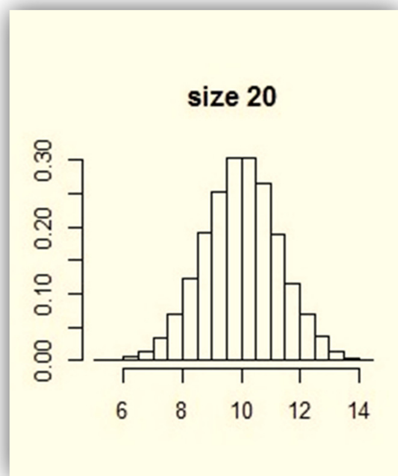
基本统计学应用于化学分析系列（IV）

总体(Population)与样本(Sample)



目录

4.1 总体(Population)与样本(Sample)之区别

定义：数理统计中把研究对象的全体称为总体，构成总体的每一个对象称为个体或样本。

实验室的分析结果是来自样本，往往样本的部分是来自于一个大量的母体。许多有代表性的个体，称样本可从母体中采集，但分析的最终对象是母体，或称总体，既通过样本的分析和统计手段来推断总体的实况。

在概念上，作为重复进行一个特定操作所得结果而可能发生的全体观测的集合称为观测的总体。统计的手段涉及到样本本身观察值的随机变化和样本与样本之间的数据概率分布。

图 4.1 表示总体与样本的关系

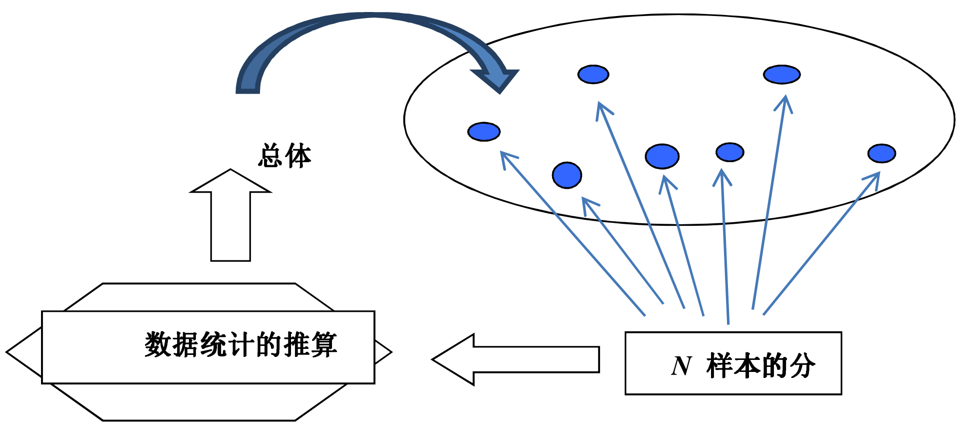


表 4.1 总体与样本统计值之比较

总体统计值	值符号	样本统计值	值符号
样品量取自总体	N	样本重复测试量	n
总体平均值	μ	样本平均值	$\bar{x}$
总体标准差	$\sigma = \sqrt{\dfrac{\sum (x_i - \mu)^2}{N}}$	样本标准差	$s = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}}$
总体方差	σ^2	样本方差	s^2

显然，一个样本测试所得的平均值 $\bar{x}$ 与总体的平均值是有差异的。当有代表性样本的测试次数无限增加时，或测试无限数量的样本时，所得的平均值逐渐接近总体平均值 μ。应用于大量测量数据的情况下，总体标准差则用 σ 表示，而 σ^2 则称总体方差。(注：所谓有代表性的样本，是指用随机抽样方法获得的样本。）同样地，在有限次测量中， 样本标准差 s 是如式 4-1，但当测量次数 n 增加时，$\bar{x}$ 越来越接近 μ，此时 s 也越来越接近 σ：

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} \qquad (4\text{-}1)$$

4.2 平均值标准误差 （Standard error of mean）

设有 m 组随机抽出的样本经过 n 次的测试给出 m 个平均值：

$$\begin{aligned}
&x_{11}, x_{12}, x_{13}, \ldots\ldots x_{1n} \rightarrow \bar{x}_1; s_{x1} \\
&x_{21}, x_{22}, x_{23}, \ldots\ldots x_{2n} \rightarrow \bar{x}_2; s_{x2} \\
&\ldots\ldots \\
&x_{m1}, x_{m2}, x_{m3}, \ldots\ldots x_{mn} \rightarrow \bar{x}_m; s_{xm}
\end{aligned}$$

m 个样本的总平均值 $\bar{\bar{x}}$ 为

$$\bar{\bar{x}} = \frac{1}{m}\sum \bar{x}_i^{\,2} \qquad (4\text{-}2)$$

如何从 m 个样本测定值得总的标准差呢？

各样本的测定平均值会不一样，但只要标准差与测定平均值无关，即于浓度相差仅几倍的范围内，测定精密度可视为相同时，每个样本的方差都可作为总体离散特性的量度。

因此 m 个样本均数 $\bar{\bar{x}}$ 的变异则以标准差表示，称均值标准误差 $s_{\bar{x}}$ 。根据随机误差的传递公式的方差为

$$s_{\bar{x}}^{\ 2} = \frac{1}{m}\left(s_{x1}^{\ 2} + s_{x2}^{\ 2} + \ldots + s_{xm}^{\ 2}\right)$$

式中 S_{xi}^2 为每个样本的方差，在相同条件下测量同一物理量，则可认为各次测量具有相同的精密度，即

$$s_{x1} = s_{x2} = s_{x3} = \ldots = s_{xm} = s$$

于是 $$s_{\bar{x}}^{\ 2} = \frac{s^2}{m} \quad 或 \quad s_{\bar{x}} = \frac{s}{\sqrt{m}}$$

数理统计学可以证明：用 m 个样本，每个样本作 n 次测量的平均值的标准差 $s_{\bar{x}}$ 与单次测量结果的标准差 s 的关系为：

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad (4\text{-}3)$$

由此可见平均值的精密度（标准差）$s_{\bar{x}}$ 是单次测量精密度 s 的 $\frac{1}{\sqrt{n}}$；当测量次数增加时，平均值得标准差减小，这也说明平均值的精密度会随着测定次数的增加而提高。在分析化学领域里，一般平行测定 3 至 4 次即可。

对总体标准差，同样有

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad (1\text{-}27)$$

在后续的“中心极限定理”一章里将有更进一步的讨论

4.3 自由度（Degree of freedom）

自由度又称自由度数(number degree of freedom)。在统计学里，自由度是指在计算某一统计量时，取值不受限制的变量个数，符号为 v。在方差的计算中，自由度为和的项数减去对和的限制数。

例 4. 1 如一组 3 个数据：10，11，12 的平均值为 11。当均值 11 被确定后，如知道了 3 减 1 或 2 个数的值，第三个数的值就可确定。因此这里的自由度为 3 减 1，或 2。

例 4. 2 估计样本的方差是利用残差平方和，$\sum_{i=1}^{n}(x_i - \bar{x})^2$ 的，于是只要有（n-1）个数的离差平方和确定了，第 n 个数的值就不能变了，因为它受到均值的约束条件所限制。因此样本方差的自由度为 n-1.

例 4.3 设对于每一个自变量 x_i 都有一个因变量 y_i。若共有 n 个数据，则其一元线性回归方程为 $y=a+bx$，有 a 为截距，b 为斜率。由于有 a 和 b 的两个约束条件，一元线性回归的自由度为 n-2。