

به نام خدا

آمار و احتمالات و کاربرد آن در مدیریت 1

Statistics and Probabilities

&

Applications in Management

دانشگاه آزاد اسلامی

واحد علوم تحقیقات

منابع:

آمار و کاربرد آن در مدیریت، عادل آذر، جلد اول، انتشارات سمت

آمار و احتمال در مهندسی و علم مدیریت، ویلیام و هاینز و داگلاس س مونتگومری، مترجم
محمد صالح اولیاء، انتشارات دانشگاه یزد

مقدمه:

علم آمار از جمله علومى است كه در قرن اخير به سرعت توسعه يافته و دامنه كاربرد آن تقريباً به تمام علوم ديگر كشيده شده است. اگر اطلاعات را مبنای تصميم گيرى در مديريت بدانيم؛ علم آمار كه وظيفه اصلى تبديل داده ها به اطلاعات را بر عهده دارد، بيشترين سهم را در تصميم گيرى مديران دارد.

امروزه به ندرت مى توان يك تحقيق علمى و آكادميك را يافت كه در آن از روش ها و فنون آمارى استفاده نشده باشد. تجزيه و تحليل اطلاعات و يافتن الگوهاى رفتارى داده هاى مورد مطالعه از ديگر کاربردهاى آمار است كه در طبقه بندي اطلاعات، كنترل كيفيت، پيش بينى آينده، بررسى صحت و سقم فرضيات علمى و بررسى اعتبار و پاىايى تحقيقات علمى مورد استفاده قرار مى گيرد.

با توسعه كاربرد كامپيوتر در آمار، کاربردهاى ديگرى همچون داده كاوى، شبیه سازی و هوش مصنوعى كه در آنها از داده هاى تصادفى براى مدل سازى بهره برده مى شود نيز به ساير کاربردهاى آمار افزوده شد.

ايميل ارسال تمرينات و پروژه ها:

srbiau.stat1@gmail.com

سرفصل‌های درس:

فصل اول: کلیات و مبانی

✓ آمار و ضرورت آن در مدیریت

✓ جامعه و نمونه

✓ سیر تحول در آمار

✓ مراحل پژوهش علمی در آمار

✓ عناصر اساسی در تحقیقات آماری

✓ آیا آمار دروغ می‌گوید؟

فصل دوم: مطالعه توصیفی داده‌های طبقه‌بندی نشده

✓ نمادگذاری مجموعه مشاهدات و عمل جمع

✓ پارامترهای مرکزی

✓ پارامترهای براکندگی

فصل سوم: طبقه‌بندی و توصیف هندسی مشاهدات

مع آما

✓ طبقه‌بندی و سازماندهی مشاهدات

✓ طبقه‌بندی مشاهدات نابیوسته

✓ توزیع فراوانی نسبی

✓ توزیع فراوانی تجمعی

✓ نمایش هندسی مشاهدات

شد طبقه‌بندی فصل چهارم: توصیف مقداری مشاهدات

ه

✓ مقدمه

✓ پارامترهای مرکزی

✓ پارامترهای براکندگی

✓ عملیات جبری میانگین و واریانس

✓ اهمیت کاربردی انحراف معیار

✓ پارامترهای تعیین انحراف از قرینگی

✓ پارامترهای تعیین انحراف از کشیدگی

ادامه سرفصل‌های درس:

فصل پنجم: احتمال

- ✓ نظریه احتمال
- ✓ برخی از مفاهیم اساسی احتمال
- ✓ احتمال یک پیشامد
- ✓ قواعد شمارش
- ✓ عملیات روی پیشامدها و قواعد احتمال
- ✓ احتمال شرطی
- ✓ قضیه بیز (احتمالات پیشین و پسین)

فصل ششم: توابع احتمال گسسته

- ✓ متغیر تصادفی گسسته، تابع احتمال و تابع توزیع
- ✓ امید ریاضی و واریانس متغیر تصادفی
- ✓ تابع احتمال توأم
- ✓ کوواریانس و استقلال دو متغیر تصادفی
- ✓ توزیع برنولی
- ✓ توزیع دوجمله‌ای
- ✓ توزیع‌های دو حمله‌ای منفی و هندسی
- ✓ توزیع چند جمله‌ای
- ✓ توزیع فوق هندسی
- ✓ توزیع بواسون

فصل هفتم: توابع احتمال پیوسته

- ✓ تابع چگالی احتمال و تابع توزیع متغیر تصادفی پیوسته
- ✓ امید ریاضی و واریانس متغیر تصادفی پیوسته
- ✓ توزیع یکنواخت
- ✓ توزیع نمایی

فصل هشتم: توزیع نرمال

- ✓ توزیع نرمال چیست؟
- ✓ استفاده مستقیم از جدول توزیع نرمال استاندارد
- ✓ استفاده معکوس از جدول توزیع نرمال استاندارد
- ✓ روش برای تشخیص نرمال بودن توزیع
- ✓ تقریب توزیع دوجمله‌ای به وسیله توزیع نرمال
- ✓ تقریب توزیع بواسون به وسیله توزیع نرمال
- ✓ قضیه حد مرکزی

(جلد دوم کتاب آمار) فصل نهم: فنون تصمیم‌گیری آماری

- ✓ شش گام در نظریه تصمیم
- ✓ انواع شرایط تصمیم‌گیری
- ✓ تصمیم‌گیری در شرایط عدم اطمینان
- ✓ تصمیم‌گیری در شرایط ریسک
- ✓ ارزش مورد انتظار اطلاعات کامل
- ✓ استفاده از اطلاعات نمونه (قضیه بیز)



فصل اول

کلیات و مبانی

Generalities and Foundations



آمار و ضرورت آن در مدیریت

کلمه آمار دو معنای کاملاً متفاوت دارد. در یک معنا عبارت است از روشی علمی که برای جمع‌آوری، تلخیص، تجزیه و تحلیل، تفسیر و به طور کلی مطالعه و بررسی مشاهدات به کار برده می‌شود. در معنای دیگر با عبارت «مشاهدات عددی» مترادف است. به طوری که از دیرباز معنای آمار با اعداد و ارقام توأم بوده است.

در ابتدا آمار به منظور فراهم نمودن اطلاعاتی که دولتمردان را در کارهای حکومتی یاری می‌کرد، استفاده می‌شد. این نوع اطلاعات از زمان ارسطو با عنوان «مسایل ایالتی» وجود داشته است. در واقع می‌توان گفت که کلمات «Statistic» و «State» از یک ریشه مشتق شده‌اند.

تصور بسیاری از مردم در مورد آمار به صورت جداول بزرگی از اعداد و نمودارهای مبهم و مبهم کننده، می‌باشد. در حالیکه نظریه و روش‌های جدید آماری، از حد ساختن جداول و نمودارها فراتر رفته و آمار به عنوان یک علم شامل مفاهیم و روش‌هایی است که در تمام پژوهش‌هایی که مستلزم جمع‌آوری داده (نه الزاماً اعداد) به وسیله فرایند آزمایش و مشاهده، استنباط و تجزیه و تحلیل داده‌ها انجام می‌شوند، نقش‌آفرینی می‌کند.



جامعه و نمونه

جامعه بزرگترین مجموعه موجودات (نه الزاما زنده) است که در یک زمان معین مورد مطالعه قرار می‌گیرند. به عنوان مثال نمرات دانشجویان دانشکده مدیریت دانشگاه آزاد اسلامی واحد علوم تحقیقات، یک جامعه آماری می‌باشد.

جامعه آماری عبارتست از: **«تعدادی از عناصر مطلوب مورد نظر که حداقل دارای یک صفت مشخصه باشند.»** هر جامعه بر حسب تعداد عناصر آن می‌تواند محدود یا نامحدود باشد.

هر بخشی از جامعه آماری که مورد مطالعه خاص قرار می‌گیرد، نمونه نام دارد. به عنوان مثال نمرات دانشجویان رشته مدیریت بازرگانی دانشکده مدیریت دانشگاه آزاد اسلامی واحد علوم تحقیقات، یک نمونه از جامعه فوق‌الذکر است.

نمونه عبارت است از: **«تعداد محدودی از عناصر جامعه آماری که بیان‌کننده (معرف) ویژگی‌های اصلی یک جامعه باشد.»** به این ترتیب از یک جامعه آماری می‌توان نمونه‌های متعددی استخراج کرد.

شاخص‌هایی که با اندازه‌گیری تمام عناصر جامعه (سرشماری) به دست می‌آیند را **«پارامتر»** و شاخص‌هایی که از اندازه‌گیری تمام عناصر یک نمونه آماری به دست می‌آیند را **«آماره»** می‌نامند.



سیر تحول در آمار

با اینکه آمار از دیرباز مورد استفاده بشر بوده، اما شاید بتوان پیشرفت‌های شایان توجه آن را به قرن شانزدهم مربوط دانست که در آن آمار حیاتی، بیمه و اقتصاد مورد مطالعه قرار گرفتند. در سال 1733 توزیع نرمال برای اولین بار به وسیله «دموآور» تعریف شد و بعد از آن «گوس» و «لاپلاس» آن را توسعه دادند. در ربع آخر قرن نوزدهم «فرانسیس گالتون» و «کارل پیرسون» با طرح مسایل ژنتیک، نظرات وابسته به رگرسیون و همبستگی را گسترش دادند. مهم‌ترین ویژگی علم آمار در اوایل قرن بیستم به «استنباط» یا «تخمین» آماری مربوط می‌شود و پس از آن توسعه «نظریه نمونه‌گیری» صورت گرفت که به کمک آن بتوان خواص نمونه که از آماره‌ها استخراج می‌شود را به پارامترهای جامعه آماری تعمیم داد. «توزیع نمونه‌برداری» و «حدود اطمینان» در آمار استنباطی و بیان «فرض‌های آماری» و روش آزمودن آن‌ها نیز از پیشرفت‌های دیگر آمار در قرن بیستم بودند. پیشرفت مهم دیگر «طراحی آزمایش‌های آماری» بود. رواج آمار در رشته‌های غیرکمی مثل علوم انسانی در اوایل قرن حاضر با مقالات ویلیام گاست با نام مستعار استیودنت در مورد روش‌های تجزیه آماری، به خصوص تحلیل و تفسیر مشاهداتی که از طریق نمونه‌گیری به دست می‌آیند، صورت گرفت. تجدید بنای مهم‌تر در آمار بعد از جنگ جهانی دوم، مربوط به آمار ناپارامتریک است که در آن مفروضات آمار کلاسیک مانند برخورداری از توزیع نرمال، ضروری نیستند.



ادامه سیر تحول در آمار

از لحاظ موضوعی سیر تحول آمار به سه مرحله تقسیم می‌شود.

1- آمار توصیفی: توصیف جامعه با هدف محاسبه پارامترهای جامعه

2- آمار استنباطی: تعمیم مقادیر آماره‌های محاسبه شده از نمونه آماری به مقادیر پارامترهای جامعه

آماري

3- آمار ناپارامتریک: مطالعه و بررسی رفتار متغیرهایی که با مقیاس‌های کیفی سنجیده شده‌اند و فاقد

توزیع آماری یا آزاد از توزیع هستند



مراحل پژوهش علمی در آمار

✓ مشخص کردن هدف: اثبات نظریه جدید، بررسی نظریه موجود، انعکاس وضع موجود، تعیین امکانات و کاربرد آنها

✓ جمع‌آوری داده‌ها: با توجه به هدف تحقیق ممکن است شامل آزمایش‌های پیچیده در شرایط کنترل شده باشد.

✓ تجزیه و تحلیل داده‌ها: به کار بستن فنون آماری روی داده‌ها برای بررسی معلومات جدید و تعیین نقاط قوت و ضعف

✓ بیان یافته‌ها: پاسخ به سوالاتی همچون: «آیا فرضیه یا نظریه موجود با داده‌ها در تناقض است؟»، «آیا داده‌ها نظریه جدیدی برای تبیین پدیده مورد مطالعه القا می‌کنند؟» یا «میزان عدم قطعیت در یافته‌های تحقیق چقدر است؟»



عناصر اساسی در تحقیقات آماری

دو عنصر اساسی «متغیرهایی برای آزمودن فرضیه‌ها» و «فرضیه‌های تحقیق» در تحقیقات آماری وجود دارند.

1- متغیرهای خصیصه‌ها: (مقدار آن از یک عضو به عضو دیگر جامعه ممکن است تغییر کند).

متغیر مستقل (درونداد یا محرک)، متغیر وابسته (برونداد یا پاسخ)، متغیر تعدیل کننده، متغیر کنترل

مقیاس‌های اندازه‌گیری متغیرها:

مقیاس اسمی، مقیاس ترتیبی (رتبه‌ای)، مقیاس فاصله‌ای، مقیاس نسبی (نسبی)

مراتب نوع مقیاس	ترتیب	فواصل	مبدأ صفر قراردادی	مبدأ صفر واقعی (مطلق)
اسمی	ندارد	ندارد	ندارد	ندارد
رتبه‌ای	دارد	ندارد	ندارد	ندارد
فاصله‌ای	دارد	دارد	دارد	ندارد
نسبی	دارد	دارد	دارد	دارد

تمرین 1: انواع متغیرها و انواع مقیاس‌ها را با ذکر توضیح و مثال



ادامه عناصر اساسی در تحقیقات آماری

2- فرضیه‌های تحقیق: «فرضیه» عبارتی آزمایشی است که رابطه قابل انتظار بین دو یا چند متغیر را به صورت دقیق و روشن بیان می‌کند و بیانگر حدس هوشمندانه محقق در یک تحقیق یا پژوهش است.
ویژگی‌های فرضیه:

1- واضح و بدون ابهام و بیشتر در قالب جملات خبری 2- قابل تبیین (علتیابی)

3- بیان‌کننده رابطه مورد انتظار بین متغیرها 4- قابل آزمودن

انواع فرضیه‌ها:

1- توصیفی در مقابل استنباطی 2- تک متغیره در مقابل چند متغیره 3- همبستگی در مقابل تجربی

4- با گروه‌های جور شده در مقابل گروه‌های مستقل 5- پارامتریک در مقابل ناپارامتریک



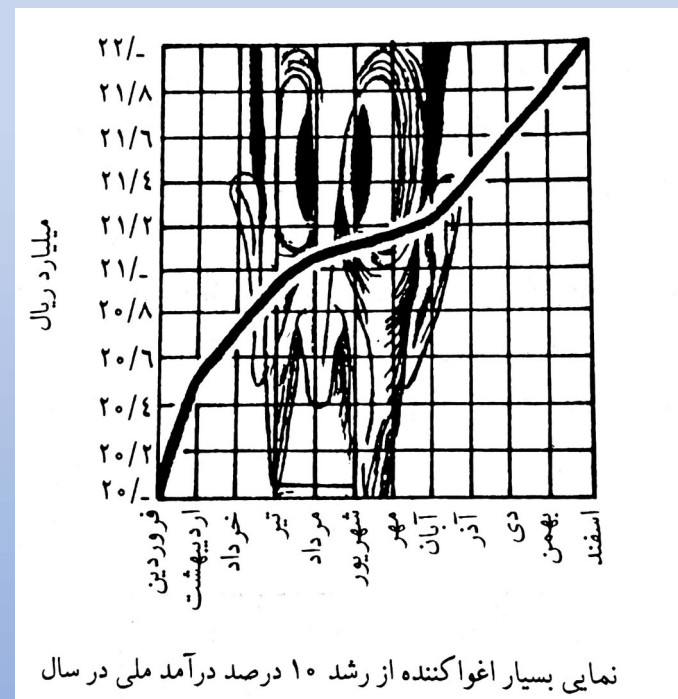
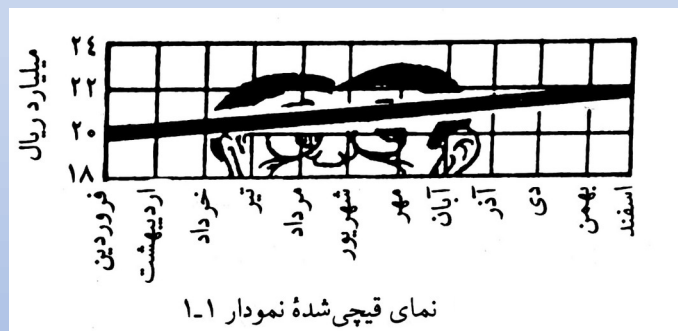
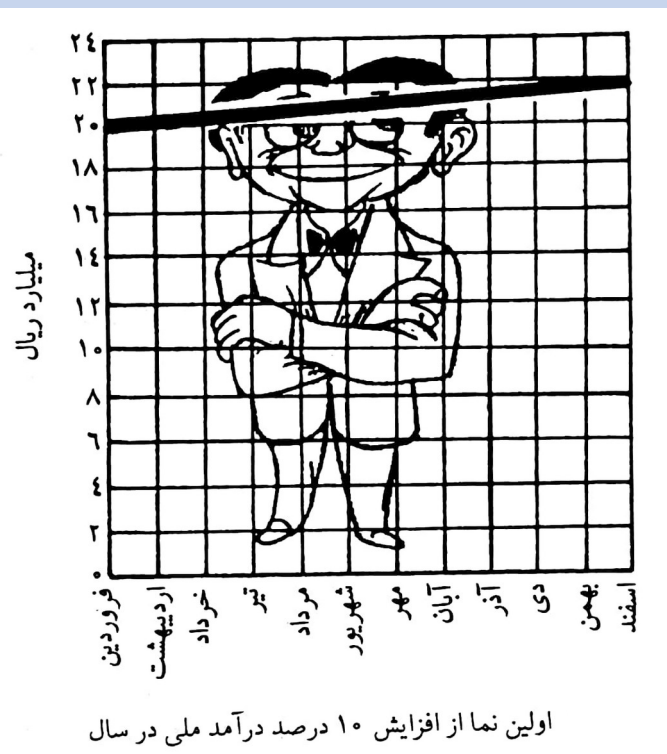
ادامه عناصر اساسی در تحقیقات آماری

پارامتر یا ناپارامتریک	گروه‌های جورشد یا مستقل	همبستگی یا تجربی	تک متغیره یا چندمتغیره	توصیفی یا استنباطی	فرضیه
؟			تک متغیره	توصیفی	(1) متوسط نمره مسوولیت‌پذیری مدیران در ایران حداقل 50 است.
			تک متغیره	توصیفی / استنباطی	(2) بیش از 60 درصد کارکنان سازمان «الف» از کار خود راضی هستند.
		همبستگی	دومتغیره		(3) بین میزان ارضای نیازهای اساسی کارکنان سازمان «الف» و عملکرد آن‌ها ارتباط وجود دارد.
		تجربی	دومتغیره		(4) روش آزمون متمرکز مدیران از نظر اثربخشی هیچ تفاوتی با روش غیرمتمرکز ندارد.
			دومتغیره		(5) دوره‌های ضمن خدمت موجب افزایش مهارت‌های انسانی در مدیران می‌گردد.
			چندمتغیره		(6) عوامل نگهدارنده هرزبرگ به طور یکسان بر تمایل به ترک خدمت موثر هستند.
ناپارامتریک					(7) نسبت رضایت از روش‌های ارزشیابی کارکنان در تمامی وزارتخانه‌های کشور مساوی است.
ناپارامتریک					(8) بلوغ پیروان در سازمان‌های کشور از توزیع نرمال برخوردار است.



آیا آمار دروغ می گوید؟

ارقام دروغ نمی گویند اما دروغگویان رقم سازی می کنند.





فصل دوم

مطالعه توصیفی داده‌های طبقه‌بندی

نشده

Unclassified Data Descriptive Study



نمادگذاری مجموعه مشاهدات و عمل جمع

برای جمع بستن n عدد $a_1, a_2, a_3, \dots, a_n$ می‌توان از نماد خلاصه شده $\sum$ (سیگما) استفاده نمود. به این ترتیب عملیات ریاضی را می‌توان به صورت مختصر و فشرده نوشت.

خواص نماد جمع‌بندی :

$$1) \sum_{i=1}^n a_i = a_1 + a_2 + a_3 + \dots + a_n$$

$$2) \sum_{i=1}^n (k a_i) = k \sum_{i=1}^n a_i \quad (k \in R)$$

$$3) \sum_{i=1}^n (a_i \pm b_i) = \sum_{i=1}^n a_i \pm \sum_{i=1}^n b_i$$

$$4) \sum_{i=1}^n k = nk$$

$$5) \sum_{i=1}^n a_i = \sum_{j=1}^n a_j = \sum_{k=1}^n a_k = \dots$$

$$6) \sum_{i=1}^n \left(\sum_{j=1}^m a_{ij} \right) = \sum_{j=1}^m \left(\sum_{i=1}^n a_{ij} \right)$$

$$7) \sum_{i=1}^n (a_i - a_{i+1}) = a_1 - a_{n+1}$$



پارامترهای مرکزی

در یک جامعه آماری، پارامترهایی که معرف مرکز داده‌ها باشند به طوریکه سایر داده‌های جامعه در اطراف آن‌ها توزیع شده باشند، را پارامترهای مرکزی می‌نامند. پارامترهایی مانند میانگین، میانه، مد و چارک‌ها از جمله پارامترهای مرکزی می‌باشند.

میانگین: یکی از مهم‌ترین پارامترهای مرکزی یک جامعه که اطلاعاتی در مورد همه داده‌ها در آن وجود دارد، میانگین می‌باشد و بسته به نوع داده‌ها دارای انواع مختلفی مانند میانگین حسابی، میانگین هندسی، میانگین پیراسته و میانگین هارمونیک است.

میانگین حسابی (معدل): این میانگین از تقسیم مجموع داده‌ها بر تعداد آن‌ها به دست می‌آید.

$$\mu_x = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_i + \dots + x_n}{n}$$

مثال: نمره مسوولیت‌پذیری پنج مدیر 13، 8، 14، 15 و 10 می‌باشد. میانگین مسوولیت‌پذیری این پنج مدیر را حساب کنید.

$$\mu_x = \frac{\sum_{i=1}^5 x_i}{5} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{10 + 15 + 14 + 8 + 13}{5} = \frac{60}{5} = 12$$



ادامه پارامترهای مرکزی

میانگین حسابی موزون: در صورتیکه هر یک از مشاهدات بیش از یک بار تکرار شده باشند، برای محاسبه میانگین جامعه از رابطه زیر استفاده می‌کنیم.

$$\mu_x = \frac{\sum_{i=1}^k w_i x_i}{\sum_{i=1}^k w_i} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_i x_i + \dots + w_k x_k}{w_1 + w_2 + \dots + w_i + \dots + w_k}$$

مثال: نمرات مسوولیت‌پذیری 20 مدیر به شرح جدول زیر به دست آمده است. میانگین مسوولیت‌پذیری این مدیران را حساب کنید.

15	12	10	6	5	(نمره)
4	6	5	2	3	(تکرار)

$$\mu_x = \frac{\sum_{i=1}^5 w_i x_i}{\sum_{i=1}^5 w_i} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5}{w_1 + w_2 + w_3 + w_4 + w_5} = \frac{3 \times 5 + 2 \times 6 + 5 \times 10 + 6 \times 12 + 4 \times 15}{3 + 2 + 5 + 6 + 4} = \frac{209}{20} = 10,45$$



ادامه پارامترهای مرکزی

خواص میانگین حسابی:

(1) جمع جبری اختلاف مجموعه‌ای از اعداد از میانگین حسابی آن‌ها، برابر صفر است. (تقارن)
$$\sum_{i=1}^N (x_i - \mu_x) = 0$$

(2) هرگاه هر یک از مشاهدات با عدد ثابت a جمع شوند، میانگین اعداد حاصل شده به اندازه a از میانگین قبلی بیشتر است.
$$y_i = x_i + a \Rightarrow \mu_y = \mu_x + a$$

(3) هرگاه هر یک از مشاهدات در عدد ثابت b ضرب گردند، میانگین اعداد حاصل برابر حاصل ضرب میانگین قبلی در عدد b خواهد بود.
$$y_i = b x_i \Rightarrow \mu_y = b \mu_x$$

(4) اگر x و y دو مجموعه از مشاهدات از یک جامعه آماری باشند و مجموعه z از جمع نظیر به نظیر اعداد دو مجموعه x و y به دست آمده باشد، میانگین مجموعه اعداد z برابر مجموع میانگین‌های دو مجموعه x و y می‌باشد.
$$z_i = x_i + y_i \Rightarrow \mu_z = \mu_x + \mu_y$$



ادامه پارامترهای مرکزی

میانگین پیراسته: اگر در یک مجموعه از مشاهدات تعداد اندکی از داده‌ها با سایرین همخوانی و تجانس نداشته باشند، در محاسبه میانگین این مجموعه تاثیرات نامطلوبی بر جای می‌گذارند. به منظور یافتن پارامتر میانگین جامعه که معرف آن جامعه باشد، داده‌های نامتجانس که به طور غیرعادی خیلی بزرگ یا خیلی کوچک هستند را حذف می‌کنیم. مراحل محاسبه میانگین پیراسته به شرح زیر است.

الف) داده‌ها را به صورت صعودی مرتب می‌کنیم.

ب) تمام مشاهدات کوچک‌تر از LN درصد پایین و بزرگتر از LN درصد بالا را حذف می‌کنیم.

ج) میانگین مشاهدات باقیمانده را محاسبه می‌کنیم.

میانگین وینزوری: اگر در محاسبه میانگین پیراسته، به جای داده‌های حذف شده، مقدار عددی LN درصد بالا و پایین جایگزین شوند، میانگین پیراسته دیگری به نام وینزوری به دست می‌آید.



ادامه پارامترهای مرکزی

میانگین هندسی: برای محاسبه میانگین اندازه‌های نسبی همچون نسبت‌ها، درصدها، نرخ‌های رشد و . . . از میانگین هندسی که به صورت زیر تعریف می‌شود استفاده می‌کنیم.

$$\mu_G = (x_1 x_2 x_3 \cdots x_i \cdots x_n)^{\frac{1}{n}}$$

مثال: نسبت سود شرکت زمزم در سال 1367 به 1366 مساوی 3، سال 1368 به 1367 مساوی 2 و سال 1369 به 1368 برابر 4/5 می‌باشد. میانگین هندسی برای نسبت سود شرکت زمزم در این سه سال چقدر است؟

$$\mu_G = (x_1 x_2 x_3)^{\frac{1}{3}} = (3 \times 2 \times 4/5)^{\frac{1}{3}} = (27)^{\frac{1}{3}} = 3$$

$$\mu_G = (x_1^{w_1} x_2^{w_2} x_3^{w_3} \cdots x_i^{w_i} \cdots x_n^{w_n})^{\frac{1}{N}} = \left(\prod_{i=1}^k x_i^{w_i} \right)^{\frac{1}{N}}$$

میانگین هندسی موزون:

55	65	80	88	(درصد کارایی)	میانگین هندسی = $(55^2 \times 65^7 \times 80^5 \times 88^6)^{\frac{1}{20}}$	میانگین حسابی = $\frac{55^2 + 65^7 + 80^5 + 88^6}{20}$
2	7	5	6	(تکرار)		

$$\log \mu_G = \frac{1}{20} (6 \log 88 + 5 \log 80 + 7 \log 65 + 2 \log 55) = 1,8676 \Rightarrow \mu_G = 10^{1,8676} = 73,74$$



ادامه پارامترهای مرکزی

میانگین هارمونیک: برای محاسبه میانگین مشاهدات با مقیاس‌های ترکیبی مانند متر بر ثانیه، دور در ثانیه، نفر روز و مانند آن‌ها از میانگین هارمونیک استفاده می‌کنیم. میانگین هارمونیک، معکوس میانگین معکوس داده‌های آماری است.

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

مثال: دنده‌ای یک سوم ابتدای مسیری مستقیم را با سرعت 8 متر بر ثانیه، یک سوم دوم را با سرعت 10 متر بر ثانیه و باقیمانده مسیر را با سرعت 6 متر بر ثانیه طی کرده است. سرعت میانگین این دنده را حساب کنید.

$$\mu_H = \frac{1}{\frac{1}{V_1} + \frac{1}{V_2} + \frac{1}{V_3}} = \frac{1}{\frac{1}{8} + \frac{1}{10} + \frac{1}{6}} = \frac{1}{\frac{15+12+20}{120}} = \frac{120}{47} \approx 2.55 \frac{m}{s}$$



ادامه پارامترهای مرکزی

میانگین هارمونیک موزون:

$$\mu_H = \frac{w_1 + w_2 + \dots + w_i + \dots + w_k}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \dots + \frac{w_i}{x_i} + \dots + \frac{w_k}{x_k}} = \frac{\sum_{i=1}^k w_i}{\sum_{i=1}^k \frac{w_i}{x_i}}$$

مثال: سرعت 30 ماشین بر حسب دور در ثانیه در جدول زیر آمده است. میانگین سرعت ماشین‌ها را محاسبه نمایید.

سرعت	5	6	10	15	x_i	
$\sum w_i = 30$	6	10	8	6	w_i	

$$\mu_H = \frac{6+10+8+6}{\frac{6}{5} + \frac{10}{6} + \frac{8}{10} + \frac{6}{15}} = \frac{30}{4.067} \simeq 7.376$$



ادامه پارامترهای مرکزی

مد (نما): معمولا مد برای رایج‌ترین سبک یا لباس استفاده می‌شود. در آمار نیز مد نشان‌دهنده داده‌ای است که بیشترین تکرار را در بین داده‌های یک جامعه آماری داشته باشد. در یک جامعه آماری معمولا داده‌ها و مشاهدات حول مد (نما) توزیع می‌شوند. به عبارت دیگر مد (MO) یکی از پارامترهای مرکزی در یک جامعه آماری می‌باشد.

یک جامعه آماری ممکن است بیش از یک مد داشته باشد و همچنین در صورتی که تعداد تکرار همه داده‌ها یکسان باشد، جامعه فاقد مد خواهد بود.

به عنوان مثال در مشاهدات 2، 3، 5، 2، 5، 4، 6، 5، 7، 5 عدد 5 به عنوان مد در نظر گرفته می‌شود. ($MO=2$) و در داده‌های 2، 4، 5، 2، 4، 4، 6، 5، 7، 5 اعداد 4 و 5 هر دو مد هستند. در مشاهدات 2، 3، 4، 5، 5، 4، 3، 2 به دلیل تعداد تکرار برابر همه مشاهدات، جامعه فاقد مد است.



ادامه پارامترهای مرکزی

چارک‌ها: چارک‌ها یک جامعه آماری را به چهار قسمت تقسیم می‌کنند. در یک جامعه که داده‌های آن مرتب شده‌اند، داده‌ای که یک چهارم داده‌ها از آن کوچکتر و سه چهارم داده‌ها از آن بزرگتر باشند را چارک اول (Q_1)، داده‌ای که دو چهارم (نصف) داده‌ها از آن کوچکتر و دو چهارم از آن بزرگتر باشند را چارک دوم (Q_2) و داده‌ای که سه چهارم داده‌ها از آن کوچکتر و یک چهارم داده‌ها از آن بزرگتر باشند را چارک سوم (Q_3) می‌نامند.

میانه: چارک دوم در یک جامعه آماری را میانه می‌نامند و آن را با نماد (Md) نمایش می‌دهند. از نظر آماری میانه به خصوص در جامعه‌های آماری نامتقارن، پارامتر بهتری برای نشان دادن مرکز داده‌ها است زیرا کمتر تحت تاثیر مقادیر داده‌ها قرار دارد.

خاصیت مهم میانه: مجموع قدرمطلق اختلاف داده‌ها با میانه از مجموع قدرمطلق اختلاف داده‌ها با هر داده دیگر در جامعه آماری کوچکتر است.



ادامه پارامترهای مرکزی

مراحل محاسبه چارک‌ها: (1) داده‌ها را به ترتیب صعودی مرتب می‌کنیم.

(2) داده‌های مرتب شده را از 1 تا N کدگذاری می‌کنیم.

(3) محل چارک a ام ($a = 1, 2, 3$) را از رابطه مقابل محاسبه می‌کنیم. $C_{Q_a} = \frac{aN}{4} + \frac{1}{2}$

(4) با استفاده از محل چارک، مقدار آن را بدست می‌آوریم.

مثال: چارک‌های اول، دوم و سوم مشاهدات 85، 140، 120، 80، 90 و 100 را بدست آورید.

داده‌های مرتب شده	80	85	90	100	120	140
کد مشاهدات	1	2	3	4	5	6

$$\left\{ \begin{array}{l} C_{Q_1} = \frac{6}{4} + \frac{1}{2} = 2 \\ C_{Q_2} = \frac{2 \times 6}{4} + \frac{1}{2} = 3.5 \\ C_{Q_3} = \frac{3 \times 6}{4} + \frac{1}{2} = 5 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} Q_1 = 85 \\ Q_2 = \frac{90 + 100}{2} = 95 \\ Q_3 = 120 \end{array} \right.$$



ادامه پارامترهای مرکزی

مثال: چارک‌های اول، دوم و سوم مشاهدات 85، 140، 120، 80، 90، 100 و 160 را بدست آورید.

داده‌های مرتب شده	80	85	90	100	120	140	160
کد مشاهدات	1	2	3	4	5	6	7

$$\left\{ \begin{array}{l} C_{Q_1} = \frac{7}{4} + \frac{1}{2} = 2,25 \\ C_{Q_2} = \frac{2 \times 7}{4} + \frac{1}{2} = 4 \\ C_{Q_3} = \frac{3 \times 7}{4} + \frac{1}{2} = 5,75 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} Q_1 = 85 + 0,25 \times (90 - 85) = 86,25 \\ Q_2 = 100 \\ Q_3 = 120 + 0,75 \times (140 - 120) = 135 \end{array} \right.$$



پارامترهای پراکندگی

برای تفسیر و توصیف بهتر داده‌ها علاوه بر پارامترهای مرکزی، از پارامترهای پراکندگی استفاده می‌شود. به خصوص اگر در دو جامعه آماری پارامترهای مرکزی یکسان باشند، برای مقایسه دو جامعه از پارامترهای پراکندگی که نحوه توزیع داده‌ها حول پارامترهای مرکزی را توصیف می‌کنند، استفاده می‌کنیم. فرض کنیم میانگین وزن محصولات تولید شده در دو دستگاه یکسان باشد. برای اینکه بتوانیم یکی از این دو دستگاه را برای خط تولید انتخاب کنیم، نیاز به اطلاعات بیشتری از این دو دستگاه هستیم. اطلاعاتی که به ما خطای دستگاه را در وزن محصولات تولید شده نشان دهد. به طور قطع دستگاهی مطلوب ما خواهد بود که خطای کمتری داشته باشد یا به عبارت دیگر دستگاهی که وزن محصولات تولید شده آن نسبت به میانگین، اختلاف کمتری داشته باشد، گزینه انتخابی ما خواهد بود.

پارامترهای پراکندگی متفاوتی برای یک جامعه قابل تعریف است که از این میان پارامترهای «دامنه تغییرات»، «دامنه میان چارکی»، «انحراف متوسط از میانگین»، «واریانس و انحراف معیار»، «نیمه واریانس» و «ضریب تغییرات» در این درس معرفی می‌شوند.



ادامه پارامترهای پراکندگی

دامنه تغییرات: اختلاف بیشترین و کمترین داده‌ها در یک جامعه آماری، دامنه تغییرات نام دارد.
 $R = \max - \min$

دامنه میان چارکی: دامنه تغییرات 50 درصد وسط در مشاهدات را دامنه میان چارکی می‌نامند.
 $IQR = Q_3 - Q_1$

البته در مواردی هم به جای دامنه میان چارکی از «نیمه میان چارکی» یا «انحراف چارکی» استفاده می‌شود.
 $SIQR = \frac{Q_3 - Q_1}{2}$

انحراف متوسط از میانگین: متوسط اختلاف‌های هر یک از داده‌ها از میانگین را انحراف متوسط از

میانگین می‌نامند.

$$A.D_{\mu} = \frac{\sum_{i=1}^N |x_i - \mu|}{N}$$

گاهی در برخی تحقیقات از پارامترهای «انحراف متوسط از میان» یا «انحراف متوسط از مد» نیز استفاده می‌شود که به طرز مشابه قابل تعریف هستند.

$$A.D_{Mo} = \frac{\sum_{i=1}^N |x_i - Mo|}{N} \quad A.D_{Md} = \frac{\sum_{i=1}^N |x_i - Md|}{N}$$



ادامه پارامترهای پراکندگی

واریانس و انحراف معیار: یکی از اشکالات پارامتر «انحراف متوسط از میانگین»، ناکارآمدی در تشخیص داده‌های پرت است. به عبارت دیگر اگر تعداد کمی از داده‌ها، اختلاف زیادی با میانگین داشته باشند و تعداد زیادی به میانگین نزدیک باشند، در محاسبات اثر داده‌های دور از میانگین از بین می‌رود. از طرف دیگر در

فضای اقلیدس $D = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$ فاصله دو نقطه از رابطه

به دست می‌آید که با الهام از این رابطه می‌توان پارامتر «انحراف معیار» یا «انحراف استاندارد» را به صورت زیر تعریف نمود.

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_N - \mu)^2}{N} = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

که در آن واریانس و انحراف معیار می‌باشند.

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_N - \mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$



ادامه پارامترهای پراکندگی

مثال: برای داده‌های 85، 140، 120، 80، 90، 100 و 160 پارامترهای دامنه تغییرات، دامنه میان چارکی،

$$R = \max - \min = 160 - 80 = 80$$

انحراف چارکی، انحراف متوسط از میانگین، واریانس و انحراف معیار را حساب کنید

$$IQR = Q_3 - Q_1 = 135 - 86,25 = 48,75$$

$$SIQR = \frac{Q_3 - Q_1}{2} = 24,375$$

$$\mu = \frac{\sum x_i}{N} = \frac{80 + 85 + 90 + 100 + 120 + 140 + 160}{7} = \frac{775}{7} \simeq 110,7142857$$

$$A.D._\mu = \frac{\sum |x_i - \mu|}{N} = \frac{|80 - 110,7142857| + |85 - 110,7142857| + \dots + |160 - 110,7142857|}{7} =$$

$$\frac{30,7142857 + 25,7142857 + \dots + 49,28571429}{7} = \frac{175,7142857}{7} = 25,10204082$$

$$\sigma^2 = \frac{(80 - 110,7142857)^2 + (85 - 110,7142857)^2 + \dots + (160 - 110,7142857)^2}{7} = \frac{5521,428571}{7} = 788,7755102$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{788,7755102} \simeq 28,0851475$$



ادامه پارامترهای پراکندگی

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \frac{\sum_{i=1}^N (x_i^2 - 2\mu x_i + \mu^2)}{N} = \frac{\sum_{i=1}^N x_i^2 - 2\mu \sum_{i=1}^N x_i + \sum_{i=1}^N \mu^2}{N} =$$

روش معادل برای محاسبه واریانس:

$$\frac{\sum_{i=1}^N x_i^2}{N} - 2\mu \frac{\sum_{i=1}^N x_i}{N} + \frac{N\mu^2}{N} = \frac{\sum_{i=1}^N x_i^2}{N} - \mu^2$$

مثال: واریانس داده‌های مثال قبل را با رابطه اخیر به دست آورید.

$$\mu = \frac{\sum x_i}{N} = \frac{80+85+90+100+120+140+160}{7} = 110,7142857$$

$$\sigma^2 = \frac{\sum x_i^2}{N} - \mu^2 = \frac{80^2 + 85^2 + 90^2 + 100^2 + 120^2 + 140^2 + 160^2}{7} - (110,7142857)^2$$

$$\sigma^2 = \frac{91325}{7} - 12257,65306 = 13046,42857 - 12257,65306 = 788,7755102$$



ادامه پارامترهای پراکندگی

خواص واریانس:

1) $y_i = x_i + b \Rightarrow \sigma_y^2 = \sigma_x^2$

2) $y_i = a x_i \Rightarrow \sigma_y^2 = a^2 \sigma_x^2$

نیمه واریانس: میانگین مجذور انحرافات نامطلوب از میانگین را نیمه واریانس می‌گویند.

$$S.V = \frac{\sum_{i=1}^K (x_i - \mu_x)^2}{K}$$

در محاسبات سود، داده‌های کمتر از میانگین نامطلوب هستند و در محاسبات زیان مقادیر بیشتر از میانگین



ادامه پارامترهای پراکندگی

ضریب پراکندگی: نسبت انحراف معیار به میانگین داده‌ها را ضریب پراکندگی می‌نامند. $C.V = \frac{\sigma_x}{\mu_x}$

کاربردهای ضریب پراکندگی:

مقایسه دو جامعه آماری که:

- (1) دارای مقیاس یکسان نیستند.
- (2) دارای مقیاس یکسان هستند ولی تفاوت زیادی در بزرگی مشاهدات وجود دارد.
- (3) واریانس‌های یکسان دارند ولی میانگین آنها متفاوت است.



ادامه پارامترهای پراکندگی

مثال: سود ده شرکت به صورت زیر گزارش شده است. مقادیر واریانس، نیمه واریانس و ضریب پراکندگی را حساب کنید.

$$x_i = 0, 1, 5, 7, 4, 3, 2, 8, 6, 4$$

$$\mu = \frac{\sum x_i}{10} = \frac{0+1+5+7+4+3+2+8+6+4}{10} = \frac{40}{10} = 4$$

ابتدا میانگین را محاسبه می‌کنیم:

با توجه به اینکه مقادیر گزارش شده مربوط به سود است، مقادیر کمتر از میانگین، انحراف نامطلوب ایجاد می‌کنند. بنابراین:

$$S.V = \frac{(0-4)^2 + (1-4)^2 + (2-4)^2 + (3-4)^2}{4} = 7,5$$

$$\sigma^2 = \frac{\sum_{i=1}^{10} (x_i - \mu)^2}{10} = \frac{60}{10} = 6$$

$$C.V = \frac{\sigma}{\mu} = \frac{\sqrt{6}}{4} = 0,6124$$

واریانس و ضریب پراکندگی نیز به این صورت به دست می‌آیند.



فصل سوم

طبقه‌بندی و توصیف هندسی

مشاهدات جامعه آماری

Classification & Geometric Description

of Statistical Population Observations



طبقه‌بندی و سازماندهی مشاهدات

در اکثر تحقیقات با خیل زیاد داده‌ها مواجه هستیم که تفسیر و توصیف آن‌ها را مشکل و حتی گاهی غیرممکن می‌کند. برای حل این مساله، معمولا محققان اقدام به طبقه‌بندی داده‌ها می‌کنند. الگوریتم نه چندان دقیق برای طبقه‌بندی به شرح زیر است.

(1) داده‌های جمع آوری شده را مرتب و دامنه تغییرات را بدست می‌آوریم $R = \max - \min$

(2) تعداد طبقات را از یکی از دو رابطه تجربی مقابل به دست می‌آوریم $K = \sqrt{N}$, $K = 1 + \frac{\log N}{3}$

(3) طول طبقات را محاسبه می‌کنیم. R

(4) در روش پیوسته که بیشتر در مورد $I = \frac{R}{K}$ داده‌های با رقم اعشار به کار می‌رود، از کوچکترین داده شروع می‌کنیم و طبقات را با طول طبقه محاسبه شده می‌سازیم به طوری که حد بالای هر طبقه منطبق بر حد پایین طبق بعدی باشد. در هر دسته تعدادی از داده‌ها که در فاصله بسته حد پایین و باز حد بالای طبقه هستند را شمارش کرده و در ستون مربوط به فراوانی (مطلق) می‌نویسیم. لازم به ذکر است هر یک از داده‌ها تنها و تنها می‌توانند عضو یکی از دسته‌ها باشند. در روش گسسته عرض طبقات با طول طبقات برابر نیستند و بسته به دقت اندازه‌گیری مشاهدات حد بالای هر طبقه کوچکتر از حد پایین طبقه بعدی است البته اختلاف این دو عدد باید طوری انتخاب شود که هیچ داده‌ای بین این دو مقدار قرار نگیرد.



ادامه طبقه‌بندی و سازماندهی مشاهدات

۲۰/۵	۱۹/۵	۱۵/۶	۲۴/۱	۹/۹
۱۵/۴	۱۲/۷	۵/۰	۱۷/۰	۲۸/۶
۱۶/۹	۷/۸	۲۳/۳	۱۱/۸	۱۸/۴
۱۳/۴	۱۴/۳	۱۹/۲	۹/۲	۱۶/۸
۸/۸	۲۲/۱	۲۰/۸	۱۲/۶	۱۵/۹

مثال: یکی از شاخص‌های محاسبه ریسک سرمایه‌گذاری در سهام شاخص نسبت قیمت به درآمد $\frac{P}{E}$ () است. اصولاً سهامی که نسبت $\frac{P}{E}$ کمتری دارند، ریسک کمتری دارند. داده‌های مقابل نرخ سهام 25 شرکت در بورس تهران هستند. برای طبقه‌بندی داده‌ها

$$(1) \text{ دامنه تغییرات: } R = \max - \min = 28,6 - 5 = 23,6$$

$$(2) \text{ تعداد طبقات: } K = 1 + 3,32 \log 25 = 5,64 \Rightarrow K = 6$$

$$(3) \text{ فاصله طبقات: } I = \frac{R}{K} = \frac{23,6}{6} = 3,933 \Rightarrow I = 4$$

(4) تشکیل جدول: با توجه به نوع کمیت از روش طبقه‌بندی پیوسته استفاده می‌کنیم.

طبقه i	حدود طبقات (C - L)	خط و نشان	فراوانی مطلق (F _i)
۱	۵-۹		۳
۲	۹-۱۳	++++	۵
۳	۱۳-۱۷	++++	۷
۴	۱۷-۲۱	++++	۶
۵	۲۱-۲۵		۳
۶	۲۵-۲۹		۱



ادامه طبقه‌بندی و سازماندهی مشاهدات

مثال: مشاهدات زیر در مورد میزان مایع موجود در 70 بطری بر حسب میلی لیتر می باشد.

۳۶/۹	۳۶/۹	۳۶/۸	۳۶/۴	۳۷/۰	۳۶/۷	۳۶/۹
۳۶/۸	۳۷/۰	۳۶/۷	۳۶/۵	۳۶/۹	۳۶/۵	۳۶/۸
۳۶/۸	۳۶/۹	۳۶/۹	۳۶/۷	۳۶/۸	۳۷/۰	۳۶/۰
۳۷/۰	۳۶/۶	۳۶/۹	۳۶/۹	۳۶/۴	۳۷/۸	۳۶/۵
۳۷/۰	۳۶/۷	۳۶/۸	۳۶/۹	۳۷/۰	۳۶/۲	۳۷/۰
۳۶/۸	۳۶/۵	۳۷/۰	۳۶/۹	۳۷/۰	۳۶/۷	۳۶/۹
۳۷/۰	۳۶/۹	۳۷/۷	۳۶/۹	۳۷/۰	۳۶/۸	۳۶/۸
۳۷/۲	۳۶/۹	۳۷/۰	۳۶/۶	۳۶/۶	۳۶/۶	۳۶/۴
۳۶/۹	۳۶/۶	۳۷/۰	۳۷/۰	۳۷/۰	۳۶/۹	۳۶/۹
۳۷/۰	۳۷/۰	۳۶/۹	۳۶/۰	۳۶/۹	۳۷/۰	۳۷/۰

اگر بخواهیم این مشاهدات را در ده طبقه با فاصله طبقاتی 0/20 سازماندهی کنیم، جدولی مطابق شکل زیر خواهیم داشت که عرض هر طبقه برابر با طول طبقه مساوی 0/20 خواهد بود. در واقع طبقه‌بندی گسسته با تقریب انجام شده است.

فرآوانی مطلق (F_i)	خط و نشان	حدود طبقات (C - L)	طبقه i
۲		۳۶/۰۰-۳۶/۱۰	۱
۱		۳۶/۲۰-۳۶/۳۰	۲
۷		۳۶/۴۰-۳۶/۵۰	۳
۱۱		۳۶/۶۰-۳۶/۷۰	۴
۲۷		۳۶/۸۰-۳۶/۹۰	۵
۱۹		۳۷/۰۰-۳۷/۱۰	۶
۱		۳۷/۲۰-۳۷/۳۰	۷
۰	—	۳۷/۴۰-۳۷/۵۰	۸
۱		۳۷/۶۰-۳۷/۷۰	۹
۱		۳۷/۸۰-۳۷/۹۰	۱۰



ادامه طبقه‌بندی و سازماندهی مشاهدات

۱۰	۲۱	۳۲	۳۸	۵۰	۵۳	۶۴	۷۳	۸۵	۹۸
۱۰	۲۴	۳۳	۴۹	۵۱	۵۴	۶۴	۷۳	۸۷	۱۰۰
۱۴	۲۵	۳۴	۴۹	۵۱	۵۷	۶۵	۷۵	۸۹	۱۰۶
۱۵	۲۷	۳۵	۵۰	۵۱	۶۰	۶۸	۷۷	۹۲	۱۱۰
۱۷	۳۰	۳۷	۵۰	۵۱	۶۳	۷۰	۸۰	۹۵	۱۲۰

مثال: داده‌های مرتب شده مقابل مربوط به سود سالانه 50 شرکت به 10 میلیون ریال است. برای طبقه‌بندی چهار مرحله یاد شده را انجام

(1) دامنه تغییرات: $R = \max - \min = 120 - 10 = 110$

(2) تعداد طبقات: $K = \sqrt{N} = \sqrt{50} = 7,071 \Rightarrow K = 8$

(3) فاصله طبقات: $I = \frac{R}{K} = \frac{110}{8} = 13,75 \Rightarrow I = 14$

(4) تشکیل جدول فراوانی:

طبقه i	سود سالانه (C - L)	خط و نشان	فراوانی مطلق (F _i)
۱	۱۰-۲۳		۶
۲	۲۴-۳۷		۹
۳	۳۸-۵۱		۱۰
۴	۵۲-۶۵		۸
۵	۶۶-۷۹		۶
۶	۸۰-۹۳		۵
۷	۹۴-۱۰۷		۴
۸	۱۰۸-۱۲۱		۲
N = ۵۰			Σ F _i = ۵۰



طبقه‌بندی مشاهدات ناپیوسته

در برخی از موارد مشاهدات جنبه کاملاً ناپیوسته دارند و نمی‌توان مانند مثال‌های قبل آنها را طبقه‌بندی کرد. به عنوان مثال پرسش‌نامه‌ای که در مورد ترجیح مشتریان یک رستوران برای استفاده از چهار نوع غذا تنظیم شده، اطلاعاتی در مورد تعداد افرادی که هر یک از غذاها را ترجیح داده‌اند به ما می‌دهد. در این موارد هر یک از انواع غذا را به عنوان یک دسته انتخاب و فراوانی متناظر را در جدول توزیع فراوانی مشخص می‌کنیم.

مثال: به منظور بررسی تعداد غایبین یک شرکت در دو نوبت صبح و عصر، دفاتر حضور و غیاب مورد بررسی قرار گرفت و نتیجه آن مطابق جدول زیر به دست آمد

تعداد روز (F_i)	تعداد غایبین عصر (y_i)	تعداد روز (F_i)	تعداد غایبین صبح (X_i)
۳	۰	۱۰	۰
۵	۱	۱۵	۱
۱۰	۲	۸	۲
۱۲	۳	۱۲	۳
۲۵	۴	۱۰	۴
۵	۵	۵	۵
$\sum F_i = 60$		$\sum F_i = 60$	



توزیع فراوانی نسبی

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

$$x_i = \frac{U_i + L_i}{2}$$

فراوانی نسبی: نسبت فراوانی مطلق به تعداد کل داده‌ها را فراوانی نسبی گویند.

متوسط طبقه: نقطه میانی حد بالا و حد پایین هر طبقه را متوسط آن طبقه گویند.

C-L	X_i (متوسط طبقه)	F_i	f_i
۵-۹	۷	۳	۰/۱۲
۹-۱۳	۱۱	۵	۰/۲۰
۱۳-۱۷	۱۵	۷	۰/۲۸
۱۷-۲۱	۱۹	۶	۰/۲۴
۲۱-۲۵	۲۳	۳	۰/۱۲
۲۵-۲۹	۲۷	۱	۰/۰۴
		$N=۲۵$	۱

$$\frac{3}{25} = 0,12$$

$$\frac{6}{25} = 0,24$$



توزیع فراوانی تجمعی

فراوانی تجمعی: فراوانی تجمعی هر طبقه برابر است با مجموع فراوانی همان طبقه

$$FC_i = \sum_{j=1}^i F_j$$

با فراوانی‌های طبقات قبلی

$$fc_i = \frac{FC_i}{N}$$

فراوانی تجمعی نسبی: فراوانی تجمعی نسبی برابر است با نسبت فراوانی تجمعی به تعداد کل داده‌ها

C-L	F_i	فراوانی تجمعی	فراوانی تجمعی
		(FC_i)	نسبی (fc_i)
۵-۹	۳	۳	۰/۱۲
۹-۱۳	۵	۸	۰/۳۲
۱۳-۱۷	۷	۱۵	۰/۶۰
۱۷-۲۱	۶	۲۱	۰/۸۴
۲۱-۲۵	۳	۲۴	۰/۹۶
۲۵-۲۹	۱	۲۵	۱



نمایش هندسی مشاهدات

یکی از روش‌های کارآمد در نمایش اطلاعات، نمایش هندسی می‌باشد که به دو دسته کمی و وصفی قابل ارائه است.

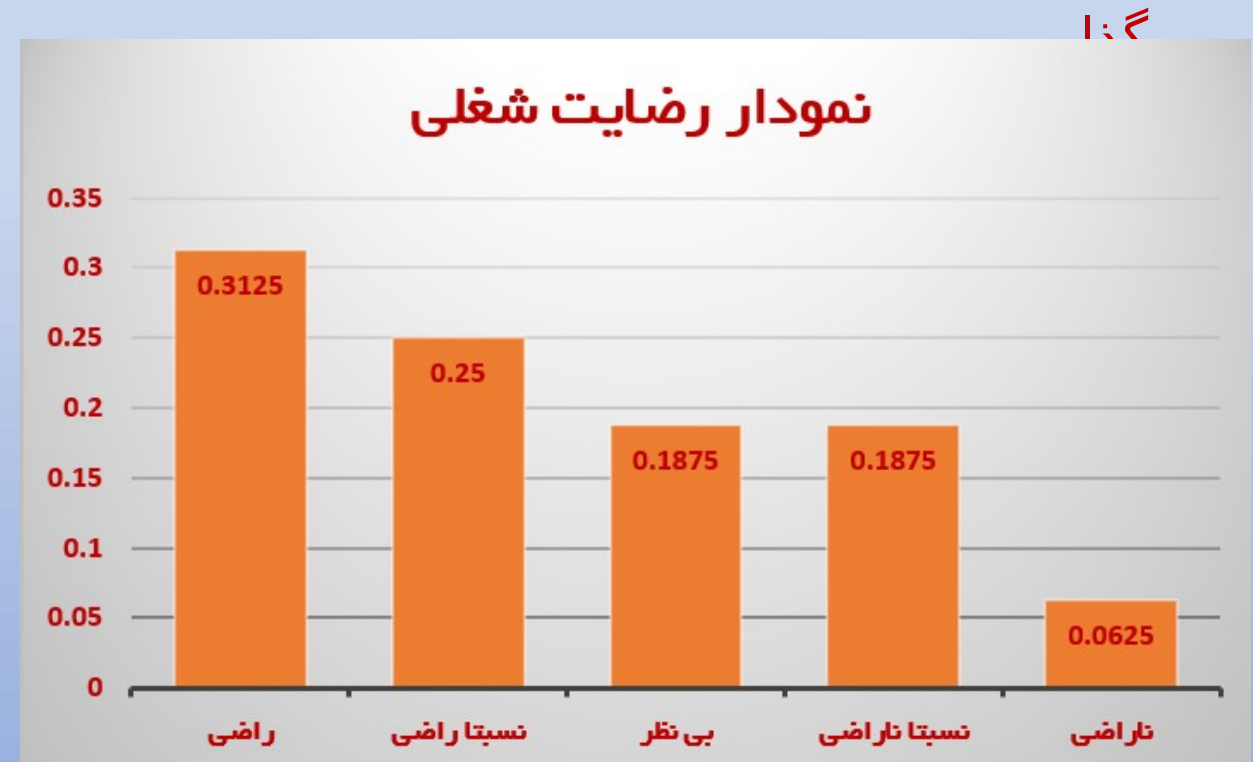




نمایش هندسی مشاهدات - نمودارهای وصفی - نمودار ستونی

نمودار ستونی معمولا برای نمایش داده‌های کیفی مانند (عالی، خوب، متوسط، بد و خیلی بد) یا (گروه A، B و C) به کار می‌رود. در این موارد هر گروه را در یک طبقه قرار می‌دهیم و فراوانی آن را در نمودار به نمایش

نوع پاسخ (x_i)	تعداد تفرات (F_i)	فراوانی نسبی (f_i)
راضی	۲۵۰	۰.۳۱۲۵
نسبتا راضی	۲۰۰	۰.۲۵
بی نظر	۱۵۰	۰.۱۸۷۵
نسبتا ناراضی	۱۵۰	۰.۱۸۷۵
ناراضی	۵۰	۰.۰۶۲۵
مجموع	۸۰۰	۱





نمایش هندسی مشاهدات - نمودارهای وصفی - نمودار دایره‌ای

یکی دیگر از نمودارهای مورد استفاده در نمایش داده‌های کیفی، نمودار دایره‌ای است. در این نمودار براساس فراوانی نسبی هر داده، زاویه اختصاص یافته به آن در نمودار دایره‌ای را محاسبه می‌کنیم.

$$\alpha_i = f_i \times 360^\circ$$

زاویه (α_i)	فراوانی نسبی (f_i)	تعداد تفرات (F_i)	سطح نیاز (x_i)
۱۱۲.۵	۰.۳۱۲۵	۲۵۰	راضی
۹۰	۰.۲۵	۲۰۰	نسبتاً راضی
۶۷.۵	۰.۱۸۷۵	۱۵۰	بی‌نظر
۶۷.۵	۰.۱۸۷۵	۱۵۰	نسبتاً ناراضی
۲۲.۵	۰.۰۶۲۵	۵۰	ناراضی
۳۶۰	۱	۸۰۰	مجموع





نمایش هندسی مشاهدات - نمودارهای وصفی - نمودار پاره‌تو

نمودار پاره‌تو نوعی نمودار ستونی است که در آن داده‌ها از بزرگ به کوچک مرتب شده‌اند و علاوه بر نمایش فراوانی مطلق، درصد فراوانی نسبی تجمعی نیز نمایش داده می‌شود و به این ترتیب سطح توجه و تمرکز بر داده‌ها مشخص می‌شود. این نمودار کاربرد زیادی در کنترل کیفیت دارد.

(x_i)	سطح نیاز	(F_i)	تعداد نفرات	(fc_i)	درصد فراوانی نسبی تجمعی
راضی		۲۵۰		۳۱۰۲۵	
نسبتاً راضی		۲۰۰		۵۶۰۲۵	
بی‌نظر		۱۵۰		۷۵	
نسبتاً ناراضی		۱۵۰		۹۳۰۷۵	
ناراضی		۵۰		۱۰۰	
مجموع		۸۰۰			

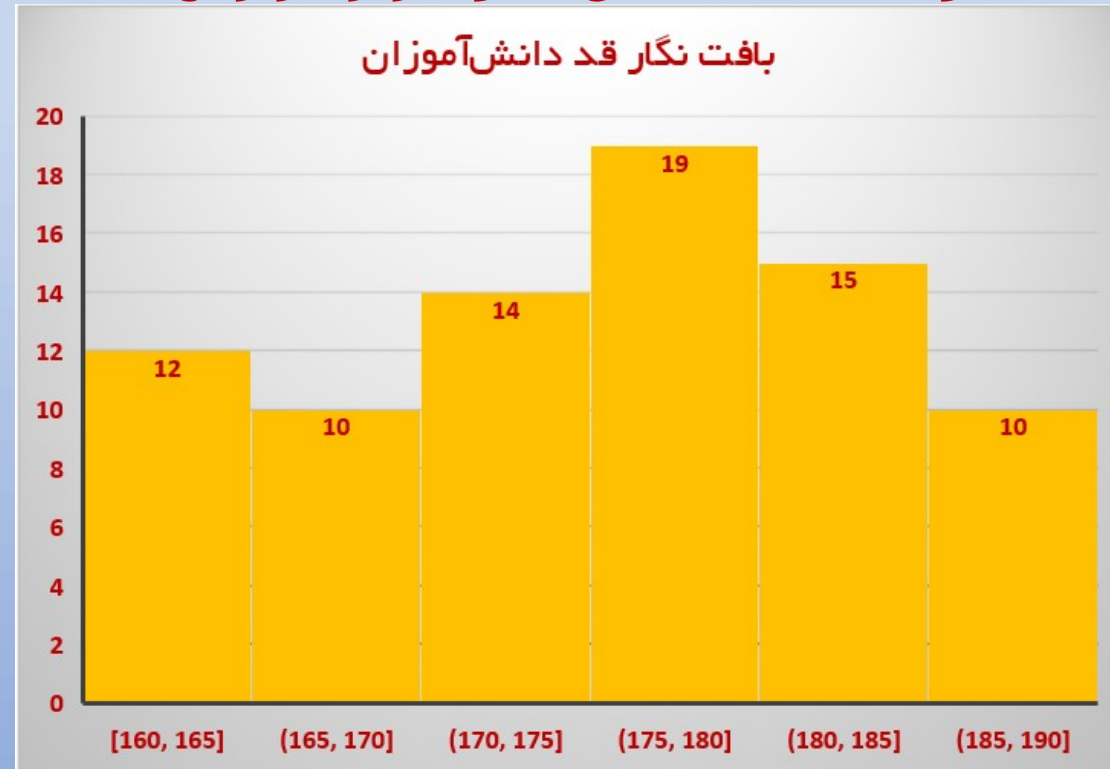




نمایش هندسی مشاهدات - نمودارهای کمی - نمودار بافت نگار

معمولا وقتی تعداد داده‌ها زیاد باشد، توزیع فراوانی داده‌ها در دسته‌های با طول برابر مشخص و نمودار بافت‌نگار برای آن رسم می‌شود. به عنوان مثال؛ قد 80 دانش‌آموز در یک مدرسه بر حسب سانتیمتر اندازه‌گیری شده و نتیجه آن مطابق جدول زیر گزارش شده است.

حدود دسته‌ها		نماینده دسته	فراوانی
۱۶۰	۱۶۵	۱۶۲.۵	۱۲
۱۶۵	۱۷۰	۱۶۷.۵	۱۰
۱۷۰	۱۷۵	۱۷۲.۵	۱۴
۱۷۵	۱۸۰	۱۷۷.۵	۱۹
۱۸۰	۱۸۵	۱۸۲.۵	۱۵
۱۸۵	۱۹۰	۱۸۷.۵	۱۰
مجموع			۸۰

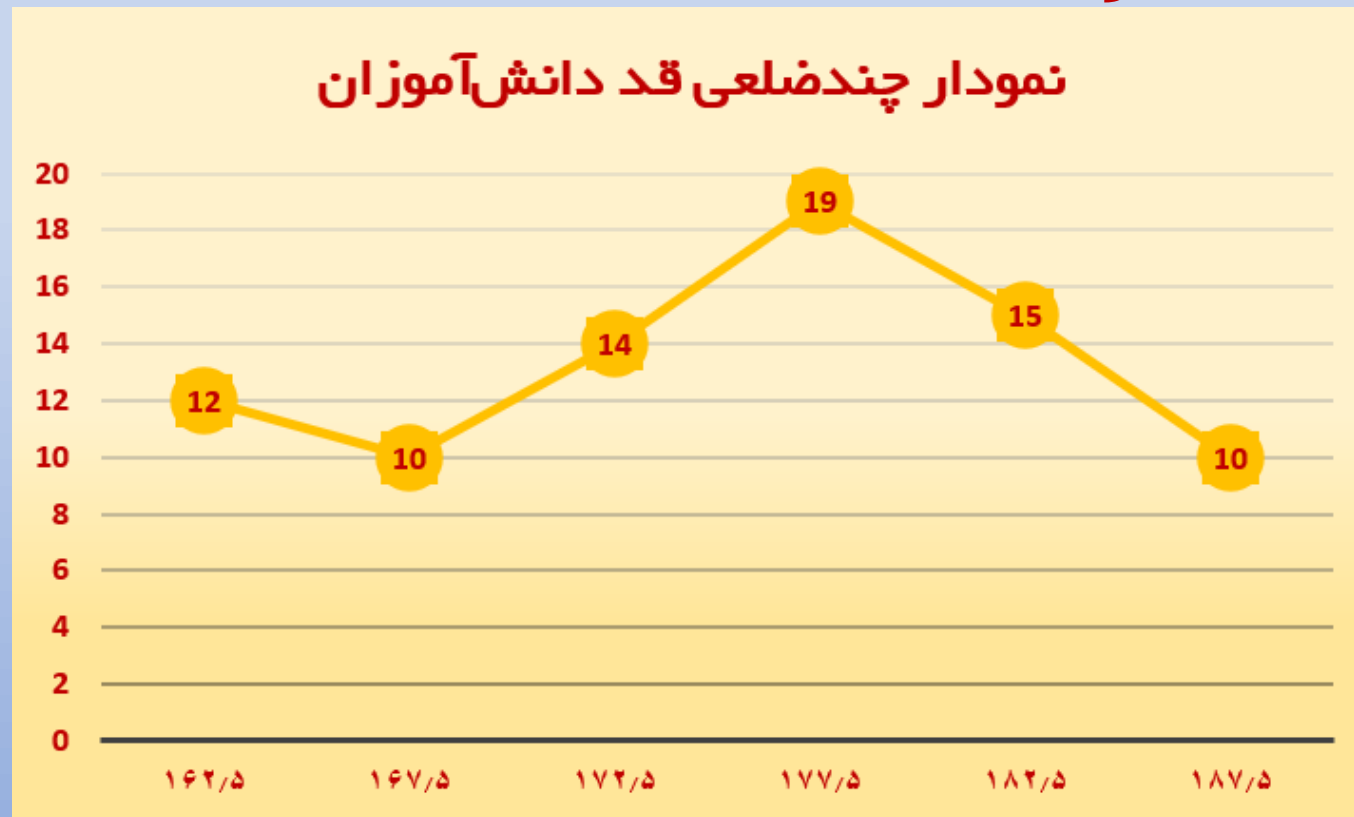




نمایش هندسی مشاهدات - نمودارهای کمی - نمودار چندضلعی

در نمودار چندضلعی، نقطه میانی هر طبقه روی محور افقی و فراوانی نسبی یا مطلق آن روی محور عمودی نمایش داده می‌شود.

حدود دسته‌ها	نماینده دسته	فراوانی
۱۶۰ ۱۶۵	۱۶۲.۵	۱۲
۱۶۵ ۱۷۰	۱۶۷.۵	۱۰
۱۷۰ ۱۷۵	۱۷۲.۵	۱۴
۱۷۵ ۱۸۰	۱۷۷.۵	۱۹
۱۸۰ ۱۸۵	۱۸۲.۵	۱۵
۱۸۵ ۱۹۰	۱۸۷.۵	۱۰
مجموع		۸۰

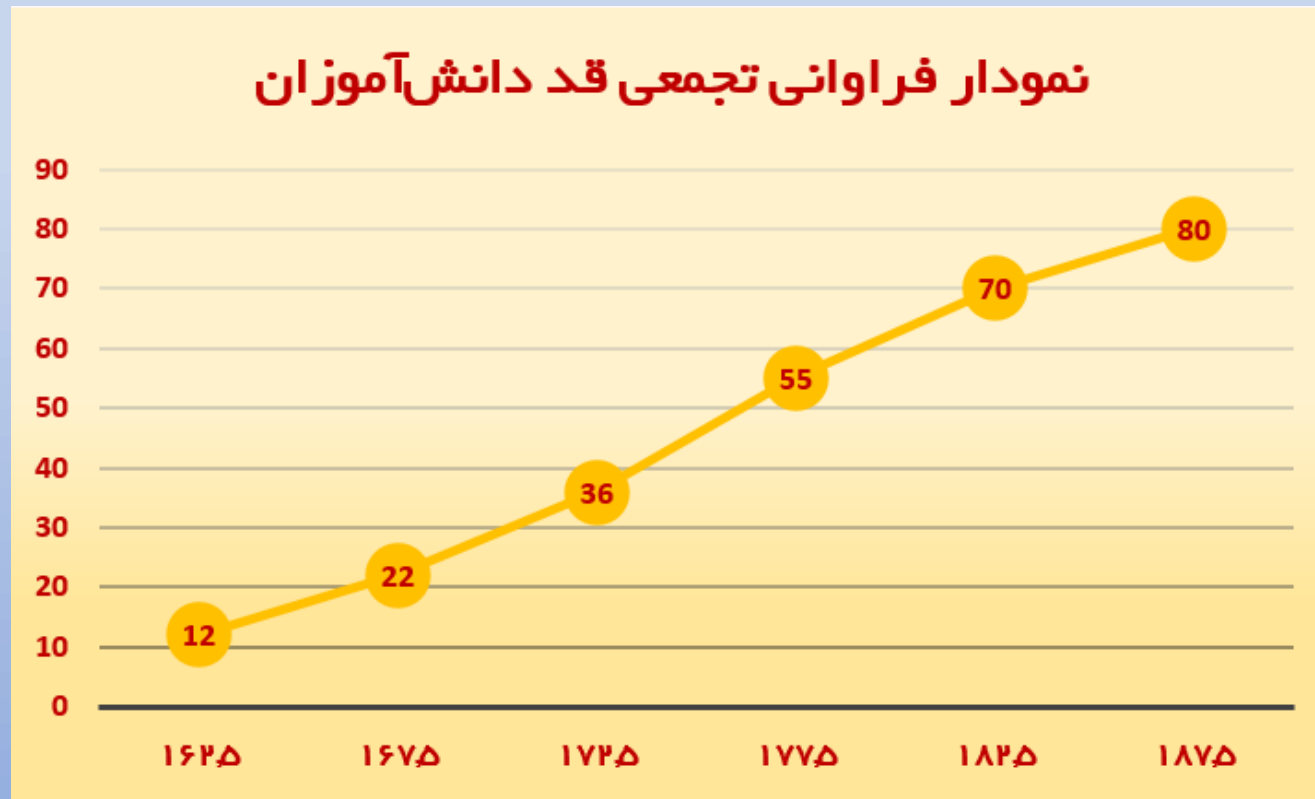




نمایش هندسی مشاهدات - نمودارهای کمی - نمودار فراوانی تجمعی

در نمودار فراوانی تجمعی، نقطه میانی هر طبقه روی محور افقی و فراوانی تجمعی آن روی محور عمودی نمایش داده می‌شود. با توجه به این نمودار می‌توان اطلاعات مربوط به صدکها، دهکها و چارکها را استخراج نمود.

فراوانی	فراوانی	نماینده دسته	حدود دسته‌ها
۱۲	۱۲	۱۶۲-۵	۱۶۵ ۱۶۰
۲۲	۱۰	۱۶۷-۵	۱۷۰ ۱۶۵
۳۶	۱۴	۱۷۲-۵	۱۷۵ ۱۷۰
۵۵	۱۹	۱۷۷-۵	۱۸۰ ۱۷۵
۷۰	۱۵	۱۸۲-۵	۱۸۵ ۱۸۰
۸۰	۱۰	۱۸۷-۵	۱۹۰ ۱۸۵
	۸۰		مجموع





نمایش هندسی مشاهدات - نمودارهای کمی - نمودار شاخه و برگ

نمودار شاخه و برگ نوعی نمایش اطلاعات که غالباً برای داده‌های عددی به کار می‌رود و در آن هر عدد به دو بخش تقسیم شده و در جدول دو بعدی نمایش داده می‌شود. به عنوان مثال میزان بارش در 40 شهر در یکسال به صورت زیر گزارش شده است. نمودار شاخه و برگ نمایش می‌دهیم. اگر چه برای نمایش اطلاعات به صورت شاخه و برگ، لزومی به مرتب‌سازی داده‌ها وجود ندارد، اما برای استخراج چنکها، بهتر است داده‌ها را

میزان بارش در ۴۰ شهر در یکسال بر حسب میلی‌متر

۱۶	۱۶	۱۷	۱۸	۲۰	۲۰	۲۰	۲۱	۲۲	۲۲
۲۳	۲۴	۲۴	۲۶	۲۷	۲۷	۲۸	۳۰	۳۰	۳۱
۳۴	۳۴	۳۵	۳۶	۳۹	۴۰	۴۰	۴۱	۴۲	۴۲
۴۳	۴۳	۴۳	۴۴	۴۶	۴۷	۴۹	۴۹	۵۰	۵۲

نمودار شاخه و برگ میزان بارش سالانه ۴۰ شهر بر حسب میلی‌متر

شاخه	برگ												
۱	۶	۶	۷	۸									
۲	۰	۰	۰	۱	۲	۲	۳	۴	۴	۶	۷	۷	۸
۳	۰	۰	۱	۴	۴	۵	۶	۹					
۴	۰	۰	۱	۲	۲	۳	۳	۳	۴	۶	۷	۹	۹
۵	۰	۲											



نمایش هندسی مشاهدات - نمودارهای کمی - نمودار جعبه‌ای

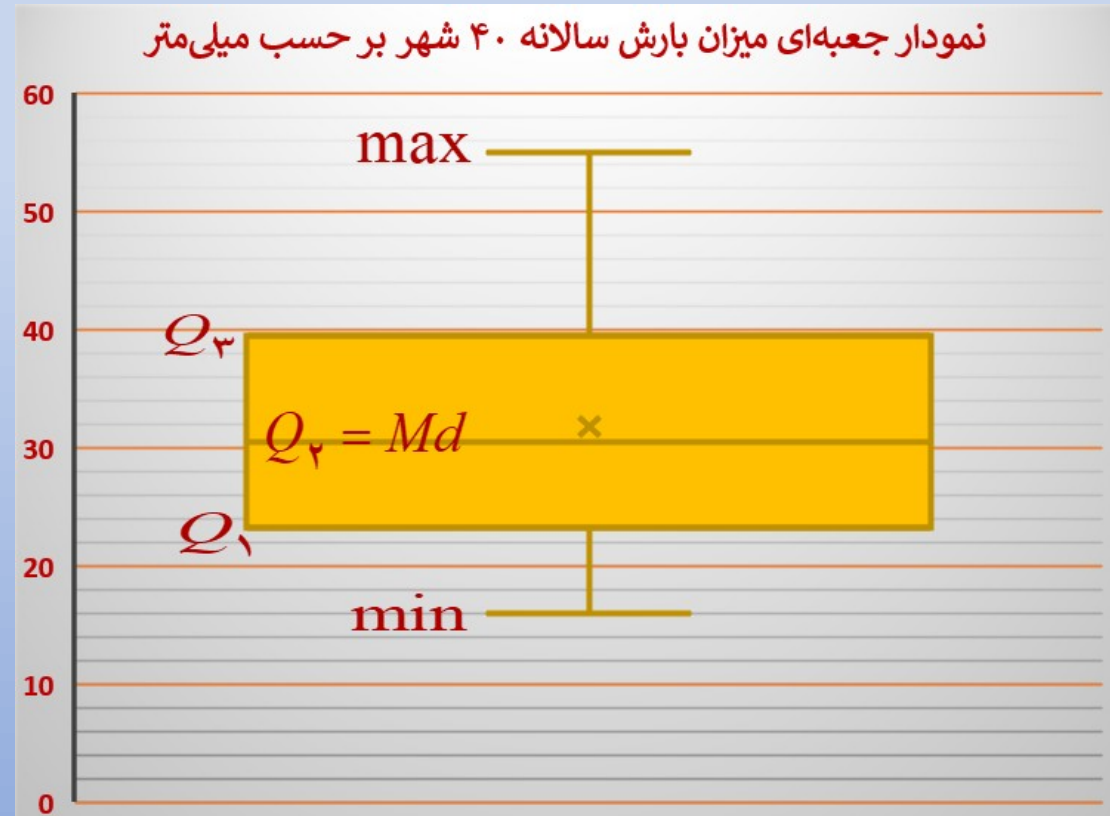
نمودار جعبه‌ای نشان‌دهنده نحوه توزیع داده‌ها در مقایسه با چارکها می‌باشد. معمولاً از این نمودار برای مقایسه دو یا چند جامعه آماری استفاده می‌شود.

میزان بارش در ۴۰ شهر در یکسال بر حسب میلی‌متر									
۱۶	۱۶	۱۷	۱۸	۲۰	۲۰	۲۰	۲۱	۲۲	۲۲
۲۳	۲۴	۲۴	۲۶	۲۷	۲۷	۲۸	۳۰	۳۰	۳۱
۳۴	۳۴	۳۵	۳۶	۳۹	۴۰	۴۰	۴۱	۴۲	۴۲
۴۳	۴۳	۴۳	۴۴	۴۶	۴۷	۴۹	۴۹	۵۰	۵۲

$$C_{Q_1} = \frac{40}{4} + \frac{1}{2} = 10,5 \Rightarrow Q_1 = 22 + \frac{1}{2}(23 - 22) = 22,5$$

$$C_{Q_2} = C_{Md} = \frac{2 \times 40}{4} + \frac{1}{2} = 20,5 \Rightarrow Q_2 = Md = 32,5$$

$$C_{Q_3} = \frac{3 \times 40}{4} + \frac{1}{2} = 30,5 \Rightarrow Q_3 = 42,5$$





فصل چهارم

توصیف مقداری

مشاهدات طبقه‌بندی شده

Numerical Description

of Classified Observations



مقدمه:

در مسایل واقعی معمولاً حجم جامعه آماری به اندازه‌ای زیاد است که توصیف تک تک داده‌ها از یک سو مستلزم صرف زمان و هزینه زیادی است و از سوی دیگر باعث ایجاد یک دید کلی و تحلیل جامع از الگوی رفتاری داده‌ها نمی‌گردد.

در این فصل پس از طبقه‌بندی داده‌ها و تشکیل جدول فراوانی به چهار سوال اساسی در مورد جامعه آماری پاسخ خواهیم داد.

- 1- مرکز توزیع فراوانی کجاست؟ (محاسبه پارامترهای مرکزی)
- 2- پراکندگی توزیع فراوانی چه مقدار است؟ (محاسبه پارامترهای پراکندگی)
- 3- مشاهدات جامعه به کدام سو تمایل دارد؟ (محاسبه انحراف از قرینگی)
- 4- پراکندگی جامعه در مقایسه با توزیع‌های مشابه چه وضعیتی دارد؟ (محاسبه شاخص کشیدگی)



پارامترهای مرکزی - میانگین

مثال: درآمد ماهیانه 100 کارمند شرکت الف در زیر داده شده است. میانگین، مد، چارک‌ها، دهک‌ها و صدک‌های درآمد کارمندان این شرکت را به دست آورید.

19 20 18 14 19 7 18 3 5 16 13 13 4 9 12 13 13 14 20 10
14 16 6 4 12 12 2 11 11 8 15 20 16 5 9 19 17 5 14 14
8 3 8 11 8 15 16 3 7 13 19 14 10 6 12 7 5 20 6 13
16 8 5 2 15 10 11 16 4 11 5 5 11 5 3 15 5 7 9 20
4 18 3 6 3 5 4 18 3 20 6 15 14 9 16 9 12 7 13 5

$$\mu = \frac{\sum x_i}{N} = \frac{1067}{100} = 10,67$$

max=	20	بزرگترین داده
min=	2	کوچکترین داده
R=	18	دامنه تغییرات
N=	100	تعداد داده ها
K=	10	تعداد طبقات
l=	2	طول طبقات

حد بالایی طبقات	حد پایینی طبقات	نماینده طبقات x	فراوانی مطلق F	F.x
1.5	3.5	2.5	9	22.5
3.5	5.5	4.5	16	72
5.5	7.5	6.5	10	65
7.5	9.5	8.5	10	85
9.5	11.5	10.5	9	94.5
11.5	13.5	12.5	12	150
13.5	15.5	14.5	12	174
15.5	17.5	16.5	8	132
17.5	19.5	18.5	8	148
19.5	21.5	20.5	6	123
			100	1066

$$\mu \simeq \frac{\sum F_i x_i}{N} = \frac{1066}{100} = 10,66$$



پارامترهای مرکزی - میانگین

روش غیر مستقیم (کدگذاری) در محاسبه میانگین: در این روش عدد دلخواه A به عنوان میانگین در نظر گرفته

می شود و سپس از رابطه زیر میانگین محاسبه

$$\mu \simeq A + \left(\frac{\sum F_i d_i}{N} \right) I$$

$$d_i = \frac{x_i - A}{I}$$

حد پایین طبقات	حد بالای طبقات	نماینده طبقات x	فراوانی مطلق F	d	F.d
1.5	3.5	2.5	9	-4	-36
3.5	5.5	4.5	16	-3	-48
5.5	7.5	6.5	10	-2	-20
7.5	9.5	8.5	10	-1	-10
9.5	11.5	10.5	9	0	0
11.5	13.5	12.5	12	1	12
13.5	15.5	14.5	12	2	24
15.5	17.5	16.5	8	3	24
17.5	19.5	18.5	8	4	32
19.5	21.5	20.5	6	5	30
			100		8

به عنوان مثال اگر فرض کنیم $A = 10/5$ خواهیم داشت:

$$\mu \simeq A + \left(\frac{\sum F_i d_i}{N} \right) I = 10/5 + \frac{8}{100} \times 2 = 10/66$$



پارامترهای مرکزی - مد

در جدول توزیع فراوانی تشخیص دقیق داده‌ای که بیشترین تکرار را دارد، امکان پذیر نیست و به همین دلیل از رابطه تقریبی زیر برای محاسبه مد استفاده می‌شود.

$$Mo \simeq L_{Mo} + \left(\frac{d_1}{d_1 + d_2} \right) I$$

که در آن (L_{Mo}) حد پایین طبقه مد (F_i) $d_1 = F_i - F_{i-1}$

فراوانی مطلق طبقه ماقبل از طبقه مد (F_i) $d_2 = F_i - F_{i+1}$

فراوانی مطلق طبقه ما بعد از طبقه مد (F_i) و () فاصله طبقات

فراوانی مطلق F	نماینده طبقات x	حد بالای طبقات	حد پایین طبقات
9	2.5	3.5	1.5
16	4.5	5.5	3.5
10	6.5	7.5	5.5
10	8.5	9.5	7.5
9	10.5	11.5	9.5
12	12.5	13.5	11.5
12	14.5	15.5	13.5
8	16.5	17.5	15.5
8	18.5	19.5	17.5
6	20.5	21.5	19.5
100			

$$\left. \begin{array}{l} d_1 = 16 - 9 = 7 \\ d_2 = 16 - 10 = 6 \\ I = 2, L_{Mo} = 3.5 \end{array} \right\} \Rightarrow Mo \simeq L_{Mo} + \left(\frac{d_1}{d_1 + d_2} \right) I = 3.5 + \frac{7}{7+6} \times 2 = 4.57$$



پارامترهای مرکزی - چارک‌ها و میانه

حد پایین طبقات	حد بالای طبقات	نماینده طبقات x	فراوانی مطلق F	فراوانی تجمعی FC
1.5	3.5	2.5	9	9
3.5	5.5	4.5	16	25
5.5	7.5	6.5	10	35
7.5	9.5	8.5	10	45
9.5	11.5	10.5	9	54
11.5	13.5	12.5	12	66
13.5	15.5	14.5	12	78
15.5	17.5	16.5	8	86
17.5	19.5	18.5	8	94
19.5	21.5	20.5	6	100
			100	

مراحل محاسبه چارک‌ها:

(1) جدول توزیع فراوانی تجمعی را تشکیل می‌دهیم.

(2) محل چارک a ام را از رابطه $C_{Q_a} = \frac{aN}{4}$ پیدا می

(3) در طبقه شامل چارک مقدار چارک را از رابطه زیر بدست

$$Q_a \simeq L_{Q_a} + \left(\frac{\frac{aN}{4} - FC_{i-1}}{F_i} \right) I$$

$$Q_1 \simeq 3.5 + \left(\frac{\frac{100}{4} - 9}{16} \right) \times 2 = 5.5$$

$$Md = Q_2 = 9.5 + \left(\frac{\frac{200}{4} - 45}{9} \right) \times 2 = 10.611$$

$$Q_3 = 13.5 + \left(\frac{\frac{300}{4} - 66}{12} \right) \times 2 = 15$$



پارامترهای مرکزی - دهک‌ها

مراحل محاسبه دهک‌ها:

(1) جدول توزیع فراوانی تجمعی را تشکیل می‌دهیم.

(2) محل دهک a را از رابطه $C_{D_a} = \frac{aN}{10}$ پیدا می‌کنیم.

(3) در طبقه شامل دهک مقدار دهک را از رابطه زیر بدست می‌آوریم.

$$D_a \simeq L_{D_a} + \left(\frac{\frac{aN}{10} - FC_{i-1}}{F_i} \right) I$$

$$D_1 \simeq 3,5 + \left(\frac{\frac{100}{10} - 9}{16} \right) \times 2 = 3,625$$

$$D_2 \simeq 3,5 + \left(\frac{\frac{200}{10} - 9}{16} \right) \times 2 = 4,875$$

$$Md = D_5 = 9,5 + \left(\frac{\frac{500}{10} - 45}{9} \right) \times 2 = 10,611$$

$$D_7 = 13,5 + \left(\frac{\frac{700}{10} - 66}{12} \right) \times 2 = 16,167$$

حد پایین طبقات	حد بالای طبقات	نماینده طبقات x	فراوانی مطلق F	فراوانی تجمعی FC
1.5	3.5	2.5	9	9
3.5	5.5	4.5	16	25
5.5	7.5	6.5	10	35
7.5	9.5	8.5	10	45
9.5	11.5	10.5	9	54
11.5	13.5	12.5	12	66
13.5	15.5	14.5	12	78
15.5	17.5	16.5	8	86
17.5	19.5	18.5	8	94
19.5	21.5	20.5	6	100
			100	



پارامترهای مرکزی - صدک‌ها

مراحل محاسبه صدک‌ها:

(1) جدول توزیع فراوانی تجمعی را تشکیل می‌دهیم.

(2) محل صدک a را از رابطه $C_{P_a} = \frac{aN}{100}$ پیدا می‌کنیم.

$$P_{20} \simeq 3,5 + \left(\frac{\frac{20 \times 1000}{100} - 90}{160} \right) \times 2 = 4,875$$

(3) در طبقه شامل صدک، مقدار صدک را از رابطه زیر بدست می‌آوریم.

$$P_a \simeq L_{P_a} + \left(\frac{\frac{aN}{100} - FC_{i-1}}{F_i} \right) I$$

$$Md = P_{50} = 9,5 + \left(\frac{\frac{50 \times 1000}{100} - 450}{90} \right) \times 2 = 10,611$$

$$P_{75} = 13,5 + \left(\frac{\frac{75 \times 1000}{100} - 660}{120} \right) \times 2 = 15$$

حد پایین طبقات	حد بالای طبقات	نماینده طبقات x	فراوانی مطلق F	فراوانی تجمعی FC
1.5	3.5	2.5	90	90
3.5	5.5	4.5	160	250
5.5	7.5	6.5	100	350
7.5	9.5	8.5	100	450
9.5	11.5	10.5	90	540
11.5	13.5	12.5	120	660
13.5	15.5	14.5	120	780
15.5	17.5	16.5	80	860
17.5	19.5	18.5	80	940
19.5	21.5	20.5	60	1000
			1000	



پارامترهای پراکندگی - انحراف متوسط از میانگین

برای محاسبه انحراف متوسط از میانگین از رابطه $A.D_\mu = \frac{\sum F_i |x_i - \mu|}{N}$ استفاده می‌کنیم.

C-L	F	x	Fx	F x-μ
۱۰ - ۲۰	۱۵	۱۵	۲۲۵	۲۷۰
۲۰ - ۳۰	۳۰	۲۵	۷۵۰	۲۴۰
۳۰ - ۴۰	۲۵	۳۵	۸۷۵	۵۰
۴۰ - ۵۰	۲۰	۴۵	۹۰۰	۲۴۰
۵۰ - ۶۰	۱۰	۵۵	۵۵۰	۲۲۰
Σ	۱۰۰		۳۳۰۰	۱۰۲۰

مثال: دستمزد ماهیانه کارکنان شرکت ایران دارو
در صورت مقابل دسته‌بندی شده است. میانگین دستمزد
ماهیانه و انحراف متوسط از میانگین را محاسبه کنید.

$$\mu \simeq \frac{\sum F_i x_i}{N} = \frac{3300}{100} = 33$$

$$A.D_\mu = \frac{\sum F_i |x_i - \mu|}{N} = \frac{1020}{100} = 10.2$$



پارامترهای پراکندگی - واریانس و انحراف معیار

روش مستقیم: با استفاده از یکی از روابط زیر واریانس

C-L	F	x	Fx	Fx ^۲	F(x-μ) ²
۱۰ - ۲۰	۱۵	۱۵	۲۲۵	۳۳۷۵	۴۸۶۰
۲۰ - ۳۰	۳۰	۲۵	۷۵۰	۱۸۷۵۰	۱۹۲۰
۳۰ - ۴۰	۲۵	۳۵	۸۷۵	۳۰۶۲۵	۱۰۰
۴۰ - ۵۰	۲۰	۴۵	۹۰۰	۴۰۵۰۰	۲۸۸۰
۵۰ - ۶۰	۱۰	۵۵	۵۵۰	۳۰۲۵۰	۴۸۴۰
Σ	۱۰۰		۳۳۰۰	۱۲۳۵۰۰	۱۴۶۰۰

در داده‌های طبقه‌بندی شده قابل محاسبه است

$$\sigma_x^2 = \frac{\sum F_i (x_i - \mu)^2}{N}$$

$$\sigma_x^2 = \frac{\sum F_i x_i^2}{N} - \mu_x^2$$

$$\mu = \frac{\sum F_i x_i}{N} = \frac{3300}{100} = 33, \quad \sigma_x^2 = \frac{\sum F_i (x_i - \mu)^2}{N} = \frac{14600}{100} = 146$$

$$\sigma_x^2 = \frac{\sum F_i x_i^2}{N} - \mu_x^2 = \frac{123500}{100} - (33)^2 = 1235 - 1089 = 146$$

$$\sigma_x = \sqrt{146} \simeq 12,083$$



پارامترهای پراکندگی - واریانس و انحراف معیار

روش غیرمستقیم (کد گذاری): اگر میانگین جامعه را با عدد A تقریب بزنیم،

واریانس از رابطه زیر قابل محاسبه است.

$$\sigma_x^2 = I^2 \left(\frac{\sum F_i d_i^2}{N} - \left(\frac{\sum F_i d_i}{N} \right)^2 \right)$$

که در آن $d_i = \frac{x_i - A}{I}$ و I فاصله طبقات

C-L	F	x	d	Fd	Fd ^۲
۱۰ - ۲۰	۱۵	۱۵	-۱.۵	-۲۲.۵	۳۳.۷۵
۲۰ - ۳۰	۳۰	۲۵	-۰.۵	-۱۵	۷.۵
۳۰ - ۴۰	۲۵	۳۵	۰.۵	۱۲.۵	۶.۲۵
۴۰ - ۵۰	۲۰	۴۵	۱.۵	۳۰	۴۵
۵۰ - ۶۰	۱۰	۵۵	۲.۵	۲۵	۶۲.۵
Σ	۱۰۰			۳۰	۱۵۵

$$A = 30$$

$$\sigma_x^2 = I^2 \left(\frac{\sum F_i d_i^2}{N} - \left(\frac{\sum F_i d_i}{N} \right)^2 \right) = 10^2 \times \left(\frac{155}{100} - \left(\frac{30}{100} \right)^2 \right) = 100 \times (1.55 - 0.09) = 146$$

$$\sigma_x = \sqrt{146} \simeq 12.083$$



پارامترهای پراکندگی - دامنه میان چارکی و انحراف چارکی

در صورتی که مشاهدات دارای طبقه بندی باز باشند و تعداد مشاهدات زیاد نباشد، محاسبه میانگین و انحراف معیار ممکن نیست. در چنین شرایطی از دامنه میان چارکی و انحراف چارکی برای

C-L	F	FC
< ۵	۲	۲
۵ - ۸	۱۰	۱۲
۸ - ۱۱	۲۵	۳۷
۱۱ - ۱۴	۲۳	۶۰
۱۴ - ۱۷	۱۵	۷۵
> ۱۷	۵	۸۰
Σ	۸۰	

$$IQR = Q_3 - Q_1$$

سنجش پراکندگی جامعه استفاده می شود:

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

مثال: شاخص شایستگی کارکنان از صفر تا 20 در جدول مقابل داده شده است.

دامنه میان چارکی و انحراف چارکی را محاسبه کنید.

$$\left. \begin{aligned} C_{Q_1} &= \frac{1 \times 80}{4} = 20 \Rightarrow Q_1 \simeq L_{Q_1} + \frac{\frac{1 \times N}{4} - FC_{i-1}}{F_i} \times I = 8 + \frac{20 - 12}{25} \times 3 = 8,96 \\ C_{Q_3} &= \frac{3 \times 80}{4} = 60 \Rightarrow Q_3 \simeq L_{Q_3} + \frac{\frac{3 \times N}{4} - FC_{i-1}}{F_i} \times I = 14 + \frac{60 - 60}{15} \times 3 = 14 \end{aligned} \right\} \Rightarrow \begin{cases} IQR = Q_3 - Q_1 = 14 - 8,96 = 5,04 \\ SIQR = \frac{Q_3 - Q_1}{2} = \frac{5,04}{2} = 2,52 \end{cases}$$



پارامترهای پراکندگی - تصحیح شپارد در واریانس

همانطور که می‌دانیم طبقه‌بندی داده‌ها باعث از بین رفتن برخی اطلاعات و در نتیجه به وجود آمدن خطا در محاسبه پارامترهای آماری می‌گردد. هنگام محاسبه میانگین، به جای هر عضو در یک طبقه داده‌ای که در وسط طبقه قرار دارد را جایگزین می‌کنیم و بنابراین خطای به وجود آمده در محاسبه میانگین برای برخی داده‌ها مثبت و برای برخی دیگر منفی می‌باشد به طوری که از آن صرف نظر می‌کنیم. اما در محاسبه واریانس تمام خطاها مثبت و در نتیجه قابل اغماض نیستند.

برای تصحیح این خطا، شپارد پیشنهاد کرده از واریانس به دست آمده، مقدار $\frac{I^2}{12}$ کم شود. (I فاصله طبقات است)

$$\sigma_c^2 = \sigma_x^2 - \frac{I^2}{12}$$

این تصحیح معمولاً هنگامی استفاده می‌شود که اولاً متغیر پیوسته باشد و ثانیاً تعداد داده‌ها بیشتر از 1000 باشد.



پارامترهای پراکندگی - تصحیح شپارد در واریانس

19 20 18 14 19 7 18 3 5 16 13 13 4 9 12 13 13 14 20 10
14 16 6 4 12 12 2 11 11 8 15 20 16 5 9 19 17 5 14 14
8 3 8 11 8 15 16 3 7 13 19 14 10 6 12 7 5 20 6 13
16 8 5 2 15 10 11 16 4 11 5 5 11 5 3 15 5 7 9 20
4 18 3 6 3 5 4 18 3 20 6 15 14 9 16 9 12 7 13 5

max=	20	بزرگترین داده
min=	2	کوچکترین داده
R=	18	دامنه تغییرات
N=	100	تعداد داده ها
K=	10	تعداد طبقات
I=	2	طول طبقات

مثال: درآمد ماهیانه 100 کارمند شرکت

الف در جدول مقابل داده شده است

واریانس درآمد کارمندان این شرکت

را به دست آورید.

$$\sigma_c^2 = \sigma_x^2 - \frac{I^2}{12} = 29,8944 - \frac{4}{12} = 29,561067$$

حد پایین طبقات	حد بالای طبقات	نماینده طبقات x	فراوانی مطلق F	F*x	F*x^۲
1.5	3.5	2.5	9	22.5	56.25
3.5	5.5	4.5	16	72	324
5.5	7.5	6.5	10	65	422.5
7.5	9.5	8.5	10	85	722.5
9.5	11.5	10.5	9	94.5	992.25
11.5	13.5	12.5	12	150	1875
13.5	15.5	14.5	12	174	2523
15.5	17.5	16.5	8	132	2178
17.5	19.5	18.5	8	148	2738
19.5	21.5	20.5	6	123	2521.5
			100	1066	14353

واریانس	29.8944
واریانس واقعی	28.3811
واریانس شپارد	29.56106667



عملیات جبری میانگین و واریانس

فرض کنیم K جامعه آماری با تعداد مشاهدات $N_1, N_2, \dots, N_K$ ، با میانگین‌های $\mu_1, \mu_2, \dots, \mu_K$ و واریانس‌های $\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2$ تشکیل یک جامعه آماری واحد با تعداد مشاهدات $N = N_1 + N_2 + \dots + N_K$ را داده باشند. در این صورت میانگین و واریانس جامعه آماری به وجود آمده از روابط زیر به دست می‌آیند.

$$\mu = \frac{N_1\mu_1 + N_2\mu_2 + \dots + N_K\mu_K}{N_1 + N_2 + \dots + N_K} = \frac{\sum N_i \mu_i}{N}$$

$$\sigma^2 = \frac{\sum N_i \sigma_i^2}{N} + \frac{\sum N_i (\mu_i - \mu)^2}{N}$$

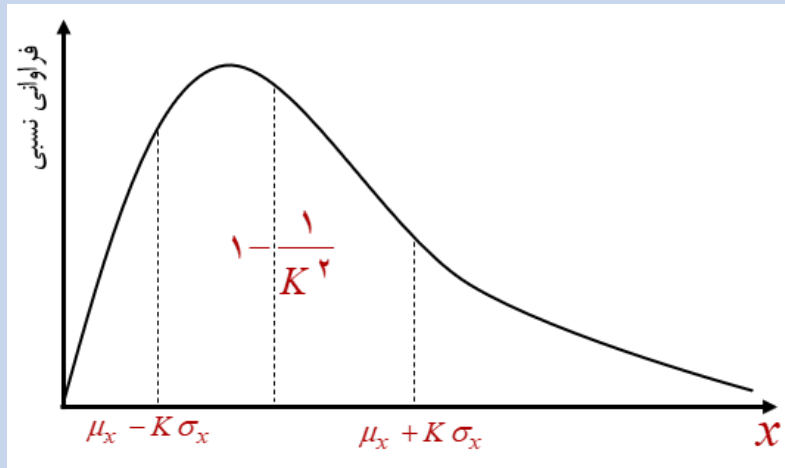
همانطور که مشاهده می‌شود واریانس جامعه به وجود آمده بزرگتر یا مساوی میانگین واریانس جوامع اولیه است.

$$\sigma^2 \geq \frac{\sum N_i \sigma_i^2}{N}$$

تساوی در حالتی برقرار است که میانگین همه جوامع با هم برابر باشند.



اهمیت کاربردی انحراف معیار



$$\frac{1}{K^2} \text{ حداقل}$$

قضیه چبی شف: برای N مشاهده $X_1, X_2, \dots, X_N$

از مشاهدات، در دامنه K انحراف معیار از میانگین قرار می گیرند.

تاکید بر «حداقل» در قضیه به دلیل قابلیت تعمیم آن به انواع توزیع های فراوانی می باشد.

K	$1 - \frac{1}{K^2}$
1	0.0%
1.5	55.6%
2	75.0%
2.5	84.0%
3	88.9%
3.5	91.8%

به عنوان مثال در سازمانی که میانگین نمرات مسوولیت پذیری برابر 65 و

انحراف معیار نمرات برابر 10 باشد، حداقل 75 درصد نمرات در بازه (45 ، 85) قرار



اهمیت کاربردی انحراف معیار - ادامه

K	$1 - \frac{1}{K^2}$
1	۰.۰%
1.۵	۵۵.۶%
۲	۷۵.۰%
۲.۵	۸۴.۰%
۳	۸۸.۹%
۳.۵	۹۱.۸%

C-L	X	F	f	$F \cdot X$	$F \cdot X^2$
۱۰ - ۲۰	۱۵	۱۵	۰.۱۵	۲۲۵	۳۳۷۵
۲۰ - ۳۰	۲۵	۳۰	۰.۳	۷۵۰	۱۸۷۵۰
۳۰ - ۴۰	۳۵	۲۵	۰.۲۵	۸۷۵	۳۰۶۲۵
۴۰ - ۵۰	۴۵	۲۰	۰.۲	۹۰۰	۴۰۵۰۰
۵۰ - ۶۰	۵۵	۱۰	۰.۱	۵۵۰	۳۰۲۵۰
Σ		۱۰۰		۳۳۰۰	۱۲۳۵۰۰
		$\mu = ۳۳$		$\sigma = ۱۲.۱$	

W

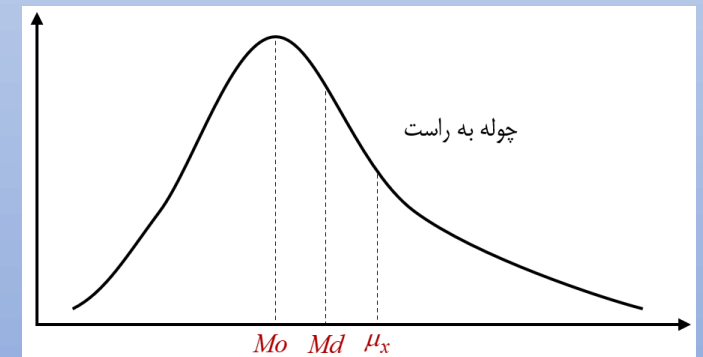
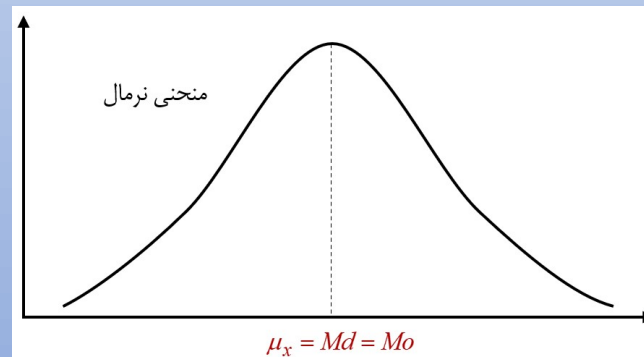
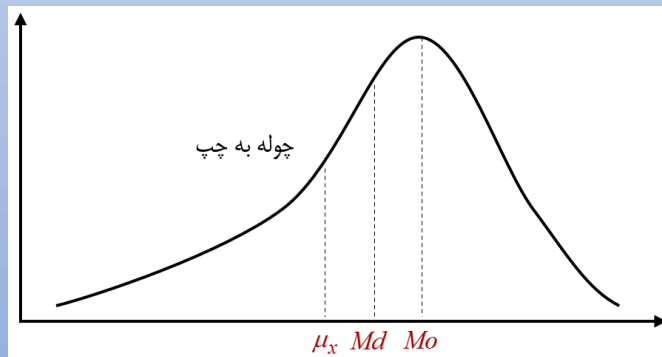
K	$\mu_x \pm K\sigma_x$	F	f	درصدهای قضیه چبیشف	درصدهای منحنی نرمال	اختلاف با درصدهای نرمال
1	۲۰.۹ - ۴۵.۰۸۳	۶۵	۰.۶۵	۰.۰%	۶۸.۲۶%	۳.۲۶%
۲	۸.۸۳ - ۵۷.۱۶۶	۹۵	۰.۹۵	۵۵.۶%	۹۵.۴۴%	۰.۴۴%
۳	-۳.۲۵ - ۶۹.۲۴۹	۱۰۰	۱	۷۵.۰%	۹۹.۷۳%	۰.۲۷%



پارامترهای تعیین انحراف از قرینگی

در مقایسه دو یا چند جامعه آماری ممکن است پارامترهای مرکزی و پارامترهای پراکندگی به دلیل برابر بودن یا داشتن اختلاف جزئی کفایت نکنند. در این شرایط می‌توان از پارامترهایی انحراف از قرینگی استفاده کرد.

در مقایسه با منحنی نرمال که کاملاً متقارن است (میانه، مد و میانگین بر هم منطبق هستند)، انحراف از قرینگی (چولگی) سایر منحنی‌های توزیع‌های آماری تعریف می‌گردد. همانطور که می‌دانیم میانه در وسط داده‌های آماری قرار دارد. در صورتیکه مد بزرگتر و میانگین کوچکتر از میانه باشد، منحنی دارای چوله به چپ و پارامتر چولگی منفی و در حالت عکس منحنی دارای چوله به راست و پارامتر چولگی مثبت است.





پارامترهای تعیین انحراف از قرینگی - ادامه

برای محاسبه ضریب چولگی (Skewness)، روابط مختلفی ارائه شده است که مهمترین آنها با توجه به گشتاور

مرتبه سوم به صورت زیر تعریف می شود:

$$S.K = \frac{r_3}{\sigma_x^3}, \quad r_3 = \frac{\sum F_i (x_i - \mu_x)^3}{N}$$

همچنین با توجه به رابطه تجربی پیرسون $(\mu_x - Mo) = 3(\mu_x - Md)$ دو رابطه دیگر برای محاسبه ضریب چولگی به صورت زیر ارائه شده است.

$$S.K_1 = \frac{(\mu_x - Mo)}{\sigma_x}, \quad S.K_2 = \frac{3(\mu_x - Md)}{\sigma_x}$$

در مواردی که محاسبه میانگین و واریانس امکان پذیر نباشد، با توجه به چندکها، ضریب چولگی به صورت زیر محاسبه می گردد.

$$S.K_Q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}, \quad S.K_P = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$



پارامترهای تعیین انحراف از قرینگی - ادامه

۱۰	۲۱	۳۲	۳۸	۵۰	۵۳	۶۴	۷۳	۸۵	۹۸
۱۰	۲۴	۳۳	۴۹	۵۱	۵۴	۶۴	۷۳	۸۷	۱۰۰
۱۴	۲۵	۳۴	۴۹	۵۱	۵۷	۶۵	۷۵	۸۹	۱۰۶
۱۵	۲۷	۳۵	۵۰	۵۱	۶۰	۶۸	۷۷	۹۲	۱۱۰
۱۷	۳۰	۳۷	۵۰	۵۱	۶۳	۷۰	۸۰	۹۵	۱۲۰

مثال: داده‌های مرتب شده مقابل مربوط به سود سالانه 50 شرکت بزرگ 10 میلیون ریال است. میانگین، مد، میانه، چارک اول و سوم، صدک دوم و نودم، انحراف معیار، ضریب چولگی شماره 1 و 2 پیرسون و ضریب چارکی را حساب کنید.

$$\mu_x = 55,7, \quad Mo = 42,7, \quad Md = 51,5$$

$$Q_1 = 23,5 + \left(\frac{12,5 - 6}{9} \right) \times 14 = 36,61$$

$$Q_3 = 65,5 + \left(\frac{37,5 - 33}{6} \right) \times 14 = 76$$

$$P_{10} = 21,17, \quad P_{90} = 97, \quad \sigma_x = 27,434$$

$$S.K_1 = \frac{(55,7 - 42,17)}{27,434} = 0,493, \quad S.K_2 = \frac{3(55,7 - 51,5)}{27,434} = 0,459$$

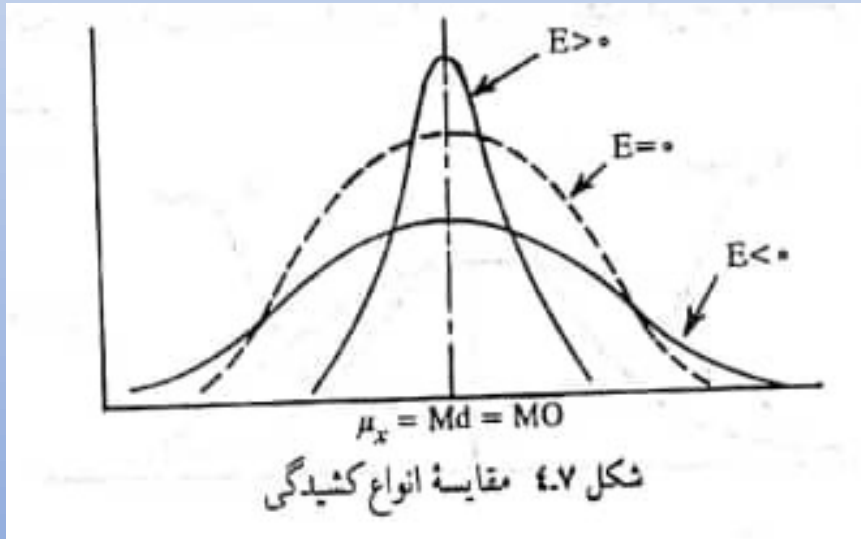
$$S.K_Q = \frac{76 - 2(51,5) - 36,61}{27,434} = 0,35, \quad S.K_P = \frac{97 - 2(51,5) + 21,17}{27,434} = 0,55$$

C-L	حدود کمرانه	F_i	FC_i	x_i	$F_i x_i$
۱۰-۲۳	۹/۵-۲۳/۵	۶	۶	۱۶/۵	۹۹
۲۴-۳۷	۲۳/۵-۳۷/۵	۹	۱۵	۳۰/۵	۲۷۴/۵
۳۸-۵۱	۳۷/۵-۵۱/۵	۱۰	۲۵	۴۴/۵	۴۴۵
۵۲-۶۵	۵۱/۵-۶۵/۵	۸	۳۳	۵۸/۵	۴۶۸
۶۶-۷۹	۶۵/۵-۷۹/۵	۶	۳۹	۷۲/۵	۴۳۵
۸۰-۹۳	۷۹/۵-۹۳/۵	۵	۴۴	۸۶/۵	۴۳۲/۵
۹۴-۱۰۷	۹۳/۵-۱۰۷/۵	۴	۴۸	۱۰۰/۵	۴۰۲
۱۰۸-۱۲۱	۱۰۷/۵-۱۲۱/۵	۲	۵۰	۱۱۴/۵	۲۲۹
N = ۵۰					۲۷۸۵



پارامترهای تعیین انحراف از کشیدگی Parameters Deviation from Elongation

منظور از مقایسه کشیدگی توزیع‌ها، مقایسه مقدار اوج (بلندی) آنها می‌باشد. در مقایسه با منحنی نرمال اگر بلندی یک توزیع بیشتر باشد، دارای کشیدگی مثبت و اگر بلندی آن کمتر باشد، دارای کشیدگی منفی است. اگر مقدار کشیدگی که از روابط زیر به دست می‌آید را با $|E| \leq 0.10$ نمایش دهیم، آنگاه در صورتی که باشد، توزیع جامعه از نظر پراکندگی تقریباً نرمال است و اگر $0.10 < |E| \leq 0.50$ باشد، توزیع از نظر کشیدگی تفاوت کمی با نرمال دارد و اگر باشد، تفاوت توزیع با توزیع نرمال از نظر پراکندگی فاحش



است. برای محاسبه ضریب کشیدگی $\sum_{i=1}^N (x_i - \mu_x)^4$ از گشتاور مرتبه چهارم استفاده

$$E = \frac{r_4}{\sigma_x^4} - 3, \quad r_4 = \frac{\sum_{i=1}^N (x_i - \mu_x)^4}{N}$$

در صورتی که بخواهیم ضریب کشیدگی $\frac{SFQR}{P_{90} - P_{10}}$ از چندگاه محاسبه کنیم:



پارامترهای تعیین انحراف از کشیدگی - ادامه

C-L	F	x	Fx	$F(x-\mu)^2$	$F(x-\mu)^4$
۱۰ - ۲۰	۱۵	۱۵	۲۲۵	۴۸۶۰	۱۵۷۴۶۴۰
۲۰ - ۳۰	۳۰	۲۵	۷۵۰	۱۹۲۰	۱۲۲۸۸۰
۳۰ - ۴۰	۲۵	۳۵	۸۷۵	۱۰۰	۴۰۰
۴۰ - ۵۰	۲۰	۴۵	۹۰۰	۲۸۸۰	۴۱۴۷۲۰
۵۰ - ۶۰	۱۰	۵۵	۵۵۰	۴۸۴۰	۲۳۴۲۵۶۰
Σ	۱۰۰		۳۳۰۰	۱۴۶۰۰	۴۴۵۵۲۰۰

مثال: ضریب کشیدگی گشتاوری و ضریب کشیدگی چندکی داد
جدول مقابل را محاسبه و آن را با منحنی نرمال مقایسه کنید.

$$\mu_x = \frac{3300}{100} = 33, \quad \sigma_x^2 = \frac{14600}{100} = 146, \quad \sigma_x = \sqrt{146} = 12,083$$

$$r_4 = \frac{4455200}{100} = 44552, \quad E = \frac{44552}{(12,083)^4} - 3 = -0,91, \quad |E| = |-0,91| > 0,5$$

تفاوت فاحش با توزیع نرمال وجود دارد.



پارامترهای تعیین انحراف از کشیدگی - ادامه

C-L	F	x	FC
۱۰ - ۲۰	۱۵	۱۵	۱۵
۲۰ - ۳۰	۳۰	۲۵	۴۵
۳۰ - ۴۰	۲۵	۳۵	۷۰
۴۰ - ۵۰	۲۰	۴۵	۹۰
۵۰ - ۶۰	۱۰	۵۵	۱۰۰
Σ	۱۰۰		

برای محاسبه ضریب کشیدگی چندی از رابطه زیر در جدول توزیع فراوانی استفاده

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

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

$$P_a \simeq L_{P_a} + \left(\frac{\frac{aN}{100} - FC_{i-1}}{F_i} \right) I$$

$$P_{10} \simeq 10 + \left(\frac{\frac{10 \times 100}{100} - 0}{15} \right) \times 10 = 16,67$$

$$Q_1 \simeq 20 + \left(\frac{\frac{1 \times 100}{4} - 15}{25} \right) \times 10 = 24$$

$$P_{90} \simeq 40 + \left(\frac{\frac{90 \times 100}{100} - 70}{45} \right) \times 10 = 44,44$$

$$Q_3 \simeq 40 + \left(\frac{\frac{3 \times 100}{4} - 70}{45} \right) \times 10 = 41,11$$

$$SIQR = \frac{41,11 - 24}{2} = 8,555$$

$$E = \frac{8,555}{44,44 - 16,67} - 0,263 = 0,045$$



فصل پنجم

احتمال

Probability



نظریه احتمال

در شرایط عدم اطمینان و تصمیم‌گیری در مورد آینده، نظریه احتمال با سازماندهی اطلاعات موجود می‌تواند کمک شایانی به مدیران نماید. در واقع به جای محاسبات سرانگشتی یا رها کردن تیری در تاریکی، نظریه احتمال کمک می‌کند تا بتوانیم به صورت سازمان‌یافته، اطلاعات را پردازش و در مورد وقایعی که قرار است در آینده رخ دهند، اظهار نظر کنیم.

عبارتی از قبیل «احتمال بارش باران فردا در تهران 70 درصد می‌باشد.» یا «احتمال وجود تقاضا برای محصولاتی با قیمت بیش از یک میلیون تومان، 35 درصد است» که نمونه‌ای از نتایج نظریه احتمال هستند، دید بهتری به مدیران سازمان‌ها می‌دهند تا بتوانند تصمیمات بهتری اتخاذ کنند.

دانشمندانی همچون *برنولی، دوماوآور، لاگرانژ و لاپلاس* از جمله افرادی هستند که در ایجاد و بسط نظریه احتمال نقش به سزایی داشتند.



برخی از مفاهیم اساسی احتمال

مفهوم احتمال: احتمال یعنی شانس وقوع یک پیشامد خاص و احتمال وقوع یک پیشامد برابر است با نسبت دفعاتی که پیشامد خاصی در تکرارهای زیاد رخ می‌دهد.

احتمال عینی و ذهنی: احتمال عینی، ثابت و مقدار آن از قبل مشخص و قابل محاسبه است و به عقاید اشخاص بستگی ندارد در حالی که احتمال ذهنی، متغیر و وابسته به نظر اشخاص است.

آزمایش: فعالیتی که نتیجه آن از قبل مشخص نیست ولی کل حالات ممکن آن معلوم است. مثل پرتاب یک سکه، که معلوم نیست دقیقاً شیر خواهد آمد یا خط ولی غیر از این دو حالت، حالت دیگری وجود ندارد.

فضای نمونه: مجموعه پیامدهای ممکن یک آزمایش را فضای نمونه آن آزمایش می‌گویند. فضای نمونه را با S نشان می‌دهند.

مثال: فضای نمونه پرتاب یک سکه $S = \{H, T\}$ و فضای نمونه پرتاب یک تاس $S = \{1, 2, 3, 4, 5, 6\}$ و فضای نمونه پرتاب دو سکه



برخی از مفاهیم اساسی احتمال

فضای نمونه محدود و نامحدود: اگر تعداد عضوهای فضای نمونه‌ای مشخص باشد، محدود و در غیر این صورت نامحدود است.

به عنوان مثال فضای نمونه پرتاب یک تاس محدود و دارای 6 عضو است ولی اگر فرض کنیم آزمایشی برای رسیدن به اولین محصول تولید شده معیوب در یک خط تولید طراحی شده باشد، پیشامد تعداد محصول تست شده تا رسیدن به محصول معیوب یک فضای نمونه نامتناهی خواهد داشت.

فضای نمونه پیوسته و گسسته: در صورتیکه فضای نمونه یک پیشامد محدود (با تعداد اعضای مشخص) یا نامحدود ولی شمارش‌پذیر (متناظر با اعداد طبیعی) باشد، آن را گسسته و در غیر اینصورت فضای نمونه را پیوسته می‌نامیم.

$$S = \{x \mid 0 \leq x \leq 1500\}$$

به عنوان مثال فضای نمونه برای نوعی لامپ که حداکثر عمر آن 1500 ساعت است، یک فضای نمونه پیوسته می‌باشد.



برخی از مفاهیم اساسی احتمال

پیشامد: به هر یک از زیرمجموعه‌های یک فضای نمونه، یک پیشامد گفته می‌شود. به عنوان مثال پیشامد ظاهر شدن عدد زوج در پرتاب یک تاس به صورت $A = \{2, 4, 6\}$ است.

پیشامدهای مقدماتی هم‌شانس: هرگاه تمام پیشامدهای اولیه در آزمایشی دارای شانس وقوع برابر باشند یعنی نوعی تقارن در آزمایش وجود داشته باشد، پیشامدهای مقدماتی (پیامدهای اولیه) را هم‌شانس می‌نامیم. به عنوان مثال در آزمایش پرتاب یک تاس، شانس ظاهر شدن هر یک از اعداد 1 تا 6 برابر و هیچیک از اعداد بر سایرین برتری ندارد.



مثال: فضای نمونه پرتاب دوسکه به طور همزمان به صورت

است. با توجه به اینکه شانس وقوع هر یک از چهار پیشامد ممکن یکسان است و هیچ یک از پیشامدها نسبت به بقیه برتری ندارد، هر یک از پیامدهای مقدماتی هم‌شانس هستند.



احتمال یک پیشامد

با توجه به اینکه احتمال به صورت شانس وقوع یک پیشامد تعریف می‌شود، در صورتیکه همه پیشامدهای مقدماتی در یک فضای نمونه هم‌شانس باشند، احتمال یک پیشامد مانند A در فضای نمونه S به صورت زیر محاسبه می‌شود.

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



مثال: در پرتاب دو سکه، هر یک از احتمالات زیر را محاسبه کنید.

$$A = \{(T, T)\} \Rightarrow P(A) = \frac{n(A)}{n(S)} = \frac{1}{4}$$

(الف) دقیقاً دو خط ظاهر شود.

$$A = \{(H, H), (H, T), (T, H)\} \Rightarrow P(A) = \frac{n(A)}{n(S)} = \frac{3}{4}$$

(ب) حداقل یک شیر ظاهر شود.

$$A = \{(H, H), (T, T)\} \Rightarrow P(A) = \frac{n(A)}{n(S)} = \frac{2}{4} = \frac{1}{2}$$

(ج) هر دو یک چیز باشند.



احتمال و فراوانی نسبی

در بسیاری از آزمایش‌ها شانس وقوع پیشامدهای مقدماتی یکسان نیستند و محاسبه احتمال به صورت نسبت تعداد اعضای پیشامد به تعداد اعضای فضای نمونه صحیح نمی‌باشد. به عنوان مثال احتمال بروز خطای محاسباتی یک حسابدار در شروع نوبت کاری و در ساعات پایانی آن یکسان نیست یا احتمال بروز تصادف در جاده‌های کشور در ساعات مختلف شبانه روز یا روزهای خاص سال یا هفته، یکسان نیست. در چنین شرایطی برای محاسبه احتمال یک پیشامد، آن را به دفعات (خیلی زیاد) تکرار می‌کنیم و فراوانی نسبی وقوع پیشامد مورد نظر را معادل احتمال وقوع آن پیشامد در نظر می‌گیریم. باید توجه داشت تعداد دفعات تکرار آزمایش هر چه بیشتر باشد، مقدار دقیق‌تری برای احتمال وقوع پیشامد به دست می‌آید.

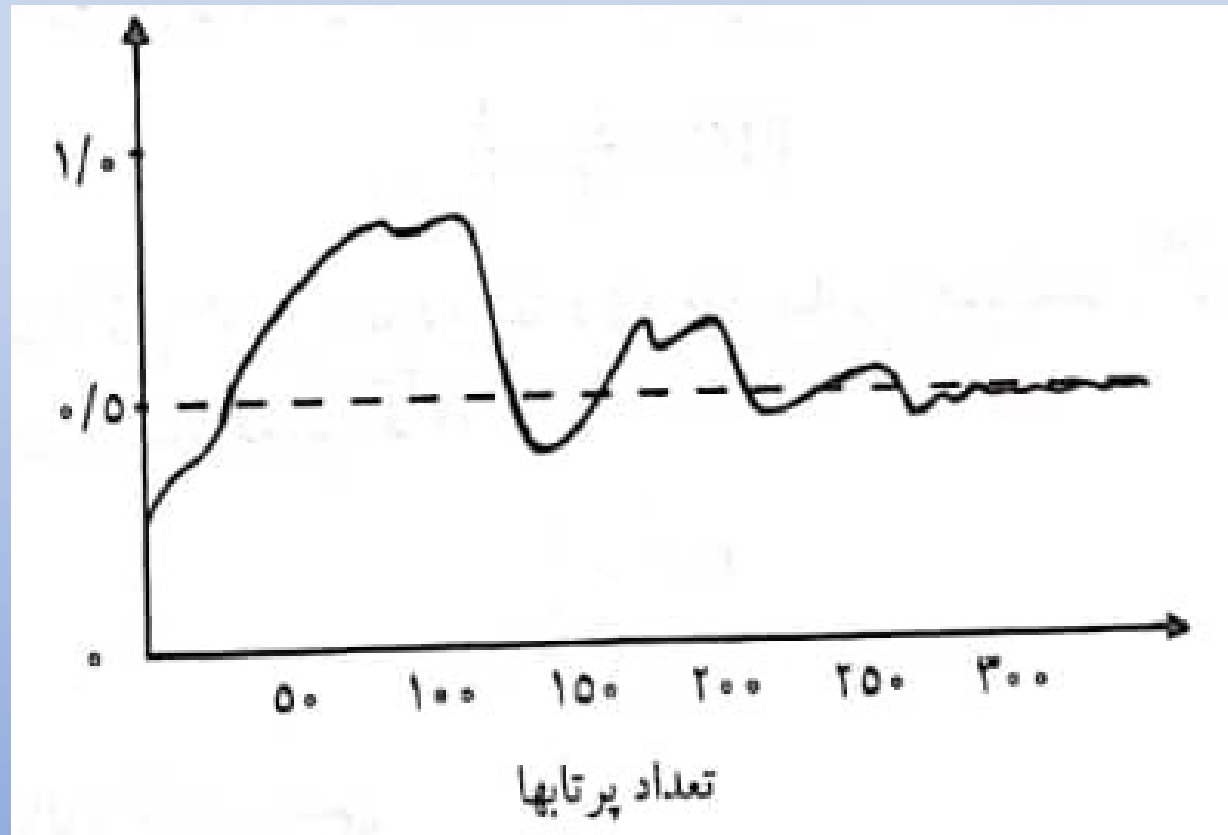
$$P(A) = \lim_{N \rightarrow \infty} \frac{\text{The number of times A occurs in N repetitions of the experiment}}{N}$$

مثلا برای محاسبه احتمال ظاهر شدن شیر در پرتاب یک سکه نمیتوان از سه پرتاب که دو بار شیر و یک بار خط ظاهر شده نتیجه گرفت که احتمال ظاهر شدن شیر دو سوم است. بلکه در تعداد زیاد تکرار این آزمایش باید نسبت به مقدار احتمال وقوع پیشامد مورد نظر قضاوت کرد.



احتمال و فراوانی نسبی

همانطور که مشاهده می‌شود بعد از 300 بار پرتاب یک سکه، فراوانی نسبی پیشامد ظاهر شدن شیر به سمت 0/5 میل کرده است. بنابراین احتمال ظاهر شدن شیر در پرتاب یک سکه را در بینهایت بار پرتاب برابر 0/5 در نظر می‌گیریم.





احتمال و فراوانی نسبی

مثال: در نمونه وسیعی که قبلا از جمعیت ایران گرفته شده است، تعداد فرزندان هر خانواده همراه با نسبت افرادی که دارای این تعداد فرزند هستند، در جدول زیر نشان داده شده است.

تعداد فرزندان:	صفر	یک	دو	سه	چهار	پنج و بیشتر
نسبت خانواددهایی که این تعداد فرزند را دارند.	۰/۰۵	۰/۱۰	۰/۳۵	۰/۲۵	۰/۱۵	۰/۱۰

اگر خانواده‌ای را به تصادف انتخاب کنیم، احتمال هر یک از پیشامدهای زیر را به دست آورید.

الف) کمتر از دو فرزند داشته باشند. (A)

$$P(A) = 0,05 + 0,10 = 0,15$$

ب) بین دو تا چهار فرزند داشته باشند. (B)

$$P(B) = 0,35 + 0,25 + 0,15 = 0,75$$

ج) چهار فرزند یا بیشتر داشته باشند. (C)

$$P(C) = 0,15 + 0,10 = 0,25$$



خواص مقدماتی احتمال

برای هر پیشامد، چه عضوهای فضای نمونه هم‌شانس باشند و چه نباشند، دو خاصیت زیر همواره برقرار هستند.

- 1) احتمال وقوع یک پیشامد مانند A در فضای نمونه S همواره عددی بزرگتر یا مساوی 0 و کوچکتر یا مساوی 1 می‌باشد.
- 2) $P(S) = 1$

$$S = \{a_1, a_2, a_3, a_4\}$$

- 2) احتمال وقوع فضای نمونه S برابر یک است.

مفروض است. کدام یک از توابع زیر فضای احتمال

- 1) $P(\{a_1\}) = \frac{1}{2}, P(\{a_2\}) = \frac{1}{3}, P(\{a_3\}) = \frac{1}{4}, P(\{a_4\}) = \frac{1}{5}$ غ
- 2) $P(\{a_1\}) = \frac{1}{2}, P(\{a_2\}) = \frac{1}{4}, P(\{a_3\}) = \frac{1}{6}, P(\{a_4\}) = \frac{1}{12}$ ص

مثال: فضای نمونه چهار عضوی S را تعریف نمی‌کند؟

- 3) $P(\{a_1\}) = \frac{4}{5}, P(\{a_2\}) = \frac{4}{5}, P(\{a_3\}) = -\frac{2}{5}, P(\{a_4\}) = -\frac{1}{5}$
- 4) $P(\{a_1\}) = \frac{3}{2}, P(\{a_2\}) = \frac{1}{2}, P(\{a_3\}) = \frac{1}{5}, P(\{a_4\}) = \frac{4}{5}$ غ



قواعد شمارش

اصل اساسی شمارش (اصل ضرب): هر گاه برای انجام یک عمل دو یا چند مرحله وجود داشته باشد به طوری که مرحله اول به n_1 طریق، مرحله دوم به n_2 طریق، ... و مرحله k ام به n_k طریق ممکن باشند، تعداد حالت‌های ممکن برای انجام این عمل $n_1 \times n_2 \times \dots \times n_k$ خواهد بود.

مثال: اگر بخواهیم به وسیله قطار، هواپیما یا اتوبوس به یکی از چهار شهر اصفهان، یزد، شیراز یا مشهد مسافرت کنیم، چند گزینه در پیش رو خواهیم داشت؟
 $n = 3 \times 4 = 12$

مثال: تعداد اعداد پنج رقمی که رقم یکان و صدگان آنها فرد و سایر رقم‌های آن زوج هستند، را به دست آورید.
 $n = 4 \times 5 \times 5 \times 5 \times 5 = 2500$

مثال: اگر در پلاک ماشینها از الگوی (ایران 13- 34 - م - 396) استفاده شده باشد، چند ماشین را می‌توان با این روش شماره‌گذاری کرد؟ (دو عدد دو رقمی، یک عدد سه رقمی و یک حرف فارسی)
 $n = (9 \times 10) \times (9 \times 10) \times 32 \times (9 \times 10 \times 10) = 233280000$



قواعد شمارش - جایگشت

تعریف فاکتوریل: حاصل ضرب اعداد طبیعی متوالی از 1 تا n را با نماد $n!$ نمایش می‌دهیم و آن را n فاکتوریل می‌خوانیم.

$$n! = 1 \times 2 \times 3 \times \cdots \times n$$

جایگشت (ترتیب): هرگاه ترتیب مراحل متوالی انجام یک عمل مهم باشد و آرایش‌های مختلف تنها برای گروهی از اشیا امکان‌پذیر باشد، ترتیب یا جایگشت داریم. مثلاً جایگشت‌های سه تایی از سه حرف a، b و c عبارتند از:

$$(n)_r = P_r^n = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}$$

برای محاسبه جایگشت r تایی از n شی متفاوت، از رابطه زیر استفاده می‌کنیم.

$$(15)_3 = P_3^{15} = \frac{15!}{(15-3)!} = \frac{15 \times 14 \times 13 \times 12!}{12!} = 15 \times 14 \times 13 = 2730$$

مثال: به چند طریق می‌توان از بین 15 نفر یک رییس، یک معاون و یک منشی انتخاب کرد؟



قواعد شمارش - جایگشت

جایگشت دوری: فرض کنیم n نفر دور یک میز نشسته‌اند و می‌خواهیم تعداد جایگشت ترتیب نشستن ایشان دور میز را محاسبه کنیم. با توجه به تقارن مساله، اگر همگی در جهت چرخش عقربه‌های ساعت (یا در خلاف آن) جابجا شوند، حالت جدیدی به وجود نخواهد آمد. در چنین حالتی تعداد جایگشت‌های $(n-1)$ دوری $\frac{n!}{n}$ شی متفاوت برابر خواهد بود با:

$$abc, acb, bca, bac, cba, cab$$

به عنوان مثال اگر جایگشت‌های سه تایی a, b, c را در نظر بگیریم (acb, bac, cba) که

دور یک دایره باشند، جایگشت‌های abc, bca, cab در صورتیکه دور یک دایره باشند، یکسان و همچنین جایگشت‌های دوری $2! = 2 = (3-1)! = \frac{3!}{3}$ نیز یکسان هستند. یعنی از هر سه حالت جایگشت ساده، یک حالت

جایگشت دوری به وجود می‌آید. بنابراین تعداد جایگشت‌های دوری سه حرف متمایز به صورت خواهد بود.



قواعد شمارش - جایگشت با عناصر تکراری

جایگشت با تکرار: فرض کنیم n شی داریم که در n_1 تین آنها شی n_1 یک نوع، شی n_2 یک نوع، ... شی n_k یک نوع دیگر، ... و شی از نوع دیگر باشند. در این صورت، اگر $n = n_1 + n_2 + \dots + n_k$ باشد، تعداد جایگشتها به صورت خواهد بود.

مثال: برای چراغانی سر در یک شرکت، 5 لامپ سبز، 4 لامپ آبی، 6 لامپ زرد و 9 لامپ سفید داریم. به چند طریق میتوانیم این لامپها را در ردیف قرار دهیم؟

$$\frac{(5+4+6+9)!}{5! \times 4! \times 6! \times 9!} = \frac{24 \times 23 \times 22 \times \dots \times 10 \times 9!}{5! \times 4! \times 6! \times 9!} = \frac{23 \times 22 \times \dots \times 12 \times 11 \times 10}{5! \times 6!} =$$

$$\frac{23 \times \overset{11}{\cancel{22}} \times \overset{7}{\cancel{21}} \times \overset{1}{\cancel{20}} \times 19 \times \overset{3}{\cancel{18}} \times 17 \times 16 \times 15 \times 14 \times 13 \times 11}{\cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times 1} = 11 \times 7 \times 17 \times 16 \times 15 \times 14 \times 13 \times 11 = 628,948,320$$



قواعد شمارش - ترکیب

ترکیب: هرگاه ترتیب مراحل متوالی انجام یک عمل مهم نباشد و بخواهیم از بین n شی متمایز r شی (بدون در نظر گرفتن ترتیب آنها) انتخاب کنیم، تعداد حالت‌های ممکن از رابطه زیر به دست می‌آید.

$$\binom{n}{r} = C_r^n = \frac{n!}{r!(n-r)!}$$

در واقع از هر $r!$ حالت (تعداد جایگشت r شی) فقط یک حالت متمایز خواهیم داشت.

مثال: به چند طریق می‌توان از بین 15 نفر یک تیم 3 نفره انتخاب کرد؟

$$\binom{15}{3} = C_3^{15} = \frac{15!}{3!(15-3)!} = \frac{\overset{5}{\cancel{15}} \times \overset{7}{\cancel{14}} \times \cancel{13} \times \cancel{12}!}{\cancel{2} \times \cancel{3} \times \cancel{12}!} = 5 \times 7 \times 13 = 455$$

مثال: به چند طریق می‌توان از بین 7 مدیر امور اداری و 5 مدیر امور مالی، یک تیم متشکل از 3 مدیر امور

اداری و 2 مدیر امور مالی انتخاب کرد؟

$$\binom{7}{3} \times \binom{5}{2} = C_3^7 \times C_2^5 = \frac{7!}{3!(7-3)!} \times \frac{5!}{2!(5-2)!} = \frac{\cancel{7} \times \cancel{6} \times \cancel{5} \times \cancel{4}!}{\cancel{3}! \times \cancel{4}!} \times \frac{\overset{2}{\cancel{5}} \times \cancel{4} \times \cancel{3}!}{\cancel{2}! \times \cancel{3}!} = 35 \times 10 = 350$$



قواعد شمارش - افرازهای مرتب

افرازهای مرتب: فرض کنیم n شی داریم و می‌خواهیم آنها را در k دسته افراز کنیم به طوری که n_1 در دسته

اول n_1 شی و در دسته دوم n_2 شی و ... و در دسته k ام n_k شی قرار گیرند و ترتیب قرار گرفتن اشیا در

دسته‌ها مهم نباشد. در این صورت تعداد افرازهای n شی به صورت $\frac{n!}{n_1! \times n_2! \times \dots \times n_k!}$ می‌باشد. $n = n_1 + n_2 + \dots + n_k$

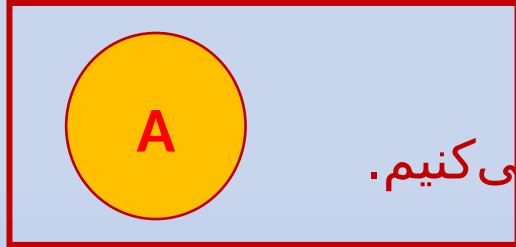
خواهد بود که در آن می‌باشد.

مثال: در یک پارکینگ سه طبقه که در طبقه اول 15 اتومبیل، در طبقه دوم 10 اتومبیل و در طبقه سوم 5 اتومبیل می‌توانند پارک کنند، به چند طریق می‌توان 30 اتومبیل را جای داد؟ $15 + 10 + 5 = 30 \Rightarrow \frac{30!}{15! \times 10! \times 5!}$

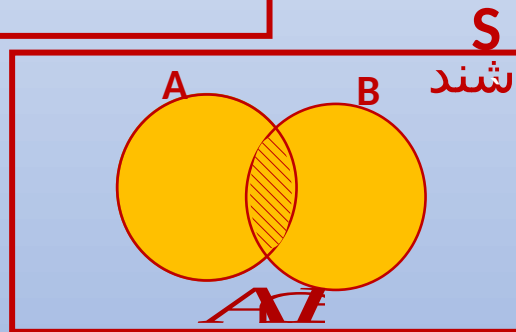


عملیات روی پیشامدها و قواعد احتمال

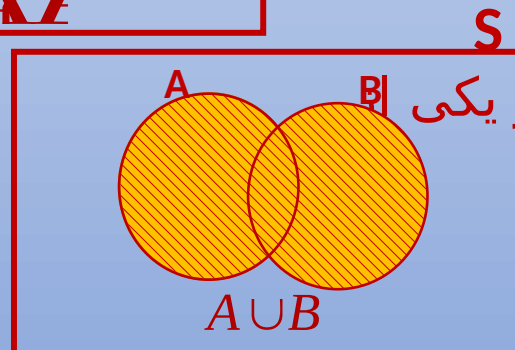
نمودار ون: برای نمایش مجموعه‌ها و روابط بین آنها به ابتکار ریاضی‌دان انگلیسی جان ون، به ازای هر مجموعه یک شکل هندسی از قبیل مستطیل، مربع، دایره، بیضی، مثلث یا امثال آن استفاده می‌کنیم. این نمودار درک بهتری از روابط بین مجموعه‌ها به ما می‌دهد.



به عنوان مثال: برای نمایش پیشامد A در فضای نمونه S از نمودار مقابل استفاده می‌کنیم.



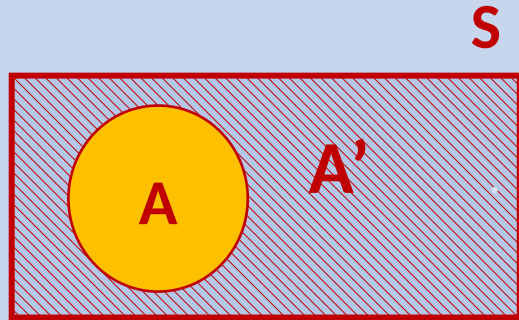
اشتراک دو پیشامد: پیشامدی که عضوهای آن در هر دو پیشامد A و B باشند را اشتراک دو پیشامد A و B گویند.



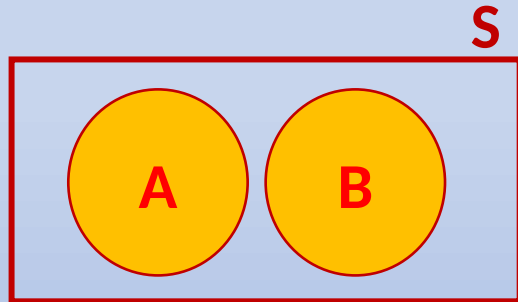
اجتماع دو پیشامد: پیشامدی که عضوهای آن در دست‌کم در یکی از دو پیشامد A و B باشند را اجتماع دو پیشامد A و B گویند.



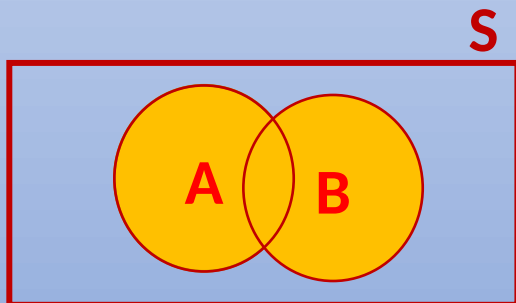
عملیات روی پیشامدها و قواعد احتمال



متمم یک پیشامد: پیشامدی که شامل همه عضوهای فضای نمونه به جز عضوهای پیشامد را متمم A گویند و آن را با نماد A' یا A^c نمایش می‌دهند.



پیشامدهای ناسازگار: دو پیشامد A و B را ناسازگار گویند اگر $A \cap B = \emptyset$ و تنها اگر



پیشامدهای سازگار: دو پیشامد A و B را سازگار گویند اگر $A \cap B \neq \emptyset$ و تنها اگر



عملیات روی پیشامدها و قواعد احتمال

(1) اگر پیشامد A زیرمجموعه پیشامد B باشد $A \subseteq B$ ، آنگاه $P(A) \leq P(B)$

(2) اگر A' متمم A باشد، آنگاه $P(A') = 1 - P(A)$

(3) با توجه به اصل شمول و عدم شمول، احتمال اجتماع دو پیشامد برابر است با: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(4) برای دو پیشامد ناسازگار $A \cap B = \emptyset$ ، آنگاه $P(A \cap B) = 0$ و در نتیجه $P(A \cup B) = P(A) + P(B)$



عملیات روی پیشامدها و قواعد احتمال

مثال 1: احتمال اینکه فردی موبایل، لپ‌تاپ یا هر دوی آنها را داشته باشد، به ترتیب $0/8$ ، $0/6$ و $0/45$ می‌باشد. احتمال پیشامدهای زیر را حساب کنید.

$$P(A) = 0/8, P(B) = 0/6, P(A \cap B) = 0/45$$

$$P(A') = 1 - P(A) = 1 - 0/8 = 0/2$$

الف) موبایل نداشته باشد.

ب) دست‌کم یکی از این دو را داشته باشد. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0/8 + 0/6 - 0/45 = 0/95$

مثال 2: از بین 20 کاندیدای انتخاب رییس یک انجمن خیریه، احتمال انتخاب آقای احمدی و خانم اکبری به ترتیب $0/06$ و $0/08$ می‌باشد. با چه احتمالی یکی از این دو نفر رییس انجمن خیریه خواهند شد؟

$$P(A) = 0/06, P(B) = 0/08, P(A \cap B) = 0$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0/06 + 0/08 - 0 = 0/14$$



عملیات روی پیشامدها و قواعد احتمال

مثال 3: چهار نفر داوطلب پست معاونت مالی سازمانی هستند. اگر احتمال انتخاب A دو برابر احتمال انتخاب B باشد و B و C هم شانس باشند و احتمال انتخاب C دو برابر D باشد، احتمالات زیر را حساب کنید.
 $P(A) + P(B) + P(C) + P(D) = 1$
 $P(A) = 2P(B)$, $P(B) = P(C)$, $P(C) = 2P(D)$

الف) احتمال انتخاب C

$$2P(B) + P(B) + P(B) + \frac{1}{2}P(B) = 1 \Rightarrow \frac{9}{2}P(B) = 1 \Rightarrow P(B) = \frac{2}{9} \Rightarrow P(C) = P(B) = \frac{2}{9}$$

ب) احتمال انتخاب A

$$P(A) = 2P(B) = 2 \times \frac{2}{9} = \frac{4}{9}$$

ج) احتمال انتخاب یکی از A، B یا C

راه اول:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

$$P(A \cup B \cup C) = \frac{4}{9} + \frac{2}{9} + \frac{2}{9} - 0 - 0 - 0 + 0 = \frac{8}{9}$$

راه دوم: اگر D انتخاب نشود یعنی یکی از A، B یا C انتخاب شده اند.

$$P(C) = 2P(D) \Rightarrow P(D) = \frac{1}{2} \times P(C) = \frac{1}{2} \times \frac{2}{9} = \frac{1}{9}$$

$$P(D') = 1 - P(D) = 1 - \frac{1}{9} = \frac{8}{9}$$



عملیات روی پیشامدها و قواعد احتمال

مثال 4: 35 درصد دانشجویان دانشگاهی در یک استان، خارج از خوابگاه زندگی می‌کنند، 55 درصد دانشجویان اهل همان استان هستند و 75 درصد دانشجویان یا اهل آن استان نیستند یا در خارج خوابگاه دانشجویی زندگی می‌کنند. احتمال اینکه یک دانشجو که به تصادف انتخاب می‌شود، اهل آن استان نباشد و در خارج از خوابگاه زندگی کند، چقدر است؟

$$P(A) = 0.35, P(B) = 0.55, P(A \cup B') = 0.75$$

$$P(A \cap B') = P(A) + P(B') - P(A \cup B') = 0.35 + (1 - 0.55) - 0.75 = 0.05$$

مثال 5: برای بررسی قابلیت اطمینان دو پمپ بنزین که در معرض دو نقص نشئت بنزین و نقص فنی پمپ قرار دارند، داده‌های زیر استخراج شده‌اند. هنگامی که یکی از این دو نقص یا هر دو رخ دهد، پمپ بنزین تعطیل

$$P(A_1 \cup B_1) = P(A_1) + P(B_1) - P(A_1 \cap B_1) = 0.17 - 0 = 0.17 \text{؟ ارد؟}$$

$$P(A_2 \cup B_2) = P(A_2) + P(B_2) - P(A_2 \cap B_2) = 0.21 - 0.06 = 0.15$$

$$P(A_1 \cup B_1) > P(A_2 \cup B_2) \Rightarrow \text{Reliability}_1 < \text{Reliability}_2$$

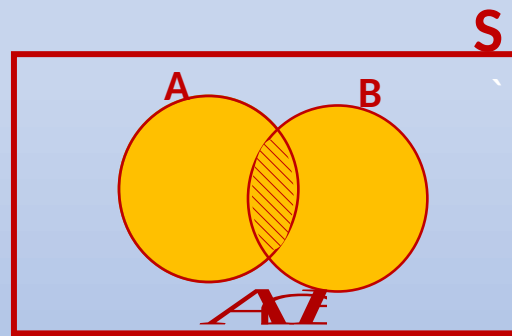
پمپ بنزین	احتمال نقص فنی پمپ	احتمال نشئت بنزین	احتمال هر دو
۱	۰/۰۷	۰/۱۰	۰
۲	۰/۰۹	۰/۱۲	۰/۰۶



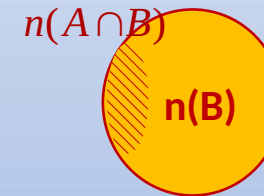
احتمال شرطی

دو پیشامد A و B را در فضای نمونه S در نظر بگیرید. می‌خواهیم احتمال پیشامد A را به شرطی که پیشامد B به وقوع پیوسته باشد را محاسبه کنیم. در چنین حالتی می‌توانیم فضای نمونه را محدود به اعضای پیشامد B و پیشامد A را محدود به عضوهای مشترک A و B در نظر بگیریم. اگر احتمال شرطی وقوع A به شرط B را با

$$P(A/B) = \frac{n(A \cap B)}{n(B)}$$



نماد $P(A/B)$ نمایش دهیم، خواهیم داشت:



$$P(A/B) = \frac{\frac{n(A \cap B)}{n(S)}}{\frac{n(B)}{n(S)}} = \frac{P(A \cap B)}{P(B)}$$

با تقسیم صورت و مخرج کسر فوق به $n(S)$ خواهیم داشت:



احتمال شرطی

مثال: جدول مقابل سود 150 شرکت که در 4 صنعت فعالیت

نشان می‌دهد. با توجه به داده‌های جدول احتمالات زیر را احسا

(1) سود شرکت انتخابی، کمتر یا مساوی 5 میلیارد ریال باشد.

(2) شرکت انتخابی در صنعت آلومینیم فعالیت کند.

(3) شرکت انتخابی در صنایع نساجی فعالیت کند و سودش بیش از 5 میلیارد ریال باشد.

(4) شرکت انتخابی یا در صنعت نساجی باشد یا سودش بیش از 5 میلیارد ریال باشد.

(5) شرکت انتخابی در صنعت نساجی باشد به شرطی که سودش بیش از 5 میلیارد ریال باشد.

(6) شرکت انتخابی سودی کمتر یا مساوی 5 میلیارد ریال داشته باشد به شرطی که در صنعت نساجی فعال

نباشد.

صنعت	میزان سود	کمتر یا مساوی ۵ میلیارد ریال (E)	بیشتر از ۵ میلیارد ریال (F)	مجموع
صنعت نساجی (A)	۱۷	۱۵	۳۲	
صنعت آلومینیوم (B)	۳۵	۳۰	۶۵	
صنعت مواد غذایی (C)	۲۸	۵	۳۳	
صنعت چوب و کاغذ (D)	۱۰	۱۰	۲۰	
مجموع	۹۰	۶۰	۱۵۰	



احتمال شرطی

پاسخ مثال:

$$1) P(E) = \frac{90}{150} = 0,6$$

$$2) P(B) = \frac{65}{150} = 0,43$$

$$3) P(A \cap F) = \frac{15}{150} = 0,1$$

$$4) P(A \cup F) = P(A) + P(F) - P(A \cap F) = \frac{32}{150} + \frac{60}{150} - \frac{15}{150} = 0,51$$

$$5) P(A / F) = \frac{P(A \cap F)}{P(F)} = \frac{\frac{15}{150}}{\frac{60}{150}} = \frac{15}{60} = 0,25$$

$$6) P(E / A') = \frac{P(E \cap A')}{P(A')} = \frac{\frac{35 + 28 + 10}{150}}{\frac{65 + 33 + 20}{150}} = 0,62$$

صنعت	میزان سود	کمتر یا مساوی ۵ میلیارد ریال (E)	بیشتر از ۵ میلیارد ریال (F)	مجموع
صنعت نساجی (A)		۱۷	۱۵	۳۲
صنعت آلومینیوم (B)		۳۵	۳۰	۶۵
صنعت مواد غذایی (C)		۲۸	۵	۳۳
صنعت چوب و کاغذ (D)		۱۰	۱۰	۲۰
مجموع		۹۰	۶۰	۱۵۰



احتمال شرطی

مثال: در خانواده‌ای که دو فرزند دارند، با فرض اینکه احتمال داشتن فرزند پسر (B) و دختر (G) برابر است، احتمالات زیر را حساب کنید.

$$S = \{GG, GB, BG, BB\}$$

(1) احتمال اینکه هر دو فرزند خانواده دختر باشند.

(2) احتمال اینکه هر دو فرزند دختر باشند، با شرط اینکه بدانیم حداقل یکی از فرزندان دختر است.

پاسخ مثال:

(1) پیشامد داشتن دو دختر را با D نمایش می‌دهیم. در این صورت $D = \{GG\}$

$$P(D) = \frac{n(D)}{n(S)} = \frac{1}{4}$$

و در نتیجه:

(2) پیشامد داشتن حداقل یک دختر را با E نمایش می‌دهیم. در این صورت $E = \{GG, GB, BG\}$

$$P(D/E) = \frac{P(D \cap E)}{P(E)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

و در نتیجه:



احتمال شرطی - قانون ضرب احتمالات

همانطور که اشاره شد، احتمال شرطی وقوع A به شرط B را با نماد $P(A/B)$ نمایش می‌دهیم. داریم $P(A/B) = \frac{P(A \cap B)}{P(B)}$

این رابطه را می‌توان به صورت $P(A \cap B) = P(B) \times P(A/B)$ نیز نوشت که به آن قانون ضرب

احتمالات گفته می‌شود. $P(A \cap B) = P(A) \times P(B/A)$

به همین ترتیب برای $P(B/A)$ نیز داریم: اشتراک دو پیشامد، استفاده می‌کنیم.

مثال: از کیسه‌ای که حاوی 5 مهره قرمز و 7 مهره سبز است، دو مهره بدون جایگزینی بیرون می‌آوریم.

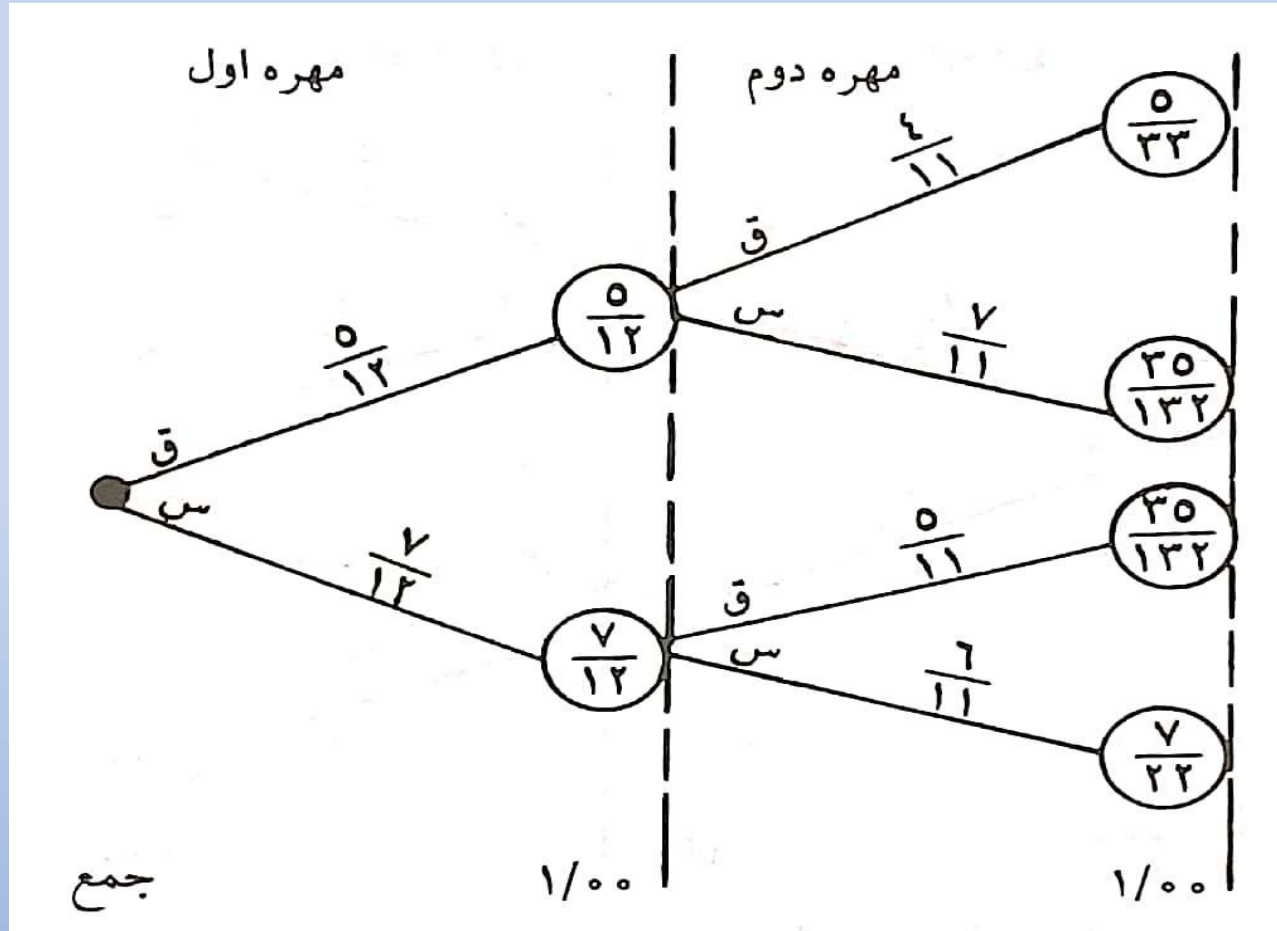
احتمال اینکه $\frac{4}{33} \times \frac{5}{12}$ دو مهره قرمز باشند، چقدر است؟

$P(\text{قرمز بودن مهره اول} / \text{قرمز بودن مهره دوم}) = P(\text{قرمز بودن مهره اول}) = P(\text{قرمز بودن هر دو مهره})$



احتمال شرطی - قانون ضرب احتمالات

تکنیک دیگر برای حل مثال قبل،
استفاده از نمودار درختی احتمالات است.



قانون ضرب احتمالات را می‌توان برای بیش از دو پیشامد نیز به کار برد

$$P(A \cap B \cap C) = P(A) \times P(B/A) \times P(C/A \cap B)$$



احتمال شرطی - پیشامدهای مستقل

هرگاه وقوع یا عدم وقوع پیشامد A هیچ اثری بر وقوع یا عدم وقوع پیشامد B نداشته باشد، دو پیشامد را مستقل گویند.

$$P(B / A) = P(B) \quad P(A / B) = P(A)$$

با توجه به قانون ضرب احتمالات

$$P(A \cap B) = P(A) \times P(B / A) \quad , \quad P(B / A) = P(B) \Rightarrow \boxed{P(A \cap B) = P(A) \times P(B)}$$

بدیهی است در چنین شرایطی خواهیم داشت:

رشته	زبان انگلیسی	زبان روسی	زبان آلمانی	مجموع
دانشجوی سال				
اول	۲۰۰	۱۸۵	۴۰	۴۲۵
دوم	۱۸۰	۱۲۰	۳۰	۳۳۰
سوم	۱۲۰	۸۵	۳۵	۲۴۰
چهارم	۱۰۵	۹۰	۲۰	۲۱۵
مجموع	۶۰۵	۴۸۰	۱۲۵	۱۲۱۰

مثال: جدول مقابل مربوط به دانشکده زبانهای خارجی می باشد.

آیا دانشجوی زبان انگلیسی بودن، مستقل از دانشجوی سال اول بودن است

پاسخ: دانشجوی زبان انگلیسی بودن را با A و دانشجوی سال اول بودن را

نمایش می دهیم.

$$P(A) = \frac{605}{1210} = 0,5 \quad , \quad P(B) = \frac{425}{1210} = 0,35124 \quad , \quad P(A \cap B) = \frac{200}{1210} = 0,16529$$

$$P(A) \times P(B) = 0,17562 \neq P(A \cap B)$$

دو پیشامد مستقل نیستند.



قضیه بیز (احتمالات پیشین و پسین)

این قضیه کاربرد زیادی در تصمیم‌گیری دارد. فرض کنید با اطلاعات اولیه، احتمال یک پیشامد را محاسبه کرده‌اید و سپس اطلاعات جدیدی به دست شما می‌رسد که مقدار محاسبه شده احتمال پیشامد را تغییر می‌دهد. به مقدار اولیه احتمال که با اطلاعات قبلی بدست آمده است، احتمال پیشین و به مقدار ثانویه که با اطلاعات جدید بدست می‌آید، احتمال پسین گفته می‌شود.

$$P(A \cap B) = P(B) \times P(A / B) = P(A) \times P(B / A) \Rightarrow$$

با توجه به قانون ضرب احتمالات داریم:

$$P(A / B) = \frac{P(A) \times P(B / A)}{P(B)}$$

این فرم ساده قضیه بیز می‌باشد.



قضیه بیز (احتمالات پیشین و پسین)

مثال: 25 درصد ماشین‌های یک شهر بیش از حد دودزا هستند. در معاینه فنی، 98 درصد ماشین‌هایی که بیش از حد دودزا هستند، بیش از حد دودزا و 7 درصد از ماشین‌هایی که دودزا نیستند، بیش از حد دودزا تشخیص داده می‌شوند. با چه احتمالی ماشینی که در معاینه فنی دودزا تشخیص داده شده است، واقعا بیش از حد دودزا است؟

پاسخ: فرض کنیم A پیشامد تشخیص بیش از حد دودزا بودن و $P(A/B) = 0.07$ باشد، $P(A/B') = 0.98$ از دودزا بودن، $P(B) = 0.25$ و $P(B') = 0.75$ است.

$$P(A) = P(A/B) \times P(B) + P(A/B') \times P(B') = 0.07 \times 0.25 + 0.98 \times 0.75 = 0.7425$$

$$P(B/A) = \frac{P(B) \times P(A/B)}{P(A)} = \frac{0.25 \times 0.07}{0.7425} = 0.023$$



فصل ششم

توابع احتمال گسسته

Discrete Probability Functions



متغیر تصادفی گسسته، تابع احتمال و تابع توزیع

متغیر تصادفی گسسته: متغیر تصادفی گسسته تابعی است که دامنه آن فضای نمونه و برد آن زیرمجموعه‌ای از اعداد حقیقی است.

مثال 1: فرض کنید اجرای یک گروه نمایش در روزهای آفتابی 10 میلیون تومان، در روزهای ابری صفر و در روزهای برفی و بارانی 5- میلیون تومان سود دارد. متغیر تصادفی میزان سود این گروه نمایش بر حسب شرایط جوی به صورت زیر است.

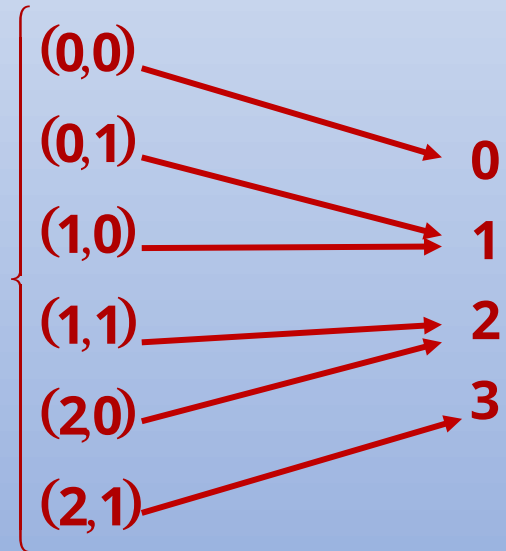
آفتابی	→	10,000,000
ابری	→	0
بارانی	↘	-5,000,000
برفی	→	
فضای نمونه		میزان سود



متغیر تصادفی گسسته، تابع احتمال و تابع توزیع

مثال 2: فروش روزانه دو بنگاه اتومبیل را در نظر می‌گیریم. تعداد اتومبیل‌های فروخته شده در بنگاه اول را x (حداکثر دو تا) و تعداد اتومبیل‌های فروخته شده در بنگاه دوم را y (حداکثر یکی) نشان می‌دهیم. فضای نمونه‌ای و متغیر تصادفی تعداد اتومبیل فروخته شده در این دو بنگاه، به شرح زیر است.

$$S = \{(0,0), (0,1), (1,0), (1,1), (2,0), (2,1)\}$$

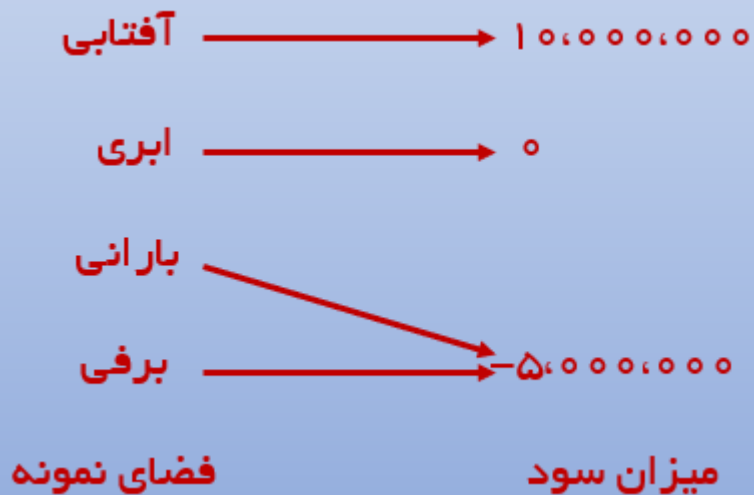




متغیر تصادفی گسسته، تابع احتمال و تابع توزیع

تابع احتمال: به تابعی که احتمال هر یک از مقادیر متغیر تصادفی را مشخص می‌کند، تابع احتمال یا توزیع احتمال گفته می‌شود.

تابع احتمال در مثال 1: اگر احتمال آفتابی، ابری، بارانی و برفی بودن هوا در مثال 1 به ترتیب $0/40$ ، $0/30$ ، $0/20$ و $0/10$ باشد، تابع احتمال میزان سود به صورت زیر خواهد بود.



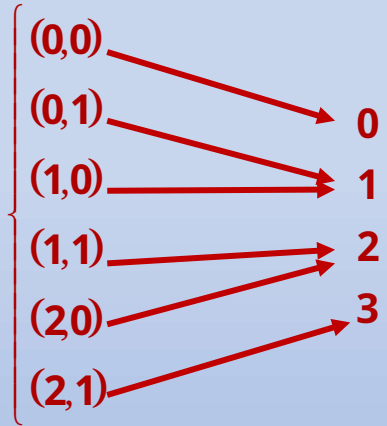
x (میزان سود به ریال)	10,000,000	0	-5,000,000	جمع
$P(X=x)$ (احتمال)	0/40	0/30	0/30	1



متغیر تصادفی گسسته، تابع احتمال و تابع توزیع

تابع احتمال در مثال 2: با توجه به احتمالات فضای نمونه فروش اتومبیل در دو بنگاه که در نمودار زیر آمده

است، تابع احتمال مربوط به تعداد اتومبیل فروخته شده در یک روز در دو بنگاه مورد نظر، $S = \{(0,0), (0,1), (1,0), (1,1), (2,0), (2,1)\}$



z	0	1	2	3	جمع
$P(Z=z)$	0.0625	0.3750	0.3750	0.0625	1



متغیر تصادفی گسسته، تابع احتمال و تابع توزیع

تابع توزیع (تابع احتمال تجمعی): به تابعی که احتمال تجمعی هر یک از مقادیر متغیر تصادفی را مشخص می‌کند، تابع توزیع یا تابع احتمال تجمعی گفته می‌شود. ~~تابع توزیع~~ را به صورت نمایش

می‌دهیم. $F(3)$ $F(1)$

مثال: با توجه به جدول تابع احتمال داده شده، جدول تابع توزیع را بنویسید و مقادیر x و $P(X=x)$ را محاسبه کنید.

x	- 2	0	3	5	10
$P(X=x)$	0/10	0/15	0/25	0/30	0/20

$$F(1) = P(X \leq 1) = 0.25, \quad F(3) = P(X \leq 3) = 0.50$$



متغیر تصادفی گسسته، تابع احتمال و تابع توزیع

تعداد روزها	تعداد غایبین
۲	۱۸
۳	۲۷
۴	۳۰
۵	۱۵
۶	۱۸
۷	۱۲
جمع	۱۲۰

مثال: تعداد غیبت کارکنان یک سازمان در 120 روز مطابق جدول مقابل داده شده

الف) تابع احتمال را بنویسید.

ب) تابع توزیع را بنویسید.

ج) احتمال اینکه در یک روز کمتر یا مساوی 5 نفر در سازمان غایب باشند

را به دست آورید.

x	2	3	4	5	6	7
$P(X=x)$	0,150	0,225	0,250	0,125	0,150	0,100

x	2	3	4	5	6	7
$F(x) = P(X \leq x)$	0,150	0,375	0,625	0,750	0,900	1

$$F(5) = P(X \leq 5) = 0,750$$



امید ریاضی و واریانس متغیر تصادفی

امید ریاضی: در بحث توابع احتمال، امید ریاضی همان میانگین وزن دار است که احتمالات به عنوان وزن در

نظر گرفته می شوند. امید ریاضی متغیر تصادفی X با نماد $E(X)$ نمایش داده می شود و به صورت

قابل محاسبه است که در آن $f(x) = P(X=x)$ می باشد.

تعداد روزها	تعداد غایبین
۱۸	۲
۲۷	۳
۳۰	۴
۱۵	۵
۱۸	۶
۱۲	۷
۱۲۰	جمع

مثال: میانگین تعداد غیبت کارکنان در یک روز در سازمانی که گزارش تعداد غایبین آر

در 120 روز در جدول مقابل آمده را بدست آورید.

x	2	3	4	5	6	7	
$f(x) = P(X=x)$	0,150	0,225	0,250	0,125	0,150	0,100	1,000
$x f(x)$	0,300	0,675	1,000	0,625	0,900	0,700	4,200



امید ریاضی و واریانس متغیر تصادفی

خواص امید ریاضی: اگر a و b مقادیر ثابتی باشند، با توجه به تعریف امید ریاضی خواهیم داشت:

$$1) E(b) = b$$

$$2) E(aX) = aE(X)$$

$$3) E(aX + b) = aE(X) + b$$

مفروض است. حاصل هر یک از مقادیر

x	- 2	1	3	5
$f(x)$	0,2	0,4	0,3	0,1

مثال: تابع احتمال

زیر $E(X)$ حساب $E(2X)$ $E(-3X + 2)$ $E(X^2)$ $E((X - 5)^2)$

$$1) E(X) = \sum_x xf(x) = 1,4 \quad 2) E(2X) = 2E(X) = 2,8$$

$$3) E(-3X + 2) = -3E(X) + 2 = -3 \times 1,4 + 2 = -2,2$$

$$4) E(X^2) = \sum_x x^2 f(x) = 6,4$$

$$5) E((X - 5)^2) = E(X^2 - 10X + 25) = E(X^2) - 10E(X) + 25 = 17,4$$

x	- 2	1	3	5	sum
$f(x)$	0,2	0,4	0,3	0,1	1
$xf(x)$	- 0,4	0,4	0,9	0,5	1,4
$x^2 f(x)$	0,8	0,4	2,7	2,5	6,4



امید ریاضی و واریانس متغیر تصادفی

واریانس: واریانس متغیر تصادفی گسسته از یکی از روابط زیر قابل محاسبه می باشد.

- 1) $V(X) = \sum_x (x - \mu)^2 f(x)$, $(\mu = E(X))$

- 2) $V(X) = E(X^2) - \mu^2$

انحراف معیار: با توجه به تعریف واریانس، انحراف معیار متغیر گسسته از رابطه $SD = \sqrt{V(X)}$ به دست می آید.

- 1) $V(a) = 0$

خواص واریانس: اگر a و b مقادیر ثابتی باشند، با توجه به تعریف واریانس خواهیم داشت:

- 2) $V(aX) = a^2 V(X)$

- 3) $V(aX + b) = a^2 V(X)$



امید ریاضی و واریانس متغیر تصادفی

مفروض است. حاصل هر یک از مقادیر

$$V(X) = E(X^2) - (E(X))^2 = 6,4 - (1,4)^2 = 4,44$$

$$1) V(3) = 0$$

$$2) V(-3X) = 3^2 V(X) = 9 \times 4,44 = 39,96$$

$$3) V(-3X + 2) = 3^2 V(X) = 9 \times 4,44 = 39,96$$

x	- 2	1	3	5
f(x)	0,2	0,4	0,3	0,1

مثال: تابع احتمال

زیر (3) حساب (3) $V(-3X + 2)$

x	- 2	1	3	5	sum
f(x)	0,2	0,4	0,3	0,1	1
xf(x)	- 0,4	0,4	0,9	0,5	1,4
x ² f(x)	0,8	0,4	2,7	2,5	6,4



تابع احتمال توأم

برای بررسی رفتار دو یا چند متغیر تصادفی به صورت همزمان از تابع احتمال توأم استفاده می‌کنیم. به عنوان مثال برای دو متغیر تصادفی X و Y که به ترتیب دارای $x_1, x_2, \dots, x_N$ و $y_1, y_2, \dots, y_K$ می‌باشند و می‌باشند، احتمال توأم شامل احتمال وقوع هر یک از زوج‌های مرتب (x_j, y_j) می‌باشند. $(j=1, 2, 3, \dots, N, j=1, 2, 3, \dots, K)$

$x \backslash y$	y_1	y_2	$\dots$	y_K
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\dots$	$f(x_1, y_K)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\dots$	$f(x_2, y_K)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_N	$f(x_N, y_1)$	$f(x_N, y_2)$	$\dots$	$f(x_N, y_K)$

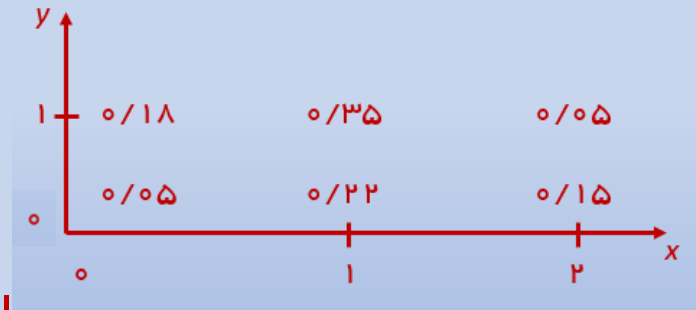


تابع احتمال توأم

مثال 1: در دو بنگاه اتومبیل، متغیر تصادفی X تعداد فروش در بنگاه 1 و Y تعداد فروش در بنگاه 2 می‌باشد. نمودار زیر نشان‌دهنده احتمالات توأم تعداد فروش در دو بنگاه می‌باشد. با رسم جدول تابع احتمال توأم فروش

در دو بنگاه، هر یک از احتمالات

$y \backslash x$	0	1
0	0,05	0,18
1	0,22	0,35
2	0,15	0,05



الف) $P(X=1)$ ، ب) $P(X>Y)$ ، ج) $P(Y \geq 0, X=1)$ ، د) تابع توزیع احتمال $Z=X+Y$

$$P(X=1) = P(X=1, Y=0) + P(X=1, Y=1) = 0,22 + 0,35 = 0,57 \quad \text{الف)}$$

$$P(X > Y) = P(X=1, Y=0) + P(X=2, Y=0) + P(X=2, Y=1) = 0,22 + 0,15 + 0,05 = 0,42 \quad \text{ب)}$$

$$P(X=2, Y \geq 0) = P(X=2, Y=0) + P(X=2, Y=1) = 0,15 + 0,05 = 0,20 \quad \text{ج)}$$

$$P(X=2, Y \geq 0) = P(X=2, Y=0) + P(X=2, Y=1) = 0,15 + 0,05 = 0,20 \quad \text{د)}$$

z	0	1	2	3
$f(z) = P(Z=z)$	0,05	0,40	0,50	0,05
$F(z) = P(Z \leq z)$	0,05	0,45	0,95	1



تابع احتمال توأم

مثال 2: در مثال 1، تابع احتمال هر یک از متغیرهای تصادفی X (تعداد فروش در بنگاه 1) و Y (تعداد فروش در بنگاه 2) را به دست آورید.

$y \backslash x$	0	1	احتمالات حاشیه‌ای X
0	0,05	0,18	0,23
1	0,22	0,35	0,57
2	0,15	0,05	0,20
احتمالات حاشیه‌ای Y	0,42	0,58	1

بنابراین خواهیم داشت:

x	0	1	2
$f(x) = P(X = x)$	0,23	0,57	0,20

y	0	1
$f(y) = P(Y = y)$	0,42	0,58



تابع احتمال توأم

مثال 3: در جعبه‌ای 8 بلبرینگ وجود دارد که 2 تای آنها معیوب است. به طور تصادفی 3 تا از این بلبرینگها را بیرون آورده‌ایم. اگر X تعداد بلبرینگ‌های سالم و Y تعداد بلبرینگ‌های معیوب باشد، جدول تابع احتمال توأم دو متغیر را نمایش دهید.

پاسخ: با توجه به اینکه X تنها می‌تواند 1، 2، 3 و Y تنها می‌تواند مقادیر 0، 1، 2 داشته باشد، 9 حالت به وجود می‌آید. اما تنها در سه حالت از این 9 حالت که مجموع دو متغیر برابر 3 است (چون ما 3 بلبرینگ انتخاب کرده‌ایم $X+Y=3$) مقدار احتمال توأم صفر نیست و در بقیه حالتها احتمال توأم برابر صفر است. $f(X=1, Y=2) = \frac{\binom{6}{1} \times \binom{2}{2}}{\binom{8}{3}} = \frac{6}{56}$

$$f(X=2, Y=1) = \frac{\binom{6}{2} \times \binom{2}{1}}{\binom{8}{3}} = \frac{30}{56}$$

$$f(X=3, Y=0) = \frac{\binom{6}{3} \times \binom{2}{0}}{\binom{8}{3}} = \frac{20}{56}$$

$X \backslash Y$	0	1	2
1	0	0	$\frac{6}{56}$
2	0	$\frac{30}{56}$	0
3	$\frac{20}{56}$	0	0



کوواریانس و استقلال دو متغیر تصادفی

کوواریانس: امید ریاضی تغییرات دو متغیر نسبت به میانگینشان را کوواریانس می‌نامیم. این معیار شدت رابطه خطی بین دو متغیر را نشان می‌دهد. در صورتی که:

1- با افزایش یک متغیر، دیگری نیز افزایش بیابد و با کاهش آن دیگری نیز کاهش یابد، بین دو متغیر رابطه مستقیم وجود خواهد داشت و کوواریانس آنها مثبت می‌باشد.

2- با افزایش یک متغیر، دیگری کاهش بیابد و با کاهش آن دیگری افزایش یابد، بین دو متغیر رابطه معکوس وجود خواهد داشت و کوواریانس آنها منفی می‌باشد.

3- افزایش یا کاهش یک متغیر هیچ تاثیری در دیگری نداشته باشد، دو متغیر مستقل خواهند بود و کوواریانس آنها صفر می‌باشد

$$Cov(X, Y) = E(X - \mu_x)(Y - \mu_y) = E(XY) - E(X)E(Y)$$

(برای متغیر تصادفی گسسته): $E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j)$, $E(X) = \mu_x$, $E(Y) = \mu_y$



کوواریانس و استقلال دو متغیر تصادفی

مثال 1: با توجه به تابع احتمال توام دو متغیر تصادفی X و Y ، کوواریانس آنها را به دست آورید.

$x \backslash y$	0	1	2		x	- 5	10	Σ		y	0	1	2	Σ
- 5	0,30	0,02	0,08	$\Rightarrow$	$f(x)$	0,40	0,60	1		$f(y)$	0,40	0,02	0,58	1
10	0,10	0	0,50		$xf(x)$	- 2	6,0	4		$yf(y)$	0	0,02	1,16	1,18

$$E(X) = 4, \quad E(Y) = 1,18, \quad E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j) = (-5) \times 0 \times 0,30 +$$

$$(-5) \times 1 \times 0,02 + (-5) \times 2 \times 0,08 + 10 \times 0 \times 0,10 + 10 \times 1 \times 0 + 10 \times 2 \times 0,50 = 9,1$$

$$\Rightarrow \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 9,1 - 4 \times 1,18 = 4,38$$



کوواریانس و استقلال دو متغیر تصادفی

استقلال دو متغیر تصادفی: همانطور که می‌دانیم دو پیشامد A و B مستقل هستند اگر و فقط اگر $P(A \cap B) = P(A) \times P(B)$

به طور مشابه دو متغیر تصادفی X و Y در صورتی مستقل هستند که داشته باشیم: $\forall i, j: f(x_i, y_j) = f(x_i) \times f(y_j)$

مثال 2: با توجه به جدول تابع احتمال توأم مقابل، آیا دو متغیر تصادفی X و Y مستقل هستند؟

$x \backslash y$	2	4	6
0	0,20	0,08	0,12
1	0,30	0,12	0,18

شرط استقلال را برای همه زوج مرتب‌های (x_i, y_j) بررسی می‌کنیم.

$x \backslash y$	2	4	6	$f(x)$
0	0,20 = 0,4 × 0,5	0,08 = 0,4 × 0,2	0,12 = 0,4 × 0,3	0,4
1	0,30 = 0,6 × 0,5	0,12 = 0,6 × 0,3	0,18 = 0,6 × 0,3	0,6
$f(y)$	0,5	0,2	0,3	1

⇒ دو متغیر مستقل هستند



کوواریانس و استقلال دو متغیر تصادفی

کوواریانس دو متغیر تصادفی مستقل: برای دو متغیر تصادفی مستقل داریم

$$E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j) = \sum_i \sum_j x_i y_j f(x_i) f(y_j) = \sum_i \sum_j x_i f(x_i) y_j f(y_j) \quad \text{زیرا}$$

$$\sum_i x_i f(x_i) \sum_j y_j f(y_j) = E(X)E(Y)$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

و بنابراین کوواریانس دو متغیر برابر صفر خواهد بود.

نکته مهم اینکه عکس این مطلب درست نیست. یعنی ممکن است کوواریانس دو متغیر صفر باشد ولی

$$\text{Cov}(X, Y) = 0 \quad \leftarrow \text{---} \text{---} \text{---} \right.$$

X و Y مستقل
هستند.

مستقل نباشند.



کوواریانس و استقلال دو متغیر تصادفی

مثال 3: کوواریانس متغیرهای مثال 2 را به دست آورید.

$x \backslash y$	2	4	6
0	0,20	0,08	0,12
1	0,30	0,12	0,18

$\Rightarrow$

x	0	1	Σ
$f(x)$	0,40	0,60	1
$xf(x)$	0	0,60	0,6

y	2	4	6	Σ
$f(y)$	0,5	0,2	0,3	1
$yf(y)$	1	0,8	1,8	3,6

$$E(X) = 0,6, \quad E(Y) = 3,6, \quad E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j) = 0 \times 2 \times 0,20 +$$

$$0 \times 4 \times 0,08 + 0 \times 6 \times 0,12 + 1 \times 2 \times 0,30 + 1 \times 4 \times 0,12 + 1 \times 6 \times 0,18 = 2,16$$

$$\Rightarrow \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 2,16 - 0,6 \times 3,6 = 2,16 - 2,16 = 0$$



کوواریانس و استقلال دو متغیر تصادفی

مثال 4: با توجه به جدول احتمال توأم داده شده، نشان دهید با اینکه دو متغیر X و Y مستقل نیستند، کوواریانس آنها صفر است.

$y \backslash x$	-1	0	1
-1	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$
1	$\frac{1}{6}$	0	$\frac{1}{6}$

$$\frac{1}{6} \neq \left(\frac{1}{6} + \frac{1}{6}\right) \times \left(\frac{1}{6} + \frac{1}{3} + \frac{1}{6}\right) = \frac{1}{3} \times \frac{2}{3}$$

x	-1	1	Σ
$f(x)$	$\frac{2}{3}$	$\frac{1}{3}$	1
$xf(x)$	$-\frac{2}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$

همانطور که ملاحظه می شود دو متغیر مستقل نیستند.

y	-1	0	1	Σ
$f(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1
$yf(y)$	$-\frac{1}{3}$	0	$\frac{1}{3}$	0

$$E(X) = -\frac{1}{3}, \quad E(Y) = 0, \quad E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j) = (-1)(-1)\left(\frac{1}{6}\right) + (-1)(0)\left(\frac{1}{3}\right) + (-1)(1)\left(\frac{1}{6}\right) +$$

$$(1)(-1)\left(\frac{1}{6}\right) + (1)(0)(0) + (1)(1)\left(\frac{1}{6}\right) = 0 \Rightarrow \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0 - \left(-\frac{1}{3}\right) \times 0 = 0$$



کوواریانس و استقلال دو متغیر تصادفی

امید ریاضی و واریانس مجموع دو متغیر تصادفی:

$$E(X + Y) = E(X) + E(Y)$$

(1) حالت کلی:

$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$$

$$V(X - Y) = V(X) + V(Y) - 2Cov(X, Y)$$

(2) اگر متغیرها مستقل باشند:

$$E(X + Y) = E(X) + E(Y)$$

$$V(X + Y) = V(X) + V(Y)$$

$$V(X - Y) = V(X) + V(Y)$$



توزیع برنولی

آزمایش برنولی: آزمایشی که دو پیامد داشته باشد که معمولا با موفقیت و شکست از آن یاد می‌شود را آزمایش برنولی گویند. به عنوان مثال احتمال پیروزی یک کاندیدا در انتخابات (p) در مقابل احتمال شکست وی ($1-p$) یا احتمال اینکه در پرتاب یک تاس اعداد 6 ظاهر شود (p) در مقابل اینکه اعداد 1 تا 5 ظاهر نگردد ($1-p$) یا احتمال اینکه یک کالا سالم باشد در مقابل اینکه معیوب باشد، نمونه‌هایی از آزمایش برنولی می‌باشند.

توزیع برنولی: در یک آزمایش برنولی توزیع تعداد موفقیت‌ها را توزیع برنولی می‌نامند. احتمال موفقیت در یک آزمایش برنولی را معمولا با p و احتمال شکست را با q نمایش می‌دهند. از آنجاییکه این دو احتمال مکمل هم هستند، داریم:

$$P + q = 1$$

مثال 1: بر اساس آمار متولدین، مشخص شده است که از هر 100 نوزاد به طور متوسط 48 نفر دختر و بقیه پسر هستند. آیا این یک آزمایش برنولی محسوب می‌شود؟

پاسخ: بله؛ اگر احتمال دختر بودن نوزاد را احتمال موفقیت بنامیم، آنگاه برای هر نوزاد احتمال موفقیت $0/48$ و احتمال شکست $0/52$ می‌باشد.



توزیع برنولی

مثال 2: در یک جعبه 20 کالا وجود دارد که 5 تای آنها معیوب است. اگر بخواهیم با جایگذاری چند کالا انتخاب کنیم، احتمال مرغوب بودن (احتمال موفقیت) در هر بار چقدر است؟ (آیا این یک آزمایش برنولی است؟)

پاسخ: چون آزمایشها به طور مستقل از هم صورت می گیرند و هر بار تعداد کالای مرغوب 15 و تعداد کل کالاها 20 می باشد، پس این یک آزمایش برنولی می باشد و احتمال موفقیت و شکست عبارتند از:

$$p = \frac{15}{20} = 0,75, \quad q = 1 - p = 0,25$$

مثال 3: اگر آزمایش مثال 2 بدون جایگذاری صورت پذیرد، آیا یک آزمایش برنولی خواهیم داشت:

پاسخ: خیر؛ زیرا احتمال موفقیت در هر آزمایش تغییر می کند. مثلاً دفعه اول $p_1 = \frac{15}{20}$ و دفعه دوم $p_2 = \frac{14}{19}$ می باشد.

سوال: آیا نمونه گیری بدون جایگذاری از جامعه بزرگ، آزمایش برنولی محسوب می شود؟

پاسخ: بله؛ زیرا **تغییرات** تعداد موفقیت در یک جامعه بزرگ نسبت به تعداد کل بسیار ناچیز و قابل اغماض است.



توزیع دوجمله‌ای

توزیع دو جمله‌ای: در n آزمایش برنولی که در آن احتمال موفقیت p است، متغیر تصادفی X که نشان‌دهنده تعداد موفقیت‌ها می‌باشد، دارای توزیعی به نام $(X \sim B(n, p))$ می‌باشد.

به عنوان مثال اگر یک آزمایش برنولی با احتمال موفقیت p را سه بار انجام دهیم، یکی از هشت حالت زیر رخ خواهد داد.

SSS, SSF, SFS, FSS, FSF, FFS, SFF, FFF

که در آن S نشان‌دهنده موفقیت و F نشان‌دهنده شکست است. از طرفی تعداد موفقیت‌ها نیز $0, 1, 2$ یا 3 می‌باشد.

بنابراین:



توزیع دوجمله‌ای

$$P(X=0) = q \times q \times q = q^3 = \binom{3}{0} p^0 q^3$$

$$P(X=1) = p \times q \times q + q \times p \times q + q \times q \times p = 3pq^2 = \binom{3}{1} p^1 q^2$$

$$P(X=2) = p \times p \times q + p \times q \times p + q \times p \times p = 3p^2q = \binom{3}{2} p^2 q^1$$

$$P(X=3) = p \times p \times p = p^3 = \binom{3}{3} p^3 q^0$$

همانطور که ملاحظه می‌شود، مقادیر احتمال به دست آمده، همان جملات بسط دو جمله‌ای $(p+q)^3$ می‌باشند. به این ترتیب در حالت کلی می‌توان نوشت:

$$X \sim B(n, p) \Rightarrow P(X=r) = \binom{n}{r} p^r q^{n-r}, (q=1-p)$$



توزیع دوجمله‌ای

مثال 1: دانشجویی می‌خواهد به 5 سوال دو گزینه‌ای پاسخ دهد. اگر احتمال پاسخگویی صحیح وی 0/7 باشد، احتمال اینکه دقیقاً به دو سوال پاسخ صحیح دهد را به دست آورید.

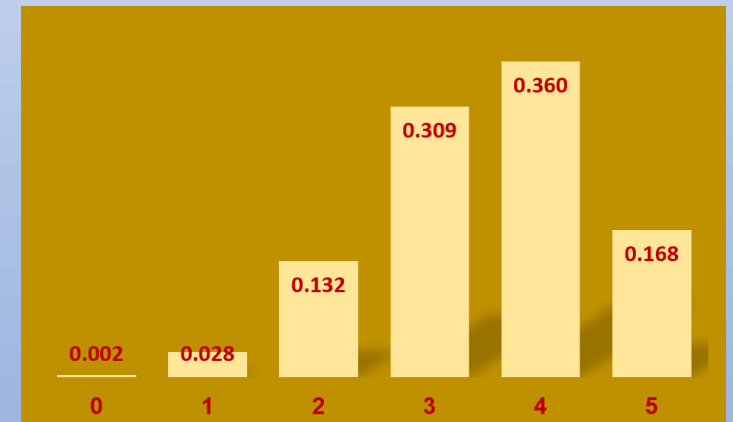
$$X \sim B(5, 0.7) \Rightarrow P(X = 2) = \binom{5}{2} 0.7^2 0.3^3 = 10 \times 0.49 \times 0.027 = 0.1323$$

مثال 2: تابع احتمال مثال 1 را برای همه مقادیر X بدست آورده و نمودار آن را رسم کنید.

$$P(X = 0) = \binom{5}{0} 0.7^0 0.3^5 = 0.002, \quad P(X = 1) = \binom{5}{1} 0.7^1 0.3^4 = 0.028$$

$$P(X = 2) = \binom{5}{2} 0.7^2 0.3^3 = 0.132, \quad P(X = 3) = \binom{5}{3} 0.7^3 0.3^2 = 0.309$$

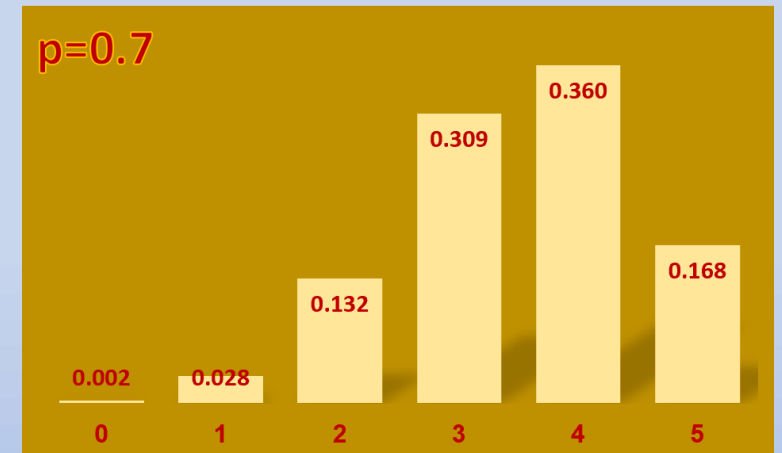
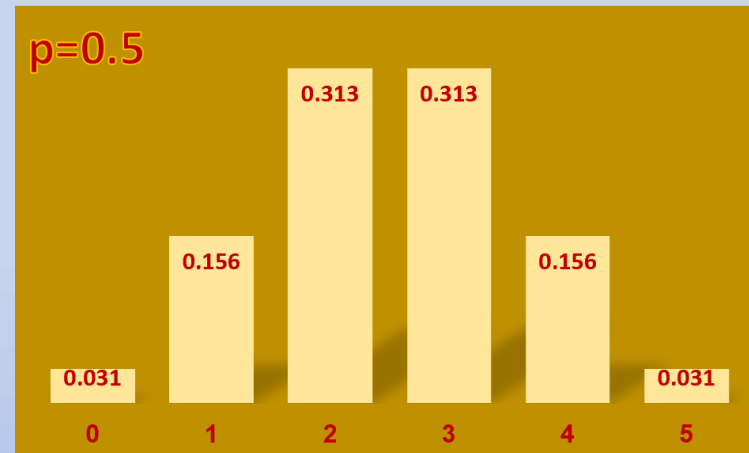
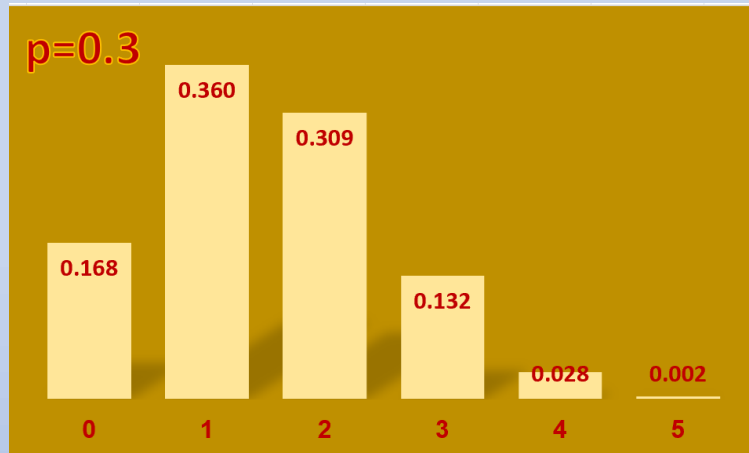
$$P(X = 4) = \binom{5}{4} 0.7^4 0.3^1 = 0.360, \quad P(X = 5) = \binom{5}{5} 0.7^5 0.3^0 = 0.168$$





توزیع دوجمله‌ای

مثال 3: نمودار تابع احتمال مثال 1 را برای حالت‌های $p=0/3$ و $p=0/5$ و $p=0/7$ رسم کنید.



همانطور که ملاحظه می‌شود در حالت $p=0/5$ نمودار کاملاً متقارن است.

همچنین در حالتی که $p > 0/5$ باشد، نمودار چوله به چپ و در حالتی که $p < 0/5$ باشد، نمودار چوله به راست دارد.



توزیع دو جمله‌ای

مثال 4: از بین شرکت‌هایی که سال گذشته حسابرسی شده‌اند، 0/25 آنها حساب‌هایشان مردود اعلام شده است. اگر 4 شرکت بر حسب تصادف انتخاب شده باشند، هر یک از احتمال‌های زیر را محاسبه کنید.

الف) حساب‌های یک شرکت رد شده باشد.

ب) حداکثر حساب‌های دو شرکت رد شده باشد.

$$X \sim B(4, 0,25) \Rightarrow P(X = 1) = \binom{4}{1} 0,25^1 0,75^3 = 0,42$$

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) = \binom{4}{0} 0,25^0 0,75^4 + \binom{4}{1} 0,25^1 0,75^3 + \binom{4}{2} 0,25^2 0,75^2 \\ &= 0,32 + 0,42 + 0,21 = 0,95 \end{aligned}$$



توزیع دوجمله‌ای

جدول ۱ احتمالات تجمعی دو جمله‌ای $P[x \leq c] = \sum_{x=0}^c \binom{n}{x} p^x (1-p)^{n-x}$

		p										
		.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
$n=1$	c											
	0	.950	.900	.800	.700	.600	.500	.400	.300	.200	.100	.050
	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=2$	0	.902	.810	.640	.490	.360	.250	.160	.090	.040	.010	.002
	1	.997	.990	.960	.910	.840	.750	.640	.510	.360	.190	.097
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=3$	0	.857	.729	.512	.343	.216	.125	.064	.027	.008	.001	.000
	1	.993	.972	.896	.784	.648	.500	.352	.216	.104	.028	.007
	2	1.000	.999	.992	.973	.936	.875	.784	.657	.488	.271	.143
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=4$	0	.815	.656	.410	.240	.130	.063	.026	.008	.002	.000	.000
	1	.986	.948	.819	.652	.475	.313	.179	.084	.027	.004	.000
	2	1.000	.996	.973	.916	.821	.688	.525	.348	.181	.052	.014
	3	1.000	1.000	.998	.992	.974	.938	.870	.760	.590	.344	.185
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=5$	0	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000
	1	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000
	2	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001
	3	1.000	1.000	.993	.969	.913	.813	.663	.472	.263	.081	.023
	4	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

برای مقادیر بزرگ n با استفاده از جدول توزیع دو ج
مقادیر احتمالات را محاسبه می‌کنیم.



توزیع دوجمله‌ای

مثال 5: فرض کنید 0/20 از مصرف‌کنندگان پودرهای شوینده، از مشتریان برند ما هستند. اگر 17 مصرف‌کننده انتخاب شوند، هر یک از احتمالات زیر را حساب کنید.

جدول ۱ احتمالات تجمعی دو جمله‌ای $P[X \leq c] = \sum_{x=0}^c \binom{n}{x} p^x (1-p)^{n-x}$

		.05	.10	.20	.30	.40	<i>P</i> .50	.60	.70	.80	.90	.95
<i>n</i>	<i>c</i>											
17	0	.418	.167	.023	.002	.000	.000	.000	.000	.000	.000	.000
	1	.792	.482	.118	.019	.002	.000	.000	.000	.000	.000	.000
	2	.950	.762	.310	.077	.012	.001	.000	.000	.000	.000	.000
	3	.991	.917	.549	.202	.046	.006	.000	.000	.000	.000	.000
	4	.999	.978	.758	.389	.126	.025	.003	.000	.000	.000	.000
	5	1.000	.995	.894	.597	.264	.072	.011	.001	.000	.000	.000
	6	1.000	.999	.962	.775	.448	.166	.035	.003	.000	.000	.000
	7	1.000	1.000	.989	.895	.641	.315	.092	.013	.000	.000	.000
	8	1.000	1.000	.997	.960	.801	.500	.199	.040	.003	.000	.000
	9	1.000	1.000	1.000	.987	.908	.685	.359	.105	.011	.000	.000
	10	1.000	1.000	1.000	.997	.965	.834	.552	.225	.038	.001	.000

الف) 3 مصرف‌کننده مشتری ما باشند.

ب) حداقل 4 مصرف‌کننده مشتری ما باشند.

ج) حداقل 2 و حداکثر 5 مصرف‌کننده مشتری ما باشند.

$$X \sim B(17, 0.2) \Rightarrow P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0.549 - 0.310 = 0.239$$

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.549 = 0.451$$

$$P(2 \leq X \leq 5) = P(X \leq 5) - P(X \leq 1) = 0.894 - 0.118 = 0.776$$



توزیع دو جمله‌ای

میانگین و واریانس توزیع دو جمله‌ای: همانطور که می‌دانیم توزیع دو جمله‌ای از تکرار n آزمایش ساده برنولی حاصل می‌شود.

y	0	1
$P(Y = y)$	q	p

در حالت $n=1$ توزیع دو جمله‌ای به صورت توزیع برنولی می‌باشد.

$$E(Y) = 0 \times q + 1 \times p = p$$

در این حالت میانگین و واریانس برابر است با:

$$E(Y^2) = 0^2 \times q + 1^2 \times p = p$$

$$V(Y) = E(Y^2) - (E(Y))^2 = p - p^2 = p(1 - p) = pq$$

حال اگر متغیر تصادفی X تعداد موفقیت‌ها در n آزمایش برنولی باشد:

$$X = Y_1 + Y_2 + \dots + Y_n$$

$$E(X) = E(Y_1 + Y_2 + \dots + Y_n) = E(Y_1) + E(Y_2) + \dots + E(Y_n) = p + p + \dots + p = np$$

$$V(X) = V(Y_1 + Y_2 + \dots + Y_n) = V(Y_1) + V(Y_2) + \dots + V(Y_n) = pq + pq + \dots + pq = npq$$



توزیع‌های دو جمله‌ای منفی و هندسی

توزیع دو جمله‌ای منفی: اگر بخواهیم احتمال x موفقیت از n آزمایش برنولی را که به توزیع دو جمله‌ای مربوط می‌شود را به دست آوریم یا تعداد آزمایشاتی که لازم است تا به K موفقیت دست پیدا کنیم را بخواهیم محاسبه کنیم، از توزیع دو جمله‌ای منفی استفاده می‌کنیم. مثلاً احتمال اینکه پنجمین فردی که یک شایعه را شنیده، دومین نفری باشد که آن را باور می‌کند یا احتمال اینکه ششمین کالای معیوب در یک کارتن، پنجمین کالای معیوبی باشد که مأمورین کنترل کیفیت متوجه آن شده باشند یا احتمال اینکه دزدی در هشتمین دزدی خود برای چهارمین بار دستگیر می‌شود.

فرمول توزیع دو جمله‌ای منفی به صورت مقابل می‌باشد.

$$P(X = x) = \binom{x-1}{K-1} p^K q^{x-K}, \quad x = K, K+1, K+2, \dots$$

که در آن K تعداد موفقیت‌ها در x آزمایش برنولی و p احتمال موفقیت در هر آزمایش برنولی است.

امید ریاضی و واریانس توزیع دو جمله‌ای منفی:

$$E(X) = \frac{K}{p}, \quad V(X) = \frac{Kq}{p^2}$$

این توزیع را با نام‌های توزیع پاسکال و توزیع زمان انتظار نیز می‌شناسند.



توزیع‌های دو جمله‌ای منفی و هندسی

توزیع هندسی: اگر در توزیع دو جمله‌ای منفی $K=1$ باشد، یعنی بخواهیم اولین موفقیت را در x آزمایش بدانیم، از توزیع هندسی که در واقع حالت خاصی از توزیع دو جمله‌ای منفی می‌باشد، استفاده می‌کنیم. مثلاً احتمال اینکه پنجمین فردی که شایعه را شنیده، اولین فردی باشد که آن را باور می‌کند یا احتمال اینکه ششمین کالای معیوب در یک کارتن، اولین کالای معیوبی باشد که ماموران کنترل کیفیت متوجه آن می‌شوند یا احتمال اینکه دزدی در هشتمین دزدی خود برای اولین بار دستگیر شود، نمونه‌هایی از توزیع هندسی هستند.

فرمول توزیع هندسی به صورت مقابل است:

$$P(X = x) = p q^{x-1}, \quad x = 1, 2, 3, \dots$$

که در آن p احتمال موفقیت در یک آزمایش و x شماره آزمایشی است که اولین موفقیت رخ داده است.

امید ریاضی و واریانس توزیع هندسی:

$$E(X) = \frac{1}{p}, V(X) = \frac{q}{p^2}$$



توزیع‌های دو جمله‌ای منفی و هندسی

مثال 1: تیراندازی 0/77 از تیرهای خود را به هدف می‌زند. احتمال اینکه سومین تیر وی اولین تیری باشد که به هدف اصابت می‌کند، چقدر است؟

$$x = 3, p = 0,77$$

$$P(X = 3) = 0,77 \times 0,23^{3-1} = 0,04$$

مثال 2: فردی متقاضی گواهینامه رانندگی است و با احتمال 0/65 در آزمون رد می‌شود. احتمال اینکه در دومین آزمون قبول شود، چقدر است؟

$$x = 2, p = 1 - 0,65 = 0,35$$

$$P(X = 2) = 0,35 \times 0,65^{2-1} = 0,23$$



توزیع چند جمله‌ای

اگر پیامدهای یک آزمایش بیشتر از دو حالت باشد، به شرطی که احتمال هر پیامد ثابت و آزمایشها از هم مستقل باشند، با یک توزیع چند جمله‌ای مواجه خواهیم بود.

اگر آزمایشی n بار به صورت مستقل انجام شود و هر آزمایش K پیامد با احتمالات $P_1, P_2, \dots, P_K$ داشته باشد به طوری که $P_1 + P_2 + \dots + P_K = 1$ داشته باشیم! ، آنگاه احتمال وقوع x_1 بار از پیامد 1، x_2 بار از پیامد 2، ... و x_K بار از پیامد K از رابطه زیر به دست می‌آید.

$$P(X_1 = x_1, X_2 = x_2, \dots, X_K = x_K) = \binom{n}{x_1, x_2, \dots, x_K} P_1^{x_1} P_2^{x_2} \dots P_K^{x_K} \quad , \quad (x_1 + x_2 + \dots + x_K = n)$$



توزیع چند جمله‌ای

مثال: 0/25 از افراد یک شهر به نامزد اول، 0/35 به نامزد دوم و 0/40 به نامزد سوم رای می‌دهند. فرض کنیم 10 نفر هم‌اکنون پای صندوق رای هستند. احتمال اینکه 2 نفر به نامزد اول، 3 نفر به نامزد دوم و 5 نفر به نامزد سوم رای دهند، چقدر است؟

$$P_1 = 0.25, P_2 = 0.35, P_3 = 0.40, x_1 = 2, x_2 = 3, x_3 = 5, n = 10$$

$$P(X_1 = 2, X_2 = 3, X_3 = 5) = \binom{10}{2, 3, 5} 0.25^2 \times 0.35^3 \times 0.40^5$$

$$= \frac{10!}{2!3!5!} \times 0.0625 \times 0.042875 \times 0.01024 = 0.0691488$$



توزیع فوق هندسی

توزیع فوق هندسی: فرض کنیم از بین N شی که k تای آن واجد شرایط هستند، n شی انتخاب کنیم به طوری که x تای آن واجد شرایط باشد. احتمال این انتخاب به صورت زیر قابل محاسبه است.

ابتدا تعداد کل حالات ممکن $\binom{N}{n}$ ترکیب n شی از N خواهد بود. حال باید از بین k شی واجد شرایط x تا $\binom{k}{x}$ و از بین $N-k$ شی که شرایط را ندارند، $n-x$ تا $\binom{N-k}{n-x}$ انتخاب کنیم. به این ترتیب خواهیم داشت:

$$P(X = x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \dots, k$$

امید ریاضی و واریانس توزیع فوق هندسی:

$$E(X) = \frac{nk}{N}$$

$$V(X) = \frac{nk(N-k)(N-n)}{N^2(N-1)}$$



توزیع فوق هندسی

مثال 1: از بین 8 مدیری که به جلسه امروز دعوت شده‌اند، 3 نفر رابطه‌مدار هستند. اگر به تصادف 4 نفر انتخاب کنیم،

الف) احتمال اینکه 2 مدیر رابطه‌مدار انتخاب شود، چقدر است؟

$$P(X=2) = \frac{\binom{3}{2} \binom{8-3}{4-2}}{\binom{8}{4}} = \frac{3 \times 5}{70} = \frac{3}{14}$$

ب) امید ریاضی و واریانس تعداد مدیران رابطه‌مدار را محاسبه کنید.

$$E(X) = \frac{nk}{N} = \frac{4 \times 3}{8} = 1,5$$

$$V(X) = \frac{nk(N-k)(N-n)}{N^2(N-1)} = \frac{4 \times 3 \times (8-3) \times (8-4)}{8^2 \times (8-1)} = \frac{12 \times 5 \times 4}{64 \times 7} \approx 0,536$$



توزیع فوق هندسی

اگر N در مقایسه با n بزرگ باشد (n کمتر از 5 درصد N باشد)، تفاوت چندانی بین نمونه‌گیری با جایگزینی و بدون جایگزینی وجود نخواهد داشت و می‌توان از توزیع دو جمله‌ای با پارامترهای n و $p=k/N$ به عنوان تقریب توزیع فوق هندسی استفاده کرد.

مثال 2: از بین 200 متقاضی یک شغل، تنها 70 نفر واجد شرایط هستند. اگر 6 نفر را بر حسب تصادف انتخاب کنیم، احتمال اینکه 3 نفر از آنها واجد شرایط باشند، چقدر است؟ ($x=3$, $k=70$, $N=200$)

$$\frac{n}{N} = \frac{6}{200} = 0,03 < 0,05 \quad \text{پارامترهای توزیع فوق هندسی:}$$

اما با توجه به بزرگ بودن N نسبت به n) می‌توان از تقریب دو جمله‌ای با پارامترهای زیر استفاده کرد.

$$n=6, \quad p = \frac{k}{N} = \frac{70}{200} = 0,35$$

$$P(X=3) = \binom{6}{3} \times (0,35)^3 \times (0,65)^{6-3} = 0,24$$



توزیع پواسون

در توزیع دو جمله‌ای، هنگامی که n بزرگ باشد، محاسبات بسیار دشوار خواهد بود. مثلاً اگر بخواهیم احتمال اینکه از بین 2000 کالا که هر یک با احتمال 0/0015 معیوب هستند، 5 کالا معیوب باشند را محاسبه کنیم باید حاصل $\binom{2000}{5} (0/0015)^5 (0/9985)^{1995}$ را بدست آوریم.

در چنین مواقعی از توزیع پواسن برای تقریب چنین توزیعی استفاده می‌کنیم. $\lambda = np$, $e \simeq 2/718$

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

مثال 1: احتمال اینکه از بین 2000 کالا که هر یک با احتمال 0/0015 معیوب هستند، 5 کالا معیوب باشند، چقدر است؟

$$\lambda = np = 2000 \times 0/0015 = 3$$

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-3} \times 3^5}{5!} = 0/1008 = 10/08\%$$



توزیع پواسون

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

$$E(X) = \lambda, \quad V(X) = \lambda$$

امید ریاضی و واریانس توزیع پواسون:

مثال 2: بر اساس تجربه مشخص شده است که یک تلفنچی 3 درصد از تماسها را اشتباه وصل می‌کند. اگر امروز 150 تماس وصل کرده باشد، هر یک از موارد زیر را محاسبه کنید.

$$E(X) = \lambda = np = 150 \times 0.03 = 4.5$$

الف) امید ریاضی (میانگین) تماس‌هایی که اشتباه وصل کرده است.

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-4.5} \times 4.5^3}{3!} = 0.1687$$

ب) احتمال اینکه 3 شماره را اشتباه وصل کرده باشد.

ج) احتمال اینکه بیش از یک تماس را اشتباه وصل کرده باشد.

$$P(X > 1) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) =$$

$$1 - \frac{e^{-4.5} \times 4.5^0}{0!} - \frac{e^{-4.5} \times 4.5^1}{1!} = 0.9389$$



توزیع پواسون

توزیع پواسون برای تعداد مراجعات:

معمولا تعداد مراجعات به یک سیستم از توزیع پواسون تبعیت می کند. اگر متوسط تعداد مراجعات در واحد زمان باشد، داریم:

$$P(X = x) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}, \quad x = 0, 1, 2, \dots$$

مثال 3: تعداد مشتریانی که به یک بانک مراجعه می کنند دارای توزیع پواسون با میانگین 2 مشتری در هر دقیقه می باشد. مطلوب است:

$$\lambda t = 2 \times 1 = 2, \quad P(X = 0) = \frac{e^{-2} (2)^0}{0!} = 0.1353$$

الف) احتمال اینکه در یک دقیقه اول هیچ مشتری مراجعه نکند.

$$\lambda t = 2 \times \frac{105}{60} = 3.5, \quad P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{e^{-3.5} (3.5)^0}{0!} + \frac{e^{-3.5} (3.5)^1}{1!} + \frac{e^{-3.5} (3.5)^2}{2!} = 0.3299$$

ج) احتمال اینکه در 1/5 دقیقه اول یک مشتری مراجعه کند.

$$\lambda t = 2 \times 1/5 = 0.4, \quad P(X = 1) = \frac{e^{-0.4} (0.4)^1}{1!} = 0.2689$$



فصل هفتم

توابع احتمال پیوسته

Continuous Probability Functions



تابع چگالی احتمال و تابع توزیع متغیر تصادفی پیوسته

متغیر تصادفی پیوسته: همانطور که می‌دانیم متغیر تصادفی، تابعی است که به هر یک از اعضای فضای نمونه یک عدد حقیقی نسبت می‌دهد. در صورتی که برد این تابع مجموعه‌ای نامتناهی و غیرقابل شمارش باشد، متغیر تصادفی را پیوسته می‌نامند.

تابع چگالی احتمال: تابعی که به همه مقادیر متغیر تصادفی پیوسته، احتمال آن را نسبت می‌دهد، تابع چگالی احتمال نام دارد. باید توجه داشت تابع چگالی احتمال در یک نقطه خاص برابر صفر است زیرا تعداد حالات ممکن بینهایت می‌باشد و معمولا از آن برای محاسبه احتمال در یک بازه استفاده می‌شود.

به عنوان مثال متغیر تصادفی مسافت پیموده شده توسط یک اتومبیل در یک اتوبان 50 کیلومتری، پیوسته

می‌باشد. برای این متغیر می‌توان تابع چگالی احتمالی به صورت زیر در نظر گرفت:

$$f(x) = \begin{cases} \frac{x}{1250}, & 0 \leq x \leq 50 \\ 0, & \text{Otherwise} \end{cases}$$



تابع چگالی احتمال و تابع توزیع متغیر تصادفی پیوسته

برای محاسبه احتمال یک متغیر تصادفی پیوسته، از رابطه زیر استفاده می‌کنیم.

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

بدیهی است در یک نقطه مشخص داریم:

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x) dx = 0$$

همچنین همه احتمالات زیر با هم برابر هستند.

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b)$$

اما دو نکته مهم در مورد تابع چگالی احتمال این است که:

$$1) \quad \forall x, f(x) \geq 0$$

$$2) \quad \int_{-\infty}^{+\infty} f(x) dx = 1$$



تابع چگالی احتمال و تابع توزیع متغیر تصادفی پیوسته

مثال 1: با توجه به تابع چگالی احتمال داده شده، مقادیر زیر را حساب کنید.

$$f(x) = \begin{cases} \frac{2}{3}x + \frac{1}{3}K, & 0 \leq x \leq 1 \\ 0, & \text{Otherwise} \end{cases}$$

الف) K

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \Rightarrow \int_0^1 \left(\frac{2}{3}x + \frac{1}{3}K\right)dx = 1 \Rightarrow \left.\frac{1}{3}x^2 + \frac{K}{3}x\right|_0^1 = 1 \Rightarrow \frac{1}{3} + \frac{K}{3} = 1 \Rightarrow K = 2$$

ب) $P(0 \leq X \leq 0.5)$

$$P(0 \leq X \leq 0.5) = \int_0^{0.5} f(x)dx = \int_0^{0.5} \left(\frac{2}{3}x + \frac{2}{3}\right)dx = \left.\frac{1}{3}x^2 + \frac{2}{3}x\right|_0^{0.5} = \frac{5}{12}$$

ج) $P(0.5 \leq X \leq 1)$

$$P(0.5 \leq X \leq 1) = \int_{0.5}^1 f(x)dx = \int_{0.5}^1 \left(\frac{2}{3}x + \frac{2}{3}\right)dx = \left.\frac{1}{3}x^2 + \frac{2}{3}x\right|_{0.5}^1 = \frac{7}{12}$$



تابع چگالی احتمال و تابع توزیع متغیر تصادفی پیوسته

$$f(x) = \begin{cases} \frac{1}{12}, & 324 \leq x \leq 326 \\ 0, & \text{Otherwise} \end{cases}$$

مثال 2: در یک شرکت مواد غذایی میزان ژبّی که در قوطی‌های 330 گرمی ریخته می‌شود،

دارای تابع چگالی احتمال مقابل است. فرض کنید یک قوطی به تصادف انتخاب شود. احتمالات زیر را محاسبه کنید.

$$P(X \leq 332) = \int_{-\infty}^{332} f(x) dx = \int_{324}^{332} \frac{1}{12} dx = \frac{1}{12} x \Big|_{324}^{332} = \frac{2}{3}$$

الف) حداکثر 332 گرم ژب داشته باشد.

$$P(X \geq 325) = \int_{325}^{+\infty} f(x) dx = \int_{325}^{336} \frac{1}{12} dx = \frac{1}{12} x \Big|_{325}^{336} = \frac{11}{12}$$

ب) حداقل 325 گرم ژب داشته باشد.

$$P(330,5 \leq X \leq 335,5) = \int_{330,5}^{335,5} f(x) dx = \int_{330,5}^{335,5} \frac{1}{12} dx = \frac{1}{12} x \Big|_{330,5}^{335,5} = \frac{5}{12}$$

ج) بین 330/5 و 335/5 گرم ژب داشته باشد.

$$P(X = 330,12) = \int_{330,12}^{330,12} f(x) dx = \int_{330,12}^{330,12} \frac{1}{12} dx = \frac{1}{12} x \Big|_{330,12}^{330,12} = 0$$

د) دقیقاً 330/12 گرم ژب داشته باشد.



تابع چگالی احتمال و تابع توزیع متغیر تصادفی پیوسته

مثال 3: متغیر تصادفی X با تابع چگالی احتمال مقابل مفروض است.
احتمالات زیر را حساب کنید.

الف) $P(1 \leq X \leq 5)$

$$P(1 \leq X \leq 5) = \int_1^5 2e^{-2x} dx = -e^{-2x} \Big|_1^5 = e^{-5} + e^{-2} = 0,135$$

ب) $P(-5 \leq X \leq 2,75)$

$$P(-5 \leq X \leq 2,75) = \int_0^{2,75} 2e^{-2x} dx = -e^{-2x} \Big|_0^{2,75} = e^{-5,5} + e^0 = 0,996$$



تابع چگالی احتمال و تابع توزیع متغیر تصادفی پیوسته

تابع توزیع متغیر تصادفی پیوسته: در بسیاری از موارد مایلیم احتمال اینکه مقادیر یک متغیر تصادفی کوچکتر یا مساوی مقدار مشخصی باشد را محاسبه کنیم. در چنین مواردی تابع توزیع از رابطه زیر قابل محاسبه می‌باشد.

$$F(X) = P(X \leq x) = \int_{-\infty}^x f(x) dx$$

$$F(+\infty) = 1 \quad F(-\infty) = 0$$

بدیهی است که $F(x)$ و $f(x)$ است.

محاسبه احتمال با استفاده از تابع توزیع: $P(a \leq X \leq b) = \int_a^b f(x) dx = F(b) - F(a)$

$$P(a \leq X \leq b) = F(b) - F(a)$$



تابع چگالی احتمال و تابع توزیع متغیر تصادفی پیوسته

مثال 4: با در نظر گرفتن تابع چگالی احتمال مثال 3، بار دیگر مقادیر خواسته شده در قسمت‌های (الف) و (ب)

را با استفاده از تابع توزیع بار دیگر محاسبه کنید.

$$F(X) = P(X \leq x) = \int_{-\infty}^x f(x) dx = \int_0^x 2e^{-2x} dx = -e^{-2x} \Big|_0^x = 1 - e^{-2x}$$

$$\Rightarrow F(X) = \begin{cases} 1 - e^{-2x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$P(1 \leq X \leq 5) = F(5) - F(1) = (1 - e^{-10}) - (1 - e^{-2}) = 0,999 - 0,864 = 0,135$$

(الف)

$$P(-5 \leq X \leq 2,75) = F(2,75) - F(-5) = (1 - e^{-5,5}) - (0) = 0,996 - 0 = 0,996$$

(ب)



امید ریاضی و واریانس متغیر تصادفی پیوسته

همانطور که می‌دانیم برای یک متغیر تصادفی گسسته امید ریاضی و واریانس به صورت زیر تعریف می‌شوند.

$$E(X) = \sum_x x \cdot f(x) \quad , \quad V(X) = \sum_x (x - E(X))^2 f(x) = \sum_x x^2 \cdot f(x) - (E(X))^2$$

در مورد متغیر تصادفی پیوسته به طرز مشابه امید ریاضی و واریانس به صورت زیر تعریف می‌شوند.

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx \quad , \quad V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx = \int_{-\infty}^{+\infty} x^2 f(x) dx - (E(X))^2$$



امید ریاضی و واریانس متغیر تصادفی پیوسته

مثال 1: سود شرکتی با تابع چگالی احتمال زیر داده شده است. امید ریاضی و واریانس سود شرکت را محاسبه کنید.

$$f(x) = \begin{cases} \frac{2}{63}(x+3), & -2 < x < 5 \\ 0, & \text{Otherwise} \end{cases}$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-2}^5 x \frac{2}{63} (x+3) dx = \frac{2}{63} \left(\frac{1}{3} x^3 + \frac{3}{2} x^2 \right) \Big|_{-2}^5 = 2,407$$

$$V(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - (E(X))^2 = \int_{-\infty}^{+\infty} x^2 \frac{2}{63} (x+3) dx - (2,407)^2$$

$$= \frac{2}{63} \left(\frac{1}{4} x^4 + x^3 \right) \Big|_{-2}^5 - (2,407)^2 = \frac{2}{63} (281,25 + 4) - (2,407)^2 = 9,056 - 5,794 = 3,262$$



امید ریاضی و واریانس متغیر تصادفی پیوسته

مثال 2: تابع چگالی احتمال مقابل را در نظر بگیرید و هر یک از موارد زیر را محاسبه کنید.

$$f(x) = \begin{cases} \frac{2x}{25}, & 0 < x < 5 \\ 0, & \text{Otherwise} \end{cases}$$

الف) گشتاور مرتبه اول نسبت به صفر $E(X)$:

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^5 x \frac{2x}{25} dx = \frac{2x^3}{75} \Big|_0^5 = \frac{10}{3}$$

ب) گشتاور مرتبه دوم نسبت به صفر $E(X^2)$:

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^5 x^2 \frac{2x}{25} dx = \frac{x^4}{50} \Big|_0^5 = \frac{25}{2}$$

ج) گشتاور مرتبه سوم نسبت به صفر $E(X^3)$:

$$E(X^3) = \int_{-\infty}^{+\infty} x^3 f(x) dx = \int_0^5 x^3 \frac{2x}{25} dx = \frac{2x^5}{125} \Big|_0^5 = 50$$

د) گشتاور مرتبه چهارم نسبت به صفر $E(X^4)$:

$$E(X^4) = \int_{-\infty}^{+\infty} x^4 f(x) dx = \int_0^5 x^4 \frac{2x}{25} dx = \frac{x^6}{75} \Big|_0^5 = \frac{625}{3}$$



توزیع یکنواخت

یکی از ساده‌ترین و در عین حال مهم‌ترین توزیع‌ها، توزیع احتمال یکنواخت است. متغیر تصادفی پیوسته X که بین دو مقدار a و b دارای مقدار می‌باشد.

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{Otherwise} \end{cases}$$

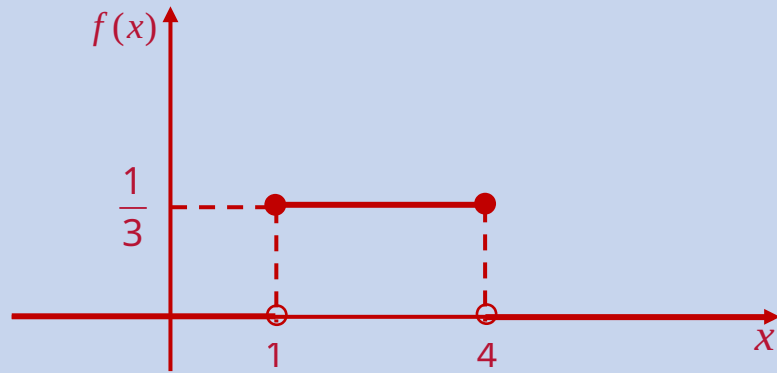
امید ریاضی و واریانس توزیع یکنواخت به صورت زیر محاسبه می‌شود.

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{2}(a+b)$$

$$V(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - (E(X))^2 = \int_a^b \frac{x^2}{b-a} dx - \frac{1}{4}(a+b)^2 = \frac{1}{12}(b-a)^2$$



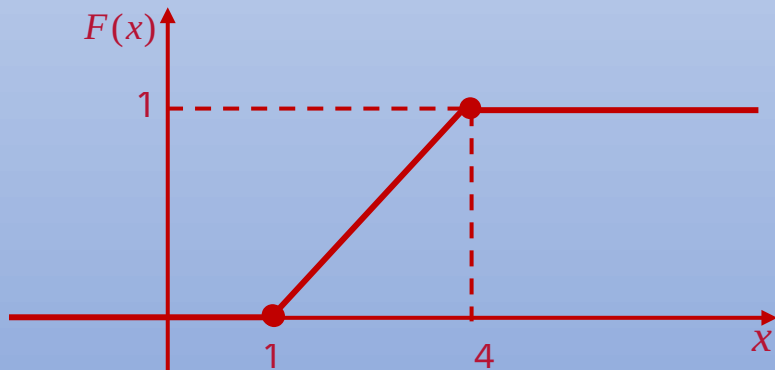
توزیع یکنواخت



مثال 1: برای تابع چگالی احتمال یکنواخت مقابل،
$$f(x) = \begin{cases} \frac{1}{3}, & 1 \leq x \leq 4 \\ 0, & \text{Otherwise} \end{cases}$$

الف) نمودار تابع چگالی را رسم کنید.

ب) تابع توزیع را به دست آورید.
$$F(X) = \int_{-\infty}^x f(x) dx = \int_{-\infty}^x \frac{1}{3} dx = \frac{1}{3} x \Big|_1^x = \frac{1}{3} (x - 1) \Rightarrow F(X) = \begin{cases} 0, & x < 1 \\ \frac{1}{3} (x - 1), & 1 \leq x \leq 4 \\ 1, & x > 4 \end{cases}$$



ج) نمودار تابع توزیع را رسم کنید.



توزیع نمایی

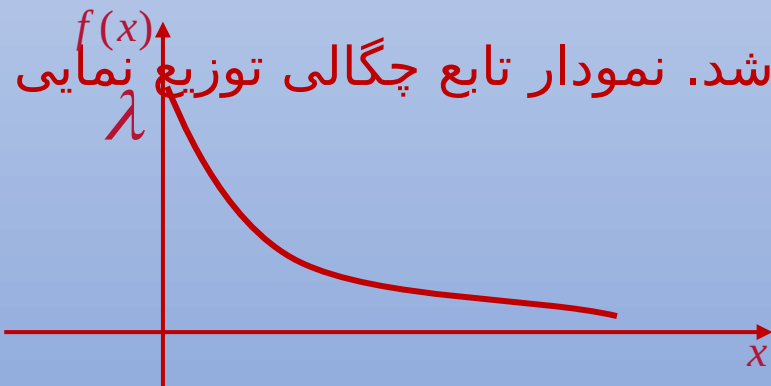
اگر تعداد موفقیت‌ها یا ورودی‌ها دارای توزیع پواسون باشد، زمان بین موفقیت‌ها یا ورودی‌های متوالی دارای «توزیع نمایی منفی» می‌باشد. از آنجائیکه زمان یک کمیت پیوسته است، نتیجه می‌گیریم که این توزیع نمایی منفی، پیوسته می‌باشد.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \text{ می‌باشد.} \\ 0 & , \text{ Otherwise} \end{cases}$$

تابع چگالی این توزیع که از این پس آن را تنها توزیع نمایی می‌نامیم، به صورت زیر می‌باشد.

λ

پارامتر λ ، متوسط تعداد موفقیت‌ها یا ورودی‌ها در واحد زمان می‌باشد. نمودار تابع چگالی توزیع نمایی نیز به صورت زیر می‌باشد.





توزیع نمایی

امید ریاضی، واریانس و تابع توزیع در توزیع نمایی:

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \underline{\frac{1}{\lambda}}$$

$$V(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - (E(X))^2 = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} = \underline{\frac{1}{\lambda^2}}$$

$$F(X) = P(X \leq x) = \int_{-\infty}^x f(x) dx = \int_0^x \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_0^x = \underline{1 - e^{-\lambda x}}$$



توزیع نمایی

مثال 1: به مدیری به طور متوسط در هر ساعت 5 بار تلفن زده می شود که تعداد تماسها از توزیع پواسون پیروی می کند.

$$E(X) = \frac{1}{\lambda} = \frac{1}{5} = 0,2$$

الف) متوسط فاصله زمانی بین دو تماس چقدر است؟

ب) احتمال اینکه فاصله زمانی بین دو تماس متوالی کمتر از 10 دقیقه باشد، چقدر است؟
 $x = 10_{\min} = \frac{10}{60} \text{ hour} = 0,167$, $P(X \leq 0,167) = F(0,167) = 1 - e^{-5 \times 0,167} = 1 - e^{-0,833} = 0,565$

ج) احتمال اینکه فاصله دو تماس متوالی بیش از 3 دقیقه باشد، چقدر است؟
 $x = 3_{\min} = \frac{3}{60} \text{ hour} = 0,05$, $P(X > 0,05) = 1 - F(0,05) = e^{-5 \times 0,05} = 0,779$



توزیع نمایی

مثال 2: به طور متوسط یک روز در میان یک کشتی با توزیع نمایی به اسکله‌ای می‌رسد. اگر امروز یک کشتی رسیده باشد، احتمال اینکه 4 روز طول بکشد تا کشتی دیگری برسد، را محاسبه کنید.

$$\lambda = \frac{1}{2}, P(X > 4) = 1 - F(4) = e^{-\frac{1}{2} \times 4} = e^{-2} = 0,135$$



فصل هشتم

توزیع نرمال

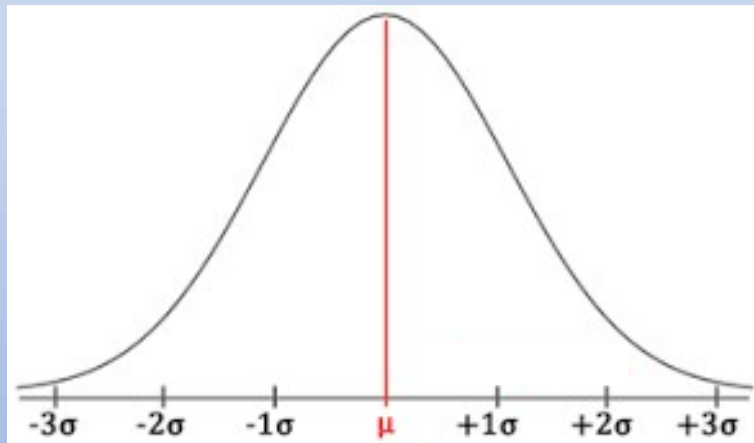
Normal Distribution



توزیع نرمال چیست؟

یکی از توزیع‌های مهم در آمار که بسیاری از پدیده‌ها در طبیعت از آن تبعیت می‌کنند، توزیع زنگی شکلی (Bell Shape) است که به همت دانشمندانی همچون لاپلاس و دموآور کشف شد و با تلاش گوس از روش ریاضی به دست آمد.

این توزیع را توزیع نرمال (Normal Distribution) یا توزیع گوسی (Gaussian Distribution) می‌نامیم. نمودار این توزیع و تابع چگالی مربوط به آن به صورت زیر می‌باشد.



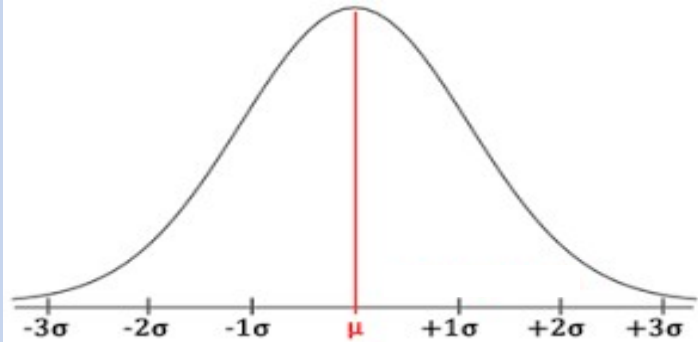
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$e \approx 2.71828$, $\pi \approx 3.141592$

در این تابع مقادیر μ و σ پارامترهای توزیع نام دارند و با مشخص بودن آنها توزیع نامده مشخص و قابل رسم است. یک متغیر تصادفی X که از توزیع نرمال با پارامترهای μ و σ تبعیت کند را با نماد $X \sim N(\mu, \sigma)$ نمایش می‌دهند.



خصوصیات توزیع نرمال:



(1) مساحت زیر منحنی نرمال و بالای محور x برابر 1 است. $\int_{-\infty}^{+\infty} f(x) dx = 1$

(2) به ازای تمام مقادیر x ، مقدار $f(x)$ بزرگتر یا مساوی صفر می باشد. $\forall x, f(x) \geq 0$

(3) حداکثر مقدار تابع در $x = \mu$ حاصل می شود. $\text{if } (X = \mu) \text{ then } (f'(x) = 0, f''(x) < 0)$

(4) نمودار تابع حول میانگین (μ) متقارن است. $f(\mu + x) = f(\mu - x)$

(5) امید ریاضی و واریانس متغیر تصادفی X به ترتیب σ^2 و μ می باشند.

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \mu, \quad V(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = \sigma^2$$

(6) با دورتر شدن از میانگین، چه در سمت راست و چه در سمت چپ، منحنی به محور x ها نزدیک می شود. $\lim_{x \rightarrow \pm\infty} f(x) = 0$

(7) در این توزیع، میانگین، میانه و مد برابر هستند. $\mu = Mo = Md$



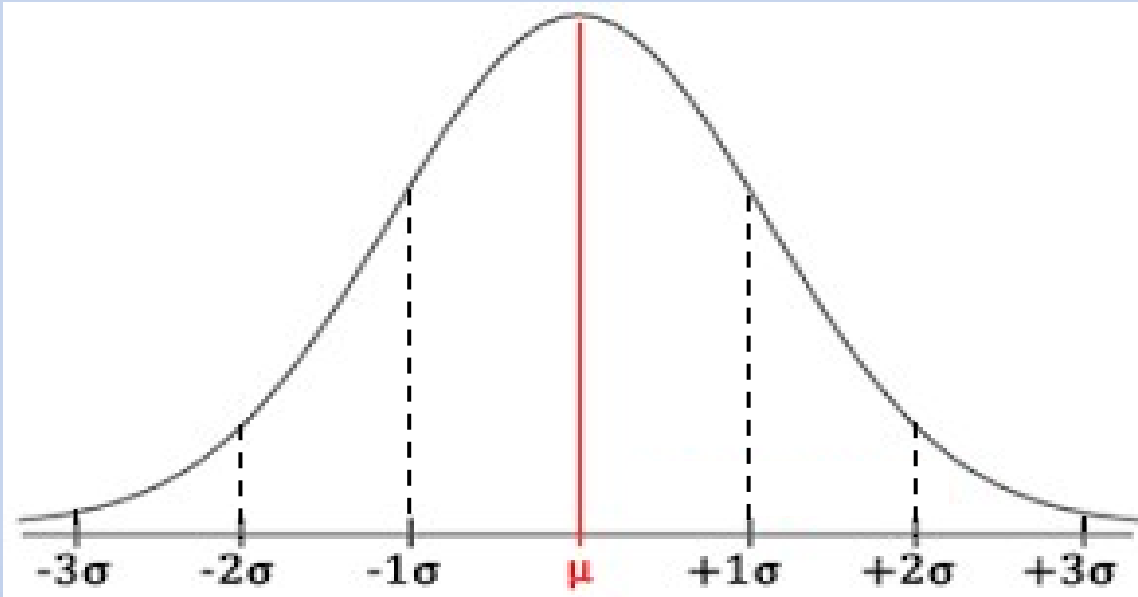
ادامه خصوصیات توزیع نرمال:

(8) احتمال فاصله‌ای در اطراف میانگین به صورت زیر

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.683$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.954$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.997$$



به عنوان مثال اگر قد دانشجویان دانشکده مدیریت از توزیع نرمال تبعیت کند و دارای میانگین 160 سانتیمتر و انحراف معیار 5 سانتیمتر باشد، به احتمال 68/3% قد یک دانشجو در این دانشکده بین 155 و 165 سانتیمتر قرار دارد. همچنین احتمال اینکه قد یک دانشجو بین 170 و 150 سانتیمتر باشد، 95/4% می‌باشد و احتمال اینکه قد یک دانشجو در این دانشکده بلندتر از 175 یا کوتاهتر از 145 سانتیمتر باشد، 0/3% است.

به همین ترتیب، احتمال اینکه قد یک دانشجو بلندتر از 170 سانتیمتر باشد، 2/3% می‌باشد.



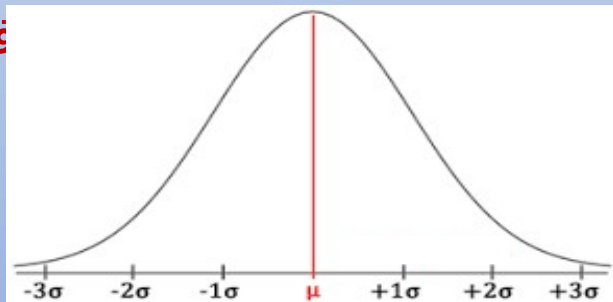
توزیع نرمال استاندارد:

به دلیل اینکه محاسبه احتمال در یک بازه‌ی دلخواه در توزیع نرمال همراه با محاسبات نسبتاً طولانی همراه است و هر توزیع نرمال دارای پارامترهای (میانگین و انحراف معیار) خاص خود می‌باشد، با استفاده از تغییر متغیر پارامترهای توزیع $\mu=0$ و $\sigma=1$ به صورت استاندارد به صورت

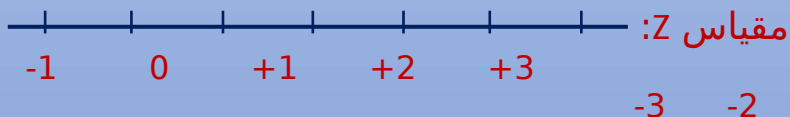
می‌آوریم و مقادیر احتمال را در توزیع نرمال استاندارد شده محاسبه و در جدولی گزارش می‌کنیم. به این ترتیب برای محاسبه احتمال در یک توزیع نرمال دلخواه دیگر لازم نیست به کمک انتگرال مساحت زیر منحنی نرمال را بدست آوریم.

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

تابع چگالی در توزیع نرمال استاندارد به صورت $Z \sim N(0,1)$ دارای توزیع نرمال با میانگین صفر و انحراف معیار یک خواهد بود.



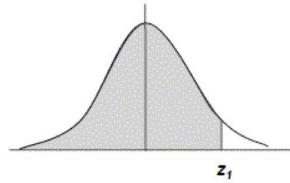
مقیاس X:



مقیاس Z:



Standard Normal Distribution



$$p(z \leq z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{z^2}{2}} dz$$

استفاده مستقیم از جدول توزیع نرمال استاندارد

z_1	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9859
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

برای محاسبه احتمال یک پیشامد که دارای توزیع نرمال می باشد

ابتدا به کمک تغییر متغیر $Z = \frac{X - \mu}{\sigma}$ مقیاس را به توزیع نرمال

استاندارد تبدیل می کنیم و سپس به کمک جدول، مقدار احتمال را

محاسبه می کنیم.

مثال: در یک شرکت میانگین حقوق کارکنان 5 میلیون تومان با انحراف

500 هزار تومان می باشد. احتمال اینکه یکی از کارکنان شرکت که به

انتخاب می شود حقوقی بیشتر از 6,075,000 تومان بگیرد، چقدر است؟

$$Z = \frac{6,075,000 - 5,000,000}{500,000} = 2.15$$

$$P(Z \leq 2.15) = 0.9842 \Rightarrow P(Z > 2.15) = 1 - 0.9842 = 0.0158 = 1.58\%$$



استفاده مستقیم از جدول توزیع نرمال استاندارد

احتمالهای نرمال استاندارد

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

مثال: در یک توزیع نرمال با میانگین 30 و انحراف معیار 9، با چه احتمالی یکی از داده‌ها بین 24 تا 43 قرار بگیرد، چقدر است؟

$$\begin{aligned} P(24 \leq X \leq 43) &= P\left(\frac{24 - 30}{9} \leq Z \leq \frac{43 - 30}{9}\right) \\ &= P(-0.67 \leq Z \leq 1.44) = P(Z \leq 1.44) - P(Z \leq -0.67) \\ &= 0.9251 - 0.2514 = 0.6737 = 67.37\% \end{aligned}$$



استفاده مستقیم از جدول توزیع نرمال استاندارد

مثال: میزان آبلیموی پر شده در شیشه‌های آبلیمو از توزیع نرمال با میانگین 330 میلی‌لیتر و انحراف معیار 5 میلی‌لیتر تبعیت می‌کند. احتمال‌های زیر را حساب کنید.

الف) احتمال اینکه شیشه‌ای بین 322 تا 328 میلی‌لیتر آبلیمو داشته باشد.

$$\begin{aligned} P(322 \leq X \leq 328) &= P\left(\frac{322 - 330}{5} \leq Z \leq \frac{328 - 330}{5}\right) = P(-1.6 \leq Z \leq -0.4) \\ &= P(Z \leq -0.4) - P(Z \leq -1.6) = 0.3446 - 0.0548 = 0.2898 = 28.98\% \end{aligned}$$

ب) احتمال اینکه شیشه‌ای بیش از 335 میلی‌لیتر آبلیمو داشته باشد.

$$P(X \geq 335) = P\left(Z \geq \frac{335 - 330}{5}\right) = P(Z \geq 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587 = 15.87\%$$

ج) در بررسی میزان آبلیموی 70 شیشه که توسط واحد کنترل کیفیت انجام شده است، انتظار می‌رود چند شیشه بیش از 335 میلی‌لیتر آبلیمو داشته باشند؟

$$n = 0.1587 \times 70 = 11.109 \simeq 11$$



استفاده مستقیم از جدول توزیع نرمال استاندارد

مثال: زمان لازم برای تعمیر یک ماشین که در خط تولید کار می‌کند دارای توزیع نرمال با میانگین 50 دقیقه و انحراف معیار 20 دقیقه است. هزینه هر بار تعمیر 5 میلیون ریال است و در صورتیکه تعمیر این ماشین بیش از 85 دقیقه طول بکشد، به علت توقف خط تولید، ضرری معادل 100 میلیون ریال به بار می‌آید. امید ریاضی (میانگین) هزینه هر بار خرابی این دستگاه را محاسبه کنید.

$$P(X \geq 85) = P\left(Z \geq \frac{85 - 50}{20}\right) = P(Z \geq 1.75) = 1 - P(Z \leq 1.75) = 1 - 0.9599 = 0.0401$$

پاسخ:

بنابراین احتمال اینکه تعمیر ماشین کمتر از 85 دقیقه طول بکشد 0.9599 می‌باشد که باید هزینه تعمیر ماشین را پردازیم و احتمال اینکه بیش از 85 دقیقه تعمیر ماشین طول بکشد، 0.0401 می‌باشد که مجبور به تحمل هزینه تعمیر و زیان توقف

خط تولید هستیم. بنابراین امید ریاضی هزینه هر بار خرابی این ماشین برابر است با:

$$E(C) = 0.9599 \times 5,000,000 + 0.0401 \times 100,000,000 = 9,010,000$$



استفاده معکوس از جدول توزیع نرمال استاندارد

همانطور که ملاحظه شد، در استفاده مستقیم از جدول توزیع نرمال استاندارد، با داشتن مقدار متغیر تصادفی، مقدار احتمال مربوط به آن را محاسبه می‌کردیم. حال می‌خواهیم با داشتن احتمال وقوع یک پیشامد، مقدار متغیر تصادفی نظیر آن را بیابیم.

به این منظور با مراجعه به جدول توزیع نرمال استاندارد، مقدار متغیر تصادفی استاندارد Z را پیدا می‌کنیم و سپس با استفاده از میانگین و انحراف معیار توزیع نرمال اصلی مقدار متغیر تصادفی X را محاسبه می‌کنیم.

مثال: میزان مصرف مواد اولیه در یک شرکت تولیدی، دارای توزیع نرمال با میانگین 752 تن و انحراف معیار 86 تن است. این شرکت مواد اولیه مورد نیاز خود را در ابتدای هر ماه سفارش می‌دهد. این شرکت چند تن مواد اولیه باید سفارش دهد تا با اطمینان 95 درصد کمبود مواد در طول ماه نداشته باشد؟

$$X = \mu + \sigma Z = 752 + 86 \times 1.645 = 893.47$$

در نتیجه

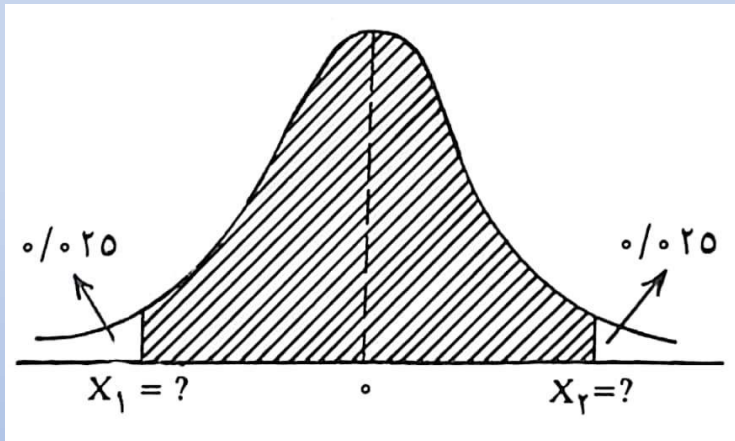
$$P(Z \leq 1.645) = 0.9500$$

پاسخ: با مراجعه به جدول



استفاده معکوس از جدول توزیع نرمال استاندارد

مثال: دستگاهی طوری تنظیم شده که به صورت خودکار و به طور متوسط پیچ‌هایی به قطر 38 میلیمتر و با انحراف معیار 0/2 میلیمتر تولید می‌کند. می‌خواهیم بدانیم 95 درصد پیچ‌های تولیدی در چه دامنه‌ای در دو طرف میانگین قرار می‌گیرند.



پاسخ: با توجه به تقارن منحنی مساحت دو طرف منحنی برابر 0/025
 $P(Z \leq -1,96) = 0,025$ و $P(Z \geq 1,96) = 0,025$
از طرفی با مراجعه به جدول

در نتیجه:

$$X_1 = \mu + \sigma Z_1 = 38 + 0,2 \times (-1,96) = 37,608$$

$$X_2 = \mu + \sigma Z_2 = 38 + 0,2 \times (+1,96) = 38,392$$

$$\Rightarrow 37,608 \leq X \leq 38,392$$

95 درصد پیچ‌ها در این بازه قرار دارند.



روش‌ی برای تشخیص نرمال بودن توزیع

روش‌های مختلفی برای بررسی نرمال بودن یک توزیع آماری وجود دارد. یکی از روش‌های نسبتاً ساده و تا حدی ذهنی برای تشخیص نرمال بودن یک توزیع آماری به شرح زیر است:

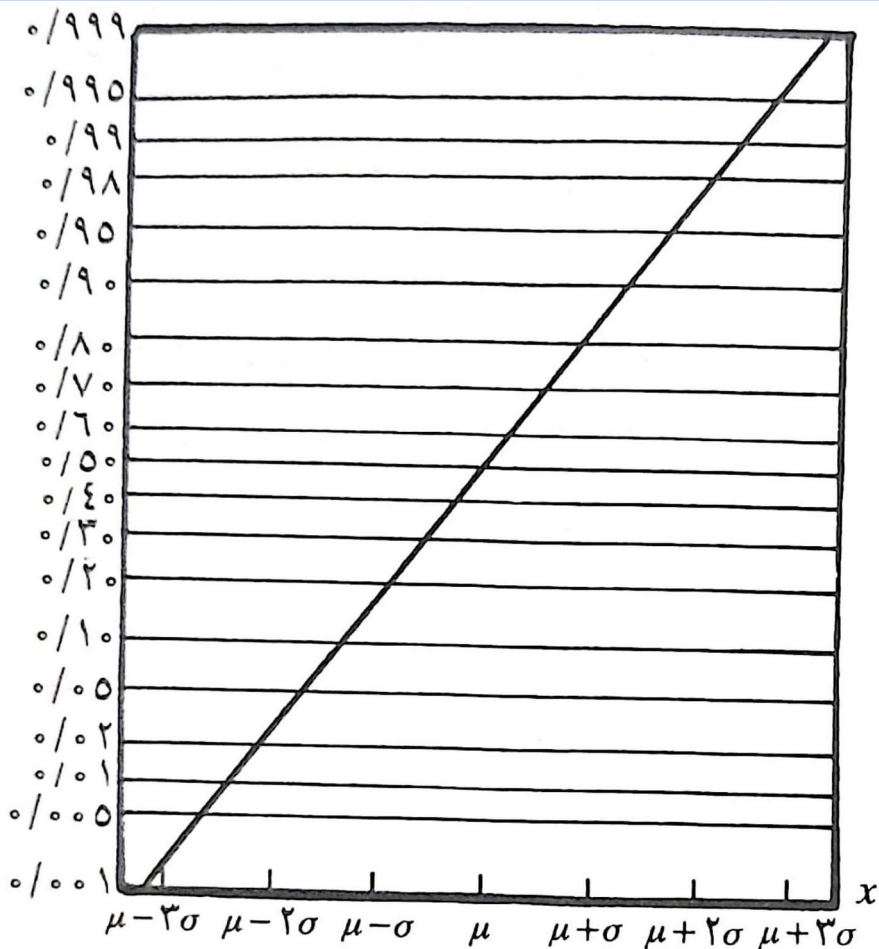
صفحه مخصوصی به نام «صفحه احتمال» مطابق نمونه زیر با

تنظیم مقیاس محور عمودی طوری رسم می‌کنیم که احتمال‌های تجمع

بر روی آن به صورت خطی راست نمایش داده شوند و در محور افقی

نیز مقیاس را طوری انتخاب می‌کنیم که شامل تمام مشاهدات ما بشود

(حداقل 15 تا 20 مشاهده در این روش مورد نیاز است).





روشی برای تشخیص نرمال بودن توزیع

مراحل رسم صفحه احتمال برای داده‌های طبقه‌بندی نشده:

(1) n مشاهده را از کوچک به بزرگ مرتب کنید.

(2) مقیاس روی محور افقی (x ها) را به منظور نمایش تمام مشاهدات تنظیم کنید.

(3) بر اساس مشاهده مرتب شده نام در محور افقی ($i = 1, 2, 3, \dots, n$) و فراوانی نسبی اصلاح شده $\frac{i - 0.5}{n}$ در محور عمودی، نقاط را روی کاغذ مشخص کنید.

(4) نمودار حاصل را با خط مستقیمی که از منحنی نرمال به دست آمده، مقایسه کنید. وجود انحراف‌های منظم، نشان‌دهنده نرمال نبودن توزیع است.

البته به جای فراوانی نسبی اصلاح شده $\frac{i - 0.5}{n}$ ، می‌توان از فراوانی نسبی تجمعی معمولی $\frac{i}{n}$ نیز استفاده کرد. در این حالت بزرگترین مشاهده که فراوانی نسبی تجمعی آن یک است، در نمودار قرار نمی‌گیرد.

در صورتی که داده‌ها طبقه‌بندی شده باشند، می‌توان حد بالای طبقات و فراوانی نسبی تجمعی را ملاک رسم نمودار در صفحه احتمال قرار داد.



روشی برای تشخیص نرمال بودن توزیع

مثال: برای مشخص شدن توزیع تعداد ضایعات روزانه یک دستگاه، 20 روز را به طور تصادفی انتخاب و تعداد ضایعات را به قرار 10، 14، 13، 18، 12، 15، 11، 17، 9، 13، 14، 8، 12، 5، 7، 14، 13، 10، 15، 8 ثبت کرده‌ایم. آیا توزیع تعداد ضایعات، از توزیع نرمال تبعیت می‌کند؟

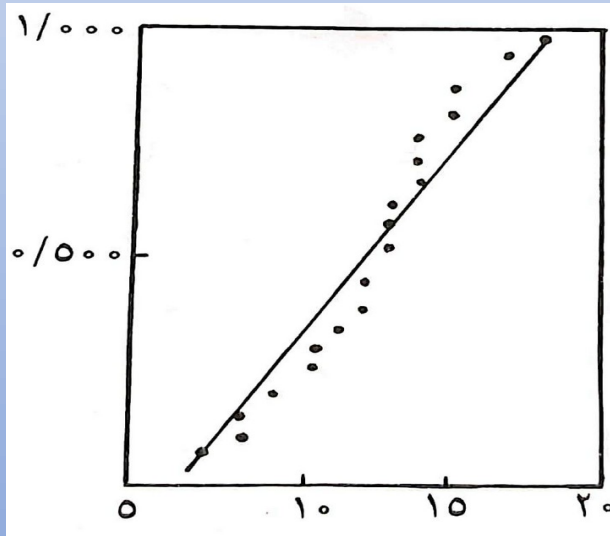
پاسخ: ابتدا داده‌ها را مرتب می‌کنیم: 10، 10، 11، 12، 12، 13، 13، 13، 14، 14، 14، 15، 15، 17، 18

$$5, 7, 8, 8, 9 \quad \text{سپس نقاط} \quad \left(5, \frac{1-0.5}{20}\right) = (5, 0.025), \left(7, \frac{2-0.5}{20}\right) = (7, 0.075), \dots, \left(18, \frac{20-0.5}{20}\right) = (18, 0.975)$$

را در صفحه احتمال رسم می‌کنیم.

با توجه به شکل مقابل و اینکه نقاط به خط نرمال نزدیک هستند، نتیجه می‌گیریم توزیع مشاهدات به توزیع نرمال نزدیک است.

یا می‌توان از توزیع نرمال به عنوان یک تقریب برای توزیع تعداد ضایعات استفاده





تقریب توزیع دوجمله‌ای به وسیله توزیع نرمال

در صورتی که مقدار n در یک توزیع دوجمله‌ای بزرگ باشد، محاسبه احتمالات نسبتاً دشوار خواهد بود. همانطور که پیش‌تر اشاره شد اگر مقدار p (احتمال موفقیت در هر آزمایش مستقل) کوچک باشد، می‌توان توزیع دوجمله‌ای را به شرطی که حاصلضرب np کوچکتر از 10 باشد، با توزیع پواسون $P(\lambda=np)$ تقریب زد. اما اگر مقدار p خیلی به صفر **(یا یک)** نزدیک نباشد، برای تقریب توزیع دوجمله‌ای می‌توان از توزیع نرمال استفاده کرد.

نکته بسیار مهم در استفاده از یک متغیر تصادفی پیوسته به عنوان تقریبی از یک متغیر تصادفی گسسته این

$P(X = x) \rightarrow P(x - 0.5 \leq X \leq x + 0.5)$ $P(X \leq x) \rightarrow P(X \leq x + 0.5)$ $P(X < x) \rightarrow P(X \leq x - 0.5)$	$P(X \geq x) \rightarrow P(X \geq x - 0.5)$ $P(X > x) \rightarrow P(X \geq x + 0.5)$ $P(x_1 \leq X \leq x_2) \rightarrow P(x_1 - 0.5 \leq X \leq x_2 + 0.5)$
--	--



تقریب توزیع دوجمله‌ای به وسیله توزیع نرمال

مثال: 65 درصد از کل افرادی که به یک فروشگاه مراجعه می‌کنند، خرید می‌کنند. اگر در یک روز 30 نفر به این فروشگاه مراجعه کنند، احتمالات زیر را محاسبه کنید.

$$B(n=30, p=0.65) \rightarrow N(\mu=np=19.5, \sigma^2=npq=6.825)$$

الف) حداقل 22 نفر خرید کنند.

$$P(X \geq 22) = P(X \geq 21.5) = P\left(Z \geq \frac{21.5 - 19.5}{\sqrt{6.825}}\right) = P(Z \geq 0.77) = 1 - P(Z \leq 0.77) = 0.2206$$

ب) کمتر از 15 نفر خرید کنند.

$$P(X < 15) = P(X \leq 14.5) = P\left(Z \leq \frac{14.5 - 19.5}{\sqrt{6.825}}\right) = P(Z \leq -1.92) = 0.0274$$

ج) دقیقاً 20 نفر خرید کنند.

$$P(X = 20) = P(19.5 \leq X \leq 20.5) = P\left(\frac{19.5 - 19.5}{\sqrt{6.825}} \leq Z \leq \frac{20.5 - 19.5}{\sqrt{6.825}}\right) = P(0 \leq Z \leq 0.38) = P(Z \leq 0.38) - P(Z \leq 0) = 0.1480$$



تقریب توزیع پواسون به وسیله توزیع نرمال

در حالتی که میانگین توزیع پواسون نسبتاً بزرگ شود، می‌توان از توزیع نرمال به عنوان یک تقریب مناسب برای آن استفاده کرد. به طور کلی در حالتی که $\lambda \geq 10$ باشد، تقریب نرمال تقریب خوبی برای توزیع پواسون است.
 $\lambda \geq 10: Po(\lambda) \rightarrow N(\mu = \lambda, \sigma^2 = \lambda)$

مثال: تعداد تلفن‌هایی که به روابط عمومی یک سازمان زده می‌شود، دارای توزیع پواسون با میانگین 10 تماس در هر دقیقه می‌باشد. هریک از احتمالات زیر را محاسبه کنید.
 $\lambda = 10: Po(10) \rightarrow N(\mu = 10, \sigma^2 = 10)$

الف) در یک دقیقه خاص، کمتر از 8 تماس برقرار شود
 $P(X < 8) = P(X \leq 7.5) = P(Z \leq \frac{7.5 - 10}{\sqrt{10}}) = P(Z \leq -0.79) = 0.2148$

$\lambda = 10 \times 2.5 = 25: Po(25) \rightarrow N(\mu = 25, \sigma^2 = 25)$

ب) در 2/5 دقیقه بین 20 تا 30 تماس برقرار شود
 $P(20 \leq X \leq 30) = P(19.5 \leq X \leq 30.5) = P(\frac{19.5 - 25}{\sqrt{25}} \leq Z \leq \frac{30.5 - 25}{\sqrt{25}}) = P(-1.1 \leq Z \leq 1.1) = 0.7286$



فصل هجدهم

فنون تصمیم‌گیری آماری

Statistical Decision Making Techniques



شش گام در نظریه تصمیم



انواع شرایط تصمیم‌گیری



تصمیم‌گیری در شرایط عدم اطمینان



تصمیم‌گیری در شرایط ریسک



ارزش مورد انتظار اطلاعات کامل



استفاده از اطلاعات نمونه (قضیه بیز)