دانشگاه علوم پزشکی ، خدمات بهداشتی و درمانی استان گیلان

مدیریت آمار و فناوری اطلاعات

گروه آمار

دوره آموزشی :

آشنایی با روش های آماری

سال97

بسم الله الرحمن الرحيم

مفهوم آمار

علم جمع آوری ، سازماندهی ، تجزیه و تحلیل داده ها

جامعه آماری

هر مجموعه از افراد یا اشیاء و... را که حداقل دارای یک خصوصیت مشترک باشند.

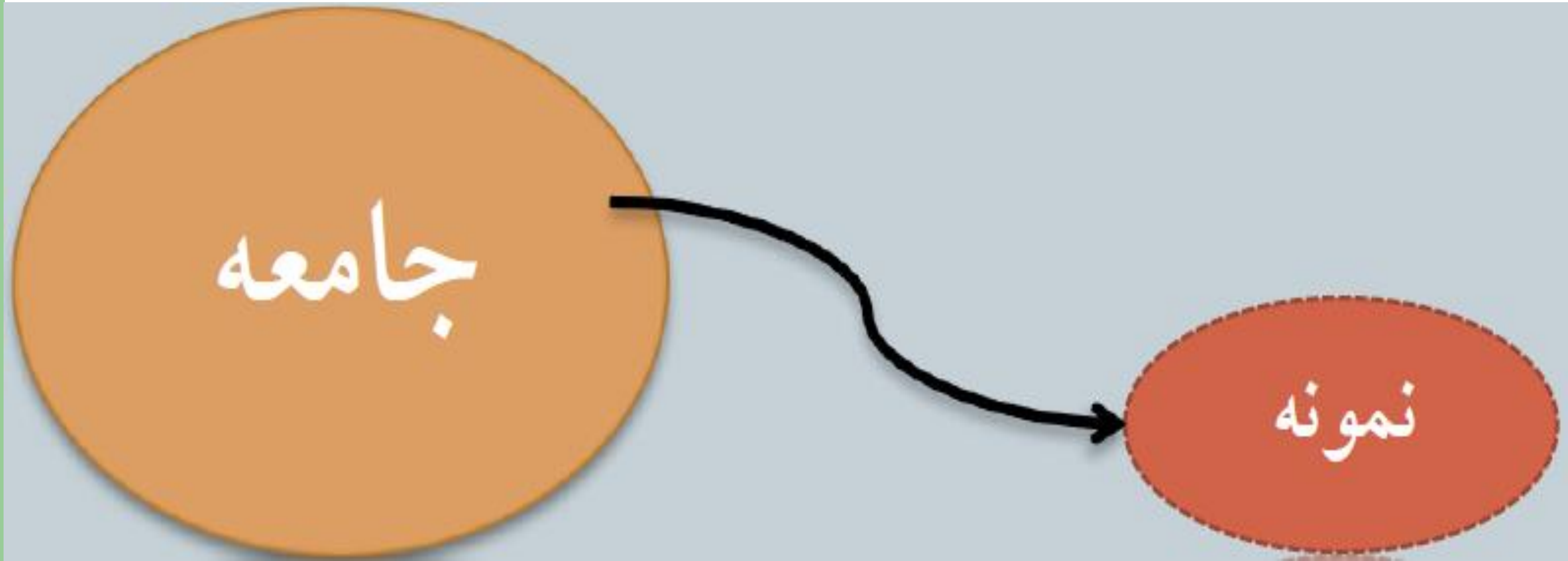
مثل:

مدیران با مدرک تحصیلی دکتری

فاکتورهای فروش کالا در یک ماه معین

تعریف نمونه

نمونه عبارتست از تعداد محدودی از آحاد جامعه آماری که بیان کننده ویژگی های اصلی جامعه باشد



مثال

قصد داریم در زمینه رضایت شغلی پزشکان پژوهش کنیم جمعیت

پژوهش ما پزشکان هستند

برای گردآوری اطلاعات باید به پزشکان مراجعه کنیم و نمونه خود

را از میان پزشکان انتخاب کنیم.

مثال

قصد داریم درباره میزان محبوبیت پزشکان در نزد بیماران به

پژوهش بپردازیم

با اینکه پژوهش درباره پزشکان است ولی جمعیت ونمونه دقیق

بیماران هستند زیرا داده ها نظرات بیماران هستند.

پارامتر، آماره

1- پارامتر : شاخص هایی که از طریق سرشماری (اندازه گیری تمامی عناصر جامعه آماری) بدست می آیند

2- آماره : شاخص هایی که از طریق نمونه گیری (اندازه گیری بخشی ازجامعه) بدست می آیند

آماربه دو نوع اصلی تقسیم می شود

1- آمار توصیفی

2- آمار استنباطی

آمار توصیفی

در آمار توصيفي يا مطالعه توصيفي صرفاً توصيف اطلاعات موجود

را بررسي مي‌كنيم، بعبارتی توصیف کل جامعه از طریق محاسبه

پارامترها

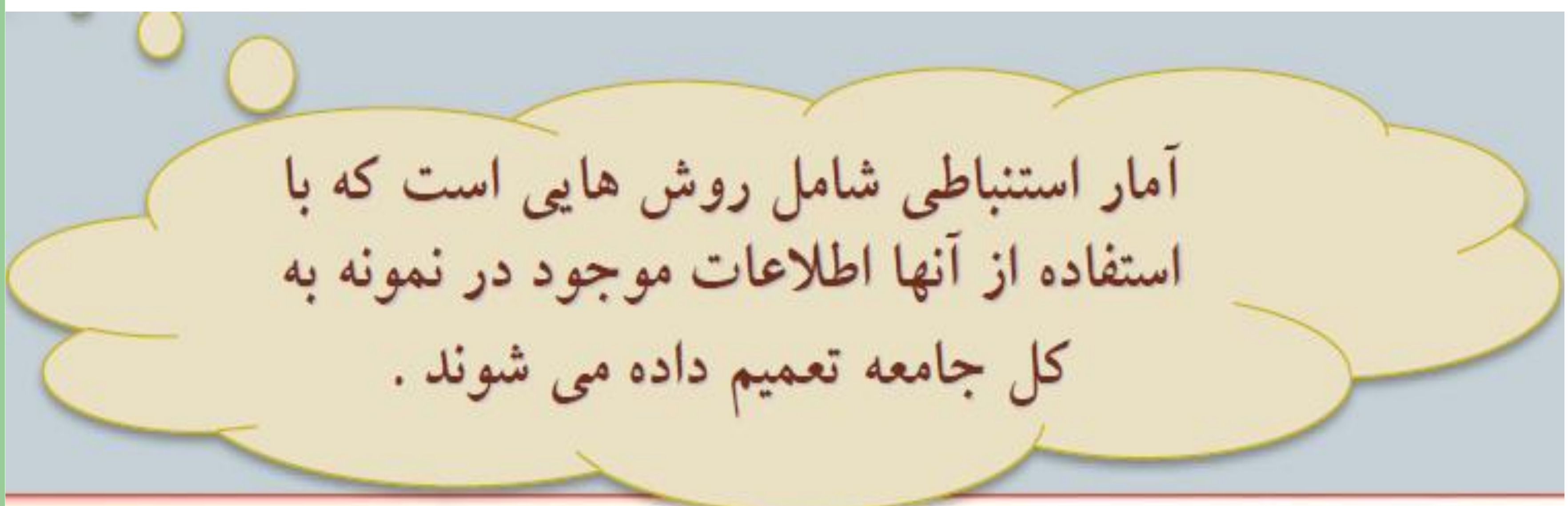
مثال آمار توصيفی

اگر مطالعه با عنوان بررسي وضعيت اقتصادي-اجتماعي ساكنين گيلان مَد نظر باشد

اطلاعات در قالب يك پرسشنامه با طرح سؤالاتي ازقبيل سن ، تعداد افراد خانواده ، شغل، ميزان درآمد و ...صورت گرفته و نتايج حاصل از مطالعه فقط وضعيت جامعه را در زمانيكه مطالعه انجام شده است را نشان مي‌دهد.

آمار استنباطی

آماری که در آن محقق ابتدا آماره ها را محاسبه و سپس به کمک تخمین و آزمون فرض آماری ، آنها را به پارامترهای جامعه تعمیم می دهد



صفت

هر عضوجامعه ویژگی هایی دارد که اعضای جامعه رابا آن ویژگی ها توصیف می کنیم. به هر ویژگی یک صفت آماری می گوییم.

مثال: تعداد ساعات حضور در کلاس، نمره کلاسی، نمره میان ترم و نمره پایان ترم صفت هایی برای ارزشیابی یک دانشجو هستند.

صفت مشخصه

صفت هایی است که در همه اعضا جامعه به طور یکسان وجود دارند

مثل:

جامعه ایرانیان که صفت مشخصه آن ایرانی بودن است.

صفت متغیر

صفت هایی هستند که از هر عضو به عضو دیگر تغییر می کنند.

مثل گروه خونی افراد شهراین متغیرها هستند که عملا در حین تحقیق

مورد سوال و اندازه گیری قرار می گیرند.

داده ها، مقادیر اندازه گیری شده یک متغیر هستند.

دسته بندی متغیرها

دسته بندی متغیرها کمک می کند مسیر مناسبی را برای رسیدن به نتایج درست، طی کنیم.

انواع متغیرها

- متغیرها
  - کیفی
    - اسمی
    - رتبه ای
  - کمی
    - گسسته
    - پیوسته

انواع متغیر ها

الف : متغیر کمی:(قابل اندازه گیری یا شمارش)

ب : متغیر کیفی:(غیر قابل اندازه گیری)

متغیرکمی

متغیرهایی هستند که قابل اندازه گیری یا شمارش یا قابل مقایسه و

سنجش هستند و همواره می توان عددی (اعشاری یا صحیح) را به

آن ها نسبت داد،این گونه متغیرها را می توان جمع و تفریق کرد.

مثل سن افراد بازنشسته

انواع متغیرهای کمی

الف : متغیر کمی پیوسته

ب : متغیر کمی گسسته

متغیر کمی پیوسته

یک متغیر کمی را پیوسته گویند، هرگاه مقدارaوb را اختیار کرد، بتواند

هر مقدار بین آن دو رانیز اختیار کند ، به عبارت دیگر، عدد اعشاری

برای آن مفهوم داشته باشد.

مانند قد،وزن،سن،متراژو...

متغیر کمی گسسته

یک متغیر کمی را گسسته گویند، هرگاه مقدارaوbرا اختیار کرد،نتواند

هر مقدار بین آن دو رانیز اختیار کند ، به عبارت دیگر، عدد اعشاری

برای آن مفهوم نداشته باشد.

مانند:تعداد افراد خانوار،تعداد تخلفات رانندگی،تعداد غایبین کلاس و..

متغیرکیفی

متغیرهایی هستند که واحد ندارند و قابل شمارش و یا اندازه گیری نیستندو فقط ممکن است به نوع یا ترتیب خاصی تعلق داشته باشند

این گونه متغیرها را نمی توان جمع و تفریق کرد.

مانند شغل(آزاد یا دولتی)،جنس(مردیازن)

تعهد شغلی و انسجام اجتماعی

انواع متغیرهای کیفی

الف : متغیر کیفی ترتیبی

ب : متغیر کیفی اسمی

متغیر کیفی ترتیبی

در این متغیرها یک نوع ترتیب طبیعی وجود دارد

مانند(بالا،وسط،پایین)،(خوب،متوسط،ضعیف)

متغیر کیفی اسمی

در این نوع متغیرها هیچ نوع ترتیبی وجود ندارد و هیچ مقایسه ای بین متغیرها امکان پذیر نیست.

مانند گروه خونی (o,AB,B,A)

مثال متغیرهای کیفی

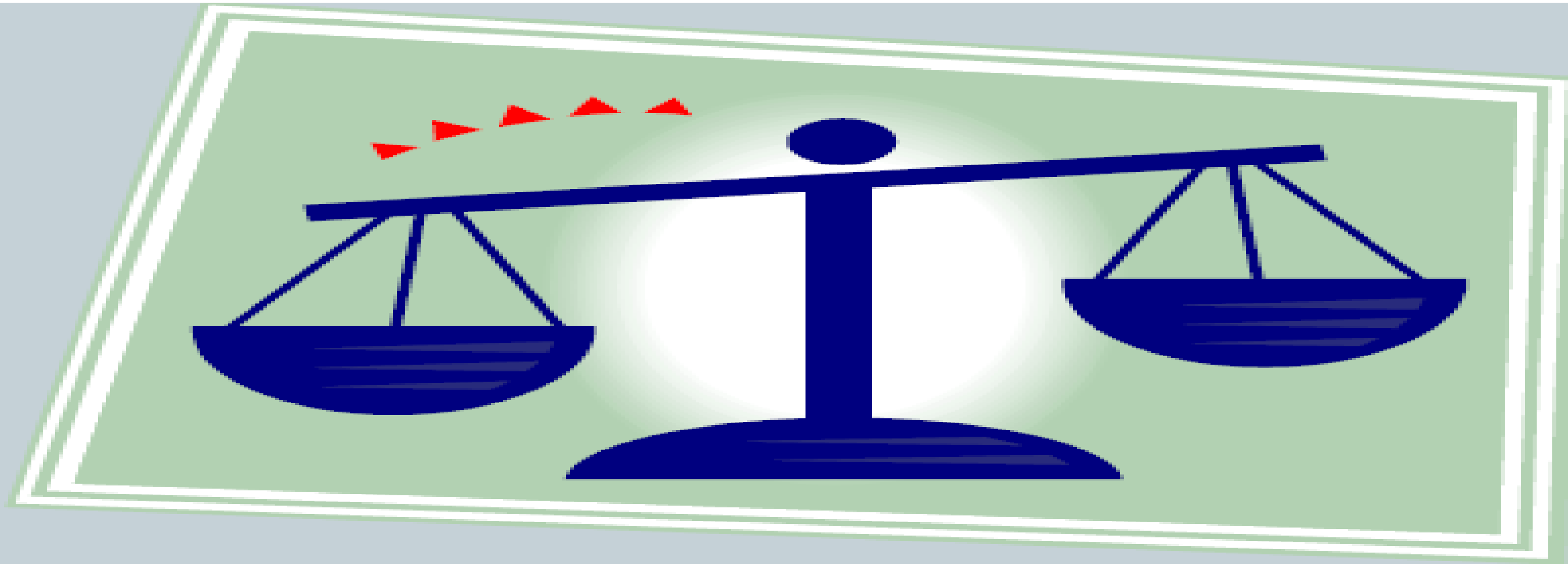
اگر به هریک از چهار روش تدریس زبان در تحقیقی اعداد 1تا 4 را اختصاص دهید، با یک متغیرکیفی اسمی روبه رو هستید و اگر میزان مهارت زبان آموز را با اعداد ا تا 3 رتبه بندی کنید، یک متغیر کیفی ترتیبی خواهید داشت.

مقایسه متغیرهای کمی و کیفی

متغیرهای کمی به دلیل قابلیت اندازه گیری،دقت بیشتری دارند و نتایج حاصل از آنها بیشتر قابل تعمیم به جامعه است.

اندازه گیری

اندازه گیری عبارت است از نسبت دادن اعداد به خصوصیات اشیاء و وقایع یا افراد بر طبق قواعدی منطقی و قابل قبول



مقیاس های اندازه گیری متغیر ها

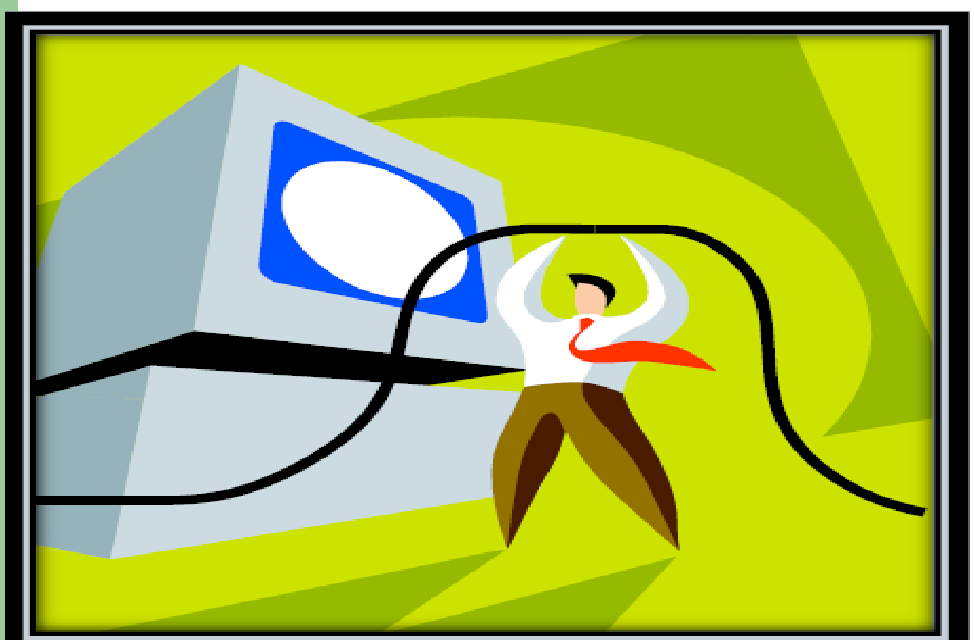
انواع متغیرهای(کمی وکیفی) با توجه به مقیاس های مختلف،قابل اندازه گیری هستند،این مقیاس ها عبارتند از:

1- مقیاس اسمی

2- مقیاس ترتیبی

3- مقیاس فاصله ای

4- مقیاس نسبی



مقیاس اسمی

متغیرهایی هستند که جنبه کیفی یک صفت را در نظر گرفته، بدین معنا که کدهایی که به پاسخ اختصاص داده می شوند واولویتی بر یکدیگر ندارند.

مثال : متغیر نوع رنگ :1: سفید ، 2:سیاه ،3: زرد

مقیاس ترتیبی

اگر بین اسامی ایجاد شده یا طبقات حاصله ناشی از مقیاس بندی اسمی یک نوع رابطه هم وجود داشته باشد پژوهشگران از مقیاس ترتیبی استفاده می نمایند.

مثال : متغیر میزان رضایت : 1 : ناراضی ، 2: متوسط ، 3: راضی

مقیاس فاصله ای

وقتی مقیاسی همه خصوصیات یک مقیاس ترتیبی را دارا باشد وعلاوه برآن، فاصله بین اعداد یا طبقات از یک نظم خاصی پیروی نماید (فواصل یکسان باشند) محققان از مقیاس فاصله ای برای اندازه گیری متغیرها استفاده می نمایند

نقطه صفر اختیاری و قراردادی است

مثال :دما برحسب سانتی گراد

مقیاس نسبی

مقیاسی است که علاوه بر داشتن همه خصوصیات مقیاس فاصله ای

، دارای نقطه صفر واقعی نیز هست .

می توان واحدهای آن را برهم تقسیم یا نسبت آن ها را حساب کرد.

مثال :،میزان موالید،میزان رشدجمعیت

رابطه بین مقیاس ها

مقیاس نسبی قویترین مقیاس و مقیاس اسمی ضعیفترین مقیاس است.

توصیف اطلاعات

اندازه های ثبت شده در مجموعه داده ها، اجزای اساسی اطلاعاتی هستند که باید به دقت آنها را مورد بررسی قرار داده و به خوبی توصیف کنیم. به همین دلیل در آمار توصیفی اهداف زیر را دنبال می کنیم.

مراحل اساسی توصیف داده ها

- خلاصه‌کردن، دسته‌بندی کردن و تهیه جدول‌های توزیع فراوانی
- رسم نمودارهای مناسب برای توصیف بصری داده ها
- بدست آوردن شاخص هایی که گرایش به مرکز داده‌ها را معین می کنند (شاخص های مرکزی)
- بدست آوردن شاخص هایی که میزان انحراف از مرکز داده‌ها را معلوم می کنند (شاخص های پراکنگی)

انواع طبقه بندی داده ها

کمی

- طبقه بندی داده ها بر اساس یک متغیر کمی مانند سن و وزن

کیفی

- طبقه بندی داده ها بر اساس یک متغیر کیفی مانند جنسیت، نژاد و شغل

زمانی

- طبقه بندی داده ها بر اساس زمان گرد آوری آنها مثل میزان بارندگی ماههای سال(داده های سری زمانی)

جغرافیایی

- طبقه بندی داده ها بر حسب مناطق جغرافیایی مانند کشورها، استانهاو شهرها

جدول توزیع فراوانی

یعنی جدول مرتب و خلاصه شده از داده ها و مشاهدات که تکرار وقوع هر داده ها در آن مشخص شده است.

دسته بندی داده ها

برای دسته بندی داده های کمی گسسته باید هریک از مقادیر متغیر را به همراه فراوانی آن ها در یک جدول قرار دهیم.

برای دسته بندی داده های کیفی کافی است فراوانی هریک از سطوح متغیر کیفی را در یک جدول قرار دهیم.

مثال

مثال

تعداد ۵۰ نفر از دانشجویان کارشناسی ارشد در تحقیقی مورد بررسی قرار دادیم. تعداد واحد های گذرانده در جدول فراوانی زیر آمده است:

تعداد واحد های گذرانده	فراوانی fi
0	10
8	13
14	17
16	10
جمع	50

طبقه بندی متغیر کمی (گسسته)

در مدیریت ، انسان ها را به لحاظ ارتباط به ۴ دسته تصویری، احساسی، صوتی و ارقامی تقسیم می‌کنند. اطلاعات جدول زیر نتایج به دست آمده از ۱۰۰ نفر است.

نوع ارتباط	فراوانی fi
تصویری	35
احساسی	20
صوتی	32
ارقامی	13
جمع	100

طبقه بندی متغیر کیفی

فراوانی ها

فراوانی مطلق

به تعداد داده ها در هر طبقه فراوانی مطلق آن طبقه می گویند و آن را با fi نشان می دهند.

فراوانی نسبی

اگر فراوانی های مطلق را بر کل فراوانی ها تقسیم کنیم، فراوانی نسبی (ri) به دست می آید.

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

فراوانی تجمعی

به مجموع فراوانی های مطلق طبقه های قبل و همان طبقه، فراوانی تجمعی آن طبقه می گویند و آن را با Fi نمایش میدهند.

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

فراوانی تجمعی نسبی

می توان از تقسیم فراوانی های تجمعی بر تعداد داده ها، این فراوانی را به دست آورد.

$$R_i = \frac{F_i}{n}$$

مثال

مثال

جدول توزیع فراوانی همراه با فراوانیهای مطلق، تجمعی، نسبی، تجمعی نسبی برای تعداد ۵۰ کارمند در مثال قبل

2+13+18

$^{13}/_{50}=0.26$

$^{2}/_{50}=0.04$

$^{43}/_{50}=0.86$

x_i	f_i	F_i	r_i	R_i
0	2	2	0/04	0/04
1	13	15	0/26	0/3
2	18	33	0/36	0/66
3	10	43	0/2	0/86
4	7	50	0/14	1

به خاطر داشته باشید که

$$\sum_{i=1}^{k} r_i = 1 \quad , \quad \sum_{i=1}^{k} f_i = n \quad , \quad R_i = \sum_{j=1}^{i} r_j$$

طبقه بندی داده های پیوسته

اگر داده ها از یک صفت کمی پیوسته اندازه گیری شده باشند، لزوما

تکرار نخواهند داشت بنابراین باید آنها را به روش زیر طبقه بندی کرد

۱- تعداد طبقه (K) را به تناسب تعداد داده ها (n) و یا با استفاده از رابطه زیر تعیین کنید.

$$\boxed{k = ۱ + ۳/۳۲۲\log(n)}$$

(دستور استورجس)

۲- دامنه تغییرات داده ها (R) را از رابطه زیر محاسبه کنید.

کوچکترین داده

$$\boxed{R = Max(x_i) - Min(x_i)}$$

بزرگترین داده

۳- فاصله طبقه ها (c) را از رابطه زیر به دست آورید:

(با تقریب اضافی)

$$\boxed{c = R/k}$$

مثال

4- حدود هر طبقه (Li - Hi) را تعیین کنید وداده هایی که در هر طبقه قرار می گیرند، شمارش کرده و به عنوان فراوانی (fi) آن طبقه ثبت کنید.

مثال

وزن ۵۰ عدد کالای تولیدی کارخانه ای با تقریب تا یک رقم اعشار، به شرح زیر است.
این داده ها را به ۸ طبقه تقسیم بندی کنید و فراوانی های مطلق نسبی تجمعی و تجمعی نسبی را به دست آورید.

2/9	1/5	1/8	2/4	2/2	2/1	2/2	1/6	1/9	2/1
2/6	2/1	2/5	2/0	2/3	2/3	1/7	1/8	2/3	1/8
1/4	2/6	2/2	1/9	2/0	1/7	1/7	1/9	1/2	1/8
2/0	2/0	2/0	2/5	2/2	2/2	1/9	1/8	2/4	2/9
1/4	1/6	1/9	2/9	2/4	1/6	1/8	1/9	2/5	1/4

برای طبقه بندی این داده ها، مراحل دسته بندی کردن داده های کمی پیوسته را پی می گیریم

بزرگترین و کوچکترین عدد در بین داده ها را مشخص کرده ایم

مثال

چهار مرحله طبقه بندی داده های پیوسته برای این مثال

حدود طبقات	x_i	f_i	r_i	F_i	R_i
[1.4-1.6)	1.5	4	0.08	4	0.08
[1.6-1.8)	1.7	6	0.12	10	0.20
[1.8-2.0)	1.9	12	0.24	22	0.44
[2.0-2.2)	2.1	9	0.18	31	0.62
[2.2-2.4)	2.3	8	0.16	39	0.78
[2.4-2.6)	2.5	6	0.12	45	0.90
[2.6-2.8)	2.7	2	0.04	47	0.94
[2.8-3.0]	2.9	3	0.06	50	1.00
جمع		50	1.00	_	_

1- تعداد طبقه در صورت مسئله، ۸ تعیین شده است

۲- دامنه تغییرات را ۲/۹ −۱/۲=۱/۷ به دست می آوریم

۳- فاصله طبقات از رابطه زیر (c) و با تقریب اضافه ۰/۲ به دست آمده است.

۴- حدود طبقات را با شروع از کمترین مقدار تعیین کرده ایم.

$$c=\frac{۲/۹-۱/۲}{۸}=۰/۲$$

انواع جدول

گزارش نتايج بيشتر در حجم كمتر

يك بعدي

دو بعدي

سه بعدي

جدول یک بعدی

تنها اطلاعات توصيفي يك متغير بيان مي‌شود

جدول ۱: توزیع فراوانی افراد برحسب استعمال سیگار

استعمال سیگار	فراوانی مطلق	فراوانی نسبی	فراوانی تجمعی
سیگار می‌کشد	....	....	....
سیگار نمی‌کشد	....	....	....
جمع	....	....	....

One-variable table (Age)

Primary and secondary syphilis morbidity
by age, United States, 1989

Age group (years)	Cases	
	Number	Percent
≤14	230	0.5
15-19	4,378	10.0
20-24	10,405	23.6
25-29	9,610	21.8
30-34	8,648	19.6
35-44	6,901	15.7
45-54	2,631	6.0
≥55	1,278	2.9
Total	**44,081**	**100.0***

*Percentages do not add to 100.0% due to rounding.

جدول دو بعدی

اطلاعات مربوط به دو متغير بيان ميگردد و ميتوان به وجود يا عدم وجود ارتباط بين دو متغير پي برد.

جدول ۲: توزیع فراوانی افراد برحسب استعمال سیگار به تفکیک جنسیت

جنس / سیگار	مرد	زن
می کشد	....	....
نمی کشد	....	....
جمع	....	....

Two-variable table

Newly reported cases of primary and secondary syphilis by age and sex, United States, 1989

Age group (years)	Number of cases by sex		
	Male	Female	Total
≤14	40	190	230
15-19	1,710	2,668	4,378
20-24	5,120	5,285	10,405
25-29	5,304	4,306	9,610
30-34	5,537	3,111	8,648
35-44	5,004	1,897	6,901
45-54	2,144	487	2,631
≥55	1,147	131	1,278
Total	**26,006**	**18,075**	**44,081**

جایگاه جمع در جدول دو بعدی کجاست؟

	بیمار		سالم		کل	
سن	تعداد	درصد	تعداد	درصد	تعداد	درصد
0-9	15		175			
10-19	10		90			
20-29	5		15			
.....						
......						

پاسخ: در جهت متغیر مستقل

جدول سه بعدی

بیان ارتباط 3 متغیر

جدول ۳: توزیع فراوانی افراد بر حسب استعمال سیگار به تفکیک سن و جنس

جنس	مرد		زن	
سن / سیگار	< ۵۰	> ۵۰	< ۵۰	> ۵۰
می کشد	....	....	....	....
نمی کشد	....	....	....	....
جمع	....	....	....	....

نمودارهای آماری

منظور از نمایش نموداری داده ها، تجسم عینی اطلاعات نهفته در آنها ست.

نمودار دايره‌اي (كلوچه‌اي) (Pie Chart)

نمودار ميله‌اي (نرده اي، ستوني) (Bar Chart)

نمودار هيستوگرام يا مستطيلی

نمودار چندضلعی (Polygon Diagram)

نمودار جعبه ای

محاسن نمودارها

استفاده از نمودارها در گزارش نویسی باعث می شود که خوانندگان با صرف کمترین زمان و با ساده ترین بیان ، گزارش را بفهمند و تصویری روشن از توزیع داشته باشند.

نمودار دایره ای

این نمودار ابزار مناسبی برای تجسم مشاهدات بوده و معمولاً بر حسب درصد تهیه می شود و به نمودارکلوچه ای نیز معروف است.

معمولا برای داده های کیفی استفاده می شود.

مراحل تهیه نمودار دایره ای

1- تبدیل فراوانی مطلق به نسبی

2- پیدا کردن مساحت هر قطاع از دایره

3- تقسیم مساحت دایره بر حسب S_i ها

4- نوشتن نوع و درصد مشاهدات بر روی دایره

فرمول مساحت هر قطاع

برای پیدا کردن مساحت هر قطاع از دایره از فرمول زیر استفاده می شود

$$S_i = 360\& \times f_i$$

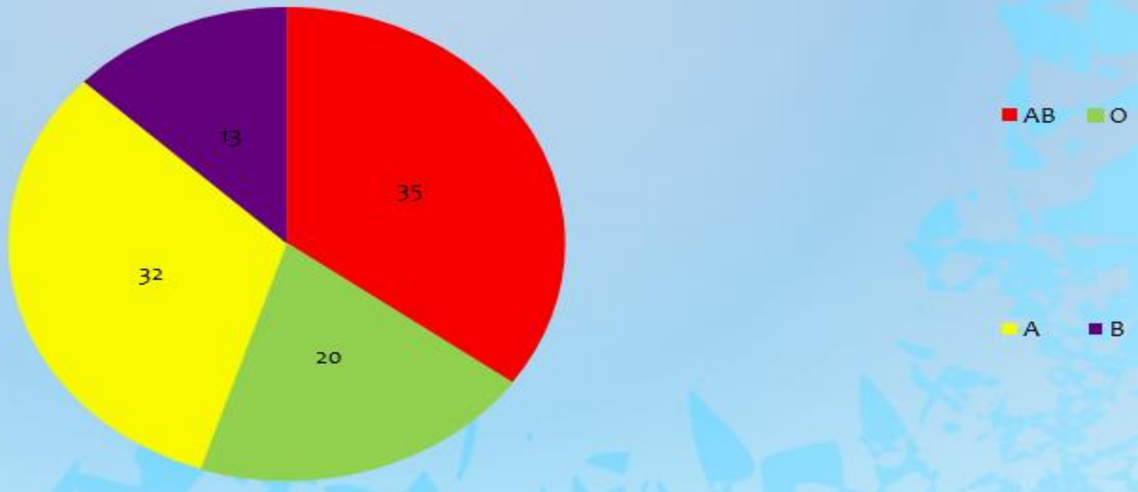
یعنی فراوانی نسبی هر مشاهده به عدد 360 ضرب می شود

مثال

فراوانی مطلق	سطوح متغیر	
35	AB	گروه خونی
20	O	
32	A	
13	B	

زاویه :

فراوانی نسبی * 360



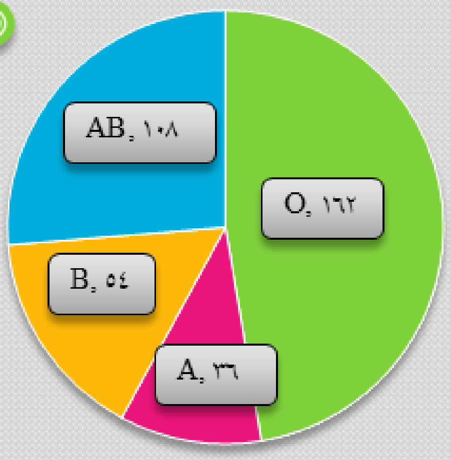
مثال

مثال: گروه خون ۱۰۰ بیمار بستری در یک بیمارستان را در اختیار داریم، نمودار دایره‌ای داده‌ها را رسم می‌کنیم.

نحوه رسم نمودار دایره‌ای را در این مثال ببینید.

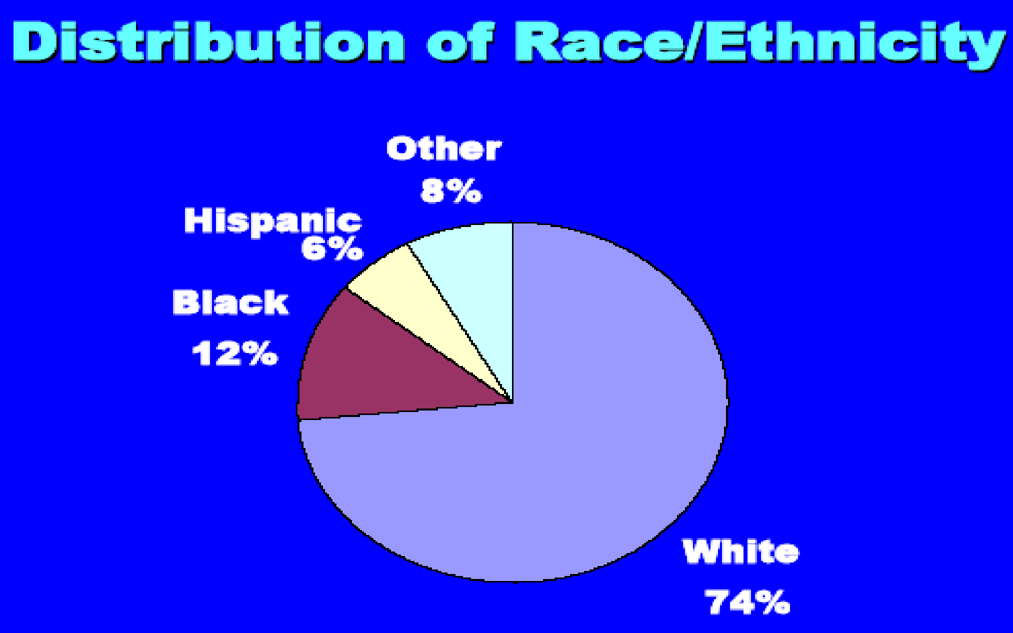
گروه خون	فراوانی fi	فراوانی نسبی ri	قطاع دایره (درجه)
O	٤٥	٠/٤٥	١٦٢
A	١٠	٠/١	٣٦
B	١٥	٠/١٥	٥٤
AB	٣٠	٠/٣	١٠٨

$٠/١٥ \times ٣٦٠ = ٥٤$



مثال

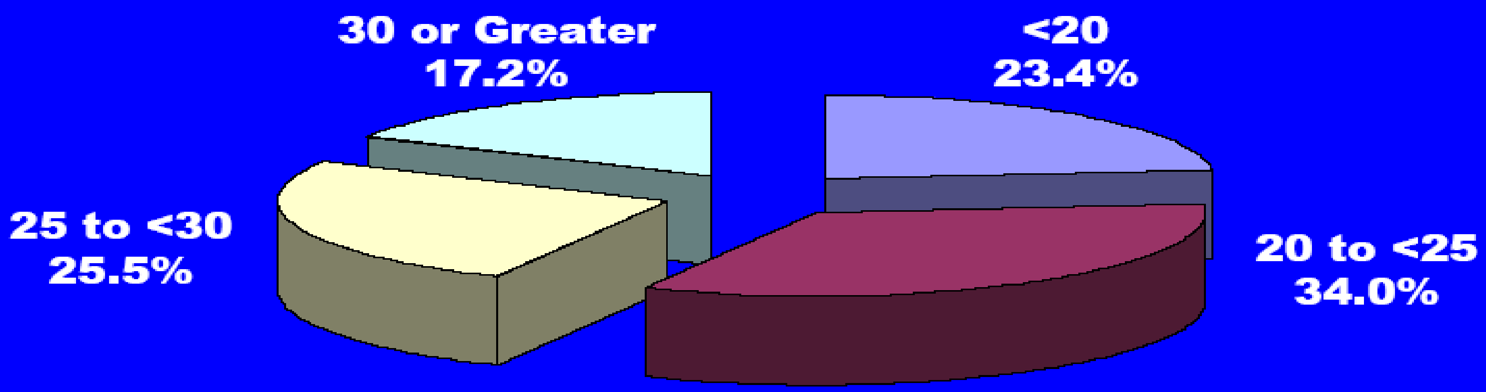
Example 1 of a Pie Chart



مثال

Example 2 of a Pie Chart

Distribution of Body Mass Index



نمودار ستونی

این نمودار در یک دستگاه مختصات که محور افقی نشان دهنده کیفیت مشاهدات و محور عمودیش نشان دهنده فراوانی مطلق یا نسبی هر گروه است ترسیم می شود

معمولا برای داده های کمی گسسته استفاده می شود.

مثال

روش رسم نمودار ستونی

برای رسم نمودار ستونی کافی است مقادیر متغیر را روی محور افقی تعیین کنید سپس مستطیل هایی با فاصله و به ارتفاع فراوانی مطلق برای هر مقدار متغیر رسم کنید.

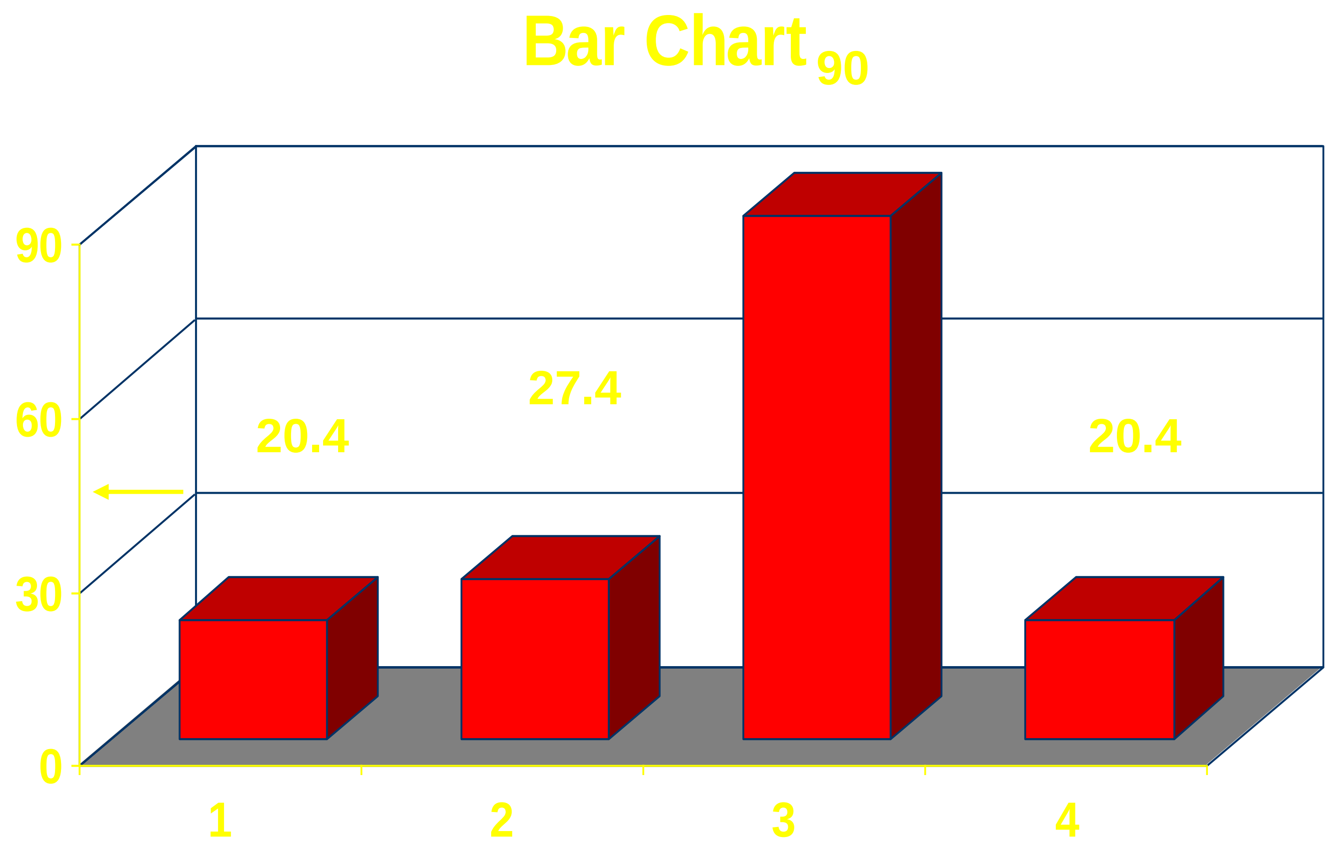
مثال برای داده های زیر نمودار ستونی رسم کنید.

3 1 2 2 0 1 2 3 2 1 1 5 2 0 1 2 2 4 4 3

x	fi
0	2
1	5
2	7
3	3
4	2
5	1

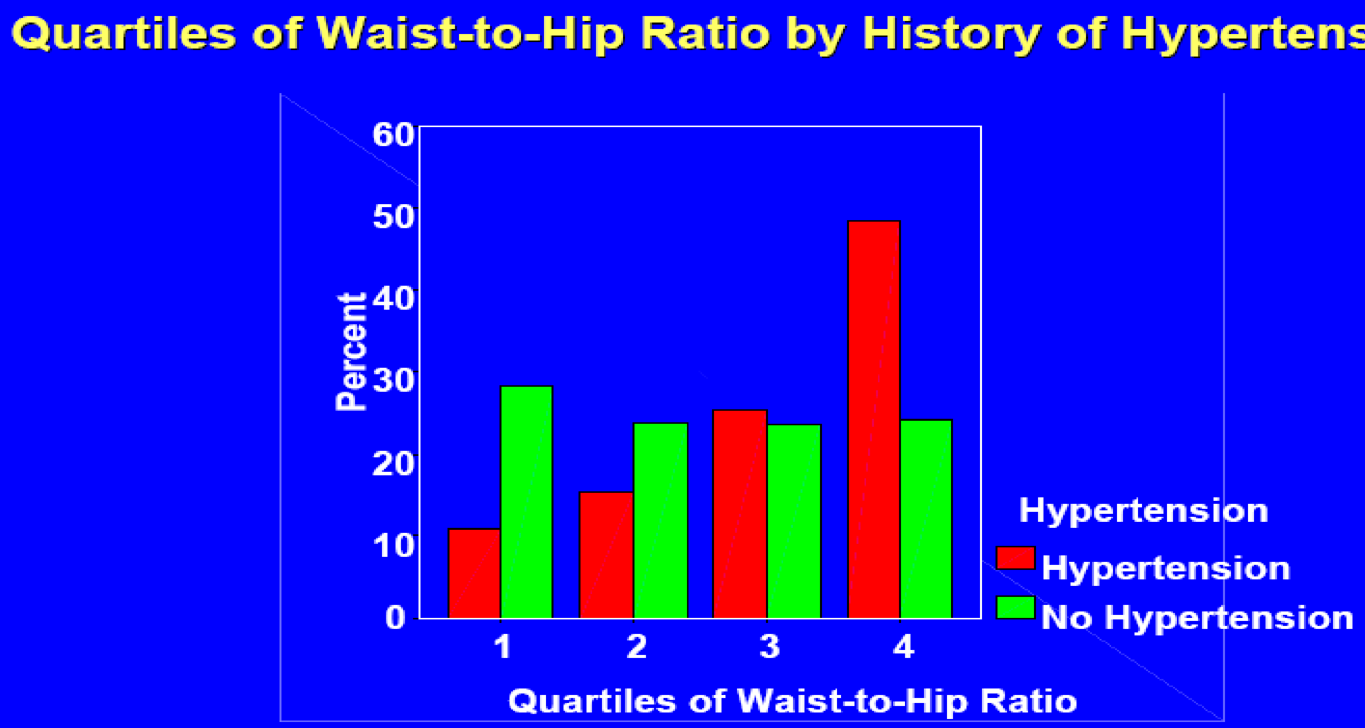
ابتدا جدول توزیع فراوانی را تشکیل می دهیم

مثال



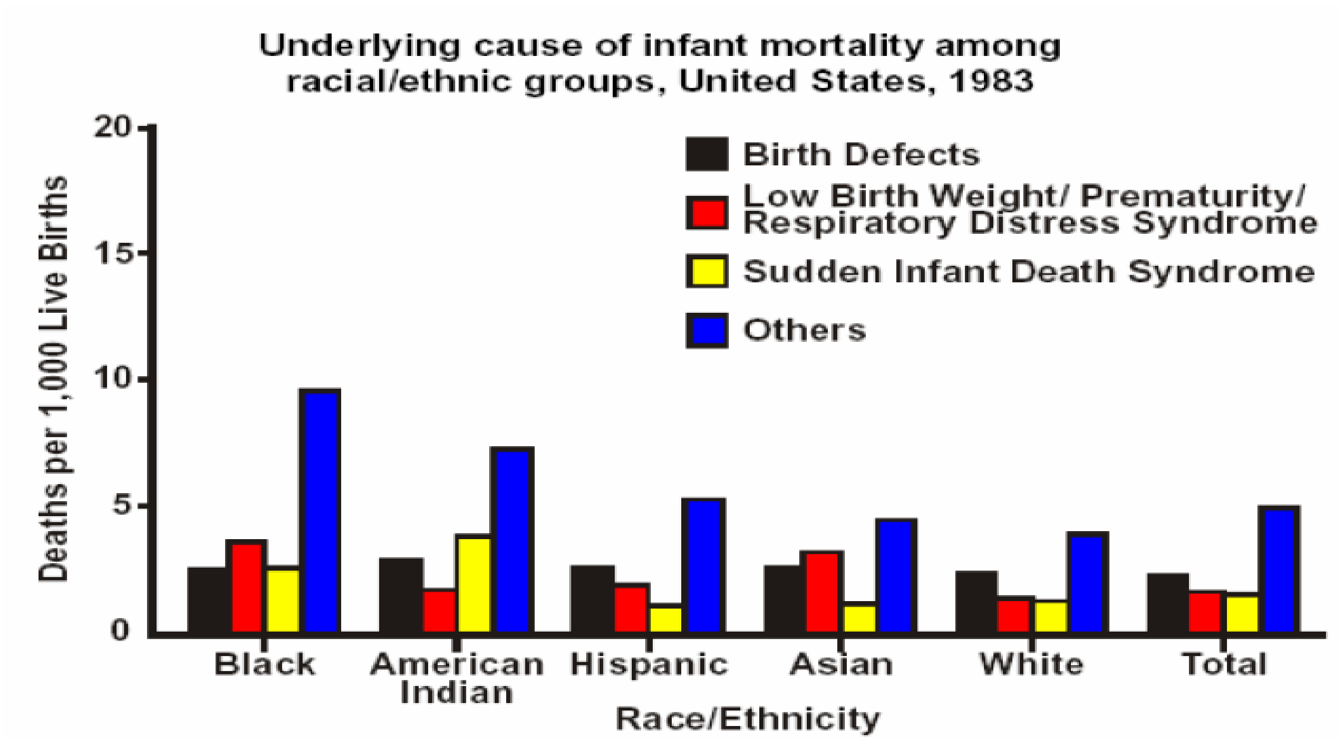
مثال

Example 2 of a Bar Chart



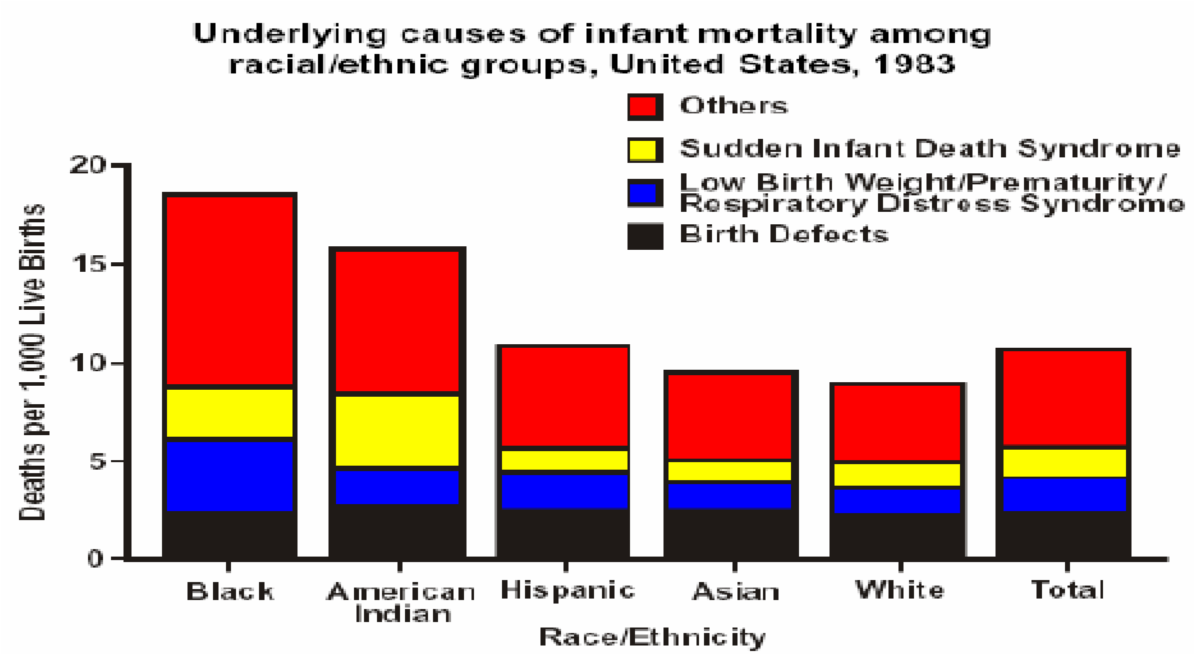
مثال

Grouped Bar Charts



مثال

Stacked Bar Charts



نمودار مستطیلی یا هیستوگرام

برای نمایش داده هاي کمي پیوسته محور عمودي فراواني هرگروه از متغیر کمي ميباشد.

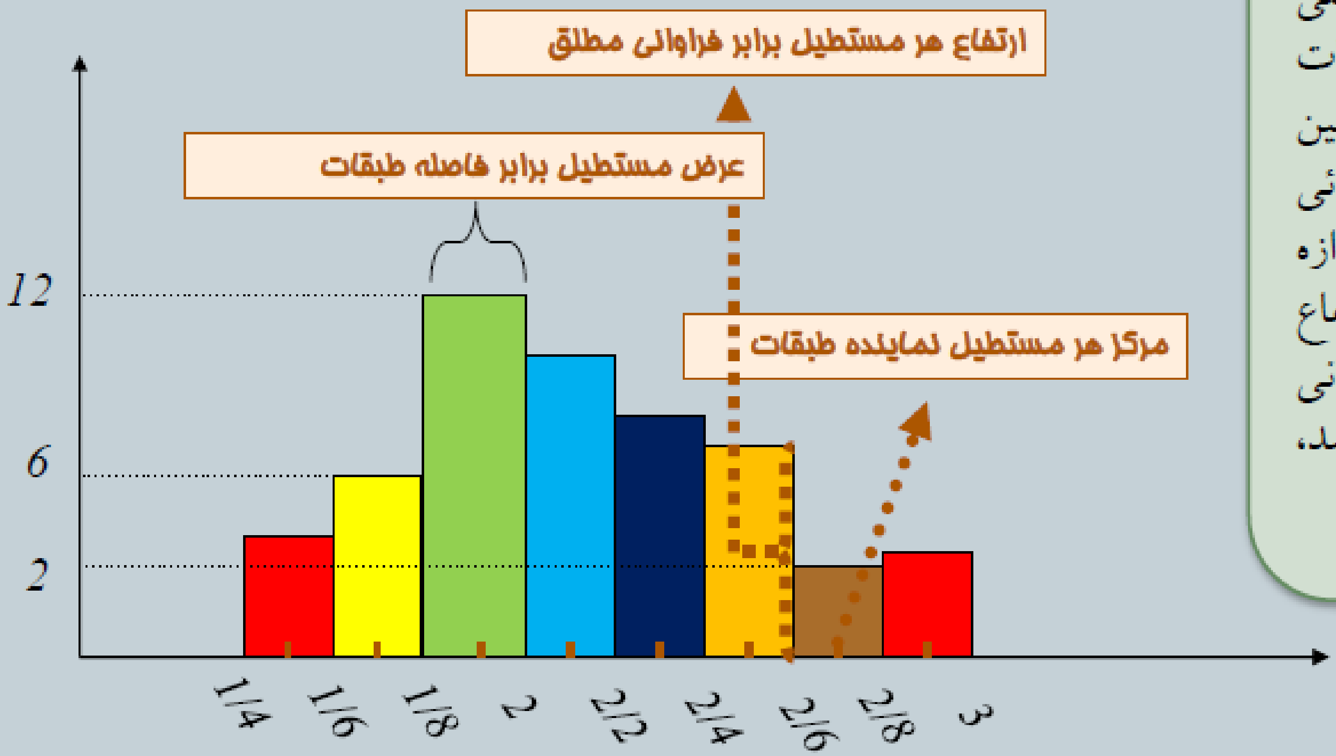
قاعده ستونها ميتواند مساوي انتخاب نشود و سطح زیر ستون متناسب با فراواني آن گروه است.

معمولا برای داده های کمی پیوسته استفاده می شود

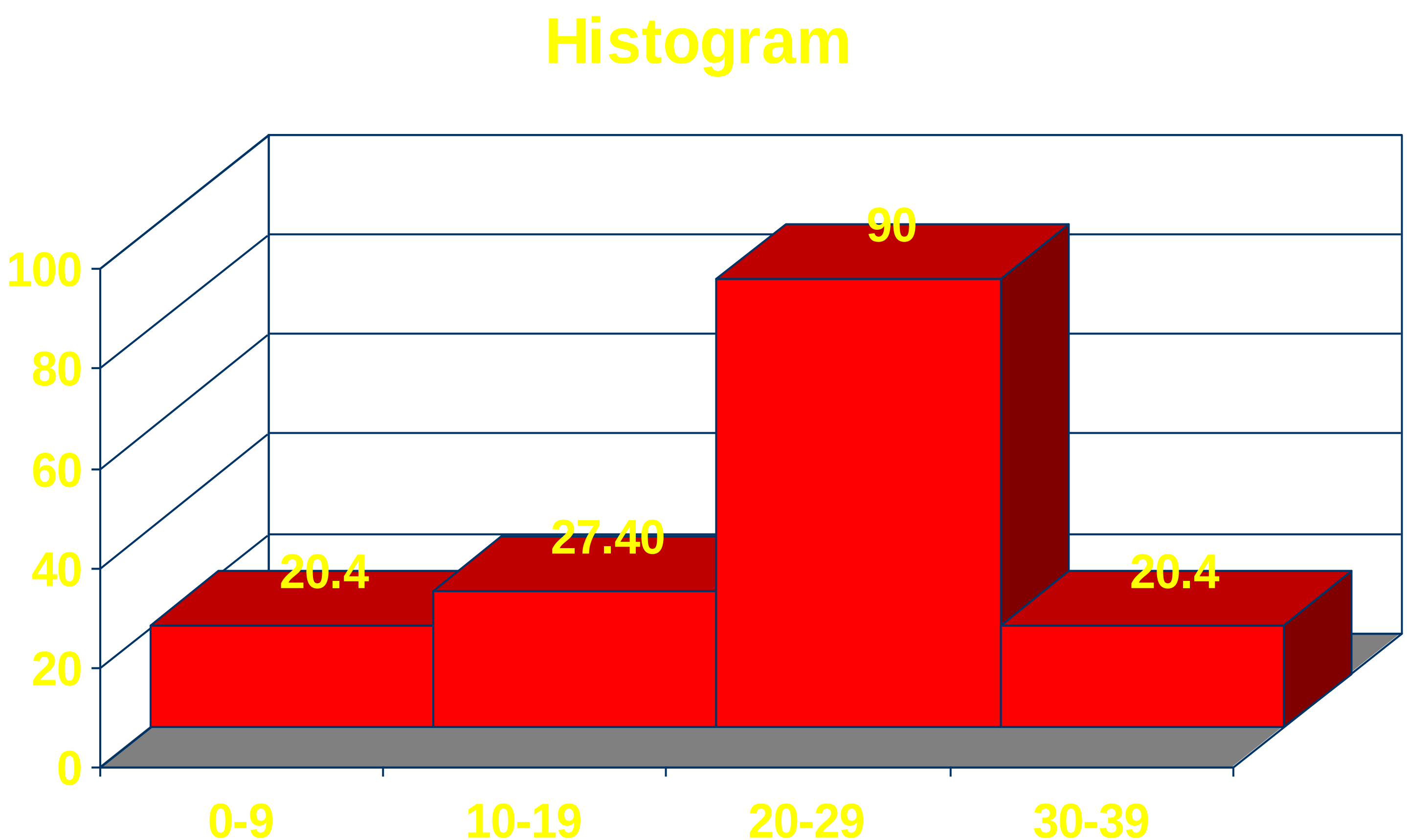
مثال

روش رسم نمودار مستطیلی

برای رسم هیستوگرام فراوانی یا نمودار مستطیلی داده ها ابتدا حدود طبقات را روی محور افقی تعیین می‌کنیم. بعد مستطیل هائی که عرض آنها به اندازه فاصله طبقات و ارتفاع آنها به اندازه فراوانی مطلق طبقه مربوطه باشد، رسم می‌کنیم.

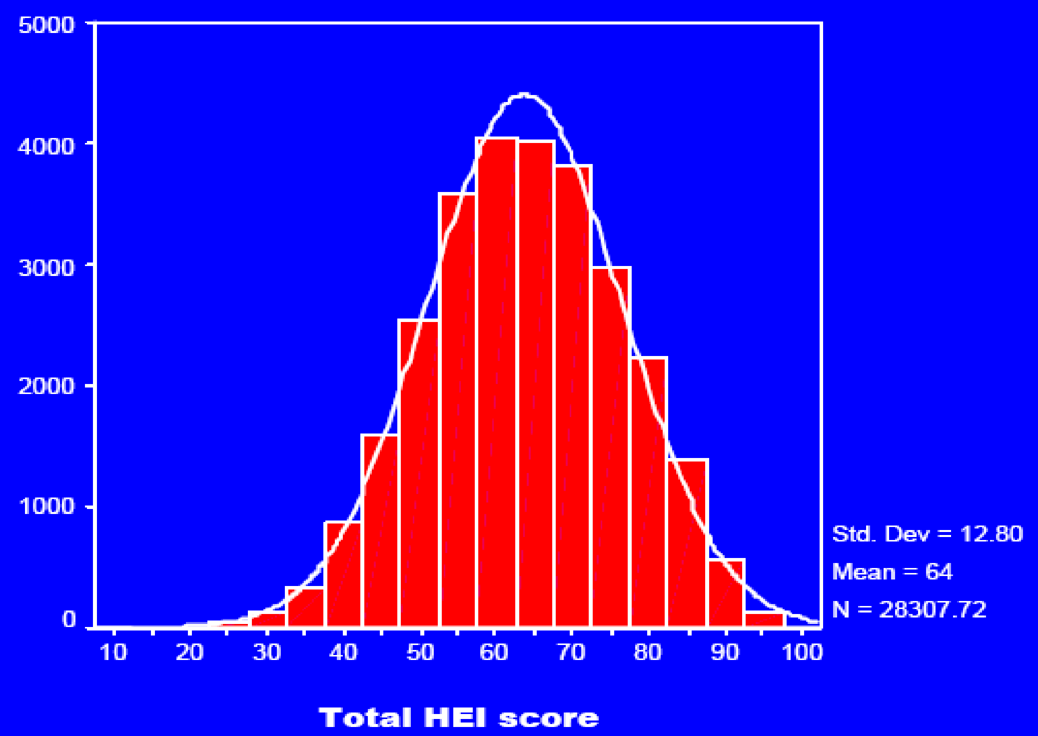


مثال



مثال

Example 3 of a Histogram



نمودار چند ضلعی

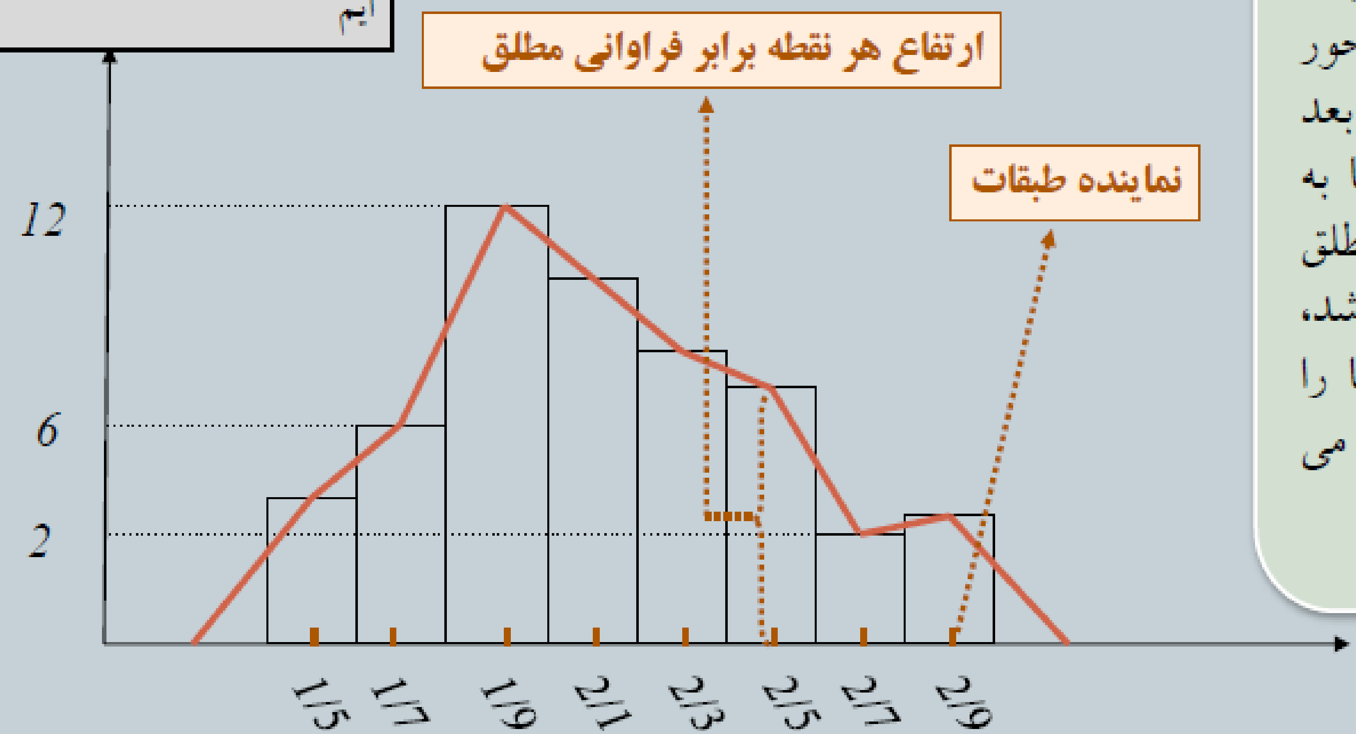
نموداریست که متناظر با هر نماینده طبقه در محور افقی و فراوانی آن در محور عمودی ، یک نقطه در صفحه مختصات ایجاد و به هم وصل می شوند

معمولا برای داده های کمی پیوسته استفاده می شود

مثال

روش رسم نمودار چند ضلعی

مستطیل ها را روی این نمودار برای درک بهتر نمودار چند ضلعی قرار داده ایم



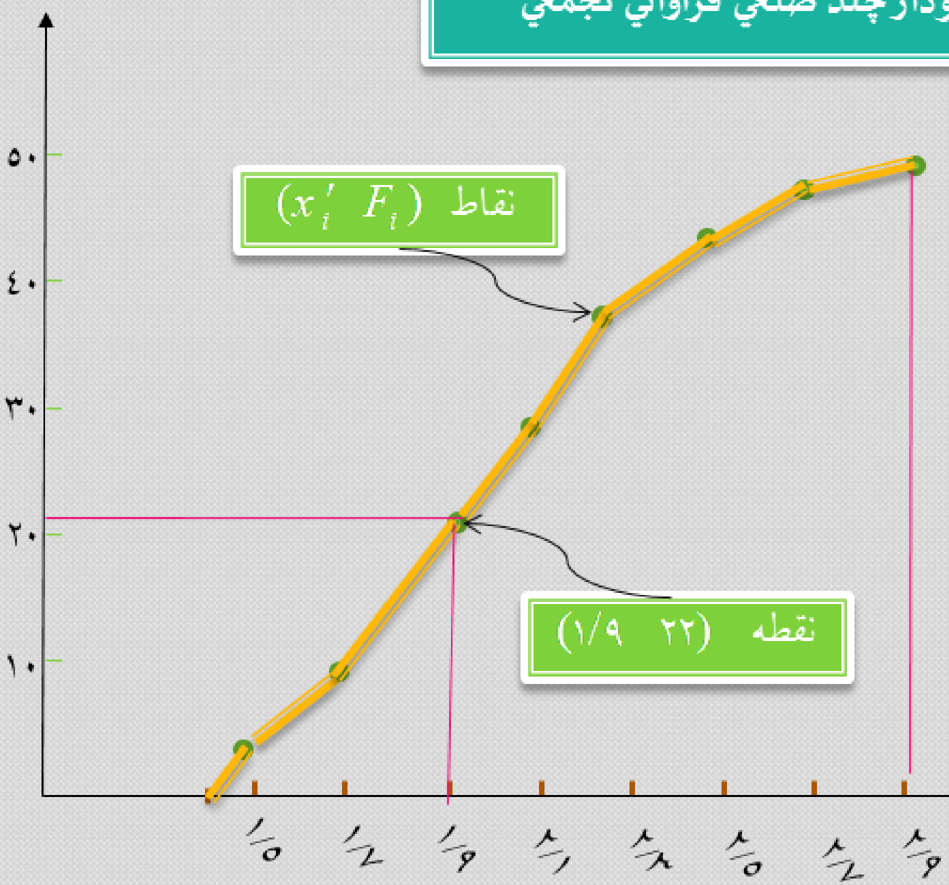
برای رسم چند ضلعی فراوانی داده ها نماینده طبقات را روی محور افقی تعیین می‌کنیم. بعد نقاطی که ارتفاع آنها به اندازه فراوانی مطلق طبقه مربوطه باشد، مشخص کرده و آنها را به یکدیگر متصل می کنیم.

چند ضلعی فراوانی تجمعی

برای ترسیم این نمودار ، از نماینده طبقات در محور افقی و فراوانی تجمعی در محور عمودی استفاده می شود ، سپس نقاط ایجاد شده به ترتیب به هم وصل می شوند.

مثال

روش رسم نمودار چند ضلعي فراواني تجمعي



	F_i
١/٤ – ١/٦	٤
١/٦ – ١/٨	١٠
١/٨ – ٢	٢٢
٢ – ٢/٢	٣١
٢/٢ – ٢/٤	٣٩
٢/٤ – ٢/٦	٤٥
٢/٦ – ٢/٨	٤٧
٢/٨ – ٣	٥٠

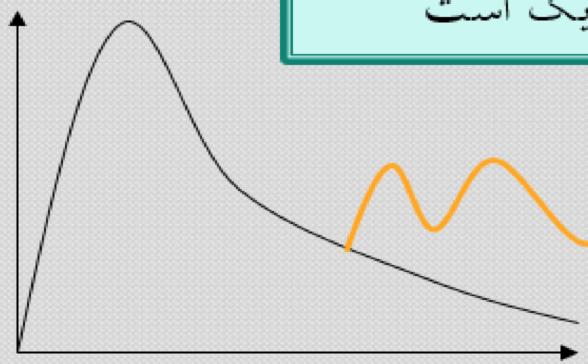
برای رسم چند ضلعی فراوانی تجمعی داده‌ها نماینده طبقات را روی محور افقی تعیین می‌کنیم. بعد نقاطی که ارتفاع آنها به اندازه فراوانی تجمعی طبقه مربوطه باشد، مشخص کرده و آنها را به یکدیگر متصل می‌کنیم.

منحنی فراوانی تجمعی

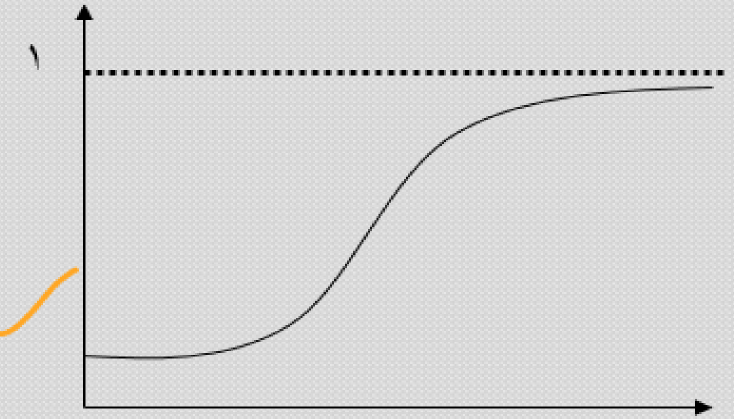
تنها فرق این نمودار با نمودار پلی گون فراوانی تجمعی در این است که در این نمودار بجای نماینده طبقات از حد بالای کرانه ها استفاده می شود

مثال

- همواره نمودار فراوانی داده ها، بالای محور یعنی در سمت مثبت است.
- میانه مساحت منحنی فراوانی را به دو قسمت مساوی تبدیل می کند.
- نمودار منحنی فراوانی تجمعی غیر نزولی و بین صفر تا یک است



نمودار منحنی فراوانی



نمودار منحنی فراوانی تجمعی

کاربردهای نمودار فراوانی تجمعی

1- برای محاسبه چندکها (چارکها ، دهکها ، صدکها)

2- برای مقایسه پدیده ها (مثل میزان رشد تورم در کشورها)

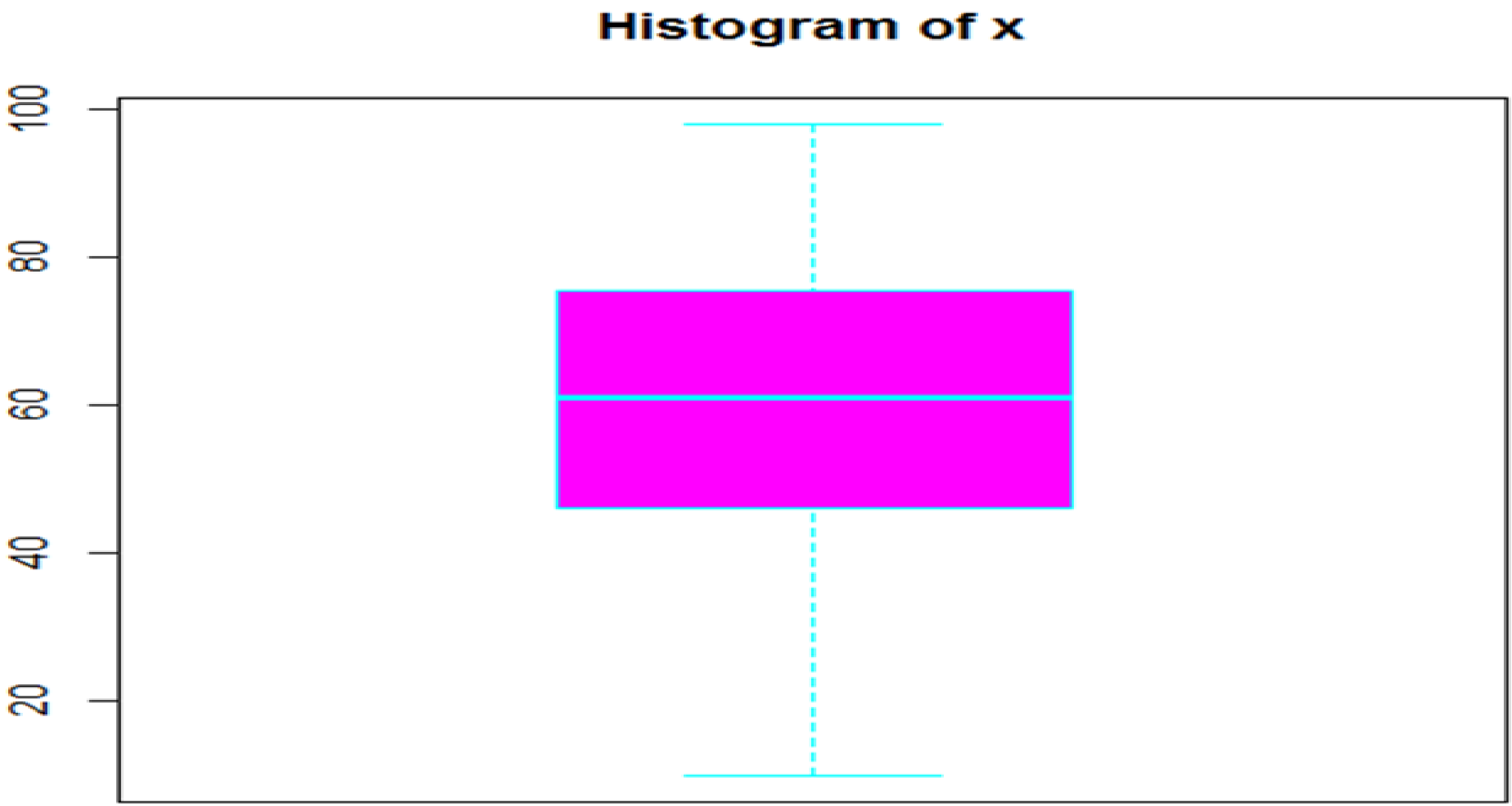
نمودار جعبه ای

این نمودار نشان دهنده چارکها و حداقل و حداکثر مشاهدات است و

برای مقایسه دو یا چند جامعه آماری مورد استفاده قرار می گیرد

معمولا برای داده های کمی پیوسته استفاده می شود

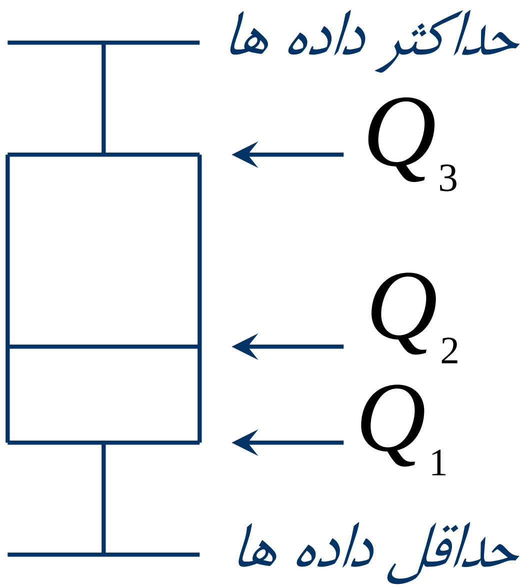
نمودار جعبه ای



مراحل تهیه نمودار جعبه ای

الف ـ پیدا کردن حداقل و حداکثر داده ها

ب ـ پیدا کردن چارکهای اول ، دوم و سوم



محورهای نمودار پاره تو

این نمودار دارای سه محور است :

1- محورافقی : نوع موضوعات

2- محور عمودی : فراوانی مطلق موضوعات

3- محور سوم (روبروی محور عمودی) :

فراوانی نسبی تجمعی موضوعات

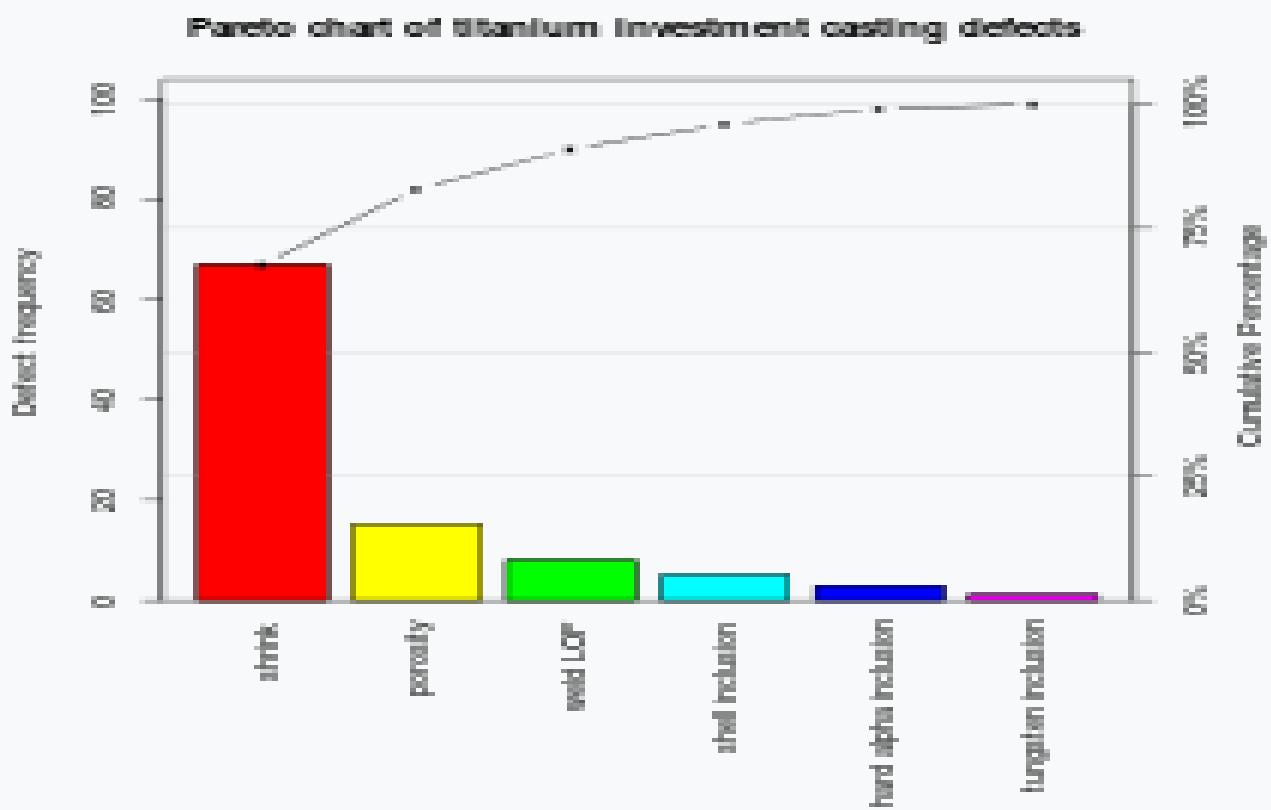
مفهوم نزولی بودن نمودار پاره تو

نموداری است که از دو نمودار میله‌ای و نمودار خط تشکیل شده‌است.

نمودار میله‌ای که به صورت نزولی نمایش داده می‌شود، نشاندهنده تک تک مقادیر به صورت مجزاست و مقدار کل با نمودار خط معرفی می‌شوداین نمودار در فرایند کنترل کیفیت اغلب نشان دهنده عوامل وقوع شایع‌ترین نقایص یا دلیل شایع‌ترین شکایات مشتریان است.

مثال

نمودار بارتو



شاخص

عدد يا نسبتي است که جهت خلاصه کردن وقايع استخراج مي شود وبعنوان مشخص کننده آن وقايع در زمان معين و يا براي مقايسه آنـها در زمانهاي مختلف بکار مي رود.

انواع شاخص

شاخص های مرکزی

شاخص هایی هستند که میزان گرایش به مرکز داده ها را اندازه می گیرند.

شاخص های پراکندگی

شاخص هایی هستند که میزان پراکندگی داده ها از مرکز را اندازه می گیرند

توجه: بسیاری از شاخص های آماری برای داده های کیفی تعریف نمی شوند

شاخص های مرکزی

- میانگین
- میانه
- نما
- چندک ها

شاخص های پراکندگی

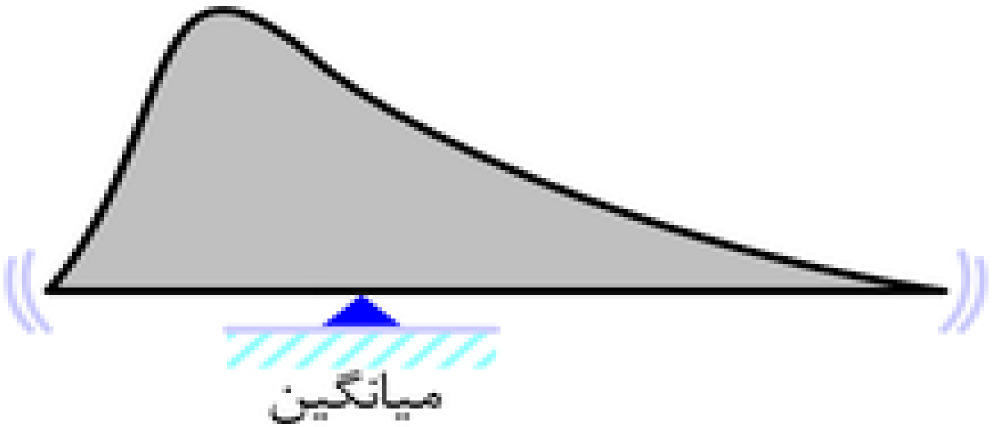
- واریانس
- انحراف معیار
- دامنه تغییرات
- ضریب تغییرات
- ضریب چولگی

پارامتر مرکزی

به هر معیار عددی که معرف مرکز مجموعه داده ها باشد ، پارامتر مرکزی اطلاق می شود یعنی همان مقدار نماینده ای که مشاهدات در اطراف آن توزیع شده اند.

میانگین

به نقطه تعادل یا مرکز ثقل توزیع ، در داده هایی که بصورت منظم بر روی یک محور ردیف شده باشند ، میانگین (*Mean*) اطلاق می شود



معايب ميانگين

ايرادي كه به ميانـگين وارد است تأثيرپذيري آن از داده‌هاي خيلي كوچك يا خيلي بزرگ مي‌باشد.

میانگین حسابی ساده

این میانگین از تقسیم مجموع مشاهدات بر تعداد آنها بدست می آید

فرمول

$$\mu_x = \frac{\sum_{i=1}^{n} X_i}{N}$$

Xi = مقدار هريك از متغيرهاي تصادفي

N = تعداد متغيرها

ميانگين نمونه با نماد x نشان داده مي شود .

ميانگين جامعه با نماد μ نشان داده مي شود

میانگین حسابی موزون

اگر هر یک از مشاهدات دارای تکرار باشند ، در این صورت تعداد تکرارها بعنوان وزن مشاهدات تلقی شده و آنها را با w_i نشان می دهند.

فرمول میانگین حسابی موزون

$$\mu_w = \frac{\sum_{i=1}^{k} W_i X_i}{\sum_{i=1}^{k} W_i} = \frac{\sum W_i X_i}{N}$$

میانگین هندسی ساده

از این میانگین برای محاسبه اندازه های نسبی همانند نسبت ها ، در صدها ، شاخص ها و نرخ های رشد استفاده می شود.

فرمول میانگین هندسی ساده

میانگین هندسی یک رشته عدد همانند X_1، X_2، . . .، X_N برابر

است با ریشه N ام حاصلضرب آن اعداد

$$\mu_G = (X_1 \times X_2 \times \ldots \times X_N)^{\frac{1}{N}}$$

میانگین هارمونیک

از این نوع ، برای محاسبه میانگین مشاهداتی استفاده می شود که از مقیاس های ترکیبی همانند « کیلو در ساعت » یا « دور در ثانیه » برخوردار هستند

فرمول میانگین هارمونیک ساده

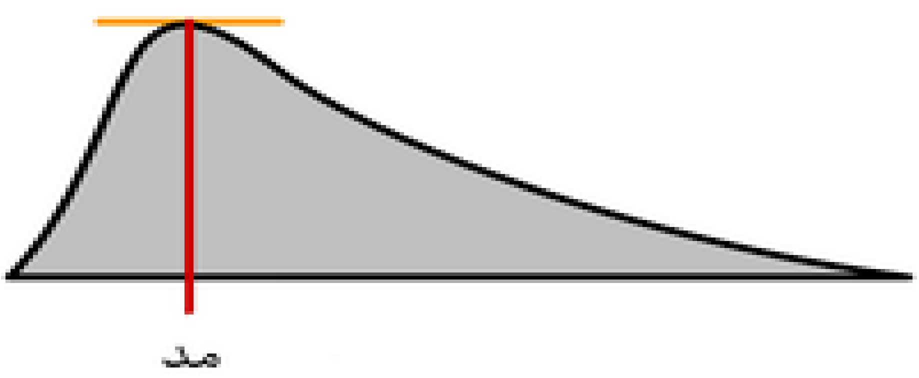
این میانگین برای چند اندازه یا مقدار برابر است با عکس میانگین حسابی معکوس آن اندازه ها

فرمول

$$\mu_H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_N}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$$

مد (نما)

به مقداری گفته می شود که در میان سایر مقادیر توزیع ، بیشترین تکرار را داشته باشد ، مد را با *Mo* نشان می دهند



مثال

مثلاً مي‌خواهيم شاخص رشته تحصيلي دوره دبيرستان را براي دانشجويان اين كلاس محاسبه كنيم. فرض مي‌كنيم بدين صورت باشد :

انساني , رياضي ,كامپيوتر ,نقشه‌كشي , هنر , رياضي ,تجربي , رياضي ,كامپيوتر , رياضي

در اين مثال ديگر اعداد و ارقامي وجود ندارد كه بتوانيم از ميانگين يا ميانه استفاده كنيم. بنابراين از شاخص مُد استفاده كرده و رشته رياضي را شاخص انتخابي كلاس قرار مي‌دهيم.

دراپیدمیولوژی بمنظورمبارزه ویا پیشگیری علیه یک بیماری، شناخت سنی که دارای بیشترین فرآوانی است(نما) برمیانگین ومیانه ارجحیت دارد.

چارک

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

انواع چارک ها

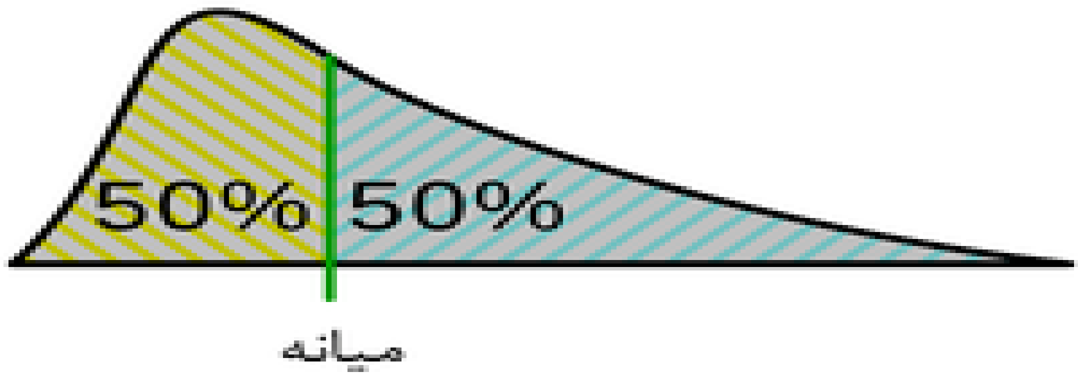
Q_1 : مقداری که 25% مشاهدات ، پایین تر از آن است

Q_2 : مقداری که 50% مشاهدات ، پایین تر از آن است

Q_3 : مقداری که 75% مشاهدات ، پایین تر از آن است

میانه

برابر است با مقداري كه نيمي از افراد، از نظرداشتن آن متغير، از آن بزرگتر و نيم ديگر از آن كوچكتر باشند.



مثال

مثال: میانه مجموعه داده‌های زیر را بدست می آوریم.

۳ ۰ ۳ ۵ ۱ ۰ ۷ ۲ ۱ ۳ ۲ ۱ ۴ ۳ ۶

ابتدا داده‌ها را مرتب می کنیم.

۰ ۰ ۱ ۱ ۱ ۲ ۲ ۳ ۳ ۳ ۳ ۴ ۵ ۶ ۷

چون تعداد داده ها ۱۵و عددی فرد است، داده هشتم میانه است زیرا:

$$۱۵ \times \frac{۱}{۲} = ۷/۵ \cong ۸ \rightarrow Md = ۳$$

موارد خاص

در موارد خاص مثلاً زماني که مي‌خواهيم سطح درآمد و يا سطح مصرف را در جامعه اي که اختلاف طبقاتي زياد است (به گونه‌اي که عده اي درآمد کلان و اکثريت درآمد محدود دارند) تعيين کنيم، ميانه شاخص مناسبتري خواهد بود. چون ميانه درآمدي را که نصف مردم کمتر و نيم ديگر بيشتر از آن را دارند مشخص مي کند ولي ميانگين چنانکه درآمدهاي کلان مقادير بسيار افراطي را شامل شود نمي‌تواند شاخص مناسبي از وضعيت درآمد باشد.

غالب موارد

درمطالعات بیولوژیک معمولاً توزیع نتیجه مطالعات،نسبتاً متقارن می باشد

ودرنتیجه میانگین ومیانه هردوشاخص های خوبی جهت نشان دادن مرکزتوزیع

می باشند،ولی بااین وجوددرغالب مواردمیانگین بعنوان شاخص مرکزی استفاده

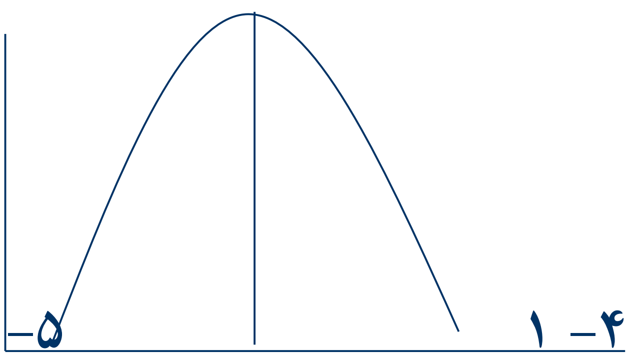
می شودزیرا تعبیروتفسیر اطلاعات وانجام آزمونهای آماری بوسیله میانگین

آسانتروقابل اعتمادترازمیانه است وبعلاوه میانگین ازاندازه همه افراد موردمطالعه

متاثر است درحالی که درمیانه این خاصیت وجودندارد.

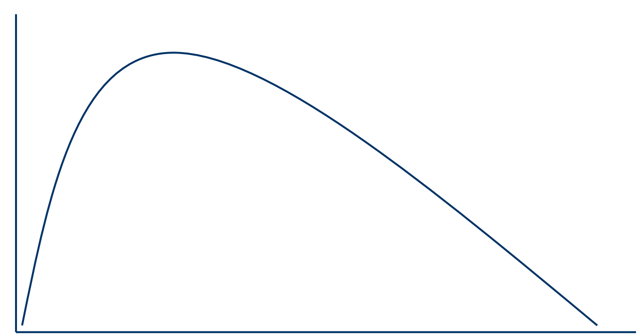
توزیع متقارن

درتوزیع متقارن (شکل5- 1) اندازه میانگین ومیانه برابر است ولی درتوزیع نامتقارن(شکل4- 1) بسته به درجه عدم تقارن توزیع ممکن است اختلاف قابل ملاحظه ای بین میانگین ومیانه مشاهده شود.



۵–۱شکل

شکل۴– ۱



تذکر

اگرچه شاخص های مرکزی مهمترین مشخص کننده برای یک توزیع می باشند،ولی بسیاراتفاق می افتد که باوجود یکسان بودن مشخص کننده های مرکزی،بین دوتوزیع تفاوت های اساسی وجوددارد.

بدین منظورشاخص های مهم پراکندگی مورد بررسی قرار می گیرد.

پارامترهای پراکندگی

شاخص هایی هستند که متوسط میزان دوری و نزدیکی داده های توزیع را نسبت به میانگین شان نشان می دهند.

محاسن پارامترهای پراکندگی

1- کمک به توصیف واقعی تر یک سری از داده ها

2- کمک به قابلیت مقایسه دو یا چند سری از داده ها

انواع شاخص های پراکندگی

دامنه تغییرات (R)

ساده ترین شاخص پراکندگی است و با کم کردن کوچکترین مشاهده

از بزرگترین آنها در یک سری توزیع بدست می آید

فرمول

$$R = MAX_{X_i} - MIN_{X_i}$$

معایب

چون درمحاسبات تنها ازمقدار ماکزیمم ومینیمم اندازه صفات استفاده میگردد وتغییرات صفت برای افراد داخل این دواندازه در آن موثرنیست،بنابراین نمیتواند بنحو مطلوبی بیانگر پراکندگی صفت باشد.

دامنه میان چارکی (IQR)

این شاخص ، پراکندگی داده ها را در فاصله چارک اول و چارک سوم نشان می دهد و کاری به مقادیر کوچکتر از Q_1 و بزرگتر Q_3 ندارد

فرمول دامنه میان چارکی

برای محاسبه این شاخص ، کافیست که مقادیر Q_1 و Q_3 را بدست آورده و از هم کم کنیم .

$$IQR = Q_3 - Q_1$$

نیمه میان چارکی

برای بدست آوردن این شاخص ، که به انحراف چارکی نیز معروف است ، کافیست ، مقدار دامنه میان چارکی را بر عدد 2 تقسیم نماییم

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

شاخص های مناسب برای توزیع های نا متقارن

1- استفاده از میانه بعنوان بهترین شاخص مرکزی

2- استفاده از انحراف چارکی بعنوان بهترین شاخص پراکندگی

انحراف متوسط از میانگین

این شاخص از تقسیم مجموع قدر مطلق انحرافات تک تک مشاهدات از میانگین شان بر تعداد مشاهدات بدست

$$A \cdot D_{\mu} = \frac{\sum \left| X_i - \mu_X \right|}{N}$$

$A.D_\mu$ محاسن و معایب

محاسن : در نظر گرفتن تغییرات کل داده ها

معایب : 1- نشان ندادن تأثیر انحرافات بزرگ

2- بی بهره بودن از بعضی از خواص مطلوب میانگین حسابی

واریانس

در این شاخص پراکندگی ، بر خلاف شاخص انحراف متوسط از میانگین

بجای قدر مطلق از مجذور (توان 2) انحرافات استفاده می شود

فرمول

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

انحراف معیار

انحراف معیار بهترین شاخص ها جهت بررسی پراکندگی داده های کمی هستند.

انحراف معیار نماینده‌ی پخش‌شدگی مقادیر متغیر تصادفی حول مقدار میانگین

ریشه دوم واریانس انحراف معیار نامیده می‌شود.

فرمول انحراف معیار

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

ویا

$$\delta_x = \sqrt{\frac{\Sigma(X_i - \mu_x)^2}{N}}$$

مقادیر نامطلوب

در داده های مربوط به سود و در آمد مقادیر کوچک تر از میانگین
ودر داده های مربوط به زیان و هزینه مقادیر بزرگتر از میانگین ،
نامطلوب قلمداد می شوند

ضریب پراکندگی(ضریب تغییرات)

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

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

$\delta_X =$ انحراف معیار مشاهدات

$\mu_X =$ میانگین مشاهدات

کاربردهای ضریب پراکندگی

برای مقایسه دو جامعه در مواردی که :

1- مقیاس ها یکسان نیستند

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

3- واریانسهای جوامع یکسان ولی میانگین هایشان متفاوت است

تذکر

بنظرمی رسد که درغالب موارد می توان ازانحراف معیار به عنوان مناسبترین شاخص پرآکندگی استفاده نمود ولی چون این کمیت ازنوع خودصفت است،درنتیجه اگرمقایسه تغییرات درصفت یا یک صفت با دوواحد مختلف باشد

مطالعه انحراف معیار به تنهایی می تواند گمراه کننده باشد.بنابراین از کمیت نسبی ضریب تغییرات که بصورت درصد بیان می شود استفاده می گردد

توزیع نرمال

یکی ازمهمترین توزیع های فرآوانی،(برای کمیتهای پیوسته وهمچنین بطورکلی) توزیع نرمال است.اهمیت این توزیع نه تنها این است که درطبیعت بسیاری ازصفات تقریباً دارای توزیع نرمال می باشند بلکه بسیاری ازروشهای آماری براساس این توزیع عرضه وحتی پاسخ بسیاری ازمسایل علمی آماربرپایه فرض نرمال بودن توزیع جامعه ، آسانتر ویا اصولاً امکان پذیر می گردد.

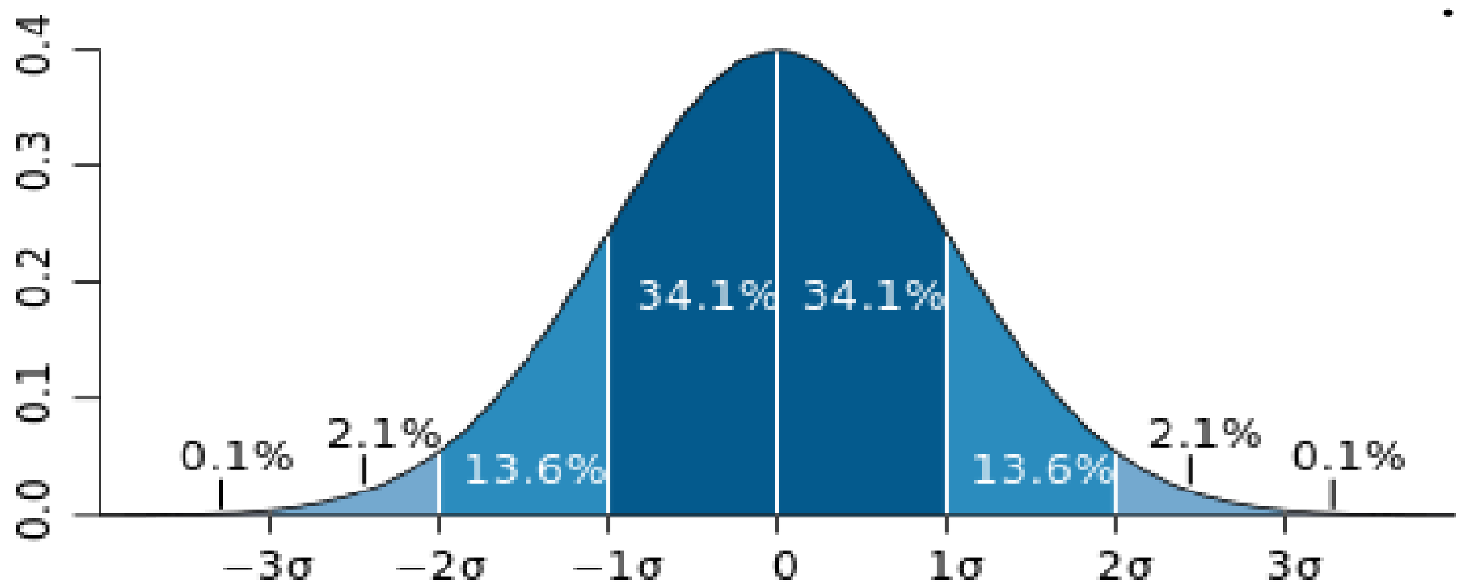
توزیع نرمال

بدیهی است هرقدرانحراف معیار توزیعی کمترباشد،تمرکزسطح زیرمنحنی بیشتر دراطراف میانگین خواهد بود.
شکل ظاهری این توزیع زنگی شکل، متقارن ودامنه تغییرات آن ازمنهای بینهایت تا مثبت بینهایت ادامه دارد ومانند هرمنحنی توزیع دیگر سطح زیرمنحنی نرمال بین دومقدارصفت، معرف فرآوانی نسبی ودرنتیجه سطح کل زیرمنحنی همواره برابر یک خواهد برد.

توزیع نرمال

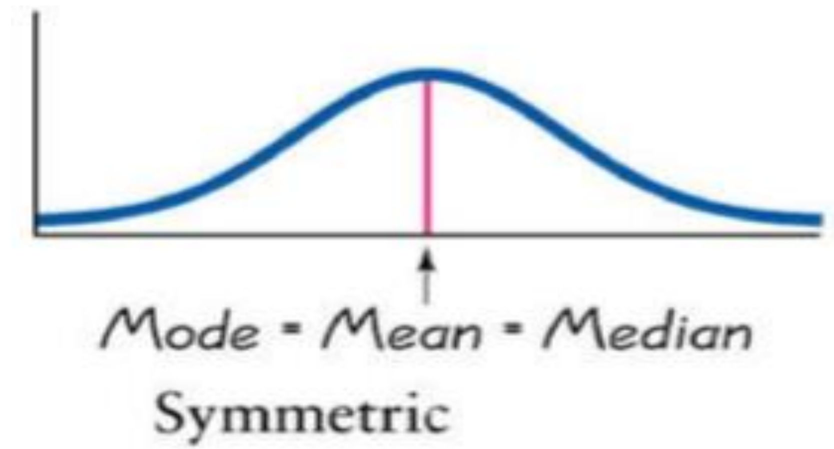
شکل زیر یک منحنی نرمال را نشان می دهد.

در یک توزیع نرمال %68.2 از داده ها به فاصله یک انحراف معیار از میانگین قرار دارند. %95.4 در فاصله دو انحراف معیار و %99.6 در فاصله 3 انحراف معیار از میانگین قرار دارند.



مقادیر مختلف ضریب چولگی(SK)

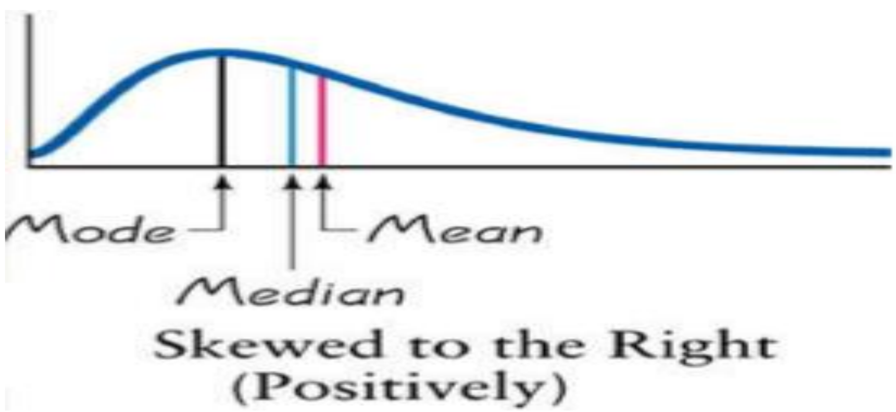
۱- صفر : در صورت متقارن بودن توزیع جامعه



متقارن (نرمال) : مد = میانه = میانگین

مقادیر مختلف ضریب چولگی(SK)

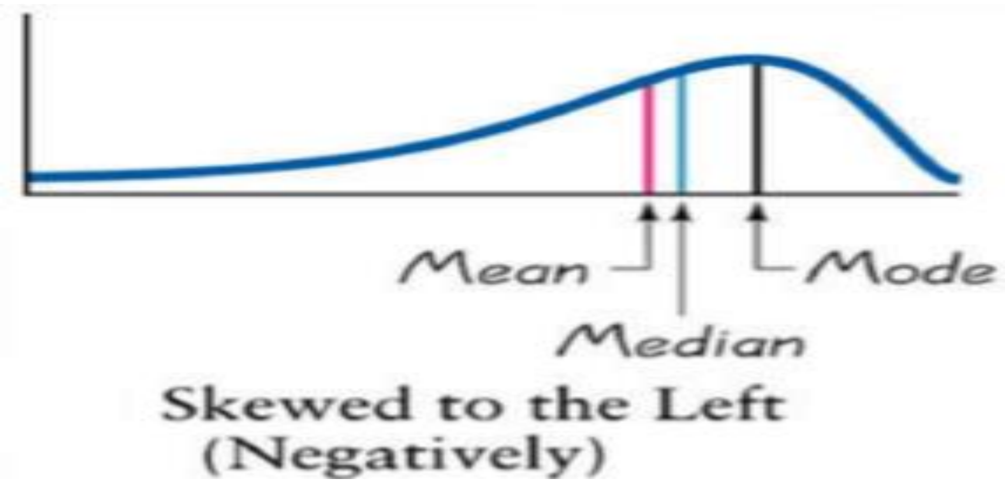
۲- مثبت : در صورت چوله به راست بودن توزیع جامعه



چوله به راست :   مد < میانه < میانگین

مقادیرمختلف ضریب چولگی(SK)

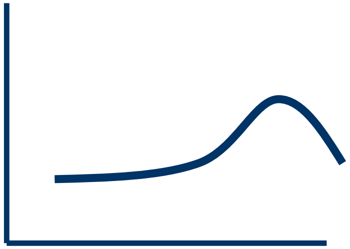
۳– منفی : در صورت چوله به چپ بودن توزیع جامعه



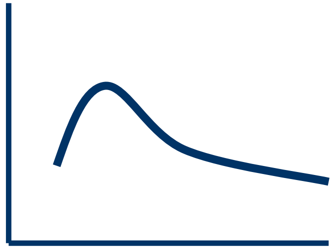
چوله به چپ : مد > میانه > میانگین

مفهوم چولگی

اگر دم توزیع جامعه به سمت راست باشد ، توزیع را چوله به راست و در صورت عکس ، آن را چوله به چپ می نامند



چوله به چپ

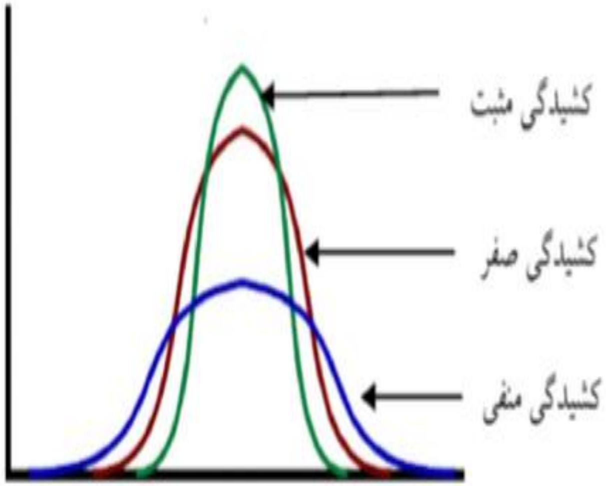


چوله به راست

پارامترهای تعیین انحراف از کشیدگی

این پارامترها برای مقایسه توزیع جوامع مورد نظر با توزیع جامعه نرمال به لحاظ کشیدگی (کوتاهی و بلندی توزیع) مورد استفاده قرار می گیرد

مقایسه انواع کشیدگی (E)



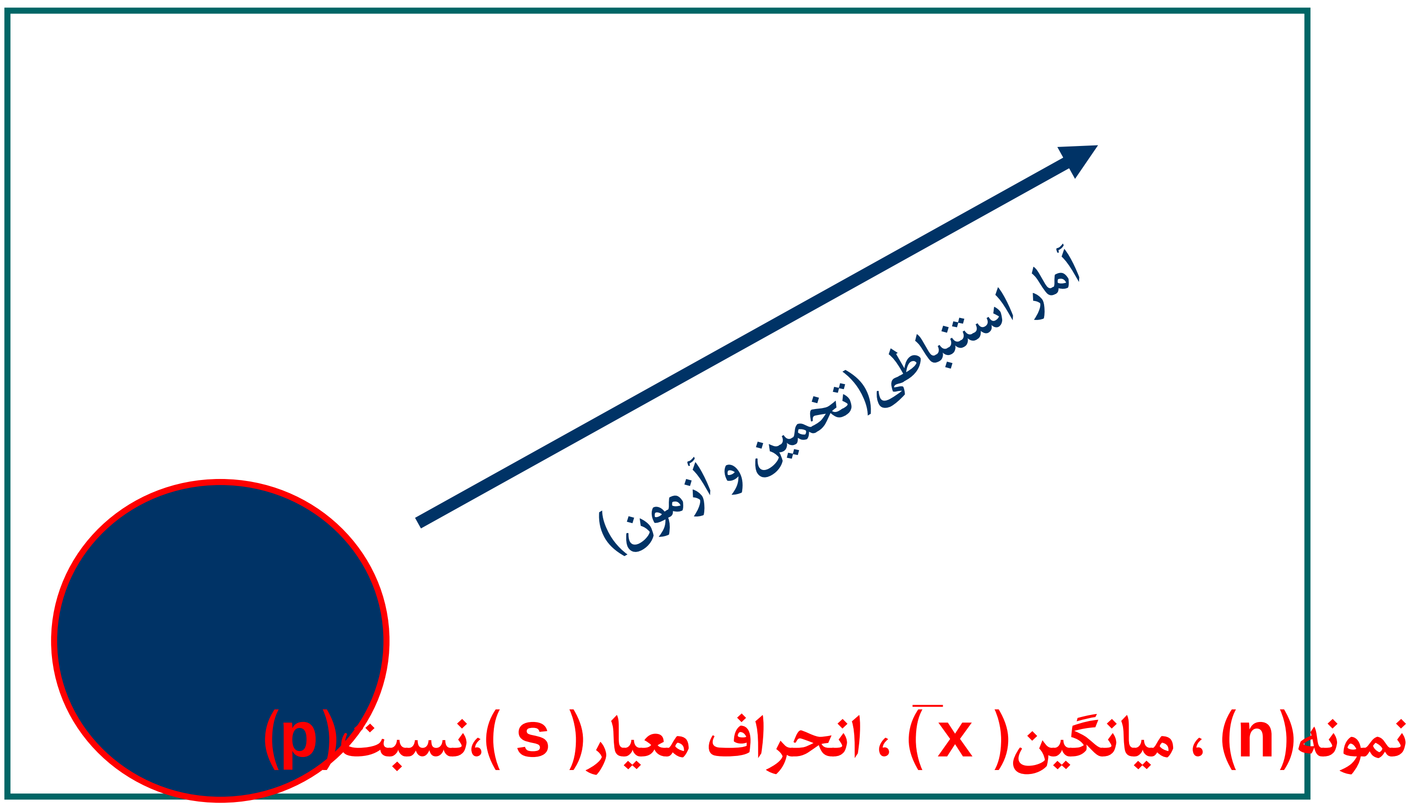
1- مساوی توزیع نرمال ($E=0$)

2- بلندتر از توزیع نرمال ($E>0$)

3- کوتاه تر از توزیع نرمال ($E<0$)

آمار استنباطی

جامعه(N) ، میانگین(μ) ، انحراف معیار(σ) ، نسبت (π)



بخش های اصلی آمار استنباطی

آمار استنباطی شامل دو بخش اصلی است:

برآورد پارامتر ها

آزمون فرض

تحقيق

اگر تحقيق از نوع سوالي و صرفا حاوي پرسش درباره پارامترها باشد براي پاسخ به سوالات از تخمين آماري استفاده مي شود و اگر حاوي فرضيه ها بوده و از مرحله سوال گذر كرده باشد آزمون فرضيه ها و فنون آماري آن بكار برده مي شود.

برآورد

یکی از مباحث آمار استنباطی موضوع برآورد است.همیشه موضوع برآورد یا تخمین با عدم قطعیت همراه است.

هرگاه از روی داده های نمونه یک مقدار عددی و یا فاصله ای شامل پارامتر به دست بیاوریم به این کار برآورد پارامتر گویند

روش برآورد

برآورد کردن به دو روش صورت می گیرد:

1-برآورد نقطه ایی

2-برآورد فاصله ای

برآورد نقطه ای

وقتی است که با استفاده از یک نمونه و توسط یک برآوردگر به نام آماره مقداری برای یک پارامتر برآورد می کنیم.

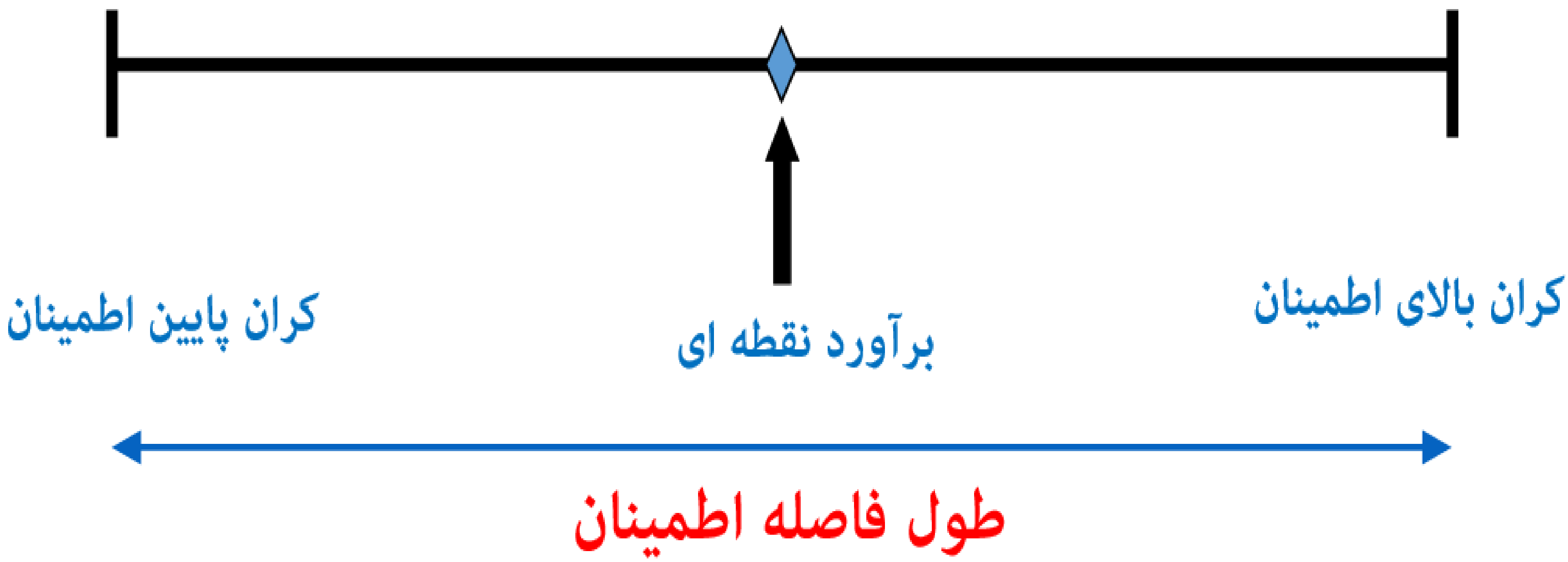
برآورد نقطه ای یک مقدار عددی است.

برآورد فاصله ای

اگر یک فاصله به دست بیاوریم طوری که انتظار داشته باشیم که پارامتر در این فاصله قرار گیرد برآورد فاصله ای گوییم.

برآورد فاصله ای اطلاعات بیشتری راجع به کمیت تحت بررسی فراهم می کند.

برآورد


کران پایین اطمینان
برآورد نقطه ای
کران بالای اطمینان
طول فاصله اطمینان

ویژگی برآورد

1-ناریب باشد(امید ریاضی آن برابربا پارامتری باشد که باید برآورد شود)

2-حداقل پراکندگی را داشته باشد(واریانس آن حداقل باشد)

برآوردکننده ناریب

برآوردکننده ای مانند $\hat{\Theta}$ را یک برآوردکنندهٔ ناریب پارامتر θ نامند اگر وتنها اگر:

$$E(\hat{\Theta}) = \theta$$

تذکر

اگر قرار باشد که بین چندین برآورد کننده ناریب یکی را انتخاب کنیم،معمولا آن برآورد کننده را انتخاب می کنیم که توزیع نمونه گیری آن دارای کمترین واریانس باشد.

خطای استاندارد میانگین نمونه

خطای استاندارد میانگین نمونه عبارت است از انحراف معیار توزیع

میانگین نمونه ها

اگر σ جامعه آماری معلوم **باشد**:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

اگر σ جامعه آماری معلوم **نباشد** (رایج):

$$s_{\bar{X}} = \frac{s}{\sqrt{n}}$$

مثال خطای استاندارد میانگین نمونه

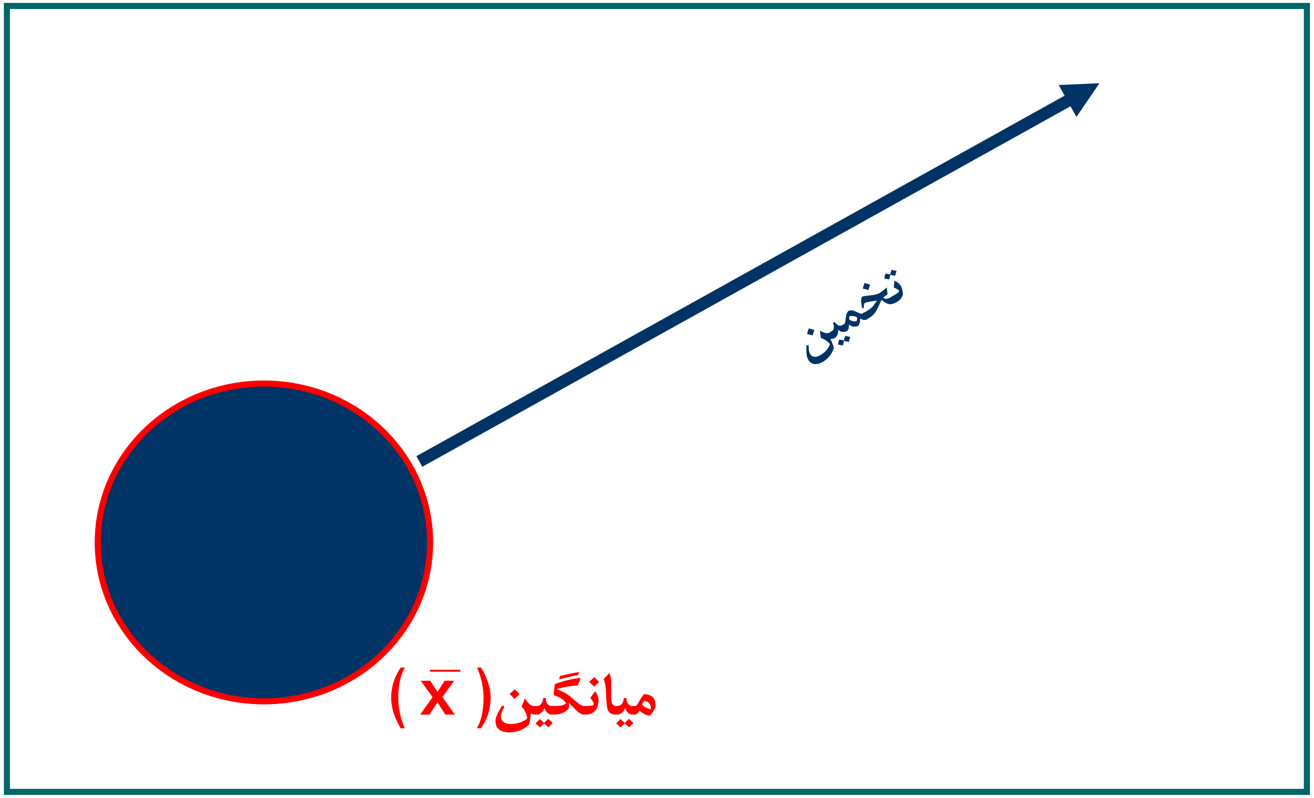
میانگین نمونه ای از 41 شرکت برابر 19 است و انحراف معیارجامعه آماری نیز برابر 6/6 می باشد. خطای استاندارد میانگین نمونه چیست؟

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6.6}{\sqrt{41}} = 1.03$$

تفسیر: برای نمونه هایی با اندازه n=41 توزیع میانگین نمونه ها دارای میانگین 19 خواهد بود وانحراف معیار این توزیع برابر1/03 است.
برآورد نقطه ای میانگین برابر 19 است.

برآورد میانگین

میانگین جامعه(μ=؟)



رابطه عمومی برآورد فاصله ای

(انحراف استاندارد)(مقدار بحرانی) ± مقدار آماره

ضریب اطمینان

در آمار استنباطی مفهوم ضریب اطمینان حائز اهمیت است.

ضریب اطمینان رایج در تحقیقات علوم پزشکی 95% است.

بطور استثناء در موارد کم اهمیت تر از ضریب اطمینان 90% و در مواردی که اهمیت زیادی دارد از ضریب اطمینان 99% استفاده می شود.

سطوح اطمینان متداول

Confidence Level	Confidence Coefficient, $1-\alpha$	$Z_{\alpha/2}$ value
80%	0.80	1.28
90%	0.90	1.645
95%	0.95	1.96
98%	0.98	2.33
99%	0.99	2.58
99.8%	0.998	3.08
99.9%	0.999	3.27

معیارهای طبقه بندی

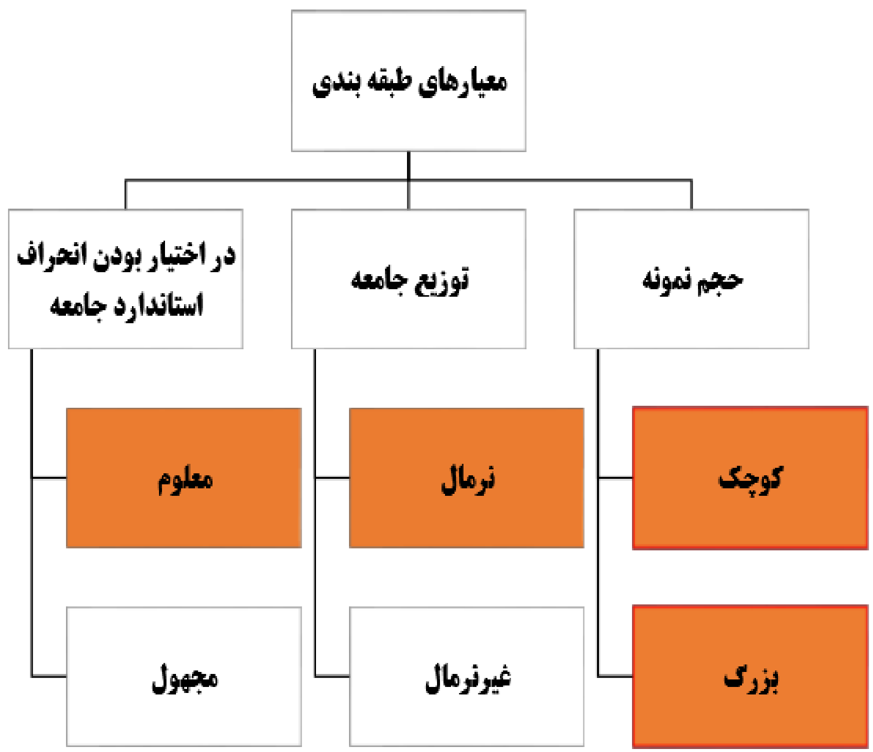
توزیع جامعه	واریانس	نمونه کوچک (n<۳۰)	نمونه بزرگ (n>۳۰)
نرمال	معلوم	آماره z	آماره z
نرمال	نامعلوم	آماره t	آماره ❋t
غیرنرمال	معلوم	ناممکن	آماره z
غیرنرمال	نامعلوم	ناممکن	آماره ❋t

❋ آماره z در این موارد قابل استفاده است اما استفاده از آماره t محافظه کارانه تر می باشد.

برآورد فاصله ای برای میانگین یک جامعه

$$\overline{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

توزیع نرمال با
واریانس معلوم



حالت اول) توزیع جامعه نرمال با انحراف معیار معلوم:

- در این حالت از فرمول زیر استفاده می شود:

$$p(\bar{x} - z_{\frac{\alpha}{2}}\sigma_{\bar{x}} \le \mu_x \le \bar{x} + z_{\frac{\alpha}{2}}\sigma_{\bar{x}}) = 1 - \alpha$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

مقدار Z از جدول بدست می آید.

سطح
اطمینان

مثال

میانگین نمونه ای از41 شرکت برابر 19 است و خطای استاندارد میانگین نمونه برابر1.03 است و توزیع آن نرمال است.

تخمین نقطه ای میانگین برابر 19 است.

فاصله اطمینان 90% عبارت است از میانگین:

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

فاصله اطمینان 90% : $z_{\alpha/2} = 1.645$

19±1.645*(1.03)

مثال

مطالعات نشان مي دهد توزيع معدل دانشجويان در واحد علوم و تحقيقات نرمال بوده و از انحراف معيار 1.5 برخوردار است. يك نمونه 25 نفري از دانشجويان رشته هاي مختلف انتخاب شده كه ميانگين نمرات آنها 13.75 مي باشد. در سطح اطمينان 95 درصد ميانگين واقعي نمرات دانشجويان را تخمين بزنيد.

مثال

$$p(\bar{x} - z_{\frac{\alpha}{2}}\sigma_{\bar{x}} \le \mu_x \le \bar{x} + z_{\frac{\alpha}{2}}\sigma_{\bar{x}}) = 1 - \alpha$$

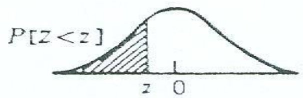
$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{1.5}{\sqrt{25}} = 0.3$$

$$z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = z_{0.025} = 1.96$$

$$p(13.75 - 1.96 * 0.3 \le \mu_x \le 13.75 + 1.96 * 0.3) = 0.95$$

$$P(13.162 \le \mu_x \le 14.338) = 0.95$$

تفسير: با اطمينان 95 درصد مي توان ادعا نمود که ميانگين نمرات دانشجويان بين 13.16و 14.33 مي باشد و فقط5 درصد احتمال دارد که ميانگين مورد نظر خارج از اين محدوده باشد. 2.5درصد احتمال دارد اين ميانگين بالاتر از 14.33بوده و 2.5درصد احتمال دارد کمتر از 13.16باشد.



جدول ۲ احتمالهای نرمال استاندارد

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
−.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

جدول z

سطح اطمینان 90% عبارت است از

مثال)

$$z_{\frac{\alpha}{2}} = z_{\frac{0.1}{2}} = z_{0.05} = 1.645$$

حالت دوم) توزيع جامعه نرمال با انحراف معيار نا معلوم

- درصورتي كه تعداد نمونه كوچكتر يا مساوي 30 باشد از فرمول زير استفاده مي كنيم:

$$p(\bar{x} - t_{\frac{\alpha}{2},df} s_{\bar{x}} \leq \mu_x \leq \bar{x} + t_{\frac{\alpha}{2},df} s_{\bar{x}}) = 1 - \alpha$$

$$df = n - 1$$

- درصورتي كه تعداد نمونه بزرگتر از 30 باشد از فرمول زير استفاده مي كنيم:

$$p(\bar{x} - z_{\frac{\alpha}{2}} s_{\bar{x}} \leq \mu_x \leq \bar{x} + z_{\frac{\alpha}{2}} s_{\bar{x}}) = 1 - \alpha$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

مثال

بازاريابي در صدد بررسي و برآورد قدرت خريد ساكنان يك محله در رشت مي باشد او ناچار بايد يك نمونه تصادفي 10 تايي از بين خريداران انتخاب و قدرت خريد هر يك را اندازه گيري كند. قدرت خريد نمونه فوق بر حسب ده هزار تومان چنين است:

$x=8,7,5,4,12,15,10,13,14,12$

قدرت خريد ساكنان محله از توزيع نرمال برخوردار است . در سطح اطمينان 95 درصد ميانگين قدرت خريد را برآورد نماييد.

مثال

$$p(\bar{x}-t_{\frac{\alpha}{2},df}s_{\bar{x}} \le \mu_x \le \bar{x}+t_{\frac{\alpha}{2},df}s_{\bar{x}}) = 1-\alpha$$
$$df = n-1$$

چون تعداد نمونه کمتر از 30 است لذا از فرمول

$$S_x = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{132}{9}} = 3.83 \quad s_{\bar{x}} = \frac{3.83}{\sqrt{10}} = 1.211 \quad t_{\frac{\alpha}{2},df} = t_{\frac{0.05}{2},9} = t_{0.025,9} = 2.262$$

$$df = n-1 = 10-1 = 9$$

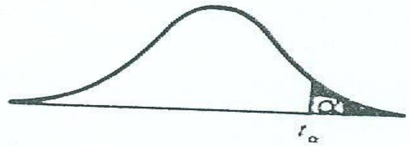
$$p(10-2.262*1.211 \le \mu_x \le 10+2.262*1.211) = 0.95$$

$$p(7.261 \le \mu_x \le 12.739) = 0.95$$

تفسیر: با اطمینان ۹۵ درصد می توان ادعا نمود که قدرت خرید ساکنان محله بین ۷۲۶۱۰تومان و ۱۲۷۳۹۰ تومان می باشد.و فقط ۵درصد احتمال دارد که قدرت خرید ساکنان محله خارج از این محدوده باشد.

جدول T

جدول ۳ سطح زیر منحنی دنبالهٔ راست توزیع t



d.f.	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	d.f.
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

From "Table of Percentage Points of the t-Distribution," computed by Maxine Merrington. *Biometrika*, Vol. 32 (1941), p. 300. Reproduced by permission of the Biometrika Trustees.

مثال)

$$t_{\frac{\alpha}{2},df} = t_{\frac{0.05}{2},9} = t_{0.025,9} = 2.262$$

حالت سوم) توزیع جامعه غیر نرمال باشد:

- درصورتی که تعداد نمونه کوچکتر یا مساوی ۳۰ باشد از فرمول زیر استفاده می کنیم:

$$p(\bar{x} - ks_{\bar{x}} \le \mu_x \le \bar{x} + ks_{\bar{x}}) \ge 1 - \alpha$$

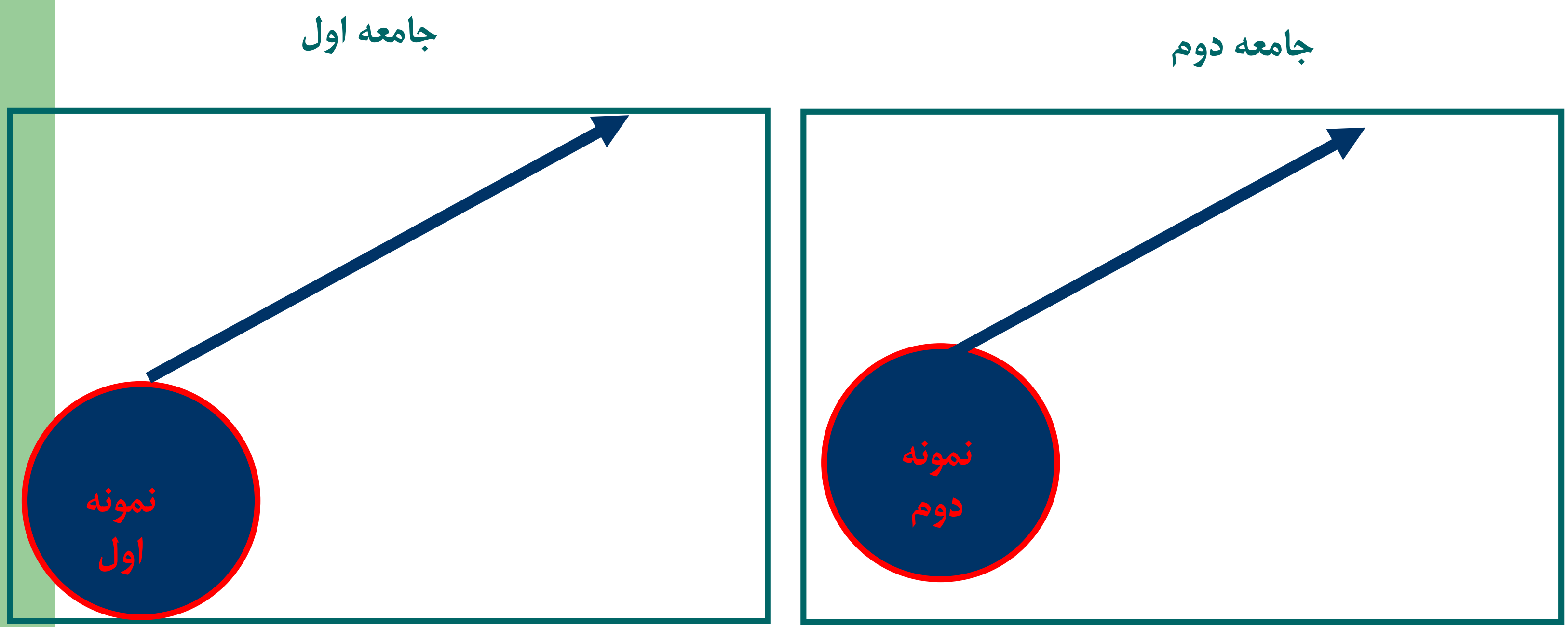
$$k = \sqrt{\frac{1}{\alpha}}$$

- درصورتی که تعداد نمونه بزرگتر از ۳۰ باشد از فرمول زیر استفاده می کنیم:

$$p(\bar{x} - z_{\frac{\alpha}{2}} s_{\bar{x}} \le \mu_x \le \bar{x} + z_{\frac{\alpha}{2}} s_{\bar{x}}) = 1 - \alpha$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

تخمین فاصله ای تفاضل میانگین دو جامعه



حالت اول) توزیع دو جامعه نرمال و انحراف معیارهای دو جامعه معلوم:

$$p\left[(\overline{x}_1-\overline{x}_2)-z_{\frac{\alpha}{2}}\sigma_{\overline{x}_1,\overline{x}_2}\le\mu_{x_1}-\mu_{x_2}\le(\overline{x}_1-\overline{x}_2)+z_{\frac{\alpha}{2}}\sigma_{\overline{x}_1,\overline{x}_2}\right]=1-\alpha$$

$$\sigma_{\overline{x}_1,\overline{x}_2}=\sqrt{\frac{\sigma_{x_1}^2}{n_1}+\frac{\sigma_{x_2}^2}{n_2}}$$

نحوه تفسیر نتایج:

- اگر هر دو دامنه مثبت باشد: $\mu 1>\mu 2$
- اگر هر دو دامنه منفی باشد: $\mu 1<\mu 2$
- در غیر اینصورت : اختلاف معنی دار نیست

حالت دوم) توزیع دو جامعه نرمال و انحراف معیارهای دو جامعه نا معلوم

۲-۱- اگر چنانچه انحراف معیار های دو جامعه برابر باشند از فرمول زیر استفاده می کنیم:

$$p\left[(\overline{x_1}-\overline{x_2})-t_{\frac{\alpha}{2},df}\,s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le \mu_{x_1-}\mu_{x_2}\le(\overline{x_1}-\overline{x_2})+t_{\frac{\alpha}{2},df}\,s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]=1-\alpha$$

$$s_p=\sqrt{\frac{(n_1-1)s_1^{\,2}+(n_2-1)s_2^{\,2}}{n_1+n_2-2}}$$

۲-۲- اگر چنانچه انحراف معیار های دو جامعه برابر نباشند از فرمول زیر استفاده می کنیم:

$$p\left[(\overline{x_1}-\overline{x_2})-z_{\frac{\alpha}{2}}\sqrt{\frac{s_{_1}^{\,2}}{n_1}+\frac{s_{_2}^{\,2}}{n_2}}\le \mu_{x_1-}\mu_{x_2}\le(\overline{x_1}-\overline{x_2})+z_{\frac{\alpha}{2}}\sqrt{\frac{s_{_1}^{\,2}}{n_1}+\frac{s_{_2}^{\,2}}{n_2}})\right]=1-\alpha$$

حالت سوم) توزیع دو جامعه مورد بررسی غیر نرمال باشد:

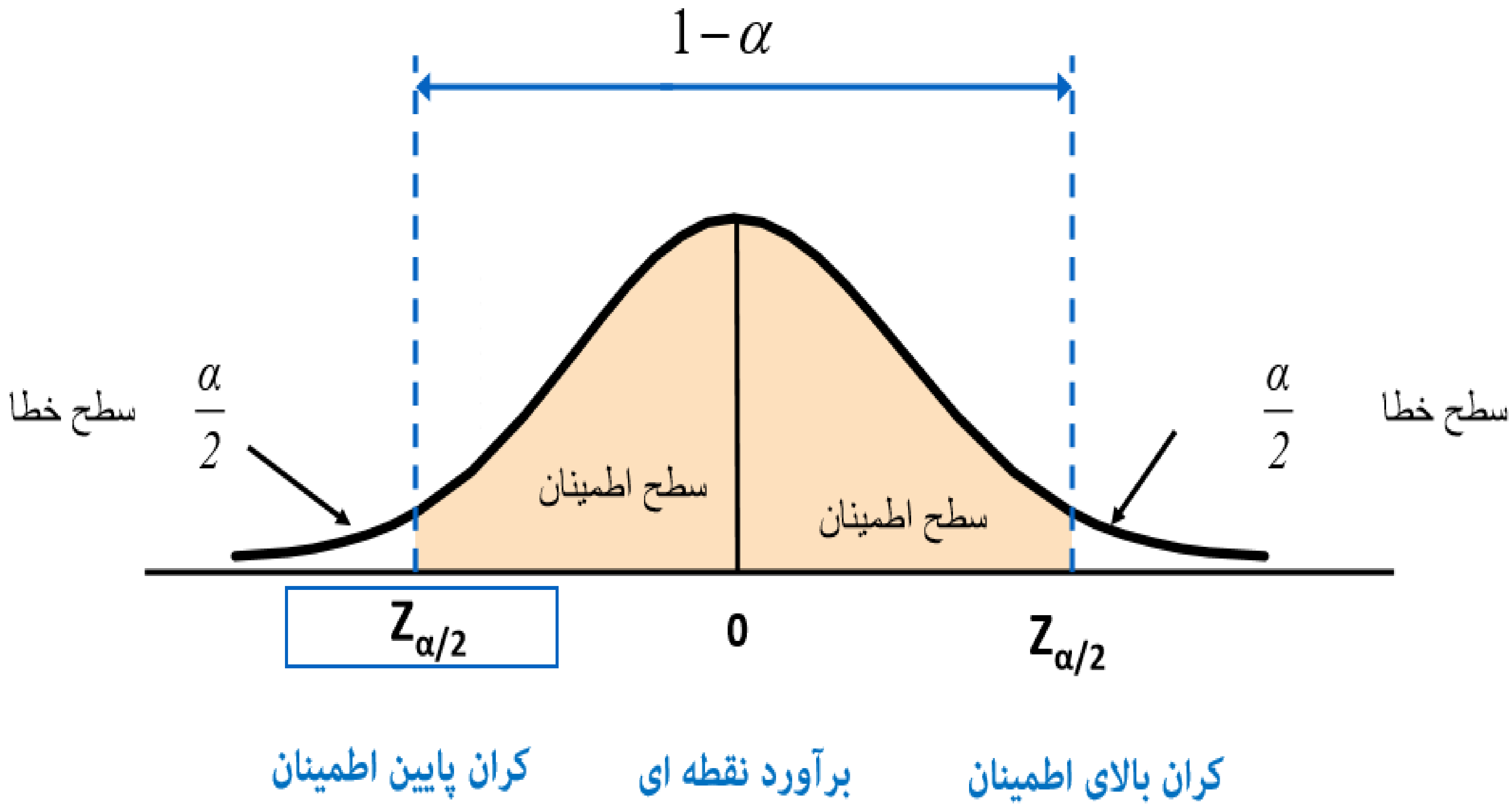
۲-۱- اگر درجه آزادی بالاتر از ۳۰ باشد از فرمول زیر استفاده می کنیم:

$$p\left[(\overline{x_1}-\overline{x_2})-z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^{\,2}}{n_1}+\frac{s_2^{\,2}}{n_2}}\leq\mu_{x_1-}\mu_{x_2}\leq(\overline{x_1}-\overline{x_2})+z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^{\,2}}{n_1}+\frac{s_2^{\,2}}{n_2}})\right]=1-\alpha$$

۲-۱- خلاف حالت فوق در دنیای عمل به ندرت اتفاق می افتد لذا بدلیل اهمیت کم از این حالت صرف نظر می شود.

- تفسیر مثل حالت دوم می باشد.

فاصله وسطح اطمینان



فرض آماری

آزمون آماری شامل آزمون فرض است

فرض آماری، حدسی در مورد پارامترجامعه یا حتی توزیع شکل جامعه آماری است

هر پاسخي كه به پرسش‌هاي يك تحقيق داده مي‌شود، يك فرضيه است. اين فرضيه ها منجر به يك يا گاهي چند آزمون آماري مي‌شود.

مثال

1. ميانگين مدت اقامت بستري در يك بيمارستان 5 روز است.
2. عامل سيگار بر سرطان ريه مؤثر است.
3. برنامه‌هاي آموزش بهداشت منجر به بهبود وضع سلامت جامعه مي‌شود.

آزمون فرض آماری

روندي است که در طي مراحل انجام آن و با کمك روش‌هاي آماري

مي‌خواهيم درستي يا نادرستي يك فرضيه (ادعا) را آزمون كنيم.

انواع فرضیه

فرضيه صفر:

و آن را با H_0 نشان میدهیم.

فرضيه يك:

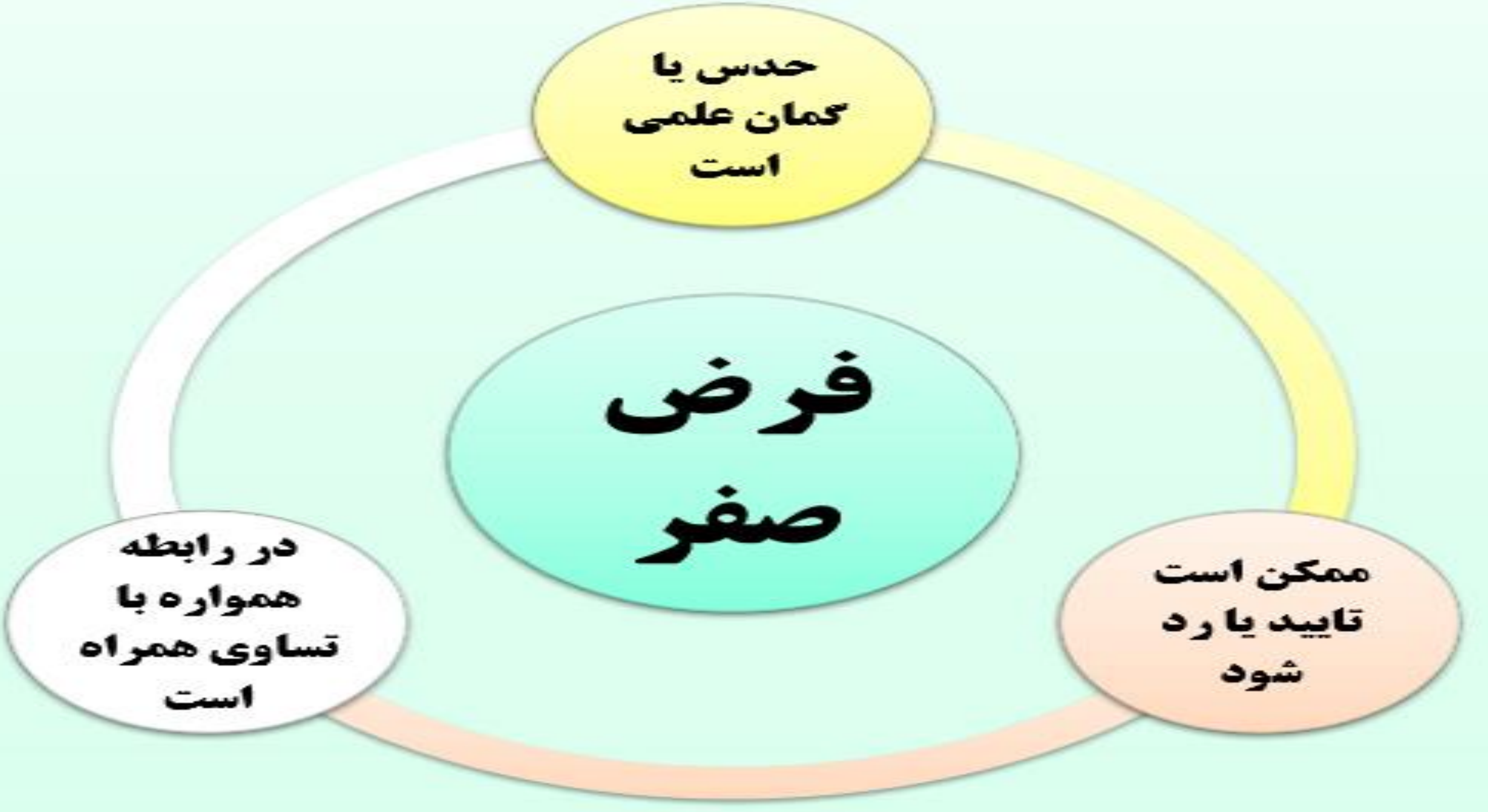
آن را با H_1 نشان مي دهيم.

فرضیه صفر

حدس اولیه برای پارامتر جامعه(هدف محقق از انجام تحقیق رد کردن این فرضیه است)

و آن را با H_0 نشان میدهیم.

فرضیه صفر


حدس یا
گمان علمی
است
فرض
صفر
ممکن است
تایید یا رد
شود
در رابطه
همواره با
تساوی همراه
است

فرضیه یک

حدس محقق برای پارامتر جامعه(هدف محقق از انجام تحقیق نشان دادن تغییر در پارامتر جامعه به کمک فرض مقابل است)

آن را با H_1 نشان مي دهیم.

مثال آزمون فرضيه

ميانگين سن مردانی که اولين بار نشانه ريزش مو را مشاهده مي‌کنند، 36 سال است.

ميانگين سن مرداني که اولين بار نشانه ريزش مو را مشاهده مي‌کنند، بيشتر از 36 سال است.

اين دو فرضيه را به صورت خلاصه زير مي‌نويسند:

$$\begin{cases} H_{۰} : \mu = ۳۶ \\ H_{۱} : \mu > ۳۶ \end{cases}$$

نمونه ای ازآزمون های فرض

آیا متوسط قد دانشجویان برابر ۱۶۷ سانتی‌متر است؟ در مقابل آن‌که میانگین قد برابر با ۱۶۷ سانتی‌متر نیست!

آیا انحراف از معیار قد دانشجویان برابر ۳۰سانتی‌متر است؟ در مقابل آن که انحراف معیار قد دانشجویان برابر با ۳۰ سانتی‌متر نیست!

ایا متوسط قد دانشجویان دختر و پسر با یکدیگر برابر است؟ در مقابل آن که متوسط قد دانشجویان دختر و پسر با یکدیگر تفاوت معنی‌داری دارد؟

نمونه ای ازآزمون های فرض

میانگین وزن نوزادان در طبقات مرفه جامعه، حداقل 3000gr است .

$$H_0 : \mu \geq 3000$$

$$H_1 : \mu < 3000$$

خطاهای آزمون

چون آزمون فرضیه بر مبنای داده های نمونه می باشد بنابراین ممکن است در تصمیم گیری دچار خطا شویم .

انواع خطا ها

1- خطاي نوع اول (α)

2- خطاي نوع دوم (β)

خطای نوع اول

احتمال رد فرض صفر است وقتي فرض صفر درست باشد که

همان سطح معني داري آزمون است.

بديهي است سطح اطمينان $1-\alpha$ خواهد بود.

خطای نوع دوم

احتمال رد فرض یك است وقتي فرض يك درست باشد(قبول

فرض صفردر صورتی که فرض یک درست است).

خطا های آزمون فرض آماری

	وضعیت واقعی جامعه	
تصمیم محقق	H_0 درست	H_0 نادرست
H_0 پذیرش	$1 - \alpha$:تصمیم درست	β: خطای نوع دوم
H_0 رد	α: خطای نوع اول	$1 - \beta$:تصمیم درست

تمرین

چه مقادیری از آلفا و بتا در یک آزمون مناسب است؟

آزمونی معتبراست که خطای نوع اول و دومش کم باشد.معمولا در تحقیقات تجربی یا شبه تجربی که احتمال ارتکاب خطای نوع اول به صورت اختیاری در دست محقق می باشد مقدار سطح معنی داری آزمون برابر 5 درصد یا 1 درصد در نظر گرفته می شود.

سطح معنا داری

در اغلب پژوهش‌ها سطح معني‌داري α را مقدار 0/1 يا 0/05 يا 0/01 در نظر مي‌گيرند.

اين مقدار را محقق قبل از انجام آزمون تعيين مي‌كند.

مراحل آزمون یک فرض آماری

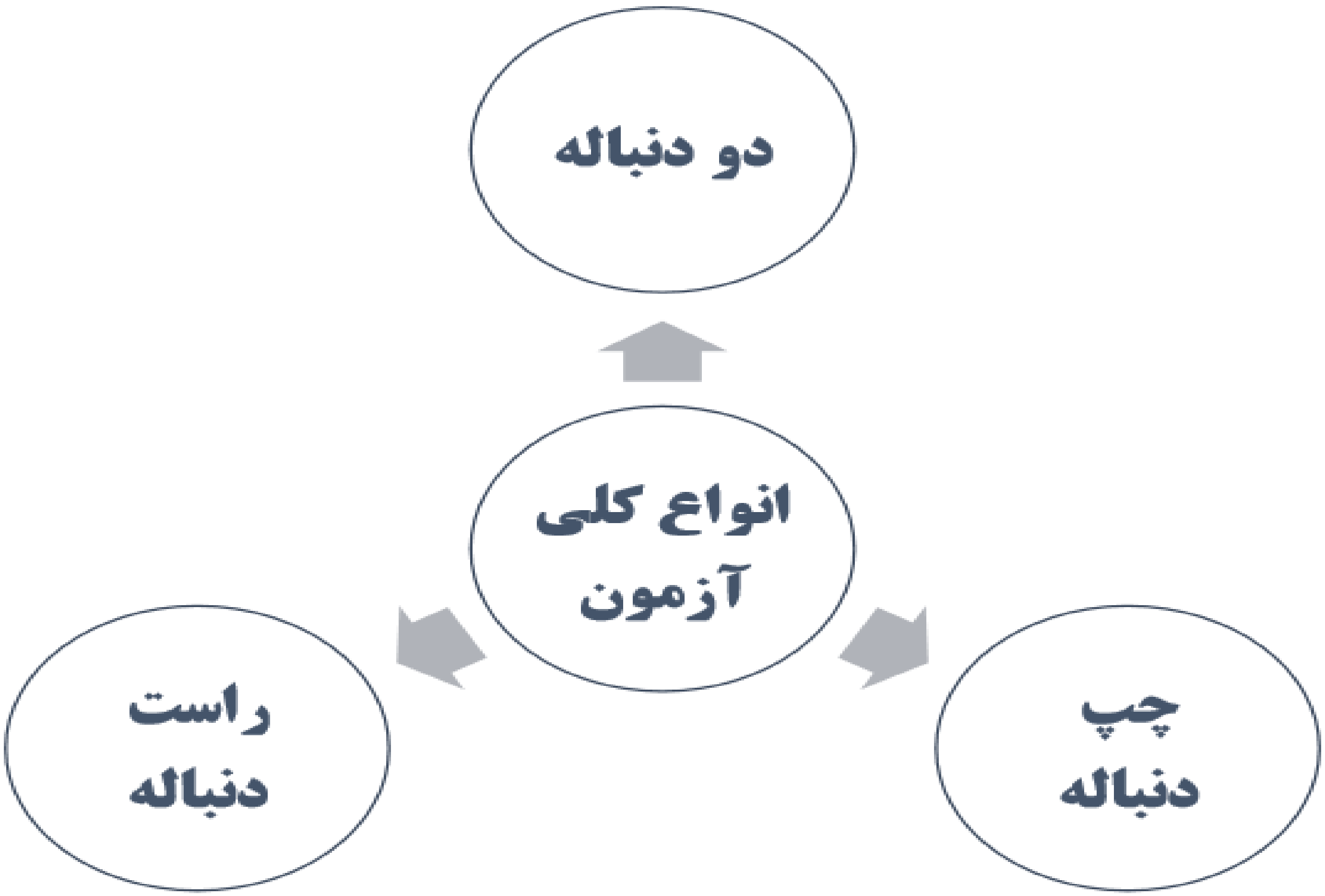
1.تعیین فرض صفر و یک آزمون

2. تعیین آماره آزمون(ملاک آزمون)

3.تعیین ناحیه بحرانی

4.مقایسه آماره آزمون با ناحیه بحرانی

طبقه بندی کلی آزمون های فرض



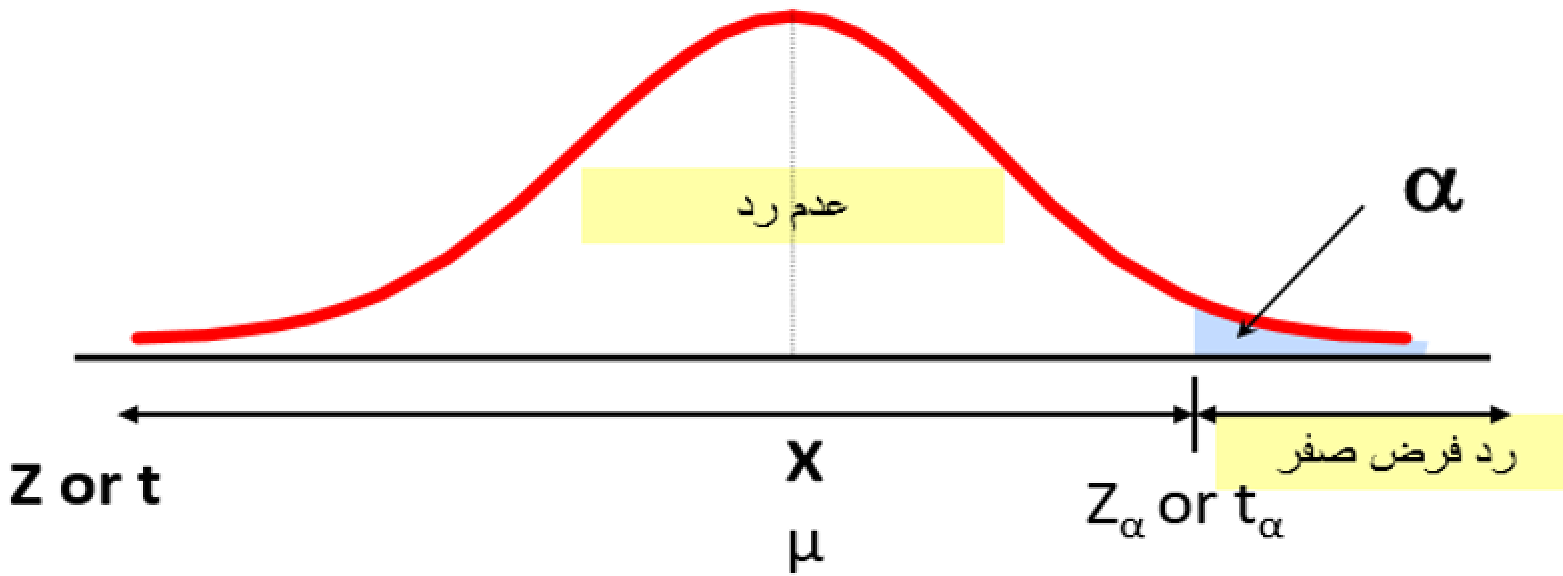
آزمون فرض

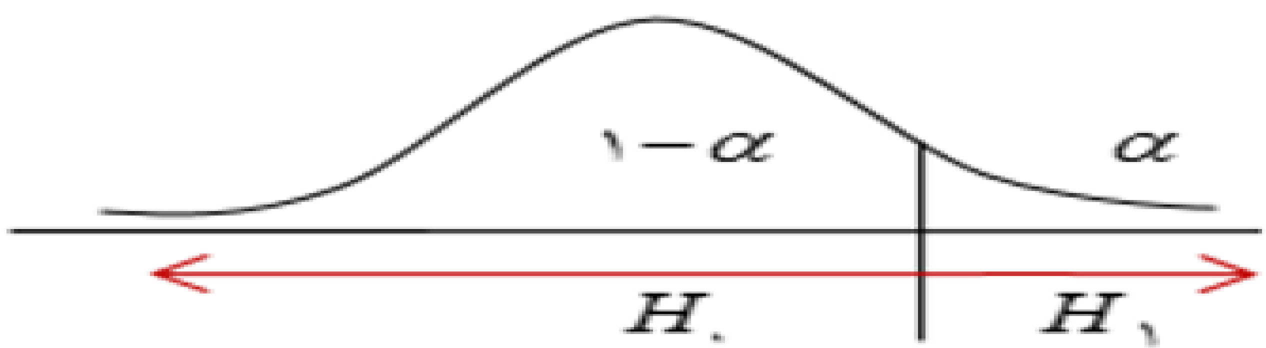
میانگین سن ازدواج در جوانان30سال است.
فردي ادعا ميكند این مقدار در سال گذشته افزایش داشته است.

آزمون يكطرفه راست

$H_0: \mu = 30 \quad H_1: \mu > 30$

منطقه پذیرش یا رد

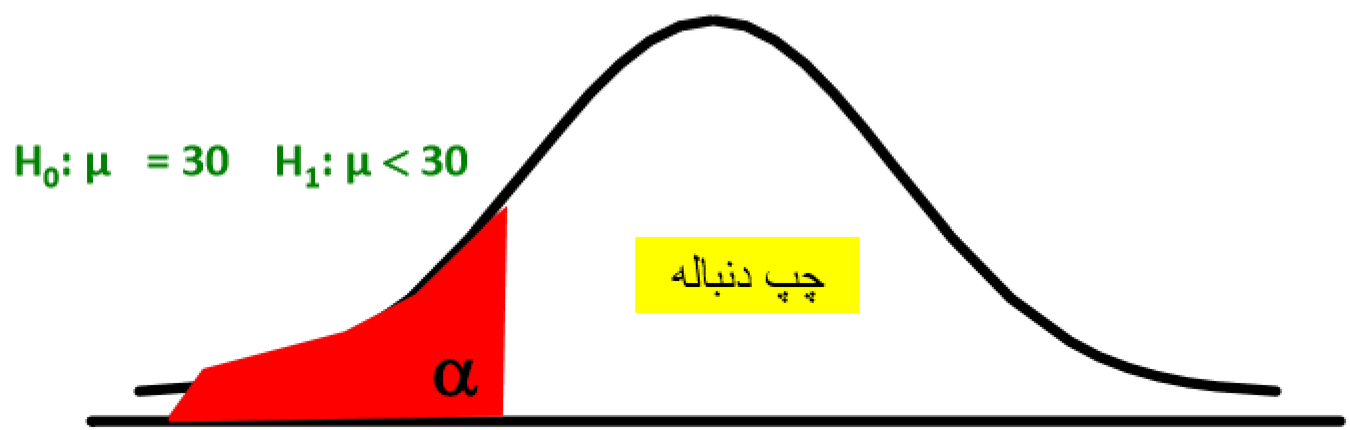




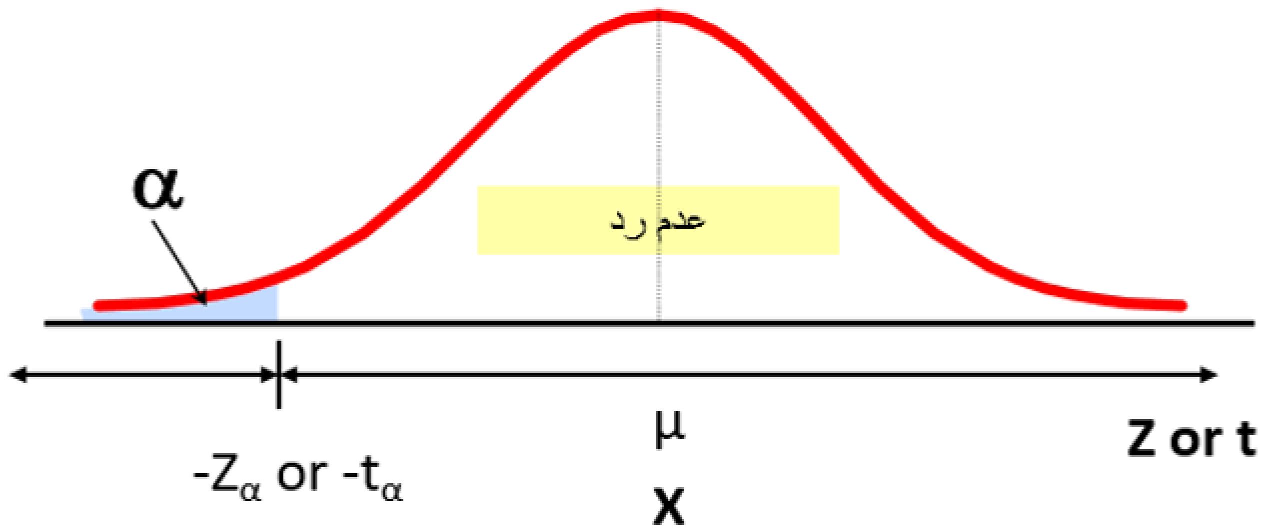
آزمون فرض

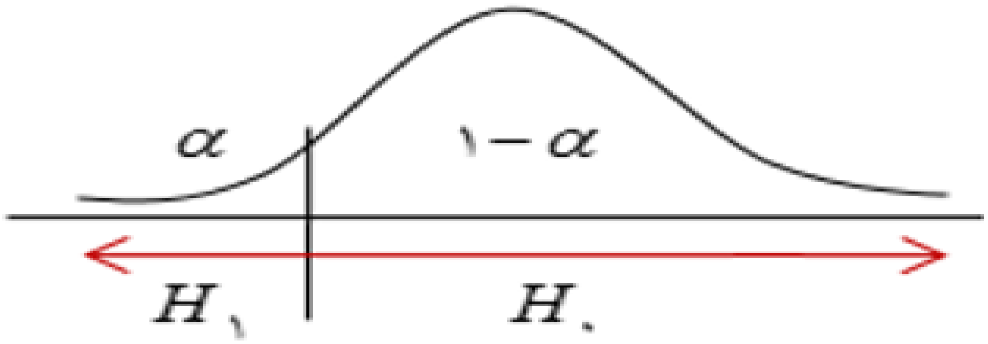
میانگین سن ازدواج در جوانان30سال است.
فردي ادعا مي‌كند اين مقدار در سال گذشته كاهش داشته است.

آزمون يكطرفه چپ



منطقه پذیرش یا رد



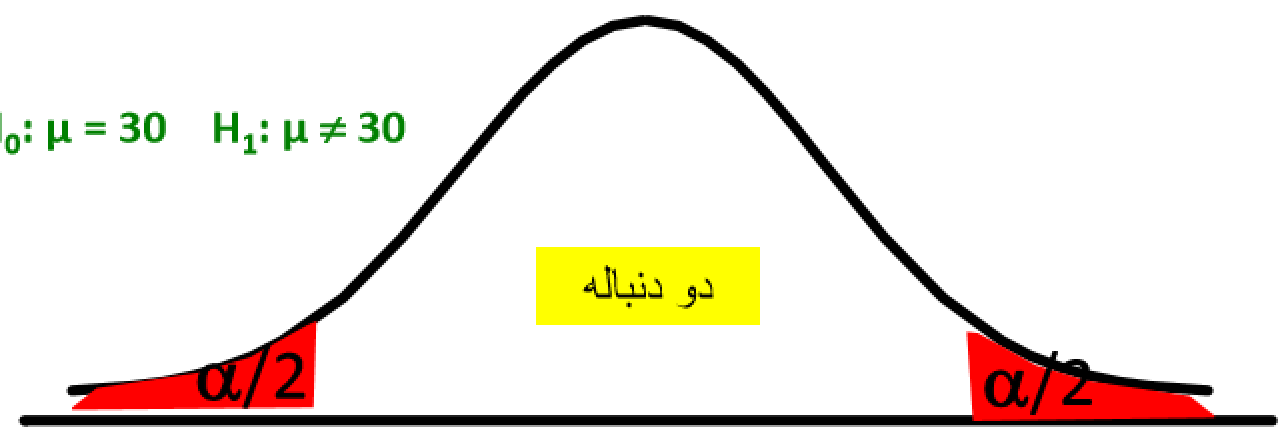


آزمون فرض

میانگین سن ازدواج در جوانان30سال است.
فردي می خواهد بررسی کند آیا این میزان تغییر یافته است؟

آزمون
دو طرفه

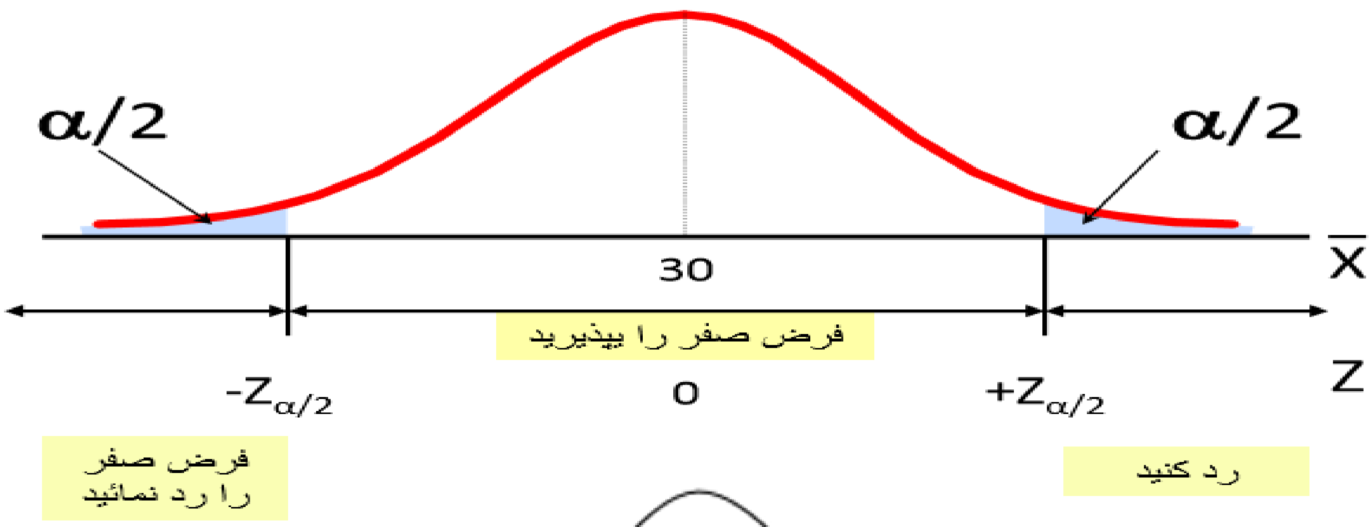
$H_0: \mu = 30 \quad H_1: \mu \neq 30$

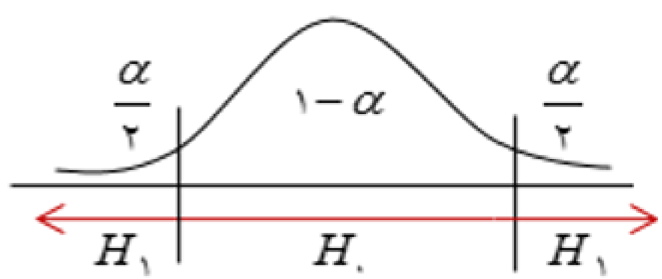


منطقه پذیرش یا رد

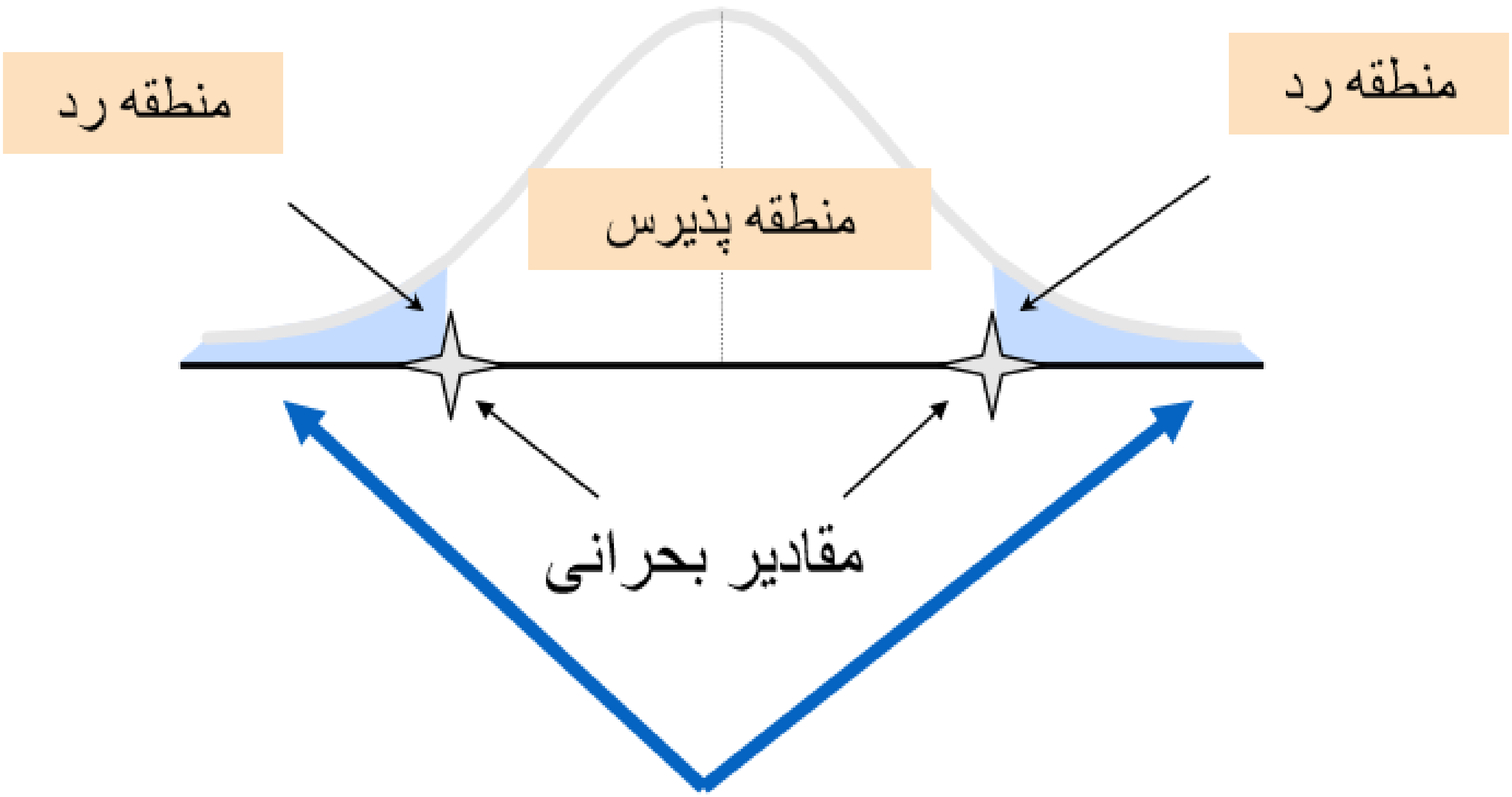
$H_0: \mu = 30$

$H_1: \mu \neq 30$





پذیرش یا رد



ناحیه بحرانی

ملاکی برای رد یا عدم رد فرض صفر.

انتخاب ملاک مناسب برای حداقل کردن خطای نوع اول و دوم

براساس مقدار خطای نوع اول، ناحیه بحرانی تعیین می‌شود.

نتایج آزمون‌های آماری بوسیله نرم‌افزارهای رایانه‌ای، دارای ناحیه بحرانی است که در سطح ثابت میزان خطای نوع اول، کمترین میزان خطای نوع دوم را دارد. (آزمون‌هایی با بیشترین توان)

آماره آزمون

$$\text{آماره آزمون} = \frac{\text{آماره نمونه} - \text{مقدار پارامتر جامعه آماری تحت آزمون صفر}}{\text{خطای استاندارد آماره نمونه}}$$

آماره آزمون

1. از داده های نمونه محاسبه می شود
2. با مقادیر بحرانی مربوط به فرض صفر قیاس می شود

در صورتیکه آماره آزمون از مقادیر بحرانی تجاوز کند (خارج محدوده مقادیر بحرانی باشد)، محقق فرض صفر را رد می کند

تفاوت آماره آزمون با عدد بحرانی

یکی از مشخصه هایی که به ما برای تصمیم گیری کمک میکند آماره آزمون است .آماره تابعی از مشاهدات است که به پارامتر مجهول بستگی ندارد بنابراین با داشتن یک نمونه تصادفی می توان مقدار آماره را تعیین کرد و از روی آن برای رد یا عدم رد فرضیه اولیه در مقابل فرضیه مقابل تصمیم گیری کرد اگر مقدار آماره ازمون محتمل باشد ان فرض که پذیرفته ایم مبتنی بر فرضیه H0 درست می باشد و در صورتی که مقدار آماره آزمون غیر محتمل باشد فرضیه H0 نادرست است

این که چه مقادیری از آماره آزمون محتمل وبا چه مقادیری غیر محتمل است بستگی به ناحیه بحرانی دارد ناحیه بحرانی ازمون مجموعه ای است که به ازای تمام مقادیرش فرضیه اولیه رد شود بنابر این ناحیه بحرانی مقادیری از آماره آزمون است که به ازای ان فرضیه H0 رد میشود.

از توجه شما

سپاسگزارم