IV 통계

IV
통계

1 대푯값과 산포도

학습한 내용: 중학교 수학 1 도수분포와 그래프 → 본 단원의 내용: 1. 대푯값과 산포도 → 학습할 내용: 고등학교 확률과 통계 확률분포

■ 단원 배경

'전자 상거래 동향', '청소년의 운동 시간', '사람들의 하루 수면 시간', '소비자가 가장 선호하는 제품', '선수의 경기 결과의 분석' 등 우리는 주변에서 다양한 자료를 접할 수 있다. 이러한 자료들을 정리하고 분석하여 실생활에 유용한 정보를 얻으려면 그 자료의 대푯값뿐만 아니라 자료가 어떻게 분포하고 있는지 등을 정확하게 파악하여야 한다.

1 단원을 들어가면서

경제 성장률, 평균 수명, 취업률, 학업 성취율, 수면 시간, 옷의 치수 등 통계는 일상생활의 곳곳에서 활용되고 있다.

예를 들면 평균 수명에 관한 통계 자료는 보험 회사에서 보험 상품을 개발할 때 이용할 수 있고, 경제 성장률, 취업률, 학업 성취율에 관한 통계 자료는 국가 정책을 수립할 때 활용할 수 있다.

이 밖에도 통계는 전자 및 통신 공학, 컴퓨터 관련 업무, 회계 · 금융업, 여론 조사 등 다양한 분야에서 활용되고 있다.

이 단원에서는 통계의 가장 기본적인 내용인 대푯값과 산포도로 구성되어 있다.

대푯값의 의미를 이해하고, 평균, 중앙값, 최빈값을 구할 수 있도록 한다. 또 산포도의 의미를 이해하고, 편차, 분산, 표준편차를 구할 수 있도록 한다.

2 단원의 지도 목표

1. 대푯값과 산포도

(1) 평균, 중앙값, 최빈값의 의미를 이해하고, 이를 구할 수 있게 한다.
(2) 편차, 분산, 표준편차의 의미를 이해하고, 이를 구할 수 있게 한다.

3 단원의 교수·학습상의 유의점

1. 대푯값과 산포도

(1) 다양한 상황에서 자료를 수집하게 하고, 수집한 자료가 적절한지 판단하는 활동을 하게 한다.
(2) 자료의 특성에 따라 적절한 대푯값을 선택하여 구할 수 있게 한다.
(3) 공학적 도구를 이용하여 대푯값과 산포도를 구할 수 있게 한다.

4 단원의 지도 계통

학습한 내용		본 단원의 내용	학습할 내용	
초등학교 수학 5~6	- 가능성과 평균 - 자료의 표현 - 비율그래프	**1. 대푯값과 산포도** 1-1 대푯값 1-2 산포도	고등학교 확률과 통계	- 확률분포 - 통계적 추정
중학교 수학 ❶	- 줄기와 잎 그림 - 도수분포표, 히스토그램, 도수분포다각형 - 도수분포표로 주어진 자료의 평균			

5 단원의 이론적 배경

1. 통계의 역사적 배경

통계학(Statistics)이란 단어는 프랑스어인 status(신분, 상태)에서 유래되었으며, 이 단어는 원래 국가(state)의 상태를 조사하고 연구하는 것을 의미하며, 중세기 이후에는 정치적인 의미로서 국가(state)를 지칭하게 되었다.

영국의 그랜트(Graunt, J. : 1620~1674)는 자료를 정리하면 '어떤 병으로 사망하는 사망자 수와 전체 사망자 수의 비가 일정하다.', '남녀의 수는 전체 인구에 대하여 거의 같다.' 등의 법칙을 발견할 수 있다고 주장하였다.

한편, 헬리(Halley, E. ; 1656~1742)는 자료를 정리하여 생명표를 만들어서 당시 선원들의 생명 보험의 보험료 산정에 크게 기여하였다.

그 이후 18세기 중엽부터는 확률의 개념이 통계학에 영향을 주어 통계학의 이론을 발전시켰다.

프랑스에서는 파스칼(Pascal, P. ; 1601~1665) 등의 수학자에 의해 정립된 도박의 수리 이론이나 베르누이(Bernoulli, J. ; 1654~1705)의 큰 수의 발견 등이 종합되어 라플라스(Laplace, P. S. ; 1749~1827)에 이르러 고전 확률론의 완성을 보게 되었다. 가우스(Gauss, K. F.; 1777~1855)는 오차분포에 대한 정규곡선 식을 발표하였고, 피어슨(Pearson, K. ; 1857~1936)은 기술통계학을 완전히 정립하였다. 현대 통계 이론인 추측통계학은 통계학자인 피셔(Fisher, R. A. ; 1890~1962)의 실험 계획에 관한 이론과 통계적 추정 이론의 발표로 발달하게 되었다.

20세기부터 통계학의 이론은 급속히 발전하기 시작하여 제2차 세계 대전 이후에는 대량 생산에서의 품질 관리와 표본조사 등에 통계적 방법을 적용하였다. 피셔의 추측통계학은 기존의 기술통계학과 구별되어 오늘날 수리 통계학의 발전에 크게 기여하였다.

2. 통계학의 분류

통계학은 기술통계학과 추측통계학의 두 가지 분야로 구분할 수 있다.

기술통계학은 수집한 자료를 표나 그림 또는 대푯값으로 요약하는 방법을 말한다.

추측통계학은 모집단에서 추출한 표본의 정보를 이용하여 모집단의 여러 가지 특성을 과학적으로 추론하는 방법을 말한다. 일반적으로 표본에서 얻어진 정보는 모집단의 특성에 대하여 완전한 정보를 갖고 있지 않다. 따라서 모집단의 특성을 추정할 때에는 약간의 오차가 발생할 수도 있다. 추측통계학은 확률적인 방법을 이용하여 모집단의 일반적인 특성을 찾아내는 데에서 발생할 수 있는 이러한 오차까지를 포괄한 것이다.

3. 여러 가지 평균

(1) 산술평균

n개의 변량 x_1, x_2, x_3, ……, x_n의 산술평균을 $\overline{x}$라 할 때,

$$\overline{x}=\frac{x_1+x_2+x_3+\cdots\cdots+x_n}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

이때, 변량 x_1, x_2, x_3, ……, x_n의 도수가 각각 f_1, f_2, f_3, ……, f_n이고 총 도수가 $\sum_{i=1}^{n}f_i=N$일 때,

$$\overline{x}=\frac{x_1f_1+x_2f_2+x_3f_3+\cdots\cdots+x_nf_n}{N}$$

$$=\frac{\sum_{i=1}^{n}x_if_i}{N}$$

(2) 기하평균

n개의 변량 x_1, x_2, x_3, ……, x_n의 기하평균을 G라 하면

$$G=\sqrt[n]{x_1x_2x_3\cdots\cdots x_n}$$

이다.

(3) 조화평균

n개의 변량 x_1, x_2, x_3, ……, x_n의 조화평균을 H라 하면

$$H=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots\cdots+\frac{1}{x_n}}$$

$$=\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

이다.

(4) 제곱평균

n개의 변량 x_1, x_2, x_3, ……, x_n의 제곱평균을 R라 하면

$$R=\sqrt{\frac{x_1^{\ 2}+x_2^{\ 2}+x_3^{\ 2}+\cdots\cdots+x_n^{\ 2}}{n}}$$

$$=\sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n}}$$

이다.

일반적으로 모든 변량이 양수일 때, 위의 평균들 사이에는 다음과 같은 대소 관계가 성립한다.

$H\le G\le \overline{x}\le R$

4. 산포도

자료를 활용할 때 대푯값만으로는 그 자료의 특징을 충분히 알 수 없는 경우가 있다. 이러한 자료에서는 자료의 분포 상태를 알아보기 위하여 분산과 표준편차를 많이 이용한다.

분산이란 (편차)=(변량)−(평균)을 구한 뒤, 편차의 제곱의 평균을 구한 것이다. 이때 분산의 음이 아닌 제곱근을 표준편차라고 한다.

자료에서 표준편차가 클수록 변량들은 평균을 중심으로 멀리 흩어져 있고, 표준편차가 작을수록 변량들은 평균을 중심으로 몰려 있다.

n개의 변량 x_1, x_2, x_3, ……, x_n의 평균을 $\overline{x}$라 할 때, 분산 s^2, 표준편차 s는 다음과 같다.

$$s^2=\frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+(x_3-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2}{n}$$

$$=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\overline{x}^2$$

$$s=\sqrt{\frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+(x_3-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2}{n}}$$

$$=\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2-\overline{x}^2}$$

도수분포표에서 분산과 표준편차를 구하면 다음과 같다.

계급값	x_1	x_2	$\cdots$	x_n	합계
도수	f_1	f_2	$\cdots$	f_n	N

$$s^2=\frac{(x_1-\overline{x})^2f_1+(x_2-\overline{x})^2f_2+(x_3-\overline{x})^2f_3+\cdots\cdots+(x_n-\overline{x})^2f_n}{N}$$

$$=\frac{1}{N}\sum_{i=1}^{n}x_i^2f_i-\overline{x}^2$$

$$s=\sqrt{\frac{(x_1-\overline{x})^2f_1+(x_2-\overline{x})^2f_2+(x_3-\overline{x})^2f_3+\cdots\cdots+(x_n-\overline{x})^2f_n}{N}}$$

$$=\sqrt{\frac{1}{N}\sum_{i=1}^{n}x_i^2f_i-\overline{x}^2}$$

$$\left(\text{단, } N=\sum_{i=1}^{n}f_i,\ \overline{x}=\frac{1}{N}\sum_{i=1}^{n}x_i^2f_i\right)$$

5. 최근 통계 교육의 경향

최근 통계 교육은 이론적인 정당성보다는 실제의 자료를 과학적인 방법으로 정리하고 분석하여, 최대한 오차를 줄이면서 신뢰성 있는 통계적 결론을 추론하는 능력을 키우는 데 중점을 두고 있다.

또한 어떤 통계 자료를 접했을 때, 이 통계 자료가 어떤 경로로 수집되었는지를 육하원칙(언제, 어디서, 누가, 무엇을, 어떻게, 왜)에 의하여 따져 보고 해석해 보아야만 통계의 함정에 빠지지 않을 수 있음을 강조하고 있다.

한편 통계 교육을 할 때에는 통계 자료를 쉽게 처리하기 위하여 계산기나 컴퓨터 및 컴퓨터 프로그램 등을 사용하는 방법을 알려 주고, 이의 사용을 적극 권장해야 한다.

참고 문헌

- 우정호, 통계 교육의 개선 방향 탐색, 대한 수학교육 학회지 학교 수학 2(1), 2000.
- 김응태 · 김용국, 수학 교육 교재론, 이우출판사, 1980.
- 박세희, 수학의 세계, 서울대학교 출판부, 1985.
- 이해용, 이필영, 통계학 입문, 자유아카데미, 1996.
- H. Eves(이우영, 신항균 역), 수학사, 경문사, 2005.
- T. M. Porter, The rise of statistical thinking, Princeton University Press, 1986.

6 단원의 지도 계획

본문 내용		지도 내용	용어와 기호	쪽수	차시
1. 대푯값과 산포도	이야기로 들려주는 통계, 준비하기			156～158	1
	1-1. 대푯값	• 대푯값 • 중앙값 • 최빈값	대푯값, 중앙값, 최빈값	159～164	2
	1-2. 산포도	• 산포도 • 편차, 분산, 표준편차 • 도수분포표로 정리된 자료의 분산과 표준편차	산포도, 분산, 표준편차	165～172	3
	중단원 마무리하기, 창의 · 인성 키우기, 컴퓨터로 하는 수학			173～178	1
	단원 마무리하기			179～180	1
	수행 과제, 수학으로 세상 읽기, 스스로 평가하기			181～183	1
	소계				9

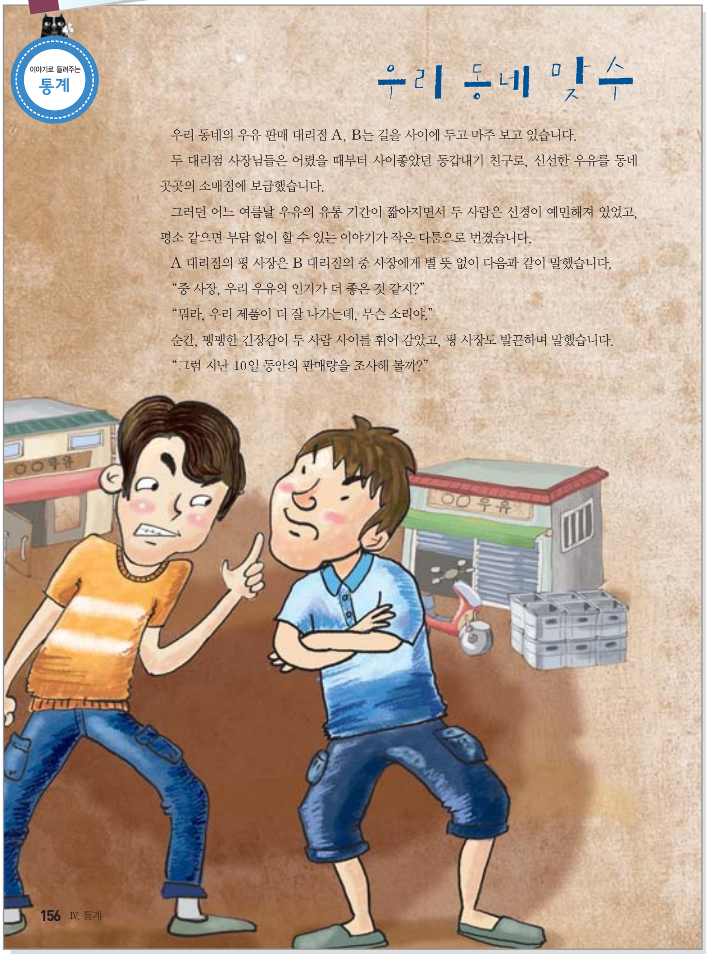
7 교수 · 학습 과정 예시안

단원	Ⅳ. 통계 1. 대푯값과 산포도 01. 대푯값	교과서 쪽수	162~163	차시	3/9
학습 주제	최빈값이란 무엇일까?	학습 목표	최빈값의 의미를 이해하고, 이를 구할 수 있다.		
준비물	활동지, PPT 자료, 과제물			교과 관련	

단계(시간)	학습 내용 및 학습 방법	교수 · 학습 활동		학습 자료 및 유의점
		교사	학생	
도입(7분)	**개념 학습**(전체 학습) – 전시 학습 제시	지난 시간에 **생각 열기**로 함께 학습한 대푯값을 상기시키고, 중앙값에 대한 학습 정도를 파악한다.	지난 시간에 함께 풀어본 문제를 통해 전시 내용을 상기한다.	
	탐구 학습 – 생각 열기	**생각 열기**를 통하여 중앙값 이외의 가장 많이 나온 바지의 치수가 또 다른 대푯값으로 쓰일 수 있음을 소개한다.	주어진 자료를 보고 중앙값이 아닌 또 다른 값이 대푯값으로 쓰일 수 있음을 생각해 본다.	▶ PPT 자료
	– 학습 목표 제시	**생각 열기**와 관련하여 학습 목표를 각자 생각해 보게 한 후 제시한다.	이번 시간에 학습할 내용이 무엇인지 생각해 보고, 학습 목표를 선생님과 함께 제시한다.	▶ PPT 자료
전개(30분)	**개념 학습**(전체 학습) – 최빈값	실생활 속의 소재에서 가장 많이 나타나는 수로 최빈값이라는 것을 이해하도록 설명한다.	최빈값은 자료에서 가장 많이 나타나는 값으로 이해한다.	최빈값은 두 개 이상 나올 수도 있음을 알게 한다.
	문제 해결 학습(개별 학습) – 문제 4	문제를 통해 최빈값을 정확히 인지하고 다른 소재의 문제에도 적용할 수 있도록 지도한다.	문제 4를 해결하면서 최빈값을 정확히 이해한다.	
	문제 해결 학습(협력 학습) – 문제 5	• 실생활 문제를 통해 최빈값과 중앙값을 구해 보면서 차이를 알 수 있도록 유도한다. • 모둠별로 발표를 통해 활동 내용을 확인할 수 있도록 한다.	• 모둠별로 최빈값과 중앙값을 구하고 그 특징을 토의한다. • 활동한 내용을 발표하고, 다른 모둠의 발표를 경청하며 비교해 본다.	• 모둠별 활동을 하는 동안 순회하며 지도한다. ▶ 활동지

단계(시간)	학습 내용 및 학습 방법	교수 · 학습 활동		학습 자료 및 유의점
		교사	학생	
	수준별 학습(개별 학습)	하 간단한 예시 문제를 통하여 최빈값의 의미를 알고 최빈값을 구할 수 있도록 지도한다. 상 생활 속에서 사례를 말하고, 그것이 평균, 중앙값, 최빈값 중 어떤 것이 대푯값으로 적절한지 토론하게 한다.	실생활에서 대푯값으로 사용되는 평균, 중앙값, 최빈값의 예를 찾아보고 발표한다.	중앙값과 최빈값의 이용의 차이를 확실히 이해시킨다.
	문제 해결 학습(전체 학습) – 함께 풀기 2	문제를 통하여 도수분포표에서 중앙값과 최빈값을 구할 수 있도록 한다.	**함께 풀기** 문제의 설명을 주의 깊게 들으며 선생님의 풀이에 함께 참여한다.	
	문제 해결 학습(개별 학습) – 문제 6, 7	• 여러 가지 문제를 통해 도수분포표에서 중앙값과 최빈값을 구할 수 있는지 확인한다. • 문제 해결 후 학생들이 발표할 수 있도록 한다.	각자 문제를 해결하고 해결하기 힘든 학생은 손을 들어 선생님께 질문한다.	
정리 및 평가 (8분)	**개념 정리**(전체 학습) – 내용 정리	학생들과 함께 배운 내용을 정리할 수 있도록 적절히 발문한다. – 발문: 중앙값과 최빈값은 어떠한 값인가요?	선생님과 함께 배운 내용을 정리한다. – 예상 답변 ① 중앙값은 중앙에 위치한 값이에요. ② 최빈값은 가장 많이 나오는 값이에요. ③ 극단적인 값이 있는 경우 평균 대신 중앙값을 쓸 수 있어요.	▶ PPT 자료
	문제 해결 학습(개별 학습) – 형성 평가 문제 (기초 1문제 / 기본 1문제 / 실력 1문제)	형성 평가를 통해 학생들이 수업 목표에 얼마나 도달했는지 확인한다.	형성 평가 문제를 각자 해결하면서 학습 도달 여부를 파악한다.	▶ 형성 평가지
	수준별 수업(개별 학습) – 수준별 과제 제시	수준별 과제를 부여한다.	수준별 과제를 받아간다.	▶ 과제물
	차시 예고 – 산포도란 무엇일까?	다음 시간에 학습할 내용을 안내한다.	다음 시간에 학습할 내용을 확인한다.	

156쪽



우리 동네의 우유 판매 대리점 A, B는 길을 사이에 두고 마주 보고 있습니다.

두 대리점 사장님들은 어렸을 때부터 사이좋았던 동갑내기 친구로, 신선한 우유를 동네 곳곳의 소매점에 보급했습니다.

그러던 어느 여름날 우유의 유통 기간이 짧아지면서 두 사람은 신경이 예민해져 있었고, 평소 같으면 부담 없이 할 수 있는 이야기가 작은 다툼으로 번졌습니다.

A 대리점의 평 사장은 B 대리점의 중 사장에게 별 뜻 없이 다음과 같이 말했습니다.

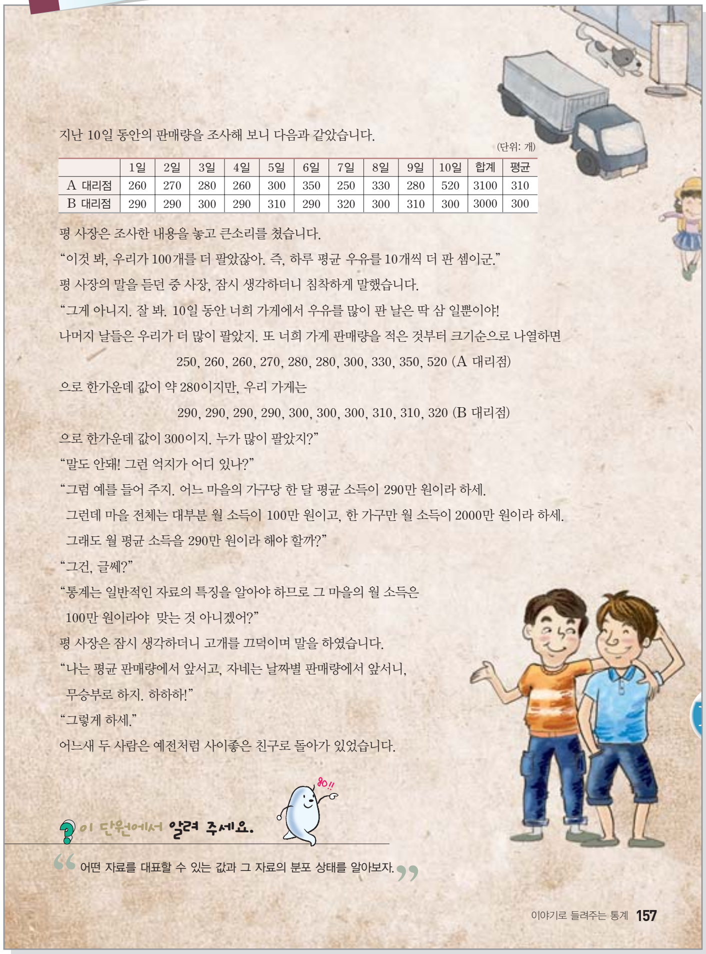
"중 사장, 우리 우유의 인기가 더 좋은 것 같지?"

"뭐라, 우리 제품이 더 잘 나가는데, 무슨 소리야."

순간, 팽팽한 긴장감이 두 사람 사이를 휘어 감았고, 평 사장도 발끈하며 말했습니다.

"그럼 지난 10일 동안의 판매량을 조사해 볼까?"

157쪽



지난 10일 동안의 판매량을 조사해 보니 다음과 같았습니다.

(단위: 개)

	1일	2일	3일	4일	5일	6일	7일	8일	9일	10일	합계	평균
A 대리점	260	270	280	260	300	350	250	330	280	520	3100	310
B 대리점	290	290	300	290	310	290	320	300	310	300	3000	300

평 사장은 조사한 내용을 놓고 큰소리를 쳤습니다.

"이것 봐, 우리가 100개를 더 팔았잖아. 즉, 하루 평균 우유를 10개씩 더 판 셈이군."

평 사장의 말을 듣던 중 사장, 잠시 생각하더니 침착하게 말했습니다.

"그게 아니지. 잘 봐. 10일 동안 너희 가게에서 우유를 많이 판 날은 딱 삼 일뿐이야!
나머지 날들은 우리가 더 많이 팔았지. 또 너희 가게 판매량을 적은 것부터 크기순으로 나열하면

250, 260, 260, 270, 280, 280, 300, 330, 350, 520 (A 대리점)

으로 한가운데 값이 약 280이지만, 우리 가게는

290, 290, 290, 290, 300, 300, 300, 310, 310, 320 (B 대리점)

으로 한가운데 값이 300이지. 누가 많이 팔았지?"

"말도 안돼! 그런 억지가 어디 있나?"

"그럼 예를 들어 주지. 어느 마을의 가구당 한 달 평균 소득이 290만 원이라 하세.
그런데 마을 전체는 대부분 월 소득이 100만 원이고, 한 가구만 월 소득이 2000만 원이라 하세.
그래도 월 평균 소득을 290만 원이라 해야 할까?"

"그건, 글쎄?"

"통계는 일반적인 자료의 특징을 알아야 하므로 그 마을의 월 소득은
100만 원이라야 맞는 것 아니겠어?"

평 사장은 잠시 생각하더니 고개를 끄덕이며 말을 하였습니다.

"나는 평균 판매량에서 앞서고, 자네는 날짜별 판매량에서 앞서니,
무승부로 하지. 하하하!"

"그렇게 하세."

어느새 두 사람은 예전처럼 사이좋은 친구로 돌아가 있었습니다.

이 단원에서 알려 주세요.

"어떤 자료를 대표할 수 있는 값과 그 자료의 분포 상태를 알아보자."

이야기로 들려주는 통계

"이야기 배경"

이 이야기는 일상생활에서 통계가 어떻게 활용되고 있으며 해석되고 있는지를 알아보는 예이다.

같은 자료라도 어떻게 정리하고 어떤 의도를 가지고 해석하느냐에 따라 그 결과가 달라진다는 것을 알려 주고 있다.

이야기 속에서 같은 자료를 가지고 이야기하고 있지만 평 사장은 평균을 이용하고, 중사장은 중앙값을 이용하여 자신들의 대리점의 우유 판매량이 높다는 것을 자랑하고 있다.

즉, 이 이야기를 통하여 통계의 자료는 어떤 대푯값을 선택하느냐에 따라 전혀 다른 결과로 설명될 수 있음을 알 수 있다.

아직 학생들이 이 단원에 대하여 배우지 않았지만, 이 내용을 읽고 다양한 해석을 할 수 있다면 충분한 학습 동기가 이루어졌다고 할 수 있다.

이 이야기 속에서 알려준 대푯값으로의 평균과 중앙값 이외에 자료의 분포를 알 수 있는 표준편차를 구해 보면 다음과 같다.

두 대리점 A, B의 편차를 구하면 다음과 같다.

	1일	2일	3일	4일	5일	6일	7일	8일	9일	10일
A	−50	−40	−30	−50	−10	40	−60	20	−30	210
B	−10	−10	0	−10	10	−10	20	0	10	0

A 대리점:

$$(\text{분산})=\frac{(-50)^2+(-40)^2+(-30)^2+(-50)^2+(-10)^2+40^2+(-60)^2+20^2+(-30)^2+210^2}{10}$$

$$=5820$$

$(\text{표준편차})=\sqrt{5820}=76.29\cdots(\text{개})$

B 대리점:

$$(\text{분산})=\frac{(-10)^2+(-10)^2+(-10)^2+10^2+(-10)^2+20^2+10^2}{10}$$

$$=100$$

$(\text{표준편차})=10(\text{개})$

따라서 B 대리점의 변량들이 평균을 중심으로 가깝게 분포되어 있음을 알 수 있다.

1 대푯값과 산포도

중단원 지도 목표

1. 중앙값, 최빈값, 평균의 의미를 이해하고 이를 구할 수 있게 한다.
2. 분산과 표준편차의 의미를 이해하고 이를 구할 수 있게 한다.

중단원의 구성

소단원 명	지도 내용
1. 대푯값	• 중앙값 • 최빈값
2. 산포도	• 산포도 • 편차, 분산, 표준편차
중단원 마무리하기	• 스스로 정리하기 • 기초 다지기, 기본 익히기, 실력 기르기
창의 · 인성 키우기	• 개념 바루기 • 문제 만들기 • 생각 키우기
컴퓨터로 하는 수학	• 스프레드시트 프로그램을 이용한 통계

158쪽

1 대푯값과 산포도

162 cm

52 kg, 60 kg

1. 대푯값
2. 산포도

준비하기

▶주어진 자료의 평균을 구할 수 있는가?

1. 오른쪽 표는 효정이의 5회에 걸친 영어 점수이다. 영어 점수의 평균을 구하여라.

회	1	2	3	4	5
점수(점)	75	72	76	83	94

▶줄기와 잎 그림을 이해하고 있는가?

2. 오른쪽은 10일 동안 정서의 인터넷 학습 시간을 조사하여 줄기와 잎 그림으로 나타낸 것이다. 다음 물음에 답하여라.
(1) 평균을 구하여라.
(2) 인터넷 학습 시간이 두 번째로 큰 값을 구하여라.

인터넷 학습 시간 (3 | 5는 35분)

줄기	잎
3	5 8 9
4	3 5 7 7
5	0 2 4

▶주어진 도수분포표에서 평균을 구할 수 있는가?

3. 오른쪽은 정규네 반 학생 20명의 수면 시간을 조사하여 나타낸 도수분포표이다. 수면 시간의 평균을 구하여라.

수면(시간)	학생 수(명)
4 이상 ~ 5 미만	4
5 ~ 6	6
6 ~ 7	7
7 ~ 8	2
8 ~ 9	1
합계	20

158 Ⅳ. 통계

준비하기

▶주어진 자료의 평균을 구할 수 있는가?

1. 이 단원에서는 대푯값으로서의 평균을 알아야 하므로 간단한 자료의 평균을 구할 수 있어야 한다.

풀이 영어 점수의 평균은 다음과 같다.

$$(\text{평균})=\frac{75+72+76+83+94}{5}=\frac{400}{5}=80(\text{점})$$

답 80점

▶줄기와 잎 그림을 이해하고 있는가?

2. 이 단원에서는 줄기와 잎 그림을 이용하므로 줄기와 잎 그림을 이해하고 있어야 한다.

풀이 (1)

$$(\text{평균})=\frac{35+38+39+43+45+47+47+50+52+54}{10}$$

$$=45(\text{분})$$

(2) 크기순으로 나열했을 때 두 번째로 큰 값은 52분이다.

답 (1) 45분 (2) 52분

▶주어진 도수분포표에서 평균을 구할 수 있는가?

3. 이 단원에서는 도수분포표로 주어진 자료의 분산, 표준편차를 구하게 된다. 따라서 도수분포표에서 평균을 구할 수 있어야 한다.

풀이 계급값이 4.5, 5.5, 6.5, 7.5, 8.5이므로 구하는 평균은 다음과 같다.

$$(\text{평균})=\frac{4.5\times4+5.5\times6+6.5\times7+7.5\times2+8.5\times1}{20}$$

$$=6(\text{시간})$$

답 6시간

보충 문제

다음 자료의 평균을 구하여라.

(1) 4, 6, 8, 2, 5, 11
(2) 10, 20, 15, 25, 35, 45
(3) 8, 12, 16, 15, 17, 19, 13, 84

답 (1) 6 (2) 25 (3) 23

Ⅳ

01 대푯값

지도 목표

1. 대푯값의 뜻을 이해하게 한다.
2. 중앙값, 최빈값의 의미를 이해하고 이를 구할 수 있게 한다.

지도상의 유의점

1. 대푯값은 자료의 특징을 하나의 수로 나타낸 값으로 이미 배운 평균 이외에도 중앙값과 최빈값이 있음을 알게 한다.
2. 중앙값과 최빈값의 의미를 이해하고, 상황에 따라 중앙값과 최빈값이 필요함을 알게 한다.
3. 주어진 자료에서 적절한 대푯값이 무엇인지 판단할 수 있게 한다.
4. 도수분포표에서 평균을 구할 때에는 계급값을 먼저 구하도록 지도한다.

1/2차시 차시별 학습 지도 방법 [교과 교실]

생각 열기	주어진 자료의 평균을 구하여 보고, 그 평균이 자료의 중심 경향이 잘 나타낸다고 볼 수 있는지 생각해 보게 한다.
본문, 함께 풀기 1	생각 열기를 이용하여 그 자료에 필요한 대푯값이 있음을 인식하게 하고, 중앙값에 대하여 설명한다.
문제 1, 3	스스로 문제를 풀어 보고, 친구들과 결과를 비교해 볼 수 있도록 한다.
문제 2	수학적인 방법을 활용하여 자신의 생각을 논리적으로 정확하게 표현하고, 다른 사람을 이해시킬 수 있도록 한다.

중앙값이란 무엇일까?

생각 열기 극단적인 값이 있을 때 평균은 대푯값으로 적절하지 않음을 알게 하는 생각 열기이다.

(1) $(평균)=\dfrac{8+9+6+10+7+8+71}{5}=17(시간)$

(2) 자료를 작은 값부터 크기순으로 나열하면

6, 7, 8, 8, 9, 10, 71

이므로 중앙에 있는 값은 8시간이다.

159쪽

01 대푯값

학습 목표 중앙값, 최빈값, 평균의 의미를 이해하고 이를 구할 수 있다.
배울 용어 대푯값, 중앙값, 최빈값

오른쪽 보도 자료의 연평균 미세 먼지 농도는 매일 관측한 수치를 평균값으로 발표한 것이다. 이 보도 자료에서 황사 관측일을 제외한 것은 황사 관측일이 기간은 짧지만 미세 먼지 농도가 매우 높으므로 전체 미세 먼지 농도의 평균값에 영향을 주기 때문이다.

보도 자료에서 황사 관측일을 제외하지 않았다면 평균은 의미가 있을까?

서울시 대기질 목표 수준인 '제주도처럼 맑은 날' ($45\mu g/m^3$)도 지난해보다 9일 늘어난 202일을 기록했다. 황사 관측일을 제외하면 연평균 미세 먼지 농도는 $43\mu g/m^3$로 황사 관측일을 제외한 제주도의 2006년부터 2008년간의 평균값과 같았다.

서울시 기후 대기 환경 정보
(http://cleanair.seoul.go.kr)

1/2차시 중앙값이란 무엇일까?

생각 열기 다음은 연서네 반 학생 7명에 대한 여름방학 동안의 봉사 활동 시간을 조사한 자료이다.

(단위: 시간)

8	9	6	10	7	8	71

(1) 봉사 활동 시간의 평균을 구하여 보자.
(2) 위의 자료를 작은 값부터 크기순으로 나열하였을 때, 중앙에 있는 값을 구하여 보자.

우리 반 학생들의 평균 키는 162cm, 평균 몸무게는 52kg과 같이 자료의 특징을 하나의 수로 나타내면 자료의 중심 경향을 쉽게 알아볼 수 있다.

대푯값에는 평균, 중앙값, 최빈값 등이 있다.

이와 같이 자료 전체의 중심 경향이나 특징을 하나의 수로 나타내어 자료 전체를 대표하는 값을 **대푯값**이라고 한다.

1 대푯값에는 여러 가지가 있으나 가장 많이 사용하는 것은 평균이다. 평균은 변량 전체의 합을 총 도수로 나눈 것이다.

1. 대푯값과 산포도 159

1 대푯값은 자료의 중심 경향이나 특징을 하나의 수로 나타내어 자료 전체를 대표하는 값이다. 대푯값은 평균, 중앙값, 최빈값 등이 있는데, 평균은 가장 많이 쓰는 대푯값이다.

2 자료 중에서 매우 크거나 매우 작은 극단적인 값이 있는 경우 평균에 영향을 주어, 평균이 한 쪽으로 치우치는 경향이 있다. 이러한 경우에는 평균이 대푯값으로 적절하지 않다. 따라서 자료 중에 극단적인 값이 있을 경우에는 중앙값이 대푯값으로 적절함을 이해하게 한다.

연구 자료

중앙값의 사용

중앙값은 극단적인 값이 나올 때, 그 값에 영향을 받지 않는다는 장점이 있다. 그러나 중앙값 이외의 변량에 대하여는 아무런 정보가 없다. 따라서 중앙값은 극단적인 값이 통계에 영향을 미쳐 정보가 왜곡될 우려가 있을 때, 주로 사용한다.

160쪽

생각 열기에서 봉사활동 시간의 평균은

$$(\text{평균})=\frac{8+9+6+10+7+8+71}{7}=17(\text{시간})$$

이다.

평균은 대체로 자료의 중심 경향을 잘 나타내어 주지만 그렇지 않은 경우도 있다.

그런데 위의 자료의 대부분은 6에서 10까지로 평균보다 낮고, 71은 평균보다 월등히 높다. 따라서 평균 17은 이 자료의 중심 경향을 잘 나타낸다고 볼 수 없다.

앞의 자료에서는 자료를 작은 값부터 크기순으로 나열하면

6 7 8 8 9 10 71

이고, 가장 중앙에 위치한 값인 8이 대푯값으로 적당하다.

이와 같이 자료를 작은 값부터 크기순으로 나열할 때, 중앙에 위치한 값을 **중앙값**이라고 한다.

자료에 있는 극단적인 값에 영향을 덜 받는 대푯값이 중앙값이다.

2 일반적으로 중앙값은 자료 중에서 매우 크거나 매우 작은 극단적인 값이 있는 경우에 주로 쓰인다.

3 중앙값을 구할 때에는 자료를 작은 값부터 크기순으로 나열하여 자료의 개수가 홀수이면 중앙에 위치한 한 개의 값을 중앙값으로 하고, 자료의 개수가 짝수이면 두 개의 값이 중앙에 있으므로 이 두 수의 값의 평균을 중앙값으로 한다.

함께 풀기 1 다음 자료의 중앙값을 구하여라.

(1) 3, 4, 6, 58, 2, 8, 7

(2) 5, 4, 3, 6, 6, 8, 72, 4, 9, 3

풀이》 자료를 작은 값부터 크기순으로 나열한다.

(1) 자료의 개수가 홀수이면 가장 중앙에 있는 값이 중앙값이다.

2 3 4 6 7 8 58

중앙값

따라서 중앙값은 6이다.

(2) 자료의 개수가 짝수이면 중앙에 있는 두 수의 평균이 중앙값이다.

3 3 4 4 5 6 6 8 9 72

중앙값: $\frac{5+6}{2}=5.5$

따라서 중앙값은 5.5이다.

답》 (1) 6 (2) 5.5

160 Ⅳ. 통계

161쪽

문제 1 다음은 수정이네 모둠 학생 11명의 1분당 윗몸일으키기 횟수를 조사한 자료이다. 이 자료의 중앙값을 구하여라.

(단위: 회)

21	28	26	20	22	25
64	26	33	32	32	

의사소통

문제 2 현수의 삼촌은 행사가 있는 곳에서 주문이 오면 직접 현장으로 출장을 나가 요리를 하는 출장 요리사이다. 다음은 현수의 삼촌이 출장을 나간 행사장에서 예약을 받은 인원 수를 나타낸 것이다. 물음에 답하여라.

(단위: 명)

40	50	30	40	680
50	40	60	60	40

(1) 예약 인원수의 평균과 중앙값을 구하여라.

(2) (1)에서 구한 값 중 어떤 값이 대푯값으로 적절한지 말하고, 그 이유를 설명하여라.

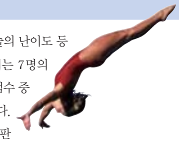
문제 3 오른쪽 그림은 영신이네 반 학생 20명의 과학 점수를 조사하여 나타낸 줄기와 잎 그림이다. 이 자료의 중앙값을 구하여라.

과학 점수 (5 | 2는 52점)

줄기	잎
5	2 4
6	4 6 2
7	8 0 2 6 4 2 8 0 2
8	6 4 4 0
9	0 4

가까운 수학

올림픽 종목과 대푯값



올림픽 종목 중 하나인 다이빙에서는 선수가 물로 뛰어들 때의 동작과 기술의 난이도 등을 심사하여 점수를 매김으로써 등수가 정해진다. 다이빙의 개인 경기에서는 7명의 심판이 1회당 10점 만점으로 채점을 하는데, 점수는 7명의 심판이 채점한 점수 중 가장 높은 점수와 가장 낮은 점수를 제외한 나머지 점수의 평균 값으로 정한다. 이와 같이 가장 높은 점수와 가장 낮은 점수를 제외하는 것은 어떤 특정한 심판이 극단적인 점수를 주어 평균 점수를 왜곡하는 것을 방지하기 위한 것이다. 이러한 점수의 채점 방식은 다이빙 경기뿐만 아니라 체조, 피겨스케이팅 등 다른 종목에서도 자주 활용된다.

1. 대푯값과 산포도 161

3 중앙값은 자료를 크기순으로 나열했을 때 자료의 개수가 홀수인 경우와 짝수인 경우로 나누어 설명한다.

(i) 자료의 개수 N이 홀수이면

$\frac{N+1}{2}$번째 값이 중앙값이다.

(ii) 자료의 개수 N이 짝수이면

$\frac{N}{2}$번째와 $\left(\frac{N}{2}+1\right)$번째의 값의 평균이 중앙값이다.

문제 1 중앙값 구하기

풀이 자료를 작은 값부터 크기순으로 나열하면

20, 21, 22, 25, 26, 26, 28, 32, 32, 33, 64

이므로 구하는 중앙값은 26회이다.

문제 2 적절한 대푯값 찾기

풀이 (1) 예약 인원수의 평균은

$$(\text{평균})=\frac{40+50+30+40+680+50+40+60+60+40}{10}$$

$$=\frac{1090}{10}=109(\text{명})$$

이다.

예약 인원수의 중앙값을 구하기 위하여 자료를 작은 값부터 크기순으로 나열하면

30, 40, 40, 40, 40, 50, 50, 60, 60, 680

이다. 이때 자료의 수가 짝수이므로 구하는 중앙값은

$\frac{40+50}{2}=45(\text{명})$이다.

(2) 자료 중에는 680명과 같은 극단적인 값이 있으므로 대푯값으로는 중앙값이 적절하다.

문제 3 줄기와 잎으로 나타낸 자료의 중앙값 구하기

풀이 20명의 과학 점수를 정리하면 다음과 같다.

과학 점수 (5 | 2는 52점)

줄기	잎
5	2 4
6	2 4 6
7	0 0 2 2 2 4 6 8 8
8	0 4 4 6
9	0 4

자료의 개수는 짝수이고, 중앙에 있는 두 값은 72점, 74점이므로 중앙값은 $\frac{72+74}{2}=73(\text{점})$이다.

수준별 교수·학습 방법

생활 속의 소재로 중앙값을 이해하고 이를 구할 수 있다.

하 대푯값에는 평균과 더불어 중앙값이 있음을 상세히 설명하고, 중앙값을 구할 수 있도록 충분히 지도한다. 이때 간단한 자료를 이용한다.

[문제] 다음 자료의 중앙값을 구하여라.

(1) 2, 3, 4, 5, 6　　　　(2) 5, 6, 7, 8, 9, 10

답 (1) 4　(2) 7.5

상 복잡한 자료를 제시하여 평균과 중앙값을 구하여 보게 하고 평균과 중앙값 중 어떤 것이 대푯값으로 적절한지 토론하게 한다. 이때 교과서 161쪽의 '가까운 수학'을 참조할 수 있다.

2/2차시 차시별 학습 지도 방법 [교과 교실]

생각 열기	지난 시간에 배운 대푯값을 상기시키고, 친숙한 소재를 통해 다른 대푯값이 있음을 소개한다.
본문	생활 속에서의 소재를 통하여 최빈값을 이해하도록 설명한다.
문제 4, 5	스스로 문제를 풀어 보게 하면서 최빈값의 특징을 알아보게 한다.
함께 풀기 2	도수분포표에서의 중앙값과 최빈값을 구하는 방법을 설명한다.
문제 6, 7	발표와 토론 등을 통하여 학생들이 스스로 문제 해결 방법을 발견하도록 유도한다.

최빈값이란 무엇일까?

생각 열기 바지의 치수를 조사한 자료에서 가장 많이 나오는 바지의 치수가 대푯값으로 쓰일 수 있음을 알게 하는 생각 열기이다.

(1) 주어진 자료를 보고 빈칸을 채우면 다음과 같다.

치수(cm)	72	74	76	78	80	82	84	86	88	합계
학생 수(명)	1	3	4	7	10	2	1	1	1	30

(2) 가장 많이 나타나는 바지의 치수는 80cm이다.

4 최빈값은 옷의 치수, 신발의 크기 등과 같은 자료에서 가장 많이 나오는 값을 대푯값으로 정할 때 주로 쓰인다. 예를 들면 운동화 제조 회사에서 재고를 줄이기 위해서는 학생들의 발 치수 중 가장 많이 나오는 치수를 알아야 한다.

162쪽

2/2차시 최빈값이란 무엇일까?

생각 열기 다음은 현수네 반 학생 30명의 바지 치수를 조사한 자료이다.

(단위: cm)

72	78	80	76	74	78	80	78	80	76
80	86	80	76	80	84	82	80	76	78
82	80	78	80	74	80	78	88	78	74

(1) 위의 치수를 보고 빈칸을 채워 보자.

치수(cm)	72	74	76	78	80	82	84	86	88	합계
학생 수(명)										30

(2) 가장 많이 나타나는 바지 치수를 말하여 보자.

소비자의 요구, 선호도 등의 자료에서 대푯값을 구할 때, 자료의 값 중에서 가장 많이 나타나는 값이 의미가 있는 경우가 있다. 예를 들어 생각 열기에서 가장 많이 나타나는 바지 치수는 80cm이다.

4 이와 같이 자료의 값 중에서 가장 많이 나타나는 값을 **최빈값**이라고 한다.

일반적으로 최빈값은 가장 많이 팔리는 음료수, 옷의 치수, 신발의 크기 등과 같은 자료의 대푯값을 구할 때 주로 쓰인다.

자료가 모두 같거나 다른 경우의 최빈값을 찾는 것은 의미가 없다.

5 보기 (1) 자료가 2, 2, 3, 5, 7, 9, 9, 9, 10일 때, 최빈값은 9이다.
(2) 자료가 20, 15, 5, 10, 15, 5, 5, 10, 10일 때, 최빈값은 5, 10이다.
(3) 자료가 호두, 땅콩, 밤, 호두, 대추, 잣, 호두일 때, 최빈값은 호두이다.

최빈값은 자료에 따라 두 개 이상 나올 수도 있다.

문제 4 다음 자료의 최빈값을 구하여라.
(1) 5, 8, 6, 12, 8, 8, 6, 7, 10, 10, 8
(2) 80, 95, 90, 95, 100, 95, 85, 95, 105, 95, 85

문제 5 다음은 영애가 하루의 기온의 변화를 알아보기 위하여 학교에 설치된 백엽상의 온도계의 온도를 오전 6시부터 오후 6시까지 매시간 측정하여 시간 순으로 기록한 것이다. 이 자료의 중앙값과 최빈값을 구하여라.

백엽상은 기온을 재기 위하여 통풍이 잘되는 평탄한 곳에 나무로 설치한 상자이다.

(단위: ℃)

3	4	5	6	6	7	8
8	12	16	11	9	6	

162 Ⅳ. 통계

5 최빈값은 보기 (2)와 같이 두 개 이상 나올 수도 있음을 알게 한다.

문제 4 최빈값 구하기

풀이 (1) 주어진 자료에서 가장 많이 나타나는 수는 8로 모두 네 번 나타난다. 따라서 최빈값은 8이다.
(2) 주어진 자료에서 가장 많이 나타나는 수는 95로 모두 다섯 번 나타난다. 따라서 최빈값은 95이다.

문제 5 중앙값과 최빈값 구하기

풀이 이 자료를 작은 값부터 크기순으로 나타내면
3, 4, 5, 6, 6, 6, 7, 8, 8, 9, 11, 12, 16
이므로 중앙값은 7℃이다.
최빈값은 가장 많이 나타난 6℃이다.

163쪽

이제 도수분포표에서 중앙값과 최빈값을 구하여 보자.

6 도수분포표에서는 각 계급에 속하는 변량들의 정확한 값을 알 수 없으므로 도수분포표에서 중앙값과 최빈값을 구할 때에는 변량이 속하는 계급의 계급값을 이용한다.

함께 풀기 2 오른쪽 표는 수정이네 반 학생 40명의 한 달간 휴대 전화 문자 사용 횟수를 조사하여 나타낸 도수분포표이다. 이 자료의 중앙값과 최빈값을 구하여라.

문자 사용 횟수(회)	학생 수(명)
0 이상 ~ 50 미만	6
50 ~ 100	13
100 ~ 150	9
150 ~ 200	6
200 ~ 250	4
250 ~ 300	2
합계	40

풀이》 중앙값은 40개의 변량을 문자 사용 횟수가 적은 순서로 나열하였을 때, 20번째 학생과 21번째 학생이 사용한 문자 사용 횟수의 평균이다.
따라서 도수분포표에서 이 학생들이 속한 계급은 100회 이상 150회 미만이므로 중앙값은 이 계급의 계급값인 125회이다.
최빈값은 도수가 가장 큰 계급인 50회 이상 100회 미만의 계급값이므로 75회이다.

답》 중앙값: 125회, 최빈값: 75회

문제 6 오른쪽 표는 컴퓨터 동아리 학생 40명의 분당 한글 타수를 조사하여 나타낸 도수분포표이다. 이 자료의 중앙값과 최빈값을 구하여라.

타수(타)	학생 수(명)
100 이상 ~ 150 미만	6
150 ~ 200	7
200 ~ 250	9
250 ~ 300	12
300 ~ 350	4
350 ~ 400	2
합계	40

문제 해결

문제 7 다음 표는 연수네 집의 최근 2년간 매월 전기 사용량을 조사하여 나타낸 도수분포표인데, 일부가 얼룩이 져서 보이지 않는다. 이 자료의 중앙값과 최빈값을 구하여라.

전기 사용량(kWh)	도수(개월)
100 이상 ~ 200 미만	9
200 ~ 300	6
300 ~ 400	
400 ~ 500	2
500 ~ 600	1
합계	24

1. 대푯값과 산포도 163

수준별 교수·학습 방법

생활 속의 소재로 최빈값을 이해하고, 이를 구할 수 있다. 또 도수분포표로 나타낸 자료의 중앙값과 최빈값을 이해하고 이를 구할 수 있다.

하

- 대푯값에는 평균, 중앙값과 더불어 최빈값이 있음을 상세히 설명하고, 최빈값을 구할 수 있도록 충분히 지도한다.

  [문제 1] 다음 자료의 최빈값을 구하여라.
  (1) 1, 2, 3, 3, 4, 5
  (2) 2, 3, 2, 3, 4, 5, 4, 4
  (3) 6, 7, 8, 6, 7, 9, 10, 11

  답 (1) 3 (2) 4 (3) 6, 7

- 도수분포표를 자세히 설명하고, 이러한 자료의 중앙값과 최빈값을 구할 수 있도록 상세히 지도한다.

- 간단한 자료에서 적당한 대푯값을 생각해 보게 한다.

  [문제 2] 다음 자료의 대푯값으로 적당한 것을 말하여라.
  (1) 65, 67, 72, 73, 75
  (2) 36, 38, 42, 38, 50, 60, 600, 24
  (3) 30, 35, 40, 35, 50, 35, 40, 35, 35, 65, 35, 25

  답 (1) 평균 (2) 중앙값 (3) 최빈값

상 생활 속에서 사례를 말하고, 그것이 평균, 중앙값, 최빈값 중 어떤 것이 대푯값으로 적절한지 토론하게 한다.
교과서 167쪽 확인하기 4번을 참조할 수도 있다.

6 도수분포표에서의 중앙값과 최빈값은 중앙값과 최빈값이 속하여 있는 계급의 계급값으로 정함을 지도한다.

문제 6 도수분포표에서 중앙값과 최빈값 구하기

풀이 중앙값은 20번째 학생과 21번 째 학생의 분당 한글 타수의 평균이다. 따라서 이 학생들이 속한 계급은 200타 이상 250타 미만이므로 중앙값은 이 계급의 계급값인 225타이다.
최빈값은 도수가 가장 큰 계급인 250타 이상 300타 미만의 계급값이므로 275타이다.

문제 7 도수분포표에서 중앙값과 최빈값 구하기

풀이 총 도수가 24개월이므로 300kWh 이상 400kWh 미만인 계급의 도수는 6이다.
중앙값은 12번째와 13번째의 평균이다. 따라서 이것이 속한 계급은 200kWh 이상 300kWh 미만이므로 중앙값은 이 계급의 계급값인 250kWh이다.
최빈값은 도수가 가장 큰 계급인 100kWh 이상 200kWh 미만의 계급값이므로 150kWh이다.

참고 **평균과 최빈값의 실생활의 예**

식품 의약품 안전청은 국민들이 즐겨 먹는 외식 음식의 중량을 다음과 같이 밝혔다.

음식점 조사에서 짜장면 1인분으로 650g을 주는 경우가 가장 많았고(최빈값), 최댓값은 840g, 최솟값은 400g으로 약 2.1배의 차이가 났다. 평균 중량은 607g이었다.

확인하기

1 평가의 주안점 평균, 중앙값, 최빈값을 구할 수 있다.

풀이 (1) 평균을 구하면

$$(\text{평균})=\frac{4+9+10+7+6+5+9+8+9+8}{10}$$

$$=7.5(\text{점})$$

(2) 작은 값부터 크기순으로 나열하면

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

이므로 중앙값은 8점이다.

(3) 가장 많이 나타나는 값은 9이므로 최빈값은 9점이다.

2 평가의 주안점 줄기와 잎 그림을 보고 중앙값과 최빈값을 구할 수 있다.

풀이 홈런을 친 개수의 줄기와 잎이 작은 것부터 순서대로 나열 되어 있으므로 가장 중앙에 위치한 값인 18개가 중앙값이다. 또 가장 많이 나타나는 값은 15개이므로 최빈값은 15개이다.

3 평가의 주안점 도수분포표에서 중앙값과 최빈값을 구할 수 있다.

풀이 키가 작은 학생부터 크기순으로 나열하면 10번째와 11번째 학생의 평균이 중앙값이다.
따라서 키가 10번째와 11번째 학생이 속하여 있는 계급 155cm 이상 160cm 미만의 계급값인 157.5cm가 중앙값이다.
최빈값은 가장 많은 도수가 있는 계급 160cm 이상 165cm 미만의 계급값인 162.5cm이다.

4 평가의 주안점 주어진 자료에서 어떤 것이 대푯값으로 적절한지 구별할 수 있고, 토론할 수 있다.

풀이 학생 10명의 수학 점수, 신발 크기, 이웃 돕기 성금의 적절한 대푯값은 다음과 같다.

- 수학 점수: 이 자료에서는 10명의 점수가 골고루 분포되어 있으므로 중심 경향을 알려면 중앙값보다 평균이 더 적당하다.
- 신발의 크기: 이 자료의 평균은 254.5mm이다. 하지만 신발의 사이즈는 5mm 단위로 크기가 정해져 있으므로 구한 평균 254.5mm 보다는 가장 많이 나온 최빈값인 255mm가 대푯값으로 적절하다. 이와 같은 자료는 평균이나 중앙값보다 많이 나타나는 변량이 더 의미가 있으므로 최빈값이 적당하다.

164쪽

확인하기

1 아래는 정아가 다트 던지기를 10회하여 얻은 점수를 조사한 자료이다. 다음을 구하여라.

(단위: 점)

4	9	10	7	6
5	9	8	9	8

(1) 평균 (2) 중앙값 (3) 최빈값

2 오른쪽 그림은 어느 해의 프로 야구 시즌에 홈런을 많이 친 선수 15명의 홈런 개수를 조사하여 나타낸 줄기와 잎 그림이다. 이 자료의 중앙값과 최빈값을 구하여라.

홈런 수 (1 | 5는 15개)

줄기	잎
1	5 5 5 5 6 7 7 8 8 9 9
2	0 0 7
3	0

한국 야구위원회(http://www.koreabaseball.com)

3 오른쪽 표는 민수네 반 학생 20명의 키를 조사하여 나타낸 도수분포표이다. 이 자료의 중앙값과 최빈값을 구하여라.

키(cm)	학생 수(명)
145 이상 ~ 150 미만	1
150 ~ 155	4
155 ~ 160	6
160 ~ 165	7
165 ~ 170	1
170 ~ 175	1
합계	20

수학적 과정 의사소통 추론 문제 해결

4 다음은 학생 10명의 수학 점수, 신발의 크기, 이웃 돕기 성금을 조사한 자료이다.

학생	혜원	승기	호동	민서	진수	수연	동민	찬수	은송	형민
수학 점수(점)	90	80	75	65	70	85	55	60	90	70
신발의 크기(mm)	250	255	255	240	255	245	265	255	250	275
이웃 돕기 성금(원)	3000	2000	3000	50000	1000	4500	2500	3500	2500	1500

위의 세 가지 자료 각각에 대하여 어떤 대푯값이 적절한지 추측하여 토론하여라.

164 Ⅳ. 통계

- 이웃 돕기 성금: 이 자료는 대부분의 값이 1000원과 4000원 사이의 값이지만 평균을 구하면 7350원이다. 왜냐하면 다른 값보다 훨씬 큰 50000원이 있기 때문이다.
따라서 이 자료는 중앙값이 대푯값으로 적당하다.

창의 · 인성 모둠별로 토론하게 하고, 원하는 답이 아니더라도 자기 주장의 타당성 있는 논리를 전개하면 인정하여 다양한 사고를 할 수 있게 한다.

연구 자료

최빈값의 사용

평균과 중앙값은 키, 몸무게, 사람의 수명 등과 같이 양적 자료의 대푯값으로 사용하지만 최빈값은 양적 자료보다는 좋아하는 음악, 좋아하는 스포츠 등과 같은 질적 자료에서 더욱 효과적으로 사용할 수 있다. 그러나 자료의 수가 너무 적거나 최빈값이 너무 많이 나올 때는 자료의 중심 경향을 잘 반영하지 못할 때도 있다.

02 산포도

지도 목표

1. 산포도의 뜻을 이해할 수 있게 한다.
2. 편차, 분산, 표준편차의 의미를 이해하고 이를 구할 수 있게 한다.

지도상의 유의점

1. 일상생활의 다양한 상황을 소재로 이용하여 산포도의 의미를 이해할 수 있도록 지도한다.
2. 필요한 경우 자료의 분산, 표준편차를 구할 때 계산기 또는 컴퓨터를 이용하도록 지도한다.

1/3차시 차시별 학습 지도 방법

생각 열기	주어진 자료의 평균을 각각 구하여 보고, 누구의 점수가 평균과 가깝게 분포되어 있는지 생각해 보게 한다.
본문	생각 열기의 두 자료의 분포를 나타낸 그래프를 보면서 자료의 분포 상태를 비교하여 설명하고, 산포도의 의미를 알게 한다.
문제 1	수학적 의사소통을 활성화시키는 과제이므로 자신의 풀이를 설명할 수 있도록 지도한다.

산포도란 무엇일까?

생각 열기 두 학생의 사격 점수의 평균을 구하고, 두 학생의 점수 분포를 비교하여 설명하게 하는 생각 열기이다.

(1) $(\text{연서의 평균})=\dfrac{7+7+8+9+9+8+9+8+7+8}{10}$

$=8(\text{점})$

$(\text{진호의 평균})=\dfrac{10+9+9+5+10+6+8+5+10+8}{10}$

$=8(\text{점})$

(2) 두 학생의 평균은 8점으로 같고, 연서의 점수가 진호의 점수보다 평균에 더 가까이 분포되어 있다.

165쪽

02 산포도

○ 학습 목표 분산과 표준편차의 의미를 이해하고 이를 구할 수 있다.
○ 배울 용어 산포도, 편차, 분산, 표준편차

사격은 표적을 총으로 맞혀 득점이 높은 사람이 이기는 경기이다. 사격은 정적인 운동이지만 선수가 되기 위해서는 근력 강화 운동, 심폐 능력 향상 운동 등의 기초 체력을 키우기 위한 꾸준한 반복 훈련이 필요하다. 사격 연습을 한 두 선수가 있다고 할 때, 기록된 평균 점수가 같으면 누구를 대표 선수로 뽑아야 할까?

1/3차시 **산포도란 무엇일까?**

생각 열기 연서와 진호는 동아리 활동 시간에 사격을 배웠다. 다음 표는 두 학생의 사격 기록을 조사하여 나타낸 것이다.

[표 1] (단위: 점)

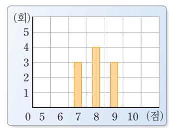
회	1	2	3	4	5	6	7	8	9	10
연서	7	7	8	9	9	8	9	8	7	8
진호	10	9	9	5	10	6	8	5	10	8

(1) 연서와 진호의 점수의 평균을 각각 구하여 보자.
(2) 두 학생의 점수 분포를 비교해 보고, 누구의 점수가 평균에 가까이 분포되어 있는지 말하여 보자.

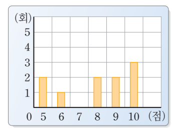
생각 열기에서 연서와 진호의 평균 점수는 8점으로 서로 같다. 이때 두 자료의 분포 상태는 어떤지 살펴보자.

[표 1]에서 사격 기록의 분포 상태를 막대그래프로 나타내면 다음과 같다.

❶



[연서]



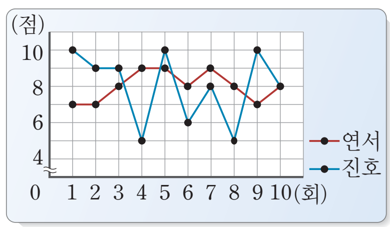
[진호]

위의 그래프에서 연서의 점수는 7점, 8점, 9점으로 평균에 가까이 모여 있고, 진호의 점수는 5점에서 10점까지 평균의 좌우로 넓게 흩어져 있다.

1. 대푯값과 산포도 165

❶ 생각 열기 자료의 그래프를 보여 주고 평균과 각 자료의 분포 상태를 파악하게 한다. 즉, 두 학생의 사격 기록의 평균은 같지만 분포 상태는 다르다. 따라서 자료의 특징을 알기 위해서는 평균뿐만 아니라 자료의 분포 상태도 필요함을 알 수 있게 한다.

생각 열기의 연서와 진호의 사격 기록 점수를 꺾은선 그래프로 나타내면 다음과 같다.



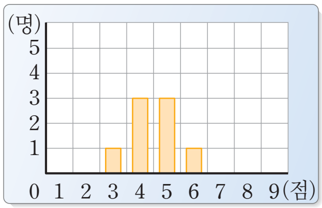
위의 그래프에서 연서의 점수가 진호의 점수보다 평균에 가깝게 모여 있고, 고른 분포를 나타낸다.
따라서 연서가 더 안정적인 경기를 한다고 할 수 있다.

IV

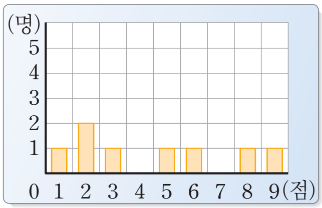
문제 1 산포도의 뜻 알기

풀이 A 팀과 B 팀의 평균은 4.5점으로 같다.
양팀의 득점을 막대그래프로 나타내어 보면 다음과 같다.

[A 팀]



[B 팀]



위의 막대그래프에서 A 팀의 개별 득점이 B 팀의 개별 득점보다 평균에 가까이 몰려 있다.

166쪽

평균이 대푯값으로 쓰이기는 하지만 자료의 분포 상태를 살펴보는 것으로는 충분하지 못함을 알 수 있다.

이와 같이 평균은 같아도 변량이 흩어져 있는 정도는 다를 수 있다.

따라서 자료의 분포 상태를 알기 위해서는 변량들이 평균 주위에 어떻게 흩어져 있는지를 알아야 한다.

이때 변량이 흩어진 정도를 하나의 수로 나타낸 값을 **산포도**라고 한다

의사소통

문제 1 다음은 핸드볼 경기에서 A팀의 선수 8명과 B팀의 선수 8명의 득점을 조사한 자료이다. 아래의 표를 막대그래프로 나타내고, 어떤 팀의 각 개인별 득점이 평균에 더 가까운지 말하여라.

(단위: 점)

4	5	4	6
4	5	5	3

[A팀]

(단위: 점)

2	3	9	5
1	6	2	8

[B팀]

(명) 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9(점)

(명) 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9(점)

편차, 분산, 표준편차란 무엇일까?

2/3차시 생각 열기

다음 표는 두 병원 A, B에서 예약 환자 10명의 진료 대기 시간을 조사하여 나타낸 것이다.

(단위: 분)

환자 순서(번)	1	2	3	4	5	6	7	8	9	10	평균
A 병원	15	18	15	15	13	14	15	16	17	12	15
B 병원	24	6	8	13	16	19	8	11	14	31	15

(1) A 병원의 예약환자 10명의 (진료 대기 시간) − (평균 대기 시간)을 각각 구하여 보자.

(2) B 병원의 예약환자 10명의 (진료 대기 시간) − (평균 대기 시간)을 각각 구하여 보자.

(3) (1), (2)의 결과 진료 대기 시간이 더 고른 병원이 어디인지 생각하여 보자.

166 Ⅳ. 통계

수준별 교수·학습 방법

산포도의 의미를 알 수 있다.

하 평균을 구하고, 자료의 분포를 그래프로 이해할 수 있게 한다.
[문제] 다음은 리듬체조 A, B 선수의 점수이다. 어떤 선수의 점수가 평균에 더 몰려있는지 구하여라.

A	6	7	9	8	5	6	7	8
B	7	7	8	7	6	6	7	8

답 B 선수

상 생활 속에서 산포도가 뚜렷이 나타나는 예를 찾아보게 한다.

2/3차시 차시별 학습 지도 방법 [교과 교실]

생각 열기	주어진 자료의 편차를 구하여 보게 하고, 두 자료를 비교하여 어떤 점이 다른지 발표하게 한다.
본문	생각 열기의 자료를 이용하여 분산과 표준편차를 설명한다.
함께 풀기 1, 문제 2	하 수준의 학생을 고려하여 표에 대한 자세한 설명을 하고, 분산과 표준편차를 구할 수 있도록 지도한다.
문제 3	두 자료의 분산과 표준편차를 구하고, 모둠별로 의견을 나누어 봄으로써 통계 자료의 분석에 대한 능력을 높일 수 있도록 지도한다.
문제 4	줄기와 잎 그림으로 나타낸 두 모둠의 역사 수행평가 점수의 평균과 표준편차를 구하고, 어떤 모둠의 성적이 더 고른지 이해할 수 있도록 지도한다.

연구 자료

산포도의 종류

(1) 범위(Ranger): 자료의 최댓값과 최솟값의 차
(2) 평균편차: 편차의 절댓값의 평균
(3) 사분편차: 도수분포 곡선에서 전체 넓이를 사등분하는 분기점을 Q_1, Q_2, Q_3이라 하면 사분편차 Q는

$$Q=\frac{Q_3-Q_1}{2}$$

(4) 표준편차: 분산의 음이 아닌 제곱근

167쪽

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

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

(편차)=(변량)−(평균)

생각 열기에서 두 병원의 환자 대기 시간에 대한 편차와 편차의 합을 구하면 다음과 같다.

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

[표 1] (단위: 분)

환자 순서(번)	1	2	3	4	5	6	7	8	9	10	합계
A 병원	0	3	0	0	−2	−1	0	1	2	−3	0
B 병원	9	−9	−7	−2	1	4	−7	−4	−1	16	0

위의 표에서 편차의 절댓값이 클수록 그 변량은 평균에서 멀리 떨어져 있고, 편차의 절댓값이 작을수록 평균에 가까이 있음을 알 수 있다.

2 [표 1]에서 편차의 합은 0이므로 편차의 합으로는 변량들이 흩어져 있는 정도를 나타낼 수 없다. 따라서 편차의 합이 0이 되지 않도록 편차의 제곱의 평균을 이용하여 산포도를 구한다.

[표 1]에서 편차의 제곱의 합을 구하면 다음과 같다.

A 병원 $0^2+3^2+0^2+0^2+(-2)^2+(-1)^2+0^2+1^2+2^2+(-3)^2=28$

B 병원 $9^2+(-9)^2+(-7)^2+(-2)^2+1^2+4^2+(-7)^2$
$+(-4)^2+(-1)^2+(16)^2=554$

이때, 두 자료에서 편차의 제곱의 평균을 각각 구하면

A 병원 편차의 제곱의 평균: $\frac{28}{10}=2.8$

B 병원 편차의 제곱의 평균: $\frac{554}{10}=55.4$

이고, 이 값은 평균을 중심으로 각 변량이 흩어져 있는 정도를 나타낸다.

▸ 표준편차의 단위는 변량의 단위와 같다.

이와 같이 어떤 자료의 편차의 제곱의 평균을 **분산**이라 하고, 분산의 음이 아닌 제곱근을 **표준편차**라고 한다. 3

A 병원과 B 병원의 분산과 표준편차를 구하면

A 병원 (분산)=2.8, (표준편차)=$\sqrt{2.8}=1.673\cdots$(분)

B 병원 (분산)=55.4, (표준편차)=$\sqrt{55.4}=7.443\cdots$(분)

4 이므로 A 병원의 표준편차는 약 1.67분이고, B 병원의 편차는 약 7.44분이다.

1. 대푯값과 산포도 167

편차, 분산, 표준편차란 무엇일까?

생각 열기 두 병원의 10명의 진료 대기 시간에서

(진료 대기 시간)−(평균 대기 시간)

을 구하여 편차의 의미를 알게 하려는 생각 열기이다.

(1), (2) A, B 두 병원의 진료 대기 시간의 편차, 즉

(진료 대기 시간)−(평균 대기 시간)

을 구하면 다음과 같다.

(단위: 분)

환자 순서(번)	1	2	3	4	5	6	7	8	9	10
A 병원	0	3	0	0	−2	−1	0	1	2	−3
B 병원	9	−9	−7	−2	1	4	−7	−4	−1	16

(3) A 병원의 진료 대기 시간이 B 병원의 진료 대기 시간보다 더 고르다.

2 생각 열기에서 환자의 진료 대기 시간의 편차의 합을 구하여 보면 A, B 병원 모두 0이 됨을 알 수 있다.
일반적으로 편차의 합은 항상 0이 됨을 알 수 있다.

따라서 편차의 평균은 의미가 없음을 알게 하고, 편차의 합이 0이 되지 않도록 편차의 제곱의 평균(분산)을 구함을 이해하게 한다. 또 분산의 음이 아닌 제곱근이 표준편차임을 알게 한다.

연구 자료

편차의 합은 항상 0이다.

n개의 변량 $x_1, x_2, x_3, \cdots\cdots, x_n$의 평균을 $\overline{x}$라 할 때,

(각 변량의 편차)$=x_i-\overline{x}$

(각 변량의 편차의 합)

$$=\frac{(x_1-\overline{x})+(x_2-\overline{x})+\cdots\cdots+(x_n-\overline{x})}{n}$$

$$=\frac{x_1+x_2+\cdots\cdots+x_n-n\overline{x}}{n}=\frac{x_1+x_2+\cdots\cdots+x_n}{n}-\overline{x}$$

$$=\overline{x}-\overline{x}=0$$

3 표준편차를 구하는 순서는 다음과 같다.

① 자료의 평균을 구한다.
② 각 변량의 편차를 구한다.
③ 편차의 제곱의 총합을 구한다.
④ ③에서 구한 총합을 도수의 총합으로 나누어 분산을 구한다.
⑤ 분산의 음이 아닌 제곱근을 구하면 표준편차를 얻는다.

연구 자료

자료의 개수가 n개일 때, 변량이 모두 같으면

$$x_1=x_2=\cdots\cdots=x_n$$

으로 놓을 수 있고, 이 변량의 편차는 모두 0이므로 분산은 0이다. 반대로 분산이 0이면 임의의 $i(1\le i\le n)$에 대하여 $x_i=\overline{x}$이므로

$$x_1=x_2=\cdots\cdots=x_n$$

이다. 이러한 자료의 표준편차는 0이다.

이와 같이 자료가 모두 같을 때는 표준편차는 0이고, 자료가 모두 같지 않을 때는 분산은 양수이다. 따라서 표준편차를 분산의 음이 아닌 제곱근으로 정의하기로 한다.

4 일반적으로 자료의 분산 또는 표준편차가 작을수록 자료가 평균을 중심으로 몰려 있음을 뜻하고, 그것은 자료의 분포가 고르다고 할 수 있음을 이해하게 한다.
또 자료의 분산 또는 표준편차가 클수록 자료가 평균을 중심으로 넓게 흩어져 있음을 뜻하고, 그것은 자료의 분포가 고르지 않고 평균을 경계로 변동이 크다고 할 수 있음을 이해하게 한다.

문제 2 **두 자료의 분산과 표준편차를 구하고 비교하기**

풀이 (1) 경대와 영서의 득점의 평균을 구하면 다음과 같다.

$$(\text{경대의 평균})=\frac{3+6+7+5+9}{10}=\frac{30}{5}=6(\text{점})$$

$$(\text{영서의 평균})=\frac{4+7+3+6+10}{10}=\frac{30}{5}=6(\text{점})$$

(2) 경대의 분산과 표준편차를 구하면 다음과 같다.

$$(\text{분산})=\frac{(3-6)^2+(6-6)^2+(7-6)^2+(5-6)^2+(9-6)^2}{5}$$

$$=\frac{20}{5}=4$$

$$(\text{표준편차})=\sqrt{4}=2(\text{점})$$

영서의 분산과 표준편차를 구하면 다음과 같다.

$$(\text{분산})=\frac{(4-6)^2+(7-6)^2+(3-6)^2+(6-6)^2+(10-6)^2}{5}$$

$$=\frac{30}{5}=6$$

$$(\text{표준편차})=\sqrt{6}=2.449\cdots(\text{점})$$

따라서 경대의 분산은 4, 표준편차는 2점이고, 영서의 분산은 6, 표준편차는 2.45점이다.

(3) 경대의 표준편차는 2점이고, 영서의 표준편차는 2.45점이므로 경대의 점수가 더 고르다.

문제 3 **두 자료의 분산과 표준편차를 구하고 비교하기**

풀이 (1) 자료 A의 분산과 표준편차를 구하면 다음과 같다.

$$(\text{평균})=\frac{2+3+4+5+6}{5}=\frac{20}{5}=4$$

$$(\text{분산})=\frac{(2-4)^2+(3-4)^2+(4-4)^2+(5-4)^2+(6-4)^2}{5}$$

$$=\frac{10}{5}=2$$

$$(\text{표준편차})=\sqrt{2}$$

(2) 자료 B의 분산과 표준편차를 구하면 다음과 같다.

$$(\text{평균})=\frac{112+113+114+115+116}{5}=\frac{570}{5}=114$$

$$(\text{분산})=\frac{(112-114)^2+(113-114)^2+(114-114)^2+(115-114)^2+(116-114)^2}{5}$$

$$=\frac{10}{5}=2$$

$$(\text{표준편차})=\sqrt{2}$$

(3) 두 자료의 평균은 다르지만 분산이 같으므로 표준편차는 같다. 즉, 두 자료는 평균을 중심으로 흩어져 있는 정도가 같다.

5 **문제 3**에서 자료 B는 자료 A의 각 변량에 110을 더한 값이다. 따라서 자료 A의 변량을 $x_i(i=1, 2, 3, 4, 5)$라 하면 자료 B의 변량은 $y_i=x_i+110$으로 나타낼 수 있다.

168쪽

따라서 A 병원의 진료 대기 시간의 표준편차가 B 병원의 진료 대기 시간의 표준편차보다 작으므로 A 병원의 진료 대기 시간이 B 병원의 진료 대기 시간보다 평균을 중심으로 가까이 모여 있음을 알 수 있다.

이상을 정리하면 다음과 같다.

분산과 표준편차

$$(\text{분산})=\frac{(\text{편차})^2\text{의 총합}}{(\text{변량})\text{의 개수}} \qquad (\text{표준편차})=\sqrt{(\text{분산})}$$

표준편차가 작을수록 평균을 중심으로 모여 있으므로 자료의 분포가 더 고르다고 할 수 있다.

일반적으로 분산 또는 표준편차가 작을수록 자료가 평균을 중심으로 모여 있음을 뜻하고, 분산 또는 표준편차가 클수록 자료가 평균으로부터 넓게 흩어져 있음을 뜻한다.

함께 풀기 1 다음은 어느 농장에서 나온 달걀 10개의 무게를 조사한 자료이다. 물음에 답하여라.

(단위: g)

45	48	49	47	44
46	42	45	41	43

(1) 달걀 무게의 평균을 구하여라.

(2) 달걀 무게의 분산과 표준편차를 구하여라.

풀이》 (1) $(\text{평균})=\frac{45+48+49+47+44+46+42+45+41+43}{10}$

$=\frac{450}{10}=45(\text{g})$

(2) 각 변량에 대한 편차와 편차의 제곱을 구하면 다음과 같다.

변량(g)	45	48	49	47	44	46	42	45	41	43	합계
편차	0	3	4	2	-1	1	-3	0	-4	-2	0
$(\text{편차})^2$	0	9	16	4	1	1	9	0	16	4	60

따라서 구하는 분산과 표준편차는

$(\text{분산})=\frac{(\text{편차})^2\text{의 총합}}{(\text{변량})\text{의 개수}}=\frac{60}{10}=6$

$(\text{표준편차})=\sqrt{(\text{분산})}=\sqrt{6}(\text{g})$

답》 (1) 45g (2) 분산: 6, 표준편차: $\sqrt{6}$g

168 Ⅳ. 통계

두 자료의 평균을 $\bar{x}$, $\bar{y}$라 할 때,

$\bar{y}=\bar{x}+110$,

$y_i-\bar{y}=(x_i+110)-(\bar{x}+110)=x_i-\bar{x}$

이므로 두 자료의 편차가 모두 같기 때문에 편차의 제곱의 평균인 분산이 같다.

이와 같이 변량의 크기가 매우 다른 자료라도 각각의 자료에서 변량의 간격이 일정하면, 즉 편차가 일정하면 표준편차는 같다.

문제 4 **두 자료의 표준편차를 구하고 비교하기**

풀이 (1) $(\text{평균})=\frac{8+10+12+16+18+20}{6}=\frac{84}{6}=14(\text{점})$

$$(\text{분산})=\frac{(8-14)^2+(10-14)^2+(12-14)^2+(16-14)^2+(18-14)^2+(20-14)^2}{6}$$

$$=\frac{56}{3}$$

$$(\text{표준편차})=\sqrt{\frac{56}{3}}=4.320\cdots(\text{점})$$

따라서 평균은 14점, 표준편차는 4.32점이다.

(2) $(\text{평균})=\frac{6+8+10+12+14+16+18+20}{8}=\frac{104}{8}=13(\text{점})$

169쪽

문제 2 다음 표는 농구 선수인 경대와 영서가 농구 자유투를 매회 10번씩 5회 던져 얻은 득점을 나타낸 것이다. 물음에 답하여라.

(단위: 점)

회	1	2	3	4	5
경대	3	6	7	5	9
영서	4	7	3	6	10

(1) 경대와 영서의 득점의 평균을 각각 구하여라.
(2) 경대와 영서의 득점의 분산과 표준편차를 각각 구하여라.
(단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)
(3) 누구의 자유투 점수가 더 고른지 말하여라.

의사소통

5 **문제 3** 두 자료 A, B에 대하여 다음 물음에 답하여라.

2 3 4 5 6	112 113 114 115 116
[자료 A]	[자료 B]

(1) 자료 A의 분산과 표준편차를 구하여라.
(2) 자료 B의 분산과 표준편차를 구하여라.
(3) (1), (2)의 결과에서 알 수 있는 것을 토론하여라.

문제 4 다음은 인영이네 모둠 6명과 현지네 모둠 8명의 역사 수행 평가 점수를 조사하여 나타낸 줄기와 잎 그림이다. 만점이 20점일 때, 다음 물음에 답하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)

인영이네 모둠 점수

줄기	잎
0	8
1	0 2 6 8
2	0

현지네 모둠 점수

줄기	잎
0	6 8
1	0 2 4 6 8
2	0

(1) 인영이네 모둠의 평균과 표준편차를 구하여라.
(2) 현지네 모둠의 평균과 표준편차를 구하여라.
(3) 어떤 모둠의 성적이 더 고른가?

1. 대푯값과 산포도 169

$$(\text{분산})=\frac{(6-13)^2+(8-13)^2+(10-13)^2+(12-13)^2+(14-13)^2+(16-13)^2+(18-13)^2+(20-13)^2}{8}$$

$$=\frac{168}{8}=21$$

$(\text{표준편차})=\sqrt{21}=4.582\cdots(\text{점})$

따라서 평균은 13점, 표준편차는 4.58점이다.

(3) 인영이네 모둠이 표준편차가 더 작으므로 역사 수행 평가 점수가 더 고르다.

수준별 교수·학습 방법

분산과 표준편차의 의미를 알고 이를 구할 수 있다.

하 간단한 자료의 분산과 표준편차를 구할 수 있게 한다. 또 두 자료의 편차가 같다면 분산과 표준편차가 같음을 예를 들어 설명한다.

[문제] 다음 두 자료의 표준편차를 비교하여라.

(1) 10, 20, 30, 40, 50　　(2) 110, 120, 130, 140, 150

답 두 자료의 표준편차는 같다.

상 표준편차의 의미를 다양하게 사고할 수 있는 토론을 하게 한다. 교과서 172쪽 확인하기 4를 참조할 수 있다.

참고 자료

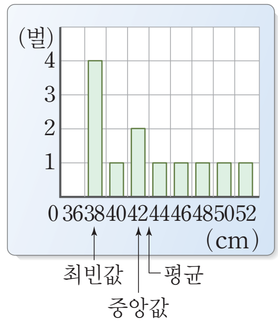
다음은 두 옷가게 A, B에서 하루 동안 판매된 셔츠의 목둘레를 조사하여 줄기와 잎 그림으로 나타낸 것이다.

[옷가게 A]

줄기와 잎 그림 (3 | 8은 38cm)

줄기	잎
3	8 8 8 8
4	2 0 4 2 6 8
5	0 2

막대 그래프



대푯값: 최빈값: 38cm, 중앙값: 42cm, 평균: 43cm

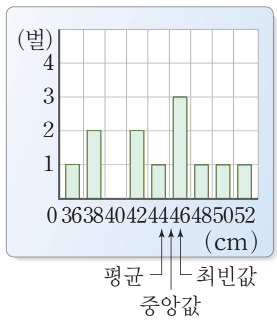
산포도: (분산)$=23$, (표준편차)$=\sqrt{23}$cm

[옷가게 B]

줄기와 잎 그림 (3 | 6은 36cm)

줄기	잎
3	6 8 8
4	2 2 4 6 6 6 8
5	0 2

막대 그래프



대푯값: 평균: 44cm, 중앙값: 45cm, 최빈값: 46cm

산포도: (분산)$=\frac{68}{3}$, (표준편차)$=\sqrt{\frac{68}{3}}$cm

위의 그림과 같이 자료의 분포 상태에 따라 평균, 중앙값, 최빈값의 위치가 달라짐을 알 수 있다.

또 옷가게 B의 표준편차가 더 작으므로 옷가게 B의 자료가 옷가게 A의 자료보다 판매된 옷의 사이즈가 평균에 가깝게 분포되어 있음을 알 수 있다.

IV

3/3차시 차시별 학습 지도 방법 [교과 교실]

생각 열기	도수분포표에서 평균을 구하는 것을 상기시키면서 빈칸을 채우고 풀어 보게 한다.
본문	생각 열기의 도수분포표를 이용하여 분산과 표준편차를 구하는 순서와 그 방법이 충분히 이해되도록 설명한다.
문제 5, 6	**문제 5**는 스스로 풀어 보게 하고, **문제 6**은 모둠별로 해결하고 토론하게 한다.

도수분포표에서 분산, 표준편차는 어떻게 구할까?

생각 열기 도수분포표에서 편차를 구하기 위한 생각 열기이다.

지도상의 유의점 각각의 변량을 알 수 없으므로 계급의 계급값이 변량을 대신함을 알게 한다.

(1) 빈칸을 채우면 다음과 같다.

계급(점)	도수(명)	계급값	(계급값)×(도수)
70 이상 ~ 90 미만	3	80	80×3
90 ~ 110	5	100	100×5
110 ~ 130	1	120	120×1
130 ~ 150	1	140	140×1
합계	10		1000

(2) $(\text{평균})=\dfrac{\{(\text{계급값})\times(\text{도수})\}\text{의 총합}}{(\text{도수})\text{의 총합}}$

$=\dfrac{1000}{10}=100(\text{점})$

(3) (편차)=(계급값)−(평균)이므로 다음과 같다.

계급값	80	100	120	140
편차	−20	0	20	40

6 도수분포표에서의 편차는

(편차)=(계급값)−(평균)

이다. 또 도수분포표에서 분산과 표준편차는

$(\text{분산})=\dfrac{\{(\text{편차})^2\times(\text{도수})\}\text{의 총합}}{(\text{도수})\text{의 총합}}$

$(\text{표준편차})=\sqrt{(\text{분산})}$

임을 이해하게 한다.

170쪽

3/3차시 도수분포표에서 분산, 표준편차는 어떻게 구할까?

생각 열기 다음 표는 볼링 반 학생 10명의 점수를 조사하여 나타낸 도수분포표이다.

계급(점)	도수(명)	계급값	(계급값)×(도수)
70 이상 ~ 90 미만	3	80	80×3
90 ~ 110	5		
110 ~ 130	1		
130 ~ 150	1		
합계	10		

(1) 빈칸을 채워 보자.
(2) 볼링 점수의 평균을 구하여 보자.
(3) 이 도수분포표에서 편차를 어떻게 구할 수 있는지 말하여 보자.

도수분포표로 주어진 자료의 편차는 다음과 같이 계급값에서 평균을 빼서 구한다.

(편차)=(계급값)−(평균)

6 이때 도수분포표에서 분산과 표준편차는 다음과 같은 순서로 구한다.

❶ 도수분포표의 각 계급에 대하여 각각

(계급값)×(도수)

의 값을 적고, 그 총합을 구한다.

❷ ❶에서 구한 총합을 도수의 총합으로 나누어 평균을 구한다.

❸ 각 계급값에서 평균을 뺀 편차를 구한다.

❹ 각 계급에 대하여 각각 $(\text{편차})^2\times(\text{도수})$의 값을 적고, 그 총합을 구한다.

❺ ❹에서 구한 총합을 도수의 총합으로 나누어 분산을 구한다.

❻ ❺에서 구한 분산의 음이 아닌 제곱근을 구하면 표준편차를 얻는다.

실제로 생각 열기의 도수분포표에서 분산과 표준편차를 구하면 다음과 같다.

❷ $(\text{평균})=\dfrac{\{(\text{계급값})\times(\text{도수})\}\text{의 총합}}{(\text{도수})\text{의 총합}}=\dfrac{1000}{10}=100$

❸ (편차)=(계급값)−(평균)을 이용하여 편차를 구한다.

❹ $(\text{편차})^2\times(\text{도수})$의 값을 적고, 그 총합을 구한다.

170 Ⅳ. 통계

7 도수분포표를 이용한 평균, 표준편차는 원래 자료의 평균, 표준편차보다 정확도가 떨어진다. 왜냐하면 한 계급의 계급값으로 그 자료의 변량들을 대신하기 때문이다.

참고 **도수분포표를 이용한 자료의 의미**

도수분포표와 같이 계급을 이용하여 조사한 자료를 그룹화 자료라고 한다.

이러한 그룹화 자료는 극히 개인적인 성향의 자료를 조사할 때 쓰인다.

예를 들어, 개인 소득, 정치적 성향 등 남들에게 말하기 어려운 사실들을 조사하여 통계적 수치로 나타낼 때 유용한 방법이다.

문제 5 **도수분포표에서 분산과 표준편차 구하기**

풀이 각 계급의 계급값과 (계급값)×(도수)를 구한 후, 평균을 구하고 (편차)=(계급값)−(평균)을 구한다. 또 $(\text{편차})^2\times(\text{도수})$를 구하여 표를 만들면 다음과 같다.

171쪽

계급(점)	도수(명)	계급값	(계급값)×(도수)	❸ 편차	$(편차)^2\times(도수)$
70 이상 ~ 90 미만	3	80	80×3	−20	$(-20)^2\times3=1200$
90 ~ 110	5	100	100×5	0	0
110 ~ 130	1	120	120×1	20	$(20)^2\times1=400$
130 ~ 150	1	140	140×1	40	$(40)^2\times1=1600$
합계	10		❶ 1000		❹ 3200

따라서 구하는 분산과 표준편차는

❺ $(분산)=\dfrac{\{(편차)^2\times(도수)\}의\ 총합}{(도수)의\ 총합}=\dfrac{3200}{10}=320$

❻ $(표준편차)=\sqrt{(분산)}=\sqrt{320}=8\sqrt{5}=17.888\cdots(점)$

이므로 표준편차를 반올림하여 소수 둘째 자리까지 구하면 17.89이다.

7 이상을 정리하면 다음과 같다.

도수분포표에서의 분산과 표준편차

$(분산)=\dfrac{\{(편차)^2\times(도수)\}의\ 총합}{(도수)의\ 총합}$ $(표준편차)=\sqrt{(분산)}$

문제 5 오른쪽 표는 지연이네 모둠 10명의 월요일부터 금요일까지의 TV 시청 시간을 조사하여 나타낸 도수분포표이다. 분산과 표준편차를 구하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)

TV 시청(시간)	학생 수(명)
0 이상 ~ 2 미만	1
2 ~ 4	2
4 ~ 6	4
6 ~ 8	2
8 ~ 10	1
합계	10

의사소통

문제 6 오른쪽 표는 민영이네 반 20명과 지원이네 반 20명의 지난 일 년 동안의 공연 또는 연극 관람 횟수를 조사하여 나타낸 도수분포표이다. 물음에 답하여라.

(1) 두 반의 평균과 표준편차를 모두 구하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)

(2) 두 반을 비교하여 알 수 있는 것을 모두 말하여라.

관람 횟수(회)	학생 수(명)	
	민영이네 반	지원이네 반
0 이상 ~ 4 미만	6	2
4 ~ 8	3	4
8 ~ 12	1	8
12 ~ 16	5	4
16 ~ 20	5	2
합계	20	20

1. 대푯값과 산포도 171

계급(시간)	도수(명)	계급값	(계급값)×(도수)	편차	$(편차)^2\times(도수)$
0 이상 ~ 2 미만	1	1	1	−4	16
2 ~ 4	2	3	6	−2	8
4 ~ 6	4	5	20	0	0
6 ~ 8	2	7	14	2	8
8 ~ 10	1	9	9	4	16
합계	10		50		48

$(평균)=\dfrac{\{(계급값)\times(도수)\}의\ 총합}{(도수)의\ 총합}=\dfrac{50}{10}=5(시간)$

$(분산)=\dfrac{\{(편차)^2\times(도수)\}의\ 총합}{(도수)의\ 총합}=\dfrac{48}{10}=4.8$

$(표준편차)=\sqrt{4.8}=2.190\cdots(시간)$

따라서 분산 4.8, 표준편차는 2.19시간이다.

문제 6 도수분포표에서 분산과 표준편차 구하기

풀이 (1) 민영이네 반의 각 계급의 계급값, (계급값)×(도수), (편차), $(편차)^2\times(도수)$를 구하여 표를 만들면 다음과 같다.

계급(회)	도수(명)	계급값	(계급값)×(도수)	편차	$(편차)^2\times(도수)$
0 이상 ~ 4 미만	6	2	12	−8	384
4 ~ 8	3	6	18	−4	48
8 ~ 12	1	10	10	0	0
12 ~ 16	5	14	70	4	80
16 ~ 20	5	18	90	8	320
합계	20		200		832

평균, 분산, 표준편차를 구하면

$(평균)=\dfrac{200}{20}=10(회)$, $(분산)=\dfrac{832}{20}=41.6$

$(표준편차)=\sqrt{41.6}=6.449\cdots(회)$

따라서 민영이네 반의 평균은 10회, 표준편차는 6.45회이다.

지원이네 반의 각 계급의 계급값, (계급값)×(도수), (편차), $(편차)^2\times(도수)$를 구하여 표를 만들면 다음과 같다.

계급(회)	도수(명)	계급값	(계급값)×(도수)	편차	$(편차)^2\times(도수)$
0 이상 ~ 4 미만	2	2	4	−8	128
4 ~ 8	4	6	24	−4	64
8 ~ 12	8	10	80	0	0
12 ~ 16	4	14	56	4	64
16 ~ 20	2	18	36	8	128
합계	20		200		384

평균, 분산, 표준편차를 구하면

$(평균)=\dfrac{200}{20}=10(회)$, $(분산)=\dfrac{384}{20}=19.2$

$(표준편차)=\sqrt{19.2}=4.381\cdots(회)$

따라서 지원이네 반의 평균은 10회, 표준편차는 4.38회이다.

(2) 민영이네 반과 지원이네 반의 관람 회수 평균은 같다. 이때 표준편차가 더 작은 지원이네 반의 학생들의 관람 횟수가 평균에 가까이 몰려있다고 할 수 있다.

수준별 교수·학습 방법

분산과 표준편차의 의미를 알고 구할 수 있다.

하 간단한 도수분포표에서의 분산과 표준편차를 구할 수 있게 한다.

[문제] 오른쪽 표는 10명의 학생이 물속에서 숨을 참는 시간을 도수분포표로 나타낸 것이다. 표준편차를 구하여라.

계급(초)	도수(명)
0 이상 ~ 20 미만	2
20 ~ 40	6
40 ~ 60	2
합계	10

답 $4\sqrt{10}$초

상 도수분포표의 두 자료를 비교하여 다른 점을 설명하게 한다.
교과서 172쪽 **문제 4**를 참조할 수 있다.

IV

확인하기

1 평가의 주안점 평균, 분산, 표준편차를 구할 수 있다.

풀이 (1) $(\text{평균})=\dfrac{5+10+15+20+25}{5}=\dfrac{75}{5}=15$

각 변량에 대한 편차와 편차의 제곱을 구하면 다음과 같다.

변량	5	10	15	20	25	합계
편차	-10	-5	0	5	10	0
$(\text{편차})^2$	100	25	0	25	100	250

$(\text{분산})=\dfrac{250}{5}=50$, $(\text{표준편차})=\sqrt{50}=7.071\cdots$

따라서 평균은 15, 분산은 50, 표준편차는 7.07이다.

(2) $(\text{평균})=\dfrac{6235+6240+6245+6250+6255}{5}=6245$

각 변량에 대한 편차와 편차의 제곱을 구하면 다음과 같다.

변량	6235	6240	6245	6250	6255	합계
편차	-10	-5	0	5	10	0
$(\text{편차})^2$	100	25	0	25	100	250

$(\text{분산})=\dfrac{250}{5}=50$, $(\text{표준편차})=\sqrt{50}=7.071\cdots$

따라서 평균은 6245, 분산은 50, 표준편차는 7.07이다.

2 평가의 주안점 평균과 분산을 구할 수 있다.

풀이 (1) $\dfrac{20+25+\boxed{30}+\boxed{35}+40}{5}=\boxed{30}$ (분)

(2) 편차를 구하면 다음과 같다.

자료(분)	20	25	30	35	40	합계
편차	-10	-5	$\boxed{0}$	$\boxed{5}$	10	0
$(\text{편차})^2$	100	25	$\boxed{0}$	$\boxed{25}$	100	$\boxed{250}$

(3) $(\text{분산})=\dfrac{\{(\text{편차})^2\text{의 총합}\}}{5}=\dfrac{\boxed{250}}{5}=\boxed{50}$

3 평가의 주안점 도수분포표에서 분산과 표준편차를 구할 수 있다.

풀이

계급(분)	도수(명)	계급값	(계급값)×(도수)	편차	$(\text{편차})^2$×(도수)
0 이상 ~ 10 미만	6	5	30	-12	864
10 ~ 20	8	15	120	-2	32
20 ~ 30	3	25	75	8	192
30 ~ 40	2	35	70	18	648
40 ~ 50	1	45	45	28	784
합계	20		340		2520

$(\text{평균})=\dfrac{340}{20}=17(\text{분})$, $(\text{분산})=\dfrac{2520}{20}=126$

$(\text{표준편차})=\sqrt{126}=11.224\cdots(\text{분})$

따라서 분산은 126, 표준편차는 11.22분이다.

172쪽

확인하기

1 다음 자료의 평균, 분산, 표준편차를 각각 구하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)

(1) 5, 10, 15, 20, 25　　(2) 6235, 6240, 6245, 6250, 6255

2 오른쪽 그림은 연지네 반 모둠 5명의 하루 동안의 가족 간 대화 시간을 조사하여 나타낸 줄기와 잎 그림이다. □ 안에 알맞은 것을 써넣어라.

대화 시간 (2 | 0은 20분)

줄기	잎
2	0 5
3	0 5
4	0

(1) $(\text{평균})=\dfrac{20+25+\square+\square+40}{5}=\square(\text{분})$

(2)

자료(분)	20	25	30	35	40	합계
편차	-10	-5	$\square$	$\square$	10	0
$(\text{편차})^2$	100	25	$\square$	$\square$	100	$\square$

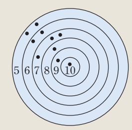
(3) $(\text{분산})=\{(\text{편차})^2\text{의 평균}\}=\dfrac{\{(\text{편차})^2\text{의 총합}\}}{5}=\dfrac{\square}{5}=\square$

3 오른쪽 표는 윤서네 반 학생 20명이 등교에 걸리는 시간을 조사하여 나타낸 도수분포표이다. 분산과 표준편차를 구하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)

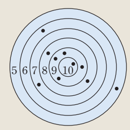
등교에 걸리는 시간(분)	학생 수(명)
0 이상 ~ 10 미만	6
10 ~ 20	8
20 ~ 30	3
30 ~ 40	2
40 ~ 50	1
합계	20

수학적 과정 의사소통 추론 문제 해결

4 현진이와 종신이는 동아리 활동으로 사격 대회 경기를 관람하였다. 다음은 두 선수 A, B가 10발씩 사격한 표적이라 할 때, 어떤 선수의 점수가 더 고른지 말하여라.



[A 선수의 표적]



[B 선수의 표적]

172 Ⅳ. 통계

4 평가의 주안점 표준편차의 의미를 토론할 수 있다.

풀이 [A 선수의 점수]

점수(점)	5	5	6	6	7	7	7	8	9	10	합계
편차	-2	-2	-1	-1	0	0	0	1	2	3	0
$(\text{편차})^2$	4	4	1	1	0	0	0	1	4	9	24

$(\text{평균})=\dfrac{5+5+6+6+7+7+7+8+9+10}{10}=\dfrac{70}{10}=7(\text{점})$

$(\text{분산})=\dfrac{24}{10}=2.4$, $(\text{표준편차})=\sqrt{2.4}(\text{점})$

[B 선수의 점수]

점수(점)	5	6	7	8	8	9	9	9	9	10	합계
편차	-3	-2	-1	0	0	1	1	1	1	2	0
$(\text{편차})^2$	9	4	1	0	0	1	1	1	1	4	22

$(\text{평균})=\dfrac{5+6+7+8+8+9+9+9+9+10}{10}=\dfrac{80}{10}=8(\text{점})$

$(\text{분산})=\dfrac{22}{10}=2.2$, $(\text{표준편차})=\sqrt{2.2}(\text{점})$

두 선수 중 B 선수의 점수의 표준편차가 더 작으므로 A 선수의 점수보다 B 선수의 점수가 평균에 더 가까이 몰려 있고 점수의 분포가 더 고르다.

173쪽

중단원 마무리하기

스스로 정리하기

1. 다음 □ 안에 알맞은 것을 써넣어라.

(1) 자료 전체의 중심적인 경향이나 특징을 하나의 수로 나타내어 자료 전체를 대표하는 값을 □이라고 한다.

(2) 자료를 작은 값부터 크기순으로 나열할 때 중앙에 위치한 값을 □이라고 한다.

(3) 자료의 값 중에서 가장 많이 나타나는 값을 □이라고 한다.

(4) 자료의 분포 상태에서 변량이 흩어진 정도를 하나의 수로 나타내는 값을 □라고 하고, 각 변량에서 평균을 뺀 값을 그 변량의 □라고 한다.

(5) 산포도에서 편차의 제곱의 평균을 □이라 하고, 분산의 음이 아닌 제곱근을 □라고 한다.

기초 다지기

1 아래는 어느 인터넷 카페 회원 12명의 나이를 조사한 자료이다. 다음을 구하여라.

(단위: 세)

14	16	17	18	19	15
15	15	15	19	16	25

(1) 평균 (2) 중앙값 (3) 최빈값

대푯값

2 오른쪽 표는 1반에서 5반까지의 과학 성적을 조사하여 얻은 것이다. 다음 중 옳은 것을 모두 찾아라.

(단위: 점)

구분	1반	2반	3반	4반	5반
평균	68	68	70	72	64
표준편차	4.2	10.5	6.5	8.5	2.4

ㄱ. 1반이 2반보다 성적이 고르다.
ㄴ. 편차의 총합이 가장 큰 반은 2반이다.
ㄷ. 표준편차로는 분산이 가장 큰 반을 알 수 없다.
ㄹ. 가장 성적이 고른 반은 5반이다.

산포도

1. 대푯값과 산포도 173

중단원 마무리하기

스스로 정리하기

1. (1) 대푯값 (2) 중앙값
(3) 최빈값 (4) 산포도, 편차
(5) 분산, 표준편차

기초 다지기

1 **평가의 주안점** 평균, 분산, 표준편차를 구할 수 있다.

풀이 (1) $(\text{평균})=\dfrac{14+16+17+18+19+15+15+15+15+19+16+25}{12}$

$=17(\text{세})$

(2) 인터넷 카페 회원 12명을 나이순으로 나열하면
14, 15, 15, 15, 15, 16, 16, 17, 18, 19, 19, 25
따라서 중앙값은 16세이다.

(3) 가장 많이 나타나는 나이는 15세이다.
따라서 최빈값은 15세이다.

2 **평가의 주안점** 평균과 표준편차의 의미를 이해할 수 있다.

풀이 ㄱ. 표준편차는 2반보다 1반이 작으므로 1반의 성적이 더 고르다.
ㄴ. 편차의 총합은 항상 0이다.
ㄷ. 표준편차로 분산이 가장 큰 반을 알 수 있다.
ㄹ. 5반이 표준편차가 가장 작으므로 성적이 가장 고르다.
따라서 옳은 것은 ㄱ, ㄹ이다.

기본 익히기

3 **평가의 주안점** 평균과 최빈값의 의미를 이해하고 추론할 수 있다.

풀이 설명에 나타나지 않은 치수를 $x$cm라 하면 평균이 98cm이므로

$\dfrac{90+100+100+x+105}{5}=98$, $x=95$

따라서 설명에 나타나지 않는 티셔츠의 치수는 95cm이다.

4 **평가의 주안점** 편차의 의미를 이해할 수 있다.

풀이 모든 편차의 총합은 항상 0이므로
$4+(-3)+x+2+(-2x)=0$
$x=3$
따라서 현수의 점수를 a점이라 하면
$3=a-20$이므로 $a=23$이다.
미성이의 점수를 b점이라 하면
$-2\times3=b-20$이므로 $b=14$이다.
따라서 현수와 미성이의 점수는 각각 23점, 14점이다.

5 **평가의 주안점** 자료에서 분산과 표준편차를 구할 수 있다.

풀이 평균을 구하면

$(\text{평균})=\dfrac{2+5+6+12+14+15}{6}=9$

	2	5	6	12	14	15	합계
편차	−7	−4	−3	3	5	6	0
$(\text{편차})^2$	49	16	9	9	25	36	144

$(\text{분산})=\dfrac{144}{6}=24$, $(\text{표준편차})=\sqrt{24}=4.898\cdots$

따라서 분산은 24, 표준편차는 4.90이다.

174쪽

기본 익히기

추론

3 동아리 활동으로 방송반 활동을 하고 있는 종수는 방송반 학생 5명의 티셔츠 치수를 다음과 같이 설명하였다. 주어진 설명에 나타나지 않은 나머지 하나의 티셔츠 치수를 구하여라.

최빈값

> ㈎ 가장 작은 티셔츠의 치수는 90cm이다.
> ㈏ 회원들의 티셔츠 치수의 평균은 98cm이다.
> ㈐ 한 티셔츠의 크기는 105cm이다.
> ㈑ 최빈값이 100cm인 변량이 2개가 있다.

4 다음 표는 학생 5명의 미술 실기 점수의 편차를 나타낸 것이다. 미술 실기 점수의 평균이 20점일 때, 현수와 미성이의 점수를 구하여라.

편차

학생	인혜	두식	현수	미실	미성
편차	4	-3	x	2	$-2x$

5 다음 자료의 분산과 표준편차를 구하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)

분산과 표준편차

2 5 6 12 14 15

서술형

6 오른쪽 그림은 학생 12명의 줄넘기 횟수를 조사하여 나타낸 줄기와 잎 그림이다. 분산과 표준편차를 구하고, 그 과정을 서술하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)

분산과 표준편차

줄넘기 횟수 (1 | 3은 13회)

줄기	잎
1	3 6
2	8 4
3	4 6 2
4	2 1 6
5	5 3

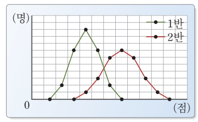
175쪽

추론

7 오른쪽 그림은 학생 수가 같은 1반과 2반의 음악 실기 점수를 나타낸 도수분포다각형이다. 물음에 답하여라.

(1) 어떤 반의 평균 점수가 더 높은가?

(2) 어떤 반이 점수가 더 고른가?



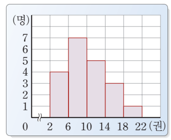
산포도

실력 기르기

8 오른쪽 그림은 연수네 반 20명이 1년 동안 읽은 독서량을 조사하여 나타낸 히스토그램이다. 물음에 답하여라.

(1) 독서량의 평균을 구하여라.

(2) 독서량의 분산과 표준편차를 구하여라.
(단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)



히스토그램을 이용하여 도수분포표를 만들고 평균, 분산, 표준편차를 구한다.

문제 해결 서술형

9 현선이는 어머니 심부름으로 사과를 사러 마트에 갔다. 사과 5개를 하나씩 저울에 달아 재어 보고, 최대한 비슷한 크기로 골랐다. 집에 와서 사과의 무게의 평균을 구했더니 평균은 130g, 분산은 10.8이었다. 그런데 검산을 해 보니, 127g, 132g 사과의 무게를 125g, 134g으로 잘못 적어 계산하였다. 물음에 답하고, 그 과정을 서술하여라.

(1) 실제 사과 무게의 평균을 구하여라.

(2) 실제 사과 무게의 분산과 표준편차를 구하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)

처음 계산한 사과 무게의 분산과 실제 사과 무게의 분산을 구하는 식을 이용하여 실제 사과의 분산을 구한다.

---

6 **평가의 주안점** 줄기와 잎 그림의 분산과 표준편차를 구할 수 있다.

풀이

$$(\text{평균})=\frac{13+16+28+24+34+36+32+42+41+46+55+53}{12}$$

$=35(\text{회})$ ······ ❶

회	13	16	24	28	32	34	36	41	42	46	53	55	합계
편차	-22	-19	-11	-7	-3	-1	1	6	7	11	18	20	0
$(\text{편차})^2$	484	361	121	49	9	1	1	36	49	121	324	400	1956

······ ❷

$(\text{분산})=\frac{1956}{12}=163$

$(\text{표준편차})=\sqrt{163}=12.767\cdots(\text{회})$ ······ ❸

따라서 분산은 163, 표준편차는 12.77회이다.

단계	채점 기준	배점 비율
❶	평균을 구한다.	30%
❷	편차와 $(\text{편차})^2$을 구한다.	40%
❸	분산과 표준편차를 구한다.	30%

7 **평가의 주안점** 그래프를 보고 평균과 표준편차를 추론하여 볼 수 있다.

풀이 (1) 2반의 그래프가 1반의 그래프보다 더 오른쪽으로 높은 점수에 분포하므로 2반의 평균 점수가 더 높다.

(2) 1반은 평균 점수를 기준으로 편차가 작고, 평균에 더 몰려 있다. 따라서 1반이 2반보다 점수가 더 고르다.

실력 기르기

8 **평가의 주안점** 히스토그램에서의 분산과 표준편차를 구할 수 있다.

풀이 각 계급의 계급값, (계급값)×(도수)를 구한 후 평균을 구한다. 또 (편차)=(계급값)−(평균), $(\text{편차})^2\times(\text{도수})$를 구한다.

독서량(권)	도수(명)	계급값	(계급값) ×(도수)	편차	$(\text{편차})^2$ ×(도수)
2 이상 ~ 6 미만	4	4	16	-6	144
6 ~ 10	7	8	56	-2	28
10 ~ 14	5	12	60	2	20
14 ~ 18	3	16	48	6	108
18 ~ 22	1	20	20	10	100
합계	20		200		400

(1) (평균)$=\dfrac{200}{20}=10$(권)

(2) (분산)$=\dfrac{400}{20}=20$

(표준편차)$=\sqrt{20}=4.472\cdots$(권)

따라서 분산은 20, 표준편차는 4.47권이다.

9 **평가의 주안점** 실생활에서 평균, 분산, 표준편차를 구할 수 있다.

풀이 (1) 처음 계산한 사과의 무게를 $a, b, c, 125, 134$ 라 하면 실제 무게는 $a, b, c, 127, 132$ 이므로 전체 무게는 변화가 없다. 즉 평균은 130g으로 같다. …… ❶

(2) 처음 계산한 사과 무게의 분산은

$$\frac{(a-130)^2+(b-130)^2+(c-130)^2+(125-130)^2+(134-130)^2}{5}$$

$=10.8$ …… ㉠

실제 사과의 분산은

$$\frac{(a-130)^2+(b-130)^2+(c-130)^2+(127-130)^2+(132-130)^2}{5}$$

$=\boxed{}$ …… ㉡

㉠−㉡을 하면

$$\frac{(125-130)^2+(134-130)^2}{5}-\frac{(127-130)^2+(132-130)^2}{5}$$

$=10.8-\boxed{}$

$\boxed{}=10.8-5.6=5.2$

따라서 구하는 실제 사과의 분산은 5.2이다. …… ❷

이때, $\sqrt{5.2}=2.280\cdots$이므로 표준편차는 2.28g이다. …… ❸

단계	채점 기준	배점 비율
❶	평균을 구한다.	30%
❷	두 가지 분산의 식을 세워 실제 분산을 구한다.	50%
❸	표준편차를 구한다.	20%

창의·인성 키우기

개념 바루기

지도상의 유의점 도수가 주어진 표에서의 분산은 (편차)^{2}의 평균으로 구하는 것이 아니라 (편차)$^2\times$(도수)의 평균으로 구함에 유의하도록 지도한다.

올바른 풀이 도수분포표에서의 분산은 {(편차)$^2\times$(도수)}의 평균이므로

(분산)$=\dfrac{(-4)^2\times 2+(-1)^2\times 3+2^2\times 4+3^2\times 1}{5}=\dfrac{60}{10}=6$

(표준편차)$=\sqrt{6}$(회)

이다.

176쪽

창의·인성 키우기

개념 바루기 윤후네 동아리 학생 10명은 고궁에 가서 투호 놀이를 하였다. 윤후는 친구들과 함께 화살을 15번씩 던져 얻은 결과를 다음과 같이 표로 나타내었고, 이를 이용하여 분산과 표준편차를 구하였다. 윤후의 풀이에서 옳지 않은 부분을 찾아 바르게 고쳐라.

투호 성공 횟수(회)	학생 수(명)
3	2
6	3
9	4
10	1
합계	10

변량의 합은 $3\times2+6\times3+9\times4+10\times1=70$이므로 평균은

(평균)$=\dfrac{70}{10}=7$(회)

각 변량의 편차를 구하면 $-4, -1, 2, 3$이므로 분산과 표준편차는

(분산)$=\dfrac{(-4)^2+(-1)^2+2^2+3^2}{10}=\dfrac{30}{10}=3$

(표준편차)$=\sqrt{3}$(회)

생각 키우기 수정이는 가족들과 함께 농장 체험 테마 여행을 갔다. 그곳에서 수정이는 방울토마토를 따서 그 무게를 잰 후 평균과 편차를 구하였다. 집에 돌아온 후 일주일이 지나 여행 가방을 정리해 보니 노트에 적어 두었던 자료가 오른쪽 그림과 같이 훼손되었다. 물음에 답하여라.

방울토마토의 무게

무게(g)	도수(개)	편차
0 이상 ~ 3 미만	4	−2
3 ~ 5		0
5 ~ 7		2
7 ~ 9	1	4
합계		

(1) 5g 이상 7g 미만의 방울토마토의 개수를 구하여라.

(2) 3g 이상 5g 미만의 방울토마토의 개수에 따라 표준편차가 어떻게 변하는지 설명하여라.

176 Ⅳ. 통계

창의 · 인성 위의 문항 외에도 일어날 수 있는 또 다른 오류를 발표해 볼 수 있는 시간을 제공하여 서로의 경험을 공유할 수 있도록 한다.

참고 투호란 병(壺)을 놓고 일정한 거리의 양쪽에 각각 한 사람씩 푸른 살(矢)과 흰 살(矢) 혹은 붉은 살(矢)을 던지는 놀이이다. 이 살을 던져 병 속에 많이 넣는 사람이 이긴다.

생각 키우기

지도상의 유의점 자료의 도수와 편차의 의미를 이해하고 표준편차를 활용하며 표준편차와 미지수의 관계를 이해할 수 있도록 지도한다.

풀이 (1) (편차)$\times$(도수)의 합은 항상 0이므로

5g 이상 7g 미만의 방울토마토의 개수를 a라 놓으면

$4\times(-2)+0+a\times2+1\times4=0$이다.

따라서 $a=2$이다.

177쪽

컴퓨터로 하는 수학

스프레드시트 프로그램을 이용하여 대푯값과 산포도 구하기

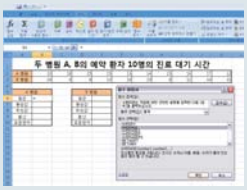
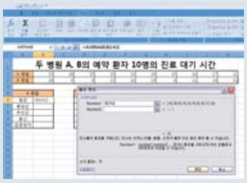
스프레드시트 프로그램을 이용하여 166쪽의 두 병원 A, B의 예약 환자 10명에 대한 진료 대기 시간의 평균, 중앙값, 최빈값 및 분산과 표준편차를 구하여 보자.

(단위: 분)

A 병원	15	18	15	15	13	14	15	16	17	12
B 병원	24	6	8	13	16	19	8	11	14	31

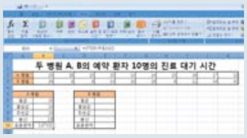
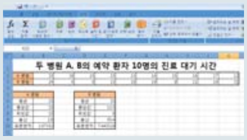
❶ 위의 표에서 A 병원의 자료를 시트에 입력한 후 평균을 구하고자 하는 셀을 선택한다. 그리고 메뉴에서 '수식' → '함수 삽입'을 클릭하여 나오는 함수 마법사 창에서 범주는 '통계', 함수는 [AVERAGE]를 선택하고 확인을 클릭한다.

❷ 함수 인수 창에서 자료가 입력된 셀인 B2 : K2를 직접 입력하거나 자료를 드래그한 후 확인을 클릭하여 평균을 구한다.

❸ 마찬가지 방법으로 함수 마법사 창에서 함수를 [MEDIAN]으로 선택하여 중앙값을, [MODE]로 선택하여 최빈값을, [VARP]로 선택하여 분산을, [STDEVP]로 선택하여 표준편차를 구하여 본다.

❹ B 병원에 대해서도 위의 절차를 거쳐 평균, 중앙값, 최빈값, 분산, 표준편차를 구한다.

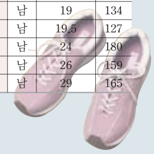



178쪽

다음은 학생 60명의 성별, 발의 크기, 키를 조사하여 정리한 자료이다. 남학생과 여학생의 두 그룹으로 나누고, 스프레드시트 프로그램을 이용하여 다음을 구하여 보자.

(단위: cm)

번호	구분	발의 크기	키	번호	구분	발의 크기	키	번호	구분	발의 크기	키
1	여	25	160	21	여	20	133	41	남	25	184
2	여	23	160	22	여	20	131	42	남	18	125
3	여	23.5	152	23	여	17	134	43	남	27.5	170
4	여	24	146	24	여	25	170	44	남	23	131
5	여	24	157	25	여	15	125	45	남	23	149
6	여	23	143	26	여	21	135	46	남	22	149
7	여	25	153	27	여	19	138	47	남	20	126
8	여	20	150	28	여	25	171	48	남	24	150
9	여	20	140	29	여	24	181	49	남	26	170
10	여	25.5	168	30	여	19.5	139	50	남	20	125
11	여	21	156	31	여	24	164	51	남	25	165
12	여	19.5	130	32	여	23	138	52	남	25	158
13	여	22	142	33	여	23.5	151	53	남	20.5	134
14	여	24	159	34	남	15	111	54	남	22	145
15	여	21.5	145	35	남	21	136	55	남	25	147
16	여	25	162	36	남	20	147	56	남	19	134
17	여	24.5	169	37	남	20	133	57	남	19.5	127
18	여	21	141	38	남	23	148	58	남	24	180
19	여	20	123	39	남	20	125	59	남	26	159
20	여	19	122	40	남	28	183	60	남	29	165



과제 1 두 그룹에 대하여 발의 크기, 키의 평균, 중앙값, 최빈값, 분산, 표준편차를 구하여 보자.
(단, 평균, 분산, 표준편차는 소수 둘째 자리까지 구한다.)

과제 2 두 그룹의 발의 크기와 키의 대푯값과 분산, 표준편차를 고려하여 자료의 특징에 대해 토론하여 보자.

(2) 3g 이상 5g 미만의 방울토마토의 개수를 b라 놓으면 평균은 4g이므로

$$(\text{분산})=\frac{(2-4)^2\times4+(4-4)^2\times b+(6-4)^2\times2+(8-4)^2\times1}{7+b}$$

$$=\frac{40}{7+b}$$

$$(\text{표준편차})=\sqrt{\frac{40}{7+b}}\,(\text{g})$$

평균 4g의 도수인 b의 값이 커지면 표준편차는 작아지고, b의 값이 작아지면 표준편차는 커짐을 알 수 있다.

창의 · 인성 통계에서 배운 수학적 지식을 적절하게 활용하여 스스로 수학적 추측이나 주장을 만들어 적용할 수 있도록 한다.

컴퓨터로 하는 수학

지도상의 유의점 자료를 정리하고, 대푯값을 구할 때, 스프레드시트 프로그램이나 그 밖의 컴퓨터 프로그램을 이용하여 자료를 정리하고, 이를 확인할 수 있도록 지도한다.

풀이 **과제 1** 컴퓨터 프로그램을 이용하여 남학생과 여학생 각각에 대하여 대푯값을 구하면 다음과 같다.

	여학생	
	발(cm)	키(cm)
평균	22.02	148.12
중앙값	23	146
최빈값	25	160
분산	6.63	227.38
표준편차	2.57	15.08

	남학생	
	발(cm)	키(cm)
평균	22.61	147.26
중앙값	23	147
최빈값	20	125
분산	10.65	379.45
표준편차	3.26	19.48

과제 2 예시 답안 자료의 특징은 다음과 같다.

- 평균으로 봤을 때 여학생의 발은 남학생의 발보다 작은 반면 키는 여학생이 더 큰 편이다.
- 분산과 표준편차는 발 크기나 키 모두 남학생의 값이 여학생보다 더 크다.

이를 통해 남학생의 변량의 분포가 더 넓게 퍼져 있음을 알 수 있고, 키의 경우에는 여학생의 평균이 더 크게 나타나지만 남학생의 경우 더 넓은 분포를 가지고 있기 때문에 키가 더 큰 경우나 더 작은 경우의 학생이 남학생 중에 있다는 것을 알 수 있다.

창의 · 인성 컴퓨터 프로그램을 이용한 과제 해결을 통하여 공학적 도구를 이용하는 능력을 키우도록 한다.

학년	반 번호:	이름:	/ 점수:

선다형은 각 4점, 나머지 문항은 각 문항에 표시함.

01 다음 표는 현지네 모둠 7명에 대한 한 달 동안의 독서량을 나타낸 것이다. 독서량의 평균과 중앙값을 차례로 쓰면?

학생	현지	민주	진환	영주	명환	지상	만득
독서량(권)	6	2	1	2	4	1	5

① 2권, 2권 ② 2권, 3권 ③ 3권, 2권
④ 3권, 3권 ⑤ 3권, 4권

02 다음 자료는 우석이네 모둠 5명의 윗몸일으키기 횟수를 조사한 것이다. 중앙값이 32회일 때, x의 값은? (단, $x>0$)

(단위: 회)

$2x-2,\quad 2x+15,\quad 2x-5,\quad 2x+10,\quad 2x+8$

① 12 ② 13 ③ 14
④ 15 ⑤ 16

03 다음 자료의 중앙값이 10일 때, 최빈값은?

$12,\quad 8,\quad 9,\quad 12,\quad 11,\quad 8,\quad x+1,\quad 9,\quad 15$

① 8 ② 9 ③ 10
④ 11 ⑤ 12

04 오른쪽 표는 학생 20명의 가족 수를 나타낸 것이다. 이 때, 이 자료의 최빈값은?

가족 수(명)	학생 수(명)
2	1
3	8
4	6
5	3
6	2
합계	20

① 2명 ② 3명
③ 4명 ④ 5명
⑤ 6명

05 다음 중 옳은 것은?

① 평균은 유일한 대푯값이다.
② 편차의 제곱의 합은 분산이다.
③ 편차의 합은 0이 아닐 때도 있다.
④ (편차)=(변량)−(평균)이다.
⑤ 표준편차는 분산의 음의 제곱근이다.

06 다음 표는 학생 5명의 중간고사 사회 점수에 대한 편차를 나타낸 것이다. 이 학생들의 평균이 75점일 때, x의 값은?

학생	A	B	C	D	E
편차	-4	x	0	2	5

① -4 ② -3 ③ -2
④ 2 ⑤ 3

07 다음은 어떤 자료의 편차를 나타낸 것일 때, 이 자료의 표준편차는?

$-4,\quad -2,\quad 0,\quad 2,\quad 4$

① $\sqrt{3}$ ② $\sqrt{2}$ ③ 2
④ $2\sqrt{2}$ ⑤ $3\sqrt{2}$

08 변량 a, b, c, d, e의 평균을 M, 표준편차를 S라 할 때, $2a$, $2b$, $2c$, $2d$, $2e$의 평균과 표준편차를 차례로 쓰면?

① M, S ② M, $2S$ ③ $2M$, S
④ $2M$, $\sqrt{2}S$ ⑤ $2M$, $2S$

단답형

09 다음 자료에서 평균, 중앙값, 최빈값을 크기가 큰 순서로 차례대로 써라. [8점]

8 16 7 9 12 14 15 9

10 다음 줄기와 잎 그림은 연두네 반 학생 18명의 시험 점수를 나타낸 것이다. 중앙값과 최빈값을 구하여라. [8점]

시험 점수 (6 | 6은 66점)

줄기	잎
6	6 8 0
7	0 2 0 4 0
8	2 6 0 2 5 4
9	0 2 4
10	0

11 다음 자료는 민서네 모둠 5명의 과학 점수에 대한 편차이다. 과학 점수의 평균이 72점일 때, **보기**에서 옳은 것을 모두 골라라. [10점]

학생	A	B	C	D	E
편차	-3	-2	1	0	4

보기

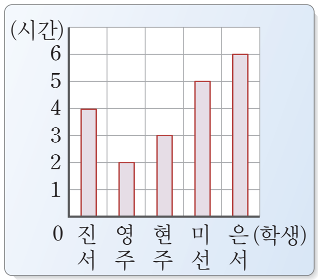
ㄱ. B 학생의 점수는 74점이다.
ㄴ. D 학생의 점수는 알 수 없다.
ㄷ. C 학생의 점수가 A 학생의 점수보다 4점 높다.
ㄹ. E 학생의 점수는 76점이다.

12 다음은 어떤 자료의 편차를 나타낸 것이다. 표준편차가 $\sqrt{6}$일 때, xy의 값을 구하여라. [8점]

-3 x 0 y 4

서술형

13 다음 그림은 진서네 모둠 5명에 대한 일주일 동안의 인터넷 사용 시간을 나타낸 것이다. 평균과 표준편차를 구하고, 그 과정을 서술하여라. (단, 표준편차는 반올림하여 소수 첫째 자리까지 구한다.) [10점]



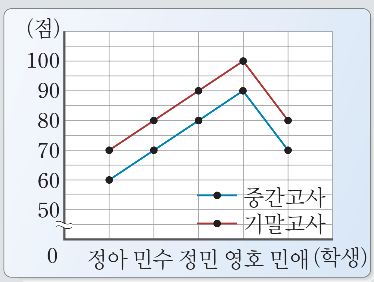
14 오른쪽 표는 진철이네 모둠 10명의 여름방학 봉사 활동 시간을 조사한 것이다. 평균과 표준편차를 구하여라. (단, 표준편차는 반올림하여 소수 첫째 자리까지 구한다.) [10점]

봉사 활동(시간)	학생 수(명)
0 이상 ~ 3 미만	1
3 ~ 5	3
5 ~ 7	2
7 ~ 9	3
9 ~ 11	1
합계	10

수리 논술형

15 다음 제시문을 읽고, 5명의 표준편차의 변화를 설명하여라. [14점]

정아, 민수, 정민, 영호, 인애는 기말고사 대비 스터디그룹을 만들었다. 방과 후 또는 주말에 도서관에 함께 모여 공부한 결과 5명 모두 성적이 향상되었다. 중간고사에 비해 기말고사의 성적이 얼마나 향상되었는지 알아보기 위하여 꺾은선 그래프를 그렸더니 오른쪽 그림과 같았다.



중단원 평가 문제

01 ③　02 ①　03 ②　04 ②
05 ④　06 ②　07 ④　08 ⑤
09 평균, 중앙값, 최빈값
10 중앙값: 81점, 최빈값: 70점
11 ㄷ, ㄹ　12 -2　13~15 풀이 참조

01 평가 기준　평균과 중앙값을 이해하고 구할 수 있는가?

풀이　(평균)$=\dfrac{6+2+1+2+4+1+5}{7}=3$(권)

자료를 작은 것부터 크기순으로 나열하면

$1,\quad 1,\quad 2,\quad 2,\quad 4,\quad 5,\quad 6$

이므로 중앙값은 2권이다.

02 평가 기준　중앙값을 이해하고, 구할 수 있는가?

풀이　자료를 작은 것부터 크기순으로 나열하면

$2x-5,\ 2x-2,\ 2x+8,\ 2x+10,\ 2x+15$

이므로 중앙값은 $2x+8=32$이고 $x=12$이다.

03 평가 기준　최빈값을 구할 수 있는가?

풀이　중앙값이 10이므로 $x+1=10$, $x=9$

따라서 최빈값은 9이다.

04 평가 기준　도수분포표에서 최빈값을 구할 수 있는가?

풀이　3명의 가족이 있는 학생이 8명이므로 최빈값은 3명이다.

05 평가 기준　통계의 용어를 알고 있는가?

풀이　① 대푯값에는 평균, 중앙값, 최빈값 등이 있다.

② 편차의 제곱의 평균이 분산이다.

③ 편차의 합은 항상 0이다.

⑤ 표준편차는 분산의 음이 아닌 제곱근이다.

06 평가 기준　편차의 의미를 알고 있는가?

풀이　편차의 합은 항상 0이므로

$-4+x+2+5=0$

$x=-3$

07 평가 기준　편차를 이용하여 표준편차를 구할 수 있는가?

풀이　(분산)$=\dfrac{(-4)^2+(-2)^2+2^2+4^2}{5}=\dfrac{40}{5}=8$

(표준편차)$=\sqrt{8}=2\sqrt{2}$

08 평가 기준　표준편차의 의미를 알고, 이를 구할 수 있는가?

풀이　$M=\dfrac{a+b+c+d+e}{5}$

$S=\sqrt{\dfrac{(a-M)^2+(b-M)^2+(c-M)^2+(d-M)^2+(e-M)^2}{5}}$

이므로 $2a$, $2b$, $2c$, $2d$, $2e$의 평균과 표준편차는

(평균)$=\dfrac{2a+2b+2c+2d+2e}{5}$

$=2\times\dfrac{a+b+c+d+e}{5}$

$=2M$

(표준편차)

$=\sqrt{\dfrac{(2a-2M)^2+(2b-2M)^2+(2c-2M)^2+(2d-2M)^2+(2e-2M)^2}{5}}$

$=\sqrt{4\times\dfrac{(a-M)^2+(b-M)^2+(c-M)^2+(d-M)^2+(e-M)^2}{5}}$

$=2S$

09 평가 기준　평균, 중앙값, 최빈값을 구할 수 있는가?

풀이　(평균)$=\dfrac{8+16+7+9+12+14+15+9}{8}$

$=\dfrac{90}{8}=11.25$

작은 것부터 크기순으로 나열하면

$7,\quad 8,\quad 9,\quad 9,\quad 12,\quad 14,\quad 15,\quad 16$

이므로 중앙값은 $\dfrac{9+12}{2}=10.5$이다.

최빈값은 9이다.

따라서 크기가 큰 순서대로 쓰면 평균, 중앙값, 최빈값이다.

10 평가 기준 줄기와 잎 그림에서 중앙값과 최빈값을 구할 수 있는가?

풀이 줄기와 잎 그림을 순서대로 나열하면 다음 표와 같다.

줄기	잎
6	0 6 8
7	0 0 0 2 4
8	0 2 2 4 5 6
9	0 2 4
10	0

따라서 중앙값은 81점이고 최빈값은 70점이다.

11 평가 기준 편차를 이해하고 있는가?

풀이 ㄱ. B 학생의 점수는 70점이다.
ㄴ. D 학생의 점수는 72점이다.
따라서 옳은 것은 ㄷ, ㄹ이다.

12 평가 기준 표준편차를 이해하고 있는가?

풀이 모든 편차의 합은 0이므로
$-3+x+0+y+4=0$에서
$x+y=-1$ ······ ㉠
(분산)$=\dfrac{(-3)^2+x^2+y^2+4^2}{5}=\dfrac{25+x^2+y^2}{5}$
$\sqrt{\dfrac{25+x^2+y^2}{5}}=\sqrt{6}$에서 $\dfrac{25+x^2+y^2}{5}=6$이므로
$x^2+y^2=5$ ······ ㉡
㉡을 변형하면 $(x+y)^2-2xy=5$ ······ ㉢
㉠을 ㉢에 대입하면 $(-1)^2-2xy=5$
따라서 $xy=-2$이다.

13 평가 기준 그래프로 나타낸 자료의 평균과 표준편차를 구할 수 있는가?

풀이 (평균)$=\dfrac{4+2+3+5+6}{5}$
$=\dfrac{20}{5}=4$(시간) ······ ❶
(분산)$=\dfrac{(2-4)^2+(3-4)^2+(5-4)^2+(6-4)^2}{5}$
$=\dfrac{10}{5}=2$ ······ ❷
(표준편차)$=\sqrt{2}=1.41\cdots$(시간)
따라서 평균은 4시간, 표준편차는 1.4시간이다. ······ ❸

단계	채점 기준	배점
❶	평균을 구한다.	3점
❷	분산을 구한다.	5점
❸	표준편차를 구한다.	2점

14 평가 기준 도수분포표로 나타낸 자료의 평균과 표준편차를 구할 수 있는가?

풀이 주어진 도수분포표에서 계급값, 편차, (편차)^2×(도수)를 구하여 표를 만들면 다음과 같다.

계급(시간)	도수(명)	계급값	편차	(편차)^2×(도수)
0 이상 ~ 3 미만	1	2	-4	16
3 ~ 5	3	4	-2	12
5 ~ 7	2	6	0	0
7 ~ 9	3	8	2	12
9 ~ 11	1	10	4	16
합계	10		0	56

(평균)$=\dfrac{2\times1+4\times3+6\times2+8\times3+10\times1}{5}$
$=\dfrac{60}{10}=6$(시간) ······ ❶
(분산)$=\dfrac{56}{10}=5.6$ ······ ❷
(표준편차)$=\sqrt{5.6}=2.36\cdots$(시간)
따라서 평균은 6시간, 표준편차는 2.4시간이다. ······ ❸

단계	채점 기준	배점
❶	평균을 구한다.	3점
❷	분산을 구한다.	4점
❸	표준편차를 구한다.	3점

15 평가 기준 막대그래프를 보고 평균과 표준편차를 설명할 수 있는가?

예시 답안 5명의 학생이 모두 똑같이 10점씩 점수가 올랐으므로 평균은 10점이 올랐다. ······ ❶
또한 5명이 똑같이 10점의 점수가 올랐기 때문에 5명의 점수의 편차는 중간고사 때와 같다.
따라서 표준편차는 변화가 없다. ······ ❷

단계	채점 기준	배점
❶	두 자료의 평균을 이해하고 비교한다.	7점
❷	두 자료의 표준편차를 이해하고 비교한다.	7점

단원 마무리하기

1 **평가의 주안점** 평균, 중앙값, 최빈값을 구할 수 있다.

풀이 $(\text{평균})=\dfrac{2.5+4.5+4.5+3.5+5+4}{6}=\dfrac{24}{6}=4(\text{시간})$

자료를 작은 것부터 크기순으로 나열하면

2.5, 3.5, 4, 4.5, 4.5, 5

이므로 $(\text{중앙값})=\dfrac{4+4.5}{2}=4.25(\text{시간})$

최빈값은 가장 많이 나온 값인 4.5시간이다.

2 **평가의 주안점** 도수분포표에서의 평균, 중앙값, 최빈값을 구할 수 있다.

풀이 계급값과 (계급값)×(도수)를 표로 나타내면

계급(년)	도수(명)	계급값	(계급값)×(도수)
0 이상 ~ 10 미만	8	5	40
10 ~ 20	8	15	120
20 ~ 30	4	25	100
30 ~ 40	3	35	105
40 ~ 50	3	45	135
50 ~ 60	1	55	55
합계	27		555

이때 평균을 반올림하여 소수 첫째 자리까지 구하면 20.6(년)이다.
중앙값은 중앙인 14번째 재위 기간에 속한 계급의 계급값이므로 15년이다.
최빈값은 계급의 도수가 가장 큰 계급의 계급값이다.
따라서 최빈값은 5년, 15년이다.
이 자료는 극단적인 자료가 있지 않고, 최빈값도 2가지가 나오므로 대푯값으로 평균이 적당하다.

3 **평가의 주안점** 실생활에서 평균과 중앙값을 구할 수 있다.

풀이 (1) (태양열의 평균)

$$=\frac{37.1+34.8+32.9+36.1+34.7+33.0+29.4+28.0+30.7}{9}$$

$$=\frac{296.7}{9}=32.96\cdots(\text{toe})$$

이므로 반올림하여 소수 첫째 자리까지 구하면 33.0toe이다.
이 자료를 작은 것부터 크기순으로 나열하면
28.0, 29.4, 30.7, 32.9, 33.0, 34.7, 34.8, 36.1, 37.1
이므로 가장 중앙에 위치한 33.0toe가 태양열의 중앙값이다.

179쪽

단원 마무리하기

1 다음은 현지네 반 모둠 6명의 일주일 동안의 운동 시간을 조사한 것이다. 이 자료의 평균, 중앙값, 최빈값을 구하여라.

(단위: 시간)

2.5 4.5 4.5 3.5 5 4

2 오른쪽 표는 조선 시대 왕의 재위 기간을 나타낸 도수분포표이다. 조선 시대 왕의 재위 기간의 평균, 중앙값, 최빈값을 구하여라. (단, 평균은 반올림하여 소수 첫째 자리까지 구한다.)

재위 기간(년)	도수(명)
0 이상 ~ 10 미만	8
10 ~ 20	8
20 ~ 30	4
30 ~ 40	3
40 ~ 50	3
50 ~ 60	1
합계	27

국사편찬위원회
http://www.history.go.kr

의사소통

3 태양 에너지는 무공해 에너지로 각 국가에서는 태양 에너지 보급을 늘리기 위하여 노력하고 있다.
다음은 2001년부터 2009년까지의 '태양열'과 '태양광'의 보급 현황이다. 물음에 답하여라. (단, 평균은 반올림하여 소수 첫째 자리까지 구한다.)

(단위: 천 toe)

연도(년)	2001	2002	2003	2004	2005	2006	2007	2008	2009
태양열	37.1	34.8	32.9	36.1	34.7	33.0	29.4	28.0	30.7
태양광	1.5	1.8	1.9	2.5	3.6	7.8	15.3	61.1	121.7

통계청 http://kostat.go.kr

'태양열'은 태양의 열 에너지를 변환하여 에너지원으로 사용하는 설비의 약칭이고, '태양광'은 태양의 빛에너지를 변환하여 전기를 생산하는 설비의 약칭으로 그 규모는 모두 'toe'라는 단위를 이용하여 표시한다.
'toe'는 원유 1톤의 발열량인 10^7 kcal를 기준으로 표준화한 단위이다.

(1) 태양열의 평균과 중앙값을 구하여라.
(2) 태양광의 평균과 중앙값을 구하여라.
(3) (1), (2)에서 태양열과 태양광 보급 현황의 대푯값으로 적당한 것을 각각 말하여라.

단원 마무리하기 179

(2) (태양광의 평균)

$$=\frac{1.5+1.8+1.9+2.5+3.6+7.8+15.3+61.1+121.7}{9}$$

$$=\frac{217.2}{9}=24.13\cdots(\text{toe})$$

이므로 반올림하여 소수 첫째 자리까지 구하면 24.1toe이다.
작은 것부터 크기순으로 나열하면
1.5, 1.8, 1.9, 2.5, 3.6, 7.8, 15.3, 61.1, 121.7
이므로 가장 중앙에 위치한 3.6toe가 태양광의 중앙값이다.

(3) 태양열은 연도별 변량이 고른 편이므로 평균이 대푯값으로 적당하다. 태양광은 극단적으로 보급량이 늘어난 변량이 있으므로 중앙값이 대푯값으로 적당하다.

4 **평가의 주안점** 분산과 표준편차의 의미를 이해할 수 있다.

풀이 (1) (기수의 평균)

$$=\frac{9+5+6+10+10+7+6+9+8+10+7+9}{12}$$

$$=\frac{96}{12}=8(\text{점})$$

IV

(명진의 평균)

$$=\frac{6+8+7+8+9+9+8+8+7+9+9+8}{12}=\frac{96}{12}=8(\text{점})$$

(2)

점수(점)	화살의 개수(개)	
	기수	명진
5	1	0
6	2	1
7	2	2
8	1	5
9	3	4
10	3	0
합계	12	12

(3) (기수의 분산)

$$=\frac{(5-8)^2\times1+(6-8)^2\times2+(7-8)^2\times2+(9-8)^2\times3+(10-8)^2\times3}{12}$$

$$=\frac{34}{12}=\frac{17}{6}$$

$$(\text{기수의 표준편차})=\sqrt{\frac{17}{6}}=1.683\cdots(\text{개})$$

$$(\text{명진이의 분산})=\frac{(6-8)^2\times1+(7-8)^2\times2+(9-8)^2\times4}{12}$$

$$=\frac{10}{12}=\frac{5}{6}$$

$$(\text{명진이의 표준편차})=\sqrt{\frac{5}{6}}=0.912\cdots(\text{개})$$

따라서 기수의 분산은 $\frac{17}{6}$, 표준편차는 1.68개이고, 명진이의 분산은 $\frac{5}{6}$, 표준편차는 0.91개이다.

(4) 명진이의 표준편차가 기수의 표준편차보다 작으므로 명진이의 점수가 더 고르다.

5 **평가의 주안점** 히스토그램으로 나타낸 자료의 분산과 표준편차를 구할 수 있다.

풀이 계급값, 편차, $(\text{편차})^2\times$ 도수를 표로 나타내면 다음과 같다.

줄넘기 횟수(회)	도수(명)	계급값	편차	$(\text{편차})^2\times(\text{도수})$
1 이상 ~ 3 미만	1	2	−4	16
3 ~ 5	2	4	−2	8
5 ~ 7	4	6	0	0
7 ~ 9	2	8	2	8
9 ~ 11	1	10	4	16
합계	10		0	48

(1) $(\text{평균})=\frac{2\times1+4\times2+6\times4+8\times2+10\times1}{10}=\frac{60}{10}=6(\text{회})$ …… ❶

(2) $(\text{분산})=\frac{48}{10}=4.8$ …… ❷

$(\text{표준편차})=\sqrt{4.8}=2.190\cdots(\text{회})$

따라서 표준편차는 2.19회이다. …… ❸

단계	채점 기준	배점
❶	평균을 구한다.	40%
❷	분산을 구한다.	40%
❸	표준편차를 구한다.	20%

180쪽

4 다음 표는 양궁 동아리 활동 시간에 기수와 명진이가 각각 3발씩 4엔드 12발의 화살을 쏘아 얻은 점수를 나타낸 것이다. 물음에 답하여라.

(단위: 점)

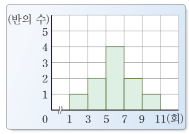
구분	1엔드			2엔드			3엔드			4엔드		
	1	2	3	4	5	6	7	8	9	10	11	12
기수	9	5	6	10	10	7	6	9	8	10	7	9
명진	6	8	7	8	9	9	8	8	7	9	9	8

(1) 기수와 명진이의 점수의 평균을 각각 구하여라.
(2) 위의 자료를 이용하여 오른쪽 도수분포표를 완성하여라.
(3) 두 학생의 점수의 분산과 표준편차를 각각 구하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)
(4) 두 학생의 점수 중 누구의 점수가 더 고른가?

점수(점)	화살의 개수(개)	
	기수	명진
5		
6		
7		
8		
9		
10		
합계		

서술형

5 오른쪽 그림은 희수네 학교 체육대회에서 3학년 10개 반의 단체 줄넘기 횟수를 조사하여 나타낸 히스토그램이다. 물음에 답하고, 그 과정을 서술하여라.
(1) 평균을 구하여라.
(2) 표준편차를 구하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)



문제 해결

6 오른쪽 표는 창수네 반 남학생들의 윗몸일으키기 횟수에 대한 도수분포표인데, 일부분이 찢어져 보이지 않는다. 윗몸일으키기 횟수의 평균이 20회일 때, 윗몸일으키기 횟수의 표준편차를 구하여라. (단, 표준편차는 반올림하여 소수 둘째 자리까지 구한다.)

윗몸일으키기 횟수(회)	학생 수(명)
0 이상 ~ 10 미만	2
10 ~ 20	9
20 ~ 30	6
30 ~ 40	
합계	

180 Ⅳ. 통계

6 **평가의 주안점** 표준편차를 활용한 문제를 구할 수 있다.

풀이 계급 30회 이상 40회 미만의 학생 수를 a명이라 하면

$$(\text{평균})=\frac{5\times2+15\times9+25\times6+35\times a}{17+a}=20\text{에서}$$

$10+135+150+35a=340+20a$

$a=3$

따라서 30회 이상 40회 미만의 학생 수는 3명이다.

주어진 도수분포표에서 계급값, 편차, $(\text{편차})^2\times$ 도수를 구하면 다음과 같다.

계급(회)	도수(명)	계급값	편차	$(\text{편차})^2\times(\text{도수})$
0 이상 ~ 10 미만	2	5	−15	450
10 ~ 20	9	15	−5	225
20 ~ 30	6	25	5	150
30 ~ 40	3	35	15	675
합계	20		0	1500

$(\text{분산})=\frac{1500}{20}=75$,

$(\text{표준편차})=\sqrt{75}=8.660\cdots(\text{회})$

따라서 표준편차는 8.66회이다.

181쪽

수행과제

내가 조사한 자료의 분산과 표준편차

통계 자료로 의미가 있는 한 가지 주제를 정하여 조사하고 다음과 같이 정리하여 보자.

(1) 주 　　제: ______________________

조사 대상: ______________________

(2) 조사한 자료를 아래 표에 나타내거나 또는 적절한 표를 만들어 나타내어 보아라.

구분	자료	구분	자료	구분	자료	구분	자료
1		7		13		19	
2		8		14		20	
3		9		15		21	
4		10		16		22	
5		11		17		23	
6		12		18		24	

◁ 작업 순서
- 주제를 정하기 위한 다양한 정보 수집하기
- 모둠별 토론 후 주제 정하기
- 조사 대상에 대한 정밀 조사하기
- 중간 점검하기
- 주제에 대한 자료 정리하기

◁ 유의 사항
조사한 자료가 특정한 자료인지 일반적인 자료인지 면밀히 검토한다.

(3) 조사한 자료를 다음과 같이 정리하고, 히스토그램과 도수분포다각형을 그려라.

계급	도수	계급값	(계급값)×(도수)	편차	$(\text{편차})^2\times(\text{도수})$
합계					

[히스토그램] 　[도수분포다각형]

(4) 위의 자료의 평균, 분산, 표준편차를 구하여라.

(5) 위의 자료들을 종합하여 내가 조사한 자료는 어떤 특징을 갖는지 토론하여 보자.

창의·인성

조사한 자료를 통계적 지식을 활용하여 해석해 보고, 이를 근거로 하여 자신의 의견을 조리 있게 말할 수 있다.

수행 과제 181

182쪽

수학으로 세·상·읽·기

통계의 함정

오늘날 우리는 인터넷, 방송, 신문이나 잡지 등을 통하여 수많은 정보를 접하고 있다. 정보를 제공할 때에는 이 정보들에 대한 신뢰를 심어 주기 위해 다양한 통계 자료를 이용하여 그 근거를 제시하기도 한다. 하지만 근거로 제시한 통계 자료에도 함정이 있을 수 있다.

다음은 대푯값의 한 가지인 평균을 사용했을 때 함정에 빠진 예를 나타낸 것이다.

어느 전쟁에서 장군이 병사들을 이끌고 이동하다가 큰 강을 만났다.

장군은 즉시 참모를 불러 강의 수심을 재어 올 것을 명령하였다. 참모는 배를 타고 여기 저기 수심을 재어 본 후 평균 수심을 구하고, 장군에게 평균 수심은 142cm라고 보고하였다. 이것을 들은 장군은

"병사들의 평균키가 167cm이므로 모두 쉽게 건널 수 있을 것이다. 모든 병사들은 즉시 강을 건너 이동하라."

라고 명령하였다.

하지만 강의 평균 수심은 평균적으로 142cm이었지만 강의 가운데는 더 깊은 곳도 많아 물에 빠지는 병사들이 많았다고 한다. 장군이 평균 수심에 대한 통계 자료를 잘못 이해하여 병사들이 큰 낭패를 보게 된 것임을 알 수 있다.

위와 같이 평균만으로는 자료의 전체적인 상황을 이해할 수 없으므로 자료의 분산, 표준편차, 최댓값 등이 필요하다. 또 이러한 통계 자료를 어떻게 구별하고 해석해야 하는지 스스로 판단할 수 있는 능력을 길러야 한다.

182 Ⅳ. 통계

수행과제

예시 답안

(1) 주제: 우리반 20명의 영어 성적

조사 대상: 우리반

(2) 영어 점수 조사표를 만들면 오른쪽과 같다.

(단위: 점)

번호	점수	번호	점수
1	75	11	90
2	65	12	95
3	60	13	90
4	55	14	85
5	85	15	80
6	90	16	80
7	50	17	80
8	75	18	65
9	70	19	70
10	95	20	75

(3) 자료의 정리 및 히스토그램과 도수분포다각형

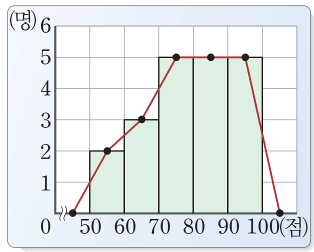
(단위: 점)

계급	도수	계급값	계급값×도수	편차	$(\text{편차})^2\times$도수
50 이상 ~ 60 미만	2	55	110	-24	1152
60 ~ 70	3	65	195	-14	588
70 ~ 80	5	75	375	-4	80
80 ~ 90	5	85	425	6	180
90 ~ 100	5	95	475	16	1280
합계	20		1580		3280

(4) $(\text{평균})=\dfrac{1580}{20}=79(\text{점})$

$(\text{분산})=\dfrac{3280}{20}=164$

$(\text{표준편차})=2\sqrt{41}\,(\text{점})$



(5) 다른 반의 영어 점수를 조사해 보고 표준편차를 구하여 우리반과 비교해 볼 수 있다. 또 전체 평균을 보고 우리반의 영어 성적을 가늠할 수 있다.

창의 · 인성 프로젝트 과제를 통하여 자기 주도적 실천 능력을 키우도록 한다.

수학으로 세·상·읽·기

통계 자료는 어떤 사람이 어떠한 목적으로 자료를 이용하느냐에 따라 잘못된 정보를 줄 수도 있고, 사용 목적에 따라 왜곡될 수도 있다. 이 이야기는 잘못된 자료에 의하여 실수를 저지르는 하나의 예를 보여 주고 있다. 상황에 맞는 통계 자료를 이용하기 위하여는 필요한 자료와 그것을 해석하는 능력을 길러야 할 것이다.