرافات را برای توزیع فراوانیهای صفت

متغیر که با جدول زیر بیان شده است، محاسبه می‌کنیم:

X	۳	۴	۵	۶	۷	۸	—
F_i	۱	۶	۱۳	۱۱	۱۰	۹	۵۰

تمامی محاسبات در جدول ۸، آورده شده است:

جدول ۸

X	F _i	F _i X _i	F _i X _i ^۲	X _i - \bar{X}	(X _i - \bar{X}) ^۲	F _i (X _i - \bar{X}) ^۲
۳	۱	۳	۹	-۳	-۲۷	-۲۷
۴	۶	۲۴	۹۶	-۲	-۸	-۴۸
۵	۱۳	۶۵	۳۲۵	-۱	-۱	-۱۳
۶	۱۱	۶۶	۳۹۶	۰	۰	۰
۷	۱۰	۷۰	۴۹۰	۱	۱	۱۰
۸	۹	۷۲	۵۷۶	۲	۸	۷۲
-	۵۰	۳۰۰	۱۸۹۲	-	-	-۶

$$\bar{X} = \frac{300}{50} = 6$$

$$V(X) = \frac{\sum F_i X_i^2}{N} - \bar{X}^2$$

$$= \frac{1892}{50} - 6^2 = 37.84 - 36 = 1.84$$

$$= \sqrt{V(X)} = \sqrt{1.84} = 1.36$$

$$\sigma^2 = 1.36 \times 1.84 = 2.50$$

با استفاده از کمیتهای محاسبه شده، طبق فرمول ۲۰، ضریب چولگی را محاسبه می کنیم:

$$M_{AS} = \frac{\sum F_i (X_i - \bar{X})^3}{\sum F_i} = \frac{-6}{2.5} = -0.048$$

بنابراین، چون ضریب چولگی، کوچکتر از ۱/۰ به دست آمده است، می توان گفت، توزیع تقریباً قرینه است.

سؤالها و تمرینها ؟

- ۱- چرا در مطالعه صفت متغیر، برای اعضای جامعه، تنها به کمیتهای متوسط نمی توان اکتفا کرد؟
- ۲- اندازه پراکندگی چیست؟ ساختن اندازه پراکندگی انحرافات مقادیر صفت از متوسط صفت، چه اهمیتی دارد؟
- ۳- طول دامنه تغییرات چیست؟
- ۴- طول دامنه تغییرات را برای داده های زیر به دست آورید :
 $3, 4, 7, 8, 10, 9, 5, 7, 11, 6, 13, 10, 15$
- ۵- طول دامنه تغییرات را برای طول قد دانش آموزان که در جدول زیر داده شده است، محاسبه کنید.

X طول قد	۱۴۴-۱۴۸	۱۴۸-۱۵۲	۱۵۲-۱۵۶	۱۵۶-۱۶۰	۱۶۰-۱۶۴	
F _i فراوانی	۲	۶	۱۰	۱۵	۷	۴۰

- ۶- نقص طول دامنه تغییرات به عنوان یک مشخصه پراکندگی در چیست؟
- ۷- چرا از انحرافات مقادیر صفت، نسبت به متوسط حسابی، نمی توانیم متوسط انحرافات را محاسبه کنیم؟
- ۸- متوسط قدرمطلق انحرافات را تعریف کنید.
- ۹- متوسط قدرمطلق انحرافات را برای توزیع قد دانش آموزان در تمرین ۵ محاسبه کنید.
- ۱۰- چرا در عمل، از متوسط قدرمطلق انحرافات، کمتر استفاده می شود؟
- ۱۱- متوسط مجذور انحرافات یا واریانس، چه مشخصه ای است؟
- ۱۲- واریانس به عنوان مشخصه پراکندگی، برای برطرف کردن چه مشکلی در آمار، وارد گردیده است؟
- ۱۳- فرمولهای کاربردی یا فرمولهای محاسباتی واریانس، کدامها هستند؟ آنها را بنویسید.
- ۱۴- واریانس را برای طول قد دانش آموزان در تمرین ۵ محاسبه کنید.

۱۵- واریانس را برای اندازه‌های صفت که در زیر داده شده‌اند، محاسبه کنید.

$X: 3, 5, 3, 6, 7, 2, 8, 10$

۱۶- چرا به جای واریانس، برای نشان دادن پراکندگی صفت، از انحراف معیار، استفاده

می‌شود؟

۱۷- انحراف معیار چیست؟ تعریف کنید.

۱۸- انحراف معیار را برای تمرین ۵ (طول قد دانش‌آموزان)، محاسبه کنید.

۱۹- برای توزیع فراوانیهای صفت که در زیر داده شده، انحراف معیار و متوسط قدر مطلق

انحرافات را محاسبه کرده، آنها را باهم مقایسه کنید. به نظر شما کدام دقیق‌تر است؟

X	۲	۴	۶	۸	۱۰	۱۲	-
F_i	۴	۷	۱۰	۶	۲	۱	۳۰

۲۰- انحراف چارکی چیست؟ تعریف کنید.

۲۱- لزوم وارد کردن انحراف چارکی به عنوان مشخصه پراکندگی، در چیست؟

۲۲- چارکها، چه مفهومی را بیان می‌کنند؟

۲۳- چارک اول چیست؟

۲۴- چارک دوم چیست؟

۲۵- چارک سوم چیست؟

۲۶- انحراف چارکی را برای توزیع فراوانیهای زیر محاسبه کنید :

X	۴-۸	۸-۱۲	۱۲-۱۶	۱۶-۲۰	۲۰-۲۴	۲۴ به بالا
F_i	۳	۵	۸	۱۲	۶	۶

۲۷- مشخصه‌های نسبی پراکندگی، چه مفهومی دارند؟

۲۸- چرا برای بیان پراکندگی صفت، از مشخصه‌های نسبی پراکندگی استفاده می‌شود؟

۲۹- ضریب تغییرات چیست؟

۳۰- ضریب تغییرات را برای تمرین ۱۵ محاسبه کنید.

۳۱- ضریب تغییرات را برای تمرین ۱۹ محاسبه کنید.

- ۳۲- ضریب «انحراف خطی نسبی»، چیست؟ در چه مواردی از آن استفاده می‌کنند؟
- ۳۳- انحراف خطی نسبی را برای تمرین ۵ محاسبه کنید.
- ۳۴- انحراف خطی نسبی را برای تمرین ۱۹ محاسبه کنید.
- ۳۵- انحراف چارکی نسبی چیست؟ در چه مواردی از آن استفاده می‌کنند؟
- ۳۶- انحراف چارکی نسبی را برای تمرین ۲۶ محاسبه کنید.
- ۳۷- برای پی‌بردن به خصوصیات عمومی توزیعها، از چه نوع مشخصه‌هایی استفاده می‌شود؟
- ۳۸- چه نوع توزیعی را، توزیع قرینه گویند؟ تعریف کنید.
- ۳۹- توزیع غیرقرینه یا توزیع چوله، چه نوع توزیعی است؟
- ۴۰- مشخصه چولگی یا ضریب چولگی چیست؟ چه خصوصیات از توزیع صفت را بیان می‌کند؟

- ۴۱- ضریب چولگی پیرسن، بر چه اساسی ساخته می‌شود؟
- ۴۲- ضریب چولگی پیرسن را برای تمرین ۱۵ محاسبه کنید.
- ۴۳- ضریب چولگی پیرسن را برای تمرین ۱۹ محاسبه کنید.
- ۴۴- ضریب چولگی منفی، چگونه تعبیر می‌شود؟
- ۴۵- ضریب چولگی مثبت، چگونه تعبیر می‌شود؟
- ۴۶- ضریب چولگی براساس متوسط توان سوم انحرافات چیست؟ فرمول آن را بنویسید.
- ۴۷- ضریب چولگی براساس متوسط توان سوم انحرافات را برای تمرین ۱۵ محاسبه کنید.
- ۴۸- ضریب چولگی براساس متوسط توان سوم انحرافات را برای تمرین ۱۹ محاسبه کنید.

خودآزمونهای چهارگزینه‌ای

- ۱- کدام یک از مشخصه‌های پراکندگی زیر، از نظر دقت، اعتبار بیشتری دارد؟
- الف - دامنه تغییرات
ب - واریانس
ج - متوسط قدرمطلق انحرافات
د - انحراف معیار
- ۲- از روی یک نمونه $n = 100$ در جامعه، کمیت‌های زیر محاسبه شده است:
- $$n = 100 \quad F_i X_i = 1200 \quad F_i (X_i - \bar{X})^2 = 36$$
- واریانس جامعه کدام است؟

الف - ۰/۳۶ ب - ۰/۱۶ ج - ۰/۶ د - ۰/۱۲

۳- اگر مجموع توان دوم، 10 اندازه از صفت متغیر $X^2 = 1700$ و میانگین آنها $\bar{X} = 12$ باشد، واریانس X کدام است؟

الف - ۶ ب - ۱۶ ج - ۲۶ د - صفر

۴- با معلوم بودن $\bar{X} = 10$ و $X^2 = 1200$ ، $V(X) = 20$ ، حجم جامعه چقدر بوده است؟

الف - ۸ ب - ۲۰ ج - ۱۲ د - ۱۰

۵- انحراف چارکی برای توصیف کدام مشخصه صفت، به کار می‌رود؟

الف - مشخصه مرکزی ب - مشخصه پراکندگی

ج - مشخصه چولگی د - هیچ کدام

۶- اگر بین مشخصه‌های مرکزی رابطه $Mo \# Me \# \bar{X}$ برقرار باشد، کدام گزینه برای صفت متغیر درست است؟

الف - چولگی به چپ است. ب - چولگی ندارد.

ج - چولگی به راست است. د - توزیع قرینه است.

۷- کدام یک از روابط زیر، فرمول ضریب چولگی بر پایه میانگین توان سوم انحرافات است؟

الف - $\frac{\sum F_i (X_i - \bar{X})^3}{N}$ ب - $\frac{\sum F_i (X_i - \bar{X})^2}{N}$

ج - $\frac{\sum F_i (X_i - \bar{X})^3}{N}$ د - $\frac{\sum F_i (X_i - \bar{X})^3}{N}$

۸- کدام گزینه تعریف ضریب تغییرات صفت می‌باشد؟

الف - نسبت میانگین بر انحراف معیار صفت

ب - نسبت میانه بر انحراف معیار صفت

ج - نسبت انحراف معیار بر میانه صفت

د - نسبت انحراف معیار بر میانگین صفت

۹- انحراف خطی نسبی از کدام گزینه به دست می آید؟

الف - $\frac{AD}{\bar{X}}$ ب - $\frac{QD}{\bar{X}}$ ج - $\frac{\dot{\quad}}{Me}$ د - $\frac{V(X)}{\bar{X}}$

۱۰- کدام یک از گزینه‌های زیر ضریب انحراف چارکی را بیان می کند؟

الف - $\frac{Q_3 - Q_2}{2}$ ب - $\frac{QD}{Me}$

ج - $\frac{Q_3 - Q_1}{2}$ د - $\frac{Q_3 - Q_1}{Me}$

فصل پنجم

توزیع صفت متغیر کیفی

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند:

- ۱- جامعه را برحسب صفات کیفی گروه‌بندی کند.
- ۲- مشخصه‌های مرکزی را برای صفات کیفی محاسبه نماید.
- ۳- مشخصه‌های پراکندگی را برای صفات کیفی محاسبه نماید.
- ۴- جامعه‌ها را با استفاده از مشخصه‌های پراکندگی باهم مقایسه کند و همگنی آنها را بیان نماید.
- ۵- جامعه‌ها را از نظر دو صفت متغیر کیفی که هر یک می‌توانند حالات مختلفی را اختیار کنند، گروه‌بندی نماید.
- ۶- جدول توزیع فراوانیهای نسبی را برحسب حالات دو صفت کیفی، تشکیل دهد.

توزیع صفت کیفی

همان‌طور که قبلاً در طبقه‌بندی صفت‌های متغیر گفته شد، صفت متغیر کیفی، صفتی است که آن را برای اعضای جامعه با عدد یا کمیت نتوان بیان نمود. برای بیان توزیع صفت کیفی فقط می‌توان حالت‌های آن را از هم متمایز ساخته، تعداد اعضای هر یک از حالتها را با فراوانیهای (مطلق یا نسبی) بیان نمود که جدول توزیع فراوانی آن به صورت زیر نشان داده می‌شود:

جدول ۱

حالات صفت	X	A_1	A_2	A_3	...	A_s	-
فراوانی	F_i	F_1	F_2	F_3	...	F_s	= N

A_1, A_2, \dots, A_s حالت‌های صفت کیفی برای اعضای جامعه می‌باشد. مثلاً اگر بخواهیم

طبقه‌بندی کارکنان یک مؤسسه را برحسب مشاغل بیان کنیم، توزیع فراوانیهای کارکنان برحسب صفت شغل به صورت جدول زیر بیان می‌شود:

جدول ۲

مدیر عامل ...	کارمند	راننده	انباردار	نگهبان	مستخدم	X مشاغل
... ۱ = N	۶	۳	۲	۱	۳	F _i فراوانی

و یا اگر بخواهیم طبقه‌بندی بیماران یک بیمارستان را برحسب نوع بیماری نشان دهیم، از جدول توزیع فراوانی بیماران برحسب نوع بیماری (X) استفاده می‌کنیم:

جدول ۳

-	...	بیماران گوارشی	بیماران عصبی	بیماران قلبی	بیماران عفونی	X نوع بیماری
= N	...	۱۲	۶	۱۵	۱۰	F _i فراوانی

بر طبق معمول در مطالعه صفت کیفی، آن حالت از صفت متغیر کیفی که مورد نظر می‌باشد از سایر حالت‌های صفت کیفی مجزا کرده، به صورت توزیع دو حالتی بیان می‌کنند. در این صورت، توزیع صفت کیفی را با دو حالت متضاد می‌نامند حالت مورد نظر را با A و بقیه حالتها را با \bar{A} نشان داده، توزیع فراوانیها را برای صفت دو حالت متضاد به صورت زیر بیان می‌کنند:

جدول ۴

X	\bar{A}	A	-
F _i	F \bar{A}	F _A	= N

یا به صورت توزیع فراوانیهای نسبی:

جدول ۵

X	\bar{A}	A	-
f _i	f \bar{A}	f _A	= ۱

مثال ۱: در مجموعه ۴۰ سند صادره در یک واحد حسابداری، برای تعیین کیفیت اسناد صادره، اسناد تحت بازرسی قرار گرفته‌اند و معلوم شده است که ۲۹ فقره از آنها «بدون نقص» (حالت \bar{A}) و بقیه «نقص‌دار» (حالت A) بوده‌اند. بنابراین مجموعه اسناد را می‌توان برحسب نقص‌دار و

بدون نقص به دو گروه تقسیم کرد که نتیجه آن در جدول ۶ بیان شده است :

جدول ۶

حالات صفت	بدون نقص	نقص دار	
	\bar{A}	A	-
فراوانی	۲۹	۱۱	= ۴۰

اگر بخواهیم جدول توزیع فراوانیهای نسبی را برای مجموعه اسناد بیان کنیم، کافی است هر یک از فراوانیهای مطلق را به N یعنی ۴۰ تقسیم کنیم، در نتیجه خواهیم داشت :

حالات صفت	\bar{A}	A	-
فراوانی نسبی f_i	$\frac{۲۹}{۴۰}$	$\frac{۱۱}{۴۰}$	= ۱

مثال ۲: فرض کنیم جامعه مورد مطالعه $N=۲۰۰$ واحد محصول تولیدشده در یک مؤسسه تولیدی باشد که واحدهای محصول از سوی کارشناسان، از نظر کیفیت تحت بازرسی قرار گرفته و براساس آن، محصول به چهار طبقه، درجه بندی شده است :

پس از شمارش واحدهای محصول در هر گروه، نتایج گروه بندی به صورت جدول ۷ درآمده است.

جدول ۷

کیفیت محصول	درجه ۱	درجه ۲	درجه ۳	درجه ۴	-
فراوانی	۶۰	۸۰	۴۰	۲۰	۲۰۰

اگر فراوانیهای هر گروه را بر حجم جامعه ($N=۲۰۰$) تقسیم نماییم، جدول بالا به صورت زیر درمی آید (جدول ۸):

جدول ۸

کیفیت محصول	درجه ۱	درجه ۲	درجه ۳	درجه ۴	-
فراوانی نسبی	۰/۳۰	۰/۴۰	۰/۲۰	۰/۱۰	۱

این دو جدول، توزیع فراوانیهای واحدهای محصول را برحسب کیفیت بیان می کند.

محاسبه مشخصه های عددی برای صفات کیفی

از آنجا که صفت متغیر، کیفی است و با حالت بیان می شود، نمی توان برای آنها مشخصه های عددی محاسبه نمود. برای این کار لازم است، حالات صفات کیفی را به صورت کمی بیان کرد و با آنها مانند صفات کمی برخورد نمود.

فرض کنیم صفت متغیر کیفی، حالت های A_1, A_2, \dots, A_S را بتواند در جامعه اختیار کند. اگر حالت های مختلف صفت را به دو گروه، حالت مورد نظر (A) و حالت غیر مورد نظر (\bar{A}) تقسیم بندی کنیم، در نتیجه، توزیع صفت به صورت زیر در خواهد آمد:

$$X \begin{pmatrix} A & \bar{A} \\ F_A & F_{\bar{A}} \end{pmatrix}$$

حال با تبدیل حالت مورد نظر (A) به عدد ۱ و حالت غیر مورد نظر (\bar{A}) به عدد صفر، توزیع به صورت زیر در خواهد آمد:

$$X \begin{pmatrix} 1 & 0 \\ F_A & F_{\bar{A}} \end{pmatrix}$$

اینک می توان برای چنین صفتی که به صورت کمی درآمده است، مشخصه مرکزی و مشخصه پراکندگی محاسبه نمود.

میانگین صفت کیفی خواهد شد:

$$\bar{X} = \frac{F_A}{N} = f_A \quad (3)$$

بنابراین، میانگین صفت کیفی، فراوانی نسبی حالت مورد نظر خواهد بود. واریانس صفت کیفی، به صورت زیر به دست می آید:

$$V(X) = f_A \cdot f_{\bar{A}} \quad (4)$$

بنابراین، واریانس صفت کیفی، حاصل ضرب فراوانی نسبی حالت مورد نظر و فراوانی نسبی حالت غیرمورد نظر خواهد بود.

مثال ۳: فرض کنید در مثال ۲ می خواهیم متوسط محصولات درجه ۱ (A) را محاسبه کنیم.

برای این کار، کافی است محصولات درجه ۱ را، حالت مورد نظر (A)، در نظر گرفته و بقیه محصولات را، حالت غیرمورد نظر (\bar{A})، در نظر بگیریم.
در نتیجه، توزیع به صورت زیر درمی آید:

$$X \begin{pmatrix} \text{محصولات درجه ۱} & \text{سایر محصولات} \\ A & \bar{A} \\ ۶۰ & ۱۴۰ \\ ۱ & ۰ \\ ۶۰ & ۱۴۰ \end{pmatrix} \begin{matrix} ۲۰۰ \\ ۲۰۰ \end{matrix}$$

بنابراین، متوسط محصولات درجه ۱، خواهد شد:

$$\bar{X} = \frac{۶۰}{۲۰۰} = ۰/۳۰$$

برای محاسبه واریانس، از فرمول ۴ استفاده می کنیم:

$$V(X) = f_A \cdot f_{\bar{A}} \\ = ۰/۳۰ \times ۰/۷ = ۰/۲۱$$

کمترین مقدار واریانس صفت کیفی، صفر و بیشترین مقدار آن ۰/۲۵ خواهد بود. غالباً برای پی بردن به همگنی جامعه‌های با صفات کیفی، از واریانس آنها استفاده می شود. اگر مقدار واریانس صفت کیفی، برابر صفر باشد، گویند جامعه از همگنی کامل برخوردار است و اگر مقدار واریانس صفت کیفی برابر ۰/۲۵ باشد، گویند جامعه از ناهمگنی کامل برخوردار است. بنابراین هرچه مقدار واریانس صفت کیفی به صفر نزدیکتر باشد، نشانه همگنی بیشتر در جامعه است و هرچه مقدار واریانس صفت کیفی به ۰/۲۵ نزدیکتر باشد، نشانه ناهمگنی بیشتر در جامعه خواهد بود.

معمولاً در مقایسه همگنی جامعه‌ها، با استفاده از واریانس آنها، می توان نسبت به همگنی آنها قضاوت نمود.

مثال ۴: دو گروه محصول یکسان را برحسب مرغوبیت، دسته بندی نموده ایم و نتایج آن، به صورت زیر به دست آمده است:

I

کیفیت محصول	مرغوب \bar{A}	نامرغوب A
f_i فراوانی نسبی	۰/۷۵	۰/۲۵

II

کیفیت محصول	مرغوب \bar{A}	نامرغوب A
f_i فراوانی نسبی	۰/۸۲	۰/۱۸

تعیین کنید، کدام گروه، همگن تر است؟

برای پاسخ این سؤال، لازم است واریانس هریک از گروه‌ها را محاسبه کنیم، هر کدام واریانس کمتری نسبت به دیگری داشته باشد، آن گروه محصول همگنی بیشتری دارد.

$$V(X_1) = f_A \cdot f_{\bar{A}} = 0.75 \times 0.25 = 0.1875$$

$$V(X_2) = f_A \cdot f_{\bar{A}} = 0.82 \times 0.18 = 0.1476$$

ملاحظه می‌شود که واریانس گروه دوم کمتر از واریانس گروه اول شده، بنابراین نتیجه می‌گیریم که گروه دوم از نظر کیفیت همگن تر است.

گاهی اوقات، توزیع صفت کیفی برحسب فراوانیهای مطلق داده می‌شود. برای محاسبه واریانس، لازم است، ابتدا، فراوانیهای نسبی را برای حالتیهای صفت محاسبه نموده، سپس با استفاده از فراوانیهای نسبی، واریانس را محاسبه کنیم.

مثال ۵: جمعیت دو روستا برحسب وضع سواد، به صورت جداول زیر گروه‌بندی شده‌اند:

روستای A

وضع سواد	باسواد	بی‌سواد	جمع
فراوانی	۲۴۹	۵۱	۳۰۰

روستای B

وضع سواد	باسواد	بی‌سواد	جمع
فراوانی	۳۲۴	۷۶	۴۰۰

کدام یک از دو روستای فوق از نظر سواد همگن تر هستند؟

برای تعیین چگونگی همگنی دو روستای فوق از نظر وضع سواد، می‌باید واریانس هریک از آنها را محاسبه کرد. واریانس صفت کیفی، فقط با فراوانیهای نسبی می‌تواند محاسبه شود، از این رو، ابتدا فراوانیهای نسبی را برای هر دو روستا محاسبه نموده، سپس با استفاده از آنها، واریانس را محاسبه می‌کنیم.

روستای A

وضع سواد	باسواد	بی سواد	جمع
F_i	۲۴۹	۵۱	۳۰۰
f_i	۰/۸۳	۰/۱۷	۱

روستای B

وضع سواد	باسواد	بی سواد	جمع
F_i	۳۲۴	۷۶	۴۰۰
f_i	۰/۸۱	۰/۱۹	۱

$$V(X)_A = 0/83 \times 0/17 = 0/1411 \quad V(X)_B = 0/81 \times 0/19 = 0/1539$$

از این رو، می توان ادعا نمود که روستای A از نظر وضع سواد، همگن تر از روستای B است، چون واریانس کمتری از روستای B دارد.

مطالعه جامعه از نظر دو صفت کیفی A و B

حال، جامعه ای را در نظر می گیریم که عناصر آن از نظر دو صفت A و B که هر یک با حالات متضاد مشاهده شده اند، مورد مطالعه قرار گرفته اند.

در این وضعیّت ممکن است یکی از چهار حالت زیر وجود داشته باشد :

– برای عنصر جامعه، هم صفت A و هم صفت B مشاهده شده است. که این حالت ترکیبی را، برای عنصر مشاهده شده با علامت AB نشان می دهیم.

– برای عنصر جامعه، صفت A مشاهده نشده ولی صفت B مشاهده شده است. این حالت ترکیبی را با نماد \overline{AB} نشان می دهیم.

– برای عنصر جامعه صفت A مشاهده شده ولی صفت B مشاهده نشده است. این حالت ترکیبی را با نماد $A\overline{B}$ نشان می دهیم.

– برای عنصر جامعه، صفت A و صفت B مشاهده نشده است، که این حالت ترکیبی را با نماد $\overline{A}\overline{B}$ نشان می دهیم.

گروه بندی را که نسبت به دو صفت باهم انجام می گیرد، «گروه بندی مرتبه دوم» خواهیم نامید. گروه بندی جامعه بر حسب یک صفت متغیر را «گروه بندی مرتبه اول» می نامند. حال اگر تعداد عناصر (فراوانیهای) چهار گروه نامبرده را با F_{AB} ، $F_{\overline{A}\overline{B}}$ ، $F_{A\overline{B}}$ و $F_{\overline{A}B}$ نشان دهیم، نتیجه این گروه بندی را می توان به صورت جدول چهارخانه ای زیر بیان کرد :

جدول ۹

$\begin{matrix} \text{B} \\ \text{A} \end{matrix}$	B	\bar{B}	جمع
A	F_{AB}	$F_{A\bar{B}}$	F_A
\bar{A}	$F_{\bar{A}B}$	$F_{\bar{A}\bar{B}}$	$F_{\bar{A}}$
	F_B	$F_{\bar{B}}$	N

تمامی فراوانیها را در جدول ۹ برحجم جامعه (N) تقسیم کرده، جدول چهارخانه‌ای مشابه را با فراوانیهای نسبی به دست می‌آوریم (جدول ۱۰):

جدول ۱۰

$\begin{matrix} \text{B} \\ \text{A} \end{matrix}$	B	\bar{B}	جمع
A	f_{AB}	$f_{A\bar{B}}$	f_A
\bar{A}	$f_{\bar{A}B}$	$f_{\bar{A}\bar{B}}$	$f_{\bar{A}}$
	f_B	$f_{\bar{B}}$	۱

مثال ۶: فرض کنیم جامعه مورد مطالعه، ۱۰۰ کارگر یک مؤسسه تولیدی است که از نظر داشتن سواد (صفت A) و داشتن بچه (صفت B) مورد مشاهده قرار گرفته‌اند. پس از تعیین داشتن سواد و داشتن بچه برای هر کارگر، نتایج مشاهدات به ۴ گروه طبقه‌بندی گردید:

AB	باسواد و با بچه	گروه اول
$A\bar{B}$	باسواد و بدون بچه	گروه دوم
$\bar{A}B$	بی سواد و با بچه	گروه سوم
$\bar{A}\bar{B}$	بی سواد و بدون بچه	گروه چهارم

نتایج مشاهدات به صورت جدول چهارخانه‌ای زیر بیان می‌شود (جدول ۱۱):

جدول ۱۱

	بچه دار B	بدون بچه \bar{B}	جمع
باسواد A	۳۵	۲۵	۶۰
بی سواد \bar{A}	۳۰	۱۰	۴۰
جمع	۶۵	۳۵	۱۰۰

جدول بالا، توزیع فراوانیهای کارگران را بر حسب دو صفت A و B بیان می کند. اگر فراوانیهای مطلق هر گروه را بر حجم جامعه ($N=100$) تقسیم نماییم، جدول توزیع فراوانیهای نسبی کارگران بر حسب دو صفت A و B به دست می آید که تشکیل آن را به عنوان تمرین، به دانش آموزان واگذار می کنیم.

به همین شیوه که با دو صفت با حالات متضاد، گروه بندی مرتبه دوم انجام می گرفت، می توان گروه بندیهای مراتب بالاتر را نیز بر حسب ۳ و بیش از ۳ صفت تشکیل داد. در حالت دو صفت A و B که هریک از آنها با بیش از دو حالت ($A_1, A_2, A_3, \dots, A_S$) برای صفت A و ($B_1, B_2, B_3, \dots, B_t$) برای صفت B مشاهده می شوند، در نتیجه گروه بندی مرتبه اول، S گروه برای صفت A و t گروه برای صفت B، تشکیل خواهد شد. نتایج این گروه بندیها با جدول دوبعدی (یا دوطرفه)، که هریک از بعدها را یکی از صفتهای متغیر کیفی، تشکیل می دهند، بیان خواهد شد (جدول ۱۲):

جدول ۱۲

A \ B	B					جمع
	B_1	B_2	B_3	...	B_t	
A_1	F_{11}	F_{12}	F_{13}	...	F_{1t}	F_{A_1}
A_2	F_{21}	F_{22}	F_{23}	...	F_{2t}	F_{A_2}
\vdots	\vdots	\vdots	\vdots		\vdots	\vdots
A_S	F_{S1}	F_{S2}	F_{S3}	...	F_{St}	F_{A_S}
جمع	F_{B_1}	F_{B_2}	F_{B_3}	...	F_{Bt}	N

این جدول، صورت تعمیم یافته جدول چهارخانه ای ۹ می باشد و نام «جدول توافق» به خود گرفته است.

مثال ۷: جامعه مورد مطالعه کارگران یک مؤسسه تولیدی با حجم $N = 50$ هستند که برحسب درجه مهارت (A) و میزان تحصیلات (B) مورد مشاهده قرار می گیرند.

درجه مهارت (A) می تواند به ۴ حالت :

A_1 - کارگر ساده

A_2 - کارگر نیمه ماهر

A_3 - کارگر ماهر

A_4 - استادکار

و میزان تحصیلات می تواند به ۴ حالت :

B_1 - بی سواد

B_2 - تحصیلات ابتدایی

B_3 - تحصیلات متوسطه

B_4 - تحصیلات دیپلم

در نظر گرفته شود. عناصر جامعه (کارگران) برحسب دو صفت متغیر A و B مورد مشاهده قرار گرفته اند و نتایج به صورت جدول زیر طبق حالت های فوق گروه بندی گردیده است :

جدول ۱۳

A \ B	B				جمع
	بی سواد	ابتدایی	متوسطه	دیپلم	
کارگر ساده	۶	۵	۱	۰	۱۲
کارگر نیمه ماهر	۳	۵	۳	۱	۱۲
کارگر ماهر	۲	۴	۷	۳	۱۶
استادکار	۰	۲	۵	۳	۱۰
جمع	۱۱	۱۶	۱۶	۷	۵۰

جدول بالا، توزیع فراوانی های کارگران مؤسسه مذکور را برحسب دو صفت A و B نشان می دهد که جدول توافق دو صفت متغیر A و B می باشد.

بیان هندسی صفات کیفی

برای نمایش هندسی یا نمودار آماری صفات کیفی، از نمودار دایره‌ای یا قطاعی استفاده می‌شود.

– نمودار دایره‌ای یا قطاعی، شکل دیگری از نمودارهای آماری است که برای نمایش سهم

هریک از حالت‌های صفت کیفی در جامعه، به کار می‌رود.

برای رسم نمودار دایره‌ای، یک دایره با شعاع دلخواه در نظر گرفته، فراوانی یا درصد فراوانی

هریک از حالات صفت را بر حسب درجه، مطابق فرمول زیر محاسبه می‌کنیم و آن را به صورت کمانی

از دایره، رسم می‌نماییم:

$$\frac{F_i \times 360}{N}$$

مثال ۸: نمودار دایره‌ای را برای هزینه خانوارها که از هزینه‌های مختلف مانند: هزینه خوراک،

هزینه پوشاک، هزینه مسکن، هزینه بهداشت و درمان، هزینه آموزش و هزینه‌های متفرقه، تشکیل شده

است و ارقام آن در جدول زیر داده شده، رسم می‌کنیم: (ارقام هزینه، متوسط هزینه‌های خانوارهای

یک منطقه می‌باشند).

هزینه‌های متفرقه	هزینه آموزش	هزینه بهداشت و درمان	هزینه مسکن	هزینه پوشاک	هزینه خوراک	هزینه‌های خانوار
۲۰	۷	۵	۳۰	۱۰	۲۸	مبلغ بر حسب هزار تومان

ابتدا، هزینه‌ها را بر حسب درجه، محاسبه می‌کنیم. در اینجا $N = 100$ می‌باشد.

$$\frac{28 \times 360}{100} = 100.8^\circ$$

$$\frac{10 \times 360}{100} = 36^\circ$$

$$\frac{30 \times 360}{100} = 108^\circ$$

$$\frac{5 \times 360}{100} = 18^\circ$$

$$\frac{7 \times 360}{100} = 25.2^\circ$$

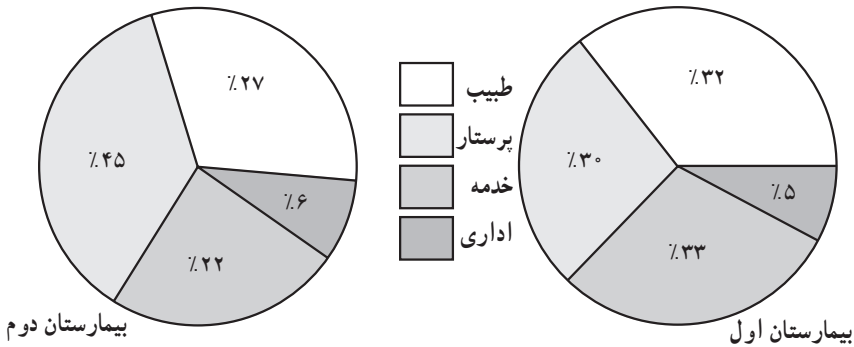
$$\frac{20 \times 360}{100} = 72^\circ$$

باید توجه نمود که جمع درجه‌ها بایستی 360° باشد.



شکل ۱- نمودار دایره‌ای هزینه خانوارها

مناسب است، بخشهای دایره را با رنگهای مختلف و یا با هاشورهای متفاوت از هم تفکیک کرد. از نمودار دایره‌ای برای مقایسه ساختار دو جامعه (یا بیشتر) نیز استفاده می‌شود. به عنوان مثال، ترکیب خدمه دو بیمارستان را به منظور مقایسه، به صورت نمودار قطاعی می‌آوریم (شکل ۵-۲):



از نمودار قطاعی برای بیان هندسی صفات متغیر کمی نیز استفاده می‌شود. همین‌طور می‌توان از نمودارهای دیگری مانند نمودار ستونی برای بیان هندسی صفت متغیر کیفی نیز استفاده نمود.

سؤالها و تمرینها ؟

- ۱- نتایج مشاهدات روی صفات کیفی، چگونه بیان می شود؟
- ۲- برای گروه بندی نتایج مشاهدات صفات کیفی، چگونه عمل می کنند؟
- ۳- اگر صفت کیفی در جامعه، حالت‌های مختلفی را اختیار کند، جدول توزیع صفت متغیر، به چه صورت بیان می شود؟
- ۴- برای تبدیل صفات کیفی به صفات کمی، از چه شیوه‌ای استفاده می شود؟
- ۵- مشخص کننده‌های عددی صفت کیفی، چگونه محاسبه می شوند؟
- ۶- میانگین صفت کیفی برابر با چه مقداری است؟
- ۷- واریانس صفت کیفی از چه رابطه‌ای محاسبه می شود؟
- ۸- کمترین مقدار واریانس صفت کیفی، چیست؟
- ۹- بیشترین مقدار واریانس صفت کیفی، چیست؟
- ۱۰- همگنی کامل، چه زمانی روی می دهد؟
- ۱۱- ناهمگنی کامل، چه زمانی روی می دهد؟
- ۱۲- برای مقایسه جامعه‌ها از نظر صفات کیفی، از چه مفهومی استفاده می کنند؟
- ۱۳- توزیع گروه خونی جمعیت دو منطقه، به صورت زیر به دست آمده است :

منطقه I

گروه خونی	A	B	AB	O
فراوانی نسبی	۰/۲۰	۰/۱۵	۰/۲۵	۰/۴۰

منطقه II

گروه خونی	A	B	AB	O
فراوانی نسبی	۰/۱۸	۰/۲۲	۰/۳۰	۰/۳۰

- کدام یک از دو منطقه مذکور، از نظر گروه خونی O، همگن ترند؟
- ۱۴- اگر دو صفت کیفی با حالات متضاد، در یک جامعه برای عناصر آن مشاهده شده باشد، نتایج آن مشاهدات، چگونه بیان می شود؟
 - ۱۵- این گروه بندی را چه می نامند؟
 - ۱۶- جدول چهارخانه‌ای برای چه منظوری تشکیل می شود؟
 - ۱۷- اگر دو صفت کیفی A و B، که هریک از آنها برای عناصر جامعه با بیش از دو حالت مشاهده شوند، نتیجه چنین مشاهداتی، چگونه گروه بندی می گردند؟

۱۸- در یک استان، ۴ نوع نوشابه تولید می‌شود، که با علامتهای A_1, A_2, A_3 و A_4 مشخص شده‌اند. از ۴۰ دانش‌آموز پسر و ۴۰ دانش‌آموز دختر نظرخواهی شد که چه نوع نوشابه‌ای را می‌پسندند، نتایج به‌قرار زیر به‌دست آمده است:

پسران	دختران
$A_1 = 12$	$A_1 = 10$
$A_2 = 10$	$A_2 = 14$
$A_3 = 15$	$A_3 = 8$
$A_4 = 3$	$A_4 = 8$

نتایج را در یک جدول دو بُعدی، گروه‌بندی نموده، جدول فراوانیهای نسبی را تشکیل دهید.
۱۹- درصد انواع محصولات کشاورزی تولید شده در یک استان به‌قرار زیر است:

حبوبات	ذرت	جو	گندم	نوع محصول
۱۶	۱۷	۲۲	۴۵	درصد

برای بیان ترکیب محصول تولید شده در این استان، نمودار قطاعی رسم کنید.
۲۰- درصد وامهای پرداختی بانک A در سالهای ۱۳۷۳ و ۱۳۷۵ به‌صورت زیر بوده است (ارقام ساختگی است):

نوع وام	سال ۱۳۷۳	سال ۱۳۷۵
ساختمان	۳۱	۲۴
صنعت	۳۰	۲۸
کشاورزی	۲۰	۲۸
خدمات	۱۹	۲۰

برای مقایسه وامهای پرداختی در سالهای ۱۳۷۳ و ۱۳۷۵ نمودار دایره‌ای رسم کنید.

فصل هشتم

توزیع مشترک دو صفت متغیر

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند:

- ۱- جامعه را برحسب دو صفت متغیر کمی گروه‌بندی نماید.
- ۲- جدول توزیع فراوانی‌های دو صفت متغیر را تشکیل دهد.
- ۳- توزیعهای حاشیه‌ای x و y را به‌دست آورد.
- ۴- توزیعهای شرطی y برحسب x و x برحسب y را به‌دست آورد.
- ۵- مشخصه‌های عددی برای دو صفت متغیر را به تفکیک محاسبه نماید.
- ۶- مفهوم پراکندگی توأم دو صفت متغیر را بیان نماید.

در فصلهای گذشته، راجع به مطالعه جامعه‌ها با یک صفت متغیر بحث کردیم. ولی اغلب برای یک محقق نیاز می‌شود که اعضاء جامعه را توسط چند صفت متغیر با هم مورد مطالعه قرار دهد. در این فصل فقط به مطالعه دو صفت متغیر از اعضاء جامعه می‌پردازیم.

توزیع مشترک دو صفت متغیر x و y

اگر برای اعضاء جامعه دو صفت متغیر x و y با هم اندازه‌گیری شوند، آنگاه برای هریک از اعضاء یک زوج اندازه یا مقدار به‌دست خواهد آمد. مجموع زوجهای به‌دست آمده از نتایج مشاهدات مقادیر دو صفت را می‌توان به‌صورت زیر نشان داد.

شماره اعضاء	۱	۲	۳	n
اندازه صفت x	x_1	x_2	x_3	x_n
اندازه صفت y	y_1	y_2	y_3	y_n

چنانچه حجم جامعه بزرگ باشد، به‌طوری که هر زوج از اندازه‌های x و y بیش از یک بار

تکرار شوند، لزوماً نتایج مشاهدات را در جدول توزیع فراوانی دو بعدی وارد می‌کنند که در آن فراوانیها (F) با دو اندیس i برای صفت x و j برای صفت y مشخص می‌گردند، یعنی به صورت F_{ij} نشان داده می‌شوند. مجموعه دو صفت متغیر را دستگاه دو متغیر نیز می‌نامند. جدول ۱ توزیع فراوانیها را برای دو صفت متغیر x و y نشان می‌دهد:

جدول ۱

x \ y	y						$F_{i\cdot}$
	y_1	y_2	y_j	y_t	
x_1	F_{11}	F_{12}	F_{1j}	F_{1t}	$F_{1\cdot}$
x_2	F_{21}	F_{22}	F_{2j}	F_{2t}	$F_{2\cdot}$
x_i	F_{i1}	F_{i2}	F_{ij}	F_{it}	$F_{i\cdot}$
x_s	F_{s1}	F_{s2}	F_{sj}	F_{st}	$F_{s\cdot}$
$F_{\cdot j}$	$F_{\cdot 1}$	$F_{\cdot 2}$	$F_{\cdot j}$	$F_{\cdot t}$	n

در این جدول اگر فراوانیهای درون جدول (F_{ij}) را روی سطر جمع کنیم خواهیم داشت:

$$\sum_{j=1}^t F_{ij} \cdot F_{i1} \cdot F_{i2} \cdot \dots \cdot F_{it} \cdot F_{i\cdot}$$

چون فراوانیها روی اندیس j جمع شده‌اند بنابراین جواب حاصل به صورت $F_{i\cdot}$ نوشته خواهد شد. هم‌چنین اگر فراوانیها را روی اندیس i جمع بیندیم جواب حاصل به صورت $F_{\cdot j}$ نوشته خواهد شد.

$$\sum_{i=1}^s F_{ij} \cdot F_{1j} \cdot F_{2j} \cdot \dots \cdot F_{sj} \cdot F_{\cdot j}$$

بنابراین جمع سطرهای جدول به صورت $F_{1\cdot}, F_{2\cdot}, \dots, F_{s\cdot}$ در خواهد آمد و اگر آنها را روی اندیس i جمع کنیم حاصل حجم جامعه یعنی n را خواهد داد. و اگر $F_{\cdot j}$ ها را روی j جمع کنیم نتیجه برابر حجم جامعه یعنی n خواهد شد.

$$\begin{matrix} s \\ i, 1 \end{matrix} : F_{i0} \cdot F_{i1} \cdot F_{i2} \cdot \dots \cdot F_{is} \cdot n$$

$$\begin{matrix} t \\ j, 1 \end{matrix} : F_{j0} \cdot F_{j1} \cdot F_{j2} \cdot \dots \cdot F_{jt} \cdot n$$

مثال ۱: فرض کنید از جامعه دانش‌آموزان یک دبیرستان دو صفت متغیر x و y ، که در آن x سن دانش‌آموزان و y تعداد افراد خانوار دانش‌آموزان، پرسش گردیده نتایج به صورت جدول توزیع فراوانیها (جدول ۲) به دست آمده باشد.

جدول ۲

جمع سطر	تعداد افراد خانوار y					x سن
	۲	۳	۴	۵	۶	
۳۰	۳	۴	۶	۱۲	۵	۱۴
۳۵	۲	۷	۱۰	۱۱	۵	۱۵
۳۲	۴	۵	۹	۱۰	۴	۱۶
۳۷	۶	۴	۱۲	۱۰	۵	۱۷
۲۸	۳	۳	۷	۸	۷	۱۸
۱۸	۲	۶	۵	۴	۱	۱۹
جمع ستون	۲۰	۲۹	۴۹	۵۵	۲۷	۱۸۰

اعداد داخل جدول نشانه فراوانیهای هر زوج از دو صفت متغیر می‌باشد. برای مثال ۴ دانش‌آموز ۱۴ ساله تعداد نفرات خانوارشان ۳ نفره است و هم چنین ۱۰ دانش‌آموز ۱۵ ساله تعداد نفرات خانوارشان ۴ نفره می‌باشد.

توزیعهای حاشیه‌ای

از این جدول دوبعدی می‌توان توزیع هر یک از صفات متغیر را جداگانه به دست آورد. مثلاً بدون در نظر گرفتن صفت y ، می‌توان توزیع صفت متغیر x را به دست آورد.

جدول ۳

x	x_1	x_2	x_i	x_s	جمع
فراوانی $F_{i.}$	$F_{1.}$	$F_{2.}$	$F_{i.}$	$F_{s.}$	n

برای مثال ۱ توزیع سن دانش‌آموزان را بدون در نظر گرفتن تعداد افراد خانوار تشکیل می‌دهیم :

جدول ۴

x سن	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	جمع
فراوانی $F_{i.}$	۳۰	۳۵	۳۲	۳۷	۲۸	۱۸	۱۸۰

چون فراوانیها برای مقادیر صفت x (سن) از ستون کناری جدول قرار داده شده، به چنین توزیعی، توزیع کناری یا توزیع حاشیه‌ای x گویند. این توزیع نشان می‌دهد که توزیع سن دانش‌آموزان این دبیرستان چگونه است. همین‌طور می‌توان توزیع صفت y را به‌طور جداگانه از روی جدول دوبعدی ۱ به‌دست آورد :

جدول ۵

y	y_1	y_2	y_j	y_t	جمع
فراوانی $F_{.j}$	$F_{.1}$	$F_{.2}$	$F_{.j}$	$F_{.t}$	n

چنین توزیعی را توزیع حاشیه‌ای y نامند. چون در برابر مقادیر صفت y فراوانیها را از سطر پایین جدول ۱ قرار داده‌ایم.

توزیع حاشیه‌ای y را برای مثال ۱ به‌صورت جدول (۶) زیر می‌نویسیم :

جدول ۶

y (تعداد افراد خانوار)	۲	۳	۴	۵	۶	جمع
فراوانی $F_{.j}$	۲۰	۲۹	۴۹	۵۵	۲۷	۱۸۰

این توزیع به ما نشان می‌دهد که دانش‌آموزان این دبیرستان از نظر تعداد افراد خانوارشان چگونه توزیع شده‌اند.

توزیعهای شرطی

هرگاه بخواهیم برای اندازه معینی از صفت y ، توزیع صفت x را بدانیم و یا برعکس برای اندازه یک x معین، توزیع صفت y را به دست آوریم. می‌توان با قرار دادن مقادیر صفت x در جدول و فراوانیهای متناظر با y معین را در برابر آنها، توزیع صفت x برحسب y معین را به دست آورد و آن را به صورت $X|Y$ نوشت و برای توزیع y برحسب x معین آن را به صورت $Y|X$ می‌نویسیم. به چنین توزیعهایی، توزیع شرطی x برحسب y و توزیع شرطی y برحسب x گویند. مثلاً اگر بخواهیم توزیع سن دانش‌آموزان را برای خانوارهای ۴ نفره به دست آوریم (با توجه به جدول ۲) خواهیم داشت:

جدول ۷

$X y$. ۴	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	جمع
فراوانی F_{i4}	۶	۱۰	۹	۱۲	۷	۵	۴۹

ملاحظه می‌شود که فراوانیها از ستون ۴. y در برابر مقادیر صفت x نوشته شده است. هم‌چنین اگر توزیع نفرات خانوار برای دانش‌آموزان ۱۷ ساله موردنظر باشد، می‌توان آن را به صورت زیر به دست آورد:

جدول ۸

$Y x$. ۱۷	۲	۳	۴	۵	۶	جمع
فراوانی F_{17j}	۶	۴	۱۲	۱۰	۵	۲۷

در اینجا نیز فراوانیها از سطر ۱۷. x در برابر مقادیر y نوشته شده است. از توزیعهای شرطی برای پی بردن به وجود بستگی بین دو صفت متغیر x و y استفاده می‌شود.

مشخصه‌های عددی برای توزیع مشترک دو صفت متغیر x و y

برای محاسبه میانگین صفت متغیر x و میانگین صفت متغیر y از توزیعهای حاشیه‌ای x و y استفاده نموده و مطابق طریقه محاسبه میانگین حسابی، می‌توان آنها را به دست آورد:

فرمولهای محاسباتی میانگینها به قرار زیر است:

$$\bar{X} = \frac{\sum_{i=1}^s F_{i\cdot} X_i}{n} \quad (1)$$

$$\bar{Y} = \frac{\sum_{j=1}^t F_{\cdot j} Y_j}{n} \quad (2)$$

برای مثال ۱، میانگین صفت متغیر x (سن دانش‌آموزان) و میانگین صفت متغیر y (تعداد افراد خانوار) را محاسبه می‌کنیم:

$$\bar{X} = \frac{\sum_{i=1}^s F_{i\cdot} X_i}{n} = \frac{30 \cdot 14 + 35 \cdot 15 + 32 \cdot 16 + 37 \cdot 17 + 28 \cdot 18 + 18 \cdot 19}{180}$$

$$= \frac{2932}{180} = 16/29$$

به عبارت بهتر میانگین سن دانش‌آموزان ۱۶/۲۹ سال می‌باشد.

$$\bar{Y} = \frac{\sum_{j=1}^t F_{\cdot j} Y_j}{n} = \frac{20 \cdot 2 + 29 \cdot 3 + 49 \cdot 4 + 55 \cdot 5 + 27 \cdot 6}{180}$$

$$= \frac{760}{180} = 4/22$$

یعنی متوسط تعداد افراد خانوار دانش‌آموزان ۴/۲۲ نفر می‌باشد.

هم‌چنین می‌توان پراکندگی (واریانس) هر یک از صفات متغیر x و y را توسط روابط زیر محاسبه نمود:

$$V(X) = \frac{\sum_{i=1}^s F_{i\cdot} X_i^2}{n} - \bar{X}^2 \quad (3)$$

$$V(Y) = \frac{\sum_{j=1}^t F_{\cdot j} Y_j^2}{n} - \bar{Y}^2 \quad (4)$$

برای مثال ۱ واریانس صفت متغیر X و واریانس صفت متغیر Y را محاسبه می‌کنیم:

$$V(X) = \frac{\sum_{i=1}^s F_i X_i^2}{n} - \bar{X}^2$$

$$= \frac{30 \cdot 14^2 + 35 \cdot 15^2 + 32 \cdot 16^2 + 37 \cdot 17^2 + 28 \cdot 18^2 + 18 \cdot 19^2}{180} - (16/29)^2$$

$$= \frac{48210}{180} - 265/36 = 267/18 - 265/36 = 2/47$$

$$V(Y) = \frac{\sum_{j=1}^t F_j Y_j^2}{n} - \bar{Y}^2$$

$$= \frac{20 \cdot 2^2 + 29 \cdot 3^2 + 49 \cdot 4^2 + 55 \cdot 5^2 + 27 \cdot 6^2}{180} - (4/22)^2$$

$$= \frac{3472}{180} - 17/81 = 19/29 - 17/81 = 1/48$$

علاوه بر پراکندگی صفات متغیر X و Y، پراکندگی توأم دو صفت متغیر X و Y را نیز محاسبه می‌کنند که آنرا کواریانس X و Y نامند و به صورت Cov(x,y) نوشته می‌شود. فرمول محاسبه کواریانس X و Y به قرار زیر است:

$$\text{Cov}(x,y) = \frac{\sum_{i,j} F_{ij} X_i Y_j}{n} - \bar{X} \cdot \bar{Y} \quad (5)$$

از کواریانس X و Y برای آگاهی از همبستگی بین دو صفت X و Y استفاده می‌کنند. هرگاه مقدار کواریانس X و Y برابر صفر شود یعنی بین دو صفت X و Y همبستگی خطی وجود ندارد و هرگاه مقدار کواریانس X و Y بزرگتر از صفر (مثبت) باشد یعنی بین دو صفت X و Y همبستگی به طور مستقیم وجود دارد. و چنانچه کوچکتر از صفر (منفی) باشد یعنی بین دو صفت X و Y همبستگی به طور معکوس وجود دارد.

برای مثال ۱ کواریانس X و Y را محاسبه می‌کنیم:

$$\text{Cov}(x,y) = \frac{\sum_{i,j} F_{ij} X_i Y_j}{n} - \bar{X} \bar{Y}$$

$$= \frac{3 \cdot 14 \cdot 2 + 4 \cdot 14 \cdot 3 + 6 \cdot 14 \cdot 4 + 4 \cdot 12 \cdot 14 + 5 \cdot 5 \cdot 14 + 6 \cdot 2 \cdot 15 + 2 \cdot 2 \cdot 15}{180} - 16/29 \cdot 4/22$$

$\underline{7. 15. 3. 10. 15. 4. 11. 15. 5. 5. 11. 6. 4. 16. 2. 5. 16. 3.}$
 $\underline{9. 16. 4. 10. 16. 5. 4. 16. 6. 6. 17. 2. 4. 17. 3. 12. 17. 4.}$
 $\underline{10. 17. 5. 5. 17. 6. 3. 18. 2. 3. 18. 3. 7. 18. 4. 8. 18. 5.}$
 $\underline{. 7. 18. 6. 2. 19. 2. 6. 19. 3. 5. 19. 4. 4. 19. 5. 1. 19. 6.}$
 180

$$\frac{12352}{180} \cdot 16/29. 4/22. 68/62. 68/74 \dots 0/12$$

برای سهولت محاسبات می‌توان کلیه محاسبات را توسط جدول زیر انجام داد: (جدول ۹)

جدول ۹

$\begin{matrix} y \\ \backslash \\ x \end{matrix}$	۲	۳	۴	۵	۶	$F_{i.}$	$F_{i.}X_i$	$F_{i.}X_i^2$	$F_{.j}Y_j$	$X_i \cdot F_{ij}Y_j$
۱۴	۳	۴	۶	۱۲	۵	۳۰	۴۲۰	۵۸۸۰	۱۳۲	۱۸۴۸
۱۵	۲	۷	۱۰	۱۱	۵	۳۵	۵۲۵	۷۸۷۵	۱۵۰	۲۲۵۰
۱۶	۴	۵	۹	۱۰	۴	۳۲	۵۱۲	۸۱۹۲	۱۳۳	۲۱۲۸
۱۷	۶	۴	۱۲	۱۰	۵	۳۷	۶۲۹	۱۰۶۹۳	۱۵۲	۲۵۸۴
۱۸	۳	۳	۷	۸	۷	۲۸	۵۰۴	۹۰۷۲	۱۲۵	۲۲۵۰
۱۹	۲	۶	۵	۴	۱	۱۸	۳۲۲	۶۴۹۸	۶۸	۱۲۹۲
$F_{.j}$	۲۰	۲۹	۴۹	۵۵	۲۷	۱۸۰	۲۹۳۲	۴۸۲۱۰	۷۶۰	۱۲۳۵۲
$F_{.j}Y_j$	۴۰	۸۷	۱۹۶	۲۷۵	۱۶۲	۷۶۰	
$F_{.j}Y_j^2$	۸۰	۲۶۱	۷۸۴	۱۳۷۵	۹۷۲	۳۴۷۲				

محاسبه اندازه همبستگی بین دو صفت متغیر x و y و تفسیر آنرا در جلد دوم کتاب روشهای

آماري از نظر خواهید گذراند.

سؤالها و تمرینها ؟

- ۱- توزیع مشترک دو صفت متغیر چیست؟
- ۲- هدف از مطالعه جامعه‌ها با بیش از یک صفت متغیر چیست؟
- ۳- فراوانی گروه z نام چیست؟
- ۴- توزیع حاشیه‌ای صفت متغیر x چگونه به دست می‌آید؟
- ۵- توزیع حاشیه‌ای صفت متغیر y چگونه به دست می‌آید؟
- ۶- توزیع شرطی صفت متغیر x بر حسب y چه مفهومی دارد؟
- ۷- توزیع شرطی صفت متغیر y بر حسب x چه مفهومی دارد؟
- ۸- از توزیع‌های شرطی برای چه منظوری استفاده می‌شود؟
- ۹- در توزیع مشترک دو صفت متغیر، میانگین حسابی دو صفت چگونه محاسبه می‌شود؟
- ۱۰- در توزیع مشترک دو صفت متغیر، پراکندگی صفات چگونه محاسبه می‌شود؟
- ۱۱- کواریانس دو صفت متغیر چیست؟ فرمول آن را بنویسید.
- ۱۲- از کواریانس برای آگاهی از چه خصوصیتی از دو صفت متغیر استفاده می‌کنند؟
- ۱۳- هرگاه کواریانس دو صفت متغیر برابر صفر گردد چه مفهومی دارد؟
- ۱۴- نتایج مشاهدات دو صفت x و y در جدول زیر آمده است :

x	۸	۱۰	۱۱	۱۲	۱۴
y	۱۰	۶	۷	۳	۵

- میانگین x و میانگین y را محاسبه کنید.
 واریانس x و واریانس y را محاسبه کنید.
 کواریانس x و y را محاسبه کنید.

۱۵- توزیع مشترک دو صفت متغیر x و y به شرح جدول زیر است :

$y \backslash x$	۳	۴	۵	۶	$F_{i.}$
۵	۲	۳	۳	۲	۱۰
۱۰	۳	۱۰	۱۰	۲	۲۵
۱۵	۱	۲	۷	۵	۱۵
$F_{.j}$	۶	۱۵	۲۰	۹	۵۰

مطلوبست :

الف - توزیع حاشیه‌ای x و توزیع حاشیه‌ای y

ب - توزیع شرطی y بر حسب $x = ۱۰$

ج - توزیع شرطی x بر حسب $y = ۵$

د - میانگین x و میانگین y

هـ - واریانس x و واریانس y

و - کواریانس x و y و تفسیر آن

منابع و مأخذ

- ۱- مفاهیم اساسی آمار تألیف علی مدنی - انتشارات فروردین.
- ۲- آنالیز آماری (جلد اول) تألیف علی مدنی.
- ۳- آمار توصیفی تألیف ژرار کالت - ترجمه حسن صادقی از انتشارات دانشگاه فردوسی

مشهد.

- ۴- آمار و کاربرد آن در مدیریت تألیف مهدی صفاری - نشر دانا.
- ۵- آمار مقدماتی (جلد اول) تألیف وونا کات - ترجمه دکتر مشکانی.
- ۶- Statistical Methods Snedecor & Cochran.
- ۷- Statistical Analysis for Managerial Decisions Boot and Cox.
- ۸- Introduction to Mathematical Statistics Paul . G . Hoel.

