* **들어가기 전에**



 = 표준편차

는 기대값 입니다

기대값의 정의는

[확률 변수의 기댓값(期待값, 영어: expected value,{\displaystyle \operatorname {E} }\operatorname {E})은 각 사건이 벌어졌을 때의 이득과 그 사건이 벌어질 확률을 곱한 것을 전체 사건에 대해 합한 값이다] 인데 간단하게

주사위를 생각하시면 주사위의 기대값은

입니다

마찬가지로 위의 표에서

이고

밑에서 나올 것인데 E[X^2]은

 가 되는 것입니다

|  | 임의의 국가 |
| --- | --- |
| Co2 | A |
| GDP | B |

1. **일반적인 표준 편차 (Numpy의 std())**

위의 표를 표준화 하면 A는



B는

로 바뀝니다

여기서 표준편차  의 정의



로부터 표준편차를 구하면

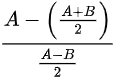






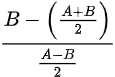
가 됩니다

그러면 표준화의 결과는

가 되고

이것은 1 입니다

B의 경우는

이 되고

이것은 -1이 될 것입니다.

이렇게 변량이 두개일때 표준화를 하면 무조건 1과 -1이 되는 것을 확인 할 수 있습니다

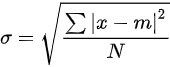
1. **판다스의 표준편차(베셀보정)**

그런데 그러면 아까 0.707107 은 무엇인가? 하면

판다스의 표준편차는 베셀 보정이라는 것을 사용하는데 이 방식은

표준편차를 구할 때 사용하는 모집단의 크기 N 대신에 N-1을 사용한다고 합니다

모집단의 크기에 의한 차이를 확인하기 위하여 또 다른 표준편차 정의를 가져 왔습니다.



이 표준편차의 정의에는 모집단의 크기 N이 들어가 있습니다.

판다스에서 표준편차를 구할때는 기본값으로 N이 N-1로 들어가 있다는 의미입니다

그러면 지금 위의 표와 같이 모집단의 크기가 2인 상황에서 판다스 표준편차를 구하면

N = 1이고

이 됩니다

이것으로부터 A가 B보다 크다고 가정할때 표준편차는

이 되고

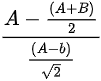
이것은

가 됩니다

그러면 표준편차  는

가 됩니다.

이것으로 표준화 계산을 해보면

가 되고 이것은

가 됩니다

그리고 이것이 0.707107 입니다