آمار و احتمالات مهندسی

نظریه برآورد

دکتر بهناز بیگدلی

دانشکده مهندسی عمران

دانشگاه صنعتی شاهرود

ساختار درس

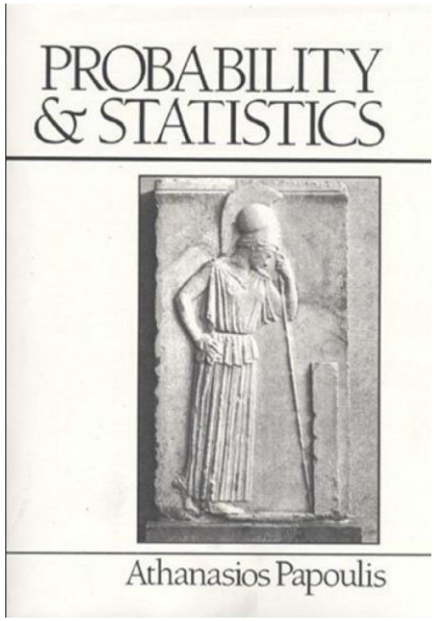
نحوه ارزیابی

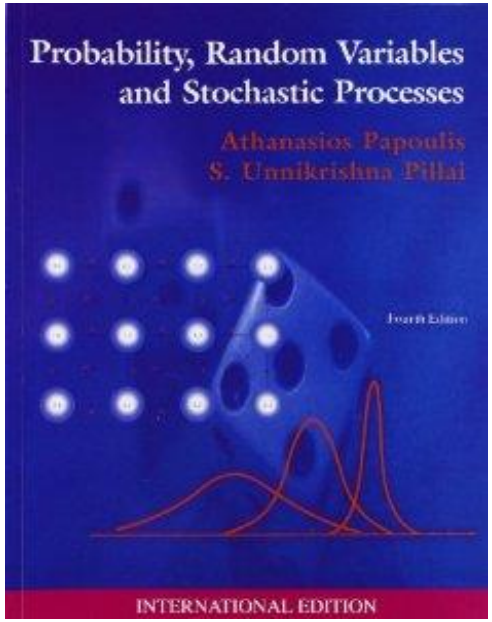
امتحان پایان ترم (۱۴)

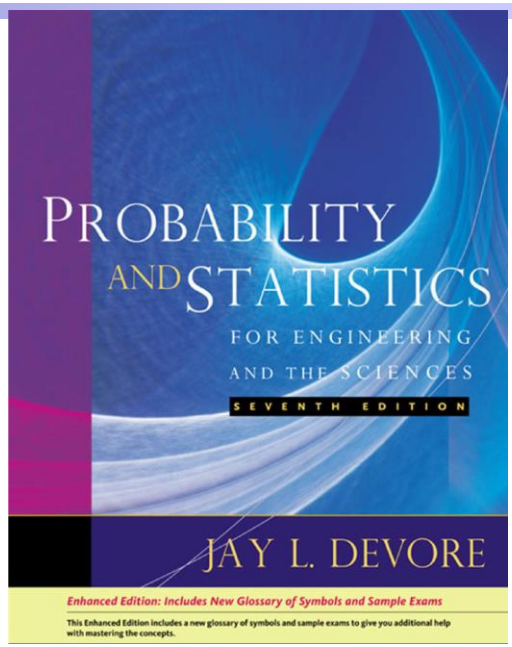
امتحان میان ترم (۴)

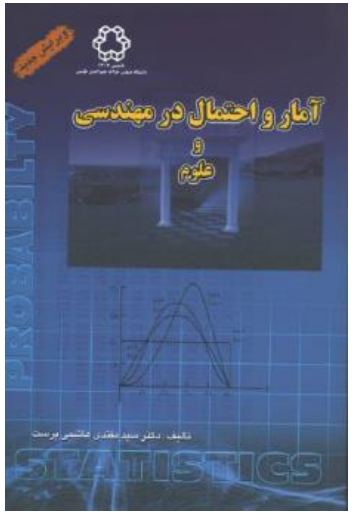
تمرینات و کوئیزها (۲)

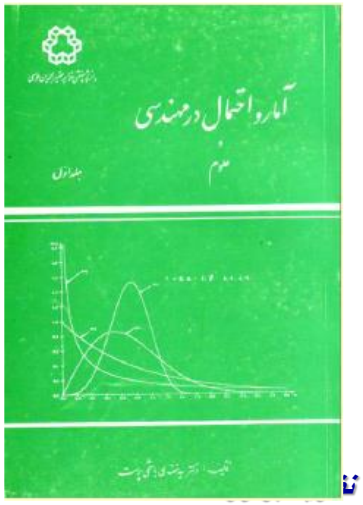
منابع


PROBABILITY
& STATISTICS
Athanasios Papoulis


Probability, Random Variables
and Stochastic Processes
Athanasios Papoulis
S. Unnikrishna Pillai
Fourth Edition
INTERNATIONAL EDITION


PROBABILITY
AND STATISTICS
FOR ENGINEERING
AND THE SCIENCES
SEVENTH EDITION
JAY L. DEVORE
Enhanced Edition: Includes New Glossary of Symbols and Sample Exams
This Enhanced Edition includes a new glossary of symbols and sample exams to give you additional help with mastering the concepts.


آمار و احتمال در مهندسی
و
علوم


آمار و احتمال در مهندسی

فهرست و عناوین درس

I. بخش اول: آمار مقدماتی (توصیفی)

1. آمار مقدماتی
2. اندازه‌گیری پراکندگی، گشاورها، چولگی و کشیدگی
3. برازش خط و منحنی بر داده‌ها
4. توزیع‌های دو بعدی و ضریب همبستگی

II. احتمال

1. احتمال و فضای نمونه
2. فضای نمونه با عناصر متعدد
3. احتمالات شرطی و نابستگی
4. تابع چگالی احتمال، تابع توزیع و امید ریاضی
5. توزیع‌های گسسته
6. توزیع‌های پیوسته مهم
7. نظریه برآورد
8. آزمون‌های فرض

نظریه برآورد

فهرست مطالب این فصل:

۱- توزیع مشترک
۲- توابع خطی از متغیرهای تصادفی مستقل
۳- توزیع میانگین
۴- قضیه حد مرکزی
۵- تقریب نرمال برای توزیع دو جمله ای
۶- توزیع واریانس نمونه
۷- توزیع t
۸- توزیع نسبت واریانس دو نمونه
۹- برآوردگر

توزیع مشترک

- فرض کنید متغیر تصادفی گسسته X دارای تابع احتمال $f(x)=p(X=x)$ باشد و $X_1, X_2, \ldots, X_n$ نمونه‌های تصادفی مستقل از هم باشند. اگر $x_1, x_2, \ldots, x_n$ مقادیر متناظر مشاهده برای $X_1, X_2, \ldots, X_n$ باشند. پیشامدهای $X_1=x_1, X_2=x_2, \ldots, X_n=x_n$ به طور مجزا از هم مستقل‌اند و احتمال توام آنها برابر است با:

$$P[X_1=x_1, X_2=x_2, \ldots, X_n=x_n]=P(X_1=x_1)P(X_2=x_2)\ldots P(X_n=x_n)$$

اگر از نماد $f(x_1, x_2, \ldots, x_n)$ به جای $P[X_1=x_1, X_2=x_2, \ldots, X_n=x_n]$ استفاده کنیم:

توزیع مشترک

$$f(x_1,x_2,\ldots,x_n)=f(x_1)\,f(x_2)\,\ldots\,f(x_n)=\prod_{i=1}^{n} f\left(x_i\right)$$

- برای $n=۲$، $f(x_1,x_2)=f(x_1)\,f(x_2)$ تابع چگالی احتمال توام X_1 و X_2 می‌باشد. در حالتی که متغیر تصادفی X از نوع پیوسته است تابع چگالی مشترک یا توام را با $f(x_1,x_2,\ldots,x_n)$ نیز نمایش می‌دهند.

توابع خطی از متغیرهای تصادفی مستقل

- فرض کنید $X_1, X_2, \ldots, X_n$ نمونه تصادفی مستقل با توزیع مشترک $f(x_1, x_2, \ldots, x_n)$ باشند. یک تابع خطی یا آماره را می‌توان در حالت کلی به صورت زیر تعریف کرد.

$$\sum_{i=1}^{n} a_i X_i = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

که a_i ها مقادیر ثابت هستند.

توابع خطی از متغیرهای تصادفی مستقل

اگر X_1 و X_2 به ترتیب دارای میانگین های μ_1 و μ_2 و واریانسهای σ_1^2، σ_1^2 باشند. میانگین و واریانس Y به طریق زیر محاسبه می‌شود:

$$\mu_y = E(a_1X_1 + a_2X_2) = \sum_{x_1}\sum_{x_2}(a_1x_1 + a_2x_2)f_1(x_1)f_2(x_2)$$

$$= a_1\left[\sum_{x_1} x_1 f_1(x_1)\right]\left[\sum_{x_2} f_2(x_2)\right] + a_2\left[\sum_{x_1} f_1(x_1)\right]\left[\sum_{x_2} x_2 f_2(x_2)\right]$$

$$= a_1E(X_1) + a_2E(X_2) = a_1\mu_1 + a_2\mu_2$$

$$\sigma_y^2 = V(Y) = E\left[Y - \mu_y^2\right] = E\left[(a_1X_1 + a_2X_2) - a_1\mu_1 - a_2\mu_2\right]^2$$

$$= E\left[(a_1(X_1 - \mu_1) + a_2(X_2 - \mu_2)\right]^2$$

$$= E\left[(a_1^2(X_1 - \mu_1)^2 + a_2^2(X_2 - \mu_2)^2 + 2a_1a_2(X_1 - \mu_1)(X_2 - \mu_2)\right]$$

$$= a_1^2V(X_1) + a_2V(X_2) + 0$$

$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2$$

توزیع میانگین

- ✓ در آمار توصیفی، میانگین نمونه تصادفی به صورت $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ تعریف شده بود. در این بخش، توزیع $\overline{X}$ را با توجه به توزیع جامعه‌ای که نمونه از آن گرفته شده بدست می‌آوریم.

توزیع میانگین

قضیه ۱

✓ اگر $X_1, X_2, \ldots, X_n$ نمونه‌های مستقل و هم توزیع از جامعه‌ای با میانگین μ و واریانس σ^2 باشند آنگاه میانگین نمونه X دارای میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.

برهان: چون $\overline{X}$ یک ترکیب خطی از X_i هاست، پس:

$$E(\overline{X}) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu$$

$$V(\overline{X}) = V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$\overline{X}$ تابعی از X_i هاست و به پارامترهای جامعه μ و σ^2 بستگی ندارد. یک آماره نااریب و دارای کمترین واریانس است.

توزیع میانگین

قضیه ۲

✓ اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی n تایی از جامعه نرمال با میانگین μ و واریانس σ^2 باشند آنگاه $\overline{X}$ دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.

برهان: $X \sim N(\mu, \sigma^2)$ است. تابع مولد گشتاورهای آن برابر است با

$$M_X(t) = e^{\mu_t + \frac{1}{2}\sigma^2 t^2}$$

$$M_{\overline{X}}(t) = E\left[e^{t\overline{X}}\right] = E\left[e^{t.\frac{1}{n}\sum X_i}\right] = E\left[e^{\frac{t}{n}(X_1 + X_2 + \ldots + X_n)}\right]$$

$$= E\left[e^{\frac{tX_1}{n} + \frac{tX_2}{n} + \ldots + \frac{tX_n}{n}}\right] = \left(E\left[e^{\frac{tX}{n}}\right]\right)^n$$

توزیع میانگین

قضیه ۲

- ✓ چون X_iها مستقل و هم توزیع اند.

$$=\left[M_x(\frac{t}{n})\right]^n=\left[e^{\mu\frac{t}{n}+\frac{1}{2}\frac{\sigma^2t^2}{n^2}}\right]^n=e^{\mu t+\frac{\sigma^2t^2}{2n}}$$

- ✓ $e^{\mu t+\frac{\sigma^2t^2}{2n}}$ تابع مولد گشتاورهای متغیر تصادفی نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.

توزیع میانگین

قضیه ۳

✓ اگر شرایط قضیه ۲ برقرار باشد متغیر $Z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد است.

$$E(Z) = E\left(\frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}\right) = \frac{\sqrt{n}}{\sigma} E(\overline{X} - \mu) = \frac{\sqrt{n}}{\sigma}(E\overline{X} - \mu) = 0$$

$$V(\overline{X}) = V\left(\frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}\right) = \frac{n}{\sigma^2} V(\overline{X} - \mu) = \frac{n}{\sigma^2} . \frac{\sigma^2}{n} = 1$$

توزیع‌های نمونه‌ای

- ✓برای استنباط در مورد پارامترهای جامعه باید از آماره‌های مناسب استفاده کرد، بنابراین متناظر با هر پارامتر در جامعه یک آماره وجود دارد که خود این آماره یک متغیر تصادفی می‌باشد. آماره خود دارای یک تابع احتمال می‌باشد که براساس نمونه‌های تصادفی nتایی که به تکرار از جامعه آماری انتخاب شده است بدست می‌آید این تابع احتمال را توزیع نمونه‌گیری آماره گویند.

قضیه حد مرکزی

- اگر $\bar{X}$ میانگین نمونه تصادفی $X_1, X_2, \ldots, X_n$ از توزیعی (جامعه‌ای) با میانگین μ و واریانس متناهی $\sigma^2 < \infty$ باشند آنگاه توزیع متغیر تصادفی $Z = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ میل می‌کند به توزیع نرمال استاندارد اگر $n \to \infty$
- این قضیه با استفاده از تابع مولد گشتاورها به راحتی اثبات می‌شود.

قضیه حد مرکزی

✓ مثال: فرض کنید وزن افراد یک جامعه بزرگ دارای میانگین ۸۰ و انحراف‌معیار ۲۰ کیلوگرم است. نمونه‌ای تصادفی و ۱۶تایی انتخاب شده است، احتمال آنکه میانگین این نمونه بیش از ۸۵ کیلوگرم باشد، کدام است؟

✓ مثال: چون $\bar{X}$ دارای توزیع نرمال است(قضیه حد مرکزی) داریم:

$$\bar{X} \sim N(\mu = 80, \frac{\sigma^2}{n} = \frac{(20)^2}{16})$$

$$P(\bar{X} > 85) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{85 - 80}{\frac{20}{\sqrt{16}}}\right) =$$

$$P(Z > 1) = 1 - P(Z < 1) = 1 - 0.8413 = 0.1578$$

خلاصه توزیع‌های نمونه‌ای

۱- در صورتی که از جامعه‌ای حجم N، نمونه‌ای nتایی انتخاب شود:

اگر حجم جامعه نامتناهی باشد.

$$\begin{cases} E(\bar{X}) = \mu \\ Var(\bar{X}) = \dfrac{\sigma^2}{n} \end{cases}$$

اگر حجم جامعه متناهی باشد.

$$\begin{cases} E(\bar{X}) = \mu \\ Var(\bar{X}) = \dfrac{N-n}{N-1} \cdot \dfrac{\sigma^2}{n} \end{cases}$$

۲- در صورتی که نمونه‌ای nتایی از جامعه نرمال با واریانس معلوم σ^2 انتخاب شود:

$$\bar{X} \sim N(\mu_X, \frac{\sigma_X^2}{n}) \Longrightarrow \begin{cases} E(\bar{X}) = \mu_X \\ Var(\bar{X}) = \dfrac{\sigma_X^2}{n} \end{cases}$$

خلاصه توزیع‌های نمونه‌ای

۳- در صورتی که نمونه‌ای nتایی از جامعه نرمال با واریانس نامعلوم σ^2 انتخاب شود:

اگر حجم نمونه $n \geq 30$ باشد.

$$\bar{X} \rightsquigarrow N\left(\mu_X, \frac{S_X^2}{n}\right) \Rightarrow Z = \frac{\bar{X} - \mu}{\frac{S_X}{\sqrt{n}}} \sim N(0,1)$$

اگر حجم نمونه $n < 30$ باشد.

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma^2}{n}\right) \Rightarrow t = \frac{\bar{X} - \mu}{\frac{S_X}{\sqrt{n}}} \sim t_{(n-1)}$$

۴- در صورتی که نمونه‌ای nتایی از جامعه‌ای با توزیع نامعلوم و واریانس معلوم σ^2 انتخاب شود:

اگر حجم نمونه $n \geq 30$ باشد.

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right) \Rightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma_X}{\sqrt{n}}} \sim N(0,1)$$

اگر حجم نمونه $n < 30$ در این صورت توزیع $\bar{X}$ مانند توزیع جامعه نامعلوم بوده و از نامساوی چی بی شف میانگین جامعه μ را برآورد می‌کنیم.

تقریب نرمال برای توزیع دوجمله‌ای

- در توزیع دوجمله‌ای با پارامترهای n و p برای nهای بزرگ محاسبه احتمال گاهی اوقات با استفاده از جداول ضمیمه خسته کننده و گاهی ممکن است جدولی با چنین nای در دسترس نباشد.

- می دانیم اگر Y دارای توزیع دوجمله‌ای باشد، می‌توان Y را به صورت جمعی از متغیرهای برنولی یعنی $Y = \sum_{i=1}^{n} X_i$ نوشت که X_iها متغیرهای برنولی با میانگین p و واریانس $p(1-p)$ می‌باشند و مقادیری که Y اختیار می‌کند اعداد صحیح ۰، ۱، ۲، ...، n است.

تقریب نرمال برای توزیع دوجمله‌ای

- متغیر Y را که از نوع گسسته است می‌توان با توجه به نتیجه قضیه حد مرکزی به وسیله متغیر نرمال استاندارد تقریب زد.

احتمال پیشامد $Y=k$ را می‌توان به صورت زیر تقریب زد.

$$p[Y=k]=p\left[k-\frac{1}{2}<Y<k+\frac{1}{2}\right]\approx\int_{k-\frac{1}{2}}^{k+\frac{1}{2}}f(y)\,dy$$

$$P\left[Y=k\right]\approx P\left[\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}<\frac{Y-np}{\sqrt{np(1-p)}}<\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]$$

تقریب نرمال برای توزیع دوجمله‌ای

✓ادامه

$$P[Y=k] \approx P\left[\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}} < Z < \frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]$$

$$= \phi\left[\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right] - \phi\left[\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]$$

✓که تابع $\Phi(t)$ برابر است با:

$$\phi(t) = p[Z<t] = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

توزیع واریانس نمونه

- واریانس نمونه nتایی در آمار توصیفی به صورت

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

- تعریف شده بود. اکنون برای ناریب بودن، آن را به صورت تعریف می‌کنیم.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

توزیع واریانس نمونه

✓ قضیه ۱: اگر متغیر Z دارای توزیع نرمال استاندارد باشد آنگاه Z^2 دارای توزیع کی‌دو با یک درجه آزادی است.

✓ برهان: با استفاده از تابع مولد گشتاورها

✓ برای متغیر Z

$$M_z(t) = E[e^{tz}] = e^{\frac{1}{2}t^2}$$

✓ برای متغیر Z^2

$$M_{z^2}(t) = E[e^{tz^2}] = \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[1-2t]z^2} dz$$

✓ با فرض $u = \sqrt{1-2t}\,z$

$$M_{z^2}(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \frac{du}{\sqrt{1-2t}} = (1-2t)^{-\frac{1}{2}}$$

توزیع واریانس نمونه

- قضیه ۲: اگر متغیرهای مستقل $Z_1, Z_2, \ldots, Z_n$ دارای توزیع نرمال استاندارد باشند، آنگاه $\sum_{i=1}^{n} Z_i^2$ دارای توزیع کی‌دو با n درجه آزادی است.

- اثبات این قضیه با استفاده از تابع مولد گشتاورها آسان است که در اینجا بدون اثبات می‌پذیریم. از این قضیه استنتاج می‌شود که اگر دو متغیر مستقل دارای توزیع کی‌دو باشند جمع آنها نیز توزیع کی‌دو است. در مورد تفاضل هم در شرایط خاص درست است. یعنی اگر دو متغیر مستقل دارای توزیع کی‌دو باشند تفاضل آنها نیز دارای توزیع کی‌دو است با تفاضل درجه آزادی دو متغیر.

توزیع واریانس نمونه

✓ قضیه ۳: اگر $\overline{X}$ و S^2 به ترتیب میانگین و واریانس نمونه $X_1, X_2, \ldots, X_n$ از جامعه نرمال با میانگین μ و واریانس σ^2 باشد آنگاه

الف- $\overline{X}$ و S^2 از هم مستقل‌اند.

ب- متغیر $\frac{(n-1)S^2}{\sigma^2}$ دارای توزیع کی‌دو با $n-1$ درجه آزادی است.

توزیع واریانس نمونه

۱- اگر از جامعه نرمال با میانگین مجهول μ و واریانس σ^2 نمونه‌های تصادفی به حجم n انتخاب شود.

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \qquad S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1} \qquad \begin{cases} E(S^2) = \sigma^2 \\ Var(S^2) = \dfrac{2\sigma^4}{n-1} \end{cases}$$

۲- اگر از جامعه‌ای با میانگین مجهول μ و واریانس σ^2 نمونه‌های تصادفی به حجم بزرگ n انتخاب شود ($n \geq 30$).

$$X^2 = \frac{n\tilde{S}^2}{\sigma^2} \sim \chi^2_{n-1} \qquad \tilde{S}^2 = \frac{\sum(X_i - \bar{X})^2}{n} \qquad \begin{cases} E(\tilde{S}^2) = \dfrac{n-1}{n}\sigma^2 \\ Var(\tilde{S}^2) = \dfrac{2(n-1)\sigma^4}{n^2} \end{cases}$$

توزیع واریانس نمونه

۳- توزیع نمونه‌ای واریانس نمونه‌ای وقتی میانگین جامعه μ معلوم باشد.

$$X^2 = \frac{nS_*^2}{\sigma^2} \sim \chi^2_{(n)}$$

$$S_*^2 = \frac{\sum (X_i - \mu)^2}{n}$$

$$\begin{cases} E(S_*^2) = \sigma^2 \\ Var(S_*^2) = \dfrac{2\sigma^4}{n} \end{cases}$$

توزیع t

- فرض کنید که $X_1, X_2, \ldots, X_n$ یک نمونه nتایی از توزیع نرمال با میانگین μ و و واریانس σ^2 باشد.
- می‌دانیم متغیر $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد و
- متغیر $\dfrac{(n-1)S^2}{\sigma^2}$ دارای توزیع کی‌دو با $n-1$ درجه آزادی است.

متغیر T را که تابعی از دو متغیر است به صورت زیر تعریف می‌کنیم.

$$T=\frac{\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\dfrac{S}{\sigma}}=\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$$

توزیع نسبت واریانس دو نمونه

- فرض کنید از جامعه اول که دارای توزیع نرمال با واریانس σ_1^2 است نمونه $X_1, X_2, \ldots, X_{n_1}$ به حجم n_1 و از جامعه دوم که دارای توزیع نرمال با واریانس σ_2^2 است نمونه $Y_1, Y_2, \ldots, Y_{n_2}$ به حجم n_2 موجود باشند.

- اگر $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$ و $S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$ به ترتیب واریانس‌های نمونه اول و دوم باشند. متغیرهای $\frac{(n_1-1)S_1^2}{\sigma_1^2}$ و $\frac{(n_2-1)S_2^2}{\sigma_2^2}$ طبق قضیه قبل دارای توزیع کی دو با درجه آزادی $n_1 - 1$ و $n_2 - 1$ هستند.

توزیع نسبت واریانس دو نمونه

- قضیه ۱: اگر متغیرهای تصادفی مستقل X, Y به ترتیب دارای توزیع کی دو با m, n درجه آزادی باشند، آنگاه متغیر $F = \frac{X/m}{Y/n}$ دارای توزیع فیشر با m, n درجه آزادی است. معمولاً m را درجه آزادی صورت و n را درجه آزادی مخرج می‌گویند و آن را با علامت $F(m,n)$ نشان می‌دهند.
- برای بدست آوردن توزیع نسبت واریانس دو نمونه، متغیر F را به صورت زیر تعریف می‌کنیم.

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$$

- متغیر F شرایط قضیه یک را دارا می‌باشد و دارای توزیع فیشر با $n_1 - 1$ و $n_2 - 1$ درجه آزادی است.

توزیع نسبت واریانس دو نمونه

✓ تبصره : اگر متغیر تصادفی F دارای توزیع فیشر با m , n درجه آزادی باشد، آنگاه متغیر $\frac{1}{F}$ دارای توزیع فیشر با n, m درجه آزادی است. یعنی :

$$F_{(\alpha,m,n)} = \frac{1}{F_{(1-\alpha,n,m)}}$$

✓ این نتیجه به ما اجازه می‌دهد توزیع F را تنها برای دنباله سمت راست در جداول ضمیمه درج کنیم.

توزیع نسبت واریانس دو نمونه

✓ مثال: با توجه به جداول ضمیمه مقدار C را برای احتمالات زیر محاسبه کنید.

$$P\left(F_{(9,10)} > C\right) = 0.05 \Longrightarrow C = 3.02$$

$$P\left(F_{(15,15)} > C\right) = 0.01 \Longrightarrow C = 3.52$$

$$P\left(F_{(8,10)} < C\right) = 0.95 \Longrightarrow P\left(F_{(8,10)} > C\right) = 0.05 \Longrightarrow C = 3.07$$

توزیع نمونه‌ای اختلاف یا مجموع میانگین دو نمونه

✓ فرض کنید از دو جامعه، دو نمونه تصادفی با حجم‌های n_1,n_2 انتخاب کرده و میانگین آنها برابر با $\overline{X_2}$ و $\overline{X_1}$ می‌باشد.

۱- اگر واریانس σ_1^2,σ_2^2 دو جامعه معلوم باشد و دو جامعه نرمال باشند و یا در صورت غیرنرمال بودن دو جامعه $n_1,n_2 > 30$ باشد در اینصورت توزیع نمونه‌ای $\overline{X_1} \pm \overline{X_2}$ برابر است با:

$$Z = \frac{(\overline{X_1} \pm \overline{X_2}) - (\mu_1 \pm \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

توزیع نمونه‌ای اختلاف یا مجموع میانگین دو نمونه

۲- اگر واریانس دو جامعه نامعلوم اما مساوی باشند $\sigma_1 = \sigma_2$.

اگر S_1,S_2 انحراف معیار متعلق به نمونه اول و دوم باشند در اینصورت واریانس مشترک دو نمونه تصادفی به صورت زیر تعریف می‌شود و توزیع نمونه‌ای $\overline{X_1} \pm \overline{X_2}$ به صورت زیر است.

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$T = \frac{(\overline{X_1} \pm \overline{X_2}) - (\mu_1 \pm \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}$$

توزیع نمونه‌ای نسبت نمونه

✓ اگر نمونه‌ای به حجم n از جامعه انتخاب کنیم و X تعداد موفقیت های در نمونه باشد، آنگاه $\bar{P} = \frac{X}{n}$ را نسبت نمونه‌ای می‌نامند. اگر حجم نمونه بزرگ باشد آنگاه توزیع نمونه‌ای $\bar{P}$ برابر خواهد بود با:

$$Z = \frac{\bar{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

توزیع نمونه‌ای تفاضل یا مجموع نسبت دو نمونه

✓ فرض کنید دو متغیر تصادفی به صورت $X_1 \sim B(n_1,p_1)$ و $X_2 \sim B(n_2,p_2)$ باشند خواهیم داشت:

$$\overline{P_1} = \frac{X_1}{n_1} \Rightarrow \begin{cases} E(\overline{P_1}) = p_1 \\ Var(\overline{P_1}) = \dfrac{p_1(1-p_1)}{n_1} \end{cases}$$

$$\overline{P_2} = \frac{X_2}{n_2} \Rightarrow \begin{cases} E(\overline{P_2}) = p_2 \\ Var(\overline{P_2}) = \dfrac{p_2(1-p_2)}{n_2} \end{cases}$$

اگر نمونه به اندازه کافی بزرگ باشد.

$$Z = \frac{(\overline{P_1} \pm \overline{P_2}) - (p_1 \pm p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

نامساوی چبیشف

نامساوی چبیشف

✓ اگر متغیر تصادفی X دارای میانگین μ و واریانس σ^2 باشد آنگاه برای هر $k>0$

$$P[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2} \qquad P(|X-\mu| \leq k) \geq 1-\frac{\sigma^2}{k^2}$$

نتیجه:

اگر متغیر تصادفی X از جامعه‌ای پیوسته با توزیع نامعلوم دارای میانگین μ و واریانس σ^2 باشد آنگاه حداقل $(1-\frac{1}{k^2})$ درصد از مشاهدات در دامنه (فاصله) k انحراف معیار از میانگین قرار دارد$(k \geq 1)$.

حداقل

$$P(|X-\mu| \leq k\sigma) \geq 1-\frac{1}{k^2} \Rightarrow P(\mu-k\sigma \leq X \leq \mu+k\sigma) \geq 1-\frac{1}{k^2}$$

حداکثر

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \Rightarrow P\left(X \leq \mu-k\sigma \text{ یا } X \geq \mu+k\sigma\right) \leq \frac{1}{k^2}$$

نامساوی چبیشف

نامساوی چبیشف

- در صورتی که $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی مستقل nتایی از توزیعی نامعلوم و $\bar{X}$ میانگین نمونه که بجای μ از $\mu_{\bar{X}}$ و بجای X از $\bar{X}$ و بجای واریانس σ^2 از $\frac{\sigma_{\bar{X}}^2}{n}$ استفاده کنیم خواهیم داشت.

$$P(|\bar{X} - \mu_{\bar{X}}| \leq k) \geq 1 - \frac{\sigma_{\bar{X}}^2}{nk^2}$$

$$P(|\bar{X} - \mu_{\bar{X}}| \geq k) \leq \frac{\sigma_{\bar{X}}^2}{nk^2}$$

برآورد نقطه‌ای و فاصله‌ای پارامتر

برآوردگر

- ✓ هر تابعی از نمونه را که به پارامتر یا پارامترهای جامعه بستگی نداشته باشد، برآوردگر یا آماره گویند. چون مقدار برآوردگر از نمونه‌ای به نمونه دیگر تغییر می‌کند، بنابراین متغیری است تصادفی. مقدار عددی آماره یا برآوردگر را برآورد گویند.

ویژگی‌های برآوردگر کارا

۱- نااریب باشد.

۲- دارای کمترین واریانس باشد.

مقدمه

✓ تئوری برآورد به دو دسته تقسیم بندی می‌شود:

۱- برآورد نقطه‌ای: استفاده از داده‌های حاصل از نمونه‌گیری و بدست آوردن عددی که آن را بتوان به عنوان برآورد پارامتر جامعه در نظر گرفت.

۲- برآورد فاصله‌ای (فاصله اطمینان): بدست آمدن یک فاصله احتمالی که احتمال قرار گرفتن پارامتر در این فاصله را بیان می‌کند.

مقدمه

✓ تعریف اول:

هر ویژگی یک جامعه را پارامتر آن جامعه می‌گویند. پارامتر را با θ نشان می‌دهند.

مانند میانگین جامعه $\theta = \mu$ و واریانس جامعه $\theta = \sigma^2$

✓ تعریف دوم:

هر ویژگی متناظر در نمونه را یک آماره می‌نامند. آماره را با $\hat{\theta}$ نشان می‌دهند.

مانند میانگین نمونه $\hat{\theta} = \bar{X}$ و واریانس نمونه $\hat{\theta} = S^2$

برآورد نقطه‌ای

- فرض کنید متغیر تصادفی X دارای تابع چگالی احتمال $f_\theta(x)$ باشد به طوریکه $\theta \in \Theta$. اگر $X_1, X_2, \ldots, X_n$ مشاهدات نمونه باشند برآورد نقطه‌ای برای θ، مقدار عددی برآوردگر براساس نمونه مشاهده شده خواهد بود. اگر فرض کنیم $\bar{x}$ دقیقاً برابر با θ باشد، صفر است. به همین خاطر از $\bar{x}$ به عنوان یک برآورد برای θ یاد می‌کنند.

- برای اینکه فرقی بین θ و برآورد آن قایل شویم، برآورد θ را با $\hat{\theta}$ نمایش می‌دهیم. اختلاف $|\hat{\theta} - \theta|$ را خطای برآورد نقطه‌ای گویند. در ادامه برآورد گشتاورها و روش درستنمایی ماکزیمم ارائه می‌شود.

روش گشتاورها

- ✓ برآورد پارامتر به روش گشتاورها یکی از روش‌های قدیمی است که در سال ۱۸۹۴ توسط کارل پیرسون پیشنهاد شد و هم اکنون در برآورد بیشتر توزیع‌ها قابل استفاده است.
- ✓ فرض کنید متغیر تصادفی X دارای تابع چگالی احتمال $f_\theta(x)$ و $x_1, x_2, \ldots, x_n$ مشاهدات نمونه تصادفی $X_1, X_2, \ldots, X_n$ باشد. برآورد پارامتر θ به روش گشتاورها از برابری گشتاورهای نمونه و جامعه به صورت زیر بدست می‌آید.

گشتاور مرتبه rام جامعه:

$$\mu_r = E(X^r)$$

گشتاور غیر مرکزی مرتبه rام نمونه حول نقطه صفر:

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$$

روش درستنمایی ماکزیمم

- ✓ روش درستنمایی ماکزییم در سال ۱۹۱۲ توسط فیشر ارائه شد و در مواردی که برآورد به روش گشتاورها و روش درستنمایی ماکزیمم یکسان نیستند، برآورد به روش درستنمایی ماکزییم را به روش گشتاورها ترجیح می‌دهند.
- ✓ اگر $X_1, X_2, \ldots, X_n$ مشاهدات متناظر نمونه تصادفی از تابع چگالی احتمال $f_\theta(x)$ ، $\theta \in \Theta$ باشد تابع چگالی توام X_iها برابر است با:

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

روش درستنمایی ماکزیمم

- در رابطه اخیر تنها متغیر θ است و می‌توان رابطه را فقط تابعی از θ نوشت.

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i), \qquad \theta \in \Theta$$

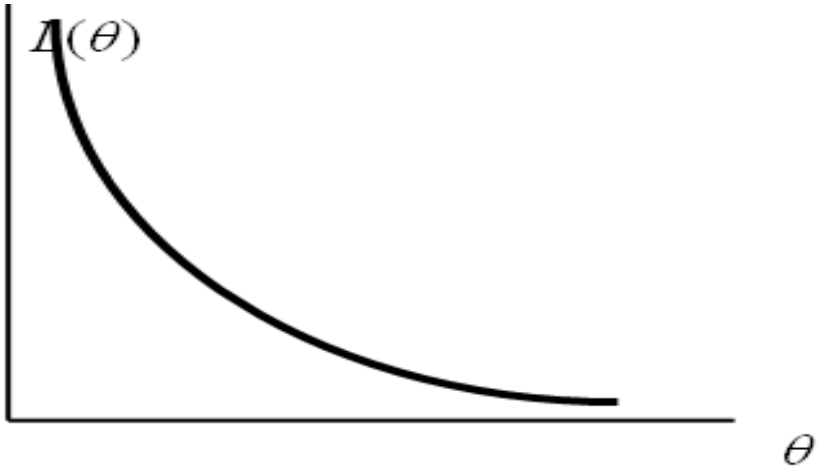
- در ادبیات آماری $L(\theta)$ را **تابع درستنمایی** گویند. هدف روش درستنمایی ماکزیمم این است که θ را طوری پیدا کند که $L(\theta)$ ماکزیمم شود. از آنجا که $L(\theta)$ و $lnL(\theta)$ در مقدار یکسانی از θ ماکزیمم می‌شود برای راحتی θ را طوری پیدا می‌کنیم که $lnL(\theta)$ را ماکزیمم کند.

روش درستنمایی ماکزیمم

✓ مثال: فرض کنید X دارای تابع چگالی احتمال یکنواخت روی بازه $(0,\theta)$ باشد θ را براساس یک نمونه nتایی برآورد کنید.

$$f_\theta(x) = \frac{1}{\theta}, \qquad 0 < x < \theta, \qquad \Theta = (0, \infty)$$

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \frac{1}{\theta^n}$$



✓ با توجه به شیب $L(\theta)$، $L(\theta)$ در هیچ جا صفر نمی‌شود لذا احتیاجی به مشتق گرفتن از $L(\theta)$ و مساوی صفر قرار دادن آن با صفر نیست.

خواص برآوردگرها

1. نااریبی
2. کارایی
3. میانگین توان دوم خطاها (MSE)
4. برآوردگر سازگار

خواص برآوردگرها-نااریبی

✓ نااریبی:

برآورد کننده $\hat{\theta}$ را یک برآورد کننده نااریب برای θ گویند، اگر و تنها اگر $E(\hat{\theta}) = \theta$ و اگر $E(\hat{\theta}) \neq \theta$ باشد $\hat{\theta}$ را یک برآوردگر اریب گویند و مقدار اریبی برابر است با:

$$\text{مقدار اریبی} = E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta$$

خواص برآوردگرها-کارایی

کارایی:

شرط کارایی آماره‌ها است، به شرط برابری میانگین‌های دو آماره، آماره‌ای بهتر است که واریانس کمتری داشته باشد.

اگر $\widehat{\theta_1}$ و $\widehat{\theta_2}$ دو برآورد ناریب برای θ باشند $\widehat{\theta_1}$ کاراتر از $\widehat{\theta_2}$ است هرگاه:

$$Var(\widehat{\theta_1}) < Var(\widehat{\theta_2})$$

و به مقدار

$$\frac{Var(\widehat{\theta_2})}{Var(\widehat{\theta_1})}$$

کارایی نسبی $\widehat{\theta_1}$ به $\widehat{\theta_2}$ گویند.

خواص برآوردگرها-میانگین توان دوم خطاها

میانگین توان دوم خطاها:

مهمترین و بهترین معیار برای مقایسه برآورد کننده‌ها، میانگین توان دوم خطاهاست چرا که همزمان واریانس و مقدار اریبی برآوردگر را محاسبه می‌کند.

$$MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right] = Var(\hat{\theta}) + \left(E(\hat{\theta}) - \theta\right)^2$$

واریانس مقدار اریبی

خواص برآوردگرها-برآوردگر سازگار

برآوردگر سازگار:

برآوردگر $\hat{\theta}$ را یک برآوردگر سازگار برای پارامتر θ گویند هرگاه:

$$\lim_{n\to\infty} MSE\left(\hat{\theta}\right) = 0$$

به عبارت دیگر آماره $\hat{\theta}$ سازگار است اگر با افزایش حجم نمونه میانگین توان دوم خطاها به سمت صفر میل نماید.

برآورد فاصله‌ای

- در بخش برآورد نقطه‌ای اشاره به این حقیقت شد که برآورد نقطه‌ای نمی‌تواند برابر با مقدار واقعی پارامتر باشد. به عنوان مثال فرض کنید مدیر کارخانه‌ای ادعا می‌کند که لامپ‌های تولیدی این کارخانه دارای عمر متوسط بین ۱۵۰۰ ± ۱۰ ساعت است.

- اگر براساس یک نمونه nتایی برآورد نقطه‌ای برای ادعای مدیر داشته باشیم مسلماً برآورد ما یک نقطه از بازه $(1500 - 10, 1500 + 10)$ خواهد بود که به خودی خود متضمن اطلاعاتی درباره میزان احتمال اینکه برآوردگر مقداری نزدیک به مقدار واقعی مجهول قبول کند، نیست.

برآورد فاصله‌ای

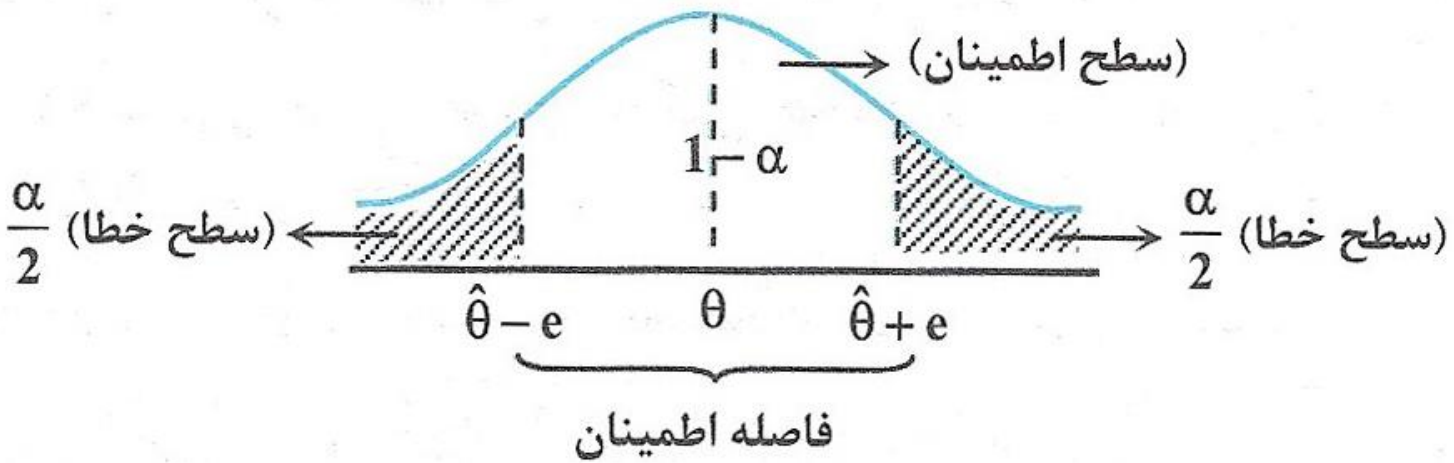
- ✓ تعریف: اگر $X_1, X_2, \dots, X_n$ یک نمونه تصادفی از تابع چگالی احتمال $f_\theta(x)$ باشد و $T_1(X_1, \dots, X_n)$ و $T_2(X_1, \dots, X_n)$ دو آماره باشد به طوری که $T_1 < T_2$ بازه (T_1, T_2) را یک بازه اطمیان با ضریب اطمینان $\%100(1-\alpha)$ برای θ گوییم. اگر احتمال اینکه دو آماره T_1 و T_2، θ در برداشته باشد مستقل از θ باشد.

$$P[T_1(X_1, \dots, X_n) < \theta < T_2(X_1, \dots, X_n)] = 1 - \alpha$$

- ✓ T_1 و T_2 را به ترتیب حدود اطمینان پایینی و بالایی θ و $T_2 - T_1$ را طول بازه اطمینان گویند.
- ✓ برای بدست آوردن بازه اطمینان برای θ یا تابعی از θ، $\tau(\theta)$ لازم است کمیت محوری تعریف کنیم.

برآورد فاصله‌ای

$$\boxed{P(\hat{\theta}-e<\theta<\hat{\theta}+e)=1-\alpha}$$



- ✓ با احتمال $(1-\alpha)$ اطمینان داریم که فاصله $(\hat{\theta}-e,\hat{\theta}+e)$ پارامتر θ را در بر می‌گیرد و با احتمال $\frac{\alpha}{2}+\frac{\alpha}{2}=\alpha$ احتمال می‌دهیم که پارامتر θ در فاصله $(\hat{\theta}-e,\hat{\theta}+e)$ قرار نگیرد.
- ✓ همچنین می‌خواهیم حداکثر به اندازه مقدار e برآورد θ یعنی $\hat{\theta}$ از پارامتر θ تفاوت داشته باشد.
- ✓ بنابراین $1-\alpha$ سطح اطمینان و α سطح خطا و خطای برآورد $e=|\hat{\theta}-\theta|$ می‌باشد.

برآورد فاصله‌ای

✓ در مسائل آمار استنباطی معمولاً با سه سطح خطا و سه سطح اطمینان بیشتر از بقیه سر و کار داریم. برای توزیع نرمال این سه سطح به شرح زیر است.

سطح اطمینان	سطح خطا	ضریب اطمینان
$(1-\alpha)=0/90$	$\alpha=0/1 \quad ; \quad \frac{\alpha}{2}=\frac{0/1}{2}=0/05$	$Z_{\alpha}=1/28 \quad ; \quad Z_{\frac{\alpha}{2}}=1/645$
$(1-\alpha)=0/95$	$\alpha=0/05 \quad ; \quad \frac{\alpha}{2}=\frac{0/05}{2}=0/025$	$Z_{\alpha}=1/645 \quad ; \quad Z_{\frac{\alpha}{2}}=1/96$
$(1-\alpha)=0/99$	$\alpha=0/01 \quad ; \quad \frac{\alpha}{2}=\frac{0/01}{2}=0/005$	$Z_{\alpha}=2/32 \quad ; \quad Z_{\frac{\alpha}{2}}=2/58$

افزایش ↓ کاهش ↓ افزایش ↓

برآورد فاصله‌ای

✓ تعریف: هر تابعی از نمونه و پارامتر را که توزیع آن مستقل از پارمتر باشد کمیت محوری گویند. به عنوان مثال $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ برای نمونه nتایی از جامعه نرمال با میانگین μ و واریانس معلوم σ^2 یک کمیت محوری است. چون $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد است. از آنجا که برای یافتن فاصله اطمینان نیاز به داشتن آمار است، دانستن توزیع جامعه‌ای که نمونه از آن استخراج شده است، ضرروی است. فرض کنید متغیر تصادفی Z دارای توزیع نرمال استاندارد باشد. احتمال اینکه دو عدد ۱/۹۶ و ۱/۹۶- متغیر تصادفی Z را در برداشته باشند، برابر با ۰/۹۵ است.

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

فاصله اطمینان برای میانگین توزیع نرمال

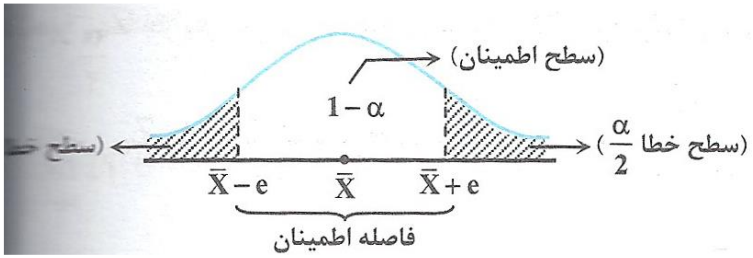
- ✓ حالت اول: اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد، براساس یک نمونه nتایی یک فاصله اطمینان، با ضریب اطمینان $\%100(1-\alpha)$ برای μ به صورت زیر بدست می‌آید.
- ✓ در این حالت می‌دانیم که متغیر $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ یک کمیت محوری و دارای توزیع نرمال استاندارد است. لذا:

$$P\left[-Z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right] = 1-\alpha$$

فاصله اطمینان برای میانگین توزیع نرمال

✓ یعنی احتمال رخداد پیشامد $-Z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}$ برابر با $1-\alpha$ است. اگر پیشامد فوق را نسبت به μ حل کنیم داریم:

$$\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$



✓ برای یک نمونه nتایی بازه اطمینان برای μ برابر است با:

$$\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

برای $\alpha = 0.05$

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

فاصله اطمینان برای میانگین توزیع نرمال

✓ حالت دوم: جامعه نرمال و واریانس σ^2 نامعلوم باشد.

براساس یک نمونه nتایی یک فاصله اطمینان، با ضریب اطمینان $\%100(1-\alpha)$ برای μ به صورت زیر بدست می‌آید.

$$P\left[-t_{\frac{\alpha}{2},n-1} < \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} < t_{\frac{\alpha}{2},n-1}\right] = 1-\alpha$$

الف: $n<30$

$$\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}$$

ب: $n>30$

$$\bar{X} - Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$$

فاصله اطمینان برای تقاضل/جمع میانگین دو جامعه

- فرض کنید $X_1, X_2, \ldots, X_m$ یک نمونه mتایی از جامعه‌ای نرمال با میانگین μ_1 و واریانس σ_1^2 و $Y_1, Y_2, \ldots, Y_n$ یک نمونه nتایی از جامعه دیگر نرمال با میانگین μ_2 و واریانس σ_2^2 باشند. می‌دانیم که $\bar{X}, \bar{Y}$ به ترتیب برآوردگرهای μ_1 و μ_2 می‌باشد. لذا هرگونه استناط روی μ_1 و μ_2 براساس $\bar{X}, \bar{Y}$ خواهد بود. در این بخش هدف، برآورد فاصله اطمینان برای $\mu_1 \pm \mu_2$ است. باتوجه به خواص برآوردگرها، $\bar{X} \pm \bar{Y}$ یک برآوردگر برای $\mu_1 \pm \mu_2$ است. چون $\bar{X}, \bar{Y}$ تک تک دارای توزیع نرمال هستند تفاضل/جمع آنها نیز نرمال است که میانگین و واریانس آن به صورت زیر محاسبه می شود.

$$E(\bar{Y} \pm \bar{X}) = E(\bar{Y}) \pm E(\bar{X}) = \mu_2 \pm \mu_1$$

$$V(\bar{Y} \pm \bar{X}) = V(\bar{Y}) + V(\bar{X}) = \frac{\sigma_2^2}{n} + \frac{\sigma_1^2}{m}$$

فاصله اطمینان برای تفاضل/جمع میانگین دو جامعه

✓ حالت اول: پس $\bar{Y}\pm\bar{X}$ دارای توزیع نرمال با میانگین $\mu_2 \pm \mu_1$ و واریانس $\frac{\sigma_2^2}{n} + \frac{\sigma_1^2}{m}$ است. اگر متغیر Z ، متغیر استاندارد شده $\bar{Y}\pm\bar{X}$ باشد، Z برابر خواهد بود با:

$$Z = \frac{\bar{Y} \pm \bar{X} - (\mu_2 \pm \mu_1)}{\sqrt{\frac{\sigma_2^2}{n} + \frac{\sigma_1^2}{m}}}$$

✓ چون Z نرمال استاندارد است لذا داریم:

$$P\left[-Z_{\frac{\alpha}{2}} < \frac{\bar{Y} \pm \bar{X} - (\mu_2 \pm \mu_1)}{\sqrt{\frac{\sigma_2^2}{n} + \frac{\sigma_1^2}{m}}} < Z_{\frac{\alpha}{2}}\right] = 1 - \alpha$$

✓ از رابطه اخیر داریم:

$$(\bar{Y} \pm \bar{X}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_2^2}{n} + \frac{\sigma_1^2}{m}} < \mu_2 \pm \mu_1 < (\bar{Y} \pm \bar{X}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_2^2}{n} + \frac{\sigma_1^2}{m}}$$

فاصله اطمینان برای تفاضل/جمع میانگین دو جامعه

✓ این رابطه یک فاصله اطمینان $\%100(1-\alpha)$ برای $\mu_2 \pm \mu_1$ است. اگر دو جامعه دارای واریانس مشترک σ^2 و حجم نمونه‌ها مساوی n باشند فاصله اطمینان برای $\mu_2 \pm \mu_1$ برابر است با:

$$(\bar{Y} \pm \bar{X}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{2\sigma^2}{n}} < \mu_2 \pm \mu_1 < (\bar{Y} \pm \bar{X}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{2\sigma^2}{n}}$$

فاصله اطمینان برای تفاضل/جمع میانگین دو جامعه

✓ حالت دوم: دو جامعه نرمال و واریانس‌ها نامعلوم

الف: اگر $n,m > 30$

$$(\bar{Y} \pm \bar{X}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{S_2^2}{n} + \frac{S_1^2}{m}} < \mu_2 \pm \mu_1 < (\bar{Y} \pm \bar{X}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{S_2^2}{n} + \frac{S_1^2}{m}}$$

ب: اگر $n,m \leq 30$ واریانس‌های نامعلوم دو جامعه برابر است.

$$(\bar{Y} \pm \bar{X}) - t_{\frac{\alpha}{2},n+m-2} S_p\sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_2 \pm \mu_1 < (\bar{Y} \pm \bar{X}) + t_{\frac{\alpha}{2},n+m-2} S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{n+m-2}$$

فاصله اطمینان برای تفاضل/جمع میانگین دو جامعه

✓ حالت دوم: دو جامعه نرمال و واریانس‌ها نامعلوم

ج: اگر $n,m \leq 30$ واریانس‌های نامعلوم دو جامعه نابرابر است.

$$(\bar{Y} \pm \bar{X}) - t_{df,\frac{\alpha}{2}}\sqrt{\frac{S_2^2}{n}+\frac{S_1^2}{m}} < \mu_2 \pm \mu_1 < (\bar{Y} \pm \bar{X}) + t_{df,\frac{\alpha}{2}}\sqrt{\frac{S_2^2}{n}+\frac{S_1^2}{m}}$$

$$df = r = \frac{\left(\frac{S_2^2}{n}+\frac{S_1^2}{m}\right)^2}{\frac{\left(\frac{S_1^2}{m}\right)^2}{m-1}+\frac{\left(\frac{S_2^2}{n}\right)^2}{n-1}}$$

فاصله اطمینان برای واریانس جامعه

✓ حالت اول: فرض کنید متغیر تصادفی X دارای توزیع نرمال با میانگین نامعلوم μ و واریانس σ^2 باشد. اگر $X_1, X_2, \ldots, X_n$ یک نمونه از این جامعه باشد، طبق بخش‌های قبل $\frac{(n-1)S^2}{\sigma^2}$ دارای توزیع کی دو با $n-1$ درجه آزادی است. برای یافتن فاصله اطمینان می‌توان از $\frac{(n-1)S^2}{\sigma^2}$ بعنوان یک کمیت محوری استفاده کرد.

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

فاصله اطمینان برای واریانس جامعه

- ✓ که اعداد $\chi^2_{[\frac{\alpha}{2},n-1]}$ و $\chi^2_{[1-\frac{\alpha}{2},n-1]}$ از جداول کی دو قابل محاسبه است.

از پیشامد احتمال اخیر می‌توان σ^2 را برحسب S^2 به صورت زیر بدست آورد.

$$\frac{(n-1)S^2}{\chi^2_{[\frac{\alpha}{2},n-1]}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{[1-\frac{\alpha}{2},n-1]}}$$

- ✓ رابطه اخیر یک فاصله اطمینان $\%100(1-\alpha)$ برای σ^2 است. فاصله اطمینان $\%100(1-\alpha)$ برای σ یا انحراف معیار برابر است با:

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{[\frac{\alpha}{2},n-1]}}} \leq \sigma \leq \sqrt{\frac{(n-1)S^2}{\chi^2_{[1-\frac{\alpha}{2},n-1]}}}$$

فاصله اطمینان برای واریانس جامعه

✓ حالت دوم: فرض کنید متغیر تصادفی X دارای توزیع نرمال با میانگین معلوم μ و واریانس σ^2 باشد. اگر $X_1, X_2, \ldots, X_n$ یک نمونه از این جامعه باشد، طبق بخش‌های قبل $\frac{nS^2}{\sigma^2}$ دارای توزیع کی دو با n درجه آزادی است. برای یافتن فاصله اطمینان می‌توان از $\frac{nS^2}{\sigma^2}$ بعنوان یک کمیت محوری استفاده کرد.

$$P\left[\chi^2_{[1-\frac{\alpha}{2},n]} < \frac{nS^2}{\sigma^2} < \chi^2_{[\frac{\alpha}{2},n]}\right] = 1-\alpha$$

فاصله اطمینان برای واریانس جامعه

- ✓ که اعداد $\chi^2_{[\frac{\alpha}{2},n]}$ و $\chi^2_{[1-\frac{\alpha}{2},n]}$ از جداول کی دو قابل محاسبه است. از پیشامد احتمال اخیر می‌توان σ^2 را برحسب S^2 به صورت زیر بدست آورد.

$$\frac{nS^2}{\chi^2_{[\frac{\alpha}{2},n]}} \leq \sigma^2 \leq \frac{nS^2}{\chi^2_{[1-\frac{\alpha}{2},n]}}$$

- ✓ رابطه اخیر یک فاصله اطمینان $\%100(1-\alpha)$ برای σ^2 است. فاصله اطمینان $\%100(1-\alpha)$ برای σ یا انحراف معیار برابر است با:

$$\sqrt{\frac{nS^2}{\chi^2_{[\frac{\alpha}{2},n]}}} \leq \sigma \leq \sqrt{\frac{nS^2}{\chi^2_{[1-\frac{\alpha}{2},n]}}}$$

فاصله اطمینان برای نسبت دو واریانس

✓ حالت اول: فرض کنید $X_1, X_2, \ldots, X_m$ یک نمونه mتایی از جامعه‌ای نرمال با میانگین نامعلوم μ_1 و واریانس σ_1^2 و $Y_1, Y_2, \ldots, Y_n$ یک نمونه nتایی از جامعه دیگر نرمال با میانگین نامعلوم μ_2 و واریانس σ_2^2 باشند. در بخش قبل دیدیم که $F = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2}$ دارای توزیع فیشر با $m-1$ و $n-1$ درجه آزادی است. متغیر F یک کمیت محوری است چون توزیع آن مستقل از σ_1^2 و σ_2^2 است. لذا:

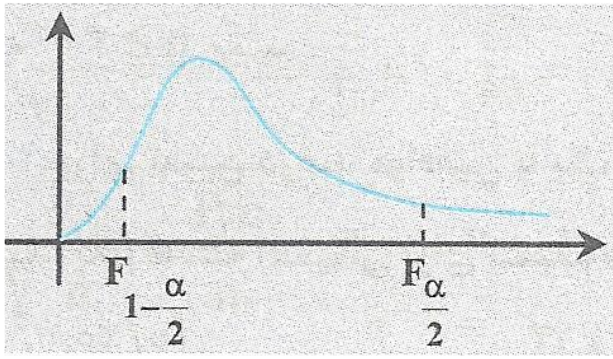
$$P\left[f_{(1-\frac{\alpha}{2},m-1,n-1)} < F < f_{(\frac{\alpha}{2},m-1,n-1)}\right] = 1-\alpha$$

$$P\left[f_{(1-\frac{\alpha}{2},m-1,n-1)} < \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2} < f_{(\frac{\alpha}{2},m-1,n-1)}\right] = 1-\alpha$$

فاصله اطمینان برای نسبت دو واریانس

✓ که در $f_{(\frac{\alpha}{2},m-1,n-1)}$ و $f_{(1-\frac{\alpha}{2},m-1,n-1)}$ از جداول توزیع فیشر بدست می‌آید. اگر $S_1^2=\frac{1}{m-1}\sum(x_i-\bar{x})^2$ و $S_2^2=\frac{1}{n-1}\sum(y_i-\bar{y})^2$ به ترتیب واریانس‌های نمونه اول و دوم باشند فاصله اطمینان برای $\frac{\sigma_1^2}{\sigma_2^2}$ از رابطه زیر بدست می‌آید.

$$\frac{1}{f_{(\frac{\alpha}{2},m-1,n-1)}}\frac{S_1^2}{S_2^2}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{1}{f_{(1-\frac{\alpha}{2},m-1,n-1)}}\frac{S_1^2}{S_2^2}$$



$$F_{\frac{\alpha}{۲},(n_۱-۱,n_۲-۱)}=\frac{۱}{F_{۱-\frac{\alpha}{۲},n_۲-۱,n_۱-۱}}$$

فاصله اطمینان برای نسبت دو واریانس

✓ حالت دوم: فرض کنید $X_1, X_2, \ldots, X_m$ یک نمونه mتایی از جامعه‌ای نرمال با میانگین معلوم μ_1 و واریانس σ_1^2 و $Y_1, Y_2, \ldots, Y_n$ یک نمونه nتایی از جامعه دیگر نرمال با میانگین معلوم μ_2 و واریانس σ_2^2 باشند. در بخش قبل دیدیم که $F = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2}$ دارای توزیع فیشر با m و n درجه آزادی است. متغیر F یک کمیت محوری است چون توزیع آن مستقل از σ_1^2 و σ_2^2 است. لذا:

$$P\left[f_{(1-\frac{\alpha}{2},m,n)} < F < f_{(\frac{\alpha}{2},m,n)}\right] = 1 - \alpha$$

$$P\left[f_{(1-\frac{\alpha}{2},m,n)} < \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2} < f_{(\frac{\alpha}{2},m,n)}\right] = 1 - \alpha$$

فاصله اطمینان برای نسبت دو واریانس

✓ که در $f_{(\frac{\alpha}{2},m,n)}$ و $f_{(1-\frac{\alpha}{2},m,n)}$ از جداول توزیع فیشر بدست می‌آید.

اگر $S_1^2 = \frac{1}{m}\sum(x_i - \bar{x})^2$ و $S_2^2 = \frac{1}{n}\sum(y_i - \bar{y})^2$ به ترتیب واریانس‌های نمونه اول و دوم باشند فاصله اطمینان برای $\frac{\sigma_1^2}{\sigma_2^2}$ از رابطه زیر بدست می‌آید.

$$\frac{1}{f_{(\frac{\alpha}{2},m,n)}}\frac{S_1^2}{S_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{f_{(1-\frac{\alpha}{2},m,n)}}\frac{S_1^2}{S_2^2}$$

فاصله اطمینان برای نسبت دو واریانس

✓ مثال: نمرات زیر نمونه‌ای از نمرات برنامه نویس در دو گروه ۱و۲ می‌باشد. اگر فرض نرمال بودن نمرات در دو گروه پذیرفته شود یک فاصله اطمینان ۹۰٪ برای نسبت واریانس دو جامعه بدست آورید.

گروه اول	12	10	14	13	11		
گروه دوم	17	15	14	16	17	17	16

✓ جواب:

$$m = 5 \quad , \quad n = 7$$

$$S_1^2 = \frac{1}{m-1}\sum (x_i - \bar{x})^2 = 2.5$$

$$S_2^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2 = 1.33$$

$$\alpha = 0.1 \quad , \quad \frac{\alpha}{2} = 0.05 \quad , \quad 1 - \frac{\alpha}{2} = 0.95$$

فاصله اطمینان برای نسبت دو واریانس

✓ جواب:

$$f_{(\frac{\alpha}{2},\, m-1,\, n-1)} = f_{(0.05,4,6)} = 4.53$$

$$f_{(1-\frac{\alpha}{2},\, m-1,\, n-1)} = f_{(0.95,4,6)} = \frac{1}{f_{(0.95,6,4)}} = \frac{1}{6.16} = 0.162$$

$$\frac{1}{f_{(\frac{\alpha}{2},m-1,n-1)}}\frac{S_1^2}{S_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{f_{(1-\frac{\alpha}{2},m-1,n-1)}}\frac{S_1^2}{S_2^2}$$

$$\frac{1}{4.53}\frac{2.5}{1.33} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{0.162}\frac{2.5}{1.33}$$

$$0.415 \le \frac{\sigma_1^2}{\sigma_2^2} \le 11.603$$

فاصله اطمینان برای p در توزیع دوجمله‌ای-نسبت جامعه

✓ اگر متغییر تصادفی X دارای توزیع دوجمله‌ای با پارامتر مجهول p و n معلوم باشد، $E(X) = np,\ V(X) = np(1-p)$ است.

✓ از رابطه $E(X) = np$ داریم:

$$E\left(\frac{X}{n}\right) = p \qquad V\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}$$

✓ متغیر تصادفی X مجموع چند متغیر تصادفی برنولی است، $\frac{X}{n}$ یک برآوردگر برای p جامعه است، لذا :

$$\hat{p} = \frac{X}{n} \qquad E(\hat{p}) = p \qquad V(\hat{p}) = \frac{p(1-p)}{n}$$

فاصله اطمینان برای p در توزیع دوجمله‌ای-نسبت جامعه

✓ با توجه به رابطه اخیر نمونه از حدی بزرگتر باشد، کمیت محوری زیر دارای توزیع تقریبی نرمال استاندارد خواهد بود.

$$Z = \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

✓ از رابطه اخیر:

$$P\left[-Z_{\frac{\alpha}{2}} < \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} < Z_{\frac{\alpha}{2}}\right] \approx 1 - \alpha$$

$$\frac{X}{n} - Z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} < p < \frac{X}{n} + Z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$$

فاصله اطمینان برای p در توزیع دوجمله‌ای-نسبت جامعه

✓ اگر X مقدار مشاهده X باشد، فاصله اطمینان اخیر قابل ارزیابی نیست چون واریانس $\frac{X}{n}$ شامل پارامتر P است. چون $\hat{p} = \frac{X}{n}$ یک برآورد نقطه‌ای برای p است. برآورد نقطه‌ای واریانس $\frac{X}{p}$ برابر است با:

$$\frac{\hat{p}(1-\hat{p})}{n} = \frac{\frac{X}{n}(1-\frac{X}{n})}{n}$$

✓ فاصله اطمینان تقریبی برای P برابر است با:

$$\frac{X}{n} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{X}{n}(1-\frac{X}{n})}{n}} < p < \frac{X}{n} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{X}{n}(1-\frac{X}{n})}{n}}$$

فاصله اطمینان برای p در توزیع دوجمله‌ای-نسبت دو جامعه

✓ اگر p_1 و p_2 نسبت دو جامعه و $\overline{P_1}$ و $\overline{P_2}$ نسبت در نمونه باشد، یک فاصله اطمینان $\%100(1-\alpha)$ برای تفاضل و مجموع نسبت دو جامعه برابر است با:

$$\overline{P_1} \pm \overline{P_2} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{P_1}(1-\overline{P_1})}{m} + \frac{\overline{P_2}(1-\overline{P_2})}{n}} < p_1 \pm p_2 < \overline{P_1} \pm \overline{P_2} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{P_1}(1-\overline{P_1})}{m} + \frac{\overline{P_2}(1-\overline{P_2})}{n}}$$

✓ به شرط آنکه $m,n \geq 30$

برآورد تعداد نمونه

- ✓ یکی از مهمترین مبحث‌های آماری، تعیین تعداد نمونه است، اگر تعداد نمونه درست درنظر گرفته نشود، نتایج بدست آمده دقت کافی را نداشته و نمی‌توان براساس آن نتیجه‌گیری و استنباط کرد.
- ✓ تعیین تعداد نمونه برای میانگین جامعه

الف: واریانس جامعه σ^2 معلوم

$$n \geq \left(\frac{\sigma}{e} \times Z_{\frac{\alpha}{2}}\right)^2$$

برآورد تعداد نمونه

✓ مثال: مدیر کیفیت کارخانه‌ای می‌خواهد برآورد میانگین وزن محصولات کارخانه را بدست آورد. او نمونه‌ای انتخاب می‌کند. اگر میانگین ۵۰ گرم و انحراف معیار ۱۲ گرم باشد و دقت برآورد ۱/۴ ± در نظر گرفته شود، حجم نمونه مناسب کدام است؟ همچنین $\alpha = 1\%$ می‌باشد.

✓ جواب:

$$\alpha = 0.01 \Longrightarrow \frac{\alpha}{2} = 0.005, \qquad Z_{0.005} = 2.57$$

$$n \geq \left(\frac{\sigma}{e} \times Z_{\frac{\alpha}{2}}\right)^2 = \left(\frac{12}{1.4} \times 2.57\right)^2 = 486$$

برآورد تعداد نمونه

✓ تعیین تعداد نمونه برای میانگین جامعه

ب: واریانس جامعه σ^2 نامعلوم

۱-اگر $n_\circ > 30$ باشد تعداد نمونه مورد نظر است.

$$n_\circ = \left(\frac{S}{e} \times Z_{\frac{\alpha}{2}}\right)^2$$

۲-اگر $n_\circ \leq 30$ باشد، آنگاه:

مرحله اول: ابتدا n_1 را محاسبه می‌کنید

$$n_1 = \left(\frac{S}{e} \times t_{\frac{\alpha}{2}, n_\circ - 1}\right)^2$$

در صورتی که $n_\circ \geq n_1$ باشد $n_\circ$ همان تعداد مورد نظر است.

مرحله دوم: در صورتی که $n_\circ < n_1$ باشد $\widehat{n_1} = n_\circ + 1$

$$n_2 = \left(\frac{S}{e} \times t_{\frac{\alpha}{2}, \widehat{n_1} - 1}\right)^2$$

اگر $n_1 \geq n_2$ باشد، n_1 همان تعداد مورد نظر است.

مرحله سوم: در صورتی که $n_1 \geq n_2$ نباشد $\widehat{n_2} = n_\circ + 2$

$$n_3 = \left(\frac{S}{e} \times t_{\frac{\alpha}{2}, \widehat{n_2} - 1}\right)^2$$

اگر $n_2 \geq n_3$ باشد، n_2 همان تعداد مورد نظر است.

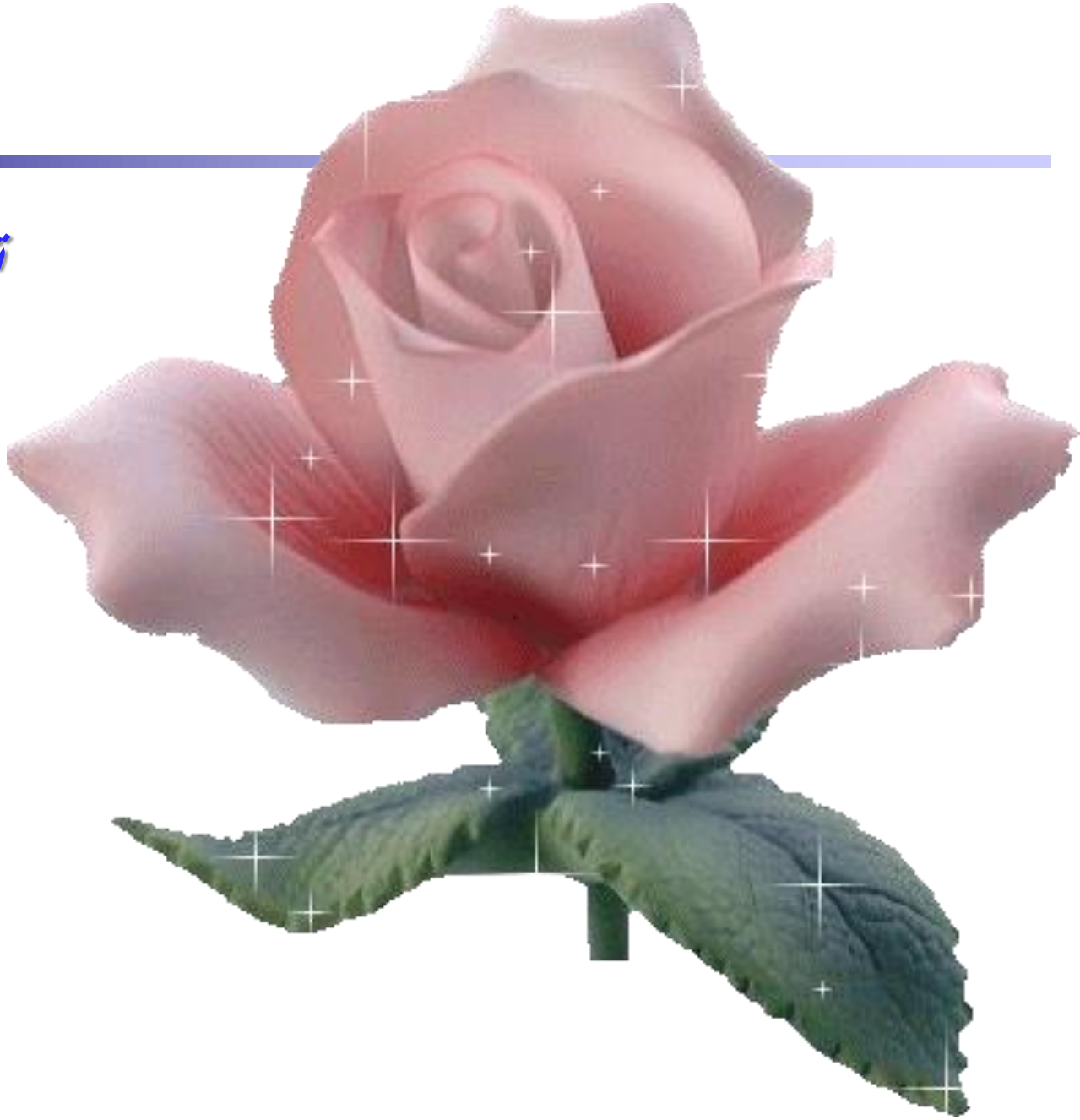
مرحله چهارم: روش فوق را تا برقرار شرط تکرار می‌کنیم.

برآورد تعداد نمونه

- ✓ تعیین تعداد نمونه برای نسبت جامعه

$$n \geq \bar{P}(1-\bar{P})\left(\frac{Z_{\frac{\alpha}{2}}}{e}\right)^2$$

- ✓ توجه: در صورتی که برآورد نسبت جامعه معلوم نباشد ($\bar{P}$)، بهترین مقدار برای آن عدد نیم ۰/۵ می‌باشد.

$$n \geq \frac{1}{4}\left(\frac{Z_{\frac{\alpha}{2}}}{e}\right)^2$$

تشکر از توجه شما