

주제: 평균 이해

(average)

평균의 종류

- ① 평균값 (mean)
- ② 중앙값 (median)
- ③ 최빈값 (mode)

P.88

평균값

예) 파워워터아웃 교실 참가자들의 나이  
19, 20, 20, 20, 21

질문: 참가자들 나이 평균값은?

답:  $\frac{19 + 20 + 20 + 20 + 21}{5}$  ↔ 나이의 총합  
참가자 수

P.90

일반화:

$$\text{평균값} = \frac{\sum X}{n}$$

$$= \mu$$

← 평균값을 나타내는 기호  
("뮤"라 읽음)

단, ①  $\sum X = x_1 + x_2 + \dots + x_n$

← 합을 나타내는 기호 (Sigma)

$x_i$ : i번째 대상을 가리키는 숫자.

(이 예제에서는 i번째 사람의 나이)

②  $n$  = 전체 조사대상 숫자.

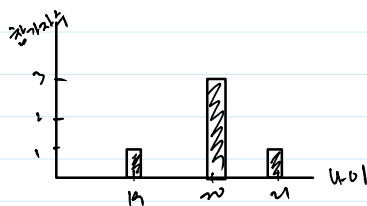
(이 예제에서는 표가교실 참가자 수)

P.91  
(상반)

예) 파워워터아웃 교실 참가자들의 나이를 평균으로 갖는 표 설정하기

나이	19	20	21
횟수	1	3	1

⇒ 막대 그래프 활용은 평균값 정보를 주지 않는다.



$$\text{평균 나이} (\mu) = \frac{19 \times 3 + 20 \times 6 + 21 \times 3 + 145 + 147}{5}$$

3인도 2인도  
 5인도 3인도

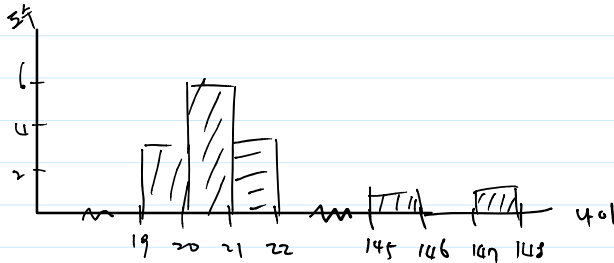
이상치

평균에 관해:

예) 공룡교실 참가자들 나이의 평균

나이	19	20	21	145	147
소수	3	6	3	1	1

⇒ (히스토그램)



$$\text{평균 연령} = \frac{19 \times 3 + 20 \times 6 + 21 \times 3 + 145 + 147}{14} = 38$$

문제점: 평균 연령이 38세이지만 대부분은 20대 초반.

즉 명의 두면의 오류.

50대 초반의 인간 참가자가 공룡교실은  
잘 따라갈 수 있을까?

평균만 보고 신청하면 매우 위험할 수 있음!

p.97~98

원인: 두 개의 이상치 존재!  
145와 147

두 개의 이상치가 평균값을 왜곡시켰음!

즉, **편향된 데이터**의 평균값은 **조심해야**  
사라진 값.

이상치 = 극단적인 값!

진문: 이상치를 고려하여 평균 정하는 방법

⇒ ① 이상치를 제거한 평균내기 (사본위수 등)

② 중앙값 활용 가능.

② 중앙값

p.101

정의: 가운데에 위치한 값.

예 1) 19 19 20 20 20 21 21 100 102

예 2) 19 20 20 20 21 21 10 102

두 수의  
평균값

p.102

(정렬)

중앙값 구하는 3단계

① 주어진  $N$  개의 데이터를 오름차순으로 정렬

②  $N$  이 홀수인 경우 : 중앙값 =  $\frac{N+1}{2}$  번째 데이터

③  $N$  이 짝수인 경우 :

중앙값 =  $\frac{N}{2}$  번째와  $\frac{N}{2} + 1$  번째 데이터의  
평균값.

p.104 포인팅된 데이터 평균값 vs. 중앙값

① 오른쪽으로 포인팅된 데이터

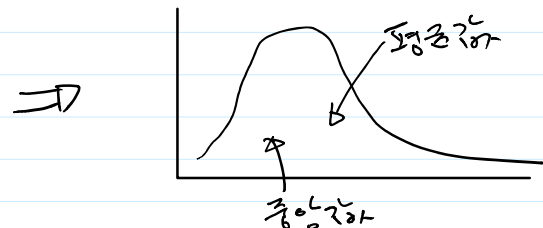
예)

값	1	2	3	4	5	6	7	8
도수	4	6	4	4	3	2	1	1

중앙값 : 3

평균값 :  $\frac{4 + 12 + 12 + 16 + 15 + 12 + 7 + 8}{25}$

$$= \frac{86}{25} = 3.44$$

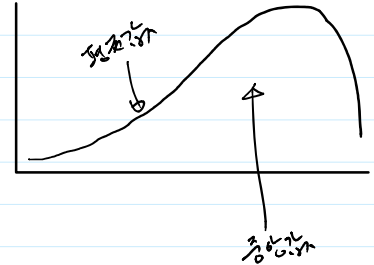


## ② 원형으로 편향된 데이터

예)

점수	1	4	6	8	9	10	11	12
횟수	1	1	2	3	4	4	5	5

⇒



중앙값: 10

$$\text{평균값} = \frac{1 + 4 + 12 + 24 + 36 + 40 + 55 + 60}{25} = 9.28$$

데이터가 왼쪽으로 편향되어 있음.

⇒ 평균값 < 중앙값

주의: 데이터가 오른쪽으로 편향되어 있을 때

⇒ 평균값 > 중앙값

P.108  
5

9.110

평균값과 중앙값의 관계

예) 부모와 함께 하는 아이 수명조사

나이	1	2	3	31	32	33
횟수	3	4	2	2	4	3

$$\text{① 평균값} = \frac{1 \times 3 + 2 \times 4 + 3 \times 2 + 31 \times 2 + 32 \times 4 + 33 \times 3}{18}$$

$$= 17$$

⇒ 질문: 참가자들의 평균 연령 17세?

전혀 고려하지  
않음!

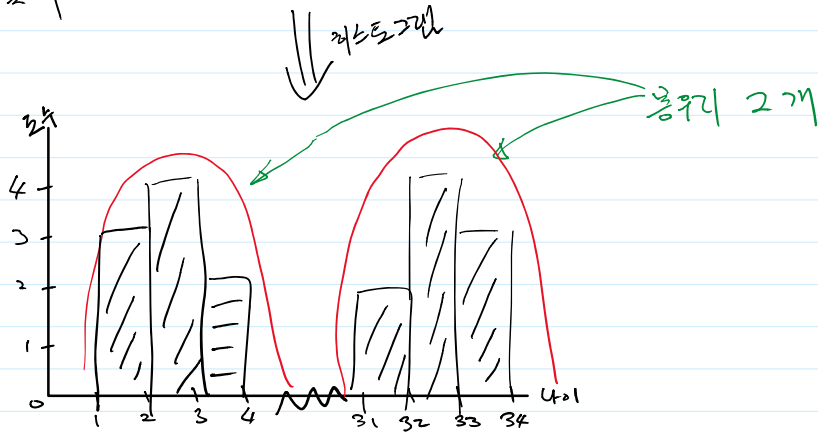
② 중앙값: 총 18명 따라서

중앙값은 아래서 9번째 값인 3과  
10번째 값인 31의

평균값인 17세이다.

⇒ 이 경우에도 17세가 모임을 선택하지 않는다는.

따라서, 이 경우에는 평균값, 중앙값 모두 전혀 선택성을 갖지 못한다.



⇒ 따라서 한 개의 나이가 아니라 두 개의 나이가 선택성을 갖는다.

↓  
최빈값!

가장 높은 도수를 갖는 값 또는 범위.

이런이 수명 2종의 경우 두 개의 최빈값 존재,  
부담 항제하는  
즉, 2 살과 32 살

P. 114

변수

①

값	1	2	3	4	5	6	7	8
도수	4	6	4	4	3	2	1	1

$$\text{평균값} = \frac{1+12+12+16+15+12+7+8}{25}$$

$$= \frac{86}{25} = 3.44$$

$$\text{중앙값} = 3$$

$$\text{최빈값} = 2$$

②

빈도	과방	방방	초록	분홍	노랑
도수	4	5	8	1	3

평균값 : X  
 중앙값 : X  
최빈값 : 3

③

간격	1	2	3	4	5
빈도	2	3	3	3	3

평균값  $\frac{2+6+9+12+15}{14} = \frac{44}{14} = 3.1...$

중앙값 : 3

최빈값 : 2, 3, 4, 5

↑  
 치이현상!

주의: 평균값이 중앙값은 수치관련 데이터에만  
 사용 가능.

P.118

요약 정리

평균	계산 방법	사용 가능 경
평균값 (M)	$\frac{\sum X}{n}$	데이터가 비교적 작고 대칭일 때
중앙값	① 모든 값을 오름차순으로 나열 ② 홀수 개수인 경우 중앙에 위치한 값 ③ 짝수인 경우 중앙에 위치한 두 수의 평균	데이터가 이상치로 인해 편향되었을 경우
최빈값	① 도수가 가장 높은 값 ② 두 종류 이상의 데이터가 동일한 도수를 가질 각 그룹에 대한 최빈값 선택	① 빈도적 데이터에서도 사용 가능 ② 데이터가 하나 이상의 데이터 그룹을 포함하고 있을 때.

P.119  
회사에 사람이 직원들의 연봉을 올려주어야 함.

방식 ① : 보수의 연봉을 동일하게 2,000원씩 올리기

방식 ② : 각자의 연봉을 10%씩 올리기

설정 : 연봉 평균액 = 50,000원

중상액 = 20,000원

최저액 = 10,000원

질문 a) 연봉을 동일하게 2,000원씩 올린 경우  
평균액, 중상액, 최저액을 계산하라.

평균액 = 52,000원

중상액 = 22,000원

최저액 = 12,000원

b) 각자의 연봉을 10% 올린 경우  
평균액, 중상액, 최저액을 계산하라

평균액 = 55,000원

중상액 = 22,000원

최저액 = 11,000원

c) 자신의 연봉이 평균액에 가까운 경우와  
최저액에 가까운 경우에 따라  
연봉 인상 순서 방식을 결정하는 수 있다.  
어떻게 결정하나?

- 중상액 이하인 경우 : 20% 인상

- " 이상인 " : 10% 인상