$M(t) = (pe^t + q)^n$

$M'(t) = npe^t(pe^t + q)^{n-1}$

$\mu = M'(0) = np$

$M''(t) = npe^t(pe^t + q)^{n-1} + (npe^t)^2(pe^t + q)^{n-1}$

$\sigma^2 = M''(0) - \mu^2 =$

مثال زیر چگونگی استفاده از توزیع تجمعی دو جمله‌ای را که برای n های مختلف $n = 1, 2, \ldots, 25$ و p های مختلف $(0.1, 0.15, \ldots, 0.9)$ نوشته شده است را نشان می‌دهد.

مثال: از ظرفی که ۸ مهرهٔ سفید و ۲ مهرهٔ سیاه دارد، ۲۰ مهره را به تصادف و با جایگذاری بیرون می‌آوریم. مطلوب است محاسبهٔ احتمال موارد زیر:

الف: دقیقاً ۱۸ مهره سفید باشد.

ب: حداقل ۱۵ مهره سفید باشد.

ج: حداکثر ۱۶ مهره سفید باشد

د: از ۱۰ تا ۱۸ مهره سفید باشد (با در نظر گرفتن خود ۱۰ و ۱۸)

این آزمایش، دو جمله‌ای است. اگر X را تعداد مهرهٔ سفید تعریف کنیم:

$b(x; 20, 0.8)$ ($n = 20$، $p = 0.8$)

$P_X(x) = \binom{20}{x}(0.8)^x(0.2)^{20-x} \qquad x = 0, 1, \ldots, 20$

الف: $P(X=18) = f_X(18) = \binom{20}{18}(0.8)^{18}(0.2)^2 =$ که مشکل است

$P(X=18) = \sum_{x=0}^{18} b(x; 20, 0.8) - \sum_{x=0}^{17} b(x; 20, 0.8)$

$= F_X(18) - F_X(17) = 0.9308 - 0.7940$

ب: $P(X \geq 15) = 1 - P(X \leq 14) = 1 - F_X(14) = 1 - 0.1958$

ج: $P(X \leq 16) = F(16) = 0.5886$

د: $P(10 \leq X \leq 18) = F(18) - F(9) = 0.9308 - 0.0006$

Subject:
Year: Month: Day:

n		p: ۰/۱	...	۰/۸
۲				
⋮	⋮			
۲۰	۵		X	
⋮	⋮			
	۲۰			

$F(r) = \sum_{x=0}^{r} b(x; n, p)$

توزیع فوق هندسی

آزمایش فوق هندسی: به یک آزمایش تصادفی که دارای ۲ ویژگی زیر باشد، آزمایش فوق هندسی

می‌گوییم:

۱- یک نمونه تصادفی n تایی بدون جایگذاری از N کالا انتخاب شود.

۲- از N کالای موجود، k کالا موفقیت و N-k کالا شکست تلقی شود.

مثال: از ظرفی که ۸ مهره سفید و ۲ مهره سیاه دارد، سه مهره را همزمان انتخاب

می‌کنیم. یا، خرید ۵ دستگاه تلویزیون از فروشگاهی که ۱۰ دستگاه تلویزیون سالم و

Subject:
Year: Month: Day:

۴ دستگاه تلویزیون خراب دارد.

متغیر تصادفی فوق هندسی: متغیر تصادفی X که شمارنده تعداد موفقیت در آزمایش فوق هندسی

باشد را متغیر فوق هندسی می‌گویند.

توزیع احتمال فوق هندسی: توزیع متغیر فوق هندسی را توزیع فوق هندسی گویند و این تابع احتمال

بصورت زیر است:

$$f_X(x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

قضیه: میانگین و واریانس توزیع فوق هندسی برابر است با

$$\mu = n\cdot\frac{k}{N} \qquad \sigma^2 = \left(\frac{N-n}{N-1}\right) n\cdot\frac{k}{N}\left(1-\frac{k}{N}\right)$$

نتیجه: اگر N به اندازه کافی بزرگ باشد و n کوچک باشد،

$$\frac{k}{N} = p \qquad 0 < p < 1$$

$$\mu = np \qquad \sigma^2 = npq$$

اگر تعداد کالا زیاد و تعداد نمونه کم باشد، نمونه گیری بدون جایگذاری با نمونه گیری با جایگذاری برابر است. در کل در حالت نمونه گیری بدون جایگذاری است و با توجه به زیاد بودن کالا از توزیع دوجمله ای استفاده می کنند.

توزیع دوجمله ای منفی

آزمایش دوجمله ای منفی: اگر یک آزمایش برنولی را آنقدر ادامه دهیم تا r موفقیت رخ دهد و سپس آزمایش را قطع کنیم، چنین آزمایشی را دوجمله ای منفی می گویند.

مثال: یک تاس را آنقدر پرتاب می کنیم تا سومین (۳امین) عدد ۶ مشاهده شود.

متغیر تصادفی دوجمله ای منفی: متغیر تصادفی X که نشان دهنده تعداد آزمایش تا رسیدن به r امین موفقیت است را متغیر دوجمله ای منفی می گویند.

توزیع این متغیر را توزیع دوجمله ای منفی نامیده و با علامت زیر نشان داده و فرمول آن به قرار زیر است

$$f_X(x)=b^*(x;r,p)=\binom{x-1}{r-1}p^r q^{x-r}$$

X: تعداد آزمایش $\quad x=r, r+1, \ldots$

قضیه: میانگین و واریانس توزیع دوجمله ای منفی برابر است با

$$\mu=\frac{r}{p} \qquad \sigma^2=\frac{rq}{p^2}$$

اثبات: به کمک تابع مولد گشتاور، ابتدا تابع مولد گشتاور را محاسبه می کنیم

$$M(t)=E(e^{tx})=\sum_{x=0} e^{tx}\binom{x-1}{r-1}p^r q^{x-r}$$

$$=p^r e^{rt}(1-qe^t)^{-r} \qquad t<e.q$$

حالت خاص:

اگر $r=1$ باشد، یعنی آزمایش را آنقدر ادامه دهیم تا اولین موفقیت بدست آید، سپس آزمایش را قطع می کنیم. این آزمایش را آزمایش هندسی می گوییم.

$$b^*(x;1,p)=g(x;p)=pq^{x-1} \qquad (x=1,2,\ldots)$$

مثال: تاسی را مستقلاً پرتاب می کنیم. مطلوب است احتمال آنکه سومین عدد ۶ در دهمین پرتاب مشاهده شود.

$r=3 \quad p=\frac{1}{6}$ پرتاب منجر به موفقیت

$x=10$ تعداد آزمایش

$$\Rightarrow b^*(10;3,\tfrac{1}{6})=\binom{10-1}{3-1}\left(\tfrac{1}{6}\right)^3\left(\tfrac{5}{6}\right)^{10-3}$$

توزیع پواسن:

آزمایش پواسن: اگر نتیجه یک آزمایش تصادفی در یک فاصله زمانی یا ناحیه‌ای معین رخ دهد چنین آزمایشی را پواسن گویند. (فاصله زمانی: ... ثانیه، دقیقه، ساعت، روز ...) (ناحیه: طول، سطح، حجم ...)

مثال: تعداد تصادف در طول روز (ماه، سال و ...) در یک شهر. یا تعداد دفعاتی که تلفن به صدا در می‌آید. یا تعداد اشتباهاتی که یک تایپیست در یک صفحه مرتکب می‌شود.

در تمام آزمایش‌های پواسن λ = متوسط تعداد موفقیت، همواره مقداری معلوم است.

متغیر پواسن: متغیر تصادفی X که نشان‌دهنده تعداد موفقیت در آزمایش پواسن باشد را متغیر پواسن می‌گویند.

$X: x = 0, 1, 2, 3, \ldots$

توزیع پواسن: توزیع متغیر پواسن را توزیع پواسن می‌گویند و با علامت $f_X(x), P(x;\lambda)$ نشان داده و فرمول آن به صورت مقابل است.

$$f_X(x) = P(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

قضیه: میانگین و واریانس توزیع پواسن برابر است با:

$$\mu = \lambda \quad , \quad \sigma^2 = \lambda$$
$$\Downarrow$$
$$\sigma = \lambda$$

اثبات: (به کمک تابع مولد گشتاور)

قبل از اثبات می‌خواهیم ببینیم که فرمول پواسن آیا تابع احتمال هست یا خیر.

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda}\cdot\lambda^x}{x!} \stackrel{?}{=} 1$$

$$e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{\lambda} \qquad \left(e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}\right) \text{ بسط تیلور}$$

$$e^{-\lambda}\cdot e^{\lambda} = 1 \ \checkmark$$ تابع احتمال هست.

ادامه اثبات:

$$M(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} f_X(x) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\cdot\lambda^x}{x!}$$

$$e^{-\lambda}\sum\frac{(\lambda e^t)^x}{x!} = e^{\lambda e^t} \quad = e^{-\lambda}\cdot e^{\lambda e^t}$$

$$M(t) = e^{\lambda(e^t-1)}$$

هر تابع احتمالی که تابع مولد گشتاورش به شکل مقابل باشد، مربوط به توزیع پواسن است.

Subject:
Year: Month: Day:

$\mu = M'(0) = \lambda e^t . e^{\lambda(e^t-1)} \Big|_{t=0} \rightarrow \boxed{\mu = \lambda}$

قضیهٔ زیر نشان می‌دهد که توزیع پواسن حالت خاصی از توزیع دو جمله‌ای است. همچنانکه

$n \to \infty$ و $p \to 0$ (یا $p \to 1$)

$q \to 1$ $q \to 0$

قضیه - تقریب دو جمله‌ای به کمک پواسن

اگر X یک متغیر تصادفی دو جمله‌ای با توزیع $b(x; n, p)$ باشد، آنگاه شکل حدی توزیع این متغیر هنگامیکه $n \to +\infty$ و $p \to 1$ یک توزیع پواسن با پارامتر $\lambda = np$ خواهد شد.

به عبارت دیگر:

$$\lim_{\substack{n \to \infty \\ p \to 1}} \binom{n}{x} p^x (1-p)^{n-x} = \frac{e^{-np} \cdot (np)^x}{x!}$$

⚠ اگر در توزیع دو جمله‌ای $n \geq 25$ باشد مشروط بر آنکه $p \to 0$ یا $p \to 1$ باشد، می‌توان از توزیع پواسن استفاده کرد.

تذکر: جدول توزیع تجمعی این متغیر برای مقادیر مختلف λ (از ۰/۱ تا ۱۸) نوشته شده که

SAMAND

Subject:
Year: Month: Day:

در محاسبات از آن استفاده می‌شود.

مثال: گزارش شده است که به طور متوسط ۶ تصادف در طول روز در یک شهر رخ می‌دهد. مطلوبست محاسبهٔ احتمال آنکه:

الف: در یک روز معین هیچ تصادفی رخ ندهد.

ب: حداقل ۲ تصادف رخ دهد.

ج: حداکثر ۱۰ تصادف رخ دهد.

حل: این آزمایش یک آزمایش پواسن است. اگر x را تعداد تصادف تعیین کنیم،

$p(x,6) = \frac{e^{-6} \times 6^x}{x!}$

الف: $p(X=0) = p(0,6) = \frac{e^{-6} \times 1}{1} = e^{-6} = \sum_{x=0}^{0} p(x,6) = F_X(0) = 0.0025$ (جدول تجمعی)

ب: $p(X \geq 2) = 1 - p(X \leq 1) = 1 - F_X(1) = 1 - 0.0174$

ج: $p(X \leq 10) = F_X(10) = 0.9332$

SAMAND

بررسی چند توزیع پیوسته

توزیع گاما:

تابع گامای α:

تابع گامای α که با علامت $\Gamma(\alpha)$ نشان داده می‌شود به صورت زیر تعریف می‌شود.

$$\Gamma(\alpha)=\int_0^{+\infty} x^{\alpha-1} e^{-x}\,dx$$

$$\alpha=1 \Rightarrow \Gamma(1)=\int_0^{+\infty} e^{-x}\,dx=1 \Rightarrow \Gamma(1)=1$$

$$\alpha=\frac{1}{2} \Rightarrow \Gamma(1/2)=\int_0^{+\infty} x^{-1/2} e^{-x}\,dx \overset{?}{=} \sqrt{\pi}$$

$$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$$

$$\Gamma(\alpha+1)=\alpha!$$

توزیع گاما:

متغیر تصادفی X که دارای تابع احتمال پیوسته به صورت زیر است،

$$f_X(x)=\begin{cases} \dfrac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & x\geq 0 \\ 0 & x<0 \end{cases}$$

را متغیر گاما و توزیع این متغیر را توزیع گاما با پارامترهای α و β می‌گویند.

تذکر: میانگین و واریانس توزیع گاما برابر است با:

$$\mu=\alpha\beta \quad , \quad \sigma^2=\alpha\beta^2$$

مثال/ متغیر تصادفی X دارای توزیع نرمال با میانگین معلوم μ و واریانس معلوم σ^2 است

به کمک جدول احتمال های زیر را محاسبه کنید

$$P(\mu-\sigma < X < \mu+\sigma) = ?$$

$$= P\left(-1 < \frac{X-\mu}{\sigma} < 1\right) = P(-1 < Z < 1) = F_Z(1) - F_Z(-1) = ٠/۶٨٢۶$$

$$P(\mu-2\sigma < X < \mu+2\sigma) = P\left(-2 < \frac{X-\mu}{\sigma} < 2\right) = F_Z(2) - F_Z(-2) = ٠/٩۵۴۴$$

$$P(\mu-3\sigma < X < \mu+3\sigma) = P\left(-3 < \frac{X-\mu}{\sigma} < 3\right) = F_Z(3) - F_Z(-3) = ٠/٩٩٧۴$$

نتیجه گیری:

در یک توزیع نرمال ۶٨/٢٧ ٪ از مشاهدات جامعه در بازه $(\mu-\sigma , \mu+\sigma)$ قرار دارند

قضیه زیربنایی دموآور لاپلاس: تحت شرایطی، توزیع گسسته دوجمله‌ای به توزیع پیوسته میل می‌کند.

قضیه - تقریب دوجمله‌ای به کمک نرمال

اگر X یک متغیر تصادفی دوجمله‌ای با توزیع $b(x; n, p)$ باشد، آن‌گاه شکل حدی توزیع متغیر تصادفی $Z = \frac{X-np}{\sqrt{npq}}$ یک توزیع نرمال استاندارد خواهد بود هنگامی که $n \to +\infty$

(n به قدر کافی بزرگ باشد)

از این قضیه هنگامی که $n \geq ۳۰$ باشد می‌توان استفاده کرد.

مثال: متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامترهای $n=۱۲$ و $p=\frac{۱}{۲}$ است. (توزیع متغیر X، توزیع دوجمله‌ای $b(x; ۱۲, \frac{۱}{۲})$ است) مطلوب است محاسبه احتمالات زیر به دو روش

۱- به کمک جدول ۲- استفاده از تقریب نرمال

الف/ $P(۴ \leq X \leq ۷) = ?$ ب/ $P(X=۵) = ?$

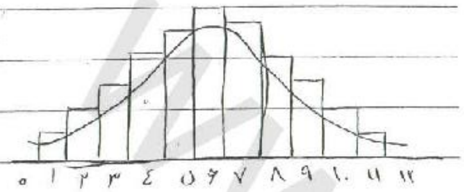
$b(x, ۱۲, \frac{۱}{۲}) = \binom{۱۲}{x}\left(\frac{۱}{۲}\right)^{x}\left(\frac{۱}{۲}\right)^{۱۲-x}$

الف) $P(۴ \leq X \leq ۷) = F_X(۷) - F_X(۴) = ۰/۷۳۳$

ب) $P(X=۵) = F_X(۵) - F_X(۴) = ۰/۱۹۳۰$

$\mu = np = ۱۲ \times \frac{۱}{۲} = ۶$

$\sigma = \sqrt{۱۲ \times \frac{۱}{۲} \times \frac{۱}{۲}} = \sqrt{۳}$



$n(x, ۶, \sqrt{۳})$

الف) $P(۴ \leq X \leq ۷) \simeq P(۳/۵ < X < ۷/۵) = P\left(\frac{۳/۵-۶}{\sqrt{۳}} < Z < \frac{۷/۵-۶}{\sqrt{۳}}\right)$

$= ۰/۷۳۴۲$

ب) $P(X=۵) \simeq P(۴/۵ < X < ۵/۵) = P\left(\frac{۴/۵-۶}{\sqrt{۳}} < Z < \frac{۵/۵-۶}{\sqrt{۳}}\right)$

$= ۰/۱۹۳۷$

۴۷

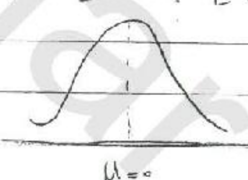
توزیع t

این توزیع پرکاربرد است و بصورت تئوری از قضیه زیر نتیجه میشود.

اگر متغیر تصادفی V دارای توزیع توان دوم کای با درجه آزادی ν و متغیر تصادفی Z دارای توزیع نرمال استاندارد باشد و V و Z مستقل باشند، آنگاه توزیع متغیر تصادفی $T = \frac{Z}{\sqrt{\frac{V}{\nu}}}$ به توزیع t معروف است. (با درجه آزادی ν)

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)} \qquad t \in \mathbb{R}$$

اگر $f(t)$ را به ازای یک مقدار دلخواه ν ($\nu \geq 30$) رسم کنیم بصورت زیر خواهد بود.



ضمناً ثابت میشود که $\mu=0$ و $\sigma^2 = \frac{\nu}{\nu-2}$

$(\nu \geq 30 \Rightarrow f_T(t) = f_Z(z))$

یک بازه اطمینان ۱۰۰(۱-α)٪ برای μ عبارت است از:

$$\bar{x} - z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

از فرمول فوق در سه صورت می‌توان استفاده کرد:

۱) $n \geq 30$ ، جامعه دلخواه (تابع احتمال جامعه نامشخص) ، σ معلوم

۲) $n \geq 30$ ، " " " ، σ نامعلوم (استفاده از s به جای σ)

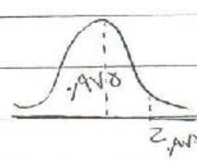
۳) $n < 30$ ، جامعه نرمال ، σ معلوم

مثال: یک نمونه تصادفی ۶۴ تایی از جامعه‌ای انتخاب و مشخص شده میانگین این نمونه $\bar{x}_{64} = 76$ و واریانس این نمونه $s^2 = 9$ می‌باشد. یک فاصله اطمینان ۹۵٪ برای μ، میانگین واقعی این جامعه بر اساس نمونه بدست آورید.

$$\bar{x} - z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

$z_{1-\frac{\alpha}{2}} = z_{0.975} = ?$



$P(Z < z) = 0.975$

$z = 1.96$

صفحهٔ مقابل به اشتباه در این قسمت نوشته شده است.

در حالیکه محل صحیح آن صفحهٔ دارای * است.

$$74 - 1.96\left(\frac{3}{\sqrt{64}}\right) < \mu < 74 + 1.96\left(\frac{3}{\sqrt{64}}\right)$$

$a < \mu < b \Rightarrow \mu \in (a,b)$ با ۹۵٪ اطمینان

محاسبه خطا در برآورد μ بوسیله $\bar{x}$

می‌دانیم یک بازه اطمینان $100(1-\alpha)\%$ برای μ عبارت است از

$$\bar{x} - z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \xrightarrow{+(-\bar{x})}$$

$$-z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu - \bar{x} < z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\underbrace{\boxed{|\mu - \bar{x}|}}_{\text{خطا}} < \underbrace{z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)}$$ حداکثر خطا

در برآورد μ بوسیله $\bar{x}$ با $100(1-\alpha)\%$ اطمینان می‌توان گفت حداکثر اندازه خطا $e = z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$ است.

$$e = z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow \sqrt{n} = \frac{z_{1-\frac{\alpha}{2}}(\sigma)}{e} \Rightarrow$$

$$n = \underbrace{\left[\left(\frac{z_{1-\frac{\alpha}{2}}(\sigma)}{e}\right)^2\right]}_{\text{جزء صحیح}}$$

جامعه‌ها و نمونه‌ها (Population & Samples)

تعریف (جامعه یا جمعیت): به مجموعه مشاهداتی که پس از آزمایش تصادفی گردآوری شده و مورد تجزیه و تحلیل قرار می‌گیرد، جامعه یا جمعیت می‌گویند.

اندازه یا حجم جامعه: به تعداد عضوهای مجموعه جامعه، اندازه جامعه یا حجم جامعه می‌گویند و با N نشان می‌دهند.

مثال: فرض کنید می‌خواهیم گروه خونی ۱۰۰ دانشجو را مورد بررسی قرار دهیم. این عمل یک آزمایش تصادفی است. پس از این آزمایش، ۱۰۰ عدد خواهیم داشت که این اعداد مقادیری از یک متغیر تصادفی مانند X می‌باشند که X، هشت مقدار را اختیار می‌کند.

X: نشان‌دهنده گروه خونی هر نفر $\quad X: x = 1, 2, \ldots, 8 \quad (1 \to O^+ \quad 2 \to AB^- \ldots$

$1, 2, 8, 7, 2, 6, 4, 5, 1, 2, \ldots$

جامعه مجموعه‌ای از مشاهدات است ← مشاهدات مقادیری از یک متغیر تصادفی مانند X است.

هر متغیری مانند X دارای یک تابع احتمال $f_X(x)$ است ← هر تابع احتمال دارای μ و σ^2 است ← پس هر جامعه دارای μ و σ^2 است.

هر جامعه‌ای دارای یک اسم است. اسم جامعه همان اسم توزیع احتمال است.

جامعه: ۱- متناهی (در عمل کوچک)

۲- نامتناهی (در عمل بزرگ)

نامعلوم؟ بزرگ ← به جز در مشاهدات دسترسی نیست ← به متغیر X دسترسی نیست ← به تابع احتمال $f_X(x)$ دسترسی نیست ← به μ و σ^2 دسترسی نیست

در جامعه بزرگ به μ و σ^2 پارامترهای جامعه می‌گویند.

نمونه چیست؟ هر زیرمجموعه از یک جامعه را نمونه می‌گویند. نمونه هم از مشاهدات تشکیل شده است.

نمونه تصادفی: اگر $X_1, X_2, \dots, X_n$، n متغیر تصادفی مستقل با توزیع یکسان

$f_X(x)$ باشند، آنگاه این n متغیر را یک نمونه تصادفی از جامعه‌ای با تابع احتمال $f_X(x)$ می‌گوییم

واضح است که $\mu_{X_1}=\mu_{X_2}=\dots=\mu_{X_n}=\mu_X=\mu_{\text{جامعه}}$ و

$\sigma^2_{X_1}=\sigma^2_{X_2}=\dots=\sigma^2_{X_n}=\sigma^2_X=\sigma^2_{\text{جامعه}}$

$f_{X_1,X_2,\dots,X_n}(x_1,x_2,\dots,x_n)=f_X(x_1)f_X(x_2)\dots f_X(x_n)$

بصورت کاربردی نمونه تصادفی را می‌توان به شکل زیر نیز تعریف کرد

نمونه تصادفی: یک نمونه تصادفی با اندازه n، نمونه‌ای است که همه نمونه‌های دیگر همان شانس انتخاب شدن را داشته باشند.

نمونه تصادفی (مشاهدات) ← یک متغیر تصادفی تعریف می‌کنیم به نام آماره ← تابع احتمال آماره ← میانگین و واریانس (نمونه)

آماره:

هر تابعی از یک نمونه تصادفی n تایی، که فاقد پارامتر مجهول باشد را آماره یا آمار می‌گویند.

به عبارت دیگر اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی n تایی از جامعه‌ای با توزیع احتمال $f_X(x)$ باشد، هر تابعی مانند $g(x)$ از متغیرهای $X_1, \ldots, X_n$ را آماره می‌گوئیم. مثلاً:

$g(x) = X_1 + X_n$ $\quad$ $g(x) = \frac{\sum x_i}{n}$

$k(x) = \frac{\sum (x_i - \mu)^2}{n}$ $\Longrightarrow$ چون μ مجهول است، پس آماره نیست.

توجه: از یک نمونه تصادفی، می‌توان آماره‌های ($g(x)$) متوالی ساخت. با توجه به تعریف آماره‌ها متغیر تصادفی می‌باشند.

آماره‌های مهم عبارتند از:

۱- آماره میانگین ۲- آماره میانه ۳- آماره مد ۴- آماره برد

۵- آماره واریانس (S^2) ۶- آماره واریانس (S'^2)

آماره میانگین (Mean)

آماره میانگین را با علامت $\bar{X}_n$ نشان داده و برای نمونه‌ای تصادفی به صورت $X_1, \ldots, X_n$ به صورت زیر تعریف می‌شود:

$$\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$$

آماره میانه (Median)

میانه را با علامت $\tilde{X}$ (ایکس تیلدا) نشان می‌دهند و برای نمونه تصادفی n تایی که به ترتیب صعودی $X_1 \le X_2 \le \cdots \le X_n$ مرتب شده باشند برابر است با مقدار میانی مشاهدات اگر n فرد باشد و نصف مجموع دو مقدار میانی اگر n زوج باشد.

آماره مد - نما (Mode)

آماره مد را با M نشان داده و مقدار این آماره برای یک نمونه تصادفی n تایی با مشاهدات $X_1, X_2, \ldots, X_n$ برابر است با مشاهداتی که بیشترین فراوانی را دارند.

از این سه آماره مخصوصاً $\bar{X}_n$ در قسمت‌های بعدی برای محاسبه μ میانگین جامعه استفاده خواهیم کرد.

آماره برد: Range

آماره برد را با R نشان داده و مقدار این آماره برای نمونه‌ای تصادفی با اندازه n که به ترتیب صعودی مرتب شده باشند $x_1 \le x_2 \le \dots \le x_n$ برابر $x_n - x_1$

آماره واریانس S^2

این آماره را با S^2 نشان داده و برای نمونه تصادفی به صورت $x_1, x_2, \dots, x_n$ برابر

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

است.

آماره واریانس S'^2

این آماره به صورت زیر تعریف می‌شود:

$$S'^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

تذکر: با توجه به تعریف S^2 و S'^2 اگر n به اندازه کافی بزرگ باشد ($n \ge 30$)، S^2 با S'^2 یکسان می‌شود. اما اگر $n < 30$ باشد نشان می‌دهیم که S^2 آماره بهتری از S'^2 است.

از این سه آماره (R, S^2, S'^2) در قسمت بعد برای برآورد σ^2 واریانس جامعه استفاده خواهیم کرد.

قضیه: اگر $x_1, x_2, \dots, x_n$ یک نمونه تصادفی از جامعه‌ای با میانگین μ و واریانس σ^2 باشد، آنگاه:

$$\mu_{\bar{X}_n} = E(\bar{X}_n) = \mu_{\text{جامعه}}$$

$$\sigma^2_{\bar{X}_n} = \frac{\sigma^2_{\text{جامعه}}}{n}$$

اثبات:

$$E(\bar{X}_n) = E\left(\frac{\sum x_i}{n}\right) = E\left(\frac{1}{n}(x_1 + \dots + x_n)\right)$$

$$= \frac{1}{n} E(x_1 + \dots + x_n)$$

$$= \frac{n\mu}{n} = \mu$$

$$\sigma^2_{\bar{X}_n} = \sigma^2_{\frac{1}{n}\sum x_i} = \frac{1}{n^2}\sigma^2_{x_1 + x_2 + \dots + x_n} = \frac{1}{n^2}\left(\sigma^2_{x_1} + \dots + \sigma^2_{x_n}\right)$$

$$= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

نتیجه: اگر S^2 واریانس یک نمونه تصادفی از جامعه‌ای با میانگین μ و واریانس σ^2 باشد آنگاه

$E(S^2) = \sigma^2$ $\qquad$ $E(S'^2) \neq \sigma^2$ $\qquad$ $E(S'^2) = \frac{n-1}{n}\sigma^2$

در برآورد μ به وسیله $\bar{x}$ با $100(1-\alpha)\%$ اطمینان می‌توان گفت که خطا

کمتر از مقدار e است اگر تعداد نمونه $n=\left[\left(\frac{z_{1-\frac{\alpha}{2}}\,\sigma}{e}\right)^2\right]$ اختیار کنیم.

محاسبه فاصله اطمینان برای μ نامعلوم $n<30$، جامعه نرمال، σ نامعلوم

می‌دانیم در این شرایط اگر به جای σ از S استفاده کنیم توزیع آماره $T=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ (1)

توزیع t با درجه آزادی $n-1$ دارا است.

$$P\left(-t_{1-\frac{\alpha}{2}}<T<t_{1-\frac{\alpha}{2}}\right)=1-\alpha \quad (2)$$

$$(1),(2) \Rightarrow P\left(-t_{1-\frac{\alpha}{2}}<\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}<t_{1-\frac{\alpha}{2}}\right)=1-\alpha \quad (3)$$

از تساوی ۳ نتیجه زیر بدست می‌آید.

$$P\left(\bar{x}-t_{1-\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)<\mu<\bar{x}+t_{1-\frac{\alpha}{2}}\left(\frac{S}{n}\right)\right)=1-\alpha \quad (4)$$

از ۴ نتیجه زیر بدست می‌آید.

نتیجه و یک فاصله اطمینان $100(1-\alpha)\%$ برای μ (در شرایطی که $n<30$، جامعه نرمال،

σ نامعلوم) عبارت است از:

$$\bar{x}-t_{1-\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)<\mu<\bar{x}+t_{1-\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)$$

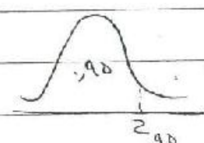
مثال: یک بازه اطمینان ۹۰٪ برای μ جامعه نرمالی بدست آورید در صورتی که بدانیم

$\sigma^2=4$ و میانگین یک نمونه ۲۵ تایی از این جامعه $\bar{x}=20$ گزارش شده است.

$$\bar{x}-z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)<\mu<\bar{x}+z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$1-\alpha=0.9 \Rightarrow \alpha=0.1 \Rightarrow \frac{\alpha}{2}=0.05$$

$$z_{1-0.05}=z_{0.95}=?$$



$$P(Z<z)=0.95$$

$$z_{0.95}=1.645$$

$$20-\left(1.645\times\frac{2}{\sqrt{25}}\right)<\mu<20+\left(1.645\times\frac{2}{\sqrt{25}}\right)$$

حداکثر خطا: $e=\frac{1.645\times 2}{\sqrt{25}}$ (کران بالای خطا)

Subject:
Year: Month: Day:

مثال: از جامعه‌ی نرمالی یک نمونه تصادفی انتخاب و نتایج زیر بدست آمده است.

$x_1=3 \,,\, x_2=1 \,,\, x_3=3 \,,\, x_4=4 \,,\, x_5=4$

$3,1,3,4,4 \Rightarrow n=5$

یک فاصله‌ی اطمینان 0.99 برای μ میانگین واقعی این جامعه بدست آورید.

$x_1,x_2,\dots,x_5 \to \bar{x}_5=3$

$\to S^2=\frac{3}{2} \Rightarrow S=\sqrt{\frac{3}{2}}$

$1-\alpha=0.99 \Rightarrow \alpha=0.01 \Rightarrow \frac{\alpha}{2}=0.005$

$t_{1-0.005}=t_{0.995}=$ از جدول $=4.6$

$\nu=4$

$\bar{x}-t_{1-\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)<\mu<\bar{x}+t_{1-\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)$

$3-\left(4.6\times\frac{\sqrt{3/2}}{\sqrt{5}}\right)<\mu<3+\left(4.6\times\frac{\sqrt{3/2}}{\sqrt{5}}\right)$

طول بازه اطمینان: $L=b-a=\left(\bar{x}_1+Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)-\left(\bar{x}_1-Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

$L=2Z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)=2e \qquad e\to 0 \Rightarrow \ell\to 0$

SAMAND

Subject:
Year: Month: Day:

یک بازه اطمینان چه ویژگی‌ای باید داشته باشد؟

۱- ضریب اطمینان بالا داشته باشد

۲- طول بازه کم باشد.

تعیین فاصله‌ی اطمینان برای پارامتر $\theta=\sigma^2$ و $\theta=\sigma$

می‌دانیم یک برآوردگر خوب برای σ^2، برآوردگر S^2 است. بنابراین یک مقدار برای این برآوردگر هنگامی که نمونه‌ی تصادفی به صورت $x_1,x_2,\dots,x_n$ اختیار شود و مقدار S_n^2 که محاسبه گردد، برآورد نقطه‌ای خوبی برای σ^2 است.

از طرفی می‌دانیم (تحت شرایطی): $X=\frac{(n-1)S^2}{\sigma^2}$ ① توزیع کای دو با درجه آزادی $n-1$ دارد و نیز می‌دانیم:

$P\left(\chi^2_{\frac{\alpha}{2}}<X<\chi^2_{1-\frac{\alpha}{2}}\right)=1-\alpha$ ②

①,② $\Rightarrow P\left(\chi^2_{\frac{\alpha}{2}}<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{1-\frac{\alpha}{2}}\right)=1-\alpha$

$P\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}<\sigma^2<\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}\right)=1-\alpha$ ③

SAMAND

از ۳ رو نتیجه می‌گیریم:

۱- یک بازه اطمینان $100(1-\alpha)\%$ برای σ^2 واریانس واقعی جامعه (که از یک نمونه تصادفی

با اندازه n ، $n<30$ از جامعه‌ای با توزیع نرمال اختیار شود) عبارت است با:

$$\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}$$

۲- یک بازه اطمینان $100(1-\alpha)\%$ برای σ انحراف معیار عبارت است از:

$$S\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}}} < \sigma < S\sqrt{\frac{n-1}{\chi^2_{\alpha/2}}}$$

مثال: در مثال قبلی یک بازه اطمینان ۹۹٪ برای σ^2 و σ (واریانس و انحراف معیار واقعی جامعه) بدست آورید.

$$S^2 = \frac{3}{2} \Rightarrow S = \sqrt{\frac{3}{2}}$$

x_i	x_i^2
3	9
1	1
3	9
4	16
4	16
$\sum x_i = 15$	$\sum x_i^2 = 51$

$$S^2 = \frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)}$$

$$S^2 = \frac{5\times 51 - (15)^2}{5\times 4} = \frac{255-225}{20} = \frac{3}{2}$$

$1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005$

$\chi^2_{0.005} =$ از جدول $= 0.207$ $\therefore$ $\chi^2_{0.995} =$ از جدول $= 14.9$

$\nu = 4$

$$\frac{4\times\frac{3}{2}}{14.9} < \sigma^2 < \frac{4\times\frac{3}{2}}{0.207}$$

تعیین کرانه نقطه‌ای برآورد و حدود اطمینان برای پارامتر $\theta = \mu_1-\mu_2$ (تفاضل میانگین دو جامعه)

می‌دانیم بهترین برآوردگر خوب برای پارامتر $\mu_1-\mu_2$ ، برآوردگر $\bar{X}_1-\bar{X}_2$ است.

$$E(\bar{X}_1-\bar{X}_2) = E(\bar{X}_1)-E(\bar{X}_2) = \mu_1-\mu_2$$

این برآوردگر نااریب است. (امید ریاضی = پارامتر جامعه)

در نتیجه یک مقدار از این برآوردگر هنگامیکه دو نمونه تصادفی مستقل با اندازه‌های n_1 و n_2 از این جامعه

انتخاب و $\bar{x}_1-\bar{x}_2$ آن محاسبه شود، برآورد نقطه‌ای خوبی برای $\mu_1-\mu_2$ است (مقدار $\bar{x}_1-\bar{x}_2$

تخمین نقطه‌ای عددی برای $\mu_1-\mu_2$ است)

می‌دانیم تحت شرایطی توزیع برآوردگر (آماره) $\bar{X}_1-\bar{X}_2$ توزیع نرمال با میانگین $\mu_1-\mu_2$ و

Subject:
YEAR: MONTH: DAY:

واریانس $\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$ است. به عبارت دیگر:

$$Z = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \quad \underline{1}$$

از طرف دیگر می‌دانیم $P(-Z_{1-\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}}) = 1-\alpha$ $\underline{2}$

$$\underline{1},\underline{2} \Rightarrow P\left(-Z_{1-\frac{\alpha}{2}} < \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

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

$$= 1-\alpha \quad \underline{3}$$

از فرمول ۳ نتیجه زیر حاصل می‌شود:

یک بازه اطمینان $100(1-\alpha)\%$ برای پارامتر $\mu_1-\mu_2$ (تفاضل میانگین دو جامعه) عبارت است از:

$$(\bar{x}_1-\bar{x}_2) - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

۳۲

SAMAND

Subject:
YEAR: MONTH: DAY:

از این فرمول در سه حالت می‌توان استفاده کرد. این سه حالت همان شرایطی است که $\bar{X}_1-\bar{X}_2$ دارای توزیع نرمال می‌شود.

۱) $n_1,n_2 \geq 30$ ، دو جامعه دلخواه مستقل (یا دو نمونه مستقل): σ_1^2 و σ_2^2 معلوم

۲) $n_1,n_2 \geq 30$ ، " " " " ، σ_1^2 و σ_2^2 نامعلوم

به جای σ_1^2 و σ_2^2 از S_1^2 و S_2^2 در فرمول استفاده شود.

۳) $n_1,n_2 < 30$ ، دو جامعه نرمال مستقل یا دو نمونه مستقل، σ_1^2 و σ_2^2 حتماً باید معلوم باشند.

تعیین برآورد نقطه‌ای و برآورد فاصله اطمینان برای پارامتر $\theta=\frac{\sigma_1^2}{\sigma_2^2}$ و $\theta=\frac{\sigma_1}{\sigma_2}$

می‌دانیم یک برآوردگر خوب برای $\frac{\sigma_1^2}{\sigma_2^2}$ ، برآوردگر $\frac{S_1^2}{S_2^2}$ است. بنابراین یک مقدار از این برآوردگر هنگامی که دو نمونه تصادفی مستقل با اندازه‌های n_1 و n_2 از دو جامعه نرمال اختیار و مقدار $\frac{S_1^2}{S_2^2}$ آن محاسبه شود، این مقدار برآورد خوبی برای نسبت $\frac{\sigma_1^2}{\sigma_2^2}$ است.

می‌دانیم توزیع برآوردگر $\frac{S_1^2}{S_2^2}\times\frac{\sigma_2^2}{\sigma_1^2}$ یک توزیع F با درجات آزادی به ترتیب $\nu_1=n_1-1$ و $\nu_2=n_2-1$ است.

$$F(\nu_1,\nu_2) = \frac{S_1^2}{S_2^2}\times\frac{\sigma_2^2}{\sigma_1^2} \quad (1)$$

SAMAND

Subject:
YEAR: Month: Day:

از طرفی داریم:

$$P\left(f_{\frac{\alpha}{2}}(\nu_1,\nu_2) < F < f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2) \right) = 1-\alpha \quad (2)$$

$$(1),(2) \Rightarrow P\left(\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{\frac{\alpha}{2}}(\nu_1,\nu_2)} \right) = 1-\alpha \quad (3)$$

نتایج زیر از تساوی ۳ بدست می‌آید.

۱) یک بازه اطمینان $(1-\alpha)$٪ برای پارامتر $\frac{\sigma_1^2}{\sigma_2^2}$ (نسبت واریانس‌های دو جامعه نرمال مستقل) عبارتست از

$$\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{\frac{\alpha}{2}}(\nu_1,\nu_2)}$$

$f_{\frac{\alpha}{2}}(\nu_2,\nu_1)$
سمت راست

$f_{1-\frac{\alpha}{2}}(\nu_2,\nu_1)$
سمت چپ

۲- یک بازه اطمینان $(1-\alpha)$٪ برای $\frac{\sigma_1}{\sigma_2}$ برابر است با:

$$\frac{S_1}{S_2}\sqrt{\frac{1}{f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2)}} < \frac{\sigma_1}{\sigma_2} < \frac{S_1}{S_2}\sqrt{f_{1-\frac{\alpha}{2}}(\nu_2,\nu_1)}$$

SAMAND

Subject:
YEAR: Month: Day:

مثال: از دو جامعه مستقل دو نمونه تصادفی با اندازه‌های $n_1=50$ و $n_2=10$ اختیار کرده و نتایج زیر بدست آمده است.

$n_1=50 \Rightarrow \bar{X}_{50}=7.82 \ , \ S_1=0.24$

$n_2=10 \Rightarrow \bar{X}_{10}=4.75 \ , \ S_2=0.3$

یک بازه اطمینان ۹۵٪ برای پارامتر $\theta=\mu_1-\mu_2$ (تفاضل میانگین واقعی جامعه) تعیین کنید.

$$(\bar{X}_1-\bar{X}_2)-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2)+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

در این فرمول به جای σ_1 و σ_2 از S_1^2 و S_2^2 استفاده می‌کنیم چون نمونه‌ها بیشتر از ۳۰ تا هستند.

$1-\alpha=0.95 \Rightarrow \alpha=0.05 \Rightarrow \frac{\alpha}{2}=0.025 \qquad Z_{0.975}=1.96$ از جدول

$$(7.28-4.75)-1.96\left(\sqrt{\frac{(0.24)^2}{50}+\frac{(0.3)^2}{10}}\right) < \mu_1-\mu_2 < (7.28-4.75)-1.96\left(\sqrt{}\right)$$

مثال: یک بازه اطمینان ۹۸٪ برای $\frac{\sigma_1}{\sigma_2}$ تعیین کنید (نسبت انحراف معیار دو جامعه) احتمالاً در دو نمونه

SAMAND

تفاوت دو میانگین مستقل با اندازه‌های $n_1=12$ و $n_2=25$ از دو جامعه نرمال انتخاب شده‌اند نتایج زیر بدست آمده باشد.

$S_1=4 , S_2=3$

$$\frac{S_1}{S_2}\sqrt{\frac{1}{f_{1-\alpha/2}(v_1,v_2)}} < \frac{\sigma_1}{\sigma_2} < \frac{S_1}{S_2}\sqrt{f_{1-\alpha/2}(v_2,v_1)}$$

$1-\alpha = 0.98 \Rightarrow \alpha = 0.02 \Rightarrow \alpha/2 = 0.01$

$$\left.\begin{array}{l} f_{0.99}(15,24) = 2.89 \\ f_{0.99}(24,15) = 3.29 \end{array}\right\} \text{از جدول}$$

تعاریف: برای تخمین یک پارامتر مانند θ از یک جامعه آماری، مراحل زیر انجام میشود:
نقطه‌ای و فاصله‌ای

۱- برای پارامتر مورد نظر که θ نام گذاری شده، تعدادی برآوردگر تعریف می‌کنیم. $(\hat\theta_1, \hat\theta_2, \dots)$

۲- بین برآوردگرهای تعریف شده، نااریب‌ها را انتخاب می‌کنیم. $(E(\hat\theta_i)=\theta, \dots)$

۳- بین برآوردگرهای نااریب، کاراترین را انتخاب می‌کنیم (برآوردگری که کمترین واریانس را دارد)

۴- برای کاراترین برآوردگر، تابع احتمال یا توزیع احتمال تعریف می‌کنیم. (فرض کنیم توزیع برآوردگر نرمال است)

۵- برای یک توزیع آماری برآوردگر و فرمول $P(W_{\alpha/2} < W < W_{1-\alpha/2}) = 1-\alpha$ یک بازه اطمینان بدست می‌آوریم.

تعیین برآورد نقطه و فاصله‌ای برای پارامتر $\theta = P$ (نسبت در آزمایش دو جمله‌ای)

$b(x;n,p) = \binom{n}{x} p^x q^{n-x} \quad x = 0,1,2,\dots$

یک برآوردگر خوب برای پارامتر p، برآوردگر $\hat p = \frac{X}{n}$ است. که در آن X متغیر تصادفی دو جمله‌ای با توزیع $b(x;n,p)$ است.

$$\mu_{\hat p} = E(\hat p) = E\left(\frac{X}{n}\right) = E\left(\frac{1}{n}X\right) = \frac{1}{n}\underbrace{E(X)}_{np} = \frac{np}{n} = p$$

پس برآوردگر $\hat p$ نااریب است.

$$\sigma^2_{\hat p} = \sigma^2_{\frac{1}{n}X} = \frac{1}{n^2}\underbrace{\sigma^2_X}_{npq} = \frac{pq}{n}$$

و اما توزیع $\hat p$،

$$Z = \frac{X - np}{\sqrt{npq}} = \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}} = \frac{\hat p - \mu_{\hat p}}{\sigma_{\hat p}}$$

$\hat p$ دارای توزیع نرمال است (به شرط $n \geq 30$)

Subject:
Year: Month: Day:

$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1-\alpha \quad (2)$$

$$1, 2 \Rightarrow P\left(-z_{1-\alpha/2} < \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} < z_{1-\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{pq}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{pq}{n}}\right) =$$

در طرفین این نامساوی بجای p و q، مقدار تخمین زده شده آن را قرار میدهیم. به عبارت دیگر
بجای p و q، $\hat{p}$ و $\hat{q}$ را به کار میبریم)

$$P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \approx 1-\alpha \quad (3)$$

نتیجه: یک بازه اطمینان $(1-\alpha)$٪ برای پارامتر $\theta = p$ (نسبت در آزمایش دوجمله‌ای)
عبارتست از: تقریبی

$$\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$\hat{p} = \frac{x}{n}$, $\hat{q} = 1-\hat{p}$ ($n \geq 30$ حجم تعداد نمونه)

SAMAND

Subject:
Year: Month: Day:

مثال: یک نمونه تصادفی ۲۰۰ تایی از باطری های قلمی اختیار کرده‌ایم و مشخص شده که ۱۰ عدد از این
باطری ها ناسالم میباشند. یک بازه اطمینان ۹۵٪ برای نسبت واقعی تعداد باطری قلمی سالم
از این کارخانه را تعیین کنید.

$$\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$\hat{p} = \frac{x}{n}$, $\hat{q} = 1-\hat{p}$ (x تعداد کالای سالم در نمونه ۲۰۰ تایی)

$n = 200$, $x = 190$

$\hat{p} = \frac{190}{200} = \frac{19}{20}$ $\quad \hat{q} = 1 - \frac{19}{20} = \frac{1}{20}$

$1-\alpha = 0.95 \longrightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$

$z_{0.975}$ = به کمک جدول = 1.96

$$\frac{19}{20} - \left(1.96 \times \sqrt{\frac{\frac{19}{20}\times\frac{1}{20}}{200}}\right) < p < \frac{19}{20} + \left(1.96 \times \sqrt{\frac{\frac{19}{20}\times\frac{1}{20}}{200}}\right)$$

SAMAND

آزمون فرض ها: testing hypotesis

در این قسمت ابتدا به فرض آماری و آزمون فرض پرداخته سپس به بیان چگونگی انجام مراحل یک آزمون فرض با
هر یک نمونه تصادفی مشخص می کنیم.

تعریف: فرض آماری

یک فرض آماری عبارتست از بیان یا فرضی در مورد پارامتر یک جامعه گفته شود بدون توجه به
اینکه درست یا نادرست باشد.

مثال: ۶۰٪ دانشجویان از خوابگاه استفاده می کنند.

($P = 0/6$: احتمال اینکه یک دانشجو به تصادف انتخاب شود خوابگاهی باشد ۰/۶ است)

مثال: ۳۰٪ دانشجویان از خوابگاه استفاده می کنند.

مثال: میانگین طول عمر لامپ نوع A : $\mu_A = 20$ ساعت است یا $\mu_A < 20$ ساعت

فرض آماری:

۱) ساده: اگر پارامتر جامعه به طور دقیق مشخص شود. $\mu_A = 20$ یا $P = 0/6$

۲) مرکب: اگر پارامتر جامعه به طور دقیق مشخص نشود. $\mu_A < 20$ یا $P < 0/3$

درستی یا نادرستی یک فرض آماری وقتی به طور دقیق مشخص میشود که تمام جامعه مورد بررسی قرار گیرد
(کاری غیر ممکن است) بنابراین تحقیق در مورد درستی یا نادرستی یک فرض آماری را اجباراً توسط
نمونه تصادفی انجام میشود. نمونه تصادفی ← سازگاری ← قبول فرض آماری
رد فرض آماری

آزمون فرض:

تحقیق در مورد درستی یا نادرستی یک فرض آماری به کمک نمونه تصادفی را آزمون فرض می گویند

یک فرض آماری چگونه باید بیان شود؟

یک فرض آماری به گونه ای بیان میشود که در صدد رد آن باشیم، زیرا رد یک فرض آماری با قاطعیت
بیشتری انجام می گیرد.

مثال: فرض کنید می خواهیم بررسی کنیم که یک سکه سالم است یا خیر

فرض آماری: سکه سالم است (احتمال شیر یا خط یکسان است) $P = 1/2$

فرض آماری: سکه سالم نیست (احتمال شیر یا خط یکسان نیست) $P \neq 1/2$

$b(x, n, p) =$ $n = 10$ و $P = \frac{1}{2}$ $x = 0, 1, \dots, 10$

$P(X \le 35) = P(X < 35.5) = P\left(Z < \frac{35.5 - 50}{5}\right) = 0.002$

دوجمله‌ای نرمال

احتمال آنکه سر ≤ ۳۵ در ۱۰۰ پرتاب یک سکه سالم،

$\mu = np = 100 \times \frac{1}{2} = 50$

که در آن ۳۵ شیر مشاهده شود: ۰/۰۰۲ است.

$\sigma = \sqrt{100 \times \frac{1}{2} \times \frac{1}{2}} = 5$

چون «سکه سالم است» رد میشود فرض

در هر آزمون فرض دو نوع فرض آماری تعریف میشود.

۱- فرض صفر (فرض خنثی)

یک فرض آماری که برای روشن شدن تصمیم میشود، فرض صفر میگوییم و با علامت H_0 نشان میدهیم. (فرض صفر چون شرایط موجود را نشان میدهد معمولاً یک فرض ساده است)

$H_0: \theta = \theta_0$

۲- فرض مقابل

یک فرض آماری که در مقابل فرض صفر قرار میگیرد را فرض مقابل میگوییم و با H_1 نشان میدهیم

$H_1: \theta \neq \theta_0$

$H_0: p = \frac{1}{2}$ سکه سالم

$H_1: p \neq \frac{1}{2}$ سکه ناسالم

در هر آزمون فرض همواره دو نوع خطا وجود دارد. عامل خطا نمونه است.

۱- خطای نوع اول ۲- خطای نوع دوم

خطای نوع اول:

اگر فرض H_0 را رد کنیم در صورتیکه درست باشد، خطای نوع اول را مرتکب شده‌ایم. (این خطا یک پیشامد است.) احتمال مرتکب شدن این نوع خطا را سطح تشخیص یا میزان معنی‌دار بودن گویند و با α نشان میدهند.

$\alpha = P($رد فرض H_0 | H_0 درست باشد$)$

خطای نوع دوم:

اگر فرض H_0 را قبول کنیم در صورتیکه نادرست باشد، خطای نوع دوم را مرتکب شده‌ایم. احتمال وقوع این خطا را با β نشان میدهند.

$\beta = P($قبول H_0 | H_0 درست نباشد$)$

یا $\beta = P($ قبول H_0 | H_1 درست باشد $)$

در هر آزمون فرض حوزه دو ناحیه تعریف می‌شود.

۱- ناحیه رد (بحرانی)

ناحیه‌ای که در صورت قرار گرفتن آماره در آن ناحیه باعث تصمیم‌گیری در رد فرض H_0 شود را ناحیه رد یا بحرانی می‌گوئیم

۲- ناحیه قبولی:

ناحیه‌ای که در صورت قرار گرفتن آماره در آن ناحیه باعث تصمیم‌گیری در قبول فرض H_0 شود را ناحیه قبولی می‌نامیم

نقاط مرزی بین این دو ناحیه را نقاط بحرانی می‌گوئیم.

فرض کنید می‌خواهیم آزمون فرض

$$\begin{cases} H_0: p = \frac{1}{2} & \text{سکه سالم است} \\ H_1: p \neq \frac{1}{2} & \text{سکه ناسالم است} \end{cases}$$

را ناحیه‌بندی کنیم

$x=0$ ناحیه رد — ۳۵ — ناحیه قبولی — ۴۴ — ناحیه رد $x=10$

$$\begin{cases} X \leq 35 \\ X \geq 44 \end{cases}$$ ناحیه رد

ناحیه قبولی $35 < X < 66$

آیا α و β را می‌توان کم کرد؟ آیا می‌توان α و β را همزمان کاهش داد؟

مثال: گزارش شده است که میانگین قد دانشجویان یک دانشگاه $\mu = 170$ cm و با انحراف معیار $\sigma = 9$ cm است. برای بررسی این گزارش (درستی یا نادرستی این فرض آماری) آزمون‌های زیر را انجام داده‌ایم.

1) $\begin{cases} H_0: \mu = 170 \\ H_1: \mu \neq 170 \end{cases}$ 2) $\begin{cases} H_0: \mu = 170 \\ H_1: \mu = 175 \end{cases}$

در نمونه تصادفی $n=36$ و $n=64$ اعتبار در هر کدام یک ناحیه بحرانی در حالت α و β را محاسبه کنید.

در آزمون ۱: ابتدا برای $n=36$ محاسبه می‌کنیم.

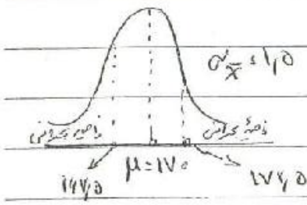
می‌دانیم یک آماره (برآوردگر) مناسب برای μ، آماره $\bar{X}$ است. چون نمونه $n=36$ است،

بنا به قضیه حد مرکزی، $\bar{X}$ دارای توزیع نرمال است.

$\mu_{\bar{X}} = 170$

$\sigma_{\bar{X}} = \frac{9}{\sqrt{36}} = \frac{9}{6} = \frac{3}{2} = ۱/۵$

۲۵



ناحیه بحرانی دلخواه
$$\begin{cases} \bar{X} < 167.5 \\ \bar{X} > 172.5 \end{cases}$$

ناحیه قبولی $167.5 < \bar{X} < 172.5$

$$\alpha = P(\bar{X} < 167.5 \mid \mu = 170) + P(\bar{X} > 172.5 \mid \mu = 170) =$$

$$= 2P(\bar{X} < 167.5 \mid \mu = 170) = 2P\left(Z < \frac{167.5 - 170}{1.5}\right) = 0.097$$

$$\beta = P(167.5 < \bar{X} < 172.5 \mid \mu \neq 170) =$$ قابل محاسبه نیست

در آزمون ۲ :

$\alpha = 0.097$

$$\beta = P(167.5 < \bar{X} < 172.5 \mid \mu = 175) = P\left(\frac{167.5 - 170}{1.5} < Z < \frac{172.5 - 170}{1.5}\right)$$

$$= 0.04$$

اگر $n = 64$ را در نظر بگیریم:
$$\sigma = \frac{9}{\sqrt{64}} = \frac{9}{8} = 1.125$$

$$\alpha = P(\bar{X} < 167.5) + P(\bar{X} > 172.5) = 2P\left(Z < \frac{167.5 - 170}{1.125}\right) = 0.02$$

$$\beta = P(167.5 < \bar{X} < 172.5) = P\left(\frac{167.5 - 170}{1.125} < Z < \frac{172.5 - 170}{1.125}\right) = 0.01$$

$$n = 36 : \begin{cases} \alpha = 0.09 \\ \beta = 0.04 \end{cases} \qquad n = 64 : \begin{cases} \alpha = 0.02 \\ \beta = 0.01 \end{cases}$$

در توزیع‌های پیوسته به جای آنکه ناحیه بحرانی را دلخواه انتخاب کرده و مقدار α را محاسبه کنیم می‌توان برعکس عمل کرد. به عبارت دیگر در چنین آزمون فرض‌هایی ابتدا α را انتخاب و به کمک آن ناحیه بحرانی مشخص می‌شود. معمولاً در مورد α را ۰٫۰۵ یا ۰٫۰۱ انتخاب می‌کنیم. اگر در مسئله آزمون فرض α را مشخص نکرده، $\alpha = 0.05$ اختیار می‌شود. (آزمون در سطح تشخیص ۰٫۰۵ انجام می‌شود.)

* آزمون‌های یک طرفه و دو طرفه:

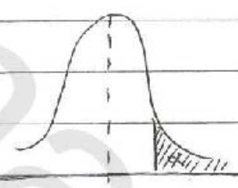
آزمون یک طرفه به آزمون فرضی می‌گویند که فرض مقابل آن یک طرفه باشد. به عبارت دیگر آزمون

بصورت زیر باشد:

$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_1 \end{cases}$ یا $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$

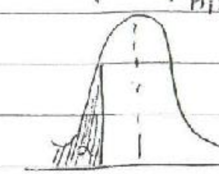
مثال: $\begin{cases} H_0: \mu = 170 \\ H_1: \mu > 170 \end{cases}$ یا $\begin{cases} H_0: \mu = 170 \\ H_1: \mu < 170 \end{cases}$:مثال

از ویژگی های یک آزمون یک طرفه آن است که ناحیه بحرانی آن یک طرفه بوده و در یک سمت توزیع قرار می‌گیرد. (اگر آزمون بصورت $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$ باشد، ناحیه بحرانی سمت چپ توزیع قرار می‌گیرد. و اگر $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ باشد، ناحیه بحرانی در سمت راست توزیع قرار می‌گیرد.



$H_0: \mu_0$

$H_1: \mu > \mu_0$



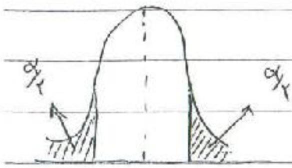
$H_0 = \mu_0$

$H_1: \mu < \mu_0$

* آزمون فرض دوطرفه به آزمونی گفته می‌شود که فرض مقابل آن دو طرفه باشد، به عبارت دیگر

$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$ باشد

از ویژگی آزمون های دو طرفه آن است که ناحیه بحرانی در دو سمت توزیع قرار می‌گیرد.



$H_0: \mu_0$

$H_1: \mu \neq \mu_0$

بطور کلی برای انجام یک آزمون فرض آماری، هنگامی که از توزیع های بهینه استفاده میشود (آزمونهای مبتنی بر توزیع های بهینه را بالا باشند)، مراحل زیر اجرا میشود:

۱_ تشکیل فرض H_0 $H_0: \theta = \theta_0$

۲_ تشکیل فرض مقابل $H_1: \theta < \theta_0$ یا $\theta > \theta_0$ یا $\theta \neq \theta_0$

۳_ انتخاب سطح مشخص آزمون (انتخاب مقدار α)

۴_ انتخاب آماره مناسب برای θ و مشخص نوع توزیع آماره یا آورده‌گر

۵_ تشکیل ناحیه بحرانی به کمک α و توزیع آماره

۶_ محاسبات: به کمک نمونه منتخب مقدار آماره مشخص میشود.

۷- نتیجه‌گیری: اگر مقدار آماره در ناحیه بحرانی قرار گیرد، فرض H_0 رد می‌شود. در غیر این صورت می‌گوییم دلیلی برای رد H_0 پیدا نشد.

مثال: گزارش شده است که میانگین طول عمر نوعی لامپ، $\mu = 1200$ ساعت با انحراف معیار $\sigma = 300$ ساعت است. کارخانه‌ای مدعی است لامپ جدیدی از این نوع تولید کرده که دارای میانگین طول عمر بیشتر از ۱۲۰۰ ساعت است. برای بررسی این ادعا یک نمونه تصادفی ۱۰۰ تایی از این نوع لامپ جدید انتخاب و مشخص شده است که میانگین طول عمر این نمونه $\bar{x}_{100} = 1265$ ساعت است. در مورد این ادعا چه اظهار نظری می‌توان کرد. به عبارت دیگر می‌خواهیم آزمون آماری داشته باشیم

که $\begin{cases} H_0: \mu = 1200 \\ H_1: \mu > 1200 \end{cases}$ باشد. (سطح تشخیص $\alpha = 0.05$ است)

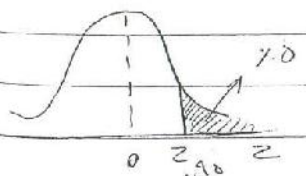
۱) $H_0: \mu = 1200$ → آزمون یک طرفه

۲) $H_1: \mu > 1200$

۳) $\alpha = 0.05$ (آزمون یک طرفه)

۴) μ، آماره مناسب برای میانگین $= \bar{x}$

۵) توزیع $\bar{x}$ نرمال است چون ($n = 100 \geq 30$) است. با نرمال استاندارد $Z_1 = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$



$Z_{0.95}$ = با کمک جدول $= 1.645$

$Z > 1.645$ ناحیه بحرانی

۶) محاسبات: $n = 100$ ، $\bar{x} = 1265$ ، $\sigma = 300$

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1265 - 1200}{\frac{300}{\sqrt{100}}} = 2.16$$

۷) نتیجه‌گیری: فرض H_0 رد می‌شود. می‌پذیریم ادعای این تولید کننده درست است.

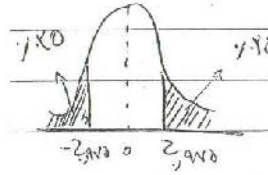
مثال: میانگین یک نمونه ۱۰۰ تایی از نوعی لامپ دارای طول عمر $\bar{x}_{100} = 1570$ و انحراف معیار $S = 120$ ساعت است. آزمون فرض‌های زیر را در سطح تشخیص $\alpha = 0.05$ و $\alpha = 0.01$ انجام دهید.

$H_0: \mu = 1600$ → آزمون دو طرفه

$H_1: \mu \neq 1600$

$\alpha = 0.05 \longrightarrow \frac{\alpha}{2} = 0.025$

$\bar{x}$ آماره مناسب ، توزیع نرمال ← نرمال استاندارد $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$



ناحیه بحرانی $\begin{cases} Z > z_{0.975} \\ Z < -z_{0.975} \end{cases}$ $\xrightarrow{\text{جدول}}$ $\begin{cases} Z > 1.96 \\ Z < -1.96 \end{cases}$

محاسبات : $n = 100$ ، $\bar{x} = 1570$ ، $S = 120$

$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1570-1600}{\frac{120}{\sqrt{100}}} = -2.5$ ← این عدد در ناحیه بحرانی هست

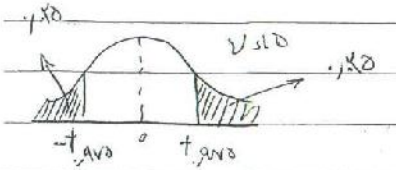
نتیجه گیری: فرض H_0 رد میشود. چون Z منفی شده ، پس کمتر شده است.

مثال: میانگین یک نمونه ۱۶تایی از یک جامعه نرمال دارای $\bar{x}_{16} = 50$ و انحراف معیار $S=5$ است. به کمک این نمونه آزمون را بصورت زیر انجام داده ایم:

$\begin{cases} H_0 : \mu = 42 \\ H_1 : \mu \neq 42 \end{cases}$

آزمون دوطرفه : $\alpha = 0.05 \rightarrow \frac{\alpha}{2} = \%.25$

(توزیع t با $df=15$ است.) $T = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$: توزیع آماره ، آماره مناسب $\bar{x}$ است



ناحیه بحرانی $\begin{cases} t > t_{0.975} \\ t < -t_{0.975} \end{cases}$ ← از جدول $\begin{cases} t > 2.13 \\ t < -2.13 \end{cases}$

محاسبات : جامعه نرمال، $n = 16$ ، $\bar{x}_{16} = 50$ ، $S = 5$

$T = \frac{50-42}{\frac{5}{\sqrt{16}}} = 6.4$

نتیجه گیری: فرض H_0 رد میشود

مثال: از جامعه نرمالی یک نمونه تصادفی انتخاب و نتایج زیر بدست آمده است: $(n=11)$

۱.۲ ، ۰.۹ ، ۱.۴ ، ۱ ، ۱.۷ ، ۰.۹ ، ۰.۱ ، ۱ ، ۱.۸ ، ۰.۸

در سطح تشخیص ۰.۰۵ : الف: $\alpha = 0.05$ آزمون را بصورت $\begin{cases} H_0 : \mu = 1 \\ H_1 : \mu > 1 \end{cases}$ انجام دهید.

ب: $\alpha = 0.1$ آزمون را بصورت $\begin{cases} H_0 : \sigma^2 = 0.3 \\ H_1 : \sigma^2 \neq 0.3 \end{cases}$ انجام دهید.

ب: به کمک نمونه $\bar{x}$ و S^2 را بدست میآوریم:

$\bar{x} = 1.08$

$S^2 = 0.24$

۲۷

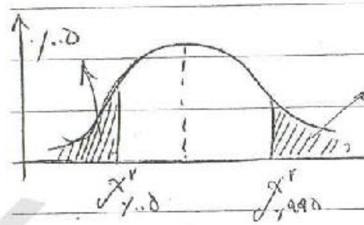
Subject:
Year: Month: Day:

$H_0: \sigma^2 = ۰٫۳$

$H_1: \sigma^2 \neq ۰٫۳$

آزمون دوطرفه → $\alpha = ۰٫۱$ → $\alpha/۲ = ۰٫۰۵$

آماره مناسب S^2 ، توزیع آماره $X = \frac{(n-1)S^2}{\sigma^2}$ در توزیع X توان دوم کای با ۱۰ د.آ است



ناحیه بحرانی $\begin{cases} X > \chi^2_{0.95} \\ X < \chi^2_{0.05} \end{cases}$

$\cancel{\chi^2_{0.995}}$

از جدول $\begin{cases} X > ۲۵٫۳ \\ X < \end{cases}$

محاسبات: $n = ۱۱$ ، $\bar{x} = ۱٫۰۸$ ، $S^2 = ۰٫۲۴$

$X = \frac{(n-1)S^2}{\sigma^2} = \frac{۱۰ \times ۰٫۲۴}{۰٫۳} = ۸$

نتیجه‌گیری: فرض H_0 رد نمی‌شود.

Subject:
Year: Month: Day:

مثال: دو فرمول غذایی با اندازه $n_1 = ۱۶$ و $n_2 = ۱۴$ از دو جامعه نرمال انتخاب و نتایج زیر بدست آمده است.

$\bar{x}_1 = ۷۸$ ، $S_1 = ۶٫۱$ ، $n_1 = ۱۶$

$\bar{x}_2 = ۶۰$ ، $S_2 = ۵٫۳$ ، $n_2 = ۱۴$

آزمون زیر را در سطح معنی‌داری ۰٫۱ در صورت زیر انجام دهید

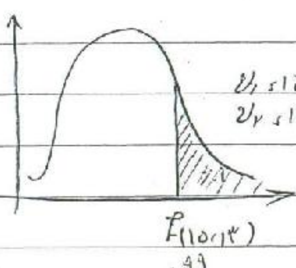
$\begin{cases} H_0: \sigma_1 = \sigma_2 \\ H_1: \sigma_1 > \sigma_2 \end{cases}$

$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 > \sigma_2^2 \end{cases} \iff \begin{cases} H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_1: \frac{\sigma_1^2}{\sigma_2^2} > 1 \end{cases}$

$\begin{cases} H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_1: \frac{\sigma_1^2}{\sigma_2^2} > 1 \end{cases}$ → آزمون یک طرفه → $\alpha = ۰٫۱$

آماره مناسب $\frac{S_1^2}{S_2^2}$ ، توزیع آماره $F(\nu_1, \nu_2) = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2}$

$\nu_1 = ۱۵$ $\nu_2 = ۱۳$



$\nu_1 = ۱۵$
$\nu_2 = ۱۳$

ناحیه بحرانی: $F > F_{0.99}(15,13) = ۳٫۸۲$

$F > ۳٫۸۲$

Subject:
YEAR: MONTH: DAY:

Subject:
YEAR: MONTH: DAY:

$$F = \frac{(۶,۱)^۲}{(۵,۳)^۲} \times ۱ = ۱,۱۵$$

نتیجه‌گیری: فرض H_0 رد نمی‌شود. یعنی واریانس یا انحراف معیار این دو جامعه یکی است