آمار و کاربرد آن در مدیریت ۱

نام منبع درس : آمار و کاربرد آن در مدیریت (جلد ۱ و ۲)

مؤلف : عادل آذر – منصور مؤمنی

تهیه کننده اسلایدها : حسن الوداری

جایگاه و هدف درس

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

چارچوب کلی درس

برای این درس ، کلیه فصل های هشتگانه جلد اول و فصل ۱۸ جلد دوم کتاب به استثناء برخی حذفیات در نظر گرفته شده است

فصل اوّل

کلیات

آشنایی با مفاهیم پایه ای علم آمار و ضرورت بکارگیری فنون و تکنیک های آن در دانش مدیریت به منظور اداره بهتر سازمان ها

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

۲- برای بررسی صحت و سقم فرضیات

۳- برای تعیین اعتبار و پایایی تحقیقات

تعریف جامعه

جامعه بزرگترین مجموعه از موجودات است که در یک زمان معین ، مطلوب ما قرار می گیرند مثل جامعه فرهنگیان ایران 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

۲- نا محدود : یعنی جامعه از یک ردیف بی انتهایی از مقادیر تشکیل شده باشد

تعریف نمونه

نمونه عبارتست از تعداد محدودی از آحاد جامعه آماری که بیان کننده ویژگی های اصلی جامعه باشد

انواع شاخص های آماری

۱- پارامتر : شاخص هایی که از طریق سرشماری (اندازه گیری تمامی عناصر جامعه آماری) بدست می آیند

۲- آماره : شاخص هایی که از طریق نمونه گیری (اندازه گیری بخشی از جامعه) بدست می آیند

روش های ناپارامتریک

آزمون هایی که مشروط به مفروضات آمار کلاسیک نیستند و کاربرد اصلی آنها در بررسی جوامع آماری غیر نرمال ، جوامع با داده های کیفی و نمونه های کوچک آماری می باشد

سیر تحول علم آمار از نظر موضوعی عبارتند از :

۱- *آمار توصیفی*

۲- *آمار استنباطی*

۳- *آمار ناپارامتریک*

آمار توصیفی

یعنی محاسبه مقادیر و شاخص های جامعه آماری با استفاده از سرشماری تمامی عناصر آن ، بعبارتی توصیف کل جامعه از طریق محاسبه پارامترها

آمار استنباطی

آماری که در آن محقق ابتدا آماره ها را محاسبه و سپس به کمک تخمین و آزمون فرض آماری ، آنها را به پارامترهای جامعه تعمیم می دهد

آمار ناپارامتریک

این نوع آمار در مقابل آمار پارامتریک یعنی آمارهای توصیفی و استنباطی دارای توزیع نرمال قرار می گیرد و برای مشاهدات فاقد توزیع آماری کاربرد دارد

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

متغیر مستقل

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

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

متغیر تعدیل کننده (واسطه ای)

یک متغیر ثانوی است که رابطه بین متغیر مستقل و متغیر وابسته را تحت تأثیر قرار می دهد

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 (Rank scale)

۳- مقیاس فاصله ای (Interval scale)

۴- مقیاس نسبی (Ratio scale)

مقیاس رسمی یا طبقه ای

محققان از این مقیاس ، صرفاً برای طبقه بندی اشیاء ، اشخاص و یا خصوصیات استفاده می کنند ، مثل استفاده از یک سری اعداد یا سمبول ها برای نام گذاری سبک های رهبری

مقیاس ترتیبی

اگر بین اسامی ایجاد شده یا طبقات حاصله ناشی از مقیاس بندی اسمی یک نوع رابطه هم وجود داشته باشد پژوهشگران از مقیاس ترتیبی استفاده می نمایند

مقیاس فاصله ای

اگر در مقیاس ترتیبی ، فاصله بین اعداد یا طبقات از یک نظم خاصی پیروی نماید (فواصل یکسان باشند) محققان از مقیاس فاصله ای برای اندازه گیری متغیرها استفاده می نمایند

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

فرضیه حدسی است زیرکانه در مورد رابطه بین دو یا چند متغیر که بصورت دقیق و روشن بیان شده و پس از آزمایش ، صحت یا سقم آن مشخص می شود

ویژگی های یک فرضیه خوب

۱- واضح و بدون ابهام

۲- بیان در قالب جملات خبری

۳- قابل تبیین (علت یابی)

۴- توضیح دهنده رابطه مورد انتظار بین متغیرها

۵- قابل آزمون بودن (آزمون پذیری)

انواع فرضیه های پژوهشی

۱- توصیفی

۲- استنباطی

۳- تک متغیره

۴- دو متغیره

۵- چند متغیره

۶- همبستگی

۷- تجربی

۸- با گروه های جور شده

۹- با گروه های مستقل

۱۰- پارامتریک

۱۱- ناپارامتریک

فرضیه توصیفی

فرضیه ای است که در مورد کل جامعه آماری تدوین شده بعبارتی ادعایی را در مورد کل جامعه آماری بیان می نماید

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

فرضیه های چند متغیره

فرضیه هایی که به ظاهر دارای دو متغیر بوده (مستقل و وابسته) ولی فرضیه مستقل آن خودش از چند متغیر دیگر تشکیل شده است

فرضیه همبستگی

به فرضیه ای گفته می شود که پژوهشگر هیچ کنترلی بر روی متغیرهای مستقل و وابسته آن ندارد ، چرا که اتفاق قبلاً رخ داده و دیگر قابل دستکاری نمی باشد

فرضیه تجربی

فرضیه ای است که محقق در آن بر روی هر دو متغیر کنترل دارد یعنی پدیده هنوز روی نداده و پژوهشگر می تواند متغیر مستقل را دستکاری نماید

فرضیه با گروه های جور شده

در این نوع فرضیه سازی ، پژوهشگران یک گروه نمونه دارند که در آن هر آزمون شونده را از لحاظ یک متغیر واحد دو بار اندازه گیری می کنند

فرضیه با گروه های مستقل

در این حالت ، 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

فرضیه های پارامتریک

فرضیه هایی هستند که در آنها از متغیرهای نسبی یا فاصله ای استفاده شده و توزیع جامعه (و یا نمونه) نرمال می باشد

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

شاخص های عددی

اعدادی هستند که به منظور بیان کمی توزیع اندازه ها از آن استفاده می شود . این شاخص ها توصیف کننده مجموعه داده ها می باشند

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 هندسی ، میانگین هارمونیک

۲- مد (نما)

۳- چارکها ؛ شامل چارک اول ، چارک دوم ، چارک سوم

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

میانگین حسابی ساده

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

فرمول

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

میانگین حسابی موزون

اگر هر یک از مشاهدات دارای تکرار باشند ، در این صورت تعداد تکرارها بعنوان وزن مشاهدات تلقی شده و آنها را با W_i *نشان می دهند*

فرمول میانگین حسابی موزون

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

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

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

میانگین هندسی ساده

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

فرمول میانگین هندسی ساده

میانگین هندسی یک رشته عدد همانند X_1، X_2، . . . ، X_N برابر است با ریشه N ام حاصلضرب آن اعداد

$$\mu_G = (X_1 \times X_2 \times \ldots \times X_N)^{\frac{1}{N}}$$

میانگین هندسی موزون

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

فرمول

$$\mu_G = (X_1^{w_1} \times X_2^{w_2} \times \ldots \times X_K^{w_K})^{\frac{1}{N}}$$

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

فرمول میانگین هارمونیک ساده

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

فرمول

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

میانگین هارمونیک موزون

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

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

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

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_1 : مقداری که ۲۵٪ مشاهدات ، پایین تر از آن است

Q_2 : مقداری که ۵۰٪ مشاهدات ، پایین تر از آن است

Q_3 : مقداری که ۷۵٪ مشاهدات ، پایین تر 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

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

فرمول تعیین محل چارک

$$CQ_a = \frac{aN}{4} + \frac{1}{2}$$

چارک مورد نظر ۱و۲و۳ =a

تعداد مشاهدات =N

پارامترهای پراکندگی

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

محاسن پارامترهای پراکندگی

۱- کمک به توصیف واقعی تر یک سری از داده ها

۲- کمک به قابلیت مقایسه دو یا چند سری از داده ها

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

دامنه تغییرات (R)

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

فرمول $R = MAX_{X_i} - MIN_{X_i}$

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_1 و بزرگتر Q_3 ندارد

فرمول دامنه میان چارکی

برای محاسبه این شاخص ، کافیست که مقادیر Q_1و Q_3 را بدست آورده و از هم کم کنیم .

$$IQR = Q_3 - Q_1$$

نیمه میان چارکی

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

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

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

انحراف متوسط از میانگین

این شاخص از تقسیم مجموع قدر مطلق انحرافات تک تک مشاهدات از میانگین شان بر تعداد مشاهدات بدست

$$A \cdot D_{\mu} = \frac{\Sigma \left| X_i - \mu_X \right|}{N}$$

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

واریانس

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

فرمول

$$\delta_X^2 = \frac{\sum (X_i - \mu_x)^2}{N}$$

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 δ_X^2

فرمول انحراف معیار

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

و یا

$$\delta_X = \sqrt{\frac{\Sigma(X_i - \mu_x)^2}{N}}$$

خواص واریانس

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

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

نیمه واریانس

$$S.V=\frac{\sum_{i=1}^{k}(X_i-\mu_x)^2}{K}$$ یعنی متوسط مجذور مقادیر نامطلوب

N= تعداد مشاهدات جامعه

K= تعداد مقادیر نامطلوب

μ_X = میانگین کل مشاهدات

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

ضریب پراکندگی

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

$$C \cdot V = \frac{\delta_X}{\mu_X}$$

δ_X= انحراف معیار مشاهدات

μ_X= میانگین مشاهدات

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

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

فرمول های محاسبه تعداد طبقات

۱- فرمول تجربی استورجس $K = 1 + 3/32 LogN$

۲- روش تقریبی $K = \sqrt{N}$

(N تعداد مشاهدات می باشد)

تعیین فاصله طبقات

فاصله طبقات از تقسیم مقدار R (دامنه تغییرات) بر مقدار محاسبه شده برای تعداد طبقات (K) به شکل زیر بدست می آید

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

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 پیوسته : برای داده های اعشاری ، یعنی مساوی بودن طول ، عرض و فاصله طبقات

۲- طبقه بندی گسسته : برای داده های غیر اعشاری ، یعنی برابر نبودن طول و عرض طبقات

مهم ترین تقریب ها در طبقه بندی گسسته

۱- تقریب ۱/۰

۲- تقریب ۵/۰

۳- تقریب ۱ (واحد)

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

طبقه بندی مشاهدات ناپیوسته

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

توزیع فراوانی نسبی

چنانچه در جدول طبقه بندی داده ها بجای فراوانی مطلق (F_i) از فراوانی نسبی (f_i) استفاده شود ، به آن توزیع فراوانی نسبی گویند

فرمول فراوانی نسبی

$$\text{فراوانی نسبی هر طبقه} = \frac{\text{فراوانی مطلق آن طبقه}}{\text{تعداد کل مشاهدات (فراوانی ها)}} \longrightarrow f_i = \frac{F_i}{N}$$

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

فرمول محاسبه نماینده یا متوسط طبقات

$$\text{متوسط یا نماینده هر طبقه} = \frac{\text{(حد بالا + حد پایین) طبقه مورد نظر}}{۲}$$

توزیع فراوانی تجمعی

اگر در جدول طبقه داده ها ، بجای فراوانی های مطلق و نسبی از فراوانی تجمعی استفاده شود ، به جدول بدست آمده ، توزیع فراوانی تجمعی گویند

فراوانی تجمعی طبقه

فـراوانی تجـمعی هر طـبقه ، عـبارتست از مجموع فراوانی های مطلق از اولین طبقه تا طبقه مورد نظر که آن را با FC_i نشان می دهند

$$Fc_i = \sum_{i=1}^{i} F_i$$

فراوانی نسبی تجمعی

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

یعنی

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

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

محاسن نمودارها

استفاده از نمودارها 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

مهم ترین نمودارهای کمی

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

مدرج کردن 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

نمودار بافت نگار

پس از مدرج کردن محورها بر روی حدود واقعی (کرانه های هر طبقه) مستطیلی عمودی رسم می شود که مساحت آن مساوی با فراوانی نسبی آن طبقه می باشد

نمودار چند ضلعی

نموداریست که متناظر با هر نماینده طبقه در محور افقی و فراوانی آن در محور عمودی ، یک نقطه در صفحه مختصات ایجاد و به هم وصل می شوند

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

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

کاربردهای نمودار فراوانی تجمعی

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

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

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

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

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

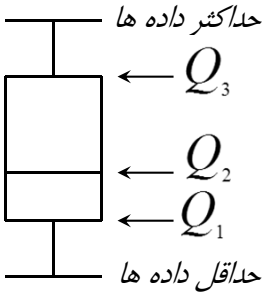
نمودار جعبه ای

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



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*

نمودار ستونی

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

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

۳- تقسیم مساحت دایره بر حسب S_i ها

۴- نوشتن نوع و درصد مشاهدات بر روی دایره

فرمول مساحت هر قطاع

برای پیدا کردن مساحت هر قطاع از دایره از فرمول زیر استفاده می شود

$$S_i = 360^{\circ} \times f_i$$

یعنی فراوانی نسبی هر مشاهده

به عدد ۳۶۰ ضرب می شود

محورهای نمودار پاره تو

این نمودار دارای سه محور است :

۱- محورافقی : نوع موضوعات

۲- محور عمودی : فراوانی مطلق موضوعات

۳- محور سوم (روبروی محور عمودی) : فراوانی نسبی تجمعی موضوعات

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- و در بررسی نحوه توزیع درآمد و توزیع پرسنل مؤسسات

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 ؟

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

فرمول میانگین به روش مستقیم

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

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

F_i = فراوانی مطلق

X_i = متوسط طبقات

N = کل مشاهدات

فرمول میانگین به روش غیرمستقیم (کد گذاری)

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

A = عدد دلخـواه

d_i = کد هر طـبقه

I = فاصله طبقات

نقش Aدر فرمول میانگین

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

فرمول d_i

d_i *که کد هر طبقه است ، به شکل زیر قابل می باشد*

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

موارد استفاده از فرمول میانگین به روش کد گذاری

زمانی 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

۲- فراوانی مطلق

۳- نماینده طبقات

۴- حاصلضرب فراوانی در نماینده طبقه

۵- کد طبقات

۶- حاصلضرب فراوانی در کد طبقه

علت استفاده از علامت$\cong$

از علامت $\cong$ (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

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

فرمول مد در داده های طبقه بندی شده

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

اجزاء تشکیل دهنده فرمول مد

L_{Mo}= حد پایین واقعی طبقه مد دار

d_1 = فراوانی مطلق طبقه مدار منهای فراوانی طبقه ماقبل

d_2 = فراوانی مطلق طبقه مدار منهای فراوانی طبقه ما بعد

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

مراحل محاسبه چندکها

۱- اضافه کردن ستون فراوانی تجمعی به جدول

۲- پیدا کردن محل چارک مورد نظر با استفاده از فرمول مربوطه

۳- پیدا کردن طبقه چارک دار و استفاده از فرمول چارک

فرمول تعیین محل چارک

$$CQ_a = \frac{aN}{4}$$

a = شماره چارک (۱، ۲ یا ۳)

N = تعداد کل مشاهدات

فرمول چارک در داده های طبقه بندی شده

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

اجزاء تشکیل دهنده فرمول چارک

Q_a = مقدار چارک

L_{Q_a} = حد پایین واقعی طبقه چارک دار

Fc_{i-1} = فراوانی تجمعی طبقه ما قبل طبقه چارک دار

F_i = فراوانی مطلق طبقه چارک دار

مراحل محاسبه دهکها (D_a)

۱- اضافه کردن F_C به جدول

۲- پیدا کردن محل دهک با استفاده از $C_{D_a} = \frac{aN}{10}$

۳- محاسبه دهک با استفاده از مراحل قبلی و فرمول چارک

فرمول دهک

$$D_a \cong L_{D_a} + \left(\frac{\frac{aN}{4} - Fc_{i-1}}{F_i}\right) I$$

اجزاء فرمول دهک

L_{D_a} = حد پایین واقعی طبقه دهک دار

F_i = فراوانی مطلق طبقه دهک دار

Fc_{i-1} = فراوانی تجمعی طبقه ما قبل طبقه دهک دار

I = فاصله طبقات

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

فرمول صدکها

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

از $C_{P_a} = \frac{aN}{100}$ برای پیدا کردن محل صدک استفاده می شود

نکته مهم در محاسبه صدکها

اگر در توزیع فراوانی طول و عرض طبقات مساوی نباشد در این صورت ، باید کرانه ها را محاسبه نموده و از حد پایین آنها استفاده نمود

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 (δ_x^2)و واریانس تصحیح شده (δ_c^2)

فرمول انحراف متوسط از میانگین

این فرمول در داده های طبقه بندی به شکل زیر می باشد در این فرمول F_i فراوانی طبقه i ام می باشد

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

دامنه میان چارکی و انحراف چارکی

از این پارامترهای پراکندگی زمانی استفاده می شود که دنباله های توزیع نا معین و باز باشد (در این حالت محاسبه میانگین و واریانس امکان پذیر نیست)

فرمول های واریانس

۱- روش مستقیم

فرمول اول $\delta_x^2 = \frac{\sum F_i (X_i - \mu_x)^2}{N}$

فرمول دوم $\delta_x^2 = \frac{\sum F_i X_i^2}{N} - \mu_x^2$

۲- روش غیر مستقیم $\delta_x^2 = I^2 \left[\frac{\sum F_i d_i^2}{N} - \left(\frac{\sum F_i d_i}{N} \right)^2 \right]$

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

فرمول میانگین حسابی جامعه کل

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

N = تعداد مشاهدات هر جامعه

μ = میانگین هر جامعه

فرمول واریانس جامعه کل

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

اجزاء واریانس جامعه کل

δ_i^2 = واریانس جامعه i ام

N = تعداد مشاهدات جامعه کل

N_i = تعداد مشاهدات جامعه i ام

μ = میانگین جامعه کل

μ_i = میانگین جامعه i ام

پارامترهای تعیین انحراف از قرینگی

در هنگام مقایسه دو یا چند جامعه ، در صورت مساوی بودن پارامترهای مرکزی و پراکندگی ، این پارامترها با بهره گیری از ضریب چولگی کارساز خواهند بود

انواع حالات توزیع ها

۱- متقارن (نرمال) : مد = میانه = میانگین

۲- چولـه به راسـت : مد < میانه < میانگین

۳- چـولـه بـه چـــپ : مد > میانه > میانگین

مقادیر مختلف ضریب چولگی (SK)

۱- صفر : در صورت متقارن بودن توزیع جامعه

۲- مثبت : در صورت چوله به راست بودن توزیع جامعه

۳- منفی : در صورت چوله به چپ بودن توزیع جامعه

مفهوم چولگی

اگر دم توزیع جامعه به سمت راست باشد ، توزیع را چوله به راست و در صورت عکس ، آن را چوله به چپ می نامند

چوله به راست　　چوله به چپ

تفسیر مقادیر SK

۱- $|SK| \leq 0/1$ ، جامعه تقریباً نرمال

۲- $0/1 < |SK| \leq 0/5$ ، تفاوت اندک با توزیع نرمال

۳- $|SK| > 0/5$ ، تفاوت فاحش با توزیع نرمال

فرمول های محاسبه ضریب چولگی (SK)

۱- ضریب چولگی گشتاوری

۲- ضریب های چولگی پیرسون

۳- ضریب های چولگی چندکی

ضریب چولگی گشتاوری

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

گشتاور مرتبه سوم به مبدأ میانگین $r_3 = \frac{\sum F_i (X_i - \mu_x)^3}{N}$

δ_x انحراف معیار

ضریب چولگی پیرسون

$$S \cdot K_1 = \frac{(\mu_X - Mo)}{\delta_x}$$ فرمول شماره ۱

$$S \cdot K_2 = \frac{3(\mu_X - Md)}{\delta_x}$$ فرمول شماره ۲

ضرایب چولگی چندکی

ضریب چولگی چارکی $SK_Q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$

ضریب چولگی صدکی $SK_P = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$

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

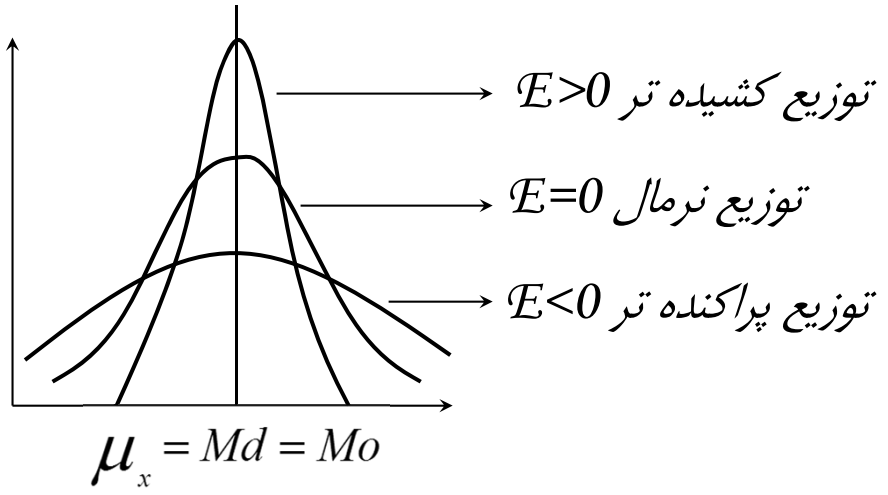
انواع توزیع به لحاظ کشیدگی و مقدار ضریب (Eآن

۱- مساوی توزیع نرمال ($E=0$)

۲- بلندتر از توزیع نرمال ($E>0$)

۳- کوتاه تر از توزیع نرمال ($E<0$)

مقایسه انواع کشیدگی



تفسیر ضریب کشیدگی (E)

۱- توزیع نرمال $|E| \leq 0/1$

۲- توزیع نسبتاً بلندتر از نرمال $0/1 < |E| \leq 0/5$

۳- توزیع کاملاً کشیده تر از نرمال $|E| > 0/5$

انواع ضرایب کشیدگی

۱- ضریب کشیدگی گشتاوری ؛ با استفاده از گشتاور مرتبه چهارم به مبدأ میانگین

۲- ضریب کشیدگی چندکی ؛ با استفاده از انحراف چارکی و صدکهای دهم و نودم

فرمول ضریب کشیدگی گشتاوری

$$E = \frac{r_4}{\delta_x^4} - 3 \quad \rightarrow \quad r_4 = \frac{\sum F_i (X_i - \mu_x)^4}{N}$$

عدد ثابت $= 3$

ضریب کشیدگی چندکی

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

مخصوص توزیع هایی که بالاجبار با استفاده از چندکها توصیف می شوند

فصل پنجم
مبادی احتمال

در این فصل دانشجو با برخی از مفاهیم احتمال و انواع احتمالات آشنا می شود

مفهوم احتمال ( ) P

احتمال یعنی شانس وقوع یک پیشامد خاص و احتمال وقوع یک پیشامد برابر است با نسبت دفعاتی که پیشامد خاصی در تکرارهای زیاد رخ می دهد

احتمال عینی و ذهنی

- احتمال ذهنی ، متغیر و وابسته به نظر اشخاص است
- احتمال عینی ، ثابت و مقدار آن از قبل مشخص است و به عقاید اشخاص بستگی ندارد

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

فضای نمونه محدود و نامحدود

محدود – یعنی این که فضای نمونه تعداد کمی عضو داشته باشد

نامحـدود – یعنی اینکه فضـای نمونه آزمایش (تعداد اعضاء آن) نامتناهی است

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

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 A و B و C و . . . نشان می دهند

پیامدهای مقدماتی هم شانس

پیامدهای مقدماتی یا پیشامدهای اولیه هم شانس یعنی این که تمام پیشامدهای اولیه در آزمایشی دارای شانس وقوع برابر باشند یعنی وجود نوعی تقارن در آزمایش

احتمال یک پیشامد در پیامدهای مقدماتی هم شانس

احتمال وقوع پیشامدی مثل A برابر می شود با تعدادهای عضو های پیشامد A به تعداد عضوهای فضای نمونه

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

احتمال یک پیشامد در پیامدهای مقدماتی غیر هم شانس

برای پیشامدی مثل A ⟵ فراوانی نسبی پیشامد A در N تکرار = $P(A)$

بشرطی که N به سمت بی نهایت میل کند

خواص اولیه احتمال

1- $0 \le P(A) \le 1$

2- $P(S) = 1$

برای هر پیشامدی مثل A چه هم شانس و چه غیر هم شانس

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

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 :

$$n_1 \times n_2 \times \ldots \times n_K$$

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)

جایگشت (ترتیب)

یعنی تعداد طرقی که می توان r شی را از بین n شی انتخاب نمود بطوریکه $r \leq n$ 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

۳- تعداد جایگشت های r شی انتخابی از بین N شی متمایز

فرمول تعداد کل جایگشت های Nشی متمایز

۱- در صورت ردیفی بودن بشکل $n!=n\times(n-1)\times\dots3\times2\times1$

۲- در صورت دایره ای بودن $(n-1)!=(n-1)(n-2)\times\dots3\times2\times1$

جایگشت های Nشی نا متمایز

$$\frac{n!}{n_1 ! n_2 ! \ldots n_K !}$$

شروط

۱- از n شی ، n_1 تای آنها از یک نوع

n_2 تای آنها از نوع دیگر و ...

۲- $n_1 + n_2 + \ldots + n_K = n$

جایگشت های rشی از بین nشی

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

شروط : ۱- r و n هر دو متمایز

۲- $r < n$

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

ترکیب

تعداد طرق انتخاب r شی متمایز از بین n شی بشرطی که ترتیب قرار گرفتن اشیاء باعث افزایش تعداد طرق نگردد

فرمول

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

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

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

افرازهای مرتب

فرمول

$$\binom{n}{n_1, n_2, \cdots, n_K} = \frac{n!}{n_1!\, n_2!\, \cdots\, n_K!}$$

ویژگی ها :

۱- تفکیک n شی به گونه ای خاص

۲- در حکم یک مجموعه بودن هر ترکیب

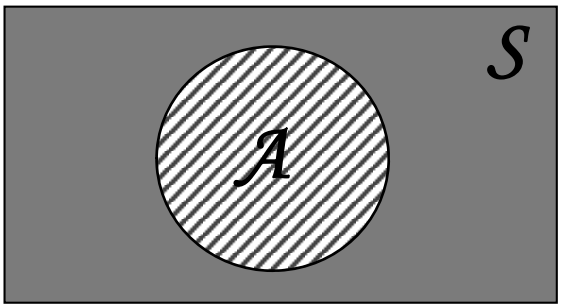
۳- مهم نبودن ترتیب اشیاء در هر زیر مجموعه

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 Aدر نمودار ون

این احتمال برابر است با سطحی که پیشامد A از S (فضای نمونه) اشغال کرده است

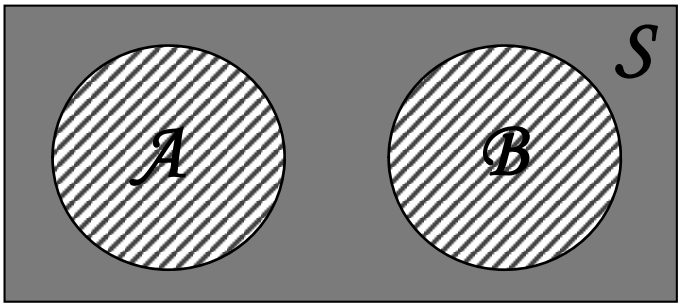


دو پیشامد نا سازگار

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

نمودار نشان می دهد که پیشامدهای A و B هیچ وجه اشتراکی با هم ندارند

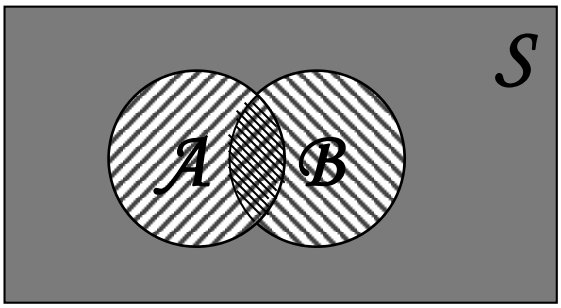


دو پیشامد سازگار

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 سازگار

محل تلاقی دو پیشامد ، نقطه مشترک آنهاست (جائیکه دو بار هاشور خورده است)



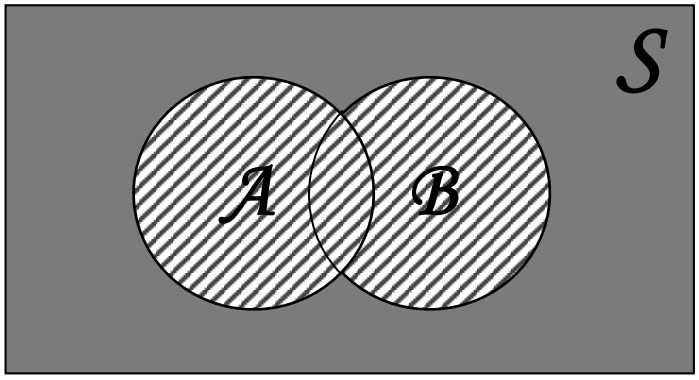
اجتماع دو پیشامد

اجتماع دو پیشامدی مثل A و B ، مجموعه تمام عضوهایی است که در A یا در B یا هم در A و هم در B قرار دارند

علامت و معنی اجتماع

اجتماع دو پیشامد A و 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 $A \cup B$

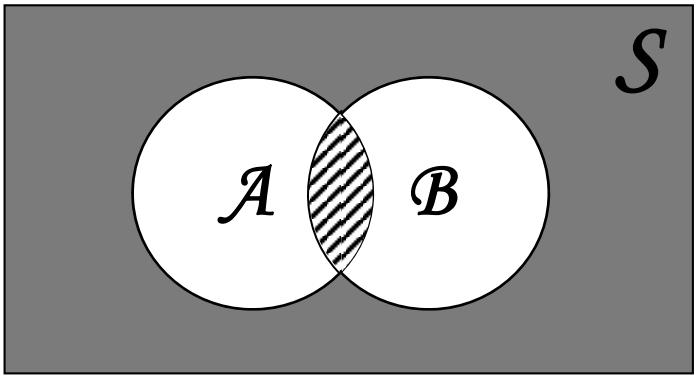


اشتراک دو پیشامد

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

نمودار ون برای اشتراک دو پیشامد

یعنی این که هم پیشامد A و هم پیشامد B رخ داده است

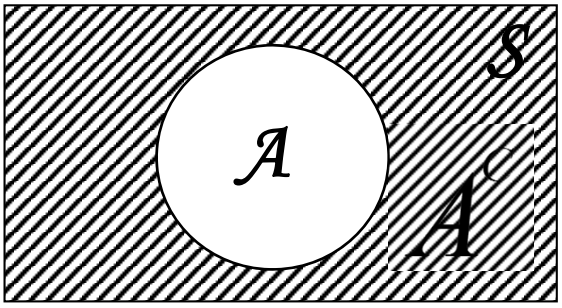


متمم یک پیشامد

متمم پیشامدی مثل A که با A^C نشان داده می شود مجموعه تمام عضوهایی است که در فضای نمونه است ولی در خود پیشامد A نیست

نمودار ون برای A^C

وقوع متمم A (A^C) به معنی عدم وقوع پیشامد A می باشد



قاعده متمم گیری

$$A \cup A^C = S \Rightarrow P(S) = 1 \Rightarrow P(A) + P(A^C) = 1$$
$$\Rightarrow P(A^C) = 1 - P(A) \quad , P(A) = 1 - P(A^C)$$

قاعده جمع پیشامدها

$$1-\ P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

برای دوپیشامد سازگار

$$2-\ P(A\cup B)=P(A)+P(B)$$

برای دوپیشامد ناسازگار

احتمال شرطی

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

شروط : ۱- *وقوع* A *به* B *مربوط بوده*

۲- B *قبلاً رخ داده*

۳- $P(B)\neq 0$

احتمال Bبه شرط A

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

شروط : ۱- A قبلاً رخ داده

۲- $P(A)\neq 0$

قانون ضرب احتمالات

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

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

یا

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

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

احتمال برای پیشامدهای مستقل

چون A و B هیچ تأثیری بر روی هم ندارند برای محاسبه احتمال اشتراک آنها بشکل زیر عمل می شود :

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

شروط مربوط به احتمالات

شرط ناسازگار بودن دو پیشامد

$$A \cap B = \phi \Rightarrow P(A \cap B) = 0$$

شرط مستقل بودن دو پیشامد

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

حالات مختلف دو پیشامد نسبت به هم

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

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

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

فصل ششم

توابع احتمال گسسته

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

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

دامنه و حوزه متغیر تصادفی

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

انواع متغیر تصادفی

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

۲- متغیر تصادفی پیوسته ؛ با تعداد مقادیر ممکن نامتناهی و غیر قابل شمارش

نمایش متغیرهای تصادفی

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

دامنه و حوزه تابع احتمال

تابع احتمال تابعی است که دامنه آن مقادیر ممکن متغیر تصادفی و حوزه آن احتمالات مربوط به هر مقدار از متغیر تصادفی 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

امید ریاضی

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

فرمول امید ریاضی

امید ریاضی یک متغیر تصادفی از حاصل جمع ضرب هر متغیر تصادفی در مقدار احتمال خودش بدست می آید

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

واریانس متغیر تصادفی X

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

فرمول های واریانس متغیر تصادفی X

$$1- V(X) = \sum (X-\mu)^2 f(X)$$

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

$$\mu = E(X)$$

تابع احتمال توأم

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

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

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

کواریانس

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

انواع رابطه بین دو متغیر

۱- رابطه مستقیم: حرکت هم جهت متغیرها

۲- رابطه معکوس: حرکت بصورت خلاف جهت هم

۳- عدم وجود رابطه: عدم تأثیر متغیرها بر هم

مقادیر مختلف کواریانس

۱- مثبت: در صورت وجود رابطه مستقیم بین متغیرها

۲- منفی: در صورت وجود رابطه معکوس بین متغیرها

۳- صفر: در صورت عدم وجود رابطه بین آنها (X و Y)

فرمول کواریانس

$$1- \ COV(X,Y) = E(X-\mu_X)(Y-\mu_Y)$$

یا

$$2- \ COV(X,Y) = E(XY) - E(X)E(Y)$$

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

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

$$f(x_i, y_j) = f(x_i) \times f(y_j)$$

رابطه استقلال و کواریانس

x و y مستقل اند $\Leftarrow$ $COV(X,Y)=0$ $\not\Rightarrow$

اگر x و y مستقل باشند ، کواریانشان حتماً صفر است ولی عکس قضیه همیشه صادق نیست

توزیع برنولی

ویژگی ها :

۱- آزمایش فقط یک بار صورت می گیرد

۲- فقط دو پیامد ممکن دارد

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

۴- آزمایش ها مستقل از یکدیگر انجام می شوند

مفاهیم pو qدرتوزیع 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

نمونه گیری و توزیع برنولی

نمونه گیری های بدون جایگزینی از جامعه باعث متغیر شدن احتمال موفقیت و شکست در آزمایش ها شده و توزیع را از حالت برنولی خارج می کند

نمونه گیری از جامعه بزرگ

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

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

فرمول توزیع دو جمله ای

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

$$x = 0,1,2,\ldots,n$$

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

مقدار p و نوع توزیع

۱- اگر $p = 0/5$ باشد ، توزیع متقارن

۲- اگر $p > 0/5$ باشد ، توزیع چوله به چپ

۳- و اگر $p < 0/5$ باشد ، توزیع چوله به راست است

جداول توزیع دو جمله ای

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

میانگین و واریانس توزیع دو جمله ای

1- $E(X) = np$

2- $V(X) = npq$

n و p و q پارامترهای توزیع دو جمله ای هستند

توزیع فوق هندسی

پیدا کردن احتمال این که ، از بین N شی که K تای آنها واجد شرایط است ، n شی را برگزینیم بطوریکه X تای آن واجد شرایط باشد

فرمول توزیع فوق هندسی

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

استفاده از توزیع دو جمله ای بعنوان تقریب

اگر $n < \%5N$ باشد می توان در محاسبه میزان احتمالات توزیع فوق هندسی از توزیع دوجمله ای کمک گرفت

علت تقریب

توزیع فوق هندسی برای نمونه گیری های بدون جایگذاری است ولی وقتی N بزرگ و n کوچک است p تقریباً ثابت می ماند ($p = \frac{k}{N}$) و حالت توزیع دو جمله ای بخود می گیرد

توزیع پواسون

اگر n به سمت بی نهایت و p به سمت صفر میل کند و در عین حال مقدار np ثابت بماند ، می توان بجای توزیع دو جمله ای از توزیع پواسون استفاده نمود

فرمول توزیع پواسون

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

$\lambda = np$ بعنوان پارامتر توزیع

و $e \cong 2/718$

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

$$E(X)=\lambda$$

$$V(X)=\lambda$$

کاربردهای توزیع پواسون

۱- بعنوان تقریب یا برآورد کننده میزان احتمال توزیع های دو جمله ای تحت شرایط خاص

۲- محاسبه احتمالات مربوط به تعداد مراجعات به سیستم با λ در واحد زمان (t)

توزیع پواسون برای تعداد مراجعات

فرمول $P(X = x) = \dfrac{e^{-\lambda t}(\lambda t)^{x}}{x!}$

$x = 0,1,2,\ldots$

$\lambda t =$ میانگین مراجعات در واحد زمانی معین

فصل هفتم

توابع احتمال پیوسته

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

احتمال در توابع 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

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

نقش علامت مساوی در احتمالات پیوسته

$$P(a < X < b) = P(a \le X < b) = P(a < x \le b) = P(a \le X \le b)$$

پس علامت مساوی در این توزیع ها نقشی ایفاء نمی کند

امید ریاضی متغیر تصادفی پیوسته

یعنی ضرب متغیر تصادفی در تابع چگالی خود و سپس انتگرال گیری به ازای مقادیر ممکن متغیر

$$E(X) = \int_{-\infty}^{+\infty} X \cdot f(X)\,dx$$

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

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

فصل هشتم

توزیع نرمال

در این فصل دانشجویان با مهم ترین و کاربردی ترین نوع توزیع از زیر مجموعه توزیع های پیوسته آشنا می شوند

تعریف توزیع نرمال

متغیر تصادفی پیوسته X در صورت داشتن تابع چگالی زیر دارای توزیع نرمال است

$$f(X)=\frac{1}{\sqrt{2\pi\delta^2}}e^{-\frac{1}{2}[(X-\mu)/\delta]^2}$$

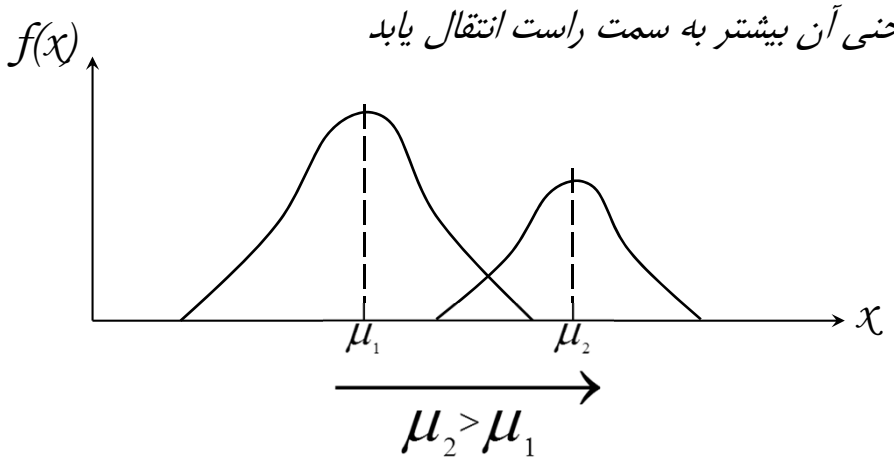
پارامترهای توزیع نرمال

$X \approx N(\mu, \delta)$

μ (میانگین) و δ (انحراف معیار) دو پارامتر توزیع نرمال بوده و با مشخص بودن آنها ، منحنی توزیع قابل ترسیم می باشد

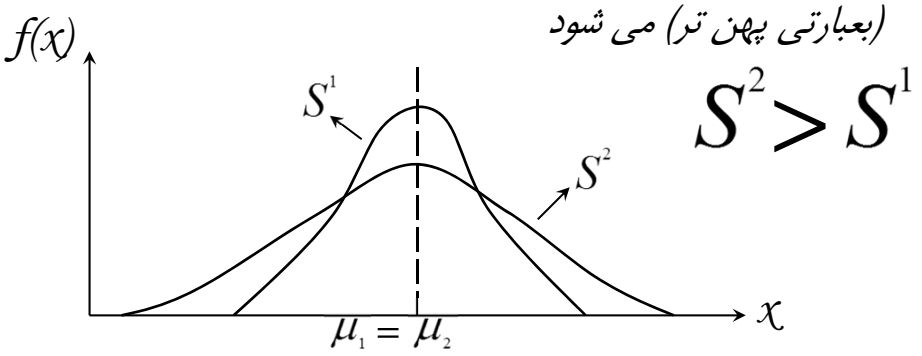
نقش میانگین در منحنی توزیع نرمال

در یک توزیع نرمال هر قدر میانگین افزایش یابد ، باعث می شود که منحنی آن بیشتر به سمت راست انتقال یابد



نقش انحراف معیار در توزیع نرمال

هر قدر انحراف معیار افزایش یابد ، منحنی توزیع نرمال کوتاه تر (بعبارتی پهن تر) می شود



$$S^2 > S^1$$

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 $X = \mu$ می باشد

۴- تابع حول میانگین ، متقارن است

ادامه ویژگی های توزیع نرمال

۵- میانگین و واریانس X به ترتیب μ و δ^2 می باشد

۶- منحنی در محور X ها ، هیچ گاه به صفر نمی رسد

۷- میانگین ، میانه و مد با هم برابرند

ادامه ویژگی ها

۸- احتمال x با توجه به انحراف معیارهای مختلف بشرح ذیل است :

$$P(\mu - \delta \le x \le \mu + \delta) = 0/683$$

$$P(\mu - 2\delta \le x \le \mu + 2\delta) = 0/954$$

$$P(\mu - 3\delta \le x \le \mu + 3\delta) = 0/997$$

توزیع نرمال استاندارد

یعنی استاندارد کردن متغیر $X \approx N(\mu, \delta)$ با استفاده از متغیر نرمالی که میانگین آن صفر و واریانس اش یک است و سپس استفاده از جدول مربوطه

نحوه تبدیل متغیر Xبه متغیر نرمال استاندارد z

با کم کردن میانگین از متغیر x و تقسیم نتیجه آن بر انحراف معیار ، z بدست می آید

$$z = \frac{X - \mu}{\delta}$$

روش های استفاده از جدول توزیع نرمال استاندارد

۱- استفاده مستقیم : مقدار Z مشخص است و احتمال آن را بدست می آوریم

۲- استفاده معکوس : احتمال Z مشخص است و مقدار آن را بدست می آوریم

استفاده معکوس از جدول

پس از پیدا کردن Z با توجه به میزان احتمال اعلام شده ، مقدار X را با استفاده از این رابطه پیدا می کنیم

$$X = \mu + \delta z$$

تقریب توزیع دو جمله ای بوسیله توزیع نرمال

اگر n بزرگ و p در نزدیکی های صفر یا یک نباشد ، تقریب نرمال با پارامترهای $\mu = np$ و $\delta = \sqrt{npq}$ تقریب خوبی برای توزیع دو جمله ای است

تصحیح پیوستگی

چون توزیع دو جمله ای ، توزیعی گسسته و توزیع نرمال ، توزیعی پیوسته است بنابراین هنگام استفاده از تقریب نرمال باید تصحیح پیوستگی صورت پذیرد (افزایش دامنه)

تقریب پواسون بوسیله نرمال

وقتی که میانگین توزیع پواسون نسبتاً بزرگ ($\lambda \geq 10$) می شود ، می توان بجای توزیع پواسون از فرمول توزیع نرمال استفاده کرد

پارامترهای توزیع نرمال هنگام برآورد تقریبی پواسون

میانگین (μ) و انحراف معیار (δ) عبارت خواهند بود از :

$$\mu = \lambda \quad , \delta = \sqrt{\lambda}$$

ضمناً استفاده از تصحیح پیوستگی نیز لازم است

فصل نهم

نظریه تصمیم

در این فصل دانشجویان با فرآیند تصمیم ، انواع شرایط تصمیم گیری و معیارهای مختلف اخذ تصمیم در شرایط عدم اطمینان آشنا می شوند

شش گام در نظریه تصمیم

۱- تعریف روشن مسأله

۲- تعیین گزینه های ممکن

۳- تعیین پیامدهای ممکن

۴- تعیین بازده ناشی از هر ترکیب

۵- انتخاب یک مدل کمی

۶- بکارگیری مدل و اتخاذ تصمیم

شکل کلی جدول بازده یا جدول تصمیم

حالات / طبیعت / گزینه ها	S_1	S_2	$\cdots$	S_H
a_1	M_{11}	M_{12}	$\cdots$	M_{1H}
a_2	M_{21}	M_{22}	$\cdots$	M_{2H}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_K	M_{K1}	M_{K2}	$\cdots$	M_{KH}

انواع شرایط تصمیم گیری

۱- تصمیم گیری تحت شرایط اطمینان کامل

۲- تصمیم گیری در شرایط عدم اطمینان

۳- تصمیم گیری در شرایط ریسک

تصمیم گیری در شرایط اطمینان کامل

در این شرایط تصمیم گیرندگان با اطمینان ، پیامدهای هر گزینه یا تصمیمی را می دانند ، بنابرین گزینه ای را انتخاب می کنند که منافع آنها را حداکثر نماید

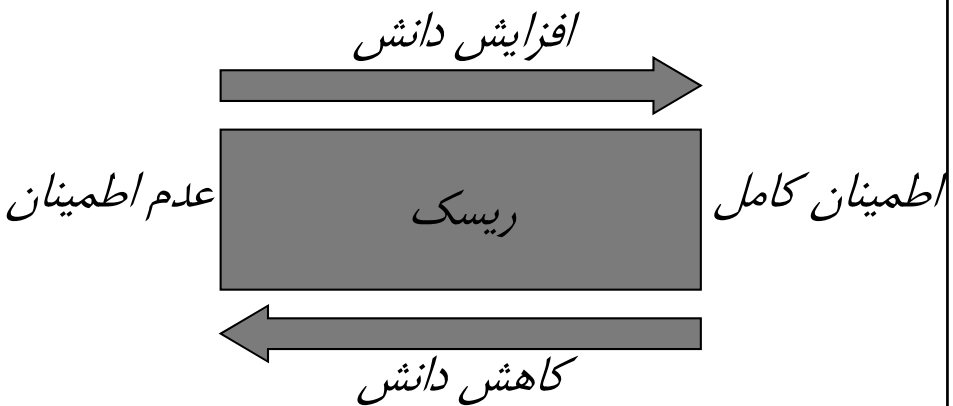
تصمیم گیری در شرایط عدم اطمینان

در این نوع تصمیم گیری ، تصمیم گیرنده نمی داند کدامیک از حالات طبیعت رخ می دهد در ضمن نمی تواند احتمال وقوع هر یک را مشخص کند

تصمیم گیری در شرایط ریسک

در این شرایط تصمیم گیرنده نمی داند کدام یک از حالات طبیعت واقع می شود ، ولی می تواند احتمال وقوع هر یک را مشخص کند

شکل انواع شرایط تصمیم گیری



مهم ترین معیارهای تصمیم گیری در شرایط عدم اطمینان

۱- حداکثر حداکثر

۲- حداکثر حداقل

۳- احتمالات مساوی

۴- واقعگرایی

۵- حداقل حداکثر غبن

معیار حداکثر حداکثر

معیار خوش بینانه ای است و تصمیم گیرنده تصور می کند که همه گزینه ها بیشترین بازدهی خود را خواهند داشت ، لذا سعی می کند از بین بهترها ، بهترین را انتخاب نماید

معیار حداکثر حداقل

به معیار بد بینانه معروف است و تصمیم گیرنده ، ابتدا بدترین (حداقل ترین) حالت هر گزینه را انتخاب و سپس از بین آنها بهترین (بازدهی حداکثر) را بر می گزیند

معیار احتمالات مساوی (لاپلاس)

این معیار شانس حالات مختلف طبیعت را یکسان فرض کرده و بدین سبب بدنبال گزینه ای می گردد که متوسط بازده آن از بقیه گزینه ها بیشتر باشد

معیار واقع گرایی (هورتیز)

بر اساس این معیار یک تصمیم گیرنده منطقی نه زیاد خوش بین و نه زیاد بد بین است و سعی می کند تعادلی بین معیارهای خوشبنیانه و بد بنیانه ایجاد نماید

فرمول معیار واقع گرایی هر گزینه

معیار واقع گرایی گزینه i =

(حداقل بازده گزینه i)(α −۱) + (حداکثر بازده گزینه i) α

معیارحداقل حداکثر غبن

جدولی بنام « جدول غبن » تشکیل و حداکثر هر سطر را مشخص می کنیم سپس حداقل آنها را تعیین و گزینه متناظر با آن را انتخاب می نماییم

تصمیم گیری در شرایط ریسک

در صورتی که تصمیم گیرنده بتواند احتمال وقوع حالات مختلف طبیعت را برای مسأله تصمیم ، تعیین کند ، تصمیم گیری از نوع ریسک خواهد بود

معیار تصمیم گیری در شرایط ریسک

مهم ترین معیار در این شرایط ، « معیار ارزش پولی مورد انتظار » یا همان EMV است . ارزش پولی مورد انتظار همان امید ریاضی بازده است

درخت تصمیم

زمانی استفاده می شود که در مسأله ای باید تصمیمات « پیچیده » و « متوالی » اتخاذ شود و در آن از معیار EMV و نمادهایی مثل دایره و مربع استفاده می شود

تحلیل حساسیت

در اغلب موارد تصمیم گیرنده ، موقع استفاده از معیار EMV ، درباره بازده ها و احتمالات نامطمئن است ، لذا از تحلیل حساسیت برای تعیین دامنه مطلوب برای بهترین گزینه استفاده می کند

ارزش مورد انتظار با اطلاعات کامل (EVPI)

EVPI بازده مورد انتظار است اگر اطلاعات کامل ، قبل از اخذ تصمیم موجود باشند

ارزش مورد انتظار اطلاعات کامل

تفاوت بین ارزش مورد انتظار با اطلاعات کامل و ارزش مورد انتظار بهترین گزینه است ، بعبارتی :

ارزش مورد انتظار اطلاعات کامل = $EVPI - EMV_{\max}$