Lec-23. 两个正态总体的区间估计、0-1 分布参数的区间估计、正态总体的单侧置信区间

主讲教师: 吴利苏 (**wulisu@sdust.edu.cn**)
主　　页: wulisu.cn

本次课内容

两个正态总体参数的区间估计

- σ_1^2, σ_2^2 已知, $\mu_1 - \mu_2$ 的置信区间
- $\sigma_1^2 = \sigma_1^2 = \sigma^2$ 未知, $\mu_1 - \mu_2$ 的置信区间
- μ_1, μ_2 未知, σ_1^2/σ_2^2 的置信区间

0-1 分布参数的区间估计

正态总体均值与方差的单侧置信区间

两个正态总体均值差 $\mu_1 - \mu_2$ 的置信区间

设总体 $X \sim N\left(\mu_1, \sigma_1^2\right)$, $Y \sim N\left(\mu_2, \sigma_2^2\right)$, σ_1^2, σ_2^2 已知. $X_1, \cdots, X_{n_1}$ 和 $Y_1, \cdots, Y_{n_2}$ 分别为来自总体 X, Y 的样本, 这两个样本相互独立. $\bar{X}, \bar{Y}$ 分别为 X, Y 的样本均值, S_1^2, S_2^2 分别为 X, Y 的样本方差. 取枢轴量

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

- 则 $\mu_1 - \mu_2$ 的一个置信水平为 $1 - \alpha$ 的区间估计为

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

$\bar{X}$ 和 $\bar{Y}$ 分别为 μ_1, μ_2 的无偏估计，故 $\bar{X} - \bar{Y}$ 是 $\mu_1 - \mu_2$ 的无偏估计. $\bar{X}, \bar{Y}$ 相互独立，$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$，$\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$.

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

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

$$P\{-z_{\alpha/2} < G < z_{\alpha/2}\} = 1 - \alpha$$

解得置信区间为 $\left(\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$.

两个正态总体均值差 $\mu_1-\mu_2$ 的置信区间

设总体 $X\sim N\left(\mu_1,\sigma_1^2\right)$, $Y\sim N\left(\mu_2,\sigma_2^2\right)$, $\sigma_1^2=\sigma_1^2=\sigma^2$ 未知.

取枢轴量

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2),$$

- 则 $\mu_1-\mu_2$ 的一个置信水平为 $1-\alpha$ 的区间估计为

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

其中 $S_W=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$, $S_W=\sqrt{S_W^2}$.

例

两个正态总体中，方差未知且相等. 设样本独立且
$n_1 = 10$, $\overline{x}_1 = 500$, $s_1 = 1.10$;
$n_2 = 20$, $\overline{x}_2 = 496$, $s_2 = 1.20$.
求 $\mu_1 - \mu_2$ 的一个置信水平为 0.95 的置信区间.

例

两个正态总体中，方差未知且相等. 设样本独立且

$n_1 = 10,\ \bar{x}_1 = 500,\ s_1 = 1.10;$

$n_2 = 20,\ \bar{x}_2 = 496,\ s_2 = 1.20.$

求 $\mu_1 - \mu_2$ 的一个置信水平为 0.95 的置信区间.

得到的置信区间的下限大于零，则推断 $\mu_1 > \mu_2$.

例

两个正态总体中，方差未知且相等．设样本独立且

$n_1 = 8,\ \bar{x}_1 = 91.73,\ s_1^2 = 3.89;$

$n_2 = 8,\ \bar{x}_2 = 93.75,\ s_2^2 = 4.02.$

求 $\mu_1 - \mu_2$ 的一个置信水平为 0.95 的置信区间．

例

两个正态总体中，方差未知且相等. 设样本独立且
$n_1 = 8,\ \bar{x}_1 = 91.73,\ s_1^2 = 3.89;$
$n_2 = 8,\ \bar{x}_2 = 93.75,\ s_2^2 = 4.02.$
求 $\mu_1 - \mu_2$ 的一个置信水平为 0.95 的置信区间.

得到的置信区间的包含零，则推断 μ_1 和 μ_2 没有显著差别.

设两个正态总体均值差 $\mu_1 - \mu_2$ 的置信区间为 $(\underline{\theta}, \overline{\theta})$,

- 若 $0 \in (\underline{\theta}, \overline{\theta})$, 则推断 $\mu_1 = \mu_2$;

$$\underline{\theta} < 0 < \overline{\theta} \xrightarrow{\text{推断}} \mu_1 = \mu_2$$

- 若 $(\underline{\theta}, \overline{\theta})$ 在 0 的右侧, 则推断 $\mu_1 > \mu_2$;

$$0 < \underline{\theta} \xrightarrow{\text{推断}} \mu_1 > \mu_2$$

- 若 $(\underline{\theta}, \overline{\theta})$ 在 0 的左侧, 则推断 $\mu_1 < \mu_2$.

$$\overline{\theta} < 0 \xrightarrow{\text{推断}} \mu_1 < \mu_2.$$

两个正态总体方差 σ_1^2/σ_2^2 的置信区间

设总体 $X \sim N\left(\mu_1, \sigma_1^2\right)$, $Y \sim N\left(\mu_2, \sigma_2^2\right)$, μ_1, μ_2 未知. 取枢轴量

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

- 则 σ_1^2/σ_2^2 的一个置信水平为 $1-\alpha$ 的区间估计为

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

例

两个正态总体中，均值方差均未知. 设样本独立且
$n_1 = 18,\ s_1^2 = 0.34;\ n_2 = 13,\ s_2^2 = 0.29.$
求 σ_1^2/σ_2^2 的一个置信水平为 0.90 的置信区间.

例

两个正态总体中，均值方差均未知. 设样本独立且 $n_1 = 18,\ s_1^2 = 0.34;\ n_2 = 13,\ s_2^2 = 0.29$.
求 σ_1^2/σ_2^2 的一个置信水平为 0.90 的置信区间.

得到的置信区间的包含 1, 则推断 σ_1^2 和 σ_2^2 没有显著差别.

设两个正态总体方差商 σ_1^2/σ_2^2 的置信区间为 $(\underline{\theta}, \overline{\theta})$,

- 若 $1 \in (\underline{\theta}, \overline{\theta})$, 则推断 $\underline{\theta} = \overline{\theta}$;

$$\underline{\theta} < 1 < \overline{\theta} \xrightarrow{\text{推断}} \sigma_1^2 = \sigma_2^2$$

- 若 $(\underline{\theta}, \overline{\theta})$ 在 1 的右侧, 则推断 $\sigma_1^2 > \sigma_2^2$;

$$\underline{\theta} > 1 \xrightarrow{\text{推断}} \sigma_1^2 > \sigma_2^2$$

- 若 $(\underline{\theta}, \overline{\theta})$ 在 1 的左侧, 则推断 $\sigma_1^2 < \sigma_2^2$.

$$0 < \overline{\theta} < 1 \xrightarrow{\text{推断}} \sigma_1^2 < \sigma_2^2$$

单个正态总体参数的区间估计

待　估参数	其他参数	枢轴量	置信区间
μ	σ^2 已知	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\left(\overline{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$
μ	σ^2 未知	$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$\left(\overline{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\right)$
σ^2	μ 未知	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$

两个正态总体参数的区间估计

待估参数	其他参数	枢轴量	置信区间
$\mu_1-\mu_2$	σ_1^2,σ_2^2 已知	$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$\left(\bar{X}-\bar{Y}\pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$
$\mu_1-\mu_2$	$\sigma_1^2=\sigma_2^2=\sigma^2$ 未知	$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$	$\left(\bar{X}-\bar{Y}\pm t_{\alpha/2}(n_1+n_2-2)S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$
σ_1^2/σ_2^2	μ_1,μ_2 未知	$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$	$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right)$

其他总体均值的区间估计

总体 X 的均值为 μ, 方差为 σ^2, 非正态分布或不知分布形式. 样本为 $X_1, \ldots, X_n$. 当 n 充分大 (一般 $n > 50$) 时, 由中心极限定理知,

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

设 $\bar{X}$ 和 S^2 分别为样本均值和样本方差. μ 的置信水平为 $1 - \alpha$ 的置信区间

- σ^2 已知时, 置信区间近似为 $\left(\overline{X} \pm z_{\alpha/2}\sigma/\sqrt{n}\right)$.
- σ^2 未知时, 置信区间近似为 $\left(\overline{X} \pm z_{\alpha/2}S/\sqrt{n}\right)$.

例

某市随机抽取 100 个家庭, 调查知道其中有 60 家拥有私家车. 试根据此调查结果, 求该市拥有私家车比例 p 的置信水平为 95% 近似置信区间.

解:

$\hat{p} = \bar{x} = \frac{60}{100} = 0.6, s^2 \approx \hat{p}(1-\hat{p}) = 0.24, z_{0.025} = 1.96.$

代入近似置信区间

$$\left(\bar{X} - z_{0.025} S/\sqrt{n}, \quad \bar{X} + z_{0.025} S/\sqrt{n}\right)$$

得近似置信区间为 $(0.512, 0.688)$. □

0-1 分布参数的区间估计

总体 $X \sim b(1, p)$, $X_1, \cdots, X_n$ $(n > 50)$ 为样本.

- 则未知参数 p 的一个置信水平为 $1-\alpha$ 的置信区间近似为

$$\left(\frac{-b-\sqrt{b^2-4ac}}{2a}, \frac{-b+\sqrt{b^2-4ac}}{2a}\right)$$

其中 $a = n + z_{\alpha/2}^2$, $b = -(2n\bar{X} + z_{\alpha/2}^2)$, $c = n\bar{X}^2$.

总体 $X \sim b(1,p)$ 的分布律为

$$P\{X=x\}=p^x(1-p)^{1-x}, x=0,1,$$

其中 p 未知参数.

$$\mu=p, \sigma^2=p(1-p).$$

由中心极限定理

$$\frac{\sum X_i - E(\sum X_i)}{\sqrt{D(\sum X_i)}}=\frac{n\bar{X}-np}{\sqrt{np(1-p)}} \sim N(0,1).$$

则

$$P\left\{-z_{\alpha/2} < \frac{n\bar{X} - np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right\} \approx 1 - \alpha$$

等价于

$$P\left\{\left|\frac{n\bar{X} - np}{\sqrt{np(1-p)}}\right| < z_{\alpha/2}\right\} \approx 1 - \alpha.$$

故

$$(n + z_{\alpha/2}^2)p^2 - (2n\bar{X} + z_{\alpha/2})p + n\bar{X}^2 < 0.$$

记 $a=n+z_{\alpha/2}^2$, $b=-(2n\bar{X}+z_{\alpha/2}^2)$, $c=n\bar{X}^2$,
则由

$$ap^2+bp+c<0$$

解得

$$\frac{-b-\sqrt{b^2-4ac}}{2a}<p<\frac{-b+\sqrt{b^2-4ac}}{2a}.$$

例

现从一批产品中取 100 个样本, 得一级品 60 个. 求这批产品得一级品率 p 的置信水平为 0.95 的置信区间.

例

现从一批产品中取 100 个样本，得一级品 60 个. 求这批产品得一级品率 p 的置信水平为 0.95 的置信区间.

解: $n = 100$, $\bar{x} = 0.6$, $1 - \alpha = 0.95$, $z_{\alpha/2} = z_{0.025} = 1.96$. $a = n + z_{\alpha/2}^2 = 103.84$, $b = -(2n\bar{x} + z_{\alpha/2}^2) = -123.84$, $c = n\bar{x}^2 = 36$. 则 $p_1 = \frac{-b-\sqrt{b^2-4ac}}{2a} = 0.5$, $p_2 = \frac{-b+\sqrt{b^2-4ac}}{2a} = 0.69$. 故 p 的置信水平为 0.95 的置信区间为 $(0.5, 0.69)$. □

单侧置信区间

定义

若

$$P\{\theta > \underline{\theta}(X_1, ..., X_n)\} \geq 1 - \alpha,$$

则 $(\underline{\theta}, \infty)$ 称为参数 θ 的置信水平为 $1-\alpha$ 的单侧置信区间, $\underline{\theta}$ 称为单侧置信下限.

若

$$P\left\{\theta < \overline{\theta}(X_1, \ldots, X_n)\right\} \geq 1 - \alpha,$$

则 $(-\infty, \overline{\theta})$ 称为参数 θ 的置信水平为 $1-\alpha$ 的单侧置信区间, $\overline{\theta}$ 称为单侧置信上限.

正态总体均值的单侧置信区间 (σ^2 未知)

设总体 $X \sim N(\mu, \sigma^2)$, μ, σ^2 均未知, $X_1, \ldots, X_n$ 是一个样本.

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

则

- μ 的一个置信水平为 $1-\alpha$ 的单侧置信区间为
$$\left(\bar{X} - \frac{S}{\sqrt{n}} t_\alpha(n-1), +\infty\right).$$
- 单侧置信下限 $\underline{\mu} = \bar{X} - \frac{S}{\sqrt{n}} t_\alpha(n-1)$.

- μ 的一个置信水平为 $1-\alpha$ 的单侧置信区间为

$$\left(-\infty, \bar{X}+\frac{S}{\sqrt{n}}t_\alpha(n-1)\right).$$

- 单侧置信上限 $\overline{\mu}=\bar{X}+\frac{S}{\sqrt{n}}t_\alpha(n-1)$.

正态总体方差的单侧置信区间 (μ 未知)

设总体 $X \sim N(\mu, \sigma^2)$, μ, σ^2 均未知, $X_1, \ldots, X_n$ 是一个样本.

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

- 则 σ^2 的置信水平为 $1-\alpha$ 的单侧置信区间

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

- 单侧置信上限 $\overline{\sigma^2} = \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}$.

- σ^2 的置信水平为 $1-\alpha$ 的单侧置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_\alpha(n-1)}, +\infty\right).$$

- 单侧置信下限 $\underline{\sigma}^2 = \frac{(n-1)S^2}{\chi^2_\alpha(n-1)}$.

例

从一批灯泡中随机地取 5 只作寿命试验, 测得寿命 (以 h 计) 为

$$1050, 1100, 1120, 1250, 1280$$

设灯泡寿命服从正态分布, 求灯泡寿命均值地置信水平为 0.95 的单侧置信下限.

例

从一批灯泡中随机地取 5 只作寿命试验, 测得寿命 (以 h 计) 为

$$1050, 1100, 1120, 1250, 1280$$

设灯泡寿命服从正态分布, 求灯泡寿命均值地置信水平为 0.95 的单侧置信下限.

解: $1-\alpha=0.95$, $\alpha=0.05$,
$t_\alpha(n-1)=t_{0.05}(4)=2.1318$. $\bar{x}=1160$, $s^2=9950$.
$\underline{\mu}=\bar{x}-\frac{s}{\sqrt{n}}t_\alpha(n-1)=1065$. □