# 2.4　区间估计

一、区间估计的概念

二、正态总体数学期望的置信区间

三、正态总体方差的区间估计

四、两个正态总体均值差的区间估计

五、两个正态总体方差比的区间估计

六、单侧置信区间

七、非正态总体参数的区间估计

# 一、区间估计基本概念

## 1. 问题的提出

点估计法：找一个统计量 $\hat{\theta}$ 来估计参数 $\theta$,

$$\theta \approx \hat{\theta}_0 \quad (\hat{\theta}\text{的观察值})$$

不足之处：未给出估计值量 $\hat{\theta}_0$ 与参数真值 $\theta$ 的误差及估计的可靠程度.

例如 $E(X) = \mu$,

$$\mu \approx \hat{\mu} = \overline{x} \text{ — 样本均值的观察值}$$

问：误差 $|\overline{x}-\mu|$ 多大？

用 $\overline{X}$ 估计 $\mu$ 的可靠程度如何？

即　给定 $\alpha$：　$0<\underline{\alpha<<1}$ 很小

使　$P\{|\overline{X}-\mu|<\varepsilon\}=\underline{1-\alpha}$ 较大

中的 $\varepsilon=?$

区间估计解决了上述问题，从而克服了点估计的不足之处.

## 2. 置信区间与置信度

**定义2.11** 设总体 $X\sim F(x;\theta)$, $\theta$为未知参数,

$(X_1, X_2, \cdots, X_n)$是来自总体 $X$ 的样本.

如果存在两个统计量

$$\hat{\theta}_1(X_1, X_2, \cdots, X_n) \text{ 和 } \hat{\theta}_2(X_1, X_2, \cdots, X_n)$$

使得对于给定的 $\alpha$ $(0<\alpha<1)$,

$$P\{\hat{\theta}_1(X_1, X_2, \cdots, X_n) \le \theta \le \hat{\theta}_2(X_1, X_2, \cdots, X_n)\} = 1-\alpha.$$

则称区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 为 $\theta$ 的置信度为 $1-\alpha$的**置信区间,**

置信下限　　　　置信上限

## 关于定义的说明

被估计的参数 $\theta$ 虽然未知,但它是一个常数,没有随机性, 而区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 是随机的.

因此定义中下表达式

$$P\{\hat{\theta}_1 \le \theta \le \hat{\theta}_2\} = 1 - \alpha$$

的本质是:

随机区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 以 $1-\alpha$ 的概率包含着参数 $\theta$ 的真值,

而不能说参数 $\theta$ 以 $1-\alpha$ 的概率落入随机区间 $[\hat{\theta}_1, \hat{\theta}_2]$.

另外定义中的表达式

$$P\{\hat{\theta}_1 \le \theta \le \hat{\theta}_2\} = 1-\alpha$$

还可以描述为:

若反复抽样多次(各次得到的样本容量相等,都是$n$)

每个样本值确定一个区间$[\hat{\theta}_1, \hat{\theta}_2]$,

每个这样的区间或包含$\theta$的真值或不包含$\theta$的真值,

按贝努利大数定理,当抽样次数充分大时,在这些区间中包含$\theta$真值的频率接近置信度$1-\alpha$,即

包含$\theta$真值的约占$100(1-\alpha)\%$,

不包含的约占$100\alpha\%$.

由定义可见，

对参数$\theta$作区间估计，就是要设法找出两个只依赖于样本的界限(构造统计量)

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

$$\hat{\theta}_2 = \hat{\theta}_2(X_1,\ldots X_n)$$

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

一旦有了样本，就把$\theta$估计在区间$[\hat{\theta}_1, \hat{\theta}_2]$内. 这里有两个要求:

1. 要求 $\theta$ 以很大的可能被包含在区间$[\hat{\theta}_1, \hat{\theta}_2]$内，就是说，概率$P\{\hat{\theta}_1 \le \theta \le \hat{\theta}_2\}$要尽可能大. 即要求估计尽量可靠.

2. 估计的精度要尽可能的高. 如要求区间长度 $\hat{\theta}_2 - \hat{\theta}_1$ 尽可能短，或能体现该要求的其它准则.

可靠度与精度是一对矛盾，一般是在保证可靠度的条件下尽可能提高精度.

## 3.求置信区间的一般步骤(共3步)

1寻求一个样本 $X_1, X_2, \cdots, X_n$ 和待估参数 $\theta$ 的函数:

$$u = u(X_1, X_2, \cdots, X_n; \theta)$$

枢轴量

（仅包含待估参数 $\theta$), 并且 $u$ 的分布已知且不依赖于任何未知参数(包括 $\theta$).

2 对于给定的置信度 $1-\alpha$, 定出两个常数 $a, b$, 使

$$P\{a \le u(X_1, X_2, \cdots, X_n; \theta) \le b\} = 1-\alpha.$$

3求解不等式

$$a \le u(X_1, X_2, \cdots, X_n; \theta) \le b \Leftrightarrow \hat{\theta}_1 \le \theta \le \hat{\theta}_2,$$

$$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n; a, b), \hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n; a, b)$$

那么 $[\hat{\theta}_1, \hat{\theta}_2]$ 就是 $\theta$ 的一个置信度为 $1-\alpha$ 的置信区间.

# 二、正态总体数学期望的置信区间

设给定置信度为$1-\alpha$,并设$X_1, X_2, \cdots, X_n$为总体$N(\mu, \sigma^2)$的样本,$\bar{X}, S_n^{*2}$分别是样本均值和修正样本方差.

**(1) $\sigma^2$为已知,** 求$\mu$的置信区间

因为$\bar{X}$是$\mu$的估计，又因为

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

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

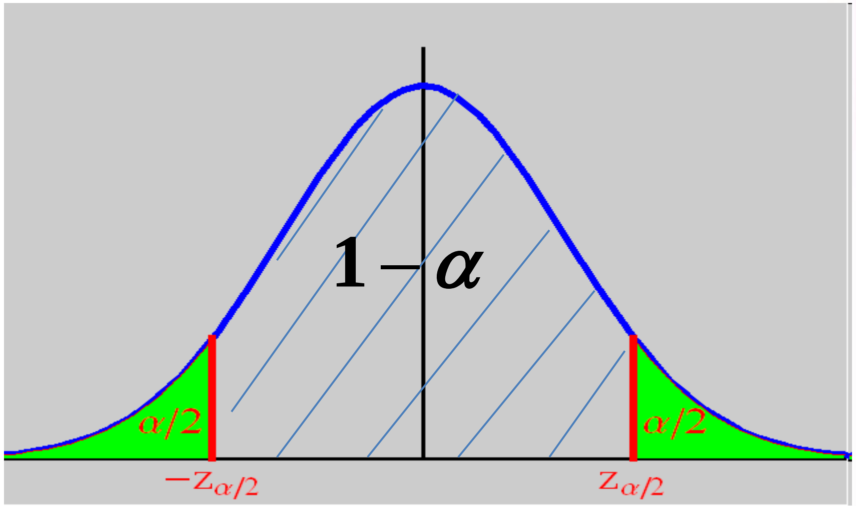
U满足枢轴量的条件

$2^\circ$ 对于给定的置信度$1-\alpha$,

定出常数$a,b$,使

$$P\{a\le U\le b\}=1-\alpha.$$

如何选$a,b$,能使得精度最高?




显然,$b=u_{\frac{\alpha}{2}},a=-u_{\frac{\alpha}{2}}$

满足区间精度最高。因此 $P\{\,|U|\le u_{\alpha/2}\,\}=1-\alpha$,

3° 作等价变形

$$|U| \le u_{\frac{\alpha}{2}} \Leftrightarrow \left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \le u_{\frac{\alpha}{2}} \Leftrightarrow |\bar{X}-\mu| \le \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}}$$

$$\Leftrightarrow \bar{X}-\frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}} \le \mu \le \bar{X}+\frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}}$$

$\therefore$ $\mu$的一个置信度为$1-\alpha$的置信区间

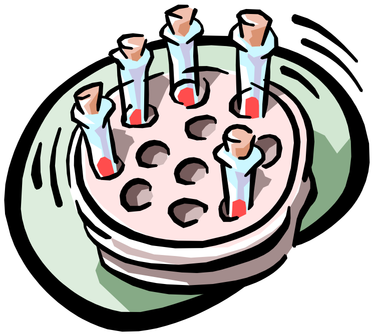
$[\bar{X}-\frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}},\ \bar{X}+\frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}}]$, 简写成 $\left(\bar{X} \pm \frac{\sigma}{\sqrt{n}} u_{\alpha/2}\right)$.

其置信区间的长度为 $2\times\frac{\sigma}{\sqrt{n}} u_{\alpha/2}$ .

**注** 上述区间长度最小，即最优

例1 包糖机某日开工包了12 包糖,称得重量(单位:克)分别为506, 500, 495, 488, 504, 486, 505, 513, 521, 520, 512, 485. 假设重量服从正态分布,且标准差为$\sigma=10$, 试求糖包的平均重量 $\mu$ 的 $1-\alpha$ 置信区间(分别取 $\alpha=0.10$ 和 $\alpha=0.05$).

**解** $\sigma=10,\quad n=12,$



计算得 $\bar{x}=502.92,$

(1) 当 $\alpha=0.10$时, $1-\dfrac{\alpha}{2}=0.95,$

查表得 $u_{\alpha/2}=u_{0.05}=1.645,$

$$\bar{x}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2} = 502.92-\frac{10}{\sqrt{12}}\times 1.645 = 498.17,$$

$$\bar{x}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2} = 502.92+\frac{10}{\sqrt{12}}\times 1.645 = 507.67,$$

即$\mu$的置信度为$90\%$的置信区间为

[498.17, 507.67].

(2) 当$\alpha = 0.05$时，$1-\dfrac{\alpha}{2}=0.975,$

查表得 $u_{\alpha/2} = u_{0.025} = 1.96,$

同理可得$\mu$的置信度为$95\%$的置信区间为

[497.26, 508.58].

显然，

$1-\alpha$越大，

置信区间

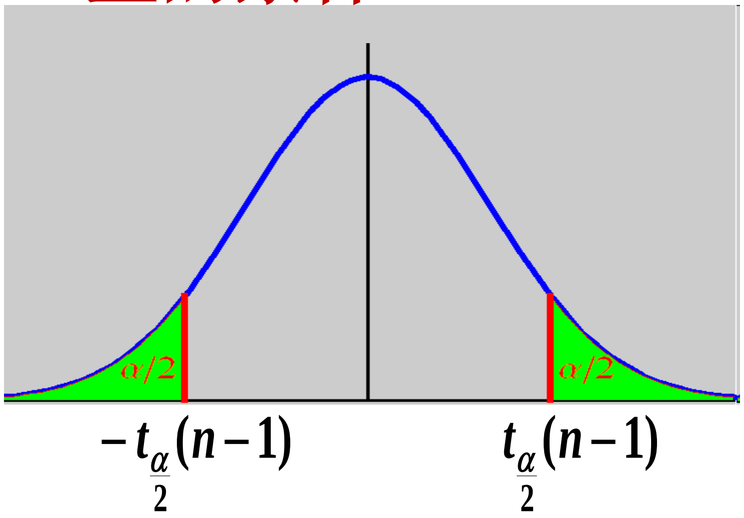
越大。

## (2) $\sigma^2$为未知, 求 $\mu$ 的置信区间

$1^\circ$ 取 $T=\dfrac{\overline{X}-\mu}{S_n^*/\sqrt{n}}\sim t(n-1)$, *T*满足枢轴量的条件

$2^\circ$ 给定定$1-\alpha$，选取

$b=-a=t_{\frac{\alpha}{2}}(n-1)$,




$$P\left\{\,|T|\le t_{\frac{\alpha}{2}}(n-1)\right\}=1-\alpha,$$

$3^\circ$ 作等价变形

$$|T|\le t_{\frac{\alpha}{2}}(n-1)\Leftrightarrow\left|\frac{\overline{X}-\mu}{S_n^*/\sqrt{n}}\right|\le t_{\frac{\alpha}{2}}(n-1)$$

$$\Leftrightarrow \quad \left|\frac{\overline{X}-\mu}{S_n^*/\sqrt{n}}\right| \le t_{\frac{\alpha}{2}}(n-1)$$

$$\Leftrightarrow \quad \overline{X}-\frac{S_n^*}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1) \le \mu \le \overline{X}+\frac{S_n^*}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)$$

$\mu$的置信度为$1-\alpha$的置信区间

$$[\overline{X}-\frac{S_n^*}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1),\ \overline{X}+\frac{S_n^*}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)],$$

简写成 $\left(\overline{X}\pm\frac{S_n^*}{\sqrt{n}}t_{\alpha/2}(n-1)\right).$

例2 有一大批糖果,现从中随机地取16袋, 称得重量(克)如下:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 506 | 508 | 499 | 503 | 504 | 510 | 497 | 512 |
| 514 | 505 | 493 | 496 | 506 | 502 | 509 | 496 |

设袋装糖果的重量服从正态分布, 试求总体均值 $\mu$ 的置信度为 0.95 的置信区间.

**解** $\alpha = 0.05,\quad n-1=15,$

查 $t(n-1)$ 分布表可知: $t_{\frac{\alpha}{2}}(15) = t_{0.025}(15) = 2.1315,$

计算得 $\bar{x} = 503.75,\quad s_n^* = 6.2022,$

得$\mu$的置信度为95%的置信区间

$$\left(503.75 \pm \frac{6.2022}{\sqrt{16}} \times 2.1315\right) \text{即} \quad [500.4, \quad 507.1].$$

就是说估计袋装糖果重量的均值在500.4克与507.1克之间, 这个估计的可信程度为95%.

若依此区间内任一值作为$\mu$的近似值,

其误差不大于 $\dfrac{6.2022}{\sqrt{16}} \times 2.1315 \times 2 = 6.61$(克).

这个误差的可信度为95%.

(续例1)如果只假设糖包的重量服从正态分布 $N(\mu,\sigma^2)$, 试求糖包重量 $\mu$ 的95%的置信区间.

$$\left(\bar{X} \pm \frac{S_n^*}{\sqrt{n}} t_{\alpha/2}(n-1)\right).$$

证　此时$\sigma$未知, $n=12$,

$\alpha=0.05$, $\bar{x}=502.92$, $s_n^*=12.35$,

查 $t(n-1)$ 分布表可知: $t_{0.025}(11)=2.201$,

于是 $\frac{s_n^*}{\sqrt{n}} t_{\alpha/2}(n-1)=\frac{12.35}{\sqrt{12}}\times 2.201=7.85$,

得$\mu$的置信度为95%的置信区间 [495.07, 510.77].

# 三、正态总体方差的区间估计

根据实际需要，只介绍 $\mu$ 未知的情况.

方差$\sigma^2$的置信度为$1-\alpha$的置信区间

$$\left( \frac{(n-1)S_n^{*2}}{\chi^2_{\alpha/2}(n-1)},\ \frac{(n-1)S_n^{*2}}{\chi^2_{1-\alpha/2}(n-1)} \right).$$

推导过程如下：

因为 $S_n^{*2}$ 是 $\sigma^2$ 的无偏估计，

根据第1章第三节定理1.12可知 $\dfrac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1)$,

则 $$P\left\{\chi^2_{1-\alpha/2}(n-1) < \frac{(n-1)S_n^{*2}}{\sigma^2} < \chi^2_{\alpha/2}(n-1)\right\} = 1-\alpha,$$

即 $$P\left\{\frac{(n-1)S_n^{*2}}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S_n^{*2}}{\chi^2_{1-\alpha/2}(n-1)}\right\} = 1-\alpha,$$

于是得方差$\sigma^2$的置信度为$1-\alpha$的置信区间

$$\left(\frac{(n-1)S_n^{*2}}{\chi^2_{\alpha/2}(n-1)},\ \frac{(n-1)S_n^{*2}}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

进一步可得:

标准差$\sigma$的一个置信度为$1-\alpha$的置信区间

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

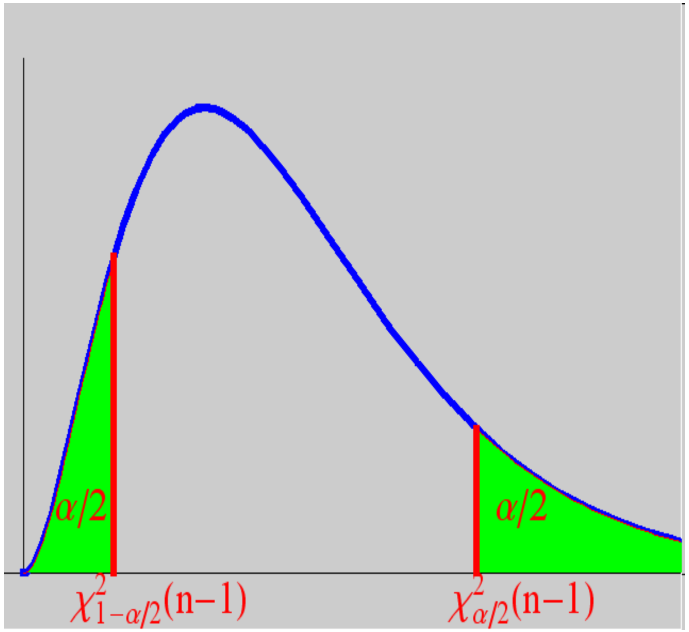
注 此置信区间长度并非最短

注意: 在密度函数不对称时,

如 $\chi^2$ 分布和 $F$分布,

习惯上仍取对称的分位点

来确定置信区间(如图).



(续例2) 求例2中总体标准差σ的置信度为0.95的置信区间.

**解** $\frac{\alpha}{2}=0.025,\quad 1-\frac{\alpha}{2}=0.975,\quad n-1=15,$

查 $\chi^2(n-1)$ 分布表可知:

$$\chi^2_{0.025}(15)=27.488,\quad \chi^2_{0.975}(15)=6.262,$$

计算得 $s_n^*=6.2022,$ $\left(\frac{\sqrt{n-1}S_n^*}{\sqrt{\chi^2_{\alpha/2}(n-1)}},\ \frac{\sqrt{n-1}S_n^*}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}}\right).$

代入公式得标准差的置信区间

$[4.58,\ 9.60].$

# 四、两个正态总体均值差的区间估计

设给定置信度为$1-\alpha$，并设$X_1, X_2, \cdots, X_{n_1}$为第一个总体$N(\mu_1, \sigma_1^2)$的样本，$Y_1, Y_2, \cdots, Y_{n_2}$为第二个总体$N(\mu_2, \sigma_2^2)$的样本，$\overline{X}, \overline{Y}$分别是第一、二个总体的样本均值，$S_1^{*2}, S_2^{*2}$分别是第一、二个总体的修正样本方差.

本章将讨论两个总体均值差和方差比的估计问题.

(1) $\sigma_1^2$ 和 $\sigma_2^2$ 均为已知,

$\mu_1-\mu_2$ 的一个置信度为 $1-\alpha$ 的置信区间

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

推导过程如下:

因为 $\overline{X}$, $\overline{Y}$ 分别是 $\mu_1,\mu_2$ 的无偏估计,

所以 $\overline{X}-\overline{Y}$ 是 $\mu_1-\mu_2$ 的无偏估计,

由 $\overline{X}, \overline{Y}$ 的独立性及

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

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

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

于是得$\mu_1-\mu_2$的一个置信度为$1-\alpha$的置信区间

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

(2) $\sigma_1^{\ 2}$和$\sigma_2^{\ 2}$均为未知,

只要$n_1$和$n_2$都很大(实用上$>50$即可),则有

$\mu_1-\mu_2$的一个置信度为$1-\alpha$的近似置信区间

$$\left(\overline{X}-\overline{Y}\pm u_{\alpha/2}\sqrt{\frac{S_1^{*2}}{n_1}+\frac{S_2^{*2}}{n_2}}\right).$$

(3) $\sigma_1^2=\sigma_2^2=\sigma^2$, 但 $\sigma^2$ 为未知,

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

$\mu_1-\mu_2$的一个置信度为$1-\alpha$的置信区间

$$\left(\overline{X}-\overline{Y}\pm t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right).$$

其中 $S_w^{\ 2}=\dfrac{(n_1-1)S_{1n_1}^{*\ 2}+(n_2-1)S_{2n_2}^{*\ 2}}{n_1+n_2-2}$, $S_w=\sqrt{S_w^{\ 2}}$.

例5 为比较I, II两种型号步枪子弹的枪口速度，随机地取I型子弹10发，得到枪口速度的平均值为 $\overline{x}_1 = 500(m/s)$，修正标准差 $s^*_{1n_1} = 1.10(m/s)$，随机地取II型子弹20发，得枪口速度平均值为 $\overline{x}_2 = 496(m/s)$，修正标准差 $s^*_{2n_2} = 1.20(m/s)$，假设两总体都可认为近似地服从正态分布，且由生产过程可认为它们的方差相等，求两总体均值差 $\mu_1 - \mu_2$ 的置信度为0.95的置 信区间.

**解** 由题意，两总体样本独立且方差相等(但未知)，

$$\frac{\alpha}{2}=0.025,\quad n_1=10,\quad n_2=20,\quad n_1+n_2-2=28,$$

查 $t(n-1)$ 分布表可知： $t_{0.025}(28)=2.0484,$

$$S_w{}^2=\frac{9\times1.10^2+19\times1.20^2}{28},\quad S_w=\sqrt{S_w{}^2}=1.1688,$$

于是得 $\mu_1-\mu_2$ 的一个置信度为0.95的置信区间

$$\left(\overline{x}_1-\overline{x}_2\pm S_w\times t_{0.025}(28)\sqrt{\frac{1}{10}+\frac{1}{20}}\right)=(4\pm0.93),$$

即所求置信区间为 [3.07, 4.93].

# 五、两个正态总体方差比的区间估计

仅讨论总体均值 $\mu_1, \mu_2$ 为未知的情况.

$\dfrac{\sigma_1^2}{\sigma_2^2}$ 的一个置信度为 $1-\alpha$ 的置信区间

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

推导过程如下:

由于 $\dfrac{(n_1-1){S^*_{1n_1}}^2}{\sigma_1^2} \sim \chi^2(n_1-1)$, $\dfrac{(n_2-1){S^*_{2n_2}}^2}{\sigma_2^2} \sim \chi^2(n_2-1)$,

且由假设知 $\dfrac{(n_1-1)S_{1n_1}^{*\ 2}}{\sigma_1^{\ 2}}$ 与 $\dfrac{(n_2-1)S_{2n_2}^{*\ 2}}{\sigma_2^{\ 2}}$ 相互独立,

根据$F$分布的结构, 知 $\dfrac{S_{1n_1}^{*\ 2}\Big/\sigma_1^{\ 2}}{S_{2n_2}^{*\ 2}\Big/\sigma_2^{\ 2}} \sim F(n_1-1, n_2-1)$,

即 $$F=\frac{S_{1n_1}^{*\ 2}\Big/\sigma_1^{\ 2}}{S_{2n_1}^{*\ 2}\Big/\sigma_2^{\ 2}}=\frac{\dfrac{(n_1-1)S_{1n_1}^{*\ 2}}{\sigma_1^{\ 2}}\Big/(n_1-1)}{\dfrac{(n_2-1)S_{2n_2}^{*\ 2}}{\sigma_2^{\ 2}}\Big/(n_2-1)} \sim F(n_1-1, n_2-1),$$

$$P\left\{F_{1-\alpha/2}(n_1-1,n_2-1)\le\frac{S_{1n_1}^{*\,2}\big/\sigma_1^2}{S_{2n_2}^{*\,2}\big/\sigma_2^2}\le F_{\alpha/2}(n_1-1,n_2-1)\right\}$$

$$=1-\alpha,$$

$$P\left\{\frac{S_{1n_1}^{*\,2}}{S_{2n_2}^{*\,2}}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{S_{1n_1}^{*\,2}}{S_{2n_2}^{*\,2}}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right\}$$

$$=1-\alpha,$$

于是得$\dfrac{\sigma_1^2}{\sigma_2^2}$的一个置信度为$1-\alpha$的置信区间

$$\left[\frac{S_{1n_1}^{*\,2}}{S_{2n_2}^{*\,2}}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\ \frac{S_{1n_1}^{*\,2}}{S_{2n_2}^{*\,2}}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right].$$

**例6(p69例2.30)** 为了考察温度对某物体断裂强力的影响，在70度和80度分别重复做了8次试验，测得的断裂强力的数据如下(单位Pa):

**70度：20.5, 18.8, 19.8, 21.5, 19.5, 21.0, 21.2**

**80度：17.7, 20.3, 20.0, 18.8, 19.0, 20.1, 20.2, 19.1**

假定$70^0 C$下的断裂强力用$X$表示，且服从$N(\mu_1,\sigma_1^2)$，$80^0 C$下的断裂强力用$Y$表示，且服从$N(\mu_2,\sigma_2^2)$，试求方差比$\dfrac{\sigma_1^2}{\sigma_2^2}$的置信度为90%的置信区间.

**解** $\bar{x}=20.4,\ {s_{1n_1}^{*}}^2=0.8857,\ \bar{y}=19.4,\ {s_{1n_2}^{*}}^2=0.8286,$

$$F_{\alpha/2}(n_1-1, n_2-1) = F_{0.05}(7,7) = 3.79$$

$$F_{1-\alpha/2}(7,7) = F_{0.95}(7,7) = \frac{1}{F_{0.05}(7,7)} = \frac{1}{3.79} = 0.263$$

于是得$\dfrac{\sigma_1{}^2}{\sigma_2{}^2}$的一个置信度为**0.90**的置信区间

$$\left([\frac{0.8857}{0.8286}\times 0.2639,\ \frac{0.8857}{0.8286}\times 3.79]\right) = (0.2821, 4.0512)$$

由于在$\dfrac{\sigma_1{}^2}{\sigma_2{}^2}$的置信区间中包含**1**,在实际中我们认为$\sigma_1{}^2$和$\sigma_2{}^2$两者没有显著差别.

# 六、单侧置信区间

在以上各节的讨论中,对于未知参数$\theta$, 我们给出两个统计量$\hat{\theta}_1$, $\hat{\theta}_2$, 得到$\theta$的双侧置信区间$(\hat{\theta}_1, \hat{\theta}_2)$.

但在某些实际问题中, 例如, 对于设备、元件的寿命来说, 平均寿命长是我们希望的, 我们关心的是平均寿命$\theta$的“下限”; 与之相反, 在考虑产品的废品率$p$时, 我们常关心参数$p$的“上限”, 这就引出了单侧置信区间的概念.

## 1. 单侧置信区间的定义

对于给定值$\alpha$ $(0<\alpha<1)$, 若由样本$X_1, X_2, \cdots,$ $X_n$ 确定的统计量 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ , 对于任意 $\theta \in \Theta$满足

$$P\{\theta > \hat{\theta}_1\} \geq 1-\alpha,$$

则称随机区间$(\hat{\theta}_1, \infty)$是$\theta$的置信度为$1-\alpha$的单侧置信区间, $\hat{\theta}_1$称为$\theta$的置信度为$1-\alpha$的单侧置信下限.

又如果统计量 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$, 对于任意 $\theta \in \Theta$满足

$$P\{\theta < \hat{\theta}_2\} \geq 1 - \alpha,$$

则称随机区间 $(-\infty, \hat{\theta}_2)$ 是$\theta$的置信度为 $1-\alpha$ 的单侧置信区间, $\hat{\theta}_2$称为 $\theta$ 的置信度为 $1-\alpha$的**单侧置信上限.**

**2.** **正态总体均值与方差的单侧置信区间**

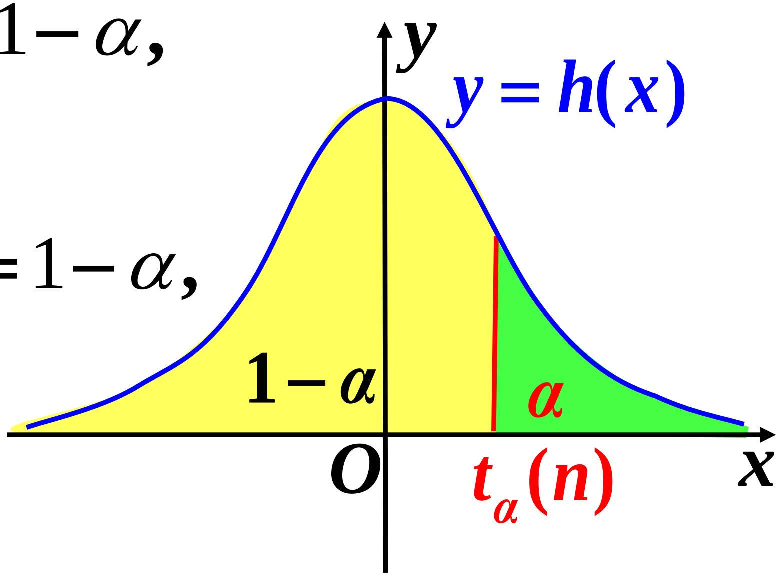
**均值的单侧置信区间**

设正态总体 $X$ 的均值是$\mu$, 方差是$\sigma^2$, (均为未知)

$X_1, X_2, \cdots, X_n$ 是一个样本, 由 $\dfrac{\bar{X}-\mu}{S_n^*/\sqrt{n}} \sim t(n-1)$,

有 $P\left\{\dfrac{\bar{X}-\mu}{S_n^*/\sqrt{n}} < t_\alpha(n-1)\right\} = 1-\alpha$,

即 $P\left\{\mu > \bar{X} - \dfrac{S_n^*}{\sqrt{n}} t_\alpha(n-1)\right\} = 1-\alpha$,



于是得$\mu$的一个置信度为$1-\alpha$的单侧置信区间

$$\left(\bar{X}-\frac{S_n^*}{\sqrt{n}}t_\alpha(n-1),\infty\right),$$

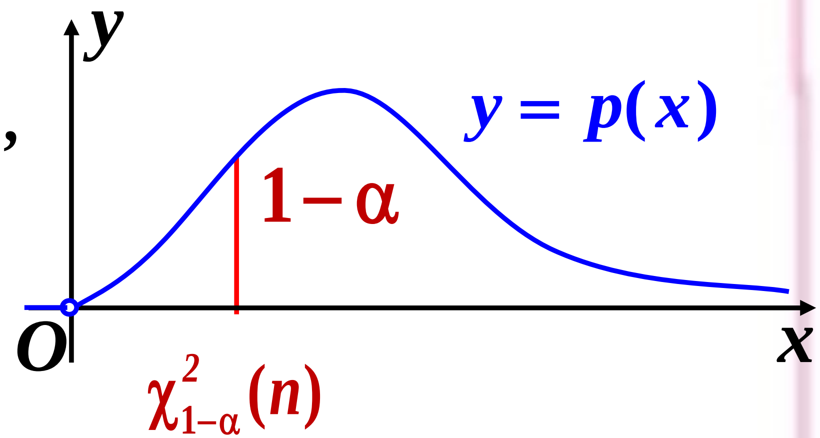
$\mu$的置信度为$1-\alpha$的置信下限 $\hat{\mu}=\bar{X}-\frac{S_n^*}{\sqrt{n}}t_\alpha(n-1).$

## 方差的单侧置信区间

又根据 $\dfrac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1)$,

有 $P\left\{\dfrac{(n-1)S_n^{*2}}{\sigma^2} > \chi^2_{1-\alpha}(n-1)\right\} = 1-\alpha$,

即 $P\left\{\sigma^2 < \dfrac{(n-1)S_n^{*2}}{\chi^2_{1-\alpha}(n-1)}\right\} = 1-\alpha$,




于是得$\sigma^2$的一个置信度为$1-\alpha$的单侧置信区间

$$\left(0, \frac{(n-1)S_n^{*2}}{\chi^2_{1-\alpha}(n-1)}\right),$$

$\underline{\dfrac{(n-1)S_n^{*2}}{\chi^2_{1-\alpha}(n-1)}}$ $\sigma^2$的置信度为$1-\alpha$的单侧置信上限

**注** 其他结果可以参见p72表2.3.

**例7(p71例2.31)** 设从一批灯泡中, 随机地取10只作寿命试验,测得样本寿命均值(以小时计)为 1500h ,样本的修正均方差为 20h, 设灯泡寿命服从正态分布, 求灯泡寿命平均值的置信度为0.95的单侧置信下限.

**解** $1-\alpha=0.95,\quad n=5,\quad \bar{x}=1500,\quad s_n^*=20,$

$$t_\alpha(n-1)=t_{0.05}(9)=1.833$$

$\mu$的置信度为 0.95的置信下限

$$\underline{\mu}=\bar{x}-\frac{s_n^*}{\sqrt{n}}t_\alpha(n-1)=1488.$$

# 七、非正态总体参数的区间估计

## 1、利用渐近正态性取代精确分布

由于统计量的精确抽样分布很难计算，因而通常可以利用近似分布取代精确分布。

**一般总体均值的置信区间：**

首先回顾定理**1.18**

**定理1.18** 设总体$X$的分布是任意的，其均值为$\mu$，且有有限方差$\sigma^2>0$，$X_1,X_2,\cdots,X_n$来自总体的一个样本,则当$n\to\infty$时，有

$$Y=\frac{\bar{X}-\mu}{\sqrt{S_n^2/n}}\xrightarrow{L}N(0,1)$$

由定理可得：

$$P\{|\frac{\bar{X}-\mu}{S_n/\sqrt{n}}|<u_{\frac{\alpha}{2}}\}\approx 1-\alpha$$

由此可得总体期望μ置信度为1-α置信区间为

$$(\bar{X}-u_{\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}},\bar{X}+u_{\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}})$$

设有一容量$n>50$的大样本，它来自$(0-1)$分布的总体$X$，$X$的分布律为 $f(x;p)=p^x(1-p)^{1-x}$，$x=0,1$，其中$p$为未知参数，则$p$的置信度为$1-\alpha$的置信区间是

$$\left(\frac{m}{n}-u_{\frac{\alpha}{2}}\sqrt{\frac{1}{n}\frac{m}{n}(1-\frac{m}{n})},\ \frac{m}{n}+u_{\frac{\alpha}{2}}\sqrt{\frac{1}{n}\frac{m}{n}(1-\frac{m}{n})}\right)$$

这是因为

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{m}{n},(\text{因为}X_i\text{取值为}0,1)$$

$$S_n^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2=\frac{m}{n}-(\frac{m}{n})^2,(\text{因为}X_i\text{取值为}0,1)$$

将这个结果代入置信区间公式即得参数p的置信区间

**例8(p74例2.33)** 在试验的**1000**个电子元件中，共

**100**个失效，试以**95%**的概率估计整批产品的失效率.

**解** 由题意可知，每个元件服从两点分布**B(1,p)**,

其中，**n=1000,m=100,1-α=0.95,**因而失效率**p**的置信区间为

$$\left(\frac{m}{n}-u_{\frac{\alpha}{2}}\sqrt{\frac{1}{n}\frac{m}{n}(1-\frac{m}{n})},\ \frac{m}{n}+u_{\frac{\alpha}{2}}\sqrt{\frac{1}{n}\frac{m}{n}(1-\frac{m}{n})}\right)$$

$$=(0.0914, 0.1186)$$

**例9 (p75例2.34)**　设总体X的分布密度为

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

$(X_1,X_2,\cdots,X_n)^T$为总体$X$的样本，求参数$\theta$的置信度为$1-\alpha$的置信区间.

**解**　先寻找参数$\theta$的枢轴量

此分布为指数分布，容易证明:

$\theta$是此分布的均值，其无偏估计为$\bar{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$.

由于$X_i \sim \Gamma(1,\frac{1}{\theta})$,因而$n\bar{X}=\sum_{i=1}^{n}X_i \sim \Gamma(n,\frac{1}{\theta})$.

$n\bar{X}=\sum_{i=1}^{n}X_i$此时还不能满足枢轴量的条件。需要再构造。

不难证明:

$$Y=\frac{2n\bar{X}}{\theta}$$(利用随机变量函数的密度函数

计算公式可以得到）服从自由度为2$n$的$\chi^2$分布，即

$$Y \sim \chi^2(2n)$$

$Y=\frac{2n\bar{X}}{\theta}$是我们要找的枢轴量。

再确定$a$，$b$.

对于给定的$1-\alpha$，需要寻找a, b，使得

$$P\{a<\frac{2n\bar{X}}{\theta}<b\}=1-\alpha$$

则有

$$P\{\underbrace{\chi^2_{1-\frac{\alpha}{2}}(2n)}_{a}<\frac{2n\bar{X}}{\theta}<\underbrace{\chi^2_{\frac{\alpha}{2}}(2n)}_{b}\}=1-\alpha$$

最后解不等式.

得$\theta$的置信度为$1-\alpha$的置信区间为

$$\left[\frac{2n\bar{X}}{\chi^2_{\frac{\alpha}{2}}(2n)},\frac{2n\bar{X}}{\chi^2_{1-\frac{\alpha}{2}}(2n)}\right]$$

**例10(p76例2.35)** 设总体X服从(0,θ)上的均匀分布，$(X_1, X_2, \cdots, X_n)^T$ 为总体$X$的样本，求参数$\theta$形如$(aX_{(n)}, bX_{(n)})$的置信度为$1-\alpha$的平均区间长度最短的置信区间.

**解** 先寻找参数$\theta$的枢轴量

$\theta$的最大似然估计是$X_{(n)}$，就从$X_{(n)}$入手寻找枢轴量。

$$X_{(n)} \sim f_{X_{(n)}}(x) = \begin{cases} \dfrac{n}{\theta^n} x^{n-1}, & 0 < x < \theta \\ 0, & \text{其他} \end{cases}$$

不满足枢轴量的条件。需要再构造。

令Y=$\dfrac{X_{(n)}}{\theta}$，算得

$$f_{\frac{X_{(n)}}{\theta}}(y)=f_{X_{(n)}}(\theta y)|(\theta y)'|=\begin{cases} ny^{n-1}, & 0<y<1 \\ 0, & 其他 \end{cases}$$

Y=$\dfrac{X_{(n)}}{\theta}$是枢轴量。

再确定$a$，$b$.　选取$b>a\geq 1$,使得$0<\dfrac{1}{b}<\dfrac{1}{a}<1$,且

$$P\{\frac{1}{b}<\frac{X_{(n)}}{\theta}<\frac{1}{a}\}=P\{aX_{(n)}<\theta<bX_{(n)}\}=1-\alpha$$

上述不等式可得 $aX_{(n)}<\theta<bX_{(n)}$　$\boxed{(aX_{(n)},bX_{(n)})}$

注：这里不能找到惟一的$a, b$!

最后求平均长度最短的置信区间。

设$L=E(bX_{(n)}-aX_{(n)})$,求$a,b$使得$L$最小.

由于 $P\{aX_{(n)}<\theta<bX_{(n)}\}=P\{\frac{1}{b}<\frac{X_{(n)}}{\theta}<\frac{1}{a}\}$

$$=\int_{\frac{1}{b}}^{\frac{1}{a}} ny^{n-1}\mathrm{d}y=\frac{1}{a^n}-\frac{1}{b^n}=1-\alpha$$

又因为$L=(b-a)E(X_{(n)})=(b-a)\dfrac{n\theta}{n+1}$

则 $\begin{cases} \min\limits_{a,b}(b-a) \\ \dfrac{1}{a^n}-\dfrac{1}{b^n}=1-\alpha \end{cases}$ 其解为$a=1,\ b=\dfrac{1}{\sqrt[n]{\alpha}}$.

则参数$\theta$型如$(aX_{(n)},bX_{(n)})$的置信度为$1-\alpha$的平均区间长度最短的置信区间为

$$(aX_{(n)},bX_{(n)})=(X_{(n)},\frac{X_{(n)}}{\sqrt[n]{\alpha}})$$

区间的平均长度为

$$L=(b-a)E(X_{(n)})=(b-a)\frac{n\theta}{n+1}$$

$$=(\frac{1}{\sqrt[n]{\alpha}}-1)\frac{n\theta}{1+n}\xrightarrow{n\to\infty}0$$

由此可见：α越小，精确度越低，

n越大，精确度越大。

**证明其解为$a=1, b=\dfrac{1}{\sqrt[n]{\alpha}}$.**

记$1-\alpha=c$ 则$\dfrac{1}{a^n}-\dfrac{1}{b^n}=c$ 解得$b=\left(\dfrac{1}{\dfrac{1}{a^n}-c}\right)^{\frac{1}{n}}$

记 $f(a)=\left(\dfrac{1}{\dfrac{1}{a^n}-c}\right)^{\frac{1}{n}}-a=\left(\dfrac{a^n}{1-ca^n}\right)^{\frac{1}{n}}-a=a\left(\dfrac{1}{1-ca^n}\right)^{\frac{1}{n}}-a$

$$f(a)=a\left(\frac{1}{1-ca^n}\right)^{\frac{1}{n}}-a \qquad f'(a)=\left(\frac{1}{1-ca^n}\right)^{\frac{1}{n}+1}-1$$

因为$\frac{1}{a^n}-c<\frac{1}{a^n}$ 即$\frac{1}{\frac{1}{a^n}-c}>a^n$ $\Rightarrow\frac{1}{1-ca^n}>1$

$$\Rightarrow\left(\frac{1}{1-ca^n}\right)^{\frac{1}{n}+1}>1 \;\Rightarrow f'(a)>0.$$

$f(a)$在$[1,\infty)$单调递增。

$f(a)$在$a=1$处取得最小值。

带入$\frac{1}{a^n}-\frac{1}{b^n}=1-\alpha$,

得$b=\frac{1}{\sqrt[n]{\alpha}}$。

**例11** 设总体 $X$ 在 $[0,\theta]$ 上服从均匀分布,其中 $\theta$ $(\theta>0)$ 未知,$(X_1,X_2,\cdots,X_n)$ 是来自总体 $X$ 的样本,给定 $\alpha$, 求 $\theta$ 的置信度为 $1-\alpha$ 的置信下限和置信上限.

**解** 令 $X_{(n)}=\max\{X_1,X_2,\cdots,X_n\}$,

对于给定的 $\alpha$, 找 $0<\underline{\theta}\le 1$, 使 $P\left\{\theta>\dfrac{X_{(n)}}{\underline{\theta}}\right\}=1-\alpha$,

$$P\left\{\underline{\theta}>\frac{X_{(n)}}{\theta}\right\}=1-\alpha,$$

即 $1-\alpha=\int_0^{\underline{\theta}} nz^{n-1}\mathrm{d}z=\underline{\theta}^n$, 于是 $\underline{\theta}=\sqrt[n]{1-\alpha}$,

所以 $P\left\{\dfrac{X_{(n)}}{\sqrt[n]{1-\alpha}}<\theta\right\}=1-\alpha$,

$\theta$的置信度为$1-\alpha$的置信下限 $\underline{\theta}=\dfrac{X_{(n)}}{\sqrt[n]{1-\alpha}}$.

对于给定的$\alpha$, 找$0<\overline{\theta}<1$,使 $P\left\{\theta<\dfrac{X_{(n)}}{\overline{\theta}}\right\}=1-\alpha$,

即 $1-\alpha=\int_{\overline{\theta}}^{1} nz^{n-1}\mathrm{d}z=1-\overline{\theta}^{n}$, 于是 $\overline{\theta}=\sqrt[n]{\alpha}$,

所以 $P\left\{\theta<\dfrac{X_{(n)}}{\sqrt[n]{\alpha}}\right\}=1-\alpha$,

$\theta$的置信度为$1-\alpha$的置信上限 $\overline{\theta}=\dfrac{X_{(n)}}{\sqrt[n]{\alpha}}$.

Thank You!