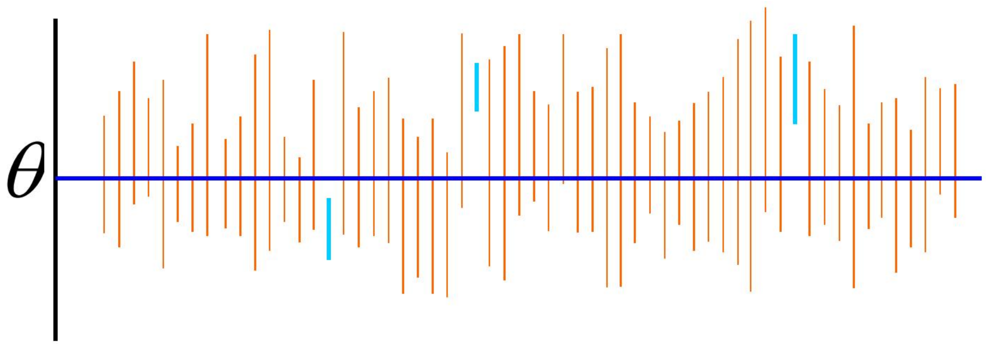
Chpt.7 Statistical Inference: Parameter Estimation

第七章 参数估计

上节回顾

称随机区间$(\hat{\theta}_1, \hat{\theta}_2)$ 为参数θ的置信水平为1-α的**双侧置信区间**

$$P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} \geq 1-\alpha$$



如反复抽样10000次, 当$\alpha = 0.05$,即置信水平为95%时, 10000个区间中包含θ真值的约为9500个;

当$\alpha = 0.01$,即置信水平为99%时, 10000个区间中包含θ的真值的约为9900个.

区间估计的一般步骤：

STEP 1：找一个合适的枢轴量$\widehat{W}$，要求满足

[1] 含有样本与待估参数；

[2] 其分布是已知的，且不依赖于样本与待估参数

STEP2：对于置信水平1−α，定义常数a和b，使得概率$P\{a < \widehat{W} < b\} = 1 - \alpha$

STEP3：反解出 $P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} \geq 1-\alpha$

STEP4：$(\hat{\theta}_1, \hat{\theta}_2)$ 就是置信水平1-**α**的置信区间（用样本观测值表示）

几类抽样分布 $X \sim N(\mu, \sigma^2)$

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

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

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

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

正态总体的区间估计 $X \sim N(\mu, \sigma^2)$

待估参数	前提条件	枢轴量	置信区间
均值	方差已知	$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\left(\overline{x}-\dfrac{\sigma_0}{\sqrt{n}}z_{\alpha/2},\ \overline{x}+\dfrac{\sigma_0}{\sqrt{n}}z_{\alpha/2}\right)$
	方差未知	$\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$\left(\overline{x}-\dfrac{s}{\sqrt{n}}t_{\alpha/2},\ \overline{x}+\dfrac{s}{\sqrt{n}}t_{\alpha/2}\right)$
方差	均值已知	$\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$	$\left(\dfrac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{\alpha/2}}, \dfrac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{1-\alpha/2}}\right)$
	均值未知	$\dfrac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$	$\left(\dfrac{(n-1)s^2}{\chi^2_{\alpha/2}}, \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right)$

习题1：

设总体$X \sim N(\mu, 9)$，$(X_1, \cdots, X_n)$是取自总体X的样本。欲使μ的$1-\alpha$置信区间长度L不超过2，问在以下两种情况下，样本容量n至少应取多少？

（1）$\alpha = 0.1$ （2）$\alpha = 0.01$

习题：置信区间与n的关系

设总体$X \sim N(\mu, 9)$，$(X_1, \cdots, X_n)$是取自总体X的样本欲使μ的$1-\alpha$置信区间长度L不超过2，问在以下两种情况下两种情况下样本容量n至少应取多少？

（1）$\alpha = 0.1$　　（2）$\alpha = 0.01$

解 给出$\sigma^2 = 9$已知，μ的置信区间

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

其区间长度

$$L = 2z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

对于给定α，欲使

$$L = 2z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le 2$$

必须

$$n \ge (z_{\frac{\alpha}{2}} \cdot \sigma)^2$$

对于α = 0.1与α = 0.01，

$z_{0.05} = 1.64$，$z_{0.005} = 2.58$，

计算可得

$$n_1 \ge \sigma^2 \mu_{0.05}^2 = 24.2$$

$$n_2 \ge \sigma^2 \mu_{0.005}^2 = 59.9$$

故样本容量n分别至少应取25与60.

请讨论区间长度与n的关系

习题2：

随机从某毛纺厂生产的羊毛锭中抽测10个样品的含脂率%，得到样本均值$\bar{x}$=7.7，样本方差S^2=0.64，假定含脂率服从正态分布。试分别在下面的置信度下给出平均含脂率的置信区间。（1）1-α=90%；（2）1-α=95%；

习题：不同置信度下区间的变化

随机从某毛纺厂生产的羊毛锭中抽测10个样品的含脂率%，得到样本均值$\bar{x}$=7.7，样本方差S^2=0.64，假定含脂率服从正态分布。试分别在下面的置信度下给出平均含脂率的置信区间。（1）1-α=90%；（2）1-α=95%；

解 由题意，σ未知，平均含脂率μ的1-α置信区间为

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

这里n=10，$\bar{x}$=7.7，S=0.8，对于$\alpha = 0.10$与$\alpha = 0.05$，

查附表$t_{0.05}(9) = 1.833$，$t_{0.025}(9) = 2.262$，分别代入计算可得

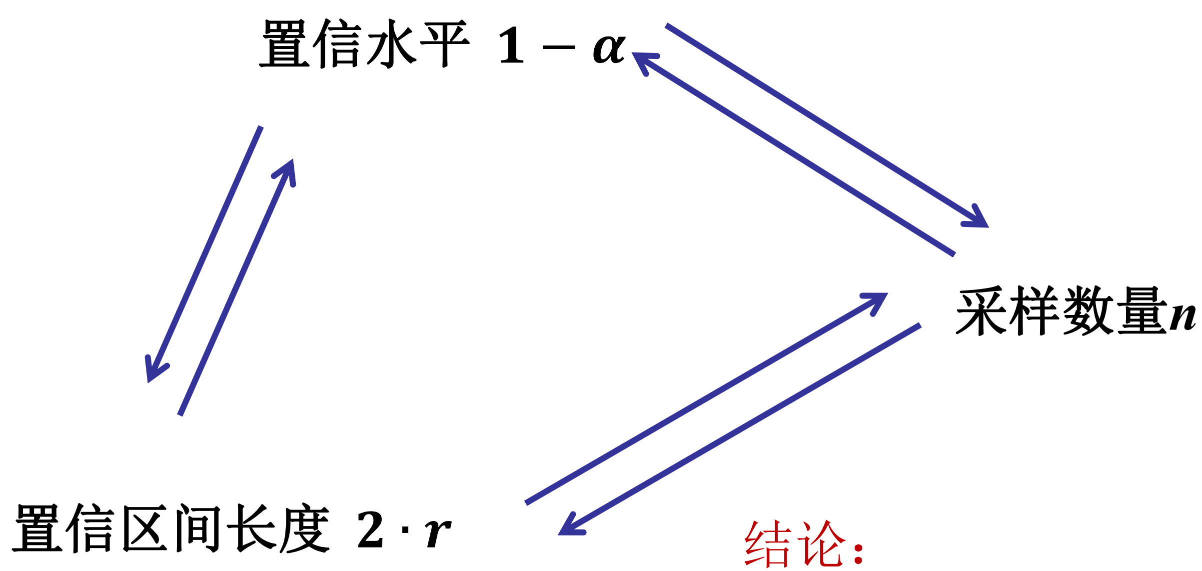
$$\bar{x} \pm t_{0.05}(9) \cdot \frac{S}{\sqrt{n}} = 7.7 \pm 0.46 \qquad \bar{x} \pm t_{0.025}(9) \cdot \frac{S}{\sqrt{n}} = 7.7 \pm 0.57$$

故所求置信区间分别为（7.24,8.16），（7.13,8.27）

请比对理解区间长度和置信度

置信水平、置信区间长度、样本容量的关系

置信水平越高，可靠度越大



置信区间越短，精确度越高

结论：

- 样本数量不变，置信水平越高，就会导致置信区间越长；反之，置信区间越短，就会导致置信水平越低。
- 在相同的置信水平下，要让置信区间越短，就要增加样本数量，这样采样成本就会越大。

单侧置信区间

前面讨论的估计量的置信区间都是双侧的，在有些实际问题中，例如某元件的使用寿命，平均寿命越长越好，在这种情况下，可将置信上限取为$+\infty$，而只关心置信下限。与之相反，在考虑化学药品中杂质含量时，杂质越少越好，只关心置信上限。因此引出了单侧置信区间的概念。

单侧置信区间

对于给定的(0<α<1),根据样本确定的统计量$\underline{\theta}(X_1,\cdots,X_n)$，对于任意的 $\theta\in\Theta$ 有 $P\{\theta>\underline{\theta}\}\geq 1-\alpha$ ，则随机区间 $(\underline{\theta},+\infty)$ 称做参数θ的置信水平为1-α的单侧置信区间；其中 $\underline{\theta}$ 叫做置信水平为1-α的单侧置信下限。

又若，统计量 $\overline{\theta}(X_1,\cdots,X_n)$，对于任意的 $\theta\in\Theta$ 有 $P\{\theta<\overline{\theta}\}\geq 1-\alpha$ ，则随机区间 $(-\infty,\overline{\theta})$ 称做参数θ的置信水平为1-α的单侧置信区间；其中 $\overline{\theta}$ 叫做置信水平为1-α的单侧置信上限.

单侧置信区间

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

$$1\,050 \quad 1\,100 \quad 1\,120 \quad 1\,250 \quad 1\,280$$

设灯泡寿命服从正态分布．求灯泡寿命平均值的置信水平为 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

1 050　1 100　1 120　1 250　1 280

设灯泡寿命服从正态分布. 求灯泡寿命平均值的置信水平为 0.95 的单侧置信下限.

解 $1-\alpha=0.95, n=5, t_{\alpha}(n-1)=t_{0.05}(4)=2.1318, \bar{x}=1\,160, s^2=9\,950.$ 由(7.4)式得所求单侧置信下限为

$$\underline{\mu}=\bar{x}-\frac{s}{\sqrt{n}}t_{\alpha}(n-1)=1\,065. \qquad \square$$

7.5 两个正态总体均值差与方差比的区间估计

不同工艺生产的两批同类产品，可以认为是来自两个相互独立的不同总体. 有时我们要对其某个质量指标作比较，分析它们是否有显著的差异.这时可观察 $\mu_1-\mu_2$ 和 σ_1^2/σ_2^2 的置信区间.

以下讨论两个正态总体均值差与方差比的区间估计问题.

设总体 $X\sim N(\mu_1,\sigma_1^2)$，且 $Y\sim N(\mu_2,\sigma_2^2)$ 和 $(x_1,x_2,\cdots,x_{n_1})$ $(y_1,y_2,\cdots,y_{n_2})$

分别是总体 X 与 Y 的样本观察值.

几类抽样分布

$X \sim N(\mu_1, \sigma_1^2)$， $Y \sim N(\mu_2, \sigma_2^2)$

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

[2] 当$\sigma_1=\sigma_2=\sigma$时
$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$$

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

[3]
$$\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2 \Big/ n_1\sigma_1^2}{\sum_{j=1}^{n_2}(y_j-\mu_2)^2 \Big/ n_2\sigma_2^2} \sim F(n_1,n_2)$$

[4]
$$\frac{S_1^2 \big/ \sigma_1^2}{S_2^2 \big/ \sigma_2^2}=\frac{S_1^2 \big/ S_2^2}{\sigma_1^2 \big/ \sigma_2^2} \sim F(n_1-1,n_2-1)$$

两个正态总体均值差的区间估计

(1)设两个正态总体 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$ 已知 σ_1, σ_2 ，求 $\mu_1 - \mu_2$ 的区间估计

选择包含$\mu_1 - \mu_2$随机变量 $$V = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

对于给定的置信水平1-α，取区间$(-z_{\alpha/2}, z_{\alpha/2})$，使

$$P\left\{-z_{\alpha/2} \le \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \le z_{\alpha/2}\right\} = 1 - \alpha$$

两个正态总体均值差的区间估计

(1)设两个正态总体 $X \sim N(\mu_1, \sigma_1^2)$,$Y \sim N(\mu_2, \sigma_2^2)$ 已知 σ_1, σ_2 ，求 $\mu_1 - \mu_2$ 的区间估计

把关于随机事件的概率描述转化为关于 $\mu_1 - \mu_2$ 的概率描述

$$P\left\{ \overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\} = 1 - \alpha$$

得置信水平为1-**α**的 $\mu_1 - \mu_2$ 的置信区间：

$$\left(\overline{x} - \overline{y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \overline{x} - \overline{y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

注：若 σ_1, σ_2 未知且 n_1 与 n_2 很大时，可用 s_1^2, s_2^2 分别代替 σ_1^2, σ_2^2 ，仍使用上式作 $\mu_1 - \mu_2$ 的区间估计.

两个正态总体均值差的区间估计

(2)设两个正态总体 $X \sim N(\mu_1,\sigma_1^2)$,$Y \sim N(\mu_2,\sigma_2^2)$ 未知 σ_1,σ_2，但是

$\sigma_1=\sigma_2$，求 $\mu_1-\mu_2$的区间估计。

使用随机变量
$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

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

对于给定的置信水平1-α，我们取区间$(-t_{\alpha/2}，\ t_{\alpha/2})$，使

$$P\left\{\frac{\left|(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\right|}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\le t_{\alpha/2}\right\}=1-\alpha$$

两个正态总体均值差的区间估计

(2)设两个正态总体 $X \sim N(\mu_1, \sigma_1^2)$,$Y \sim N(\mu_2, \sigma_2^2)$未知 σ_1, σ_2， 但是 $\sigma_1 = \sigma_2$， 求$\mu_1 - \mu_2$ 的区间估计。

把关于随机事件的概率描述转化为关于 $\mu_1 - \mu_2$

$$P\left\{\bar{X} - \bar{Y} - t_{\alpha/2} S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \bar{X} - \bar{Y} + t_{\alpha/2} S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right\} = 1 - \alpha$$

得置信水平为1-α 的 $\mu_1 - \mu_2$ 的置信区间（用样本观测值）

$$\left(\bar{x} - \bar{y} - t_{\alpha/2} S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x} - \bar{y} + t_{\alpha/2} S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

两个正态总体均值差的区间估计

例2 两台机床生产同一个型号的滚珠,从甲机床生产的滚珠中抽取8个,从乙机床生产的滚珠中抽取9个,测得这些滚珠的直径(毫米)如下:

甲机床: **15.0, 14.8, 15.2, 15.4, 14.9, 15.1, 15.2, 14.8**;

乙机床: **15.2, 15.0, 14.8, 15.1, 15.0, 14.6, 14.8, 15.1, 14.5**.

设两台机床生产的滚珠直径服从正态分布

求: 这两台机床生产的滚珠直径均值差 $\mu_1 - \mu_2$ 的对应于置信水平**0.90**的置信区间, 如果:

(1) 已知两台机床生产的滚珠直径的标准差分别是

σ_1=0.18(毫米)及σ_2=0.24(毫米);

(2) 未知σ_1及σ_2,但假设$\sigma_1=\sigma_2$.

解 (1) σ_1及σ_2已知，估计 $\mu_1-\mu_2$，采用统计量

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

置信水平为1-α的 $\mu_1-\mu_2$ 的置信区间

$$\left(\overline{x}-\overline{y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \ \overline{x}-\overline{y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

查正态分布表得 $z_{0.05}$=1.645，代入上式得所求的置信区间

（-0.018，0.318）

解 (2) σ_1，σ_2未知，但$\sigma_1=\sigma_2$，估计 $\mu_1-\mu_2$，采用统计量

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

$$S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

置信水平为1-α的$\mu_1-\mu_2$ 的置信区间

$$\left(\overline{x}-\overline{y}-t_{\alpha/2}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \overline{x}-\overline{y}+t_{\alpha/2}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

取自由度k＝8+9-2=15,查t 分布表得$t_{0.05}$=1.753，再计算S_w=0.228，

代入上式得所求的置信区间为（-0.044，0.344）.

两个正态总体方差比的区间估计

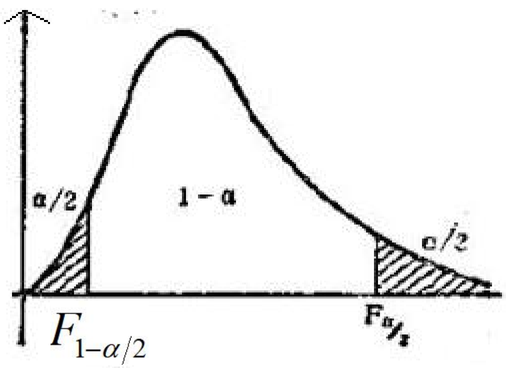
(1)设两个正态总体 $X\sim N(\mu_1,\sigma_1^2)$ **,** $Y\sim N(\mu_2,\sigma_2^2)$ **已知** μ_1,μ_2 **，求** σ_1^2/σ_2^2 **的区间估计**

利用随机变量

$$F=\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2\Big/ n_1\sigma_1^2}{\sum_{j=1}^{n_2}(Y_j-\mu_2)^2\Big/ n_2\sigma_2^2}\sim F(n_1,n_2)$$

对于给定的置信水平1-**α**，构造置信区间 $(F_{1-\alpha/2},F_{\alpha/2})$

$$P\left\{F_{1-\alpha/2}\le\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2\Big/ n_1\sigma_1^2}{\sum_{j=1}^{n_2}(Y_j-\mu_2)^2\Big/ n_2\sigma_2^2}\le F_{\alpha/2}\right\}=1-\alpha$$



两个正态总体方差比的区间估计

(1)设两个正态总体$X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$ 已知μ_1, μ_2， 求 σ_1^2 / σ_2^2的区间估计

把上述对于事件的描述转化为关于对方差比的描述

$$P\left(\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2 \Big/ n_1}{F_{\alpha/2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2 \Big/ n_2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2 \Big/ n_1}{F_{1-\alpha/2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2 \Big/ n_2} \right) = 1-\alpha$$

得置信水平为1-**α**的 σ_1^2 / σ_2^2的置信区间：

$$\left(\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2 \Big/ n_1}{F_{\alpha/2}\sum_{j=1}^{n_2}(y_j-\mu_2)^2 \Big/ n_2}, \frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2 \Big/ n_1}{F_{1-\alpha/2}\sum_{j=1}^{n_2}(y_j-\mu_2)^2 \Big/ n_2} \right)$$

两个正态总体方差比的区间估计

(2)设两个正态总体 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$未知 μ_1,μ_2 ，求 σ_1^2/σ_2^2 的区间估计

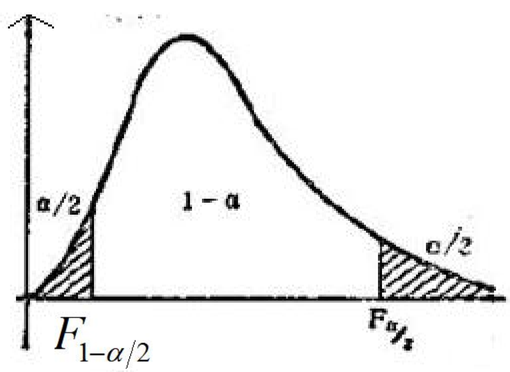
由于二者的总体均值未知，替代为样本均值，采用随机变量：

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

对于给定的置信水平1-α，构造置信区间 $(F_{1-\alpha/2},F_{\alpha/2})$

得置信水平为1-α的σ_1^2/σ_2^2的置信区间

$$\left(\frac{s_1^2}{F_{\alpha/2}s_2^2},\ \frac{s_1^2}{F_{1-\alpha/2}s_2^2}\right)$$



两个正态总体方差比的区间估计

例3 两台机床生产同一个型号的滚珠,从甲机床生产的滚珠中抽取8个,从乙机床生产的滚珠中抽取9个,测得这些滚珠的直径(毫米)如下:

甲机床: **15.0, 14.8, 15.2, 15.4, 14.9, 15.1, 15.2, 14.8**;

乙机床: **15.2, 15.0, 14.8, 15.1, 15.0, 14.6, 14.8, 15.1, 14.5**.

设两台机床生产的滚珠直径服从正态分布

求: 这两台机床生产的滚珠直径方差比 σ_1^2/σ_2^2 的对应于置信水平 $1-\alpha=0.90$ 的置信区间, 如果:

(1) 已知两台机床生产的滚珠直径的均值分别是 $\mu_1=15.0$(毫米)及 $\mu_2=14.9$(毫米);

(2) 未知 μ_1 及 μ_2.

解 已知 $n_1=8, n_2=9, \alpha=0.10$,

(1) 取自由度 $n_1=8, n_2=9$，查F分布表得 $F_{\alpha/2}=F_{0.05}(8,9)=3.23$

利用F分布的性质计算 $F_{1-\alpha/2}=F_{0.95}(8,9)=\dfrac{1}{F_{0.05}(9,8)}=\dfrac{1}{3.39}=0.295$

再计算

$$\sum_{i=1}^{8}(x_i-\mu_1)^2=0.34,\ \sum_{j=1}^{9}(y_j-\mu_2)^2=0.46$$

代入求得置信区间（**0.257**，**2.819**）.

解 已知 $n_1=8$, $n_2=9$, $\alpha=0.10$,

(2) 取自由度 $n_1=8-1$, $n_2=9-1$，查F分布表得 $F_{\alpha/2}=F_{0.05}(7,8)=3.50$

利用F分布的性质计算

$$F_{1-\alpha/2}=F_{0.95}(7,8)=\frac{1}{F_{0.05}(8,7)}=\frac{1}{3.73}=0.268$$

再计算

$$s_1^2=0.0457,\quad s_2^2=0.0575$$

代入求得置信区间（0.227，2.966）.

待估参数	前提条件	枢轴量
均值 $\mu_1-\mu_2$	方差已知	$$V=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$$
	方差未知，但两个方差相等	$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$
方差 σ_1^2/σ_2^2	均值已知	$$\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2\Big/n_1\sigma_1^2}{\sum_{j=1}^{n_2}(y_j-\mu_2)^2\Big/n_2\sigma_2^2}\sim F(n_1,n_2)$$
	均值未知	$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$$

表 7—1　正态总体均值、方差的置信区间与单侧置信限(置信水平为 $1-\alpha$)

	待估参数	其他参数	枢轴量 W 的分布	置信区间	单侧置信限
一个正态总体	μ	σ^2 已知	$Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\left(\overline{X}\pm\dfrac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$	$\overline{\mu}=\overline{X}+\dfrac{\sigma}{\sqrt{n}}z_{\alpha}$　$\underline{\mu}=\overline{X}-\dfrac{\sigma}{\sqrt{n}}z_{\alpha}$
	μ	σ^2 未知	$t=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$	$\left(\overline{X}\pm\dfrac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right)$	$\overline{\mu}=\overline{X}+\dfrac{S}{\sqrt{n}}t_{\alpha}(n-1)$　$\underline{\mu}=\overline{X}-\dfrac{S}{\sqrt{n}}t_{\alpha}(n-1)$
	σ^2	μ 未知	$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$	$\overline{\sigma^2}=\dfrac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}$　$\underline{\sigma^2}=\dfrac{(n-1)S^2}{\chi^2_{\alpha}(n-1)}$
两个正态总体	$\mu_1-\mu_2$	σ_1^2,σ_2^2 已知	$Z=\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$\left(\overline{X}-\overline{Y}\pm z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}\right)$	$\overline{\mu_1-\mu_2}=\overline{X}-\overline{Y}+z_{\alpha}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$ $\underline{\mu_1-\mu_2}=\overline{X}-\overline{Y}-z_{\alpha}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$
	$\mu_1-\mu_2$	$\sigma_1^2=\sigma_2^2=\sigma^2$ 未知	$t=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$ $S_w^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$	$\left(\overline{X}-\overline{Y}\pm t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}\right)$	$\overline{\mu_1-\mu_2}=\overline{X}-\overline{Y}+t_{\alpha}(n_1+n_2-2)S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$ $\underline{\mu_1-\mu_2}=\overline{X}-\overline{Y}-t_{\alpha}(n_1+n_2-2)S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$
	$\dfrac{\sigma_1^2}{\sigma_2^2}$	μ_1,μ_2 未知	$F=\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$	$\left(\dfrac{S_1^2}{S_2^2}\dfrac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\dfrac{S_1^2}{S_2^2}\dfrac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right)$	$\overline{\dfrac{\sigma_1^2}{\sigma_2^2}}=\dfrac{S_1^2}{S_2^2}\dfrac{1}{F_{1-\alpha}(n_1-1,n_2-1)}$ $\underline{\dfrac{\sigma_1^2}{\sigma_2^2}}=\dfrac{S_1^2}{S_2^2}\dfrac{1}{F_{\alpha}(n_1-1,n_2-1)}$

7.6 非正态总体参数的区间估计

若总体不服从正态分布时，一般很难确定总体中的未知参数的区间估计;

但当样本容量 n 很大时，中心极限定理告诉我们 $\boxed{\dfrac{\overline{X}-\mu(\theta)}{\sigma(\theta)/\sqrt{n}}}$ 近似

服从标准正态分布 **N(0,1)**. 可利用此作出近似的区间估计.

设总体X服从某一分布,其概率函数或密度函数中含有未知参数θ,则总

体均值与方差都依赖于参数θ. 对于样本 $X_1, X_2, \cdots, X_n$，它们相互

独立且与总体同分布, 设

$$E(X_i)=\mu(\theta),\quad D(X_i)=\sigma^2(\theta),\quad (i=1,2,\cdots,n)$$

7.6 非正态总体参数的区间估计

当样本容量n充分大(≥50)时，由列维定理知，样本函数 $\dfrac{\overline{X}-\mu(\theta)}{\sigma(\theta)/\sqrt{n}}$ 近似

服从标准正态分布**N(0,1).**

因此，对给定的置信水平1-**α**, 有

$$P\left\{\frac{\left|\overline{X}-\mu(\theta)\right|}{\sigma(\theta)/\sqrt{n}} \le z_{\alpha/2}\right\} \approx 1-\alpha$$

若能从不等式 $\dfrac{\left|\overline{x}-\mu(\theta)\right|}{\sigma(\theta)/\sqrt{n}} \le z_{\alpha/2}$ 解出参数θ，把关于随机事件的概

率描述转换为关于参数θ 的描述，从而得到参数θ的近似置信水平为1-**α**的

置信区间。

服从**0-1分布**的总体参数*p*的区间估计

设总体*X*服从“ 0-1 ”分布：x=0 或者 x=1，其中参数 p 未知. 则$E(X)=p$

$D(X)=p(1-p)$。对给定的置信水平1-**α**,得

$$P\left\{\frac{\left|\overline{X}-p\right|}{\sqrt{p(1-p)}\Big/\sqrt{n}}\le z_{\alpha/2}\right\}\approx 1-\alpha$$

对不等式 $\dfrac{\left|\overline{x}-p\right|}{\sqrt{p(1-p)}\Big/\sqrt{n}}\le z_{\alpha/2}$ 两边平方并整理得

$$n(\overline{x}-p)^2\le p(1-p)z_{\alpha/2}^2$$

再化作关于*p*的二次不等式

$$(n+z_{\alpha/2}^2)p^2-(2n\overline{x}+z_{\alpha/2}^2)p+n\overline{x}^2\le 0$$

服从0-1分布的总体参数*p*的区间估计

$$(n+z_{\alpha/2}^2)p^2-(2n\overline{x}+z_{\alpha/2}^2)p+n\overline{x}^2\le 0$$

令 $a=n+z_{\alpha/2}^2,\quad b=-(2n\overline{x}+z_{\alpha/2}^2),\quad c=n\overline{x}^2$

因为各x_i取值0或1，故 $0\le\overline{x}\le 1$ ，从而判别式

$$b^2-4ac=4n\overline{x}(1-\overline{x})z_{\alpha/2}^2+z_{\alpha/2}^4>0$$

令

$$\widehat{p_1}=\frac{-b-\sqrt{b^2-4ac}}{2a}$$

$$\widehat{p_2}=\frac{-b+\sqrt{b^2-4ac}}{2a}$$

则参数 p 的近似置信区间为$(\widehat{p_1},\widehat{p_2})$

服从0-1分布的总体参数p的区间估计

例4：从一批产品中,抽取100个样品,发现其中有75个优质品

求 这批产品的优质品率 p 对应于置信水平 0.95 的置信区间.

解 设总体 $X=\begin{cases}0 & \text{取到非优质品}\\ 1 & \text{取到优质品}\end{cases}$

则 X 服从 "0-1"分布: $P\{X=x\}=p^x(1-p)^{1-x}$ $x=0$ 或者 $x=1$

其中 p 为这批产品的优质品率.

按题意，样本容量$n=100$,在样本观测中恰有25个0与75个1,所以 $\bar{x}=0.75$

查表得$z_{\alpha/2}=z_{0.025}=1.96$，于是代入公式计算得

$$a=100+1.96^2=103.8406$$

$$b=-(2\times100\times0.75+1.96^2)=-153.8416$$

$$c=100\times0.75^2=56.25$$

由此得$\widetilde{p_1}=0.657,\widetilde{p_2}=0.825$ 参数 p 的近似置信区间为(0.657,0.825)

服从指数分布的总体参数θ的区间估计

总体X服从指数分布 $e(\theta)$，其中参数θ未知, 则$E(X)=\theta$, $D(X)=\theta^2$.

对给定的置信水平1-α, 有 $$P\left\{\frac{\left|\overline{X}-\theta\right|}{\theta/\sqrt{n}}\le z_{\alpha/2}\right\}\approx 1-\alpha$$

$$P\left\{\frac{\overline{X}}{1+\frac{z_{\alpha/2}}{\sqrt{n}}}\le\theta\le\frac{\overline{X}}{1-\frac{z_{\alpha/2}}{\sqrt{n}}}\right\}\approx 1-\alpha$$

故参数θ的近似置信区间为 $$\left(\frac{\overline{x}}{1+\frac{z_{\alpha/2}}{\sqrt{n}}},\ \frac{\overline{x}}{1-\frac{z_{\alpha/2}}{\sqrt{n}}}\right)$$

服从指数分布的总体参数θ的区间估计

例5：从一批电子元件中,随机抽取 **50** 个样品,测得它们的平均寿命为 **1200** 小时,设电子元件的使用寿命服从指数分布 $e(\theta)$,求参数θ相应于置信水平 **0.99** 的置信区间.

解 已知 $n=50,\ \bar{x}=1200,\ \alpha=0.01$ ，查正态分布表得 $Z_{0.005}=2.576$.

$$\hat{\theta}_1=\frac{\bar{x}}{1+\frac{z_{\alpha/2}}{\sqrt{n}}}=\frac{1200}{1+\frac{2.576}{\sqrt{50}}}=879.571$$

$$\hat{\theta}_2=\frac{\bar{x}}{1-\frac{z_{\alpha/2}}{\sqrt{n}}}=\frac{1200}{1-\frac{2.576}{\sqrt{50}}}=1887.687$$

故所求参数θ的置信区间为 **(879.571, 1887.687).**

Review1

统计任务:

[1] 参数估计
- -- 点估计
- -- 区间估计

[2] 假设检验
- -- 关于参数的假设
- -- 关于分布的假设

基本概念

- 对某一数量（或几个）指标进行随机实验、观察，将试验的全部可能的观察值称为<u>总体</u>。
- 每个可能的观察值称为<u>个体</u>

总体 ⟷ 随机变量

Review2

抽样：对总体进行一次观察并记录其结果，称为一次抽样；

对X独立进行n次观察，并将结果按顺序记为

$$X_1, \cdots, X_n$$

样本：随机抽取部分个体，以用于推断总体的特性。

样本与总体是同分布的

样本之间是独立的

统计量：样本的函数，除了样本、样本的参数外，不含有其他未知量

抽样分布：统计量的分布称为抽样分布

Review3

参数估计

- 点估计：总体X的形式已知，但有参数未知；

  借助总体X的样本来估计总体分布中的未知参数θ

矩估计

最大似然估计

- 区间估计：确定两个估计量 $\hat{\theta}_1(X_1,\cdots,X_n),\hat{\theta}_2(X_1,\cdots,X_n)$

并给出落在此区间的概率1-α。

称 $(\hat{\theta}_1,\hat{\theta}_2)$ 为置信水平为1-α的置信区间。