区间估计

定义

设 θ 是待估计参数，α 是一个给定的数 $(0<\alpha<1)$，若能构造两个统计量 $\hat{\theta}_1(X_1,X_2,\cdots,X_n)$，$\hat{\theta}_2(X_1,X_2,\cdots,X_n)$ 使得 $P(\hat{\theta}_1<\theta<\hat{\theta}_2)=1=\alpha$ 则称概率值 $1-\alpha$ 为置信水平或者置信度，由两个统计量构成的区间 $(\hat{\theta}_1,\hat{\theta}_2)$ 是一个随机区间，称该随机区间为参数 θ 的置信度为 $1-\alpha$ 的置信区间，并称 $\hat{\theta}_1$，$\hat{\theta}_2$ 为置信上界和置信下界。

NOTE:
1. α 确定后，置信区间的选取方式不唯一，常选最小的一个，以提高估计精度。
2. α 反映了估计的可靠度（α 越小越可靠），$\hat{\theta}_2-\hat{\theta}_1$ 反映了估计精度。当 α 减小时，估计的可靠度增大，但此时 $\hat{\theta}_2-\hat{\theta}_1$ 往往增大，即精度降低。

求解方式

1. 寻找一个样本的函数 $g(X_1,X_2,\cdots,X_n,\theta)$ 称作枢轴量。这个函数含有待估计参数，不含有其他参数，且分布不依赖于待估计参数。

2. 对于给定的置信度 $1-\alpha$，确定出常数 a 和 b，使得

$$P(a<g(X_2,X_2,\cdots,X_n,\theta)<b)=1-\alpha$$

3. 由不等式

$$P(a<g(X_2,X_2,\cdots,X_n,\theta)<b)=1-\alpha$$

   解出置信上下界：$\hat{\theta}_1(X_1,X_2,\cdots,X_n)$，$\hat{\theta}_2(X_1,X_2,\cdots,X_n)$

求置信区间常用公式

记 u_α， $t_\alpha(n)$， $\chi^2_\alpha(n)$， $F_\alpha(n,m)$ 分别为标准正态分布，$t(n)$ 分布，$\chi^2(n)$ 分布，$F(n,m)$ 分布的 α 下侧分位数。

一个正态总体的情形

方差 σ^2 已知，求 μ 的置信区间

由 $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$，选取枢轴量为

$$g(X_1,X_2,\cdots,X_n,\mu)=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

则由：

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

确定出 μ 的置信度为 $1-\alpha$ 的置信区间为

$$(\bar{X}-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \ \bar{X}+u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})$$

方差 σ^2 未知，求 μ 的置信区间

选取枢轴量

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

，得到置信区间为

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

μ 已知，求方差 σ^2 的置信区间

选取枢轴量为

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

，得到置信区间为

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

μ 未知，求方差 σ^2 的置信区间

选取枢轴量为

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

，得到置信区间为

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

两个（相互独立）正态总体的情形

σ_1^2，σ_2^2 已知，求$\mu_1-\mu_2$ 的置信区间

由于 $\bar{X}\sim N(\mu_1,\frac{\sigma_1^2}{n}),\ \ \bar{Y}\sim N(\mu_2,\frac{\sigma_2^2}{m})$，选取枢轴量为

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

可得置信区间为

$$(\ (\bar{X}-\bar{Y})-u_{1-\frac{\mu}{2}}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}},\ \ (\bar{X}-\bar{Y})+u_{1-\frac{\mu}{2}}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\ )$$

σ_1^2，σ_2^2 未知，但（$\sigma_1^2=\sigma_2^2=\sigma^2$），求$\mu_1-\mu_2$ 的置信区间

选取枢轴量为

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}}\sim t(n+m-2)$$

可得置信区间为

$$(\ (\bar{X}-\bar{Y})\pm t_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}\ )$$

σ_1^2，σ_2^2 未知，但（$n,m>50$），求$\mu_1-\mu_2$ 的置信区间

此时

$$\frac{S_1^2}{n}+\frac{S_2^2}{m}\approx\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}$$

选取枢轴量为

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}}\sim N(0,1)$$

可得置信区间为

$$(\ (\bar{X}-\bar{Y})-u_{1-\frac{\mu}{2}}\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}},\ \ (\bar{X}-\bar{Y})+u_{1-\frac{\mu}{2}}\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}\ )$$

σ_1^2，σ_2^2 未知，但（$n=m$），求$\mu_1-\mu_2$ 的置信区间

令 $Z=X-Y$ 则

$$\begin{aligned}\bar{Z}&=\bar{X}-\bar{Y},\\ S_Z^2&=\frac{1}{n-1}\sum_{i=1}^{n}[(X_i-Y_i)-(\bar{X}-\bar{Y})]^2\end{aligned}$$

选取枢轴量为

$$T=\frac{\bar{Z}-\mu}{S_z/\sqrt{n}}\sim t(n-1)$$

可得置信区间为

$$(\ (\bar{X}-\bar{Y})\pm t_{1-\frac{\alpha}{2}}(n-1)\frac{S_z}{\sqrt{n}}\ )$$

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

选取枢轴量为

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

可得置信区间为

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

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

选取枢轴量为

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

可得置信区间为

$$(\frac{\frac{m}{n}\frac{\sum_{i=1}^{n}(X_i-\mu_1)^2}{\sum_{i=1}^{n}(Y_i-\mu_2)^2}}{F_{1-\frac{\alpha}{2}}(n,m)},\ \frac{\frac{m}{n}\frac{\sum_{i=1}^{n}(X_i-\mu_1)^2}{\sum_{i=1}^{n}(Y_i-\mu_2)^2}}{F_{\frac{\alpha}{2}}(n,m)})$$

单侧区间估计

给定 $\alpha\ (0<\alpha<1)$，θ 是待估参数，$(X_1,X_2,\cdots,X_n)$ 是总体 X 的样本，若能确定一个统计量：

$$\underline{\theta}=\underline{\theta}(X_1,X_2,\cdots,X_n)$$

或

$$\overline{\theta}=\overline{\theta}(X_1,X_2,\cdots,X_n)$$

使得 $P(\theta>\underline{\theta})=1-\alpha$ 或 $P(\theta<\overline{\theta})=1-\alpha$ 则称 $(\underline{\theta},\ +\infty)$（或 $(-\infty,\ \overline{\theta})$）为待估参数 θ 的置信度为 $1-\alpha$ 的单侧置信区间。并称 $\underline{\theta}$ 为单侧置信下界（$\overline{\theta}$ 为单侧置信上界）。