

평균 분산 표준편차

평균은 데이터를 나타내는 대푯값으로써 mean을쓴다.

$$m=a+b+c/n$$

문제 [1,2,3] 의 평균은?

$$1+2+3/3=2$$

모든값을 더한후 넘버로 나눈다 즉 데이터의개수

분산은 데이터가 얼마나 퍼져있는지 나타내는값

$$v=(a-m)**2+(b-m)**2+(c-m)**2/n$$

왜 제곱을 나타내냐? 제곱이없으면 뭐든지 0이되기 때문에 제곱을한다.

참고로 그값에서 평균을빼준값은 편차라한다.

얼마나 퍼져있는지 그래서 분산은 편차제곱의 평균이라고도한다.

분산에도 문제가발생을한다. 제곱을쓰면 제곱이된다.

그러면 데이터의 크기가커진다.

그래서 루트를 씌운다.

표준편차(std): 데이터와 비슷한트기로 분산을 변환

문제) [1,2,3] 의 분산 표준편차

$$(1-2)**2+(2-2)**2+(3-2)**2/3$$

$$1+0+1/3=2/3$$

$$\sigma=\sqrt{2/3}$$

$m-\sigma$ and $m+\sigma$ 은 전체데이터의 68%이다.

$m-\sigma$ 은 평균에서 34% 작은범위 하위 16%

$m+\sigma$ 은 큰범위에서 34% 상위 16% 이다.