第六章　　参数的估计

(Estimation of parameter)

张爱莲

新疆大学生命科学与技术学院　新疆生物资源基因工程重点实验室

zalxju@gmail. com

20101209

课前复习

参数估计的概念与内容

- **参数估计的定义：**

参数估计就是用样本统计量来估计总体参数。

- **参数估计的内容：**

点估计 （point estimation）

区间估计（interval estimation）

本章内容

6.1 点估计

6.2 区间估计

6.3 两个样本的差异显著性检验

6.1 点估计

- **定义：**

将样本统计量直接作为总体相应参数的估计值叫点估计。

样本平均数可估计总体平均数。

样本均方可估计总体方差。

一个好的估计量应满足三个条件：

1.无偏估计量：

如果一个统计量的数学期望等于总体参量值，则这个统计量就称为无偏估计量。 $E\left(\overline{x}\right)=\mu,\qquad E\left(S^2\right)=\sigma^2$

2.有效估计量：

在样本含量相同的情况下，如果一个估计量的方差小于另一个统计量的方差，则前一个统计量是更有效的估计量。

3.相容估计量：

若统计量的取值任意接近于参量值，则该统计量称为参数的相容估计量。

6.2　区间估计

- **定义：**

在一定概率保证下指出总体参数的可能范围，所给出的可能范围叫置信区间（confidence interval），给出的概率保证称为置信度或置信概率（confidence probability）。 以L_1、L_2分别表示置信区间的上、下限。保证参数在该区间的概率以1-α表示，称为置信水平（置信系数，置信度）。这样的估计方法称为参数的区间估计。

6.2.1 区间估计的一般原理

- 在例题中 H_0： $\mu=10.00g$，$\bar{x}=10.23g$ 所得 $u=1.82$

  $u<1.96$，接受H_0。

- 如果H_0不是 $\mu=10.00$，而是 $\mu=10.20$, 得$u=0.24$接受H_0

- 如果H_0不是 $\mu=10.00$，而是 $\mu=10.40$, 得$u=-1.34$接受H_0

这些H_0全都落在接受域内。由此可见，当用样本平均数估计总体平均数时所得到的结果不是单一值而是一个区间。

只要标准化的样本平均数落在$-u_{\alpha/2}$和$u_{\alpha/2}$区间内，所有H_0都将被接受，于是得到一个包括总体平均数的区间，用这种方法对总体参数所做的估计称为区间估计。

6.2.2 总体平均数 μ 的置信区间

设有一来自正态总体（μ，σ²）的样本，包含*n*个观测值，样本平均数 $\bar{x}$ ，标准误 $\sigma_{\bar{x}}$,样本平均数 $\bar{x}$ 服从N（ μ，$\sigma_{\bar{x}}^2$ ），当概率水平α=0.05或0.01时，置信度p=1- α=0.95的条件下，有：

$$P(-1.96\langle u\langle 1.96)=0.95 \rightarrow P\left(-1.96\langle \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\langle 1.96\right)=0.95$$

$$\rightarrow P\left(\bar{x}-1.96\frac{\sigma}{\sqrt{n}}\langle \mu\langle \bar{x}+1.96\frac{\sigma}{\sqrt{n}}\right)=0.95$$

因此，对于某一概率标准 α，则有通式：

$$P\left(\bar{x}-u_{\alpha/2}\sigma_{\bar{x}}\langle \mu\langle \bar{x}+u_{\alpha/2}\sigma_{\bar{x}}\right)=1-\alpha$$

一般式：$P(\bar{x}-u_{a/2}\sigma_{\bar{x}} \le \mu \le \bar{x}+u_{a/2}\sigma_{\bar{x}})=1-a$

$$或P(\bar{x}-u_{a/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x}+u_{a/2}\frac{\sigma}{\sqrt{n}})=1-\alpha$$

称为总体平均数 μ 置信度为1-*a*的置信区间。

其中，$\bar{x}-u_{\frac{\alpha}{2}}\sigma_{\bar{x}}$ 和 $\bar{x}+u_{\frac{\alpha}{2}}\sigma_{\bar{x}}$ 为置信下限和置信上限；置信上下限之差称为置信距，置信距越小，估计的精确度就越高。

总体平均数μ的95%和99%的置信区间如下：

$$\bar{x} - 1.96\sigma_{\bar{x}} \le \mu \le \bar{x} + 1.96\sigma_{\bar{x}}$$

$$\bar{x} - 2.58\sigma_{\bar{x}} \le \mu \le \bar{x} + 2.58\sigma_{\bar{x}}$$

【例】测得某批小麦25个样本得平均蛋白质含量 $\bar{x}=14.5\%$ ，已知 σ =2.50%，试估计总体蛋白质含量平均数的95%置信区间。

先求出：$\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}=\frac{2.50}{\sqrt{25}}=0.50$

区间估计为：$L_1=\bar{x}-u_{a(双侧)}\sigma_{\bar{x}}=14.5-1.96\times0.50=13.52\%$

$L_2=\bar{x}+u_{a(双侧)}\sigma_{\bar{x}}=14.5+1.96\times0.50=15.48\%$

σ未知，μ的置信区间

未知时，用S代替，变量 $t=(\bar{x}-\mu)/S_{\bar{x}}$ 服从n-1自由度的t分布， 因此μ的置信区间为：

$$P(\bar{x}-t_{df,\frac{\alpha}{2}}S_{\bar{x}}\le\mu\le\bar{x}+t_{df,\frac{\alpha}{2}}S_{\bar{x}})=1-a$$

或$$P(\bar{x}-t_{df,\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\le\mu\le\bar{x}+t_{df,\frac{\alpha}{2}}\frac{s}{\sqrt{n}})=1-\alpha$$

【例】已知玉米单交种群105的平均穗重μ_0=300g ，喷药后，随机抽取9个果穗，其穗重为：308，305，311，298，315，300，321，294，320克。问喷药前后的果穗重，差异是否显著？

解1：检验法：α=0.05

$$H_0: \mu = \mu_0 = 300 \quad H_A: \mu \neq \mu_0$$

$$\bar{x} = 308, \quad S = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}} = 9.62.$$

$$t = \frac{308 - 300}{\frac{9.62}{\sqrt{9}}} = 2.49$$

$t_{8,\frac{0.05}{2}} = 2.306$ **t>$t_{0.05/2}$，P<0.05，**拒绝 H_0，即喷药前后，果穗重有显著性差异。

解2：区间估计法，取1−α = **0.95**

$$\bar{x} = 308,\quad S = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}} = 9.62.$$

$$t_{8,\frac{0.05}{2}} = 2.306$$

∴置信区间为 $308 \pm 2.306 \times \frac{9.62}{\sqrt{9}} = 300.6 - 315.4$

即喷药后果穗重在**300.6-315.4**之间，此估计的可靠度为**95%**。

此区间不包括H_0**=300**克。因此在**=0.05**水平上喷药的效果是显著的，这与**t**检验的结果是一致的。

在实际应用中，置信区间越小，估计的结果越精确。

根据 $\bar{x} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{x} \pm u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ 决定置信区间长度的因素有三个：

（1）总体中天然存在的变异性σ或S。

（2）样本容量n。n越大，区间长度越短。

（3）α值越大，区间长度越短。不可取。

缩短区间长度，采用扩大样本容量和改善技术比较好。

6.2.3 两平均数差（成组数据）的置信区间

1. 标准差σ_i未知

根据 $$t=\frac{\left(\bar{x}_1-\bar{x}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\left(n_1-1\right)S_1^2+\left(n_2-1\right)S_2^2}{\left(n_1-1\right)\left(n_2-1\right)}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{\left(\bar{x}_1-\bar{x}_2\right)-\left(\mu_1-\mu_2\right)}{S_{\bar{x}_1-\bar{x}_2}}$$

$\mu_1-\mu_2$的$1-\alpha$置信区间

$$\left(\bar{x}_1-\bar{x}_2\right)\pm t_{\left(n_1+n_2-2\right),\frac{\alpha}{2}}\sqrt{\frac{\left(n_1-1\right)S_1^2+\left(n_2-1\right)S_2^2}{\left(n_1-1\right)+\left(n_2-1\right)}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

6.2.4 配对数据的置信区间

根据：$t = \dfrac{\bar{d} - \mu_d}{\dfrac{S_{\bar{x}}}{\sqrt{n}}}$

则μ_d的$1-\alpha$置信区间 $\bar{d} \pm t_{(n-1),\frac{\alpha}{2}} \dfrac{S_d}{\sqrt{n}}$

【例】用家兔10只试验某批注射液对体温的影响，测定每只家兔注射前后的体温。

$$\bar{d} = -0.73, \qquad S_{\bar{d}} = S_d \big/ \sqrt{n} = 0.445 \big/ \sqrt{10} = 0.141$$

注射前后体温差的95%的置信区间：

$$\bar{d} \pm t_{(n-1),\frac{\alpha}{2}} \frac{S_d}{\sqrt{n}} \Rightarrow -0.73 \pm 2.228 \times 0.141 \Rightarrow (-1.044, -0.416)$$