



非正态分布或总体分布未知的参数区间估计 (大样本)

总体均值的区间估计 (大样本)

1. 假定条件

- 总体服从正态分布, 且方差(σ^2) 已知
- 如果不是正态分布, 可由正态分布来近似 ($n \geq 30$)

2. 使用正态分布统计量 z

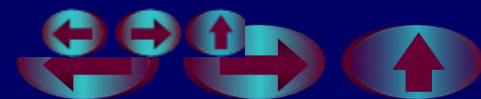
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

定理1 (独立同分布下的中心极限定理)

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 服从同一分布, 且具有数学期望和方差: $E(X_k) = \mu, D(X_k) = \sigma^2$ ($k = 1, 2, \dots$), 则随机变量之和 $\sum_{k=1}^n X_k$ 的标准化变量

$Y_n = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma}$ 的分布函数 $F_n(x)$ 对于任意 x 满足

$$\begin{aligned} \lim_{n \rightarrow \infty} F_n(x) &= \lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x \right\} \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x) \end{aligned}$$



总体均值的区间估计 (大样本)

1. 假定条件

- 总体服从正态分布, 且方差(σ^2) 已知
- 如果不是正态分布, 可由正态分布来近似 ($n \geq 30$)

2. 使用正态分布统计量 z

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

3. 总体均值 μ 在 $1-\alpha$ 置信水平下的置信区间为

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{或} \quad \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} (\sigma \text{ 未知})$$



【例】一家食品生产企业以生产袋装食品为主，为对产量质量进行监测，企业质检部门经常要进行抽检，以分析每袋重量是否符合要求。现从某天生产的一批食品中随机抽取了25袋，测得每袋重量如下表所示。已知产品重量的分布服从正态分布，且总体标准差为10g。试估计该批产品平均重量的置信区间，置信水平为95%

25袋食品的重量

112.5	101.0	103.0	102.0	100.5
102.6	107.5	95.0	108.8	115.6
100.0	123.5	102.0	101.6	102.2
116.6	95.4	97.8	108.6	105.0
136.8	102.8	101.5	98.4	93.3



解：已知 $X \sim N(\mu, 10^2)$, $n=25$, $1-\alpha = 95\%$, $z_{\alpha/2}=1.96$ 。根据样本数据计算得： $\bar{x} = 105.36$

总体均值 μ 在 $1-\alpha$ 置信水平下的置信区间为

$$\begin{aligned}\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} &= 105.36 \pm 1.96 \times \frac{10}{\sqrt{25}} \\ &= 105.36 \pm 3.92 \\ &= (101.44, 109.28)\end{aligned}$$

该食品平均重量的置信区间为 101.44g~109.28g



σ 未知，用样本标准差 S 近似代替.

$$U = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad \text{近似 } N(0,1) \text{ 分布}$$

使
$$P\left\{\left|\frac{\bar{X} - \mu}{S/\sqrt{n}}\right| \leq Z_{\alpha/2}\right\} = 1 - \alpha$$

得均值 μ 的置信水平为 $1 - \alpha$ 的区间估计为

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



总体分布未知!!!

【例】

某单位要估计平均每天职工的总医疗费，观察了30天，其总金额的平均值是170元，标准差为30元，试决定职工每天总医疗费用平均值的区间估计（置信水平为0.95）。

解 设每天职工的总医疗费为 X ，则有

$$E(X)=\mu, D(X)=\sigma^2$$

总体方差未知



均值 μ 的置信水平为 $1-\alpha$ 的区间估计为

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

将 $\bar{X}=170, S=30, u_{\alpha/2}=1.96, n=30$ 代入得,

μ 的置信水平为 0.95 的置信区间是

$$[159.27, 180.74]$$



【例】一家保险公司收集到由36投保个人组成的随机样本，得到每个投保人的年龄(周岁)数据如下表。试建立投保人年龄90%的置信区间

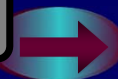
36个投保人年龄的数据

23	35	39	27	36	44
36	42	46	43	31	33
42	53	45	54	47	24
34	28	39	36	44	40
39	49	38	34	48	50
34	39	45	48	45	32

7-32-88



作答



解：已知 $n=36$, $1-\alpha = 90\%$, $z_{\alpha/2}=1.645$ 。根据样本数据计算得： $\bar{x} = 39.5$, $s = 7.77$

总体均值 μ 在 $1-\alpha$ 置信水平下的置信区间为

$$\begin{aligned}\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} &= 39.5 \pm 1.645 \times \frac{7.77}{\sqrt{36}} \\ &= 39.5 \pm 2.13 \\ &= (37.37, 41.63)\end{aligned}$$

投保人平均年龄的置信区间为37.37岁~41.63岁

