بخش اول

کلیات و تعاریف

با نام خدا

آمار

علمی است که به جمع آوری، تجزیه تحلیل و تفسیر نتایج می پردازد و در کلیه علوم بیولوژیکی، روانشناسی، اجتماعی و ... کاربرد دارد.

آمار خود شامل ۲ شاخه توصیفی و استنباطی است. در آمار توصیفی، انبوهی از داده ها به صورت خلاصه شده در می آید. برای خلاصه کردن این داده ها می توان از جدول، نمودار و پارامترهای وسط گرایی (متمایل به مرکز) استفاده کرد. در حقیقت در آمار توصیفی، داده ها به صورت توصیف شده به کار برده می شود، بدون اینکه هیچ گونه پیش گویی برای آینده داشته باشیم.

اما شاخه مهمتر علم آمار، شاخه استنباطی است که در آن از اطلاعات آمار توصیفی به منظور تجزیه تحلیل و تفسیر نتایج استفاده می شود. در حقیقت به کمک آمار استنباطی اولاً، می توانیم پیش بینی کنیم در آینده چه اتفاقی می افتد و ثانیاً می توان نتیجه مربوط به نمونه را به کل جامعه تعمیم دهیم.

جامعه

مجموعه افرادی که حداقل در یک صفت، مشترک هستند. مانند جامعه ایرانیان، جامعه گیاهان گلدار و ...

نمونه

زیر مجموعه ای از افراد جامعه می باشد که از آن برای تجزیه تحلیل جامعه استفاده می شود.

از آنجایی که بررسی یک جامعه بزرگ هم زمان و هم هزینه بالایی را می طلبد، بنابراین در اکثر تحقیقات علمی، نتیجه مربوط به نمونه را به کل جامعه تعمیم می دهیم.

نمونه باید شرایط زیر را دارا باشد:

۱- **نمونه نا اریب باشد:** به این مفهوم است که درصد افراد موجود در نمونه به تناسب درصد افراد موجود در کل جامعه باشد.

۲- **نمونه تصادفی باشد:** یعنی اینکه هر یک از افراد جامعه، شانس مساوی برای انتخاب شدن را داشته باشند.

اگر نمونه دارای این دو شرط باشد، نتایج مربوط به نمونه قابل تعمیم به کل جامعه می باشد.

انواع متغیرها

شرط انجام محاسبات آماری در یک تحقیق، وجود تغییرات برای صفات مورد نظر است. متغیرها دو نوع می باشند:

۱- **متغیرهای کمی**

۲- متغیرهای کیفی

متغیرهای کمی

این متغیرها عدد پذیرند و خود به دو دسته تقسیم می شوند:

۱- متغیرهای کمی پیوسته یا متصل: که قابل اندازه گیری بوده و در یک محدوده معین، هر عددی را اختیار می کنند. مانند اندازه قد و وزن.

۲- متغیرهای کمی گسسته یا منفصل: این نوع متغیرها قابل شمارش هستند و هر عددی را نمی توانند اختیار کنند . مانند تعداد فرزندان یک خانواده.

متغیرهای کیفی

این متغیرها عدد پذیر نمی باشند ولی می توانیم آنها را بر حسب شدت، گروهبندی نموده و به هر گروه، یک رتبه دهیم. مانند رنگ گل که بر اساس شدت رنگ از تیره تا روشن به هر رنگ یه رتبه اعطا می‌کنیم:

قرمز تیره : ۱ ، قرمز روشن : ۲ ، صورتی : ۳ و سفید : ۴.

نکته: همانگونه که ذکر شد، در آمار توصیفی به یکی از ۴ روش زیر می توان داده ها را خلاصه کرد:

۱- استفاده از جدول

۲- استفاده از نمودار

۳- استفاده از پارامترهای وسط گرایی یا متمایل به مرکز (میانگین، مد و میانه)

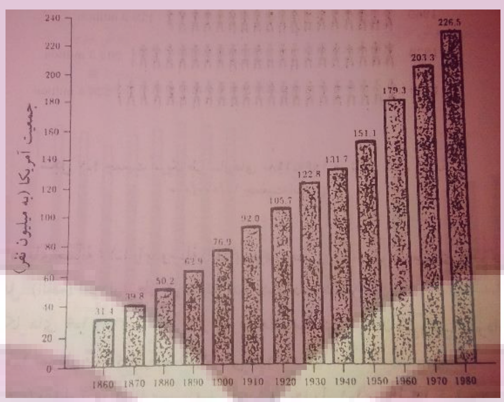
۴- استفاده از پارامترهای پراکنش یا پراکندگی (واریانس، انحراف معیار، میانگین انحرافات و دامنه)

انواع نمودارها

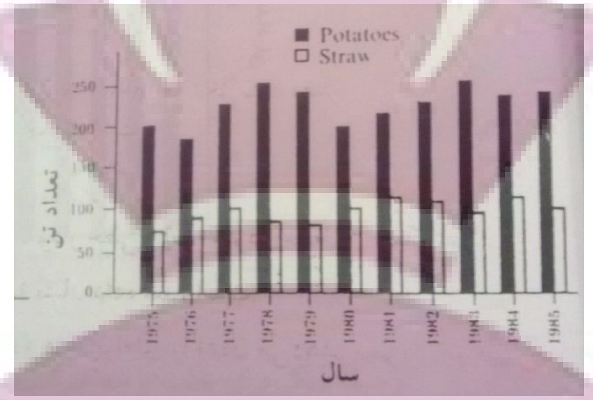
۱- نمودار ستونی: در این نمودارها از ستون یا میله برای نشان دادن فراوانی یا میزان یک عامل در طول زمان یا عامل دیگر استفاده می شود. در نمودار ستونی، فاصله بین دو ستون مجاور برابر با نصف عرض ستون است. تقسیم بندی انواع نمودارهای ستونی به شکل زیر می باشد:

- نمودار ستونی
  - افقی
    - ساده
    - غیر ساده
      - گروهی
      - مرکب
  - عمودی
    - ساده
    - غیر ساده
      - گروهی
      - مرکب

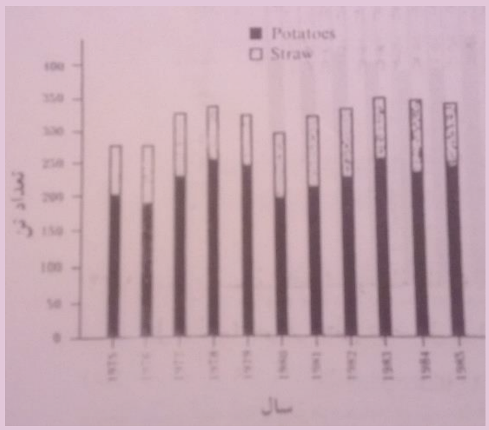
به عنوان مثال نمونه ای از نمودار ستونی عمودی ساده را در زیر مشاهده می کنید که جمعیت کشور امریکا را به میلیون نفر طی دهه های ۱۸۶۰ تا ۱۹۸۰ میلادی نشان می دهد:



شکل زیر یک نمودار ستونی عمودی غیر ساده گروهی را نشان می دهد که میزان محصول سیب زمینی و پوشال را به تن طی سالهای ۱۹۷۵ تا ۱۹۸۵ نشان می دهد:

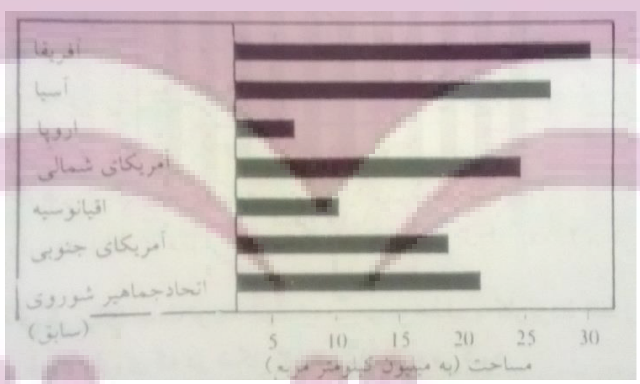


و شکل زیر یک نمودار ستونی عمودی غیر ساده مرکب برای همان داده های فوق می باشد:



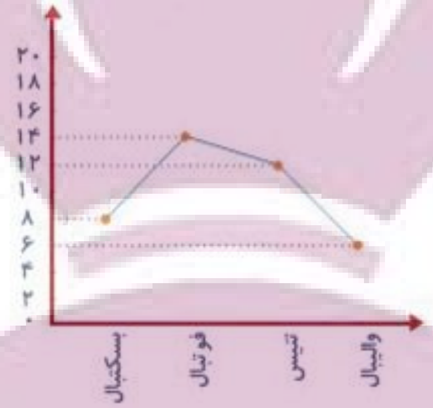
نکته: اگر هدف، مقایسه جزء به جزء باشد از نمودار گروهی و اگر هدف، مقایسه کلی باشد، از نمودار مرکب استفاده می شود.

شکل زیر هم یک نمودار ستونی افقی ساده می باشد که مساحت قاره ها را به میلیون کیلومتر مربع نشان می دهد:

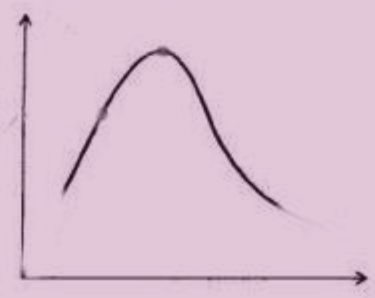


۲- نمودار خط شکسته: این نمودار از اتصال قسمت بالای ستون های نمودار ستونی با خطوط مستقیم به هم، به وجود می آید.

در نمودار خط شکسته زیر، امتیازات کسب شده توسط تیمهای مختلف یک باشگاه نشان داده شده است:



۳- نمودار منحنی: با افزایش تعداد ستونها در نمودار ستونی و همراه با کوچکتر شدن عرض ستونها، نمودار خط شکسته به تدریج تبدیل به نمودار منحنی می شود.



۴- نمودار دایره ای (قطاعی): زمانی از این نمودار استفاده می شود که اولاً با درصد یا نسبت سروکار داشته باشیم و ثانیاً تعداد، کم باشد.

برای محاسبه مقدار زاویه هر قطاع، از رابطه زیر استفاده می کنیم:

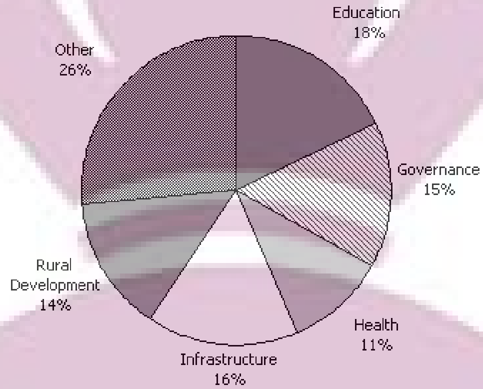
مقدار زاویه هر قطاع = نسبت یا درصد هر قسمت * ۳۶۰ درجه

یا:

$$\alpha = 360° \times \frac{a}{N}$$

که a مقدار هر قسمت یا قطاع و N مقدار کل است.

نمونه ای از نمودار دایره ای در شکل زیر نشان داده شده است تأثیر عوامل مختلف را در توسعه جامعه نشان می دهد:

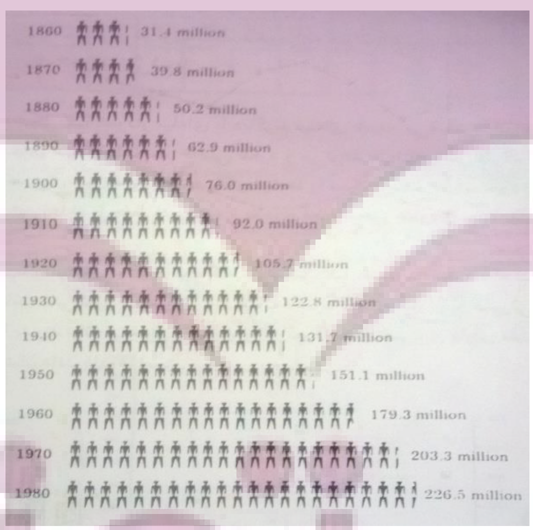


به عنوان مثال زاویه قطاع مربوط به عامل آموزش و پرورش (Education) در نمودار فوق به صورت زیر محاسبه می شود:

$$\alpha = 360° \times \frac{18}{100} = 64.8° \simeq 65°$$

۵- نمودار تصویری: در این نمودار از تصاویر برای نمایش اعداد استفاده می شود. مثلاً در نقشه های جغرافیایی، هر آدم یا آدمک نشان دهنده ۱۰۰هزار نفر یا ۱ میلیون نفر می باشد.

به عنوان مثال، شکل زیر یک نمودار تصویری است که جمعیت آمریکا را بین سالهای ۱۸۶۰ تا ۱۹۸۰ نشان می دهد و هر شکل نشان دهنده ۱۰ میلیون نفر است.



بخش دوم

(جدول توزیع فراوانی)

داده خام

داده های جمع آوری شده است که به طور عددی سازمان نیافته است. مانند دسته ای از وزن ۱۰۰ دانشجوی پسر.

داده منظم

داده های خام که در ترتیب صعودی یا نزولی است.

دامنه داده

تفاوت بزرگترین و کوچکترین عدد، دامنه داده نامیده می شود. به عنوان مثال اگر بیشترین وزن ۱۰۰ دانشجوی پسر، ۷۴ کیلوگرم و کمترین وزن آنها ۶۰ کیلوگرم باشد، دامنه داده ها برابر است با:

$D = 74 - 60 = 14$

توزیع فراوانی

زمانی که توده بزرگی از داده های خام را در دسته ها یا گروه هایی توزیع کنیم، به طوری که تعداد افراد متعلق به هر دسته را تعیین کنیم (فراوانی دسته)، این مرتب کردن جدولی داده ها در دسته ها همراه با فراوانی دسته متناظر را توزیع فراوانی یا جدول توزیع می نامیم. جدول زیر توزیع فراوانی وزن ۱۰۰ دانشجوی پسر (x) را که با تقریب کیلوگرم ثبت شده اند را نشان می دهد:

وزن (کیلوگرم)	تعداد دانشجویان (فراوانی) f
۶۰-۶۲	۵
۶۳-۶۵	۱۸
۶۶-۶۸	۴۲
۶۹-۷۱	۲۷
۷۲-۷۴	۸
کل	۱۰۰

این یعنی، به عنوان مثال سومین دسته شامل وزن های ۶۸-۶۶ کیلوگرم است که دارای دامنه ۶۶ تا ۶۸ کیلوگرم بوده و ۴۲ دانشجو، دارای وزن هایی در این محدوده می باشند.

نکته ۱: اگرچه گروه بندی داده ها در جدول توزیع فراوانی، تفصیل اولیه داده ها را معمولاً از بین می برد، اما تصویر کلی از داده ها به دست می آید.

نکته ۲: داده های انتهایی هر دسته مثلاً ۶۳ و ۶۵ در دسته سوم، **حدود دسته** نامیده می شوند. کوچکترین عدد (۶۳) را **حد پایین** و بزرگترین عدد (۶۵) را **حد بالای** دسته می نامند.

نکته ۳: چنانچه یک دسته شامل حد بالا یا حد پایین نباشد، **دسته باز** (**فاصله دسته باز**) نامیده می شود. به عنوان مثال دسته ۷۲ کیلوگرم و بالاتر، دسته باز می باشد.

نکته ۴: در مثال فوق، چون وزن ها با تقریب کیلوگرم ثبت شده اند، به عنوان مثال فاصله دسته ۷۱–۶۹ شامل تمام اندازه ها از ۶۸/۵ تا ۷۱/۵ خواهد بود. اعداد ۶۸/۵ و ۷۱/۵ ، **حدود دسته واقعی** یا **حدود دسته پیوسته** نامیده می شوند.

برای محاسبه حدود پیوسته یا واقعی، حاصل جمع حد بالای یک دسته با حد پایین دسته بعدی را تقسیم بر ۲ می کنیم:

$$\frac{68+69}{2}=68.5$$

$$\frac{71+72}{2}=71.5$$

نکته ۵: **اندازه فاصله دسته** یا **طول دسته** که با c نشان داده می شود، از تفاضل حدود پیوسته بالا و پایین یک دسته یا از تفاضل بین دو حد پایین دسته های متوالی یا دو حد بالای دسته های متوالی به دست می آید:

$$c=62.5-59.5=65.5-62.5=\ldots$$

یا

$$c=63-60=66-63=\ldots or 65-62=68-65=\ldots$$

نکته ۶: **متوسط دسته**، وسط فاصله دسته است که از حاصل جمع حد پایین و بالای هر دسته تقسیم بر ۲ به دست می آید. به عنوان مثال متوسط دسته اول به صورت زیر محاسبه می شود:

$$\frac{60+62}{2}=61$$

در محاسبات پارامترهای متمایل به مرکز و پراکندگی مانند انواع میانگین یا واریانس و ... از متوسط دسته که غالباً با x نشان داده می شود، استفاده می شود.

نحوه تشکیل جدول توزیع فراوانی

۱– ابتدا بزرگترین و کوچکترین عدد را در داده های خام مشخص کرده و دامنه را محاسبه می کنیم.

۲- دامنه را به تعداد دسته مناسب تقسیم می کنیم. معمولاً تعداد دسته مناسب بین ۵ تا ۲۰ دسته می باشد. فاصله دسته ها طوری انتخاب می شوند که متوسط دسته ها مطابق با داده های مشاهده شده باشند. با این کار حدود دسته ها با داده های مشاهده شده مطابقت نخواهند داشت.

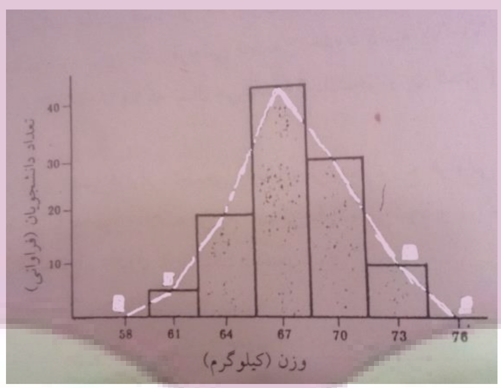
۳- تعداد مشاهده ها در هر دسته یعنی فراوانی دسته ها را تعیین می کنیم.

هیستوگرام

هیستوگرام نوعی نشان دادن نموداری توزیع فراوانی است. هیستوگرام یک نمودار ستونی عمودی است که در آن بین ستونها فاصله ای وجود ندارد. محور xها فاصله بین دسته ها و محور yها فراوانی دسته ها را نشان می دهد.

در واقع هیستوگرام شامل دسته ای از مستطیل هایی است که پایه آنها در محور xها قرار داشته و وسط آنها متوسط دسته را نشان می دهد. طول این قسمت، مساوی اندازه دسته می باشد و مساحت مستطیل معادل فراوانی دسته است. البته اگر همه دسته ها حدود دسته مساوی داشته باشند، ارتفاع مستطیل ها معادل فراوانی است.

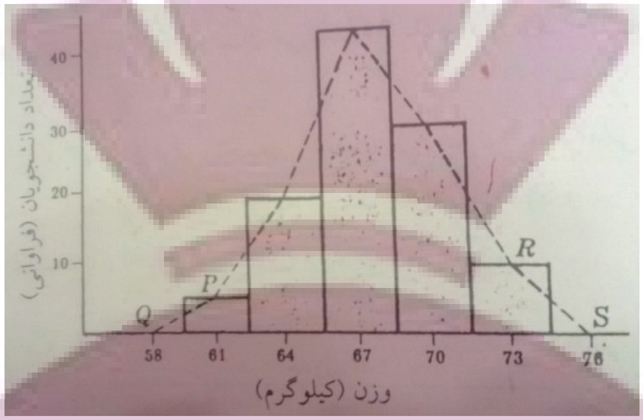
در شکل زیر هیستوگرام متناظر با توزیع فراوانی وزن دانشجویان نشان داده شده است:



پلی گن فراوانی

نمودار خطی است که در آن، فراوانی دسته نسبت به متوسط دسته رسم شده است. پلی گن فراوانی را می توان از وصل کردن نقاط وسط بالای مستطیل های هیستوگرام به دست آورد.

در شکل زیر پلی گن فراوانی متناظر با توزیع فراونی وزن دانشجویان رسم شده است. معمولاً قسمت های الحاقی RQ و RS را به متوسط دسته های ما قبل دسته اول و ما بعد دسته آخر که فراوانی های صفر دارند، وصل می کنند. در چنین حالتی، جمع سطوح مستطیل ها در هیستوگرام با جمع کل سطوح محدود شده توسط پلی گن فراوانی و محور xها مساوی خواهد بود.



فراوانی نسبی

فراوانی نسبی هر دسته از تقسیم فراوانی آن دسته بر کل فراوانی تمام دسته ها حاصل می شود و معمولاً به صورت درصد بیان می شود. مجموع فراوانی نسبی تمام دسته ها برابر با ۱ یا ۱۰۰ ٪ است.

مثلاً فراوانی نسبی دسته سوم جدول توزیع فراوانی وزن دانشجویان به صورت زیر محاسبه می شود:

$$\frac{42}{100} = 42\%$$

نکته: اگر فراوانی ها در جدول توزیع فراوانی با فراوانی های نسبی متناظر جایگزین شود، **توزیع فراوانی نسبی** یا **توزیع درصد** یا **جدول فراوانی نسبی** حاصل می شود.

وزن (کیلوگرم)	تعداد دانشجویان (فراوانی) f	فراوانی نسبی
۶۰–۶۲	۵	۰/۰۵
۶۳–۶۵	۱۸	۰/۱۸
۶۶–۶۸	۴۲	۰/۴۲
۶۹–۷۱	۲۷	۰/۲۷
۷۲–۷۴	۸	۰/۰۸
کل	۱۰۰	۱

برای نمایش تصویری این توزیع هم کافی است با تغییر معیار محور عمودی هیستوگرام یا پلی گن فراوانی از فراوانی به فراوانی نسبی اقدام کنیم که در این صورت نمودارهای حاصله، به ترتیب **هیستوگرام فراوانی نسبی** یا **هیستوگرام درصد** و **پلی گن فراوانی نسبی** یا **پلی گن درصد** نامیده می شوند.

فراوانی تجمعی

مجموع فراوانی های تمام دسته های ماقبل به علاوه فراوانی دسته مورد نظر، فراوانی تجمعی دسته مذکور نامیده می شود.

به عنوان مثال فراوانی تجمعی دسته دوم جدول توزیع فراوانی وزن دانشجویان به صورت زیر محاسبه می شود:

$$\sum f_2 = 5+18 = 23$$

نکته: جدولی که چنین فراوانی تجمعی را نشان بدهد، **توزیع فراوانی تجمعی**، **جدول فراوانی تجمعی** یا **توزیع تجمعی** نامیده می شود.

وزن (کیلوگرم)	تعداد دانشجویان (فراوانی) f	فراوانی تجمعی کمتر	فراوانی تجمعی بیشتر
۶۲-۶۰	۵	۰+۵=۵	۰+۸+۲۷+۴۲+۱۸+۵=۱۰۰
۶۵-۶۳	۱۸	۰+۵+۱۸=۲۳	۰+۸+۲۷+۴۲+۱۸=۹۵
۶۸-۶۶	۴۲	۰+۵+۱۸+۴۲=۶۵	۰+۸+۲۷+۴۲=۷۷
۷۱-۶۹	۲۷	۰+۵+۱۸+۴۲+۲۷=۹۲	۰+۸+۲۷=۳۵
۷۴-۷۲	۸	۰+۵+۱۸+۴۲+۲۷+۸=۱۰۰	۰+۸=۸
کل	۱۰۰		

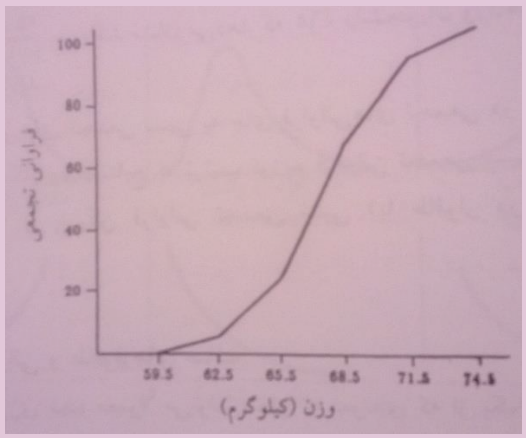
نکته ۱: در حالت فوق، در واقع فراوانی کل تمامی مقادیر کمتر از حد بالای هر دسته مفروض محاسبه می شود که کاربرد آن در تعیین تعداد افرادی است که صفت مورد نظر در آنها (در مثال های فوق صفت وزن) از یک حد مشخص کمتر است یا حداکثر آن مقدار مشخص را دارا هستند.

وزن (کیلوگرم)	تعداد دانشجویان (فراوانی) f
کمتر از ۵۹/۵	۰
کمتر از ۶۲/۵	۵
کمتر از ۶۵/۵	۲۳
کمتر از ۶۸/۵	۶۵
کمتر از ۷۱/۵	۹۲

کمتر از ۷۴/۵	۱۰۰

بر این اساس می توان مشخص کرد به عنوان مثال تعداد دانشجویانی که وزن آنها کمتر از ۶۸/۵ کیلوگرم است، ۶۵ نفراست (یا به عبارتی ۶۵ نفر حداکثر دارای وزن ۶۸/۵ کیلوگرم هستند).

نکته ۲: نموداری که نشان دهنده فراوانی تجمعی کمتر از حد بالای هر دسته نسبت به فراوانی تجمعی رسم می شود، **پلی گن فراوانی تجمعی** یا **طاقوار** نامیده می شود که در شکل زیر برای توزیع وزن دانشجویان رسم شده است:



نکته ۳: گاهی توزیع فراوانی تجمعی تمامی مقادیر بزرگتر یا مساوی با حد پایین هر دسته را در نظر می گیرند که **توزیع فراوانی بیشتر** نامیده می شود (حالت قبلی توزیع تجمعی کمتر نامیده می شود).

وزن (کیلوگرم)	تعداد دانشجویان (فراوانی) f
بیشتر از ۵۹/۵	۱۰۰
بیشتر از ۶۲/۵	۹۵
بیشتر از ۶۵/۵	۷۷
بیشتر از ۶۸/۵	۳۵
بیشتر از ۷۱/۵	۸
بیشتر از ۷۴/۵	۰

در اینجا هم کاربرد در تعیین تعداد افرادی است که صفت مورد نظر در آنها (در مثال های فوق صفت وزن) از یک حد مشخص بیشتر است یا حداقل آن مقدار مشخص را دارا هستند. مثلاً در جدول فوق ۷۷ نفر از دانشجویان بیشتر از ۶۵/۵ کیلوگرم وزن دارند یا ۷۷ نفر حداقل ۶۵/۵ کیلوگرم هستند.

نکته ۵: طاقوارهای متناظر با توزیع های تجمعی کمتر و بیشتر نیز **طاقوارهای کمتر و بیشتر** نامیده می شوند. اما زمانی که توزیع تجمعی یا طاقوار بدون ذکر کمتر و بیشتر مورد اشاره قرار می گیرد، منظور نوع کمتر است (حالت اول).

فراوانی تجمعی نسبی (تجمعی درصد)

این فراوانی از تقسیم فراوانی تجمعی به کل فراوانی به دست می آید.

به عنوان مثال فروانی تجمعی نسبی وزن های کمتر از ۶۵/۵ به صورت زیر به دست می آید:

$$\frac{23}{100} = 23\%$$

و نشان می دهد ۲۳ درصد دانشجویان وزن هایی کمتر از ۶۵/۵ کیلوگرم دارند.

نکته: چنانچه فراوانی های تجمعی نسبی به جای فراوانی های تجمعی در جدول قرار بگیرند، **توزیع فراوانی تجمعی نسبی** یا **توزیع تجمعی درصد** نامیده می شوند و به همین ترتیب اگر جای پلی گن فراوانی تجمعی قرار بگیرند، **پلی گن فراوانی تجمعی نسبی** یا **طاقوار درصد** نامیده می شوند.

منحنی های فراوانی و طاقوارهای صاف

چنانچه تعداد مشاهدات بسیار زیاد شود، در داده های متصل به تدریج فاصله دسته ها کوچکتر و پاره خطهای نمودار خطی پلی گن فراوانی یا پلی گن فراوانی نسبی، خیلی کوچک و شکسته و به سمت منحنی میل می کنند و در نتیجه منحنی های فراوانی یا منحنی های فراوانی نسبی به وجود می آیند. به عبارتی این منحنی ها با صاف کردن نمودارهای پلی گن تقریب می شوند و با افزایش اندازه نمونه این تقریب افزایش می یابد. به همین دلیل منحنی فراوانی، پلی گن فراوانی صاف نیز نامیده می شوند.

به همین ترتیب طاقوارهای صاف از صاف کردن طاقوارها به دست می آیند.

انواع منحنی ها

منحنی ها شکل های ویژه ای دارند، بعضی متقارن یا زنگی شکل، برخی نامتقارن یا چوله، بعضی به شکل J یا معکوس شکل J، برخی U شکل، بعضی دارای دو مُد و برخی دارای چند مُد (بیش از دو ماگزیمم) هستند.

جزوه تکمیلی

آمار کاربردی ۱

(رشته حسابداری)

تهیه و تدوین

امیرحسین نوربخش

عضو هیأت علمی

دانشگاه پیام نور مرکز تاکستان

بخش سوم

(شاخص‌های مرکزی)

اندیس

علامت x_i (که x اندیس i خوانده می شود)، هر N مقدار x، یعنی $x_1, x_2, x_3, ..., x_N$ را که در مورد متغیر x فرض شده است را نشان می دهد. i می تواند مقادیر ۱، ۲، ۳، ... و N را شامل شود.

علامت جمع

علامت جمع در ریاضی به صورت Σ (سیگما) معرفی شده است. اگر متغیری با x نشان داده شود و مشاهدات متوالی متغیر به صورت $x_1, x_2, x_3, ..., x_N$ نوشته شود، برای جمع این مشاهدات از علامت Σ به صورت زیر استفاده می کنیم:

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + x_3 + ... + x_N$$

مثال: اگر $x_1 = 4, x_2 = 5, x_3 = 7, x_4 = 9$ باشد، مطلوب است محاسبه عبارات زیر:

$$\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4 = 4 + 5 + 7 + 9 = 25$$

$$\sum_{i=2}^{3} x_i = x_2 + x_3 = 5 + 7 = 12$$

قوانین جمع

۱- جمع حاصل ضرب یک عدد ثابت در یک متغیر برابر است با حاصل ضرب عدد ثابت در جمع متغیر.

$$\sum_{i=1}^{N} a x_i = a \sum_{i=1}^{N} x_i$$

مثال:

$$\sum_{i=2}^{3} 5 x_i = 5 \sum_{i=2}^{3} x_i$$

۲- جمع N مرتبه از یک عدد ثابت برابر است با حاصل ضرب N در عدد ثابت.

$$\sum_{i=1}^{N} a = N.a$$

مثال:

$$\sum_{i=1}^{4} 5 = 4 \times 5 = 20$$

$$\sum_{i=2}^{3} 6 = 2 \times 6 = 12$$

۳- جمع یک جمله که خود شامل دو عبارت یا بیشتر می باشد، برابر است با مجموع تک تک عبارات.

$$\sum_{i=1}^{N}(x_i+y_i+z_i)=\sum_{i=1}^{N}x_i+\sum_{i=1}^{N}y_i+\sum_{i=1}^{N}z_i$$

مثال: چنانچه $x_1=2, x_2=4, x_3=3, x_4=1$ باشد، حاصل عبارات زیر را به دست آورید:

$$\sum x_i^{2}=x_1^{2}+x_2^{2}+x_3^{2}+x_4^{2}=2^2+4^2+3^2+1^2$$

$$(\sum x_i)^2=(x_1+x_2+x_3+x_4)^2=(2+4+3+1)^2$$

$$\sum(x_i+3)=(x_1+3)+(x_2+3)+(x_3+3)+(x_4+3)=(2+3)+(4+3)+(3+3)+(1+3)$$

$$\sqrt{\sum(x_i+3)}=\sqrt{(x_1+3)+(x_2+3)+(x_3+3)+(x_4+3)}=\sqrt{(2+3)+(4+3)+(3+3)+(1+3)}$$

$$\sum(x_i+3)^2=(x_1+3)^2+(x_2+3)^2+(x_3+3)^2+(x_4+3)^2=(2+3)^2+(4+3)^2+(3+3)^2+(1+3)^2$$

شاخص‌های مرکزی (متوسط ها)

یکی از بهترین روش ها برای توصیف داده ها، استفاده از معیارهای وسط گرایی است که مهمترین آنها عبارتند از میانگین، میانه و مد.

هریک از این معیارها در دو حالت قابل بررسی هستند:

۱- وقتی رشته ای از اعداد داشته باشیم.

۲- در جدول توزیع فراوانی

تذکر: فرمولهایی که برای معیارهای وسط گرایی یا پراکندگی جداول توزیع فراوانی ذکر می شوند، هم برای متغیرهای پیوسته و هم برای متغیرهای گسسته یکسانند.

میانگین حسابی

این میانگین را با x_N یا $\bar{x}$ یا m نشان می دهند و

الف- برای رشته ای از اعداد مانند $x_1, x_2, x_3, \ldots, x_N$ به صورت زیر تعریف می کنند:

$$\bar{x}=\frac{x_1+x_2+x_3+\ldots+x_N}{N}=\frac{\sum_{i=1}^{N}x_i}{N}=\frac{\sum x}{N}$$

ب- و در جدول توزیع فراوانی به صورت زیر تعریف می شود:

$$\bar{x}=\frac{f_1x_1+f_2x_2+f_3x_3+\ldots+f_Nx_N}{N}=\frac{\sum_{i=1}^{N}f_ix_i}{N}=\frac{\sum fx}{N}$$

$$N=\sum f$$

در رابطه فوق، f فراوانی هر دسته جدول توزیع فراوانی و xi متوسط هر دسته می باشد.

مثال: میانگین اعداد ۳، ۵، ۷ و ۹ را به دست آورید.

$$\bar{x} = \frac{\sum x}{N} = \frac{3+5+7+9}{4} = 6$$

مثال: در جدول توزیع فراوانی زیر میانگین حسابی را محاسبه کنید.

دسته	حد وسط (متوسط دسته)	فراوانی (f_i)
۱۷-۲۰	۱۸/۵	۳۰
۲۰-۲۳	۲۱/۵	۲۰
۲۳-۲۶	۲۴/۵	۷
۲۶-۲۹	۲۷/۵	۳
کل		۶۰

$$\bar{x} = \frac{\sum fx}{N} = \frac{(30\times 18.5)+(20\times 21.5)+(7\times 24.5)+(3\times 27.5)}{60}$$

میانگین وزنی

یکی از انواع میانگین حسابی، میانگین وزنی است که در محاسبه آن تعداد افراد یا اشیاء (وزن افراد یا اشیاء به معنای درجه اهمیتشان) تأثیر دارد. میانگین وزنی را با m_w نشان داده و به صورت زیر تعریف می کنند:

$$m_w or \bar{x} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_k\, x_k}{w_1 + w_2 + w_3 + \ldots + w_k} = \frac{\sum wx}{\sum w}$$

مثال: اگر در یک کلاس، اکیپ اول شامل ۲۰ دانشجو با معدل ۱۵، اکیپ دوم شامل ۳۰ دانشجو با معدل ۱۴ و اکیپ سوم شامل ۴۰ دانشجو با معدل ۱۶ باشند، میانگین معدل کلاس چند است؟

$$m_w = \frac{w_1x_1 + w_2x_2 + w_3x_3}{w_1 + w_2 + w_3}$$

$$m_w = \frac{(20\times 15)+(30\times 14)+(40\times 16)}{20+30+40} = 15.11$$

اصول میانگین حسابی

۱- جمع جبری انحرافات دسته ای از اعداد از میانگین حسابی صفر است.

$$\sum\left(x_i - \bar{x}\right) = 0$$

به عنوان مثال اگر رشته ای از اعداد شامل ۳، ۵ و ۷ داشته باشیم، خواهیم داشت:

$$\bar{x}=\frac{\sum x}{N}=\frac{3+5+7}{3}=5$$

$$\sum\left(x_i-\bar{x}\right)=(3-5)+(5-5)+(7-5)=-2+0+2=0$$

۲- مجموع مربعات انحرافات دسته ای از اعداد از میانگین حداقل است (مجموع مربعات انحرافات دسته ای از اعداد از هر عددی مانند a حداقل است اگر و فقط اگر $a=\bar{x}$ باشد).

$$\sum\left(x_i-\bar{x}\right)^2=Min$$

در مثال قبلی این اصل را را بررسی می کنیم:

$$\begin{cases}\bar{x}=\frac{\sum x}{N}=\frac{3+5+7}{3}=5\\ a=\bar{x}=5\rightarrow\sum\left(x_i-\bar{x}\right)^2=(3-5)^2+(5-5)^2+(7-5)^2=4+0+4=8\\ a=3\rightarrow\sum\left(x_i-\bar{x}\right)^2=(3-3)^2+(5-3)^2+(7-3)^2=0+4+16=20\quad\rightarrow\sum\left(x_i-\bar{x}\right)^2=Min\\ a=7\rightarrow\sum\left(x_i-\bar{x}\right)^2=(3-7)^2+(5-7)^2+(7-7)^2=16+4+0=20\\ a=1\rightarrow\sum\left(x_i-\bar{x}\right)^2=(3-1)^2+(5-1)^2+(7-1)^2=4+16+36=56\end{cases}$$

۳- اگر f_1 عدد میانگین m_1، f_2 عدد میانگین m_2 و f_k عدد میانگین m_k داشته باشند، میانگین تمام اعداد عبارتند از:

$$\bar{x}=\frac{f_1m_1+f_2m_2+f_3m_3+\ldots+f_km_k}{f_1+f_2+f_3+\ldots+f_k}$$

مثال: شرکتی ۸۰ کارمند دارد. ۶۰ نفر از آنها ۷ دلار در ساعت و ۲۰ نفر دیگر ۴ دلار در ساعت دریافت می کنند. میانگین دریافتی کارمندان در ساعت چقدر است؟

$$\bar{x}=\frac{(60\times7)+(20\times4)}{80}=6.25\$/\mathrm{h}$$

۴- اگر A میانگین فرضی یا هر عدد فرضی دیگری باشد و $d_i=x_i-A$ انحراف x_i از میانگین فرضی A باشد، میانگین حسابی عبارت خواهد بود با:

$$\bar{x}=A+\frac{\sum_{i=1}^{N}d_i}{N}$$

و در جدول توزیع فراوانی خواهیم داشت:

$$\bar{x} = A + \frac{\sum_{i=1}^{k} f_i d_i}{\sum f_i} = A + \frac{\sum fd}{N}$$

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

محاسبه میانگین در جداول توزیع فراوانی به روشهای کوتاه و کدگذاری

چنانچه در جدول توزیع فراوانی، xi متوسط دسته و A متوسط هر دسته مفروض باشد، $d_i = x_i - A$ را به عنوان انحراف x_i از A در نظر گرفته و میانگین را از رابطه زیر محاسبه می کنیم:

$$\bar{x} = A + \frac{\sum fd}{N}$$

محاسبه میانگین به این روش را روش کوتاه نامیده و محاسبه میانگین با فرمول اصلی میانگین ($\bar{x} = \frac{\sum fx}{N}$) را روش طولانی می نامند. در روش کوتاه، با توجه به اینکه از مقادیر متوسط هر دسته، مقدار A کم می شود، در محاسبات، اعداد کوچکتر و کار ساده تری داریم.

حال در صورتی که فاصله دسته ها همگی اندازه مساوی c داشته باشند، در این صورت انحراف $d_i = x_i - A$ برابر با cu_i خواهد بود ($d_i = cu_i$). u_i اعداد صحیح مثبت، منفی یا صفر می باشد ($\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$). در این حالت خواهیم داشت:

$$\bar{x} = A + c\left(\frac{\sum_{i=1}^{k} f_i u_i}{N}\right) = A + c\left(\frac{\sum fu}{N}\right)$$

محاسبه میانگین به روش فوق را روش کدگذاری می نامند که روش خیلی کوتاهی است و با اعداد بسیار کوچکتر از اعداد اصلی جدول سروکار خواهیم داشت، در نتیجه محاسبات ساده تر هستند. مجدداً تاکید می شود که این روش فقط در جداول توزیع فراوانی که فاصله دسته ها مساوی می باشند، به کار می رود.

مثال: در مثال وزن ۱۰۰ دانشجو که در فصل ۲ به آن اشاره شد، میانگین را به روشهای طولانی (با فرمول اصلی)، کوتاه و کد گذاری محاسبه کنید.

وزن (کیلوگرم)	متوسط دسته (x)	فراوانی (f)	fx	d=x-A	fd	$u=\frac{d}{c}$	fu
۶۰-۶۲	۶۱	۵	۳۰۵	-۶	-۳۰	-۲	-۱۰
۶۳-۶۵	۶۴	۱۸	۱۱۵۲	-۳	-۵۴	-۱	-۱۸
۶۶-۶۸	۶۷	۴۲	۲۸۱۴	۰	۰	۰	۰
۶۹-۷۱	۷۰	۲۷	۱۸۹۰	۳	۸۱	۱	۲۷
۷۲-۷۴	۷۳	۸	۵۸۴	۶	۴۸	۲	۱۶
کل		$N=\sum f=100$	$\sum fx=6745$		$\sum fd=45$		$\sum fu=15$

برای محاسبه میانگین به روش طولانی یا همان روش اصلی به داده های ستون چهارم جدول فوق نیاز داریم:

$$\bar{x}=\frac{\sum fx}{N}\rightarrow \bar{x}=\frac{6745}{100}=67.45$$

ملاحظه می شود که در محاسبه ستون چهارم با اعداد نسبتاً بزرگی مواجهیم.

برای محاسبه میانگین به روش کوتاه، ابتدا باید یک عدد فرضی A از میان متوسط دسته ها انتخاب کنیم. مجاز به انتخاب هر کدام از متوسط دسته ها هستیم، اما برای راحتی محاسبات بهتر است، متوسط دسته وسطی یعنی ۶۷ را انتخاب کنیم. سپس در ستون پنجم انحراف متوسط دسته ها را از عدد انتخابی (از ۶۷) محاسبه (d) و سپس در ستون ششم حاصل ضرب فراوانی را در انحراف های محاسبه شده (fd) به دست می آوریم:

$$\bar{x}=A+\frac{\sum fd}{N}\rightarrow \bar{x}=67+\frac{45}{100}=67.45$$

دیده شد که در محاسبه ستون ششم با اعداد کوچکتری مواجه بودیم.

با توجه به اینکه فاصله دسته ها همه مساوی و برابر با مقدار c=3 است، می توانیم از روش کدگذاری استفاده کنیم. ابتدا در ستون هفتم مقدار u را با توجه به رابطه $d=cu\rightarrow u=\frac{d}{c}$ محاسبه و سپس در ستون هشتم حاصل ضرب فراوانی در u را به دست می آوریم:

$$\bar{x}=A+c(\frac{\sum ud}{N})\rightarrow \bar{x}=67+3(\frac{15}{100})=67.45$$

در روش کدگذاری دیده شد، در محاسبه ستون آخر با اعداد ساده ای مواجه بودیم.

تذکر مهم: اشتباهی که اکثر دانشجویان مرتکب می شوند، در انتخاب عدد A می باشد. توجه شود که A فقط از اعداد ستون متوسط داده ها (x) انتخاب می شود نه از فراوانی یا fها.

نکته: چنانچه در محاسبه uها، اعداد به صورت صحیح و پشت سر هم ($\ldots,-3,-2,-1,0,1,2,3,\ldots$) به دست نیایند، یعنی جایی در محاسبات دچار اشتباه شده ایم، به خصوص در انتخاب عدد فرضی A.

میانگین هندسی

میانگین هندسی دسته ای از اعداد را با m_g یا G نمایش داده و به صورت زیر تعریف می کنند:

$$m_g or G = \sqrt[N]{x_1.x_2.x_3...x_N}$$

$$\log G = \frac{\log x_1 + \log x_2 + ... + \log x_N}{N} = \frac{\sum \log x_i}{N}$$

میانگین هندسی برای مشاهداتی به کار می رود که در آن نسبت هر دو عدد متوالی ثابت یا تقریباً ثابت باشد.

مثال: میانگین هندسی اعداد ۲۰۰۰، ۴۰۰۰، ۸۵۰۰ و ۱۵۰۰۰ را به دست آورید.

$$\log G = \frac{\log 2000 + \log 4000 + \log 8500 + \log 15000}{4}$$

میانگین هارمونیک

میانگین هارمونیک اغلب برای اندازه گیری متوسط سرعت ها وقتی فواصل مربوط به آنها مساوی باشد، به کار می رود. میانگین هارمونیک را با H یا m_h نشان داده و به صورت زیر تعریف می کنند:

$$m_h or H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_N}} = \frac{N}{\sum \frac{1}{x}}$$

مثال: هواپیمایی یک فاصله ۹۰۰ مایلی را می پیماید. اگر این هواپیما، ثلث اول و ثلث سوم این فاصله را با سرعت ۲۵۰ مایل در ساعت و ثلث دوم را با سرعت ۳۰۰ مایل در ساعت طی کند، متوسط سرعت هواپیما چقدر است؟

$$H = \frac{N}{\sum \frac{1}{x}}$$

$$H = \frac{3}{\frac{1}{250} + \frac{1}{300} + \frac{1}{250}} = 264.70 mil / h$$

میانه

وقتی رشته ای از اعداد داریم، برای محاسبه میانه، ابتدا اعداد را از کوچک به بزرگ مرتب می کنیم. چنانچه تعداد اعداد فرد باشد، میانه برابر با عدد وسطی و اگر تعداد اعداد زوج باشد، میانه برابر با میانگین دو عدد وسطی است.

مثال: میانه اعداد ۳، ۵، ۳، ۲ و ۷ را به دست بیاورید.

$$2,3,3,5,7 \rightarrow M_d = 3$$

مثال: میانه اعداد ۶، ۳، ۴، ۷، ۳ و ۸ را به دست بیاورید.

$$3,3,4,6,7,8 \rightarrow M_d = \frac{4+6}{2} = 5$$

برای محاسبه میانه در جدول توزیع فراوانی از رابطه زیر استفاده می شود:

$$M_d = L_1 + c\left(\frac{\frac{N}{2} - (\sum f)}{f_{M_d}}\right)$$

L_1 حد پایین دسته میانه (دسته ای که شامل میانه است)، N تعداد کل داده ها، $(\sum f)$ جمع فراوانی های تمامی دسته های ماقبل دسته میانه ، f_{M_d} فراوانی دسته میانه و c فاصله دسته میانه است.

مثال: در جدول فراوانی وزن ۱۰۰ دانشجو میانه را محاسبه کنید.

وزن (کیلوگرم)		فراوانی (f)	فراوانی تجمعی
۶۰-۶۲	۵۹/۵-۶۲/۵	۵	۵
۶۳-۶۵	۶۲/۵-۶۵/۵	۱۸	۲۳
۶۶-۶۸	۶۵/۵-۶۸/۵	۴۲	۶۵
۶۹-۷۱	۶۸/۵-۷۱/۵	۲۷	۹۲
۷۲-۷۴	۷۱/۵-۷۴/۵	۸	۱۰۰
کل		۱۰۰	

برای محاسبه میانه در جدول توزیع فراوانی، ابتدا فراوانی تجمعی را محاسبه می کنیم و دسته ای را که میانه در آن قرار می گیرد، تعیین می کنیم. در این مثال کل فراوانی ۱۰۰ است، پس داده پنجاهم میانه است و داده پنجاهم در دسته سوم قرار می گیرد. ضمن اینکه حدود پیوسته دسته ها را به دست می آوریم.

$$M_d = L_1 + c\left(\frac{\frac{N}{2} - (\sum f)}{f_{M_d}}\right)$$

$$\mathrm{M}_d = 65.5 + 3\left(\frac{\frac{100}{2} - 23}{42}\right) = 67.43$$

تذکر: دسته میانه اولین دسته ای است که فراوانی تجمعی آن از $\frac{N}{2}$ بیشتر است.

مد

عدد یا اعدادی که بیش از همه تکرار شده باشد، مد نام دارند. ممکن است در رشته ای از اعداد یک مد یا بیش از یک مد وجود داشته باشد. همچنین ممکن است مد وجود نداشته باشد.

مثال: در رشته اعداد زیر مد کدام است؟

$$3,5,7,9,2,2 \rightarrow M_o = 2$$
$$3,3,5,7,9,2,2 \rightarrow M_o = 3,2$$

در جدول توزیع فراوانی مد ار طریق رابطه زیر به دست می آید:

$$M_o = L_1 + c\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)$$

در رابطه فوق، L_1 حد پایینی دسته مددار (دسته ای که بیشترین فراوانی را دارد)، Δ_1 افزونی فراوانی دسته مددار روی فراوانی دسته ماقبل، Δ_2 افزونی دسته مد دار روی فراوانی دسته مابعد و c فاصله دسته مددار می باشد.

مثال: در جدول توزیع فراوانی وزن ۱۰۰ دانشجو، مد را پیدا کنید.

وزن (کیلوگرم)		فراوانی (f)
۶۰-۶۲	۵۹/۵-۶۲/۵	۵
۶۳-۶۵	۶۲/۵-۶۵/۵	۱۸
۶۶-۶۸	۶۵/۵-۶۸/۵	۴۲
۶۹-۷۱	۶۸/۵-۷۱/۵	۲۷
۷۲-۷۴	۷۱/۵-۷۴/۵	۸
کل		۱۰۰

در این جدول، مد در دسته سوم قراردارد (بیشترین فراوانی در این دسته وجود دارد) و:

$$M_o = L_1 + c\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)$$

$$\Delta_1 = 42 - 18 = 24$$

$$\Delta_2 = 42 - 27 = 15$$

$$M_o = 65.5 + 3\left(\frac{24}{24+15}\right) = 67.35$$

رابطه بین میانگین، میانه و مد

در منحنی های فراوانی تک مددار که کمی نامتقارن (چوله) هستند، رابطه زیر برقرار است:

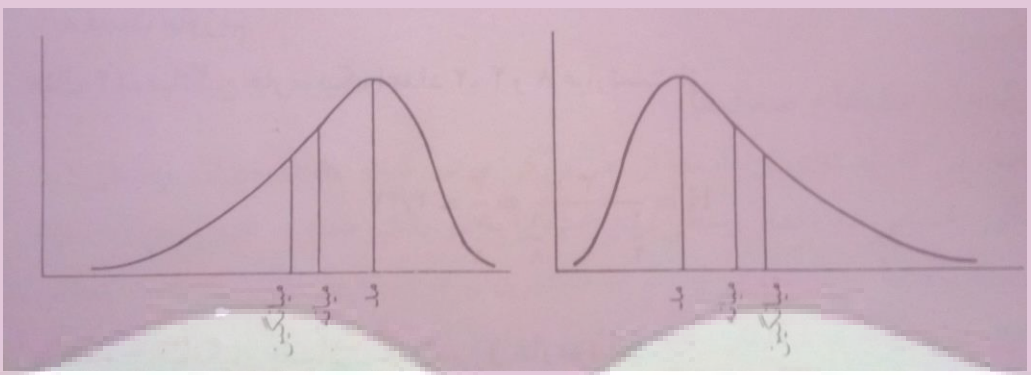
$$\bar{x} - M_o = 3(\bar{x} - \mathrm{M}_d)$$

مثال: در یک منحنی نامتقارن، چنانچه میانه ۶۲ و میانگین ۶۰ باشد، مد چقدر است؟

$$\bar{x} - M_o = 3(\bar{x} - \mathrm{M}_d)$$

$$60 - M_o = 3(60 - 62) \rightarrow M_O = 66$$

موقعیت نسبی میانگین، میانه و مد در منحنی های نامتقارن در شکل زیر نشان داده شده است:



در شکل فوق، منحنی سمت راست چوله مثبت و منحنی سمت چپ چوله منفی است.

مثال: در مثال قبلی، با توجه به مقادیر به دست آمده برای میانه، مد و میانگین، نوع چولگی را مشخص کنید.

با توجه به اینکه مقدار مد از میانگین بیشتر است، بنابراین منحنی چوله به سمت چپ یا چوله منفی است. به عبارت دیگر دم آن به سمت چپ کشیده شده است.

نکته: در منحنی های نامتقارن، مقدار میانه همواره بین میانگین و مد قرار می گیرد. اگر مد بزرگتر از همه باشد، منحنی چوله منفی یا چوله به سمت چپ است و چنانچه میانگین از همه بزرگتر باشد، منحنی چوله مثبت یا چوله به سمت راست می باشد.

رابطه بین میانگین حسابی، هندسی و هارمونیک

در دسته ای از اعداد مثبت، همواره رابطه زیر بین میانگین های حسابی، هندسی و هارمونیک برقرار است:

$$H \le G \le \bar{x}$$

یعنی میانگین حسابی بزرگتر مساوی میانگین هندسی و میانگین هندسی بزگتر مساوی میانگین هارمونیک می باشد.

ریشه میانگین مربعات (RMS)

ریشه میانگین مربعات یا میانگین درجه دوم دسته ای از اعداد به صورت زیر تعریف می شود:

$$RMS = \sqrt{\overline{x^2}} = \sqrt{\frac{\sum x^2}{N}}$$

مثال: ریشه میانگین مربعات دسته اعداد ۱، ۳، ۴، ۵ و ۷ را به دست بیاورید.

$$RMS = \sqrt{\frac{\sum x^2}{N}}$$

$$RMS = \sqrt{\frac{1^2 + 3^2 + 4^2 + 5^2 + 7^2}{5}} = 4.47$$

چارک ها، دهک ها و صدک ها

در دسته ای از اعداد که به ترتیب بزرگی مرتب شده اند، همان گونه که قبلاً اشاره شد، مقدار وسطی یا میانگین دو مقدار وسطی میانه است که دسته اعداد را به دو قسمت مساوی تقسیم می کند. با بسط این ایده، مقادیری که دسته داده ها را به چهار قسمت مساوی تقسیم می کنند، چارک نامیده شده که به ترتیب با Q_1، Q_2 و Q_3 نشان داده می شوند (به ترتیب چارک اول، دوم و سوم). همچنین مقادیری که دسته داده ها را به ۱۰ بخش مساوی تقسیم می کنند، دهک نامیده شده که که به ترتیب با D_1، D_2، ... و D_9 نشان داده می شوند و در نهایت مقادیری که دسته داده ها را به ۱۰۰ قسمت مساوی تقسیم می کنند، صدک نامیده شده که به ترتیب با P_1، P_2، ... و P_{99} نشان داده می شوند.

به طور کلی چارک ها، دهک ها، صدک ها و کلاً مقادیری که دسته ای از اعداد را به قسمتهای مساوی تقسیم می کنند، چندک نامیده می شوند.

نکته: از بین پارامترهای وسط گرایی، میانگین برای گزارش نمودن داده ها از بقیه بهتر است، زیرا در محاسبه میانگین از تک تک داده ها استفاده می شود. اگر در بین داده ها، داده پرت داشته باشیم، میانه و مد پارامترهای مناسب تری خواهند بود.

مثال: در همان جدول توزیع فراوانی وزن ۱۰۰ دانشجو، چارک سوم، دهک دوم و صدک هشتادم را به دست آورید.

وزن (کیلوگرم)		فراوانی (f)	فراوانی تجمعی
۶۰-۶۲	۵۹/۵-۶۲/۵	۵	۵
۶۳-۶۵	۶۲/۵-۶۵/۵	۱۸	۲۳
۶۶-۶۸	۶۵/۵-۶۸/۵	۴۲	۶۵
۶۹-۷۱	۶۸/۵-۷۱/۵	۲۷	۹۲
۷۲-۷۴	۷۱/۵-۷۴/۵	۸	۱۰۰
کل		۱۰۰	

برای محاسبه چندک ها هم مانند محاسبه میانه عمل می شود. ابتدا حدود پیوسته (واقعی) دسته ها را به دست آورده وسپس فراوانی های تجمعی را محاسبه می کنیم.

در مرحله بعدی باید پیدا کنیم چندک مورد نظر در کدام دسته قرار می گیرد. در مرحله آخر با رابطه ای شبیه به فرمول میانه، چندک مذکور را محاسبه می کنیم:

به عنوان مثال برای محاسبه چارک سوم ابتدا مشخص می کنیم چندمین داده در بین داده های مثال فوق است:

$$\frac{3N}{4}=\frac{3\times100}{4}=75$$

یعنی چارک سوم، ۷۵اُمین داده می باشد. حال باید پیدا کنیم در کدام دسته قرار می گیرد که این مستلزم داشتن فراوانی های تجمعی است.

با توجه به ستون چهارم جدول فوق، تا دسته سوم، ۶۵ داده وجود دارد، پس چارک سوم در دسته چهارم قرار می گیرد (۷۵اُمین داده). یعنی چارک سوم عددی بین ۶۸/۵ و ۷۱/۵ (حدود واقعی دسته) می باشد. به عبارت دیگر باید مقداری به ۶۸/۵ (حد پایین دسته شامل چندک مورد نظر) افزود که این همان اساس فرمول میانه است. یعنی ۱۰اُمین داده ($75-65=10$) بعد از داده ۶۵اُم از ۲۷ داده (فراوانی دسته ای که چندک مورد نظر در آن قرار گرفته) که متعلق به دسته چهارم است ($\frac{75-65}{27}$) را با توجه به فاصله دسته c=3 محاسبه ($3\times(\frac{75-65}{27})$) و به حد پایینی دسته چهارم (۶۸/۵) اضافه می کنیم:

$$Q_3=68.5+3\times(\frac{75-65}{27})=69.61$$

یعنی چارک سوم یا ۷۵اُمین داده در جدول توزیع فراوانی فوق، عدد ۶۹/۶۱ می باشد. اگر توضیحات فوق را بخواهیم با فرمولی شبیه آنچه در مورد میانه ارائه شد، بیان کنیم، خواهیم داشت:

$$Q_3 = L_1 + c \times \left(\frac{\frac{3N}{4} - \left(\sum f\right)}{f_{Q_3}}\right)$$

که فرمولی کاملاً شبیه فرمول میانه است و در آن L1 حد پایین دسته چارک سوم (دسته ای که شامل چارک سوم است)، N تعداد کل داده ها، $\left(\sum f\right)$ جمع فراوانی های تمامی دسته های ماقبل دسته چارک سوم ، f_{Q_3} فراوانی دسته چارک سوم و c فاصله دسته چارک سوم است.

با این منطق، برای تمام چندک ها می توان فرمول ارائه داد.

برای دهک دوم و صدک هشتادم هم به همین ترتیب خواهیم داشت:

$$\frac{2N}{10} = \frac{2 \times 100}{10} = 20$$

$$D_2 = 62.5 + 3 \times \left(\frac{20-5}{18}\right) = 65$$

یعنی دهک دوم در دسته دوم قرا گرفته است، به عبارت دیگر داده ۲۰اُم ۶۵ است.

$$\frac{80N}{100} = \frac{80 \times 100}{100} = 80$$

$$P_{80} = 68.5 + 3\left(\frac{80-65}{27}\right) = 70.17$$

در اینجا هم صدک هشتادم، ۸۰اُمین داده است که در دسته چهارم قرار گرفته و مقدار آن ۷۰/۱۷ می باشد.

بخش چهارم

(شاخص‌های پراکندگی)

شاخص‌های پراکندگی (پراکنش)

میزان پراکندگی داده ها در حول میانگین یا مرکز را نشان می دهند. در یک تحقیق علمی حتماً باید به همراه یک پارامتر وسط گرایی، یک پارامتر پراکنش نیز ذکر شود.

به عنوان مثال دو گروه از داده ها را در نظر می گیریم:

گروه ۱:	$1,5,9$	$m=5$
گروه ۲:	$3,5,7$	$m=5$

در دو گروه داده های بالا، اگرچه میانگین این دو گروه با هم برابر است، ولی میزان پراکندگی داده ها در اطراف میانگین متفاوت است و در گروه اول این پراکندگی بیشتر است.

انواع پارامترهای پراکنش

۱- دامنه (Range)

برای محاسبه دامنه که با D نشان داده می شود، تفاضل بزرگترین و کوچکترین داده را به دست می آوریم.

مثال: در رشته اعداد ۳، ۵، ۷، ۳ و ۲، دامنه را محاسبه کنید.

$D=7-2=5$

۲- میانگین انحرافات

میانگین انحرافات (انحرافات متوسط) در دسته ای از اعداد شامل $x_1, x_2, \dots, x_N$ با MD نشان داده می شود و به صورت زیر تعریف می شود:

$$MD=\frac{\sum_{i=1}^{N}|x_i-\bar{x}|}{N}=\frac{\sum|x-\bar{x}|}{N}=\overline{|x-\bar{x}|}$$

مثال: میانگین انحرافات ۱۰، ۱۶، ۲۰، ۱۷، ۱۴ و ۱۱ را به دست آورید.

$$\bar{x}=\frac{\sum x}{N}\rightarrow \bar{x}=\frac{11+14+17+20+16+10}{6}=14.67$$

$$MD=\frac{\sum|x-\bar{x}|}{N}\rightarrow MD=\frac{|11-14.67|+\dots+|10-14.67|}{6}=3$$

در جدول توزیع فراوانی میانگین انحرافات به صورت زیر محاسبه می شود:

$$MD = \frac{\sum_{i=1}^{k} f_i \,|\, x_i - \bar{x} \,|}{N} = \frac{\sum f \,|\, x - \bar{x} \,|}{N}$$

$$N = \sum_{i=1}^{k} f_i = \sum f$$

مثال: در جدول توزیع فراوانی زیر میانگین انحرافات را به دست آورید.

دسته	حد وسط (متوسط دسته)	فراوانی (f_i)
۲۰-۱۷	۱۸/۵	۳۰
۲۳-۲۰	۲۱/۵	۲۰
۲۶-۲۳	۲۴/۵	۷
۲۹-۲۶	۲۷/۵	۳
کل		۶۰

$$\bar{x} = \frac{\sum fx}{N} = \frac{(30\times18.5)+(20\times21.5)+(7\times24.5)+(3\times27.5)}{60} = 20.65$$

$$MD = \frac{\sum f \,|\, x - \bar{x} \,|}{N}$$

$$MD = \frac{[30\times|18.5-20.65|]+[20\times|21.5-20.65|]+[7\times|24.5-20.65|]+[3\times|27.5-20.65|]}{60} = 2.15$$

نکته: گاهی میانگین انحرافات به صورت قدر مطلق انحرافات از میانه یا متوسط دیگری به جای میانگین تعریف می شود و نکته جالب این که جمع ، زمانی حداقل خواهد بود که a میانه باشد.

۳- انحراف معیار (Standard Deviation)

انحراف معیار دسته ای از اعداد شامل $x_1, x_2, ..., x_N$ با s نشان داده می شود و به صورت زیر تعریف می شود:

$$s = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N}} = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2}{N}} = \sqrt{\overline{\left(x - \bar{x}\right)^2}}$$

مثال: در رشته اعداد ۱۰، ۱۶، ۲۰، ۱۷، ۱۴ و ۱۱ انحراف معیار را محاسبه کنید.

$$\bar{x} = \frac{\sum x}{N} \rightarrow \bar{x} = \frac{11+14+17+20+16+10}{6} = 14.67$$

$$s=\sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{N}}$$

$$s=\sqrt{\frac{\left(11-14.67\right)^2+\left(14-14.67\right)^2+\ldots+\left(10-14.67\right)^2}{6}}=3.45$$

در جدول توزیع فراوانی انحراف معیار به صورت زیر تعریف می شود:

$$s=\sqrt{\frac{\sum_{i=1}^{k} f_i\left(x_i-\bar{x}\right)^2}{N}}=\sqrt{\frac{\sum f\left(x-\bar{x}\right)^2}{N}}=\sqrt{\overline{\left(x-\bar{x}\right)^2}}$$

$$N=\sum_{i=1}^{k} f_i=\sum f$$

که f همان فراوانی هر دسته در جدول توزیع فراوانی است.

نکته: گاهی انحراف معیار داده های نمونه با جایگذاری N-1 به جای N در مخرج تعریف می شود، زیرا مقدار حاصله برآورد بهتری از انحراف معیار جامعه ای که نمونه از آن گرفته شده است را ارائه می دهد.

زمانی که تعداد نمونه از ۳۰ بزرگتر باشد ($N>30$)، تفاوتی در دو نوع تعریف وجود ندارد و در مخرج فرمول انحراف معیار همان N قرار داده می شود.

۴- واریانس

واریانس دسته ای از داده ها از مجذور انحراف معیار به دست آمده و با s^2 نشان داده می شود:

$$s^2=\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}{N}=\frac{\sum\left(x-\bar{x}\right)^2}{N}=\overline{\left(x-\bar{x}\right)^2}$$

در جدول توزیع فراوانی، واریانس به صورت زیر تعریف می شود:

$$s^2=\frac{\sum_{i=1}^{k} f_i\left(x_i-\bar{x}\right)^2}{N}=\frac{\sum f\left(x-\bar{x}\right)^2}{N}=\overline{\left(x-\bar{x}\right)^2}$$

$$N=\sum_{i=1}^{k} f_i=\sum f$$

که f فراوانی هر دسته در جدول توزیع فراوانی است.

نکته: فرمول های ارائه شده برای انحراف معیار و واریانس در بالا، فرمول های تعریفی هستند. برای آسان تر کردن محاسبات، می توان از فرمولهای کاربردی زیر نیز استفاده کرد:

$$s=\sqrt{\frac{\sum_{i=1}^{N}x_i^2-\frac{(\sum_{i=1}^{N}x_i)^2}{N}}{N}}=\sqrt{\frac{\sum_{i=1}^{N}x_i^{\,2}}{N}-\left(\frac{\sum_{i=1}^{N}x_i}{N}\right)^2}=\sqrt{\frac{\sum x^2}{N}-\left(\frac{\sum x}{N}\right)^2}$$

در جدول توزیع فراوانی انحراف معیار با فرمول کاربردی به صورت زیر می باشد:

$$s=\sqrt{\frac{\sum_{i=1}^{k}f_ix_i^2-\frac{(\sum_{i=1}^{k}f_ix_i)^2}{N}}{N}}=\sqrt{\frac{\sum_{i=1}^{k}f_ix_i^{\,2}}{N}-\left(\frac{\sum_{i=1}^{k}f_ix_i}{N}\right)^2}=\sqrt{\frac{\sum fx^2}{N}-\left(\frac{\sum fx}{N}\right)^2}$$

$$N=\sum_{i=1}^{k}f_i=\sum f$$

همانند فرمولهای تعریفی، در اینجا هم فرمولهای کاربردی واریانس از مجذور فرمولهای کاربردی انحراف معیار به دست می آیند.

توضیح: برای بهتر به خاطر سپردن روابط فوق، باید توجه داشت که صورت فرمول تعریفی واریانس یعنی $\sum\left(x-\bar{x}\right)^2$ را مجموع مربعات یا به اختصار SS (Sum of Squares) می نامند که برابر با $\sum x^2-\frac{\left(\sum x\right)^2}{N}$ می باشد و برای ساده تر شدن محاسبات به کار برده می شود (فرمول کاربردی مجموع مربعات).

روش کوتاه و کدگذاری برای محاسبه انحراف معیار (و واریانس)

همان گونه که در مبحث محاسبه میانگین به روشهای کوتاه و کدگذاری اشاره شد، برای محاسبه واریانس و انحراف معیار نیز می توان از روشهای کوتاه و کدگذاری به شکل زیر استفاده کرد.

چنانچه $d_i=x_i-A$ ، انحراف xi (متوسط دسته) از مقدار ثابت A باشد، برای محاسبه انحراف معیار و واریانس در روش کوتاه از روابط زیر استفاده می شود:

$$s=\sqrt{\frac{\sum_{i=1}^{N}d_i^{\,2}}{N}-\left(\frac{\sum_{i=1}^{N}d_i}{N}\right)^2}=\sqrt{\frac{\sum d^2}{N}-\left(\frac{\sum d}{N}\right)^2}$$

$$s^2=\frac{\sum_{i=1}^{N}d_i^{\,2}}{N}-\left(\frac{\sum_{i=1}^{N}d_i}{N}\right)^2=\frac{\sum d^2}{N}-\left(\frac{\sum d}{N}\right)^2$$

در جدول توزیع فراوانی، روابط انحراف معیار و واریانس در روش کوتاه به صورت زیر می باشند:

$$s=\sqrt{\frac{\sum_{i=1}^{k} f_i d_i^{\,2}}{N}-\left(\frac{\sum_{i=1}^{k} f_i d_i}{N}\right)^2}=\sqrt{\frac{\sum fd^2}{N}-\left(\frac{\sum fd}{N}\right)^2}$$

$$s^2=\frac{\sum_{i=1}^{k} f_i d_i^{\,2}}{N}-\left(\frac{\sum_{i=1}^{k} f_i d_i}{N}\right)^2=\frac{\sum fd^2}{N}-\left(\frac{\sum fd}{N}\right)^2$$

$$N=\sum_{i=1}^{k} f_i=\sum f$$

در صورتی که داده ها در جدول توزیع فراوانی دسته بندی شده و فاصله دسته آنها برابر و مساوی c باشد، $d_i = cu_i$ بوده و در نتیجه برای محاسبه انحراف معیار و واریانس به روش کدگذاری از روابط زیر استفاده می کنیم:

$$=c\sqrt{\frac{\sum_{i=1}^{k} f_i u_i^{\,2}}{N}-\left(\frac{\sum_{i=1}^{k} f_i u_i}{N}\right)^2}=c\sqrt{\frac{\sum fu^2}{N}-\left(\frac{\sum fu}{N}\right)^2}$$

$$s^2=c^2\left[\frac{\sum_{i=1}^{k} f_i u_i^{\,2}}{N}-\left(\frac{\sum_{i=1}^{k} f_i u_i}{N}\right)^2\right]=c^2\left[\frac{\sum fu^2}{N}-\left(\frac{\sum fu}{N}\right)^2\right]$$

مثال: انحراف معیار را در جدول توزیع فراوانی وزن ۱۰۰ دانشجو به روش های معمولی، کوتاه و کدگذاری محاسبه کنید.

مطابق جدول زیر ابتدا به روش معمولی، انحراف معیار را محاسبه می کنیم:

وزن (کیلوگرم)	متوسط دسته (x)	فراوانی (f)	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
۶۰–۶۲	۶۱	۵	−۶/۴۵	۴۱/۶۰۲۵	۲۰۸/۰۱۲۵
۶۳–۶۵	۶۴	۱۸	−۳/۴۵	۱۱/۹۰۲۵	۲۱۴/۲۴۵۰
۶۶–۶۸	۶۷	۴۲	−۰/۴۵	۰/۲۰۲۵	۸/۵۰۵۰
۶۹–۷۱	۷۰	۲۷	۲/۵۵	۶/۵۰۲۵	۱۷۵/۵۶۷۵
۷۲–۷۴	۷۳	۸	۵/۵۵	۳۰/۸۰۲۵	۲۴۶/۴۲۰۰
کل		$N=\sum f=100$			$\sum f(x-\bar{x})^2=825.7500$

یادآوری: در مثال مربوط به محاسبه میانگین به روشهای معمولی، کوتاه و کدگذاری، میانگین وزن ۱۰۰ دانشجو ۶۷/۴۵ کیلوگرم محاسبه شد.

$$s=\sqrt{\frac{\sum f\left(x-\bar{x}\right)^2}{N}}=\sqrt{\frac{825.75}{100}}=\sqrt{8.5275}=2.92Kg$$

در محاسبه انحراف معیار به روشها کوتاه و کدگذاری به صورت زیر عمل می کنیم. در اینجا هم متوسط یک دسته مانند ۶۷ را A فرض می کنیم:

وزن (کیلوگرم)	متوسط دسته (x)	فراوانی (f)	d=x-A	fd	fd^2	$u=\frac{d}{c}$	fu	fu^2
۶۰-۶۲	۶۱	۵	-۶	-۳۰	۱۸۰	-۲	-۱۰	۲۰
۶۳-۶۵	۶۴	۱۸	-۳	-۵۴	۱۶۲	-۱	-۱۸	۱۸
۶۶-۶۸	۶۷	۴۲	۰	۰	۰	۰	۰	۰
۶۹-۷۱	۷۰	۲۷	۳	۸۱	۲۴۳	۱	۲۷	۲۷
۷۲-۷۴	۷۳	۸	۶	۴۸	۲۸۸	۲	۱۶	۳۲
کل		$N=\sum f=100$		$\sum fd=45$	$\sum fd^2=873$		$\sum fu=15$	$\sum fu^2=97$

در روش کوتاه، انحراف معیار با توجه به مقادیر ستون های پنجم و ششم جدول فوق به صورت زیر محاسبه می شود:

$$s=\sqrt{\frac{\sum fd^2}{N}-\left(\frac{\sum fd}{N}\right)^2}=\sqrt{\frac{873}{100}-\left(\frac{45}{100}\right)^2}=\sqrt{8.5275}=2.92Kg$$

با توجه به مساوی بودن فاصله دسته ها (c=3) می توان انحراف معیار را به روش کدگذاری و با توجه به مقادیر ستون های هشتم ونهم جدول فوق، به صورت زیر محاسبه کرد:

$$s=c\sqrt{\frac{\sum fu^2}{N}-\left(\frac{\sum fu}{N}\right)^2}=3\times\sqrt{\frac{97}{100}-\left(\frac{15}{100}\right)^2}=3\times\sqrt{0.9475}=2.92Kg$$

۵- دامنه نیمه چارک داخلی (انحرافات چارکی)

دامنه نیمه چارک داخلی یا انحرافات چارکی دسته ای از داده ها را که با Q نشان می دهند، به صورت زیر تعریف می شود:

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

که در رابطه فوق Q_1 و Q_3 به ترتیب چارک های اول و سوم هستند.

نکته: گاهی دامنه چارک داخلی Q_3-Q_1 مورد استفاده قرار می گیرد، ولی انحرافات چارکی به عنوان پارامتر پراکندگی بیشتر متداول است.

مثال: دامنه انحرافات چارکی وزن ۱۰۰ دانشجو را به دست آورید.

وزن (کیلوگرم)		فراوانی (f)	فراوانی تجمعی
۶۰-۶۲	۵۹/۵-۶۲/۵	۵	۵
۶۳-۶۵	۶۲/۵-۶۵/۵	۱۸	۲۳
۶۶-۶۸	۶۵/۵-۶۸/۵	۴۲	۶۵
۶۹-۷۱	۶۸/۵-۷۱/۵	۲۷	۹۲
۷۲-۷۴	۷۱/۵-۷۴/۵	۸	۱۰۰
کل		۱۰۰	

همانند آنچه در مثال مربوط به چندک ها ذکر شد، ابتدا فراوانی تجمعی را مشخص کرده، سپس محل قرار گرفتن چارک های اول و سوم در جدول را تعیین و در نهایت مقادیر آنها را محاسبه می کنیم.

$$\begin{cases} \frac{1N}{4} = \frac{1\times 100}{4} = 25 \\ Q_1 = 65.5 + 3\times\left(\frac{25-23}{42}\right) = 65.64KG \end{cases}$$

$$\begin{cases} \frac{3N}{4} = \frac{3\times 100}{4} = 75 \\ Q_3 = 68.5 + 3\times\left(\frac{75-65}{27}\right) = 69.61Kg \end{cases}$$

$$Q = \frac{Q_3 - Q_1}{2} \rightarrow Q = \frac{69.61-65.64}{2} = 1.98Kg$$

$$\frac{1}{2}\left(Q_1 + Q_3\right) = 67.63Kg$$

از آنجایی که فاصله بین چارک سوم و اول، ۵۰ درصد داده ها را شامل می شود، بنابراین می توان گفت ۵۰ درصد دانشجویان وزنی در حدود $67.63 \pm 1.98Kg$ دارند.

دامنه صدک ۱۰ تا ۹۰

دامنه صدک ۱۰ تا ۹۰ دسته ای از داده ها به صورت رابطه زیر تعریف می شود:

$P = P_{90} - P_{10}$

در رابطه فوق P_{10} و P_{90} به ترتیب صدک های دهم و نودم داده ها هستند.

نکته: دامنه نیمه صدک ۱۰ تا ۹۰ یعنی $\frac{P_{90}-P_{10}}{2}$ نیز مورد استفاده قرار می گیرد ولی متداول نیست.

مثال: دامنه صدک های ۱۰ تا ۹۰ وزن ۱۰۰ دانشجو را به دست آورید.

وزن (کیلوگرم)		فراوانی (f)	فراوانی تجمعی
۶۰-۶۲	۵۹/۵-۶۲/۵	۵	۵
۶۳-۶۵	۶۲/۵-۶۵/۵	۱۸	۲۳
۶۶-۶۸	۶۵/۵-۶۸/۵	۴۲	۶۵
۶۹-۷۱	۶۸/۵-۷۱/۵	۲۷	۹۲
۷۲-۷۴	۷۱/۵-۷۴/۵	۸	۱۰۰
کل		۱۰۰	

در اینجا هم مانند آنچه در مثال قبل و فصل قبل توضیح داده شد، عمل می کنیم:

$$\begin{cases} \frac{10N}{100}=10 \\ P_{10}=62.5+3\times\left(\frac{10-5}{18}\right)=63.33Kg \end{cases}$$

$$\begin{cases} \frac{90N}{100}=90 \\ P_{90}=68.5+3\times\left(\frac{90-65}{27}\right)=71.27Kg \end{cases}$$

$$\frac{P_{90}-P_{10}}{2}=\frac{71.27-63.33}{2}=3.97Kg$$

$$\frac{1}{2}\left(P_{10}+P_{90}\right)=67.30Kg$$

از آنجایی که فاصله بین صدک نودم و دهم، ۸۰ درصد داده ها را شامل می شود، بنابراین می توان گفت ۸۰ درصد دانشجویان وزنی در دامنه $67.30\pm3.97Kg$ دارند.

۶- ضریب تغییرات

ضریب تغییرات را که با .C.V نشان داده می شود به صورت زیر محاسبه می شود:

$$C.V.=\frac{s}{\bar{x}}$$

و معمولاً به صورت درصد بیان می شود.

ضریب تغییرات میزان تغییرات در یک آزمایش را نشان داده و برای سنجش میزان دقت آزمایش، از آن استفاده می شود.

مثال: ضریب تغییرات در مثال وزن ۱۰۰ دانشجو را به دست آورید.

با توجه به اینکه میانگین و انحراف معیار وزن ۱۰۰ دانشجو به ترتیب ۶۷/۴۵ کیلو گرم و ۲/۹۲ کیلوگرم به دست آمده، ضریب تغییرات برابر است با:

$$C.V.=\frac{s}{\bar{x}} \rightarrow C.V.=\frac{2.92}{67.45}=0.043 \rightarrow C.V.=0.043\times 100=4.3\%$$

بخش پنجم

(گشتاورها)

گشتاورها

اگر متغیر x1 به صورت $x_1, x_2, ..., x_N$ تعریف شده باشد، rاُمین گشتاور را با $\overline{x^r}$ نشان داده که به صورت زیر تعریف می‌شود:

$$\overline{x^r} = \frac{x_1^r + x_2^r + ... + x_3^r}{N} = \frac{\sum_{i=1}^{N} x_i^r}{N} = \frac{\sum x^r}{N}$$

اولین گشتاور (r = 1) برابر با میانگین حسابی است.

rامین گشتاور نسبت به میانگین به صورت زیر تعریف می شود که آن را با mr نشان داده می شود:

$$m_r = \frac{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^r}{N} = \overline{\left(x - \bar{x}\right)^r}$$

چنانچه r=1 باشد، $m_1=0$ خواهد بود و اگر r=2 باشد، $m_2=s^2$ (واریانس) است.

rامین گشتاور نسبت به هر مبدا مانند A را با m_r' نشان داده و به صورت زیر تعریف می شود:

$$m'_r = \frac{\sum_{i=1}^{N} (x_i - A)^r}{N} = \frac{\sum d^r}{N} = \overline{(x - A)^r}$$

بنابراین، فرمول اول حالت خاصی از فرمول فوق است زمانی که A=0 باشد.

در صورتی که بخواهیم برای داده های گروه بندی شده در جدول توزیع فراوانی گشتاورها را محاسبه کنیم، روابط فوق به صورت زیر درخواهند آمد:

$$\overline{x^r} = \frac{\sum fx^r}{N}$$

$$m_r = \frac{\sum f\left(x - \bar{x}\right)^r}{N}$$

$$m'_r = \frac{\sum f\left(x - A\right)^r}{N}$$

به طوری که $N = \sum f$ خواهد بود.

نکته:

$$m'_r = \frac{\sum (x - A)^r}{N}$$

$$m'_1 = \frac{\sum (x-A)^1}{N} = \frac{\sum x - \sum A}{N} = \frac{\sum x}{N} - \frac{NA}{N} = \bar{x} - A \Rightarrow m'_1 = \bar{x} - A$$

روابط بین گشتاورها

بین گشتاورهای نسبت به میانگین(m_r) و گشتاورهای نسبت به مبدا خاص، روابط زیر برقرار است:

$$m_2 = m'_2 - m'^2_1$$
$$m_3 = m'_3 - 3m'_1 m'_2 + 2m'^3_1$$
$$m_4 = m'_4 - 4m'_1 m'_3 + 6m'^2_1 m'_2 - 3m'^4_1$$

ضمن اینکه می دانیم $m_1 = 0$ می باشد.

مثال: گشتاورهای اول و دوم داده های ۲، ۳، ۷، ۸ و ۱۰ را محاسبه کنید. سپس گشتاورهای اول و دوم این داده ها را نسبت به میانگین و نسبت به مبدا ۴ محاسبه کنید.

$$\bar{x} = \frac{\sum x}{N} = \frac{2+3+7+8+10}{5} = 6$$
$$\overline{x^2} = \frac{\sum x^2}{N} = \frac{2^2+3^2+7^2+8^2+10^2}{5} = \frac{226}{5} = 45.2$$
$$m_1 = \frac{\sum (x-\bar{x})}{N} = \frac{(2-6)+(3-6)+\ldots+(10-6)}{5} = 0$$
$$m_2 = \frac{\sum (x-\bar{x})^2}{N} = \frac{(2-6)^2+(3-6)^2+\ldots+(10-6)^2}{5} = \frac{46}{5} = 9.2 = s^2$$
$$A = 4 \rightarrow \begin{cases} m'_1 = \dfrac{\sum (x-A)}{N} = \dfrac{(2-4)+(3-4)+\ldots+(10-4)}{5} = \dfrac{10}{5} = 2 \\ m'_2 = \dfrac{\sum (x-A)^2}{N} = \dfrac{(2-4)^2+(3-4)^2+\ldots+(10-4)^2}{5} = \dfrac{66}{5} = 13.2 \end{cases}$$

محاسبه گشتاورها در جدول توزیع فراوانی

همانگونه که در بحث میانگین به روش کدگذاری ذکر شد $\bar{x} = A + c\dfrac{\sum fu}{N}$، برای محاسبه گشتاورها نیز از همین رابطه به صورت زیر استفاده می شود:

$$m'_r = c^r \frac{\sum fu^r}{N} = c^r \overline{u^r}$$

مثال: گشتاورهای مرتبه اول تا چهارم را نسبت به میانگین در جدول توزیع فراوانی وزن ۱۰۰ دانشجو محاسبه کنید.

fu^4	fu^3	fu^2	fu	$u=\frac{d}{c}$	d=x-A	f	x
۸۰	−۴۰	۲۰	−۱۰	−۲	−۶	۵	۶۱
۱۸	−۱۸	۱۸	−۱۸	−۱	−۳	۱۸	۶۴
۰	۰	۰	۰	۰	۰	۴۲	۶۷
۲۷	۲۷	۲۷	۲۷	۱	۳	۲۷	۷۰
۱۲۸	۶۴	۳۲	۱۶	۲	۶	۸	۷۳
$\sum fu^4=253$	$\sum fu^3=33$	$\sum fu^2=97$	$\sum fu=15$			$N=\sum f=100$	

$$m_1'=c\frac{\sum fu}{N}=3\times\frac{15}{100}=0.45$$

$$m_2'=c^2\frac{\sum fu^2}{N}=3^2\times\frac{97}{100}=8.73$$

$$m_3'=c^3\frac{\sum fu^3}{N}=3^3\times\frac{33}{100}=8.91$$

$$m_4'=c^4\frac{\sum fu^4}{N}=3^4\times\frac{253}{100}=204.93$$

بنابراین با توجه به روابط بین گشتاورهای نسبت به یک مبدأ خاص و گشتاورهای نسبت به میانگین خواهیم داشت:

$$m_1=0$$

$$m_2=m_2'-m_1'^2=8.73-(0.45)^2=8.5275$$

$$m_3=m_3'-3m_1'm_2'+2m_1'^3=8.91-3(0.45)(8.73)+2(0.45)^3=-2.6932$$

$$m_4=m_4'-4m_1'm_3'+6m_1'^2m_2'-3m_1'^4=204.93-4(0.45)(8.91)+6(0.45)^2(8.73)-3(0.45)^4=199.3759$$

گشتاورها در فرم داده های بدون بعد

می توان گشتاورها را بدون بعد حول میانگین، به صورت زیر تعریف کرد:

$$a_r=\frac{m_r}{s^r}=\frac{m_r}{\left(\sqrt{m_2}\right)^r}=\frac{m_r}{\sqrt{m_2^r}}$$

با توجه به اینکه می دانیم $m_2=s^2$ می باشد و در نتیجه:

$$m_2=s^2\rightarrow s=\sqrt{m_2}\rightarrow s^r=\left(\sqrt{m_2}\right)^r$$

چولگی

عدم تقارن یا انحراف از تقارن در یک توزیع را چولگی می نامند.

در توزیع های چوله (همانگونه که در فصل سوم ذکر گردید)، میانگین تمایل دارد به همان سمت دم طویل تر قرار گیرد. بنابراین چولگی از تفاوت میانگین و مد به دست می آید و با تقسیم بر انحراف معیار بدون بعد می شود. این رابطه تحت عنوان ضریب اول چولگی پیرسون نامیده می شود (از آنجایی که تفاوت میانگین و مد برابر با سه برابر تفاوت میانگین و میانه می باشد، می توان ضریب دوم چولگی پیرسون را نیز به صورت زیر تعریف کرد):

ضریب اول چولگی پیرسون $$=\frac{\bar{x}-M_o}{s}$$

ضریب دوم چولگی پیرسون $$=\frac{3\left(\bar{x}-M_d\right)}{s}$$

ضریب گشتاور چولگی که با کاربرد گشتاور مرتبه سوم حول میانگین در فرم بدون بعد بیان می شود، به صورت زیر است:

ضریب گشتاور چولگی $$=a_3=\frac{m_3}{s^3}=\frac{m_3}{\left(\sqrt{m_2}\right)^3}=\frac{m_3}{\sqrt{m_2^3}}$$

این اندازه گاهی با b_1 نمایش داده می شود که برای منحنی های متقارن نظیر منحنی نرمال صفر می باشد.

کشیدگی

درجه نوک تیزی یک توزیع را در مقایسه با توزیع نرمال نشان می دهد. بنابراین منحنی می تواند کشیده باریک، پهن یا کشیده میانی (نرمال) باشد.

ضریب گشتاور کشیدگی که با کاربرد گشتاور مرتبه چهارم نسبت به میانگین در فرم بدون بعد بیان می شود، به صورت زیر است:

ضریب گشتاور کشیدگی $$=a_4=\frac{m_4}{s^4}=\frac{m_4}{\left(\sqrt{m_2}\right)^4}=\frac{m_4}{m_2^2}$$

که این اندازه با b_2 نیز نشان داده می شود. برای توزیع نرمال، $a_4=b_2=3$ است. b_2 برای توزیع کشیده باریک بیشتر از ۳ و برای توزیع پهن، کوچکتر از ۳ است.

مثال: در مثال حل شده فوق، ضریب گشتاور چولگی و ضریب گشتاور کشیدگی را محاسبه کنید.

$$a_3=\frac{m_3}{\sqrt{m_2^3}}=\frac{-2.6932}{\sqrt{(8.5275)^3}}=-0.1082=-0.11$$

$$a_4=\frac{m_4}{m_2^2}=\frac{199.3759}{(8.5275)^2}=2.7418=2.74$$

بخش ششم

(احتمالات)

در فصول گذشته، در مورد آمار توصیفی توضیح داده شد. همچنان که قبلا ذکر شد، در آمار توصیفی انبوهی از داده ها به یکی از روش هایی که قبلا توضیح داده شده است، خلاصه می شوند.
ما نیاز داریم که در مورد داده ها استنباطی داشته باشیم. پایه آمار استنباطی، احتمالات است.

تعریف احتمال

اگر حادثه ای به s صورت (s مساوی است با تعداد موفقیت) رخ دهد و به f صورت (f برابر است با تعداد عدم موفقیت) رخ ندهد، احتمال رخ دادن حادثه (احتمال موفقیت) برابر است با:

$$p=\frac{s}{s+f}$$
$$q=\frac{f}{s+f}$$
$$p+q=1$$

p احتمال موفقیت و q احتمال عدم موفقیت می باشد.

مثال: در پرتاب یک تاس، احتمال آمدن عدد ۶ چقدر است؟

آمدن عدد ۶ = موفقیت = ۱ $\Longrightarrow$ p = احتمال آمدن عدد ۶ $=\frac{1}{6}$

نیامدن عدد ۶ = عدم موفقیت = ۵ $\Longrightarrow$ q = احتمال نیامدن عدد ۶ $=\frac{5}{6}$

مثال: در کیسه ای ۵ مهره آبی و ۶ مهره قرمز و۴ مهره سفید وجود دارد. یک مهره به تصادف از کیسه خارج می شود. احتمال آبی بودن مهره چقدر است؟

آبی بودن مهره = موفقیت = ۵ $\Longrightarrow$ p = احتمال آبی بودن $=\frac{5}{10+5}=\frac{5}{15}$

آبی نبودن مهره = عدم موفقیت = ۱۰ $\Longrightarrow$ q = احتمال آبی نبودن $=1-\frac{5}{15}=\frac{10}{15}$

مثال: در یک مرتبه ریختن دو تاس احتمال به دست آوردن مجموع عدد ۹ چقدر است؟

(3,6) , (6,3) , (4,5) , (5,4)

احتمال آنکه مجموع ۹ باشد $=p=\frac{4}{36}$ احتمال آنکه مجموع ۹ نباشد $=q=1-\frac{4}{36}=\frac{32}{36}$

قوانین مقدماتی احتمالات

سه قانون مقدماتی احتمالات عبارتند از:

۱- قانون مانعه الجمع یا پیشامدهای ناسازگار

۲- قانون ضرب یا پیشامدهای مستقل

۳- قانون شرطی

پیشامدهای ناسازگار

پیشامدهایی را ناسازگار می گویند که وقوع یکی مانع از وقوع دیگری شود. یعنی در حقیقت، این پیشامدها مانعه الجمع هستند. در این حالت با "یا" سرو کار داریم و باید از جمع احتمالات هر پیشامد استفاده کنیم.

(به عنوان مثال یک سکه را در نظر بگیرید)

$$P(A\bigcup B)=P(A)+P(B)-P(A\bigcap B)$$

مثال: در پرتاب یک تاس، احتمال آمدن عدد ۳ یا ۵ چقدر است؟

احتمال آمدن ۳ + احتمال آمدن ۵ $=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$

مثال: از کیسه ای که شامل ۵ مهره آبی، ۶ مهره قرمز و ۴ مهره سفید است، یک مهره به تصادف خارج می شود. احتمال آبی بودن یا قرمز بودن مهره چقدر است؟

$$P=\frac{5}{15}+\frac{6}{15}$$

مثال: در پرتاب دو تاس به طور همزمان، احتمال اینکه مجموع دو تاس، عدد ۸ یا زوج بودن هر دو تاس چقدر است؟

مجموع عدد ۸:	(2,6) , (6,2) , (3,5) , (5,3) , (4,4)
زوج بودن:	(2,2) , (4,4) , (6,6) , (2,4) , (4,2) , (2,6) , (6,2) , (4,6) , (6,4)

$$P(A\bigcup B)=P(A)+P(B)-P(A\bigcap B)=\frac{5}{36}+\frac{9}{36}-\frac{3}{36}=\frac{11}{36}$$

قانون ضرب یا پیشامدهای مستقل

دو یا چند پیشامد را زمانی مستقل می نامند که احتمال وقوع هر یک از آنها تحت تاثیر وقوع دیگری قرار نگیرد. اگر $p_1, p_2,..., p_r$ احتمال وقوع r حادثه مستقل باشد، در این صورت احتمال وقوع کلیه این حوادث برابر است با:

$$p_1\times p_2\times...\times p_r$$

برای پیشامدهای مستقل از " و" و علامت ضرب استفاده می شود (دو سکه را در نظر بگیرید).

مثال: در پرتاب دو تاس احتمال اینکه اولی عدد ۶ و دومی عدد ۵ باشد، چقدر است؟

$$P=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36}$$

مثال: در داخل یک کیسه، ۵ مهره قرمز، ۶ مهره سبز و ۴ مهره سفید وجود دارد.

الف- دو مهره با جایگزینی خارج می شود. احتمال اینکه مهره اول، قرمز و مهره دوم سبز باشد، چقدر است؟

$$P=\frac{5}{15}\times\frac{6}{15}$$

ب- دو مهره با جایگزینی خارج می شود. احتمال اینکه هردو قرمز باشند، چقدر است؟

$$P=\frac{5}{15}\times\frac{5}{15}$$

ج- دو مهره بدون جایگزینی خارج می شود. احتمال اینکه اولی قرمز و دومی سبز باشد، چقدر است؟

$$P=\frac{5}{15}\times\frac{6}{14}$$

د- دو مهره بدون جایگزینی خارج می شود. احتمال اینکه هردو قرمز باشند، چقدر است؟

$$P=\frac{5}{15}\times\frac{4}{14}$$

مثال: در کشویی، دو جفت جوراب (۴ لنگه) به رنگ های قرمز و آبی وجود دارد. دو لنگه جوراب بدون جایگزینی (بترتیب یا پشت سرهم، یعنی بدون جایگزینی) خارج می شود. احتمال اینکه دو لنگه جوراب همرنگ باشد، چقدر است؟

اولی قرمز و دومی قرمز یا اولی آبی و دومی آبی

$$\frac{2}{4}\times\frac{1}{3}+\frac{2}{4}\times\frac{1}{3}=\frac{1}{3}$$

مثال: در جعبه ای، ۵ مهره فلزی، ۵ مهره پلاستیکی و ۵ مهره چوبی قرار دارد. سه مهره از هرکدام از شماره ۱ تا ۳ شماره گذاری شده و ۲ مهره بدون شماره اند. یک مهره به تصادف از این جعبه خارج می شود. احتمال اینکه عدد ۳ یا بدون شماره باشد، چقدر است؟

$$P=\frac{3}{15}+\frac{6}{15}=\frac{9}{15}$$

احتمال شرطی

اگر E_1 و E_2 دو واقعه باشند، احتمال اینکه E_2 اتفاق بیفتد به شرط اینکه E_1 اتفاق افتاده باشد را با $P\{E_2|E_1\}$ نشان داده و آن را احتمال شرطی می نامند.

$$P\{E_1E_2\}=P\{E_1\}P\{E_2|E_1\}$$

مثال: اگر جعبه ای شامل ۴ توپ قرمز و ۵ توپ آبی باشد، اگر دو توپ بدون جایگزینی از جعبه خارج شوند، اگر توپ اول آبی باشد، احتمال اینکه توپ دوم آبی باشد، چقدر است؟

احتمال اینکه توپ اول آبی باشد $P(E_1)=\frac{5}{9}$

احتمال اینکه توپ دوم آبی باشد به شرطی که توپ اول آبی باشد (احتمال شرطی) $P\{E_2|E_1\}=\frac{4}{8}=\frac{1}{2}$

امید ریاضی

اگر p احتمال دریافت جمع مبلغ s توسط فردی باشد، امید ریاضی به عنوان ps تعریف می شود. مثلا اگر احتمال این که فردی ۱۰۰ تومان جایزه ببرد $\frac{1}{5}$ باشد، امید برنده شدن او $\frac{1}{5}\times 100=20$ تومان خواهد بود.

بسط(گسترش) مفهوم امید ریاضی

اگر x_i متغیر تصادفی منفصل $x_1, x_2, ..., x_k$ با احتمال های متناظر $p_1, p_2, ..., p_k$ باشد، با فرض اینکه $p_1+p_2+...+p_k=1$ باشد، امید ریاضی برابر است با:

$$E(x)=p_1x_1+p_2x_2+...+p_kx_k=\sum_{i=1}^{k}p_ix_i=\sum px$$

$$p=\frac{f_i}{N}$$

با فرض اینکه $\sum f_i=N$ باشد، خواهیم داشت:

$$E(\mathrm{x})=\sum_{i=1}^{k}p_ix_i=\frac{\sum f_ix_i}{N}=\overline{X}$$

که نشان می دهد E(x) میانگین جامعه ای است که نمونه از آن انتخاب شده است.

n فاکتوریل

n فاکتوریل که با !n نشان داده می شود، به صورت زیر تعریف می شود:

n!=n(n-1)(n-2)…1

به عنوان مثال !4 به صورت زیر محاسبه می شود:

$$4!=4\times3\times2\times1=24$$

نکته ۱: 1=!0 می باشد.

نکته ۲: وقتی n بزرگ باشد، محاسبه !n عملی نیست. در این موارد از فرمول تقریبی زیر که توسط استرلینگ توسعه یافته استفاده می شود:

$n! = \sqrt{2\pi n} n^n e^{-n}$

ترتیب، ترکیب و تبدیل

از این سه برای تعیین تعداد حالات مورد نظر یا تعداد کل حالات استفاده می شود (تعداد حالات نه احتمال).

ترتیب

حالت های گروه بندی n پدیده یا واقعه را n به n، ترتیب گویند. به عنوان مثال، ترتیب قرار گرفتن ۵ کتاب در یک قفسه.

مثال: به چند طریق ۴ نفر می توانند روی ۴ صندلی بنشینند؟

$4\times3\times2\times1 = 4! = 24$

ترکیب(combination)

حالت های گروه بندیr رویداد از n واقعه(n>r) مدنظر می باشد و تقدم و تأخر و یا پس و پیش شدن وقایع(ترتیب قرار گرفتن) مورد توجه نمی باشد.

$$\begin{bmatrix} n \\ r \end{bmatrix} = C_r^n = \frac{n!}{(n-r)!r!}$$

مثال: از بین ۴ حرف A، B، C و D چند کلمه سه حرفی می توان ایجاد کرد به طوری که ترتیب قرار گرفتن حروف، حالات متفاوتی را ایجاد نکند.

$$C_r^n = C_3^4 = \frac{n!}{(n-r)!r!} = \frac{4!}{(4-3)!3!} = \frac{4\times3!}{1!\times3!} = 4$$

ABC, ABD, BCD, ACD

تبدیل یا جایگشت (permutation)

چنانچه حالتهای گروه بندی r رویداد از n واقعه (n>r) مدنظر باشد و تقدم وتأخر وقایع نیز حالتهای متفاوتی ایجاد کند، از تبدیل یا جایگشت استفاده می شود.

$${}_nP_r = P_{(n,r)} = P_{n,r} = P_r^n = \frac{n!}{(n-r)!}$$

مثال: از بین ۴ حرف A، B، C و D به چند طریق می توان ۳ حرف را کنار هم قرار داد، به شرط این که ترتیب قرار گرفتن حروف، حالات متفاوتی ایجاد کند؟

$$P_{r=3}^{n=4} = \frac{n!}{(n-r)!} = \frac{4!}{(4-3)!} = 24$$

ABC $\Longrightarrow$ ABC, ACB, BAC, BCA, CAB, CBA

همانگونه که ملاحظه می شود، ترکیب ABC (در مثال قبل)، در تبدیل یا جایگشت ۶ حالت مختلف ایجاد می کند. به همین ترتیب سه ترکیب دیگر در مثال قبلی نیز هر کدام ۶ حالت مختلف ایجاد کرده و در نهایت در تبدیل، ۲۴ حالت خواهیم داشت.

مثال: به چند طریق می توان از بین ۵ مرد، ۳ مرد را انتخاب کرد؟

$$C_{r=3}^{n=5}=\frac{n!}{(n-r)!r!}=\frac{5!}{(5-3)!3!}=\frac{5!}{2!\times 3!}=10$$

مثال: به چند طریق می توان از بین ۵ مرد و ۶ زن، ۲ مرد و ۲ زن را انتخاب کرد؟

$$C_{r=2}^{n=5}=\frac{n!}{(n-r)!r!}=\frac{5!}{3!2!}=10$$

$$C_{r=2}^{n=6}=\frac{n!}{(n-r)!r!}=\frac{6!}{4!2!}=15$$

$$P=10\times 15=150$$

مثال: در جعبه ای ۱۰ کارت شماره دار از شماره ۱ تا ۱۰ وجود دارد. ۳ کارت بدون جایگزینی از جعبه خارج می شود.

الف– احتمال این که عدد ۳، ۴ و ۵ بیرون بیایند؟

$$P=\frac{3}{10}\times\frac{2}{9}\times\frac{1}{8}$$

or

$$P=\frac{1}{10}\times\frac{1}{9}\times\frac{1}{8}\times 3!$$

ب– احتمال این که عدد ۳، ۴ و ۵ به ترتیب بیرون بیایند؟

$$P=\frac{1}{10}\times\frac{1}{9}\times\frac{1}{8}$$

بخش هفتم

(توزیع‌های احتمال گسسته و پیوسته)

توزیع دوجمله ای

اگر سه شرط زیر به طور همزمان وجود داشته باشد، برای محاسبه احتمالات از توزیع دوجمله ای استفاده می شود:

۱- آزمایش دو حالتی باشد (p موفقیت و q عدم موفقیت).

۲- آزمایش n مرتبه تکرار شود.

۳- در هر دفعه تکرار آزمایش، احتمال p و q ثابت باشد.

به عنوان مثال، ۱۰ بار پرتاب یک سکه سالم از توزیع دوجمله ای تبعیت می کند، زیرا اولاً آزمایش دو حالتی است (شیر یا خط)، آزمایش ۱۰ بار تکرار شده و در هر دفعه تکرار پرتاب سکه، احتمال p و q ثابت مانده و برابر است.

در این حالت برای محاسبه احتمالات از بسط دوجمله ای $(p+q)^n$ به صورت زیر استفاده می شود:

$$p_{(x)}=\begin{bmatrix} n \\ x \end{bmatrix} p^x q^{n-x}=\frac{n!}{(n-x)!x!}p^x q^{n-x}$$

که در رابطه فوق، p احتمال موفقیت، q احتمال عدم موفقیت، x تعداد دفعات موفقیت (تعداد دفعات مورد نظر) و $\frac{n!}{(n-x)!x!}$ ضریب جمله مورد نظر در دوجمله ای بسط یافته است.

به عنوان مثال بسط دو جمله ای $(p+q)^4$ به صورت زیر می باشد:

$$(p+q)^4=p^4+4p^3q+6p^2q^2+4pq^3+q^4$$

نکته ۱: همچنین برای تعیین ضرایب دو جمله ای می توان از مثلث خیام (یا مثلث خیام-پاسکال یا مثلث خیام-پاسکال-نیوتن) استفاده کرد:

$$\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1
\end{array}$$

در مثلث خیام در هر سطر، ضرایب با ۱ آغاز و با ۱ پایان می یابند و سایر ضرایب از جمع دو عدد بالایی خود به دست می آیند. عدد ۱ در رأس مثلث، ضریب دو جمله ای $(p+q)^0$ می باشد (یعنی زمانی که n=0 است). در سطر دوم، ضرایب دوجمله ای $(p+q)^1$ دیده می شود (n=1). سطر سوم مربوط به ضرایب دو جمله ای $(p+q)^2$ می باشد(n=2) و به همین ترتیب ضرایب دوجمله ای ها با مقادیر n بزرگتر به دست می آیند. در مثلث فوق ضرایب دو جمله ای های تا n=5 نوشته شده است.

$(a+b)^0$	1
$(a+b)^1$	$1a+1b$
$(a+b)^2$	$1a^2+2ab+1b^2$
$(a+b)^3$	$1a^3+3a^2b+3ab^2+1b^3$
$(a+b)^4$	$1a^4+4a^3b+6a^2b^2+4ab^3+1b^4$
$(a+b)^5$	$1a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+1b^5$

نکته ۲: به طور کلی برای بسط دو جمله ای بدون آگاهی از روابط فوق می توان از اصول زیر استفاده کرد:

۱- بسط هر دوجمله ای از n+1 جمله تشکیل شده است.

۲- ضریب جمله اول و جمله آخر بسط ۱ می باشد.

۳- جمله اول بسط p^n و جمله آخر q^n می باشد (همان طور که ذکر شد، ضرایب این جملات ۱ می باشد).

۴- در جملات بعدی از توان p یکی کاسته و به توان q یکی اضافه خواهد شد.

۵- مجموع توان های p و q در هر جمله بسط برابر با n می باشد (در واقع جمله اول بسط در اصل به صورت p^nq^0 و جمله آخر بسط به صورت p^0q^n می باشد).

۶- برای نوشتن ضرایب جملات دیگر (غیر از جملات اول و آخر که ضریبشان ۱ است) به این ترتیب عمل می کنیم که ضریب هر جمله (مثلاً جمله سوم در بسط دوجمله ای $(p+q)^5$) عبارت است از حاصل ضرب توان p در ضریب در جمله قبلی (در این مثال، جمله دوم) تقسیم بر تعداد جملات نوشته شده (در این مثال ۲ جمله). به عنوان مثال برای نوشتن ضریب جمله سوم در بسط دوجمله ای $(p+q)^5$ به صورت زیر عمل می شود:

$$(p+q)^5 = p^5 + \underline{5p^4q} + \ldots$$

$$\frac{4\times 5}{2} = 10$$

۷- ضریب جملات متقارن در بسط دو جمله ای با هم برابرند.

مثال: بسط دوجمله ای $(p+q)^7$ را بنویسید.

$$(p+q)^7 = p^7 + 7p^6q + 21p^5q^2 + 35p^4q^3 + 35p^3q^4 + 21p^2q^5 + 7pq^6 + q^7$$

مثال: تاسی را ۵ مرتبه پرتاب می کنیم، احتمال اینکه :

الف- درست ۳ مرتبه عدد ۴ بیاید؟

$$(p+q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$$

$$p=\frac{1}{6},q=\frac{5}{6},x=3\Rightarrow 10p^3q^2=10\times(\frac{1}{6})^3(\frac{5}{6})^2$$

ب- احتمال اینکه هر ۵ مرتبه عدد ۴ بیاید؟

$$x=5\rightarrow 1p^5=(\frac{1}{6})^5$$

ج- احتمال اینکه حداقل ۳ مرتبه عدد ۴ بیاید؟

$$x=3,4,5\rightarrow p^5+5p^4q+10p^3q^2=(\frac{1}{6})^5+5\times(\frac{1}{6})^4(\frac{5}{6})+10\times(\frac{1}{6})^3(\frac{5}{6})^2$$

د- احتمال اینکه حداکثر ۳ مرتبه عدد ۴ بیاید؟

$$p=0.1,q=0.9,\mathrm{n}=7,x=4,5,6,7$$
$$P=p^7+7p^6q+21p^5q^2+35p^4q^3=(0.1)^7+7(0.1)^6(0.9)+21(0.1)^5(0.9)^2+35(0.1)^4(0.9)^3$$
$$(p+q)^7=p^7+7p^6q+21p^5q^2+35p^4q^3+35p^3q^4+21p^2q^5+7pq^6+q^7$$
$$P=p^4=(\frac{3}{4})^4$$
$$p_{(x)}=\begin{bmatrix} n \\ x \end{bmatrix}p^xq^{n-x}=\frac{n!}{(n-x)!x!}p^xq^{n-x}\rightarrow P_{(4)}=\begin{bmatrix} 4 \\ 4 \end{bmatrix}p^4q^{4-4}=\frac{4!}{0!4!}p^4q^0=p^4=(\frac{3}{4})^4$$
$$C_2^5=\frac{5!}{3!2!}=10$$
$$p_{(x)}=\begin{bmatrix} n \\ x \end{bmatrix}p^xq^{n-x}\rightarrow P_{(2)}=\begin{bmatrix} 5 \\ 2 \end{bmatrix}p^2q^3=10(\frac{1}{2})^2(\frac{1}{2})^3=10(\frac{1}{2})^5$$
$$P=C_2^5p^2q^3+C_3^5p^3q^2+C_4^5p^4q+C_5^5p^5=10(0.4)^2(0.6)^3+10(0.4)^3(0.6)^2+5(0.4)^4(0.6)+(0.4)^5$$

ه- احتمال اینکه بیش از ۳ مرتبه عدد ۴ بیاید؟

$$x=4,5\rightarrow p^5+5p^4q=\left(\frac{1}{6}\right)^5+5\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)$$

و- احتمال اینکه کمتر از ۳ مرتبه عدد ۴ بیاید؟

$$x=0,1,2\rightarrow 10p^2q^3+5pq^4+q^5=10(\frac{1}{6})^2(\frac{5}{6})^3+5(\frac{1}{6})(\frac{5}{6})^4+(\frac{5}{6})^5$$

مثال: احتمال اینکه در یک گاوداری یک گاو مریض وجود داشته باشد، برابر ۰/۱ می باشد. اگر ۷ گاو به تصادف از این گاوداری انتخاب شوند، احتمال اینکه حداقل ۴ گاو بیمار باشند چقدر است؟

$p=0.1, q=0.9, \mathrm{n}=7, x=4,5,6,7$

$(p+q)^7=p^7+7p^6q+21p^5q^2+35p^4q^3+35p^3q^4+21p^2q^5+7pq^6+q^7$

$P=p^7+7p^6q+21p^5q^2+35p^4q^3=(0.1)^7+7(0.1)^6(0.9)+21(0.1)^5(0.9)^2+35(0.1)^4(0.9)^3$

مثال: زن و مردی که ناقل بیماری هموفیلی هستند، با هم ازدواج می کنند. اگر احتمال داشتن فرزند سالم برابر $\frac{3}{4}$ و فرزند بیمار $\frac{1}{4}$ باشد، حساب کنید احتمال آن را که از ۴ فرزندی که متولد می شوند همه فرزندان سالم باشند.

$p=\frac{3}{4}$, $q=\frac{1}{4}$, n=4 , x=4

$(p+q)^4=p^4+4p^3q+6p^2q^2+4pq^3+q^4$

$P=p^4=(\frac{3}{4})^4$

or

$$p_{(x)}=\begin{bmatrix} n \\ x \end{bmatrix}p^xq^{n-x}=\frac{n!}{(n-x)!x!}p^xq^{n-x}\to P_{(4)}=\begin{bmatrix} 4 \\ 4 \end{bmatrix}p^4q^{4-4}=\frac{4!}{0!4!}p^4q^0=p^4=(\frac{3}{4})^4$$

مثال: زوجی می خواهند ۵ فرزند داشته باشند:

الف- به چند طریق می توانند ۲ پسر و ۳ دختر داشته باشند؟

ب- احتمال داشتن ۲ فرزند پسر چقدر است؟

$p=\frac{1}{2}$, $q=\frac{1}{2}$, $n=5$

الف-

$C_2^5=\frac{5!}{3!2!}=10$

ب-

$$p_{(x)}=\begin{bmatrix} n \\ x \end{bmatrix}p^xq^{n-x}\to P_{(2)}=\begin{bmatrix} 5 \\ 2 \end{bmatrix}p^2q^3=10(\frac{1}{2})^2(\frac{1}{2})^3=10(\frac{1}{2})^5$$

مثال: اگر احتمال گرفتن تقلب در یک امتحان ۰/۴ باشد، حساب کنید احتمال آن را که از ۵ نفری که مشکوک به تقلب کردن هستند، حداقل ۲ تقلب گرفته شود.

$p=0.4$, $q=0.6$, $n=5$, x=2,3,4,5

$$P=C_2^5p^2q^3+C_3^5p^3q^2+C_4^5p^4q+C_5^5p^5=10(0.4)^2(0.6)^3+10(0.4)^3(0.6)^2+5(0.4)^4(0.6)+(0.4)^5$$

نکته: بر اساس جدول توزیع فراوانی متغیر گسسته ما می توانیم میانگین و واریانس را مطابق آنچه قبلاً گفته شده محاسبه کنیم. این مقدار برای توزیع دو جمله ای برابر است با:

$\mu=np$
$\sigma^2=npq$

مثال: در آزمایش ۱۰۰ پرتاب یک سکه میانگین تعداد شیر و انحراف معیار را بدست آورید.

$$\mu=np=100\times\frac{1}{2}=50$$

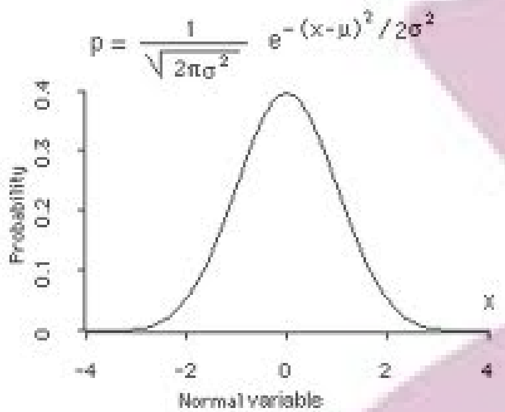
$$\sigma=\sqrt{npq}=\sqrt{100\times\frac{1}{2}\times\frac{1}{2}}=5$$

مثال: اگر میانگین یک توزیع دوجمله ای برابر ۴ و واریانس برابر ۳ باشد، مقادیر n، p و q را محاسبه کنید.

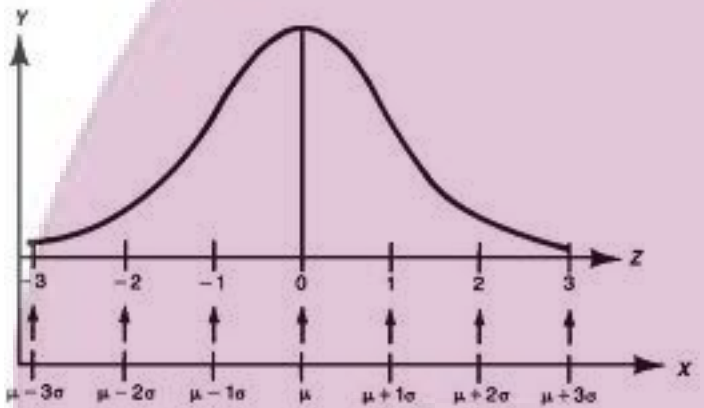
$\mu=np \Longrightarrow 4=np$
$\sigma^2=npq \Longrightarrow 3=npq \Longrightarrow 3=4q \Longrightarrow q=\frac{3}{4}$
$p+q=1 \Longrightarrow p=1-q=1-\frac{3}{4}=\frac{1}{4}$
$4=np \Longrightarrow 4=n\times\frac{1}{4} \Longrightarrow n=16$

توزیع نرمال

توزیع احتمال یک متغیر تصادفی پیوسته است (مثل قد، وزن و .....) و دارای خصوصیاتی است:
منحنی نرمال یک منحنی متقارن است که به وسیله میانگین μ و انحراف معیار σ مشخص می شود. نیمی از داده ها بیشتر از میانگین و نیمی از داده ها کمتر از میانگین هستند. اگر متغیری دارای توزیع نرمال باشد، بدین مفهوم است که بیشترین تعداد افراد در حول و حوش میانگین قرار دارند و هر چه از میانگین به دو سمت دور می شویم، تعداد افراد کم می شوند. منحنی نرمال از رابطه زیر تعریف می شود:



برای محاسبه احتمال از طریق توزیع نرمال باید سطح زیر منحنی را محاسبه کنیم. برای این منظور ابتدا توزیع نرمال را به توزیع نرمال استاندارد تبدیل می کنیم که دارای میانگین صفر و انحراف معیار یک می باشد و به صورت زیر محاسبه می شود:



The Translation of X to Z by the Transformation $Z = (X-\mu)/\sigma$

$$z = \frac{x-\mu}{\sigma}$$

بنابراین کافی است برای نقطه مورد نظر z را محاسبه کرده و سپس از روی جدول توزیع نرمال استاندارد (جدول۳ صفحه ۳۴۶ کتاب)، مقدار A یا سطح زیر منحنی یا همان احتمال مورد نظر را بخوانیم (در این جدول احتمال فاصله ۰ تا z محاسبه شده به دست می آید).

از آنجایی که منحنی نرمال استاندارد یک منحنی متقارن است، سطح زیر منحنی در دو نیمه با هم برابر و مساوی ۰/۵ است.

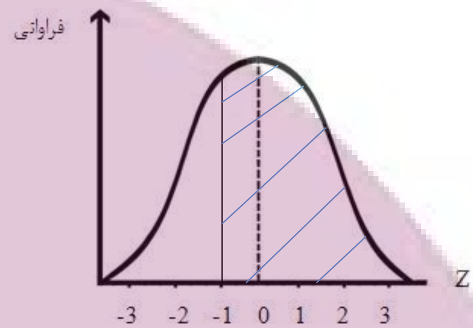
الف- توزیع نرمال به عنوان یک متغیر تصادفی پیوسته

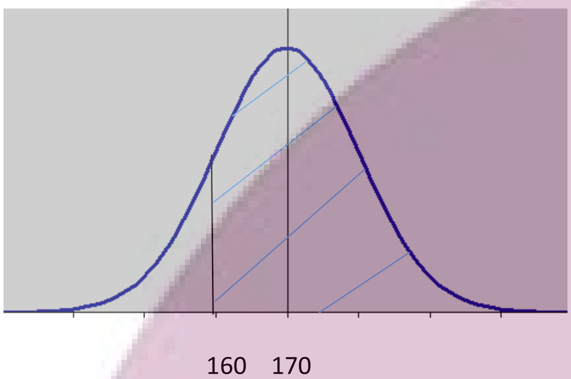
در این حالت متغیر پیوسته است و μ و σ را داریم.

مثال: اگر قد افراد موجود در کلاس دارای توزیع نرمال با میانگین ۱۷۰cm و انحراف معیار ۱۰cm باشد، به سؤالات زیر پاسخ دهید.

الف– اگر فردی به تصادف انتخاب شود، احتمال اینکه قد او از ۱۶۰ cm بیشتر باشد، چقدر است؟

نکته: حتماً مسائل توزیع نرمال را با شکل حل کنید. قسمت هاشور زده در شکل، پاسخ مسأله خواهد بود.



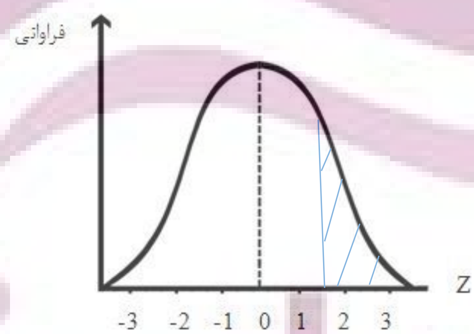


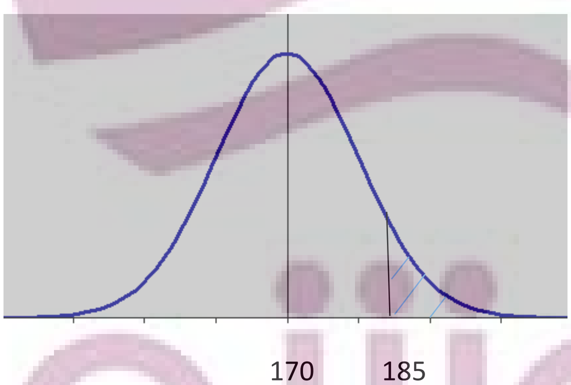
$$z=\frac{x-\mu}{\sigma}=\frac{160-170}{10}=-1$$

z	A
-1	0.3418

P(x>160) $\Longrightarrow$ P(z>-1) $\Longrightarrow$ P= 0.3418+0.5=0.8418

ب– اگر فردی به تصادف انتخاب شود، احتمال اینکه قد او از ۱۸۵ cm بیشتر باشد چقدر است؟



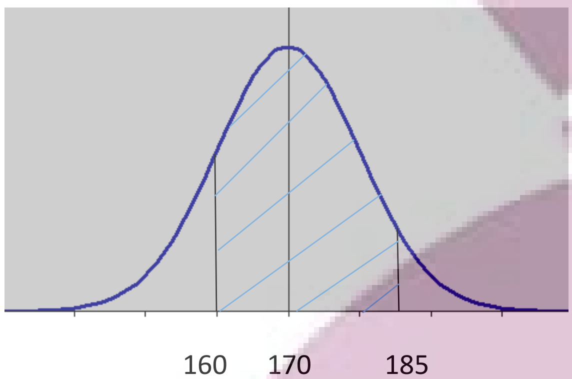


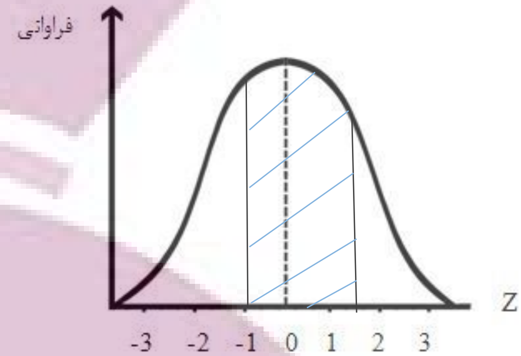
$$z=\frac{185-170}{10}=1.5$$

z	A
1.5	0.4332

P(x>185) $\Longrightarrow$ P(z>1.5) $\Longrightarrow$ P= 0.5-0.4332= 0.0668

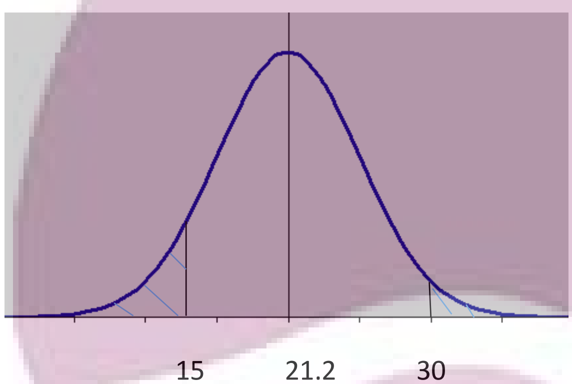
ج- احتمال اینکه قد او بین ۱۶۰ cm و ۱۸۵ cm باشد؟

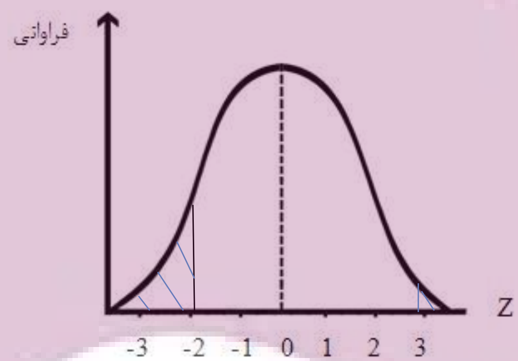




P(160<x<185) $\Longrightarrow$ P(-1<x<1.5) $\Longrightarrow$ P= 0.3418+0.4332= 0.775

مثال: اگر در یک توزیع نرمال یک متغیر پیوسته مقدار میانگین و انحراف معیار به ترتیب ۲۱/۲ و ۳/۱ باشد، پیدا کنید احتمال آن را که یک داده به طور تصادفی انتخاب و بزرگتر از ۳۰ و کوچکتر از ۱۵ باشد.





$z=\frac{x-\mu}{\sigma}=\frac{15-21.2}{3.1}=-2$

$z=\frac{30-21.2}{3.1}=2.84$

z	A
-2	0.4772
2.84	0.4977

P(x<15 , x>30) $\Longrightarrow$ P(z<-2 , z>2.84) $\Longrightarrow$ P= (0.5-0.4772) + (0.5-0.4977)= 1- (0.4772+0.4977)= 0.0251

نکته: چنانچه فرض بر این باشد که مقادیر متغیر پیوسته (قد، وزن، ارتفاع گیاه، عملکرد و .....) با تقریب درج شده اند، بنابر نیاز از مقادیر x ، ۰/۵ واحد کم یا به آن ۰/۵ واحد اضافه می کنیم (به تمرین ۶-۷ صفحه ۱۸۷ حتماً توجه شود).

ب- توزیع نرمال به عنوان تقریبی از توزیع دوجمله ای

در این حالت متغیر گسسته است. ۳ شرط توزیع دو جمله ای صادق است و μ و σ را نداریم که از روابط زیر محاسبه می کنیم.

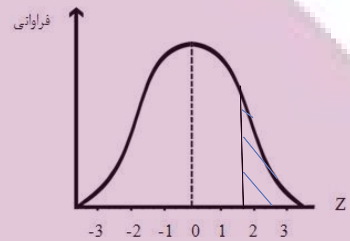
$\mu=np$

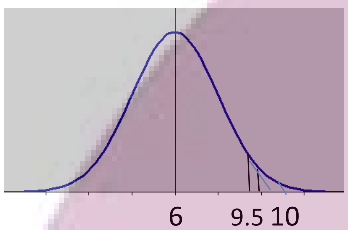
$\sigma=\sqrt{npq}$

در صورت نیاز باید ۰/۵ واحد از x کم شود یا به آن اضافه گردد.

مثال: تاسی را ۳۶ مرتبه پرتاب می کنیم. حساب کنید:

الف- احتمال آن را که حداقل ۱۰ مرتبه عدد ۴ بیاید.





$\mu=np=36\times\frac{1}{6}=6$

$\sigma=\sqrt{36\times\frac{1}{6}\times\frac{5}{6}}=2.24$

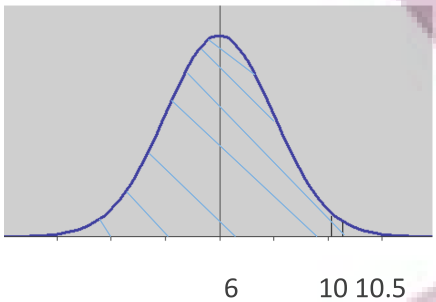
چون احتمال یک متغیر گسسته را از طریق یک توزیع پیوسته (نرمال) بدست می آوریم و احتمال حداقل ۱۰ مرتبه را می خواهیم یعنی خود ۱۰ را نیز شامل می شود، از x (۱۰) مقدار ۰/۵ واحد کم می کنیم.

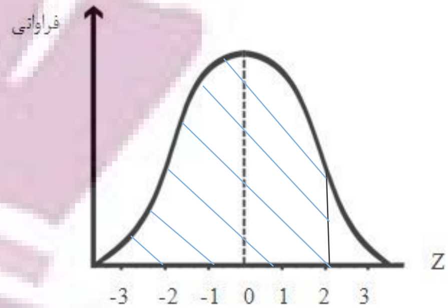
$z=\frac{x-\mu}{\sigma}=\frac{9.5-6}{2.24}=1.56$

z	A
1.56	0.4406

$P(x\geq 10)\Longrightarrow P(z\geq 1.56)\Longrightarrow P=0.5-0.4406=0.0594$

ب- احتمال آنکه حداکثر ۱۰ مرتبه ۴ بیاید.





چون احتمال یک متغیر گسسته را از طریق یک توزیع پیوسته (نرمال) بدست می آوریم و احتمال حداکثر ۱۰ مرتبه را می خواهیم یعنی خود ۱۰ را نیز شامل می شود، به x (۱۰) مقدار ۰/۵ واحد اضافه می کنیم.

$z=\frac{10.5-6}{2.24}=2.01$

z	A
2.01	0.4778

$P(x\leq10) \Longrightarrow P(z\leq2.01) \Longrightarrow P=0.5+0.4778=0.9778$

ج- احتمال اینکه بیش از ۱۰ مرتبه عدد ۴ بیاید.

چون احتمال بیش از ۱۰ مرتبه را می خواهیم و خود ۱۰ مدنظر نیست، پس نیازی به ۰/۵ واحد کم یا زیاد کردن نیست.

$z=\frac{10-6}{2.24}=1.78$

z	A
1.78	0.4625

$P(x>10) \Longrightarrow P(z>1.78) \Longrightarrow P=0.5-0.4625=0.0375$

د- احتمال اینکه درست ۱۰ مرتبه عدد ۴ بیایید.

چون احتمال خود ۱۰ مرتبه را می خواهیم احتمال $9.5<x<10.5$ را محاسبه می کنیم (یکبار ۰/۵ واحد کم و یکبار ۰/۵ واحد اضافه می کنیم).

$z=\frac{9.5-6}{2.24}=1.56$

$z=\frac{10.5-6}{2.24}=2.01$

z	A
1.56	0.4406
2.01	04778

$P(9.5<x<10.5) \Longrightarrow P(1.56<z<2.01) \Longrightarrow P= 0.4778-.04406= 0.0372$

نکته (رابطه توزیع دوجمله ای و توزیع نرمال)

توزیع های دوجمله ای که در آنها $p=q=\frac{1}{2}$ می باشد، شکل متقارن دارند. چنانچه n افزایش یابد، شکل این توزیع به تدریج به منحنی نرمال شباهت می یابد به طوری که در حالت $n=\infty$، این دو منحنی کاملاً بر هم منطبق می شوند. یعنی منحنی نرمال حد توزیع دوجمله ای در حالت p=q است.

توزیع پواسن (پویسن) یا وقایع نادر

این توزیع برای وقایعی که به ندرت اتفاق می افتند صادق است، مثلاً احتمال پنچری لاستیک ماشین ها در یک روز در یک خیابان خاص یا احتمال گزارش یک بیماری نادر. زمانی که احتمال وقوع پیشامدی (p) کوچک باشد به طوری که $np<5$ باشد، آن پیشامد را نادر می گویند. در توزیع دوجمله ای اگر به تدریج که n افزایش می یابد p کوچک تر شود، مقدار $\lambda=np$ ثابت می ماند. یعنی زمانی که n به سمت بی نهایت و p به سمت صفر میل می کند و np ثابت می ماند، توزیع دوجمله ای به توزیع پواسن تبدیل می شود.

در توزیع پواسن میانگین و واریانس با هم مساوی و برابر λ می باشند. برای محاسبه احتمال وقوع x پیشامد در n آزمایش از فرمول زیر استفاده می شود:

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

$e\approx 2.71828$

$$\begin{cases}\mu=\lambda\\ \sigma^2=\lambda\end{cases} \Longrightarrow \sigma=\sqrt{\lambda}$$

مثال: میزان مرگ و میر برای یک نوع بیماری ۶ در هزار است ، احتمال وقوع۴ مرگ را در یک گروه ۵۰۰ نفری محاسبه نمایید.

$\lambda=np=500(0.006)=3 \qquad P(x)=\frac{\lambda^x e^{-\lambda}}{x!}$

$P=\frac{\lambda^4 e^{-\lambda}}{x!}=\frac{3^4 e^{-3}}{4!}=\frac{81\times(0.0498)}{24}=0.168$

چنانچه این مثال از طریق بسط دوجمله ای حل شود به صورت زیر عمل می کنیم (از آن جایی که n زیاد و p کوچک می باشد

توزیع دو جمله ای به توزیع پواسن تبدیل شده و محاسبه فرمول دو جمله ای دشوار است):

$\binom{n}{x} p^x q^{n-x}$

$\binom{500}{4}(p)^4(q)^{496}=\frac{500!}{496!4!}(0.006)^4(0.994)^{496}$

توزیع چندجمله ای

چنانچه وقایع E_1 و E_2 ، ... و E_k به ترتیب با احتمال P_1 ، P_2 ، ... و P_k اتفاق بیفتد ، احتمال اینکه وقایع E_1 و E_2 ، ... و E_k به ترتیب X_1 و X_2 ، ... و X_k بار اتفاق بیفتد عبارتست از :

$\frac{N!}{X_1!X_2!\dots X_k!}P_1^{X_1}P_2^{X_2}\dots P_k^{X_k}$

$X_1+X_2+\dots+X_k=N$

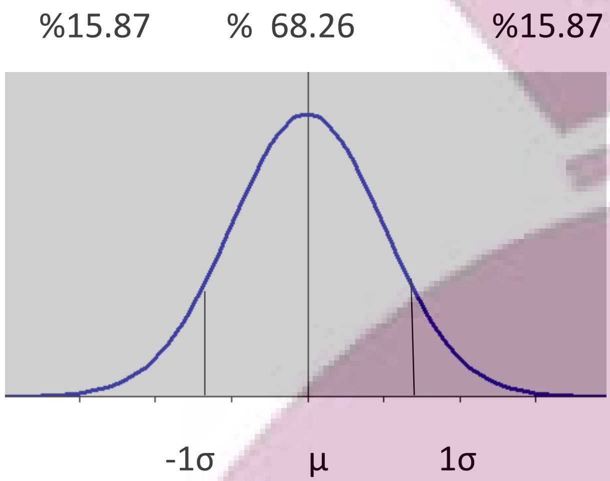
مثال: تاس متعادلی ۶ بار پرتاب می شود. احتمال اینکه عدد یک ۱ بار، عدد دو ۲ بار و عدد سه ۳ بار ظاهر شود چقدر است؟

$\frac{N!}{X_1!X_2!\dots X_k!}P_1^{X_1}P_2^{X_2}\dots P_k^{X_k}$

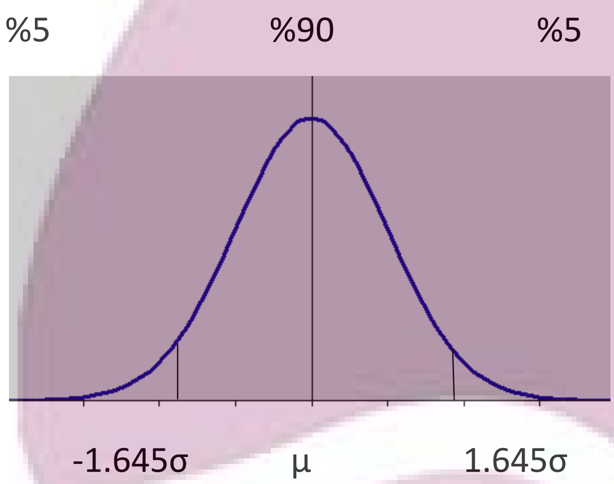
$\frac{6!}{1!2!3!}=(\frac{1}{6})^1(\frac{1}{6})^2(\frac{1}{6})^3=60(\frac{1}{6})^6$

نکته: تقسیمات متعارف سطح زیر منحنی نرمال به صورت زیر می باشد:

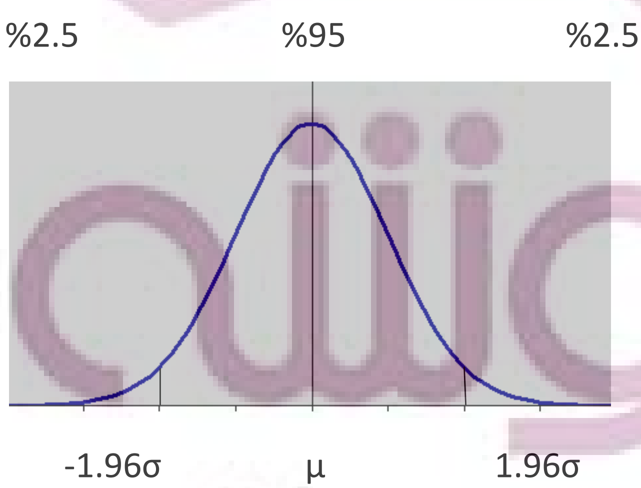
۱- نقاط واقع به فاصله یک انحراف معیار در اطراف میانگین



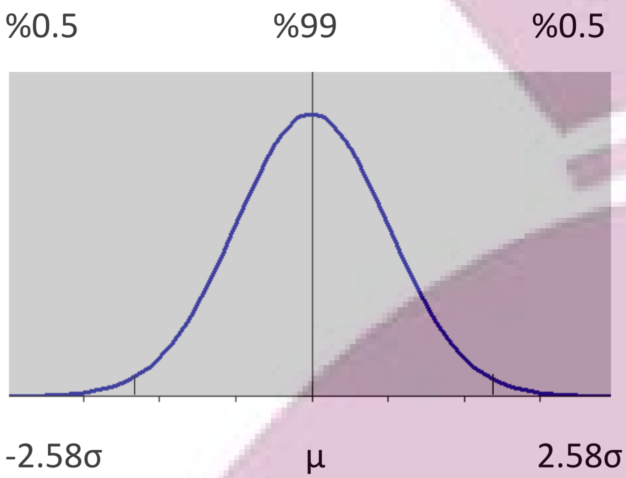
۲- نقاط واقع به فاصله ۱/۶۴۵ انحراف معیار در اطراف میانگین



۳- نقاط واقع به فاصله۱/۹۶ انحراف معیار در اطراف میانگین



۴- نقاط واقع به فاصله ۲/۵۸ انحراف معیار در اطراف میانگین



پایان