



시험 점수를 너무 잘 받은 친구 또는
시험 점수를 너무 못 받은 친구 때문에,
'평균'은 신뢰성이 낮은 대표값이 될 수 있다.

→ 평균과 중앙값의 차이가 많이 날 경우엔,
'중앙값'이 신뢰성이 높은 대표값이 될 수
있다. (즉, 일반적으로 중앙값이 '평균'과 신뢰성이 높은 '대표값'으로 여겨진다.)

→ 극적인 데이터들의 분포가 대칭 모형이 아닌 (즉,
'평균분포' 모형이 아닌), '평균'은 신뢰성이 낮은
대표값이 된다. 이에 따라 일반적인
'점점 통계량' (표본평균을 통해 구해지는 통계량)
도 사용할 수 있게 된다.

평균, 중앙값, 최빈값의 장단점

대푯값에서 평균과 중앙값, 최빈값을 알아봤는데, 각각이 어떤 장단점이 있는지 알아야겠죠? 어떤 자료들의 특징을 대표할 때 어떤 값을 사용하는 것이 대표성을 가장 잘 나타내는지 말이에요.

평균은 모든 자료의 값을 다 이용한다는 장점이 있어요.

중앙값은 1, 1, 1, 2, 2, 2, 100처럼 자료의 값 중 어느 하나가 너무 크거나 너무 작을 때 자료의 특징을 잘 대표할 수 있어요.

최빈값은 가장 많이 발생하는 값을 구할 때 유용하고, 특히 자료가 숫자가 아니어도 사용할 수 있지요. 대신 최빈값은 없을 수도 있고, 2개 이상일 수도 있다는 단점이 있어요.

다음 자료의 평균, 중앙값, 최빈값을 구하여라.

19, 20, 21, 19, 26

평균 = $(19 + 20 + 21 + 19 + 26) \div 5 = 21$

중앙값을 구하기 위해서 작은 거부터 순서대로 써보죠. 19, 19, 20, 21, 26이네요. 전체 자료의 수가 5로 홀수 개니까 $(n + 1) \div 2 = 3$ 번째 값인 20이 중앙값입니다.

최빈값은 도수가 가장 높은 값이에요. 19의 도수는 2, 나머지 20, 21, 26의 도수는 1이니까 도수가 2인 19가 모두 최빈값이라고 할 수 있겠네요.