第二十课

7．3 区间估计

前面学习了 X（总体）中未知参数 θ，由样本 $X_1,X_2,\ldots,X_n$ 所做的估计 $\hat{\theta}(X_1,X_2,...,X_n)$（矩法、极大似然法）。但于估计值 $\hat{\theta}(X_1,X_2,...,X_n)$ 是随机的，它是在真值附近摆动的。到底摆动多少？即要更确切地知道 $\hat{\theta}(X_1,X_2,...,X_n)$ 摆动的范围。

定义：设 X 的分布中含参数 θ，如果由样本 $X_1,X_2,\ldots,X_n$ 确定的两个统计量 $\hat{\theta}_1(X_1,X_2,...,X_n)$ 和 $\hat{\theta}_2(X_1,X_2,...,X_n)$ 对给定的 a (0<a<1)，满足 $P\{\hat{\theta}_1<\theta<\hat{\theta}_2\}=1-\alpha$，称随机区间 $(\hat{\theta}_1,\hat{\theta}_2)$ 是参数 θ 的置信水平（置信度）为 1-a 的置信区间，$\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别称为置信下限和置信上限。

式子 $P\{\hat{\theta}_1<\theta<\hat{\theta}_2\}=1-\alpha$ 的意义：

若反复抽样多次，每组样本观察值确定一个区间 $(\hat{\theta}_1,\hat{\theta}_2)$，每个区间要么包含真值 θ，要么不包含真值 θ。在这些区间中，包含真值的占约占 100(1-a)%；不包含真值的点仅 100a%

一般 a 很小，如取 a=0.05，那么包含真值的可能性就应为 1-a=95%

对于一般的分布问题越难，下面只讨论正态分布中的参数 μ 和 σ 的区间估计。

7．4 正态总体均值与方差的区间估计

一、 单个总体 N(μ，σ²) 的情况

1. 若 σ^2 为已知，求 μ 的置信区间

$X_1,X_2,\ldots,X_n$ 为样本，$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 作为 μ 的点估计量，而 $\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$，所以 $U=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，

给定 a ，由正态分布表可查得 $Z_{a/2}$

使 $P\{|U|<Z_{a/2}\}=P\{|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}|<Z_{a/2}\}=1-a$ ，即 $P\{\overline{X}-Z_{a/2}\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+Z_{a/2}\frac{\sigma}{\sqrt{n}}\}=1-a$

所以 μ 的置信水平为 1-a 的置信区间为：$(\overline{X}-Z_{a/2}\frac{\sigma}{\sqrt{n}},\overline{X}+Z_{a/2}\frac{\sigma}{\sqrt{n}})$

例 1：设总体 X~N(μ ,0.09)，随机抽得 4 个独立观察值 x_1,x_2,x_3,x_4，求总体均值 μ 的 95%的置信区间。

解：a=1-0. 95=0. 05，$\sigma=\sqrt{0.09}=0.3$，样本容量 n=4，$\overline{X}=\frac{1}{4}\sum_{i=1}^{4}x_i$，按双侧分位点 $Z_{a/2}=Z_{0.05/2}=1.96$

所以有 $\overline{X}-\frac{0.3}{2}Z_{0.05/2}<\mu<\overline{X}+\frac{0.3}{2}Z_{0.05/2}$ 得 $(\overline{X}-0.294,\overline{X}+0.294)$ 为参数 μ 的置信度为 0. 95 下的区间估计。

如：$x_1=12.6,x_2=13.4,x_3=12.8,x_4=13.2$，则得 $\overline{X}=13$，所以有（12.71，13.29）

2. 若 σ^2 为未知，求 μ 的置信区间

由于 σ 未知，可以用 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 作为 σ^2 的估计。那么 $t=\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$，给定 a，可

查得 $t_{a/2}(n-1)$，使得

$P\{|\frac{\overline{X}-\mu}{S/\sqrt{n}}|<t_{a/2}(n-1)\}<1-a$，可解得 $P\{\overline{X}-\frac{S}{\sqrt{n}}t_{a/2}(n-1)<\mu<\overline{X}+\frac{S}{\sqrt{n}}t_{a/2}(n-1)\}=1-a$

所以μ的置信水平为 1-a 的置信区间为：

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

在实际应用中，这种情况更实用，因为在大多数情况下方差都不知道。

例 2：为确定某种溶液中的甲醛浓度，取样得 4 个独立测定值的平均值 $\overline{X}=8.34\%$，样本标准差为 S=0.03%，没被测总体近似地服从正态分布，求总体均值μ的 95%置信区间。

解：因为 1-a=0.95 所以 a=0.05，n=4，查表有 $t_{0.05/2}(3)=3.1824$，$\frac{S}{\sqrt{n}}=\frac{0.03}{\sqrt{4}}=0.015$，于是μ的 95%置信区间为：

（(8.34-3.1824X0.015）%，（8.34+3.1824X0.015）%=(8.292%，8.388%)

即可认为这种溶液中甲醛浓度在(.292%到 8.388%之间，这个估计的可靠度为 95%

3. 方差 σ^2 的区间估计

$X\sim N(\mu\ \sigma^2)$，参数 $\mu\ \sigma^2$ 未知，要找一个与 σ^2，S^2 有关的表达式。

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

由第六章定理 2 得这式子

给定 a 查 $\chi^2_{a/2}(n-1)$ 和 $\chi^2_{1-a/2}(n-1)$，使 $P\{\chi^2_{1-a/2}(n-1)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{a/2}(n-1)\}=1-a$

即 $P\{\frac{(n-1)S^2}{\chi^2_{a/2}(n-1)}<\sigma^2<\frac{(n-1)S^2}{\chi^2_{1-a/2}(n-1)}\}=1-a$，所以 σ ^{2} 的置信水平为 1-a 的置信区间为：

$(\frac{(n-1)S^2}{\chi^2_{a/2}(n-1)}, \quad \frac{(n-1)S^2}{\chi^2_{1-a/2}(n-1)})$，而 σ 的置信水平为 1-a 的置信区间为：$(\frac{\sqrt{n-1}S}{\sqrt{\chi^2_{a/2}(n-1)}}, \quad \frac{\sqrt{n-1}S}{\sqrt{\chi^2_{1-a/2}(n-1)}})$

例 3：在例 2 中，求 σ ^{2} 的 95%的置信区间

解：因为 a/2=0.05/2=0.025，1-a/2=0.975，n=4，S^2=0.0009，$\chi^2_{0.025}(3)=9.348$，$\chi^2_{0.975}(3)=0.216$

所以 σ ^{2} 的 95%的置信区间为：$(\frac{3\times 0.0009}{9.348}, \quad \frac{3\times 0.0009}{0.216})=(0.00029, \quad 0.0125)$

二、 两总体 N(μ $_1$, σ $_1^2$), N(μ $_2$, σ $_2^2$) 的情况

$\bar{X}_1$，S_1^2 为总体 N(μ $_1$, σ $_1^2$) 的容量为 n_1 样本均值和方差；$\bar{X}_2$，S_2^2 为总体 N(μ $_2$, σ $_2^2$) 的容量为 n_2 样本均值和方差，设两样本相互独立。

1、μ $_1$−μ $_2$ 的区间估计

因为 $E(\bar{X}_1-\bar{X}_2)=\mu_1-\mu_2$，$D(\bar{X}_1-\bar{X}_2)=D(\bar{X}_1)+D(\bar{X}_2)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$，而 $\bar{X}_1-\bar{X}_2 \sim N(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$

1^0 σ $_1^2$， σ $_2^2$ 都已知

由前面对于单个正态时的σ ^{2}为已知情形有：$(\bar{X}_1-\bar{X}_2\pm Z_{a/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_1^2}{n_1}})$

2^0 σ $_1{}^2$， σ $_2{}^2$未知

这时只要当 n_1, n_2>50 有$(\bar{X}_1-\bar{X}_2\pm Z_{a/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_1^2}{n_1}})$

3^0 σ $_1{}^2$= σ $_2{}^2$= σ ^{2}， 但σ ^{2}为未知

由第六章定理 3 知 $\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$，

所以有$(\bar{X}_1-\bar{X}_2\pm t_{a/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})$， 其中$S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$

例 4：为提高某一化学生产过程的得率，试图采用一种新的催化剂。为慎重起见，在实验工厂先进行试验。设采用原来的催化剂进行了 n_1=8 次试验，得到得率的平均值$\bar{x}_1=91.73$。样本方差 S_1^2=3.89；又采用新的催化剂进行了 n_2=8 次试验，得到得率的均值$\bar{x}_2=93.75$，样本方差 S_2^2=4.02。假设两总体都可认为服从正态分布，且方差相等，试求两总体均值差μ $_1$- μ $_2$的置信度为 0.95 的置信区间。

解：这是属于σ$_1^2$=σ$_2^2$=σ^2，但σ^{2}为未知的情况，因为 $S_w=\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}=\sqrt{3.96}$，所以

有 $(\overline{X}_1-\overline{X}_2\pm t_{a/2}(n_1+n_2-2)S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}})=(91.73-93.75\pm t_{0.025}(14)\sqrt{3.96}\sqrt{\dfrac{1}{8}+\dfrac{1}{8}})=(-2.02\pm 2.13)$

即（-4.15，0.11）

由于所得置信区间包含零，所以在实际中我们就认为采用这两种催化剂所得的得率均值没有显著差别。当然，如果区间的下限大于零，那么说明μ$_1$大于μ$_2$；如果区间的上限大于零，那么说明μ$_1$小于μ$_2$

2、$\dfrac{\sigma_1^2}{\sigma_2^2}$ 的区间估计

对于 F 分布，参看第六章内容。

因为 $\dfrac{(n_1-1)S_1^2}{\sigma_1^2}\sim\chi^2(n_1-1)$， $\dfrac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_2-1)$

所以 $\dfrac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)}=\dfrac{\dfrac{S_1^2}{S_2^2}}{\dfrac{\sigma_1^2}{\sigma_2^2}}\sim F(n_1-1,n_2-1)$，得：$P\{F_{1-a/2}(n_1-1,n_2-1)<\dfrac{\dfrac{S_1^2}{S_2^2}}{\dfrac{\sigma_1^2}{\sigma_2^2}}<F_{a/2}(n_1-1,n_2-1)\}=1-a$

所以有 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 在置信度为 1-a 下的区间估计为：$(\dfrac{S_1^2}{S_2^2F_{a/2}(n_1-1,n_2-1)},\dfrac{S_1^2}{S_2^2F_{1-a/2}(n_1-1,n_2-1)})$

注意$\frac{\sigma_1^2}{\sigma_2^2}$的意义

我们知道方差是考查波动性的，若σ_1^2/σ_2^2区间上限小于 1，则 N(μ，若σ_1^2)的波动性较小；当σ_1^2/σ_2^2区间下限大于 1 时则 N(μ，若σ_1^2)的波动性较大；等于 1 时不能判定两个总体的波动性。

例 5：两正态总体的参数都未知，依次取容量为 25，15 的两独立样本，测得样本方差依次为 6.38，5.15，求总体方差比σ_1^2/σ_2^2的 90%置信区间。

解：因为 $n_1=25$，$n_2=15$，所以 $n_1-1=24$，$n_2-1=14$，又 1-a=0.9，a/2=0.05，1-a/2=0.95，查表得 $F_{0.05}(24,14)=2.35$，$F_{0.95}(24,14)=1/F_{0.05}(14,24)=1/2.13$

而 $S_1^2/S_1^2=6.38/5.15=1.24$

所以σ_1^2/σ_2^2的 90%置信区间为：（1.24/2.35，1.24X2.13）=(0.528，2.64)

7．6 单侧置信区间

前面都是双侧的情况，有的问题只要单侧，如产品寿命大于某个时间（T，+∞）。

方法：

在双侧的基础上只取一边，但注意分位点也应取单侧，如：

σ^2为已知：$P\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<Z_a\}=1-a$，解出 $\bar{X}-Z_a\frac{\sigma}{\sqrt{n}}<\hat{\mu}<+\infty$

σ^2为未知：$P\{\frac{\bar{X}-\mu}{S/\sqrt{n}}<t_a(n-1)\}=1-a$，解得 $\bar{X}-t_a(n-1)\frac{S}{\sqrt{n}}<\hat{\mu}<+\infty$

例 1：已知某种电子原件的的寿命 X 服从正态分布，现抽取 5 只测得寿命为：1050，1100，1120，1250，1280，试求其寿命的 95%的置信下限。

解：因为 $\bar{x}=1160$，$S^2=9950$，又因为 1-a=0.95，a=0.05，$t_{0.05}(5-1)=t_{0.05}(4)=2.1318$，估寿命的 95%的单侧置信区间下限为

$(1160-2.1318\frac{\sqrt{9950}}{\sqrt{5}},\quad +\infty)=(1065,+\infty)$，即可认为这批原件的寿命有 95%的可能性是大于 1065 的。