优良估计量三个常用准则

无偏性(没有系统误差,消除随机误差, 即$E(\hat{\theta})=\theta$)和

有效性(无偏性下,控制随机误差大小,即$D(\hat{\theta}_1)$)都是固定样本容量n:

$\overline{X}$和S^2分别是μ(总体均值)和σ^2(总体方差)的最小方差无偏估计。

μ已知,$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$是$\sigma^2$的无偏估计; $M_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$不是$\sigma^2$的无偏估计;

相合性是样本容量在变化, 较大时(无无偏性的要求):

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

满足无偏性情况下,证明可借助*Chebyshev*不等式

$\overline{X}$是μ的相合估计; S^2和M_2都是σ^2的相合估计

$\boxed{\overline{X}}$和S^2分别是μ和σ^2的最小方差无偏和相合估计.

相合性:不满足无偏性情况下,无法用Chebyshev不等式不等式

$$\lim_{n\to\infty} P\{|\hat{\theta}-\theta| \ge \varepsilon\} = 0 \quad 即: \hat{\theta}-\theta \xrightarrow{P} 0$$

设φ是$(0,\infty)$上的严格正值增函数, $\varphi(u)=\varphi(-u)$,

$E[\varphi(X)]<\infty$, 则对每个$u>0$: $P\{|X| \ge u\} \le \dfrac{E[\varphi(X)]}{\varphi(u)}$.

比如教材: $P\{|M_2-\sigma^2| \ge \varepsilon\} \le \dfrac{E[(M_2-\sigma^2)^2]}{\varepsilon^2} = \dfrac{2n-1}{n^2\varepsilon^2}\sigma^4 \xrightarrow{n\to\infty} 0$

则: $M_2-\sigma^2 \xrightarrow{P} 0$

统计推断三个方面:

抽样分布(精确分布);

参数估计;(已知分布类型)

假设检验。

参数估计
- 点估计
  - 矩估计
  - 极大似然估计
- 区间估计

区间估计 （双侧）

$$P\left\{\hat{\theta}_1(X_1,\ldots,X_n)\le\theta\le\hat{\theta}_2(X_1,\ldots,X_n)\right\}=1-\alpha$$

则称随机区间 $\left[\hat{\theta}_1,\hat{\theta}_2\right]$ 为θ的置信度为1-α的置信区间.

1－α又称置信系数或置信概率或置信水平
α又称显著性水平，通常取值为0.01，0.05.

注：反复抽多次，每个样本值确定一个区间，这些区间中要么包含待估参数的真值，要么不包含待估参数的真值，按伯努利大数定律，其中包含待估参数的真值占(1-α)%，不包含待估参数真值仅占α%.

难点和重点：由样本$X_1,\ldots,X_n$确定两个统计量：$\hat{\theta}_2=\hat{\theta}_2(X_1,\ldots,X_n)$

$\hat{\theta}_1=\hat{\theta}_1(X_1,\ldots,X_n)$

Neyman原则： 在置信度达到一定要求前提下,寻求精确度尽可能高的区间估计,平均长度尽可能短.

区间估计的优点与缺点(与点估计比较)

置信区间的枢轴变量法(寻找置信区间的步骤):

1) 选取待估参数θ 的估计量; 原则:优良性准则; 常用: $\bar{X} \to \mu$, $S^2 \to \sigma^2$

2) 建立枢轴变量

构造关于待估参数θ和样本的函数$W(X_1, X_2, \ldots, X_n, \theta)$,
其中W不含任何其他未知参数, W 称为枢轴变量(不是统计量).

枢轴变量W的要求: 1.为待估参数θ的连续、严格单调函数;
2. 枢轴变量W通常具有经典分布
(主要有: 正态、χ^2、T、F分布), 分布与θ 无关;
3. 尽量包含已经信息, 如均值或者方差信息.

例如:

$$U = \frac{\bar{X} - \mu}{\sigma_0 / \sqrt{n}} \sim N(0,1)$$

3) 根据W的分布,查上侧分位数$w_{1-\alpha/2}$和$w_{\alpha/2}$(Neyman原则), 则

$$P\{w_{1-\alpha/2} \le W \le w_{\alpha/2}\} = 1-\alpha = P\{A(X_1, \ldots, X_n) \le \theta \le B(X_1, \ldots, X_n)\}$$

4) 改写不等式, 其中A、B是不含未知参数的统计量.

上面工作过程的关键是构造枢轴变量W, 并以它为轴心, 由 $a \le W \le b$ 旋转出所需不等式$A \le \theta \le B$.

1. 随机变量X的$E(X),D(X)$存在, 且$D(X)>0$,则$X^*=\dfrac{X-E(X)}{\sqrt{D(X)}}$

$E(X^*)=0,\ D(X^*)=1$

样本: $(X_1, X_2, \cdots, X_n)$

(1) X_i与总体X相同分布；(2) $X_1, X_2, \cdots, X_n$相互独立.

如果矩存在的话: $E(X_i^k)=E(X^k)$,

$E[X_i-E(X_i)]^k=E[X-E(X)]^k$

$E(X_i)=E(X),\ D(X_i)=D(X)$

$$E\left(\overline{X}_n\right)=E\left(X\right) \qquad D\left(\overline{X}_n\right)=\frac{1}{n}D\left(X\right)$$

$$\Leftarrow \overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$$

$\overline{X}$ 和S^2分别是μ(总体均值)和σ^2(总体方差)的最小方差无偏估计。

μ已知, $\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$是$\sigma^2$的无偏估计；

$M_2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$不是$\sigma^2$的无偏估计；

统计学中最常用的公式

(1) $\sum\limits_{i=1}^{n}(X_i-\overline{X})=0$;

(2) $\sum\limits_{i=1}^{n}(X_i-\overline{X})^2=\sum\limits_{i=1}^{n}X_i^2-n\overline{X}^2$ (计算多用右式)

复习: 抽样分布定理(正态总体的统计量分布)

定理6.2.4 设 $X_1, X_2, ..., X_n$ 是正态总体 $X \sim N(\mu, \sigma^2)$ 的样本,

$\overline{X}, S^2$ 分别是样本均值和样本方差，则

(1) $\overline{X}$ 与 S^2 相互独立; $\Leftrightarrow \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$; (2) $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$;

(3) $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$; $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$; (4) $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

定理6.2.5: 设正态总体 X 与 Y 相互独立,

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

(2)当 $\sigma_1^2 = \sigma_2^2$ 时, $T = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$

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

思考: $\sigma_1^2 \neq \sigma_2^2$ 时的情况？提示: $n_1 = n_2$ (成对抽取)

区间估计(双侧)

$X \sim N(\mu, \sigma^2)$ $P\{w_{1-\alpha/2} \le W \le w_{\alpha/2}\} = 1-\alpha$

原则:选取最简单的优良估计量

原则:无偏、有效、相合

被估参数	条件	枢轴变量	优良估计量
μ	已知 σ^2	$U = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\bar{X}$
μ	未知 σ^2	$T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$\bar{X}$
σ^2	已知 μ	$\dfrac{n}{\sigma^2}\cdot\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$	$\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$
σ^2	未知 μ	$\dfrac{(n-1)}{\sigma^2}\cdot S^2 = \sum_{i=1}^{n}\left(\dfrac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$	$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$

一、单个正态总体：$X \sim N(\mu,\sigma^2)$

1. μ的估计 1) 已知$\sigma=\sigma_0$

$$U=\frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}}\sim N(0,1)$$

$$P\{-u_{\alpha/2}\le\frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}}\le u_{\alpha/2}\}=1-\alpha$$

$$\Rightarrow[\overline{X}-\frac{\sigma_0}{\sqrt{n}}u_{\alpha/2},\overline{X}+\frac{\sigma_0}{\sqrt{n}}u_{\alpha/2}]$$

2) σ^2未知:

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

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

$$[\overline{X}-t_{\alpha/2}(n-1)\bullet\frac{S}{\sqrt{n}},\overline{X}+t_{\alpha/2}(n-1)\bullet\frac{S}{\sqrt{n}}]$$

一、单个正态总体：$X\sim N(\mu,\sigma^2)$

2. σ^2 的估计 1) μ 已知:

$$\chi^2=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\frac{n}{\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n)$$

$$P\{\underline{\chi^2_{1-\alpha/2}(n)}\le\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\le\underline{\chi^2_{\alpha/2}(n)}\}=1-\alpha$$

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)},\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}\right]$$

2) μ未知:

$$\chi^2=\frac{n-1}{\sigma^2}S^2\Leftrightarrow\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\sim\chi^2(n-1);$$

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

$$\left[(n-1)S^2\big/\chi^2_{\alpha/2}(n-1),(n-1)S^2\big/\chi^2_{1-\alpha/2}(n-1)\right]$$

区间估计(双侧)

$X \sim N(\mu, \sigma^2)$ $P\{w_{1-\alpha/2} \le W \le w_{\alpha/2}\} = 1-\alpha$

被估参数	条件	枢轴变量	置信区间
μ	已知σ^2	$U=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$[\overline{X}-\dfrac{\sigma_0}{\sqrt{n}}u_{\alpha/2}, \overline{X}+\dfrac{\sigma_0}{\sqrt{n}}u_{\alpha/2}]$
μ	未知σ^2	$T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$[\overline{X}-t_{\alpha/2}(n-1)\bullet\dfrac{S}{\sqrt{n}}, \overline{X}+t_{\alpha/2}(n-1)\bullet\dfrac{S}{\sqrt{n}}]$
σ^2	已知μ	$\dfrac{n}{\sigma^2}\cdot\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2=\sum\limits_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$	$\left[\dfrac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)}, \dfrac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}\right]$
σ^2	未知μ	$\dfrac{(n-1)S^2}{\sigma^2}=\sum\limits_{i=1}^{n}\left(\dfrac{X_i-\overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$	$\left[(n-1)S^2/\chi^2_{\alpha/2}(n-1), (n-1)S^2/\chi^2_{1-\alpha/2}(n-1)\right]$

零件长度的方差 例: 从自动机床加工的同类零件中任取16件测得长度值为(单位: mm)

12.15	12.12	12.01	12.28	12.09	12.16	12.03	12.01
12.06	12.13	12.07	12.11	12.08	12.01	12.03	12.06

求方差的估计值和置信区间(α=0.05).

解: 设零件长度为X, 可认为X 服从正态分布.

$$\bar{x}=\frac{1}{16}\sum_{i=1}^{16}x_i=12.08,\qquad \sum_{i=1}^{16}(x_i-\bar{x})^2=0.0761$$

$$s^2=\frac{1}{15}\sum_{i=1}^{16}(x_i-\bar{x})^2=0.005$$

$$s^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)=\frac{1}{15}\left(\sum_{i=1}^{16}x_i^2-16\bar{x}^2\right)$$

求方差的置信区间:

由于μ未知, S^2 是σ^2 的优良估计, 选取枢轴变量:

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

相应的置信区间为:

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

$$\left[\,(n-1)S^2\big/\chi^2_{\alpha/2}(n-1),(n-1)S^2\big/\chi^2_{1-\alpha/2}(n-1)\right]$$

$$\left[(n-1)S^2 \big/ \chi^2_{\alpha/2}(n-1),(n-1)S^2 \big/ \chi^2_{1-\alpha/2}(n-1) \right]$$

查χ^2分布表可得：

$$\chi^2_{\alpha/2}(n-1) = \chi^2_{0.025}(15) = 27.488$$

$$\chi^2_{1-\alpha/2}(n-1) = \chi^2_{0.975}(15) = 6.262$$

σ^2 的一个区间估计为：$[\frac{0.0761}{27.488}, \frac{0.0761}{6.26}]$ 即 [0.002768,0.012]

注：该区间不一定包含σ.

比较：σ^2 的点估计值为 s^2= 0.005

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) = \frac{1}{15}\left(\sum_{i=1}^{16} x_i^2 - 16\overline{x}^2\right)$$

脉搏次数的估计

例2、某人自测每分钟脉搏次数, 测得如下数据(15项):

71,72,64,68,60,79,61,66,72,73,56,82,70,66,71

试以 95% 的置信度估计脉搏次数平均值及方差的置信区间.

解：假定每分钟脉搏数X服从正态分布, 则 $X \sim N(\mu, \sigma^2)$

(1) 需估计μ,而σ^2未知: α= 0.05 , n= 15 ,

取 $T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ 有 $t_{0.025}(14)$= 2. 145 ,

$$\left[\overline{X}-\frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1), \overline{X}+\frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)\right]$$

$\because \overline{x} \approx 68.73,$

$$s^2=\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2-n\overline{x}^2\right)$$

$\therefore s \approx 6.91$

平均脉搏次数的一个置信区间为:[64.90, 72.56].

脉 搏 次 数 的 估 计

例2、某人自测每分钟脉搏次数, 测得如下数据(15项):

71,72,64,68,60,79,61,66,72,73,56,82,70,66,71

试以 95% 的置信度估计脉搏次数平均值及方差的置信区间.

(2) 需估计σ^2 ,而 μ 未知

取 $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

有 $\chi^2_{0.025}(14)=$26.119, $\chi^2_{0.975}(14)=$5.629 ,

$$\left[\, (n-1)S^2 \big/ \chi^2_{\alpha/2}(n-1), (n-1)S^2 \big/ \chi^2_{1-\alpha/2}(n-1) \right]$$

$\because 14 \times S^2 = 668.47$

$\therefore \sigma^2$的一个置信区间为: [25.59, 118.76]

两个正态总体的情况

- 实际中存在这样的问题：
- 已知产品的某一指标服从正态分布，但由于原料、设备条件、操作人员不同，或工艺过程的改变等因素的影响，而引起总体均值、方差的改变。
- 我们要考察这些变化的大小，这就涉及两个正态总体均值差或方差比的估计问题。
- 设有两个正态总体 $N(\mu_1,\sigma_1^2),N(\mu_2,\sigma_2^2)$，样本均值和方差分别为 $\overline{X},\overline{Y},S_1^2,S_2^2$

二、两个正态总体：$X\sim N(\mu_1,\sigma_1^2)$， $Y\sim N(\mu_2,\sigma_2^2)$

1. $\mu_1-\mu_2$ 的估计

总体均值差的置信区间的含义是：

- 若$\mu_1-\mu_2$的置信下限大于零，则可认为 $\mu_1>\mu_2$；

  若$\mu_1-\mu_2$的置信上限小于零，则可认为 $\mu_1<\mu_2$.

2. σ_2^2/σ_1^2的估计

总体方差的置信区间的含义是：

- 若σ_2^2/σ_1^2的置信下限大于1，则可认为 $\sigma_2^2>\sigma_1^2$；

  若σ_2^2/σ_1^2的置信上限小于1，则可认为 $\sigma_2^2<\sigma_1^2$.

复习: 抽样分布定理(正态总体的统计量分布）

定理6.2.4 设 $X_1, X_2, \dots, X_n$ 是正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，

$\overline{X}, S^2$ 分别是样本均值和样本方差，则

(1) $\overline{X}$ 与 S^2 相互独立; $\Leftrightarrow \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$; (2) $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$;

(3) $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$; $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$; (4) $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

定理6.2.5: 设正态总体 X 与 Y 相互独立，

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

(2) 当 $\sigma_1^2 = \sigma_2^2$ 时，

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

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

思考: $\sigma_1^2 \neq \sigma_2^2$ 时的情况？提示: $n_1 = n_2$ (成对抽取)

被估参数	条件	枢轴变量	优良估计量
$\mu_1-\mu_2$	已知σ_1^2与σ_2^2	$U=\dfrac{(\bar X-\bar Y)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$\bar X-\bar Y$
$\mu_1-\mu_2$	未知σ_1^2和σ_2^2 $(\sigma_1^2=\sigma_2^2)$	$\dfrac{(\bar X-\bar Y)-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$	$\bar X-\bar Y$
$\dfrac{\sigma_2^2}{\sigma_1^2}$	未知μ_1和μ_2	$\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\dfrac{\sum_{i=1}^{n_1}\left(\dfrac{X_i-\bar X}{\sigma_1}\right)^2\Big/n_1-1}{\sum_{j=1}^{n_2}\left(\dfrac{Y_j-\bar Y}{\sigma_2}\right)^2\Big/n_2-1}\sim F(n_1-1,n_2-1)$	$\dfrac{S_2^2}{S_1^2}=\dfrac{\dfrac{1}{n_1-1}\sum_{j=1}^{n_2}(Y_j-\bar Y)^2}{\dfrac{1}{n_2-1}\sum_{i=1}^{n_1}(X_i-\bar X)^2}$
$\dfrac{\sigma_2^2}{\sigma_1^2}$	已知μ_1和μ_2	$\dfrac{\sum_{i=1}^{n_1}\left(\dfrac{X_i-\mu_1}{\sigma_1}\right)^2\Big/n_1}{\sum_{j=1}^{n_2}\left(\dfrac{Y_j-\mu_2}{\sigma_2}\right)^2\Big/n_2}\sim F(n_1,n_2)$	$\dfrac{\dfrac{1}{n_2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2}{\dfrac{1}{n_1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2}$

1. $\mu_1-\mu_2$的区间估计

1) σ_1^2和σ_2^2已知：$\overline{X}\sim N(\mu_1,\dfrac{\sigma_1^2}{n_1}),\overline{Y}\sim N(\mu_2,\dfrac{\sigma_2^2}{n_2})$（相互独立）

线性函数$Z=\overline{X}-\overline{Y}\sim N(\mu_1-\mu_2,\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2})$

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

$\overline{X}-\overline{Y}$为$\mu_1-\mu_2$优良估计量

$$P\{-u_{\frac{\alpha}{2}}\le\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\le u_{\frac{\alpha}{2}}\}=1-\alpha$$

$$[\overline{X}-\overline{Y}-u_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\overline{X}-\overline{Y}+u_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}]$$

1. $\mu_1-\mu_2$的区间估计

2) σ_1^2和σ_2^2未知， $\sigma_1^2=\sigma_2^2=\sigma^2$;

$\overline{X}-\overline{Y}$为$\mu_1-\mu_2$优良估计量

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

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

$$P\{-t_{\frac{\alpha}{2}}(n_1+n_2-2)\le\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\le t_{\frac{\alpha}{2}}(n_1+n_2-2)\}=1-\alpha$$

$$[\overline{X}-\overline{Y}-t_{\frac{\alpha}{2}}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\overline{X}-\overline{Y}+t_{\frac{\alpha}{2}}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}]$$

2. $\dfrac{\sigma_2^2}{\sigma_1^2}$ 的区间估计

1）μ_1、μ_2未知

$$\text{或者}F=\frac{\dfrac{1}{n_1-1}}{\dfrac{1}{n_2-1}}\cdot\frac{\sum_{i=1}^{n_1}\left(\dfrac{X_i-\overline{X}}{\sigma_1}\right)^2}{\sum_{j=1}^{n_2}\left(\dfrac{Y_j-\overline{Y}}{\sigma_2}\right)^2}$$

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

$$P\{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)\le\frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{S_1^2}{S_2^2}\le F_{\frac{\alpha}{2}}(n_1-1,n_2-1)\}=1-\alpha$$

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

2. $\dfrac{\sigma_2^2}{\sigma_1^2}$ 的区间估计

2）已知μ_1 与μ_2

$$\frac{n_1}{\sigma_1}\frac{1}{n_1}\sum_{i=1}^{n_1}\left(X_i-\mu_1\right)^2\sim\chi^2(n_1),\quad \frac{n_2}{\sigma_2}\frac{1}{n_2}\sum_{j=1}^{n_2}\left(Y_j-\mu_2\right)^2\sim\chi^2(n_2)$$

$$\boxed{F=\frac{\dfrac{1}{n_1}\cdot\displaystyle\sum_{i=1}^{n_1}\left(\frac{X_i-\mu_1}{\sigma_1}\right)^2}{\dfrac{1}{n_2}\displaystyle\sum_{j=1}^{n_2}\left(\frac{Y_j-\mu_2}{\sigma_2}\right)^2}\sim F(n_1,n_2)}$$

两稻种产量的期望差的置信区间

例3: 甲、乙两种稻种分别种在10块试验田中，每块田中甲、乙稻种各种一半. 假设两种稻种产量X、Y 服从正态分布，且方差相等.

10块田中的产量如下表 (单位：公斤)，求两稻种产量的期望差 $\mu_1 - \mu_2$ 的置信区间(α=0.05).

甲	140	137	136	140	145	148	140	135	144	141
乙	135	118	115	140	128	131	130	115	121	125

解: 设$X \sim N(\mu_1,\sigma_1^2)$，$Y \sim N(\mu_2,\sigma_2^2)$， $\sigma_1^2=\sigma_2^2=\sigma^2$

要估计$\mu_1 - \mu_2$，取枢轴变量

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

解：设$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，$\sigma_1^2=\sigma_2^2=\sigma^2$

要估计$\mu_1-\mu_2$，取枢轴变量

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

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

由样本表可计算得：

$$\bar{x}=140.6 \qquad s_1^2=16.933 \qquad n_1=10$$

$$\bar{y}=126.8 \qquad s_2^2=71.956 \qquad n_2=10$$

从而，$S_w=\sqrt{\dfrac{9\times16.933+9\times71.956}{18}}=6.667$

查t分布表得：$t_{0.025}(18)=$ 2.1009

可得两稻种产量期望差的置信度为95%的置信区间为：

$$\left[\overline{X}-\overline{Y}-t_{\frac{\alpha}{2}}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\overline{X}-\overline{Y}-t_{\frac{\alpha}{2}}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

可得两稻种产量期望差的一个区间估计为：

$$\left[140.6-126.8-2.1009\times6.667\sqrt{\frac{2}{10}},140.6-126.8-2.1009\times6.667\sqrt{\frac{2}{10}}\right]$$

即 $[7.536,\quad 20.064]$

若$\sigma_1^2\neq\sigma_2^2$,但 $n_1=n_2=n$时，能否用其它的方法来估计？

例4: 设$X_1, X_2, \cdots, X_n$为来自泊松分布总体 $P(\lambda)$的样本, 求参数λ的置信度为1-α的置信区间.

解: 由于泊松分布$P(\lambda)$的数学期望和方差都是λ.

由中心极限定理, 当n足够大时

$$\left(\sum_{i=1}^{n} X_i - n\lambda\right) / \sqrt{n\lambda}$$ 近似地服从$N(0,1)$,

所以有

$$P\left\{-u_{\frac{\alpha}{2}} \le \frac{\sum_{i=1}^{n} X_i - n\lambda}{\sqrt{n\lambda}} \le u_{\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

等价于 $P\{A \le \lambda \le B\} \approx 1-\alpha$

其中, A和B是下列二次方程的两个根,

$$\left(\sum_{i=1}^{n} X_i - n\lambda\right)^2 = n\lambda u_{\alpha/2}^2$$

即

$$A = \overline{X} + \frac{1}{2n}u_{\frac{\alpha}{2}}^2 - u_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}}{n} + \frac{1}{4n^2}u_{\frac{\alpha}{2}}^2},$$

$$B = \overline{X} + \frac{1}{2n}u_{\frac{\alpha}{2}}^2 + u_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}}{n} + \frac{1}{4n^2}u_{\frac{\alpha}{2}}^2}.$$

故得到λ的置信区间为$[A, B]$, 置信度近似地为$1-\alpha$.

三、大样本方法构造置信区间（非正态总体）

大样本方法就是本质上这是利用近似分布代替精确分布以构造近似置信区间.

其主要思想是中心极限定理. 利用极限分布确定枢轴变量的分布, 进而构造出置信区间.

称为大样本区间估计

在统计学中, 一个统计方法如果依据的是有关变量的精确分布, 不论样本容量多大都称为小样本方法(n为固定的).

而一个统计方法如果是基于有关变量的极限分布($n\to\infty$), 则称这个方法是大样本方法.

四、单侧置信区间

前面讨论的区间估计问题, 其置信区间都有两个有限的端点, 这样的置信区间称为双侧置信区间.

在有些实际问题中, 我们常常关心的是未知参数至少有多大(例如设备、元件的使用寿命等), 或者是未知参数至多是多少(例如产品的不合格品率、杂质含量等,

这就引出了只有一个有限端点的单侧置信区间概念.

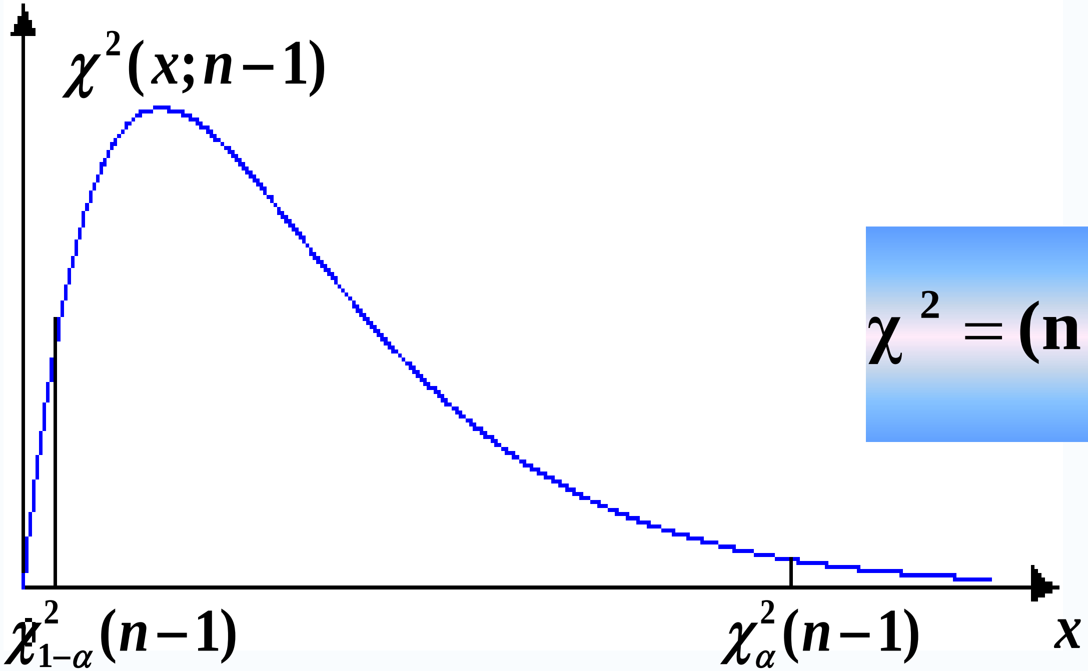
定义：设 $X_1, X_2, \cdots, X_n$ 是来自某个总体的样本，总体分布包含未知参数 θ . $\hat{\theta}_1(X_1, \cdots, X_n)$ 和 $\hat{\theta}_2(X_1, \cdots, X_n)$ 是 θ 的统计量．如果对 θ 的一切可能取值，有

$$P\{\hat{\theta}_1(X_1, X_2, \cdots, X_n) \le \theta\} = 1-\alpha$$

则称随机区间[$\hat{\theta}_1$, +∞]为参数 θ 的置信度为1-α 的单侧置信区间. $\hat{\theta}_1$ 称为单侧置信下限.

$$P\{\hat{\theta}_2(X_1, X_2, \cdots, X_n) \ge \theta\} = 1-\alpha$$

则称随机区间[—∞ , $\hat{\theta}_2$]为参数 θ 的置信度为1-α 的单侧置信区间. $\hat{\theta}_2$ 称为单侧置信上限.



$$\chi^2=(n-1)\frac{S^2}{\sigma_0^2}\sim\chi^2(n-1)$$

$$P\{\frac{n-1}{\sigma^2}S^2>\underline{\chi^2_{1-\alpha}(n-1)}\}=1-\alpha \qquad \Rightarrow P\{\frac{n-1}{\chi^2_{1-\alpha}(n-1)}S^2>\sigma^2\}=1-\alpha$$

或

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