به نام خدا

عنوان درس: آمار و احتمال (4 واحد)
منبع مورد تدریس:

مفاهیم و روش‌های آماری

جلد اول

تالیف گوری ک. باتاچاریا
ریچارد ا. جانسون

ترجمه مرتضی ابن شهرآشوب
فتاح میکائیلی

تهیه کننده:

محمدهادی خسروی راد

فصل اول: مقدمه

فصل دوم: مطالعه توصیفی داده‌ها

فصل سوم: مبادی احتمال

فصل چهارم: متغیرهای تصادفی و توزیع‌های احتمال

فصل پنجم: توزیع‌های متغیرهای تصادفی گسسته

فصل ششم: مفاهیم اساسی آزمون فرض‌ها

فصل هفتم: توزیع نرمال و نمونه های تصادفی

فصل ۱

مقدمه

آمار چیست؟

آمار در زبان انگلیسی Statistics و از کلمه لاتین Status گرفته شده است.

آمار به عنوان یک موضوع علمی، امروزه شامل مفاهیم و روش‌هایی است که در تمام پژوهش‌هایی‌که مستلزم جمع‌آوری داده‌ها به وسیله‌ی یک فرآیند آزمایش و مشاهده، و انجام استنباط و نتیجه‌گیری به وسیله تجزیه و تحلیل این داده‌ها هستند، اهمیت بسیار دارند.

مراحل یک تحقیق علمی

مشخص کردن هدف

جمع آوری اطلاعات

تجزیه و تحلیل داده‌ها

بیـــان یافتـــه‌هـــا

جامعه:

جامعه(آماری)، عبارت است از مجموعه‌ی کامل اندازه‌های ممکن یا اطلاعات ثبت شده از یک صفت کیفی، در مورد گردآورده‌ی کامل واحدها، که می‌خواهیم استنباط‌هایی راجع به آن انجام دهیم. جامعه، آماج تحقیق است، و منظور از عمل گردآوری داده‌ها استحراج نتایج درباره‌ی جامعه می‌باشد.

نمونه:

نمونه‌ای از جامعه‌ی آماری، مجموعه‌ی اندازه‌هایی است که عملاً در جریان یک تحقیق گردآوری می‌شود.

هدف های اصلی آمار عبارتند از:

الف) انجام استنباط درباره‌ی جامعه، از طریق تجزیه و تحلیل اطلاعات موجود در داده‌های نمونه‌ای.

ب) سنجش میزان عدم حتمیتی که در این استنباط‌ها وجود دارد.

عملی که برای رسیدن به هدفهای فوق اهمیت دارد، عبارت است از طرح‌ریزی فرآیند و دامنه‌ی نمونه‌گیری به طوری که مشاهدات، مبنایی برای استخراج استنباط‌های معتبر تشکیل دهند.

فصل ۲

مطالعه توصیفی داده‌ها

مراحل اساسی توصیف مجموعه‌ی داده ها

الف) خلاصه کردن و توصیف الگوی کلی داده ها به وسیله‌ی:

۱) ارائه جداول و نمودارها

۲) بررسی کلی نمودار داده‌ها از لحاظ خصوصیات مهم، از جمله تقارن یا انحراف از تقارن.

۳) بررسی اجمالی نمودار داده‌ها برای ملاحظه‌ی مشاهدات غیرمنتظره‌ای که به نظر می‌رسد از توده‌ی اصلی داده‌ها دورند.

ب) محاسبه معیارهای عددی برای به دست آوردن:

۱) یک مقدار نوعی یا معرف ، که مرکز داده‌ها را نشان دهد.

۲) مقدار پراکندگی داده‌ها.

توصیف داده‌ها به وسیله نمودارها و جدول

عمدتاً دو نوع نمودار برای نشان دادن داده‌ها به کار می‌روند که عبارت است از:

- نمودار نقطه‌ای
- بافت‌نگار

نمودار نقطه‌ای

وقتی تعداد داده‌ها کم است، از نمودار نقطه‌ای استفاده می‌شود.

روش ترسیم نمودار نقطه‌ای:

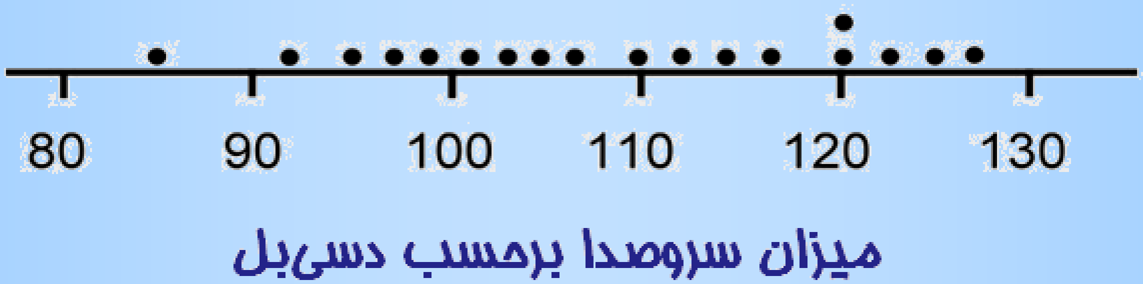
ابتدا خطی را رسم کرده و روی آن مقیاسی که حوزه اندازه‌ها را در بر گیرد، در نظر گرفته و هر یک از اندازه‌ها بر روی این خط به صورت نقطه‌ی پر رنگی مشخص می‌کنیم.

مثال: برای بررسی مقررات جدیدی که در مورد حد مجاز سر و صدا وضع شده، ۱۸ ماشین چمن‌زنی مورد مشاهده قرار گرفته‌اند و میزان صدای آنها برحسب نزدیکترین دسی‌بل (واحد اندازه‌گیری میزان صدا) ثبت گردیده است.

داده‌ها:

۹۵	۱۲۰	۱۱۷	۹۹	۱۱۰	۱۰۷	۱۲۵	۹۸	۸۵
۱۲۷	۱۰۵	۱۱۴	۱۰۳	۱۱۲	۹۲	۱۰۱	۱۲۲	۱۲۰

نمودار نقطه‌ای:



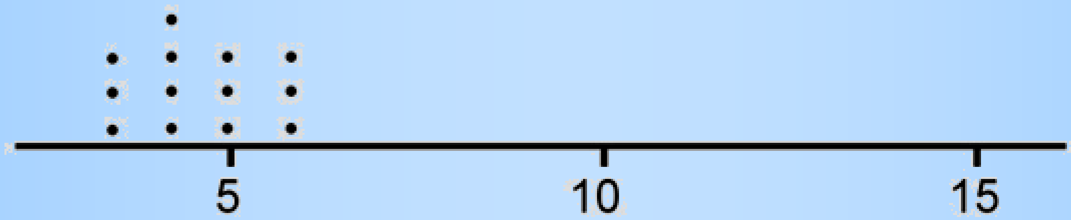
این نمودار نقطه‌ای توزیع نسبتاً یکنواختی را نشان می‌دهد که مرکز آن بین ۱۰۰ و ۱۱۰ است.

مثال: شخصی که منتظر فرصتی برای عبور از عرض یک خیابان بود متوجه شد ماشین‌هایی که در یک جهت حرکت می کنند با فواصل زمانی زیر بر حسب ثانیه، از مقابل او می‌گذرند:

داده ها: ۶ ۳ ۵ ۶ ۴ ۳ ۵ ۴ ۶ ۳ ۴ ۵ ۴ ۱۸

نمودار نقطه‌ای آن را رسم کنید.

داده ها: ۶ ۳ ۵ ۶ ۴ ۳ ۵ ۴ ۶ ۳ ۴ ۵ ۴ ۱۸



نمودار نقطه‌ای فاصله‌های زمانی برحسب ثانیه

نمودار نشان می‌دهد که زمانها در اطراف مقدار مرکزی ۴ یا ۵ جمع شده‌اند بجز مشاهده ۱۸ که می‌توان تشخیص داد که چراغ قرمز چهارراه قبلی باعث بوجود آمدن این مشاهده شده است.

توزیع فراوانی

مراحل تشکیل توزیع فراوانی:

الف) پیدا کردن مقدار کمینه و بیشینه در مجموعه داده‌ها.

ب) انتخاب تعدادی زیر فاصله یا خانه‌هایی به طول مساوی، بطوری که دامنه مقادیر بین بیشینه و کمینه را بدون داشتن فصل مشترک در بر گیرند. هر یک از این زیر فاصله‌ها را رده و دوسر هر رده را مرزهای آن رده می‌نامند.

ج) شمارش تعداد مشاهدات موجود در هر رده. هر یک از اعداد حاصل از شمارش ها را فراوانی رده‌ای یا فراوانی خانه‌ای می‌نامند.

د) تعیین فراوانی نسبی هر رده با تقسیم فراوانی رده ای آن رده به تعداد کل مشاهدات:

$$\text{فراوانی نسبی یک رده} = \frac{\text{فراوانی رده ای}}{\text{تعداد کل مشاهدات}}$$

فراوانی نسبی یک رده: نسبت تعداد مشاهدات موجود در آن رده به تعداد کل مشاهدات.

مثال:

داده‌ها: اندازه تمرکز ازن در نواحی جنوب شهر در طول تابستان‌های ۱۳۴۵-۱۳۴۸.

۵/۶	۵/۳	۴/۴	۴/۲	۶/۰	۶/۶	۱/۴	۳/۵
۴/۰	۳/۵	۴/۷	۵/۴	۴/۴	۵/۴	۲/۵	۶/۸
۳/۴	۴/۱	۳/۰	۶/۵	۴/۷	۵/۶	۳/۰	۲/۴
۱۱/۷	۶/۷	۶/۰	۷/۴	۴/۷	۵/۳	۱/۷	۶/۸
۶/۶	۳/۹	۱/۴	۵/۵	۵/۶	۵/۱	۱/۱	۵/۵
۴/۱	۶/۱	۲/۸	۵/۸	۶/۰	۶/۲	۷/۵	۶/۲
۶/۶	۸/۱	۲/۵	۱/۶	۵/۸	۳/۱	۵/۸	۵/۷
۳/۷	۲/۰	۳/۷	۱/۴	۷/۶	۵/۸	۳/۴	۹/۴
۷/۶	۶/۲	۳/۳	۵/۹	۳/۸	۴/۷	۳/۱	۶/۸
		۹/۴	۳/۷	۴/۵	۵/۷	۴/۴	۶/۶

توزیع فراوانی داده های تمرکز ازن

<table>
<tr><th>رده ها</th><th>خط نشانه</th><th>فراوانی</th><th>فراوانی نسبی</th></tr>
<tr><td>۱/۰۵-۱/۵۵</td><td>||||</td><td>۴</td><td>۰/۰۵۱</td></tr>
<tr><td>۱/۵۵-۲/۰۵</td><td>|||</td><td>۳</td><td>۰/۰۳۸</td></tr>
<tr><td>۲/۰۵-۲/۵۵</td><td>|||</td><td>۳</td><td>۰/۰۳۸</td></tr>
<tr><td>۲/۵۵-۳/۰۵</td><td>|||</td><td>۳</td><td>۰/۰۳۸</td></tr>
<tr><td>۳/۰۵-۳/۵۵</td><td>|| 𝍸</td><td>۷</td><td>۰/۰۹۰</td></tr>
<tr><td>۳/۵۵-۴/۰۵</td><td>| 𝍸</td><td>۶</td><td>۰/۰۷۷</td></tr>
<tr><td>۴/۰۵-۴/۵۵</td><td>|| 𝍸</td><td>۷</td><td>۰/۰۹۰</td></tr>
<tr><td>۴/۵۵-۵/۰۵</td><td>||||</td><td>۴</td><td>۰/۰۵۱</td></tr>
<tr><td>۵/۰۵-۵/۵۵</td><td>|| 𝍸</td><td>۷</td><td>۰/۰۹۰</td></tr>
<tr><td>۵/۵۵-۶/۰۵</td><td>||| 𝍸 𝍸</td><td>۱۳</td><td>۰/۱۶۷</td></tr>
<tr><td>۶/۰۵-۶/۵۵</td><td>𝍸</td><td>۵</td><td>۰/۰۶۴</td></tr>
<tr><td>۶/۵۵-۷/۰۵</td><td>||| 𝍸</td><td>۸</td><td>۰/۱۰۳</td></tr>
<tr><td>۷/۰۵-۷/۵۵</td><td>||</td><td>۲</td><td>۰/۰۲۶</td></tr>
<tr><td>۷/۰۵-۸/۵۵</td><td>||</td><td>۲</td><td>۰/۰۲۶</td></tr>
<tr><td>۸/۵۵-۸/۰۵</td><td>|</td><td>۱</td><td>۰/۰۱۳</td></tr>
<tr><td>۸/۰۵-۹/۵۵</td><td></td><td>۰</td><td></td></tr>
<tr><td>۹/۰۵-۹/۵۵</td><td>||</td><td>۲</td><td>۰/۰۲۶</td></tr>
<tr><td>۹/۵۵-۱۰/۰۵</td><td></td><td>۰</td><td></td></tr>
<tr><td>۱۰/۰۵-۱۰/۵۵</td><td></td><td>۰</td><td></td></tr>
<tr><td>۱۰/۰۵-۱۱/۵۵</td><td></td><td>۰</td><td></td></tr>
<tr><td>۱۱/۵۵-۱۱/۰۵</td><td></td><td>۰</td><td></td></tr>
<tr><td>۱۱/۰۵-۱۲/۵۵</td><td>|</td><td>۱</td><td>۰/۰۱۳</td></tr>
<tr><td>مجموع</td><td></td><td>۷۸</td><td>۱/۰۰۱</td></tr>
</table>

توزیع فراوانی داده های تمرکز ازن طول ۱

فراوانی نسبی	فراوانی	رده ها
۰/۰۹۰	۷	۱/۰۵-۲/۰۵
۰/۰۷۷	۶	۲/۰۵-۳/۰۵
۰/۱۶۷	۱۳	۳/۰۵-۴/۰۵
۰/۱۴۱	۱۱	۴/۰۵-۵/۰۵
۰/۲۵۶	۲۰	۵/۰۵-۶/۰۵
۰/۱۶۷	۱۳	۶/۰۵-۷/۰۵
۰/۰۵۱	۴	۷/۰۵-۸/۰۵
۰/۰۱۳	۱	۸/۰۵-۹/۰۵
۰/۰۲۶	۲	۹/۰۵-۱۰/۰۵
	۰	۱۰/۰۵-۱۱/۰۵
۰/۰۱۳	۱	۱۱/۰۵-۱۲/۰۵
۱/۰۰۱	۷۸	مجموع

توزیع فراوانی داده های تمرکز ازن طول ۲

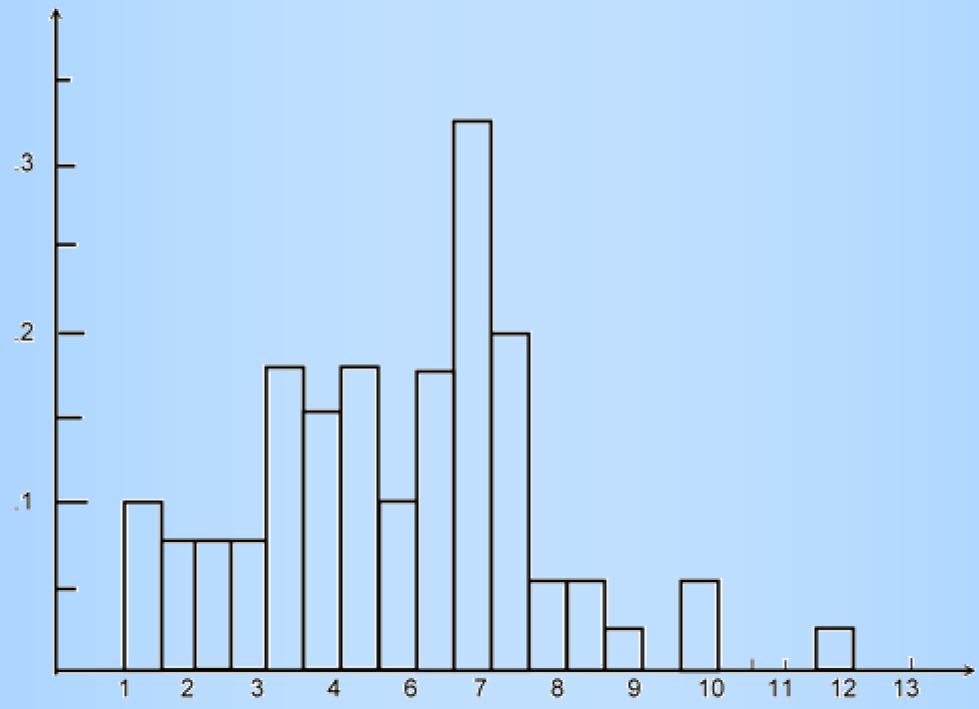
رده ها	فراوانی	فراوانی نسبی
۰/۰۵-۲/۰۵	۷	۰/۰۹۰
۲/۰۵-۴/۰۵	۱۹	۰/۲۴۴
۴/۰۵-۶/۰۵	۳۱	۰/۳۹۷
۶/۰۵-۸/۰۵	۱۷	۰/۲۱۸
۸/۰۵-۱۰/۰۵	۳	۰/۰۳۸
۱۰/۰۵-۱۲/۰۵	۱	۰/۰۱۳
مجموع	۷۸	۱/۰۰۰

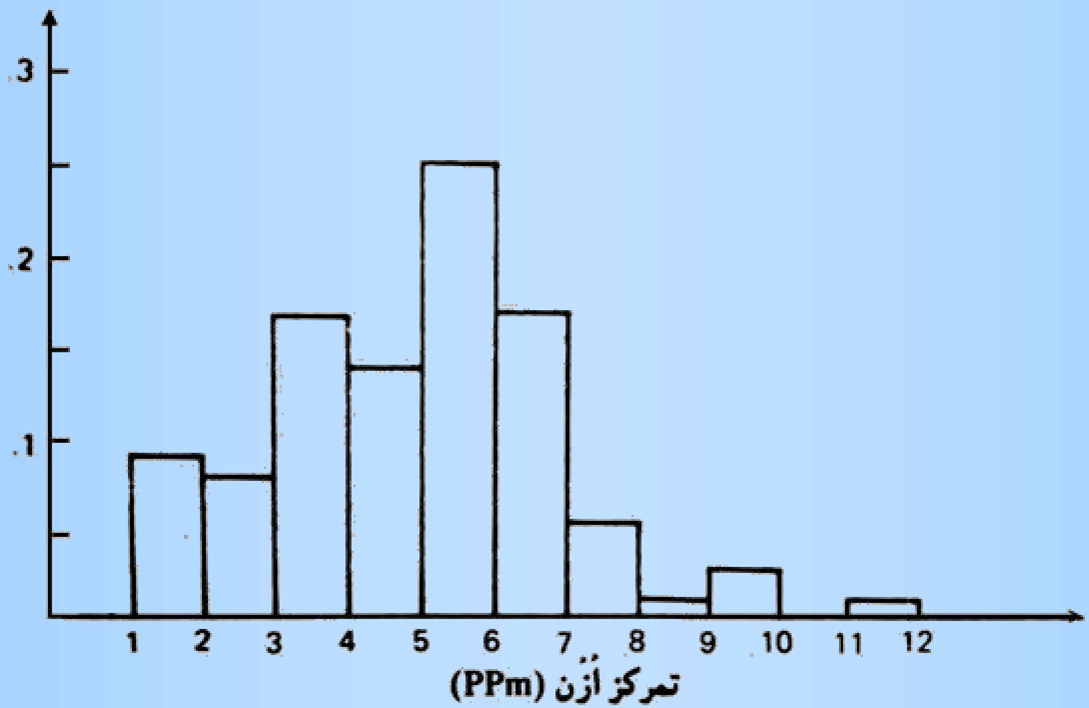
بافت نگار فراوانی نسبی

برای رسم بافت نگار فراوانی نسبی رده ها را روی محـــور افقی نمودار مشخص می کنیم. آنگـــاه روی هــر رده ، مستطیلی همودی رسم می کنیم که مساحت آن مساوی با فراوانی نسبی آن رده باشد.

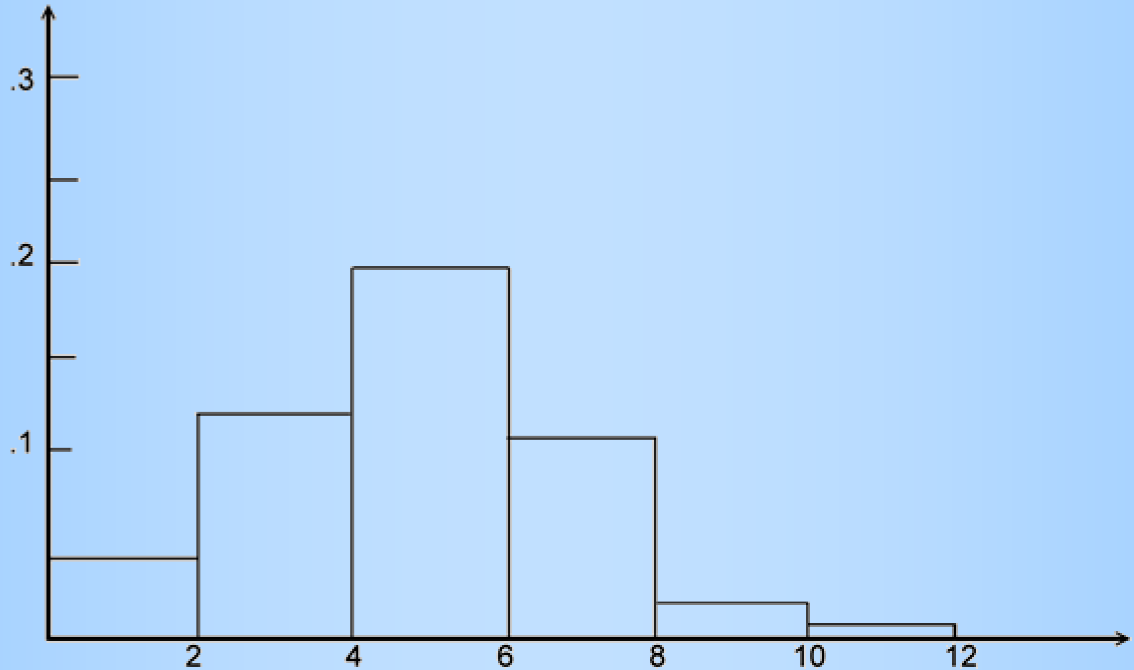
$$\text{ارتفاع مستطیل} = \frac{\text{فراوانی نسبی رده}}{\text{طول رده}}$$

بافت‌نگار فراوانی نسبی برای داده‌های تمرکز ازن(PPm)

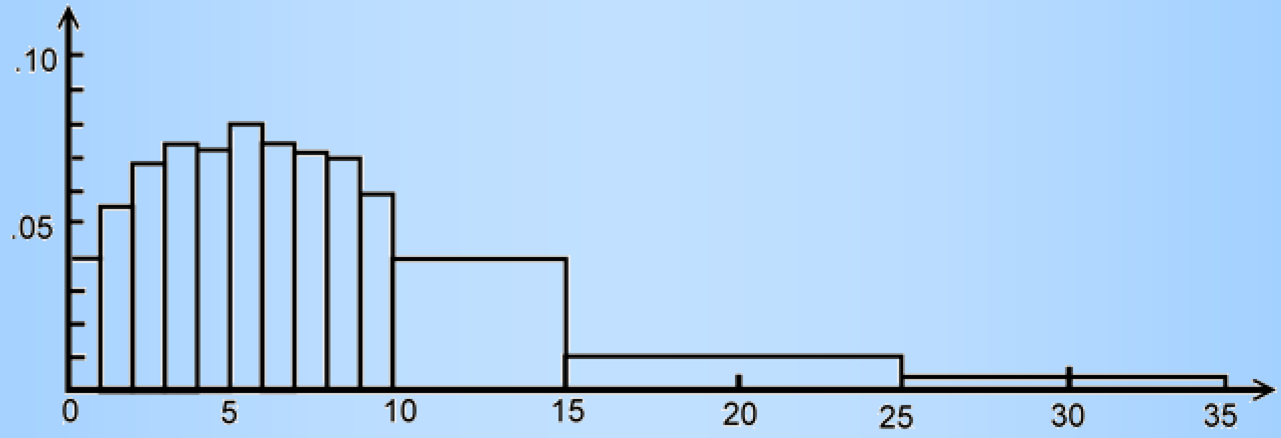




.3
.2
.1
1 2 3 4 5 6 7 8 9 10 11 12
تمرکز أزن (PPm)



تمرکز اُزُن(PPm)



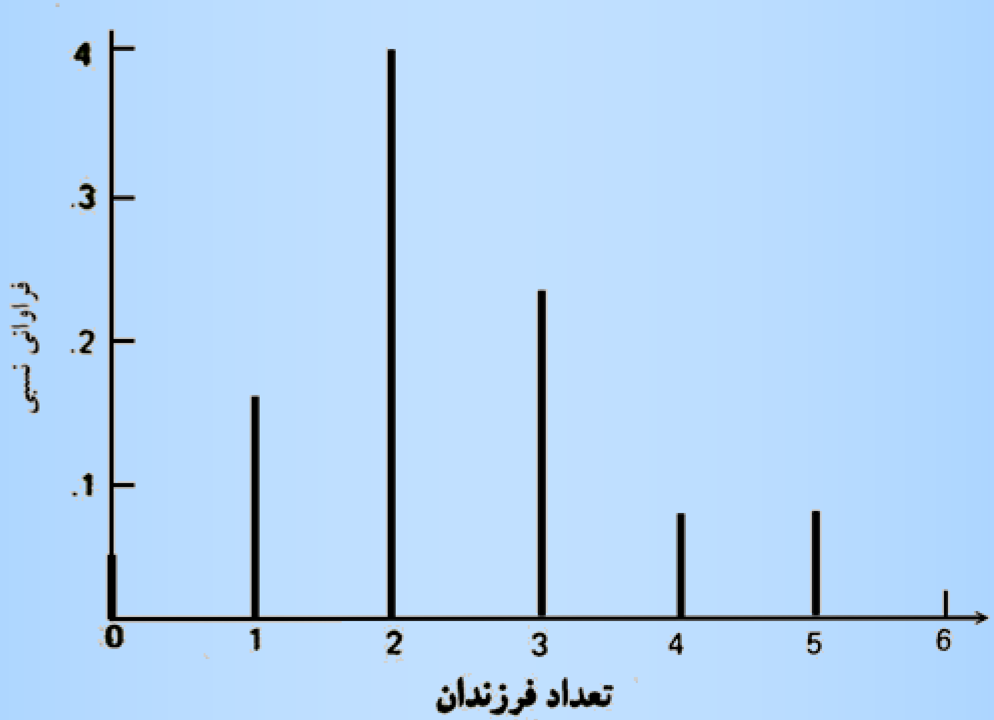
درآمد برحسب هزار دلار

مثال: توزیع فراوانی درآمد ۲۶۴۳۲۲۶ خانوار سفیدپوست مناطق روستایی آمریکا در سال ۱۳۴۹

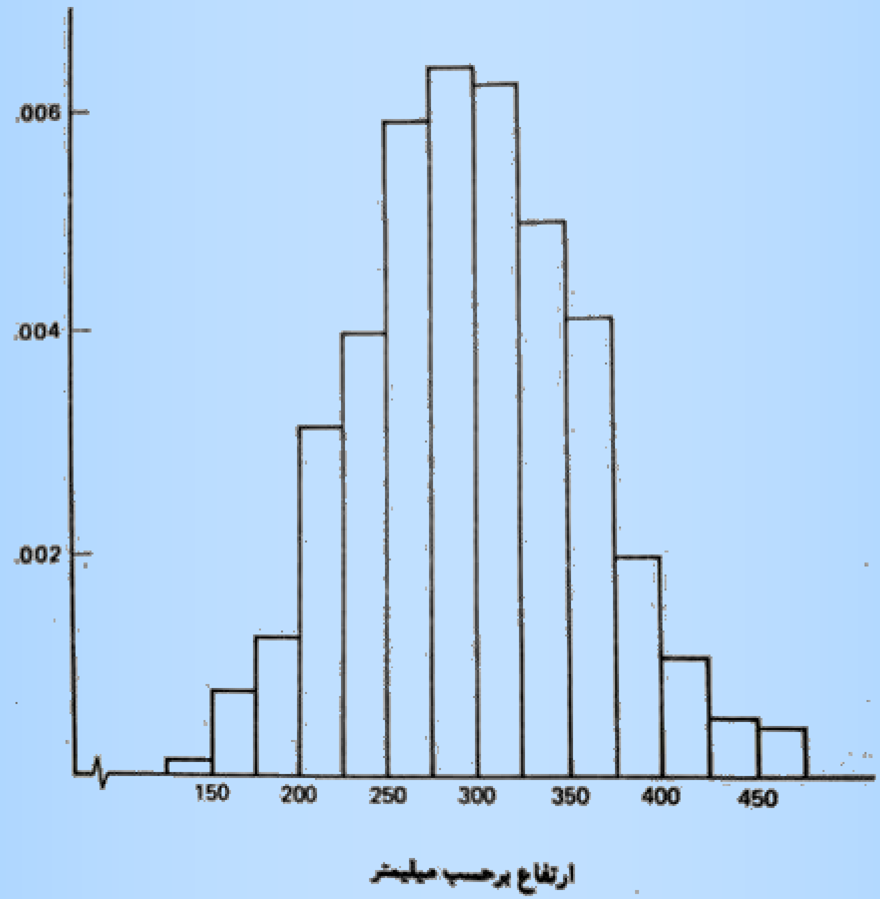
درآمد خانوار (برحسب هزار دلار)	فراوانی نسبی (٪)
۰−۱	۴/۰
۱−۲	۵/۴
۲−۳	۶/۸
۳−۴	۷/۳
۴−۵	۷/۲
۵−۶	۸/۰
۶−۷	۷/۴
۷−۸	۷/۲
۸−۹	۷/۰
۹−۱۰	۵/۹
۱۰−۱۵	۱۹/۹
۱۵−۲۵	۱۰/۳
۲۵ به بالا	۳/۵
مجموع	۹۹/۹

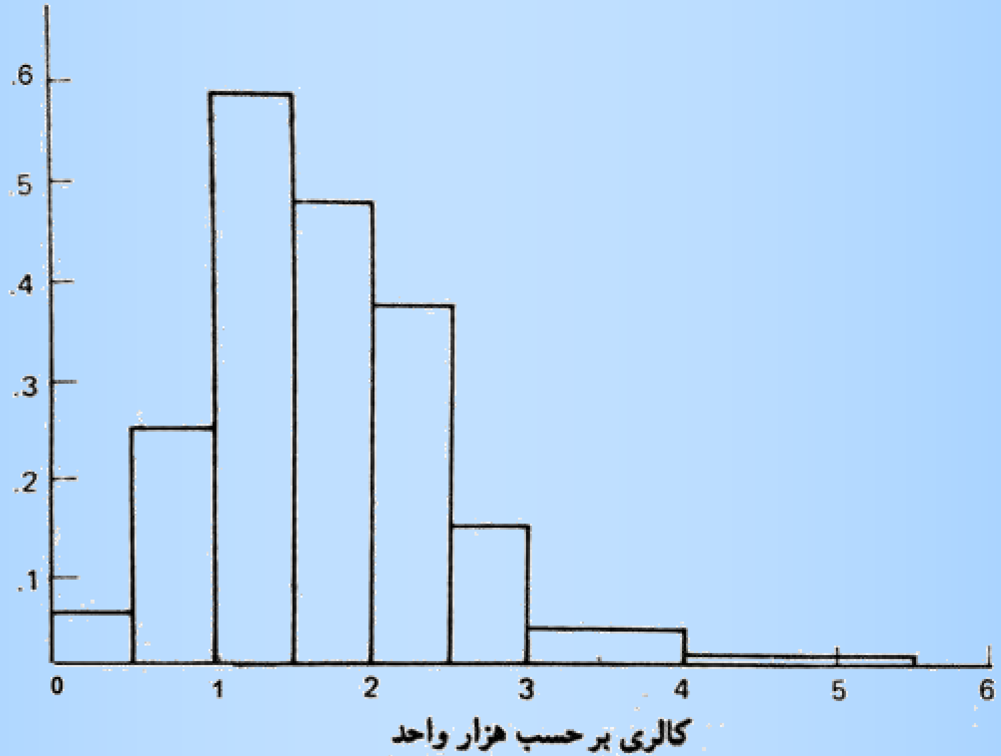
نمودار خطی فراوانی نسبی

مقادیر متمایز را به صــورت نقــاطی روی محــور افقــی مشخص می‌کنیم و سپس از نقاط حاصل، خطهایی عمــود بر محور رسم می‌کنیم که ارتفاع هر یک برابر با فراوانــی نسبی مقدار مربوطه باشد.



4
.3
.2
.1
فراوانی نسبی
0
1
2
3
4
5
6
تعداد فرزندان


.006
.004
.002
150
200
250
300
350
400
450
ارتفاع برحسب میلیمتر

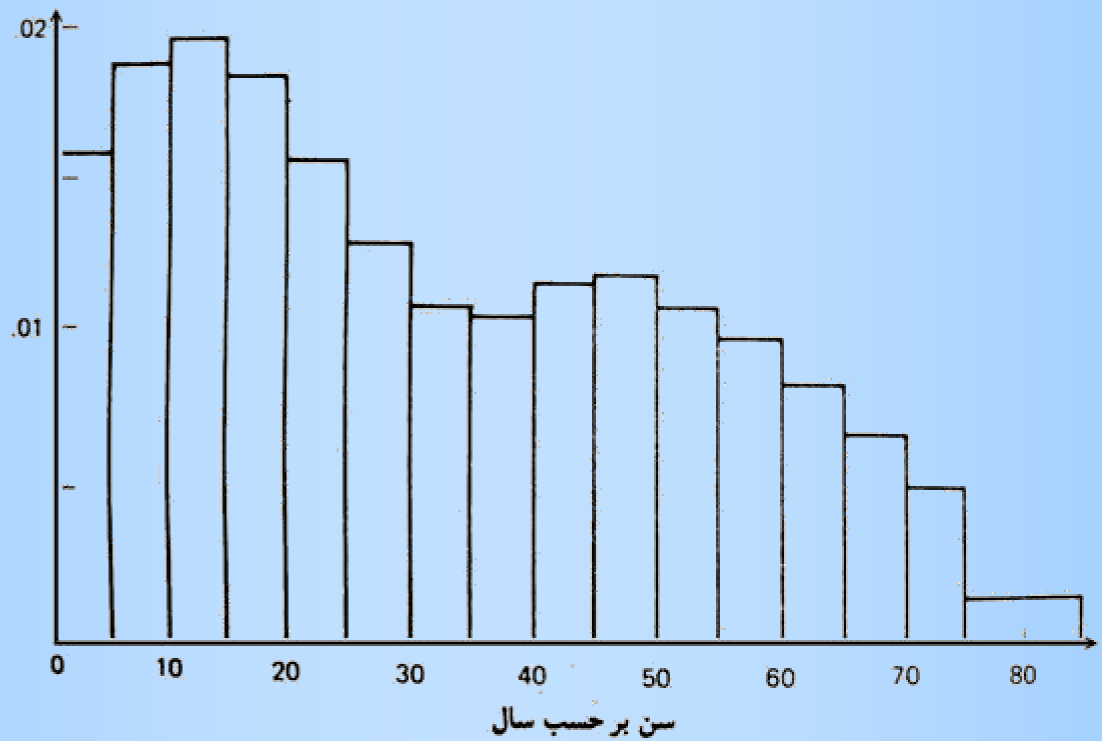

.6
.5
.4
.3
.2
.1
0
1
2
3
4
5
6
کالری بر حسب هزار واحد

مثال:

داده‌ها:

فراوانی تعداد فرزندان ۲۵ خانواده

تعداد فرزندان	فراوانی	فراوانی نسبی
۰	۱	۰/۰۴
۱	۴	۰/۰۱۶
۲	۱۰	۰/۴۰
۳	۶	۰/۲۴
۴	۲	۰/۰۸
۵	۲	۰/۰۸
مجموع	۲۵	۱/۰۰۰


.02
.01
0
10
20
30
40
50
60
70
80
سن بر حسب سال

نمادگذاری مجموعه‌های داده‌ها و عمل جمع

نماد جمع ($\sum$)

نماد $\sum_{i=1}^{n} x_i$ نشان دهنده‌ی مجمــوع n عــدد x_1 ، x_2 ، ... ، x_n است و این نماد به این صورت خوانده می‌شود:

مجموع تمام x_iها ، که i از ۱ تا n تغییر می‌کند.

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

مثال: فرض کنید: $x_1=3$ ، $x_2=5$ ، $x_3=4$ ، $x_4=3$

$$\sum_{i=1}^{4} x_i = x_1+x_2+x_3+x_4=3+5+4+3=15$$

$$\sum_{i=1}^{4} 3x_i = x_1+x_2+x_3+x_4=3+5+4+3=15$$

$$\sum_{i=۱}^{۴}(x_i-۲)=(x_۱-۲)+(x_۲-۲)+(x_۳-۲)+(x_۴-۲)$$

$$=\sum_{i=۱}^{۴}x_i-۴(۲)=۱۵-۸=۷$$

$$\sum_{i=۱}^{۴}(x_i-۲)^۲=(x_۱-۲)^۲+(x_۲-۲)^۲+(x_۳-۲)^۲+(x_۴-۲)^۲$$

$$=(۳-۲)^۲+(۵-۲)^۲+(۴-۲)^۲+(۳-۲)^۲$$

$$=۱+۹+۴+۱=۱۵$$

بعضی از خواص اصلی جمع

اگر a و b اعداد ثابتی باشند ، داریم

$$\sum_{i=1}^{n} bx_i = b\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} (bx_i + a) = b\sum_{i=1}^{n} x_i + na$$

$$\sum_{i=1}^{n} (x_i - a) \;= \sum_{i=1}^{n} x_i \;- ۲\,a\sum_{i=1}^{n} x_i + na^{۲}$$

معیارهای گرایش به مرکز

هر معیار عددی را که معرف مرکز مجموعه داده‌ها باشد، معیار گرایش به مرکز گویند.

متداول‌ترین معیارهای گرایش به مرکز عبارتند از:

میانگین

میانه

میانگین یا متوسط نمونه‌ای مرکب از n اندازه‌ی x_1، x_2، ...، x_n، عبارت است از خارج قسمت مجموع این اندازه ها بر n. میانگین را با $\bar{x}$ نشان می‌دهند که در عمیلات، به صورت زیر نوشته می‌شود:

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

میانه نمونه‌ای مرکب از n انــدازه x_1، x_2، . . .، x_n عبــارت است، از اندازه‌ی وسطی، در صورتی که اندازه‌ها را بترتیــب از کوچکترین به بزرگترین مقدار مرتب کرده باشیم. اگــر n عدد فردی باشد، یک مقدار وسطی منحصر بفرد وجود دارد که میانه است. اگر n زوج باشد دو مقــدار وســطی وجــود دارند که متوسط آنها به عنوان میانه تعریف می‌شود.

مثال: میانه داده‌های مربوط به وزن نوزادان را بدست آورید.

داده‌ها: ۳/۵۴، ۳/۶۷، ۴/۷۶، ۲/۹، ۴/۱۷

مرتب کردن از کوچک به بزرگ

۴/۷۶ ۴/۱۷ ۳/۶۷ ۳/۵۴ ۲/۹۰

مثال: درآمد ماهانه ی ۸ کارمند رسمی یک شرکت به تومان عبارت است از :

۵۰۰۰، ۷۵۰۰، ۶۰۰۰، ۵۵۰۰، ۷۰۰۰، ۲۰۰۰۰ ، ۵۵۰۰

میانگین و میانه ی درآمدها را محاسبه کنید.

ابتدا داده‌ها را از کوچک به بزرگ مرتب می‌کنیم

۵۰۰۰ ۵۵۰۰ ۵۵۰۰ $\boxed{۵۵۰۰ \quad ۶۰۰۰}$ ۷۰۰۰ ۷۵۰۰ ۲۰۰۰۰

میانگین:

$$\bar{x} = \frac{۶۲۰۰۰}{۸}$$ =۷۷۵۰ تومان

میانه :

$$\frac{۵۵۰۰+۶۰۰۰}{۲}$$ =۵۷۵۰ تومان

صدک(۱۰۰p)ام نمونه، مقداری است که وقتی داده ها از کوچکترین تا بزرگترین مقدار مرتــب شــدند ، حــداقل ۱۰۰p٪ از مشاهدات منطبق بر این مقــدار یــا در ســمت چپ(زیر) آن و حداقل (۱−p)۱۰۰٪ از مشاهدات منطبق بر این مقدار یا در سمت راست (بالای) آن باشند.

چارکهای نمونه

چارک (اول) کوچکتر	صدک ۲۵ ام= Q_1
چارک دوم (یا میانه)	صدک ۵۰ ام= Q_2
چارک (سوم) بزرگتر	صدک ۷۵ ام= Q_3

مثال: ۹ کارگر صنعتی تحت آزمون قدرت پنجه قرار گرفتند و اندازه‌های زیر به دست آمدند:

۱۲۸/۳ ۱۰۶/۵ ۱۱۶/۶ ۱۵۲/۴ ۱۲۵/۰ ۱۳۲/۱ ۱۰۵/۸ ۱۳۶/۷

چارک‌ها را برای این داده‌ها محاسبه کنید.

ابتدا داده‌ها را بترتیب از کوچکترین به بزرگترین اندازه مرتب می‌کنیم:

۱۵۲/۴ ۱۳۶/۷ ۱۳۲/۱ ۱۲۸/۳ ۱۲۵/۰ ۱۱۶/۶ $\boxed{\text{۱۰۶/۵}}$ ۱۰۵/۸ ۹۳/۹

چارک اول:

تعدادمقادیر نابیشتر از Q_1 حداقل برابر است با۰/۲۵×۹=۲/۲۵

تعدادمقادیر ناکمتر از Q_1 حداقل‌برابراست با۰/۷۵×۹=۶/۷۵

مشاهده ۱۰۶/۵ دارای چنین شرایطی است و چارک اول است.

۱۵۲/۴ ۱۳۶/۷ ۱۳۲/۱ ۱۲۸/۳ ۱۲۵/۰ ۱۱۶/۶ ۱۰۶/۵ ۱۰۵/۸ ۹۳/۹

با شمارش سه مشاهده از بزرگترین مقدار به طرف پایین،

۱۳۲/۱ چارک سوم است.

۱۵۲/۴ ۱۳۶/۷ ۱۳۲/۱ ۱۲۸/۳ ۱۲۵/۰ ۱۱۶/۶ ۱۰۶/۵ ۱۰۵/۸ ۹۳/۹

میانه یا صدک ۵۰ ام مقدار وسطی یعنی ۱۲۵/۰ است.

توجه: معمولاً چارک‌ها را در حالاتی که تعداد مشاهدات کمتر از ۲۵ است، محاسبه نمی‌کنند.

مثال: ۵۰ اندازه مربوط به سر و صدای ناشی از ترافیک در یک چهارراه، که بترتیب از کوچکترین به بزرگترین مقدار مرتب شده‌اند، در جدول زیر آمده است. چارک‌ها و همچنین صدک ۱۰ ام را محاسبه کنید.

۵۲/۰	۵۵/۹	۵۶/۷	۵۹/۴	۶۰/۲	۶۱/۰	۶۳/۱	۶۲/۸	۶۵/۷	۶۷/۹
۵۴/۴	۵۵/۹	۵۶/۸	۵۹/۴	۶۰/۳	۶۱/۴	۶۲/۶	۶۴/۰	۶۶/۲	۶۸/۲
۵۴/۵	۵۶/۲	۵۷/۲	۵۹/۵	۶۰/۵	۶۱/۷	۶۲/۷	۶۴/۶	۶۶/۸	۶۸/۹
۵۵/۷	۵۶/۴	۵۷/۶	۵۹/۸	۶۰/۶	۶۱/۸	۶۳/۱	۶۴/۸	۶۷/۰	۶۹/۴
۵۵/۸	۵۶/۴	۵۸/۹	۶۰/۰	۶۰/۸	۶۲/۰	۶۳/۶	۶۴/۹	۶۷/۱	۷۷/۱

۵۲/۰	۵۵/۹	۵۶/۷	۵۹/۴	۶۰/۲	۶۱/۰	۶۳/۱	۶۲/۸	۶۵/۷	۶۷/۹
۵۴/۴	۵۵/۹	۵۶/۸	۵۹/۴	۶۰/۳	۶۱/۴	۶۲/۶	۶۴/۰	۶۶/۲	۶۸/۲
۵۴/۵	۵۶/۲	۵۷/۲	۵۹/۵	۶۰/۵	۶۱/۷	۶۲/۷	۶۴/۶	۶۶/۸	۶۸/۹
۵۵/۷	۵۶/۴	۵۷/۶	۵۹/۸	۶۰/۶	۶۱/۸	۶۳/۱	۶۴/۸	۶۷/۰	۶۹/۴
۵۵/۸	۵۶/۴	۵۸/۹	۶۰/۰	۶۰/۸	۶۲/۰	۶۳/۶	۶۴/۹	۶۷/۱	۷۷/۱

برای چارک اول حداقل ۰/۲۵×۵۰=۱۲/۵ مشاهده را از کوچکترین اندازه به بالا و حداقل ۰/۷۵×۵۰=۳۷/۵ مشاهده را از بزرگترین مشاهده به پایین باید شمارش شود. سیزدهمین مشاهده مرتب شده یعنی ۵۷/۲، مقداری است که ۱۳ تا از مقادیر نابیشتر از آن و ۳۸ تا از آنها ناکمتر از آن هستند. پس چارک اول برابر ۵۷/۲ است.

۵۲/۰	۵۵/۹	۵۶/۷	۵۹/۴	۶۰/۲	۶۱/۰	۶۳/۱	۶۲/۸	۶۵/۷	۶۷/۹
۵۴/۴	۵۵/۹	۵۶/۸	۵۹/۴	۶۰/۳	۶۱/۴	۶۲/۶	۶۴/۰	۶۶/۲	۶۸/۲
۵۴/۵	۵۶/۲	۵۷/۲	۵۹/۵	۶۰/۵	۶۱/۷	۶۲/۷	۶۴/۶	۶۶/۸	۶۸/۹
۵۵/۷	۵۶/۴	۵۷/۶	۵۹/۸	۶۰/۶	۶۱/۸	۶۳/۱	۶۴/۸	۶۷/۰	۶۹/۴
۵۵/۸	۵۶/۴	۵۸/۹	۶۰/۰	۶۰/۸	۶۲/۰	۶۳/۶	۶۴/۹	۶۷/۱	۷۷/۱

با شمارش ۱۳مشاهده از بزرگترین اندازه به طرف پایین

چارک سوم را بدست می آوریم. یعنی ۶۴/۶

۵۲/۰	۵۵/۹	۵۶/۷	۵۹/۴	۶۰/۲	۶۱/۰	۶۳/۱	۶۲/۸	۶۵/۷	۶۷/۹
۵۴/۴	۵۵/۹	۵۶/۸	۵۹/۴	۶۰/۳	۶۱/۴	۶۲/۶	۶۴/۰	۶۶/۲	۶۸/۲
۵۴/۵	۵۶/۲	۵۷/۲	۵۹/۵	۶۰/۹		۶۲/۷	۶۴/۶	۶۶/۸	۶۸/۹
۵۵/۷	۵۶/۴	۵۷/۶	۵۹/۸	۶۰/۶	۶۱/۸	۶۳/۱	۶۴/۸	۶۷/۰	۶۹/۴
۵۵/۸	۵۶/۴	۵۸/۹	۶۰/۰	۶۰/۸	۶۲/۰	۶۳/۶	۶۴/۹	۶۷/۱	۷۷/۱

میانه عبارت است از (۶۰/۸+۶۱/۰)/۲=۶۰/۹

۵۲/۰	۵۵/۹	۵۶/۷	۵۹/۴	۶۰/۲	۶۱/۰	۶۳/۱	۶۲/۸	۶۵/۷	۶۷/۹
۵۴/۴	۵۵/۹	۵۶/۸	۵۹/۴	۶۰/۳	۶۱/۴	۶۲/۶	۶۴/۰	۶۶/۲	۶۸/۲
۵۴/۵	۵۶/۲	۵۷/۲	۵۹/۵	۶۰/۵	۶۱/۷	۶۲/۷	۶۴/۶	۶۶/۸	۶۸/۹
۵۵/۷	۵۶/۴	۵۷/۶	۵۹/۸	۶۰/۶	۶۱/۸	۶۳/۱	۶۴/۸	۶۷/۰	۶۹/۴
۵۵/۸	۵۶/۴	۵۸/۹	۶۰/۰	۶۰/۸	۶۲/۰	۶۳/۶	۶۴/۹	۶۷/۱	۷۷/۱

برای محاسبه صدک ۱۰ ام باید حداقل ۰/۱۰×۵۰=۵ مشاهده نابیشتر از آن ، و ۰/۹۰×۵۰=۴۵ مشاهده ناکمتر از آن باشند. پنجمین و ششمین مشاهده از لحاظ کوچکی هر دو این شرایط را دارا هستند بنابراین متوسط آنها یعنی (۵۵/۸+۵۵/۹)/۲=۵۵/۸۵ صدک دهم است. پس فقط ۱۰٪ از ۵۰ اندازه میزان سر و صدا خفیفتر از ۵۵/۸۵ دسی‌بل هستند.

طرز به دست آوردن میانگین نمونه‌ی پیراسته

الف) تمام مشاهدات کوچکتر از چارک اول را برمی‌داریم.

تمام مشاهدات بزرگتر از چارک سوم را بر مـــی‌داریـــم.

ب) میانگین مشاهدات باقیمانده را حساب می کنیم.

طرز به دست آوردن میانگین نمونه‌ی وینزوری

الف) به جای هر یک از مشاهدات کوچکتر از چارک اول، مقدار چارک اول را می‌گذاریم. به جای هر یـــک از مشاهدات بزرگتر از چارک سوم، مقدار چارک ســـوم را می‌گذاریم. بقیه مشاهدات را تغییر نمی‌دهیم.

ب) پس از این اصلاحات، میـــانگین تمـــام مـــشاهدات را، محاسبه می‌کنیم.

مثال: میانگین های نمونه پیراسته را برای داده های زیر بدست آورید.

۱۵۲/۴ ۱۳۶/۷ ۱۳۲/۱ ۱۲۸/۳ ۱۲۵/۰ ۱۱۶/۶ ۱۰۶/۵ ۱۰۵/۸ ۹۳/۹

چارک سوم ↑ (۱۳۲/۱)

چارک اول ↑ (۱۰۶/۵)

$$\text{میانگین پیراسته} = \frac{۱۰۶/۵+۱۱۶/۶+۱۲۵/۰+۱۲۸/۳+۱۳۲/۱}{۵}$$

$$= \frac{۶۰۸/۵}{۵} = ۱۲۱/۷$$

مثال: میانگین های نمونه وینزوری را برای داده های زیر بدست آورید.

۱۵۲/۴ ۱۳۶/۷ ۱۳۲/۱ ۱۲۸/۳ ۱۲۵/۰ ۱۱۶/۶ ۱۰۶/۵ ۱۰۵/۸ ۹۳/۹

چارک سوم

چارک اول

$$\text{میانگین وینزوری} = \frac{۱۰۶/۵+۱۰۶/۵+۱۰۶/۵+۱۱۶/۶+۱۲۵/۰+۱۲۸/۳+۱۳۲/۱+۱۳۲/۱+۱۳۲/۱}{۹}$$

$$= \frac{۱۰۸۵/۷}{۹} = ۱۲۰/۶$$

معیارهای پراکندگی عبارتند از:

- واریانس
- انحراف معیار
- دامنه نمونه
- دامنه میان چارکی نمونه

واریــانس: واریــانس($S^۲$) نمونــه‌ای مرکــب از n انــدازه

$x_۱$، $x_۲$،...،x_n بصورت زیر تعریف می‌شود:

$$s^۲ = \frac{\sum_{i=۱}^{n}(x_i - \bar{x})^۲}{n-۱}$$

انحراف معیار: انحراف معیار نمونه عبارت است از:

$$s = \sqrt{\text{واریانس}} = \sqrt{\frac{\sum_{i=\text{۱}}^{n}(x_i - \bar{x})^{\text{۲}}}{n - \text{۱}}}$$

دامنه :

کــوچکترین مــشاهده – بزرگتــرین مــشاهده = دامنــه نمونــه

دامنه‌میان‌چارکی:

چـــارک اول – چـــارک ســـوم = دامنـــه میـــان چـــارکی نمونـــه

مثال: در یک آزمایش روانشناسی، یک علامت محرکه با شدت ثابت بر روی ۶ فردی که مورد آزمایش قرار می گیرند، به کار می‌رود. زمانهای عکس العمل آنها، که برحسب ثانیـه ثبـت شده، عبارت اند از ۴، ۲، ۳، ۳، ۶ ، ۳. میانگین، واریـانس و انحراف معیار داده‌ها را محاسبه کنید.

x	۴	۲	۳	۳	۶	۳	۲۱ $=\sum x_i$ $\bar{x} = \frac{۲۱}{۶} =$ ۳/۵
$(x-\bar{x})$	۰/۵	-۱/۵	-۰/۵	-۰/۵	۲/۵	-۰/۵	۰
$(x-\bar{x})^۲$	۰/۲۵	۲/۲۵	۰/۲۵	۰/۲۵	۶/۲۵	۰/۲۵	۹/۵۰ $=(x-\bar{x})^۲$

میانگین نمونه: $\bar{x} =$ ۳/۵ثانیه

واریانس نمونه: $s^۲ = \frac{\sum_{i=۱}^{n}(x_i-\bar{x})^۲}{n-۱} = \frac{۹/۵۰}{۵} =$ ۱/۹(ثانیه)^۲

انحراف معیار نمونه: $s=\sqrt{۱/۹}=$ ۱/۴

قاعده چبیشف

برای هرمجموعه از داده ها:

الف)فاصله $\bar{x}-2s$ تا $\bar{x}+2s$ شامل حداقل $\frac{3}{4}=\left(1-\frac{1}{2^2}\right)$ از داده‌هاست.

ب) فاصله $\bar{x}-3s$ تا $\bar{x}+3s$ شامل حداقل $\frac{8}{9}=\left(1-\frac{1}{3^2}\right)$ از داده‌هاست.

ج) به طور کلی به ازای هر ضریب $k>1$ ، فاصله $\bar{x}-ks$ تا $\bar{x}+ks$ شامل حداقل $\left(1-\frac{1}{k^2}\right)$ از داده هاست.

مثال: قاعده چبیشف را در مورد داده های مربوط به تمرکز ازن امتحان کنید:

۳/۵	۱/۴	۶/۶	۶/۰	۴/۲	۴/۴	۵/۳	۵/۶
۶/۸	۲/۵	۵/۴	۴/۴	۵/۴	۴/۷	۳/۵	۴/۰
۲/۴	۳/۰	۵/۶	۴/۷	۶/۵	۳/۰	۴/۱	۳/۴
۶/۸	۱/۷	۵/۳	۴/۷	۷/۴	۶/۰	۶/۷	۱۱/۷
۵/۵	۱/۱	۵/۱	۵/۶	۵/۵	۱/۴	۳/۹	۶/۶
۶/۲	۷/۵	۶/۲	۶/۰	۵/۸	۲/۸	۶/۱	۴/۱
۵/۷	۵/۸	۳/۱	۵/۸	۱/۶	۲/۵	۸/۱	۶/۶
۹/۴	۳/۴	۵/۸	۷/۶	۱/۴	۳/۷	۲/۰	۳/۷
۶/۸	۳/۱	۴/۷	۳/۸	۵/۹	۳/۳	۶/۲	۷/۶
۶/۶	۴/۴	۵/۷	۴/۵	۳/۷	۹/۴		

$\bar{x} =$ ۴/۹۷

$S =$ ۲/۰۰

$\bar{x} - 1/5S$: $1 - \frac{1}{(1/5)^2} = 0/56$

$\bar{x} - 2S$: $1 - \frac{1}{2^2} = 0/75$

$\bar{x} - 3S$: $1 - \frac{1}{3^2} = 0/89$

نسبتی از ۷۸ مشاهده که در هر یک از سه فاصله قرار می گیرند		فاصله		
عملاً مشاهده شده	قاعده چبیشف	$\bar{x}+ks$	$\bar{x}-ks$	k
$\frac{۶۸}{۷۸}=۰/۸۷$	حداقل ۰/۵۶	۷/۹۷	۱/۹۷	۱/۵
$\frac{۷۵}{۷۸}=۰/۹۶$	حداقل ۰/۷۵	۸/۹۷	۰/۹۷	۲
$\frac{۷۷}{۷۸}=۰/۹۹$	حداقل ۰/۸۹	۱۰/۹۷	-۱/۰۳	۳

دامنه نمونه: معیار دیگری برای پراکندگی

کوچکترین مشاهده – بزرگترین مشاهده

مثال: دامنه نمونه داده‌های ازن را بدست آورید.

۳/۵	۱/۴	۶/۶	۶/۰	۴/۲	۴/۴	۵/۳	۵/۶
۶/۸	۲/۵	۵/۴	۴/۴	۵/۴	۴/۷	۳/۵	۴/۰
۲/۴	۳/۰	۵/۶	۴/۷	۶/۵	۳/۰	۴/۱	۳/۴
۶/۸	۱/۷	۵/۳	۴/۷	۷/۴	۶/۰	۶/۷	۱۱/۷
۵/۵	۱/۱	۵/۱	۵/۶	۵/۵	۱/۴	۳/۹	۶/۶
۶/۲	۷/۵	۶/۲	۶/۰	۵/۸	۲/۸	۶/۱	۴/۱
۵/۷	۵/۸	۳/۱	۵/۸	۱/۶	۲/۵	۸/۱	۶/۶
۹/۴	۳/۴	۵/۸	۷/۶	۱/۴	۳/۷	۲/۰	۳/۷
۶/۸	۳/۱	۴/۷	۳/۸	۵/۹	۳/۳	۶/۲	۷/۶
۶/۶	۴/۴	۵/۷	۴/۵	۳/۷	۹/۴		

کوچکترین مشاهده= ۱/۱

بزرگترین مشاهده= ۱۱/۷

دامنه نمونه=۱۱/۷-۱/۱=۱۰/۶

چارک اول – چارک سوم= دامنه میان چارکی نمونه

مثال: دامنه و دامنه میان چارکی داده های میزان سروصدا را محاسبه کنید:

۵۲/۰	۵۵/۹	۵۶/۷	۵۹/۴	۶۰/۲	۶۱/۰	۶۳/۱	۶۲/۸	۶۵/۷	۶۷/۹
۵۴/۴	۵۵/۹	۵۶/۸	۵۹/۴	۶۰/۳	۶۱/۴	۶۲/۶	۶۴/۰	۶۶/۲	۶۸/۲
۵۴/۵	۵۶/۲	۵۷/۲	۵۹/۵	۶۰/۵	۶۱/۷	۶۲/۷	۶۴/۶	۶۶/۸	۶۸/۹
۵۵/۷	۵۶/۴	۵۷/۶	۵۹/۸	۶۰/۶	۶۱/۸	۶۳/۱	۶۴/۸	۶۷/۰	۶۹/۴
۵۵/۸	۵۶/۴	۵۸/۹	۶۰/۰	۶۰/۸	۶۲/۰	۶۳/۶	۶۴/۹	۶۷/۱	۷۷/۱

کوچکترین مشاهده-بزرگترین مشاهده = دامنه نمونه

=۷۷/۱-۵۲/۰

=۲۵/۱ دسی‌بل

دامنه میان چارکی = Q_3-Q_1

=۶۴/۶-۵۷/۲

=۷/۴ دسی‌بل

تذکراتی درباره‌ی معیار های عددی

در انتخاب معیار مناسب توجه به نکات زیر مفید است:

الف) منظوری که از توصیف اختصاری داده‌ها داریم.

ب) سادگی تعبیر.

ج) درجه‌ی تاثیر از مشاهدات غیر منتظره.

د) میزان بالقوه کاربرد در فرآیند استنباط آماری

بطورخلاصه:

الف)ارائه نموداری برای نمایش شکل کلی داده‌هـــا، یکـــی از مراحـــل اساسی توصیف اختصاری داده‌هاست.

ب)اگر با توجه به هدفی‌که از توصیف‌اختصاری‌داده‌ها داریم، معیارهـــای گوناگونی به صورت رقیب یکدیگر موجود باشند، بهتر است به‌جـــای استفاده دلبخواه از یک معیار، همه‌ی آنها را به کار بریم.

ج) در تعبیر معیارهای توصیفی، شکل کلی توزیع داده‌ها را همیشه بایـــد در نظر داشت.

د) هر یک از مقادیر مشاهدات غیرعادی را باید به صورت پانوشت در صفحه‌ی گزارش ذکر کرد.

معیارهای توصیفی برای داده‌های گروه‌بندی شده

میانگین نمونه گروه بندی شده

اگر توزیع فراوانی دارای k رده باشد و نقاط وسط رده‌ها m_1، m_2، ....، m_k و فراوانی‌هـــای متنـــاظر f_1، f_2، ....، f_k باشند آنگاه:

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{n} = \frac{\sum_{\text{رده‌ها}} (\text{نقطه وسط} \times \text{فراوانی})}{\text{کل فراوانی}}$$

واریاس نمونه گروه بندی شده عبارت است از:

$$s^{۲} = \frac{\sum_{i=۱}^{k} (m_i - \bar{x})^{۲} f_i}{n}$$

فرمول دیگر برای واریانس نمونه گروه‌بندی شده:

$$s^{۲} = \frac{\sum_{i=۱}^{k} m_i^{۲} f_i}{n} - \bar{x}^{۲}$$ (برای محاسبه دستی)

مراحل محاسبه صدک(۱۰۰p)ام داده‌های گروه‌بندی شده

الف) ستون فراوانی تجمعی را محاسبه می کنیم.

ب) رده‌ای را که فراوانی تجمعی آن مساوی np یا اولین عدد صحیح بزرگتر از آن است، پیدا می‌کنیم (n تعداد کل مشاهدات است). این رده را رده‌ی صدک (۱۰۰p) ام می‌نامیم و ابتدای آن را با L نشان می‌دهیم.

مراحل محاسبه صدک($100p$)ام داده‌های گروه‌بندی شده

ج) کسر $\frac{np-a}{f}$ را محاسبه می‌کنیم، که در آن a فراوانی تجمعی رده ای است که بلافاصله قبل از رده‌ی صدک ($100p$)ام قرار دارد، و f فراوانی رده‌ی صدک ($100p$)ام است.

د) سپس محاسبه‌ی زیر را انجام می‌دهیم:

$$\text{صدک } (100p)\text{ام} = L + \frac{(np-a)}{f} h$$

که در آن، h مساوی است با طول رده‌ی صدک ($100p$)ام.

مثال: میانه و دامنه میان چاکی را برای داده‌های زیر بدست آورید:

رده (سن برحسب سال)	فراوانی	فراوانی تجمعی
۱۴/۵–۱۹/۵	۱۸	۱۸
۱۹/۵–۲۴/۵	۷۴	۹۲
۲۴/۵–۲۹/۵	۶۲	۱۵۴
۲۹/۵–۳۴/۵	۲۶	۱۸۰
۳۴/۵–۳۹/۵	۲۰	۲۰۰
مجموع	۲۰۰	

برای محاسبه میانه،

$p=۰/۵$ $\qquad$ $np=۲۰۰(۰/۵)=۱۰۰$

$\boxed{\text{رده‌ی میانه} \quad ۲۴/۵ - ۲۹/۵}$ $\qquad$ فراوانی آن $f=۶۲$

$$\text{میانه} = ۲۴/۵ + \frac{(۱۰۰-۹۲)}{۶۲} \times ۵ = ۲۵/۱۵ \text{ سال}$$

$$Q_۱ = ۱۹/۵ + \frac{۳۲}{۷۴} \times ۵ = ۲۱/۶۶ \text{ سال}$$

$$Q_۳ = ۲۴/۵ + \frac{(۱۵۰-۹۲)}{۶۲} \times ۵ = ۲۹/۱۸ \text{ سال}$$

$$Q_۳ - Q_۱ = ۲۹/۱۸ - ۲۱/۶۶ = ۷/۵۲ \text{ سال}$$

کدگذاری داده‌ها

کدگذاری مجموعه‌ای از داده‌ها، عبارت است از عملیـــاتی که در طی آن از هر مشاهده، عدد ثابتی را کـــم(یـــا بـــه آن اضافه) کرده و نتیجه را بر عدد ثابتی تقسیم (یا در آن ضرب) می‌نمایند. این عمل به منظور ساده‌کردن محاسبه معیارهـــای توصیفی انجام می‌گیرد، ولی‌در حال حاضر این کـــار توســـط کامپیوتر انجام می‌گیرد. کدگذاری باعث صرفه‌جویی در وقت نمی‌شود. معهذا نتایج حاصل از آن برای نشان دادن اثراتـــی که **تغییرات واحد اندازه‌گیری** می‌تواند بـــر روی معیارهـــای توصیفی داشته باشد مفید است.

اگر $x_i = au_i + b$ یا $u_i = \frac{1}{a}(x_i - b)$ آنگاه

$$\bar{x} = a\bar{u} + b \qquad \bar{u} = \frac{1}{a}(\bar{x} - b)$$

$$s_x^2 = a^2 s_u^2 \qquad s_u^2 = \frac{1}{a^2} s_x^2$$

$$s_x = |a| s_u \qquad s_u = \frac{1}{|a|} s_x$$

مثال: اندازه گیری طول یک قسمت از بزرگراه جدیـدی را ۵ مرتبه تکرار کرده و مقادیر زیر بدست آمده:

۱۰۰/۱۸ ، ۱۰۰/۱۷، ۱۰۰/۱۸، ۱۰۰/۱۹، ۱۰۰/۱۷

می‌خواهیم داده‌ها را به طـور مناسـب کدگـذاری کـرده و مقدار میانگین و انحراف معیار را پیدا کنیم.

با نگاه به داده‌ها، عددی نزدیک به مرکز مشاهدات مثلا ۱۰۰/۱۸=b را انتخاب می‌کنیم. چون تفاوت مشاهدات با این مقدار مضارب ۰/۰۱ هستند، a را برابر با ۰/۰۱ می‌گیریم. یعنی،

$$u_i = \frac{۱}{۰/۰۱}\left(x_i - ۱۰۰/۱۸\right)$$

بنابراین مقادیر کدها عبارتند از: ۰، ۱-، ۰، ۱، ۱-

$$\bar{u} = \frac{-۱+۱+۰-۱+۰}{۵} = -۰/۲$$

و واریانس برابر است با:

$$S_u^{۲}=\frac{(-۱+۰/۲)^{۲}+(۱+۰/۲)^{۲}+(۰/۲)^{۲}+(-۱+۰/۲)^{۲}+(۰/۲\ )^{۲}}{۵-۱}=۰/۷۰$$

$$S_u=\sqrt{S_u^{۲}}=۰/۸۳۷$$

با مراجعه مجدد به مقادیر اولیه x، اکنون داریم $X_i=(۰/۰۱)\ u_i+۱۰۰/۱۸$

$$\bar{x}=(۰/۰۱)\bar{u}+۱۰۰/۱۸=۱۰۰/۱۷۸$$

$$S_x^{۲}=(۰/۰۱)^{۲}\ S_u^{۲}=(۰/۰۱)^{۲}(۰/۷)=۰/۰۰۰۰۷$$

$$S_x=(۰/۰۱)\ \ S_u=(۰/۰۱)\ (۰/۸۳۷)=۰/۰۰۸۳۷$$

فصل ۳

مبادی احتمال

مبادی احتمال

فضای نمونه و پیشامدها

آزمایش، عبارت از فرایند گردآوری داده‌هـــای مربـــوط بـــه پدیده‌ای است که برآمدهای آن متفاوتند.

مجموعه‌ی تمام برآمدهای متمایز ممکن آزمایش، **فضای نمونه** برآمدها نامیده می‌شود و هر برآمد متمایز یک **پیشامد ساده** یا یک **برآمد مقدماتی** یا یک **عنصر فضای نمونـــه** خوانـــده مـــی‌شـــود.
فضای نمونه را S با نشان می دهند.

مجموعه‌ای از برآمدهای مقدماتی را که با توصیف بعـــضی از خصوصیاتشان مشخص شده‌اند **پیـــشامد** گـــوییم. پیـــشامد، زیـــر مجموعه‌ای از فضای نمونه S است. چند حرف بزرگ اول الفبای لاتین(A، B ، C، ...)، معمولاً برای نشان دادن پیشامدها به کار می روند.

وقتی یکی از برآمدهای مقدماتی موجود در A رخ می‌دهـــد، گوییم پیشامد A رخ می‌دهد.

فضای نمونه‌ای را که تعداد عناصر آن متناهی یا به‌طور شمارش پذیر نامتناهی باشد، فضای نمونه‌ی گسسته گوییم.

وقتی که فضای نمونه شامل تمام اعداد متعلق به یک فاصله باشد ، آن را فضای نمونه‌ی پیوسته گوییم.

احتمال یک پیشامد

تصور شهودی ما از اندازه‌ی عددی احتمال یـــک پیـــشامد، نسبت دفعاتی است که انتظار می رود آن پیشامد رخ بدهـــد، وقتی که آزمایش تحت شرایط یکسان تکرار گردد.

علامت $P(A)$ برای نشان دادن احتمال پیشامد A بـــه کـــار می‌رود.

برآمدهای مقدماتی همشانس

اگر یک فضای نمونه شامل k برآمد مقـــدماتی $\{e_1,\ldots,e_k\}$ باشد که به طور همشانس رخ بدهد، احتمال برآمد مقـــدماتی برابر با $1/k$ است اگر پیشامد A شامل m برآمد از k برآمـــد مقدماتی باشد، داریم

$$P(A)=\frac{m}{k}$$

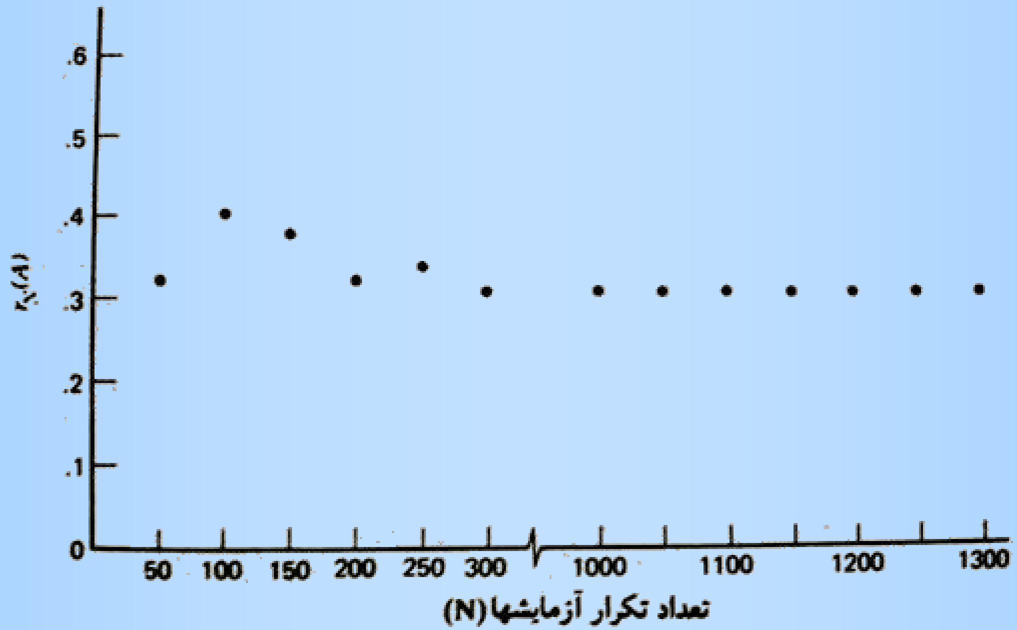
گوییم فضای نمونه‌ای که عناصـــرش همـــشانس‌انـــد دارای **مدل احتمال یکنواخت** است.

پایداری فراوانی نسبی

فراوانی نسبی پیشامد A در N بار تکرار آزمایش

$$r_N(A) = \frac{\text{تعداد دفعاتی که } A \text{ در } N \text{ بار تکرار آزمایش رخ می‌دهد}}{N}$$

گواه تجربی برای انتساب مقدار عددی $P(A)$ بـــه احتمـــال پیشاند A، از مشاهده پایداری فراوانی نسبی A بعد از تکرار بسیار آزمایش به دست می‌آید.



.6
.5
.4
.3
.2
.1
0
$r_N(A)$
50
100
150
200
250
300
1000
1100
1200
1300
تعداد تکرار آزمایشها (N)

شرایط مدل احتمال برای فضاهای نمونه گسسته

احتمال، تابعی است که روی پیشامدها تعریف می‌شود، و در شرایط زیـــر صدق می کند:

یک) برای هر پیشامد A **،** $0 \le P(A) \le 1$

دو) P(A)**، برابر مجموع احتمالهای برآمدهای مقدماتی متعلق به** A **است؛ به عبارت دیگر:**

$$P(A) = \sum_{\text{تمام } e_i \text{ های متعلق به } A} P(e_i)$$

سه) $P(\mathcal{S}) = \sum_{\text{تمام } e_i \text{ های متعلق به } \mathcal{S}} P(e_i) = 1$

مثال: آزمایش جنس فرزند اول و دوم در خانواده‌های دارای دو فرزند، چهار برآمد دارد(پسر، پسر)$=e_1$، (پسر، دختر)$=e_2$، (دختر، پسر)$=e_3$، (دختر، دختر)$=e_4$. آیا هر دسته از احتمال‌های نسبت داده شده زیر در شرایط مدل احتمال صدق می کند؟

الف) $P(e_1)=P(e_2)=P(e_3)=P(e_4)=\frac{1}{4}$

ب) $P(e_1)=\frac{3}{16}, P(e_2)=\frac{3}{8}, P(e_3)=\frac{1}{4}, P(e_4)=\frac{3}{16}$

ج) $P(e_1)=\frac{1}{2}, P(e_2)=\frac{1}{4}, P(e_3)=\frac{1}{2}, P(e_4)=\frac{1}{2}$

عملیات روی پیشامدها و قوانین پایه ای احتمال

نمودار ون: فضای نمونه S را به عنوان مجموعه ای از نقاط در یک نمودار در نظر بگیرید که در آن نمودار، هـــر نقطـــه متناظر است با یک برآمد مقدماتی مشخص آزمایش. در این صورت هر پیشامد در نمودار به صورت مجموعه نقاطی که در خصوصیت توصیف کننده‌ی پیشامد صدق مـــی‌کننـــد معرفـــی می‌شود.

مثال: وقتی یک محرک آزمایشی روی حیوانی به کار می‌رود، یا حیوان پاسخ می‌دهد یا در پاسخ دادن عاجز می‌ماند. به عبارت دیگر، تنها دو برآمد ممکن وجود دارد: یا حیوان پاسخ می‌دهد (R) یا پاسخ نمی‌دهد (N). آزمایش، عبارت است از اعمال تحریک‌های متوالی بر سه حیوان و ثبت R یا N برای هر یک از آنهاست. یک نمودار ون برای فضای نمونه بسازید و پیشامدهای زیر را نشان دهید:

A: تنها یک حیوان پاسخ می‌دهد.

B: در اولین تحریک، پاسخ وجود دارد.

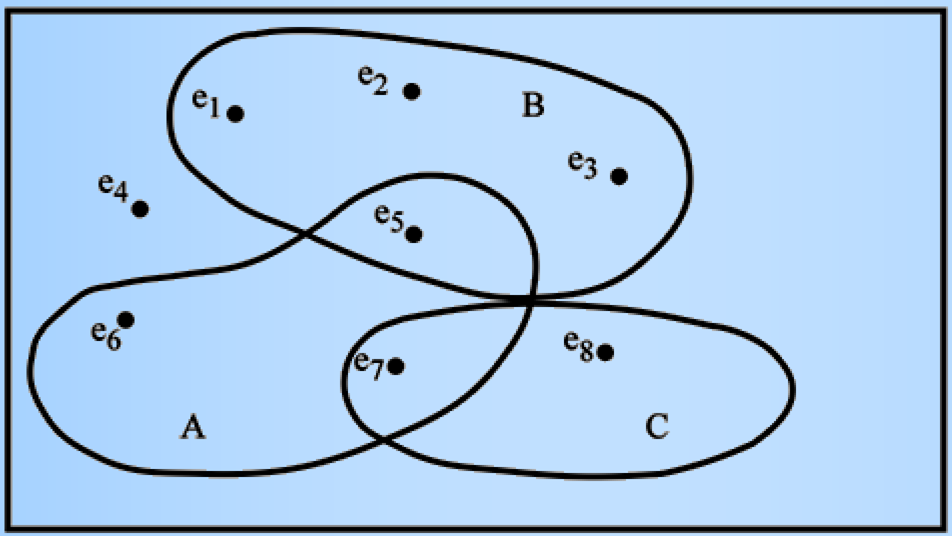
C: هر دو حیوان اول و دوم در پاسخ دادن عاجزند.

تعداد برآمد مقدماتی ۸=۲×۲×۲

RRR($e_{۱}$) RRN($e_{۲}$) RNR($e_{۳}$) NRR($e_{۴}$)

RNN($e_{۵}$) NRN($e_{۶}$) NNR($e_{۷}$) NNN($e_{۸}$)

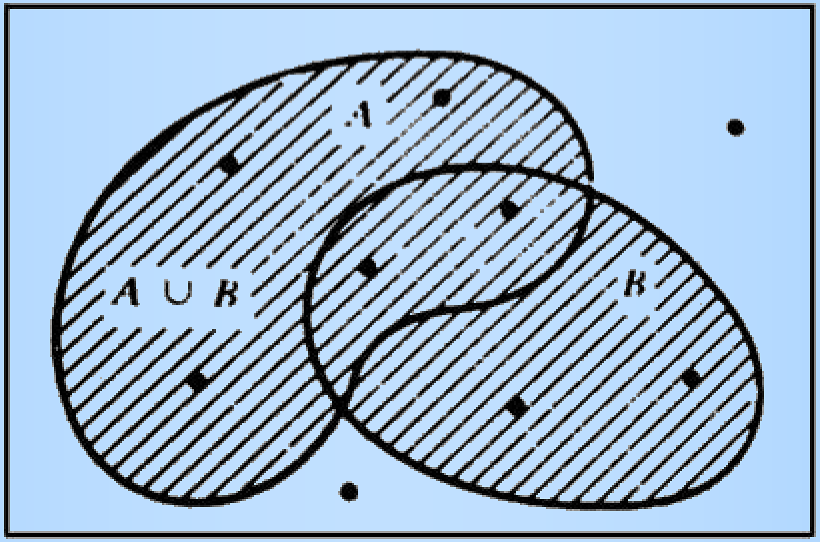
$A=\{e_{۵},e_{۶},e_{۷}\}$ $B=\{e_{۱}, e_{۲},e_{۳},e_{۵}\}$ $C=\{e_{۷},e_{۸}\}$



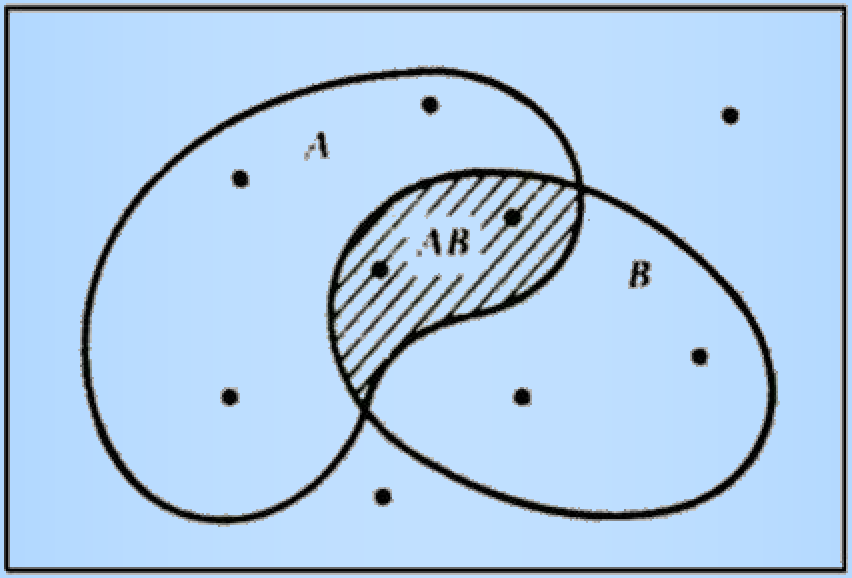
سه عمل پایه ای روی پیشامدها عبارت است از :

- اجتماع
- اشتراک
- متمم

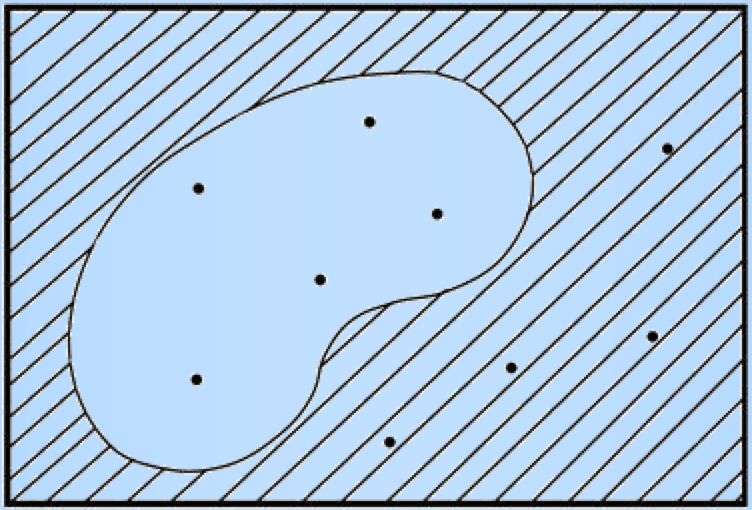
اجتماع دو پیشامد A و B ، مجموعه تمام برآمدهای مقدماتیی است که در A، یا در B و یا هم در A و هم در B ، قــرار دارنـــد. اجتماع A و B به صورت $A \bigcup B$نشان داده می‌شود، که با $B \bigcup A$ یکی است . وقوع $A \bigcup B$ به معنی این است که حداقل یکی از دو پیشامد A و B رخ می دهد.



اشتراک دو پیشاند A و B ، مجموعه تمام برآمدهای مقدماتیی است که هم به A و هم به B تعلق دارند. اشتراک A و B به صورت AB نشان داده می شود ، که با BA یکی است. وقوع AB به معنی این است که هر دو پیشامد Aو B رخ می دهند.



متمم پیشامد A، مجموعه‌ی تمام برآمدهای مقدماتیی است که در A نیستند. متمم A به صورت Ac نشان داده می‌شود. این حکم که A رخ نمی‌دهد مترادف با این حکم است که A^c رخ می‌دهد.



مثال: فرض کنید که پیشامدهای A، B، C به صورت زیر است

$A=\{e_5,e_6,e_7\}$ $B=\{e_1,e_2,e_3,e_5\}$ $C=\{e_7,e_8\}$

ساختار هر یک از پیشامدهای زیر را مشخص کنید:

$A\cup B, AC, BC, B^c, (A\cup B)^c, A^c\cup B^c$

$A \bigcup B = \{e_{\text{۱}}, e_{\text{۲}}, e_{\text{۳}}, e_{\text{۵}}, e_{\text{۶}}, e_{\text{۷}}\}$

$AC = \{e_{\text{۷}}\}$

$BC = \text{تهی}$

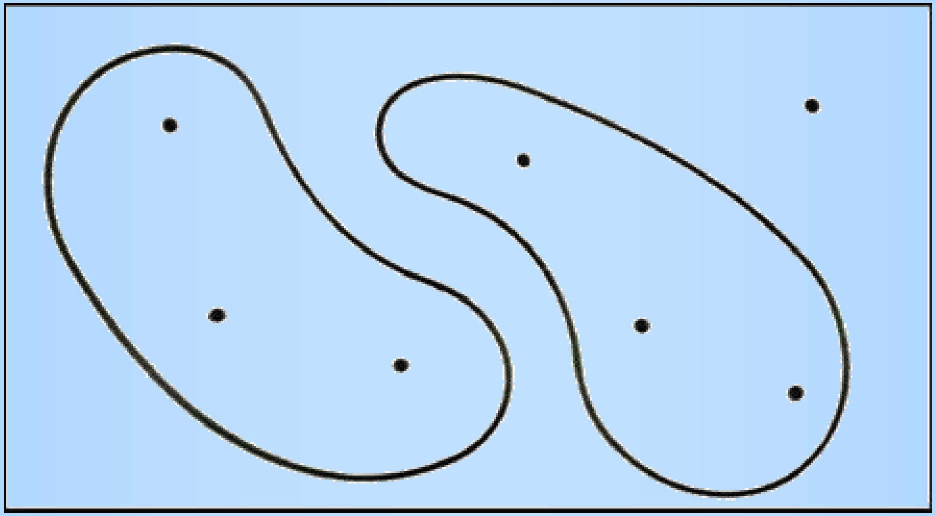
$B^c = \{e_{\text{۴}}, e_{\text{۶}}, e_{\text{۷}}, e_{\text{۸}}\}$

$(A \bigcup B)^c = \{e_{\text{۴}}, e_{\text{۸}}\}$

$A^c = \{e_{\text{۱}}, e_{\text{۲}}, e_{\text{۳}}, e_{\text{۴}}, e_{\text{۸}}\}$

$A^c \bigcup B^c = \{e_{\text{۱}}, e_{\text{۲}}, e_{\text{۳}}, e_{\text{۴}}, e_{\text{۶}}, e_{\text{۷}}, e_{\text{۸}}\}$

یکی از حالات ممکن دو پیشامد A و B این است کــه هــیچ برآمــد مقدماتی بطور همزمان متعلق به هر دو پیشامد A و B نباشد.



دوپیشامد A و B را پیشامدهای **جدا از هم** یا **دوبدو ناسازگار** گوییم اگر هیچ برآمد مقدماتی مشترکی نداشته باشند؛ یعنـــی، اگر اشتراک آنها، AB، تهی باشد. پیـــشامدهای جـــدا از هـــم نمی‌توانند به طور همزمان رخ دهند.

عبارت لفظی	عبارت نمادی
حداقل یکی از A یا B	$A \bigcup B$
هم A و هم B	AB
چیزی غیر از A	A^C

قانون متمم گیری

$$P(A^c)= ۱-P(A)$$

قانون جمع برای اجتماع پیشامدهای جدا از هم

اگر دو پیشامد A و B جدا از هم باشند. آنگاه،

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

قانون کلی جمع

برای هر دو پیشامد A و B ،

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

مثال: در امتحانی، ۶ عبارت به شاگردی داده می‌شود و از او خواسته می‌شود که ببیند هر عبارت درست است یا نادرست و جلوی عبارت درست علامت R و جلوی عبارت نادرست علامت W بگذارد. فرض کنید که شاگرد به جای اینکه قبل از جواب دادن عبارات را بخواند و درباره آنها فکر کند، تصمیم بگیرد که علامت R و W را به طور تصادفی جلو هر عبارت بگذارد.

الف) احتمال اینکه در مجموعه‌ی جواب‌های شاگرد حداقل یک علامت R وجود داشته باشد چقدر است؟

تعداد برآمدها برای مجموعه‌ی ۶ عبارت، برابر ۶۴=۲×۲×۲×۲×۲×۲ است. مدل احتمال یکنواختی با ۶۴ برآمد مقدماتی داریم.

A= حداقل یک علامت R در مجموعه جواب‌ها باشد.

A^c= همه علامتها W هستند.

$$P(A^c) = \frac{۱}{۶۴} \qquad P(A^c) = ۱ - \frac{۱}{۶۴}$$

ب) احتمال اینکه در مجموعه جواب‌های شاگرد پنج یا شش علامت R وجود داشته باشد چقدر است؟

فرض کنید: $A_۵$ = دقیقاً ۵ تا R.

(WRRRRR),(RWRRRR),(RRWRRR)
(RRRWRR),(RRRRWR),(RRRRRW)

$A_۶$= دقیقاً ۶ تا R. (RRRRRR)

$$P(A_۵)= \frac{۶}{۶۴} \qquad P(A_۶)= \frac{۱}{۶۴}$$

$$P(R\text{ علامت ۵ یا ۶}) = P(A_۵ \cup A_۶)$$

$$= P(A_۵) + P(A_۶)$$

$$= \frac{۶}{۶۴} + \frac{۱}{۶۴} = \frac{۷}{۶۴}$$

مثال: گروهی از شاگردان دبیرستانی را برگزیده و آنها را برحسب قابلیتشان در درس‌های ریاضیات و ورزش رده‌بندی کرده‌ایم نسبت‌های تعداد شاگردان بر حسب قابلیتشان در درس‌های ریاضیات و ورزش.

اگر شاگردی را بطور تصادفی از این گروه برگزینیم، احتمال آنکه

		ورزش		
		ضعیف	متوسط	خوب
ریاضیات	خوب	۰/۱۸	۰/۱۱	۰/۰۶
	متوسط	۰/۱۰	۰/۱۸	۰/۱۲
	ضعف	۰/۰۴	۰/۰۵	۰/۱۶

الف) حداقل یک درس شاگرد خوب باشد چیست؟

فرض کنید A این پیشامد باشد که درس ریاضی شاگرد خوب است و B

که حداقل یک درس شاگرد خوب است، و داریم

	ورزش			
ضعیف	متوسط	خوب		
۰/۱۸	۰/۱۱	۰/۰۶	خوب	ریاضیات
۰/۱۰	۰/۱۸	۰/۱۲	متوسط	ریاضیات
۰/۰۴	۰/۰۵	۰/۱۶	ضعف	ریاضیات

P(A)=۰/۰۶+۰/۱۱+۰/۱۸=۰/۳۵

P(B)=۰/۰۶+۰/۱۲+۰/۱۶=۰/۳۴

P(AB)=۰/۰۶

P(A∪B)=P(A)+(B)-P(AB)

=۰/۳۵+۰/۳۴-۰/۰۶

=۰/۶۳

		ورزش		
		ضعیف	متوسط	خوب
ریاضیات	خوب	۰/۱۸	۰/۱۱	۰/۰۶
	متوسط	۰/۱۰	۰/۱۸	۰/۱۲
	ضعف	۰/۰۴	۰/۰۵	۰/۱۶

ب) شاگرد در ریاضیات ضعیف نباشد چیست؟

فرض کنید D، پیشامد ضعیف بودن شاگرد در ریاضیات باشد، می‌خواهیم $P(D^c)$ را محاسبه کنیم.

$P(D)$=۰/۱۶+۰/۰۵+۰/۰۴=۰/۲۵

$P(D^c)$=۱−۰/۲۵=۰/۷۵

قواعد شمارش و کاربرد آنها در مدلهای احتمال یکنواخت

احتمال یک پیشامد A

$$P(A)=\frac{\text{تعداد عناصر در } A}{\text{تعداد عناصر در } S}$$

قاعده ضرب

هرگاه آزمایشی مرکب از دو مرحله باشد به‌طوری که در مرحله اول بتواند دارای k نتیجه متمایز باشد و به ازای هر نتیجه مرحله اول، امکان l نتیجه متمایز در مرحله دوم وجود داشته باشد. تعداد کل نتاج ممکن آزمایش ، $k\times l$ خواهد بود.

مثال: شرکتی برای مطرح کردن دعوایی، می تواند در مرحله‌ی اول دادرسی، یک وکیل را از میان ۳ وکیل، و یک دادگاه را از میان دو دادگاه انتخاب کند. این شرکت در مورد گزینش وکیل و دادگاه چند نوع تصمیم می‌تواند بگیرد؟

k=۳ انتخاب برای وکیل وجود دارد، و به ازای هر یک از اینها، l=۲ انتخاب برای دادگاه. بنابراین تعداد تصمیم‌های مختلف ممکن، برابر با ۶=۲×۳=$k \times l$ است.

قاعده جایگشت‌ها یا ترتیبها

تعداد ترتیب‌های مختلفی که می‌توان با r شی، برگزیده از گروهی مرکب از n شی متمایز ساخت، با P^{n}_{r} نشان داده می‌شود و به صورت "تعداد جاگشت‌های r از n" خوانده می‌شود.

تعداد ترتیبهای ممکن r شی، برگزیده از n شی متمایز برابر است با

$$P_r^n = \underbrace{n(n-۱)(n\text{-}۲)\ldots}_{r\ \text{عامل}}$$

$$=n(n\text{-}۱)(n\text{-}۲)\ldots(n\text{-}r+۱)$$

مثال: پانزده نفر در یک مسابقه دوچرخه‌سواری شرکت می‌کنند. به چند طریق می‌توان جوایز اول و دوم و سوم را به افراد شرکت کننده در مسابقه اعطاء نمود؟

با در نظر گرفتن جوایز اول، دوم، سوم به عنوان سه جعبه، می خواهیم تعداد ترتیبهای ممکن سه نفر برگزیده شده از پانزده مسابقه دهنده را که بتوان در این سه جعبه قرار داد بیابیم. بنا به قاعده‌ی جایگشت‌ها، جواب برابر است با:

$$P_{۳}^{۱۵}=۱۵\times۱۴\times۱۳=۲۷۳۰$$

تعداد ترتیبهای ممکن n شی متمایز برابر است با:

$$P_n^n = n(n-1)(n-2)\times\ldots\times 2\times 1$$

این حاصل ضرب را با !n نیز نشان داده و آن را n فاکتوریل می‌خوانند.

مثال: تمام حالتهای ممکن (الف) ترکیبهای و (ب) ترتیبهای سه حرف برگزیده شده از میان چهار حرف A، B، C، D را به دست آورید:

ترکیب‌ها	ترتیب‌ها					
{A,B,C}	ABC	ACB	BAC	BCA	CAB	CBA
{A,B,D}	ABD	ADB	DAB	DBA	BAD	BDA
{A,C,D}	ACD	ADC	CAD	CDA	DAC	DCA
{B,C,D}	BCD	BDC	CBD	CDB	DBC	DCB

تعداد ترکیب‌های ممکن را می‌توان به صورت زیر تعیین نمود:

$$P^{۴}_{۳} = \frac{۲۴}{۶} = ۴$$

قاعده ترکیب‌ها

تعداد ترکیب‌های ممکن r شیء برگزیده شده از میان n شیء متمایز را با نمایش می‌دهند و آن را **تعداد ترکیب‌های r از n** می خوانند . فرمول به صورت

$$\binom{n}{r} = \frac{P_r^n}{r!}$$

$$= \frac{n(n-1)\dots(n-r+1)}{r!}$$

داده می شود.

مثال: مقادیر $\binom{۵}{۳}$ و $\binom{۷}{۲}$ و $\binom{۷}{۵}$ و $\binom{۱۰}{۳}$ را محاسبه کنید.

$$\binom{۵}{۳} = \frac{۵\times۴\times۳}{۱\times۲\times۳} = ۱۰$$

$$\binom{۷}{۵} = \frac{۷\times۶\times۵\times۴\times۳}{۱\times۲\times۳\times۴\times۵} = ۲۱$$

$$\binom{۷}{۲} = \frac{۷\times۶}{۱\times۲} = ۲۱$$

$$\binom{۱۰}{۳} = \frac{۱۰\times۹\times۸}{۱\times۲\times۳} = ۱۲۰$$

خواص فرمول ترکیب

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

$$\binom{n}{n} = ۱ \quad , \binom{n}{۱} = n$$

تعریف: ۱=!۰

درنتیجه، $\binom{n}{0} = 1 = \binom{n}{n}$

مثال: جعبه‌ای شامل ۱۲ کالا داریم از آنها ۸ تا سالم (با علایم G_1، G_2،....،G_8) و ۴ تا معیوب(با علایم D_1،D_2،D_3،D_4) است. حال فرض کنید که ۳ کالا به طور همزمان از جعبه استخراج شوند و علایم اشیای خارج شده را یادداشت نماییم.

الف) امکان چند نتیجه‌ی متمایز وجود دارد؟

تعداد نمونه های مرتب نشده ممکن است به حجم ۳ کالا از میان ۱۲ کالا، برابر است با:

$$\binom{12}{3} = \frac{12 \times 11 \times 10}{1 \times 2 \times 3} = 220$$

ب) با این شرط که دو تا از کالاهای استخراج شده معیوب باشند و یکی سالم باشد چند نتیجه ممکن وجود دارد؟

تعداد ترکیب‌های ۲ کالای معیوب از میان ۴ کالای معیوب برابر است با $\binom{۴}{۲}$، و تعداد ترکیبهای ۱ کالای سالم از میـان ۸ کالای سالم ۸ کالای سالم برابر است با $\binom{۸}{۱}$. تعداد نتایج(بنا به قاعده ضرب) ممکن برابر است با:

$$\binom{۴}{۲}\binom{۸}{۱}=\frac{۴\times۳}{۱\times۲}\times\frac{۸}{۱}=۴۸$$

ج) اگر سه کالا بتصادف استخراج شوند(یعنی، به روشی که همه‌ی مجموعه‌های سه عنصری از کالاها از لحاظ گزینش دارای شانس برابر باشند) احتمال به دست آوردن دوکالای معیوب و یک کالای سالم چقدر است؟

با استفاده از مدل احتمال یکنواخت داریم:

$$P(\text{۲ کالای معیوب و ۱ کالای سالم}) = \frac{\binom{۴}{۲}\binom{۸}{۱}}{\binom{۱۲}{۳}} = \frac{۴۸}{۲۲۰}$$

به عنوان یک راهنمایی کلی، بهتر است به خاطر داشته باشید که قاعده جایگشت‌ها وقتی به کار می‌رود که لازم است پیشامدی برحسب ترتیب مشخصی از نتایج بیان شود. زمانی که ترتیب رخ دادن نتایج مورد نظر نیست، قاعده ترکیب‌ها به کار می‌رود.

مثال: یک کمیته مشورتی که درباره‌ی اصلاح وضع زندانی‌ها مطالعه می‌کند، مرکب از ۱۵ عضو است. درباره برنامه خاصی ۹ نفر موافق، ۴ نفر مخالف، و ۲ نفر بیطرف هستند. خبرنگاری علاقمند است که ۳ نفر را بطور تصادفی از این کمیته برگزیند و نظرات آنها را در یک برنامه خبری تلویزیون پخش کند.

الف) احتمال اینکه حداقل دو نفر از افراد برگزیده شده موافق برنامه باشند چقدر است؟

ترتیب افراد در نمونه گیری مطرح نیست. فرض کنید:

$A_۲$: پیشامد اینکه دقیقاً ۲ نفر موافق است.

$A_۳$: پیشامد اینکه دقیقاً ۳ نفر موافق است.

$$P(A_۲ \cup A_۳) = P(A_۲) + P(A_۳)$$

$$P(A\ ) = \frac{\binom{۹}{۳}\binom{۶}{۰}}{\binom{۱۵}{۳}} = \frac{۸۴}{۴۵۵}$$

$$P(A\ ) = \frac{\binom{۹}{۲}\binom{۶}{۱}}{\binom{۱۵}{۳}} = \frac{۲۱۶}{۴۵۵}$$

$$P(\text{حداقل دو نفر موافق}) = \frac{۲۱۶}{۴۵۵} + \frac{۸۴}{۴۵۵} = \frac{۳۰۰}{۴۵۵}$$

ب) احتمال اینکه دو نفر اول موافق برنامه باشند و سومی مخالف با آن باشد، چقدر است؟

$$P(\text{دو نفر اول موافق و نفر سوم مخالف}) = \frac{۹\times۸\times۴}{۱۵\times۱۴\times۱۳} = \frac{۲۸۸}{۲۷۳۰}$$

نمونه‌ای به حجم n، برگزیده شده از جامعـــه ای شـــامل N شیء متمایز ، نمونه ای تصادفی نامیده می شود اگر گــزینش هر مجموعه ای به حجم n دارای احتمال $1 \Big/ \binom{N}{n}$ باشد.

احتمال شرطی

احتمال شرطی A به شرط B با $P(A|B)$ نشان داده می‌شود و با فرمول

$$P(A\,|\,B)=\frac{P(AB)}{P(B)}$$

تعریف می‌گردد، که در آن $P(B)>0$. این فرمول را می‌توان به صورت زیر نوشت:

$$P(AB)=P(B)P(A|B)$$

که آن را **قانون ضرب احتمال‌ها** گوییم.

مثال: گروهی از افراد بر طبق وزن بدن و ابتلا یا عدم ابتلا به بیماری فشار خون رده‌بندی شده اند.

فرض کنید:

A: نشان دهنده پیشامد ابتلای فرد به فشار خون

B: نشان دهنده پیشامد وزن فرد بیشتر از حد معمول

	وزن کمتر از حد معمول	وزن معمولی	وزن بیشتر از حد معمول
فشار خون دارند	۰/۰۲	۰/۰۸	۰/۱۰
فشار خون ندارند	۰/۲۰	۰/۴۵	۰/۱۵
مجموع	۰/۲۲	۰/۵۳	۰/۲۵

	وزن بیشتر از حد معمول	وزن معمولی	وزن کمتر از حد معمول	مجموع
فشار خون دارند	۰/۱۰	۰/۰۸	۰/۰۲	۰/۲۰
فشار خون ندارند	۰/۱۵	۰/۴۵	۰/۲۰	۰/۸۰
مجموع	۰/۲۵	۰/۵۳	۰/۲۲	۱/۰۰

الف) احتمال اینکه فردی که بتصادف از این گروه برگزیده شده، مبتلا به فشار خون باشد چقدر است؟

۲۰٪ از افراد مبتلا به فشار خون هستند و یک نفر بتصادف از گروه برگزیده می شود، $P(A)=۰/۲$.

مجموع	وزن کمتر از حد معمول	وزن معمولی	وزن بیشتر از حد معمول	
۰/۲۰	۰/۰۲	۰/۰۸	۰/۱۰	فشار خون دارند
۰/۸۰	۰/۲۰	۰/۴۵	۰/۱۵	فشار خون ندارند
۱/۰۰	۰/۲۲	۰/۵۳	۰/۲۵	مجموع

ب) احتمال اینکه فردی که بتصادف از این گروه برگزیده شده و وزن بیش از حد معمول دارد، مبتلا به فشار خون باشد، چقدر است؟

$$P(A|B)=\frac{۰/۱۰}{۰/۲۵}=۰/۴$$

قانون ضرب را می‌توان برای بیش از دو پیشامد نیز تعمیم داد.

$$P(ABC)=P(A)P(B|A)P(C|AB)$$

مثال: کشاورزی سبدی شامل ۳۰ تخم‌مرغ دارد که ۵ تا از آنها دارای لکه‌های خونی هستند. سه تخم‌مرغ را به طور تصادفی یکی بعد از دیگری از سبد برداشته و بررسی می‌کند. این احتمال که دو تخم‌مرغ اول دارای لکه‌های خونی باشند و تخم‌مرغ سوم سالم باشد چقدر است؟

فرض کنید:

S_1 = پیشامد گزینش تخم مرغ لکه دار در استخراج اول

S_2 = پیشامد گزینش تخم مرغ لکه دار در استخراج دوم

C_3 = تخم مرغ سالم در استخراج سوم

احتمال اشتراک $(S_1S_2C_3)$ این سه پیشامد را بیابید:

$$P(S_۱C_۲C_۳)=P(S_۱)P(S_۲ \mid S_۱)\,P(C_۳ \mid S_۱S_۲)$$

چون استخراج ها به طور تصادفی انجام می‌گیرند، داریم

$P(S_۱)=\frac{۵}{۳۰}$.

$$P(S_۲|S_۱)=\frac{۴}{۲۹} \qquad P(C_۲|S_۱S_۲)=\frac{۲۵}{۲۸}$$

$$P(S_۱C_۲C_۳)=\frac{۵}{۳۰}\times\frac{۴}{۲۹}\times\frac{۲۵}{۲۸}=\frac{۲۵}{۱۲۱۸}$$

$$P(S_۱C_۲C_۳)=\frac{P_{۲}^{۵}\times P_{۱}^{۲۵}}{P_{۳}^{۳۰}}=\frac{۵\times ۴\times ۲۵}{۳۰\times ۲۹\times ۲۸}=\frac{۲۵}{۱۲۱۸}$$

دو پیشامد A و B مستقل اند اگر

$$P(A|B)=P(A)$$

شرطهای زیر، هم ارز با شرط بالا هستند:

$$P(B|A)=P(B)$$

یا

$$P(AB)=P(A)P(B)$$

مثال: آزمایش دوبار پرتاب یک سکه سالم را در نظر گیرید . از تقارن آزمایش چنین برمی آید که برآمدهای $\{HH,HT,TH,TT\}$ همشانس هستند. سه پیشامد زیر را تعریف می‌کنیم:

A: آمدن شیر در پرتاب اول

B: آمدن شیر در پرتاب دوم

C: آمدن شیر در دوپرتاب یا آمدن خط در دو پرتاب

کدام زوج از پیشامدهای فوق مستقل از هم‌اند؟

پیشامدهای A و B مستقل از هم‌اند. چون هر برآمد مقدماتی در S دارای احتمالی برابر با $\frac{۱}{۴}$ است، داریم

$$P(A)=\frac{۱}{۲} \qquad P(AB)=\frac{۱}{۴}$$

$$P(B)=\frac{۱}{۲} \qquad P(AC)=\frac{۱}{۴}$$

$$P(C)=\frac{۱}{۲} \qquad P(BC)=\frac{۱}{۴}$$

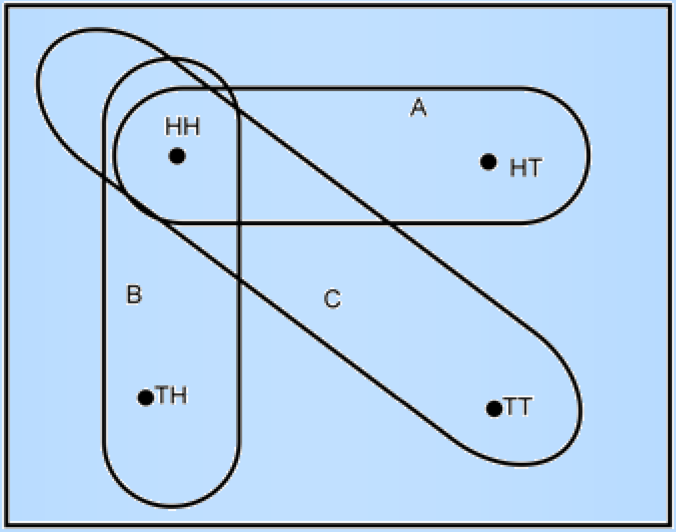
با توجه به این احتمال‌ها داریم:

$$P(A)P(B)=\frac{۱}{۴}=P(AB)$$

$$P(A)P(C)=\frac{۱}{۴}=P(AC)$$

$$P(B)P(C)=\frac{۱}{۴}=P(BC)$$

بنابراین برطبق فرمول ضربی استقلال، هر سه زوج از پیشامدهای فوقۀ مستقل اند.



نمودار ون برای دوبار پرتاب یک سکه

مثال: آیا دو پیشامد

B={وزن بیش از حد معمول}   و A={مبتلا به فشار خون}

از جدول زیر مستقل از هم‌اند؟

	وزن بیشتر از حد معمول	وزن معمولی	وزن کمتر از حد معمول	مجموع
فشار خون دارند	۰/۱۰	۰/۰۸	۰/۰۲	۰/۲۰
فشار خون ندارند	۰/۱۵	۰/۴۵	۰/۲۰	۰/۸۰
مجموع	۰/۲۵	۰/۵۳	۰/۲۲	۱/۰۰

$$P(A)=\text{۰/۲}$$

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{\text{۰/۱۰}}{\text{۰/۲۵}}=\text{۰/۴}$$

دوپیشامد مستقل نیستند.

مثال: یک آزمون بالینی ناکامل: یک بیماری خونی در ۲٪ از جامعه‌ای به طور شدید و در ۱۰٪ آن به طور خفیف وجود دارد، و در ۸۸٪ باقیمانده به هیچ وجه وجود ندارد. یک آزمون‌بالینی ناکامل در تشخیص بیماری در حالت‌های شدید، ۹۰٪ موفقیت‌آمیز است. به عبارت دیگر، اگر شخصی شدیداً بیمار باشد احتمال اینکه آزمون بالینی مثبت باشد ۰/۹ است و احتمال اینکه آزمون بالینی منفی باشد ۰/۱ است. بعلاوه، نرخ موفقیت آزمون بالینی برای تشخیص بیماری در حالتهای خفیف ۰/۶ می‌باشد. همچنین احتمال اینکه آزمون در مورد افراد سالم به نتیجه مثبت منجر گردد برابر با ۰/۱ است. آزمون در مورد شخصی که به طور تصادفی از این جامعه برگزیده شده انجام می‌شود و نتیجه مثبت است. احتمال اینکه این شخص شدیداً بیمار باشد چقدر است؟

چند علامت مناسب برای نشان دادن پیشامدهای مختلف

+: آزمون مثبت

-: آزمون منفی

S: حالت شدید

L: حالت خفیف

N: سالم

احتمال‌های موجود در مسئله:

$P(S)=۰/۰۲$ $P(+|S)=۰/۹$

$P(L)=۰/۱۰$ $P(+|L)=۰/۶$

$P(N)=۰/۸۸$ $P(+|N)=۰/۱$

$$P(S\mid +)=\frac{P(S+)}{P(+)}$$

$$P(S+)=P(S)P(+\mid S)=۰/۰۲\times۰/۹=۰/۰۱۸$$

$$P(L+)=P(L)P(+\mid L)=۰/۱۰\times۰/۶=۰/۰۶۰$$

$$P(N+)=P(N)P(+\mid N)=۰/۸۸\times۰/۱=۰/۰۸۸$$

پیشامدهای +S، +L و +N، سه پیشامد جدا از هم هستند که اجتماع آنها پیشامد + است.

$$P(+) = P(+S) + P(L+) + P(N+)$$

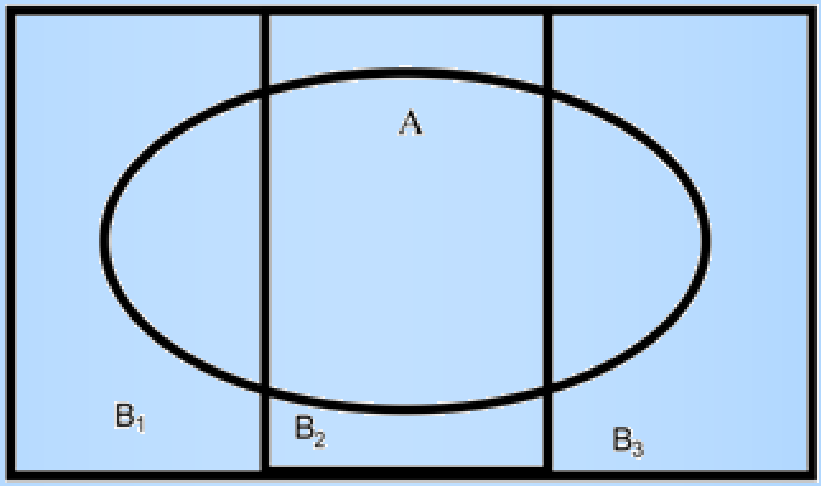
$$= ۰/۰۱۸ + ۰/۰۶۰ + ۰/۰۸۸$$

$$= ۰/۱۶۶$$

$$P(S \mid +) = \frac{P(S+)}{P(+)} = \frac{۰/۰۱۸}{۰/۱۶۶} = ۰/۱۱$$

ساختار قضیه بیز

$$k=۳$$



$$P(B_{۱} \mid A) = \frac{P(AB_{۱})}{P(A)}$$

$$P(AB_j) = P(B_j)P(A \mid B_j)$$

قضیه بیز

$$P(B_۱ | A) = \frac{P(B_۱)P(A | B_۱)}{\sum_{j=۱}^{k} P(B_j)P(A | B_j)}$$

در ابتدا محقق از چند وضعیت‌طبیعی ممکن B_1 ، B_2، ...، B_k بااطلاع است ولی نمی‌داند که کدامیک از آنها براستی پیش می‌آید بر مبنای دانش موجود درباره‌ی یک وضعیت، یا براساس یک گواه آزمایشی که از وضعیت‌های مشابه به دست آمده‌است، محقق ممکن است درباره احتمال‌های $P(B_1)$، $P(B_2)$، ...، $P(B_k)$ ارزیابی‌هایی نماید که در واقع بازتابی از احساس شخصی او در مورد میزان محتمل بودن هر یک از وضعیتهای طبیعی است. چنین احتمال‌هایی را **احتمال پیشین** یا **پیش از آزمایش** برای وضعیت‌های طبیعی گویند.

سپس محقق به انجام مشاهده یا اجرای آزمایش می‌پردازد و داده‌ها را گردآوری می‌کند. او می‌تواند احتمال گواه آزمایشی A را بــه شرط وقوع هر وضعیت مشهود B_j تعیین کند. آنگاه قضیه بیز به محقق امکان می‌دهد که احتمال‌هــای شــرطی $P(B_j|A)$ ، ($j=1,...,k$) را محاسبه نماید، که این کار، چیزی نیست جز نوعی تجدیــد نظــر در احتمال‌های وضعیت‌های طبیعی مختلف، بعد از بدست‌آمــدن گــواه آزمایشی. این احتمال‌های تجدید نظر شده را احتمال‌های پسین یــا پس از آزمایش گویند.

قواعد کلی احتمال

احتمال P، تابعی است با مقادیر عـددی کـه روی پیـشامدهای موجود در یک فضای نمونه S تعریف می‌شـود و در شـرایط زیـر صدق می کند:

یک) برای تمام پیشامدهای A، $0 \leq P(A) \leq 1$

دو) $P(S)=1$

سه) برای پیشامدهای جدا از هم A_1، A_2، ...

$$P(A_1 \cup A_2 \cup \ldots)=P(A_1)+P(A_2)+\ldots$$

سه قانون مهم احتمال برای یک فضای نمونه در حالت کلی

$$P(A) = ۱ - P(A^c)$$

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

$$P(AB) = P(B)P(A \mid B)$$

فصل 4

متغیرهای تصادفی و توزیع‌های احتمال

متغیرهای تصادفی

متغیر تصادفی X تابعی با مقادیر عددی است که روی فضای نمونه تعریف می‌شود. به عبارت دیگر، به هر پیشامد ساده e از فضای نمونه، یک عدد X(e)، که اندازه‌ای از مشخصه‌ی مورد نظر را به دست می‌دهد نسبت داده می‌شود.

مثال: فرض کنید که دو محصول A و B به وسیله چهار مصرف کننده از لحاظ برتری یکی بر دیگری سنجیده می‌شوند.

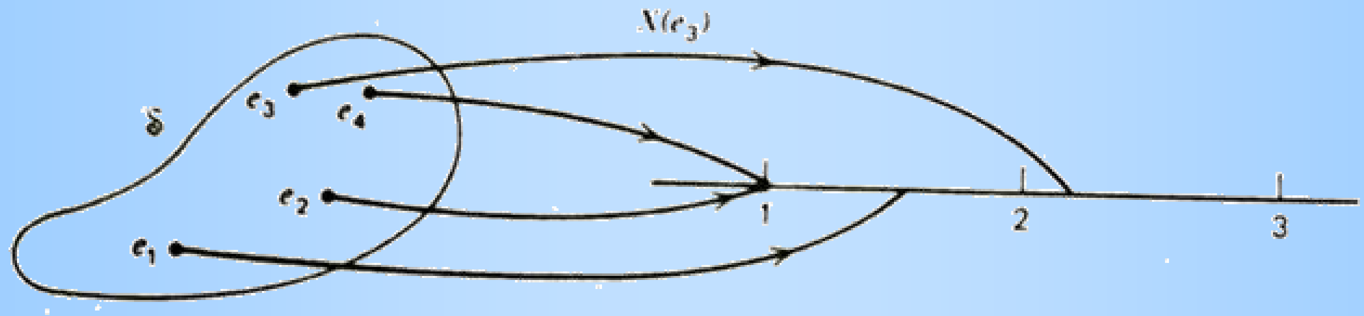
مصرف کننده دوم و چهارم B را ترجیح می‌دهد.

ABAB

مصرف کننده اول و سوم A را ترجیح می‌دهد.

تعداد برآمدها $۱۶ = ۲^{۴} = ۲\times۲\times۲\times۲$ است و عبارتند از:

AAAA	AAAB	AABB	ABBB	BBBB
	AABA	ABAB	BABB	
	ABAA	ABBA	BBAB	
	BAAA	BAAB	BBBA	
		BBAA		
		BABA		



هر پیشامد ساده ی e متشکل از دنباله ای از چهار نماد A یا B است.

توزیع احتمال متغیر X(تعداد اشخاصی که A را به B ترجیح می‌دهند)

مقادیر متمایز X	۰	۱	۲	۳	۴
احتمال	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

$$P[X\geq 2]=\frac{6}{16}+\frac{4}{16}+\frac{1}{16}=\frac{11}{16}$$

$$P[1\leq X\leq 3]=\frac{4}{16}+\frac{6}{16}+\frac{4}{16}=\frac{14}{16}$$

توزیع احتمال یک متغیر تصادفی گسسته، یا به طور خلاصه، **توزیع یک متغیر تصادفی عبارت است از فهرست مقادیر متمایز x_i از متغیر تصادفی X همراه با احتمال منسوب به هر یک از این مقادیر، $f(x_i) = P[X = x_i]$. اغلب می‌توان به جای استفاده از یک فهرست مفصل، از یک فرمول استفاده کرد.**

جدول توزیع احتمال

مقادیر متمایز X	x_1	x_2	$\dots$	x_k
احتمال	$f(x_1)$	$f(x_2)$	$\dots$	$f(x_k)$

$$\sum_{i=1}^{k} f(x_i) = 1$$

مثال: نمونه‌گیری از دسته‌ای کالای سالم و خراب:

یک فروشگاه بزرگ ۱۵ رادیوی ساعتی باقیمانده خود را برای فروش به حراج می‌گذارد. هیچکس نمی‌داند که ۵ تا از این رادیوها خراب است. اگر خریداری سه رادیوی ساعتی مختلف را که به طور تصادفی برگزیده آزمایش کند، توزیع متغیر تصادفی X را که عبارت است از تعداد رادیوهای خراب در نمونه، پیدا کنید.

حل: چون فقط تعداد کل رادیوهای خراب در نمونه مورد نظر است، ترتیب قرار گرفتن رادیوهای سالم و خراب مطرح نیست.

$$f(۰)=P[X=۰]=\frac{\binom{۱۰}{۳}}{\binom{۱۵}{۳}}=\frac{۱۲۰}{۴۵۵}=\frac{۲۴}{۹۱}$$

$$f(۱)=P[X=۱]=\frac{\binom{۱۰}{۲}\binom{۵}{۱}}{\binom{۱۵}{۳}}=\frac{۴۵\times۵}{۴۵۵}=\frac{۴۵}{۹۱}$$

$$f(۲)=P[X=۲]=\frac{\binom{۱۰}{۱}\binom{۵}{۲}}{\binom{۱۵}{۳}}=\frac{۱۰\times۱۰}{۴۵۵}=\frac{۲۰}{۹۱}$$

$$f(۳)=P[X=۳]=\frac{\binom{۵}{۳}}{\binom{۱۵}{۳}}=\frac{۱۰}{۴۵۵}=\frac{۲}{۹۱}$$

جدول احتمال تعداد رادیوهای خراب

مقدار x	۰	۱	۲	۳
احتمال $f(x)$	$\frac{۲۴}{۹۱}$	$\frac{۴۵}{۹۱}$	$\frac{۲۰}{۹۱}$	$\frac{۲}{۹۱}$

مثال: شانس اینکه میمونها طی انجام یک آزمایش واکنش مثبت(موفقیت‌آمیز) نشان دهند برابر با p است. انجام این آزمایش را تا مشاهده یک واکنش مثبت تکرار می‌کنیم، و متغیر تصادفی X را به عنوان انجام آزمایش تعریف می کنیم. فضای نمونه از پیشامدهای ساده‌ای تشکیل شده است، که هر یک عبارت از تعدادی F(شکست) و به دنبال آنها فقط یک S(موفقیت) است، به عبارت دیگر:

$$\mathcal{S}=\{S,FS,FFS,FFFS,\ldots\}$$

فرض این است که آزمایش‌ها مستقل‌اند.

توزیع احتمال تعداد آزمایشهای لازم برای به دست آوردن یک واکنش مثبت

مقدار x	۱	۲	۳	۴	...
احتمال $f(x)$	$\frac{1}{3}$	$\left(\frac{2}{3}\right)\frac{1}{3}$	$\left(\frac{2}{3}\right)^{2}\frac{1}{3}$	$\left(\frac{2}{3}\right)^{3}\frac{1}{3}$	...

$$f(x)=P[X=x]=\left(\frac{2}{3}\right)^{x-1}\times\frac{1}{3} \qquad x=1,2,\ldots$$

$$P[X\geq 3]=1-\frac{1}{3}-\frac{2}{3}\times\frac{1}{3}=\frac{4}{9}$$

نمایش توزیع احتمال به صورت نمودار

- نمودار خطی توزیع احتمال
- بافت نگار احتمال توزیع

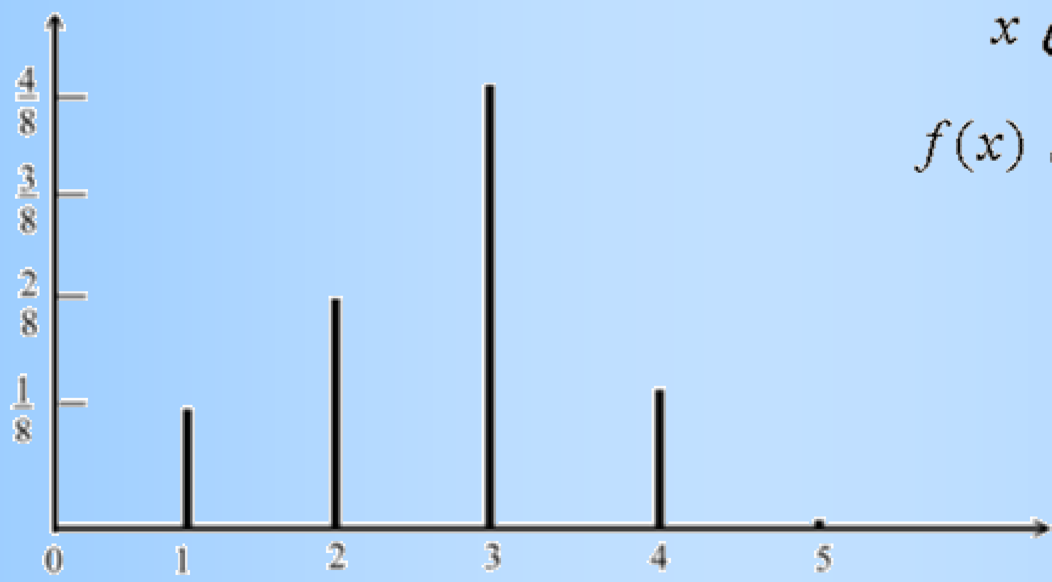
جدول توزیع احتمال

مقدار x	۱	۲	۳	۴
احتمال $f(x)$	$\frac{۱}{۸}$	$\frac{۱}{۴}$	$\frac{۱}{۲}$	$\frac{۱}{۸}$

نمودار خطی توزیع احتمال

محور افقی متغیر تصادفی x

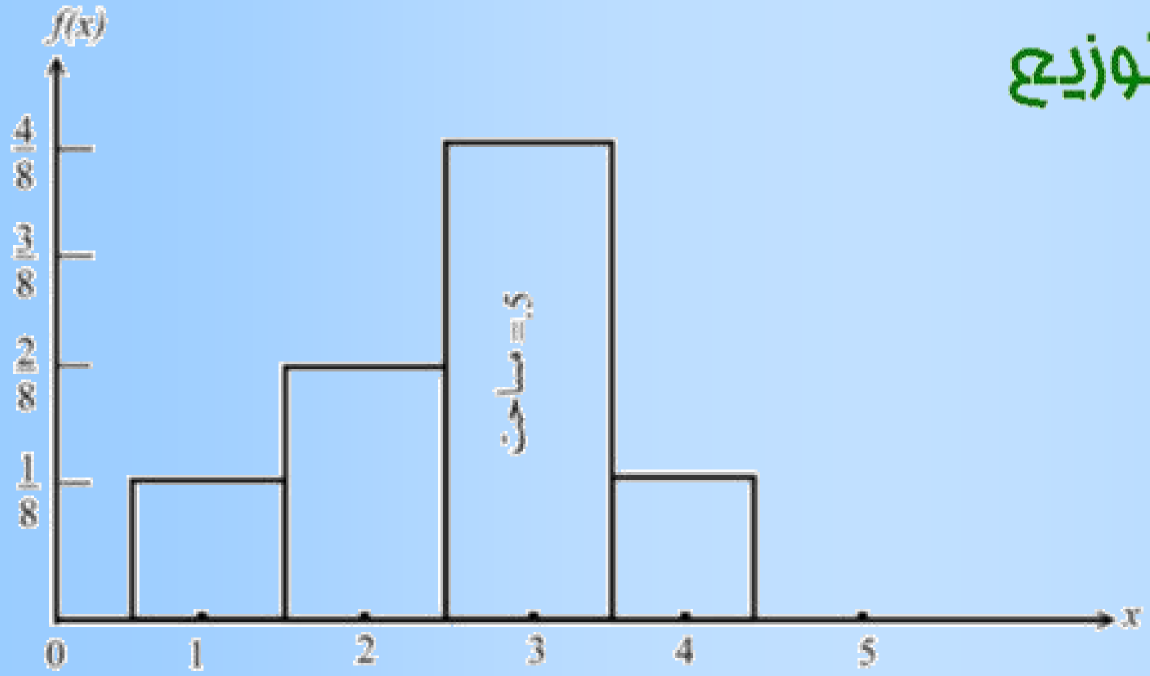
محور عمودی برابر احتمال $f(x)$



جدول توزیع احتمال

مقدار x	۱	۲	۳	۴
احتمال $f(x)$	$\frac{۱}{۸}$	$\frac{۱}{۴}$	$\frac{۱}{۲}$	$\frac{۱}{۸}$

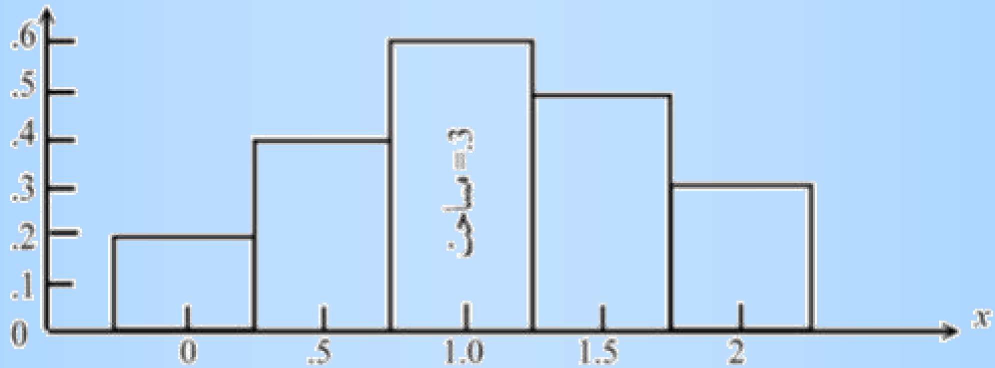
بافت نگار احتمال توزیع



مساحت=عرض × ارتفاع =۰/۵×۰/۶ = ۰/۳

جدول توزیع احتمال

مقدار x	۰	۰/۵	۱/۰	۱/۵	۲/۰
احتمال $f(x)$	۰/۱	۰/۲	۰/۳	۰/۲۵	۰/۱۵



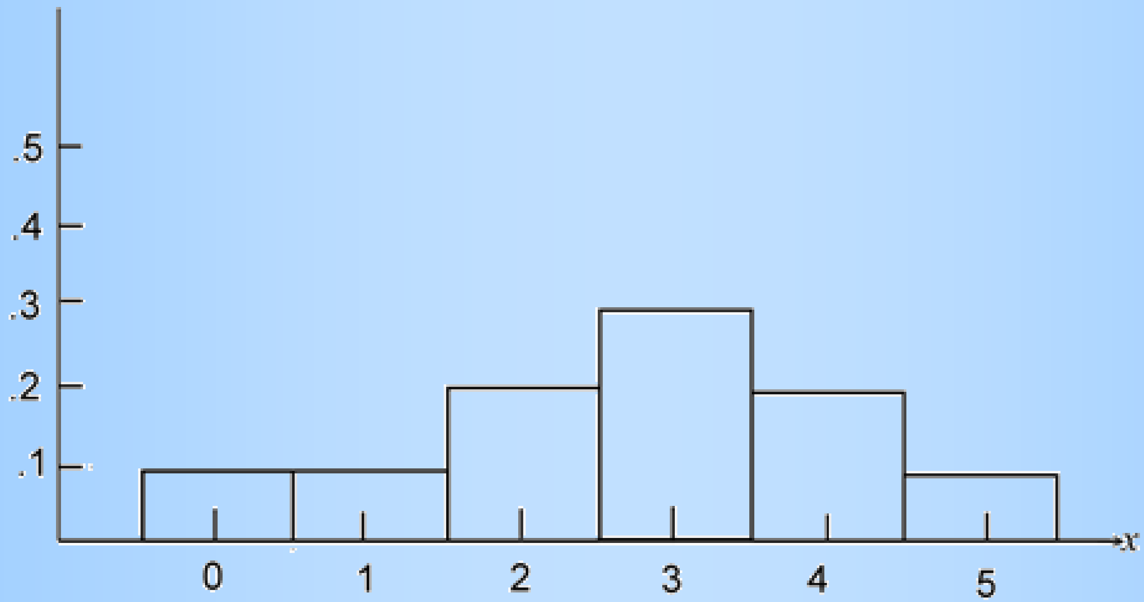
بافت نگار احتمال توزیع داده های جدول

توصیه می‌شود که بافت‌نگار احتمال برای توزیع‌هایی به کار رود که در آنها، فواصل بین xها برابر باشند. در حالتی که فواصل بین x ها نابرابرند باید از نمودار خطی استفاده کرد.

مثال: مدیر فروشگاهی با توجه به فیشهای فروش دو نوع کالای A و B که در مدت زمان طولانی گردآوری شده، توزیعهای احتمال فروش آنها را تعیین می کند. توزیع های احتمال X(تعداد کالای A) و Y(تعداد کالای B)ای که در مدت یک هفته فروخته شده، در جدول مقابل آمده است. با رسم بافت نگار های احتمال، آن دو را مقایسه کنید.

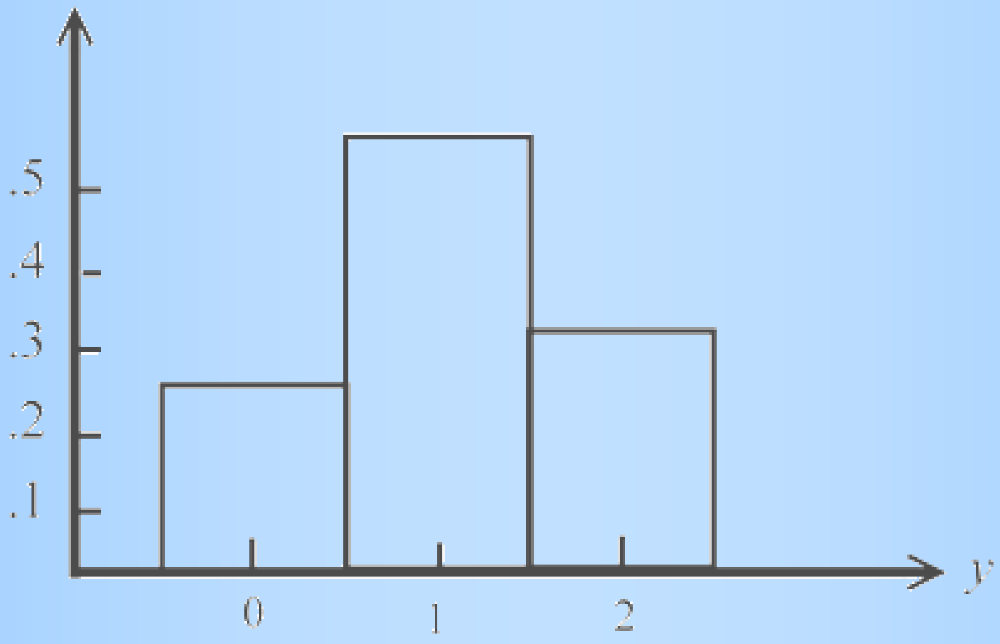
نوع A	مقدار x	۰	۱	۲	۳	۴	۵
	احتمال $f(x)$	۰/۱	۰/۱	۰/۲	۰/۳	۰/۲	۰/۱
نوع B	مقدار y	۰	۱	۲			
	احتمال $f(y)$	۰/۲۳	۰/۴۸	۰/۲۹			

نوع A	مقدار x	۰	۱	۲	۳	۴	۵
	احتمال $f(x)$	۰/۱	۰/۱	۰/۲	۰/۳	۰/۲	۰/۱



نوع A

نوع B	مقدار y	۰	۱	۲
	احتمال $f(y)$	۰/۲۳	۰/۴۸	۰/۲۹



نوع B

امید ریاضی و خواص آن

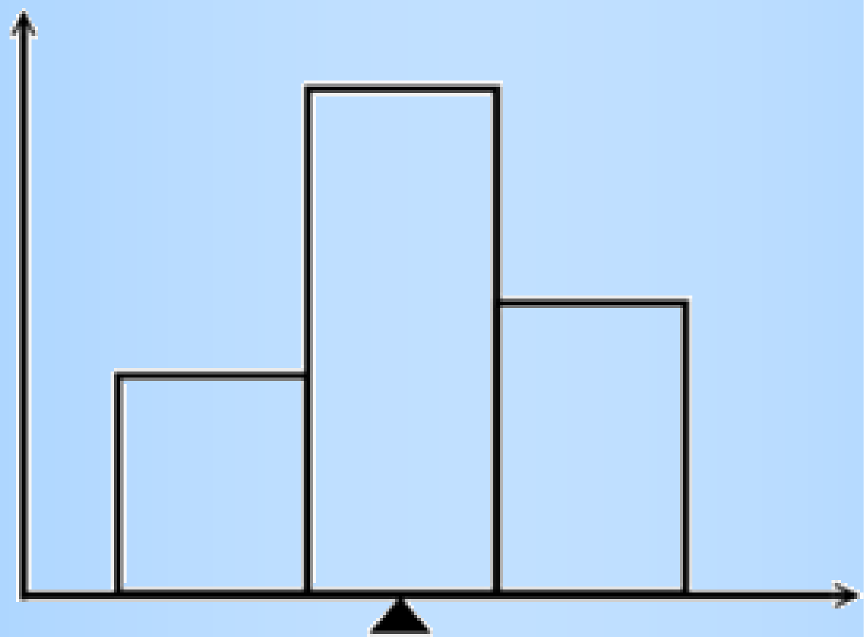
میانگین X، امید ریاضی X نیز نامیده می‌شود و با E(X) نشان داده می‌شود.

امید ریاضی X عبارت است از :

$$E(X)=\sum_{\text{تمام مقادیر متمایز } X} x_i f(x_i)$$

میانگین جامعه برای متغیر تصادفی X با μ یا μ_x نشان می‌شود.

$$\mu = E(X)$$



تعبیر $E(X)$ به عنوان مرکز ثقل

مثال: *توضیح اصلاح امید ریاضی:* در یک بازی به یک بازیکن اجازه می دهند که تاس سالمی را بریزد و به تعداد نقاط روس تاس، پول به تومان دریافت کند. بازیکن مجبور است برای هر دفعه بازی کردن c تومان بپردازد. بدون در نظر گرفتن این مبلغ، درآمد بازیکن در هر دفعه بازی متغیری است تصادفی با توزیع احتمالی که در جدول مقابل داده شده است.

درآمد x	۱	۲	۳	۴	۵	۶
احتمال	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۶}$

بازی عادلانه است اگر $c=$ ۳/۵ باشد زیرا در این صورت متوسط برد درصورت تکرار بازی به دفعات زیاد، برابر صفر است. اگر بازیکن برای شرکت در هر بازی مبلغ ۴ تومان بپردازد، امید ریاضی باخت خالص بازیکن ۰/۵ تومان در هر بازی است.

فرض کنید به دلایلی متغیر تصادفی بصورت $g(X)=(X-۲)^۲$ باشد که تابعی از X است. به دو روش امید ریاضی را می توان محاسبه کرد:

روش اول: توزیع احتمال $(X-۲)^۲$ را با استفاده از توزیع احتمال X محاسبه شود و سپس نتایج را در دو سطر نوشته و به روش حاصلضرب‌های اجزای متناظر آن را محاسبه کرد:

مقادیر متمایز $(X-۲)^۲$	۰	۱	۴	مجموع
احتمال	$\frac{۳}{۸}$	$\frac{۴}{۸}$	$\frac{۱}{۸}$	۱
احتمال × مقدار	۰	$\frac{۴}{۸}$	$\frac{۴}{۸}$	۱

$$E[(X-۲)^۲]=۱$$

روش دوم: مستقیماً از توزیع احتمال X استفاده شود. مقادیر توزیع احتمال را در دو سطر محاسبه کرده در احتمال X=x ضرب شود آنگاه از جمع کردن حاصل ضرب‌ها امید ریاضی به دست می آید .

مقادیر متمایز X	۰	۱	۲	۳	۴	مجموع
$f(x)$	$\frac{۱}{۱۶}$	$\frac{۴}{۱۶}$	$\frac{۶}{۱۶}$	$\frac{۴}{۱۶}$	$\frac{۱}{۱۶}$	۱
$(x-۲)^۲$	۴	۱	۰	۱	۴	
$(x-۲)^۲f(x)$	$\frac{۴}{۱۶}$	$\frac{۴}{۱۶}$	۰	$\frac{۴}{۱۶}$	$\frac{۴}{۱۶}$	۱

$$E[(X-۲)^{۲}]=۱$$

امید ریاضی یک تابع $g(X)$ از متغیر تصادفی X به صورت مقابل تعریف می شود.

$$E(X)=\sum_{x} g(x_i)f(x_i)$$

تمام مقادیر متمایز x

خواص امید ریاضی

$$E(a) = a$$

$$E(bX) = bE(X)$$

$$E(X + a) = E(X) + a$$

$$E(a + bX + cX^{2}) = a + bE(X) + cE(X^{2})$$

مثال: فرض کنید X تعداد واحدهای کالای نوع A باشد که در یک هفته فروخته می‌شود و توزیع احتمال آن در جدول مقابل داده شده‌است و فرض کنید که برای هر واحد فروخته شده ۵۰ تومان سود به دست آمده و هزینه ثابت هفتگی فروشگاه ۲۰ تومان است. امید ریاضی سود چقدر است؟

حل: سود حال از فروش X واحد در یک هفته، $۵۰X$ تومان است.

با کم کردن هزینه ثابت، سود $۵۰X-۲۰$ است.

$$E(۵۰X-۲۰)=۵۰E(x)-۲۰$$

$$=۵۰\times۲/۷-۲۰$$

$$=۱۱۵$$ تومان

واریانس: معیاری برای پراکندگی

واریانس X عبارت است از :

$$Var(X) = E[(X-\mu)^2] = E(X^2) - \mu^2$$

واریانس X با σ^2 یا σ_X^2 نیز نشان داده می‌شود.

انحراف معیار X عبارت است از:

$$sd(X) = \sqrt{Var(X)}$$

انحراف معیار X با σ یا σ_X نیز نشان داده می شود.

مثال: محاسبه واریانس: فروش هفتگی کالای نوع A

x	۰	۱	۲	۳	۴	۵	مجموع
$f(x)$	۰/۱	۰/۱	۰/۲	۰/۳	۰/۲	۰/۱	۱
$xf(x)$	۰	۰/۱	۰/۴	۰/۹	۰/۸	۰/۵	۲/۷=$E(X)$
$x^{۲}f(x)$	۰	۰/۱	۰/۸	۲/۷	۳/۲	۲/۵	۹/۳=$E(X^{۲})$

$\mathrm{Var}(X)$=۹/۳−(۲/۷)^{۲}=۲/۰۱

$\mathrm{sd}(X)$=$\sqrt{۲/۰۱}$=۱/۴۲

خواص

واریانس	انحراف معیار
Var(X) نمی تواند منفی باشد	Sd(X) نمی تواند منفی باشد
$Var(X+a)=Var(X)$	$sd(X+a)=sd(X)$
$Var(bX)=b^{2}Var(X)$	$sd(bX)=\lvert b\rvert sd(X)$
$Var(a+bX)=b^{2}Var(X)$	$sd(a+bX)=\lvert b\rvert sd(X)$

برای متغیر تصادفی استاندارد شده‌ی $Z = \frac{(X - \mu_x)}{\sigma_x}$ داریم

$E(Z) = ۰$ و $Var(Y) = ۱$

مثال: با توجه به مثال قبل داشتیم X برابر با تعداد واحدهای کالای نوع Aی فروخته شده در هفته بود و برای هر واحد فروخته شده ۵۰ تومان سود به دست آمد. با کم کردن هزینه ثابت ۲۰ تومان، سود خالص هفتگی از فروش این نوع کالا، متغیر تصادفی $Y=50X-20$ می‌باشد. پس

$E(X)=50E(X)-20=115$ تومان

$Var(Y) =(50)^2 Var(X)=(50)^2(2/01)=5025$

$sd(Y) = |50| sd(X) =(50)(1/42) =71$ تومان

توزیع توام دو متغیر تصادفی

		y_1	y_2	$\cdots$	y_l
	x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\cdots$	$f(x_1, y_l)$
مقادیر X	x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\cdots$	$f(x_2, y_l)$
	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	x_k	$f(x_k, y_1)$	$f(x_k, y_2)$	$\cdots$	$f(x_k, y_l)$

توزیع احتمال توام دو متغیر تصادفی گسسته X و Y با جـــدولی مشخص می‌شود که در آن، مقادیر X و Y در دو کنــــاره جــــدول نشان داده می‌شوند و احتمال‌های مربوط به زوج‌های مقـــادیر X و Y در خانه‌ها قرار می‌گیرند. احتمال‌های خانه‌ها را اغلب به جـــای اینکه در جدول بیاورند به صورت فرمولی ارائه می‌دهند.

مثال: ده ماشین وجود دارند که پنج عـدد از آنهـا در وضـعیت خـوب (G) هستند، دو عدد دارای نقص در قـسمت انتقـال قـدرت(DT)، و سـه عـدد دیگر دارای نقص در مکـانیزم فرمان(DS) مـی‌باشـند. فـرض کنیـد کـه دو ماشین به‌طور تصادفی برگزیده شده‌اند، و متغیرهای تصادفی مـورد نظـر عبارت اند از X (تعداد ماشین‌هایی که در قـسمت انتقـال قـدرت نقـص دارند) و Y(تعداد ماشین‌هایی که در مکانیزم فرمان نقص دارنـد). توزیـع احتمال توام X و Y را به دست آوریم.

نمای نمونه گیری

۲	DT
۳	DS
۵	G

نمونه‌ی به حجم ۲

X= تعداد DT در نمونه

Y= تعداد DS در نمونه

توزیع توام

		مقادیر Y		
		۰	۱	۲
مقادیر X	۰	۱۰/۴۵	۱۵/۴۵	۳/۴۵
	۱	۱۰/۴۵	۶/۴۵	۰
	۲	۱/۴۵	۰	۰

برای مثال

$$f(0,0)=P[X=0,Y=0]=P[\text{هر دو G هستند}]=\frac{\binom{5}{2}}{\binom{10}{2}}=\frac{10}{45}$$

$$f(1,0)=P[X=1,Y=0]=P[\text{یکی DT و یکی G است}]=\frac{\binom{2}{1}\binom{5}{1}}{\binom{10}{2}}=\frac{10}{45}$$

$$f(1,2)=P[X=1,Y=2]=0$$

مثال: کارخانه کوچکی روزانه در دو نوبت کار می کند . در بررسی الگوی غیبت کارگران این کارخانه، متغیرهای مورد نظر عبارتند از: X(تعداد غایبین در نوبتکاری صبح) و Y(تعداد غایبین در نوبتکاری عصر همان روز). سرپرست کارگران بر مبنای تعداد زیادی از مشاهدات حضور و غیاب کارگران در گذشته، مقادیر توزیع احتمال توام X و Y را تعیین می‌کند که نتیجه آن به صورت مقابل است:

توزیع توام غایبین روزانه در دو نوبت کاری

x \ y	۰	۱	۲	۳	مجموع سطر
۰	۰/۰۵	۰/۰۵	۰/۱۰	۰	۰/۲۰
۱	۰/۰۵	۰/۱۰	۰/۲۵	۰/۱۰	۰/۵۰
۲	۰	۰/۱۵	۰/۱۰	۰/۰۵	۰/۳۰
مجموع ستون	۰/۱۰	۰/۳۰	۰/۴۵	۰/۱۵	۱

الف) احتمال هر پیشامد مربوط به X و Y:

PX=۲]=۰/۱۵+۰/۱۰+۰/۰۵=۰/۳۰

P[X>Y]=۰/۰۵+۰+۰/۱۵=۰/۲۰

ب) توزیع احتمال تابعی از دو متغیر تصادفی X و Y:

جدول توزیع Z=X+Y

z	۰	۱	۲	۳	۴	۵	مجموع
$f_Z(z)$	۰/۰۵	۰/۱۰	۰/۲۰	۰/۴۰	۰/۲۰	۰/۰۵	۱/۰۰

ج) *توزیع احتمال X بتنهایی*: این توزیع را می‌توان مستقیماً از توزیع توام X و Y بدست آورد. البته به توزیع X توزیع کناری X نیز می‌گویند که همان مجموع سطرهاست که در کناره سمت راست نوشته شده است.

د) *میانگین و انحراف معیار X یا Y*: برای توزیع X، هر گونه معیار عددی از قبیل امید ریاضی μ_X و واریانس σ^2_X تنها از روی توزیع کناری X و بدون مراجعه به توزیع توام محاسبه می‌شود. همین روش برای توزیع Y نیز صادق است.

ه) جمع‌پذیری امید ریاضی: امید ریاضی دارای خاصیت جمع‌پذیری است. یعنی برای دو متغیر تصادفی X و Y داریم:

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

x	۰	۱	۲	مجموع
$f_x(x)$	۰/۲	۰/۵	۰/۳	۱/۰
$xf_x(x)$	۰	۰/۵	۰/۶	۱/۱
$x^2 f_x(x)$	۰	۰/۵	۱/۲	۱/۷

y	۰	۱	۲	مجموع
$f_y(y)$	۰/۲	۰/۵	۰/۳	۱/۰
$yf_y(y)$	۰	۰/۵	۰/۶	۱/۱
$y^2 f_y(y)$	۰	۰/۵	۱/۲	۱/۷

μ_X =۱/۱

σ^2_X=۱/۷−(۱/۱)^2=۰/۴۹

σ_X=۰/۷۰

μ_Y =۱/۶۵

σ^2_Y=۳/۴۵−(۱/۶۵)^2=۰/۷۲۷۵

σ_Y=۰/۸۵

کوواریانس و همبستگی

کوواریانس بین X و Y عبارت است از:

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

$$E(XY) = \sum_{\text{تمام خانه ها}} (\text{احتمال خانه}) \times (\text{مقدار } XY \text{ برای هر خانه})$$

ضریب همبستگی بین X و Y عبارت است از :

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

مثال: محاسبه کوواریانس و همبستگی

x \ y	۰	۱	۲	۳
۰	۰/۰۵ (۰)	۰/۰۵ (۰)	۰/۱۰ (۰)	۰ (۰)
۱	۰/۰۵ (۰)	۰/۱۰ (۱)	۰/۲۵ (۲)	۰/۱۰ (۳)
۲	۰ (۰)	۰/۱۵ (۲)	۰/۱۰ (۴)	۰/۰۵ (۶)

$$\text{۰/۱۰}\times\text{۱}=\text{۰/۱۰}$$

$$\text{۰/۲۵}\times\text{۲}=\text{۰/۵۰}$$

$$\text{۰/۱۰}\times\text{۳}=\text{۰/۳۰}$$

$$\text{۰/۱۵}\times\text{۲}=\text{۰/۳۰}$$

$$\text{۰/۱۰}\times\text{۴}=\text{۰/۴۰}$$

$$\underline{\text{۰/۰۵}\times\text{۶}=\text{۰/۳۰}}$$

$$E(XY)=\text{۱/۹۰}$$

$$Cov(X,Y)=\text{۱/۹۰}-\text{۱/۶۵}\times\text{۱/۱}-\text{۰/۰۸۵}$$

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=\frac{\text{۰/۰۸۵}}{\text{۰/۷۰}\times\text{۰/۸۵}}=\text{۰/۱۴}$$

خواص همبستگی

الف) $Corr(X,Y)$ همیشه عددی بین ۱– و ۱+ است. فـــرین ۱– و ۱+ زمانی است که X و Y به وسیله خط مستقیمی بترتیب بـــا شـــیب منفی و یا مثبت به هم مربوط باشند.

ب) اگر مقادیر ثابتی با متغیرها جمع شـــوند یـــا اگـــر متغیرهـــا در مقادیری ثابت با علامت یکسان ضرب شوند، $Corr(X,Y)$ تغییری نمی‌کند.

مثال:

U=۵X+2

V=۲Y+۳

$$Corr(U,V)=Corr(X,Y)$$

استقلال دو متغیر تصادفی

متغیرهای تصادفی X و Y مستقل‌اند اگر برای تمام زوج‌های مقادیر (x_i, y_i) در توزیع توام، داشته باشیم $f(x_i, y_i) = f_X(x_i) f_Y(y_i)$. به عبارت دیگر، اگر هر احتمال توام برابر با حاصل ضرب دو احتمال کناری مربوطه باشد.

مثال: سکه ای قرار است ۳ بار پرتاب شود. فرض کنید:

X= تعداد شیرها در پرتابهای اول و دوم.

Y= تعداد شیرها در پرتاب سوم.

توزیع احتمال توام X و Y بصورت مقابل است. آیا X و Y مستقل اند؟

x \ y	۰	۱	مجموع سطر
۰	۱/۸	۱/۸	۱/۴
۱	۲/۸	۲/۸	۱/۲
۲	۱/۸	۱/۸	۱/۴
مجموع ستون	۱/۸	۱/۸	۱

X و Y مستقل نیستند چون $f(۰ , ۰) = ۱۰/۴۵$ است در حالی که

$$f_X(۰)f_Y(۰) = (۲۸/۴۵) \times (۲۱/۴۵) \neq (۱۰/۴۵)$$

اگر X و Y مستقل باشند داریم:

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

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

جمع‌پذیری امید ریاضی

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

فرمول‌های کلی برای واریانس مجموع و تفاضل:

$$Var(X+Y) = Var(X) + Var(Y) + ۲Cov(X,Y)$$

$$Var(X-Y) = Var(X) + Var(Y) - ۲Cov(X,Y)$$

اگر X و Y مستقل باشند،

$$Var(X+Y) = Var(X) + Var(Y)$$

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

فصل ۵

توزیع‌های متغیرهای تصادفی گسسته

ایده‌ی مدل احتمال

مدل احتمال متغیر تصادفی X، شکل معینی از توزیع احتمال است که فرض می‌شود رفتار X را نشان دهد. احتمال‌ها برحسب پارامترهای نامعلومی بیان می‌شوند که به مشخصه‌های جامعه و روش نمونه‌گیری مربوط‌اند.

خواص مطلوب یک مدل احتمال

کفایت: مدل باید مبتنی بر فرض‌هایی باشد که در مورد آن مکانیزم شانسی که موجب تغییر در مشاهدات می‌گردد، حاوی اطلاعات کافی باشند.

سادگی: در صورت امکان باید از تقریب‌هایی که مدل را ساده‌تر می‌کنند استفاده شود تا مدل، بدون آنکه کفایت آن لطمه بخورد، برای تحلیل آماری مناسب شود.

امساک در تعداد پارامترها: وجود تعدادی پارامتر زاید (در مدل مورد نظر)، کیفیت تحلیل آماری را پایین می‌آورد. در انتخاب بین دو مدل که هر دو تقریب واقعبینانه‌ای از رفتار شانسی مورد بحث هستند، برتری با مدلی که پارامترهای مجهول کمتری دارد.

امتحان های برنولی

الف) هر امتحان به یکی از دو برآمد ممکن می‌انجامد که در اصطلاح فنی موفقیت(S) و شکست(F) نامیده می شود.

ب) برای تمام امتحان‌ها ، احتمال موفقیت $P(S)$، یکی است و به صورت $P=P(S)$ نشان داده می‌شود. بنابراین احتمال شکست برای هر امتحان $P(F)=۱-p$ است که آن را با q نشان می‌دهیم، بطوری که $p+q=۱$.

ج) امتحان‌ها مستقل از یکدیگرند. احتمال موفقیت در یک امتحان با داشتن هر مقدار اطلاعات از برآمدهای سایر امتحان‌ها تغییر نمی‌کند.

ساده ترین مثال برای امتحان‌های برنولی پرتاب سکه سالم و ناریب است. شیر و خط را می‌توان بترتیب با S و F نامید: $p=q=۱/۲$

نمونه‌گیری از جامعه دوحالتی

الف) نمونه گیری با جایگذاری

جامعه شامل ۱۵ کالا

۵ کالا معیوب و ۱۰ کالا سالم

S معیوب بودن کالا

$$P(S)=p=\frac{۵}{۱۵}$$

نمونه‌گیری از جامعه دوحالتی

ب) نمونه گیری بدون جایگذاری

۳ کالا یکی بعد از دیگری بدون جایگذاری

برای اولین کالا $P(S)=\frac{۵}{۱۵}$

برای دومین کالا به شرط اینکه اولین گزینش به موفقیت(S) بیانجامد

احتمال شرطی دومین استخراج $P(S)=\frac{۵}{۱۵}$ است که شرط استقلال نقض

شده چون $\frac{۴}{۱۴} \neq \frac{۵}{۱۵}$

نمونه‌گیری از جامعه دوحالتی

چ) نمونه گیری بدون جایگذاری از جامعه بزرگ

مجموعه شامل ۱۵۰۰ کالا که ۵۰۰ کالای آن معیوب است

نمونه‌گیری بدون جایگذاری ۳ کالا

$$P(S_{۱}) = \frac{۵۰۰}{۱۵۰۰} = \frac{۵}{۱۵}$$

$$P(S_{۲} \mid S_{۱}) = \frac{۴۹۹}{۱۴۹۹}$$

اگر از جامعه دوحالتی نمونه‌ای تصادفی با روش جایگذاری بگیریم، شرایط امتحانهای برنولی برقرار خواهد بود. اگر نمونه‌گیری بدون جایگذاری انجام گیرد، شرط استقلال آزمایش‌ها نقض مـــی‌شـــود. لکن، اگر حجم جامعه در مقایسه با حجم نمونه خیلی بـــزرگ باشـــد تاثیر این نقض استقلال، قابل صـــرف نظـــر کـــردن اســـت و مـــدل امتحان‌های برنولی را می‌توان به عنوان یک تقریب در نظر گرفت.

توزیع دوجمله‌ای

توزیع احتمال X با n امتحان و احتمال موفقیت p

مثال: $n=۴$

FFFF	SFFF	SSFF	SSSF	SSSS
	FSFF	SFSF	SSFS	
	FFSF	SFFS	SFSS	
	FFFS	FSSF	FSSS	
		FSFS		
		FFSS		

مقدار X	۰	۱	۲	۳	۴
احتمال هر دنباله	$q^۴$	$pq^۳$	$q^۲p^۲$	$q^۳p$	$q^۴$
تعداد دنباله	$۱=\binom{۴}{۰}$	$۴=\binom{۴}{۱}$	$۶=\binom{۴}{۲}$	$۴=\binom{۴}{۳}$	$۱=\binom{۴}{۴}$

$$P[X=2]=P(SSFF)=P(S)P(S)P(F)P(F)= q^۲p^۲$$

X	۰	۱	۲	۳	۴
$P[X=x]$	$\binom{۴}{۰}p^{۰}q^{۴}$	$\binom{۴}{۱}p^{۱}q^{۳}$	$\binom{۴}{۲}p^{۲}q^{۲}$	$\binom{۴}{۳}p^{۳}q^{۱}$	$\binom{۴}{۴}p^{۴}q^{۰}$

$$P[X=x]=\binom{n}{x}p^{x}q^{n-x} \qquad x=\text{۱،۰،.....،n}$$

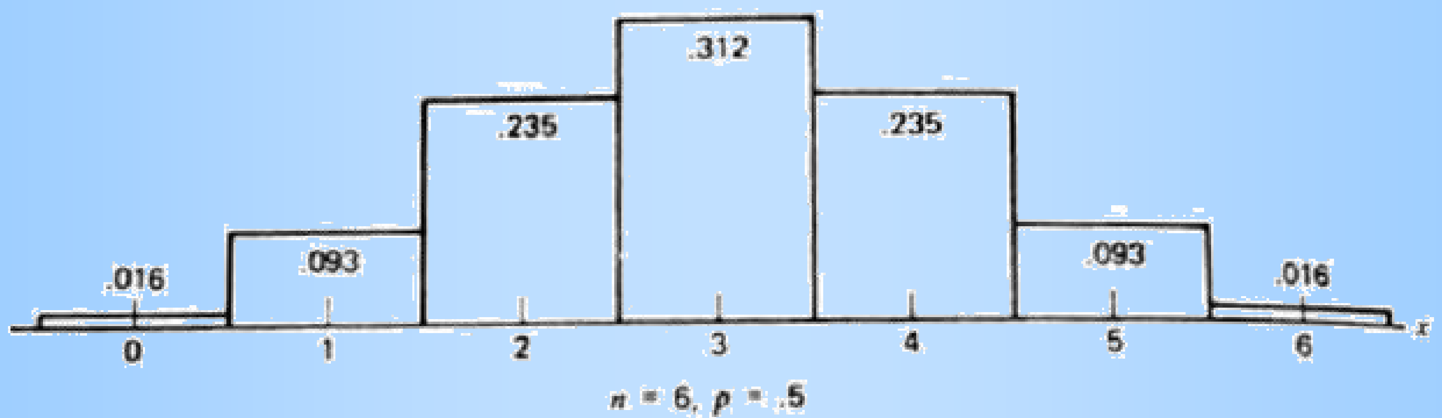
توزیع دوجمله‌ای

n امتحان $p=P(S)$

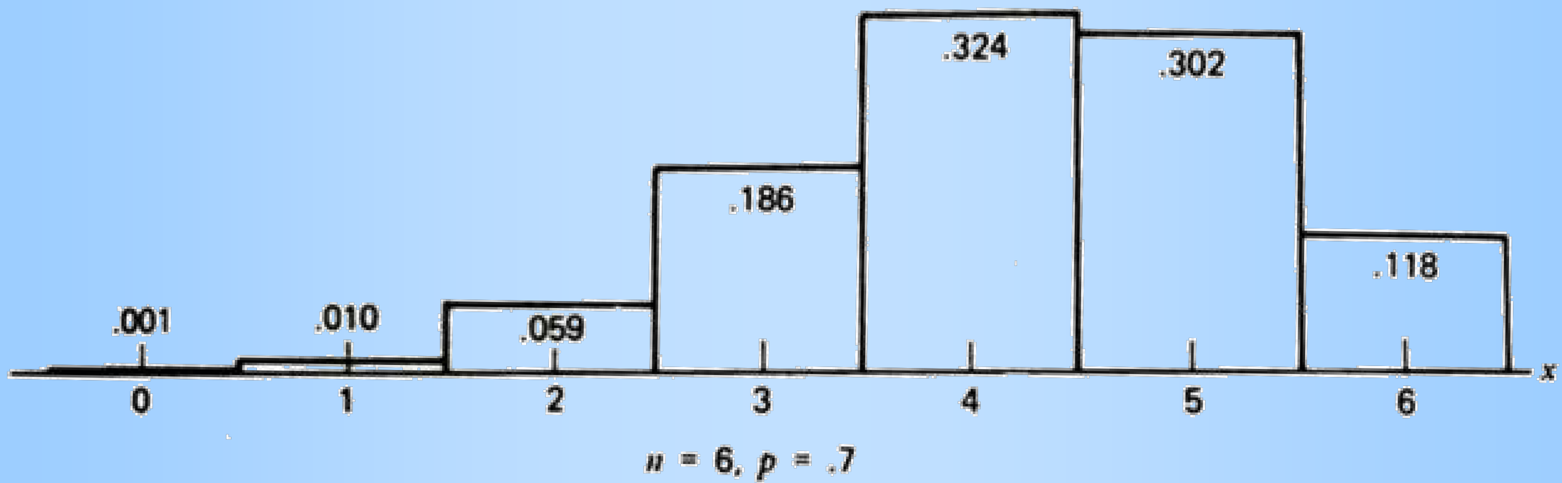
$$P[X=x]=b(x;n,p)=\binom{n}{x}p^{x}q^{n-x}$$

برای مقادیر ممکن n،...،۱،۰=x

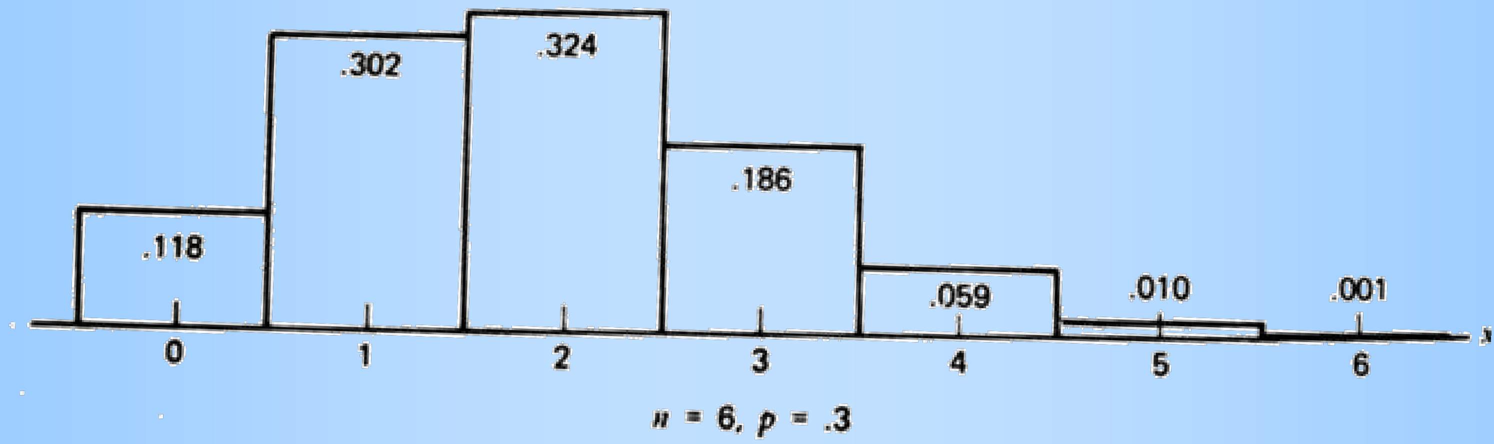
تاثیر مقادیر p در شکل توزیع



تاثیر مقادیر p در شکل توزیع



تاثیر مقادیر p در شکل توزیع



مثال: مطابق نظریه مندل درباره خصوصیات وراثتی بر اثر آمیزش مصنوعی بین گلهای قرمز و سفید از یک نوع، ۲۵ درصد گل قرمـز تولید می شود. فرض کنید باغبان می‌خواهد ۵ جفت از این گلهای قرمز و سفید را با یکدیگر بیامیزد احتمال اینکـه از ۵ گـل حاصـل

الف) هیچیک قرمز نباشد، چقدر است؟

ب) ۴ تا یا بیشتر قرمز باشند، چقدر است؟

طبیعی است که ۵ امتحان مستقل از هم‌اند و شرایط امتحان‌های برنولی برقرار است.

متغیر تصادفی X تعداد گل‌های قرمز. قرمز بودن را با S نشان می‌دهیم. $P(S)=p=\frac{۱}{۴}$

X دارای توزیع دوجمله‌ای با $n=۵$ و $p=\frac{۱}{۴}$ است.

$$P[X=۰]=b(۰)=\left(\frac{۳}{۴}\right)^{۵}=\frac{۲۴۳}{۱۰۲۴}=۰/۲۳۷$$

$$P[X\geq ۴]=b(۴)+b(۵)$$

$$=\binom{۵}{۴}\left(\frac{۱}{۴}\right)^{۴}\left(\frac{۳}{۴}\right)^{۱}+\binom{۵}{۵}\left(\frac{۱}{۴}\right)^{۵}\left(\frac{۳}{۴}\right)^{۰}$$

$$=\frac{۱۶}{۱۰۲۴}=۰/۰۱۶$$

چگونگی کاربرد جدول‌های دوجمله‌ای

جداول احتمال تجمعی $P[X \leq c] = \sum_{x=0}^{c} b(x; n, p)$ برای بعضی مقادیر n و p و c موجود است روش بدست آوردن سایر پیشامدها این است که آنها را بصورت $[X \leq c]$ بنویسیم . قوانین احتمال مناسب را به کار بریم.

$$P[X = x] = P[X \leq x] - P[X \leq (x-1)]$$

$$P[a \leq X \leq b] = P[X \leq b] - P[X \leq (a-1)]$$

$$P[X > x] = 1 - P[X \leq x]$$

مثال: فرض کنید شانس جوانه زدن نوع مخصوصی از دانه گیاه چنانچه در فصل معینی کاشته شود ۸۰درصد است. اگر ۱۵ دانه این گیاه را بکاریم تعیین کنید احتمال این را که:

الف) حداکثر ۸ دانه جوانه بزنند.

ب) ۱۰ دانه یا بیشتر جوانه بزنند.

ج) تعداد جوانه ها از ۸ کمتر و از ۱۲ بیشتر نباشد.

جوانه زدن با S نشان می‌دهیم و فرض می‌شود که نتایج مربوط به دانه‌ها مستقل از هم‌اند. X متغیر تصادفی برابر تعداد دانه‌هایی که جوانه می‌زنند. توزیع دوجمله‌ای $n=۱۵$ و $p=۰/۸$ برای متغیر X مناسب است.

الف) $P[X \leq ۸] = ۰/۰۱۸$

ب) $P[X \geq ۱۰] = ۱ - P[X \leq ۹] = ۰/۰۱۸$

ج) $P[۸ \leq X \leq ۱۲] = P[X \leq ۱۲] - P[X \leq ۷] = ۰/۶۰۲ - ۰۰۴ = ۰/۵۹۸$

توجه:

$$b(x;n,p) = b(n-x;n,p^*)$$

$$p^* = 1-p$$

میانگین و واریانس توزیع دوجمله‌ای

برای یک توزیع دوجمله ای با n امتحان؛ $p=P(S)$

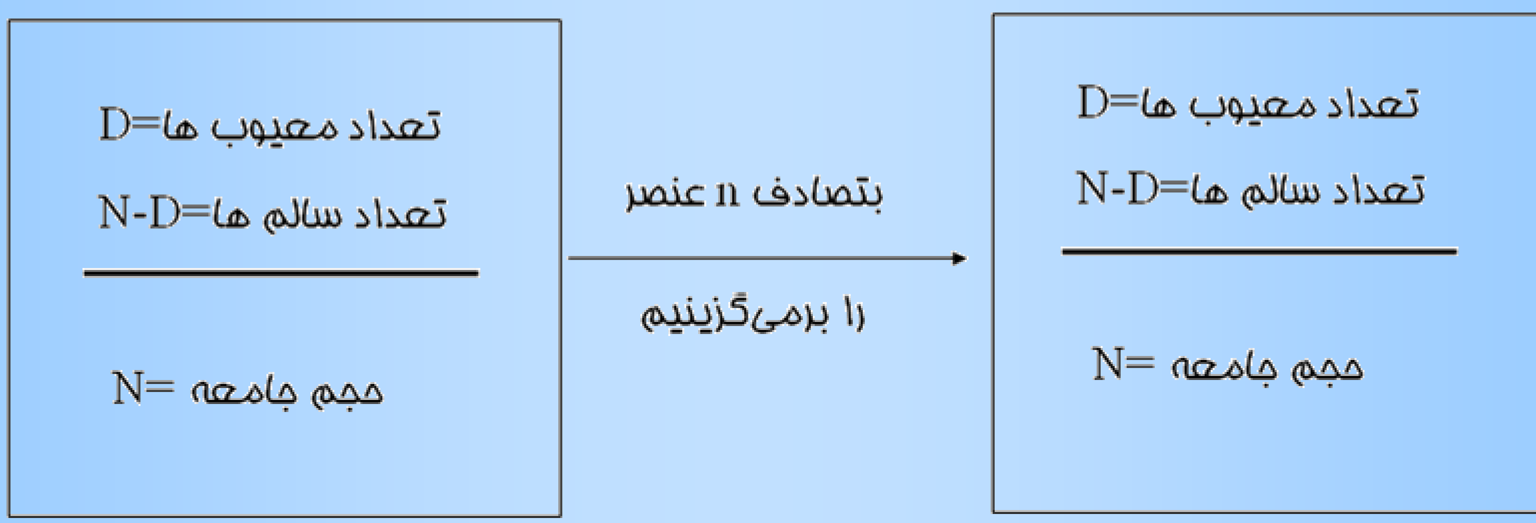
میانگین= np

واریانس= $np(۱-p)=npq$

$$sd=\sqrt{np(۱-p)}=npq$$

توزیع فوق هندسی

شمای نمونه‌گیری بدون جایگذاری



توزیع فوق هندسی

تعداد معیوب ها در جامعه=D

تعداد معیوب ها در نمونه=X

تعداد معیوب ها در جامعه=D

حجم نمونه=n

$$P[X=x]=\frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} \qquad x=0,1,\ldots,n$$

وقتی که $n \le D$ و $n \le N-D$

میانگین مساوی np، که در آن p=D/N (نسبت معیوبهای جامعه)، و

واریانس مساوی $npq\left(\frac{N-n}{N-1}\right)$، که در آن q=1-p.

توزیع هندسی:

X تعداد امتحان‌های برنولی تا بدست آوردن اولین موفقیت

$$p = P[S], \qquad q = 1 - p$$

$$P[X = x] = q^{x-1} p, \qquad x = 1, 2, \ldots$$

$$\text{میانگین} = \frac{1}{p}, \quad \text{واریانس} = \frac{q}{p^2}$$

توزیع هندسی را گاهی توزیع زمان انتظار گسسته می گویند.

پیشامدهای کمیاب *انتظار می‌رود هر یک از این پیـــشامدها در تعداد زیادی از امتحان‌های بالقوه، فقط چندبار رخ دهد.*

تعریف رسمی توزیع پواسن

پیشامد S با پیروی از فرض‌های زیر در زمان رخ می دهد:

الف) **استقلال:** تعداد دفعاتی که پیشامد S در هر فاصله زمانی روی می‌دهد، مستقل از تعداد دفعاتی است که آن دو پیشامد در هر فاصله زمانی مجزای دیگر، اتفاق می‌افتد.

ب) **با هم نیامدن:** شانس دو یا چند رخداد همزمان را می‌توان صفر فرض کرد.

ج) **نرخ:** متوسط تعداد رخدادهای پیشامد در واحد زمان مقدار ثابتی است که با m نشان داده می‌شود، و در اثر تغییر زمان، تغییر نمی‌کند.

X متغیری تصادفی است که تعداد دفعات رخ دادن S را در فاصله زمانی واحد نشان می‌دهد.

توزیع پواسن

$$P[X=x]=\frac{e^{-m}m^{x}}{x!},\qquad x=۰,۱,۲,...$$

e عدد نمایی و مقدارش برابر ۲/۷۱۸۲۸

تقریب پواسن برای دوجمله‌ای

	پواسن (m)	دوجمله‌ای (n,p)
میانگین	m	np
واریانس	m	npq

مثال: فرض کنید متوسط تعداد سیکلوپ‌ها در هر لیتر از آب رودخانه‌ای برابر ۲ است. احتمال اینکه ۵ سیکلوپ یا بیشتر در نمونه‌ای ۳ لیتری از آب دریاچه یافت شود، چقدر است؟

متوسط تعداد سیکلوپ‌ها در هر لیتر ۲ است، در حجم ۳ لیتری به طور متوسط تعداد $۳\times۲=۶$ سیکلوپ وجود دارد.

$$P[X\geq ۵]=۱-P[X\leq ۴]=۱-۰/۲۸۵=۰/۷۱۵$$

فصل ۶

مفاهیم اساسی آزمون فرض‌ها

مفاهیم اساسی آزمون فرضها

فرض آماری حکمی درباره جامعه است. قابل قبول بودن آن باید بر مبنای اطلاعات از نمونه‌گیری از جامعه بررسی شود.

دو فرض مکمل:

فرض H: ادعا درست است.

فرض H': ادعا نادرست است.

دو تصمیم یا استنباط:

H' را رد کند و نتیجه بگیرد که H به وسیله داده ها قویا تایید می گردد.

H' را رد نکند و نتیجه بگیرد که داده ها H را قویا تایید نمی کنند.

انتخاب یکی از دو شق را آزمون فرض آماری نامند.

توجه: نتیجه حاصل از آزمون فرض آماری به وسیله تحلیل داده‌های تجربی، نوعی عدم حتمیت به همراه دارد.

گزاره ریاضی

نتایج تحقیق

نتیجه

یک فرض آماری

نتایج تحقیق

نتیجه نوعی

فرض صفر و فرض مقابل

مسئله:

تجربه نشان داده است که نرخ شفابخشی نوعی داروی استاندارد برای بیماری مفروضی ۶۰درصد است. انتظار می رود که نرخ شفابخشی داروی جدیدی از داروی استاندارد بیشتر باشد. فرض کنید که قرار است داروی جدیدی را بر روی نمونه‌ای مرکب از ۲۰ بیمار آزمایش کنند و تعداد بیماران شفا یافته از بین ۲۰ بیمار (X) ثبت شود. از داده‌های به دست آمده از این آزمایش، برای پاسخ داده به سوال زیر چگونه می‌توان استفاده کرد؟ «آیا گواه قانع کننده‌ای برای این مدعا وجود دارد که نرخ شفا بخشی داروی جدید بیشتر از داروی استاندارد است؟»

گزینش H_0 و H_1

هرگاه بخواهیم یک ادعا را از طریق تایید آن به وسیله اطلاعات حاصل از نمونه ثابت کنیم، نفی آن ادعا را فرض صفر H_0 و خود آن را فرض مقابل H_1 می‌گیریم.

در آزمون فرض صفر H_0 در برابر فرض مقابل H_1 ، روش ما این است که H_0 را درست بدانیم مگر اینکه داده های بدست آمده قویاً برخلاف آن حکم کنند، که در آن صورت H_0 باید بنفع H_1 رد شود. رد کردن H_0 درست است، نسبت به رد نکردن H_0 وقتی که H_1 درست است ، خطای مهمتری است.

Close Full Screen

$H_0: p \leq ۰/۶$ (داروی جدید بهتر نیست)

$H_1: p > ۰/۶$ (داروی جدید بهتر است)

آزمون فرض صفر فرایندی است که طی آن مجموعه‌ای از مقادیر متغیر تصادفی X که به ازای آنها H_0 باید رد شود، تعیین می گردد . متغیری تصادفی که مقادیر آن برای انجام این عمل به کار می‌روند آماره آزمون نامیده می شود و مجموعه مقادیری از این متغیر تصادفی را که به ازای آنها H_0 باید رد شود ناحیه آزمون می‌نامند. آزمون به وسیله آماره آزمون و ناحیه رد به طور کامل مشخص می شود.

دو نوع خطا و تابع توان آزمون

H_0 غلط ($p>$۰/۶)	H_0 درست ($p\leq$۰/۶)	نتایج آزمون
غلط(خطای نوع II)	صحیح	H_0 رد نمی‌شود
صحیح	غلط(خطای نوع I)	H_0 رد می‌شود

خطای نوع I: رد کردن H_0 وقتی که H_0 درست است.

خطای نوع II: رد نکردن H_0 وقتی که H_1 درست است.

احتمال های دو نوع خطا

$\alpha=P[$خطای نوع $I]=P[$ وقتی H_0 رد کردن H_1 درست است $]$

$\beta=\alpha=P[$خطای نوع $II]=P[$ رد نکردن H_0 وقتی H_1 درست است $]$

احتمال α بستگی به مقدار مشخص پارامتر در دامنه‌ای دارد که H_0 آن را در بر می گیرد و حال آنکه β بستگی به مقدار پارامتر در دامنه‌ای دارد که H_1 آن را در بر می‌گیرد.

تابع احتمال رد فرض در یک آزمون $p(\gamma)$

$\gamma(p)=P$[در آزمون، H_0 رد شود در حالی که مقدار واقعی پارامتر p است]

منحنی تابع احتمال رد کردن (γ) برای یک آزمون، نشانگر نحوه انجام آزمون از طریق نمایش احتمال خطا به ازای تمـــام مقادیر ممکن پارامتر می‌باشد. این منحنـــی را منحنـــی تـــوان آزمون می‌نامند.

عرض منحنی در ناحیه‌ای که H_0 صحیح است، احتمال خطای نوع I را می‌دهد؛ در حالی که، در ناحیه دیگر، با کـــم کـــردن مقدار عرض منحنی از عدد یک، احتمال خطای نوع II به دست می آید.

گزینش یک آزمون از میان چند آزمون

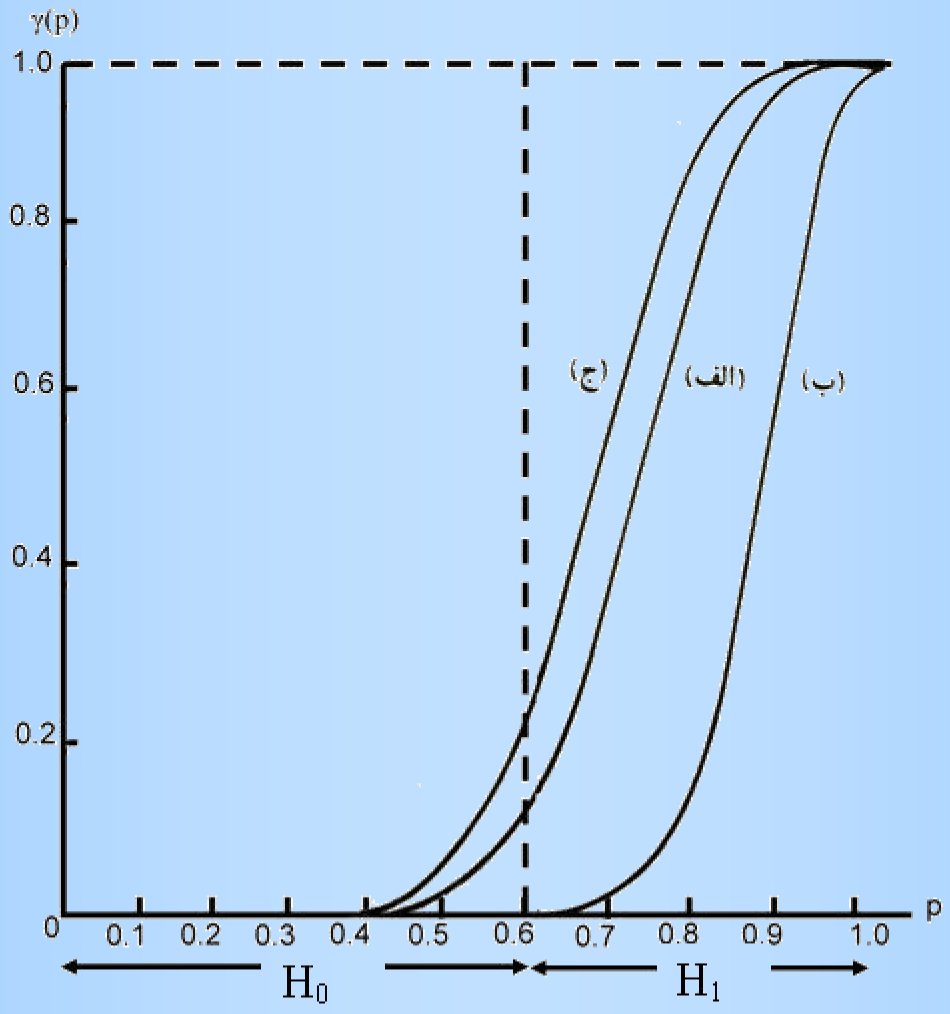
آزمون الف: ناحیه رد $X \geq ۱۵$

آزمون ب: ناحیه رد $X \geq ۱۸$

آزمون ج: ناحیه رد $X \geq ۱۴$

احتمال‌های رد برای آزمون‌های الف،ب و ج

P	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
$\gamma(p)=P[X\geq ۱۵]$	۰/۰۰۰	۰/۰۰۲	۰/۰۲۱	۰/۱۲۶	۰/۶۱۴	۰/۸۰۴	۰/۹۸۹
$\gamma(p)=P[X\geq ۱۸]$	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۴	۰/۰۳۵	۰/۲۰۶	۰/۶۷۷
$\gamma(p)=P[X\geq ۱۴]$	۰/۰۰۰	۰/۰۰۶	۰/۰۵۸	۰/۲۵۰	۰/۶۰۸	۰/۹۱۳	۰/۹۹۸


γ(p)
1.0
0.8
0.6
0.4
0.2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
p
(ج)
(الف)
(ب)
H_0
H_1

استخراج نتایج از یک آزمون

نتیجه آزمون به طریق زیر بیان شود:

در سطح معنی دار بودن α

H_0 رد می‌شود

یا اینکه

H_0 رد نمی‌شود

علاوه بر اجرای آزمون فرض‌ها با مقدار از قبل تعیین شده α، عمل آماری مفیدی که می‌توان انجام داد این است که احتمال معنی‌دار بودن را هم ثبت کنیم.

علاوه بر اجرای آزمون فرض‌ها با مقدار از قبل تعیین شده α، عمل آماری مفیدی که می‌توان انجام داد این است که احتمال معنی‌دار بودن را هم ثبت کنیم.

آزمون، در حالتی که فرض مقابل دوطرفه است

H_0 **رد می شود اگر** $X\leq c_1$ **یا** $X\geq c_2$

c_1 و c_2 **باید به نحوی تعیین کرد که احتمال خطای نوع** I **محدود به سطح معینی باشد.**

مراحل کلی آزمون فرضها

الف) با توجه به ماهیت داده‌های آزمایشی و با در نظر گرفتن ادعاهایی که باید آزمایش گردند، مدل احتمال مناسب را تعیین می کنیم و هر مدعا را بر حسب دامنه‌ای از مقادیر پارامتر(های) مربوطه θی مدل، بیان می‌کنیم.

ب) وقتی در جستجوی شواهدی برای اثبات مدعای خاصی از طریق تایید داده های آزمایشی هستیم، نفی این مدعا را به صورت فرض صفر H_0 و خود مدعا را به صورت فرض مقابل H_1 فرمولبندی می‌کنیم. H_0 و H_1 هر دو باید برحسب پارامترهای مدل بیان گردند. به عنوان مثال:

$H_0: \theta=\theta_0$ در برابر $H_1: \theta>\theta_0$ یا $H_0: \theta=\theta_0$ در برابر $H_1: \theta\neq\theta_0$

ج) یک آماره آزمون T انتخاب می‌کنیم که مقدار آن موجه بودن فرض تحت آزمون را به بهترین صورت نشان دهد. این آمـــاره آزمون باید تابعی از داده‌های قابل مشاهده باشد و نباید شامل هیچ پارامتر نامعلومی باشد. دامنه مقادیر ممکن T را مشخص می‌کنیم و مقادیری را تعیین می‌کنیم که به ازای آنها H_0 به نفع H_1 رد می‌شود. به عبارت دیگر، ساختار ناحیه رد را با بیان این مطلب فرمولبندی می‌کنیم که H_0 باید رد شود اگر، برحـــسب مورد مقدار مشاهده شده T خیلی زیاد یا خیلی کم باشد.

د) سطح اغماض احتمال خطای نوع I برای فرایند آزمون کردن را مشخص می‌کنیم. از جدولهایی که توزیع های احتمال T را نشان می‌دهند برای تعیین ناحیه رد استفاده می‌کنیم، به طوری که بیشینه احتمال خطای نوع I از حد اغماض مشخص شده بیشتر نباشد. معمولا برای این منظور کافی است که مطمئن شویم که انتخاب حجم نمونه و α طوری است که خطای نوع II به طور معقولی کم است.

ه) بعد از اینکه آزمون یا قاعده تصمیم‌گیری به طور واضح فرمولبندی شد و سطح معنی‌دار بودن معین گشت، مقدار مشاهده شده T را از روی داده‌های آزمایشی محاسبه کرده و معلوم می کنیم که آیا این مقدار در ناحیه رد واقع است یا نه. اگر در آن ناحیه واقع باشد، نتیجه می‌گیریم که اعتبار H_1 در سطح مفروض معنی‌دار بودن به وسیله داده‌ها نشان داده نشده است. در اینجا باز ادعا نمی‌کنیم که اعتبار H_0 از روی شواهد آماری اثبات شده است، بلکه باید بگوییم که شواهد تجربی کافی برای رد H_0 وجود ندارد.

فصل هفتم

توزیع نرمال و نمونه های تصادفی :

توزیع نرمال و نمونه های تصادفی

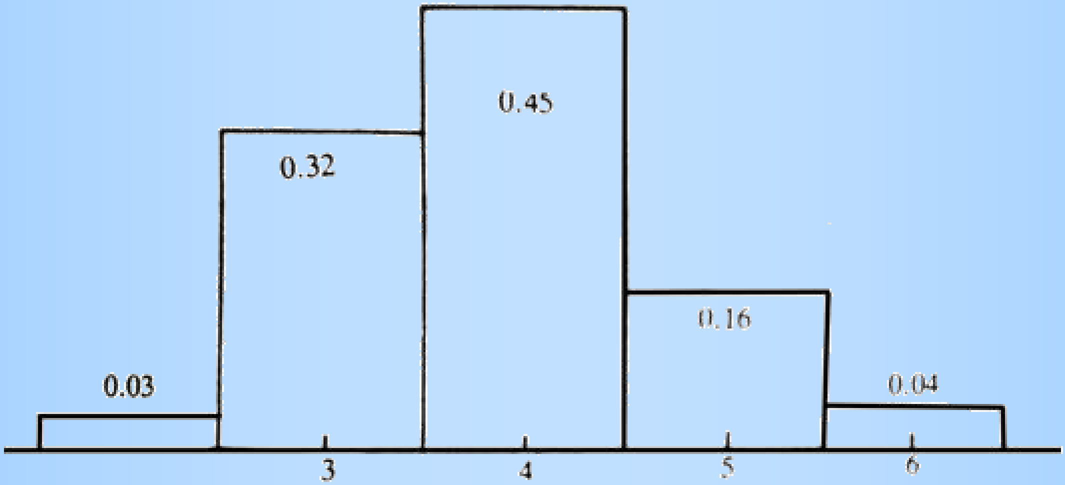
مدلهای احتمال برای متغیرهای پیوسته

توزیع احتمال یک متغیر تصادفی پیوسته را می توان به صورت شکل هموار از بافت‌نگار فراوانی نسبی برای تعداد زیادی مساده تصور نمود.

یادآوری: خواص بافت نگار فراوانی

الف) مساحت کل سطح زیر بافت نگار برابر با یک است.

ب) به ازای هر دو مقطه a و b ، که هر یک نقطه ای کرانه ای از یک رده است ، فراوانی نسبی اندازه ها در فاصله a تا b برابر با مساحت زیر بافت نگار محدود به این فاصله است.

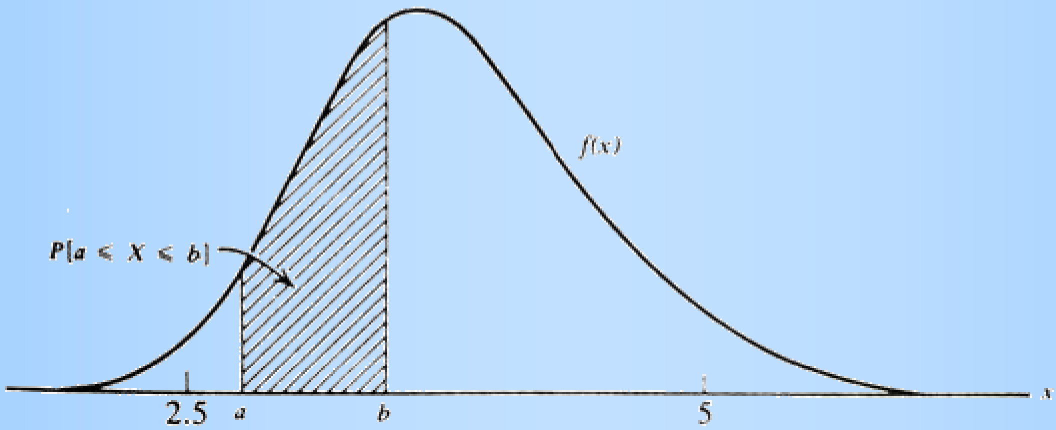


بافت نگار فراوانی نسبی وزن ۱۰۰ نوزاد در

موقع تولد با رده به طول یک کیلو

بافت نگار فراوانی نسبی وزن ۵۰۰۰ نوزاد در

موقع تولد با رده به طول ۱/۸ کیلو



منحنی چگالی احتمال برای متغیر تصادفی پیوسته X(وزن نوزاد در موقع تولد)

تابع چگالی احتمال $f(x)$ ، توزیع احتمال یک متغیر تصادفی پیوسته را توصیف می‌کند و دارای خواص زیر است:

الف) مساحت کل زیر منحنی چگالی برابر با یک است.

ب) مساحت زیر منحنی چگالی بین a و b مساوی با $P[a \leq X \leq b]$.

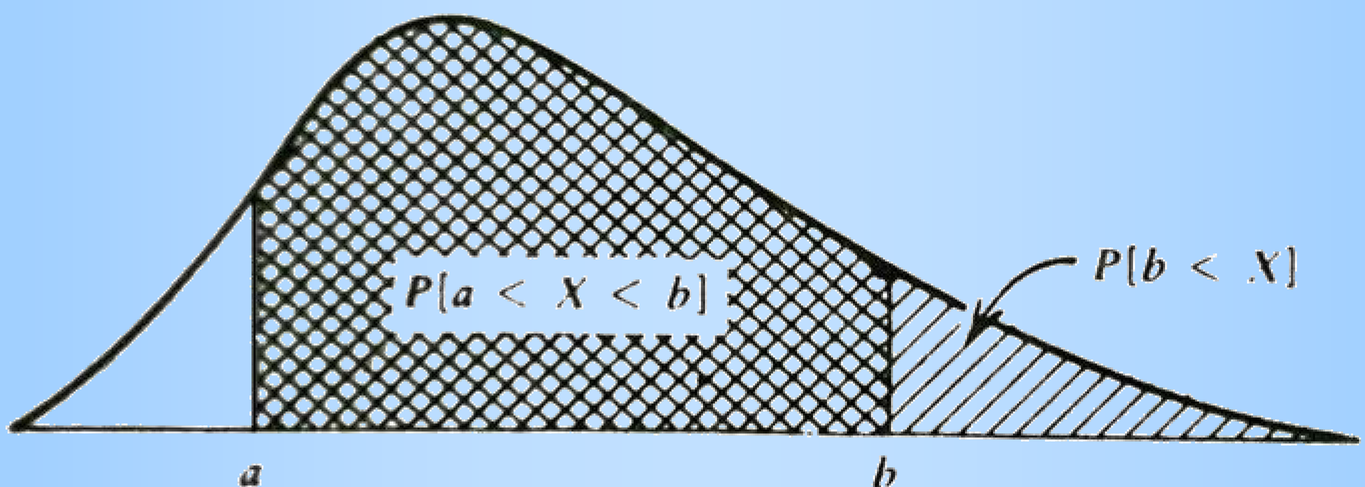
ج) $f(x)$ مثبت یا صفر است.

برای یک متغیر تصادفی پیوسته احتمال آنکه $X = x$، همیشه صفر است و فقط از احتمال قرار گرفتن X در یک فاصله ، صحبت به میان می آید.

در بیشتر جدول‌ها، مساحت سطح واقع در سمت چپ هر نقطه در جدول آمده‌است. برای‌به دست‌آوردن احتمال فواصل دیگر، باید قواعد زیر را بکار برد:

$P[a \le X \le b] = ($مساحت سمت چپ $b) - ($مساحت سمت چپ $a)$

$P[b < X] = 1 - ($مساحت سمت چپ $b)$



مشخص کردن مدل احتمال

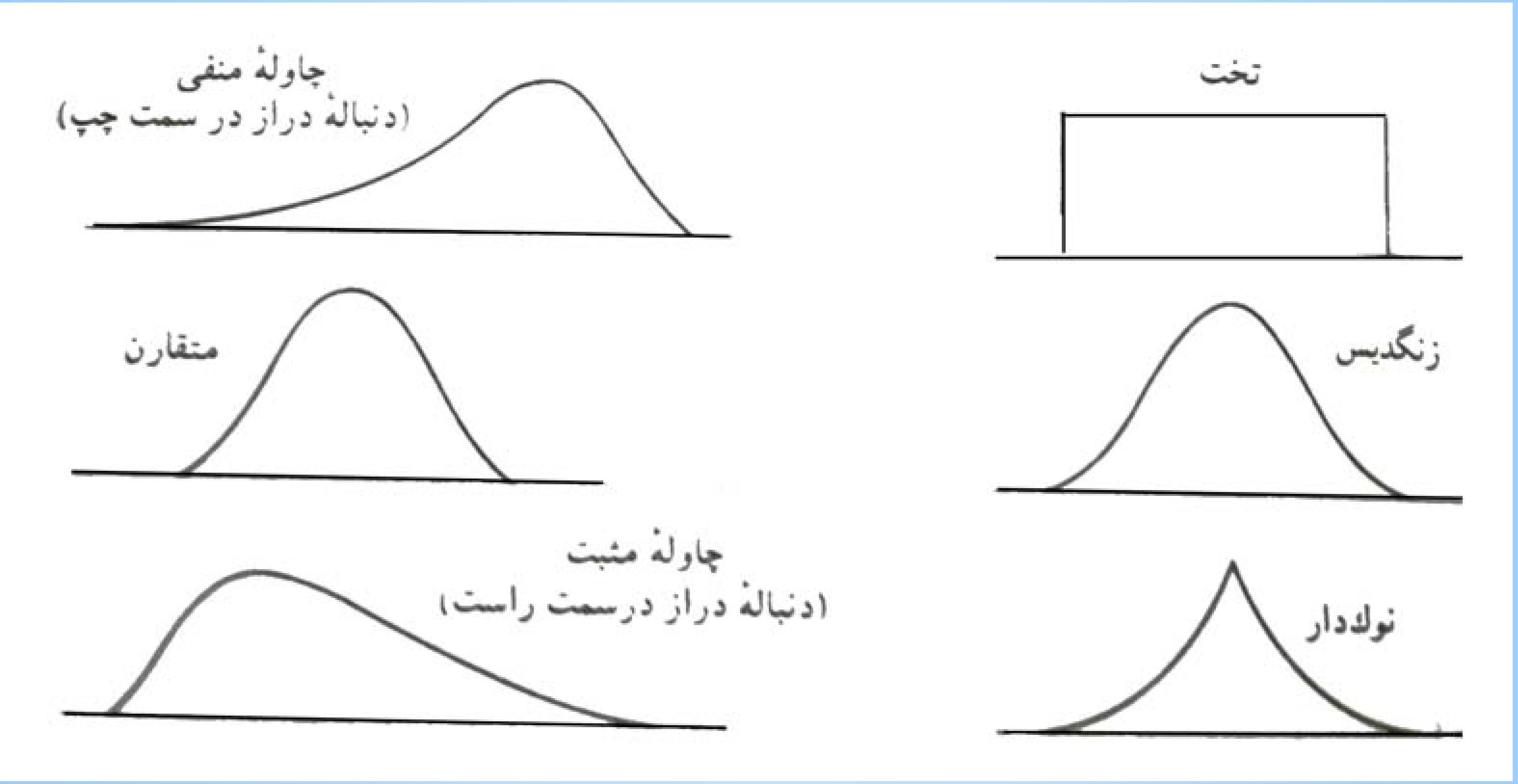
مدل احتمال متغیر تصادفی پیوسته به وسیله شکل ریاضی تابع چگالی احتمال مشخص می شود.

وقتی که تعداد نسبتاً زیادی از مشاهدات مربوط به یک متغیر تصادفی در دسترس باشند، سعی می‌کنیم قسمت بالای نردبانی شکل بافت‌نگار فراوانی نسبی را به وسیله یک منحنی ریاضی تقریب کنیم در تعیین یک چنین منحنی تقریبی، باید معیارهای سادگی شکل و امساک در استفاده از پارامترها را در نظر گرفت تا مدل در روش‌های استنباط آماری قابل استفاده باشد.

وقتی تعداد زیادی از دادها در دسترس نباشد، ممکن است به طور آزمایشی مدل قابل قبولی را بپذیریم که داده‌های خاص از یک منبع مشابه، پیش پای ما می گذارند.

خصوصیات پیوسته

شکلهای گوناگون منحنی های چگالی احتمال



چاوله: یکی از دنباله ها ، درازتر از دیگری است.

یک **متغیر تصادفی پیوسته** X نیز می‌تواند دارای میانگین یا امیـــد ریاضی $E(X)$ و نیز واریانس و انحراف معیار باشد.

میانگین $\mu = E(X)$ نقطه تعادل توزیع احتمالی را نشان می‌دهد.

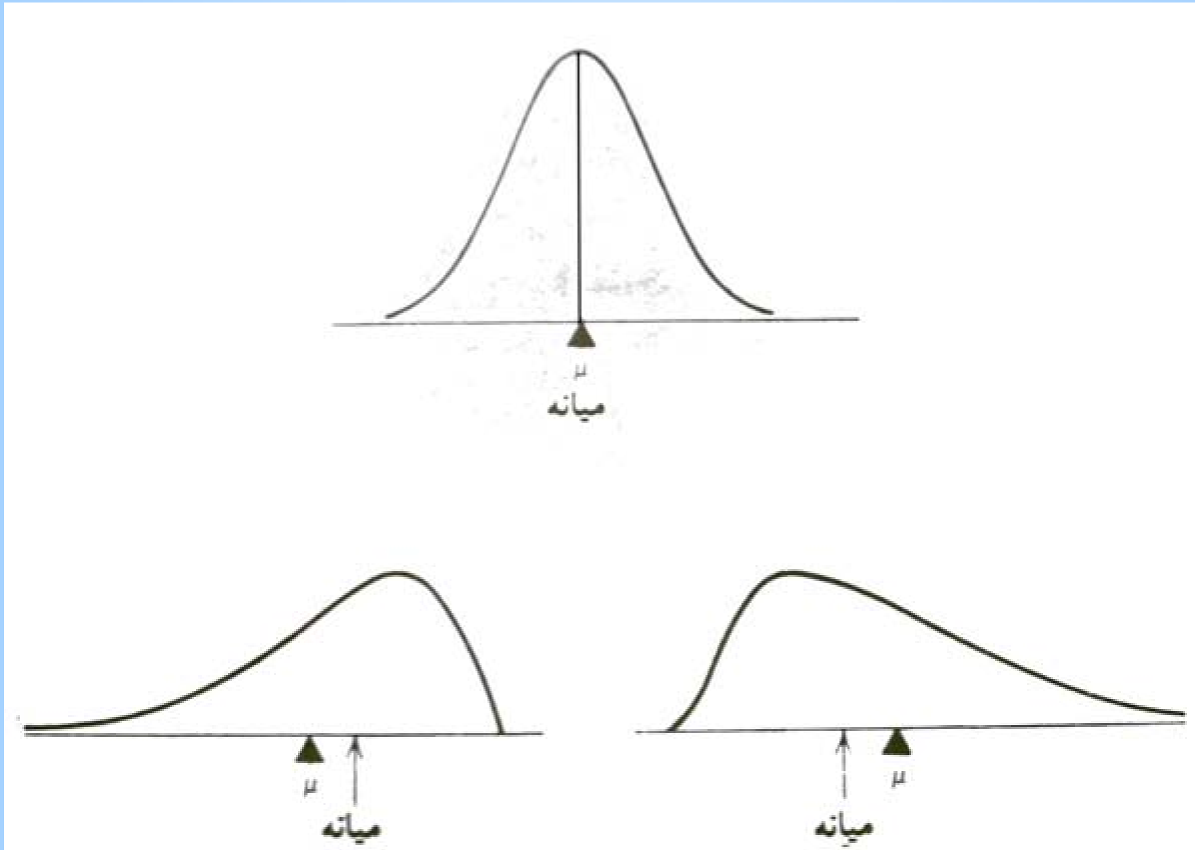
میانه معیار دیگر به مرکـــز مقـــداری از X اســـت کـــه مـــساحت زیرمنحنی را به دو قسمت برابر تقسیم می‌کند.

صدک($100p$)ام جامعه ، یک مقدار X است که مقدار مساحت در سمت چپ آن p و مقدار مساحت در سمت راست آن $1-p$ است.

چارک اول (پایینی) عبارت است از صدک ۲۵ام

چارک دوم (میانه) عبارت است از صدک ۵۰ام

چارک سوم (بالایی) عبارت است از صدک ۷۵ام


μ
میانه
μ
میانه
میانه
μ

چند خاصیت مهم

برای n متغیر تصادفی $X_1, \ldots, X_n$

$$E(a_1 X_1 + \ldots + a_n X_n) = a_1 E(X_1) + \ldots + a_n E(X_n)$$

اگر $X_1, \ldots, X_n$ مستقل باشند،

$$Var(a_1 X_1 + \ldots + a_n X_n) = a_1^2 Var(X_1) + \ldots + a_n^2 E(X_n)$$

توزیع نرمال

توزیع نرمال دارای چگالی زنگدیس زیر است:

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$ **برای**

که در آن ، μ میانگین و σ انحراف معیار است.

احتمال فاصله‌ای که به اندازه یک احراف معیار در هر طرف میانگین امتداد دارد برابر است با

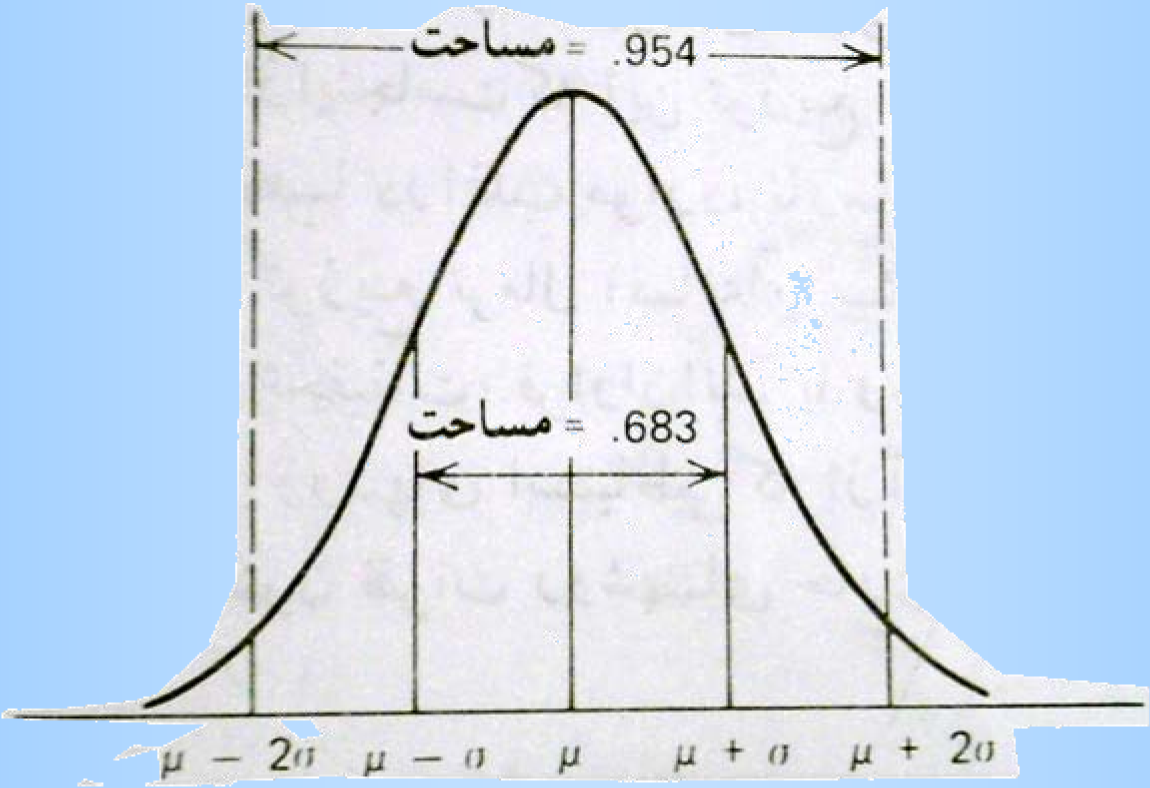
$$P[\mu - \sigma < X < \mu + \sigma] = ۰/۶۸۳$$

دو انحراف معیار در هر طرف میانگین امتداد دارد برابر است با

$$P[\mu - 2\sigma < X < \mu + 2\sigma] = ۰/۹۵۴$$

سه انحراف معیار در هر طرف میانگین امتداد دارد برابر است با

$$P[\mu - 3\sigma < X < \mu + 3\sigma] = ۰/۹۹۷$$

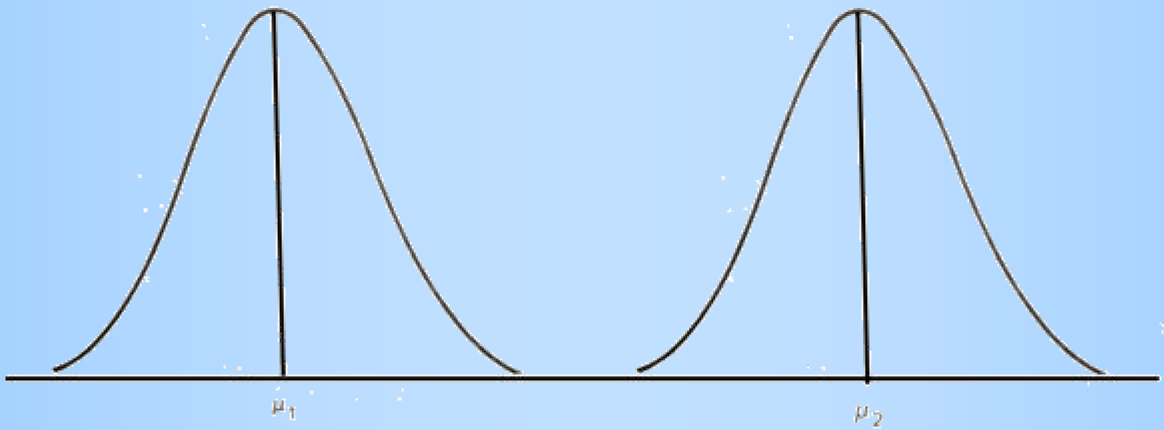


مساحت = .954
مساحت = .683
$\mu - 2\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 2\sigma$

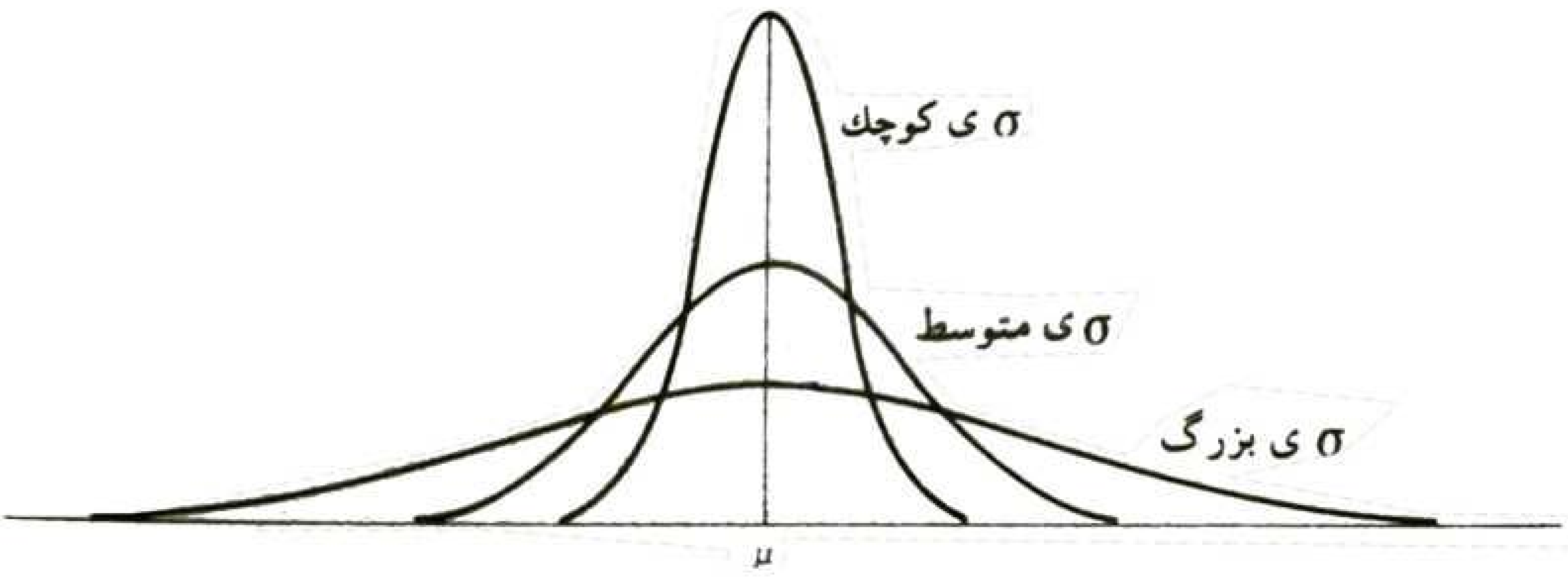
نمادگذاری

توزیع نرمال با میانگین μ و انحراف معیار σ به صورت $N(\mu,\sigma)$ نشان داده می شود.

دو توزیع نرمال با میانگین‌های مختلف و با انحراف معیارهای برابر



کاهش مقدار σ، عرض نقطه بیشینه و تمرکز احتمال در اطراف μ را افزایش می دهد.

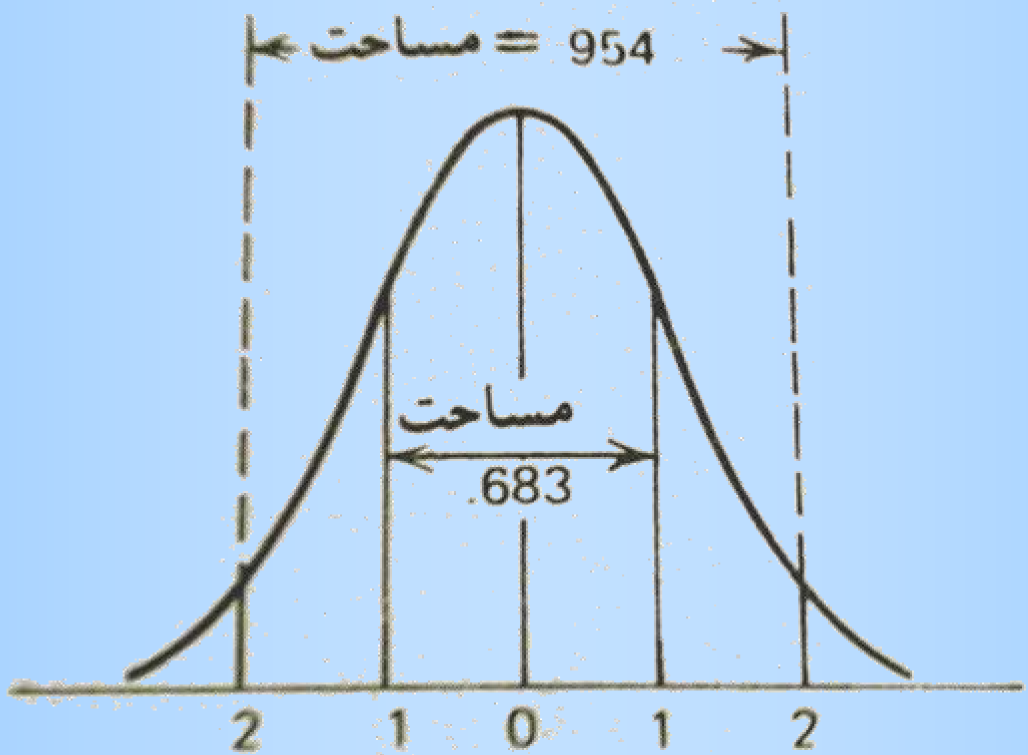


توزیع نرمال استاندارد ، دارای یک منحنی زنگدیس با

(میانگین) $\mu = 0$

(انحراف معیار) $\sigma = 1$

توزیع نرمال استاندارد به صورت $N(0,1)$ **نشان داده می شود.**



مساحت = 954
مساحت
.683
2
1
0
1
2

استفاده از جدول نرمال

جدول مساحت زیر منحنی در سمت چپ هر مقدار مشخص z را ارائه می‌دهد.

$P[Z \le z] =$ **مساحت زیر منحنی در سمت چپ** z

احتمال یک فاصله $[a,b]$ ،

$P[a \le Z \le b] =$ [**مساحت سمت چپ** b] $-$ [**مساحت سمت چپ** a]

خواص ناشی از خاصیت تقارن تابع چگالی حول مرکز صفر

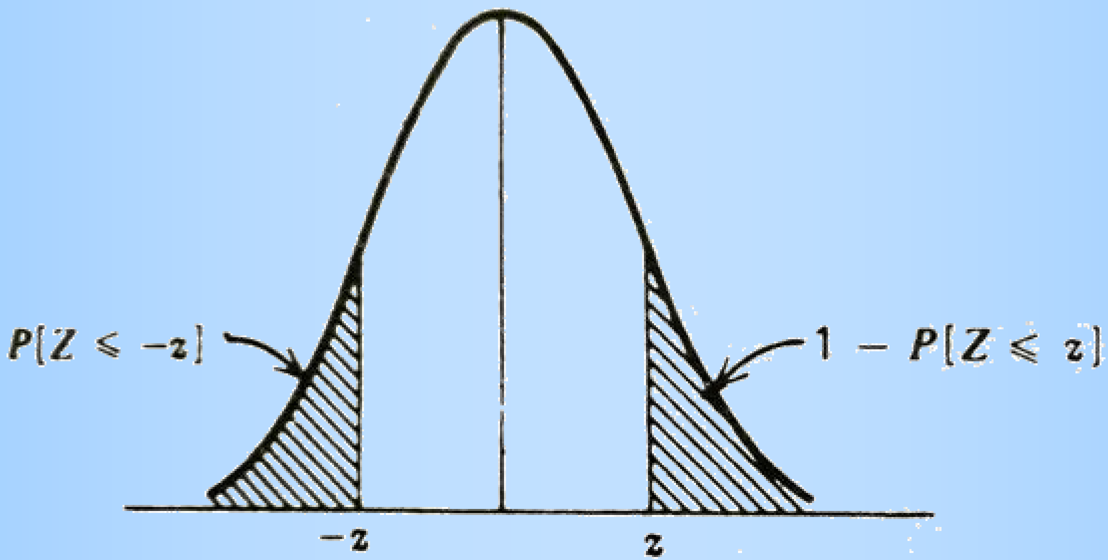
الف) $P[Z \le 0] = 0.5$

ب) $P[Z \le -z] = 1 - P[Z \le z]$

ج) اگر $Z > 0$ داریم

$$P[Z \le z] = 0.5 + P[0 < Z \le z]$$

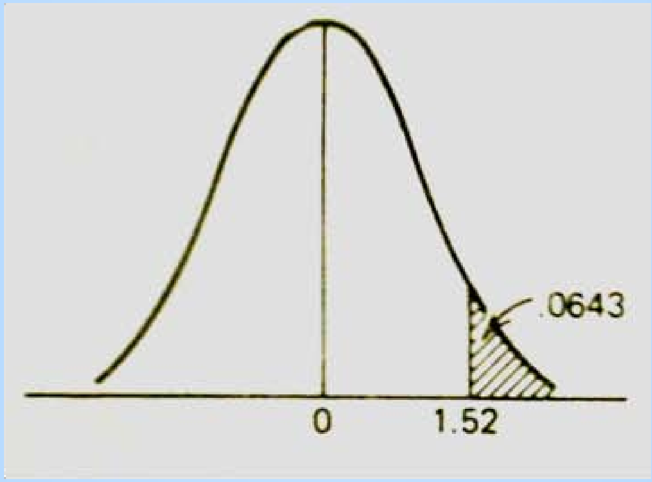
$$P[Z \le -z] = 0.5 - P[0 < Z \le z]$$


P[Z ≤ −z]
1 − P[Z ≤ z]
−z
z

مثال:

$$P[Z > ۱/۵۲] = ?$$

$$P[Z > ۱/۵۲] = ۱ - P[Z \le ۱/۵۲] = ۱ - ۰/۹۳۵۷ = ۰/۰۶۴۳$$



روش دیگر استفاده از خاصیت تقارن

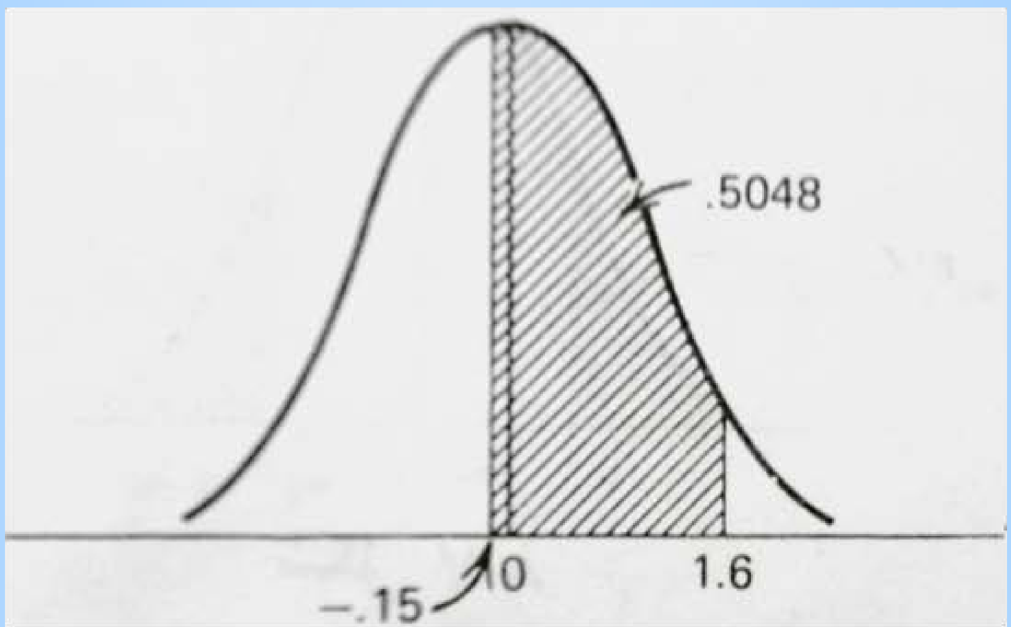
$$P[Z > ۱/۵۲] = P[Z < -۱/۵۲]$$

مثال: $P[-۰/۱۵ < Z < ۱/۶۰] = ?$

$P[Z \leq ۱/۶۰] =$ مساحت واقع در سمت چپ ۱/۶۰ $=$ ۰/۹۴۵۲

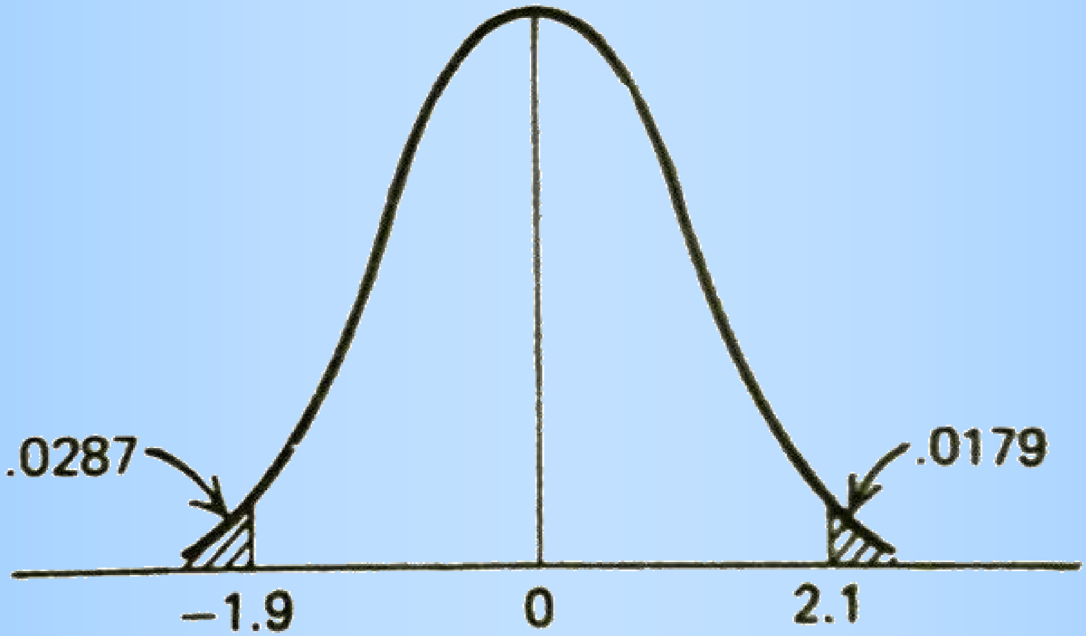
$P[Z \leq -۰/۱۵] =$ مساحت واقع در سمت چپ ۰/۱۵- $=$ ۰/۴۴۰۴

$P[-۰/۱۵ < Z < ۱/۶۰] = ۰/۹۴۵۲ - ۰/۰۴۴۰۴ = ۰/۵۰۴۸$



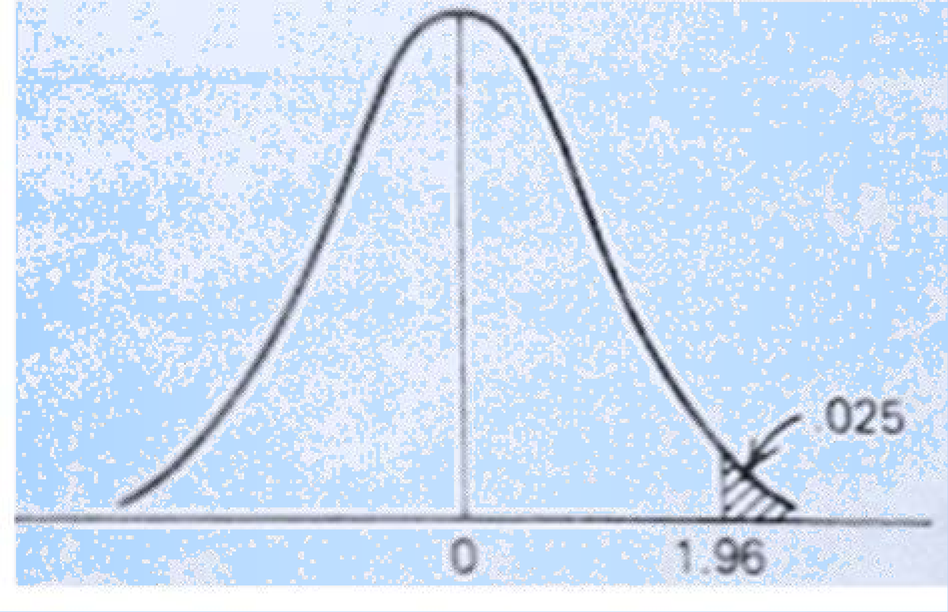
مثال: $P[Z < -۱/۹$ یا $Z > ۲/۱] = ?$

$P[Z < -۱/۹$ یا $Z > ۲/۱] = P[Z < -۱/۹] + P[Z > ۲/۱] = ۰/۰۲۸۷ + ۰/۰۱۷۹ = ۰/۰۴۶۶$



مثال: مقدار z را طوری پیدا کنید که در $P[Z > z] = ۰/۰۲۵$ صدق کند.

می‌دانیم مساحت کل برابر با یک است. مساحت سمت چپ z برابر ۰/۹۷۵۰=۱-۰/۰۲۵۰ باشد. مقدار کناری برای درایه ۰/۹۷۵۰ از جدول برابر با z=۱/۹۶ است.



مثال: مقدار $z>0$ را بدست آورید هرگاه $P[-z \leq Z \leq z]=0/90$.

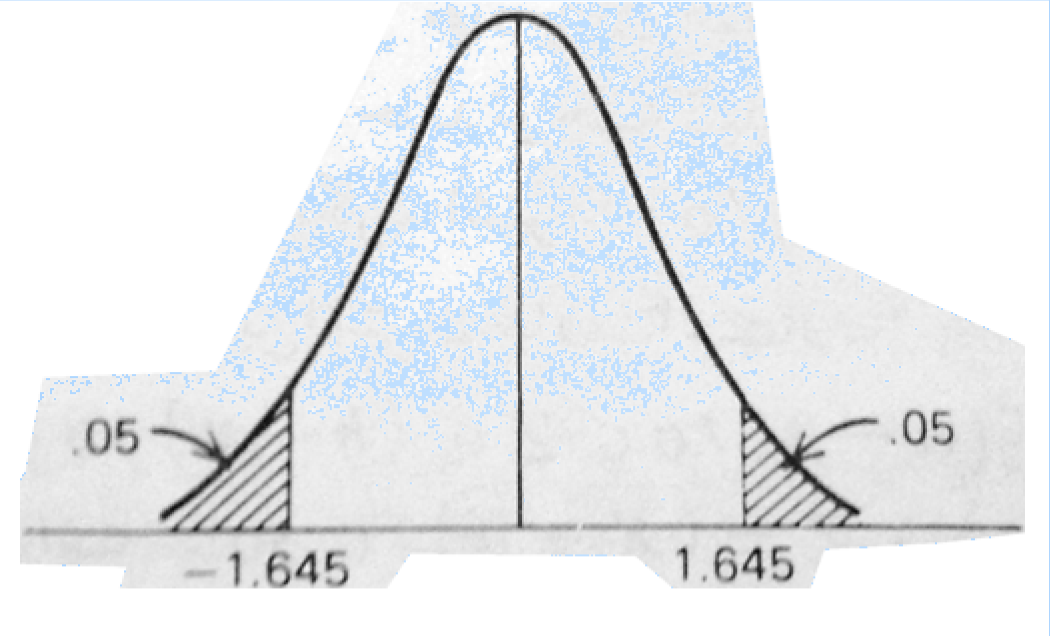
با تئجه به تقارن منحنی داریم

$$P[Z < -z] = P[Z > z] = 0/05$$

$z=1/65$ منجر به $P[Z \leq -1/65] = 0/0495$

$z=1/64$ منجر به $P[Z \leq -1/64] = 0/0505$

با درون یابی بین این دو مقدار $Z=1/645$ را بدست می آوریم



.05
.05
−1.645
1.645

اگر X دارای توزیع $N(\mu,\sigma)$ باشد آنگاه $Z=\frac{(X-\mu)}{\sigma}$ دارای توزیع $N(0,1)$ خواهد بود.

بنابراین

$$P[X\le b]=P\left[\frac{X-\mu}{\sigma}\le\frac{b-\mu}{\sigma}\right]=P\left[Z\le\frac{b-\mu}{\sigma}\right]$$

$$P[a\le X\le b]=P\left[\frac{a-\mu}{\sigma}\le\frac{X-\mu}{\sigma}\le\frac{b-\mu}{\sigma}\right]$$

$$=P\left[\frac{a-\mu}{\sigma}\le Z\le\frac{b-\mu}{\sigma}\right]$$

که در آن احتمالهای مربوط به Z از جدول نرمال استاندارد می آیند.

مثال: در صورتی که X دارای توزیـع N(۱,۴) باشـد، $P[X>-۳]$ و $P[-۳<X<۱۰]$ را محاسبه کنید.

در اینجا $\mu=۱$ و $\sigma=۴$، بنابراین عدد ۱ را از طرفین کم می کنیم و سپس بر ۴ تقسیم می‌نماییم:

$$P[X>-۳]=P[X-۱>-۳-۱]=P\left[\frac{X-۱}{۴}>\frac{-۳-۱}{۴}\right]=P[Z>-۱]=۰/۸۴۱۳$$

$$P[-۳<X<۱۰]=P\left[\frac{-۳-۱}{۴}<\frac{X-۱}{۴}<\frac{۱۰-۱}{۴}\right]=P[-۱<Z<۲/۲۵]=۰/۸۲۹۱$$

خواص دیگر توزیع نرمال

اگر X دارای توزیع $N(\mu,\sigma)$ باشد آنگاه $Y=a+bX$ دارای توزیع $N(a+b\mu,|b|\sigma)$ خواهد بود. یعنی با ضرب کردن X در مقدار ثابت b و اضافه کردن مقدار ثابت a به نتیجه حاصل ، تنها میانگین و واریانس توزیع نرمال تغییر می کنند.

توزیع مجموع دو متغیر نرمال مستقل ، نرمال است. اگر دارای توزیع $N(\mu_1,\sigma_1)$ و Y دارای توزیع $N(\mu_2,\sigma_2)$ باشد و X و Y مستقل باشند. آنگاه $X+Y$ دارای توزیع $N(\mu,\sigma)$ خواهد بود که در آن

$$\mu=\mu_1+\mu_2$$
$$\sigma^2=\sigma_1^2+\sigma_2^2$$
$$\sigma=\sqrt{\sigma_1^2+\sigma_2^2}$$

تقریب نرمال توزیع دوجمله ای

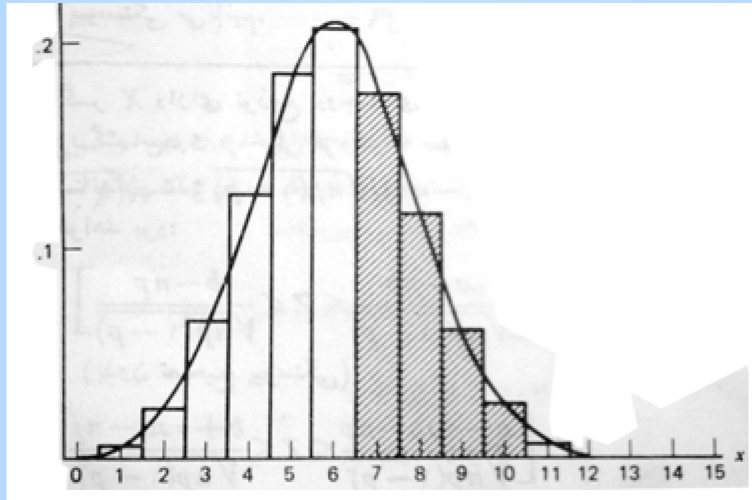
X دارای توزیع دوجمله ای b(n,p) دارای میانگین $\mu = np$ و انحراف معیار $\sigma = \sqrt{np(1-p)}$است.

$$X \sim N(np, \sqrt{np(1-p)})$$

$$P[a \le X \le b] = \sum_{x=a}^{b} \binom{n}{x} p^x (1-p)^{n-x}$$

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

$$P[a \le X \le b] \approx P\left[\frac{a - np}{\sqrt{np(1-p)}} \le Z \le \frac{b - np}{\sqrt{np(1-p)}}\right]$$



بافت نگار احتمال توزیع دوجمله ای b(۱۵ , ۰/۴) و منحنی نرمال تقریبی آن

$$P[۷ \leq X \leq ۱۰] = ۰/۹۹۱ - ۰/۶۱۰ = ۰/۳۸۱$$

$$P[\ ۷ \leq X \leq ۱۰\ ] \approx P\left[\frac{۷-۶}{۱/۹} \leq Z \leq \frac{۱۰-۶}{۱/۹}\right]$$

$$= P[\,۰/۵۲۶ \leq X \leq ۲/۱۰۵\,]$$

$$= ۰/۹۸۲ - ۰/۷۰۰$$

$$= ۰/۲۸۲$$

اگر X دارای توزیع دوجملـــه ای $b(n,p)$ باشـــد، کـــه در آن n بزرگ است و p خیلی نزدیک به صفر یا یک نزدیک نیست، توزیع متغیر استاندارد شده $Z=(X-np)/\sqrt{np(1-p)}$ تقریباً $N(0,1)$ خواهد بود:

بدون تصحیح پیوستگی:

$$P[a \le X \le b] \approx P\left[\frac{a-np}{\sqrt{np(1-p)}} \le Z \le \frac{b-np}{\sqrt{np(1-p)}}\right]$$

با تصحیح پیوستگی:

$$P[a \le X \le b] \approx P\left[\frac{a-0.5-np}{\sqrt{np(1-p)}} \le Z \le \frac{b+0.5-np}{\sqrt{np(1-p)}}\right]$$

مثال: بررسی وسیعی که چند سال قبل انجام شد نشان داد که ۳۰ درصد افراد بالغ جامعه معتاد به سیگار هستند. اگر نرخ جاری معتادان به سیگار نیز همین قدر باشد، احتمال اینکه در یک نمونه تصادفی مرکب از ۱۰۰۰ فرد بالغ، تعداد معتادان به سیگار (الف) کمتر از ۲۸۰ باشد، (ب) ۳۱۶ یا بیشتر باشد، چقدر است؟

الف) $n=$ ۱۰۰۰ و $p=$ ۰/۳ ، $np=$ ۳۰۰ و ۱۴/۵ $= \sqrt{np(۱-p)} = \sqrt{۲۱۰}$

X تعداد معتادان به سیگار دارای توزیع $N($۳۰۰ ،۱۴/۵$)$ است.

$$P[X \le ۲۷۹] \approx P\left[Z \le \frac{۲۷۹-۳۰۰}{۱۴/۵}\right]$$

$$= P[Z \le -۱/۴۴۸]$$

$$= ۰/۰۷۴$$

با استفاده از تصحیح پیوستگی

$$P[X \le ۲۷۹] \approx P\left[Z \le \frac{۲۷۹/۵-۳۰۰}{۱۴/۵}\right]$$

$$= P[Z \le -۱/۴۱۴]$$

$$= ۰/۰۷۹$$

ب)

$$P[X \geq ۳۱۶] \approx P\left[Z \geq \frac{۳۱۶-۳۰۰}{۱۴/۵}\right]$$

$$= P[Z \geq ۱/۱۰۳]$$

$$= ۱ - ۰/۸۶۵$$

$$= ۰/۱۳۵$$

نمونه‌گیری تصادفی، آماره و توزیع نمونه‌ای

مفاهیم بنیادی آمار عبارتند از :

الف) نمونه گیری تصادفی از یک توزیع احتمال

ب) توزیع های نمونه ای

مثال: محققی قصد دارد در شهری که ۵۰۰۰ خانواده در آن ساکن هستند، تعداد دوچرخه‌هایی را که هر خانواده دارد، بررسی نماید.

توزیع جامعه X، تعداد دوچرخه ها در هر خانواده

x	۰	۱	۲	۳
$f(x)$	۰/۲	۰/۴	۰/۳	۰/۱

قبل از آنکه یک مشاهده به عمل آید ، مدل مشاهده به صورت یک متغیر تصادفی X با توزیع $f(x)$ در نظر گرفته می شود؛ مقدار عددی مشاهده که در عمل به دست می آید ، یک مقدار تحقق یافته متغیر تصادفی است.

نمونه تصادفی:

نمونه ای تصادفی به حجم n از یک جامعه $f(x)$، گردآورده ای از n متغیر تصادفی مستقل $X_1,\ldots,X_n$ است که هر کدام دارای توزیع $f(x)$ است.

آماره: تابعی از مشاهدات است.

هر آماره، خود متغیری تصادفی است . توزیع احتمال آن توزیع نمونه‌ای آماره نامیده می شود.

مثال: توزیع جامعه‌ای که در جدول زیر آمده است. قرار است نمونه تصادفی $(X_۱ , X_۲)$ به حجم ۲ از این جامعه اختیار شود. توزیع نمونه‌ای آماره $\bar{X} = (X_۱ + X_۲)/۲$ چیست؟

x	۰	۱	۲	۳
$f(x)$	۰/۲	۰/۴	۰/۳	۰/۱

$$P[X_۱=۰،X_۲=۱]=P[X_۱=۰]P[X_۲=۱]=۰/۲\times۰/۴=۰/۰۸$$

x_2 \ x_1	۰	۱	۲	۳	مجموع سطر
۰	۰/۰۴	۰/۰۸	۰/۰۶	۰/۰۲	۰/۲
۱	۰/۰۸	۰/۱۶	۰/۱۲	۰/۰۴	۰/۴
۲	۰/۰۶	۰/۱۲	۰/۰۹	۰/۰۳	۰/۳
۳	۰/۰۲	۰/۰۴	۰/۰۳	۰/۰۱	۰/۱
مجموع ستون	۰/۲	۰/۴	۰/۰۳	۰/۱	۱

$$P[\bar{X} = ۱/۵] = ۰/۰۲ + ۰/۱۲ + ۰/۱۲ + ۰/۰۲ = ۰/۲۸$$

مقدار $\bar{X}$	۰	۰/۵	۱	۱/۵	۲	۲/۵	۳	مجموع
احتمال	۰/۰۴	۰/۱۶	۰/۲۸	۰/۲۸	۰/۱۷	۰/۰۶	۰/۰۱	۱

توزیع میانگین نمونه و قضیه حد مرکزی

قرارداد:

میانگین جامعه $=\mu$

واریانس جامعه $=\sigma^2$

$$E(X_1)=\ldots=E(X_n)=\mu$$

$$Var(X_1)=\ldots=Var(X_n)=\sigma^2$$

$$\overline{X}=\frac{1}{n}(X_1+\ldots+X_n)$$

$$
\begin{aligned}
Var(\overline{X}) &= \frac{1}{n^2} Var(X_1 + \ldots + X_n) \\
&= \frac{1}{n^2}[Var(X_1) + \ldots + Var(X_n)] \\
&= \frac{1}{n^2}[\sigma^2 + \ldots + \sigma^2] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}
$$

$$
\begin{aligned}
E(\overline{X}) &= \frac{1}{n} E(X_1 + \ldots + X_n) \\
&= \frac{1}{n}[E(X_1) + \ldots + E(X_n)] \\
&= \frac{1}{n}[\mu + \ldots + \mu] = \frac{n\mu}{n} = \mu
\end{aligned}
$$

میانگین و انحراف معیار $\bar{X}$

برای توزیع میانگین نمونه مبتنی بر نمونه ای تصادفی به حجم n، داریم:

$E(\bar{X}) = \mu$ **(میانگین جامعه)**

$Var(\bar{X}) = \dfrac{\sigma^2}{n} \left(= \dfrac{\text{واریانس جامعه}}{n}\right)$

$sd(\bar{X}) = \dfrac{\sigma}{\sqrt{n}} \left(= \dfrac{\text{انحراف معیار جامعه}}{\sqrt{n}}\right)$

مثال: توزیع جامعه ای در جدول زیر آمده است میانگین و واریانس را در حالت n=۲ محاسبه و درستی روابط $E(\bar{X})=\mu$ و $Var(\bar{X})=\frac{\sigma^2}{n}$ را بررسی کنید.

توزیع جامعه

x	۰	۱	۲	۳	مجموع
$f(x)$	۰/۲	۰/۴	۰/۳	۰/۱	۱
$xf(x)$	۰	۰/۴	۰/۶	۰/۳	۱/۳
$x^2 f(x)$	۰	۰/۴	۱/۲	۰/۹	۲/۵

$\mu =$ ۱/۳

$\sigma^2 =$ ۲/۵-(۱/۳)^۲

=۲/۵-۱/۶۹

=۰/۸۱

توزیع $\bar{X} = (X_۱ + X_۲)/۲$

x	۰	۰/۵	۱	۱/۵	۲	۲/۵	۳	مجموع
$f(x)$	۰/۰۴	۰/۱۶	۰/۲۸	۰/۲۸	۰/۱۷	۰/۰۶	۰/۰۱	۱
$xf(x)$	۰	۰/۰۸	۰/۲۸	۰/۴۲	۰/۳۴	۰/۱۵	۰/۰۳	۱/۳
$x^2 f(x)$	۰	۰/۰۴	۰/۳۸	۰/۶۳	۰/۶۸	۰/۳۷۵	۰/۰۹	۲/۰۹۵

$$E(\bar{X}) = ۱/۳ = \mu$$

$$Var(\bar{X}) = ۲/۰۹۵ - (۱/۳)^۲$$

$$= ۲/۰۹۵ - ۱/۶۹$$

$$= ۰/۴۰۵ = \sigma^۲/۲$$

$\bar{X}$ حاصل از یک جامعه نرمال دارای توزیع نرمال است.

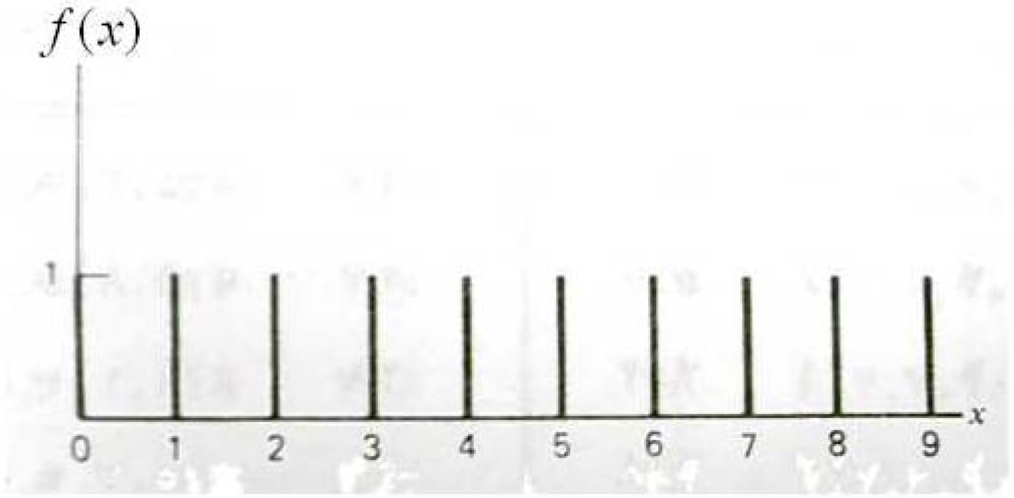
برای نمونه‌ای تصادفی به حجم n از جامعه‌ای نرمال با میانگین μ و انحراف از معیار σ، میانگین نمونه $\bar{X}$ دارای توزیع نرمال با میانگین μ و انحراف معیار $\sigma / \sqrt{n}$ است.

قضیه حد مرکزی

در نمونه گیری تصادفی از جامعه دلخواهی با میانگین μ و انحراف معیار σ، وقتی n بزرگ است توزیع $\bar{X}$ توزیعی است تقریبْ نرمال با میانگین μ و انحراف معیار $\sigma / \sqrt{n}$. به عبارت دیگر $Z = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ تقریباً (۰،۱) N است.

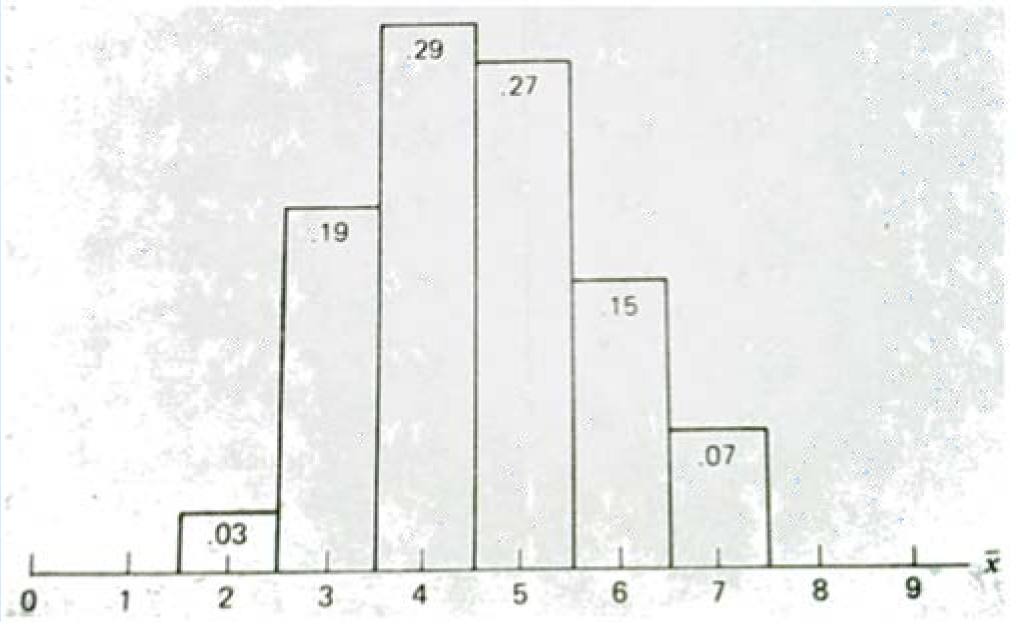
قضیه حد مرکزی به این مورد اشاره دارد که خواه توزیع جامعه پیوسته، گسسته، متقارن، یا چاوله باشد مادام که واریانس جامعه متناهی است، توزیع میانگین نمونه μ، اگر حجم نمونه بزرگ باشد تقریبا نرمال است.

مثال: نمایش قضیه حد مرکزی: جامعه‌ای را در نظر بگیرید که دارای توزیع یکنواخت گسسته است که احتمال ۰/۱ را به هر عدد صحیح ۰، ۱، ...، ۹ نسبت می‌دهد. این مدل ممکن است برای توزیع رقم آخر شماره‌های تلفن یا اولین رقمی که در محاسبات سرریز می‌شود مدل مناسبی باشد.

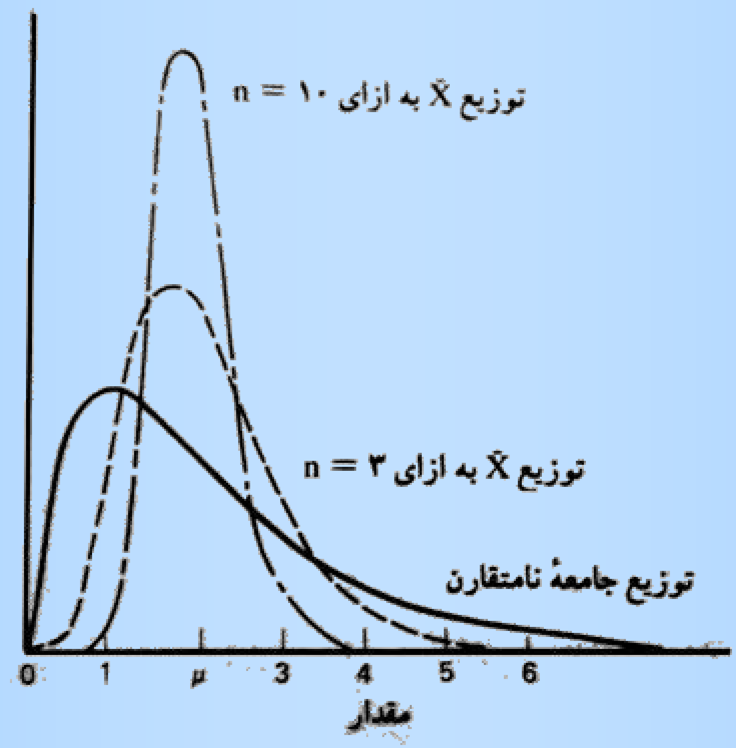


نمودار خطی توزیع یکنواخت بر روی اعداد صحیح ۰، ۱، ...، ۹

با استفاده از کامپیوتر ۱۰۰ نمونه تصادفی به حجم ۵ از این توزیع تولید می‌کنیم و مقدار $\bar{X}$ را برای هر نمونه محاسبه می‌نماییم.



بافت نگار فراوانی نسبی مقادیر $\bar{X}$



توزیع $\bar{X}$ برای n=۳ و n=۱۰ در نمونه‌گیری از یک جامعه چاوله

مثال: فرض کنید معلوم شده که توزیع جامعه نیروی بازوی کارگران صنعتی دارای میانگین ۱۱۰ و انحراف از معیار ۵ می‌باشد برای نمونه ای تصادفی به حجم ۷۵، احتمال اینکه نمونه بین ۱۰۹ و ۱۱۰ باشد چیست؟

$\bar{X}$ میانگین=۱۱۰

$\bar{X}$ انحراف معیار $= \frac{\sigma}{\sqrt{n}} = \frac{۵}{\sqrt{۷۵}} = ۰/۵۷۷$

$$P[۱۰۹ < \bar{X} < ۱۱۱] = P\left[\frac{۱۰۹-۱۱۰}{۰/۵۷۷} < Z < \frac{۱۱۱-۱۱۰}{۰/۵۷۷}\right] = P[-۱/۷۳ < Z < ۱/۷۳] = ۰/۹۱۶$$

توجه: دلیل مناسب بودن توزیع نرمال به عنوان تقریبی برای توزیع دوجمله‌ای، در قضیه حد مرکزی نهفته است.

برای درک این موضوع به هر یک از امتحان‌های برنولی، یک متغیر شمارشگر به صورت زیر نسبت می‌دهیم:

اگر نتیجه امتحان i ام موفقیت باشد، $X_i = ۱$

اگر نتیجه امتحان i ام شکست باشد، $X_i = ۰$

توزیع هر یک از متغیرهای مستقل X1,….,Xn بصورت زیر است

x	۰	۱
$f(x)$	۱-p	p

$$\text{تعداد موفقیتها در n امتحان} = \sum_{i=1}^{n} X_i = T \text{(جمع مقادیر نمونه)}$$

$$\frac{\text{تعداد موفقیتها در n امتحان}}{n} = \frac{T}{n} = \overline{X} \text{(میانگین نمونه)}$$

وقتی که n بزرگ است، نسبت تعداد موفقیت‌ها، $\hat{p}$ در n امتحان برنولی تقریباً به صورت $N(p, \sqrt{p(1-p)/n})$ توزیع می‌شود و $Z = (\hat{p} - p)/\sqrt{p(1-p)/n}$ تقریباً $N(0,1)$ است.

مثال: فرض کنید که ۶۰ درصد افراد یک جامعه شهری ، موافق با سرمایه گذاری برای ایجاد تسهیلات تفریحی باشند. اگر ۱۵۰ نفر به تصادف برای مصاحبه برگزیده شوند ، احتمال اینکه نسبت افرادی از نمونه که موافق با این طرح باشند کمتر از ۰/۵۲ باشد چقدر است؟

در اینجا $n=۱۵۰$ و $p=۰/۶$ است و عبارت $P[\hat{p} < ۰/۵۲]$ باید محاسبه شود

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{۰/۲۴}{۱۵۰}} = ۰/۰۴۰$$

$$P[\hat{p} < ۰/۵۲] \approx P\left[Z < \frac{۰/۵۲-۰/۶۰}{۰/۰۴۰}\right] = P[Z < -۲] = ۰/۰۲۲۸$$

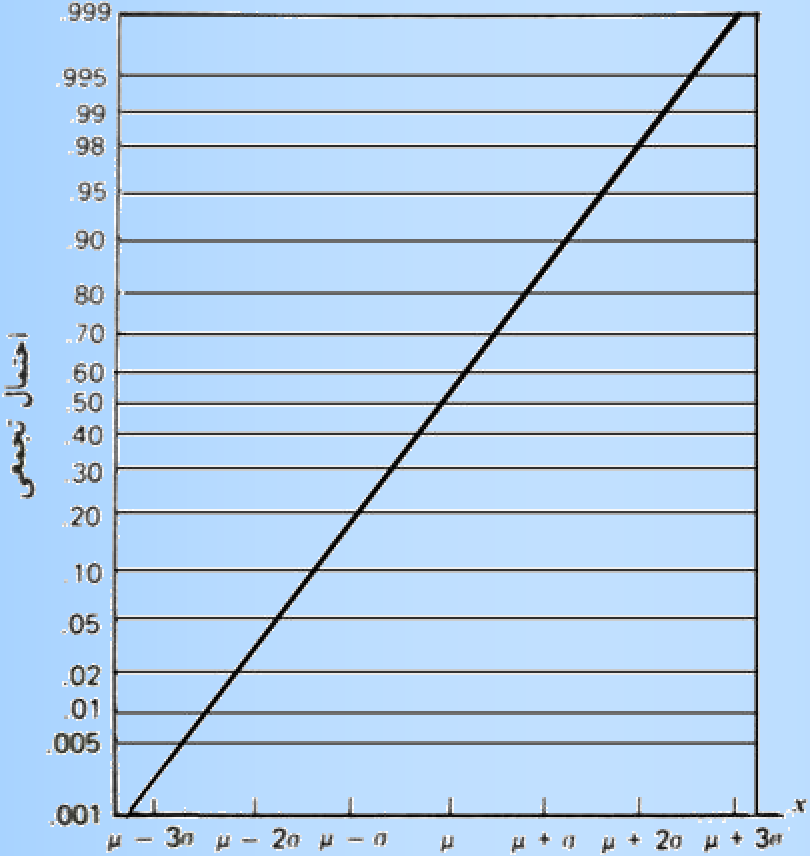
بررسی فرض نرمال بودن یک جامعه

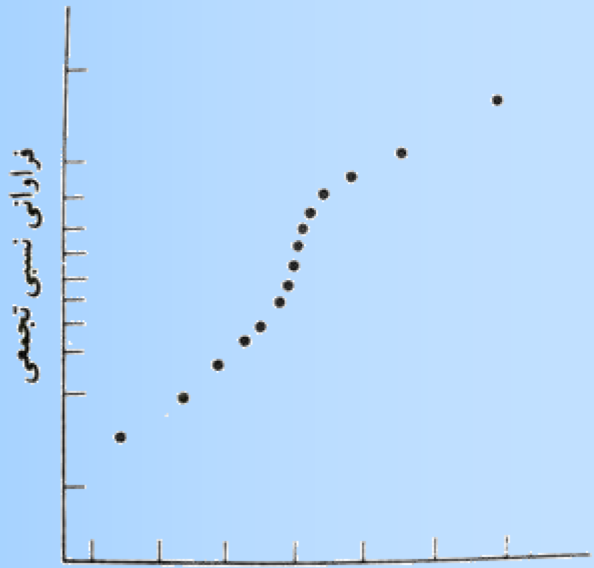
۱- بررسی نسبتهای مشاهدات در فواصل حول میانگین

توزیع $N(\mu,\sigma)$	$(\mu-\sigma,\mu+\sigma)$	$(\mu-2\sigma,\mu+2\sigma)$	$(\mu-3\sigma,\mu+3\sigma)$
مقدار احتمال در فواصل	۰/۶۸۳	۰/۹۵۴	۰/۹۹۷
مقدار احتمال درخارج فواصل	۱/۳	۱/۲۰	۱/۳۰۰

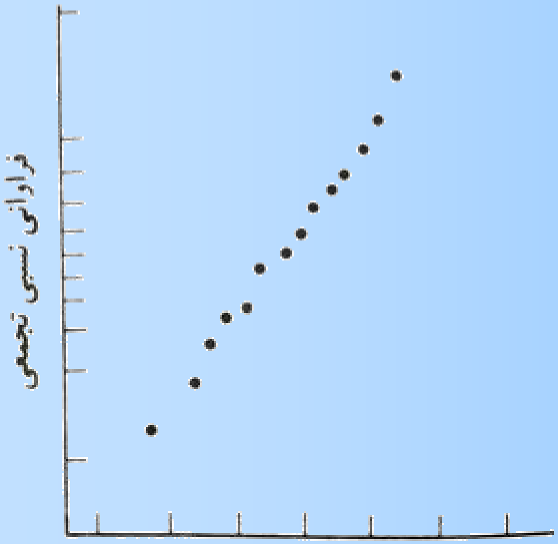
اگر حجم نمونه نسبتاً بزرگ باشد می‌توان انتظار داشت که میانگین نمونه $\bar{X}$ به μ و انحراف معیار نمونه s به σ نزدیک باشند.

۲ – کاغذ احتمال نرمال





نرمال نیست؛ دنباله‌ها ضخیم به نظر می‌رسند.



به‌طور معقولی نرمال است

تبدیل مشاهدات برای دستیابی به حالت نزدیک به نرمال

تغییر مقیاس

برخی از تبدیل‌های مفید

تبدیل‌هایی که مقادیر بزرگ را بزرگتر می‌سازند:

$$x^2, x^3$$

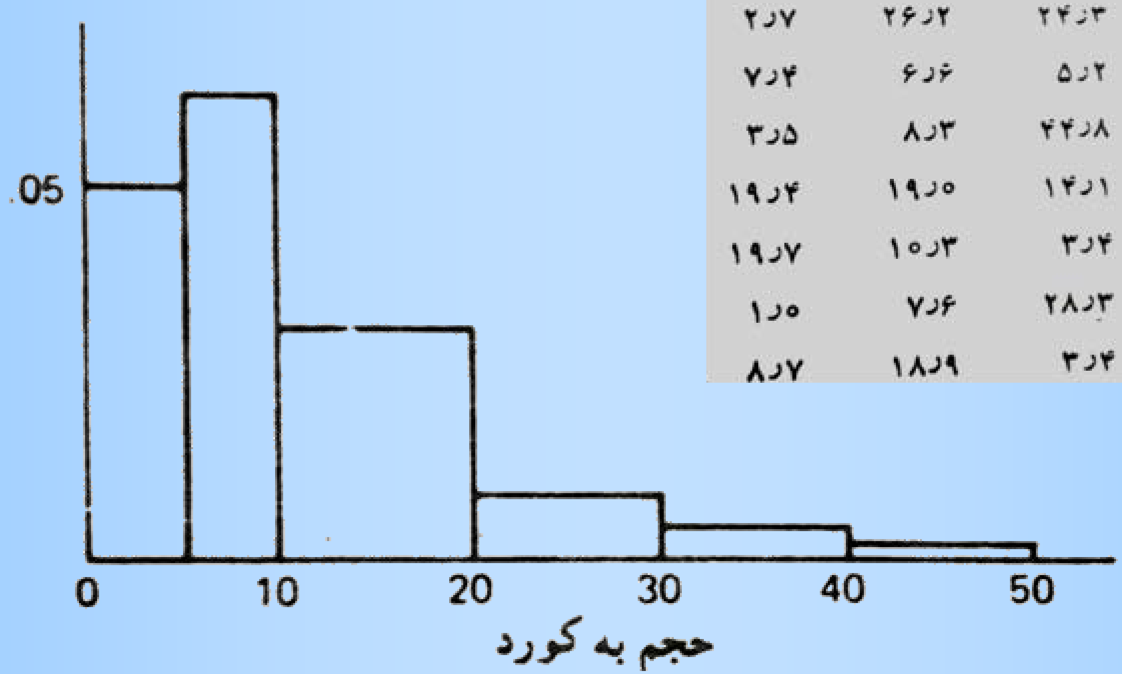
تبدیل‌هایی که مقادیر بزرگ را کوچکتر می‌سازند:

$$\frac{1}{x}, \log_{10} x, \sqrt[4]{x}, \sqrt{x}$$

مثال: (فن تبدیل) جنگلبانی به طور تصادفی ۴۹ قطعه زمین را در جنگل بزرگی برمی‌گزیند و حجم چوب در هر قطعه زمین را بر حسب کورد(نام واحد حجم چوب است برابر با ۱/۲۸ فوت مکعب)تعیین می‌کند. بافت نگارهای حجم چوب با استفاده از تبدیل‌های $\sqrt{x}, \sqrt[4]{x}, \log_{10} x, \frac{1}{x}$

حجم چوب بر حسب کورد

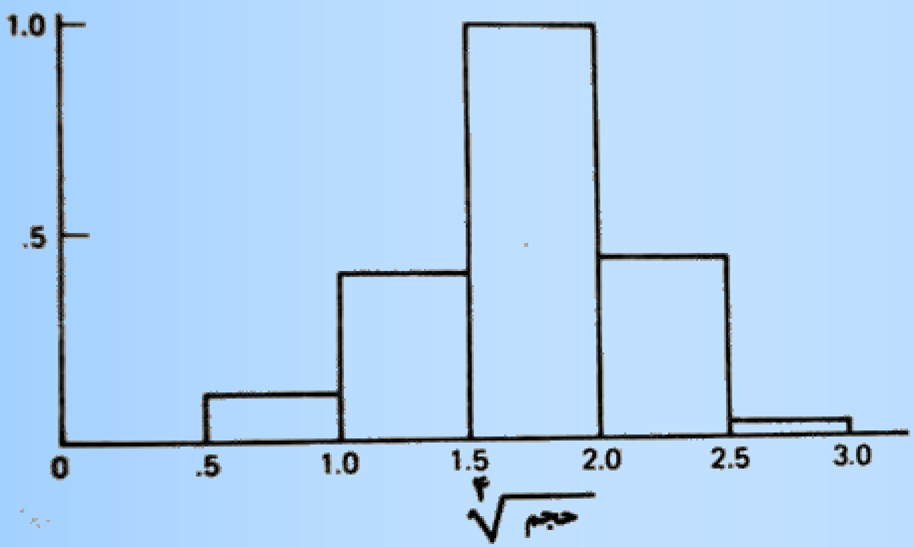
۶ر۵	۰ر۹	۶ر۳	۱۴ر۸	۳۹ر۳
۷ر۱	۱ر۳	۱۰ر۰	۸ر۳	۳ر۵
۷ر۹	۰ر۷	۱۶ر۸	۱۷ر۱	۶ر۰
۳ر۲	۱۷ر۷	۲۴ر۳	۲۶ر۲	۲ر۷
۵ر۹	۸ر۳	۵ر۲	۶ر۶	۷ر۴
۱۳ر۴	۸ر۳	۴۴ر۸	۸ر۳	۳ر۵
۱۲ر۰	۱ر۹	۱۴ر۱	۱۹ر۰	۱۹ر۴
۴ر۳	۱۶ر۷	۳ر۴	۱۰ر۳	۱۹ر۷
۳۱ر۷	۲۶ر۲	۲۸ر۳	۷ر۶	۱ر۰
	۱۰ر۰	۳ر۴	۱۸ر۹	۸ر۷



بافت نگار حجم چوب

داده های بدست آمده از تبدیل $\sqrt[4]{\text{حجم}}$

۲٫۵۰	۱٫۹۶	۱٫۵۸	۰٫۹۷	۱٫۶۰
۱٫۳۷	۱٫۷۵	۱٫۷۸	۱٫۰۷	۱٫۶۳
۱٫۵۷	۲٫۰۳	۲٫۰۲	۰٫۹۱	۱٫۶۸
۱٫۲۹	۲٫۲۶	۲٫۲۲	۲٫۰۵	۱٫۳۴
۱٫۶۴	۱٫۶۵	۱٫۵۱	۱٫۷۵	۱٫۵۶
۱٫۳۷	۱٫۷۵	۲٫۵۹	۱٫۷۵	۱٫۹۱
۲٫۰۷	۲٫۱۰	۱٫۹۳	۱٫۱۷	۱٫۸۶
۲٫۱۱	۱٫۷۹	۱٫۳۶	۲٫۰۲	۱٫۴۴
۱٫۰۰	۱٫۶۶	۲٫۳۱	۲٫۲۶	۲٫۳۷
۱٫۷۲	۲٫۰۹	۱٫۳۶	۱٫۷۸	



پایان