فصل اول............... کلیات

فصل دوم............... سازماندهی داده‌ها

فصل سوم............... توصیف عددی مشاهدات

فصل چهارم............... احتمال

فصل پنجم............... متغیرهای تصادفی و تابع احتمال

فصل ششم............... امید ریاضی و کاربردهای آن

فصل هفتم............... توزیع‌های آماری

فصل اول

کلیات

معنای انگلیسی آمار، Statistics است که از کلمه لاتین Status مشتق شده که یکی از معانی آن دولت است. در ابتدا آمار به عنوان یک هنر و علم حکومت‌داری مطرح بوده و اولین استادان این علم دانشمندان و کارگزاران علوم سیاسی قلمداد می‌شدند. گرچه علم آمار یکی از اختراعات مهم قرن بیستم به شمار می‌آید، مبنای آن یعنی نظریه احتمال توسط پاسکال و فرما در قرن ۱۷ پایه‌ریزی شد. البته پیش از آن کاردان و گالیله در قرن ۱٦ به چنین مسائلی پرداخته بودند.

علم آمار: به مجموعه روش‌های علمی جهت جمع‌آوری، تلخیص، تنظیم، تجزیه و تحلیل و نتیجه‌گیری از داده‌ها که به منظور تامین اهدافی از قبل تعیین شده انجام می‌گیرد، علم آمار می‌گویند.

به طور کلی علم آمار را می‌توان از دو دیدگاه مورد بررسی قرار داد.

۱- *آمار توصیفی:* به مجموعه روش‌هایی که برای سازماندهی، خلاصه کردن و توصیف اطلاعات به کار می‌رود، گفته می‌شود و شامل بخش‌های(طبقه‌بندی داده‌ها، شاخص‌های مرکزی، شاخص‌های پراکندگی و رسم نمودارها) است.

۲- *آمار استنباطی:* به مجموعه روش‌هایی که به وسیله آنها نتایج به دست آمده از نمونه را به جامعه تعمیم می‌دهند، گفته می‌شود و شامل بخش‌هایی همانند (احتمال، توابع توزیع، نمونه‌گیری، برآورد، آزمون فرض) است.

جامعه آماری: مجموعه افراد یا چیزهایی که حداقل در یک ویژگی مشترک باشند و به خوبی تعریف و مشخص شده باشند. مثال: دانشجویان دانشگاه مازندران.

نمونه: مطالعه یک یک افراد جامعه بعلت هزینه زیاد و کمی وقت و یا نداشتن امکانات کافی امکان‌پذیر نیست، بنابراین تنها بخشی از جامعه را که طبق ضوابطی مقبول انتخاب می‌شود و مطالعه آن به جای تمام جامعه مقدور است، نمونه می‌نامیم.

صفت مشخصه: صفت(هایی) هستند که در همه اعضای جامعه مشترک‌اند و جامعه آماری و محدوده مطالعه با این صفت‌ها مشخص می‌شود. مثال: بررسی متوسط قد تمامی دانشجویان دانشگاه مازندران که در آن صفت مشخصه دانشجوی دانشگاه مازندران بودن است و برای تمامی اعضا یکسان است.

صفت متغیر: صفت(هایی) هستند که از هر عضو به عضو دیگر جامعه تغییر می‌کند. از این رو آنها را به اختصار، متغیر می‌گویند. این متغیرها هستند که عملاً در حین تحقیق مورد سوال و اندازه گیری قرار می‌گیرند. متغیرها را معمولاً با حروف بزگ مانند X، Y، Z ،.... نمایش می‌دهند.

برای مثال در بررسی متوسط قد دانشجویان دانشگاه مازندران، قد یک صفت متغیر است. زیرا از هر عضو جامعه به عضو دیگر متغیر می‌باشد.

توجه: داده‌ها مقادیر اندازه‌گیری شده یک متغیر هستند.

متغیرها خود بر دو نوع هستند:

الف) متغیرهای کمی: متغیرهایی هستند که قابل شمارش و اندازه‌گیری هستند و همیشه نتیجه اندازه‌گیری آنها یک عدد است. همانند قد، وزن، فشار خون و دمای هوا.

ب) متغیرهای کیفی: متغیرهایی هستند که غیرقابل شمارش و اندازه‌گیری‌اند. اینگونه متغیرها ماهیت اندازه‌پذیری ندارند و تنها با وضعیت‌هایی که دارند، ثبت می‌شوند. همانند: رنگ چشم، گروه خونی، درجه کیفیت کالا، برچسب انرژی، رنگ‌ها.

به طور کلی می‌توان برای متغیرها تقسیم‌بندی زیر را در نظر گرفت:

- متغیر
  - کمی
    - گسسته : متغیرهایی که قابل شمارش هستند. همانند: تعداد فرزندان، تعداد تصادفات
    - پیوسته: متغیرهایی که اندازه پذیرند و مقادیر آنها متعلق به اعداد حقیقی است. همانند: وزن، قد، دمای هوا
  - کیفی
    - اسمی : اختصاص یک عدد یا اسم یا ... به هر سطح متغیر تنها برای شناسایی و تمیز دادن آنها. همانند اختصاص اعداد ۱ تا ٤ به هر یک از گروههای خونی.
    - رتبه ای: هر گاه عدد یا اسم یا ... اختصاص داده شده به هر سطح برتری آنها را نشان دهد.
    - مثال: رتبه بندی کارگران یک کارخانه از نظر مهارت کاری، برچسب‌های انرژی.

پارامتر: اگر شاخص‌های جامعه را با اندازه‌گیری تمامی عناصر جامعه به دست آوریم آنها را پارامتر می‌نامند.

آماره: اگر شاخص‌های مورد نظر جامعه با استفاده از بخشی از جامعه(نمونه گیری) به دست آمده باشند، آماره نامیده می‌شوند.

مثال: بررسی متوسط درآمد خانوارهای یک کشور از طریق سرشماری و نمونه‌گیری.

اندازه‌گیری و سطوح مختلف متغیرها:

اندازه‌گیری: تخصیص اندازه‌هایی معین به متغیرهایی که قصد مطالعه آنها را داریم اندازه‌گیری نامیده می‌شود.

فرض کنید بخواهیم وزن سیب را اندازه بگیریم. اینکار با اندازه‌گیری و تخصیص عددی مثل x به آن انجام می‌شود. حال اگر بخواهیم تُردی سیب را اندازه بگیریم به چه صورت خواهد بود. آیا روشی برای اندازه گیری میزان تردی سیب وجود دارد؟

استیونز، استاد روانشناسی دانشگاه هارواد، در سال ۱۹٤٦ برای اندازه‌گیری سطوح مختلف متغیرها، چهار نوع مقیاس را به شرح زیر معرفی کرد:

الف) مقیاس اسمی: هر گاه x، که معمولاً یک عدد طبیعی است، تنها برای شناسایی افراد یا اشیاء یا مکان به کار رود و هیچ کمیتی را بیان نکند، آنرا یک مقیاس اسمی می‌نامند. بعنوان مثال: اسامی کتاب‌ها، مقاطع تحصیلی، رنگ چشم.

ب) مقیاس ترتیبی: هرگاه x، برتری یا ترتیب سطوح را نشان دهد. بعنوان مثال: رتبه بندی کارگران یک کارخانه بر اساس مهارت کاری، برچسب‌های انرژی.

ج) مقیاس فاصله‌ای: هرگاه x، بیانگر یک کمیت باشد ولی مبدا معینی نداشته باشد. لازم به ذکر است که مقیاس فاصله‌ای نسبت دو تفاضل یا دو فاصله را حفظ می‌کند. به سخنی دیگر اگر با این مقیاس برای $a_۱, a_۲, a_۳, a_۴$ ویژگی t با اعداد $x_۱,x_۲,x_۳,x_۴$ اندازه‌گیری شود بایستی نسبت $\frac{x_۴-x_۳}{x_۲-x_۱}$ ثابت بماند و به واحد اندازه‌گیری بستگی نداشته باشد.

به عنوان مثال فرض کنید ویژگی t دمای چهار جسم باشد و بر حسب درجه سانتیگراد داشته باشیم $x_۱=۱۰, x_۲=۵, x_۳=۲۰, x_۴=۴۵$ برای این چهار عدد داریم $\frac{x_۴-x_۳}{x_۲-x_۱}=\frac{۴۵-۲۰}{۱۵-۱۰}=۵$. اینک برای این داده‌ها با مقیاس فارنهایت داریم: $y=\frac{۹}{۵}x+۳۲$. در نتیجه: $y_۱=۵۰, y_۲=۵۹, y_۳=۶۸, y_۴=۱۱۳$. برای این داده‌های جدید داریم: $\frac{y_۴-y_۳}{y_۲-y_۱}=\frac{۱۱۳-۶۸}{۵۹-۵۰}=۵$. همانطور که مشاهده می‌شود نسبت $\frac{x_۴-x_۳}{x_۲-x_۱}$ در هر دوحالت برابر است. بنابراین حرارت یک مقیاس فاصله‌ای است.

این مقیاس بیشتر در علوم روانشناسی، تعلیم وتربیت، فیزیک و برای سنجش هوش و ارزیابی آزمون‌ها کاربرد دارد.

د) مقیاس نسبتی: هرگاه x، دقیقا بیانگر یک کمیت باشد و شامل یک مبداء معین باشد. توجه داشته باشید که در مقیاس نسبتی، نسبت حفظ می‌شود و به واحد اندازه‌گیری بستگی ندارد.

برای مثال: $\frac{۶۰۰۰gr}{۲۰۰۰gr}=\frac{۶kg}{۲kg}=۳$

به طور کلی می‌توان خصوصیات انواع مقیاس‌ها را در جدول زیر خلاصه کرد.

انواع مقیاس	ترتیب	ارزش	مبداء صفر قراردادی	مبداء صفر مطلق
اسمی	ندارد	ندارد	ندارد	ندارد
رتبه‌ای	دارد	ندارد	ندارد	ندارد
فاصله‌ای	دارد	دارد	دارد	ندارد
نسبتی	دارد	دارد	دارد	دارد

مسائل

۱- آمار استنباطی و توصیفی را تعریف کنید؟

۲- تفاوت آماره و پارامتر را بنویسید؟

۳- مقیاس های اندازه‌گیری را با ذکر مثال توضیح دهید؟

٤- مقیاس هر یک از متغیر های زیر را تعیین کنید؟

نژاد، زمان، درآمد، شغل، گروه خونی

۵- در هر یک از موراد زیر جمعیت(جامعه آماری)، نمونه، متغیر و مقیاس را مشخص کنید؟

الف- دانش آموزان یک مدرسه و قدرت دید آنان

ب- تعداد غلط‌ها در یک صفحه از کتاب ریاضی دوم دبیرستان.

فصل دوم

سازماندهی داده‌ها

هنگامی که توده‌ای از اطلاعات برای تحقیق گردآوری می‌شود، ابتدا سازمان‌بندی و خلاصه کردن آنها به طریقی که به صورت معنی‌داری قابل درک و ارتباط باشند، ضروری است. روشهای آمار توصیفی به همین منظور بکار برده می‌شوند. غالباً مفیدترین و در عین حال اولین قدم در سازمان دادن به داده‌ها، مرتب کردن داده‌ها بر اساس یک ملاک منطقی است و سپس استخراج شاخص‌های مرکزی و پراکندگی و در صورت لزوم محاسبه همبستگی میان دو دسته اطلاعات و استفاده از تحلیل‌های پیشرفته‌تر نظیر رگرسیون و پیش‌بینی می‌باشد. در یک جمع‌بندی، با استفاده مناسب از روشهای آمار توصیفی می‌توان دقیقاً ویژگی‌های یک دسته از اطلاعات را بیان کرد. آمار توصیفی همیشه برای تعیین و بیان ویژگی‌های اولیه اطلاعاتِ پژوهش‌ها، بکار برده می‌شوند.

برای شروع بحث در مورد آمار توصیفی ابتدا به بیان چند تعریف که در طول کتاب مکرراً با آنها برخورد خواهیم کرد، می‌پردازیم.

داده: فرض کنید می‌خواهیم ویژگی(هایی) از یک جمعیت را که معمولاً یک متغیر است مطالعه کنیم، اگر این متغیر را در مورد یک یک افراد جمعیت یا نمونه گرفته شده از جمعیت، اندازه‌گیری کنیم، یک مجموعه از اعداد به دست می‌آید که آن را داده می‌نامند. داده ها بر دو نوع گسسته و پیوسته‌اند.

الف) داده‌های گسسته: داده‌هایی هستند که از راه اندازه‌گیری با مقیاس اسمی، ترتیبی(رتبه‌ای) یا شمارش به دست می‌آیند. بعنوان مثال تعداد فرزندان خانواده، تعدا تصادفات در یک چهارراه.

ب) داده‌های پیوسته: داده‌هایی هستند که از راه اندازه‌گیری فاصله‌ای یا نسبتی به دست می‌آیند. بعنوان مثال طول عمر یک لامپ، قد یا وزن افراد.

جدول‌های آماری: داده‌ها اغلب به صورت انبوهی از اعداد ارائه می‌شوند و به خودی خود خام هستند، برای اینکه بتوان آنها را پخته کرد و حقایق را جویا شد بایستی سه کار عمده زیر را انجام داد.

۱- دسته‌بندی اطلاعات و تهیه جدول توزیع فراوانی
۲- از روی جدول‌ها، نمودارهایی برای توصیف بصری اطلاعات رسم کنیم.
۳- آنها را در یک یا چند عدد که بتواند ویژگی داده‌ها را به خوبی آشکار کند خلاصه کنیم.

تنها پس از طی این مراحل می‌توان قوانین شانسی حاکم بر آنان را پیدا کرده و به برداشت‌های آماری و تهیه گزارش نهایی درباره ویژگی مورد مطالعه پرداخت. نمایش داده‌ها با نظمی خاص، در چند سطر یا ستون را یک جدول آماری می‌نامند. در کارهای آماری برای اینکه داده‌ها را خلاصه کنند آنها را در جدولی به نام جدول توزیع فراوانی تنظیم می‌کنند. در ادامه نحوه تنظیم چنین جداولی را برای داده‌های کیفی و کمی توضیح خواهیم داد.

جدول توزیع فراوانی برای داده‌های کیفی و کمی گسسته:

برای دسته‌بندی داده‌های کیفی و کمی گسسته باید هریک از سطوح(مقادیر) متغیر را به همراه فراوانی آنها در یک جدول قرار دهیم.

مثال۱: فرض کنید پزشکی ۲۰ بیمار قلبی دارد و گروه خونی آنان را به شرح زیر تعیین نموده است. جدول فراوانی برای این داده‌ها به صورت زیر خواهد بود.

B A O AB O A A A O O A A B B AB O AB AB O O

گروه خونی	متغیر اسمی	خط و نشان	فراوانی (f_i)
A	۱	~~////~~ /	٦
B	۲	///	۳
AB	۳	////	٤
O	٤	~~////~~ //	۷
جمع			۲۰

توجه داشته باشید که همواره مجموع فراوانی‌ها برابر با تعداد داده‌ها می‌باشد. یعنی $\sum f_i = n$.

جدول توزیع فراوانی برای داده‌های کمی پیوسته:

برای داده‌های پیوسته از آنجا که تعداد داده‌ها زیاد و فراوانی کم است، استفاده از روش قبلی برای تشکیل جدول توزیع فراوانی دیگر پاسخگو نیست. برای دسته‌بندی داده‌های کمی پیوسته بایستی مراحل زیر را انجام داد.

الف) تعیین تعداد طبقات (K): تعداد طبقات را به تناسب تعداد داده‌ها و یا با استفاده از روابط $n = ۲^K$ و یا $K = ۱ + ۳/۳۲۲\log(n)$ تعیین می‌کنیم. که در آن n تعداد داده‌ها و K تعداد طبقات است. آماردانان توصیه می‌کنند که، تعداد طبقات به گونه‌ای انتخاب شود که تعداد داده‌های هر طبقه کمتر از ۵ و بیشتر از ۲۰ نباشد. معمولاً تعداد طبقات را بین ۵ الی ۱۵ اختیار می‌کنند.

ب) تعیین فاصله طبقات (C): برای یافتن فاصله طبقات ابتدا بایستی دامنه تغییرات داده‌ها (برد داده‌ها) را تعیین کنیم.

دامنه تغییرات(برد): عبارت است از اختلاف بین کوچکترین و بزرگترین عدد. $R = Max(x_i) - Min(x_i)$

حال فاصله طبقات را از رابطه $C=\frac{R}{K}$ محاسبه خواهیم کرد. در صورتیکه حاصل $\frac{R}{K}$ عدد صحیحی نباشد آن را با اولین عدد صحیح بزرگتر از آن تقریب خواهیم زد.

ج) حدود طبقات: حدود طبقات را با شروع از کمترین مقدار تعیین کرده و داده‌هایی را که در هر طبقه قرار می‌گیرند شمارش کرده و بعنوان فراوانی آن طبقه ثبت می‌کنیم.

مثال۲: داده‌های زیر میزان پرش ارتفاع ۲۵ دانشجوی رشته تربیت بدنی را بر حسب سانتی‌متر نشان می‌دهد. جدول فراوانی برای این داده‌ها به صورت زیر خواهد بود.

۹۸ ۱۰۱ ۹۷ ۱۰۴ ۱۰۲ ۱۱۰ ۱۰۳ ۱۰۶ ۱۱۰ ۱۰۴ ۱۰۳ ۹۸ ۱۰۵ ۱۰۰ ۱۰۹ ۱۰۳ ۱۰۴ ۹۹ ۹۸

۱۰۹ ۱۰۵ ۱۰۳ ۱۱۰ ۱۰۴ ۱۰۵

برای تعیین تعداد طبقات با استفاده از رابطه $n=2^K$ داریم $25=2^K$. برای $k=4$ حاصل 2^4، عدد شانزده است که از ۲۵ کوچکتر است و برای $k=5$ داریم: $2^5=32$ که بزرگتر از ۲۵ است بنابراین تعداد طبقات را برابر با ۵ انتخاب خواهیم کرد. برای تعیین فاصله طبقات داریم:

$$R=Max(x_i)-Min(x_i)=110-97=13 \quad \Rightarrow \quad C=\frac{R}{K}=\frac{13}{5}=2/6\simeq 3$$

حال می‌توان جدول توزیع فراوانی را به صورت زیر نوشت.

فراوانی (f_i)	خط و نشان	حدود طبقات
۴	////	۹۷–۱۰۰
۴	////	۱۰۰–۱۰۳
۱۱	~~////~~ ~~////~~	۱۰۳–۱۰۶
۱	/	۱۰۶–۱۰۹
۵	~~////~~	۱۰۹–۱۱۲
۲۵		جمع

در یک جدول توزیع فراوانی برای به دست آوردن اطلاعات بیشتر از داده‌ها می‌توان ستون‌های دیگری را به شرح زیر به آن اضافه کرد.

فراوانی مطلق (f_i): به تعداد داده‌ها در هر طبقه فراوانی مطلق آن طبقه می‌گویند. توجه داشته باشید که مجموع فراوانی‌های مطلق برابر با تعداد داده‌ها است. به عبارت ریاضی: $\sum f_i=n$.

فراوانی نسبی (r_i): اگر فراوانی مطلق هر طبقه را بر تعداد کل داده‌ها (n)تقسیم کنیم، فراوانی نسبی به دست می‌آید، $r_i=\frac{f_i}{n}$.

درصد فراوانی نسبی ($\%r_i$): به صورت $(r_i \times 100)\%$، محاسبه می‌شود. توجه داشته باشید که مجموع درصدهای فراوانی نسبی، برابر ۱۰۰٪ است.

فراوانی تجمعی (F_i): فراوانی تجمعی هر طبقه، برابر است با مجموع فراوانی‌های مطلق تمامی طبقات ماقبل آن بعلاوه فراوانی مطلق خود آن طبقه. $F_i = \sum_{j=1}^{i} f_j$

فراوانی تجمعی نسبی ($R_i = \frac{F_i}{n}$): فراوانی تجمعی نسبی هر طبقه، از تقسیم فراوانی تجمعی آن طبقه بر تعداد کل داده‌ها به دست می‌آید، $R_i = \frac{F_i}{n}$.

نماینده طبقه (X'_i): نقطه وسط هر طبقه را نماینده آن طبقه می‌گویند و از رابطه $X'_i = \frac{L_i + H_i}{2}$ به دست می‌آید که در آن L_i حد پایین طبقه iام و H_i حد بالای طبقه iام است.

مثال۳: برای جدول توزیع فراوانی داده‌های مثال پرش ارتفاع، ستون‌های فراوانی نسبی، تجمعی، تجمعی نسبی و نقاط میانی را به دست آورید.

نقطه میانی (X'_i)	فراوانی تجمعی نسبی (R_i)	فراوانی تجمعی (F_i)	درصد فراوانی نسب ($\%r_i$)	فراوانی نسبی (r_i)	فراوانی (f_i)	خط و نشان	حدود طبقات
۹۸	۰/۱٦	٤	٪۱٦	۰/۱٦	٤	////	۹۷-۱۰۰
۱۰۱	۰/۳۲	۸	٪۱٦	۰/۱٦	٤	////	۱۰۰-۱۰۳
۱۰٤	۰/۷٦	۱۹	٪٤٤	۰/٤٤	۱۱	卌 卌	۱۰۳-۱۰٦
۱۰۷	۰/۸۰	۲۰	٪٤	۰/۰٤	۱	/	۱۰٦-۱۰۹
۱۱۰	۱	۲۵	٪۲	۰/۲۰	۵	卌	۱۰۹-۱۱۲
			٪۱۰۰	۱	۲۵		جمع

مثال٤: مدت زمانی لازم برای طی کردن مسیر منزل تا محل شخصی در یک دوره ۵۰ روزه به صورت زیر است. جدول توزیع فراوانی را برای این داده‌ها به دست آورید.

۶۰ ٤۷ ۸۲ ۹۵ ۸۸ ۷۲ ٦۷ ٦٦ ٦۸ ۱۰۳ ۹۰ ۷۷ ۸٦ ۵۸ ٦٤ ۹۵ ۷٤
۷۲ ۸۸ ۷٤ ۷۷ ۳۵ ۹۰ ٦۳ ٦۸ ۹۷ ۷۰ ٦٤ ۷۰ ۷۰ ۵۸ ۷۸ ۸۹ ٤٤
۵۵ ۸۵ ۸۲ ۸۳ ۷۲ ۷۷ ۷۲ ۸٦ ۵۰ ۹٤ ۹۲ ۸۰ ۹۱ ۷۵ ۷٦ ۷۸

پاسخ: با توجه به ماهیت پیوسته بودن داده‌ها نیاز به طبقه‌بندی داده‌ها داریم. ابتدا دامنه داده‌ها را به صورت، $R = Max(x_i) - Min(x_i) = 103 - 35 = 68$ به دست می‌آوریم. برای یافتن تعداد طبقات با استفاده از

رابطه با استفاده از رابطه $n = ۲^k$، تعداد طبقات، k، برابر با ٦ به دست می‌آید($۶۴ = ۲^۶$، اولین عدد بزرگتر یا مساوی ۵۰ است). فاصله طبقات را به صورت $C = \frac{R}{k} = \frac{۵۰}{۶} = ۸/۳۳ \simeq ۹$ به دست می‌آوریم.

طبقات	حدود واقعی طبقات	خط و نشان	f_i	r_i	F_i	R_i	X'_i
۳۵ - ٤٤	۳٤/٥ - ٤٤/٥	//	۲	۰/۰٤	۲	۰/۰٤	۳۹/٥
٤٥ - ٥٤	٤٤/٥ - ٥٤/٥	//	۲	۰/۰٤	٤	۰/۰۸	٤۹/٥
٥٥ - ٦٤	٥٤/٥ - ٦٤/٥	卌 //	۷	۰/۱٤	۱۱	۰/۲۲	٥۹/٥
٦٥ - ۷٤	٦٤/٥ - ۷٤/٥	卌 卌 ///	۱۳	۰/۲٦	۲٤	۰/٤۸	٦۹/٥
۷٥ - ۸٤	۷٤/٥ - ۸٤/٥	卌 卌 /	۱۱	۰/۲۲	۳٥	۰/۷۰	۷۹/٥
۸٥ - ۹٤	۸٤/٥ - ۹٤/٥	卌 卌 /	۱۱	۰/۲۲	٤٦	۰/۹۲	۸۹/٥
۹٥ - ۱۰٤	۹٤/٥ - ۱۰٤/٥	////	٤	۰/۰۸	٥۰	۱	۹۹/٥
جمع			٥۰	۱			

حدود واقعی طبقات: با توجه به ماهیت پیوسته بودن داده‌ها، این داده‌ها می‌توانند هر مقداری را در مجموعه اعداد حقیقی اختیار کنند. برای مثال اگر مدت زمان رسیدن شخص از منزل تا محل کارش برابر با ٤٤/۷ دقیقه باشد، در کدام طبقه جای می‌گیرد. برای رفع این مشکل از حدود واقعی طبقات استفاده خواهیم کرد. برای به دست آوردن حدود واقعی طبقات به صورت زیر عمل می‌کنیم که: نصف اختلاف بین حد پایین طبقه$i+۱$ام و حد بالای طبقه iام را از حد پایین هر طبقه کم کرده و به حد بالای هر طبقه اضافه می‌کنیم. توجه داشته باشید که پس از انجام این کار بایستی در ستون حدود واقعی طبقات، حد بالای طبقه iام با حد پایین طبقه $i+۱$ام برابر باشد.

توجه: پس از تشکیل ستون حدود واقعی طبقات تمامی کارها از قبیل رسم نمودارها و یافتن نقاط میانی با استفاده از حدود واقعی طبقات خواهد بود.

رسم نمودارها

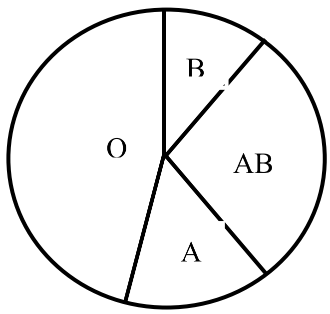
همانطور که قبلاً نیز بدان اشاره شد، دومین مرحله در آمار توصیفی بعد از تهیه جدول توزیعی فراوانی از داده‌ها، رسم نمودارهایی از روی این جداول برای یک نتیجه‌گیری بصری از داده‌ها می‌باشد. در این مرحله با توجه به نوع متغیرها از جداول خاصی برای نمایش داده‌ها استفاده می‌شود که در ادامه به شرح و چگونگی رسم این نمودارها خواهیم پرداخت.

نمودار دایره‌ای: از این نمودار برای نمایش متغیرهای کیفی استفاده می‌شود. در این نمودار اگر مجموع فراوانی‌ها، سطح یک دایره را اشغال کنند، در آن صورت فراوانی هریک از اندازه‌ها یا طبقه‌ها به قطاعی از سطح دایره اختصاص پیدا می‌کند. برای تعیین قطاع مربوط به هر سطح متغیر، فراوانی آن سطح را بر فراوانی کل تقسیم کرده و حاصل را در ۳٦۰ ضرب می‌کنیم. در نتیجه خواهیم داشت:

$$\text{قطاع مربوط به سطح } i \text{ام متغیر} = \frac{f_i}{n} \times ۳۶۰ = r_i \times ۳۶۰$$

مثال۵: گروه خونی ۱۰۰ بیمار بستری در یک بیمارستان به صورت زیر در اختیار است، برای رسم نمودار دایره‌ای این داده‌ها داریم:

گروه خونی	فراوانی f_i	فراوانی نسبی r_i	قطاع دایره (درجه)
A	۱۰	۰/۱۰	۳۶°
B	۲۰	۰/۲۰	۷۲°
AB	۲۵	۰/۲۵	۹۰°
O	٤۵	۰/٤۵	۱۶۲°

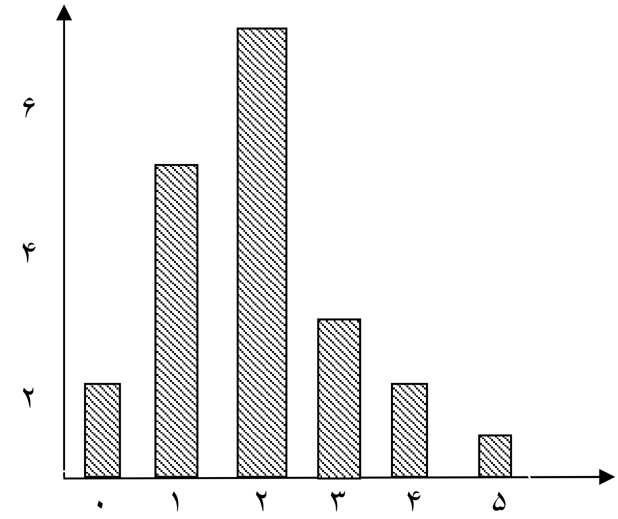


نمودار دایره‌ای برای گروه خونی ۱۰۰ بیمار

نمودار ستونی: از این نمودار برای نمایش داده‌های کمی گسسته استفاده می‌شود. برای رسم این نمودار روی محور افقی (محور Xها)، مقادیر متغیرها و روی محور عمودی (محور Yها)، فراوانی درج می‌گردد. سپس مستطیل‌هایی بافاصله و به ارتفاع فراوانی مطلق برای هر مقدار متغیر رسم می‌کنیم.

مثال٦) برای داده‌های جدول فراوانی زیر نمودار ستونی به صورت زیر خواهد بود.

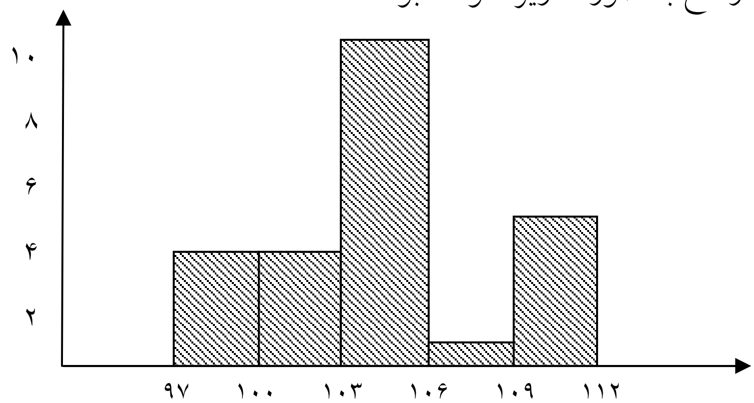
X	۰	۱	۲	۳	٤	۵
f_i	۲	۵	۷	۳	۲	۱



توجه: در رسم نمودار ستونی، اگر بجای مستطیل از میله استفاده کنیم، نمودار بدست آمده را نمودار میله‌ای می‌گویند.

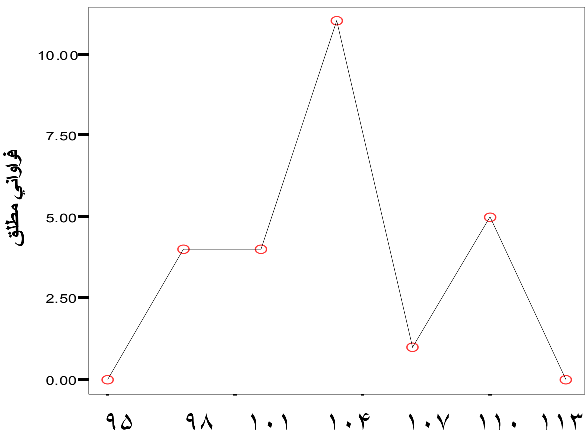
نمودار مستطیلی(هیستوگرام): این نمودار برای نمایش داده‌های کمی پیوسته(داده‌های طبقه‌بندی شده) به کار می‌رود.

برای رسم این نمودار، روی محور افقی مستطیل‌هایی بنا می‌کنیم به طوریکه قاعده هر مستطیل از حد پایین تا حد بالایی هر طبقه باشد و ارتفاع هر مستطیل برابر فراوانی مطلق همان طبقه باشد. بعنوان مثال هیستوگرام داده‌های پرش ارتفاع به صورت زیر خواهد بود.



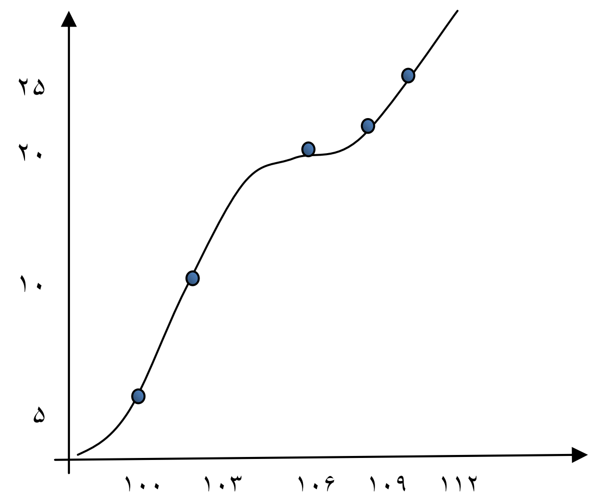
نمودار چند ضلعی(شکسته): در این نمودار روی محور افقی، نقطه میانی طبقات و روی محور عمودی، فراوانی طبقات قرار داده می‌شود. سپس نقاطی که ارتفاع آنها به اندازه فراوانی مطلق طبقه مربوطه باشد را مشخص کرده و آنها را به یکدیگر متصل می‌کنیم.

نمودار چند ضلعی برای داده‌های پرش ارتفاع به صورت زیر است.



برای رسم نمودار چند ضلعی شکسته اغلب پیشنهاد می‌گردد با فرض وجود یک طبقه قبل از طبقه اول و یک طبقه بعد از طبقه آخر نقاط میانی این طبقات فرضی را یافته و با مشخص نمودن آنها بر روی محور افقی و در نظر گرفتن فراوانی صفر برای هر دوی آنها ابتدا و انتهای نمودار را به این نقاط فرضی متصل نماییم، آنچنانکه در شکل بالا مشاهده می‌کنید.

نمودار فراوانی تجمعی: برای رسم این نمودار روی محور افقی، حد بالایی هر طبقه و روی محور عمودی، فراوانی تجمعی داده‌ها را قرار می‌دهیم. سپس نقاطی را که ارتفاع آنها به اندازه فراوانی تجمعی طبقه مربوطه باشد را مشخص کرده و آنها را به یکدیگر متطل می‌نماییم.بعنوان مثال نمودار فراوانی تجمعی برای داده‌های مثال پرش ارتفاع به صورت زیر است.



مسائل

۱- تعداد کتاب‌های ریاضی، آمار، زبان و فیزیک موجود در یک کتابخانه به ترتیب ۳۸، ۳۰، ۲۵ و ۱۷ کتاب است. نمودار دایره‌ای را برای این کتاب‌ها رسم کنید؟

۲- در یک نمونه ۵۰ تایی از جمعیت ساکنان یک منطقه تعداد فرزندان خانواده‌ها به صورت زیر به دست آمده است. یک جدول توزیع فراوانی کامل برای این داده‌ها بدست آوردید ؟

١ ١ ٠ ٣ ٤ ٦ ١ ٠ ٢ ٥ ٢ ٥ ٣ ٤ ٠ ١ ٢ ٤ ٥ ٣ ٣ ٤ ٤ ٢ ٦ ٥ ٢ ٣ ٣

٢ ١ ٣ ٢ ٠ ٢ ٢ ٣ ٤ ٦ ٥ ١ ٢ ٤ ٣ ٥ ٢ ٠ ١ ٢ ١

۳- وزن‌های ٤٠ کیسه گندم تا نزدیکترین کیلو در زیر داده شده است. یک جدول توزیع فراوانی کامل با تعداد ۸ طبقه، برای این داده‌ها به دست آورید؟

١٣٢ ١٤٤ ١٢٥ ١٤٩ ١٥٧ ١٤٦ ١٥٨ ١٤٠ ١٤٧ ١٣٦ ١٤٨ ١٥٠ ١٣٨ ١٦٤

١٥٢ ١٤٤ ١٦٨ ١٢٦ ١٣٨ ١٧٦ ١٦٣ ١١٩ ١٥٤ ١٦٥ ١٤٦ ١٧٣ ١٤٢

١٤٧ ١٣٥ ١٥٣ ١٤٠ ١٣٥ ١٦١ ١٣٥ ١٤٥ ١٤٢ ١٥٠ ١٥٦ ١٤٥ ١٢٨

فصل سوم

توصیف عددی مشاهدات

در فصل قبل دیدیم که با استفاده از جدول فراوانی و نمودارهای آماری، می‌توان تا حدودی دانسته‌های نهفته در داده‌ها را مختصر و محسوس کرد. با این حال سعی می‌شود تا این دانسته‌ها را به صورت چند شاخص معقول درآورد تا بتوان ایده‌ای کلی درباره ویژگی مورد مطالعه به دست آورد و هم نتیجه مطالعات را به سادگی گزارش داد. این شاخص‌ها را می‌توان به طور کلی به دو دسته تقسیم کرد.

الف) شاخص‌های گرایش به مرکز: شاخص‌هایی را که معمولاً در حوالی مرکز منحنی فراوانی می‌باشند و میزان گرایش به مرکز داده‌ها را تعیین می‌کنند. همانند: میانگین، نما(مُد)، میانه و چندک‌ها.

ب) شاخص‌های پراکندگی: شاخص‌هایی که میزان پراکندگی داده‌ها را از مرکز توزیع داده‌ها نمایش می‌دهند. همانند: دامنه تغییرات، انحراف میانگین، واریانس، انحراف معیار، ضریب تغییرات، ضریب چولگی و کشیدگی.

در این فصل نحوه محاسبه این شاخص‌ها را در داده‌های خام و داده‌های طبقه‌بندی شده به تفضیل مورد بحث قرار خواهیم داد. در این فصل ابتدا شاخص‌های گرایش به مرکز را مورد بررسی قرار خواهیم داد.

شاید مهمترین نکته در مطالعه توزیع یک نمونه از داده‌ها، تعیین یک مقدار مرکزی باشد، یعنی، یک مقدار نماینده که داده‌ها در اطراف آن توزیع شده‌اند. هر شاخص عددی را که معرف مرکز مجموعه‌ی داده‌ها باشد، شاخص گرایش به مرکز می‌نامند. متداولترین شاخص‌های گرایش به مرکز عبارتند از : میانگین، نما، میانه و چندک‌ها.

میانگین حسابی(میانگین): میانگین حسابی یا متوسط نمونه‌ای مرکب از n اندازه $X_1, X_2, ..., X_n$ عبارت است از خارج قسمت مجموع این اندازه ها بر n. میانگین را با $\bar{X}$ یا μ نشان می‌دهند که در عملیات، به صورت $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ نوشته می‌شود.

همانطور که از مفهوم "متوسط" برمی‌آید، میانگین، مرکز مجموعه داده‌ها را نمایش می‌دهد. اگر نمودار نقطه‌ای مجموعه داده‌ها را این طور تجسم کنیم که روی میله افقی نازکی، گویهای هم اندازه‌ایی در محل داده‌ها قرار دارند، آنگاه، میانگین نشان دهنده نقطه‌ای است که این میله در آن نقطه به حال تعادل در می‌آید.

مثال۱: قیمت ۵ کتاب یه صورت ۱۵۰۰۰، ۱۲۵۰۰، ۱۰۵۰۰، ۹۵۰۰ و ۸۳۲۰ تومان است. میانگین قیمت کتاب‌ها چقدر است؟

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{15000 + ... + 8320}{5} = 11/164 \text{ تومان}$$

میانگین وزنی: در برخی از داده ها برای محاسبه میانگین حسابی، به دلیل اینکه مقادیر مشاهده شده ارزش‌های(وزن‌های) متفاوت دارند از میانگین وزنی به صورت زیراستفاده می‌شود:

میانگین وزنی اعداد $x_۱,x_۲,...,x_n$ با وزن‌های $w_۱,w_۲,...,w_n$ را با $\bar{X}_w$ نشان داده و به صورت $\bar{X}_w = \frac{\sum_{i=۱}^{n} w_i x_i}{\sum_{i=۱}^{n} w_i}$ محاسبه می‌شود.

مثال۲: نمرات یک دانشجو در یک ترم در ۳ درس ریاضی، آمار و فیزیک به صورت جدول زیر داده شده است. معدل این دانشجو را به دست آورید؟

نمرات (x_i)	تعداد واحد (w_i)	
۱۳	۳	ریاضی
۱۰	٤	آمار
۱۵	۲	فیزیک

پاسخ:

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} = \frac{(۳)(۱۳)+(۴)(۱۰)+(۲)(۱۵)}{۳+۴+۲} = ۱۲/۱۱$$

میانگین هندسی: میانگین هندسی را با $\bar{X}_G$ نشان داده و برابر است با ریشه nاُم حاصلضرب آنها. یعنی: $\bar{X}_G = \sqrt[n]{x_۱x_۲...x_n}$.

مثال۳: میانگین هندسی دو عدد ۲ و ۸ برابر است با: $\bar{x}_G = \sqrt{۲\times۸} = ۴$.

در صورتی که $x_۱,x_۲,...,x_n$ همگی مثبت با وزن‌های $w_۱,w_۲,...,w_n$ باشند، میانگین هندسی به صورت $\bar{X}_G = \sqrt[n]{x_۱^{w_۱}x_۲^{w_۲}...x_n^{w_n}}$ محاسبه می‌شود.

اگر داده‌ها به صورت کسری باشند، بطوریکه واحد صورت و مخرج یکسان باشد از میانگین هندسی و اگر واحد صورت و مخرج یکسان نباشد از میانگین هارمونیک استفاده می‌کنیم.

معمولاً هرگاه x_iها از درصدها یا نسبت‌ها تشکیل شده باشند، مثلاً در کارهای اقتصادی یا جمعیت شناسی، میانگین هندسی به کار می‌برند.

مثال٤: میزان تولید برنج در ۳ سال متوالی به ترتیب ۳۰٪، ۲۸٪ و ٤۰٪ نسبت به سال پایه افزایش داشته است. نرخ متوسط رشد را محاسبه کنید؟

پاسخ: برای پاسخ به این سوال نیاز به دانستن میزان تولید در این ۳ سال، نسبت به سال پایه داریم. اگر سال پایه را برابر با یک قرار دهیم، مقدار تولید در این ۳ سال متوالی نسبت به سال پایه به ترتیب ۱/۳، ۱/۲۸ و ۱/٤ خواهد بود. حال نرخ متوسط رشد را با استفاده از میانگین هندسی به صورت زیر به دست خواهیم آورد: $\bar{x}_G = \sqrt[3]{(1/3)(1/28)(1/4)} \simeq 1/32$.

مثال۵: میزان تولیدات یک کارخانه بر حسب تن، در چند سال متوالی به صورت جدول زیر داده شده است. نرخ متوسط رشد تولیدات کارخانه را به دست آورید؟

سال	میزان تولید	میزان تولید نسبت به سال پایه
۱۳۸۰	۲۰	$\frac{20}{20}=1$
۱۳۸۱	۳۰	$\frac{30}{20}=1/5$
۱۳۸۲	٤۰	۲
۱۳۸۳	۵۰	۲/۵
۱۳۸٤	۲۰	۱
۱۳۸۵	٦۰	۳

پاسخ: در صورت مسئله فقط دو ستون اول جدول یعنی شماره سال و میزان تولید داده شده است.برای محاسبه متوسط نرخ رشد بایستی میزان تولید هر سال نسبت به سال پایه را حساب کنیم. با در نظر گرفتن سال ۱۳۸۰ به عنوان سال پایه داریم:

$$\bar{x}_G = \sqrt[5]{(1/5)(2)(2/5)(2)(3)} \simeq 2/14$$

میانگین هارمونیک(توافقی): میانگین هارمونیک را با $\bar{X}_H$ نشان داده و برابر است با تعداد داده‌ها بر مجموع معکوس داده‌ها در یک مجموعه داده. به عبارت ریاضی داریم: $\bar{X}_H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$.

در صورتیکه $x_1,x_2,...,x_n$ همگی غیر صفر با وزن‌های $w_1,w_2,...,w_n$ باشند. میانگین هارمونیک از رابطه $\bar{X}_H = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n}\frac{w_i}{x_i}}$ محاسبه می‌شود.

کاربرد این میانگین در موارد خاص می باشد. مثلاً برای محاسبه متوسط سرعت اتومبیل وقتی که اتومبیل فاصله بین دو شهر را با سرعت های متفاوت طی می کند، سرعت متوسط را نمی توان از میانگین حسابی

به دست آورد. فرض کنید راننده ای مسافتی را با سرعت ۸۰ کیلومتر در ساعت طی می کند و در برگشت، همان مسافت را با سرعت ۹۰ کیلومتر در ساعت، در این صورت سرعت متوسط راننده از میانگین هارمونیک برابر است با: $\bar{x}_H = \frac{۲}{\frac{۱}{۸۰}+\frac{۱}{۹۰}} = \frac{(۲)(۸۰)(۹۰)}{(۸۰)+(۹۰)} = ۸۴/۷$.

مثال٦: اتومبیلی مسیری را با سرعت ۱۰۰ کیلومتر رفته و $\frac{۱}{۳}$ مسیر را با سرعت ۸۰ کیلومتر و باقیمانده را با سرعت ۱۲۰ کیلومتر برگشته، سرعت متوسط این اتومبیل چقدر بوده است؟

پاسخ: از آنجا که سرعت اتومبیل در تکه‌های مختلف از مسیر متفاوت بوده است، می‌توان سرعت را به عنوان متغیر و قسمت مربوطه را به عنوان وزن آن به صورت جدول زیر در نظر گرفت:

مسیر مربوطه w_i	سرعت X_i
۱	۱۰۰
$\frac{۱}{۳}$	۸۰
$\frac{۲}{۳}$	۱۲۰

در نتیجه: $\bar{x}_H = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{۱+\frac{۱}{۳}+\frac{۲}{۳}}{\frac{۱}{۱۰۰}+\frac{۱/۳}{۸۰}+\frac{۲/۳}{۱۲۰}} = ۱۰۱/۴$.

میانگین مرتبه دو: این میانگین به صورت $M_۲ = \left(\sum_{i=۱}^{n} \frac{w_i x_i^۲}{n}\right)^{\frac{۱}{۲}}$ تعریف می‌شود و در حقیقت برابر است با جذر میانگین حسابی اعداد $x_۱^۲, x_۲^۲, \ldots, x_n^۲$.

می‌توان ثابت کرد که برای داده‌های مثبت میان این چهار نوع میانگین، نامساوی زیر برقرار است:

$$\boxed{\bar{X}_H \le \bar{X}_G \le \bar{x} \le M_۲}$$

محاسبه میانگین در جدول توزیع فراوانی:

فرض کنید f_i فراوانی طبقه iام و X'_i نقطه میانی طبقه iام باشد، در اینصورت میانگین از رابطه $\bar{X} = \frac{\sum_{i=۱}^{n} f_i X'_i}{n}$ به دست خواهد آمد.

مثال۷: در جدول توزیع فراوانی زیر میانگین حسابی را به دست آورید؟

حدود طبقات	f_i	X_i'	f_iX_i'
۳٤/۵ – ٤۱/۵	۱	۳۸	۳۸
٤۱/۵ – ٤۸/۵	۲	٤۵	۹۰
٤۸/۵ – ۵۵/۵	۲	۵۲	۱۰٤
۵۵/۵ – ٦۲/۵	۳	۵۹	۱۷۷
٦۲/۵ – ٦۹/۵	۲	٦٦	۱۳۲
	$n=\sum f_i = ۱۰$		$\sum f_iX_i' = ۵۴۱$

$$\bar{X} = \frac{\sum f_i X_i'}{n} = \frac{۵۴۱}{۱۰} = ۵۴/۱$$

میانه: میانه عددی است که تقریباً ۵۰ درصد داده‌ها از آن کوچکتر و ۵۰ درصد داده‌ها از آن بزرگتر هستند. در واقع میانه، مقادیر داده‌ها را به دو نیمه مساوی در دو طرف خود تقسیم می‌کند و آن را با Me یا Mn نشان می‌دهند.

الف) محاسبه میانه برای داده‌های طبقه‌بندی نشده:

فرض کنید n داده داشته باشیم که به صورت غیر نزولی یعنی به صورت $X_۱ \le X_۲ \le \dots \le X_n$ مرتب شده باشند. در اینصورت، با توجه به تعداد داه‌ها میانه را از یکی از روابط زیر محاسبه میکنیم.

$$Mn = \begin{cases} X_{(\frac{n+۱}{۲})} & \text{اگر } n \text{ فرد باشد} \\ \dfrac{X_{(\frac{n}{۲})} + X_{(\frac{n+۲}{۲})}}{۲} & \text{اگر } n \text{ زوج باشد} \end{cases}$$

مثال۸: میانه داده‌های ۱۷، ۸، ۹، ۵، ٤، ۱۲ و ۲۰ را به دست آورید؟

حل: ابتدا داده‌ها را به صورت غیر نزولی مرتب خواهیم کرد: ۲۰، ۱۷، ۱۲، ۹، ۸، ۵، ٤

با توجه به اینکه $n=۷$ یعنی تعداد داده ها فرد است بنابراین میانه را از رابطه $Mn = X_{(\frac{n+۱}{۲})}$ محاسبه خواهیم کرد. در نتیجه داریم $Mn = X_{(\frac{n+۱}{۲})} = X_۴ = ۹$.

مثال۹: میانه داده‌های ۵، ٤، ۳، ۳، ۲، ۲، ۲ و ۱ را به دست آورید؟

حل: در این مسئله داده ها خود به صورت غیر نزولی در صورت مسله داده شده‌اند. با توجه به اینکه تعدا داده‌ها $n=۸$ زوج است برای محاسبه میانه داریم. $Mn = \dfrac{X_{(\frac{n}{۲})} + X_{(\frac{n+۲}{۲})}}{۲} = \dfrac{x_۴ + x_۵}{۲} = \dfrac{۲+۳}{۲} = ۲/۵$

ب) محاسبه میانه در جدول توزیع فراوانی: برای محاسبه میانه در جدول توزیع فراوانی مراحل زیر را به ترتیب انجام می‌دهیم.

۱- $\frac{n}{۲}$ را محاسبه می‌کنیم.

۲- در ستون فراوانی تجمعی اولین عدد بزرگتر یا مساوی $\frac{n}{۲}$ را یافته و طبقه مربوط به آن را طبقه میانه (i) می‌نامیم.

۳- میانه را از رابطه $Mn = L_i + \frac{(\frac{n}{۲} - F_{i-۱})w}{f_i}$ محاسبه می‌کنیم که در آن L_i حد پایین طبقه میانه، $F_{i-۱}$ فراوانی تجمعی طبقه ماقبل طبقه میانه، f_i فراوانی مطلق طبقه میانه و w طول طبقه میانه است.

تذکر: در جدول توزیع فراوانی طول تمامی طبقات با هم برابر است و آن را با کسر حد بالای یکی از طبقات از حد پایین همان طبقه می‌توان به دست آورد. به عبارت ریاضی داریم: $w = H_i - L_i$

مثال۱۰: با توجه به داده‌های جدول مقابل مقدار میانه (R) را محاسبه کنید؟

حدود طبقات	f_i	F_i
۱-۵	۲	۲
۵-۹	۳	۵
۹-۱۳	۳	۸
۱۳-۱۷	۲	۱۰

فراوانی تجمعی طبقه ماقبل طبقه میانه

طبقه میانه

اولین عدد بزرگتر یا مساوی $\frac{n}{۲} = ۵$

۱) $\frac{n}{۲} = \frac{۱۰}{۲} = ۵$

۲) در ستون فراوانی تجمعی اولین عدد بزرگتر یا مساوی ۵ خود عدد ۵ است که متعلق به طبقه دوم (۵-۹) یعنی $i = ۲$ است. همچنین داریم $w = ۵ - ۱ = ۴$.

۳) اکنون می‌توان به راحتی میانه را به صورت زیر محاسبه کرد:

$$Mn = L_۲ + \frac{(\frac{n}{۲} - F_۱) \times w}{f_۲} = ۵ + \frac{(۵ - ۲) \times ۴}{۳} = ۹$$

مثال۱۱: در جدول مقابل میانه را محاسبه کنید؟

حدود طبقات	f_i	F_i
۳–۷	۱۰	۱۰
۷–۱۱	۳۰	٤۰
۱۱–۱۵	۲۰	٦۰

$$\frac{n}{۲}=\frac{۶۰}{۲}=۳۰$$

۱- در ستون فراوانی تجمعی اولین عدد بزرگتر یا مساوی ۳۰ عدد٤۰ است که متعلق به طبقه دوم (۷–۱۱) یعنی $i=۲$ است. همچنین داریم $w=۷-۳=۴$.

۲- $Mn=L_۲+\dfrac{(\frac{n}{۲}-F_۱)\times w}{f_۲}=۷+\dfrac{(۳۰-۱۰)\times ۴}{۳۰}=۹/۶$.

موارد استفاده از میانه و میانگین :

وجود معدودی مشاهده خیلی بزرگ یا خیلی کوچک ، در میانه تاثیر ندارد، در حالی که وجود اینگونه مقادیر فرین در میانگین اثر قابل ملاحظه‌ای دارد. به نظر می‌رسد برای توزیع‌هایی که خیلی نامتقارن هستند، میانه معیار معقولتری از گرایش به مرکز است تا میانگین. به این دلیل در گزارشهای دولتی راجع به توزیع درآمد، به جای میانگین ، میانه درآمدها را ذکر می‌کنند. وقتی توزیع خیلی نامتقارن نیست، میانگین به میانه ترجیح داده می‌شود و خیلی بیشتر از میانه بکار می‌رود، زیرا در روشهای استنباطی ، میانگین از لحاظ نظری دارای امتیازاتی است که میانه فاقد آنهاست.

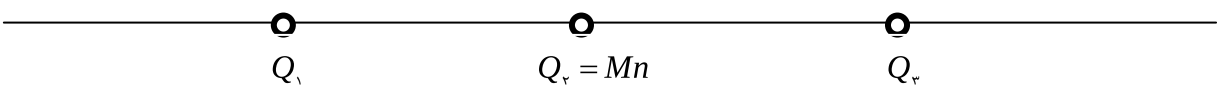
چندک‌ها: عدد X_p را که در آن $۰<p<۱$است، چندک مرتبه p می‌نامند هرگاه تقریباً $۱۰۰p$درصد داده‌ها کوچکتر از آن باشند. مثلاً $X_{۰/۱۵}$ را چندک مرتبه ۰/۱۵ می‌نامند هرگاه تقریباً ۱۵درصد داده‌ها کوچکتر از آن باشند.

توجه: چندک‌ها کلی‌تر از میانه هستند، $X_{۰/۵۰}$ در واقع همان میانه است.

چندک‌های معروف:

الف) چارک‌ها: که به ازای $p=۰/۲۵,۰/۵۰,۰/۷۵$ به دست می‌آیند و داده ها را به چهار بخش مساوی تقسیم میکنند و به ترتیب آنها را $Q_۱,Q_۲,Q_۳$نمایش می‌دهند.

$Q_۱,Q_۳$را به ترتیب چارک اول و سوم و $Q_۲$را میانه می‌نامند.



ب) *دهک‌ها:* که به ازای $p=۰/۱,۰/۲,...,۰/۹$ به دست می‌آیند که داده‌ها با به ۱۰ قسمت مساوی تقسیم کرده و به ترتیب آنها را با $D_۱,D_۲,....,D_۹$ نشان می‌دهند.

ج) *صدک‌ها:* که به ازای $p=۰/۰۱,۰/۰۲,...,۰/۹۹$ به دست می‌آیند و داده‌ها را به ۱۰۰قسمت مساوی تقسیم کرده و به ترتیب آنها را با $P_۱,P_۲,...,P_{۹۹}$ نمایش می‌دهند.

محاسبه چندک‌ها برای داده‌های طبقه بندی نشده:

فرض کنید n داده داشته باشیم که به ترتیب غیرنزولی $X_۱ \le X_۲ \le ... \le X_n$ مرتب شده باشند، برای به دست آوردن X_p بدین ترتیب عمل خواهیم کرد. حاصل عبارت $(n+۱)p$ را به دست می‌آوریم اگر حاصل مساوی عدد صحیح r باشد، فرض می‌کنیم $X_p=X_r$. در غیر اینصورت $[(n+۱)p]=r$ قرار داده و اختلاف ان با $(n+۱)p$ را برابر w می‌گیریم. اکنون X_p را از رابطه $X_p=(۱-w)X_r+wX_{r+۱}$ محاسبه خواهیم کرد.

مثال۱۲: برای داده‌های زیر $X_{۰/۶۰}$ را محاسبه کنید؟

۰ ۱ ۲ ۳ ۴ ۴ ۵ ۷ ۷ ۸ ۱۲

پاسخ:

$$n=۱۱ \& p=۰/۶ \Rightarrow (n+۱)p=۱۲\times ۰/۶۰=۷/۲ \notin Z$$
$$[(n+۱)p]=۷=r \quad \& \quad ۷/۲-۷=۰/۲=w$$
$$\Rightarrow X_{۰/۶۰}=(۱-۰/۲)X_۷+۰/۲X_۸=۰/۸\times ۵+۰/۲\times ۷=۵/۴$$

ب) *محاسبه چندک برای داده‌های جدول توزیع فراوانی:*

برای محاسبه چندک در جدول توزیع فراوانی مراحل زیر را به ترتیب انجام دهید.

۱- np را محاسبه می‌کنیم.

۲- در ستون فراوانی تجمعی اولین عدد بزرگتر یا مساوی np را یافته و آن را طبقه iام می‌نامیم.

۳- مقدار چندک را از رابطه $X_p=L_i+\frac{(np-F_{i-۱})w}{f_i}$ محاسبه می‌کنیم.

مثال۱۳: در جدول فراوانی زیر $X_{۰/۱۵}$، $X_{۰/۲۰}$ و چارک سوم را به دست آورید؟

حدود طبقات	f_i	F_i
۸ – ۱۰	۵	۵
۱۰ – ۱۲	۵	۱۰
۱۲ – ۱۴	۸	۱۸
۱۴ – ۱۶	۱۲	۳۰

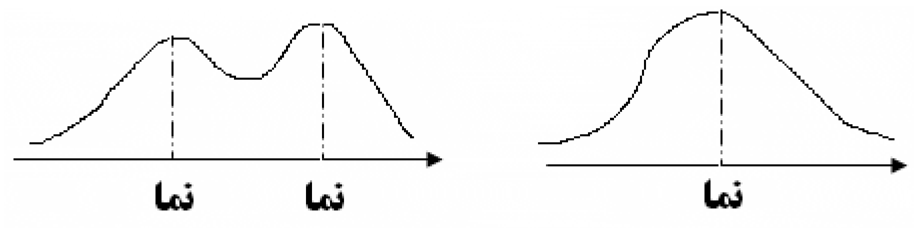
$$X_{\cdot/۱۵}\begin{cases} NP=۳۰\times ۰/۱۵=۴/۵ \\ (۸-۱۰) \qquad i=۱ \end{cases} \Rightarrow X_{\cdot/۱۵}=۸+\frac{(۴/۵-۰)\times ۲}{۵}=۸/۹$$

$$X_{\cdot/۲۰}\begin{cases} NP=۳۰\times ۰/۲۰=۶ \\ (۱۰-۱۲) \qquad i=۲ \end{cases} \Rightarrow X_{\cdot/۲۰}=۱۰+\frac{(۶-۵)\times ۲}{۵}=۱۰/۴$$

$$Q_۳=X_{\cdot/۷۵}\begin{cases} NP=۳۰\times ۰/۷۵=۲۵/۵ \\ (۱۴/۱۶) \qquad i=۴ \end{cases} \Rightarrow Q_۳=۱۴+\frac{(۲۵/۵-۱۸)\times ۲}{۱۲}=۱۵/۲۵$$

نما:

نما ترجمه کلمه مد است، که یک لغت فرانسوی و به معنای "متداول‌ترین" است. نما، برای مجموعه‌ای از داده‌ها عبارتست از اندازه‌ای که بیشترین فراوانی را دارا می‌باشد. برخلاف میانگین و میانه که برای مجموعه‌ای از داده‌ها وجود داشته و یکتا است، نما، لزوماً چنین خاصیتی را ندارد. اگر فراوانی داده‌ها یکسان باشد، توزیع آنها نما ندارد. به عبارت دیگر داده ها بدون نما هستند. اگر دو اندازه از داده‌ها فراوانی یکسان و بیشترین فراوانی را داشته باشند توزع آنها دو نمائی است. به همین ترتیب ممکن است توزیع چند نمائی برای مجموعه‌ای از داده‌ها داشته باشیم. نما را با حرف Mo نمایش می دهیم.شکل منحنی‌های یک نمایی و دو نمایی در زیر آمده است.



محاسبه نما برای داده های خام:

۱- پیدا کردن فراوانی داده ها.
۲- داده‌ای که فراوانی آن بیشتر باشد را به عنوان نما انتخاب می کنیم.

توضیح: اگر دو داده دارای فراوانی مساوی باشند، بیشتر از سایر فراوانی‌ها، هر دو را به عنوان نما انتخاب می‌کنیم. مشروط بر اینکه این دو داده کنار هم نباشند. اگر کنار هم بودند، نصف مجموع آنها را نما می‌خوانیم.

مثال۱٤: برای داده‌های ۳،۱،۳،۱،۳،۱،۳،۲،٤،۵،۳،۳،۱ نما برابر است با $Mo=۳$.

مثال۱۵: برای داده‌های ۱،۱،۳،۲،۱،٤،۳،۳،۵ دو داده ۱ و ۳ که کنار هم نیستند و فراوانی مشترک آنها بیش از سایر فراوانی‌هاست، هر دو به عنوان نما اختیار می‌شوند.

مثال۱٦: برای داده‌های۲،۵،٤،۲،۲،۳،۱،۲،۱،۱،۳،۱ نصف دو داده ۱ و ۲ که کنار هم هستند و دارای فراوانی ٤ (بیش از سایر فراوانی‌ها) هستند به عنوان نما اختیار می‌شود. یعنی $Mo = \frac{۱+۲}{۲} = ۱/۵$.

محاسبه نما برای داده های طبقه بندی شده:

۱- خلاصه کردن داده‌ها در یک جدول فراوانی.

۲- نماینده رده‌ای را که دارای بیشترین فراوانی می‌باشد و رده نمایی نامیده می‌شود، را به عنوان نما اختیار می کنیم.

۳- برای دقت بیشتر می توان نما را از فرمول $Mo = L_i + (\frac{d_۱}{d_۱ + d_۲})w$ بدست آورد. در این فرمول L_i حد پایین طبقه‌ای که فراوانی مطلق آن بیشترین است، $d_۱$ اختلاف فراوانی‌های مطلق طبقه نمادار و طبقه بلافاصله قبل از آن، $d_۲$ اختلاف فراوانی‌های مطلق طبقه نمادار و طبقه بلافاصله بعد از آن و w طول طبقه است.

مثال۱۷: در جدول توزیع فراوانی زیر نما را به دست آورید؟

f_i	حدود طبقات
۱۰	۳ − ۷
۳۰	۱۱ − ۷
۲۰	۱۱ − ۱۵

پاسخ:

$$\begin{cases} L_i = ۷ \\ w = ۷ - ۳ = ۴ \\ d_۱ = f_۲ - f_۱ = ۳۰ - ۱۰ = ۲۰ \\ d_۲ = f_۲ - f_۳ = ۳۰ - ۲۰ = ۱۰ \end{cases} \Rightarrow Mo = ۷ + \left(\frac{۲۰}{۲۰+۱۰}\right) \times ۴ = ۹/۳$$

مقایسه اندازه های مرکزی میانگین، میانه و نما:

برای مجموعه‌ای از داده‌ها نمی‌توان به سادگی نتیجه گرفت که کدامیک از اندازه‌های مرکزی میانگین، میانه و نما بهترین معیار است. لذا در این قسمت به ویژگی‌ها و موارد استفاده آنها اشاره می‌شود تا استفاده کننده با شناخت بهتری بتواند معیار مناسب را انتخاب نماید.

الف) میانگین حسابی

۱- میانگین با استفاده از ارزش همه داده‌ها محاسبه می‌گردد.
۲- مقدار میانگین نسبت به دو معیار دیگر، در نمونه‌گیری‌های متفاوت از یک جمعیت، کمتر تغییر می‌یابد.
۳- از میانگین برای محاسبه معیارهای پراکندگی استفاده می‌شود.
٤- میانگین را نمی‌توان برای یک جدول توزیع فراوانی که طبقه اول و طبقه آخر آن محدود نمی‌باشند محاسبه کرد.
۵- میانگین برای مجموعه‌ای از داده‌ها یکتا است.
٦- اندازه میانگین تحت تأثیر مقادیر بسیار بزرگ و یا مقادیر بسیار کوچک قرار می‌گیرد و به همین دلیل می‌تواند معیار مرکزی نامناسبی باشد.

ب) میانه

۱- میانه زمانی محاسبه می‌گردد که نیاز به شناخت ارزش میانی داده‌ها باشد.
۲- برای تشخیص اینکه اندازه‌ای از داده‌ها در نیمه بالا و یا نیمه پایین توزیع قرار می‌گیرد، محاسبه میانه لازم است.
۳- میانه معیار مرکزی مناسبی برای داده‌های جدول توزیع فراوانی است که طبقه اول و طبقه آخر آن محدود نیست.
٤- میانه کمتر تحت تأثیر مقادیر بسیار کوچک و بسیار بزرگ قرار می‌گیرد.

ج) نما

۱- برای تعیین متداول‌ترین اندازه داده‌ها از معیار نما استفاده می‌شود.
۲- محاسبه معیار مرکزی نما از سایر معیارها ساده‌تر است.
۳- نما را می‌توان به عنوان یک معیار مرکزی برای داده‌های کیفی نیز به کار برد.
٤- ممکن است برای مجموعه‌ای از داده‌ها نما وجود نداشته باشد و یا بیش از یک نما موجود باشد.

این بخش را با بیان چند نوع میانگین به پایان خواهیم برد.

همانگونه که پیش از این ذکر شد اندازه میانگین تحت تأثیر مقادیر بسیار بزرگ و یا مقادیر بسیار کوچک قرار می‌گیرد و به همین دلیل می‌تواند معیار مرکزی نامناسبی باشد در چنین شرایطی شاید مناسب باشد از معیار میانگین‌های دیگری که از تاثیر مقادیر فرین مصون می‌باشند استفاده نماییم. در ادامه به تعدادی از آنان اشاره خواهد شد.

میانگین پیراسته: اگر داده‌های کوچکتر از مقدار چارک اول و بزرگتر از مقدار چارک سوم را از مجموعه داده‌ها حذف کنیم و میانگین داده‌های باقیمانده را به دست آوریم آن را میانگین پیراسته می‌گوییم.

میانگین وینزوری: اگر بجای هریک از داده‌های کوچکتر از چارک اول، مقدار چارک اول و بجای هر یک از داده‌های بزرگتر از چارک سوم، مقدار چارک سوم را قرار دهیم و سپس میانگین داده‌های جدید را به دست آوریم به آن میانگین وینزوری می‌گوییم.

میانگین اصلاح شده: روش کاملتری برای حذف اثر سوء داده‌های پرت، این است که p درصد ابتدا و p درصد انتهای داده‌ها را حذف کنیم و میانگین مابقی داده‌ها را به دست آوریم. این میانگین به میانگین اصلاح شده p درصد معروف است.

اگر تعداد نمونه‌ها کم و مقدار دورافتاده قابل تشخیص باشند، میانگین اصلاح شده را از رابطه $T_k = \frac{۱}{n-۲k}\sum_{i=k+۱}^{n-k} X_i$ به دست می‌آوریم و به آن میانگین اصلاح شده مرتبه k گویند. در این رابطه k تعداد داده‌هایی است که دور افتاده‌اند و بایستی از مجموعه حذف شوند. البته باید داده‌ها مرتب شده و $k < \frac{n}{۲}$ باشد.

توجه:میانگین پیراسته یک میانگین اصلاح شده ۲۵ درصد است.

مثال۱۸: میانگین پیراسته و میانگین وینزوری را برای مجموعه داده‌های زیر به دست آورید؟

۱ ۱۲ ۱۴ ۱۵ ۱۷ ۲۰ ۲۱ ۲۵ ۲۸ ۳۰

حل: ابتدا چارک اول و سوم را محاسبه می‌کنیم

$$Q_۱\begin{cases} n=۱۰ \,\&\, p=۰/۲۵ \Rightarrow (n+۱)p = ۱۱\times ۰/۲۵ = ۲/۷۵ \notin Z \\ \left[(n+۱)p\right] = ۲ = r \quad \& \quad ۲/۷۵-۲=۰/۷۵=w \\ \Rightarrow Q_۱ = (۱-۰/۷۵)X_۲ + ۰/۷۵X_۳ = ۰/۲۵\times ۱۲ + ۰/۷۵\times ۱۴ = ۱۳/۵ \end{cases}$$

$$Q_۳\begin{cases} n=۱۰ \,\&\, p=۰/۷۵ \Rightarrow (n+۱)p = ۱۱\times ۰/۷۵ = ۸/۲۵ \notin Z \\ \left[(n+۱)p\right] = ۸ = r \quad \& \quad ۸/۲۵-۸=۰/۲۵=w \\ \Rightarrow Q_۱ = (۱-۰/۲۵)X_۸ + ۰/۲۵X_۹ = ۰/۷۵\times ۲۵ + ./۲۵\times ۲۸ = ۲۵/۷۵ \end{cases}$$

برای محاسبه میانگین پیراسته، اعداد ۱ و ۱۲ را کمتر از مقدار چارک اول و اعداد ۲۸ و ۳۰ را که بیشتر از مقدار چارک سوم هستند را حذف کرده و سپس میانگین مابقی اعداد را به صورت زیر محاسبه می‌کنیم.

$$\text{میانگین پیراسته} = \frac{(۱۴+۱۵+۱۷+۲۰+۲۱+۲۵)}{۶} = ۱۸/۶۶$$

برای محاسبه میانگین وینزوری، به جای مقادیر ۱و ۱۲ مقدار چارک اول، یعنی ۱۳/۵ و به جای مقادیر ۲۸ و ۳۰ مقدار چارک سوم، یعنی ۲۵/۷۵ را قرار می‌دهیم و سپس میانگین این اعداد را به عنوان میانگین وینزوری به صورت روبرو محاسبه خواهیم کرد.

$$\text{میانگین وینزوری} = \frac{(۱۳/۵+۱۳/۵+۱۴+۱۵+۱۷+۲۰+۲۱+۲۵+۲۵/۷۵+۲۵/۷۵)}{۱۰} = ۱۹/۰۵$$

اندازه های پراکندگی

شاخص‌های مرکزی مانند میانگین، میانه و نما توصیف کننده وضعیت کامل توزیع داده‌ها نیستند. به عبارت دیگر دو مجموعه داده که دارای میانگین‌های یکسان هستند ممکن است پراکندگی متفاوتی داشته باشند و زمانی می‌توان توزیع داده‌ها را دقیقاً توصیف نمود که علاوه بر شناخت شاخص مناسبی برای مرکزیت آن، شاخصی را هم برای پراکندگی آنها تعیین نمود. معیارهای پراکندگی متداول عبارتند از: برد (دامنه) داده‌ها، برد چارک‌ها، نصف دامنه تغییرات، واریانس و انحراف معیار، ضریب تغییرات، ضریب چولگی و کشیدگی.

برد (دامنه) داده‌ها:

ساده‌ترین شاخص پراکندگی برای یک مجموعه داده، برد آنها است که آن را با R نشان داده و به صورت $R = Max(X_i) - Min(X_i)$ تعریف می‌شود که $Max(X_i)$، بزرگترین داده و $Min(X_i)$، کوچکترین داده مشاهده شده است. برد نیز همانند میانگین تحت تأثیر داده‌های پرت قرار می‌گیرد و در چنین حالاتی یک معیار مناسب پراکندگی نیست. به علاوه، چون برای محاسبه برد فقط از دو اندازه بزرگترین مشاهده و کوچکترین مشاهده استفاده می شود معمولاً معیار رضایت بخشی برای پراکندگی به حساب نمی‌آید.

مثال۱۹: برد داده های ٤،٤،٥،٧،٧،٨،١٢،٢،٣ عبارت است از: ۱۰=۲−۱۲.

برد میان چارکی:

این معیار در حقیقت برد پیراسته داده‌ها است و از داده‌های پرت تاثیر نمی‌گیرد . معمولا این معیار را با IQR نمایش می‌دهند که عبارتند از : $IQR = Q_۳ - Q_۱$. که در آن $Q_۱$ و $Q_۳$ به ترتیب چارکهای اول و سوم هستند. برد میان چارکی طول بازه‌ای است که پنجاه درصد داده‌ها در آن قرار دارند. برخی پژوهشگران از نیم برد میان چارکی استفاده می‌کنند که به صورت $\frac{Q_۳ - Q_۱}{۲}$ تعریف شده و انحراف چارکی نیز نامیده می‌شود. یک اندازه پراکندگی نسبی، که ضریب تغییر چارکی نام دارد به صورت نسبت نیم‌برد میان چارکی بر میانگین $Q_۱$ و $Q_۳$ ضرب در ۱۰۰، یعنی به صورت $۱۰۰ \times \frac{Q_۳ - Q_۱}{Q_۳ + Q_۱}$، تعریف می‌شود.

انحراف متوسط(میانگین انحراف): میانگین قدرمطلق تفاوت هر یک از داده‌ها از میانگین اعداد را انحراف متوسط می‌گویند و آن را با MD نمایش می‌دهند.

انحراف متوسط با استفاده از روابط زیر محاسبه میگردد.

انحراف متوسط برای داده های خام	$MD = \frac{\sum_{i=1}^{n} \lvert X_i - \bar{X} \rvert}{n}$
انحراف متوسط برای جدول فراوانی دارای رده	$MD = \frac{\sum_{i=1}^{n} f_i \lvert X_i - \bar{X} \rvert}{n}$
انحراف متوسط برای جدول فراوانی دارای طبقه با حدود	$MD = \frac{\sum_{i=1}^{n} f_i \lvert X'_i - \bar{X} \rvert}{n}$

مثال۲۰: برای داده‌های ۹، ۵، ۱، ٤، ۳ و ۸ مقدار انحراف متوسط را به دست آورید؟

حل: در حل اینگونه مسائل بهترین راه برای پرهیز از اشتباهات محاسباتی استفاده از جدولی به مانند روبرو است.

X_i	$X_i - \bar{X}$	$\lvert X_i - \bar{X} \rvert$
۹	٤	٤
۵	۰	۰
۱	٤-	٤
٤	۱-	۱
۳	۲-	۲
۸	۳	۳
جمع		١٤

$$\bar{x} = \frac{\sum x_i}{n} = \frac{۳۰}{۶} = ۵$$

$$MD = \frac{\sum \lvert x_i - \bar{x} \rvert}{n} = \frac{۱۴}{۶} = ۲/۲$$

مثال۲۱: برای جدول فراوانی زیر مقدار انحراف متوسط را به دست آورید؟

X_i	f_i	f_iX_i	$X_i-\bar{X}$	$\lvert X_i-\bar{X}\rvert$	$f_i\lvert X_i-\bar{X}\rvert$
۱	۱	۱	۴-	۴	۴
۴	۱	۴	۱-	۱	۱
۵	۶	۳۰	۰	۰	۰
۶	۱	۶	۱	۱	۱
۹	۱	۹	۴	۴	۴
جمع	$n=$۱۰	۵۰			۱۰

جواب

$$\bar{x}=\frac{\sum f_i x_i}{n}=۵$$

$$MD=\frac{\sum f_i\lvert x_i-\bar{x}\rvert}{n}=۱$$

مثال۲۲: برای داده‌های جدول زیر مقدار انحراف متوسط را به دست آورید؟

X_i	f_i	X_i'	f_iX_i'	$X_i'-\bar{X}$	$\lvert X_i'-\bar{X}\rvert$	$f_i\lvert X_i'-\bar{X}\rvert$
۴ − ۰	۸	۲	۱۶	۸-	۸	۶۴
۸ − ۴	۱۲	۶	۷۲	۴-	۴	۴۸
۱۲ − ۸	۲۰	۱۰	۲۰۰	۰	۰	۰
۱۶ − ۱۲	۱۲	۱۴	۱۶۸	۴	۴	۴۸
۲۰ − ۱۶	۸	۱۸	۱۴۴	۸	۸	۶۴
جمع	۶۰		۶۰۰			۲۲۴

جواب

$$\bar{x}=\frac{\sum f_i x_i'}{n}=۱۰$$

$$MD=\frac{\sum f_i\lvert x_i'-\bar{x}\rvert}{n}=۳/۷۳$$

واریانس و انحراف معیار

مفیدترین اندازه پراکندگی واریانس و یا جذر آن، انحراف معیار داده‌ها، است. اندازه انحراف معیار به ما می‌گوید که مشاهدات تا چه مقدار در اطراف میانگین آنها قرار دارند، یک اندازه کم برای انحرف معیارِ مجموعه‌ای از داده‌ها، نشان دهنده این واقعیت است که داده‌ها در دامنه کوچکی حول میانگین پراکنده شده‌اند و بالعکس انحراف معیار بزرگ، بیان کننده دامنه گسترده‌تری است که داده‌ها حول میانگین پراکنده گردیده‌اند. انحراف معیار ریشه دوم مثبت واریانس است که برای جامعه آن را با علامت σ و برای نمونه آن را با علامت S نشان می‌دهند. به عبارت دیگر علامت واریانس در جامعه $\sigma^۲$ و در نمونه $S^۲$ است. برای محاسبهٔ واریانس با استفاده از داده‌های خام از روابط نیز استفاده می شود.

واریانس جمعیت با اندازه N	$\sigma^۲ = \dfrac{\sum_{i=۱}^{N}(X_i-\mu)^۲}{N}$
واریانس نمونه با اندازه n	$S^۲ = \dfrac{\sum_{i=۱}^{n}(X_i-\bar{X})^۲}{n-۱}$

دلیل تقسیم مجموع مربعات انحراف از میانگین در نمونه بر $(n-۱)$، بر اساس خواص برآورد کردن واریانس جمعیت به وسیله واریانس نمونه است.

مثال۲۳: برای داده‌های ۹، ۵، ۷، ۳ و ۱ مقدار واریانس را به دست آورید؟

$$\bar{x} = \frac{\sum x_i}{n} = ۵$$

$$S^۲ = \frac{\sum_{i=۱}^{n}(X_i-\bar{X})^۲}{n-۱} = \frac{(۱-۵)^۲+\ldots(۹-۵)^۲}{۵-۱} = ۱۰$$

محاسبه واریانس در جدول توزیع فراوانی:

برای جدواول توزیع فراوانی واریانس را از رابطه $S^۲ = \dfrac{\sum_{i=۱}^{n} f_i(X_i-\bar{X})^۲}{n-۱}$ و اگر طبقات دارای حدود باشند در رابطه بالا به جای X_i از نماینده طبقات یعنی X'_i استفاده خواهیم کرد. در اینصورت داریم: $S^۲ = \dfrac{\sum_{i=۱}^{n} f_i(X'_i-\bar{X})^۲}{n-۱}$.

مثال۲٤: برای داده‌های جدول مقابل، مقدار واریانس را به دست آورید؟

X_i	f_i	X_i'	$f_i X_i'$	$X_i'-\bar{X}$	$\|X_i'-\bar{X}\|$	$f_i\|X_i'-\bar{X}\|$
۱ - ۳	٤	۲	۸	۲/۲	٤/۸٤	۱۹/۳٦
۳ - ۵	۳	٤	۱۲	۰/۲	۰/۰٤	۰/۱۲
۵ - ۷	۱	٦	٦	۱/۸	۳/۲٤	۳/۲٤
۷ - ۹	۲	۸	۱٦	۳/۲	۱۰/۲٤	۲۰/٤۸
جمع	۱۰		٤۲			٤۳/۲

پاسخ:

$$\bar{x}=\frac{\sum f_i x_i'}{n}=۴/۲$$

$$S^{۲}=\frac{\sum f_i (x_i'-\bar{x})^{۲}}{n-۱}=۴/۸$$

همانطور که قبلا نیز بیان کردیم به جذر واریانس انحراف استاندارد می‌گویند و برای نمونه آن را با S نمایش می‌دهند. بعبارت ریاضی داریم که: $S=\sqrt{S^{۲}}$

به عنوان نمونه در مثال بالا انحراف استاندارد برابر است با $S=\sqrt{۴/۸}=۲/۱۹$.

خواص واریانس:

۱- اگر تمامی داده‌ها با هم برابر باشند، واریانس داده‌ها صفر است و بالعکس.

$$x_{۱}=x_{۲}=\ldots=x_n \Leftrightarrow S_x^{۲}=۰$$

۲- واریانس داده‌ها مقداری مثبت است.$S_x^{۲}\geq ۰$

۳- اگر به داده‌ها عدد ثابتی را اضافه یا کم کنیم، واریانس تغییری نمی‌کند.

$$Y_i=X_i \pm a \quad \Rightarrow S_y^{۲}=S_x^{۲}$$

٤- اگر تمامی داده‌ها در عدد ثابتی مانند a ضرب (تقسیم) شود آنگاه واریانس در مجذور a ضرب (تقسیم) خواهد شد.

$$Y_i=aX_i \quad \Rightarrow S_y^{۲}=a^{۲}S_x^{۲}$$

$$Y_i=\frac{X_i}{a} \quad \Rightarrow S_y^{۲}=\frac{S_x^{۲}}{a^{۲}}$$

ضریب تغییرات:

برای مقایسه پراکندگی در دو جمعیت و یا دو نمونه اگر صفت‌های مورد بررسی دارای واحد اندازه‌گیری یکسان باشند می‌توان از اندازه‌های واریانس و یا انحراف معیار استفاده نمود. فرض کنید در تحقیقی معدل دیپلم و دانشگاه و سن افراد مورد سؤال است. برای مقایسه پراکندگی در معدل‌های دیپلم یا پراکندگی در معدل‌های دانشگاه، از اندازه واریانس می‌توان استفاده کرد. اما اگر بخواهیم پراکندگی سن را با پراکندگی معدل دانشگاه مقایسه کنیم، امکان استفاده از اندازه‌های واریانس و انحراف معیار به تنهایی نیست و لذا ضریب تغییرات که به وسیله رابطه $CV = \frac{S}{\bar{X}} \times ۱۰۰$ محاسبه می‌گردد را می‌توان برای چنین مقایسه‌هایی به کار برد. نمونه یا جمعیت‌ای که دارای ضریب تغییرات کوچکتری باشد از پراکندگی کمتری برخوردار است. با این کار عملاً نقش واحد اندازه‌گیری را از بین می‌بریم و امکان مقایسه داده‌ها با ماهیت مختلف و یا واحدهای متفاوت را فراهم می‌آوریم.

مثال۲۵: اگر میانگین حقوق کارگران ۱۸۰ هزار تومان در ماه با انحراف معیار۴۳۰۰ تومان و میانگین حقوق کارمندان ۱۶۰هزار تومان با انحراف معیار ۳۲۰۰ تومان باشد. ضریب تغییرپذیری کدام دسته بیشتر است؟

$$\text{ضریب تغییرات حقوق کارگران} = \frac{۴۳۰۰}{۱۸۰۰۰۰} \times ۱۰۰ = ۲/۳۹$$

$$\text{ضریب تغییرات حقوق کارمندان} = \frac{۳۲۰۰}{۱۶۰۰۰۰} \times ۱۰۰ = ۲$$

در نتیجه حقوق کارگران از پراگندگی بیشتری برخوردار است.

مثال۲۶: در جدول فراوانی زیر ضریب تغییرپذیری را به دست آورید؟

X_i	f_i	$(X_i - \bar{X})^۲$	$f_i(X_i - \bar{X})^۲$
۱	۱	۱۶	۱۶
٤	۱	۱	۱
٥	٦	۰	۰
٦	۱	۱	۱
۹	۱	۱۶	۱۶
جمع	۱۰		۳٤

$$\bar{X} = ۵$$

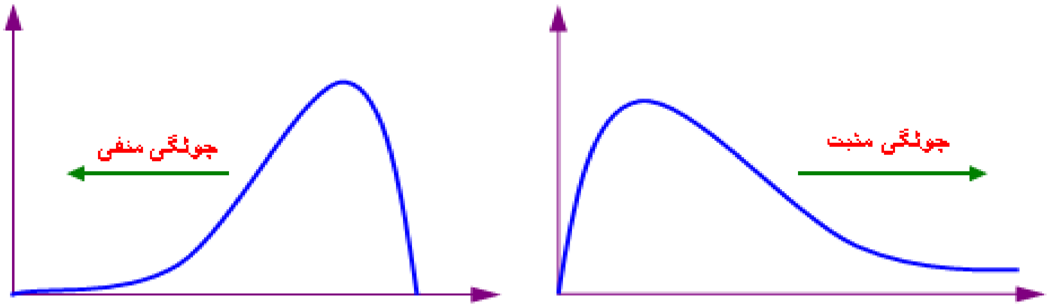
$$S^۲ = ۳/۷۶$$

$$S = \sqrt{۳/۷۶} = ۱/۹۴$$

$$CV = \frac{۱/۹۴}{۵} \times ۱۰۰ = ۳۸/۸$$

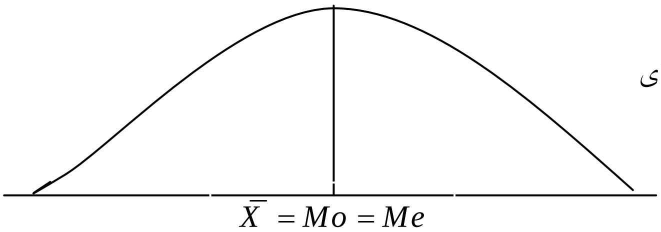
تقارن و چولگی

ممکن است دو مجموعه داده از لحاظ گرایش به مرکز و پراکندگی تفاوت چندانی با هم نداشته باشند ولی از نظر تقارن و چولگی با یکدیگر متفاوت باشند. از این رو لازم است تا شاخص‌های دیگری برای اندازه‌گیری میزان تقارن یا عدم تقارن داده‌ها معرفی شوند. در شکل زیر دو نمونه از داده‌های نامتقارن نشان داده شده است.

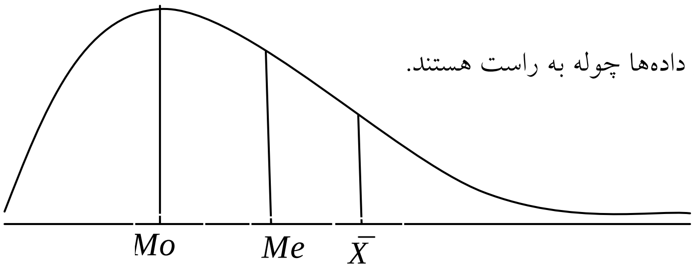


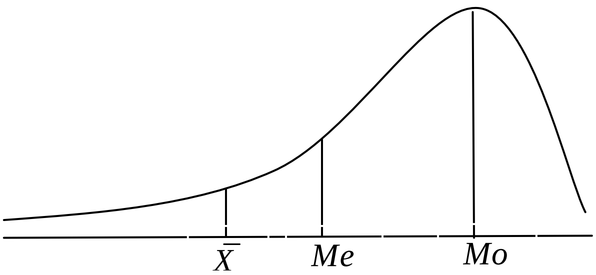
به میزان انحراف توزیع داده‌ها از یک توزیع متقارن با میانگین صفر و واریانس یک(نرمال استاندارد)، چولگی یا کجی می‌گویند. برای تعیین مقدار و نوع انحراف از توزیع متقارن از شاخصی به نام ضریب چولگی که آن را با SK نمایش می‌دهند، استفاده می‌شود. در ادامه چندین روش برای محاسبه ضریب چولگی بیان خواهد شد. ضریب چولگی ممکن است مقداری مثبت، منفی و یا صفر باشد که بسته به مقدار آن نوع چولگی داده‌ها مشخص می‌شود.

$SK = ۰$ در این حالت توزیع داده ها متقارن و شاخص‌های میانه، میانگین و مد بر هم منطبق هستند.



$SK > ۰$ در این حالت توزیع داده‌ها نامتقارن و داده‌ها چوله به راست هستند.





$SK > ۰$ در این حالت توزیع داده‌ها متقارن و داده ها چوله به چپ هستند.

ضریب چولگی بر حسب چارک‌ها:

اگر چارک‌های اول، دوم و سوم یک توزیع را در اختیار داشته باشیم، ضریب چولگی چارک‌ها را می‌توان به طریق زیر محاسبه کرد.

$$sk = \frac{(Q_۳ - Q_۲) - (Q_۲ - Q_۱)}{(Q_۳ - Q_۲) + (Q_۲ - Q_۱)} = \frac{Q_۳ - Q_۱ - ۲Q_۲}{Q_۳ - Q_۱}$$

مثال۲۷: اگر در یک مجموعه داده چارک‌های اول، دوم و سوم به ترتیب $Q_۱ = ۵$، $Q_۲ = ۹$ و $Q_۳ = ۱۱$ باشند، ضریب چولگی چارکی را یافته و در مورد تقارن داده‌ها بحث کنید؟

حل: $sk = \frac{۱۱ + ۵ - ۱۸}{۱۱ - ۵} = -\frac{۱}{۳}$ با توجه به منفی شدن ضریب چولگی می‌توان نتیجه گرفت که داده‌ها نامتقارن بوده و چوله به چپ هستند.

ضریب چولگی پیرسن:

یکی دیگر از شاخص‌های اندازه‌گیری میان تقارن یا عدم تقارن داده‌ها، ضریب چولگی پیرسن است. اگر $\bar{X}$، میانگین، Mo، نما، Md، میانه و S انحراف معیار داده‌ها باشند، دو ضریب چولگی که هر دو منصوب به پیرسن است از رابطه‌های زیر به دست می‌آیند.

ضریب چولگی اول پیرسن	$sk = \frac{(\bar{X} - Mo)}{S}$
ضریب چولگی دوم پیرسن	$sk = \frac{۳(\bar{X} - Md)}{S}$

مثال۲۸: در جدول توزیع فراوانی زیر ضریب چولگی پیرسن و ضریب چولگی چارک‌ها را به دست آورید.

حدود طبقات	f_i
۷ - ۱۱	۲
۱۱ - ۱۵	۳
۱۵ - ۱۹	۵
۱۹ - ۲۳	۱۰
۲۳ - ۲۷	۲

$$\begin{cases} \bar{x} = ۱۸/۲۷ \\ s^۲ = ۱۹/۵۷ \\ s = \sqrt{۱۹/۵۷} = ۴/۴۲ \\ Md = ۱۹/۴ \end{cases} \Rightarrow sk = \frac{۳(\bar{x} - Md)}{s} = -۰/۷۷$$

$$\begin{cases} Q_۱ = ۱۵/۴ \\ Q_۲ = ۱۹/۴ \\ Q_۳ = ۲۱/۶ \end{cases} \Rightarrow sk = -۰/۳۹$$

برجستگی (کشیدگی)

میزان کشیدگی یا پخی منحنی فراوانی را نسبت به یک منحنی متقارن با میانگین صفر و واریانس یک (منحنی نرمال استاندارد)، برجستگی آن منحنی می‌نامند. فرض کنید $m_۴$ گشتاور مرکزی چهارم و S انحراف استاندارد باشد. چون بر اساس خصوصیات توزیع فراوانی نرمال استاندارد مقدار $\frac{m_۴}{S^۴}$ به عدد ۳ نزدیک است، معیار برجستگی را از رابطهٔ $k = \frac{m_۴}{S^۴} - ۳ = \frac{\frac{\sum_{i=۱}^{n}(X_i - \bar{X})^۴}{n}}{S^۴} - ۳$ به دست می‌آورند. بر حسب آنکه k مثبت یا منفی باشد منحنی فراوانی نسبت به منحنی نرمال استاندارد کشیده‌تر (ارتفاع بیشتر) یا پخ‌تر (ارتفاع کمتر) است. اکر kنزدیک صفر باشد، برجستگی منحنی فراوانی طبیعی است.

مثال۲۹: در بررسی طول عمر ۱۰۰ باطری اتومبیل اگر میانگین، میانه، انحراف استاندارد و کشیدگی به ترتیب ۳/۵، ۳/٤۸، ۱/٦۵، ۱/۷۵– سال باشد. دربارهٔ شکل توزیع (نمودار هیستوگرام) آن چه می‌توان گفت؟

پاسخ: با اطلاعات داده شده ضریب چولگی دوم پیرسون عبارتند از: $sk = \frac{۳(۳/۵ - ۳/۴۸)}{۰/۶۵} = ۰/۳۶$.
بنابراین منحنی فراوانی طول عمر باطری‌ها کمی چوله به راست می‌باشد. با توجه به مقدار ضریب برجستگی منحنی فراوانی عمر باطری‌ها نسبت به منحنی نرمال استاندارد پخ‌تر است.

داده‌های استاندارد

در این بخش به معرفی یکی از کاربردهای مفید میانگین و انحراف استاندارد در مقایسهٔ واحدهای جمعیت برای موضوعات مختلف می‌پردازیم. فرض کنید $x_۱, x_۲, ..., x_n$ متغیرهای مشاهده‌ای با میانگین $\bar{x}$ و انحراف استاندارد S باشند. داده‌های $z_i = \frac{x_i - \bar{x}}{s}$ $\quad i = ۱,۲,...,n$ را داده‌های استاندارد نامند. داده‌های استاندارد دارای میانگین صفر و واریانس یک هستند و شکل توزیع آنها متقارن می‌باشد. کاربرد داده‌های استاندارد در مثال زیر واضح‌تر است.

مثال۳۰: نمره کارکنان یک کلاس آموزشی در آزمون کامپیوتر دارای میانگین ۷۲ و انحراف استاندارد ۱۵ و در آزمون نگارش دارای میانگین ۵۰ و انحراف استاندارد ۲۰ است. اگر نمره فردی در کامپیوتر ٦۰ و در نگارش ۳۵ باشد، آن‌گاه معلومات فرد در کدام موضوع بیشتر است؟

پاسخ: چون این دو آزمون با مقیاسهای مختلف به‌عمل آمده‌اند، مقایسه اعداد ٦٠ و ۳۵ مفهومی ندارند. اگر نمره‌های دو آزمون تقریباً دارای منحنی فراوانی نرمال باشند، تنها بعد از استاندارد کردن می‌توان آنها را با هم مقایسه کرد.

نمره استاندارد فرد در آزمون کامپیوتر
$$z_{camputer}=\frac{۶۰-۷۲}{۱۵}=-۰/۸$$

نمره استاندارد فرد در آزمون نگارش
$$z_{negaresh}=\frac{۳۵-۵۰}{۲۰}=-۰/۷۵$$

بنابراین معلومات فرد در آزمون نگارش بیشتر است، زیرا داریم $z_{negaresh}>z_{computer}$.

مسائل

۱- تعداد تصادفات ده روز در یک چهارراه به صورت زیر است.

الف- میانگین و نمای این داده‌ها را به دست آورید.

ب- ضریب تغییرات آنها را محاسبه کنید.

٤ ۳ ۰ ۲ ۳ ۳ ۱ ۳ ۰ ۱

۲- میانه و نمای داده‌ها را برای داده‌های زیرمحاسبه کنید.

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

۳- داده‌های مربوط به یک مطالعه را ۵ برابر کرده و با عدد ۳ جمع کرده‌ایم. میانگین جدید ۲۰/۵ و واریانس جدید ۱۰۰ است. میانگین و واریانس اولیه را به دست آورید؟

٤- جدول زیر سطح هموگلوبین در ۲۵ زن تحت مطالعه را نشان می‌دهد.

الف- میانگین سطوح هموگلوبین در بین زنان چقدر است؟

ب- ضریب چولگی داده‌ها را به دست آورده و تفسیر کنید؟

ج- چارک سوم داده‌ها معلوم کنید؟

د- نمودار چند ضلعی فراوانی تجمعی را رسم کنید؟

فراوانی	حدود طبقات
٤	۷ - ۹
۷	۹ - ۱۱
۱۱	۱۱ - ۱۳
۳	۱۳ - ۱۵

۵- در جدول زیر واریانس، انحراف استاندارد، متوسط انحراف و ضریب تغییر پذیری را به دست آورید؟

فراوانی	حدود طبقات
۲	۱ - ۵
۷	٥ - ۹
۱۱	۹ - ۱۳
١٦	۱۳ - ۱۷
۱۰	۱۷ - ۲۱
٤	۲۱ - ۲۵

٦- اگر میانگین اعداد $4X_i$ برابر A باشد: میانگین اعداد $(4X_i)-(8X_i)+(4X_i-2)$ چه مقدارخواهد بود ؟

۷- اگر $Var(X)=4$ باشد، انحراف معیار $2aX+b$ کدام است ؟

۸- در یک نمونه آماری اگر انحراف معیار سه برابر و میانگین چهار برابر شود ضریب تغییرات جدید چند درصد ضریب تغییرات اولیه است؟

۹- اگر $\bar{x}=۲/۴$، $MD=۰/۴۸$، $CV=۲۰$ و $Mo=۲۱$ باشد. کشیدگی (چولگی) این توزیع نسبت به توزیع نرمال چگونه است؟

۱۰- در یک جدول توزیع فراوانی، اگر حد پایین طبقه اول ۳۵ و حد پایین طبقه دوم ٤۲ باشد. نقطه میانی طبقه سوم را به دست آورید؟

۱۱- برای جدول توزیع فراوانی روبرو مقدار میانگین و انحراف متوسط را به دست آورید؟

X_i	۱	٤	٥	٦	۹
فراوانی	۱	۱	٦	۱	۱

۱۲- چارک اول داده‌های جدول مقابل کدام است؟

حدود طبقات	فراوانی
۱ - ۲	٤
۲ - ۳	۳
۳ - ٤	۱
٤ - ٥	۲

۱۳- ضریب تغییرپذیری داده‌های ٥، ٤، ۳، ۲ و ۱ را به دست آورید؟

١٤- در یک کنفراس، تعداد ۲۰۰ مقاله ارائه شده است. اگر سهم کشور ایران در مقالات ارائه شده ١٦ باشد، در نمودار دایره‌ای زاویه اندازه قطاع مربوط به ایران چقدر خواهد بود؟

١٥- در دسته‌بندی داده‌ها، اگر تعداد طبقات برابر ۱۰، فاصله طبقات برابر۷ و کوچکترین عدد ٣٤/٥ باشد. آنگاه عدد ٤١ در کدام طبقه قرار دارد؟

فصل چهارم

احتمال

پیدایش رسمی احتمال از قرن هفدهم به عنوان متدی برای محاسبه شانس در بازی‌های قمار بوده است. اگر چه ایده‌های احتمال شانس و تصادفی بودن از تاریخ باستان در رابطه با افسونگری و بخت‌آزمایی و بازیهای شانسی و حتی در تقسیم کار بین راهبان در مراسم مذهبی وجود داشته است و به علاوه شواهدی از بکارگیری این ایده‌ها در مسائل حقوق، بیمه، پزشکی و نجوم نیز یافت می‌شود. اما بسیار عجیب است که حتی یونانیان اثری از خود در رابطه با استفاده از تقارنی که در هندسه بکار می برده اند در زمینه احتمال یا اصولی که حاکم بر مسائل شانس باشد بجا نگذاشته‌اند. ارسطو پیشامدها را به سه دسته تقسیم می‌نمود:

۱- پیشامدهای قطعی که لزوماً اتفاق می‌افتند.
۲- پیشامدهای احتمالی که در بیشتر موارد اتفاق می‌افتند.
۳- پیشامدهای غیر قابل پیش‌بینی و غیر قابل شناسایی که فقط با شانس محض رخ می‌دهند.

اما ارسطو به تعبیرهای مختلف احتمال اعتقاد نداشته و فقط احتمال شخصی که مربوط به درجه اعتقاد افراد نسبت به وقوع پیشامدهاست را معتبر می‌دانسته است. پاسکال و فرما اولی کسانی هستند که در اوایل قرن هفدهم مسائل مربوط به بازیهای شانسی را مورد مطالعه قرار دادند و این دو نفر به عنوان بنیانگزاران تئوری ریاضی احتمال لقب گرفته‌اند.

قبل از آنکه به بحث در مورد احتمال بپردازیم لازم است با برخی از مفاهیم و تعاریف اولیه احتمالات آشنا شویم.

آزمایش تصادفی: انجام دادن هر نوع فعالیتی که به رسیدن نتیجه منجر شود، آزمایش نام دارد و آزمایش تصادفی فرآیندی است که نتیجه آن را قبل از انجام آزمایش با قطعیت نمی‌توان معین نمود و فقط پیش‌بینی در مورد نتیجه آن صورت می‌گیرد. برای مثال تعداد شیرها در پرتاب دو سکه.

فضای نمونه: مجموعه تمام حالات ممکن و متمایز یک آزمایش تصادفی را فضای نمونه می‌گویند و آن را با S نمایش می‌دهند.

برای نمونه در پرتاب یک تاس فضای نمونه به صورت $S=\{۱,۲,۳,۴,۵,۶\}$ خواهد بود. و یا در پرتاب دو سکه فضای نمونه به صورت $S=\{HH, HT, TH, TT\}$است که در آن حرف H بیانگر شیر یا روی سکه و حرف T بیانگر خط یا پشت سکه است.

فضای نمونه گسسته: اگر اعضای S قابل شمارش باشند(متناهی یا نامتنهایی)، فضای نمونه را گسسته می‌گویند. $S=\{۳,۱۲,۷,۲,۲۵,۴۸\}$

فضای نمونه پیوسته: اگر فضای نمونه قابل شمارش نباشد. برای مثال اگر زمان را با T نمایش دهیم، فضای نمونه $S=\{T \mid ۵<t<۲۰\}$ یک فضای نمونه پیوسته است.

نقطه نمونه: هر یک از اعضای S را یک نقطه نمونه (برآمد) می‌گویند.

پیشامد: به هر زیر مجموعه‌ای از فضای نمونه یک پیشامد می‌گویند و آنها را حروف بزرگ $A,B,C,....$ نمایش می‌دهند.

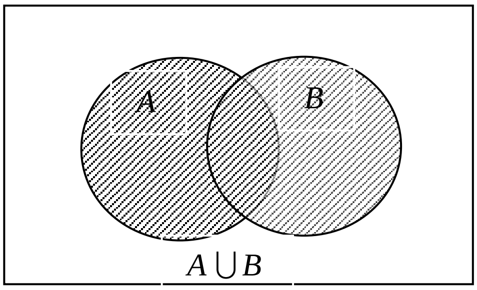
پیشامد ساده و مرکب: اگر پیشامدی فقط یک عضو داشته باشد(یک نقطه نمونه) آن را یک پیشامد ساده و چنانچه بیش از یک عضو داشته باشد آن را پیشامد مرکب می‌گویند. برای مثال $A=\{HH\}$ یک پیشامد ساده و $B=\{H,H\}$ یک پیشامد مرکب است.

پیشامد تهی را با $\{\}$ یا $\varnothing$ نمایش داده و آن را پیشامد غیر ممکن می‌نامند و پیشامد S(فضای نمونه) را پیشامند حتمی می‌گویند.

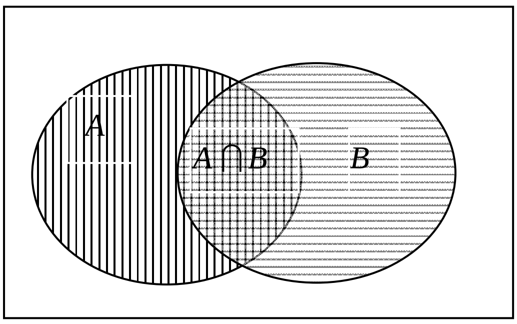
پیشامدهای ناسازگار(جدا از هم): دو پیشامد A و B را ناسازگار گویند هرگاه هیچ نقطه نمونه مشترکی نداشته باشند، به عبارت دیگر $A\cap B=\varnothing$.

مثال۱: در پرتاب دو سکه فضای نمونه به صورت $S=\{HH,HT,TH,TT\}$ خواهد بود حال اگر دو پیشامد A و B را به صورت $A=\{HH\}$ و $B=\{TT\}$ تعریف کنیم آنگاه $A\cap B=\varnothing$. در نتیجه دو پیشامد ناسازگار هستند.

اجتماع دو پیشامد: اجتماع دو پیشامد A و B را با $A\cup B$ نمایش داده و فقط و فقط زمانی رخ می‌دهد که A یا B (یا هر دو) رخ دهند. در مثال قبل داریم: $A\cup B=\{HH,TT\}$.



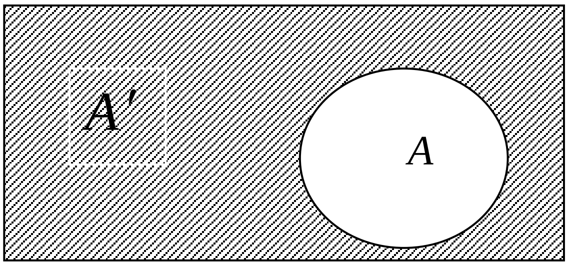
اشتراک دو پیشامد: اجتماع دو پیشامد A و B را با $A\cap B$ نمایش داده و فقط و فقط زمانی رخ می‌دهد که A و B هر دو با هم رخ دهند.



مثال۲: اشتراک دو پیشامد $A=\{HH,HT,TH\}$ و $B=\{HT,TH,TT\}$ به صورت $A\cap B=\{HT,TH\}$ خواهد بود.

پیشامد متمم: متمم یک پیشامد، با تعریف فضای نمونه (مجموعه‌ی مرجع) معنی پیدا می کند. اگر A یک زیر مجموعه از فضای نمونه S باشد، متمم A در S را با نماد A^C یا A' نمایش می‌دهند و شامل آن عضوهایی از فضای نمونه است که در A نباشند. به عبارت ساده‌تر اگر از فضای نمونه، اعضای پیشامد A را برداریم ، آنچه باقی می‌ماند را A^c می‌نامیم .

به زبان سورها متمم A به صورت زیر تعریف می شود: $A^c=\{x\in S \mid x\notin A\}$.



مثال۳: در پرتاب دو سکه اگر پیشامد A را آمدن دو شیر تعریف کنیم آنگاه A^cرا به دست آورید؟

$$\begin{cases} S=\{HH,HT,TH,TT\} \\ A=\{HH\} \end{cases} \Rightarrow A^c=\{HT,TH,TT\}$$

مثال٤: در پرتاب یک تاس سالم اگر A و B را به ترتیب پیشامد رخ دادن عدد زوج و فردو پیشامد C را رخ دادن عدد اول تعریف کنیم آنگاه، S، A، B، C، $A\cap B$، $A\cup C$، $B\cap C$ و C' را تعیین کنید؟

$$\begin{cases} S=\{۱,۲,۳,۴,۵,۶\} \\ A=\{۲,۴,۶\} \\ B=\{۱,۳,۵\} \\ C=\{۲,۳,۵\} \end{cases} \& \begin{cases} A\cap B=\varnothing \\ A\cup C=\{۲,۳,۴,۵,۶\} \\ B\cap C=\{۳,۵\} \\ C'=\{۱,۴,۶\} \end{cases}$$

مفهوم احتمال:

وقتی در زندگی روزمره از عباراتی همچون احتمال یا شانس صحبت می‌کنیم به خوبی مفهوم عامیانه آن را می‌دانیم. مثلاً وقتی می‌گوییم که احتمال برد تیم الف ۷۰٪ است از معنی آن آگاه هستیم. حال اگر این اعداد را به اعدادی در بازه [۰,۱] ببریم، در این صورت با احتمال سروکار داریم. هر چه این عدد به صفر نزدیکتر باشد، امکان وقوع آن کمتر و هرچه به یک نزدیکتر باشد، امکان وقوع آن پیشامد بیشتر است. در مفهوم کلی احتمال یعنی حد فراوانی نسبی، وقتی که تعداد داده‌ها به بی‌نهایت میل می‌کند $\left(\lim_{n\to\infty}\frac{f_i}{n}\right)$.

تعریف احتمال: احتمال تابعی است مانند P، که به هر پیشامد A از فضای نمونه S، عدد حقیقی $P(A)$ را به گونه‌ای نسبت می‌دهد که در سه اصل معروف کلموگروف زیر صدق کند.

۱- $P(S)=1$
۲- برای هر پیشامد A در فضای نمونه، $0 \le P(A) \le 1$
۳- اگر A_1، A_2، A_3،... پیشامدهای دو به دو ناسازگار باشند آنگاه اجتماع پیشامدها برابر با مجموع پیشامدها خواهد بود. به عبارت ریاضی داریم: $P(\bigcup_i A_i)=\sum_i P(A_i)$.

احتمال کلاسیک

چنانچه فرض کنیم کلیه نقطه نمونه‌های فضای نمونه هم‌شانس باشند و تعداد حالات مساعد پیشامد A را با $n(A)$ و تعداد حالات ممکن فضای نمونه را با $n(S)$ نمایش دهیم در این صورت احتمال پیشامد A را که با $P(A)$ نشان می‌دهیم به صورت زیر تعریف می‌شود: $P(A)=\frac{n(A)}{n(S)}=\frac{\text{تعداد حالات مساعد}}{\text{تعداد کل حالات}}$.

مثال۵: در پرتاب دو سکه اگر پیشامد A را آمدن یک شیر تعریف کنیم. احتمال پیشامد A را حساب کنید؟

پاسخ:

$$S=\{HH,TH,HT,TT\}$$
$$A=\{HT,TH\}$$
$$P(A)=\frac{n(A)}{n(S)}=\frac{2}{4}$$

قضیه‌های پایه‌ای احتمال

از این پس، مبنای همه تعاریف، سه اصل اصولی احتمال بیان شده در بالا خواهند بود. اکنون چند قضیه پایه‌ای احتمال را با استفاده از سه اصل اساسی احتمال بیان می‌کنیم.

قضیه۱: برای هر پیشامد دلخواه A داریم: $P(A^c)=1-P(A)$.

مثال۶: اگر در پرتاب یک سکه، $P(A)=P(\{H\})=0/75$ باشد، آنگاه:

$$P(A^c)=P(\{T\})=1-0/75=0/25$$

قضیه۲: برای هر پیشامد تهی $\varnothing$ داریم: $P(\varnothing)=0$.

قضیه۳: اگر $E\subseteq F$ باشد، آنگاه $P(E)\le P(F)$ خواهد بود.

قضیه۴: اگر A و B دو پیشامد دلخواه باشند داریم: $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

تذکر: اگر A و B دو پیشامد ناسازگار باشند، یعنی $A \cap B = \varnothing$، آنگاه $P(A \cap B) = ۰$. در این حالت داریم: $P(A \cup B) = P(A) + P(B)$.

مثال۷: تاسی را ۲ بار پرتاب می‌کنیم. مطلوب است محاسبه احتمال اینکه هر دو عدد، زوج باشند؟

پاسخ: ابتدا فضای نمونه را مشخص می‌کنیم:

$$S = \{(۱,۱),(۱,۲),(۱,۳),...,(۶,۵),(۶,۶)\} \quad \& \quad n(S) = ۶ \times ۶ = ۳۶$$

الف: اگر A را پیشامد اینکه هر دو عدد زوج باشند، تعریف کنیم داریم:

$$A = \{(۲,۲),(۲,۴),(۲,۶),(۴,۲),(۴,۴),(۴,۶),(۶,۲),(۶,۴),(۶,۶)\} \quad \Rightarrow n(A) = ۹$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{۹}{۳۶} = ۰/۲۵$$

مثال۸: سکه‌ای را پرتاب می‌کنیم. اگر احتمال آمدن شیر ۲ برابر احتمال آمدن خط باشد، احتمال آمدن شیر را محاسبه کنید؟

پاسخ: فضای نمونه پرتاب یک سکه به صورت $S = \{H, T\}$ داریم. از اصل اول احتمال کلموگروف داریم: $P(S) = ۱ \quad \Rightarrow \quad P(H) + P(T) = ۱$. از طرفی طبق صورت مسئله داریم: $P(H) = ۲P(T)$. در نتیجه:

$$P(S) = P(H) + P(T) = ۲P(T) + P(T) = ۳P(T) = ۱$$

$$\Rightarrow P(T) = \frac{۱}{۳} \quad \& \quad P(H) = ۲P(T) = \frac{۲}{۳}$$

مثال۹: تاسی را پرتاب می‌کنیم. اگر احتمال آمدن عدد زوج، ۲ برابر احتمال آمدن عدد فرد باشد. مطلوبست محاسبه احتمال آمدن عدد زوج؟

پاسخ: اگر احتمال آمدن عدد فرد را a در نظر بگیریم، آنگاه احتمال آمدن عدد زوج $۲a$ خواهد بود. تابع توزیع احتمال را می‌توان به صورت زیر نوشت:

X_i	۱	۲	۳	٤	٥	٦
$P(X_i)$	a	$۲a$	a	$۲a$	a	$۲a$

بنا بر اصل اول احتمال کلموگروف داریم:

$$P(S) = P(۱) + P(۲) + P(۳) + P(۴) + P(۵) + P(۶)$$
$$= a + ۲a + a + ۲a + a + ۲a = ۹a = ۱ \quad \Rightarrow a = \frac{۱}{۹}$$

در نتیجه: $P(\text{زوج}) = ۲P(\text{فرد}) = \frac{۲}{۹}$.

مثال۱۰: دو سکه با هم انداخته می‌شود. مطلوب است محاسبه احتمالات زیر:

الف- احتمال اینکه یک شیر بیاید؟

ب- احتمال اینکه حداقل یک شیر بیاید؟

ج- احتمال اینکه هر دو شیر یا هردو خط بیایند؟

د- احتمال اینکه حداقل یک شیر یا حداقل یک خط بیاید؟

پاسخ: ابتدا فضای نمونه را مشخص می‌کنیم: $S=\{HH,HT,TH,TT\} \quad \& \quad n(S)=۴$

الف: اگر A را پیشامد آمدن یک شیر تعریف کنیم داریم: $A=\{HT,TH\}$ و $n(A)=۲$. در نتیجه داریم:

$$P(A)=\frac{n(A)}{n(S)}=\frac{۲}{۴}.$$

ب: اگر B را پیشامد آمدن حداقل یک شیر تعریف کنیم داریم: $B=\{HH,HT,TH\}$ و $n(B)=۳$. در نتیجه داریم: $P(B)=\frac{n(B)}{n(S)}=\frac{۳}{۴}$.

ج: اگر E را پیشامد آمدن دو شیر و F را پیشامد آمدن دو خط تعریف کنیم داریم:

$$\begin{cases} E=\{HH] \\ F=\{TT\} \\ E\bigcap F=\varnothing \end{cases} \qquad \begin{aligned} P(E\bigcup F)&=P(E)+P(F)-P(E\bigcap F) \\ &=\frac{n(E)}{n(S)}+\frac{n(F)}{n(S)}-\frac{n(E\bigcap F)}{n(S)} \\ &=\frac{۱}{۴}+\frac{۱}{۴}-\frac{۰}{۴}=\frac{۲}{۴} \end{aligned}$$

توجه: مسئله بالا را می‌توانستیم به طور مستقیم نیز به صورت زیر حل نماییم.

$$E\bigcup F=\{HH,TT\}$$
$$P(E\bigcup F)=\frac{n(E\bigcup F)}{n(S)}=\frac{۲}{۴}$$

د: : اگر G را پیشامد آمدن حداقل یک شیر و K را پیشامد آمدن حداقل یک خط تعریف کنیم داریم:

$$\begin{cases} G=\{HH,HT,TH] \\ K=\{TT,TH,HT\} \\ G\bigcap K=\{TH,HT\} \end{cases} \qquad \begin{aligned} P(G\bigcup K)&=P(G)+P(J)-P(G\bigcap J) \\ &=\frac{n(G)}{n(S)}+\frac{n(J)}{n(S)}-\frac{n(G\bigcap J)}{n(S)} \\ &=\frac{۳}{۴}+\frac{۳}{۴}-\frac{۲}{۴}=۱ \end{aligned}$$

مسئله بالا را نیز می‌توانستیم به طور مستقیم به صورت زیر حل نماییم.

$$G\bigcup K=\{HH,HT,TH,TT\}$$
$$P(G\bigcup K)=\frac{n(G\bigcup K)}{n(S)}=۱$$

مثال۱۱: ظرفی حاوی ۱۰ گوی است که با شماره‌های ۳ تا ۱۲ شماره‌گذاری شده‌اند. توپی را به تصادف انتخاب می‌کنیم. مطلوب است محاسبه احتمال اینکه عدد روی توپ بر ۳ یا ٤ بخش‌پذیر باشد؟

پاسخ: اگر Aرا پیشامد بخش‌پذیر بودن عدد بر ۳ و B را پیشامد بخش‌پذیر بودن عدد بر ٤ تعریف کنیم داریم:

$$S=\{۳,۴,۵,...,۱۱,۱۲\} \quad ,n(S)=۱۰$$
$$A=\{۳,۶,۹,۱۲\} \quad ,n(A)=۴$$
$$B=\{۴,۸,۱۲\} \quad ,n(B)=۳$$
$$A\cap B=\{۱۲\} \quad n(A\cap B)=۱$$
$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{۴}{۱۰}+\frac{۳}{۱۰}-\frac{۱}{۱۰}=\frac{۶}{۱۰}$$

مثال۱۲: تعداد تصادفات ده چهار راه در بهمن ماه به صورت ۲۹، ۲۸، ۲۷، ٢٦، ٢٦، ٢٦، ۲۵، ٢٤، ٢٤ و ۲۳ ثبت شده است. اگر به طور تصادفی یکی از از این چهارراه‌ها را انتخاب کنیم، مطلوب است محاسبه احتمالات زیر:

الف- احتمال اینکه این چهارراه دقیقاً داری ٢٦ تصادف باشد؟

ب- احتمال اینکه این چهارراه دارای ٢٤ یا ۲۵ تصادف باشد؟

ج- احتمال اینکه تعداد تصادفات این چهارراه کمتر از ۲۷ تصادف باشد یا اینکه دقیقا داری ٢٤ تصادف باشد؟

حل: برای فضای نمونه داریم: $S=\{۲۳,۲۴,۲۴,۲۵,۲۶,۲۶,۲۶,۲۷,۲۸,۲۹\}$

الف: اگر A را پیشامد ٢٦ تصادف تعریف کنیم داریم:

$$A=\{۲۶,۲۶,۲۶\}$$
$$P(A)=\frac{n(A)}{n(S)}=\frac{۳}{۱۰}$$

ب: : اگر E را پیشامد ۲۵ تصادف و F را پیشامد ٢٤ تصادف تعریف کنیم داریم:

$$E=\{۲۵\}$$
$$F=\{۲۴,۲۴\}$$
$$E\cap F=\varnothing$$
$$P(E\cup F)=P(E)+P(F)-P(E\cap F)$$
$$=\frac{n(E)}{n(S)}+\frac{n(F)}{n(S)}-\frac{n(E\cap F)}{n(S)}$$
$$=\frac{۱}{۱۰}+\frac{۲}{۱۰}-۰=۰/۳$$

ج: اگر G را پیشامد کمتر از ۲۷ تصادف و K را پیشامد ٢٤ تصادف تعریف کنیم داریم:

$$G=\{۲۳,۲۴,۲۴,۲۵,۲۶,۲۶,۲۶\}$$
$$K=\{۲۴,۲۴\}$$
$$G\cap K=\{۲۴,۲۴\}$$
$$P(G\cup K)=P(G)+P(K)-P(G\cap K)$$
$$=\frac{n(G)}{n(S)}+\frac{n(K)}{n(S)}-\frac{n(G\cap K)}{n(S)}$$
$$=\frac{۷}{۱۰}+\frac{۲}{۱۰}-\frac{۲}{۱۰}=./۷$$

مثال۱۳: اگر شخصی در یک جایگاه بنزین توقف کند، احتمال آنکه لاستیک را مورد بازدید قرار دهد برابر ۰/۵ و احتمال انکه روغن ماشین را مورد بازدید قرار دهد ۰/۳ و احتمال اینکه هر دو را با هم بازدید کند برابر ۰/۱۲ است.مطلوب است محاسبه احتمال آنکه اگر فردی در جایگاه بنزین توقف کند:

الف- حداقل یکی از این دو مورد را بازدید کند؟

ب- هیچکدام را بازبینی نکند؟

پاسخ: اگر A را پیشامد بازدید از لاستیک و B را پیشامد بازدید از روغن تعریف کنیم، داریم:

الف: $P\left(A\cup B\right)=P\left(A\right)+P\left(B\right)-P\left(A\cap B\right)=۰/۵+۰/۳-۰/۱۲=۰/۶۸$

ب: برای حل این قسمت از احتمال متمم استفاده می‌کنیم،

$P($هیچکدام را بازدید نکند$)=۱-P($هر دو را بازدید کند$)=۱-P(A\cap B)=۱-۰/۱۲=۰/۸۸$

روش‌های شمارش و کاربرد آن در احتمالات:

در بخش قبل دیدیم که برای محاسبه احتمال یک پیشامد دلخواه همانند A، طبق فرمول احتمال، $p(A)=\frac{n(A)}{n(S)}$، نیاز به شمارش تعداد اعضای پیشامد A و تعداد اعضای فضای نمونه داریم. در مثال‌هایی که تاکنون ارائه کردیم بعلت کم بودن تعداد اعضا اینکار به راحتی با نوشتن اعضای فضای نمونه و پیشامد مورد نظر انجام‌پذیر بود. انجام این کار زمانی که مجموعه‌ها بزرگ باشند کاری بس مشکل و گاهی غیرممکن است. برای رفع این مشکل به دنبال یافتن روشی برای شمارش تعداد اعضای پیشامدها و فضای نمونه بدون نیاز به نوشتن آن هستیم که اصول شمارش نامیده شده و در ادامه به تشریح آن خواهیم پرداخت.

اصل جمع: اگر برای انجام کاری دو طریق مختلف وجود داشته باشد و طریق اول را بتوان به n حالت و طریق دوم را بتوان به m حالت انجام داد، آنگاه تعداد کل حالاتی که می‌توان کار را انجام داد $n+m$ حالت است.

مثال۱٤: برای رفتن از شهر (الف) به شهر (ب) دو راه زمینی و هوایی وجود دارد. روش زمینی را می‌توان با سه وسیله اتوبوس، قطار و اتومبیل شخصی طی کرد و مسیر هوایی را می‌توان با دو وسیله هواپیما و بالن طی کرد. در نتیجه طبق اصل جع برای رفتن از شهر (الف) به شهر (ب) کلاً، $۵=۳+۲$ روش وجود دارد.

تعمیم اصل جمع: اگر برای انجام کاری K طریق مختلف وجود داشته باشد و طریق اول را بتوان به $n_۱$ حالت، طریق دوم را بتوان به $n_۲$ حالت انجام داد،...و طریق Kام را بتوان به n_k حالت انجام داد، آنگاه تعداد کل حالاتی که می‌توان کار را انجام داد $n_۱+n_۲+...+n_k$ حالت است.

اصل ضرب: اگر عملی به دو مرحله اول و دوم تقسیم شود و اگر در مرحله اول n طریق ممکن و برای هر یک از این طرق، m حالت در مرحله دوم وجود داشته باشد، آنگاه کل عمل می‌تواند با ترتیب یاد شده، به nm طریق انجام شود.

مثال۱۵: فرض کنید از شهر A به شهر B، ۵ راه وجود دارد و از شهر B به شهر C، ٤ راه وجود دارد. آنگاه برای رفتن از شهر A به شهر C طبق اصل ضرب $۲۰=۵×۴$ راه وجود دارد.

تعمیم اصل ضرب: اگر برای انجام عملی بایستی K کار به طور همزمان انجام شوند و کار اول را بتوان به $n_۱$ حالت، کار دوم را به $n_۲$،.... و کار Kام را به n_k حالت بتوان انجام داد، آنگاه کل کار را می‌توان به $n_۱n_۲...n_k$ حالت انجام داد.

مثال۱٦: کمیته طرح‌ریزی دانشکده‌ای مرکب از ۳ دانشجوی سال اول، ٤ دانشجوی سال دوم، ۵ دانشجوی سال سوم و ۲ دانشجوی سال آخر است. می‌خواهیم زیر کمیته‌ای که در آن ٤ نفر و از هر کلاسی یک نفر شرکت دارند تشکیل دهیم. چند زیر کمیته می‌توان تشکیل داد؟

پاسخ: طبق اصل ضرب داریم: $۳\times۴\times۵\times۲=۱۲۰$

مثال۱۷: چند نمره پلاک اتومبیل مختلف می‌توان ساخت در صورتی‌که بدانیم از هفت جای در نظر گرفته شده، ۳ جای اول با حرف و ٤ جای بعدی با اعداد پر می‌شوند؟

پاسخ: طبق اصل ضرب داریم $۲۶\times۲۶\times۲۶\times۱۰\times۱۰\times۱۰\times۱۰=۱۷۵,۷۶۰,...$

مثال۱۸: در مثال بالا اگر تکرار مجاز نباشد چند پلاک می‌توان نوشت؟

پاسخ: $۲۶\times۲۵\times۲۴\times۱۰\times۹\times۸\times۷=۷۸,۶۲۴,...$

مثال۱۹: با اعداد ۳، ۵، ٦ و ۸:

الف- چند عدد سه رقمی بدون تکرار می‌توان نوشت؟ طریق۲٤= $\boxed{۲}\times\boxed{۳}\times\boxed{٤}$

ب- چند عدد سه رقمی با تکرار می‌توان نوشت؟ طریق٦٤= $\boxed{٤}\times\boxed{٤}\times\boxed{٤}$

مثال۲۰: با اعداد ۱،۵،٤،۹ و ۰ :

الف- چند عدد سه رقمی بدون تکرار می‌توان نوشت؟ عدد $۴۸ = ۳\times۴\times۴$

ب- چند عدد زوج ۳ رقمی بدون تکرار می‌توان نوشت؟ عدد $۲۱ = (۱\times۳\times۳) + (۱\times۳\times۴)$

ج- چند عدد فرد ۳ رقمی بدون تکرار می‌توان نوشت؟ عدد $۲۷ = ۳\times۳\times۳$

مثال۲۱: می‌خواهیم از بین ۵ بازرگان، ٦ دبیر و ٤ قاضی یک جلسه دو نفری تشکیل دهیم بطوریکه اعضای جلسه دارای مشاغل مختلف باشند، جلسه را به چند طریق می‌توان تشکیل داد؟

پاسخ: تعداد حالاتی که جلسه شامل بازرگان و دبیر باشد: $۳۰ = ۶\times۵$

تعداد حالاتی که جلسه شامل بازرگان و قاضی باشد: $۲۰ = ۴\times۵$

تعداد حالاتی که جلسه شامل قاضی و دبیر باشد: $۲۴ = ۶\times۴$

در نتیجه تعداد کل حالات برابر با $۷۴ = ۲۴ + ۲۰ + ۳۰$ است.

جایگشت

اگر $a_۱, a_۲, ..., a_n$ به عنوان n شی متمایز باشند آنگاه کنار هم قرار گرفتن این n شی در یک ردیف را یک جایگشت از این n شی می‌گوییم. برای ردیف کردن این n شی کنار هم به n مکان نیاز است. برای قرار دادن اولین شی در خانه اول، n حالت انتخاب داریم. برای قرار دادن دومین شی در خانه دوم، $n-۱$ حالت انتخاب داریم و به همین ترتیب برای قرار داردن n امین شی باقی مانده در خانه nام (خانه آخر) ۱ حالت انتخاب داریم. به این ترتیب بر طبق اصل ضرب برای قرار دادن این n شی در کنار هم در یک ردیف $n\times(n-۱)\times(n-۲)\times...\times۱$ حالت وجود دارد که برابر می‌باشد با $n!$.

به این ترتیب تعداد جایگشت‌های n شی متمایز برابر با $n!$ است.

فاکتوریل: حاصلضرب اعداد صحیح مثبت از ۱ تا n را به صورت $n!$ نمایش داده و آن را nفاکتوریل می‌نامند.

داریم: $n! = n\times(n-۱)\times(n-۲)\times...\times۱$.

تعریف می‌کنیم که: $۰! = ۱! = ۱$.

مثال۲۲: به چندطریق می‌توان ۵ کتاب متفاوت را کنار هم در یک قفسه قرار داد؟

پاسخ: برطبق توضیحات داده شده جواب برابر است با: $۱۲۰ = ۱ \times ۲ \times ۳ \times ۴ \times ۵ = !۵$

مثال۲۳: دانشجویی ۱۰ کتاب دارد که می‌خواهد آنها را در قفسه‌ای بچیند. از این کتاب‌ها ٤ کتاب ریاضی، ۳ کتاب شیمی، ۲ کتاب تاریخ و یک کتاب زبان است. او می‌خواهد کتاب‌هایش را طوری مرتب کند که کتاب‌های مربوط به یک موضوع در کنار هم باشند. چند آرایش مختلف امکان دارد؟

پاسخ: ٤ کتاب ریاضی را می‌توان به !۴، ۳ کتاب شیمی را به !۳، ۲ کتاب تاریخ را به !۲ و یک کتاب زبان را به !۱ می‌توان چید و خود این ٤ نوع کتاب نیز بین خودشان جایگشتی به اندازه !۴ دارند. بنابراین تعداد کل جایگشت‌ها برابر با $۶۹۱۲ = (!۱ \times !۲ \times !۳ \times !۴)!۴$ خواهد بود.

جایگشت با تکرار

در قسمت قبل در مورد گونه‌ای جایگشت توضیح دادیم که در آن اشیا متمایز بودند اما گاهی ممکن است این اشیا متمایز نباشند و مثلا ۳ عدد از آنها از یک نوع باشند. چنین حالاتی را جایگشت باتکرار بررسی می‌کند. با یک مثال روش محاسبه را توضیح می‌دهیم و سپس فرمولی برای محاسبه حالات، بیان می‌کنیم. فرض کنید می‌خواهیم فقط با ارقام ۳،۲،۲،۱ اعداد چهار رقمی بسازیم. یعنی عدد ۱ یکبار، عدد ۲ دو بار و عدد ۳ یکبار آمده باشد. بدیهی است که اگر این چهار رقم متمایز و به غیر صفر بودند تعداد اعداد برابر ۲٤=!٤ عدد می‌شد ولی اصل ضرب در این مورد ناخواسته دو عدد ۲ را متمایز در نظر گرفته است. با نوشتن تعداد حالات متوجه می‌شویم که تعداد حالات واقعی این جایگشت !۲ برابر مقدار محاسبه شده با اصل ضرب است. به این ترتیب تعداد حالات واقعی برابر $\frac{۴!}{۲!}$است. پس به این ترتیب تعداد k شی از یک نوع، به اندازه$k!$ حالات اضافه تولید می‌کنند که باید از کل حالات که با اصل ضرب محاسبه می‌شود برداشته شوند.

اگر n شی در اختیار داشته باشیم که n_1 تا از نوع اول، n_2 تا از نوع دوم، n_3 تا از نوع سوم،....و n_k تا از نوع k ام باشند به گونه‌ای که $n_1+n_2+...+n_k=n$ این n شی به $\frac{n!}{n_1!n_2!...n_k!}$ طریق می‌توانند در کنار هم قرار بگیرند. در فرمول فوق علت تـــقسیم‌ها حذف حالات اضافی بوجود آمده است.

واضح است که در این سوال پرچم‌های سفید، !٤ و پرچم‌های قرمز !۳ و پرچم‌های آبی !۲ حالت اضافی تولید می کنند که باید از حالات کل یعنی !۹ حذف شوند .

مثال۲۵: با استفاده از حروف P,E,P,P,E,R چند کلمه ٦ حرفی بدون توجه به معنی آن می‌توان ساخت؟

پاسخ: با توجه به اینکه ۳ حرف P و ۲ حرف E داریم و با استفاده از رابطه جایگشت با تکرار خواهیم داشت: کلمه $\frac{۶!}{۳!۲!}=۶۰$

مثال۲٦: به چند طریق می‌توان ٤کتاب ریاضی، ۳ کتاب فیزیک و ٥ کتاب ادبیات را در یک کتبخانه کنار هم چید؟

پاسخ: $\frac{۱۲!}{۴!۳!۵!}=۲۷۷۲۰$.

ترتیب:

هرگاه از n چیز متمایز، r تا برگزیده و از چپ به راست پهلوی هم بچینیم، آنرا یک ترتیب r تایی از n چیز می‌گویند و با $p_{(n,r)}$ نمایش می‌دهند و از رابطه زیر محاسبه می‌شود. $P_{(n,r)}=\frac{n!}{(n-r)!}$

مثال۲۷: از بین ۷ نفر اعضای هیات مدیره می‌خواهیم کمیته‌ای ۳ نفره تشکیل بدهیم که به ترتیب، اولی مدیر، دومی بعنوان معاون و سومی بعنوان منشی باشد. به چند طریق می‌توان این کار را انجام داد؟

پاسخ: چون ترتیب انتخاب در این مسئله اهمیت داری پس با مسئله ترتیب مواجه هستیم. در نتیجه طبق رابطه بالا داریم: $P_{(۷,۳)}=\frac{۷!}{۴!}=۲۱۰$.

ترکیب:

هر گاه از n چیز متمایز یک گروه r تایی را با هم یا یک به یک بدون توجه به ترتیب آنها برگزینیم، آن را یک ترکیب r تایی از این n چیز می‌گویند و آن را با $\binom{n}{r}$ یا C_n^r نمایش می‌دهند و به صورت زیر محاسبه می‌گردد. $C_n^r=\binom{n}{r}=\frac{n!}{r!(n-r)!}$

مثال۲۸: از بین ۷ نفر اعضای هیات مدیره می‌خواهیم کمیته‌ای ۳ نفره تشکیل بدهیم. به چند طریق می‌توان این کار را انجام داد؟

پاسخ: چنانچه از مسئله مشخص است ترتیب انتخاب اعضا اهمیتی ندارد بنابراین در این مسئله با ترکیب روبرو هستیم. در نتیجه خواهیم داشت: $\binom{۷}{۳}=\frac{۷!}{۳!۴!}=۳۵$.

مثال۲۹: از گروهی مرکب از ۵ مرد و ۷ زن چند تیم مختلف که از ۲ مرد و ۳ زن تشکیل شده باشد می‌توان تشکیل داد؟ اگر ۲ نفر از زنان با هم قهر باشند و از کار در یک تیم خودداری کنند، چند تیم می‌توان تشکیل داد؟

پاسخ: برای پاسخ به قسمت اول داریم که: تعداد راههای انتخاب ۲ مرد از بین ۵ مرد به صورت $\binom{۵}{۲}$ و تعداد راههای انتخاب ۳ زن از بین ۷ زن بصورت $\binom{۷}{۳}$ می‌باشد. پس طبق اصل ضرب تعداد کل گروه‌هایی که می‌توان تشکیل داد عبارتند از: $\binom{۵}{۲}\binom{۷}{۳}=\frac{۵!}{۲!۳!}\times\frac{۷!}{۳!۴!}=۳۵۰$. برای پاسخ به قسمت دوم یکبار گروه را بدون حضور این دو زن و یکبار تنها با حضور یکی از آنان تشکیل خواهیم داد. در نتیجه داریم: $\left[\binom{۲}{۰}\binom{۵}{۳}+\binom{۲}{۱}\binom{۵}{۲}\right]\binom{۵}{۲}=۳۰۰$.

مثال۳۰: از ۵ زمین کشاورزی که در ۳ تای آنها گندم و در ۲ تا جو کاشته شده است، ۲ زمین به تصادف انتخاب می‌کنیم. مطلوب است:

الف- احتمال اینکه در هر دو زمین گندم کاشته شده باشد؟

ب- احتمال اینکه در هر دو زمین جو کاشته شده باشد؟

ج- احتمال اینکه در یکی گندم و در یکی جو کاشته شده باشد؟

پاسخ: در مجموع ۵ زمین داریم و از این زمین‌ها ۲ تا را به تصادف انتخاب میکنیم چون ترتیب انتخاب زمین‌ها اهمیتی ندارد. فضای نمونه بر اساس اصل ترکیب دارای $n(S)=\binom{۵}{۲}=۱۰$ عضو است.

الف: احتمال اینکه در هر دو زمین گندم کاشته شده باشد بدین معنی است که در نمونه ۲ تایی انتخابی، از بین ۳ زمین گندم بایستی ۲ زمین را انتخاب کرده و از زمین‌های جو، هیچ زمینی انتخاب نشود. در نتیجه داریم: $$p(A)=\frac{n(A)}{n(S)}=\frac{\binom{۳}{۲}\binom{۲}{۰}}{\binom{۵}{۲}}=۰/۳$$

ب: احتمال اینکه در هر دو زمین جو کاشته شده باشد بدین معنی است که در نمونه ۲ تایی انتخابی، از بین ۲ زمین جو بایستی هر دو انتخاب شوند و از زمین‌های گندم، هیچ زمینی انتخاب نشود. در نتیجه داریم:

$$p(B)=\frac{n(B)}{n(S)}=\frac{\binom{۳}{۰}\binom{۲}{۲}}{\binom{۵}{۲}}=./۱.$$

ج: احتمال اینکه در یکی گندم و در یکی جو کاشته شده باشد بدین معنی است که در نمونه ۲ تایی انتخابی، از بین ۳ زمین گندم بایستی یک انتخا و از بین ۲ زمین جو نیز بایستی یکی انتخاب شوند. در نتیجه داریم:

$$p(C)=\frac{n(C)}{n(S)}=\frac{\binom{۳}{۱}\binom{۲}{۱}}{\binom{۵}{۲}}=./۶.$$

مثال۳۱: جنگلی ۲۰ گوزن دارد که از آنها ۵ تا گرفته شده و پس از علامت گذاری رها می‌شوند. بعد از مدتی ٤ گوزن از ۲۰ گوزن گرفته می‌شود. احتمال اینکه ۲ گوزن از ٤ گوزن گرفته شده علامت‌دار باشند چقدر است؟

پاسخ: کلاً ۲۰ گوزن داریم و از آنها نمونه چهار تایی استخراج می‌کنیم، پس تعداد اعضای فضای نمونه عبارت است از: $n(S)=\binom{۲۰}{۴}=۴۸۴۵$. اگر پیشامد A را وجود دو گوزن علامت دار در نمونه استخراجی تعریف کنیم بدین معنی است که از بین ۵ گوزن علامت دار بایستی دو گوزن و از مابقی گوزن‌ها نیز دو گوزن بایستی انتخاب شوند. در نتیجه داریم: $P(A)=\frac{n(A)}{n(S)}=\frac{\binom{۵}{۲}\binom{۱۵}{۲}}{\binom{۲۰}{۴}}=\frac{۱۰\times۱۰۵}{۴۸۴۵}=۰/۲۱۶$.

مثال۳۲: ظرفی حاوی ۱۰ عدد لامپ می‌باشد که ٤ لامپ آن معیوب است. ۲ لامپ به تصادف انتخاب می‌کنیم، مطلوب است محاسبه احتمال اینکه:

الف- هیچکدام از لامپ‌ها معیوب نباشند؟

ب- یکی از لامپ‌ها معیوب باشد؟

پاسخ: $n(S)=\binom{۱۰}{۲}=۴۵$

الف: اگر A را پیشامد اینکه معیوب بودن دو لامپ انتخاب شده تعریف کنیم داریم:

$$n(A)=\binom{۶}{۲}=۱۵$$

$$\Rightarrow P(A)=\frac{n(A)}{n(S)}=\frac{۱۵}{۴۵}=\frac{۱}{۳}$$

ب: اگر B را پیشامد اینکه یکی از دولامپ انتخابی معیوب باشد تعریف کنیم داریم:

$$n(B)=\binom{۶}{۱}\binom{۴}{۱}=۲۴$$
$$\Rightarrow P(B)=\frac{n(B)}{n(S)}=\frac{۲۴}{۴۵}=\frac{۸}{۱۵}$$

احتمال شرطی:

احتمال پیشامد A مشروط بر اینکه قبلاً پیشامد B رخ داده است را احتمال شرطی A به شرط B می‌نامیم و آن را با $P(A|B)$ نمایش می‌دهیم و به صورت زیر تعریف می‌کنیم.

$$\boxed{P(A|B)=\frac{P(A\cap B)}{P(B)}}$$

توجه: با توجه به تعریف احتمال می‌توان نوشت: $P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{n(A\cap B)}{n(S)}}{\frac{n(B)}{n(S)}}=\frac{n(A\cap B)}{n(B)}$

مثال۳۳: یک جفت تاس را باهم پرتاب می‌کنیم. اگر مجموع شماره‌های ظاهر شده ٦ باشد. احتمال اینکه شماره یکی از تاس‌ها ۲ باشد، چقدر است؟

پاسخ: A را پیشامد اینکه یکی از تاس‌ها ۲ باشد و B را پیشامد اینکه مجموع شماره‌ها ۲ باشد تعریف می‌کنیم

$$\begin{cases} A=\{(۱,۲),(۲,۲),(۳,۲),(۴,۲)(۵,۲),(۶,۲),(۲,۱),(۲,۳),(۲,۴),(۲,۵),(۲,۶)\} \\ B=\{(۱,۵),(۲,۴),(۳,۳),(۴,۲),(۵,۱)\} \\ A\cap B=\{(۲,۴),(۴,۲)\} \\ P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{۲/۳۶}{۵/۳۶}=./۴ \end{cases}$$

مثال٣٤: تاسی را ۲ بار پرتاب می‌کنیم. اگر بدانیم یکی از تاس‌ها عدد ٤ را نشان می‌دهد، مطلوبست احتمال اینکه تاس دیگر عدد ۳ را نشان دهد؟

پاسخ: A را پیشامد اینکه یکی از تاس‌ها عدد ٤ را نشان دهد و B را پیشامد اینکه تاس‌ها عدد ۳ را نشان دهد، تعریف می‌کنیم. در نتیجه داریم:

$$\begin{cases} B=\{(۱,۴),(۲,۴),(۳,۴),(۴,۴)(۵,۴),(۶,۴),(۴,۱),(۴,۲),(۴,۳),(۴,۵),۴,۶)\} \\ A=\{(۱,۳),(۲,۳),(۳,۳),(۴,۳)(۵,۳),(۶,۳),(۳,۱),(۳,۲),(۳,۴),(۳,۵),۳,۶)\} \\ A\bigcap B=\{(۳,۴),(۴,۳)\} \\ P(A|B)=\dfrac{P(A\bigcap B)}{P(B)}=\dfrac{۲/۳۶}{۱۱/۳۶}=\dfrac{۲}{۱۱} \end{cases}$$

قاعده ضرب احتمال: $P(A|B)=\dfrac{P(A\bigcap B)}{P(B)} \Rightarrow P(A\bigcap B)=P(B).P(A|B)$

این رابطه به قاعده ضرب احتمال موسوم است. به کمک این قاعده می‌توان احتمال رخداد هم زمان دو پیشامد را تعیین کرد.

استقلال دو پیشامد:

دو پیشامد A و B را مستقل گوییم هرگاه، وقوع یا عدم وقوع یکی از پیشامدها، در وقوع یا عدم وقوع دیگری تاثیری نداشته باشد. به زبان ریاضی داریم: $P(A|B)=P(A)$.
اگر A و B مستقل باشند داریم: $P(A\bigcap B)=P(B).P(A|B)=P(A).P(B)$.
در نتیجه:

دو پیشامد A و B مستقلند $\Leftrightarrow$ $P(A\bigcap B)=P(A).P(B)$

مثال۳۵: دو سکه سالم را پرتاب می‌کنیم. اگر A، پیشامد اینکه در سکه اول شیر و B، پیشامد اینکه در سکه دوم خط ظاهر شود، تعریف شوند. استقلال A و B را بررسی کنید؟
پاسخ:

$$\begin{cases} S=\{HH,HT,TH,TT\} \\ A=\{HH,HT\} \quad \Rightarrow P(A)=\dfrac{۱}{۲} \\ B=\{HT,TT\} \quad \Rightarrow P(B)=\dfrac{۱}{۲} \\ A\bigcap B=\{HT\} \quad \Rightarrow P(A\bigcap B)=\dfrac{۱}{۴} \end{cases} \Rightarrow P(A\bigcap B)=P(A).P(B)=\dfrac{۱}{۴}$$

در نتیجه A و B مستقلند

مثال۳۶: در پرتاب دو تاس، A را پیشامد اینکه مجموع تاس‌ها ٦ و B را پیشامد اینکه تاس اول ٤ باشد تعریف می‌کنیم. استقلال A و B را بررسی کنید؟

پاسخ:

$$\begin{cases} S=\{(۱,۱),(۱,۲),(۱,۳),.....,(۵,۶),(۶,۶)\} \qquad n(S)=۳۶ \\ A=\{(۱,۵),(۲,۴),(۳,۳),(۴,۲),(۵,۱)\} \quad \Rightarrow P(A)=\frac{۵}{۳۶} \\ B=\{(۴,۱),(۴,۲),(۴,۳),(۴,۴),(۴,۵),(۴,۶)\} \quad \Rightarrow P(B)=\frac{۶}{۳۶} \\ A\cap B=\{(۴,۲)\} \quad \Rightarrow P(A\cap B)=\frac{۱}{۳۶} \end{cases}$$

$$\Rightarrow \frac{۱}{۳۶}=P(A\cap B)\neq P(A).P(B)=\frac{۵}{۲۱۶}$$

در نتیجه A و B مستقل نیستند

مثال۳۷: احتمال آنکه تیر فرد الف به هدف اصابت کند برابر ۰/٤ و احتمال آنکه تیر فرد ب به هدف اصابت کند برابر ۰/۳ است. این دو نفر تیری به سمت هدف پرتاب می‌کنند. مطلوب است احتمال آنکه حداقل یک تیر به هدف اصابت کند؟

پاسخ: اگر A را پیشامد اینکه تیر فرد الف به هدف اصابت کند و Bرا پیشامد اینکه اینکه تیر فرد الف به هدف اصابت کند، تعریف کنیم. با توجه به استقلال این دو پیشامد داریم:

$$P\left(A\cup B\right)=P\left(A\right)+P\left(B\right)-P\left(A\cap B\right)$$
$$P\left(A\cap B\right)=P\left(A\right).P\left(B\right)=۰/۴\times ۰/۳=۰/۱۲$$
$$\Rightarrow P\left(A\cup B\right)=۰/۴+۰/۳-۰/۱۲=۰/۵۸$$

قضیه بیز:

فرض کنید n پیشامد $A_۱,A_۲,...,A_n$ به گونه‌ای باشند که داشته باشیم:

۱- $A_۱\cup A_۲\cup ...\cup A_n=S$

۲- پیشامدها دو به دو ناسازگار باشند، یعنی: $A_i\cap A_j=\varnothing \qquad \forall i,j=۱,۲,...,n$

در این صورت برای هر پیشامد $B\subset S$ داریم:

قضیه اول بیز
$$P(B)=\sum_{i=۱}^{n}P(A_i)P(B\,|\,A_i)$$

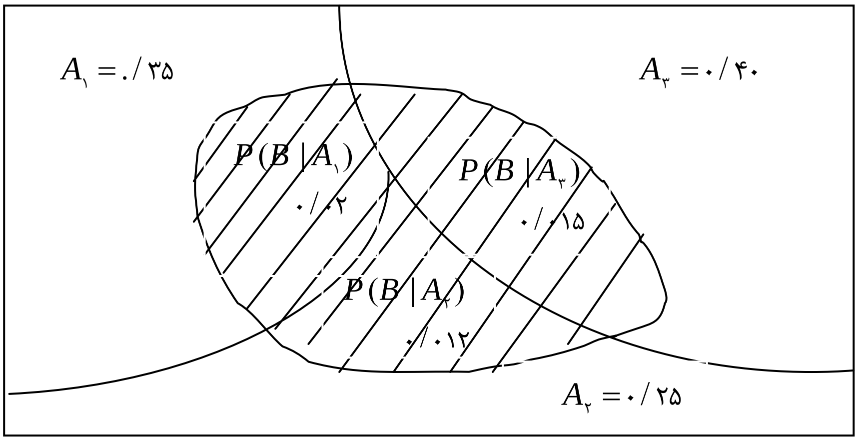
قضیه دوم بیز
$$P(A_i\,|\,B)=\frac{P(A_i)P(B\,|\,A_i)}{\sum_{i=۱}^{n}P(A_i)P(B\,|\,A_i)}$$

مثال۳۸: فرض کنید در یک کارخانه تولیدی، ماشین A_1، ۳۵ درصد و ماشین A_2، ۲۵ درصد و ماشین A_3، ۴۰ درصد از کل تولیدات کارخانه را تولید می‌کنند. همچنین فرض کنید که ماشین A_1، ۲ درصد، ماشین A_2، ۱/۲ درصد و ماشین A_3، ۱/۵ درصد از تولیداتشان خراب است. قطعه‌ای به تصادف از بین تولیدات کارخانه انتخاب می‌کنیم. مطلوب است:

الف- احتمال معیوب بودن قطعه انتخابی؟

ب- اگر قطعه انتخابی معیوب باشد با چه احتمالی متعلق به ماشین A_2 است؟

پاسخ: برای حل اینگونه مسائل بیزی ساده‌ترین راه استفاده از نمودار است. اگر A_i را پیشامد اینکه کالای انتخاب شده به ماشین iام تعلق داشته باشد و B را پیشامد خراب بودن تولیدات تعریف کنیم و در شکل زیر با قسمت هاشور خورده نمایش دهیم، آنگاه داریم:



الف:

$$P(B) = \sum_{i=1}^{3} P(A_i)P(B|A_i)$$
$$= ۰/۳۵ \times ۰/۰۲ + ۰/۲۵ \times ۰/۰۱۲ + ۰/۴۰ \times ۰/۰۱۵$$
$$= ۰/۰۱۶$$

ب:

$$P(A_2|B) = \frac{P(A_2)P(A_2|B)}{P(B)}$$
$$= \frac{./۲۵ \times ./.۱۲}{./.۱۶} = ./۱۸۷۵$$

مسائل

۱- جعبه‌ای دارای ۵ مهره قرمز، ٦ مهره آبی و ۸ مهره سبز است. اگر ۳ مهره به تصادف انتخاب کنیم، مطلوب است محاسبه احتمالات زیر:

الف- هر سه مهره همرنگ باشند؟

ب- هر مهره از یک رنگ باشد؟

ج- دو تا قرمز و یکی سبز باشد؟

۲- از گروهی شامل ۳ دانش‌آموز دبستانی، ٤ دانش‌آموز راهنمایی، ٤ دبیرستانی و ۳ دانشگاهی یک گروه ٤ نفری به تصادف انتخاب می‌شود. مطلوب است محاسبه احتمال اینکه این گروه شامل:

الف- یک نفر از هر کلاس باشد؟

ب- ۲ دانش‌آموز راهنمایی و ۲ دانش‌آموز دبیرستانی باشد؟

ج- فقط دانش‌آموزان راهنمایی و دبیرستانی باشند؟

۳- کیسه‌ای شامل ۸ گلوله قرمز و ٤ گلوله سفید است. دو گلوله بدون جایگذاری از کیسه بیرون می-اوریم. اگر در هر استخراج احتمال استخراج هر گلوله در کیسه برابر باشد. احتمال اینکه هر دو گلوله استخراج شده قرمز باشند چقدر است؟

٤- شرکت بیمه‌ای بر این باور است که افراد را می‌توان به دو گروه تقسیم کرد: گروهی که مستعد تصادفند و گروهی که نیستند. آمارهی این شرکت نشان می‌دهد که یک فرد مستعد تصادف با احتمال ۰/٤ تصادفی در یک دوره یکساله خواهد داشت. در صورتیکه این احتمال برای فرد فاقد این استعداد به ۰/۲ کاهش می‌یابد. اگر فرض کنیم که ۳۰ درصد جامعه‌ای مستعد تصادف هستند.

الف- احتمال اینکه یک بیمه گذار جدید در ظرف یک سال یک تصادف داشته باشد، چقدر است؟

ب- اگر بیمه‌گذار جدیدی در مدت یک سال تصادف داشته باشد، احتمال اینکه این فرد مستعد تصادف باشد، چقدر است؟

۵- اگر از کیسه‌ای شامل ۱۲ مهره شماره‌گذاری شده از ۱ تا ۱۲ یک مهره به تصادف استخراج کنیم. احتمال اینکه مهره خارج شده مضرب ۲ یا ۳ باشد، چقدر است؟

٦- از کیسه‌ای حاوی ۵ مهره سفید، ۳ مهره آبی و ۲ مهره زرد، ۲ مهره به تصادف انتخاب می‌کنیم. مطلوب است محاسبه احتمالات زیر:

الف-ا مهره آبی و یک مهره زرد باشد؟

ب- هر دو مهره همرنگ باشند؟

ج- هر دو مهره همرنگ نباشند؟

۷- از کیسه‌ای که ٦ مهره سفید و ٤ مهره سیاه دارد، ٤ مهره به تصادف و بدون جایگذاری خارج می‌کنیم.

الف- احتمال اینکه رنگه مهره‌های خارج شده به تناوب سفید و سیاه باشد؟

ب- احتمال اینکه ٤ مهره استخراجی هم رنگ نباشند؟(راهنمایی: از احتمال متمم حل کنید)

۸- سه پایانه $A_۱$، $A_۲$ و $A_۳$ در بخش رایانه کتابخانه یک دانشگاه وجود دارند. تجربیات قبلی نشان می‌دهد که حدوداً ٤٥ درصد از مراجعه کنندگان از پایانه $A_۱$، ٤٠ درصد از پایانه $A_۲$ و ۱۵ درصد از پایانه $A_۳$استفاده می‌کنند. با استفاده از اطلاعات پرسش‌نامه‌ای مشخص گردید که حدود ۱۲ درصد از مراجعان، از پاسخ کنندگان از پایانه $A_۱$، ٤ درصد از پایانه $A_۲$ و ۱۱ درصد از پایانه $A_۳$ ناراضی بوده‌اند. یکی از پرسش‌نامه‌ها را به تصادف انتخاب می‌کنیم. مطلوب است محاسبه احتمالات زیر:

الف- با چه احتمالی مراجعه کننده ناراضی بوده است؟

ب- اگر مراجعه کننده ناراضی بوده باشد، با چه احتمالی این نارضایتی از پایانه $A_۱$ بوده است؟

۹- یک جفت تاس را با هم پرتاب می‌کنیم. اگر مجموع تاس‌های ظاهر شده ۷ باشد، احتمال ۲ بودن شماره یکی از تاس‌ها را به دست آورید؟

۱۰- در جعبه‌ای ۲۷ لامپ موجود است که ٦ عدد از آنها سوخته‌اند، ۲ لامپ را به طور تصادفی از جعبه خارج می‌کنیم، احتمال سوخته بودن هر دو لامپ چقدر است؟

۱۱- از میان ۱۵۰ کارت کتاب، ۱۳۵ کارت درست تایپ شده‌اند. ٤ کارت را به تصادف انتخاب می‌کنیم. احتمال اینکه همه آنها اشتباه تایپ شده باشند را به دست آورید؟

۱۲- اگر انحراف استاندارد حقوق کارمندان ٤۳ و ضریب تغییرپذیری حقوق کارمندان ۲/۳۹ باشد، میانگین حقوق آنان را به دست آورید؟

۱۳- اگر $S=\{۱,۲,۳,۴,۵,۶,۷,۸,۹\}$ و $A=\{۱,۲,۳,۴\}$ و $B=\{۲,۴,۶,۸\}$ باشد، $P\left(A\cup B\right)$ و $P\left(B'\right)$ را محاسبه کنید؟

١٤- یک تاس سالم را دو بار پرتاب می‌کنیم. احتمال اینکه مجموع شماره‌ها ۹ باشد را به دست آورید؟

۱۵- در پرتاب دو تاس اگر بدانیم که مجموع شماره‌ها٦ است، احتمال اینکه هر دو شماره کمتر از ٤ باشد، چقدر است؟

۱٦- به چند طریق می‌توان یک نفر فرماندار، یک نفر شهردار و یک نفر بخشدار را از بین ۵ نفر انتخاب کرد؟

۱۷- با انجام یک آزمایش تصادفی، تنها یکی از پیشامدهای A، B یا C رخ می‌دهد. اگر $P(A)=2P(B)=3P(C)$ باشد، آنگاه احتمال‌های A، B و C را به دست آورید؟

۱۸- به طور تصادفی از بین اسامی ۵ دانشجوی پسر و ٤ دانشجوی دختر:

الف- ۳ اسم با جایگذاری بر می‌داریم. احتمال اینکه هر سه نفر پسر باشند چقدر است؟

ب- ۳ اسم یکجا برمی‌داریم. احتمال اینکه دو پسر انتخاب شود، چقدر است؟

ج- ۳ اسم یکی پس از دیگری و بدون جایگذاری برمیداریم. احتمال اینکه دو نفر اول دختر و نفر سوم پسر باشد، چقدر است؟

۱۹- از یک جمعیت ۱۰۰ نفری اطلاعات زیر در دست است. یک نفر به تصادف از این جمعیت انتخاب می‌کنیم:

الف- اگر بدانیم مرد است، احتمال اینکه ترک کرده باشد؟

ب- احتمال اینکه یا زن باشد یا غیرسیگاری باشد؟

	مرد	زن
سیگاری	۲۵	۵
غیر سیگاری	۳۵	۱۵
ترک کرده	۱۸	۲

فصل پنجم

متغیرهای تصادفی

و تابع احتمال

در اغلب آزمایش‌هایی که انجام می‌شود به تابعی از برآمد حاصل در مقبل خود برآمد علاقه‌مندیم. مثلاً در پرتاب دو تاس اغلب می‌خواهیم مجموع دو تاس را بدانیم تا مقادیر هریک از آنها را. به طور مثال ممکن است فقط مجموع ۷ مورد توجه ما باشد به برآمدهای (۱،٦)، (٦،۱)، (۲،۵)، (۵،۲)، (٤،۳)، (۳،٤). و یا در پرتاب سکه ممکن است تعداد شیرها مدنظر باشد نه دنباله شیر و خط‌های حاصل از آزمایش.

متغیر تصادفی: متغیر تصادفی تابعی است که به هر نقطه از فضای نمونه یک عدد حقیقی نسبت می‌دهد، پس متغیر تصادفی‌ای مانند X، تابعی است از فضای نمونه به مجموعه اعداد حقیقی: $X: S \to R$.

متغیرهای تصادفی معمولاً با حروف بزرگ مانند X, Y, Z,...نمایش داده می‌شود و مقادیر آن با حروف کوچک نظیر نمایش داده می‌شود.

مثال۱: دو سکه را همزمان پرتاب می‌کنیم. می‌دانیم $S = \{HH, TH, HT, TT\}$. با تعریف متغیر تصادفی X بعنوان تعداد شیرها در آزمایش، به ترتیب مقادیر ۰، ۱، ۱، ۲ را به اعضای نمونه نسبت می‌دهد.

متغیرهای تصادفی بر دو نوع هستند:

متغیر تصادفی گسسته: متغیری را گویند که جمیع مقادیر ممکن آن، مجموعه محدود یانامحدود شمارش پذیر تشکیل دهد و برای هر یک از این متغیرها، نظیر به نظیر یک احتمال مشخص وجود داشته باشد.

متغیر تصادفی پیوسته: متغیری را گویند که جمیع مقادیر ممکن آن، مجموعه تمامی اعداد حقیقی یا فاصله‌هایی از مجموعه اعداد حقیقی باشد، یا به گفته بهتر مجموعه نظیر آن شمارش ناپذیر باشد.

تابع احتمال

در عمل نه تنها باید بدانیم که متغیری تصادفی مانند X، چه مقادیری را اختیار می‌کند بلکه بایستی بدانیم که این مقادیر را با چه احتمالی اختیار می‌کند.

به تابعی که بتوان با استفاده از آن احتمال هر یک از مقادیر ممکن متغیر تصادفی را مشخص کرد تابع احتمال یا توزیع احتمال گفته می شود. در مورد متغیرهای گسسته و پیوسته تابع احتمال به طور متفاوت تعریف می‌شود. ما در این کتاب تنها به بررسی متغیرهای تصادفی گسسته خواهیم پرداخت.

تابع احتمال گسسته:

یک متغیر تصادفی گسسته هر یک از مقادیر خود را با احتمالی معین اختیار می‌کند. حال، جدول یا فرمولی که تمام مقادیر متغیر تصادفی را همراه با احتمال‌های مربوطه نشان دهد تابع احتمال نامیده و آن را با $f(x), g(x), ...$ نشان می‌دهیم. در حالتی که متغیر تصادفی گسسته باشد، تابع احتمال را تابع توزیع احتمال می‌نامیم و به صورت $f(x_i) = P(X = x_i)$ تعریف می‌شود.

یک تابع توزیع احتمال بایستی در دو شرط زیر صدق کند:

۱- $\forall x_i \quad P(X = x_i) \geq 0$

۲- $\sum_{i=1}^{n} P(X = x_i) = 1$.

مثال۲: آیا تابع $x = 0,1,2,3 \quad f(x) = \frac{x-1}{5}$، یک تابع توزیع احتمال است؟

پاسخ: یک تابع توزیع احتمال نیت. زیرا برای مقداری $x = 0$ داریم: $f(0) = \frac{-1}{5}$. که در شرط اول صدق نمی‌کند.

مثال۳: سکه‌ای را سه مرتبه پرتاب می‌کنیم، اگر متغیر تصادفی Xرا تعداد شیرها در این سه پرتاب تعریف کنیم، تابع توزیع احتمال X را به دست آورید؟

پاسخ: می‌دانیم که در این آزمایش $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$، تابع احتمال X را به صورت زیر داریم:

X	۰	۱	۲	۳
$f(x) = P(X = x)$	$1/8$	$3/8$	$3/8$	$1/8$

توجه داشته باشید که در تابع احتمال بالا هر دو شرط تابع احتمال برقرار است.

مثال٤: مقدار k را چنان بیابید که جدول زیر یک تابع توزیع احتمال باشد.

X	۱	۲	۳	٤
$f(x) = P(X = x)$	$0/2$	$2k-1$	$0/3$	$0/3$

پاسخ: $\sum f(x_i) = 1 \quad \Rightarrow 0/2 + 2k - 1 + 0/3 + 0/3 = 1 \quad \Rightarrow k = 0/6$

تابع احتمال توام:

تاحالا با فضاهای نمونه یک بعدی و متغیرهای تصادفی مربوط به این فضاها آشنا شدیم، ولی آزمایش‌های زیادی وجود دارند که به طور همزمان دو یا چندین نتیجه خواهند داشت و یا اینکه می‌توان روی یک فضای نمونه متغیرهای متفاوتی تعریف کرد. برای مثال:

۱- در پرتاب دو تاس، X تعداد تاس های زوج آمده و Y مجموع دو تاس.
۲- در یک خانواده ٤ فرزندی، X تعداد فرزندان پسر و Y تعداد فرزندان دختر.
۳- در یک کلاس آمار، X ضریب هوشی و Y معدل آنها.

اگر بخواهیم رفتار چند متغیر را به طور همزمان بررسی کنیم، باید تابع احتمالی در دست باشد که احتمالات رخ دادن همزمان چندین متغیر را تولید کند. به چنین تابعی، تابع احتمال توام می‌گویند.

اگر X و Y دو متغیر تصادفی باشند، تابع احتمال توام آنها را علامت تابعی $f(x,y)$ نمایش می‌دهند. در حالتی که X و Y هر دو گسسته باشند، $f(x,y)$ را تابع توزیع توام می‌نامند.

خصوصیات تابع توزیع احتمال توام

$$\begin{aligned}
&۱)\ f(x,y)\geq ۰\\
&۲)\ \sum_x\sum_y f(x,y)=۱\\
&۳)\ f(x,y)=P\left(X=x\ ,Y=y\right)
\end{aligned}$$

حال این سوال مطرح است که اگر توزیع احتمال توام دو یا چند متغیر تصادفی معلوم باشد، آیا می توان توزیع هر کدام از این متغیرها را به صورت جداگانه حساب کرد؟ جواب این سوال همواره مثبت است. بدین معنی که با در دست داشتن تابع احتمال توام، همواره می‌توان تابع احتمال هر یک از متغیرهای تصادفی را پیدا کرد.

توزیع احتمال حاشیه‌ای:

اگر X و Y دو متغیر تصادفی با تابع احتمال توام $f(x,y)$ باشند، آنگاه تابع احتمال هر یک از متغیرهای تصادفی X و Y تابع احتمال حاشیه‌ای نامیده شده و به صورت زیر محاسبه می شوند:

$$\begin{cases}
f(x)=\sum_y f(x,y)\\
f(y)=\sum_x f(x,y)
\end{cases}$$

مثال۵: در تابع داده شده در زیر، مطلوب است:

الف- m را چنان بیابید که تابع مقابل یک تابع احتمال توام باشد.

ب- توزیع توام متغیرهای تصادفی X و Y را در قالب یک جدول نشان دهید.

ج- $P(X \geq ۱, Y = ۱)$ را به دست آورید.

د- توابع احتمال حاشیه‌ای X و Y را معلوم کنید.

پاسخ:

الف: در یک تابع احتمال توام بایستی مجموع احتمالات برابر یک باشد، پس داریم:

$$\sum_{y=۱}^{۲}\sum_{x=۰}^{۲} f(x,y) = f(۰,۱)+f(۰,۲)+f(۱,۱)+f(۱,۲)+f(۲,۱)+f(۲,۲) =$$

$$= m + \frac{m}{۲} + ۲m + m + ۳m + \frac{۳m}{۲} = ۱ \quad \Rightarrow m = \frac{۱}{۹}$$

اگر $m = \frac{۱}{۹}$ را در رابطه بالا قرار دهیم، تابع احتمال توام به صورت روبرو خواهد شد:

$$\begin{cases} f(x,y) = \dfrac{x+۱}{۹y} \\ X = ۰,۱,۲ \\ Y = ۱,۲ \end{cases}$$

ب:

Y \ X	۰	۱	۲
۱	$\frac{۱}{۹}$	$\frac{۲}{۹}$	$\frac{۳}{۹}$
۲	$\frac{۱}{۱۸}$	$\frac{۲}{۱۸}$	$\frac{۳}{۱۸}$

ج:

$$P(X \geq ۱, Y = ۱) = f(۱,۱) + f(۲,۱) = \frac{۲}{۹} + \frac{۳}{۹} = \frac{۵}{۹}$$

د:

Y	۱	۲
$f(Y)$	$\frac{۱۲}{۱۸}$	$\frac{۶}{۱۸}$

X	۰	۱	۲
$f(X)$	$\frac{۳}{۱۸}$	$\frac{۶}{۱۸}$	$\frac{۹}{۱۸}$

مثال۶: شیشه ای حاوی ۳ قرص آسپرین، ۲ قرص خواب آور و ۲ قرص مسکن مفروض است. شخصی به تصادف دو قرص از این ظرف خارج می‌کند. اگر فرض کنیم X، معرف تعداد قرص‌های آسپرین و Y، معرف تعداد قرص خواب آور باشد. مطلوب است:

الف- توزیع احتمال توام برای دو متغیر تصادفی X و Y ؟

ب- توزیع های حاشیه ای X و Y ؟

پاسخ:

نمایش تابع توزیع احتمال به صورت جدول زیر خواهد بود:

$Y \backslash X$	۰	۱	۲	$f(y)$
۰	$\frac{1}{21}$	$\frac{6}{21}$	$\frac{3}{21}$	$\frac{10}{21}$
۱	$\frac{4}{21}$	$\frac{6}{21}$	۰	$\frac{10}{21}$
۲	$\frac{1}{21}$	۰	۰	$\frac{1}{21}$
$f(x)$	$\frac{6}{21}$	$\frac{12}{21}$	$\frac{3}{21}$	۱

برای محاسبه هر یک از مقادیر این جدول، از ترکیب استفاده می‌کنیم. برای مثال:

$$\begin{cases} f(0,0) = \dfrac{\binom{3}{0}\binom{2}{0}\binom{2}{2}}{\binom{7}{2}} = \dfrac{1}{21} \\ f(2,0) = \dfrac{\binom{3}{2}\binom{2}{0}\binom{2}{0}}{\binom{7}{2}} = \dfrac{3}{21} \end{cases}$$

توجه داشته باشید که برای $f(2,1)$ یا $f(2,2)$، چون مجموع دو مقدار بیشتر از تعداد نمونه استخراجی یعنی ۲ است. پس یک پیشامد غیرممکن می‌باشد، که بنا بر تعریف، احتمال آن برابر صفر است.

مثال۷: توزیع احتمال توام زیر را در نظر بگیرید. مطلوب است محاسبه:

$Y \backslash X$	۱	۳	۵
۲	۰/۱	۰/۲	۰/۱
۴	۰/۱۵	۰/۳	۰/۱۵

الف- $P(X=2)$.

ب- $P(X>Y)$.

ج- $P(x=4, Y\geq 3)$.

پاسخ:

الف:

$$P(X=۲)=P(x=۲,y=۱)+P(x=۲,y=۳)+P(x=۲,y=۵)$$
$$=۰/۱+۰/۲+۰/۱=۰/۴$$

ب: $P(X>Y)=P(۲,۱)+P(۴,۱)+P(۴,۳)=۰/۱+۰/۱۵+۰/۳=۰/۵۵$

ج: $P(X=۴,Y\geq ۳)=P(۴,۳)+P(۴,۵)=۰/۳+۰/۱۵=۰/۴۵$

مثال۸: توزیع احتمال توام زیر را در نظر بگیرید. مطلوب است محاسبه:

الف- $P(X=۱)$

ب- $P(X>۱)$

ج- $P(Y<۲)$

د- $P(X<Y)$.

X \ Y	۱	۲	۳
۰	۱/۱۵	۳/۱۵	۲/۱۵
۱	۴/۱۵	۲/۱۵	۳/۱۵

پاسخ:

الف: $P(X=۱)=\sum_{y}f(x=۱,Y=y)=f(۱,۱)+f(۱,۲)+f(۱,۳)=\frac{۴}{۱۵}+\frac{۲}{۱۵}+\frac{۳}{۱۵}=\frac{۹}{۱۵}$

ب: $P(X>۱)=p(\varnothing)=۰$

ج: $P(Y<۲)=P(Y=۱)=\sum_{x}f(X=x,۱)=f(۰,۱)+f(۱,۱)=\frac{۱}{۱۵}+\frac{۴}{۱۵}=\frac{۵}{۱۵}$

د:

$$P(X<Y)=f(۰,۱)+f(۰,۲)+f(۰,۳)+f(۱,۲)+f(۱,۳)$$
$$=\frac{۱}{۱۵}+\frac{۳}{۱۵}+\frac{۲}{۱۵}+\frac{۲}{۱۵}+\frac{۳}{۱۵}=\frac{۱۱}{۱۵}$$

تابع احتمال شرطی توام:

اگر X و Y دو متغیر تصادفی با تابع احتمال توام $f(x,y)$ و توزیع‌های حاشیه‌ای $f(x)$ و $f(y)$ باشند، تابع احتمال شرطی متغیر تصادفی X در صورتیکه $Y=y$ داده شده باشد، احتمال شرطی X به شرط $Y=y$ خوانده شده و به صورت $f(X|Y=y)$ نشان داده می‌شود و عبارت است از:

$$f(X|Y=y)=\frac{f(x,y)}{f(y)}\quad f(y)>۰.$$

به همین ترتیب تابع احتمال شرطی متغیر تصادفی Y در صورتیکه $X=x$ داده شده باشد، احتمال شرطی Y به شرط $X=x$ خوانده شده و به صورت $f(Y\mid X=x)$ نشان داده می‌شود و عبارت است از:

$$f(x)>۰ \quad f(Y\mid X=x)=\frac{f(x,y)}{f(x)}.$$

مثال۹: برای تابع توزیع توام داده شده در زیر، مطلوب است محاسبه:

الف- $f\left(X\mid y=۱\right)$.

ب- $f\left(Y\mid x=۲\right)$

ج- $P\left(X\le ۱\right)$

X \ Y	-۱	۰	۱
۰	۰/۱	۰/۲	۰/۱
۱	۰	۰/۱	۰/۱
۲	۰/۱	۰	۰/۳

الف:

$$f\left(X\mid y=۱\right)=\frac{f\left(X,۱\right)}{f\left(y=۱\right)}=\frac{f\left(X,۱\right)}{۰/۵}$$

$$x=۰ \quad f\left(x=۰\mid y=۱\right)=\frac{f\left(۰,۱\right)}{f\left(y=۱\right)}=\frac{۰/۱}{۰/۵}=\frac{۱}{۵}$$

$$x=۱ \quad f\left(x=۱\mid y=۱\right)=\frac{f\left(۱,۱\right)}{f\left(y=۱\right)}=\frac{۰/۱}{۰/۵}=\frac{۱}{۵}$$

$$x=۲ \quad f\left(x=۲\mid y=۱\right)=\frac{f\left(۲,۱\right)}{f\left(y=۱\right)}=\frac{۰/۳}{۰/۵}=\frac{۳}{۵}$$

X	۰	۱	۲
$f\left(X\mid y=۱\right)$	$\frac{۱}{۵}$	$\frac{۱}{۵}$	$\frac{۳}{۵}$

ب:

$$f\left(Y\mid x=۲\right)=\frac{f\left(۲,Y\right)}{f\left(x=۲\right)}=\frac{f\left(۲,Y\right)}{۰/۴}$$

$$y=-۱ \quad f\left(y=-۱\mid x=۲\right)=\frac{f\left(۲,-۱\right)}{f\left(x=۲\right)}=\frac{۰/۱}{۰/۴}=\frac{۱}{۴}$$

$$y=۰ \quad f\left(y=۰\mid x=۲\right)=\frac{f\left(۲,۰\right)}{f\left(x=۲\right)}=\frac{۰}{۰/۴}=۰$$

$$y=۱ \quad f\left(y=۱\mid x=۲\right)=\frac{f\left(۲,۱\right)}{f\left(x=۲\right)}=\frac{۰/۳}{۰/۴}=\frac{۳}{۴}$$

Y	۰	۱	۲
$f\left(Y\mid x=۲\right)$	$\frac{۱}{۴}$	۰	$\frac{۳}{۴}$

ج:

$$P\left(X\le ۱\right)=P\left(x=۰\right)+P\left(x=۱\right)=۰/۴+۰/۲=۰/۶$$

استقلال دو متغیر تصادفی:

اگر X و Y دو متغیر تصادفی با تابع احتمال توام $f(x,y)$ و توزیع‌های حاشیه‌ای $f(x)$ و $f(y)$ باشند، آنگاه X و Y را از لحاظ آماری مستقل گوییم اگر و تنها اگر برای هر (x,y) داشته باشیم: $f(x,y)=f(x).f(y)$

مثال۱۰: تابع توزیع توام زیر داده شده است، استقلال X و Y را بررسی کنید؟

Y \ X	۱	۲
۱−	$\frac{۱}{۴}$	$\frac{۱}{۴}$
۰	$\frac{۱}{۴}$	$\frac{۱}{۴}$

پاسخ: برای بررسی استقلال دو متغیر تصادفی ابتدا بایستی توزیع حاشیه‌ای آنها را به دست آوریم.

Y \ X	۱	۲	$f(Y)$
۱−	$\frac{۱}{۴}$	$\frac{۱}{۴}$	$\frac{۲}{۴}$
۰	$\frac{۱}{۴}$	$\frac{۱}{۴}$	$\frac{۲}{۴}$
$f(X)$	$\frac{۲}{۴}$	$\frac{۲}{۴}$	۱

حال بایستی درستی رابطه $f(x,y)=f(x)f(y)$ را برای تمامی مقادیر موجود بررسی کنیم.

$$\begin{cases} f(-۱,۱)=f(-۱)f(۱) \quad \Rightarrow \frac{۱}{۴}=\frac{۲}{۴}\times\frac{۲}{۴} \\ f(-۱,۲)=f(-۱)f(۲) \quad \Rightarrow \frac{۱}{۴}=\frac{۲}{۴}\times\frac{۲}{۴} \\ f(۰,۱)=f(۰)f(۱) \quad \Rightarrow \frac{۱}{۴}=\frac{۲}{۴}\times\frac{۲}{۴} \\ f(۰,۲)=f(۰)f(۲) \quad \Rightarrow \frac{۱}{۴}=\frac{۲}{۴}\times\frac{۲}{۴} \end{cases}$$

چون این رابطه برای تمامی مقادیر برقرار است در نتیجه X و Y مستقل‌اند.

مثال۱۱: تابع توزیع توام زیر داده شده است، استقلال X و Y را بررسی کنید؟

$Y \backslash X$	-۱	۰	۱	$f(Y)$
۰	$\frac{۱}{۳}$	۰	$\frac{۱}{۳}$	$\frac{۲}{۳}$
۱	۰	$\frac{۱}{۳}$	۰	$\frac{۱}{۳}$
$f(X)$	$\frac{۱}{۳}$	$\frac{۱}{۳}$	$\frac{۱}{۳}$	۱

پاسخ: برای اسنکه دو متغیر X و Y مستقل باشند بایستی رابطه $f(x,y)=f(x)f(y)$ برای تمامی مقادیری ممکن برقرار باشد. ولی در تابع چگالی داده شده این رابطه برای تمامی مقادیر برقرار نیست. در نتیجه X و Y مستقل نیستند. بعنوان مثال برای نقطه $(۰,۰)$ داریم: $f(۰,۰)\neq f_x(۰)f_y(۰) \Rightarrow ۰\neq\frac{۱}{۳}\times\frac{۲}{۳}$

مثال۱۲: جدول توزیع احتمال داده شده زیر را در نظر بگیرید.

$Y \backslash X$	۰	۲	٤
۱	۰/۱	۰/۳	۰
۳	۰	۰/۲	۰/۱
۵	۰/۱	۰	۰/۲

الف- آیا X و Y مستق‌اند؟

ب- $f\left(X \mid y=۲\right)$.

ج-$P\left(X=۳\right)$.

د-$P\left(X>Y\right)$.

الف: X و Y مستقل نیستند، زیرا: $f\left(۱,۰\right)=۰/۱\neq f\left(۱\right).f\left(۰\right)=۰/۴\times۰/۲=۰/۸$

ب:

$$f\left(X \mid y=۲\right)=\frac{f\left(X,۲\right)}{f\left(y=۲\right)}=\frac{f\left(X,۲\right)}{۰/۵}$$

$$x=۱ \quad f\left(x=۱ \mid y=۲\right)=\frac{f\left(۱,۲\right)}{f\left(y=۲\right)}=\frac{۰/۳}{۰/۵}=\frac{۳}{۵}$$

$$x=۳ \quad f\left(x=۳ \mid y=۲\right)=\frac{f\left(۳,۲\right)}{f\left(y=۲\right)}=\frac{۰/۲}{۰/۵}=\frac{۲}{۵}$$

$$x=۵ \quad f\left(x=۵ \mid y=۲\right)=\frac{f\left(۵,۲\right)}{f\left(y=۲\right)}=\frac{۰}{۰/۵}=۰$$

X	۱	۳	۵
$f\left(X \mid y=۲\right)$	$\frac{۳}{۵}$	$\frac{۲}{۵}$	۰

ج:

$$P\left(x=۳\right)=\sum_y f\left(x=۳,Y=y\right)=f\left(۳,۰\right)+f\left(۳,۲\right)+f\left(۳,۴\right)=۰+۰/۲+۰/۱=۰/۳$$

د:

$$P\left(X>Y\right)=f\left(۱,۰\right)+f\left(۳,۰\right)+f\left(۳,۲\right)+f\left(۵,۰\right)+f\left(۵,۲\right)+f\left(۵,۴\right)=$$
$$=۰/۱+۰+۰/۲+۰/۱+۰+۰/۲=۰/۷$$

تابع توزیع تجمعی:

تابع توزیع تجمعی تابعی است صعودی و یکنواخت که برد آن بازه [۰،۱] بوده و احتمال آنکه متغیر تصادفی X دارای مقداری کوچک‌تر از X باشد را نشان می‌دهد، یعنی: $F(x) = P(X \le x)$.

اگر X یک متغیر تصادفی گسسته با تابع توزیع احتمال $f(x)$ باشد آنگاه: $F(x) = P(X \le x) = \sum_{t<x} f(t)$

توجه: تابع توزیع تجمعی، F، تابعی صعودی و یکنواخت خواهد بود.

مثال۱۳: شرکت بیمه‌ای اصلاعات مربوط به تعداد تصادفات در ۵۰ روز یکی از مناطق شهری را به صورت زیر جمع‌آوری کرده است:

الف- تابع احتمال را تعیین کنید؟

ب- تابع توزیع را تعیین کنید؟

ج- احتمال اینکه در یک روز کمتر از سه تصادف رخ داده باشد، چقدر است؟

تعداد روزها	تعداد تصادفات
۱۰	۰
۲۰	۱
۸	۲
۷	۳
۵	٤
۵۰	جمع

پاسخ: اگر متغیر تصادفی X را تعداد تصادفات تعریف کنیم، داریم:

الف:

X	۰	۱	۲	۳	٤
$P(X = x)$	۰/۲	۰/٤	۰/۱٦	۰/۱٤	۰/۱

ب:

X	۰	۱	۲	۳	٤
$F(X) = P(X \le x)$	۰/۲	۰/٦	۰/۷٦	۰/۹۰	۱

پ: $P(X < ۳) = P(X \le ۲) = P(X = ۲) + P(X = ۱) + P(X = ۰) = ۰/۲ + ۰/۴ + ۰/۱۶ = ۰/۷۶$

خواص تابع توزیع تجمعی:

۱- تابع توزیع تجمعی کمیتی غیر منفی است، یعنی: $F(X) \ge ۰$

۲- تابع توزیع تجمعی، تابعی غیر نزولی است. یعنی برای هر $a < b$ داریم: $F(a) \le F(b)$

۳- همواره:
$$\begin{cases} Lim_{x \to +\infty} F(X) = F(+\infty) = ۱ \\ Lim_{x \to -\infty} F(X) = F(-\infty) = ۰ \end{cases}$$

۴-
$$\begin{cases} P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a) \\ P(X = a) = F(a) - F(a^-) \end{cases}$$

۵- تابع $F(X)$ در تمام نقاط از سمت راست پیوسته است.

مثال۱۴: با توجه به جدول احتمال مقابل مطلوب است $F(\sqrt{۵})$؟

X	-۱	۰	۱	۲	٤
$f(x)$	۰/۱	۰/۲	۰/۱	k	۰/٤

پاسخ: ابتدا باید مقدار مجهول k را پیدا کنیم، از خاصیت $\sum f(x) = ۱$ استفاده می‌کنیم.

$$\sum f(x) = ۰/۱ + ۰/۲ + ۰/۱ + k + ۰/۴ = ۱ \quad \Rightarrow k = ۰/۲$$
$$F(\sqrt{۵}) = \sum_{\forall x \le \sqrt{۵}} f(x) = ۰/۱ + ۰/۲ + ۰/۱ + ۰/۲ = ۰/۶$$

مسائل

۱- در جدول زیر مقدار k چقدر باشد تا $f(X)$ یک توزیع احتمال باشد؟

X	۱-	۰	۱
$f(X)$	$\frac{۱}{۹}$	k	$\frac{۷}{۱۵}$

۲- در جدول زیر مقدار k چقدر باشد تا $f(X)$ یک توزیع احتمال باشد؟

X	۱	۲	٤	
$f(X)$	$\frac{۱}{۵}$	k	$\frac{۱}{۲}$	۲k

۳- از کیسه‌ای که در آن ۳ مهره سفید و ۲ مهره سیاه وجود دارد، ۲ مهره به تصادف خارج می‌کنیم. اگر X تعداد مهره‌های سفید در بین مهره‌های خارج شده باشد، توزیع احتمال متغیر تصادفی X را به دست آورید؟

٤- در یک خانواده ۳ فرزندی اگر X، تعداد فرزندان پسر و Y، تعداد فرزندان دختر باشد. مطلوب است:

الف: توزیع احتمال توام X و Y؟

ب: توزیع حاشیه‌ای X؟

ج: توزیع حاشیه‌ای Y ؟

۵- در جعبه‌ای ۱۰ کالا وجود دارد که ۳ تای آنها معیوب است. به طور تصادفی ٤ تا از این کالاها را بیرون می‌آوریم. X را تعداد کالاهای سالم و Y را تعداد کالاهای معیوب در بین ٤ کالای خارج شده تعریف می‌کنیم.مطلوب است:

الف- تابع توزیع احتمال توام X و Y را به دست آورید؟

ب- تابع توزیع احتمال حاشیه‌ای X و Y؟

فصل ششم

امید ریاضی

و کاربردهای آن

امید ریاضی یا میانگین یک متغیر تصادفی مانند X، پارامتری است که نشان دهنده مقدار مورد انتظار برای آن متغیر تصادفی در اثر تکرار آن آزمایش به دفعات زیاد است و آن را با $E(X)$ یا μ_X نشان داده و به صورت زیر تعریف می‌کنیم.

$$\mu_X = E(X) = \sum_{i=۱}^{n} x_i f(x_i)$$

مثال۱: توزیع احتمال زیر مفروض است، $E(X)$ را محاسبه کنید؟

X	۰	۱	۲
$f(X)$	۰/۲	۰/۵	۰/۳

پاسخ:
$$E(X) = \sum_{i=۱}^{۳} x_i f(x_i) = (۰\times ۰/۲) + (۱\times ۰/۵) + (۲\times ۰/۳) = ۱/۱$$

این بدین معنی است که اگر آزمایش به دفعات زیاد انجام شود، انتظار داریم که به طور متوسط متغیر تصادفی X، مقدار ۱/۱ را اختیار کند.

مثال۲: توزیع احتمال زیر مفروض است، $E(X)$ را محاسبه کنید؟

X	-۲	-۱	۰	۱	۲
$f(X)$	۰/۲	۰/۳	۰/۱	۰/۲	۰/۳

پاسخ:
$$E(X) = \sum x_i f(x_i) = (-۲\times ۰/۲) + (-۱\times ۰/۳) + (۰\times ۰/۲) + (۱\times ۰/۵) + (۲\times ۰/۳) = ۰/۱$$

این بدین معنی است که اگر آزمایش به دفعات زیاد انجام شود، انتظار داریم که به طور متوسط متغیر تصادفی X، مقدار ۰/۱ را اختیار کند.

قضیه: اگر متغیر تصادفی X دارای تابع احتمال $f(x)$ باشدف امید ریاضی هر تابعی از X مانند $g(X)$ عبارتست از:

$$E\left(g(x)\right) = \sum_{x} g(x) f(x)$$

مثال۳: تابع توزیع احتمال متغیر تصادفی X در زیر داده شده است، $E(X^۲)$ را به دست آورید؟

X	-۱	۰	۱	۲
$f(x)$	/۳	۰/۲	۰/٤	۰/۱

پاسخ: در این مثال $g(x) = x^۲$ است. در نتیجه:

$$E(x^۲) = \sum x^۲ f(x) = (-۱)^۲(۰/۳) + (۰)^۲(۰/۲) + (۱)^۲(۰/۴) + (۲)^۲(۰/۱)$$
$$= ۰/۳ + ۰ + ۰/۴ + ۰/۴ = ۱/۱$$

مثال٤: متغیر تصادفی X می‌تواند یکی از سه مقدار ٥، ٤ و $x_۳$ را انتخاب کند که احتمال‌های آنها به ترتیب ۰/۲، ۰/٥ و $p_۳$ است. اگر میانگین متغیر تصادفی X برابر ٦ باشد، مقدار $x_۳$ چقدر است؟

پاسخ: می‌توان اطلاعات داده شده در صورت مسئله را در جدول زیر خلاصه کرد:

X	٤	٥	$x_۳$
$f(x)=p(x_i)$	۰/۲	۰/٥	$p_۳$

با استفاده از شرط $\sum P(x_i)=۱$ داریم: $۰/۵+۰/۲+p_۳=۱ \quad \Rightarrow p_۳=./۳$

حال با توجه به اینکه $E(x)=۶$ داریم:

$$\begin{cases} E(X)=\sum xf(x)=۶ \\ \Rightarrow ۴(./۵)+۵(./۲)+x_۳(./۳)=۶ \\ \Rightarrow x_۳=۱۰ \end{cases}$$

امید ریاضی تابع احتمال توام:

اگر X و Y دو متغیر تصادفی با تابع احتمال توام $f(x,y)$ باشند، امید ریاضی هر تابعی از X و Y مانند $g(x,y)$ عبارتست از:

$$E\left(g(x,y)\right)=\sum_x\sum_y g(x,y)f(x,y)$$

مثال٥: تابع توزیع احتمال توام دو متغیر تصادفی X و Y در زیر داده شده است. مطلوب است محاسبه $E(X)$، $E(Y)$ و $E(XY)$.

$Y \backslash X$	۰	۱	$f(y)$
۰	۰/۱٥	۰/۳٥	۰/٥
-۱	۰/۱٥	۰/۳٥	۰/٥
$f(x)$	۰/۳	۰/۷	۱

پاسخ:

$$\begin{cases} E(x)=\sum xf(x)=(۰\times۰/۳)+(۱\times۰/۷)=۰/۷ \\ E(y)=\sum yf(y)=(۰\times۰/۵)+(-۱\times۰/۵)=-۰/۵ \\ E(xy)=\sum xyf(x,y)=(۰\times۰\times۰/۱۵)+(۰\times۱\times۰/۳۵)+(-۱\times۰\times۰/۱۵)+(-۱\times۱\times۰/۳۵)=-۰/۳۵ \end{cases}$$

خواص امید ریاضی:

اگر X و Y دو متغیر تصادفی، a و b دو عدد ثابت و توابع g و h توابعی از X و Y باشند، داریم:

$$\begin{cases} E(a)=a \\ E(aX+b)=aE(X)+b \\ E(a\,g(X)\pm b\,h(X))=a\,E(g(X))\pm b\,E(h(X)) \\ E(a\,g(X,Y)\pm b\,h(X,Y))=a\,E(g(X,Y))\pm b\,E(h(X,Y)) \\ E(X,Y)=E(X)E(Y) \quad \longleftarrow \text{اگر } X \text{ و } Y \text{ مستقل باشند} \end{cases}$$

مثال۶: در جدول توزیع احتمال زیر مطلوب است محاسبه:

X	-۲	-۱	۰	۱	۲
$f(X)=P(X=x)$	۰/۱	۰/۲	۰/٤	۰/۲	۰/۱

الف- $E(X)$.

ب- $E(X^{۲})$.

ج- $E(۳X+۵)$.

د- $E\left[(X-۱)^{۲}\right]$.

الف: $E(X)=\sum x_i f(x_i)=(-۲\times ۰/۱)+(-۱\times ۰/۲)+(۰\times ۰/۴)+(۱\times ۰/۲)+(۲\times ۰/۱)=۰$

ب: $E(X^{۲})=\sum x_i^{۲} f(x_i)=(-۲)^{۲}(۰/۱)+(-۱)^{۲}(۰/۲)+(۰)^{۲}(۰/۴)+(۱)^{۲}(۰/۲)+(۲)^{۲}(۰/۱)=۱/۲$

ج: $E(۳X+۵)=۳E(X)+۵=۳(۰)+۵=۵$

د: $E\left[(X-۱)^{۲}\right]=E(X^{۲}-۲X+۱)=E(X^{۲})-۲E(X)+۱=۱/۲-۲(۰)+۱=۲/۲$

مثال۷: در جدول توزیع احتمال زیر مطلوب است محاسبه:

X	۱	۲	۳	٤
$f(X)$	$\frac{۱}{۸}$	$\frac{۳}{۸}$	$\frac{۲}{۸}$	$\frac{۲}{۸}$

الف- $E(X)$.

ب- $E(X^{۳})$.

ج- $E(X^{۲}+۳X)$.

د- $E\left[(X+۱)^{۲}\right]$.

الف: $E(X)=\sum x_i f(x_i)=\left(۱\times\frac{۱}{۸}\right)+\left(۲\times\frac{۳}{۸}\right)+\left(۳\times\frac{۲}{۸}\right)+\left(۴\times\frac{۲}{۸}\right)=\frac{۲۱}{۸}$

ب:

$$E(X^{۳})=\sum x_i^{۳} f(x_i)=\left(۱^{۳}\times\frac{۱}{۸}\right)+\left(۲^{۳}\times\frac{۳}{۸}\right)+\left(۳^{۳}\times\frac{۲}{۸}\right)+\left(۴^{۳}\times\frac{۲}{۸}\right)=\frac{۱۷۵}{۸}$$

ج:

$$E(X^{2})=\sum x_i^{2} f(x_i)=\left(1^{2}\times\frac{1}{8}\right)+\left(2^{2}\times\frac{3}{8}\right)+\left(3^{2}\times\frac{2}{8}\right)+\left(4^{2}\times\frac{2}{8}\right)=\frac{47}{8}$$

$$E\left(X^{2}+3X\right)=E\left(X^{2}\right)+3E\left(X\right)=\frac{47}{8}+3\left(\frac{21}{8}\right)=\frac{99}{8}$$

د: $$E\left[\left(X+1\right)^{2}\right]=E\left(X^{2}+2X+1\right)=E\left(X^{2}\right)+2E\left(X\right)+1=\frac{47}{8}+2\times\frac{21}{8}+1=\frac{113}{8}$$

توجه داشته باشید که قسمت د را می‌توان از روش مستقیم نیز به صورت زیر حل کرد:

$$E\left[\left(X+1\right)^{2}\right]=\sum\left(x+1\right)^{2}.f\left(x\right)=\left(\left(1+1\right)^{2}\times\frac{1}{8}\right)+\left(\left(2+1\right)^{2}\times\frac{3}{8}\right)+\left(\left(3+1\right)^{2}\times\frac{2}{8}\right)+\left(\left(4+1\right)^{2}\times\frac{2}{8}\right)=\frac{113}{8}$$

امید شرطی:

اگر X و Y دو متغیر تصادفی با تابع احتمال توام $f(x,y)$ باشند، آنگاه امید ریاضی X به شرط $Y=y$ را با $E(X|Y=y)$ نمایش داده و به صورت زیر محاسبه می‌کنند:

$$E(X|Y=y)=\sum_x xf(x|Y=y)=\sum_x x\frac{f(x,y)}{f(y)}$$

و نیز همینطور امید ریاضی Y به شرط $X=x$ را با $E(Y|X=x)$ نمایش داده و به صورت زیر محاسبه می‌کنند:

$$E(Y|X=x)=\sum_Y Yf(Y|X=x)=\sum_Y y\frac{f(x,y)}{f(x)}$$

مثال۸: توزیع احتمال توام دو متغیر تصادفی X و Y در جدول زیر داده شده است. $E(Y|X=6)$ را به دست آورید؟

X \ Y	۲	۳	٤	۵	$f(x)$
٤	۰/۱	۰/۰۵	۰/۰۳	۰/۰۲	۰/۲
٦	۰/۰۵	۰/۲	۰/۲	۰/۰۵	۰/۵
۸	۰/۰۵	۰/۱	۰/۱	۰/۰۵	۰/۳
$f(y)$	۰/۲	۰/۳۵	۰/۳۳	۰/۱۲	۱

پاسخ:

$$E(Y\mid X=۶)=\frac{\sum y_i\, f(x=۶,y_i)}{f(x=۶)}=\frac{(۲\times ۰/۰۵)+(۳\times ۰/۲)+(۴\times ۰/۲)+(۵\times ۰/۰۵)}{۰/۰۵+۰/۲+۰/۲+۰/۰۵}=\frac{۱/۷۵}{۰/۵}=۳/۵$$

مثال۹: توزیع احتمال توام دو متغیر تصادفی X و Y در زیر داده شده است. مطلوب است محاسبه:

X \ Y	−۱	۰	۱
۲	۰/۱۲	۰/۱۸	۰/۲
٤	۰/۱۱	۰/۲۲	۰/۱۷

الف- $E(X\mid Y=۱)$.

ب- $E(Y\mid X=۲)$.

پاسخ: ابتدا باید توزیع‌های حاشیه‌ای را به دست آوریم:

X \ Y	−۱	۰	۱	$f(X)$
۲	۰/۱۲	۰/۱۸	۰/۲	۰/۵
٤	۰/۱۱	۰/۲۲	۰/۱۷	۰/۵
$f(Y)$	۰/۲۳	۰/٤	۰/۳۷	۱

الف:

$$E(X\mid Y=۱)=\sum_x X\, f(X\mid Y=۱)=\sum_x X\,\frac{f(X,Y=۱)}{f(Y=۱)}$$
$$=\left(۲\times\frac{۰/۲}{۰/۳۷}\right)+\left(۴\times\frac{۰/۱۷}{۰/۳۷}\right)=۲/۹۲$$

ب:

$$E(Y\mid X=۲)=\sum_y Y\, f(Y\mid X=۲)=\sum_y Y\,\frac{f(X=۲,Y)}{f(X=۲)}$$
$$=\left(-۱\times\frac{۰/۱۲}{۰/۵}\right)+\left(۰\times\frac{۰/۱۸}{۰/۵}\right)+\left(۱\times\frac{۰/۲}{۰/۵}\right)=۰/۱۶$$

واریانس ($Var(X), V(X), \sigma_x^{۲}$):

چنانچه در فصل دوم نیز بیان شد، واریانس یکی از شاخص‌های پراکندگی است که برای اندازه‌گیری میزان پراکندگی داده‌ها حول میانگین(امید ریاضی) به کار می‌رود. هرچه واریانس بزرگتر باشد یعنی پراکندگی داده‌ها بیشتر بوده و هر چه واریانس کمتر باشد، نشان دهنده تمرکز داده‌ها حول میانگین خواهد بود. داریم: $\sigma_X^{۲} = Var(X) = E\left(X - \mu_X\right)^{۲} = E(X^{۲}) - \mu_X^{۲}$.

اگر X یک متغیر تصادفی با تابع احتمال $f\left(x\right)$ باشد، واریانس آن به صورت زیر محاسبه خواهد شد:

$$\boxed{Var(x) = E\left(X - \mu_X\right)^{۲} = E(x^{۲}) - \left(E(x)\right)^{۲}}$$

مثال۱۰: تابع احتمال زیر مفروض است. امید ریاضی و واریانس X را به دست آورید.

X	۱-	۰	۱
$P(x)$	۰/٤	۰/۲	۰/٤

پاسخ:

$$E(x) = \sum x_i P(x_i) = (-۱)(۰/۴) + (۰)(۰/۲) + (۱)(۰/۴) = ۰$$
$$E(x^{۲}) = \sum x_i^{۲} P(x_i) = (-۱)^{۲}(۰/۴) + (۰)^{۲}(۰/۲) + (۱)^{۲}(۰/۴) = ./۸$$
$$\mathrm{var}(x) = E(x^{۲}) - E^{۲}\left(x\right) = ۰/۸ - ۰ = ۰/۸$$

ویژگی‌های واریانس:

اگر X یک متغیر تصادفی و a و b دو مقدار ثابت باشند، آنگاه:

$$Var(a) = ۰$$
$$Var(ax) = a^{۲} Var(x)$$
$$Var(ax \pm b) = a^{۲} Var(x)$$

مثال۱۱: در جدول داده شده زیر، مطلوب است محاسبه:

الف- $Var\left(X\right)$.

ب- $Var\left(۳X + ۴\right)$.

X	۳-	۱-	۰	۱	۳
$f\left(X\right)$	$\frac{۱}{۹}$	$\frac{۳}{۹}$	$\frac{۱}{۹}$	$\frac{۳}{۹}$	$\frac{۱}{۹}$

پاسخ:

الف:

$$E(X)=\sum xf(x)=\left(-۳\times\frac{۱}{۹}\right)+\left(-۲\times\frac{۳}{۹}\right)+\left(۰\times\frac{۱}{۹}\right)+\left(۱\times\frac{۳}{۹}\right)+\left(۳\times\frac{۱}{۹}\right)=۰$$

$$E(X^۲)=\sum x^۲f(x)=\left((-۳)^۲\times\frac{۱}{۹}\right)+\left((-۱)^۲\times\frac{۳}{۹}\right)+\left((۰)^۲\times\frac{۱}{۹}\right)+\left((۱)^۲\times\frac{۳}{۹}\right)+\left((۳)^۲\times\frac{۱}{۹}\right)=\frac{۸}{۳}$$

$$Var(X)=E(X^۲)-(E(X))^۲=\frac{۸}{۳}$$

ب:

$$Var(۳X+۴)=۹Var(X)=۹\times\frac{۸}{۳}=۲۴$$

مثال۱۲: متغیر تصادفی X دارای میانگین ۵ و واریانس ۹ است. میانگین و واریانس $\frac{x-۵}{۳}$را به دست آورید؟

پاسخ:

$$E\left(\frac{x-۵}{۳}\right)=\frac{E(x)}{۳}-\frac{۵}{۳}=\frac{۵}{۳}-\frac{۵}{۳}=۰$$

$$\text{var}\left(\frac{x-۵}{۳}\right)=\frac{\text{var}(x)}{۹}=\frac{۹}{۹}=۱$$

مثال۱۳: تابع توزیع احتمال توام دو متغیر تصادفی X و Y در زیر داده شده است، مطلوب است محاسبه:

الف- $E(X)$ & $E(Y)$ ؟

ب- $Var(X)$ & $Var(Y)$ ؟

ج- $E(۲X+۳Y)$ & $Var(۲X-۴)$ ؟

X \ Y	-۱	۰	۳
۲	۰/۱	۰/٤	۰
٤	۰/۱۵	۰/۲	۰/۱۵

پاسخ: ابتدا بایستی توزیع‌های حاشیه‌ای را محاسبه کنیم.

X \ Y	-۱	۰	۳	$f(X)$
۲	۰/۱	۰/٤	۰	۰/۵
٤	۰/۱۵	۰/۲	۰/۱۵	۰/۵
$f(Y)$	۰/۲۵	۰/٦	۰/۱۵	۱

الف:

$$\begin{cases} E(X)=(۲)(۰/۵)+(۴)(۰/۵)=۳ \\ E(Y)=(-۱)(۰/۲۵)+(۰)(۰/۶)+(۳)(۰/۱۵)=۰/۲ \end{cases}$$

ب:

$$E(X^۲)=\sum x^۲f(x)=(۲)^۲(۰/۵)+(۴)^۲(۰/۵)=۱۰$$

$$Var(X)=E(X^۲)-[E(X)]^۲=۱۰-۳^۲=۱$$

$$E(Y^۲)=\sum y^۲f(y)=(-۱)^۲(۰/۲۵)+(۰)^۲(۰/۶)+(۳)^۲(۰/۱۵)=۱/۶$$

$$Var(Y)=E(Y^۲)-[E(Y)]^۲=۱/۶-(۰/۲)^۲=۱/۵۶$$

ج:

$$E(۲X+۳Y)=۲E(X)+۳E(Y)=(۲)(۳)+(۳)(۰/۲)=۶/۶$$

$$Var(۲X-۴)=۴Var(X)=۴$$

مثال۱۴: تابع توزیع احتمال توام دو متغیر تصادفی X و Y در زیر داده شده است، مطلوب است محاسبه:

الف- $E(X)\quad \&\quad E(Y)$؟

ب- $Var(X)\quad \&\quad Var(Y)$؟

$X \backslash Y$	۰	۲	٤
۱	۰/۱	۰	۰/۲
۳	۰/۳	۰/۱	۰
۵	۰	۰/۱	۰/۲

پاسخ: ابتدا بایستی توزیع‌های حاشیه‌ای را محاسبه کنیم.

$X \backslash Y$	۰	۲	٤	$f(X)$
۱	۰/۱	۰	۰/۲	۰/۳
۳	۰/۳	۰/۱	۰	۰/٤
۵	۰	۰/۱	۰/۲	۰/۳
$f(Y)$	۰/٤	۰/۲	۰/٤	۱

الف:

$$\begin{cases} E(X)=(۱)(۰/۳)+(۳)(۰/۴)+(۵)(۰/۳)=۳ \\ E(Y)=(۰)(۰/۴)+(۲)(۰/۲)+(۴)(۰/۴)=۲ \end{cases}$$

$$E(X^{۲})=(۱^{۲})(۰/۳)+(۳^{۲})(۰/۴)+(۵^{۲})(۰/۳)=۱۱/۴$$

$$Var(X)=E(X^{۲})-[E(X)]^{۲}=۱۱/۴-۳^{۲}=۲/۴$$

ب:

$$E(Y^{۲})=\sum y^{۲}f(y)=(۰^{۲})(۰/۴)+(۲^{۲})(۰/۲)+(۴^{۲})(۰/۴)=۷/۲$$

$$Var(Y)=E(Y^{۲})-[E(Y)]^{۲}=۷/۲-(۲)^{۲}=۳/۲$$

توجه: برای داده‌های آماری واریانس را می‌توان از رابطه زیر به دست آورد:

$$\boxed{Var(X)=\sum x_i^{۲}-n(\bar{x})^{۲}}$$

مثال۱۵: اگر برای ۱۰ داده، مقادیر $\sum x_i=۵۰$ و $\sum x_i^{۲}=۳۰۰$ به دست آمده باشد. $Var(X)$ را به دست آورید؟

پاسخ:

$$Var(X)=\sum x_i^{۲}-n(\bar{x})^{۲}=۳۰۰-۱۰(۵)^{۲}=۵۰$$

$$\bar{x}=\frac{\sum x_i}{n}=\frac{۵۰}{۱۰}=۵$$

جذر واریانس را انحراف معیار نامیده و برای متغیر تصادفی X با σ_X نمایش می‌دهند. بنابراین داریم: $\sigma_X=\sqrt{\sigma_X^{۲}}=\sqrt{Var(X)}$.

کوواریانس:

برای اینکه کوواریانس را تعریف کنیم نخست با استفاده از تعریف واریانس به محاسبه مجموع دو متغیر تصادفی X و Y با میانگین‌های μ_X و μ_Y می‌پردازیم. چون امید ریاضی $X+Y$ برابر است با $\mu_X+\mu_Y$، پس داریم:

$$\begin{aligned} Var(X+Y) &= E\left(\left[(X+Y)-(\mu_x+\mu_y)\right]^{۲}\right) \\ &= E\left[(X-\mu_X)^{۲}+(Y-\mu_Y)^{۲}+۲(X-\mu_X)(Y-\mu_Y)\right] \\ &= E\left[(X-\mu_X)^{۲}\right]+E\left[(Y-\mu_Y)^{۲}\right]+۲E\left[(X-\mu_X)(Y-\mu_Y)\right] \end{aligned}$$

آخرین امیدریاضی را با $Cov(X,Y)$ نشان داده و آن را کواریانس X و Y می‌نامیم. بنابراین برای $Var(X+Y)$ فرمول زیر به نام فرمول واریانس مجموع به دست می‌آید:

$$\boxed{Var(X+Y)=Var(X)+Var(Y)+۲Cov(X,Y)}$$

کواریانس را امیدریاضی تغییرات دو متغیر بر حسب میانگین‌شان تعریف میکنیم.

کواریانس معیاری عددی است که نوع و جهت رابطه خطی بین دو متغیر تصادفی را نشان می‌دهد و به صورت زیر محاسبه می‌شود:

$$Cov(X,Y)=E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

کواریانس می‌تواند مقادیر منفی، مثبت و صفر را اختیار کند که با توجه به مقدار به دست آمده برای آن نوع و جهت رابطه بین بین دو متغیر را به صورت زیر می‌توان توضیح داد.

با افزایش (کاهش) X، Y نیز افزایش (کاهش) می‌یابد $\longrightarrow$	$\sigma_{XY}=Cov(X,Y)>۰$
با افزایش (کاهش) X، Y کاهش (افزایش) می‌یابد $\longrightarrow$	$\sigma_{XY}=Cov(X,Y)<۰$
افزایش (کاهش) یک متغیر هیچ تاثیری در دیگری نداشته باشد $\longrightarrow$	$\sigma_{XY}=Cov(X,Y)=۰$

نکته: اگر دو متغیر تصادفی X و Y مستقل باشند آنگاه $Cov(X,Y)=۰$ خواهد بود. اما عکس این مطلب همواره برقرار نیست. یعنی ممکن است کواریانس دو متغیر برابر صفر باشد اما دو متغیر مستقل نباشند.

مثال۱۶: تابع توزیع احتمال توام در جدول زیر داده شده است. $Cov(X,Y)$ را محاسبه کنید؟

$X \backslash Y$	۰	۱	$f(X)$
۱-	۰/۳	۰/۳	۰/۶
۰	۰/۳	۰/۱	۰/۴
$f(Y)$	۰/۶	۰/۴	۱

پاسخ:

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$
$$E(X)=\sum x.f(x)=(-1)(0/6)+(0)(0/4)=-0/6$$
$$E(Y)=\sum y.f(y)=(0)(0/6)+(1)(0/4)=0/4$$
$$E(XY)=\sum xy.f(xy)=(-1)(0)(0/3)+(-1)(1)(0/3)+(0)(0)(0/3)+(0)(1)(0/1)=-0/3$$
$$\Rightarrow Cov(X,Y)=-0/3-(-0/6)(0/4)=-0/06$$

با توجه به اینکه مقدار کواریانس عددی منفی شده است، نتیجه می‌شود که رابطه بین دو متغیر یک رابطه معکوس و در خلاف جهت یکدیگر است.

ویژگی‌های کواریانس:

با استفاده از فرمول محاسبه کواریانس می‌توان به آسانی ویژگی‌های زیر را ثابت کرد:

۱) $Cov(X,Y)=Cov(Y,X)=\sigma_{XY}=\sigma_{YX}$

۲) $Cov(X,X)=Var(X)=\sigma_X^2$

۳) $Cov(X,a)=0$

۴) $Cov(X+a,Y+b)=Cov(X,Y)$

۵) $Cov(aX,bY)=abCov(X,Y)$

۶) $Cov(aX+b,cY+d)=acCov(X,Y)$

۷) $Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)$

۸) $Cov(aX+bY,cZ)=acCov(X,Z)+bcCov(Y,Z)$

۹) $Var(aX\pm bY+c)=a^2Var(X)+b^2Var(Y)\pm 2abCov(X,Y)$

۱۰) $Var(aX\pm bY\pm cZ+d)=a^2Var(X)+b^2Var(Y)+c^2Var(Z)\pm 2abCov(X,Y)$
$\pm 2acCov(X,Z)\pm 2bcCov(Y,Z)$

اگر متغیرهای X، Y و Z مستقل باشند،

۱۱) $Var(aX\pm bY\pm cZ+d)=a^2Var(X)+b^2Var(Y)+c^2Var(Z)$

مثال۱۷: تابع توزیع توام داده شده زیر را در نظر بگیرید. مطلوب است محاسبه:

$X \backslash Y$	۲−	۱۰
۰	۰/۳۳	۰/۱۷
٤	۰/۲۷	۰/۲۳

الف- $Cov(X,Y)$.

ب- $Cov(2X+1, 3Y-4)$.

ج- $Var(X-Y)$.

پاسخ: ابتدا توزیع حاشیه‌ای را به دست می‌آوریم:

$X \backslash Y$	۲−	۱۰	$f(X)$
۰	۰/۳۳	۰/۱۷	۰/۵
٤	۰/۲۷	۰/۲۳	۰/۵
$f(Y)$	۰/٦	۰/٤	۱

الف:

$$E(X)=\sum_x X.f(X)=(0\times 0/5)+(4\times 0/5)=2$$
$$E(Y)=\sum_y Y.f(Y)=(-2\times 0/6)+(10\times 0/4)=2/8$$
$$E(XY)=\sum_x\sum_y XY.f(X,Y)=(0\times(-2)\times 0/33)+(0\times 10\times 0/17)+(4\times(-2)\times 0/27)+(4\times 10\times 0/23)=7/04$$
$$Cov(X,Y)=E(XY)-E(X)E(Y)=7/04-(2)(2/8)=1/44$$

ب: بنا بر نکته ٦ بالا داریم: $Cov(2X+1, 3Y-4)=6Cov(X,Y)$. در نتیجه:

$$Cov(2X+1, 3Y-4)=6Cov(X,Y)=6\times 1/44=8/64$$

ج: با توجه به نکته ۹ داریم: $Var(X-Y)=Var(X)+Var(Y)-2Cov(X,Y)$. بنابراین نیاز به محاسبه $Var(X)$ و $Var(Y)$ به صورت زیر داریم.

$$E(X^2)=\sum X^2 f(X)=(0^2\times 0/5)+(4^2\times 0/5)=8$$
$$Var(X)=E(X^2)-(E(X))^2=8-2^2=4$$
$$E(Y^2)=\sum Y^2 f(Y)=((-2)^2\times 0/6)+(10^2\times 0/4)=42/4$$
$$Var(Y)=E(Y^2)-(E(Y))^2=42/4-(2/8)^2=34/56$$

$$\Rightarrow Var(X-Y)=Var(X)+Var(Y)-2Cov(X,Y)$$
$$=4+34/56-2(1/44)=35/68$$

توجه: قبل از محاسبه کواریانس بهتر است که مستقل بودن را چک کنیم. چون اگر مستقل باشند، آنگاه کواریانس برابر صفر است.

مثال۱۸: تابع توریع احتمال دو متغیر تصادفی X و Y در زیر داده شده است. $Cov(X,Y)$ را محاسبه کنید؟

$Y \backslash X$	۰	۱	$f(y)$
۰	۰/۱۵	۰/۳۵	۰/۵
−۱	۰/۱۵	۰/۳۵	۰/۵
$f(x)$	۰/۳	۰/۷	۱

پاسخ:

$$E(x)=\sum xf(x)=(0\times 0/3)+(1\times 0/7)=0/7$$
$$E(y)=\sum yf(y)=(0\times 0/5)+(-1\times 0/5)=-0/5$$
$$E(xy)=\sum xyf(x,y)=(0\times 0\times 0/15)+(0\times 1\times 0/35)+(-1\times 0\times 0/15)+(-1\times 1\times 0/35)=-0/35$$
$$\Rightarrow\quad Cov(X,Y)=E(XY)-E(X)E(Y)=-0/35-(0/7)(-0/5)=0$$

قبل از انجام این محاسبات با کمی دقت، متوجه می‌شدیم که در جدول بالا برای تمامی مقادیر ممکن رابطه $f(x,y)=f(x).f(y)$ برقرار است. در نتیجه بدون محاسبه کواریانس، می‌توانستیم نتیجه بگیریم که:

$$Cov(X,Y)=0$$

ضریب همبستگی:

معیار کواریانس برای سنجش تغییرات X و Y به واحد اندازه‌گیری بستگی دارد. در نتیجه تغییر در یکی از متغیرهای X و Y در محاسبه کواریانس دخالت می‌کند. برای رفع این مشکل از معیار دیگری تحت عنوان ضریب همبستگی، که تحت تاثیر واحدهای اندازه‌گیری نمی‌باشد، به منظور تعیین شدت رابطه و همچنین نوع رابطه (مستقیم یا معکوس) بین دو متغیر استفاده می‌شود.

اگر به جای X و Y، مقادیر $\frac{X}{\sigma_X}$ و $\frac{Y}{\sigma_Y}$ را در فرمول کواریانس به کار بریم، به رابطه ضریب همبستگی خواهیم رسید. ضریب همبستگی را با ρ_{XY} یا r_{XY} نشان داده و به صورت زیر محاسبه می‌شود:

$$\boxed{\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}=\frac{\sigma_{XY}}{\sigma_x\sigma_Y}}$$

مثال۱۹: با توجه به جدول توزیع احتمال توام زیر، کواریانس و ضریب همبستگی را به دست آورید؟

$X \backslash Y$	−۱	۰	۱
۰	۰/۱	۰	۰/۲
۱	۰/۲	۰/۱	۰/۱
۲	۰	۰/۲	۰/۱

پاسخ: ابتدا باید توزیع حاشیه‌ای را به صورت زیر به دست آوریم:

$X \backslash Y$	−۱	۰	۱	$f(X)$
۰	۰/۱	۰	۰/۲	۰/۳
۱	۰/۲	۰/۱	۰/۱	۰/٤
۲	۰	۰/۲	۰/۱	۰/۳
$f(Y)$	۰/۳	۰/۳	۰/٤	۱

$$E(X)=\sum xf(x)=(۰\times ۰/۳)+(۱\times ۰/۴)+(۲\times ۰/۳)=۱$$
$$E(Y)=\sum yf(y)=(-۱\times ۰/۳)+(۰\times ۰/۳)+(۱\times ۰/۴)=۰/۱$$
$$E(XY)=\sum xy\,f(x,y)=(۰)(-۱)(۰/۱)+(۰)(۰)(۰)+(۰)(۱)(۰/۲)+(۱)(-۱)(۰/۲)+(۱)(۰)(۰/۱)+$$
$$(۱)(۱)(۰/۱)+(۲)(-۱)(۰)+(۲)(۰)(۰/۲)+(۲)(۱)(۰/۱)=۰/۱$$
$$Cov(X,Y)=E(XY)-E(X)E(Y)=۰/۱-۰/۱=۰$$

با توجه به صفر شدن کواریانس، نتیجه می‌گیریم که ضریب همبستگی نیز صفر است. یعنی دو متغیر هیچ رابطه‌ای با هم ندارند.

توجه: ضریب همبستگی برای داده‌های آماری به صورت زیر محاسبه می‌شود:

$$\rho_{XY}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}=\frac{\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n}}{\sqrt{\frac{\sum(x_i-\bar{x})^۲}{n}\times\frac{\sum(y_i-\bar{y})^۲}{n}}}=\frac{\frac{\sum x_iy_i}{n}-\left(\frac{\sum x_i}{n}\right)\left(\frac{\sum y_i}{n}\right)}{\sqrt{\left(\frac{\sum x_i^۲}{n}-\left(\frac{\sum x_i}{n}\right)^۲\right)\left(\frac{\sum y_i^۲}{n}-\left(\frac{\sum y_i}{n}\right)^۲\right)}}$$

مثال۲۰: برای داده‌های جدول زیر مقدار ضریب همبستگی را محاسبه کنید؟

X_i	۷	۱۰	٤	۱۱
Y_i	۱٤	۲۰	۸	۲۲

پاسخ:

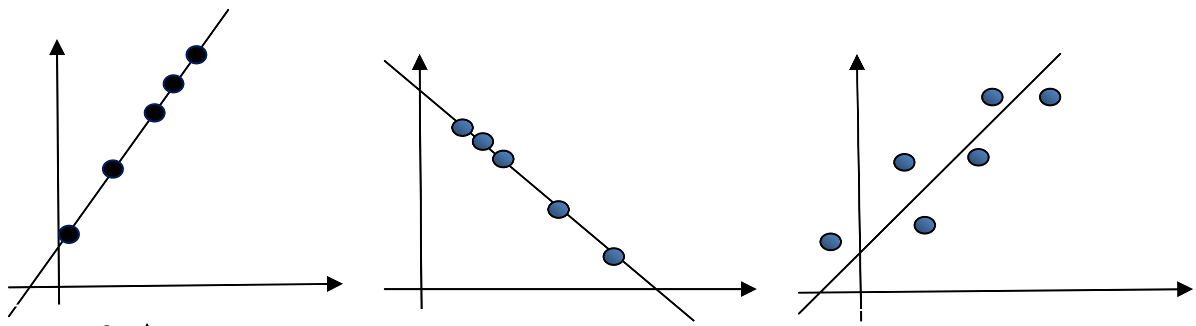
x_i	y_i	$x_i^۲$	$y_i^۲$	$x_i y_i$
۷	۱٤	٤۹	۱۹٦	۹۸
۱۰	۲۰	۱۰۰	٤۰۰	۲۰۰
٤	۸	۱٦	٦٤	۳۲
۱۱	۲۲	۱۲۱	٤۸٤	۲٤۲
$\sum x_i = ۳۲$	$\sum y_i = ۶۴$	$\sum x_i^۲ = ۲۸۶$	$\sum y_i^۲ = ۱۱۴۴$	$\sum x_i y_i = ۵۷۲$

$$\rho_{X\,Y} = \frac{\dfrac{\sum x_i y_i}{n} - \left(\dfrac{\sum x_i}{n}\right)\left(\dfrac{\sum y_i}{n}\right)}{\sqrt{\left(\dfrac{\sum x_i^۲}{n} - \left(\dfrac{\sum x_i}{n}\right)^۲\right)\left(\dfrac{\sum y_i^۲}{n} - \left(\dfrac{\sum y_i}{n}\right)^۲\right)}}$$

$$= \frac{\dfrac{۵۷۲}{۴} - \dfrac{۳۲}{۴} \times \dfrac{۶۴}{۴}}{\sqrt{\left(\dfrac{۲۸۶}{۴} - \left(\dfrac{۳۲}{۴}\right)^۲\right)\left(\dfrac{۱۱۴۴}{۴} - \left(\dfrac{۶۴}{۴}\right)^۲\right)}} = \frac{۱۴۳ - ۸ \times ۱۶}{\sqrt{(۷۱/۵ - ۶۴)(۲۸۶ - ۲۵۶)}} = \frac{۱۵}{\sqrt{۲۲۵}} = ۱$$

خواص ضریب همبستگی:

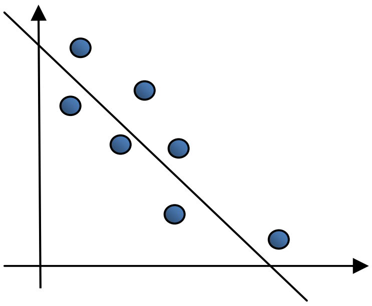
۱- همواره رابطه $-۱ \le \rho \le ۱$ برقرار است.



$\rho = ۱$ — رابطه مستقیم و کامل

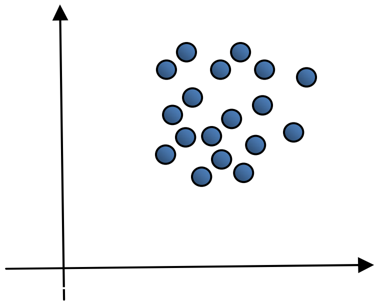
$\rho = -۱$ — رابطه معکوس و کامل

$۰ < \rho < ۱$ — رابطه مستقیم و ناقص



$-1 < \rho < 0$

رابطه معکوس و ناقص



$\rho = 0$

رابطه ندارند

۱) $\rho_{X,-X} = -1$

۲) $\rho_{X,X} = 1$

۳) $\rho_{X,a} = 0$

۴) $\rho_{X,Y} = \rho_{Y,X}$

۵) $\rho_{aX \pm b, cY \pm d} = \begin{cases} \rho_{X,Y} & \longleftarrow \text{اگر } a \text{ و } c \text{ هم علامت باشند} \\ -\rho_{X,Y} & \longleftarrow \text{اگر } a \text{ و } c \text{ مختلف العلامهٔ باشند} \end{cases}$

مثال۲۱: اگر $Cov(X,Y) = 10$، $\sigma_X^2 = 25$ و $\sigma_Y^2 = 9$ باشد، ضریب همبستگی را به دست آورید؟

پاسخ:
$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X . \sigma_Y} = \frac{10}{(5)(3)} = 0/66$$

مثال۲۲: ضریب همبستگی $-5x + \frac{2}{3}$ و $3x$ را به دست آورید؟

پاسخ: بنا بر نکته ۵ و ۲ داریم:
$$\rho_{3x, -5x+\frac{2}{3}} = -\rho_{x,x} = -1$$

ضریب تعیین:

ضریب تعیین، یکی از مهمترین معیارهایی که می‌توان به کمک آن رابطه بین دو متغیر X و Y را توضیح داد، ضریب تعیین است. در واقع ضریب تعیین بیان کننده درصد تغییرات متغیر وابسته به وسیله تغییرات متغیر مستقل است. به عبارت دیگر ضریب تعیین معلوم می‌کند که چند درصد از تغییرات Y ناشی از

تغییرات X است.ضریب تعیین را با $R^{۲}$ یا $r^{۲}$ نشان داده و به صورت مجذور ضریب همبستگی، $\rho(r)$، تعریف می‌شود. به عبارت ریاضی: $R^{۲} = r^{۲} = \rho^{۲}$.

مثال۲۳: اگر $Cov\,(X\,,Y\,) = ۳۲/۴$ و $\sigma_X^{۲} = \sigma_Y^{۲} = ۳۶$ باشد، ضریب تعیین را به دست آورید؟

پاسخ: $$r = \rho = \frac{Cov\left(X\,,Y\,\right)}{\sigma_X\,.\sigma_Y} = \frac{۳۲/۴}{(۶)(۶)} = ۰/۹ \quad \longrightarrow \quad r^{۲} = ۰/۸۱$$

مثال۲٤: اگر کواریانس دو متغیر تصادفی X و Y برابر $Cov\left(X\,,Y\,\right) = -۲۷$ باشد و و اریانس هر یک به ترتیب $Var\left(X\,\right) = ۳۶$ و $Var\left(Y\,\right) = ۲۵$ به دست آمده باشد، چند درصد تغییرات Y به وسیله X بیان می-شود؟

پاسخ: $r = \frac{Cov\left(X\,,Y\,\right)}{\sigma_X\,.\sigma_Y} = \frac{-۲۷}{\sqrt{۳۶}\sqrt{۲۵}} = -۰/۹ \quad \longrightarrow \quad r^{۲} = ۰/۸۱$ در نتیجه ۸۱٪ تغییرات Y به وسیله X بیان می‌شود.

توجه داشته باشید که، $۱-R^{۲}$، نشان دهنده درصدی از تغییرات Y است که توسط X بیان نمی‌شود.

مسائل

۱- متغیر تصادفی X دارای توزیع احتمال زیر است. امید ریاضی و واریانس آن را به دست آورید؟

X	۰	۱	۲	۳
$f(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

۲- متغیر تصادفی X دارای توزیع احتمال زیر است. امید ریاضی ، واریانس و انحراف استاندارد آن را به دست آورید؟

X	۰	۱	۲	۳	۴
$f(X)$	$\frac{1}{6}$	$1-2a$	a	$\frac{1}{3}$	$\frac{1}{6}$

۳- در پرتاب دو تاس، اگر متغیر تصادفی X مجموع شماره‌های دو تاس باشد، امیدریاضی، واریانس و انحراف استاندارد X را بیابید؟

٤- از جعبه‌ای که شامل ٥ گوی قرمز و ٤ گوی سفید است، ۳ گوی به تصادف خارج می‌کنیم. اگر متغیر تصادفی X را تعداد گوی‌های سفید در بین گوی‌های خارج شده تعریف کنیم، مطلوب است محاسبه امید ریاضی و واریانس X.

۵- اگر $E(X)=3/4$ و $E(X^2)=12$ باشد، $Var\left(-\frac{1}{2}x+3\right)$ را به دست آورید؟

٦- اگر میانگین و انحراف معیار X برابر ۲ باشد، میانگین X^2 چقدر است؟

۷- توزیع احتمال توام دو متغیره X و Y به صورت زیر است. $E(Y \mid X=1)$ را به دست آورید؟

X \ Y	۰	۱	۲
۰	۰/۲	۰/۱۵	۰/۰۵
۱	۰/۰۵	۰/۲	۰/۰۵
۲	۰/۰۵	۰/۰۵	۰/۲

۸- توزیع احتمال توام دو متغیره X و Y به صورت زیر است. $Cov(X,Y)$ و $Var(X+Y)$ را حساب کنید؟

Y \ X	۱	۲	۳
۱	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$
۲	$\frac{2}{8}$	۰	۰

۹- توزیع احتمال توام دو متغیره X و Y به صورت زیر است. $E(X)$، $E(Y)$، $E(XY)$ و $Cov(X,Y)$ را به دست آورید؟

$Y \backslash X$	۱	۲	۳
۱	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
۲	$\frac{1}{6}$	۰	$\frac{1}{6}$
۳	۰	$\frac{1}{3}$	۰

۱۰- در تمرین ۹، $Cov(X+Y, 2X)$ را به دست آورید؟

۱۱-توزیع احتمال توام دو متغیره X و Y به صورت زیر است. $Cov(X,Y)$ و $\rho_{X,Y}$ را به دست آورید؟

$Y \backslash X$	۱	۲	۳
۱-	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$
۰	۰	$\frac{1}{8}$	$\frac{1}{8}$

فصل هفتم

توزیع های گسسته

بسیاری از آزمایش‌ها، در شرایط یکسان دارای نتایج مشابهی هستند. از این‌رو بهتر است که این آزمایش‌های مشابه را دسته‌بندی کرده و خواصی را برای آنها معرفی کنیم. در فصل پنج با صورت کلی توزیع‌های احتمال آشنا شدیم. در این فصل به توصیف چند توزیع احتمال خاص که به دلیل ویژگی‌های خاص خود در بسیاری از کاربردها پیش می‌آیند، خواهیم پرداخت.

توزیع برنولی:

در بسیاری از آزمایش‌ها، توجه ما به وقوع یا عدم وقوع پیشامد معینی است. برای مثال: بررسی افراد در مورد وجود یا عدم وجود بیماری خاص، بررسی یک کالا در مورد سالم یا معیوب بودن آن و ....

هر آزمایشی که دو برآمد داشته باشد یا نتایج آن در دو رده پیروزی یا شکست، دسته‌بندی شوند را یک آزمایش برنولی می‌نامیم. اگر احتمال موفقیت را با p و احتمال شکست را با q که در آن $q=(1-p)$ است، نمایش دهیم، آنگاه داریم: $p+q=p+(1-p)=1$.

توجه داشته باشید که آزمایش برنولی تنها یکبار انجام می‌گیرد. اگر متغیر تصادفی X، نشان دهنده تعداد پیروزی در یک آزمایش برنولی با احتمال پیروزی p، باشد، آن را به صورت $X \sim Ber(p)$ نمایش می‌دهیم. توزیع احتمال X، به صورت زیر خواهد بود:

X	۰	۱
$f(X)=P(X=x)$	$1-p$	p

حال اگر بخواهیم این توزیع احتمال را به صورت یک تابع توزیع احتمال بیان کنیم، به صورت زیر خواهد بود: $P(X=x)=p^x(1-p)^{1-x} \qquad X=0,1$.

اگر $X \sim Ber(p)$ باشد، امید ریاضی و واریانس آن به صورت زیر خواهد بود:

$$\mu_x=E(X)=p$$
$$\sigma_x^2=Var(X)=p(1-p)$$

توزیع دوجمله‌ای:

n آزمایش برنولی را در نظر بگیرید. بطوریکه:

۱- احتمال موفقیت در هر آزمایش مقداری ثابت باشد.
۲- آزمایش‌ها از هم مستقل باشند.

حال اگر متغیر تصادفی X را تعداد موفقیت‌ها در n آزمایش مستقل برنولی با احتمال پیروزی p در هر آزمایش، در نظر بگیریم. آنگاه متغیر تصادفی X دارای توزیع دوجمله‌ای با تابع توزیع احتمالی به صورت زیر است و به صورت $X \sim Bin(n,p)$ نشان می‌دهیم.

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x} \qquad x=0,1,2,3,\cdots,n$$

اگر $X \sim Bin(n,p)$ باشد، امید ریاضی و واریانس آن به صورت زیر خواهد بود:

$$\mu_x=E(X)=np$$
$$\sigma_x^2=Var(X)=np(1-p)$$

به طور خلاصه، توزیع دوجمله‌ای برای پاسخ دادن به پرسش‌هایی که در آن، احتمال تعداد X پیشامد در n آزمایش مستقل برنولی مد نظر می‌باشد، مفید است.

مثال۱:احتمال مرگ در یک عمل جراحی ۰/۱ است. اگر قرار باشد این عمل روی ۵ نفر انجام شود، احتمال مرگ ۲ نفر چقدر است؟

پاسخ: اگر متغیر تصادفی X را تعداد افرادی که در این ۵ عمل جراحی فوت می‌کنند و پیروزی را مرگ فرد، تعریف کنیم، آنگاه: $X \sim Bin(n=5, p=0/1)$. در نتیجه احتمال مرگ ۲ نفر برابر است با:

$$P(X=x)=\binom{5}{x}(0/1)^x(1-0/1)^{5-x} \qquad x=0,1,2,3,4,5$$

$$\Rightarrow \quad P(X=2)=\binom{5}{2}(0/1)^2(1-0/1)^{5-2}=\binom{5}{2}(0/1)^2(0/9)^3=0/729$$

مثال۲: فرض کنید که حدود٦٠٪ از اتومبیل‌های سطح کشور بیمه شخص ثالث داشته باشند. یک مامور راهنمایی ۵ اتومبیل را متوقف و کارت بیمه آنان را مطالبه می‌کند.مطلوب است محاسبه:

الف- احتمال اینکه دقیقاً ۲ تا از این ماشین‌ها بیمه داشته باشند ؟

ب- امید ریاضی و واریانس X را به دست آورید؟

پاسخ: اگر متغیر تصادفی X را تعداد ماشین‌هایی که دارای بیمه شخص ثالث هستند و پیروزی را داشتن بیمه شخص ثالث تعریف کنیم، آنگاه: $X \sim Bin(n=۵ , p=۰/۶)$. در نتیجه احتمال اینکه دقیقاً ۲ نفر دارای بیمه شخص ثالث باشند، برابر است با:

الف:

$$P(X=x)=\binom{۵}{x}(۰/۶)^{x}(۱-۰/۶)^{۵-x} \qquad x=۰,۱,۲,۳,۴,۵$$

$$\Rightarrow \quad P(X=۲)=\binom{۵}{۲}(۰/۶)^{۲}(۱-۰/۶)^{۵-۲}=\binom{۵}{۲}(۰/۱)^{۲}(۰/۴)^{۳}=۰/۲۳$$

ب:

$$E(X)=np=۵\times ۰/۶=۳$$

$$Var(X)=npq=۵\times ۰/۶\times ۰/۴=۱/۲$$

مثال۳: ۸۰ درصد تیرهای یک تیرانداز به هدف برخورد می‌کند. این تیرانداز ۸ تیر به سمت هدف شلیک می‌کند. احتمال آنکه ٦ تیر به هدف برخورد کند، چقدر است؟

پاسخ: اگر متغیر تصادفی X را تعداد تیرهایی که به هدف برخورد می‌کنند و پیروزی را برخورد تیر به هدف، تعریف کنیم، آنگاه: $X \sim Bin(n=۸ , p=۰/۸)$. در نتیجه احتمال اینکه دقیقاً ٦ تیر به هدف برخورد کند، برابر است با:

$$P(X=x)=\binom{۸}{x}(۰/۸)^{x}(۱-۰/۸)^{۵-x} \qquad x=۰,۱,۲,۳,۴,۵,۶,۷,۸$$

$$\Rightarrow \quad P(X=۶)=\binom{۸}{۶}(۰/۸)^{۶}(۱-۰/۸)^{۸-۶}=\binom{۸}{۶}(۰/۸)^{۶}(۰/۲)^{۲}\simeq ۰/۲۹$$

مثال٤: یک سکه سالم را ۵ بار مستقلاً پرتاب می‌کنیم. مطلوب است محاسبه:

الف- احتمال اینکه دقیقاً ۲ بار شیر بیاید؟

ب- امید ریاضی و واریانس این توزیع را به دست آورید؟

پاسخ: اگر متغیر تصادفی X را تعداد شیرها و پیروزی را آمدن شیر، تعریف کنیم، آنگاه: $X \sim Bin(n=۵ , p=۰/۵)$.

الف: احتمال اینکه دقیقاً ۲ شیر بیاید، برابر است با:

$$P(X=x)=\binom{۵}{x}(۰/۵)^{x}(۱-۰/۵)^{۵-x} \qquad x=۰,۱,۲,۳,۴,۵$$

$$\Rightarrow \quad P(X=۲)=\binom{۵}{۲}(۰/۵)^{۲}(۱-۰/۵)^{۵-۲}=\binom{۵}{۲}(۰/۵)^{۲}(۰/۵)^{۳}=۰/۳۱$$

ب:

$$E(X) = np = ۵\times ۰/۵ = ۲/۵$$
$$Var(X) = npq = ۵\times ۰/۵\times ۰/۵ = ۱/۲۵$$

مثال۵: در یک امتحان تستی که در آن ۲۰ سوال چهارگزینه‌ای داده شده است، دانشجویی تمامی سوالات را شانسی پاسخ می‌دهد.مطلوب است محاسبه:

الف- این دانشجو به طور متوسط چند سوال را پاسخ می‌دهد؟

ب- انحراف معیار تعداد پاسخ‌های صحیح چقدر است؟

پاسخ: اگر متغیر تصادفی X را تعداد سوالاتی که پاسخ صحیح می‌دهد و پیروزی را دادن پاسخ صحیح، تعریف کنیم، آنگاه: $X \sim Bin\left(n = ۲۰, p = \frac{۱}{۴}\right)$.

الف: این دانشجو به طور متوسط به ۵ سوال پاسخ صحیح می‌دهد. $E(X) = np = ۲۰\times\frac{۱}{۴} = ۵$

ب: $\sigma_X = \sqrt{\sigma_X^۲} = \sqrt{Var(X)} = \sqrt{np(۱-p)} = \sqrt{۲۰\times\frac{۱}{۴}\times\frac{۳}{۴}} = ۱/۹۴$

مثال٦: یک تولید کننده وسائل برقی، کالاهای فروخته شده را به مدت ٦ ماه ضمانت می‌کند. در یک هفته ١٦ دستگاه از این وسائل برقی به فروش رفته است. اگر بدانیم کالایی در مدت ضمانت با احتمال ۲/. احتیاج به تعمیر دارد، مطلوب است محاسبه:

الف- احتمال اینکه هیچیک از این دستگاه‌ها در دوره ضمانت خراب نشوند؟

ب- احتمال اینکه حداقل یک دستگاه خراب شود؟

ج- احتمال اینکه دقیقاً ۲ دستگاه خراب شود؟

پاسخ: اگر متغیر تصادفی X را تعداد کالاهای خراب شده و پیروزی را خراب شدن کالا، تعریف کنیم، آنگاه: $X \sim Bin(n = ۱۶, p = ./۲)$.

الف:

$$P(X = x) = \binom{۱۶}{x}(۰/۲)^x(۱-۰/۲)^{۱۶-x} \qquad x = ۰,۱,۲,\ldots,۵,۱۶$$
$$\Rightarrow \quad P(X = ۰) = \binom{۱۶}{۰}(۰/۲)^۰(۱-۰/۲)^{۱۶-۰} = \binom{۱۶}{۰}(۰/۲)^۰(۰/۸)^{۱۶} \simeq ۰/۰۲۸$$

ب: در اینگونه مسائل که محاسبه احتمال مستقیم شامل جملات زیادی است استفاده از احتمال متمم آسانتر خواهد بود. $P(X \geq ۱) = ۱ - P(X = ۰) = ۱ - ۰/۰۲۸ = ۰/۹۷۲$

ج: $$P(X = ۲) = \binom{۱۶}{۲}(۰/۲)^{۲}(۱-۰/۲)^{۱۶-۲} = \binom{۱۶}{۲}(۰/۲)^{۲}(۰/۸)^{۱۴} \simeq ۰/۲۱$$

مثال۷: در یک خانواده ۵ فرزندی با فرض اینکه احتمال داشتن دختر و پسر یکسان باشد. مطلوب است محاسبه احتمال اینکه آن خانواده:

الف- دارای یک فرزند پسر باشد؟

ب- دارای حداقل یک فرزند پسر باشد؟

ج- دارای حداکثر ۲ فرزند پسر باشد؟

پاسخ: اگر متغیر تصادفی X را تعداد فرزندان پسر و پیروزی را پسر بودن، تعریف کنیم، آنگاه: $X \sim Bin(n = ۱۶, p = ./۲)$.

الف:

$$P(X = x) = \binom{۵}{x}(۰/۵)^{x}(۱-۰/۵)^{۵-x} \qquad x = ۰,۱,۲,۳,۴,۵$$

$$\Rightarrow \quad P(X = ۱) = \binom{۵}{۱}(۰/۵)^{۱}(۱-۰/۵)^{۵-۱} = \binom{۵}{۱}(۰/۵)^{۱}(۰/۵)^{۴} \simeq ۰/۱۵$$

ب: $$P(X \geq ۱) = ۱ - P(X < ۱) = ۱ - P(X = ۰) = ۱ - \left[\binom{۵}{۰}(۰/۵)^{۰}(۰/۵)^{۵}\right] \simeq ۰/۹۶$$

ج:

$$\begin{aligned} P(X \leq ۲) &= P(X = ۰) + p(X = ۱) + p(X = ۲) \\ &= \left(\binom{۵}{۰}(۰/۵)^{۰}(۰/۵)^{۵}\right) + \left(\binom{۵}{۱}(۰/۵)^{۱}(۰/۵)^{۴}\right) + \left(\binom{۵}{۲}(۰/۵)^{۲}(۰/۵)^{۳}\right) \\ &= ۰/۰۳ + ۰/۱۵ + ۰/۳ = ۰/۴۸ \end{aligned}$$

توزیع پواسن:

توزیع پواسن یک توزیع گسسته است که برای محاسبه تعداد پیشامدها در یک فاصله زمانی یا مکانی خاص به کار می‌رود. برای مثال: تعداد تصادفات در یک چهارراه، تعداد غلط‌ها در یک صفحه، تعداد تلفن‌ها به مرکز ۱۱۵ و ... .

اگر متغیر تصادفی X را تعداد پیشامدها در یک فاصله زمانی یا مکانی خاص و نرخ متوسط تعداد پیشامدها را در یک فاصله زمانی یا مکانی با λ نمایش دهیم، آنگاه می‌گوییم X دارای توزیع پواسن با میانگین λ است و به صورت $X \sim Pos(\lambda)$ نمایش می‌دهیم و تابع توزیع احتمال به صورت زیر خواهد بود:

$$P(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad X=۰,۱,۲,....$$

$$E(X)=Var(X)=\lambda$$

نکته۱: توزیع پواسون تنها توزیعی است که میانگین و واریانس آن با هم برابر هستند.

نکته۲: در حل مسائل توجه داشته باشید که بازه زمانی یا مکانی تعریف شده برای X و λ بایستی حتماً با هم برابر باشند. در غیر این صورت بایستی λ را بر حسب بازه زمانی یا مکانی خواسته شده برای X، به دست آوریم.

مثال۸: اگر تعداد مراجعات به یک باجه بلیط به طور متوسط در طی یک ساعت ٤ نفر باشد. احتمال اینکه در یک ساعت خاص دقیقاً ۲ نفر به این باجه مراجعه کنند، چقدر است؟

پاسخ: متغیر تصادفی X را تعداد مراجعین به باجه بانک در یک ساعت تعریف می‌کنیم. از مسئله نیز نرخ متوسط تعداد مراجعات به باجه بانک در یک ساعت برابر ٤ می‌باشد. در نتیجه داریم:

$$X \sim Pos(۴) \Rightarrow P(X=x)=\frac{e^{-۴}۴^{x}}{x!} \qquad x=۰,۱,۲,...$$

$$\Rightarrow \quad P(X=۲)=\frac{e^{-۴}۴^{۲}}{۲!}=۸e^{-۴}$$

مثال۹: یک دستگاه چاپگر کامپیوتری به طور متوسط هر ماه دو بار چک می‌شود. مطلوب است محاسبه:

الف- احتمال اینکه در یک ماه کمتر از ۲ بار چک شود؟

ب- احتمال اینکه در ۳ ماه، این چاپگر دقیقاً ۳ بار چک شود؟

ج- احتمال اینکه در ۳ ماه، این چاپگر حداقل ۲ بار چک شود؟

پاسخ: متغیر تصادفی X را تعداد چک شدن‌های چاپگر در یک ماه تعریف می‌کنیم. از مسئله نیز نرخ متوسط تعداد چک کردن‌های چاپگر در یک ماه برابر ۲ می‌باشد.

الف:

$$X \sim Pos(۲) \Rightarrow P(X=x)=\frac{e^{-۲}۲^{x}}{x!} \qquad x=۰,۱,۲,...$$

$$\Rightarrow \quad P(X<۲)=P(x=۰)+p(X=۱)=\left(\frac{e^{-۲}۲^{۰}}{۰!}\right)+\left(\frac{e^{-۲}۲^{۱}}{۱!}\right)=e^{-۲}+۲e^{-۲}=۳e^{-۲}$$

ب: در این قسمت توجه داشته باشید که بازه زمانی تعریف شده برابر ۳ ماه است که با بازه زمانی یک ماهه تعریف شده برای λ، یکسان نیست. در نتیجه مقدار λ را برای یک بازه زمانی ۳ ماه به صورت زیر محاسبه خواهیم کرد:

مقدار λ	بازه زمانی(ماه)
۲	۱
$\lambda=?$	۳

$$\Rightarrow \quad \lambda=\frac{۲\times۳}{۱}=۶$$

حال با توجه به مقدار λی جدید داریم:

$$P(X=۳)=\frac{e^{-۶}۶^{۳}}{۳!}=۳۶e^{-۶}$$

ج:

$$P(X\geq۲)=۱-P(X<۲)=۱-P(X\leq۱)=۱-\left[P(X=۰)+P(X=۱)\right]$$

$$=۱-\left[\frac{e^{-۶}۶^{۰}}{۰!}+\frac{e^{-۶}۶^{۱}}{۱!}\right]=۱-۷e^{-۶}$$

مثال۱۰: به طور متوسط در هر صفحه از یک کتاب مشخص، ۲ غلط تایپی وجود دارد. مطلوب است محاسبه:

الف- احتمال اینکه در یک صفحه، حداکثر ۳ غلط تایپی وجود داشته باشد؟

ب- در ۲ صفحه‌ی مشخص، حداقل دو غلط تایپی وجود داشته باشد؟

پاسخ: متغیر تصادفی X را تعداد غلط‌ها در یک صفحه تعریف می‌کنیم. با توجه به صورت مسئله $\lambda=۲$ برای هر صفحه است.

الف:

$$P(X\le 3)=P(X=0)+P(X=1)+P(X=2)+P(X=3)$$
$$=\frac{e^{-2}2^{0}}{0!}+\frac{e^{-2}2^{1}}{1!}+\frac{e^{-2}2^{2}}{2!}+\frac{e^{-2}2^{3}}{3!}=\frac{19}{3}e^{-2}$$

ب-

مقدار λ	بازه مکانی(صفحه)
۲	۱
$\lambda=?$	۲

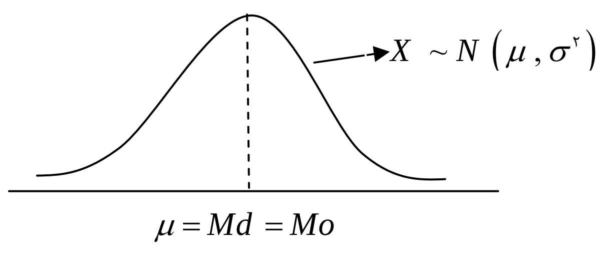
$$\Rightarrow\quad \lambda=\frac{2\times 2}{1}=4$$

$$P(X\ge 2)=1-P(X\le 1)=1-\left[P(X=0)+P(X=1)\right]$$
$$=1-\left[\frac{e^{-4}4^{0}}{0!}+\frac{e^{-4}4^{1}}{1!}\right]=1-5e^{-4}$$

توزیع نرمال:

اگر X یک متغیر پیوسته باشد، احتمال در یک نقطه برابر صفر است. به عنوان مثال احتمال اینکه قد فردی دقیقاً برابر با ۱۶۰ سانتی متر باشد، برابر صفر است. متغیرهای تصادفی پیوسته دارای توزیع‌های منحنی شکل هستند. مهمترین توزیع پیوسته که بسیاری از وقایع زندگی روزمره از این توزیع پیروی می‌کنند و مهمترین توزیع آماری می‌باشد، توزیع نرمال است. توزیع نرمال با دانستن میانگین و واریانس مشخص می‌شود. اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد. آن را به صورت $X\sim N(\mu,\sigma^2)$ نشان می‌دهیم.

منحنی توزیع نرمال زنگوله‌ای شکل و نسبت به میانگین آن، متقارن است.



محاسبه $P(a<X<b)$:

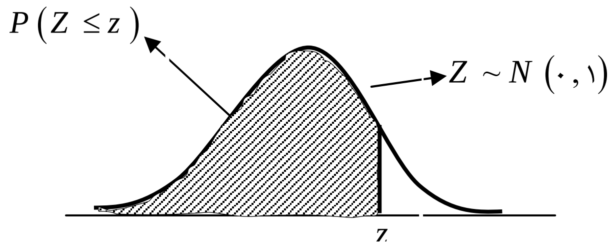
همانطور که قبلاً نیز گفته شد، برای متغیرهای تصادفی پیوسته احتمال یک نقطه برابر صفر است. برای متغیرهای پیوسته احتمال برای یک بازه محاسبه خواهد شد و معنی‌دار است.

اگر X یک متغیر تصادفی نرمال با میانگین μ و واریانس $\sigma^{۲}$ باشد آنگاه $P\left(a<X<b\right)$، برابر با مساحت سطح زیر منحنی نرمال بین دو خط $x=a$ و $x=b$ است. محاسبه این مساحت کاری مشکل و بدون استفاده از نرم‌افزار غیرممکن است. برای رفع این مشکل، آماردانان جدولی را برای محاسبه این احتمال، برای تابع چگالی نرمالی با میانگین صفر و واریانس یک، که به توزیع نرمال استاندارد معروف استف تشکیل داده‌اند. بنابراین برای محاسبه احتمال متغیرهای تصادفی نرمال، ابتدا آنها را تبدیل به نرمال استاندارد کرده و سپس از جدول توزیع نرمال استاندارد که در پایان کتاب آورده شده است، به راحتی این احتمالات را محاسبه می‌کنیم.

نرمال استاندارد: اگر X هر متغیر تصادفی نرمال با میانگین μ و انحراف معیار σ باشد، آنگاه با تغییر متغیر $Z=\dfrac{X-\mu}{\sigma}$، متغیر تصادفی نرمال دیگری به دست می‌آید که میانگین آن صفر و واریانس آن یک است(نرمال استاندارد).

در جدول نرمال استاندارد، احتمالات را برای مقادیر متعلق به بازه $[-۳/۴,۳/۴]$ دارا می‌باشد.

در جدول نرمال استاندارد، به ازای یک مقدار همانند $z\in[-۳/۴,۳/۴]$، سطح زیر منحنی نرمال و سمت چپ خط $x=z$ داده شده است، یعنی: $P\left(Z\le z\right)$.



برای استفاده از جدول نرمال استاندارد دانستن دو نکته زیر مفید می‌باشد:

$$۱)\ P\left(Z>z\right)=۱-P\left(Z<z\right)$$
$$۲)\ P\left(z_{۱}<Z<z_{۲}\right)=P\left(Z<z_{۲}\right)-P\left(Z<z_{۱}\right)$$

مثال۱۱: با استفاده از جدول نرمال استاندارد برای احتمالات زیر داریم:

$$P(Z<۱/۹۶)=۰/۹۷۵$$
$$P(Z<-۲/۱۳)=۰/۰۱۶۶$$
$$P(Z>۳/۰۴)=۱-P(Z<۳/۰۴)=۱-۰/۹۹۸۸=۰/۰۰۱۲$$
$$P(۰/۸۷<Z<۱/۲۸)=P(Z<۱/۲۸)-P(Z<۰/۸۷)=۰/۸۹۹۷-۰/۸۰۷۸=۰/۰۹۱۹$$
$$P(-۰/۳۴<Z<۰/۶۲)=P(Z<۰/۶۲)-P(Z<-۰/۳۴)=۰/۷۳۲۴-۰/۳۶۶۹=۰/۳۶۵۵$$

حال فرض کنید که X، یک متغیر تصادفی نرمال با میانگین μ و واریانس $\sigma^۲$ است. برای محاسبه $P(a<X<b)$، ابتدا آن را به صورت زیر تبدیل به نرمال استاندارد خواهیم کرد: $P(a<X<b)=P\left(\frac{a-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{b-\mu}{\sigma}\right)=P\left(\frac{a-\mu}{\sigma}<Z<\frac{b-\mu}{\sigma}\right)$. که در آن طبق تعریف، Z یک متغیر تصادفی نرمال استاندارد است و احتمال به راحتی از جدول قابل محاسبه است.

مثال۱۲: اگر متغیر تصادفی X، دارای توزیع نرمال با میانگین ۱۰ و واریانس ۱٦ باشد. مطلوب است محاسبه:

الف- $P(۸<X<۱۵)$

ب- $P(X>۹)$

الف:
$$P(۸<X<۱۵)=P\left(\frac{۸-۱۰}{۴}<\frac{X-۱۰}{۴}<\frac{۱۵-۱۰}{۴}\right)=P(-۰/۵<Z<۱/۲۵)$$
$$=P(Z<۱/۲۵)-P(Z<-۰/۵)=۰/۸۹۴۴-۰/۳۰۸۵=۰/۵۸۵۹$$

ب:
$$P(X>۹)=P\left(\frac{X-۱۰}{۴}>\frac{۹-۱۰}{۴}\right)=P(Z>-۰/۲۵)=۱-P(Z<۰/۲۵)=۱-۰/۵۹۸۷=۰/۴۰۱۳$$

مثال۱۳: نمرات دانشجویی دارای توزیع نرمال با میانگین ۵۰۰ و انحراف معیار ۱۰۰ است. مطلوب است:

الف- $P(۶۰۰<X<۷۰۰)$

ب- $P(X<۶۰۰)$

الف:
$$P(۶۰۰<X<۷۰۰)=P\left(\frac{۶۰۰-۵۰۰}{۱۰۰}<\frac{X-۵۰۰}{۱۰۰}<\frac{۷۰۰-۵۰۰}{۱۰۰}\right)=P(-۱<Z<۲)$$
$$=P(Z<۲)-P(Z<-۱)=۰/۹۷۷۲-۰/۸۴۱۳=۰/۱۳۵۹$$

ب:
$$P(X<۶۰۰)=P\left(\frac{X-۵۰۰}{۱۰۰}<\frac{۶۰۰-۵۰۰}{۱۰۰}\right)=P(Z<۱)=۰/۸۴۱۳$$

مسائل

۱- اگر ۲۰٪ از قطعات تولیدی یک ماشین معیوب باشند و از بین قطعات تولیدی ٤ قطعه را به تصادف اتخاب کنیم. مطلوب محاسبه احتمال اینکه:

الف- یک قطعه معیوب باشد؟

ب- هیچ قطعه‌ای معیوب نباشد؟

ج- حداکثر دو قطعه معیوب باشند؟

۲- تجربه نشان داده که ۱۰٪ از لامپ‌های تولیدی یک کارخانه معیوب هستند. به طور تصادفی ۱۰ عدد لامپ را انتخاب می‌کنیم. مطلوب است محاسبه احتمال اینکه:

الف- همه سالم باشند؟

ب- همه معیوب باشند؟

ج- فقط یکی سالم باشد؟

۳- در بیمارستانی ٤ موتور برق اضطراری وجود دارد که احتمال کار کردن هر کدام مستقل از هم و برابر ۰/۹ است. مطلوب است محاسبه احتمال اینکه:

الف- حداقل یک موتور برق کار کند؟

ب- دقیقاً دو موتور برق کار کند؟

٤- اگر X یک متغیر تصادفی دوجمله‌ای با میانگین ٦ و واریانس ٤/٥ باشد. توزیع احتمال X را تعیین کنید؟ (راهنمایی: مقادیر n و p را به دست آورید)

۵- فرض کنید متغیر تصادفی X دارای توزیع نرمال با میانگین ۲ و انحراف معیار ۱/۵ باشد. $P\left(۳/۱۴ < X < ۵/۵\right)$ را محاسبه کنید؟

٦- اگر توزیع مقدار قند خون افراد جامعه‌ای نرمال با میانگین ۱۰۰ و انحراف معیار ۲/۵ سانتی‌گرم در لیتر باشد. مطلوب است محاسبه احتمال اینکه قند خون یک نفر از این جامعه در فاصله (۹۵, ۱۰۲/۵) سانتی‌گرم در لیتر باشد؟

۷- میانگین قد افراد یک جامعه دارای توزیع نرمال با میانگین ۱٦۷ سانتی‌متر و انحراف معیار ۵ سانتی‌متر است. مطلوب است محاسبه:

الف- احتمال اینکه قد فردی بیشتر از ۱۸۰ سانتی‌متر باشد؟

ب- احتمال اینکه قد فردی بین ۱۷۰ و ۱۸۰ سانتی‌متر باشد؟

ج- احتمال اینکه قد فردی کمتر از ۱۵۷ سانتی‌متر باشد؟

۸- اگر قطر پیستون‌هایی که توسط یک ماشین ساخته می‌شود دارای توزیع نرمال با میانگین ۵۰ میلی‌متر و انحراف معیار ۰/۵ میلی‌متر باشند:

الف- احتمال اینکه قطر پیستونی بین (۵۱ , ۵۰/۵) میلی‌متر باشد، چقدر است؟

ب- اگر پیستون‌هایی که قطر آنها کمتر از ۴۹/۵ میلی‌متر باشد، معیوب به حساب آیند، در بین ۱۰۰۰ پیستون، چند پیستون معیوب خواهد بود؟(راهنمایی: $P\left(X < ۴۹/۵\right)$ را به دست آورده و در ۱۰۰۰ ضرب کنید.)