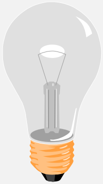
07

参数估计

参数估计

估计新生儿的平均体重

估计废品率



估计池塘中鱼数

参数估计

θ 是$F(x,\theta)$中的未知参数，$\theta \in \Theta$

样本 $(X_1, X_2, \cdots X_n)$

估计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$

估计区间 $\hat{\theta}_1(X_1, \ldots, X_n)$，$\hat{\theta}_2(X_1, \ldots, X_n)$

估计值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$

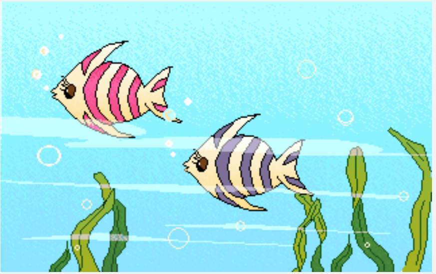
$\theta \in \left(\hat{\theta}_1,\ \hat{\theta}_2\right)$

点估计

区间估计

参数估计

如何估计池塘中的鱼数 N？

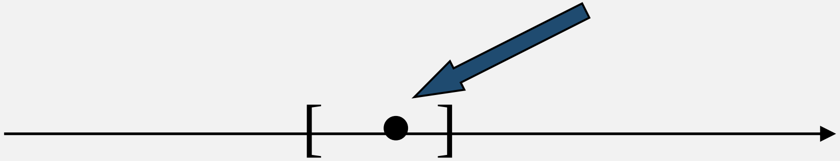


做法

- 第一次捕上 r 条，做上记号后放回.
- 再捕出 S 条鱼，发现其中有 k 条标有记号.

鱼数 N 的真值

$$\hat{N}=\left[\frac{Sr}{k}\right]$$



点估计

区间估计

07

参数估计

矩估计

➢ 理论依据：大数定律

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i^{\boldsymbol{k}} - E(X^{\boldsymbol{k}})\right| < \varepsilon\right) = 1$$

K.Pearson

➢ 基本原则：

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{\text{估计}} a_k = E(X^k) \qquad k = 1, 2, \ldots$$

可实现　　　　　　未知

矩估计

例 设总体 $X \sim f(x)=\begin{cases}(\alpha+1)x^{\alpha} & 0<x<1 \\ 0 & \text{其它}\end{cases}$，求参数 α 的矩估计.

解 由矩法有

$$\hat{E}X = \overline{X}$$

$$EX = \int_0^1 x\cdot(\alpha+1)x^{\alpha}dx = \frac{\alpha+1}{\alpha+2}$$

$$\hat{\alpha} = \frac{2\overline{X}-1}{1-\overline{X}}$$

矩估计

例　试求总体期望$\theta_1=EX$和方差 $\theta_2=DX$ 的矩估计.

解　$\hat{E}X=\bar{X}$,　$\hat{E}X^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$

$\hat{\theta}_1=\bar{X}$,　$DX=EX^2-(EX)^2$

$$\hat{\theta}_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=\tilde{S}^2$$

注1　$\hat{E}(X)=\bar{X}$,　$\hat{D}(X)=\tilde{S}^2$

注2　估计不唯一，如对总体$P(\lambda)$，有 $\hat{\lambda}=\bar{X}$，$\hat{\lambda}=\tilde{S}^2$

注3　基本原则可增加，用 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$ 估计 $\beta_k=E(X-EX)^k$

矩估计

例 设总体 $X \sim U[a, b]$, 试由样本$(X_1, X_2, \dots, X_n)$, 求未知参数 a, b 的矩估计.

解

$$\begin{cases} \hat{E}X = \bar{X} \\ \hat{D}X = \tilde{S}^2 \end{cases} \Rightarrow \begin{cases} \dfrac{a+b}{2} = \bar{X} \\ \dfrac{(b-a)^2}{12} = \tilde{S}^2 \end{cases} \Rightarrow \begin{cases} \hat{a} = \bar{X} - \sqrt{3\tilde{S}^2} \\ \hat{b} = \bar{X} + \sqrt{3\tilde{S}^2} \end{cases}$$

缺点 如 $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots, \frac{1}{11}$ 为来自 $X \sim U[a,b]$ 的样本观察值,

则a, b的矩估计值为 $\hat{a} = -0.01$, $\hat{b} = 0.414$.

矩估计

思考题 如何用矩法估计事件发生的概率 p？

解

X	0	1
P	$1-p$	p

$$\hat{E}X = \bar{X},$$

$$\hat{p} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{m}{n}$$

事件发生的频率估计概率.

07

参数估计

极大似然估计

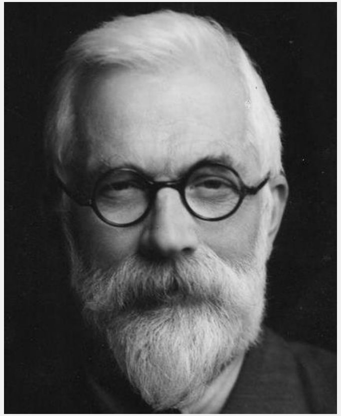
总体的分布类型已知，对未知参数做估计．

Lambert，1760年；

Gauss，1821年；

Fisher，1912年重新发现了这一方法，

并首先研究了极大似然估计法的一些性质．



极大似然估计

引例 设$X \sim B(1, p)$，p未知. 设想我们事先知道 p只有两种可能: $p=0.8$ 或 $p=0.1$

抽样结果: 1, 1, 0, 1. 问: 应如何估计参数 p?

解 $$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1) = p^3(1-p)$$

参数 p	观测值出现的概率
0.8	0.1024
0.1	0.0009

故参数 p 的极大似然估计值是 0.8.

极大似然估计

极大似然原则

已发生的事件，其概率应该最大.

极大似然估计 (Maximum Likelihood Estimation, MLE)

总体 X 的分布形式已知，参数 θ 未知

$$x_1, x_2, \ldots, x_n$$

参数 θ 的选择应有利于样本观测值的发生，即让这组数据发生的概率达到最大.

极大似然估计

步骤

- 似然函数

$$L(x_1, x_2, \cdots, x_n; \theta) = \begin{cases} \prod_{i=1}^{n} f(x_i; \theta), & f(x;\theta) \text{ 为} X \text{ 的概率密度} \\ \prod_{i=1}^{n} P(X = x_i; \theta), & P(X = x;\theta) \text{ 为} X \text{ 的分布律} \end{cases}$$

- θ的极大似然估计值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$

$$L(x_1, x_2, \cdots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n; \theta)$$

- 求解似然方程

$$\frac{\partial \ln L(\theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, k,$$

极大似然估计

例 设$X_1, X_2, \ldots, X_n$是取自总体 X 的一个样本，求θ的极大似然估计.

$$X \sim f(x) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{否} \end{cases} \qquad \text{其中}\theta > 0$$

解

$$L(\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1}$$

$$\ln L(\theta) = n \ln \theta + (\theta - 1)\sum_{i=1}^{n} \ln x_i \qquad (0 < x_i < 1)$$

$$\frac{d \ln L(\theta)}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i = 0 \qquad \hat{\theta} = -n \Big/ \sum_{i=1}^{n} \ln x_i$$

07

参数估计

极大似然估计

极大似然原则

已发生的事件，其概率应该最大.

极大似然估计 (Maximum Likelihood Estimation, MLE)

总体 X 的分布形式已知，未知的仅仅是参数

$$x_1, x_2, \ldots, x_n$$

参数 θ 的选择应有利于样本观测值的发生，即让这组数据发生的概率达到最大.

极大似然估计

例 设总体 $X \sim N(\mu, \sigma^2)$，求 μ 和 σ^2 的极大似然估计.

解

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\ln L=-\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\begin{cases}\dfrac{\partial}{\partial\mu}\ln L=\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)=\dfrac{1}{\sigma^2}(n\bar{x}-n\mu)=0\\[2ex]\dfrac{\partial}{\partial\sigma^2}\ln L=-\dfrac{n}{2}\cdot\dfrac{1}{\sigma^2}+\dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2=0\end{cases}$$

$$\begin{cases}\hat{\mu}=\bar{x}\\[2ex]\hat{\sigma}^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2=\tilde{S}^2\end{cases}$$

极大似然估计

例 设总体 $X \sim U[a, b]$, 试求 a 和 b 的极大似然估计.

解

$$L(a,b)=\prod_{i=1}^{n} f(x_i;a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a\le x_1,x_2,\cdots,x_n\le b\\ 0, & \text{否}\end{cases}$$

在 $a\le x_1,x_2,\cdots,x_n\le b$ 下,

$$a\uparrow, b\downarrow \Rightarrow (b-a)\downarrow \Rightarrow L(a,b)\uparrow$$

a, b的极大似然估计为 $\hat{a}=x_1^*$, $\hat{b}=x_n^*$

极大似然估计

不变性

若$\hat{\theta}$为θ的极大似然估计，$u=u(\theta)$有反函数$\theta=\theta(u)$，则$\hat{u}=u(\hat{\theta})$为$u=u(\theta)$的极大似然估计.

例 设总体$X\sim N(\mu,\sigma^2)$, 求 σ的极大似然估计.

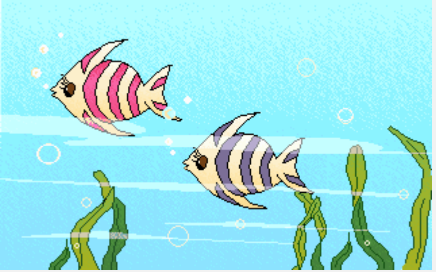
解 $\sigma>0,\sigma=\sqrt{\sigma^2}$ 有反函数，且 $\hat{\sigma}^2=\tilde{S}^2$

$$\hat{\sigma}=\tilde{S}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

极大似然估计

应用

极大似然法估计池塘中鱼的数量



- 第一次捕上r条，做上记号后放回
- 再捕上S条鱼，其中k条有记号.

分析 第二次捕出的有记号的鱼数X是随机变量，服从超几何分布：

$$L(N;k)=P\{X=k\}=\frac{C_r^k C_{N-r}^{S-k}}{C_N^S},\quad N\in\{\max(r,s),\max(r,s)+1,\cdots\}$$

$$\frac{P(X=k;N)}{P(X=k;N-1)}=\frac{(N-S)(N-r)}{N(N-r-S+k)}\begin{cases}>1, & N<Sr/k\\ <1, & N>Sr/k\end{cases}$$

$$\hat{N}=\left[\frac{Sr}{k}\right]$$

07

参数估计

极大似然估计

不变性

若$\hat{\theta}$为θ的极大似然估计，$u=u(\theta)$有反函数$\theta = \theta(u)$，则$\hat{u} = u(\hat{\theta})$为$u=u(\theta)$的极大似然估计.

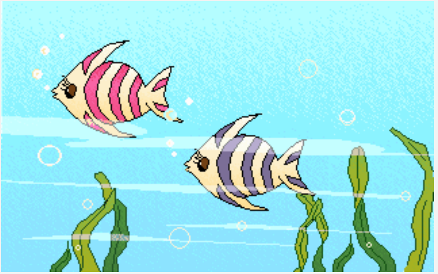
例 设总体$X \sim N(\mu, \sigma^2)$, 求σ的极大似然估计.

解 $\sigma > 0, \sigma = \sqrt{\sigma^2}$ 有反函数，且 $\hat{\sigma}^2 = \tilde{S}^2$

$$\hat{\sigma} = \tilde{S} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

极大似然估计

应用　极大似然法估计池塘中鱼的数量



- 第一次捕上r条，做上记号后放回
- 再捕上S条鱼，其中k条有记号.

分析　第二次捕出的有记号的鱼数X是随机变量，服从超几何分布：

$$L(N;k)=P\{X=k\}=\frac{C_r^k C_{N-r}^{S-k}}{C_N^S},\quad N\in\{\max(r,s),\max(r,s)+1,\cdots\}$$

$$\frac{P(X=k;N)}{P(X=k;N-1)}=\frac{(N-S)(N-r)}{N(N-r-S+k)}\begin{cases}>1, & N<Sr/k\\<1, & N>Sr/k\end{cases}$$

$$\hat{N}=\left[\frac{Sr}{k}\right]$$

07

参数估计

评选原则

无偏性　若 $E(\hat{\theta})=\theta$，则称 $\hat{\theta}$ 为 θ 的无偏估计量.

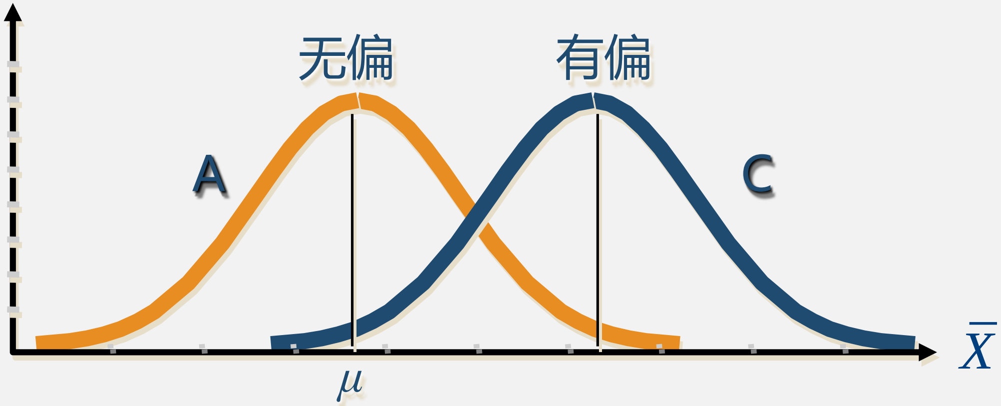
例　证明样本均值 $\bar{X}$ 为总体期望 $\mu=E(X)$ 的无偏估计.

证

$$E(\bar{X})=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=E(X)=\mu$$

一般

$$E(A_k)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^k)=E(X^k)=\alpha_k$$



评选原则

例 证明样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 是总体方差 $\mathrm{D}(X)=\sigma^2$ 的无偏估计.

证

$$E(S^2)=E[\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\bar{X}^2)]$$

$$=\frac{1}{n-1}[nE(X^2)-nE(\bar{X}^2)]$$

$$\xrightarrow{D(X)=E(X^2)-(EX)^2} \quad =\frac{1}{n-1}[n(\sigma^2+\mu^2)-n(D(\bar{X})+\mu^2)]$$

$$=\frac{1}{n-1}(n\sigma^2+n\mu^2-\sigma^2-n\mu^2)$$

$$=\sigma^2$$

评选原则

注1　$\tilde{S}^2$ 不是 σ^2 的无偏估计.

$$E(\tilde{S}^2)=E(\frac{n-1}{n}S^2)=\frac{n-1}{n}\sigma^2 \xrightarrow[n\to\infty]{} \sigma^2$$ 渐近无偏估计

注2　$\hat{\theta}$ 为 θ 的无偏估计，但 $u(\hat{\theta})$不一定是 $u(\theta)$的无偏估计.

例 若 $D(X)>0$，则 $E[(\hat{\mu})^2]=E(\bar{X}^2)=D(\bar{X})+[E(\bar{X})]^2>[E(\bar{X})]^2=\mu^2$

注3　无偏估计不唯一，如X_1和 $\bar{X}$ 均为 $\mu=E(X)$ 的无偏估计.

事实上对任何 $c_1, c_2, \ldots, c_n$，当 $c_1+c_2+\cdots+c_n=1$ 时，

$$E(\sum_{i=1}^{n}c_iX_i)=\mu$$

评选原则

应用 对某种产品从1开始用整数连续编号，试估计产品的总量.

分析 设产品编号为X, 则

X	1	2	$\cdots$	N
p	$\frac{1}{N}$	$\frac{1}{N}$	$\cdots$	$\frac{1}{N}$

样本：20, 96, 190, 255

无偏估计：317.75

$$L(N)=\frac{1}{N^n},\quad 1\le x_1^*\le x_2^*\le\cdots\le x_n^*\le N,\qquad \hat{N}=x_n^*$$

$$P(X_n^*=k)=F_{X_n^*}(k)-F_{X_n^*}(k-1)=\frac{k^n-(k-1)^n}{N^n}$$

$$E(X_n^*)=\frac{n}{n+1}(N+1),\qquad E(\frac{n+1}{n}X_n^*-1)=N$$

07

参数估计

评选原则

有效性

定义　设 $\hat{\theta}_1,\ \hat{\theta}_2$ 都是参数θ的无偏估计量，若

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

则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效.

设 $\hat{\theta}_0$是参数θ的无偏估计量，若对θ的任一无偏估计$\hat{\theta}$有

$$D(\hat{\theta}_0) \le D(\hat{\theta})$$

则称 $\hat{\theta}_0$ 是θ的最小方差无偏估计.

评选原则

有效性

例 证明 $\overline{U} = \{ \hat{\mu} = \sum_{i=1}^{n} k_i X_i \mid \sum_{i=1}^{n} k_i = 1\}$ 是μ=$E(X)$的线性无偏估计类.

并在此无偏估计类里寻找μ的最小方差无偏估计.

证

$$E(\hat{\mu}) = \sum_{i=1}^{n} k_i E(X_i) = \sum_{i=1}^{n} k_i \mu = \mu, \qquad \mu \in \overline{U}$$

评选原则

例 证明 $\bar{U}=\{\hat{\mu}=\sum_{i=1}^{n}k_iX_i \mid \sum_{i=1}^{n}k_i=1\}$ 是μ=$E(X)$的线性无偏估计类.

在此无偏估计类里寻找μ的最小方差无偏估计.

分析 $(\sum_{i=1}^{n}a_ib_i)^2 \le (\sum_{i=1}^{n}a_i^2)(\sum_{i=1}^{n}b_i^2)$

$$D(\hat{\mu})=D(X)\sum_{i=1}^{n}k_i^2=\frac{1}{n}(\sum_{i=1}^{n}1^2)(\sum_{i=1}^{n}k_i^2)D(X)\ge\frac{1}{n}\left(\sum_{i=1}^{n}1\times k_i\right)^2D(X)$$

$$=\frac{1}{n}D(X)$$

而 $D(\bar{X})=\dfrac{1}{n}D(X)$　故 $\hat{\mu}_0=\bar{X}$ 是μ的最小方差无偏估计.

07

参数估计

评选原则

一致性

定义 设 $\hat{\theta}$ 是参数θ的估计量，若对任意$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| < \varepsilon\right) = 1 \quad \text{即} \ \hat{\theta} \xrightarrow{P} \theta$$

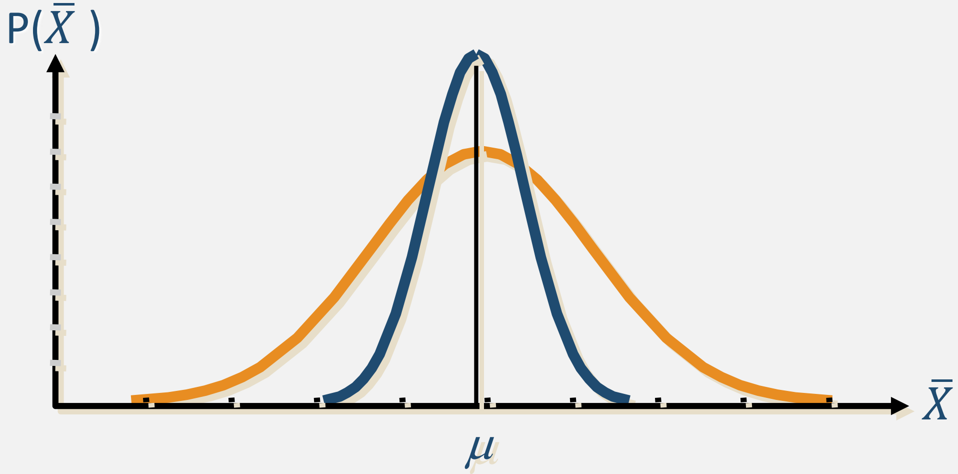
则称 $\hat{\theta}$为 θ 的 一致估计量.

评选原则

一致性

例 样本均值$\bar{X}$是总体均值 μ 的一致估计.

证 大数定律 $$\lim_{n\to\infty} P(|\bar{X}-\mu|<\varepsilon)=1$$



评选原则

一致性

例 证明正态总体的样本方差S^2是σ^2的一致估计.

证 $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$, $\quad E(S^2)=\sigma^2$

$$D(S^2) = D\left(\frac{\sigma^2}{n-1}\cdot\frac{n-1}{\sigma^2}S^2\right) = \frac{\sigma^4}{(n-1)^2}D\left(\frac{n-1}{\sigma^2}S^2\right) = \frac{2\sigma^4}{(n-1)}$$

由切比雪夫不等式

$$P(|S^2-\sigma^2|<\varepsilon) \geq 1-\frac{2\sigma^4}{(n-1)\varepsilon^2} \xrightarrow[n\to\infty]{} 1$$

07

参数估计

区间估计

问题 如何让$\hat{\theta}$与θ的误差体现在估计中?

办法 对给定的置信水平(置信度)1 - α

置信下限$\hat{\theta}_1(X_1, X_2, \ldots, X_n)$和置信上限$\hat{\theta}_2(X_1, X_2, \ldots, X_n)$

使

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) \geq 1 - \alpha$$

称$(\hat{\theta}_1, \hat{\theta}_2)$为未知参数$\theta$的置信度为1 - α的置信区间.

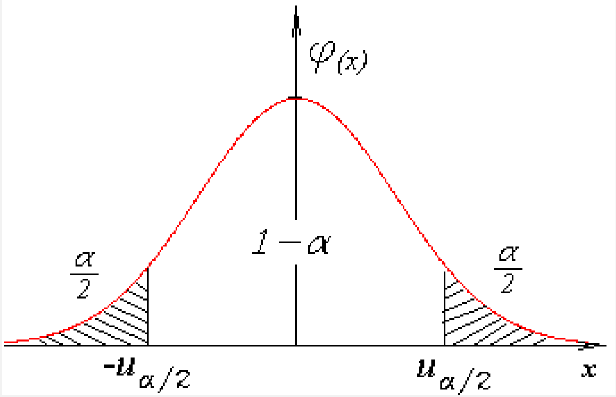
含义 若1 - α =0.95，抽样100次中约有95个$(\hat{\theta}_1, \hat{\theta}_2)$包含$\theta$.

区间估计

单个正态总体均值的区间估计

σ^2已知

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



$$P(\bar{X}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}})=1-\alpha$$

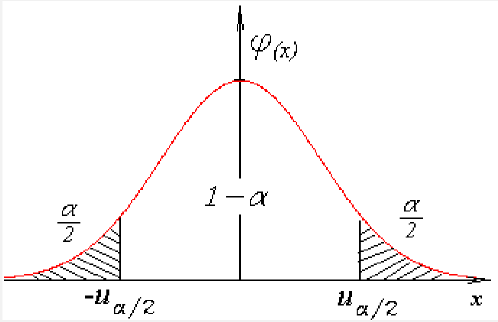
μ的置信度为1 - α的置信区间为$(\bar{X}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}})$; $(\bar{X}\pm u_{\alpha/2}\frac{\sigma}{\sqrt{n}})$

区间估计

单个正态总体均值的区间估计

注

$$\left(\bar{X} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$



- 置信区间长度 $l = \dfrac{2\sigma}{\sqrt{n}} u_{\alpha/2}$
- 相同置信水平下，置信区间选取不唯一.

区间估计

单个正态总体均值的区间估计

例 滚珠直径$X \sim N(\mu, 0.0006)$，从某天生产的滚珠中随机抽取6个，测得直径为

(单位：mm) 1.46 1.51 1.49 1.48 1.52 1.51

求 μ 的置信度为95%的置信区间.

解 $$\left(\bar{X} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{x} = 1.495, \qquad \alpha = 0.05, \qquad u_{0.05/2} = u_{0.025} = 1.96$$

$$\left(1.495 \pm 1.96 \times \sqrt{\frac{0.0006}{6}}\right) = (1.4754,\ 1.5146)$$

07

参数估计

区间估计

单个正态总体均值的区间估计

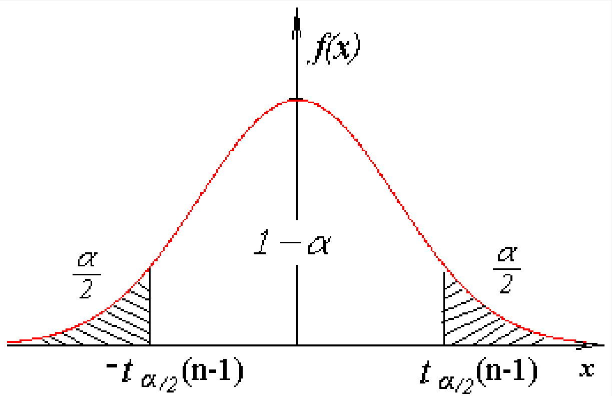
σ^2未知

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

$$P(|T|<t_{\alpha/2}(n-1))=1-\alpha$$

$$P(\bar{X}-t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}<\mu<\bar{X}+t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}})=1-\alpha$$

μ的置信度为1-α的置信区间为 $(\bar{X}\pm t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}})$



参数估计

单个正态总体均值的区间估计

例 滚珠直径$X\sim N(\mu,\sigma^2)$，从某天生产的滚珠中随机抽取6个，测得直径为(单位：mm) 1.46 1.51 1.49 1.48 1.52 1.51

求μ的置信度为95%的置信区间.

解 $S=0.02258$, $t_{0.025}(5)=2.5706$,

$$\left(\bar{X}\pm t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right)=(1.4716\,,\ 1.5187)\qquad l=0.0474$$

区间估计

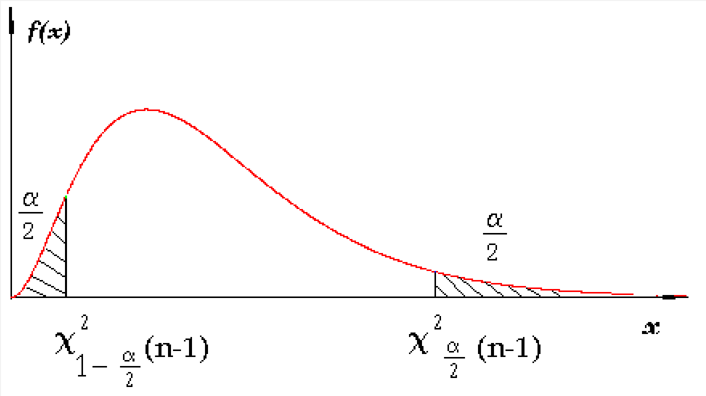
单个正态总体方差的区间估计

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

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

$$P(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}) = 1-\alpha$$

σ^2 的置信度为1-α 的置信区间为 $\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$



区间估计

单个正态总体方差的区间估计

例 从自动车床加工的一批零件中随机的抽取16件，测得各零件长度为(单位：cm)

2.15　2.10　2.12　2.10　2.14　2.11　2.15　2.13

2.13　2.11　2.14　2.13　2.12　2.13　2.10　2.14

设零件长度服从正态分布，求零件长度标准差σ的置信度为95%的置信区间.

解 计算 $\overline{x}=2.125$, $S^2=0.000293$, $\alpha=0.05$, $\chi^2_{0.025}(15)=27.488$, $\chi^2_{0.975}(15)=6.262$

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}},\ \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\right)$$

$=(0.01265，0.02651)$

参数估计

单侧置信区间

设总体 X 的分布函数为$F(x;\theta)$，样本 $X_1, X_2, \dots, X_n$

给定置信度1-α

- θ 的单侧置信下限　　$P(\theta > \hat{\theta}_1) \geq 1-\alpha$
- θ 的单侧置信上限　　$P(\theta < \hat{\theta}_2) \geq 1-\alpha$

思考　单个正态总体的均值和方差的单侧置信区间

07

参数估计

区间估计

双正态总体均值差的区间估计

$(X_1, X_2, \ldots, Xn)$和$(Y_1, Y_2, \cdots, Y_n)$为分别取自两个相互独立的正态总体

$N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$的两个样本，其样本方差分别记为S_1^2和S_2^2，则

- σ_1^2，σ_2^2均已知，$\mu_1 - \mu_2$的区间估计

$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

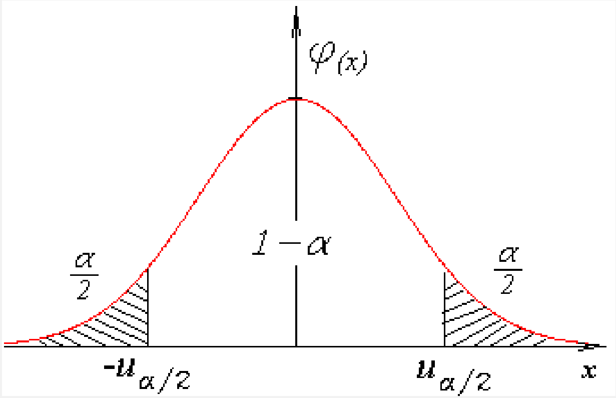
$$U = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

区间估计

双正态总体均值差的区间估计

σ_1^2，σ_2^2均已知，$\mu_1-\mu_2$的置信度为1 - α的置信区间为

$$\left(\overline{X}-\overline{Y}-\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}u_{\alpha/2},\quad \overline{X}-\overline{Y}+\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}u_{\alpha/2}\right)$$



参数估计

双正态总体均值差的区间估计

例 为比较两位银行职员为新顾客办理个人结算账目的平均时间长度，分别给两位职员随机安排了10位顾客，并记录下为每位顾客办理账单所需的时间

$$\overline{x}_1=22.2,\ \overline{x}_2=28.5 \qquad \text{(单位：分钟)}$$

假定两位职员办理账单所需时间分别服从方差为$\sigma_1^2=16$和$\sigma_2^2=19$的正态分布，试求两位职员办理账单的服务时间之差的95%的区间估计.

参数估计

双正态总体均值差的区间估计

解

$$X_1 \sim N(\mu_1, 16)$$

$$X_2 \sim N(\mu_2, 19)$$

$$\bar{x}_1 = 22.2$$

$$\bar{x}_2 = 28.5$$

$$n_1 = n_2 = 10$$

$$u_{\alpha/2}=1.96$$

$\mu_1 - \mu_2$的置信度为 95%的置信区间为

$$(22.2 - 28.5) \pm (1.96)\sqrt{\frac{16}{10} + \frac{19}{10}}$$

$$= (-10.0, -2.6)$$

07

参数估计

区间估计

双正态总体均值差的区间估计

$(X_1, X_2, \ldots, Xn)$和$(Y_1, Y_2, \cdots, Y_n)$为分别取自两个相互独立的正态总体$N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$的两个样本，其样本方差分别记为S_1^2和S_2^2，则

- 总体方差均已知，$\mu_1 - \mu_2$的置信区间

$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

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

区间估计

双正态总体均值差的区间估计

$(X_1, X_2, \ldots, Xn)$和$(Y_1, Y_2, \cdots, Y_n)$为分别取自两个相互独立的正态总体

$N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$的两个样本，其样本方差分别记为S_1^2和S_2^2，则

➢ 总体方差未知但相等，$\sigma_1^2 = \sigma_2^2 = \sigma^2$，$\mu_1 - \mu_2$的区间估计

$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

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

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

区间估计

双正态总体均值差的区间估计

$(X_1, X_2, \ldots, Xn)$和$(Y_1, Y_2, \cdots, Y_n)$为分别取自两个相互独立的正态总体

$N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$的两个样本，其样本方差分别记为S_1^2和S_2^2，则

- 总体方差未知但相等，$\sigma_1^2 = \sigma_2^2 = \sigma^2$，$\mu_1 - \mu_2$的区间估计

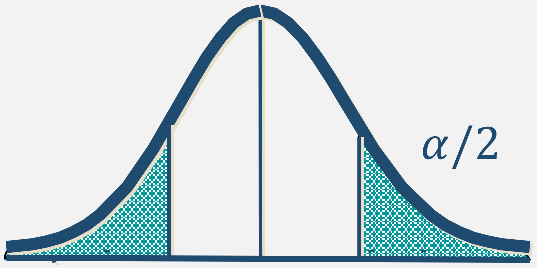
$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

区间估计

双正态总体均值差的区间估计

$$P\left(\left|\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right|<\mathrm{t}_{\alpha/2}(n_1+n_2-2)\right)=1-\alpha$$



$\mu_1-\mu_2$ 的置信度为1 - α 的置信区间为

$$\left(\overline{X}-\overline{Y}-S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\mathrm{t}_{\alpha/2}(n_1+n_2-2),\ \overline{X}-\overline{Y}+S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\mathrm{t}_{\frac{\alpha}{2}}(n_1+n_2-2)\right)$$

参数估计

双正态总体均值差的区间估计

例 为比较两位银行职员为新顾客办理个人结算账目的平均时间长度，分别给两位职员随机安排了10位顾客，并记录下为每位顾客办理账单所需的时间

$$\overline{x}_1 = 22.2,\ \overline{x}_2 = 28.5 \quad \text{（单位：分钟）}$$

$$S_1^2 = 16.36,\ S_2^2 = 18.92$$

假定两位职员办理账单所需时间均服从正态分布，且方差相等. 试求两位职员办理账单的服务时间之差的95%的置信区间.

参数估计

双正态总体均值差的区间估计

解

$$X_1 \sim N(\mu_1, \sigma^2)$$

$$X_2 \sim N(\mu_2, \sigma^2)$$

$$\bar{x}_1 = 22.2$$

$$\bar{x}_2 = 28.5$$

$$S_1^2 = 16.36$$

$$S_2^2 = 18.92$$

$$n_1 = n_2 = 10$$

$$\mathrm{t}_{\alpha/2}(18)=2.1$$

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

$$= 4.2$$

$\mu_1 - \mu_2$置信度为 95%的置信区间为

$$(22.2 - 28.5) \pm (2.1)(4.2)\sqrt{\frac{1}{10} + \frac{1}{10}}$$

$$= (-10.2, -2.4)$$

07

参数估计

区间估计

双正态总体均值差的区间估计

$(X_1, X_2, \ldots, Xn)$和$(Y_1, Y_2, \cdots, Y_n)$为分别取自两个相互独立的正态总体$N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$的两个样本，其样本方差分别记为S_1^2和S_2^2，则

- 总体方差均已知，$\mu_1 - \mu_2$的置信区间

$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

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

区间估计

双正态总体均值差的区间估计

$(X_1, X_2, \ldots, Xn)$和$(Y_1, Y_2, \cdots, Y_n)$为分别取自两个相互独立的正态总体 $N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$的两个样本，其样本方差分别记为S_1^2和S_2^2，则

➢ 总体方差未知且$\sigma_1^2 \neq \sigma_2^2$，$\mu_1 - \mu_2$的置信区间

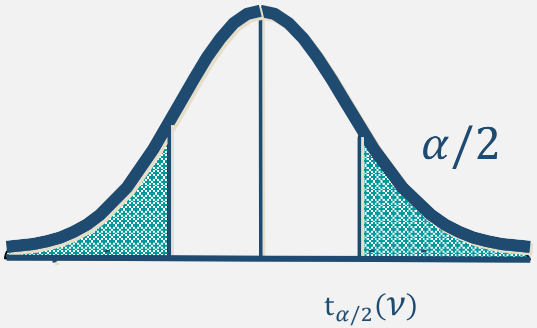
$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \sim t(\nu)$$

$$\nu = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}}$$

区间估计

双正态总体均值差的区间估计

$$P\left(\left|\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\right| < t_{\alpha/2}(\nu)\right)=1-\alpha$$



$\mu_1-\mu_2$ 的置信度为1 - α 的置信区间为

$$\left(\overline{X}-\overline{Y}-\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\,t_{\frac{\alpha}{2}}(\nu),\quad \overline{X}-\overline{Y}+\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\,t_{\frac{\alpha}{2}}(\nu)\right)$$

参数估计

双正态总体均值差的区间估计

例 为比较两位银行职员为新顾客办理个人结算账目的平均时间长度，分别给两位职员随机安排了10位顾客，并记录下为每位顾客办理账单所需的时间

$$\overline{x}_1 = 22.2, \ \overline{x}_2 = 28.5 \qquad (\text{单位：分钟})$$

$$S_1^2 = 16.36, \ S_2^2 = 18.92$$

假定两位职员办理账单所需时间均服从正态分布，且方差不等. 试求两位职员办理账单的服务时间之差的95%的区间估计.

双正态总体均值差的区间估计

解 $X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

$\sigma_1^2 \neq \sigma_2^2$

$\bar{x}_1 = 22.2$

$\bar{x}_2 = 28.5$

$S_1^2 = 16.36$

$S_2^2 = 18.92$

$n_1 = n_2 = 10$

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} \approx 18$$

$t_{\frac{\alpha}{2}}(\nu)$=$t_{0.025}(18)$=2.1

$\mu_1 - \mu_2$置信度为 95%的置信区间为

$$(22.2 - 28.5) \pm (2.1)\sqrt{\frac{16.36}{10} + \frac{18.92}{10}}$$

$$= (-10.25, -2.35)$$

07

参数估计

区间估计

双正态总体方差比的区间估计

$(X_1, X_2, \dots, Xn)$和$(Y_1, Y_2, \cdots, Y_n)$为分别取自两个相互独立的正态总体$N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$的两个样本，其样本方差分别记为S_1^2和S_2^2，则

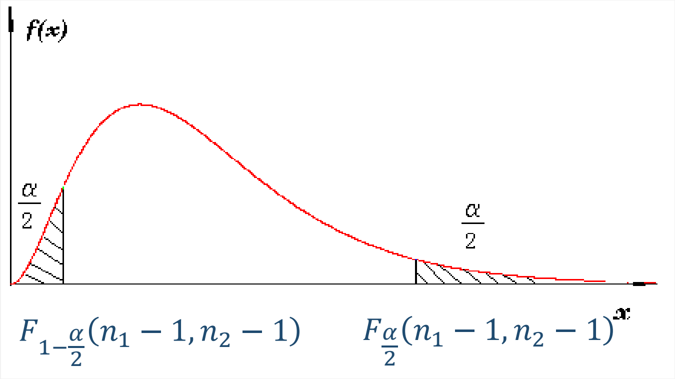
- 方差比σ_1^2/σ_2^2的区间估计

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

区间估计

双正态总体方差比的区间估计

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



$$P\left(F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1) < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{\frac{\alpha}{2}}(n_1-1, n_2-1)\right) = 1-\alpha$$

方差比σ_1^2/σ_2^2的置信度为1 - α的置信区间为

$$\left(\frac{S_1^2/S_2^2}{F_{\frac{\alpha}{2}}(n_1-1, n_2-1)}, \frac{S_1^2/S_2^2}{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)}\right)$$

参数估计

双正态总体方差比的区间估计

$$\left(\frac{S_1^2/S_2^2}{F_{\frac{\alpha}{2}}(n_1-1,n_2-1)},\quad \frac{S_1^2/S_2^2}{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)}\right)$$

例 从正态总体$N(\mu_1,\sigma_1^2)$和正态总体$N(\mu_2,\sigma_2^2)$中独立地抽取容量为10的样本，其样本方差分别为0.541, 0.6065，求方差比σ_1^2/σ_2^2的置信度为90%的置信区间.

解 $\alpha=0.1,\quad F_{\frac{\alpha}{2}}(n_1-1,n_2-1)=F_{0.05}(9,9)=3.18$

$$F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)=F_{0.95}(9,9)=\frac{1}{F_{0.05(9,9)}}=0.31$$

方差比σ_1^2/σ_2^2的置信度为90%的置信区间为 (0.28，2.84).