2 参数的区间估计

当样本观测值给定以后，点估计给出未知参数θ一个确定的数值．但估计值只是θ的一个近似值，究竟它与θ的真值有没有误差，误差是多少并不知道．而在实际问题中，这种误差的大小往往是人们比较关心的．为此，引入参数估计的另一种形式--区间估计.

例如：天气预报

明天的最高温度：12°C. ——点估计

明天的最高温度：11°C —13°C. ——区间估计

置信区间定义

设θ是一个待估参数，给定$\alpha > 0$,若由样本$X_1,X_2,\ldots X_n$确定的两个统计量 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$, $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$, $(\hat{\theta}_1 < \hat{\theta}_2)$,满足

$$P(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2) = 1 - \alpha$$

则称区间 $[\hat{\theta}_1, \hat{\theta}_2]$是$\theta$的置信水平(Confidence level,置信度、置信概率)为 $1 - \alpha$的置信区间(Confidence Intervals).

$\hat{\theta}_1$和$\hat{\theta}_2$分别称为置信下限和置信上限. 置信区间是一个随机区间, 它是否包含了未知参数 θ 是一个随机事件.

说明

1) 要求 $[\hat{\theta}_1, \hat{\theta}_2]$ 以较大的概率包含θ，

置信度$P(\hat{\theta}_1 \le \theta \le \hat{\theta}_2) = 1 - \alpha$**要满足要求，即要求估计可靠.**

2) 估计的精度要尽可能的高，即要求区间长度
$\hat{\theta}_2 - \hat{\theta}_1$尽可能短.

置信度与精度是一对矛盾，当样本容量固定时，置信度越高，则精度越差.

处理"可靠性与精度关系"的原则:
求参数置信区间，要先保证可靠性，再提高精度

求置信区间的步骤

例 设$X_1,\dots X_n$是取自$N(\mu,\sigma^2)$的样本，σ^2已知，求参数μ的置信水平为$1-\alpha$的置信区间.

明确问题:求什么参数的置信区间?置信水平是多少?

解 选μ的点估计为$\overline{X}$

寻找未知参数的一个良好估计量.

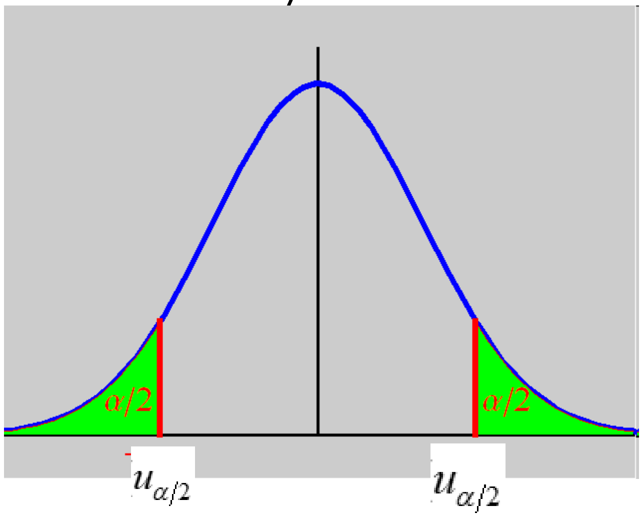
取 $U=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

寻找一个待估参数和估计量的函数U，要求其分布已知
(知道分布，就可以求出U取值于任意区间的概率).

对于给定的置信水平，根据U的分布，确定一个区间，使得U取值位于该区间的概率为置信水平.

对给定的置信水平$1-\alpha$,查正态分布表得到$u_{\alpha/2}$,

使 $P\{\left|\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \le u_{\alpha/2}\} = 1-\alpha$



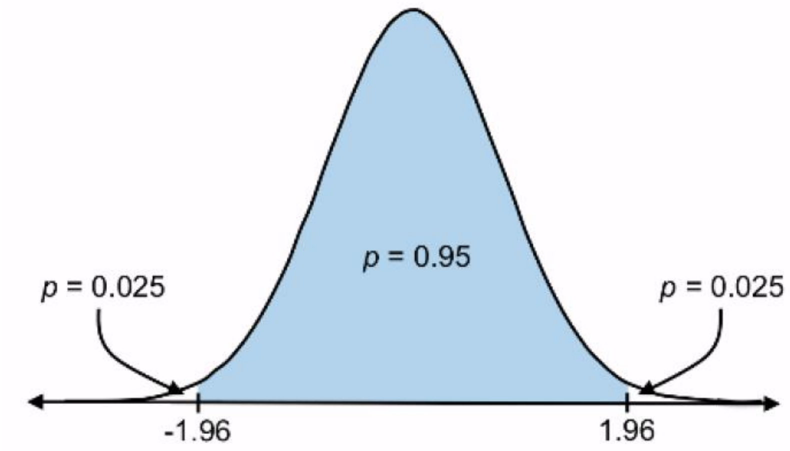
$$\Phi(u_{\alpha/2}) - \Phi(u_{-\alpha/2}) = 1-\alpha$$

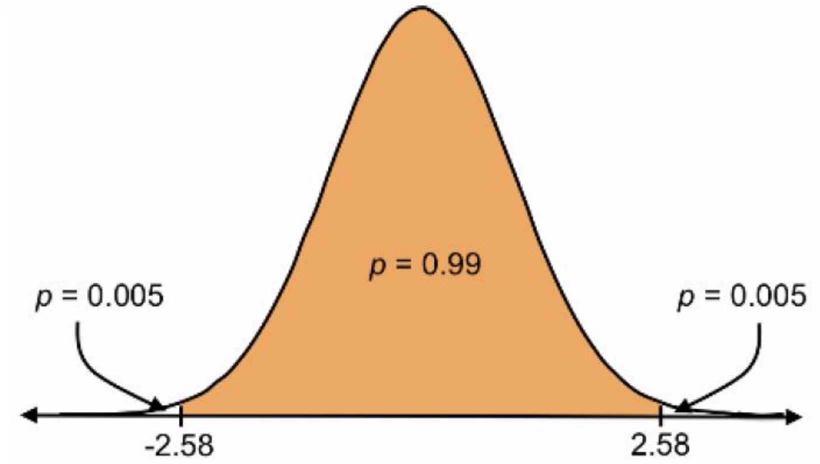
$$2\Phi(u_{\alpha/2}) - 1 = 1-\alpha$$

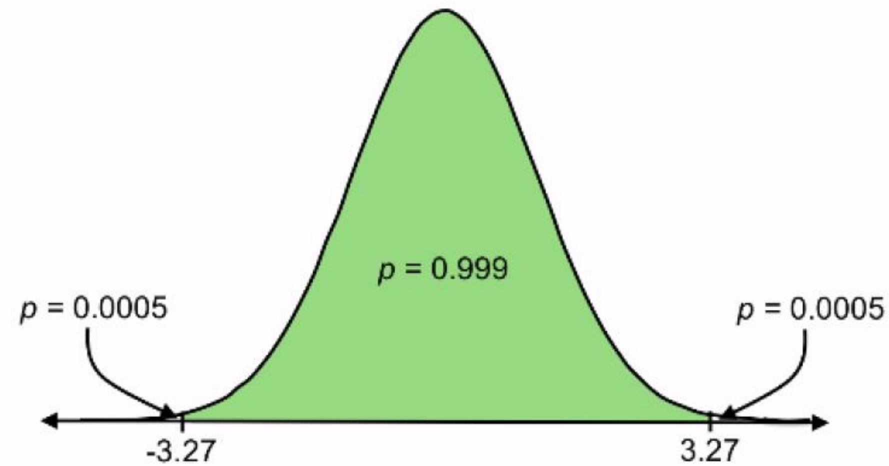
$$\Phi(u_{\alpha/2}) = 1-\frac{\alpha}{2}$$

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

所求μ的置信区间为 $[\bar{X} - \dfrac{\sigma}{\sqrt{n}}u_{\alpha/2},\ \bar{X} + \dfrac{\sigma}{\sqrt{n}}u_{\alpha/2}]$ 简记为$\bar{X} \pm \dfrac{\sigma}{\sqrt{n}}u_{\alpha/2}$



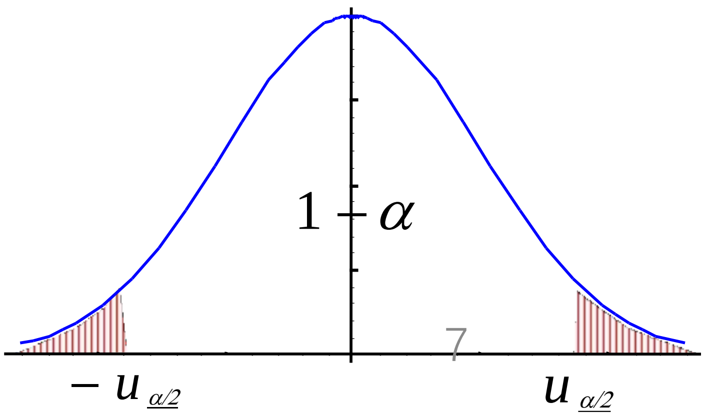




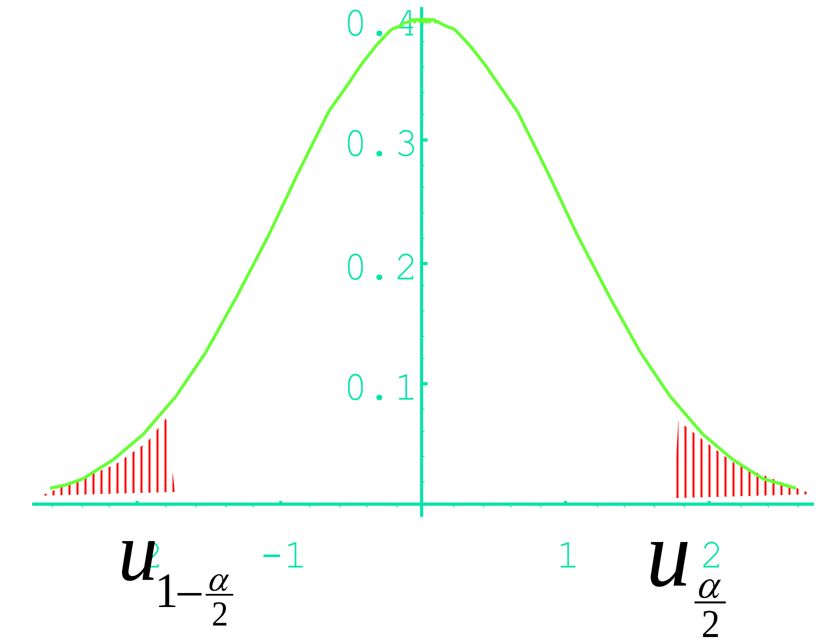
α	1-α	$u_{\alpha/2}$
0.10	0.90	$u_{0.05} = 1.645$
0.05	**0.95**	$\boldsymbol{u_{0.025} = 1.96}$
0.01	0.99	$u_{0.005} = 2.58$
0.001	0.999	$u_{0.0005} = 3.27$

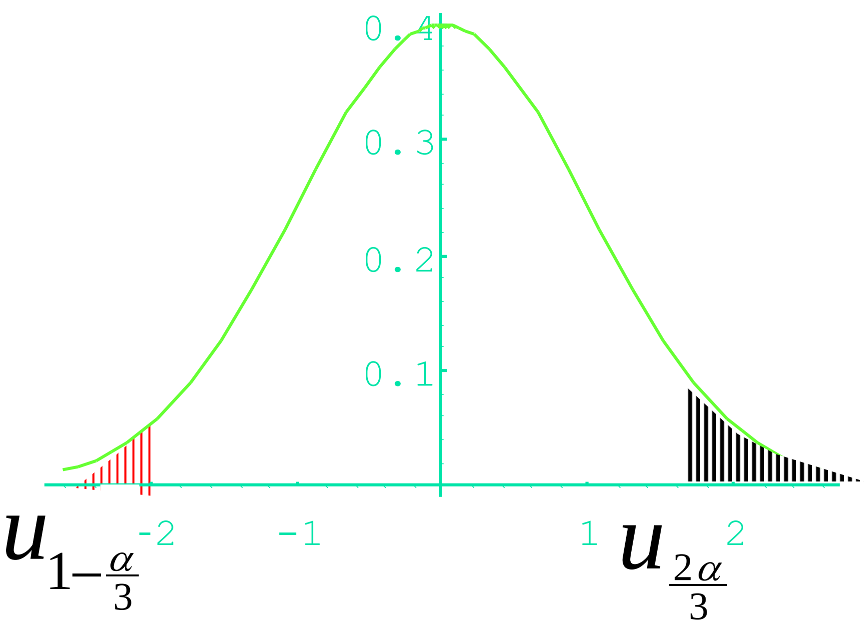
说明

- 给定样本和置信水平，置信区间不是唯一的.对同一个参数，可以构造许多置信区间.为什么要取 $u_{\alpha/2}$?



区间的长度为 $2\frac{\sigma}{\sqrt{n}}u_{\alpha/2}$，达到最短，精度最高。



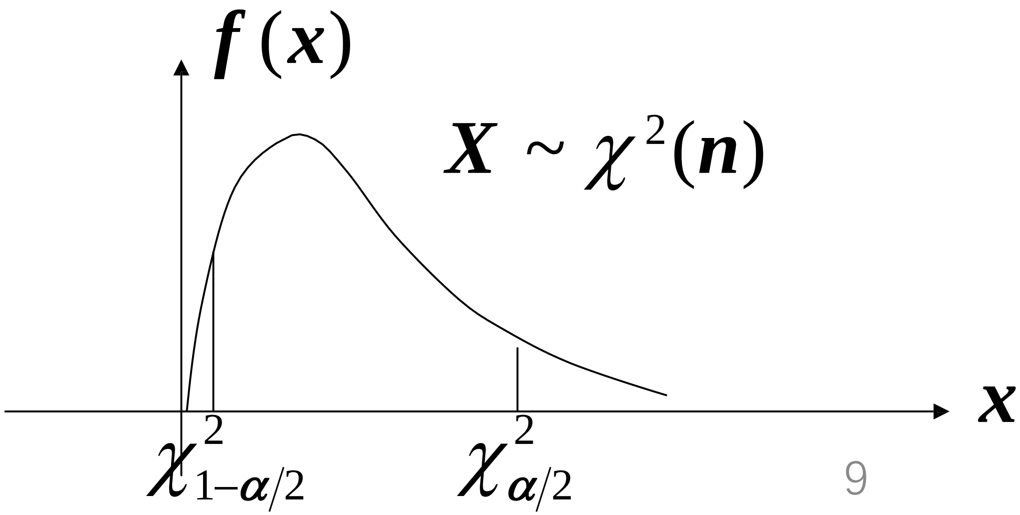


取 $\alpha = 0.05$

$$u_{\frac{\alpha}{2}} - u_{1-\frac{\alpha}{2}} = 1.96-(-1.96)$$
$$= 3.92$$

$$u_{\frac{2\alpha}{3}} - u_{1-\frac{\alpha}{3}} = 1.84-(-2.13)$$
$$= 3.97$$

即使在概率密度不对称的情形，如χ^2分布、F分布，习惯上仍取对称的分位点来计算未知参数的置信区间.



在保证足够可靠的前提下，尽量使区间的长度短一些.

求置信区间的一般步骤:

1). 明确问题, 求什么参数的置信区间?置信水平$1-\alpha$是多少?

2). 寻找参数θ的一个良好的点估计$T(X_1, X_2, \ldots X_n)$

3). 寻找一个待估参数θ和估计量T的函数 $S(T, \theta)$,
且其分布为已知，称$S(T, \theta)$为枢轴量(pivot).

4). 对于给定的置信水平$1-\alpha$，根据$S(T, \theta)$的分布，
确定常数a, b，使得

$$P(a \le S(T, \theta) \le b) = 1-\alpha$$

5). 对 “$a \le S(T, \theta) \le b$” 作等价变形,得到:

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

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

区间估计的关键：要寻找一个待估参数θ和估计量T 的函数$S(T, \theta)$, 且$S(T, \theta)$的分布为已知, 不依赖于任何其它未知参数。(这样才能确定一个概率区间).

这与总体分布有关，总体分布的形式，至关重要.

我们主要讨论总体分布为正态的情形.
若样本容量很大，即使总体分布未知，应用中心极限定理，可得总体的近似分布，也可以求得参数的近似区间估计.

一个正态总体 $X \sim N(\mu,\sigma^2)$参数的置信区间

(1) 方差σ^2已知, μ的置信区间

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

(推导过程见前面)

(2) 方差σ^2未知, μ的置信区间（常用)

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

2. 正态总体均值的置信区间(置信水平为 $1-\alpha$)

待估参数	其他参数	枢轴量的分布	置信区间
μ	σ^2 已知	$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$
μ	σ^2 未知	$t=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$	$\left(\bar{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\bar{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right)$

推导 选取枢轴量 $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$

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

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

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

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

(3) 当 μ 已知时，方差 σ^2 的置信区间

σ^2 的极大似然估计量为 $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$,

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

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

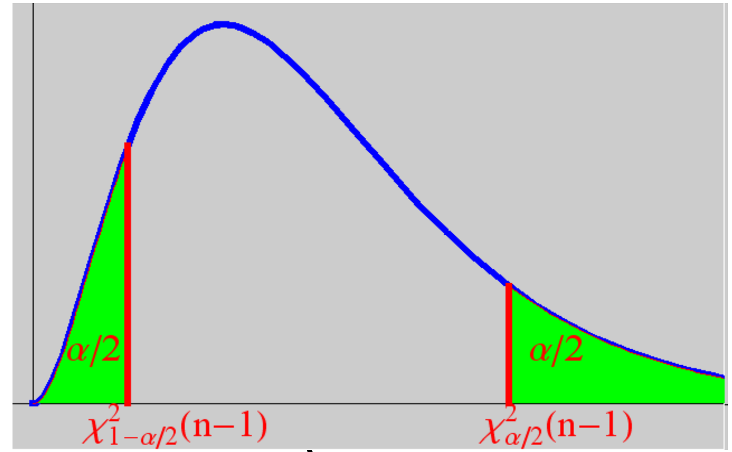
得 σ^2 的置信度为 $1-\alpha$ 置信区间为

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

(4) 当μ未知时，方差σ^2的置信区间（常用）

方差 σ^2 的一个良好点估计为 S^2，

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



对于给定的 α，

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

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

于是方差 σ^2 的置信水平为 $1-\alpha$ 的置信区间为

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

例 某工厂生产一批滚珠，其直径 X 服从正态分布 $N(\mu,\sigma^2)$，现从某天的产品中随机抽取6件，测得直径为:

$$15.1\,,\ 14.8\,,\ 15.2\,,\ 14.9\,,\ 14.6\,,\ 15.1$$

(1) 若 $\sigma^2=0.06$，求 μ 的置信区间
(2) 若 σ^2 未知，求 μ 的置信区间
(3) 求方差 σ^2 的置信区间.

} 置信度均为0.95

解： (1) $\left[\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2},\ \overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right]$ $\qquad u_{\frac{\alpha}{2}}=u_{0.025}=1.96$

由给的数据算得 $\quad \bar{x}=\frac{1}{6}\sum_{i=1}^{6}x_i=14.95$

μ 的置信区间为 $(14.95-1.96\times0.1, 14.95+1.96\times0.1)$
$=(14.75, 15.15)$

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

查表 $t_{0.025}(5)=2.5706$ $\bar{x}=14.95$

$$s^2=\frac{1}{5}(\sum_{i=1}^{6}x_i^2-6\bar{x}^2)=0.051.\quad s=0.226$$

μ 的置信区间为

$$(\bar{x}-\frac{s}{\sqrt{6}}t_{0.025}(5),\ \bar{x}+\frac{s}{\sqrt{6}}t_{0.025}(5)\ )=(14.71, 15.187\ )$$

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

$s^2=0.051.$ 查表得 $\chi^2_{0.025}(5)=12.833$， $\chi^2_{0975}(5)=0.831$

σ^2 的置信区间为 $(\ \frac{5s^2}{\chi^2_{0.025}(5)},\ \ \frac{5s^2}{\chi^2_{0.975}(5)}\ )=(0.0199,\ \ 0.3069\ )$

填空题 2分

设置

已知一批零件的长度 X(单位: cm) 服从正态分布 $N(\mu,1)$, 从中随机地抽取 16 个零件, 得到长度的平均值为 40(cm), 则 μ 的置信度为 0.95 的置信区间是 ([填空1], [填空2]) [40-1.96/4,40+1.96/4]=[39.51,40.49]

(注: 标准正态分布函数值 $\Phi(1.96)=0.975, \Phi(1.645)=0.95$, 保留2位小数)

由题设，1-a=0.95，，可见a=0.05，于是查标准正态分布表知 $u_{\frac{\alpha}{2}}=1.96$。

本题n=16，$\bar{x}=40$，因此，根据

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

正常使用填空题需3.0以上版本雨课堂

作答

填空题 2分

设置

从一批钉子中抽取16枚，测得长度（单位：厘米），$\bar{X} = 2.125, S = 0.017$，设钉长分布为正态，总体期望的置信度为0.90的置信区间是([填空1]，[填空2])。

[2.118, 2.132]

(精确到三位小数)

标准正态分布数值表：$\Phi(1.96) = 0.975, \Phi(1.645) = 0.95$ t分布数值表：$t_{0.025}(15) = 2.1315$，$t_{0.05}(15) = 1.7531$

χ^2 分布数值表：$\chi^2_{0.025}(14) = 26.119, \chi^2_{0.975}(14) = 5.629.$

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

（2）因为σ未知，故 $\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t(n-1)$，

即：$\dfrac{2.125-\mu}{\frac{0.017}{\sqrt{16}}} = \dfrac{4(2.125-\mu)}{0.017} \sim t(15)$.

由已知条件，$t_{0.05}(15) = 1.7531$，

作答

两个正态总体参数的置信区间

$(X_1, X_2, \cdots, X_m)$ 为取自总体 $N(\mu_1, {\sigma_1}^2)$的样本,

$(Y_1, Y_2, \cdots, Y_n)$ 为取自总体 $N(\mu_2, {\sigma_2}^2)$的样本,

$\bar{X}, S_1^2$； $\bar{Y}, S_2^2$分别表示两样本的均值与方差,

置信度为 $1-\alpha$

(1) σ_1^2, σ_2^2 已知，$\mu_1 - \mu_2$ 的置信区间

$$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{m}), \quad \bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n})$$

$\bar{X}, \bar{Y}$ 相互独立，

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$$

$\mu_1 - \mu_2$ 的置信区间为

$$\left((\bar{X} - \bar{Y}) - u_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}, \quad (\bar{X} - \bar{Y}) + u_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \right)$$

(2) $\sigma_1^2 = \sigma_2^2 = \sigma^2$未知， $\mu_1 - \mu_2$ 的置信区间

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2)$$

$$\mathrm{S_w} = \sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}}$$

$$P\left(\left|\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\right| < t_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$\mu_1-\mu_2$的置信区间为 $\left((\bar{X}-\bar{Y}) \pm t_{\frac{\alpha}{2}} S_w\sqrt{\frac{1}{m}+\frac{1}{n}}\right)$

$\boldsymbol{\sigma_1^2}$、$\boldsymbol{\sigma_2^2}$ 均未知，且$\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$时，求 $\boldsymbol{\mu_1 - \mu_2}$ 的置信区间.

$\mu_1 - \mu_2$ 的极大似然估计是 $\bar{X} - \bar{Y}$, 用样本方差 S_1^2、S_2^2 分别代替 σ_1^2、σ_2^2, 令

$$J = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}},$$

可以证明, 当 m、n 都很大时, 近似地有 $J \sim N(0;1)$.

$\mu_1 - \mu_2$ 的双侧 (近似) $1 - \alpha$ 置信区间为

$$\left[(\bar{X} - \bar{Y}) - u_{\alpha/2}\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}, (\bar{X} - \bar{Y}) + u_{\alpha/2}\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}\right].$$

例 从某地区随机地选取男、女各 100 名测量身高. 测得男子高度的平均值为 171 厘米, 标准差为 3.5 厘米; 女子高度的平均值为 161 厘米, 标准差为 3.8 厘米. 假定身高服从正态分布, 且男女身高相互独立, 求该地区男女平均身高之差的双侧 95% 置信区间.

解 因为 $m = n = 100, \alpha = 0.05, \bar{x} = 171,\ s_1^2 = (3.5)^2 = 12.25, \bar{y} = 161, s_2^2 = (3.8)^2 = 14.44$, 查标准正态分布表得 $u_{0.975} = 1.96$, 得到该地区男女平均身高之差的双侧 95% 置信区间为

$$\left[(171-161) - 1.96 \times \sqrt{\frac{12.25}{100} + \frac{14.44}{100}}, (171-161) + 1.96 \times \sqrt{\frac{12.25}{100} + \frac{14.44}{100}}\right]$$

$= [8.99, 11.01]$.

(3) 方差比 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信区间 (μ_1，μ_2 已知)

取枢轴量

$$F=\frac{\dfrac{1}{m}\dfrac{\sum_{i=1}^{m}(X_i-\mu_1)^2}{\sigma_1^2}}{\dfrac{1}{n}\dfrac{\sum_{j=1}^{n}(Y_j-\mu_2)^2}{\sigma_2^2}}=\frac{A}{\dfrac{\sigma_1^2}{\sigma_2^2}}\sim F(m,n)$$

$$A=\frac{\frac{1}{m}\sum_{i=1}^{m}(X_i-\mu_1)^2}{\frac{1}{n}\sum_{j=1}^{n}(Y_j-\mu_2)^2}$$

方差比 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信区间为 $\left(\dfrac{A}{F_{\frac{\alpha}{2}}(m,n)},\ \dfrac{A}{F_{1-\frac{\alpha}{2}}(m,n)}\right)$

(4) 方差比$\dfrac{\sigma_1^2}{\sigma_2^2}$的置信区间 (μ_1, μ_2 未知)

取枢轴量

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

方差比 $\dfrac{\sigma_1^2}{\sigma_2^2}$的置信区间为

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\frac{\alpha}{2}}(m-1,n-1)},\quad \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\frac{\alpha}{2}}(m-1,n-1)}\right)$$

例 某厂利用两条自动化流水线罐装番茄酱. 现分别从两条流水线上抽取了容量分别为13与17的两个相互独立的样本，已知 $\bar{X}=10.6\mathrm{g}$，$\bar{Y}=9.5\mathrm{g}$，$S_1^2=2.4g^2$，$S_2^2=4.7g^2$

假设两条流水线上罐装的番茄酱的重量都服从正态分布，其均值分别为 μ_1 与 μ_2

(1) 若它们的方差相同，$\sigma_1^2=\sigma_2^2=\sigma^2$，求均值差 $\mu_1-\mu_2$ 的置信度为0.95 的置信区间;

(2) 求它们的方差比的置信度为 0.95 的置信区间

解　(1)　查表得 $t_{0.025}(28) = 2.0484$

由公式 $\mu_1 - \mu_2$ 的置信区间为

$$\left(\bar{X} - \bar{Y}\right) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} = (-0.3545, 2.545)$$

(2)　查表得 $F_{0.025}(16, 12) = 3.16$, $F_{0.975}(16, 12) = \dfrac{1}{F_{0.025}(12, 16)} \approx \dfrac{1}{2.89}$

由公式得方差比 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信区间为

$$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{0.025}(n_1 - 1, n_2 - 1)}, \frac{S_1^2}{S_2^2} F_{0.025}(n_2 - 1, n_1 - 1)\right) = (0.1767, \ 1.6136)$$

非正态总体的区间估计

设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的样本，要求未知参数 $\boldsymbol{\mu = E(X)}$ **的双侧 $\boldsymbol{1-\alpha}$ 置信区间.**

尽管不知道总体 X 究竟服从什么分布，由于当n充分大(样本容量 $n > 30$)时，$J = \sqrt{n}\frac{\bar{X}-\mu}{S}$ 近似地服从标准正态分布 $N(0;1)$，

在大样本时，μ 的双侧 $1-\alpha$ 置信区间近似地为

$$\left[\bar{X} - \mu_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + u_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

例 从一大批产品中随机地抽取 100 件进行检查, 发现有 4 件是不合格品, 求不合格率 p 的单侧 95% 置信区间上限的近似值.

解 由于总体 X 服从 0-1 分布 $B(1;p)$, 其中 $0 \leqslant p \leqslant 1$, 但 p 末知, 所以 p 的极大似然估计量是 $\bar{X}$.

当 n 充分大时, $\sqrt{n}\frac{\bar{X}-p}{\sqrt{\bar{X}(1-\bar{X})}}$ 近似地服从标准正态分布 $N(0;1)$,

p 的单侧 95% 置信区间上限的近似值为 $\bar{x}+u_{0.05}\sqrt{\frac{1}{n}\bar{x}(1-\bar{x})}$.

现 $n=100, \bar{x}=0.04$, 上限的近似值为

$$0.04+1.645\times\sqrt{\frac{1}{100}\times 0.04\times 0.96}=0.072.$$

讨论两个正态总体均值之间的差异时, 我们假定要么 σ_1^2 与 σ_2^2 均已知, 要么 $\sigma_1^2=\sigma_2^2=\sigma^2$ 但 σ^2 未知.

若取消两个总体方差齐性的假定, 即 $\sigma_1^2\neq\sigma_2^2$, 且 σ_1^2、σ_2^2 均未知时, 如何求 **$\mu_1-\mu_2$ 的置信区间**呢?

这时, $\mu_1-\mu_2$ 的极大似然估计仍是 $\bar{X}-\bar{Y}$, 而用样本方差 S_1^2、S_2^2 分别代替 σ_1^2、σ_2^2, 令

$$J=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}},$$

可以证明, 当 m、n 都很大时, 近似地有 $J\sim N(0;1)$. 从而 $\mu_1-\mu_2$ 的双侧 (近似) $1-\alpha$ 置信区间为

$$\left[(\bar{X}-\bar{Y})-u_{\alpha/2}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}},(\bar{X}-\bar{Y})+u_{\alpha/2}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}\right].$$

例 从某地区随机地选取男、女各 100 名测量身高. 测得男子高度的平均值为 171 厘米, 标准差为 3.5 厘米; 女子高度的平均值为 161 厘米, 标准差为 3.8 厘米. 假定身高服从正态分布, 且男女身高相互独立, 求该地区男女平均身高之差的双侧 95% 置信区间.

解 因为 $m=n=100, \alpha=0.05, \bar{x}=171, s_1^2=(3.5)^2=12.25, \bar{y}=161, s_2^2=(3.8)^2=14.44$, 查标准正态分布表得 $u_{0.025}=1.96$, 得到该地区男女平均身高之差的双侧 95% 置信区间为

$$\left[(171-161)-1.96\times\sqrt{\frac{12.25}{100}+\frac{14.44}{100}}, (171-161)+1.96\times\sqrt{\frac{12.25}{100}+\frac{14.44}{100}}\right]$$

$$=[8.99,11.01].$$