

가을을 대표하는 것의 하나로 단풍을 꼽을 수 있듯이
통계 자료에서도 대표하는 값을 찾을 수 있다.

VII

통계

- 1 대푯값
- 2 산포도
- 3 산점도와 상관관계





되짚어 보기

평균

초 5~6

- ▶ 자료의 평균은 변량의 총합을 변량의 개수로 나눈 값이다.

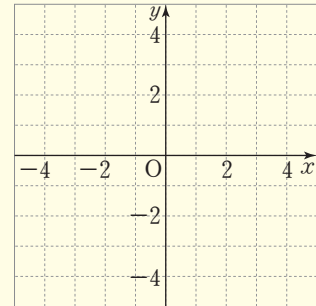
$$(\text{평균}) = \frac{(\text{변량의 총합})}{(\text{변량의 개수})}$$

좌표평면

중 1

- 2 다음 점을 오른쪽 좌표평면 위에 나타내시오.

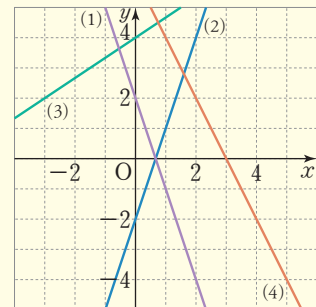
- (1) A(2, 4)
- (2) B(-3, 1)
- (3) C(-2, -3)
- (4) D(0, 1)



일차함수의 그래프의 성질

중 2

- 3 오른쪽 그림과 같은 일차함수의 그래프에서 x 의 값이 증가할 때 y 의 값도 증가하는 직선을 모두 찾으시오.



제곱근

중 3

- ▶ 어떤 수 x 를 제곱하여 a 가 될 때, x 를 a 의 제곱근이라고 한다.

- 4 다음 수의 양의 제곱근을 구하시오.

- (1) 16
- (2) 24
- (3) 0.36
- (4) $\frac{25}{169}$



단원을 시작하며

형찬이의 일기

2020년 8월 ○○일

오늘은 친구들과 '통계 물놀이 공원'에 놀러 갔다. 안으로 들어가 보니 두 수영장에 다음과 같은 안내문이 있었는데, '왜 평균 수심에 대한 정보는 없을까?' 하는 의문이 들었다.



최저 수심: 0.8 m
최고 수심: 2.0 m



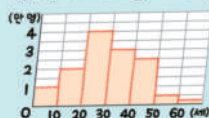
최저 수심: 1.3 m
최고 수심: 1.5 m

안전 요원에게 물어보니 두 수영장 모두 평균 수심은 1.4 m로 같은데, 수심의 차이가 크게 나서 평균만으로는 수심의 정보를 충분히 알 수 없다고 하였다. 이 일로 평균 수심이 같아도 물의 깊이가 다를 수 있다는 것을 알았다. 오늘 평균 수심만 안다고 해서 안전하게 놀 수 있는 것은 아님을 통계 물놀이 공원에서 체험했다.

❶ 이 단원에서는 통계 자료를 분석하는 데 필요한 대푯값과 산포도의 의미를 이해하고, 산점도와 상관관계를 알아본다.

통계 물놀이 공원

입장객 나이에 대한 히스토그램



일주일간 낮 최고 기온 (단위: °C)

5월	6월	7월	8월	9월	10월	11월
33.9	35.3	35.9	35.3	32.5	36.3	36.0

대푯값

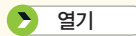
중앙값, 최빈값, 평균의 의미를 이해하고, 이를 구할 수 있다.

반 대표를 뽑을 때는 투표를 하여 가장 많은 표를 받은 사람을 뽑는다.



탐구 학습

대푯값은 무엇인가요?

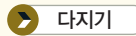


열기

다음은 영철이네 반 학생 7명의 오래 매달리기 시간을 조사하여 나타낸 자료이다. 이 자료의 평균을 구하여 보자.

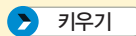
오래 매달리기 시간 (단위: 초)

4, 6, 10, 7, 9, 12, 64



다지기

이 자료의 평균은 $\frac{4+6+10+7+9+12+64}{7} = \square$ (초)



키우기

위에서 구한 평균이 이 자료를 대표한다고 볼 수 있을까?

대푯값

자료의 전체적인 특징을 알아보려면 먼저 자료를 대표할 수 있는 값을 찾는 것이 필요하다.

일반적으로 자료 전체의 특징을 대표적으로 나타내는 값을 그 자료의 **대푯값**이라고 한다. 대푯값에는 여러 가지가 있으나 평균이 가장 많이 쓰인다. 평균은 변량의 총합을 변량의 개수로 나눈 것이다.

이전에 배운 내용

변량은 자료를 수량으로 나타낸 것이다.



개념 확인

변량이 4, 5, 2, 9일 때,
(평균) $= \frac{4+5+2+9}{4} = 5$



평균은 $\frac{(\text{변량의 총합})}{(\text{변량의 개수})}$ 이야.

문제 1

다음은 학생 10명의 줄넘기 2단 뛰기 횟수를 조사하여 나타낸 자료이다. 이 자료의 평균이 7회라고 할 때, x 의 값을 구하시오.

줄넘기 2단 뛰기 횟수 (단위: 회)

x , 8, 9, 4, 6, 6, 7, 5, 10, 9



중앙값

탐구 학습의 자료와 같이 변량 중 매우 크거나 매우 작은 값이 있는 경우에 평균은 그 극단적인 값의 영향을 많이 받는다.

이 같은 경우에는 변량을 작은 값부터 크기순으로 나열한

4, 6, 7, 9, 10, 12, 64

에서 중앙에 있는 값인 9초를 대푯값으로 사용하기도 한다. 이처럼 변량을 작은 값부터 크기순으로 나열할 때 중앙에 놓인 값을 **중앙값**이라고 한다.

이때 변량의 개수가 짝수이면 중앙에 놓이는 값이 2개이므로 두 값의 평균을 중앙값으로 한다.

변량 중 극단적인 값이 있는 경우에 대푯값으로 중앙값을 쓰기도 해.



개념 확인

자료의 중앙값 구하기

변량의 개수가 5이므로
3번째 값인 2가 중앙값이야.



1 2 2 3 5

→ 중앙값: 2

변량의 개수가 6이므로
3번째와 4번째 값의
평균인 2.5가 중앙값이야.



1 2 2 3 5 5

→ 중앙값: $\frac{2+3}{2} = 2.5$

문제 2 다음 자료의 중앙값을 구하시오.

(1) 4, 7, 10, 9, 16, 13, 14

(2) 11, 18, 18, 16, 15, 13

문제 3 다음은 학생 8명이 한 학기 동안 읽은 책의 수를 조사하여 나타낸 표이다. 물음에 답하시오.

읽은 책의 수							(단위: 권)
민호	지혜	상철	용민	지민	민수	지환	희원
6	5	9	8	4	6	28	6

(1) 학생 8명이 한 학기 동안 읽은 책의 수의 평균과 중앙값을 구하시오.

(2) 평균과 중앙값 중 이 자료의 특징을 더 잘 나타내는 것을 말하고, 그 이유를 설명하시오.

최빈값

대푯값을 구할 때 신발의 크기와 같은 자료에서는 가장 많이 나타나는 값이 중요한 경우가 있다. 이때 변량 중 가장 많이 나타나는 값을 그 자료의 **최빈값**이라고 한다.

예를 들어 어느 신발 가게에서 하루 동안 팔린 신발의 크기(단위: mm)가

255, 260, 265, 275, 255, 280, 255, 255

일 때, 가장 많이 팔린 신발의 크기는 255 mm이므로 최빈값은 255 mm이다.

최빈값은 값이 하나로 정해지는 평균과 중앙값과는 달리 자료에 따라 두 개 이상일 수 있다.

최빈값은 자료가 수치로 주어지지 않은 경우에도 사용할 수 있다.

개념 확인

자료의 최빈값 구하기



문제 4 다음 자료의 최빈값을 구하시오.

(1) 지수 친구들이 사용하는 유에스비(USB)의 용량

(단위: GiB)

16, 32, 4, 16, 8, 16, 4, 8, 16, 32



(2) 미술 동아리 학생들이 좋아하는 색깔

빨강, 파랑, 검정, 파랑, 검정, 검정, 파랑, 빨강, 검정



의사소통



다음 두 학생의 대화를 읽고 일상생활에서 평균, 중앙값, 최빈값을 대푯값으로 하는 적절한 예를 각각 하나씩 찾아 그 이유를 말하여 보자.

자료에 극단적인 값이 없는 경우는 일반적으로 평균을 사용해.



자료에 극단적인 값이 있는 경우는 어떻게 하지?





1

다음 자료의 평균, 중앙값, 최빈값을 구하시오.

(1) 7, 8, 1, 6, 8

(2) 19, 17, 2, 19, 14, 19

2

다음은 학생 10명이 1분 동안 실시한 윗몸 일으키기 기록을 조사하여 나타낸 줄기와 잎 그림이다. 이 자료의 평균, 중앙값, 최빈값을 구하시오.

윗몸 일으키기 기록 (0이9는 9회)

줄기	잎
0	9
1	3 3 6 7
2	4 8 8
3	2 5

3

다음은 6개의 변량을 작은 값부터 크기순으로 나열한 것이다. 이 자료의 중앙값이 8일 때, x 의 값을 구하시오.

3, 5, x , 9, 12, 13

4

다음은 예원이 친구들의 취미 활동을 조사하여 나타낸 표이다. 이 표를 보고, 최빈값을 구하시오.

취미 활동	학생 수(명)
음악 감상	5
독서	6
영화 감상	16
춤	4

5

다음은 학생 8명이 일주일 동안 운동한 시간을 조사하여 나타낸 자료이다. 이 자료의 평균이 8시간이라고 할 때, x 의 값을 구하고, 중앙값과 최빈값을 구하시오.

운동한 시간 (단위: 시간)

x , 5, 8, 14, 1, 5, 14, 12



6 창의·융합

영수네 학교 농구 팀 13명의 키는 평균이 190 cm이고, 중앙값은 188 cm이다. 그런데 이 선수 중에서 한 선수가 다른 학교로 전학을 가고, 다른 학교에서 한 선수가 전학을 온 후에 키의 평균이 191 cm가 되었다. 물음에 답하시오. (단, 13명의 키는 모두 다르다.)

- (1) 전학을 온 선수의 키는 전학을 간 선수의 키보다 몇 cm가 더 큰지 구하시오.
- (2) 전학을 온 선수의 키가 189 cm라고 할 때, 중앙값은 어떻게 변화하는지 말하시오.

산포도

분산과 표준편차의 의미를 이해하고, 이를 구할 수 있다.

사하라 사막의 1월 일평균 기온은 17℃이지만 밤에는 눈이 내리기도 한다.



탐구 학습

산포도는 무엇인가요?



열기

다음은 두 양궁 선수 A와 B가 9번의 화살을 쏘아 얻은 점수를 나타낸 표이다. 물음에 답하여 보자.

점수										(단위: 점)
회	1	2	3	4	5	6	7	8	9	
A 선수	10	5	8	6	10	5	9	9	10	
B 선수	7	8	9	8	9	9	8	7	7	

- (1) 두 선수가 얻은 점수의 평균을 각각 구하여 보자.
- (2) 두 선수가 얻은 점수의 분포 상태를 각각 그래프로 나타내 보고, 어느 선수의 점수가 평균을 중심으로 더 모여 있는지 말하여 보자.

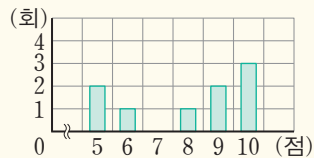


다지기

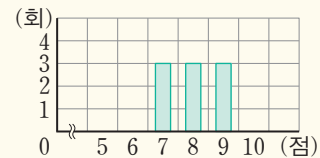
(1) A 선수: $\frac{10+5+8+6+10+5+9+9+10}{9} = \square$ (점)

B 선수: $\frac{7+8+9+8+9+9+8+7+7}{9} = \square$ (점)

- (2) 두 선수가 얻은 점수의 분포 상태를 각각 그래프로 나타내면 다음과 같다.



〈A 선수〉



〈B 선수〉

따라서 ☐ 선수의 점수가 평균을 중심으로 더 모여 있다.



키우기

그림이나 표 이외에 변량이 흩어져 있는 정도를 나타내는 다른 방법이 있을까?

산포도

탐구 학습에서 A 선수와 B 선수가 얻은 점수의 평균은 같지만 분포 상태가 서로 다르다는 사실을 알 수 있다.

이처럼 두 자료의 평균은 같아도 흩어져 있는 정도는 서로 다를 수 있다. 따라서 자료의 분포 상태를 알아보려면 대푯값 이외에도 변량이 흩어져 있는 정도를 살펴볼 필요가 있다.

산포도(散布度)에서 산포(散布)는 '흩어져 퍼진다.'는 뜻이다.

변량이 흩어져 있는 정도를 하나의 수로 나타낸 값을 **산포도**라고 한다. 일반적으로 변량이 대푯값에 모일수록 산포도가 작아지고, 변량이 대푯값에서 멀리 흩어질수록 산포도가 커진다.

개념 확인

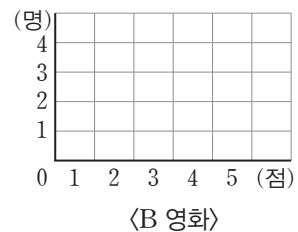


문제 1 다음은 두 영화 A, B를 관람한 10명의 평점을 조사하여 나타낸 자료이다. 물음에 답하시오.

A 영화 평점 (단위: 점)					B 영화 평점 (단위: 점)				
3	4	3	3	4	1	2	5	4	4
2	2	3	4	2	3	5	1	4	1

(1) 두 영화 A, B의 평점의 평균을 각각 구하시오.

(2) 두 영화 A, B의 평점의 분포 상태를 각각 그래프로 나타내고, 어느 영화의 평점이 산포도가 더 작은지 말하시오.



분산과 표준편차

산포도에는 여러 가지가 있으나 여기에서는 평균을 중심으로 변량이 흩어져 있는 정도를 나타내는 분산과 표준편차를 알아보자.

어떤 자료가 있을 때, 각 변량에서 평균을 뺀 값을 그 변량의 **편차**라고 한다.

- 평균보다 큰 변량의 편차는 양수이고, 평균보다 작은 변량의 편차는 음수이다.

$$(\text{편차}) = (\text{변량}) - (\text{평균})$$

탐구 학습의 두 선수의 점수에 대한 편차와 그 편차의 총합

(단위: 점)

회	1	2	3	4	5	6	7	8	9	총합
A 선수의 편차	2	-3	0	-2	2	-3	1	1	2	0
B 선수의 편차	-1	0	1	0	1	1	0	-1	-1	0

위의 표에서 편차의 총합은 항상 0이므로 편차의 평균도 0이 되어 이 값으로는 변량이 평균을 중심으로 흩어져 있는 정도를 알 수 없다. 따라서 편차의 제곱의 평균을 산포도로 이용한다.

두 선수의 점수에 대한 편차의 제곱의 평균

$$\begin{aligned} \text{A 선수: } & \frac{2^2 + (-3)^2 + 0^2 + (-2)^2 + 2^2 + (-3)^2 + 1^2 + 1^2 + 2^2}{9} = 4 \\ \text{B 선수: } & \frac{(-1)^2 + 0^2 + 1^2 + 0^2 + 1^2 + 1^2 + 0^2 + (-1)^2 + (-1)^2}{9} = \frac{2}{3} \end{aligned}$$

이때 4가 $\frac{2}{3}$ 보다 크므로 A 선수의 점수가 B 선수의 점수보다 평균을 중심으로 더 멀리 흩어져 있다는 것을 알 수 있다. 즉, B 선수의 점수가 A 선수의 점수보다 더 고르다고 할 수 있다.

- 자료의 분산과 표준편차가 클수록 그 자료의 분포 상태는 평균을 중심으로 더 멀리 흩어져 있다고 할 수 있다.

이처럼 어떤 자료의 편차의 제곱의 평균을 그 자료의 **분산**이라 하고, 분산의 음이 아닌 제곱근을 **표준편차**라고 한다.

이상을 정리하면 다음과 같다.

분산과 표준편차

$$(\text{분산}) = \frac{\{(\text{편차})^2 \text{의 총합}\}}{(\text{변량의 개수})}$$

$$(\text{표준편차}) = \sqrt{(\text{분산})}$$

예제 1

다음은 학생 5명이 일주일 동안 도서관을 이용한 시간을 조사하여 나타낸 자료이다. 도서관 이용 시간의 분산과 표준편차를 구하시오.

도서관 이용 시간 (단위: 시간)

4, 8, 6, 7, 10

$$\begin{aligned}\text{평균} &= \frac{4+8+6+7+10}{5} = 7(\text{시간}) \\ \text{분산} &= \frac{(-3)^2 + 1^2 + (-1)^2 + 0^2 + 3^2}{5} = 4 \\ \text{표준편차} &= \sqrt{4} = 2(\text{시간})\end{aligned}$$

☞ 분산: 4, 표준편차: 2시간

따라 하기

다음은 학생 5명의 어느 날 통학 시간을 조사하여 나타낸 자료이다. 통학 시간의 분산과 표준편차를 구하시오.

통학 시간 (단위: 분)

17, 18, 9, 15, 11

$$\begin{aligned}\text{평균} &= \quad = \quad (\text{분}) \\ \text{분산} &= \quad = \quad \\ \text{표준편차} &= \quad = \quad (\text{분})\end{aligned}$$

☞ 분산: , 표준편차: 분

문제 2

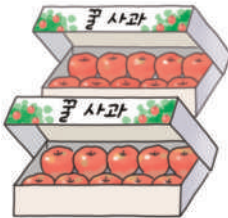
다음은 10개씩 포장된 두 사과 상자 A, B에 들어 있는 사과 무게를 조사하여 나타낸 표이다. 물음에 답하시오. (단, 표준편차는 반올림하여 0.01 g 단위까지 구한다.)

사과 무게

(단위: g)

상자 A	302	294	280	313	306	305	316	286	301	297
상자 B	320	275	292	315	298	279	287	324	306	304

- (1) 각 상자에 들어 있는 사과 무게의 분산과 표준편차를 각각 구하시오.
- (2) 두 상자 중에서 어느 상자에 들어 있는 사과 무게가 더 고른지 말하시오.



의사소통



대포깃 이외에도 산포도를 알아야 하는 경우를 우리 주변에서 찾아 그 이유를 설명하여 보자.

선수 A, B의 최근 5년 동안의 타율 실적

연도	2015	2016	2017	2018	2019	통산 타율
A 선수	0.23	0.32	0.33	0.28	0.24	0.28
B 선수	0.27	0.28	0.29	0.27	0.29	0.28

통산 타율은 똑같은데
누구를 뽑는 것이 좋을까?



1

다음 보기의 설명 중에서 옳은 것을 모두 찾으시오.

보기

- ㄱ. 각 변량의 편차의 총합은 항상 0이다.
- ㄴ. 평균이 클수록 산포도가 커진다.
- ㄷ. 편차는 변량에서 평균을 뺀 값이다.
- ㄹ. 분산은 0이 될 수 없다.

2

다음은 학생 6명이 줄넘기를 한 횟수에 대한 편차를 나타낸 표이다. 줄넘기를 한 횟수의 평균이 90회일 때, C 학생이 줄넘기를 한 횟수를 구하시오.

줄넘기 횟수 (단위: 회)						
학생	A	B	C	D	E	F
편차	-3	-2		4	-8	5



3



다음은 학생 5명이 일주일 동안 휴대 전화로 사용한 무선 데이터 통신량을 조사하여 나타낸 자료이다. 무선 데이터 통신량의 평균과 분산을 구하시오.

무선 데이터 통신량 (단위: MiB)

280, 160, 530, 370, 410



4

다음은 A, B, C 세 반의 중간고사 성적의 평균과 표준편차를 나타낸 표이다. 이때 성적이 가장 우수한 반과 성적이 가장 고른 반을 각각 구하시오.

(단위: 점)

반	A	B	C
평균	74	83	78
표준편차	4	5	2

5

다음은 학생 5명이 1분 동안 측정한 맥박 수에 대한 편차를 나타낸 자료이다. 맥박 수의 표준편차를 구하시오.

(단위: 회)

3, , -6, 5, -4



6

창의·융합



오른쪽은 위도가 비슷한 두 도시 콜카타와 홍콩의 어느 해 1월부터 12월까지 월평균 기온을 조사하여 나타낸 표이다. 어느 도시의 월평균 기온이 더 고른지 설명하시오.

월평균 기온 (단위: °C)

월	콜카타	홍콩
1	20.3	16.4
2	23.3	16.5
3	27.9	19.2
4	30.7	22.9
5	31.2	26.2
6	30.3	28.3
7	29.2	29
8	29.1	28.8
9	29	27.9
10	28	25.6
11	24.4	21.7
12	20.7	18

인도의 콜카타



홍콩





산점도와 상관관계

자료를 산점도로 나타내고, 이를 이용하여 상관관계를 말할 수 있다.

여름에는 기온이 오를수록 대체로 전력 사용량이 증가한다.



탐구 학습

산점도는 무엇인가요?



열기

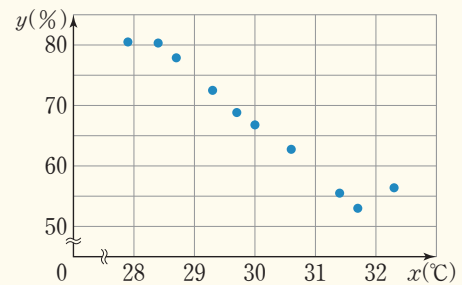
다음은 2018년 8월 1일부터 8월 10일까지 어느 도시의 일평균 기온 x °C와 일평균 습도 y %를 조사하여 나타낸 표이다. 두 변량 x, y 의 순서쌍 (x, y) 를 좌표로 하는 점을 좌표 평면 위에 나타내 보자.

날짜	1일	2일	3일	4일	5일	6일	7일	8일	9일	10일
일평균 기온(°C)	31.4	31.7	32.3	30.6	29.3	28.7	30	29.7	27.9	28.4
일평균 습도(%)	55.5	53	56.4	62.8	72.5	77.9	66.8	68.8	80.5	80.3



다지기

두 변량 x, y 의 순서쌍 (x, y) 를 좌표로 하는 점을 좌표평면 위에 나타내면 오른쪽 그림과 같다.



〈그림 1〉



키우기

통계 자료에서 두 변량 사이의 관계를 그래프로 나타낸 것을 무엇이라고 할까?

산점도

탐구 학습의 〈그림 1〉과 같이 통계 자료에서 두 변량 사이의 관계를 그래프로 나타내면 그 관계를 쉽게 알아볼 수 있는 경우가 있다.

이처럼 두 변량의 순서쌍을 좌표로 하는 점을 좌표평면 위에 나타낸 그래프를 **산점도**라고 한다.

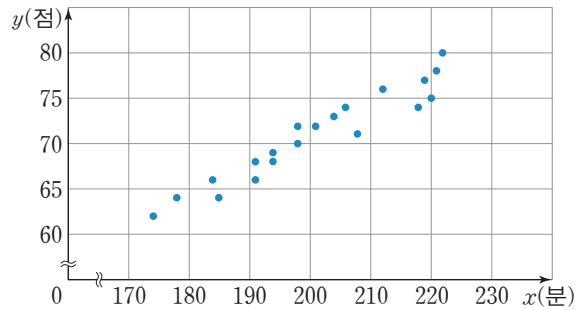
〈그림 1〉의 산점도에서 점들이 어느 정도 흩어져 있기는 하지만 대체로 일평균 기온이 높을수록 일평균 습도가 낮아지는 경향이 있음을 알 수 있다.

예제 1

다음은 효정이네 반 학생들의 하루 평균 공부 시간과 수학 점수를 조사하여 나타낸 표이다. 하루 평균 공부 시간과 수학 점수에 대한 산점도를 그리시오.

시간(분)	점수(점)	시간(분)	점수(점)	시간(분)	점수(점)	시간(분)	점수(점)
208	71	206	74	191	68	220	75
218	74	221	78	201	72	194	68
174	62	198	70	212	76	194	69
185	64	222	80	178	64	219	77
204	73	198	72	184	66	191	66

풀이 하루 평균 공부 시간을 x 분, 수학 점수를 y 점이라고 할 때, 하루 평균 공부 시간과 수학 점수에 대한 산점도를 그리면 다음과 같다.

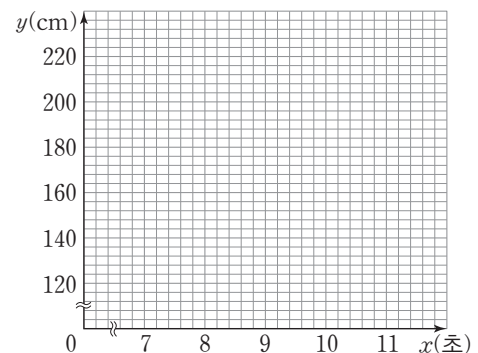


답 풀이 참조

문제 1

다음은 정훈이네 반 학생들의 50 m 달리기 기록 x 초와 제자리멀리뛰기 기록 y cm를 조사하여 나타낸 표이다. 50 m 달리기 기록과 제자리멀리뛰기 기록에 대한 산점도를 그리시오.

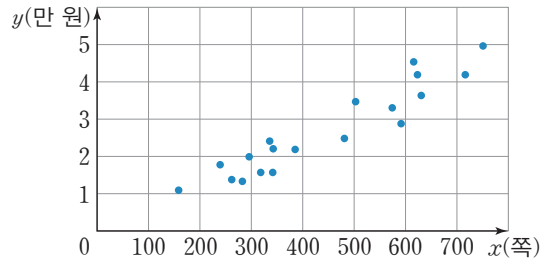
x (초)	y (cm)	x (초)	y (cm)
9.8	140	7.8	208
8.8	184	8	188
9.8	156	8.6	180
10.2	132	9.6	168
10.4	120	10.2	156
10.6	124	11	144
11	116	7	228
9	176	7.6	204



상관관계는 무엇인가요?

상관관계

다음 그림은 어느 서점에서 판매하는 책의 쪽수 x 와 가격 y 만 원을 조사하여 산점도로 나타낸 것이다.



이 산점도에서 점들이 어느 정도 흩어져 있기는 하지만 대체로 책의 쪽수가 많을수록 책의 가격은 비싸진다는 것을 알 수 있다. 즉, 변량 x 의 값이 증가하면 변량 y 의 값도 증가하는 경향이 있음을 알 수 있다.

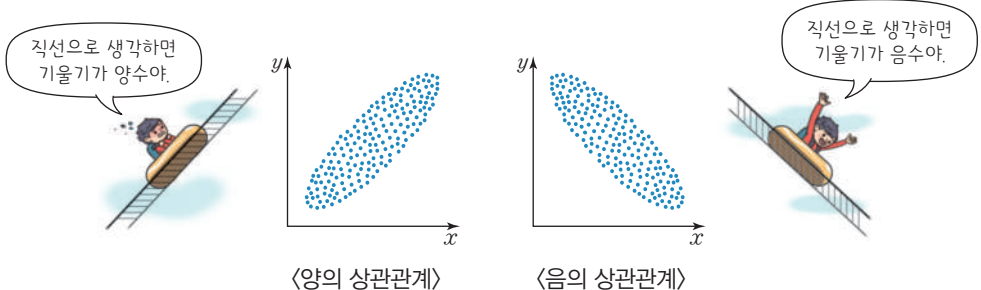
이처럼 두 변량 x, y 사이에 x 의 값이 증가함에 따라 y 의 값이 증가하거나 감소하는 경향이 있을 때, 두 변량 x, y 사이에 **상관관계**가 있다고 한다.

두 변량 x, y 에 대한 산점도에서 x 의 값이 증가함에 따라 y 의 값도 대체로 증가하는 경향이 있을 때, 두 변량 x, y 사이에 양의 상관관계가 있다고 한다.

이와 반대로 두 변량 x, y 에 대한 산점도에서 x 의 값이 증가함에 따라 y 의 값이 대체로 감소하는 경향이 있을 때, 두 변량 x, y 사이에 음의 상관관계가 있다고 한다.

이전에 배운 내용

- 일차함수 $y = ax + b$ 의 성질
- ① $a > 0$ 일 때, 그래프는 오른쪽 위로 향하는 직선이다.
 - ② $a < 0$ 일 때, 그래프는 오른쪽 아래로 향하는 직선이다.



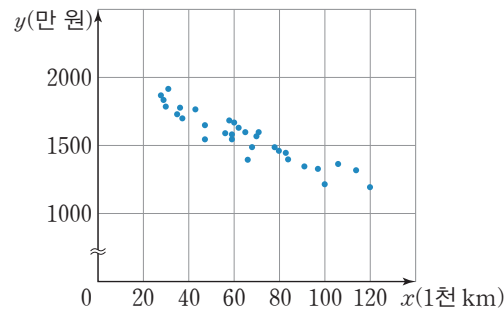
예제 2

다음은 2015년에 생산된 A 자동차의 3년 후 주행 거리에 따른 중고 자동차 가격을 조사하여 나타낸 표이다. 주행 거리와 중고 자동차 가격에 대한 산점도를 그리고 상관관계를 조사하시오.

(단, 표에서 거리의 단위는 1천 km, 가격의 단위는 1만 원이다.)

거리	가격	거리	가격	거리	가격	거리	가격
28	1870	47	1550	65	1600	84	1400
29	1840	47	1650	66	1397	91	1350
30	1790	56	1590	68	1490	97	1330
31	1920	58	1690	70	1570	97	1330
35	1730	59	1580	71	1600	100	1220
36	1780	59	1550	78	1490	106	1370
37	1700	60	1670	80	1465	114	1320
43	1770	62	1630	83	1450	120	1200

풀이 주행 거리 x 천 km와 중고 자동차 가격 y 만 원에 대한 산점도를 그리면 다음과 같다. 주행 거리가 늘어남에 따라 중고 자동차 가격은 대체로 낮아지는 경향이 있으므로 음의 상관관계가 있다.



풀이 참조

문제 2

다음은 신정이가 10일 동안 하루에 걸은 거리와 걸으면서 소모한 열량을 조사하여 나타낸 표이다. 걸은 거리 x km와 소모한 열량 y kcal에 대한 산점도를 그리고 상관관계를 조사하시오.

거리(km)	3.2	1.2	2.4	1.6	5.4	1.4	3	2.6	1.2	4.2
열량(kcal)	170	70	110	90	290	70	150	130	60	220

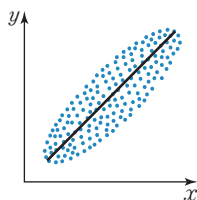
여러 가지 상관관계

두 변량 사이에 양 또는 음의 상관관계가 있는 산점도에서 점들이 한 직선 주위에 가까이 몰려 있을수록 상관관계는 강하다고 하고, 멀리 흩어져 있을수록 상관관계는 약하다고 한다.

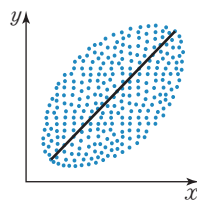
다음 산점도에서 ①, ②인 경우는 두 변량 x, y 사이에 양의 상관관계가 있으며, ①이 ②보다 강한 상관관계가 있다. 또, ③, ④인 경우는 두 변량 x, y 사이에 음의 상관관계가 있으며, ③이 ④보다 강한 상관관계가 있다.

양의 상관관계

① 강한 경우

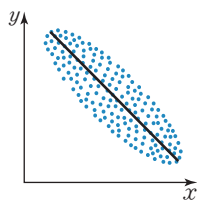


② 약한 경우

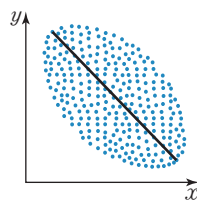


음의 상관관계

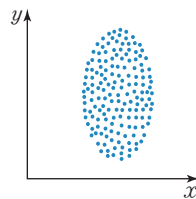
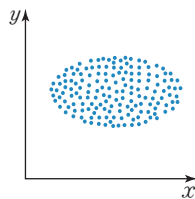
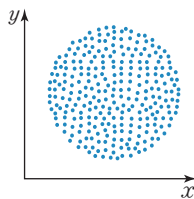
③ 강한 경우



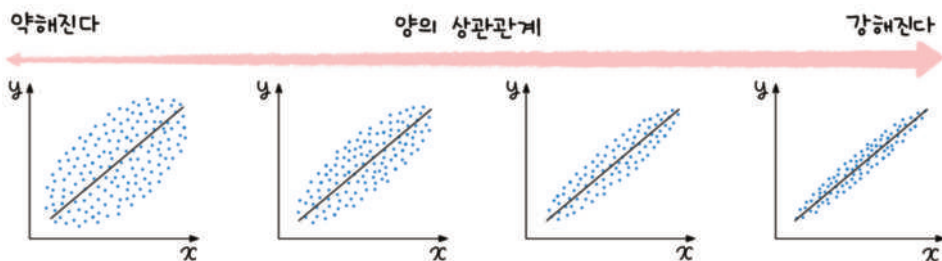
④ 약한 경우



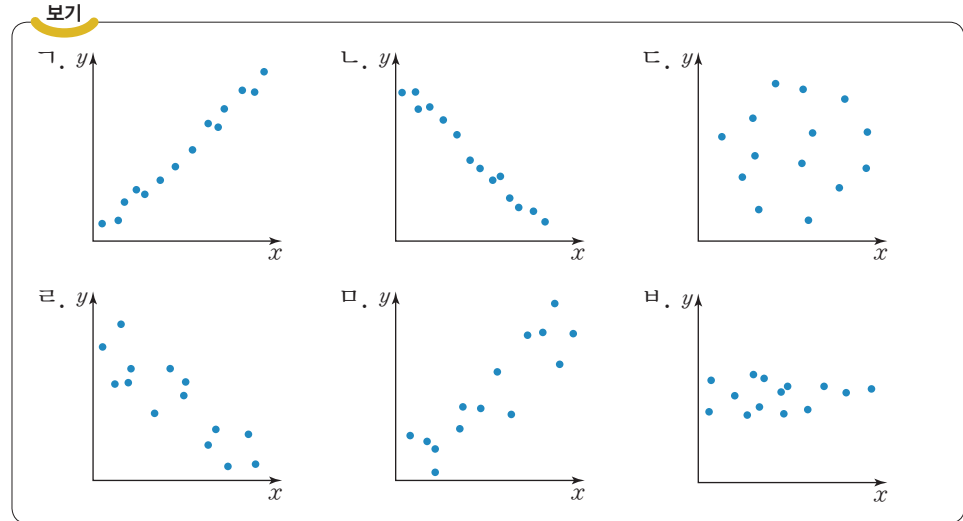
한편, x 의 값이 증가함에 따라 y 의 값이 증가하는 경향이 있는지 감소하는 경향이 있는지 분명하지 않을 때는 두 변량 x, y 사이에는 상관관계가 없다고 한다. 다음 그림과 같이 점들이 흩어져 있거나 좌표축에 평행한 경우에는 상관관계가 없다.



개념 확인



문제 3 다음 보기의 산점도에 대하여 물음에 답하십시오.



- (1) 양의 상관관계가 있는 산점도를 모두 찾으시오.
- (2) 음의 상관관계가 있는 산점도를 모두 찾으시오.
- (3) 상관관계가 없는 산점도를 모두 찾으시오.

數

수준수준



상관관계와 인과 관계

여름에 기온이 오르면 시원한 것을 찾는 사람들이 늘면서 아이스크림의 판매량이 늘고, 또 더위를 피해 여행을 떠나는 피서객의 수도 늘어난다. 따라서 기온과 아이스크림 판매량 사이에는 양의 상관관계가 있고, 기온과 피서객의 수 사이에도 양의 상관관계가 있다. 이처럼 기온이 오르는 것은 아이스크림의 판매량과 피서객의 수가 늘어나는 것의 원인이 될 수 있다.

여기서 아이스크림의 판매량과 피서객의 수 사이에는 어떤 관계가 있을까? 기온이 오르면서 아이스크림의 판매량도 늘고 피서객의 수도 같이 늘어 양의 상관관계가 있음을 알 수 있다. 그러나 아이스크림의 판매량이 늘어났기 때문에 피서객의 수가 늘어난다고는 할 수 없다. 즉, 아이스크림의 판매량 증가가 피서객의 수가 늘어나는 원인이라고는 할 수 없다. 이를 상관관계는 있으나 인과 관계는 없다고 한다.





스스로 확인하기

정답 및 풀이 269쪽

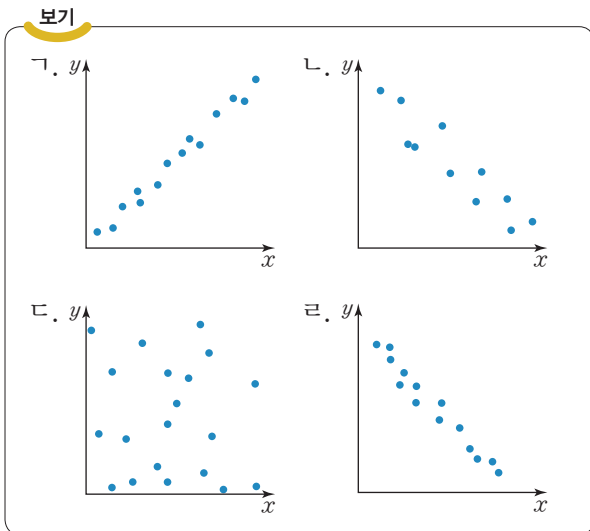
1

오른쪽은 프로 야구 선수 10명이 한 시즌 동안 출전한 경기 수와 안타 수를 조사하여 나타낸 표이다. 출전한 경기 수 x 회와 안타 수 y 개에 대한 산점도를 그리시오.

경기 수(회)	안타 수(개)
116	128
122	136
128	126
130	136
132	154
134	162
136	148
138	175
142	176
144	175

2

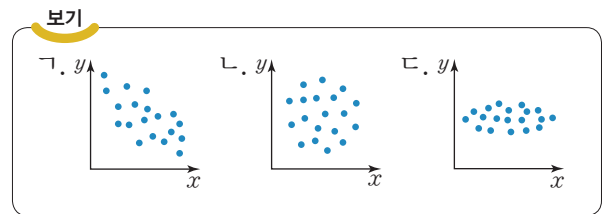
다음 보기의 산점도에 대하여 물음에 답하시오.



- (1) 양의 상관관계가 있는 산점도를 모두 찾으시오.
- (2) 음의 상관관계가 있는 산점도를 모두 찾으시오.

3

다음 보기의 산점도 중에서 상관관계가 없는 것을 모두 찾으시오.



4 창의·융합

다음은 어느 분식집에서 15일 동안 판매된 떡볶이, 순대, 라면의 양을 조사하여 나타낸 표이다. 물음에 답하시오.

- (1) 떡볶이의 판매량 x 인분과 순대의 판매량 y 인분에 대한 산점도를 그리고 상관관계를 조사하시오.
- (2) 떡볶이의 판매량 x 인분과 라면의 판매량 y 인분에 대한 산점도를 그리고 상관관계를 조사하시오.

떡볶이(인분)	순대(인분)	라면(인분)
19	23	27
27	26	21
38	36	16
37	45	18
44	41	10
42	48	15
44	45	17
39	41	17
29	33	21
31	31	25
27	36	22
40	36	12
23	31	24
32	40	19
23	27	27

대푯값과 산포도 구하기

이지통계 소프트웨어를 이용하면 자료의 대푯값과 산포도를 편리하게 구할 수 있다.

오른쪽은 2016년 리우데자네이루 올림픽 예선에 출전한 우리나라 여자 배구 선수 20명의 키를 조사하여 나타낸 자료이다.

(대한민국배구협회, 2016년)

여자 배구 선수의 키 (단위: cm)

192	181	186	177	176
177	182	177	179	185
189	168	187	190	173
170	180	190	179	163



이 자료의 대푯값과 산포도를 다음 순서에 따라 구하여 보자.

- ① 이지통계 소프트웨어에 접속하여 아래와 같이 자료를 입력한다.

No.	자료
1	192
2	181
3	186
4	177
5	176
6	177
7	182
8	177
9	179
10	185
합계	3601

- ② **통계값**을 선택하면 아래와 같이 대푯값을 구할 수 있는 메뉴가 나타난다.

No.	자료
1	192
2	181
3	186
4	177
5	176
합계	3601

- ③ 각 메뉴를 선택하면 대푯값을 구할 수 있다.

No.	자료
1	192
2	181
3	186
4	177
5	176
합계	3601

- ④ ③에서 **↓**를 선택한 후 분산, 표준편차를 선택하면 아래와 같은 결과를 얻을 수 있다.

No.	자료	편차	(편차) ²
1	192	11.95	142.8025
2	181	0.95	0.9025
3	186	5.95	35.4025
4	177	-3.05	9.3025
5	176	-4.05	16.4025
합계	3601	0	1166.95

수행 과제

문제 해결

스포츠와 관련된 자료를 조사한 후 이지통계 소프트웨어를 이용하여 대푯값과 산포도를 구하시오.

개념 콕콕

1 대푯값

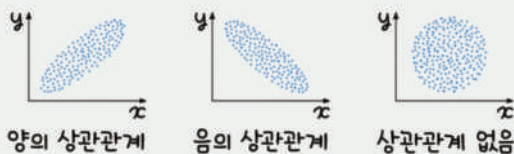
- (1) 대푯값: 자료 전체의 특징을 대표적으로 나타내는 값
- (2) 중앙값: 변량을 작은 값부터 크기순으로 나열할 때 중앙에 놓인 값
 - ① 변량의 개수가 홀수이면 중앙에 놓인 값
 - ② 변량의 개수가 짝수이면 중앙에 놓이는 두 값의 평균
- (3) 최빈값: 변량 중 가장 많이 나타나는 값

2 산포도

- (1) 산포도: 변량이 흩어져 있는 정도를 하나의 수로 나타낸 값
- (2) 분산과 표준편차
 - ① (편차) = (변량) - (평균)
 - ② (분산) = $\frac{\{(\text{편차})^2\text{의 총합}\}}{(\text{변량의 개수})}$
 - ③ (표준편차) = $\sqrt{(\text{분산})}$

3 산점도와 상관관계

- (1) 산점도: 두 변량의 순서쌍을 좌표로 하는 점을 좌표 평면 위에 나타낸 그래프
- (2) 상관관계: 두 변량 x, y 사이에 x 의 값이 증가함에 따라 y 의 값이 증가하거나 감소하는 경향이 있을 때, 두 변량 x, y 사이에 상관관계가 있다고 한다.



01 다음 자료의 중앙값과 최빈값을 구하시오.

- (1) 1, 4, 2, 1, 3, 3, 1, 3, 2
- (2) 34, 25, 13, 12, 50, 60, 20, 25, 15, 46

02 다음은 민찬이네 반 학생들의 오래 매달리기 기록을 조사하여 나타낸 줄기와 잎 그림이다. 중앙값을 a 초, 최빈값을 b 초라고 할 때, $a - b$ 의 값을 구하시오.

오래 매달리기 기록 (0이9는 9초)	
줄기	잎
0	9
1	2 2 6 7 7 7 9
2	3 3 5 5 6
3	0 2

03 다음은 변량 8개를 작은 값부터 크기순으로 나열한 것이다. 이 자료의 중앙값이 20일 때, x 의 값을 구하시오.

11, 13, 15, x , 23, 25, 28, 30



- 04 다음은 지혜네 모둠 10명이 여름 방학 동안 읽은 책의 수를 조사하여 나타낸 자료이다. 이 자료의 평균이 6권일 때, 최빈값을 구하시오.

읽은 책의 수 (단위: 권)

7, 8, 3, 5, 5, x , 6, 4, 9, 8



- 05 다음은 학생 6명이 농구의 자유투를 10회 던졌을 때, 성공한 횟수에 대한 편차를 나타낸 표이다. 물음에 답하시오.

(단위: 회)

학생	민혁	정욱	중훈	승민	경모	재규
편차	-2	4	-1	-1	2	x

- x 의 값을 구하시오.
- 학생 6명의 자유투 성공 횟수의 분산을 구하시오.

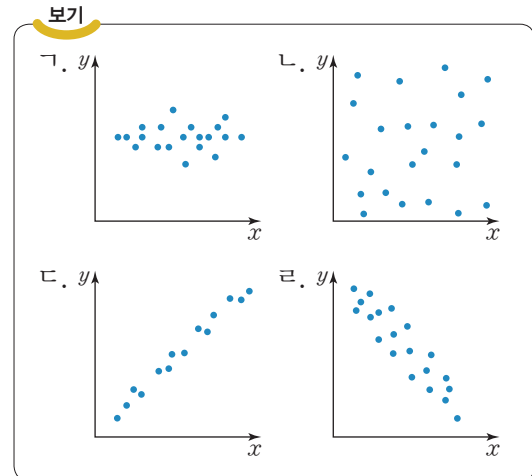
- 06 다음은 학생 5명이 제기차기를 한 횟수를 조사하여 나타낸 자료이다. 제기차기 횟수의 분산과 표준편차를 구하시오.

제기차기 횟수 (단위: 회)

12, 13, 16, 19, 20

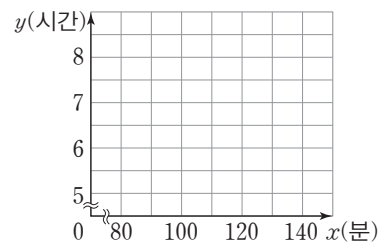


- 07 다음 보기 중에서 양의 상관관계가 있는 산점도를 찾으시오.



- 08 다음은 일주일 동안 영철이의 컴퓨터 사용 시간 x 분과 잠자는 시간 y 시간을 조사하여 나타낸 표이다. 컴퓨터 사용 시간과 잠자는 시간에 대한 산점도를 그리고 상관관계를 조사하시오.

x (분)	80	90	100	115	120	125	140
y (시간)	8	7.5	7.5	6.5	6	5.5	5



서술형

- 09 다음은 학생 8명이 텃걸이를 한 횟수를 조사하여 나타낸 자료이다. 이 자료의 중앙값이 11회, 최빈값이 9회일 때, $a+b+c$ 의 값을 구하시오.

텃걸이 횟수 (단위: 회)

$a, b, c, 9, 12, 15, 15, 20$



- 10 3개의 수 a, b, c 의 평균이 16, 분산이 14일 때, 5개의 수 $a, b, c, 13, 19$ 의 평균과 분산을 구하시오.

사고력 높이기

- 11 6개의 수 9, 12, 13, 18, x, y 의 중앙값과 평균이 모두 14일 때, x, y 의 값을 구하시오. (단, $x < y$)

- 12 다음은 두 반 A, B의 수학 성적의 평균과 표준편차를 나타낸 표이다. 물음에 답하시오.

반	학생 수(명)	평균(점)	표준편차(점)
A	26	75	9
B	24	75	16

- (1) 두 반 중에서 어느 반의 수학 성적이 더 큰지 말하시오.
 (2) 두 반 전체 학생의 수학 성적의 평균과 분산을 구하시오.

학습 내용 점검

1. 대푯값 ▶01, 02, 03, 04, 09, 11번
 2. 산포도 ▶05, 06, 10, 12번
 3. 산점도와 상관관계 ▶07, 08번



학습 태도 점검

흥미도 ☆☆☆☆☆

집중도 ☆☆☆☆☆

참여도 ☆☆☆☆☆

협동심 ☆☆☆☆☆

나의 학습 일기

이 단원을 배우고 나서 새롭게 알게 된 점이나 부족한 점을 적어 보세요.

다음과 같은 순서로 자료를 조사한 후 산점도를 그려 상관관계를 알아보고 보고서를 작성하여 보자.

순서

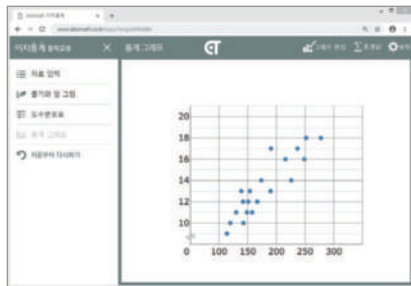
- ① 상관관계가 있을 것으로 예상하는 두 변량을 정한다.



- ② 조사 대상을 선정하여 자료를 조사한다.

전기 사용량(kWh)	수도물 사용량(m ³)
143	11
149	12
174	15
139	14
⋮	⋮

- ③ 이지통계 소프트웨어를 이용하여 산점도를 그려 상관관계를 알아본다.



- ④ 조사 목적, 조사 대상, 결과 등을 넣어 보고서를 작성한다.



탐구 과제 위와 같은 방법으로 모둠을 나누어 자료를 조사한 후 산점도를 그려 상관관계를 알아보고, 보고서를 작성하여 보자.

