

## جزوه آمار و احتمالات

مدرس : اسماعیلی

شماره تماس: ۰۹۱۴۹۲۳۸۸۲۶

[Collage.miyaneh@yahoo.com](mailto:Collage.miyaneh@yahoo.com)

تهیه کننده : بهروز اکرمی

منابع :

- آمار و احتمالات دکتر مسعود نیکوکار
  - آمار ریاضی فروند، والپول
  - آمار و احتمال مهندسی دانشگاه پیام نور
- نمره پایانی درس:
- نمره پایان ترم (۱۰ نمره)
  - نمره میان ترم (۵ نمره)
  - حل تمرینات و فعالیت کلاسی (۵ نمره) {تمرینات به صورت متناوب در هر جلسه ارائه می شود که دانشجو تا جلسه آینده مهلت تحویل دارد}

## سرفصلهای آمار و احتمالات

- فصل اول: آمار توصیفی
- فصل دوم: نظریه مجموعه ها، آنالیز ترکیبی
- فصل سوم: احتمال
- فصل چهارم: متغیرهای تصادفی
- فصل پنجم: امید ریاضی و گشتاورها
- فصل ششم: بررسی چند توزیع متغیرهای تصادفی گسسته
- فصل هفتم: بررسی چند توزیع پیوسته
- فصل هشتم: بررسی چند توزیع پیوسته
- فصل نهم: جامعه و نمونه آماری
- فصل دهم: نظریه برآورد کردن (تئوری تخمین)
- فصل یازدهم: آزمون فرض ها
- فصل دوازدهم: رگرسیون و همبستگی
- فصل سیزدهم: تحلیل واریانس

## فصل اول: آمار توصیفی

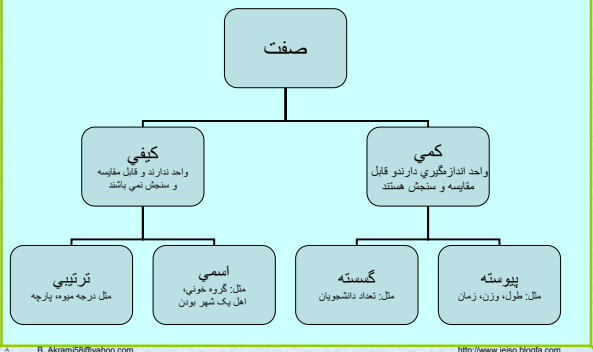
**علم آمار**: به مجموعه روش های علمی اطلاق می شود، که برای جمع آوری اطلاعات اولیه، مرتب و خلاصه کردن، طبقه بندی و تجزیه و تحلیل اطلاعات اولیه و تفسیری آنها به کار می رود.

**جامعه آماری**: هر مجموعه از اشیاء یا افراد که لاقط یک صفت مشترک داشته باشند را جامعه آماری گویند.

**صفت مشخصه**: صفت مشترک بین اعضای جامعه آماری را صفت مشخصه گویند.

مثال: مجموعه دانشجویان **ترم ۵** دانشگاه تهران یک جامعه آماری است که صفت مشخصه این جامعه آماری ترم ۵ بودن آنهاست ولی در جامعه آماری هدف بررسی **صفات غیر مشترک** مثل قد، وزن، تاریخ تولد و... می باشد.

صفات متغیر که در آمارگیری مورد استفاده قرار می‌گیرند به صورت زیر تقسیم بندی می‌گردند:



**داده های آماری:** مجموعه مقادیر صفت متغیر که بوسیله اعداد یا نشان ها نمایش داده می شود، را داده های آماری می‌نامیم.

برای مثال اعداد زیر با واحد سانتیمتر می‌تواند داده‌های آماری برای طول قد یک جامعه آماری باشد:

۱۷۰ ۱۶۷ ۱۵۴ ۱۶۱ ۱۵۵ ۱۵۲ ۱۷۸ ۱۶۲ ۱۶۵ ۱۷۱

یا داده های زیر می‌تواند نشانگر گروه خونی جامعه آماری مورد مطالعه باشد.

A B AB O A A AB A

**آمار توصیفی:** آن قسمت از علم آمار که مشتمل بر خلاصه کردن داده ها در قالب جداول، نمایش ترسیمی آنها بوسیله نمودار و محاسبه شاخصهای عددی گرایش به مرکز، پراکنندگی، چولگی و کشیدگی می باشد را آمار توصیفی نامند.

**آمار استنباطی:** آن قسمت از علم آمار که درباره تخمین پارامترهای جامعه از روی پارامترهای نمونه آماری بحث میکند را آمار استنباطی گویند.

### آمارگیری

تمام فعالیتی که برای جمع آوری داده ای آماری به کار می رود.

#### نمونه برداری

بخشی از افراد جامعه طبق اصول خاص\* مورد مطالعه قرار می‌گیرد.

#### سرشماری

همه افراد جامعه مورد مطالعه قرار می‌گیرد.

\* بهترین روش نمونه برداری، روش تصادفی است که باید به تمام افراد جامعه شانس انتخاب یکسان داد.

### نمودارهای داده های کیفی:

**نمودار میله ای، سوزنی، ستونی:** این نمودار در یک دستگاه مختصات که محور افقی نشان دهنده کیفیت مشاهدات و محور عمودیش نشان دهنده فراوانی مطلق یا نسبی (فراوانی که در آن نسبت فراوانی هر دسته به فراوانی کل در نظر گرفته می‌شود). هر گروه است ترسیم می شود. مقدار فراوانی را میتوان با نقطه (سوزنی) میله یا ستون مشخص کرد.

**نمودار دایره‌ای:** در این نمودار، دایره را به چند بخش متناسب با فراوانی هر دسته تقسیم میکنیم که هر بخش برای هر دسته از رابطه زیر محاسبه می گردد.

$$X_i = \frac{360}{\sum f_i} \times f_i$$

### نمایش جدولی داده های کیفی:

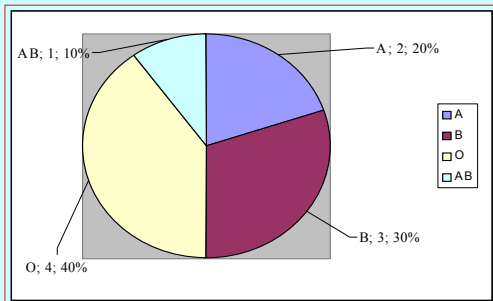
فرض کنید گروه خونی ۱۰ دانشجو تعیین و نتایج زیر به دست آمده باشد.

A O B AB A B B O O O

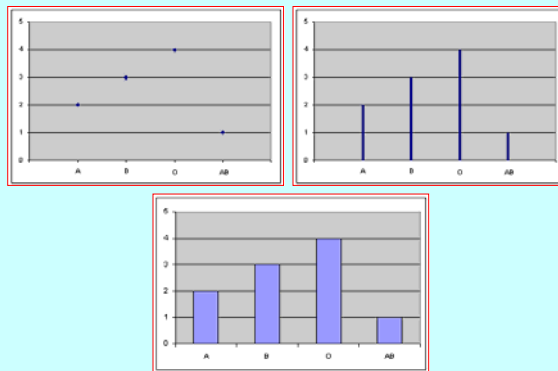
گروه خون	فراوانی
A	۲
B	۳
O	۴
AB	۱

نمایش جدولی نتایج به صورت زیر می‌باشد.

### نمودار دایره ای برای مثال گروه خونی



### نمودار میله ای، سوزنی، ستونی برای مثال گروه خونی



در اکثر موارد تعداد داده ها زیادند یا اینکه صفت مورد بررسی از نوع پیوسته اند (قد، وزن، زمان...) که در این صورت داده ها باید طبقه بندی گردند. اولین قدم پیدا کردن پارامترهای زیر است:

• **دامنه:** عبارت است از تفاضل بزرگترین داده از کوچکترین داده که با  $R$  نمایش داده می شود. در محاسبه پارامترها چون معمولاً داده ها گرد میشوند با توجه به دقت گرد باید مقادیر واقعی مورد استفاده قرار گیرند.

$$R = \text{Max}(X_i) - \text{Min}(X_i)$$

### نمایش جدولی داده های کمی

اطلاعات به دست آمده از اندازه گیری یا شمارش، همواره به صورت اعداد بیان می شوند که به آنها **داده های خام** می گوئیم. داده های خام باید مرتب و طبقه بندی گردند تا قابل تفسیر و تجزیه و تحلیل آماری شوند.

داده های مربوط به نمره درس ۱۰ دانشجوی

اولین قدم مرتب کردن داده ها از کوچک به بزرگ باشد.

۱۲ ۱۷ ۱۳ ۱۹ ۱۴ ۱۶ ۱۵ ۷ ۶ ۵

حالا با نگاه اجمالی به داده ها می توان اطلاعاتی چون کمترین یا بیشترین نمره را به دست آورد.

### نمایش جدولی داده های کمی طبقه بندی شده:

مثال: داده های زیر طول عمر ۲۵ لامپ بر حسب ساعت می باشد و با دقت کمتر از ۱ گرد شده اند داده ها را طبقه بندی و جدول آن را رسم کنید؟

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

$$R = \text{Max}(X_i) - \text{Min}(X_i) = 110.5 - 96.5 = 14$$

$$K \cong \sqrt{N} = \sqrt{25} = 5$$

$$C \cong \frac{R}{K} = \frac{14}{5} \cong 3$$

### تعداد طبقه: با توجه به موضوع مورد بررسی باید تعداد طبقات

تعیین گردد و معمولاً بین ۵ تا ۲۰ انتخاب می کنند. تعداد طبقات باید طوری انتخاب شود که اطلاعات زیاد از دست نرود و فرمول نیز به صورت زیر برای تعیین تعداد طبقات داریم که اولی به فرمول استورجس معروف است که در آن  $N$  تعداد داده ها و  $K$  تعداد طبقات می باشد.

$$K \cong 1 + 3.322 \text{Log}N$$

$$K \cong \sqrt{N}$$

**فاصله طبقات:** از تقسیم دامنه بر تعداد طبقات با تقریب اضافی محاسبه می شود که با  $C$  نمایش داده می شود.

$$C \cong \frac{R}{K}$$

### نمودار های داده های کمی

• **هیستوگرام فراوانی:** نموداریست در دستگاه مختصات که محور افقی آن با حدود واقعی طبقات و محور عمودی آن با فراوانی مطلق یا فراوانی نسبی داده ها مشخص می شود.

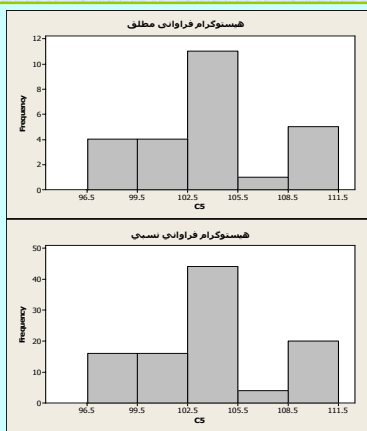
• **چند ضلعی فراوانی:** نموداریست که متناظر با هر نماینده طبقه در محور افقی و فراوانی آن در محور عمودی، یک نقطه در صفحه مختصات ایجاد و به هم وصل می شوند که معمولاً هیستوگرام فراوانی با چند ضلعی فراوانی را در یک دستگاه رسم می کنند.

• **پلی گون فراوانی تجمعی:** برای ترسیم این نمودار، از نماینده طبقات در محور افقی و فراوانی تجمعی در محور عمودی استفاده می شود، سپس نقاط ایجاد شده به ترتیب به هم وصل می شوند.

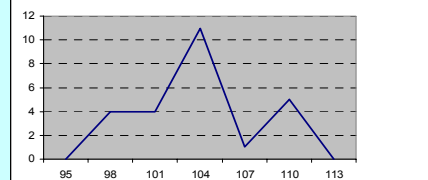
حدود واقعی	حدود طبقات	نماینده دسته	فراوانی مطلق $f_i$	فراوانی تجمعی	فراوانی نسبی	فراوانی نسبی
		میانگین کران بالا و پایین طبقه $x_i = \frac{L_i + U_i}{2}$	تعداد دفعات تکرار داده ها در هر دسته	$F_i = f_i + F_{i-1}$	به درصد نیز بیان میشود $f_{c_i} = \frac{f_i}{N}$	به درصد نیز بیان میشود $F_{c_i} = f_{c_i} + F_{c_{i-1}}$
۹۶,۵-۹۹,۵	۹۷-۹۹	۹۸	۴	۴	۰,۱۶	۰,۱۶
۹۹,۵-۱۰۲,۵	۱۰۰-۱۰۲	۱۰۱	۴	۸	۰,۱۶	۰,۳۲
....	۱۰۳-۱۰۵	۱۰۴	۱۱	۱۹	۰,۴۴	۰,۷۶
...	۱۰۶-۱۰۸	۱۰۷	۱	۲۰	۰,۰۴	۰,۸۰
....	۱۰۹-۱۱۱	۱۱۰	۵	۲۵	۰,۲	۱
			$\sum f_i = 25$		$\sum f_{c_i} = 1$	

### نمودار های داده های کمی

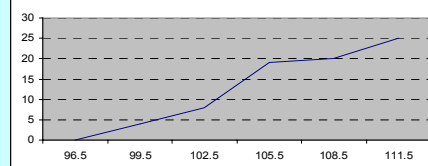
نمودار فراوانی تجمعی: تنها فرق این نمودار با نمودار پلی گون فراوانی تجمعی در این است که در این نمودار بجای نماینده طبقات از حدود واقعی استفاده می شود این نمودار در محاسبه چندکها (چارکها، دهکها، صدکها) و مقایسه پدیده هایی (مثل میزان رشد تورم در کشورها) به کار می رود.



### نمودار چند ضلعی فراوانی

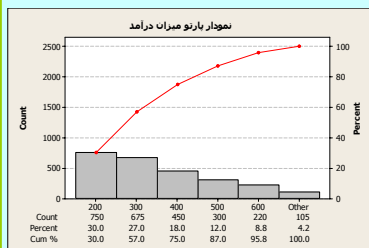


### نمودار فراوانی تجمعی

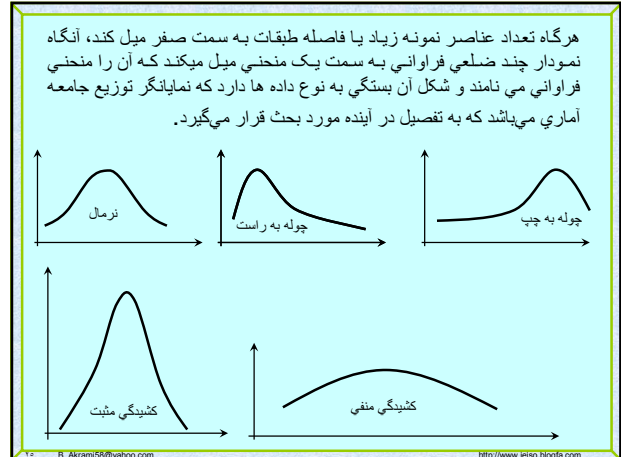
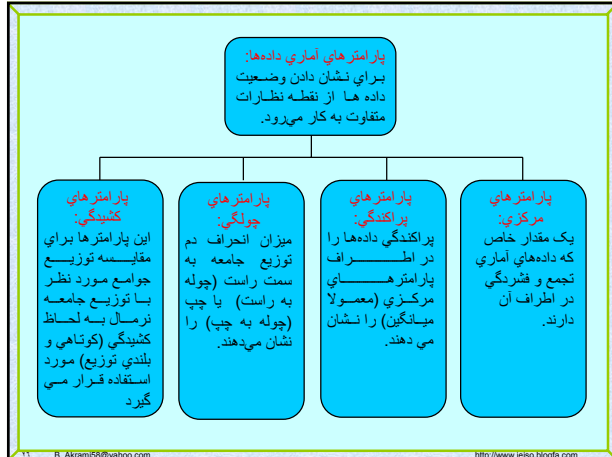


نمودار پارتنو: یکی از مهمترین نمودارهای آماری که بیشتر برای داده های کیفی می باشد، نمودار پارتنو است. این نمودار دارای سه محور است

- ۱- محور افقی: نوع موضوعات
- ۲- محور عمودی: فراوانی مطلق موضوعات
- ۳- محور سوم (روبروی محور عمودی): فراوانی نسبی تجمعی موضوعات



مثال: اگر میزان درآمد ۲۵۰۰ نفر از یک نمونه تصادفی با واحد هزار تومان به صورت زیر باشد نمودار پارتنو را برای آن رسم می کنیم.



**پارامترهای مرکزی:**

**• میانگین هندسی:** این میانگین حد متوسط شاخصها، نسبتها و درصدها را بیان میکند مثلاً افزایش جمعیت، کشت باکتری، تجزیه رادیواکتیو...

برای محاسبه معمولاً از  $\log$  استفاده می‌کنیم.

$$A) G = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

$$B) G = \sqrt[k]{x_1^{f_1} \times x_2^{f_2} \times \dots \times x_N^{f_N}} = \sqrt[k]{\prod_{i=1}^K x_i^{f_i}}$$

**پارامترهای مرکزی:**

قرارداد: در محاسبه پارامترها دو حالت وجود دارد اگر داده‌ها طبقه بندی شده باشند که با  $B$  در غیر اینصورت با  $A$  نشان می‌دهیم.

**• میانگین حسابی:** در بین پارامترهای مرکزی نوسان کمتری دارد و از ثبات بیشتری برخوردار است.

$$A) \bar{x} = \frac{1}{N} \sum_{i=1}^N X_i$$

$$B) \bar{x} = \frac{1}{N} \sum_{i=1}^k f_i x_i$$

**• میانگین حسابی وزنی:** اگر داده‌های  $X_1, X_2, \dots, X_N$  دارای ضریب وزنی  $w_1, w_2, \dots, w_N$  آنگاه

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

**پارامترهای مرکزی:**

**• میانگین درجه دوم:**

$$A) Q = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$$

$$B) Q = \sqrt{\frac{1}{N} \sum_{i=1}^K f_i x_i^2}$$

• رابطه بین میانگینهای بیان شده به صورت زیر است

$$H \leq G \leq \bar{X} \leq Q$$

**• میانگینهای پیراسته و وینزوری بعد از معرفی چارکها بیان می‌شود.**

**پارامترهای مرکزی:**

**• میانگین همساز (هارمونیک):** اگر هیچکدام از داده‌ها صفر نباشد میانگین همساز از رابطه زیر محاسبه می‌گردد.

این میانگین برای محاسبه حد متوسط سرعتها، مطالعه در شبکه‌های برق و عینک شناسی بکار می‌رود.

$$A) H = N / \sum_{i=1}^N \frac{1}{x_i}$$

$$B) H = N / \sum_{i=1}^k \frac{f_i}{x_i}$$

پارامترهاي مركزي:

• **ميانه:** ابتدا داده ها را به طور غير نزولي مرتب مي كنيم اگر تعداد داده ها فرد باشد ميانه عدد وسطي است در غير اينصورت ميانگين حسابي دو عدد وسطي معرف ميانه مي باشد. در داده هاي طبقه بندي شده از رابطه زير بدست مي آيد.

$i$ : شماره طبقه ميانه دار و عبارتست از اولين طبقه اي كه فراواني تجمعي آن بزرگتر يا مساوي  $N/2$  باشد.  
ميانه مشاهدات را به دو بخش مساوي تقسيم مي كند و تحت تاثير داده هاي پرت قرار نمي گيرد.

$$M_d = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \times C$$

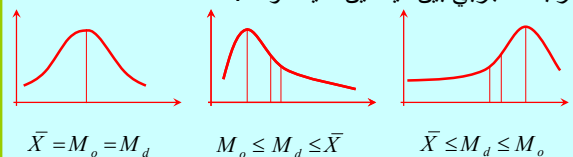
پارامترهاي مركزي:

• **مد يا نما:** اندازه اي از متغير است كه فراواني آن ماكسيم باشد كه در داده هاي كيفي مهمترين شاخص مركزي مد مي باشد.

$i$ : شماره طبقه مد دار و عبارتست از طبقه اي كه فراواني آن ماكسيم است.

$$M_o = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \times C$$

رابطه تجربي بين ميانگين، ميانه و مد:



$$\bar{X} = M_o = M_d \quad M_o \leq M_d \leq \bar{X} \quad \bar{X} \leq M_d \leq M_o$$

در توزيع هايي كه چولگي زياد نباشد رابطه تجربي زير كه به رابطه **پيرسن** معروف است، برقرار مي باشد.

$$\{\text{ميانه} - \text{ميانگين}\} = 3 \times \{\text{مد} - \text{ميانگين}\}$$

پارامترهاي مركزي:

• **چنگها:** هرگاه داده ها را به طور غير نزولي مرتب كنيم. عددي را كه لااقل  $p$  درصد داده ها كوچكتر از آن و  $(100-p)$  درصد داده ها بزرگتر از آن باشند، صدك  $p$  نامند. صدك ۵۰، ميانه است حال صدك هاي معروف را نام مي بريم.

• **چارک ها:** آن ها را  $Q_1, Q_2, Q_3$  نشان مي دهند كه به ترتيب برابر است با  $P_{25}, P_{50}, P_{75}$ . در اينجا نيز  $Q_2$  همان ميانه است.

• **دهک ها:** آن را با  $D_1, D_2, \dots, D_9$  نشان مي دهيم كه به ترتيب برابر است با  $P_{10}, P_{20}, \dots, P_{90}$ . در اينجا نيز  $D_5$  همان ميانه است.

پارامترهاي مركزي:

• **ميانگين پيراسته:** براي محاسبه آن تمام داده هاي بزرگتر از چارك سوم و كوچكتر از چارك اول را کنار مي گذاريم سپس ميانگين حسابي باقي داده ها را حساب مي كنيم.

• **ميانگين وينزوري:** به جاي تمام داده هاي كوچكتر از چارك اول، مقدار چارك اول و بجاي تمام داده هاي بزرگتر از چارك سوم مقدار چارك سوم را قرار مي دهيم سپس ميانگين حسابي مجموعه داده هاي جديد را حساب مي كنيم.

محاسبه صدكها: داده ها را بطور غير نزولي مرتب مي كنيم اگر  $N$  تعداد كل داده ها باشد براي محاسبه صدك  $p$  ابتدا  $(N+1)p$  را محاسبه مي كنيم.

$$\begin{aligned} [(N+1)p] &= K \\ \text{if } K \in \mathbf{Z} &\Rightarrow \mathbf{x}_K = P_p \quad \text{else} \\ &K = k + r \quad 0 < r < 1 \\ &P_p = (1-r)x_k + rx_{k+1} \end{aligned}$$

اگر داده ها طبقه بندي باشند:

$$P_p = L_i + \frac{(N \times p / 100) - F_{i-1}}{f_i} \times C$$

كه در آن  $i$  طبقه صدك دار مي باشد و عبارتست از اولين طبقه اي كه فراواني تجمعي آن بزرگتر يا مساوي  $NP/100$  مي باشد.

### پارامترهای پراکندگی

• **دامنه تغییرات:** عبارتست از اختلاف بزرگترین و کوچکترین داده که از اهمیت کمتری برخوردار است چون تحت تأثیر دو داده از مجموعه داده‌ها قرار دارد.

$$R = \text{Max}_{\forall i} (X_i) - \text{Min}_{\forall i} (X_i)$$

• **انحراف چارکها:** به صورت زیر تعریف می‌شود. که این پارامتر تحت تأثیر ۵۰ درصد داده‌هاست

$$Q = \frac{1}{2}(Q_3 - Q_1)$$

### پارامترهای پراکندگی

• **انحراف متوسط یا انحراف از میانگین:** به صورت زیر تعریف می‌شود که تحت تأثیر تمام داده‌هاست. مشکلی که این پارامتر دارد وجود عبارت قدر مطلق می‌باشد که محاسبات را پیچیده می‌کند.

$$A) A.D = \frac{1}{N} \sum_{i=1}^N |x_i - \bar{X}|$$

$$B) A.D = \frac{1}{N} \sum_{i=1}^K f_i |x_i - \bar{X}|$$

### پارامترهای پراکندگی

• **واریانس:** در یک نمونه به صورت زیر تعریف می‌شود که تحت تأثیر تمام داده‌هاست و پیچیدگی محاسبات قدر مطلق نیز در آن نیست.

$$A) S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$$

$$B) S^2 = \frac{1}{N-1} \sum_{i=1}^K f_i (x_i - \bar{X})^2$$

حال فرض کنید داده‌ها مربوط به طول قد افراد بر حسب سانتی‌متر باشد ولی واحد  $(x_i - \bar{X})^2$  سانتی‌متر مربع خواهد شد که ایراد واریانس نیز در این مورد می‌باشد.

### پارامترهای پراکندگی

• چنانچه مطالعات روی جامعه باشد و یا تعداد نمونه زیاد گردد در این صورت میانگین را با  $\mu$  و واریانس با  $\sigma^2$  نمایش داده و به شکل زیر محاسبه می‌گردد.

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

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

### پارامترهای پراکندگی

• **انحراف معیار:** جذر مثبت واریانس را انحراف معیار می‌نامیم که مشکل مربع شدن واحد در واریانس را نیز ندارد و به عنوان بهترین پارامتر پراکندگی می‌باشد.

$$S = \sqrt{S^2} \leftarrow \text{انحراف معیار نمونه}$$

$$\sigma = \sqrt{\sigma^2} \leftarrow \text{انحراف معیار جامعه}$$

• **ضریب تغییرات:** برای مقایسه پراکندگی دو جمعیت استفاده می‌گردد و مزیت آن در بدون واحد بودن آن است.

$$C = \frac{\sigma}{\bar{X}} \times 100\%$$

### پارامترهای پراکندگی

• **گشتاور:** یکی دیگر از پارامترهای پراکندگی می‌باشد. گشتاور مرتبه  $r$ ام پیرامون نقطه ثابت  $a$  را برای جامعه به شکل زیر نمایش و تعریف می‌کنند.

$$A) m_r^a = \frac{1}{N} \sum_{i=1}^N (x_i - a)^r$$

$$B) m_r^a = \frac{1}{N} \sum_{i=1}^K f_i (x_i - a)^r$$

گشتاور حول میانگین جامعه را گشتاور مرکزی نامند و آن را با  $\mu_r$  نشان می‌دهند

$$A) m_r^\mu = \mu_r = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^r$$

$$B) m_r^\mu = \mu_r = \frac{1}{N} \sum_{i=1}^K f_i (x_i - \mu)^r$$

### پارامترهای پراکنندگی

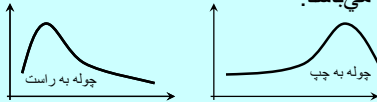
مواضع است که واریانس جامعه برابر است با گشتاور مرکزی مرتبه دوم.

$$\mu_2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 = \sigma^2$$

گشتاور حول مبدا را نیز به اختصار با  $\mu'_r$  نشان می‌دهند در نتیجه رابطه گشتاور به صورت زیر در می‌آید.

$$\mu'_r = \frac{1}{N} \sum_{i=1}^N (x_i)^r$$

**پارامترهای چولگی:** آن را با  $SK$  نشان می‌دهند و اگر دم توزیع جامعه به سمت راست باشد، توزیع را چوله به راست و در صورت عکس، آن را چوله به چپ می‌نامند. این مقدار طوری تعریف گردیده است که در صورت صفر بودن گویند توزیع بدون چولگی است و نرمال می‌باشد در غیر اینصورت با مثبت بودن مقدار چولگی بسته به عدد چولگی، توزیع چوله به راست می‌باشد و در صورت منفی بودن، باز بسته به مقدار آن توزیع چوله به چپ می‌باشد.



### پارامترهای چولگی:

• **ضریب چولگی گشتاوری:**

$$SK = \frac{\mu_3}{\sigma^3}$$

• **ضریب های چولگی پیرسون:**

$$1) SK = \frac{\mu - M_o}{\sigma}$$

$$2) SK = \frac{3(\mu - M_d)}{\sigma}$$

• **ضریب های چولگی چندکی:**

$$1) SK = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

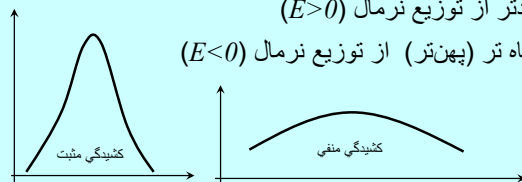
$$2) SK = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

**پارامترهای کشیدگی:** این پارامترها برای مقایسه توزیع جوامع مورد نظر با توزیع جامعه نرمال به لحاظ کشیدگی (کوتاهی و بلندی توزیع) مورد استفاده قرار می‌گیرد که آن را با  $E$  نشان می‌دهیم. که باز سه حالت زیر پیش می‌آید:

• مساوی توزیع نرمال ( $E=0$ )

• بلندتر از توزیع نرمال ( $E>0$ )

• کوتاه تر (پهن‌تر) از توزیع نرمال ( $E<0$ )



### روش کد گذاری در محاسبه پارامترها:

در بعضی مواقع در محاسبه پارامترها راحت‌تر این است که مشاهدات را تغییر دهیم و به جای محاسبه پارامترها با داده‌های اصلی، با داده‌هایی جدید حساب کنیم ولی این تغییر باید طوری گردد که بتوان با پارامتر محاسبه شده به پارامتر اصلی برسیم.

یکی از این تغییرات به شکل زیر است

اگر  $x_i$  ها مجموعه داده‌های ما در جامعه آماری باشند (در داده‌های طبقه بندی شده نماینده طبقات در نظر گرفته می‌شود) داده‌های جدید را به صورت زیر در نظر می‌گیریم:

$$u_i = \frac{x_i - A}{C}$$

### پارامترهای کشیدگی:

• **ضریب کشیدگی گشتاوری**

$$E = \frac{\mu_4}{\sigma^4} - 3$$

### روش کد گذاری در محاسبه پارامترها:

که در آن  $A$  و  $C$  اعداد ثابتی هستند. در داده‌های طبقه بندی شده  $C$  را فاصله طبقات  $A$  را  $x_i$  مربوط به نماینده طبقه وسط یا طبقه مد دار در نظر می‌گیرند. حال داریم:

$$u_i = \frac{x_i - A}{C} \Rightarrow x_i = C u_i + A$$

$$\bar{X} = C \bar{U} + A$$

$$S_x^2 = C^2 S_u^2$$

$$S_x = C^2 S_u$$

$$m_r^a(x) = C^r m_r^a(u)$$

**تمرین:** جدول زیر مربوط به آب مصرفی ۲۰ خانوار میباشد مطالبه:

- تکمیل جدول با پارامترهای نماینده، فراوانی جمعی، فراوانی نسبی و فراوانی جمعی نسبی هر طبقه.
- رسم نمودار هیستوگرام و چندضلعی فراوانی
- رسم نمودار دایره ای بر اساس درجه میزان مصرف خانوار
- محاسبه میانگین، مد و میانه دادها
- محاسبه انحراف چارک ها، واریانس و ضریب تغییرات
- محاسبه ضریب چولگی پیرسون

درجه میزان مصرف	فراوانی	حدود طبقات
بسیار کم مصرف	۲	۱۰ - ۱۲
کم مصرف	۴	۱۳ - ۱۵
مصرف متوسط	۶	۱۶ - ۱۸
پر مصرف	۳	۱۹ - ۲۱
بسیار پر مصرف	۵	۲۲ - ۲۴

استفاده از روش کد گذاری در محاسبه پارامترها مجاز می باشد.

**تمرین:** جدول زیر مربوط به طول قد ۱۰۰ دانشجو میباشد مطالبه:

- تکمیل جدول با پارامترهای نماینده، فراوانی جمعی، فراوانی نسبی و فراوانی جمعی نسبی هر طبقه.
  - رسم نمودار های هیستوگرام و چند ضلعی فراوانی
  - قد چند درصد از دانشجویان بالای ۱۶۸ سانتی متر می باشد.
  - محاسبه میانگین، مد و میانه قد دانشجویان
  - محاسبه انحراف چارک ها، انحراف از میانگین واریانس و انحراف معیار
  - محاسبه ضریب چولگی
- استفاده از روش کد گذاری در محاسبه پارامترها مجاز می باشد.

فراوانی	حدود طبقات
۱۵	۱۵۰ - ۱۵۶
۲۰	۱۵۷ - ۱۶۳
۳۰	۱۶۴ - ۱۷۰
۲۵	۱۷۱ - ۱۷۷
۱۰	۱۷۸ - ۱۸۴

**تمرین:** فرض کنید فاصله سه شهر A, B, C از یکدیگر برابر باشد، اتومبیلی فاصله بین A تا B را با سرعت ۳۰ کیلومتر در ساعت و فاصله B تا C را با سرعت ۴۰ کیلومتر در ساعت و فاصله C تا A را با سرعت ۵۰ کیلومتر در ساعت پیموده است، سرعت متوسط حرکت این اتومبیل را حساب نمایید.

**تمرین:** فرض کنید میزان تولید کارخانه ای در چهار سال متوالی ۲، ۴، ۶ و ۲۷ برابر نسبت به سال قبل باشد. مطالبه است میزان افزایش متوسط تولید کارخانه

**تمرین:** میانگین درجه دوم داده های زیر را حساب کنید.

۱۲	۱۰	۷	۶	۶	۵	۳
----	----	---	---	---	---	---

**مجموعه:** عبارتست از یک دسته از اشیاء که کاملاً مشخص باشند و بطوری که هر شئی مفروض، یا متعلق (عضو) به مجموعه هست و یا نیست.

• بحثهای نمودار ون، اشتراک، اجتماع، زیر مجموعه، متمم مجموعه، مکمل مجموعه، مجموعه مرجع، دو مجموعه جدا از هم، تقاضل دو مجموعه مفاهیم اصلی نظریه مجموعه ها هستند.

• خاصیتهای جابجایی، شرکت پذیری، توزیع پذیری و قوانین دمрган نیز از خواص مجموعه ها میباشد.

- 1)  $A \cup B = B \cup A$
- 2)  $A \cup (B \cap C) = (A \cup B) \cap C$       $A \cap (B \cup C) = (A \cap B) \cup C$
- 3)  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$       $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- 4)  $\left(\bigcup_{i=1}^n A_i\right)' = \bigcap_{i=1}^n A_i'$       $\left(\bigcap_{i=1}^n A_i\right)' = \bigcup_{i=1}^n A_i'$

## فصل دوم: نظریه مجموعه‌ها،

### آنالیز ترکیبی

### فضاي نمونه:

**آزمایش تصادفي:** فعاليتي که نتیجه آن از قبل مشخص نیست ولي کل حالات ممکن آن معلوم است، مثل پرتاب یک سکه، که معلوم نیست شیر خواهد آمد یا خط. چون در نظریه احتمال فقط آزمایشهاي تصادفي مورد نظر می باشد لذا برای سادگی کلمه تصادفي ذکر نمی گردد.

**پیشامد تصادفي:** پیش آمدي که در اثر یک آزمایش تصادفي می تواند رخ دهد یا رخ ندهد.

**فضاي نمونه:** مجموعه پیامدهاي ممکن یک آزمایش تصادفي را فضاي نمونه آن آزمایش می گویند که آن را با  $S$  نشان می دهند.

• تعداد اعضاي فضاي نمونه را با  $n(S)$  نشان می دهیم.

**فضاي نمونه محدود:** یعنی این تعداد اعضاي فضاي نمونه آزمایش متناهي باشد.

**فضاي نمونه نامحدود:** یعنی اینکه تعداد اعضاي فضاي نمونه آزمایش نامتناهي است.

• هر عضو فضاي نمونه را **نقطه نمونه** می نامیم.

• پیشامدي که شامل یک نقطه نمونه باشد را **پیشامد ساده** و اگر بیش از یک نقطه نمونه داشته باشد را **پیشامد مرکب** می نامیم.

• پیشامد  $A=S$  را **پیشامد حتمي** و پیشامد تهی را **پیشامد غیر ممکن** نامیم.

### آنالیز ترکیبي:

**کاربردهاي قواعد شمارش:** از این قواعد در وضعیت هايي استفاده می شود که فهرست نمودن تمام حالات ممکن آزمایش مقدور نمی باشد، یا نیازی به فهرست نمودن آنها نمی باشد، لذا فقط به ذکر تعداد حالات ممکن و مختلف اکتفا می شود.

**اصل اساسي شمارش:** اگر عملي به  $m_1$  طریق و برای هر کدام از آنها عمل دیگری را به  $m_2$  طریق و برای هر یک از این دو عمل سومی را به  $m_3$  طریق و ... و عمل  $K$  امی را به  $m_k$  طریق بتوان انجام داد، آنگاه  $m_1 \times m_2 \times \dots \times m_k$  عمل را با هم به طریق میتوان انجام داد.

**تمرین:** سکه را سه بار پرتاب می کنیم، فضاي نمونه آن را مشخص کنید؟

**تمرین:** یک سکه را با تاس همزمان پرتاب می کنیم فضاي نمونه آزمایش را مشخص کنید.

تمام ترتیب هاي ممکن دسته اي از اشیاء و یا قسمتي از آن را **تبدیل** یا **جایگشت** گوئیم.

**قضیه:** تعداد تبدیلهاي  $n$  شیئي متمایز برابر با  $n!$

**مثال:** به چند طریق می توان یک صف ۵ نفری برای سوار شدن به اتوبوس تشکیل داد

**مثال:** چند عدد سه رقمي از ارقام ۴، ۳، ۲، ۱، ۰ میتوان نوشت بطوریکه تکرار ارقام مجاز نباشد؟

**مثال ۱:** سه سکه را پرتاب می کنیم مطلوبست تعداد اعضاي فضاي نمونه؟

**مثال ۲:** دو تاس را با چهار سکه پرتاب می کنیم مطلوبست تعداد اعضاي فضاي نمونه؟

**مثال ۲:** چند عدد سه رقمي از ارقام ۴، ۳، ۲، ۱، ۰ میتوان نوشت بطوریکه تکرار ارقام مجاز باشد؟

**قضیه:** تعداد تبدیلیهای  $r$  تایی از  $n$  شیئی متمایز برابر است با

$$P_n^r = \frac{n!}{(n-r)!} = n(n-1)(n-2)\dots(n-r+1)$$

**قضیه:** تعداد تبدیلیهای دوری (تبدیلیهایی که بوسیله یک دسته از اشیا روی محیط دایره مرتب می‌شوند)،  $n$  شیئی متمایز برابر است با  $(n-1)!$

**قضیه:** تعداد تبدیلیهای مختلف  $n$  شیئی که  $n_1$  شیئی آن از نوع اول،  $n_2$  شیئی آن از نوع دوم و .... و  $n_r$  شیئی آن از نوع  $r$  باشد، برابر است با

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \dots n_r!}, \quad n_1 + n_2 + \dots + n_r = n$$

**مثال ۱:** به چند طریق میتوان ۳ لامپ قرمز، ۲ لامپ سبز و ۴ لامپ آبی را روی یک صفحه نصب نمود.

**مثال ۲:** به چند طریق ۵ نفر می‌توانند در اطراف یک میز دایره‌ای بنشینند؟

**مثال ۳:** ۶ زوج مختلف با هم دور یک میز دایره‌ای می‌نشینند تعداد حالت آن را در صورتیکه هر زوج کنار هم بنشینند را بیابید؟

**مثال ۴:** به چند طریق می‌توان ۷ نفر با یک وسیله نقلیه که به ترتیب گنجایش ۲، ۳ و ۷ نفر را دارند، از محلی به محل دیگر منتقل نمود؟

چگونگی انتخاب  $r$  شیئی از  $n$  شیئی بدون در نظر گرفتن ترتیب، **ترکیب** نامیده می‌شود.

**قضیه:** تعداد ترکیبهای  $r$  تایی از  $n$  شیئی متمایز برابر با:

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

مثال: از یک گروه مرکب از ۵ پزشک و ۳ پرستار، چند کمیته ۳ نفره می‌توان تشکیل داد؟

$$\binom{8}{3} = \frac{8!}{3!(5)!} = 56$$

**تمرین:** از ۱۰ دستگاه تلویزیون موجود در یک فروشگاه ۳ تلویزیون نقص فنی دارند تعداد حالات انتخاب ۴ دستگاه از این تلویزیون‌ها بطوریکه حداقل ۲ تلویزیون نقص فنی داشته باشند؟

**تمرین:** به چند طریق ۳ مرد و ۲ زن می‌توانند در یک صف قرار گیرند، اگر

• محدودیتی نداشته باشیم.

• زن‌ها کنار هم باشند.

• مردها کنار هم و زن‌ها کنار هم باشند.

**تمرین:** فرض کنید ۱۱ دوست صمیمی دارید، به چند طریق میتوان ۵ نفر از آن‌ها را به مهمانی دعوت کرد اگر

• محدودیتی نداشته باشیم.

• دو نفر از آنها بخواهند با هم در مهمانی شرکت کنند.

• دو نفر از آنها نمی‌خواهند با هم در مهمانی شرکت کنند.

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

در نظریه احتمالات به هر نقطه از فضای نمونه متناهی عددی نسبت داده می‌شود که «وزن» آن نقطه نامیده می‌شود.

مجموع وزن‌های تمام نقاط موجود در پیشامد  $A$  را «احتمال پیشامد  $A$ » گوئیم و آن را با  $P(A)$  نشان می‌دهیم. با توجه به تعریف پیشامد  $A$ ، داریم

- 1)  $P(\Phi) = 0$
- 2)  $P(S) = 1$
- 3)  $\forall A \subseteq S \Rightarrow 0 \leq P(A) \leq 1$

## فصل سوم: احتمال

اگر پیشامدهای  $A_1, A_2, \dots, A_n$  هر سه خاصیت ناسازگار، فرسا و همترازی را داشته باشند در نظریه احتمال آنها را **حالتها یا شانسیها** نامیده و می‌گوییم توسط **مدل کلاسیک** بیان می‌شوند.

اگر یک آزمایش را بتوان توسط مدل کلاسیک بیان نمود، در اینصورت احتمال هر پیشامد  $A$  در این آزمایش برابر است با

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

پیشامدهای  $A_1, A_2, \dots, A_n$  را **ناسازگار** می‌گوییم اگر وقوع همزمانی هر دو پیشامدی، غیر ممکن باشد.

$$A_i \cap A_j = \Phi \quad \forall i \neq j, 1 \leq i, j \leq n \Rightarrow \\ P(A_i \cap A_j) = 0$$

اگر پیشامدهای  $A_1, A_2, \dots, A_n$  همه نتایج ممکن در اثر آزمایش را در برگیرند، در اینصورت پیشامدها را **فرسا** می‌گوییم.

اگر شرایط آزمایش به گونه ای باشد که احتمال وقوع هر یک از پیشامدهای  $A_1, A_2, \dots, A_n$  برابر باشند، در اینصورت پیشامدها را **همتراز یا همشانس** می‌گوییم.

مثال: سکه سالمی را سه بار پرتاب می‌کنیم مطلوبت احتمال اینکه دقیقاً یکبار شیر بیاید

مثال: جعبه ای محتوی ۶ توپ آبی و ۱۲ توپ زرد می‌باشد، سه توپ بدون جایگذاری به تصادف برمی‌داریم مطلوبت:

- احتمال اینکه هر سه توپ آبی باشند
- دو توپ آبی و دیگری زرد باشد
- حداقل دو توپ آبی باشد.

اگر آزمایش تصادفی، یک آزمایش با فضای نمونه پیوسته باشد تعریف احتمال برای یک پیشامد دلخواه به صورت زیر در می‌آید.

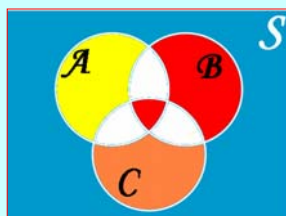
$$P(A) = \frac{L(A)}{L(S)} \longrightarrow \text{اگر آزمایش مربوط به طول باشد:}$$

$$P(A) = \frac{S(A)}{S(S)} \longrightarrow \text{اگر آزمایش مربوط به طول باشد:}$$

$$P(A) = \frac{V(A)}{V(S)} \longrightarrow \text{اگر آزمایش مربوط به حجم باشد:}$$

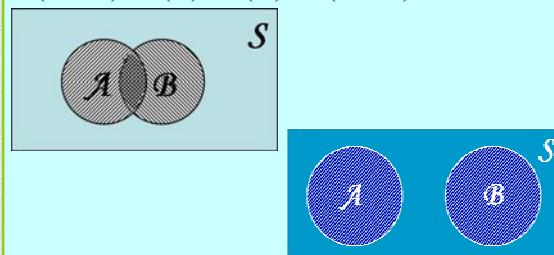
**قضیه:** اگر  $A$  و  $B$  و  $C$  سه پیشامد باشند آنگاه

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



**قضیه:** اگر  $A$  و  $B$  دو پیشامد باشند آنگاه

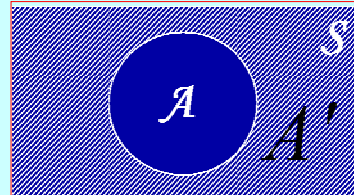
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



$$\text{if } A \cap B = \Phi \Rightarrow P(A \cup B) = P(A) + P(B)$$

**قضیه:** اگر  $A$  و  $A'$  دو پیشامد متمم باشند آنگاه

$$P(A') = 1 - P(A)$$



**تمرین:** ظرفی محتوی ۲۰ کارت در چهار رنگ مختلف است. بطوریکه از هر رنگ، ۵ کارت و کارتهای هم رنگ از ۱ تا ۵ شماره گذاری شده‌اند. ۲ کارت به تصادف بر می داریم. مطلوبست احتمال اینکه ۲ کارت دارای یک شماره باشند؟

**تمرین:** از ظرفی محتوی ۳ مهره سفید و ۴ مهره سیاه، ۳ مهره به تصادف خارج می کنیم، مطلوبست احتمال اینکه:

ا- ۳ مهره هم رنگ باشند

ب- ۳ مهره هم رنگ نباشند

اگر رخ دادن پیشامد  $A$ ، اثری در رخ دادن  $B$  نداشته باشد در اینصورت دو پیشامد را **مستقل** گوئیم.

دو پیشامد  $A$  و  $B$  را مستقل گوئیم، اگر و تنها اگر

$$P(A \cap B) = P(A)P(B)$$

پیشامدهای  $A_1, A_2, \dots, A_n$  مستقل گوئیم اگر و تنها اگر

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \forall i \neq j$$

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) \quad \forall i \neq j \neq k$$

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

**احتمال شرطی:** احتمال رخ دادن پیشامد  $B$ ، مشروط بر آنکه بدانیم  $A$  رخ داده است را احتمال  $B$  به شرط  $A$  گوئند و به صورت زیر تعریف می گردد.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B) \neq 0$$

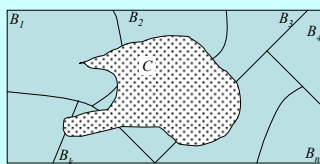
**قضیه:** اگر وقوع پیشامدهای  $A$  و  $B$  بطور همزمان امکان پذیر باشد، آنگاه

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

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

**قضیه (بیز):** فرض کنید مجموعه  $\{B_1, B_2, \dots, B_n\}$  یک افراز از فضای نمونه  $S$  باشد. اگر  $C$  پیشامدی دلخواه از  $S$  باشد و  $P(C)$  مخالف صفر باشد آنگاه به ازای  $K = 1, 2, \dots, n$  داریم:

$$P(B_k|C) = \frac{P(B_k \cap C)}{\sum_{i=1}^n P(B_i \cap C)} = \frac{P(B_k)P(C|B_k)}{\sum_{i=1}^n P(B_i)P(C|B_i)}$$



مثال: دو تاس را پرتاب می کنیم، در صورتیکه بدانیم مجموع اعداد ظاهر شده برابر ۶ است. مطلوبست احتمال اینکه یکی از تاسها عدد ۲ را نشان دهند؟

مثال: از جعبه ای محتوی ۴ مهره سفید و ۳ مهره سیاه، ۲ مهره را با جایگذاری خارج می کنیم (یک مهره را برداشته پس از مشاهده رنگ آن، مهره را به جعبه برمی گردانیم) مطلوبست محاسبه احتمال اینکه:

ا- هر دو مهره سفید باشد.

ب- مهره اول سفید و مهره دوم سیاه باشد.

مثال: کارخانه ای دارای سه ماشین است که ۵۰٪، ۳۰٪، و ۲۰٪ محصول آن کارخانه را تولید می کنند و میدانیم درصد کالاهای معیوب این سه ماشین به ترتیب ۳٪، ۴٪ و ۵٪ است. مطلوبست محاسبه احتمال اینکه اگر کالای انتخاب شده معیوب باشد، این کالا توسط ماشین اول تولید شده باشد:

**تمرین:** از شش زوج خواهر و برادر متفاوت، ۲ نفر را به تصادف انتخاب میکنیم، مطلوبست محاسبه احتمال اینکه این دو نفر الف: با هم خواهر و برادر باشند.  
ب: یکی مرد و یکی زن باشد.

**تمرین:** در کلاسی که ۱۵ دانشجو دارد در مورد موضوع خاصی، ۹ نفر موافق، ۴ نفر موافق و ۲ نفر ممتنع هستند، ۳ نفر را به تصادف انتخاب می کنیم و نظرشان را میپرسیم مطلوبست محاسبه احتمال اینکه:  
الف: حداقل دو نفر موافق باشند.  
ب: دو نفر اول موافق و نفر سوم مخالف باشد.

**تمرین:** از جعبه ای که محتوی ۵ مهره سفید و ۲ مهره سیاه است، یک مهره را به تصادف خارج کرده و بدون آنکه رنگ آن را مشاهده کنیم آنرا کنار میگذاریم و سپس مهره دیگری را خارج میکنیم، مطلوبست محاسبه احتمال اینکه این مهره سفید باشد.

**تمرین:** جعبه A محتوی ۶ توپ سفید و ۴ توپ سیاه است و جعبه B محتوی ۲ توپ سفید و ۲ توپ سیاه میباشد. از جعبه A، ۲ توپ به تصادف برداشته و در جعبه B میگذاریم، سپس دو توپ از جعبه B بدون جایگزینی خارج میکنیم. مطلوبست محاسبه احتمال اینکه فقط یکی از دو توپ سفید باشد.

**تمرین:** در یک مهمانی سه زوج ازدواج کرده باهم دور یک میز گرد می نشینند مطلوبست محاسبه احتمال اینکه  
الف: سه زن کنار هم نشسته باشند.  
ب: سه زوج کنار هم بنشینند.

**تمرین:** فرض کنید ۴ لامپ از ۱۲ لامپ موجود در جعبه ای سوخته باشند ۲ لامپ را به تصادف و بدون جایگزینی بر میداریم مطلوبست محاسبه احتمال اینکه:  
الف: هر دو لامپ سالم باشد.  
ب: حداقل یک لامپ سوخته باشد.

**تمرین:** سه ظرف کاملاً مشابه با محتویات زیر مفروض هستند. ظرفی را به تصادف انتخاب و مهره ای را از آن خارج میکنیم، اگر این مهره سفید باشد، مطلوبست محاسبه احتمال اینکه ظرف اول را انتخاب کرده باشیم.

ظرف سوم	ظرف دوم	ظرف اول
۵	۳	۲
مهره سفید		
۷	۵	۴
مهره سیاه		

## فصل چهارم: متغیرهای تصادفی

با فرض اینکه هر آزمایش تصادفی دارای فضای نمونه  $S$  باشد **متغیر تصادفی** تابعی است که به وسیله آن به هر نقطه از فضای نمونه ای یک عدد حقیقی نسبت میدهد پس **متغیر تصادفی**  $X$  تابعی است از فضای نمونه به مجموعه اعداد حقیقی:

$$X: S \rightarrow R$$

برای مثال دو سکه را همزمان پرتاب می کنیم،

$$S = \{HH, TH, HT, TT\}$$

حال میتوانیم  $X$  را یک متغیر تصادفی که نشان دهنده تعداد شیرها در آزمایش باشد در نظر بگیریم. که به ترتیب مقادیر ۰، ۱، ۲ را به اعضای فضای نمونه نسبت می دهد.

**توزیع احتمال گسسته:** یک متغیر تصادفی گسسته هر یک از مقادیر خود را با احتمالی معین اختیار می کند حال، **جدول** یا **فرمولی** که تمام مقادیر متغیر تصادفی را همراه با احتمالاتی مربوطه نشان دهد تابع احتمال نامیده و آن را با  $f(x)$ ،  $g(x)$  و ... نشان می دهیم.

در حالتی که متغیر تصادفی گسسته باشد، تابع احتمال را **توزیع احتمال** می نامیم.

نوشتن توزیع احتمال هم به صورت جدول و هم فرمول برای یک متغیر تصادفی ممکن است امکان پذیر نباشد و باید به یکی اکتفا کرد.

فضای نمونه  $S$  را **گسسته** گوئیم، اگر تعداد عضوهای آن متناهی یا نامتناهی شمارا باشد. مثل فضای نمونه پرتاب یک سکه یا پرتاب یک سکه تا شیر بیاید.

فضای نمونه  $S$  را **پیوسته** گوئیم، اگر تعداد عضوهای آن نامتناهی ناشمارا باشد مثل فضای نمونه آزمایش طول قد یا وزن افراد یک شهر.

اگر متغیر تصادفی  $X$  روی فضای نمونه گسسته تعریف گردد **متغیر تصادفی گسسته** و اگر روی فضای نمونه پیوسته تعریف گردد **متغیر تصادفی پیوسته** نامیم.

تابع  $f(x)$  را یک تابع توزیع احتمال برای متغیر تصادفی  $X$  گوئیم اگر:

- 1)  $\forall x_i, f(x_i) \geq 0$
- 2)  $\sum_i f(x_i) = 1$
- 3)  $P(X = x_i) = f(x_i)$

مثال: اگر  $f(x)$  یک تابع توزیع احتمال مطلوبیت مقدار  $k$ ؟

$$f(x) = \frac{k+1}{2^x} \quad x = 0, 1, 2, \dots$$

$$f(x) > 0 : 2^x > 0 \Rightarrow k+1 > 0 \Rightarrow k > -1$$

$$\sum f(x) = 1 \Rightarrow (k+1) \sum_{x=0}^{\infty} 2^{-x} = 1 \Rightarrow (k+1) \left(1 + \frac{1}{2} + \frac{1}{4} + \dots\right) = 1$$

$$\Rightarrow (k+1) \left(\frac{1}{1-\frac{1}{2}}\right) = 1 \Rightarrow 2k+2=1 \Rightarrow k = -\frac{1}{2}$$

**مثال:** سکه ای را سه مرتبه پرتاب می کنیم، اگر متغیر تصادفی  $X$  نشان دهنده شیر در این آزمایش باشد توزیع احتمال  $X$  را به صورت جدول و فرمول بیان می کنیم؟

$$S = \{ TTT, TTH, THT, HTT, THH, HTH, HHT, HHH \}$$

$X$	0	1	2	3
$f(x) = P(X=x)$	1/8	3/8	3/8	1/8

$$f(x) = \frac{n(A)}{n(S)} = \frac{\binom{3}{x}}{2^3}, \quad x = 0, 1, 2, 3$$

اگر  $X$  یک متغیر تصادفی گسسته با توزیع احتمال  $f(x)$  باشد تابع توزیع تجمعی  $X$  به شکل زیر نمایش و تعریف می شود.

$$F_x : R \longrightarrow R$$

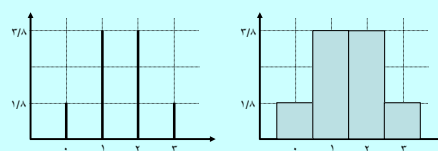
$$F_x(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

مثال: جدول زیر، توزیع احتمال متغیر تصادفی  $X$  را نشان می دهد، تابع توزیع تجمعی  $X$  را تعیین و نمودار آن را رسم کنید:

$X$	0	1	2	3
$f(x) = P(X=x)$	1/8	3/8	3/8	1/8

نمایش ترسیمی توزیع احتمال معمولاً درک مطلب را ساده تر می نماید دو نمودار **میله ای** و **هیستوگرام** برای نمایش توزیع احتمال به صورت نمودار به کار می رود. که برای مثال نمودار میله ای و هیستوگرام پرتاب سه سکه را رسم می کنیم.

$X$	0	1	2	3
$f(x) = P(X=x)$	1/8	3/8	3/8	1/8



تابع توزیع تجمعی دارای خواص زیر می‌باشد:

$$(1) \text{ بازای هر } x \text{ } 0 \leq F_x(x) \leq 1$$

(۲)  $F(x)$  تابعی غیر نزولی است.

$$(3) F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, \quad F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$$

(۴) تابع  $F(x)$  در تمام نقاط از سمت راست پیوسته است.

(۵) به ازای  $a$  و  $b$  ثابت داریم

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

$$P(X = a) = F(a) - F(a^-)$$

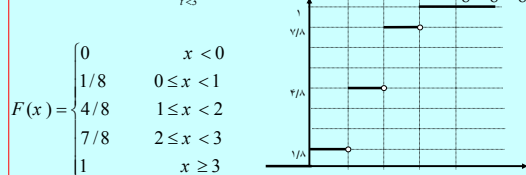
$$x < 0 \Rightarrow F(x) = 0$$

$$0 \leq x < 1 \Rightarrow F(x) = \sum_{t < x} f(t) = f(0) = \frac{1}{8}$$

$$1 \leq x < 2 \Rightarrow F(x) = \sum_{t < x} f(t) = f(0) + f(1) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$$

$$2 \leq x < 3 \Rightarrow F(x) = \sum_{t < x} f(t) = f(0) + f(1) + f(2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

$$3 \leq x < 4 \Rightarrow F(x) = \sum_{t < x} f(t) = f(0) + f(1) + f(2) + f(3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$



همانطور که دیدیم به تابع احتمال یک متغیر تصادفی گسسته تابع توزیع احتمال گفتیم حال اگر متغیر تصادفی پیوسته باشد به تابع احتمال آن **تابع چگالی احتمال** می‌گوییم و فرق اساسی آن با حالت گسسته در این است که احتمال اینکه متغیر تصادفی پیوسته دقیقاً یکی از مقادیر خود را اختیار کند برابر صفر می‌باشد عبارت دیگر اگر  $X$  یک متغیر تصادفی پیوسته باشد آنگاه

$$P(X = a) = 0$$

یعنی در تابع چگالی بحث روی بازه هست نه نقطه.

**تمرین:** یک جفت تاس را پرتاب می‌کنیم، اگر متغیر تصادفی  $X$  نشان دهنده مجموع اعداد ظاهر شده باشد، مطلوبست تعیین:

• توزیع احتمال  $X$

• نمودار توزیع احتمال

• تابع توزیع تجمعی

• نمودار توزیع تجمعی  $X$

• احتمالهای زیر:

$$P(2 < X \leq 5), \quad P(2 \leq X \leq 5)$$

$$P(X \geq 4), \quad P(X < 5)$$

مثال: در تابع زیر پارامتر  $a$  را طوری بیابید که تابع  $f(x)$  یک تابع چگالی احتمال باشد:

$$f(x) = \frac{1}{a(1+x^2)}, \quad x \geq 0$$

$$\frac{1}{a} \int_0^{+\infty} \frac{dx}{1+x^2} = 1 \Rightarrow \frac{1}{a} \tan^{-1} x \Big|_0^{+\infty} = 1 \Rightarrow$$

$$\frac{1}{a} (\tan^{-1} x - \tan^{-1} 0) = 1 \Rightarrow \frac{1}{a} \left( \frac{\pi}{2} - 0 \right) = 1 \Rightarrow$$

$$\frac{\pi}{2a} = 1 \Rightarrow a = \frac{\pi}{2}$$

تمامی خواصی که برای تابع توزیع احتمال گفته شد در تابع چگالی احتمال نیز صادق است که به آنها اشاره می‌کنیم.

تابع  $f(x)$  را یک **تابع چگالی** می‌گوییم هرگاه:

$$1) \forall x \in R, f(x) \geq 0$$

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

$$3) P(a \leq X \leq b) = P(a < X < b) = \int_a^b f(x) dx$$

تمامی خواص تابع توزیع تجمعی در حالت گسسته در حالت پیوسته نیز برقرار است:

$$0 \leq F_x(x) \leq 1: x \text{ برای هر } (1)$$

$F(x)$  تابعی غیر نزولی است. (2)

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, \quad F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1 \quad (3)$$

اگر  $X$  یک متغیر تصادفی پیوسته با تابع چگالی احتمال  $f(x)$  باشد، تابع توزیع تجمعی  $X$  را با  $F(x)$  نشان داده و به صورت زیر تعریف می‌کنیم:

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

با توجه به تعریف بالا داریم:

$$P(a < X < b) = P(X < b) - P(X < a) = F(b) - F(a)$$

از تعریف انتگرال معین نیز میتوان نوشت:

$$P(a < X < b) = \int_a^b f(x) dx = F(b) - F(a) \Rightarrow F'(x) = f(x)$$

### احتمال

تا حالا با فضاهای نمونه یک بعدی و متغیرهای تصادفی مربوط به این فضاها آشنا شدیم ولی آزمایشهای زیادی وجود دارند که بطور همزمان دو یا چندین نتیجه خواهند داشت.

بعنوان مثال اگر یک اسید را با فلزی ترکیب کنیم بطور همزمان میخواهیم مقدار رسوب حاصل و مقدار گاز متصاعد شده را مورد بررسی کنیم.

یا اگر کیسه محتوی ۵ مهره قرمز ۴ مهره آبی و ۳ مهره زرد باشد ۳ توپ به تصادف خارج کنیم  $X$  میتواند تعداد توپ قرمز،  $Y$  تعداد توپ آبی و  $Z$  تعداد توپ زرد باشد که هر کدام از این متغیرها می‌توانند اعداد ۰، ۱، ۲، ۳ را بگیرند.

**تمرین:** اگر نقطه ای به تصادف در داخل دایره ای به شعاع  $r$  انتخاب کنیم و اگر متغیر تصادفی  $X$  نشان دهنده فاصله این نقطه از مرکز دایره باشد مطلوبست محاسبه  $F(x)$  و  $f(x)$  و  $P(a < X < b)$  ( $a$  و  $b$  داخل دایره اند).

**تمرین:** اگر متغیر تصادفی  $X$  دارای تابع چگالی احتمال زیر باشد مقدار  $a$  و  $F(x)$  را بیابید.

$$f(x) = \begin{cases} \frac{a}{x^2}, & x > 1 \\ 0, & x < 1 \end{cases}$$

### خصوصیات تابع احتمال:

- 1)  $f(x, y) \geq 0, \quad \forall (x, y)$  اگر  $X$  و  $Y$  گسسته باشند.  $\left\langle \sum_x \sum_y f(x, y) = 1 \right\rangle$
- 2) اگر  $X$  و  $Y$  پیوسته باشند.  $\left\langle \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dy dx = 1 \right\rangle$
- 3) اگر  $X$  و  $Y$  گسسته باشند.  $\left\langle P[(X, Y) \in A] = \sum_A f(x, y) \right\rangle$
- اگر  $X$  و  $Y$  پیوسته باشند  $\left\langle P[(X, Y) \in A] = \int_A f(x, y) dy dx \right\rangle$

### احتمال

اگر  $X$  و  $Y$  دو متغیر تصادفی باشند، احتمال وقوع همزمانی آنها با علامت تابعی  $f(x, y)$  نشان داده و آن را **تابع احتمال  $X$  و  $Y$**  می‌نامیم.

در حالتی که  $X$  و  $Y$  هر دو **گسسته** باشند  $f(x, y)$  را **توزیع احتمال** و در حالتی که  $X$  و  $Y$  هر دو **پیوسته** باشند  $f(x, y)$  را **تابع چگالی احتمال** می‌نامیم.

اگر  $X$  و  $Y$  دو متغیر تصادفی با توزیع احتمال توأم (تابع چگالی توأم)،  $f(x,y)$  باشند، توزیع احتمال هر یک از متغیرهای تصادفی  $X$  و  $Y$  به صورت زیر نمایش و محاسبه می کنیم:

$$g(x) = \sum_y f(x,y) \quad \leftarrow \text{در حالت گسسته}$$

$$h(y) = \sum_x f(x,y)$$

$$g(x) = \int_{-\infty}^{+\infty} f(x,y) dy \quad \leftarrow \text{در حالت پیوسته}$$

$$h(y) = \int_{-\infty}^{+\infty} f(x,y) dx$$

$g(x)$  و  $h(y)$  را **توزیع های حاشیه ای**  $X$  و  $Y$  می نامیم.

حال این سوال مطرح است اگر توزیع احتمال توأم دو یا چند متغیر تصادفی معلوم باشد آیا می توان توزیع هر کدام از این متغیرها را به صورت جداگانه حساب کرد یا بالعکس، اگر توزیع های چند متغیر معلوم باشد آیا امکان محاسبه توزیع احتمال توأم آنها (در صورت امکان وقوع همزمانشان) وجود دارد.

جواب: پاسخ قسمت اول همواره مثبت است ولی قسمت دوم فقط در مواردی امکان پذیر است.

ادامه مثال:

X \ Y	0	1	2	$h(y)$
0	1/21	6/21	3/21	10/21
1	4/21	6/21	0	10/21
2	1/21	0	0	1/21
$g(x)$	6/21	12/21	3/21	

$$g(x) = \sum_{y=1}^2 f(x,y) = f(x,0) + f(x,1) + f(x,2) \quad , x = 0,1,2$$

$$h(y) = \sum_{x=1}^2 f(x,y) = f(0,y) + f(1,y) + f(2,y) \quad , y = 0,1,2$$

**مثال:** شیشه ای حاوی ۳ قرص آسپرین، ۲ قرص خواب آور و ۲ قرص مسکن مفروض است. شخصی به تصادف دو قرص از این ظرف خارج می کند. اگر فرض کنیم  $X$  معرف قرص آسپرین و  $Y$  معرف قرص خواب آور باشد. مطلوب است:

• توزیع احتمال توأم را برای دو متغیر تصادفی  $X$  و  $Y$  هم به صورت جدول و هم به صورت فرمول.

• توزیع های حاشیه ای  $X$  و  $Y$

$$f(x,y) = P(X=x, Y=y) = \frac{\binom{3}{x} \binom{2}{y} \binom{2}{2-x-y}}{\binom{7}{2}} \quad x=0,1,2, y=0,1,2$$

اگر  $X$  و  $Y$  دو متغیر تصادفی با تابع احتمال توأم  $f(x,y)$  و توزیع های حاشیه ای  $g(x)$  و  $h(y)$  باشند، **تابع احتمال شرطی** متغیر تصادفی  $Y$  در صورتیکه  $X=x$  داده شده باشد، و **تابع احتمال شرطی** متغیر تصادفی  $X$  در صورتیکه  $Y=y$  داده شده باشد، و عبارتست از:

$$k(x|y) = \frac{f(x,y)}{h(y)} \quad , h(y) > 0$$

$$k(y|x) = \frac{f(x,y)}{g(x)} \quad , g(x) > 0$$

**تمرین:** تابع چگالی احتمال توأم دو متغیر تصادفی  $X$  و  $Y$  به صورت زیر تعریف شده است. مطلوب است:

• تعیین مقدار  $K$

• توزیع های حاشیه ای  $X$  و  $Y$

• محاسبه  $P(X>3, Y<2)$

$$f(x,y) = \begin{cases} Kxy & 0 < x < 4, 1 < y < 5 \\ 0 & \text{other points} \end{cases}$$

اگر  $X$  و  $Y$  دو متغیر تصادفی با تابع احتمال توأم  $f(x,y)$  و توزیع‌های حاشیه‌ای  $g(x)$  و  $h(y)$  باشند،  $X$  و  $Y$  را از لحاظ آماری **مستقل** گوئیم **اگر و تنها اگر** برای هر  $(x,y)$  داشته باشیم:

$$f(x,y) = g(x)h(y)$$

پس فقط در حالت استقلال دو متغیر تصادفی است که می‌توان از توزیع‌های حاشیه‌ای به تابع توزیع احتمال توأم رسید.

**مثال:** اگر  $X$  و  $Y$  دارای تابع چگالی احتمال توأم زیر باشند مطلوبست محاسبه  $k(x|y)$  و  $k(y|x)$

$$f(x,y) = \begin{cases} 8xy & , 0 \leq x \leq 1, 0 \leq y \leq x \\ 0 & \text{other points} \end{cases}$$

$$g(x) = \int_{-\infty}^{+\infty} f(x,y) dy = \int_0^x 8xy dy = \begin{cases} 4x^2 & , 0 \leq x \leq 1 \\ 0 & \text{other points} \end{cases}$$

$$h(y) = \int_{-\infty}^{+\infty} f(x,y) dx = \int_y^1 8xy dx = \begin{cases} 4y(1-y^2) & , 0 \leq y \leq 1 \\ 0 & \text{other points} \end{cases}$$

$$k(y|x) = \frac{f(x,y)}{g(x)} = \frac{8xy}{4x^2} = \begin{cases} \frac{2y}{x} & , 0 \leq y \leq x \\ 0 & \text{other points} \end{cases}$$

$$k(x|y) = \frac{f(x,y)}{h(y)} = \frac{8xy}{4y(1-y^2)} = \begin{cases} \frac{2x}{1-y^2} & , y \leq x \leq 1 \\ 0 & \text{other points} \end{cases}$$

## فصل پنجم: امید ریاضی و گشتاورها

**تمرین:** تابع چگالی احتمال توأم دو متغیر تصادفی  $X$  و  $Y$  به صورت زیر مفروض است مطلوبست:

$$f(x,y) = \begin{cases} K(x^2 + y^2) & , 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{other points} \end{cases}$$

• مقدار ثابت  $K$

• توزیع‌های حاشیه‌ای  $X$  و  $Y$

•  $K(x|y)$  و  $K(y|x)$

• بررسی استقلال دو متغیر تصادفی

• محاسبه  $P(0 < X < 1/2 | Y = 1/2)$  و  $P(X < 1/2, Y > 1/2)$

**مثال:** متغیر تصادفی  $X$  با تابع چگالی احتمال زیر مفروض است امید ریاضی آن را به دست آورید؟

$$f(x) = \begin{cases} \frac{x}{8} & , 0 < x < 4 \\ 0 & , \text{other points} \end{cases}$$

$$E(X) = \int_{-\infty}^{+\infty} xf(x) dx = \int_0^4 x \frac{x}{8} dx = \frac{1}{24} x^3 \Big|_0^4 = \frac{8}{3}$$

**امید ریاضی یا میانگین یا متوسط** یک متغیر تصادفی مانند  $X$ ، پارامتری است که نشان دهنده مقدار مورد انتظار برای آن متغیر تصادفی در اثر تکرار آن آزمایش به دفعات زیاد، می‌باشد که به صورت زیر محاسبه می‌گردد:

$$E(X) = \sum_x xf(x) \longleftarrow \text{اگر } X \text{ گسسته باشد}$$

$$= \int_{-\infty}^{+\infty} xf(x) dx \longleftarrow \text{اگر } X \text{ پیوسته باشد}$$

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

اگر  $X$  و  $Y$  هر دو گسسته باشد

$$E(h(x,y)) = \sum_x h(x,y) f(x,y)$$

اگر  $X$  و  $Y$  هر دو پیوسته باشد

$$E(h(x,y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y) f(x,y) dx dy$$

**قضیه:** اگر متغیر تصادفی  $X$  دارای تابع احتمال  $f(x)$  باشد، امید ریاضی هر تابعی از  $X$  مانند  $h(x)$  عبارتست از:

اگر  $X$  گسسته باشد  $\leftarrow E(h(x)) = \sum_x h(x) f(x)$

اگر  $X$  پیوسته باشد  $\leftarrow = \int_{-\infty}^{+\infty} h(x) f(x) dx$

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

$$\| E(b) = b \|$$

$$\| E(aX + b) = aE(X) + b \|$$

$$\| E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)] \|$$

$$\| E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)] \|$$

$$\| E(XY) = E(X)E(Y) \text{ اگر } X \text{ و } Y \text{ مستقل باشند} \|$$

**مثال:** فرض کنید  $X$  و  $Y$  دو متغیر تصادفی مستقل با توابع چگالی زیر باشند مطلوبست امید ریاضی  $h(x,y)=xy$

$$g(x) = \begin{cases} \frac{8}{x^3}, & x > 2 \\ 0, & \text{other points} \end{cases}, \quad h(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{other points} \end{cases}$$

چون دو متغیر تصادفی مستقلند پس چگالی توأم آنها برابر است با حاصل ضرب چگالی های حاشیه ای:

$$f(x,y) = \begin{cases} \frac{16y}{x^3}, & x > 2, 0 < y < 1 \\ 0, & \text{other points} \end{cases}$$

$$E(xy) = \int_2^{\infty} \int_0^1 xy \frac{16y}{x^3} dx dy = \frac{8}{3}$$

**گشتاورها:** اگر  $X$  یک متغیر تصادفی با تابع احتمال  $f(x)$  باشد گشتاورهای مختلف آن را به صورت زیر نشان داده و تعریف می‌کنیم:

گشتاور مرتبه  $r$ ام پیرامون نقطه  $a$   $\leftarrow \mu_r^a = E(X - a)^r$

گشتاور مرتبه  $r$ ام پیرامون مبدأ  $\leftarrow \mu_r^0 = \mu_r' = E(X^r)$  if  $(a=0)$

گشتاور مرکزی مرتبه  $r$ ام  $\leftarrow \mu_r^\mu = \mu_r = E(X - \mu)^r$  if  $(a = \mu)$

واریانس  $X$   $\leftarrow \mu_2 = V(X) = \sigma_X^2 = E(X - \mu)^2$  if  $\begin{cases} a = \mu \\ r = 2 \end{cases}$

**تمرین:** اگر متغیر تصادفی  $X$  دارای توزیع احتمال زیر باشد مطلوبست محاسبه

$x$	-2	-1	1	2
$f(x)$	1/4	1/4	1/4	1/4

$$E(3x^2 - 4x + 1/2)$$

اگر  $Y = X^2$  توزیع احتمال  $Y$  و  $E(Y)$

**تمرین:** اگر متغیر تصادفی  $X$  دارای تابع چگالی احتمال زیر باشد مطلوبست محاسبه  $E(X)$

$$f(x) = \begin{cases} 2e^{-2x}, & x > 0 \\ 0, & \text{other point} \end{cases}$$

جذر مثبت واریانس را **انحراف معیار** گوئیم.

واریانس متغیر تصادفی به صورت زیر راحتتر محاسبه می‌گردد.

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

**کواریانس**  $X$  و  $Y$  را به صورت زیر تعریف می‌کنیم

$$\sigma_{XY} = \text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

کواریانس دو متغیر تصادفی به صورت زیر راحتتر محاسبه می‌گردد.

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y$$

• برای بررسی چگونگی **همبستگی خطی** بین دو متغیر تصادفی  $X$  و  $Y$  از کواریانس آنها استفاده می‌کنیم اگر تغییرات  $X$  و  $Y$  همسو باشند **کواریانس مثبت**، فاقد تغییرات مشترک باشند **کواریانس صفر** و تغییرات مخالف یکدیگر باشند **کواریانس منفی** خواهد بود.

• اگر  $X$  و  $Y$  مستقل باشند کواریانس آنها صفر ولی عکس مطلب صحیح نیست.

اگر  $X$  و  $Y$  دو متغیر تصادفی باشند. **ضریب همبستگی خطی** بین  $X$  و  $Y$  به صورت زیر نشان داده و تعریف می‌کنیم:

$$\rho = \rho(X, Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

• برای بررسی چگونگی همبستگی خطی بین دو متغیر تصادفی  $X$  و  $Y$  استفاده می‌کنیم.

• برتری ضریب همبستگی به کواریانس در آنست که ضریب همبستگی بستگی به واحد اندازه گیری ندارد

• اگر  $X$  و  $Y$  مستقل باشند ضریب همبستگی بین آنها صفر است.

• ضریب همبستگی همواره در فاصله  $[-1, 1]$  قرار دارد.

• کواریانس و ضریب همبستگی همیشه هم علامتند.

**مثال:** برای جدول توزیع احتمال زیر واریانس  $X$  و  $Y$  و کواریانس آنها را پیدا کرده و تفسیر نمایید:

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \sum_{x=0}^2 x^2 g(x) - \left[ \sum_{x=0}^2 xg(x) \right]^2 = (0 \times 6/21 + 1 \times 12/21 + 4 \times 3/21) - (0 \times 6/21 + 1 \times 12/21 + 2 \times 3/21)^2 = 20/49$$

$$\sigma_Y^2 = E(Y^2) - \mu_Y^2$$

$$\sigma_{XY} = \text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y =$$

$$E(XY) = \sum_y \sum_x xyf(x, y) = \frac{6}{21} = \frac{2}{7}$$

$$E(X) = \sum_x xg(x) = \frac{12}{21} + \frac{6}{21} = \frac{6}{7}$$

$$E(Y) = \sum_y yh(y) = \frac{10}{21} + \frac{2}{21} = \frac{4}{7}$$

$$\sigma_{XY} = \frac{2}{7} - \frac{6}{7} \times \frac{4}{7} = \frac{-10}{49} < 0$$

X \ Y	0	1	2	h(y)
0	1/21	6/21	3/21	10/21
1	4/21	6/21	0	10/21
2	1/21	0	0	1/21
g(x)	6/21	12/21	3/21	

**خواص واریانس و کواریانس و ضریب همبستگی:** اگر  $a, b, c$  و  $d$  اعداد ثابت و  $X$  و  $Y$  دو متغیر تصادفی با توابع احتمال حاشیه‌ای  $g(x)$  و  $h(y)$  و تابع احتمال توأم  $f(x, y)$ ، و  $z(h)$  تابعی دلخواه از  $X$  باشد داریم:

$$1) \sigma_{h(x)}^2 = E\left\{ [h(x) - \mu_{h(x)}]^2 \right\} \quad 2) \sigma_b^2 = 0$$

$$3) \sigma_{ax+b}^2 = a^2 \sigma_x^2 \quad 4) \sigma_{ax+by+c}^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2ab \sigma_{xy}$$

$$5) \text{Cov}(aX + b, dY + c) = ad \text{Cov}(X, Y)$$

$$6) \rho(aX + b, dY + c) = \rho(X, Y)$$

**مثال:** اگر  $X$  و  $Y$  دو متغیر تصادفی با توزیع احتمال زیر باشند، ضریب همبستگی بین آنها را حساب نمایید؟

$$f(x, y) = \begin{cases} \frac{1}{18}(x+2y), & x=1,2, y=1,2 \\ 0, & \text{other points} \end{cases}$$

$$\mu_X = \frac{14}{9}, \mu_Y = \frac{29}{18}, \sigma_X^2 = \frac{14}{9}, \sigma_Y^2 = \frac{29}{18}$$

$$E(XY) = \sum_y \sum_x xy \frac{x+2y}{18} = \frac{45}{18}$$

$$\sigma_{xy} = \frac{45}{18} - \left( \frac{14}{9} \times \frac{29}{18} \right) =$$

$$\rho = \rho(X, Y) = \frac{\sigma_{xy}}{\sigma_X \sigma_Y} = \frac{-\frac{1}{162}}{\sqrt{\frac{20}{81} \times \frac{77}{324}}} = -0.025$$

**تمرین:** اگر  $X$  و  $Y$  دو متغیر تصادفی با تابع چگالی احتمال توأم زیر باشند. مطلوبست

$$f(x, y) = \begin{cases} x + y & , 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & , \text{other points} \end{cases}$$

• ضریب همبستگی بین  $X$  و  $Y$

• محاسبه  $E(Y | X=x)$

• محاسبه  $V(Y | X=x)$

• محاسبه  $V(4X+3Y-2)$

اگر  $X$  و  $Y$  دو متغیر تصادفی با تابع احتمال توأم  $f(x, y)$  باشند امید ریاضی  $Y$  به شرط آنکه  $X=x$  باشد، عبارتست از:

$$E(Y | X=x) = \int_{-\infty}^{+\infty} y k(y | x) dy \quad \text{اگر } X \text{ و } Y \text{ پیوسته باشند:}$$

$$= \sum_y y k(y | x) \quad \text{اگر } X \text{ و } Y \text{ گسسته باشند:}$$

**گشتاور مشروط  $Y$  بشرط آنکه  $X=x$  باشد** حول نقطه دلخواه  $a$  عبارتست از:

$$E((Y-a)^r | X=x) = \int_{-\infty}^{+\infty} (y-a)^r k(y | x) dy \quad \text{اگر } X \text{ و } Y \text{ پیوسته باشند:}$$

$$= \sum_y (y-a)^r k(y | x) \quad \text{اگر } X \text{ و } Y \text{ گسسته باشند:}$$

**قضیه:** اگر  $X$  یک متغیر تصادفی با تابع مولد گشتاور  $M(t)$  باشد آنگاه:

$$\left. \frac{d^r M(t)}{dt^r} \right|_{t=0} = \mu_r'$$

**قضیه:** اگر  $X$  و  $Y$  دو متغیر تصادفی با توابع مولد گشتاور  $M_x(t)$  و  $M_y(t)$  باشند، آنگاه  $X$  و  $Y$  دارای توابع توزیع یکسان هستند اگر و فقط اگر  $M_x(t) = M_y(t)$

**قضیه:** اگر  $X$  و  $Y$  دو متغیر تصادفی مستقل با توابع مولد گشتاور  $M_x(t)$  و  $M_y(t)$  باشند، آنگاه  $M_{x+y}(t) = M_x(t)M_y(t)$

**تابع مولد گشتاور:** تابع مولد گشتاور  $X$  با تابع احتمال  $f(x)$  را با نماد  $M_x(t)$  یا  $M(t)$  نشان داده و به صورت زیر تعریف می‌کنیم:

$$M_x(t) = E(e^{tx}), \quad t \in R$$

$$= \begin{cases} \sum_x e^{tx} f(x) & \text{اگر } X \text{ گسسته باشند:} \\ \int_{-\infty}^{+\infty} e^{tx} f(x) dx & \text{اگر } X \text{ پیوسته باشند:} \end{cases}$$

یکی از کاربردهای این تابع محاسبه میانگین و واریانس متغیرهای تصادفی است.

**تمرین:** اگر متغیر تصادفی  $X$  دارای تابع چگالی احتمال زیر باشد، مطلوبست محاسبه  $M(t)$  و میانگین و واریانس  $X$  روی آن؟

$$f(x) = \begin{cases} xe^{-x} & , x \geq 0 \\ 0 & , \text{other points} \end{cases}$$

**مثال:** اگر متغیر تصادفی  $X$  دارای توزیع احتمال زیر باشد مطلوبست تعیین  $M(t)$ ، میانگین و واریانس  $X$ :

$$f(x) = 2\left(\frac{1}{3}\right)^x, \quad x = 1, 2, 3, \dots$$

$$M(t) = E(e^{tx}) = \sum_{x=1}^{\infty} 2\left(\frac{1}{3}\right)^x e^{tx} = 2 \sum_{x=1}^{\infty} \left(\frac{e^t}{3}\right)^x$$

$$\text{if } \frac{e^t}{3} < 1 \Rightarrow M(t) = 2 \frac{e^t}{3} \times \frac{1}{1 - e^t/3} = \frac{2e^t}{3 - e^t}$$

$$\mu_x = \mu_1' = M'(t)|_{t=0}$$

$$\sigma_x^2 = \mu_2' - (\mu_1')^2 = \left[ M''(t) - (M'(t))^2 \right]_{t=0}$$

### قضیه چیبیشف:

طبق این قضیه اطلاعاتی از میزان پراکندگی داده ها در اطراف میانگین بدست خواهیم آورد. اگر متغیر تصادفی  $X$  دارای واریانس کوچکی باشد انتظار می رود که مشاهدات در اطراف میانگین متمرکزتر باشند.

**قضیه:** اگر  $X$  یک متغیر تصادفی و  $h(X)$  یک تابع نامنفی از  $X$  باشد آنگاه:

$$P(h(X) \geq c) \leq \frac{1}{c} E(h(X)), \quad c > 0$$

**قضیه چیبیشف:** اگر  $X$  یک متغیر تصادفی با واریانس محدود باشد آنگاه:

$$1) P[|X - \mu| \geq k\sigma] = P[(X - \mu)^2 \geq k^2\sigma^2] \leq \frac{1}{k^2}, \quad k > 0$$

$$P[|X - \mu| < k\sigma] \geq 1 - \frac{1}{k^2}$$

**مثال:** اگر متغیر تصادفی  $X$  دارای میانگین ۲۵ و واریانس ۱۶ باشد، با استفاده از قضیه چیبیشف احتمال زیر را تعیین کنید؟

$$P(17 < X < 33) = P(-8 < X - 25 < 8) = P(|X - 25| < 2 \times 4) \geq 1 - \frac{1}{4} = 3/4$$

$$P(|X - 25| \geq 12) = P(|X - 25| \geq 3 \times 4) \leq 1/9$$

### فصل ششم: بررسی چند توزیع متغیرهای تصادفی گسسته

میدانیم هر آزمایش تصادفی فضایی نمونه ای را ایجاد خواهد که می توان متغیر تصادفی را روی این فضا تعریف نمود. همچنین هر متغیر تصادفی دارای تابع احتمالی است که میتوان آن را مشخص نمود حال این سؤال مطرح است:

آیا برای هر آزمایش آماری باید یک تابع احتمال تعیین کرد یا اینکه می توان آزمایشهای آماری را به دسته های مختلفی را تقسیم بندی کرد که هر دسته دارای خواص مشترکی باشند؟

دو فصل آینده سعی در دسته بندی آزمایشهای آماری است به طوری که مطالعه این آزمایشات را اصولی تر کند.

**مثال:** جعبه ای شامل صفحه کلید با شماره های ۱ تا  $k$  است. که شانس انتخاب شماره های این صفحه کلید یکسان می باشد. اگر شماره ای از این صفحه کلید به تصادف انتخاب شود و  $X$  را شماره کلید انتخابی در نظر بگیریم آنگاه  $X$  دارای تابع **توزیع احتمال یکنواخت** است

$$f(x_i; k) = \frac{1}{k}, \quad X = 1, 2, \dots, k$$

$$\mu = \frac{1}{k} \sum_{i=1}^k x_i = \frac{1}{k} \times \frac{k(k+1)}{2} = \frac{(k+1)}{2}$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = E(X^2) - \mu^2 = \frac{(k+1)(2k+1)}{6} - \frac{(k+1)^2}{4}$$

$$\sigma^2 = \frac{(k+1)(k-1)}{12}$$

$$E(X^2) = \frac{1}{k} \sum_{i=1}^k x_i^2 = \frac{1}{k} \frac{k(k+1)(2k+1)}{6} = \frac{(k+1)(2k+1)}{6}$$

**توزیع یکنواخت:** اگر متغیر تصادفی  $X$  مقادیر  $x_1, x_2, \dots, x_n$  را با احتمال مساوی قبول کند، توزیع احتمال این متغیر تصادفی را توزیع یکنواخت می نامیم و آن را به صورت زیر نمایش و تعریف می کنیم.

$$f(x_i; n) = \frac{1}{n}, \quad X = x_1, x_2, \dots, x_n$$

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

**مثال:** احتمال موفقیت برای يك داوطلب در جهت اخذ گواهینامه ۷۰٪ می باشد، توزیع احتمال برای این شخص که در آزمون رانندگی شرکت می کند بنویسید .

$$f(x) = \begin{cases} (0.7)^x (0.3)^{1-x}, & x = 0, 1 \\ 0, & \text{other points} \end{cases}$$

$$\mu = p = 0.7$$

$$\sigma^2 = pq = 0.7 \times 0.3 = 0.21$$

$$M(t) = E(e^{tx}) = \sum_x e^{tx} (p)^x (q)^{1-x} = q + e^t p = 0.3 + 0.7e^t$$

**توزیع برنولی:** اگر نتیجه آزمایشی فقط دو وضعیت را معرفی کند (پیروزی یا شکست) آن آزمایش را آزمایش برنولی می‌گویند. توزیع منتسب به این آزمایش، **توزیع برنولی** نام دارد. اگر احتمال موفقیت P و احتمال شکست q=1-P فرض شود. در این صورت توزیع مورد نظر به صورت زیر خواهد بود:

$$f(x) = P(X=x) = \begin{cases} p^x q^{1-x}, & x = 0, 1 \\ 0, & \text{other points} \end{cases}$$

$$\mu = E(X) = p$$

$$\sigma^2 = E(X^2) - (E(X))^2 = pq = p(1-p)$$

**مثال:** يك فوتبالبست با احتمال ۸۰٪ تویی را با موفقیت به درون دروازه می فرستد . این فوتبالبست ۶ توپ را در اختیار دارد . مطلوبست:

• معرفی توزیع احتمال مربوطه

• احتمال اینکه درست ۲ توپ را به درون دروازه بفرستد .

• احتمال اینکه حداکثر ۲ توپ را به درون دروازه بفرستد .

• احتمال اینکه حداقل ۲ توپ را به درون دروازه بفرستد .

$$f(x) = P(X=x) = \binom{6}{x} (0.8)^x (0.2)^{6-x}$$

$$f(2) = P(X=2) = \binom{6}{2} (0.8)^2 (0.2)^4 = 15 \times 0.64 \times 0.0016$$

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$P(X \geq 2) = P(X=2) + \dots + P(X=6) = 1 - P(X < 2) =$$

$$1 - [P(X=0) + P(X=1)]$$

**توزیع دو جمله‌ای:** اگر يك آزمایش برنولی n بار به صورت مستقل تکرار شود و در این آزمایش متغیر تصادفی X نشان دهنده تعداد پیروزی باشد، توزیع منتسب به این آزمایش، **توزیع دو جمله‌ای** نام دارد و به این شکل تعریف و نمایش داده می‌شود:

$$f(x) = b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \dots, n$$

$$M(t) = E(e^{tx}) = \sum_x e^{tx} \binom{n}{x} p^x q^{n-x} =$$

$$\sum_x e^{tx} \binom{n}{x} (e^t p)^x q^{n-x} = (e^t p + q)^n$$

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

**تمرین:** يك صفحه هدف زنی دایره ای شکل به ۱۵ قطاع مساوی تقسیم شده و با شماره های ۱ تا ۱۵ متمایز گردیده است، فرض کنید متغیر تصادفی X برابر عددی باشد که سوزن در قطاع مربوط به آن اصابت می‌کند، توزیع احتمال X را مشخص کرده و میانگین و واریانس آن را بدست آورید؟

**تمرین:** احتمال زدن تیر به هدف ۱/۳ است، اگر ۵ بار تیر شلیک شود مطلوبست محاسبه احتمال اینکه:

الف: حداقل دو بار تیر به هدف بخورد

ب: لااقل چند بار تیر را شلیک کرد تا با احتمال بیش از ۹۰٪ تیر به هدف بخورد.

برای محاسبه راحت تر احتمالهایی  $P(X \leq r)$  با توزیع دو جمله ای، جدولهایی آماده برای n های مختلف و به ازای P های متفاوت با rهایی مشخص احتمال آن محاسبه شده است که در محاسبات می توان به آنها مراجعه نمود.

### توزیع چند جمله ای:

$$f(x_1, x_2, \dots, x_k, p_1, p_2, \dots, p_k, n) = \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

$$\sum_{i=1}^k x_i = n, \quad \sum_{i=1}^k p_i = 1$$

توزیع چند جمله ای: آزمایش برنولی را تعمیم می دهیم، به این صورت که نتیجه آزمایش به یکی از پیشامدهای  $E_1, E_2, \dots$ ،  $E_k$  منجر شود به طوری که این پیشامدها ناسازگار و فرسا (در جریان آزمایش الزاماً یکی از آنها رخ دهد) باشد.

حال آزمایش را  $n$  بار به طور مستقل انجام می دهیم اگر احتمال وقوع پیشامدهای  $E_1, E_2, \dots, E_k$  به ترتیب برابر با  $p_1, p_2, \dots, p_k$ ، در نظر بگیریم آنگاه توزیع احتمال توأم متغیرهای تصادفی  $X_1, X_2, \dots, X_k$  که بترتیب نشان دهنده تعداد نتایج پیشامدهای  $E_1, E_2, \dots, E_k$  در این  $n$  آزمایش مستقل را **توزیع چند جمله ای** می نامیم.

**توزیع فوق هندسی:** اگر آزمایشی دو شرط زیر را داشته باشد آن را **فوق هندسی** می نامیم:

- از جمعیتی با  $N$  عضو، یک نمونه تصادفی  $n$  تایی (بدون جایگزاری) انتخاب کنیم.
- $k$  عضو از  $N$  عضو بنام موفقیت و  $N-k$  عضو دیگر بنام عدم موفقیت باشند.

توزیع احتمال آن به صورت زیر نمایش و تعریف می گردد.

$$h(x; N, n, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \dots, n \quad (n < k)$$

$$x = 0, 1, 2, \dots, k \quad (k < n)$$

**مثال:** در یکی از شهرهای کشور، ۴ شبکه تلویزیون قابل استفاده است. بدین ترتیب شهروندان از شبکه یک به میزان ۳۰٪، از شبکه دو به میزان ۱۰٪ و از شبکه سه به میزان ۴۰٪ و بقیه از شبکه چهار می توانند استفاده مطلوب ببرند. اگر ۱۰ نفر از جمعیت این شهرستان به طور تصادفی به مصاحبه دعوت شوند، احتمال آن پیشامدی را بیابید که ۳ نفر از شبکه یک و ۲ نفر از شبکه دو و ۴ نفر از شبکه سه و بقیه از شبکه چهار استفاده نموده اند:

ج: طبق توزیحاتی که داده شد توزیع، یک توزیع چند جمله ای است پس:

$$P(X_1 = 3, X_2 = 2, X_3 = 4, X_4 = 1) = \binom{10}{3, 2, 4, 1} (0.3)^3 (0.1)^2 (0.4)^4 (0.2)^1$$

اگر در توزیع فوق هندسی  $N$  خیلی بزرگ باشد به طوری که نسبت  $n/N$  خیلی به چشم نیاد این توزیع به سمت توزیع دو جمله زیر میل خواهد کرد

$$h(x; N, n, k) \approx b(x; n, \frac{k}{N})$$

برای مثال فرض کنید در یک شهر ۱۰۰۰۰ نفری نسبت رأی موافق به مخالف در مورد موضوع خاصی ۶۰ به ۴۰ است. اگر ۱۵ نفر از این افراد را به تصادف انتخاب کنیم:

اگر این انتخاب با جایگزاری باشد یعنی هر شخص بعد از انتخاب شانس دوباره انتخاب شدن را داشته باشد توزیع دو جمله ای است ولی اگر بدون جایگزاری باشد توزیع فوق هندسی است، از آنجایی که  $N$  بزرگ است و نسبت  $n/N$  کوچک می توان با توزیع دو جمله ای تقریب زد.

**مثال:** محموله ای شامل ۱۰ تلویزیون که ۳ تای آن معیوب است به فروشگاه ای ارسال می شود. شخصی به تصادف ۴ تلویزیون را می خرد اگر  $X$  تعداد تلویزیونهای خرابی باشد که در خرید این شخص باشد توزیع احتمال آن را به دست آورید.

$$h(x; 10, 4, 3) = \frac{\binom{3}{x} \binom{10-3}{4-x}}{\binom{10}{4}}, \quad x = 0, 1, 2, 3$$

**مثال:** از گروهی متشکل از ۷۵ شیمیدان، ۱۵ پزشک و ۳۵ ریاضیدان یک کمیته ۶ نفره به تصادف انتخاب می‌کنیم. مطلوبست محاسبه احتمال اینکه ۲ شیمیدان، ۳ پزشک و ۱ ریاضیدان شرکت کنند:

$$f(2, 3, 1; 25, 15, 35, 75, 6) = \frac{\binom{25}{2} \binom{15}{3} \binom{35}{1}}{\binom{75}{6}}$$

**تعمیم توزیع فوق هندسی:** در توزیع فوق هندسی فقط دو حالت داشتیم **موفقیت و عدم موفقیت**.

حال اگر جمعیت N عضوی به دسته  $A_1, A_2, \dots, A_k$  افزا شود به طوری که از N عضو جمعیت،  $a_1$  عضو در  $A_1$ ،  $a_2$  عضو در  $A_2$ ، ...، و  $a_k$  عضو در  $A_k$  باشد، از این جمعیت یک نمونه تصادفی n تایی بدون جایگذاری برداریم. میخواهیم احتمال توأم آن را حساب کنیم که  $x_1$  عضو از  $A_1$ ،  $x_2$  عضو از  $A_2$ ، ...، و  $x_k$  عضو از  $A_k$  انتخاب شود که به صورت زیر است.

$$f(x_1, x_2, \dots, x_k; a_1, a_2, \dots, a_k, N, n) = \frac{\binom{a_1}{x_1} \binom{a_2}{x_2} \dots \binom{a_k}{x_k}}{\binom{N}{n}}, \quad \sum_{i=1}^k a_i = N, \quad \sum_{i=1}^k x_i = n$$

**مثال:** احتمال اینکه راننده ای از یک چراغ قرمز عبور نماید و پلیس آنرا متوقف کند ۴۰٪ می‌باشد، احتمال آن پیشامد را بیابید که در حین عبور از چراغ قرمز چهارم پلیس او را متوقف کند؟

$$b^*(4; 1, 0.4) = f(4) = \binom{3}{0} (0.4)^1 (0.6)^3 = 0.09$$

**مثال:** احتمال اینکه اگر فردی شایعه ای را بشنود آنرا باور کند ۰,۷ می‌باشد. احتمال آن پیشامدی را بیابید که ۵ امین فردی که این شایعه را می‌شنود سومین فردی باشد که آنرا باور می‌کند؟

$$b^*(5; 3, 0.7) = f(5) = \binom{4}{2} (0.7)^3 (0.3)^2$$

**توزیع دو جمله ای منفی:** اگر آزمایش برنولی را آنقدر ادامه دهیم تا r موفقیت داشته باشیم و پس از وقوع r امین موفقیت آزمایش متوقف گردد، این آزمایش را **دو جمله ای منفی** می‌نامیم. حال اگر متغیر تصادفی X نشان دهنده تعداد دفعات آزمایش دو جمله ای منفی باشد توزیع آن به صورت زیر نمایش و تعریف می‌گردد.

$$b^*(x; r, p) = \binom{x-1}{r-1} p^r q^{x-r}, \quad x = r, r+1, \dots$$

$$M(t) = E(e^{tx}) = p^r e^{tr} (1 - qe^t)^{-r}, \quad t < -\ln q$$

$$\mu = \frac{r}{p}, \quad \sigma^2 = \frac{rq}{p^2}$$

**مثال:** اگر موفقیت جهت اخذ گواهینامه داوطلب ۸۰٪ باشد. اولاً توزیع احتمال متناسب به این آزمایش را معرفی کنید. ثانیاً احتمال آن پیشامدی را بیابید که این داوطلب در مرحله سوم موفق به اخذ گواهینامه شده است.

$$g(x; p) = g(x; 0.8) = 0.8 \times 0.2^{x-1} \quad x = 1, 2, \dots$$

$$g(3; 0.8) = f(3) = (0.8)(0.2)^2 = 0.032$$

**توزیع هندسی:** اگر در آزمایش دو جمله ای منفی  $r=1$  یعنی آزمایش را تا حصول اولین پیروزی ادامه دهیم این آزمایش را **آزمایش هندسی** می‌نامیم و X را که نشان دهنده تعداد دفعات آزمایش هندسی است متغیر تصادفی هندسی نامیم و توزیع آن را به صورت زیر نمایش و تعریف می‌کنیم.

$$g(x; p) = pq^{x-1} \quad x = 1, 2, \dots$$

**توزیع پواسن:** برخی آزمایشها به گونه ای هستند که نتایج حاصل از آنها، تعداد وقایعی است که در فواصل زمانی یا در ناحیه مکانی مشخص به وقوع می پیوندد، چنین آزمایشهایی به آزمایش پواسن معروفند. در واقع بررسی پیشامدهای جدا در در یک فاصله زمانی یا یک مکان مشخص فرایند پواسن نامیده می شود.

مثل تعداد تلفنهایی که در ساعات مشخص به یک مرکز زده می شود. یا تعداد تصادفاتی که در یک چهارراه مشخص رخ می دهد. یا تعداد اشتباهاتی که یک تایپیست در یک صفحه دارد و ...

**تمرین:** از جعبه ای شامل ۵ توپ قرمز و ۴ توپ سبز یک نمونه تصادفی ۶ تایی انتخاب می کنیم مطلوبست محاسبه اینکه ۴ توپ قرمز و ۲ توپ سبز باشد؟

**تمرین:** احتمال اینکه در آزمایش پرتاب سکه، سومین شیر در هفتمین بار پرتاب بدست آید را محاسبه کنید؟

**تمرین:** احتمال اینکه تیراندازی به هدف بزند برابر ۰٫۸ است مطلوبست محاسبه احتمال اینکه:

الف: کمتر از ۵ بار لازم باشد تا اولین تیر به هدف بخورد؟

ب: تعداد زوج بار لازم باشد تا اولین تیر به هدف بخورد؟

**مثال:** پزشکی به طور متوسط در هر ساعت ۴ بیمار را ویزیت می کند. احتمال پیشامدهای زیر را برای او محاسبه کنید.

(۱) در یک ساعت خاص پزشک بیکار باشد.

(۲) در یک ساعت خاص درست دو بیمار را ویزیت کند.

(۳) در یک ساعت خاص حداکثر دو بیمار را ویزیت کند.

(۴) حداقل دو بیمار را ویزیت کند.

(۵) در ۱۵ دقیقه اول درست سه بیمار را ویزیت کند.

**توزیع پواسن:** متغیر تصادفی  $X$  که نشان دهنده تعداد موفقیت در یک آزمایش پواسن باشد را متغیر تصادفی پواسن نامیده و توزیع احتمال آن در صورتیکه میانگین تعداد موفقیت در فاصله زمانی یا ناحیه مکانی مشخص ( $\lambda$ ) معلوم باشد، به صورت زیر نمایش و تعریف می گردد

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

$$M(t) = E(e^{tx}) = e^{\lambda(e^t - 1)}$$

$$\mu = \lambda, \quad \sigma^2 = \lambda$$

**قضیه:** اگر متغیر تصادفی  $X$  دارای توزیع دو جمله ای با توزیع احتمال  $b(x; n, p)$  باشد، آنگاه شکل حدی این توزیع وقتی که  $n \rightarrow \infty$  و  $p \rightarrow 0$  یا  $p \rightarrow 1$  توزیع پواسن با پارامتر  $\lambda = np$  می باشد بعبارت دیگر:

$$b(x; n, p) \approx P(x; np)$$

$$n \rightarrow \infty$$

$$p \rightarrow 0 \text{ or } p \rightarrow 1$$

**جواب:**  $P(x, \lambda) = P(x, 4) = \frac{e^{-4} \cdot 4^x}{x!}, \quad x = 0, 1, \dots$

$$1) P(0, 4) = f(0) = e^{-4}$$

$$2) P(2, 4) = f(2) = \frac{e^{-4} \cdot 4^2}{2!} = 8e^{-4}$$

$$3) P(x \leq 2, 4) = f(x \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = e^{-4} + 4e^{-4} + 8e^{-4}$$

$$4) P(x \geq 2, 4) = f(x \geq 2) = 1 - P(X < 2) = 1 - (e^{-4} + 4e^{-4}) = 1 - 5e^{-4}$$

$$5) \lambda = \lambda \cdot t = 4 \times \frac{1}{4} = 1 \Rightarrow P(x, 1) = \frac{e^{-1} \cdot 1^x}{x!}$$

$$P(3, 1) = f(3) = \frac{e^{-1}}{3!}$$

**مثال:** در يك استودیوم ۱۰۰۰۰ نفری احتمال اینکه فردی در اثر گرمادگی بیمار شود ۰,۰۰۰۳ می باشد. احتمال آن پیشامدی را بیابید که در این جمع ۴ نفر مبتلا به گرمادگی شوند؟

$$b(4; 10000, 0.0003) = f(4) = \binom{10000}{4} (0.0003)^4 (0.997)^{9996}$$

$$b(x; n, p) \square P(x; np)$$

$$n \rightarrow \infty$$

$$p \rightarrow 0 \text{ or } p \rightarrow 1$$

$$b(4; 10000, 0.0003) \square P(4; 3) = f(4) = \frac{e^{-3} \cdot 3^4}{4!} = \frac{81}{24} e^{-3}$$

**تمرین:** فرض که در هر سال، در بین هر ۵۰۰۰۰ نفر، ۲ نفر خودکشی میکنند، احتمال اینکه در يك شهر، ۱۰۰۰۰۰ نفری در يك سال معین،

الف: صفر خودکشی

ب: يك خودکشی

ت: ۲ خودکشی یا بیشتر نتجم گیرد؟

**تمرین:** در يك جاده به خصوص به طور متوسط ۶ تصادف در سال رخ می دهد، مطلوبست محاسبه احتمال اینکه در يك سال مشخص

الف: کمتر از ۴ تصادف رخ دهد؟

ب: حداقل ۴ تصادف رخ دهد؟

**توزیع یکنواخت پیوسته:** متغیر تصادفی X در فاصله [a , b] داراي توزیع یکنواخت است اگر تابع چگالی آن به صورت زیر باشد که به صورت U [a , b] نشان می دهند.

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

$$\mu = \frac{a+b}{2} , \sigma^2 = \frac{(b-a)^2}{12}$$

بعبارت دیگر، در توزیع یکنواخت احتمال قرار گرفتن يك نقطه در قسمتی از این فاصله متناسب با طول آن قسمت بوده و قرار گرفتن آن نقطه در درون این فاصله حتمی است.

## فصل هفتم: بررسی چند توزیع پیوسته

**توزیع نمایی:** گفتیم توزیع پواسن تعداد وقایع در يك ناحیه پیوسته یا يك فاصله زمانی است حال اگر متغیر تصادفی X زمان اولین اتفاق یا زمان بین دو اتفاق متوالی در توزیع پواسن، باشد X يك توزیع نمایی است با پارامتر  $\theta$  که در آن  $1/\theta$  تعداد وقایع در واحد زمان می باشد.

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad x \geq 0 , \theta > 0$$

$$M_X(t) = \frac{1}{1-t\theta} , t < \frac{1}{\theta}$$

$$\mu_X = E(X) = \theta , \sigma_X^2 = \theta^2$$

**مثال:** اگر X ، U[3 , 10] باشد مطلوبست محاسبه

الف: میانگین و واریانس توزیع

$$b: P(4 < X < 8) = \begin{cases} \frac{1}{b-a} = \frac{1}{7} & , 3 \leq x \leq 10 \\ 0 & , \text{other points} \end{cases}$$

$$\mu = \frac{a+b}{2} = \frac{3+10}{2} = 6.5$$

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(10-3)^2}{12} = 24.5$$

$$P(4 < X < 8) = \int_4^8 \frac{1}{7} dx = \frac{4}{7}$$

جواب:

$$\frac{1}{\theta} = 20/60 = 1/3$$

$$f(x) = \frac{1}{3} e^{-\frac{x}{3}} \quad x > 0$$

$$P(X > 5) = \int_5^{\infty} \frac{1}{3} e^{-\frac{x}{3}} dx = e^{-\frac{5}{3}}$$

$$P(X < 10) = \int_0^{10} \frac{1}{3} e^{-\frac{x}{3}} dx = \frac{1}{3} (1 - e^{-\frac{10}{3}})$$

$$\left\{ \begin{array}{l} P(X > t_1 + t_2 | X > t_1) = P(X > t_2) \\ P(X < t_1 + t_2 | X > t_1) = P(X < t_2) \end{array} \right\} \rightarrow \text{توزیع نمایی فاقد حافظه است}$$

$$P(X > 25 | X > 25) = P(X > 5)$$

$$P(X < 25 | X > 20) = P(X < 5)$$

**مثال:** به طور متوسط در هر ساعت تعداد ۲۰ اتومبیل وارد يك پارکینگ می‌شوند اگر ساعت ۶ صبح پارکینگ باز شود مطلوبست محاسبه احتمال اینکه

الف: حداقل ۵ دقیقه طول بکشد تا اولین اتومبیل وارد پارکینگ شود؟

ب: زمان بین ورود دو ماشین به پارکینگ حداقل ۵ دقیقه باشد؟

پ: تا ساعت ۶:۱۰ صبح اولین اتومبیل وارد پارکینگ شود؟

چ: اگر تا ساعت ۶:۲۰ اتومبیلی وارد پارکینگ نشده باشد بعد از ۶:۲۵ وارد شود؟

ج: اگر تا ساعت ۶:۲۰ اتومبیلی وارد پارکینگ نشده باشد تا ۶:۲۵ وارد شود؟

### توزیع گاما

تابع گاما: به صورت زیر تعریف می‌گردد

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

$$\Gamma(1) = 1$$

$$\Gamma(1/2) = \sqrt{\pi}$$

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$$

$$\Gamma(\alpha + 1) = \alpha!$$

که دارای خواص مقابل می‌باشد.

**توزیع گاما:** در بخش قبل زمان لازم برای اولین رخداد توزیع پواسن مطرح گردید که گفتیم توزیع نمایی است با پارامتر  $\theta$ . حال مطلب فوق را تعمیم می‌دهیم و زمان لازم برای  $\alpha$  امین رخداد را بررسی می‌کنیم.

چنین توزیعی را توزیع گاما با پارامترهای  $\alpha$  و  $\theta$  می‌نامیم که تابع چگالی آن به شکل زیر می‌باشد

$$f(x) = \begin{cases} \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\theta}} & 0 \leq x < \infty \\ 0 & x < 0 \end{cases}$$

$$M_x(t) = \left(\frac{1}{1-\theta t}\right)^\alpha = (1-\theta t)^{-\alpha}, \quad t < \frac{1}{\theta}$$

$$\mu_x = \alpha\theta, \quad \sigma_x^2 = \alpha\theta^2$$

**مثال:** فرض کنید در هر ساعت به طور متوسط ۳۰ اتومبیل وارد پارکینگ میشوند، مطلوبست محاسبه

الف: احتمال اینکه مأمور پارکینگ لافل ۵ دقیقه منتظر بماند تا دومین ماشین وارد پارکینگ شود؟

ب: میانگین و واریانس متغیر تصادفی X

$$\frac{1}{\theta} = 30/60 = 1/2 \Rightarrow \theta = 2$$

$$\alpha = 2$$

$$f(x) = \begin{cases} \frac{1}{2^2 \Gamma(2)} x^{2-1} e^{-\frac{x}{2}} & 0 \leq x < \infty \\ 0 & x < 0 \end{cases}$$

$$P(X > 5) = \int_5^{\infty} \frac{1}{4} x e^{-\frac{x}{2}} dx = \frac{7}{2} e^{-\frac{5}{2}}$$

$$\mu_x = \alpha\theta = 2 \times 2 = 4, \quad \sigma_x^2 = 2 \times 2^2 = 8$$

**توزیع توان دوم کای (خی دو):** اگر در توزیع گاما  $\theta=2$  و  $\alpha=v/2$  (عدد صحیح و مثبت) باشد می‌گوییم متغیر تصادفی X دارای توزیع  $X^2$  با درجه آزادی v است.

$$f(x) = \begin{cases} \frac{1}{2^{\frac{v}{2}} \Gamma(\frac{v}{2})} x^{\frac{v}{2}-1} e^{-\frac{x}{2}} & 0 \leq x < \infty \\ 0 & x < 0 \end{cases}$$

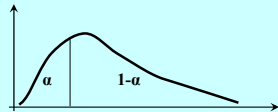
$$M_x(t) = (1-2t)^{-\frac{v}{2}}$$

$$\mu_x = v, \quad \sigma_x^2 = 2v$$

**مثال:** اگر متغیر تصادفی  $X$  دارای توزیع توان دوم کای با  $v=10$  باشد مطلوبست محاسبه  $P(X>3.5)$  و  $P(3.5<X<20.5)$ ؟

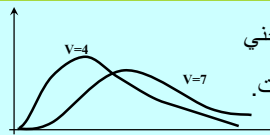
$$P(2.5 < X < 9.34) = P(X < 9.34) - P(X < 2.5) = 0.5 - 0.01 = 0.49$$

$$P(X > 2.5) = 1 - P(X < 2.5) = 1 - 0.01 = 0.99$$



v	$X^2_{0.005}$	$X^2_{0.01}$	$X^2_{0.25}$	$X^2_{0.05}$	$X^2_{0.1}$	$X^2_{0.25}$	$X^2_{0.5}$
1							
2							
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34

مقادیر  $X^2$  منفی نمی باشند و منحنی توزیع آن حول  $x=0$  متقارن نیست.



چون این توزیع در عمل مورد استفاده فراوان قرار می‌گیرد سطح زیر منحنی این توزیع به ازای مقادیر مختلف  $X^2$  و درجات آزادی مختلف محاسبه و جدول آن تهیه شده است. چگونگی استفاده از این جدول را با مثالی شرح می‌دهیم:

در سالهای اخیر توزیع بتا کاربردهای مهمی در استنباطهای بیزی پیدا کرده است که در آن پارامترها به عنوان متغیر تصادفی در نظر گرفته می‌شوند که با تغییر پارامترها تابع چگالی شکلهای گوناگونی به خود می‌گیرد.

در حالتی که  $\alpha=\beta=1$  داریم

$$f(x) = \begin{cases} \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} x^0 \cdot (1-x)^0 = 1, & 0 < x < 1 \\ 0, & \text{other points} \end{cases}$$

که همان تابع چگالی یکنواخت می‌باشد.

**تابع بتا:** به صورت زیر تعریف می‌گردد و دارای خواص زیر می‌باشد:

$$\beta(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$$

$$\beta(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

**توزیع بتا:** متغیر تصادفی  $X$  دارای توزیع بتا است هرگاه تابع چگالی احتمال آن به صورت زیر باشد:

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} \cdot (1-x)^{\beta-1}, & 0 < x < 1, \alpha, \beta > 0 \\ 0, & \text{other points} \end{cases}$$

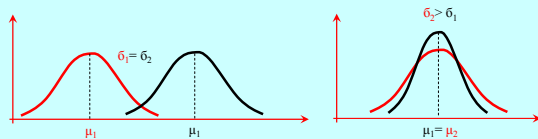
$$\mu = \frac{\alpha}{\alpha + \beta}, \quad \sigma^2 = \frac{\alpha\beta^2}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

**توزیع نرمال:** متغیر تصادفی  $X$  دارای توزیع نرمال است هرگاه تابع چگالی احتمال آن به صورت زیر باشد:

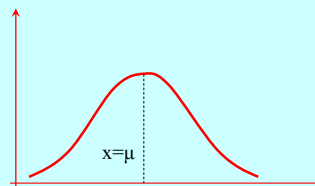
$$n(x; \mu, \sigma) = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in \mathbb{R}$$

$$M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

$$\mu_X = \mu, \quad \sigma_X^2 = \sigma^2$$



**توزیع نرمال:** نمودار این توزیع به منحنی نرمال معروف است و زنگی شکل بوده و بیشتر وقایعی که در طبیعت و تحقیقات علمی بوقوع می‌پیوندد از این منحنی پیروی میکند. متغیر تصادفی  $X$  که منحنی آن زنگی شکل باشد را متغیر تصادفی نرمال می‌نامیم. این منحنی نسبت به خط متقارن  $x=\mu$  است.



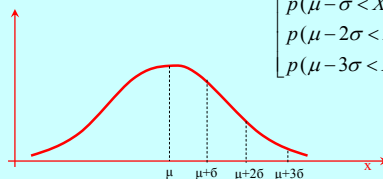
### خواص توزیع نرمال:

این منحنی نسبت به خط متقارن  $x=\mu$  است .  
میان، میانگین و مد بر هم منطبقند.

منحنی دارای دو نقطه عطف در نقاط  $x=\mu \pm \sigma$

محور  $x$  ها مجانب افقی منحنی است یعنی  $\lim_{x \rightarrow \pm\infty} f(x)=0$  وقتی

$$\left[ \begin{aligned} p(\mu - \sigma < X < \mu + \sigma) &= 0.68 \\ p(\mu - 2\sigma < X < \mu + 2\sigma) &= 0.95 \\ p(\mu - 3\sigma < X < \mu + 3\sigma) &= 0.997 \end{aligned} \right]$$



**متغیر تصادفی نرمال استاندارد:** متغیر تصادفی با میانگین  $=0$  و واریانس  $\sigma^2=1$  را متغیر تصادفی نرمال استاندارد می‌گوییم و آن را با  $Z$  نشان می‌دهیم.

• اگر  $X$  یک متغیر تصادفی با میانگین  $\mu$  و واریانس  $\sigma^2$  باشد آنگاه متغیر تصادفی  $Z = \frac{X - \mu}{\sigma}$  دارای توزیع نرمال استاندارد است .

**مثال:** اگر  $X$  دارای توزیع نرمال با  $\mu=1$  و  $\sigma^2=4$  مطلوبست محاسبه احتمالات  $P(X < 2.3)$  و  $P(0 < X < 3)$

$$P(0 < X < 3.5) = P\left(\frac{0-1}{2} < Z < \frac{3.5-1}{2}\right) = P\left(-\frac{1}{2} < Z < 1.25\right) = \phi(1.25) - \phi(-0.5) = 0.8944 - 0.3085 = 0.5859$$

$$P(X > 4) = P\left(Z > \frac{4-1}{2}\right) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 - \phi(1.5) = 1 - 0.9332 = 0.0668$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
-0.5	0.3085						
1.2						0.8944	
1.5	0.9332						

برای محاسبه احتمال  $P(x_1 < X < x_2)$  که در آن  $X$  نرمال با میانگین و واریانس مشخص باشد به صورت زیر عمل می‌کنیم:

$$\mu_x = \mu, \quad \sigma_x^2 = \sigma^2$$

$$p(x_1 < X < x_2) = p\left(\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma}\right) =$$

$$P(z_1 < Z < z_2) = P(Z < z_2) - P(Z < z_1) = \phi(z_2) - \phi(z_1)$$

مقدار تابع  $\Phi$  که مقادیر متغیر تصادفی نرمال استاندارد برای احتمال‌هایی به صورت  $P(Z < z)$  می‌باشد در جداولی به صورت آماده محاسبه می‌گردد که چگونگی استفاده از این جدول با مثالی شرح داده می‌شود.

**مثال:** اگر  $X$  دارای توزیع نرمال با  $\mu=25$  و  $\sigma=6$  باشد مطلوبست تعیین ثابت  $c$  در احتمال زیر؟

$$P(|X - 25| \leq c) = 0.9544$$

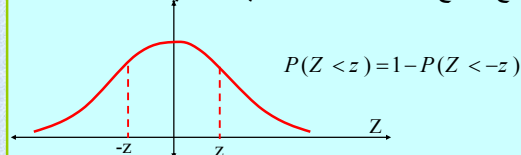
$$P(|X - 25| \leq c) = 0.9544 = P(-c \leq X - 25 \leq c) =$$

$$P\left(\frac{-c}{6} \leq \frac{X - 25}{6} \leq \frac{c}{6}\right) = P\left(-\frac{c}{6} \leq Z \leq \frac{c}{6}\right) =$$

$$2P\left(Z \leq \frac{c}{6}\right) - 1 = 0.9544 \Rightarrow P\left(Z \leq \frac{c}{6}\right) = 0.9772$$

$$\frac{c}{6} = 2 \Rightarrow c = 12$$

چون تابع توزیع نرمال متقارن هست پس برای هر  $Z$ :



$$P(Z < -1) = 1 - P(Z < 1) = 1 - 0.8413 = 0.1587$$

$$P(-0.5 < Z < 0.5) = P(Z < 0.5) - P(Z < -0.5) =$$

$$P(Z < 0.5) - (1 - P(Z < 0.5)) = 2P(Z < 0.5) - 1 = (2 \times 0.6915) - 1 = 0.383$$

**تقریب توزیع نرمال برای توزیع دو جمله ای:**

زمانی که  $n$  در توزیع دو جمله ای بزرگ باشد، عملاً محاسبه احتمالات دو جمله ای ممکن نیست در این حالت می توانیم از توزیع نرمال برای تقریب دو جمله ای استفاده کنیم. وقتی  $np > 5$ ، تقریب نرمال برای توزیع دو جمله ای بسیار خوب است.

برای آن که بتوانیم توزیع دو جمله ای را به کمک توزیع نرمال تقریب بزنیم باید توجه داشته باشیم که چون دو جمله ای یک توزیع گسسته می باشد و آن را به کمک توزیع نرمال که یک توزیع پیوسته می باشد تقریب می زنیم باید از تصحیح پیوستگی به صورت زیر استفاده کنیم:

$$P(X = k) = P(k - 0.5 \leq X \leq k + 0.5)$$

**تقریب توزیع نرمال برای توزیع دو جمله ای:**

اگر  $X$  یک متغیر تصادفی گسسته از نوع دو جمله ای با میانگین  $\mu = np$  و واریانس  $\sigma^2 = npq$  باشد، شکل حدی توزیع متغیر تصادفی زیر وقتی  $n$  به سمت بینهایت میل می کند توزیع نرمال استاندارد می باشد.

$$Z = \frac{X - np}{\sqrt{npq}}$$

**مثال:** فرض می کنیم  $X$  دارای توزیع دو جمله ای به صورت  $b(x; 15, 0.4)$  باشد، مطلوب است محاسبه  $P(7 < X < 10)$ ؟  
حل:

$$\begin{aligned} np &= 6 > 5 \\ p(7 \leq X \leq 10) &= p(6.5 \leq X \leq 10.5) \\ p\left(\frac{6.5-6}{\sqrt{6(0.6)}} \leq \frac{X-np}{\sqrt{np(1-p)}} \leq \frac{10.5-6}{\sqrt{6(1-0.4)}}\right) &= \\ p(0.26 < Z < 2.37) &= \Phi(2.37) - \Phi(0.26) = \\ &= 0.9911 - 0.6026 = 0.3885 \end{aligned}$$

**قضیه:** اگر  $X$  یک متغیر تصادفی نرمال با میانگین  $\mu$  و واریانس  $\sigma^2$  باشد، آنگاه متغیر تصادفی  $V$  دارای توزیع توان دوم کای با  $k$  درجه آزادی است.

$$V = \left(\frac{X - \mu}{\sigma}\right)^2$$

$$P(|Z| < 1.96) = P(Z^2 < 3.842) = P(X^2 < 3.84) = 0.95$$

**قضیه:** اگر  $X_1, X_2, \dots, X_n$  متغیر تصادفی مستقل نرمال با میانگین های  $\mu_1, \mu_2, \dots, \mu_n$  و واریانس های  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$  باشند، آنگاه  $U$  دارای توزیع توان دوم کای با  $n$  درجه آزادی است.

$$U = \sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$$

**تمرین:** معدل نمره ۳۰۰ دانشجوی یک دانشکده تقریباً دارای توزیع نرمال با میانگین  $2/1$  و انحراف معیار  $1/2$  است. در صورتیکه معدلها تا یکدهم تقریب محاسبه شده باشند، مطلوب است

الف: معدل چند نفر از دانشجویان در فاصله  $[2/5, 2/5]$  قرار دارد؟

الف: احتمال اینکه معدل یک دانشجو که به تصادف انتخاب می شود دقیقاً ۲,۵ باشد؟

**تمرین:** تاس همگنی را ۷۲۰ بار پرتاب می کنیم اگر متغیر تصادفی  $X$  نشان دهنده تعداد ۶ های ظاهر شده باشد، مطلوب است محاسبه  $P(100 < X < 125)$

**تمرین:** اگر متغیر تصادفی  $X$  دارای توزیع توان دوم کای با  $v=4$  باشد، مقدار  $x$  را در حالات زیر محاسبه کنید؟

الف:  $P(X < x) = 0.99$

ب:  $P(X > x) = 0.75$

**تمرین:** اگر متغیر تصادفی  $X$  دارای توان دوم کای با  $v=18$  باشد مطلوب است:

الف: محاسبه میانگین و واریانس توزیع  $X$

ب:  $P(X > 17.3)$

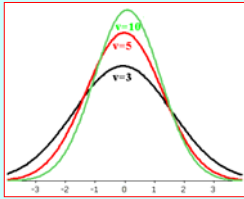
**تمرین:** اگر متغیر تصادفی  $X$  دارای توزیع نمایی با میانگین  $0.5$  باشد مطلوب است

محاسبه  $P(X > 4)$

**تمرین:** اگر متغیر تصادفی  $X$  دارای توزیع نرمال با میانگین ۳ و واریانس ۱۶ مطلوب است محاسبه  $P(4 < X < 8)$

**تمرین:** طول عمر یک نوع وسیله الکتریکی دارای توزیع نرمال با میانگین ۲ و انحراف معیار  $0.3$  سال است. احتمال اینکه یک وسیله از این نوع دارای طول عمر کمتر از  $2/3$  سال باشد، چقدر است؟

توزیع t مانند توزیع نرمال زنگی شکل است و همانند توزیع نرمال استاندارد نسبت به خط  $x=0$  متقارن است و هر چه درجه آزادی بزرگتر باشد توزیع t بسمت نرمال استاندارد میل خواهد کرد. و وقتی  $v \rightarrow +\infty$  منحنی توزیع t همانند توزیع نرمال استاندارد است.



**توزیع student t:** فرض کنید  $Z$  یک متغیر تصادفی نرمال استاندارد و  $\chi^2$  یک متغیر تصادفی توان دوم کای با درجه آزادی  $v$  باشد و همچنین  $Z$  و  $\chi^2$  مستقل باشند، آنگاه

$$T = \frac{Z}{\sqrt{\chi^2/v}} =$$

دارای تابع چگالی به صورت زیر می‌باشد که به توزیع  $t$  با  $v$  درجه آزادی می‌نامیم.

$$f(t) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-\frac{1}{2}(v+1)}, t \in \mathbb{R}$$

$$\mu = 0, \sigma^2 = \frac{v}{v-2}, v > 2$$

**مثال:**

$$P(T < 1.75), v = 16 \Rightarrow = 0.95$$

$$P(T > 1.75), v = 16 = 1 - P(T < 1.75) = 1 - 0.95 = 0.05$$

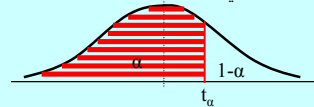
$$P(T < t) = 0.95, v = 24 \Rightarrow t = 1.71$$

$$P(T > t) = 0.65, v = 24 \Rightarrow 1 - P(T < t) = 0.65$$

$$P(T < t) = 0.35 \Rightarrow t = 0.65$$

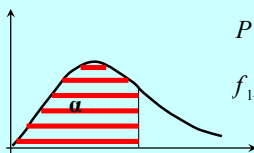
v	$t_{0.55}$	$t_{0.6}$	$t_{0.7}$	$t_{0.75}$	$t_{0.8}$	$t_{0.9}$	$t_{0.95}$
16	0.128					1.34	
24		0.256	0.531	0.685			1.71

**محاسبه احتمال در توزیع t:** چون توزیع  $t$  دارای تابع چگالی احتمال پیچیده‌ای است و انتگرال گیری از آن مشکل می‌باشد، احتمالهای این توزیع نیز مانند توزیع نرمال در جداولی برای  $P(T < t)$  با درجات آزادی مختلف محاسبه شده است و می‌توان از آن استفاده نمود.



$P(T < t_\alpha) = \alpha$	$P(Z < z_\alpha) = \alpha$
$P(T < -t_\alpha) = 1 - \alpha$	$P(Z < -z_\alpha) = 1 - \alpha$
$P(T < t_{1-\alpha}) = P(T < -t_\alpha)$	$P(Z < z_{1-\alpha}) = P(Z < -z_\alpha)$
$t_{1-\alpha} = -t_\alpha$	$z_{1-\alpha} = -z_\alpha$

**محاسبه احتمال در توزیع F:** چون توزیع  $F$  تابع چگالی پیچیده‌ای دارد و محاسبه انتگرال ساده نمی‌باشد احتمالهای این توزیع نیز بر حسب  $v_1$  و  $v_2$  های مختلف محاسبه می‌گردد. که  $F_\alpha(v_1, v_2)$  نقطه‌ای از توزیع  $F$  با درجات  $v_1$  و  $v_2$  می‌باشد که مساحت  $\alpha$  در سمت چپ آن واقع شده است.



$$P(F(v_1, v_2) \leq f_\alpha(v_1, v_2)) = \alpha$$

$$f_{1-\alpha}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}$$

**توزیع F:** اگر  $U$  و  $V$  دو متغیر تصادفی مستقل توان دوم کای با درجات آزادی  $v_1$  و  $v_2$  باشند آنگاه متغیر تصادفی  $F$  را توزیع  $F$  با درجه آزادی  $v_1$  و  $v_2$  می‌نامیم که ترتیب درجات آزادی با توجه به شکل تابع توزیع مهم می‌باشد. که تابع چگالی آن به شکل زیر است:

$$F = \frac{U/v_1}{V/v_2}$$

$$f(x) = \begin{cases} \frac{\Gamma((v_1+v_2)/2) (v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \cdot \frac{x^{v_1-1}}{(1+v_1x/v_2)^{(v_1+v_2)/2}}, & x > 0 \\ 0, & \text{other points} \end{cases}$$

$$\mu_x = \frac{v_2}{v_2-2}, \sigma_x^2 = \frac{2v_2^2(v_1+v_2-2)}{v_1(v_2-4)(v_2-2)^2}, Mo = \left(\frac{v_1-2}{v_1}\right)\left(\frac{v_2}{v_2+2}\right)$$

رابطه زیر در محاسبه احتمالات مختلف در سه تابع توزیع نرمال، t و F رابطه ای مفید می باشد.

$$P\left(z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(t_{\frac{\alpha}{2}} < T < t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(f_{\alpha/2}(v_1, v_2) < F < f_{1-\alpha/2}(v_1, v_2)\right) = 1 - \alpha$$

مثال:

$$P(F(7,4) \leq 6.09) = 0.95$$

$$P(F(3,5) > a) = 0.05 \Rightarrow 1 - P(F(3,5) < a) = 0.05$$

$$P(F(3,5) < a) = 0.95 \Rightarrow a = 5.41$$

		F <sub>0.95</sub>						
v <sub>2</sub> \ v <sub>1</sub>		1	2	3	4	5	6	7
4								6.09
5				5.41				

**مقدمه:** در این فصل هدف پیدا کردن تابع توزیع یا چگالی احتمال توابعی مانند  $y = u(x_1, x_2, \dots, x_n)$  با فرض معلوم بودن توابع احتمال متغیرهای تصادفی  $x_1, x_2, \dots, x_n$  است. برای این کار چندین روش موجود است:

۱- تکنیک تابع توزیع

۲- تکنیک تبدیل متغیرها

۳- تکنیک تابع مولد گشتاورها

که بسته به نوع مسأله می توان از یک یا چند تکنیک مختلف برای پیدا کردن توابع احتمال استفاده کرد.

## فصل هشتم: تابع های متغیرهای تصادفی

**مثال:** اگر توزیع توأم  $x_1$  و  $x_2$  به صورت زیر باشد مطلوبست چگالی احتمال توأم  $y = x_1 + x_2$

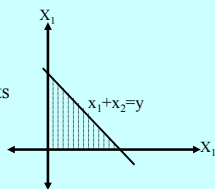
$$f(x_1, x_2) = \begin{cases} 6e^{-3x_1 - 2x_2} & ; x_1, x_2 > 0 \\ 0 & ; \text{other points} \end{cases}$$

$$y = x_1 + x_2 :$$

$$F(y) = P(Y < y) = P(x_1 + x_2 < y) = \int_0^y \int_0^{y-x_1} 6e^{-3x_1 - 2x_2} dx_1 dx_2 =$$

$$= 1 + 2e^{-3y} - 3e^{-2y}$$

$$f(y) = \frac{dF(y)}{dy} = \begin{cases} 6(e^{-2y} - e^{-3y}) & ; y > 0 \\ 0 & ; \text{other points} \end{cases}$$



**تکنیک تابع توزیع:**

روشی برای به دست آوردن چگالی احتمال تابعی از متغیرهای تصادفی پیوسته است که اگر  $x_1, x_2, \dots, x_n$  متغیرهای تصادفی پیوسته با چگالی احتمال مفروضی باشند آنگاه:

$$Y = u(x_1, x_2, \dots, x_n)$$

$$F(y) = P(Y \leq y) = P[u(x_1, x_2, \dots, x_n) \leq y] \Rightarrow$$

$$f(y) = \frac{dF(y)}{dy}$$

### تكنيك تبديل متغير:

اين تكنيك در هر دو حالت گسسته و پيوسته کاربرد دارد:

**الف:** در حالت گسسته مادامي كه رابطه بين مقادير  $x$  (متغير تصادفي با تابع توزيع معلوم) و تابع  $y=u(x)$  (متغير تصادفي با تابع توزيع مجهول) يك رابطه يك به يك است آنچه بايد انجام دهيم فقط يك جايگذاري است ولي اگر اين رابطه يك به يك نباشد بايد دقت شود اين جايگذاري به ازاي تمام روابط جايگزين شود.

**مثال:** اگر  $x$  تعداد شيرهاي باشد كه در چهار پرتاب يك سكه همگن به دست مي آيد توزيع احتمال  $y=1/(1+x)$  و  $z=(x-2)^2$  را بيابيد؟

$$X : b(x; 4, 0.5) \Rightarrow f(x) = \binom{4}{x} \cdot 0.5^x \cdot 0.5^{4-x} = \binom{4}{x} \cdot 0.5^4 ; x = 0, 1, 2, 3, 4$$

$$y = \frac{1}{1+x} \Rightarrow x = \frac{1}{y} - 1 \Rightarrow g(y) = f\left(\frac{1}{y} - 1\right) = \binom{4}{\frac{1}{y} - 1} \cdot 0.5^4 ; y = 1, 1/2, 1/3, 1/4$$

$$z = (x-2)^2 = \begin{cases} 4 ; x=0 \\ 1 ; x=1 \\ 0 ; x=2 \\ 1 ; x=3 \\ 4 ; x=4 \end{cases} \Rightarrow h(z) = \begin{cases} f(2) ; z=0 \\ f(1)+f(3) ; z=1,3 \\ f(0)+f(4) ; z=0 \end{cases}$$

### تكنيك تبديل متغير:

**ب:** در حالت پيوسته، فرض كنيم  $f(x)$  مقدار چگالي احتمال متغير تصادفي  $x$  باشد اگر تابعي كه به صورت  $y=u(x)$  داده شده است مشتقپذير و به ازاي تمامي مقادير برد  $x$  كه براي آنها  $f(x) \neq 0$ ، صعودي يا نزولي باشد آنگاه، براي اين مقادير  $x$  معادله  $y=u(x)$  را مي توان به صورت يكتا بر حسب  $x$  حل كرد تا  $x=w(y)$  به دست آيد و چگالي احتمال  $y$  به صورت زير است:

$$g(y) = \begin{cases} f[w(y)] \cdot |w'(y)| ; u'(x) \neq 0 \\ 0 ; \text{other points} \end{cases}$$

**مثال:** اگر  $x$  داراي توزيع نمائي به صورت زير باشد مطلوبست چگالي احتمال متغير تصادفي  $y$ ؟

$$f(x) = \begin{cases} e^{-x} ; x > 0 \\ 0 ; \text{other points} \end{cases}$$

$$y = \sqrt{x} \Rightarrow f(y)?$$

$$y^2 = x \Rightarrow x' = 2y$$

$$f(y) = f(y^2) \cdot |2y| = \begin{cases} 2ye^{-y^2} ; y > 0 \\ 0 ; \text{other points} \end{cases}$$

### تكنيك تبديل متغير:

اين تكنيك را مي توان براي پيدا كردن توزيع متغير تصادفي كه تابعي از دو يا چند متغير تصادفي است به كار برد به اين صورت كه اگر تابع احتمال توأم  $x_1, x_2, \dots, x_n$ ، مفروض باشد و بخواهيم تابع احتمال  $y=u(x_1, x_2, \dots, x_n)$  را به دست آوريم رابطه يكي از  $x_i$  ها و  $y$  را با ثابت نگه داشتن ساير  $x_i$  ها حساب مي كنيم:

$$y = u(x_1, x_2, \dots, x_n) : \exists x_k : x_k = v(y, x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$$

حال در حالت گسسته توزيع توأم  $y$  و ساير  $x_i$  ها را با جايگذاري به دست مي آوريم سپس توزيع حاشيه اي  $y$  را مي توان از توزيع توأم بدست آورد ولي در حالت پيوسته از رابطه زير استفاده مي كنيم:

$$g(y, x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) = f(x_1, x_2, \dots, x_n) \cdot \left| \frac{\partial x_k}{\partial y} \right|$$

حال مي توان چگالي احتمال حاشيه اي  $y$  را از روي چگالي توأم موجود به دست آورد.

**مثال:** اگر  $x_1$  و  $x_2$  متغيرهاي تصادفي مستقل باشند كه توزيع هاي پواسن با پارامترهاي  $\lambda_1$  و  $\lambda_2$  دارند مطلوبست توزيع احتمال متغير تصادفي  $y$ ؟

$$y = x_1 + x_2$$

$$f(x_1, x_2) = \frac{e^{-\lambda_1} \lambda_1^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2} \lambda_2^{x_2}}{x_2!} = \frac{e^{-\lambda_1 + \lambda_2} \lambda_1^{x_1} \lambda_2^{x_2}}{x_1! x_2!} ; x_1, x_2 = 0, 1, \dots$$

$$y = x_1 + x_2 \Rightarrow x_1 = y - x_2 \Rightarrow$$

$$g(y, x_2) = \frac{e^{-\lambda_1 + \lambda_2} \lambda_1^{y-x_2} \lambda_2^{x_2}}{x_2! (y-x_2)!} \Rightarrow h(y) = \sum_{x_2=0}^y \frac{e^{-\lambda_1 + \lambda_2} \lambda_1^{y-x_2} \lambda_2^{x_2}}{x_2! (y-x_2)!} =$$

$$= \frac{e^{-\lambda_1 + \lambda_2}}{y!} \sum_{x_2=0}^y \frac{y!}{x_2! (y-x_2)!} \lambda_1^{y-x_2} \lambda_2^{x_2} = \frac{e^{-\lambda_1 + \lambda_2}}{y!} (\lambda_1 + \lambda_2)^y ; y = 0, 1, \dots$$

**تکنیک تابع مولد گشتاورها:**

این تکنیک در تعیین توزیع احتمال یا چگالی احتمال تابعی از متغیرهای تصادفی، وقتی تابع مزبور ترکیب خطی از متغیرهای تصادفی مستقل است نقش مهمی دارد به این صورت که اگر  $X_1, X_2, \dots, X_n$  متغیرهای تصادفی مستقل و مقدار تابع مولد گشتاورهای  $X_i$  به ازای  $t$  باشند و آنگاه  $y = X_1 + X_2 + \dots + X_n$

$$M_y(t) = \prod_{i=1}^n M_{X_i}(t)$$

**مثال:** اگر چگالی توأم  $X_1, X_2$  و  $X_3$  به صورت زیر باشد مطلوبست چگالی احتمال متغیر تصادفی  $y$ ؟

$$y = x_1 + x_2 + x_3$$

$$f(x_1, x_2, x_3) = \begin{cases} e^{-(x_1+x_2+x_3)} & ; x_1, x_2, x_3 > 0 \\ 0 & ; \text{other points} \end{cases}$$

$$x_1 = y - x_2 - x_3 \Rightarrow \frac{\partial x_1}{\partial y} = 1$$

$$g(y, x_2, x_3) = e^{-y} \cdot |1| = e^{-y} ; x_2, x_3 > 0 \text{ \& } y > x_2 + x_3 \Rightarrow$$

$$h(y) = \int_0^y \int_0^{y-x_3} e^{-y} dx_2 dx_3 = \frac{1}{2} y^2 e^{-y}$$

**تمرین:** اگر چگالی احتمال  $x$  به صورت زیر باشد مطلوبست چگالی احتمال  $y = x^3$

$$f(x) = \begin{cases} 6x(1-x) & ; 0 < x < 1 \\ 0 & ; \text{other points} \end{cases}$$

**تمرین:** اگر  $y = |x|$  مطلوبست تابع احتمال  $y$ ؟

**تمرین:** اگر  $x$  دارای توزیع نرمال استاندارد باشد مطلوبست چگالی احتمال  $z = x^2$

**تمرین:** اگر  $F(x)$  تابع توزیع تجمعی متغیر تصادفی  $x$  باشد مطلوبست چگالی احتمال  $y = F(x)$ ؟

**مثال:** اگر  $X_1, X_2, \dots, X_n$  متغیرهای تصادفی مستقل از توزیع نمایی با پارامتر  $\theta$  باشند مطلوبست چگالی احتمال  $y$ ؟

$$y = x_1 + x_2 + \dots + x_n :$$

$$M_{X_i}(t) = (1 - \theta t)^{-1} \Rightarrow M_y(t) = \prod_{i=1}^n (1 - \theta t)^{-1} = (1 - \theta t)^{-n}$$

پس  $y$  يك توزیع گاما با  $\alpha = n$  و  $\beta = \theta$  است

**مثال:** اگر  $X_1, X_2, \dots, X_n$  متغیرهای تصادفی مستقل از توزیع پواسن با پارامتر  $\lambda$  باشند مطلوبست چگالی احتمال  $y$ ؟

$$y = x_1 + x_2 + \dots + x_n :$$

$$M_{X_i}(t) = e^{\lambda_i (e^t - 1)} \Rightarrow M_y(t) = \prod_{i=1}^n (e^{\lambda_i (e^t - 1)}) = e^{(\lambda_1 + \dots + \lambda_n)(e^t - 1)}$$

پس  $y$  يك توزیع پواسن با پارامتر  $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$  است

**فصل نهم: جامعه و نمونه آماری**

**تمرین:** اگر چگالی توأم دو متغیر تصادفی به صورت زیر باشد مطلوبست تابع چگالی  $y$ ؟

$$f(x_1, x_2) = \begin{cases} e^{-(x_1+x_2)} & ; x_1, x_2 > 0 \\ 0 & ; \text{other points} \end{cases}$$

$$y = \frac{x_1}{x_1 + x_2} ; h(y)?$$

### جامعه آماری:

- به مجموعه کل مشاهداتی که مورد بررسی هستند، جامعه آماری اطلاق می‌شود.
- تعداد عضوهای جامعه را حجم جامعه می‌نامیم.
- مشاهدات آماری مقادیری از یک متغیر تصادفی هستند پس به هر جامعه آماری می‌توان یک توزیع احتمال متناسب با متغیر تصادفی آن، نسبت داد.
- جامعه می‌تواند طبق حجم آن، منتهای یا نامنتهای باشد.
- در جوامع نامنتهای از میانگین و واریانس آن به دلیل عدم امکان دسترسی به تمام اعضای آن، اطلاعی نداریم از این رو میانگین و واریانس جامعه را پارامترهای جامعه می‌نامیم.
- برای تخمین (برآورد) پارامترهای جامعه نامنتهای، از نمونه (تصادفی) استفاده می‌کنیم
- نمونه باید معرف جامعه باشد (مشت نمونه خروار باشد).

**نمونه تصادفی:** یک نمونه تصادفی به حجم  $n$ ، نمونه‌ایست که هر زیر مجموعه  $n$  عضوی از جامعه دارای شانس انتخاب یکسان باشند بعبارت دیگر:

اگر  $X_1, X_2, \dots, X_n$  متغیر تصادفی با تابع احتمال یکسان  $f(x)$  باشند، آنگاه  $x_1, x_2, \dots, x_n$  را یک نمونه تصادفی با حجم  $n$  از جامعه می‌نامیم و توزیع احتمال آن عبارتست از:

$$f(x_1, x_2, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n)$$

هدف تخمین پارامترهای جامعه از روی نمونه تصادفی است  
**(آمار استنباطی)**

**آماره:** هر تابعی از عضوهای نمونه تصادفی که شامل پارامترهای مجهول نباشد را یک آماره می‌نامیم اگر  $X_1, X_2, \dots, X_n$ ، یک نمونه تصادفی از متغیر تصادفی  $X$  باشند آنگاه توابع زیر یک آماره هستند:

$$X_1 + 3X_2 - 1, \frac{\sum_{i=1}^n X_i}{n}, \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}, \frac{X_1}{X_4}$$

آماره‌ها را با یکی از حروف لاتین نشان داده و روی آن علامتی به صورت  $(- , ^ , \sim)$  قرار می‌دهیم. هدف تعیین توابع توزیع آماره‌ها برای بررسی آنها و تخمین پارامترهای جامعه از روی آنهاست.

### بررسی چند آماره مفید

**گشتاور نمونه ای:** اگر  $X_1, X_2, \dots, X_n$  یک نمونه تصادفی از متغیر تصادفی  $X$  باشد آنگاه روابط زیر به ترتیب معرف گشتاور نمونه مرتبه  $r$  ام حول نقطه  $a$ ، حول میانگین نمونه و حول مبدأ تعریف می‌گردد.

$$M_r^a = \frac{1}{n} \sum_{i=1}^n (X_i - a)^r, r=1,2,3,\dots$$

$$a = \bar{X} \Rightarrow M_r^{\bar{X}} = M_r, r=1,2,3,\dots$$

$$a = 0 \Rightarrow M_r^0 = M_r', r=1,2,3,\dots$$

در گشتاور نمونه مرتبه  $r$  ام حول مبدأ اگر  $r=1$  میانگین نمونه بدست می‌آید

$$M_1' = \frac{1}{n} \sum_{i=1}^n (X_i) = \bar{X}$$

### یادآوری:

**گشتاور:** یکی از پارامترهای پراکنندگی جامعه می‌باشد. گشتاور مرتبه  $r$  ام پیرامون نقطه ثابت  $a$  را برای یک جامعه به شکل زیر نمایش و تعریف می‌کنند.

$$m_r^a = \frac{1}{N} \sum_{i=1}^N (x_i - a)^r$$

گشتاور حول میانگین جامعه را گشتاور مرکزی نامند و آن را با  $\mu_r$  نشان می‌دهند

$$m_r^\mu = \mu_r = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^r$$

گشتاور حول مبدأ را نیز به اختصار به صورت زیر نمایش داده می‌شود

$$m_r^0 = \mu_r' = \frac{1}{N} \sum_{i=1}^N (x_i)^r$$

**واریانس نمونه:** اگر  $X_1, X_2, \dots, X_n$  یک نمونه تصادفی از جامعه‌ای با متغیر تصادفی  $X$  و تابع احتمال  $f(x)$  باشند، واریانس نمونه به صورت زیر تعریف می‌شود:

$$S_n^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, n > 1$$

**قضیه:** اگر  $X_1, X_2, \dots, X_n$  یک نمونه تصادفی از جامعه‌ای نامتناهی با متغیر تصادفی  $X$  و تابع احتمال  $f(x)$  باشند آنگاه:

$$E(S^2) = \sigma_x^2$$

$$\text{var}(S^2) = \frac{2\sigma^4}{n-1}$$

**قضیه:** اگر  $X_1, X_2, \dots, X_n$  یک نمونه تصادفی از جامعه‌ای با متغیر تصادفی  $X$  و تابع احتمال  $f(x)$  باشند آنگاه:

if  $(\mu'_r \text{ exist}) \Rightarrow$

$$E(M'_r) = \mu'_r$$

$$V(M'_r) = \frac{1}{n} [\mu'_{2r} - (\mu'_r)^2]$$

در حالت خاص اگر  $r=1$ :

$$** E(\bar{X}) = \mu_x$$

$$** V(\bar{X}) = \frac{\sigma_x^2}{n}$$

**مثال:** یک نمونه تصادفی ۲۵ تایی از یک جامعه نرمال با واریانس ۶ انتخاب می‌کنیم. مطلوبست احتمال زیر؟

$$P(S^2 > 9.1)?$$

$$P(S^2 > 9.1) = 1 - P(S^2 < 9.1) =$$

$$1 - P\left(\frac{(25-1)S^2}{6} < \frac{(25-1)9.1}{6}\right) = 1 - P(\chi^2 < 36.4) = 0.05$$

**توزیع نمونه آماره  $S^2$ :** می‌دانیم که آماره  $S^2$  واریانس نمونه است و به صورت زیر تعریف می‌شود:

$$S_n^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, n > 1$$

**قضیه:** توزیع آماره  $S^2$  مشخص نیست ولی اگر  $S^2$  واریانس یک نمونه تصادفی با حجم  $n$  از یک جامعه نرمال با واریانس  $\sigma^2$  باشد آنگاه توزیع آماره

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

توزیع توان دوم کای با  $v=n-1$  درجه آزادی است.

**مثال:** اگر  $S_1^2$  و  $S_2^2$  واریانس‌های دو نمونه تصادفی با حجم‌های  $n_1=25$  و  $n_2=31$  از دو جامعه نرمال با واریانس‌های  $\sigma_1^2=10$  و  $\sigma_2^2=15$  باشند مطلوبست محاسبه احتمال

$$P\left(\frac{S_1^2}{S_2^2} > 1.26\right)?$$

$$P\left(\frac{S_1^2}{S_2^2} > 1.26\right) = P\left(\frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} > \frac{15}{10} \times 1.26\right) =$$

$$P(F(24, 30) > 1.89) = 1 - P(F(24, 30) < 1.89) = 1 - 0.95 = 0.05$$

**توزیع آماره  $S_1^2/S_2^2$ :** در بسیاری از مواقع لازم است که واریانس‌های دو جامعه با یکدیگر مقایسه شوند برای این منظور از قضیه زیر استفاده می‌کنیم:

**قضیه:** توزیع آماره  $S_1^2/S_2^2$  مشخص نیست اما اگر  $n_1$  و  $n_2$  دو حجم نمونه مستقل از دو جامعه نرمال با واریانس‌های  $\sigma_1^2$  و  $\sigma_2^2$  باشند آنگاه توزیع آماره

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$$

توزیع  $F$  با  $v_1=n_1-1$  و  $v_2=n_2-1$  درجه آزادی است.

### توزیع نمونه میانگین

**قضیه:** اگر  $X_1, X_2, \dots, X_n$  یک نمونه تصادفی "با جایگذاری" از جامعه‌ای متناهی با میانگین  $\mu$  و واریانس متناهی  $\sigma^2$  باشد آنگاه توزیع نمونه  $\bar{X}$ :

وقتی  $n \geq 30$  باشد، بدون توجه به شکل توزیع جامعه، نرمال با میانگین  $\mu$  و واریانس  $\sigma^2/n$  می‌باشد یعنی  $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  نرمال استاندارد است.

ولی اگر  $n < 30$  باشد، در صورت نرمال بودن توزیع جامعه، نرمال با میانگین  $\mu$  و واریانس  $\sigma^2/n$  می‌باشد.

### توزیع نمونه میانگین

**قضیه:** اگر  $X_1, X_2, \dots, X_n$  یک نمونه تصادفی از یک توزیع نرمال با میانگین  $\mu$  و واریانس متناهی  $\sigma^2$  باشد آنگاه توزیع نمونه  $\bar{X}$  نرمال با میانگین  $\mu$  و واریانس  $\sigma^2/n$  می‌باشد.

### توزیع نمونه میانگین

**قضیه حد مرکزی:** اگر  $X_1, X_2, \dots, X_n$  یک نمونه تصادفی (چه با جایگذاری و چه بدون جایگذاری) از جامعه‌ای نامتناهی (بزرگ) با میانگین  $\mu$  و واریانس متناهی  $\sigma^2$  باشد آنگاه شکل حدی توزیع:

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

وقتی  $n \rightarrow \infty$  ( $n \geq 30$ ) باشد، بدون توجه به شکل توزیع جامعه، نرمال استاندارد است یعنی توزیع نمونه  $\bar{X}$  نرمال با میانگین  $\mu$  و واریانس  $\sigma^2/n$  می‌باشد.

### توزیع نمونه میانگین

**قضیه:** اگر  $X_1, X_2, \dots, X_n$  یک نمونه تصادفی "بدون جایگذاری" از جامعه‌ای متناهی به حجم  $N$  با میانگین  $\mu$  و واریانس متناهی  $\sigma^2$  باشد آنگاه توزیع نمونه:

وقتی  $n \geq 30$  و  $N \geq 2n$  باشد، بدون توجه به شکل توزیع جامعه، توزیع نمونه نرمال با میانگین  $\mu$  و واریانس  $(\sigma^2/n)[(N-n)(N-1)]$  می‌باشد.

ولی اگر  $n < 30$  باشد، در صورت نرمال بودن توزیع جامعه، نرمال با میانگین  $\mu$  و واریانس  $(\sigma^2/n)[(N-n)(N-1)]$  می‌باشد.

**مثال:** یک جامعه متناهی به حجم ۸۵ با میانگین ۱.۵ و واریانس ۱۶ مفروض است یک نمونه تصادفی ۶۴ تایی از این جامعه، {الف: با جایگذاری} ب: بدون جایگذاری} انتخاب می‌کنیم مطلوب است  $P(0.5 < \bar{X} < 2)$  در هر دو حالت؟

**حل:** با توجه به حجم نمونه ( $n > 30$ ) توزیع نمونه  $\bar{X}$  نرمال است اما در حالت الف توزیع آن نرمال با میانگین  $\mu$  و واریانس  $\sigma^2/n$  می‌باشد ولی در حالت ب توزیع آن نرمال با میانگین  $\mu$  و واریانس  $(\sigma^2/n)(N-n/N-1)$  می‌باشد پس

$$\mu = 1.5, \quad n = 64, \quad N = 124, \quad \sigma^2 = 16$$

$$A: \bar{X} : n(\mu, \sigma/\sqrt{n}) = n(1.5, 0.5)$$

$$P(0.5 < \bar{X} < 2) = P\left(\frac{0.5-1.5}{0.5} < \frac{\bar{X}-1.5}{0.5} < \frac{2-1.5}{0.5}\right) = P(-2 < Z < 1) = ?$$

$$B) \bar{X} : n(\mu, \frac{\sigma}{\sqrt{n}} \sqrt{(N-n)/(N-1)}) = n(1.5, 0.5 \times 0.5)$$

$$P(0.5 < \bar{X} < 2) = P\left(\frac{0.5-1.5}{0.25} < \frac{\bar{X}-1.5}{0.25} < \frac{2-1.5}{0.25}\right)$$

**مثال:** میانگین طول عمر یک نوع لامپ الکتریکی ۷۸۰ ساعت و انحراف معیار آن ۳۶ ساعت است. یک نمونه تصادفی ۳۶ تایی از این نوع لامپ خریداری می‌شود، احتمال اینکه میانگین طول عمر این نمونه، از ۷۷۴ ساعت بیشتر باشد، چقدر است.

حل: طبق قضیه حد مرکزی حل می‌کنیم

$$\mu_{\bar{X}} = \mu = 780, \quad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{(36)^2}{36} \Rightarrow \sigma_{\bar{X}} = 6$$

$$P(\bar{X} > 774) = 1 - P(\bar{X} < 774) =$$

$$1 - P\left(\frac{\bar{X} - 780}{6} < \frac{774 - 780}{6}\right) = 1 - P(Z < -1) = 1 - 0.1587 = 0.8413$$

### توزیع نمونه میانگین

هنگام استفاده از قضیه حد مرکزی هرگاه واریانس جامعه معلوم نباشد میتوان به جای واریانس جامعه ( $\sigma^2$ ) از واریانس نمونه ( $S^2$ ) استفاده کرد اما در این حالت توزیع  $\left[ \frac{\bar{X} - \mu}{S / \sqrt{n}} \right]$  اگر  $n \geq 30$  باشد، توزیع نرمال استاندارد است.  
ولی اگر  $n < 30$  باشد، دارای توزیع  $t$  با  $n-1$  درجه آزادی است..

**مثال:** تولیدکننده‌ای ادعا می‌کند که محصولات کارخانه‌اش دارای حد متوسط ۱۸٫۳ کیلوگرم است، یک نمونه ۸ تایی از محصولات کارخانه را وزن کردیم که میانگین واریانس نمونه آن به ترتیب، ۱۹٫۵ و ۴٫۲۸ شد، آیا با ادعای کارخانه‌دار موافق هستید؟

حل: چون  $n < 30$  و واریانس جامع معلوم نمی‌باشد پس برای استفاده از واریانس نمونه به جای واریانس جامعه باید از توزیع  $t$  استفاده کنیم یعنی توزیع نمونه، توزیع  $t$  با ۷ درجه آزادی است

$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}} = \frac{19.5 - 18.3}{0.73} = 1.64$$

$$\text{From table: } (t_{0.975}, \nu = 7) = 2.36 \Rightarrow$$

$$P(-2.36 < T < 2.36) = 2P(T < 2.36) - 1 = 0.95$$

$$-2.36 < 1.64 < 2.36$$

پس با احتمال ۹۵٪ ادعای کارخانه دار درست است.

**آماره  $(\bar{X}_1 - \bar{X}_2)$ :** در بسیاری از مواقع لازم است که میانگینهای دو جامعه با یکدیگر مقایسه شوند برای این منظور قضیه زیر بیان می‌گردد.

**قضیه:** اگر نمونه‌های تصادفی با حجم‌های  $n_1$  و  $n_2$  از دو جامعه با میانگین‌های  $\mu_1$  و  $\mu_2$  و واریانس‌های  $\sigma_1^2$  و  $\sigma_2^2$  انتخاب شوند آنگاه توزیع نمونه  $\bar{X}_1$

اگر  $n_1, n_2 \geq 30$ ، بدون توجه به توزیع دو جامعه، نرمال با میانگین  $\mu_1 - \mu_2$  و واریانس  $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$  خواهد بود وگرنه در صورت نرمال بودن دو جامعه، نرمال با میانگین و واریانس ذکر شده خواهد بود.

**مثال:** دو نمونه‌گیری از دو جامعه نرمال طبق جدول زیر انجام می‌دهیم مطلوبست

جامعه اول	جامعه دوم
$\mu_1 = 80$	$\mu_2 = 75$
$\sigma_1 = 5$	$\sigma_2 = 3$
$n_1 = 25$	$n_2 = 36$

$$P(3.4 < \bar{X}_1 - \bar{X}_2 < 8.9) = ?$$

$$\bar{X}_1 - \bar{X}_2 \text{ is Normal \& } \mu_{\bar{X}_1 - \bar{X}_2} = 80 - 75,$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{25/25 + 9/36} = \sqrt{5/4}$$

$$P(3.4 < \bar{X}_1 - \bar{X}_2 < 8.9) = P\left(\frac{3.4 - 5}{\sqrt{5/4}} < Z < \frac{8.9 - 5}{\sqrt{5/4}}\right) =$$

$$P(-1.34 < 3.48) = ?$$

**قضیه:** به ازای  $n$ های بزرگ، توزیع نمونه‌ای میانه نمونه تصادفی به اندازه  $2n+1$ ، تقریباً نرمال با میانگین  $\mu$  و واریانس  $\pi\sigma^2/4n$  می‌باشد

اگر واریانس میانه را برای نمونه‌هایی به اندازه  $2n+1$  از یک جامعه نامتناهی با واریانس میانگین نمونه یعنی  $(\sigma^2/2n+1)$  مقایسه کنیم در می‌یابیم که در نمونه‌های بزرگ از یک جامعه نرمال میانگین قابل اعتمادتر از میانه است.

**آماره ترتیب:** اگر  $X_1, X_2, \dots, X_n$  یک نمونه تصادفی از متغیر تصادفی  $X$  باشند و آنها را به طور غیر نزولی مرتب کنیم

$$Y_1 \leq Y_2 \leq \dots \leq Y_n$$

(که  $Y_i$ ها مرتب شده  $X_i$ ها هستند) آنگاه  $Y_i$  را  $i$ امین آماره ترتیب می‌نامیم که میانه نمونه اگر داده‌ها فرد باشند برابر است با داده وسطی و اگر زوج باشند برابر با میانگین حسابی دو داده

وسطی است. و برد نمونه نیز برابر است با  $R = Y_n - Y_1$

**تمرین:** میانگین نمرات تست هوش دانشجویان سال اول يك دانشكده ۵۴۰ و انحراف معیار آن ۵۰ است دو نمونه تصادفی با حجم  $n_1=32$  و  $n_2=50$  انتخاب می‌کنیم مطلوبست احتمال اینکه تفاضل میانگین نمرات این دو نمونه

الف: بیش از ۲۰ باشد

ب: بین ۵ و ۱۰ باشد

**تمرین:** از يك جامعه نرمال با ادعای میانگین ۲۰ و واریانس مجهول، يك نمونه تصادفی ۹ تایی با میانگین و انحراف معیار ۲۴ و ۴,۱ انتخاب می‌کنیم آیا ادعای میانگین ۲۰ مورد قبول است؟

**تمرین:** از جامعه ای نرمال با میانگین  $43/2$  و واریانس  $39/69$  يك نمونه تصادفی ۳۶ تایی انتخاب می‌کنیم، احتمال اینکه میانگین این نمونه بین ۴۲ و ۴۵ باشد چقدر است؟

**تمرین:** طول قد ۳۰۰۰ كودك دارای توزیع نرمال با میانگین ۶۸ و انحراف معیار ۳ می‌باشد، اگر ۸۰ نمونه تصادفی ۲۵ تایی از این جامعه انتخاب کنیم، میانگین و واریانس  $\bar{X}$  را حساب کنید در صورتیکه نمونه برداری

الف با جایگذاری باشد ب: بدون جایگذاری باشد

پ: چند میانگین نمونه در فاصله  $68/8 - 66/3$  قرار می‌گیرند

ت: چند میانگین نمونه کمتر از  $66/4$  قرار گیرند؟

**تمرین:** اگر  $S_1^2$  و  $S_2^2$  به ترتیب واریانس‌های دو نمونه تصادفی مستقل از دو جامعه نرمال باشند، در صورتیکه واریانس‌های دو جامعه مساوی بوده و نمونه‌های با حجم  $n_1=8$  و  $n_2=12$  انتخاب شده باشند، مطلوبست محاسبه

$$P(S_1^2 > 4.89 S_2^2)$$

## فصل دهم: نظریه برآورد کردن (تئوری تخمین)

در این فصل هدف برآورد پارامترهای جامعه از روی آماره‌ها و قضایای معرفی شده در فصل قبل می‌باشد.

**نظریه تصمیم‌گیری:** به روشهایی اطلاق می‌شود که به کمک آن از نمونه منتخب جامعه، بتوان استنباطهایی را در مورد پارامترهای جامعه به دست آورد. که دارای دو شاخه اصلی می‌باشد:

• نظریه برآورد کردن (تئوری تخمین)

• آزمون فرض‌ها

پارامترهای جامعه را به دو طریق می‌توان برآورد کرد (تخمین زد):

**الف: برآورد نقطه‌ای:** مقدار يك آماره را برای پارامتر يك جامعه به کار می‌بریم.

**ب: برآورد فاصله‌ای:** پارامترهای جامعه را به صورت يك بازه قابل اعتماد برآورد می‌کنیم.

### برآورد نقطه‌ای

به طور کلی پارامتر جامعه را با  $\theta$ ، يك برآورد نقطه‌ای آن را  $\hat{\theta}$  و آماره‌ای که برای بررسی این پارامتر استفاده خواهد شد  $\hat{\theta}$  با نشان می‌دهیم به عنوان مثال اگر  $\mu$  میانگین جامعه مورد بررسی باشد:

$$\hat{\theta} = \bar{X}, \quad \hat{\theta} = \bar{X}, \quad \theta = \mu$$

**تعریف:** برآوردگر  $\hat{\theta}$  را يك برآوردگر ناريب پارامتر  $\theta$  گوييم اگر و فقط اگر

$$E(\hat{\theta}) = \theta$$

براي مثال:  $\bar{X}$  و  $S^2$  به ترتيب برآورد كننده هاي ناريب براي میانگين و واريانس جامعه خواهد بود زیرا:

$$E(\bar{X}) = \mu, \quad E(S^2) = \sigma^2$$

**تعریف:** از میان برآوردگرهاي ناريب پارامتر  $\theta$ ، برآوردي که کمترین واريانس را داشته باشد، کاراترين برآوردگر ناميده مي شود.

**برآوردگر:** آماره اي را که براي تخمين نقطه اي پارامتر ي استفاده شود، برآوردگر مي ناميم يك پارامتر ممکن است داراي چند برآوردگر باشد براي مثال براي  $\theta = \mu$  سه برآوردگر شناخته شده میانگين، ميانه و مد نمونه، وجود دارد حال اين سؤال مطرح است که بين برآوردگرهاي يك پارامتر، کداميك بهتر است؟

براي جواب دادن به سؤال بالا تعريف زیر را انجام مي دهيم

**قضيه:** اگر  $\hat{\theta}$  يك برآوردگر ناريب پارامتر  $\theta$  باشد و

$$\text{var}(\hat{\theta}) = \frac{1}{nE\left[\left(\frac{\partial \ln f(x)}{\partial \theta}\right)^2\right]}$$

آنگاه  $\hat{\theta}$  يك برآوردگر ناريب با کمترین واريانس از پارامتر  $\theta$  است.

**مثال:** در برآورد میانگين يك جامعه نرمال بر مبناي يك نمونه تصادفي به اندازه  $2n+1$ ، کاري ميانه نسبت به میانگين چيست؟

حل: مي دانيم که هر دو برآورد ناريب است ولي

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{2n+1}, \quad \text{Var}(\bar{X}) = \frac{\pi\sigma^2}{4n}$$

$$\frac{\text{Var}(\bar{X})}{\text{Var}(\bar{X})} = \frac{2n+1}{4n} = \Rightarrow$$

$$\lim_{n \rightarrow \infty} \frac{4n}{\pi(2n+1)} = \frac{2}{\pi} = 0.64 < 1$$

**تعریف:** آماره  $\hat{\theta}$  يك برآوردکننده سازگار پارامتر  $\theta$  است اگر و فقط اگر به ازاي هر ثابت  $c$ ,

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \geq c) = 0 \quad \text{OR} \quad \lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| < c) = 1$$

**قضيه:** آماره  $\hat{\theta}$  يك برآوردکننده سازگار پارامتر  $\theta$  است اگر

الف:  $\hat{\theta}$  ناريب باشد

$$\text{b: } \lim_{n \rightarrow \infty} \text{var}(\hat{\theta}) = 0$$

**مثال:** نشان دهيد که میانگين نمونه يك برآورد کننده ناريب با کمترین واريانس میانگين جامعه نرمال است؟

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

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

$$E\left(\left(\frac{\partial \ln f(x)}{\partial \mu}\right)^2\right) = \frac{1}{\sigma^2} E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma^2} \cdot 1 = \frac{1}{\sigma^2}$$

$$\frac{1}{nE\left[\left(\frac{\partial \ln f(x)}{\partial \theta}\right)^2\right]} = \frac{1}{n} \frac{1}{\sigma^2} = \frac{\sigma^2}{n}$$

$$\bar{X} \text{ is unbiased \& } V(\bar{X}) = \frac{\sigma^2}{n}$$

**مثال:** واریانس نمونه  $(S^2)$  یک برآوردگر سازگار است زیرا:

$$E(S^2) = \sigma^2 \Rightarrow S^2 \text{ is unbiased}$$

$$\lim_{n \rightarrow \infty} \text{Var}(S^2) = \lim_{n \rightarrow \infty} \frac{2\sigma^4}{n-1} = 0$$

شرط گفته شده برای سازگاری برآوردگر یک شرط کافی است نه لازم یعنی برآورد کننده ای می تواند سازگار باشد بدون اینکه ناریب باشد. ولی در این حالت نیز باید مجاناً ناریب باشد. برای مثال برآوردکننده

یک برآوردکننده پارامتر دو جمله ای است ولی

$$E\left[\frac{(X+1)}{(n+2)}\right] = \frac{1}{(n+2)} E(X+1) = \frac{1}{(n+2)} (E(X)+1) \neq \theta$$

$$\lim_{n \rightarrow \infty} E\left[\frac{(X+1)}{(n+2)}\right] = E \lim_{n \rightarrow \infty} \frac{(X+1)}{(n+2)} = \theta$$

**تعریف:** آماره  $\hat{\theta}$  یک برآوردگر کافی پارامتر  $\theta$  است اگر و تنها اگر با ازای هر مقدار  $\hat{\theta}$ ، توزیع شرطی نمونه تصادفی  $X_1, X_2, \dots, X_n$  به  $\hat{\theta}$  مستقل از  $\theta$  باشد.

**قضیه:** آماره  $\hat{\theta}$  یک برآوردگر کافی پارامتر  $\theta$  است اگر و تنها اگر، بتوان تابع احتمال توأم نمونه تصادفی را تجزیه کرد به طوری که

$$f(x_1, x_2, \dots, x_n | \theta) = g(\hat{\theta}, \theta) \cdot h(x_1, x_2, \dots, x_n)$$

که در آن  $g$  تنها به  $\theta$  و  $\hat{\theta}$  بستگی دارد و  $h$  به  $\theta$  بستگی ندارد.

**مثال:** نشان دهید که میانگین نمونه یک برآوردگر کافی میانگین جامعه نرمال با واریانس معلوم  $\sigma^2$  است؟

$$f(x_1, x_2, \dots, x_n | \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2}$$

$$\left\{ \sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \right\}$$

$$f(x_1, x_2, \dots, x_n | \mu) = \left[ \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2} e^{-\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i - \bar{X}}{\sigma}\right)^2} \right]$$

برای برآورد پارامترهای جامعه به روش برآورد نقطه‌ای روشهای متعددی وجود دارد که به دو روش مهم آن اشاره می‌کنیم:

**روش گشتاورها:** عبارتست از حل دستگاه معادلات زیر:

$$M'_k = \mu'_k \quad k = 1, 2, \dots, p$$

که در آن  $p$  تعداد پارامتر جامعه است.

**مثال:** نمونه‌ای تصادفی به اندازه  $n$  از جامعه گاما داریم، برای برآورد کردن پارامترهای  $\alpha$  و  $\beta$  روش گشتاورها را به کار برید؟

$$M'_1 = \mu'_1, \quad M'_2 = \mu'_2$$

$$\left\{ \mu'_1 = \alpha\beta, \quad \mu'_2 = \alpha(\alpha+1)\beta^2 \right\}$$

$$\hat{\alpha} = \frac{(M'_1)^2}{M'_2 - (M'_1)^2} \quad \hat{\beta} = \frac{M'_2 - (M'_1)^2}{M'_1}$$

**روش درستنمایی ماکسیمم:** در این روش مقادیر یک نمونه تصادفی را در نظر می‌گیریم، سپس به عنوان برآورد خود از پارامتر مجهول جامعه، مقداری را بر می‌گزینیم که به ازای آن احتمال به دست آوردن داده‌های مشاهده شده، ماکسیمم گردد یعنی ماکسیمم تابع

$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta)$$

را نسبت به  $\theta$  که تابع درستنمایی نامیده می‌شود، برآورد درستنمایی ماکسیمم  $\theta$  می‌نامیم.

### برآورد فاصله اي

**مثال:** با مفروض بودن  $x$  پیروزي در  $n$  امتحان برآوردگر درستتمایي ماکسیم پارامتر  $\theta$  را در توزیع دو جمله‌اي به دست آورید؟

$$L(\theta) = b(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$\ln L(\theta) = \ln \binom{n}{x} + x \ln \theta + (n-x) \ln(1-\theta)$$

$$\frac{d \ln L(\theta)}{d \theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0 \Rightarrow \theta = \frac{x}{n}$$

$$\hat{\theta} = \frac{x}{n}$$

### برآورد فاصله‌اي میانگین جامعه ( $\mu$ ):

وقتي نمونه تصادفي با حجم  $n$  از جامعه‌اي با واریانس معلوم  $\sigma^2$  انتخاب شود فاصله اطمینان برای  $\mu$  به شکل زیر محاسبه می‌گردد (استفاده از قضیه حد مرکزی):



$$P(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}) = 1-\alpha \Rightarrow$$

$$P(-Z_{1-\alpha/2} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < Z_{1-\alpha/2}) = 1-\alpha \Rightarrow$$

$$\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$$

**برآورد فاصله اي:** فرض کنیم  $1-\alpha$  احتمال مشخص ولي بزرگي باشد (۰/۹۵ یا ۰/۹۹ یا ... )  $(0 < \alpha < 1)$ ، یک برآورد فاصله اي برای پارامتری مانند  $\theta$  عبارتست از فاصله‌ي بازی مانند  $(L, U)$ ، تابعی از نمونه تصادفي  $X_1, X_2, \dots, X_n$ ، که با احتمال  $1-\alpha$  داشته باشیم:

$$P(L < \theta < U) = 1-\alpha$$

$1-\alpha$  را **سطح اطمینان** (ضریب اطمینان) می‌نامیم و معمولاً با درصد بیان می‌شود. پس در برآورد فاصله اي باید فاصله ي بازی مانند  $(L, U)$  را بیابیم که با احتمال مشخص  $1-\alpha$ ، تعریف بالا برقرار باشد.

### نکته‌هاي کاربردي:

• با توجه به رابطه گفته شده طول فاصله اطمینان برابر است با:

$$L = 2 \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$$

پس برای اینکه فاصله اطمینان خوبی داشته باشیم باید حجم نمونه را زیاد کنیم.

• محاسبه خطا در برآورد  $\mu$ :

$$-\frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} < \mu - \bar{X} < \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \Rightarrow |\mu - \bar{X}| < \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$$

### نکته‌هاي کاربردي:

• اگر نمونه تصادفي دارای حجم  $n \geq 30$  باشد، بدون توجه به نوع توزیع جامعه می‌توان از فرمول بالا استفاده نمود.

• اگر نمونه تصادفي دارای حجم  $n < 30$  باشد، باید توزیع جامعه نرمال باشد تا بتوان از فرمول بالا استفاده نمود.

• اگر نمونه تصادفي دارای حجم  $n \geq 30$  باشد، در صورت نامعلوم بودن واریانس جامعه می‌توان به جای  $\sigma^2$  از واریانس نمونه ( $S^2$ ) استفاده نمود.

### نکته‌های کاربردی:

• حداکثر خطا در برآورد  $\mu$  برابر است با:

$$e = \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$$

• می‌توان گفت خطا کمتر از مقدار  $e$  است اگر حجم نمونه از

رابطه زیر به دست آید:

$$n = \left[ \frac{\sigma}{\sqrt{e}} Z_{1-\alpha/2} \right]^2$$

**مثال:** فرض می‌کنیم عمر کامپیوترهای تولیدی یک موسسه از توزیع نرمال با میانگین  $\mu$  و واریانس  $\sigma^2 = 0.25$  سال پیروی می‌کند. بر اساس نمونه‌ی تصادفی از این جامعه مقادیر زیر حاصل شده است:

$$10 - 7 - 6 - 5 - 8 - 5 - 9 - 6 - 7$$

یک فاصله اطمینان ۹۵٪ برای  $\mu$  بیابید. مقدار حداکثر خطا را برای این فاصله اطمینان محاسبه نمایید اگر بخواهیم با اطمینان ۹۵٪ خطای کمتر از ۰.۲۵ داشته باشیم حجم نمونه را چه تعداد باید انتخاب کنیم.

$$\bar{X} = \frac{\sum X_i}{n} = \frac{63}{9} = 7$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow Z_{1-\alpha/2} = Z_{0.975} = 1.96$$

$$\mu: 7 \pm (1.96) \left( \frac{0.5}{\sqrt{9}} \right) \Rightarrow \mu: (6.67, 7.33)$$

$$7 - (1.96) \left( \frac{0.5}{\sqrt{9}} \right) < \mu < 7 + (1.96) \left( \frac{0.5}{\sqrt{9}} \right) \Rightarrow 6.67 < \mu < 7.33$$

$$e = \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} = \frac{0.5}{3} \times 1.96 = 0.64$$

$$n = \left[ \frac{\sigma}{\sqrt{e}} Z_{1-\alpha/2} \right]^2 = \left( \frac{0.5}{0.25} \times 1.96 \right)^2 = (3.92)^2 = 15.36 \approx 16$$

### برآورد فاصله‌های میانگین جامعه ( $\mu$ ):

وقتی نمونه تصادفی با حجم  $n < 30$  از جامعه‌ای نرمال با واریانس نامعلوم  $\sigma^2$  انتخاب شود فاصله اطمینان برای به شکل زیر محاسبه می‌گردد:



$$P(t_{1-\alpha/2, n-1} < T < t_{\alpha/2, n-1}) = 1 - \alpha$$

$$P\left(t_{1-\alpha/2, n-1} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < t_{\alpha/2, n-1}\right) = 1 - \alpha \Rightarrow$$

$$\bar{X} - \frac{\sigma}{\sqrt{n}} t_{1-\alpha/2, n-1} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} t_{1-\alpha/2, n-1}$$

$$e = \frac{\sigma}{\sqrt{n}} t_{1-\alpha/2, n-1}, \quad n = \left[ \frac{\sigma}{\sqrt{e}} t_{1-\alpha/2, n-1} \right]^2$$

**مثال:** یک نمونه تصادفی ۱۰ تایی با میانگین نمونه ۴/۳۸ و انحراف معیار نمونه ۰/۰۶ از یک جامعه نرمال با میانگین  $\mu$  و واریانس نامعلوم  $\sigma^2$  انتخاب می‌کنیم مطلوبست یک فاصله اطمینان ۹۵٪ برای میانگین جامعه:

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$$

$$t_{1-\alpha/2, n-1} = t_{0.975, 9} = 2.26$$

$$\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2, n-1} < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2, n-1}$$

$$4.38 - \frac{0.06}{\sqrt{10}} \times 2.26 < \mu < 4.38 + \frac{0.06}{\sqrt{10}} \times 2.26$$

$$4.338 < \mu < 4.42$$

### برآورد فاصله‌های تفاضل میانگین دو جامعه ( $\mu_1 - \mu_2$ ):

**حالت اول:** فرض می‌کنیم دو نمونه‌ی تصادفی مستقل از هم به حجم  $n_1$  و  $n_2$  از دو جامعه‌ی نرمال با میانگین‌های  $\mu_1$  و  $\mu_2$  و واریانس‌های معلوم  $\sigma_1^2$  و  $\sigma_2^2$  استخراج کرده باشیم یک برآورد فاصله‌ای برای تفاضل میانگین دو جامعه ( $\mu_1 - \mu_2$ ) با ضریب اطمینان  $1 - \alpha$  به صورت زیر محاسبه می‌گردد:

$$(\bar{X}_1 - \bar{X}_2) - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} Z_{1-\alpha/2} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} Z_{1-\alpha/2}$$

$$\mu_1 - \mu_2: (\bar{X}_1 - \bar{X}_2) \pm \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} Z_{1-\alpha/2}$$

### نکته‌های کاربردی:

• اگر نمونه‌های تصادفی دارای حجم  $n_1, n_2 \geq 30$  باشد، بدون توجه به توزیع دو جامعه می‌توان از فرمول بالا استفاده نمود.

• اگر نمونه‌های تصادفی دارای حجم  $n_1, n_2 < 30$  باشد، باید توزیع دو جامعه نرمال باشد تا بتوان از فرمول بالا استفاده نمود.

• اگر نمونه تصادفی دارای حجم  $n_1, n_2 \geq 30$  باشد، در صورت نامعلوم بودن واریانس‌های جامعه می‌توان به جای  $\sigma_1^2$  و  $\sigma_2^2$  از واریانس‌های نمونه ( $S_1^2$  و  $S_2^2$ ) استفاده نمود.

### حالت دوم:

• اگر نمونه‌های تصادفی دارای حجم  $n_1, n_2 < 30$  از دو جامعه نرمال باشد در صورت نامعلوم بودن واریانس‌های جامعه ولی با فرض برابری واریانس‌ها ( $\sigma_1^2 = \sigma_2^2$ ) برآورد تفاضل میانگین‌ها به شکل زیر در می‌آید:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \leftarrow \text{واریانس ادغام شده}$$

$$\mu_1 - \mu_2 : (\bar{X}_1 - \bar{X}_2) \pm S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{1-\alpha/2, n_1+n_2-2}$$

### نکته‌های کاربردی

• چنانچه واریانس‌های دو جامعه مساوی نباشند ولی  $n_1 = n_2$  باز می‌توان از رابطه بالا استفاده نمود.

• چنانچه واریانس‌های دو جامعه مساوی نباشند و امکان انتخاب دو نمونه با حجم‌های مساوی نباشند می‌توانیم از رابطه زیر استفاده کنیم:

$$v \approx \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1} - \frac{S_2^2}{n_2}\right)^2 / n_1 - 1 + \left(\frac{S_2^2}{n_2} - \frac{S_1^2}{n_1}\right)^2 / n_2 - 1}$$

$$\mu_1 - \mu_2 : (\bar{X}_1 - \bar{X}_2) \pm \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} t_{1-\alpha/2, v}$$

**مثال:** از دو جامعه نرمال با میانگین‌های  $\mu_1$  و  $\mu_2$  و واریانس‌های  $\sigma_1^2 = 25$  و  $\sigma_2^2 = 9$  دو نمونه تصادفی مستقل با حجم‌های  $n_1 = 25$  و  $n_2 = 36$  انتخاب می‌کنیم اگر میانگین نمونه اول ۸۰ و نمونه دوم ۷۵ باشد مطلوبست فاصله اطمینان ۹۴٪ برای  $\mu_1 - \mu_2$ ؟

$$1 - \alpha = 0.94 \Rightarrow 1 - \alpha/2 = 0.97, Z_{0.97} = 1.88$$

$$\mu_1 - \mu_2 : (\bar{X}_1 - \bar{X}_2) \pm \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} Z_{1-\alpha/2}$$

$$\mu_1 - \mu_2 : (80 - 75) \pm \sqrt{\frac{25}{25} + \frac{9}{36}} \times 1.88$$

$$2.9 < \mu_1 - \mu_2 < 7.1$$

**مثال:** از دو جامعه نرمال با واریانس‌های مساوی  $\sigma_1^2 = \sigma_2^2 = 6$ ، دو نمونه تصادفی مستقل  $n_1 = 9$  و  $n_2 = 16$  انتخاب کرده‌ایم و نتایج زیر به دست آمده است.

$$\bar{X}_1 = 64, S_1^2 = 36, \bar{X}_2 = 59, S_2^2 = 25$$

۹۵٪ فاصله اطمینان را برای  $\mu_1 - \mu_2$  بدست آورید.

$$1 - \alpha = 0.95 \Rightarrow 1 - \alpha/2 = 0.975$$

$$v = n_1 + n_2 - 2 = 23, t_{0.975, 23} = 2.07$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{1}{23}(8 \times 36 + 15 \times 25) = 28.83$$

$$\mu_1 - \mu_2 : (\bar{X}_1 - \bar{X}_2) \pm S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{1-\alpha/2, n_1+n_2-2} \Rightarrow$$

$$\mu_1 - \mu_2 : (64 - 59) \pm 5.37 \sqrt{\frac{1}{9} + \frac{1}{25}} \times 2.07 \Rightarrow$$

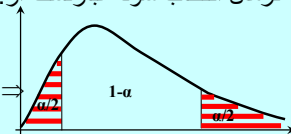
$$0.68 < \mu_1 - \mu_2 < 9.32$$

### برآورد فاصله‌ای واریانس جامعه:

برآورد فاصله‌ای با ضریب اطمینان  $(1 - \alpha)$  برای واریانس جامعه ( $\sigma^2$ ) هنگامیکه نمونه تصادفی با حجم  $n$  از یک جامعه نرمال انتخاب شود عبارتست از:

$$X^2 = \frac{(n-1)S^2}{\sigma^2}$$

$$P(\chi_{\alpha/2}^2 < X^2 < \chi_{1-\alpha/2}^2) = 1 - \alpha$$



$$\frac{(n-1)S^2}{\chi_{1-\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{\alpha/2}^2}$$

**مثال:** یک نمونه تصادفی با حجم  $n = 20$  از یک جامعه نرمال دارای میانگین نمونه  $32/8$  و انحراف معیار  $4/51$  انتخاب می‌کنیم. یک فاصله اطمینان ۹۵٪ برای  $\sigma^2$  و  $\sigma$  به دست آورید؟

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05, \alpha/2 = 0.025, 1 - \alpha/2 = 0.975$$

$$\chi_{0.975, 19}^2 = 32.9, \chi_{0.025, 19}^2 = 8.91$$

$$\frac{(n-1)S^2}{\chi_{1-\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{\alpha/2}^2}$$

$$\frac{19 \times (4.51)^2}{32.9} < \sigma^2 < \frac{19 \times (4.51)^2}{8.91} \Rightarrow$$

$$11.75 < \sigma^2 < 43.37$$

$$3.42 < \sigma < 6.58$$

### برآورد فاصله‌های نسبت واریانس‌های دو جامعه:

برآورد فاصله‌های با ضریب اطمینان  $(1-\alpha)$  برای  $(\sigma_1^2/\sigma_2^2)$ ، وقتی دو نمونه تصادفی مستقل با حجم‌های  $n_1$  و  $n_2$  از دو جامعه نرمال با واریانس‌های  $\sigma_1^2$  و  $\sigma_2^2$  انتخاب شوند، عبارتست از:

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

$$P(f_{\alpha/2}(n_1-1, n_2-1) < F < f_{1-\alpha/2}(n_1-1, n_2-1)) = 1-\alpha \Rightarrow$$



$$\frac{S_1^2}{S_2^2} \cdot \frac{1}{f_{1-\alpha/2}(n_1-1, n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{1-\alpha/2}(n_2-1, n_1-1)$$

**مثال:** يك فاصله اطمینان ۹۸٪ برای نسبت  $(\sigma_1^2/\sigma_2^2)$  و  $(\sigma_1/\sigma_2)$  از دو جامعه نرمال تعیین کنید در صورتیکه واریانس‌های دو نمونه تصادفی با حجم‌های  $n_1=9$  و  $n_2=16$  از این دو جامعه به ترتیب  $S_1^2=36$  و  $S_2^2=25$  باشد؟

$$1-\alpha = 0.98 \Rightarrow \alpha = 0.02, 1-\alpha/2 = 0.99$$

$$f_{0.99}(8, 15) = 4, f_{0.99}(15, 8) = 5.52$$

$$\frac{S_1^2}{S_2^2} \cdot \frac{1}{f_{1-\alpha/2}(n_1-1, n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{1-\alpha/2}(n_2-1, n_1-1)$$

$$\frac{36}{25} \times \frac{1}{4} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{36}{25} \times 5.52$$

$$0.36 < \frac{\sigma_1^2}{\sigma_2^2} < 7.59$$

$$0.6 < \frac{\sigma_1}{\sigma_2} < 2.81$$

### برآورد فاصله‌های نسبت p (نسبت در توزیع دو جمله‌ای):

برآورد فاصله‌های برای نسبت p با ضریب اطمینان  $(1-\alpha)$  وقتی حجم نمونه تصادفی  $n \geq 30$  باشد عبارتست:

$$Z = \frac{\hat{P} - p}{\sqrt{\hat{p}\hat{q}/n}} \sim n(0,1) \Rightarrow P(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}) = 1-\alpha$$

$$\hat{P} - \sqrt{\hat{p}\hat{q}/n} Z_{1-\alpha/2} < p < \hat{P} + \sqrt{\hat{p}\hat{q}/n} Z_{1-\alpha/2}$$

در رابطه بالا حداکثر خطا و مقدار حجم نمونه برای اینکه خطایی کمتر از e داشته باشیم از رابطه زیر به دست می‌آید:

$$e = \sqrt{\frac{\hat{p}\hat{q}}{n}} Z_{1-\alpha/2} \quad n = \frac{1}{e^2} \hat{p}\hat{q} (Z_{1-\alpha/2})^2$$

### برآورد فاصله‌های نسبت (p):

یکی از پارامترهایی که معمولاً مورد بررسی قرار می‌گیرد پارامتر p، نسبت یا احتمال موفقیت در آزمایش دو جمله‌ای است. یک برآوردگر نارایب برای این پارامتر برآوردگر  $\hat{P} = \frac{X}{n}$  است. (متغیر تصادفی X تعداد موفقیت در آزمایش دو جمله‌ای است) بنابراین این برآوردگر یک برآورد نقطه‌ای خوب برای نسبت p است می‌دانیم وقتی n بزرگ باشد این برآوردگر توزیع نرمال با میانگین و واریانس زیر می‌باشد

$$\mu_{\hat{P}} = E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p$$

$$\sigma_{\hat{P}}^2 = \text{var}\left(\frac{X}{n}\right) = \frac{1}{n^2} \sigma_X^2 = \frac{npq}{n^2} = \frac{pq}{n}$$

### برآورد تفاضل دو نسبت (p1-p2) در توزیع دو جمله‌ای:

برآورد فاصله‌های برای تفاضل پارامترهای دو جامعه دو جمله‌ای  $(p_1-p_2)$  هنگامیکه دو نمونه تصادفی مستقل با حجم‌های  $n_1, n_2 \geq 30$  انتخاب شوند عبارتست از:

$$\hat{p}_1 = \frac{X_1}{n_1}, \hat{q}_1 = 1 - \hat{p}_1, \hat{p}_2 = \frac{X_2}{n_2}, \hat{q}_2 = 1 - \hat{p}_2$$

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} \sim n(0,1) \Rightarrow P(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}) = 1-\alpha$$

$$p_1 - p_2 : (\hat{P}_1 - \hat{P}_2) \pm \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} Z_{1-\alpha/2}$$

**مثال:** در یک نمونه تصادفی با حجم  $n=640$  مشخص شده است که  $x=160$  خانواده دارای کامپیوتر شخصی هستند، یک فاصله اطمینان ۹۵٪ برای نسبت واقعی تمام خانواده‌های کامپیوتردار تعیین کنید، حداکثر خطا را تعیین کنید و مشخص کنید چقدر باید نمونه داشته باشیم تا خطایی کمتر از ۱٪ داشته باشیم؟

$$1-\alpha = 0.95 \Rightarrow 1-\alpha/2 = 0.975$$

$$\hat{p} = \frac{X}{n} = \frac{160}{640} = 0.25, \hat{q} = 1 - \hat{p} = 0.75$$

$$\hat{P} - \sqrt{\hat{p}\hat{q}/n} Z_{1-\alpha/2} < p < \hat{P} + \sqrt{\hat{p}\hat{q}/n} Z_{1-\alpha/2}$$

$$p : 0.25 \pm \sqrt{0.25 \times 0.75 / 640} \times 1.96 \Rightarrow 0.21 < p < 0.28$$

$$e = \sqrt{\hat{p}\hat{q}/n} Z_{1-\alpha/2} = \sqrt{0.25 \times 0.75 / 640} \times 1.96 = 0.3$$

$$n = \frac{1}{e^2} \hat{p}\hat{q} (Z_{1-\alpha/2})^2 = \frac{1}{(0.01)^2} \times 0.25 \times 0.75 \times (1.96)^2 = 7203$$

**تمرین:** اگر  $X_1, X_2, \dots, X_n$  مقادیر یک نمونه تصادفی از جامعه‌ای نمایی باشند، برآورد درست‌نمایی پارامتر  $\theta$  جامعه را پیدا کنید؟

**تمرین:** فرض می‌کنیم  $X_1, X_2, \dots, X_n$  مقادیر یک نمونه تصادفی از توزیع زیر باشد:

$$f(x, \theta) = \theta^x (1-\theta)^{1-x} : x = 0, 1$$

$\theta$  را به روش MLE برآورد کنید

**تمرین:** فرض می‌کنیم تابع چگالی احتمال  $X$  عبارتست از:

$$f(x, \theta) = \theta e^{-\theta x} : x > 0$$

$\theta$  را به روش گشتاورها برآورد کنید

**تمرین:** از جامعه‌ای نرمال، نمونه تصادفی با حجم  $n=5$  انتخاب و نتایج زیر به دست آمده است.

$X_1=3$	$X_2=1$	$X_3=4$	$X_4=3$	$X_5=4$
---------	---------	---------	---------	---------

فاصله اطمینان ۹۹٪ برای میانگین این جامعه را به دست آورید؟

**تمرین:** در صورتیکه میانگین نمونه یک نمونه تصادفی با حجم  $n$  از یک جامعه نرمال با انحراف معیار ۳، مفروض باشد، تعداد  $n$  را طوری تعیین کنید که ۹۹٪ فاصله اطمینان برای میانگین جامعه به صورت  $(\bar{X} - 0.1, \bar{X} + 0.1)$  باشد؟

**تمرین:** یک نمونه تصادفی  $n_1=15$  تایی از یک جامعه با میانگین نمونه ۶۰ و انحراف معیار نمونه ۳ است و نمونه تصادفی دیگری با حجم  $n_2=21$  تایی از یک جامعه دیگر دارای میانگین نمونه ۵۸ و انحراف معیار نمونه ۲ است، فرض کنید دو جامعه نرمال با واریانس‌های مساوی باشند مطلوبست

الف: فاصله اطمینان ۹۵٪ برای تفاضل میانگین دو جامعه؟

ب: اگر واریانس‌ها نامساوی ولی  $n_1=n_2=10$  و دارای همان میانگین نمونه باشد فاصله اطمینان ۹۵٪ برای تفاضل میانگین دو جامعه؟

ج: اگر واریانس‌ها نامساوی باشند فاصله اطمینان ۹۵٪ برای تفاضل میانگین دو جامعه؟

**تمرین:** یک نمونه تصادفی  $n_1=150$  تایی از یک جامعه نرمال دارای میانگین نمونه ۱۴۰۰ و انحراف معیار نمونه ۱۲۰ است و نمونه تصادفی دیگری با حجم  $n_2=200$  از یک جامعه نرمال دیگر دارای میانگین نمونه ۱۲۰۰ و انحراف معیار نمونه ۸۰ است، مطلوبست فاصله اطمینان ۹۵٪ برای تفاضل میانگین دو جامعه؟

**تمرین:** یک نمونه ۱۶ تایی از محصلان یک مدرسه انتخاب و انحراف معیار را اندازه گرفتیم،  $S=2.4$  سانتی‌متر است. فاصله اطمینان ۹۵٪ برای انحراف معیار دانش‌آموزان را در صورتیکه که جامعه نرمال فرض گردد به دست آورید؟

**تمرین:** توزیع جامعه معدل دانشجویان دو دانشگاه نرمال فرض می‌شوند مطلوبست فاصله اطمینان ۹۸٪ برای نسبت واریانس‌ها و انحراف معیارهای دو جامعه در صورتیکه دو نمونه منتخب تصادفی مستقل از این دو جامعه به شکل زیر باشد؟

A دانشگاه	3	3.2	2.9	3.4	2.5	2.8	3	3.3	2.7	3.1
A دانشگاه	3	2.6	3.1	3.5	3.2	3.9	2.7			

**تمرین:** سکه‌ای را ۴۰ بار پرتاب می‌کنیم و ۲۴ بار شیر آمده است مطلوبست یک فاصله اطمینان ۹۵٪ برای نسبت واقعی شیرها؟

**تمرین:** مدیر کارخانه‌ای که دو نوع لامپ با مارکهای A و B تولید می‌کند، مدعی است که لامپ‌های مارک A، ۸٪ بیشتر از مارک B فروش دارد. یک نمونه تصادفی با حجم  $n_1=200$  نفر از خریداران انتخاب شده که ۴۲ نفر از این افراد از لامپ نوع A استفاده می‌کنند. نمونه تصادفی دیگری با حجم  $n_2=150$  نفر از خریداران انتخاب شده که مشخص شده ۱۸ نفر از لامپ نوع B استفاده می‌کنند یک فاصله اطمینان ۹۴٪ برای تفاضل  $p_1-p_2$  تعیین کنید، آیا با ادعای مدیر کارخانه موافقت؟

همان‌طور که اشاره شد نظریه تصمیم‌گیری (روش‌های رسیدن از نمونه به جامعه) به دو دسته تقسیم می‌شود، یکی نظریه برآورد کردن (تئوری تخمین) و دیگری **آزمون فرض‌ها**.

آزمون فرض‌ها یکی از مهمترین شاخه‌های نظریه تصمیم‌گیری است.

**فرض آماری:** فرض، بیان یا حدسی است درباره‌ی توزیع جامعه یا پارامتر جامعه که درست بودن فرض آماری را باید از نتایج به دست آمده از نمونه‌گیری بررسی نمود (**آزمون فرض**).

مثال: ادعا شده است متوسط عمر تراشه‌های تولیدی یک کارخانه بزرگ برابر ۱۰۰۰ ساعت است. بنابراین فرض آماری عبارتست از:

$$\theta = \mu = 1000$$

## فصل یازدهم: آزمون فرض‌ها

### يك فرض آماری باید چگونه بیان شود؟

همان طور که بیان شد، فرض یا ادعاهایی در مورد پارامتر يك جامعه بیان می‌شود، از آنجا که این ادعا ممکن است درست یا نادرست باشد، لذا دو فرض مکمل مطرح است یکی آنکه (فرض: ادعا درست است) و دیگری آنکه (فرض: ادعا نادرست است). پس شروع يك آزمون فرض با دو فرض آماری در مقابل هم می‌باشد فرض آماری که برای رد شدن تنظیم می‌شود را **فرض صفر (خنثی)** نامیده و آن را با  $H_0$  نشان می‌دهیم که همیشه شرایط موجود را در نظر می‌گیریم و فرض آماری را که در مقابل فرض صفر قرار می‌گیرد، **فرض مقابل** نامیده و با  $H_1$  نشان می‌دهیم که این فرض همواره طوری انتخاب می‌شود که می‌خواهد شرایط موجود (فرض صفر) را تغییر دهد.

**فرض آماری:** فرض ها به دو دسته تقسیم می شوند :

یا ساده هستند یعنی مقدار دقیقی برای پارامتر جامعه مشخص می‌گردد مثل  $(\theta = \theta_0)$

یا مرکب هستند که خود سه دسته‌اند.

یک طرفه از چپ  $(\theta < \theta_0)$

یک طرفه از راست  $(\theta > \theta_0)$

دو طرفه  $(\theta \neq \theta_0)$

برای پاسخ دقیق به آزمون آماری نیاز است که کل جامعه مورد بررسی قرار گیرد ولی می‌دانیم که این کار میسر نیست و پاسخ ما از روی نمونه جامعه خواهد بود پس آزمون آماری حتماً دارای **خطا** خواهد بود و سعی ما در حداقل ساختن این خطاست.

برای حداقل ساختن این خطاها باید اول این خطاها شناسایی گردند تا بتوان در مورد آنها بحث نمود. دو نوع خطا داریم

رد فرض صفر در حالیکه درست باشد، **خطای نوع اول** و قبول کردن فرض صفر در حالیکه نادرست باشد را **خطای نوع دوم** می‌نامیم.

**مثال:**

$$H_0 : \mu = \theta_0$$

$$H_1 : \mu \neq \theta_0 \quad \text{OR} \quad \mu < \theta_0 \quad \text{OR} \quad \mu > \theta_0$$

$$H_0 : \mu_1 - \mu_2 = 0 \Leftrightarrow \mu_1 = \mu_2$$

$$H_1 : \mu_1 - \mu_2 \neq 0 \quad \text{OR} \quad \mu_1 - \mu_2 < \theta_0 \quad \text{OR} \quad \mu_1 - \mu_2 > \theta_0$$

$$H_0 : P = \theta_0$$

$$H_1 : P \neq \theta_0 \quad \text{OR} \quad P < \theta_0 \quad \text{OR} \quad P > \theta_0$$

• يك آزمون خوب آزمونی است که هر دو نوع خطا در آن کم باشند، بعبارت دیگر نه فرضی را بی‌مورد رد و نه بی‌مورد قبول کنیم.

• خطای نوع اول و دوم به یکدیگر وابسته‌اند و افزایش یکی باعث کاهش دیگری می‌شود.

• با تغییر ناحیه بحرانی می‌توان احتمال خطاها را تغییر داد.

• با افزایش مقدار حجم نمونه، می‌توان هر دو خطا را به طور همزمان کاهش داد.

احتمال خطای نوع اول را با  $\alpha$  و احتمال خطای نوع دوم را با  $\beta$  نشان می‌دهیم:

$$\alpha = P(H_0 \text{ درست باشد} | \text{فرض } H_0) = P(RH_0 | H_0)$$

$$\beta = P(H_1 \text{ درست باشد} | \text{قبول فرض } H_0) = P(AH_0 | H_1)$$

$\alpha$  را میزان معنی دار بودن یا سطح تشخیص می‌نامیم.

ناحیه‌ای که باعث رد فرض صفر می‌شود **ناحیه بحرانی** یا **ناحیه رد** و ناحیه‌ای که باعث قبول فرض صفر می‌شود را **ناحیه قبولی** و مرز بین این دو ناحیه را **نقطه بحرانی** می‌نامیم. پس آزمون فرض عبارتست از **افراز فضایی نمونه به دو قسمت ناحیه بحرانی و ناحیه قبولی**.

ملاحظه می‌کنید که  $\beta$  قابل محاسبه نمی‌باشد زیرا فرض مقابل بطور مقابل مشخص نشده ولی اگر فرض مقابل را تغییر داده و به صورت  $(H_1: p=0.8)$  در نظر بگیریم،  $\beta$  قابل محاسبه است:

$$\beta = P(AH_0 | H_1) = P(X < 15.5 | p = 0.8) = \sum_{x=0}^{15} b(x; 20, 0.8) = 0.3704$$

با تغییر نقطه بحرانی خطاها تغییر می‌کنند در مثال بالا اگر نقطه بحرانی را از ۱۵،۵ به ۱۴،۵ تغییر دهیم:

$$\alpha = P(RH_0 | H_0) = P(X > 14.5 | p = 0.7) = \sum_{x=15}^{20} b(x; 20, 0.7) = 0.4164$$

$$\beta = P(AH_0 | H_1) = P(X < 14.5 | p = 0.8) = \sum_{x=0}^{14} b(x; 20, 0.8) = 0.1958$$

**مثال:** فرض کنید یک نوع کیسول سرماخوردگی (A) در کشور تولید می‌شود که پس از ۵ روز ۷۰٪ موثر است. کارخانه‌ای مدعی است کیسولی مشابه با نام B تولید می‌کند که اثر آن پس از ۵ روز بیشتر از ۷۰٪ است. فرض‌های آماری را برای بررسی ادعای کارخانه جدید تشکیل دهید:

$$H_0: p = 0.7$$

$$H_1: p > 0.7$$

یک نمونه ۲۰ تایی انتخاب می‌کنیم آماره‌ای که تصمیم‌گیری بر اساس آن صورت می‌گیرد تعداد افرادی است که اثر کیسول B در آنها موثر بوده است اگر X تعداد افرادی فرض کنیم که کیسول B در آنها موثر بوده است اگر  $X \geq 16$  را ناحیه بحرانی (ناحیه رد) و  $X < 16$  ناحیه قبول در نظر می‌گیریم. خطاهای آزمون فرض را محاسبه می‌کنیم:

$$\alpha = P(RH_0 | H_0) = P(X > 15.5 | p = 0.7) = \sum_{x=16}^{20} b(x; 20, 0.7) = 0.2375$$

$$\beta = P(AH_0 | H_1) = P(X < 15.5 | p > 0.7) = \text{????}$$

پس اگر آزمون فرض مربوط به توزیع گسسته باشد برای بررسی آن باید نقطه بحرانی را به طور دلخواه انتخاب و به کمک آن خطاهای نوع اول و دوم را به دست آوریم و اگر این خطاها بزرگ نباشند ناحیه بحرانی را مشخص و چنانچه نتیجه آزمایش در ناحیه بحرانی قرار گرفت فرض صفر را رد و در غیر اینصورت فرض صفر پذیرفته می‌شود. اما اگر احتمال خطاهای نوع اول و دوم زیاد شد با تغییر نقطه بحرانی (با ثابت نگه داشتن حجم نمونه) و در صورت نرسیدن به مقدار مطلوب خطاها، در صورت امکان حجم نمونه را افزایش می‌دهیم.

گفتیم تصمیم‌گیری خوب وقتی انجام می‌پذیرد که هر دو نوع خطا کم باشد برای کاهش همزمان هر دو نوع خطا باید حجم نمونه را افزایش دهیم در مثال بالا اگر حجم نمونه را ۱۰۰ انتخاب کنیم و نقطه بحرانی را به دلخواه ۷۳/۵ اختیار کنیم خطاها به صورت زیر در می‌آیند:

$$\alpha = P(RH_0 | H_0) = P(X > 73.5 | p = 0.7) = \text{???}$$

$$np > 5 \ \& \ n \rightarrow \infty \Rightarrow b(x; n, p) \square n(np, \sqrt{npq}) \Rightarrow$$

$$\mu = np = 100 \times 0.7 = 70 \ \& \ \sigma^2 = npq = 21 \Rightarrow b(x; 100, 0.7) \square n(70, \sqrt{21})$$

$$\alpha \square P(Z > \frac{73.5 - 70}{4.58}) = 1 - P(Z < 0.76) = 0.2236$$

$$\beta = P(AH_0 | H_1) = P(X < 73.5 | p = 0.8) = \text{???}$$

$$np > 5 \ \& \ n \rightarrow \infty \Rightarrow b(x; n, p) \square n(np, \sqrt{npq}) \Rightarrow$$

$$\mu = np = 100 \times 0.8 = 80 \ \& \ \sigma^2 = npq = 16 \Rightarrow b(x; 100, 0.7) \square n(80, 4)$$

$$\beta \square P(Z < \frac{73.5 - 80}{4}) = P(Z < -1.625) = 0.0521$$

تابع توان یک آزمون فرض آماری  $H_0$  در برابر فرض مقابل  $H_1$  به صورت زیر است:

$$\pi(\theta) = \begin{cases} \alpha(\theta) & \text{برای مقادیر } \theta \text{ که تحت } H_0 \text{ اختیار می‌شوند} \\ 1 - \beta(\theta) & \text{برای مقادیر } \theta \text{ که تحت } H_1 \text{ اختیار می‌شوند} \end{cases}$$

بنابراین تابع توان برای مقادیر  $\theta$  تحت  $H_0$ ، احتمال ارتکاب خطای نوع اول و برای مقادیر  $\theta$  تحت  $H_1$ ، احتمال مرتکب نشدن خطای نوع دوم می‌باشد.

تابع‌های توان نقش بسیار مهمی در ارزیابی آزمون‌های آماری، بویژه در مقایسه چندین ناحیه بحرانی برای یک آزمون را دارند.

در صنعت تابع  $(1 - \pi(\theta))$  به عنوان **منحنی مشخصه عمل** کاربرد دارد که این تابع طبق تعریف احتمال‌های قبول  $H_0$  به جای رد آن است.

ولی اگر آماره‌ای که تصمیم‌گیری بر اساس آن صورت می‌پذیرد دارای توزیع پیوسته باشد می‌توان ابتدا خطای نوع اول را اختیار کرد و به کمک آن ناحیه بحرانی را تعیین نمود و نیازی نیست که مانند گسسته‌ها ابتدا ناحیه بحرانی را انتخاب و  $\alpha$  محاسبه کنیم و در صورت قابل قبول نبودن، ناحیه بحرانی را تغییر دهیم.

در مثال قبل برای تعیین احتمال خطاها یک فرض ساده را در برابر یک فرض ساده دیگر می‌آزمودیم ولی در عمل چنین نیست و در بیشتر مواقع با فرض‌های مرکب سرو کار داریم و ناچار به تعیین احتمال خطاها در این حالت هستیم برای اینکار از **تابع توان آزمون فرض** استفاده می‌کنیم.

### آزمون‌های یک طرفه و دو طرفه:

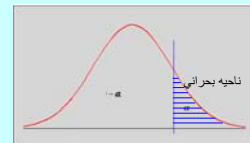
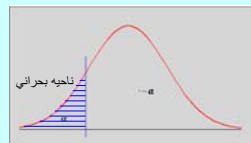
آزمون فرض را **یکطرفه** می‌نامیم اگر فرض مقابل آن یکطرفه باشد:

$$H_0: \theta = \theta_0$$

$$H_1: \theta < \theta_0$$

$$H_0: \theta = \theta_0$$

$$H_1: \theta > \theta_0$$



### آزمون‌های یک طرفه و دو طرفه:

آزمون فرض را **دو طرفه** گوئیم اگر فرض مقابل دو طرفه باشد.

$$H_0: \theta = \theta_0$$

$$H_1: \theta \neq \theta_0$$



معمولاً در آزمون فرض‌ها  $\alpha$  را  $0.05$  یا  $0.01$  اختیار می‌کنیم، اگر  $\alpha = 0.05$  آزمون را **معنی دار** و اگر  $\alpha = 0.01$  آزمون را **بسیار معنی دار** می‌نامیم. اگر مقدار  $\alpha$  مشخص نباشد آن را  $0.05$  یعنی معنی‌دار در نظر می‌گیریم.

### مراحل آزمون فرض‌ها:

- ۱- تشکیل فرض صفر  $H_0: \theta = \theta_0$
- ۲- تشکیل فرض مقابل  $H_1: \theta \neq \theta_0$  OR  $\theta < \theta_0$  OR  $H_1: \theta > \theta_0$
- ۳- انتخاب میزان معنی‌دار بودن
- ۴- انتخاب آماره مناسب برای پارمتر  $\theta$  و تشخیص نوع توزیع آن طبق قضایا و مطالب فصل قبل، و سپس تشکیل ناحیه بحرانی
- ۵- محاسبه مقدار آماره انتخابی به کمک نمونه تصادفی انتخاب شده
- ۶- نتیجه‌گیری: اگر مقدار آماره در ناحیه بحرانی قرار گیرد فرض  $H_0$  رد، در غیر اینصورت فرض  $H_0$  پذیرفته می‌شود.

$H_0$	آماره آزمون	$H_1$	ناحیه بحرانی
$\mu = \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ ( $n < 30$ معلوم، جامعه نرمال و $n \geq 30$ هر جامعه‌ای و $n \geq 30$ نامعلوم (استفاده از واریانس نمونه)، هر جامعه‌ای و $n \geq 30$ )	$\mu < \mu_0$	$Z < -Z_{1-\alpha}$
		$\mu > \mu_0$	$Z > Z_{1-\alpha}$
		$\mu \neq \mu_0$	$Z < -Z_{1-\alpha/2}$ $Z > Z_{1-\alpha/2}$
$\mu = \mu_0$	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}; v = n - 1$ ( $n < 30$ نامعلوم (استفاده از S به جای $\sigma$ )) جامعه نرمال و $n < 30$ )	$\mu < \mu_0$	$T < -t_{1-\alpha}$
		$\mu > \mu_0$	$T > t_{1-\alpha}$
		$\mu \neq \mu_0$	$T < -t_{1-\alpha/2}$ $T > t_{1-\alpha/2}$

$H_0$	آماره آزمون	$H_1$	ناحیه بحرانی
$\mu_1 - \mu_2 = d_0$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$ ( $\sigma_1, \sigma_2$ معلوم، جوامع نرمال و $n_1, n_2 < 30$ ) ( $\sigma_1, \sigma_2$ معلوم، هر جامعه‌ای و $n_1, n_2 \geq 30$ ) ( $\sigma_1, \sigma_2$ نامعلوم (استفاده از واریانس نمونه)، هر جامعه‌ای و $n_1, n_2 \geq 30$ )	$\mu_1 - \mu_2 < d_0$	$Z < -Z_{1-\alpha}$
		$\mu_1 - \mu_2 > d_0$	$Z > Z_{1-\alpha}$
		$\mu_1 - \mu_2 \neq d_0$	$Z < -Z_{1-\alpha/2}$ $Z > Z_{1-\alpha/2}$

$H_0$	آماره آزمون	$H_1$	ناحیه بحرانی
$\mu_1 - \mu_2 = d_0$	$T = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{S_p \sqrt{(1/n_1) + (1/n_2)}}$ $v = n_1 + n_2 - 2$ $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ ( $\sigma_1, \sigma_2$ نامعلوم ولی مساوی، جوامع نرمال و $n_1, n_2 < 30$ ) ( $\sigma_1, \sigma_2$ نامعلوم و نامساوی ولی $n_1 = n_2$ ) ( $\sigma_1, \sigma_2$ نامعلوم (استفاده از واریانس نمونه)، هر جامعه‌ای و $n_1, n_2 \geq 30$ )	$\mu_1 - \mu_2 < d_0$	$T < -t_{1-\alpha}$
		$\mu_1 - \mu_2 > d_0$	$T > t_{1-\alpha}$
		$\mu_1 - \mu_2 \neq d_0$	$T < -t_{1-\alpha/2}$ $T > t_{1-\alpha/2}$

H <sub>0</sub>	آماره آزمون	H <sub>1</sub>	ناحیه بحرانی
p = p <sub>0</sub>	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{X - np_0}{\sqrt{np_0 q_0}}$ $\hat{p} = \frac{X}{n}, q_0 = 1 - p_0$	p < p <sub>0</sub>	Z < -Z <sub>1-α</sub>
		p > p <sub>0</sub>	Z > Z <sub>1-α</sub>
		p ≠ p <sub>0</sub>	Z < -Z <sub>1-α/2}</sub>
p <sub>1</sub> = p <sub>2</sub> = P	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ $\hat{p}_1 = \frac{x_1}{n_1}$ $\hat{p}_2 = \frac{x_2}{n_2}, \text{ if } (p?) \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$	p <sub>1</sub> < p <sub>2</sub>	Z < -Z <sub>1-α</sub>
		p <sub>1</sub> > p <sub>2</sub>	Z > Z <sub>1-α</sub>
		p <sub>1</sub> ≠ p <sub>2</sub>	Z < -Z <sub>1-α/2}</sub>
	(n <sub>1</sub> , n <sub>2</sub> ≥ 30) n <sub>1</sub> , n <sub>2</sub> → ∞		

H <sub>0</sub>	آماره آزمون	H <sub>1</sub>	ناحیه بحرانی
σ <sup>2</sup> = σ <sub>0</sub> <sup>2</sup>	$X^2 = \frac{(n-1)S^2}{\sigma_0^2}$ $v = n - 1$	σ <sup>2</sup> < σ <sub>0</sub> <sup>2</sup>	X <sup>2</sup> < χ <sup>2</sup> <sub>α</sub>
		σ <sup>2</sup> > σ <sub>0</sub> <sup>2</sup>	X <sup>2</sup> > χ <sup>2</sup> <sub>1-α</sub>
		σ <sup>2</sup> ≠ σ <sub>0</sub> <sup>2</sup>	X <sup>2</sup> < χ <sup>2</sup> <sub>α/2}</sub>
σ <sub>1</sub> <sup>2</sup> = σ <sub>2</sub> <sup>2</sup>	$F = \frac{S_1^2}{S_2^2}$ $v_1 = n_1 - 1$ $v_2 = n_2 - 1$	σ <sub>1</sub> <sup>2</sup> < σ <sub>2</sub> <sup>2</sup>	F < f <sub>α</sub> (v <sub>1</sub> , v <sub>2</sub> )
		σ <sub>1</sub> <sup>2</sup> > σ <sub>2</sub> <sup>2</sup>	F < f <sub>1-α</sub> (v <sub>1</sub> , v <sub>2</sub> )
		σ <sub>1</sub> <sup>2</sup> ≠ σ <sub>2</sub> <sup>2</sup>	F < f <sub>α/2</sub> (v <sub>1</sub> , v <sub>2</sub> ) F < f <sub>1-α/2</sub> (v <sub>1</sub> , v <sub>2</sub> )

**آزمون‌های مربوط به تفاضل‌های بین k نسبت:** برای تفاضل دو نسبت روابط آزمون مورد بررسی قرار گرفت حال اگر آزمون فرض، مربوط به تفاضل‌های k نسبت باشد تکلیف چیست؟

برای این منظور از روش کلی زیر استفاده می‌کنیم:

- تشکیل فرض صفر H<sub>0</sub>: p<sub>1</sub> = p<sub>2</sub> = ... = p<sub>k</sub> = p
- تشکیل فرض مقابل حداقل یکی از نسبت‌ها (ها) برابر با p نیست: H<sub>1</sub>
- انتخاب α
- آماره مناسب برای این آزمون به صورت زیر محاسبه می‌گردد:

در آزمون نسبت فرض شد که (n → ∞) و آزمون مورد بررسی قرار گرفت حال اگر n کوچک باشد به صورت زیر (طبق روش تابع توزیع گسسته) عمل می‌کنیم

- تشکیل فرض صفر H<sub>0</sub>: p = p<sub>0</sub>
- تشکیل فرض مقابل H<sub>1</sub>: p < p<sub>0</sub> OR p > p<sub>0</sub> OR p ≠ p<sub>0</sub>
- انتخاب α و تشکیل ناحیه بحرانی به صورت زیر

if (H<sub>1</sub>: p < p<sub>0</sub>) ⇒ P(X ≤ x | p = p<sub>0</sub>) < α  
 if (H<sub>1</sub>: p > p<sub>0</sub>) ⇒ P(X ≥ x | p = p<sub>0</sub>) < α  
 if (H<sub>1</sub>: p ≠ p<sub>0</sub>) ⇒ { if (x < np<sub>0</sub>) ⇒ P(X ≤ x | p = p<sub>0</sub>) < α/2  
 if (x > np<sub>0</sub>) ⇒ P(X ≥ x | p = p<sub>0</sub>) < α/2

۴- محاسبات: به کمک نمونه منتخب با حجم n، مقدار x را تعیین و ناحیه بحرانی را تعیین می‌کنیم.

۵- نتیجه‌گیری: اگر x در ناحیه بحرانی قرار گرفت، فرض صفر را رد می‌کنیم.

**آزمون‌های مربوط به تفاضل‌های بین k نسبت:**

آماره X<sup>2</sup> از رابطه زیر به دست می‌آید:

$$X^2 = \sum_{i=1}^k \sum_{j=1}^2 \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

۵- ناحیه بحرانی به صورت X<sup>2</sup> ≥ χ<sup>2</sup><sub>(1-α)</sub> با v=k-1 درجه آزادی مشخص می‌شود.

۶- نتیجه‌گیری: اگر مقدار آماره در ناحیه بحرانی قرار گیرد فرض H<sub>0</sub> رد، در غیر اینصورت فرض H<sub>0</sub> پذیرفته می‌شود.

**آزمون‌های مربوط به تفاضل‌های بین k نسبت:**

داده‌ها را به صورت جدول زیر مرتب می‌کنیم:

	شکست‌ها	بیروزی‌ها
نمونه اول	n <sub>1</sub> -x <sub>1</sub>	x <sub>1</sub>
نمونه دوم	n <sub>2</sub> -x <sub>2</sub>	x <sub>2</sub>
⋮	⋮	⋮
نمونه k	n <sub>k</sub> -x <sub>k</sub>	x <sub>k</sub>

که درایه‌های آن را فراوانی‌های خانه‌ای مشاهده شده f<sub>ij</sub> می‌نامیم که در آن اولین اندیس نشانه سطر و دومین اندیس نشانه ستون این جدول 2\*k است.

امید فراوانی‌های خانه‌ای مشاهده شده را با e<sub>ij</sub> نشان داده و به صورت زیر محاسبه می‌کنیم.

$$e_{ij} = \begin{cases} n_i p & j = 1 \\ n_i (1-p) & j = 2 \end{cases} \text{ if } (p?) \Rightarrow \hat{p} = \frac{x_1 + x_2 + \dots + x_k}{n_1 + n_2 + \dots + n_k}$$

### آزمون جدول‌های توافقی:

یک جدول توافقی  $r \times c$  دارای  $c$  ستون معرف رسته‌های مختلف  $A_1, A_2, \dots, A_c$ ،  $r$  سطر معرف رسته‌های مختلف  $B_1, B_2, \dots, B_r$  از متغیر دیگر است و  $f_{ij}$  فراوانی خانه‌های مشاهده شده برای خانه‌ای است که به سطر  $i$  ام و ستون  $j$  ام تعلق دارد.

	$A_1$	$A_2$	...	$A_c$	
$B_1$	$f_{11}$	$f_{12}$	...	$f_{1c}$	$f_{10}$
$B_2$	$f_{21}$	$f_{22}$	...	$f_{2c}$	$f_{20}$
...	...	...	...	...	...
$B_r$	$f_{r1}$	$f_{r2}$	...	$f_{rc}$	$f_{r0}$
	$f_{01}$	$f_{02}$	...	$f_{0c}$	$f$

$$f_{i0} = \sum_{j=1}^c f_{ij}$$

$$f_{0j} = \sum_{i=1}^r f_{ij}$$

$$f = \sum_{i=1}^r \sum_{j=1}^c f_{ij}$$

### آزمون جدول‌های توافقی:

فرض صفری که می‌خواهیم آزمون کنیم مستقل بودن دو متغیر است یعنی اگر  $\theta_{ij}$  احتمال آن باشد که فقره‌ای در خانه‌ای قرار خواهد گرفت که به سطر  $i$  ام و ستون  $j$  ام تعلق دارد و  $\theta_{i0}$  احتمال آن باشد که فقره‌ای به سطر  $i$  ام و  $\theta_{0j}$  احتمال آن باشد که فقره‌ای به ستون  $j$  ام تعلق دارد. پس فرض صفر به صورت زیر در می‌آید:

$$\theta_{ij} = (\theta_{i0})(\theta_{0j}) \quad i = 1, 2, \dots, r \quad \& \quad j = 1, 2, \dots, c$$

و فرض مقابل هم به این شکل است که بازای حداقل یک زوج  $i$  و  $j$  مقدار احتمال  $\theta_{ij}$  برابر حاصلضرب  $\theta_{i0}$  و  $\theta_{0j}$  نیست.

### آزمون جدول‌های توافقی:

فراوانی‌های خانه‌های مورد انتظار به صورت زیر محاسبه می‌شود:

$$\hat{\theta}_{i0} = \frac{f_{i0}}{f}, \quad \hat{\theta}_{0j} = \frac{f_{0j}}{f}$$

$$e_{ij} = (\hat{\theta}_{i0})(\hat{\theta}_{0j})f = \frac{f_{i0}}{f} \cdot \frac{f_{0j}}{f} \cdot f = \frac{(f_{i0})(f_{0j})}{f}$$

آماره این آزمون نیز به شکل زیر محاسبه می‌گردد:

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

ناحیه بحرانی برای این آزمون به صورت  $X^2 \geq \chi_{1-\alpha}^2$  با  $v = (r-1)(c-1)$  درجه آزادی مشخص می‌شود

### آزمون جدول‌های توافقی:

ناحیه بحرانی برای این آزمون به صورت  $X^2 \geq \chi_{1-\alpha}^2$  با  $v = (r-1)(c-1)$  درجه آزادی مشخص می‌شود که اگر آماره آزمون در این محدوده قرار گرفت فرض صفر رد می‌شود.

این آزمون را وقتی به کار می‌بریم که هیچ یک از  $e_{ij}$  ها کمتر از 5 نباشند این امر مستلزم این است که برخی خانه‌ها را با هم ادغام کنیم که در نتیجه از درجه آزادی متناظر کاسته می‌شود.



### آزمون نیکویی برازش:

این آزمون مواقعی به کار می‌رود که می‌خواهیم تعیین کنیم آیا می‌توان مجموعه‌ای از داده‌ها را به عنوان نمونه‌هایی تصادفی از جامعه‌ای با توزیع مفروض تلقی کرد یا نه؟

برای این منظور داده‌ها را طبق اصول آمار توصیفی طبقه‌بندی می‌کنیم که معمولاً این طبقه‌بندی در داده‌های گسسته صورت نمی‌گیرد ولی اگر دامنه تغییرات داده‌ها زیاد باشد در این حالت نیز می‌توان داده‌ها را به چند طبقه بر اساس اصول آمار توصیفی طبقه‌بندی کرد سپس جدول فراوانی داده‌ها تکمیل می‌گردد.

### آزمون نیکویی برازش:

با توجه به فضایی گفته شده پارامترهای مورد نیاز توزیع مفروض از روی نمونه، تخمین می‌گردد تا احتمالهای مورد نظر محاسبه شوند.

فراوانی‌های مورد انتظار $e_{ij} = p_i \cdot f_j$	احتمال هر رده با توجه به توزیع مفروض $p_i$	فراوانی مشاهده شده $f_i$	داده‌ها یا نماینده هر دسته در حالت دسته‌بندی	حدود طبقات در حالت دسته‌بندی
...	...	...	...	...
		$\sum f_i = f$		

### آزمون نیکویی برآزش:

فرض صفر در این آزمون به این صورت است که مجموعه‌ای از داده‌های مشاهده شده، دارای توزیع مفروض می‌باشند و فرض مقابل هم این است که داده‌ها دارای توزیع دیگری است. آماره این آزمون به شکل زیر محاسبه می‌گردد:

$$X^2 = \sum_{j=1}^m \frac{(f_j - e_j)^2}{e_j}$$

ناحیه بحرانی برای این آزمون به صورت  $X^2 \geq \chi_{1-\alpha}^2$  با  $v=s+t-1$  درجه آزادی می‌باشد که با توجه به آماره محاسبه شده نتیجه لازم مشخص می‌شود.

### آزمون نیکویی برآزش:

در این آزمون درجه آزادی  $v=s-t-1$  مشخص شد. که  $s$  تعداد جملات و  $t$  تعداد پارامترهای مستقلی است که به جای آنها برآوردهایشان گذاشته می‌شود.

رابطه  $v=s-t-1$  یک رابطه کلی است هرگاه فراوانی‌های خانهای مورد انتظار در فرمول‌های  $X^2$  بر مبنای داده‌های شمارشی نمونه‌ای برآورد شوند. برای مثال در جدول توافقی  $s=rc$  و  $t=(r+c-2)$  در نتیجه:

$$s - t - 1 = rc - (r + c - 2) - 1 = (r - 1)(c - 1)$$

و در آزمون تقاض  $k$  نسبت،  $s=2k$  و  $t=k$  در نتیجه:

$$s - t - 1 = 2k - (k) - 1 = k - 1$$

**مثال:** اگر میانگین یک نمونه تصادفی با حجم  $n=64$  از یک جامعه،  $78.8$  و انحراف معیار جامعه  $12.8$  باشد آیا در میزان معنی دار بودن  $0.02$  می‌توان گفت که میانگین جامعه از  $75$  بزرگتر است؟

1)  $H_0: \mu = 75$

2)  $H_1: \mu > 75$

3)  $\alpha = 0.02$

4)  $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$  & Rejection region:  $Z > z_{1-\alpha} = z_{0.98} = 2.054$

5)  $Z = \frac{78.8 - 75}{12.8 / \sqrt{64}} = 2.375$

۶- چون  $Z$  در ناحیه بحرانی قرار می‌گیرد پس فرض  $H_0$  رد می‌شود.

**مثال:** فرض کنید  $X$  دارای توزیع نرمال با میانگین  $\mu$  و واریانس  $36$  است در آزمونی به صورت

$$H_0: \mu = 50, \quad H_1: \mu = 55$$

می‌باشد ناحیه بحرانی و  $n$  را طوری تعیین کنید که به ازای آن‌ها خطای نوع اول  $0.05$  و خطای دوم  $0.1$  باشد؟

حل: فرض کنید  $c$  نقطه بحرانی باشد:

$$\alpha = P(\bar{X} > c | \mu = 50) \Rightarrow 0.05 = P(Z > \frac{c - 50}{6 / \sqrt{n}}) \Rightarrow$$

$$P(Z < \frac{c - 50}{6 / \sqrt{n}}) = 0.95 \Rightarrow \boxed{1.645 = \frac{c - 50}{6 / \sqrt{n}}} \quad (1)$$

$$\beta = P(\bar{X} < c | \mu = 55) \Rightarrow 0.1 = P(Z < \frac{c - 55}{6 / \sqrt{n}}) \Rightarrow \boxed{-1.288 = \frac{c - 55}{6 / \sqrt{n}}} \quad (2)$$

$$(1) \& (2) \Rightarrow c = 52.8, \quad n = 13$$

**مثال:** یک نمونه  $5$  تایی از جامعه‌ای نرمال انتخاب کرده‌ایم، که میانگین نمونه و انحراف معیار نمونه آن به ترتیب  $3$  و  $0.9$  است. آیا می‌توان گفت میانگین جامعه از  $3.5$  کمتر است؟

1)  $H_0: \mu = 3.5$

2)  $H_1: \mu < 3.5$

3)  $\alpha = 0.5$

4)  $T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$  & Rejection region:  $T < -t_{1-\alpha, n-1} = -t_{0.95, 4} = -2.132$

5)  $T = \frac{3 - 3.5}{0.9\sqrt{5}} = -1.238$

۶- چون  $T$  در ناحیه بحرانی قرار نمی‌گیرد پس  $H_0$  رد نمی‌شود.

**مثال:** از دو کلاس  $40$  و  $50$  نفری امتحان مشابهی را گرفتیم، نتایج میانگین و انحراف معیار نمرات این دو گروه به صورت زیر است:

$$\mu_1 = 74, \quad s_1 = 8, \quad \mu_2 = 78, \quad s_2 = 7$$

آیا می‌توان گفت این دو کلاس با هم اختلاف دارند؟

1)  $H_0: \mu_1 - \mu_2 = 0$

2)  $H_1: \mu_1 - \mu_2 \neq 0$

3)  $\alpha = 0.5$

4)  $Z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$  & Rejection region:  $\begin{cases} Z < -z_{1-\alpha/2} = -z_{0.975} = -1.96 \\ Z > z_{1-\alpha/2} = z_{0.975} = 1.96 \end{cases}$

5)  $Z = \frac{74 - 78}{\sqrt{\frac{64}{40} + \frac{49}{50}}} = -2.49$

۶- چون  $Z$  در ناحیه بحرانی قرار می‌گیرد پس  $H_0$  رد می‌شود.

**مثال:** اگر انحراف معیار نمونه تصادفی با حجم  $n=25$  از یک جامعه نرمال  $s=0.28$  باشد آیا می‌توان ادعا کرد که انحراف معیار این جامعه کمتر از  $0.3$  است؟

- 1)  $H_0: \sigma = 0.3$
- 2)  $H_1: \sigma < 0.3$
- 3)  $\alpha = 0.5$
- 4)  $X^2 = \frac{(n-1)S^2}{\sigma^2}$  & Rejection region:  $X^2 < \chi_{\alpha, n-1}^2 = 13.8$
- 5)  $X^2 = \frac{(25-1)(0.28)^2}{(0.3)^2} = 20.907$

۶- چون  $X^2$  در ناحیه بحرانی نیست پس فرض صفر رد نمی‌شود.

**مثال:** در مثال قبل اگر  $n_1=16$  و  $n_2=25$  با فرض برابری واریانس‌های دو جامعه آزمون را دوباره انجام دهید:

- 1)  $H_0: \mu_1 - \mu_2 = 0$
- 2)  $H_1: \mu_1 - \mu_2 \neq 0$
- 3)  $\alpha = 0.5$
- 4)  $T = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{S_p \sqrt{(1/n_1) + (1/n_2)}}$  & Rejection region:  $\begin{cases} T < -t_{1-\alpha/2} = -t_{0.975} = -2.07 \\ T > t_{1-\alpha/2} = t_{0.975} = 2.07 \end{cases}$
- 5)  $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} = \frac{(16-1) \cdot 64 + (8-1) \cdot 49}{23} = 56.65$
- 5)  $T = \frac{74-78}{7.53 \times \sqrt{\frac{1}{16} + \frac{1}{8}}} = -1.23$

۶- چون  $T$  در ناحیه بحرانی قرار نمی‌گیرد پس  $H_0$  رد نمی‌شود.

**مثال:** تیراندازی مدعی است که  $60\%$  تیرهایی را که شلیک می‌کند به هدف می‌خورد اگر  $55$  تیر از  $100$  تیری که شلیک کرده به هدف خورده باشد، در مورد این ادعا چه اظهار نظری می‌توان کرد؟

- 1)  $H_0: p = 0.6$
- 2)  $H_1: p \neq 0.6$
- 3)  $\alpha = 0.05$
- 4)  $Z = \frac{x - np_0}{\sqrt{np_0q_0}}$  & Rejection region:  $\begin{cases} Z > z_{1-\alpha/2} = z_{0.975} = 1.96 \\ Z < -z_{1-\alpha/2} = -z_{0.975} = -1.96 \end{cases}$
- 5)  $Z = \frac{55 - 100 \times 0.6}{\sqrt{100 \times 0.6 \times 0.4}} = -1.02$

۶- چون  $Z$  در ناحیه بحرانی قرار نمی‌گیرد پس فرض صفر رد نمی‌شود.

**مثال:** فرض کنید معدل دانشجویان دانشگاه A و دانشگاه B دارای توزیع نرمال با واریانس‌های  $\sigma_A^2$  و  $\sigma_B^2$  باشند دو نمونه تصادفی به حجمهای  $n_1=10$  و  $n_2=7$  از دو دانشگاه انتخاب می‌کنیم واریانس‌های نمونه آن به ترتیب برابر با  $0.745$  و  $0.2028$  می‌باشد آیا می‌توان ادعا کرد که دو دانشگاه در میزان معنی دار بودن  $0.02$  دارای واریانس برابر هستند؟

- 1)  $H_0: \sigma_A = \sigma_B$
- 2)  $H_1: \sigma_A \neq \sigma_B$
- 3)  $\alpha = 0.02$
- 4)  $F^2 = \frac{S_A^2}{S_B^2}$  & Rejection region:  $\begin{cases} F < f_{\alpha/2}(n_1-1, n_2-1) = 0.172 \\ F < f_{1-\alpha/2}(n_1-1, n_2-1) = 7.98 \end{cases}$
- 5)  $F = \frac{0.0765}{0.2028} = 0.3772$

۶- چون  $f$  در ناحیه بحرانی قرار نمی‌گیرد، لذا فرض صفر را رد نمی‌کنیم.

**مثال:** بازیکنی مدعی است که  $60\%$  از توپ‌هایی را که پرتاب می‌کند وارد سبد می‌شوند، از  $15$  توپی که پرتاب کرده  $10$  توپ وارد سبد شده است، آیا در سطح تشخیص  $5\%$  می‌توان این ادعا را پذیرفت؟

- 1)  $H_0: p = 0.6$
- 2)  $H_1: p \neq 0.6$
- 3)  $\alpha = 0.05$
- Rejection region:  $\begin{cases} x=10, np_0 = 15 \times 0.6 = 9 \\ x > np_0 [10 > 9] \Rightarrow P(X > x | p = 0.6) < (\alpha/2) = 0.025 \end{cases}$
- 4)  $P(X > x | p = 0.6) = \sum_{x=10}^{15} b(x; 15, 0.6) = 1 - \sum_{x=0}^9 b(x; 15, 0.6) = 1 - 0.5968 = 0.4032$

۵- چون احتمال محاسبه شده کمتر از  $0.025$  نیست پس فرض صفر را رد نمی‌کنیم عبارت دیگر دلیلی برای شک کردن به ادعای بازیکن نیست.

**مثال:**  $56$  نفر از  $200$  نفر که مصرف کننده نوشابه هستند نوع A را ترجیح می‌دهند و  $29$  نفر از  $150$  نفر دیگر که مصرف کننده نوشابه هستند نوع B را ترجیح می‌دهند، آیا در سطح تشخیص  $0.06$  می‌توان نتیجه گرفت که نوع A از نوع B بهتر است؟

- 1)  $H_0: p_A = p_B$
- 2)  $H_1: p_A > p_B$
- 3)  $\alpha = 0.06$
- 4)  $Z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_2} + \frac{1}{n_1})}}$  & Rejection region:  $Z > z_{1-\alpha/2} = z_{0.94} = 1.555$
- 5)  $\hat{p}_1 = \frac{x_A}{n_A} = \frac{56}{200} = 0.28, \hat{p}_2 = \frac{x_B}{n_B} = \frac{29}{150} = 0.19, \hat{p} = \frac{x_A + x_B}{n_A + n_B} = \frac{56 + 29}{200 + 150} = 0.24$
- 6)  $Z = \frac{0.28 - 0.19}{\sqrt{0.27 \times 0.76 \times (\frac{1}{200} + \frac{1}{150})}} = 1.95$

۶- پس فرض صفر رد می‌شود یعنی جواب مثبت است.

ناحیه بحرانی:  $X^2 \geq \chi_{0.95,2}^2 = 5.991$

$$X^2 = \sum_{i=1}^3 \sum_{j=1}^2 \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = \frac{(232-212)^2}{212} + \dots + \frac{(203-188)^2}{188} = 6.48$$

چون  $X^2$  در ناحیه بحرانی قرار دارد پس فرض صفر باید رد شود، به عبارت دیگر، نسبت‌های واقعی مشتریانی که ماده پاک کننده A را به ماده پاک کننده B ترجیح می‌دهند، در هر شهر یکسان نیستند.

**مثال:** بر مبنای داده‌های نمونه‌ای که در جدول زیر نشان داده، تعیین کنید آیا نسبت واقعی مشتریانی که ماده پاک کننده A را به ماده پاک کننده B ترجیح می‌دهند، در هر سه شهر یکسان است یا نه؟

	ماده A	ماده B	
شهر الف	232	168	$n_1=400$
شهر ب	260	240	$n_1=500$
شهر ج	197	203	$n_1=400$

$$\begin{cases} H_0: \theta_1 = \theta_2 = \theta_3 \\ H_1: \theta \text{ همه با هم برابر نیستند} \end{cases}$$

$$e_{ij} = \begin{cases} n_i p & j=1 \\ n_i (1-p) & j=2 \end{cases} \quad \hat{p} = \frac{x_1 + x_2 + x_3}{n_1 + n_2 + n_3}$$

$$\hat{p} = \frac{232 + 260 + 197}{400 + 500 + 400} = 0.53$$

$e_{ij}$	1	2
۱	212	188
۲	265	235
۳	212	188

گنبد

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

گنبد

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

گنبد

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

گنبد

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

گنبد

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

گنبد

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

گنبد

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

گنبد

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

گنبد

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

Dr. AlirezaSG@qibqa.com

http://www.scpb.ir/

## فصل دوازدهم: رگرسیون و همبستگی

فرض کنید  $X$  و  $Y$  دو متغیر تصادفی از دو جامعه آماری مختلف باشند هر مشاهده از این دو متغیر را به صورت زوج مرتب  $(x, y)$  نشان می‌دهیم در این فصل دو مسأله زیر مورد بررسی قرار می‌گیرد:

۱- آیا بین دو متغیر  $X$  و  $Y$  همبستگی (رابطه‌ای) وجود دارد؟

۲- آیا این همبستگی یا رابطه را می‌توان به صورت معادله بیان نمود؟

برای مثال می‌خواهیم بررسی کنیم که آیا بین سن و وزن افراد رابطه‌ای وجود دارد و اگر رابطه‌ای وجود دارد معادله‌ای را که این رابطه را بیان می‌کند پیدا کنیم. معمولاً متغیری که از روی مقادیر آن پیشگویی انجام می‌شود را با  $X$  و متغیری که پیشگویی می‌شود را با  $Y$  نشان می‌دهیم.  $X$  را متغیر مستقل و  $Y$  را متغیر وابسته می‌نامیم.

برای مثال اگر جدول زیر دو نمونه از دو متغیر تصادفی  $x$  و  $y$  باشند:

سن (X)	۵۶	۴۸	۴۲	۳۵	۵۸	۴۰	۳۹	۵۰
وزن (Y)	۴۴/۰	۹۵/۲	۷۵/۰	۶۹/۹	۵۹/۰	۹۴/۹	۶۲/۰	۶۵/۵
	۶۵/۲							

هدف بررسی این موضوع است که آیا بین سن و وزن افراد رابطه‌ای وجود دارد؟

اگر  $x$  را متغیر مستقل و  $y$  را متغیر وابسته در نظر بگیریم هدف تعیین

• متوسط مقدار  $y$  به ازای مقدار مفروضی از  $x$  یا میانگین شرطی  $\mu_{y|x}$  می‌باشد.

که اگر  $f(x,y)$  تابع احتمال توأم دو متغیر تصادفی  $x$  و  $y$  باشند

$$\mu_{y|x} = E(y|x) = \int_y y \cdot k(y|x) dy$$

$$\mu_{y|x} = E(y|x) = \sum_y y \cdot k(y|x)$$

معادله رگرسیون  $y$  به روی  $x$  نامیده میشود.

**مثال:** با مفروض بودن متغیرهای تصادفی  $x$  و  $y$  با چگالی توأم زیر معادله رگرسیون  $y$  روی  $x$  را بیابید؟

$$f(x,y) = \begin{cases} xe^{-x(1+y)} & ; x, y > 0 \\ 0 & ; \text{other point} \end{cases}$$

$$g(x) = \int_y xe^{-x(1+y)} dy = \begin{cases} e^{-x} & ; x > 0 \\ 0 & ; \text{other point} \end{cases}$$

$$k(y|x) = \frac{f(x,y)}{g(x)} = \begin{cases} xe^{-xy} & ; y > 0 \\ 0 & ; \text{other point} \end{cases}$$

$$\mu_{y|x} = E(y|x) = \int_y y \cdot k(y|x) dy = \int_y y \cdot xe^{-xy} dy = \begin{cases} \frac{1}{x} & ; x > 0 \\ 0 & ; \text{other point} \end{cases}$$

**رگرسیون خطی:** اگر منحنی رگرسیون  $y$  نسبت به  $x$  خطی باشد معادله آن را به صورت

$$\mu_{y|x} = \alpha + \beta x$$

در نظر می‌گیریم که در آن  $\alpha$  و  $\beta$  ضرایب رگرسیون نامیده می‌شوند. به دلایل زیر رگرسیون خطی مورد توجه می‌باشد:

- اعمال ریاضی بر این معادلات به سادگی صورت می‌گیرد.
- اغلب تقریب بسیار خوبی برای معادلات رگرسیون پیچیده‌تر هستند.
- در حالت توزیع نرمال دو متغیره معادلات رگرسیون خطی‌اند.

**قضیه:** اگر رگرسیون  $y$  روی  $x$  خطی باشد آنگاه

$$\mu_{y|x} = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x) \Rightarrow \begin{cases} \beta = \rho \frac{\sigma_y}{\sigma_x} = \frac{\sigma_{xy}}{\sigma_x^2} \\ \alpha = \mu_y - \rho \frac{\sigma_y}{\sigma_x} \mu_x \end{cases}$$

ولی در اغلب مواقع اطلاعی از توزیع توأم دو متغیر تصادفی برای محاسبه رگرسیون خطی نداریم در این حالت از روی نمونه تصادفی مقادیر  $\alpha$  و  $\beta$  را تخمین می‌زنیم که معادله برآورد (تخمین) شده را به صورت زیر نشان می‌دهیم:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

**برآورد پارامترها به روش کمترین مربعات:**

• مدل مورد نظر برای n مشاهده به صورت زیر در نظر گرفته می‌شود:

$$Y_i = \alpha + \beta X_i + E_i$$

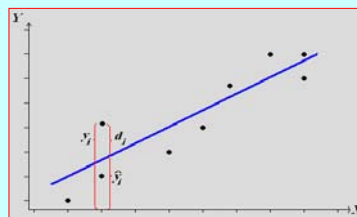
که در آن  $E_i$  ها متغیر تصادفی نرمال با میانگین صفر و واریانس  $\sigma^2$  می‌باشد.

در این روش برای تخمین ضرایب رگرسیون عبارت زیر را که مجموع مانده‌ها نمایده می‌شود، مینیمم می‌گردد:

$$SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

**برآورد پارامترها به روش کمترین مربعات: فرض کنید n نمونه**

از دو متغیر تصادفی X و Y مشاهده شده است که هر زوج از دو متغیر تصادفی به صورت  $(x_i, y_i)$  مسأله ما پیدا کردن مناسبترین خط به طوری که n نقطه تا جایی که امکان دارد به خط رگرسیون نزدیک باشد



**برآورد پارامترها به روش کمترین مربعات:**

پس از حل دستگاه معادلات نرمال ضرایب رگرسیون از روابط زیر به دست می‌آید:

$$1) \bar{x} = \frac{\sum x_i}{n} ; \bar{y} = \frac{\sum y_i}{n}$$

$$2) S_{xx} (SS_x) = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$$

$$3) S_{yy} (SS_y) = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$$

$$4) S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$$

$$5) SSE = S_{yy} - \hat{\beta}^2 S_{xx} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = S_{yy} - \hat{\beta} S_{xy}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad \& \quad \hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

**برآورد پارامترها به روش کمترین مربعات:**

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

$$SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \Rightarrow$$

$$\frac{\partial SSE}{\partial \hat{\alpha}} = -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0$$

$$\frac{\partial SSE}{\partial \hat{\beta}} = -2 \sum_{i=1}^n x_i (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0$$

این معادلات را معادلات نرمال می‌نامیم که از حل این دستگاه ضرایب رگرسیون بدست می‌آید.

$$n\hat{\alpha} + \hat{\beta} \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$\hat{\alpha} \sum_{i=1}^n x_i + \hat{\beta} \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

**تمرین:** مطلوبست معادله رگرسیونی خطی و مجموع مانده‌ها بین دو متغیر X و Y بر اساس نمونه تصادفی به اندازه ۱۵ طبق نتایج حاصله زیر از نمونه:

$$\sum x_i = 162 \quad \sum y_i = 1840.5 \quad \sum x_i y_i = 19945.7$$

$$\sum x_i^2 = 1820.2 \quad \sum y_i^2 = 1348$$

**تمرین:** برای تعیین رابطه هزینه حمل یک کالا که آن را با r نشان می‌دهیم، با فاصله فروشگاه که آن را با x نشان می‌دهیم، نمونه‌ای تصادفی شامل ۸ فروشگاه را انتخاب و براساس آن نتایج زیر را بدست آورده ایم:

(X)	۸	۱۰	۲۱	۱۵	۶	۱۲	۲۱	۱۴
(Y)	۱۱۳	۶۸	۹۱	۱۳۹	۹۵	۴۹	۵۶	۱۶۹

معادله خط رگرسیون و مجموع مانده‌ها را حساب کنید؟

**مثال:** مطلوبست خط رگرسیون و مجموع مانده‌ها برای داده‌های جدول زیر؟

$x_i$	$y_i$	$x_i y_i$	$x_i^2$	$y_i^2$	
4	31	124	16	961	
9	58	522	81	3364	
10	65	650	100	4225	
14	73	1022	196	5329	
4	37	148	16	1369	
7	44	308	49	1936	
12	60	720	144	3600	
22	91	2002	484	8281	
1	21	21	1	441	
17	84	1428	289	7056	
$\sum$	100	564	6945	1376	36562

$$\bar{x} = \frac{\sum x_i}{n} = \frac{100}{10} = 10$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{564}{10} = 56.4$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{6945 - 10 \times 10 \times 56.4}{1376 - 10 \times 10^2} = 3.471$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 56.4 - 3.471 \times 10 = 21.69$$

$$SSE = S_{yy} - \hat{\beta} S_{xy} = 222.745$$

روابط زیر در تحلیل رگرسیونی نرمال صادقند:

1)  $\hat{\alpha}$  is normal &  $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$$E(\hat{\alpha}) = \alpha \quad \& \quad \text{var}(\hat{\alpha}) = \sigma^2 \left( \frac{\sum x_i^2}{nS_{xx}} \right)$$

2)  $\hat{\beta}$  is normal &  $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$

$$E(\hat{\beta}) = \beta \quad \& \quad \text{var}(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$$

3)  $S^2 = \frac{SSE}{n-2}$  &  $E(S^2) = \sigma^2$

4)  $(n-2) \frac{S^2}{\sigma^2} \sim \chi^2_{n-2}$

**تحلیل رگرسیونی نرمال:** در این تحلیل فرض می‌شود که به ازای هر  $x_i$  ثابت، چگالی شرطی متغیرهای تصادفی  $y_i$ ، چگالی نرمال

$$w(y_i | x_i) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left[ \frac{y_i - (\alpha + \beta x_i)}{\sigma} \right]^2}, \quad y_i \in \square$$

است.

با استفاده از برآورد درست نمایی برای  $n$  مشاهده می‌توان  $\alpha$ ،  $\beta$  و  $\sigma$  را تخمین زد که با مقایسه نتایج آن مشاهده می‌گردد که همان نتایج روش کمترین مربعات برای تخمین پارامترها به دست می‌آید.

**آزمون فرض برای ضرایب رگرسیون:**

$H_0$	آماره آزمون	$H_1$	ناحیه بحرانی
$\alpha = \alpha_0$	$T = \frac{\hat{\alpha} - \alpha_0}{S \sqrt{\frac{\sum x_i^2}{nS_{xx}}}}$ $v = n - 2$	$\alpha < \alpha_0$	$T < -t_{1-\alpha}$
		$\alpha > \alpha_0$	$T > t_{1-\alpha}$
$\beta = \beta_0$	$T = \frac{\hat{\beta} - \beta_0}{S / \sqrt{S_{xx}}}$ $v = n - 2$	$\alpha \neq \alpha_0$	$T < -t_{1-\alpha/2}$ $T > t_{1-\alpha/2}$
		$\beta < \beta_0$	$T < -t_{1-\alpha}$
		$\beta > \beta_0$	$T > t_{1-\alpha}$
		$\beta \neq \beta_0$	$T < -t_{1-\alpha/2}$ $T > t_{1-\alpha/2}$

یک فاصله اطمینان  $100(1-\alpha)\%$  برای ضرایب رگرسیون و خط رگرسیون به ازای  $x$  مفروض به شکل زیر محاسبه می‌گردد:

$$T = \frac{\hat{\alpha} - \alpha}{S \sqrt{\frac{\sum x_i^2}{nS_{xx}}}} \sim t_{(n-2)}; \quad \alpha: \hat{\alpha} \pm t_{1-\alpha/2, (n-2)} S \sqrt{\frac{\sum x_i^2}{nS_{xx}}}$$

$$T = \frac{\hat{\beta} - \beta}{S / \sqrt{S_{xx}}} \sim t_{(n-2)}; \quad \beta: \hat{\beta} \pm t_{1-\alpha/2, (n-2)} \cdot S / \sqrt{S_{xx}}$$

$$T = \frac{\hat{y}_0 - y_0}{S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}}; \quad y_0: \hat{y}_0 \pm t_{1-\alpha/2, (n-2)} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

**مثال:** جدول زیر تعداد ساعات مطالعه در برابر نمره دریاقتی برای درس زبان برای ۱۰ نفر می باشد. با توجه به این داده ها مطلوبیت

• معادله خط رگرسیون  $y$  روی  $x$

• یک فاصله اطمینان ۹۵٪ برای  $\beta$

• آزمون فرض  $\beta=3$  در برابر  $\beta>3$  در سطح ۰/۰۱.

(X)	۴	۹	۱۰	۱۴	۴	۷	۱۲	۲۲	۱	۱۷
(Y)	۳۱	۵۸	۶۵	۷۳	۳۷	۴۴	۶۰	۹۱	۲۱	۸۴

**آزمون فرض برای خط رگرسیون به ازای  $x_0$  مفروض:**

$H_0$	آماره آزمون	$H_1$	ناحیه بحرانی
$y = y_0$	$T = \frac{\hat{y}_0 - y_0}{S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}}$ $v = n - 2$	$y < y_0$	$T < -t_{1-\alpha}$
		$y > y_0$	$T > t_{1-\alpha}$
		$y \neq y_0$	$T < -t_{1-\alpha/2}$ $T > t_{1-\alpha/2}$

$$\begin{cases}
 S_{xx} = \sum x_i^2 - n\bar{x}^2 = 1376 - 10 \times 10^2 = 376 \\
 S_{yy} = \sum y_i^2 - n\bar{y}^2 = 36562 - 10 \times (56.4)^2 = 4752.4 \\
 S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 6945 - 10 \times 10 \times 56.4 = \\
 SSE = S_{yy} - \hat{\beta} S_{xy} = 4752.4 - 3.47 \times 1305 = 224.05 \\
 S^2 = \frac{SSE}{n-2} = \frac{224.05}{8} = 28 \\
 \hat{\beta} - t_{1-\alpha/2, (n-2)} \cdot \frac{S}{\sqrt{S_{xx}}} < \beta < \hat{\beta} + t_{1-\alpha/2, (n-2)} \cdot \frac{S}{\sqrt{S_{xx}}} \\
 1-\alpha = 0.90 \Rightarrow 1-\alpha/2 = 0.975 \Rightarrow t_{0.975, (8)} = 2.31 \\
 3.47 - 2.31 \sqrt{\frac{28}{376}} < \beta < 3.47 + 2.31 \sqrt{\frac{28}{376}} \Rightarrow 2.84 < \beta < 4.1
 \end{cases}$$

$x_i$	$y_i$	$x_i y_i$	$x_i^2$	$y_i^2$	
4	31	124	16	961	
9	58	522	81	3364	
10	65	650	100	4225	
14	73	1022	196	5329	
4	37	148	16	1369	
7	44	308	49	1936	
12	60	720	144	3600	
22	91	2002	484	8281	
1	21	21	1	441	
17	84	1428	289	7056	
$\Sigma$	100	564	6945	1376	36562

$$\begin{aligned}
 \bar{x} &= \frac{\sum x_i}{n} = \frac{100}{10} = 10 \\
 \bar{y} &= \frac{\sum y_i}{n} = \frac{564}{10} = 56.4 \\
 \hat{\beta} &= \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \\
 &= \frac{6945 - 10 \times 10 \times 56.4}{1376 - 10 \times 10^2} = 3.471 \\
 \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} = 56.4 - 3.471 \times 10 \\
 &= 21.69 \\
 \hat{y} &= \hat{\alpha} + \hat{\beta}x = 21.7 + 3.47x
 \end{aligned}$$

**تمرین:** برای جدول مقادیر زیر معادله خط رگرسیون را بنویسید.

X	۵۲	۷۵	۳۴	۴۷	۵۷	۲۸	۳۹	۲۱	۴۳	۶۴
Y	۷۵	۹۸	۵۶	۸۹	۹۲	۷۳	۶۵	۵۲	۷۸	۸۲

یک فاصله اطمینان ۹۵٪ برای ضرایب رگرسیون به دست آورید.  
 یک فاصله اطمینان ۹۵٪ برای  $y_0$  پیدا کنید وقتی  $x_0=47$   
 آزمون فرض  $\beta=0$  را در مقابل  $\beta \neq 0$  در سطح ۰/۰۵ انجام دهید؟  
 آزمون فرض  $\alpha=40$  را در مقابل  $\alpha < 40$  در سطح ۰/۰۵ انجام دهید؟  
 آزمون فرض  $y_0=75$  را در مقابل  $y_0 > 75$  وقتی  $x_0=47$  در سطح ۰/۰۵ انجام دهید؟

- $H_0: \beta = 3$
  - $H_1: \beta > 3$
  - $\alpha = 0.01 \Rightarrow 1 - \alpha = 0.99$
  - $T = \frac{\hat{\beta} - \beta_0}{S/\sqrt{S_{xx}}}$  & Rejection region:  $T > t_{0.99, 8} = 2.90$
  - $5) T = \frac{3.47 - 3}{\sqrt{\frac{28}{376}}} = 1.722$
- ۶- چون T در ناحیه بحرانی نیست پس فرض  $H_0$  رد نمی‌شود.

**همبستگی:**  
 یادآوری: اگر X و Y دو متغیر تصادفی با تابع احتمال توأم  $f(x,y)$  باشند. ضریب همبستگی خطی بین X و Y به صورت زیر نشان داده و تعریف می‌کنیم:  
 $\rho = \rho(X, Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$   
 برای بررسی چگونگی همبستگی خطی بین دو متغیر تصادفی X و Y استفاده می‌کنیم.  
 اگر X و Y مستقل باشند ضریب همبستگی بین آنها صفر است.  
 ضریب همبستگی همواره در فاصله [-1, 1] قرار دارد.  
 ضریب همبستگی بستگی به واحد اندازه گیری ندارد.  
 همبستگی مستقیم: در این حالت اندازه های دو متغیر با هم تغییر می‌کنند یعنی اگر یکی زیاد شود دیگری هم زیاد می‌شود و برعکس  
 همبستگی معکوس: در این حالت دو متغیر در جهت مخالف هم می‌باشند.

$$\begin{aligned}
 \bar{x} &= \frac{\sum x_i}{n} = \frac{460}{10} = 46 & \bar{y} &= \frac{\sum y_i}{n} = \frac{760}{10} = 76 \\
 S_{xx} &= \sum x_i^2 - n\bar{x}^2 = 2472 & S_{yy} &= \sum y_i^2 - n\bar{y}^2 = 2051 \\
 S_{xy} &= \sum x_i y_i - n\bar{x}\bar{y} = 1894 \\
 \hat{\beta} &= \frac{S_{xy}}{S_{xx}} = \frac{1894}{2474} = 0.76 & \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} = 40/8 \\
 \hat{y} &= \hat{\alpha} + \hat{\beta}x = 40/8 + 0.76x \\
 SSE &= S_{yy} - \hat{\beta} S_{xy} = 2056 - 0.76 \times 1894 = 627/0 \\
 S^2 &= \frac{SSE}{n-2} = \frac{627/0}{8} = 78/37 \\
 T &= \frac{0.76 - 0}{\sqrt{\frac{78/37}{2474}}} = 4/47 & t_{0.025, 8} &= 2/306 \\
 \Rightarrow T &= 4/47 > 2/306 = t_{0.025, 8} \Rightarrow RH.
 \end{aligned}$$

### آزمون فرض برای معنی دار بودن همبستگی با توجه به r:

$H_0$	آماره آزمون	$H_1$	ناحیه بحرانی
$\rho = \rho_0$	$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ $v = n - 2$	$\rho < \rho_0$	$T < -t_{1-\alpha}$
		$\rho > \rho_0$	$T > t_{1-\alpha}$
		$\rho \neq \rho_0$	$T < -t_{1-\alpha/2}$ $T > t_{1-\alpha/2}$
$\rho = \rho_0$	$Z = \frac{\sqrt{n-3}}{2} \cdot \ln \frac{(1+r)(1-\rho)}{(1-r)(1+\rho)}$	$\rho < \rho_0$	$Z < -t_{1-\alpha}$
		$\rho > \rho_0$	$Z > t_{1-\alpha}$
		$\rho \neq \rho_0$	$Z < -t_{1-\alpha/2}$ $Z > t_{1-\alpha/2}$

### همبستگی:

حال اگر تابع احتمال توأم را نداشته باشیم و هدف تعیین چگونگی همبستگی بین دو متغیر تصادفی باشد برای تعیین آن دو نمونه تصادفی از دو متغیر تصادفی انتخاب و از رابطه زیر استفاده می‌کنیم.

$$\rho = \rho(X, Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$r = \frac{S_{xy}}{\sqrt{S_{yy} S_{xx}}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum y_i^2 - n\bar{y}^2)(\sum x_i^2 - n\bar{x}^2)}}$$

### فصل سیزدهم: تحلیل واریانس

**مثال:** فرض کنید یک نمونه ۱۰ تایی انتخاب و  $r=0.966$  محاسبه شده است میزان معنی دار بودن همبستگی را در سطح  $0.05$  آزمون کنید؟

1)  $H_0: \rho = 0$

2)  $H_1: \rho \neq 0$

3)  $\alpha = 0.05 \Rightarrow 1 - \alpha/2 = 0.975$

4)  $T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$  & Rejection region:  $\begin{cases} T < -t_{0.975,8} = -2.31 \\ T > t_{0.975,8} = 2.31 \end{cases}$

5)  $T = \frac{0.966\sqrt{8}}{\sqrt{1-(0.966)^2}} = 10.565$

4)  $Z = \frac{\sqrt{n-3}}{2} \cdot \ln \frac{(1+r)(1-\rho)}{(1-r)(1+\rho)}$  & Rejection region:  $\begin{cases} Z < -Z_{0.975} = -1.96 \\ Z > Z_{0.975} = 1.96 \end{cases}$

5)  $Z = \frac{\sqrt{10-3}}{2} \cdot \ln \frac{(1+0.966)(1-0)}{(1-0.966)(1+0)} = 5.37$

۶- در هر دو آزمون فرض صفر رد می‌شود، یعنی وجود همبستگی معنی‌دار است.

### تحلیل واریانس یکطرفه:

در حالت کلی در چنین مسائلی،  $k$  نمونه تصادفی به اندازه  $n$  از  $k$  جامعه داریم که مقدار  $z$  ام از جامعه  $i$  ام را با  $x_{ij}$  نشان می‌دهند.

جامعه ۱	$x_{11}$	$x_{12}$	...	$x_{1n}$
جامعه ۲	$x_{21}$	$x_{22}$	...	$x_{2n}$
...	...	...	...	...
جامعه $k$	$x_{k1}$	$x_{k2}$	...	$x_{kn}$

فرض می‌کنیم که  $x_{ij}$  ها مستقلند

و دارای توزیع‌های نرمال با

میانگین‌های  $\mu_i$  و واریانس مشترک

$\sigma^2$  باشند. پس می‌توان نوشت:

$$x_{ij} = \mu_i + e_{ij} \quad ; e_{ij} : n(0, \sigma) \Rightarrow$$

$$x_{ij} = \mu + \alpha_i + e_{ij}$$

که در اینجا به  $\mu$  میانگین کل اطلاق می‌شود و  $\alpha_i$  ها اثرهای تیماری نامیده می‌شوند که

$$\sum_{i=1}^k \alpha_i = 0$$

### مقدمه:

در مسائل تحلیل واریانس بررسی می‌کنیم که:

آیا تفاوت‌های مشاهده شده بین بیش از دو میانگین نمونه‌ای را می‌توان معلول تصادف دانست و یا بین میانگین‌های جامعه‌های مورد نمونه‌گیری تفاوت‌های واقعی وجود دارد؟

مثلاً ممکن است بخواهیم بر مبنای داده‌های نمونه‌ای تصمیم بگیریم آیا واقعاً تفاوتی بین میزان مؤثر بودن سه روش تدریس یک درس وجود دارد یا آیا واقعاً تفاوتی بین متوسط مصرف بنزین چند ماشین وجود دارد یا نه؟ یا آیا تفاوتی بین میزان مؤثر بودن چند ماده شیمیایی شوینده وجود دارد؟

### تحليل واريانس يکطرفه:

آزمون فرض ما عبارتست از فرض برابري ميانگين هاي جوامع  
 $(\mu_1 = \mu_2 = \dots = \mu_k)$  در برابر نابرابري آنها يا

$$\begin{cases} H_0: \alpha_i = 0 \\ H_1: \exists i; \alpha_i \neq 0 \end{cases}; i = 1, 2, \dots, k$$

قضيه:

$$\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{i\cdot})^2 = n \sum_{i=1}^k (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2 + \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{i\cdot})^2$$

$$SST = SS(Tr) + SSE$$

$\bar{x}_{i\cdot}$ : ميانگين مشاهدات جامعه i ام	$\bar{x}_{\cdot\cdot}$ : ميانگين همه nk مشاهده
SS(Tr): مجموع مربعات تيمارها	SSE: مجموع مربعات خطا
SST: مجموع كل مربعات	

### تحليل واريانس يکطرفه:

قضيه:

$$SST = \sum_{i=1}^k \sum_{j=1}^n x_{ij}^2 - \frac{1}{kn} T_{\cdot\cdot}^2 \quad \& \quad SS(Tr) = \frac{1}{n} \sum_{i=1}^k T_{i\cdot}^2 - \frac{1}{kn} T_{\cdot\cdot}^2$$

$T_{\cdot\cdot}$ : مجموع مشاهدات براي تيمار i ام  
 $T_{i\cdot}$ : مجموع همه nk مشاهده

$$\frac{1}{\sigma^2} \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{i\cdot})^2 : X_{k(n-1)}^2 \Rightarrow \frac{SSE}{k(n-1)} \sigma^2$$

$$\frac{n}{\sigma^2} \sum_{i=1}^k (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2 : X_{k-1}^2 \Rightarrow \frac{SS(Tr)}{k-1} \sigma^2$$

$$F = \frac{\frac{SS(Tr)}{k-1}}{\frac{SSE}{k(n-1)}} = \frac{k(n-1)SS(Tr)}{(k-1)SSE} : F(k-1, k(n-1))$$

### تحليل واريانس يکطرفه:

نتيجه آزمون:

F	ميانگين مربعات	مجموع مربعات	درجه آزادي	منبع تغيير
$\frac{MS(Tr)}{MSE}$	$MS(Tr) = \frac{SS(Tr)}{k-1}$	SS(Tr)	K-1	تيمارها
	$MSE = \frac{SSE}{k(n-1)}$	SSE	K(n-1)	خطا
		SST	Kn-1	جمع

هرگاه مقدار حاصل F از  $F_{1-\alpha}(k-1, k(n-1))$  بيشتر شود فرض صفر در سطح  $\alpha$  رد مي شود.

مثال: فرض كنيد كه بخواهيم عمل پاك كنندگي سه ماده پاك كننده را بر مبناي درجه سفيدي ۱۵ قواره پارچه سفيد كه ابتدا به مركب آلوده شده و سپس در يك ماشين لباس شوي با اين سه ماده پاك كننده شسته شده اند مقايسه كنيم.

در سطح معني دار بودن ۰/۰۱ آزمون برابري ميانگين ها را انجام دهيد؟	۸۰	۷۶	۷۱	۸۱	۷۷	جامعه (تيمار) ۱
	۷۰	۶۶	۷۴	۵۸	۷۲	جامعه (تيمار) ۲
	۷۷	۸۰	۸۲	۸۵	۷۶	جامعه (تيمار) k

$$\begin{cases} 1) H_0: \alpha_i = 0 \\ 2) H_1: \exists i; \alpha_i \neq 0 \end{cases}; i = 1, 2, \dots, k$$

$$3) \alpha = 0.01 \Rightarrow 1 - \alpha = 0.99$$

$$4) F = \frac{MS(Tr)}{MSE} \quad \& \quad \text{Rejection region:}$$

$$F > F_{1-\alpha}(k-1, k(n-1)); F > F_{0.99}(2, 12) = 6.93$$

### تحليل واريانس دو طرفه:

داده هاي زير مربوط به زمان لازم (برحسب دقيقه) براي شخصي است كه با اتومبيل خود از شنبه تا چهارشنبه با استفاده از چهار مسير مختلف به سر كار خود مي رسد:

مسير ۱	۲۲	۲۶	۲۵	۲۵	۳۱
مسير ۲	۲۵	۲۷	۲۸	۲۶	۲۹
مسير ۳	۲۹	۲۹	۳۳	۳۰	۳۳
مسير ۴	۲۶	۲۸	۲۷	۳۰	۳۰

ميانگين هاي اين چهار نمونه عبارتند از ۲۵/۸، ۲۷/۰، ۳۰/۲ و ۲۸/۲. چون اختلاف بين ميانگين ها نسبتا بزرگ است اين استنتاج معقول خواهد بود كه اختلاف بين متوسط زمان لازم براي رسيدن شخص به محل كار خود از چهار مسير مختلف وجود دارد ولي اين امر از تحليل واريانس يکطرفه نتيجه نمي شود زيرا با استفاده از محاسبات لازم داريم

$$F = 2.80 \quad \& \quad F_{0.95, 3, 16} = 3.24$$

(۵) محاسبه آماره

$T_1=385$	$T_2=340$	$T_3=400$	$T_{\cdot\cdot}=1125$	$\sum \sum x_{ij}^2 = 85041$
-----------	-----------	-----------	-----------------------	------------------------------

$$SST = 85041 - 1/15(1125)^2 = 666$$

$$SS(Tr) = 1/5(385^2 + 340^2 + 400^2) - 1/15(1125)^2 = 390$$

$$SSE = SST - SS(Tr) = 666 - 390 = 276$$

F	ميانگين مربعات	مجموع مربعات	درجه آزادي	منبع تغيير
$195/23 = 8.48$	$390/2 = 195$	390	2	تيمارها
	$276/12 = 23$	276	12	خطا
		66	14	جمع

(۶) چون مقدار F در ناحيه بحراني قرار مي گيرد پس فرض صفر رد مي گردد.

### تحليل واريانس دو طرفه:

پس اين اختلاف هاي نسبتاً بزرگ بين ميانگين ها و مقادير بين نمونه ها را در هر دسته معلول کدام علت مي توان دانست؟  
اختلاف و تغيير در داخل نمونه ها مي تواند معلول شرايط رانندگي در روزهاي مختلف هفته دانست که اگر چنين باشد، اين تغييرات، مجموع مربعات خطا را در تحليل واريانس يک طرفه تحت تاثير خود قرار مي دهد که اين اثر موجب تغيير آماره F (بزرگتر شدن مخرج آن و در نهايت کوچک شدن خود F) مي شود که اين تغيير خود باعث بي معني شدن نتيجه آزمون مي گردد.

#### تکليف چيست؟

يک راه حل اين است که عامل غير مربوط (شرايط رانندگي در روزهاي مختلف) را ثابت بگيريم براي مثال نتايج را براي يک روز در نظر بگيريم. ولي با اين کار به ندرت به اطلاعاتي که لازم داريم، ميرسيم.

### تحليل واريانس دو طرفه:

امکان ديگر اين است که عامل غير مربوط را در روي بردي به وسعت لازم تغيير دهيم به طوري که به توان تغيير ناشي از آن را اندازه گرفته و بنابر اين از مجموع مربعات خطا حذف نمود. يعني يک تحليل واريانس دو طرفه انجام داد که در آن، تغييرات کل داده ها به سه جزء تيمارها (در مثال ما چهار مسير) عامل غير مربوط (در مثال ما شرايط رانندگي در روزهاي مختلف) و خطاي آزمايش افزايش مي شود.  
روش پيشنهادي **بلوک بندي** نام دارد و به اجزاي عامل غير مربوط (در مثال ما شرايط رانندگي در روزهاي مختلف) بلوک اطلاق مي شود.  
پس بلوک ها سطوحی هستند که در آن سطوحها عامل غير مربوط ثابت در نظر گرفته مي شود و اگر هر تيمار در هر بلوک به تعداد دفعات مساوي ظاهر شود گوييم آزمايش يک **طرح بلوکی کامل** است.

### تحليل واريانس دو طرفه:

فرض مي کنيم که  $x_{ij}$  ها مستقلند و داراي توزيع هاي نرمال با ميانگين هاي  $\mu_{ij}$  و واريانس مشترك  $\sigma^2$  باشند. پس مي توان نوشت

	بلوک ۱	بلوک ۲	...	n
تيمار ۱	$x_{11}$	$x_{12}$	...	$x_{1n}$
تيمار 2	$x_{21}$	$x_{22}$	...	$x_{2n}$
...	...	...	...	...
تيمار k	$x_{k1}$	$x_{k2}$	...	$x_{kn}$

$$x_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

$$e_{ij} : n(0, \sigma)$$

که در اینجا به  $\mu$  ميانگين کل اطلاق مي شود و  $\alpha_i$  ها اثرهاي تيماري و  $\beta_j$  ها اثرهاي بلوکی ناميده مي شوند که

$$\sum_{i=1}^k \alpha_i = 0 \quad \& \quad \sum_{j=1}^n \beta_j = 0$$

### تحليل واريانس دو طرفه:

آزمون فرضي که انجام خواهيم داد عبارت است از دو فرض صفر "اثرهاي تيماري و بلوکی صفرند" در برابر دو فرض مقابل "اثرهاي تيماري و بلوکی همه برابر صفر نيستند"

$$\begin{cases} H_0: \alpha_i = 0 & ; i = 1, 2, \dots, k \\ H'_0: \beta_j = 0 & ; j = 1, 2, \dots, n \\ H_1: \exists i ; \alpha_i \neq 0 \\ H'_1: \exists j ; \beta_j \neq 0 \end{cases}$$

### تحليل واريانس دو طرفه:

قضيه:

$$\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{\cdot\cdot})^2 = n \sum_{i=1}^k (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2 + kn \sum_{j=1}^n (\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot})^2 + \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{\cdot\cdot})^2$$

$$SST = SS(Tr) + SSB + SSE$$

$\bar{x}_{i\cdot}$	ميانگين مشاهدات تيمار i ام	$\bar{x}_{\cdot\cdot}$	ميانگين همه nk مشاهده
$\bar{x}_{\cdot j}$	ميانگين مشاهدات بلوک j ام	SST	مجموع کل مربعات
SS(Tr)	مجموع مربعات تيمارها	SSE	مجموع مربعات خطا
SSB	مجموع مربعات تيمارها		

### تحليل واريانس دو طرفه:

قضيه:

$$SSB = \frac{1}{k} \sum_{j=1}^n T_{\cdot j}^2 - \frac{1}{kn} T_{\cdot\cdot}^2$$

T<sub>..</sub>: مجموع کل تمام مشاهدات | T<sub>·j</sub>: مجموع مقادير حاصل براي بلوک j ام

**مثال:** برای جدول داده‌های زیر در سطح معنی‌دار بودن  $0.05$  آزمون کنید که آیا اختلاف‌های بین میانگین‌های حاصل برای مسیرهای مختلف (تیمارها) معنی‌دار هستند یا نه، و نیز آیا اختلاف‌های بین میانگین‌های حاصل برای روزهای مختلف هفته (بلوک‌ها) معنی دارند یا نه؟

مسیر ۱	۲۲	۲۶	۲۵	۲۵	۳۱
مسیر ۲	۲۵	۲۷	۲۸	۲۶	۲۹
مسیر ۳	۲۹	۲۹	۳۳	۳۰	۳۳
مسیر ۴	۲۶	۲۸	۲۷	۳۰	۳۰

1)  $\begin{cases} H_0: \alpha_i = 0 & ; i = 1, 2, \dots, k \\ H'_0: \beta_j = 0 & ; j = 1, 2, \dots, n \end{cases}$  2)  $\begin{cases} H_1: \exists i; \alpha_i \neq 0 \\ H'_1: \exists j; \beta_j \neq 0 \end{cases}$

3)  $\alpha = 0.05 \Rightarrow 1 - \alpha = 0.95$

4)  $\begin{cases} F_{Tr} = \frac{MS(Tr)}{MSE} \\ F_B = \frac{MSB}{MSE} \end{cases} \Rightarrow \text{Rejection region: } \begin{cases} F_{Tr} > F_{1-\alpha}(k-1, (k-1)(n-1)) \\ ; F > F_{0.95}(3, 12) = 3.49 \\ F_B > F_{1-\alpha}(n-1, (k-1)(n-1)) \\ ; F > F_{0.95}(4, 12) = 3.26 \end{cases}$

**تحلیل واریانس دو طرفه: نتیجه آزمون**

منبع تغییر	درجه آزادی	مجموع مربعات	میانگین مربعات	F
تیمارها	k-1	SS(Tr)	$MS(Tr) = \frac{SS(Tr)}{k-1}$	$F_{Tr} = \frac{MS(Tr)}{MSE}$
بلوک‌ها	n-1	SSB	$MSB = \frac{SSB}{n-1}$	$F_B = \frac{MSB}{MSE}$
خطا	(K-1)(n-1)	SSE	$MSE = \frac{SSE}{(k-1)(n-1)}$	
جمع	Kn-1	SST		

فرض‌های صفر رد می‌شود اگر:

$$F_{Tr} \geq F_{1-\alpha}(k-1, (k-1)(n-1)) \text{ \& } F_B \geq F_{1-\alpha}(n-1, (k-1)(n-1))$$

**۶) نتیجه آزمون:**

چون  $F_{Tr} = 7.75$  از  $F_{0.95, 3, 12} = 3.49$  بیشتر است و همچنین  $F_B = 8.06$  از  $F_{0.95, 4, 12} = 3.26$  بیشتر است نتیجه می‌گیریم که هر دو فرض صفر باید رد گردد به عبارت دیگر، اختلاف‌های بین میانگین‌های حاصل برای چهار مسیر مختلف و همچنین اختلاف‌های بین میانگین‌های حاصل برای روزهای مختلف هفته معنی‌دار است.

**۵- محاسبه آماره**

$T_1=99$	$T_2=110$	$T_3=113$	$T_4=111$	$T_5=123$
$T_1=129$	$T_2=135$	$T_3=151$	$T_4=141$	$T_5=556$
$\sum x^2 = 15610$				
$SST = 15610 - 1/20(556)^2 = 153.20$				
$SS(Tr) = 1/5(129^2 + \dots + 141^2) - 1/20(556)^2 = 52.8$				
$SSB = 1/4(99^2 + \dots + 123^2) - 1/20(556)^2 = 73.2$				
$SSE = 153.2 - 52.8 - 73.2 = 27.2$				

منبع تغییر	درجه آزادی	مجموع مربعات	میانگین مربعات	F
تیمارها	3	52.8	$52.8/3 = 17.6$	$17.6/2.27 = 7.75$
بلوک‌ها	4	73.2	$73.2/4 = 18.3$	$18.3/2.27 = 8.06$
خطا	$3*4 = 12$	27.2	$27.2/12 = 2.27$	
جمع	$20-1 = 19$	153.2		

