**به نام خدا**

# فصل اول

درس  آمار و کاربرد آن در مدیریت یکی از دروس اصلی رشته مدیریت بوده و هدف آن آشناسازی دانشجویان با علم آمار و نحوه بکارگیری آن در دانش مدیریت است

**کلیات**

- **تعریف آمار:** روش علمی است که برای جمع آوری ، تلخیص ، تجزیه و تحلیل ، تفسیر و بطور کلی برای مطالعه و بررسی مشاهدات بکار گرفته می شود.
- **از فنون آماری در مدیریت برای چه مقاصدی استفاده می شود ؟**

۱- برای تبدیل داده ها به اطلاعات  (با بررسی قیمت سهام در مورد آینده قیمت آنها نظر دادن)

۲- برای بررسی صحت و سقم فرضیات (اجرای یک سیستم اتوماسیون نامه نگاری در سرعت رسیدگی به مراجعات تاثیری دارد یا نه؟)

۳- برای تعیین اعتبار و پایایی تحقیقات پرسشنامه ای و مصاحبه ای  ( پرسشنامه GHQ در بررسی سلامت روانی افراد به خوبی عمل می کند.)

- **تعریف جامعه:** جامعه بزرگترین مجموعه از موجودات است که در یک زمان معین ، مطلوب ما قرار می گیرند**.** مثل جامعه فرهنگیان ایران و . . .

لازم به ذکر است جامعه متناسب با هدف شما تغییر می کند. به طور مثال اگر هدف شما بررسی رضایتمندی کارکنان بانک مرکزی باشد، جامعه مورد نظر شما تمام کارکنان بانک مرکزی می باشد.

اگر هدف شما بررسی رضایتمندی معلمان باشد جامعه شما به صورت تمام معلمان تعریف می شود و..

- **جامعه آماری:** تعدادی از عناصر مطلوب مورد نظر که حداقل دارای یک صفت مشخصه باشند.
- **صفت مشخصه:** صفتی است که بین همه عناصر جامعه آماری مشترک و متمایز کننده جامعه آماری از سایر جوامع باشد.
- **انواع جامعه آماری**

۱- **محدود :** یعنی جامعه مقادیر از تعداد محدود و ثابتی تشکیل شده و پایان پذیر باشد. (کارکنان بانک مرکزی)

۲- **نا محدود :** یعنی جامعه از یک ردیف بی انتهایی از مقادیر تشکیل شده باشد. ( تمام برگ های درختان)

- **تعریف نمونه:** نمونه عبارتست از تعداد محدودی از آحاد جامعه آماری که بیان کننده ویژگی های اصلی جامعه باشد . ( نمونه انتخاب شده باشد تا حد ممکن شبیه جامعه مورد نظر باشد. برای بررسی یک روش آموزش درست نیست فقط دانشجویان با معدل بالا را انتخاب کرد بلکه انتخاب دانشجویان از هر طیف معدل الزامی است.)

- **انواع شاخص های آماری**

۱- **پارامتر** : شاخص هایی که از طریق سرشماری ( اندازه گیری تمامی عناصر جامعه آماری ) بدست می آیند.. ( محاسبه متوسط درآمد کارکنان بانک مرکزی با استفاده از اندازه گیری درآمد تمام کارکنان دولت.)

۲- **آماره** : شاخص هایی که از طریق نمونه گیری ( اندازه گیری بخشی از جامعه ) بدست می آیند. ( محاسبه متوسط درآمد کارکنان بانک مرکزی با استفاده از اندازه گیری درآمد نمونه ای از کارکنان دولت.)

- **روش های ناپارامتریک**

در جوامع آماری ای که از توزیع نرمال برخوردار نیستند و داده های غیرکمی (کیفی) با نمونه های کوچک را می توان با این فنون بررسی کرد.

- **سیر تحول علم آمار از نظر موضوعی عبارت است از:**

۱- آمار توصیفی

۲- آمار استنباطی

۳- آمار ناپارامتریک

- **آمار توصیفی**

این توع آمار به توصیف جامعه می پردازد و هدف آن محاسبه پارامترهای جامعه است. چنانچه محاسبه مقادیر و شاخص های جامعه آماری با استفاده از سرشماری تمامی عناصر آن انجام گیرد به آن آمار توصیفی گویند.

- **آمار استنباطی**

در این نوع آمار با استفاده از مقادیر نمونه، آماره ها محاسبه شده و به کمک تخمین و آزمون فرض آماری، آماره ها به پارامترهای جامعه تعمیم داده می شود. ( در بحث های آماری هرجا سخن از استنباط و استنتاج باشد، به آن آمار استنباطی گویند.)

- **آمار ناپارامتریک**

آمار ناپارامتریک در مقابل آمار پارامتریک مطرح می شود یکی از فرض های اساسی در آمار پارامتریک برخوردار بودن مشاهدات از تویع نرمال است، در حالی که در فنون ناپارامتریک این فرض ضرورتی ندارد. در بررسی هایی که متغیرهای آنهادارای مقیاس کیفی هستند، از این روش ها استفاده می شود چرا که متغیرهایی که دارای مقیاس کیفی هستند فاقد توزیع آماری بوده و به آنها آزاد توزیع گویند

- **مراحل پژوهش علمی در آمار**

۱- مشخص کردن هدف

۲- جمع آوری داده ها

۳- تجزیه و تحلیل داده ها

۴- بیان یافته ها

### ➢ دو عنصر اصلی تحقیقات رفتاری و مدیریتی

۱- فرضیه های تحقیق

۲- متغیرهایی که برای آزمودن آنها بکار گرفته می شوند.

### ➢ نقش متغیرها در فرضیات

متغیرها ، فرضیه ها را بصورتی نشان می دهند که محققان رفتاری و مدیریتی بتوانند آنها ( فرضیه ها ) را مشاهده و اندازه گیری نمایند.

### ➢ انواع متغیرها

۱- متغیر خصیصه

۲- متغیر مستقل

۳- متغیر وابسته

۴- متغیر تعدیل کننده ( واسطه ای )

۵- متغیر کنترل

- **متغیر خصیصه**

متغیری که مقدار آن از یک فرد به فرد دیگر و یا از یک عضو به عضو دیگر جامعه آماری ممکن است تغییر کند . مثل اندازه سازمان، قد افراد، رنگ چشم افراد ، نظر آنها در مورد یک موضوع خاص و ..

- **متغیر مستقل**

به علت احتمالی یا فرضی متغیر وابسته ، متغیر مستقل یا متغیر درونداد و به عبارتی محرک گفته می شود .

مثال: فرض کنید می خواهیم تاثیر مصرف شیر را روی افرایش قد افراد بررسی کنیم در این حالت مصرف شیر روی افزایش قد تاثیر دارد ولی افزایش قد روی مصرف شیر تاثیر نداردبنابراین متغیر مستقل مصرف شیر است.

- **متغیر وابسته**

به متغیری که به تبع تغییر متغیر مستقل ، مقدارش کم و زیاد می شود متغیر وابسته ، متغیر پاسخ و یا برونداد اطلاق می شود. مثال: در مثال قبل افزایش قد از آنجایی که وابسته به مصرف شیر است به عنوان متغیر وابسته شناخته می شود.

- **متغیر تعدیل کننده ( واسطه ای )**

متغر ثانوی است که پژوهشگر می خواهد تاثیر آن را در متغیر مستقل اولیه و متغیر وابسته ملاحظه کند. این متغر بدین منظور انتخاب می شود که روشن شود آیا این متغیر ، رابطه بین متغیر مستقل و متغیر وابسته را تحت تأثیر قرار می دهد یا نه.

- **متغیر کنترل**

به متغیرهایی که در موقع انجام پژوهش ، لازم است تأثیر آنها خنثی شده و یا از بین برود، متغیرهای کنترل می گویند.

**فرق متغیر تعدیل کننده با متغیر کنترل**

موقع انجام تحقیق ، پژوهشگر سعی می کند تأثیرات متغیر کنترل را از بین ببرد ولی تأثیرات متغیر تعدیل کننده را مورد بررسی قرار می دهد.

- **مقیاس های اندازه گیری متغیر ها**

۱- مقیاس اسمی ( *Nominal scale* )

۲- مقیاس ترتیبی ( رتبه ای) ( *Rank scale* )

۳- مقیاس فاصله ای ( *Interval scale* )

۴- مقیاس نسبی (*Ratio scale* )

- **مقیاس رسمی یا طبقه ای**

در این نوع مقیاس محققین از اعداد یا سمبولها صرفا برای طبقه بندی اشیا، اشخاص یا خصوصیات استفاده می کنند. مثال: در مورد جنسیت ممکن است عدد ۱ را برای مرد و عدد ۲ را برای زن انتخاب کنیم ولی این ارقام مفهومی از رتبه را در برندارند.

- **مقیاس ترتیبی**

اگر صرف نظر از تفاوت محتویات یک طبقه با طبقه دیگر، یک نوع ارتباط بین آنها برقرار باشد در آن صورت می گویند متغیر مورد نظر دارای مقیاس ترتیبی است. به طور مثال در طیف لیکرت

- **مقیاس فاصله ای**

در مقیاس ترتیبی زمانی که فاصله بین دو طبقه نیز برای ما اهمیت داشته و به طور دقیق قابل اندازه گیری باشد. در این حالت به آن مقیاس فاصله ای گویند. در این نوع اندازه گیری نسبت هر دو فاصله، مستقل از واحد اندازه گیری و مستقل از نقطه صفر است.

**مقیاس نسبی**

مقیاسی است که علاوه بر داشتن همه خصوصیات مقیاس فاصله ای ، دارای نقطه صفر واقعی نیز هست ، مثل پوند و گرم

**جدول مقادیر مقیاس های چهارگانه**

| مراتب / نوع مقیاس | ترتیب | فواصل | مبدأ صفر قراردادی | مبدأ صفر مطلق |
|---|---|---|---|---|
| اسمی | ندارد | ندارد | ندارد | ندارد |
| رتبه ای | دارد | ندارد | ندارد | ندارد |
| فاصله ای | دارد | دارد | دارد | ندارد |
| نسبتی | دارد | دارد | دارد | دارد |

- **فرضیه**

عبارتی آزمایشی است که رابطه بین دو یا چند متغیر که بصورت دقیق و روشن بیان می کند و پس از آزمایش ، صحت یا سقم آن مشخص می شود .

## ➢ ویژگی های یک فرضیه خوب

۱- واضح و بدون ابهام (بیان در قالب جملات خبری )

۲- قابل تبیین ( علت یابی )

۳- بیان کننده رابطه مورد انتظار بین متغیرها

۴- قابل آزمون بودن ( آزمون پذیری )

## ➢ انواع فرضیه های پژوهشی

۱-فرضیه های توصیفی در مقابل فرضیه های استنباطی

۲- فرضیه های تک متغیره در مقابل فرضیه های چندمتغیره

۳-فرضیه های همبستگی در مقابل فرضیه های تجربی

۴- فرضیه های پژوهشی با گروه های زوج شده در مقابل فرضیه های مستقل

۵- فرضیه های پارامتریک در مقابل ناپارامتریک

### • فرضیه توصیفی درمقابل فرضیه استنباطی

فرضیه توصیفی فرضیه ای است که در مورد کل جامعه آماری تدوین شده بعبارتی ادعایی را در مورد کل جامعه آماری بیان می نماید. در حالی که فرضیه استنباطی به فرضیه ای اطلاق می شود که در مورد یک نمونه انتخابی از کل جامعه آماری تدوین شود و صحت و سقم آن تحت تأثیر خطای نمونه گیری باشد

مثال: متوسط قد دانشجویان در ایران ۱۶۰ است. اگر برای بررسی این فرضیه متوسط قد کل دانشجویان ایران اندازه گیری شود فرضیه مورد نظر توصیفی و اگر برای بررسی این فرضیه قد یک نمونه از دانشجویان اندازه گیری شود و براساس توزیع و در نظر گرفتن خطای نمونه گیری به آن پاسخ داده شود، فرضیه استنباطی است.

- **فرضیه تک متغیره در مقابل چند متغیره**

بر اساس تعداد متغیرهایی که در یک فرضیه حضور دارند آن را تک متغیره یا چند متغیره می نامند.

مثال: در فرضیه"قد دانشجویان در ایران ۱۶۰ است" با توجه به اینکه تنها متغیری که در فرضیه حضور دارد قد دانشجویان است، فرضیه تک متغیره و در فرضیه " بین انتخاب یک سیستم تشویقی مناسب برای کارکنان و افزایش کارایی سازمان رابطه معناداری وجود دارد. " با توجه به آنکه دو متغیر "انتخاب یک سیستم تشویقی مناسب "و "افزایش کارایی سازمان" در فرضیه حضور دارند فرضیه چند متغیره است.

- **فرضیه همبستگی در مقابل فرضیه تجربی**

در فرضیه های همبستگی و تجربی هدف بررسی وجود رابطه بین دو یا چند متغیر است با این تفاوت که در فرضیه همبستگی هیچ یک از متغیرهای مورد بررسی تحت کنترل پژوهشگر نیست ولی در فرضیه های تجربی حداقل یکی از متغیرها تحت کنترل پژوهشگر است.

مثال: بین تورم و نرخ بیکاری رابطه معناداری وجود دارد. در این فرضیه بررسی همبستگی بین دو متغیر مدنظر است، ( تورم و نرخ بیکاری) که هیچ یک تحت کنترل پژوهشگر نیست . پس فرضیه مورد نظر از نوع همبستگی است.

بین تغییر محل خدمت کارکنان از محل A به محل B و میزان کارایی آنها رابطه معنا داری وجود دارد. در این فرضیه نیز بررسی از نوع همبستگی است ولی انتخاب اینکه چه کارمندانی در محل A و کدام یک از کارمندان در محل B خدمت کنند در اختیار پژوهشگر است بنابر این فرضیه از نوع تجربی است.

- **فرضیه با گروه های جور شده در مقابل گروه های مستقل**

در فرضیه با گروه های زوج شده، پژوهشگران یک گروه نمونه دارند که در آن هر آزمون شونده را از لحاظ یک متغیر واحد دو بار اندازه گیری می کنند ولی در فرضیه با گروه های مستقل، محقق برای آزمون ، دو گروه دارد که هر کدام از آنها را از لحاظ یک متغیر واحد مشابه یک بار بطور جداگانه اندازه گیری می نماید.

مثال: استفاده از سیستم آموزشی A بر روی کارایی کارمندان اثر مثبت دارد. برای بررسی چنین فرضیه ای یک گروه از کارمندان انتخاب شده و کارایی آنها قبل و بعد از آموزش بررسی می شود و بین این دو دسته از اطلاعات بررسی انجام

می شود. در این مورد ما تنها یک گروه داشتیم که به صورت قبل و بعد بررسی شدند پس فرصیه مورد نظر از نوع فرضیه با گروه های جور شده است.

کارایی کارکنان سازمان A بیشتر از کارایی کارکنان شرکت B است در این مورد ما دو دسته از افراد را مورد بررسی قرار می دهیم بنابراین فرضیه بیان شده، فرضیه با گروه های مستقل است.

- **فرضیه های پارامتریک در مقابل ناپارامتریک**

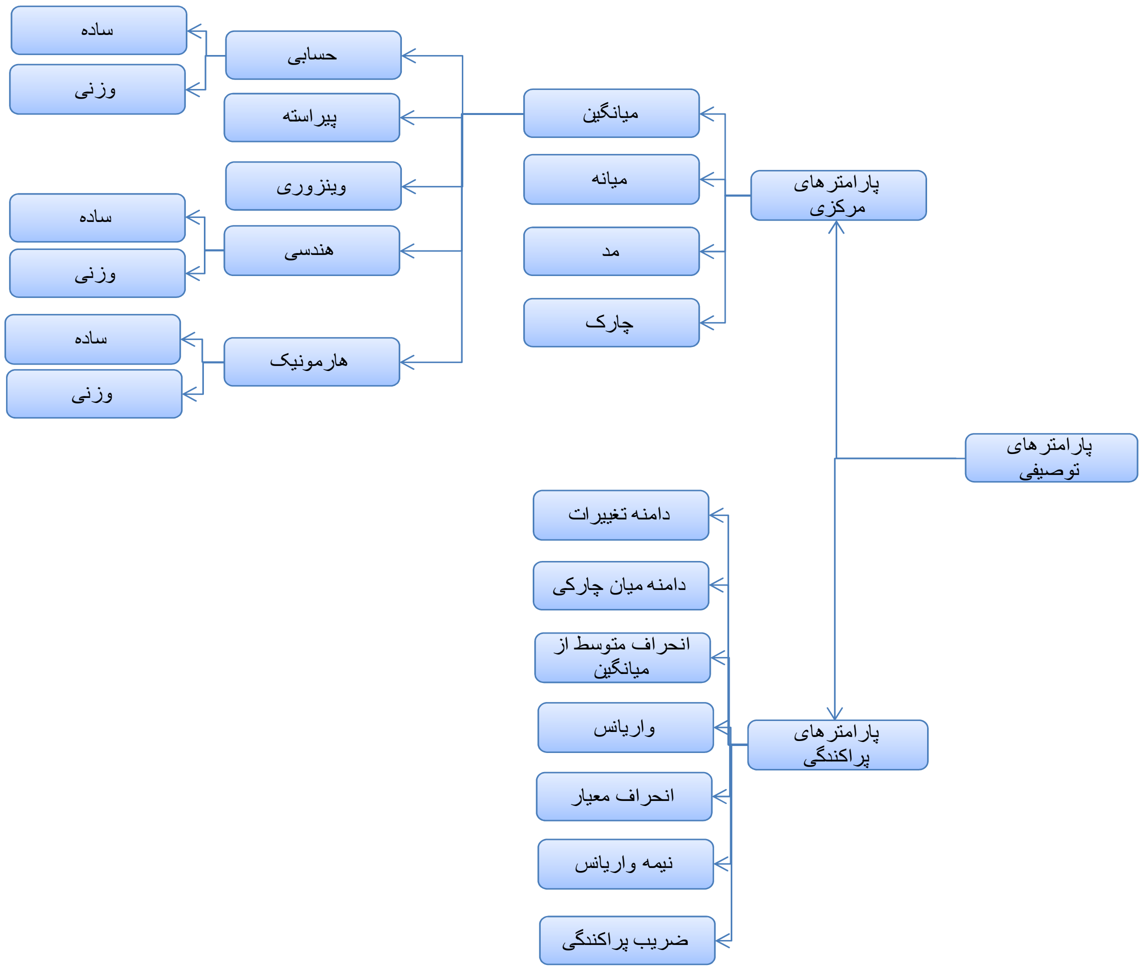
فرضیه های پارامتری فرضیه هایی هستند که در آنها از متغیرهای نسبی یا فاصله ای استفاده شده و توزیع جامعه (و یا نمونه ) نرمال می باشد. در مقابل فرضیه های ناپارامتریک، فرضیه هایی هستند که متغیرهای موجود در آنها دارای مقیاس اسمی یا رتبه ای می باشند یا این که بر اساس شواهد موجود ، محققان نمی توانند فرض نرمال بودن جامعه ( نمونه ) را بپذیرند.

## فصل دوّم: مطالعه توصیفی داده های طبقه بندی نشده

هدف این فصل آشناسازی دانشجویان با پارامترهای مرکزی و پراکندگی در جوامع کوچک ( $N \leq ۲۰$ ) می باشد.

در فعالیتهای متنوعی از قبیل تجربیات آزمایشگاهی ، نظر خواهی از مردم و بررسی سندهای تاریخی با فرآیند جمع آوری داده ها سروکار داریم. فرآیند جمع آوری داده ها هرچه باشد ، مجموعه داده های حاصل معمولاً متشکل از اندازه های عددی است که می تواند از تعدادی محدودی ارقام تا صدها و بلکه هزارها عدد را در بر داشته باشد و در مواجهه با تعداد زیادی از اندازه ها ذهن انسان نمی تواند محتوای کلی اطلاعات ثبت شده در مجموعه داده را فوراً درک کند . خلاصه کردن و توضیح خصوصیات مهم مجموعه دادها را معمولاً آمار توصیفی می نامند.

پارامترهای توصیفی که در این فصل مطالعه می شوند را می توان به صورت زیر بیان کرد.




قبل از بیان شاخص ها چند نماد معرفی می کنیم.

مجموعه داده ها متشکل از تعدادی اندازه است که به طور نمادی به صورت $x_1, x_2, \cdots, x_n$ نشان داده می شوند . آخرین زیرنویس ( $n$ ) مربوط به $x_n$ ، نشان دهنده تعداد داده هاست، و $x_1, x_2, \cdots$ به ترتیب نشان دهنده اولین مشاهده، دومین مشاهده، و الی آخر، هستند.

در مطالعه آمار ، همه جا با عمل جمع کردن داده ها یا اعداد دیگری که با استفاده از داده ها به دست می آیند ، سروکار داریم . برای اجتناب از نوشتن مکرر علامت (+) ، نماد $\sum$ (حرف بزرگ یونانی سیگما) را به عنوان اختصار ریاضی برای عمل جمع به کار می بریم.

نماد $\sum_{i=1}^{n} x_i$ نشان دهنده مجموع $n$ عدد $x_1, x_2, \cdots, x_n$ است این نماد به این صورت خوانده می شود : مجموع تمام $x_i$ ها ، که $i$ از ۱ تا $n$ تغییر می کند.

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

جمله ای که بعد از نماد $\sum$ نوشته می شود، نشان دهنده مقادیری است که باید جمع شوند، و نمادهای پایین و بالای $\sum$ دامنه زیرنویس $i$ را معین می کنند.

## پارامتر مرکزی

به هر معیار عددی که معرف مرکز مجموعه داده ها باشد ، پارامتر مرکزی اطلاق می شود یعنی همان مقدار نماینده ای که مشاهدات در اطراف آن توزیع شده اند.

### مهمترین پارامترهای مرکزی

**۱- میانگین؛ شامل میانگین حسابی، میانگین پیراسته، میانگین وینزوری، میانگین هندسی، میانگین هارمونیک**

**۲- میانه**

**۳- مد ( نما )**

**۴- چارکها؛ شامل چارک اول، چارک دوم، چارک سوم**

## ۱- میانگین

به نقطه تعادل یا مرکز ثقل توزیع، در داده هایی که بصورت منظم بر روی یک محور ردیف شده باشند، میانگین ( Mean ) اطلاق می شود.

### ۱-۱ میانگین حسابی

#### ۱-۱-۱ میانگین حسابی ساده

**این میانگین از تقسیم مجموع مشاهدات بر تعداد آنها بدست می آید.**

$$\mu_x = \frac{\sum_{i=1}^{n} X_i}{N}$$

#### ۱-۱-۲ میانگین حسابی موزون

اگر هر یک از مشاهدات دارای وزنی باشند، و این وزن ها را با $W_i$ نشان دهیم. در این صورت میانگین حسابی موزون به صورت زیر تعریف می شود.

$$\mu_w = \frac{\sum_{i=1}^{k} W_i X_i}{\sum_{i=1}^{k} W_i} = \frac{\sum W_i X_i}{N}$$

مثال۱: الف) دروس ابتدایی فردی به شرح زیر است، مطلوب است معدل او

| | | | |
|---|---|---|---|
| در | ریا | ادب | ع |
| ں | ـہ | ي | |

| | | | |
|---|---|---|---|
| | ر | ا ر | |
| نمر ه | ۱۸ | ۱۹ | ۲ |

$$\mu_x=\frac{۱۸+۱۹+۲۰}{۳}=۱۹$$

ب) نمرات درس دبیرستان به شرح زیر می باشد، دقت فرمائید که هر درس بر حسب اهمیت خود واحدی نیز دارد :

$$\mu_x=\frac{۴\times۱۸+۲\times۱۹+۳\times۲۰}{۹}=۱۸/۸۸$$

| | | | |
|---|---|---|---|
| د | ریا ـ ی | ادبیا ت | علو م |
| و | ۴ | ۲ | ۳ |
| ن | ۱۸ | ۱۹ | ۲۰ |

مثال۲: الف) نمره مسئولیت پذیری ۵ مدیر عبارت است از

$$x_i=۱۰,۱۵,۱۴,۸,۱۳$$

میانگین نمره مسئولیت پذیری این ۵ مدیر را محاسبه کنید.

$$\mu_x=\frac{۱۰+۱۵+۱۴+۸+۱۳}{۵}=\frac{۶۰}{۵}=۱۲$$

ب) نمرات مسئولیت پذیری ۲۰ مدیر به صورت زیر است.

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ (نمره) | ۵ | ۶ | ۱۰ | ۱۲ | ۱۵ |
| $w_i$ (تکرار) | ۳ | ۲ | ۵ | ۶ | ۴ |

$$\mu_w = \frac{(3\times 5)+(2\times 6)+(5\times 10)+(6\times 12)+(4\times 15)}{3+2+5+6+4} = 10/45$$

نکته: زمانی که در مشاهدات تکرار وجود دارد. تکرارها به عنوان وزن مشاهده محصوب می شوند.

➢ خواص میانگین حسابی:

۱- جمع جبری اختلاف مجموعه ای از اعداد از میانگینشان برابر صفر است. یعنی:

$$\sum_{i=1}^{N}(x_i - \mu_x) = 0$$

۲- هر گاه هر یک از مشاهدات با عدد ثابت a جمع شود، و در عدد ثابت b ضرب شود. میانگین اعداد حاصل شده برابر  میانگین مجموعه اعداد قبلی ضرب در b به اضافه a خواهد بود. یعنی:

$$y_i = b\times x_i + a \Rightarrow \mu_y = b\times \mu_x + a$$

۳- اگر x و z دو مجموعه از مشاهدات باشند. و مجموعه y از جمع دو به دو اعداد x و z حاصل شده باشد، میانگین مجموعه y برابر استبا جمع دو میانگین مشاهدات x و z. یعنی:

$$y_i = x_i + z_i \Rightarrow \mu_y = \mu_x + \mu_z$$

### ۱-۲ میانگین پیراسته

از این میانگین زمانی استفاده می شود که در توزیع مشاهدات، تعداد اندکی از آنها، با بقیه داده ها همخوانی و تجانس نداشته باشد.

طرز بدست آوردن میانگین پیراسته

۱- مرتب کردن صعودی داده ها

۲- حذف تمام مشاهدات کوچکتر از LN ٪ پایین و بزرگتر از ٪LN بالا

۳- محاسبه میانگین برای باقیمانده مشاهدات

### ۱-۳ میانگین وینزوری

در این میانگین بجای حذف کامل ٪LN ها ، مقادیر بالا و پایین آن بجای مقادیر حذف شده مورد استفاده قرار می گیرند و از تعداد داده ها کاسته نمی شود

**مثال۳: هزینه ماهانه یک خانواده تهرانی به این صورت به دست آمده است:**

**هزینه بر حسب صد هزار ریال: ۱۴/۵, ۱۵, ۳۰, ۲۵, ۱۸, ۱۷, ۱۶, ۹, ۲۰, ۱۵, ۸, ۱۰**

**میانگین پیراسته و وینزوری را در صورتی که ۲۵= LN٪ باشد، محاسبه کنید.**

$$۸, ۹, ۱۰, ۱۴/۵, ۱۵, ۱۵, ۱۶, ۱۷, ۱۸, ۲۰, ۲۵, ۳۰$$

$$N = ۱۲$$

$$\%LN \times N = ۰/۲۵ \times ۱۲ = ۳$$

$$\Rightarrow \mu'_x = \frac{۱۴/۵ + ۱۵ + ۱۵ + ۱۶ + ۱۷ + ۱۸}{۶} = ۱۵/۹۱$$

$$\Rightarrow \mu''_x = \frac{۱۴/۵ + ۱۴/۵ + ۱۴/۵ + ۱۴/۵ + ۱۵ + ۱۵ + ۱۶ + ۱۷ + ۱۸ + ۱۸ + ۱۸ + ۱۸}{۱۲} = \frac{۱۹۳}{۱۲} = ۱۶/۰۸$$

## ۱-۴ میانگین هندسی

از این میانگین برای محاسبه اندازه های نسبی همانند نسبت ها ، در صدها ، شاخص ها و نرخ های رشد استفاده می شود.

### ۱-۴-۱ میانگین هندسی ساده

**میانگین هندسی یک رشته عدد همانند $X_۱$ ، $X_۲$ ، . . . ، $X_N$ برابر است با ریشه N ام حاصلضرب آن اعداد**

$$\mu_G = (X_۱ \times X_۲ \times \ldots \times X_N)^{\frac{۱}{N}}$$

### ۱-۴-۲ میانگین هندسی موزون

اگر داده ها در میانگین هندسی دارای وزن باشند ، از این نوع میانگین استفاده می شود.

$$\mu_G = (X_۱^{W_۱} \times X_۲^{W_۲} \times \ldots \times X_K^{W_K})^{\frac{۱}{N}}$$

مثال۴: الف) شاخص قیمت ۵ نوع کالای مختلف در سال ۱۳۶۱، عبارت است از:

۱۲۰، ۱۳۰، ۱۳۵، ۱۳۷، ۱۴۰ متوسط شاخص قیمت را محاسبه کنید؟

$$\mu_G = (120 \times 130 \times 135 \times 137 \times 140)^{\frac{1}{5}} = 132/21$$

**ب)** شاخص قیمت ۱۵ نوع کالای مختلف در سال ۱۳۶۱، به صورت زیر است، متوسط شاخص قیمت را محاسبه کنید؟

| شاخص قیمت | ۱ | ۱۳ | ۱ | ۱ | ۱ |
|---|---|---|---|---|---|
| تعداد | ۲ | ۵ | ۳ | ۴ | ۱ |

$$\mu_G = (120^2 \times 130^5 \times 135^3 \times 137^4 \times 140^1)^{\frac{1}{15}} = 132/07$$

## ۱–۵ میانگین هارمونیک

از این نوع ، برای محاسبه میانگین مشاهداتی استفاده می شود که از مقیاس های ترکیبی همانند « کیلو در ساعت » یا « دور در ثانیه » برخوردار هستند.

### ۱–۵–۱ میانگین هارمونیک ساده

این میانگین برای چند اندازه یا مقدار برابر است با عکس میانگین حسابی معکوس آن اندازه ها

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

### ۱-۵-۲ میانگین هارمونیک موزون

در صورت تکرار داده ها ( وزن داشتن آنها ) از فرمول زیر استفاده می شود :

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

مثال۵: الف) وسیله‌ای مسافت تهران – کرج را با سرعت ۸۰ کیلومتر بر ساعت رفته و با ۱۲۰ کیلومتر بر ساعت برگشته است متوسط سرعت رفت و برگشت فرد چقدر است؟

$$\mu_H=\frac{1+1}{\frac{1}{80}+\frac{1}{120}}=\frac{2}{\frac{3+2}{240}}=96\,km/h$$

ب) وسیله‌ای مسافت تهران – قم را این‌گونه رفته و برگشته :

$\frac{1}{3}$مسیر رفت را با سرعت $80 km/h$ و باقیمانده را با سرعت $100 km/h$ رفته و کل مسیر برگشت را با سرعت $110 km/h$طی کرده است مطلوب است متوسط سرعت این وسیله در رفت و برگشت.

$$\mu_{W.H}=\frac{\frac{1}{3}+\frac{2}{3}+1}{\frac{\frac{1}{3}}{80}+\frac{\frac{2}{3}}{100}+\frac{1}{110}}=100/3\,km/h$$

۲- میانه

مراحل محاسبه میانه

۱- مرتب کردن اعداد.

۲- تعیین نقطه ای که نیمی از داده ها بالاتر و نیمی دیگر پایین تر از آن هستند.

نکته: در صورتی که تعداد داده ها فرد باشد میانه عددی است که در وسط قرار دارد و اما در صورتی که تعداد داده ها زوج باشد، میانه عبارتست از معدل دو نمره ای که در وسط واقع می شوند.

مثال۶: میانه مشاهدات زیر را محاسبه کنید.

**۶ - ۴- ۹ -۱۳ - ۱۷**

**۴ - ۶ - ۷ -۹**

$$\Rightarrow \frac{۶+۷}{۲}=۶/۵$$

**۳- مد ( نما )**

به مقداری گفته می شود که در میان سایر مقادیر توزیع ، بیشترین تکرار را داشته باشد ، مد را با Mo نشان می دهند.

مثال۷: مد مشاهدات ۲, ۲, ۵, ۱۵, ۱۰, ۸, ۵, ۲ را بیابید.

$Mo = ۲$

**نکته**: چنانچه کلیه مشاهدات به یک اندازه تکرار شده باشند، جامعه فاقد مد یا مد جامعه تهی خواهد بود.

**۴- چارک**

اگر جامعه آماری به چهار قسمت مساوی تقسیم شود ، به هر یک از قسمت ها یک چارک گفته می شود و آنها را با Q نشان می دهند.

## انواع چارک ها

$Q_۱$ : مقداری که ۲۵٪ مشاهدات ، کمتر از آن و ۷۵٪ بیشتر از آن است.

$Q_۲$ : مقداری که ۵۰٪ مشاهدات ، کمتر از آن و ۵۰٪ بیشتر از آن است.

$Q_۳$ : مقداری که ۷۵٪ مشاهدات ، کمتر از آن و ۲۵٪ بیشتر از آن است.

## نحوه محاسبه چارکها

۱- مرتب نمودن صعودی داده ها

۲- کد گذاری کردن آنها از ۱ تا N

۳- پیدا نمودن محل چارک مورد نظر $CQ_a = \frac{aN}{۴} + \frac{۱}{۲}$

که در فرمول فوق ۱و۲و۳ =a و تعداد مشاهدات =N

۴- تعیین نمودن مقدار چارک مورد نظر به کمک محل چارک

مثال ۸: چارک اول، دوم و سوم مشاهدات زیر را به دست آورید:

۸۰, ۹۰, ۵۰, ۶۰, ۱۰۰, ۷۰, ۷۵, ۸۰, ۷۰, ۸۰, ۶۰

حل:

۵۰ ,۶۰ ,۶۰ , ۷۰ ,۷۰ ,۷۵ ,۸۰ ,۸۰ ,۸۰ ,۹۰ ,۱۰۰

$$C_{Q_1} = \frac{1 \times 11}{4} + \frac{1}{2} = 3/25 \Rightarrow Q_1 = 60 + 0/25 \times (70 - 60) = 62/5$$

$$C_{Q_2} = \frac{2 \times 11}{4} + \frac{1}{2} = 6 \Rightarrow Q_2 = 75$$

$$C_{Q_3} = \frac{3 \times 11}{4} + \frac{1}{2} = 8/75 \Rightarrow Q_3 = 80 + 0/75 \times (80 - 80) = 80$$

### ❖ پارامترهای پراکندگی

شاخص هایی هستند که متوسط میزان دوری و نزدیکی داده های توزیع را نسبت به میانگین شان نشان می دهند.

**انواع شاخص های پراکندگی**

۱- دامنه تغییرات
۲- دامنه میان چارکی
۳- انحراف متوسط از میانگین
۴- واریانس
۵- انحراف معیار
۶- نیمه واریانس
۷- ضریب پراکندگی

**۱- دامنه تغییرات($R$)**

ساده ترین شاخص پراکندگی است و با کم کردن کوچکترین مشاهده از بزرگترین آنها در یک سری توزیع بدست می آید.

$R = MAX_x - MIN_x$

مثال۹: کارگاهی دارای ۱۵ کارگر است که حقوق ماهانه آنها(گرد شده به صد هزار تومان ) به قرار زیر است:

۱۲, ۱۱, ۱۹, ۱۶, ۲۲, ۸, ۱۳, ۱۶, ۱۷, ۱۵, ۲۰, ۱۴, ۱۷, ۱۴, ۱۵

**دامنه تغییرات مشاهدات را به دست آورید.**

$R = ۲۲ - ۸ = ۱۴$

**۲- دامنه میان چارکی $(IQR)$**

این شاخص ، پراکندگی داده ها را در فاصله چارک اول و چارک سوم نشان می دهد و کاری به مقادیر کوچکتر از $Q_۱$ و بزرگتر $Q_۳$ ندارد.

برای محاسبه این شاخص ، کافیست که مقادیر $Q_۱$ و $Q_۳$ را بدست آورده و از هم کم کنیم.

$$IQR = Q_۳ - Q_۱.$$

**مثال ۱۰: در مثال ۹ دامنه میان چارکی را محاسبه کنید.**

**۲۲, ۲۰, ۱۹, ۱۷, ۱۷, ۱۶, ۱۶, ۱۵, ۱۵, ۱۴, ۱۴, ۱۳, ۱۲, ۱۱, ۸**

$$C_{Q_۱} = \frac{۱ \times ۱۵}{۴} + \frac{۱}{۲} = ۴/۲۵ \Rightarrow Q_۱ = ۱۳ + ۰/۲۵ \times (۱۴ - ۱۳) = ۱۳/۲۵$$

$$C_{Q_۳} = \frac{۳ \times ۱۵}{۴} + \frac{۱}{۲} = ۱۱/۷۵ \Rightarrow Q_۳ = ۱۷ + ۰/۷۵ \times (۱۷ - ۱۷) = ۱۷$$

$$Q_۳ - Q_۱ = ۱۷ - ۱۳/۲۵ = ۳/۷۵$$

**۳- نیمه میان چارکی (انحراف چارکی)**

برای بدست آوردن این شاخص ، که به انحراف چارکی نیز معروف است ، کافیست ، مقدار دامنه میان چارکی را بر عدد ۲ تقسیم نماییم.

$$SIQR = \frac{Q_۳ - Q_۱}{۲}$$

**مثال ۱۱: در مثال ۹ نیمه میان چارکی را محاسبه کنید.**

$$SIQR = \frac{Q_۳ - Q_۱}{۲} = \frac{۱۷ - ۱۳/۲۵}{۲} = ۱/۸۷۵$$

**شاخص های مناسب برای توزیع های نا متقارن**

۱- استفاده از میانه بعنوان بهترین شاخص مرکزی

۲- استفاده از انحراف چارکی بعنوان بهترین شاخص پراکندگی

**۴- انحراف متوسط از میانگین**

این شاخـص از تقسیم مجموع قدر مطلـق انحـرافات تک تک مشاهدات از میانگین شان بر تعداد مشاهدات بدست می آید.

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

**مثال ۱۲: در مثال ۹ انحراف متوسط از میانگین را محاسبه کنید.**

$$A.D_{\mu}=\frac{\sum_{i=1}^{n}|X_i-\mu_x|}{N}=\frac{|15-15/27|+|14-15/27|+.....+|12-15/27|}{15}=\frac{40/27}{15}=2/68$$

**محاسن و معایب $A\cdot D_{\mu}$**

محاسن : در نظر گرفتن تغییرات کل داده ها

معایب : ۱- نشان ندادن تأثیر انحرافات بزرگ

۲- بی بهره بودن از بعضی از خواص مطلوب میانگین حسابی

**۵- واریانس**

در این شاخص پراکندگی ، بر خلاف شاخص انحراف متوسط از میانگین بجای قدر مطلق از مجذور (توان ۲) انحرافات استفاده می شود.

$$\sigma_x^2=\frac{\sum_{i=1}^{N}(X_i-\mu_x)^2}{N}$$

**مثال ۱۳: در مثال ۹ واریانس مشاهدات را محاسبه کنید.**

$$\sigma_x^2=\frac{\sum_{i=1}^{N}(X_i-\mu_x)^2}{N}=\frac{(15-15/27)^2+(14-15/27)^2+.....+(12-15/27)^2}{15}=\frac{178/93}{15}=11/93$$

**۶- انحراف معیار**

این شاخص به منظور برطرف کردن عیوب شاخص های قبلی است یعنی همان نشان ندادن تأثیر انحراف بزرگ توسط $A\cdot D_{\mu}$ و افزایش دادن تأثیر این انحراف توسط $\sigma_X^2$.

$$\sigma_X=\sqrt{\sigma_X^2}$$

**و یا**

$$\sigma_X=\sqrt{\frac{\Sigma(X_i-\mu_X)^2}{N}}$$

**مثال ۱۳: در مثال ۹ انحراف از معیار را محاسبه کنید.**

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{11/93} = 3/45$$

**خواص واریانس**

۱- اگر تمام مشاهدات با عدد ثابت $a$ جمع شوند ، واریانس جدید تغییر نمی کند.

$$Y_i = X_i + a \quad \Rightarrow \sigma_Y^2 = \sigma_X^2$$

۲- اگر تمام مشاهدات ، به عدد ثابت $b$ ضرب شوند ، واریانس جدید $b^2$ برابر افزایش می یابد

$$Y_i = bX_i \quad \Rightarrow \sigma_Y^2 = b^2\sigma_X^2$$

**۷- نیمه واریانس**

یعنی متوسط مجذور مقادیر نامطلوب

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

که $\mathcal{N}$ ، تعداد کل مشاهدات، $\mathcal{K}$ ، تعداد موارد نامطلوب و $\mu_x$، میانگین کل مشاهدات می باشد.

- **مقادیر نامطلوب** : در داده های مربوط به سود و در آمد مقادیر کوچک تر از میانگین و در داده های مربوط به زیان و هزینه مقادیر بزرگتر از میانگین ، نامطلوب قلمداد می شوند.

**مثال ۱۴: در مثال ۹ نیمه واریانس را محاسبه کنید.**

**حل: با توجه به اینکه مشاهدات مربوط به درآمد می باشد مقادیر کمتر از میانگین به عنوان مقادیر نامطلوب محسوب شده و نیمه واریانس برابر است با**

$$S.V = \frac{\sum_{i=1}^{k}(X_i - \mu_x)^2}{k} = \frac{(8-15/27)^2 + (11-15/27)^2 + (12-15/27)^2 + (13-15/27)^2 + (14-15/27)^2 + (15-15/27)^2 + (15-15/27)^2}{8} = \frac{90/30}{8} = 11/29$$

**۸- ضریب پراکندگی**

ضریب پراکندگی یکی از معیارهای پراکندگی نسبی است که با فرمول زیر بیان می شود

$$C.V = \frac{\sigma_X}{\mu_X}$$

که $\sigma_x$، انحراف معیار و $\mu_x$، میانگین مشاهدات است.

**مثال ۱۵: در مثال ۹ ضریب پراکندگی را محاسبه کنید.**

$$C.V = \frac{\sigma_X}{\mu_X} = \frac{۳/۴۵}{۱۵/۲۷} = ۰/۲۲۶$$

**کاربردهای ضریب پراکندگی**

برای مقایسه دو جامعه در مواردی که :

۱- مقیاس ها یکسان نیستند

۲- مقیاس یکسان ولی تفاوت زیادی در بزرگی مشاهدات وجود دارد

۳- واریانسهای جوامع یکسان ولی میانگین هایشان متفاوت است.

## فصل سوّم: طبقه بندی و توصیف هندسی مشاهدات جامعه

هدف این فصل آشنایی دانشجویان با طبقه بندی و سازماندهی مشاهدات و استفاده از نمودارهای مختلف برای توصیف داده هاست.

### توزیع فراوانی

یعنی جدول مرتب و خلاصه شده از داده ها و مشاهدات که تکرار وقوع هر داده ها در آن مشخص شده است. معمولا یک جدول توزیع فراوانی شامل ستون های مشخص کننده حد پایین و بالای طبقات( طبقه بندی داده ها) ، مرکز دسته ها، فراوانی مطلق، فراوانی مطلق نسبی، فراوانی تجمعی، فراوانی نسبی تجمعی است.

۱) مراحل طبقه بندی داده ها

۱- مرتب کردن داده ها و محاسبه دامنه تغییرات ( R )
۲- مشخص کردن تعداد طبقات ( K )
۳- محاسبه نمودن فاصله طبقات ( I )
۴- سازماندهی طبقات

۱- مرتب کردن داده ها و محاسبه دامنه تغییرات یعنی $R = MAX_{X_i} - MIN_{X_i}$

۲- فرمول های محاسبه تعداد طبقات

- فرمول تجربی استورجس $K = ۱ + ۳/۳۲ Log N$
- روش تقریبی $K = \sqrt{N}$ ( N تعداد مشاهدات می باشد )

۳- تعیین فاصله طبقات

فاصله طبقات از تقسیم مقدار R (دامنه تغییرات) بر مقدار محـاسبه شـده برای تعـداد طـبقـات (K) به شکل زیر بدست می آید

$$I = \frac{R}{K}$$

۴- سازماندهی داده ها

پس از مشخص شدن K و I سازماندهی یعنی تعیین نوع جدول و شیوه طبقه بندی داده ها شروع می شود که این بستگی به نوع داده های جمع آوری شده دارد.

انواع طبقه بندی داده ها

۱- طبقه بندی مشاهدات پیوسته: در این نوع مشاهدات طبقه بندی به دو صورت پیوسته و گسسته انجام می شود.

۱. طبقه بندی پیوسته: در این نوع طبقه بندی طول ، عرض و فاصله طبقات مساوی هستند. این نوع طبقه بندی برای داده های اعشاری استفاده می شود.
۲. طبقه بندی گسسته : در این نوع طبقه بندی طول و عرض طبقات با هم برابر نیستند. این نوع طبقه بندی برای داده های غیر اعشاری استفاده می شود.

- مهم ترین تقریب ها در طبقه بندی گسسته
  ۱. تقریب ۰/۱
  ۲. تقریب ۰/۵
  ۳. تقریب ۱ ( واحد )

تقریب ، اختلاف طول و عرض طبقات یا فاصله بین حد بالای یک طبقه با حد پایین طبقه بعدی است.

۲- طبقه بندی مشاهدات ناپیوسته

در مشاهدات ناپیوسته تعریف بصورت فاصله طبقات بی معناست لذا برای تشکیل توزیع فراوانی آنها کافیست یک ستون برای مشاهدات و ستون دیگری برای فراوانی آنها تنظیم شود .

۲) مرکز دسته ( $x_i$ ):

۲/ (ابتدای دسته + انتهای دسته)

۳) فراوانی مطلق ( $F_i$ ): تعداد مشاهدات در هر دسته

۴) فراوانی نسبی ( $f_i$ ): از تقسیم فراوانی مطلق هر دسته بر تعداد کل مشاهدات حاصل می شود.

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

- کاربرد فراوانی نسبی: به کمک این فراوانی می توان در صد تراکم داده ها را در هر طبقه مشخص نمود بعبارتی از $f_i$ جهت یافتن محل تمرکز داده ها استفاده می شود.

۵) فراوانی تجمعی ( $FC_i$ )

فراوانی تجمعی هر طبقه ، عبارتست از مجموع فراوانی های مطلق از اولین طبقه تا طبقه مورد نظر که آن را با $FC_i$ نشان می دهند

$$Fc_i = \sum_{j=1} F_j$$

۶) فراوانی نسبی تجمعی ( $fc_i$ )

این فراوانی از تقسیم فراوانی تجمعی هر طبقه بر تعداد مشاهدات بدست می آید.

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

➢ مفهوم فراوانی نسبی تجمعی: این فراوانی بیانگر در صد داده ها و مشاهدات واقع شده بین حد پایین اولین طبقه تا حد بالای طبقه مورد نظر است.

مثال ۱: ارقام سود روزانه ۲۵ دکه روزنامه فروشی در زیر آمده است. جدول توزیع فراوانی مشاهدات را به دست آورید.

۵۵/۳۱ ، ۸۱/۴۷ ، ۶۴/۹۰ ، ۷۰/۸۸ ، ۸۶/۰۲ ، ۷۷/۲۵ ، ۷۶/۷۳ ، ۸۴/۲۱ ، ۵۶/۰۲ ، ۸۴/۹۲ ، ۹۰/۲۳ ، ۷۸/۰۱ ، ۸۸/۰۵ ، ۷۳/۳۷ ، ۸۷/۰۹ ، ۵۷/۴۱ ، ۸۵/۴۳ ، ۷۴/۷۶ ، ۸۶/۵۱ ، ۸۶/۳۷ ، ۷۶/۱۵ ، ۸۸/۶۴ ، ۸۴/۷۱ ، ۶۶/۰۵ ، ۸۳/۹۱

حل: ابتدا مشاهدات را طبقه بندی می کنیم. برای این منظور مراحل زیر را انجام می دهیم.

۱- محاسبه دامنه مشاهدات: $R = MAX_i - MIN_i = ۹۰/۲۳ - ۵۵/۳۱ = ۳۴/۹۲$

۲- مشخص کردن تعداد طبقات: $\sqrt{N} = \sqrt{۲۵} = ۵$

۳- تعیین طول طبقات: $I = \frac{R}{k} = \frac{۳۴/۹۲}{۵} = ۶/۹۸۴ \approx ۷$

۴- سازمان دهی مشاهدات: لازم به ذکر است با توجه اعشاری بودن مشاهدات از طبقه بندی پیوسته به منظور طبقه بندی مشاهدات استفاده می شود.

با توجه به سادگی محاسبات، نتایج نهایی در جدول ارائه می شود.

| فرونی نسبی تجمعی ( $fc_i$ ) | فراوانی تجمعی ( $FC_i$ ) | فراوانی نسبی ( $f_i$ ) | فراوانی مطلق ( $F_i$ ) | مرکز دسته ( $x_i$ ) | حدود طبقات |
|---|---|---|---|---|---|
| ۰/۱۲ | ۳ | ۰/۱۲ | ۳ | ۵۸/۸ | [۵۵/۳ -۶۲/۳) |
| ۰/۲ | ۵ | ۰/۰۸ | ۲ | ۶۵/۸ | [۶۲/۳ -۶۹/۳) |
| ۰/۴ | ۱۰ | ۰/۲ | ۵ | ۷۲/۸ | [۶۹/۳ -۷۶/۳) |
| ۰/۵۲ | ۱۳ | ۰/۱۲ | ۳ | ۷۹/۸ | [۷۶/۳ -۸۳/۳) |
| ۱ | ۲۵ | ۰/۴۸ | ۱۲ | ۸۶/۸ | [۸۳/۳ -۹۰/۳] |
| | | ۱ | ۲۵ | | کل |

مثال۲: سن ۳۰ مدیر موسسه در زیر ارائه شده است جدول توزیع فراوانی آنها را به دست آورید.

۳۵ - ۴۶ – ۶۳ – ۶۹– ۵۴– ۵۰ – ۶۲ – ۶۸– ۳۸– ۴۰ – ۵۵– ۴۳– ۴۲– ۵۹– ۴۵– ۴۴– ۵۷– ۴۷– ۴۸– ۴۶–
۴۳– ۶۴– ۴۹– ۳۶– ۵۹– ۶۰ – ۴۲– ۶۰ –۴۲– ۳۸

حل: ابتدا مشاهدات را طبقه بندی می کنیم. برای این منظور مراحل زیر را انجام می دهیم.

۱– محاسبه دامنه مشاهدات: $R = MAX_i - MIN_i = ۶۹ - ۳۶ = ۳۴$

۲– مشخص کردن تعداد طبقات: $k = \sqrt{N} = \sqrt{۳۰} = ۵/۸۳ \approx ۶$

۳– تعیین طول طبقات: $I = \frac{R}{k} = \frac{۳۴}{۶} = ۵/۶۷ \approx ۶$

۴– سازمان دهی مشاهدات: لازم به ذکر است با توجه به غیراعشاری بودن مشاهدات از طبقه بندی گسسته به منظور طبقه بندی مشاهدات استفاده می شود.

با توجه به سادگی محاسبات، نتایج نهایی در جدول ارائه می شود.

| حدود طبقات | مرکز دسته ($x_i$) | فراوانی مطلق ($F_i$) | فراوانی نسبی ($f_i$) | فراوانی تجمعی ($FC_i$) | فروانی نسبی تجمعی ($fc_i$) |
|---|---|---|---|---|---|
| [۳۵ –۴۱ ] | ۳۸ | ۵ | ۰/۱۷ | ۵ | ۰/۱۷ |
| [۴۲ –۴۷ ] | ۴۴ | ۱۰ | ۰/۳۳ | ۱۵ | ۰/۵ |
| [۴۸ –۵۳ ] | ۵۰ | ۳ | ۰/۱ | ۱۸ | ۰/۶ |
| [۵۴ –۵۹ ] | ۵۶ | ۵ | ۰/۱۷ | ۲۳ | ۰/۷۷ |
| [۶۰ –۶۵ ] | ۶۲ | ۵ | ۰/۱۷ | ۲۸ | ۰/۹۳ |
| [۶۶ –۷۱ ] | ۶۸ | ۲ | ۰/۰۶۷ | ۳۰ | ۱ |
| کل | | ۳۰ | ۱ | | |

مثال۳: در ۵۰ دقیقه متوالی، تعداد تلفن هایی که به یک مرکز تلفن شده است در هریک از دقایق به شرح زیر ثبت شده است . جدول توزیع فراوانی را به دست آورید.

۱– ۰ – ۱– ۱– ۰– ۰– ۲– ۲– ۰– ۱– ۱– ۰– ۱– ۴– ۰– ۳– ۰– ۱– ۰– ۲– ۰– ۰– ۱– ۱– ۰– ۰– ۱– ۲– ۱–
۰– ۰– ۱– ۳– ۱– ۰– ۱– ۱– ۴– ۰– ۱– ۲– ۱– ۱– ۰– ۲– ۰– ۰– ۱– ۲– ۱

حل: با توجه به ناپیوسته بودن مشاهدات از طبقه بندی مربوط به داده های ناپیوسته به منظور طبقه بندی مشاهدات استفاده می شود.

| مشاهدات | فراوانی مطلق ($F_i$) | فراوانی نسبی ($f_i$) | فراوانی تجمعی ($FC_i$) | فروانی نسبی تجمعی ($fc_i$) |
|---|---|---|---|---|
| ۰ | ۱۹ | ۰/۳۸ | ۱۹ | ۰/۳۸ |
| ۱ | ۲۰ | ۰/۴ | ۳۹ | ۰/۷۸ |
| ۲ | ۷ | ۰/۱۴ | ۴۶ | ۰/۹۲ |
| ۳ | ۲ | ۰/۰۴ | ۴۸ | ۰/۹۶ |
| ۴ | ۲ | ۰/۰۴ | ۵۰ | ۱ |
| کل | ۵۰ | ۱ | | |

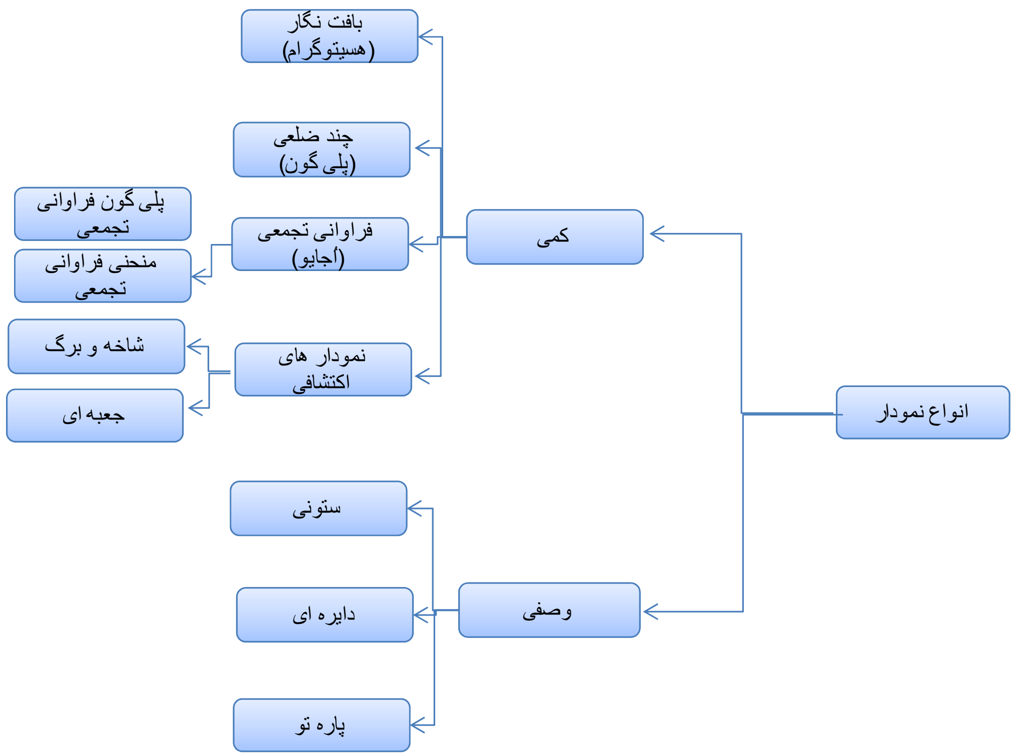
## ❖ نمایش هندسی مشاهدات

استفاده از نمودارها یا به عبارتی نمایش هندسی مشاهدات در گزارش نویسی باعث می شود که خوانندگان با صرف کمترین زمان و با ساده ترین بیان ، گزارش را بفهمند و تصویری روشن از توزیع داشته باشند

انواع نمودارها

۱– نمودارهای کمی : مخصوص داده هایی با مقیاس فاصله ای و نسبی

۲– نمودارهای وصفی : مخصوص داده هایی با مقیاس اسمی و یا رتبه ای




همانطور که در شکل فوق دیده می شود مهم ترین نمودارهای کمی عبارتند از:

۱- بافت نگار (هیستوگرام)    ۴- تحلیل اکتشافی داده ها

۲- چند ضلعی (پلی گون)    ۴-۱ نمودار شاخه و برگ

۳- فراوانی تجمعی (اُجایو)    ۴-۲ نمودار جعبه ای

۳-۱ پلی گون فراوانی تجمعی

۳-۲ منحنی فراوانی تجمعی

که در زیر هریک از ننمودار های فوق تشریح شده است.

## ۱) بافت نگار

بافت نگار نموداریست در دستگاه مختصات که محور افقی آن با حدود واقعی طبقات و محور عمودی آن با فراوانی مطلق یا نسبی درجه بندی می شود.

حدود واقعی: اگر دسته بندی مشاهدات از نوع پیوسته باشد حدود واقعی با حدود طبقات برابر است و در صورتی که طبقه بندی از نوع گسسته باشد برای پیدا کردن حدود واقعی، ۰/۵ واحد از کران پایین طبقات کم می کنیم و ۰/۵ واحد به کران بالای طبقات اضافه می کنیم تا دسته بندی مشاهدات پیوسته شود. آنگاه حدود حاصل، حدود واقعی طبقات است.

در صورتی که از فراوانی مطلق به منظور مندرج کردن محور عمودی استفاده کنیم. پس از مدرج کردن محور افقی  با استفاده از حدود واقعی (کرانه های هر طبقه) مستطیلی عمودی رسم می شود که ارتفاع آن برابر فراوانی مطلق هر طبقه می باشد.

برای رسم بافت نگار فراوانی نسبی ، رده ها (کران پایین و بالای طبقات) را روی محور افقی نمودار مشخص می کنیم . آنگاه روی هر رده ، مستطیلی عمودی رسم می کنیم که مساحت آن مساوی با فراوانی نسبی آن رده باشد . در نتیجه ارتفاع مستطیل برابر است با

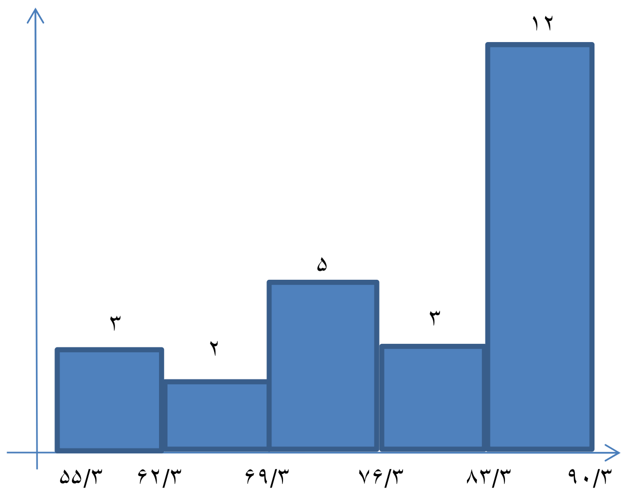
طول رده /  فراوانی نسبی رده = ارتفاع مستطیل

به این ترتیب ، مساحت هر یک از مستطیل های بافت نگار نشان دهنده نسبت مشاهدات موجود در رده ای است که مستطیل بر آن قرار دارد . بنابراین مجموع مساحت تمام مستطیل ها در یک بافت نگار برابر با یک است.

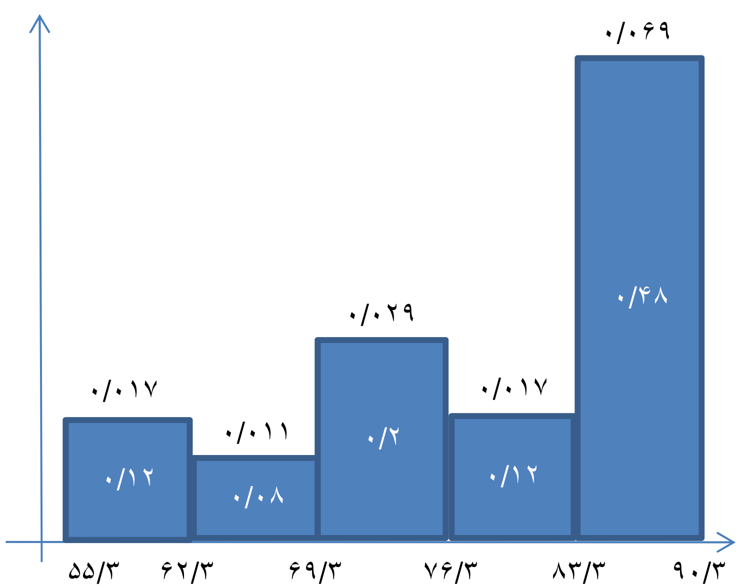
برای نمایش فراوانی های نسبی ، استفاده از مساحت مستطیل ها به جای ارتفاع آنها فایده آشکاری دارد . به نظر می رسد که در موقع مقایسه کردن دو قسمت یک بافت نگار ، یا دو بافت نگار مختلف ، چشم انسان به طور غریزی مساحت ها را با هم مقایسه می کند.

مثال ۴: برای مشاهدات مثال ۱ نمودار بافت نگار را رسم کنید.

حل: رسم نمودار بر اساس فراوانی مطلق




برای رسم نمودار فراوانی نسبی، ابتدا فراوانی نسبی هر طبقه را بر طول طبقه که در این مثال ۷ است تقسیم می کنیم. بعد از مشخص کردن کران طبقات روی محور افقی روی کران طبقات مستطیلی به ارتفاع مقادیر حاصل از تقسیم رسم می شود.

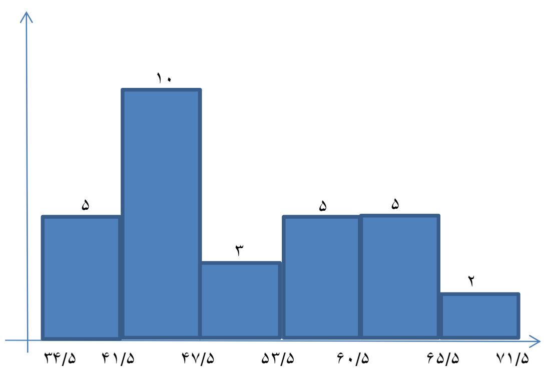


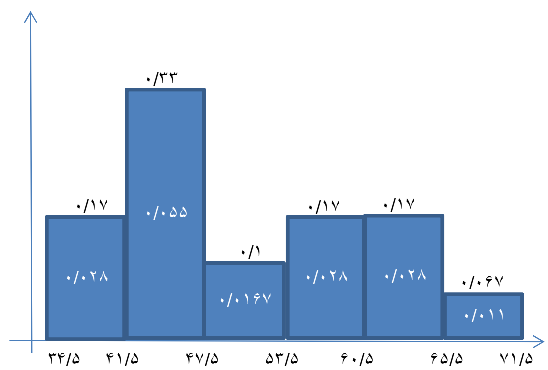

لازم به ذکر از مقادیر داخل هر مستطیل، بیانگر مساحت مستطیل هاست.

مثال ۵: برای مشاهدات مثال ۲ نمودار بافت نگار را رسم کنید.

حل:

| حدود طبقات | حدود واقعی طبقات | فراوانی مطلق $(F_i)$ | فراوانی نسبی $(f_i)$ | ارتفاع مستطیل |
|---|---|---|---|---|
| [۳۵ – ۴۱] | ۳۴/۵ – ۴۱/۵ | ۵ | ۰/۱۷ | ۰/۰۲۸ |
| [۴۲ – ۴۷] | ۴۱/۵ – ۴۷/۵ | ۱۰ | ۰/۳۳ | ۰/۰۵۵ |
| [۴۸ – ۵۳] | ۴۷/۵ – ۵۳/۵ | ۳ | ۰/۱ | ۰/۰۱۶۷ |
| [۵۴ – ۵۹] | ۵۳/۵ – ۵۹/۵ | ۵ | ۰/۱۷ | ۰/۰۲۸ |
| [۶۰ – ۶۵] | ۵۹/۵ – ۶۵/۵ | ۵ | ۰/۱۷ | ۰/۰۲۸ |
| [۶۶ – ۷۱] | ۶۵/۵ – ۷۱/۵ | ۲ | ۰/۰۶۷ | ۰/۰۱۱ |
| کل | | ۳۰ | ۱ | |



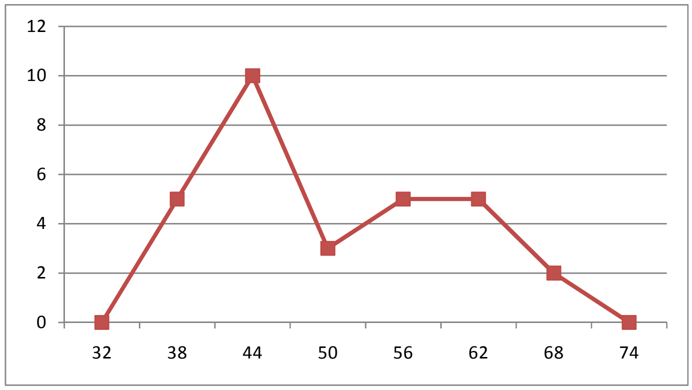




**۲) نمودار چند ضلعی**

نموداریست که متناظر با هر نماینده طبقه ( مرکز طبقه) در محور افقی و فراوانی آن در محور عمودی، یک نقطه در صفحه مختصات ایجادمی شود. به نقاط مزبور دو نقطه فرضی دیگر اضافه می کنیم اولی مرکز طبقه ماقبل

اولین طبقه و دیگری نماینده طبقه مابعد آخرین طبقه، از اتصال متوالی نقاط به یکدیگر نمودار مورد نظر حاصل می شود.

مثال ۶: برای مشاهدات مثال ۲ نمودار چندضلعی را رسم کنید.

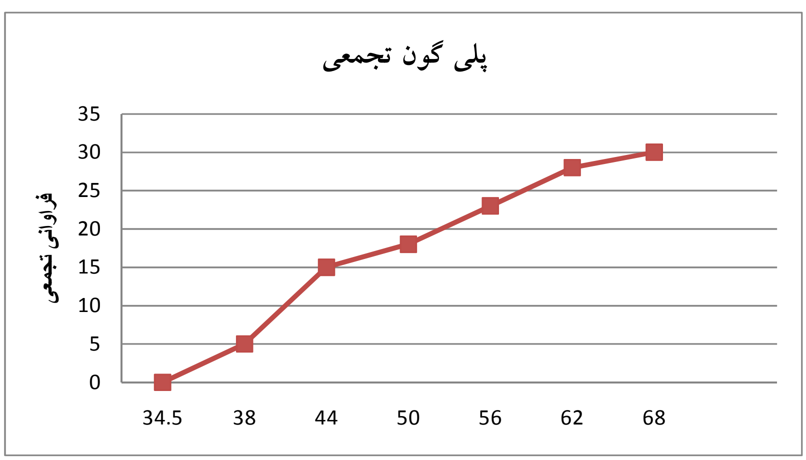



۳) نمودار فراوانی تجمعی

- 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.

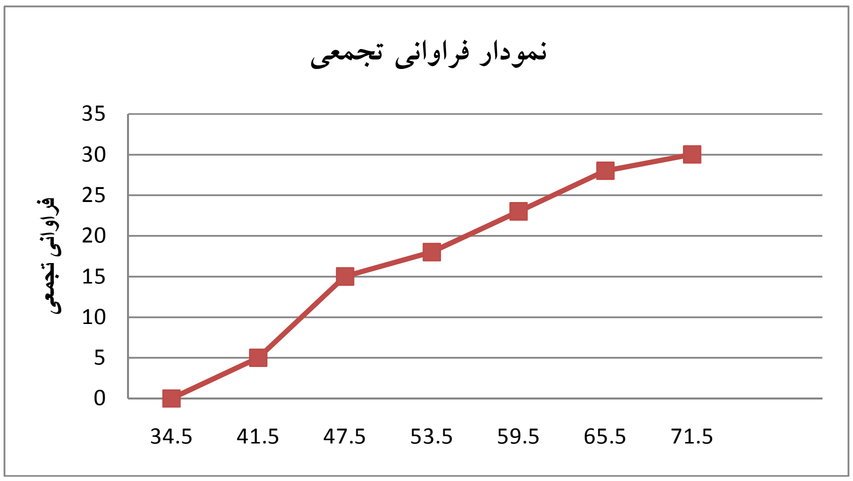
مثال ۷: برای مشاهدات مثال ۲ پلی گون فراوانی تجمعی را رسم کنید.



- منحنی فراوانی تجمعی

تنها فرق این نمودار با نمودار پلی گون فراوانی تجمعی در این است که در این نمودار بجای نماینده طبقات از حد بالای کرانه ها استفاده می شود.

مثال ۸: برای مشاهدات مثال ۲ نمودار فراوانی تجمعی را رسم کنید.




➢ کاربردهای نمودار فراوانی تجمعی

۱- برای محاسبه چندکها (چارکها ، دهکها ، صدکها)

۲- برای مقایسه پدیده ها (مثل میزان رشد تورم در کشورها)

## ۴) تحلیل اکتشافی داده ها

در بر گیرنده نمودار های جدیدی است که در مراحل اولیه تحلیل داده ها مفید هستند و اطلاعات بیشتری را در مورد تک تک داده ها به معرض نمایش می گذارند.

### ۴-۱ نمودار شاخه و برگ

برای تهیه این نمودار ، ارقام مشاهدات به دو بخش شاخه و برگ تقسیم می شوند، شاخه شامل یک یا چند رقم اولیه و برگ شامل ارقام باقی مانده است.

مثال ۹: برای مشاهدات مثال ۲ نمودار شاخه و برگ را رسم کنید.

| | |
|---|---|
| ۳ | ۵ ۶ ۸ ۸ |
| ۴ | ۰ ۲ ۲ ۳ ۳ ۴ ۵ ۶ ۶ ۷ ۸ ۹ |
| ۵ | ۰ ۴ ۵ ۷ ۹ ۹ |
| ۶ | ۰ ۰ ۲ ۳ ۴ ۸ ۹ |

محاسن نمودار شاخه و برگ

در این نمودار بر خلاف بافت نگار ، اعداد اصلی از بین نمی روند و محاسبه چندکها هم با استفاده از آن براحتی امکان پذیر است.

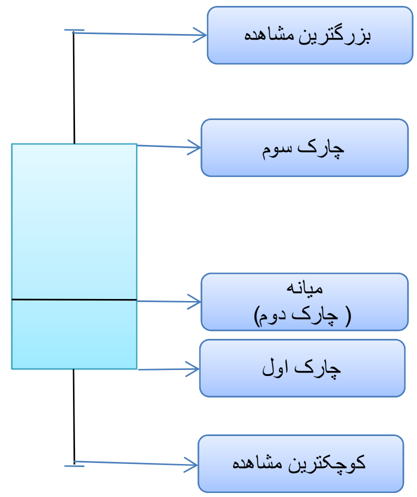
### ۴-۲ نمودار جعبه ای

این نمودار نشان دهنده چارکها و حداقل و حداکثر مشاهدات است و برای مقایسه دو یا چند جامعه آماری مورد استفاده قرار می گیرد.

مراحل تهیه نمودار جعبه ای

الف – پیدا کردن حداقل و حداکثر داده ها

ب – پیدا کردن چارکهای اول ، دوم و سوم




مثال ۹: برای مشاهدات زیر  نمودار جعبه ای را رسم کنید

۸۰, ۹۰, ۵۰, ۶۰, ۱۰۰, ۷۰, ۷۵, ۸۰, ۷۰, ۸۰, ۶۰

حل:

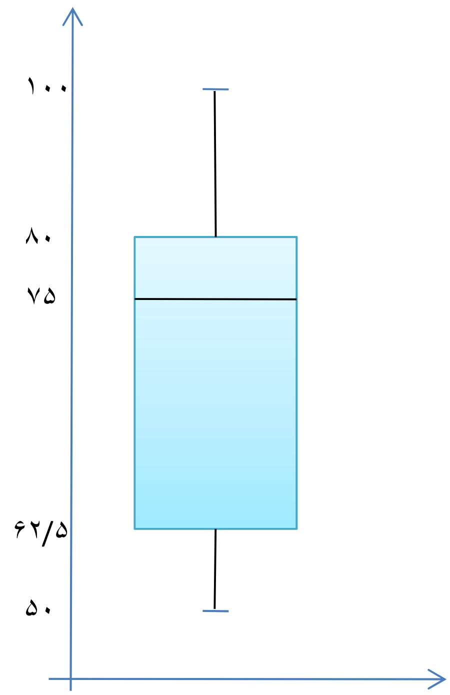
۱۰۰, ۹۰, ۸۰, ۸۰, ۸۰, ۷۵, ۷۰, ۷۰ , ۶۰, ۶۰, ۵۰

$$MIN = ۵۰ \quad , \qquad MAX = ۱۰۰$$

$$C_{Q_۱} = \frac{۱ \times ۱۱}{٤} + \frac{۱}{۲} = ۳/۲۵ \quad \Rightarrow Q_۱ = ٦۰ + ۰/۲۵ \times (۷۰ - ٦۰) = ٦۲/۵$$

$$C_{Q_۲} = \frac{۲ \times ۱۱}{٤} + \frac{۱}{۲} = ٦ \quad \Rightarrow Q_۲ = ۷۵$$

$$C_{Q_۳} = \frac{۳ \times ۱۱}{٤} + \frac{۱}{۲} = ۸/۷۵ \quad \Rightarrow Q_۳ = ۸۰ + ۰/۷۵ \times (۸۰ - ۸۰) = ۸۰$$




❖ **نمودارهای وصفی**

این دسته از نمودارها برای نمایش هندسی داده های کیفی بکار می روند، در این نمودارها هر یک از مقادیر بعنوان یک طبقه در نظر گرفته می شوند.

مهم ترین نمودارهای وصفی

۱- نمودار ستونی

۲- نمودار دایره ای

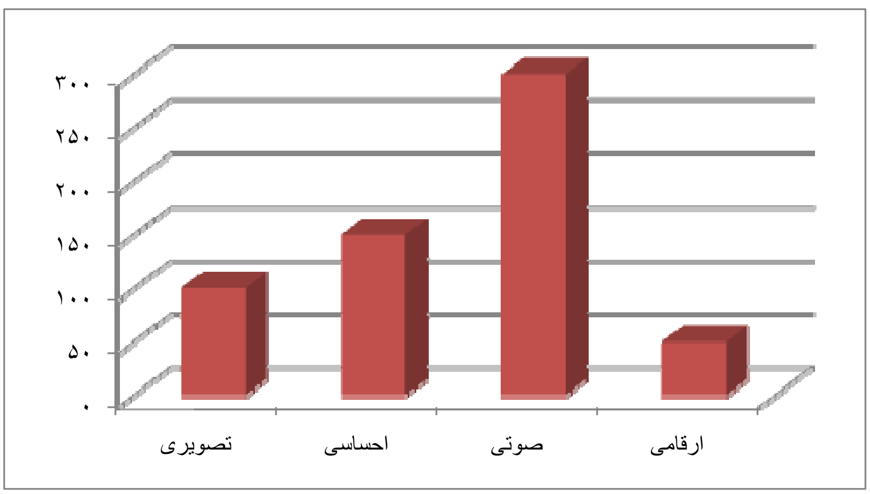
۳- نمودار پاره تو

### ۱) نمودار ستونی (میله ای)

این نمودار در یک دستگاه مختصات که محور افقی نشان دهنده کیفیت مشاهدات و محور عمودیش نشان دهنده فراوانی مطلق یا نسبی هر گروه است ترسیم می شود. در نمودار میله ای، خطوط جایگزین مستطیل ها می شوند تا بر این موضوع تاکید شود که فراوانی ها واقعا روی فاصله پخش نشده اند.

مثال۱۰: در مدیریت، انسان ها را به لحاظ ارتباطات به چهار دسته تصویری، احساسی، صوتی و ارقامی تقسیم می کنند. کارکنان یک سازمان از این لحاظ مورد بررسی قرار گرفته اند که حاصل تحقیق در این جدول آمده است:

| گروه ارتباطی | تصویری | احساسی | صوتی | ارقامی |
|---|---|---|---|---|
| تعداد کارکنان | ۱۰۰ | ۱۵۰ | ۳۰۰ | ۵۰ |




### ۲) نمودار دایره ای

این نمودار ابزار مناسبی برای تجسم مشاهدات بوده و معمولاً بر حسب در صد تهیه می شود و به نمودار کلوچه ای نیز معروف است

مراحل تهیه نمودار دایره ای

۱- تبدیل فراوانی مطلق به نسبی

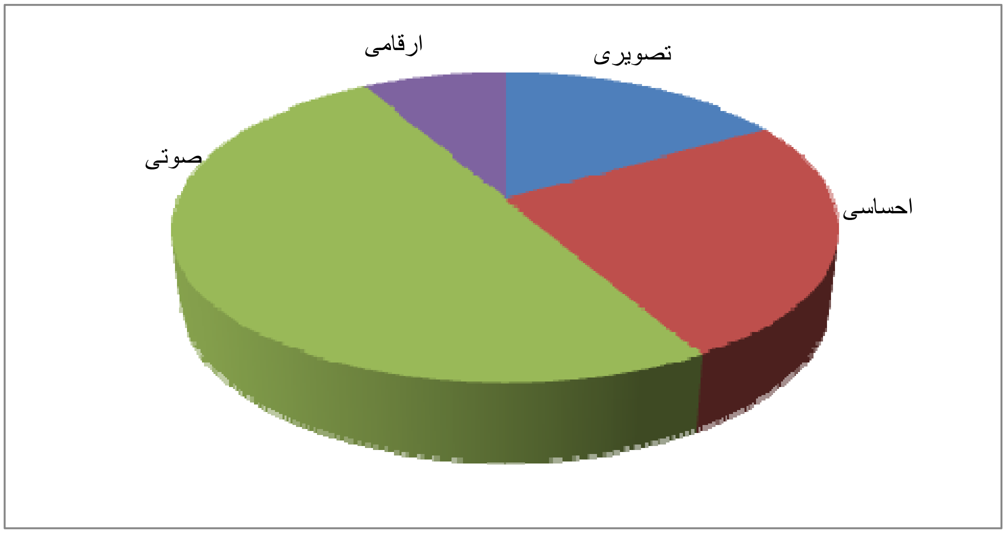
۲- پیدا کردن مساحت هر قطاع از دایره با استفاده از رابطه $S_i = ۳۶۰ \times f_i$

۳- تقسیم مساحت دایره بر حسب $S_i$ ها

۴- نوشتن نوع و درصد مشاهدات بر روی دایره

مثال۱۱: نمودار دایره ای مشاهدات مثال ۱۰ را رسم کنید.

| گروه ارتباطی | تصویری | احساسی | صوتی | ارقامی |
|---|---|---|---|---|
| تعداد کارکنان | ۱۰۰ | ۱۵۰ | ۳۰۰ | ۵۰ |
| فراوانی نسبی | ۰/۱۶۶۷ | ۰/۲۵ | ۰/۵ | ۰/۰۸۳ |
| $S_i$ | ۶۰ | ۹۰ | ۱۸۰ | ۳۰ |




### ۳) نمودار پاره تو

نحوه رسم نمودار

۱- مرتب کردن موضوعات بر اساس فراوانی مطلق آنها به صورت نزولی ( مورد اول دارای بیشترین فراوانی و مورد آخر دارای کمترین فراوانی است.)

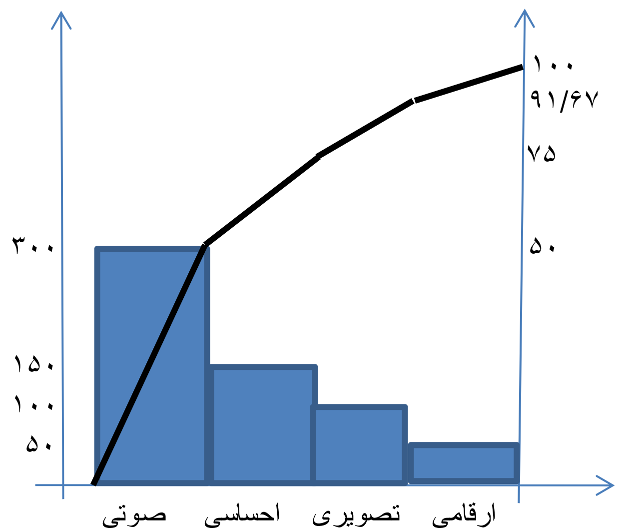
۲- مشخص کردن موارد روی نمودار افقی

۳- مشخص کردن فراوانی مطلق موارد روی نمودار عمودی با استفاده از نمودار ستونی، برای رسم نمودار، ستون ها به شکل مستطیل در نظر گرفته می شوند.

۴- مشخص کردن درصد فراوانی نسبی تجمعی موارد روی محور سوم (روبروی محور عمودی) با استفاده از نمودار فراوانی تجمعی

مثال۱۱: نمودار پاره تو مشاهدات مثال ۱۰ را رسم کنید.

| گروه ارتباطی | صوتی | احساسی | تصویری | ارقامی |
|---|---|---|---|---|
| تعداد کارکنان | ۳۰۰ | ۱۵۰ | ۱۰۰ | ۵۰ |
| فراوانی تجمعی | ۳۰۰ | ۴۵۰ | ۵۵۰ | ۶۰۰ |
| درصد فراوانی نسبی تجمعی | ۵۰ | ۷۵ | ۹۱/۶۷ | ۱۰۰ |




مفهوم نزولی بودن نمودار پاره تو

یعنی این که در این نمودار پر وقوع ترین موضوعات در سمت چپ نمودار قرار گرفته ، سپس موضوعات با فراوانی کمتر در سمت راست آنها قرار می گیرند.

کاربرد نمودار پاره تو

۱- در تحلیل موجودیهای جنسی انبارها
۲- در بررسی نواقص سیستم ها
۳- در بررسی نحوه توزیع درآمد و توزیع پرسنل مؤسسه

## فصل چهارم: توصیف مقداری مشاهدات طبقه بندی شده

هدف اصلی این فصل آشنا ساختن دانشجویان با پارامترهای مرکزی ، پراکندگی و تعیین انحراف از قرینگی و کشیدگی در داده های طبقه بندی شده می باشد

**سؤالاتی که توزیع فراوانی به آنها پاسخ می دهد.**

۱- مرکز توزیع کجاست ؟

۲- پراکندگی آن چقدر است ؟

۳- تمایل داده به کدام سمت است ؟

۴- پراکندگی توزیع در مقایسه با توزیع های مشابه چگونه است ؟

سؤال اول به کمک پارامترهای مرکزی قابل پاسخ‌گویی است به سؤال دوم نیز با محاسبهٔ پارامترهای پراکندگی جواب داده می‌شود ولی برای پارامترهای مرکزی و پراکندگی فرمول‌های جدیدی بر اساس مشاهدات طبقه بندی شده ارائه می‌شود. برای پاسخ‌گویی به سوال سوم از پارامترهایی با عنوان «پارامترهای تعیین انحراف از قرینگی» استفاده می‌شود و به سؤال چهارم با محاسبهٔ شاخص کشیدگی جواب داده می‌شود.


## ❖ انواع پارامترهای مرکزی در داده های طبقه بندی شده

۱- میانگین ؛ که به روش های مستقیم و غیرمستقیم قابل محاسبه است.

۲- مد ؛ که نشان دهنده بیشترین تکرار می باشد.

۳- چندکها ؛ شامل چارکها ، دهکها و صدکها است.

## ۱- میانگین

### ۱-۱ میانگین به روش مستقیم

این فرمول برای داده های طبقه بندی شده به شرح ذیل است :

$$\mu_x \cong \frac{\sum F_i X_i}{N}$$

که در آن $F_i$، فراوانی مطلق، $X_i$، متوسط طبقات و $N$، تعداد کل مشاهدات است.

### ۱-۲ میانگین به روش غیرمستقیم ( کد گذاری )

$$\mu_x \cong A + (\frac{\sum F_i d_i}{N}) I$$

که $A$، عدد دلخواه، $d_i$، کد هر طبقه و $I$ فاصله طبقات است.

در مقدار A می توان گفت، این عدد بعنوان میانگین تقریبی، از وسط ستون نماینده طبقات انتخاب شده و موجب تسهیل عملیات ریاضی در پیدا کردن میانگین تقریباً واقعی می شود.

در مورد $d_i$ که کد هر طبقه است ، به شکل زیر قابل محاسبه می باشد.

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

موارد استفاده از فرمول میانگین به روش کد گذاری زمانی است که مشاهدات حالت اعشار داشته یا این که به گونه ای تعریف شوند که محاسبه میانگین به روش مستقیم وقت گیر و مشکل آفرین باشد.

ستون های جدول توزیع فراوانی برای روش غیر مستقیم میانگین این ستون ها ضروری هستند :

۱- حدود طبقات	۴- حاصلضرب فراوانی در نماینده طبقه

۲- فراوانی مطلق	۵- کد طبقات

۳- نماینده طبقات	۶- حاصلضرب فراوانی در کد طبقه

- نکته: از علامت ≅ (تقریباً مساوی) در فرمول های میانگین بدین جهت استفاده می شود که پارامترها به واسطه طبقه بندی مشاهدات (استفاده از نماینده طبقات) دقیق نمی باشند.

مثال ۱:جدول توزیع فراوانی ارقام سود روزانه ۲۵ دکه روزنامه فروشی در زیر آمده است. میانگین مشاهدات را حساب کنید.

| فراوانی مطلق ($F_i$) | حدود طبقات |
|---|---|
| ۳ | [۵۵/۳ -۶۲/۳) |
| ۲ | [۶۲/۳ -۶۹/۳) |
| ۵ | [۶۹/۳ -۷۶/۳) |
| ۳ | [۷۶/۳ -۸۳/۳) |
| ۱۲ | [۸۳/۳ -۹۰/۳] |
| ۲۵ | کل |

**حل:**

| حدود طبقات | فراوانی مطلق ($F_i$) | مرکز دسته ($x_i$) | $F_i x_i$ | کد طبقه ($d_i$) | $F_i d_i$ |
|---|---|---|---|---|---|
| [۵۵/۳ –۶۲/۳) | ۳ | ۵۸/۸ | ۱۷۶/۴ | −۲ | −۶ |
| [۶۲/۳ –۶۹/۳) | ۲ | ۶۵/۸ | ۱۳۱/۶ | −۱ | −۲ |
| [۶۹/۳ –۷۶/۳) | ۵ | ۷۲/۸ | ۳۶۴ | ۰ | ۰ |
| [۷۶/۳ –۸۳/۳) | ۳ | ۷۹/۸ | ۲۳۹/۴ | ۱ | ۳ |
| [۸۳/۳ –۹۰/۳] | ۱۲ | ۸۶/۸ | ۱۰۴۱/۶ | ۲ | ۲۴ |
| کل | ۲۵ | | | | |

- میانگین مستقیم

$$\mu_x \cong \frac{\sum F_i X_i}{N} = \frac{۳\times ۵۸/۸+۲\times ۶۵/۸+۵\times ۷۲/۸+۳\times ۷۹/۸+۱۲\times ۸۶/۸}{۲۵} = \frac{۱۹۵۳}{۲۵} = ۷۸/۱۲$$

- میانگین غیرمستقیم

با توجه به اینکه وسط، مرکز دسته ها عدد ۷۲/۸ است. داریم:

$$A = ۷۲/۸$$

$$\mu_x \cong A + \left(\frac{\sum F_i d_i}{N}\right) I = ۷۲/۸ + \left(\frac{-۶-۲+۰+۳+۲۴}{۲۵}\right)\times ۷ = ۷۸/۱۲$$

**۲– مد ( نما )**

تعریف مد بصورت بیشترین تکرار برای داده های پیوسته و طبقه بندی شده بخوبی گویا و رسا نیست و رسایی آن فقط در مورد طبقه مددار می باشد.

در داده های طبقه بندی شده، مد از طریق زیر محاسبه می شود.

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

اجزاء تشکیل دهنده فرمول مد عبارت است از:

$L_{Mo}$ =حد پایین واقعی طبقه مد دار

$d_1 = F_i - F_{i-1}$ =فراوانی مطلق طبقه مدار منهای فراوانی طبقه ماقبل

$d_2 = F_i - F_{i+1}$ =فراوانی مطلق طبقه مدار منهای فراوانی طبقه ما بعد

مثال۲: برای مشاهدات ارائه در مثال ۱ مد را حساب کنید.

**حل:**

$$L_{Mo} = ۸۳/۳$$

$$d_1 = F_i - F_{i-1} = ۱۲ - ۳ = ۹$$

$$d_2 = F_i - F_{i+1} = ۱۲ - ۰$$

$$Mo \cong L_{Mo} + \left(\frac{d_1}{d_1 + d_2}\right) I = ۸۳/۳ + \left(\frac{۹}{۹ + ۱۲}\right) \times ۷ = ۸۶/۳$$

## ۳- چندکها

با تقسیم دامنه تغییرات به چهار قسمت مساوی به چارکها، به ده قسمت مساوی به دهکها و به صد قسمت مساوی به صدکها خواهیم رسید.

- **کاربرد چندکها**

**۱**- در کنترل کیفیت آماری

۲-در مدیریت

۳-در اقتصاد کلان و سایر علوم مشابه

۳-۱ چارک:

مراحل محاسبه چندکها

۱- اضافه کردن ستون فراوانی تجمعی به جدول

۲- پیدا کردن محل چارک مورد نظر با استفاده از

$$CQ_a = \frac{aN}{٤}$$

که **a= شماره چارک (۲،۱ یا ۳) و N= تعداد کل مشاهدات.**

۳- پیدا کردن طبقه چارک دار و استفاده از فرمول چارک **در داده های طبقه بندی شده**

$$Q_a \cong L_{Q_a} + \left(\frac{\frac{aN}{٤} - Fc_{i-١}}{F_i}\right) I$$

که در آن

$Q_a$ = مقدار چارک

$L_{Q_a}$ = حد پایین واقعی طبقه چارک دار

$Fc_{i-١}$ = فراوانی تجمعی طبقه ما قبل طبقه چارک دار

$F_i$ = فراوانی مطلق طبقه چارک دار

## ۲-۳ مراحل محاسبه دهکها $(D_a)$

۱- اضافه کردن $F_C$ به جدول

۲- پیدا کردن محل دهک با استفاده از $C_{D_a} = \frac{aN}{١٠}$

۳- محاسبه دهک با استفاده از مراحل قبلی و فرمول دهک برای داده های طبقه بندی شده

$$D_a \cong L_{D_a} + \left(\frac{\frac{aN}{١٠} - Fc_{i-١}}{F_i}\right) I$$

که در فرمول فوق

$L_{D_a}$ = حد پایین واقعی طبقه دهک دار

$F_i$ = فراوانی مطلق طبقه دهک دار

$Fc_{i-١}$ = فراوانی تجمعی طبقه ما قبل طبقه دهک دار

$I$ = **فاصله طبقات**

## ۳-۳ صدکها

**صدکها را با $P_a$ نشان می دهند و مراحل محاسبه آن تقریباً مشابه دهکها و چارکها است و مقدار محاسبه شده نیز همانند سایر پارامترهای مربوط به جداول تقریبی است.**

**- اضافه کردن $F_C$ به جدول**

**۲- پیدا کردن محل صدک با استفاده از** $C_{P_a} = \frac{aN}{100}$

**۳- محاسبه صدک با استفاده از مراحل قبلی و فرمول شدک برای داده های طبقه بندی شده**

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

**که در فرمول فوق**

$L_{P_a}$ = **حد پایین واقعی طبقه صدک دار**

$F_i$ = **فراوانی مطلق طبقه صدک دار**

$Fc_{i-1}$ = **فراوانی تجمعی طبقه ما قبل طبقه صدک دار**

$I$ = **فاصله طبقات**

مثال۳: برای مشاهدات ارائه در مثال ۱ موارد زیر را حساب کنید.

الف) چارک ها.

ب) دهک اول، پنجم و هفتم.

ج)صدک ۲۵، ۵۰ و ۷۰.

حل:

| حدود طبقات | فراوانی مطلق ($F_i$) | فراوانی تجمعی ($Fc_i$) |
|---|---|---|
| [۵۵/۳ -۶۲/۳) | ۳ | ۳ |
| [۶۲/۳ -۶۹/۳) | ۲ | ۵ |
| [۶۹/۳ -۷۶/۳) | ۵ | ۱۰ |
| [۷۶/۳ -۸۳/۳) | ۳ | ۱۳ |
| [۸۳/۳ -۹۰/۳] | ۱۲ | ۲۵ |
| کل | ۲۵ | |

الف)

$$CQ_{۱} = \frac{۱\times ۲۵}{۴} = ۶/۲۵, \quad Q_{۱} \cong ۶۹/۳ + \left(\frac{\frac{۱\times ۲۵}{۴} - ۵}{۵}\right)\times ۷ = ۷۱/۰۵$$

$$CQ_{۲} = \frac{۲\times ۲۵}{۴} = ۱۲/۵, \quad Q_{۱} \cong ۶۹/۳ + \left(\frac{\frac{۲\times ۲۵}{۴} - ۱۰}{۳}\right)\times ۷ = ۷۵/۱۳$$

$$CQ_{۲} = \frac{۳\times ۲۵}{۴} = ۱۸/۷۶, \quad Q_{۱} \cong ۸۳/۳ + \left(\frac{\frac{۳\times ۲۵}{۴} - ۱۳}{۱۲}\right)\times ۷ = ۸۶/۶۶$$

ب)

$$C_{D_{۱}}=\frac{۱\times ۲۵}{۱۰} \qquad D_{۱}\cong ۵۵/۲+(\frac{\frac{۱\times ۲۵}{۱۰}-۰}{۲})\times ۷=۶۱/۱۳$$

$$C_{D_{۵}}=\frac{۵\times ۲۵}{۱۰} \qquad D_{۵}\cong ۷۶/۲+(\frac{\frac{۵\times ۲۵}{۱۰}-۱۰}{۳})\times ۷=۸۲/۱۳$$

$$C_{D_{۷}}=\frac{۷\times ۲۵}{۱۰} \qquad D_{۷}\cong ۸۲/۲+(\frac{\frac{۷\times ۲۵}{۱۰}-۱۳}{۱۳})\times ۷=۸۵/۹۲۵$$

ج)

$$C_{P_{۲۵}}=\frac{۲۵\times ۲۵}{۱۰۰}=۶/۲۵ \qquad P_{۲۵}\cong ۶۹/۲+(\frac{\frac{۲۵\times ۲۵}{۱۰۰}-۵}{۵})\times ۷=۷۱/۰۵$$

$$C_{P_{۵۰}}=\frac{۵۰\times ۲۵}{۱۰۰}=۱۲/۵ \qquad P_{۲۵}\cong ۷۶/۲+(\frac{\frac{۵۰\times ۲۵}{۱۰۰}-۱۰}{۳})\times ۷=۸۲/۱۳$$

$$C_{P_{۷۰}}=\frac{۷۰\times ۲۵}{۱۰۰}=۱۷/۵ \qquad P_{۲۵}\cong ۸۲/۲+(\frac{\frac{۷۰\times ۲۵}{۱۰۰}-۱۳}{۱۳})\times ۷=۸۵/۹۲۵$$

➢ **نکته مهم :** هنگام محاسبه پارامترهای مرکزی، مثل مد و چندک ها (چارک ها، دهک ها و صدک ها) باید طول و عرض طبقه مساوی باشد. به عبارت دیگر چنانچه روش طبقه بندی با تقریب صورت گرفته باشد باید از حد پایین کرانه ها واقعی برای محاسبه این دسته از پارامترها استفاده شود.

مثال۴: سن ۳۰ مدیر موسسه در جدول توزیع فراوانی زیر ارائه شده است.

| فراوانی مطلق ($F_i$) | حدود طبقات |
|---|---|
| ۵ | ۳۵-۴۱ |
| ۱۰ | ۴۲ -۴۷ |
| ۳ | ۴۸ -۵۳ |

| | |
|---|---|
| ۵۴ -۵۹ | ۵ |
| ۶۰ -۶۵ | ۵ |
| ۶۶ -۷۱ | ۲ |
| کل | ۳۰ |

الف) برای مشاهدات فوق میانگین مستقیم و غیرمستقیم را محاسبه کنید.

ب) مد مشاهدات فوق را به دست آورید.

ج) میانه (چارک دوم)، دهک هشتم و صدک نودم را محاسبه کنید.

حل: با توجه به گسسته بودن حدود طبقات ابتدا کرانه های واقعی را محاسبه می کنیم.

| حدود طبقات | کرانه های واقعی | مرکز دسته ($x_i$) | فراوانی مطلق ($F_i$) | $d_i$ | فراوانی تجمعی ($FC_i$) |
|---|---|---|---|---|---|
| ۳۵ -۴۱ | ۳۴/۵ -۴۱/۵ | ۳۸ | ۵ | -۲/۵ | ۵ |
| ۴۲ -۴۷ | ۴۱/۵-۴۷/۵ | ۴۴ | ۱۰ | -۱/۵ | ۱۵ |
| ۴۸ -۵۳ | ۴۷/۵-۵۳/۵ | ۵۰ | ۳ | -۰/۵ | ۱۸ |
| ۵۴ -۵۹ | ۵۳/۵-۵۹/۵ | ۵۶ | ۵ | ۰/۵ | ۲۳ |
| ۶۰ -۶۵ | ۵۹/۵-۶۵/۵ | ۶۲ | ۵ | ۱/۵ | ۲۸ |
| ۶۶ -۷۱ | ۶۵/۵-۷۱/۵ | ۶۸ | ۲ | ۲/۵ | ۳۰ |
| کل | | | ۳۰ | | |

**الف)** میانگین مستقیم

$$\mu_x \cong \frac{\sum F_i X_i}{N} = \frac{۵\times ۳۸ + ۱۰\times ۴۴ + ۳\times ۵۰ + ۵\times ۵۶ + ۵\times ۶۲ + ۲\times ۶۸}{۳۰} = \frac{۱۵۰۶}{۳۰} = ۵۰/۲$$

میانگین غیرمستقیم

با توجه به اینکه وسط، مرکز دسته ها اعداد ۵۰ و ۵۶ است. داریم:

$$A = \frac{۵۰ + ۵۶}{۲} = ۵۳$$

$$\mu_x \cong A + \left(\frac{\sum F_i d_i}{N}\right) I = ۵۳ + \left(\frac{-۱۲/۵ - ۱۵ - ۱/۵ + ۲/۵ + ۷/۵ + ۵}{۳۰}\right)\times ۶ = ۵۰/۲$$

ب)

$$L_{Mo} = ۴۱/۵$$

$$d_۱ = F_i - F_{i-۱} = ۱۰ - ۵ = ۵$$

$$d_۲ = F_i - F_{i+۱} = ۱۰ - ۳ = ۷$$

$$Mo \cong L_{Mo} + \left(\frac{d_۱}{d_۱ + d_۲}\right) I = ۴۱/۵ + \left(\frac{۵}{۵+۷}\right)\times ۶ = ۴۴$$

ج)

میانه (چارک دوم):

$$CQ_۲ = \frac{۲\times ۳۰}{۴} = ۱۵, \quad Q_۱ \cong ۴۷/۵ + \left(\frac{\frac{۲\times ۳۰}{۴} - ۱۵}{۳}\right)\times ۶ = ۴۷/۵$$

**دهک هشتم:**

$$C_{D_8} = \frac{8 \times 30}{10} = 24 \qquad D_7 \cong 59/5 + \left(\frac{\frac{8 \times 30}{10} - 23}{5}\right) \times 6 = 60/7$$

صدک نودم:

$$C_{P_{90}} = \frac{90 \times 30}{100} = 27 \qquad P_{25} \cong 59/5 + \left(\frac{\frac{90 \times 30}{100} - 23}{5}\right) \times 6 = 64/3$$

### ❖ پارامترهای پراکندگی در داده های طبقه بندی شده

**۱- انحراف متوسط از میانگین** $(A \cdot D_{\mu})$

**۲- دامنه میان چارکی** $(IQR)$

**۳- انحراف چارکی** $(SIQR)$

**۴- واریانس** $(\sigma_x^2)$ **که شامل روش مستقیم و غیرمستقیم می شود.**

**۵- واریانس تصحیح شده** $(\sigma_c^2)$

### ۱- انحراف متوسط از میانگین

**فرمول محاسبه انحراف متوسط از میانگین در داده های طبقه بندی به شکل زیر می باشد در این فرمول $F_i$ فراوانی مطلق طبقه i ام می باشد.**

$$A \cdot D_{\mu} = \frac{\sum F_i \left| X_i - \mu_x \right|}{N}$$

مثال۵: برای مشاهدات مثال۱ ، انحراف متوسط ازمیانگین را جساب کنید.

حل:

$$A \cdot D_{\mu} = \frac{۳ \times \left| ۵۸/۸ - ۷۸/۱۲ \right| + ۲ \times \left| ٦۵/۸ - ۷۸/۱۲ \right| + ۵ \times \left| ۷۲/۸ - ۷۸/۱۲ \right| + ۳ \times \left| ۷۹/۸ - ۷۸/۱۲ \right| + ۱۲ \times \left| ۸٦/۸ - ۷۸/۱۲ \right|}{۲۵} = \frac{۲۱۹/۵٤}{۲۵} = ۸/۷۸$$

### ۲- دامنه میان چارکی

برای محاسبه دامنه میان چارکی مانند قبل از فرمول زیر استفاده می شود.

$$IQR = Q_۳ - Q_۱$$

مثال۶: برای مشاهدات مثال۱ ، دامنه مین چارکی را جساب کنید.

$$IQR = Q_۳ - Q_۱ = ۸٦/٦٦ - ۷۱/۰۵ = ۱۵/٦۱$$

### ۳- انحراف چارکی

برای محاسبه انحراف چارکی مانند قبل از فرمول زیر استفاده می شود.

$$SIQR = \frac{Q_۳ - Q_۱}{۲}$$

مثال۷: برای مشاهدات مثال۱ ، انحراف چارکی را جساب کنید.

$$SIQR = \frac{Q_۳ - Q_۱}{۲} = \frac{۸۶/۶۶ - ۷۱/۰۵}{۲} = \frac{۱۵/۶۱}{۲} = ۷/۸۰۵$$

- **نکته**: از پارامترهای پراکندگی دامنه میان چارکی و انحراف چارکی زمانی استفاده می شود که دنباله های توزیع نامعین و باز باشد ( در این حالت محاسبه میانگین و واریانس امکان پذیر نیست.)

**مثال۸**: در کشور آمریکا، در سال های ۱۹۰۰ تا ۱۹۷۳، تعداد تلفات انسانی در هر طوفان بزرگی ثبت شده و فراوانی این تلفات، به قرار زیر بوده است:

| تعداد مرگ | فراوانی مطلق ($F_i$) |
|---|---|
| ۲۴ و کمتر | ۴ |
| ۲۵ -۴۹ | ۱۶ |
| ۵۰ -۷۴ | ۱۶ |
| ۷۵ -۹۹ | ۱۱ |
| ۱۰۰ -۱۴۹ | ۶ |
| ۱۵۰ -۱۹۹ | ۲ |
| ۲۰۰-۲۴۹ | ۴ |
| ۲۵۰ و بیشتر | ۱ |
| کل | ۶۰ |

برای داده های فوق پارامترهای پراکندگی را محاسبه کنید.

**حل:** با توجه به مشخص نبودن حدود در طبقات اول و هشتم، امکان محاسبه میانگین و مرکز دسته های این طبقات وجود ندارد بنابراین ما از دامنه  میان چارکی و انحراف چارکی برای محاسبه پارامترهای پراکندگی استفاده می کنیم.

برای محاسبه چارک ها با توجه به گسسته بودن طبقه بندی ابتدا کرانه های واقعی را به دست می آوریم.

| تعداد مرگ | کرانه های واقعی | فراوانی مطلق ( $F_i$ ) | فراوانی تجمعی ( $Fc_i$ ) |
|---|---|---|---|
| ۲۴ و کمتر | ۲۴/۵ و کمتر | ۴ | ۴ |
| ۲۵–۴۹ | ۲۴/۵–۴۹/۵ | ۱۶ | ۲۴ |
| ۵۰–۷۴ | ۴۹/۵–۷۴/۵ | ۱۶ | ۴۰ |
| ۷۵–۹۹ | ۷۴/۵–۹۹/۵ | ۱۱ | ۵۱ |
| ۱۰۰–۱۴۹ | ۹۹/۵–۱۴۸/۵ | ۶ | ۵۷ |
| ۱۵۰–۱۹۹ | ۱۴۹/۵–۱۹۹/۵ | ۲ | ۵۹ |
| ۲۰۰–۲۴۹ | ۱۹۹/۵–۲۴۹/۵ | ۴ | ۵۶۳ |
| ۲۵۰ و بیشتر | ۲۴۹/۵ و بیشتر | ۱ | ۶۰ |
| کل | | ۶۰ | |

$$I = ۴۹ - ۲۵ + ۱ = ۲۵$$

$$CQ_{۱} = \frac{۱ \times ۶۰}{۴} = ۱۵, \quad Q_{۱} \cong ۲۴/۵ + \left(\frac{\frac{۱ \times ۶۰}{۴} - ۴}{۱۶}\right) \times ۲۵ = ۴۰/۱۲۵$$

$$CQ_{۳} = \frac{۳ \times ۶۰}{۴} = ۴۵, \quad Q_{۳} \cong ۷۴/۵ + \left(\frac{\frac{۳ \times ۶۰}{۴} - ۳۶}{۱۱}\right) \times ۲۵ = ۹۴/۹۵$$

$$IQR = Q_{۳} - Q_{۱} = ۹۴/۹۵ - ۴۰/۱۲۵ = ۵۴/۸۲۵$$

$$SIQR = \frac{Q_{۳} - Q}{۲} = \frac{۹۴/۹۵ - ۴۰/۱۲۵}{۲} = \frac{۵۴/۸۲۵}{۲} = ۲۷/۴۱$$

**۴- واریانس**

همانطور که بیان شد واریانس در داده های طبقه بندی شده به دو صورت مستقیم و غیرمستقیم محاسبه می شود.

- **روش مستقیم**

واریانس به روش مستقیم در داده های طبقه بندی شده با استفاده از فرمول زیر محاسبه می شود.

$$\sigma_x^{۲} = \frac{\sum F_i (X_i - \mu_X)^{۲}}{N}$$

به منظور سادگی در محاسبات، به جای فرمول فوق می توان از فرمول زیر جهت محاسبه **واریانس به روش مستقیم برای داده های طبقه بندی شده استفاده کرد. ( می توان نشان داد هر دو فرمول معادل هستند.)**

$$\sigma_x^{۲} = \frac{\sum F_i X_i^{۲}}{N} - \mu_X^{۲}$$

- **روش غیر مستقیم**

واریانس به روش مستقیم در داده های طبقه بندی شده با استفاده از فرمول زیر محاسبه می شود.

$$\sigma_x^{۲} = I^{۲}\left[\frac{\sum F_i d_i^{۲}}{N} - \left(\frac{\sum F_i d_i}{N}\right)^{۲}\right]$$

مثال۹: برای مشاهدات مثال۱ ، واریانس را به دو روش مستقیم و غیرمستقیم محاسبه کنید.

حل: محاسبه واریانس مستقیم با استفاده از فرمول اول

$$\sigma_x^{۲} = \frac{۳\times(۵۸/۸-۷۸/۱۲)^{۲}+۲\times(۶۵/۸-۷۸/۱۲)^{۲}+۵\times(۷۲/۸-۷۸/۱۲)^{۲}+۳\times(۷۹/۸-۷۸/۱۲)^{۲}+۱۲\times(۸۶/۸-۷۸/۱۲)^{۲}}{۲۵} = \frac{۲۴۷۷/۴۴}{۲۵} = ۹۹/۰۹۷۶$$

محاسبه واریانس مستقیم با استفاده از فرمول اول

$$\sigma_x^{۲} = \frac{۳\times۵۸/۸^{۲}+۲\times۶۵/۸^{۲}+۵\times۷۲/۸^{۲}+۳\times۷۹/۸^{۲}+۱۲\times۸۶/۸^{۲}}{۲۵} - ۷۸/۱۲^{۲} = \frac{۱۵۵۰۴۵/۸}{۲۵} - ۷۸/۸۳۲ = ۹۹/۰۹۷۶$$

و انحراف معیار مشاهدات برابر خواهد بود با $\sigma_x = \sqrt{\sigma_x^{۲}} = \sqrt{۹۹/۰۹۷۶} = ۹/۹۵$

**محاسبه واریانس به روش غیرمستقیم**

$$\sigma_x^{۲} = ۷^{۲}\times\left[\frac{۳\times۴+۲\times۱+۵\times۰+۳\times۱+۱۲\times۴}{۲۵} - \left(\frac{۳\times(-۲)+۲\times(-۱)+۵\times۰+۳\times۱+۱۲\times۲}{۲۵}\right)^{۲}\right]$$
$$= ۴۹\times\left[\left(\frac{۶۵}{۲۵}\right)-\left(\frac{۱۹}{۲۵}\right)^{۲}\right] = ۹۹/۰۹۷۶$$

### ۵- واریانس تصحیح شده ($\sigma_c^۲$):

در داده‌های طبقه‌بندی شده برای محاسبهٔ میانگین و واریانس از نمایندهٔ طبقات استفاده می‌شود نتیجهٔ این عمل ممکن است دارای اختلاف با مقادیر واقعی داده‌ها باشد در میانگین اشتباه ناشی از این تقریب به علت مثبت و منفی بودن اشتباهات جبران می‌شود از این رو از مجموع خطاها صرف‌نظر می‌گردد. در مورد واریانس چون خطاهای مثبت و منفی به توان دو می‌رسد خطاها یکدیگر را خنثی نمی‌کنند بنابراین مقدار بدست آمده برای واریانس بیش از مقدار واقعی است برای تصحیح این اشتباه، شیارد رابطهٔ زیر را تعریف کرد که در آن صورت مقدار بدست آمده دقیقتر از واریانس خواهد بود

$$\sigma_c^۲ = \sigma_x^۲ - \frac{I^۲}{۱۲}$$

$I$ ← فاصلهٔ طبقات

باید توجه داشت این تصحیح در مواردی به کار می‌رود که اولاً متغیر پیوسته باشد ثانیاً تعداد N دست‌کم هزار باشد ثالثاً توزیع فراوانی از نوع متقارن یا اندکی متقارن باشد.

**مثال۱۰:** در مورد مشاهدات مثال ۱، واریانس تصحیح شده را محاسبه کنید.

$$\sigma_c^۲ = \sigma_x^۲ - \frac{I^۲}{۱۲} = ۹۹/۹۷۶ - \frac{۷^۲}{۱۲} = ۹۵/۱$$

### ❖ عملیات جبری میانگین و واریانس

اگر جامعه آماری از ترکیب چند جامعه مستقل با میانگین ها و واریانس های مشخص تشکیل شده باشد ، می توان میانگین و واریانس جامعه کل را بدست آورد.

- **میانگین حسابی جامعه کل**

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

$N_i$ = **تعداد مشاهدات جامعه i ام برای** $i = 1,2,\dots,k$

$N$ = **تعداد مشاهدات جامعه کل که** $N = N_1 + N_2 + \dots + N_k$

$\mu_i$ = **میانگین جامعه i ام برای** $i = 1,2,\dots,k$

- **واریانس جامعه کل**

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

که در آن

$\sigma_i^2$ = **واریانس جامعه i ام برای** $i = 1,2,\dots,k$

$N$ = **تعداد مشاهدات جامعه کل**

$N_i$ = **تعداد مشاهدات جامعه i ام برای** $i = 1,2,\dots,k$

$\mu$ = **میانگین جامعه کل**

$\mu_i$ = **میانگین جامعه i ام برای** $i = 1,2,\dots,k$

**مثال۱۱:** میانگین و واریانس نمرات ۳ کلاس در درس آمار به صورت زیر است میانگین و وایانس کل کلاس‌ها را محاسبه می‌کنیم.

| کلاس | ۱ | ۲ | ۳ |
|---|---|---|---|
| میانگین | ۱۲ | ۱۴ | ۱۸ |
| واریانس | ۹ | ۹ | ۴ |
| تعداد افراد | ۱۰ | ۱۰ | ۵ |

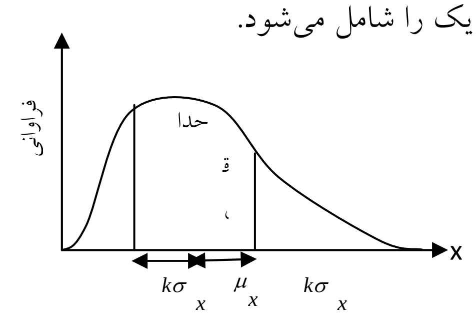
$$\mu=\frac{\sum N_i\,\mu_i}{N}=\frac{(۱۲\times۱۰)+(۱۴\times۱۰)+(۱۸\times۵)}{۲۵}=۱۴$$

$$\sigma^{۲}=\frac{\sum N_i\,\sigma_i^{۲}}{N}+\frac{\sum N_i\,(\mu_i-\mu)^{۲}}{N}=\frac{(۹\times۱۰)+(۹\times۱۰)+(۴\times۵)}{۲۵}+$$
$$+\frac{۱۰(۱۲-۱۴)^{۲}+۱۰(۱۴-۱۴)^{۲}+۵(۱۸-۱۴)^{۲}}{۲۵}=۸+\frac{۱۲۰}{۲۵}=۱۲/۸$$

- **اهمیت کاربرد انحراف معیار**

یکی از موارد استفاده از انحراف معیار تعیین درصد داده هایی است که در محدوده ای حول میانگین قرار می گیرند و در قضیه ای با نام چبیشف به صورت زیر اثبات شده است.

**قضیه چبیشف:**در هر توزیع آماری حداقل $\left[۱-\frac{۱}{k^{۲}}\right]$ ٪ (درصد) مشاهدات در فاصلهٔ $\mu\pm k\sigma$ قرار می‌گیرند. K عددی مثبت که مقادیر مساوی یا بزرگتر از یک را شامل می‌شود.



مثال ۱۲) فرض کنید توزیع قد دانشجویان کلاسی دارای میانگین ۱۶۶ و واریانس ۱۶

الف) فاصله‌ای را نسبت به میانگین محاسبه کنید که حداقل قد ۷۵ درصد افراد را دربرگیرد.

ب) در فاصلهٔ ۱۷۳–۱۵۹ قد حداقل چنددرصد دانشجویان قرار می‌گیرد.

حل :

الف) طبق قضیه چبیشف، در هر توزیع آماری حداقل $\left[۱-\frac{۱}{k^۲}\right]$ ٪ (درصد) مشاهدات در فاصلهٔ $\mu \pm k\sigma$ قرار می‌گیرند در نتیجه با توجه به سوال ما می خواهیم

$$۱-\frac{۱}{k^۲}=۰/۷۵ \rightarrow \frac{۱}{k^۲}=۰/۲۵ \rightarrow k^۲=۴ \rightarrow k=۲$$

$$\mu_x=۱۶۶ \qquad \sigma_x^۲=۱۶ \qquad \sigma_x=\sqrt{۱۶}=۴$$

$$\mu \pm k\sigma=۱۶۶ \pm ۲ \times ۴=(۱۵۸,۱۷۴)$$

ب :

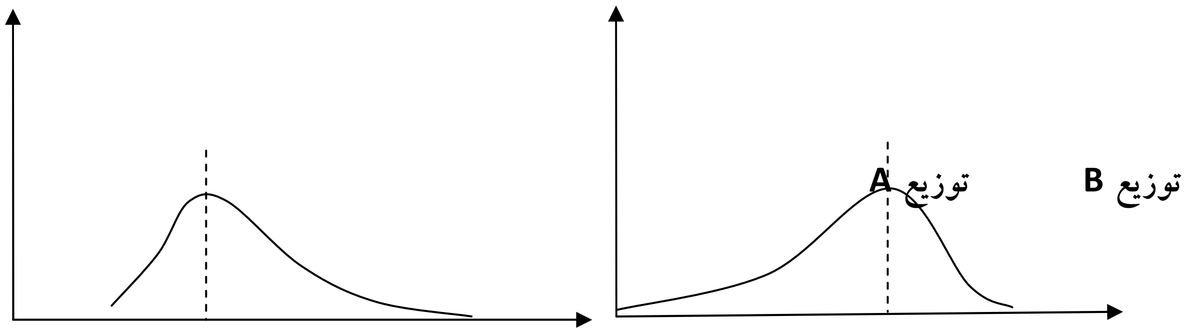
$$\mu-k\sigma=۱۵۹ \rightarrow ۱۶۶-k\times ۴=۱۵۹ \rightarrow ۴k=۷ \rightarrow k=\frac{۷}{۴}$$

$$۱-\frac{۱}{k^۲}=۱-\frac{۱}{(\frac{۷}{۴})^۲}=۱-\frac{۱۶}{۴۹}=\frac{۳۳}{۴۹}=۰/۶۷\times ۱۰۰=۶۷\%$$

یعنی ۶۷ درصد قد افراد در این فاصله قرار دارد.

## ❖ پارامترهای تعیین انحراف از قرینگی

**سوال: در هنگام مقایسه دو یا چند جامعه، در صورت مساوی بودن پارامترهای مرکزی و پراکندگی، آیا می توان گفت دو جامعه با هم برابرند؟**




جواب: لزوما دو جامعه با هم برابر نیستند و برابر بودن آنها بستگی به نوع توزیع مشاهدات حول میانگین دارد.

**مفهوم چولگی**

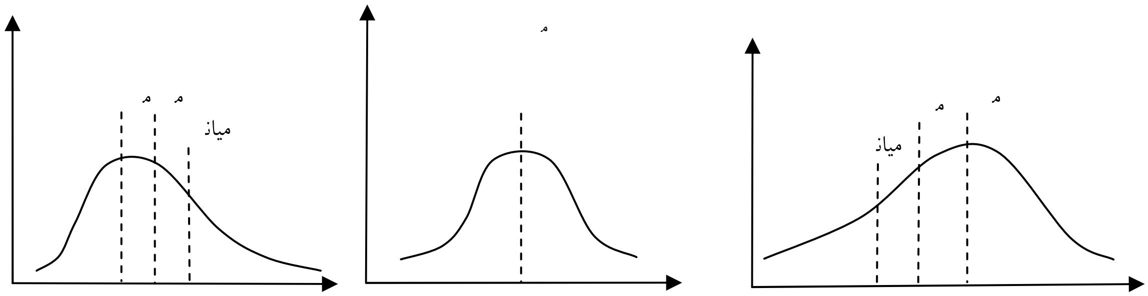
**اگر دم توزیع جامعه به سمت راست باشد ، توزیع را چوله به راست و در صورت عکـس ، آن را چوله به چپ می نامند.** چولگی توزیع‌ها در مقایسه با توزیع متقارن معین می‌شود.

**انواع حالات توزیع ها بر اساس میزان چولگی**

**۱- متقارن ( نرمال ) : مد = میانه = میانگین (مقدار ضریب چولگی (SK) برابر صفر است.)**

**۲- چـولـه به راسـت : مد < میانه < میانگین (مقدار ضریب چولگی (SK) مثبت است.)**

**۳- چـولـه بـه چـــپ : مد > میانه > میانگین (مقدار ضریب چولگی (SK) منفی است.)**



چوله به چپ $\mu_X < Md < Mo$

متقارن $\mu_X = Md = Mo$

چوله به راست $\mu_X > Md > Mo$

## ➢ تفسیر مقادیر SK

قدرمطلق ضریب چولگی نشان دهندهٔ میزان اختلاف جامعهٔ آماری با توزیع نرمال از نظر قرینگی است بدیهی است که هر چه $|sk|$ بزرگتر باشد اختلاف جامعهٔ آماری با توزیع نرمال از نظر قرینگی بیشتر خواهد بود به طوریکه :

**۱- $|SK| \leq ./۱$ ، جامعه از نظر قرینگی تقریباً نرمال است.**

**۲- $./۱ < |SK| \leq ./۵$ ، جامعه از نظر قرینگی تفاوت اندک با توزیع نرمال دارد.**

**۳- $|SK| > ./۵$ ، جامعه از نظر قرینگی تفاوت فاحش با توزیع نرمال دارد.**

- **محاسبه ضریب چولگی ( SK )**

**۱- ضریب چولگی گشتاوری**

**۲- ضریب های چولگی پیرسون**

**۳- ضریب های چولگی چندکی که شامل ضریب چولگی چارکی و ضریب چولگی صدکی است.**

- **ضریب چولگی گشتاوری**

**برای محاسبه ضریب چولگی گشتاوری از فرمول زیر استفاده می شود.**

$$SK = \frac{r_3}{\sigma_x^3}$$

که در آن $r_3 = \frac{\sum F_i (X_i - \mu_X)^3}{N}$، گشتاور مرتبه سوم به مبدأ میانگین و $\sigma_x$، انحراف معیار است.

- **ضریب چولگی پیرسون**

1. **فرمول شماره ۱** $$S \cdot K_1 = \frac{(\mu_X - Mo)}{\sigma_X}$$
2. **فرمول شماره ۲** $$S \cdot K_2 = \frac{3(\mu_X - Md)}{\sigma_X}$$

**که در روابط فوق Mo ، مد و Md میانه است.**

مثال ۱۳: برای مشاهدات مثال ۱ چولگی گشتاوری و چولگی پیرسون را محاسبه کنید.

حل: ضریب چولگی گشتاوری

$$r_{۳} = \frac{\sum F_i (X_i - \mu_X)^{۳}}{N} = \frac{-۱۸۲٦۵/۱٦}{۲۵} = -۷۳۰/٦۱ \ , \ \sigma_x^{۲} = ۹۹/۰۹۷٦ \Rightarrow \sigma_x = ۹/۹۵$$

$$SK = \frac{r_{۳}}{\sigma_x^{۳}} = \frac{-۷۳۰/٦۱}{۹/۹۵^{۳}} = -۰/۷٤$$

**ضریب چولگی پیرسون**

$$S \cdot K_{۱} = \frac{(\mu_X - Mo)}{\sigma_X} = \frac{(۷۸/۱۲ - ۸٦/۳)}{۹/۹۵} = -۰/۸۲۲$$

$$S \cdot K_{۲} = \frac{۳(\mu_X - Md)}{\sigma_X} = \frac{۳ \times (۷۸/۱۲ - ۷۵/۱۳)}{۹/۹۵} = -۰/۹۰۱$$

**با توجه به آنکه مقدار $|SK| > ۰/۵$ ، می توان گفت جامعه از نظر قرینگی تفاوت فاحش با توزیع نرمال دارد.**

- **ضرایب چولگی چندکی**

۱. **ضریب چولگی چارکی** $$SK_Q = \frac{Q_{۳} - ۲Q_{۲} + Q_{۱}}{Q_{۳} - Q_{۱}}$$

۲. **ضریب چولگی صدکی** $$SK_P = \frac{P_{۹۰} - ۲P_{۵۰} + P_{۱۰}}{P_{۹۰} - P_{۱۰}}$$

که در آن $Q_i$ چارک iام و $P_i$، صدک iام است.

➢ **نکته:** چولگی چندکی زمانی که محاسبه میانگین و واریانس برای توزیع امکانپذیر نباشد. کاربرد دارد.

مثال ۱۴: برای مشاهدات مثال ۸ چولگی چارکی و صدکی را محاسبه کنید.

**حل: ضریب چولگی چارکی**

$$CQ_{۲}=\frac{۲\times ٦۰}{٤}=۳۰,\quad Q_{۲}\cong ٤۹/٥+\left(\frac{\frac{۲\times ٦۰}{٤}-۲۰}{۱٦}\right)\times ۲٥=٦٥/۱۲٥$$

$$SK_{Q}=\frac{Q_{۳}-۲Q_{۲}+Q_{۱}}{Q_{۳}-Q_{۱}}=\frac{۹٤/۹٥-۲\times ٦٥/۱۲٥+٤۰/۱۲٥}{۹٤/۹٥-٤۰/۱۲٥}=۰/۰۸۸$$

**ضریب چولگ صدکی**

$$C_{P_{۱۰}}=\frac{۱۰\times ٦۰}{۱۰۰}=٦,\quad P_{۱۰}\cong ۲٤/٥+\left(\frac{\frac{۱۰\times ٦۰}{۱۰}-٤}{۱٦}\right)\times ۲٥=۲۷/٦۲٥$$

$$C_{P_{۹۰}}=\frac{۹۰\times ٦۰}{۱۰۰}=٥٤,\quad P_{۹۰}\cong ۱٤۹/٥+\left(\frac{\frac{۹۰\times ٦۰}{۱۰۰}-٥۳}{۲}\right)\times ۲٥=۱٦۲$$

$$P_{٥۰}=Q_{۲}=٦٥/۱۲٥$$

$$SK_{P}=\frac{P_{۹۰}-۲P_{٥۰}+P_{۱۰}}{P_{۹۰}-P_{۱۰}}=\frac{۱٦۲-۲\times ٦٥/۱۲٥+۲۷/٦۲٥}{۱٦۲-۲۷/٦۲٥}=۰/٤٤$$

## ❖ پارامترهای تعیین انحراف از کشیدگی

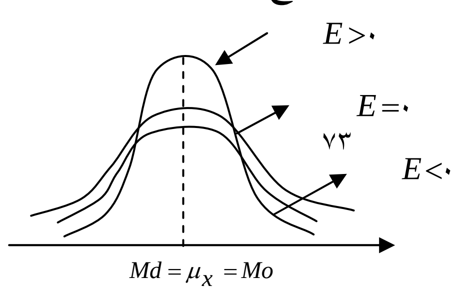
این پارامترها برای مقایسه توزیع جوامع مورد نظر با توزیع جامعه نرمال به لحاظ کشیدگی (کوتاهی و بلندی توزیع) مورد استفاده قرار می گیرد.

**انواع توزیع به لحاظ کشیدگی و مقدار ضریب ( E ) آن**

**۱-از نظر کشیدگی (بلندی توزیع) مساوی توزیع نرمال ( E=۰ )**

**۲- از نظر کشیدگی (بلندی توزیع) بلندتر از توزیع نرمال ( E>۰ )**

**۳- از نظر کشیدگی (بلندی توزیع) کوتاه تر از توزیع نرمال ( E<۰ )**



- **تفسیر ضریب کشیدگی ( E )**

**۱- اگر $|E| \le ۰/۱$ توزیع جامعه از نظر کشیدگی تقریبا نرمال است.**

**۲- اگر $۰/۱ < |E| \le ۰/۵$. توزیع جامعه از نظر کشیدگی نسبتاً بلندتر از نرمال است.**

**۳- اگر $|E| > ۰/۵$ توزیع جامعه از نظر کشیدگی کاملاً کشیده تر از نرمال است.**

- **انواع ضرایب کشیدگی**

**۱- ضریب کشیدگی گشتاوری ؛**

**۲- ضریب کشیدگی چندکی ؛**

- **ضریب کشیدگی گشتاوری**

این ضریب گشیدکی با استفاده از گشتاور مرتبه چهارم نسبت به مبدأ میانگین محاسبه می شود که فرمول آن به صورت زیر ارائه می شود.

$$E = \frac{r_{٤}}{\sigma_X^{٤}} - ۳$$

که $r_4 = \frac{\sum F_i(X_i - \mu_x)^4}{N}$ و ۳ عددی ثابت است.

مثال۱۵: برای مشاهدات مثال ۱ کشیدگی گشتاوری را محاسبه کنید.

$$r_4 = \frac{\sum F_i(X_i - \mu_x)^4}{N} = \frac{536197}{25} = 21447/88$$

$$E = \frac{r_4}{\sigma_X^4} - 3 = \frac{21447/88}{9/95^4} - 3 = 2/19 - 3 = -0/81$$

با توجه به اینکه $0/5 \leq |E| < 1/0$ می توان گفت، توزیع جامعه از نظر کشیدگی نسبتاً بلندتر از نرمال است.

- **ضریب کشیدگی چندکی**

این ضریب گشیدکی با استفاده از انحراف چارکی و صدکهای دهم و نودم محاسبه می شود که فرمول آن به صورت زیر ارائه می شود.

$$E_P = \frac{SIQR}{P_{90} - P_{10}} - 0/263$$

**نکته:**این ضریب گشیدگی مخصوص توزیع هایی که بالاجبار با استفاده از چندکها توصیف می شوند.

مثال ۱۶: برای مشاهدات مثال ۸ ضریب کشیدگی چندکی را محاسبه کنید.

$$E_P = \frac{SIQR}{P_{90} - P_{10}} - 0/263 = \frac{27/41}{162 - 27/625} - 0/263 = -0/42$$

با توجه به اینکه $0/5 \leq |E| < 1/0$،، می توان گفت توزیع جامعه از نظر کشیدگی نسبتاً بلندتر از نرمال است.

## فصل پنجم: مبادی احتمال

در این فصل دانشجو با برخی از مفاهیم احتمال و انواع احتمالات آشنا می شود.  مفاهیمی چون فضای نمونه، پیشامد و نیز اصول شمارش و جایگشت ها و ترکیب ها بیان شده است. و نیز مفهوم احتمال، تابع احتمال، اصل موضوع احتمال شرطی استقلال دو پیشامد قانون  و احتمال کل توضیح داده می شود.

❖ مفاهیم اساسی احتمال

- **مفهوم احتمال (P):** احتمال یعنی شانس وقوع یک پیشامد خاص.
- **احتمال عینی:** احتمال عینی ، ثابت و مقدار آن از قبل مشخص است و به عقاید اشخاص بستگی ندارد. (احتمال ۶ آمدن تاس در یک بار پرتاب)
- **احتمال ذهنی:** احتمال ذهنی ، متغیر و وابسته به نظر اشخاص است. (مثال: پاسخ مسافران در مورد احتمال تاخیر پرواز تهران–تبریز)
- **آزمایش** : فعالیتی که نتیجه آن از قبل مشخص نیست ولی کل حالات ممکن آن معلوم است، مثل پرتاب یک سکه ، که معلوم نیست دقیقاً شیر خواهد آمد یا خط
- **فضای نمونه** :مجموعه پیامدهای ممکن یک آزمایش را فضای نمونه آن آزمایش می گویند .فضای نمونه را با $S$ نشان می دهند.

مثال۱: فضای نمونه پرتاب یک تاس

$$S=\{١,٢,٣,٤,٥,٦\}$$

- **فضای نمونه محدود:** یعنی این که فضای نمونه تعداد کمی عضو داشته باشد. (مثل فضای نمونه پرتاب یک تاس)
- **فضای نمونه نامحدود**: یعنی اینکه فضـای نمونه آزمایش ( تعداد اعضاء آن ) نامتناهی است. مثال۲: یک سکه را آنقدر پرتاب می کنیم تا شیر ظاهر شود. فضای نمونه را بنویسید.

حل: اگر $\mathcal{H}$ نماد شیر و $\mathcal{T}$ نماد خط باشد داریم.

$$S=\{H,TH,TTH,\dots\}$$

- **فضای نمونه گسسته و پیوسته** : اگر اعضای فضای نمونه آزمایش قابل شمارش باشد ، آن را فضای نمونه گسسته ولی اگر فضای نمونه آزمایشی بصورت اعداد اعشاری باشند آن را پیوسته می نامند.

**مثال۳:** تعداد پرتاب های یک تاس تا آمدن ۶ نمونه ای از فضای نمونه گسسته و نامتناهی است.

$$S = \{۱, ۲, ۳, \ldots\}$$

فضای نمونه طول عمر لامپ های تولیدی یک کارخانه نمونه ای از فضای نمونه نامتناهی و پیوسته است.

$$S = \{x \geq ۰\}$$

فضای نمونه ظول عمر نوعی لامپ که حداکثر عمر آن ۱۷۸۰ ساعت است. نمونه ای از فضای نمونه متناهی و پیوسته است.

$$S = \{۰ \leq x \leq ۱۷۸۰\}$$

- **پیشامد** : به هر یک از زیر مجموعه های فضای نمونه ، یک پیشامد گفته می شود هر پیشامد را با یکی از حروف بزرگ انگلیسی مثل $\mathcal{A}$ و $\mathcal{B}$ و $C$ و . . . نشان می دهند.

مثال۴: B پیشامد ظاهر شدن عدد زوج در پرتاب یک تاس

$$S = \{۱, ۲, ۳, ٤, ٥, ٦\}$$
$$B = \{۲, ٤, ٦\}$$

➢ **پیامدهای مقدماتی هم شانس** : پیامدهای مقدماتی یا پیشامدهای اولیه هم شانس یعنی این که تمام پیشامدهای اولیه در آزمایشی دارای شانس وقوع برابر باشند. (مثال: احتمال رو آومدن هر سک از اعداد ۱ تا ۶ در پرتاب یک تاس)

## ❖ احتمال یک پیشامد

- **احتمال یک پیشامددر پیامدهای مقدماتی هم شانس**

احتمال وقوع پیشامدی مثل $\mathcal{A}$ برابر می شود با تعدادهای عضو های پیشامد $\mathcal{A}$ به تعداد عضوهای فضای نمونه

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

مثال۴: احتمال پیشامد ظاهر شدن عدد زوج در پرتاب یک تاس.

$$P(B)=\frac{۳}{٦}=\frac{۱}{۲}$$

- **احتمال یک پیشامد در پیامدهای مقدماتی غیر هم شانس**

برای پیشامدی مثل $\mathcal{A}$ $\qquad$ فراوانی نسبی پیشامد $\mathcal{A}$ در$\mathcal{N}$تکرار $=\mathcal{P}(\mathcal{A})$

بشرطی که $\mathcal{N}$ به سمت بی نهایت میل کند.

## ❖ خواص اولیه احتمال

برای هر پیشامدی مثل $\mathcal{A}$ (چه هم شانس و چه غیر هم شانس)

$$۱-\quad .\le P(A)\le ۱$$
$$۲-\quad P(S)=۱$$

## ❖ قواعد شمارش

این قواعد عبارتند از :

۱- قاعده ضرب ۲- جایگشت (ترتیب)

۳- ترکیب ۴- افرازهای (تفکیک های) مرتب

**کاربردهای قواعد شمارش**

از این قواعد در وضعیت هایی استفاده می شود که فهرست نمودن تمام حالات ممکن آزمایش مقدور نمی باشد ، لذا فقط به ذکر تعداد حالات ممکن و مختلف اکتفا می شود.

**۱- اصل اساسی شمارش**

اساسی ترین اصل در شمارش « قاعده ضرب » است و این اصل مختص آزمایش هایی است که در آنها عملیات در چند مرحله ( مثلاً $K$ ) مرحله انجام می پذیرد.

- **قاعده ضرب** : طرق ممکن انجام عمل در آزمایشی که مرحله اول آن به $n_1$ طریق و . . . مرحله $K$ ام آن به $n_K$ طریق انجام میگیرد ، عبارت خواهد بود از :

$$n_1 \times n_2 \times \dots \times n_K$$

مثال۵: اگر دایره بازاریابی و فورش شرکتی بخواهد یکی از ۵ متن تهیه شده را با یکی از وسیله های تبلیغاتی ( رادیو، تلویزیون، مجله و روزنامه) آگهی کند، این کار را به چند طریق می تواند انجام دهد.

حل: با توجه به اینکه این کار در ۲ مرحله انجام می شود به طوری که $n_1 = ۵$ و $n_2 = ٤$ بوده و این دو مرحله از هم مستقل هستند. پس تعداد کل حالات برابر است با:

$$n_1 \times n_2 = ۵ \times ٤ = ۲۰$$

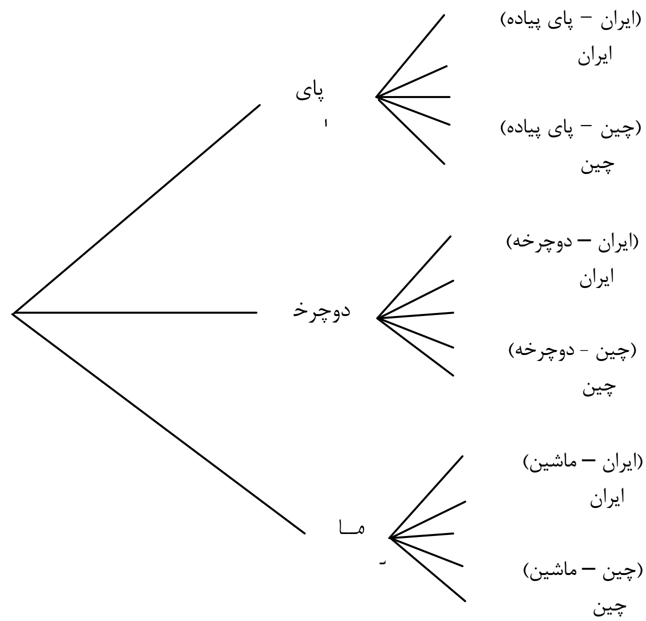
- **نمودار درختی** : این نمودار روشی است منظم برای نشان کل حالات ممکن در آزمایشاتی که عملیات در آنها طی چندین مرحله انجام می پذیرد. ( مکمل قاعده ضرب )

**مثال ۶:** جهانگردی می‌خواهد با پای پیاده یا با دوچرخه یا با ماشین به یکی از ۵ کشور ایران، چین، یونان، ترکیه یا ایتالیا سفر کند این جهانگرد به چند طریق می‌تواند این مسافرت را انجام دهد.

جواب : وی طریقهٔ مسافرت را به $n_1 = ۳$ طریق می‌تواند انتخاب کند و کشور مورد نظر را به $n_2 = ۵$ طریق انتخاب کند بنابراین تعداد طرقی که می‌تواند انتخاب کند عبارتند از :

$$n_1 \times n_2 = ۳ \times ۵ = ۱۵$$

نمودار درختی مربوط در شکل زیر رسم شده است.



پای
دوچرخ
مـا
(ایران – پای پیاده)
ایران
(چین – پای پیاده)
چین
(ایران – دوچرخه)
ایران
(چین - دوچرخه)
چین
(ایران – ماشین)
ایران
(چین – ماشین)
چین

➢ تعریف فاکتوریل: فاکتوریل که برای عددی مانند n با نماد «n!» نشان می‌دهیم عبارت است از حاصلضرب اعداد ۱ تا n، که در این صورت داریم :

۱!=۱ ۲!=۲×۱=۲ ۳!=۳×۲×۱=۶ ۴!=۴×۳×۲×۱=۲۴ ,....

و بنا به تعریف (۰!) برابر ۱ می‌باشد.

**۲- جایگشت ( ترتیب )**

یعنی تعداد طرقی که می توان $r$ شی را از بین $n$ شی انتخاب نمود بطوریکه $r \le n$ و ترتیب قرار گرفتن اشیاء نیز مهم باشد

**حالات مختلف پیدا کردن جایگشت**

۱. تعداد کل جایگشت های $n$ شی متمایز
۲. تعداد کل جایگشت های $n$ شی نامتمایز
۲. تعداد جایگشت های $r$ شی انتخابی از بین $n$ شی متمایز

**۱. تعداد کل جایگشت های n شی متمایز**

۱- در صورت ردیفی بودن بشکل $n! = n \times (n-۱) \times (n-۲) \times ... \times ۱$

۲- در صورت دایره ای بودن $(n-۱)! = (n-۱) \times (n-۲) \times ... \times ۱$

**۲. جایگشت های n شی نامتمایز**

$$\frac{n!}{n_۱! \times n_۲! \times ... n_k!}$$

که از n شی، $n_۱$ تای آنها از نوع اول، $n_۲$ تای آنها از نوع دوم و ....و $n_k$ تای آنها از نوع k ام است. و $n_۱ + n_۲ + ... + n_k = n$ است.

**۳- جایگشت های rشی از بین n شی**

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

که $r$ و $n$ هر دو متمایز و $r < n$

➢ **نکات مهم در محاسبه جایگشت ها ( ترتیب ها )**

۱- توجه به تعداد اشیاء و حجم انتخابی از بین آنها

۲- توجه به ردیفی یا دایره ای بودن اشیاء

۳- توجه به متمایز یا نامتمایز بودن اشیاء

مثال۷: به چند طریق می توان با حروف E، D ، C ، B و A یک کلمه سه حرفی ساخت؟

| | | |
|---|---|---|
| ۳ | ۴ | ۵ |

$$\frac{5!}{(5-3)!}=\frac{5!}{2!}=5\times4\times3=60$$

مثال۸: از بین ۱۵ تیم شرکت کننده در مسابقه فوتبال به چند طریق می توان سه تیم رتبه های اول، دوم و سوم را به دست آورد.

| | | |
|---|---|---|
| ۱۳ | ۱۴ | ۱۵ |

$$\frac{15!}{(15-3)!}=\frac{15!}{12!}=15\times14\times13=2730$$

مثال۹: تعداد جایگشت های حروف کلمه book را به دست آورید.

(جایگشت با n شی نامتمایز)

$$\frac{4!}{2!}=12$$

مثال۱۰: به چند طریق می توان یک کلاس ۲۰ نفری را به دسته های ۳، ۴، ۶ و ۷ نفری تقسیم کرد.

(جایگشت با n شی نامتمایز)

$$\frac{20!}{3!4!6!7!}$$

## ۳- ترکیب

تعداد طرق انتخاب $r$ شی متمایز از بین $n$ شی بشرطی که ترتیب قرار گرفتن اشیاء بی اهمیت باشد. که تعداد حالات ممکن را می توان از فرمول زیر محاسبه کرد.

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

➢ **استفاده از قاعده ضرب در ترکیب**

اگر ترکیب اول به شکل $\binom{n_1}{r_1}$ ترکیب دوم بصورت $\binom{n_2}{r_2}$ و . . . ترکیب آخر به شکل $\binom{n_K}{r_K}$ باشد

در آن صورت تعداد کل طرق برابر است با

$$\binom{n_1}{r_1}\binom{n_2}{r_2}\cdots\binom{n_K}{r_K}$$

- **ویژگی های ترکیب r شی از n شی**

۱) $\binom{n}{0}=1$      ۲) $\binom{n}{1}=n$

۳) $\binom{n}{n}=1$      ٤) $\binom{n}{n-1}=n$

مثال۱۱: از بین ۱۵ تیم شرکت کننده در مسابقه فوتبال به چند طریق می توان سه تیم را انتخاب کرد.

$$\binom{15}{3}=\frac{15!}{(15-3)!3!}=\frac{15\times14\times13}{3\times2}=455$$

مثال۱۲: به چند طریق می توان از ۱۲ کتاب که ۵ تای آن آمار و بقیه ریاضی هستند، یک کتاب آمار و ۲ کتاب ریاضی را به عنوان کتاب سال برگزید؟

$$\binom{5}{1}\binom{7}{2}=5\times21=105$$

**۴- افرازهای مرتب**

گاهی تعداد طرقی که می‌توان مجموعهٔ n شیئی را به k زیرمجموعه با $n_1$ شی در مجموعه اول $n_2$ شی در مجموعه دوم، ... و $n_k$ شی در مجموعه k افراز کرد برابر است با :

$$\binom{n}{n_1,\ n_2,\ \ldots,n_k}=\frac{n!}{n_1!\,n_2!\ldots n_k!}$$

که در آن $n_1+n_2+\ldots+n_k=n$ است.

ویژگی ها افراز :

۱- تفکیک $n$ شی به گونه ای خاص

۲- در حکم یک مجموعه بودن هر ترکیب

۳- مهم نبودن ترتیب اشیاء در هر زیر مجموعه

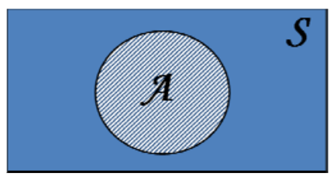
مثال۱۳: به چند طریق می توان ۹ اسباب بازی را بین ۴ بچه تقسیم کرد به شرط آنکه به کوچکترین بچه ۳ اسباب بازی و به هر کدام از بچه های دیگر ۲ اسباب بازی برسد؟

$$\binom{۹}{۳,۲,۲,۲}=\frac{۹!}{۳!۲!۲!۲!}=۷۵٦۰$$

## ❖ عملیات روی پیشامد ها و قواعد احتمال

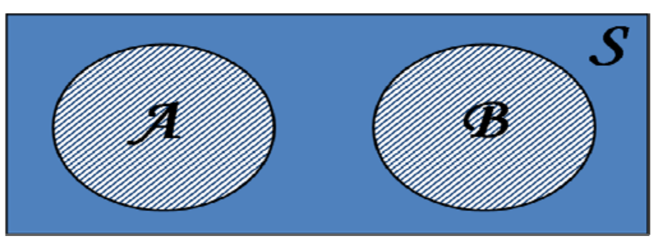
- **نمودار ون** : در این نمودار به منظور نشان دادن پیشامد ها ،کل فضای نمونه در قالب مستطیلی ارائه شده و هر پیشامدی قسمتی از این مستطیل را به خود اختصاص می دهد.

احتمال پیشامدی مانند A در نمودار ون برابر است با سطحی که پیشامد $\mathcal{A}$ از $S$ (فضای نمونه) اشغال کرده است یا بربر با نسبت تعداد اعضای $\mathcal{A}$ به $S$ است.



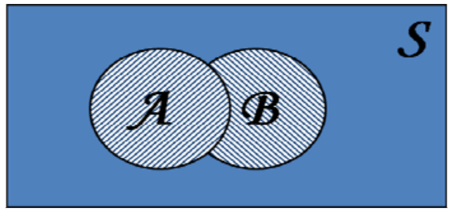

- **دو پیشامد نا سازگار** : دو پیشامد را در صورتی « نا سازگار » گویند که امکان وقوع همزمان نداشته باشند یعنی با وقوع یکی ، دیگری امکان وقوع نداشته باشد.( مثل شب و روز)

نمودار ون برای دو پیشامد نا سازگار به صورت زیر است. نمودار نشان می دهد که پیشامدهای $\mathcal{A}$ و $\mathcal{B}$ هیچ وجه اشتراکی با هم ندارند .



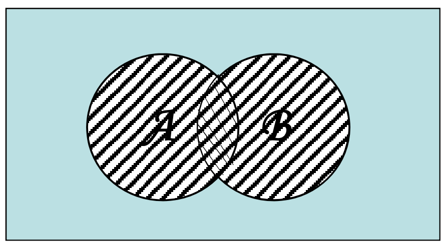

- **دو پیشامد سازگار** : دو پیشامدی را گویند که وقوع یکی مانع وقوع دیگری نیست بعبارتی این دو پیشامد دارای حداقل یک عضو مشترک هستند.

نمودار ون برای دو پیشامد سازگاربه صورت زیر است. محل تلاقی دو پیشامد ، نقطه مشترک آنهاست.



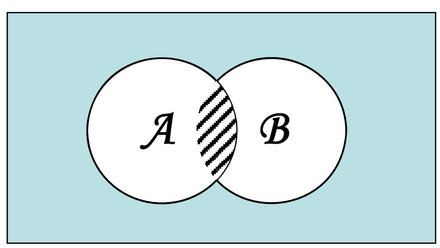

- **اجتماع دو پیشامد:** اجتماع دو پیشامدی مثل $\mathcal{A}$ و $\mathcal{B}$ ، مجموعه تمام عضوهایی است که در $\mathcal{A}$ یا در $\mathcal{B}$ یا هم در $\mathcal{A}$ و هم در $\mathcal{B}$ قرار دارند.اجتماع دو پیشامد $\mathcal{A}$ و $\mathcal{B}$ را با $A \cup B$ نشان می دهند . وقوع $A \cup B$ یعنی این که حد اقل یکی از دو پیشامد مذبور رخ داده است.

نمودار ون برای $A \cup B$ به صورت زیر است.




- **اشتراک دو پیشامد:** اشتراک دو پیشامدی مثل $\mathcal{A}$ و $\mathcal{B}$ را با $\mathcal{B} \cap \mathcal{A}$ نشان می دهند . وقوع $A \cap B$ یعنی این که هر دو پیشامد $\mathcal{A}$ و $\mathcal{B}$ رخ داده است.

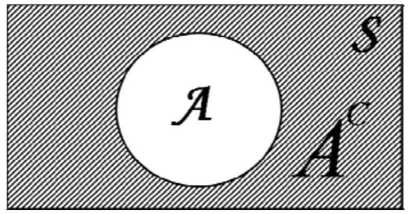
نمودار ون برای اشتراک دو پیشامدبه صورت زیر است.




یعنی این که هم پیشامد $\mathcal{A}$ و هم پیشامد $\mathcal{B}$ رخ داده است.

- **متمم یک پیشامد** : متمم پیشامدی مثل $\mathcal{A}$ که با $A^C$ نشان داده می شود مجموعه تمام عضوهایی است که در فضای نمونه است ولی در خود پیشامد $\mathcal{A}$ نیست

**نمودار ون برای $A^C$ به صورت زیر است.** وقوع متمم $\mathcal{A}$ ( $A^C$ ) به معنی عدم وقوع پیشامد $\mathcal{A}$ می باشد .




➢ برخی قواعد احتمال

۱. در صورتیکه A و B دو پیشامد از فضای نمونهٔ S باشند و A زیر مجموعهٔ B نیز باشد $(A \subset B)$ چون A دارای تعداد عضوهای کمتر یا مساوی B می‌باشد احتمال وقوع پیشامد A کوچکتر یا مساوی احتمال وقوع پیشامد B است یعنی

$$P(A) \leq P(B)$$

۲. از آنجایی که $A^c$ قسمتی از فضای نمونه است که در A نیست اجتماع A و $A^c$ برابر فضای نمونه است یعنی $A \cup A^C = S$ و مجموع احتمال آن‌ها برابر با احتمال فضای نمونه یعنی یک خواهد بود $P(A) + P(A^C) = ۱$ بنابراین :

$$P(A^C) = ۱ - P(A)$$

۳. احتمال اجتماع دو پیشامد A و B را از قاعدهٔ جمع احتمالات بدست می‌آوریم :

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

۴. در صورتی که دو پیشامد ناسازگار باشند چون دو پیشامد ناسازگار ,اشتراکی ندارند $(A \cap B = \Phi)$ و در نتیجه $P(A \cap B) = ۰$، احتمال اجتماع آنها عبارتست از:

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

مثال۱۴: اگر برای خانواده ای، احتمال دارا بودن یک دستگاه کامپیوتر، یک دستگاه لپ تاپ و یا هردو به ترتیب ۰/۸۶، ۰/۳۵ و ۰/۲۹ باشد. احتمال اینکه این خانواده یکی از دو نوع یا هر دو نوع دستگاه (حداقل یکی از دو نوع دستگاه) را داشته باشند. چقدر است؟

حل: اگر A ، پیشامد این باشد که خانواده مزبور دارای یک دستگاه کامپیوتر است. B ، پیشامد این باشد که خانواده مزبور دارای لپ تاپ است. داریم:

$$P(A)=۰/۸۶,\quad P(B)=۰/۳۵,\quad P(A\cap B)=۰/۲۹$$

بنابراین احتمال مورد نظر (حداقل یکی از دو نوع دستگاه را دارا باشند.) برابر است با:

$$P(A\cup B)=۰/۸۶+۰/۳۵-۰/۲۹=۰/۹۲$$

مثال۱۵: خودرویی در کنار خیابانی متوقف شده است. با احتمال ۰/۲۳ ، ترمزهایش معیوب است و با احتمال ۰/۲۴ فرسودگی شدید تایر دارد. همچنین با احتمال ۰/۳۸ ترمزهایش معیوب یا فرسودگی شدید دارد. احتمال اینکه این اتومبیل ترمزهایش معیوب بوده و فرسودگی شدید تایر داشته باشد. چقدر است؟ همینطور احتمال اینکه خودرو مورد نظر ترمزهایش معیوب نباشد چقدر است؟

حل: اگر B پیشامد این باشد که خودرو مورد نظر ترمزهایش معیوب باشد و T پیشامد این باشد که تایرهایش فرسوده باشد. داریم:

$$P(B)=۰/۲۳\ ,\ P(T)=۰/۲۴\ ,\ P(BUT)=۰/۳۸$$

بنابراین احتمال اینکه هم ترمزش معیوب باشد هم تایرهایش فرسوده باشند برابر است با:

$$۰/۳۸=۰/۲۳+۰/۲۴-P(B\cap T)$$
$$\Rightarrow P(B\cap T)=۰/۲۳+۰/۲۴-۰/۳۸=۰/۰۹$$

و احتمال اینکه ترمزهایش معیوب نباشد برابر است با:

$$P(B^c)=۱-P(B)=۱-۰/۲۳=۰/۷۷$$

## ❖ احتمال شرطی

اگر پیشامدی همانند A به پیشامد دیگری همانند B مربوط باشد و بدانیم پیشامد B به وقوع پیوسته است در این صورت احتمال وقوع A، به احتمال وقوع A به شرط B ($P(A|B)$) تغییر می‌یابد که آنرا احتمال شرطی می‌نامیم.

اگر A و B دو پیشامد دلخواه از فضای نمونه S باشد و $P(B)\neq ۰$ احتمال وقوع A به شرط B برابر است با :

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

مثال۱۶: دو تاس را پرتاب می کنیم در صورتی که بدانیم مجموع اعداد ظاهر شده برابر ۶ است. مطلوب است احتمال اینکه یکی از تاس ها عدد ۲ را نشان دهد.

حل: تمام حالات پرتاب دو تاس برابر است با:

$$S=\{(۱,۱),(۱,۲),\ldots,(۱,۶),\ldots,(۶,۶)\}$$

اگر B پیشامد این باشد که یکی از تاس ها عدد ۲ را نشان دهد و C پیشامد این باشد که مجموع اعداد ظاهر شده برابر ۶ است. داریم:

$$C=\{(۲,۴),(۳,۳),(۴,۲),(۵,۱),(۱,۵)\}$$
$$B\cap C=\{(۲,۴),(۴,۲)\}$$

در نتیجه احتمال مورد نظر برابر است با:

$$P(B|C)=\frac{P(B\cap C)}{P(C)}=\frac{\frac{۲}{۳۶}}{\frac{۵}{۳۶}}=\frac{۲}{۵}$$

### ❖ قانون ضرب احتمالات

با استفاده از احتمـال شرطی می توان قانـون ضرب را برای محاسبه احتمال اشتراک پیشامدها بشرح زیر بیان نمود.

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

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

- قانون ضرب احتمالات برای بیش از دو پیشامد نیز کاربرد دارد. فرمول ضرب احتمالات برای سه پیشامد این گونه است :

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

مثال۱۷: فرض کنید جعبه ای شامل ۱۰ لامپ می باشد که در بین آن ها ۴ لامپ معیوب وجود دارد. دو لامپ پشت سر هم و بدون جایگذاری استخراج می کنیم. احتمال این که هر دو لامپ معیوب باشند چقدر است ؟

حل : پیشامد اولی ناسالم :A پیشامد دومی ناسالم :B

$$p(A)=\frac{٤}{١٠},\ \ p(B|A)=\frac{٣}{٩}$$

$$\Rightarrow p(A\cap B)=p(B|A).p(A)=\frac{٣}{٩}\times\frac{٤}{١٠}=\frac{٤}{١٠}\times\frac{٣}{٩}=\frac{١٢}{٩٠}$$

حال اگر سه لامپ استخراج شود احتمال سالم بودن هر سه لامپ را حساب کنید.

حل : پیشامد اولی سالم :A پیشامد دومی سالم :B پیشامد سومی سالم:C

$$P(A\cap B\cap C)=\frac{٦}{١٠}\times\frac{٥}{٩}\times\frac{٤}{٨}$$

### ❖ دو پیشامد مستقل

دو پیشامد را « مستقل » می گوییم ، در صورتی که وقوع یا عدم وقوع یکی در وقوع و یا عدم وقوع دیگری هیچ تأثیری نداشته باشد

زمانی که دو پیشامد $\mathcal{A}$ و $\mathcal{B}$ از هم مستقل باشند هیچ تأثیری بر روی هم ندارند برای محاسبه احتمال اشتراک آنها بشکل زیر عمل می شود :

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

- شرط ناسازگار بودن دو پیشامد $A\cap B=\varphi \Rightarrow P(A\cap B)=0$
- شرط مستقل بودن دو پیشامد $P(A\cap B)=P(A)P(B)$
- پیشامدها نسبت به هم می تواند ، حالت سازگار و مستقل ، سازگار و غیر مستقل ، ناسازگار و غیر مستقل و . . . داشته باشند.

## ❖ قضیه بیز

این قضیه پژوهشگران را در تجدید نظر احتمالات ، در صورت دسترسی به اطلاعات جدید، کمک می کند فرمول این قضیه در حالتی که دو عامل مدنظر باشد به صورت زیر است..

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

در بسیاری از موارد بیش از ۲ عامل نقش تعیین کننده‌ای در وقوع پیشامدی دارند اگر $A_1$ و $A_2$، ... و $A_k$ نشان دهندهٔ K حادثهٔ ناسازگار باشند که می‌توانند حادثهٔ B را باعث شوند آنگاه احتمال اینکه $A_i$ عامل وقوع باشد برابر است با :

$$P(A_i|B)=\frac{P(A_i\cap B)}{P(B)}=\frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1)+....+P(A_i)P(B|A_i)+...+P(A_K)P(B|A_K)}$$

- **احتمالات پسین و پیشین** :به احتمال وقوع پیشامدی قبل از کسب اطلاعات جدید « احتمال پیشین » و به احتمال وقوع آن پیشامد بعد از کسب اطلاعات جدید « احتمال پسین » می گویند.

**مثال۱۸:** فرض کنید می دانیم ۳۰٪ دانشجویان چهارم ،۸۰ %دانشجویان سال سوم ،۷۰ %دانشجویان سال دوم و۵۰٪ دانشجویان سال اول از کتابخانه استفاده می کنند .اگرازهمه ی دانشجویان ، ۲۵٪ سال اول ، ۲۵٪ سال دوم ،۲۰٪ سال سوم و ۳۰٪ سال چهارم باشند ، در اینصورت

الف) احتمال اینکه دانشجویی از کتابخانه ی مرکزی استفاده کند چقدر است ؟

ب) دانشجوی سال دومی انتخاب می شود چقدر احتمال دارد از کتابخانه مرکزی استفاده کند؟

حل: تعریف می کنیم

پیشامد اینکه دانشجویی از کتابخانه ی مرکزی استفاده کند: $\mathcal{A}$

پیشامد اینکه دانشجوی سال اول باشد : $\mathcal{F}$

پیشامد اینکه دانشجوی سال دوم باشد: $O$

پیشامد اینکه دانشجوی سال سوم: $\mathcal{J}$

پیشامد اینکه دانشجوی سال چهارم باشد: $\mathcal{E}$

$$P(F)=./۲۵, P(O)=./۲۵\ ,\ P(J)=./۲\ ,\ P(E)=./۳$$
$$P(A|F)=./۵,\ P(A|O)=./۷,\ P(A|J)=./۸\ ,\ P(A|E)=./۳$$

الف)

$$\begin{aligned} p(A) &= P(A\cap F)+P(A\cap O)+P(A\cap J)+P(A\cap E) \\ &= p(A|F).p(F)+p(A|O).p(O)+p(A|J).p(J)+p(A|E).p(E) \\ &= ./۵\times./۲۵+./۷\times./۲۵+./۸\times./۲+./۳\times./۳=./۵۵ \end{aligned}$$

ب)

$$P(O|A)=\frac{P(A|O)P(O)}{P(A)}=\frac{./۷۵\times./۲۵}{./۵۵}=./۳۳$$

مثال۱۹: سه ماشین C,B,A به ترتیب ۶۰ درصد، ۳۰ درصد و ۱۰ درصد کل محصولات کارخانه‌ای را تولید می‌کنند درصد محصولات معیوب این ماشین‌ها به ترتیب برابر ۲ درصد، ۳ درصد، ۴ درصد است از میان محصولات این کارخانه محصولی به صورت تصادفی انتخاب می‌کنیم می‌خواهیم هر یک از این احتمالات را محاسبه کنیم :

الف : احتمال اینکه معیوب باشد.

ب : احتمال اینکه با ماشین C تولید شده باشد در صورتیکه بدانیم معیوب است.

حل : احتمال معیوب بودن : X

$$P(A)=۰/٦\,,\ P(B)=۰/۲\,,\ P(C)=۰/۱$$
$$P(X\mid A)=۰/۰۲\,,\ P(X\mid B)=۰/۰۳\,,\ P(X\mid C)=۰/۰٤$$

الف :

$$P(X)=P(A)P(X/A)+P(B)P(X/B)+P(C)P(X/C)$$
$$=۰/۶۰\times۰/۰۲+۰/۰۳\times۰/۰۳+۰/۱۰\times۰/۰۴=۰/۰۲۵$$

ب :

$$P(C\mid X)=\frac{P(C)P(X\mid C)}{P(A)P(X\mid A)+P(B)P(X\mid B)+P(C)P(X\mid C)}=$$
$$\frac{۰/۱۰\times۰/۰٤}{(۰/٦۰\times۰/۰۲)+(۰/۰۳\times۰/۰۳)+(۰/۱۰\times۰/۰٤)}=۰/۱٦$$

# فصل ششم: توابع احتمال گسسته

هدف این فصل آشناسازی دانشجویان با متغیرهای تصادفی گسسته ، توابع احتمال و توزیع های مربوط به آنهاست.

## ❖ متغیر تصادفی گسسته، تابع احتمال و تابع توزیع

### ▪ متغیر تصادفی گسسته

➢ **متغیر تصادفی:** تابعی است که روی فضای نمونه تعریف می شود و هر یک از مقادیر آن، متناظر با یک یا چند عضو از اعضای فضای نمونه است. با توجه به این که هر تابع دارای دامنه و حوزه می باشد دامنه یک متغیر تصادفی نیز فضای نمونه ( S ) و حوزه اش مجموعه اعداد حقیقی است.

مثال۱: فرض کنید که X قانونی باشد که اگر سکه ای پرتاب شود و شیر بیاید ، X را برابر یک تعریف کنیم و اگر خط بیاید X را برابر صفر تعریف کنیم.

$$X : \{H, T\} \rightarrow \{۱, ۰\} \Rightarrow X : S \rightarrow \Re$$

➢ **انواع متغیر تصادفی**

۱- متغیر تصادفی گسسته ؛ با تعداد مقادیر متناهی یا شمارش پذیر

۲- متغـیر تصادفی پیوسـته ؛ با تعداد مقادیر ممـکن نامتناهی و غیر قابل شمارش

➢ متغیرهای تصادفی را با حروف بزرگ لاتین مثل Z و Y و X و هر یک از مقادیر انتخابی آنها را با حروف کوچک z و y و x نشان می دهند

### ▪ تابع احتمال

به تابعی که بتوان با استفاده از آن احتمال هر یک از مقادیر ممکن متغیر تصادفی را مشخص کرد « تابع احتمال » یا « توزیع احتمال » گویند. تابع احتمال تابعی است که دامنه آن مقادیر ممکن متغیر تصادفی و حوزه آن احتمالات مربوط به هر مقدار از متغیر تصادفی است.

در حالت گسسته احتمال ها بوسیله ی تابعی به نام تابع احتمال معرفی می شوند که با نماد $p(X=x)$ نشان داده می شود. بنابر اصول موضوع احتمال، $p(X=x)$ وقتی یک تابع احتمال است که :

$P(X=x)$ در حوزه $\forall x \in X$

۱- $P(X=x) \geq .$

۲- $\sum_{x} P(X=x) = ۱$

مثال۲: تعداد فروش کت و شلوار فروشگاه لباسی در هر روز، همراه با احتمال آن، در جدول زیر ارائه شده است. آیا می توان گفت احتمال های ارائه شده یک تابع احتمال است ؟

| $X=x$ | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ |
|---|---|---|---|---|---|---|---|---|
| $p(X=x)$ | ۰/۱ | ۰/۱۵ | ۰/۱۶ | ۰/۲۰ | ۰/۱۹ | ۰/۰۸ | ۰/۰۷ | ۰/۰۵ |

حل :

۱)داریم:

$$P(X=۱)=./۱,\ P(X=۲)=./۱۵,\ p(X=۳)=./۱۶,\ P(X=۴)=./۲.$$
$$P(X=۵)=./۱۹, P(X=۶)=./.۸,\ P(X=۷)=./.۷,\ P(X=۸)=./.۵$$

پس اصل اول برقرار است ( $P(X=x) \geq ., \forall x$ )

۲)

$$\sum_x P(X=x)=۱ \Rightarrow ./۱+./۱۵+./۱۶+./۲.+./۱۹+./.۸+./.۷+./.۵=۱$$

بنابراین ، اصول موضوع برقرار است و پاسخ مثبت است.

## ▪ تابع توزیع ( تابع احتمال تجمعی )

تابع توزیع ، تابعی است که به ازای جمیع مقادیر ممکن متغیر تصادفی X ، احتمال وقوع مقداری کوچکتر یا مساوی با X را نشان می دهد. یعنی

$$F(x)=P(X\le x)$$

مثال۳: در مورد مثال ۲ تابع توزیع را محاسبه کنید.

**حل:**

$$F(۱)=P(X\le ۱)=./۱$$
$$F(۲)=P(X\le ۲)=P(X=۱)+P(X=۲)=./۱+./۱۵=./۲۵$$
$$F(۳)=P(X\le ۳)=P(X=۱)+P(X=۲)+P(X=۳)=./۱+./۱۵+./۱۶=./۴۱$$

| $X=x$ | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ |
|---|---|---|---|---|---|---|---|---|
| $p(X=x)$ | ۰/۱ | ۰/۲۵ | ۰/۴۱ | ۰/۶۱ | ۰/۸۰ | ۰/۸۸ | ۰/۹۵ | ۱/۰ |

## ❖ امید ریاضی و واریانس متغیر تصادفی

### ▪ امید ریاضی

امید ریاضی متغیر تصادفی X که با $E(X)$ نشان داده می شود همان میانگین موزون است که احتمالات در آن ، نقش ضرایب ( وزن ها ) را ایفاء می کنند.

امید ریاضی یک متغیر تصادفی از حاصل جمع ضرب هر متغیر تصادفی در مقدار احتمال خودش بدست می آید. که عبارت فوق را می توان به صورت زیر بیان کرد.

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

در عبارت فوق $f(x)$ همان $P(X=x)$ است.

- امید ریاضی را میانگین X نیز می نامیم و با $\mu_x$ یا $\mu$ نیز نشان می دهیم.
- ویژگی های امید ریاضی

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

مثال۴: برای مشاهدات مثال۲، امید ریاضی X را محاسبه کنید.

$$E(X)=۱\times./۱+۲\times./۱۵+۳\times./۱٦+٤\times./۲+٥\times./۱۹+٦\times./.۸+۷\times./.۷+۸\times./.٥=٤$$

و این مقدار به این معنی است که فروشگاه انتظار دارد به طور متوسط ۴ کت وشلوار در روز به فروش برساند.

### ▪ واریانس متغیر تصادفی X

واریانس را با نماد V(X) نشان داده و میزان پراکندگی را حول میانگین ( امید ریاضی ) نشان می دهد.

برای محاسبه واریانس می توان از دو فرمول زیر استفاده کرد.

$$۱-V(X)=\sum(X-\mu)^۲ f(X)$$
$$۲-V(X)=E(X^۲)-\mu^۲$$

که $\mu=E(X)$ و $f(x)=P(X=x)$ است.

- مانند قبل از جذر واریانس انحراف معیار محاسبه می شود. یعنی $SD(X)=\sqrt{V(X)}$

**مثال۵: برای مشاهدات مثال۲ واریانس و انحراف معیار را محاسبه کنید.**

$$\begin{aligned}V(X)&=\sum(X-\mu)^۲ f(X)\\&=(۱-٤)^۲\times ./۱+(۲-٤)^۲\times ./۱۵+(۳-٤)^۲\times ./۱٦+(٤-٤)^۲\times ./۲.\\&+(٥-٤)^۲\times ./۱۹+(٦-٤)^۲\times ./.۸+(۷-٤)^۲\times ./.۷+(۸-٤)^۲\times ./.٥\\&=۳/٦\end{aligned}$$

$$V(X)=E(X^۲)+\mu^۲=۱۹/٦-۱٦=۳/٦$$

$$\begin{aligned}E(X^۲)&=۱^۲\times ./۱+۲^۲\times ./۱۵+۳^۲\times ./۱٦+٤^۲\times ./۲.\\&+٥^۲\times ./۱۹+٦^۲\times ./.۸+۷^۲\times ./.۷+۸^۲\times ./.٥=۱۹/٦\end{aligned}$$

$$SD(X)=\sqrt{V(X)}=\sqrt{۳/٦}=۱/۸۳$$

- **ویژگی های واریانس**

$۱)V(b)=.$

$۲)V(aX)=a^۲V(X)$

$۳)V(aX+b)=a^۲V(X)$

### ❖ تابع احتمال توأم

تابع احتمال توأم عبارتست از فهرستی از زوج های $(X_i,Y_j)$ و احتمال های متناظر با آنها، یعنی $f(X_i,Y_j)=P(X_i=x_i,Y_j=y_j)$.

## ➢ موارد استفاده تابع احتمال توأم

هر گاه پژوهشگر بخواهد رفتار متغیری مثل X را در ارتباط با رفتار متغیر دیگری مثل Y بررسی نماید، از این تابع استفاده می کند.

مثال ۶: دو فروشگاه لوازم خانگی را در نظر بگیرید فرض کنید X تعداد یخچال‌های فریز فروخته شده در فروشگاه الف و Y تعداد یخچال‌های فروخته شدهٔ فروشگاه ب است نمودار زیر نشان‌دهندهٔ احتمال توام در فروشگاه در هفتهٔ گذشته است که تابع احتمال توام آنها در جدول زیر آمده است.

الف) $P(X=۱)$

ب) $P(X>Y)$

ج) $P(Y\geq ۰\,,\,X=۱)$

د) توزیع احتمال $Z=X+Y$

| X \ Y | ۰ | ۱ |
|---|---|---|
| ۰ | ۰/۰۵ | ۰/۱۸ |
| ۱ | ۰/۲۲ | ۰/۳۵ |
| ۲ | ۰/۱۵ | ۰/۰۵ |

الف) $P(X=۱)=P(X=۱\,,\,Y=۰)+P(X=۱\,,Y=۱)=۰/۲۲+۰/۳۵=۰/۵۷$

ب) $P(X>Y)=P(X=۱\,,\,Y=۰)+P(X=۲\,,Y=۰)+P(X=۲,Y=۱)=۰/۲۲+۰/۱۵+۰/۰۵=۰/۴۲$

ج) $P(Y \geq ۰ ,\ X=۱) = P(Y=۰ , X=۱) + P(Y=۱ , X=۱) = ۰/۲۲ + ۰/۳۵ = ۰/۵۷$

د)

| $Z=X+Y$ | ۰ | ۱ | ۲ | ۳ |
|---|---|---|---|---|
| $f(z)$ | ۰/۰۵ | ۰/۴۰ | ۰/۵۰ | ۰/۰۵ |

## ➢ احتمالات حاشیه ای

۱– احتمالات حاشیه ای X : برای پیدا کردن تابع احتمال متغیر تصادفی X

۲– احتمالات حاشیه ای Y : برای پیدا کردن تابع احتمال متغیر تصادفی Y

با در دست داشتن تابع احتمال توام می‌توان احتمال جداگانه هر متغیرتصادفی را پیدا کرد. در مثال قبل تابع احتمال متغیرتصادفی X را می‌توان از جمع کردن احتمالات هر سطر و نوشتن آن در حاشیهٔ سمت راست جدول بدست آورد و تابع احتمال متغیرتصادفی Y را از جمع احتمالات هر ستون و نوشتن آنها در حاشیهٔ پائین جدول بدست آورد به این احتمالات «احتمالات حاشیه‌ای» می‌گوئیم.

مثال ۷: با توجه به جدول احتمال توأم مثال قبل احتمالات حاشیه‌ای را بنویسید و تابع احتمال X و Y را در دو جدول مجزا بنویسید.

حل :

| Y \ X | ۰ | ۱ | احتمالات حاشیه‌ای x |
|---|---|---|---|
| ۰ | ۰/۰۵ | ۰/۱۸ | ۰/۲۳ |
| ۱ | ۰/۲۲ | ۰/۳۵ | ۰/۵۷ |
| ۲ | ۰/۱۵ | ۰/۰۵ | ۰/۲۰ |
| احتمالات حاشیه‌ای Y | ۰/۴۲ | ۰/۵۸ | ۱ |

پس تابع احتمال هر یک از X و Y برابر است با :

| X | ۰ | ۱ | ۲ |
|---|---|---|---|
| $f(X)$ | ۰/۲۳ | ۰/۵۷ | ۰/۲۰ |

| Y | ۰ | ۱ |
|---|---|---|
| $f(Y)$ | ۰/۴۲ | ۰/۵۸ |

## ❖ کوواریانس و استقلال دو متغیر تصادفی

### ▪ کوواریانس

معیار عددی است که نوع و شدت رابطه خطی بین دو متغیر تصادفی را نشان می دهد و عبارتست از امید ریاضی تغییرات دو متغیر بر حسب میانگین شان.

برای محاسبه کوواریانس می توان از دو فرمول زیر استفاده کرد.

$$۱- COV(X,Y) = E(X-\mu_X)(Y-\mu_Y)$$
$$۲- COV(X,Y) = E(XY) - E(X)E(Y)$$

در صورت گسسته بودن x و y
$$E(XY) = \sum_i \sum_j x_i y_j f(x_i y_j)$$

## ➢ مقادیر مختلف کواریانس

۱- کوواریانس مثبت : نشان‌دهندهٔ رابطهٔ مستقیم دو متغیراست یعنی با افزایش یکی دیگری نیز افزایش می‌یابد.

۲- کوواریانس منفی : که نشان دهندهٔ رابطهٔ معکوس دو متغیراست یعنی با افزایش یکی دیگری کاهش می‌یابد.

۳- کوواریانس صفر : که بازگو گنندهٔ استقلال دو متغیراست یعنی هیچ کدام بر یکدیگر تأثیری ندارند.

مثال ۸: در مثال ۶ کواریانس بین فروش دو فروشگاه را محاسبه کنید. در مورد نوع رابطهٔ دو متغیربحث کنید.

حل:

$$E(X) = \sum_x x f(x) = ۰\times۰/۲۳ + ۱\times۰/۵۷ + ۲\times۰/۲۰ = ۰/۹۷$$

$$E(Y) = \sum_y y f(y) = ۰\times۰/۴۲ + ۱\times۰/۵۸ = ۰/۵۸$$

$$E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j) = (۰\times۰\times۰/۰۵)$$
$$+(۰\times۱\times۰/۱۸) + (۱\times۰\times۰/۲۲) + (۱\times۱\times۰/۳۵) + (۲\times۰\times۰/۱۵)$$
$$+(۲\times۱\times۰/۰۵) = ۰/۳۵ + ۰/۱ = ۰/۴۵$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = ۰/۴۵ - (۰/۹۷\times۰/۵۸) = -۰/۱۱۲۶$$

$$Cov(XY)=E(X-\mu_x)(Y-\mu_y)$$
$$=(۰-۰/۹۷)\times(۰-۰/۵۸)\times۰/۰۵+(۰-۰/۹۷)\times(۱-۰/۵۸)\times۰/۱۸$$
$$+(۱-۰/۹۷)\times(۰-۰/۵۸)\times۰/۲۲+(۱-۰/۹۷)\times(۱-۰/۵۸)\times۰/۳۵$$
$$+(۲-۰/۹۷)\times(۰-۰/۵۸)\times۰/۱۵+(۲-۰/۹۷)\times(۱-۰/۵۸)\times۰/۰۵=-۰/۱۱۲۶$$

با توجه به منفی بودن مقدار کوواریانس رابطه بین دو متغیر تصادفی X و Y رابطه معکوس است.

- **استقلال دو متغیر تصادفی**

دو متغیر تصادفی X و Y در صورتی مستقل اند که به ازای تمام زوج های $(X_i, Y_j)$ رابطه روبرو برقرار باشد. $f(x_i, y_j)=f(x_i)\times f(y_j)$

مثال ۹: استقلال دو متغیر X و Y در مثال ۶ را بررسی کنید.

$$f(x_۱, y_۱)=p(X=۰,Y=۰)=۰/۰۵\neq P(x=۰)\times P(Y=۰)=۰/۲۳\times۰/۴۲$$

بنابراین دو متغیر مستقل نیستند.

➢ **رابطه استقلال و کواریانس دو متغیر**

اگر x و y مستقل باشند ، کواریانشان حتماً صفر است ولی عکس قضیه همیشه صادق نیست. به این معنی که اگر کواریانس بین دو متغیر صفر باشد دو متغیر لزوما مستقل نیستند.

❖ **معرفی چند توزیع گسسته**

- **توزیع برنولی**

➢ ویژگی های پیشامدهایی که دارای نوزیع برنولی هستند عبارت است :

۱- آزمایش فقط یک بار صورت می گیرد.

۲- فقط دو پیامد ممکن دارد. (موفقیت و شکست)

۳- احتمال موفقیت و شکست ثابت است. ( البته در صورت تکرار آزمایش )

۴- آزمایش ها مستقل از یکدیگر انجام می شوند.

متغیر مورد نظر: موفقیت یا شکست در آزمایش انجام شده.

- **مفاهیم p و q درتوزیع برنولی**

- P یعنی احتمال موفقیت ( احتمال وقوع پیشامد مورد نظر )
- q یعنی احتمال شکست ( احتمال عدم وقوع پیشامد مورد نظر )

p و q مکمل یکدیگر هستند یعنی $p+q=۱ \Rightarrow q=۱-p$

- **سوال: آیا پیشامد استخراج شده از نمونه گیری بدون جایگزاری دارای توزیع برنولی است یا خیر؟**

**پاسخ:** نمونه گیری های بدون جایگزینی از جامعه باعث می شود احتمال موفقیت و شکست در آزمایش ها تغییر کرده و شرط سوم از ویژگی های برنولی نقض گردد و توزیع را از حالت برنولی خارج شود. البته لازم به ذکر است نمونه گیری بدون جایگزینی از جوامع خیلی بزرگ ، می تواند به برنولی بودن توزیع کمک نماید. (مثال: در بانکی ۵۰۰۰ نفر دارای حساب بانکی هستند که ۲۰۰۰ نفر آنها دارای حساب کوتاه مدت هستند. اگر موفقیت را داشتن حساب کوتاه مدت تعریف کنیم. پیشامد انتخاب یک نفر از این ۵۰۰۰ نفر دارای توزیع برنولی است. چرا که تمام ویژگی های توزیع برنولی را داشته و در مورد شرط سوم می توان گفت احتمال موفقیت فرد اول $\frac{۲۰۰۰}{۵۰۰۰}$ و احتمال موفقیت فرا دوم $\frac{۱۹۹۹}{۴۹۹۹}$ همانطور که ملاحظه می شود این دو احتمال تفاوت چندانی با هم ندارند. پس شرط سوم نیز برقرار است.)

پس به طور کلی می توان گفت نمونه گیری با جایگزینی از کلیه جوامع آماری و نمونه گیری بدون جایگزینی صرفاً از جوامع خیلی بزرگ ، می تواند به برنولی بودن توزیع کمک نماید.

- میانگین و واریانس توزیع برنولی

$$\mu_x = E(X) = p$$
$$\sigma_x^2 = V(X) = pq$$

مثال۱۰: اگر در ریختن یک تاس سالم آمدن عدد ۴ یا ۶ را موفقیت بنامیم و بقیه حالت ها را شکست به حساب آوریم ، میانگین و واریانس توزیع را حساب کنید.

حل : ابتدا ویژگی های توزیع برنولی را در مورد آن بررسی می کنیم.

۱- آزمایش فقط یک بار صورت می گیرد. ☺

۲- فقط دو پیامد ممکن دارد. (موفقیت و شکست) ☺

۳- احتمال موفقیت و شکست ثابت است. ( البته در صورت تکرار آزمایش ) ☺

۴- آزمایش ها مستقل از یکدیگر انجام می شوند. ☺

بنابراین آزمایش انجام شده دارای توزیع برنولی بوده که احتمال موفقیت و شکست برابر است با

$$p = P(X = 1) = P(T = 4 \text{ or } T = 6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$
$$\Rightarrow q = P(X = 0) = 1 - p = \frac{2}{3}$$

**در نتیجه**

$$\mu = E(X) = p = \frac{1}{3}$$
$$\sigma_x^2 = V(X) = pq = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$$

- **توزیع دو جمله ای**

ویژگیها:

۱- تکرار آزمایش ( n بار )

۲- هر آزمایشی فقط دو پیامد دارد

۳- ثابت بودن p و q در هر آزمایش

۴- مستقل بودن آزمایش ها از همدیگر

متغیر مورد نظر:تعداد موفقیت در n آزمایش انجام شده.

در صورت برقراری شروط فوق آنگاه X تعداد موفقیت ها در n تکرار آزمایش دارای توزیع دوجمله ای است که با نماد $X \sim binomial(n,p)$ نشان داده می شود. و فرمول توزیع دو جمله ای به صورت زیر بیان می شود.

$$P(X=x)=\binom{n}{x}p^{x}q^{n-x}$$
$$x=۰,۱,۲,...,n$$

اجزاء تشکیل دهنده توزیع دو جمله ای

n = تعداد آزمایش ها

x = تعداد موفقیت های مورد نظر

p = احتمال موفقیت در هر آزمایش

q = احتمال شکست در هر آزمایش

## ➢ میانگین و واریانس توزیع دو جمله ای

$$\mu_x = E(X) = np$$
$$\sigma_x^۲ = V(X) = npq$$

n و p و q پارامترهای توزیع دو جمله ای هستند.

مثال ۱۱: رستورانی ۸ نوع خوراک ماهی ، ۱۲ نوع خوراک گوشت و۱۰ نوع خوراک مرغ درست می کند . اگر مشتریان این رستوران خوراک ها را به تصادف انتخاب کنند.

الف) احتمال اینکه دونفر ازچهار مشتری بعدی خوراک ماهی سفارش دهند چقدر است ؟

ب) احتمال اینکه حداکثر دو مشتری از چهار مشتری خوراک ماهی سفارش دهند چقدر است؟

ج) میانگین و واریانس تعداد افرادی که خوراک ماهی سفارش می دهند چقدر است؟

حل: ابتدا توزیع متغیر مورد بررسی را پیدا می کنیم. با توجه به ساختار آزمایش مورد نظر به نظر می رسد متغیر مورد نظر (تعداد مشتریانی که خوراک ماهی سفارش می دهند.) دارای توزیع دوجمله ای باشد به همین دلیل شرایط توزیع دو جمله ای بررسی می شود.

۱- تکرار آزمایش ( n بار ) ☺ (۴ مشتری بررسی شده اند.)

۲- هر آزمایشی فقط دو پیامد دارد. ☺ (مشتری یا خوراک ماهی سفارش می دهد یا خوراکی غیر ازماهی سفارش می دهد.)

۳- ثابت بودن p و q در هر آزمایش ☺ (در صورت مسئله فرض شده مشتریان این رستوران خوراک ها را به تصادف انتخاب می کنند. پس p و q ثابت است.)

۴- مستقل بودن آزمایش ها از همدیگر ☺ (نظر هر مشتری مستقل از نظر مشتری دیگر است.)

پس متغیر آزمایش X دارای توزیع دو جمله ای با پارامترهای ۴ و $p = \frac{۸}{۳۰}$ است.

الف) $$P(X=۲)=\binom{۴}{۲}\left(\frac{۸}{۳۰}\right)^{۲}\left(\frac{۲۲}{۳۰}\right)^{۴-۲}=\frac{۴!}{۲!(۴-۲)!}\left(\frac{۸}{۳۰}\right)^{۲}\left(\frac{۲۲}{۳۰}\right)^{۲}=۰/۱۱۵$$

ب) $$P(X\le ۲)=P(X=۰)+(X=۱)+P(X=۲)$$
$$=\binom{۴}{۰}\left(\frac{۸}{۳۰}\right)^{۰}\left(\frac{۲۲}{۳۰}\right)^{۴-۰}+\binom{۴}{۱}\left(\frac{۸}{۳۰}\right)^{۱}\left(\frac{۲۲}{۳۰}\right)^{۴-۱}+\binom{۴}{۲}\left(\frac{۸}{۳۰}\right)^{۲}\left(\frac{۲۲}{۳۰}\right)^{۴-۲}$$

➢ وقتی که n نسبتاً بزرگ است ، محاسبه احتمال از طریق فرمول کار خسته کننده ای می شود ، لذا برای رفع این مشکل از جدول های مخصوصی استفاده می شود. در این کتاب این جدول در پیوست (۱) ارائه شده است. در این جدول احتمال $P(X\le c)=\sum_{x=۰}^{c}\binom{n}{x}p^{x}q^{n-x}$ برای مقادیر مختلف n ، p و c آورده شده است.

در نتیجه با استفاده از جدول بیان شده و برای $n=۴$ ، $c=۲$ و $p=۰/۲۷\approx ۰/۳۰$ داریم:

$$P(X\le ۲)=۰/۹۱۶$$

ج) $$E(X)=np=۴\times\frac{۸}{۳۰}=\frac{۳۲}{۳۰}=۱/۰۷$$

**این مقدار ریاضی بیانگر ان است که انتظار می رود به طور متوسط ۱/۰۷ نفر خوراک ماهی سفارش دهند.**

$$V(X)=npq=۴\times\frac{۸}{۳۰}\times\frac{۲۲}{۳۰}=۰/۷۸$$

➢ **مقدار p و نوع توزیع**

۱- اگر p = ۰/۵ باشد، توزیع متقارن

۲- اگر $p>۰/۵$ باشد ، توزیع چوله به چپ

۳- و اگر $p<۰/۵$ باشد ، توزیع چوله به راست است

- **توزیع هندسی**

ویژگیها:

۱- تکرار آزمایش

۲- هر آزمایشی فقط دو پیامد دارد

۳- ثابت بودن p و q در هر آزمایش

۴- مستقل بودن آزمایش ها از همدیگر

متغیر مورد نظر: تعداد آزمایش تارسیدن به اولین موفقیت.

در صورت برقراری شروط فوق آنگاه X تعداد آزمایش تارسیدن به اولین موفقیت دارای توزیع هندسی است که با نماد $X \sim Geometric(p)$ نشان داده می شود. و فرمول توزیع هندسی به صورت زیر بیان می شود.

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

اجزاء تشکیل دهنده توزیع هندسی

x = تعداد آزمایش های انجام شده تا رسیدن به اولین موفقیت

p = احتمال موفقیت در هر آزمایش

q = احتمال شکست در هر آزمایش

- **میانگین و واریانس توزیع هندسی**

$$\mu_x = E(X) = \frac{۱}{p}$$

$$\sigma_x^۲ = V(X) = \frac{q}{p^۲}$$

p و q پارامترهای توزیع هندسی هستند.

مثال۱۲: : اگر احتمال قبولی در یک امتحان رانندگی که شخصی هر بار شرکت میکند ۷۵٪ باشد، احتمال اینکه این شخص سرانجام در چهارمین بار قبول شود ، چند است ؟ میانگین و واریانس متغیر مورد نظر را پیدا کنید.

حل:

$$X = ٤ ,\ p = ./ ۷۵$$

$$\Rightarrow P(X = ٤) = q^{x-۱} p = ./ ۲۵^{٤-۱} \times ./ ۷۵ = ./ .۱۲$$

$$E(X) = \frac{۱}{p} = \frac{۱}{./ ۷۵} = ۱/ ۳۳$$

$$V(X) = \frac{q}{p^۲} = \frac{./ ۲۵}{./ ۷۵^۲} = ./ ٤٤$$

- **توزیع دوجمله ای منفی**

ویژگیها:

۱- تکرار آزمایش

۲- هر آزمایشی فقط دو پیامد دارد

۳- ثابت بودن p و q در هر آزمایش

۴- مستقل بودن آزمایش ها از همدیگر

متغیر مورد نظر: تعداد آزمایش تارسیدن به k امین موفقیت.

در صورت برقراری شروط فوق آنگاه X تعداد آزمایش تارسیدن به k امین موفقیت دارای توزیع دوجمله ای منفی است که با نماد $X \sim NB(k,p)$ نشان داده می شود. و فرمول توزیع دوجمله ای منفی به صورت زیر بیان می شود.

$$P(X=x)=\binom{x-۱}{k-۱}p^k q^{x-k} \quad , \qquad x=k,k+۱,k+۲,...$$

اجزاء تشکیل دهنده توزیع دوجمله ای منفی

x = تعداد آزمایش های انجام شده تا رسیدن بهk امین موفقیت

p = احتمال موفقیت در هر آزمایش

q = احتمال شکست در هر آزمایش

- **میانگین و واریانس توزیع دوجمله ای منفی**

$$\mu_x = E(X) = \frac{k}{p}$$
$$\sigma_x^۲ = V(X) = \frac{kq}{p^۲}$$

p و q پارامترهای توزیع دو جمله ای منفی هستند.

- در صورتی که k برابر ۱ باشد توزیع دوجمله ای منفی همان توزیع هندسی است.

مثال۱۳: اگر شخصی در معرض ابتلا به یک بیماری مسری قرار داشته باشد، با احتمال ۴۰٪ به آن دچار می شود. احتمال اینکه ۱۰ امین شخص در معرض بیماری ۳ امین شخصی باشد که به آن مبتلا می شود، چقدر است.

حل:

$$X = ۱۰,\; k = ۳,\; p = ۰/۴۰$$

$$\Rightarrow P(X = ۱۰) = \binom{۱۰-۱}{۳-۱} p^{۳} q^{۱۰-۳} = ۳۶ \times ۰/۴۰^{۳} \times ۰/۶^{۷} = ۰/۰۶۴$$

$$E(X) = \frac{k}{p} = \frac{۳}{۰/۴۰} = ۷/۵$$

$$V(X) = \frac{kq}{p^{۲}} = \frac{۳ \times ۰/۶}{۰/۴^{۲}} = ۱۱/۲۵$$

مفهوم امید ریاضی در این حالت به این صورت است که انتظار می رود به طور متوسط ۷/۵ امین نفر سومین نفری باشد که به بیماری مبتلا می شود.

## ▪ توزیع چند جمله ای

ویژگیها:

۱- تکرار آزمایش ( n بار )

۲- هر آزمایشی بیش از دو پیامد دارد

۳-احتمال هر پیامد در آزمایش های مختلف ثابت است.( ثابت بودن $p_1, p_2, ..., p_k$ در هر آزمایش)

۴- مستقل بودن آزمایش ها از همدیگر

متغیرهای مورد نظر:تعداد هر یک از پیامد ها $X_1, X_2, ..., X_k$ در n آزمایش انجام شده.

در صورت برقراری شروط فوق آنگاه تعداد $X_1, X_2, ..., X_k$ در n تکرار آزمایش دارای توزیع چندجمله ای است که با نماد $X \sim Multinomial(n, p)$ نشان داده می شود. و فرمول توزیع چندجمله ای به صورت زیر بیان می شود.

$$P(X_1=x_1, X_2=x_2, \ldots, X_k=x_k)=\binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$
$$x_1+x_2+\ldots+x_k=n$$

اجزاء تشکیل دهنده توزیع چند جمله ای

تعداد آزمایش ها = n

تعداد مورد نظر در هر پیامد: $x_1, x_2, \ldots, x_k$

احتمال پیامد های مختلف در هر آزمایش = $p_1, p_2, \ldots, p_k$

مثال ۱۴: رستورانی ۸ نوع خوراک ماهی ، ۱۲ نوع خوراک گوشت و۱۰ نوع خوراک مرغ درست می کند . اگر مشتریان این رستوران خوراک ها را به تصادف انتخاب کنند.

الف) احتمال اینکه دونفر ازچهار مشتری بعدی خوراک ماهی و یک نفر خوراک گوشت و یک نفر خوراک مرغ را سفارش دهند چقدر است ؟

حل: در صورتی که $X_1$ را تعداد افرادی که خوراک ماهی، $X_2$ را تعداد افرادی که خوراک گوشت و $X_3$ را تعداد افرادی که خوراک مرغ سفارش می دهند در نظر بگیریم داریم:

$$X_1=2, X_2=1, X_3=1 \qquad , \qquad p_1=\frac{8}{30}, p_2=\frac{12}{30}, p_3=\frac{10}{30}$$
$$P(X_1=2, X_2=1, X_3=1)=\binom{4}{2,1,1}\left(\frac{8}{30}\right)^2\left(\frac{12}{30}\right)^1\left(\frac{10}{30}\right)^1=\frac{4!}{2!1!1!}\left(\frac{8}{30}\right)^2\left(\frac{12}{30}\right)^1\left(\frac{10}{30}\right)^1=0/21$$

- **توزیع فوق هندسی**
- **ویژگی های نوزیع فوق هندسی:**

حالتی را در نظر می گیریم که از جامعه ای با دو حالت ( موفقیت و شکست ) نمونه ای تصادفی بدون جایگذاری داشته باشیم .فرض می کنیم از جامعه ای با اندازه N با دو حالت موفقیت ( سالم بودن ) که به تعداد k در جامعه هستند و شکست ( معیوب بودن ) که به تعداد $N-k$ در جامعه هستند، داشته باشیم .از این جامعه نمونه ای به اندازه n انتخاب می کنیم .فرض می کنیم در این نمونه : X تا از نوع k و بقیه n-X از نوع N-k می باشند.

متغیر مورد نظر(X) : تعداد واحدهایی از نمونه که دارای ویژگی مورد نظر هستند. یعنی از k تا مورد نظر انتخاب شده اند.

در صورت برقراری شروط فوق آنگاه X تعداد واحدهای نمونه که دارای ویژگی مورد نظر هستند دارای توزیع فوق هندسی است. و فرمول توزیع فوق هندسی به صورت زیر بیان می شود.

$$P(X=x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad , x=۰,۱,...,k$$

اجزاء تشکیل دهنده توزیع فوق هندسی

x = تعداد واحدهایی از نمونه که دارای ویژگی مورد نظر هستند

K =.تعداد واحد های جامعه که دارای ویژگی مورد نظر هستند

N =تعداد کل جامعه

n=تعداد کل نمونه

- **میانگین و واریانس توزیع فوق هندسی**

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

$$\sigma_x^{۲} = V(X) = \frac{nk(N-k)(N-n)}{N^{۲}(N-۱)}$$

مثال۱۵: فرض می کنیم از جعبه ای شامل ۶ لامپ که ۳ تای آن ها سوخته است ، ۴ لامپ به تصادف و بدون جایگذاری انتخاب می کنیم احتمال اینکه بین این ۴ لامپ ، ۲ لامپ سوخته وجود داشته باشد چقدر است ؟

حل :

$$N = ۶, k = ۳ \Rightarrow N - k = ۳$$
$$n = ۴, x = ۲ \Rightarrow n - x = ۴ - ۲ = ۲$$

$$p(x = ۲) = \frac{\binom{۳}{۲}\binom{۶-۳}{۴-۲}}{\binom{۶}{۴}} = ۰.۶$$

- **توزیع پواسون**

توزیع پواسون در دو حالت استفاده می شود یکی برای تقریب توزیع دوجمله ای و دیگری برای بررسی تعداد مراجعات به سیستم با میانگین $\lambda$ در واحد زمان ( t ).

**۱. توزیع پواسون برای تقریب توزیع دوجمله ای:**

➢ **ویژگی های توزیع پواسون برای تقریب توزیع دوجمله ای:**

اگر n به سمت بی نهایت و p به سمت صفر میل کند و در عین حال مقدار np ثابت بماند که با نماد $\lambda$ نشان داده می شود ( $np = \lambda$)، می توان بجای توزیع دو جمله ای از توزیع پواسون استفاده نمود.

متغیر مورد نظر: تعداد موفقیت ها در n آزمایش.

در صورت برقراری شروط فوق آنگاه X تعداد موفقیت ها در n آزمایش دارای توزیع پواسون است. و فرمول توزیع پواسون به صورت زیر بیان می شود.

$$P(X=x)=\frac{e^{-\lambda}\cdot\lambda^{x}}{x!},x=۰,۱,۲,...$$

$\lambda=np$ بعنوان پارامتر توزیع

$e\cong ۲/۷۱۸$

## ➢ امید ریاضی و واریانس توزیع پواسون

از بین کلیه توزیع های رایج ، توزیع پواسون تنها توزیعی است که میانگین و واریانس آن با هم برابرند.

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

**مثال۱۶:** احتمال اینکه تیراندازی تیرش به خطا برود ۳ درصد است اگر وی ۱۵۰ تیر شلیک کند هر یک از موارد زیر را محاسبه کنید.

الف) احتمال اینکه ۳ تیر به خطا برود.

ب) امید ریاضی و واریانس تعداد تیرهای خطا رفته را محاسبه کنید.

ج) احتمال اینکه حداقل ۳ تیر به خطا برود.

جواب :

الف)

$$p=۰/۰۳\ ,\ n=۱۵۰\ ,\ \lambda=np=۱۵۰\times ۰/۰۳=۴/۵,X=۳$$
$$P(X=x)=\frac{e^{-\lambda}.\lambda^{x}}{x!}\Rightarrow P(X=۳)=\frac{e^{-۴/۵}(۴/۵)^{۳}}{۳!}=۰/۱۶۸۷$$

ب)

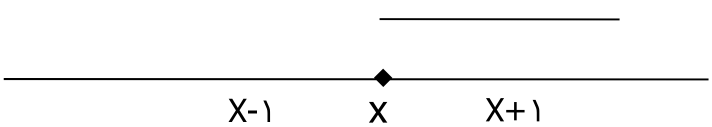
$$E(X)=V(X)=np=۱۵۰\times ۰/۰۳=۴/۵$$

➢ وقتی که x نسبتاً بزرگ است ، محاسبه احتمال از طریق فرمول کار خسته کننده ای می شود ، لذا برای رفع این مشکل از جدول های مخصوصی استفاده می شود. در این کتاب این جدول در پیوست (۲) ارائه شده است. در این جدول احتمال $P(X \le c)=\sum_{x=۰}^{c}\frac{e^{-\lambda}\lambda^{x}}{x!}$ برای مقادیر مختلف $\lambda$و c آورده شده است.

➢ همینطور با توجه به اینکه فقط برای مقادیر کوچکتر در جدول مقدار ارائه شده است. از قوانین احتمال استفاده کرده و داریم:

$$P(X > x)=۱-P(X \le x)$$

$$P(X \ge x)=۱-P(X \le x-۱)$$




$$P(a \le X \le b)=P(X \le b)-P(X \le a-۱)$$

ج)

$$P(X \ge ۲)=۱-P(X \le ۲)=۱-[P(X=۰)+P(X=۱)+P(X=۲)]$$
$$=۱-(\frac{e^{-۴/۵}(۴/۵)^{۰}}{۰!}+\frac{e^{-۴/۵}\times(۴/۵)^{۱}}{۱!}+\frac{e^{-۴/۵}\times(۴/۵)^{۱}}{۱!})=۱-۰/۱۷۴=۰/۸۲۶$$

**۲- توزیع پواسون برای تعداد مراجعات** به سیستم با میانگین $\lambda$ در واحد زمان ( t )

➢ **ویژگی های توزیع پواسون تعداد مراجعات** به سیستم با میانگین $\lambda$ در واحد زمان ( t ):

زمانی که **تعداد مراجعات** به سیستمی با میانگین مراجعه $\lambda$ در واحد زمان t انجام می شود مد نظر باشد.

متغیر مورد نظر: تعداد مراجعه.

در صورت برقراری شروط فوق آنگاه X تعداد مراجعات دارای توزیع پواسون است. و فرمول توزیع پواسون به صورت زیر بیان می شود.

$$P(X=x)=\frac{e^{-(t\lambda)}\cdot(t\lambda)^{x}}{x!},x=۰,۱,۲,...$$

که در رابطه فوق $\lambda$ متوسط تعداد مراجعات در واحد زمان است و $t$ مدت زمان مراجعات.

مثال۱۷: تعداد اتومبیل‌هایی که به پمپ بنزینی مراجعه می‌کند دارای توزیع پواسون با میانگین ۲/۵ اتومبیل در هر ۱۰ دقیقه می‌باشد. این احتمالات را محاسبه کنید.

الف) در ۱۰ دقیقه اول بیش از یک اتومبیل مراجعه کند.

ب) در ۲۰ دقیقه بیش از ۳ اتومبیل و کمتر یا مساوی ۶ اتومبیل مراجعه کند.

ج) در ۵ دقیقهٔ اول اتومبیلی مراجعه نکند.

حل:

الف)

$$\left.\begin{array}{l}\lambda=۲/۵ \quad \text{در هر ۱۰} \\ t=\frac{۱۰}{۱۰}=۱\end{array}\right\} \Rightarrow \lambda t=۲/۵\times۱=۲/۵$$

$$P(X>۲)=۱-P(X\leq۱)=۱-۰/۲۸۷=۰/۷۱۳$$

ب)

$$t=\frac{۲۰}{۱۰}=۲ \qquad \lambda t=۲/۵\times۲=۵$$

$$P(۳<X\leq۶)=P(X\leq۶)-P(X\leq۳)=۰/۷۶۲-۰/۲۶۵=۰/۴۹۷$$

ج)

$$t=\frac{۵}{۱۰}=\frac{۱}{۲} \qquad \lambda t=۲/۵\times۰/۵=۱/۲۵$$

$$P(X=۰)=\frac{e^{-۱/۲۵}\times(۱/۲۵)^{۰}}{۰!}=۰/۲۸۶$$

# فصل هفتم: توابع احتمال پیوسته

هدف اصلی این فصل آشنا ساختن دانشجویان با متغیرهای تصادفی پیوسته و تعدادی از توابع مهم آنهاست.

## ❖ احتمال متغیرهای پیوسته

بخاطر این که میزان احتمال در توابع پیوسته در یک نقطه معین مساوی صفر است ، لذا در این گونه توابع ، احتمال همیشه در قالب یک فاصله تعیین می شود

## ❖ تابع چگالی احتمال

احتمال این که متغیر تصادفی پیوسته x مقداری بین دو نقطه a و b را بگیرد برابر است با

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

که رابطه فوق f(x) تحت عنوان تابع چگالی احتمال شناخته می شود که دارای ویژگی های زیر است.

$$۱) f(x) \ge ۰, \qquad -\infty < x < +\infty$$
$$۲) \int_{-\infty}^{+\infty} f(x)dx = ۱$$

- نقش علامت مساوی در احتمالات پیوسته

$$۱) P(X = a) = \int_a^a f(x)dx = ۰$$
$$۲) P(a < X < b) = P(a \le X < b) = P(a < x \le b) = P(a \le X \le b)$$

پس علامت مساوی در این توزیع ها نقشی ایفاء نمی کند.

مثال (۱): فرض کنیم تابع چگالی احتمال X به صورت زیر باشد.

$$f(x)=\begin{cases} ke^{-2x} & , x>0 \\ 0 & , O.W \end{cases}$$

الف) k را تعیین کنید .

ب) $p(0.5\le x\le 1)$

حل :

بنابر ویژگی اول $f(x)$ همواره باید بزرگتر از صفر باشد بنابراین برای برقرار این شرط $k\ge 0$ باشد.

برای برقراری شرط دوم داریم:

$$\int_{-\infty}^{+\infty} ke^{-2x}dx = k\int_{0}^{\infty} e^{-2x}dx = 1$$

$$k[-\frac{1}{2}e^{-2x}]_{0}^{\infty}=1 \qquad -\frac{k}{2}[0-1]=\frac{k}{2}=1 \Rightarrow k=2$$

$$\Rightarrow f(x)=\begin{cases} 2e^{-2x} & x>0 \\ 0 & ow \end{cases}$$

ب)

$$p(0.5\le x\le 1)=\int_{0.5}^{1} 2e^{-2x}dx = 2\int_{0.5}^{1} e^{-2x}dx = 2[-\frac{1}{2}e^{-2x}]_{0.5}^{1} = -1(e^{-2}-e^{-1.5})=e^{-1.5}-e^{-2}$$

## ❖ تابع توزیع متغیر تصادفی پیوسته

تابع توزیع متغیر تصادفی پیوسته که گاهی تحت عنوان تابع توزیع تجمعی نیز نامیده می شود به صورت زیر تعریف می شود.

$$F_X(x)=p(X\le x)=\int_{-\infty}^{x} f(t)dt$$

بدیهی است که $F(-\infty)=0$ و $F(+\infty)=1$ است.

بنا به تعریف تابع توزیع داریم :

$$p(a \le X \le b) = \int_a^b f(x)dx = P(X \le b) - P(X \le a) = F(b) - F(a)$$

مثال(۲): در مثال (۱)، تابع توزیع را محاسبه کرده و بر اساس آن احتمال $p(0.5 \le x \le 1)$ را محاسبه کنید

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_0^x 2e^{-2t}dt = 2\left(-\frac{1}{2}e^{-2t}\right)_0^x = -(e^{-2x} - 1) = 1 - e^{-2x}$$

$$p(0.5 \le x \le 1) = F(1) - F(0.5) = (1 + e^{-2}) - (1 - e^{-1.5}) = e^{-1.5} - e^{-2}$$

## ❖ امید ریاضی متغیر تصادفی پیوسته

یعنی ضرب متغیر تصادفی در تابع چگالی خود و سپس انتگرال گیری به ازای مقادیر ممکن متغیر.

$$\mu = E(X) = \int_{-\infty}^{+\infty} X \cdot f(X)dx$$

## ❖ واریانس متغیر تصادفی پیوسته

یعنی کسر متغیر تصادفی از میانگین خود ، به توان ۲ رساندن نتیجه و ضرب نتیجه حاصله به تابع چگالی و انتگرال گیری

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

در محاسبات ، واریانس به صورت زیر نیز محاسبه می شود :

$$\sigma_X^2 = E(X^2) - (E(X))^2 = E(X^2) - \mu^2 = \int_x x^2.f(x)dx - \mu^2$$

نکته ۱ : انحراف معیار : $\sigma = \sqrt{\sigma^2}$

نکته ۲ : $Var(ax + b) = a^2\sigma_X^2$

مثال (۳): فرض می کنیم متغیر تصادفی X دارای تابع چگالی احتمال زیر است :

$$f(x)=\begin{cases}-\frac{1}{2}(2x-3x^2) & ,-1<x<1\\ 0 & ,O.W\end{cases}$$

میانگین و واریانس X را محاسبه کنید .

حل :

$$E(x)=\mu=\int_{-1}^{1}x.(-\frac{1}{2})(2x-3x^2)dx$$

$$=\int_{-1}^{1}-\frac{1}{2}(2x^2-3x^3)dx=-\frac{1}{2}[\frac{2}{3}x^3-\frac{3}{4}x^4]_{-1}^{1}=-\frac{2}{3}$$

$$\sigma^2=E(x^2)-(E(x))^2=\int_{-1}^{1}x^2.(-\frac{1}{2})(2x-3x^2)dx-(\frac{2}{3})^2$$

$$=\int_{-1}^{1}(-\frac{1}{2})(2x^3-3x^4)dx-\frac{4}{9}=-\frac{1}{2}[\frac{1}{2}x^4-\frac{3}{5}x^5]_{-1}^{1}-\frac{4}{9}=\frac{3}{5}-\frac{4}{9}=\frac{67}{90}$$

نکته: گشتاور مرتبه r ام $\mu^r=E(X^r)=\int_{-\infty}^{+\infty}x^r f(x)dx$

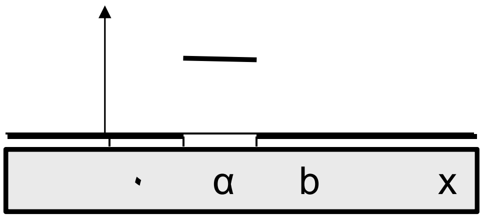
## ❖ تابع چگالی احتمال یکنواخت (مستطیلی)

متغیر تصادفی X دارای تابع چگالی احتمال یکنواخت با پارامترهای $\alpha$ و $\beta$ است اگر احتمال وقوع x در فاصله های هم اندازه در فاصله $\alpha$ و $\beta$ برابر باشد. تابع چگالی یکنواخت به صورت زیر تعریف می شود.

$$f(x)=\begin{cases}\frac{1}{\beta-\alpha} & ,\alpha<x<\beta\\ 0 & ,OW\end{cases}$$

- **ویژگیهای تابع چگالی احتمال یکنواخت**

۱- نمودار $f(x)$ برای $-\infty<\alpha<\beta<\infty$ به صورت زیر است.




۲- تابع توزیع $F(X)$ برابر است با:

$$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(t)dt=\begin{cases} ۰ & x<\alpha \\ \dfrac{x-\alpha}{\beta-\alpha} & \alpha\leq x<\beta \\ ۱ & x\geq\beta \end{cases}$$

۳- دارای میانگین $\frac{\alpha+\beta}{۲}$ و واریانس $\frac{(\beta-\alpha)^{۲}}{۱۲}$ است.

۴- برای $\alpha=۰$، $\beta=۱$ ، توزیع x را یکنواخت استاندارد گویند.

مثال (۴): تابع توزیع متغیر تصادفی x به صورت زیر است.

$$f(x)=\begin{cases} \frac{۱}{۱۰} & ۵<x<۱۵ \\ ۰ & \text{در غیر این} \end{cases}$$

مطلوب است محاسبهٔ :

الف) $p(۰\leq x\leq ۱۰)$

ب) امید ریاضی و واریانس x

حل :

الف) $p(۰\leq x\leq ۱۰)=\int_{۰}^{۱۰}\frac{۱}{۱۰}dx=\int_{۰}^{۵}\frac{۱}{۱۰}dx+\int_{۵}^{۱۰}\frac{۱}{۱۰}dx=۰+\frac{x}{۱۰}\Big]_{۵}^{۱۰}=۱-\frac{۱}{۲}=\frac{۱}{۲}$

ب) $E(x)=\frac{(\alpha+\beta)}{۲}=\frac{(۵+۱۵)}{۲}=۱۰$

$$v(x)=\frac{۱}{۱۲}(\beta-\alpha)^{۲}=\frac{۱}{۱۲}(۱۵-۵)^{۲}=\frac{۱۰۰}{۱۲}=\frac{۲۵}{۳}$$

## ❖ تابع چگالی احتمال نمایی

متغیر تصادفی X دارای تابع چگالی احتمال نمایی با پارامتر $\lambda$ است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & ,x>۰ \\ ۰ & ,O.W\end{cases}$$

توزیع نمایی کاربردهای مهمی دارد. به طور مثال اگر تعداد موفقیت ها یا ورودی ها دارای توزیع پواسون باشد، می توان نشان داد که زمان انتظار مابین موفقیت ها یا ورودی های متوالی از توزیع نمایی پیروی می کند.

- ویژگیهای توزیع نمایی

$$۱-F(x)=P(X<x)=\int_{-\infty}^{x} f(t)dt=۱-e^{-\lambda x}$$
$$۲-P(X>x)=۱-P(X<x)=۱-F(x)=e^{-\lambda x}$$
$$۳-E(X)=\frac{۱}{\lambda} \quad , \quad Var(X)=\frac{۱}{\lambda^{۲}}$$

مثال(۵): تعداد اتومبیل‌هایی که به یک رستوران بین راهی مراجعه می‌کنند به طور متوسط در هر ساعت ۵ اتومبیل می‌باشد می‌خواهیم مقادیر زیر را محاسبه کنیم.

الف) به طور متوسط فاصلهٔ زمانی بین مراجعهٔ ۲ اتومبیل.

ب) احتمال اینکه فاصلهٔ زمانی بین ۲ مراجعه متوالی کمتر از ده دقیقه باشد.

ج) احتمال اینکه ۲ مراجعهٔ متوالی بیش از ۳ دقیقه طول بکشد.

حل: چون تعداد مراجعات از توزیع پواسون برخوردار است فاصلهٔ زمانی بین دو مراجعه از توزیع نمایی برخوردار است و چون به طور متوسط ۵ اتومبیل به رستوران در هر ساعت مراجعه می‌کنند پس داریم.

$$\lambda=۵ \qquad E(x)=\frac{۱}{\lambda}=\frac{۱}{۵}$$

ب) چون واحد زمانی ۱۰ ساعت است ابتدا ده دقیقه را به ساعت تبدیل می‌کنیم.

$$\frac{۱۰}{۶۰}=\frac{۱}{۶}=۰/۱۶۷$$

حال داریم :

$$P(X\leq ۰/۱۶۷)=F(۰/۱۶۷)=۱-e^{-\lambda x}=۱-e^{-۵x}=۱-e^{-۵(۰/۱۶۷)}=۱-e^{-۰/۸۳۳}=۰/۵۶۵$$

ج) ابتدا زمان ۳ دقیقه را بر حسب ساعت بدست می‌آوریم.

$$\frac{۳}{۶۰}=\frac{۱}{۲۰}=۰/۰۵$$

$$P(X>۰/۰۵)=۱-F(۰/۰۵)=e^{-\lambda x}=e^{-۵x}=e^{-۵(۰/۰۵)}=۰/۷۷۹$$

چند نکته:

١) $\int_a^b (nx+m)dx = n\int_a^b x\,dx + \int_a^b m\,dx$

٢) $\int_a^b c\,dx = cx \Big|_a^b = cb - ca$

٣) $\int_a^b x\,dx = \frac{١}{٢}x^{٢} \Big|_a^b = \frac{١}{٢}b^{٢} - \frac{١}{٢}a^{٢}$

٤) $\int_a^b x^r dx = \frac{١}{r+١}x^{r+١} \Big|_a^b = \frac{١}{r+١}b^{r+١} - \frac{١}{r+١}a^{r+١}$

٥) $\int_a^b \sqrt{x}\,dx = \int_a^b x^{\frac{١}{٢}}dx = \frac{١}{\frac{١}{٢}+١}x^{\frac{١}{٢}+١} \Big|_a^b = \frac{٢}{٣}b^{\frac{٣}{٢}} - \frac{٢}{٣}a^{\frac{٣}{٢}}$

٦) $\int_a^b \frac{١}{\sqrt{x}}dx = \int_a^b x^{-\frac{١}{٢}}dx = \frac{١}{-\frac{١}{٢}+١}x^{-\frac{١}{٢}+١} \Big|_a^b = ٢b^{\frac{١}{٢}} - ٢a^{\frac{١}{٢}} = ٢(\sqrt{b}-\sqrt{a})$

٧) $\int_a^b e^x dx = e^x \Big|_a^b = e^b - e^a$

٨) $\int_a^b e^{-cx}dx = -\frac{١}{c}e^{-x} \Big|_a^b = \frac{١}{c}e^{-a} - \frac{١}{c}e^{-b}$

# فصل هشتم: توزیع نرمال

در این فصل دانشجویان با مهم ترین و کاربردی ترین نوع توزیع از زیر مجموعه توزیع های پیوسته یعنی توزیع نرمال آشنا می شوند.

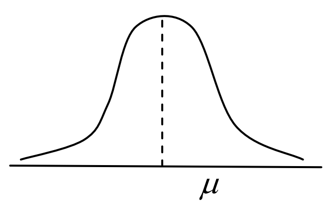
## ❖ تعریف توزیع نرمال

توزیع نرمال توزیعی زنگی شکل است که کاربردهای وسیعی است چرا که اولا خیلی از پدیده های طبیعی دارای این توزیع هستند و ثانیا شکل حدی بسیاری از توزیع های دیگر نیز نرمال است.

متغیر تصادفی پیوسته X با میانگین $\mu$ و انحراف معیار $\sigma$، در صورت داشتن تابع چگالی زیر دارای توزیع نرمال است

$$f(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma}.e^{\frac{-1}{2}\left(\frac{X-x}{\sigma}\right)^2}, & -\infty < x < \infty \\ 0 & ,ow\end{cases}$$

در این رابطه $\pi$=۳/۱۴۱۵۹... و $e$=۲/۷۱۸۲... است. شکل زیر نشان دهندۀ منحنی نرمال است.

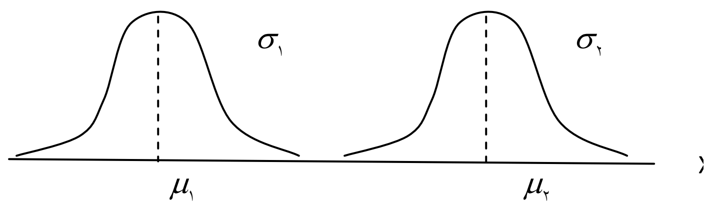



دو پارامتر توزیع نرمال $\mu$ , $\sigma$ است که با مشخص بودن آن‌ها توزیع دقیقاً مشخص و منحنی آن قابل ترسیم است. بدلیل کاربرد زیاد آن متغیر تصادفی X که دارای توزیع نرمال با میانگین $\mu$ و انحراف معیار $\sigma$ است را با $X \sim N(\mu , \sigma)$ نشان می‌دهیم و به این صورت می‌خوانیم X دارای توزیع نرمال با میانگین $\mu$ و انحراف $\sigma$ است.

### ❖ نقش میانگین در منحنی توزیع نرمال

در یک توزیع نرمال هر قدر میانگین افزایش یابد ، باعث می شود که منحنی آن بیشتر به سمت راست انتقال یابد.

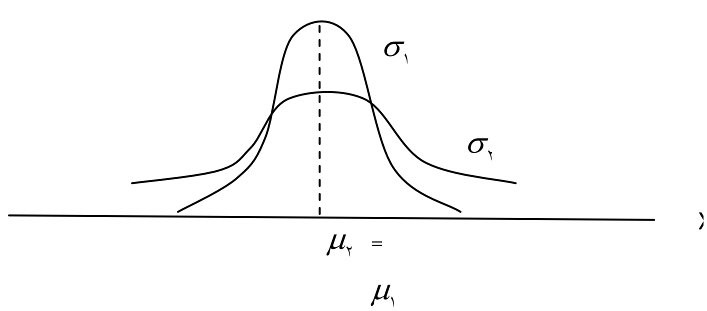
شکل زیر دو منحنی نرمال را نشان می‌دهد که انحراف معیارشان برابر ($\sigma_1=\sigma_2$) ولی میانگین اولی کوچکتر از میانگین دومی است ($\mu_1<\mu_2$)




### ❖ نقش انحراف معیار در توزیع نرمال

هر قدر انحراف معیار افزایش یابد ، منحنی توزیع نرمال کوتاه تر (بعبارتی پهن تر) می شود.

در شکل زیر دو منحنی نرمال را نشان می‌دهد که میانگین‌شان یکسان ($\mu_1=\mu_2$) ولی انحراف معیار دومی بزرگتر از انحراف معیار اولی است ($\sigma_1<\sigma_2$)



❖ **خصوصیات توزیع نرمال**

۱- سطح زیر منحنی همیشه برابر یک است.

۲- $f(x)$ همیشه بزرگتر یا مساوی صفر است.

۳- حداکثر مقدار تابع در X = μ می باشد.

۴- تابع حول میانگین ، متقارن است.

۵- میانگین و واریانس X به ترتیب μ و $\sigma^{۲}$ می باشد.

۶- منحنی در محور X ها ، هیچ گاه به صفر نمی رسد.

۷- میانگین ، میانه و مد با هم برابرند.

۸-احتمال x با توجه به انحراف معیارهای مختلف بشرح ذیل است :

۱. در صورتی که به اندازه یک انحراف معیار در طرفین میانگین جدا کنیم ، سطح حاصله تقریبا برابر ۶۸٪ سطح خواهد بود .

$$p(\mu - ۱\sigma < X < \mu + ۱\sigma) = ./۶۸$$

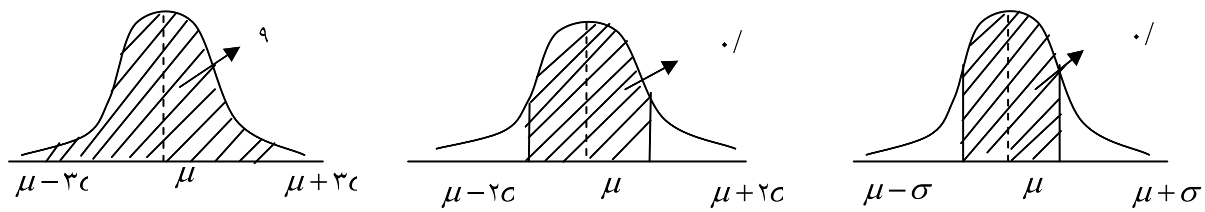
۲. ۴- اگر به اندازه ی ۲ انحراف معیار در طرفین میانگین جدا کنیم ، سطح حاصله تقریبا ۹۵٪ سطح کل می باشد .

$$p(\mu - ۲\sigma < X < \mu + ۲\sigma) = ./۹۵$$

۳. اگر به اندازه ی ۳ انحراف معیار جدا کنیم ، سطح حاصله تقریبا ۹۹/۷٪ خواهد بود .

$p(\mu - ۳\sigma < X < \mu + ۳\sigma) = ۰/۹۹۷$

این مفاهیم در شکل زیر نمایش داده شده است.




## ❖ توزیع نرمال استاندارد

برای این که در حالت کلی هر احتمالی را بتوانیم در توزیع نرمال حساب کنیم ، کافی است نرمال را به نرمال استاندارد تبدیل کرده و با استفاده از جدول های مربوطه مقدار احتمال را محاسبه کنیم.

توزیع نرمال استاندارد، توزیع نرمالی است که دارای میانگین صفر و واریانس ۱ است. اگر X متغیری باشد که دارای توزیع نرمال با میانگین $\mu$ و انحراف معیار $\sigma$ است یعنی $X \sim N(\mu,\sigma)$ . برای تبدیل آن به نرمال استاندارد Z، که دارای توزیع نرمال با میانگین صفر انحراف معیار ۱ است به صورت زیر عمل می کنیم.

با کم کردن میانگین از متغیر x و تقسیم نتیجه آن بر انحراف معیار ، z بدست می آید

$$z = \frac{X - \mu}{\sigma}$$

## ❖ استفاده از جدول توزیع نرمال استاندارد

بعد از آن که نرمال استاندارد را تعریف کردیم فورا نتیجه می گیریم :

$$p(x_۱ < X < x_۲) = p(\frac{x_۱ - \mu}{\sigma} < X < \frac{x_۲ - \mu}{\sigma}) = p(z_۱ < Z < z_۲)$$

احتمالات $p(z_۱ < Z < z_۲)$ را می توانیم با استفاده از جدول نرمال استاندارد به راحتی محاسبه کنیم ( استفاده از جدول ) .

جدول های مورد نظر کتاب مقدار احتمال را از ∞–تاعدد بخصوصی می دهد. یعنی $p(Z < z_۱)$

- چند نکته در مورد روابط احتمالات در متغیرهای پیوسته

اگر Y یک متغیر تصادفی پیوسته باشد.

$$۱) P(a < Y < b) = P(a \le Y < b) = P(a < Y \le b) = P(a \le Y \le b)$$
$$۲) P(Y > a) = ۱ - P(Y < a)$$
$$۳) P(a < Y < b) = P(Y < b) - P(Y < a)$$

- روش های استفاده از جدول توزیع نرمال استاندارد

۱– استفاده مستقیم : مقدار Z مشخص است و احتمال آن را بدست می آوریم.

مثال ۱: مطلوبست محاسبه

$$p(Z < ۱/۷۳)$$
$$P(Z > ۱/۷۳)$$
$$P(۱/۳ < Z < ۱/۷۵)$$

حل:

$$P(Z < ۱/۷۳) = ./۹۵۷۳$$
$$P(Z > ۱/۷۳) = ۱ - P(Z < ۱/۷۳) = ۱ - ./۹۵۷۳ = ./.٤۲۷$$
$$P(۱/۳ < Z < ۱/۷۵) = P(Z < ۱/۷۵) - P(Z < ۱/۳) = ./۹۵۹۹ - ./۹.۳۲ = ./.٥٦۷$$

مثال۲: فرض کنیم X دارای توزیع نرمال با میانگین ۴/۳۵ و انحراف معیار ۰/۵۹ باشد. مطلوب است احتمال آن که

$P(X<٤/٣٥)$
$p(X>٥/٢)$
$P(٢/٥<X<٥/٥)$

باشد .

حل: طبق اطلاعات مسئله $X \sim N(٤/٣٥,./٥٩)$

برای محاسبه احتمالات فوق ابتدا باید متغیر مورد نظر به نرمال استاندارد تبدیل شود.

$$P(X<٤/٣٥)=P(\frac{X-\mu}{\sigma}>\frac{٤/٣٥-٤/٣٥}{./٥٩})=p(Z<./..)=./٥$$
$$P(X>٥/٢)=P(\frac{X-\mu}{\sigma}>\frac{٥/٢-٤/٣٥}{./٥٩})=P(Z>١/٤٤)=١-P(Z<١/٤٤)=١-./٩٢٥١=./.٧٤٩$$
$$P(٢/٥<X<٥/٥)=P(\frac{٢/٥-٤/٣٥}{./٥٩}<\frac{X-\mu}{\sigma}<\frac{٥/٥-٤/٣٥}{./٥٩})$$
$$=P(-٣/١٤<Z>١/٩٥)=P(Z<١/٩٥)-P(Z<-٣/١٤)=./٩٧٤٤-./...٨=./٩٧٣٦$$

مثال۳: قطر یک بلبرینگ متغیر نرمال با میانگین ۷ سانتیمتر و انحراف معیار ۰/۱ است. استاندارد فنی قطر این کالا عبارتست از $۶/۹۱\le X\le ۷/۰۹$ و تولید یک بلبرینگ استاندارد ۱۲۵۰ ریال سود دارد. اگر قطر بلبرینگ تولیدی کمتر از ۶/۹۱ باشد غیر قابل استفاده است و ۱۰۵۰ ریال زیان دارد و در صورتیکه قطر آن بیش از ۷/۰۹ باشد با انجام دادن کار اضافی و صرف هزینه‌ای معادل ۲۰۰ ریال می‌توان آنرا به بلبرینگ استاندارد تبدیل کرد. سود مورد انتظار هر بلبرینگ را حساب کنید.

حل: با توجه به صورت مسأله ما سه حالت داریم که باید احتمال هر یک از این ۳ حالت را محاسبه کنیم و بعد از آن در هر کدام از حالت‌ها سودی را که می‌توانیم بدست آوریم محاسبه کرده و با ضرب کردن احتمال هر حالت در سود آن، سود مورد انتظار را بدست می‌آوریم.

(احتمال حالت اول $$P(۶/۹۱\le X\le ۷/۰۹)=P(\frac{۶/۹۱-۷}{۰/۱}\le Z\le\frac{۷/۰۹-۷}{۰/۱})=p(-۰/۹\le Z\le ۰/۹)=$$
$$=P(Z\le ۰/۹)-P(Z\le -۰/۹)=۰/۸/۵۹-۰/۱۸۴۱=۰/۶۳۱۸$$

احتمال حالت دوم) $P(X<۶/۹۱)=P(Z<\frac{۶/۹۱-۷}{۰/۱})=P(Z<-۰/۹)=۰/۱۸۴۱$

احتمال حالت سوم) $=P(X>۷/۰۹)=P(Z>\frac{۷/۰۹-۷}{۰/۱})=P(Z>۰/۹)$
$=۱-P(Z\leq ۰/۹)=۱-۰/۸۱۵۹=۰/۱۸۴۱$

| | حالت اول | حالت دوم | حالت سوم |
|---|---|---|---|
| سود | ۱۲۵۰ | −۱۰۵۰ | ۱۲۵۰−۲۰۰=۱۰۵۰ |

سود مورد انتظار $=(۰/۶۳۱۸\times ۱۲۵۰)+(۰/۱۸۴۱\times -۱۰۵۰)+(۰/۱۸۴۱\times ۱۰۵۰)=۷۸۹/۷۵\approx ۷۹۰$

۲– استفاده معکوس : احتمال Z مشخص است و مقدار آن را بدست می آوریم.

در استفاده مستقیم از توزیع نرمال، ابتدا Z را مشخص و سپس احتمال آن را از جدول پیدا می کردیم، در استفاده معکوس، مقدار Z برای مشخص نیست و تنها احتمال آن مشخص است، احتمال را در جدول پیدا کرده و سپس Z متناظر با آن را مشخص می کنیم.

- در روش مستقیم برای حل مسائل، ابتدا متغیر تصادفی X را به Z تبدیل و سپس به جدول مراجعه می کردیم، ولی در روش معکوس، پس از پیدا کردن Z مقدار X را با استفاده از رابطه زیر پیدا می کنیم.

$$Z=\frac{X-\mu}{\sigma}\Rightarrow X=\mu+\sigma Z$$

مثال ۴: می‌خواهیم مقدار z را در صورتیکه $P(Z\leq z)=۰/۰۴۸۵$ را محاسبه کنیم.

حل : ابتدا عدد ۰/۰۴۸۵ را از جدول پیدا می‌کنیم و سپس مقدار z آنرا با توجه به سطر و ستون مربوط پیدا می‌کنیم که برابر ۱/۶۶– است پس :

$$P(Z\leq -۱/۶۶)=۰/۰۴۸۵$$

مثال ۵: متغیر تصادفی x دارای توزیع نرمال با میانگین ۱۲ و انحراف معیار ۳ می‌باشد به ازاء چه مقداری از x رابطهٔ $P(X \leq x)=۰/۸۵۰۸$ برقرار است؟

حل : با توجه به جدول احتمال ۰/۸۵۰۸ برابر با $z=۱/۰۴$ می‌باشد یعنی $P(Z \leq ۱/۰۴)=۰/۸۵۰۸$ پس :

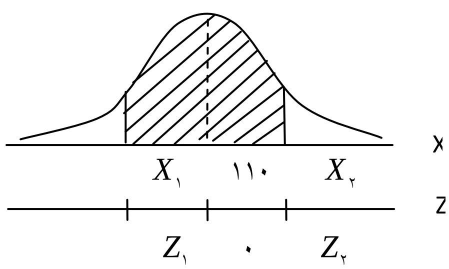
$$Z = \frac{X - \mu}{\sigma} \Rightarrow x = \mu + z\,\sigma \Rightarrow x = ۱۲ + ۱/۰٤ \times ۳ = ۱۵/۱۲$$

مثال ۶: زمان لازم برای صدور کارت برای هر داوطلب شرکت در آزمونی به طور متوسط ۱۱۰ ثانیه با انحراف معیار ۳۵ ثانیه است که به صورت توزیع نرمال است. می‌خواهیم موارد زیر را محاسبه کنیم :

الف) کار ۸۵ درصد از کارت‌های صادر شده در چه دامنه‌ای در دو طرف میانگین زمان صدور کارت هر داوطلب قرار می‌گیرد.

ب) کار ۵ درصد از داوطلبانی که بیشترین زمان را به خود اختصاص می‌دهد، حداقل چقدر طول می‌کشد.

حل :




هدف ما پیدا کردن دو مقدار $x_۱$ , $x_۲$ است. بطوریکه: $P(x_۱ \leq X \leq x_۲) = ۰/۸۵$

پس قسمت هاشور نخورده در دو طرف منحنی مساحتی برابر با ۱−۰/۸۵=۰/۱۵ دارد که چون توزیـع نرمال می‌باشد و توزیع نرمال حول میانگین متقارن است پس مساحت هـر قســمت هاشـور نخـورده برابـر است با $\frac{0/15}{2}=0/075$ حال باید مقادیر z را برای دو طرف پیدا کنیم که احتمال آن ۰/۰۷۵ است یعنی :

$$P(X > x_2)=0/075 \Rightarrow P(Z \geq z_2)=0/075$$
$$\Rightarrow P(Z \geq z_2)=1-P(Z \leq z_2)=0/075 \Rightarrow P(Z \leq z_2)=0/925$$
$$\Rightarrow P(Z \leq 1/44)=0/925 \Rightarrow z_2=1/44$$
$$P(X < x_1)=0/075 \Rightarrow P(Z \leq z_1)=0/075 \Rightarrow P(Z \leq -1/44)=0/075 \Rightarrow z_1=-1/44$$

حال با توجه به مقادیر $z_1 , z_2$ مقدار $x_1 , x_2$ را می‌یابیم پس داریم :

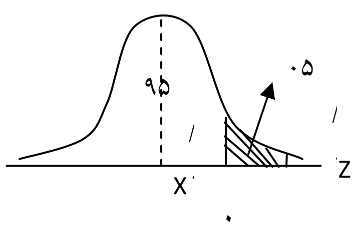
$$z_2=\frac{x_2-\mu}{\sigma} \Rightarrow x_2=\mu+z_2\sigma=110+1/44(35)=150/4$$
$$z_1=\frac{x_1-\mu}{\sigma} \Rightarrow x_1=\mu+z_1\sigma=110-1/44(35)=59/6$$

پس :

$$59/6 \leq X \leq 150/4$$

ب)

چون ۵ درصد از داوطلبان بیشترین زمان را به خود اختصاص می‌دهند پس ۹۵ درصد آن‌ها در زمان عادی کارشان انجام می‌شود با توجه به شکل هدف ما پیدا کردن مقدار x است.



$P(Z \leq z)=۰/۹۵ \Rightarrow P(Z \leq ۱/۶۴۵)=۰/۹۵$
$x=\mu+z\sigma \Rightarrow x=۱۱۰+(۱/۶۴۵\times۳۵)=۱۶۷/۵۷۵$

پس با توجه به نتایج کار ۵ درصدی از داوطلبان که بیشترین زمان را به خود احتصاص می دهند حداقل ۱۶۷/۵۷۵ ثانیه طول می کشد.

### ❖ تقریب توزیع دو جمله ای بوسیله توزیع نرمال

در فصل قبل توضیح داده شد که در توزیع دو جمله‌ای وقتی که n خیلی بزرگ شود و P به سمت صفر میل کند محاسبهٔ احتمال کار دشواری خواهد بود بنابراین در چنین حالتی ( در توزیع دوجمله ای با n بزرگ و p کوچک ) از تقریب توزیع پواسون با پارامتر $\lambda=np$ استفاده می شود.

در صورتی که n بزرگ باشد و p به صفر یا یک نزدیک نباشد توزیع نرمال تقریب خوبی برای توزیع دو جمله‌ای خواهد بود که در این صورت پارامترهای میانگین و انحراف معیار به صورت زیر محاسبه خواهد شد :

$$\mu=np \qquad \sigma=\sqrt{npq}$$

- بنا به تجربه اگر np و nq هر دو بزرگتر از ۵ باشند تقریب نرمال، تقریب خوبی برای توزیع دوجمله‌ای خواهد بود در مواردی که p به ۰/۵ نزدیک باشد تقریب نرمال برای n های کوچک نیز خوب است.
- همانطور که می‌دانید توزیع دوجمله‌ای توزیع گسسته است و توزیع نرمال توزیعی پیوسته است مثلاً در توزیع دوجمله‌ای با $n=۱۰$ و $p=۰/۴$، $P(X=۵)$ مقداری مثبت است ولی در توزیع نرمال $P(X=۵)$ برابر صفر است بنابراین وقتی تقریب نرمال را برای دوجمله‌ای به کار می‌بریم باید از «تصحیح پیوستگی» استفاده کنیم. لازم به ذکر این نوع تصحیح در تمام مواردی که یک توزیع گسسته بوسیله یک توزیع پیوسته تقریب زده می شود باید انجام شود. در جدول زیر تصحیح‌های پیوستگی مختلف آمده شده است :

| احتمال موردنظر از توزیع نرمال | احتمال مورد نظر از توزیع گسسته<br><br>(مثلا دو جمله ای یا پواسون) |
|---|---|
| $P(x-۰/۵\le X\le x+۰/۵)$ | $P(X=x)$ |
| $P(X\le x+۰/۵)$ | $P(X\le x)$ |
| $P(X\le x-۱+۰/۵)=P(X\le x-۰/۵)$ | $P(X<x)$ |
| $P(X\ge x-۰/۵)$ | $P)X\ge x)$ |
| $P(X\ge x+۱-۰/۵)=P(X\ge x+۰/۵)$ | $P(X>x)=P(X\ge x+۱)$ |
| $P(x_۱-۰/۵\le X\le x_۲+۰/۵$ | $P(x_۱\le x\le x_۲)$ |

مثال۷: یک شبکه تلویزیونی ادعا می کند که ۷۰ درصد از کل بینندگان تلویزیون وقت خود را به دیدن برنامه خاصی در شبهای سه شنبه اختصاص می دهند. به فرض درست بودن این ادعا، احتمال آنکه از بین ۴۰۰ بیننده، حداقل ۲۵۰ نفر از آنان برنامه را تماشا کرده باشند، چقدر است؟

$$X \sim Binomial(٤٠٠,٠/٧)$$
$$\Rightarrow \mu=np=٤٠٠\times ٠/٧=٢٨٠ \ , \sigma=\sqrt{npq}=\sqrt{٤٠٠\times ٠/٧\times ٠/٣}=\sqrt{٨٤}=٩/١٦$$
$$P(X\ge ٢٥٠)=P(X\ge ٢٤٩/٥)=P\left(\frac{X-\mu}{\sigma}\ge\frac{٢٤٩/٥-٢٨٠}{٩/١٦}\right)=P(Z\ge -٣/٣٣)$$
$$\Rightarrow P(Z\ge -٣/٣٣)=١-P(Z<-٣/٣٣)=١-٠/٠٠٠٤=٠/٩٩٩٦$$

مثال ۸: استادی به تجربه دریافته است که ۴۵ درصد از دانشجویان در درس خاصی نمرهٔ «ب» می‌گیرند اگر در ترم جاری ۲۵ نفر این درس را با او گرفته باشند احتمال گرفتن نمرهٔ «ب» را برای این تعداد محاسبه کنید.

الف) حداقل ۱۰ نفر　　　　ب) بین ۸ تا ۲۰ نفر　　　　ج) دقیقاً ۱۰ نفر

حل : چون $np=۲۵\times۰/۴۵=۱۱/۲۵>۵$ و $nq=۲۵\times۰/۵۵=۱۳/۷۵$ بزرگتر از ۵ هستند پس تقریب نرمال می‌تواند تقریب خوبی برای توزیع دوجمله‌ای باشد حال داریم :

$$\mu=np=۲۵\times۰/۴۵=۱۱/۲۵ \qquad \sigma=\sqrt{npq}=\sqrt{۲۵\times۰/۴۵\times۰/۵۵}=۲/۴۹$$

با استفاده از تصحیح پیوستگی داریم :

الف)

$$P(X\geq ۱۰)\Rightarrow P(X\geq ۹/۵)=P(Z\geq \frac{۹/۵-۱۱/۲۵}{۲/۴۹})=P(Z\geq -۰/۷)=۱-P(Z\leq -۰/۷)=۰/۷۵۸۰$$

ب)

$$P(۸\leq X\leq ۲۰)\Rightarrow P(۷/۵\leq X\leq ۲۰/۵)=P(\frac{۷/۵-۱۱/۲۵}{۲/۴۹}\leq Z\leq \frac{۲۰/۵-۱۱/۲۵}{۲/۴۹})=P(-۱/۵۱\leq Z\leq ۳/۷۱)$$
$$=P(Z\leq ۳/۷۱)-P(Z\leq -۱/۵۱)=۱-۰/۰۶۵۵=۰/۹۳۴۵$$

ج)

$$P(X=۱۰)\Rightarrow P(۹/۵\leq X\leq ۱۰/۵)=P(\frac{۹/۵-۱۱/۲۵}{۲/۴۹}\leq Z\leq \frac{۱۰/۵-۱۱/۲۵}{۲/۴۹})=P(-۰/۷\leq Z\leq -۰/۳)$$
$$=P(Z\leq -۰/۳)-P(Z\leq -۰/۷)=۰/۱۴۰۱$$

## تقریب پواسون بوسیله نرمال

وقتی که میانگین توزیع پواسون نسبتاً بزرگ ($\lambda\geq$۱۰) می شود ، می توان بجای توزیع پواسون از فرمول توزیع نرمال استفاده کرد. در این صورت میانگین و انحراف معیار توزیع نرمال برابر $\mu=\lambda$ , $\sigma=\sqrt{\lambda}$ خواهد بود.

به دلیل گسسته بودن توزیع پواسون و پیوسته بودن توزیع نرمال در هنگام حل مسائل باید از تصحیح پیوستگی همانند تقریب دوجمله‌ای به وسیلهٔ نرمال استفاده کرد.

مثال ۹: به طور متوسط در هر دقیقه ۰/۵ مشتری با توزیع پواسون به قسمت پرداخت فروشگاهی مراجعه می‌کند احتمال اینکه بیش از ۲۰ مشتری در طی نیم ساعت مراجعه کنند چقدر است.

جواب : میانگین تعداد مشتری‌ها را در ۳۰ دقیقه محاسبه می‌کنیم :

تقریب پواسون بوسیلۀ تقریب نرمال $\lambda = ۱۵ > ۱۰ \rightarrow$ $t = \frac{۳۰}{۱} = ۳۰ \Rightarrow \lambda = ۳۰ \times ۰/۵ = ۱۵$

$\mu = \lambda = ۱۵$ $\sigma = \sqrt{\lambda} = \sqrt{۱۵} = ۳/۸۷$

$$P(X > ۲۰) \Rightarrow P(X \geq ۲۰/۵) = P(Z \geq \frac{۲۰/۵ - ۱۵}{۳/۸۷}) = P(Z \geq ۱/۴۲) = ۱ - P(Z \leq ۱/۴۲) = ۱ - ۰/۹۲۲۲ = ۰/۰۷۷۸$$

## ❖ قضیۀ حد مرکزی

این قضیه بیانگر این موضوع است که اگر $X_۱, X_۲, ..., X_n$، $n$ متغیر تصادفی مستقل باشند به شرط اینکه n به اندازۀ کافی بزرگ باشد آن گاه متغیر تصادفی $Y = X_۱ + X_۲ + ... + X_n$ دارای توزیع نرمال با $\mu_Y = \sum_{i=۱}^{n} \mu_{X_i}$ و $\sigma^۲_Y = \sum_{i=۱}^{n} \sigma^۲_{X_i}$ خواهد بود که $\mu_{X_i}$ , $\sigma^۲_{X_i}$ به ترتیب میانگین و واریانس متغیر تصادفی $X_i$ هستند.

توزیع مجموع حداقل ۱۰ متغیر تصادفی $(n \geq ۱۰)$ تقریباً نرمال است در ضمن باید به این نکته توجه کنیم که $X_i$ ها می‌توانند هر توزیعی داشته باشند.

مثال۱۰: کشتارگاهی ۴۷۰ گوسفند را بدون وزن کردن آن‌ها یکجا می‌خرد وزن هر گوسفند به طور متوسط ۴۵ کیلوگرم با انحراف معیار ۴ است احتمال اینکه وزن این ۴۷۰ گوسفند بیش از ۲۲ تن باشد، چقدر است.

حل : برای حل این مسئله می توان از قضیه حد مرکزی استفاده کرد.

$$\sigma_{x_i} = ٤ \rightarrow \sigma^{٢}_{x_i} = ١٦$$

$$\sigma^{٢}_{y} = \sum_{i=١}^{n} \sigma^{٢}_{x_i} = ٤٧٠ \times ١٦ = ٧٥٢٠ \rightarrow \sigma_{y} = ٨٦/٧$$

$$\mu_{x_i} = ٤٥ \rightarrow \mu_{y} = \sum_{i=١}^{n} \mu_{x_i} = ٤٧٠ \times ٤٥ = ٢١١٥٠$$

$$p(Y > ٢٢٠٠٠) = p(z > \frac{٢٢٠٠٠ - ٢١١٥٠}{٨٦/٧}) = p(z > ٩/٨) = ١ - p(z < ٩/٨) = ١ - ١ = ٠$$