فصل دهم
دسته‌بندی داده‌ها و تجزیه و تحلیل آن‌ها

مقدمه

پژوهشگر، پس از مشخص ساختن ابزار مناسب برای جمع آوری داده‌ها و به کارگیری آن‌ها، به تجزیه و تحلیل داده‌ها می‌پردازد تا تکلیف فرضیه‌های پژوهش را، معین سازد. برای تجزیه و تحلیل داده‌های جمع‌آوری شده و آزمودن فرضیه‌ها، باید از تکنیک‌ها و فنون آماری مناسب با داده‌ها، استفاده کرد. پژوهشگر برای پاسخ‌گویی به پرسش پژوهش، از طریق آزمون فرضیه‌ها، داده‌های جمع‌آوری شده را به قسمت‌ها و بخش‌های گوناگون تجزیه می‌کند، اما تجزیه‌ی داده‌های پژوهش به خودی خود پاسخ پرسش‌ها را ارائه نمی‌دهد، بنابراین تحلیل داده‌ها ضروری است. تحلیل کردن، یعنی معنا بخشیدن به داده‌ها و تحلیل داده‌های خام، دشوار یا غیرممکن است. بنابراین، پژوهشگر باید ابتدا داده‌ها را تجزیه کند و سپس به تحلیل آن‌ها بپردازد.

تجزیه و تحلیل

به‌طور کلی، تجزیه و تحلیل[1]، عبارت است از روشی، که از طریق آن، کل فرایند پژوهشی، از انتخاب مسأله تا دسترسی به یک نتیجه هدایت می‌شود. تجزیه و تحلیل به معنی طبقه‌بندی، تنظیم، پردازش، دستکاری و خلاصه کردن داده‌ها برای یافتن پاسخ برای پرسش‌های پژوهش است. هدف از تحلیل، تقلیل داده‌ها به شکل قابل فهم و تفسیر است. این تفسیر به‌گونه‌ای باید بتواند روابط متغیرهای گوناگون مرتبط با مساله‌ی پژوهش را مورد مطالعه قرار دهد.

---

1. analyze

ضرورت تجزیه و تحلیل داده‌ها

هدف تحلیل داده‌ها روشن است، استفاده از داده‌ها برای پاسخ دادن به سؤالات پـژوهش. پژوهش، مجموعه‌ای از فعالیت‌ها است، که به کمک آن، اطلاعات مورد نیاز برای پاسخ دادن بـه سؤالات پژوهش یا آزمون فرضـیه‌هـا، گـردآوری مـی‌شـود. پژوهشـگر بایـد داده‌هـا و اطلاعـات گردآوری شده را تجزیه و تحلیل کند. بنابراین، تجزیه و تحلیل داده‌هـا بـه عنـوان یـک فراینـد علمی، یکی از مجموعه فعالیت‌های اساسی هر کار پژوهشی است. شـرط اساسـی و اولیـه بـرای انجام یک تحلیل درست آن است که، پژوهشگر با شیوه‌های آماری مناسب، یا بـه عبـارت دیگـر آزمون‌های آماری، آشنایی داشته باشد.

انواع روش‌های تجزیه و تحلیل

از آن‌جایی که اغلی مسائل در تعلیم و تربیت، خیلی گسترده و وسیع هستند، پژوهشـگر برای پاسخگویی به مسأله تدوین شده و یا تصمیم‌گیـری در مـورد رد یـا تأییـد فرضـیه‌ای کـه صورت بندی کرده است، از روش‌های مختلف تجزیه و تحلیل استفاده می‌کند. روش‌های تجزیـه و تحلیل داده‌ها به پنج طبقه تقسیم می‌شوند :

1. تجزیه و تحلیل توصیفی

در این نوع تجزیه و تحلیل، پژوهشگر داده‌های جمع‌آوری شده را با استفاده از شاخص‌های آمار توصیفی خلاصه و طبقه‌بندی می‌کند. به عبارت دیگر در تجزیه و تحلیل توصیفی، پژوهشگر ابتدا داده‌های جمع‌آوری شده را با تهیه و تنظیم جدول توزیع فراوانی خلاصه می‌کند و سپس به کمک نمودار آن‌ها را نمایش می‌دهد و از سایر شاخص‌های آمار توصیفی مانند انواع اندازه‌های مرکزی و پراکندگی آن‌ها را بررسی می‌کند. معروف‌ترین و در عین حال، پرمصرف‌ترین شاخص‌های آمار توصیفی عبارتند از : میانگین، میانه، انحراف استاندارد و واریانس. باید توجه داشت که در تجزیه و تحلیل توصیفی، از آزمون‌های آمار استنباطی استفاده نمی‌شود؛ زیرا در این نوع تجزیه و تحلیل، هدف، بررسی واقعیت‌های موجود است و پژوهشگر قصد پیش بینی یا استنباط آماری را ندارد.

2. تجزیه و تحلیل مقایسه‌ای

در تجزیه و تحلیل مقایسه‌ای، علاوه بر این که اطلاعات جمع‌آوری شده به صورت توصیفی مورد تجزیه وتحلیل قرار می‌گیرند، با یکدیگر نیز مقایسه می‌شوند. سؤالی که در این نوع تجزیه و تحلیل مطرح می‌شود، این است که، آیا شاخص آماری محاسبه شده، بزرگ‌تر یا کوچک‌تر از شاخص آماری دیگر است؟ به این معنی که دو یا چند شاخص آمار توصیفی(میانگین، میانه، انحراف استاندارد و واریانس)نسبت به همدیگر مورد مقایسه واقع می‌شوند. آزمون‌های آماری در تجزیه و تحلیل مقایسه‌ای جهت تصمیم‌گیری درباره‌ی آزمون فرضیه‌های تحقیق (تأیید یا عدم تأیید) به کار می‌روند.

3. تجزیه و تحلیل علّی(علیت)

در تجزیه و تحلیل علّی، روابط علت و معلولی بین متغیرهای مستقل و وابسته مورد بحث و بررسی قرار می‌گیرد در تجزیه و تحلیل علیت، ابزار یا شاخص‌های آماری به کار برده شده، مشابه شاخص‌هایی است که در تجزیه و تحلیل مقایسه‌ای مورد استفاده قرار می‌گیرد، با این تفاوت که در این نوع تجزیه و تحلیل‌ها، روابط علت و معلولی بین متغیرهای مستقل و وابسته مورد بررسی قرار می‌گیرد و پژوهشگر با رد یا تأیید فرضیه‌های آماری به تأیید یا رد ادعای وجود روابط علت و معلولی بین متغیرها می‌پردازد. تجزیه و تحلیل علیت برای پاسخ به سؤالات (چرایی) به کار می‌رود. تحلیل عاملی و تحلیل مسیر از جمله تجزیه و تحلیل علیت هستند.

4. تجزیه و تحلیل همبستگی یا همخوانی

در تحلیل همخوانی، درجه‌ی هم‌پراکنش دو متغیر مشخص می‌شود، یعنی، مشخص می‌شود که آیا اندازه‌های یک متغیر با اندازه‌های متغیر دیگر همخوانی دارد. برای مثال، آیا بین پایگاه شغلی و اعتقادات مذهبی آن‌ها همخوانی وجود دارد؟ روش‌های مختلف همبستگی، تحلیل واریانس و کواوریانس و رگرسیون از جمله تحلیل‌های همخوانی یا همبستگی هستند.

5. تجزیه و تحلیل استنباطی

آمار استنباطی برای دو هدف به کار می‌رود : برآورد ویژگی‌های جامعه (پارامتر) از روی ویژگی‌های نمونه (آماره) و پیش بینی روابط یا تفاوت(حسن‌زاده، 1392: 276).

مراحل تجزیه و تحلیل

به طور کلی تجزیه و تحلیل داده‌ها را می‌توان شامل مراحل زیر دانست :

1. ویرایش داده‌ها
2. مناسب سازی سؤال‌های بی‌پاسخ
3. کدگذاری داده‌ها
4. طبقه‌بندی داده‌ها
5. ایجاد پرونده (فایل) داده‌ها
6. تنظیم برنامه‌ی تجزیه و تحلیل(خاکی، 1391: 213).

روش‌های آماری

روش‌های آماری به دو بخش آمار توصیفی و آمار استنباطی تقسیم می‌شود.

1. آمار توصیفی

به آن قسمت از علم آمار که به جمع‌آوری، مرتب و خلاصه کردن اطلاعات و محاسبه شاخص‌ها می‌پردازد و ویژگی‌های کلی بسیاری از داده‌ها را در قالب یک عدد، جدول یا نمودار بیان می‌کند، آمار توصیفی می‌گویند.

روش‌های آمار توصیفی

آمار توصیفی شامل روش‌های زیر می‌باشد :

تهیه و تنظیم جدول فراوانی و ترسیم نمودار

برای نظم و ترتیب دادن به داده‌ها و همچنین برای نتیجه گیری بهتر، مخصوصاً وقتی داده‌ها زیاد هستند، جدول فراوانی استفاده می‌کنند. جدول متشکل از دو ستون، که یکی

نشان‌دهنده‌ی حالات و دیگری نشان‌دهنده‌ی تعداد افراد، که دارای آن حالت هستند را جدول توزیع فراوانی می‌نامند.

مثال: در کلاس 2 نفر، نمره 13، 3 نفر، نمره 14، 3 نفر، نمره 15 و 4 نفر، نمره 16 گرفته‌اند. جدول توزیع فراوانی را تشکیل دهید.

حل : تعداد کل دانشجویان این کلاس 12 نفر می‌باشد.

x_i (نمره دانشجویان)	f_i (تعداد دانشجویان)
13	2
14	3
15	3
16	4
	12

هنگامی که تعداد داده‌های گسسته زیاد هستند، اغلب آن‌ها رادسته‌بندی می‌کنند، تا به فرم خلاصه شده، قابل مشاهده باشند.

أ. مراحل تشکیل یک جدول گروه‌بندی(دسته بندی)شده

1. تعین دامنه تغییرات: دامنه تغییرات که با R نشان داده می‌شود ازفرمول زیر محاسبه می‌شود.

کوچکترین داده – بزرگترین داده = دامنه تغییرات $R = X_{max} - X_{min}$

2. تعیین تعداد دسته‌ها k با در نظر گرفتن تعداد و تنوع داده‌ها و دامنه تغییرات آن‌ها با توجه به فرمول استرژ

$$K = 1 + 3.32 \text{ Log } n$$

مثال : برای تنظیم 100 داده آماری چند گروه مناسب خواهد بود؟

$$K = 1 + 3.32 \text{ Log } 100$$
$$1 + 3.32\ (2) = 7.64 \approx 8$$

3. تعیین فاصله دسته‌های I یا C : یعنی تعداد داده‌هایی که در هر دسته باید قرار داده شوند. اگر دامنه تغییرات R و تعداد دسته‌ها k را داشته باشیم، فاصله دسته‌ها از تقسیم دامنه بر تعداد دسته‌ها بدست می‌آید.

$$I = \frac{R}{k}$$

همچنین می‌توان k و R را از روابط زیر محاسبه نمود :

$$k = \frac{R}{I}$$

$$R = k.I$$

مثال : اگر دامنه تغییرات یک‌سری داده برابر 21 و تعداد دسته ها را 3 اختیار کنیم، فاصله دسته‌ها را تعیین کنید؟

$$I = \frac{۲۱}{۳}$$

4. تعیین نماینده دسته : مرکز یا نماینده هر دسته X_i ، برابر نصف مجموع حد پایین و حد بالای همان دسته است.

مثال 1: نماینده دسته [19-10] در یک جدول فراوانی برابر است با :

$$X_i = \frac{۱۰+۱۹}{۲} = ۱٤.٥$$

مثال 2 : فرض کنید در بررسی قد دانش‌آموزان یک کلاس، قد 10 نفر را اندازه گرفته و اعداد زیر بدست آمده است R و k را محاسبه کنید؟ 157، 128، 166، 165، 164، 154، 143، 126، 142، 144

$R = 166 - 126 = 40$ $1+3.32 \text{ Log } 10 = 4.32 \approx 5$

نکته : در دسته‌بندی داده‌ها، مثلاً در دسته (134-126] اعداد بزرگتر از 126 و کوچکتر از 134 که شامل 126 هم هست می‌شود ولی در دسته آخر [166-158] خود 166 نیز قبول می‌کنیم.

$$I = \frac{R}{k} = \frac{۴۰}{۵} = ۸$$

دسته‌ها	فراوانی f_i	مرکز دسته x_i
$[۱۲۶-۱۳۴)$	2	$\frac{۱۲۶+۱۳۴}{۲}=۱۳۰$
$[۱۲۶-۱۳۴)$	0	$\frac{۱۳۴+۱۴۲}{۲}=۱۳۸$
$[۱۴۲+۱۵۰)$	3	$\frac{۱۴۲+۱۵۰}{۲}=۱۴۶$
$[۱۵۰+۱۵۸)$	2	$\frac{۱۵۰+۱۵۸}{۲}=۱۵۴$
$[۱۵۸+۱۶۶]$	3	$\frac{۱۵۸+۱۶۶}{۲}=۱۶۲$
	n = 10	

ب. انواع فراوانی‌ها

1. فراوانی مطلق (fi)

تعداد دفعات تکرار یک داده(و یا تعداد دفعات تکرار داده‌های متعلق به هر دسته) را در اصطلاح فروانی مطلق آن داده یا دسته می‌نامند. بدیهی است که حاصل جمع فراوانی‌های مطلق برابر با حجم جامعه(یا نمونه) است.

$$\sum f_i = n$$

2. فراوانی نسبی (r_i یا f_{pi})

نسبت فروانی مطلق هر دسته به کل داده(حجم جامعه) را فروانی نسبی آن دسته می‌نامند.

نکته : مجموع فروانی‌های نسبی برابر یک می‌باشد.

$r_i = \frac{f_i}{n}$ یا f_{pi}

درصد فروانی نسبی (pi) : اگر فروانی نسبی f_{pi} را در عدد 100 ضرب کنیم، درصد فراوانی نسبی بدست می‌آید.

$$p_i = f_{pi} \times ۱۰۰$$

3. فراوانی تجمعی یا انباشته یا تراکمی (F_{ci})

فراوانی تجمعی هر دسته برابر است با مجموع فراوانی مطلق دسته به اضافه فراوانی‌های مطلق دسته‌های قبلی آن. یعنی :

$$F_{ci} = f_1 + f_2 + ... + f_i$$

نکته 1 : فراوانی تجمعی دسته اول با فراوانی مطلق همان دسته برابر است و فراوانی تجمعی دسته آخر برابر حجم جامعه(N) می‌باشد.

نکته 2 : فراوانی تجمعی نسبی F_{cpi} : اگر فراوانی تجمعی هر دسته را به تعداد کل داده (N)تقسیم کنیم فراوانی تجمعی نسبی بدست می‌آید.

نکته 2 : درصد فراوانی تجمعی نسبی P_{ci} : هرگاه فراوانی تجمعی نسبی هر دسته را در عدد 100 ضرب کنیم درصد فراوانی تجمعی نسبی آن دسته بدست می‌آید.

ج. نمودارها

نمودارها، روش تصویری مناسبی برای توصیف و نمایش داده‌های جمع‌آوری شده هستند.

توزیع فراوانی را می‌توان با چند نوع نمودار نمایش داد، مانند : ستونی، چند ضلعی، دایره‌ای و هیستوگرام.

از نمودارهای هیستوگرام و چند ضلعی زمانی استفاده می‌شود، که متغیر مورد اندازه‌گیری پیوسته است. برای رسم هیستوگرام حدود واقعی طبقات را روی محور عمودی تعیین می‌کنیم. ترسیم نمودار چند ضعلی همانند هیستوگرام است با این تفاوت که، در نمودار چند ضلعی به جای حدود واقعی طبقات نقطه میانی هر طبقه (مرکز طبقه) به کار برده می‌شود.

1. شاخص‌های مرکزی

شاخص‌های عددی، پارامترهایی هستند که، برای مقایسه، بین چند جامعه به کار می‌روند. این اعداد خصوصیات کلی جامعه نمونه را برای صفت متغیر بیان می‌کنند (نیکوکار، 1386: 11).

شاخص های مهم مرکزی عبارتند از :

أ. میانگین

1. میانگین حساب ساده ($\bar{x}$)

میانگین حسابی ساده همان معدل معمولی است.

نکته 1: اگر فراوانی داده‌ها برابر یک باشند یا دسته‌بندی‌نشده باشند، میانگین را از فرمول زیر بدست می‌آوریم : (n تعداد داده‌هاست).

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

نکته 2 : اگر فراوانی داده‌ها بیشتر از یک باشند یا دسته‌بندی‌شده باشند، میانگین را از فرمول زیر حساب می‌کنیم.

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

نکته 3 : n تعداد داده‌ها، k تعداد دسته‌هاست، اگر داده‌ها دسته‌بندی‌شده باشند، در فرمول فوق به جای x_i از نشان دسته‌ها یا نماینده دسته‌ها استفاده می‌کنیم (عبدا...زاده، 1385: 42).

4، 3، 1، 9، 5، 6، 7، 4، 8، 2

حل :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}(2+8+4+7+6+5+9+1+3+4) = \frac{49}{10} = 4/9$$

مثال: میانگین داده‌های جدل فروانی زیر را بدست آورید.

حدود دسته ها	f_i	x_i	$f_i x_i$
[3-7)	4	5	20
[7-11)	8	9	72
[11-15)	6	13	78
[15-19)	7	17	119
	n =25		289

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i = \frac{۲۸۹}{۲۵} = ۱۱/۵۶$$

2. میانگین حسابی موزون(وزنی)

اگر داده‌های $x_1, x_2, ..., x_n$ به ترتیب دارای ضرایب وزنی $w_1, w_2, ..., w_n$ باشند، آن‌گاه از فرمول زیر، میانگین وزنی را حساب می‌کنیم.

$$\bar{X}_w = \frac{\sum w_i x_i}{\sum w_i}$$

مثال : نمونه معروف این میانگین، محاسبه معدل دروس اخذ شده توسط دانشجو در یک ترم می‌باشد. دانشجویی در یک ترم، درس فیزیک را نمره 12، ریاضی را نمره 15، ادبیات را 14 گرفته است. اگر فیزیک 3 واحد، ریاضی 2 واحد و ادبیات 1 واحد باشد، معدل او را محاسبه کنید.

$$\bar{x} = \frac{۳\times۱۲ + ۲\times۱۵ + ۱\times۱۴}{۳+۲+۱} = \frac{۸۰}{۶} = ۱۳/۳$$

3. میانگین هندسی(G_m)

میانگین هندسی برابر است با، ریشه n ام حاصلضرب n داده. از این میانگین برای محاسبه‌ی حد متوسط شاخص‌ها، نسبت‌ها و درصدها در یک دوره زمانی استفاده می‌کنیم. برای مثال، رشد جمعیت، کشت باکتری و غیره. اگر داده های $x_1, x_2, ..., x_n$ دسته‌بندی‌شده باشند :

$$G_m = \sqrt[n]{x_1 \times x_2 \times ... \times x_n}$$

$$G_m = \sqrt[n]{{x_1}^{f_i} \times {x_2}^{f_i} \times ... \times {x_k}^{f_k}}$$

نکته : اگر در بین داده‌ها، مقادیر صفر و منفی وجود داشته باشند، نمی‌توان از میانگین هندسی استفاده کرد.

مثال: میزان تولید کارخانه‌ای در چهار سال متوالی 2، 4، 6، 27 برابر نسبت به سال قبل می‌باشد، مطلوب است محاسبه میزان افزایش متوسط تولید؟

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

مثال: در بیست سال گذشته نرخ افزایش جمعیت شهری مطابق جدول زیر بوده است.

در این دوره به طور متوسط هر سال چند درصد بر جمعیت این شهر افزوده شده است؟

نرخ افزایش جمعیت x_i	٪4-4/19	٪4/2-4/39	٪4/4-4/59	٪4/6-4/79	٪4/8-4/99
تعداد سال‌ها (f_i)	2	4	5	8	1

حل :

$$G_m = \sqrt{(۰/۰۴۰۹۵)^{۲} \times (۰/۰۴۲۹۵)^{۴} \times (۰/۰۴۴۹۵)^{۵} \times (۰/۰۴۶۹۵)^{۸} \times (۰/۰۴۸۹۵)^{۱}} = ۰/۰۴۵ = \%۴/۵$$

4. میانگین هارمونیک (توافقی، همساز، معکوس)

میانگین هارمونیک برابر است با، عکس میانگین عکس داده‌ها. هرگاه کار ثابتی با سرعت‌های متفاوت انجام گرفته و محاسبه متوسط سرعت انجام کار، مورد نظر باشد، مناسب‌ترین معیار مرکزی برای توصیف آن‌ها، میانگین هارمونیک است، با این شرط که $x_k,...,x_۲,x_۱$ همگی غیر صفر و هم‌علامت باشند(به بیان دیگر هرگاه داده‌ها بصورت $\frac{۱}{x_k},...,\frac{۱}{x_۲},\frac{۱}{x_۱}$ باشند) میانگین هارمونیک از رابطه زیر بدست می‌آید.

$$H = \frac{n}{\sum\left(\frac{f_i}{x_i}\right)}$$

اگر در حالت $f_۱ = f_۲ = ... = f_k = ۱$ باشد، داریم :

$$H = \frac{n}{\sum\left(\frac{۱}{x_i}\right)}$$

(نیکوکار و مقسمی، 1386: 19).

مثال 1 : اتومبیلی محیط یک مزرعه مربعی شکل را که هر ضلع آن 4 کیلومتر است به شرح زیر می‌پیماید :

ضلع شمالی را با سرعت 80، ضلع شرقی را با سرعت 120، ضلع جنوبی را با سرعت 90، ضلع غربی را با سرعت 60 کیلومتر در ساعت، متوسط سرعت اتومبیل را حساب کنید.

$$H = \frac{n}{\sum\left(\frac{۱}{x_i}\right)} = \frac{۴}{\frac{۱}{۸۰}+\frac{۱}{۱۲۰}+\frac{۱}{۹۰}+\frac{۱}{۶۰}} = ۸۲/۳ km/h$$

مثال 2: از بین ماشین‌نویس‌های وزارتخانه‌ای، 40 ماشین‌نویس را بطور تصادفی برگرفته و مشاهده کرده‌ایم، که یک متن ثابت 200 کلمه‌ای را 12 نفر آن‌ها در 4 دقیقه، 10 نفر آن‌ها در 5 دقیقه، 18 نفر آن‌ها در 6 دقیقه تایپ کردند. میانگین زمان لازم برای تایپ چنین متنی را در نمونه مورد مشاهده حساب کنید.

$$H = \frac{n}{\sum\left(\frac{f_i}{x_i}\right)} = \frac{۴۰}{\left(\frac{۱۲}{۴}+\frac{۱۰}{۵}+\frac{۱۸}{۶}\right)} = \frac{۴۰}{۸} = ۵$$

ب. میانه

اگر داده‌ها را بطور صعودی یا نزولی مرتب کنیم، میانه اندازه‌ای از صفت متغیر است، که در وسط قرار گرفته باشد. میانه را با نماد M_d نشان می‌دهیم.

1. میانه در سری داده‌ها

اگر پس از مرتب کردن داده‌ها، تعداد داده‌ها فرد باشد، میانه داده‌ای است که در وسط قرار می‌گیرد، ولی اگر تعداد داده‌ها زوج باشد، در این صورت، میانه برابر با میانگین دو عدد وسطی است (حسینی‌پور، 1389: 33).

مثال 1 : میانه دادهای زیر را تعیین کنید. 9، 6، 3، 7، 5

حل : ابتدا داده‌ها را به صورت صعودی مرتب می‌کنیم. 9، 7، 6، 5، 3

چون تعداد داده‌ها فرد است، پس میانه عدد وسطی یعنی 6 می‌باشد.

$$M_d = ۶$$

مثال2 : میانه داده‌های زیر را بدست آورید. 7،4، 6، 9، 3، 10

حل : ابتدا مرتب می‌کنیم 10، 9، 7، 6، 4، 3 ، چون تعداد داده‌ها زوج است :

$$M_d = \frac{۶+۷}{۲} = ۶/۵$$

2. میانه در توزیع ساده

اگر n فرد باشد، داده شماره $\frac{n+۱}{۲}$ را میانه تلقی می‌کنیم :

$$M_d = X\,\frac{n+۱}{۲}$$

اگر n زوج باشد، داده شماره $\frac{n}{۲}$ را میانه تلقی می‌کنیم :

$$M_d = X\,\frac{n}{۲}$$

برای این منظور ابتدا باید ستون فراوانی تجمعی را تشکیل دهیم(عبدا...زاده، 1385: 49).

مثال : برای داده‌های جدول زیر میانه را حساب کنید.

x_i	2	5	7	10	11
f_i	1	2	5	4	5

حل : ابتدا ستون فروانی تجمعی را بدست می‌آوریم.

چون n فرد است داریم :

$$M_d = X\,\frac{n+۱}{۲} = X\,\frac{۱۷+۱}{۲} = X_۹ = ۱۰$$

x_i	f_i	F_i
2	1	1
5	2	3
7	5	8
10	4	12
11	5	17
	n =17	

3. میانه در توزیع دسته‌بندی‌شده

ابتدا ستون تجمعی را تشکیل می‌دهیم، پس عدد $\frac{n}{۲}$ را با فراوانی‌های تجمعی مقایسه می‌کنیم. اولین طبقه‌ای که فروانی تجمعی آن بزرگتر یا مساوی $\frac{n}{۲}$ باشد را، به عنوان طبقه‌ی میانه‌دار انتخاب و میانه را از فرمول زیر محاسبه می‌کنیم :

$$M_d = L_i + \left(\frac{\frac{n}{۲} - F_{ci-۱}}{f_i} \right) \times I$$

L_i کرانه‌ی پایین طبقه میانه‌دار، $F_{ci-۱}$ فروانی تجمعی طبقه قبل از طبقه میانه‌دار، f_i فروانی مطلق طبقه میانه‌دار و I فاصله‌ی طبقات است.

مثال : در جدول زیر میانه را محاسبه کنید.

حدود طبقات	60-62	63-65	66-68	69-71	72-74
حدود واقعی طبقات	59/5-62/5	62/5-65/5	62/5-68/5	68/5-71/5	71/5-74/5
f_i	5	18	42	27	8
F_{ci}	5	23	65	92	100

حل : $n = \frac{۱۰۰}{۲} = ۵۰$، اولین طبقه‌ای که فروانی تجمعی آن بزرگتر یا مساوی 50 است، طبقه سوم می‌باشد.

$$L_i = ۶۵/۵, \frac{n}{۲} = ۵۰, F_{ci-۱} = ۲۳, f_i = ۴۲, I = ۳ \Rightarrow M_d = ۶۵/۵ + \frac{۵۰ - ۲۳}{۴۲} \times ۳ = ۶۷/۴$$

ج. مد یا نما

در یک مجموعه از داده‌ها، مد یا نما داده‌ای است که، فراوانی تکرار آن از سایر داده‌ها بیشتر است، که با M_o نشان می‌دهیم.

مثال 1 : در داده‌های روبرو مد کدام است؟ 1، 5، 9، 5، 6، 9، 7، 5

حل : $M_o = ۵$

برای محاسبه مد در داده‌های گروه‌بندی‌شده، ابتدا طبقه‌ای که فروانی آن ماکزیمم است را مشخص و آن را طبقه مددار می‌نامیم، سپس مد را از فرمول زیر محاسبه می‌کنیم :

$$M_o = L_i + \left(\frac{d_i}{d_۱ + d_۲} \right) \times I$$

که در آن L_i کران پایین طبقه مددار، $d_۱$ تفاضل فراوانی مطلق طبقه مددار از فروانی طبقه قبل، $d_۲$ تفاضل فروانی مطلق طبقه مددار از فروانی طبقه بعد و I فاصله طبقه‌ها است(نیکوکار و مقسمی، 1386: 25).

مثال 2 : در جدول زیر مد را تعیین کنید :

حدود طبقات	60-62	63-65	66-68	69-71	72-74
حدود واقعی طبقات	59/5-62/5	62/5-65/5	65/5-68/5	68/5-71/5	71/5-74/5
f_i	5	18	42	27	8

حل : چون طبقه سوم بیشترین فراوانی را دارد آن را طبقه مددار می‌گیریم.

$L_i = ٦۵/۵$

$$d_۱ = ۴۲ - ۱۸ + ۲۴,\ d_۲ = ۴۲ - ۲۷ = ۱۵,\quad I = ۳,\quad M_o = ۶۵/۵ + \frac{۲۴}{۲۴ + ۱۵} \times ۳ = ۶۷/۳۵$$