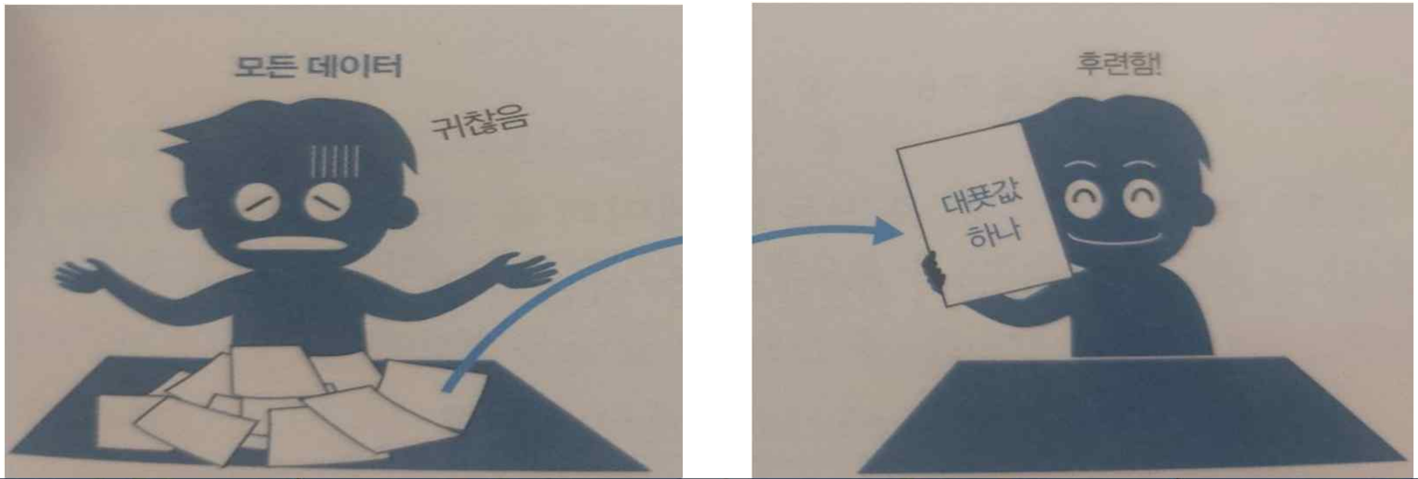
목차

1. 대푯값의 3가지 종류



B광고 7월	주문수	B광고 8월	주문수
1주차	55	1주차	50
2주차	50	2주차	45
3주차	45	3주차	50
4주차	50	4주차	55
7월 평균 주문수	74	8월 평균 주문수	50

대푯값이란?

데이터 전체의 특징을 하나의 수로 나타내는 값

자료의 중심 경향을 나타내어 전체 자료를 대표하는 값

평균

중앙값

최빈값

1) 평균

평균(average)은 어떤 값들의 집합의 적절한 특징을 나타내거나 요약하는 것을 의미한다.

표1 대푯값의 유형

일반적인 평균

복수개의 수치의 곱을 수치의 개수로 제곱근 취해서 산출되는 평균값
예) 성장률 평균

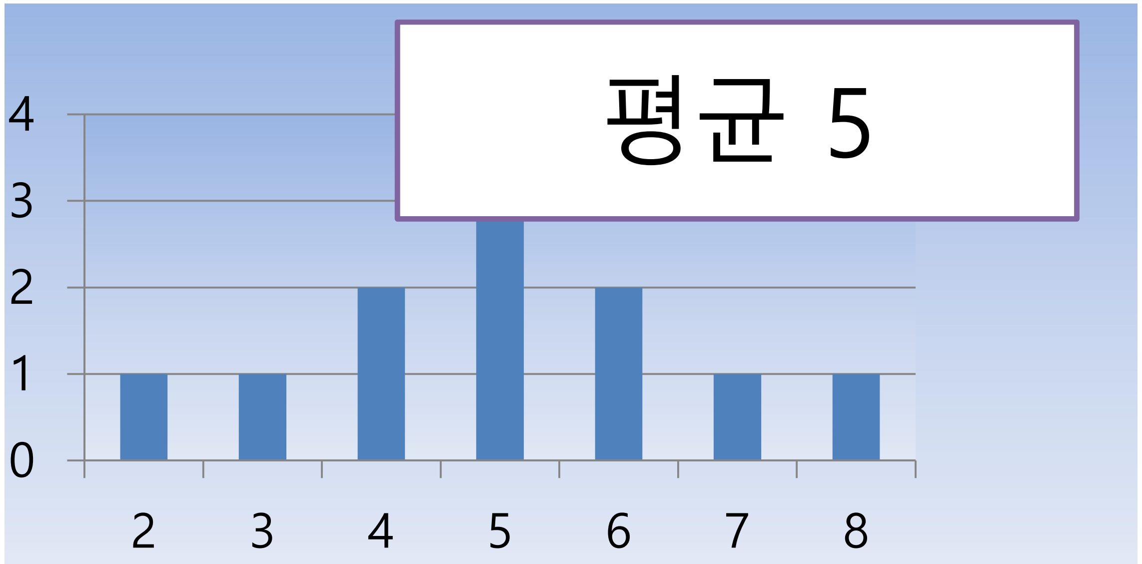
표2 a_1부터 a_n까지의 평균을 산출하는 세 방식

① 산술평균	$A=\dfrac{a_1+a_2+a_3\ \cdots\ +a_n}{n}$
② 기하평균	$G=\sqrt[n]{a_1\cdot a_2\cdot a_3\cdot\ \cdots\ \cdot a_n}$
③ 조화평균	$H=\dfrac{n}{\dfrac{1}{a_1}+\dfrac{1}{a_2}+\dfrac{1}{a_3}+\cdots+\dfrac{1}{a_n}}$

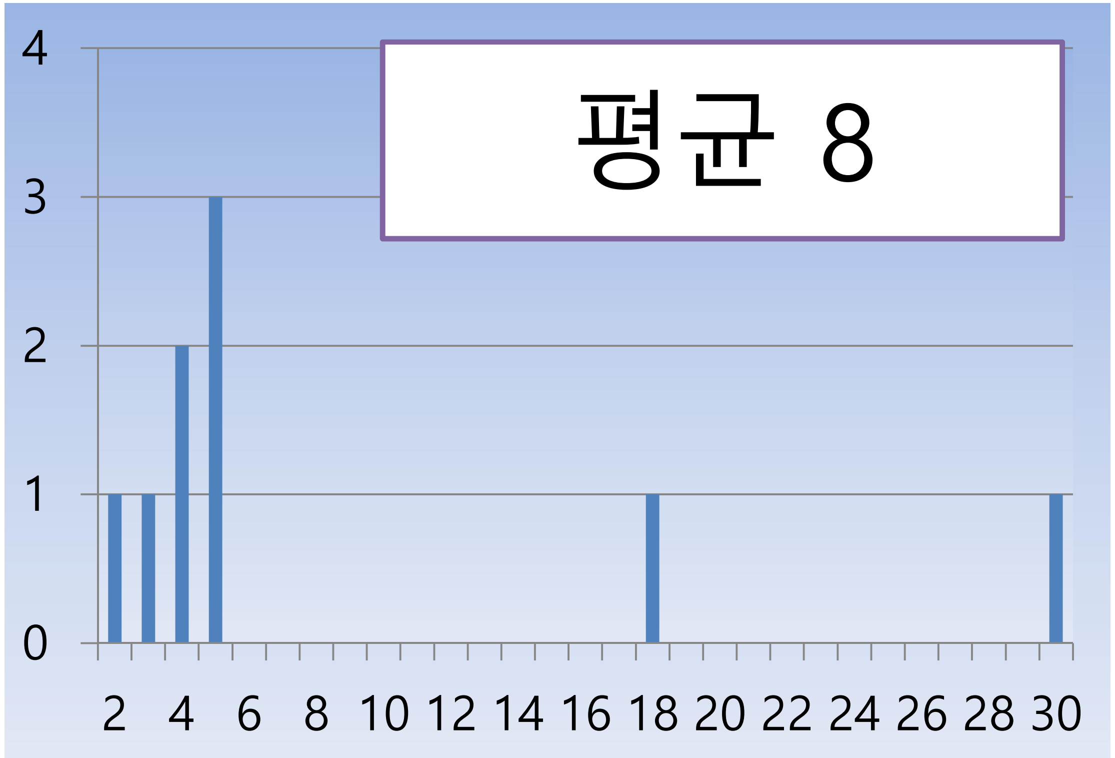
주어진 수들의 역수의 산술 평균의 역수
예) 속력 평균

평균 ≵ 중심 ?

23445556678



2344555661830



평균을 중심이라고 보기 보다는
균형을 맞추는 수라고 보는 게 맞다

2) 중앙값

강건한 대푯값이라고도 불림

(특잇값에 영향을 받지 않음)

중앙값(median) 또는 **중위수**는 어떤 주어진 값들을 크기의 순서대로 정렬했을 때 가장 중앙에 위치하는 값을 의미한다.

1) 홀수 일 때는 중간에 있는 값, 중앙값
2) 짝수 일 때 중간에 있는 두개의 값의 평균

1, 3, 3, **6**, 7, 8, 9

Median = **6**

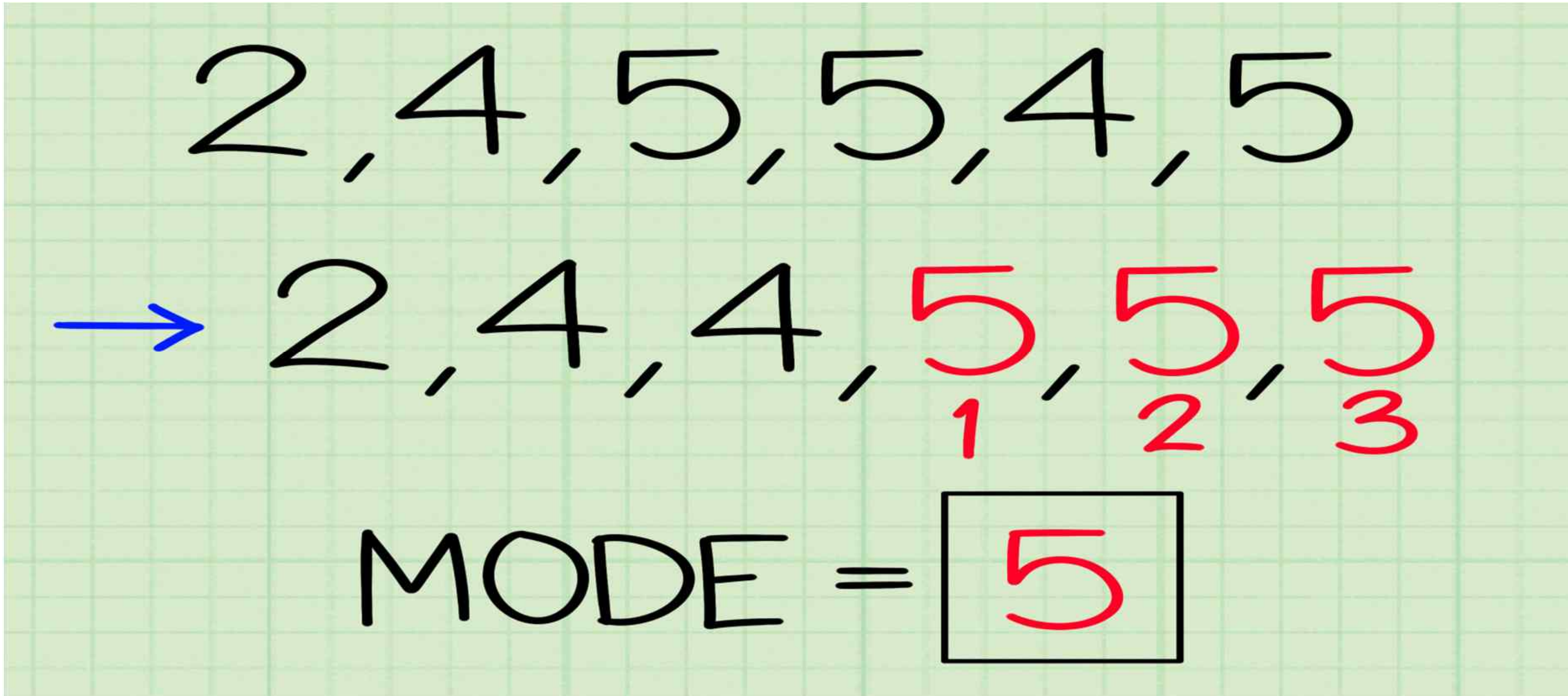
1, 2, 3, **4**, **5**, 6, 8, 9

Median = (4 + 5) ÷ 2

= **4.5**

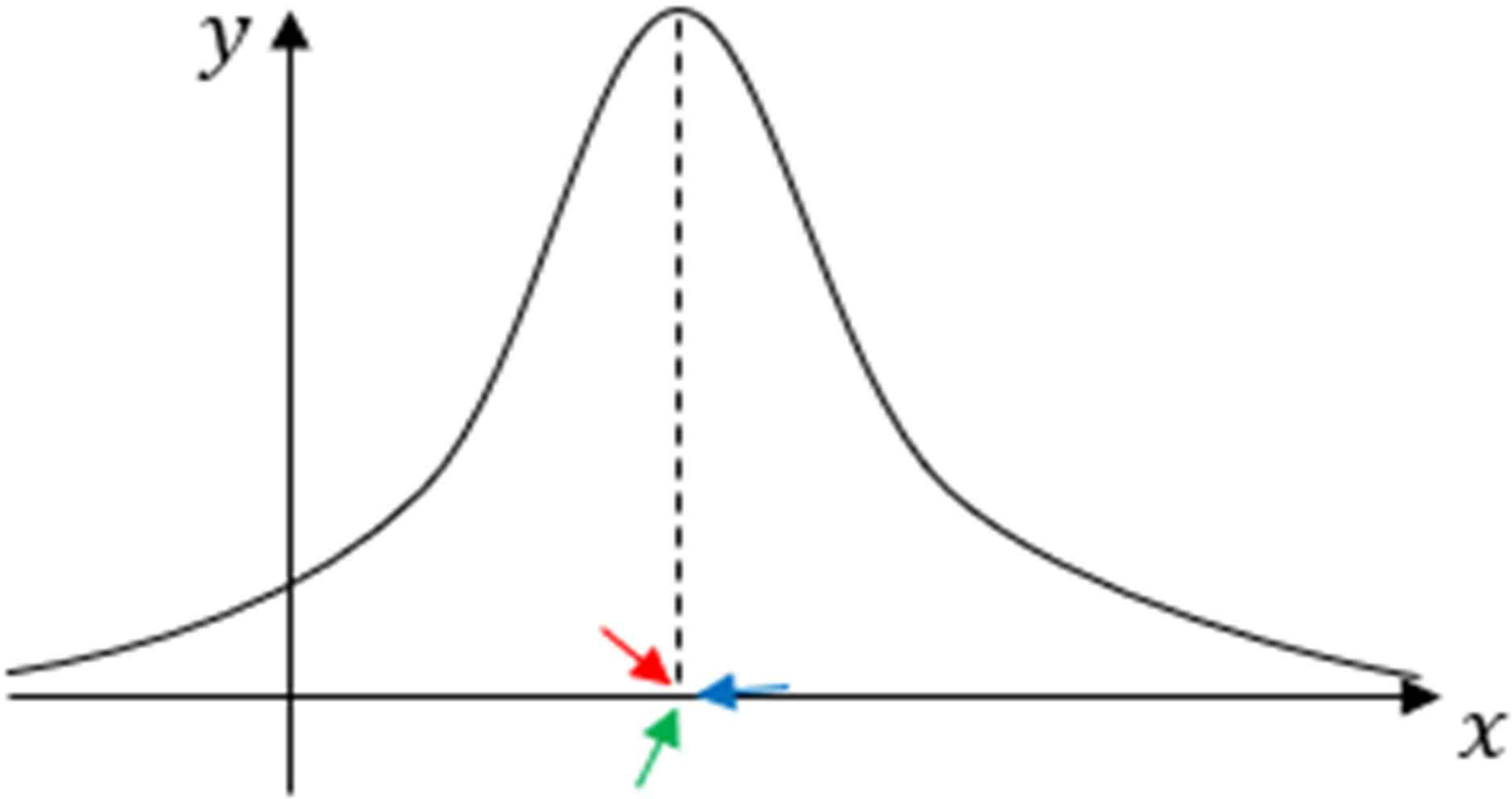
3) 최빈값

최빈값(最頻-), **모드**(mode)는 통계학 용어로, 가장 많이 관측되는 수, 즉 주어진 값 중에서 가장 자주 나오는 값이다.



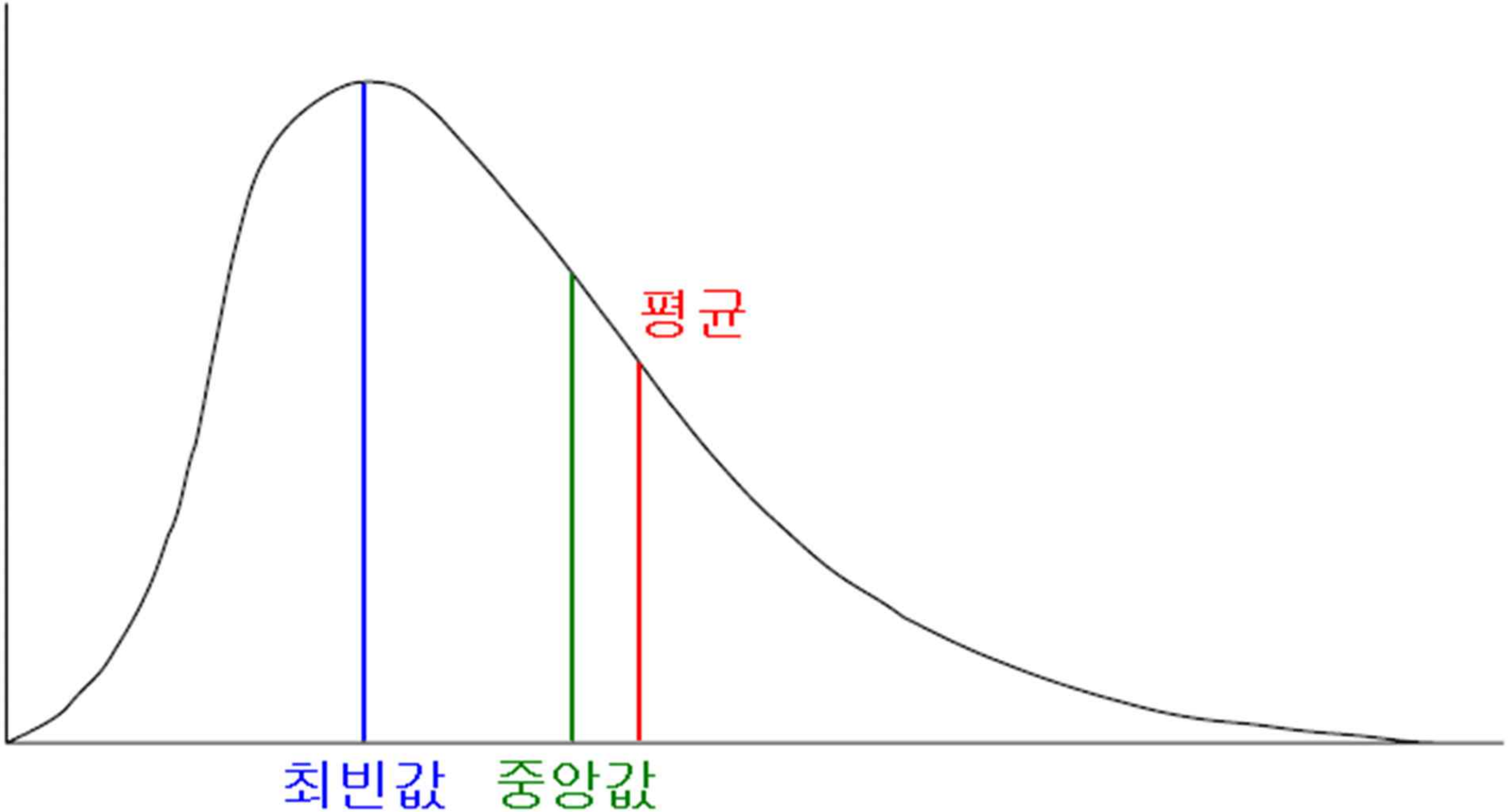
2. 위치 관계

1. 평균 = 중앙값 = 최빈값 정규분포 모양



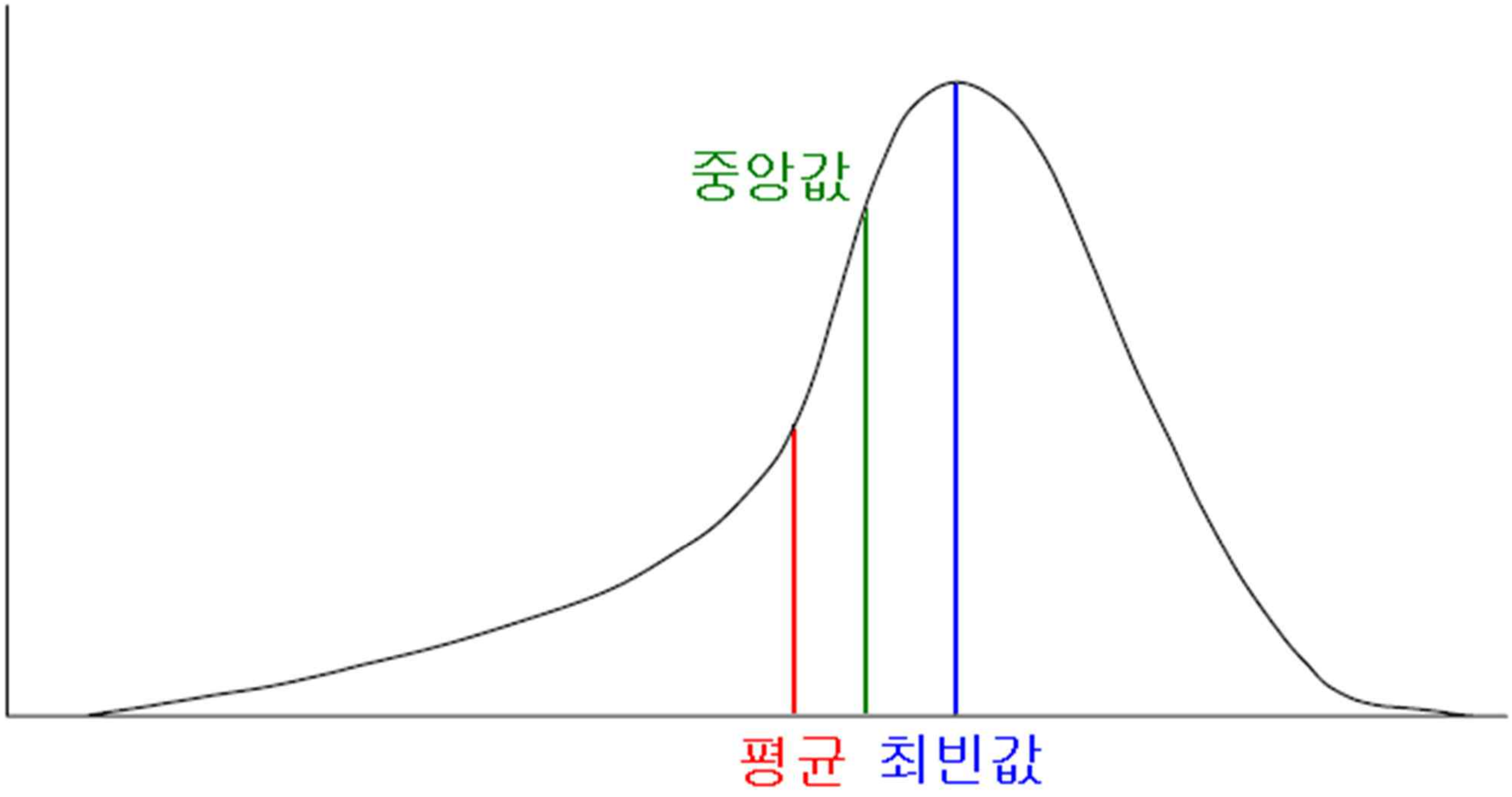
2. 위치관계

2. 평균 > 중앙값 > 최빈값 오른쪽으로 길게 늘어진 모양



2. 위치관계

3. 평균 < 중앙값 < 최빈값 왼쪽으로 길게 늘어진 모양



3. 산포도를 나타내는 사분위수

산포도란?

산포도(dispersion, variability, scatter, spread)는 데이터가 얼마나 퍼져있나를 나타내는 통계학 용어이다.

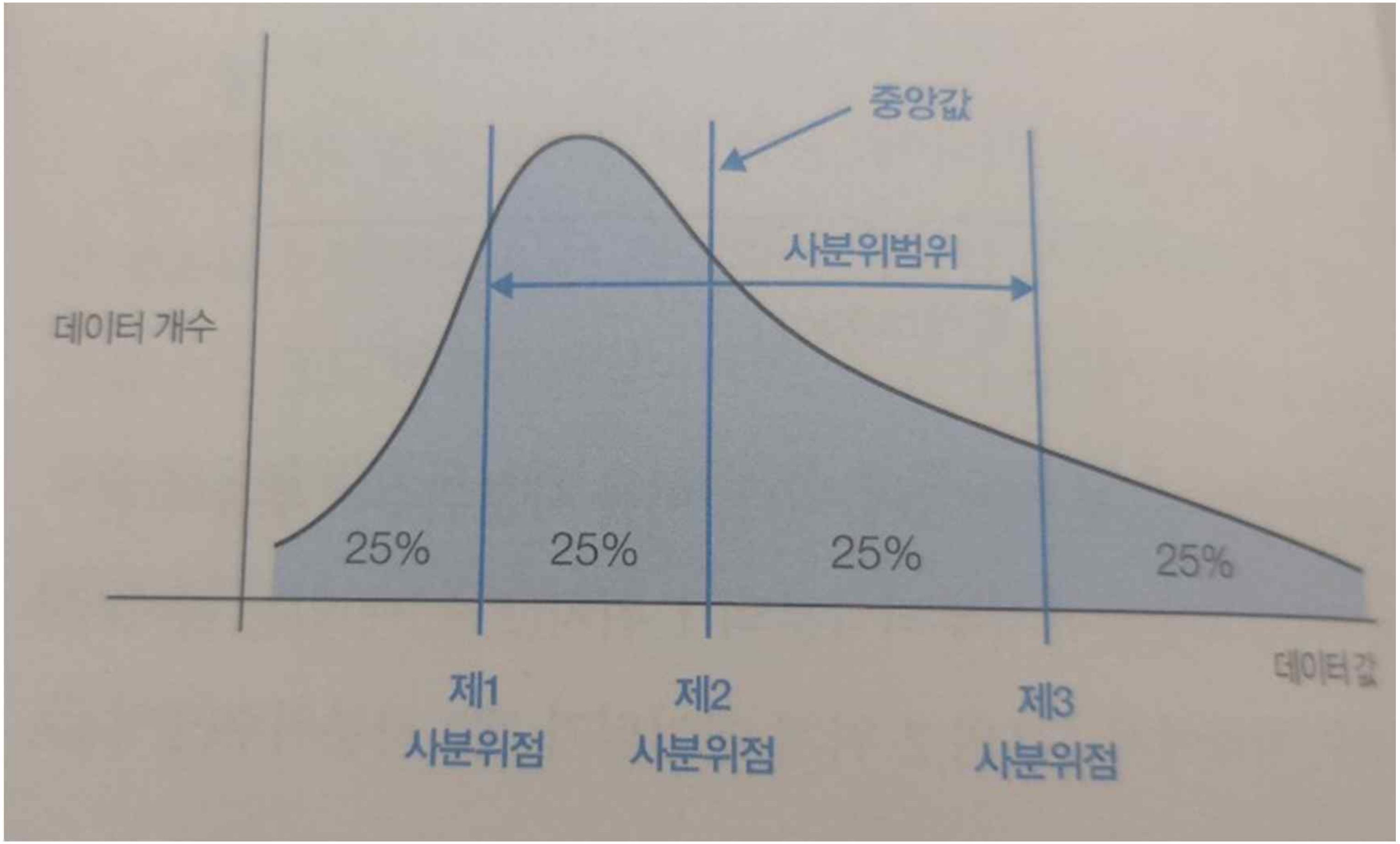
범위 = 최댓값 – 최솟값

제 1 사분위수 = 최솟값에서부터 ¼ 위치

제 2 사분위수 = 중앙값

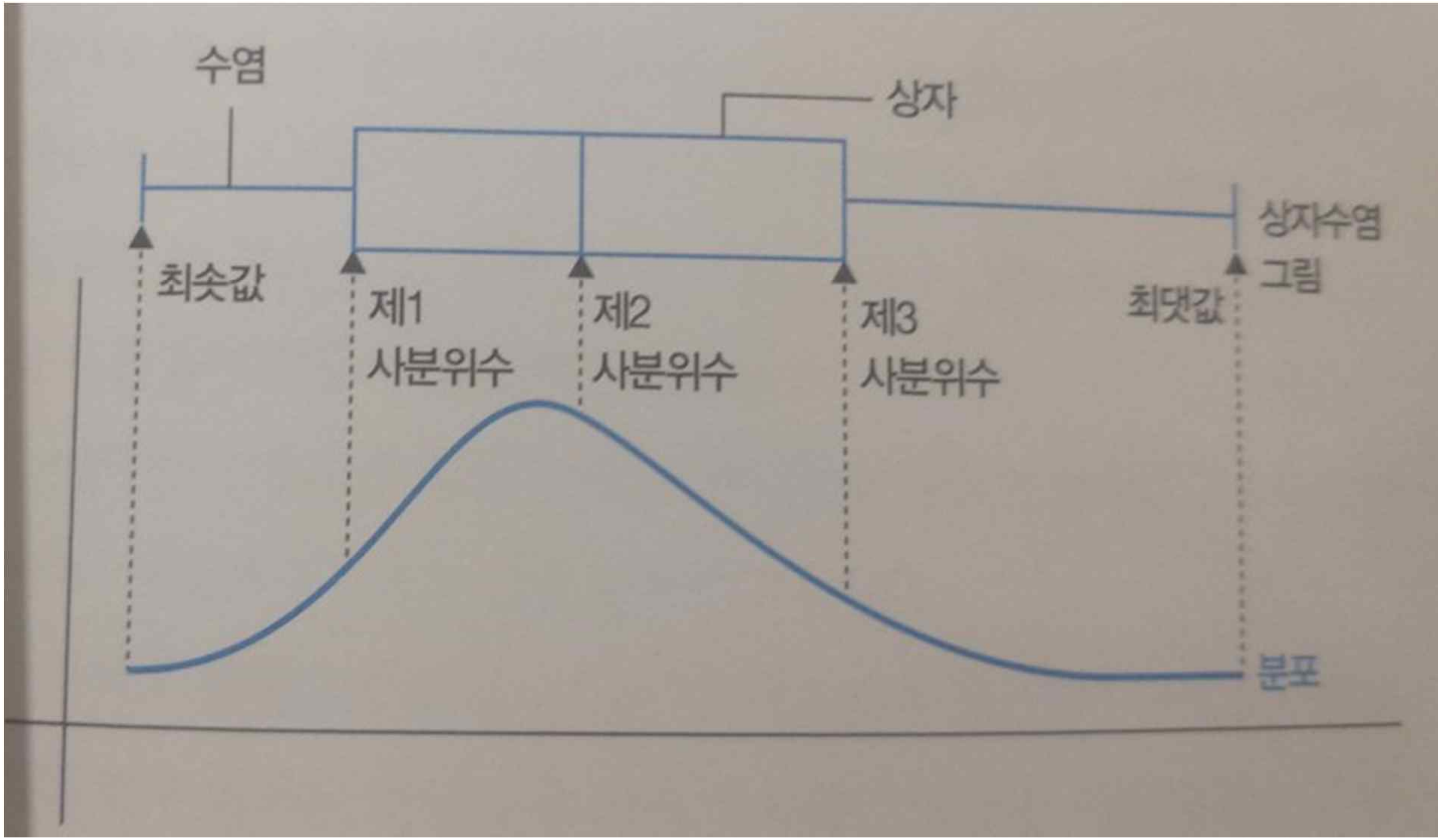
= 최솟값에서부터 2/4 위치

제 3 사분위수 = 최솟값에서부터 ¾ 위치



중앙값
사분위범위
데이터 개수
25%
25%
25%
25%
데이터 값
제1
사분위점
제2
사분위점
제3
사분위점

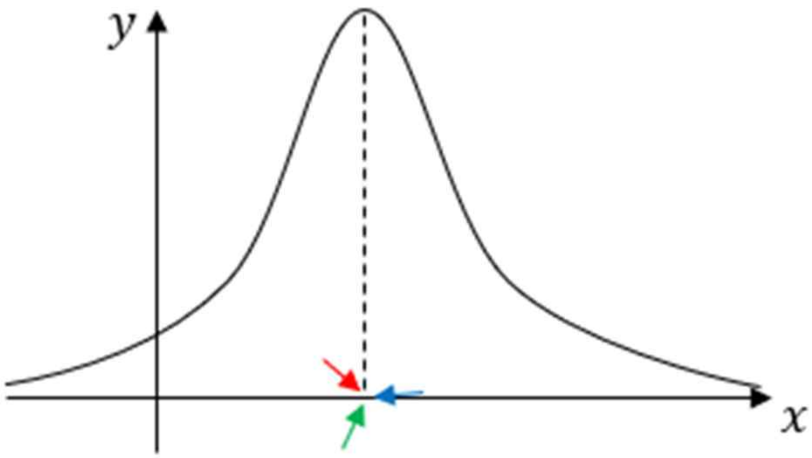
상자수염그림



-	국어	영어	수학	과학	사회
1	98	[illegible]			
2	94	[illegible]			
3	88	[illegible]			
4	75	[illegible]			
5	65	[illegible]			
6	62	[illegible]			
7	59	[illegible]			
8	55	[illegible]			
9	53	67	59	53	68
10	50	59			
11	41	56			
12	41	56			
13	38	48			
14	37	44			
15	32	40			
16	31	35			
17	26	31			
18	14	24			
19	10	23			
20	10	16			

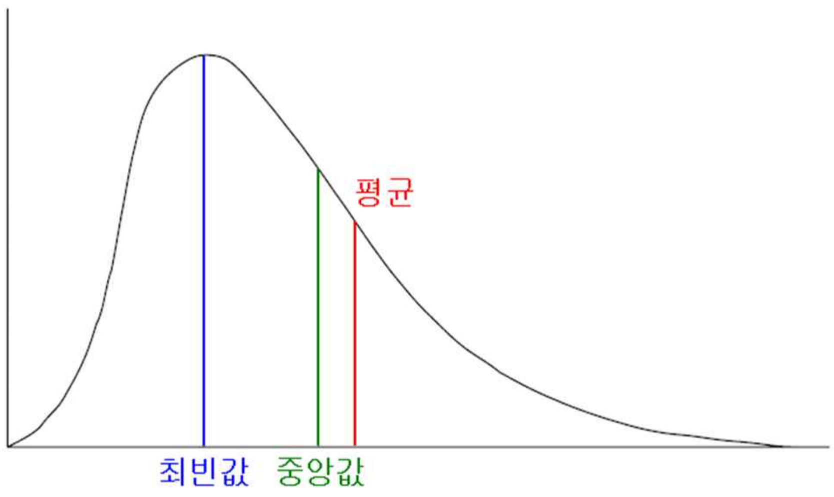
최대값	98	97	100	100	155
최소값	10	16	12	13	18
평균값	48.95	59.30	54.70	49.30	64.50
중앙값	45.50	57.50	50.50	51.00	65.50
Q3	62.75	80.25	69.00	65.50	83.50
Q1	31.75	38.75	40.25	26.25	36.25

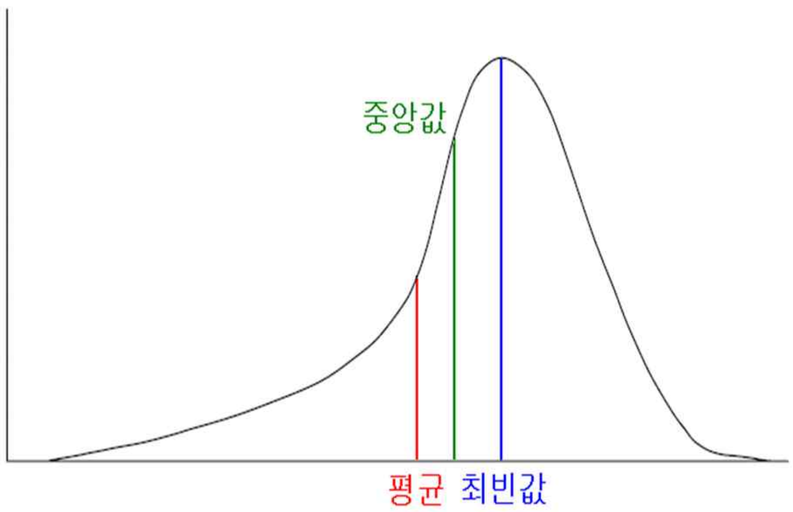
4. 분산



여기 이 그래프만 보면서
산포도를 비교하는건 힘들다

일정한 수치로 설명하면 설득력이 높아짐





4. 분산

1.편차

편차는 (각)데이터 – 평균 인데

한쪽으로 얼마나 치우쳐져 있는지를 알려준다

평균 = 균형 을 맞춘 것이므로

편차를 다 더하면 0이 됨

4. 분산

2. 편차의 절댓값으로 계산해보자

|편차|는 평균편차라고도 한다

그렇게 하면 1번의 문제점인 편차의 합이

0이 되지 않아서 비교가 가능하다

그러나

1)절댓값을 이용한 연산이 싫음

2)수학적으로 다루기 어렵다

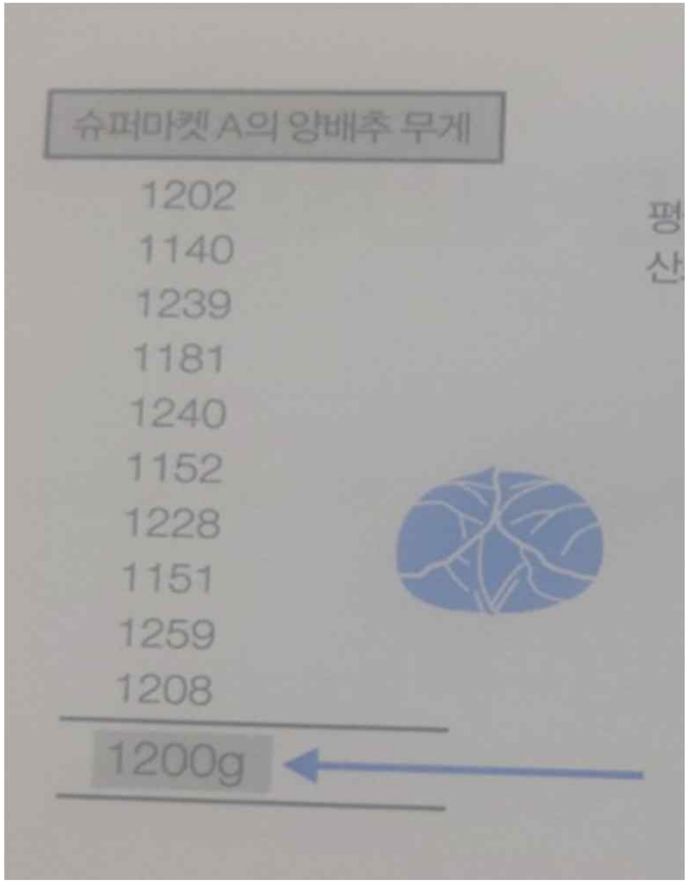
3)정규분포표를 사용할 때 분산(표준편차)가 더 편하다

3. 분산으로 이용

분산은 ‘수치의 지표’가 될 수 있다

양배추 10개 평균 1200g

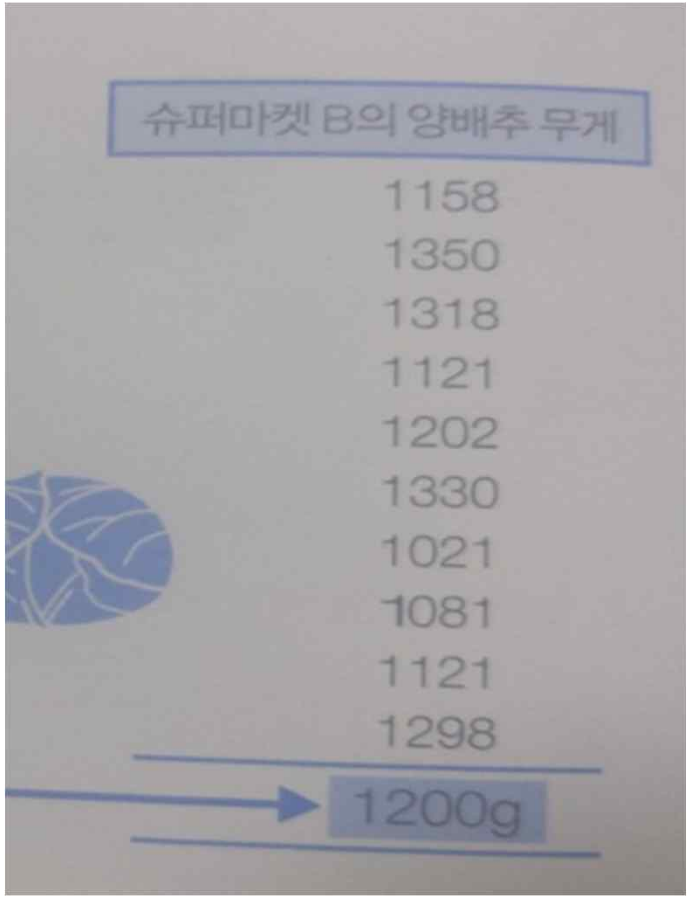
예제



슈퍼마켓 A의 양배추 무게

1202
1140
1239
1181
1240
1152
1228
1151
1259
1208

1200g

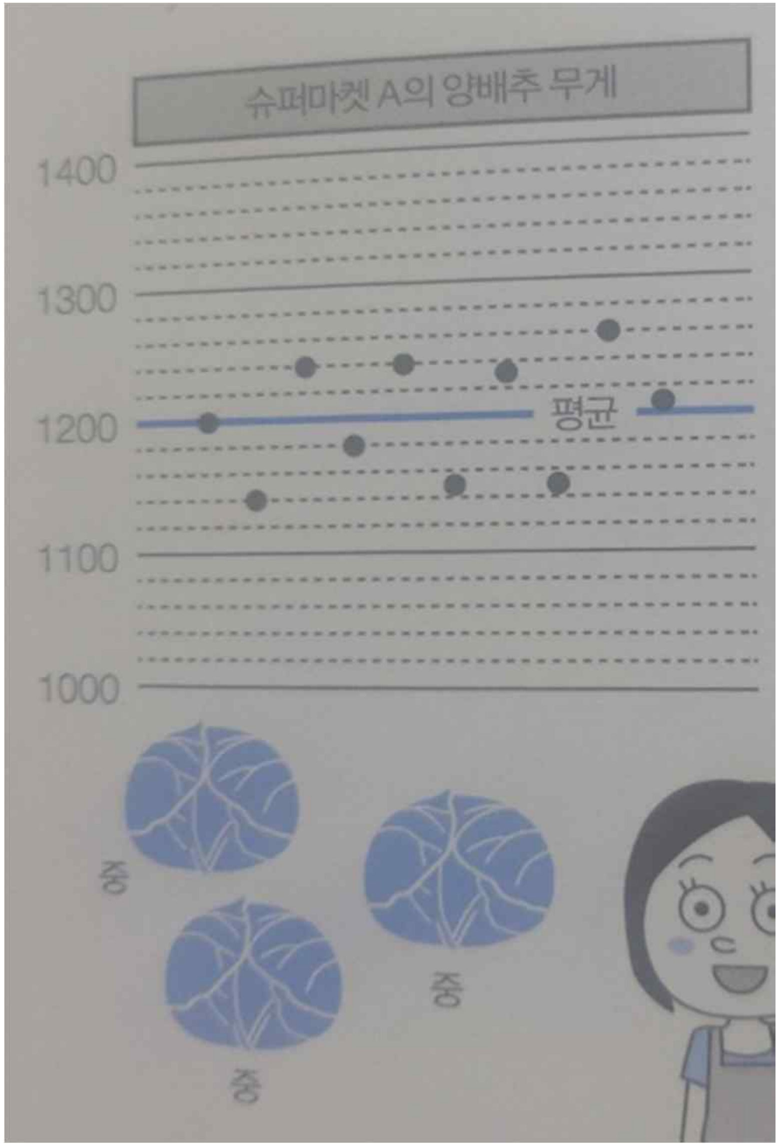


슈퍼마켓 B의 양배추 무게

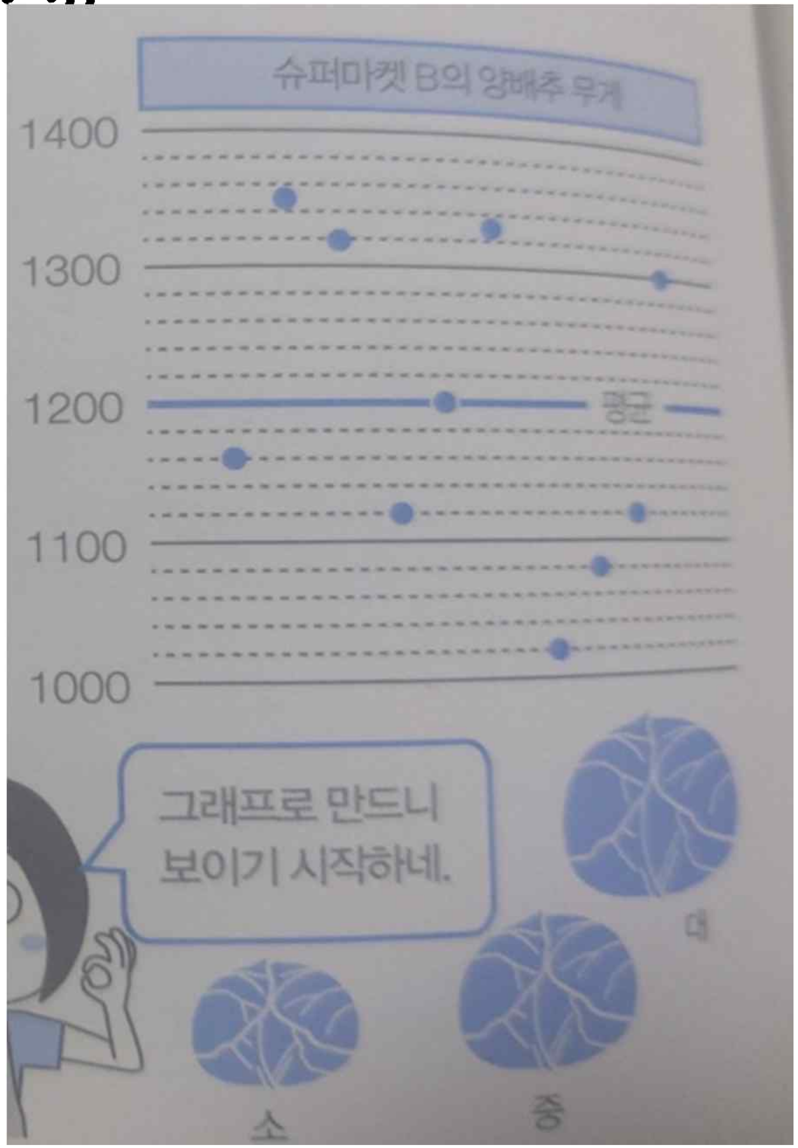
1158
1350
1318
1121
1202
1330
1021
1081
1121
1298

1200g

예제



슈퍼마켓 A의 양배추 무게
1400
1300
1200
1100
1000
평균
중
중
중



슈퍼마켓 B의 양배추 무게
1400
1300
1200
1100
1000
평균
그래프로 만드니
보이기 시작하네.
대
소
중

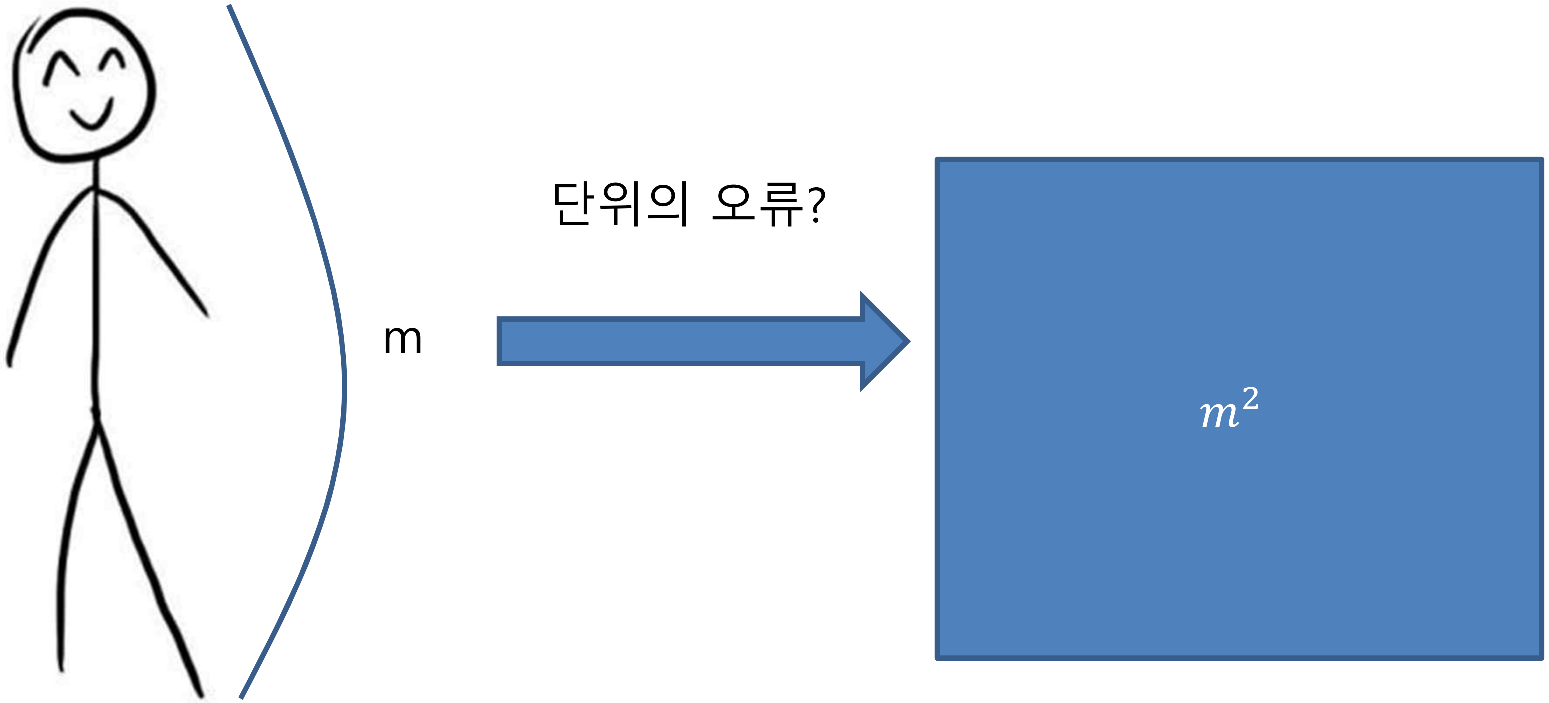
5. 분산의 단점

분산을 쓰면 크게 두가지 단점들이 존재한다.

1. 원래의 편차와 비교했을때 수가 너무 크다
2. 단위가 변함

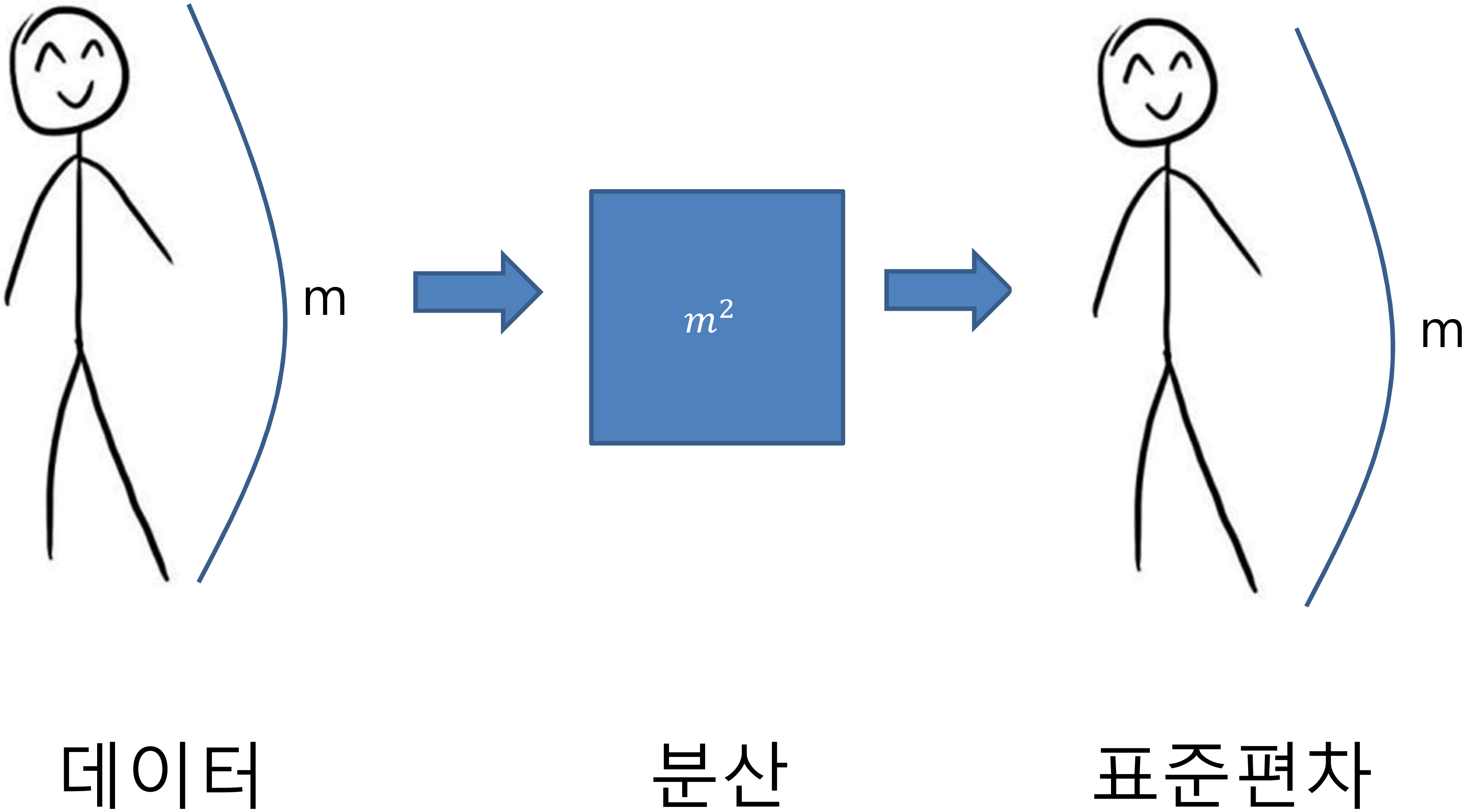
분산이 아닌 표준편차로 수치를 나타낸다

6. 표준편차



1.71m, 1.68m, 1.62m, 1.81m, 1.71m,
1.67m, 1.74m, 1.75m, 1.68m, 1.63m

6. 표준편차



끝

출처..

https://xlstory.tistory.com/m/entry/Excel-%EC%83%81%EC%9E%90-%EC%88%98%EC%97%BC-%EA%B7%B8%EB%A6%BC%EC%9C%BC%EB%A1%9C-%EC%82%B0%ED%8F%AC%EB%8F%84%EB%A5%BC-%EC%B8%A1%EC%A0%95

위키백과

http://app.goodgag.net/8989/