方差、标准差、均方差、均方误差（MSE）区别总结

- 期望

期望是衡量某一随机变量其分布的平均值。
1.对于离散型随机变量X的分布律为

$$P\{X=x_k\}=p_k, k=1,2,3...$$

若级数

$$\sum_{k=1}^{\infty}x_kp_k$$

绝对收敛，那么称$\sum_{k=1}^{\infty}x_kp_k$为随机变量$X$的数学期望并记为$E(x)$,即

$$E(x)=\sum_{k=1}^{\infty}x_kp_k$$

同样的，设连续型随机变量X的概率密度为$f(x)$,若积分

$$\int_{-\infty}^{\infty}xf(x)dx$$

绝对收敛，则称积分$\int_{-\infty}^{\infty}xf(x)dx$的值为随机变量$X$的数学期望并记为$E(x)$即

$$E(x)=\int_{-\infty}^{\infty}xf(x)dx$$

数学期望又称为 **均值**
当随机变量X是离散型时候(发生的事件可以穷举完)，某一事件x_k发生的概率p_k可以替换成它发生的次数和事件总数的比，这时候

$$E(X)=\frac{1}{N}\sum_{k=1}^{\infty}x_k(\sum_{i=1}^{N}1(x_k发生))$$

总的来说就是x_k乘上它发生的次数再求和再除以总数。

- 方差

方差是用来度量随机变量与其均值$E(X)$的偏离程度，本来是由

$$E\{|X - E(X)|\}$$

表示,但是由于上式带有绝对值，不方便运算，为了方便，用

$$D(X) = E\{[X - E(X)]^2\}$$

来度量随机变量X与其均值$E(X)$的偏离程度。
按照定义，随机变量X的方差表达了X的取值与其数学期望的偏离程度，若$D(X)$较小意味着X的取值比较集中在$E(X)$附近，反之，若$D(X)$较大则表示X的取值比较分散，因此$D(X)$是刻画X取值分散程度的一个尺度。
对于离散型随机变量X有：

$$D(X) = \sum_{k=1}^{\infty}[x_k - E(X)]^2 p_k$$

对于连续型随机变量有：

$$D(X) = \int_{-\infty}^{\infty}[x_k - E(X)]^2 f(x)dx$$

其中$f(x)$为X的概率密度.
一般地，随机变量X的方差可按下列公式计算

$$D(X) = E(X^2) - [E(X)]^2$$

对于离散型一般的，方差$D(X)$等于各个数据与均值的差的平方和再除以事件总数N

$$D(X) = \frac{1}{N}\sum_{k=1}^{N}(x_k - E(X))^2$$

- 标准差

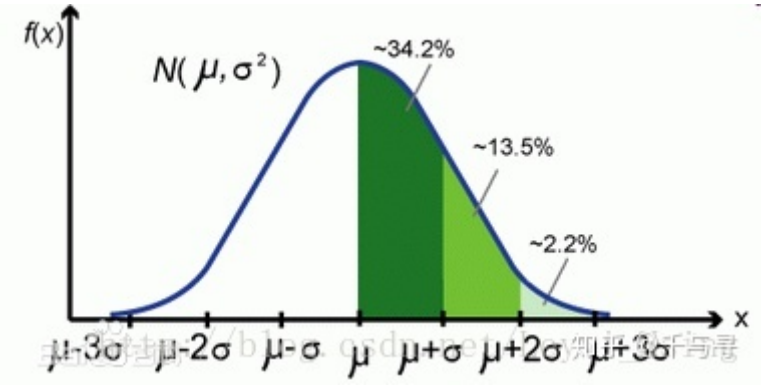
$$\sigma = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_k - E(X))^2}$$

标准差是方差的平方根。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？

原因是: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。即方差是平方的，不在一个数量级上，标准差和每个数据同属于一个数量级。

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，假设成绩服从正态分布，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68%，即约等于下图中的34.2%*2

额外说明：一个标准差约为 68%（平均值-标准差，平均值+标准差）， 两个标准差约为95%（平均值-2倍标准差，平均值+2倍标准差）, 三个标准差约为99%。它反映组内个体间的离散程度。



- 均方误差**(MSE)**

$$\frac{1}{N}\sum_{k=1}^{N}(y_{test}^{(k)} - y)^2$$

标准差（Standard Deviation），又称均方差，但不同于均方误差（mean squared error），均方误差是**各数据偏离真实值差值的平方和的平均数**，也就是误差平方和的平均数。均方误差的开方叫均方根误差，均方根误差才和标准差形式上接近。

- 总结

1、**均方差**就是标准差，**标准差**就是均方差。

2、**方差**是各数据偏离平均值的差值的平方和的平均数。

3、**均方误差(MSE**是各数据偏离真实值的差值的平方和的平均数。

总的来说，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，所以我们只需注意区分 真实值和均值 之间的关系就行了。

- 公式
  - 均值（期望）：$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
    注意这里的n是数据的总数，不是可能发生的事件的总数，注意辨别这里针对所有数据，上面介绍针对的是所有可能发生事件的个数乘以它发生的概率。
  - 方差：$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
  - 标准差：$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$
  - 均方误差：$\frac{1}{n}\sum_{i=1}^{n}(y_{test}^{(i)} - y)^2$

均值（期望）描述的是样本集合的中间点（平均值），但是它告诉我们的信息是有限的，而标准差给我们描述的是样本集合的各个样本点到均值的距离之平均。

> 以这两个集合为例，[0, 8, 12, 20]和[8, 9, 11, 12]，两个集合的均值都是10，但显然两个集合的差别是很大的，计算两者的标准差，前者是8.3后者是1.8。标准差小的距离均值较为集中。标准差描述的就是这种“散布度”。
> ps：之所以除以n-1而不是n，是因为这样能使我们以较小的样本集更好地逼近总体的标准差，即统计上所谓的“无偏估计”。而方差则仅仅是标准差的平方

参考：

- https://zhuanlan.zhihu.com/p/83410946
- https://zhuanlan.zhihu.com/p/86181679