بسم الله

آمار و کاربرد آن در مدیریت.

جزوه شماره ۲

توزیع آماری:

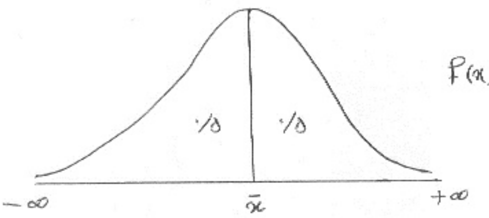
در آمار مقدماتی یکی از اولین اقداماتی که بر روی داده‌ها برای طبقه‌بندی داده‌ها بکار برده می‌شود تشکیل جدول فراوانی است که در یک ستون گروه‌بندی داده‌ها و در یکی دیگر از ستون‌ها فراوانی نسبی داده‌ها که عبارت از فراوانی هر گروه تقسیم بر تعداد کل نمونه است، نمایش داده می‌شود. چنانچه این نسبت را با تعریف احتمال ($P(A) = \frac{n(A)}{N}$) مقایسه کنیم، مشاهده می‌کنیم فراوانی نسبی هر گروه در جدول فراوانی به تعبیری، تخمینی از احتمال قرار گرفتن داده‌ها در گروه خاصی است به عبارت دیگر یک جدول فراوانی مقادیر مختلف متغیر، به همراه احتمالات مربوطه (تخمینی از احتمال) آورده شده است. که می‌توان همین اطلاعات را به صورت نمودار (مثل هیستوگرام) یا یک فرمول آماری ارائه داد. به این جدول، نمودار و یا فرمول آماری که مقادیر یک متغیر به همراه احتمالات مربوطه را با ارائه می‌دهد یک توزیع آماری می‌گوئیم. که یکی از بهترین روشهای برای نمایش توزیع آماری داده‌ها، استفاده از نمودار یا منحنی (بخصوص در رابطه با متغیرهای تصادفی پیوسته) می‌باشد. بنابراین آنچه در توزیعهای آماری اهمیت دارد، مقدار متغیر و احتمال مربوطه است که معمولاً مقدار متغیر معلوم، امّا احتمال مربوطه مجهول است. بنابراین در بیشتر مواقع بایستی برای مقادیر یک متغیر احتمال مربوطه را حساب کنیم. در آمار مقدماتی انواع توزیعهای آماری که شامل توزیعهای آماری گسسته (برای متغیرهای تصادفی قابل شمارش) و توزیعهای آماری پیوسته می‌باشد را خوانده‌ایم. در این بخش به توضیح درخصوص برخی از توزیعهای آماری پیوسته که درخصوص متغیرهای تصادفی پیوسته است می‌پردازیم. یکی از مهمترین این توزیعها توزیع نرمال است که با جزئیات بیشتری به شرح آن می‌پردازیم

توزیع نرمال :

توزیع نرمال به خاطر اینکه یکی از کاربردی ترین توزیع های آماری است و درصد بالایی از متغیرهای تصادفی پیوسته به نوعی دارای توزیع نرمال هستند، از اهمیت ویژه ای برخوردار است. توزیع های آماری پیوسته و از جمله توزیع نرمال را، با استفاده از منحنی آنها نمایش می دهند که منحنی توزیع نرمال به شکل زیر است و مهمترین خواص آن شرح داده می شود

تابع چگالی توزیع نرمال :

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



خواص توزیع نرمال :

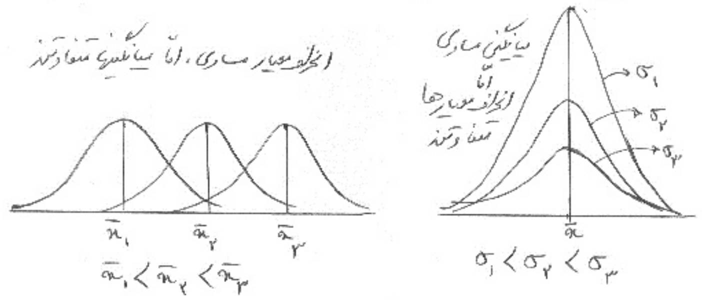
- منحنی توزیع نرمال از $-\infty$ تا $+\infty$ امتداد دارد.
- منحنی توزیع نرمال یک منحنی متقارن است و خط تقارن از میانگین داده ها می گذرد
- مساحت زیر منحنی نرمال، همانطور که مجموع فراوانی های نسبی در یک جدول فراوانی برابر با یک می باشد، برابر است با یک.

همانطور که قبلاً گفته شد مساحت زیر منحنی نشان دهنده احتمال است و مجموع احتمالات همه مقادیر ممکنه یک متغیر برابر است با یک.

- خط تقارن منحنی نرمال را به دو قسمت مساوی تقسیم می کند و این بدین مفهوم است که احتمال اینکه مقدار متغیر مربوط از میانگین داده ها بزرگتر باشد برابر است با ۰/۵ و احتمال اینکه مقدار این متغیر کوچکتر باشد نیز برابر است با ۰/۵.

ادامه توزیع نرمال:

- اگر داده‌ها دارای توزیع نرمال باشند، سه شاخص مرکزی میانگین، میانه و نما باهم مساوی هستند.
- توزیع نرمال فقط به دو پارامتر میانگین و انحراف معیار داده‌ها وابسته است که میانگین تعیین کننده موقعیت منحنی‌های متفاوت نرمال نسبت به هم و همچنین نسبت به محور افقی است و انحراف معیار نشان دهندهٔ کشیدگی یا ارتفاع منحنی است. در واقع یک رابطهٔ معکوس بین انحراف معیار داده‌ها و کشیدگی منحنی نرمال وجود دارد



- اگر داده‌ها دارای توزیع نرمال باشند، ۶۸٪ از داده‌ها در فاصله دو انحراف معیار از میانگین قرار دارند به عبارت دیگر مساحت زیر منحنی حدفاصل محدوده $[\mu-\sigma, \mu+\sigma]$ برابر است با ۶۸٪.

$\%68 \in [\mu-\sigma, \mu+\sigma]$

- به همین صورت، تقریباً ۹۵ درصد از داده‌ها در فاصله دو انحراف معیار از میانگین قرار دارد و مساحت زیر منحنی حدفاصل این دو مقدار برابر است با ۹۵٪

$\%95 \in [\mu-2\sigma, \mu+2\sigma]$

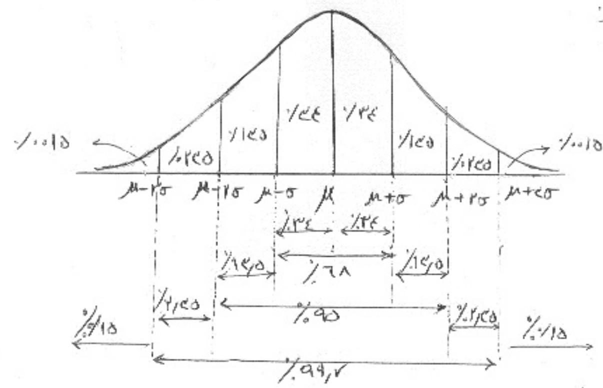
ادامه توزیع نرمال

- و نهایتاً تقریباً ۹۹٫۷ درصد از داده‌ها در فاصله سه انحراف معیار از میانگین قرار دارند، یعنی مساحت زیر منحنی بین میانگین منهای سه انحراف معیار و میانگین بعلاوه سه انحراف معیار برابر است با ۰/۹۹۷ .

$$\%99.7 \in [\mu - 3\sigma \,,\, \mu + 3\sigma]$$

این خاصیت نشان می‌دهد که علی‌رغم اینکه منحنی توزیع نرمال از $-\infty$ تا $+\infty$ امتداد دارد، اما بیش از ۹۹٫۷ درصد از داده‌ها در فاصله سه انحراف معیار از میانگین قرار دارند و فقط ۰/۳ درصد از داده‌ها خارج از این محدوده هستند.

حال با توجه به خاصیت تقارن منحنی نرمال، خواص فوق را می‌توان به صورت زیر خلاصه کرد:



اما همانطور که قبلاً گفته شد، در بیشتر مواقع احتمال قرار گرفتن متغیر در یک محدوده مشخص مورد نیاز است و می‌بایست احتمال یعنی مساحت زیر منحنی حد فاصل دو مقدار مشخص

ادامه توزیع نرمال

نیازمند استفاده از روش‌های ریاضی و انتگرال‌گیری است که با توجه به فرمول مربوط به چگالی توزیع نرمال، محاسبه انتگرال در فاصله دو مقدار مشخص به سادگی امکان‌پذیر نیست.

البته در موارد خاص که در صفحه قبل توزیع داده شد، یعنی در فاصله یک، دو و سه انحراف معیار اطراف میانگین می‌توان از مقادیر بدست آمده در صفحه قبل استفاده کرد.

برای رفع این مشکل، می‌توان با انجام تغییرات کوچکی روی داده‌ها، بدون اینکه ماهیت داده‌ها از نظر نرمال بودن مورد خدشه قرار گیرد، داده‌ها را ساده‌تر کرد و برای محاسبه مساحت زیر منحنی نرمال از جدولی که برای همین منظور تهیه شده است استفاده کرد.

برای این منظور می‌توان با استفاده از رابطه $Z_i = \frac{x_i - \bar{x}}{s}$ داده‌ها را استاندارد کرد که در این رابطه x_i متغیر مربوطه، $\bar{x}$ میانگین داده‌ها و s انحراف معیار داده‌هاست که از نمونه جمع‌آوری شده محاسبه می‌شوند. با این کار، مقادیر استاندارد شده یعنی Z_i ها همچنان دارای توزیع نرمال اما با میانگین صفر و انحراف معیار یک می‌باشند که به این توزیع نرمال استاندارد می‌گویند. این توزیع نیز همان خواص توزیع نرمال را دارا می‌باشد یعنی:

%۶۸ ∈ [-۱ , +۱]

%۹۵ ∈ [-۲ , +۲]

%۹۹٫۷ ∈ [-۳ , +۳]

به عبارت دیگر با استاندارد کردن داده‌ها، اکثر قریب به اتفاق مقادیر جدید در فاصله ۳ انحراف معیار از میانگین قرار می‌گیرد.

ادامه توزیع نرمال

از طرفی پس از استاندارد شدن داده‌ها، حتماً بسیاری از مقادیر استاندارد شده مثبت و منفی بیشتر خواهند بود و با توجه به ساده‌تر شدن داده‌ها و به تبع فرمول مربوط به چگالی توزیع نرمال، استفاده از جدول مربوط برای محاسبه مساحت زیر منحنی نرمال به سادگی امکان‌پذیر می‌شود. این جدول در پایان همه کتابهای آمار و همچنین از طریق اینترنت قابل دسترسی است. البته هنگام استفاده از جدول بایستی به راهنمای استفاده از آن توجه کرد.

لازم به توضیح است که داده‌های استاندارد شده فاقد واحد اندازه‌گیری است و این مزیت مقادیر استاندارد شده این امکان را به ما می‌دهد که مقادیر یا متغیرهای متفاوت با توزیع نرمال متفاوت (میانگین یا انحراف معیار متفاوت) و یا متغیرهای مختلف با واحدهای اندازه‌گیری متفاوت را باهم مقایسه کرد.

مثال:

در دو کلاس A و B، نمره درس آمار مربوط به دانشجویان شرکت کننده در این دو کلاس به شرح زیر است:

در کلاس A، آزمون بر مبنای 100 و میانگین و انحراف معیار نمرات کسب شده توسط دانشجویان به ترتیب ۷۴ و ۶ بدست آمده است. و در کلاس B، آزمون بر مبنای ۲۰ و میانگین و انحراف معیار نمرات به ترتیب ۱۵ و ۲ بدست آمده است. نمره ۱۸ در کلاس B معادل چه نمره‌ای در کلاس A است؟ توجه کنید که نمرات دو کلاس از یک توزیع نرمال با میانگین و انحراف معیار متفاوت برخوردارند. حتی اگر نمرات کلاس A را بر مبنای ۲۰ تغییر دهیم، نمرات قابل مقایسه نیستند.

ادامه توزیع نرمال

برای این منظور حتماً باید مقادیر استاندارد شده قابل مقایسه باشند.

استاندارد شده نمره ۱۸ در کلاس B برابر است با $\frac{۱۸-۱۵}{۲}=۱٫۵$

حال پیدا کنیم نمره معادل در کلاس A ، باید نمره‌ای را پیدا کنیم که استاندارد شده آن برابر ۱٫۵ باشد یعنی

$$۱٫۵ = \frac{x-۷۴}{۶} \Rightarrow x = ۷۴+۹ = ۸۳$$

به عبارت دیگر نمره ۱۸ در کلاس B از نظر ارزش، معادل نمره ۸۳ در کلاس A می‌باشد.

مثال در خصوص توزیع نرمال و استفاده از جدول توزیع نرمال:

- نتایج حاصل از یک نمونه ۵۰۰ تایی از خانوارهای شهری نشان می‌دهد که توزیع هزینه‌های ماهیانه مواد غذایی خانوارها نرمال با میانگین ۴۳۵ هزار تومان و انحراف معیار ۳۸ هزار تومان است. چند درصد از خانوارهای این شهر، هزینه ماهیانه مواد غذایی آنها:

- کمتر از ۳۵۰ هزار تومان است؟
- بیشتر از ۴۰۰ هزار تومان است؟
- بین ۳۸۰ هزار تومان تا ۵۰۰ هزار تومان است؟
- بین ۴۵۰ تا ۵۰۰ هزار تومان است؟

✓ $P(X<۳۵۰) = P\left(Z < \frac{۳۵۰-۴۳۵}{۳۸}\right) = P(Z<-۲٫۲۳)$
$= ۰/۵ - ۰/۴۸۷۱ = ۰/۰۱۲۹$

توجه داشته باشید در این جا ما از جدولی استفاده شده که در این جدول مساحت زیر منحنی حد فاصل محور تقارن (محوری که از صفر می‌گذرد) و محوری که از نقطه Z می‌گذرد، داده شده است. پس هنگام استفاده از جدول نرمال استاندارد بایستی از راهنمای جدول استفاده کرد

ادامه توزیع نرمال

$/۰۱۲۹ \times ۱۰۰ = \%۱,۲۹$

احتمال × صد = درصد

✓ $P(X>۴۰۰) = P(Z> \frac{۴۰۰-۴۳۵}{۳۸}) = P(Z>-/۹۲)$
$= /۵ + /۳۲۱۲ = /۸۲۱۲$

$/۸۲۱۲ \times ۱۰۰ = \%۸۲,۱۲$

✓ $P(۳۸۰<X<۴۵۰) = P(\frac{۳۸۰-۴۳۵}{۳۸}<Z<\frac{۴۵۰-۴۳۵}{۳۸}) = P(-۱,۴۴<Z</۳۹)$
$= /۴۲۵۱ + /۱۵۱۷ = /۵۷۶۸$

$/۵۷۶۸ \times ۱۰۰ = \%۵۷,۶۸$

✓ $P(۴۵۰<X<۵۰۰) = P(\frac{۴۵۰-۴۳۵}{۳۸}<Z<\frac{۵۰۰-۴۳۵}{۳۸})$
$= P(/۳۹<Z<۱,۷۱) = /۴۵۶۵ - /۱۵۱۷$
$= /۳۰۴۸$

$/۳۰۴۸ \times ۱۰۰ = \%۳۰,۴۸$

توجه داشته باشید در توزیع‌های آماری پیوسته $P(X=a)=0$ است. بنابراین قراردادن علامت تساوی در هر کدام از روابط فوق مثل $<$ یا $\leq$ هیچ تأثیری در نتیجه ندارد.

تمرین:

برای نمایش توزیع نرمال از نماد $X \sim N(\mu, \sigma)$ استفاده می‌شود که در آن نماد N معرف توزیع نرمال، μ میانگین و σ انحراف معیار است.
فرض کنید متغیر X دارای توزیع نرمال با میانگین ۱۲۸ و انحراف معیار ۱۷ باشد یعنی $X \sim N(۱۲۸, ۱۷)$. احتمالات زیر را حساب کنید.

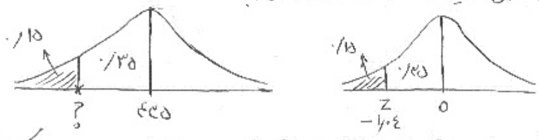
$P(X>۱۵۰)=?$ $\quad P(X<۱۸۰)=?$ $\quad P(X>۱۰۰)=?$

$P(۱۵۰<X<۱۷۰)=?$ $\quad P(۸۰<X<۱۰۰)=?$

ادامه توزیع نرمال

امّا در مواردی با داشتن احتمال مربوط و یا مساحت زیر منحنی، تعیین اندازه متغیر مربوط مورد نیاز است که در این صورت ابتدا باید با استفاده از جدول توزیع نرمال مقدار استاندارد شده متغیر را (برعکس کاری که قبلاً در محاسبه مساحت زیر منحنی انجام می‌گرفت) تعیین و سپس مقدار اصلی متغیر را محاسبه نمود.

به عنوان مثال چنانچه بخواهیم در مثال صفحه ۷ این بخش، به ۱۵ درصد از خانوارهایی که کمترین هزینه‌های ماهیانه مقدار ثابتی را دارند، کمک‌های مالی اعطا کنیم، از چه مبلغی مشمول دریافت کمک‌های مالی می‌شوند؟



ابتدا با استفاده از جدول با توجه به اطلاعات داده شده از جدول مقدار ۰/۳۵ یا نزدیکترین عدد به ۰/۳۵ را در متن جدول پیدا می‌کنیم سپس با توجه به عدد ستون سمت چپ و عدد بالای این ستون مقدار صحیح و اولین رقم بعد از علامت اعشار در ستون سمت چپ جدول و رقم دوم بعد از علامت اعشار در ردیف بالای جدول (سرستون) مشخص می‌شود که این مقدار در جدول نرمال برای شکل فوق ۱/۰۴- است که علامت منفی به خاطر قرار گرفتن این عدد در سمت چپ یا نیمه سمت چپ جدول است. سپس این مقدار را به مقدار اولیه متغیر تبدیل می‌کنیم.

$$-۱/۰۴ = \frac{x - ۴۳۵}{۳۸} \Rightarrow x = ۳۹۵/۴۸$$

یعنی هزینه‌های کمتر از ۴۹۵/۴۸ هزار ریال، مشمول دریافت کمک هزینه می‌باشند.