



## 표준편차

통계용어 중 하나인 표준편차(Standard Deviation)는 어떤 데이터 집단의 퍼져있는 정도를 나타내는 수치로, 표준편차가 작은 데이터 집단은 모든 데이터가 평균을 중심으로 모여 있음을 의미한다.

The standard deviation(represented by sigma  $\sigma$  or  $s$ ) is a measure that is used to quantify the amount of variation or dispersion of a set of data values.

다음은 A, B사에서 생산한 Mxx 볼트 각 5개를 동일한 토크로 체결한 후 축력을 측정한 결과다.

구분	시료 번호	축력	평균	편차	(편차) <sup>2</sup>	분산	표준편차
A사 볼트	1	60	80	-20	400	200	약 14.1
	2	70		-10	100		
	3	80		0	0		
	4	90		10	100		
	5	100		20	400		
B사 볼트	1	70	80	-10	100	50	약 7.1
	2	75		-5	25		
	3	80		0	0		
	4	85		5	25		
	5	90		10	100		

만일 측정된 축력의 평균(Average 또는 Mean)으로 두 회사의 제품을 비교한다면 동일한 수준이라 할 수 있다. 하지만 표준편차를 비교하여 보면 두 회사의 볼트 수준에 차이가 있음을 알 수 있는데, B사의 볼트가 축력의 산포가 적으므로 품질이 보다 안정되었다고 판단할 수 있다.

편차(Deviation)는 각각의 변량 또는 변수(Variable)가 평균으로부터 떨어진 정도를 의미한다.

$$\text{편차} = \text{변수} - \text{평균}$$

분산(Variance)과 표준편차는 모든 변수가 전반적으로 평균으로부터 떨어진 정도를 의미하는데, 아래의 공식에 의해 산출된다.

$$\text{분산} = \frac{(\text{편차})^2 \text{의 합}}{\text{변수의 개수}}$$

$$\text{표준편차} = \sqrt{\text{분산}}$$

위의 예에서 A사의 경우 축력 평균이 80이고 표준편차가 14.1이므로 A사가 생산하는 Mxx 볼트의 대부분은 동일 토크하에서 66~94 정도의 축력을 보증한다고 예상할 수 있다.

만일 A사에서 Mxx 볼트를 매일 수백만개씩을 만든다고 하면, 위에서와 같이 몇 개의 샘플을 선정하여 축력을 측정한 후 그 볼트의 축력에 대한 표준편차를 구하여 품질의 산포를 예측할 수 밖에 없을 것이다. 이 경우 선정된 샘플의 개수 즉, 변수의 개수를  $n$ 이라 하면, 분산을 산출할 때  $n$  대신  $n-1$ 을 분모로 취한다. 따라서 분산은 250이며, 표준편차는 250의 제곱근인 약 15.8이 된다.

이렇게 큰 데이터 집단의 데이터 산포를 평가하기 위한 수단으로 일부 무작위적으로 선택된 데이터만을 이용하는 경우, 위와 같은 보정(Bessel's Correction)을 통해 표준편차를 산출하는데, 이때 보정을 거치지 않고 산출된 표준편차를 Sample Standard Deviation이라 구분하여 칭하기도 한다.