آمار و احتمال

سری کتاب‌های کمک آموزشی کارشناسی ارشد و دکتری

مجموعه مدیریت،اقتصاد و حسابداری

مؤلف: دکترحسن رضاپور

ویرایش جدید

سرشناسه	: رضاپور، حسن
عنوان	: آمار و احتمال
مشخصات نشر	: تهران : مشاوران صعود ماهان، ۱۴۰۰
مشخصات ظاهری	: ۴۶۶ص
فروست	: سری کتاب های کمک آموزشی کارشناسی ارشدودکتری
شابک	: 978-600-389-112-8
وضعیت فهرست نویسی	: فیپای مختصر
یادداشت	: این مدرک در آدرس http://opac.nlai.ir قابل دسترسی است.
شماره کتابشناسی ملی	: ۳۸۸۹۳۱۲

نام کتاب: ....آمار و احتمال
مدیران مسئول: .... هادی سیاری، مجید سیاری
مولف:.. دکتر حسن رضاپور
مدیر تولید و برنامه‌ریزی .... سمیه بیگی
ناشر: ....مشاوران صعود ماهان
نوبت و تاریخ چاپ .... اول/ ۱۴۰۰
تیراژ: .... ۱۰۰۰ جلد
قیمت: ....۱/۷۹۰/۰۰۰ ریال
شابک .... ISBN: ۹۷۸-۶۰۰-۳۸۹-۱۱۲-۸

انتشارات مشاوران صعود ماهان: خیابان ولیعصر، بالاتر از تقــاطع مطهری،
روبروی قنادی هتل بزرگ تهـــران، جنب بانک ملی، پلاک ۲۰۵۰
تلفن: ۴- ۸۸۱۰۰۱۱۳

سخن ناشر

«ن والقلم و ما یسطرون»

کلمه نزد خدا بود و خدا آن را با قلم بر ما نازل کرد.

به پاس تشکر از چنین موهبت الهی، موسسه ماهان درصدد برآمده است تا در راستای انتقال دانش و مفاهیم با کمک اساتید مجرب و مجموعه کتب آموزشی خود برای شما داوطلبان ادامه تحصیل در مقطع کارشناسی ارشد گام موثری بردارد. امید است تلاش‌های خدمتگزاران شما در این موسسه پایه‌گذار گام‌های بلند فردای شما باشد.

مجموعه کتاب‌های کمک آموزشی ماهان به‌منظور استفاده داوطلبان کنکور کارشناسی ارشد سراسری و آزاد تالیف شده‌اند. در این کتاب‌ها سعی کرده‌ایم با بهره‌گیری از تجربه اساتید بزرگ و کتب معتبر داوطلبان را از مطالعه کتاب‌های متعدد در هر درس بی‌نیاز کنیم.

دیگر تالیفات ماهان برای سایر دانشجویان به‌صورت ذیل می‌باشد.

- **مجموعه کتاب‌های ۸ آزمون:** شامل ۵ مرحله کنکور کارشناسی ارشد ۵ سال اخیر به همراه ۳ مرحله آزمون تالیفی ماهان همراه با پاسخ تشریحی می‌باشد که برای آشنایی با نمونه سوالات کنکور طراحی شده است. این مجموعه کتاب‌ها با توجه به تحلیل ۳ ساله اخیر کنکور و بودجه‌بندی مباحث در هریک از دروس، اطلاعات مناسبی جهت برنامه‌ریزی درسی در اختیار دانشجو قرار می‌دهد.
- **مجموعه کتاب‌های کوچک:** شامل کلیه نکات کاربردی در گرایش‌های مختلف کنکور کارشناسی ارشد می‌باشد که برای دانشجویان جهت جمع‌بندی مباحث در ۲ ماهه آخر قبل از کنکور مفید می‌باشد.

بدین‌وسیله از مجموعه اساتید، مولفان و همکاران محترم خانواده بزرگ ماهان که در تولید و به‌روزرسانی تالیفات ماهان نقش موثری داشته‌اند، صمیمانه تقدیر و تشکر می‌نماییم.

دانشجویان عزیز و اساتید محترم می‌توانند هرگونه انتقاد و پیشنهاد درخصوص تالیفات ماهان را از طریق سایت ماهان به آدرس mahan.ac.ir با ما در میان بگذارند.

موسسه آموزش عالی آزاد ماهان

سخن مؤلف

مجموعه حاضر برای درس آمار و احتمال ویژه کنکورهای کارشناسی ارشد و دکتری رشته‌های مدیریت، حسابداری و اقتصاد و .... تالیف شده‌است. در نحوه تالیف سعی شده است علاوه‌بر شرح درس، شامل نکته‌ها و مثال‌ها، به حل تست‌های تالیفی، کنکورهای سراسری و آزاد نیز پرداخته شود. برای مطالعه و تسلط بر درس، خلاصه‌نویسی فرمول‌های مهم و مرور آنها را فراموش نکنید. فصل‌های آمار توصیفی، آنالیز ترکیبی و احتمال، متغیرهای تصادفی و توزیع‌ها، برآورد فاصله‌ای و آزمون فرض از فصل‌های بسیار مهم کتاب به‌شمار می‌آیند. پس این فصول را با دقت بیشتر مطالعه نمایید.

باعث افتخار است که همه شما عزیزان، نظرات، انتقادات و پیشنهادات خود را در راستای بهبودی این اثر به ایمیل Hassan.Rezapour@gmail.com ارسال نمایید. این اثر را به پدر فداکار و مادر مهربانم، همسر عزیزم و البرز پسر نازنینم تقدیم می‌نمایم.

مؤلف

حسن رضاپور

فهرست

فصل اول

آمار توصیفی

- دسته‌بندی داده‌ها
- نمودارها
- انواع میانگین
- مد
- میانه
- واریانس و انحراف معیار
- گشتاورها
- ضریب چولگی
- ضریب کشیدگی
- انحراف چارکی
- قضیه چی بی شف

آمار توصیفی

۱-۱تعاریف اولیه

آمار توصیفی: به روش‌های علمی برای جمع‌آوری، طبقه‌بندی، خلاصه‌کردن و نمایش داده‌های آماری برای رسیدن به نتایجی روشن و اتخاذ تصمیمات منطقی آمار می‌گوییم. اگر داده‌های آماری کل جامعه را پوشش دهد به این بررسی آمار توصیفی، می‌گوییم، در غیراین‌صورت اگر داده‌های آماری، نمونه‌ای از جامعه باشد به این بررسی آمار استنباطی گوییم.

جامعه آماری: بخشی از (یا کل) جامعه که دارای حداقل یک صفت مشخصه باشند (یعنی همه عناصر دارای صفتی باشند که این جامعه را نسبت به جامعه دیگر متمایز سازد) را جامعه آماری می‌نامیم. اگر تعداد افراد جامعه محدود باشد به آن، جامعه آماری محدود و در غیراین‌صورت به آن جامعه آماری نامحدود می‌گوییم.

انواع صفت

۱- صفت ثابت (مشترک): صفتی که بین همه اعضای جامعه مشترک و ثابت است مثل صفت مذکربودن یا مونث‌بودن.

۲- صفت متغیر: صفتی که افراد یک جامعه را از یکدیگر مشخص می‌سازد و از عضوی به عضو دیگر قابل تغییر است مثل وزن، قد و نمره اعضا در یک درس خاص و ... .

نکته: صفت متغیر به دو دسته کمی و کیفی قابل تقسیم‌بندی است. صفت کمی قابل اندازه‌گیری است مثل قد، نمره و ... و به دو دسته پیوسته (مثل قد، وزن و ...) و گسسته (مثل تعداد ورودی‌های رشته ریاضی در یک دانشگاه خاص) تقسیم می‌شود. صفت کیفی به کیفیت صفت خاص می‌پردازد مثل مرغوبیت کالا یا رنگ چشم اعضای جامعه و ...

مقیاس

برای اندازه‌گیری صفت متغیر در دو دسته کمی و کیفی، مقیاس‌های متفاوتی برای اندازه‌گیری وجود دارد:

۱- اسمی (نامی یا طبقه‌بندی): در این مقیاس تنها به نام‌گذاری و طبقه‌بندی داده پرداخته می‌شود و از اعداد و علائم برای طبقه‌بندی اشیا استفاده می‌شود، مثل شماره‌گذاری لباس اعضای تیم فوتبال.

۲- رتبه‌ای (ترتیبی یا مرتبه‌ای): علاوه بر نام‌گذاری، در این مقیاس به رتبه‌بندی اعضا نیز پرداخته می‌شود مثل استفاده از حالت ترتیبی (بزرگ‌تر، مساوی، کوچک‌تر) و یا استفاده از درجات نظامی.

۳- فاصله‌ای (بازه‌ای): در این مقیاس که همه خصوصیات مقیاس رتبه‌ای را دارد به این نکته دقت می‌شود که بین هر دو عدد، بی‌نهایت عدد وجود دارد یعنی انتخاب اعداد می‌تواند از یک بازه انجام پذیرد. در این مقیاس انتخاب محل صفر قراردادی است مثل صفر سانتی‌گراد یا فارنهایت در درجه حرارت.

۴- مقیاس نسبی (نسبتی): در این مقیاس علاوه‌بر وجود خصوصیات مقیاس‌های قبلی، صفر واقعی وجود دارد و به‌طور ثابت در نظر گرفته می‌شود. مثل مقیاس گرم، اینچ و .... .

۲-۱- دسته‌بندی داده‌ها

داده‌های یک جامعه را می‌توان برحسب نیاز به دو طریق دسته‌بندی کرد: جدول گسسته داده‌ها و یا جدول پیوسته داده‌ها. برای این کار ابتدا به تعاریف زیر احتیاج داریم:

۱- فراوانی مطلق (F_i) : تعداد دفعات تکرار هر داده i را فراوانی مطلق آن می‌نامیم و آن را با F_i نمایش می‌دهیم. واضح است که اگر حجم جامعه برابر N و تعداد داده‌ها k باشند، داریم:
$$\sum_{i=1}^{k} F_i = N$$

۲- فراوانی نسبی (f_i): نسبت فراوانی مطلق هر داده i به حجم جامعه را فراوانی نسبی می‌گوییم یعنی: $f_i = \frac{F_i}{N}$. واضح است مجموع فراوانی نسبی برابر یک است.

$$\sum_{i=1}^{k} f_i = \sum_{i=1}^{k} \frac{F_i}{N} = \frac{\sum_{i=1}^{k} F_i}{N} = \frac{N}{N} = ۱$$

ضمناً از ضرب فراوانی نسبی در صد، برای هر داده درصد فراوانی نسبی حاصل می‌گردد یعنی:

$P_i = f_i \times ۱۰۰ =$ درصد فراوانی نسبی

۳- فراوانی تجمعی (F_{c_i}): جمع فراوانی مطلق هر داده (طبقه) با فراوانی‌های مطلق طبقات ماقبل آن را فراوانی تجمعی می‌گوییم و داریم: $F_{c_i} = F_۱ + F_۲ + \cdots + F_i$..

۴- فراوانی تجمعی نسبی (f_{c_i}): نسبت فراوانی تجمعی به حجم جامعه:

$$f_{c_i} = \frac{F_{c_i}}{N}$$

مثال: اگر در درس مبانی احتمال در یک کلاس ۱۰ نفری یک میان ترم از ۵ نمره، از دانشجویان گرفته شده باشد و نمرات زیر به‌دست آمده باشد، جدول گسسته داده‌ها را می‌توان به‌صورت زیر تشکیل داد.

حل:

$x_i = ۱/۵ , ۱/۵ , ۲/۵ , ۴/۵ , ۳/۵ , ۳/۵ , ۲/۵ , ۲/۵ , ۲/۵ , ۴/۵$: نمرات دانشجویان

x_i	F_i	f_i	F_{c_i}	f_{c_i}
۱/۵	۲	۰/۲	۲	۰/۲
۲/۵	۴	۰/۴	۶	۰/۶
۳/۵	۲	۰/۲	۸	۰/۸
۴/۵	۲	۰/۲	۱۰	۱
	$N=۱۰$			

دقت کنید که می‌توان داده‌ها را به‌صورت جدول پیوسته نیز دسته‌بندی کرد. در این حالت به تعاریف دیگری نیز احتیاج است.

۵- دامنه تغییرات (R): تفاضل بین بزرگ‌ترین داده و کوچک‌ترین داده را دامنه تغییرات می‌نامند. $R = x_{max} - x_{min}$

۶- تعداد طبقات $\mu = \frac{۳Me - Mo}{۲} = \frac{۳۶۰ - ۱۰۰}{۲} = \frac{۲۶۰}{۲} = ۱۳۰ \Rightarrow$: تعداد طبقات یا براساس نیاز مساله مشخص می‌گردد یا می‌توان براساس روابط پیشنهادی زیر آن را به‌دست آورد:

الف) $k = \sqrt{N}$ ب) $k = ۱ + ۳/۳۲ \log N$

۷- فاصله طبقات (I): از تقسیم دامنه تغییرات بر تعداد دسته‌ها حاصل می‌گردد.

$$.. I = \frac{R}{K}$$

۸- مرکز طبقه (نماینده طبقه) $(\tilde{x}_i)$:

$$\tilde{x}_i = \frac{\text{حد بالا} + \text{حد پایین}}{۲}$$

تذکر: واضح است در جدول پیوسته داده‌ها، فراوانی مطلق یک طبقه (و سایر تعاریف) تعداد داده‌هایی است که در آن طبقه قرار می‌گیرند.

مثال: مثال قبل را در نظر بگیرید. جدول پیوسته داده‌ها را طوری تشکیل دهید که شامل ۴ دسته باشد.

حل:

$$R = x_{max} - x_{min} = ۴/۵ - ۱/۵ = ۳ \,,\, k = ۴$$

$$I = \frac{R}{k} = \frac{۳}{۴} = ۰/۷۵$$

حدود طبقات	F_i	F_{c_i}	f_i	f_{c_i}	$\tilde{x}_i$
۱/۵ − ۲/۲۵	۲	۲	۰/۲	۰/۲	$\frac{۳/۷۵}{۲} = ۱/۸۷۵$
۲/۲۵ − ۳	۴	۶	۰/۴	۰/۶	$\frac{۵/۲۵}{۲} = ۲/۶۲۵$
۳ − ۳/۷۵	۲	۸	۰/۲	۰/۸	$\frac{۶/۷۵}{۲} = ۳/۳۷۵$
۳/۷۵ − ۴/۵	۲	۱۰	۰/۲	۱	$\frac{۸/۲۵}{۲} = ۴/۱۲۵$

تذکر: فاصله دو مرکز طبقه متوالی نیز با حدود طبقات برابرست. یعنی:

$$\tilde{x}_{i+۱} - \tilde{x}_i = I \Rightarrow \tilde{x}_{i+۱} = I + \tilde{x}_i$$

۳-۱- نمودارها

برای خلاصه‌کردن داده‌ها می‌توان از نمودارهای آماری نیز استفاده کرد که در این بخش به مهم‌ترین آنها می‌پردازیم.

الف) نمودارهای کمی:

برای توزیع‌های آماری پیوسته و گسسته کاربرد دارد.

ب) نمودارهای وصفی (کیفی):

برای نمایش هندسی داده‌های کیفی کاربرد دارد.

الف) نمودارهای کمی:

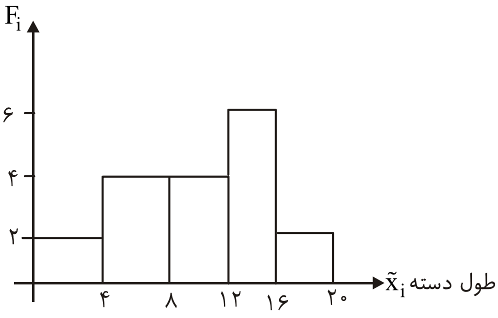
۱- نمودار هیستوگرام (بافت نگار): محور xها بیانگر طول دسته و محور yها بیانگر فراوانی مطلق (فراوانی نسبی) است.

۲- نمودار چندضلعی: محور xها بیانگر مرکز دسته و محور yها بیانگر فراوانی مطلق (فراوانی نسبی) است.

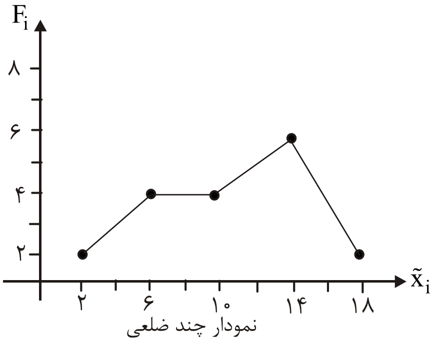
مثال: جدول پیوسته زیر را در نظر بگیرید. نمودارهای هیستوگرام و چند ضلعی را برای آن رسم کرده‌ایم.

حدود طبقات	۰ − ۴	۴ − ۸	۱۸ − ۱۲	۱۲ − ۱۶	۱۶ − ۲۰
F_i فراوانی مطلق	۲	۴	۴	۶	۲
$\tilde{x}_i$	۲	۶	۱۰	۱۴	۱۸

حل:



نمودار هیستوگرام



نمودار چند ضلعی

نمودار هیستوگرام، نمایش هندسیِ بسیار مفیدی است که به کمک آن می‌توان درک صحیحی از مشاهدات داشت و این نمودار در نمایش توزیع، مرکزیت و پراکندگی داده‌ها به کار می‌رود، با این حال نمودار هیستوگرام، امکان شناسایی نقـاط انفـرادی داده‌هـا را فراهم نمی‌کند زیرا داده‌های در یک طبقه، از هم قابل تمییز نیستند.

۳- نمودار ساقه و برگ (شاخه و برگ): این نمودار از نوع تحلیل اکتشافی داده‌هاست. برای رسـم نمـودار سـاقه و بـرگ، ابتـدا داده‌ها را به‌صورت صعودی مرتب کنید. هر داده معادل یک برگ و یک ساقه اسـت. عـدد سـمت راسـت هـر داده را بـرگ و عـدد (اعداد) باقی مانده را ساقه می‌نامیم ادامه مراحل را در مثال زیر ببینید.

مثال: نمودار ساقه و برگ داده‌های زیر را رسم کنید.

۵۹ ، ۱۴ ، ۱۱ ، ۱۸ ، ۵۸ ، ۵۹ ، ۴۸ ، ۲۲ ، ۲۹ ، ۱۴ ، ۱۵ ، ۱۲ ، ۹ ، ۱۸ ، ۲۳ ، ۱۱ ، ۸ ، ۳۶

حل:

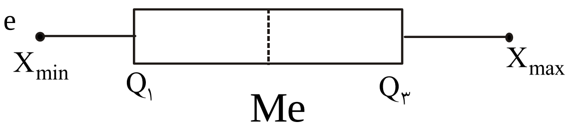
۵۹ ، ۵۹ ، ۵۸ ، ۴۸ ، ۳۶ ، ۲۹ ، ۲۳ ، ۲۲ ، ۱۸ ، ۱۸ ، ۱۵ ، ۱۴ ، ۱۴ ، ۱۲ ، ۱۱ ، ۱۱ ، ۹ ، ۸

نمودار ساقه و برگ به‌صورت زیر رسم می‌گردد:

ساقه	برگ
۰	۸ ۹
۱	۱ ۱ ۲ ۴ ۴ ۵ ۸ ۸
۲	۲ ۳ ۹
۳	۶
۴	۸
۵	۸ ۹ ۹

واضح است که اگر به‌عنوان مثال داده‌ها سه رقمی باشند، رقم یکان به‌عنوان برگ و ساقه شامل اعداد دو رقمی می‌باشد.

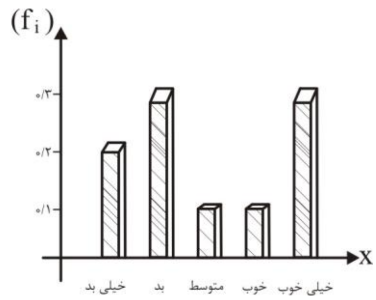
۴- نمودار جعبه‌ای: نمودار جعبه‌ای برای مقایسه‌ی دو یا چند جامعه آماری کاربرد دارد و داده‌ها به‌صورت زیر رسم می‌شود:



که در آن $Q_۱$ و $Q_۳$ به‌ترتیب چارک‌های اول و سوم و Me میانه است که نحوه محاسبه آنها را در ادامـه خـواهیم گفـت. x_{min} کوچک‌ترین داده و x_{max} بزرگ‌ترین داده است. واضح است طول جعبه برابر با $Q_۳ - Q_۱$ می‌باشد.

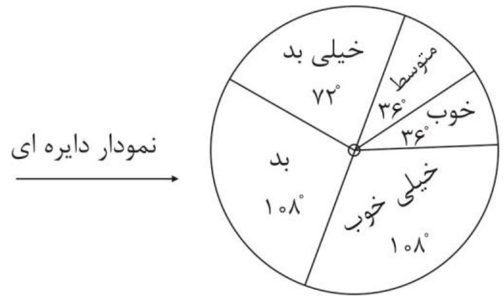
) نمودارهای وصفی:

۱- نمودار ستونی: در این نمودار محورxها نمایش‌دهنده صفت و محور yها بیانگر فراوانی مطلق (F_i) می‌باشد. (البته می‌توان از محور yها برای نشان‌دادن فراوانی نسبی هم استفاده کرد). برای مثال جدول زیر را در نظر بگیرید. نمودارهای میله‌ای آن رسم شده است.



۲- نمودار دایره‌ای: در این نمودار داده‌ها درون یک دایره به نمایش گذاشته می‌شوند که در آن زاویه هر قطاع از رابطه $f_i \times ۳۶۰$ به‌دست می‌آید. به مثال زیر دقت کنید:

x_i	خیلی‌بد	متوسط	خوب	خیلی‌خوب	
فراوانی مطلق F_i	۶	۲	۲	۶	۴
فراوانی نسبی f_i	۰/۳	۰/۱	۰/۱	۰/۳	۰/۲
زاویه هر قطاع	۱۰۸	۳۶ ۰/۱×۳۶۰	۳۶	۱۰۸	۷۲ ۰/۲×۳۶۰



۳- نمودار پاره‌تو (Pareto): نمودار پاره‌تو، نوعی نمودار ستونی برای داده‌های وصفی است. برای رسم آن، موضوعات را ابتدا نزولی کنید. سپس روی محور xها، موضوعات را نشان دهید. سپس فراوانی‌های مطلق را روی محور yها نمایش دهید.
و در نهایت یک محور عمودی دیگر، برای فراوانی‌های تجمعی در نظر بگیرید. به مثال زیر دقت کنید.

مثال: برای جدول زیر نمودار پاره‌تو را رسم نمائید.

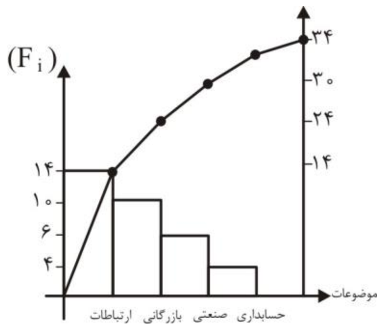
رشته	بازرگانی	حسابداری	صنعتی	ارتباطات
F	۱۰	۴	۶	۱۴

حل:

ابتدا جدول را برحسب فراوانی، به صورت نزولی مجدداً بازنویسی می‌کنیم. F_C ها را نیز به صورت زیر محاسبه می‌کنیم.

رشته / فراوانی	ارتباطات	بازرگانی	صنعتی	حسابداری
F (مطلق)	۱۴	۱۰	۶	۴
F_c (تجمعی)	۱۴	۲۴	۳۰	۳۴

پس نمودار پاره‌تو به صورت زیر خواهد بود.



یادآوری (نماد سیگما):

برای خلاصه نویسی به جای عبارت $x_۱ + x_۲ + \ldots + x_N$ می‌توان از نماد سیگما به صورت $\sum_{i=۱}^{N} x_i$ استفاده کرد. لزومی ندارد حتماً i از عدد یک شروع شود به نکات زیر دقت نمایید:

$$۱)\ \sum_{i=۳}^{۶}(x_i - ۲) = (x_۳ - ۲) + (x_۴ - ۲) + (x_۵ - ۲) + (x_۶ - ۲)$$

$$۲)\ \sum_{i=۱}^{N}(x_i - b) = \sum_{i=۱}^{N} x_i - Nb$$

$$۳)\ \sum_{i=۱}^{N} a x_i = a\sum_{i=۱}^{N} x_i$$

$$۴)\ \sum_{i=۱}^{N}(x_i + y_i) = \sum_{i=۱}^{N} x_i + \sum_{i=۱}^{N} y_i$$

مثال: اگر $x_۱ = ۱$ **و** $x_۲ = -۲$ **و** $x_۳ = ۳$ **و** $x_۴ = -۴$ **باشد حاصل عبارات زیر را بیابید.**

الف) $\sum_{i=۱}^{۴} x_i$

ب) $\sum_{i=۱}^{۴} x_i^۲$

ج) $\left(\sum_{i=۱}^{۴} x_i\right)^۲$ ..

د) $\sum_{i=۱}^{۴}(۳x_i + ۲)$

حل:

$$\text{الف)}\ \sum_{i=۱}^{۴} x_i = x_۱ + x_۲ + x_۳ + x_۴ = (۱) + (-۲) + (۳) + (-۴) = -۲$$

$$\text{ب)}\ \sum_{i=۱}^{۴} x_i^۲ = x_۱^۲ + x_۲^۲ + x_۳^۲ + x_۴^۲ = (۱)^۲ + (-۲)^۲ + (۳)^۲ + (-۴)^۲ = ۱ + ۴ + ۹ + ۱۶ = ۳۰$$

ج) $\left(\sum_{i=1}^{4} x_i\right)^2 = (-2)^2 = 4$

د) $\sum_{i=1}^{4} 3x_i + 2 = 3\sum_{i=1}^{4} x_i + 4(2) = 3(-2) + 8 = -6 + 8 = 2$

۴-۱- شاخص‌های عددی

شاخص‌های عددی پارامترهایی برای مقایسه چند جامعه هستند و می‌توان آنها را به دو دسته اصلی پارامترهای مرکزی و پارامترهای پراکندگی تقسیم کرد.

۱-۴-۱- پارامترهای مرکزی

هر شاخص و پارامتری را که بیانگر مرکز مجموعه داده باشد، پارامتر مرکزی می‌نامیم که از مهمترین آنها می‌توان به میانگین، میانه، مد و چندک‌ها اشاره کرد.

الف) میانگین حسابی: فرض کنید N حجم جامعه و k تعداد طبقات باشد، در این‌صورت میانگین جامعه (μ_x) در حالت‌های مختلف به‌صورت زیر است:

۱- بدون جدول توزیع فراوانی:

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

۲- در جدول گسسته داده‌ها:

$$\mu_x = \frac{\sum_{i=1}^{k} F_i x_i}{N} \quad \text{که در آن} \quad N = \sum_{i=1}^{k} F_i$$

۳- در جدول پیوسته داده‌ها:

$$\mu_x = \frac{\sum_{i=1}^{k} F_i \tilde{x}_i}{N} \quad \text{که در آن} \quad N = \sum_{i=1}^{k} F_i$$

مثال: در هرکدام از قسمت‌های زیر میانگین حسابی را بیابید.

الف) $x_i = 12, 15, 18, 14, 16$

حل:

$$\mu_x = \frac{\sum x_i}{N} = \frac{12+15+18+14+16}{5} = \frac{75}{5} = 15$$

ب)

x_i	۰	۱	۲	۳	۴
F_i	۴	۸	۱۶	۸	۴

حل:

$$N = 40 \Rightarrow \mu_x = \frac{\sum_{k=1}^{5} F_i x_i}{N} = \frac{0+8+32+24+16}{40} = \frac{80}{40} = 2$$

ج)

حدود طبقات	۸ − ۰	۱۶ − ۸	۲۴ − ۱۶	۳۲ − ۲۴	۴۰ − ۳۲
F_i فراوانی مطلق	۴	۸	۲۶	۸	۱۴

حل:

حدود طبقات	۰−۸	۸−۱۶	۱۶−۲۴	۲۴−۳۲	۳۲−۴۰
F_i فراوانی مطلق	۴	۸	۲۶	۸	۱۴
$\tilde{x}_i$	۴	۱۲	۲۰	۲۸	۳۶

$$\mu_x=\frac{\sum_{i=1}^{5}F_i\tilde{x}_i}{N}=\frac{4\times4+8\times12+26\times20+8\times28+14\times36}{60}=\frac{16+96+520+224+504}{60}=\frac{1360}{60}=\frac{45}{2}=22/5$$

خواص میانگین حسابی:

۱- منحصربه‌فرد است.

۲- میانگین اعداد ثابت b، برابر b است. $\mu_b=b$

۳- اگر $z=ax+by+c$ آنگاه $\mu_z=a\mu_x+b\mu_y+c$

۴- مجموع انحرافات از میانگین، برابر صفر است. $\left(\sum_{i=1}^{k}(x_i-\mu)=0\right)$

۵- مجموع مربعات انحرافات از میانگین، حداقل است (یعنی برای هر داده دلخواه A، $\sum_{i=1}^{k}(x_i-\mu_x)^2<\sum_{i=1}^{k}(x_i-A)^2$)

۶- اگر میانگین حسابی N_1 داده برابر μ_1، میانگین حسابی N_2 داده برابر μ_2 و ... و میانگین N_k داده برابر μ_k باشد، آنگاه میانگین کل μ_t برابر است با:

$$\mu_t=\frac{\sum_{i=1}^{k}N_i\mu_i}{\sum_{i=1}^{k}N_i}=\frac{N_1\mu_1+N_2\mu_2+\dots+N_k\mu_k}{N_1+\dots+N_t}$$

۷- مهمترین شاخص مرکزی، میانگین است.

مثال: میانگین نمرات ۲۰ دانشجوی خانم در درس آمار ۱۴ و میانگین ۱۵ دانشجوی آقا در همین درس و همین کلاس برابر ۱۸ می‌باشد. میانگین کل کلاس را بیابید.

حل:

$$\mu_T=\frac{N_1\mu_1+N_2\mu_2}{N_1+N_2}=\frac{20\times14+15\times18}{20+15}=\frac{280+270}{35}=\frac{550}{35}\simeq15/72$$

نکته: روابط زیر را به‌خاطر بسپارید.

۱) $\sum_{i=1}^{n}i=\frac{n(n+1)}{2}=1+2+3+\cdots+n$

۲) $\sum_{i=1}^{n}i^2=\frac{n(n+1)(2n+1)}{6}=1+2^2+3^2+\cdots+n^2$

۳) $(a+b)^n=\sum_{i=0}^{n}\binom{n}{i}a^ib^{n-i}\xrightarrow{a=1,b=-1}\binom{n}{0}-\binom{n}{1}+\dots+\binom{n}{n}=0\xrightarrow{a=b=1}\sum_{i=0}^{n}\binom{n}{i}=2^n$

۴) $\sum_{i=1}^{n}i^3=\left(\frac{n(n+1)}{2}\right)^2=1+2^3+3^3+\cdots+n^3$

مثال: میانگین حسابی را برای جدول زیر بیابید.

حدود طبقات	۰−۴	۴−۸	۸−۱۲	۱۲−۱۶	۱۶−۲۰
فراوانی	۲	۴	۸	۴	۲

حل:

ابتدا مرکز هر طبقه $F_i\tilde{x}_i, \tilde{x}_i$ را به دست می‌آوریم. پس:

حدود طبقات	۰−۴	۴−۸	۸−۱۲	۱۲−۱۶	۱۶−۲۰
F_i	۲	۴	۸	۴	۲
$\tilde{x}_i$	۲	۶	۱۰	۱۴	۱۸
$F_i\tilde{x}_i$	۴	۲۴	۸۰	۵۶	۳۶

$$\Rightarrow N=\sum F_i=۲۰$$

$$\mu=\frac{\sum F_i\tilde{x}_i}{N}=\frac{۴+۲۴+۸۰+۵۶+۳۶}{۲۰}=\frac{۲۰۰}{۲۰}=۱۰$$

ب) میانگین هندسی: میانگین هندسی را به‌صورت $\tilde{x}_G$ نمایش می‌دهیم و به‌صورت زیر محاسبه می‌کنیم:

۱- بدون جدول توزیع فراوانی:

$$\tilde{x}_G=\sqrt[N]{x_۱.x_۲.x_۳\ldots x_N}=\sqrt[N]{\prod_{i=۱}^{N}x_i}$$

۲- در جدول گسسته داده‌ها $N=\sum_{i=۱}^{k}F_i$ که:

$$\tilde{x}_G=\sqrt[N]{x_۱^{F_۱}.x_۲^{F_۲}\ldots x_k^{F_k}}$$

۳- در جدول پیوسته داده‌ها: $N=\sum_{i=۱}^{k}F_i$ و $\tilde{x}_i$ ها مرکز دسته که:

$$\tilde{x}_G=\sqrt[N]{\tilde{x}_۱^{F_۱}.\tilde{x}_۲^{F_۲}\ldots \tilde{x}_k^{F_k}}$$

مثال: در هر قسمت میانگین هندسی را بیابید.

الف) $x_i=۳,۶,۱۲$

حل:

$$\bar{x}_G=\sqrt[۳]{۳\times۶\times۱۲}=\sqrt[۳]{۳\times۳\times۲\times۲^۲\times۳}=\sqrt[۳]{۳^۳\times۲^۳}=۶$$

ب) $x_i=۱,۲^t,۳^t,\ldots,t^t$ **به‌طوری‌که** $t\in\mathbb{N}, t>۱$

حل:

$$\bar{x}_G=\sqrt[t]{\prod_{i=۱}^{t}i^t}=\sqrt[t]{(\prod_{i=۱}^{t}i)^t}=\prod_{i=۱}^{t}i=t!$$

تست:

میانگین داده‌های هندسی ۱۹۲,۷۲,۱۶۲,۱۲۸,۲۱۶ کدام است؟ (کارشناسی ارشد سراسری- ارتباطات)

۱) ۱۵۶ ۲) ۱۴۴ ۳) ۱۳۲ ۴) ۱۰۸

حل:

گزینه ۲ صحیح است.

$$\bar{x}_G=\sqrt[5]{192\times 72\times 162\times 128\times 216}=\sqrt[5]{(3\times 2^6)(3^2\times 2^3)(2\times 3^4)(2^7)(3^3\times 2^3)}=\sqrt[5]{3^{10}\times 2^{25}}=3^2\times 2^4=144$$

خواص میانگین هندسی:

۱- هنگامی از میانگین هندسی استفاده می‌شود که داده‌های آماری به‌صورت نسبت درصد و یا نرخ رشد باشند.

۲- $\bar{x}_G=\sqrt[N]{\prod_{i=1}^{N} x_i} \Leftarrow \log \bar{x}_G=\frac{1}{N}\sum_{i=1}^{N}\log x_i$

۳- متوسط نرخ رشد $=\bar{x}_G-1$

۴- اگر فروش در سال اول x_1 ، در سال دوم x_2 و ... و در سال Nام، x_N باشد داریم: $\bar{x}_G=\sqrt[N-1]{\frac{x_N}{x_1}}$ که $\bar{x}_G$ متوسط فروش هر سال نسبت به سال قبل است.

مثال: اگر متوسط نرخ رشد سهام در هر سال به‌طور متوسط ۱۰۰٪ باشد، بعد از چند دوره قیمت سهام از ۲۰۰ ریال به ۶۴۰۰ می‌رسد؟

حل:

$$\text{متوسط نرخ رشد}=\bar{x}_G-1=\sqrt[N]{\frac{x_N}{x_0}}-1\Rightarrow \bar{x}_G-1=\sqrt[N]{\frac{6400}{200}}-1\Rightarrow 2=\sqrt[N]{32}\Rightarrow N=5$$

مثال: فروش یک فروشگاه در سال گذشته ۸۰ درصد افزایش داشته و امسال ۸۰ درصد کاهش یافته است. متوسط نرخ رشد فروش سالانه در این سال کدام است؟

$$\text{نرخ رشد}=\bar{x}_G-1\Rightarrow \bar{x}_G=\text{نرخ رشد}+1$$

نرخ رشد	$\bar{x}_G$
$+0/8$	$1/8$
$-0/8$	$0/2$

$$\Rightarrow \bar{x}_G=\sqrt{1/8\times 0/2}=\sqrt{0/36}=0/6\Rightarrow \text{متوسط نرخ رشد}=\bar{x}_G-1=0/6-1=-0/4$$

۴۰ درصد کاهش داشته باشد.

ج) میانگین هارمونیک (همسازه): میانگین هارمونیک را با $\bar{x}_H$ نمایش داده و به‌صورت زیر محاسبه می‌کنیم:

۱- بدون جدول توزیع فراوانی:

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

۲- در جدول گسسته داده‌ها:

$$\bar{x}_H=\frac{N}{\sum_{i=1}^{K}F_i\left(\frac{1}{x_i}\right)} \quad \text{که} \quad \sum_{i=1}^{k}F_i=N$$

۳- در جدول پیوسته داده‌ها:

$$\bar{x}_H=\frac{N}{\sum_{i=1}^{k}F_i\left(\frac{1}{\tilde{x}_i}\right)} \quad \text{که} \quad \sum_{i=1}^{k}F_i=N$$

✓ **تست: یک هواپیما فاصله ۳ هزار کیلومتری را با سرعت ۶۰۰ کیلومتر بر ساعت، فاصله ۵ هزار کیلومتر را با سرعت ۷۵۰ کیلومتر بر ساعت و فاصله ۴ هزار کیلومتری را با سرعت ۸۰۰ کیلومتر بر ساعت طی می‌کند. سرعت متوسط آن کدام است؟**
(مدیریت و حسابداری – کارشناسی ارشد– سراسری)

۱) ۷۱۵ ۲) ۷۱۷ ۳) ۷۲۰ ۴) ۷۲۵

پاسخ: گزینه «۳»

$$\bar{x}_H=\frac{\sum_{i=1}^{k}F_i}{\sum_{i=1}^{k}F_i\left(\frac{1}{x_i}\right)}=\frac{3+5+4}{\frac{3}{600}+\frac{5}{750}+\frac{4}{800}}=\frac{12}{\frac{10}{600}}=720$$

خواص میانگین هارمونیک:

۱- از میانگین هارمونیک زمانی استفاده می‌شود که داده‌ها دارای مقیاس ترکیبی باشند. مقیاس ترکیبی مانند: نفر بر ساعت، کیلومتر بر ساعت، متر بر ثانیه و ...

۲- اگر کلیه داده‌ها با هم برابر باشند داریم: $\bar{x}_H = \bar{x}_G = \mu_x$ و درغیراین‌صورت $\bar{x}_H < \bar{x}_G < \mu_x$

۳- در حالت کلی داریم: $\bar{x}_H \le \bar{x}_G \le \mu_x$

۴- اگر کلیه داده‌ها a برابر شوند، هر سه میانگین حسابی، هندسی و هارمونیک نیز a برابر می‌شوند.

✓تست: راننده‌ای $\frac{۱}{۳}$ مسیر را با سرعت ۶۰ کیلومتر بر ساعت و بقیه را با سرعت ثابت ۹۰ کیلومتر بر ساعت طی می‌کند و همین مسیر را با سرعت ثابت ۸۰ کیلومتر بر ساعت برمی‌گردد. سرعت متوسط او در کل مسیر رفت و برگشت کدام است؟ (حسابداری و مدیریت- کارشناسی ارشد- سراسری)

۱) ۷۸/۵ ۲) ۸۰ ۳) ۸۱/۵ ۴) ۸۲

☑**پاسخ:** گزینه ۱ صحیح است.

اگر طول کل مسیر را **x** در نظر بگیریم راننده $\frac{۱}{۳}x$ را با سرعت $۶۰\frac{km}{h}$، $\frac{۲}{۳}x$ را با سرعت $۹۰\frac{km}{h}$ و مسیر برگشت (که طول آن **x** است) را با سرعت ۸۰ برگشته است یعنی کلاً ۲x را طی کرده و داریم:

x_i	۶۰	۹۰	۸۰
F_i	$\frac{x}{۳}$	$\frac{۲}{۳}x$	x

$$\Rightarrow \bar{x}_H = \frac{۲x}{\frac{\frac{۱}{۳}x}{۶۰}+\frac{\frac{۲}{۳}x}{۹۰}+\frac{x}{۸۰}} = \frac{۲}{\frac{۱}{۱۸۰}+\frac{۲}{۲۷۰}+\frac{۱}{۸۰}} = \frac{۲}{\frac{۴+۹}{۷۲۰}+\frac{۲}{۲۷۰}}$$

$$= \frac{۲}{\frac{۱۳}{۷۲۰}+\frac{۲}{۲۷۰}} = \frac{۲}{\frac{۳۹+۱۶}{۲۱۶۰}} = \frac{۴۳۲۰}{۵۵} = ۷۸/۵$$

د) مد (نماد) (Mo): مُد کلمه‌ای فرانسوی برای رایجترین لباس یا سبک است. در آمار، داده‌ای که بیشترین فراوانی را دارد، مد یا نما می‌نامند. درواقع احتمال وقوع این داده بیشتر از داده‌های دیگر است. واضح است که مد لزوما منحصر به فرد نیست.

۱- مد در حالت بدون جدول

مثال: مد را در حالت‌های زیر بیابید.

الف) $x_i = ۲,۴,۶,۴,۳,۸,۴,۱۱ \rightarrow Mo = ۴$ (مد)

ب) $x_i = ۳,۶,۲,۴,۹,۱۱,۳,۸,۶ \rightarrow Mo = ۳,۶$ (دونمایی)

ج) $x_i = ۸,۵,۲,۸,۵,۲ \rightarrow Mo =$ چون تمام داده‌ها به یک میزان تکرار شده‌اند، مد ندارد

د) $x_i = ۳,۴,۹,۸,۳,۴,۹,۱۱,۳,۴,۹,۱۲ \rightarrow Mo = ۳,۴,۹$ (سه نمایی)

۲- مد در جدول گسسته داده‌ها، داده‌ای که بیشترین فراوانی مطلق را دارد.

مثال: مد را در جدول زیر بیابید.

حل:

x_i	۰	۱	۲	۳	۴
F_i	۳	۴	۲	۶	۲

$$\Rightarrow x_i = ۳ = Mo$$

$$x_i = ۳ \Rightarrow Mo = ۳$$

۳- مد در جدول پیوسته داده‌ها

برای محاسبه مد، ابتدا طبقه مدار را (طبقه‌ای که بیشترین فراوانی مطلق را دارد) انتخاب کنید. سپس مد را از رابطه زیر بیابید:

$$Mo = L_{mo} + \left(\frac{d_۱}{d_۱ + d_۲} \times I\right)$$

که در آن L_{mo}، کران پایین طبقه مدار است و $d_۱$ تفاضل فراوانی مطلق طبقه مدار با طبقه ما قبل آن $\left(d_۱ = F_{mo} - F_{mo-۱}\right)$ و $d_۲$ تفاضل فراوانی مطلق طبقه مدار با طبقه بعد از آن $\left(d_۲ = F_{mo} - F_{mo+۱}\right)$ می‌باشد. I نیز طول طبقات است.

مثال: در جدول زیر مد را بیابید.

حل:

حدود طبقات	۹−۱۲	۱۲−۱۵	۱۵−۱۸	۱۸−۲۱
فراوانی مطلق F_i	۸	۱۲	۱۵	۹

طبقه مدار

$$\Rightarrow Mo = ۱۵ + \left(\frac{۱۵-۱۲}{(۱۵-۱۲)+(۱۵-۹)} \times ۳\right) = ۱۵ + \left(\frac{۳}{۹} \times ۳\right) = ۱۶$$

مثال: در جدول زیر مد را بیابید.

حل:

حدود طبقات	F_i
۰-۴	۴
۴-۸	۶
۸-۱۲	۱۲
۱۲-۱۶	۱۲
۱۶-۲۰	۱۰

$$\rightarrow Mo = ۸ + \left(\frac{۱۲-۶}{(۱۲-۶)+(۱۲-۱۲)} \times ۴\right) = ۸ + ۴ = ۱۲$$

$$\rightarrow Mo = ۱۲ + \left(\frac{۱۲-۱۲}{(۱۲-۱۲)+(۱۲-۱۰)} \times ۴\right) = ۱۲ + ۰ = ۱۲$$

نکته:

مد را می‌توان برحسب فراوانی نسبی از رابطه $L_{mo} + (\frac{d_۱}{d_۱ + d_۲} \times I) = L_{mo} + (\frac{f_{mo} - f_{mo-۱}}{۲f_{mo} - f_{mo-۱} - f_{mo+۱}} \times I)$ یافت که در آن f_{mo} فراوانی نسبی طبقه‌دار، $f_{mo-۱}$ فراوانی نسبی طبقه قبل از طبقه مددار است.

مثال. در جدول فراوانی نسبی زیر مد را بیابید.

حدود طبقات	۰−۴	۴−۸	۸−۱۲	۱۲−۱۶	۱۶−۲۰
فراوانی نسبی	۰/۱	۰/۲	۰/۴	۰/۱	۰/۲

طبقه مدار

$$mo = ۸ + \frac{(۰/۴ - ۰/۲)}{(۰/۴-۰/۲)+(۰/۴-۰/۱)} \times ۴ = ۸ + \frac{۰/۲}{۰/۲+۰/۳} \times ۴ = ۸ + \frac{۸}{۵} = ۸ + \frac{۱۶}{۱۰} = ۹/۶$$

نکته:

۱- اگر بیشترین فراوانی مطلق در دو طبقه متوالی برابر باشند، حد بالای طبقه بیشترین اولی یا حد پایین طبقه بیشترین دومی برابر مد است.

۲- برخلاف میانگین و میانه، مد منحصربه‌فرد نیست.

۳- برخلاف میانگین، مد تحت‌تاثیر داده‌های کوچک و بزرگ نیست.

ه) میانه (Median): میانه برای مجموعه‌ای از داده‌ها که به‌صورت صعودی و یا نزولی مرتب شده‌اند، داده‌ای است که نیمی از داده‌ها کوچک‌تر از آن و نیمی دیگر از داده‌ها بزرگ‌تر از آن باشند. به‌عبارتی داده‌ای است که در وسط توزیع داده‌های مرتب شده قرار دارد.

۱- میانه در حالت بدون جدول: ابتدا داده‌ها را به‌صورت صعودی مرتب کنید. اگر تعداد داده‌ها فرد باشد داده وسطی برابر میانه است و اگر تعداد داده‌ها زوج باشد، میانگین دو داده وسطی برابر میانه می‌باشد (یعنی در این حالت $(Me=\dfrac{x_{\frac{n}{۲}}+x_{\frac{n}{۲}+۱}}{۲}$

مثال: میانه را در موارد زیر بیابید.

حل:

الف) $x_i : ۵\,,۱۲\,,۸\,,۱۹\,,۱۷$

$$x_i : ۵,۸\,,۱۲\,,۱۷\,,۱۹ \xrightarrow{\text{تعداد داده‌ها فرد است.}} \text{داده وسطی} = ۱۲ = Me$$

ب) $x_i : ۵\,,۱۲\,,۸\,,۱۹\,,۱۷\,,۹$

$$x_i : ۵,۸\,,۹\,,۱۲\,,۱۷\,,۱۹ \xrightarrow[n=۶]{\text{تعداد داده‌ها زوج است.}} Me=\frac{x_۳+x_۴}{۲}=\frac{۹+۱۲}{۲}=\frac{۲۱}{۲}=۱۰/۵$$

۲- میانه در جدول گسسته: فراوانی تجمعی را محاسبه کنید. میانه برابر اولین داده‌ای است که فراوانی تجمعی آن بزرگ‌تر از $\frac{N}{۲}$ باشد. دقت کنید اگر داده‌ای با فراوانی تجمعی دقیقاً برابر $\frac{N}{۲}$ وجود داشته باشد، آنگاه میانه از رابطه $\dfrac{x_{\frac{N}{۲}}+x_{\frac{N}{۲}+۱}}{۲}$ به‌دست می‌آید.

مثال: در حالات زیر میانه را بیابید.

حل:

الف)

x_i	F_i	F_{c_i}
۲	۴	۴
۴	۸	۱۲
۶	۱۲	$\boxed{۲۴}$
۸	۸	۳۲
۱۰	۲	$۳۶=N$

$$> \frac{N}{۲}=\frac{۳۶}{۲}=۱۸=\frac{N}{۲} \Rightarrow Me=۶$$

ب)

x_i	F_i	F_{c_i}
۲	۴	۴
۴	۸	۱۲
۶	۶	$\boxed{۱۸}$
۸	۱۰	۲۸
۱۰	۸	$۳۶=N$

$$=\frac{N}{۲}=۱۸ \rightarrow Me=\frac{x_a+x_{۱۹}}{۲}=\frac{۶+۸}{۲}=۷$$

۳- میانه در جدول پیوسته: برای محاسبه میانه در جدول پیوسته داده‌ها، ابتدا فراوانی‌های تجمعی را محاسبه کنید. سپس طبقه میانه‌دار (اولین طبقه‌ای که فراوانی تجمعی آن بزرگ‌تر یا مساوی $\frac{N}{۲}$ است) را مشخص کنید. سپس میانه را از رابطه زیر محاسبه کنید.

$$Me = L_{Me} + \left(\frac{\frac{N}{۲} - Fc_{Me-۱}}{F_{Me}} \times I \right)$$

که در آن: Me: میانه، L_{Me}: کران پایین طبقه میانه‌دار، I: فاصله طبقات، فراوانی تجمعی طبقه ماقبل و F_{Me}: فراوانی مطلق طبقه میانه‌دار است.

مثال: در موارد زیر میانه را بیابید:

حل:

الف)

حدود طبقات	F_i	F_{c_i}
$۰-۵$	۴	$۴ < \frac{N}{۲} = ۱۰$
$۵-۱۰$	۵	$۹ < \frac{N}{۲} = ۱۰$
$۱۰-۱۵$	۳	$۱۲ \geq \frac{N}{۲} = ۱۰$
$۱۵-۲۰$	۸	۲۰
	$N = ۲۰$	$\rightarrow \frac{N}{۲} = ۱۰$

$\rightarrow$ طبقه میانه‌دار $\Rightarrow Me = ۱۰ + \left(\frac{۱۰-۹}{۳} \times ۵ \right) = \frac{۳۵}{۳}$

ب)

	F_i	F_{ci}
$۰-۵$	۴	$۴ < \frac{N}{۲} = ۱۰$
$۵-۱۰$	۶	$۱۰ \geq \frac{N}{۲} = ۱۰$
$۱۰-۱۵$	۳	۱۳
$۱۵-۲۰$	۷	۲۰
	$N = ۲۰$	$\rightarrow \frac{N}{۲} = ۱۰$

$\rightarrow$ طبقه میانه‌دار $\Rightarrow Me = ۵ + \left(\frac{۱۰-۴}{۶} \times ۵ \right) = ۱۰$

نکته:

۱- اگر فراوانی تجمعی طبقه میانه‌دار برابر $\frac{N}{۲}$ باشد، حد بالای آن طبقه برابر میانه است.

۲- میانه منحصربه‌فرد است و تحت تاثیر داده‌های بزرگ و کوچک نمی‌باشد.

۳- مجموع قدر مطلق انحرافات از میانه حداقل است (یعنی برای داده دلخواه $A: \sum_{i=۱}^{n}|x_i - A| < \sum_{i=۱}^{n}|x_i - Me|$)

تست: عبارت $\sum_{i=۱}^{n}|x_i - A|$ به ازای کدام مورد کمینه است؟ (دکتری روانشناسی و اقتصاد و حسابداری+ ارشد حسابداری مدیریت اقتصاد)

۱) $A = \bar{x}$ ۲) $A = Mod$ ۳) $A = Me$ ۴) $A = \bar{x}_H$

حل: گزینه ۳ صحیح است.

طبق نکته قبل $\sum|x_i - Me|$ کمترین مقدار را دارد. پس باید $A = Me$ باشد.

۴- میانه را برحسب فراوانی نسبی نیز می‌توان از رابطه زیر یافت:

$$Me = LMe + \frac{۰/۵ - f_{cMe-۱}}{fMe} \times I$$

که در آن L_{Me} کران پایین طبقه میانه‌دار (اولین طبقه‌ای که فراوانی تجمعی نسبی آن $\geq ۰/۵$) $fc_{Me-۱}$ فراوانی نسبی تجمعی طبقه قبل و f_{Me} فراوانی نسبی طبقه میانه‌دارست.

تست. در جدول فراوانی نسبی زیر میانه کدام است؟ (ارتباطات- کنکور کارشناسی ارشد- سراسری)

فاصله طبقات	۱۰−۱۵	۱۵−۲۰	۲۰−۲۵	۲۵−۳۰	۳۰−۳۵
فراوانی نسبی	۰/۱	۰/۱۵	۰/۲	۰/۲۵	۰/۳

۱) ۲۶/۵ ۲) ۲۶ ۳) ۲۷ ۴) ۲۸

حل: گزینه ۲ صحیح است.

فاصله طبقات	۱۰−۱۵	۱۵−۲۰	۲۰−۲۵	۲۵−۳۰	۳۰−۳۵
فراوانی نسبی f	۰/۱	۰/۱۵	۰/۲	۰/۲۵	۰/۳
فراوانی نسبی تجمعی f_c	۰/۱	۰/۲۵	۰/۴۵	$\underline{\underline{۰/۷}}$	۱

طبقه میانه‌دار است. زیرا اولین طبقه‌ای است که $f_{c \geq ۰/۵}$

$$Me = ۲۵ + \frac{۰/۵ - ۰/۴۵}{۰/۲۵} \times ۵ = ۲۵ + \frac{۰/۰۵}{۰/۲۵} \times ۵ = ۲۵ + \frac{۰/۲۵}{۰/۲۵} = ۲۵ + ۱ = ۲۶$$

و) چندک‌ها (چارک Q، دهک D و صدک P)

چندک i ام کمیتی است که $\frac{i}{k}$ ام داده‌ها کوچک‌تر از آن و بقیه داده‌ها از آن بزرگ‌تر است. در چارک i ام

$۷۲ = \bar{X}_H = \frac{\frac{a}{۳} + \frac{a}{۲} + \frac{a}{۶}}{\frac{\frac{a}{۳}}{۶۰} + \frac{\frac{a}{۲}}{۷۵} + \frac{\frac{a}{۶}}{۱۰۰}} \Rightarrow$، دهک i ام $k = ۱۰$ و در صدک i ام $k = ۱۰۰$ است. به‌عنوان مثال چارک سوم، کمیتی است که

$\frac{۳}{۴}$ داده‌ها از آن کوچک‌تر و بقیه داده‌ها از آن بزرگ‌تر است. واضح است چارک دوم، دهک پنجم و صدک پنجاهم با هم برابر هستند و با مقدار میانه نیز برابر می‌باشند.

۱- محاسبه چندک‌ها در حالت بدون جدول

۱- ابتدا داده‌ها را به‌صورت صعودی مرتب کنید.

۲- جایگاه داده‌ها را شماره‌گذاری نمایید.

۳- جایگاه چندک i ام را از رابطه $t=\frac{iN}{k}+\frac{۱}{۲}$ بیابید. (برای چارک $k=۴$، دهک $k=۱۰$ و صدک $k=۱۰۰$ است.)

۴- داده‌ای که در مکان t ام قرار دارد چندک iام است یعنی مقدار چندک را از رابطه $a_{[t]}+(t-[t])(a_{[t+۱]}-a_{[t]})$ می‌توان محاسبه کرد. (دقت کنید $t-[t]$ یعنی قسمت اعشاری t)

مثال: در داده‌های زیر:

۱۰۴, ۹۶ , ۹ , ۳۰ , ۴۰ , ۸۰

الف) چارک سوم **ب) دهک ششم** **ج) صدک چهل و چهارم را بیابید.**

حل:

الف) چارک سوم

داده‌ها x_i : ۹ , ۳۰ , ۴۰ , ۸۰ , ۹۶ , ۱۰۴

شماره‌گذاری: $a_۱ , a_۲ , a_۳ , a_۴ , a_۵ , a_۶ \quad \rightarrow N=۶$

چارک سوم: $t=\frac{۳N}{۴}+\frac{۱}{۲}=\frac{۳\times ۶}{۴}+\frac{۱}{۲}=\frac{۱۸+۲}{۴}=۵$

$\rightarrow Q_۳=a_{[a]}+(۵-[۵])\ (a_۶-a_۵)=a_۵=۹۶$

ب) دهک ششم

دهک ششم: $t=\frac{۶N}{۱۰}+\frac{۱}{۲}=\frac{۶\times ۶}{۱۰}+\frac{۱}{۲}=\frac{۳۶+۵}{۱۰}=\frac{۴۱}{۱۰}=۴/۱$

$D_۶=a_{[۴/۱]}+(۴/۱-[۴/۱])(a_{[۵/۱]}-a_{[۴/۱]})=a_۴+(۰/۱)(a_۵-a_۴)=۸۰+(۰/۱)(۹۶-۸۰)=۸۰+۱/۶=۸۱/۶$

ج) صدک چهل و چهارم

صدک چهل و چهارم $x=۲۰$

$P_{۴۴}=a_{[۳/۱۴]}+(۳/۱۴-[۳/۱۴])(a_{[۴/۱۴]}-a_{[۳/۱۴]})=a_۳+(۰/۱۴)(a_۴-a_۳)=۴۰+(۰/۱۴)(۸۰-۴۰)=۴۰+۵/۶=۴۵/۶$

۲- محاسبه چندک در حالت جدول پیوسته

۱- ابتدا فراوانی‌های تجمعی را محاسبه کنید.

۲- طبقه شامل چندک i ام را مشخص کنید. (طبقه‌ای که فراوانی تجمعی آن بزرگ‌تر یا مساوی $\frac{iN}{k}$ است که در آن N تعداد داده‌ها، k برای چارک ۴، دهک ۱۰ و صدک ۱۰۰ می‌باشد.)

۳- محاسبه چندک از رابطه زیر:

$$\text{چندک i ام}=L_{Q_{ik}}+\left(\frac{\frac{iN}{k}-F_{C_{Q_{ik-۱}}}}{F_{Q_{ik}}}\right)\times I$$

که در آن $L_{Q_{ik}}$ کران پایین طبقه شامل چندک i ام و $F_{Q_{ik}}$ فراوانی مطلق طبقه شامل چندک i ام و $F_{C_{Q_{ik-1}}}$ فراوانی تجمعی طبقه ماقبل شامل چندک i ام است.

✓ **تست: متغیرهای پیوسته در جدول زیر گروه‌بندی شده‌اند. متغیر ۸۰ درصدی داده‌ها کدام است؟**

(مدیریت و حسابداری- کارشناسی ارشد- سراسری)

حدود دسته	۱۲-۱۵	۱۵-۱۸	۱۸-۲۱	۲۱-۲۴	۲۴-۲۷	۲۷-۳۰
فراوانی	۵	۱۰	۹	۱۱	۸	۷

۱) ۲۵/۸۷۵
۲) ۲۵/۶۲۵
۳) ۲۶/۱۲۵
۴) ۲۶/۲۲۵

☑ **پاسخ:** گزینه «۱»

هدف محاسبه صدک هشتادم است.

حدود دسته	۱۲-۱۵	۱۵-۱۸	۱۸-۲۱	۲۱-۲۴	۲۴-۲۷	۲۷-۳۰
فراوانی مطلق	۵	۱۰	۹	۱۱	۸	۷
فراوانی تجمعی	۵	۱۵	۲۴	۳۵	۴۳	۵۰

$$\frac{aN}{100}=\frac{80\times 50}{100}=40<43\rightarrow$$ طبقه پنجم شامل صدک هشتادم

$$P_{80}=L_5+\left(\frac{\frac{80\times 50}{100}-F_{C_4}}{F_5}\right)=24+\left(\frac{40-35}{8}\times 3\right)=25/875$$

مثال: با توجه به جدول توزیع فراوانی زیر، چارک اول و دهک هشتم را بیابید.

فاصله طبقات	۰-۴	۴-۸	۸-۱۲	۱۲-۱۶	۱۶-۲۰
فراوانی مطلق	۵	۱۰	۲۰	۱۰	۵

حل: ابتدا فراوانی‌های تجمعی را می‌یابیم.

فاصله طبقات	۰-۴	۴-۸	۸-۱۲	۱۲-۱۶	۱۶-۲۰
F	۵	۱۰	۲۰	۱۰	۵
F_C	۵	۱۵	۳۵	۴۵	$50=N$

برای یافتن چارک اول $i=1, K=4$

پس اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر مساوی $\frac{iN}{k}=\frac{1\times 50}{4}=12/5$ است (یعنی طبقه دوم) شامل Q_1 می‌باشد. پس:

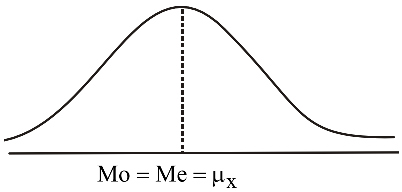
$$Q_1=4+\frac{12/5-5}{10}\times 4=7$$

برای یافتن دهک هشتم قرار دهید. $i=8, k=10$، پس اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر مساوی $\frac{iN}{k}=\frac{8\times 50}{10}=40$ است (یعنی طبقه ۴) شامل D_8 می‌باشد. پس:

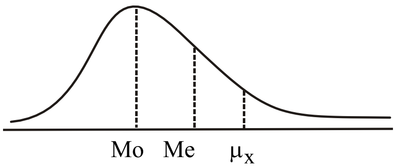
$$D_8=12+\left(\frac{40-35}{10}\times 4\right)=12+2=14$$

نکته: (رابطه میانگین – میانه – مد)

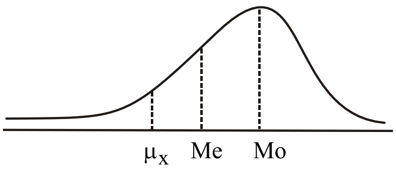
الف) در یک توزیع متقارن داریم: $Me=Mo=\mu_x$،



ب) در یک توزیع چوله به راست داریم: $Mo < Me < \mu_x$،



ج) در یک توزیع چوله به چپ داریم: $\mu_x < Me < Mo$،



نکته: اگر هریک از داده‌ها دارای وزن w_i باشند، برای محاسبه میانگین می‌توان از میانگین وزنی استفاده کرد و داریم:

$$\mu_w = \frac{\sum w_i x_i}{\sum w_i}$$

مثال: میانگین وزنی نمرات یک دانشجو را با توجه به جدول زیر بیابید.

حل:

درس	واحد	نمره	$w_i x_i$
مبانی احتمال	۳	۱۶	۴۸
مبانی آنالیز ریاضی	۳	۱۴	۴۲
زبان تخصصی	۲	۱۵	۳۰
اقتصاد	۴	۱۷	۶۸
ریاضی عمومی	۴	۱۶	۶۴

$$\Rightarrow \mu_w = \frac{\sum_{i=۱}^{۵} w_i x_i}{\sum_{i=۱}^{۵} w_i} = \frac{۴۸+۴۲+۳۰+۶۸+۶۴}{۱۶} = \frac{۲۵۲}{۱۶} = ۱۵/۷۵$$

نکته: (میانگین پیراسته) میانگین حسابی تحت‌تاثیر داده‌های کوچک و بزرگ دچار خطا می‌گردد. برای برطرف کردن این مشکل می‌توان از میانگین پیراسته استفاده کرد. برای این کار لازم است ابتدا داده‌ها را به شکل صعودی مرتب کنید. سپس داده‌های نامتعارف را حذف کنید و درنهایت قرینگی را حفظ کنید. (یعنی اگر پس از مرتب کردن از ابتدای داده‌ها ۲ داده حذف شود حتما از انتهای داده‌ها هم باید دو داده حذف گردد) سپس میانگین داده‌های باقی‌مانده را محاسبه کنید.

مثال: میانگین پیراسته را برای داده‌های ۵۰۰ ، ۶۵۰ ، ۵۵۰، ۵ ، ۴ ، ۶۰۰ ، ۷۰۰ ، ۵۵۰۰ بیابید.

حل: ابتدا داده‌ها را مرتب می‌کنیم و داریم ۵۵۰۰ ، ۷۰۰ ، ۶۵۰ ، ۶۰۰ ، ۵۵۰ ، ۵۰۰ ، ۵ ، ۴ بنابراین واضح است دو داده ابتدایی و یک داده پایانی نامتعارف است پس آنها را حذف می‌کنیم. چون از ابتدای لیست داده‌ها دو عدد حذف شده از انتها هم باید به دلیل قرینگی دو عدد حذف شود پس علاوه‌بر ۵۵۰۰ عدد ۷۰۰ نیز حذف می‌گردد تا ۶۵۰ و ۶۰۰ و ۵۵۰ و ۵۰۰ باقی بمانند که میانگین آنها برابر است با:

$$\frac{۵۰۰+۵۵۰+۶۰۰+۶۵۰}{۴} = ۵۷۵$$

نکته (میانگین وینزوری):

برای یافتن میانگین وینزوری گام‌های زیر را انجام دهید.

۱- داده‌ها را صعودی مرتب کنید.

۲- داده‌های غیرمتعارف را حذف کنید.

۳- داده اول باقی‌مانده را به تعداد داده‌هایی که از ابتدا و داده آخر باقی‌مانده را به تعداد داده‌هایی که از انتها حذف کرده‌اید تکرار کنید.

۴- میانگین داده‌های جدید را بیابید.

مثال: میانگین وینزوری را برای داده‌های ۱۰۰,۹۵,۹۰,۲,۳,۱۸۵,۱۰۲ بیابید.

حل: ابتدا داده‌ها را صعودی مرتب می‌کنیم: ۲,۳,۹۰,۹۵,۱۰۰,۱۰۲,۱۸۵

- داده‌های ۱۸۵,۳,۲ نامتعارف هستند آنها را حذف می‌کنیم. (۹۰,۹۵,۱۰۰,۱۰۲)

- چون دو داده از ابتدا و یک داده از انتها حذف کردیم ۹۰ را دوبار و ۱۰۲ را یکبار دیگر تکرار می‌کنیم.

(۹۰,۹۰,۹۰,۹۵,۱۰۰,۱۰۲,۱۰۲)

- حال میانگین این داده‌ها را می‌یابیم.

$$\frac{۹۰+۹۰+۹۰+۹۵+۱۰۰+۱۰۲+۱۰۲}{۷}=\frac{۶۶۹}{۷}=۹۵/۵۷$$

نکته: انحراف چارکی به‌صورت $\frac{Q_۳-Q_۱}{۲}$ تعریف می‌شود که گاهی آن را نیم‌دامنه چارکی (SIQR) هم می‌نامند، که شاخص بسیار مهمی است.

۲-۴-۱ - پارامترهای پراکندگی

برای تحلیل داده‌ها علاوه‌بر پارامترهای مرکزی نیاز به پارامترهای دیگری داریم که پراکندگی یا انحراف داده‌ها را نسبت به میانگین و یا میانه و یا .... بسنجد.

الف) دامنه تغییرات: دامنه تغییرات را با R نمایش داده و از رابطه زیر محاسبه می‌کنیم:

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

برای مثال دامنه تغییرات در داده‌های ۱۵,۱۶,۱۷,۱۷,۲۰,۲۲,۶۰ برابر است با:

$$R=۶۰-۱۵=۴۵$$

این شاخص پراکندگی (دامنه تغییرات) هرچه بیشتر باشد به مفهوم آن است که پراکندگی بیشتر است. هرچند محاسبه آن ساده است ولی دارای ثبات زیادی نیست زیرا فقط به دو داده (بزرگترین و کوچکترین داده) بستگی دارد.

ب) انحراف متوسط از میانگین: انحراف متوسط از میانگین از رابطه $\sum\frac{(x_i-\mu)}{N}$ حاصل می‌شود ولی این پارامتر همواره برای هر سری داده، برابر صفر است پس نمی‌تواند شاخص مناسبی باشد.

ج) قدر مطلق انحراف متوسط از میانگین:

این شاخص را از رابطه $\sum\frac{|x_i-\mu|}{N}$ محاسبه می‌کنند ولی در عمل (به خاطر وجود قدر مطلق) کارکردن با آن دشوار است. مثلاً برای مواردی که نیاز به مشتق گرفتن از آن باشد.

د) واریانس[1]

واریانس یک سری داده‌ها، میانگین مربعات انحراف‌های این مقادیر از میانگین آنها می‌باشد. به‌عبارت دیگر، اگر واریانس جامعه را با $\sigma_x^۲$ نشان دهیم، آنگاه داریم:

1 - Variance

$$\sigma_x^2 = V(x) = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N} = \frac{\sum_{i=1}^{N}x_i^2}{N} - \left(\frac{\sum_{i=1}^{N}x_i}{N}\right)^2 = \mu_{x^2} - (\mu_x)^2$$

اگر داده‌ها به‌صورت جدول پیوسته فراوانی‌ها نمایش داده شوند، آنگاه داریم:

$$\sigma_x^2 = V(x) = \frac{\sum_{i=1}^{k}F_i(\tilde{x}_i-\bar{x})^2}{\sum_{i=1}^{N}F_i = N} = \frac{\sum_{i=1}^{k}F_i\tilde{x}_i^2}{N} - \left(\frac{\sum_{i,1}^{k}F_i\tilde{x}_i}{N}\right)^2$$

که در این‌جا $\tilde{x}_i$ ها ، همان نماینده یا مرکز طبقات می‌باشند.

☑**تست: اگر $\bar{x}$ میانگین n داده آماری x_i باشد حاصل $\sum_{i=1}^{n}(x_i-\bar{x})^2 - \sum_{i=1}^{n} x_i^2 + n\bar{x}^2$ کدام است؟**

(کارشناسی ارشد- سراسری مدیریت و حسابداری)

۱) صفر **۲) همواره مثبت**

۳) نیمه واریانس **۴) انحراف متوسط از میانگین**

✓**پاسخ:** گزینه ۱ صحیح است.

$$\text{عبارت داده شده} = A \xrightarrow{\text{طرفین} \times \frac{1}{n}} \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} - \sum_{i=1}^{n}\frac{x_i^2}{n} + \bar{x}^2 = \frac{A}{n}$$

$$\Rightarrow \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} - \left(\sum_{i=1}^{n}\frac{x_i^2}{n} - \bar{x}^2\right) = \frac{A}{n} \Rightarrow \text{واریانس} - \text{واریانس} = \frac{A}{n} \quad \Rightarrow \text{صفر} = \frac{A}{n} \Rightarrow A = 0$$

خواص واریانس:

۱) واریانس اعداد ثابت برابر صفر است. $(\delta_b^2 = 0)$، ضمناً اگر واریانس یک سری داده صفر باشند، تمام آن داده‌ها با هم برابر هستند.

۲) $\delta_{ax}^2 = a^2\delta_x^2$

۳) $\delta_{ax+b}^2 = a^2\delta_x^2$

مثال. اگر در ۱۰ داده آماری $\sum_{i=1}^{10}x_i^2 = 240$ و $\sum_{i=1}^{10}x_i = 40$ ، حاصل $V(2x-1)$ را بیابید.

حل:

$$\delta_x^2 = V(x) = \sum_{i=1}^{10}\frac{x_i^2}{n} - \left(\sum_{i=1}^{10}\frac{x_i}{n}\right)^2 = \left(\frac{240}{10}\right) - \left(\frac{40}{10}\right)^2 = 24 - 16 = 8 \Rightarrow$$

$$V(2x-1) = 4V(x) = 4(8) = 32$$

نکته: به جذر واریانس، σ، انحراف معیار می‌گوییم. با استفاده از انحراف معیار و میانگین می‌توان ضریب تغییرات یا ضریب پراکندگی را به‌صورت زیر تعریف کرد که یک شاخص پراکندگی است.