第 七 章 参数估计

第三节 区间估计

1、区间估计的定义

2、单个正态总体的区间估计

3、两个正态总体的区间估计

4、小结、思考

点估计的缺陷：

1) 估计值只是真实值的近似值，它与真实值误差范围没有指明
2) 估计值的可靠性没有指明

改进：

对于θ的估计，给定一个范围：$\left[\hat{\theta}_1, \hat{\theta}_2\right]$

可以指明范围的可靠性

区间估计的一般表述：

要求构造一个仅依赖样本的区间**[A(X),B(**X**)]**，一旦得到样本**X**的观测值 x ，就把区间**[A(x),B(x)]**作为θ的估计

定义

设总体的未知参数为θ，由样本X_1，…，X_n确定两个统计量 $\hat{\theta}_1=\hat{\theta}_1(X_1,\ldots,X_n)$和 $\hat{\theta}_2=\hat{\theta}_2(X_1,\ldots,X_n)$对于给定的实数$\alpha$(0< α <1)，满足

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

则称随机区间 $[\hat{\theta}_1,\hat{\theta}_2]$为$\theta$的置信度为1-$\alpha$的置信区间。

> 1-α又称置信系数或置信概率
> α又称置信水平，通常取值为0.1，0.05。

注:

1）对 $P\{\hat{\theta}_1\le\theta\le\hat{\theta}_2\}=1-\alpha$ 的理解。

随机区间 $[\hat{\theta}_1,\hat{\theta}_2]$ 包含θ的概率为1-α。

2）评价随机区间 $\left[\hat{\theta}_1, \hat{\theta}_2\right]$ 的优劣有两个因素：

（1） $P\left\{\hat{\theta}_1 \le \theta \le \hat{\theta}_2\right\}$的大小　　可靠度

（2） $\hat{\theta}_2 - \hat{\theta}_1$ 的长度　　精度

这两者往往是相互矛盾的。

著名统计学家 奈曼 提出了处理原则：

先照顾可靠度，在此前提下，使精度尽可能高。

正态分布中μ的区间估计

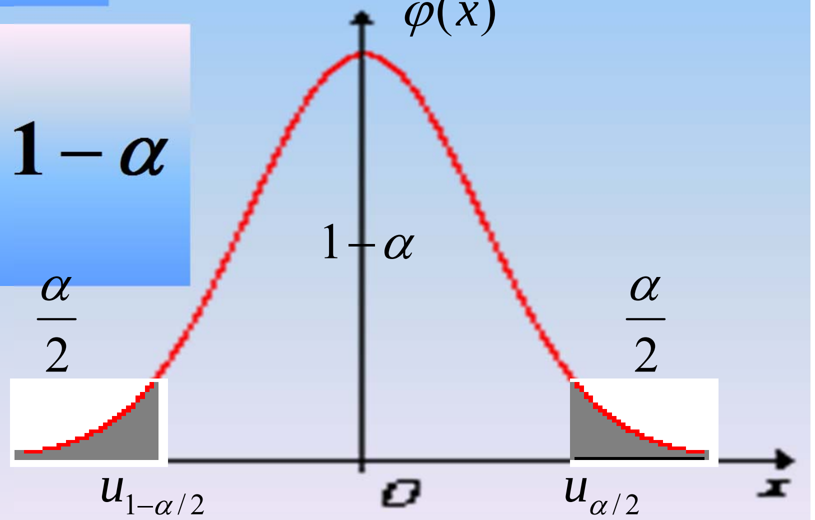
例 设$X \sim N(\mu,\sigma^2)$， $\sigma^2=\sigma_0^2$，求参数μ的置信度为$1-\alpha$的置信区间。

解：$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$是$\mu$的优良估计，且 $\bar{X}\sim N(\mu,\frac{\sigma_0^2}{n})$

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

由 $$P\left\{u_{-\alpha/2}\le U\le u_{\alpha/2}\right\}=1-\alpha$$

即 $$P\left\{-u_{\frac{\alpha}{2}}\le \frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\le u_{\frac{\alpha}{2}}\right\}=1-\alpha$$



可得：$$P\left\{\overline{X}-\frac{\sigma_0}{\sqrt{n}}u_{\frac{\alpha}{2}}\le\mu\le\overline{X}+\frac{\sigma_0}{\sqrt{n}}u_{\frac{\alpha}{2}}\right\}=1-\alpha$$

由此可得，μ的置信度为$1-\alpha$的置信区间为：

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

特别，当$\sigma_0=1$，$\alpha=0.05$，样本观测值为：

5.1	5.1	4.8	5.0	4.7	5.0	5.2	5.1	5.0

$u_{\alpha/2}=$ 1.96，μ 的置信区间为：[4.35，5.65]

#

寻找置信区间的步骤(枢轴变量法)：

1）选取待估参数θ的估计量；

原则：优良性准则

常用：$\overline{X} \to \mu$，$S^2 \to \sigma^2$

2) 考察估计量服从的分布（若有其他未知参数，则选一优良估计量来替代)；利用抽样分布定理化至常用分布(主要是：正态、χ^2、T、F分布)；相应的变换函数 W 称为枢轴变量

3) 对 $P\left\{w_{1-\alpha/2} \le W \le w_{\alpha/2}\right\} = 1-\alpha$ 查上侧分位数；

4) 代换得到 $P\{A \le \theta \le B\} = 1-\alpha$

区间 $[A，B]$ 即为所求。

估计量的选取

例 设X~N(μ,σ^2)， 求参数σ^2的置信度为1-α的置信区间。

分析:

1) 当μ未知时，应选统计量为： S^2

要化至常用分布，由抽样分布定理可知:

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

2) 当μ已知时，应选统计量为： $\because \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$

原因：它是σ^2的无偏、有效、相合估计量。

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

#

例　设X~N(μ,σ^2)， σ^2 未知，求参数μ的置信度为1-α的置信区间。

解：$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$是$\mu$的优良估计， σ^2 未知

枢轴变量选：

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

由： $$P\left\{ t_{-\alpha/2}(n-1) \le T \le t_{\alpha/2}(n-1) \right\} = 1-\alpha$$

整理可得 μ 的置信区间：

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

正态总体的区间估计：一、单个正态总体：$X \sim N(\mu, \sigma^2)$

1. μ 的估计

1) σ^2已知:

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

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

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

2) σ^2未知:

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

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

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

2. σ^2 的估计

1) μ 已知:

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

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

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

2) μ未知:

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

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

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

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，用S^2作为方差的估计

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

故 方差的估计值为 $$\hat{\sigma}^2 = \frac{1}{15}\sum_{i=1}^{16}(x_i - \bar{x})^2 = 0.005$$

下面计算方差的置信区间：

由于μ未知， S^2是σ^2的优良估计，枢轴变量选：

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

相应的置信区间为：

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

查 χ^2分布表可得： $\chi^2_{\alpha/2}(n-1)=\chi^2_{0.025}(15)=\quad 27.488$

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

σ^2的置信度为0.95的置信区间为：

$$\left[\frac{0.0761}{27.488},\frac{0.0761}{6.26}\right]$$ 即 $[0.002768,0.012]$

比较：σ^2的估计值为 0.005

#

婴儿体重的估计

例、假定初生婴儿的体重服从正态分布，随机抽取12 名婴儿，测得体重为：（单位：克）

3100，2520，3000， 3000，3600，3160，

3560，3320，2880， 2600，3400，2540

试以 95% 的置信度估计初生婴儿的平均体重以及方差。

解：设初生婴儿体重为X克，则 $X \sim N(\mu, \sigma^2)$

(1) 需估计μ ,而σ^2 未知：

α= 0.05 ，n= 12 ，

t0.025(11)= 2.201 ，

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

$\because \overline{X} \approx 3057, \quad S \approx 375.3$

$\therefore \mu$的置信区间为: $[\,3057 - \dfrac{375.3}{\sqrt{12}} \times 2.201, 3057 + \dfrac{375.3}{\sqrt{12}} \times 2.201]$

即 $[\,2818, 3296\,]$

(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 $\chi^2_{0.025}(11)=$ <u>21.92</u>, $\chi^2_{0.975}(11)=$ <u>3.816</u>,

$\because 11\times S^2=1549000$

$\therefore \sigma^2$的置信区间为:$[\dfrac{1549000}{21.92},\dfrac{1549000}{3.816}]$

即 $[70666, 405922.4]$

#

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

1) σ_1^2 和 σ_2^2 已知:

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

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

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

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

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

其中，$S_w=\sqrt{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{j=1}^{n_2}(Y_j-\overline{Y})^2}=\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$

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

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

二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 $\mu_1 < \mu_2$。

两稻种产量的期望差的置信区间

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

μ_1 与 μ_2 已知

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

$$P\{F_{1-\frac{\alpha}{2}}(n_1,n_2)\le\frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{\frac{1}{n_1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\frac{1}{n_2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2}\le F_{\frac{\alpha}{2}}(n_1,n_2)\}=1-\alpha$$

$$[\frac{\frac{1}{n_2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2}{\frac{1}{n_1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2}\cdot F_{1-\frac{\alpha}{2}}(n_1,n_2),\ \frac{\frac{1}{n_2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2}{\frac{1}{n_1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2}\cdot F_{\frac{\alpha}{2}}(n_1,n_2)]$$

μ_1、μ_2 未知

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

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

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

例　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

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

其中，$S_w=\sqrt{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{j=1}^{n_2}(Y_j-\overline{Y})^2}=\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$

由样本表可算得：

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

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

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：

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

$$=\left[140.6-126.8-2.10096.667\sqrt{\frac{2}{10}},140.6-126.8-2.10096.667\sqrt{\frac{2}{10}}\ \right]$$

即 $[7.536, 20.064]$ #

小结与思考

1、总结区间估计的思想与做法；

2、单个正态总体：$X \sim N(\mu,\sigma^2)$中，如果σ已知，能否用

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

作为枢轴变量来估计μ，*结果会怎样？*

3、假如总体的分布不是正态分布，如何对参数给出区间估计？