

بِه نام خدا

مفاهيم و روش های آماری

**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 نمونه‌های 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 پژوهش‌هایی که 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

بیان یافته‌ها

## جامعه:

جامعه (آماري)، 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 استخراج 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 داده‌ها

**الف) خلاصه کردن و توصیف الگوی کلی داده‌ها 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:

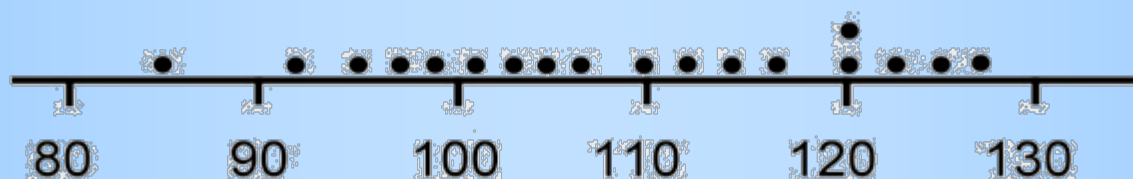
ابتدا خطی را رسم کرده و روی آن مقیاسی که حوزه اندازه‌ها را در بر گیرد، در نظر گرفته و هر یک از اندازه‌ها بر روی این خط به صورت نقطه‌ای پر رنگی مشخص می‌کنیم.

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

داده‌ها:

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

نمودار نقطه‌ای:



میزان سروصدا بر حسب دسی‌بل

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

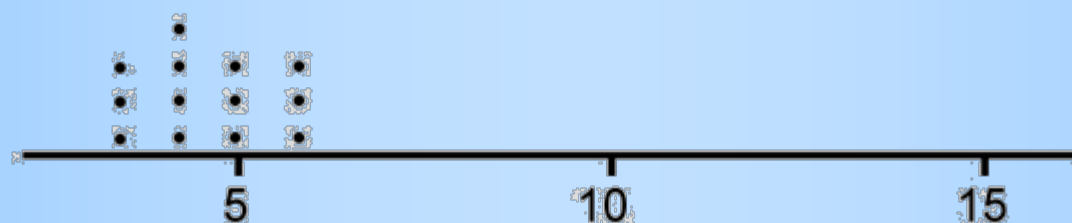
متوجه شد ماشین‌هایی که در یک جهت حرکت می‌کنند با فواصل

زمانی زیر بر حسب ثانیه، از مقابل او می‌گذرند:

داده‌ها: ۶ ۳ ۵ ۶ ۴ ۳ ۵ ۴ ۶ ۳ ۴ ۳ ۴ ۵ ۴ ۱۸

نمودار نقطه‌ای آن را رسم کنید.

داده ها: ۶ ۳ ۵ ۶ ۴ ۳ ۵ ۴ ۶ ۴ ۳ ۴ ۵ ۴ ۱۸



نمودار نقطه‌ای فاصله‌های زمانی بر حسب 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.

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



توزیع فراوانی داده های تمرکز ازن

فراوانی نسبی	فراوانی	خط نشان	(داده ها)
۰/۰۵۱	۴		۱/۰۵-۱/۵۵
۰/۰۳۸	۳		۱/۵۵-۲/۰۵
۰/۰۳۸	۳		۲/۰۵-۲/۵۵
۰/۰۳۸	۳		۲/۵۵-۳/۰۵
۰/۰۹۰	۷	+++	۳/۰۵-۳/۵۵
۰/۰۷۷	۶	+++	۳/۵۵-۴/۰۵
۰/۰۹۰	۷	+++	۴/۰۵-۴/۵۵
۰/۰۵۱	۴		۴/۵۵-۵/۰۵
۰/۰۹۰	۷	+++	۵/۰۵-۵/۵۵
۰/۱۶۷	۱۳	+++ ++	۵/۵۵-۶/۰۵
۰/۰۶۴	۵	+++	۶/۰۵-۶/۵۵
۰/۱۰۳	۸	+++	۶/۵۵-۷/۰۵
۰/۰۲۶	۲		۷/۰۵-۷/۵۵
۰/۰۲۶	۲		۷/۵۵-۸/۰۵
۰/۰۱۳	۱		۸/۵۵-۸/۰۵
	۰		۸/۰۵-۹/۵۵
۰/۰۲۶	۲		۹/۰۵-۹/۵۵
	۰		۹/۵۵-۱۰/۰۵
	۰		۱۰/۰۵-۱۰/۵۵
	۰		۱۰/۵۵-۱۱/۰۵
	۰		۱۱/۵۵-۱۱/۰۵
۰/۰۱۳	۱		۱۱/۰۵-۱۲/۵۵
۱/۰۰۱	۷۸		مجموع

## 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

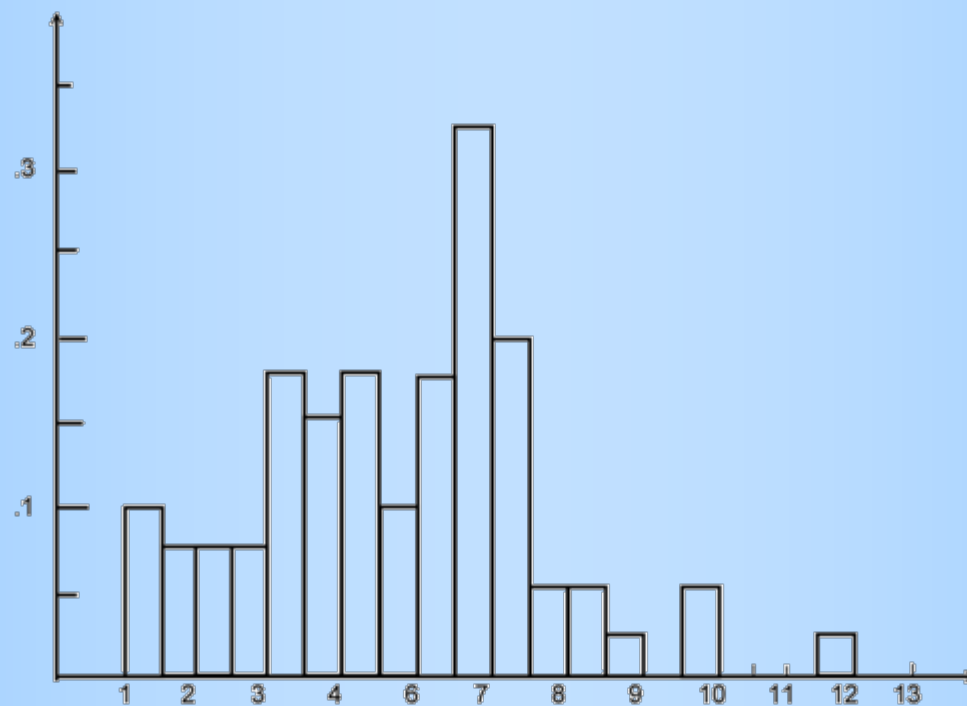
فراوانی نسبی	فراوانی	رده ها
۰/۰۹۰	۷	۰/۰۵-۲/۰۵
۰/۲۴۴	۱۹	۲/۰۵-۴/۰۵
۰/۳۹۷	۳۱	۴/۰۵-۶/۰۵
۰/۲۱۸	۱۷	۶/۰۵-۸/۰۵
۰/۰۳۸	۳	۸/۰۵-۱۰/۰۵
۰/۰۱۳	۱	۱۰/۰۵-۱۲/۰۵
۱/۰۰۰	۷۸	مجموع

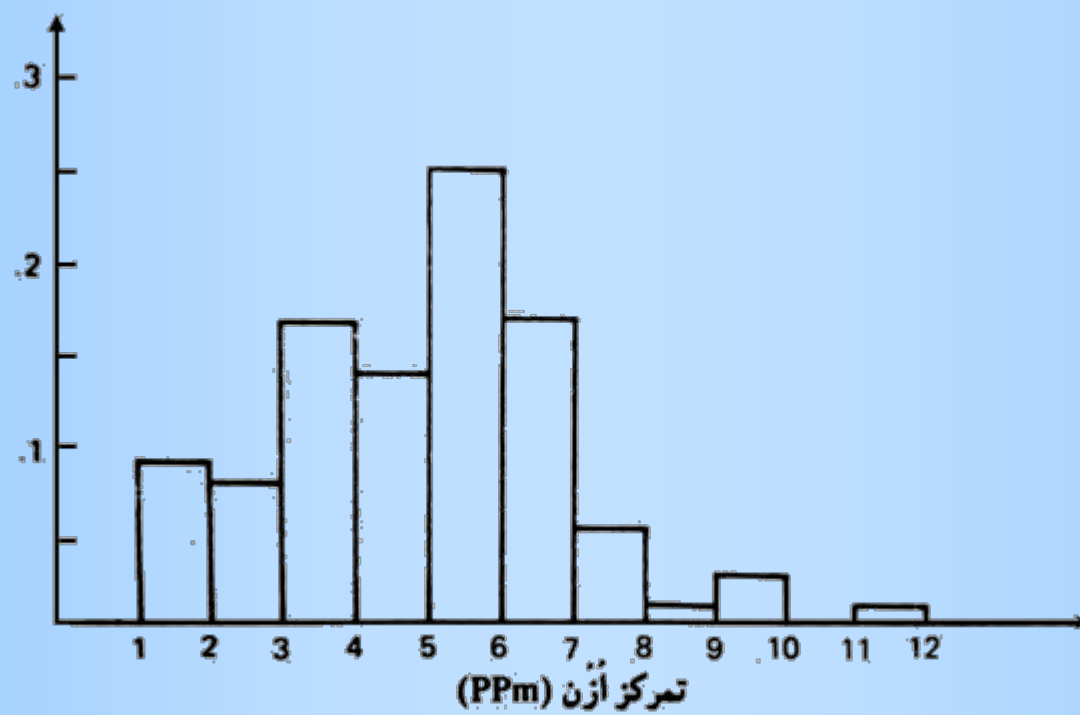
## بافت نگار فراوانی نسبی

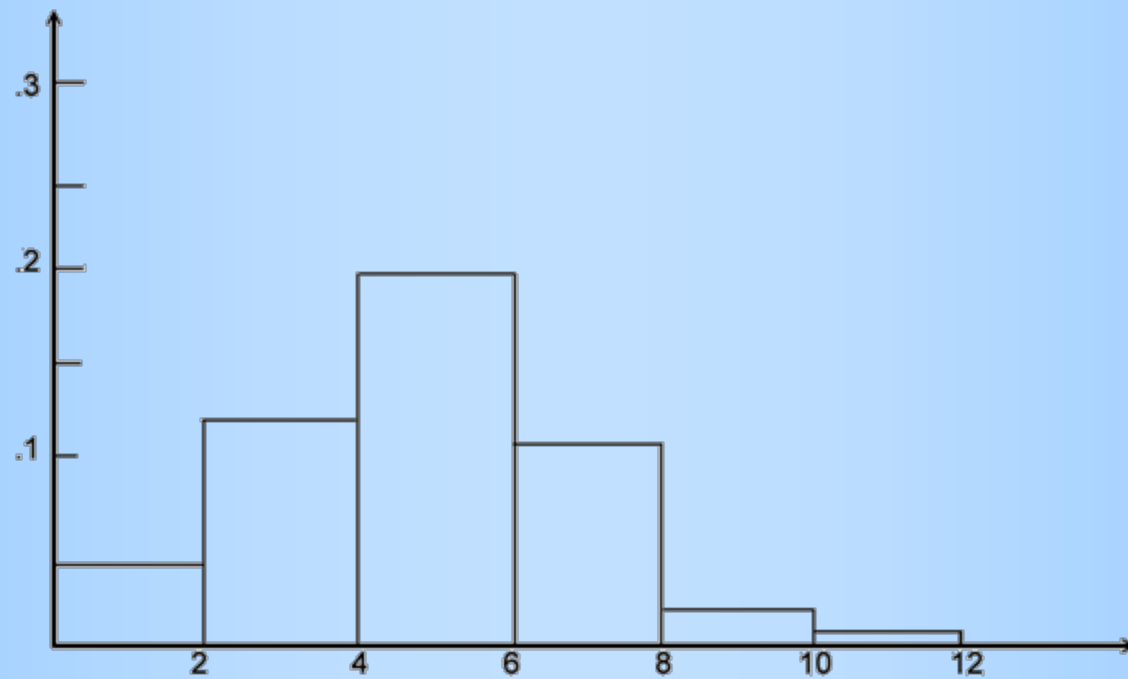
برای رسم بافت نگار فراوانی نسبی رده ها را روی محور افقی نمودار مشخص می کنیم. آنگاه روی هر رده ، مستطیلی همودی رسم می کنیم که مساحت آن مساوی با فراوانی نسبی آن رده باشد.

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

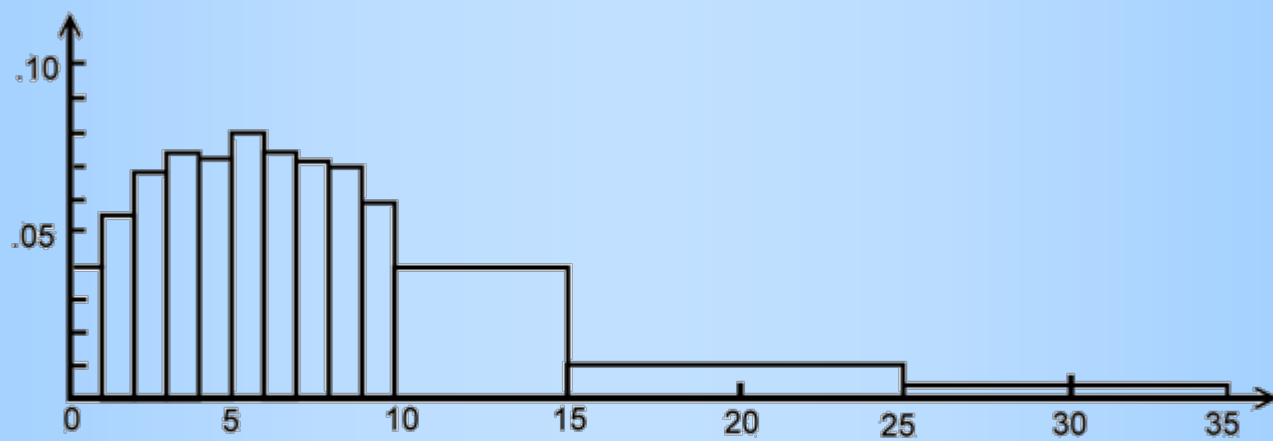
## بافت‌نگار فراوانی نسبی برای داده‌های تمرکز ازن (PPm)







تمرکز اُزن (PPm)



درآمد بر حسب هزار دلار



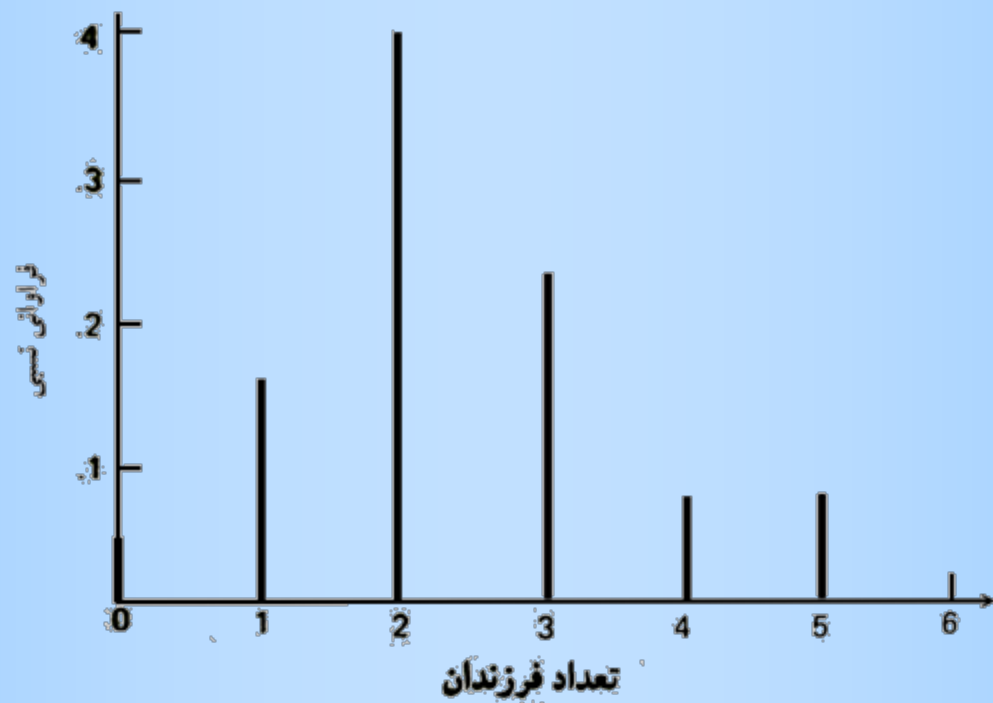
**مثال: توزیع فراوانی درآمد ۲۶۴۳۲۲۶ خانوار سفیدپوست**

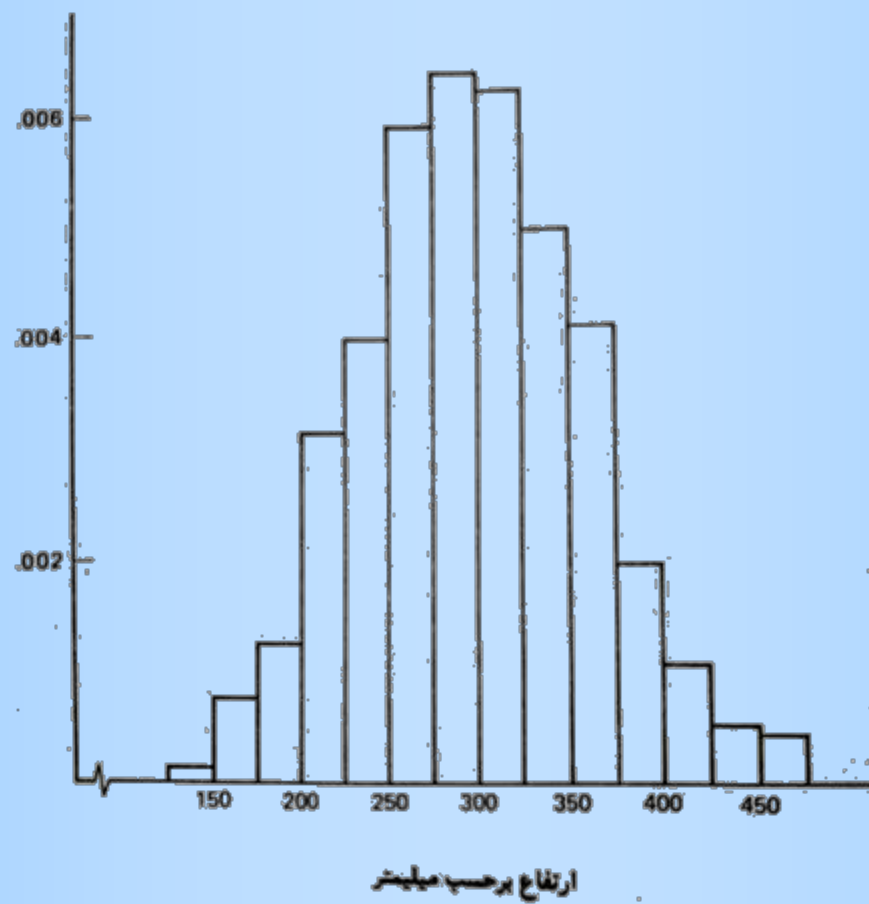
**مناطق روستایی آمریکا در سال ۱۳۴۹**

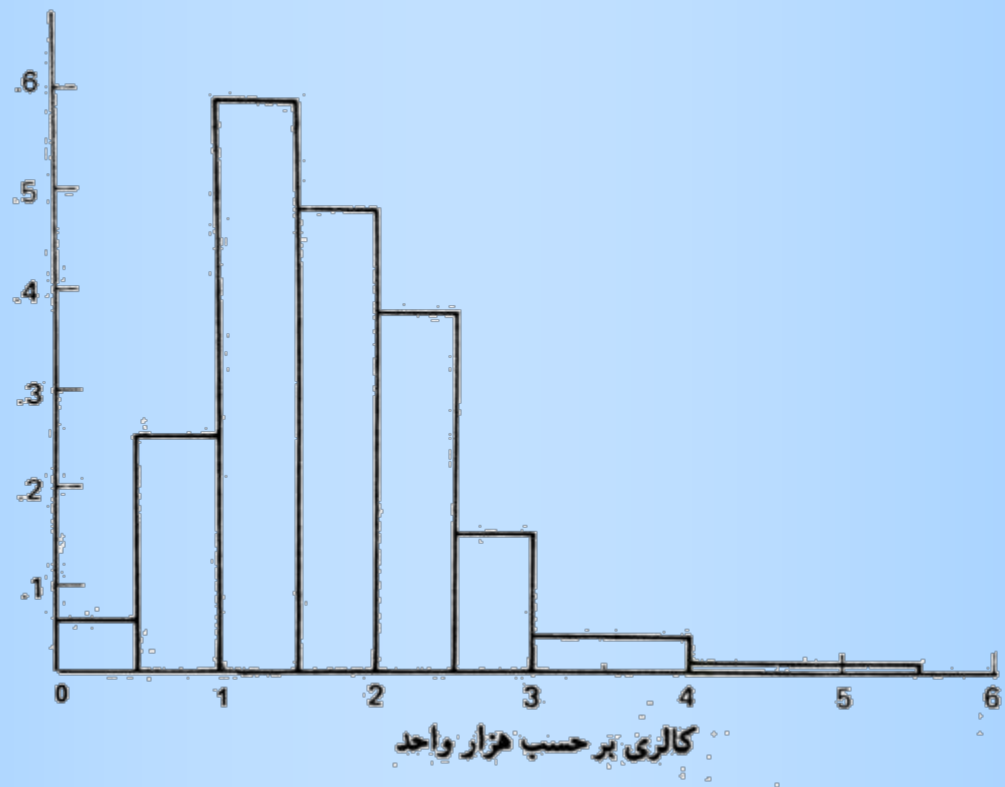
فراوانی نسبی ( % )	درآمد خانوار (برمب هزار دلار)
۴/۰	۰-۱
۵/۴	۱-۲
۶/۸	۲-۳
۷/۳	۳-۴
۷/۲	۴-۵
۸/۰	۵-۶
۷/۴	۶-۷
۷/۲	۷-۸
۷/۰	۸-۹
۵/۹	۹-۱۰
۱۹/۹	۱۰-۱۵
۱۰/۳	۱۵-۲۵
۳/۵	۲۵ و بالاتر
۹۹/۹	مجموع

## نمودار خطی فراوانی نسبی

مقادیر متمایز را به صورت نقاطی روی محور افقی مشخص می‌کنیم و سپس از نقاط حاصل، خطهایی عمود بر محور رسم می‌کنیم که ارتفاع هر یک برابر با فراوانی نسبی مقدار مربوطه باشد.





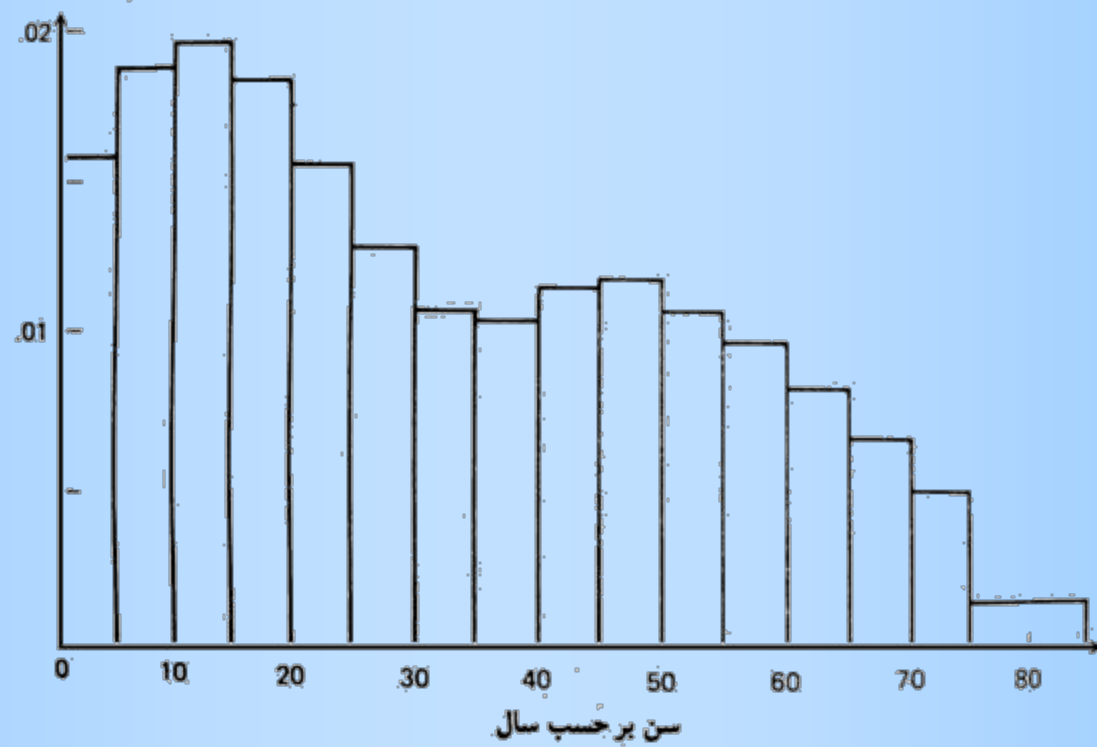


مثال:

داده‌ها:

### فراوانی تعداد فرزندان ۲۵ خانواده

تعداد فرزندان	فراوانی	فراوانی نسبی
۰	۱	۰/۰۴
۱	۴	۰/۰۱۶
۲	۱۰	۰/۰۴۰
۳	۶	۰/۰۲۴
۴	۲	۰/۰۰۸
۵	۲	۰/۰۰۸
مجموع	۲۵	۱/۰۰۰



## نمادگذاری مجموعه‌های داده‌ها و عمل جمع

### نماد جمع ( $\Sigma$ )

نماد  $\sum_{i=1}^n x_i$  نشان دهنده‌ی مجموع  $n$  عدد  $x_1, x_2, \dots, x_n$

است و این نماد به این صورت خوانده می‌شود:

مجموع تمام  $x_i$  ها، که  $i$  از ۱ تا  $n$  تغییر می‌کند.

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



**مثال:** فرض کنید:  $x_{\mu}=1$  ,  $x_{\nu}=1$  ,  $x_{\rho}=0$  ,  $x_1=1$

$$\sum_{i=1}^{\mu} x_i = x_1 + x_{\rho} + x_{\nu} + x_{\mu} = 1 + 0 + 1 + 1 = 3$$

$$\sum_{i=1}^{\mu} \nu x_i = x_1 + x_{\rho} + x_{\nu} + x_{\mu} = 1 + 0 + 1 + 1 = 3$$

$$\sum_{i=1}^K (x_i - \mu) = (x_1 - \mu) + (x_p - \mu) + (x_w - \mu) + (x_K - \mu)$$

$$= \sum_{i=1}^K x_i - K(\mu) = 10 - 4 = 6$$

$$\sum_{i=1}^K (x_i - \mu)^p = (x_1 - \mu)^p + (x_p - \mu)^p + (x_w - \mu)^p + (x_K - \mu)^p$$

$$= (1 - \mu)^p + (9 - \mu)^p + (1 - \mu)^p + (1 - \mu)^p$$

$$= 1 + 9 + 1 + 1 = 10$$

## بعضی از خواص اصلی جمع

اگر  $a$  و  $b$  اعداد ثابتی باشند ، داریم

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

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

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

## معیارهای گرایش به مرکز

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

**متداول ترین معیارهای گرایش به مرکز عبارتند از:**

**میانگین**

**میان**

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

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

**میانه نمونه‌ای** مرکب از  $n$  اندازه  $x_1, x_2, \dots, x_n$  عبارت است، از اندازه‌ی وسطی، در صورتی که اندازه‌ها را بترتیب از کوچکترین به بزرگترین مقدار مرتب کرده باشیم. اگر  $n$  عدد فردی باشد، یک مقدار وسطی منحصر بفرد وجود دارد که میانه است. اگر  $n$  زوج باشد دو مقدار وسطی وجود 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

$۲/۹$      $۳/۵۴$      $۳/۶۷$      $۴/۱۷$      $۴/۷۶$



**مثال:** درآمد ماهانه ی ۸ کارمند رسمی یک شرکت به

تومان عبارت است از :

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

میانگین و میانه ی درآمدها را محاسبه کنید.

ابتدا داده‌ها را از کوچک به بزرگ مرتب می‌کنیم

۵۰۰۰ ۵۵۰۰ ۵۵۰۰ ۵۵۰۰ ۶۰۰۰ ۷۰۰۰ ۷۵۰۰ ۲۰۰۰۰

میانگین:

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

میان:

$$\frac{۵۵۰۰+۶۰۰۰}{۲} = ۵۷۵۰ \text{ تومان}$$

**صدک  $(100p)$  ام نمونه، مقداری است که وقتی داده ها از کوچکترین تا بزرگترین مقدار مرتب شدند، حداقل  $100p\%$  از مشاهدات منطبق بر این مقدار یا در سمت چپ (زیر) آن و حداقل  $(1-p)100\%$  از مشاهدات منطبق بر این مقدار یا در سمت راست (بالای) آن باشند.**

## چارکهای نمونه

چارک (اول) کوچکتر

چارک دوم (یا میانه)

چارک (سوم) بزرگتر

صدک ۲۵ ام  $Q_1$

صدک ۵۰ ام  $Q_2$

صدک ۷۵ ام  $Q_3$

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

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

چارک‌ها را برای این داده‌ها محاسبه کنید.

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

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

چارک اول:

تعداد مقادیر نابیشتر از  $Q_1$  حداقل برابر است با  $۲/۲۵ \times ۹ = ۰/۲۵$

تعداد مقادیر ناکمتر از  $Q_1$  حداقل برابر است با  $۶/۷۵ \times ۹ = ۰/۷۵$

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

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

با شمارش سه مشاهده از بزرگترین مقدار به طرف پایین،  
۱۳۲/۱ چارک سوم است.

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

میانۀ 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 صدی 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	۶۸/۳
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  $۱۲/۵ = ۵۰ \times ۲۵/۰$  مشاهده را از  
کوچکترین اندازه به بالا و حداقل  $۳۷/۵ = ۵۰ \times ۷۵/۰$  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 ۳۱ مشاهده از بزرگترین اندازه به طرف پایین

چاری 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	۶۰/۷	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 صدی ۱۰ ام باید حداقل  $۵ = ۵۰ \times ۱۰ / ۰$  مشاهده نابیشتر از آن ، و  $۴۵ = ۵۰ \times ۹۰ / ۰$  مشاهده ناکمتر از آن باشند. پنجمین و ششمین مشاهده از لحاظ کوچکی هر دو این شرایط را دارا هستند بنابراین متوسط آنها یعنی  $۵۵/۸۵ = (۵۵/۸ + ۵۵/۹) / ۲$  صدی دهم است. پس فقط ۱۰٪ از ۵۰ اندازه میزان سر و صدا خفیفتر از ۵۵/۸۵ دسی بل هستند.

**طرز به دست آوردن میانگین نمونه‌ی پیراسته**

**الف) تمام مشاهدات کوچکتر از چارک اول را برمی‌داریم.**

**تمام مشاهدات بزرگتر از چارک سوم را برمی‌داریم.**

**ب) میانگین مشاهدات باقیمانده را حساب می‌کنیم.**

## **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 یک از مشاهدات بزرگتر از چارک سوم، مقدار چارک سوم را می‌گذاریم. بقیه مشاهدات را تغییر نمی‌دهیم.**

**ب) پس از این اصلاحات، میانگین تمام مشاهدات را، محاسبه می‌کنیم.**





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

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

چارک اول

چارک سوم

$$\text{میانگین وینزوری} = \frac{106/5 + 106/5 + 106/5 + 116/6 + 125/0 + 128/3 + 132/1 + 132/1 + 132/1}{9}$$

$$= \frac{1085/7}{9} = 120/6$$

## معیارهای پراکندگی عبارتند از:

- واریانس
- انحراف معیار
- دامنه نمونه
- دامنه میان چارگی نمونه

**واریانس:** واریانس ( $S^2$ ) نمونه‌ای مرکب از  $n$  اندازه

$x_1, x_2, \dots, x_n$  بصورت زیر تعریف می‌شود:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

**انحراف معیار:** انحراف معیار نمونه عبارت است از:

$$S = \sqrt{\text{واریانس}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

**دامنه:**

کوچکترین مشاهده – بزرگترین مشاهده = دامنه نمونه

**دامنه میان چارکی:**

چارک اول – چارک سوم = دامنه میان چارکی نمونه

**مثال:** در یک آزمایش روانشناسی، یک علامت محرکه با شدت ثابت بر روی ۶ فردی که مورد آزمایش قرار می گیرند، به کار می رود. زمانهای عکس العمل آنها، که بر حسب ثانیه ثبت شده، عبارت اند از ۴، ۲، ۳، ۳، ۶، ۳. میانگین، واریانس و انحراف معیار داده ها را محاسبه کنید.

$x$	۴	۲	۳	۳	۶	۳	$\sum x_i = ۲۱$ $\bar{x} = \frac{۲۱}{۶} = ۳.۵$
$(x - \bar{x})$	۰.۵	-۱.۵	-۰.۵	-۰.۵	۲.۵	-۰.۵	۰
$(x - \bar{x})^p$	۰.۲۵	۲.۲۵	۰.۲۵	۰.۲۵	۶.۲۵	۰.۲۵	$۹.۵۰ = (x - \bar{x})^p$

میانگین نمونه:  $\bar{x} = ۳.۵$  ثانیه

$$s^p = \frac{\sum_{i=1}^n (x_i - \bar{x})^p}{n-1} = \frac{۹.۵۰}{۵} = ۱.۹ (\text{ثانیه})^p \quad \text{واریانس نمونه:}$$

$$s = \sqrt{۱.۹} = ۱.۴ \quad \text{انحراف معیار نمونه:}$$

## قاعده چبیشف

برای هر مجموعه از داده ها:

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

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

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 ها است.

مثال: قاعده چیشف 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		

$$\bar{x} = ۴/۹۷$$

$$S = ۲/۰۰$$

$$1 - \frac{1}{(1/5)^p} = ۰/۵۶ : \bar{x} - 1/5S$$

$$1 - \frac{1}{p^p} = ۰/۷۵ : \bar{x} - ۲S$$

$$1 - \frac{1}{\mu^p} = ۰/۸۹ : \bar{x} - ۳S$$



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

**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 = دامنہ میان چارکی نمونہ**

**مثال:** دامنه و میان چارگی داده های سروسدا 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**

$$= ۷۷/۱ - ۵۲/۰$$

$$= ۲۵/۱ \text{ دسی بل}$$

**دامنه میان چارگی =  $Q_p - Q_1$**

$$= ۶۴/۶ - ۵۷/۲$$

$$= ۷/۴ \text{ دسی بل}$$

## تذکراتی درباره‌ی معیارهای 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

## بطور خلاصه:

الف) ارائه نموداری برای نمایش شکل کلی داده‌ها، یکی از مراحل اساسی توصیف اختصاری داده‌هاست.

ب) اگر با توجه به هدفی که از توصیف اختصاری داده‌ها داریم، معیارهای گوناگونی به صورت رقیب یکدیگر موجود باشند، بهتر است به جای استفاده دلخواه از یک معیار، همه‌ی آنها را به کار ببریم.

ج) در تعبیر معیارهای توصیفی، شکل کلی توزیع داده‌ها را همیشه باید 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

$m_1, m_2, \dots, m_k$  و فراوانی‌های متناظر  $f_1, f_2, \dots, f_k$

باشند آنگاه:

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



**واریانس نمونه گروه بندی شده عبارت است از:**

$$s^2 = \frac{\sum_{i=1}^k (m_i - \bar{x})^2 f_i}{n}$$

**فرمول دیگر برای واریانس نمونه گروه بندی شده:**

$$s^2 = \frac{\sum_{i=1}^k m_i^2 f_i}{n} - \bar{x}^2 \text{ (برای محاسبه دستی)}$$

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

الف) ستون فراوانی تجمعی را محاسبه می‌کنیم.

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 ( $100p$ )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 رده‌ای 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 بر حسب سال)	فراوانی تجمعی	فراوانی
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،

$$np = 200(0/5) = 100$$

$$p = 0/5$$

فراوانی آن  $f=62$

دهی میانه  $29/5 - 24/5$

$$\text{سال} = 24/5 + \frac{(100-92)}{62} \times 5 = 25/15$$

$$Q_1 = 19/5 + \frac{32}{74} \times 5 = 21/66 \text{ سال}$$

$$Q_3 = 24/5 + \frac{(150-92)}{62} \times 5 = 29/18 \text{ سال}$$

$$Q_3 - Q_1 = 29/18 - 21/66 = 7/52 \text{ سال}$$

## کدگذاری داده‌ها

**کدگذاری مجموعه‌ای از داده‌ها، عبارت است از عملیاتی که در طی آن از هر مشاهده، عدد ثابتی را کم (یا به آن اضافه) کرده و نتیجه را بر عدد ثابتی تقسیم (یا در آن ضرب) می‌نمایند. این عمل به منظور ساده کردن محاسبه معیارهای توصیفی انجام می‌گیرد، ولی در حال حاضر این کار توسط کامپیوتر انجام می‌گیرد. کدگذاری باعث صرفه‌جویی در وقت نمی‌شود. معهذا نتایج حاصل از آن برای نشان دادن اثراتی که تغییرات واحد اندازه‌گیری می‌تواند بر روی معیارهای توصیفی داشته باشد مفید است.**

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

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

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

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

**مثال:** اندازه گیری طول یک قسمت از بزرگراه جدیدی را ۵

مرتبه تکرار کرده و مقادیر زیر بدست آمده:

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

می‌خواهیم داده‌ها را به طور مناسب کدگذاری کرده و

مقدار میانگین و انحراف محیار را پیدا کنیم.



با نگاه به داده‌ها، عددی نزدیک به مرکز مشاهدات مثلاً  $b=100/18$  را انتخاب می‌کنیم. چون تفاوت مشاهدات با این مقدار مضارب  $0/01$  هستند،  $a$  را برابر با  $0/01$  می‌گیریم. یعنی،

$$u_i = \frac{1}{0/01} (x_i - 100/18)$$

بنابراین مقادیر کدها عبارتند از:  $0, -1, 0, 1, -1$

$$\bar{u} = \frac{-1+1+0-1+0}{5} = -0/2$$

و واریانس برابر است با:

$$S_u^p = \frac{(-1+o/p)^p + (1+o/p)^p + (o/p)^p + (-1+o/p)^p + (o/p)^p}{5-1} = o/Vo$$

$$S_u = \sqrt{S_u^p} = o/\wedge \wp V$$

با مراجعه مجدد به مقادیر اولیه  $x$ ، اکنون داریم  $x_i = (o/o\wedge) u_i + 1o\circ/o\wedge$

$$\bar{x} = (o/o\wedge) \bar{u} + 1o\circ/o\wedge = 1o\circ/o\wedge \wedge$$

$$S_x^p = (o/o\wedge)^p S_u^p = (o/o\wedge)^p (o/V) = o/o\circ\circ\circ V$$

$$S_x = (o/o\wedge) S_u = (o/o\wedge) (o/\wedge \wp V) = o/o\circ\wedge \wp V$$

# فصل ۳

# مبادی احتمال

## مبادی احتمال

### فضای نمونه و پیشامدها

**آزمایش**، عبارت از فرایند گردآوری داده‌های مربوط به 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 نمونه** خوانده می‌شود. فضای نمونه را  $\Omega$  با نشان می‌دهند.

مجموعه‌ای از برآمدهای مقدماتی را که با توصیف بعضی از خصوصیاتشان مشخص شده‌اند **پیشامد** گوئیم. پیشامد، زیر مجموعه‌ای از فضای نمونه  $\mathcal{S}$  است. چند حرف بزرگ اول الفبای لاتین ( $A, B, C, \dots$ )، معمولاً برای نشان دادن پیشامدها به کار می‌روند.

وقتی یکی از برآمدهای مقدماتی موجود در  $A$  رخ می‌دهد، گوئیم **پیشامد  $A$  رخ می‌دهد**.

فضای نمونه‌ای را که تعداد عناصر آن متناهی یا به طور 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 یک پیشامد، نسبت دفعاتی است که انتظار می‌رود آن پیشامد رخ بدهد، وقتی که آزمایش تحت شرایط یکسان تکرار گردد.

علامت  $P(A)$  برای نشان دادن احتمال پیشامد  $A$  به کار می‌رود.



## برآمدهای مقدماتی همشانس

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

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

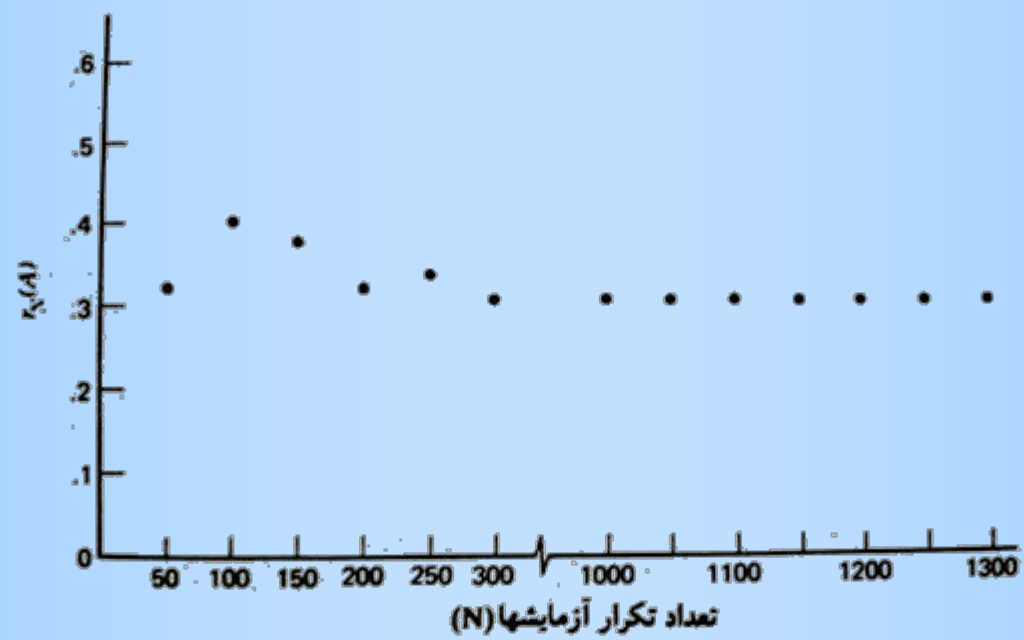
گوییم فضای نمونه‌ای که عناصرش همشانس‌اند دارای مدل احتمال یکنواخت است.

## پایداری فراوانی نسبی

فراوانی نسبی پیشامد A در N بار تکرار آزمایش

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

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



## شرایط مدل احتمال برای فضاهای نمونه گسسته

**احتمال**، تابعی است که روی پیشامدها تعریف می‌شود، و در شرایط زیر صدق می‌کند:

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

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

$$P(A) = \sum P(e_i)$$

تمام  $e_i$  های متعلق به  $A$

$$P(\mathcal{S}) = \sum P(e_i) = 1 \quad \text{(سه)}$$

تمام  $e_i$  های متعلق به  $\mathcal{S}$

**مثال:** آزمایش جنس فرزند اول و دوم در خانواده‌های دارای دو

فرزند، چهار برآمد دارد (پسر، پسر)  $e_1$ ، (پسر، دختر)  $e_p$ ، (دختر،

پسر)  $e_{ps}$ ، (دختر، دختر)  $e_{ff}$ . آیا هر دسته از احتمالات نسبت داده

شده زیر در شرایط مدل احتمال صدق می‌کند؟

$$P(e_1) = P(e_p) = P(e_{ps}) = P(e_{ff}) = \frac{1}{4} \quad \text{الف)}$$

$$P(e_1) = \frac{3}{16}, P(e_p) = \frac{3}{8}, P(e_{ps}) = \frac{1}{4}, P(e_{ff}) = \frac{3}{16} \quad \text{ب)}$$

$$P(e_1) = \frac{1}{9}, P(e_p) = \frac{1}{4}, P(e_{ps}) = \frac{1}{9}, P(e_{ff}) = \frac{1}{9} \quad \text{ج)}$$

## عملیات روی پیشامدها و قوانین پایه ای احتمال

**نمودار ون:** فضای نمونه  $\Omega$  را به عنوان مجموعه ای از نقاط در یک نمودار در نظر بگیرید که در آن نمودار، هر نقطه متناظر است با یک برآمد مقدماتی مشخص آزمایش. در این صورت هر پیشامد در نمودار به صورت مجموعه نقاطی که در خصوصیت توصیف کننده ی پیشامد صدق می کنند معرفی می شود.

**مثال:** وقتی یک مهری آزمایشی روی میوانی به کار می‌رود، یا میوان پاسخ می‌دهد یا در پاسخ دادن عاجز می‌ماند. به عبارت دیگر، تنها دو برآمد ممکن وجود دارد: یا میوان پاسخ می‌دهد (R) یا پاسخ نمی‌دهد (N). آزمایش، عبارت است از اعمال تمریک‌های متوالی بر سه میوان 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 میوان پاسخ می‌دهد.

B: در اولین تمریک، پاسخ وجود دارد.

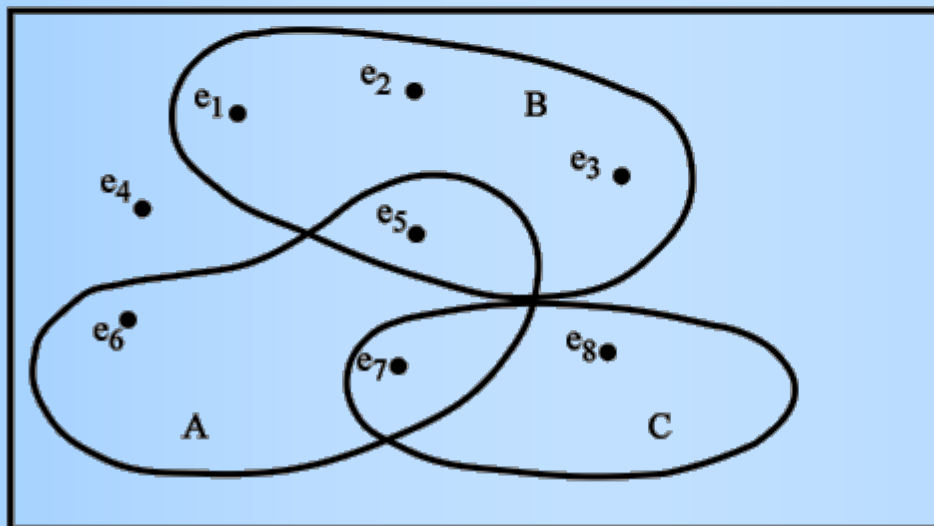
C: هر دو میوان اول و دوم در پاسخ دادن عاجزند.

تعداد برآمد مقدماتی  $۲ \times ۲ \times ۲ = ۸$

RRR( $e_1$ )    RRN( $e_p$ )    RNR( $e_{\text{پ}}$ )    NRR( $e_{\text{ف}}$ )

RNN( $e_{\delta}$ )    NRN( $e_{\psi}$ )    NNR( $e_v$ )    NNN( $e_{\lambda}$ )

$A = \{e_{\delta}, e_{\psi}, e_v\}$      $B = \{e_1, e_p, e_{\text{پ}}, e_{\delta}\}$      $C = \{e_v, e_{\lambda}\}$





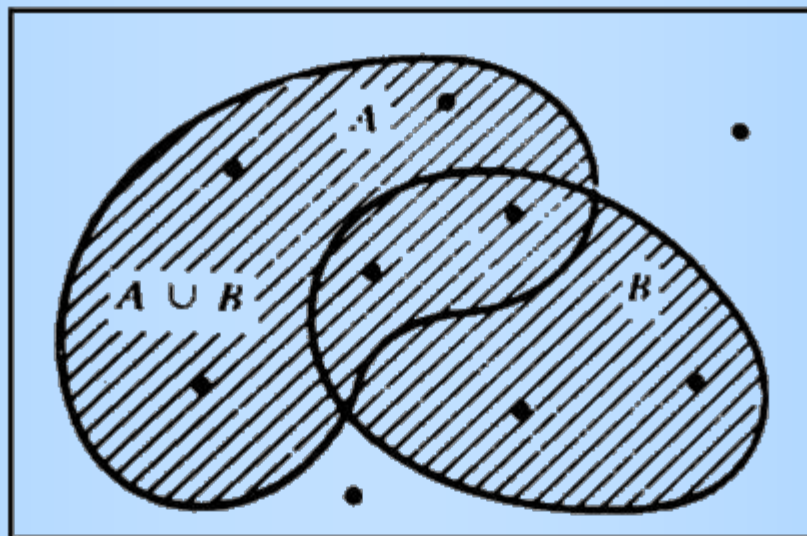
سه عمل پایه ای روی پیشامدها عبارت است از :

• اجتماع

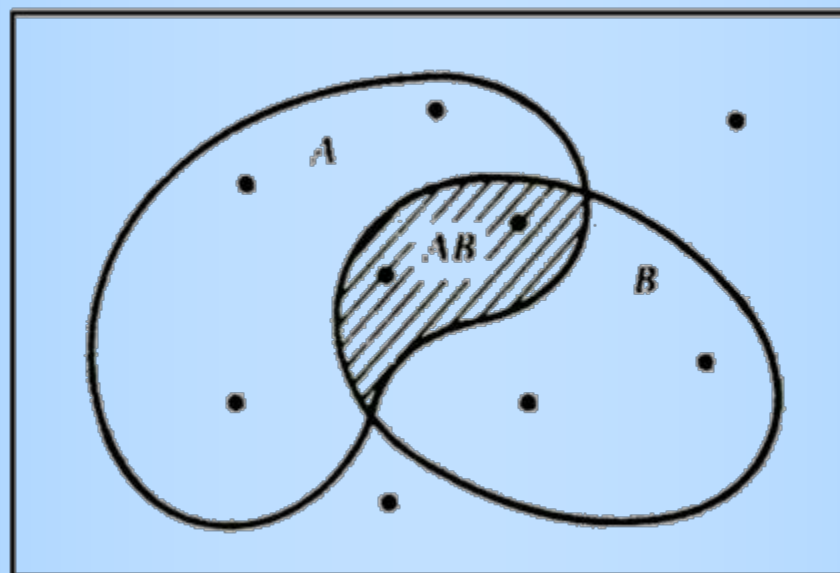
• اشتراک

• متمم

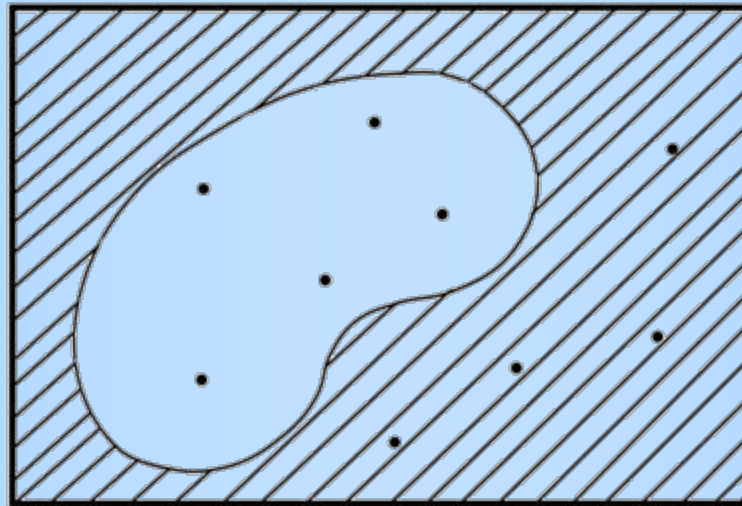
**اجتماع دو پیشامد A و B ، مجموعه تمام برآمدهای مقدماتی است که در A، یا در B و یا هم در A و هم در B ، قرار دارند. اجتماع A و B به صورت  $A \cup B$  نشان داده می‌شود، که با  $B \cup A$  یکی است. وقوع  $A \cup B$  به معنی این است که حداقل یکی از دو پیشامد A و B رخ می‌دهد.**



**اشتراک دو پیشامد A و B** ، مجموعه تمام برآمدهای مقدماتی 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 A و B رخ می دهند.



**متمم پیشامد  $A$** ، مجموعه‌ی تمام برآمدهای مقدماتی است که در  $A$  نیستند. متمم  $A$  به صورت  $A^c$  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

$$A = \{e_6, e_7, e_8\} \quad B = \{e_1, e_2, e_3, e_4\} \quad C = \{e_5, e_8\}$$

ساختار هر یک از پیشامدهای زیر را مشخص کنید:

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

$$A \cup B = \{e_b, e_p, e_{pb}, e_\Delta, e_\gamma, e_v\}$$

$$AC = \{e_v\}$$

$$BC = \emptyset$$

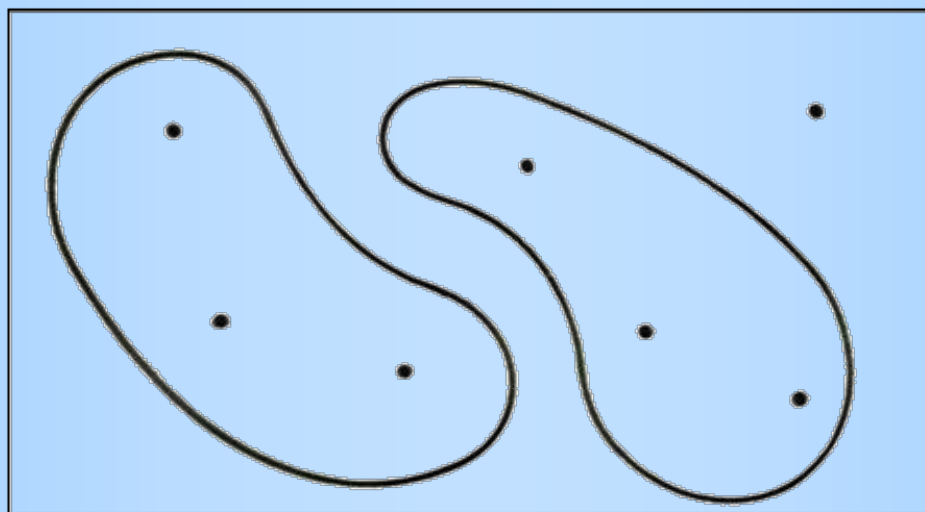
$$B^c = \{e_F, e_\gamma, e_v, e_\Lambda\}$$

$$(A \cup B)^c = \{e_F, e_\Lambda\}$$

$$A^c = \{e_b, e_p, e_{pb}, e_F, e_\Lambda\}$$

$$A^c \cup B^c = \{e_b, e_p, e_{pb}, e_F, e_\gamma, e_v, e_\Lambda\}$$

یکی از حالات ممکن دو پیشامد  $A$  و  $B$  این است که هیچ برآمد  
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؛ یعنی،  
اگر اشتراک آنها،  $AB$ ، تهی باشد. پیشامدهای جدا از هم  
نمی‌توانند به طور همزمان رخ دهند.



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

قانون متمم گیری

$$P(A^c) = 1 - P(A)$$

## قانون جمع برای اجتماع پیشامدهای جدا از هم

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

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

## قانون کلی جمع

برای هر دو پشامد A و B،

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

**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  $۶۴ = ۲ \times ۲ \times ۲ \times ۲ \times ۲ \times ۲$  است. مدل احتمال یکنواختی با ۶۴ برآمد مقدماتی داریم.

$A$  = حداقل یک علامت R در مجموعه جواب‌ها باشد.

$A^c$  = همه علامتها W هستند.

$$P(A^c) = 1 - \frac{1}{64} \qquad P(A^c) = \frac{1}{64}$$

ب) احتمال اینکه در مجموعه جواب‌های شاگرد پنج یا شش

علامت R وجود داشته باشد چقدر است؟

فرض کنید:  $A_5 =$  دقیقاً ۵ تا R.

(WRRRRR), (RWRRRR), (RRWRRR)

(RRRWRR), (RRRRWR), (RRRRRW)

$A_4 =$  دقیقاً ۴ تا R. (RRRRRR)

$$P(A_4) = \frac{1}{64}$$

$$P(A_5) = \frac{6}{64}$$

$$P(R \text{ علامت } 4 \text{ یا } 5) = P(A_5 \cup A_4)$$

$$= P(A_5) + P(A_4)$$

$$= \frac{6}{64} + \frac{1}{64} = \frac{7}{64}$$

**مثال:** گروهی از شاگردان دبیرستانی را برگزیده و آنها را بر مبنای قابلیتشان در درسهای ریاضیات و ورزش ردهبندی کردهایم نسبتهای تعداد شاگردان بر مبنای قابلیتشان در درسهای ریاضیات و ورزش. اگر شاگردی را بطور تصادفی از این گروه برگزینیم، احتمال آنکه

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



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

$$P(A) = 0/06 + 0/11 + 0/18 = 0/35$$

$$P(B) = 0/06 + 0/12 + 0/16 = 0/34$$

$$P(AB) = 0/06$$

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

$$= 0/35 + 0/34 - 0/06$$

$$= 0/63$$

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

ب) شاگرد در ریاضیات ضعیف نباشد چیست؟

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

$$P(D) = 0/16 + 0/05 + 0/04 = 0/25$$

$$P(D^c) = 1 - 0/25 = 0/75$$

## قواعد شمارش و کاربرد آنها در مدل‌های احتمال یکنواخت

### احتمال یک پیشامد $A$

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

### قاعده ضرب

هرگاه آزمایشی مرکب از دو مرحله باشد به طوری 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 نتایج ممکن آزمایش،  $k \times l$  خواهد بود.

**مثال:** شرکتی برای مطرح کردن دعوایی، می تواند در مرحله ی 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 می تواند بگیرد؟

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

## قاعده جایگشت‌ها یا ترتیبا

تعداد ترتیب‌های مختلفی که می‌توان با  $r$  شی، برگزیده از گروهی مرکب از  $n$  شی متمایز ساخت، با  $P_r^n$  نشان داده می‌شود و به صورت "تعداد جایگشت‌های  $r$  از  $n$ " خوانده می‌شود.

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

$$P_r^n = \underbrace{n(n-1)(n-2)\dots}_{r \text{ عامل}}$$

$$= n(n-1)(n-2)\dots(n-r+1)$$

**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، می‌خواهیم تعداد ترتیبهای ممکن سه نفر برگزیده شده از پانزده مسابقه دهنده را که بتوان در این سه جعبه قرار داد بدیابیم. بنا به قاعده‌ی جایگشت‌ها، جواب برابر است با:

$$P_{15}^3 = 15 \times 14 \times 13 = 2730.$$

تعداد ترتیبهای ممکن  $n$  شی متمایز برابر است با:

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

این حاصل ضرب را با  $n!$  نیز نشان داده و آن را  $n$  فاکتوریل می خوانند.



**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:

$$P_3^4 = \frac{4!}{1} = 24$$

## 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  $n$**  می‌خوانند. فرمول به صورت

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

داده می‌شود.

**مثال: مقادیر**  $\begin{pmatrix} 5 \\ \mu \end{pmatrix} \begin{pmatrix} 7 \\ 5 \end{pmatrix} \begin{pmatrix} 7 \\ \mu \end{pmatrix} \begin{pmatrix} 10 \\ \mu \end{pmatrix}$  را محاسبه کنید.

$$\begin{pmatrix} 5 \\ \mu \end{pmatrix} = \frac{5 \times 4 \times \mu}{1 \times \mu \times \mu} = 10$$

$$\begin{pmatrix} 7 \\ 5 \end{pmatrix} = \frac{7 \times 6 \times 5 \times 4 \times \mu}{1 \times \mu \times \mu \times 4 \times 5} = 21$$

$$\begin{pmatrix} 7 \\ \mu \end{pmatrix} = \frac{7 \times 6}{1 \times \mu} = 21$$

$$\begin{pmatrix} 10 \\ \mu \end{pmatrix} = \frac{10 \times 9 \times 8}{1 \times \mu \times \mu} = 120$$

## خواص فرمول ترکیب

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

$$\binom{n}{n} = 1, \binom{n}{1} = n$$

**تعریف:  $0! = 1$ .**

$$\binom{n}{n} = 1 = \binom{n}{0}, \text{ در نتیجه،}$$

**مثال:** جعبه‌ای شامل ۱۲ کالا داریم از آنها ۸ تا سالم  
(با علایم  $G_1, G_2, \dots, G_8$ ) و ۴ تا معیوب (با علایم  $D_1, D_2, D_3, D_4$ )  
است. مال 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 نمونه‌های 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 استفراخ 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  $\binom{4}{2}$ ، و تعداد ترکیب‌های ۱ کالای سالم از میان ۸

کالای سالم ۸ کالای سالم برابر است با  $\binom{8}{1}$ . تعداد نتایج (بنا

به قاعده ضرب) ممکن برابر است با:

$$\binom{4}{2} \binom{8}{1} = \frac{4 \times 3}{1 \times 2} \times \frac{8}{1} = 12$$



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:

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

به عنوان یک راهنمایی کلی، بهتر است به خاطر داشته باشید که قاعده جایگشت‌ها وقتی به کار می‌رود که لازم است پیشامدی بر حسب 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:

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

$A_3$ : پیشامد اینکه دقیقاً ۳ نفر موافق است.

$$P(A_p \cup A_3) = P(A_p) + P(A_3)$$

$$P(A_p) = \frac{\binom{9}{2} \binom{4}{1}}{\binom{15}{3}} = \frac{72}{455}$$

$$P(A_3) = \frac{\binom{9}{3} \binom{4}{0}}{\binom{15}{3}} = \frac{84}{455}$$

$$P(\text{حداقل دو نفر موافق}) = \frac{72}{455} + \frac{84}{455} = \frac{156}{455}$$

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

$$P(\text{دو نفر اول موافق و نفر سوم مخالف}) = \frac{9 \times 8 \times 4}{15 \times 14 \times 13} = \frac{288}{2730}$$

نمونه‌ای به حجم  $n$ ، برگزیده شده از جامعه‌ای شامل  $N$  شیء متمایز، نمونه‌ای تصادفی نامیده می‌شود اگر گزینش هر مجموعه‌ای به حجم  $n$  دارای احتمال  $1/\binom{N}{n}$  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 کمتر از حد معمول	وزن معمولی	وزن بیشتر از حد معمول	مجموع
فشار خون دارند	۰/۰۲	۰/۰۸	۰/۱۰	۰/۲۰
فشار خون ندارند	۰/۲۰	۰/۴۵	۰/۱۵	۰/۸۰
مجموع	۰/۲۲	۰/۵۳	۰/۲۵	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،  $P(A)=0/2$ .

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

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؟

$$P(A|B) = \frac{0/10}{0/25} = 0/4$$

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

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

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

فرض کنید:

$S_1$  = پیشامد گزینش تخته مرغ لکه دار در استخراج اول

$S_p$  = پیشامد گزینش تخته مرغ لکه دار در استخراج دوم

$C_3$  = تخته مرغ سالم در استخراج سوم

احتمال اشتراک  $(S_1 S_p C_3)$  این سه پیشامد را بیابید:

$$P(S_1 C_p C_w) = P(S_1) P(S_p | S_1) P(C_w | S_1 S_p)$$

چون استخراج ها به طور تصادفی انجام می گیرند، داریم

$$P(S_1) = \frac{5}{100}$$

$$P(C_p | S_1 S_p) = \frac{25}{28}$$

$$P(S_p | S_1) = \frac{4}{29}$$

$$P(S_1 C_p C_w) = \frac{5}{100} \times \frac{4}{29} \times \frac{25}{28} = \frac{25}{1418}$$

$$P(S_1 C_p C_w) = \frac{P_p^5 \times P_1^{25}}{P_{w_0}^{100}} = \frac{5 \times 4 \times 25}{100 \times 29 \times 28} = \frac{25}{1418}$$

**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 بگیرید. 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  $\frac{1}{4}$  است، داریم

$$P(A) = \frac{1}{4} \quad P(AB) = \frac{1}{4}$$

$$P(B) = \frac{1}{4} \quad P(AC) = \frac{1}{4}$$

$$P(C) = \frac{1}{4} \quad P(BC) = \frac{1}{4}$$

با توجه به این احتمال‌ها داریم:

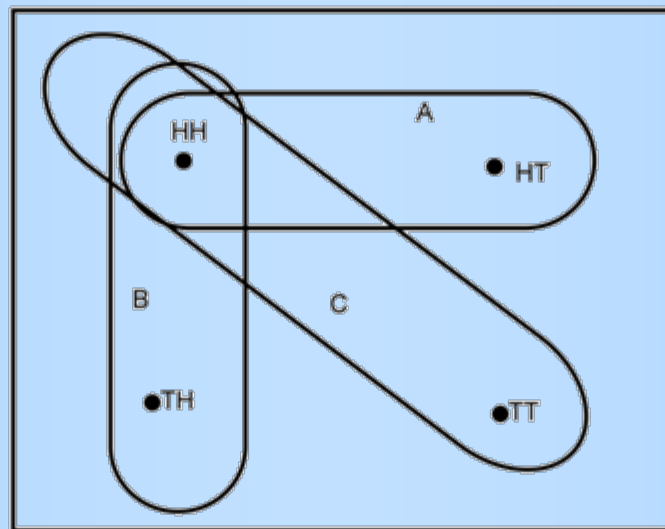
$$P(A)P(B) = \frac{1}{4} = P(AB)$$

$$P(A)P(C) = \frac{1}{4} = P(AC)$$

$$P(B)P(C) = \frac{1}{4} = P(BC)$$



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



نمودار ون برای دوبار پرتاب یک سکه

**مثال:** آیا دو پیشامد

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

از جدول زیر مستقل از هم اند؟

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

$$P(A) = ۰/۲$$

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

**دو پیشامد مستقل نیستند.**

**مثال: یک آزمون بالینی ناکامل:** یک بیماری فونی 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 آزمون بالینی 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. به علاوه، 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      S: حالت شدید

- : آزمون منفی      L: حالت خفیف

N: سالم

احتمال‌های موجود در مسئله:

$$P(S) = 0.05 \quad P(+|S) = 0.9$$

$$P(L) = 0.10 \quad P(+|L) = 0.4$$

$$P(N) = 0.85 \quad P(+|N) = 0.1$$

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

$$P(S+) = P(S)P(+ | S) = 0.05 \times 0.9 = 0.045$$

$$P(L+) = P(L)P(+ | L) = 0.10 \times 0.4 = 0.04$$

$$P(N+) = P(N)P(+ | N) = 0.85 \times 0.1 = 0.085$$

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

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

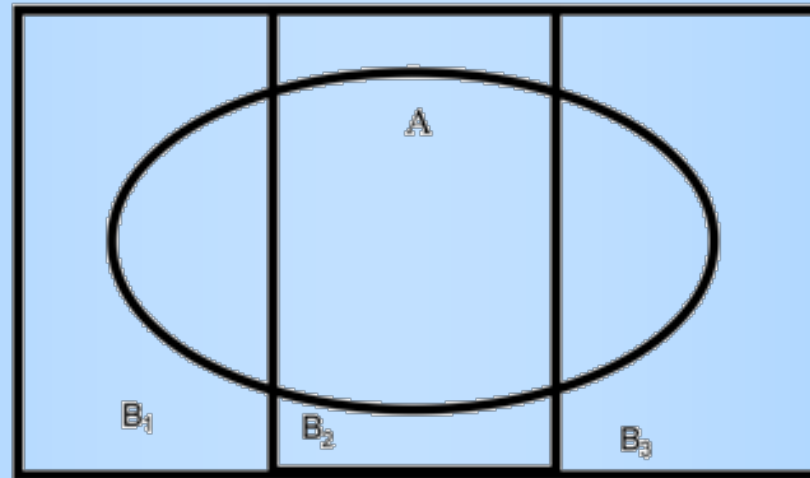
$$= 0.018 + 0.040 + 0.088$$

$$= 0.144$$

$$P(S | +) = \frac{P(+S)}{P(+)} = \frac{0.018}{0.144} = 0.11$$

## ساختار قضیه بیز

$$k=۳$$



$$P(B_i | A) = \frac{P(AB_i)}{P(A)}$$

$$P(AB_j) = P(B_j)P(A | B_j)$$

## قضيه بيز

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^k P(B_j)P(A | B_j)}$$



در ابتدا محقق از چند وضعیت طبیعی ممکن  $B_1, B_2, \dots, B_k$  با اطلاع 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  $P(B_1), P(B_2), \dots, P(B_k)$  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 به شرط وقوع هر وضعیت مشهود  $B_j$  تعیین کند. آنگاه قضیه بیز به محقق امکان می‌دهد که احتمال‌های شرطی  $P(B_j|A)$ ،  $(j=1, \dots, k)$  را محاسبه نماید، که این کار، چیزی نیست جز نوعی تجدید نظر در احتمال‌های وضعیت‌های طبیعی مختلف، بعد از بدست آمدن گواه آزمایشی. این احتمال‌های تجدید نظر شده را **احتمال‌های پسین** یا **پس از آزمایش** گویند.

## قواعد کلی احتمال

احتمال  $P$ ، تابعی است با مقادیر عددی که روی پیشامدهای موجود در یک فضای نمونه  $S$  تعریف می‌شود و در شرایط زیر صدق می‌کند:

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

دو)  $P(S) = 1$

سه) برای پیشامدهای جدا از هم  $A_1, A_2, \dots$

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

سه قانون مهم احتمال برای یک فضای نمونه در حالت کلی

$$P(A) = 1 - P(A^c)$$

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

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

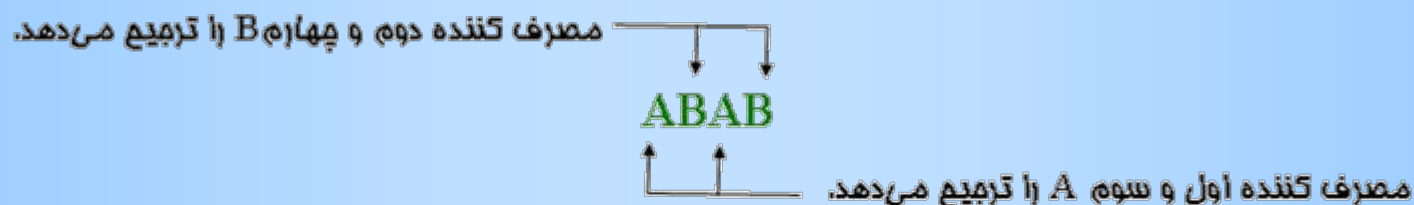
# فصل 4

# 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.

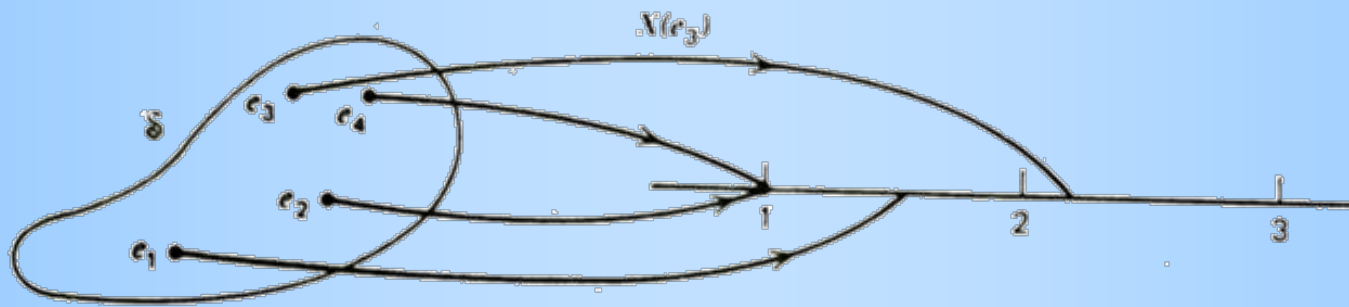
**مثال:** فرض کنید که دو محصول A و B به وسیله چهار مصرف کننده از  
لحاظ برتری یکی بر دیگری سنجیده می‌شوند.



تعداد برآمدها  $16 = 2^4 = 2 \times 2 \times 2 \times 2$  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	$\frac{۱}{۱۶}$	$\frac{۴}{۱۶}$	$\frac{۶}{۱۶}$	$\frac{۴}{۱۶}$	$\frac{۱}{۱۶}$

$$P[X \geq ۲] = \frac{۶}{۱۶} + \frac{۴}{۱۶} + \frac{۱}{۱۶} = \frac{۱۱}{۱۶}$$

$$P[۱ \leq X \leq ۳] = \frac{۴}{۱۶} + \frac{۶}{۱۶} + \frac{۴}{۱۶} = \frac{۱۴}{۱۶}$$

**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 $X$	$x_1$	$x_p$	...	$x_k$
احتمال	$f(x_1)$	$f(x_p)$	...	$f(x_k)$

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

**مثال:** نمونه‌گیری از دسته‌ای کالای سالم و فراب:

یک فروشگاه بزرگ ۱۵ رادیوی ساعتی باقیمانده خود را برای فروش به مزاج می‌گذارد. هیچکس نمی‌داند که ۵ تا از این رادیوها فراب 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 فراب 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.

$$f(0) = P[X=0] = \frac{\binom{10}{3}}{\binom{15}{3}} = \frac{120}{455} = \frac{24}{91}$$

$$f(1) = P[X=1] = \frac{\binom{10}{2} \binom{5}{1}}{\binom{15}{3}} = \frac{45 \times 5}{455} = \frac{45}{91}$$

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

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

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  $\frac{1}{2}$  میمونها طی انجام یک آزمایش واکنش مثبت (موفقیت آمیز) نشان دهند برابر با  $\frac{1}{2}$  است. انجام این آزمایش را تا مشاهده یک واکنش مثبت تکرار می‌کنیم، و متغیر تصادفی  $X$  را به عنوان انجام آزمایش تعریف می‌کنیم. 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:

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

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 $f(x)$	$\frac{1}{\mu}$	$\left(\frac{\mu}{\mu}\right) \frac{1}{\mu}$	$\left(\frac{\mu}{\mu}\right)^2 \frac{1}{\mu}$	$\left(\frac{\mu}{\mu}\right)^3 \frac{1}{\mu}$	...

$$f(x) = P[X = x] = \left(\frac{\mu}{\mu}\right)^{x-1} \times \frac{1}{\mu} \quad x=1, 2, \dots$$

$$P[X \geq 3] = 1 - \frac{1}{\mu} - \frac{\mu}{\mu} \times \frac{1}{\mu} = \frac{\mu}{\mu}$$



## نمایش توزیع احتمال به صورت نمودار

- نمودار خطی توزیع احتمال

- بافت نگار احتمال توزیع

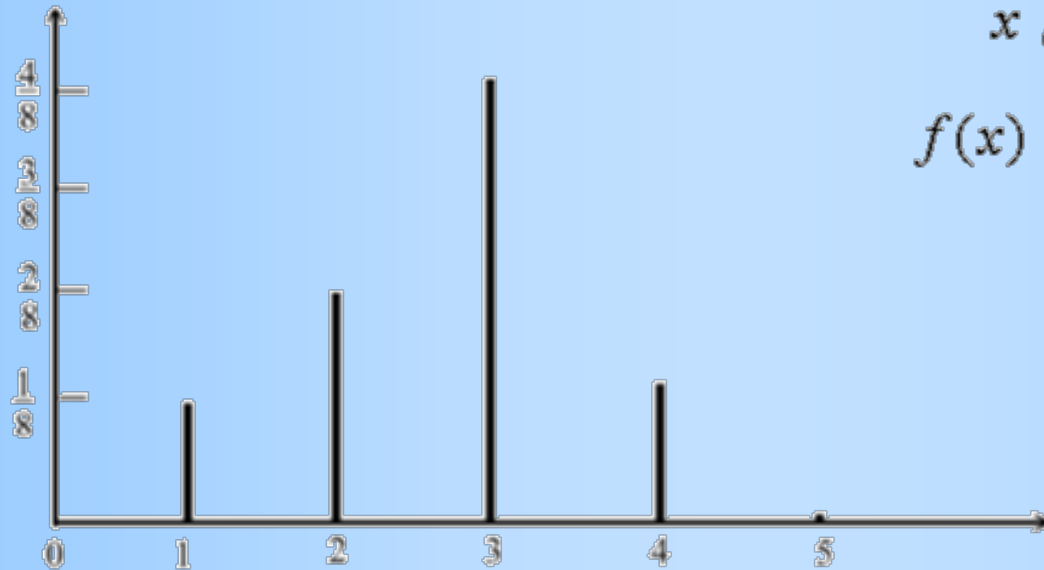
## جدول توزیع احتمال

مقدار $x$	۱	۲	۳	۴
احتمال $f(x)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$

## نمودار فطی توزیع احتمال

محور افقی متغیر تصادفی  $x$

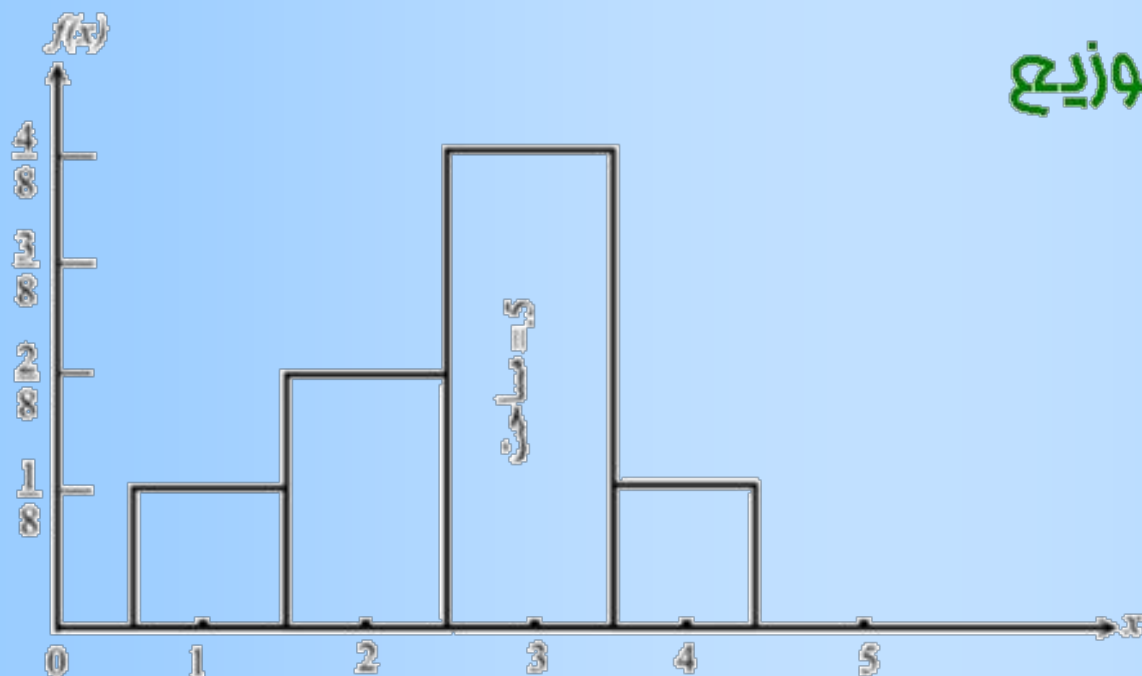
محور عمودی برابر احتمال  $f(x)$



## جدول توزیع احتمال

مقدار $x$	۱	۲	۳	۴
احتمال $f(x)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$

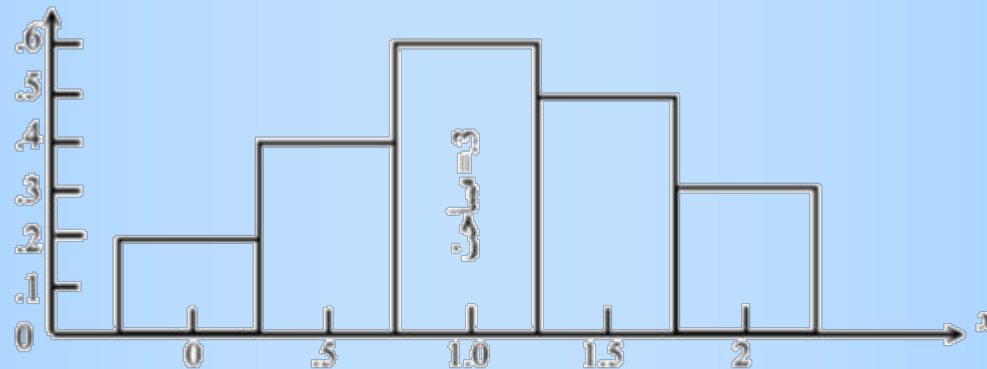
## بافت نگار احتمال توزیع



$$\text{مساحت} = \text{عرض} \times \text{ارتفاع} = 0.5 \times 0.4 = 0.2$$

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



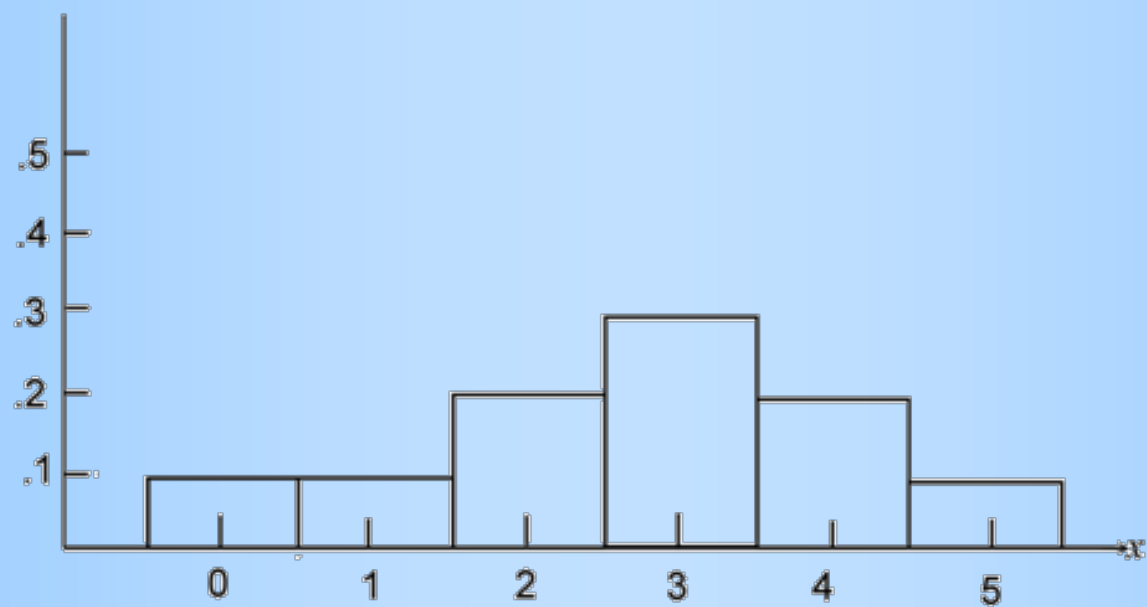
بافت نگار احتمال توزیع داده های جدول

توصیه می شود که بافت نگار احتمال برای توزیع هایی به کار رود که در آنها، فواصل بین  $x$  ها برابر باشند. در حالتی که فواصل بین  $x$  ها نابرابرند باید از نمودار فطی 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.

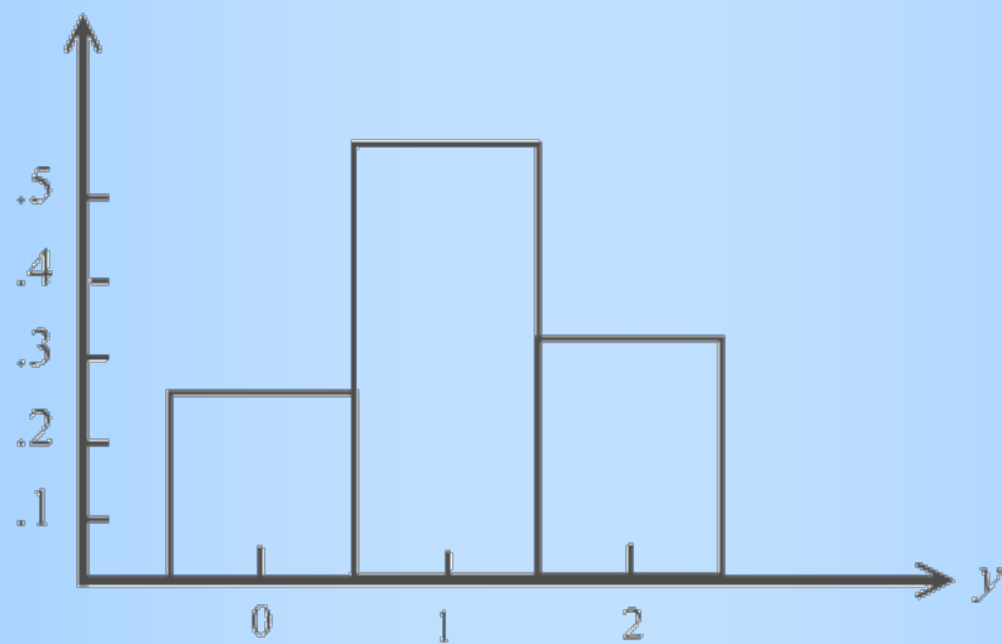
نوع A	مقدار $x$	۰	۱	۲	۳	۴	۵
	احتمال $f(x)$	۰/۱	۰/۱	۰/۲	۰/۳	۰/۲	۰/۱
نوع B	مقدار $y$	۰	۱	۲			
	احتمال $f(y)$	۰/۲۳	۰/۴۸	۰/۲۹			

نوع A	مقدار $x$	٠	١	٢	٣	٤	٥
	احتمال $f(x)$	٠/١	٠/١	٠/٢	٠/٣	٠/٢	٠/١



نوع A

نوع B	مقدار $y$	0	1	2
	احتمال $f(y)$	0.23	0.48	0.29



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 :

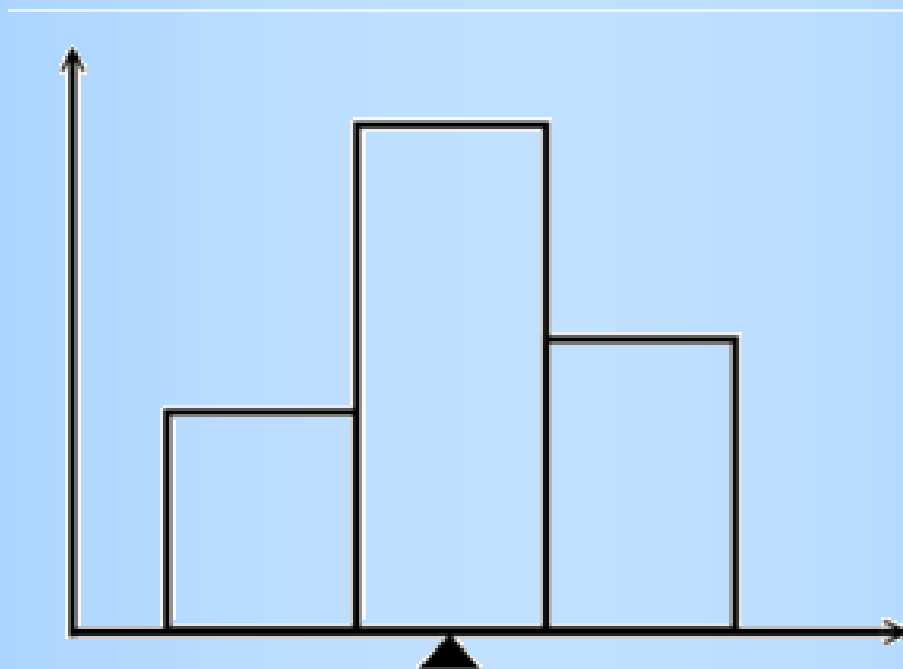
$$E(X) = \sum x_i f(x_i)$$

تمام مقادیر متمایز  $X$



میانگین جامعه برای متغیر تصادفی  $X$  با  $\mu$  یا  $\mu_x$  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	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

بازی عادلانه است اگر  $c = \frac{35}{6}$  باشد زیرا در این صورت متوسط برد در صورت تکرار بازی به دفعات زیاد، برابر صفر است. اگر بازیکن برای شرکت در هر بازی مبلغ ۴ تومان بپردازد، امید ریاضی باخت خالص بازیکن  $0/5$  تومان در هر بازی است.

فرض کنید به دلایلی متغیر تصادفی بصورت  $g(X)=(X-2)^2$  باشد که  
تابعی از  $X$  است. به دو روش امید ریاضی را می توان محاسبه کرد:

**روش اول:** توزیع احتمال  $(X-2)^2$  را با استفاده از توزیع احتمال  $X$  محاسبه شود و  
سپس نتایج را در دو سطر نوشته و به روش حاصلضرب های اجزای متناظر  
آن را محاسبه کرد:

مجموع	۴	۱	۰	مقادیر متمایز $(X-2)^2$
۱	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{3}{8}$	احتمال
۱	$\frac{4}{8}$	$\frac{4}{8}$	۰	احتمال $\times$ مقدار

$$E[(X-2)^2]=1$$

**روش دوم:** مستقیماً از توزیع احتمال  $X$  استفاده شود. مقادیر توزیع احتمال را در دو سطر مناسبه 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	
$(x-۲)^۲ f(x)$	$\frac{۴}{۱۶}$	$\frac{۴}{۱۶}$	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.**

$$E(X) = \sum g(x_i) f(x_i)$$

**تمام مقادیر متمایز  $x$**

## خواص امید ریاضی

$$E(a) = a$$

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

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

$$E(a + bX + cX^p) = a + bE(X) + cE(X^p)$$

---

**مثال:** فرض کنید  $X$  تعداد وامدهای 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 وامد فروخته شده ۵۰ تومان سود به دست آمده و هزینه ثابت هفتگی فروشگاه ۲۰ تومان است. امید ریاضی سود چقدر است؟

**حل:** سود مال از فروش  $X$  واحد در یک هفته،  $50X$  تومان است.

با کم کردن هزینه ثابت، سود  $50X - 20$  است.

$$E(50X - 20) = 50E(X) - 20$$

$$= 50 \times 2/7 - 20$$

$$= 115 \text{ تومان}$$



## واریانس: معیاری برای پراکندگی

واریانس  $X$  عبارت است از :

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

واریانس  $X$  با  $\sigma^p$  یا  $\sigma_x^p$  نیز نشان داده می شود.

**انحراف معیار  $X$  عبارت است از:**

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

**انحراف معیار  $X$  با  $\sigma$  یا  $\sigma_x$  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	$۲/۷=E(X)$
$x^2f(x)$	۰	۰/۱	۰/۸	$۲/۷$	$۳/۲$	$۲/۵$	$۹/۳۵=E(X^2)$

$$\text{Var}(X) = ۹/۳۵ - (۲/۷)^2 = ۲/۰۱$$

$$\text{sd}(X) = \sqrt{۲/۰۱} = ۱/۴۲$$

# خواص

## واریانس

**Var(X) نمی تواند منفی باشد**

$$Var(X + a) = Var(X)$$

$$Var(bX) = b^p Var(X)$$

$$Var(a + bX) = b^p Var(X)$$

## انحراف معیار

**Sd(X) نمی تواند منفی باشد**

$$sd(X + a) = sd(X)$$

$$sd(bX) = |b|sd(X)$$

$$sd(a + bX) = |b|sd(X)$$

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

$$E(Z) = 0 \text{ و } Var(Z) = 1$$

**مثال:** با توجه به مثال قبل داشتیم  $X$  برابر با تعداد وامدهای کالای نوع A فروخته شده در هفته بود و برای هر وام 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  $Y=۵۰X-۲۰$  می باشد. پس

$$E(Y)=۵۰E(X)-۲۰=۱۱۵ \text{ تومان}$$

$$\text{Var}(Y)=(۵۰)^2\text{Var}(X)=(۵۰)^2(۲/۰۱)=۵۰۲۵$$

$$\text{sd}(Y)=|۵۰|\text{sd}(X)=(۵۰)(۱/۴۲)=۷۱ \text{ تومان}$$

## توزیع توام دو متغیر تصادفی

		$y_1$	$y_p$	...	$y_l$
مقادیر $X$	$x_1$	$f(x_1, y_1)$	$f(x_1, y_p)$	...	$f(x_1, y_l)$
	$x_p$	$f(x_p, y_1)$	$f(x_p, y_p)$	...	$f(x_p, y_l)$
	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
	$x_k$	$f(x_k, y_1)$	$f(x_k, y_p)$	...	$f(x_k, y_l)$

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




**مثال:** ده ماشین وجود دارند که پنج عدد از آنها در وضعیت خوب ( $G$ ) هستند، دو عدد دارای نقص در قسمت انتقال قدرت ( $DT$ )، و سه عدد دیگر دارای نقص در مکانیزم فرمان ( $DS$ ) می‌باشند. فرض کنید که دو ماشین به‌طور تصادفی برگزیده شده‌اند، و متغیرهای تصادفی مورد نظر عبارت اند از  $X$  (تعداد ماشین‌هایی که در قسمت انتقال قدرت نقص دارند) و  $Y$  (تعداد ماشین‌هایی که در مکانیزم فرمان نقص دارند). توزیع احتمال توأم  $X$  و  $Y$  را به دست آوریم.

## نمای نمونه گیری

$\mu$	DT
$\mu$	DS
$\omega$	G

نمونه‌ی به حجم  $\mu$



$X$  = تعداد DT در نمونه

$Y$  = تعداد DS در نمونه

## توزيع توام

		مقادير Y		
		٠	١	٢
مقادير X	٠	$\frac{10}{145}$	$\frac{15}{145}$	$\frac{3}{145}$
	١	$\frac{10}{145}$	$\frac{4}{145}$	٠
	٢	$\frac{1}{145}$	٠	٠

برای مثال

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

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

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

**مثال:** کارخانه کوپکی 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 مضور و غیاب کارگران در گذشته، مقادیر توزیع احتمال توأم  $X$  و  $Y$  را تعیین می کند که نتیجه آن به صورت مقابل است:

## توزیع توام غایبین روزانه در دو نوبت کاری

$x \backslash y$	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  $Y$ :

$$P[X=۲]=۰/۱۵+۰/۱۰+۰/۰۵=۰/۳۰$$

$$P[X>Y]=۰/۰۵+۰+۰/۱۵=۰/۲۰$$

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  $\mu_x$  و واریانس  $\sigma_x^2$  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:

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

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

$$\mu_X = 1/1$$

$$\sigma_X^2 = 1/7 - (1/1)^2 = 0/49$$

$$\sigma_X = 0/70$$

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

$$\mu_Y = 1/45$$

$$\sigma_Y^2 = 13/45 - (1/45)^2 = 0/7275$$

$$\sigma_Y = 0/85$$

## کوواریانس و همبستگی

**کوواریانس بین  $X$  و  $Y$  عبارت است از:**

$$\begin{aligned}Cov(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E(XY) - \mu_X \mu_Y\end{aligned}$$

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

**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

$x \backslash y$	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)

$$\sigma_{10 \times 1} = \sigma_{10}$$

$$\sigma_{p\Delta \times p} = \sigma_{p\Delta}$$

$$\sigma_{10 \times w} = \sigma_{w0}$$

$$\sigma_{1\Delta \times p} = \sigma_{pw}$$

$$\sigma_{10 \times c} = \sigma_{c0}$$

$$\sigma_{0\Delta \times c} = \sigma_{w0}$$

$$E(XY) = 1/90$$

$$Cov(X, Y) = 1/90 - 1/4\Delta \times 1/1 - 0/0 \wedge \Delta$$

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{0/0 \wedge \Delta}{0/70 \times 0/1 \wedge \Delta} = 0/1c$$

## خواص همبستگی

الف)  $Corr(X, Y)$  همیشه عددی بین  $-1$  و  $+1$  است. فرین  $-1$  و  $+1$  زمانی است که  $X$  و  $Y$  به وسیله خط مستقیمی بترتیب با شیب منفی و یا مثبت به هم مربوط باشند.

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

مثال:

$$U = 5X + 2$$

$$V = pY + m$$

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

## استقلال دو متغیر تصادفی

متغیرهای تصادفی  $X$  و  $Y$  مستقل اند 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 توأم  $X$  و  $Y$  بصورت مقابل است. آیا  $X$  و  $Y$  مستقل اند؟

x \ y	مجموع سطر		
	۰	۱	
۰	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
۱	$\frac{2}{8}$	$\frac{2}{8}$	$\frac{1}{2}$
۲	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
مجموع ستون	$\frac{1}{8}$	$\frac{1}{8}$	۱

X و Y مستقل نیستند چون  $f(0, 0) = 10/45$  است در حالی که

$$f_X(0)f_Y(0) = (28/45) \times (21/45) \neq (10/45)$$



**اگر  $X$  و  $Y$  مستقل باشند داریم:**

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

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

## جمع پذیری امید ریاضی

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

## فرمول‌های کلی برای واریانس مجموع و تفاضل:

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

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

**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 بر حسب پارامترهای نامعلومی بیان می‌شوند که به مشخصه‌های جامعه و روش نمونه‌گیری مربوط اند.

## 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 زائد (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 می انجامد که در اصطلاح فنی موفقیت ( $S$ ) و شکست ( $F$ ) نامیده می شود.

ب) برای تمام امتحان ها، احتمال موفقیت  $P(S)$ ، یکی است و به صورت  $P=P(S)$  نشان داده می شود. بنابراین احتمال شکست برای هر امتحان  $P(F)=1-p$  است که آن را با  $q$  نشان می دهیم، بطوری که  $p+q=1$ .

ج) امتحان ها مستقل از یکدیگرند. احتمال موفقیت در یک امتحان با داشتن هر مقدار اطلاعات از برآمدهای سایر امتحان ها تغییر نمی کند.

ساده ترین مثال برای امتحان های برنولی پرتاب سکه سالم و ناریب است. شیر و خط را می توان بترتیب با  $S$  و  $F$  نامید:  $p=q=1/2$



## نمونه‌گیری از جامعه دوحالتی

الف) نمونه‌گیری با جایگذاری

جامعه شامل ۱۵ کالا

۵ کالا معیوب و ۱۰ کالا سالم

$S$  معیوب بودن کالا

$$P(S)=p=\frac{5}{15}$$

## نمونه‌گیری از جامعه دوحالتی

(ب) نمونه‌گیری بدون جایگذاری

۳ کالا یکی بعد از دیگری بدون جایگذاری

$$P(S) = \frac{5}{15} \text{ برای اولین کالا}$$

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

احتمال شرطی دومین استفرآج  $P(S) = \frac{5}{15}$  است که شرط استقلال نقض

$$\frac{5}{15} \neq \frac{4}{14} \text{ شده چون}$$

## نمونه‌گیری از جامعه دو حالتی

چ) نمونه‌گیری بدون جایگذاری از جامعه بزرگ  
مجموعه شامل ۱۵۰۰ کالا که ۵۰۰ کالای آن معیوب است  
نمونه‌گیری بدون جایگذاری ۳ کالا

$$P(S_1) = \frac{500}{1500} = \frac{5}{15}$$

$$P(S_2 | S_1) = \frac{1499}{1499}$$

اگر از جامعه دو حالتی نمونه‌ای تصادفی با روش جایگذاری بگیریم، شرایط امتحانهای برنولی برقرار خواهد بود. اگر نمونه‌گیری بدون جایگذاری انجام گیرد، شرط استقلال آزمایش‌ها نقض می‌شود. لکن، اگر حجم جامعه در مقایسه با حجم نمونه خیلی بزرگ باشد تاثیر این نقض استقلال، قابل صرف نظر کردن است و مدل امتحانهای برنولی را می‌توان به عنوان یک تقریب در نظر گرفت.

## توزیع دوجمله‌ای

توزیع احتمال  $X$  با  $n$  امتحان و احتمال موفقیت  $p$

مثال:  $n=4$

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	$q^4$	$pq^3$	$q^2p^2$	$q^3p$	$q^4$
تعداد دنباله	$1 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$	$4 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$	$6 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$	$4 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$	$1 = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$

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

X	0	1	2	3	4
$P[X = x]$	$\binom{n}{0} p^0 q^n$	$\binom{n}{1} p^1 q^{n-1}$	$\binom{n}{2} p^2 q^{n-2}$	$\binom{n}{3} p^3 q^{n-3}$	$\binom{n}{4} p^4 q^{n-4}$

$$P[X = x] = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, \dots, n$$

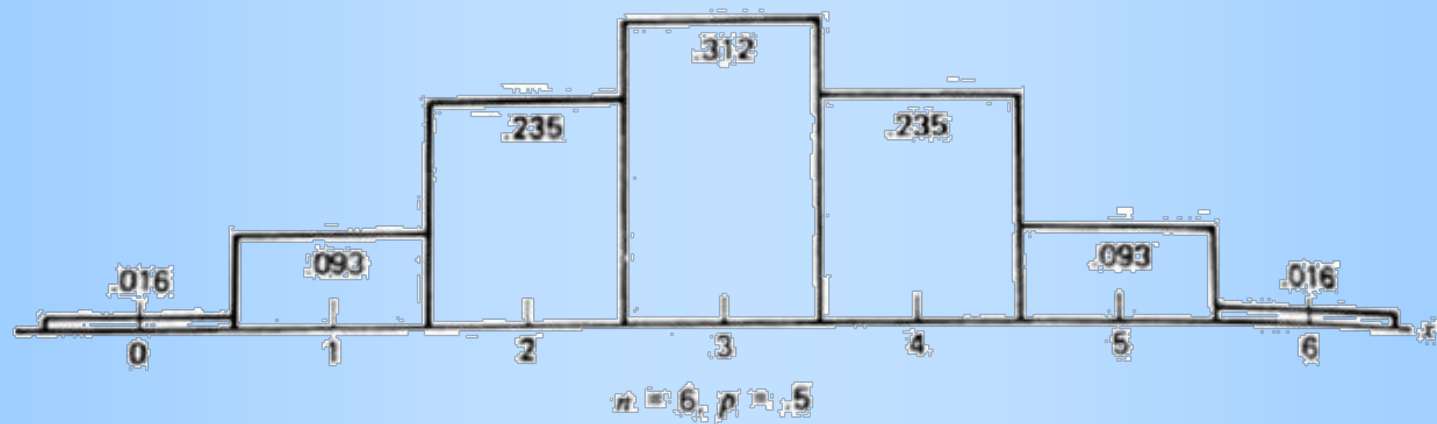
## توزیع دوجمله‌ای

$p=P(S)$  امتحان  $n$

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

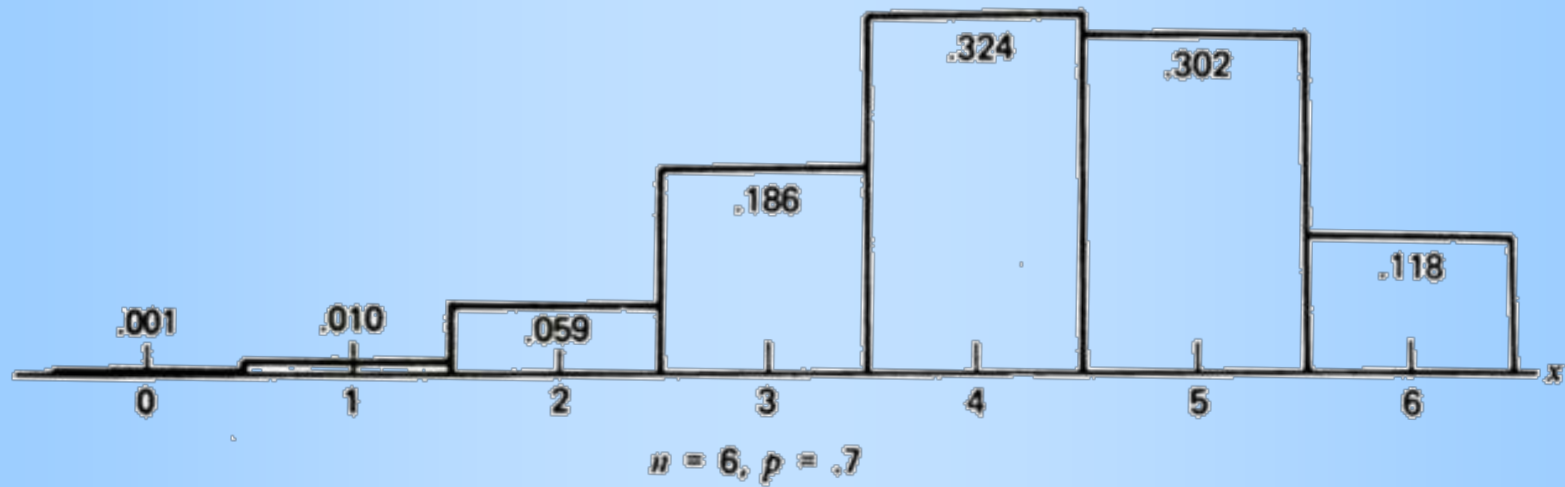
برای مقادیر ممکن  $x=0, 1, \dots, n$

## تأثیر مقادیر $p$ در شکل توزیع

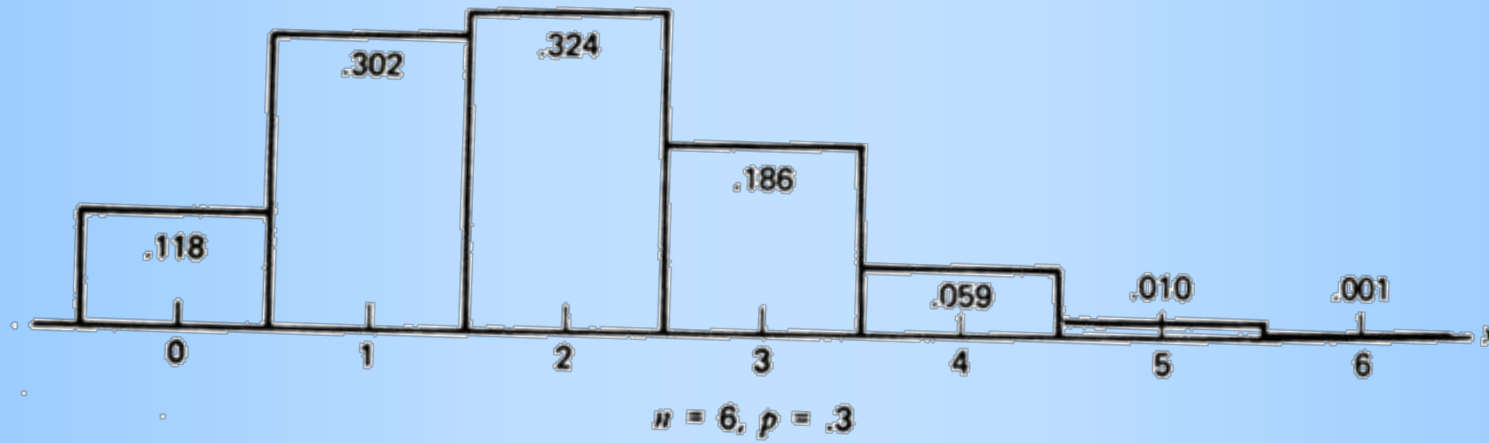




## تأثیر مقادیر $p$ در شکل توزیع



## تأثیر 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 قرمز

تولید می‌شود. فرض کنید باغبان می‌خواهد ۵ جفت از این گل‌های

قرمز و سفید را با یکدیگر بیامیزد احتمال اینکه از ۵ گل حاصل

**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.  $P(S)=p= \frac{1}{4}$ .

$X$  دارای توزیع دو جمله‌ای با  $n=5$  و  $p= \frac{1}{4}$  است.

$$P[X=0]=b(0)=\left(\frac{3}{4}\right)^5 = \frac{243}{1024} = 0.237$$

$$P[X \geq 4] = b(4) + b(5)$$

$$= \binom{5}{4} \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^1 + \binom{5}{5} \left(\frac{1}{4}\right)^5 \left(\frac{3}{4}\right)^0$$

$$= \frac{16}{1024} = 0.016$$

## 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 ۸۰ درصد 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 دو جمله‌ای  $n=15$  و  $p=0/8$  برای متغیر  $X$  مناسب است.

$$P[X \leq 8] = 0/018 \text{ (الف)}$$

$$P[X \geq 10] = 1 - P[X \leq 9] = 0/018 \text{ (ب)}$$

$$P[8 \leq X \leq 12] = P[X \leq 12] - P[X \leq 7] = 0/402 - 0/014 = 0/598 \text{ (ج)}$$

توجه:

$$b(x; n, p) = b(n - x; n, p^*)$$
$$p^* = 1 - p$$



## میانگین و واریانس توزیع دوجمله‌ای

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

$$np = \text{میانگین}$$

$$np(1-p) = npq = \text{واریانس}$$

$$sd = \sqrt{np(1-p)} = \sqrt{npq}$$

# توزیع فوق هندسی

## شمای نمونه گیری بدون جایگذاری



## توزیع فوق هندسی

D=تعداد معیوب ها در جامعه

D=تعداد معیوب ها در جامعه

n=حجم نمونه

X=تعداد معیوب ها در نمونه

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

وقتی که  $n \leq D$  و  $n \leq N - D$

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

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

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

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

توزیع هندسی را گاهی توزیع زمان انتظار گسسته می‌گویند.

**پیشامدهای کمیاب انتظار می رود هر یک از این پیشامدها در  
تعداد زیادی از امتحان های بالقوه، فقط چندبار رخ دهد.**

## تعریف رسمی توزیع پواسن

پیشامد  $S$  با پیروی از فرض‌های زیر در زمان رخ می‌دهد:

(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

$$P[X = x] = \frac{e^{-m} m^x}{x!}, \quad x = 0, 1, 2, \dots$$

**e عدد نمایی و مقدارش برابر ۲/۷۱۸۲۸**

## تقریب پواسن برای دو جمله‌ای

	پواسن (m)	دو جمله‌ای (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  $3 \times 2 = 6$  سیکلوپ وجود دارد.

$$P[X \geq 5] = 1 - P[X \leq 4] = 1 - 0.285 = 0.715$$

# فصل ٦

# 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 داده‌ها قویاً تایید می‌گردد.

$H'$  را رد نکند و نتیجه بگیرد که داده‌ها  $H$  را قویاً تایید نمی‌کنند.

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.

یک فرض آماری  
نتایج تحقیق  
نتیجه نوعی

گزاره ریاضی  
نتایج تحقیق  
نتیجه

## فرض صفر و فرض مقابل

### مسئله:

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

$H_0: p \leq 0.05$  (داروی جدید بهتر نیست)

$H_1: p > 0.05$  (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 می‌گردد. متغیری تصادفی که مقادیر آن برای انجام این عمل به کار می‌روند آماره آزمون نامیده می‌شود 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 می‌شود.

## دو نوع خطا و تابع توان آزمون

نتایج آزمون	$H_0$ درست ( $p \leq 0.05$ )	$H_0$ غلط ( $p > 0.05$ )
$H_0$ رد نمی‌شود	صحيح	غلط (خطای نوع II)
$H_0$ رد می‌شود	غلط (خطای نوع I)	صحيح

**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

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

$$\beta = 1 - \alpha = P[\text{II نوع خطای}] = P[\text{وقتی } H_0 \text{ رد نکردن } H_1 \text{ درست است}]$$

**احتمال  $\alpha$  بستگی به مقدار مشخص پارامتر در دامنه ای دارد که  $H_0$  آن را در بر می گیرد و حال آنکه  $\beta$  بستگی به مقدار پارامتر در دامنه ای دارد که  $H_1$  آن را در بر می گیرد.**

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

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

منحنی تابع احتمال رد کردن ( $\gamma$ ) برای یک آزمون، نشانگر نحوه انجام آزمون از طریق نمایش احتمال خطا به ازای تمام مقادیر ممکن پارامتر می‌باشد. این منحنی را **منحنی توان آزمون** می‌نامند.

عرض منحنی در ناحیه‌ای که  $H_0$  صحیح است، احتمال خطای نوع I را می‌دهد؛ در حالی که، در ناحیه دیگر، با کم کردن مقدار عرض منحنی از عدد یک، احتمال خطای نوع II به دست می‌آید.

## گزینش یک آزمون از میان چند آزمون

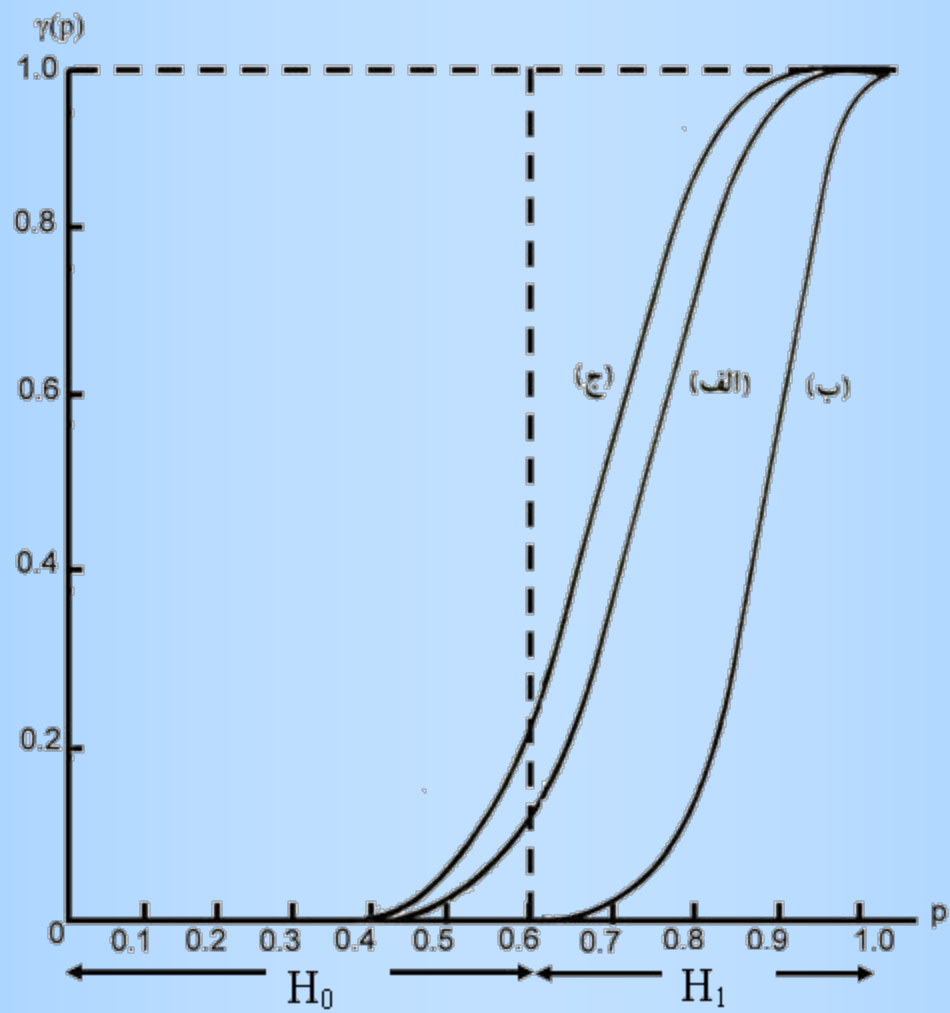
آزمون الف: ناحیه رد  $X \geq 15$

آزمون ب: ناحیه رد  $X \geq 18$

آزمون ج: ناحیه رد  $X \geq 14$

احتمال‌های رد برای آزمون‌های الف، ب و ج

P	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
$\gamma(p)=P[X \geq 15]$	۰/۰۰۰	۰/۰۰۲	۰/۰۲۱	۰/۱۲۶	۰/۶۱۴	۰/۸۰۴	۰/۹۸۹
$\gamma(p)=P[X \geq 18]$	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۴	۰/۰۳۵	۰/۲۰۶	۰/۶۷۷
$\gamma(p)=P[X \geq 14]$	۰/۰۰۰	۰/۰۰۶	۰/۰۵۸	۰/۲۵۰	۰/۶۰۸	۰/۹۱۳	۰/۹۹۸





## استخراج نتایج از یک آزمون

نتیجه آزمون به طریق زیر بیان شود:

در سطح معنی دار بودن  $\alpha$

$H_0$  رد می شود

یا اینکه

$H_0$  رد نمی شود

علاوه بر اجرای آزمون فرض‌ها با مقدار از قبل تعیین شده  $\alpha$ ، 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  $\alpha$ ، 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 می‌کنیم و هر مدعا را بر حسب دامنه‌ای از مقادیر پارامتر(های) مربوطه  $\theta$ ی 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 می کنیم.  $H_0$  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 آماره 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، برحسب 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. معمولاً برای این منظور کافی است که مطمئن شویم که انتخاب حجم نمونه و  $\alpha$  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 می‌کنیم 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

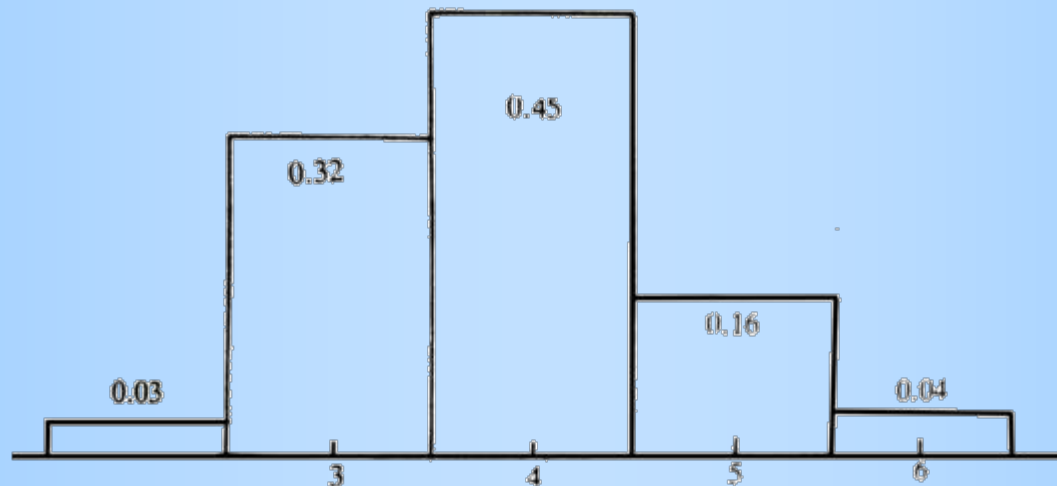
مدلهای احتمال برای متغیرهای پیوسته

توزیع احتمال یک متغیر تصادفی پیوسته را می توان به صورت شکل  
همواری 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 نقطه  $a$  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.**



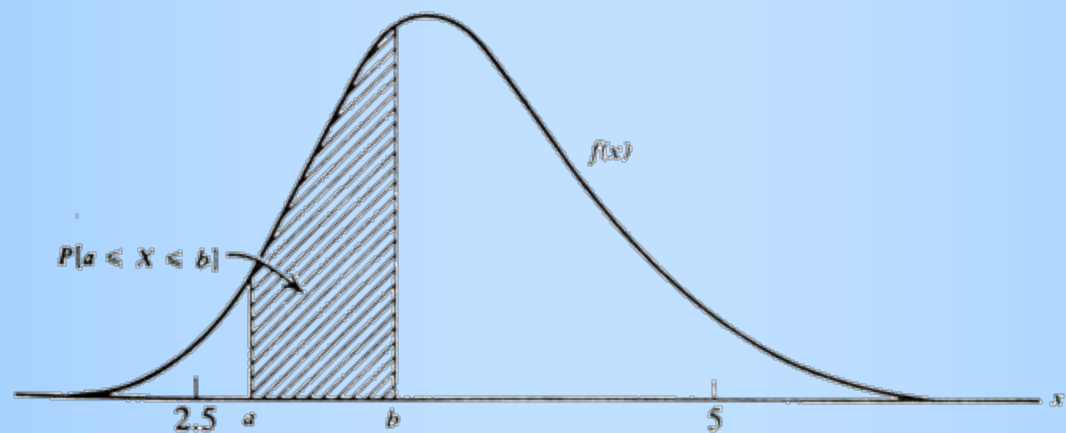
بافت نگار فراوانی نسبی وزن ۱۰۰ نوزاد در

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  $P[a \leq X \leq b]$ .

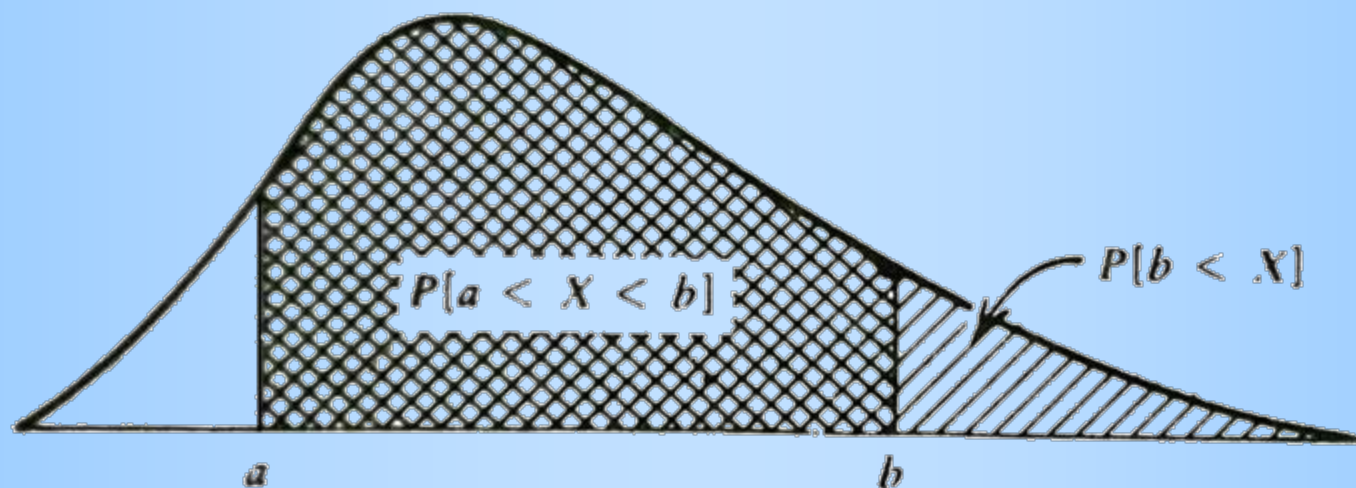
ج)  $f(x)$  مثبت یا صفر است.

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. برای به دست آوردن احتمال فواصل دیگر، باید قواعد زیر را بکار برد:

$$P[a \leq X \leq b] = (\text{مساحت سمت چپ } b) - (\text{مساحت سمت چپ } a)$$

$$P[b < X] = 1 - (\text{مساحت سمت چپ } b)$$



## مشخص کردن مدل احتمال

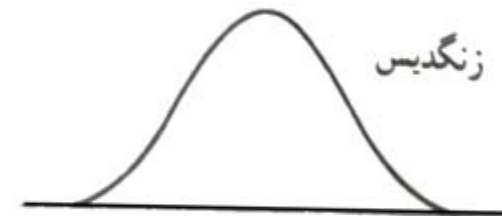
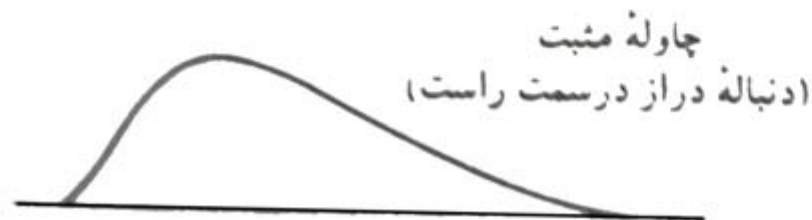
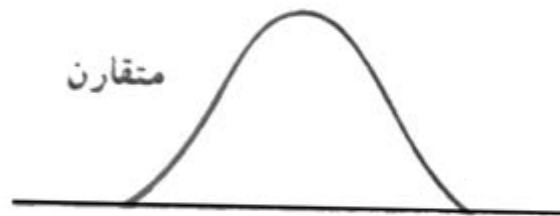
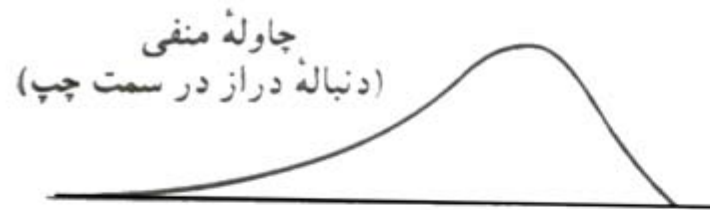
مدل احتمال متغیر تصادفی پیوسته به وسیله شکل ریاضی تابع چگالی احتمال مشخص می شود.

وقتی که تعداد نسبتاً زیادی از مشاهدات مربوط به یک متغیر تصادفی در دسترس باشند، سعی می کنیم 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 داده‌ها در دسترس نباشد، ممکن است به  
طور آزمایشی مدل قابل قبولی را بپذیریم که داده‌های خاص از  
یک منبع مشابه، پیش پای ما می‌گذارند.

## 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 امید ریاضی  $E(X)$  و نیز واریانس و انحراف معیار باشد.

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

میانۀ معیار دیگر به مرکز مقداری از  $X$  است که مساحت زیر منحنی را به دو قسمت برابر تقسیم می‌کند.

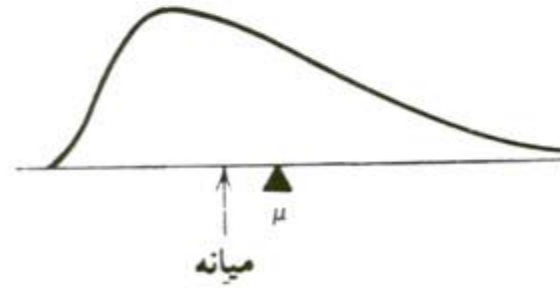
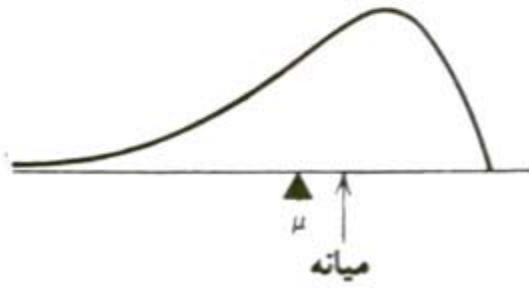
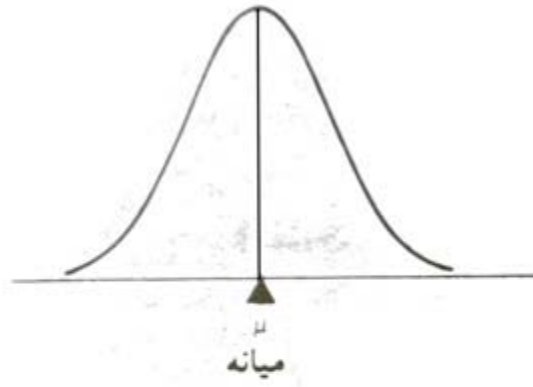
صدک  $(100p)$  ام جامعه ، یک مقدار  $X$  است که مقدار مساحت در سمت چپ آن  $p$  و مقدار مساحت در سمت راست آن  $1-p$  است.

چارک اول (پایینی) عبارت است از صدک ۲۵ام

چارک دوم (میانه) عبارت است از صدک ۵۰ام

چارک سوم (بالایی) عبارت است از صدک ۷۵ام





## چند خاصیت مهم

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

$$E(a_1X_1 + \dots + a_nX_n) = a_1E(X_1) + \dots + a_nE(X_n)$$

اگر  $X_1, \dots, X_n$  مستقل باشند،

$$Var(a_1X_1 + \dots + a_nX_n) = a_1^2Var(X_1) + \dots + a_n^2Var(X_n)$$

## توزیع نرمال

توزیع نرمال دارای چگالی زنگدیس زیر است:

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

که در آن ،  $\mu$  میانگین و  $\sigma$  انحراف معیار است.

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

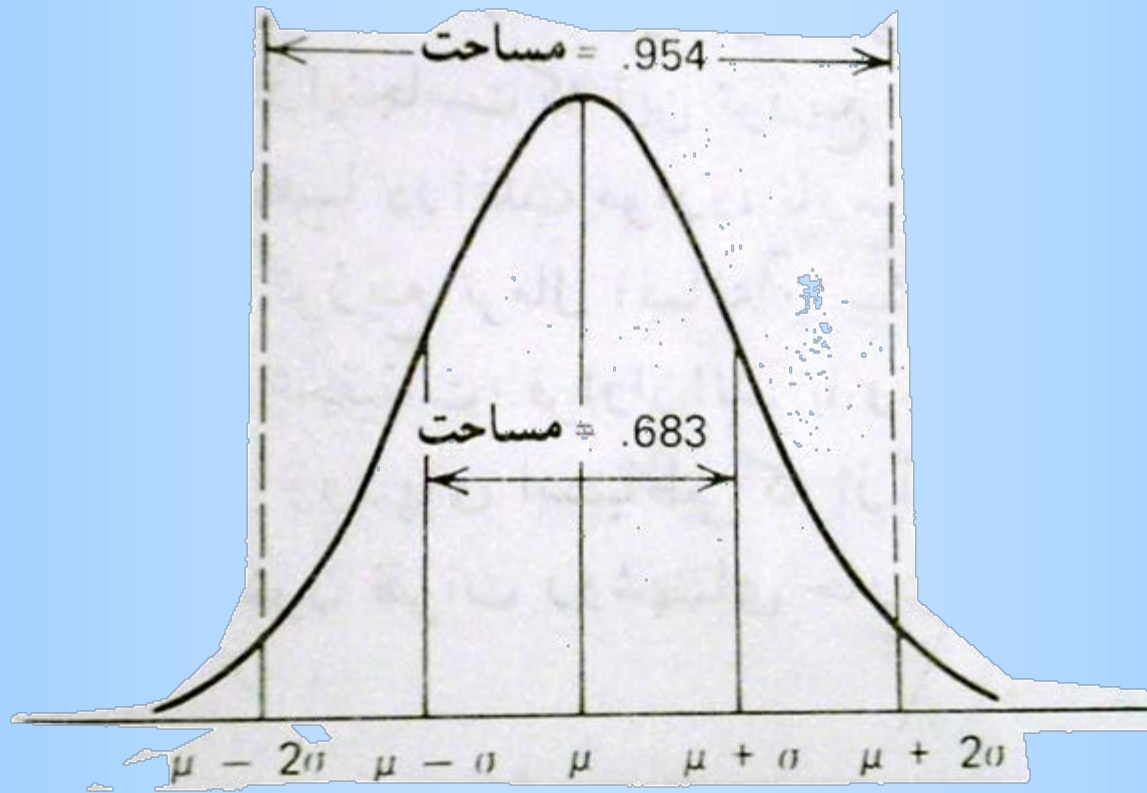
$$P[\mu - \sigma < X < \mu + \sigma] = 0.683$$

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

$$P[\mu - 2\sigma < X < \mu + 2\sigma] = 0.954$$

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

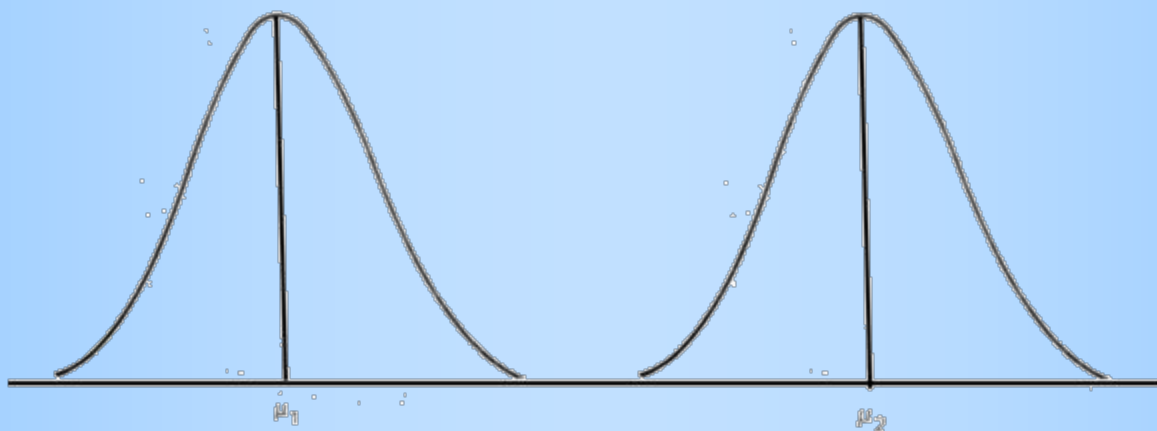
$$P[\mu - 3\sigma < X < \mu + 3\sigma] = 0.997$$



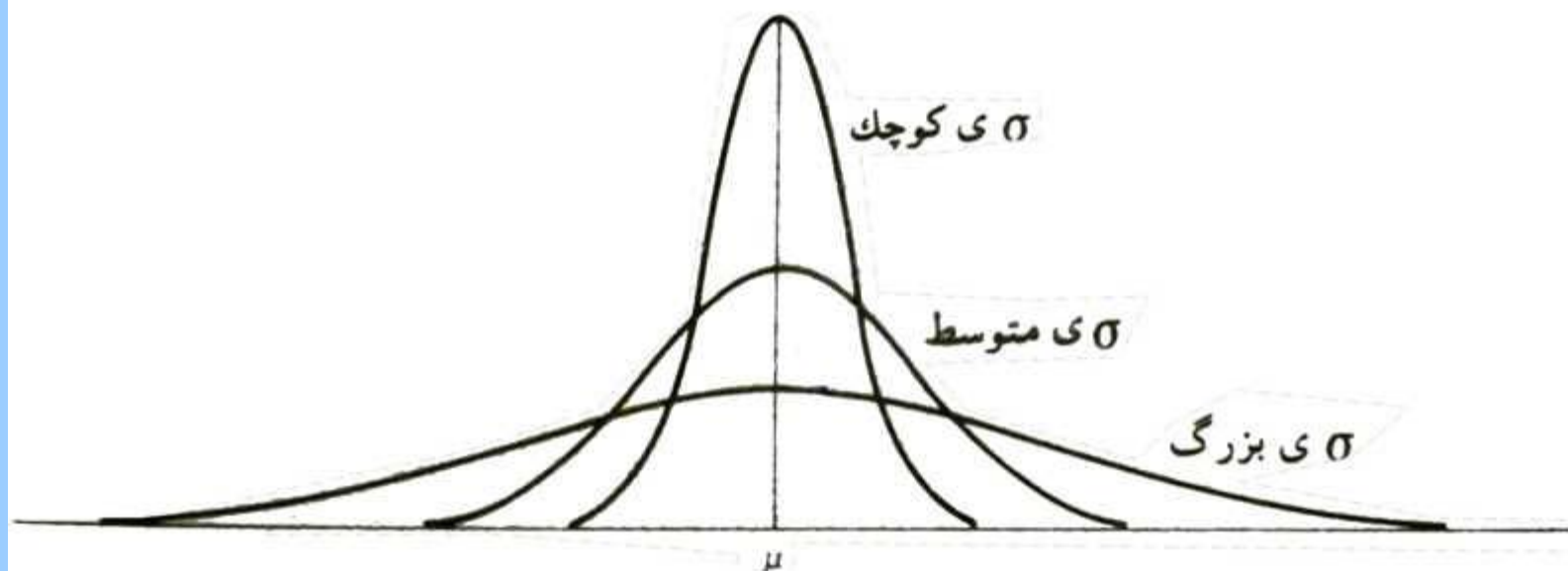
## نمادگذاری

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

## دو توزیع نرمال با میانگین‌های مختلف و با انحراف‌های معیارهای برابر



کاهش مقدار  $\sigma$ ، عرض نقطه بیشینه و تمرکز احتمال در اطراف  $\mu$  را افزایش می دهد.



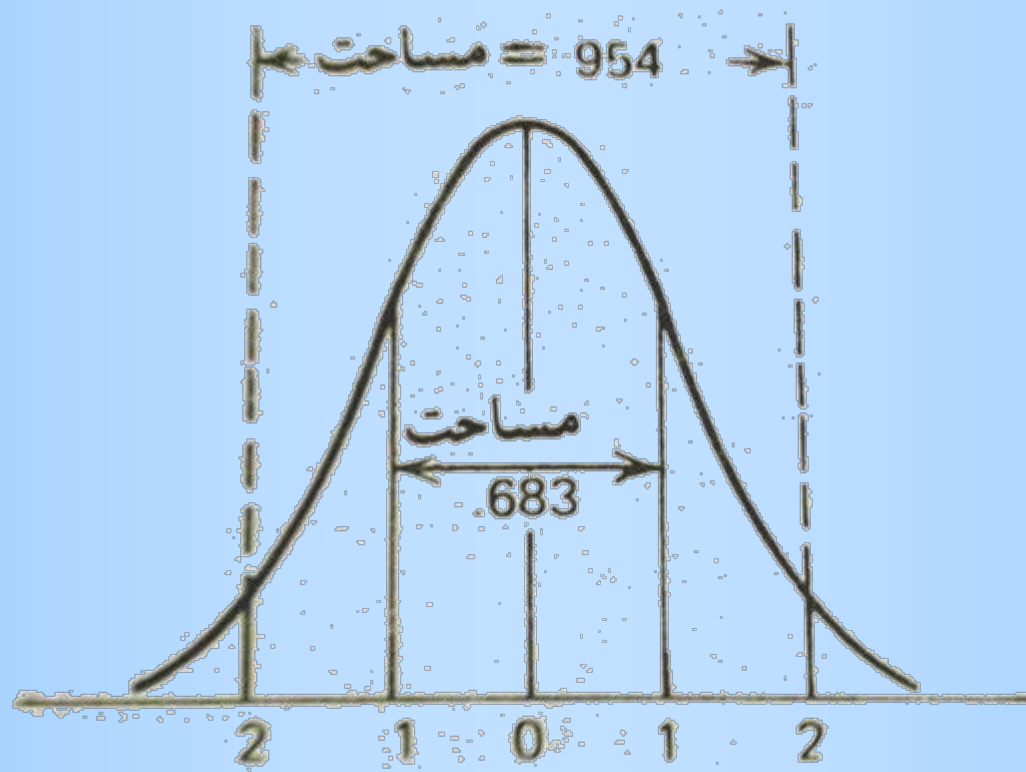


**توزیع نرمال استاندارد ، دارای یک منحنی زنگدیس با**

$$\mu = 0 \text{ (میانگین)}$$

$$\sigma = 1 \text{ (انحراف معیار)}$$

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



## استفاده از جدول نرمال

جدول مساحت زیر منحنی در سمت چپ هر مقدار مشخص  $z$  را ارائه می دهد.

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

، احتمال یک فاصله  $[a, b]$

$$P[a \leq Z \leq b] = [\text{مساحت سمت چپ } b] - [\text{مساحت سمت چپ } a]$$

## خواص ناشی از خاصیت تقارن تابع چگالی حول مرکز صفر

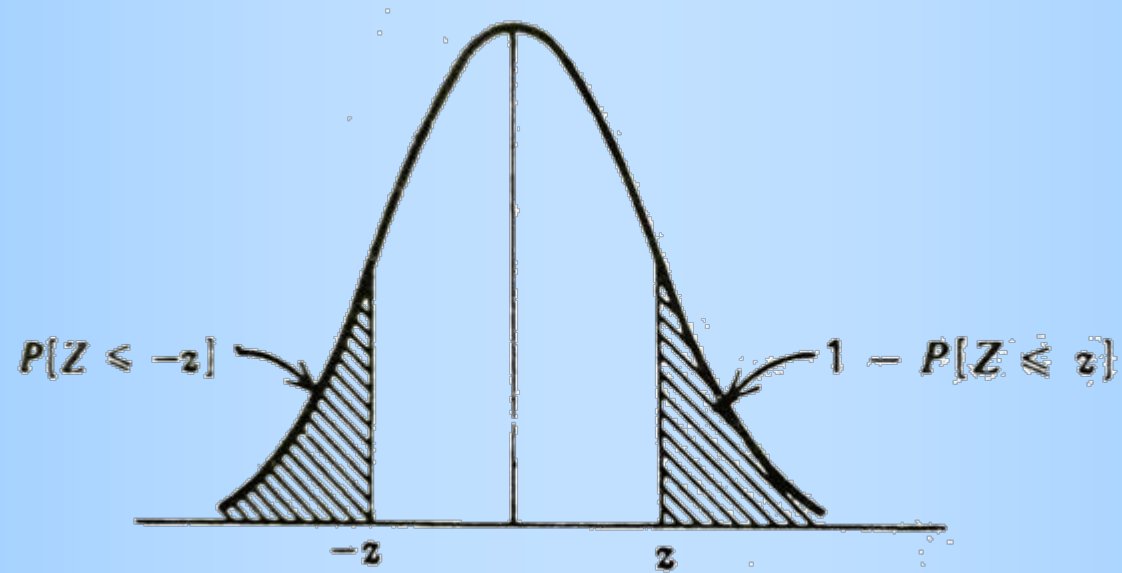
**الف)**  $P[Z \leq 0] = 0.5$

**ب)**  $P[Z \leq -z] = 1 - P[Z \leq z]$

**ج)** اگر  $Z > 0$  داریم

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

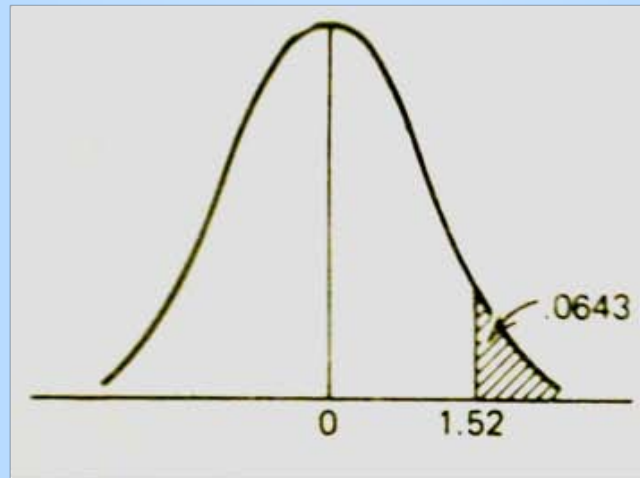
$$P[Z \leq -z] = 0.5 - P[0 < Z \leq z]$$



مثال:

$$P[Z > 1/52] = ?$$

$$P[Z > 1/52] = 1 - P[Z \leq 1/52] = 1 - 0.9357 = 0.0643$$



روش دیگر استفاده از خاصیت تقارن

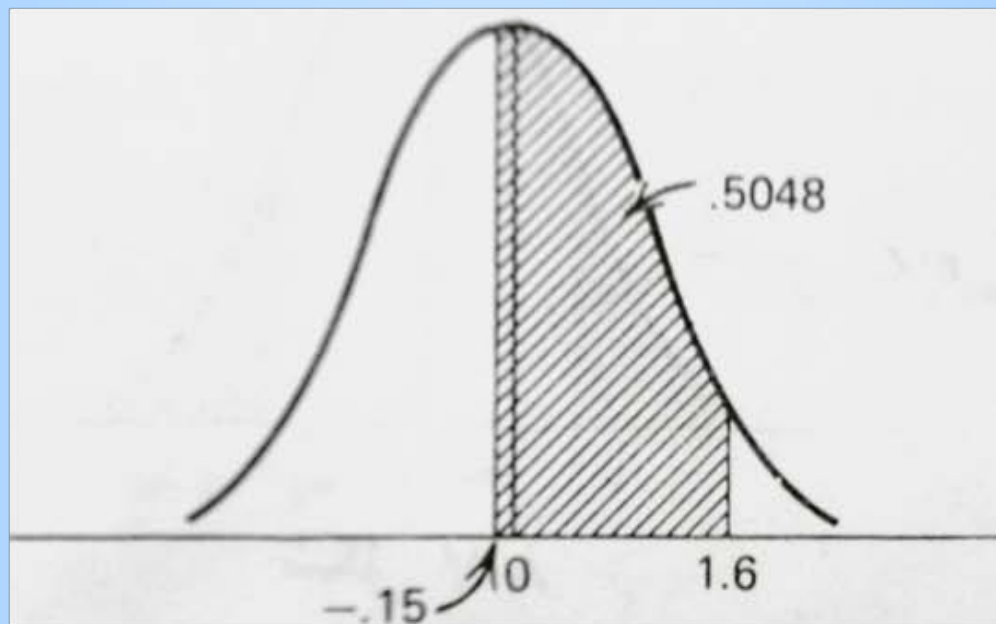
$$P[Z > 1/52] = P[Z < -1/52]$$

مثال:  $P[-0.15 < Z < 1.6] = ?$

$$P[Z \leq 1.6] = 0.9452 = \text{مساحت واقع در سمت چپ } 1.6$$

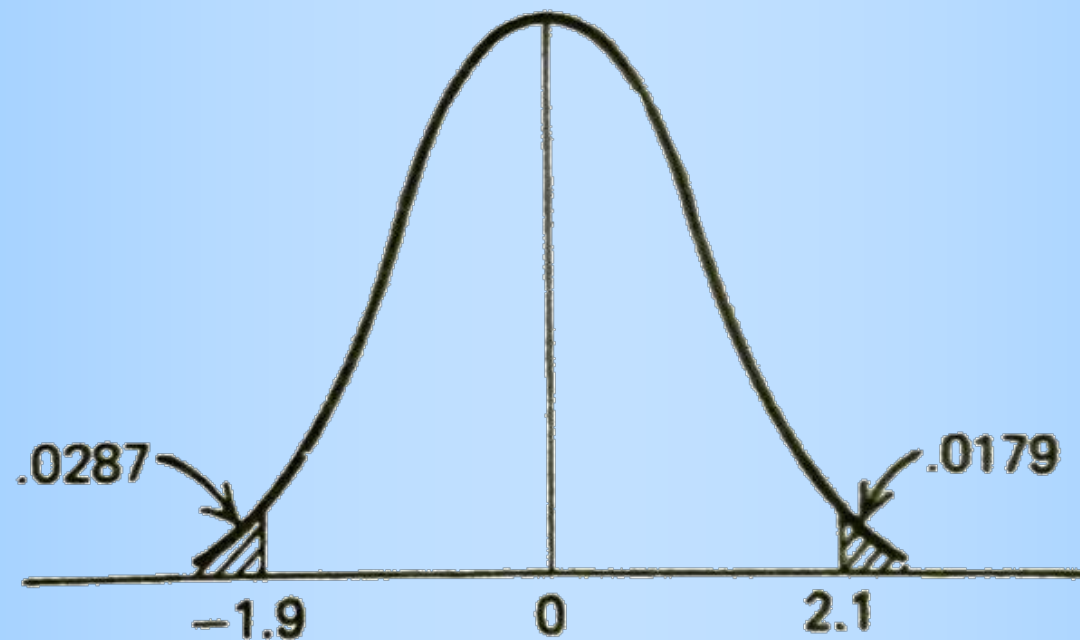
$$P[Z \leq -0.15] = 0.4404 = \text{مساحت واقع در سمت چپ } -0.15$$

$$P[-0.15 < Z < 1.6] = 0.9452 - 0.4404 = 0.5048$$



مثال:  $P[Z < -1/9 \text{ یا } Z > 2/1] = ?$

$$P[Z < -1/9 \text{ یا } Z > 2/1] = P[Z < -1/9] + P[Z > 2/1] = 0.0287 + 0.0179 = 0.0466$$



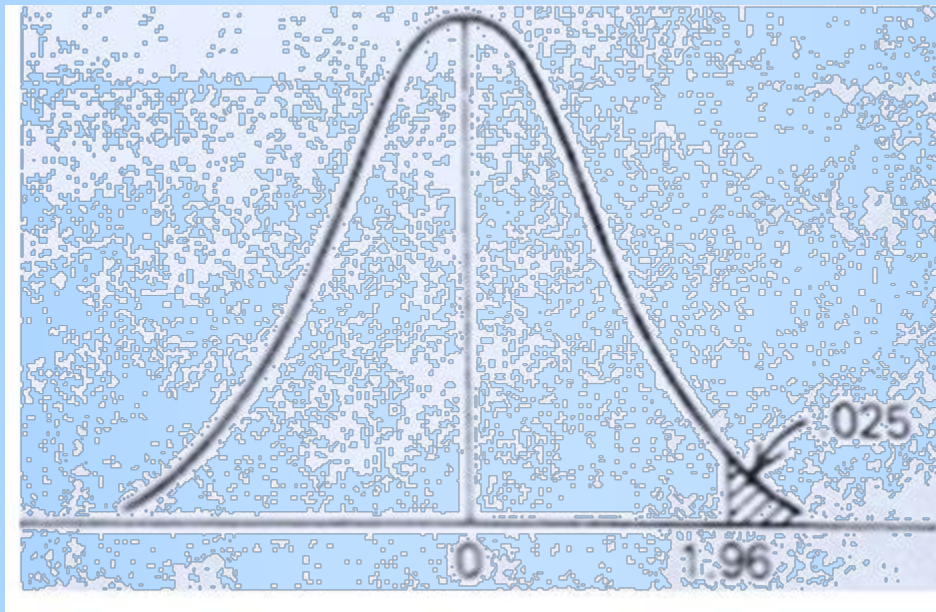


**مثال:** مقدار  $z$  را طوری پیدا کنید که در  $P[Z > z] = 0.025$  صدق کند.

می‌دانیم مساحت کل برابر با یک است. مساحت سمت چپ  $z$  برابر

$1 - 0.025 = 0.975$  باشد. مقدار کناری برای درایه  $0.975$  از جدول برابر با

$z = 1.96$  است.



**مثال:** مقدار  $z > 0$  را بدست آورید هرگاه  $P[-z \leq Z \leq z] = 0.90$ .

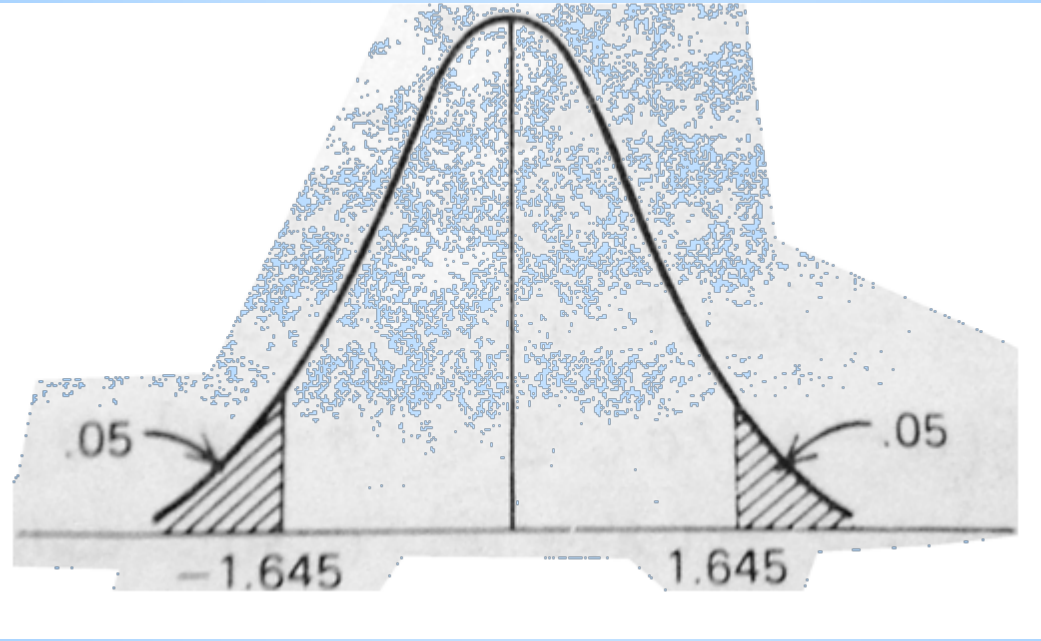
با توجه به تقارن منحنی داریم

$$P[Z < -z] = P[Z > z] = 0.05$$

$$P[Z \leq -1/45] = 0.0495 \text{ منجر به } z = 1/45$$

$$P[Z \leq -1/44] = 0.0505 \text{ منجر به } z = 1/44$$

با درون یابی بین این دو مقدار  $Z = 1/44.5$  را بدست می آوریم



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

بنابراین

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

$$\begin{aligned} P[a \leq X \leq b] &= P\left[\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right] \\ &= P\left[\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right] \end{aligned}$$

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

**مثال:** در صورتی که  $X$  دارای توزیع  $N(1, 4)$  باشد،  $P[X > -3]$  و  $P[-3 < X < 10]$  را محاسبه کنید.

در اینجا  $\mu = 1$  و  $\sigma = 2$ ، بنابراین عدد ۱ را از طرفین کم می کنیم و سپس بر ۲ تقسیم می نماییم:

$$P[X > -3] = P[X - 1 > -3 - 1] = P\left[\frac{X-1}{2} > \frac{-3-1}{2}\right] = P[Z > -2] = 0.9772$$

$$P[-3 < X < 10] = P\left[\frac{-3-1}{2} < \frac{X-1}{2} < \frac{10-1}{2}\right] = P[-2 < Z < 4.5] = 0.9999$$

## 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

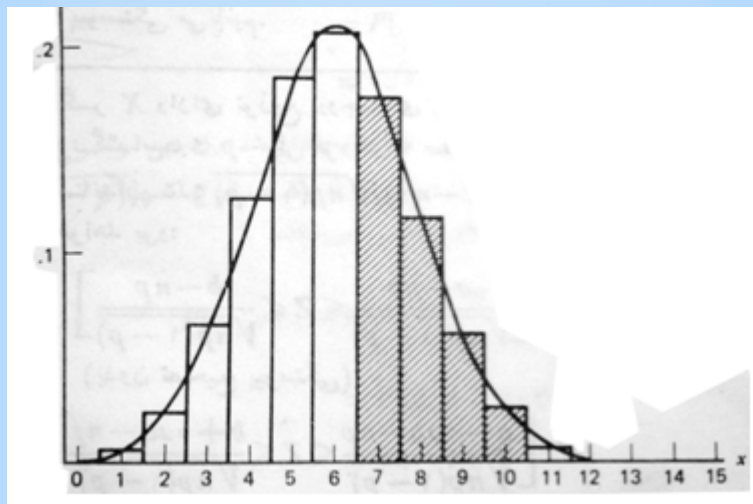
$X$  دارای توزیع دوجمله ای  $b(n,p)$  دارای میانگین  $\mu = np$  و انحراف معیار  $\sigma = \sqrt{np(1-p)}$  است.

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

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

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

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



**بافت نكار احتمال توزيع دو جمله ای  $b(15, 0/4)$  و منحنی نرمال تقریبی آن**

$$P[7 \leq X \leq 10] = 0/991 - 0/410 = 0/581$$

$$P[7 \leq X \leq 10] \approx P\left[\frac{7-6}{1/9} \leq Z \leq \frac{10-6}{1/9}\right]$$

$$= P[0/549 \leq Z \leq 3/636]$$

$$= 0/982 - 0/289$$

$$= 0/693$$



اگر  $X$  دارای توزیع دوجمله ای  $b(n,p)$  باشد، که 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:

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

با تصحیح پیوستگی:

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

**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)  $n = 1000$  و  $p = 0/3$  ،  $np = 300$  و  $\sqrt{np(1-p)} = \sqrt{210} = 14/5$

$X$  تعداد معتادان به سیگار دارای توزیع  $N(300, 14/5)$  است.

$$P[X \leq 279] \approx P\left[Z \leq \frac{279-300}{14/5}\right]$$

$$= P[Z \leq -1/448]$$

$$= 0/074$$

**با استفاده از تصحیح پیوستگی**

$$P[X \leq 279] \approx P\left[Z \leq \frac{279/5-300}{14/5}\right]$$

$$= P[Z \leq -1/414]$$

$$= 0/079$$

(ب)

$$P[X \geq 314] \approx P\left[Z \geq \frac{314 - \mu_{00}}{14/5}\right]$$

$$= P[Z \geq 1/10.3]$$

$$= 1 - 0.840$$

$$= 0.160$$

**نمونه‌گیری تصادفی، آماره و توزیع نمونه‌ای**

**مفاهیم بنیادی آمار عبارتند از :**

**الف) نمونه‌گیری تصادفی از یک توزیع احتمال**

**ب) توزیع‌های نمونه‌ای**

**مثال:** محققى قصد دارد در شهرى كه ۵۰۰۰ خانواده در آن ساكن هستند، تعداد دوپرفه‌هاى را كه هر خانواده دارد، بررسى نمايد.

توزيع جامعه  $X$ ، تعداد دوپرفه 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  $f(x)$ ، گرد آورده ای از  $n$  متغیر تصادفی مستقل  $X_1, \dots, X_n$  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  $(X_1, X_p)$  به حجم ۲ از این جامعه اختیار شود. توزیع نمونه‌ای آماره‌ای آماره  $\bar{X} = (X_1 + X_p)/۲$  چیست؟

$x$	۰	۱	۲	۳
$f(x)$	۰/۲	۰/۴	۰/۳	۰/۱

$$P[X_1=۰, X_p=۱] = P[X_1=۰]P[X_p=۱] = ۰/۲ \times ۰/۴ = ۰/۰۸$$

$x_i \backslash x_j$	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

$$P[\bar{X} = 1/5] = 0/02 + 0/12 + 0/12 + 0/02 = 0/28$$

$\bar{X}$ مقدار	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:**

**$\mu$  = میانگین جامعه**

**$\sigma^2$  = واریانس جامعه**

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

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

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$$

$$Var(\bar{X}) = \frac{1}{n^2} Var(X_1 + \dots + X_n)$$

$$= \frac{1}{n^2} [Var(X_1) + \dots + Var(X_n)]$$

$$= \frac{1}{n^2} [\sigma^2 + \dots + \sigma^2] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$E(\bar{X}) = \frac{1}{n} E(X_1 + \dots + X_n)$$

$$= \frac{1}{n} [E(X_1) + \dots + E(X_n)]$$

$$= \frac{1}{n} [\mu + \dots + \mu] = \frac{n\mu}{n} = \mu$$

## میانگین و انحراف معیار $\bar{X}$

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

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

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

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

**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

$$\mu = 1/3$$

$$\sigma^2 = 2/5 - (1/3)^2$$

$$= 2/5 - 1/9$$

$$= 8/45$$

$$\bar{X} = (X_1 + X_2) / 2 \quad \text{توزيع}$$

$x$	٠	٠/٥	١	١/٥	٢	٢/٥	٣	مجموع
$f(x)$	٠/٠٤	٠/١٦	٠/٢٨	٠/٢٨	٠/١٧	٠/٠٦	٠/٠١	١
$xf(x)$	٠	٠/٠٨	٠/٢٨	٠/٤٢	٠/٣٤	٠/١٥	٠/٠٣	١/٣
$x^2 f(x)$	٠	٠/٠٤	٠/٣٨	٠/٦٣	٠/٦٨	٠/٣٧٥	٠/٠٩	٢/٠٩٥

$$E(\bar{X}) = 1/3 = \mu$$

$$Var(\bar{X}) = 2/0.95 - (1/3)^2$$

$$= 2/0.95 - 1/9$$

$$= 0/40.5 = \sigma^2 / 2$$



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

## قضیه حد مرکزی

در نمونه گیری تصادفی از جامعه دلخواهی با میانگین  $\mu$  و انحراف معیار  $\sigma$ ، وقتی  $n$  بزرگ است توزیع  $\bar{X}$  توزیعی است تقریباً نرمال با میانگین  $\mu$  و انحراف معیار  $\sigma/\sqrt{n}$ . به عبارت دیگر  $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  تقریباً  $N(0, 1)$  است.

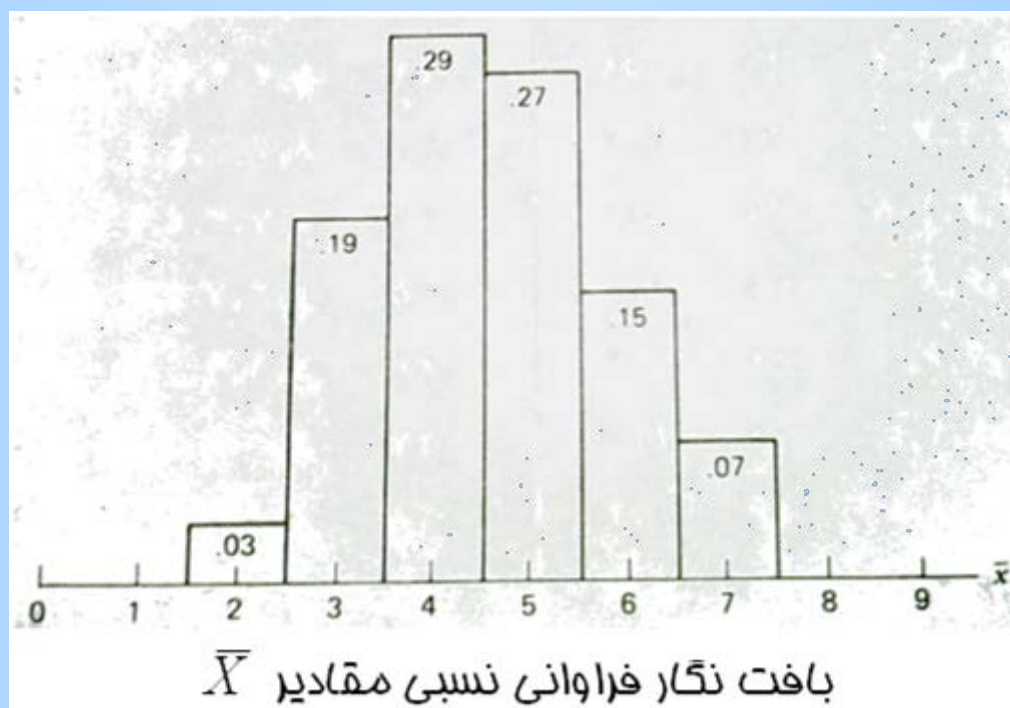
قضیه حد مرکزی به این مورد اشاره دارد که خواه توزیع جامعه پیوسته، گسسته، متقارن، یا چوله باشد مادام که واریانس جامعه متناهی است، توزیع میانگین نمونه  $\mu$ ، اگر حجم نمونه بزرگ باشد تقریباً نرمال است.

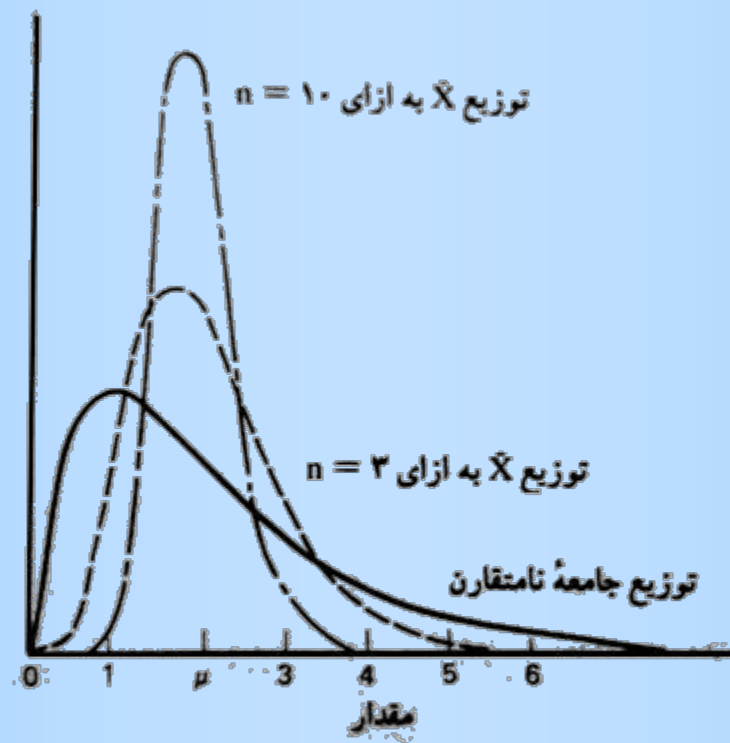
**مثال:** نمایش قضیه مد مرکزی: جامعه‌ای را در نظر بگیرید که دارای توزیع یکنواخت گسسته است که احتمال  $0/1$  را به هر عدد صحیح  $0, 1, \dots, 9$  نسبت می‌دهد. این مدل ممکن است برای توزیع رقم آخر شماره‌های تلفن یا اولین رقمی که در محاسبات سرریز می‌شود مدل مناسبی باشد.



نمودار فطی توزیع یکنواخت بر روی اعداد صحیح  $0, 1, \dots, 9$

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.





توزیع  $\bar{X}$  برای  $n=3$  و  $n=10$  در نمونه‌گیری از یک جامعه چاوله

**مثال:** فرض کنید معلوم شده که توزیع جامعه نیروی 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

$$\bar{X} \text{ انحراف معیار} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{75}} = 0.577$$

$$P[109 < \bar{X} < 111] = P\left[\frac{109-110}{0.577} < Z < \frac{111-110}{0.577}\right] = P[-1.73 < Z < 1.73] = 0.914$$

**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،  $X_i = 1$

اگر نتیجه امتحان  $i$ ام شکست باشد،  $X_i = 0$

توزیع هر یک از متغیرهای مستقل  $X_1, \dots, X_n$  بصورت زیر است

$x$	0	1
$f(x)$	$1-p$	$p$

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

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

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 ۵۲/۰ باشد چقدر است؟

در اینجا  $n=150$  و  $p=0/4$  است و عبارت  $P[\hat{p} < 0/52]$  باید محاسبه شود

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0/36}{150}} = 0/040$$

$$P[\hat{p} < 0/52] \approx P\left[Z < \frac{0/52-0/40}{0/040}\right] = P[Z < -3] = 0/0044$$

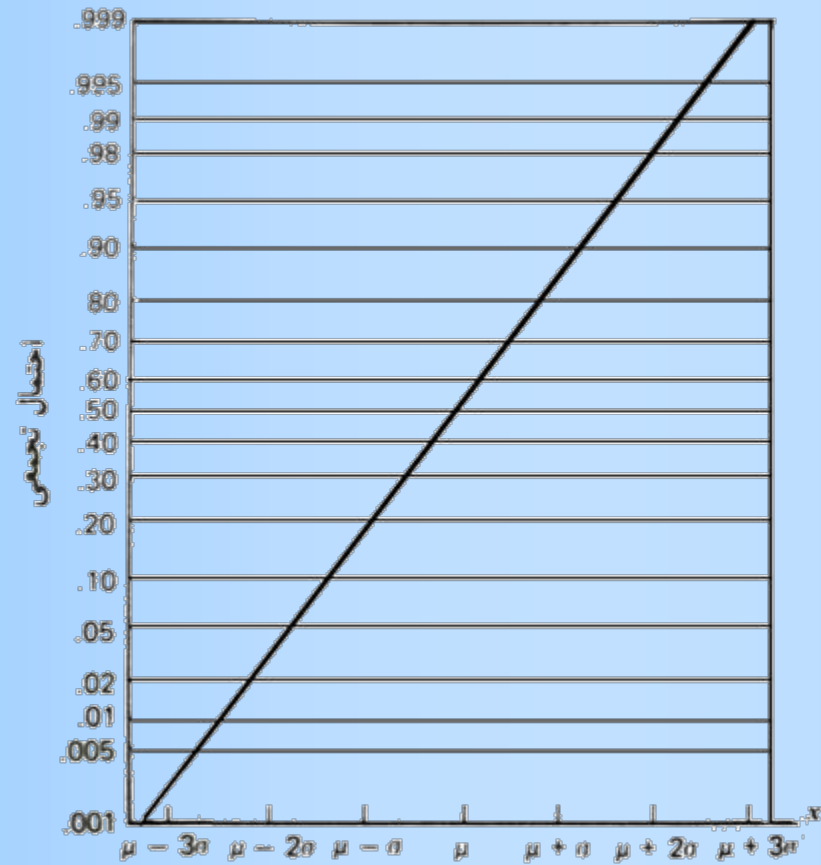
## بررسی فرض نرمال بودن یک جامعه

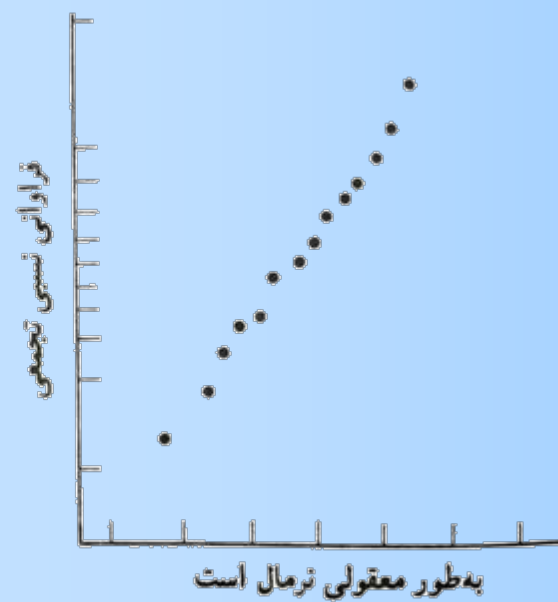
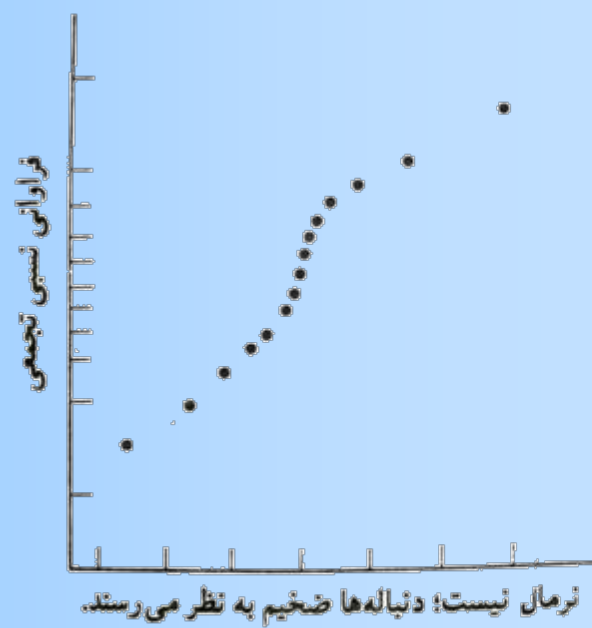
### ۱- بررسی نسبتهای مشاهدات در فواصل حول میانگین

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 در خارج فواصل	۱/۳	۱/۴۰	۱/۳۰۰

اگر حجم نمونه نسبتاً بزرگ باشد می توان انتظار داشت که میانگین نمونه  $\bar{X}$  به  $\mu$  و انحراف معیار نمونه  $s$  به  $\sigma$  نزدیک باشند.

## ۲- کاغذ احتمال نرمال





**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:**

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

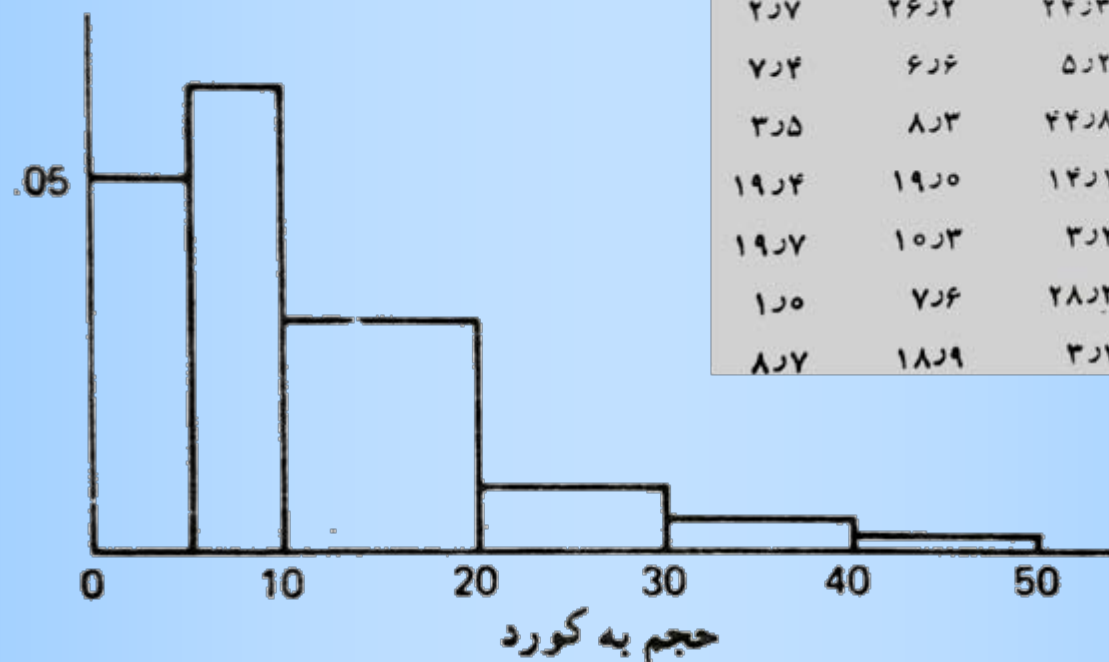


**مثال:** (فن تبدیل) جنگلانی به طور تصادفی ۴۹ قطعه زمین را در جنگل بزرگی برمی‌گزینند و حجم چوب در هر قطعه زمین را بر حسب کورد(نام واحد حجم چوب است برابر با ۱/۲۸ فوت مکعب) تعیین می‌کند. بافت نگارهای حجم چوب با استفاده از

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

## حجم چوب بر حسب گورد

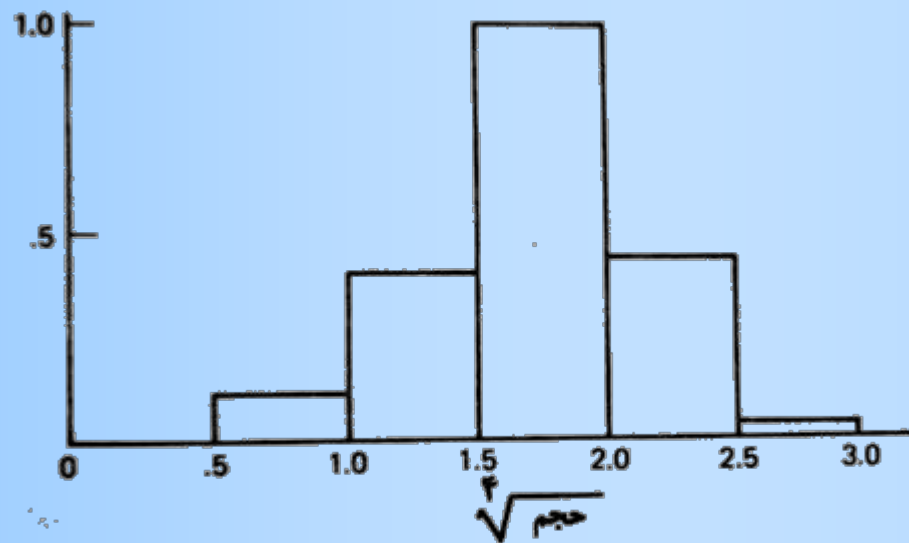
۳۹۲۳	۱۴۲۸	۶۲۳	۵۲۹	۶۲۵
۳۲۵	۸۲۳	۱۵۲۵	۱۲۳	۷۲۱
۶۲۵	۱۷۲۱	۱۶۲۸	۵۲۷	۷۲۹
۲۲۷	۲۶۲۲	۲۴۲۳	۱۷۲۷	۳۲۲
۷۲۴	۶۲۶	۵۲۲	۸۲۳	۵۲۹
۳۲۵	۸۲۳	۴۴۲۸	۸۲۳	۱۳۲۴
۱۹۲۴	۱۹۲۵	۱۴۲۱	۱۲۹	۱۲۲۵
۱۹۲۷	۱۵۲۳	۳۲۴	۱۶۲۷	۴۲۳
۱۲۵	۷۲۶	۲۸۲۳	۲۶۲۲	۳۱۲۷
۸۲۷	۱۸۲۹	۳۲۴	۱۵۲۵	



بافت نگار حجم چوب

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

۲۲۵۰	۱۲۹۶	۱۲۵۸	۵۲۹۷	۱۲۶۵
۱۲۳۷	۱۲۷۵	۱۲۷۸	۱۲۵۷	۱۲۶۳
۱۲۵۷	۲۲۵۳	۲۲۵۲	۵۲۹۱	۱۲۶۸
۱۲۲۹	۲۲۲۶	۲۲۲۲	۲۲۵۵	۱۲۳۲
۱۲۶۴	۱۲۶۵	۱۲۵۱	۱۲۷۵	۱۲۵۶
۱۲۳۷	۱۲۷۵	۲۲۵۹	۱۲۷۵	۱۲۹۱
۲۲۵۷	۲۲۱۵	۱۲۹۳	۱۲۱۷	۱۲۸۶
۲۲۱۱	۱۲۷۹	۱۲۳۶	۲۲۵۲	۱۲۴۴
۱۲۵۵	۱۲۶۶	۲۲۳۱	۲۲۲۶	۲۲۳۷
۱۲۷۲	۲۲۵۹	۱۲۳۶	۱۲۷۸	



پایان