آمار توصیفی

❍ **علم آمار:** روشی علمی است که برای جمع‌آوری، تلخیص، تجزیه و تحلیل، تفسیر و به طور کلی مطالعه و بررسی مشاهدات به کار برده می‌شود.

- **جامعهٔ آماری:** تعدادی از عناصر جامعه که حداقل دارای یک صفت مشخصه باشند، جامعه آماری را تشکیل می‌دهند.

جامعه آماری به دو دسته تقسیم می‌شود:
۱) محدود: تعداد افراد جامعه محدود است.
۲) نامحدود: تعداد افراد جامعه نامحدود است.

- **صفت مشخصه:** صفتی است که بین همه عناصر جامعهٔ آماری مشترک و متمایز کننده آن جامعهٔ آماری از سایر جوامع می‌باشد.

- **صفت:** کمیت یا کیفیتی است که متعلق به عناصر جامعه آماری بوده و همواره به دو بخش تقسیم می‌شود: صفت مشترک (ثابت) و صفت متغیر

۱) صفت مشترک (ثابت): صفتی است که بین همه عناصر جامعه آماری مشترک است. مانند صفت دانش‌آموز بودن برای دانش‌آموزان یک کشور.

۲) صفت متغیر: خاصیتی است که افراد یک جامعه را از یکدیگر متفاوت، جدا و مشخص می‌سازد. و از فردی به فرد دیگر می‌تواند تغییر کند. مانند صفات: قد، سن، وزن، . . . .

صفات متغیر به دو دسته کمی و کیفی تقسیم می‌شوند.

صفات متغیر کمی: در صفات کمی امکان اندازه‌گیری و بیان یک عدد واحددار مانند: کیلومتر ـ کیلو و ... وجود دارد. که به دو دسته : پیوسته (قد ـ وزن) و گسسته (تعداد اعضاء خانواده) تقسیم می‌شود.

صفات متغیر کیفی : در صفات کیفی امکان اندازه‌گیری با ابزارهای رایج وجود نداشته و نمی‌توان آن را به صورت عددی واحددار بیان نمود. مانند: رشته تحصیلی ـ رنگ پوست ـ کیفیت و مرغوبیت کالا

- **مقیاس:** با توجه به نوع صفات کیفی و کمی، مقیاس‌های متفاوتی برای اندازه‌گیری متغیرها وجود دارد. انواع مقیاس‌ها به شرح زیر است:

۱ـ اسمی (طبقه‌ای)

۲ـ رتبه‌ای (ترتیبی)

۳ـ فاصله‌ای

۴ـ نسبی (نسبتی)

۱ـ مقیاس اسمی(طبقه‌ای): ضعیف‌ترین شکل اندازه‌گیری است که در آن از اعداد و علائم برای طبقه‌بندی اشیاء، اشخاص یا خصوصیت استفاده می‌شود. مانند مشخص کردن سازمان‌ها با اسم‌های A و B و C و ... .

این نوع مقیاس به علت ضعف در اندازه‌گیری، در صفات کیفی استفاده می‌شود.

۲ـ مقیاس رتبه‌ای (ترتیبی): در مواردی صرف‌نظر از محتویات یک طبقه یا گروه با طبقه یا گروه دیگر نوعی ارتباط بین آن‌ها برقرار است، این روابط با توجه به نوع مقیاس نشان‌دهندهٔ حالت ترتیبی است. برای مثال طبقه‌بندی افراد جامعه به صورت (پردرآمد ـ متوسط ـ کم درآمد)، (قوی ـ متوسط ـ ضعیف)،(بالا ـ وسط ـ پایین)، (بزرگتر ـ مساوی ـ کوچکتر) نشان‌دهنده حالت ترتیبی است.

این نوع مقیاس نیز به علت ضعف در اندازه‌گیری، در صفات کیفی مورد استفاده قرار می‌گیرد.

۳ـ مقیاس فاصله‌ای: وقتی یک مقیاس همهٔ خصوصیات مقیاس ترتیبی را داشته و به علاوه فاصله بین هر دو عدد نیز در آن مشخص باشد، به یک مقیاس قوی‌تر رسیده‌ایم که می‌تواند برای صفات کمی مورد استفاده قرار گیرد.

در این مقیاس، صفر به صورت قراردادی و اختیاری است. مانند: سانتی‌گراد و فارنهایت که دارای صفرهای قراردادی مختلفی هستند.

و نسبت هر دو فاصله مستقل از واحد اندازه‌گیری و مستقل از نقطه صفراست.

۴ـ مقیاس نسبتی (نسبی): دقیق‌ترین مقیاس برای صفات کمی است که علاوه بر داشتنِ تمام خصوصیات مقیاس فاصله‌ای ، دارای صفر واقعی نیز می‌باشد. مقیاس‌هایی مثل: پوند، گرم، متر، اینچ دارای صفر واقعی هستند. همچنین نسبت هر دونقطه دلخواه مستقل از واحد اندازه‌گیری است (مانند مقیاس فاصله‌ای). برای مثال وزن دو شیء مختلف را می‌توان هم با گرم و هم با پوند اندازه‌گیری کرد. اما نسبت آن‌ها با هم فرقی نمی‌کند و به واحد اندازه‌گیری ربطی ندارد.

● **نمونه:** به هر بخش از جامعه آماری محدود یا نامحدود یک نمونه گفته می‌شود یا به عبارت دقیق‌تر به تعداد محدودی از اعضای جامعه آماری که بیان کننده تمام ویژگی‌های جامعه اصلی باشند، نمونه گویند.

با توجه به تعاریف جامعهٔ آماری و نمونه، حال می‌توانیم آماره و پارامتر را تعریف کنیم.

● **آماره:** اصطلاحی است که در مورد نمونه استفاده می‌شود و خصوصیتی از آن را بررسی می‌کند. مانند: میانگین نمونه $(\overline{x})$، واریانس نمونه (S^2)،نسبت نمونه $(\overline{p})$.

نکته: هر آماره یک متغیر تصادفی است، چرا که از یک نمونه به نمونه دیگر تغییر می‌کند.

● **پارامتر:** عددی است که خصوصیتی از یک جامعه را بیان می‌کند مانند: میانگین جامعه (μ)، واریانس جامعه (σ^2) ، میانه (Md).

نکته: پارامترها در جامعه ثابت هستند ولی مجهول و باید آن‌ها را از طریق آماره‌ها در نمونه‌گیری تخمین بزنیم.

● **پارامترها و آماره‌های مهم:**

شاخص	گروه	نماد کلی	میانگین	واریانس	نسبت
آماره	نمونه	$\hat{\theta}$	$\overline{x}$	S^2	$\overline{P}$
پارامتر	جامعه	θ	μ	σ^2	P

تفاوت در نوع و همچنین کاربردهای شاخص آماره و پارامتر، موجب تقسیم‌بندی آمار به آمار توصیفی و آمار استنباطی (استنتاجی) شده است.

آمار از نظر موضوعی به سه بخش تقسیم می‌شود: آمار توصیفی، آمار استنتاجی(استنباطی)، آمار ناپارامتریک

۱) آمار توصیفی: این آمار به توصیف جامعه می‌پردازد و هدف آن محاسبه پارامترهای جامعه است. چنانچه محاسبه مقادیر و شاخص‌های آماری برای جامعه از طریق سرشماری تمامی عناصر انجام گیرد، به آن آمار توصیفی گفته می‌شود.

۲) آمار استنتاجی (استنباطی): به قسمتی از آمار که می‌تواند نتایج حاصل از تجزیه و تحلیل نمونه را به جامعه تعمیم دهد، آمار استنتاجی گفته می‌شود. به عبارت دیگر آماره‌ها از طریق نمونه‌گیری بدست می‌آیند، سپس به کمک تخمین (برآورد) و آزمون فرض، به پارامترهای جامعه تعمیم داده می‌شوند.

۳) آمار ناپارامتریک: در مقابل آمار پارامتریک قرار دارد. در آمار پارامتریک فرض اساسی برخوردار بودن مشاهدات از توزیع نرمال است. در آمار ناپارامتریک، فرض فوق وجود نداشته و بیشتر متغیرها با مقیاس کیفی سنجیده می‌شوند و آزاد از توزیع هستند. در حقیقت در این آمار به هیچ توزیعی وابستگی وجود ندارد.

• نمونه گیری

در بسیاری از موارد پژوهشگران به دنبال تعیین پارامترهای جامعه (مانند: میانگین جامعه (μ) ، واریانس جامعه (σ^2)) هستند. البته به‌طور طبیعی این عمل امکان‌پذیر نیست. به همین دلیل با استفاده از نمونه‌گیری به استنباط پارامترهای جامعه آماری می‌پردازند.

نکته: اگر پارامتر را **شاخص** بدست آمده از طریق نمونه‌گیری بنامیم، به این شاخص در یک نمونه n تائی آماره می‌گوئیم. گفتنی است آماره از یک نمونه به نمونه دیگر تغییر می‌کند. به همین علت برای رسیدن به یک پایائی و reliability باید به یک تقریب برای **توزیع نمونه‌گیری آماره** برسیم.

•توزیع آماره

تابع احتمالی است که از نمونه‌گیری مکرر حاصل می‌شود. در شکل کامل‌تر به آن، توزیع نمونه‌گیری آماره (Statistic Sampling Distribution : (SSD)) می‌گوئیم.

● دلایل نمونه‌گیری

۱ـ هزینه

۲ـ به روز بودن

۳ـ درستی و صحت

۴ـ صرفه‌جوئی در زمان

۵ـ آزمون تخریب‌کننده

نکته: انواع روش‌های نمونه‌گیری:

۱ـ تصادفی ساده:
- الف) قرعه‌کشی
- ب) جدول اعداد تصادفی

۲ـ نمونه‌گیری منظم (Systematic Sampling)

۳ـ نمونه‌گیری گروهی (Stratified Sampling)

۴ـ نمونه‌گیری خوشه‌ای (Cluster sampling)

۵ـ نمونه‌گیری مرحله‌ای (Stage Sampling)

دقت کنید، صرف‌نظر از این‌که چه روش آماری برای استنباط آماری موردنظر است، قدرت آن به روش بکار رفته برای انتخاب نمونه بستگی دارد. در صورتی که نمونه موردنظر نماینده واقعی جامعه نباشد، نمونه موردنظر دارای **اریب (bias)** است. در این حالت پیش‌بینی صحیح و دقیق درباره پارامترهای جامعه امکان نخواهد داشت.

۱ـ نمونه‌گیری تصادفی ساده

در این حالت هر یک از عناصر جامعه برای انتخاب شدن شانس مساوی دارند (هم تراز هستند). در این حالت افراد یا اشیاء به طور تصادفی از لیست تهیه شده از جامعه انتخاب می‌شوند و باید دارای ویژگی‌هائی همانند ویژگی‌های همان جامعه‌ای که از آن انتخاب می‌شوند، باشند.

که این راه به دو روش : قرعه‌کشی و جدول اعداد تصادفی انجام می‌شود.

۲ـ نمونه‌گیری منظم (سیستماتیک)

در این روش، شکل تغییر یافته حالت تصادفی ساده به کار گرفته می‌شود. یک نقطه از فهرست افراد یا اشیاء جامعه را به طور تصادفی انتخاب می‌کنیم و بعد از آن نمونه موردنظر را به صورت منظم پشت سر هم انتخاب می‌کنیم. این روش برای آن دسته از جوامع آماری که کد از پیش تعیین شده و مرتبی دارند، کاربرد فراوان دارد با مشخص شدن اولین عضو بقیه اعضا نیز مشخص می‌شوند این خاصیت از یک سو یک حسن است اما چون شانس را از بقیه اعضا می‌گیرد عیب محسوب می‌شود. مانند: شماره کارمندی، شماره دانشجوئی

۳ـ نمونه‌گیری گروهی

در این روش جامعه را به گروه‌های متجانس تقسیم و هر گروه دارای ویژگی‌های مشابهی هستند. پس از تقسیم جامعه به گروه‌های متجانس از هر گروه نمونه موردنظر به روش تصادفی ساده و منظم گرفته می‌شود. نکته مهم این است که در جوامعی مورد استفاده قرار می‌گیرد که از نظر صفت موردنظر ناهمگون است. مانند بررسی عملکرد واحدهای مختلف یک سازمان.

۴ـ نمونه‌گیری خوشه‌ای

هر گاه جامعه موردنظر خیلی وسیع و گسترده باشد مانند وضعیت معاش یا تحصیل یک شهر بزرگ یا یک کشور برای کارمندان،برای نمونه‌گیری ابتدا سازمان‌ها یا اداراتی را به روش تصادفی ساده یا سیستماتیک (منظم) انتخاب می‌کنیم سپس کارمندان موردنیاز را با استفاده از همین روش به دست می آوریم در این‌جا واحد نمونه‌گیری خوشه‌ای، سازمان بوده است.

۵ـ نمونه‌گیری مرحله‌ای

شکل گسترش یافته نمونه‌گیری خوشه‌ای است . در این حالت نمونه‌گیری از جامعه طی چند مرحله انجام می‌شود. یعنی انتخاب نمونه از نمونه دیگر، به طور مثال چند سازمان از شهر انتخاب می‌کنیم سپس از بین هر سازمان چند واحد را معین می‌کنیم سپس عناصر نمونه را به صورت تصادفی بدست می‌آوریم.

• داده‌های آماری

به اندازه‌های صفت متغیر عناصر جامعه یا نمونهٔ آماری که با استفاده از اندازه‌گیری، آزمایش، مشاهده و غیره به دست می‌آید، داده‌های آماری می‌گویند.

داده‌های آماری به دو صورت طبقه‌بندی شده و طبقه‌بندی نشده می‌باشند.

• طبقه‌بندی صفات

از آن‌جا که اطلاعات آماری به صورت اعداد و ارقام بیان می‌شوند، اگر بتوان آن‌ها را به صورت طبقه‌بندی شده بیان کرد، به راحتی می‌توان به خصوصیات مهم آن‌ها پی برد، این داده‌ها به دو دسته پیوسته و ناپیوسته تقسیم می‌شوند.

ـ به اعدادی که طبقات یک جدول توزیع فراوانی را مشخص می‌سازند، حدود طبقات می‌گویند.

ـ مرکز یک طبقه برابر نصف مجموع حد پائین و حد بالای آن طبقه است. که به آن نماینده طبقه یا متوسط طبقه نیز می‌گویند.

ـ طول طبقه یا دسته تفاوت بین حدود بالا یا پائین دو طبقه متوالی است.

نکته :پیوسته کردن جدول با طبقات ناپیوسته:

به طور کلی در جداول با طبقات گسسته کافی است عبارت $\left(\frac{\text{حد بالای طبقه} - \text{حد پائین طبقه‌بعدی}}{2} \right)$ محاسبه شده و مقدار محاسبه شده را از حد پایین طبقات کم و به حد بالای طبقات اضافه کنیم و طبقات را به صورت پیوسته ظاهر می‌کنیم.

مثال۲: به جدول اعداد طبقه‌بندی شده (ناپیوسته) زیر توجه کنید:

طبقات	$5-9$	$10-14$	$15-19$
فراوانی	3	7	4

جدول طبقه‌بندی شده ناپیوسته

$\rightarrow$

طبقات	$4.5-9.5$	$9.5-14.5$	$14.5-19.5$
فراوانی	3	7	4

جدول طبقه‌بندی شده پیوسته

توجه: در این جدول، حد بالای طبقه اول با حد پائین طبقه دوم برابر نمی‌باشند، اصطلاحاً به این‌گونه طبقه‌بندی، طبقه‌بندی ناپیوسته می‌گویند. برای پیوسته کردن این جدول لازم است به اندازه 0.5 واحد از حد پائین طبقات کم و به حد بالای آن‌ها اضافه نمائیم. بقیه محاسبات مانند مثال قبل است.

• انواع فراوانی‌ها

۱ـ فراوانی مطلق(F_i)

تعداد دفعات تکرار هر داده i را فراوانی مطلق دادهٔ i گوئیم که با F_i نشان داده می‌شود.

نکته :مجموع فراوانی‌های مطلق داده‌ها برابر با حجم جامعه می‌شود:

در صورتی‌که K داده مختلف داشته باشیم :

$$\boxed{\sum_{i=1}^{K} F_i = N}$$

۲ـ فراوانی نسبی (f_i)

نسبت فراوانی مطلق هر داده i به حجم جامعه را فراوانی نسبی می‌نامند.

$$\boxed{f_i = \frac{F_i}{N}}$$

نکته :مجموع فراوانی‌های نسبی برابر یک است .

$$\boxed{\sum_{i=1}^{k} f_i = f_1 + f_2 + ... + f_k = 1}$$

۳ـ درصد فراوانی نسبی

به حاصلضرب فراوانی نسبی هر داده i در 100، درصد فراوانی نسبی گفته می‌شود.

۴ـ فراوانی تجمعی (F_{C_i})

فراوانی تجمعی هر داده (طبقه) برابر است با جمع فراوانی مطلق همان طبقه به علاوه فراوانی‌های مطلق طبقات ماقبل آن :

$$\boxed{F_{C_i} = F_1 + F_2 + \cdots + F_i}$$

نکته :فراوانی تجمعی طبقه آخر برابر با حجم کل جامعه است.

۵ـ فراوانی نسبی تجمعی $\left(f_{C_i}\right)$

فراوان نسبی تجمعی درصد مشاهدات واقع شده بین حد پایین اولین طبقه و حد بالای i امین طبقه را نشان می‌دهد. یعنی: $\boxed{f_{c_i} = \frac{F_{c_i}}{N}}$

نکته :فراوانی نسبی تجمعی آخرین طبقه یک است.

مثال : به جدول زیر توجه کنید:

C − L	0 − 5	5 − 10	10 − 15	15 − 20
F_i	10	20	30	40
F_{C_i}	10	30	60	100
f_i	$\frac{10}{100}$	$\frac{20}{100}$	$\frac{30}{100}$	$\frac{40}{100}$
f_{c_i}	$\frac{10}{100}$	$\frac{30}{100}$	$\frac{60}{100}$	$\frac{100}{100}$

$\longrightarrow N = \sum_{i=1}^{K} F_i = 100$

$\longrightarrow$ حجم کل جامعه = فراوانی تجمعی طبقه آخر $= N = 100$

$\longrightarrow \sum_{i=1}^{k} f_i = \frac{10 + 20 + 30 + 40}{100} = 1$

$\longrightarrow$ فراوانی نسبی تجمعی آخرین طبقه $= 1$

❍ مشخص کننده‌های عددی

پارامترهایی هستند که برای مقایسه بین چند جامعه به کار می‌روند و به سه بخش تقسیم مـی‌شـوند: ۱ــ پـارامترهـای مرکـزی ، ۲ــ پارامترهای پراکندگی، ۳ـ پارامترهای نسبی پراکندگی

۱) پارامترهای مرکزی: هر معیاری که معرف مرکز مجموعهٔ داده‌ها باشد، پارامتر مرکزی نامیـده مـی‌شـود. از جملـه ایـن معیارهـا عبارتند از: میانگین ـ میانه ـ مد(نما) ـ چندک‌ها (چارک‌ها ـ دهک‌ها ـ صدک‌ها)

ـ میانگین: اصلی‌ترین شاخص مرکزی میانگین است. در حقیقت اگر داده‌ها روی یک محـور بـه صـورت مـنظم ردیـف شـوند، مقـدار میانگین دقیقاً نقطه تعادل و مرکز ثقل توزیع است.

❍ انواع میانگین

۱ـ میانگین حسابی $(\mu = \overline{x})$

۲ـ میانگین هندسی $(\overline{X}_G)$

۳ـ میانگین هارمونیک $(\overline{X}_H)$

۴ـ میانگین پیراسته

۵ـ میانگین وزنی $(\overline{X}_\omega)$

نکته: همیشه $\overline{X} > \overline{X}_G > \overline{X}_H$ است و فقط زمانی که داده‌ها با یکدیگر برابر باشند $\overline{X} = \overline{X}_G = \overline{X}_H$ خواهد بود.

۱ـ میانگین حسابی

معدل مجموعه‌ای از مشاهدات را میانگین حسابی می‌نامند که از تقسیم مجموع مشاهدات بر تعداد آن‌ها محاسبه می‌شود.

۱ـ محاسبه میانگین حسابی داده‌های طبقه‌بندی نشده و نیمه طبقه‌بندی شده

$$\boxed{\mu = \bar{X} = \frac{\sum_{i=1}^{n} x_i}{N} = \frac{\sum_{i=1}^{n} F_i x_i}{N} = \sum_{i=1}^{n} f_i x_i}$$

۲) محاسبهٔ میانگین حسابی داده‌های طبقه‌بندی شده: ابتدا مرکز طبقات را بدست آورده و سپس به فرم زیر عمل می‌کنیم:

$$\bar{X} = \frac{\sum F_i\ x_i}{N} = \sum f_i\ x_i$$ (x_i: مرکز طبقه)

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

مثال : مطلوبست میانگین جدول زیر:

C − L	0 − 5	5 − 10	10 − 15	
F_i	2	3	4	$\Sigma F_i = N = 9$

ابتدا مرکز طبقات را بدست می‌آوریم:

$$\text{مرکز هر طبقه} = \frac{\text{حد بالا + حد پائین}}{2}$$

x_i	$\overbrace{\frac{5+0}{2}}^{2.5}$	$\overbrace{\frac{5+10}{2}}^{7.5}$	$\overbrace{\frac{10+15}{2}}^{12.5}$
F_i	2	3	4

$$\Rightarrow \frac{\sum F_i x_i}{N} = \frac{(2\times 2.5)+(3\times 7.5)+(4\times 12.5)}{9} = \frac{5+22.5+50}{9} = \frac{77.5}{9}$$

• خواص میانگین حسابی

۱) مجموع انحرافات از میانگین همیشه صفر است. $\boxed{\sum (x_i - \mu) = 0}$

۲) مجموع مجذور انحرافات از میانگین همیشه می‌نیمم است. a عدد دلخواه است ; $\boxed{\sum\ (x_i - \mu)^2 < \sum\ (x_i - a)^2}$

۳) اگر a و b اعداد ثابتی باشند:

الف) $\mu(\pm a) = \pm a$ ب) $\mu(x \pm a) = \mu_x \pm a$ ج) $\mu_{(\pm ax)} = \pm a\mu_{(x)}$

د) $\mu\left(\frac{x}{\pm a}\right) = \frac{1}{\pm a}\ \mu_x$ هـ) $\mu(\pm x \pm y) = \pm\mu_x \pm \mu_y$ و) $\mu(\pm ax \pm b) = \pm a\mu_x \pm b$

۴) در جامعه آماری فقط یک میانگین داریم.

۵) مقادیر بزرگ و کوچک به سهم خود در میانگین سهم دارند.

۶) میانگین تنها پارامتری است که اگر به جای کلیه داده‌ها قرار گیرد، مجموع آن‌ها تغییری نمی‌کند، یعنی:

$$\underbrace{1,2,3}_{\text{مجموع}=6} \to \bar{x} = \frac{1+2+3}{3} = 2 \Rightarrow \underbrace{2,2,2}_{\text{مجموع}=6}$$

۲ـ میانگین هندسی (μ_G) :

اگر داده‌های بدست آمده نسبت،درصد، شاخص نرخ رشد و ... باشد، برای بدست آوردن مقدار متوسط از میانگین هندسی استفاده می‌کنیم:

برای داده‌های طبقه‌بندی نشده:

$$\boxed{\mu_G = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n} = \left(x_1 \times x_2 \times \ldots \times x_n\right)^{\frac{1}{n}}}$$

برای داده‌های طبقه‌بندی شده:

$$\boxed{\mu_G = \sqrt[\sum F_i]{x_1^{F_1} \times x_2^{F_2} \times \ldots \times x_n^{F_n}} = \left(x_1^{F_1} \times x_2^{F_2} \times \ldots \times x_n^{F_n}\right)^{\frac{1}{\sum F_i}}} \quad \text{یا} \quad \boxed{\mu_G = x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_n^{f_n}}$$

نکته :باید توجه داشت که اگر در بین داده‌ها عدد صفر یا منفی وجود داشته باشد، نمی‌توان از این روش استفاده کرد.

نکته :در میانگین هندسی اگر به جای نسبت از اعداد استفاده شود، برای به دست آوردن متوسط رشد و یا میانگین به درصد باید در انتها مقدار به دست آمده را از یک کسر کنیم، اگر فقط چند برابر شدن مدنظر سؤال باشد نیازی به این محاسبه نمی‌باشد.

مثال۳: فرض کنید شاخص قیمت خرده فروشی از 200 در سال ۷۸ به 450 در سال ۸۰ رسیده باشد، متوسط نرخ تورم در این فاصله زمانی چه بوده است؟

$$\bar{X}_G = \sqrt{\frac{\text{سال } 79}{\text{سال } 78} \times \frac{\text{سال } 80}{\text{سال } 79}} = \sqrt{\frac{450}{200}} = 1.5$$

$$1.5 - 1 = 0.5 = \%50$$

مثال۴: تعداد کارکنان کارخانه‌ای طی چهار سال متوالی 190، 180، 160، 150 بوده است. متوسط رشد سالانه تعداد کارکنان چند درصد است؟

$$\bar{X}_G = \sqrt[\text{تعداد سال‌ها} - 1]{\frac{\text{سال آخر}}{\text{سال اول}}} = \sqrt[4-1]{\frac{190}{150}} = 1.082 \rightarrow 1.082 - 1.000 = 0.082 = \%8.2$$

مثال۵: قیمت کالایی در سال گذشته 20% کاهش و امسال 20% افزایش داشته است. متوسط نرخ رشد قیمت این کالا در این دو سال چیست؟

با توجه به رابطه سوم میانگین هندسی و این‌که: در واقع این تغییرات برای سه سال متوالی صورت گرفته است داریم:

سال اول		سال دوم		سال سوم
x تومان	$\xrightarrow[x - 0.2x]{\text{کاهش } \%20}$	0.8x تومان	$\xrightarrow[0.8x + 0.2(0.8x)]{\text{افزایش } \%20}$	0.96 x تومان

$$\bar{X}_G = \sqrt[\text{تعداد سال‌ها} - 1]{\frac{\text{سال آخر}}{\text{سال اول}}} = \sqrt[3-1]{\frac{0.96x}{x}} = \sqrt[2]{0.96} \ \text{برابر} \ \rightarrow \left(\sqrt[2]{0.96} - 1\right) \text{درصد}$$

توجه شود که برای میانگین هندسی پیش فرض درصد است.

۳ـ میانگین هارمونیک (توافقی یا معکوس یا همساز، μ_H)

اگر مقیاس داده‌ها به صورت ترکیبی باشد از این میانگین استفاده می‌کنیم. مانند: متر در ثانیه، کیلومتر بر ساعت و.....

$$\boxed{\bar{X}_H = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}}$$

نکته :ممکن است داده‌ها دارای وزن باشند، آنگاه میانگین هارمونیک به فرم زیر محاسبه می‌شود:

$$\boxed{\bar{X}_H = \frac{\sum_{i=1}^{n} w_i}{\frac{w_1}{x_1}+\cdots+\frac{w_k}{x_k}} = \frac{n}{\sum_{i=1}^{k}\frac{w_i}{x_i}}}$$

مثال ۱: اتومبیلی مسیری را با سرعت 100 کیلومتر رفته، و $\frac{1}{3}$ مسیر را با سرعت 80 کیلومتر و باقیمانده را با سرعت 120 کیلومتر برگشته، سرعت متوسط این اتومبیل چقدر بوده است؟

$$\bar{X}_H = \frac{\sum_{i=1}^{n} w_i}{\frac{w_1}{x_1}+\ldots+\frac{w_n}{x_n}} = \frac{1+\frac{1}{3}+\frac{2}{3}}{\frac{1}{100}+\frac{1/3}{80}+\frac{2/3}{120}} = 101.4$$

مثال ۳:در یک کارگاه 5 ماشین با سرعت 4 دور در ثانیه و 3 ماشین با سرعت 6 دور در ثانیه کار می‌کنند. سرعت متوسط این ماشین‌ها چند دور در ثانیه است؟

با توجه به واحد ترکیبی (دور در ثانیه) از میانگین هارمونیک استفاده می‌کنیم:

$$\bar{x}_H = \frac{5+3}{\frac{5}{4}+\frac{3}{6}} = \frac{8}{\frac{21}{12}} = \frac{96}{21} = 4.57$$

۴ـ میانگین وزنی (μ_w):

در صورتی که داده‌ها دارای وزن‌های متفاوتی باشند، از میانگین وزنی استفاده می‌کنیم:

$$\bar{X}_w = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

مثال : در صورتی که نمرات یک دانشجو در یک ترم به فرم زیر باشدمطلوبست محاسبه معدل یا میانگین معدل او در این ترم.

x_i	w_i	
10	3 واحد	آمار
15	2 واحد	ریاضی
10	3 واحد	فیزیک

$$\bar{X}_\omega = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} = \frac{3\times10+2\times15+3\times10}{3+2+3} = \frac{90}{8} = 11.25$$

۵ـ میانگین پیراسته

این میانگین زمانی مورد استفاده قرار می‌گیرد که در بین داده‌ها، داده‌هائی وجود داشته باشند که به علت بزرگی یا کوچکی زیاد با سایر داده‌ها همخوانی نداشته باشند.

محاسبه میانگین این روش به صورت دستی انجام شده و زمان‌بر است. ترتیب آن به فرم زیر است:

۱) داده‌ها به صورت صعودی مرتب شوند.

۲) تمام مشاهدات پائین‌تر و بالاتر از درصد مشخصی حذف شوند.

۳) میانگین بقیه داده‌ها را به صورت حسابی محاسبه می‌کنیم.

❍ مُد یا نما (MO)

متغیری که دارای بیشترین فراوانی است مد (نما) نامیده می‌شود. گفتنی است نما کم اهمیت‌ترین پارامتر مرکزی نیز به حساب می‌آید.

۱ـ محاسبه مد در داده‌های طبقه‌بندی نشده:

مد در این حالت داده‌ای است که فراوانی مطلق آن از سایر داده‌ها بیشتر است.

مثال :

5 و 4 و 3 و 1 و 1   MO = 1

2 و 2 و 3 و 1 و 1   MO = 1و 2

2 و 2 و 1 و 1   MO = نداریم   (چون تمام داده‌ها به یک اندازه تکرار شده‌اند)

۲ـ محاسبه مد در داده‌های طبقه‌بندی شده:

الف) ابتدا در ستون فراوانی مطلق طبقه‌ای را پیدا می‌کنیم که بیشترین فراوانی مطلق را دارد.

ب) سپس مقدار مد را از رابطه زیر بدست می‌آوریم:

$$\boxed{MO = \text{حد پائین طبقه مددار} + \frac{d_1}{d_1+d_2} \times \text{طول طبقه}}$$

$d_1 =$ فراوانی مطلق طبقه مددار ـ فراوانی مطلق طبقه ماقبل طبقه مددار

$d_2 =$ فراوانی مطلق طبقه مددار ـ فراوانی مطلق طبقه بعداز طبقه مددار

مثال ۲: در جدول زیر مد کدام مقدار است؟

C-L	3-5	6-8	9-11
F_i	4	20	12

حل : ابتدا جدول را پیوسته می‌کنیم:

C-L	2.5-5.5	5.5-8.5	8.5-11.5	جمع
F_i	4	20	12	N = 36

الف) طبقه مددار = $(5.5\text{-}8.5) \rightarrow$ طبقه‌ای که دارای بیشترین فراوانی است.

ب) مقدار مد = $5.5+\dfrac{(20-4)}{(20-4)+(20-12)}\times 3=5.5+\dfrac{16}{16+8}\times 3=7.5$

نکته :مد نیز همانند دیگر پارامترهای مرکزی خاصیت خطی بودن دارد یعنی:

$$y=\pm ax\pm b\rightarrow Mo_y=\pm aMo_x\pm b$$

❍ تفاوت‌های اساسی بین میانگین، میانه و مد

۱) میانگین برحسب مقیاس داده‌ها است و در محاسبه آن فراوانی و کمیت داده، در نظر گرفته می‌شود، اما میانه و مد تابع ترتیب و فراوانی داده‌ها هستند.

۲) میانگین از ترکیب داده‌ها حاصل نشده و هر افزایش یا کاهش داده‌ها مقدار میانگین را عوض می‌کند، اما اگر افزایش یا کاهش ترتیب داده‌ها را عوض نکند در مقدار مد و میانه تأثیری ندارد.

❍ چندک‌ها (چارک‌ها ـ دهک‌ها ـ صدک‌ها)

چندک‌ها مقادیری از مشاهدات هستند که دامنهٔ تغییرات را در فواصل چندکی تقسیم می‌کنند. به طوری که فراوانی‌ها در هر یک از این فواصل، درصد معینی از فراوانی کل را تشکیل می‌دهد.

چارک‌ها: دامنهٔ تغییرات را به چهار قسمت مساوی تقسیم می‌کنند.

دهک‌ها: دامنهٔ تغییرات را به ده قسمت مساوی تقسیم می‌کنند.

صدک‌ها: دامنهٔ تغییرات را به صد قسمت مساوی تقسیم می‌کنند.

۱ـ محاسبهٔ چندک‌ها برای داده‌های طبقه‌بندی نشده :

الف) ابتدا داده‌ها را به ترتیب صعودی مرتب می‌کنیم.

ب) سپس با توجه به نوع چندک محل آن را با استفاده از:

Q= چارک: $\dfrac{aN}{4}+\dfrac{1}{2}\ ;\ a=1,2,3$

D= دهک: $\dfrac{aN}{10}+\dfrac{1}{2}\ ;\ a=1,2,\ldots,9$

P= صدک: $\dfrac{aN}{100}+\dfrac{1}{2}\ ;\ a=1,2,\ldots,99$

پیدا می‌کنیم.

مثال : دهک چهارم داده‌های 9، 6، 2، 0، 4، 1- چقدر است؟

مرتب‌سازی : $-1\,,0\,,2\,,4\,,6\,,9$

محل دهک چهارم = $\dfrac{4\times 6}{10}+\dfrac{1}{2}=\dfrac{24}{10}+\dfrac{1}{2}=\dfrac{29}{10}=2.9$

داده دوم

مقدار دهک چهارم = $\underset{\text{داده دوم}}{0}+0.9(\underset{\text{داده سوم}}{2}-0)=1.8$

۲ـ محاسبه چندک‌ها برای داده‌های طبقه‌بندی شده:

الف) ابتدا از روی جدول، فراوانی تجمعی را محاسبه می‌کنیم.

ب) با استفاده از $\frac{aN}{4}\ (a=1,2,3)$ یا $\frac{aN}{10}\ (1,2,\ldots,9)$ یا $\frac{aN}{100}\ (1,2,\ldots,99)$ اولین طبقه‌ای که فراوانی تجمعی‌اش بیشتر یا مساوی یکی از مقادیر فوق باشد را با توجه به چارک، دهک یا صدک پیدا می‌کنیم.

ج) با استفاده از فرمول زیر آن را محاسبه می‌نماییم:

$$\text{چارک} = \text{مقدار چندک} = \text{حد پائین طبقه چارک‌دار} + \frac{\left(\frac{aN}{4} - \text{فراوانی تجمعی طبقه ماقبل}\right)}{\text{فراوانی مطلق طبقه چارک‌دار}} \times \text{طول طبقه}$$

$$\text{دهک} = \text{مقدار چندک} = \text{حد پائین طبقه دهک‌دار} + \frac{\left(\frac{aN}{10} - \text{فراوانی تجمعی طبقه ماقبل}\right)}{\text{فراوانی مطلق طبقه دهک‌دار}} \times \text{طول طبقه}$$

$$\text{صدک} = \text{مقدار چندک} = \text{حد پائین طبقه صدک‌دار} + \frac{\left(\frac{aN}{100} - \text{فراوانی تجمعی طبقه ماقبل}\right)}{\text{فراوانی مطلق طبقه صدک‌دار}} \times \text{طول طبقه}$$

نکته :دهک پنجم = چارک دوم = صدک پنجاهم = میانه است.

مثال ۱: مطلوبست دهک دوم جدول زیر

C – L	40-50	50-60	60-70
F_i	5	18	7

ابتدا باید فراوانی تجمعی جدول فوق را محاسبه نمائیم:

حد پائین طبقه دهک‌دار

C – L	40 – 50	50 – 60	60 – 70	
F_i	5	18	7	$N=\Sigma F_i=30$
F_{c_i}	5	23	30	

فراوانی مطلق طبقه دهک‌دار

فراوانی تجمعی طبقه ماقبل

$\frac{aN}{10}=\frac{2\times 30}{10}=6 \quad \rightarrow$: محل دهک دوم ، طبقه $50-60$ محل دهک دوم می‌باشد.

$50+\frac{(6-5)}{18}\times 10 = 50+\frac{10}{18}=50.55$: مقدار دهک دوم

○ پارامترهای پراکندگی

به‌طور کلی پارامترهای پراکندگی معیارهایی برای تعیین میزان پراکندگی داده از یکدیگر یا میزان پراکندگی آن‌ها نسبت به میانگین است، که از پارامترهای پراکندگی می‌توان به دامنه تغییرات ـ دامنه میان چارکی ـ نیمه میان‌چارکی (انحراف چارکی) ـ انحراف متوسط از میانگین ـ واریانس ـ انحراف معیار ـ نیمه واریانس،اشاره نمود.

الف) دامنه تغییرات

ساده‌ترین پارامتر پراکندگی، دامنه تغییرات است. این شاخص با تفاضل کوچک‌ترین مشاهده از بزرگ‌ترین مشاهده محاسبه می‌شود، یعنی:

$$\boxed{R = \text{Max } x_i - \text{Min } x_i}$$

مثال : دامنهٔ تغییرات برای داده‌های 9 و 0 و 1 و 7 و 4 و 1- کدام است؟

مرتب‌سازی: 1- و 0 و 1 و 4 و 7 و 9 $\Rightarrow R = 9 - (-1) = 10$

● نکات مهم دامنه تغییرات

۱ـ دامنه تغییرات کم اهمیت‌ترین پارامتر پراکندگی است، زیرا تنها تابع تغییرات دو اندازه است و وضعیت اعداد وسط را مشخص نمی‌کند.

۲ـ دامنهٔ تغییرات علی‌رغم سادگی در محاسبه، از ثبات چندانی برخوردار نیست چون:

ـ در تعیین این شاخص، از مجموعه مشاهدات فقط به کوچکترین و بزرگترین مشاهده توجه می‌شود.

ـ با این شاخص چگونگی توزیع سایر مشاهدات نشان داده نمی‌شود، یعنی اگر داده‌ها به مقدار بزرگ‌تر یا مقدار کوچک‌تر نزدیک باشند، دامنه تغییرات این ویژگی مشاهدات را نشان نمی‌دهد.

ب) دامنه میان چارکی (IQR) :

دامنه تغییرات 50% از مشاهدات است. در این تعریف، میانه را که 50% داده‌هاست ملاک قرار داده و از پائین تا 25% و از بالا تا 75% گسترش می‌دهیم.به عبارت بهتر IQR چارک اول تا سوم را شامل می‌شود و برابر است با:

$$\boxed{IQR = Q_3 - Q_1}$$

○ نیمه میان چارکی (انحراف چارکی، SIQR)

انحراف چارکی مساوی است با نصف دامنه میان چارکی، یعنی:

$$\boxed{SIQR = \frac{Q_3 - Q_1}{2} = \frac{IQR}{2}}$$

- **نکات مهم میان چارکی**

۱ـ اصولاً انحراف چارکی از دامنه تغییرات و دامنه میان‌چارکی با ثبات‌تر است.

۲ـ در توزیع‌هائی که دارای تعداد اندکی مقدار در ابتدا و انتها هستند از انحراف چارکی به عنوان شاخص پراکندگی استفاده می‌شود.

۳ـ در توزیع‌های نامتقارن اغلب از میانه به عنوان شاخص مرکزی و از انحراف چارکی (SIQR) به عنوان شاخص پراکندگی استفاده می‌شود.

۴ـ اگر انحراف چارکی اندازه‌ها برابر صفر باشد، 50% اندازه‌هایی که در وسط قرار گرفته‌اند، با هم برابرند. در نتیجه چارک‌های اول و دوم و سوم با هم برابرند و برعکس، یعنی:

$$\boxed{Q = 0 \leftrightarrow Q_1 = Q_2 = Q_3}$$

ج) انحراف متوسط از میانگین (انحراف میانگین، انحراف متوسط، $A.D_\mu$)

هیچ‌کدام از شاخص‌هائی که تا به حال در مورد آن‌ها صحبت شده قادر به بیان تمامی تغییرات نیستند. تغییر زمانی مفهوم پیدا می‌کند که هر یک از داده‌ها نسبت به مبدأ مقایسه شوند و بهترین مبدأ یا مرکز داده‌ها میانگین است. اما از آن‌جا که مجموع انحرافات از میانگین همیشه صفر است، برای محاسبه انحراف متوسط از میانگین از رابطه زیر استفاده می‌کنیم:

داده‌های طبقه‌بندی نشده: $A.D_\mu = \dfrac{\sum_{i=1}^{n} |x_i - \mu_x|}{N}$: (μ_x : میانگین و N : تعداد مشاهدات)

داده‌های طبقه بندی شده: $A.D_\mu = \dfrac{\sum F_i |x_i - \mu_x|}{N}$ یا $A.D_\mu = \sum f_i |x_i - \mu_x|$

- **نکات مهم انحراف متوسط از میانگین**

۱ـ این فرمول از علامت جبری داده‌ها صرف‌نظر می‌کند.

۲ـ تأثیر انحرافات بزرگ را در شرایطی که تعداد زیادی انحرافات کوچک در برابر تعداد کمی انحرافات بزرگ وجود داشته باشد، نشان نمی‌دهد. که نقص اساسی این پارامتر می‌باشد.

۳ـ در صورتی که تمام داده‌ها با هم برابر باشند انحراف متوسط از میانگین، صفر است: $\boxed{x_1 = x_2 = \dots = x_n \Leftrightarrow A.D_\mu = 0}$

۴ـ اگر به متغیرها عدد ثابتی اضافه یا کم شود، انحراف متوسط تغییری نمی‌کند: $\boxed{A.D(x \pm a) = A.D_\mu}$

۵ـ اگر متغیرها را در عدد ثابتی ضرب یا تقسیم کنیم، انحراف متوسط در قدر مطلق آن عدد ضرب یا بر قدرمطلق آن عدد تقسیم می‌شود.

$$\boxed{\begin{array}{l} A.D_{(\pm ax)} = |\pm a| . A.D_x \\ A.D_{\left(\frac{x}{\pm a}\right)} = \dfrac{A.D_x}{|\pm a|} \end{array}}$$

۶ ـ انحراف متوسط اعداد ثابت صفر است.

د) واریانس (پراش، σ_x^2، var (x) و D(x)) و انحراف معیار (σ_x)

یکی از شاخص‌های پراکندگی داده نسبت به میانگین، واریانس است. هنگام محاسبه انحراف متوسط از میانگین، برای جلوگیری از خنثی شدن انحرافات منفی از قدرمطلق استفاده می‌شود، اما در واریانس از مجذور انحرافات استفاده می‌کنیم که شاخص بهتری نسبت به انحراف متوسط از میانگین است.

واریانس و انحراف معیار می‌توانند تأثیرات انحرافات بزرگ را به راحتی نشان دهند، بنابراین بهترین شاخص برای نشـان دادن انحرافـات بزرگ هستند.

❍ خواص واریانس

۱) واریانس اعداد مساوی صفر است.

۲) فرمول محاسبه:

محاسبهٔ واریانس برای داده‌های طبقه‌بندی نشده: $V(x)=\sigma_x^2=\frac{\sum(x_i-\mu_x)^2}{N}=\frac{\sum x_i^2}{N}-\left(\frac{\sum x_i}{N}\right)^2=\frac{\sum x_i^2}{N}-\mu_x^2$

محاسبه واریانس برای داده‌های طبقه‌بندی شده: $V(x)=\sigma_x^2=\frac{\sum F_i(x_i-\mu_x)^2}{N}=\frac{\sum F_i x_i^2}{N}-\left(\frac{\sum F_i x_i}{N}\right)^2=\frac{\sum F_i x_i^2}{N}-\mu_x^2$

$$V(\mathbf{x})=\sigma_x^2=\sum \mathbf{f}_i\ (\mathbf{x}_i-\mu_x)^2=\sum f_i\ \mathbf{x}_i^2-\left(\sum \mathbf{f}_i\ \mathbf{x}_i\right)^2=\sum f_i x_i^2-\mu_x^2$$

● نکات مهم واریانس

۱ـ اگر همه متغیرها با هم برابر باشند یا اعداد ثابت داشته باشیم، واریانس و انحراف معیار صفر است.

۲ـ
$$\left.\begin{array}{l}\sigma^2(x\pm a)=V(x\pm a)=V(x)=\sigma_x^2\\ \sigma^2(\pm bx\pm a)=(\pm b)^2\sigma_x^2\\ \sigma^2\left(\frac{x}{\pm a}\right)=\frac{\sigma_x^2}{(\pm a)^2}\end{array}\right\}\Rightarrow \text{انحراف معیار}\left\{\begin{array}{l}\sigma(x\pm a)=\sigma_x\\ \sigma(\pm bx\pm a)=|\pm b|\sigma_x\\ \sigma=\left(\frac{x}{\pm a}\right)=\frac{\sigma_x}{|\pm a|}\end{array}\right.$$

۳ـ **محاسبه واریانس یک نمونه n تائی:** اگر واریانس نمونه مورد نظر باشد، در مخرج کسر تمامی فرمول‌های واریانس، بـه جـای n از

(n-1) استفاده می‌شود. یعنی:

$$\boxed{V(x) = S^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1} = \frac{1}{n-1}\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right]}$$

۴ـ واحد اندازه‌گیری واریانس، مجذور واحد اصلی متغیر است، بنابراین اگر x برحسب ریال اندازه‌گیری شده باشد، واحد واریانس جامعه، ریال به توان 2 خواهد بود.

۵ـ در مقایسه دو یا چند جامعهٔ آماری، آن‌که انحراف معیارش کم‌تر است، مقادیر صفات متغیر مورد مطالعه آن جامعه یکنواخت‌تر از جامعه‌های دیگر می‌باشد.

مثال ۱: مطلوبست واریانس و انحراف معیار داده‌های زیر:

12451 , 12452 , 12453 , 12454 , 12455

حل : برای محاسبه واریانس چند داده می‌توانیم قسمت مشترک را حذف کنیم بنابراین واریانس داده‌های زیر را محاسبه می‌کنیم.

1 , 2 , 3 , 4 , 5

$$\mu = \frac{\sum x_i}{N} = \frac{1+2+3+4+5}{5} = 3$$

$$\sigma^2 = \frac{\sum\left(x_i - \mu\right)^2}{N} = \frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5} = \frac{4+1+0+1+4}{5} = 2$$

$\sigma^2 = 2$ در نتیجه انحراف معیار $\sigma = \sqrt{2}$

❍ پارامترهای پراکندگی نسبی:

❍ ضریب پراکندگی (ضریب تغییرات) (CV) :

عبارتست از نسبت انحراف معیار به میانگین:

$$\boxed{CV = \frac{\sigma_x}{\mu_x}}$$

توجه کنید که در شرایط زیر از معیار ضریب پراکندگی استفاده می‌کنیم که میانگین و واریانس فاقد آن هستند:

۱) دو یا چند جامعه در مقایسه با هم دارای مشاهدات ناهمگون از نظر واحد اندازه‌گیری می‌باشند. مانند: یک جامعه برحسب متر و یک جامعه بر حسب اینچ

گاهی نیز مقیاس صفت مورد اندازه‌گیری در دو جامعه یکسان است، ولی بزرگی مشاهدات آن‌ها به طور قابل ملاحظه‌ای تفاوت دارد. مانند: مقایسه پراکندگی سود و زیان در صنایع دستی با صنایع سنگین.

۲) زمانی که دو یا چند جامعه دارای میانگین‌های متفاوتی باشند.

- **نکات مهم ضریب تغییرات**

۱ـ ضریب تغییرات اعداد ثابت، برابر صفر است: (a عدد ثابت دلخواه است) : $\boxed{CV(a)=0}$

۲ـ اگر همهٔ متغیرها با هم برابر باشند، ضریب تغییرات برابر صفر است و برعکس.

$$\boxed{x_1=x_2=...=x_N \Leftrightarrow \sigma_x=0 \Leftrightarrow CV_x=0}$$

۳ـ اگر به متغیرها عدد ثابتی مانند a را اضافه یا کم کنیم، ضریب تغییرات عبارتست از:

$$\boxed{CV(x\pm a)=\frac{\sigma(x\pm a)}{\mu(x\pm a)}=\frac{\sigma_x}{\mu_x\pm a}}$$

۴ـ اگر متغیرها را در عدد ثابتی ضرب کنیم، ضریب تغییرات عبارتست از:

$$CV(ax)=\frac{\sigma_{(ax)}}{\mu_{(ax)}}=\frac{|a|\sigma_x}{a\mu_x}=\begin{cases} CV(x) & \text{اگر } a \text{ مثبت باشد:} \\ -CV(x) & \text{اگر } a \text{ منفی باشد:} \end{cases}$$

۵ـ اگر متغیرها را بر عدد ثابتی مانند a تقسیم کنیم، ضریب تغییرات عبارتست از:

$$Cv\left(\frac{x}{a}\right)=\frac{\sigma\left(\frac{x}{a}\right)}{\mu\left(\frac{x}{a}\right)}=\frac{\frac{\sigma_x}{|a|}}{\frac{\mu_x}{a}}=\begin{cases} CV(x) & \text{اگر } a \text{ مثبت باشد:} \\ -CV(x) & \text{اگر } a \text{ منفی باشد:} \end{cases}$$

۶ـ ضریب تغییرات واحد ندارد.

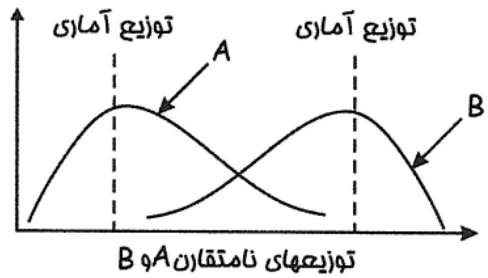
۷ـ در مسائل اقتصادی از بین دو جامعه آن جامعه‌ای که cv اش کمتر است بهتر است.

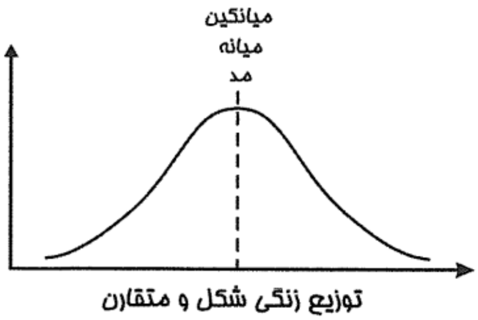
۸ـ فرمول‌های دیگر ضریب پراکندگی عبارتند از:

- **درصد ضریب پراکندگی** $\boxed{CV\times 100=\frac{\sigma_x}{\mu_x}\times 100}$

❍ چولگی (عدم قرینگی، ضریب انحراف توزیع از حالت قرینگی)

در مقایسه دو یا چند جامعه با یکدیگر ابتدا از پارامترهای مرکزی استفاده می‌شود. اما در صورت تساوی برخی از پارامترهای مرکزی (مانند: میانگین) اختلاف جوامع آماری به کمک شاخص‌های پراکندگی (مانند: انحراف معیار) مشخص می‌گردد. گاهی پارامترهای پراکندگی نیز به علت مساوی بودن جوابگو نیستند. بنابراین برای رفع این مشکل از معیاری به نام چولگی (skewness) استفاده می‌کنیم. برای مثال در شکل زیر دو توزیع دارای میانگین و واریانس مساوی هستند، اما توزیع یکسانی ندارند. توزیع جامعه A دارای تراکم در حول وحوش مبدأ مختصات است، در حالی که مد جامعه B در نقطه مقابل آن قرار دارد. این تفاوت را چولگی (skewness) یا انحراف از قرینگی می‌گویند. چولگی توزیع‌ها در مقایسه با توزیع متقارن معین می‌شود. توزیع متقارن توزیعی است که پارامترهای مرکزی آن (مد، میانگین و میانه) با همدیگر مساوی باشند. هر چه یک توزیع با توزیع متقارن تفاوت بیشتری داشته باشد، انحراف از قرینگی آن بیشتر خواهد بود.





❍ ضریب چولگی (Sk)

شاخص اندازه‌گیری پارامترهای تعیین انحراف از قرینگی، ضریب چولگی است که با Sk نمایش داده می‌شود. توجه کنید که این معیار بدون واحد است. برای محاسبه ضریب چولگی از فرمول‌های زیر استفاده می‌شود:

$$SK = \frac{\mu_3}{\sigma^3_x} = \frac{\frac{\sum_{i=1}^{N}(x_i-\mu)^3}{N}}{\left(\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}\right)^{\frac{3}{2}}} = \frac{\frac{\sum_{i=1}^{N}F_i(x_i-\mu)^3}{N}}{\left(\frac{\sum_{i=1}^{N}F_i(x_i-\mu)^2}{N}\right)^{\frac{3}{2}}} = \frac{\sum_{i=1}^{N}f_i(x_i-\mu)^3}{\left(\sum_{i=1}^{N}f_i(x_i-\mu)^2\right)^{\frac{3}{2}}}$$

ضریب چولگی به طریقه گشتاورها : فرمول (1)

دو فرمول مهم، به نام فرمول‌های پیرسون به صورت زیر وجود دارد:

فرمول (2) : ضریب چولگی اول پیرسون : $SK_1 = \frac{(\bar{x}-MO)}{\sigma_x}$

فرمول (3) : ضریب چولگی دوم پیرسون : $SK_2 = \frac{3(\bar{x}-Md)}{\sigma_x}$

نکته :زمانی که چولگی توزیع متعادل یا متناسب یا خفیف باشد، فرمول‌های پیرسون با یکدیگر برابر شده و رابطه زیر را به وجود می‌آورند:

$$\boxed{\bar{x}-MO = 3(\bar{x}-Md)}$$

● انواع چولگی

۱ـ چوله به راست (تمایل توزیع به سمت چپ)

۲ـ چوله به چپ (تمایل توزیع به سمت راست)

چوله به راست: در توزیع‌هایی که ضریب چولگی مثبت دارند ($Sk > 0$)، چوله به راست می‌باشد یعنی $Mo < Md < \mu_X$

چوله به چپ: در توزیع‌هایی که ضریب چولگی منفی دارند ($Sk < 0$)، چوله به چپ می‌باشد یعنی $Mo > Md > \mu_X$

به شکل‌های زیر توجه کنید:

● تفسیر ضریب چولگی

قدرمطلق ضریب چولگی نشاندهنده میزان اختلاف جامعه آماری با توزیع نرمال از نظر قرینگی است. بدیهی است هر چه |SK| بزرگتر باشد، تفاوت جامعه از نظر قرینگی با توزیع نرمال بیشتر خواهد بود. طوری که:

۱ـ تقریباً چولگی وجود ندارد و جامعه از نظر قرینگی تقریباً نرمال است. $|SK| \le 0.1$

۲ـ چولگی موجود، اندک ولی غیر قابل اغماض است در حقیقت جامعه از نظر قرینگی دارای تفاوت اندکی با توزیع نرمال است. $0.1 < |SK| \le 0.5$

۳ـ چولگی زیاد و غیر قابل اغماض است. به عبارت دیگر جامعه از نظر قرینگی دارای تفاوت فاحشی با توزیع نرمال است. $|SK| > 0.5$

❍ کشیدگی و ضریب کشیدگی

کشیدگی معیاری است بدون واحد که ارتفاع منحنی را مورد بحث قرار می‌دهد و رابطه معکوس با پراکندگی دارد.

چنانچه داده‌ها دارای طبقه‌بندی مشخص و کمی باشند، بهترین روش برای محاسبه کشیدگی استفاده از گشتاورهاست.

برای محاسبه کشیدگی توزیع‌ها از رابطه زیر استفاده می‌شود:

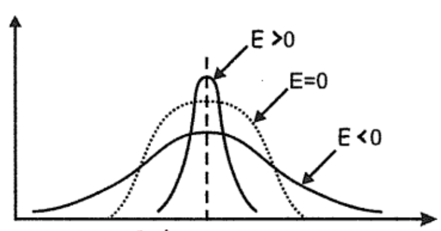
کشیدگی (گشتاوری) ؛ $\frac{\mu_4}{\sigma^4} = \dfrac{\frac{\sum (x_i - \mu)^4}{N}}{(\text{واریانس})^2} = \dfrac{\frac{\sum F_i (x_i - \mu)^4}{N}}{(\text{واریانس})^2}$

در این رابطه μ_4 گشتاور مرتبه چهارم نسبت به میانگین و $\frac{\mu_4}{\sigma^4}$ نشاندهنده کشیدگی هر توزیع دلخواه است.

معیار ضریب کشیدگی اختلاف کشیدگی توزیع‌ها را نسبت به کشیدگی توزیع نرمال بدست می‌آورد.

نکته :کشیدگی گشتاوری توزیع نرمال همیشه 3 است و کشیدگی چندکی توزیع نرمال 0.263 می‌باشد.

ضریب کشیدگی ؛ $E = \frac{\mu_4}{\sigma^4} - 3$



مقایسه انواع کشیدگی

● تفسیر ضریب کشیدگی:

چنانچه $E = 0$ باشد، کشیدگی توزیع هم اندازه و هم ارتفاع توزیع نرمال است.
چنانچه $E>0$ باشد، کشیدگی توزیع از نرمال بلندتر و پراکندگی آن از نرمال کمتر است.
چنانچه $E<0$ باشد، کشیدگی توزیع از نرمال کوتاه‌تر و پراکندگی آن از نرمال بیشتر است.

قدرمطلق ضریب کشیدگی میزان اختلاف ارتفاع را با توزیع نرمال بیان می‌کند:

ـ اگر $\lvert E \rvert \leq 0.1$ باشد، توزیع جامعه از نظر پراکندگی تقریباً نرمال است.
ـ اگر $0.1 < \lvert E \rvert \leq 0.5$ باشد، توزیع از نظر کشیدگی دارای تفاوت اندکی با توزیع نرمال است. اما غیر قابل اغماض است.
ـ اگر $\lvert E \rvert > 0.5$ باشد، توزیع از نظر کشیدگی اختلاف فاحش با توزیع نرمال دارد و تفاوت غیرقابل اغماض است.

مثال ۱: اگر $N = 1000$, $\sum F_i \left(x_i - \mu_x\right)^4 = 5000$ و انحراف معیار جامعه 2 باشد، مقدار ضریب کشیدگی کدام است؟

$$\begin{cases} E = \dfrac{\mu_4}{\sigma^4} - 3 = \dfrac{\dfrac{\sum F_i \left(x_i - \mu\right)^4}{N}}{\left(\sigma^2\right)^2} - 3 = \dfrac{\dfrac{5000}{1000}}{(4)^2} - 3 = -2.69 \\ \sigma_x = 2 \quad \rightarrow \quad \sigma^2 = 4 \end{cases}$$

❍ نمایش هندسی مشاهدات:

ـ برای نمایش توزیع‌های فراوانی، اغلب از نمودارها استفاده می‌شود.

ـ برای نمایش داده‌ها با توجه به نوع مقیاس داده‌ها، روش‌های مختلفی وجود دارد:

۱ـ نمودارهای کمی برای مقیاس‌های نسبی و فاصله‌ای استفاده می‌شود.

۲ـ نمودارهای کیفی برای مقیاس‌های اسمی و رتبه‌ای استفاده می‌شود.

نکته: مهمترین نمودارهای کمی و کیفی به شرح زیر هستند.

• نمودارهای کمی:

۱) نمودار بافت‌نگار (Histogram chart)

۲) نمودار چندضلعی

۳) نمودار فراوانی تجمعی (اجایو) (ogive) (cumulative frequency chart)

۴) نمودارهای تحلیل اکتشافی داده‌ها
- a) شاخه و برگ
- b) جعبه‌ای (box)

• نمودارهای کیفی :(وصفی)

۱) نمودار ستونی (میله‌ای) (bar chart)

۲) نمودار دایره‌ای (pie Chart)

۳) نمودار پارتو (pareto chart)

نمودارهای داده‌های کمی

نمودار میله‌ای

این نمودار عموماً برای داده‌های گسسته استفاده می‌شود. در این نمودار محور افقی نشان دهنده سطوح متغیر x_i است و محور عمودی نشان دهنده r_i یا f_i ها می‌باشد.

نمودار بافت نگار یا هیستوگرام

این نمودار عموماً برای داده‌های پیوسته استفاده می‌شود و مناسب برای توصیف هندسی مشاهدات است. در این نمودار محور افقی معرف سطوح متغیر x_i است که به صورت حدود بازه‌ها در جدول فراوانی تعریف شده‌اند و محور عمودی بر اساس r_i یا f_i ها تعریف می‌شود، برای تعیین ارتفاع مستطیل‌ها فراوانی یا فراوانی نسبی هر طبقه را بر طول طبقه تقسیم می‌کنیم. به عبارت دیگر

$$\text{ارتفاع هر طبقه} = \frac{\text{فراوانی ساده (نسبی) طبقه مورد نظر}}{\text{طول طبقه مورد نظر}}$$

نمودار چندبر یا چند ضلعی (فراوانی یا فراوانی نسبی)

این نمودار عموماً برای داده‌های پیوسته استفاده می‌شود. در این نمودار محور افقی معرف نماینده طبقات است و محور عمودی براساس r_i یا f_i ها مانند آنچه در بافت نگار بیان‌شد به‌دست می‌آید (به عبارت دیگر برای رسم این نمودار بهتر است ابتدا بافت نگار رسم شود سپس نقاط وسط عرض مستطیل‌ها را به هم وصل کنیم). رسم همزمان این نمودار برای چند جامعه در یک دستگاه امکان مقایسه جوامع را فراهم می‌کند.

نمودار فراوانی تجمعی

این نمودار عموماً برای داده‌های پیوسته استفاده می‌شود. در این نمودار محور افقی نشان دهنده حدود طبقات است و محور عمودی نشان دهنده F_i یا R_i است. در این نمودار نقطه صفر کران پایین طبقه اول منهای طول طبقه اول است و به کران پایین هر طبقه فراوانی تجمعی یا فراوانی نسبی تجمعی آن طبقه را نسبت می‌دهیم.

نمودارهای داده‌های کیفی

نمودار کلوچه‌ای یا دایره‌ای

در این نمودار مساحت هر قطاع به شکل زیر محاسبه می‌شود:

$$\text{مساحت قطاع} = ۳۶۰ \times r_i$$

در این نمودار معمولاً قطاع‌ها به ترتیب نزولی مرتب می‌شوند.

نمودار ستونی

در این نمودار محور عمودی نشان دهنده r_i یا f_i است و محور افقی بیانگر سطوح متغیر می‌باشد، که عرض ستون‌ها به صورت دلخواه تعیین می‌شود.

نکته ۲۰.۱: این نمودار امکان مقایسه دو یا چند جامعه را در سطوح مختلف در یک زمان امکان پذیر می‌سازد.

نمودار پارتو

این نمودار که برای داده‌های کیفی مناسب است اهمیت خاصی در کنترل کیفیت دارد. نمودار دارای دو محور عمودی و یک محور افقی می‌باشد. محور افقی شامل کیفیت مشاهدات است که به ترتیب کاهش فراوانی مرتب می‌شوند. محور عمودی اول فراوانی هر مشاهده بوده و محور عمودی دوم نشان دهنده فراوانی نسبی تجمعی (درصد) است که در انتهای محور افقی در سمت راست رسم می‌شود.

تمرینات

۷) در مدیریت، انسان‌ها را به لحاظ ارتباطات به چهار دسته تصویری، احساسی، صوتی و ارقامی تقسیم می‌کنند. کارکنان یک سازمان مورد بررسی قرار گرفته‌اند و حاصل بررسی در جدول زیر خلاصه شده است.

گروه ارتباطی	تصویری	احساسی	صوتی	ارقامی
تعداد کارکنان	۱۰۰	۱۵۰	۳۰۰	۵۰

نمودار مناسب را برای مشاهدات فوق رسم کنید.

۸) قیمت مسکن در سال ۷۸ نسبت به سال ۷۷، ۱۰٪ رشد داشته است و در سال ۷۹ نسبت به سال ۷۸، ۲۰٪ و در سال ۸۰ نسبت به سال ۷۹، ۱۵٪ و در سال ۸۱ نسبت به سال ۸۰، ۱۲٪ رشد داشته است. متوسط نسبت رشد قیمت مسکن در این سالها را به‌دست آورید.

۱۲) مشاهدات زیر مربوط به تعداد دفعات مراجعه مادران به یک مرکز بهداشتی مادر و کودک است.

۳ ۴ ۵ ۳ ۱ ۵ ۲ ۱ ۰ ۱ ۳ ۴ ۵ ۴

۹ ۸ ۶ ۷ ۷ ۱ ۸ ۲ ۹ ۸ ۴ ۳ ۷ ۶

۹ ۳ ۵ ۹ ۷ ۴ ۶ ۲ ۳ ۳ ۷ ۸

الف) داده‌های فوق را در یک جدول فراوانی خلاصه کنید.

ب) نمودار مناسب را رسم کنید.

ج) معیارهای مرکزی مشاهدات فوق را محاسبه کنید.

د) واریانس و انحراف معیار مشاهدات فوق را محاسبه کنید.

هـ) توزیع مشاهدات را از نظر شکل توزیع تفسیر کنید.

۱۸) جدول زیر توزیع سنی موارد بیماری معینی را که در طول سال در استان خاصی گزارش شده نشان می‌دهد.

حدود طبقات	f_i	F_i
۵ − ۱۵	۶	
۱۵ − ۲۵		۱۶
۲۵ − ۳۵		۴۵
۳۵ − ۴۵	۲۲	
۴۵ − ۵۵		
جمع	۸۰	

الف) جدول را کامل کنید.
ب) نمودارمناسب را رسم کنید.
ج) معیارهای مرکزی وپراکندگی مشاهدات را محاسبه کنید.
د) شکل توزیع را با استفاده از معیارهای مناسب تفسیر کنید.

۱۹) جدول زیر توزیع نمرات درس امار ۱۰۰ دانشجو را نمایش می‌دهد.

حدود طبقات	f_i	F_i	r_i
۷ − ۱۰	۵		
۱۰ − ۱۲		۲۳	
۱۲ − ۱۵			۰/۶۵
۱۵ − ۱۷			
۱۷ − ۲۰	۳		

الف) جدول را کامل کنید.
ب) بهترین معیار مرکزی را برای این مشاهدات به‌دست آورید.
ج) بهترین معیار پراکندگی را برای این مشاهدات به‌دست آورید.
د) شکل توزیع را با استفاده از معیارهای مناسب تفسیر کنید.