جزوه آمار و کاربرد آن در مدیریت

۱/۵

متغیر تصادفی اعداد خود را بصورت تصادفی می‌گیرد   مثلا تعداد شیرهای ظاهر شده در پرتاب دو سکه

بر خلاف اسمش متغیر نیست بلکه تابع است که روی فضای نمونه تعریف می‌شود

متغیر تصادفی، تابعی است که دامنه آن فضای نمونه و حوزه آن مجموعه‌ای از اعداد حقیقی است.

به عنوان مثال در پرتاب دو سکه:

تعداد شیرها متغیر تصادفی که مقادیر ۰ و ۱ و ۲ را به خود می‌گیرد

$(TT) \to ۰$

$(HT),\ (TH) \to ۱$

$(HH) \to ۲$

تابع احتمال: تابعی که می‌توان با استفاده از آن احتمال هر یک از مقادیر ممکن متغیر تصادفی را مشخص کرد

تابع احتمال یا توزیع احتمال گفته می‌شود. در واقع تابع احتمال تابعی است که دامنه آن مقادیر ممکن متغیر تصادفی و حوزه آن احتمالات مربوط به هر متغیر تصادفی است.

تابع احتمال غیر منفی و مجموع احتمالات همیشه ۱ می‌باشد.

x	۰	۱	۲
$P(X=x)$	/۲۵	/۵	/۲۵

تابع توزیع (تابع احتمال تجمعی): تابعی است که به ازای هر مقادیر ممکن متغیر تصادفی X، احتمال وقوع مقادیری کوچکتر یا مساوی x را نشان می‌دهد. که به صورت $P(X\leq x)$ یا $F(x)$ نشان می‌دهند

مثال: جدول تابع احتمال زیر را در نظر بگیرید:

x	-۲	۰	۳	۵	۱۰
$P(X=x)$	/۱	/۱۵	/۲۵	/۳	/۲

x	-۲	۰	۳	۵	۱۰
$F(x)=P(X\leq x)$	/۱	/۲۵	/۵	/۸	۱

به عنوان مثال $F(۳)=P(X\leq ۳)=/۵$

امید ریاضی:

امید ریاضی همان میانگین موزون است که در آن احتمالات نقش وزنها را بازی می کنند.

که با $E(x)$ نشان داده می شود.

امید ریاضی متغیر تصادفی x از طریق رابطه زیر محاسبه میشود

$$E(x) = \sum_x x f(x)$$

$$f(x) = P(X = x)$$

مثال: شرکتی تولیدکننده آبگرمکن گازی است. تقاضای ماهانه همراه با احتمالات مربوط در این جدول آورده شده است. امید ریاضی تقاضا را بدست آورید.

x (تقاضا)	۱۰۰	۲۰۰	۳۰۰	۴۰۰	۵۰۰	جمع
احتمال $f(x)$	۰/۱۵	۰/۲۵	۰/۳	۰/۲	۰/۱	۱
$x f(x)$	۱۵	۵۰	۹۰	۸۰	۵۰	۲۸۵

بنابراین می توان گفت این شرکت انتظار دارد که برای کالایش ۲۸۵ واحد تقاضا در ماه داشته باشد.

با توجه به تعریف امید ریاضی اگر a و b مقادیر ثابتی باشند دارای این خواص خواهد بود

$$E(b) = b$$

$$E(aX) = a E(x)$$

$$E(ax+b) = a E(x) + b$$

تابع احتمال زیر مفروض است

x	۲-	۱	۳	۵
$f(x)$	۰/۲	۰/۴	۰/۳	۰/۱

مطلوب است:

الف) $E(x)$ ب) $E(-3x+2)$ ج) $E[(x-5)^2]$ محاسبه شود.

۲/۵

واریانس

واریانس متغیر تصادفی گسسته X که مقدار پراکندگی را حول میانگین (امید ریاضی) نشان می‌دهد و آن را با $V(X)$ نشان می‌دهیم.

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

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

اگر a و b مقادیر ثابتی فرض گردند این خواص برای واریانس متغیر تصادفی قابل اثبات است

$$V(b)=0$$

$$V(aX)=a^2V(x)$$

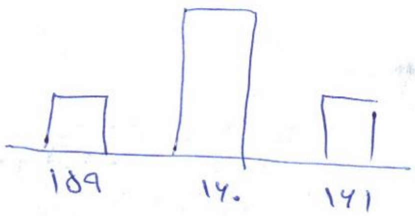
$$V(aX+b)=a^2V(X)$$

مثال: اگر $E(x)=2/5$ و $E(x^2)=8$ باشد مقادیر زیر را حساب کنید

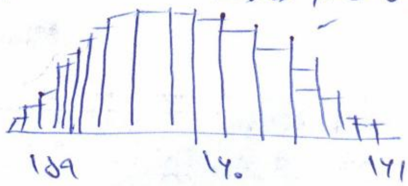
$V(3)$ $V(x)$ $V(-2x+3)$

۴/۵

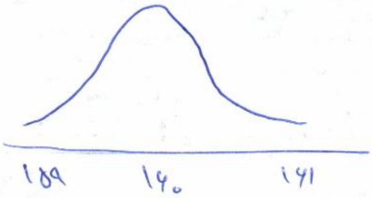
فرض کنید می خواهیم قد ۴۰ دانش آموز را اندازه بگیریم



با خطای ۱۰ سانتیمتر



با خطای یک سانتیمتر



واضح است اگر مقادیر گرد نشوند و مقادیر بسیار کوچکی همچون یک صدم میلیمتر محاسبه شوند توزیع احتمال از حالت گسسته به پیوسته تبدیل می شود.

تابعی که معرف متغیر تصادفی پیوسته است تابع چگالی احتمال خوانده می شود که مساحت زیر منحنی آن برابر یک است.

تابع چگالی احتمال را با $f(x)$ نشان می دهند

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

تابعی با مقادیر $f(x)$ را تابع چگالی احتمال گوییم اگر

$$f(x) \geq 0$$

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

مثال: در یک شرکت مواد غذایی میزان ربی که در قوطی های ۳۳۰ گرمی ریخته می شود دارای تابع چگالی زیر است. احتمالات زیر را حساب کنید

$$f(x) = \begin{cases} \frac{1}{12} & ۳۲۴ < x < ۳۳۶ \\ 0 & \text{در غیر اینصورت} \end{cases}$$

الف) حداکثر ۳۳۲ گرم رب داشته باشد

ب) حداقل ۳۲۵ گرم رب داشته باشد

ج) بین ۳۳۰٫۵ تا ۳۳۵٫۵ گرم رب داشته باشد. د) دقیقاً ۳۳۰٫۱۲ گرم رب داشته باشد

تابع توزیع متغیر تصادفی پیوسته:

در بسیاری موارد مایلیم که احتمال اینکه متغیر تصادفی پیوسته از عدد حقیقی x کوچکتر یا مساوی باشد را بدانیم. در این موارد باید از تابع توزیع که $F(x)$ نامیده می‌شود استفاده کرد:

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

بدیهی است که $F(-\infty)=0$ و $F(\infty)=1$ است.

$$P(a \leq x \leq b) = \int_{a}^{b} f(x)\,dx \Rightarrow$$

$$P(a \leq x \leq b) = F(b) - F(a)$$

امید ریاضی و واریانس متغیر تصادفی پیوسته:

(میانگین)

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

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

مثال: اگر سود شرکتی را متغیری پیوسته در نظر بگیریم که دارای چگالی احتمال زیر است. سود مورد انتظار این شرکت چقدر است؟

$$f(x) = \begin{cases} \frac{2}{63}(x+3) & -2 < x < 5 \\ 0 & \text{در غیر اینصورت} \end{cases}$$

$$E(x) = \int_{-\infty}^{+\infty} x.f(x).dx = \int_{-2}^{5} x\,\frac{2}{63}(x+3)\,dx = \ldots$$

۹/b

توزیع نرمال

مهمترین توزیع پیوسته، توزیع نرمال است، توزیع نرمال، توزیعی زنگی شکل است که اولین بار در قرن هجدهم مورد مطالعه قرار گرفت. معمولاً پدیده‌های واقعی دارای توزیع نرمال هستند به همین دلیل این توزیع نقش اساسی در آمار دارد چراکه خیلی از پدیده‌های طبیعی دارای این توزیع هستند وهمچنین شکل حدی بسیاری از توزیع‌های دیگر نیز نرمال است.

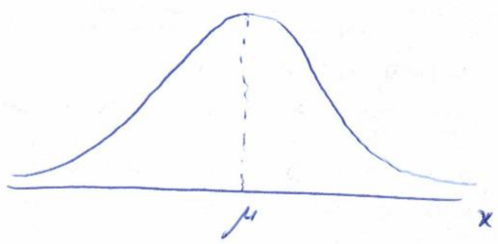
توزیع نرمال را می‌توان چنین تعریف کرد:

متغیر تصادفی پیوسته X با میانگین μ و انحراف معیار σ دارای توزیع نرمال است اگر تابع چگالی آن به صورت زیر باشد

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

$e = ۲٫۷۱۸$

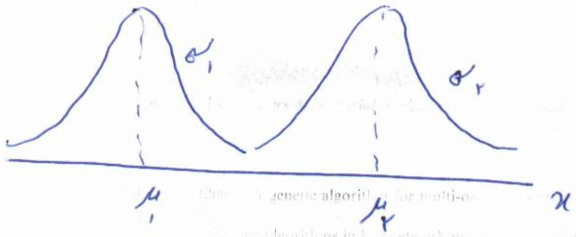
$\pi = ۳٫۱۴$



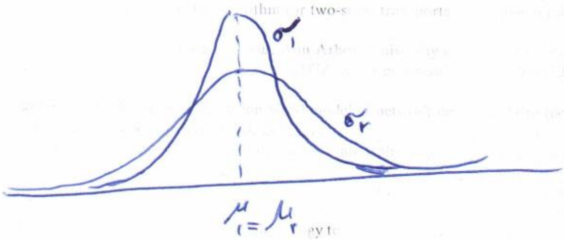
$X \sim N(\mu, \sigma)$

به دلیل استفاده زیاد از این توزیع، توزیع نرمال متغیر تصادفی X با انحراف معیار σ و میانگین μ را به صورت مقابل نشان می‌دهیم.

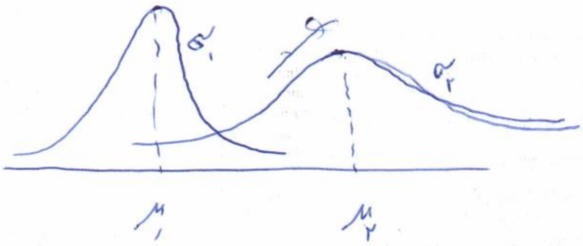
V/b



$\mu_1 < \mu_2 \;,\; \sigma_1 = \sigma_2$



$\mu_1 = \mu_2 \;,\; \sigma_2 > \sigma_1$



$\sigma_1 < \sigma_2 \;,\; \mu_1 \leq \mu_2$

خصوصیات توزیع نرمال

توزیع نرمال

۱- سطح زیر منحنی توزیع نرمال برابر است با ۱ یعنی $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

۲- به ازای تمام مقادیر X مقدار $f(x)$ بزرگتر یا مساوی صفر است.

۳- حداکثر مقدار تابع در $x = \mu$ حاصل می‌شود $\Leftarrow$ $f'(x) = 0$ و $f''(x) < 0$

۴- تابع حول میانگین μ متقارن است یعنی به عبارت دیگر $f(x+\mu) = f(-x+\mu)$

۵- $\mu = E(x) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$ $\qquad$ $\sigma^2 = V(x) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx$

۶- با دور شدن از μ، چه به سمت راست منحنی و چه در سمت چپ منحنی، منحنی به محور Xها نزدیکتر می‌شود ولی هیچگاه به صفر نمی‌رسد. به بیان دیگر

$$\lim_{x \to +\infty} f(x) = 0 \qquad \lim_{x \to -\infty} f(x) = 0$$

۷- در این توزیع میانه، مد، میانگین با هم برابرند.

۹/ط

توزیع نرمال استاندارد:

۱) تابع چگالی احتمال نرمال به گونه‌ای است که محاسبه احتمال مورد نظر از آن، به این دلیل که نمی‌توان از این تابع به سادگی انتگرال گرفت، کار مشکلی است؛ بنابراین جدولی تهیه شده که فقط برای توزیع نرمال با میانگین صفر و واریانس یک قابل استفاده است. حال اگر بخواهیم به جای متغیر $X \sim N(\mu, \sigma^2)$ از متغیر نرمالی استفاده کنیم که میانگین آن صفر و واریانس آن یک باشد می‌توانیم برای پیدا کردن احتمال مورد نظر از جدول توزیع نرمال استفاده کنیم

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[(x-\mu)/\sigma\right]^2} \qquad -\infty < x < +\infty$$

در این تابع متغیر تصادفی دارای میانگین μ و واریانس σ^2 است. حال اگر متغیری همچون Z را به صورت زیر تعریف کنیم

$$Z = \frac{X-\mu}{\sigma}$$

آنگاه تابع چگالی احتمال به صورت زیر است

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$$E(z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma} E(X-\mu) = \frac{1}{\sigma}\left[E(x) - E(\mu)\right] = \frac{1}{\sigma}(\mu-\mu) = 0$$

$$V(z) = V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} V(X-\mu) = \frac{1}{\sigma^2}\left[V(x)\right] = \frac{1}{\sigma^2}(\sigma^2) = 1$$

بنابراین هر توزیع متغیر تصادفی نرمالی همانند Z که دارای میانگین صفر و واریانس یک باشد توزیع نرمال استاندارد است که به زبان ریاضی چنین نوشته می‌شود

$$Z \sim N(0, 1)$$

پس با داشتن میانگین و واریانس هر متغیری که توزیع آن نرمال باشد می‌توان آن را به توزیع نرمال استاندارد تبدیل کرد و با استفاده از جدول احتمال آن را پیدا کرد.

۱۰/b

نحوه استفاده از جدول توزیع نرمال استاندارد

در جدول توزیع نرمال استاندارد، هدف پیدا کردن احتمال یک مقدار مشخص است. در اینگونه استفاده ابتدا متغیر تصادفی نرمال را به متغیر تصادفی نرمال استاندارد تبدیل می‌کنیم و سپس احتمال مورد نظر را پیدا می‌کنیم.

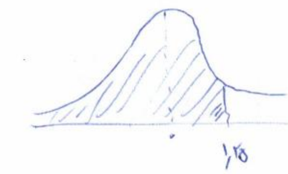
توضیح step ها جدول

مثال: می‌خواهیم مقادیر هر یک از این احتمالات را پیدا کنیم

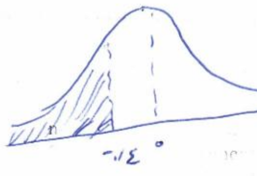
الف) $P(Z \leq 1.25)$ ب) $P(Z \leq -0.4)$ ج) $P(Z \geq 1.59)$

د) $P(-1.1 < Z < 2.76)$

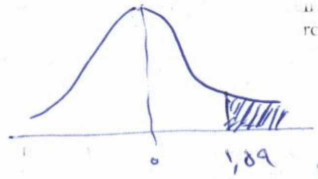
در پاسخ هر احتمالی در توزیع نرمال بهتر است منحنی نرمال برای آن رسم کنیم و سطح مورد نظر را به صورت تقریبی نشان دهیم.



$$P(Z \leq 1.25) = 0.8944$$

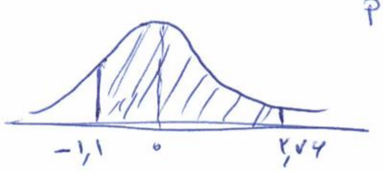


$$P(Z \leq -0.4) = 0.3446$$



$$P(Z \geq 1.59) = 1 - P(Z \leq 1.59) = 1 - 0.9441 = 0.0559$$

$P(Z < 2.76) = P(Z \leq 2.76)$ چون توزیع نرمال پیوسته است



$$P(-1.1 < Z < 2.76) = P(Z \leq 2.76) - P(Z < -1.1) = 0.9971 - 0.1357 = 0.8614$$

مثال: توزیع نرمال با $\mu=30$ و $\sigma=9$ را در نظر می‌گیریم، می‌خواهیم بدانیم احتمال اینکه متغیر تصادفی X مقداری بین ۲۴ تا ۴۳ بگیرد چقدر است.

با تغییر متغیر X به Z می‌توان به جای $P(x_1 \leq X \leq x_2)$ مقدار $P(z_1 \leq Z \leq z_2)$ را محاسبه و از جدول حساب کرد.

$$P(24 \leq x \leq 43) = P\left(\frac{24-30}{9} \leq Z \leq \frac{43-30}{9}\right)$$
$$= P(-0.67 \leq Z \leq 1.44)$$
$$= P(Z \leq 1.44) - P(Z \leq -0.67)$$
$$= 0.9251 - 0.2514 = 0.6737$$

مثال: دستگاه پرکننده شیشه‌های آبلیمو طوری تنظیم شده است که فقط ۳۳۰ گرم آبلیمو را داخل هر شیشه بریزد، با وجود این میزان آبلیمویی که وارد هر شیشه می‌شود دارای توزیع نرمال با میانگین ۳۳۰ گرم و انحراف معیار ۵ گرم است. می‌خواهیم این احتمالات را محاسبه کنیم:

الف) شیشه بین ۳۲۲ تا ۳۲۸ گرم آبلیمو داشته باشد

ب) شیشه‌ای بیش از ۳۳۵ گرم آبلیمو داشته باشد

ج) اداره کنترل کیفیت، میزان آبلیموی ۷۰ شیشه را به صورت تصادفی وزن می‌کند، انتظار می‌رود چند شیشه بیش از ۳۳۵ گرم آبلیمو داشته باشد

الف)
$$P(322 \leq x \leq 328) = P\left(\frac{322-330}{5} \leq Z \leq \frac{328-330}{5}\right)$$
$$= P(-1.6 \leq Z \leq -0.4)$$
$$= P(Z \leq -0.4) - P(Z \leq -1.6)$$
$$= 0.3446 - 0.0548 = 0.2898$$

ب)
$$P(x \geq 335) = P\left(Z \geq \frac{335-330}{5}\right)$$
$$= P(Z \geq 1)$$
$$= 1 - P(Z \leq 1)$$
$$= 1 - 0.8413 = 0.1587$$

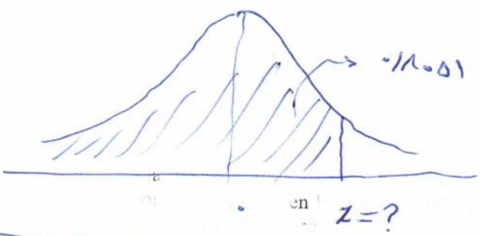
ج)
$$0.1587 \times 70 = 11.109 \simeq 11$$

۱۲/c

استفاده معکوس از جدول توزیع نرمال استاندارد.

در استفاده مستقیم از توزیع نرمال ابتدا Z را مشخص و سپس احتمال آن را از جدول پیدا می‌کردیم، در استفاده معکوس، مقدار Z برای ما مشخص نیست و تنها احتمال آن مشخص است احتمال را در جدول پیدا کرده، سپس Z متناظر با آن را مشخص می‌کنیم.

مثال: می‌خواهیم مقدار Z را در صورتی که $P(Z \le z) = 0.8051$ می‌باشد بیابیم.



$P(Z \le 0.86)$

در روش مستقیم برای حل مسائل، ابتدا متغیر تصادفی X را به Z تبدیل و سپس به جدول مراجعه می‌کردیم، ولی در روش معکوس، پس از پیدا کردن Z مقدار X را با استفاده از این رابطه پیدا می‌کنیم

$$Z = \frac{X - \mu}{\sigma} \Rightarrow x = \mu + \sigma Z$$

مثال: فرض کنید $X \sim N(10, 2)$ است، می‌خواهیم بدانیم به ازای چه مقداری از X رابطه $P(X \le x) = 0.6700$

$$Z_{0.6700} = 0.44$$

$$X = \mu + \sigma Z \Rightarrow X = 10 + 2 \times 0.44 = 10.88$$

$$P(X \le 10.88) = P(Z \le 0.44) = 0.6700$$

۱۳/۶

مثال: میزان مصرف مواد اولیه ماهانه در یک شرکت تولیدی، دارای توزیع نرمال با میانگین ۷۵۲ تن و انحراف معیار ۸۶ است. این شرکت مواد اولیه مورد نیاز را در ابتدای هر ماه تهیه می‌کند. می‌خواهیم حساب کنیم شرکت باید چند تن مواد اولیه تهیه کند تا با ۹۵ درصد اطمینان بداند از نظر مواد اولیه کمبودی نخواهد داشت.

$$P(Z \le z) = .95$$

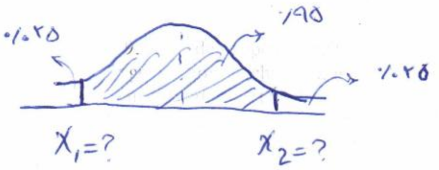
$$P(Z \le 1.645) = .95$$

با مراجعه به جدول

$$X = \mu + \sigma Z \Rightarrow X = 752 + 86(1.645) = 893.47$$

یعنی شرکت با خرید ۸۹۳٫۴۷ تن مواد اولیه با ۹۵ درصد اطمینان کمبودی نخواهد داشت.

مثال: دستگاهی طوری تنظیم شده است که به صورت خودکار و به طور متوسط پیچ‌هایی به قطر ۳۸ میلیمتر و با انحراف معیار ۰٫۲ تولید می‌کند. می‌خواهیم بدانیم ۹۵ درصد پیچ‌های تولیدی در چه دامنه‌ای در دو طرف میانگین قرار می‌گیرند.



هدف پیدا کردن مقدار X_1 و X_2 است. جمع مساحت‌های هاشور نخورده در سمت چپ و راست منحنی برابر است با $1 - .95 = .05$. به دلیل متقارن بودن توزیع نرمال حول میانگین، آن دو با هم برابرند و مساحت هر یک برابر با .025 است. پس

$$P(Z \ge 1.96) = .025$$

$$P(Z \le -1.96) = .025$$

$$X_1 = 38 - .2(1.96) = 37.608$$

$$X_2 = 38 + .2(1.96) = 38.392$$

۹۵ درصد پیچ‌های تولیدی دارای قطر بین ۳۷٫۶۰۸ و ۳۸٫۳۹۲ هستند.

14/C

نمونه‌گیری

در بسیاری از زمینه‌های کاربردی، محققان در صدد تخمین پارامترهای جامعه هستند ولی دسترسی به آنها به طور مستقیم با استفاده از شمارش(سرشماری) جامعه آماری امکان پذیر نیست. در چنین موقعیتی محققان ناچارند برای استنباط پارامترهای مورد نظر به نمونه‌هایی از جوامع آماری اکتفا کنند.

اگر پارامتر را شاخص به دست آمده از جامعه آماری با استفاده از سرشماری بنامیم، پس شاخص به دست آمده از یک نمونه n تایی از جامعه را آماره گویند.

یکی از مهمترین جنبه‌های آمار استنباطی تعمیم آماره به پارامتر است.

قبل از ورود به بحث توزیع آماره و خواص آن ابتدا روشهای نمونه‌گیری را که الفبای هر نوع کاربرد آمار استنباطی است را بررسی می‌کنیم.

دلایل نمونه‌گیری

دلایل متعددی وجود دارد که محققان از سرشماری کل جامعه صرف نظر کنند و به تعداد محدودی از عناصر جامعه اکتفا می‌کنند که عبارتند از:

۱- هزینه: نمونه می‌تواند اطلاعات قابل اعتماد و مفیدی با هزینه کمتر از سرشماری فراهم کند.

۲- به روز بودن: نمونه اغلب اطلاعات به‌هنگام‌تر از سرشماری به دست می‌دهد زیرا داده‌های کمتری جمع‌آوری و تجزیه و تحلیل می‌شوند. این جنبه از نمونه به خصوص زمانی که اطلاعات برای تصمیم‌گیری سریع مورد نیاز است اهمیت بیشتری می‌یابد. به عنوان مثال چنانچه هدف بررسی خط تولید برای کنترل آن باشد، بررسی کل محصولات تولید شده در یک روز غیرعملی است، پس نمونه‌گیری ناچار می‌شود.

۱۵/۴

۳ـ درستی: نمونه اغلب اطلاعاتی را درستی سرشماری فراهم می‌کند. زیرا خطاهای جمع‌آوری داده‌ها در یک کار تحقیقاتی کوچک کمتر از یک کار تحقیقاتی بزرگ است.

۴ـ زمان: سرشماری کل جامعه آماری به زمان طولانی نیاز دارد، به همین دلیل با استفاده از نمونه‌گیری می‌توان با زمان کمتر اطلاعات لازم را بدست آورد

۵ـ آزمون تخریب کننده:

وقتی که آزمون باعث خراب شدن یک کالا می‌شود باید نمونه‌گیری به کار برد

این امر در صنایع لامپ‌سازی، کنسروسازی کالاها با تخریب همراه است بیشتر مصداق دارد

به عنوان مثال تولیدکننده ظروف شیشه‌ای نشکن

روش‌های نمونه‌گیری

۱ـ نمونه‌گیری تصادفی ساده:

در نمونه‌گیری تصادفی ساده هر یک از عناصر جامعه مورد نظر برای انتخاب شدن شانس مساوی دارد

الف) قرعه‌کشی

ب) جدول اعداد تصادفی

۲ـ نمونه‌گیری منظم:

روش نمونه‌گیری منظم روش تغییر شکل یافته نمونه‌گیری تصادفی است. این روش برای آن دسته از جوامع آماری که از قبل به ترتیب معین شده مرتب شده‌اند (ملاک، شماره دانشجویی) کاربرد فراوان دارد

۱۶/۵

۳- نمونه‌گیری گروهی:

برای بیشتر کردن شباهت نمونه و جامعه و افزایش دقت نمونه برداری برای برآورد پارامترهای جامعه و دخالت دادن ویژگی‌های جامعه در نمونه، در این روش، جامعه به گروه‌های همگن تقسیم و هر گروه از افرادی تشکیل می‌شود که ویژگی‌های مشابه دارند

واحد تولیدی　واحد مالی　واحد اداری

۴- نمونه‌گیری خوشه‌ای:

هرگاه جامعه مورد بررسی خیلی وسیع و گسترده باشد و تهیه فهرست از اعضای جامعه امکان پذیر نباشد انتخاب نمونه از طریق خوشه‌ای شکل است. به عنوان مثال فرض کنید می‌خواهیم میزان تحصیلات کارمندان شهر تهران را بررسی کنیم، ابتدا از سازمان‌ها انتخاب می‌کنیم سپس از سازمان‌ها کارمندان را از این سازمان‌ها انتخاب می‌کنیم

۵- نمونه‌گیری مرحله‌ای:

شکل گسترش یافته نمونه‌گیری خوشه‌ای است. در این روش عناصر نمونه اصلی طی چند مرحله انتخاب می‌شوند

به عنوان مثال ابتدا چند سازمان انتخاب، بعد از بین هر سازمان چند واحد سازمانی را معین و پس از آن عناصر نمونه را از هر واحد به صورت تصادفی انتخاب می‌کنیم.

۱۷/۵

روش نمونه‌گیری تصادفی با جایگذاری و بدون جایگذاری

اگر فرض کنیم که در یک جامعه آماری با N عضو، یک نمونه تصادفی به حجم n انتخاب شود چنانچه هر یک از عناصر نمونه پس از انتخاب و یادداشت صفت مورد نظر مجدداً به جامعه برگردد و دوباره بخت عضو نمونه شدن داشته باشد و این روند تا انتخاب n امین عضو نمونه ادامه داشته باشد به این نمونه‌گیری تصادفی با جایگذاری معروف است.

واضح است که احتمال انتخاب هر یک از عناصر نمونه از جامعه آماری مساوی $\frac{1}{N}$ است.

حال چنانچه عناصر نمونه پس از انتخاب به جامعه برنگردند و این روند تا انتخاب n امین عضو نمونه ادامه داشته باشد آن را نمونه‌گیری تصادفی ساده بدون جایگذاری می‌نامیم.

واضح است که احتمال انتخاب اولین عضو برابر $\frac{1}{N}$ و دومین عضو نمونه برابر $\frac{1}{N-1}$ و سرانجام n امین عضو نمونه برابر $\frac{1}{N-n+1}$ است.

در نمونه‌گیری با جایگذاری ~~امکان~~ امکان انتخاب یک فرد برای چندین مرتبه وجود دارد اما در نمونه‌گیری بدون جایگذاری امکان تکرار یک عضو از جامعه وجود ندارد.

نکته مهم: اگر نمونه با جایگذاری انجام بشود مشاهدات تصادفی $x_1, x_2, \dots, x_n$ مستقل خواهند بود. در جامعه‌های بزرگ حتی اگر نمونه‌گیری بدون جایگذاری انجام شود در عمل شبیه نمونه‌گیری با جایگذاری است.

نمونه‌گیری از جامعه‌های بزرگ را نمونه‌گیری از جامعه نامحدود گویند که براساس این قاعده بیان می‌شود

$$\frac{n}{N} \le \%5$$

در نمونه‌گیری با جامعه کوچک با فرض ~~...~~ بدون جایگذاری مشاهدات را می‌توان مستقل از هم فرض کرد. در تمامی بحث‌ها ~~...~~ که در ادامه پیش رو داریم تعریف از نمونه‌ها نمونه‌گیری تصادفی با جایگذاری است.

۱۸/۵

توزیع‌های نمونه‌گیری

برای استنباط درباره پارامترهای جامعه باید از یک آماره استفاده کرد. بنابراین متناظر با هر پارامتر یک آماره وجود دارد که خود یک متغیر تصادفی است. هر آماره دارای یک توزیع نمونه‌گیری (تابع احتمال) است. تابع احتمال یک آماره، تابعی است که بر اساس نمونه‌های تصادفی n تایی که به طور مکرر از جامعه آماری انتخاب شده است به دست می‌آید. این تابع را توزیع نمونه‌گیری آماره گویند.

~~براساس قضیه~~

قضیه حد مرکزی:

بر اساس قضیه حد مرکزی، مجموع و میانگین مقادیر یک نمونه تصادفی n تایی که از یک جامعه آماری انتخاب می‌شوند، به طور تقریبی به یک توزیع نمونه‌گیری قرینه گرایش دارند.

در قضیه حد مرکزی اگر یک نمونه تصادفی n تایی از یک جامعه غیر نرمال با میانگین μ و انحراف معیار σ معین انتخاب شود، وقتی n بزرگ باشد، توزیع نمونه‌گیری میانگین نمونه، $\bar{x}$، تقریباً به صورت نرمال توزیع خواهد شد و این میانگین و انحراف معیار را خواهد داشت:

$$\mu_{\bar{x}} = \mu_x$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

هرچه حجم نمونه بزرگتر و بزرگتر می‌شود، تقریب بیشتر و بیشتر به توزیع نرمال نزدیک شده و دقیق‌تر خواهد شد.

بسیاری از محققان معتقدند که صرف نظر از توزیع جامعه آماری حداقل یک نمونه ۳۰ تایی لازم است تا بتوان گفت توزیع آماره $\bar{x}$ نرمال است.

۱۹/۵

مثال) توزیع نمره‌های ارزشیابی کارمندان یک سازمان نرمال است. متوسط نمره‌های ارزشیابی آنان ۱۵ و انحراف معیارشان ۳ است. احتمال اینکه نمره یکی از کارمندان حداقل ۱۸ باشد چقدر است

توزیع نرمال است پس

$$P(X \geq ۱۸) = P\left(Z \geq \frac{۱۸-۱۵}{۳}\right)$$

$$= P(Z \geq ۱) = ۱ - ۰/۸٤۱۳ = ۰/۱۵۸۷$$

مثال) فرض کنید از جامعه کارمندان مثال قبل یک نمونه تصادفی ۹ نفره انتخاب کرده‌ایم. احتمال اینکه میانگین نمره ارزشیابی آنها حداقل ۱۸ باشد چقدر است؟

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

می‌دانیم

$$\mu_{\bar{x}} = \mu_x$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

پس

$$Z = \frac{\bar{X} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}} = \frac{۱۸-۱۵}{\frac{۳}{\sqrt{۹}}} = ۳$$

$$P(\bar{x} \geq ۱۸) = P(Z \geq ۳) = ۱ - ۰/۹۹۸۷ = ۰/۰۰۱۳$$

۲۰/۴

الگوی رفتار $\bar{x}$

هرگاه یک نمونه n تایی از یک جامعه محدود با N عضو که دارای میانگین μ و واریانس σ^2 است با همه مشخصات انتخاب شود، توزیع میانگین نمونه این خواص را دارد

of ... their demand ... known, (2) the ... DCs and
the ... known, (3) ... presents a
sin... SCM

$$\mu_{\bar{x}} = \mu_x$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

در نمونه‌گیری از جامعه‌های محدود

$$\text{ضریب تصحیح} = \frac{N-n}{N-1}$$

عامل اصلاح جامعه محدود

بیشتر اوقات به آن توجه نمی‌شود به خصوص زمانی که نسبت n به N حداکثر ۵ درصد باشد.

در یک جامعه بسیار بزرگ فرض بر آن است که اندازه نمونه بزرگ یا نمونه‌برداری از جامعه نامحدود است یعنی ضریب $\frac{N-n}{N-1}$ برابر با یک فرض می‌شود.

~~۲۱~~

مثال) میانگین طول عمر محصولات کارخانه ۱۰۰ ساعت و پراکندگی آن ۱۶ ساعت است. یک نمونه تصادفی ۶۴ تایی از محصولات تولیدی انتخاب شده است. احتمال اینکه میانگین طول عمر آنها دست کم ۹۵ ساعت باشد چقدر است

براساس قضیه حد مرکزی، توزیع $\bar{x}$ برای نمونه ۶۴ تایی صرف نظر از توزیع عمر محصولات نرمال است

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$P(\bar{x} \geq 95) = P\left(z \geq \frac{95 - 100}{\frac{16}{\sqrt{64}}}\right) = P(z \geq -2.5)$$

$$= 1 - 0.0062 = 0.9938$$

یعنی ۹۹.۳۸٪ احتمال دارد که متوسط عمر نمونه‌های ۶۴ تایی دست کم ۹۵ ساعت باشد

۲۲/۴

مثال) یک نمونه ۱۰۰ تایی از یک جامعه غیرنرمال با میانگین ۵۰ و انحراف معیار ۱۰ انتخاب شده است مفروض است

الف) مقدار میانگین و انحراف معیار توزیع نمونه‌گیری $\bar{x}$ چقدر است

ب) با استفاده از قضیه حد مرکزی این احتمالات را محاسبه کنید

$$P(\bar{x} > 52) \quad , \quad P(\bar{x} \leq 52)$$

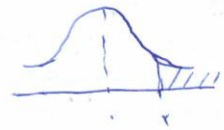
$$\mu_{\bar{x}} = \mu_x = 50$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \Rightarrow \sigma_{\bar{x}} = \frac{10}{\sqrt{100}} = 1$$

$$P(\bar{x} > 52) = P\left(z > \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$$

$$P\left(z > \frac{52 - 50}{1}\right) = P(z > 2) =$$

$$1 - P(z < 2) = 1 - .97772 =$$



$$P(\bar{x} \leq 52) = P\left(z \leq \frac{52 - 50}{1}\right) = P(z \leq 2) = .97772$$

توزیع نمونه‌گیری نسبت موفقیت در نمونه ($\bar{P}$)

بسیاری از تحقیقات از مقیاس اسمی یا رتبه‌ای برخوردارند که با استفاده توصیف آن از نسبت موفقیت‌ها به مواردی همچون درصد کالاهای معیوب، درصد کارمندانی که از کار خود راضی هستند و ... . همچنان که مشاهده می‌شود این تحقیقات و موارد آن به می‌توانند در قالب توزیع دوجمله‌ای بیان شوند

در توزیع دوجمله‌ای P نسبت موفقیت در جامعه است و q نسبت شکست در جامعه است.

در یک آزمایش دوجمله‌ای با n بار تکرار عمل، تعداد موفقیت‌ها را x و با نسبت موفقیت در نمونه $\frac{x}{n}$ است

$$\bar{P} = \frac{X}{n} \quad \text{(نسبت موفقیت در نمونه)}$$

طبق قضیه حد مرکزی، توزیع دوجمله‌ای از توزیع نرمال برخوردار است.

در صورتی که نمونه‌گیری از یک جامعه آماری محدود انجام گیرد ($\frac{n}{N} \geq 0/05$) در این صورت رابطه‌های زیر برقرار است

$$E(\bar{P}) = \mu_{\bar{P}} = P$$

$$\sigma^2_{\bar{P}} = \frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$$

اثبات مثال صفحه ۳۵-۳۶

واضح است که همانند توزیع $\bar{x}$ واریانس آن را باید با ضریب کاهش $\frac{N-n}{N-1}$ تصحیح نمود.

چنانچه نمونه‌گیری از جامعه نامحدود (با جایگذاری) انجام گیرد، به راحتی می‌توان از ضریب کاهش $\left(\frac{N-n}{N-1}\right)$ در محاسبه واریانس $\bar{P}$ صرف نظر کرد. پس داریم

$$\mu_{\bar{P}} = P$$

$$\sigma^2_{\bar{P}} = \frac{P(1-P)}{n}$$

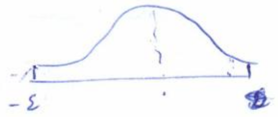
مثال) مدیر کارخانه‌ای درصدد تولید کالای جدید است. او می‌خواهد یک نمونه تصادفی ۱۰۰ تایی از بین مشتریان بالقوه انتخاب کند و با آن‌ها مصاحبه کند تا درصد تمایل در نمونه‌های ۱۰۰ تایی برای او مشخص شود. فرض کنید ۵۰ درصد از مشتریان بالقوه به خرید کالا تمایل دارند. احتمال اینکه دست‌کم ۳۰ درصد نمونه انتخاب شده به مصرف کالا تمایل داشته باشند، چقدر است؟

با توجه به فرض جامعه نامحدود و حجم نمونه است، از اساس قضیه حد مرکزی می‌توان از توزیع نرمال برای حل این مسئله استفاده کرد

$$P(\bar{p} \geq 0.30) = P\left(Z \geq \frac{\bar{p} - \mu_{\bar{p}}}{\sigma_{\bar{p}}}\right) =$$

$$P\left(Z \geq \frac{\bar{p} - P}{\sqrt{\frac{P(1-P)}{n}}}\right) = P\left(Z \geq \frac{0.3 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}}\right)$$

$$P(Z \geq -4) = 1 - P(Z < -4) = 1$$



یعنی با احتمال ۱۰۰ درصد ۳۰ درصد از نمونه انتخاب شده به مصرف کالا تمایل دارند

۲۵/۴

مثال) تحقیقات نشان می‌دهد که نسبت مدیران وظیفه‌مدار در سازمان‌های کشور ۶۰ درصد است از بین مدیران کشور یک نمونه تصادفی ۴۰۰ نفره انتخاب شده است. احتمال اینکه کمتر از ۵۵ درصد آنها وظیفه‌مدار باشند چقدر است

حجم نمونه نسبت به جامعه
با توجه به حجم نمونه فرض جامعه نامحدود است پس با استفاده از قضیه حد مرکزی

$$P(\bar{P} < 0.55) = P\left(Z < \frac{\bar{P} - \mu_{\bar{P}}}{\sigma_{\bar{P}}}\right) =$$

$$P\left(Z < \frac{\bar{P} - P}{\sqrt{\frac{P(1-P)}{n}}}\right) = P\left(Z < \frac{0.55 - 0.6}{\sqrt{\frac{0.6(1-0.6)}{400}}}\right) =$$

$$P\left(Z < \frac{-0.05}{\sqrt{\frac{0.24}{400}}}\right) = P\left(Z < \frac{-0.05}{0.02}\right) =$$

$$P(Z < -2.5) = 0.0062$$

با احتمال ۰٫۰۰۶۲ ~~[illegible]~~ افراد وظیفه‌مدار کمتر از ۵۵٪ نمونه انتخاب شده هستند

۲۸/ج

هدف از تحلیل و توصیف در آمار، جامعه آماری است که به دو طریق توصیف می‌شود

۱ـ با استفاده از سرشماری کل جامعه آماری و محاسبه پارامتر ← فنون آمار توصیفی

۲ـ با استفاده از تخمین زننده برای برآورد پارامتر.

شاخه‌ای از آمار را استنباطی گویند که شامل فنون تخمین آماری و آزمون فرضیه‌ها می‌شود که به نوع تحقیق بستگی دارد

اگر تحقیق از نوع سوالی و صرفاً حاوی پرسش در مورد پارامتر باشد ← از تخمین آماری استفاده می‌شود

اگر تحقیق حاوی فرضیه‌ها بود و از مرحله سوال گذر کرده بود ← از آزمون فرضیه استفاده می‌شود

تخمین آماری (برآورد کردن)

در این فصل چگونگی بدست آوردن پارامترهای جامعه را از روی داده‌های نمونه تصادفی را توضیح می‌دهیم

میانگین و انحراف معیار جامعه دو پارامتر مهم هستند که در صورت نامعلوم [illegible] بودن مقادیرشان باید آنها را برآورد کنیم.

برآورد فاصله‌ای:

با استفاده از برآورد نقطه‌ای پارامتر جامعه، حدودی را برای پارامتر مورد نظر پیدا می‌کنیم که این حدود [illegible] بازه یا فاصله را مشخص می‌کنند.

حد پایین ← $l < \theta < u$ → حد بالا

θ: پارامتر مجهول

فرض کنید متغیر تصادفی Z دارای توزیع نرمال استاندارد باشد. احتمال اینکه

$$P(-1.96 < Z < 1.96) = 0.95$$

در حالت کلی برای متغیر تصادفی Z رابطه زیر همواره برقرار است

$$P[-Z_{\alpha/2} < Z < Z_{\alpha/2}] = 1-\alpha$$

به $(1-\alpha)$ فاصله اطمینان یا سطح اطمینان گویند

و α را سطح خطا گویند به عنوان مثال در سطح اطمینان ۹۵ درصد، سطح خطا ۵ درصد خواهد بود

به عنوان مثال $\alpha = 0.05$ رابطه زیر برقرار است

$$P[-Z_{0.05/2} < Z < Z_{0.05/2}] = 1 - 0.05$$

$$P[-Z_{0.025} < Z < Z_{0.025}] = 0.95$$

$$P[-1.96 < Z < 1.96] = 0.95$$

اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس معلوم σ^2 باشد بر اساس یک نمونه n تایی یک فاصله اطمینان با ضریب اطمینان $100(1-\alpha)\%$ برای μ به صورت زیر بدست می‌آید

می‌دانیم که متغیر $\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ دارای توزیع نرمال استاندارد است پس:

$$P\left[-z_{\alpha/2} < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right] = 1-\alpha \Rightarrow$$

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

به عنوان مثال برای $\alpha = 0.05$

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

۱۲/G

سطح اطمینان	$Z_{\alpha/2}$
٪۹۰ $\longrightarrow$	$Z_{۰٫۰۵} = ۱٫۶۴۵$
٪۹۵ $\longrightarrow$	$Z_{۰٫۰۲۵} = ۱٫۹۶$
٪۹۹ $\longrightarrow$	$Z_{۰٫۰۰۵} = ۲٫۵۷۵$

~~برآورد فاصله‌ای μ برای نمونه‌های بزرگ~~

توزیع جامعه آماری نرمال با انحراف معیار معلوم (نمونه بزرگ)

فرض اساسی در این حالت این است که نمونه از یک جامعه آماری نرمال با انحراف معیار معلوم انتخاب شده است

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

$$P\left(-Z_{\alpha/2} < \frac{\bar{x}-\mu_x}{\frac{\sigma_x}{\sqrt{n}}} < Z_{\alpha/2}\right) = 1-\alpha$$

$$\bar{x} - Z_{\alpha/2} \times \frac{\sigma_x}{\sqrt{n}} < \mu_x < \bar{x} + Z_{\alpha/2} \times \frac{\sigma_x}{\sqrt{n}}$$

مثال: بررسی‌ها نشان می‌دهد که توزیع وزن محصولات تولید شده یک کارخانه بزرگ، نرمال و انحراف معیار آن ۲۱ تن است. از آنجا که اندازه‌گیری وزن محصولات بطور روزانه امکان‌پذیر نیست، یک نمونه ۵۰ روزه از تولیدات انتخاب شده که میانگین وزن آن ۸۷۱ تن است. در سطح اطمینان ۹۰ درصد میانگین واقعی وزن محصولات تولید شده در یک روز را حساب کنید.

$$1-\alpha = ۰٫۹ \Rightarrow \alpha/۲ = ۰٫۰۵ \Rightarrow Z_{۰٫۰۵} = ۱٫۶۴۵$$

$$P\left(\bar{x} - Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}} < \mu_x < \bar{x} + Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right) = ۰٫۹$$

$$۸۷۱ - ۱٫۶۴۵ \times \frac{۲۱}{\sqrt{۵۰}} < \mu_x < ۸۷۱ + ۱٫۶۴۵ \times \frac{۲۱}{\sqrt{۵۰}}$$

$$۸۶۶٫۱۱ < \mu_x < ۸۷۵٫۸۹$$

با احتمال ۹۰ درصد میانگین وزن محصولات کارخانه بین این دو مقدار است که فقط ۱۰ درصد احتمال دارد که خارج از این دو مقدار باشد.

مثال: ... برای نمونه‌ای به اندازه $n=400$ بسته که در روزی خاص به وسیله اداره پست جابجا می‌شود ... ثبت شده است. میانگین نمونه عبارت است از $\bar{x}=364.7$ ...

حال اگر انحراف معیار جامعه معلوم باشد و برابر با ۳۰۰ باشد μ را در سطح اطمینان ۹۵٪ بدست آورید

$$1-\alpha=0.95 \Rightarrow \alpha=0.05 \Rightarrow \alpha/2=0.025 \Rightarrow Z_{0.025}=1.96$$

$$P\left(\bar{x}-Z_{\alpha/2}\times\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)=0.95$$

$$364.7-1.96\times\frac{300}{20} < \mu < 364.7+1.96\times\frac{300}{20}$$

نکته: هرچه ضریب اطمینان افزایش می‌یابد ← محدوده اعداد بزرگتر می‌شود ← از دقت برآورد کاسته می‌شود

٪۹۰	۱٫۶۴۵
	۱٫۹۶
٪۹۵	۲٫۵۷۵
٪۹۹	

توزیع میانگین آماری نرمال و انحراف معیار نامعلوم

هرگاه انحراف معیار جامعه نامعلوم باشد در تخمین فاصله‌ای μ ناچار $S_{\bar{x}}$ جایگزین $\sigma_{\bar{x}}$ خواهد شد یعنی $t=\frac{\bar{x}-\mu}{S_{\bar{x}}}$

آماره $\frac{\bar{x}-\mu}{S_{\bar{x}}}$ دارای توزیع t استیودنت است که توزیعی قرینه ولی دقت کمتری نسبت به توزیع نرمال دارد (پراکندگی بیشتر)

درجه آزادی: مجموع ۵ عدد ۲۵ ← ۴ تاش آزادیم هرچی بگیم – ولی عدد پنجم بستگی به انتخاب ماقبلی دارد

$$df=5-1=4 \Rightarrow df=n-1$$

در جدول در صورتی که درجه آزادی ۳۰ یا بیشتر بود از سطر آخر جدول t همان سطر Inf استفاده می‌کنیم

$$P\left(-t_{\alpha/2,df} \leqslant t \leqslant +t_{\alpha/2,df}\right) = 1-\alpha \Rightarrow$$

$$P\left(-t_{\alpha/2,df} \leqslant \frac{\bar{x}-\mu_x}{S_{\bar{x}}} \leqslant t_{\alpha/2,df}\right) = 1-\alpha \Rightarrow$$

$$\boxed{\bar{x} - t_{\alpha/2,df}\, S_{\bar{x}} \leqslant \mu_x \leqslant \bar{x} + t_{\alpha/2,df} \times S_{\bar{x}}}$$

در این رابطه $S_{\bar{x}} = \frac{S_x}{\sqrt{n}}$

براساس قضیه حد مرکزی هرچه حجم نمونه بزرگ‌تر شود، توزیع t استیودنت به سمت توزیع نرمال میل می‌کند

اگر $n > 30$ باشد می‌توان به جای توزیع t استیودنت از توزیع نرمال استفاده کرد

$$Z = \frac{\bar{x}-\mu_x}{S_{\bar{x}}} = t$$

پس

$$\boxed{P\left(\bar{x} - Z_{\alpha/2} \cdot S_{\bar{x}} \leqslant \mu_x \leqslant \bar{x} + Z_{\alpha/2} \cdot S_{\bar{x}}\right) = 1-\alpha}$$

مثال) نمونه‌ای مرکب از ۵ قوطی رب به تصادف از خط تولید کارخانه‌ای انتخاب می‌شود، محتوای ماده هر قوطی وزن می‌شود. تجربه گذشته نشان می‌دهد که توزیع وزن ماده هر قوطی نرمال است، در فاصله اطمینان ۹۹ درصد μ چقدر است؟

نتایج نمونه به صورت زیر است:

$$x_i = 230,\ 235,\ 235,\ 250,\ 245$$

$$\bar{x} = \frac{\sum x_i}{n} \Rightarrow \bar{x} = \frac{1195}{5} = 239$$

$$S^2 = \frac{\sum(\bar{x}-x_i)^2}{n-1} = \frac{81+16+16+121+36}{4} = \frac{270}{4} = 67.5 \Rightarrow S = \sqrt{67.5} = 8.21$$

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = 3.67$$

$$1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \alpha/2 = 0.005$$

نمونه کوچک - جامعه نرمال - σ نامعلوم $\Leftarrow$

$$\bar{x} - t_{\alpha/2,df} \cdot S_{\bar{x}} \leqslant \mu \leqslant \bar{x} + t_{\alpha/2,df} \cdot S_{\bar{x}}$$

$$239 - 4.604 \times 3.67 \leqslant \mu \leqslant 239 + 4.604 \times 3.67$$

$$222 \leqslant \mu \leqslant 256$$

توزیع جامعه غیرنرمال باشد.

چنانچه حجم نمونه کمتر از ۳۰ باشد $(n<30)$ و جامعه به طور غیرنرمال توزیع شده باشد (جامعه نرمال نباشد) برای تنظیم فاصله اطمینان نمی‌توان از توزیع نرمال و t استیودنت استفاده کرد. در این حالت از قضیه چبی‌شف استفاده می‌کنیم.

$$P\left(-K\sigma_{\bar{x}} \leq \bar{x}-\mu_x \leq K\sigma_{\bar{x}}\right) \geq 1-\frac{1}{K^2}$$

$$P\left(\bar{x}-K\sigma_{\bar{x}} \leq \mu_x \leq \bar{x}+K\sigma_{\bar{x}}\right) \geq 1-\alpha$$

برای پیدا کردن K ،

$$1-\frac{1}{K^2}=1-\alpha$$

اگر σ_x نامعلوم باشد از $S_{\bar{x}}$ استفاده می‌کنیم

$$S_{\bar{x}}=\frac{S}{\sqrt{n}}$$ نمونه

$$P\left(\bar{x}-KS_{\bar{x}} \leq \mu_x \leq \bar{x}+KS_{\bar{x}}\right) \geq 1-\alpha$$

مثال) از شیشه‌هایی که با یک دستگاه پر شده است، یک نمونه تصادفی ۲۵ تایی انتخاب می‌کنیم، میانگین آن ۲۵۰ میلی‌لیتر و انحراف معیار آن ۱۰ میلی‌لیتر است. هیچ دلیلی بر نرمال بودن توزیع مایع ریخته شده در شیشه‌ها وجود ندارد. در سطح اطمینان ۹۵ درصد، میانگین کل مایع ریخته شده در شیشه‌ها را برآورد کنید.

$$1-\frac{1}{K^2}=1-\alpha \Rightarrow -\frac{1}{K^2}=-\alpha \Rightarrow K^2=\frac{1}{\alpha} \Rightarrow K=\sqrt{\frac{1}{\alpha}}$$

$$K=\sqrt{\frac{1}{0.05}}=4.47$$

σ از جامعه معلوم نیست ⇐

$$S_{\bar{x}}=\frac{10}{\sqrt{25}}=2$$

$$P\left(250-4.47\times 2 \leq \mu_x \leq 250+4.47\times 2\right) \geq 0.95$$

$$\left(241.06 \leq \mu_x \leq 258.94\right)$$

با حداقل ۹۵ درصد اطمینان می‌توان گفت میانگین کل مایع بین

تخمین فاصله‌ای تفاضل میانگین دو جامعه. $(\mu_1 - \mu_2)$

همانطور که برآورد میانگین واقعی یک جامعه آماری اهمیت دارد، به همان اندازه شاید هم بیشتر، مقایسه دو جامعه آماری با استفاده از میانگین آنها نیز برای تصمیم‌گیری اهمیت دارد. به عنوان مثال مقایسه دو شرکت از نظر سود، درآمد، هزینه، کیفیت، [redacted] و ...

$\mu_1 \rightarrow$ میانگین جامعه اول $\qquad n_1 \rightarrow$ تعداد نمونه جامعه اول $\qquad \bar{x}_1 \qquad S_1$

$\mu_2 \rightarrow$ میانگین جامعه دوم $\qquad n_2 \rightarrow$ تعداد نمونه جامعه دوم $\qquad \bar{x}_2 \qquad S_2$

برای مقایسه میانگین دو جامعه اولین روشی که به ذهن هر تصمیم‌گیرنده می‌رسد، تخمین فاصله‌ای است

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2} = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n} \Rightarrow \sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

توزیع دو جامعه آماری نرمال و σ_1 و σ_2 معلوم

چون توزیع جامعه نرمال و σ_1 و σ_2 معلوم است پس $\bar{x}_1$ و $\bar{x}_2$ نیز نرمال است.

بنابراین متغیر استاندارد را می‌توان به این صورت نوشت

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}}$$

با جایگذاری در رابطه

$$P\left(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\right) = 1 - \alpha$$

داریم

$$P\left(-Z_{\alpha/2} \leq \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}} \leq Z_{\alpha/2}\right) = 1-\alpha$$

$$\boxed{P\left((\bar{x}_1-\bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \leq \mu_1-\mu_2 \leq (\bar{x}_1-\bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right) = 1-\alpha}$$

مثال: در تحقیقی که در خصوص مقایسه عملکرد کارکنان دو سازمان بوده است از هر دو سازمان نمونه‌ای تهیه شده است از سازمان A یک نمونه تصادفی ۲۵ نفره با میانگین ۶۰ و از سازمان B یک نمونه تصادفی ۲۰ نفره با میانگین ۵۵ انتخاب شده است. توزیع نمره‌های عملکرد هر دو سازمان نرمال و انحراف معیار عملکرد هر دو جامعه به ترتیب ۱۰ و ۱۲ است. در سطح اطمینان ۹۹ درصد میانگین عملکرد دو جامعه را باهم مقایسه کنید.

A	B
$n_1 = 25$	$n_2 = 20$
$\bar{x}_1 = 60$	$\bar{x}_2 = 55$
$\sigma_1 = 10$	$\sigma_2 = 12$

$$Z_{\alpha/2} = Z_{0.005} = 2.575$$

2α چون دو دامنه است

$$P\left((60-55) - 2.575\left(\underbrace{\sqrt{\frac{100}{25}+\frac{144}{20}}}_{8.634}\right) \leq \mu_1-\mu_2 \leq (60-55) + 2.575\sqrt{\frac{100}{25}+\frac{144}{20}}\right) = 0.99$$

$$-3.634 \leq (\mu_1-\mu_2) \leq 13.634$$

۱) اگر هر دو دامنه مثبت بودند ← در سطح اطمینان مورد نظر $\mu_2 < \mu_1$

۲) اگر هر دو دامنه منفی بودند ← " " " $\mu_1 < \mu_2$

۳) حکم در غیر این صورت (موارد ۱ و ۲) بین μ_1 و μ_2 اختلاف معناداری وجود ندارد چون فاصله اطمینان مورد نظر از نقطه صفر می‌گذرد $\mu_1 - \mu_2 = 0$

پس در سطح اطمینان ۹۹ درصد بین میانگین عملکرد کارکنان دو سازمان اختلاف معناداری وجود ندارد.

۳۴/۸

توزیع درجایی که آماره نرمال دارای انحراف معیارهای نامعلوم

زمانی که تخمین واریانس نمونه کوچک است باید از توزیع t استفاده

در بیشتر تحقیقات فرض معلوم بودن σ_1 و σ_2 تقریباً نامعقول است. بنابراین اگر تعداد نمونه کمتر از ۳۰ باشد

پس باید از توزیع t استفاده کنیم. بنابراین

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_{\bar{x}_1 - \bar{x}_2}}$$

$$S_{\bar{x}_1 - \bar{x}_2} \xrightarrow[\text{واریانس دو جامعه برابر باشد}]{\text{فرض } \sigma_1^2 = \sigma_2^2} \xrightarrow[\text{واریانس مشترک آنها } (\sigma^2) \text{ براساس } s_1^2 \text{ و } s_2^2 \text{ تخمین زده می‌شود}]{\text{هر دو جامعه یک جامعه تلقی می‌شوند}}$$

$$S_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

و

$$S_{\bar{x}_1 - \bar{x}_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

پس

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \cancel{(\mu_1 - \mu_2)}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

درجه آزادی این متغیر $df = n_1 + n_2 - 2$

$$P\left(-t_{\alpha/2, df} \leq t \leq t_{\alpha/2, df}\right) = 1 - \alpha$$

$$P\left((\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, df} \cdot S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, df} \cdot S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1 - \alpha$$

واضح است که t براساس حجم نمونه‌ها (بیشتر از ۳۰) طبق قضیه حد مرکزی از توزیع نرمال برخوردار است و می‌توان از $Z_{\alpha/2}$ استفاده

مثال - هدف از تحقیقی مقایسه سطح آمادگی کارمندان در سازمان A و B است. در این تحقیق از سازمان A یک نمونه ۹ نفره با میانگین ۴۵ و انحراف معیار ۱۲ انتخاب شده است. از سازمان B یک نمونه ۱۵ نفره با میانگین ۵۵ و انحراف معیار ۱۴ انتخاب شده است. فرض کنید توزیع نمرات سطح آمادگی در دو سازمان نرمال و $\sigma_1^2=\sigma_2^2$ است. در سطح اطمینان ۹۵ درصد تخمین لازم را برای مقایسه دو جامعه به عمل آورید.

$$S_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{8\times(12)^2+14\times(14)^2}{9+15-2}} = 13.31$$

$$\begin{cases} \alpha/2 = 0.05 \\ df = n_1+n_2-2 = 9+15-2 = 22 \end{cases} \Rightarrow t_{\alpha/2,\,df} = 1.717$$

$$P\left((\bar{x}_1-\bar{x}_2) - t_{\alpha/2,df}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1-\mu_2 \le (\bar{x}_1-\bar{x}_2) + t_{\alpha/2,df}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right) = 1-\alpha$$

$$(45-55) - 1.717\left(13.31\times\sqrt{\tfrac{1}{9}+\tfrac{1}{15}}\right) \le \mu_1-\mu_2 \le (45-55)+1.717\left(13.31\times\sqrt{\tfrac{1}{9}+\tfrac{1}{15}}\right)$$

$$-19.625 \le (\mu_1-\mu_2) \le -0.364$$

چون هر دو دامنه منفی هستند پس در سطح ~~خطای~~ اطمینان ۹۵ درصد می‌توان گفت μ_1 کوچکتر μ_2 است

حال اگر $\sigma_1^2 \neq \sigma_2^2$ باشد

$$S_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

و درجه آزادی از فرمول زیر محاسبه می‌شود

$$df = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

۲۶

توزیع از جامعه نمونه‌گیری غیرنرمال

در این حالت با فرض عدم نرمال بودن بر اگر دو جامعه، توزیع $\bar{x}_1-\bar{x}_2$ تحت تأثیر تعداد درجه آزادی کم

اگر $df > 30$ باشد توزیع $\bar{x}_1-\bar{x}_2$ نرمال می‌شود و در غیر این صورت می‌توان از قاعده چی‌بی‌شف استفاده

کرد که به نسبتی که اهمیت بودن آن در برآورد $\mu_1-\mu_2$ از زیر آن صرف‌نظر می‌کنیم.

chebysheff

تخمین فاصله‌ای نسبت موفقیت در جامعه (P)

محققان اغلب برای تعیین نسبت موفقیت در جامعه از نمونه آماری استفاده می‌کنند. برای مثال دولتمردان با استفاده از نمونه می‌توانند نرخ بیکاری یا نسبت افراد شاغل به کل افراد را مشخص کنند.

از آنجائی که بسیاری از تحقیقات در مدیریت از مقیاس کیفی برخوردارند، استنباط نسبت موفقیت در جامعه آماری (P) با استفاده از یک نمونه n تایی اهمیت ویژه‌ای دارد. $\left(\bar{P} = \frac{x}{n}\right)$

می‌دانیم که توزیع نمونه‌گیری $\bar{P}$ در نمونه‌های بزرگ از تقریب نرمال برخوردار است و متغیر استانداردآن بصورت زیر است:

$$Z = \frac{\bar{P} - P}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n}}}$$

که در آن $P = \frac{x}{n}$ و $S_{\bar{P}} = \sqrt{\frac{\bar{P}(1-\bar{P})}{n}}$ است.

می‌دانیم که:

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

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

$$\boxed{P\left(\bar{P} - Z_{\alpha/2} \times \sqrt{\frac{\bar{P}(1-\bar{P})}{n}} < P < \bar{P} + Z_{\alpha/2} \times \sqrt{\frac{\bar{P}(1-\bar{P})}{n}}\right) = 1-\alpha}$$

مثال: هدف از تحقیقی تعیین نسبت افراد ناراضی در سازمان است. از آنجا که دسترسی به تمام افراد سازمان میسر نیست، محقق ۴۰۰ نفر از کارمندان را بطور تصادفی انتخاب کرده‌اند که فقط ۳۲ نفر از آنها از کار خود ناراضی هستند. نسبت افراد ناراضی را در سازمان در سطح خطای یک درصد برآورد کنید.

$n = ۴۰۰$

$x = ۳۲$

$\bar{P} = \frac{x}{n} = \frac{۳۲}{۴۰۰} = ۰/۰۸$

$Z_{\alpha/2} = Z_{۰/۰۰۵} = ۲/۵۷۵$

$\sqrt{\frac{\bar{P}(1-\bar{P})}{n}}$

$$P\left(۰/۰۸ - ۲/۵۷۵\left(\sqrt{\frac{۰/۰۸(۰/۹۲)}{۴۰۰}}\right) < P < ۰/۰۸ + ۲/۵۷۵\left(\sqrt{\frac{۰/۰۸(۰/۹۲)}{۴۰۰}}\right)\right) = ۰/۹۹$$

$۰/۰۳۵$

$$۰/۰۴۵ < P < ۰/۱۱۵$$

با اطمینان ۹۹ درصد می‌توان گفت تعداد افراد ناراضی در سازمان بین ۴/۵ تا ۱۱/۵ درصد کارمندان هستند.

تخمین فاصله‌ای تفاضل نسبت موفقیت در دو جامعه (P_1-P_2)

اگر داده‌های جمع آوری شده از جامعه‌ها از نوع میانگین پذیر باشند از تخمین فاصله‌ای $\mu_1-\mu_2$ برای مقایسه آنها استفاده می‌شود. ولی اگر داده‌ها از نوع کیفی باشند به ناچار باید از مقایسه نسبت موفقیت دو جامعه آماری استفاده کرد.

$P_1 \rightarrow$ نسبت موفقیت در جامعه اول   آماره $\bar{P}_1 = \frac{x_1}{n_1}$

$P_2 \rightarrow$ " " " دوم   آماره $\bar{P}_2 = \frac{x_2}{n_2}$

فرضیات : ۱- نمونه‌های تصادفی مستقل از دو جامعه [illegible] بگیرند

۲- در نمونه تصادفی هر دو بزرگ هستند

$$\mu_{\bar{P}_1-\bar{P}_2} = \mu_{\bar{P}_1} - \mu_{\bar{P}_2} = P_1 - P_2$$

$$\sigma^2_{\bar{P}_1-\bar{P}_2} = \sigma^2_{\bar{P}_1} + \sigma^2_{\bar{P}_2} = \frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}$$

آماره $\bar{P}_1-\bar{P}_2$ دارای توزیع نرمال است پس

$$Z = \frac{(\bar{P}_1-\bar{P}_2)-(P_1-P_2)}{\sigma_{\bar{P}_1-\bar{P}_2}}$$

$$P\left((\bar{P}_1-\bar{P}_2) - Z_{\alpha/2}\,\sigma_{\bar{P}_1-\bar{P}_2} \leqslant P_1-P_2 \leqslant (\bar{P}_1-\bar{P}_2) + Z_{\alpha/2}\,\sigma_{\bar{P}_1-\bar{P}_2}\right) = 1-\alpha$$

۳۹/ج

مثال: تحقیقی در ارتباط با مقایسه درصد مدیران برخوردار از روش A در سازمانهای دولتی و خصوصی انجام شده است.
بدین منظور از میان مدیران دولتی یک نمونه ۵۰۰ تایی و از مدیران خصوصی یک نمونه ۴۰۰ تایی انتخاب شده است.
۱۰۰ نفر از مدیران دولتی و ۵۰ نفر از مدیران سازمانهای خصوصی از روش A ~~...~~ استفاده می‌کنند. در سطح اطمینان ۹۵ درصد
نسبت واقعی مدیران برخوردار از روش A را در دو جامعه آماری مقایسه کنید.

$n_1 = 500 \qquad n_2 = 400$

$x_1 = 100 \qquad x_2 = 50$

$\bar{P}_1 = 0.2 \qquad \bar{P}_2 = \frac{50}{400} = 0.125$

$Z_{\alpha/2} = Z_{0.025} = 1.96$

$$P\left((0.2 - 0.125) - 1.96 \sqrt{\frac{0.2(0.8)}{500} + \frac{(0.125)(0.875)}{400}} < P_1 - P_2 < (0.2 - 0.125) + 1.96\sqrt{\bigcirc} \right) = 0.95$$

$$(0.027 < P_1 - P_2 < 0.124)$$

چون هر دو دامنه مثبت هستند پس در سطح اطمینان ۹۵ درصد نسبت واقعی مدیران برخوردار از روش A
در سازمانهای دولتی بیشتر از نسبت ~~...~~ مدیران برخوردار از روش A در سازمانهای خصوصی است.

آزمون فرض آماری

بسیاری از تحقیقات از مرحله سوال گذر کرده و به مرحله فرضیه می رسند. فرضیه حدسی زیرکانه درمورد دنیا (رابطه) ما بعد است. فنون آماری مناسب برای بررسی صحت فرضیه ها، فنون آزمون فرض آماری هستند

هر حکم یا ادعایی درباره جامعه را یک فرض آماری می نامند که قابل قبول بودن آن بر مبنای اطلاعات حاصل از نمونه گیری از جامعه بررسی می شود. در واقع فرضیه اثبات نمی شود بلکه تائید یا رد می شود.

چون ادعا ممکن است صحیح یا غلط باشد، بنابراین دو فرض مقابل در ذهن به وجود می آید

فرض اول) ادعا صحیح است   فرض دوم) ادعا غلط است

مثال) نسبت مدیران برخوردار از شیوه A در کشور بیش از ۶۰ درصد است. فرض کنید یک نمونه مرکب از ۲۰ مدیر انتخاب می شود و تعداد مدیران برخوردار از شیوه A ، X ثبت شود

فرضیات:

۱ـ نسبت مدیران برخوردار از شیوه A بیش از ۶۰ درصد است $P > /۶$

۲ـ نسبت مدیران برخوردار از شیوه A حداکثر ۶۰ درصد است $P \le /۶$

باید یکی از دو فرض بالا را H_0 و دیگری را H_1 بگیریم / حال کدام؟

هرگاه بخواهیم یک ادعا را از طریق تائید آن به وسیله اطلاعات حاصل از نمونه آزمون کنیم، نفی آن ادعا را فرض صفر و خود آن را فرض H_1 در نظر می گیریم. پس روش کامل

$$\begin{cases} H_0: P \le /۶ & \text{نقیض ادعا} \\ H_1: P > /۶ & \text{ادعا} \end{cases}$$

با توجه به اینکه H_0 باید تائید یا رد شود و اگر حد مشخص نداشته باشد آزمون پذیر نیست پس حتماً در فرض H_0 باید تساوی وجود داشته باشد $\Leftarrow$ پس گاهی اوقات H_0 ادعا و H_1 نقیض ادعا است

فرض صفر (H_0): به فرضی گویند که تائید یا رد می شود و در آن علامت (=) قرار دارد

فرض مقابل (H_1): به فرض مخالف H_0 گویند که در صورت عدم تائید H_0 پذیرفته می شود

* همواره باید H_0 را تائید کنیم و در صورت رد آن H_1 را بپذیریم
* ادعا ممکن است فرض H_0 یا فرض مقابل (H_1) باشد.

٤١/k

آزمون فرض یک دنباله و دو دنباله (یک طرفه، دو طرفه)

به صورت کلی H_0 را باید پذیرفت مگر آنکه دلیل محکمی بر رد آن وجود داشته باشد. این بدین معناست که فرض صفر همیشه در برگیرنده سطح اطمینان $100(1-\alpha)$ درصد است: بنابراین H_1 در برگیرنده سطح معنی داری α است.

پس H_1 همیشه به اندازه α در دنباله ها توزیع خواهد شد.

انواع آزمون فرض

آزمون دو طرفه

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

آزمون یک طرفه راست

$$\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

آزمون یک طرفه چپ

$$\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

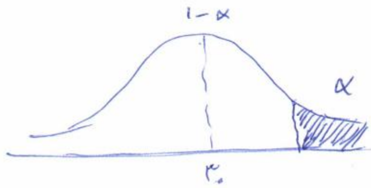
اگر میانگین نمونه در این قسمت قرار بگیرد فرض صفر رد می‌شود

مثال:

میانگین سن اساتید دانشگاه بیشتر از ۳۰ سال است

آزمون یک طرفه راست

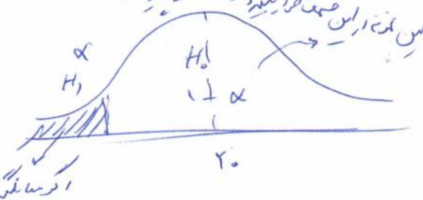
$$\begin{cases} H_0: \mu \leq 30 \\ H_1: \mu > 30 \end{cases}$$



میانگین دستمزد کارگران یک کارخانه صنعتی کمتر از ۲۰ هزار تومان است

$$\begin{cases} H_0: \mu_x \geq 20 & \text{نقیض ادعا} \\ H_1: \mu_x < 20 & \text{ادعا} \end{cases}$$

اگر میانگین نمونه در این قسمت قرار بگیرد فرض صفر نیز پذیرفته می‌شود

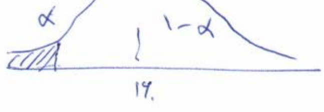


اگر میانگین نمونه در این قسمت قرار بگیرد فرض صفر رد می‌شود

میانگین قد دانشجویان دانشگاه حداقل ۱۶۰ می‌باشد

آزمون یک طرفه چپ

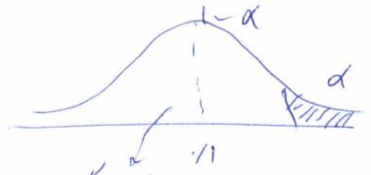
$$\begin{cases} H_0: \mu \geq 160 \\ H_1: \mu < 160 \end{cases}$$



٤٢/K

حداکثر ۱۰ درصد محصولات کارخانه الف معیوب هستند

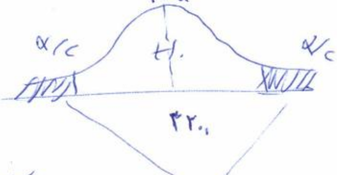
$$\begin{cases} H_0: P \le 0/1 & \text{ادعا} \\ H_1: P > 0/1 & \text{نقیض ادعا} \end{cases}$$



اگر میانگین نمونه در این قسمت قرار بگیرد فرض صفر را می‌پذیریم

میانگین هزینه‌های نگهداری هواپیما ۳۲۰۰ دلار است

آزمون دوطرفه
$$\begin{cases} H_0: \mu = ۳۲۰۰ \\ H_1: \mu \neq ۳۲۰۰ \end{cases}$$



اگر میانگین نمونه در این ناحیه قرار بگیرد فرض صفر رد نمی‌شود

ناحیه قبول و رد: ناحیه رد یک آزمون ناحیه رد فرض H_0 می‌باشد و ناحیه قبول یک آزمون در واقع ناحیه قبول فرض H_0 است

مراحل عمومی آزمون فرض:

برای تمامی آزمون فرض‌های آماری، باید مراحل چهارگانه زیر را انجام داد

۱- تعریف فرضیه‌های آماری H_0 و H_1

۲- تعیین توزیع نمونه‌گیری آزمون و نوع آماره آزمون (نرمال یا T / μ و P یا دو جامعه)

۳- تعیین سطح زیر منحنی H_0 و H_1 و مقدار بحرانی

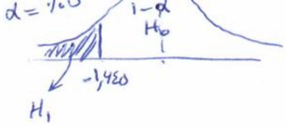
۴- تصمیم‌گیری.

۲۳/K

مثال: فرضیه‌ای به این صورت بیان شده است: میانگین نمره مسئولیت‌پذیری مدیران در کشور دست کم ۵۰ است.
محقق برای بررسی فرضیه فوق یک نمونه ۶۴ تایی از بین مدیران کشور به صورت تصادفی انتخاب کرده که میانگین و انحراف معیار آن به ترتیب ۴۵ و ۱۶ است. در سطح خطای ۵ درصد صحت فرضیه فوق را بررسی کنید.

$$\begin{cases} H_0: \mu_x \geq 50 \\ H_1: \mu_x < 50 \end{cases}$$

$$Z = \frac{\bar{x} - \mu}{S_{\bar{x}}} = \frac{45 - 50}{\frac{16}{\sqrt{64}}} = -2.5$$



$Z_\alpha = Z_{.05} = -1.645 \longrightarrow$ مقدار بحرانی

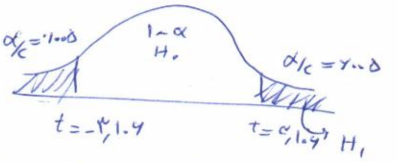
آماره آزمون در ناحیه H_1 قرار می‌گیرد.

تصمیم‌گیری: پس در سطح اطمینان ۹۵ درصد می‌توان گفت شواهد دلالت کافی بر تأیید H_0 ندارند.
در سطح خطای ۵ درصد فرض H_0 رد و فرض H_1 پذیرفته می‌شود.

مثال:
یک دانشجوی مدیریت فرضیه‌ای به این صورت طراحی کرده است: میانگین مهارت ادراکی مدیران سازمان A ۵۵ است.
به منظور بررسی فرضیه فوق، دانشجو از بین مدیران سازمان A یک نمونه تصادفی ۱۲ نفره انتخاب کرده که میانگین و انحراف معیار آن به ترتیب ۶۰ و ۱۵ است. فرض کنید توزیع امتیاز مهارت ادراکی مدیران سازمان A نرمال است. در سطح اطمینان ۹۹ درصد صحت فرضیه فوق را بررسی کنید.

$$\begin{cases} H_0: \mu_x = 55 \\ H_1: \mu_x \neq 55 \end{cases}$$

آماره آزمون $t = \frac{\bar{x} - \mu}{S_{\bar{x}}} = \frac{60 - 55}{\frac{15}{\sqrt{12}}} = 1.15$



مقادیر بحرانی: $t = \pm 3.106$

تصمیم‌گیری: چون آماره آزمون $t = 1.15$ در ناحیه H_0 قرار می‌گیرد پس فرض H_0 در سطح اطمینان ۹۹ درصد پذیرفته می‌شود.

۲۲/۸

مثال: نمونه‌ای تصادفی به اندازه $n=49$ از جامعه‌ای که دارای میانگین ۳۰ و انحراف معیار ۴ است انتخاب شده است

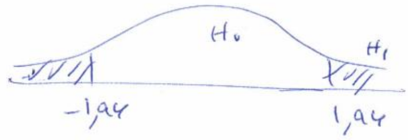
با استفاده از اطلاعات حاصل از این نمونه، آزمون آماری زیر را در سطح معنی‌دار بودن $\alpha = 0.05$ انجام دهید

$$\begin{cases} H_0: \mu = 32 \\ H_1: \mu \neq 32 \end{cases}$$

$$\alpha/2 = \frac{0.05}{2} \Rightarrow Z_{\alpha/2} = Z_{0.025} = 1.96$$

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{30 - 32}{\frac{4}{7}} = -5.25$$

$$Z_{\alpha/2} = \pm 1.96$$



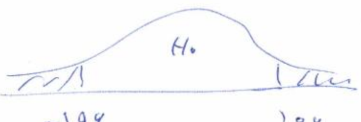
با اطمینان ۹۵ درصد فرض H_0 رد و فرض H_1 پذیرفته می‌شود.

مثال: یک سازنده داروهای دامی، قرص‌های هورمون رشد را برای گاوها در بسته‌های بزرگ تولید می‌کند. مشخصه‌های تولید ادعا می‌کنند که میانگین محتوای هورمونی قرص‌ها معادل $\mu = 1280$ میلی‌گرم برای هر قرص باشد. از هر بسته نمونه‌ای تصادفی مرکب از ۸۱ قرص انتخاب می‌کنند تا آزمون کنند که هر بسته دارای این مشخصه‌ها است یا خیر. از روی تجربه می‌دانند که انحراف معیار نمونه $110 = \sigma$ است. اگر میانگین نمونه‌ای ۱۲۹۰ میلی‌گرم باشد آزمون را در سطح اطمینان ۹۵ درصد انجام دهید

$$\begin{cases} H_0: \mu = 1280 \\ H_1: \mu \neq 1280 \end{cases}$$

$$\alpha = 0.05 \Rightarrow Z_{\alpha/2} = \pm 1.96$$

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{1290 - 1280}{\frac{110}{\sqrt{81}}} = \frac{10}{\frac{110}{9}} \Rightarrow Z = 0.81$$



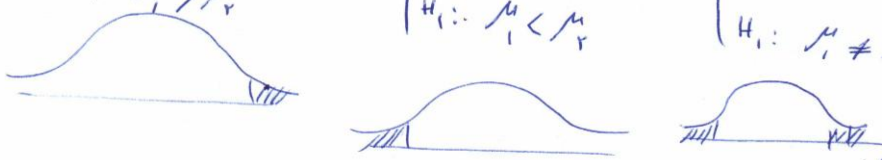
در سطح اطمینان ۹۵ درصد H_0 تائید و H_1 رد می‌شود.

آزمون مقایسه میانگین دو جامعه آماری

بخش اعظم فرضیه‌های پژوهشی در مدیریت و علوم رفتاری به منظور مقایسه دو جامعه آماری انجام می‌گیرد. این نوع فرضیه‌ها را فرضیه‌های تطبیقی گویند. برای آزمون این نوع فرضیه‌ها، در صورتی که میانگین نیاز باشد می‌توان از مراحل آزمون فرض آماری برای میانگین دو جامعه استفاده کرد.

۱- فرضیه‌ها

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases} \qquad \begin{cases} H_0: \mu_1 \geq \mu_2 \\ H_1: \mu_1 < \mu_2 \end{cases} \qquad \begin{cases} H_0: \mu_1 \leq \mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases}$$



۲- آماره آزمون:

اگر نمونه‌ها از دو جامعه نرمال با انحراف معیار معلوم انتخاب شوند:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

اگر نمونه‌ها از دو جامعه آماری نرمال با انحراف معیار نامعلوم انتخاب شوند، توزیع نمونه‌گیری به n_1+n_2-2 بستگی دارد:

الف) اگر درجه آزادی کمتر از ۳۰ باشد توزیع نمونه‌گیری آماره t استیودنت است

که اگر $\sigma_1=\sigma_2$ باشد آماره آزمون عبارتست از

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad , \quad S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

واریانس مشترک

اگر فرض $\sigma_1 \neq \sigma_2$ پذیرفته شود

$$d.f = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

ب) اگر درجه آزادی بزرگتر از ۳۰ باشد آماره آزمون به صورت زیر تعریف می‌شود

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

مثال: تحقیقی برای مقایسه دو روش A و B که مدیران استفاده می‌کنند انجام شده است. بدین منظور فرضیه‌ای به صورت زیر بیان شده است. روش A برای مدیران بهتر از روش B است. سؤال: مسئله این تحقیق مقایسه میانگین نمره‌های مدیرانی است که از روش A و B استفاده می‌کنند. چون به کل مدیران دسترسی نبوده است از هر گروه نمونه‌هایی تهیه شده است. بررسی‌ها نشان می‌دهد که توزیع نمره‌ها در هر دسته نرمال است. اطلاعات حاصل از نمونه‌ها به شرح زیر است. با فرض تساوی واریانس‌ها جامعه صحت فرضیه فوق را در سطح خطای یک درصد بررسی کنید.

A	B
$n_1 = 10$	$n_2 = 15$
$\bar{x}_1 = 52$	$\bar{x}_2 = 45$
$S_1 = 12$	$S_2 = 8$

$H_0: \mu_1 \le \mu_2$ انتظار اثربخشی روش A بهتر از روش B نیست

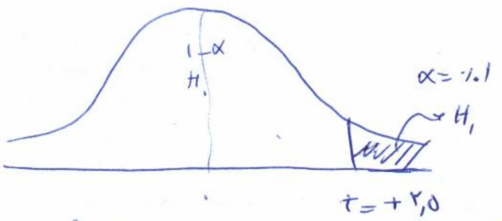
$H_1: \mu_1 > \mu_2$ انتظار اثربخشی روش A بهتر از روش B است

$$S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\frac{(10-1)(12)^2+(15-1)(8)^2}{10+15-2}} = 9.762$$

$$t = \frac{(52-45)-0}{9.762\sqrt{\frac{1}{10}+\frac{1}{15}}} = 1.757$$

مقادیر بحرانی:

$$t_{\alpha, df} = t_{0.01, 23} = \pm 2.5$$



آماره آزمون $t = 1.757$ در ناحیه H_0 قرار می‌گیرد بنابراین فرض H_0 در سطح خطای یک درصد پذیرفته می‌شود.

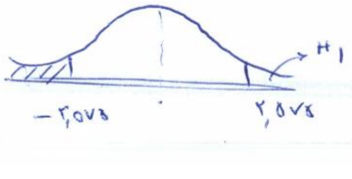
برای بررسی اینکه تفاوتی بین میانگین های دو گروه A و B وجود دارد یا خیر نتایج حاصل از نمونه های از این دو گروه به شرح زیر آمده است. فرضیه متقابل را در سطح خطای یک درصد بررسی کنید.

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \ne \mu_2 \end{cases}$$

A	B
$n_1 = 200$	$n_2 = 250$
$\bar{x}_1 = 15530$	$\bar{x}_2 = 14910$
$s_1 = 5170$	$s_2 = 5840$

$$z_{\alpha/2} = z_{0.005} = 2.575$$

$$z = \frac{\overbrace{(14910 - 15530)}^{620} - 0}{\underbrace{\sqrt{\frac{(5170)^2}{200}+\frac{(5840)^2}{250}}}_{519.5}} = 2.66$$



در سطح خطای یک درصد فرض H_0 رد و فرض H_1 پذیرفته می‌شود.