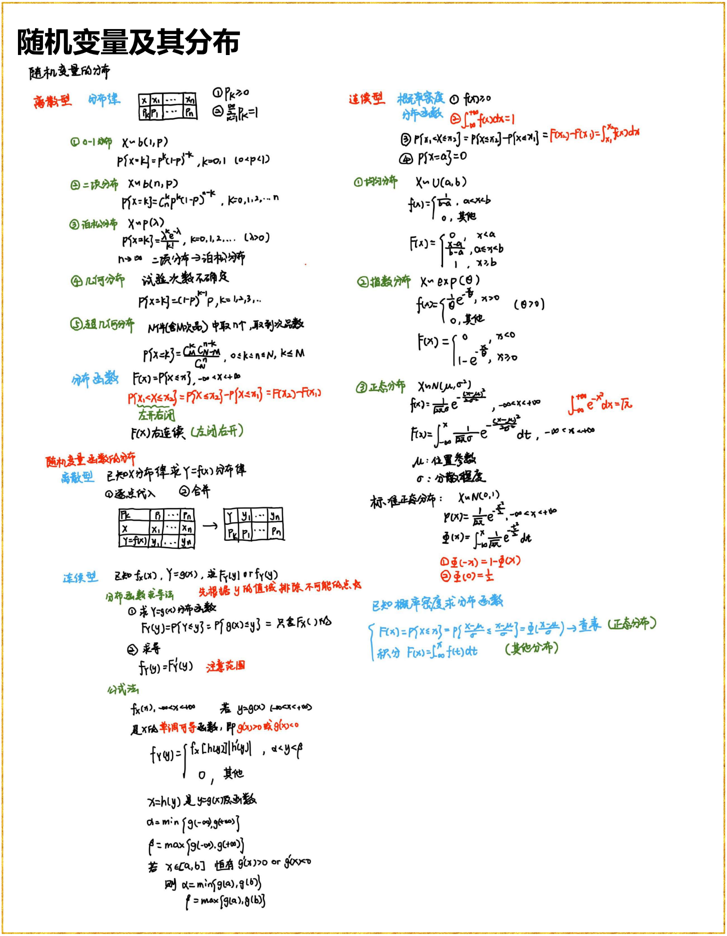
背公式

2024年12月1日 星期日　20:13



随机变量及其分布

随机变量的分布

离散型　分布律

X	x_1	$\cdots$	x_n
P_k	p_1	$\cdots$	p_n

① $P_k \geq 0$

② $\sum_{k=1}^{\infty} P_k = 1$

① 0-1分布　$X \sim b(1,p)$

$P\{X=k\} = p^k(1-p)^{1-k}, k=0,1 \quad (0<p<1)$

② 二项分布　$X \sim b(n,p)$

$P\{X=k\} = C_n^k p^k (1-p)^{n-k}, \ k=0,1,2,\cdots n$

③ 泊松分布　$X \sim P(\lambda)$

$P\{X=k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \ k=0,1,2,\cdots \quad (\lambda>0)$

$n \to \infty$　二项分布 → 泊松分布

④ 几何分布　试验次数不确定

$P\{X=k\} = (1-p)^{k-1} p, \ k=1,2,3,\ldots$

⑤ 超几何分布　N件(含M次品)中取n个，取到次品数

$P\{X=k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \ 0 \leq k \leq n \leq N, \ k \leq M$

分布函数　$F(x) = P\{X \leq x\}, -\infty < x < +\infty$

$P\{\underbrace{x_1 < X \leq x_2}_{\text{左开右闭}}\} = P\{X \leq x_2\} - P\{X \leq x_1\} = F(x_2) - F(x_1)$

$F(x)$右连续（左闭右开）

随机变量函数的分布

离散型　已知X分布律求$Y=f(x)$分布律

①逐点代入　②合并

P_k	p_1	$\cdots$	p_n
X	x_1	$\cdots$	x_n
$Y=f(X)$	y_1	$\cdots$	y_n

→

Y	y_1	$\cdots$	y_n
P_k	p_1	$\cdots$	p_n

连续型　已知 $f_X(x)$, $Y=g(X)$, 求 $F_Y(y)$ or $f_Y(y)$

分布函数求导法　先根据 y 的值域排除不可能的点☆

① 求 $Y=g(x)$ 分布函数

$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\}$ = 只含$F_X(\ )$的

② 求导

$f_Y(y) = F_Y'(y)$　注意范围

公式法

$f_X(x), -\infty<x<+\infty$　若 $y=g(x) \ (-\infty<x<+\infty)$

是x的单调可导函数，即 $g'(x)>0$ 或 $g'(x)<0$

$$f_Y(y) = \begin{cases} f_X[h(y)]\,|h'(y)|, & \alpha<y<\beta \\ 0, & 其他 \end{cases}$$

$x=h(y)$ 是 $y=g(x)$ 反函数

$\alpha = \min\{g(-\infty), g(+\infty)\}$

$\beta = \max\{g(-\infty), g(+\infty)\}$

若 $x \in [a,b]$ 恒有 $g'(x)>0$ or $g'(x)<0$

则 $\alpha = \min\{g(a), g(b)\}$

$\beta = \max\{g(a), g(b)\}$

连续型　概率密度　分布函数

① $f(x) \geq 0$

② $\int_{-\infty}^{+\infty} f(x)dx = 1$

③ $P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx$

④ $P\{X=a\} = 0$

① 均匀分布　$X \sim U(a,b)$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a<x<b \\ 0, & 其他 \end{cases}$$

$$F(x) = \begin{cases} 0, & x<a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

② 指数分布　$X \sim \exp(\theta)$

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x>0 \\ 0, & 其他 \end{cases} \quad (\theta>0)$$

$$F(x) = \begin{cases} 0, & x<0 \\ 1-e^{-\frac{x}{\theta}}, & x \geq 0 \end{cases}$$

③ 正态分布　$X \sim N(\mu, \sigma^2)$

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty<x<+\infty$　　$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$

$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dt, \ -\infty<x<+\infty$

μ：位置参数

σ：分散程度

标准正态分布：　$X \sim N(0,1)$

$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty<x<+\infty$

$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

① $\Phi(-x) = 1 - \Phi(x)$

② $\Phi(0) = \frac{1}{2}$

已知概率密度求分布函数

$$\begin{cases} F(x) = P\{X \leq x\} = P\{\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\} = \Phi(\frac{x-\mu}{\sigma}) \rightarrow \text{查表（正态分布）} \\ \text{积分}\ F(x) = \int_{-\infty}^{x} f(t)dt \quad \text{（其他分布）} \end{cases}$$

多维随机变量及其分布

联合分布函数

$F(x,y)=P\{(X\le x)\cap(Y\le y)\}=P\{X\le x,\ Y\le y\}$

性质：① $F(x,y)$ 不减

② $\lim\limits_{x\to-\infty}F(x,y)=0$　　$\lim\limits_{\substack{x\to-\infty\\ y\to-\infty}}F(x,y)=0$

$\lim\limits_{y\to-\infty}F(x,y)=0$　　$\lim\limits_{\substack{x\to+\infty\\ y\to+\infty}}F(x,y)=0$

$\left.\begin{array}{l}\lim\limits_{x\to+\infty}F(x,y)\\ \lim\limits_{y\to+\infty}F(x,y)\end{array}\right\}$ 不确定

③ 右连续

二维离散型

联合分布律

Y \ X	x_1	…	x_n
y_1			
⋮			
y_n			

$P\{X=x_i,y=y_i\}=P_{ij}\quad ,ij=1,2,\cdots$

① $P_{ij}\ge0$　② $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}P_{ij}=1$

$F(x,y)=P\{X\le x,Y\le y\}=\sum\limits_{x_i\le x}\sum\limits_{y_j\le y}P\{X=x_i,Y=y_j\}=\sum\limits_{x_i\le x}\sum\limits_{y_j\le y}P_{ij}$

二维连续型

联合概率密度

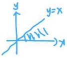
$f(x,y)$ 使　$F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x}f(u,v)\,du\,dv,\quad -\infty<x<+\infty\quad -\infty<y<+\infty$

① $f(x,y)\ge0$

② $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,dx\,dy=1$

③ $P\{(X,Y)\in G\}=\iint\limits_{G}f(x,y)\,dx\,dy$

例：$f(x,y)=\begin{cases}2e^{-(2x+y)}, & x>0,y>0\\ 0, & 其他\end{cases}$　　求 $P\{Y\le X\}$



$P\{Y\le X\}=\iint\limits_{G}f(x,y)\,dx\,dy=\int_0^{+\infty}dx\int_0^{x}2e^{-(2x+y)}dy$

$=\int_0^{+\infty}2e^{-2x}\int_0^{x}e^{-y}dy=\frac{1}{3}$

二维正态分布

$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

边缘分布 $\begin{cases}X\sim N(\mu_1,\sigma_1^2)\\ Y\sim N(\mu_2,\sigma_2^2)\end{cases}$

$\rho=0$ 时，X.Y 相互独立

二维均匀分布

G 的面积为 A，概率密度为

$f(x,y)=\begin{cases}\frac{1}{A}, & (x,y)\in G\\ 0, & 其他\end{cases}$

边缘分布函数

$F_X(x)=P\{X\le x\}=\lim\limits_{y\to+\infty}F(x,y)$

$F_Y(y)=P\{Y\le y\}=\lim\limits_{x\to+\infty}F(x,y)$

离散型　边缘分布律

$P\{X=x_i\}=\sum_{j=1}^{\infty}P_{ij}$

$P\{Y=y_j\}=\sum_{i=1}^{\infty}P_{ij}$

连续型　边缘概率密度

$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,dy$

$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\,dx$

条件分布函数

离散型　条件分布律

$P\{X=x_i\mid Y=y_j\}=\dfrac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\dfrac{P_{ij}}{\sum_{i=1}^{\infty}P_{ij}}\ ,i=1,2,\cdots$

$P\{X=x_i\mid Y=y_j\}=\dfrac{P\{X=x_i,Y=y_j\}}{P\{X=x_i\}}=\dfrac{P_{ij}}{\sum_{j=1}^{\infty}P_{ij}}\ ,j=1,2,\cdots$

连续型　条件概率密度

$f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}=\dfrac{联合概率密度}{边缘概率密度}$

$f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}$

条件分布函数

$Y=y$ 时，$F_{X|Y}(x|y)=P\{X\le x\mid Y=y\}=\int_{-\infty}^{x}\dfrac{f(x,y)}{f_Y(y)}dx$

$X=x$ 时，$F_{Y|X}(y|x)=P\{X=x\mid Y\le y\}=\int_{-\infty}^{y}\dfrac{f(x,y)}{f_X(x)}dy$

相互独立的随机变量

离散型

$P\{X=x_i,Y=y_j\}=P\{X=x_i\}\cdot P\{Y=y_j\}\quad \forall i,j$

① 联合分布律 = 边缘分布律乘积

② $i=1,2,\ldots m$　$n\times m$ 个等式全成立
$j=1,2,\ldots n$

连续型

$f(x,y)=f_X(x)f_Y(y)\quad \forall x,y$

随机变量的数字特征

数学期望

离散型　$E(X)=\sum_{k=1}^{\infty}x_kP_k$

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

一维随机变量的函数　$Y=g(X)$

离散型　$E(Y)=\sum_{k=1}^{\infty}g(x_k)P_k$

连续型　$E(Y)=\int_{-\infty}^{+\infty}g(x)f(x)\,dx$

二维随机变量的函数　$Z=g(X,Y)$

离散型　$E(Y)=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}g(x_i,y_j)P_{ij}$

可先求分布律再求 E(Z)

连续型　$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\,dx\,dy$

性质：① $E(C)=C$

方差

$D(X)=E\{[X-E(X)]^2\}$

计算公式　$D(x)=E(x^2)-E^2(x)$

离散型　$D(X)=\sum_{k=1}^{\infty}[X_k-E(X)]^2P_k$

① 求 $E(X)$

② 求 X^2 分布律

③ $D(X)=E(X^2)-E^2(X)$

连续型　$D(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2f(x)\,dx$

① $E(X)=\int_{-\infty}^{+\infty}xf(x)\,dx$

② $E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)\,dx$

③ $D(X)=E(X^2)-E^2(X)$

性质：① $D(C)=0$（C为常数）

② $D(CX)=C^2D(X)$　$D(X+C)=D(X)$

③ $D(X+Y)=D(X)+D(Y)+2E\{[X-E(X)][Y-E(Y)]\}$

② $E(CX)=CE(X)$

③ $E(X+Y)=E(X)+E(Y)$

④ X、Y相互独立 ⇒ $E(XY)=E(X)E(Y)$

(0-1)分布 $X\sim b(1,p)$

$E(X)=p \quad D(X)=p(1-p)$

二项分布 $X\sim b(n,p)$

$E(X)=np \quad D(X)=np(1-p)$

泊松分布 $X\sim \pi(\lambda)$

$E(X)=\lambda \quad D(X)=\lambda$

均匀分布 $X\sim U(a,b)$

$E(X)=\frac{a+b}{2} \quad D(X)=\frac{(b-a)^2}{12}$

指数分布 $X\sim \exp(\theta)$

$E(X)=\theta \quad D(X)=\theta^2$

正态分布 $X\sim N(\mu,\sigma^2)$

$E(X)=\mu \quad D(X)=\sigma^2$

切比雪夫不等式（只知 $E(X)$、$D(X)$，不知分布）

$E(X)=\mu \quad D(X)=\sigma^2$，对任意正数 ε，有

$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}$ （记住它 $\geq \varepsilon \leq$）

变1：$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}$

变2：$P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{D(X)}{\varepsilon^2}$

矩

1. k阶（原点）矩

$E(X^k) \quad k=1,2,\cdots$

$E(X)$ 是一阶矩

2. k阶中心矩

$E\{[X-E(X)]^k\}, \quad k=2,3\ldots$

$D(X)$ 是二阶中心矩

3. k+l阶 混合矩

$E(X^kY^l) \quad k,l=1,2,\ldots$

4. k+l阶 混合中心矩

$E\{[X-E(X)]^k[Y-E(Y)]^l\} \quad k,l=1,2,\cdots$

$Cov(X,Y)$ 协方差是二阶混合中心矩

$=D(X)+D(Y)+2[E(XY)-E(X)E(Y)]$

若X、Y独立，则 $D(X+Y)=D(X)+D(Y)$

$D(X-Y)=D(X)\underline{+}D(Y)$

协方差及相关系数

$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$ 定义式

$=E(XY)-E(X)E(Y)$ 计算式

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

Cov 性质：① $Cov(X,X)=D(X)$

② $Cov(X,Y)=Cov(Y,X)$

$Cov(X,C)=0$

③ $Cov(aX,bY)=ab\,Cov(X,Y)$

④ $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

⑤ $D(X+Y)=D(X)+D(Y)+2Cov(X,Y)$

ρ_{XY} 性质：① $|\rho_{XY}|\leq 1$

存在线性关系

② $|\rho_{XY}|=1 \Leftrightarrow$ 存在a、b使 $P\{Y=a+bX\}=1$

正相关：$\rho_{XY}=1 \Rightarrow b>0$

负相关：$\rho_{XY}=-1 \Rightarrow b<0$

$|\rho_{XY}|\uparrow$ 相关性好

$\rho_{XY}=0 \Leftrightarrow$ 不相关

X、Y相互独立 $\Leftrightarrow F(x,y)=F_X(x)F_Y(y) \quad \forall x,y$ 分布函数

$\Leftrightarrow p_{ij}=\sum_{j=1}^{\infty}p_{ij}\cdot\sum_{i=1}^{\infty}p_{ij} \quad \forall i,j$ 分布律

$\Leftrightarrow f(x,y)=f_X(x)\cdot f_Y(y) \quad \forall x,y$ 概率密度

X、Y不相关 $\Leftrightarrow \rho_{XY}=0$

$\Leftrightarrow Cov(X,Y)=0$

$\Leftrightarrow E(XY)=E(X)E(Y)$

$\Leftrightarrow D(X+Y)=D(X)+D(Y)$

对于正态分布，不相关 ⇔ 相互独立

大数定律及中心极限定理

辛钦大数定理

$\lim_{n\to\infty}P\{|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu|<\varepsilon\}=1 \quad \bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{P}\mu$

伯努利大数定理

$\lim_{n\to\infty}P\{|\frac{f_A}{n}-p|<\varepsilon\}=1 \quad \frac{f_A}{n}\xrightarrow{P}p$

中心极限定理（核心：向正态分布变换）

定理1.（独立同分布）

$E(X_k)=\mu \quad D(X_k)=\sigma^2>0$

$\sum_{k=1}^{n}X_k$ 标准化变量为

$Y_n=\frac{\sum_{k=1}^{n}X_k-E(\sum_{k=1}^{n}X_k)}{\sqrt{D(\sum_{k=1}^{n}X_k)}}=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$

$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\{\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\leq x\}=\Phi(x)$

$\Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

定理3. 棣莫弗-拉普拉斯定理

$\lim_{n\to\infty}P\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\leq x\}=\Phi(x)$

$\Rightarrow \frac{\eta_n-np}{\sqrt{np(1-p)}}\sim N(0,1)$

拆解二项分布：

(1) 当 $n\leq 5$，直接算

$P\{X=k\}=C_n^kp^k(1-p)^{n-k}, \quad k=0,1,2,\cdots,n$

(2) 当n很大，p很小，且 $np=\lambda<10$ 时，用泊松定理

$P\{X=k\}\approx\frac{\lambda^k\cdot e^{-\lambda}}{k!}, \quad k=0,1,2,\cdots,n, \quad \lambda=np$

(3) 当n很大（$n\geq 50$），p不是很小时，用棣-拉

$\frac{X-np}{\sqrt{np(1-p)}}\overset{\text{近似}}{\sim}N(0,1)$

例 1 一加法器同时收到 20 个噪声电压 $V_k(k=1,2,\cdots,20)$，设它们是相互独立的随机变量，且都在区间 $(0,10)$ 上服从均匀分布，记 $V=\sum_{k=1}^{20}V_k$，求 $P\{V>105\}$ 的近似值.

$X\sim U(0,10)\Rightarrow E(V_k)=5 \quad D(V_k)=\frac{(b-a)^2}{12}=\frac{100}{12}$

相互独立

$E(\sum_{k=1}^{20}V_k)=n\mu=20\times5=100$

$D(\sum_{k=1}^{20}V_k)=20\times\frac{100}{12}$

$V=\frac{\sum_{k=1}^{20}V_k-E(\sum_{k=1}^{20}V_k)}{\sqrt{D(\sum_{k=1}^{20}V_k)}}=\frac{V-100}{10\sqrt{\frac{5}{3}}}$

$P\{V>105\}=P\{\frac{V-100}{10\sqrt{\frac{5}{3}}}>\frac{105-100}{10\sqrt{\frac{5}{3}}}\}=1-\Phi(0.387)=0.348$

例 2 一船舶在某海区航行，已知每遭受一次波浪的冲击，纵摇角大于 3°的概率为 $p=1/3$，若船舶遭受了 90 000 次波浪冲击，问其中有 29 500～30 500 次纵摇角度大于 3°的概率是多少？

$X\sim b(90000,\frac{1}{3})$

$E(X)=np=30000$

$D(X)=np(1-p)=20000$

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

$P\{29500\leq X\leq 30500\}=P\{-\frac{5}{\sqrt2}\leq\frac{X-30000}{100\sqrt2}\leq\frac{5}{\sqrt2}\}$

$=2\Phi(\frac{5}{\sqrt2})-1=0.9995$

样本及抽样分布

统计量

$x_1, x_2, \dots x_n$ 是 $X_1, X_2, \dots X_n$ 的样本值

则 $g(x_1, x_2, \dots x_n)$ 是 $g(X_1, X_2, \dots X_n)$ 的观察值（可计算出结果）

常用的统计量

样本平均值 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$

样本标准差 $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$

样本k阶原点矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$，$k=1,2,\dots$

$A_1 = \bar{X}$

样本k阶中心矩 $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$　$k=2,3,\dots$

$B_1 = \frac{1}{n}\sum_{i=1}^{n}(X-\bar{X})^2 \neq S^2$

经验分布

$F_n(x) = \frac{1}{n}S(x) = \frac{x_1, x_2, \dots x_n\text{中不大于}x\text{的个数}}{n}$，$x \in (-\infty, +\infty)$

按大小排列样本值，根据定义写出分布函数

三大抽样分布

(χ^2) N方和卡：$X_1, \dots X_n$ 服从 $N(0,1)$，则 $\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$

性质：① 可加 $X \sim \chi^2(n_1)$ $Y \sim \chi^2(n_2) \Rightarrow X+Y \sim \chi^2(n_1+n_2)$

② $E(\chi^2(n)) = n$

$D(\chi^2(n)) = 2n$

(t) N开卡t：$X \sim N(0,1)$ $Y \sim \chi^2(n)$ X、Y相互独立，则 $\frac{X}{\sqrt{Y/n}} \sim t(n)$

性质：① $t_{1-\alpha}(n) = -t_\alpha(n)$

(F) 卡除自F：$X \sim \chi^2(n_1)$ $Y \sim \chi^2(n_2)$，X、Y相互独立，则 $\frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$

（自由度）

性质：① $F(n_1, n_2) = \frac{1}{F(n_2, n_1)}$

② $F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}$

正态总体样本： 总体X，$E(X)=\mu$，$D(X)=\sigma^2$

样本 $X_1, X_2, \dots X_n$，$\bar{X}$为样本均值，S^2为样本方差

① $E(\bar{X}) = \mu$

② $D(\bar{X}) = \frac{\sigma^2}{n}$

$\Rightarrow$ ④ $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

③ $E(S^2) = D(X) = \sigma^2$

⑤ $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

⑥ $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

参数估计

点估计

矩估计： 样本矩 = 总体矩，反解留 $\bar{X}$ 加尖，量大写，值计算

1个未知参数 θ：

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ = $E(X) = f(\theta)$　$\theta = \varphi(E(X)) = \varphi(\bar{x})$

$\hat{\theta} = \varphi(\bar{X})$ 矩估计量

数值代入 矩估计值

2个未知参数 a, b：

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ = $E(x) = f_1(a,b)$　$a = \varphi_1(\bar{x})$

$\frac{1}{n}\sum_{i=1}^{n} x_i^2$ = $E(X^2) = f_2(a,b)$　$b = \varphi_2(\bar{x})$

$$\begin{cases} \hat{a} = \varphi_1(\bar{X}) \\ \hat{b} = \varphi_2(\bar{X}) \end{cases}$$

矩估计量

数值代入 矩估计值

注：$E(X^2)$不好算，常代换 $E(X^2) = \underbrace{E^2(X)}_{\bar{X}^2} + D(X)$

得到 $\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \bar{X}^2 + D(X)$

$\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 \Rightarrow \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = D(X)$

2个未知参数

设总体X在[a, b]服从均匀分布，$X_1, X_2, \dots X_n$是来自X的样本，a, b未知，求a, b的矩估计量

$X \sim U[a,b]$　$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{其它} \end{cases}$

$E(X) = \frac{a+b}{2}$　$D(X) = \frac{(b-a)^2}{12}$

①
$$\begin{cases} E(X) = \frac{a+b}{2} \\ E(X^2) = D(X) + E^2(X) = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2 \end{cases}$$

② 令
$$\begin{cases} A_1 = \bar{X} = E(X) \\ A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = E(X^2) \end{cases} \Rightarrow \begin{cases} \bar{X} = \frac{a+b}{2} & ① \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \left(\frac{a+b}{2}\right)^2 + \frac{(b-a)^2}{12} & ② \end{cases}$$

①代入②中　$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \bar{X}^2 + \frac{(b-a)^2}{12}$

$\Rightarrow b-a = \sqrt{\frac{12}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$

由①知 $b+a = 2\bar{X}$

$$\begin{cases} b-a = \sqrt{\frac{12}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2} \\ a+b = 2\bar{X} \end{cases} \Rightarrow \begin{cases} \hat{a} = \bar{X} - \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2} \\ \hat{b} = \bar{X} + \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2} \end{cases}$$

最大似然估计

全乘取对，导数为零，值小写，量大写

离散型：

$L(x_1, x_2, \dots x_n; \theta_1, \theta_2, \dots \theta_m) = \prod_{i=1}^{n} P\{X = x_i\}$

连续型：

$L(x_1, x_2, \dots x_n; \theta_1, \theta_2, \dots \theta_m) = \prod_{i=1}^{n} f(x_i)$

$\ln L(\theta)$

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1} = 0 \\ \frac{\partial \ln L}{\partial \theta_2} = 0 \\ \dots \end{cases} \Rightarrow \begin{cases} \hat{\theta}_1 = \varphi_1(x_1 \dots x_n) \\ \hat{\theta}_2 = \varphi_2(x_1 \dots x_n) \end{cases}$$

最大似然估计值

$\Downarrow$

$$\begin{cases} \hat{\theta}_1 = \varphi_1(X_1, \dots X_n) \\ \hat{\theta}_2 = \varphi_2(X_1, \dots X_n) \end{cases}$$

最大似然估计量

若无解则利用单调性

估计量评选标准

无偏　$\hat{\theta}$为θ的估计量，$E(\hat{\theta})=\theta$

有效　$D(\hat{\theta}_1)<D(\hat{\theta}_2)$　$\hat{\theta}_1$较$\hat{\theta}_2$有效

相合　$n\to\infty$，$\hat{\theta}\xrightarrow{P}\theta$

（基本要求）　样本k阶矩A_k是总体k阶矩$E(X^k)$的相合估计量

样本均值$\bar{X}$是总体均值$E(X)$的相合估计量

样本方差S^2是总体方差$D(X)$的相合估计量

区间估计

置信区间（找枢轴量，列不等式，不用背置信区间，置信水平为$1-\alpha$）

单个总体

估μ知σ N限Z　求μ，σ已知 $\Rightarrow -Z_{\frac{\alpha}{2}}<\overset{N限Z}{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}<Z_{\frac{\alpha}{2}} \Rightarrow \left(\bar{X}\pm\frac{\sigma}{\sqrt{n}}Z_{\frac{\alpha}{2}}\right)$

σ若未知需St　求μ，σ未知 $\Rightarrow -t_{\frac{\alpha}{2}}(n-1)<\frac{\bar{X}-\mu}{S/\sqrt{n}}<t_{\frac{\alpha}{2}}(n-1) \Rightarrow \left(\bar{X}\pm\frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)\right)$

σ.μ不知方除　求σ，μ未知 $\Rightarrow \chi^2_{1-\frac{\alpha}{2}}(n-1)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{\frac{\alpha}{2}}(n-1) \Rightarrow \left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$

$\mu_1-\mu_2$ ｛置信下限大于0：$\mu_1>\mu_2$

　　　　　 置信区间包含0：μ_1，μ_2无显著变化

$\frac{\sigma_1^2}{\sigma_2^2}$：置信区间包含1：$\sigma_1^2$与$\sigma_2^2$无显著差别

减1　间2级牢记

$t(n-1)$　置信区间写$\frac{\alpha}{2}$

$\chi^2(n-1)$　单侧置信限写α

两个总体

μ差形同单　σ已知用Z，σ未知用t，单侧α 区间$\frac{\alpha}{2}$

σ方和开根　求$\mu_1-\mu_2$，σ_1^2.σ_2^2已知 $\Rightarrow -Z_{\frac{\alpha}{2}}<\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}<Z_{\frac{\alpha}{2}} \Rightarrow \left(\bar{X}-\bar{Y}\pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$

↖ $\sqrt{\left(\frac{\sigma_1}{\sqrt{n_1}}\right)^2+\left(\frac{\sigma_2}{\sqrt{n_2}}\right)^2}$

S_w算术平均　求$\mu_1-\mu_2$，$\sigma_1^2=\sigma_2^2=\sigma^2$未知 $\Rightarrow -t_{\frac{\alpha}{2}}(n_1+n_2-2)<\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}<t_{\frac{\alpha}{2}}(n_1+n_2-2) \Rightarrow \left(\bar{X}-\bar{Y}\pm t_{\frac{\alpha}{2}}(n_1+n_2-2)\cdot S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$

$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ 算术平均

σ除则用F　求$\frac{\sigma_1^2}{\sigma_2^2}$，$\mu_1$.$\mu_2$未知 $\Rightarrow F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)<\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}<F_{\frac{\alpha}{2}}(n_1-1,n_2-1) \Rightarrow \left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\frac{\alpha}{2}}(n_1-1,n_2-1)},\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)}\right)$

注：F是由χ^2分布除χ^2分布得到的，因此单个σ^2用χ，$\frac{\sigma_1^2}{\sigma_2^2}$用F。

假设检验

假设

H_0　原假设或零假设

H_1　备择假设

$H_0:\mu=\mu_0$　$H_1:\mu\neq\mu_0$　双边检验

$H_0:\mu\leq\mu_0$　$H_1:\mu>\mu_0$　右边检验

$H_0:\mu\geq\mu_0$　$H_1:\mu<\mu_0$　左边检验

显著性水平为α

检验统计量与置信区间枢轴量相同

双边$\frac{\alpha}{2}$单边α

单边符号同备择且带等

μ双边绝对值大于等于

σ 大于α ($\frac{\alpha}{2}$) 小于 $1-\alpha$ ($1-\frac{\alpha}{2}$)

	H_0	H_1	统计量	拒绝域
① σ^2已知	$\mu\leq\mu_0$ $\mu\geq\mu_0$ $\mu=\mu_0$	$\mu>\mu_0$ $\mu<\mu_0$ $\mu\neq\mu_0$	$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	$Z\geq Z_\alpha$ $Z\leq -Z_\alpha$ $\lvert Z\rvert\geq Z_{\frac{\alpha}{2}}$
② σ^2未知	$\mu\leq\mu_0$ $\mu\geq\mu_0$ $\mu=\mu_0$	$\mu<\mu_0$ $\mu<\mu_0$ $\mu\neq\mu_0$	$t=\frac{\bar{X}-\mu}{S/\sqrt{n}}$	$t\geq t_\alpha(n-1)$ $t\leq -t_\alpha(n-1)$ $\lvert t\rvert\geq t_{\frac{\alpha}{2}}(n-1)$
③ σ^2已知	$\mu_1-\mu_2\leq\delta$ $\mu_1-\mu_2\geq\delta$ $\mu_1-\mu_2=\delta$	$\mu_1-\mu_2>\delta$ $\mu_1-\mu_2<\delta$ $\mu_1-\mu_2\neq\delta$	$Z=\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$Z\geq Z_\alpha$ $Z\leq -Z_\alpha$ $\lvert Z\rvert\geq Z_{\frac{\alpha}{2}}$
④ σ^2未知 $\sigma_1^2=\sigma_2^2=\sigma^2$	$\mu_1-\mu_2\leq\delta$ $\mu_1-\mu_2\geq\delta$ $\mu_1-\mu_2=\delta$	$\mu_1-\mu_2>\delta$ $\mu_1-\mu_2<\delta$ $\mu_1-\mu_2\neq\delta$	$t=\frac{\bar{X}-\bar{Y}-\delta}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$	$t\geq t_\alpha(n_1+n_2-2)$ $t\leq -t_\alpha(n_1+n_2-2)$ $\lvert t\rvert\geq t_{\frac{\alpha}{2}}(n_1+n_2-2)$
⑤ μ未知	$\sigma^2\leq\sigma_0^2$ $\sigma^2\geq\sigma_0^2$ $\sigma^2=\sigma_0^2$	$\sigma^2>\sigma_0^2$ $\sigma^2<\sigma_0^2$ $\sigma^2\neq\sigma_0^2$	$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2\geq\chi^2_\alpha(n-1)$ $\chi^2\leq\chi^2_{1-\alpha}(n-1)$ $\chi^2\geq\chi^2_{\frac{\alpha}{2}}(n-1)$或$\chi^2\leq\chi^2_{1-\frac{\alpha}{2}}(n-1)$
⑥ μ_1.μ_2未知	$\sigma_1^2\leq\sigma_2^2$ $\sigma_1^2\geq\sigma_2^2$ $\sigma_1^2=\sigma_2^2$	$\sigma_1^2>\sigma_2^2$ $\sigma_1^2<\sigma_2^2$ $\sigma_1^2\neq\sigma_2^2$	$F=\frac{S_1^2}{S_2^2}$ 注：不是$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$	$F\geq F_\alpha(n_1-1,n_2-1)$ $F\leq F_{1-\alpha}(n_1-1,n_2-1)$ $F\geq F_{\frac{\alpha}{2}}(n_1-1,n_2-1)$或$F\leq F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)$

方差分析及回归分析

单因素试验

假设问题的拒绝域

$H_0: \mu_1=\mu_2=\cdots=\mu_s$ or $H_0: \delta_1=\delta_2=\cdots=\delta_s=0$

$H_1: \mu_1,\mu_2\ldots\mu_s$ 不全相等 $\quad H_1: \delta_1,\delta_2\ldots\delta_s$ 不全为零

$T_{\cdot j}=\sum_{i=1}^{n_j}X_{ij},\ j=1,2,\cdots,s \qquad T_{\cdot\cdot}=\sum_{j=1}^{s}\sum_{i=1}^{n_j}X_{ij}$

总偏差平方和 $S_T=\sum_{j=1}^{s}\sum_{i=1}^{n_j}X_{ij}^2-\frac{T_{\cdot\cdot}^2}{n}$

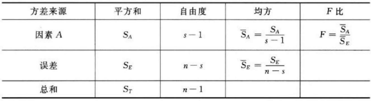
效应平方和 $S_A=\sum_{j=1}^{s}\frac{T_{\cdot j}^2}{n_j}-\frac{T_{\cdot\cdot}^2}{n}$

误差平方和 $S_E=S_T-S_A$

$$F=\frac{S_A/(s-1)}{S_E/(n-s)}=\frac{S_A/\sigma^2}{s-1}\Big/\frac{S_E/\sigma^2}{n-s}\sim F(s-1,n-s)$$

拒绝域 $F=\frac{S_A/(s-1)}{S_E/(n-s)}\geq F_\alpha(s-1,n-s)$

表 9—5　单因素试验的方差分析表



方差来源	平方和	自由度	均方	F 比
因素 A	S_A	$s-1$	$\bar{S}_A=\frac{S_A}{s-1}$	$F=\frac{\bar{S}_A}{\bar{S}_E}$
误差	S_E	$n-s$	$\bar{S}_E=\frac{S_E}{n-s}$	
总和	S_T	$n-1$		

参数的点估计和区间估计

$$\begin{cases}\hat{\sigma}=\frac{S_E}{n-s}\\ \hat{\mu}=\bar{X}\\ \hat{\mu}_j=\bar{X}_{\cdot j}\\ \hat{\delta}_j=\bar{X}_{\cdot j}-\bar{X}\end{cases}$$

均值差 $\mu_j-\mu_k=\delta_j-\delta_k$ 的置信区间（置信水平 $1-\alpha$）

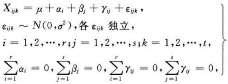
$$\frac{(\bar{X}_{\cdot j}-\bar{X}_{\cdot k})-(\mu_j-\mu_k)}{\sqrt{S_E(\frac{1}{n_j}+\frac{1}{n_k})}}=\frac{(\bar{X}_{\cdot j}-\bar{X}_{\cdot k})-(\mu_j-\mu_k)}{\sigma\sqrt{\frac{1}{n_j}+\frac{1}{n_k}}}\Big/\sqrt{\frac{S_E}{\sigma^2}/(n-s)}\sim t(n-s)$$

$$-t_{\frac{\alpha}{2}}(n-s)\leq\frac{(\bar{X}_{\cdot j}-\bar{X}_{\cdot k})-(\mu_j-\mu_k)}{\sqrt{S_E(\frac{1}{n_j}+\frac{1}{n_k})}}\leq t_{\frac{\alpha}{2}}(n-s)$$

$$\Rightarrow\ \left((\bar{X}_{\cdot j}-\bar{X}_{\cdot k})\pm t_{\frac{\alpha}{2}}(n-s)\sqrt{S_E(\tfrac{1}{n_j}+\tfrac{1}{n_k})}\right)$$

双因素试验

考虑交互作用



$$\left.\begin{aligned}&X_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+\varepsilon_{ijk},\\&\varepsilon_{ijk}\sim N(0,\sigma^2),\text{各 }\varepsilon_{ijk}\text{ 独立},\\&i=1,2,\cdots,r;j=1,2,\cdots,s;k=1,2,\cdots,t,\\&\sum_{i=1}^{r}\alpha_i=0,\sum_{j=1}^{s}\beta_j=0,\sum_{i=1}^{r}\gamma_{ij}=0,\sum_{j=1}^{s}\gamma_{ij}=0,\end{aligned}\right\}$$

$$\begin{cases}H_{01}: \alpha_1=\alpha_2=\cdots=\alpha_r=0\\ H_{11}: \alpha_1,\alpha_2\ldots\alpha_r\text{ 不全为零}\end{cases}$$

$$\begin{cases}H_{02}: \beta_1=\beta_2=\cdots=\beta_s=0\\ H_{12}: \beta_1,\beta_2\ldots\beta_s\text{ 不全为零}\end{cases}$$

$$\begin{cases}H_{03}: \gamma_{11}=\gamma_{12}=\cdots=\gamma_{rs}=0\\ H_{13}: \gamma_{11},\gamma_{12}\ldots\gamma_{rs}\text{ 不全为零}\end{cases}$$

总偏差平方和 $S_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}^2-\frac{T_{\cdots}^2}{rst}$

A的效应平方和 $S_A=\frac{1}{st}\sum_{i=1}^{r}T_{i\cdot\cdot}^2-\frac{T_{\cdots}^2}{rst}$

B的效应平方和 $S_B=\frac{1}{rt}\sum_{j=1}^{s}T_{\cdot j\cdot}^2-\frac{T_{\cdots}^2}{rst}$

A.B交互效应平方和 $S_{A\times B}=\left(\frac{1}{t}\sum_{i=1}^{r}\sum_{j=1}^{s}T_{ij\cdot}^2-\frac{T_{\cdots}^2}{rst}\right)-S_A-S_B$

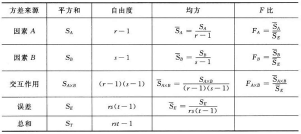
误差平方和 $S_E=S_T-S_A-S_B-S_{A\times B}$

拒绝域 $F_A\geq F_\alpha(r-1,rs(t-1))$

$F_B\geq F_\alpha(s-1,rs(t-1))$

$F_{A\times B}\geq F_\alpha((r-1)(s-1),rs(t-1))$

表 9—9　双因素试验的方差分析表

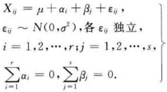


方差来源	平方和	自由度	均方	F 比
因素 A	S_A	$r-1$	$\bar{S}_A=\frac{S_A}{r-1}$	$F_A=\frac{\bar{S}_A}{\bar{S}_E}$
因素 B	S_B	$s-1$	$\bar{S}_B=\frac{S_B}{s-1}$	$F_B=\frac{\bar{S}_B}{\bar{S}_E}$
交互作用	$S_{A\times B}$	$(r-1)(s-1)$	$\bar{S}_{A\times B}=\frac{S_{A\times B}}{(r-1)(s-1)}$	$F_{A\times B}=\frac{\bar{S}_{A\times B}}{\bar{S}_E}$
误差	S_E	$rs(t-1)$	$\bar{S}_E=\frac{S_E}{rs(t-1)}$	
总和	S_T	$rst-1$		

已知无交互作用

$$\begin{cases}H_{01}: \alpha_1=\alpha_2=\cdots=\alpha_r=0\\ H_{11}: \alpha_1,\alpha_2,\ldots\alpha_r\text{ 不全为零}\end{cases}$$

$$\begin{cases}H_{02}: \beta_1=\beta_2=\cdots=\beta_s=0\\ H_{12}: \beta_1,\beta_2,\ldots\beta_s\text{ 不全为零}\end{cases}$$



$$\left.\begin{aligned}&X_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij},\\&\varepsilon_{ij}\sim N(0,\sigma^2),\text{各 }\varepsilon_{ij}\text{ 独立},\\&i=1,2,\cdots,r;j=1,2,\cdots,s,\\&\sum_{i=1}^{r}\alpha_i=0,\sum_{j=1}^{s}\beta_j=0.\end{aligned}\right\}$$

总偏差平方和 $S_T=\sum_{i=1}^{r}\sum_{j=1}^{s}X_{ij}^2-\frac{T_{\cdot\cdot}^2}{rs}$

A的效应平方和 $S_A=\frac{1}{s}\sum_{i=1}^{r}T_{i\cdot}^2-\frac{T_{\cdot\cdot}^2}{rs}$

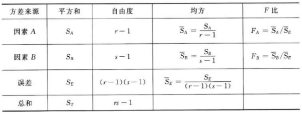
B的效应平方和 $S_B=\frac{1}{r}\sum_{j=1}^{s}T_{\cdot j}^2-\frac{T_{\cdot\cdot}^2}{rs}$

误差平方和 $S_E=S_T-S_A-S_B$

拒绝域 $F_A\geq F_\alpha(r-1,(r-1)(s-1))$

$F_B\geq F_\alpha(s-1,(r-1)(s-1))$

⟸



方差来源	平方和	自由度	均方	F 比
因素 A	S_A	$r-1$	$\bar{S}_A=\frac{S_A}{r-1}$	$F_A=\bar{S}_A/\bar{S}_E$
因素 B	S_B	$s-1$	$\bar{S}_B=\frac{S_B}{s-1}$	$F_B=\bar{S}_B/\bar{S}_E$
误差	S_E	$(r-1)(s-1)$	$\bar{S}_E=\frac{S_E}{(r-1)(s-1)}$	
总和	S_T	$rs-1$		

一元线性回归

a.b的估计

$$\begin{cases}S_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-\frac{1}{n}(\sum_{i=1}^{n}x_i)^2\\ S_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-\frac{1}{n}(\sum_{i=1}^{n}y_i)^2\\ S_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_iy_i-\frac{1}{n}(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}y_i)\end{cases}$$

$$\begin{cases}\hat{b}=\frac{S_{xy}}{S_{xx}}\\ \hat{a}=\frac{1}{n}\sum_{i=1}^{n}y_i-(\frac{1}{n}\sum_{i=1}^{n}x_i)\hat{b}\end{cases}$$

σ^2的估计

$\hat{\sigma}^2=\frac{Q_e}{n-2}=\frac{1}{n-2}(S_{yy}-\hat{b}S_{xy})$

显著性检验

$H_0: b=0$ 回归效果不显著

$H_1: b\neq0$ 回归效果显著

$t=\frac{\hat{b}}{\hat{\sigma}}\sqrt{S_{xx}}\sim t(n-2)$

多元线性回归

$Y=b_0+b_1x_1+\cdots+b_px_p+\varepsilon,\quad \varepsilon\sim N(0,\sigma^2)$

正规方程组的矩阵形式

$$X=\begin{bmatrix}1&x_{11}&x_{12}&\cdots&x_{1p}\\1&x_{21}&x_{22}&\cdots&x_{2p}\\\vdots&\vdots&\vdots&&\vdots\\1&x_{n1}&x_{n2}&\cdots&x_{np}\end{bmatrix},\quad Y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}\quad B=\begin{bmatrix}b_0\\b_1\\\vdots\\b_p\end{bmatrix}$$

$$\hat{B}=\begin{bmatrix}\hat{b}_0\\\hat{b}_1\\\vdots\\\hat{b}_p\end{bmatrix}=(X^TX)^{-1}X^TY$$

↓

$(b_0,b_1,\ldots b_p)^T$ 的最大似然估计

$\hat{b}_0+\hat{b}_1x_1+\cdots+\hat{b}_px_p\overset{\text{记为}}{=\!=}\hat{y}$

↓

$\mu(x_1,x_2,\ldots x_p)=b_0+b_1x_1+\cdots+b_px_p$ 的估计

$\hat{y}=\hat{b}_0+\hat{b}_1x_1+\cdots+\hat{b}_px_p$

↳ Y 关于 x 的 p 元回归方程

拒绝域 $|t|=\frac{|\hat{b}|}{\hat{\sigma}}\sqrt{S_{xx}} > t_{\frac{\alpha}{2}}(n-2)$

系数b的置信区间

回归效果显著时对b作区间估计

b的置信水平为 $1-\alpha$ 的置信区间

$$\left(\hat{b} \pm t_{\frac{\alpha}{2}}(n-2)\frac{\hat{\sigma}}{\sqrt{S_{xx}}}\right)$$

$\mu(x)=a+bx$ 的点估计和置信区间

点估计 $\hat{y}_0=\widehat{\mu(x_0)}=\hat{a}+\hat{b}x_0$

置信区间 $\frac{\hat{Y}_0-(a+bx_0)}{\sigma\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}} \sim t(n-2)$

$\mu(x_0)=a+bx_0$ 的置信水平为 $1-\alpha$ 的置信区间

$$\left(\hat{Y}_0 \pm t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$$

or $\left(\hat{a}+\hat{b}x_0 \pm t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$

多一个 ε
因此+1

Y的点预测和区间预测 $(Y_0=a+bx_0+\varepsilon_0)$

点预测 $\hat{Y}_0=\hat{a}+\hat{b}x_0$

区间预测 $\left(\hat{Y}_0 \pm t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$

or $\left(\hat{a}+\hat{b}x_0 \pm t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$

可化为一元线性回归的例子

1° $Y=\alpha e^{\beta x}\cdot\varepsilon$ $\ln\varepsilon \sim N(0,\sigma^2)$

$$\Rightarrow \underbrace{\ln Y}_{\downarrow\, Y'}=\underbrace{\ln\alpha}_{\downarrow\, a}+\underbrace{\beta x}_{\downarrow\, bx'}+\underbrace{\ln\varepsilon}_{\downarrow\, \varepsilon'}$$

$Y'=a+bx'+\varepsilon'$ $\varepsilon'\sim N(0,\sigma^2)$

2° $Y=\alpha x^{\beta}\cdot\varepsilon$ $\ln\varepsilon \sim N(0,\sigma^2)$

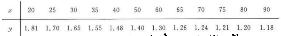
$$\Rightarrow \underbrace{\ln Y}_{\downarrow\, Y'}=\underbrace{\ln\alpha}_{\downarrow\, a}+\underbrace{\beta\ln x}_{\downarrow\, bx'}+\underbrace{\ln\varepsilon}_{\downarrow\, \varepsilon'}$$

$Y'=a+bx'+\varepsilon'$ $\varepsilon'\sim N(0,\sigma^2)$

3° $Y=\alpha+\beta h(x)+\varepsilon$ $\varepsilon\sim N(0,\sigma^2)$

$Y=a+bx'+\varepsilon$ $\varepsilon\sim N(0,\sigma^2)$

例：



x	20	25	30	35	40	50	60	65	70	75	80	90
y	1.81	1.70	1.65	1.55	1.48	1.40	1.30	1.26	1.24	1.21	1.20	1.18

选取模型 $Y=b_0+b_1x+b_2x^2+\varepsilon$，$\varepsilon\sim N(0,\sigma^2)$

$$\boldsymbol{X}=\begin{array}{ccc} x^0 & x^1 & x^2 \\ \hline 1 & 20 & 400 \\ 1 & 25 & 625 \\ 1 & 30 & 900 \\ 1 & 35 & 1\,225 \\ 1 & 40 & 1\,600 \\ 1 & 50 & 2\,500 \\ 1 & 60 & 3\,600 \\ 1 & 65 & 4\,225 \\ 1 & 70 & 4\,900 \\ 1 & 75 & 5\,625 \\ 1 & 80 & 6\,400 \\ 1 & 90 & 8\,100 \end{array},\ \boldsymbol{Y}=\begin{pmatrix} 1.81 \\ 1.70 \\ 1.65 \\ 1.55 \\ 1.48 \\ 1.40 \\ 1.30 \\ 1.26 \\ 1.24 \\ 1.21 \\ 1.20 \\ 1.18 \end{pmatrix},\ \boldsymbol{B}=\begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \Rightarrow \text{计算 } \hat{B}=(X^TX)^{-1}X^TY$$