

一、百度百科上方差是这样定义的：

(variance)是在概率论和统计方差衡量随机变量或一组数据时离散程度的度量。概率论中方差用来度量随机变量和其数学期望（即均值）之间的偏离程度。统计中的方差（样本方差）是各个数据分别与其平均数之差的平方的和的平均数。在许多实际问题中，研究方差即偏离程度有着重要意义。

看这么一段文字可能有些绕，那就先从公式入手，

对于一组随机变量或者统计数据，其期望值我们由 $E(X)$ 表示，即随机变量或统计数据的均值，

然后对各个数据与均值的差的平方求和

$$\sum [X - E(X)]^2$$

这里写图片描述

，最后对它们再求期望值就得到了方差公式。

$$D(x) = E \left\{ \sum [X - E(X)]^2 \right\}$$

这里写图片描述

这个公式描述了随机变量或统计数据与均值的偏离程度。

二、方差与标准差之间的关系就比较简单了

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

这里写图片描述

根号里的内容就是我们刚提到的

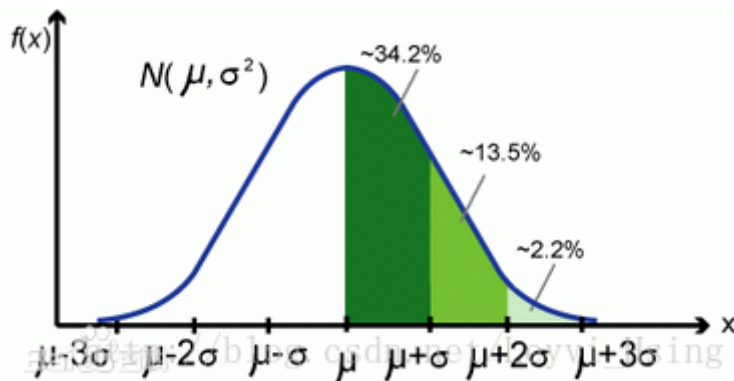
$$D(x) = E\left\{\sum [X - E(X)]^2\right\}$$

这里写图片描述

那么问题来了，既然有了方差来描述变量与均值的偏离程度，那又搞出来个标准差干什么呢？

发现没有，方差与我们要处理的数据的量纲是不一致的，虽然能很好的描述数据与均值的偏离程度，但是处理结果是不符合我们的直观思维的。

举个例子：一个班级里有60个学生，平均成绩是70分，标准差是9，方差是81，成绩服从正态分布，那么我们通过方差不能直观的确定班级学生与均值到底偏离了多少分，通过标准差我们就很直观的得到学生成绩分布在[61,79]范围的概率为0.6826，即约等于下图中的34.2%*2



这里写图片描述

三、均方差、均方误差又是什么？

标准差（Standard Deviation），中文环境中又常称均方差，但不同于均方误差（mean squared error，均方误差是各数据偏离真实值的距离平方和的平均数，也即误差平方和的平均数，计算公式形式上接近方差，它的开方叫均方根误差，均方根误差才和标准差形式上接近），标准差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。

从上面定义我们可以得到以下几点：

- 1、均方差就是标准差，标准差就是均方差
- 2、均方误差不同于均方误差
- 3、均方误差是各数据偏离真实值的距离平方和的平均数

举个例子：我们要测量房间里的温度，很遗憾我们的温度计精度不高，所以需要测量5次，得到一组数据 $[x_1, x_2, x_3, x_4, x_5]$ ，假设温度的真实值是 x ，数据与真实

值的误差 $e = x - x_i$

那么均方误差MSE=

$$\frac{\sum \varepsilon_i^2}{n}$$

这里写图片描述

总的来说，均方差是数据序列与均值的关系，而均方误差是数据序列与真实值之间的关系，所以我们只需要搞清楚真实值和均值之间的关系就行了。