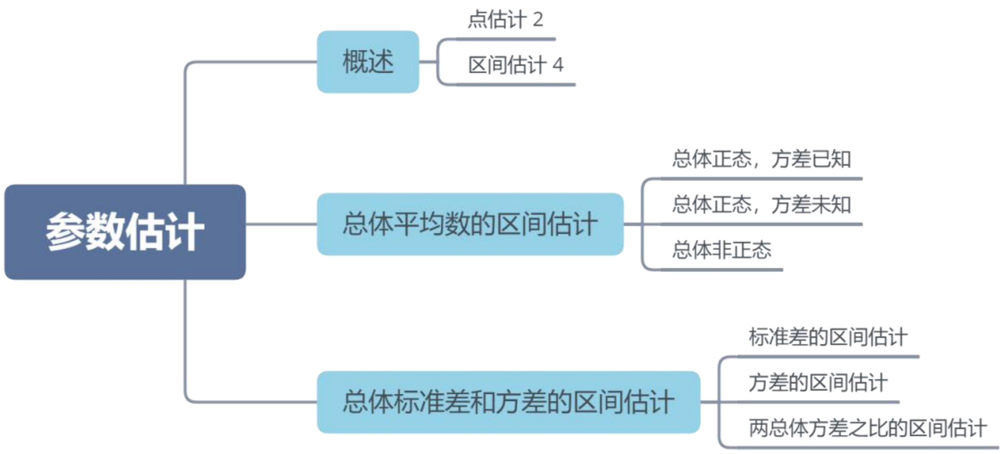
# 第七章　参数估计




## 一、概述

当在研究中当在研究中获得一组样本数据后，如何通过这组数据，对总体特征进行估计，也就是如何从局部结果推论总体的情况，称为总体参数估计。总体参数估计可以分为点估计与区间估计。

通过参数估计将得到一个数值（点估计）或一个区间（区间估计）来表示未知的总体参数或参数可能的变化范围。

### （一）点估计

1.含义

用单一的数值对总体的未知参数进行估计。即用样本统计量来估计总体参数，因为样本统计量为数轴上某一点的值，估计结果也以一个点的数值表示。

点估计能够提供总体参数的精确估计值，但是它总以误差的存在为前提，即可信度不高。

2.良好估计量的标准

（1）无偏性

用多个样本的统计量估计总体参数的估计值，其偏差的平均数为零，即样本量围绕着总体参数变化。

（2）有效性

当总体参数的无偏估计量不止一个时，无偏估计变异小者有效性高，变异大者有效性低，而应该选择变异小者，即方差越小越好。

（3）一致性

当样本容量无限增大时，估计值应该能够越来越接近它所估计的总体参数。

（4）充分性

样本的统计量是否充分地反映了全部 n 个数据所反映总体的信息。

### （二）区间估计

1.含义

根据估计量以一定可靠程度推断总体参数所在的区间范围，即用数轴上的一段距离表示

未知参数可能落入的范围。

它能在点估计的基础上，不仅给出一个估计范围，还能说明估计结果的把握程度（正确率），但是区间估计的精确度不高。

2.显著性水平、置信水平与置信区间

（1）显著性水平

总体参数落在某一区间，可能犯错误的概率，用α表示，又称意义阶段、信任系数。

（2）置信水平

又称置信度，即估计的正确率，用 1-α表示。

（3）置信区间

在某一置信水平时，总体参数所在的区域距离或长度。区间上下端点为置信界限。

影响置信区间的因素包括：

①样本容量：n 越大，标准误越小，置信区间越窄；本质即样本量越大，获得的信息就越多，于是估计得越准确。

②置信水平：置信水平越高，置信区间越宽。

③样本方差：样本数据变异性越大，对于相同置信度，所需置信区间越宽。

3.区间估计的原理

（1）原理

区间估计依据样本分布理论，利用样本分布提供概率解释，以标准误（SE）的大小为单位决定区间估计的长度。

所有总体参数的估计原理相同，但根据样本分布以及标准误的不同，计算方法不同。

（2）实际问题

若想提高正确估计的概率，则必然会增大置信区间，即降低了估计的精确性；而若想提高估计的精确性，则必然会减小置信区间，即降低了正确估计的概率。

实际上的解决方法是：在保证置信水平（正确率）的前提下，尽量提高精确度。通常是固定置信水平为 0.95 或 0.99，即α=0.05 或 0.01。

4.区间估计的步骤

（1）列出已知条件；

（2）计算标准误；

（3）查表获得临界值；

（4）根据样本分布、标准误以及通过查表获得的临界值来估计总体多数的范围。

## 二、总体平均数的区间估计

### （一）总体分布正态，总体方差已知

样本均值服从 Z 分布：

$$SE_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \qquad \overline{X} - Z_{\alpha/2} SE_{\overline{X}} < \mu < \overline{X} + Z_{\alpha/2} SE_{\overline{X}}$$

### （二）总体分布正态，总体方差未知

样本均值服从 t 分布：

$$SE_{\overline{X}}=\frac{s}{\sqrt{n-1}}=\frac{s_{n-1}}{\sqrt{n}} \qquad \overline{X}-t_{\alpha/2}SE_{\overline{X}}<\mu<\overline{X}+t_{\alpha/2}SE_{\overline{X}}$$

### （三）总体非正态

方差已知，但样本容量 n>30 时，样本均值服从 Z 分布，与第一种情况相同。

方差未知，但样本容量 n>30 时，样本均值服从 t 分布，与第二种情况相同。

| 总体分布形态 | 总体方差 | 样本量 | 样本平均数抽样分布 | 标准误 |
| --- | --- | --- | --- | --- |
| 正态分布 | 已知 | 无论大小 | 正态分布 | $SE_{\overline{X}}=\frac{\sigma}{\sqrt{n}}$ |
| 正态分布 | 未知 | 无论大小 | t分布 | $SE_{\overline{X}}=\frac{s}{\sqrt{n-1}}=\frac{s_{n-1}}{\sqrt{n}}$ |
| 非正态分布 | 已知 | n>30 | 渐进正态分布 | $SE_{\overline{X}}=\frac{\sigma}{\sqrt{n}}$ |
| 非正态分布 | 未知 | n>30 | t分布 | $SE_{\overline{X}}=\frac{s}{\sqrt{n-1}}=\frac{s_{n-1}}{\sqrt{n}}$ |
| 非正态分布 | | n<30 | | 不能用 |

## 三、总体标准差与方差的区间估计

### （一）标准差的区间估计

当 n>30 时，样本标准差服从 Z 分布：

$$SE=\frac{\sigma}{\sqrt{2n}} \qquad \sigma \text{ 未知时，可用 } s_{n-1} \text{ 代替}$$

$$s_{n-1}-Z_{\alpha/2}SE<\sigma<s_{n-1}+Z_{\alpha/2}SE$$

### （二）方差的区间估计

样本方差与总体方差的比值服从卡方分布：

$$\chi^2=\frac{ns^2}{\sigma^2}=\frac{(n-1)s_{n-1}^2}{\sigma^2} \qquad \frac{(n-1)s_{n-1}^2}{\chi^2_{\alpha/2}}<\sigma^2<\frac{(n-1)s_{n-1}^2}{\chi^2_{(1-\alpha/2)}} \qquad df=n-1$$

### （三）两总体方差之比的区间估计

两样本方差的比值服从 F 分布：

$$F=\frac{s_{n_1-1}^2}{s_{n_2-1}^2} \qquad \frac{1}{F_{\alpha/2}}\times\frac{s_{n_1-1}^2}{s_{n_2-1}^2}<\frac{\sigma_1^2}{\sigma_2^2}<F_{\alpha/2}\times\frac{s_{n_1-1}^2}{s_{n_2-1}^2}$$