برآوردهای آماری

در مطالعات آماری در اکثر موارد به دلایل متفاوت از جمله تعداد جمعیت هدف (جامعه آماری)، ملاحظات مالی، صرفه‌جویی در زمان و قابل دسترس نبودن همهٔ واحدهای جمعیت، نمونه‌ای از جمعیت، با استفاده از روشهای نمونه‌گیری انتخاب و با مطالعه بر روی نمونه انتخابی، نتایج حاصله به جمعیت تعمیم داده می‌شود و بدین روش استنباط‌های آماری انجام می‌گیرد.

یکی از اولین اقدامات در تحلیل و استنباط داده‌های آماری، تخمین پارامترهای جمعیت با استفاده از شاخصهای آماری بدست آمده از نمونه n تایی می‌باشد. به این شاخصهای آماری که از نمونه جمع‌آوری شده از جمعیت هدف محاسبه می‌شوند، برآوردگر آماری می‌گویند.

برآورد نقطه‌ای (Point Estimate):

هر شاخص آماری که با استفاده از نمونه n تایی بدست می‌آید مانند میانگین ($\bar{x}$)، واریانس (s^2)، انحراف معیار (s) یا نسبت ($\bar{p}$) برای برآورد یا تخمین میانگین یک ویژگی در جمعیت (μ)، واریانس (σ^2)، انحراف معیار جمعیت (σ) و یا نسبت یک ویژگی در جمعیت (P)، برآورد نقطه‌ای گفته می‌شود که در همه تحلیلهای آماری از آنها استفاده می‌شود. اما ایراد اصلی این شاخصها این است که با تکرار نمونه‌گیری، چون واحدهای نمونه تغییر می‌کنند، این شاخصها نیز در نمونه‌گیری‌های متعدد، مقادیر متفاوتی خواهند بود و نوعی اعتماد کامل محقق از مقادیر بدست آمده برآورده نمی‌شود. طریقه محاسبه دقیق کلیه شاخصهای فوق‌الذکر در جلسات قبل شرح داده شده‌اند.

از آنجایی که معمولاً نمونه‌گیری از جمعیت هدف فقط یک بار انجام می‌گیرد، برآورد نقطه‌ای را می‌توان به یک برآورد فاصله‌ای تبدیل کرد طوری که با اطمینان مشخصی این فاصله همه برآوردهای نقطه‌ای بدست آمده حاصل از نمونه‌گیری‌های متعدد را در برگیرد و به عبارت صحیح‌تر بتوان ادعا کرد با اطمینان مشخصی، فاصله بدست آمده، پارامتر جمعیت را در برگیرد یا شامل بشود. به این فاصله، برآورد فاصله‌ای یا فاصله اطمینان می‌گوئیم.

برآورد فاصله‌ای (Interval Estimate) :

یکی از اولین اقدامات در تحلیل آماری، تعمیم نتیجه حاصل از نمونه به کل جمعیت با استفاده از محاسبه برآورد نقطه‌ای یا فاصله اطمینان (Confidence Interval) است.

طبیعی است در برآورد هر پارامتر جمعیت، نتیجه حاصل از نمونه n تایی به دلایل مختلف مانند خطای نمونه‌گیری، نمی‌تواند دقیقاً برابر با پارامتر جمعیت باشد و شاخص بدست آمده از نمونه متفاوت از مقدار واقعی خواهد بود که به تفاوت بین شاخص بدست آمده از نمونه و پارامتر جمعیت، خطای برآورد گفته می‌شود که با d نشان داده می‌شود. اما خطای برآورد با توجه به اینکه مقدار پارامتر جمعیت نامعلوم است، مقداری است مجهول و دقیقاً مشخص نیست مقدار خطای برآورد چقدر است اما با استفاده از توزیع آماری شاخص آماری بدست آمده از نمونه و خواص توزیع مربوطه می‌توان حداکثر خطای برآورد را با اطمینان مشخصی تعیین کرد و با اضافه کردن و کم کردن این مقدار به برآورد نقطه‌ای، آن را به یک فاصله تبدیل کرد یعنی مثلاً برای میانگین حاصل از نمونه، برآورد فاصله‌ای برابر خواهد بود با:

$\bar{x} \pm d$ یا $[\bar{x}-d \,,\, \bar{x}+d]$ که d مقدار ثابتی است که با توجه به توزیع آماری شاخص بدست آمده از نمونه و میزان اطمینان مورد نظر قابل محاسبه است. لازم به توضیح است که هر شاخص آماری که از نمونه بدست می‌آید از آنجایی که با تغییر نمونه گیری، مقدار این شاخص نیز تغییر می‌کند، یک متغیر محسوب می‌شود و هر متغیر از توزیع آماری خاص برخوردار است. نوع توزیع آماری مربوط به شاخص بدست آمده به میزان اطلاعات ما از پارامترهای جمعیت (که معمولاً مجهول هستند) بستگی دارد که در این جزوه وارد جزئیات آن نمی‌شویم. به عنوان نمونه اگر برای محاسبه شاخص میانگین ($\bar{x}$) از جمعیتی با میانگین μ و واریانس σ^2، نمونه‌گیری انجام گیرد و متغیر مورد نظر در این جمعیت را با X نشان دهیم با استفاده از قضایای آماری (مانند قضیه حد مرکزی) می‌توان نشان داد که توزیع نمونه برداری میانگین - نرمال با پارامترهای μ و $\frac{\sigma}{\sqrt{n}}$ می‌باشد. یعنی:

$$\bar{X} \sim N\left(\mu , \frac{\sigma}{\sqrt{n}}\right)$$

برآورد فاصله‌ای میانگین یک جمعیت (μ):

حالت اول: توزیع جامعه آماری نرمال و واریانس جمعیت (σ^2) معلوم است.

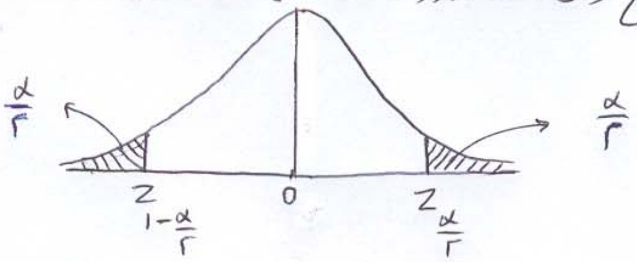
در این حالت، همچنانکه قبلاً توضیح داده شد، توزیع $\bar{X}$، نرمال با میانگین μ و انحراف معیار $\frac{\sigma}{\sqrt{n}}$ خواهد بود و تعداد نمونه تاثیری در این توزیع آماری ندارد و به سادگی با استفاده از خواص توزیع نرمال مقدار استاندارد شده $\bar{X}$ برابر خواهد بود با

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

و از آنجایی که توزیع نرمال یک توزیع متقارن است، اینکه مقدار استاندارد شده میانگین با احتمال $(1-\alpha)$ در فاصله مشخصی قرار گیرد را می‌توان به صورت زیر نوشت

$$P\left(Z_{1-\frac{\alpha}{2}} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq Z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

که $Z_{\frac{\alpha}{2}}$ و $Z_{1-\frac{\alpha}{2}}$ دو مقدار قرینه هستند و می‌توان به سادگی با استفاده از جدول توزیع نرمال استاندارد آنها را بدست آورد.



و با اندکی محاسبات جبری رابطه فوق را می‌توان به صورت زیر نوشت.

$$P\left(\bar{x}+Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}+Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

که در این رابطه $Z_{\frac{\alpha}{2}}$ و $Z_{1-\frac{\alpha}{2}}$ قرینه همدیگر هستند یعنی از نظر قدر مطلق با هم مساوی هستند. بدین ترتیب مقدار خطای برآورد یعنی d برابر است با

$$d = Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$$

بنابراین فاصله اطمینان $(1-\alpha)$ درصد برای میانگین جمعیت (μ) برابر است با

$$\bar{x} \pm Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} \quad \text{یا} \quad \left[\bar{x}-Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x}+Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right]$$

مثال: تجربیات گذشته نشان می‌دهد که توزیع آماری سن کارکنان یک سازمان نرمال با انحراف معیار ۶ سال می‌باشد. نتایج حاصل از یک نمونه ۶۴ تایی از کارکنان این سازمان نشان می‌دهد که میانگین

سن کارکنان ۳۶ سال است. فواصل اطمینان ۹۰، ۹۵ و ۹۹ درصد برای میانگین سن کارکنان این سازمان را حساب کنید

$\%90 \Rightarrow \alpha = 1 - \%9 = 0.1 \Rightarrow \frac{\alpha}{2} = \frac{0.1}{2} = 0.05$

با استفاده از جدول توزیع نرمال استاندارد $Z_{\frac{\alpha}{2}} = 1.64$، پس فاصله اطمینان ۹۰ درصد برای میانگین سن کارکنان برابر است با:

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow 36 \pm 1.64 \times \frac{6}{\sqrt{64}} \to [34.77 , 37.23]$$

$$36 \pm 1.23$$

با اطمینان ۹۰٪، میانگین سن کارکنان این سازمان در این فاصله قرار دارد

به همین ترتیب می‌توان فواصل اطمینان ۹۵ و ۹۹ درصد را نیز محاسبه نمود

$$\%95 \qquad \alpha = 1 - \%95 \Rightarrow \frac{\alpha}{2} = \%0.025 \Rightarrow Z_{\%0.025} = 1.96$$

$$36 \pm 1.96 \times \frac{6}{\sqrt{64}} \longrightarrow [34.53 , 37.47]$$

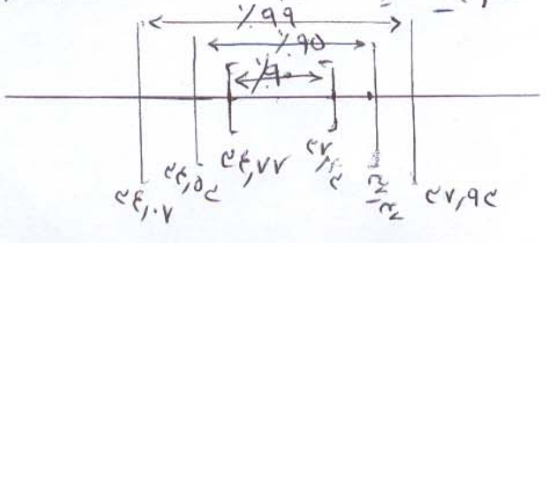
$$36 \pm 1.47$$

$$\%9 \qquad \alpha = 1 - \%99 \Rightarrow \alpha = \%0.1 \Rightarrow \frac{\alpha}{2} = \%0.05 \Rightarrow Z_{\%0.05} = 2.58$$

$$36 \pm 2.58 \frac{6}{\sqrt{64}} \longrightarrow [34.07 \quad 37.93]$$

$$36 \pm 1.93$$

مقایسه فواصل اطمینان فوق نشان می‌دهد که با افزایش میزان اطمینان، دامنه فاصله اطمینان نیز افزایش می‌یابد و این یک قاعده کلی است که در همه فواصل اطمینان صدق می‌کند یعنی یک رابطه مستقیم بین میزان اطمینان و دامنه فاصله اطمینان وجود دارد.



ادامه مثال :

اگر همین نتیجه از یک نمونه ۱۰۰ تایی از کارکنان آن سازمان بدست آمده باشد فواصل اطمینان فوق چقدر خواهند بود؟

در این قسمت فقط تعداد نمونه تغییر کرده است. پس فواصل اطمینان مذکور برابر خواهند بود با

$\%90 \quad 36 \pm 1.64 \frac{6}{\sqrt{100}} \Rightarrow [35.02 \,,\, 36.98]$

$\%95 \quad 36 \pm 1.96 \frac{6}{\sqrt{100}} \Rightarrow [34.82 \,,\, 37.18]$

$\%99 \quad 36 \pm 2.58 \frac{6}{\sqrt{100}} \Rightarrow [34. \quad ,\, 37.55]$

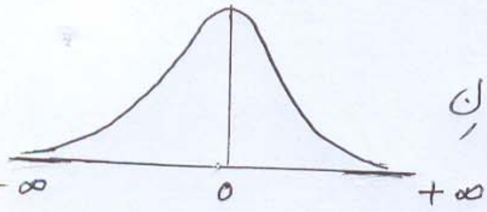
مقایسه فواصل اطمینان با نمونه بزرگتر نشان می دهد که دامنه فاصله اطمینان کاهش می یابد و این یک قاعده کلی است که یک رابطه معکوس بین تعداد نمونه و دامنه فاصله اطمینان وجود دارد. بنابراین برای داشتن یک فاصله اطمینان با میزان اطمینان بالا بهتر است تا آنجایی که ممکن است حجم نمونه را افزایش داد.

حالت دوم : توزیع جامعه آماری نرمال و واریانس جمعیت (σ^2) معلوم نیست.

در این حالت ، برای برآورد واریانس جمعیت ، از نمونه انتخاب شده استفاده می شود و علاوه بر میانگین ، واریانس نمونه یعنی s^2 نیز محاسبه می شود. در این حالت مقدار استاندارد شده $\bar{x}$ یعنی $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ بجای توزیع Z (نرمال استاندارد) از توزیع دیگری بنام توزیع t پیروی می کند. در اینجا لازم است قبل از ادامه این بحث ، مختصری در خصوص توزیع t گفته شود.

توزیع t .

توزیع t یکی از توزیع‌های پیوسته آماری است که از نظر شکل ظاهری، منحنی آن
شبیه به منحنی نرمال استاندارد است.



توزیع t نیز یک توزیع متقارن است و خط تقارن
از نقطه صفر می‌گذرد.

یکی از مهمترین خواص توزیع t این است که توزیع t، برخلاف توزیع نرمال که دارای
دو پارامتر میانگین و انحراف معیار است، توزیع t فقط دارای یک پارامتر به نام
درجه آزادی (Degree of freedom) است که با df نشان داده می‌شود.

و خاصیت مهم دیگر توزیع t این است که با افزایش تعداد نمونه به عبارت دیگر
افزایش درجه آزادی توزیع t، توزیع t به توزیع Z (نرمال استاندارد) میل
می‌کند. به عبارت دیگر

$$n \longrightarrow \infty$$
$$t \longrightarrow Z$$

که از آنجایی که وقتی تعداد نمونه بیشتر از ۳۰ باشد. این دو توزیع خیلی بهم نزدیک
می‌شوند، می‌توان در حالاتی که $n>30$ است. از جدول توزیع نرمال استفاده کرد
به عبارت دیگر می‌توان گفت در این حالت $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ تقریباً (با تقریب
خیلی بالا) از توزیع Z پیروی می‌کند.

برای تعیین مقادیر مختلف $t_{\frac{\alpha}{2}}$ و $t_{1-\frac{\alpha}{2}}$ (که این دو مقدار نیز قرینه هستند)
می‌توان از جدول توزیع t با توجه به درجه آزادی توزیع که معمولاً $df = n-1$
می‌باشد

مثال:

نتایج حاصل از یک نمونه ۱۶ تایی از مدیران دستگاههای دولتی در رابطه با میزان آگاهی آنها از تئوری مدیریت، بر اساس یک روش استاندارد سنجش فنون مدیریت مدیران، نشان می دهد که توزیع امتیاز مدیریت آنها نرمال با میانگین ۷۸ و انحراف معیار ۵ می باشد. فواصل اطمینان ۹۰، ۹۵ و ۹۹ درصد برای میانگین امتیاز مدیران دستگاههای دولتی را حساب کنید.

فرمول محاسبه فاصله اطمینان برای میانگین $\bar{x} \pm t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$

که مقادیر s و $\bar{x}$ از نمونه ۱۶ تایی محاسبه شده اند و درجه آزادی توزیع t در این مثال برابر ۱۵ می باشد. $df = 16 - 1 = 15$

مقادیر $t_{\frac{\alpha}{2}}$ و $t_{1-\frac{\alpha}{2}}$ که قرینه هستند با استفاده از جدول توزیع t بدست می آید. فواصل اطمینان مذکور برابر خواهند بود

$$\%90 \rightarrow \alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow t_{\frac{\alpha}{2}} = 1.75$$

$$78 \pm 1.75 \times \frac{5}{\sqrt{16}} \longrightarrow [75.81 \quad 80.19]$$
$$78 \pm 2.19$$

$$\%95 \quad \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow t_{\alpha/2} = 2.13$$

$$78 \pm 2.13 \frac{5}{\sqrt{16}} \longrightarrow [75.34 \quad 80.66]$$
$$78 \pm 2.66$$

$$\%99 \quad \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005 \Rightarrow t_{\frac{\alpha}{2}} = 2.94$$

$$78 \pm 2.94 \times \frac{5}{\sqrt{16}} \longrightarrow [74.33, \ 81.67]$$
$$78 \pm 3.67$$