

标准差 (Standard Deviation) , 也称均方差 (mean square error) , 是各数据偏离平均数的距离的平均数, 它是离均差平方和平均后的方根, 用 σ 表示。标准差是方差的[算术平方根](#)。标准差能反映一个数据集的离散程度。平均数相同的, 标准差未必相同。标准差可以反映平均数不能反映出的东西 (比如稳定度等) 。

样本标准差

在真实世界中, 除非在某些特殊情况下, 找到一个总体的真实的标准差是不现实的。大多数情况下, 总体标准差是通过随机抽取一定量的样本并计算样本标准差估计的。



从一大组数值当中取出一样本数值组合, 常定义其样本标准差:

样本方差 s 是对总体方差 σ 的无偏估计。 s 中分母为 $n-1$ 是因为 的自由度为 $n-1$, 这是由于存在约束条件。

这里示范如何计算一组数的标准差。例如一群儿童年龄的数值为 { 5, 6, 8, 9 } :

第一步, 计算平均值

$(A1+A2+.....+An)/n$

在这里, 5, 6, 8, 9 的平均值为 $(5+6+8+9) / 4 = 7$

第二步, 计算标准差

标准差 $\sigma = \sqrt{0.25 * \{ (5-7) * (5-7) + (6-7) * (6-7) + (8-7) * (8-7) + (9-7) * (9-7) \}}$

$= \sqrt{10 / 4} = 1.58$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

$n = 4$ (因为集合里有 4 个数), 分别设为:

$$\begin{aligned} x_1 &= 5 \\ x_2 &= 6 \\ x_3 &= 8 \\ x_4 &= 9 \end{aligned}$$
$$\bar{x} = \frac{1}{4} \sum_{i=1}^4 x_i \quad \text{用 4 取代 } N$$
$$\bar{x} = \frac{1}{4} (x_1 + x_2 + x_3 + x_4)$$
$$\bar{x} = \frac{1}{4} (5 + 6 + 8 + 9)$$
$$\bar{x} = 7 \quad \text{此为平均值。}$$

标准差