分布函数、期望、方差

两点分布 $P\{X=k\}=p^k(1-p)^{1-k}$ $E=p, D=p(1-p)$

二项分布 $X\sim B(n,p)$ $P\{X=k\}=C_n^k p^k(1-p)^{n-k}$ $E=np, D=np(1-p)$

泊松分布 $X\sim P(\lambda)$ $P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$ $E=\lambda, D=\lambda$

几何分布 $P\{X=k\}=(1-p)^{k-1}p$ $E=1/p, D=(1-p)/p^2$

均匀分布 $E=\frac{a+b}{2}, D=\frac{(b-a)^2}{12}$

正态分布 $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

指数分布 $X\sim Exp(\lambda)$ $P(X)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$ $E=\frac{1}{\lambda}, D=\frac{1}{\lambda^2}$

伽马分布 $f(x;\alpha,\lambda)=\begin{cases}\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}, & x\ge 0\\ 0, & x<0\end{cases}$ $\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx$ $E=\frac{\alpha}{\lambda}, D=\frac{\alpha}{\lambda^2}$

Beta 分布 $f(x;a,b)=\frac{x^{a-1}(1-x)^{b-1}}{\int_0^1 x^{a-1}(1-x)^{b-1}dx}=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}$ $E=\frac{a}{a+b}, D=\frac{ab}{(a+b)^2(a+b+1)}$

顺序统计量

$f_{(k)}(x)=\frac{n!}{(k-1)!(n-k)!}\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k}f(x)$

$f_{(1)}(x)=n\left[1-F(x)\right]^{n-1}f(x)$

$f_{(n)}(x)=n\left[F(x)\right]^{n-1}f(x)$

$f_{(k)(j)}(x,y)=\frac{n!}{(k-1)!(j-1-k)!(n-j)!}$

$\cdot\left[F(x)\right]^{k-1}\left[F(y)-F(x)\right]^{j-1-k}\left[1-F(y)\right]^{n-j}f(x)f(y)$

$f_{(1)(n)}(x,y)=\begin{cases}n(n-1)\left[F(y)-F(x)\right]^{n-2}f(x)f(y), & x<y\\ 0 & x\ge y\end{cases}$

抽样分布

$\chi^2(n)$： $E(X)=n, D(X)=2n$ $u_{1-\alpha}=-u_\alpha$

$t(n)$： $X\sim N(0,1), Y\sim\chi^2(n)$ 则 $T=\frac{X}{\sqrt{Y/n}}\sim t(n)$ $t_{1-\alpha}(n)=-t_\alpha(n)$

$F(n_1,n_2)$： $X\sim\chi^2(n_1)$ $Y\sim\chi^2(n_2)$ 则 $F=\frac{X/n_1}{Y/n_2}\sim F(n_1,n_2)$

$F_\alpha(n_1,n_2)=\frac{1}{F_{1-\alpha}(n_2,n_1)}$

$\bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$ $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1)$

第二章

点估计

矩估计： $E(X)=\bar{X}$ 最大似然估计： $\frac{\partial\ln L(\theta)}{\theta}=0, L(\theta)=\prod p(x)$

顺序统计量法： $\hat{\mu}=Me$, $\hat{\sigma}=R/d_n$

无偏、有效、相合

无偏 $E(\hat{\theta})=\theta$ ，渐进无偏 $\lim_{n\to\infty}E(\hat{\theta})=\theta$

均方误差： $E(\hat{\theta}-\theta)^2=D(\hat{\theta})+\left(E(\hat{\theta}-\theta)\right)^2$ ， $D(\hat{\theta})$ 越小越有效

一致最小方差无偏估计： $D_\theta(\hat{\theta}^*)=\min D_\theta(\hat{\theta})$ $E_\theta(\hat{\theta}^*\theta)=0$

RC 不等式： $D(\hat{\theta})\ge\frac{1}{nI(\theta)}=D_0(\theta)$, $I(\theta)=E\left[\frac{\partial}{\partial\theta}\ln p(X,\theta)\right]^2=-E\left[\frac{\partial^2}{\partial\theta^2}\ln p(X,\theta)\right]$

若 $D(\hat{\theta})=D_0(\theta)$ ，则 $\hat{\theta}$ 为有效估计（有效估计条件比一致最小方差估计严格）

有效估计充要条件: $\frac{\partial\ln L(\theta)}{\partial\theta}=C(\theta)\left[\hat{\theta}-\theta\right]$ ， $D(\hat{\theta})=\frac{1}{C(\theta)}, I(\theta)=\frac{C(\theta)}{n}$

均方相合估计： $\lim_{n\to\infty}E\left[\hat{\theta}-\theta\right]^2=0$ ，均方相合估计一定是一致相合估计

第三章假设检验

μ	σ^2 已知	$U=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1)$	$\bar{X}\mp\frac{\sigma}{\sqrt{n}}u_{\alpha/2}$	$\lvert U\rvert\ge u_{\alpha/2}, U\ge u_\alpha, U\le -u_\alpha$
μ	σ^2 未知	$T=\frac{\sqrt{n}(\bar{X}-\mu)}{S^*}\sim t(n-1)$	$\bar{X}\mp\frac{S^*}{\sqrt{n}}t_{\alpha/2}(n-1)$	$\lvert T\rvert\ge t_{\alpha/2}(n-1), U\ge t_\alpha(n-1), U\le -t_\alpha(n-1)$
σ^2	μ 已知	$\chi^2=\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$	$\frac{\sum_{i=1}^n(x_i-\mu)^2}{\chi^2_{\alpha/2}\ \&\ \chi^2_{1-\alpha/2}(n)}$	*大样本下检验* $\mu_1-\mu_2=c$ *接第五行第五列*
σ^2	μ 未知	$\chi^2=\frac{(n-1)S^{*2}}{\sigma^2}\sim\chi^2(n-1)$	$\frac{(n-1)S^{*2}}{\chi^2_{\alpha/2}(n-1)\ \&\ \chi^2_{1-\alpha/2}(n-1)}$	$\chi^2\le\chi^2{}_{1-\alpha/2}(n-1)$或$\ge\chi^2{}_{\alpha/2}$ $\chi^2\ge\chi^2{}_\alpha(n-1), \chi^2\le\chi^2{}_{1-\alpha}(n-1)$
$\mu_1-\mu_2$	σ_1^2,σ_2^2 均已知	$U=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$(\bar{X}-\bar{Y})\mp u_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$	$\frac{(X-Y)-c}{\sqrt{\frac{S_{1n_1}^2}{n_1}+\frac{S_{2n_2}^2}{n_2}}}\sim N(0,1)$ *拒绝域* $\lvert u\rvert\ge u_{\alpha/2}$
$\mu_1-\mu_2$	σ_1^2,σ_2^2 均未知但 $\sigma_1^2=\sigma_2^2$	$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$	$(\bar{X}-\bar{Y})\mp t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	$\lvert T\rvert\ge t_{\alpha/2}(n_1+n_2-2), T\ge t_\alpha, T\le -t_\alpha$ (T 的分子变为 $(\bar{X}-\bar{Y})-c$)
$\mu_1-\mu_2$	σ_1^2,σ_2^2 均未知但 $n_1=n_2$	$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_Z^*/\sqrt{n}}\sim t(n-1)$	$\bar{Z}\mp\frac{S_Z^*}{\sqrt{n}}t_{\alpha/2}(n-1)$	$\bar{Z}=\bar{X}-\bar{Y}$ $S_Z^*=\sqrt{\frac{1}{n-1}\sum_{i=1}^n(Z_i-\bar{Z})^2}$
σ_1^2/σ_2^2	μ_1,μ_2 均未知	$\frac{\sigma_2^2S_{1n_1}^{*2}}{\sigma_1^2S_{2n_2}^{*2}}\sim F(n_1-1,n_2-1)$	$\frac{S_{1n_1}^{*2}/S_{2n_2}^{*2}}{F_{\alpha/2}\ \&\ F_{1-\alpha/2}(n_1-1,n_2-1)}$	$F\le F_{1-\alpha/2}(n_1-1,n_2-1)$或$\ge F_{\alpha/2}$ $F\ge F_\alpha, F\le F_{1-\alpha}$
指数分布 λ	$2n\lambda\bar{X}\sim\chi^2(2n)$	$\frac{\chi^2_{1-\alpha/2}\ \&\ \chi^2_{\alpha/2}(2n)}{2n\bar{X}}$	$2n\lambda_0\bar{X}\le\chi^2{}_{1-\alpha/2}(2n)$或$\ge\chi^2{}_{\alpha/2}$	*补充：* $H_0:\sigma_1{}^2=\sigma_2^2$ *且* μ *已知* $F=\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2}\sim F(n_1,n_2)$ *拒绝域：* $F\ge F_{\frac{\alpha}{2}}(n_1,n_2)$或$\le F_{1-\frac{\alpha}{2}}(n_1,n_2)$
二项分布 p	$\frac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}}\sim N(0,1)$	$\frac{1}{2a}\left(b\mp\sqrt{b^2-4ac}\right)$ 大样本 0-1: $\frac{m}{n}\mp\sqrt{\frac{1}{n}\frac{m}{n}\left(1-\frac{m}{n}\right)}\cdot u_{\alpha/2}$	$S_w=\sqrt{\frac{(n_1-1)S_{1n_1}^{*2}+(n_2-1)S_{2n_2}^{*2}}{n_1+n_2-2}}$ $a=n+u_{\alpha/2}^2$, $b=2n\bar{X}+u_{\alpha/2}^2$, $c=n\bar{X}^2$	
大样本 μ	$\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim N(0,1)$	$\bar{X}\mp\frac{S}{\sqrt{n}}u_{\alpha/2}$	$\lvert u\rvert\ge u_{\alpha/2}$	

第五列表示假设检验 $H_0=,\le,\ge$ 的拒绝域

χ^2 检验 $H_0:X\sim N(0,1)$ ，最大似然估计参数（均值方差），设拒绝域： $K_n=\sum_{i=1}^r\frac{m_i^2}{np_i}-n\ge\chi_\alpha^2(r-k-1)$ ，m 为频数，n 为样本数，p 为理论概率值，r 为分区数,k 为未知参数个数

独立性检验 $H_0:p_{ij}=p_{i\cdot}p_{\cdot j}$ ，拒绝域： $K_n=\sum_{i=1}^r\sum_{j=1}^s\left(n_{ij}-\frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2\Big/\left(\frac{n_{i\cdot}n_{\cdot j}}{n}\right)\ge\chi_\alpha{}^2\left((r-1)(s-1)\right)$

第四章

单因素方差分析表

$H_0:\delta_1=...=\delta_r=0$

方差来源	平方和	自由度	均方和	F 值
因素 A（组间）	$Q_A=\sum_{i=1}^{r} n_i(\bar{x}_i-\bar{x})^2$	$r-1$	$\bar{Q}_A=\frac{Q_A}{r-1}$	$F=\frac{\bar{Q}_A}{\bar{Q}_E}$
误差 E（组内）	$Q_E=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$	$n-r$	$\bar{Q}_E=\frac{Q_E}{n-r}$	
总和	$Q_T=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2=Q_A+Q_E$	$n-1$	拒绝域 : $F\ge F_\alpha(r-1,n-r)$	

点估计： $\hat{\mu}=\bar{X},\hat{\mu}_i=\bar{X}_i,\hat{\delta}_i=\bar{X}_i-\bar{X},\hat{\sigma}^2=\frac{Q_E}{n-r}=\bar{Q}_E$

均值差区间估计：统计量 $T=\frac{(X_i-X_k)-(\mu_i-\mu_k)}{\sqrt{(\frac{1}{n_i}-\frac{1}{n_k})\bar{Q}_E}}\sim t(n-r)$ ，置信区间 $(X_i-X_k\pm t_{\frac{\alpha}{2}}(\mathrm{n-r})\sqrt{(\frac{1}{n_i}-\frac{1}{n_k})\bar{Q}_E})$

多因素方差有重复试验分析表

方差来源	平方和	自由度	均方和	F 值
因素 A	$Q_A=sl\sum_{i=1}^{r}(\bar{X}_{i\cdot\cdot}-\bar{X})^2$	$r-1$	$\bar{Q}_A=\frac{Q_A}{r-1}$	$F_A=\frac{\bar{Q}_A}{\bar{Q}_E}$
因素 B	$Q_B=rl\sum_{j=1}^{s}(\bar{X}_{\cdot j\cdot}-\bar{X})^2$	$s-1$	$\bar{Q}_B=\frac{Q_B}{s-1}$	$F_B=\frac{\bar{Q}_B}{\bar{Q}_E}$
交互作用	$Q_I=l\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{X}_{ij\cdot}-\bar{X}_{i\cdot\cdot}-\bar{X}_{\cdot j\cdot}+\bar{X})^2$	$(r-1)\cdot(s-1)$	$\bar{Q}_I=\frac{Q_I}{(r-1)(s-1)}$	$F_I=\frac{\bar{Q}_I}{\bar{Q}_E}$
误差	$Q_E=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}(X_{ijk}-\bar{X}_{ij\cdot})^2$	$rs(l-1)$	拒绝域：$F_A\ge F_\alpha(r-1,rs(l-1))$ $F_B\ge F_\alpha(s-1,rs(l-1))$ $F_I\ge F_\alpha((r-1)(s-1),rs(l-1))$	
总和	$Q_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}(X_{ijk}-\bar{X})^2$	$rsl-1$		

无重复试验

方差来源	平均和	自由度	均方和	F 值
因素 A	$Q_A=s\sum_{i=1}^{r}(\bar{X}_{i\cdot}-\bar{X})^2$	$r-1$	$\bar{Q}_A=\frac{Q_A}{r-1}$	$F_A=\frac{\bar{Q}_A}{\bar{Q}_E}$
因素 B	$Q_B=r\sum_{j=1}^{s}(\bar{X}_{\cdot j}-\bar{X})^2$	$s-1$	$\bar{Q}_B=\frac{Q_B}{s-1}$	$F_B=\frac{\bar{Q}_B}{\bar{Q}_E}$
误差	$Q_E=\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{X}_{ij}-\bar{X}_{i\cdot}-\bar{X}_{\cdot j}-\bar{X})^2$	$(r-1)(s-1)$	$\bar{Q}_E=\frac{Q_E}{(r-1)(s-1)}$	
总和	$Q_T=\sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij}-\bar{X})^2=Q_A+Q_B+Q_E$	$rs-1$	$F_A\ge F_\alpha(r-1,(r-1)(s-1))$ $F_B\ge F_\alpha(s-1,(r-s)(s-1))$	

正交表方差分析

$Q_j=\frac{s_j}{n}\sum_{i=1}^{s_j}T_{ij}^2-\frac{1}{n}\left(\sum_{i=1}^{s_j}T_{ij}\right)^2$,sj 为水平数,自由度 $r_j=s_j-1$

$Q_T=\sum_{i=1}^{n}(y_i-\bar{y})^2$ 自由度 $r_T=n-1$ $Q_e=Q_T-\sum Q_j$ 自由度 $r_e=r_T-\sum r_j$

方差分析表同上，计算 F 值与拒绝域

第五章

一元线性回归

一元最小二乘估计： $\hat{b}=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\,\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}$ $\hat{a}=\bar{y}-\hat{b}\bar{x}$ $\hat{\sigma}^2=\frac{1}{n}\sum(y_i-\bar{y})^2-\frac{\hat{b}^2}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

假 设 检 验 ： $l_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2$ $l_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2$ $l_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$ $\hat{\sigma}^{*2}=\frac{n}{n-2}\hat{\sigma}^2$ $H_0:b=0$ $t=\frac{\hat{b}}{\hat{\sigma}^*}\sqrt{l_{xx}}$ 拒 绝 域 ： $|t|\ge t_{\alpha/2}(n-2)$ 样本相关系数： $r=\frac{l_{xy}}{\sqrt{l_{xx}l_{yy}}}$

预测区间： $\frac{y_0-\hat{y}_0}{\hat{\sigma}^*\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}}}\sim t(n-2)$，记 $\delta(x_0)=t_{\alpha/2}(n-2)\hat{\sigma}^*\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}}$,区间为 $\hat{y}_0\mp\delta(x_0)$，n 很大时 $\delta(x_0)\approx u_{\alpha/2}\hat{\sigma}^*$

多元线性回归 $X=\begin{bmatrix}1\ x_{11}\ \dots\ x_{1k}\\ \vdots\\ 1\ x_{n1}\ \dots\ x_{nk}\end{bmatrix}$ 其中 k 为变量种类数，n 为变量样本数

多元最小二乘估计： $B=(X^TX)^{-1}X^TY$ $\hat{\sigma}^{*2}=\frac{Q_e}{n-k-1}$ 为 σ^2 的无偏估计

回归方程显著性检验

H_0：$b_1=...=b_k=0$ 拒绝域 $F\ge F_\alpha(k,n-k-1)$

方差来源	平方和	自由度	均方和	F 值
回归	$Q_r=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	k	$\bar{Q}_r=\frac{Q_r}{k}$	$F=\frac{\bar{Q}_r}{\bar{Q}_e}$
误差	$Q_e=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	$n-k-1$	$\bar{Q}_e=\frac{Q_e}{n-k-1}$	
总和	$Q_t=\sum_{i=1}^{n}(y_i-\bar{y})^2=Q_r+Q_e$	$n-1$		

回归系数显著性检验

H_{0_j}：$b_j=0$ $C=(X^TX)^{-1}=(c_{ij})_{(k+1)\times(k+1)}$， c_{jj} 为第 $j+1$ 个元素

检验统计量 $t_j=\frac{\hat{b}_j/\sqrt{c_{jj}}}{\sqrt{Q_e/(n-k-1)}}\sim t(n-k-1)$，拒绝域 $|t_j|\ge t_{\alpha/2}(n-k-1)$

预测区间， $\hat{y}_0\mp t_{\alpha/2}(n-k-1)\sqrt{1+\sum_{i=0}^{k}\sum_{j=0}^{k}c_{ij}x_{0i}x_{0j}}\sqrt{\frac{Q_e}{n-k-1}}$

第六章

后验期望估计 $\pi(\theta|x)\propto\pi(\theta)p(x|\theta)$ $p(x|\theta)=\prod_{i=1}^{n}f(x_i|\theta)$ $\hat{\theta}=E(\theta|x)$

最大后验估计 $\frac{\partial\ln(\pi(\theta)p(x|\theta))}{\partial\theta}=0$ （核取对数求导为 0）

后验中位数估计 $x|\theta\sim U(0,\theta)$ $\pi(\theta)=aM^a\theta^{-(a+1)}$ $F(\theta)=1-(M_0/\theta)^{a+n}$

$M_0=\max(x_1,...,x_n,M)$ 令 $F(\theta)=1/2$， 中位数估计 $\hat{\theta}(x)=2^{\frac{1}{a+n}}M_0$

区间估计 $X\sim N(\mu,\sigma^2)$ $\mu\sim N(\mu_0,\sigma_0^2)$ $x(\mu|x)\propto\exp\left\{-\frac{(\mu-a)^2}{2b^2}\right\}$ $a=\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)/\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)$ $b^2=\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)^{-1}$ 区间 $a\mp bu_{\alpha/2}$

假设检验 $P(H_0|x)/P(H_1|x)$，大于 1 则 H0 成立

先验分布选取:贝叶斯假设 $\pi(\theta)\propto1$ 共轭先验分布 $\pi(x|\theta)$与$\pi(\theta)$ 同一核函数

课后题 11.5,12.5；11.5,11.6 投票人数 $X\sim B(1000,\theta)$