فصل چهاردهم

احصائیه

روش جمع آوری معلومات:

وقت که برای جمع آوری اطلاعات ، سوال می کنید ، سوال می کنید ، سوال را میتوان به شکل شفاهی یا کتبی بپرسید. برخی اوقات بهتر است سوال نکینم در این مورد به مشاهده آن بپردازیم تا اطلاعات بهتر به دست آوریم و بعضی اوقات باید آزمایشی را انجام داد تا اطلاعات را جمع آوری نماییم.

مثال:

(1) اگر بخواهیم در آمد یک خانواده را بدانیم ، از چه روش بهتر است اطلاعات را جمع آوری نماییم یا از اطلاعاتی که از قبل ثبت شده ، استفاده نماییم؟

جواب: هرگاه در آمد کم باشد ، ممکن است شاگردان خوش نداشته باشد که معلومات دهد ، پس بهتر است بدون نام ازآنها بپرسیم.

(2) اگر بخواهیم نمره ریاضی شاگردان صنف ششم را بدانم ، از چه روش بهتر است اطلاعات را جمع نماییم؟

جواب: چون ممکن است شاگردان نمره واقعی خود را نگوید ، پس بهتر است از دفاتر ثبت شده استفاده نمایم.

③ اگر بخواهیم تعداد خواهران و برادران شاگردان را بدانم از چه روش اطلاعات را جمع آوری مینماییم؟

جواب: از پرسش شفاهی یا کتبی استفاده مینماییم.

④ اگر بخواهیم وزن نوزادان را بررسی کنیم ، از چه روش بهتر است اطلاعات را جمع آوری نماییم؟

جواب: باید وزن نوزادان را اندازه گیری نماییم.

- ✓ جمع آوری داده ها را اطلاعات می گویند.
- ✓ روش های جمع آوری اطلاعات عبارت اند از: پرسش (شفاهی ، مصاحبه) مشاهده و انجام آزمایش و یا استفاده از اطلاعات ثبت شده.
- ✓ جامعه احصائیوی یا به طور خلاصه جامعه ، مجموعه یی از افراد و اشیای است که در بارهٔ اعضای آن اطلاعات مورد نیاز را دریافت می کنیم.

جامعه و نمونه:

در یک بررسی ، مجموعه همه افراد و یا اشیا که از آن ها اطلاعات مورد نیاز را دریافت می کنیم جامعه می نامیم. هرگاه اطلاعات از همه اعضای جامعه به دست آوریم ، این عمل را رأی پرسی همگانی می گویند. برخی اوقات به دلیل مشکلات چون کمبود وقت ، مشکلات اقتصادی ، امکان نداشتن دسترسی به همه افراد جامعه مجبور هستیم فقط اطلاعات بخشی از اعضای جامعه را به دست آوریم.

نمونه بخشی از جامعه است ، یک نمونه از جامعه باید خاصیت و صفات کل جامعه را داشته باشد.

- ✓ بخشی از جامعه را نمونه می گویند.
- تعداد اعضای جامعه را اندازه جامعه و تعداد اعضای نمونه را اندازه نمونه می نامند.
- ✓ به خاطر شناخت جامعه ، نمونه یی را که از آن جامعه انتخاب می کنیم ، باید نمونه تصادفی باشد روش انتخاب نمونه به گونه یی باشد که:
  - ○ انتخاب هر فرد به عنوان عضوی از نمونه امکان پذیر باشد.
  - ○ پیش از انتخاب یک نمونه باید نتوانیم دربارهٔ مشخصات آن قضاوت کنیم.

مثال های از جامعه و موضوع مورد بررسی آن ها:

جامعه		موضوع مورد بررسی
سابقه تدریس معلمان در هرات	....	معلمان ولایت هرات
محصول پنبه سمت شمال	....	میزان تولید پنبه
محصول زراعتی افغانستان	....	انواع محصولات افغانستان

مثال های از نمونه:

- یک مثلث برنج ، نمونه یی از یک بوجی برنج است.
- شاگردان صنف هفتم مکتب شما نمونه یی از شاگردان صنف هفتم افغانستان است.
- معلمان ریاضی کندز نمونه یی از معلمان کندز است.
- گندم نمونه یی از محصولات زراعتی افغانستان است.

نمونه تصادفی:

برای آن که یک نمونه ، نشان دهندۀ یک جامعه و دارای خصوصیات جامعه باشد باید خصوصیات ذیل را داشته باشد.

1. امکان انتخاب هر فرد و یا شی به عنوان عضوی از نمونه امکان پذیر باشد.
2. قبل از انتخاب نمونه نتوانیم در بارۀ اعضای نمونه قضاوت کنیم.
3. تمام اعضای جامعه به عنوان یک نمونه سهم برابر داشته باشد.

مثال: کدام یک از نمونه های ذیل یک نمونه تصادفی است.

موضوع: بررسی سواد اهالی شهر

جامعه: اهالی شهر

نمونه اول: آن فرد که ساعت 5 بعد از ظهر از سرک عبور کند؟

جواب: این یک نمونه تصادفی است زیرا از قبل سواد فردی را که از سرک می گذرد نمی توان پیش بینی کرد.

نمونه دوم: دکتوران یک شفاخانه؟

جواب: این یک نمونه غیر تصادفی است زیرا از قبل می توان نتیجه را پیش بینی کرد و این نمونه نشان دهنده همه جامعه نیست.

متحول تصادفی و انواع آن:

اگر اطلاعات جمع آوری شده از موضوع مورد مطالعه از یک عضو جامعه به عضو دیگر قابل پیش بینی نباشد موضوع را متحول تصادفی می نامیم.

این دسته از متحول ها را بنام متحول کمی یا عددی می نامیم ، هرگاه در متحول کمی نتوانیم بین دو واحد پشت سر هم ، عددی پیدا کنیم آن را کمی مجزا می نامیم. اگر بین دو واحد پشت سر هم بتوانیم عددی را پیدا نماییم آن را کمی پیوسته می نامیم. در صورت که اطلاعات را توصیف و بدون عدد بیان کنیم متحول کیفی یا توصیفی می نامیم.

- ✓ اطلاعات جمع آوری شده از یک موضوع را متحول های تصادفی می گویند.

متحول های تصادفی دو نوع است.

- ❖ کمی یا عددی که قابل اندازه گیری باشد.
- ❖ کیفی یا غیر عددی که قابل اندازه گیری نمی باشد.

متحول های کمی دو نوع است.

A. پیوسته: که بین هر دو مقدار آن میتوان مقدار دیگر را دریافت کرد.

B. مجزا: که پیوسته نباشد.

مثال:

متحول تصادفی با شمارش	متحول تصادفی با اندازه گیری	متحول تصادفی با توصیف
کمی مجزا	کمی پیوسته	کیفی
① تعداد اعضای خانواده	① طول قد شاگردان	① رنگ چشم شاگردان
② تعداد صنف های مکتب	② درجه حرارت شهر شما	② میزان سواد کارگران
③ تعداد موتر های که از سرک می گذرد	③ وزن گوسفندان	③ موسیقی مورد علاقه مردم

جدول کثرت $\boldsymbol{Frequency\ Table}$:

در یک بررسی ، داتا های جمع آوری شده را که روی آن هیچ عمل انجام نشده باشد ، داتای خام می نامند. در هر بررسی با منظم کردن داتاها جدولی تشکیل نموده که آن را جدول کثرت می نامیم. برخی اوقات یک جدول را به شکل سطری ترتیب نموده و مقدار دفعات را که یک داتا تکرار شده است ، کثرت آن داتا می نامیم.

مجموع کثرت داتاها در یک نمونه برابر با کل داتاها یا تعداد اعضای نمونه می باشد، اگر f_1 کثرت داتای اول ، f_2 کثرت داتای دوم .... f_n کثرت داتای n—ام و تعداد کل داتاها مساوی به n باشد پس:

$$n = f_1 + f_2 + \cdots + f_n$$

اطلاعات جمع آوری شده را داتا $Data$ می گویم.

- ✓ اگر $x_1 , x_2 , \dots \dots , x_n$ تعداد داده های یک موضوع باشد ، تعداد دفعاتی که یک داده تکرار می شود ، به نام کثرت آن داده ها می گویند ، و معمولاً آن را به $f_1 , f_2 , \dots \dots , f_n$ نشان می دهند.

مثال:

تعداد اعضای خانواده	**1**	**2**	**3**	**4**	**5**	6	7	8	مجموع
تعداد خانواده	4	**7**	**9**	8	6	3	2	1	**40**

جدول بالا نشان می دهد که تعداد 6 خانواده 5 نفر عضو داشته و تعداد یک خانواده 8 نفر عضو دارد.

گراف تصویری:

گاهی برای دانستن اطلاعات داده شده از سمبول ها و اشکال استفاده مینماییم ، این روش را به نام گراف تصویری می گویند. در صورت که کثرت داده ها زیاد باشد ، از مقیاس استفاده مینماییم.

مـــــــــد $Mode$:

داتایی که بیشترین کثرت را دارد $Mode$ می نامیم که میتوان در موضوعات رأی گیری ها ، فروش کالا و غیر استفاده نمود.

اوســـط $Mean$:

اوسط داده ها را میتوان از تقسیم حاصل جمع داده ها بر تعداد داده ها پیدا کرد.

$$\bar{X} = \frac{x_1+x_2+x_3+\cdots+x_n}{n}$$

جدول کثرت داتا های گسسته:

بعضی اوقات وقت تعداد داتاها زیاد یا کثرت بسیاری از داتاها صفر یا عدد کم باشد ، جدول کثرت مجزا کمک زیادی ننموده ، یا تشکیل آن خیلی مشکل است ، بنابرین در این حالت از جدول کثرت به شکل دسته بندی استفاده می کنیم، کثرت هر دسته نشان می دهند که در دسته چند داتا وجود دارد ، ولی نمیتوانیم بگوییم آن داتا ها کدام اند.

خواص و اجزای جدول کثرت:

کم ترین مقدار که میتواند در یک دسته قرار گیرد ، سرحد پائینی و بیشترین مقدار را که میتوان در یک دسته قرار گیرد ، سرحد بالایی می نامیم. تعداد داتا های که می تواند در یک دسته قرار گیرد به نام طول دسته ، نامگذاری نموده اند، برای پیدا کردن طول دسته کافی است تفاوت بین دو سرحد بالایی دو دسته متواتر را دریافت مینماییم.

کثرت تجمعی:

مجموع کثرت مطلق هر دسته و کثرت دسته های قبل از آن را کثرت تجمعی آن دسته می نامند، و کثرت آخرین دسته برابر است با تعداد کل داتا ها.

یادداشت:

- برای یکسان نشان دادن هر دسته از عددی که اوسط هر دسته است استفاده می نماییم.
- تکرار هر داتا را کثرت مطلق آن داتا در جدول کثرت می گوییم.
- در هر دسته بندی داتا ها تمام داتاهای واقع در یک دسته را برابر مرکز آن دسته در نظر می گیریم.
- مرکز دسته ها به تعداد اعضای که در آن دسته قرار دارند تکرار می شود یعنی کثرت مرکز دسته ها برابر تعداد اعضای است که در آن دسته قرار گرفته و این کثرت را کثرت مطلق آن دسته می گوییم.

کثرت نسبی:

در بعضی حالات برای مقایسه دو وضعیت نمیتوان کثرت های مطلق را به هم مقایسه کرد. در چنین حالت از $\frac{\text{کثرت مطلق}}{\text{کثرت کل داتا}}$ استفاده مینماییم. مقدار این نسبت را کثرت نسبی می گوییم. برای مقایسه بهتر ، این عدد را با فیصدی نشان می دهیم و فیصد کثرت نسبی می نامیم.

$Note$:

1. آن دسته که $Data$ معلوم نمی شود و یا هم کثرت مطلق ندارد قیمت کثرت نسبی مساوی به صفر است.
2. به هر دسته (طبقه) مجموع تمام کثرت نسبی در یک دسته یا طبقه مساوی به یک است.
3. قیمت کثرت نسبی یک طبقه همیشه مساوی به یک عدد مثبت که از 1 کوچک است میگردد.

گراف میله ای (گراف نواری):

یک گراف میله یی باید دارای عنوان ، مقیاس و مشخصه محور باشد ، در گراف میله یی محل قرار گرفتن داتاها مهم نیست ، طول میله ها نشان دهندۀ کثرت داتاها می باشد. از گراف میله یی بیشتر برای ترسیم متحول مجزا و کیفی استفاده می کند ، در ترسیم گراف میله یی ترتیب قرار گرفتن میله ها اهمیت ندارد ، آن چه که در این گراف مهم است ، کثرت داتاها است.

گراف خط منکسر:

اگر اطلاعات جمع آوری شده را توسط نقاط در مستوی مختصات ترسیم نموده و این نقاط را به کمک خطوط منکسر با هم وصل نماییم ، گراف که توسط نقاط وصل شده به دست می آید به نام گراف خط منکسر یاد می شود.

اوسط داتاهای گسسته با کثرت:

برای پیدا کردن اوسط داتاها در صورت تکرار داتاها میتوان به جای جمع داتاها از ضرب کثرت در داتاها استفاده نمود اگر در حالت کل یک داتا را به X و کثرت آن به f نشان دهیم ، این حاصل ضرب برابر $f\,.\,x$ خواهد بود. اگر داتا اول و کثرت آن را به x_1 و f_1 ، داتا دوم و کثرت آن را به x_2 و f_2 .... داتا آخر را به x_n و کثرت آن را به f_n نشان دهیم ، اوسط داتا ها عبارت است از:

$$\bar{X}=\frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{f_1+f_2+\cdots+f_n}=\frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{n}$$

اوسط داتاهای پیوسته:

در داتاهای پیوسته ، مرکز دسته ها را در کثرت های آن ضرب و با هم جمع کرده و پس از آن بر مجموع کثرت (که همان تعداد داتاها است) تقسیم کرد.

$$\bar{X}=\frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{f_1+f_2+\cdots+f_n}$$

روش دسته بندی داتاها:

برای دسته بندی داتاها مراحل زیرا به ترتیب انجام می دهیم.

1. وسعت بیشترین مقدار و کم ترین مقدار داتاها را به دست آورید.
2. این وسعت را بر تعداد دسته ها تقسیم و طول دسته را به دست آورید در صورت که حاصل عددی طبیعی نباشد ، میتوان آن را به بالا گرد $(Round\ Off)$ کرد.
3. دسته ها را به این مقدار تشکیل دهید.

مراحل دسته بندی داتا:

A. ساحه تحول: وسعت بین بیشترین و کمترین داتا.

B. طول دسته: نسبت ساحه تحول بر تعداد دسته ها.

C. کثرت دسته: تعداد داتاهایی که در هر دسته قرار دارد.

D. کثرت دسته: محاسبه وسط هر دسته.

دسته بندی داتاهای پیوسته:

در دسته بندی اگر متحول ها پیوسته باشد ، سرحد بالایی دسته بعدی برابر سرحد پایانی دسته قبلی است. در صورتیکه داتا برابر به سرحد بالایی دسته باشد آن داتا متعلق به دسته بعدی می باشد.

اوسط وزنی:

اگر داتاها با ضریب خاص بیان شده باشند ، به این معنا است که تأثیر داتاها یکسان نبوده ، بسته گی به ضریب آن دارد ، در این حالت در جدول کثرت ضریب ها به عنوان کثرت آن داتا به حساب آمده و به w نشان داده می شوند ، اوسط به دست آمده در این حالت را اوسط وزنی می نامیم.

$$\text{اوسط وزنی} = \bar{x} = \frac{w_1x_1+w_2x_2+\cdots+w_nx_n}{w_1+w_2+\cdots+w_n}$$

داتا ها	x_1	x_2	$\cdots\cdots\cdots$	x_n
ضریب (وزن)	w_1	w_2	$\cdots\cdots\cdots$	w_n

گراف مستطیلی $Histogram$:

در نمایش داتاهای پیوسته ، از گراف مستطیلی استفاده می کنیم، در این گراف عرض مستطیل ها برابر طول دسته ها است در گراف مستطیلی ، مساحت های مستطیل نشان دهنده کثرت هر دسته بوده ، و با یکدیگر خوبتر مقایسه می شوند، مساحت مستطیل ها کثرت دسته ها را نشان می دهند. اگر طول دسته ها با هم برابر باشند ، میتوان عوض مساحت ها کثرت را مستقیماً مقایسه نماییم ، در این وضعیت محور عمودی کثرت را نشان می دهد.

- ✓ هــتوگرام یا گراف مستطیلی عبارت از گراف است که در آن توزیع کثرت توسط مستطیل ها نشان داده می شود. عرض مستطیل یا قاعده مستطیل برابر به طول دسته و طول مستطیل یا ارتفاع مستطیل برابر به کثرت دسته است. مساحت هر مستطیل برابر به حاصل ضرب طول دسته و کثرت دسته می باشد ، در گراف مستطیلی ، مستطیل ها با هم پیوسته و از متحولین پیوسته برای نشان دادن گراف استفاده می شود.

گراف دایره یی:

نشان دادن داتاها به کمک دایره را بنام گراف دایره یی میگویند. درگراف دایره یی ابتدا نسبت کثرت هر دسته را بر تعداد کل داتاها تقسیم و ضرب در $360°$ که زاویه مرکزی همان دسته را نشان می دهد، مینماییم.

$$\text{کثرت بر حسب درجه} = \frac{\text{کثرت داتاها}}{\text{تعداد کل داتاها}} \cdot 360°$$

- ✓ دایره یی را به شعاع اختیاری به وسیله زاویه مرکزی به n قسمت تقسیم می کنیم به قسم که اندازه زاویه مرکزی هر یک از این قسمت ها متناسب به کثرت آن قسمت باشد، در این صورت زاویه مرکزی نظر به دسته اول عبارت است از:

$$\text{کثرت بر حسب درجه} = \frac{\text{کثرت داتاها}}{\text{تعداد کل داتاها}} \cdot 360°$$

میانه $Median$:

پس از مرتب کردن داده ها مقداری را که تعداد داتاهای بعد از آن با تعداد داتاهای قبل از آن برابر باشد میانه می گوییم. اگر تعداد داتاها تاق باشد ، میانه خود یکی از داتاهای مابینی است ، ولی اگر تعداد داتا جفت باشد ، میانه وسط دو داتا مابینی است.

ساحه تحول $Range$:

طول فاصله یی که متحول در آن امکان تغییر را دارد ، به نام ساحه تحول یاد می کند. این معیار ، وسعت بین بیشترین و کم ترین داتا را نشان می دهد ، متوجه باید بود که بزرگی ساحه تحول نشان دهنده فرق یا پراگنده گی زیاد در جامعه است ، هر اندازه این فرق کم تر باشد پراگندگی افراد کم تر است. افراد جامعه از لحاظ این خصوصیت به هم نزدیکتر اند ، اگر ساحه تحول صفر باشد ، خصوصیت مورد بررسی همه افراد با هم برابر و یکسان اند در آن حالت جامعه را یک جامعه متجانس می نامیم.

- ✓ ساحه تحول عبارت از تفاوت کوچکترین داتا از بزرگترین داتا در مجموع داتاها می باشد یا به عبارهٔ دیگر ساحه تحول عبارت است از طول فاصله است که متحول در آن ساحه تغییر مینماید.

اوسط انحراف $Average\ deriation$:

وسعت که بین داتاها و اوسط موجود است آن را انحراف از اوسط می گویند. مجموع انحراف از اوسط داتاها همیشه صفر است ، به این دلیل برای بررسی داتاها از قیمت مطلقه انحراف ها استفاده می کنیم. اگر قیمت مطلق همه انحراف ها را جمع نموده و تقسیم بر تعداد داتاها نماییم اوسط انحراف گفته می شود.

$$\text{اوسط انحراف} = \frac{|x_1-\bar{x}|+|x_2-\bar{x}|+\cdots+|x_n-\bar{x}|}{n}$$

هر چه انحراف اوسط عدد بزرگتر باشد ، به همان اندازه پراگندگی داتاها از اوسط بیشتر است.

- ✓ تفاضل اوسط از هر داتا را انحراف از اوسط گویند.

گراف چند ضلعی کثرت:

در گراف چند ضلعی ، مرکز هر دسته روی محور افقی و کثرت مطلق یا کثرت نسبی هر یک از دسته ها روی محور عمودی نشان داده می شود. متقابل با مرکز هر دسته و کثرت آن یک نقطه در مستوی مشخص میگردد که عرض آن مرکز دسته و طول آن برابر با کثرت آن دسته است. به تعداد دسته های جدول در مستوی سیستم مختصات نقطه به وجود می آید. اگر به نقاط مذکور دو نقطه اختیاری دیگر $(x_1 - c, 0)$ و $(x_n + c\ ,0)$ را در اول و آخر دسته ها اضافه کنیم ، طوریکه c وسعت هر صنف (سرحد بالایی منفی سرحد پائینی صنف) است از اتصال این نقاط به یکدیگر ، یک گراف حاصل می شود که آن را گراف چند ضلعی کثرت می نامند.

- ✓ هر یک از راس های گراف چند ضلعی کثرت در نقاط مابینی ضلع بالایی یک مستطیل مربوطه به جدول کثرت مورد مطالعه قرار دارد.
- ✓ مساحت سطح زیر گراف چند ضلعی کثرت و مساحت گراف مستطیلی با هم برابر است.
- ✓ گراف چند ضلعی کثرت نسبی بیش تر برای دیتا $Data$ پیوسته یا متصل به کار می رود.
- ✓ جوره های مرتب نقاطی را که عرض آن ها مرکز دسته ها و طول آن ها برابر کثرت همان دسته باشد با هم وصل می کنیم گراف چند ضلعی کثرت به وجود می آید ، در گراف چند ضلعی کثرت دو نقطه با کثرت صفر به ابتدأ و انتهای دسته ها اضافه می شود تا گراف چند ضلعی کثرت به محور x متصل شود.

گراف ساقه و برگ:

برای رسم گراف ساقه و برگ از اعداد استفاده می شود. دیتای احصائیوی را به صورت اعداد در آورده و پس از این اعداد گراف ساقه و برگ را تشکیل می دهیم. این گراف برای دیتای که تفاوت کوچکترین و بزرگترین دیتا از نظر تعداد رقم ها اندک باشد ، مناسب است.

مثلاً: در عدد 37 ، عدد 3 ساقه و 7 برگ می باشد. به خاطر نمایش گراف ساقه و برگ عدد 8.3 را به صورت 083 عدد 11.2 را به صورت 112 و 12 را به صورت 120 می نویسیم.

چارک ها:

عددی که جامعهٔ مرتب را به دو قسمت مساوی تقسیم می کند ، میانه نامیده می شود. حال اعدادی را در نظر بگیرید که جامعه مرتب را به چهار قسمت مساوی تقسیم می کنند. این اعداد را با Q_1 ، Q_2 و Q_3 نشان می دهند و آن ها را به ترتیب چارک های اول تا سوم می نامند، واضع است که Q_2 میانه است.

✓ اعداد که دیتای مرتب را به چهار قسمت مساوی تقسیم می کنند آن ها را چارک های اول ، دوم و سوم می نامند ، و به Q_1 ، Q_2 و Q_3 نشان می دهند.

- چارک اول مقداری است که 25% دیتای جامعه پائین تر آن و 75% بالاتر از آن قرار می گیرد.
- چارک دوم مقداری است که 50% دیتای جامعه پائین تر از آن و 50% دیتا بالاتر از آن قرار می گیرد.
- چارک سوم مقداری است که 75% دیتای جامعه پائین تر از آن و 25% دیتا بالاتر از آن واقع است.
- اگر دیتا را به صورت صعودی مرتب کنیم میانه دیتا مساوی به Q_2 و میانه نیمه اول دیتا مساوی به Q_1 و میانه دوم دیتا مساوی به Q_3 است.

محاسبه چارک ها:

① دیتای مرتب شده را از 1 تا n شماره گذاری مینماییم.

② محل $P-$ام $(P = 1,2,3)$ را با استفاده از رابطه ذیل به دست می آوریم.

$$C_{QP} = \frac{P.n}{4} + \frac{1}{2}$$

③ با استفاده از محل چارک ، مقدار چارک ها را تعیین مینماییم.

گراف صندوقچه یی:

گراف تصویری است که پراگندگی دیتا را نسبت به گراف های دیگر بهتر نشان می دهد. این گراف دیتا را بر اساس مقادیر ذیل نمایش می دهند.

① کمترین دیتا ② چارک اول ③ میانه

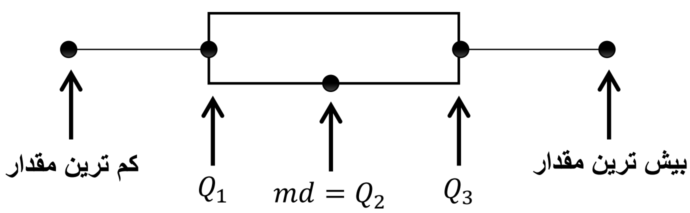
④ چارک سوم ⑤ بیش ترین دیتا

گراف صندوقچه یی نشان دهندهٔ چارک ها ، حد اقل و حد اکثر دیتا است.

مراحل تهیه گراف صندوقچه یی:

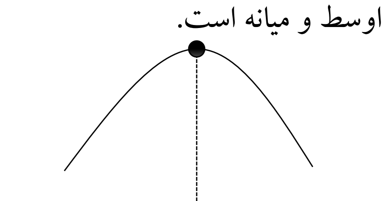
① کوچکترین دیتا ② بیش ترین دیتا ③ میانه

④ چارک اول ⑤ چارک سوم ⑥ ترسیم گراف



مقایسه شاخص های مرکزی توسط منحنی نارمل:

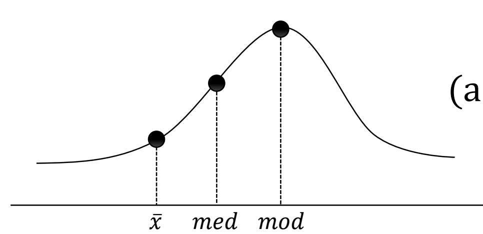
اگر منحنی نارمل متناظر باشد ، در این صورت موقعیت میانه و اوسط در منحنی نارمل یکسان می باشد، و چون منحنی نارمل نقطه اعظمی دارد ، بنابر این موقعیت مـــــود آن نیز ، برابر اوسط و میانه است.



$\bar{X} = mod = med$

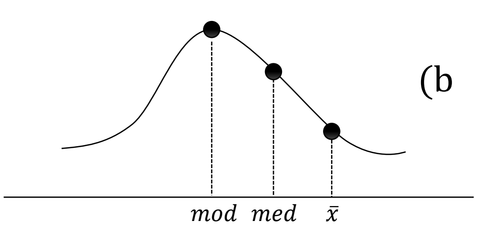
اگر منحنی نارمل متناظر نباشد در این صورت داریم که:

(a



$\bar{x} < med < mod$

(b



$mod < med < \bar{x}$

اگر اوسط و میانه مساوی باشد ، تعداد دیتای که قبل و بعد از اوسط و میانه قرار دارند مساوی باشند، اگر اوسط در سمت چپ میانه واقع باشد ، تعداد دیتای که در سمت راست اوسط قرار دارند، بیشتر از تعداد دیتای اند که در سمت چپ اوسط قرار دارند. مانند شکل (a) .

اگر اوسط در سمت راست میانه واقع باشد، تعداد دیتای که در سمت راست اوسط قرار دارند کمتر از تعداد دیتای اند که در سمت چپ اوسط قرار گرفته. مانند شکل (b) .

انحراف چارک ها :

ساحه تحول در بعضی مواقع به علت موجودیت دو مقدار خیلی کوچک و خیلی بزرگ در جامعه ممکن است تعبیر های نامناسب از جامعه را ارائه کند، بنابر این در همچو مواقع از شاخص دیگری به نام انحراف چارک ها که بتواند ساحه تحول جامعه را بهتر مشخص نماید استفاده مینماییم.

اگر Q_1 و Q_3 به ترتیب چارک اول و سوم مجموعه ای از دیتا باشند ، انحراف چارک ها را به Q نمایش داده و قرار ذیل تعریف می کنند.

$$Q = Q_3 - Q_1$$

انحراف چارک ها یکی از شاخص های نشان دهندهٔ پراگندگی دیتا است ، زیرا از روی تعریف چارک اول و سوم بر می آید که 50% جامعه در فاصله $Q_3 - Q_1$ قرار دارند ، هر مقدار این فاصله کوچکتر باشد، دیتا جمع تر و به عبارت دیگر پراگندگی آن کم تر است.

گاهی انحراف چارک ها را به صورت $Q = \frac{Q_3 - Q_1}{2}$ نیز تعریف می کند، و آن را نیم چارک می نامند.

واریانس $Variance$:

شاخص های پراگندگی ، اندازه هایی هستند که وضع پراگندگی دیتا را نسبت به یکدیگر و نسبت به اوسط مشخص می کنند. واریانس یکی از مهم ترین شاخص های پراگندگی است که با S^2 نشان داده می شود و از رابطه ذیل به دست می آید.

$$S^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n} = \frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n}$$

محاسبه واریانس از جدول کثرت توسط فورمول $S^2 = \dfrac{\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^2}{\sum_{i=1}^{n} f_i}$ به دست می آید.

x_i: مرکز دسته ها

مراحل محاسبه واریانس:

1. اوسط دیتا را دریافت می نماییم. $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

2. مجموع مربع های انحراف ها یعنی: $\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \left(x_1 - \bar{x}\right)^2 + \left(x_2 - \bar{x}\right)^2 + \cdots\cdots + \left(x_n - \bar{x}\right)^2$

3. مجموع فوق را بر تعداد اعضای مجموعه ، n تقسیم نموده و مساوی به S^2 نشان می دهیم. $S^2 = \dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}$

برخی اوقات برای محاسبه واریانس از فورمول ذیل نیز استفاده مینماییم. $S^2 = \dfrac{\sum_{i=1}^{n} x_i^{\,2}}{n} - \bar{x}^2$

انحراف معیاری:

جذر مربع واریانس را به S نشان می دهند و آن را انحراف معیاری می گویند. $S = \sqrt{\dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}}$

محاسبه انحراف معیاری از جدول کثرت توسط فورمول ذیل به دست می آید. $S = \sqrt{\dfrac{\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^2}{\sum_{i=1}^{n} f_i}}$

ضریب تغییرات $Coefficient\ Variations$:

ضریب تغییرات یا پراگندگی نسبی موارد استعمال زیاد دارد که واریانس و انحراف معیاری فاقد آن ها است. یکی از کاربرد های آن مقایسه نمودن دو جامعه احصائیوی نا متجانس و نا همگون است. ضریب تغییرات که به سمبول CV نشان داده می شود ، عبارت از خارج قسمت انحراف معیاری بر اوسط که عدد مطلق (بدون واحد) است .

یعنی:

$$\text{ضریب تغییرات} = \frac{\text{انحراف معیاری}}{\text{اوسط}} \quad \text{یا} \quad CV = \frac{S}{\bar{X}}$$

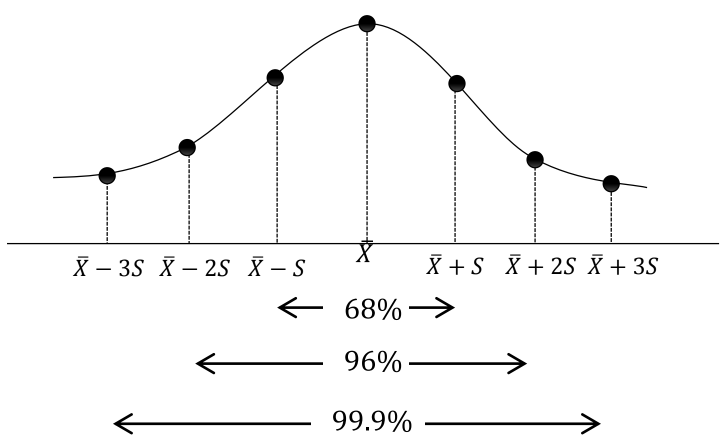
ضریب تغییرات را اکثراً به صورت فیصدی می نویسند. اگر ضریب تغییرات به 100 ضرب بشود ضریب تحول بدست می آید.

$$\text{ضریب تحول} \qquad CV\% = 100\ \frac{S}{\bar{X}}$$

- ضریب تغییرات فقط برای $data$ مثبت تعریف می شود.
- اگر همه $data$ با هم برابر باشند ، ضریب تغییرات صفر است.
- اگر همه دیتا را در یک عدد مثبت ضرب کنیم ضریب تغییرات تغییر نمی کند.
- اگر به همه دیتا یک عدد مثبت را اضافه کنیم ، ضریب تغییرات جدید کوچکتر از ضریب تغییرات دیتا اولیه است.

پراگندگی در منحنی نورمال $Normal\ Curve$:

منحنی نورمال وسیله مهم برای توصیف از مجموع های آماری است. در توزیع نورمال در حالت که اداره ها دارای توزیعی نورمال و منحنی کثرت متناظر باشد ، واریانس نقشی عمده یی دارد. در واقع با مشخص بودن دو پارامتر اوسط و انحراف معیاری در توزیع نورمال ، این توزیع در کل مشخص خواهد بود و محاسبه هر نوع شاخص مساعد است.



شاخص های شکل توزیع نورمال:

شاخص های شکل توزیع را میتوان در دو حالت مطالعه نمود.

① شاخص خمیدگی $Skewness$:

توزیع که در اطراف اوسط متناظر نباشد خمیدگی گفته می شود این شاخص را توسط دو ضریب زیر نشان می دهند.

I. **ضریب خمیدگی:** شاخص است که برای تعیین میزان خمیدگی به کار می رود.

$$\alpha_3 = \frac{\frac{1}{n}\sum\left(x_i - \bar{x}\right)^3}{S^3}$$

اگر $\alpha_3 = 0$ توزیع متناظر ، اگر $\alpha_3 > 0$ توزیع خمیدگی مثبت و اگر $\alpha_3 < 0$ توزیع دارای خمیدگی منفی است.

II. **ضریب خمیدگی پیرسون:** ضریب پیرسون به صورت زیر تعریف می شود.

$$Sk_{(P)} = \frac{3\left(\bar{x} - med\right)}{S}$$

در توزیع متناظر ضریب خمیدگی پیرسون برابر صفر است ، کمیت های مثبت و منفی ضریب خمیدگی پیرسون به ترتیب نشان دهنده خمیدگی مثبت و یا منفی توزیع است.

② شاخص کشیدگی $Kurtosis$:

شاخص کشیدگی نشان دهنده آن است که یک توزیع چه وقت دارای اوج و چه وقت دارای پخشی می باشد.

ضریب کشیدگی معمولترین شاخص است که برای اندازه گیری کشیدگی به کار رفته به صورت ذیل تعریف می گردد. $\alpha_4 = \dfrac{\frac{1}{n}\sum\left(x_i - \bar{x}\right)^4}{S^4}$

در صورت جدول کثرت فورمول شاخص کشیدگی $\dfrac{\frac{1}{n}\sum f_i\left(x_i - \bar{x}\right)^4}{S^4}$ است.

شاخص کشیدگی بستگی به موقعیت و پراگندگی توزیع ندارد ، این شاخص برای مقایسه مورد استفاده قرار می گیرد.

جامعه های چند متحوله:

یکی از اهداف عمده در اکثر تحقیقات احصائیوی پیش بینی و پیش گویی نمودن و تعیین یک متحول از جنس متحول دیگر است، زمان که ارتباط بین دو شی را مورد بررسی قرار می دهیم منظور از جامعه دو متحوله می باشد. مانند رابطه بین حجم و فشار گاز ، ارتباط بین صحت و میزان مرگ و میر رابطه بین سطح کشت و مقدار محصول ، رابطه بین شعاع دایره و مساحت آن.

برای سهولت ، معمولاً ارتباط بین دو یا چندین متحول را به وسیله معادلات ریاضی ارائه می دارند. در قدم اول به منظور تشخیص و تشکیل معادلات مورد ضرورت معلومات لازم جمع آوری می گردد، در قدم دوم معلومات جمع آوری شده به شکل ارزش متحول های مورد مطالعه در یک مستوی مختصات قایم، نقطه گذاری می گردد، که از وصل این نقاط به دست می آید یک گراف را به ما می دهد.

گراف پراگندگی $Scater\ diagram$:

برای ترسیم نمودن گراف پراکنش (پراگندگی) داده ها را به صورت جوره های مرتب ارائه و توسط نقاط در یک مستوی محور های مختصات نمایش می دهیم.
گراف پراکنش می تواند سه نوع اطلاعات را در اختیار ما قرار دهد.

- نمونه یی که نشان دهندهٔ نوعی ارتباط بین مشاهدات باشد موجود است یا نه؟
- در صورت موجودیت نوعی ارتباط آیا ارتباط خطی است یا غیر خطی؟
- در صورت که رابطه خطی باشد نوع ارتباط چگونه است؟

همبسته گی و ضریب همبسته گی:

همبسته گی به سنجش و دریافت درجه ارتباط بین متحول ها می باشد، ارتباط بین متحول ها می تواند به صورت خطی توسط یک خط مستقیم و یا به صورت غیر خطی به وسیله یک منحنی ارائه می گردد.

همبسته گی عموماً به دو صورت مثبت و منفی بین دو متحول بیان می شود، اگر اندازهٔ دو متحول در یک جهت تغییر کند یعنی x و y هر دو بزرگ یا هر دو کوچک شوند همبسته گی مثبت (خط مستقیم) است.

بهترین معیاری تشخیص وجود همبسته گی یا عدم آن و حتی نوع ، جهت و میزان همبسته گی خط ، ضریب همبسته گی است که توسط فورمول ذیل ارائه می گردد.

$$r=\frac{\frac{1}{n}\sum xy-\left(\bar{x}\right)\left(\bar{y}\right)}{S_x\cdot S_y}$$

S_x: انحراف معیاری x

S_y: انحراف معیاری y

اگر نقاط به شکل یک خط مستقیم هر قدر که نزدیکتر باشد، خطای متحول y نظر به x کمتر است ، و برعکس هر قدر که از خط دورتر باشند، خطای y بیشتر است.

خط رگرسیون $Regression$:

رگرسیون (تخمین) ، سنجش و دریافت ارزش یک متحول تابع نظر به ارزش یک یا چند متحول مستقل می باشد. معادلهٔ که ارتباط بین متحول ها را افاده می نماید به نام معادله یی رگرسیون یا معادله سنجش یاد می شود و این معادله را میتوان به روش کمترین مربعات محاسبه و ضرایب a و b را به کمک این روش به صورت ذیل به دست آورد.

$$(y=ax+b)$$

r: ضریب همبسته گی

$$a=r\,\frac{S_y}{S_x}$$
$$b=\bar{y}-a\bar{x}$$

- ✓ بهترین خط مناسب خطی است که مجموع مربعات خط هایش از بقیه خطوط ممکن دیگر کمتر باشد، چنین خط (خط رگرسیون) می گویند.

مجموع توان های خط های دوم $=\sum\left(y-\bar{y}\right)^2=\sum\left[y-\left(ax+b\right)\right]^2=\sum\left(y-b-ax\right)^2$

$$=\left(ax_1+b-y_1\right)^2+\left(ax_2+b-y_2\right)^2+\cdots$$

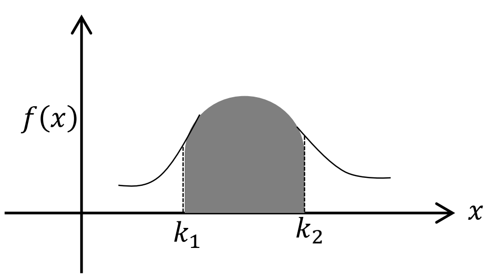
نتیجه: خط رگرسیون وسیله یا ابزاری است برای پیش بینی مقدار یک متحول بر حسب متحول دیگر که به آن وابسته است مورد استفاده قرار می گیرد.

توزیع تابع احتمال:

توزیع تابع احتمال که در احصائیه و احتمالات مورد بحث قرار می گیرد عبارت از تابع است که ناحیه تعریف آن فضای نمونه و ناحیه قیمت های آن اعداد حقیقی است.

- اگر $P(x = x_i) = f(x_i)$ داشته باشیم در این صورت جوره های مرتب $[x_1, f(x_1)]\ , [x_2, f(x_2)]\ , \dots\dots , [x_n, f(x_n)]$ را تابع احتمال مجزا (گسسته) می گویند.
- تابع احتمال پیوسته و تجمعی را میتوان به شکل $F(x) = P(x \le x)$ ارائه نمود.
- اگر $f(x)$ تابع احتمال و x متحول تصادفی باشد در این صورت احتمال اینکه x بین k_1 و k_2 قرار گیرد برابر است به:

$$P(k_1 \le x \le k_2) = \int_{k_1}^{k_2} f(x)dx$$



- اگر x متحول تصادفی پیوسته و $k_1 < k_2$ باشد در این صورت :

$$P(k_1 \le x \le k_2) = F(k_2) - F(k_1)$$

- اگر x متحول تصادفی گسسته باشد در این صورت اوسط $Expected\ Value$ یک متحول تصادفی x که $E(x)$ نشان داده می شود برابر است به:

$$E(x) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i)$$

$E(x)$ اوسط x نامیده می شود که آن را به $\bar{x}$ نمایش می دهیم و همچنان اگر x متحول تصادفی گسسته باشد در این صورت وریانس که به شکل S^2 نمایش داده می شود مساویست به:

$$S^2 = \sum_{i=1}^{n} \left[x_i - E_{(xi)}\right]^2 f(x_i)$$

- ✓ فرق بین متحول تصادفی در احصائیه و احتمالات با متحول در الجبر این است که متحول احصائیه و احتمالات از فضای نمونه و متحول الجبر از اعداد حقیقی انتخاب می شود.
- ✓ متحول تصادفی اصطلاح است که به عنوان یک تابع در احصائیه و احتمالات مورد استفاده قرار می گیرد.

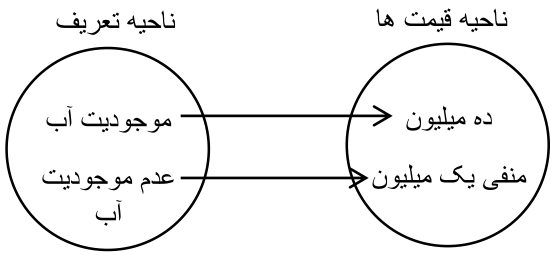
- ✓ تابع احتمال یک متحول تصادفی گسسته ، تابعی است که ناحیه تعریف آن اعدادیست که متحول تصادفی می تواند اختیار کند و ناحیه قیمت های آن شامل احتمال های مربوط به عناصر ناحیهٔ تعریف است.
- ✓ تابع احتمال تجمعی و پیوسته تابع است که ناحیه تعریف آن شامل آن اعدادیست که متحول تصادفی x اختیار می کند و ناحیه قیمت آن هم تصاویر $f(x)$ می باشد.

اوسط

$$E(x)=\sum_{i=1}^{n} x_i f(x_i)=\bar{x}\cdots$$

$$V(x)=S^2=\sum_{i=1}^{n}(x_i-E(x_i))^2\cdot f(x_i)\cdots$$

وریانس تصادفی گسسته



آزمایش برنولی و توزیع دو جمله یی:

توزیع احتمال دو جمله یی یک توزیع گسسته است که برای توصیف حوادث مختلف به کار می رود ، بیشترین اتفاقاتی که در دنیا رخ می دهد دو حالت دارد. آزمایش برنولی یک آزمایش تصادفی است که نتیجه آن را میتوان به یکی از دو حالت کامیابی و ناکامی دسته بندی کرد.

توزیع برنولی را میتوان به صورت $P(x=m)=P^m(1-P)^{1-m}=P^m\cdot q^{1-m}$ نشان داد.
در حالیکه P احتمال کامیابی و $q=1-P$ احتمال عدم کامیابی می باشد . هرگاه یک آزمایش را n دفعه تکرار کنیم یک ترادف به دست می آید طوری که اگر احتمال کامیابی هر آزمایش P و احتمال ناکامی آن q باشد در این صورت احتمال m کامیابی در این n آزمایش عبارت از:

$$P(x\le m)=\binom{n}{m}P^m\cdot q^{n-m}\qquad\qquad 0\le m\le n$$

رابطه فوق را میتوان به صورت $B(m,n,P)$ نیز ارائه نمود.

اوسط توزیع دو جمله یی $\bar{x}=n\,.\,P$ و انحراف معیاری این توزیع $S=\sqrt{n\,.\,P\,.\,q}$ است.

توزیع احتمال پواسن:

فورمول پواسن می تواند برای محاسبه تقریبی m اشکال کامیابی از n آزمایش وقت که n بزرگ و احتمال کامیابی P کوچک باشد مورد استفاده قرار می گیرد.

فورمول پواسن برای محاسبه تقریبی احتمال m اشکال در n آزمایش عبارت از:

$$P(x=m)=\frac{e^{-\lambda}.\lambda^{m}}{m!}$$

$$\bar{x}=n\,.P\ \ ,e=2.718182$$

زمان که در توزیع دو جمله یی قیمت $P\to 0$ و $n\to\infty$ از توزیع احتمال پواسن استفاده می گردد.

یادداشت: فورمول پواسن را برای محاسبه احتمال تعداد مراجعات در زمان مشخص میتوان به صورت ذیل ارائه کرد.

λ : اوسط تعداد مراجعه ها در واحد زمان

t : نسبت زمان عنوان شده به کل زمان

$$P(x=m)=\frac{e^{-\lambda t}\cdot(\lambda\cdot t)^{m}}{m!}$$

توزیع نورمال:

منحنی های توزیع نورمال می تواند به چهار طریق با یکدیگر تفاوت داشته باشد. شکل ریاضی معادل توزیع نورمال که نشان دهندۀ تابع توزیع احتمال آن $f(x)$ است به صورت زیر نوشته می شود.

S: انحراف معیاری

$$f(x)=N(x\,,\bar{x}\,,S)$$

$f(x)$: ارتفاع منحنی

$$f(x)=\frac{1}{S\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{S}\right)^{2}}$$

توزیع نورمال از جمله توزیع های پیوسته است ، اختلاف در اندازه گیری ها را میتوان توسط توزیع به خوب تقرب نمود.

- ✓ شکل توزیع نورمال متناظر و شبه زنگوله است ، در توزیع نورمال شاخص های مرکزی با هم برابر اند و متحول های تصادفی پیوسته دارای ناحیه تعریف محدود می باشد ، تابع احتمال آن:

μ: اوسط جامعه

$$f(x)=\frac{1}{\delta\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\delta}\right)^{2}}$$

δ: انحراف معیار جامعه

مساحت تحت منحنی توزیع نورمال و استاندرد کردن آن:

برای محاسبه مساحت تحت منحنی تابع احتمال $f(x)$ در فاصله های a الی b میتوان از انتیگرال زیر استفاده نمود.

$$P(a \le x \le b) = \int_a^b f(x)dx = \int_a^b \frac{a}{\delta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2}$$

✓ توسط رابطهٔ $Z = \frac{x-\mu}{\delta}$ میتوان هر مجموعهٔ احصائیوی را که دارای توزیع نورمال است به نورمال معیاری ستاندرد تبدیل کرد.

Z: متحول معیاری نورمال و منحنی را بنام منحنی نورمال معیاری یا منحنی احتمال نورمال یاد مینماید.

بخاطر داشته باشید که متحول معیاری شده Z همیشه دارای اوسط صفر و انحراف معیاری 1 می باشد. همچنان مساحت بین منحنی نورمال و محور افقی برابر به واحد انتخاب شده می باشد. به صورت ساده:

$$f(z_1 < z < z_2) = \int_{z_1}^{z_2} \frac{1}{S\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{z_1}^{z_2} N(z,0,1)dz$$

$$Z = \frac{x-\bar{x}}{S}$$

نمونه گیری: نمونه به دو دسته تقسیم می شود ، نمونه ساده و نمونه تصادفی.

روش های نمونه گیری:

نمونه گیری تصادفی: عناصر جامعه همه برای انتخاب شدن هم چانس باشد.

نمونه گیری سیستماتیک: عناصر جامعه به صورت منظم شماره $(Code)$ گذاری شده باشد.

نمونه گیری طبقه یی: جامعه به گروه های متجانس تقسیم شده باشند.

نمونه گیری خوشه یی: جامعه خیلی بزرگ باشد ، آن را به خوشه های مختلف تقسیم و از هر خوشه نمونه ای را انتخاب می کنند.

هر ویژه گی عددی یک جامعه (اوسط و انحراف معیار) را پارامتر جامعه می گویند.

هر ویژه گی عددی یک نمونه (اوسط و انحراف معیار) را آمار می گویند.

متحول های تصادفی $x_1, x_2, \dots \dots, x_n$ را یک نمونه تصادفی از متحول تصادفی x می گویند. اگر تابع مربوط آن به صورت $f(x_1, x_2, \dots \dots, x_n) = f(x_1) \cdot f(x_2) \cdots\cdots f(x_n)$ تعریف شده باشد.

- ✓ روش های نمونه گیری به صورت عموم عبارت است از:
  - نمونه گیری تصادفی
  - نمونه گیری منظم
  - نمونه گیری گروهی
  - نمونه گیری خوشه یی
- ✓ کمیت نمونه را اوسط نمونه و کمیت جامعه را پارامتر جامعه می گویند.

توزیع اوسط نمونه:

اگر $x_1 , x_2 , \ldots\ldots , x_n$ یک نمونه تصادفی از جامعه تابع احتمال $f(x)$ باشد ، در این صورت توزیع احتمال نمونه تصادفی عبارت است از:

x	x_1	x_2	$\cdots\cdots\cdots$	x_n
$f(x)$	$\frac{1}{n}$	$\frac{1}{n}$	$\cdots\cdots\cdots$	$\frac{1}{n}$

اوسط متحول $\overline{x_n}$ $\quad E(\overline{x_n}) = \mu$

واریانس متحول $\overline{x_n}$ $\quad V(\overline{x_n}) = \frac{1}{n}\delta^2$

واریانس نمونه $\quad S_n^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$

اوسط واریانس نمونه $\quad E(S^2) = \delta^2$

اوسط نمونه $\quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

δ^2 : واریانس جامعه S^2 واریانس نمونه

قضیه لمیت مرکزی:

اگر از یک جامعه بزرگ N با اوسط متناهی μ و واریانس متناهی δ^2 یک نمونه تصادفی n تایی انتخاب کنیم ، در این صورت اوسط نمونه یعنی $\bar{x}$ دارای توزیع تقریباً نورمال با اوسط $\mu_{\bar{x}} = \mu$ و واریانس $\delta_{\bar{x}}^2 =$

$\frac{\delta^2}{n}$ و متحول تصادفی $Z = \frac{\bar{x}-\mu}{\frac{\delta}{\sqrt{n}}}$ دارای توزیع نورمال معیاری است. در حالیکه ضریب $\frac{N-n}{N-1}$ برای قیمت های بزرگ N به عدد 1 نزدیک شود در حقیقت لمیت آن وقت که $n \to \infty$ مساوی به یک است.

توزیع نمونه نسبت:

اگر x متحول تصادفی ، n مجموعه آزمایش های برنولی و P احتمال موفقیت هر آزمایش باشد در این صورت آماره نسبت نمونه $\hat{P} = \frac{x}{n}$ و اوسط $E(x) = n\,.\,P$ و واریانس متحول تصادفی $V(x) = n.\,P\,.\,q$ می باشد.

توزیع $\hat{P}$: $\hat{P} = 0\,, \frac{1}{n}\,, \frac{2}{n}\,, \ldots\ldots$ $\quad f\left(n\hat{P}\right) = \binom{n}{n\hat{P}} P^{n\hat{P}} \left(1-P\right)^{n\left(1-\hat{P}\right)}$

اوسط $\quad \mu_P = E\left(\hat{P}\right) = P$

واریانس متحول تصادفی $\quad \delta^2_{\hat{P}} = V\left(\hat{P}\right) = \frac{Pq}{n} = \frac{P(1-P)}{n}$

توزیع های نورمال معیاری $\quad Z = \frac{x-nP}{\sqrt{nPq}} = \frac{\hat{P}-P}{\sqrt{\frac{P.q}{n}}}$