خلاصه درس آمار 2 به همره فرمولها

خلاصه درس آمار ۲

از آمار ۱ به خاطر داریم که با یک آزمایش که درآن یک کمیت خاص را (مقدار کالا های معیوب ، قد افراد و ...) مشاهده می‌کنیم یک متغیر تصادفی X به آن مربوط است که توزیع احتمال آن با تابع توزیع $F(x) = p(X \leq x)$ داده می‌شود که برای هر x احتمالی داده می‌شود که X مقادیری را بگیرد که بیش از x نباشد.

در آمار ما نمونه های تصادفی $x_1 \dots x_n$ (با سایزn) را می‌گیریم. با انجام آن آزمایش nبار و آنگاه از ویژگی های آن نمونه ، استنباطهایی درباره توزیع متغیر تصادفی X داشته باشیم . ما این را با برآورد نقطه ای ، یا فاصله اطمینان یا با اعمال یک آزمون برای پارامتر های (μ و σ^2 درتوزیع‌نرمال ، توزیع دو جمله ای و...) یا با یک آزمون برای توابع توزیع انجام می‌دهیم.

یک برآورد نقطه ای تقریبی از یک پارامتر توزیع X است که از یک نمونه می‌آید .

مثلا میانگین نمونه :

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} xj = \frac{1}{n}(x_1 + \cdots + x_n)$$

تخمین از μ متغیر X است و واریانس نمونه ی

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(xj - \bar{x})^2 = \frac{1}{n-1}((x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2)$$

تخمین از واریانس σ^2 متغیر X است. برآورد نقطه ای می‌تواند با روش ماکسیمم درستی M.L.E انجام شود.

فاصله اطمینان $\theta_1 \leq \theta \leq \theta_2$ که از یک نمونه حاصل شده است طوری که ما با یک احتمال بالای γ بازه ای به دست می‌آوریم مشتمل بر مقدار صحیح با پارامتر نامعلوم θ از توزیع X .

اینجا γ در ابتدا انتخاب شده است و معمولا بین 95٪ تا 99٪ . ما چنین بازه ای را با $CONF_{\gamma}\{\theta_1 \leq \theta \leq \theta_2\}$ نمایش می‌دهیم.

در یک آزمون برای یک پارامتر ما فرض $\theta = \theta_0$ را در برابر فرض جایگزین $\theta = \theta_1$ می‌آزماییم و آنگاه بر مبنای یک نمونه، فرض را می‌پذیریم یا آن را به نفع فرض جایگزین رد می‌کنیم . مانند هر استنباطی از نمونه ها درباره X این ممکن است مشتمل بر خطاهایی باشد که به تصمیم غلط منجر شود. اینجا یک احتمال کوچک α هست مثلا 5٪ یا 1 ٪ که یک فرض درست را رد کنیم و یک احتمال β هست که می‌توانیم با گرفتن یک نمونه بزرگتر کاهش اش دهیم که یک فرض غلط را بپذیریم .

α را سطح اطمینان نامند و $1 - \beta$ را قدرت آزمون نامند . در میان سایر کاربردهای مهندسی ، آزمودن در کنترل کیفیت و نمونه گیری قابل قبول کاربرد دارند .

نه تنها یک پارامتر از X بلکه نوع توزیع X مجهول است ، ما می‌توانیم آزمون مربع کای را برای آزمودن این فرضیه به کار بریم که تابع $F(x)$ تابع توزیع مجهول X است . این با معلوم کردن عدم تطابق بین $F(x)$ و تابع توزیع $\tilde{F}(x)$ از یک نمونه داده شده ، محاسبه می‌شود .

آزمونهای مستقل از توزیع یا آزمون های ناپارامتریک ، آزمون هایی هستند که اعمال می‌شوند به هر توزیعی ، از آنجا که آن ها مبتنی هستند بر ایده های ترکیباتی. این آزمون ها اغلب خیلی ساده هستند .

آخرین بخش مرتبط است با زوج هایی از مقادیر ، که بر می‌خیزند از یک آزمایشی وقتی ما همزمان دو کمیت را مشاهده می‌کنیم . در آنالیز رگرسیون ، یکی از کمیت ها x ، یک مقدار معمولی است و دیگری y ، یک متغیر تصادفی است که میانگین آن μ وابسته به x است ، بگویید: $\mu(x) = k_0 + k_1 x$

در آنالیز همبستگی ارتباط بین x و y در یک متغیر تصادفی دو بعدی (x,y) مورد پژوهش قرار می‌گیرد ، به طور قابل توجه بر حسب ضریب همبستگی ρ .

برخی توزیع‌های احتمال

توزیع احتمال	تعریف	تابع احتمال
برنولی	آزمایشی که دارای دو نتیجه موفقیت با احتمال p و شکست با احتمال $q=1-p$ باشد	$f_X(x)=p^x(1-p)^{1-x}$
دو جمله‌ای	آزمایشی که از انجام چند آزمایش مستقل برنولی به وجود می‌آید	$f_X(x)=\binom{n}{x}p^x(1-p)^{n-x} \quad , \quad x=0,1,2,\ldots,n$
فوق هندسی	اگر آزمایشات برنولی تکرار شده از یکدیگر مستقل نباشند	$f_X(x)=\dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$
پواسون	آزمایشی که تعداد موفقیت‌ها در یک فاصله زمانی یا ناحیه مشخص را به دست دهد	$f_X(x)=\dfrac{e^{-\mu}\mu^x}{x!}$
دو جمله‌ای منفی	آزمایش مستقل برنولی را آنقدر تکرار کنیم تا به یک تعداد معینی از موفقیت دست یابیم	$f_X(x)=\binom{x-1}{r-1}p^r q^{x-r} \quad , \quad x=r,r+1,r+2,\ldots$
هندسی	تعداد آزمایشات مستقل برنولی تا رسیدن به یک موفقیت	$f_X(x)=p\ q^{x-1} \quad , \quad x=1,2,3,\ldots$
توزیع یکنواخت گسسته	متغیر تصادفی گسسته X تمام مقادیرش را با احتمالات یکسان اختیار کند	$f_X(x)=P(X=x)=\dfrac{1}{K} \quad , \quad x=x_1,x_2,\ldots,x_k$
توزیع یکنواخت پیوسته	متغیر تصادفی X در فاصله (a,b) دارای توزیع یکنواخت است ، هرگاه دارای تابع چگالی احتمال روبرو باشد	$f_X(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b \\ 0 & \text{سایر نقاط}\end{cases}$

توزیع نمایی	متغیر تصادفی X بیشتر نمایانگر طول عمر یک قطعه می‌باشد.	$f_X(x)=\begin{cases}\frac{1}{\theta}e^{-x/\theta} & x>0,\quad \theta>0\\ 0 & \text{سایر نقاط}\end{cases}$
توزیع نرمال	مهمترین توزیع در آمار و احتمال است و اغلب متغیرهای تصادفی پیوسته در طبیعت و صنعت دارای این نمودار توزیع می‌باشند.	$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\quad ,-\infty<x<+\infty,\quad \mu\in R,\quad \sigma>0$
نرمال استاندارد	توزیع نرمالی که میانگین آن صفر و واریانس آن یک باشد و نماد آن $Z\sim N(0,1)$ است.	اگر $X\sim N(\mu,\sigma^2)$ آنگاه $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$

تقریب توزیع فوق هندسی به وسیله توزیع دوجمله‌ای	اگر M و N به سمت بی‌نهایت میل کنند و مقدار $\frac{M}{N}=p$ ثابت باشد، آنگاه:	$\lim\limits_{n\to\infty}\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}=\binom{n}{x}p^x(1-p)^{n-x}$
تقریب توزیع دو جمله‌ای به وسیله توزیع پواسون	اگر $n\to\infty$ و $p\to 0$ و مقدار $\mu=np$ مقداری ثابت باشد آنگاه:	$\lim\limits_{n\to\infty}\binom{n}{x}p^x(1-p)^{n-x}=\frac{e^{-\mu}\mu^x}{x!}$
تقریب توزیع دو جمله‌ای به وسیله توزیع نرمال	اگر X یک متغیر تصادفی دو جمله‌ای با میانگین $\mu=np$ و واریانس $\sigma^2=npq$ و $n\to\infty$ ($n>30$) آنگاه:	$Z=\frac{X-np}{\sqrt{npq}}$

توزیع نمونه‌ای

انتخاب قسمتی از جمعیت را **نمونه‌گیری** و قسمت انتخاب شده را **نمونه** گویند.

هر ویژگی یک جمعیت را یک پارامتر جمعیت و یا به اختصار **پارامتر** گویند و ویژگی از یک نمونه تصادفی را **آماره** گویند.

توزیع احتمال دو آماره را **توزیع نمونه‌ای** گویند.

قضیه حد مرکزی:

فرض کنید $X_1,X_2...$ دنباله‌ای از متغیرهای تصادفی هم توزیع با میانگین μ و واریانس σ^2 متناهی باشند به قسمی که هر تعداد متناهی از این متغیرهای تصادفی از یکدیگر مستقل باشند. در این صورت توزیع حدی $Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ موقعی که $n\to\infty$ توزیع نرمال استاندارد است.

Matlab20.blog.ir

توزیع نمونه‌ای میانگین	$(n>30)\; n\to\infty$	$Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$
توزیع کای دو $Y\sim\chi^2_{(n)}$	اگر $Z_1,\ldots,Z_n$ یک نمونه تصادفی n تایی از جمعیت نرمال استاندارد باشد(یعنی Z_i ها از یکدیگر مستقل باشند و $Z_i\sim N(0,1)$) آنگاه توزیع $Y=Z_1^2+\ldots+Z_n^2$ را یک توزیع مربع-کای با n درجه آزادی گویند	$f_Y(y)=\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}y^{\frac{n}{2}-1}e^{-\frac{y}{2}} \qquad y>0$
توزیع t $T\sim t_{(n)}$	فرض کنید $Z\sim N(0,1)$ و $Y\sim\chi^2_{(n)}$ و X و Y از یکدیگر مستقل باشند. در این صورت توزیع متغیر تصادفی $T=\dfrac{Z}{\sqrt{\frac{Y}{n}}}$ را یک توزیع t با n درجه آزادی گویند	$f_T(t)=\dfrac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}(1+\dfrac{t^2}{n})^{-\frac{n+1}{2}} \qquad -\infty<t<\infty$
قضیه	اگر $\overline{X}$ و S^2 به ترتیب میانگین و واریانس یک نمونه تصادفی به اندازه n از یک جمعیت نرمال با میانگین μ و واریانس σ^2 باشند آنگاه:	$T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$
توزیع نمونه‌ای اختلاف میانگین‌ها	حالت اول: واریانس دو جمعیت σ_1^2 و σ_2^2 معلوم باشند	$Z=\dfrac{(\overline{X}_1+\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$
توزیع نمونه‌ای اختلاف میانگین‌ها	حالت دوم: واریانس دو جمعیت نامعلوم اما مساوی باشند	$S_P^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)}$
توزیع F $F\sim F_{(m,n)}$	اگر $U\sim\chi^2_{(m)}$ و $V\sim\chi^2_{(n)}$ و متغیرهای تصادفی U و V از یکدیگر مستقل باشند آنگاه توزیع متغیر تصادفی $F=\dfrac{U/m}{V/n}$ را یک توزیع F با m و n درجه آزادی گویند	$f_F(x)=\dfrac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}(\dfrac{m}{n})^{\frac{m}{2}}\dfrac{x^{\frac{m}{2}-1}}{(1+\frac{m}{n}x)^{\frac{m+n}{2}}} \qquad x>0$
قضیه نسبت واریانس‌ها	اگر S_1^2 و S_2^2 به ترتیب واریانس‌های نمونه‌های تصادفی مستقل به اندازه‌های n_1 و n_2 از جمعیت‌های نرمال با واریانس‌های σ_1^2 و σ_2^2 باشند آنگاه:	$F=\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F_{(n_1-1,n_{12}-1)}$

نظریه برآوردیابی

استنباط آماری
- 1- برآورد پارامترهای مجهول جمعیت
  - الف- برآورد نقطه‌ای
  - ب- برآورد فاصله‌ای
- 2- آزمون فرض‌های آماری در مورد پارامترهای مجهول جمعیت

برآورد نقطه‌ای: تنها مقدار مشاهده شده یک آماره را به عنوان تقریبی از پارامتر مجهول جمعیت ارائه می‌دهیم.اگر براساس نمونه تصادفی $X_1,...,X_n$ آماره مورد نظر ما برای تخمین پارامتر مجهول $T=T(X_1,...,X_n)$ باشد. آنگاه متغیر تصادفی $T=T(X_1,...,X_n)$ برآوردگر پارامتر θ و عدد $t=T(x_1,...,x_n)$ را یک برآورد پارامتر θ گویند.

برآوردگر نااریب: برآوردگر $T=T(X_1,...,X_n)$ را یک برآوردگر نااریب برای θ گویند هرگاه $E(T)=\theta$ **برآورد فاصله‌ای:** یک فاصله (L,U) از اعداد حقیقی را به عنوان تقریبی از پارامتر مجهول θ ارائه می‌دهیم که این فاصله با یک احتمال زیاد پارامتر θ را در برداشته باشد. این فاصله را **فاصله اطمینان** گویند و اگر احتمال قرار گرفتن پارامتر θ در این فاصله $1-\alpha$ باشد، آن را یک فاصله اطمینان $1-\alpha$ ٪100 گویند.

برآورد میانگین جمعیت	الف– واریانس جمعیت معلوم است: یک فاصله اطمینان $1-\alpha$ ٪100 برای میانگین جمعیت نرمال μ موقعی که واریانس σ^2 معلوم است عبارت است از:	$\mu \in (\bar{x}-z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}}, \bar{x}+z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}})$
	ب– واریانس جمعیت نامعلوم است: یک فاصله اطمینان $1-\alpha$ ٪100 برای میانگین جمعیت نرمال μ موقعی که واریانس σ^2 نامعلوم است عبارت است از:	$\mu \in (\bar{x}-t_{(1-\frac{\alpha}{2})}(n-1)\frac{S}{\sqrt{n}}, \bar{x}+t_{(1-\frac{\alpha}{2})}(n-1)\frac{S}{\sqrt{n}})$
خطای برآورد میانگین	چون اغلب مقدار برآورد نقطه‌ای $\bar{x}$ دقیقا مساوی μ نیست بنابراین برآورد نقطه‌ای دارای خطا است. با استفاده از حدود فاصله اطمینان می‌توان میزان این خطا یعنی $\lvert\bar{x}-\mu\rvert$ را در دوحالت تعیین کرد: الف– اگر واریانس σ^2 معلوم باشدآنگاه:	$-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \bar{x}-\mu < Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \lvert\bar{x}-\mu\rvert < Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$
	ب– اگر واریانس σ^2 نامعلوم باشدآنگاه:	$-t_{1-\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} < \bar{x}-\mu < t_{1-\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} \quad \Rightarrow \quad \lvert\bar{x}-\mu\rvert < t_{1-\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}$
تعیین اندازه	الف– اگر واریانس σ^2 معلوم باشدآنگاه:	$Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le e \quad \Rightarrow \quad n \ge (Z_{1-\frac{\alpha}{2}}\sigma/e)$

Matlab20.blog.ir

نمونه	ب– اگر واریانس σ^2 نامعلوم باشدآنگاه:	$t_{1-\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} \le e \quad \Rightarrow \quad n \ge (Z_{1-\frac{\alpha}{2}}(n-1)\, s/e)$
برآورد واریانس جمعیت	برای به دست آوردن یک فاصله اطمینان برای σ^2	$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$
	یک فاصله اطمینان $1-\alpha$ 100٪ برای واریانس جمعیت نرمال σ^2 عبارت است،که در آن S^2 واریانس یک نمونه تصادفی n تایی است.	$\sigma^2 \in (\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)})$

برآورد تفاضل میانگین دو جمعیت	**الف– واریانس دو جمعیت معلوم باشند:** یک فاصله اطمینان $1-\alpha$ 100٪ برای تفاضل میانگین دو جمعیت نرمال که واریانس‌های آن‌ها معلوم است عبارت است از:	$\mu_1 - \mu_2 \in (\bar{x}_1 - \bar{x}_2 - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}})$
	الف– واریانس دو جمعیت نامعلوم اما مساوی باشند: یک فاصله اطمینان $1-\alpha$ 100٪ برای تفاضل میانگین دو جمعیت نرمال که واریانس‌های آن‌ها نامعلوم اما مساوی هستند عبارت است از:	$\mu_1-\mu_2 \in (\bar{x}_1-\bar{x}_2-t_{1-\frac{\alpha}{2}}(n_1-n_2-2)S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \bar{x}_1-\bar{x}_2-t_{1-\frac{\alpha}{2}}(n_1-n_2-2)S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})$
برآورد نسبت واریانس دو جمعیت	برای به دست آوردن یک فاصله اطمینان برای $\frac{\sigma_2^2}{\sigma_1^2}$ اگر دو جمعیت نرمال باشند	$F = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F(n_1-1, n_2-1)$
	یک فاصله اطمینان $1-\alpha$ 100٪ برای نسبت واریانس دو جمعیت نرمال عبارت است از:	$\frac{\sigma_2^2}{\sigma_1^2} \in (\frac{S_2^2}{S_1^2}\frac{1}{F_{1-\frac{\alpha}{2}}(n_2-1, n_1-1)}, \frac{S_2^2}{S_1^2}F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1))$

Matlab20.blog.ir

آزمون فرض های آماری

تعریف: آماره $T = T(x_1, x_2, ..., x_n)$ که براساس مقادیر مشاهده شده آن یک فرض را رد یا قبول می‌کنیم؛ آماره آزمون می‌گویند و به مجموعه مقادیری از این آماره که به ازای آن H_o را بایستی رد کرد ناحیه بحرانی آزمون گویند و با c نمایش می‌دهند.

خطای آزمون:

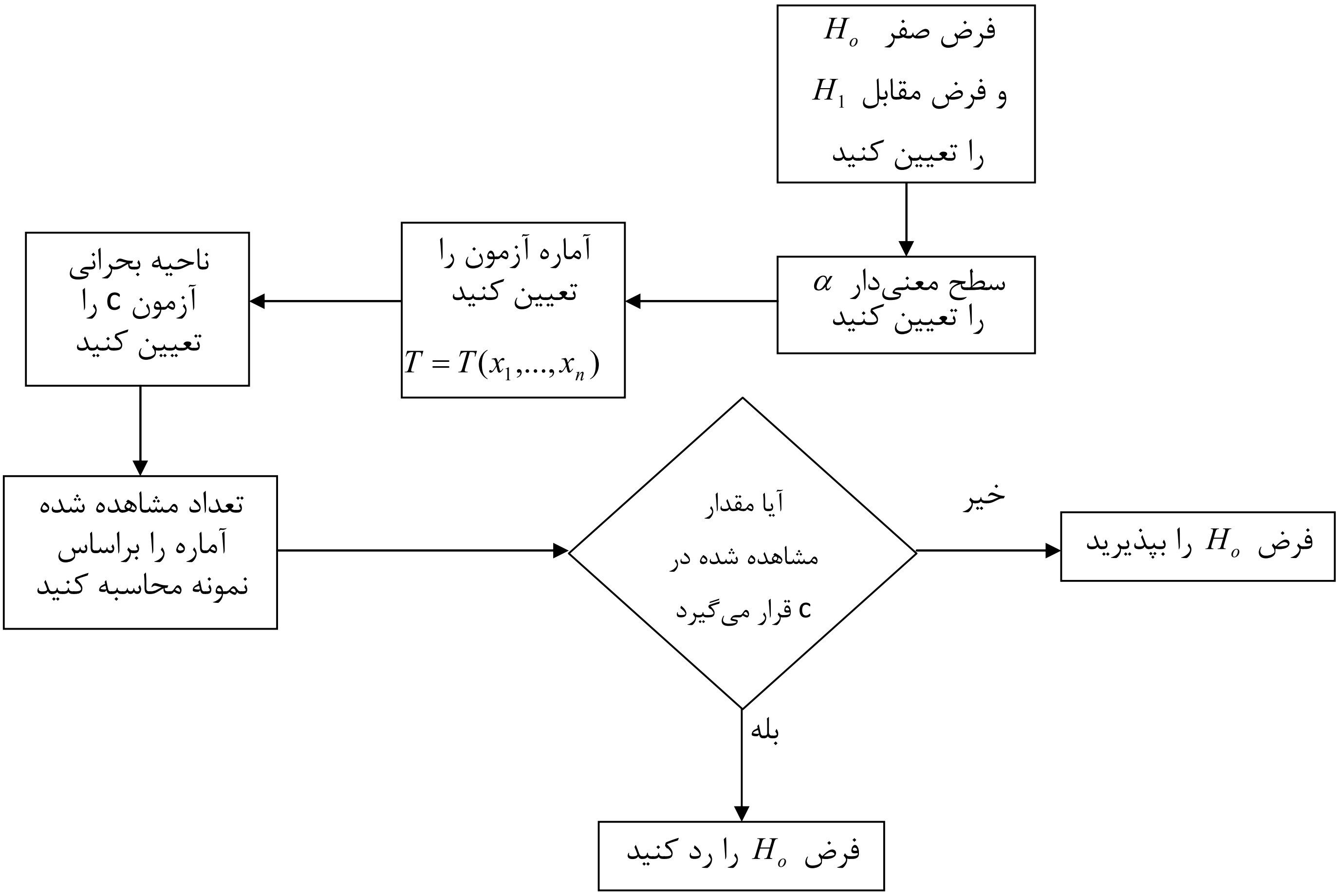
خطای نوع اول: احتمال خطای نوع اول را با α نمایش می‌دهند و آن را **سطح معنی‌دار** یا سطح تشخیص آزمون گویند.

$$\alpha = P\left(\text{فرض } H_o \text{ رد شود} \mid H_o \text{ درست است}\right)$$

خطای نوع دوم: $\beta = P\left(\text{فرض } H_o \text{ قبول شود} \mid H_1 \text{ درست است}\right)$

توان آزمون: $\beta^* = P\left(\text{فرض } H_o \text{ رد شود} \mid H_1 \text{ درست است}\right) = 1 - P\left(\text{فرض } H_o \text{ قبول شود} \mid H_1 \text{ درست است}\right) = 1 - \beta$

مراحل انجام یک آزمون:



Matlab20.blog.ir

آزمون های آماری روی میانگین و واریانس یک جمعیت نرمال

شماره آزمون	H_0	H_1	آماره آزمون	ناحیه بحرانی
1	$\mu \le \mu_0$ یا $\mu = \mu_0$	$\mu > \mu_0$	σ معلوم $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$	$Z > z_{1-\alpha}$
۲	$\mu \ge \mu_0$ یا $\mu = \mu_0$	$\mu < \mu_0$	σ معلوم $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$	$Z < -z_{1-\alpha}$
3	$\mu = \mu_0$	$\mu \ne \mu_0$	σ معلوم $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$	$\lvert Z \rvert > z_{1-\frac{\alpha}{2}}$
4	$\mu \le \mu_0$ یا $\mu = \mu_0$	$\mu > \mu_0$	σ نامعلوم $T = \dfrac{\overline{X} - \mu_0}{S / \sqrt{n}}$	$T > t_{1-\alpha}(n-1)$
5	$\mu \ge \mu_0$ یا $\mu = \mu_0$	$\mu < \mu_0$	σ نامعلوم $T = \dfrac{\overline{X} - \mu_0}{S / \sqrt{n}}$	$T < -t_{1-\alpha}(n-1)$
6	$\mu = \mu_0$	$\mu \ne \mu_0$	σ نامعلوم $T = \dfrac{\overline{X} - \mu_0}{S / \sqrt{n}}$	$\lvert T \rvert > t_{1-\alpha}(n-1)$
7	$\sigma^2 \le \sigma_0^2$ یا $\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$X^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	$X^2 > \chi^2_{1-\alpha}(n-1)$
8	$\sigma^2 \ge \sigma_0^2$ یا $\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$X^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	$X^2 < \chi^2_{\alpha}(n-1)$
9	$\sigma^2 = \sigma_0^2$	$\sigma^2 = \sigma_0^2$	$X^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	$X^2 > \chi^2_{1-\frac{\alpha}{2}}(n-1)$ یا $X^2 < \chi^2_{\frac{\alpha}{2}}(n-1)$

زمون های آماری روی تفاضل میانگین ها و نسبت واریانس های دو جمعیت نرمال

10				
10	$\mu_1-\mu_2=d_0$ یا $\mu_1-\mu_2\le d_0$	$\mu_1-\mu_2>d_0$	$Z=\dfrac{(\overline{X}_1-\overline{X}_2)-d_0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$	$Z>z_{1-\alpha}$
11	$\mu_1-\mu_2=d_0$ یا $\mu_1-\mu_2\ge d_0$	$\mu_1-\mu_2<d_0$	$Z=\dfrac{(\overline{X}_1-\overline{X}_2)-d_0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$ σ_1^2,σ_2^2 معلوم	$Z<-z_{1-\alpha}$
۱۲	$\mu_1-\mu_2=d_0$	$\mu_1-\mu_2\neq d_0$	$Z=\dfrac{(\overline{X}_1-\overline{X}_2)-d_0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$ σ_1^2,σ_2^2 معلوم	$\lvert Z\rvert>z_{1-\frac{\alpha}{2}}$
۱۳	$\mu_1-\mu_2=d_0$ یا $\mu_1-\mu_2\le d_0$	$\mu_1-\mu_2>d_0$	$T=\dfrac{(\overline{X}_1-\overline{X}_2)-d_0}{S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ ولی $\sigma_1=\sigma_2$ نامعلوم	$T>t_{1-\alpha}(n_1+n_2-2)$
۱٤	$\mu_1-\mu_2=d_0$ یا $\mu_1-\mu_2\ge d_0$	$\mu_1-\mu_2<d_0$	$T=\dfrac{(\overline{X}_1-\overline{X}_2)-d_0}{S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ ولی $\sigma_1=\sigma_2$ نامعلوم	$T<-t_{1-\alpha}(n_1+n_2-2)$
۱۵	$\mu_1-\mu_2=d_0$	$\mu_1-\mu_2\neq d_0$	$T=\dfrac{(\overline{X}_1-\overline{X}_2)-d_0}{S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ ولی $\sigma_1=\sigma_2$ نامعلوم	$\lvert T\rvert>t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$
۱٦	$\sigma_1^2=\sigma_2^2$ یا $\sigma_1^2\le\sigma_2^2$	$\sigma_1^2>\sigma_2^2$	$F=\dfrac{S_1^2}{S_2^2}$	$F>F_{1-\alpha}(n_1-1,n_2-1)$
۱۷	$\sigma_1^2=\sigma_2^2$ یا $\sigma_1^2\ge\sigma_2^2$	$\sigma_1^2<\sigma_2^2$	$F=\dfrac{S_1^2}{S_2^2}$	$F<F_{\alpha}(n_1-1,n_2-1)$
۱۸	$\sigma_1^2=\sigma_2^2$	$\sigma_1^2\neq\sigma_2^2$	$F=\dfrac{S_1^2}{S_2^2}$	$F<F_{\alpha}(n_1-1,n_2-1)$ یا $F>F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)$

Matlab20.blog.ir

آزمون برازندگی داده ها (آزمون نکویی برازش)

روش های کلی آزمون نکویی:
- آزمون کای دو
- آزمون‌های پارامتریک
- آزمون‌های غیرپارامتریک

الگوریتم آزمون نکویی برازش:

گام 1: توزیع احتمالی که تصور می‌رود مجموعه داده‌ها را داراست، عنوان کنید.

گام 2: فرض صفر H_o و فرض مقابل H_1 را عنوان کنید.

گام 3: پارامتر یا پارامترهای توزیع احتمال مدنظر را برآورد کنید.

گام 4: فراوانی تئوریک هر گروه (طبقه) را براساس توزیع احتمال به دست آورید.

گام 5: قرار دهید $X^2 = \sum \frac{(f_i - f_o)}{f_o}$ که f_i تعداد داده‌های تجربی و f_o تعداد مشاهده شده

گام٦: اگر $X^2 > \chi^2_{1-\alpha}(K-P-1)$ آنگاه H_o رد می‌شود.

رگرسیون خطی

تعریف رگرسیون ساده خطی: هرگاه بین متغیر مستقل x و متغیر وابسته $Y = Y|x$ یک رابطه خطی برقرار باشد گوییم یک مدل رگرسیون ساده خطی بین Y,x برقرار است.

ضرایب رگرسیون: β و α تعریف $\hat{\alpha}$ و $\hat{\beta}$: $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ $\quad \hat{\beta} = \dfrac{S_{xy}}{S_{xx}}$	مجموع مربعات باقیمانده $SSE = \sum_{i=1}^{n} e_i^2 = \sum (y_i - \alpha - \beta x_i)^2$
نکته: می‌توان نشان داد $\hat{\alpha}$ و $\hat{\beta}$ مقدار SSE را مینیمم می‌کنند و $\hat{\alpha}$ و $\hat{\beta}$ را برآوردهای حداقل مربعات گویند.	

$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} \qquad S_{xx} = \sum (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$$

استنباط آماری روی ضرایب رگرسیونی

قضیه: در مدل رگرسیونی ساده خطی با فرض $i = 1,2,\ldots,n \quad , \quad E_i \sim N(0,\sigma^2)$ داریم که: الف) $\hat{\alpha} \sim N(\alpha, \dfrac{\sigma^2 \sum x_i^2}{nS_{xx}})$ $\qquad$ ب) $\hat{\beta} \sim N(\beta, \dfrac{\sigma^2}{S_{xx}})$ ج) اگر $S^2 = \dfrac{SSE}{n-2}$ آنگاه $\dfrac{SSE}{\sigma^2} = \dfrac{(n-2)S^2}{\sigma^2} \sim \chi^2_{(n-2)}$ د) $\hat{\alpha}$ و S^2 و همچنین $\hat{\beta}$ و S^2 از یکدیگر مستقل هستند.	با استفاده از این قضیه می توان نتایج زیر را به دست آورد: $\dfrac{\hat{\alpha} - \alpha}{S\sqrt{\dfrac{\sum x_i^2}{nS_{xx}}}} \sim t_{(n-2)}$ $\qquad$ $\dfrac{\hat{\beta} - \beta}{\dfrac{s}{\sqrt{S_{xx}}}} \sim t_{(n-2)}$ $S^2 = \dfrac{SSE}{n-2} = \dfrac{1}{n-2}\sum (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 = \dfrac{1}{n-2}\left[S_{yy} - \hat{\beta}S_{xy}\right]$

محاسبه فواصل اطمینان $100(1-\gamma)\%$ برای $\hat{\alpha}$ و $\hat{\beta}$:

$$\alpha \in \left[\bar{\alpha} - t_{1-\frac{\gamma}{2}}(n-2)S\sqrt{\frac{\sum x_i^2}{nS_{xx}}}, \bar{\alpha} + t_{1-\frac{\gamma}{2}}(n-2)S\sqrt{\frac{\sum x_i^2}{nS_{xx}}}\right]$$

$$\beta \in \left[\hat{\beta} - t_{1-\frac{\gamma}{2}}(n-2)\frac{s}{\sqrt{S_{xx}}}, \hat{\beta} + t_{1-\frac{\gamma}{2}}(n-2)\frac{s}{\sqrt{S_{xx}}}\right]$$

آزمون های آماری روی ضرایب رگرسیونی $\hat{\alpha}$ و $\hat{\beta}$:

H_0	آماره آزمون	H_1	ناحیه بحرانی
$\alpha = \alpha_0$	$T = \dfrac{\hat{\alpha} - \alpha_0}{S\sqrt{\dfrac{\sum x_i^2}{nS_{xx}}}}$	$\alpha > \alpha_0$	$T > t_{1-\gamma}(n-2)$
		$\alpha < \alpha_0$	$T < -t_{1-\gamma}(n-2)$
		$\alpha \neq \alpha_0$	$\lvert T\rvert > t_{1-\frac{\gamma}{2}}(n-2)$
$\beta = \beta_0$	$T = \dfrac{\hat{\beta} - \beta_0}{\dfrac{S}{\sqrt{S_{xx}}}}$	$\beta > \beta_0$	$T > t_{1-\gamma}(n-2)$
		$\beta < \beta_0$	$T < -t_{1-\gamma}(n-2)$
		$\beta \neq \beta_0$	$\lvert T\rvert > t_{1-\frac{\gamma}{2}}(n-2)$

ضریب همبستگی خطی

تعریف ضریب همبستگی خطی: $\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}$ نکته: برای تشخیص میزان وابستگی دو متغیر تصادفی Y , X از معیاری به نام ضریب همبستگی خطی استفاده می شود.
همواره $-1\le\rho\le1$ میباشد.حالتهای مختلف از همبستگی خطی: الف- $\rho=1$ همبستگی کامل مثبت ب- $0\le\rho\le1$ ج- $\rho=0$ عدم همبستگی د- $-1\le\rho\le0$ همبستگی منفی هـ- $\rho=-1$ همبستگی کامل منفی
برآورد ضریب همبستگی: برای برآورد ضریب همبستگی یک نمونه تصادفی $(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)$ از $(X\ ,Y\ )$ را انتخاب میکنیم و از روی این نمونه تصادفی کوواریانس Y , X، واریانس X و واریانس Y را به صورت زیر برآورد میکنیم: $\hat{\sigma}_{XY}=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{n-1}S_{xy}$ $\hat{\sigma}_Y^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2=\frac{1}{n-1}S_{yy}$ $\hat{\sigma}_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n-1}S_{xx}$
حال با قرار دادن این برآوردگرها به جای پارامترهای $\hat{\sigma}_{XY},\hat{\sigma}_Y^2,\hat{\sigma}_X^2$ در فرمول ضریب همبستگی خطی، برآوردگر ضریب همبستگی به صورت زیر به دست میآید: $\hat{\rho}=R=\frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$
R^2 : به مقدار R^2 ضریب تعیین گویند و R^2 ۱۰۰٪ از تغییرات مقادیر Y جهت رابطه خطی آن با متغیر X به حساب میآید.

استنباط آماری روی ρ :

آماره $W=\frac{1}{2}Ln\frac{1+R}{1-R}$ برای اندازه نمونه تقریبا بزرگ n تقریبا دارای توزیع نرمال با میانگین تقریبی $\mu_w\approx\frac{1}{2}Ln\frac{1+\rho}{1-\rho}$ و واریانس تقریبی $\sigma_w^2=\frac{1}{n-3}$ است. یعنی به طور تقریبی داریم: $Z=\frac{W-\mu_w}{\sigma_w}\sim N(0,1)$

آزمون های آماری روی ρ که $\mu_{w_0}\approx\frac{1}{2}Ln\frac{1+\rho}{1-\rho}$ میباشد:			
ناحیه بحرانی	H_1	آماره آزمون	H_0
$Z>z_{1-\alpha}$	$\rho>\rho_0$	$Z=\frac{W-\mu_{w_0}}{\sigma_w}$	$\rho=\rho_0$
$Z<-z_{1-\alpha}$	$\rho<\rho_0$		
$\|Z\|>z_{1-\frac{\alpha}{2}}$	$\rho\neq\rho_0$		

درس آمار 2. دانشگاه مفید. احمد سعیدی

13940618

Matlab20.blog.ir