

单变量样本

假设数值 x_1, \dots, x_n 是重复实验的结果。根据 n 次实验的进行方式（在相同的初始条件下进行，而且没有前一次实验的“记忆”），我们可以合理地认为这 n 个数值是单变量样本 X_1, \dots, X_n 的实现；这些随机变量是独立同分布(i.i.d.)的。这样，统计模型的很大一部分已经确定。剩下的问题是：我们应该使用哪一个（边际）分布的集合？在接下来的描述性统计的部分讨论一些可以提供帮助的数值和图形方法。

分布的两个重要数值特征是位置和离散度。期望和中位数常用于表示分布的位置；当分布是对称时，它们相等。当分布向右（或向左）偏时，期望大于（或小于）中位数。为了根据观察值 x_1, \dots, x_n 得到分布位置的一个概念，我们可以使用样本均值或样本中位数。样本中位数是排序后的观察值序列中的中间值。

🔗 样本均值

样本 X_1, \dots, X_n 的样本均值是随机变量

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

分布的离散度可以通过方差（或标准差）和四分位距表示。四分位距是分布的上四分位数和下四分位数之间的距离。通过观察值 x_1, \dots, x_n ，我们可以计算样本的四分位距和样本方差，以获得离散度的概念。样本四分位距是数据的上四分位数和下四分位数之间的距离。

🔗 样本方差

样本 X_1, \dots, X_n 的样本方差是随机变量

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

注意样本方差这里分母上是 $\frac{1}{n-1}$ 在实际中，当分布看起来对称时，通常使用观察到的样本的均值和观察到的样本的标准差。对于非对称分布，优先使用观察到的样本中位数和观察到的样本四分位距。当然确定分布是否对称的最佳方法是使用统计图形方法。