

평균 분산 표준편차

평균 그리고 분산 표준편차는 데이터 분석에서 많이 쓰입니다.

공학에서 중요한 역할을 하지요 대푯값이 평균으로 얼마나 퍼져있는지 분산 또는 표준편차가 됩니다.

a b c d 가 있다고 해볼까요

평균 구하기 == $\text{mean} = a + b + c + d / \text{데이터 개수}(n)$ 대푯값으로 평균을 많이 쓴다

분산은 variance으로 퍼져있기 전에 어떻게 구할까 생각이 드는데요

그렇게 사람들이 생각이 드는게 편차입니다 .

a-m b-m c-m d-m 으로 이것들의 제곱들의 합

$(a-m)^2 + (b-m)^2 + (c-m)^2 + (d-m)^2 / n$

그래서 분산은 편차의 제곱의 합을 데이터 개수로 나누는 것

그런데 이거 의문점이 뭐냐면 제곱이 되니까 스케일이 너무 커진다.

스케일이라는 것은 규모가 커진다 데이터 분석할 때 오류가 나는 거는

규모가 커진다는 것이다. 표준편차는 $\sigma = \sqrt{v}$ 으로 규모를 만져준다.

그러면 대푯값이 평균인거는 알겠습니다. 다만 도데체 얼마나 퍼져있는지는

어떻게 아나?

즉 평균에서 시그마를 하나 더하고 뺀 값이 있었다고 쳤을 때에

$m - \sigma < \leq m + \sigma$ 빈칸에 있는 값들이 전체 데이터에서 68퍼센트를 설명한다.

$m - 2\sigma < \leq m + 2\sigma$ 빈칸에 있는 값들은 전체 데이터에서 95퍼센트를 설명합니다.

```
[7]: import numpy as np

[9]: A=np.array([1,2,3,4])
A
[9]: array([1, 2, 3, 4])

* [11]: mean=(A[0]+A[1]+A[2]+A[3])/4
mean # 평균구하기
[11]: 2.5

* [13]: vari=((A[0]-mean)**2+(A[1]-mean)**2+(A[2]-mean)**2+(A[3]-mean)**2)/4
vari # 분산구하기
[13]: 1.25

* [15]: std=np.sqrt(vari)
std # 표준편차
[15]: 1.118033988749895

* [21]: mean=np.mean(A) # 줄여서 평균과 분산 표준편차
vari=np.var(A)
std=np.std(A)
```

< 평균 분산 표준편차 코드 >