به نام خدا

دانشگاه فرهنگیان

واحد تربیت معلم نیشابور

Ph.D. Student Vahid Hajimirza

Email: vhc1359@gmail.com

Mobile: 091515170119

http://www.htcnet.ir

به نام خدا

شیوه های تحلیل کمی و کیفی داده ها با کاربرد نرم افزارها

بارم بندی نمرات:

۱- مقاله کنفرانسی ۵ نمره

۲- سوالات و تمرین کلاسی ۲ نمره

۳- میان ترم و پایان ترم ۸ نمره

۴- پروژه spss به همراه مکتوبات ۵ نمره

نمره اضافی :

۱-مقاله علمی ترویجی یا علمی پژوهشی ۱۰ نمره

۲-مقاله اضافی کنفرانسی ۵ نمره

۳-ارفاق به کل دانشجویان ۱ نمره

۴- ترجمه بخشی از کتاب ۳ نمره

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 می‌رسندو در طول اخبار، به جریانهای بازار بورس و سهام اشاره می‌شود و روزنامه‌ها خبر از افزایش نرخ اجناس می‌دهندو...

آمار به عنوان پایه یک 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، معمولا سعی می‌شود که اطلاعات موجود در یک 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 یک جمله آمار مجموعه‌ای از روشهای جمع آوری، تهیه وتنظیم و تجزیه و تحلیل اطلاعات است که برای کسب یک یا چند نتیجه به خدمت گرفته می‌شود.**

فهرست مطالب:

آمار توصیفی

جدول‌های آماری

نمودارهای آماری

معیارهای مرکزی

معیارهای پراکندگی

منحنی‌های فراوانی

آشنایی با آمار توصیفی

Ph.D. Student Vahid Hajimirza

مروری بر مفاهیم اصلی و اهم کاربردهای آمار

Ph.D. Student Vahid Hajimirza

مقدمه

آمار چیست و چگونه به ما کمک می کند؟

تعریف آمار

آمار مجموعه ایی از روشها را برای جمع آوری، و خلاصه کردن داده ها، طبقه بندی آنها و روشهای تحلیلی برای پیش بینی، برآورد و تصمیم گیری در شرایط مختلف ارائه می دهد.

تعریف

داده :

داده ها واقعیتها یا ارقامی هستند که می توان از آنها نتایجی را بیرون کشید. در واقع داده هامواد خام آماری هستند.

مجموعه داده :

داده هایی را که برای مطالعه ایی خاص گردآوری شده باشند مجموعه داده ها می نامیم.

عنصر:

هر عنصر اطلاعات یک یا چند مشخصه مجموعه داده را در بر می گیرد(مثل مشخصه های یک دانش آموز)

تعریف

متغییر :

متغیر مشخصه مربوط به یک عنصر است که می تواند برآمدهای مختلف را قبول کند. وقتی برآمدها مستقیما به صورت عددی بیان شود متغییر را کمی در غیر این صورت کیفی است.(مثل وزن و جنسیت)

مورد:

اطلاعات مربوط به تمام متغییرها برای یک عنصر از مجموعه داده ها را یک مورد می نامیم. مثل اطلاعات مربوط به پنج متغییربرای یک دانش آموز

مشاهده:

داده مربوط به یک عنصر از مجموعه داده ها در باره یک متغییر را یک مشاهده یا یک برامد می نامند. مثلا عدد ۱۲که سن دانش آموز را نشان می دهد یک مشاهده از متغییر سن برای این دانش آموز است

انواع داده :

۱- داده های اندازه گیری شده: مثل وزن

۲- داده های شمارشی: مثل تعداد افراد خانواده

۳- داده های رتبه ایی: مثل رتبه دانشجویان از ۱تا ۴

۴- داده های رده بندی شده: مثل جنسیت دانشجوها

تعریف

جامعه :

مجموعه عناصر مورد نظر برای مسأله ایی مفروض است(مثل مجموعه تولیدات یک کارخانه)

نمونه :

بخشی از جامعه تحت بررسی است، به قسمی که بتوان از آن نتایجی را در مورد جامعه استخراج نمود.

اندازه جامعه:

تعداد عناصر جامعه را اندازه جامعه می گویند.اگر تعداد عناصر جامعه متناهی باشد جامعه را متناهی در غیر این صورت نامتناهی می گوییم.

اندازه نمونه:

تعداد عناصر نمونه را اندازه نمونه می گویند.

آمار توصیفی:

برای اینکه نتایج مناسب و مطلوب از اطلاعات که در آمار گیری‌ها جمع آوری می‌کنیم، به دست آید باید:

اعداد نماینده واقعی مشاهدات بوده و غیرواقع یا غلط نباشند

به نحو مفیدی تهیه و تنظیم شوند

به نحو صحیح تجزیه و تحلیل گردند

قابل نتیجه گیری صحیح باشند

آمار توصیفی:

به طور کلی، روش‌هایی را که به وسیله آنها می‌توان اطلاعات جمع آوری شده را تنظیم کرده و خلاصه نمود، آمار توصیفی می‌نامیم و در یک کلام آمار توصیفی عبارت از مجموعه روشهایی است که پردازش داده‌ها را فراهم می‌سازد. اطلاع از اصطلاحات زیر در آمار ضروری است.

جمعيت

مجموعه افراد يا اشيايی را که می‌خواهيم يک يا چند خصوصيت مشترک آنها را مورد بررسی قرار دهيم، جمعيت يا جمعيت آماری می‌ناميم.

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 که 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.

نمونه

نمونه

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

نکته

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.

متغير

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.

متغير

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

مقیاسهاي اندازه‌گیري

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

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

مقياسهاي اندازه‌گيري

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.

مقياسهاي اندازه‌گيري

مقياس فاصله اي:

اين نوع مقياس اندازرزه‌گيري 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.

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.

داده

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

داده

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

Ph.D. Student Vahid Hajimirza

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 $f_1,f_2,\ldots,f_k$ تشکيل شده باشند،آنگاه f_i را فراواني x_i مي‌گوييم. به عبارت ديگر تعداد دفعاتي را که x_i در داده‌هاي $y_1,y_2,\ldots,y_n$ تکرار مي‌شود، فراواني x_i مي‌ناميم و آن را با نماد f_i نمايش مي‌دهيم.

به خاطر داشته باشيد که

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

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

مثال:

داده‌های زیر میزان تصادف منجر به مرگ رد ۳۰ منطقه را نشان می‌دهد. فراوانی دادها 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

فراواني

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

x_i	f_i
۲	۱
۳	۴
۴	۳
۵	۸
۶	۷
۷	۳
۸	۴

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:

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

فراواني نسبي

به خاطر داشته باشيد كه

براي $j = 1,\ldots,k$

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

$$0 \le r_i \le 1$$

Ph.D. Student Vahid Hajimirza

فراواني نسبي

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

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

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

با توجه به تعریف فراوانی، فراوانی تجمعی ردیف i را با نماد F_i نمایش می‌دهیم و به صورت زیر تعریف می‌کنیم:

فراواني تجمعي

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

به خاطر داشته باشيد كه

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

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

فراواني تجمعي

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

$$1+4=5$$

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

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

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

فراواني نسبي تجمعي

به خاطر داشته باشيد كه

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

$$r_{i=} R_1 \le R_2 \le \dots \le R_k = 1$$

فراواني نسبي تجمعي

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	$4/30$	$30/30$

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

مثال:

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

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

تشكيل جدول فراوانی
برای
داده های پيوسته

کمترین عدد ۱,۴ است و بیشترین عدد ۲,۹

چون داده‌ها تا یک رقم اعشار گرد شده‌اند، بنابراین می‌توان گفت که اندازه واقعی معدلها در فاصله [1/35,2/95]

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

تعداد ارقام گرد شده

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

Ph.D. Student Vahid Hajimirza

تعداد طبقات

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

کلاس	x_i	f_i	r_i	F_i	R_i
۱,۳۵-۱,۵۵	۱,۴۵	۴	۰,۰۸	۴	۰,۰۸
۱,۵۵-۱,۷۵۵	۱,۶۵	۶	۰,۱۲	۱۰	۰,۲۰
۱,۷۵-۱,۹۵	۱,۸۵	۱۲	۰,۲۴	۲۲	۰,۴۴
۱,۹۵-۲,۱۵	۲,۰۵	۹	۰,۱۸	۳۱	۰,۶۲
۲,۱۵-۲,۳۵	۲,۲۵	۸	۰,۱۶	۳۹	۰,۷۸
۲,۳۵-۲,۵۵	۲,۴۵	۶	۰,۱۲	۴۵	۰,۹۰
۲,۵۵-۲,۷۵	۲,۶۵	۲	۰,۰۴	۴۷	۰,۹۴
۲,۷۵-۲,۹۵	۲,۸۵	۳	۰,۰۶	۵۰	۱,۰۰
جمع		۵۰	۱,۰۰	–	–

Ph.D. Student Vahid Hajimirza

نمودار های به صورت زیر دسته بندی می شوند

۱. گسسته
   ۱. میله ای
   ۲. دایره ای
۲. پیوسته
   ۱. هیستوگرام
   ۲. چندبر فراوانی
   ۳. چندبر فراوانی تجمعی
   ۴. منحنیهای فراوانی و فراوانی تجمعی
   ۵. نمایش نمودار تنه و شاخه
   ۶. نمودار جعبه‌ای

نمودارهاي آماري

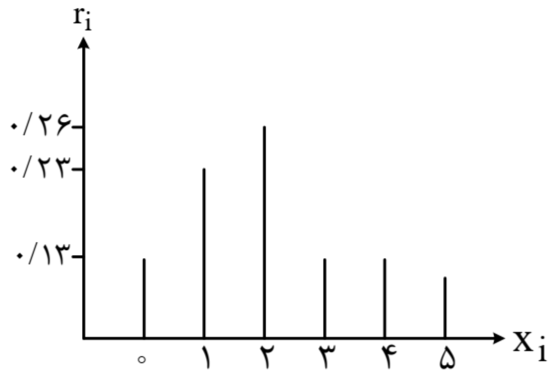
Ph.D. Student Vahid Hajimirza

معمولا داده‌ها را با نمودارهای مختلف نمایش می‌دهند. عموما این نمودارها در ارتباط با داده‌های پیوسته به کار گرفته می شود و منظور از نمایش آنها، تجسم عینی اطلاعات نهفته در داده‌ها است. در این بخش به معرفی چند نمودار معروف اکتفا می‌کنیم:

- هیستوگرام
- چندبر فراوانی
- چندبر فراوانی تجمعی
- منحنیهای فراوانی و فراوانی تجمعی
- نمایش نمودار تنه و شاخه
- نمودار جعبه‌ای

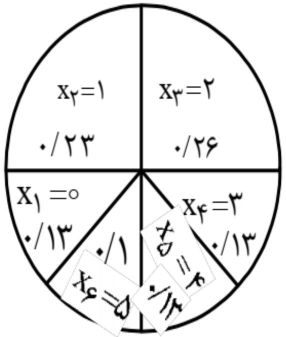
نمودار میله ای

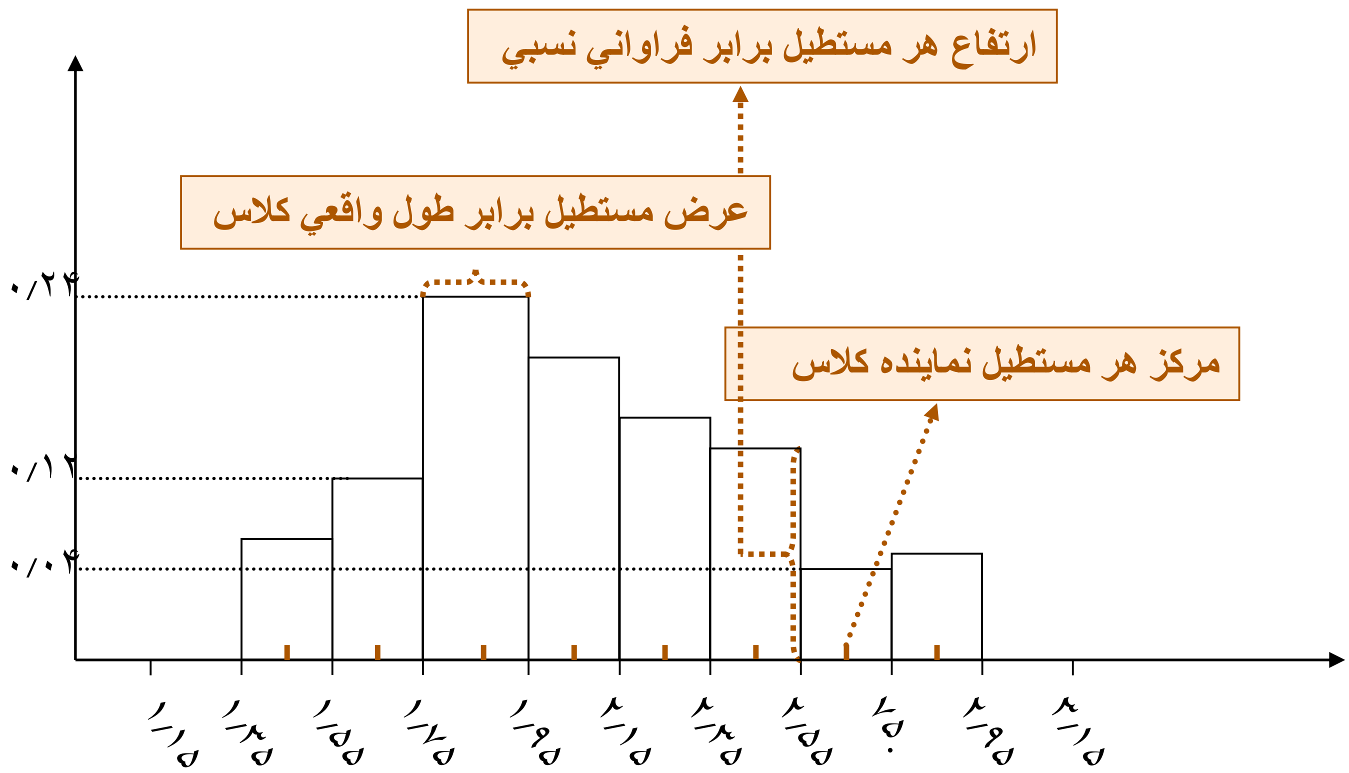
در این نمودار محور x ها نمایش دهنده مقادیر داده‌ها و محور y ها نمایش دهنده فراوانی نسبی می‌باشد. چون فراوانی نسبی داده‌های هر جامعه آماری مقادیر بین صفر و یک را می‌گیرد بنابر این می‌توان چندین نمودار را با یکدیگر مقایسه نمود. شکل نمودار میله‌ای را برای مثال ۱ نمایش می‌دهد.

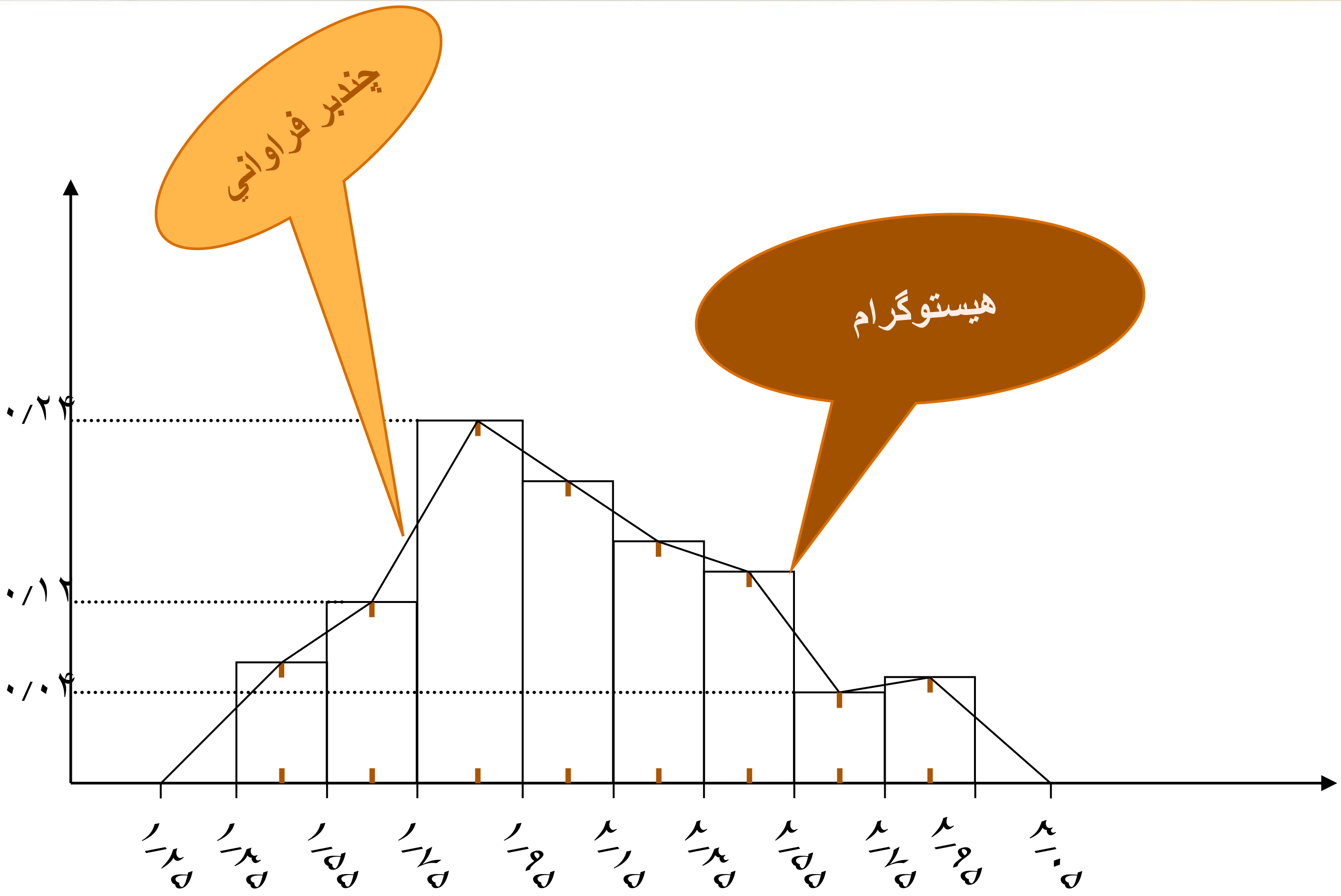


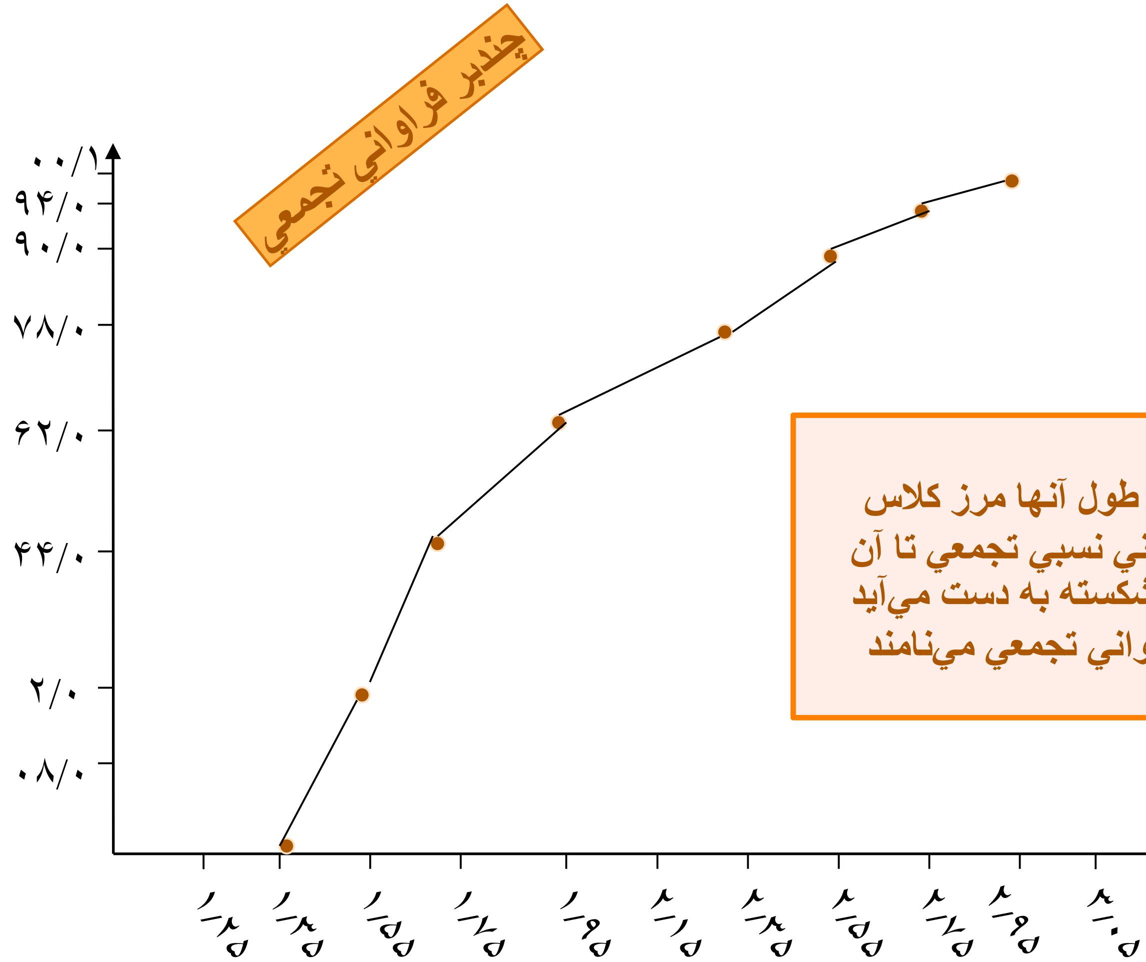
نمودار دایره ای

نمودار دایره‌ای : این نمودار درون دایره‌ای رسم می‌شود که به قطاع‌هایی تقسیم شده است و هر قطاع متناسب با مساحت خود معرف یکی از فراوانی‌های نسبی در جدول فراوانی می‌باشد. شکل زیر نمایش دهنده نمودار دایره‌ای برای مثال ۱ می‌باشد.




ارتفاع هر مستطيل برابر فراواني نسبي
عرض مستطيل برابر طول واقعي كلاس
مركز هر مستطيل نماينده كلاس
۰/۲۴
۰/۱۲
۰/۰۴

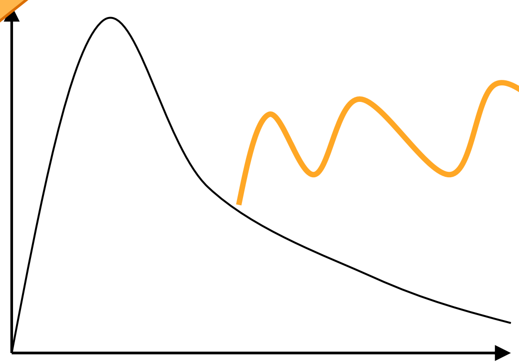

چندبر فراواني
هيستوگرام
۰/۲۴
۰/۱۲
۰/۰۴
۱/۲۵
۱/۳۵
۱/۵۵
۱/۷۵
۱/۹۵
۲/۱۵
۲/۳۵
۲/۵۵
۲/۷۵
۲/۹۵
۳/۰۵



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

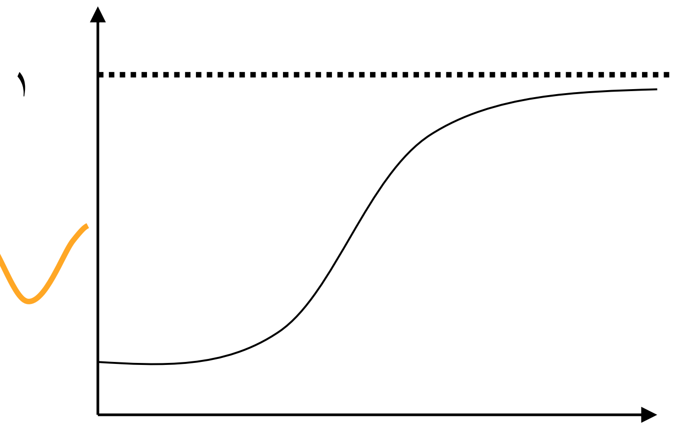
Ph.D. Student Vahid Hajimirza

منحني‌هاي فراواني و.....



نمودار منحني فراواني

نمودار منحني فراواني تجمعي



منحني‌هاي فراواني و.....

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

فراواني

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

نمایش نمودار تنه و شاخه

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

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

نمایش نمودار تنه و شاخه

پس از ساختن نمودار اولیه معمولا بهتر است مقادیر هر شاخه را از کوچك به بزرگ، با تعداد دفعات تکرار،مرتب کرد، به صورت زیر:

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

معيارهاي مركزي

Ph.D. Student Vahid Hajimirza

با استفاده از جدول فراوانی و رسم نمودارها می‌توانیم داده‌ها را به نحو مطلوبی **تنظیم کرده و اطلاعات نهفته را تا حدودی مشخص کنیم**. با این حال برای ارایه یک گزارش مناسب،بهتر است آنها را در یک یا چند عدد مناسب نیز خلاصه کنیم. چنین عددی می‌تواند معیار مرکزی باشد. مهمترین معیارهای مرکزی **میانگین**، **میانه و نما** است که در بخش این به شرح هر یک از آنها خواهیم 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 ترتیب با تعدادهاي $f_1, f_2, \ldots, f_k$ تشکیل شده باشند، آنگاه f_i را فراواني x_i مي‌گوييم.

میانگین

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

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 و گاها با me يا md نشان می دهند

ميانه

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

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

مثال میانه داده های گسسته

میانه

مثال ۱: ۵ ، ۸ ، ۲ ، ۱۲ ، ۳

شکل صعودی: ۲ ، ۳ ، ۵ ، ۸ ، ۱۲

$$\mathrm{m}d = \frac{n}{2} + \frac{1}{2} = \frac{5}{2} + \frac{1}{2} = 3$$

مثال میانه داده های گسسته

میانه

مثال ۲: ۵ ، ۸ ، ۲ ، ۱۲ ، ۳ ، ۶

شکل صعودی: ۲ ، ۳ ، ۵ ، ۶ ، ۸ ، ۱۲

me = ½* (6/2+(6/2+2/2))=3.5

میانه

محاسبه میانه براي داده‌هاي پيوسته

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

$$n/2 = 25$$

$$m = L_m + \frac{\frac{n}{2} - F_i}{f_i} * w$$

طول هر رده

L_m کران پایین

n تعداد کل

F_i فراوانی تجمعی رده قبل

f_i فراوانی همان رده

w طول رده

M=1.95+((25-22)/9)*0.2

M=2.02

چندک یک معیار کلی‌تر از میانه است و درعنوان حالت خاص میانه را نیز در بر می‌گیرد. اگر pیک عدد حقیقی بین صفر و یک باشد،آنگاه عدد Q_p را چندک مرتبه p می‌نامیم هر گاه p ۱۰۰٪ داده‌ها سمت چپ و (p -۱) ۱۰۰٪ داده‌ها سمت راست Q_p باشند.

چندکهاي معروف عبارتند از :

چندکها

چارکها اگر جامعه آماری به چهار قسمت مساوی تقسیم شود ، به هر یک از قسمت ها یک چارک گفته می شود و آنها را با Qنشان می دهند

چارکها به ازای p= ۰,۲۵ ، ۰,۵ ، ۰,۷۵ 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 ازای p=۱/۰، ۲/۰،..... ۹/۰، به دست می‌آیند و آنها را به ترتیب با نماد D_1 (دهک اول)، D_2 (دهک دوم)،...... و D_9(دهک نهم) نشان می‌دهند.

صدکها

صدکها به ازای p= 0.01,0.02,0.03 به دست می‌آیند و آنها را به ترتیب با نماد P_1 (صدک اول)، P_2 (صدک دوم)،.....و P_{99}(صدک نود و نهم) نشان می‌دهند.

Q_1 : مقداری که ۲۵٪ مشاهدات ، پایین تر از آن است

Q_2 : مقداری که ۵۰٪ مشاهدات ، پایین تر از آن است

Q_3 : مقداری که ۷۵٪ مشاهدات ، پایین تر از آن است

نحوه بدست آوردن چهارک :

برای بدست آوردن چارک ها نظیر میانه عمل می کنیم یعنی ابتدا اعداد را به صورت صعودی نوشته و سپس به کمک فرمول زیر محل چارک و نهایتا خود چارک پیدا می شود.

$$CQ_a = \frac{aN}{4} + \frac{1}{2}$$

چارک مورد نظر ۱و۲و۳ =a

تعداد مشاهدات =$\mathcal{N}$

چندکها

محاسبه چندك براي داده‌هاي 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 چندک

$$(n+1)p \begin{cases} \text{صحیح باشد} & r=(n+1)p, \quad Q_p = y_{(r)} \\ \text{صحیح نباشد} & r=[(n+1)p], \quad \omega=(n+1)p-r \longrightarrow Q_p = (1-\omega)y_{(r)} + \omega y_{(r+1)} \end{cases}$$

محاسبه چندك براي داده‌هاي پيوسته

چندكها

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

(p) (n)

$$0.25 \times 50 = 12.5$$

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

محاسبه نما براي داده‌هاي 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

مثال: براي داده‌هاي ۲، ۲، ۵، ۷، ۹، ۹، ۹، ۱۰، ۱۰، ۱۱، ۱۲و ۱۸ نما برابر ۹=M 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=4.5.

نما

محاسبه نما براي داده‌هاي پيوسته

کلاس	x_i	f_i	r_i
۱/۳۵-۱/۵۵	۴۵/۱	۴	۰۸/۰
۱/۵۵-۱/۷۵	۶۵/۱	۶	۱۲/۰
۱/۷۵-۱/۹۵	۸۵/۱	۱۲	۲۴/۰
۱/۹۵-۲/۱۵	۰۵/۲	۹	۱۸/۰
۲/۱۵-۲/۳۵	۲۵/۲	۸	۱۶/۰
۲/۳۵-۲/۵۵	۲/۴۵	۶	۱۲/۰
۲/۵۵-۲/۷۵	۶۵/۲	۲	۰۴/۰
۲/۷۵-۲/۹۵	۸۵/۲	۳	۰۶/۰
جمع		۵۰	۰۰/۱

از روی جدول ملاحظه می‌شود که فراوانی رده ۱.۷۵-۱.۹۵ دارای بیشترین فراوانی است بنابراین به عنوان رده نما در نظر می‌گیریم.

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

۰/۲۴-۰/۱۲

۲۴/۰-۱۸/۰

$$M = 1.75 + \frac{0.12}{0.12 + 0.06} \times 0.2 \quad = 1.9$$

با وجود این که در بسیاری از موارد، میانگین توصیف نسبتا کاملی از مجموعه داده‌ها ارائه می‌دهد، اما گاهی وجود اطلاعات بیشتر در مورد داده‌ها ضروری است. یک 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 تغیرات به عنوان معیارهای پراکندگی خواهیم پرداخت. مهمترین معیارهای پراکندگی عبارتند از **دامنه، میانگین انحراف ها از میانگین یا از میانه، میان دامنه چارکها، دامنه صدکی، واریانس و انحراف معیار است. علاوه بر مطالب فوق، در این بخش داده‌های استاندارد و ضریب تغیرات را نیز معرفی خواهیم کرد.**

واریانس

در این شاخص پراکندگی ، از مجذور (توان ۲) انحرافات استفاده می شود

فرمول

$$\delta_X^2 = \frac{\sum (X_i - \mu_X)^2}{N}$$

Ph.D. Student Vahid Hajimirza

مثال

Xi	Xi- μ
1	-2
2	-1
6	3
9 جمع	0

$$\delta_X^2 = \frac{\sum (X_i - \mu_X)^2}{N}$$

$$\mu = \frac{9}{3} = 3$$

$$\delta_X^2 = \frac{4+1+9}{3} = \frac{14}{3}$$

Ph.D. Student Vahid Hajimirza

خواص واریانس

۱ – اگر تمام مشاهدات با عدد ثابت b جمع شوند ، واریانس جدید تغییر نمی کند

۲– اگر تمام مشاهدات ، به عدد ثابت b ضرب شوند ، واریانس جدید b^2 برابر افزایش می یابد

انحراف معیار

این شاخص به منظور برطرف کردن عیوب شاخص واریانس و افزایش دادن تأثیر این انحراف توسط

$$\delta_X^2$$

فرمول انحراف معیار

$$\delta_X = \sqrt{\delta_X^2}$$

و یا

$$\delta_X = \sqrt{\frac{\Sigma (X_i - \mu_X)^2}{N}}$$

معيارهاي پراكندگي

Ph.D. Student Vahid Hajimirza

ضریب پراکندگی

ضریب پراکندگی یکی از معیارهای پراکندگی نسبی است که با فرمول زیر بیان می شود

$$C \cdot V = \frac{\delta_X}{\mu_X}$$

انحراف معیار مشاهدات = δ_X

میانگین مشاهدات = μ_X

کاربردهای ضریب پراکندگی

برای مقایسه دو جامعه در مواردی که :

۱- مقیاس ها یکسان نیستند

۲- مقیاس یکسان ولی تفاوت زیادی در بزرگی مشاهدات وجود دارد

۳- واریانسهای جوامع یکسان ولی میانگین هایشان متفاوت است

سری اعداد طبقه بندی شده معیارهای مرکزی میانگین

این فرمول برای داده های طبقه بندی شده به شرح ذیل است :

$$\mu_x \cong \frac{\sum F_i X_i}{N}$$

فراوانی مطلق = F_i

متوسط طبقات = X_i

کل مشاهدات = N

Ph.D. Student Vahid Hajimirza

مثال

C_I	F_i	X_i	F_iX_i
30-50	۸	۴۰	۳۲۰
50-70	۱۵	۶۰	۹۰۰
70-90	۲۵	۸۰	۲۰۰۰
90-110	۸	۱۰۰	۸۰۰
110-130	۴	۱۲۰	۴۸۰
جمع	۶۰		۴۵۰۰

$$\mu_x \cong \frac{\sum F_i X_i}{N}$$

$$\mu_x \cong \frac{4500}{60} = 75$$

Ph.D. Student Vahid Hajimirza

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 آن هستند كه دامنه، معرف واقعي ميزان انحراف باشد. در چنين مواردي، واريانس يك معيار مورد قبول همگان به شمارمي‌رود.

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 اندازه‌گيري به كار 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	$y_{(1)}$ کوچکترین و $y_{(n)}$ بزرگترین داده هستند $R = y_{(n)} - y_{(1)}$
میانگین انحرافها از میانگین	$MD_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{k} f_i \lvert x_i - \bar{x} \rvert$
میانگین انحرافها از میانه	$MD_m = \sum_{i=1}^{k} f_i \lvert x_i - m \rvert$
دامنه چارکي	$IQR = Q_3 - Q_1$
دامنه صدکي	$IPR = P_{90} - P_{10}$
واریانس داده‌ها	$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2$
انحراف معیار	$s = \sqrt{\frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}$
ضریب تغییرات	$v = \frac{s}{\bar{x}}$

اگر $y_1, y_2, \ldots, y_n$ نمايانگر داده‌هاي خام باشند، براساس جدول فراواني f_1 برابر x_1 و f_2 برابر x_2 تا f_k ها برابر x_k .

مي‌دانيم كه $\bar{x}$ و s به ترتيب ميانگين و انحراف معيار داده است. اگر از هر داده $\bar{x}$ را كم و بر s تقسيم كنيم ، يعني

داده‌هاي استاندارد

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

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.

معیارهای پراکندگی

فرض کنید یک دسته از دانشجویان در دو امتحان شرکت کرده‌اند و خلاصه نتاسج 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 مقیسه 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

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

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

چون $v_1 > v_2$ ، بنابراین نتیجه مي‌گیریم که دانشجویان در امتحان دوم نمرات مطلوبتري را کسب کرده‌اند.

ب) براي مقاسه، ابتدا نمرات دانشجو را استاندارد مي‌کنیم

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

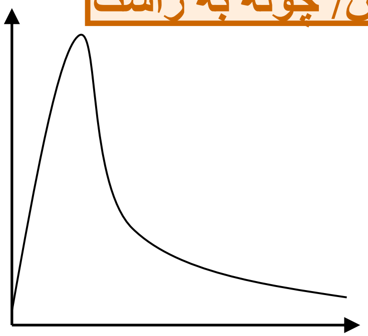
چون $z_1 < z_2$، بنابراین نمره آزمون دوم دانشجو د رمقایسه از موقعیت بهتري برخوردار است.

منحني هاي فراواني

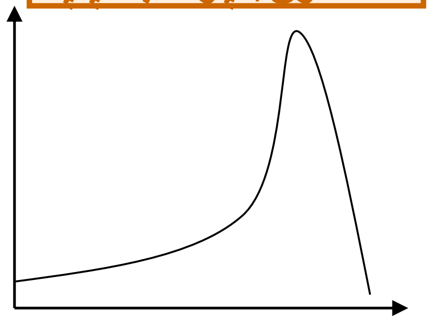
Ph.D. Student Vahid Hajimirza

منحني هاي 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 آل ترين 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.

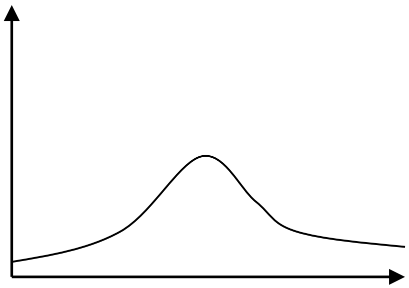
نامتقارن/ چوله به راست



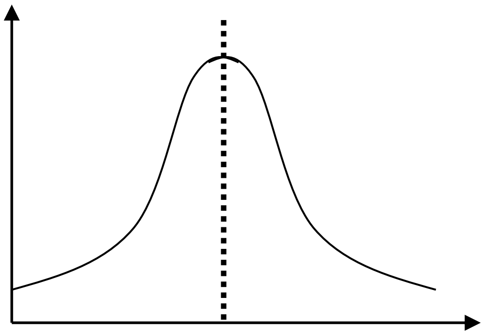
نامتقارن/ چوله به چپ



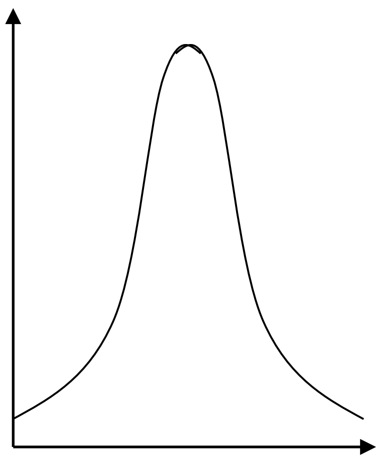
پخ



منحنی نرمال استاندارد



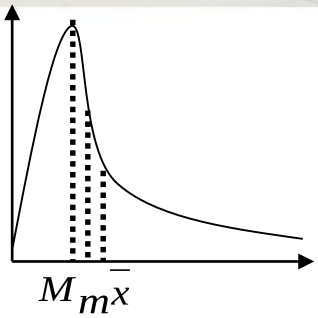
برجسته



چولگي:

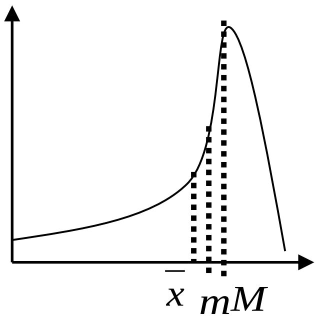
به راست يا مثبت

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



به چپ يا منفي

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



معیار اندازه گیری چولگی:

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

میزان کشیدگی یا پخی منحنی فراوانی را نسبت 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.