



پیام نوریها
public channel



✓ کانال پیام نوریها در سال 95 با هدف تهیه جزوات و نمونه سوالات افتتاح و از همان ابتدای تاسیس کوشیده است با تکیه بر تلاش بی وقفه ، کارگروهی و فعالیت های بدون چشمداشت کاربران متمایز خود ، قدمی کوچک در راه پیشرفت ارائه خدمات به دانشجویان این مرز و بوم بردارد.

@Payamnoria

telegram.me/Payamnoria

رایگان است و همیشه رایگان میماند



اطلاع از اخبار و دانلود جزوات و نمونه سوالات

[برای ورود به کانال تلگرامی پیام نوریها کلیک کنید](#)

" کانال و خانواده تلگرامی پیام نوریها "

با عضویت در کانال و به آرشیو زیر دسترسی پیدا کنید

✓ تمام نمونه سوالات به روز تا آخرین دوره

✓ جزوات درسی

✓ بیش از ۱۰۰ فلش کارت دروس

✓ اخبار به روز پیام نور

✓ فیلم و فایل آموزشی اختصاصی

✓ انجام انتخاب واحد و حذف و اضافه

✓ پاسخگویی به سوالات دانشجویان

✓ معرفی گروه و انجمن های پیام نوری

✓ طنز و توییت دانشجویی

به یکی از بزرگترین کانال های پیام نوری بپیوندید

[برای ورود به کانال تلگرامی پیام نوریها کلیک کنید](#)

نام کتاب الکترونیکی

آمار توصیفی

تهیه کننده

مادسیج

بزرگترین درس زندگی اینست که گاهی احم قها هم درست می
گویند.
وینستون چرچیل



آشنایی با آمار توصیفی



پیشگفتار:

در عصر حاضر کسی نمی‌تواند منکر این واقعیت باشد که آمار نقشی لاینفک در زندگی روزمره ما بازی می‌کند. اخبار روزانه رسانه‌های گروهی با گزارشی از وضع هوا به پایان می‌رسند و در طول اخبار، به جریانهای بازار بورس و سهام اشاره می‌شود و روزنامه‌ها خبر از افزایش نرخ اجناس می‌دهند...

آمار به عنوان پایه يك روش و راه موثر در بررسی مسائل موجود، در بسیاری از زمینه‌های علمی از جمله جامعه شناسی، کشاورزی، فیزیک و... به کار گرفته می‌شود. در دانش امروزی، معمولاً سعی می‌شود که اطلاعات موجود در يك زمینه خاص، در قالب اعداد نمایش داده شود تا به هنگام تجزیه و تحلیل اطلاعات، فهم بهتری از پدیده مورد مطالعه به دست آمده و امکان مقایسه فراهم گردد. در يك جمله آمار مجموعه‌ای از روشهای جمع آوری، تهیه و تنظیم و تجزیه و تحلیل اطلاعات است که برای کسب يك یا چند نتیجه به خدمت گرفته می‌شود.

فهرست مطالب:

■ آمار توصیفی

■ جدول های آماری

■ نمودار های آماری

■ معیار های مرکزی

■ معیار های پراکندگی

■ منحنی های فراوانی

آمار توصیفی:

برای اینکه نتایج مناسب و مطلوب از اطلاعات که در آمار گیری ها جمع آوری می کنیم، به دست آید باید:

- اعداد نماینده واقعی مشاهدات بوده و غیر واقع یا غلط نباشند
- به نحو مفیدی تهیه و تنظیم شوند
- به نحو صحیح تجزیه و تحلیل گردند
- قابل نتیجه گیری صحیح باشند



به طور کلی، روش هایی را که به وسیله آنها می توان اطلاعات جمع آوری شده را تنظیم کرده و خلاصه نمود، آمار توصیفی می نامیم و در يك کلام آمار توصیفی عبارت از مجموعه روش هایی است که پردازش داده ها را فراهم می سازد. اطلاع از اصطلاحات زیر در آمار ضروری است.

آمار توصیفی:

داده
مقیاسهای اندازه گیری
نمونه
متغیر
جمعیت

مجموعه افراد یا اشیایی را که می‌خواهیم یک یا چند خصوصیت مشترک آنها را مورد بررسی قرار دهیم، جمعیت یا جمعیت آماری می‌نامیم.

مثال:

اندازه قد یا وزن دانشجویان بیست ساله یک شهر، تعداد لامپهای سالم و یا ناسالم تولید شده در یک کارخانه و در یک روز معین، مثالهایی از جمعیتهای آماری هستند.

نکته:

معمولا مطالعه ویژگی‌های مورد نظر، به هنگامی که جمعیت آماری بسیار گسترده باشد، مستلزم صرف هزینه و وقت زیادی می‌باشد و در بسیاری از مواقع، این امر اصولا امکان پذیر نیست. بنابراین در چنین موردی، برای مطالعه ویژگی مورد نظر، به قسمتی از جمعیت آماری اکتفا می‌کنیم.

آمار توصیفی:

داده
مقیاسهای اندازه گیری
متغیر

نمونه

جمعیت

قسمتی از جمعیت را که طبق قاعده و ضوابط خاصی، برای مطالعه خصوصیتی از جمعیت انتخاب می‌شود، یک نمونه از جمعیت می‌نامیم.

نمونه

این نمونه وقتی مفید و قابل قبول خواهد بود که بتواند نماینده خوبی برای کل جمعیت مورد مطالعه باشد. با توجه به اهمیت این موضوع شاخه‌ای از آمار تحت عنوان نظریه نمونه‌گیری با بررسی نمونه‌ای به این امر مهم می‌پردازد. در بسیاری از موارد، معمولاً نمونه تصادفی ساده را در نظر می‌گیرند.

نکته

برای بررسی اندازه قد دانشجویان بیست ساله یک شهر، انتخاب مثال ۱۵۰ نفر از بین این جمعیت به طور تصادفی، یا انتخاب ۱۰۰ لامپ به تصادف از لامپهای تولیدی یک کارخانه در یک روز معین، برای تعیین کیفیت لامپهای تولیدی این کارخانه مثالهایی از نمونه تصادفی هستند.

مثال:

آمار توصیفی:

خصوصیت مورد مطالعه، از فردی به فرد دیگر، یا از شی به شی دیگر در جمعیت آماری تغییر می‌کند، که آن را اصطلاحاً متغیر می‌نامیم.

معمولاً دو نوع متغیر در آمار مورد نظر هستند:

متغیرهای گروهی، نظیر رنگ، نژاد، شغل و گروه خونی که شامل چند گروه یا طبقه می‌باشند.

متغیرهای عددی که ممکن است نتیجه شمارش باشد، مانند تعداد احشام هر خانوار در یک روستا، تعداد حوادث در یک کارخانه در روزهای مختلف، و یا نتیجه اندازه‌گیری باشد، مثل قد دانشجویان بیست ساله در یک شهر، حجم شربت مولتی ویتامین با استاندارد خاص.

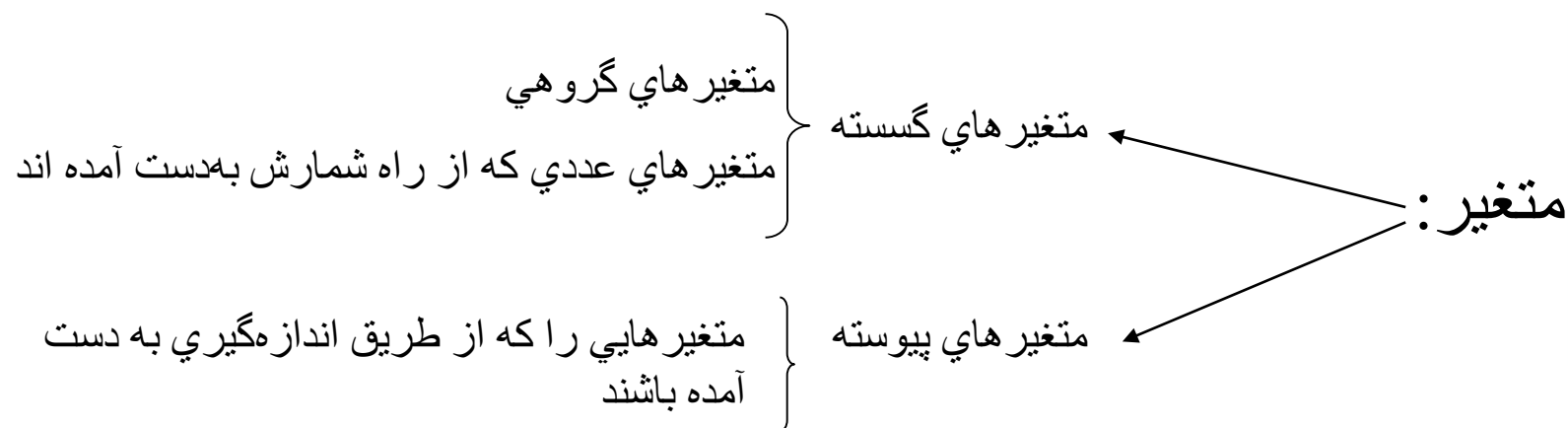
داده
مقیاس‌های اندازه‌گیری
متغیر

نمونه
جمعیت

آمار توصیفی:

داده
مقیاسهای اندازه‌گیری
متغیر

نمونه
جمعیت



آمار توصیفی:

داده

مقیاسهای اندازه گیری

متغیر

نمونه

جمعیت

در بسیار از مسائل پیش‌رو، اندازه‌گیری ویژگی يك متغیر مستلزم آگاهی و شناخت خاصی است. به طور کلی چهار نوع مقیاس برای اندازه گیری وجود دارد:

- مقیاس اسمی
- مقیاس ترتیبی
- مقیاس فاصله‌ای
- مقیاس نسبی

آمار توصیفی:

مقیاس اسمی:

این نوع مقیاس اندازه‌گیری عمدتاً برای طبقه‌بندی داده‌ها به کار می‌رود و منظور از آن اتلاق يك عدد طبیعی به داده‌های متفاوت است.

اختصاص اعداد ۱ تا ۴ به گروه‌های خونی A, B, AB, O.

مثال:

توجه داشته باشید که:

این اعداد را نمی‌توان برای مقایسه یا چهار عمل اصلی به کار برد

داده

مقیاسهای اندازه‌گیری

متغیر

نمونه

جمعیت

آمار توصیفی:

مقیاس ترتیبی:

این نوع مقیاس اندازه‌گیری عموماً برای طبقه‌بندی داده‌ها به منظور یک نوع برتری به کار می‌رود.

مثال:

در یک کارخانه ممکن است کارگران را به سه دسته ساده، نیمه ماهر و ماهر تقسیم‌بندی کنیم. اتلاق به ترتیب اعداد ۱ تا ۳ به این سه دسته یک مقیاس ترتیبی است.

نوجه داشته باشید که:

این اعداد تنها برای مقایسه به کار می‌روند و نمی‌توان با آنها چهار عمل اصلی را انجام داد.

داده

مقیاسهای اندازه‌گیری

متغیر

نمونه

جمعیت

آمار توصیفی:

مقیاس فاصله ای:

این نوع مقیاس اندازه‌گیری عموماً در زمینه‌هایی که علاوه بر حفظ ترتیب به نحوی فاصله بین ویژگی‌ها را نیز حفظ می‌کند. به عبارت دیگر در چنین مقیاسی نسبت تفاضلها ثابت می‌ماند.

مثال:

اندازه‌گیری ضریب هوشی دانش آموزان کلاس اول دبستان در شهر اصفهان.

توجه داشته باشید که:

در این نوع مقیاس، عدد صفر يك مفهوم قرار داده است.

داده
مقیاسهای اندازه‌گیری

متغیر

نمونه

جمعیت

آمار توصیفی:

داده

مقیاسهای اندازه گیری

متغیر

نمونه

جمعیت

مقیاس نسبتي:

این نوع مقیاس اندازه گیری علاوه بر حفظ فاصله، نسبت را نیز حفظ می کند. به عبارت دیگر در این نوع اندازه گیری نسبت دو مقدار بستگی به واحد اندازه گیری ندارد.

آمار توصیفی:

داده

مقیاسهای اندازه گیری

متغیر

نمونه

جمعیت

اطلاعاتی که از مطالعه يك متغیر به دست می آیند، معمولاً شامل انبوهی عدد یا علامت می باشند که آنها را داده می نامیم. داده ها را نسبت به نوع تغییری که اندازه گیری می کنیم به دو دسته داده گسسته و داده های پیوسته تقسیم می کنیم.

معمولاً به داده های جمع آوری شده که انبوهی عدد است و هیچ نوع پردازشی روی آنها انجام نشده است داده خام می گویند.

داده خام



آمار توصیفی:

داده

مقیاسهای اندازه گیری

متغیر

نمونه

جمعیت

مواردی که در ارتباط با يك مجموعه از داده‌های می‌بایستی مد نظر قرار داد، عبارت‌اند از:

= خلاصه کردن و توضیح داده‌ها به وسیله تنظیم جداول و رسم نمودارها.
= محاسبه مقادیر عددی، برای دست یافتن به معیارهایی که تمرکز و یا پراکندگی داده‌ها را نشان دهد.

در آمار، برای اینکه از داده‌های خام واقعیتهای موجود را استخراج کنیم، آنها را به نحوی مناسب دسته‌بندی کرده و جدولهایی به نام جدولهای آماری تهیه می‌نماییم. متداولترین جدول در آمار، جدول فراوانی است.

پیش از آنکه نحوه تنظیم جدول فراوانی را بیان نماییم، اطلاع از اصطلاحات زیر ضروری است.

جدول‌های آماری:

فراوانی نسبی
فراوانی نسبی
فراوانی نسبی
فراوانی نسبی

هرگاه n داده $y_1, y_2, \dots, y_n$ از k نوع $x_1, x_2, \dots, x_k$ ، با فرض $2 \leq k \leq n$ ،
به ترتیب با تعدادهای $f_1, f_2, \dots, f_k$ تشکیل شده باشند، آنگاه f_i را فراوانی x_i ،
می‌گوییم. به عبارت دیگر تعداد دفعاتی را که x_i در داده‌های $y_1, y_2, \dots, y_n$
تکرار می‌شود، فراوانی x_i می‌نامیم و آن را با نماد f_i نمایش می‌دهیم.

به خاطر داشته باشید که

اگر اندازه نمونه برابر n باشد، آنگاه برای $j = 1, \dots, k$

$$\sum_{i=1}^k f_i = n$$

$$1 \leq f_i \leq n$$

جدول‌هاي آماری:

مثال:

داده‌هاي زیر میزان تصادف منجر به مرگ رد ۳۰ منطقه را نشان می‌دهد. فراواني داده‌ها را تعیین نمایید.

۸ ۶ ۵ ۵ ۳ ۴ ۳ ۶ ۶ ۷
۳ ۵ ۵ ۸ ۵ ۷ ۴ ۸ ۴ ۳
۲ ۸ ۷ ۶ ۵ ۶ ۶ ۵ ۵ ۶

مشاهده می‌شود که داده‌هاي تکرار اعداد ۲، ۳، ۴، ۵، ۶، ۷، ۸ می‌باشند، بنابراین جدول زیر را برای فراواني داده‌ها خواهیم داشت:

x_i	f_i
۲	۱
۳	۴
۴	۳
۵	۸
۶	۷
۷	۳
۸	۴

جدول‌هاي آماری:

نسبت فراواني به اندازه نمونه را فراواني نسبي مي‌ناميم. اگر فراواني x_i در يك نمونه با اندازه n ، برابر f_i باشد، آنگاه فراواني نسبي x_i را با نماد r_i نمايش خواهيم داد، به طوري كه:

$$r_i = \frac{f_i}{n}$$

به خاطر داشته باشيد كه

براي $j = 1, \dots, k$

$$\sum_{i=1}^k r_i = n$$

$$0 \leq r_i \leq 1$$

جدول‌هاي آماری:

فراواني نسبي تجمعي
فراواني تجمعي
فراواني نسبي

فراواني

x_i	f_i	r_i
۲	۱	۰,۰۳۳
۳	۴	۰,۱۳۳
۴	۳	۰,۱۰۰
۵	۸	۰,۲۶۷
۶	۷	۰,۲۳۳
۷	۳	۰,۱۰۰
۸	۴	۰,۱۳۳

$$\frac{1}{30} = 0.033$$

$$\frac{4}{30} = 0.133$$

جدول‌هاي آماری:

فراواني نسبي تجمعي
فراواني تجمعي

فراواني نسبي
فراواني

با توجه به تعريف فراواني، فراواني تجمعي ردیف i را با نماد F_i نمایش می‌دهیم و به صورت زیر تعریف می‌کنیم:

$$F_i = \sum_{j=1}^i f_j$$

به خاطر داشته باشید که

برای اندازه نمونه n و $i = 1, \dots, k$ آنگاه

$$f_i = F_1 \leq F_1 \leq \dots \leq F_k = n$$

جدول‌هاي آماری:

فراوانی نسبی تجمعی
فراوانی تجمعی

فراوانی نسبی
فراوانی

x_i	f_i	F_i
۲	۱	۱
۳	۴	۵
۴	۳	۸
۵	۸	۱۶
۶	۷	۲۳
۷	۳	۲۶
۸	۴	۳۰

$$1 + 4 = 5$$

$$1 + 4 + 3 + 8 + 7 = 23$$

جدول‌هاي آماری:

فراواني نسبي تجمعي

فراواني تجمعي

فراواني نسبي

فراواني

با توجه به تعريف فراواني نسبي، فراواني نسبي تجمعي ردیف i را با نماد R_i نماد نمايش مي‌دهيم و به صورت زیر تعريف مي‌کنيم:

$$R_i = \sum_{j=1}^i r_j$$

به خاطر داشته باشید که

براي اندازه نمونه n و $i = 1, \dots, k$ آنگاه

$$r_i R_1 \leq R_2 \leq \dots \leq R_k = 1$$

جدول های آماری:

فراوانی نسبی تجمعی

فراوانی تجمعی

فراوانی نسبی

فراوانی

x_i	r_i	R_i
۲	$\frac{1}{30}$	$\frac{1}{30}$
۳	$\frac{4}{30}$	$\frac{5}{30}$
۴	$\frac{3}{30}$	$\frac{8}{30}$
۵	$\frac{8}{30}$	$\frac{16}{30}$
۶	$\frac{7}{30}$	$\frac{23}{30}$
۷	$\frac{3}{30}$	$\frac{26}{30}$
۸	$\frac{4}{30}$	$\frac{30}{30}$

$$\frac{1}{30} + \frac{4}{30} + \frac{3}{30}$$

جدول های آماری:

مثال:

معدل ۵۰ دانشجوی دانشگاه با تقریب تا يك رقم اعشار، به شرح زیر است:

۹/۲	۵/۱	۸/۱	۴/۲	۲/۲	۱/۲	۲/۲	۶/۱	۹/۱	۱/۲
۶/۲	۱/۲	۵/۲	۰/۲	۳/۲	۳/۲	۷/۱	۸/۱	۳/۲	۸/۱
۴/۱	۶/۲	۲/۲	۹/۱	۰/۲	۷/۱	۷/۱	۹/۱	۱/۲	۸/۱
۰/۲	۰/۲	۰/۲	۵/۲	۲/۲	۲/۲	۹/۱	۸/۱	۴/۲	۹/۲
۴/۱	۶/۱	۹/۱	۹/۲	۴/۲	۶/۱	۸/۱	۹/۱	۵/۲	۴/۱

تشکیل جدول فراوانی
برای
داده های پیوسته

چون داده ها تا يك رقم اعشار گرد شده اند، بنابراین می توان گفت که اندازه واقعی معدلها

در فاصله

[1/35,2/95]

$$U = MAX + 0.5 \times 10^{-r}$$

$$L = MIN - 0.5 \times 10^{-r}$$

تعداد ارقام گرد شده

جدول های آماری:

فراوانی نسبی تجمعی
فراوانی نسبی
فراوانی نسبی
فراوانی نسبی

$$\frac{U - L}{c} = \frac{2.95 - 1.35}{8} = 0.2$$

تعداد طبقات

R_i	F_i	r_i	f_i	x_i	کلاس
۰.۸/۰	۴	۰.۸/۰	۴	۴۵/۱	۳۵/۱ _ ۵۵/۱
۲۰/۰	۱۰	۱۲/۰	۶	۶۵/۱	۵۵/۱ _ ۷۵/۱
۴۴/۰	۲۲	۲۴/۰	۱۲	۸۵/۱	۷۵/۱ _ ۹۵/۱
۶۲/۰	۳۱	۱۸/۰	۹	۰۵/۲	۹۵/۱ _ ۱۵/۲
۷۸/۰	۳۹	۱۶/۰	۸	۲۵/۲	۱۵/۲ _ ۳۵/۲
۹۰/۰	۴۵	۱۲/۰	۶	۲/۴۵	۳۵/۲ _ ۵۵/۲
۹۴/۰	۴۷	۰.۴/۰	۲	۶۵/۲	۵۵/۲ _ ۷۵/۲
۰۰/۱	۵۰	۰.۶/۰	۳	۸۵/۲	۷۵/۲ _ ۹۵/۲
—	—	۰۰/۱	۵۰		

جمع

نمودارهاي آماری:

معمولا داده‌ها را با نمودارهاي مختلف نمایش می‌دهند. عموما این نمودارها در ارتباط با داده‌هاي پیوسته به کار گرفته می‌شود و منظور از نمایش آنها، **تجسم عيني اطلاعات نهفته در داده‌ها است.** در این بخش به معرفی چند نمودار معروف اکتفا می‌کنیم:

= هیستوگرام

= چندبر فراواني

= چندبر فراواني تجمعي

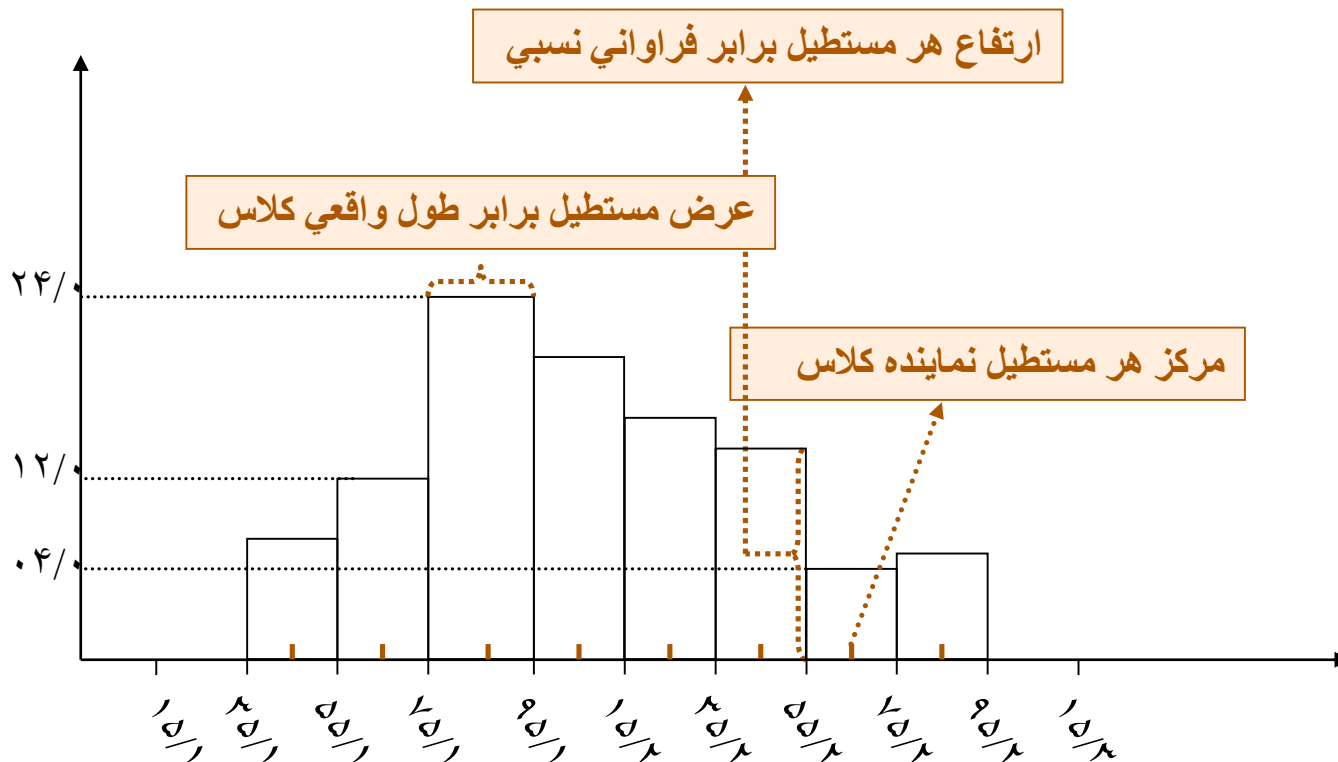
= منحنیهاي فراواني و فراواني تجمعي

= نمایش نمودار تنه و شاخه

= نمودار جعبه‌اي

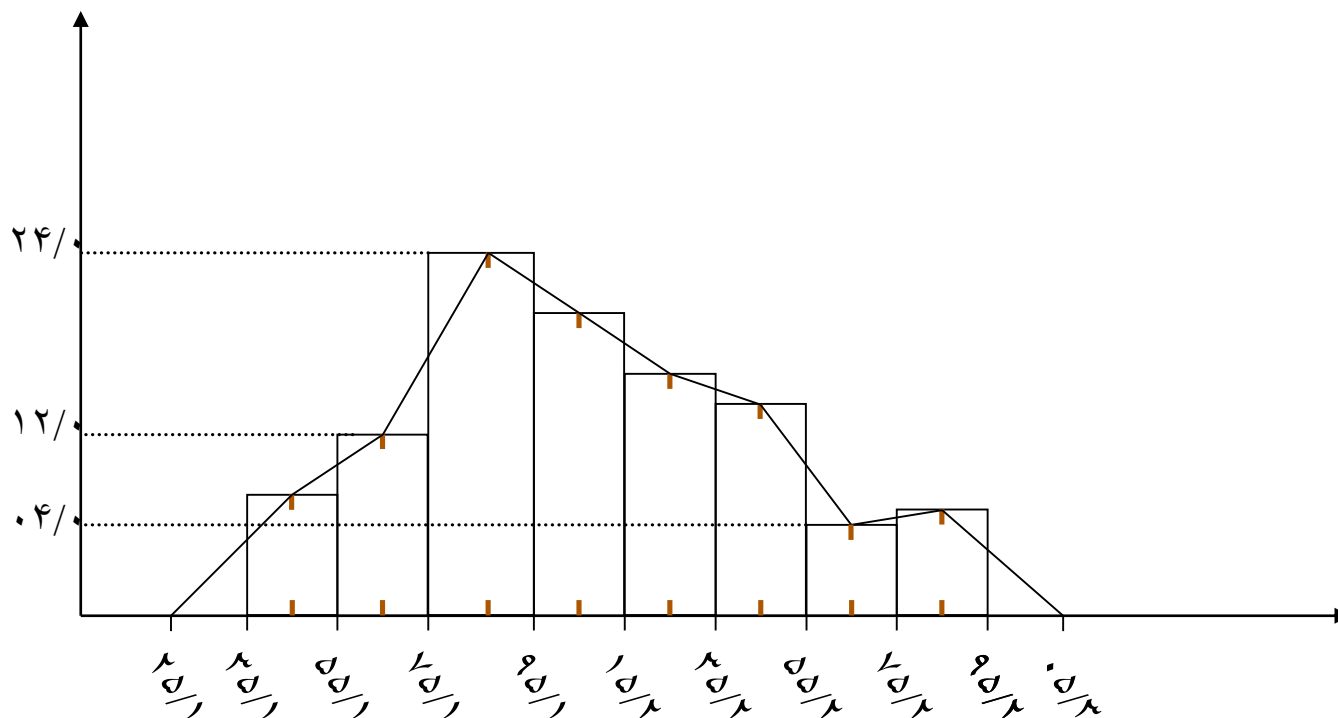
نمودارهاي آماری:

نمایش نمودار تته و شاخه
منحنی های فراوانی و.....
چندبر فراوانی، تجمعی
چندبر فراوانی
هیستوگرام



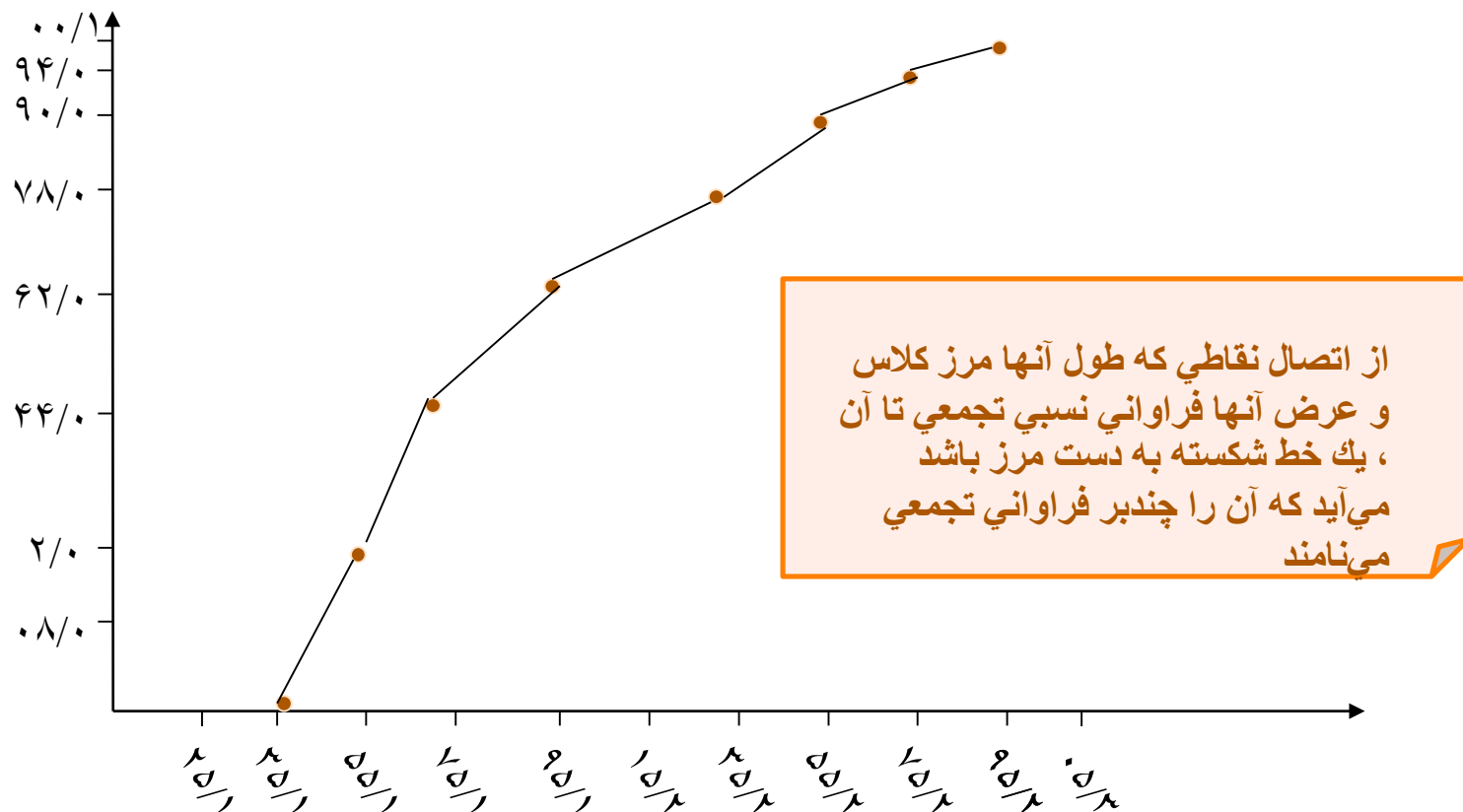
نمودارهاي آماري:

نمایش نمودار تته و شاخه
منحنی های فراوانی و.....
چندبر فراوانی تجمعی
چندبر فراوانی
هیستوگرام



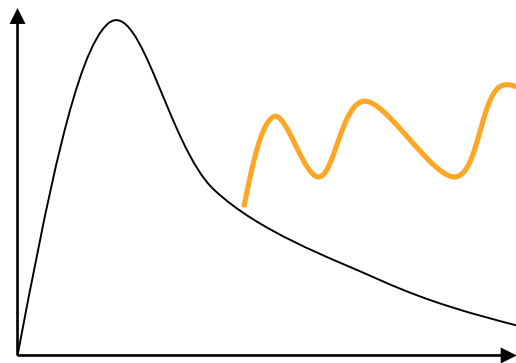
نمودارهاي آماری:

نمایش نمودار تنه و شاخه
منحنی های فراوانی و.....
چندبر فراوانی تجمعی
هیستوگرام
چندبر فراوانی



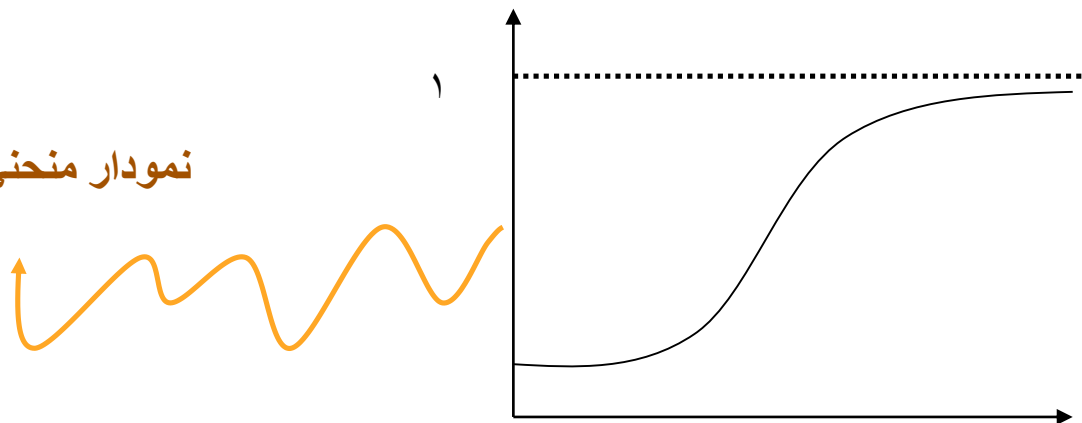
نمودارهاي آماري:

نمایش نمودار تته و شاخه
منحنی های فراوانی و.....
چندبر فراوانی تجمعی
چندبر فراوانی
هیستوگرام



نمودار منحنی فراوانی

نمودار منحنی فراوانی تجمعی



نمودارهاي آماري:

نمایش نمودار تنه و شاخه
منحني هاي فراواني و.....
چندبر فراواني تجمعي
چندبر فراواني
هستوگرام

۲ **
۳ ****
۴ ***
۵ *****
۶ *****
۷ ***
۸ ****



فراواني

۴/۱ ***
۵/۱ *
۶/۱ ***
۷/۱ ***
۸/۱ *****
۹/۱ *****
۰/۲ *****
۱/۲ *****
۲/۲ *****
۳/۲ ***
۴/۲ ***
۵/۲ ***
۶/۲ **
۷/۲
۸/۲
۹/۲ ***



نمودارهاي آماری:

نمایش نمودار تنه و شاخه
منحنی های فراوانی و.....
چندبیر فراوانی تجمعی
چندبیر فراوانی
هیستوگرام

نمرات ۸۰ دانشجو در امتحانات نهایی درس احتمال و آمار به شرح زیر است:

۶۸ ۸۴ ۷۵ ۸۲ ۶۸ ۹۰ ۶۲ ۸۸ ۷۶ ۹۳
۷۳ ۷۹ ۸۸ ۷۳ ۶۰ ۹۳ ۷۱ ۵۹ ۸۵ ۷۵
۶۱ ۶۵ ۷۵ ۸۷ ۷۴ ۶۲ ۹۵ ۷۸ ۶۳ ۷۲
۶۶ ۷۸ ۸۲ ۷۵ ۹۴ ۷۷ ۶۹ ۷۴ ۶۸ ۶۰
۹۹ ۷۸ ۸۹ ۶۱ ۷۵ ۹۵ ۶۰ ۷۹ ۸۳ ۷۱
۷۹ ۶۲ ۶۷ ۹۷ ۷۸ ۸۵ ۷۶ ۶۵ ۷۱ ۷۵
۶۵ ۸۰ ۷۳ ۵۷ ۸۸ ۷۸ ۶۲ ۷۶ ۵۰ ۷۴
۸۶ ۶۷ ۷۳ ۸۱ ۷۲ ۶۳ ۷۶ ۷۵ ۸۵ ۷۷



نمودارهاي آماری:

نمایش نمودار تته و شاخه
منحني هاي فراواني و.....
چندبر فراواني تجمعي
چندبر فراواني
هستوگرام

پس از ساختن نمودار اولیه معمولاً بهتر است مقادير هر شاخه را از كوچك به بزرگ، با تعداد دفعات مرتب كرد، به صورت زیر: تکرار،

۵	۰ ۷ ۹
۶	۰ ۰ ۰ ۱ ۱ ۲ ۲ ۲ ۲ ۳ ۳ ۵ ۵ ۵ ۶ ۷ ۸ ۸ ۹
۷	۱ ۱ ۱ ۲ ۲ ۳ ۳ ۳ ۳ ۴ ۴ ۴ ۵ ۵ ۵ ۵ ۵ ۵ ۵ ۶ ۶ ۶ ۶ ۷ ۷ ۸ ۸ ۸ ۸ ۸ ۹ ۹ ۹
۸	۰ ۱ ۲ ۲ ۳ ۴ ۵ ۵ ۵ ۶ ۷ ۸ ۸ ۸ ۹
۹	۰ ۳ ۳ ۴ ۵ ۵ ۷ ۹

نما

چندگانه

میانگین

میانگین

معیارهای مرکزی:

با استفاده از جدول فراوانی و رسم نمودارها می‌توانیم داده‌ها را به نحو مطلوبی **تنظیم کرده و اطلاعات** بهتر است آنها را در یک **نهفته را تا حدودی مشخص کنیم**. با این حال برای آرایه یک گزارش مناسب، یا چند عدد مناسب نیز خلاصه کنیم. چنین عددی می‌تواند معیار مرکزی باشد. مهمترین معیارهای ، میانه و نما است که در بخش این به شرح هر یک از آنها خواهیم پرداخت. مرکزی میانگین

هرگاه n داده $y_1, y_2, \dots, y_n$ از k نوع $x_1, x_2, \dots, x_k$ ، با فرض $2 \leq k \leq n$ ، به ترتیب با تعدادهای $f_1, f_2, \dots, f_k$ تشکیل شده باشند، آنگاه f_i را فراوانی x_i می‌گوییم.

$\bar{x} = \frac{1}{n} \sum_{i=1}^n f_i x_i$	میانگین حسابی
$\bar{x}_w = \frac{\sum_{i=1}^k w_i x_i}{\sum_{i=1}^k w_i}$	میانگین وزنی
$G = \left(\prod_{i=1}^k x_i^{f_i} \right)^{\frac{1}{n}}$ کلیه داده‌ها بزرگتر از صفر باشند	میانگین هندسی

معیارهای مرکزی:

نما

چندکها

میانه

میانه‌گین

اگر داده‌ها را از کوچک به بزرگ مرتب نماییم، عدد m را میانه این داده‌ها می‌نامیم، اگر نصف داده‌ها در سمت چپ و نصف داده در سمت راست این عدد قرار گیرد

محاسبه میانه برای داده‌های گسسته

فرض کنید $y_1, y_2, \dots, y_n$ داده‌های ما باشند و شکل مرتب شده آنها را با $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ نمایش دهیم آنگاه

$$M = \begin{cases} y_{(\frac{n+1}{2})} & \text{اگر } n \text{ فرد باشد} \\ \frac{1}{2} [y_{(\frac{n}{2})} + y_{(\frac{n}{2}+1)}] & \text{اگر } n \text{ زوج باشد} \end{cases}$$

نما

چندکها

میانها

میانگین

معیارهای مرکزی:

محاسبه میانه برای داده‌های پیوسته

کلاس	x_i	f_i	F_i
۳۵/۱ - ۵۵/۱	۴۵/۱	۴	۴
۵۵/۱ - ۷۵/۱	۶۵/۱	۶	۱۰
۷۵/۱ - ۹۵/۱	۸۵/۱	۱۲	۲۲
۹۵/۱ - ۱۵۵/۲	۰۵/۲	۹	۳۱
۱۵۵/۲ - ۳۵۵/۲	۲۵۵/۲	۸	۳۹
۳۵۵/۲ - ۵۵۵/۲	۲/۴۵	۶	۴۵
۵۵۵/۲ - ۷۵۵/۲	۶۵۵/۲	۲	۴۷
۷۵۵/۲ - ۹۵۵/۲	۸۵۵/۲	۳	۵۰
جمع		۵۰	—

$$\frac{n}{2} = 25$$

$$m = L_m + \left(\frac{\frac{n}{2} - F_b}{f_m} \right) w$$

طول هر رده

معیارهای مرکزی:

چندك يك معیار کلی‌تر از میانه است و در عنوان حالت خاص میانه را نیز در بر می‌گیرد. اگر p يك عدد حقیقی بین صفر و يك باشد، آنگاه عدد Q_p را چندك مرتبه p می‌نامیم هر گاه 100%

داده‌ها سمت چپ و $(1-p) \cdot 100\%$ داده‌ها سمت راست Q_p باشند.

چندكهای معروف عبارتند از :

چارکها

چارکها به ازاي $p = 25\%, 50\%, 75\%$ به دست می‌آیند و آنها را به ترتیب با نماد Q_1 (چارک اول)، Q_2 (چارک دوم) و Q_3 (چارک سوم) نشان می‌دهند.

دهکها

دهکها به ازاي $p = 10\%, 20\%, \dots, 90\%$ به دست می‌آیند و آنها را به ترتیب با نماد D_1 (دهک اول)، D_2 (دهک دوم)، $\dots$ و D_9 (دهک نهم) نشان می‌دهند.

صدکها به ازاي $p = 1\%, 2\%, \dots, 99\%$ به دست می‌آیند و آنها را به ترتیب با نماد P_1 (صدک اول)، P_2 (صدک دوم)، $\dots$ و P_{99} (صدک نود و نهم) نشان می‌دهند.

معیارهای مرکزی:

محاسبه چنك براي داده‌های گسسته

فرض کنید $y_1, y_2, \dots, y_n$ داده‌های ما باشند و شکل مرتب شده آنها را با $y_{(1)}, y_{(2)}, \dots, y_{(n)}$

نمایش دهیم. برای محاسبه چنك

$(n+1)p$ صحیح باشد $r = (n+1)p, \quad Q_p = y_{(r)}$

$(n+1)p$ صحیح نباشد $r = [(n+1)p], \quad \omega = (n+1)p - r \rightarrow Q_p = (1-\omega)y_{(r)} + \omega y_{(r+1)}$

نمودارهاي آماری:

محاسبه چنك براي داده‌هاي پيوسته

با توجه به ستون فراواني جمعي در جدول فراواني،
كلاسي را كه چنك در آن قرار دارد مشخص
مي‌كنيم.

(p) (n)

$$0.25 \times 50 = 12.5$$

كلاس	x_i	f_i	F_i
۳۵/۱ _ ۵۵/۱	۴۵/۱	۴	۴
۵۵/۱ _ ۷۵/۱	۶۵/۱	۶	۱۰
۷۵/۱ _ ۹۵/۱	۸۵/۱	۱۲	۲۲
۹۵/۱ _ ۱۵/۲	۰۵/۲	۹	۳۱
۱۵/۲ _ ۳۵/۲	۲۵/۲	۸	۳۹
۳۵/۲ _ ۵۵/۲	۲/۴۵	۶	۴۵
۵۵/۲ _ ۷۵/۲	۶۵/۲	۲	۴۷
۷۵/۲ _ ۹۵/۲	۸۵/۲	۳	۵۰
جمع		۵۰	—

$$Q_p = L_{Q_p} + \left(\frac{np - F_b}{f_{Q_p}} \right) w$$

$$Q_p = 1.75 + \left(\frac{50 \times 0.25 - 10}{12} \right) \times 0.2$$

نما

چندکها

میانه

میانتگین

نمودارهای آماری:

محاسبه نما برای داده‌های گسسته

داده‌ای که فراوانی آن نسبت به دیگر داده‌ها بیشتر باشد، نما یا مد نامیده می‌شود و آن را با نماد M نمایش می‌دهیم. برای به دست آوردن نما، نخست فراوانی داده‌ها را پیدا می‌کنیم و داده‌ای را که فراوانی آن بیشتر باشد، به عنوان نما اختیار می‌کنیم و اگر دو داده، دارای فراوانی یکسان و بیش از دیگر فراوانی‌ها باشند، هر دو را به عنوان نما اختیار می‌کنیم و داده‌ها را دو نمایی می‌گوییم، به شرط آن که این دو داده در یک صف غیر نزولی، کنار هم نباشند. در صورتی که این دو داده در یک صف غیر نزولی، کنار هم باشند نصف مجموع آنها را به عنوان نما اختیار می‌کنیم. اگر تمام داده دارای فراوانی یکسان باشند، می‌گوییم داده‌ها بدون نما هستند. به یاد داشته باشید که نما، به عنوان یک معیار تمرکز در داده‌های گروهی به کار گرفته می‌شود.

نما

چندکها

میانه

میانیگین

نمودارهای آماری:

مثال: برای داده‌های ۲، ۲، ۵، ۷، ۹، ۹، ۹، ۱۰، ۱۰، ۱۱، ۱۲ و ۱۸ نما برابر $M=9$ است، زیرا فراوانی داده ۹ بیش از فراوانی دیگر داده‌ها است.

مثال: برای داده‌ها ۲، ۳، ۴، ۴، ۴، ۵، ۵، ۷، ۷ و ۹، دو داده ۴ و ۷ به عنوان نما اختیار می‌شوند، زیرا فراوانی این دو داده، بیش از فراوانی داده‌های دیگر است.

مثال: برای داده‌های ۳، ۵، ۸، ۱۰، ۱۲، ۱۵ و ۱۶، نما وجود ندارد، زیرا تمام داده‌ها دارای فراوانی یکسان هستند.

مثال: برای داده‌ها ۲، ۳، ۴، ۴، ۴، ۵، ۵، ۵، ۷ و ۹ دو داده ۴ و ۵ را که بیشترین فراوانی هستند به عنوان نما بر می‌گزینیم، اما از آنجا که این دو داده در یک صف غیر نزولی در کنار یکدیگر قرار دادند، نصف مجموع دو داده به عنوان نما اختیار می‌شود، یعنی $M=5/4$.

نما

چندگه

میان

میانگین

نمودارهای آماری:

محاسبه نما برای داده‌های پیوسته

کلاس	X_i	f_i	r_i
۳۵/۱_۵۵/۱	۴۵/۱	۴	۰.۸/۰
۵۵/۱_۷۵/۱	۶۵/۱	۶	۱۲/۰
۷۵/۱_۹۵/۱	۸۵/۱	۱۲	۲۴/۰
۹۵/۱_۱۵/۲	۰.۵/۲	۹	۱۸/۰
۱۵/۲_۳۵/۲	۲۵/۲	۸	۱۶/۰
۳۵/۲_۵۵/۲	۲/۴۵	۶	۱۲/۰
۵۵/۲_۷۵/۲	۶۵/۲	۲	۰.۴/۰
۷۵/۲_۹۵/۲	۸۵/۲	۳	۰.۶/۰
جمع		۵۰	۰.۰/۱

از روی جدول ملاحظه می‌شود که فراوانی رده ۷۵/۱_۹۵/۱ دارای بیشترین فراوانی است بنابراین به عنوان رده نما در نظر می‌گیریم.

$$M = L_M + \frac{d_1}{d_1 + d_2} \times \omega$$



$$M = 1.75 + \frac{0.12}{0.12 + 0.06} \times 0.2$$

معیارهای پراکندگی:

با وجود این که در بسیاری از موارد، میانگین توصیف نسبتاً کاملی از مجموعه داده‌ها ارائه می‌دهد، اما گاهی وجود اطلاعات بیشتر در مورد داده‌ها ضروری است. يك مفهوم مهم در ارتباط با داده‌های آماری، میزان تغییرات آنهاست، بدین معنی که اندازه‌گیری‌ها تا چه اندازه از فردی به فرد دیگر یا شیی به شیی دیگر تغییر می‌کنند. در این بخش، به بررسی و محاسبه میزان تغییرات به عنوان معیارهای پراکندگی خواهیم پرداخت. مهمترین معیارهای پراکندگی عبارتند از **دامنه، میانگین انحراف ها از میانگین یا از میانه، میان دامنه چارکها، دامنه صدکي، واریانس و انحراف معیار** است. علاوه بر مطالب فوق، در این بخش داده‌های استاندارد و ضریب تغییرات را نیز معرفی خواهیم کرد.

معیارهای پراکندگی:

اگرچه دامنه يك وسیله ساده برای اندازه‌گیری اختلاف و پراکندگی در يك سری از داده‌ها است، اما در بیشتر موارد رضایتبخش نیست. داده‌های بسیار بزرگ یا بسیار کوچک مانع از آنند که دامنه، معرف واقعی میزان انحراف باشد. در چنین مورد قبول همگان به

با توجه به اینکه در محاسبه واریانس داده‌ها را مربع می‌کنیم، بدین جهت ریشه دوم مثبت آن را که انحراف معیار یا انحراف استاندارد می‌نامیم، به عنوان يك معیار پراکندگی بر مبنای مقیاس می‌بریم.

میانگین انحرافها از میانگین

میانگین انحرافها از میانه

ضریب تغییر عبارت است از اندازه نسبی انحراف معیار در مقایسه با میانگین. ضریب همبستگی به واحد اندازه‌گیری وابسته نیست و برای مقایسه جمعیت‌های یکسان به کار می‌رود. در مقایسه هر اندازه که ضریب تغییر ویژگی جمعیتی کمتر باشد، ویژگی آن جمعیت بهتر ارزیابی می‌شود.

دامنه	$y_{(1)} \text{ کوچکترین و } y_{(n)}$
میانگین انحرافها از میانگین	
میانگین انحرافها از میانه	
انحراف معیار	$MD_m = \sum_{i=1}^k f_i x_i - \bar{x} $
ضریب تغییرات	$IQR = Q_3 - Q_1$
	$IPR = P_{90} - P_{10}$
	$s^2 = \frac{1}{n} \sum_{i=1}^k f_i (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k f_i x_i^2 - \bar{x}^2$
	$s = \sqrt{\frac{1}{n} \sum_{i=1}^k f_i (x_i - \bar{x})^2}$
	$v = \frac{s}{\bar{x}}$

معیارهای پراکندگی:

اگر $y_1, y_2, \dots, y_k$ نمایانگر داده‌های خام باشند، براساس جدول فراوانی f_1 تا f_k تا از y_i ها برابر x_1 ، f_2 تا برابر x_2 ، و f_k تا برابر x_k است. می‌دانیم که S و $\bar{x}$ به ترتیب میانگین و انحراف معیار داده است. اگر از هر داده $\bar{x}$ را کم و بر S تقسیم کنیم، یعنی

داده‌های استاندارد

$$z_i = \frac{x_i - \bar{x}}{S}, \quad i=1, \dots, k$$

آنگاه $z_1, z_2, \dots, z_k$ با فراوانی‌های به ترتیب $f_1, f_2, \dots, f_k$ را داده‌های استاندارد می‌نامیم. به سادگی می‌توان نشان داد که داده‌های استاندارد دارای میانگین برابر با صفر و واریانس برابر با یک هستند و به واحد اندازه‌گیری بستگی ندارند.

معیارهای پراکندگی:

چون $V_1 > V_2$ ، بنابراین نتیجه می‌گیریم که دانشجویان در امتحان دوم نمرات مطلوب‌تری را کسب کرده‌اند.

ب) برای مقایسه، ابتدا نمرات دانشجو را استاندارد می‌کنیم

$$z_1 = \frac{65 - 60}{6} = \frac{5}{6}, \quad z_2 = \frac{720 - 700}{7} = \frac{20}{7}$$

چون $z_1 < z_2$ ، بنابراین نمره آزمون دوم دانشجو در مقایسه از موقعیت بهتری برخوردار است.

معیارهای پراکندگی

• فرض کنید يك دسته از دانشجویان در دو امتحان شرکت کرده‌اند و خلاصه نتایج آزمونها به شرح زیر است.

• آزمون اول: میانگین نمرات برابر ۶۰، انحراف معیار برابر ۶، ماکزیمم نمره از ۱۰۰

• آزمون دوم: میانگین نمرات ۷۰۰، انحراف معیار برابر ۷، ماکزیمم نمره از ۱۰۰۰

• الف- چگونه این دو نتیجه را با هم مقایسه و ارزیابی می‌کنید؟

• ب- اگر دانشجویی در آزمون اول نمره ۶۵ و در آزمون دوم نمره ۷۲۰ را کسب کرده باشد، وضعیت دانشجو در کدام آزمون مطلوبتر است؟

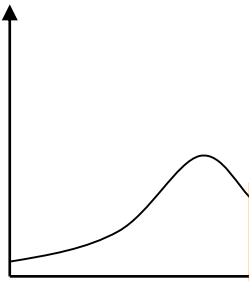
• حل: الف) با محاسبه ضریب تغییر دو آزمون معلوم می‌شود که

$$v_1 = \frac{s_1}{x_1} = \frac{6}{60} = \frac{1}{10} = 10\%$$

$$v_2 = \frac{s_1}{x_2} = \frac{7}{700} = \frac{1}{100} = 1\%$$

منحنیهای فراوانی:

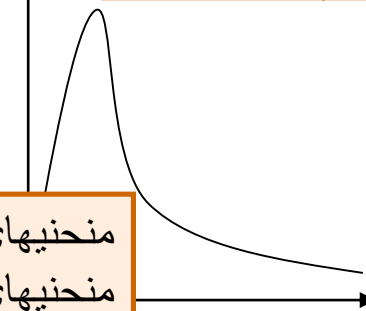
پخ



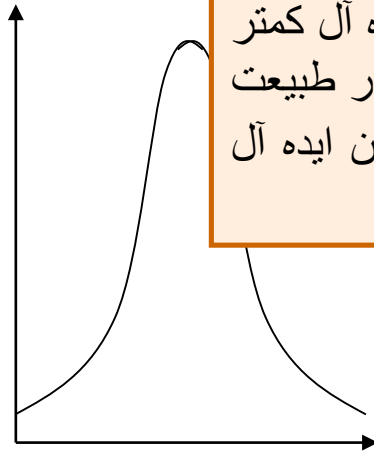
نامتقارن / چوله به چپ



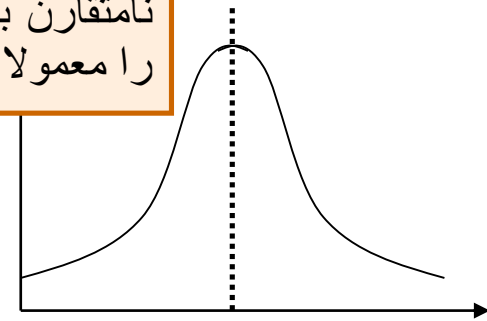
نامتقارن / چوله به راست



برجسته

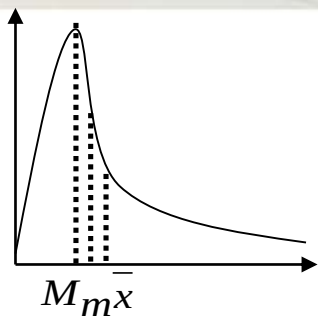


منحنی نرمال



منحنیهای فراوانی در طبیعت تنوع زیادی دارند، اما بسیاری از منحنیهای فراوانی تک نمایی یا متقارن هستند یا چوله و یا برجسته و یا پخ. ایده آلترین منحنی فراوانی متقارن، منحنی فراوانی نرمال استاندارد است. برای منحنیهای فراوانی کاملاً متقارن تک نمایی مقادیر میانگین، میانه و نما بر هم منطبق می‌شوند. در طبیعت، عموماً منحنی فراوانی متقارن ایده آل کمتر یافت می‌شود و بسیاری از منحنیهای فراوانی موجود در طبیعت نامتقارن برجسته یا پخ هستند. میزان انحراف از تقارن ایده آل را معمولاً با دو معیار چولگی و برجستگی می‌سنجند.

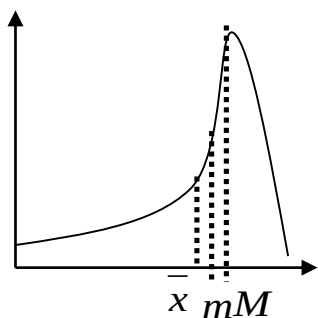
منحنیهای فراوانی:



$$M < m < \bar{x}$$

به راست یا مثبت

چولگی:



$$M > m > \bar{x}$$

به چپ یا منفی

معیار اندازه گیری چولگی:

$$\text{ضریب چولگی اول پیرسن} = \frac{\bar{x} - M}{s}$$

منحنیهای فراوانی:

میزان کشیدگی یا پخی منحنی فراوانی را یسبت به منحنی نرمال استاندارد، برجستگی منحنی فراوانی می نامیم. فرمول زیر را می توان به عنوان معیار برجستگی به کار برد.

$$\text{ضریب برجستگی} = k = \frac{(Q_3 - Q_1) / 2}{P_{90} - P_{10}}$$

نشان داده شده است که برای منحنی فراوانی نرمال استاندارد $k=0.263$ ، بنابراین معمولا ضریب برجستگی را به صورت زیر تعریف می کنند:

$$\text{ضریب برجستگی} = k = \frac{(Q_3 - Q_1) / 2}{P_{90} - P_{10}} - 0.263$$

برحسب آن که این مقدار مثبت یا منفی باشد گوییم منحنی فراوانی برجسته یا پخ است.