100문제를 푸는 것보다 1문제를 푸는 1가지 원리를 가르칩니다!

원리탐구 중등수학

중3수학(하)

내신대비 및 심화 문제 탐구편

Mathmatics The discovery of dharma

최경호 지음

홈페이지(www.m1239.co.kr) 동영상 강의 및 풀이

원리탐구

중3수학(하) 탐구 차례

	생활 속의 수학	읽을거리
제1강	1. 통계를 신뢰할 것인가?	1. 여러 가지 산포도(Dispersion)
제2강	1. 화면의 크기	1. 피타고라스 정리의 증명방법은 몇 가지나 될까?
제3강	1. 음료수 캔이 원기둥인 이유	1. 피타고라스는 피타고라스 정리를 어떻게 발견하였을까?
제4강	1. 직접 재지 않고도 높이를 구할 수 있어요!	1. 삼각비의 기호 이야기
제5강	1. 어디까지 볼 수 있을까?	1. 태양까지의 거리는 달까지의 거리의 몇 배일까?
제6강	1. 이상혁의 산술관견(算術管見)	1. 원이야기
제7강	1. 아벨로스(Arbelos)	1. 원의 접선과 반지름
제8강	1. 아르키메데스의 샐리논 (Salinon, 소금그릇)	1. 수레바퀴는 왜 둥글게 만드는가?

제 1강 대푯값과 산포도

점수(점)	학생수(명)
50점	2
60점	5
70점	12
80점	13
90점	7
100점	1
합계	40

1 탐구예제

원리탐구

다음 중 옳은 것을 모두 고르면?

① 대푯값은 자료들이 흩어져 있는 정도를 하나의 수로 나타낸 값이다.
② 대푯값에서 평균, 중앙값, 최빈값 등이 있다.
③ 평균은 변량의 총합을 변량의 개수로 나눈 값이다.
④ 평균은 극단적인 값에 영향을 받지 않는다.
⑤ 평균은 자료를 작은 값부터 크기순으로 나열했을 때 중앙에 위치한 값과 일치한다.
⑥ 최빈값은 없을 수도 있다.

풀이 답: ②, ③, ⑥

① 대푯값에서 평균, 중앙값, 최빈값 등이 있다. 표준편차는 산포도를 나타낸다.
④ 평균은 극단적인 값에 영향을 받는다.
⑤ 자료를 작은 값에서부터 크기순으로 나열했을 때, 중앙에 위치한 값은 중앙값이다.
⑥ 각 자료의 도수가 모두 같으면 최빈값은 없다. 옳다.

유제 1

다음 〈보기〉 중 옳지 <u>않은</u> 것을 모두 골라라.

① 산포도에는 평균, 중앙값, 최빈값이 있다.
② 평균은 어떤 자료에도 쓸 수 있는 대푯값이다.
③ 최빈값은 여러 개일 수도 있다.
④ 중앙값은 항상 주어진 자료 중에 존재한다.
⑤ 자료의 값 중에 매우 크거나 매우 작은 값이 있는 경우 대푯값으로 중앙값을 사용한다.

2 탐구예제

원리탐구

다음 |보기|를 모두 만족하는 자연수 a의 개수를 구하여라.

|보기|

① 다섯 개의 수 5, 10, 17, 20, a의 중앙값은 17이다.
② 여섯 개의 수 11, 33, 41, 48, 51, a의 중앙값은 37이다.

풀이 답: 17 (개)

①에서 $a \geq 17$이고, ②에서 $\frac{33+41}{2}=37$이므로 $a \leq 33$이다. $\therefore 17 \leq a \leq 33$

따라서 자연수 a의 개수는 $33-17+1=17$ (개)다.

유제 2

두 자연수 a, b에 대하여 변량 3, 4, a, b, 7의 중앙값이 5이고, 변량 10, a, b, 15의 중앙값이 12일 때, $a+b$의 값을 구하여라.(단, $a<b$)

3 탐구예제 원리탐구

변량 $a, b, c, 9, 2, 4, 3, 2, 9$ 의 중앙값이 7이고, 최빈값이 9일 때, a, b, c의 값을 구하여라.(단, a, b, c는 $a<b<c$인 자연수이다.)

① $a=5, b=7, c=8$ ② $a=7, b=8, c=10$ ③ $a=7, b=8, c=9$
④ $a=5, b=6, c=8$ ⑤ $a=4, b=5, c=7$

풀이 답: ③

순서대로 나열하면 $2,2,3,4,a,b,c,9,9$ 또는 $2,2,3,4,a,b,9,9,c$ 또는 $2,2,3,4,a,9,9,b,c$ 이고, 중앙값이 7이므로 $a=7$, 최빈값이 9이므로 b 또는 c가 9이다. $a<b<c$이므로 $a=7, b=8, c=9$이다.

유제 3

다음 자료의 중앙값, 최빈값, 평균을 각각 a, b, c라 할 때, a, b, c의 크기를 바르게 비교한 것은?

6, 7, 4, 5, 6, 4, 1, 6, 3, 9

① $a<b<c$ ② $c<b<a$ ③ $a<c<b$ ④ $b<c<a$ ⑤ $c<a<b$

4 탐구예제 원리탐구

다음 표는 수돌이네 반 학생 A, B, C, D, E 5명의 수학 성적의 평균이 91점일 때의 편차를 나타낸 것이다. 다음을 구하여라.

학생	A	B	C	D	E
편차(점)	-2	-5	x	4	1

(1) A학생의 성적 (2) 수학 성적의 표준편차

풀이 답: (1) 89 (점) (2) $\sqrt{10}$ (점)

(1) 평균이 91점이고 A의 편차는 -2점이므로 A의 성적은 $91+(-2)=89$ (점)이다.
(2) 편차의 합은 항상 0이므로 $(-2)+(-5)+4+x+1=0$, $x=2$ 이다.

(분산)$=\dfrac{(-2)^2+(-5)^2+4^2+2^2+1^2}{5}=\dfrac{50}{5}=10$이므로 (표준편차)$=\sqrt{10}$ (점)이다.

유제 4

다음 표는 A, B, C, D, E 다섯 명의 수학 성적의 편차를 나타낸 것이다. 평균이 76 점일 때, 다음을 구하여라.

학 생	A	B	C	D	E
편 차	-3	-4	3	6	x

(1) 학생 E 의 성적 (2) 수학 성적의 표준편차

5 탐구예제 원리탐구

다음은 어느 학급 학생 10명의 수학 성적에 대한 도수분포표이다. 성적의 분산을 구하여라.

수학 성적	학생 수
$65^{이상}$ ~ $75^{미만}$	4
75 ~ 85	4
85 ~ 95	2
합 계	10

풀이 답: 56

평균이 $\frac{70\times4+80\times4+90\times2}{10}=\frac{780}{10}=78$점이므로 분산은 다음과 같다.

$\frac{(-8)^2\times4+2^2\times4+12^2\times2}{10}=\frac{560}{10}=56$

유제 5

다음 표는 수돌이네 반 학생 10명의 수학 점수를 조사하여 나타낸 도수분포표이다. 물음에 답하여라.

계급(점)	도수	편차
$60^{이상}$ ~ $70^{미만}$	1	-20
70 ~ 80	2	-10
80 ~ 90	3	a
90 ~ 100	4	10
합계	10	

(1) 위 도수분포표에서 a값을 구하여라.

(2) 위 도수분포표에서 표준편차를 구하여라.

6 탐구예제 원리탐구

다음 도수분포표는 100 점 만점의 수학 시험에 대한 성적을 나타낸 것이다. 평균이 80 점, 분산이 120 점일 때, $x+y$ 의 값을 구하여라.

점수(점)	60	70	80	90	100
학생 수(명)	2	x	y	4	1

풀이 답: 8

평균이 80점이므로 $\frac{60\times2+70\times x+80\times y+90\times4+100\times1}{2+x+y+4+1}=80$, $\frac{70x+80y+580}{x+y+7}=80$,

$\frac{7x+8y+58}{x+y+7}=8$, $7x+8y+58=8x+8y+56$, $x=2$이다.

$x=2$이고, 분산이 120점이므로 $\frac{(-20)^2\times2+(-10)^2\times2+0+10^2\times4+20^2\times1}{2+2+y+4+1}=120$,

$\frac{1800}{9+y}=120$, $9+y=15$, $y=6$이다. $\therefore x+y=2+6=8$

유제 6

다음 10개의 자료의 평균이 5이고 분산이 9일 때 $b-a$의 값은?(단, $a<b$이다.)

3, 4, a, 2, 7, b, 8, 3, 8, 5

7 탐구예제 원리탐구

4개의 변량 a, b, c, d 의 평균이 5, 분산이 2일 때, 변량 $2a-5,\ 2b-5,\ 2c-5,\ 2d-5$ 에서 다음을 구하여라.

(1) 평균 (2) 분산

풀이 답: (1) 5 (2) 8

(ⅰ) a, b, c 의 평균을 m 이라하면 $\dfrac{a+b+c}{3}=m$가 된다.

a, b, c 에 각각 2를 곱한 $2\times a,\ 2\times b,\ 2\times c$의 평균은 $\dfrac{2\times a+2\times b+2\times c}{3}=2\times\dfrac{a+b+c}{3}=2\times m$ 에서 알 수 있듯이 변량에 같은 값을 곱하면 **평균도 그 값을 곱한 만큼 커진다.**

(ⅱ) a, b, c 의 분산은 $\dfrac{(a-m)^2+(b-m)^2+(c-m)^2}{3}=S^2$이다.

a, b, c 에 각각 2를 곱한 $2\times a,\ 2\times b,\ 2\times c$의 분산은 얼마일까?

$$\frac{(2\times a-2\times m)^2+(2\times b-2\times m)^2+(2\times c-2\times m)^2}{3}=2^2\times\frac{(a-m)^2+(b-m)^2+(c-m)^2}{3}=2^2\times S^2$$

에서 알 수 있듯이 변량에 같은 값을 곱하면 **분산은 그 값을 제곱한 만큼 커진다.** 따라서 표준편차는 2를 곱한 만큼 커진다. 변량에 같은 값을 합하는 경우에 평균은 그 값을 합한 만큼 커지고, 변량에 같은 값을 빼는 경우에 평균은 그 값을 뺀 만큼 작아진다. 그러나 **변량에 같은 값을 합하거나 빼는 경우에 분산과 표준편차는 평균값에서 얼마만큼 차이가 나는 가의 문제이므로 변화가 없다.**

이것을 정리하면 오른쪽과 같다.

(변량의 변화에 따른 평균, 분산, 표준편차의 변화)

변량	평균	분산	표준편차
x_i	m	V	σ
$ax_i\pm b$	$am\pm b$	a^2V	$\lvert a\rvert\sigma$

a, b, c, d의 평균과 분산이 각각 $5, 2$이므로 $2a-5,\ 2b-5,\ 2d-5$의 평균과 분산은 다음과 같다.

(1) (평균)$=2\times5-5=5$이다. (2) (분산)$=2^2\times2=8$이다.

유제 7

네 수 $a,\ b,\ c,\ d$의 평균을 30, 표준편차를 5라고 할 때, 네 수 $a+20,\ b+20,\ c+20,\ d+20$의 평균과 분산을 각각 구하여라.

8 탐구예제 원리탐구

변량 $a,\ b,\ c,\ d$의 평균이 5이고 분산이 6일 때, 변량 $a^2,\ b^2,\ c^2,\ d^2$의 평균은 구하여라.

풀이 답: 31

다음 도수분포표에서 평균을 m 이라고 하면, 분산S^2 을 구하는 방법

변량	x_1	x_2	$\cdots$	x_n

분 산: $S^2=\dfrac{(x_1-m)^2+(x_2-m)^2+\cdots+(x_n-m)^2}{n}$

$=\dfrac{(x_1^2+x_2^2+\cdots+x_n^2)-2m(x_1+x_2+\cdots+x_n)+m^2n}{n}$

$=\dfrac{(x_1^2+x_2^2+\cdots+x_n^2)-2m\times mn+m^2n}{n}=\dfrac{x_1^2+x_2^2+\cdots+x_n^2}{n}-m^2$

즉, (분산)=(변량의 제곱의 평균)-(평균)^{2}이다.

(해1) $a+b+c+d=20$

$(a-5)^2+(b-5)^2+(c-5)^2+(d-5)^2=24$

$a^2+b^2+c^2+d^2-10(a+b+c+d)+100=24$

$a^2+b^2+c^2+d^2=10\times20-100+24$

$\therefore a^2+b^2+c^2+d^2=124$

따라서 $a^2,\ b^2,\ c^2,\ d^2$의 평균은 $\dfrac{a^2+b^2+c^2+d^2}{4}=\dfrac{124}{4}=31$이다.

(해2) 분산이 6이고, (분산)=(변량의 제곱의 평균)-(평균)^{2}이므로

(변량의 제곱의 평균)=(분산)+(평균)$^2=6+5^2=31$이다.

유제 8

변량 $3,\ x,\ y,\ 2$의 평균이 1, 분산이 5일 때, x^2+y^2의 값을 구하여라.

종합문제〉〉

1. 중앙값에 대한 다음 설명 중 옳지 않은 것은?
① 중앙값은 항상 주어진 자료 중에 존재한다.
② 자료의 개수가 짝수인 경우에는 두 개의 값이 중앙에 있게 되며, 이 경우 중앙에 있는 두 자료의 값의 평균을 중앙값으로 정한다.
③ 자료의 값 중에 매우 크거나 매우 작은 값에 영향을 받지 않는다.
④ 자료를 크기순으로 나열할 때, 한가운데 놓이는 값이다.
⑤ 중앙값은 항상 단 하나로 값이 정해진다.

2. 다음 표는 A, B, C 세 동아리의 회원 수와 나이의 평균이다. 이 때, 세 동아리 전체 회원의 나이의 평균을 구하여라.

동아리	회원 수(명)	평균(세)
A	8	20
B	12	15
C	10	17

3. 최빈값에 대한 다음 설명 중 옳지 않은 것은?
① 자료 중에서 가장 많이 나오는 값이다.
② 최빈값은 항상 존재한다.
③ 선거에서 가장 많은 표를 얻은 후보자가 당선되는 것은 최빈값을 이용한 경우이다.
④ 자료의 개수가 적은 경우 자료 전체의 특징을 반영하지 못할 수도 있다.
⑤ 최빈값은 여러 개 있을 수도 있으며 매우 작거나 매우 큰 값의 영향을 받지 않는다.

4. 아래 표는 수순이 학급 50명의 학생이 영국에서 활약하고 있는 기성용 선수에 대한 인기투표 결과의 성적을 나타낸 것이다. 다음을 구하여라.

계급(점)	50 이상 ~60 미만	60 이상 ~70 미만	70 이상 ~80 미만	80 이상 ~90 미만	90 이상 ~100 미만	합 계
도수(명)	4	8	12	16	10	50

(1) 평균 (2) 분산 (3) 표준편차

5. 다음 표는 어느 독서 동아리 회원 20명이 겨울 방학 동안 읽은 책의 권수에 대한 도수분포표이다. 다음을 각각 구하여라.

책의 수(권)	도수(명)
1^이상^ ~ 3^미만^	4
3 ~ 5	8
5 ~ 7	3
7 ~ 9	4
9 ~ 11	1
합계	20

(1) 평균

(2) 중앙값

(3) 최빈값

6. 다음은 수순이네 반 학생들의 수학 시험 점수를 조사한 것이다. 옳은 것은?

점수(점)	학생수(명)
50점	2
60점	5
70점	12
80점	13
90점	7
100점	1
합계	40

① 평균은 72점이다.
② 중앙값은 알 수 없다.
③ 최빈값은 13개이다.
④ 주어진 자료의 대푯값으로 75점도 적합하다.
⑤ 이 자료에서 분산으로 자료의 분포상태를 알 수 없다.

7. 다음 8개 자료의 중앙값은5, 최빈값은 7이라고 할 때 $a+b+c$의 값을 구하면?

2, a, 4, 5, b, 4, 7, c

8. 세 변량 a, b, c의 평균이 8, 분산이 14일 때, $a-2$, $b-2$, $c-2$의 평균 m과 분산 s^2의 합을 구하여라.

9. 학생 6명의 수학 점수를 작은 값에서부터 크기순으로 나열할 때, 3번째 학생의 점수는 76점이고, 중앙값은 79점이라고 한다. 이 집단에 수학 점수가 83점인 학생이 들어왔을 때, 7명의 수학 점수의 중앙값은?

① 79 ② 80 ③ 81 ④ 82 ⑤ 83

10. 어느 마을의 남녀 전체의 평균 나이가 45세이다. 이 마을 남자의 평균 나이가 50세이고, 여자의 평균 나이가 40세일 때, 남자와 여자의 수의 비를 구하여라.

11. 다음 자료는 어느 고등학교 1학년 4개 반 학생들의 국어 성적의 평균과 표준편차를 나타낸 것이다. 다음 설명 중 옳은 것을 모두 고르면?(정답2개)

반	1	2	3	4
평균(점)	75	75	76	74
표준편차(점)	$3\sqrt{2}$	$2\sqrt{3}$	$4\sqrt{3}$	$\sqrt{17}$

① **최저득점자가 4반에 있다.**
② **편차의 총합은 3반이 가장 크다.**
③ **2반의 성적이 3반의 성적보다 더 고르다고 할 수 있다.**
④ **전체적으로 1반의 성적이 2반의 성적보다 더 높다고 할 수 있다.**
⑤ **표준편차만으로 편차를 제곱한 값들의 평균이 가장 큰 반을 알 수 있다.**

12. 한 상자에 들어있는 20개의 달걀을 12개와 8개의 묶음으로 나누었다. 12개 묶음에 있는 달걀 무게의 평균은 $50g$, 분산은 5이고 8개 묶음에 있는 달걀 무게의 평균은 $50g$, 분산은 6이다. 이 때 달걀 20개 전체의 무게의 분산을 구하여라.

13. 변량 2, 4, x, y, z의 평균이 3이고, 표준편차가 1일 때, $x^2+y^2+z^2$의 값을 구하여라.

14. 변량 6, 10, a, b, c의 중앙값과 최빈값이 12이고, 평균이 13일 때, 분산을 구하여라.

15. 수학시험 결과 A반과 B반의 분산의 비가 $3:1$이 되었다. A반의 표준편차가 $\sqrt{2}$일 때, B반의 표준편차를 구하여라.

16. 네 수 a, b, c, d의 평균을 7, 분산을 5라고 할 때, 다음 설명 중 옳은 것을 두 가지 고르면?

① $a+3$, $b+3$, $c+3$, $d+3$**의 평균은** 7**이다.**

② $a+3$, $b+3$, $c+3$, $d+3$**의 분산은** 8**이다.**

③ $2a$, $2b$, $2c$, $2d$**의 평균은** 14**이다.**

④ $2a$, $2b$, $2c$, $2d$**의 분산은** 10**이다.**

⑤ $2a+3$, $2b+3$, $2c+3$, $2d+3$**의 분산은** 20**이다.**

17. 어떤 시험의 합격자의 평균은 76점, 불합격자의 평균은 52점, 전체 평균은 64점 이었다. 이 시험의 합격률을 구하면?

① 35% ② 40% ③ 45% ④ 50% ⑤ 55%

18. 다음 두 자료 A, B에 대하여 A의 중앙값이 17이고, 두 자료 A, B를 섞은 전체 자료의 중앙값이 18일 때, $a+b$의 값은? (단, a, b는 자연수이고 $a<b$이다.)

자료 A: 10, 12, a, b, 20
자료 B: 15, $b-1$, 20, 21, a

19. 세 변량 a, b, c의 평균을 m이라고 할 때, $\frac{3a+m}{3}$, $\frac{3b+m}{3}$, $\frac{3c+m}{3}$의 평균은?

① $4m$ ② $3m$ ③ $\frac{4m}{3}$ ④ $\frac{m}{3}$ ⑤ $\frac{m}{4}$

20. 다음 표는 5명의 학생 A, B, C, D, E의 수학 수행평가 성적에 대한 편차를 나타낸 것이다. 5명의 평균이 8점일 때, B와 C 두 학생의 수행평가 점수의 평균을 구하면?

학생	A	B	C	D	E
편차	-3	a	b	1	4

① 5 ② 6 ③ 7 ④ 8 ⑤ 9

탐구문제〉〉〉

1. 5 개의 변량 4, 10, x, y, 5 의 평균이 6 이고 분산이 4.4 일 때, xy 의 값을 구하여라.

2. 다섯 개의 수 x, y, z, 2, 3의 중앙값은 5, 평균이 6, 분산이 12일 때, $x-y+z$의 값은?(단, $x<y<z$)

3. 다음 표는 A, B, C, D 네 명의 용돈과 평균의 차를 나타낸 것이다. 학생 A보다 용돈이 1000원 적은 학생 한 명을 더 포함하여 A, B, C, D, E 다섯 학생 용돈의 평균을 구하였더니 원래 평균보다 5% 감소하였다고 할 때, 가장 용돈이 많은 학생의 용돈을 구하여라.

학생	A	B	C	D
(용돈)−(평균)(원)	500	1000	−500	−1000

4. 3개의 변량 x, y, z의 평균은 3이고 표준편차는 1이다. 여기에 평균이 5이고 표준편차가 3인 세 변량 a, b, c를 추가 했을 때, 다음을 구하여라.

(1) 6개의 변량의 평균 **(2) 6개의 변량의 분산**

[사고력 퀴즈와 퍼즐]

1. 서울광장에 모인 인원수는 어떻게 셀까?

아래 기사를 읽고 물음에 답하여라.

가수 싸이(본명 박재상)는 2012년 10월 4일 오후 서울광장에서 `서울시와 함께 하는 싸이 글로벌 석권 기념 콘서트'를 성황리에 가졌다.

인터넷과 소셜미디어를 타고 전 세계를 강타한 싸이의 ` '강남스타일' 열풍이 서울광장에 8만여 명이 동시에 말춤을 추는 대 장관을 연출했다. 싸이는 특유의 열정적인 무대 매너로 관객들의 참여를 이끌어내며 콘서트장인 서울광장을 거대한 축제의 용광로로 달궈냈다. 싸이가 전 세계에 과거와는 전혀 다른 형태의`한류 열풍을 촉발시키며 대한민국을 `문화수출국으로 전환시키는데 촉매제 역할을 할 것이란 평가다.

특히 이날 공연은 세계적인 소셜미디어인 유튜브와 유스트림, 서울시청 라이브 방송 등을 통해 한국은 물론 전세계 네티즌들에게 생중계돼 전 세계인이 즐기는 축제의 장이 됐다. 싸이는"전 세계에서 유튜브를 통해 감상중인 해외 팬들에게 감사하다"면서 세계 네티즌들에 인사말을 건넸다.

서울광장에 모인 인원수 8만여 명은 어떻게 셀까? 방법과 이유를 써라.

[생활 속의 수학]

1. 통계를 신뢰할 것인가?

한 관광객이 낯선 도시에서 밤에 택시 한 대가 주차된 자동차를 들이받고 달아나는 것을 목격한다. 그는 파란색 택시의 소행이라고 경찰에 진술한다. 그 도시에는 택시 회사가 두 곳 뿐인데 한 곳은 파란색 택시를 운영하고 다른 곳은 초록색 택시를 운영하므로, 곧바로 파란색 택시 회사가 용의선상에 오른다. 그러나 경찰관들은 그 증인을 믿어도 되는지 확인하고 싶다. 사건이 발생한 시각은 캄캄한 밤이므로 증인이 파란색과 초록색을 헷갈렸을 수도 있으니까 말이다. 그리하여 그들은 사건 당시와 비슷하게 어두운 이튿날 밤에 증인의 시력을 검사한다. 검사 결과로 볼 때 증인은 초록색 택시와 파란색 택시를 각각 80%의 확률로 옳게 식별한다. 판사는 그 정도 확률이면 충분하다고 보고 파란색 택시 회사의 사장에게 유죄를 선고한다.

과연 옳은 선고일까? 아니다. 왜냐하면 그 도시에 초록색 택시는 25대나 있지만 파란색 택시는 5대 뿐이라는 사실을 감안하지 않고 확률을 계산했기 때문이다. 택시의 대수와 증인의 적중률을 함께 고려하면 아래와 같은 표를 얻을 수 있다.

	증언: "파란색 택시다!"	증언: "초록색 택시다!"
파란색 택시를 본 경우	4	1
초록색 택시를 본 경우	5	20

경찰에서 증인의 시력을 검사한 결과에 따르면, 증인은 20%의 확률로 오류를 범한다. 쉽게 말해서 증인은 파란색 택시 5대 가운데 1대를 초록색 택시로, 초록색 택시 25대 가운데 5대를 파란색 택시로 오인한다. 만일 그 도시의 택시 30대 전부가 차례로 증인 앞을 지나갔다면, 통계학적으로 볼 때 증인은 파란색 택시가 9대 지나갔다고 증언할 것이다. 그러나 그 9대 가운데 5대는 초록색 택시다!(적중률 $\frac{4}{9}\times 100 = 44.44\%$) 따라서 추가 증거들이 없다면, 경찰은 그 증언을 무시해야 마땅하다. 증인의 시력 (적중률 80%)만 가지고 증언의 가치를 평가할 수는 없다.

[출전: 크리스토프 드뢰서 지음, 수학 시트콤, 해나무]

[읽을거리]

1. 여러 가지 산포도(Dispersion)

같은 대표값으로 두 집단을 비교하거나, 또는 한 집단의 자료의 특징을 기술함에 있어서 대표값만으로 부족할 때가 많다. 그 보다는 오히려 자료값이 그 대표값 주위에 어느 정도 흩어져 있는가를 알아보는 측도가 필요한 경우가 더 많다. 이 측도를 바로 **산포도**라고 한다.

대표값을 산술평균 $\overline{X}$로 할 때의 주어진 변량 x_1, x_2, $\cdots$, x_n에 대하여. 자료에서 평균을 뺀 값 즉 $x_1-\overline{X}$, $x_2-\overline{X}$, $\cdots$, $x_n-\overline{X}$을 편차라고 하는데 산포도로서 이들 편차들의 평균을 구하는 것이 가장 바람직할 것이나 편차의 합은 항상 0이 나오므로 편차의 평균을 구하면 항상 0이 되어 산포도로 사용할 수가 없고, 다른 방법을 이용하여 산포도를 구해야 한다.

산포도에는 **분산, 표준편차, 평균편차, 사분편차, 범위** 등이 있다.

(1) **분산, 표준편차** : 변량 x_i와 평균 $\overline{X}$의 편차들의 제곱의 평균을 분산이 라고 하며, 분산의 양의 제곱근을 **표준편차**라고 하며 가장 대표적인 산포도이다. 분산은 편차의 제곱을 포함하므로 변량의 극단값에 영향을 많이 받으며, 표준편차의 경우도 마찬가지이나, 수리적 이론 전개에 편리하므로 가장 많이 사용되는 산포도이다.

분산; $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{X})^2 f_i$, 표준편차; $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{X})^2 f_i}$

(2) **평균편차** : 편차의 절대값의 평균을 평균편차라고 하지만 식에 절대값의 기호가 있기 때문에 수리적으로 다루기가 불편하여 별로 사용되지 않는다.

$$\text{평균편차; } M.D = \frac{1}{n}\sum_{i=1}^{n}\left|x_i-\overline{X}\right| f_i$$

(3) **사분편차** : n개의 자료를 작은 것부터 큰 것의 순으로 배열하여 $\frac{n}{4}$번째에 해당되는 자료의 값을 **제1사분위수** Q_1, $\frac{3n}{4}$번째에 해당되는 자료의 값을 **제3사분위수** Q_3이라고 할 때, $Q=\frac{Q_3-Q_1}{2}$를 **사분편차**라고 한다.

(4) **범위** : 자료 중에서 최대값과 최소값의 차를 **범위**라고 한다. 이는 산포도로서 가장 간단히 구할 수 있으나, 극단값에 영향을 많이 받는다는 단점이 있다.