东北大学应用数理统计笔记

概率论基础

0.1 概率 $\mathrm{P}(\mathrm{A})$

0.1.1 事件间的关系

事件独立：$\mathrm{P}(\mathrm{AB})=\mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B})$

事件互斥：$\mathrm{P}(\mathrm{AB})=0$

条件概率：$\mathrm{P}(\mathrm{B}\mid\mathrm{A})=\frac{\mathrm{P}(\mathrm{AB})}{\mathrm{P}(\mathrm{A})}$

0.1.2 概率的计算公式

加法公式：

1. $\mathrm{P}(\mathrm{A}\cup\mathrm{B})=\mathrm{P}(\mathrm{A})+\mathrm{P}(\mathrm{B})-\mathrm{P}(\mathrm{AB})$
2. $\mathrm{P}(\mathrm{A}\cup\mathrm{B}\cup\mathrm{C})=\mathrm{P}(\mathrm{A})+\mathrm{P}(\mathrm{B})+\mathrm{P}(\mathrm{C})-\mathrm{P}(\mathrm{AB})-\mathrm{P}(\mathrm{AC})-\mathrm{P}(\mathrm{BC})+\mathrm{P}(\mathrm{ABC})$
3. 如果事件互斥：$\mathrm{P}(\mathrm{A}\cup\mathrm{B})=\mathrm{P}(\mathrm{A})+\mathrm{P}(\mathrm{B})$

减法公式：

1. $\mathrm{P}(\mathrm{A}-\mathrm{B})=\mathrm{P}(\mathrm{A})-\mathrm{P}(\mathrm{AB})$
2. 如果事件互斥：$\mathrm{P}(\mathrm{A}-\mathrm{B})=\mathrm{P}(\mathrm{A})$

乘法公式：

1. $\mathrm{P}(\mathrm{AB})=\mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B}\mid\mathrm{A})$
2. $\mathrm{P}(\mathrm{ABC})=\mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B}\mid\mathrm{A})\mathrm{P}(\mathrm{C}\mid\mathrm{AB})$
3. 如果事件独立：$\mathrm{P}(\mathrm{AB})=\mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B})$

全概率公式：$\mathrm{P}(\mathrm{A})=\sum_{\mathrm{i}=1}^{\mathrm{N}}\mathrm{P}(\mathrm{B_i})\mathrm{P}(\mathrm{A}\mid\mathrm{B_i})$

贝叶斯公式：$\mathrm{P}(\mathrm{B_j}\mid\mathrm{A})=\frac{\mathrm{P}(\mathrm{B_j})\mathrm{P}(\mathrm{A}|\mathrm{B_j})}{\sum_{\mathrm{i}=1}^{\mathrm{N}}\mathrm{P}(\mathrm{B_i})\mathrm{P}(\mathrm{A}|\mathrm{B_i})}$

0.2 随机变量 X

0.2.1 随机变量的概率分布 P

离散型——分布律：$\mathrm{P}\{\mathrm{X}=\mathrm{x_k}\}=\mathrm{P_k}$

特别地：

$$\mathrm{P}\{\mathrm{N}=\mathrm{n}\}=\mathrm{P}\{\mathrm{N}\leqslant\mathrm{n}\}-\mathrm{P}\{\mathrm{N}\leqslant\mathrm{n}-1\}=\mathrm{F}(\mathrm{n})-\mathrm{F}(\mathrm{n}-1)$$

$$\mathrm{P}\{\mathrm{N}=\mathrm{n}\}=\mathrm{P}\{\mathrm{N}\geqslant\mathrm{n}\}-\mathrm{P}\{\mathrm{N}\geqslant\mathrm{n}+1\}$$

连续型——概率密度：$\mathrm{f}(\mathrm{x})=\frac{\mathrm{dF}(\mathrm{x})}{\mathrm{dx}}$

0.2.2 随机变量的分布函数 $\mathrm{F}(\mathrm{x})$

离散型：$\mathrm{F}(\mathrm{x})=\mathrm{P}\{\mathrm{X}\leqslant\mathrm{x_k}\}=\sum_{\mathrm{x_k}\leq\mathrm{x}}\mathrm{P_k}$

连续型：$\mathrm{F}(\mathrm{x})=\mathrm{P}\{\mathrm{X}\leqslant\mathrm{x}\}=\int_{-\infty}^{\mathrm{x}}\mathrm{f}(\mathrm{t})\mathrm{dt}$

0.2.3 随机变量的数学期望 $\mathrm{E}(\mathrm{X})$

离散型：$\mathrm{E}(\mathrm{X})=\sum_{\mathrm{k}=1}^{\infty}\mathrm{x_k}\mathrm{p_k}$

连续型：$E(X)=\int_{-\infty}^{\infty}xf(x)dx$

性质：

1. $E(C)=C$
2. $E(CX)=CE(X)$
3. $E(X\pm Y)=E(X)\pm E(Y)$
4. 如果 X 与 Y 互不相关：$E(XY)=E(X)E(Y)$

0.2.4 随机变量的方差 $D(X)$

定义：$D(X)=E\left\{[X-E(X)]^2\right\}$

性质：

1. $D(X)=E(x^2)-[E(x)]^2$
2. $D(C)=0$
3. $D(aX+b)=a^2D(X)$
4. $D(X\pm Y)=D(X)+D(Y)\pm 2Cov(X,Y)$
5. 如果 X 与 Y 互不相关：$D(X\pm Y)=D(X)+D(Y)$

0.2.5 随机变量的矩 a_k,b_k

k 阶原点矩：$a_k=E(X^k)$

k 阶中心矩：$b_k=E\left\{[X-E(X)]^k\right\}$

$k+l$ 阶混合矩：$E(X^kY^l)$

$k+l$ 阶中心矩：$E\left\{[X-E(X)]^k[X-E(X)]^l\right\}$

性质：

1. $a_1=E(X)$
2. $a_2=E(X^2)$
3. $b_2=D(X)$

0.2.5 随机变量的协方差 $Cov(X,Y)$

定义：$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$

性质：

1. $Cov(X,Y)=E(XY)-E(X)E(Y)$
2. $Cov(X,Y)=Cov(Y,X)$
3. $Cov(aX,bY)=abCov(X,Y)$
4. $Cov(X_1\pm X_2,Y)=Cov(X_1,Y)\pm Cov(X_2,Y)$

相关系数：$\rho_{xy}=\frac{Cov(X,Y)}{\sqrt{D(X)D(Y)}}$

独立条件：X，Y 都服从正态分布，且协方差为 0，可以推 X，Y 独立

0.3 随机向量 η

随机向量：$\eta=\begin{bmatrix}X_1 & X_2 & \ldots & X_n\end{bmatrix}^T$

随机向量的期望向量：$\theta = \begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_n \end{bmatrix}^T$

随机向量的协方差矩阵：

$$\Sigma = \begin{bmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Cov(X_2, X_2) & \dots & Cov(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & Cov(X_n, X_n) \end{bmatrix}$$

$\eta \sim N(\theta, \Sigma)$

性质：$A\eta \sim N(A\theta, A\Sigma A^T)$

0.4 $\mathrm{Chebyshev}$ 不等式

1. $P\{|X - \mu| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2}$
2. $P\{|X - \mu| < \varepsilon\} \geqslant 1 - \frac{\sigma^2}{\varepsilon^2}$

0.5 中心极限定理

$\frac{\sum_{k=1}^{\mu} X_i - nE(X)}{\sqrt{nD(X)}} \sim N(0, 1)$

1 抽样分布

1.1 统计量 T

1.1.1 样本均值 $\bar{X}$

定义：$\bar{X} = \frac{1}{n}\sum_{i=1}^{n1} X_i$

性质：

1. $E(\bar{X}) = E(X)$
2. $D(\bar{X}) = \frac{D(X)}{n}$

1.1.2 样本方差 S^2

定义：$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

性质：

1. $E(S^2) = D(X)$
2. $D(S^2) = \frac{2D(X)^2}{n-1}$

1.1.3 样本矩 A_k, B_k

k 阶样本原点矩：$A_k = \frac{1}{n}\sum_{i=1}^{n1} X_i^k$

k 阶样本中心矩：$B_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^k$

性质：

1. $A_1 = \bar{X}$
2. $A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$
3. $B_2 = \frac{n-1}{n}S^2$

1.1.4. 顺序统计量 $X_{[i]}$

极小统计量：$X_{[1]} = \min(X_i)$

极大统计量：$X_{[n]} = \max(X_i)$

经验分布：$F_n(X) = \frac{k}{n}, \quad X_{[k]} \leqslant X < X_{[k+1]}$

性质:

1. $P\{X_{[1]} \leqslant x\} = 1 - P\{X_{[1]} > x\} = 1 - [P\{X > x\}]^n = 1 - [1 - P\{X \leqslant x\}]^n$

2. $P\{X_{[n]} \leqslant x\} = [P\{X \leqslant x\}]^n$
3. $P\{X_{[1]} = x\} = n[1 - P\{X \leqslant x\}]^{n-1} P\{X = x\}$
4. $P\{X_{[n]} = x\} = n[P\{X \leqslant x\}]^{n-1} P\{X = x\}$

1.2 常用的分布

1.2.1 常用的离散型分布

分布	记作	$P\{X=k\}$	$E(X)$	$D(X)$
$0-1$分布	$X \sim B(1,p)$	$P\{X=k\} = p^k(1-p)^{1-k}$	$E(X) = p$	$D(X) = p(1-p)$
二项分布	$X \sim B(n,p)$	$P\{X=k\} = C_n^k p^k (1-p)^{n-k}$	$E(X) = np$	$D(X) = np(1-p)$
几何分布	$X \sim G(p)$	$P\{X=k\} = p(1-p)^{k-1}$	$E(X) = \frac{1}{p}$	$D(X) = \frac{1-p}{p^2}$
超几何分布	$X \sim H(n,M,N)$	$P\{X=k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^k}$	$E(X) = \frac{nM}{N}$	$D(x) = \frac{nM}{N}(1-\frac{M}{N})\frac{N-n}{N-1}$
Poisson分布	$X \sim P(\lambda)$	$P\{X=k\} = \frac{\lambda^k}{k!}e^{-\lambda}$	$E(X) = \lambda$	$D(X) = \lambda$

1.2.2 常用的连续型分布

连续型分布	记作	$f(x)$	$E(X)$	$D(X)$
均匀分布	$X \sim U(a,b)$	$f(x) = \frac{1}{b-a}, \quad a \leqslant x \leqslant b$	$E(X) = \frac{a+b}{2}$	$D(X) = \frac{(b-a)^2}{12}$
指数分布	$X \sim E(\lambda)$	$f(x) = \lambda e^{-\lambda x}, \quad a \leqslant x \leqslant b$	$E(X) = \frac{1}{\lambda}$	$D(X) = \frac{1}{\lambda^2}$
正态分布	$X \sim N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$E(X) = \mu$	$D(X) = \sigma^2$
Γ 分布	$X \sim \Gamma(\alpha,\lambda)$	$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$	$E(X) = \frac{\alpha}{\lambda}$	$D(X) = \frac{\alpha}{\lambda^2}$
$I\Gamma$ 分布	$X \sim I\Gamma(\alpha,\lambda)$	$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\lambda}{x}}, \quad x > 0$	$E(X) = \frac{\lambda}{\alpha-1}$	$D(X) = \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)}$
B 分布	$X \sim B(\alpha,\beta)$	$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x > 0$	$E(X) = \frac{\alpha}{\alpha+\beta}$	$D(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

1.2.3 常用的统计分布

分布	记作	$f(x)$	$E(X)$	$D(X)$
χ^2 分布	$X \sim \chi^2(n)$	$k_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$	$E(X) = n$	$D(X) = 2n$
t 分布	$X \sim t(n)$	$t_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$	$E(X) = 0$	$D(X) = \frac{n}{n-2}$
F 分布	$X \sim F(m,n)$	$f_{m,n}(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}}$	$E(X) = \frac{n}{n-2}$	$D(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$

1.2.4 具有可加性的分布

前提：X 与 Y 独立

分布	分布 X	分布 Y	分布 $X+Y$
二项分布	$X\sim B(n_1,p)$	$Y\sim B(n_2,p)$	$X+Y\sim B(n_1+n_2,p)$
$Poisson$分布	$X\sim P(\lambda_1)$	$Y\sim P(\lambda_2)$	$X+Y\sim P(\lambda_1+\lambda_2)$
正态分布	$X\sim N\left(\mu_1,\sigma_1^2\right)$	$Y\sim N\left(\mu_2,\sigma_2^2\right)$	$X+Y\sim N\left(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2\right)$
Γ 分布	$X\sim\Gamma(\alpha_1,\lambda)$	$Y\sim\Gamma(\alpha_2,\lambda)$	$X+Y\sim\Gamma(\alpha_1+\alpha_2,\lambda)$
χ^2 分布	$X\sim\chi^2(n_1)$	$Y\sim\chi^2(n_2)$	$X+Y\sim\chi^2(n_1+n_2)$

1.2.5 具有无记忆性的分布

$P(X>s+t|X>t)=P(X>s)$

分布	事件
几何分布	“扔了9次硬币正面，第10次反面概率还是 1/2”
指数分布	“等了9小时没出现客人，接下来的1小时出现第一位客人的概率还是不变”

1.2.6 $0-1$ 分布 $X\sim B(1,p)$

事件：掷 1 次硬币，出现正面的概率

1.2.7 二项分布 $X\sim B(n,p)$

事件：掷 n 次硬币，出现 k 次正面的概率

1.2.8 几何分布

事件：掷到第 k 次硬币，才出现正面的概率

1.2.9 超几何分布

事件：在 N 件产品中有 M 件次品，从中一次性抽取 n 件产品，有 k 件次品的概率

1.2.10 $Poisson$分布 $X\sim P(\lambda)$

事件：一段时间内，发生 k 次的概率

$Poisson$定理：n 很大，p 很小时：$B(n,p)\approx P(np)$

1.2.11 均匀分布 $X\sim U(a,b)$

分布函数：$F(x)=\begin{cases}0 & , \quad x<a\\ \frac{x-a}{b-a} & , \quad a\leqslant x<b\\ 1 & , \quad b\leqslant x\end{cases}$

1.2.12 指数分布 $X\sim E(\lambda)$

事件：发生一次事件，所需要的时间。

和 $Poisson$分布 一同理解：假如 $\lambda=2$，一小时平均发生两次，发生一次平均需要半小时。

分布函数：$F(x)=\begin{cases}1-e^{-\lambda x} & ,x>0\\ 0 & ,x\leqslant 0\end{cases}$

建立服从 χ^2 分布检验量：$2\lambda n\bar{X}\sim\chi^2(2n)$

1.2.13 正态分布 $X\sim N\left(\mu,\sigma^2\right)$

分布函数：$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$

1.2.14 二维正态分布 $(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$

概率密度函数：$f(x,y)=$
$\left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^{-1}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right]$

1.2.15 Γ 分布 $X\sim\Gamma(\alpha,\lambda)$

Γ 分布性质：

1. $cX\sim\Gamma\left(\alpha,\frac{\lambda}{c}\right)$
2. $E(\lambda)=\Gamma(1,\lambda)$
3. $\chi^2(n)=\Gamma\left(\frac{n}{2},\frac{1}{2}\right)$

Γ 函数：$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx,\alpha>0$

Γ 函数性质：

1. $\Gamma(1)=1$
2. $\Gamma(\frac{1}{2})=\sqrt{\pi}$
3. $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$
4. $\Gamma(n+1)=n\Gamma(n)=n!$

1.2.16 $I\Gamma$ 分布 $X\sim I\Gamma(\alpha,\lambda)$

$X\sim\Gamma(\alpha,\lambda)$，则 $\frac{1}{X}\sim I\Gamma(\alpha,\lambda)$

1.2.17 B 分布 $X\sim B(\alpha,\beta)$

B 函数：$B(\alpha,\beta)=\int_0^1x^{\alpha-1}(1-x)^{\beta-1}dx\quad(\alpha>0)$

B 函数性质：

1. $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

1.2.18 χ^2 分布 $X\sim\chi^2(n)$

1.2.19 t 分布 $X\sim t(n)$

t 分布性质：

1. $t_{1-\alpha}(n)=-t_\alpha(n)$
2. $t(n)^2=X(n)$

1.2.20 F 分布 $X\sim F(m,n)$

F 分布性质：

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

1.3 常用的抽样分布

1.3.1 一个正态总体的抽样分布

1. $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1)$
2. $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$
3. $\frac{\sqrt{n}(\bar{X}-\mu)}{S}\sim t(n-1)$
4. $\frac{\sum_{i=1}^n(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$

1.3.2 两个正态总体的抽样分布

1. $\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
2. 如果 $\sigma_1^2=\sigma_2^2$：

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

   $\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$
3. $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1,n_2-1)$
4. $\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2/n_1\sigma_1^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2\sigma_2^2} \sim F(n_1,n_2)$

1.3.3 一个指数总体的抽样分布

1. $2\lambda n\bar{X} \sim \chi^2(2n)$

1.3.4 一个二项总体的抽样分布

1. $\frac{X-np}{\sqrt{np(1-p)}} \sim N(0,1)$
2. $\frac{p_s-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$

1.3.5 一个非正态总体均值的抽样分布

1. $\frac{\bar{X}-E(X)}{\sqrt{D(X)/n}} \sim N(0,1)$
2. $\frac{\bar{X}-E(X)}{\sqrt{S^2/n}} \sim N(0,1)$

1.3.6 两个总体的组合的抽样分布

1. $\chi^2(n_1)+\chi^2(n_2)=\chi^2(n_1+n_2)$
2. $\frac{\chi^2(n_1)/n_1}{\chi^2(n_2)/n_2}=F(n_1,n_2)$
3. $\frac{N(0,1)}{\sqrt{\chi^2(n)/n}}=t(n)$

2 参数估计

2.1 点估计

2.1.1 矩估计

1. A_k 估计 a_k，$A_k=a_k$
2. B_k 估计 b_k，$B_k=b_k$

2.1.2 极大似然估计

1. 似然函数取对数，再求导
2. 前后项比较，求出极值点
3. 边界条件与极小极大统计量的关系

2.1.3 评价估计量好坏的标准

无偏性：$E(\hat{\theta})=\theta$

有效性：

一致性：$\hat{\theta}$ 依概率收敛于 θ

2.2 区间估计

2.2.1 一个总体的置信区间

$P\{k_1 < \theta < k_2\} = 1 - \alpha$

2.2.2 两个总体的置信区间

1. $P\{k_1 < \theta_1 - \theta_2 < k_2\} = 1 - \alpha$
2. $P\{k_1 < \theta_1/\theta_2 < k_2\} = 1 - \alpha$

2.3 Bayes 估计

2.3.1 核

1. $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$：$X \sim N\left(\mu, \sigma^2\right)$
2. $x^{\alpha-1}e^{-\lambda x}$：$X \sim \Gamma(\alpha, \lambda)$
3. $x^{-\alpha-1}e^{-\frac{\lambda}{x}}$：$X \sim I\Gamma(\alpha, \lambda)$
4. $x^{\alpha-1}(1-x)^{\beta-1}$：$X \sim B(\alpha, \beta)$

2.3.2 损失函数

1. $(\theta - d)^2$：$\hat{\theta} = E(\theta)$
2. $\lambda(\theta)(\theta - d)^2$：$\hat{\theta} = \frac{E[\theta\lambda(\theta)]}{E[\lambda(\theta)]}$

author: virgilwjj

3 假设检验

3.1 拒绝域

H_0	H_1	拒绝域
$a = a_0$	$a \neq a_0$	$\hat{a} \neq a_0$
$a = a_0$	$a > a_0$	$\hat{a} > a_0$
$a = a_0$	$a = a_1(a_0 < a_1)$	$\hat{a} > a_0$
$a \leqslant a_0$	$a > a_0$	$\hat{a} > a_0$
$a = a_0$	$a < a_0$	$\hat{a} < a_0$
$a = a_0$	$a = a_1(a_0 > a_1)$	$\hat{a} < a_0$
$a \geqslant a_0$	$a < a_0$	$\hat{a} < a_0$

3.2 两类错误

3.1.1 第一类错误 弃真

$P\{拒绝了H_0 \mid H_0为真\} = \alpha$

3.1.2 第二类错误 采假

$P\{接受了H_0 \mid H_0为假\} = \beta$

3.2 参数检验

3.3 非参数检验

3.3.1 χ^2 检验

H_0：$P(X) = P_0(X)$

H_1：$P(X) \neq P_0(X)$

检验统计量：$K^2 = \sum_{i=1}^{k} \frac{n}{\hat{p}_i}\left(\frac{v_i}{n} - \hat{p}_i\right)^2 = \sum_{i=1}^{k} \frac{(v_i - n\hat{p}_i)^2}{n\hat{p}_i} = \frac{1}{n}\sum_{i=1}^{k} \frac{v_i^2}{\hat{p}_i} - n$
拒绝域：$K^2 > \chi_\alpha^2(k - r - 1)$

r：未知的参数的个数，即需要做点估计的参数的个数；不需要做点估计的参数或题目告诉你的，算已知。

3.3.2 χ^2 分析

H_0：$P(AB) = P(A)P(B)$

H_1：$P(AB) \neq P(A)P(B)$

检验统计量：$K^2 = \sum_{i=1}^{s} \sum_{j=1}^{t} \frac{(n_{ij} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j}$

拒绝域：$K^2 > \chi_\alpha^2((s-1)(t-1))$

当 $2 * 2$ 时：

检验统计量：$K^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}$

拒绝域：$K^2 > \chi_\alpha^2(1)$

3.3.3 秩和检验

检验统计量：第二个样本的秩和 W

拒绝域：

1. $F(x)$，$G(x)$ 是两个总体分布函数

H_0	H_1	拒绝域
$F(x) \leqslant G(x)$	$F(x) > G(x)$	$W \geqslant d$
$F(x) \geqslant G(x)$	$F(x) < G(x)$	$W \leqslant c$
$F(x) = G(x)$	$F(x) \neq G(x)$	$W \leqslant c \cup W \geqslant d$

2. μ_1，μ_2 是两个总体的均值

H_0	H_1	拒绝域
$\mu_1 \geqslant \mu_2$	$\mu_1 < \mu_2$	$W \geqslant d$
$\mu_1 \leqslant \mu_2$	$\mu_1 > \mu_2$	$W \leqslant c$
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$W \leqslant c \cup W \geqslant d$

建立服从正态分布检验量：$R_1 \sim N\left(\frac{n_1(n1+n2+1)}{2}, \frac{n_1 n_2(n1+n2+1)}{12}\right)$

R_1 为第一个样本的秩和

3.3.4 符号检验

单样本：与中位数的差的绝对值的秩和检验

双样本：对应的差的绝对值的秩和检验

4 方差分析

4.1 方差分析的常用统计量

误差方差估计：$\hat{\sigma}^2 = \frac{RSS}{n-r}$

总平方和：$TSS = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = (n-1)S^2$

自变量平方和：$CSS = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2$

残差平方和：$RSS = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 = \sum_{i=1}^{r}(n_i-1)S_i^2$

性质：

1. $TSS = CSS + RSS$
2. $\frac{RSS}{\sigma^2} = \frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-r)$
3. $\frac{CSS}{\sigma^2} \sim \chi^2(r-1)$

4.2 方差分析

H_0：$\mu_1 = \mu_2 = \ldots = \mu_r$

H_1：$\mu_1, \mu_2, \ldots, \mu_r$ 不完全相等

检验统计量：$F = \frac{CSS/(r-1)}{RSS/(n-r)}$

拒绝域：$F > F(r-1, n-r)$

方差来源	平方和	自由度	均方	F
分类变量	CSS	$r-1$	$CSS/(r-1)$	$\frac{CSS/(r-1)}{RSS/(n-r)}$
残差变量	RSS	$n-r$	$RSS/(n-r)$	
总计	TSS	$n-1$	$TSS/(n-1)$	

5 线性回归模型

5.1 一元线性回归

5.1.1 一元回归分析

$Y = X\beta + \varepsilon$

$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \ldots & \ldots \\ 1 & x_n \end{bmatrix}$

$\beta = [\beta_0 \quad \beta_1]^T$

$\varepsilon \sim N(0, \sigma^2 I_n)$

$Y \sim N(X\beta, \sigma^2 I_n)$

$S = X^T X$

$\hat{\beta} = [\hat{\beta}_0 \quad \hat{\beta}_1]^T$

$\hat{\beta} \sim N(\beta, \sigma^2 S^{-1})$

$S^{-1} = \frac{1}{L_{xx}}\begin{bmatrix} \frac{\sum_{i=1}^{n} x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$

5.1.1 最小二乘法

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$

回归系数估计：$\hat{\beta}_1 = \frac{L_{xy}}{L_{xx}}$

误差方差估计：$\hat{\sigma}^2 = \frac{RSS}{n-2}$

总平方和：$TSS = \sum_{i=1}^{n}(y_i-\bar{y})^2 = L_{yy}$

回归平方和：$RegSS = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 = \frac{L_{xy}^2}{L_{xx}}$

残差平方和 $RSS = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2$

相关系数 $r^2=\frac{RegSS}{TSS}=\frac{L_{xy}^2}{L_{xx}L_{yy}}$

性质：

1. $TSS=RegSS+RSS$
2. $\frac{RSS}{\sigma^2}=\frac{(n-2)\hat{\sigma}^2}{\sigma^2}\sim\chi^2(n-2)$
3. $\frac{RegSS}{\sigma^2}\sim\chi^2(1)$
4. $\hat{\beta}_0\sim N\left(\beta_0,\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{L_{xx}}\right)\right)=N\left(\beta_0,\frac{\sigma^2\sum_{i=1}^n x_i^2}{nL_{xx}}\right)$
5. $\hat{\beta}_1\sim N\left(\beta_1,\frac{\sigma^2}{L_{xx}}\right)$
6. $\hat{\beta}_0$ 与 $\hat{\beta}_1$ 不独立，协方差为 $Cov\left(\hat{\beta}_0,\hat{\beta}_1\right)=-\sigma^2\frac{\bar{x}}{L_{xx}}$

5.1.2 回归关系检验—— F 检验法

H_0：$\beta_1=0$

H_1：$\beta_1\neq0$

检验统计量：$F=\frac{RegSS}{RSS/(n-2)}=\frac{(n-2)L_{xx}^2}{L_{xx}L_{yy}-L_{xy}^2}=\frac{(n-2)r^2}{1-r^2}$

拒绝域：$F>F(1,n-2)$

方差来源	平方和	自由度	均方	F
回归变量	$RegSS$	1	$RegSS$	$\frac{RegSS}{RSS/(n-2)}$
残差变量	RSS	$n-2$	$RSS/(n-2)$	
总计	TSS	$n-1$	$TSS/(n-1)$	

5.1.3 回归关系检验—— t 检验法

$\hat{\beta}_1\sim N\left(\beta_1,\frac{\sigma^2}{L_{xx}}\right)$

$\frac{(n-2)\hat{\sigma}^2}{\sigma^2}\sim\chi^2(n-2)$

检验统计量：$t=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}}\sqrt{L_{xx}}\sim t(n-2)$

拒绝域：$|t|>t_{\alpha/2}(n-2)$

5.1.4 回归关系检验—— r 检验法

检验统计量：$r=\sqrt{\frac{RegSS}{TSS}}=\sqrt{\frac{L_{xy}^2}{L_{xx}L_{yy}}}$

拒绝域：$|r|>r_\alpha(n-2)$

5.1.5 利用回归方程进行预测（y_0 的区间估计，x_0 对区间的控制）

$\hat{y}_0=\hat{\beta}_0+\hat{\beta}_1x_0=(1,x_0)(\hat{\beta}_0,\hat{\beta}_1)^T$

$\hat{y}_0\sim N(\beta_0+\beta_1x_0,\sigma^2[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{L_{xx}}])$

$y_0=\hat{\beta}_0+\hat{\beta}_1x_0+\varepsilon_0$

$y_0\sim N(\beta_0+\beta_1x_0,\sigma^2)$

$y_0-\hat{y}_0\sim N(0,\sigma^2[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{L_{xx}}])$

$\frac{(n-2)\hat{\sigma}^2}{\sigma^2}\sim\chi^2(n-2)$

检验统计量：$t=\frac{y_0-\hat{y}_0}{\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{L_{xx}}}}\sim t(n-2)$

置信区间：$|t|<t_{\alpha/2}(n-2)$

5.1.6 β_0 的区间估计

$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{L_{xx}}\right)\right)$

$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$

检验统计量：$t = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{L_{xx}}}} \sim t(n-2)$

置信区间：$|t| < t_{\alpha/2}(n-2)$