آشنایی با آمار توصیفي

پیشگفتار:

در عصر حاضر كسي نمي‌تواند منكر اين واقعيت باشد كه آمار نقشي لاينفک در زندگي روزمره ما بازي مي‌كند. اخبار روزانه رسانه‌هاي گروهی با گزارشی از وضع هوا به پايان مي‌رسندو در طول اخبار، به جريانهای بازار بورس و سهام اشاره مي‌شود و روزنامه‌ها خبر از افزايش نرخ اجناس مي‌دهندو...

آمار به عنوان پايه يك روش و راه موثر در بررسی مسائل موجود، در بسياری از زمينه‌هاي علمي از جمله جامعه شناسي، كشاورزی، فيزيك و....به كار گرفته مي‌شود. در دانش امروزي، معمولا سعی مي‌شود كه اطلاعات موجود در يك زمينه خاص، در قالب اعداد نمايش داده شود تا به هنگام تجزيه و تحليل اطلاعات، فهم بهتری از پديده مورد مطالعه به دست آمده و امكان مقايسه فراهم گردد. **در يك جمله آمار مجموعه‌اي از روشهای جمع آوری، تهيه وتنظيم و تجزيه و تحليل اطلاعات است كه براي كسب يك يا چند نتيجه به خدمت گرفته مي‌شود.**

فهرست مطالب:

آمار توصيفي:

براي اينكه نتايج مناسب و مطلوب از اطلاعات كه در آمار گيري‌ها جمع آوري مي‌كنيم، به دست آيد بايد:

- اعداد نماينده واقعي مشاهدات بوده و غيرواقع يا غلط نباشند
- به نحو مفيدي تهيه و تنظيم شوند
- به نحو صحيح تجزيه و تحليل گردند
- قابل نتيجه گيري صحيح باشند

آمار توصيفي:

به طور كلي، روش‌هايي را كه به وسيله آنها مي‌توان اطلاعات جمع آوري شده را تنظيم كرده و خلاصه نمود، آمار توصيفي مي‌ناميم و در يك كلام آمار توصيفي عبارت از مجموعه روشهايي است كه پردازش داده‌ها را فراهم مي‌سازد. اطلاع از اصطلاحات زير در آمار ضروري است.

آمار توصيفي:

داده
مقياسهاي اندازه‌گيري
نمونه
متغير
جمعيت

مجموعه افراد يا اشيايي را كه مي‌خواهيم يك يا چند خصوصيت مشترك آنها را مورد بررسی قرار دهيم، جمعيت يا جمعيت آماري مي‌ناميم.

مثال:

اندازه قد يا وزن دانشجويان بيست ساله يك شهر، تعداد لامپهاي سالم و يا ناسالم توليد شده در يك كارخانه و در يك روز معين، مثالهايي از جمعيتهاي آماري هستند.

نكته:

معمولا مطالعه ويژگي‌هاي مورد نظر، به هنگامی كه جمعيت آماری بسيار گسترده باشد، مستلزم صرف هزينه و وقت زيادي مي‌باشد و در بسياری از مواقع، اين امر اصولا امكان پذير نيست. بنابراين در چنين موردی، براي مطالعه ويژگی مورد نظر، به قسمتی از جمعيت آماری اكتفا مي‌كنيم.

داده
مقياسهاي اندازه‌گيري
متغير
نمونه
جمعيت

آمار توصيفي:

نمونه

قسمتي از جمعيت را كه طبق قاعده و ضوابط خاصي، براي مطالعه خصوصيتي از جمعيت انتخاب مي‌شود، يك نمونه از جمعيت مي‌ناميم.

نكته

اين نمونه وقتي مفيد و قابل قبول خواهد بود كه بتواند نماينده خوبي براي كل جمعيت مورد مطالعه باشد. با توجه به اهميت اين موضوع شاخه‌اي از آمار تحت عنوان نظريه نمونه‌گيري با بررسي نمونه‌اي به اين امر مهم مي‌پردازد. در بسياري از موارد، معمولا نمونه تصادفي ساده را در نظر مي‌گيرند.

مثال:

براي بررسی اندازه قد دانشجويان بيست ساله يك شهر، انتخاب مثلا ۱۵۰ نفر از بين اين جمعيت به طور تصادفی، يا انتخاب ۱۰۰ لامپ به تصادف از لامپهاي توليدي يك كارخانه در يك روز معين، براي تعيين كيفيت لامپهاي توليدي اين كارخانه مثالهايي از نمونه تصادفی هستند.

داده
مقياسهاي اندازه‌گيري
متغير
نمونه
جمعيت

آمار توصيفي:

خصوصيت مورد مطالعه، از فردی به فرد ديگر، يا از شي به شي ديگر در جمعيت آماری تغيير مي‌كند، كه آن را اصطلاحا متغير مي‌ناميم.

معمولا دو نوع متغير در آمار مورد نظر هستند:

- متغيرهاي گروهي، نظير رنگ، نژاد، شغل و گروه خوني كه شامل چند گروه يا طبقه مي‌باشند.
- متغيرهاي عددي كه ممكن است نتيجه شمارش باشد، مانند تعداد احشام هر خانوار در يك روستا،تعداد حوادث در يك كارخانه در روزهاي مختلف، و يا نتيجه اندازه‌گيري باشد، مثل قد دانشجويان بيست ساله در يك شهر، حجم شربت مولتي ويتامين با استاندارد خاص.

آمار توصيفي:

متغير:

- متغيرهاي گسسته
  - متغيرهاي گروهي
  - متغيرهاي عددي كه از راه شمارش به‌دست آمده اند
- متغيرهاي پيوسته
  - متغيرهايي را كه از طريق اندازه‌گيري به دست آمده باشند

آمار توصيفي:

در بسيار از مسائل پيش‌رو، اندازه‌گيري ويژگی يك متغير مستلزم آگاهی و شناخت خاصي است. به طور كلي چهار نوع مقياس براي اندازه گيري وجود دارد:

- مقياس اسمي
- مقياس ترتيبي
- مقياس فاصله‌اي
- مقياس نسبتي

آمار توصيفي:

مقياس اسمي:

اين نوع مقياس اندازه‌گيري عمدتا براي طبقه بندي داده‌ها به كار مي‌رود و منظور از آن اتلاق يك عدد طبيعي به داده‌هاي متفاوت است.

مثال: اختصاص اعداد ۱ تا ۴ به گروه‌هاي خوني A,B, AB, O.

توجه داشته باشيد كه:

اين اعداد را نمي‌توان براي مقايسه يا چهار عمل اصلي به كار برد

آمار توصيفي:

مقياس ترتيبي:

اين نوع مقياس اندازه‌گيري عموما براي طبقه بندي داده‌ها به منظور يك نوع برتري به كار مي‌رود.

مثال:

در يك كارخانه ممكن است كارگران را به سه دسته ساده، نيمه ماهر و ماهر تقسيم بندي كنيم. اتلاق به ترتيب اعداد ۱ تا ۳ به اين سه دسته يك مقياس ترتيبي است.

توجه داشته باشيد كه:

اين اعداد تنها براي مقايسه به كار مي‌روند و نمي‌توان با آنها چهار عمل اصلي را انجام داد.

آمار توصيفي:

مقياس فاصله اي:

اين نوع مقياس اندارزه‌گيري عموما در زمينه‌هاي كه علاوه بر حفظ ترتيب به نحوي فاصله بين ويژگي‌ها را نيز حفظ مي‌كند. به عبارت ديگر در چنين مقياسي نسبت تفاضلها ثابت مي‌ماند.

مثال:

اندازه‌گيري ضريب هوشي دانش آموزان كلاس اول دبستان در شهر اصفهان.

توجه داشته باشيد كه:

در اين نوع مقياس، عدد صفر يك مفهوم قراردادي است.

آمار توصيفي:

مقياس نسبتي:

اين نوع مقياس اندازه‌گيري علاوه بر حفظ فاصله، نسبت را نيز حفظ مي‌كند. به عبارت ديگر در اين نوع اندازه‌گيري نسبت دو مقدار بستگي به واحد اندازه‌گيري ندارد.

آمار توصيفي:

اطلاعاتي كه از مطالعه يك متغير به دست مي‌آيند، معمولا شامل انبوهي عدد يا علامت مي‌باشند كه آنها را داده مي‌ناميم. داده‌ها را نسبت به نوع متغيري كه اندازه گيري مي‌كنيم به دو دسته داده گسسته و داده‌هاي پيوسته تقسيم مي‌كنيم.

معمولا به داده‌هاي جمع آوري شده كه انبوهي عدد است و هيچ نوع پردازشی روي آنها انجام نشده است داده خام مي‌گويند.

داده خام

آمار توصيفي:

مواردي كه در ارتباط با يك مجموعه از داده‌هاي مي‌بايستي مد نظر قرار داد،عبارت‌اند از:

- خلاصه كردن و توضيح داده‌ها به وسيله تنظيم جداول و رسم نمودارها.
- محاسبه مقادير عددي، براي دست يافتن به معيارهايي كه تمركز و يا پراكندگي داده‌ها را نشان دهد.

در آمار،براي اينكه از داده‌هاي خام واقعيتهاي موجود را استخراج كنيم،آنها را به نحوي مناسب دسته‌بندي كرده و جدولهايي به نام جدولهاي آماري تهيه مي‌نماييم.
متداولترين جدول در آمار، جدول فراواني است.

پيش از آنكه نحوه تنظيم جدول فراواني را بيان نماييم،اطلاع از اصطلاحات زير ضروري است.

جدول‌هاي آماري:

فراواني نسبي تجمعي
فراواني تجمعي
فراواني نسبي
فراواني

هرگاه n داده $y_1, y_2, \ldots, y_n$ از k نوع $x_1, x_2, \ldots, x_k$ ، با فرض $2 \le k \le n$ ، به ترتيب با تعدادهاي $f_1, f_2, \ldots, f_k$ تشكيل شده باشند،آنگاه f_i را فراواني x_i مي‌گوييم. به عبارت ديگر تعداد دفعاتي را كه x_i در داده‌هاي $y_1, y_2, \ldots, y_n$ تكرار مي‌شود، فراواني x_i مي‌ناميم و آن را با نماد f_i نمايش مي‌دهيم.

به خاطر داشته باشيد كه

اگر اندازه نمونه برابر n باشد، آنگاه براي $j = 1, \ldots, k$

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

جدول‌هاي آماري:

فراواني نسبي تجمعي
فراواني تجمعي
فراواني نسبي
فراواني

مثال:

داده‌هاي زير ميزان تصادف منجر به مرگ رد ۳۰ منطقه را نشان مي‌دهد. فراواني دادها را تعيين نماييد.

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 مي‌شود كه داده‌هاي تكرار اعداد ۲،۳،۴،۵،۶،۷،۸ مي‌باشند،بنابراين جدول زير را براي فراواني داده‌ها خواهيم داشت:

x_i	f_i
۲	۱
۳	۴
۴	۳
۵	۸
۶	۷
۷	۳
۸	۴

19
21
23
31

فراواني نسبي تجمعي
فراواني تجمعي
فراواني نسبي
فراواني

جدول‌هاي آماري:

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 $j = 1, \ldots, k$

$$0 \le r_i \le 1$$

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

جدول‌هاي آماري:

فراواني نسبي تجمعي

فراواني تجمعي

فراواني نسبي

فراواني

$1/30 = 0.033$

$4/30 = 0.133$

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

فراوانی نسبی تجمعی
فراوانی تجمعی
فراوانی نسبی
فراوانی

جدول‌هاي آماري:

با توجه به تعريف فراواني، فراواني تجمعي رديف i را با نماد F_i نمايش مي‌دهيم و به صورت زير تعريف مي‌كنيم:

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

به خاطر داشته باشيد كه

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

$$f_{i=} F_1 \le F_1 \le \ldots \le F_k = n$$

فراوانی نسبی تجمعی
فراوانی تجمعی
فراوانی نسبی
فراوانی

جدول‌هاي آماري:

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

$1+4=5$

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

فراواني نسبي تجمعي
فراواني تجمعي
فراواني نسبي
فراواني

جدول‌هاي آماري:

با توجه به تعريف فراواني نسبي، فراواني نسبي تجمعي رديف iرا با نماد R_i نماد نمايش مي‌دهيم و به صورت زير تعريف مي‌كنيم:

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

به خاطر داشته باشيد كه

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

$$r_i = R_1 \le R_2 \le \ldots \le R_k = 1$$

فراواني نسبي تجمعي
فراواني تجمعي
فراواني نسبي
فراواني

جدول‌هاي آماري:

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

$1/30 + 4/30 + 3/30$

فراوانی نسبی تجمعی
فراوانی تجمعی
فراوانی نسبی
فراوانی

جدول‌هاي آماري:

مثال:

معدل ۵۰ دانشجوي دانشگاه با تقريب تا يك رقم اعشار، به شرح زير است:

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

تشکیل جدول فراوانی برای داده های پیوسته

چون داده‌ها تا يك رقم اعشار گرد شده‌اند، بنابراين مي‌توان گفت كه اندازه واقعي معدلها در فاصله [1/35,2/95]

$$U = MAX + 0.5\times10^{-r}$$

$$L = MIN - 0.5\times10^{-r}$$

تعداد ارقام گرد شده

جدول‌هاي آماري:

فراواني نسبي تجمعي | فراوانی تجمعی | فراوانی نسبی | فراوانی

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

تعداد طبقات

کلاس	x_i	f_i	r_i	F_i	R_i
۳۵/۱_۵۵/۱	۴۵/۱	۴	۰۸/۰	۴	۰۸/۰
۵۵/۱_۷۵/۱	۶۵/۱	۶	۱۲/۰	۱۰	۲۰/۰
۷۵/۱_۹۵/۱	۸۵/۱	۱۲	۲۴/۰	۲۲	۴۴/۰
۹۵/۱_۱۵/۲	۰۵/۲	۹	۱۸/۰	۳۱	۶۲/۰
۱۵/۲_۳۵/۲	۲۵/۲	۸	۱۶/۰	۳۹	۷۸/۰
۳۵/۲_۵۵/۲	۲/۴۵	۶	۱۲/۰	۴۵	۹۰/۰
۵۵/۲_۷۵/۲	۶۵/۲	۲	۰۴/۰	۴۷	۹۴/۰
۷۵/۲_۹۵/۲	۸۵/۲	۳	۰۶/۰	۵۰	۰۰/۱
جمع		۵۰	۰۰/۱	_	_

27

29

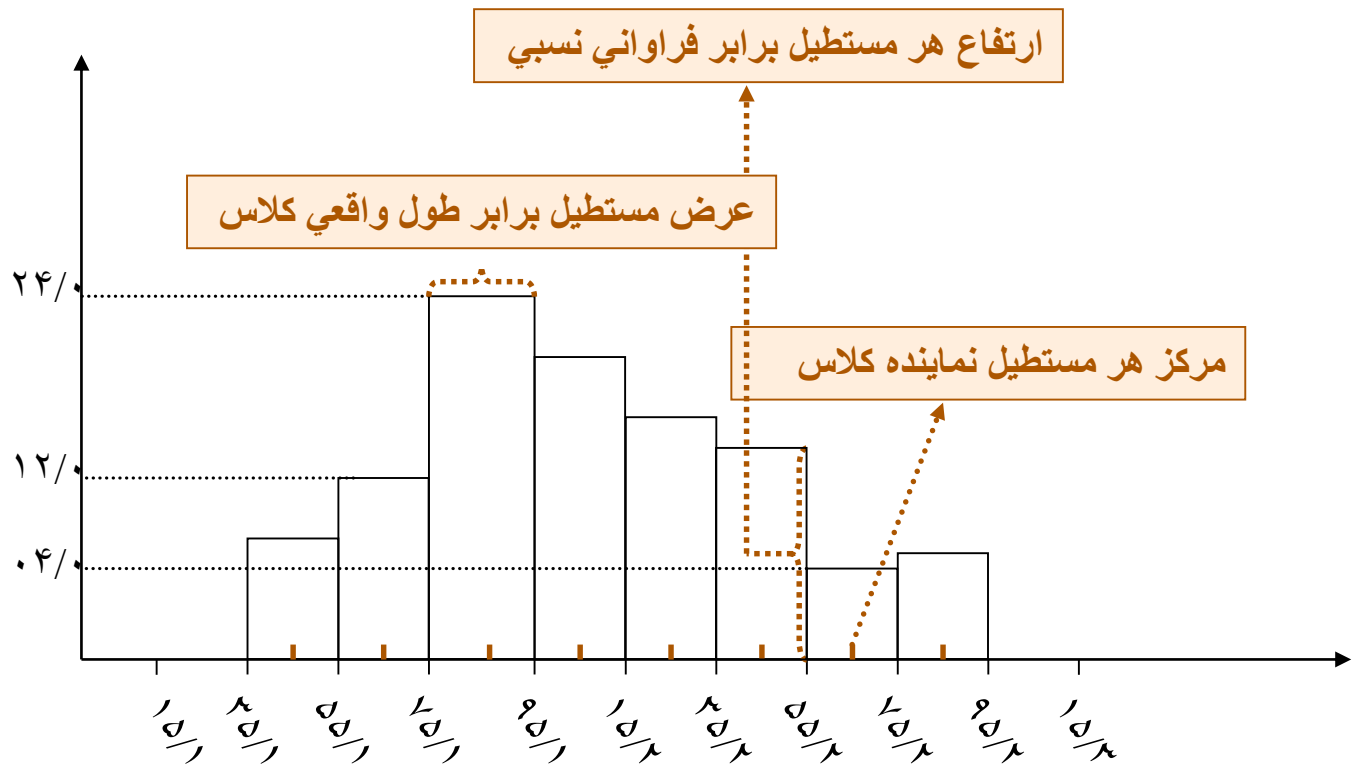
نمودارهاي آماري:

معمولا داده‌ها را با نمودارهاي مختلف نمايش مي‌دهند. عموما اين نمودارها در ارتباط با داده‌هاي پيو.سته به كار گرفته مي شود و منظور از نمايش آنها، تجسم عيني اطلاعات نهفته در داده‌ها است.
در اين بخش به معرفي چند نمودار معروف اكتفا مي‌كنيم:

- هيستوگرام
- چندبر فراواني
- چندبر فراواني تجمعي
- منحنيهاي فراواني و فراواني تجمعي
- نمايش نمودار تنه و شاخه
- نمودار جعبه‌اي

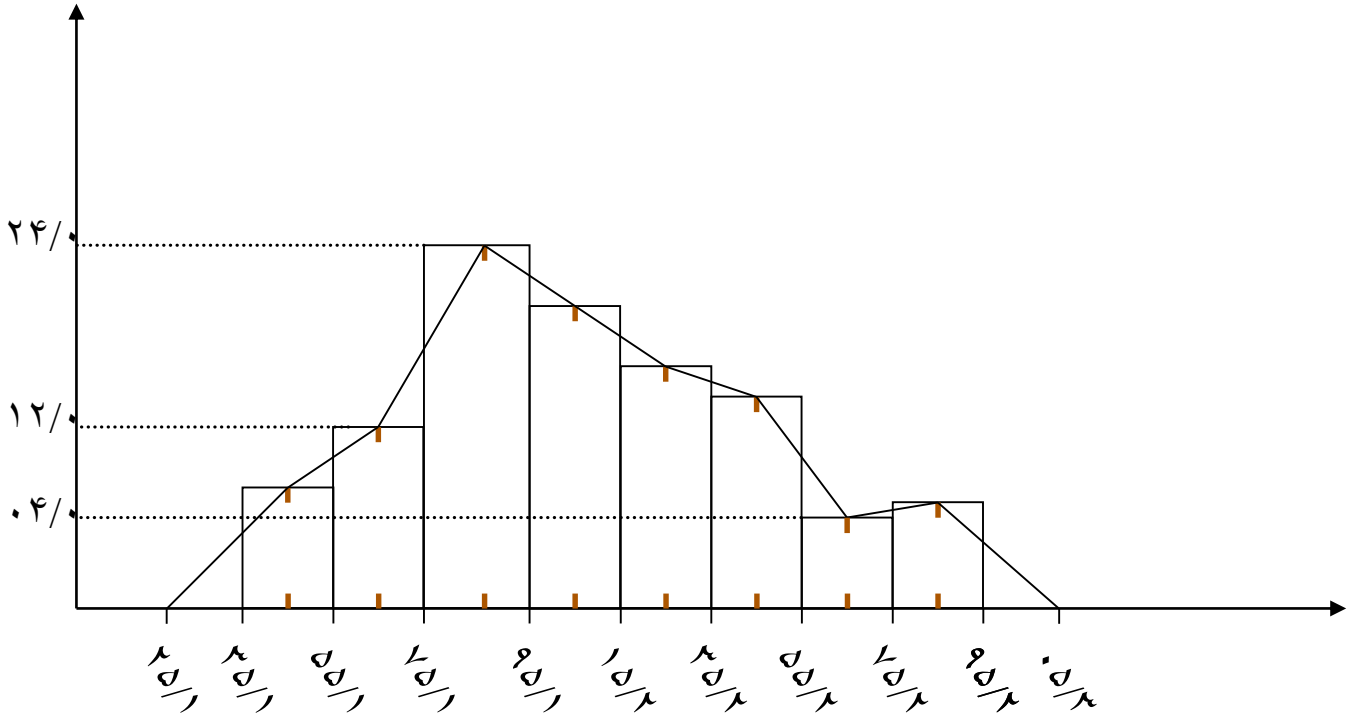
نمودارهاي آماري:

- نمايش نمودار تنه و شاخه
- منحني‌هاي فراواني و.....
- چندبر فراواني تجمعي
- چندبر فراواني
- هيستوگرام



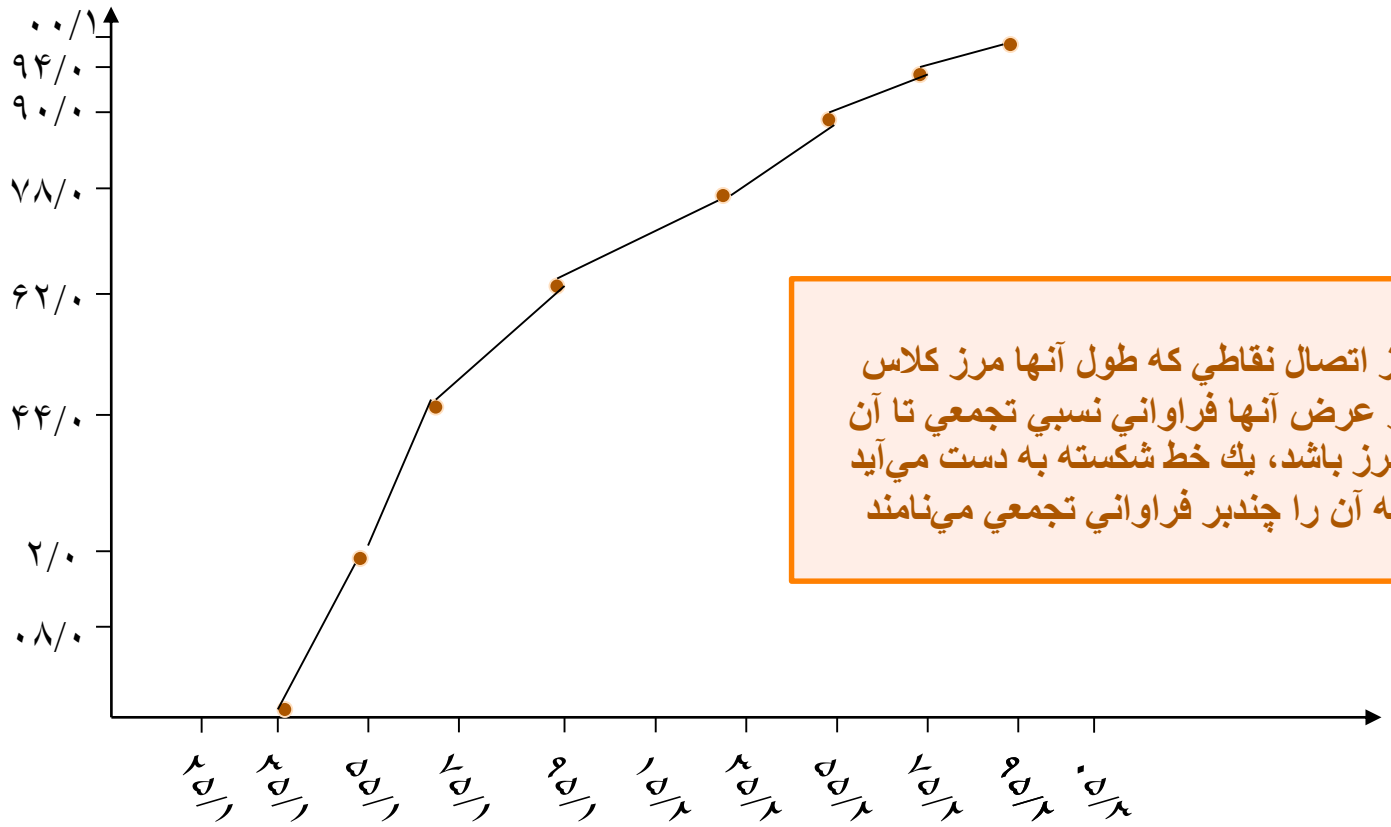
نمودارهاي آماري:

- نمايش نمودار تنه و شاخه
- منحني‌هاي فراواني و.....
- چندبر فراواني تجمعي
- چندبر فراواني
- هيستوگرام



نمودارهاي آماري:

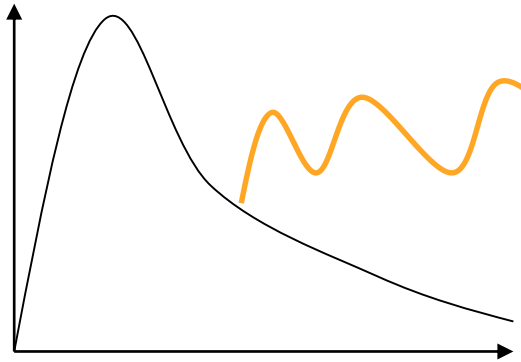
نمايش نمودار تنه و شاخه
منحني‌هاي فراواني و.....
چندبر فراواني تجمعي
چندبر فراواني
هيستوگرام



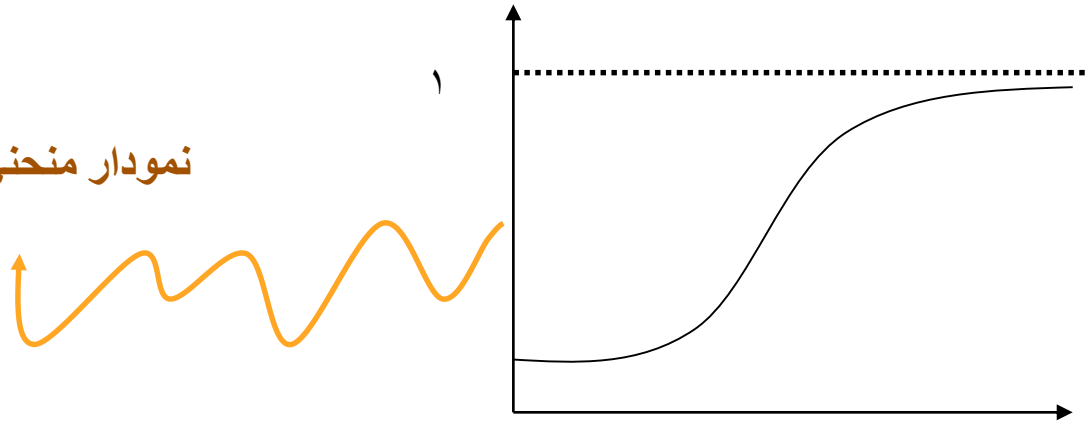
از اتصال نقاطي كه طول آنها مرز كلاس و عرض آنها فراواني نسبي تجمعي تا آن مرز باشد، يك خط شكسته به دست مي‌آيد كه آن را چندبر فراواني تجمعي مي‌نامند

نمودارهاي آماري:

نمايش نمودار تنه و شاخه
منحني‌هاي فراواني و.....
چندبر فراواني تجمعي
چندبر فراواني
هيستوگرام



نمودار منحني فراواني



نمودار منحني فراواني تجمعي

نمودارهاي آماري:

- نمايش نمودار تنه و شاخه
- منحني‌هاي فراواني و.....
- چندبر فراواني تجمعي
- چندبر فراواني
- هيستوگرام

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

فراواني

۴/۱	***
۵/۱	*
۶/۱	***
۷/۱	***
۸/۱	******
۹/۱	******
۰/۲	*****
½	****
۲/۲	*****
۳/۲	***
۴/۲	***
۵/۲	***
۶/۲	**
۷/۲	
۸/۲	
۹/۲	***

نمودارهاي آماري:

نمايش نمودار تنه و شاخه

منحني‌هاي فراواني و.....

چندبر فراواني تجمعي

چندبر فراواني

هيستوگرام

نمرات ۸۰ دانشجو در امتحانات نهايي درس احتمال و آمار به شرح زير 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

تنه

شاخه

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

نمودارهاي آماري:

- نمايش نمودار تنه و شاخه
- منحني‌هاي فراواني و.....
- چندبر فراواني تجمعي
- چندبر فراواني
- هيستوگرام

پس از ساختن نمودار اوليه معمولا بهتر است مقادير هر شاخه را از كوچك به بزرگ، با تعداد دفعات تكرار،مرتب كرد، به صورت زير:

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

ميانگين | ميانه | چندكها | نما

معيارهاي مركزي:

با استفاده از جدول فراواني و رسم نمودارها مي‌توانيم داده‌ها را به نحو مطلوبي **تنظيم كرده و اطلاعات نهفته را تا حدودي مشخص كنيم**. با اين حال براي ارايه يك گزارش مناسب،بهتر است آنها را در يك يا چند عدد مناسب نيز خلاصه كنيم. چنين عددي مي‌تواند معيار مركزي باشد. مهمترين معيارهاي مركزي ميانگين، ميانه و نما است كه در بخش اين به شرح هر يك از آنها خواهيم پرداخت.

هرگاه n داده $y_1, y_2, \ldots, y_n$ از k نوع $x_1, x_2, \ldots, x_k$ ، با فرض $2 \le k \le n$ ، به ترتيب با تعدادهاي $f_1, f_2, \ldots, f_k$ تشكيل شده باشند، آنگاه f_i را فراواني x_i مي‌گوييم.

ميانگين حسابي	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i$
ميانگين وزني	$\bar{x}_w = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_k}$
ميانگين هندسي	$G = \left(\prod_{i=1}^{k} x_i^{f_i}\right)^{\frac{1}{n}}$ كليه داده‌ها بزرگتر از صفر باشند

معيارهاي مركزي:

اگر داده‌ها را از كوچك به بزرگ مرتب نماييم،عدد m را ميانه اين داده‌ها مي‌ناميم،اگر نصف داده‌ها در سمت چپ و نصف داده در سمت راست اين عدد قرار گيرد

محاسبه ميانه براي داده‌هاي گسسته

فرض كنيد $y_1, y_2, \ldots, y_n$ داده‌هاي ما باشند و شكل مرتب شده آنها را با $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ نمايش دهيم آنگاه

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

معيارهاي مركزي:

محاسبه ميانه براي داده‌هاي پيوسته

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_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	_

$$n/2 = 25$$

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

طول هر رده

معيارهاي مركزي:

چندك يك معيار كلي‌تر از ميانه است و درعنوان حالت خاص ميانه را نيز در بر مي‌گيرد. اگر pيك عدد حقيقي بين صفر و يك باشد،آنگاه عدد Q_p را چندك مرتبه p مي‌ناميم هر گاه p ١٠٠%

داده‌ها سمت چپ و (١- p) ١٠٠% داده‌ها سمت راست Q_p باشند.

چندكهاي معروف عبارتند از :

چاركها

چاركها به ازاي ٧٥/٠ ، ٥/٠ ، ٢٥/٠ =p به دست مي‌آيند و آنها را به ترتيب با نماد Q_1 (چارك اول)، Q_2 (چارك دوم) و Q_3 (چارك سوم)نشان مي‌دهند.

دهكها

دهكها به ازاي ٩/٠، .....،٢/٠ ،١/٠=p به دست مي‌آيند و آنها را به ترتيب با نماد D_1 (دهك اول)، D_2 (دهك دوم)،..... و D_9 (دهك نهم) نشان مي‌دهند.

صدكها

صدكها به ازاي ٩٩/٠،.....٠٢/٠ ،٠١/٠=p به دست مي‌آيند و آنها را به ترتيب با نماد P_1 (صدك اول)، P_2 (صدك دوم)،.....وP_{99} (صدك نود و نهم) نشان مي‌دهند.

معيارهاي مركزي:

محاسبه چندك براي داده‌هاي گسسته

فرض كنيد $y_1, y_2, \ldots, y_n$ داده‌هاي ما باشند و شكل مرتب شده آنها را با $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ نمايش دهيم. براي محاسبه چندك

$(n+1)p$ صحيح باشد $\rightarrow$ $r = (n+1)p, \qquad Q_p = y_{(r)}$

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

نمودارهاي آماري:

محاسبه چندك براي داده‌هاي پيوسته

با توجه به ستون فراواني تجمعي در جدول فراواني، كلاسي را كه چندك در آن قرار دارد مشهص 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	_

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

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

نمودارهاي آماري:

محاسبه نما براي داده‌هاي گسسته

داده‌اي كه فراواني آن نسبت به ديگر داده‌ها بيشتر باشد، نما يا مد ناميده مي‌شود و آن را با نماد M نمايش مي‌دهيم. براي به دست آوردن نما، نخست فراواني داده‌ها را پيدا مي‌كنيم و داده‌اي را كه فراواني آن بيشتر باشد، به عنوان نما اختيار مي‌كنيم و اگر دو داده، داراي فراواني يكسان و بيش از ديگر فراواني‌ها باشند، هر دو را به عنوان نما اختيار مي‌كنيم و داده‌ها را دو نمايي مي‌گوييم، به شرط آن كه اين دو داده در يك صف غير نزولي،كنار هم نباشند. در صورتي كه اين دو داده در يك صف غير نزولي،كنار هم باشند نصف مجموع آنها را به عنوان نما اختيار مي‌كنيم. اگر تمام داده داراي فراواني يكسان باشند،مي‌گوييم داده‌ها بدون نما هستند. به ياد داشته باشيد كه نما، به عنوان يك معيار تمركز در داده‌هاي گروهي به كار گرفته مي‌شود.

نمودارهاي آماري:

مثال: براي داده‌هاي ۲، ۲، ۵، ۷، ۹، ۹، ۹، ۱۰، ۱۰، ۱۱، ۱۲و ۱۸ نما برابر M=۹ است، زيرا فراواني داده ۹ بيش از فراواني ديگر داده‌ها است.

مثال: براي داده‌ها ۲، ۳، ۴، ۴، ۴، ۵، ۵، ۷، ۷، ۷و ۹، دو داده ۴ و ۷ به عنوان نما اختيار مي‌شوند، زيرا فراواني اين دو داده، بيش از فراواني داده‌هاي ديگر است.

مثال: براي داده‌هاي ۳، ۵، ۸، ۱۰، ۱۲، ۱۵و ۱۶، نما وجود ندارد، زيرا تمام داده‌ها داراي فراواني يكسان هستند.

مثال: براي داده‌ها ۲، ۳، ۴، ۴، ۴، ۵، ۵، ۵، ۷، ۷و ۹دو داده ۴و ۵ را كه داراي بيشترين فراواني هستند به عنوان نما بر مي‌گزينيم،اما از آنجا كه اين دو داده در يك صف غير نزولي در كنار يكديگر قرار دادند،نصف مجموف دو داده به عنوان نما اختيار مي‌شود، يعني M=۵/۴.

نمودارهاي آماري:

محاسبه تما براي داده‌هاي پيوسته

از روي جدول ملاحظه مي‌شود كه فراواني رده ۹۵/۱_۷۵/۱ داراي بيشترين فراواني است بنابراين به عنوان رده نما در نظر مي‌گيريم.

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

۲۴/۰-۱۲/۰

۲۴/۰-۱۸/۰

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

کلاس	x_i	f_i	r_i
۳۵/۱_۵۵/۱	۴۵/۱	۴	۰۸/۰
۵۵/۱_۷۵/۱	۶۵/۱	۶	۱۲/۰
۷۵/۱_۹۵/۱	۸۵/۱	۱۲	۲۴/۰
۹۵/۱_۱۵/۲	۰۵/۲	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	۰۴/۰
۷۵/۲_۹۵/۲	۸۵/۲	۳	۰۶/۰
جمع		۵۰	۰۰/۱

معيارهاي پراكندگي:

با وجود اين كه در بسياری از موارد، ميانگين توصيف نسبتا كاملي از مجموعه داده‌ها ارائه مي‌دهد، اما گاهي وجود اطلاعات بيشتر در مورد داده‌ها ضروری است. يك مفهوم مهم در ارتباط با داده‌هاي آماری، ميزان تغييرات آنهاست،بدين معني كه اندازه‌گيريها تا چه اندازه از فردي به فرد ديگر يا شيي به شيي ديگر تغيير مي‌كنند. در اين بخش، به بررسي و محاسبه ميزان تغيرات به عنوان معيارهای پراكندگی خواهيم پرداخت. مهمترين معيارهای پراكندگي عبارتند از **دامنه، ميانگين انحراف ها از ميانگين يا از ميانه، ميان دامنه چاركها، دامنه صدكي، واريانس و انحراف معيار است. علاوه بر مطالب فوق، در اين بخش داده‌هاي استاندارد و ضريب تغيرات را نيز معرفی خواهيم كرد.**

معيارهاي پراكندگي:

اگرچه دامنه يك وسيله ساده براي اندازه‌گيري اختلاف و پراكندگي در يك سري از داده‌ها است، اما در بيشتر موارد رضايتبخش نيست. داده‌هاي بسيار بزرگ يا بسيار كوچك مانع از آنند كه دامنه، معرف واقعي ميزان انحراف باشد. در چنين مورد قبول همگان به

با توجه به اينكه در محاسبه واريانس داده‌ها را مربع مي‌كنيم، بدين جهت ريشه دوم مثبت آن را كه انحراف معيار يا انحراف استاندارد مي‌ناميم، به عنوان يك معيار پراكندگي بر مبناي مقياس مي‌بريم.

ضريب تغيير عبارت است از اندازه نسبي انحراف معيار در مقايسه با ميانگين. ضريب همبستگي به واحد اندازه‌گيري وابسته نيست و براي مقايسه جمعيتهاي يكسان به كار مي‌رود. در مقايسه هر اندازه كه ضريب تغيير ويژگي جمعيتي كمتر باشد،ويژگي آن جمعيت بهتر ارزيابي مي‌شود.

دامنه	$y_{(1)}$ كوچكترين و $y_{(n)}$
ميانگين انحرافها از ميانگين	
ميانگين انحرافها از ميانه	$MD_m = \sum_{i=1}^{k} f_i \lvert x_i$
	$IQR = Q_3 - Q_1$
	$IPR = P_{90} - P_{10}$
	$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2$
انحراف معيار	$s = \sqrt{\frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}$
ضريب تغييرات	$v = \frac{s}{\bar{x}}$

معيارهاي پراكندگي:

اگر $y_1, y_2, \ldots, y_n$ نمايانگر داده‌هاي خام باشند، براساس جدول فراواني f_1 تا از y_i ها برابر x_1 ، f_2 تا برابر x_2 ،.....و f_k تا برابر x_k است. مي‌دانيم كه s و $\bar{x}$ به ترتيب ميانگين و انحراف معيار داده است. اگر از هر داده $\bar{x}$ را كم و بر s تقسيم كنيم ، يعني

داده‌هاي استاندارد

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

آنگاه $z_1, z_2, \ldots, z_k$ با فراواني‌هاي به ترتيب $f_1, f_2, \ldots, f_k$ را داده‌هاي استاندارد مي‌ناميم. به سادگي مي‌توان نشان داد كه داده‌هاي استاندارد داراي ميانگين برابر با صفر و واريانس برابر با يك ۱ هستند و به واحد اندازه‌گيري بستگي ندارند.

معيارهاي پراكندگي:

چون $v_1 > v_2$ ، بنابراين نتيجه مي‌گيريم كه دانشجويان در امتحان دوم نمرات مطلوبتري را كسب كرده‌اند.

ب) براي مقاسه، ابتدا نمرات دانشجو را استاندارد مي‌كنيم

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

چون $z_1 < z_2$، بنابراين نمره آزمون دوم دانشجو د رمقايسه از موقعيت بهتري برخوردار است.

معیارهای پراکندگی

- فرض کنید يك دسته از دانشجويان در دو امتحان شركت كرده‌اند و خلاصه نتاسج آزمونها به شرح زير است.
- آزمون اول: ميانگين نمرات برابر ۶۰، انحراف معيار برابر ۶، ماكزيمم نمره از ۱۰۰
- آنمون د.م: ميانگين نمرات ۷۰۰، انحراف معيار برابر ۷، ماكزيمم نمره از ۱۰۰۰
- الف- چگونه اين دو نتيجه را با هم مقيسه و ارزيابي مي‌كنيد؟
- ب- اگر دانشجويي در آزمون اول نمره ۶۵ و در آزمون دوم نمره ۷۲۰ را كسب كرده باشد، وضعيت دانشجو در كدام آزمون مطلوبتر است؟
- حل: الف) با محاسبه ضريب تغيير دو آزمون معلوم مي‌شود كه

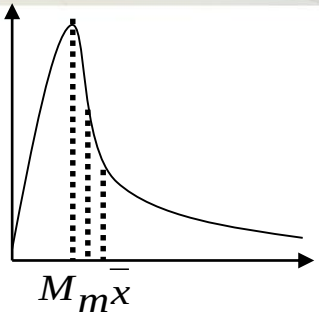
$$v_1 = \frac{s_1}{\bar{x}_1} = \frac{6}{60} = \frac{1}{10} = \%10$$

$$v_2 = \frac{s_1}{\bar{x}_2} = \frac{7}{700} = \frac{1}{100} = \%1$$

منحنيهاي فراواني:

منحنيهاي فراواني در طبيعت تنوع زيادي دارند، اما بسياري از منحنيهاي فراواني تك نمايي يا متقارن هستند يا چوله و يا برجسته و يا پخ. ايده آلترين منحني فراواني متقارن، منحني فراواني نرمال استاندارد است. براي منحنيهاي فراواني كاملا متقارن تك نمايي مقادير ميانگين،ميانه و نما بر هم منطبق مي‌شوند. در طبيعت،عموما منحني فراواني متقارن ايده آل كمتر يافت مي‌شود و بسياري از منحنيهاي فراواني موجود در طبيعت نامتقارن برجسته يا پخ هستند. ميزان انحراف از تقارن ايده آل را معمولا با دو معيار چولگي و برجستگي مي‌سنجند.

منحنيهاي فراواني:

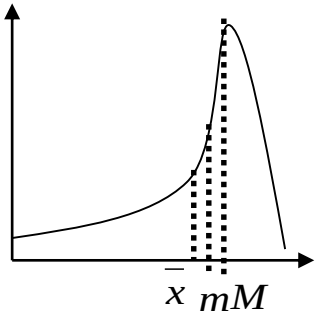


$M < m < \bar{x}$

به راست یا مثبت

چولگي:

به چپ یا منفي



$M > m > \bar{x}$

معیار اندازه گیری چولگی:

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

منحنيهاي فراواني:

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.

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

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

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

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

نمودار جعبه ای:

همان گونه که گفته شد، روشهای نموداری و خلاصه کردن داده ها به صورت مقادیر عددی موضوعی اساسی در تجزیه و تحلیلهای آماری است. پیش از این دیدیم که چگونه نمایش نمودار تنه و شاخه را می توان به عنوان ابزاری ساده و مهم در نمایش و استنباط از داده ها به کار گیریم که چنین نموداری بسیار همانند نمودار هستوگرام بود.

نموار با ارزش دیگری که برخی از امتیازهای دیگر را در مقایسه با نمودار تنه و شاخه دارد نمودار جعبه ای است که برخی از امتیازهای دیگر را در مقایسه با نمودار تنه وشاخه دارد .

نمایش نمودار جعبه ای بر پایه داده های مرتب شده از کوچک به بزرگ و تعیین میانه، چارک اول و چارک سوم است.

نمودار جعبه ای:

گام اول

نمایش نمودار تنه و شاخه،

۹	۰ ۲
۱۰	۰ ۰ ۹
۱۱	۰ ۵ ۶ ۶ ۷ ۷ ۸ ۹
۱۲	۵ ۹ ۹ ۹ ۹ ۹
۱۳	۰ ۰
۱۴	۰
۱۵	۰ ۴

گام دوم

تعیین مکان میان، چارک اول و چارک سوم،

مکان میانه $= \frac{n+1}{2} = \frac{24+1}{2} = 12.5$

با توجه به مقدار به دست امده میانگین داده های دوازدهم و سیزدهم را به عنوان میانه در نظر می گیریم، یعنی

$$m = \frac{118+119}{2} = 118.5$$

نمودار جعبه ای:

برای تعیین مکان چارک اول و سوم به صورت زیر عمل می کنیم

مکان چارکها $= \frac{\lceil \text{مکان میانه} \rceil + 1}{2} = \frac{\lceil 12.5 \rceil + 1}{2} = \frac{12+1}{2} 6.5$

با توجه به مقدار به دست آمده، میانگین داده های ششم و هفتم از پایین به بالا را به ترتیب به عنوان چارک اول و چارک سوم در نظر می گیریم، یعنی

چارک اول $Q_1 = \frac{110+115}{2} = 112.5$

چارک سوم $Q_2 = \frac{125+129}{2} = 127$

نکته

در صورتی که مقادیر به دست آمده در مکانها اعداد صحیح باشند، داده همان مرتبه به عنوان میانه، چارک اول و چارک سوم در نظر گرفته می شود.

میانه و چارکها به دست آمده در ارایه نمایشی برای نمودار جعبه ای با روش بیان شده در بخش های قبل فرق دارد. در حقیقت معیارهای به دست آ»ده از این روش را هینج می نامند که کمی با معیارهای گفته شده متفاوت است.

نمودار جعبه ای:

گام سوم

تعیین دو فاصله به عنوان حصارهای درونی و بیرونی است،

نخست دامنه چارکها را محاسبه می کنیم ،

$$IQR = Q_3 - Q_1 = 125 - 112.5 = 12.5$$

کرانهای حصار درونی را به صورت زیر تعریف می کنیم:

کران پایین حصار درونی $= LIF = Q_1 - 1.5IQR$

کران بالای حصار درونی $= UIF = Q_3 + 1.5IQR$

بنابراین حصار درونی در این مثال به صورت زیر تعریف می شود

$$(LIF, UIF) = (93.75, 145.75)$$

نمودار جعبه ای:

کرانهای حصار بیرونی را به صورت زیر تعریف می کنیم:

$$LOF = Q_1 - 3IQR$$
$$UOF = Q_3 + 3IQR$$

در نتیجه فاصله زیر، حصار بیرونی در این مثال است،

$$(LOF, UOF) = (75, 164.5)$$

گام چهارم **تعیین مقادیری از داده ها که در همسایگی کرانهای حصار درونی است.**

در حقیقت مقادیر این داده ها در حصار درونی قرار دارد و کمینه بیشینه مقدار ممکن از داده ها در حصار درونی است که نزدیک به کران بالا و پایین حصار درونی است.
همسایگی کران پایین را با نماد LA و همسایگی کران بالا را با نماد UA نمایش می دهیم. بنابراین در این مثال،

$$LA = 100$$
$$UA = 140$$

نمودار جعبه ای:

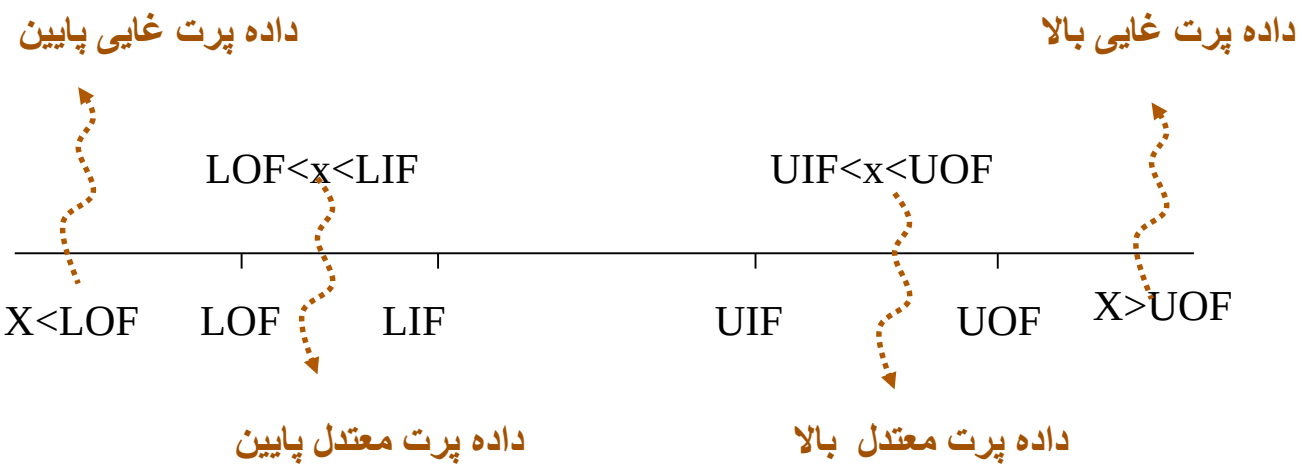
گام پنجم

تعیین داده های پرت،

هر داده بیرون از حصار درونی را داده پرت می نامیم. در صورتی که این داده ها بیرون از حصار بیرونی نباشد، آن را داده پرت معتدل و به جز این صورت آن را داده پرت غایی می نامیم.

نمادی که برای نمایش داده های پرت در نمودار جعبه ای به کار خواهیم برد، دایره توخالی برای داد پرت معتدل و دایره توپر برای داده پرت غایی.

در نتیجه



نمودار جعبه ای:

بنابراین در این مثال داده های پرت عبارتند از: ۱۵۴، ۱۵۰، ۹۲و ۹۰ و داده پرت غایی نداریم.