2. 확률과 그 기본 성질

2.1. 이론적 배경

(1) 확률론의 역사

봉건시대에는 성 안에서의 놀이로서 주사위, 트럼프가 성행하여 자연히 중세의 귀족 사이에도 도박이 대단히 성행하였다. 한편, 항구 도시는 대부분 상인들과 선원들로 왕래가 빈번해졌고, 이들은 남는 시간 동안에 술집 등에 둘러앉아 도박을 하곤 했다. 도박의 목적은 돈을 따는 것이므로 사람들은 어느 쪽이 이길 수 있는지 그 가능성에 관심을 갖게 된다. 도박은 사람들로 하여금 확률의 개념을 생각해내게 하였다.

확률론의 본격적인 연구가 진행된 것은 17세기 프랑스이다. 당시 유명한 도박꾼인 드 메레(Chevalier de Mere)는 도박을 하면서 수학적으로 그 승패를 생각하여 많은 이익을 얻었다. 그는 도박 중에 돈을 잃게 되는 경우나 상금 분배 문제를 파스칼(Pascal, B. 1623~1662)에게 물어보았고 파스칼은 당시의 유명한 수학자인 페르마(Fermat, P. 1601~1665)와 편지를 주고받으면서 그 문제를 논하였다.

베르누이(Bernoulli, J. 1654~1705)는 큰수의 법칙의 기본이 되는 베르누이의 정리를 발표하였고, 이는 초기 확률론에 큰 공헌을 하였다. 그는 우연에 의하여 지배되는 개개의 사건에 대하여 실험을 반복함으로써 하나의 법칙을 발견할 수 있다고 하였다. 또한 확률론의 이론적 전개에 필요한 많은 기본인 공식들(이항분포, 초기하분포, 음이항분포, 이항분포의 푸아송 근사, 이항분포에 관한 대수의 법칙 등)이 유도되었다.

드 무아브르(De Moivre, A. 1667~1754)는 도박에 관한 수학적인 원리를 체계적으로 밝히면서 확률론은 이후 급속한 발전을 이루었다. 그는 이항분포의 정규분포로의 근사, 순환급수의 이론 등을 연구하였다.

19세기 초에 이르러 프랑스의 수학자이자 천문학자인 라플라스(Laplace, P. S. 1749~1827)에 의해 확률은 하나의 학문적 체계로 조직화된다. 라플라스는 1812년 '확률의 해석적 이론'이라는 책을 펴냈는데 이때부터 확률론이 수학의 한 분과 역할을 했다.

일어날 수 있는 경우의 수가 유한일 때만 생각할 수 있던 확률이 연속변수인 경우까지 확장되면서 뉴턴(Newton, I. 1642~1727)과 라이프니츠(Leibnitz, G. W. 1646~1716)에 의한 미적분의 응용에 의하여 확률은 해석적으로 다루어지게 되었다.

20세기 들어와서 확률론은 그 면모를 달리했다. 칸토어(Cantor, G. ; 1845~1918)의 집합론과 힐베르트(Hilbert, D. ; 1862~1943)의 공리적 이론에 영향을 받아 콜모고로프(Kolmogorov, A. 1903~1987)는 확률론의 공리적 기초를 확립하기 위하여 '확률론의 기초 개념'을 발표하였다. 콜모고로프가 제안한 확률론의 견고한 공리적 구성은 확률론을 순수 수학적 이론으로 정비하였을 뿐만 아니라 확률개념의 적용 범위를 넓히고 응용에 대해서도 새로운 개척을 하였다. 이것으로 확률론은 수학이론으로 확고한 자리를 잡게 되었고, 자연과학이나 사회과학에도 많이 응용되고 있다.

(2) 확률의 정의

(가) 수학적 확률

같은 조건에서 여러 번 반복할 수 있는 어떤 실험에서 일어날 가능성이 있는 모든 결과를 원소로 하는 집합 S를 표본공간이라 하고 각각의 원소를 그 실험의 근원사건이라 한다. 이때 각각의 근원사건이 일어날 가능성이 같은 정도일 때 라플라스는 확률을 다음과 같이 정의하였다.

N개의 실험 결과로 구성된 표본공간 S에서 각각의 실험 결과가 일어날 가능성이 같을 때 m개의 실험 결과로 구성된 사건 A의 확률은 P(A)= $\frac{m}{N}$ 이다.

(나) 통계적 확률

실험 횟수 N을 한없이 크게 하면 어떤 사건 A의 상대도수가 어떤 일정한 이상적인 값에 가까워지는데 이것이 사건 A의 통계적 확률이다. 통계적 확률은 경험으로부터 얻을 수 있는 경우가 많으므로 '경험적 확률'이라고 하기도 한다.

(다) 공리적 확률

공리적 확률은 콜모고로프에 의해 20세기 초에 제안되었는데, 확률의 의미보다는 확률의 형식화, 즉 공리를 부여하는 것에 관심을 둔 것이었다.

확률변수를 표본공간의 측도 가능 부분 집합 전체로부터 (0, 1)로 가는 함수로 보고, 다음의 ①, ②, ③을 만족시킬 때 P(A)를 사건 A의 확률이라고 한다.

① 표본공간 S에서의 임의의 사건 A에 대하여 확률 P(A)는 $0 \le P(A) \le 1$이다.

② $P(S) = 1$ (단, S는 표본공간)

③ 서로 배반인 사건 A_1, A_2, $\dots$에 대하여

$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$이다.

확률의 공리적 정의는 다양한 확률의 정의를 모두 수용할 수 있다.

(라) 주관적 확률

여러 가지 객관적 근거에 의해 얻은 확률의 계산 결과가 항상 직관적인 느낌과 일치하는 것은 아니다. 간혹 우리는 확률의 계산 과정에 직접 사용되지는 않았지만 상황에 대한 예감, 경험 등을 반영하고자 할 때가 있다. 이때 전문가 등의 주관적인 판단에 근거한 확률을 객관적으로 계산된 확률에 혼합시키기를 원한다. 전문가 개인의 직관적 판단에 의해 부여되는 확률을 주관적인 확률이라 한다.

주관적 확률 역시 확률의 공리를 충족시켜야 하며 효용이론(utility theory) 등의 도움을 받기도 한다.

(3) 확률 이론

(가) 시행과 사상

동전이나 주사위를 던지는 것과 같이 동일한 상태에서 어떤 실험이나 관측을 반복하는 것을 시행(施行)이라 하고 시행에 의하여 나타날 수 있는 가능한 결과의 집합을 표본공간 (sample space)이라 하며 표본공간의 부분집합을 사상(事象; event)이라 한다. 또한 전혀 일어나지 않는 사상을 공사상(空事象)이라 공집합으로 나타낸다. 중학교에서는 사상을 사건으로 부른다.

두 사상 A, B에 대하여 A 또는 B가 일어날 때 이 사상을 A와 B의 합사상(Union)이라 하고 A∪B로 나타내고, A와 B가 동시에 일어날 때 이 사상을 A와 B의 곱사상(intersection)이라 하고 A∩B로 나타낸다. 두 사상 A, B에 대하여 어느 한 사상이 일어날 때 다른 한 사상은 결코 일어나지 않으면 이 사상을 서로 배반사상이라 한다. 또한 사상 A가 일어나지 않는 사상을 A의 여사상이라 하며 $\overline{A}$로 나타낸다.

(나) 독립사상과 종속사상

두 사상 A, B에 대하여 A와 B가 서로 영향을 주지 않는 경우, 두 사상 A와 B는 서로 독립이라고 한다. 그리고 서로 독립이 아닌 두 사상을 종속이라고 한다.

어떤 시행에서 일어나는 사상이 같은 정도로 기대될 때, 일어날 수 있는 모든 경우의 수를 n이라 할 때, 사상 E가 일어날 경우의 수를 k라 하면, 사건 E가 일어날 확률 P(E)는

$$P(E)=\frac{k}{n}$$

라고 정의한다. 이와 같이 정의된 확률(確率, probability)이 바로 수학적 확률이다.

수학적 확률에는 다음과 같은 기본 성질이 성립한다.

(i) $0 \le P(E) \le 1$

(ii) $P(E)=0 \rightleftarrows$ 사상 E가 결코 일어나지 않는다.

(iii) $P(E)=1 \rightleftarrows$ 사상 E가 반드시 일어난다.

(iv) $P(\overline{A})=1-P(A)$

(v) 사상 A가 사상 B의 부분집합 즉, A⊂B이면,

$P(A) \le P(B)$

(vi) A, B, C를 사상이라 하면,

$$P(A \cup B)=p(A)+P(B)-P(A \cap B)$$

$$\begin{aligned}P(A \cup B \cup C)=&p(A)+P(B)+P(C)\\&-P(A \cap B)-P(B \cap C)-P(C \cap A)\\&+P(A \cap B \cap C)\end{aligned}$$

(vii) 만일 $A_1, A_2, \cdots, A_n$이 서로 배반사상이면,

$$P(A_1 \cup A_2 \cup \cdots \cup A_n)=P(A_1)+P(A_2)+\cdots+P(A_n)$$

(다) 조건부 확률

확률이 0이 아닌 두 사상 A와 B에 대하여 사상 A가 일어났다는 가정 아래 사상 B가 일어날 확률을 A가

일어났을 때의 B의 조건부 확률이라 하며 P(B/A) 또는 P_A(B)로 나타낸다.

(ⅷ) P(B/A)=$\frac{P(A \cap B)}{P(A)}$

(ix) 만일 A_1, A_2, ⋯, A_n이 서로 독립이면, $P(A_1 \cap A_2 \cap \cdots \cap A_n)=P(A_1)P(A_2)\cdots P(A_n)$

수학적 확률은 표본공간에서 사상이 같은 정도로 일어난다는 가정 아래 정의된 것이다. 그러나 실제로 자연 현상이나 사회 현상 등에서는 각 시행의 결과가 같은 정도로 일어난다고 기대할 수 없는 경우가 많다. 일반적으로 같은 n번 시행하여 사상 E가 r번 일어났을 때, n을 충분히 크게 하면 상대도수 $\frac{r}{n}$은 일정한 값 p에 가까워진다. 즉, p=lim$\lim_{n \to \infty} \frac{r}{n}$이다. 이 p를 사상 E의 통계적 확률 또는 경험적 확률이라고 한다.

2.2. 교육과정 및 교과서 내용

2 확률과 그 기본 성질

① 경우의 수를 구할 수 있다.

> **새로운 지구촌의 축제**
>
> 1920년대 무렵 유럽과 남미에서 실력 있는 축구 선수들이 프로화를 선언하였다. 올림픽 무대는 아마추어만 출전할 수 있어 실력 있는 프로 선수들이 뛸 수 있는 세계 무대가 절실히 필요하게 되었다.
>
> 국제 축구 연맹(FIFA) 회장 줄리메가 올림픽과는 달리 프로 선수들도 참가할 수 있는 축구 대회 개최 계획을 발표하였다. 이에 1930년 우루과이에서 단일 종목으로는 가장 먼저 세계 선수권 대회인 월드컵이 처음으로 개최되었다.
>
> 생각해 봅시다
>
> ❶ 2010년 월드컵의 예선 리그는 한 조에 4팀씩으로 구성되었다. 한 팀이 치러야 할 예선 경기의 수는 몇 경기일까?
>
> ❷ 만일 한 조에 5팀씩으로 구성되었다면 한 팀이 치러야 할 예선 경기의 수는 몇 경기일까?

◦ 사건과 경우의 수의 뜻을 알게 한다.

주사위를 던지거나 동전을 던지는 경우와 같이, 동일한 상태에서 반복하여 시행할 수 있는 실험이나 관찰의 결과로 나타나는 현상을 사건이라 함을 알게 한다.

> 압정 1개, 주사위 1개, 동전 2개를 한 번씩 던지는 실험을 해 보자.
>
> 
>
> ❶ 각각의 실험에서 얻은 각자의 결과를 모둠별로 적어 보자.
>
> ❷ 위 ❶의 각 실험에서 얻은 결과와 다른 경우가 일어날 수 있는지 토론해 보자.

◦ 사건 A 또는 B가 일어날 수 있는 경우의 수를 구할 수 있게 한다.

두 사건 A, B가 동시에 일어나지 않을 때, A 또는 B가 일어날 수 있는 경우의 수를 구체적으로 세어 보게 하고, 그것이 각각의 사건이 일어날 수 있는 경우의 수의 합이 됨을 이해하게 한다.

> 소영이는 현장 체험 학습으로 역사 체험이나 자연 탐험 중 한 가지를 선택할 수 있다고 한다. 역사 체험은 경주 신라 문화, 부여 백제 문화, 서울 고궁 문화 중 한 곳을, 자연 탐험은 설악산 수목 탐사, 보령 갯벌, 창녕 우포늪, 순천만 갈대숲 중 한 곳을 선택할 수 있다고 한다.
>
> ❶ 다음 그림을 완성해 보자.
>
> 
>
>
> ❷ 역사 체험지를 선택하는 방법은 몇 가지인가?
>
> ❸ 자연 탐험지를 선택하는 방법은 몇 가지인가?
>
> ❹ 역사 체험지 또는 자연 탐험지 중 한 곳을 선택하는 방법은 몇 가지인가?

◦ 사건 A와 B가 동시에 일어날 수 있는 경우의 수를 구할 수 있게 한다.

두 사건이 서로에 영향을 끼치지 않을 때, 두 사건이 동시에 일어날 수 있는 경우의 수를 구체적으로 세어 보게 하고, 그것이 각각의 사건이 일어날 수 있는 경우의 수의 곱이 됨을 이해하게 한다. 이 경우 수형도를 적절하게 사용하여 이해를 돕도록 한다.

경우의 수를 구할 때, 합의 법칙, 곱의 법칙등과 같이 형식화하여 지도하지 않도록 하며 지나치게 복잡한 경우는 다루지 않는다.

자동차 공장에서 오른쪽 표와 같이 용량이 다른 3종류의 엔진과 2종류의 변속기 중 각각 1종류의 부품을 이용하여 자동차를 만들려고 한다.

엔진	변속기
1.8 L 2.0 L 2.4 L	자동 수동

❶ 엔진과 변속기를 선택하는 방법은 각각 몇 가지인가?

❷ 이 공장에서 만들 수 있는 서로 다른 자동차를 (엔진, 변속기)의 순서쌍으로 모두 나열해 보자.

② 확률의 의미와 그 기본 성질을 이해한다.

혈액형이 같은 사람은 얼마나 되나?

1667년 프랑스 국왕 루이 14세의 주치의였던 장 밥티스트 데니가 4명의 빈혈 환자에게 새끼 양의 혈액을 수혈했다는 기록이 있다. 이후 19세기가 되어 사람 사이의 수혈이 시작되었지만 혈액형에 대한 지식이 없었기 때문에 불행한 결과를 낳는 경우가 많았다.

오스트리아의 란트슈타이너는 1901년에 ABO식 혈액형을, 1940년에 Rh식 혈액형을 발견하여 안전한 수혈로 가는 길을 열었다. 현재 우리나라 사람의 혈액형의 비율은 대략 A형, B형, O형, AB형이 차례대로 34 %, 27 %, 28 %, 11 %이고 Rh 음성인 사람의 비율은 전체 인구의 0.1 %인 것으로 알려져 있다.

생각해 봅시다

❶ 한 사람의 한국인이 A형일 확률은 얼마일까?

❷ 한 사람의 한국인이 Rh 양성일 확률은 얼마일까?

◦ 확률의 뜻을 알게 한다.

어떤 사건이 일어날 가능성을 수로 나타낸 것이 확률임을 알게 한다. 확률 개념은 간단한 경우의 수 또는 상대도수와 관련된 소재를 통하여 도입한다.

오른쪽 그림과 같이 빨간 구슬 1개, 노란 구슬 2개, 파란 구슬 3개가 들어 있는 주머니로 다음과 같은 실험을 해 보자.

❶ 주머니에서 구슬 한 개를 꺼낼 때, 각 색 구슬이 나올 가능성을 추측하여 다음 표를 완성해 보자.

구슬의 색	빨간색	노란색	파란색	합계
가능성	$\frac{1}{6}$			

❷ 주머니에서 구슬 한 개를 꺼내어 색을 확인하고 다시 주머니에 넣는 활동을 각자 10번 반복하여 다음 표를 완성해 보자.

<나의 실험 결과>

구슬의 색	빨간색	노란색	파란색	합계
나온 횟수				10
상대도수				

<4명의 실험 결과의 합>

구슬의 색	빨간색	노란색	파란색	합계
나온 횟수				40
상대도수				

<8명의 실험 결과의 합>

구슬의 색	빨간색	노란색	파란색	합계
나온 횟수				80
상대도수				

❸ 위의 ❷에서 구한 각 색 구슬이 나온 횟수의 상대도수가 ❶에서 추측한 가능성과 같은지 확인해 보자.

▶ 확률

어떤 실험이나 관찰에서 각각의 경우가 일어날 가능성이 같다고 할 때, 일어날 수 있는 모든 경우의 수를 n, 어떤 사건 A가 일어날 경우의 수를 a라고 하면 사건 A가 일어날 확률 p는

$$p = \frac{(\text{사건 } A\text{가 일어날 경우의 수})}{(\text{모든 경우의 수})} = \frac{a}{n}$$

◦ 확률의 기본 성질을 이해하게 한다.

구체적인 예를 들어 다음과 같은 확률의 기본 성질을 이해하게 한다.

- 어떤 사건이 일어날 확률을 p라고 하면 $0 \le p \le 1$이다.
- 반드시 일어나는 사건의 확률은 1이다.
- 절대로 일어나지 않는 사건의 확률은 0이다.
- 사건 A가 일어날 확률이 p라고 하면, A가 일어나지 않을 확률은 $1-p$이다.

이 때, 여사건이란 용어는 사용하지 않는다.

다음의 사건들에 대하여 아래 물음에 답하여 보자.

A : 타율이 3할인 타자가 타석에서 안타를 치는 사건
B : 비가 내릴 가능성이 60 %인 날에 비가 내리는 사건
C : 오렌지 주스 10병이 들어 있는 상자에서 사과 주스를 꺼내는 사건
D : 사은품이 모두 들어 있는 과자를 살 때, 과자에서 사은품이 나오는 사건

❶ 절대로 일어날 수 없는 사건을 찾고 그 사건의 확률은 얼마라고 할 수 있는지 말해 보자.

❷ 반드시 일어나는 사건을 찾고 그 사건의 확률은 얼마라고 할 수 있는지 말해 보자.

❸ 사건 A, B, C, D의 확률 a, b, c, d를 각각 구하고 다음 수직선 위에 그 확률에 해당하는 점을 표시해 보자.

민선이는 감자 과자, 고구마 과자, 옥수수 과자, 양파 과자 중 두 가지를 사려고 한다.

❶ 선택할 수 있는 경우를 모두 나열해 보자.

❷ 감자 과자가 선택될 경우를 조사하고 그 확률을 말해 보자.

❸ 감자 과자가 선택되지 않을 경우를 조사하고 그 확률을 말해 보자.

❹ 감자 과자가 선택될 확률과 선택되지 않을 확률의 합은 얼마인가?

③ 확률의 계산을 할 수 있다.

◦ 사건 A 또는 B가 일어날 확률을 구할 수 있게 한다.

두 사건 A, B가 동시에 일어나지 않을 때, A 또는 B가 일어날 확률을 경우의 수를 이용하여 구할 수 있게 한다. 그리고 이 확률은 사건 A가 일어날 확률을 p, 사건 B가 일어날 확률을 q라고 할 때 $p+q$가 됨을 알게 한다.

혜원이가 가지고 있는 MP3 재생기에는 가곡이 7곡, 가요가 12곡, 외국곡이 11곡 들어 있다. 이 MP3 재생기에서 임의로 한 곡을 선택하여 들으려고 한다.

❶ 가곡이 나올 확률과 외국곡이 나올 확률을 각각 구해 보자.

❷ 가곡 또는 외국곡이 나오는 경우의 수를 구해 보자. 또, 그 확률을 구해 보자.

❸ 위의 ❶, ❷에서 구한 확률 사이에는 어떤 관계가 있는지 말해 보자.

◦ 사건 A와 B가 동시에 일어날 확률을 구할 수 있게 한다.

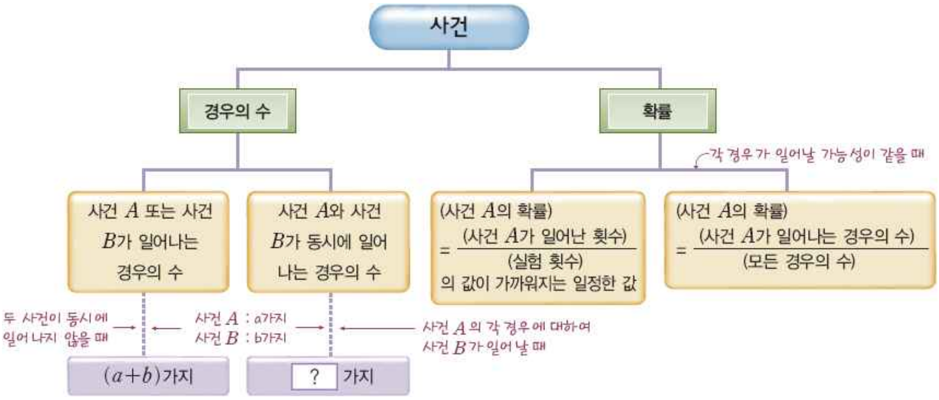
두 사건 A, B가 서로 영향을 끼치지 않을 때, A와 B가 동시에 일어날 확률을 경우의 수를 이용하여 구할 수 있게 한다. 그리고 이 확률은 사건 A가 일어날 확률을 p, 사건 B가 일어날 확률을 q라고 할 때 $p\times q$가 됨을 알게 한다.

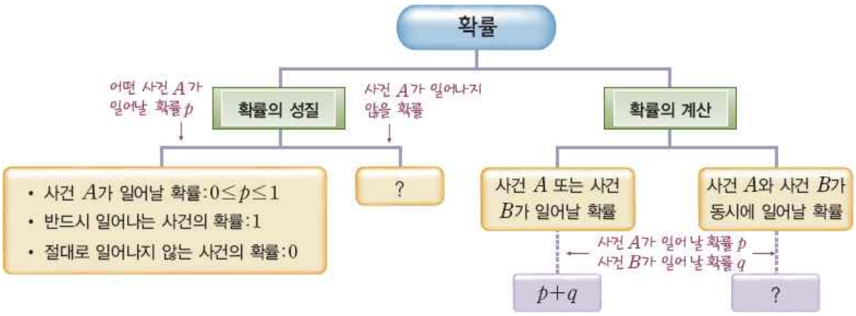
현규는 야외 체험 학습에 입을 수 있는 바지 3개와 셔츠 4개가 있다. 이 중에서 임의로 바지 한 개와 셔츠 한 개를 선택하려고 한다.

❶ 청바지가 선택될 확률과 흰 셔츠가 선택될 확률을 각각 구해 보자.

❷ 청바지와 흰 셔츠가 동시에 선택될 경우의 수를 구해 보자. 또, 그 확률을 구해 보자.

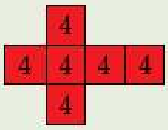
❸ 위의 ❶, ❷에서 구한 확률 사이에는 어떤 관계가 있는지 말해 보자.

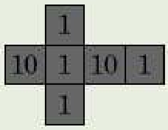


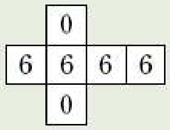


다음은 “왓슨, 내가 이겼네!”의 한 구절이다.

나(왓슨)는 발간색, 검은색, 흰색의 주사위를 손에 올려놓고 이리저리 돌려 보았다. 보통 주사위와는 다른 방식으로 주사위 눈이 그려져 있었다.







홈스가 내게 물었다.

“이 게임은 2명이 하는데 주사위를 하나씩 고른 다음 주사위를 동시에 굴려서 높은 숫자가 나오는 사람이 이기는 거야. 빨간색과 검은색 주사위를 고른 사람이 게임한다고 하세. 어느 쪽이 이기겠나?”

나는 머릿속으로 계산해 보았다. ‘발간색이 여섯 번 중 (①)번 이기는군.’

홈스가 또 물었다. “이번엔 검은색과 흰색으로 게임하면?”

이번엔 조금 까다로웠다.

“검은색 주사위의 눈이 10이 나오면 틀림없이 이기니까 $\frac{1}{3}$정도는 틀림없이 이기네. 하지만 검은색에는 1도 있으니까 $\frac{2}{3}$정도는 흰색이 이기는군.”

“그러면 검은색 주사위를 고른 사람이 이길 확률이 (②) 이라는 말이군.”

홈스가 말했다. “이번엔 빨간색과 흰색으로 게임해 보게.”

(출처 : 콜린 부르스, 이은희 옮김, “왓슨, 내가 이겼네!”, 경문사, 2006)

1. 위의 글에서 ①의 빈칸을 알맞게 채워라.

2. 위의 글에서 ②의 빈칸을 알맞게 채워라.

3. 위의 글의 마지막 문장에서 빨간색 주사위와 흰색 주사위로 게임할 때, 흰색 주사위를 고른 사람이 이길 확률을 구하여라.

2.3. 교수학습 참고자료

표본공간

통계 조사에서 가능한 모든 실험 결과의 집합을 표본공간이라고 한다. 예를 들어 가위바위보를 할 때 나오는 사건의 경우의 수를 생각하면 일어날 수 있는 모든 경우는 '가위', '바위', '보'이므로 표본공간은 {가위, 바위, 보}이다. 그러나 중학교 수준에서는 표본공간이라는 용어는 쓰지 않는다.

통계적 확률과 수학적 확률

1. 통계적 확률 : 일어나는 경우의 가능성이 서로 같지 않고 전체 경우의 수가 유한하지 않을 때, 일정한 조건 아래에서 n회의 시행을 반복하여 사건 A가 a회 일어났다고 하자. n을 충분히 크게 할 때, 상대도수 $\frac{a}{n}$가 일정한 값 p에 가까워지면 이 값 p를 사건A가 일어날 확률이라고 한다.

예) 흡연자가 비흡연자보다 폐암에 걸릴 확률이 높다.

2. 수학적 확률 : 사건이 일어나는 모든 경우의 수를 n이라고 하자. 그들의 어느 두 개도 중복해서 일어나지 않고 모든 경우가 일어날 가능성이 같을 때, 이 n가지 중에서 사건A가 일어나는 경우의 수가 a라면, 사건 A가 일어나는 확률 p는 $\frac{a}{n}$이다.

예) 로또에서 1등에 당첨될 확률은 $\frac{1}{45\times44\times43\times42\times41\times40}=\frac{1}{8145060}$이다.

역사 속의 수학 _ 공정한 분배

프랑스의 수학자인 파스칼(Pascal, B. 1623-1662)은 페르마(Fermat, P. 1601-1665)는 게임의 상금을 분배하는 문제를 연구하여, 확률의 수학적 이론을 창시한 사람으로 인정받고 있다. 이들이 연구한 분배 문제는 다음과 같다.

A와 B 두 사람이 같은 돈을 걸고 게임을 해서 먼저 5점을 얻는 사람이 돈을 모두 가지기로 하였다. A와 B 두 사람의 게임 실력은 비슷하다. A와 B가 4:3의 득점 상황에서 게임을 중단해야 한다면 돈을 어떻게 분배해야 하는가?

두 사람은 위의 문제를 다음과 같이 해결하였다고 한다.

이 게임을 계속 진행한다면 나올 수 있는 경우는 아래의 세 가지이다.

(1) A가 이길 경우: A의 5승 3패(A의 승리) 확률은 $\frac{1}{2}$

(2) B가 이기고 A가 이길 경우: B의 4승 4패 → A의 5승 4패 (A의 승리) 확률은 $\frac{1}{4}$

(3) B가 이기고 B가 이길 경우: B의 4승 4패 → B의 5승 4패 (B의 승리) 확률은 $\frac{1}{4}$

A가 이길 확률은 (1)과 (2)의 경우이므로 $\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$이고, B가 이길 확률은 (3)의 경우이므로 $\frac{1}{4}$이

다. 따라서 A에게 상금의 $\frac{3}{4}$을, B에게 상금의 $\frac{1}{4}$을 분배하면 공평하다.

음악 속의 수학 _ 미뉴에트

하이든, 베토벤과 함께 고전파의 3대 작곡가인 볼프강 아마데우스 모차르트(Wolfgang Amadeus Mozart)는 36세도 채 되지 않은 짧은 생애였으나 어려서부터 창작활동을 해 왔기 때문에 그가 남긴 작품은 성악·기악의 모든 영역에 걸쳐 매우 다양하다.

그의 뛰어난 작품들 중 <음악의 주사위 놀이>(Musikalisches Wurfelspiel, K.516f)라는 작품은 주사위를 던져서 곡을 만든 것이다. 이 곡은 미리 176개의 마디를 작곡해 놓고 주사위를 이용하여 조합하면 누구나 손쉽게 미뉴에트를 작곡할 수 있도록 만든 것이다. 이러한 방식으로 작곡할 수 있는 경우의 수는 129,629,238,163,050,258,624,287,932,416(12자, 1경의 약 12조 배)가지나 된다.

이처럼 어떤 창작에 우연을 도입하는 것을 '알레아토릭(Aleatorik)'이라고 부른다. ('알레아'는 라틴어로 '주사위'라는 뜻이다.)

이것을 직접 경험해 볼 수 있는 인터넷 사이트의 주소는 아래와 같다.

http://sunsite.univie.ac.at/Mozart/dice/collaborate.cgi?tables=yes

우리 반에 생일이 같은 사람이 있을 확률은?

나와 생일이 같은 사람을 만날 확률은 아주 드물다. 1년은 365일이고 우리 반 학생들이 25명이라고 생각하자.

내 짝의 생일이 나와 다를 확률은 $\frac{364}{365}$이고, 세 번째 학생의 생일이 앞의 두 사람과 다를 확률은 $\frac{364}{363}$이다. 이런 식으로 하면 25명의 생일이 모두 다를 확률은 $\frac{364}{365}\times\frac{363}{365}\times\frac{362}{365}\times\ldots\times\frac{341}{365}$이고, 이를 계산하면 약 0.43이다. 따라서 25명 중 나와 생일이 같은 학생이 있을 확률은 1−0.43=0.57이 된다.

몬티홀 딜레마

무대에 커튼으로 가려진 세 개의 문이 있는데, 이 중에 한 개의 문 뒤에는 비싼 상품(자동차나 밍크코트 등)이 숨어 있고, 나머지 두 개의 문 뒤에는 이상한 물건(애완견 먹이나 삐쩍 마른 염소 등)이 숨어 있다. 예를 들어, 출연자가 1번 문을 선택했을 때 사회자는 염소가 있는 2번 문을 열어 보이며 출연자에게 한 번의 기회를 더 준다. 그렇다면 출연자가 상품을 갖기 위해서는 사회자의 유혹대로 다른 문으로 옮기는 것이 유리할까? 아니면 처음에 선택했던 그 문을 고집하는 것이 유리할까?

이것이 몬티홀 딜레마 또는 몬티홀 문제이다.

이 문제를 해결하기 위해 편의상 자동차가 1번 문 뒤에 있다고 하자. 출연자가 어느 문을 선택하는가에 따라 다음과 같은 경우가 있다.

ⅰ) 출연자가 1번 문을 선택했을 때, 사회자는 2번(또는 3번) 문을 열어서 염소를 보여줄 것이다. 이

때, 출연자가 1번 문을 고수한다면 당첨될(자동차를 타게 될) 것이고, 3번(또는 2번) 문으로 옮긴다면 낙첨될 (염소를 보게 될) 것이다.

ii) 출연자가 2번 문을 선택했을 때, 사회자는 3번 문을 열어서 염소를 보여 줄 것이다. (1번 문을 열어줄 수는 없을 테니까.) 이때, 출연자가 2번 문을 고수한다면 낙첨될 것이고, 1번 문으로 옮긴다면 당첨될 것이다.

iii) 출연자가 3번 문을 선택한 경우는 ii와 마찬가지이다.

따라서, 원래 문을 고수할 경우 자동차를 탈 확률은 $\frac{1}{3}$이고, 다른 문으로 옮길 경우 자동차를 탈 확률은 $\frac{2}{3}$이다.

3. 대푯값과 산포도

3.1. 이론적 배경

(1) 대푯값

대푯값은 주어진 자료의 중심 위치를 나타내는 값으로서, 자료 전체를 하나의 수치로 요약한 값이다. 대푯값은 크게 두 가지로 나누어 볼 수 있는데 한 가지는 계산에 의해 결정되는 산술적 대푯값(산술평균, 기하평균, 조화평균)이 있고 자료가 크기에 의해 놓인 순서나 빈도에 의해 결정되는 위치적 대푯값(중앙값, 최빈값, 사분위수)이 있다(남궁필선, 2005).

이 단원에서 다루는 대푯값은 중앙값, 최빈값, 평균이다. 중앙값은 변량을 작은 값부터 크기순으로 나열할 때 중앙에 나타나는 값이다. N개의 변량을 크기순으로 나열하였을 때 N이 홀수이면 중앙값은 $\frac{N+1}{2}$번째 값이고, N이 짝수이면 중앙값은 $\frac{N}{2}$번째 값과 $\frac{N}{2}+1$번째 값의 평균이다. 키, 몸무게, 나이, 용돈, 강수량, 적설량, 독서량 등 양적 자료로 표현 가능한 모든 상황에서 중앙값이 사용될 수 있다. 특히 극단적으로 너무 크거나 작은 값이 포함된 자료의 경우에는 평균보다 중앙값이 더 적절하다.

최빈값은 자료에서 가장 많이 나타나는 값이다. 모든 자료가 같은 횟수씩 나타날 때 최빈값은 존재하지 않을 수 있으며, 가장 많이 나타나는 자료가 여러 개일 때 최빈값은 여러 개 존재할 수 있다. 최빈값은 자료의 수가 많고, 자료에 같은 값이 많은 경우 쉽게 구할 수 있다. 모든 자료가 같은 횟수씩 나타날 때 최빈값은 존재하지 않을 수 있으며, 가장 많이 나타나는 자료가 여러 개일 때 최빈값은 여러 개 존재할 수도 있다. 저금통을 깨서 동전을 분류하고 가장 많이 저금된 동전을 찾는다든지, 우리 반 아이들이 좋아하는 연예인을 가장 인기가 좋은 것으로 간주한다던지, 기성복 제작, 투표, 설문조사 등 최빈값이 사용되는 실생활 상황은 풍부하다. 최빈값은 평균이나 중앙값과는 달리 양적 자료뿐만 아니라 질적 자료에서도 사용 가능하다. 그러나 연속형 자료의 경우 최빈값은 평균이나 중앙값에 비해 자주 쓰이지 않는데, 그 이유는 다음과 같은 단점이 있기 때문이다. 첫째, 도수분포표를 사용할 때 계급의 크기에 따라 최빈값이 달라질 수 있다. 둘째, 이봉형 분포를 갖는 자료에서와 같은 경우 여러 개의 최빈값이 존재할 수 있는데, 이때 중심 위치로서의 의미가 줄어들게 된다(구자홍 외, 2003: 54).

평균은 변량의 합을 변량의 개수로 나눈 값으로, 변량 전체의 중심 위치를 나타낼 때 가장 많이 쓰이는 대푯값이다. 평균은 수학뿐만 아니라 사회 전반적인 모든 분야에서 널리 응용되고 있다. 학생들은 성적, 용돈, 각종 요금, 나이와 같은 양적 자료들을 다루거나 사회나 과학 등 타 과목의 학습에서도 평균을 사용할 수 있다. 평균에는 산술평균, 기하평균, 조화평균이 있으나 이 단원에서는 산술평균에 국한한다.

(가) 최빈값(mode)

도수분포표에서 최빈값은 도수가 가장 큰 계급에 대응하는 계급값을 말하며, 최빈값의 기호는 일반적으로 M_o로 나타낸다.

다음과 같은 도수분포표가 주어졌을 때 최빈값 M_o는 $f_i = Max(f_1, f_2, f_3, \ldots, f_n)$에 대응하는 계급값 x_i이다.

계급값	x_1	x_2	$\cdots$	x_i	$\cdots$	x_n	합계
f	f_1	f_2	$\cdots$	f_i	$\cdots$	f_n	N

(나) 중앙값(median)

도수분포표에서 중앙값은 전체 도수의 한 가운데 오는 변량이 속한 계급의 계급값이다. 일반적으로 중앙값은 기호 M_e로 나타낸다.

아래와 같이 도수분포표가 주어졌을 때 중앙값 M_e는 다음과 같이 유도할 수 있다.

계급	$a_1 \sim a_2$	$a_2 \sim a_3$	$\cdots$	$a_i \sim a_{i+1}$	$\cdots$	$a_n \sim a_{n+1}$	합계
도수	f_1	f_2	$\cdots$	f_i	$\cdots$	f_n	N

계급의 크기를 w라 하면, 오른쪽 히스토그램에서 넓이를 이등분하는 곳의 변량의 값이 M_e이다.

$f_1+f_2+\cdots+f_{i-1} < \frac{N}{2} < f_1+f_2+\cdots+f_{i-1}+f_i$라고 하면, 위의 그림에서

$$(f_1+f_2+\cdots+f_{i-1})w+zf_i = \frac{1}{2}(f_1+f_2+...+f_n)w$$

$$z=\frac{\frac{N}{2}-(f_1+f_2+...+f_{i-1})}{f_i}\times w$$

따라서 $M_e = a_i+\frac{\frac{N}{2}-(f_1+f_2+...+f_{i-1})}{f_i}\times w$이다.

확률분포에서 최빈값은 확률밀도함수 $f(x)$에서 $f(x)$가 최대가 되게 하는 x의 값을 구하는 것이고, 중앙값은 $f(x)$의 그래프 좌우의 넓이를 50%씩 나누는 x의 값을 찾는 것이다.

(2) 산포도

산포도는 변량들이 흩어져 있는 정도를 하나의 수로 나타낸 값이다. 자료가 대푯값 주위에 어느 정도 흩어져 있는 측도를 알아보는 것이 필요한 경우가 있다. 산포도에는 변량의 최댓값과 최솟값의 차인 범위, 분산과 표준편차, 그리고 사분편차와 평균편차가 있다. 산포도는 저울이나 자 등의 측정도구가 얼마나 일관성 있게 재어주고 있느냐 하는 정도나 한 개인의 판매량의 변화, 작업 능률의 변화 정도 등을 알아볼 때 이용되며, 추리통계에 있어서 표집에 따른 산포도를 측정함으로써 전집에 관한 전집치를 추정할 수 있게 된다. 이 단원에서는 평균을 중심으로 흩어진 정도를 나타내는 분산과 표준편차를 다룬다. 분산은 편차의 제곱의 평균이고 표준편차는 분산의 양의 제곱근이다. 두 값은 편차의 제곱을 포함하므로 변량의 극단적인 값에 영향을 많이 받으나 이론 전개에 편리하므로 가장 많이 사용되는 산포도이다.

(가) 분산과 표준편차

분산과 표준편차는 평균을 대푯값으로 하였을 때, 변량의 흩어진 정도를 나타낸다. 도수분포표가 주어졌을 때의 분산과 표준편차는 다음과 같이 계산한다.

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

(분산) $= \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})f_i$ (단, $\overline{x} = \frac{1}{N}\sum_{i=1}^{N}x_i f_i$)

(표준편차) $= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})f_i}$

(나) 그 외의 산포도

(평균편차)=(분산) $= \frac{1}{N}\sum_{i=1}^{N}|x_i - M_e|f_i$

범위(R): 변량의 최댓값과 최솟값의 차

사분편차(Q): 모든 자료를 크기순으로 배열할 때, $\frac{1}{4}$번째 변량의 값 Q_1과 $\frac{3}{4}$번째 변량의 값 Q_3과의 차의 반이다. 즉, $Q = \frac{Q_3 - Q_1}{2}$.

3.2. 교육과정 및 교과서 내용

③ 대푯값과 산포도

혼자 만든 운동 경기

캐나다인 네이스미스가 창안한 농구 경기는 1892년 1월 20일에 미국 매사추세츠 주에서 첫 공식 경기가 행해졌다. 초창기에는 인원 제한을 설정하지 않는 바람에 100명 가까이 경기하는 고등학교가 나오기도 하고, 바구니를 걸어 놓은 2층에 있던 관객이 골을 방해하는 일도 있었다. 우리나라에는 1907년 미국인 질레트가 처음 소개하였으며, 1925년에 대한농구협회가 조직되면서 인기 있는 종목으로 발전하였다. 이 후 1997년에 남자 프로 농구가, 1999년에 여자 프로 농구가 출범하였다.

선수	득점
김재성	30
강원식	19
박정진	15
손영훈	6
윤준상	4
최익준	4
오희진	2
김영신	2
조진규	1
허훈성	1
남대식	0
정영문	0
합계	84

생각해 봅시다

❶ 위의 표에서 선수들의 평균 득점은 얼마인가?
❷ 평균 득점보다 많이 넣은 선수는 몇 명인가?

① 중앙값, 최빈값, 평균의 의미를 이해하고, 이를 구할 수 있다.

◦대푯값으로서 중앙값, 최빈값, 평균의 의미를 알고 이를 구할 수 있게 한다.

대푯값은 자료 전체의 특징을 대표적으로 나타내는 값으로서, 초등학교에서 배운 평균 이외에도 각 변량을 크기순으로 나열할 때 중앙에 오는 값인 중앙값, 각 변량 중에 도수가 가장 높은 값인 최빈값이 있음을 알고, 그 필요성을 이해하며 구할 수 있게 한다. 특히, 자료의 개수가 짝수인 경우에 중앙값을 구하는 방법도 알게 한다. 예를 들어 변량을 크기순으로 나열하여 얻어진 20개의 자료 중에 10번째 값이 6, 11번째 값이 7인 경우에는 중앙값이 6.5임을 알게 한다.

다음은 새로운 품종의 벼를 시험 재배하기 위하여 10개의 시험 구획 중 6개의 구획에는 품종 A를, 나머지 4개의 구획에는 품종 B를 심어 그 수확량을 조사한 표이다.

(단위: kg)

품종	수확량						합계
A	37	36	27	26	23	25	174
B	36	35	28	29			128

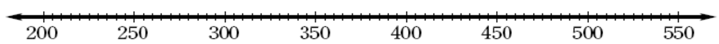
❶ 품종 A의 구획당 평균 수확량을 구해 보자.
❷ 품종 B의 구획당 평균 수확량을 구해 보자.
❸ 두 품종 중에서 어느 품종을 재배하는 것이 구획당 더 많은 수확량을 얻을 수 있는지 말해 보자.

다음은 어느 상점에서 판매하는 9가지 종류의 과자 50 g에 포함된 열량을 나타낸 것이다. 이때, 9가지 종류의 과자 50 g에 포함된 열량의 평균은 310 kcal이다.

(단위: kcal)

305	295	255	230	265	315	285	290	550

❶ 위 9개의 자료의 값과 그 평균을 아래 수직선 위에 나타내어 보자.



❷ 평균보다 작은 자료의 개수와 큰 자료의 개수를 각각 구해 보자. 또, 290 kcal보다 작은 자료의 개수와 큰 자료의 개수를 각각 구해 보자.
❸ 평균과 290 kcal 중 어떤 값이 위 자료 전체의 특징을 잘 나타낸다고 할 수 있는 지 말해 보자.

② 분산과 표준편차의 의미를 이해하고, 이를 구할 수 있다.

같지만 서로 다른 것

달걀을 사려고 어느 시장에 간 호진이는 두 양계장의 달걀 상자를 비교하였다. 각 상자에는 10개씩의 달걀이 들어 있었고 모두 600 g 이상이라고 적혀 있었다.

호진이는 저울을 가져다 각 달걀의 무게를 일일이 재어 다음과 같은 표를 만든 다음 한참 생각하다가 가산 양계장의 달걀을 샀다.

양계장	달걀 10개의 무게 (g)					합계
가산 양계장	57	62	59	58	61	605
	64	63	61	58	62	
나산 양계장	67	65	64	53	62	605
	57	54	65	57	61	

과연 호진이는 어떤 생각을 하였을까?

생각해 봅시다

❶ 달걀들의 무게의 최댓값과 최솟값의 차는 어느 상자가 더 클까?

❷ 어느 상자의 달걀들의 무게가 더 고르다고 할 수 있을까?

◦ 산포도로서 편차, 분산, 표준편차를 이해하고 계산할 수 있게 한다.

산포도는 자료의 흩어져 있는 정도를 측정하는 값으로서 편차, 분산, 표준편차가 있음을 알고, 그 필요성을 이해하며, 구할 수 있게 한다.

다음 표는 다이빙 시합에서 동수가 5회 다이빙하여 심판으로부터 얻은 점수를 기록한 것이다.

회	1	2	3	4	5	평균
점수 (점)	8.1	8.8	8.5	8.9	8.2	8.5

❶ 동수의 각 점수에서 평균을 뺀 값을 구하여 오른쪽 표를 완성해 보자.

❷ (점수)−(평균)의 값이 양수일 때와 음수일 때의 의미를 각각 말해 보자.

❸ (점수)−(평균)의 값의 합이 0이 됨을 확인하고, 그 이유를 말해 보자.

점수 (점)	(점수)−(평균)
8.1	−0.4
8.8	
8.5	
8.9	
8.2	
합계	

n개의 변량 $x_1, x_2, x_3, \cdots, x_n$에 대하여 이들의 평균이 m일 때

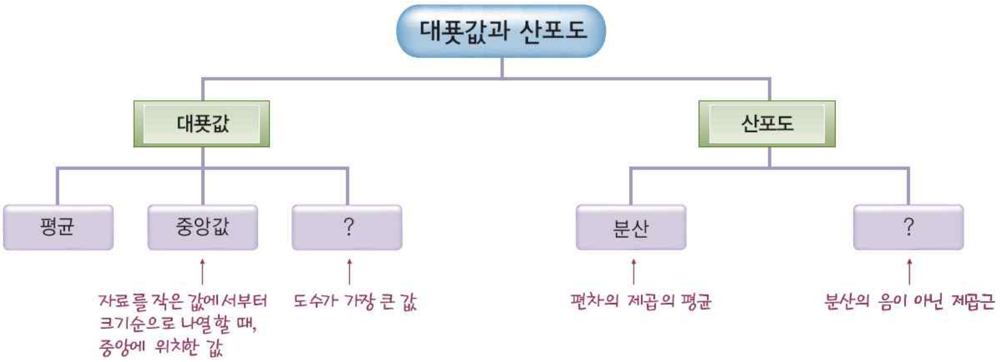
❶ (편차)=(변량)−(평균)

$=x_1 - m \ (i=1, 2, 3, \cdots, n)$

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

$$=\frac{1}{n}\{(x_1-m)^2+(x_2-m)^2+(x_3-m)^2+\cdots+(x_n-m)^2\}$$

❸ $(\text{표준편차})=\sqrt{(\text{분산})}$


대푯값과 산포도
대푯값
산포도
평균
중앙값
?
분산
?
자료를 작은 값에서부터 크기순으로 나열할 때, 중앙에 위치한 값
도수가 가장 큰 값
편차의 제곱의 평균
분산의 음이 아닌 제곱근

다음 글을 읽고 물음에 답하여라.

평균이 항상 자료의 중심 경향을 잘 나타내어 준다고 할 수 있을까? 그렇지는 않다. 평균은 상황에 따라 사람들에게 착각이나 심각한 오해를 불러 일으킬 수 있다. 다음의 각 상황을 살펴보자.

1. 학생 열 명이 모두 합하여 10달러를 가졌다면 각 학생은 평균 1달러를 가진 셈이다. 그러나 ㉠학생 한 명이 10달러를 가지고 있고 다른 학생 아홉 명은 돈을 하나도 가지고 있지 않을 수도 있다. 그래도 학생 한 명당 가지고 있는 돈의 평균은 1달러이다.

2. ㉡어떤 집단에서 한 사람의 소득이 연간 6백만 달러에서 6억 달러로 늘어나고, 1억 명이 연간 2만 달러에서 5달러씩 수입이 감소하여 총 5억 달러가 감소하였다 하더라도 전체 집단의 평균 소득은 올라간다. 그러나 이 경우에 실제로 사람들의 소득이 늘어났다고 말할 수는 없다.

평균이 그 집단의 중심 경향을 제대로 나타낼 수 없는 경우에는 다른 대푯값인 중앙값 또는 최빈값을 생각하여야 한다. 그러나 어떤 대푯값을 구해야 하는지를 일러 주는 수학 법칙은 없다. 가장 적합한 대푯값에 대한 판단은 자료의 특성과 관련 지식을 바탕으로 스스로 내려야 한다.

(참고 자료: 스티븐 제이 굴드, 이명희 옮김,“ 풀하우스”, 사이언스북스, 2002)

1. 위의 밑줄 친 문장 ㉠이 사실이라고 할 때, 그 자료의 중앙값과 최빈값을 구하여라.

2. 위의 밑줄 친 문장 ㉡이 사실이라고 할 때, 그 자료의 중심 경향을 잘 나타내어 주는 대푯값은 무엇인지 말하고, 그 이유를 설명하여라.

3. 위의 밑줄 친 문장 ㉠을 ‘학생 열 명이 각각 1달러씩 가지고 있다.’로 고쳤을 때, 고치기 전과 고친 후의 자료의 분산을 각각 구하고, 그 값을 비교하여라.

컴퓨터로 푸는 수학!!

통계 소프트웨어를 이용하여 대푯값(평균, 중앙값, 최빈값)과 산포도(분산, 표준편차)를 구하여 보자.

통계 소프트웨어를 이용하여 다음 자료의 대푯값과 산포도를 구하여 보자.

34, 32, 28, 30, 31, 36, 34, 30, 28, 32, 33, 31, 29, 38, 32

통계 소프트웨어의 초기 화면에서 다음 순서에 따라 실행한다.

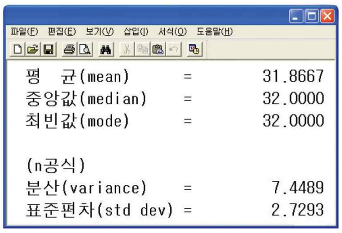
❶ 위의 자료를 오른쪽 그림과 같이 세로로 입력한다.

❷ [메뉴] → [통계분석] → [기초 분석] → [기초 통계량]을 차례로 선택한다.

❸ 기초 통계량 도구 상자에서 변수 칸의 1을 선택하고 ▶▶을 눌러 분석변수 칸으으로 보낸다.

[17,1]		V1	V2
1		34	
2		32	
3		28	
4		30	
5		31	
6		36	
7		34	
8		30	
9		28	
10		32	
11		33	
12		31	
13		29	
14		38	
15		32	

❹ 자료분석 단추를 누르면 다음과 같이 평균, 중앙값, 최빈값과 분산, 표준편차가 나 타난다.



조선 시대 왕 수명 자료를 인터넷으로 검색하여 그 자료의 대푯값과 산포도를 통계 소프트웨어를 이용하여 구하여라.

통계 소프트웨어는 여러 사이트에서 무료로 내려받을 수 있다.
(http://www.s-link.com → [다운로드] → [S-Link 평가판]

3.3. 교수학습 참고자료

통계의 교수 학습 방법

통계는 자료를 모으고 그 자료를 잘 정리하여 통계적 처리를 하여 그 결과를 분석하고 해석하는 학문이다. 따라서 많은 사실에서 공통되는 일반적인 법칙이나 경향을 찾아내는 귀납적 방법을 학습하기에 좋다. 그러므로 통계를 학습할 때는 귀납적 사고를 충분히 경험할 수 있도록 하고, 연역적 사고와 방법을 강조하지 않는다.

교과서에는 자료가 제시되어 있지만, 학생들 스스로 자료를 수집하고 수집한 자료가 적절한지 판단하면서 그 자료를 활용하여 대푯값과 산포도를 학습하게 할 수 있다. 또 프로젝트 과제로 제시된 과제를 단원 도입부터 시작하게 할 수도 있다.

학생들이 대푯값과 산포도의 필요성을 인식할 수 있도록 하면서 도입한다. 대푯값에는 평균, 중앙값, 최빈값이 있으며 각각이 어떤 상황에서 적절한지 판단하게 한다.

대푯값은 자료의 위치를 나타내는 것으로 자료를 하나의 값으로 요약할 수 있지만, 자료의 분포 상태를 알려주지는 않는다. 산포도의 개념을 도입하여 분산과 표준편차의 의미를 알고 상황에 따라 해석하게 한다.

또 다른 평균

인구가 만 명인 어느 도시의 인구가 1년 후 2만 명으로 2배, 2년 후 16만 명으로 1년 만에 8배 증가하였다. 이 때 인구의 매년 평균 증가율을 구해보자. 평균을 $\frac{2+8}{2}=5$으로 계산하면, 2년 후 1만$\times 5\times 5=25$만명이 되어야 하지만, 실제로 2년 후에는 16만 명이다. 이런 경우에는 우리가 알고 있는 평균(즉, 산술평균)으로 답을 구하면 안 된다. 대신에 두 수를 곱한 후 제곱근을 취하는 '기하평균'을 사용하여, 인구가 1년에 평균 $\sqrt{2\times 8}=4$배가 늘어났다고 본다. 그러면, 도시의 인구는 2년 후에 1만$\times 4\times 4=16$만 명이 된다.

또, A 지점에서 100m 떨어진 B 지점까지 직선 길을 따라 갈 때는 10m/s의 속력으로, 올 때는 20m/s의 속력으로 왕복했을 때 평균 속력을 구해보자. A에서 B까지 갈 때 10초, 올 때 5초가 걸리므로 전체 걸린 시간은 15초이며, 전체 이동한 거리는 200m이다. 따라서 평균 속력은 $\frac{40}{3}$m/s이며, 산술평균과 다른 값이 나온다. 이것은 '조화평균'이라 불린다.

평균이 적절하지 않은 사례

어떤 회사는 배불러씨와 그의 동생, 그리고 다섯 명의 친척들로 임원을 맡고 있다. 사원으로는 다섯 명의 작업반장과 열 명의 직공이 있다. 그런데 일손이 부족하여 한 사람의 직공이 더 필요하게 되었다.

고생해: 저는 이 회사에서 일하고 싶습니다. 이곳의 월급이 꽤 높다고 들었거든요.
배불러: 우리 회사의 보수는 아주 높은 편이지요. 1인당 평균 월급이 100만원이나 되니까요.

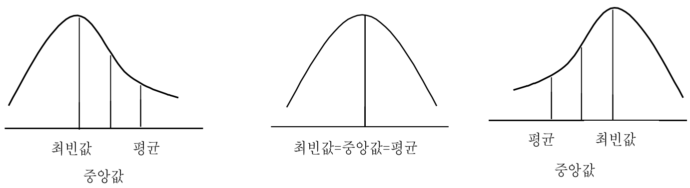
수습기간을 보낸 후 봉급을 몇 번 받은 후 고생해는 사장을 찾아갔다.

고생해: 당신은 나를 속였어요. 다른 직공들에게 물어 보았는데, 월급이 50만원이 넘는 사람이 하나도 없었어요. 그런데 어떻게 평균 월급이 100만원이라는 거지요?

배불러: 내 말을 잘 들어보시게. 나는 매달 (　　　　)만원을 받네. 그리고 내 동생은 (　　　　)만원, 5명의 내 친척들은 각각 (　　)만원씩, 또 5명의 작업반장들은 (　　　　)만원씩, 10명의 직공들은 각각 (　　　)만원이니, 평균은 100만원인 거지.

자료의 분포에 따른 대푯값의 대소 관계

자료의 중심 경향을 나타내는 대푯값에는 평균, 중앙값, 최빈값이 있다. 어떤 것을 자료의 대푯값으로 선택할지는 자료의 분포에 따라 달라질 수 있다. 일반적으로 자료의 분포곡선 모양에 따른 평균, 중앙값, 최빈값의 위치는 다음과 같다. 대칭 구조를 갖는 경우는 세 개의 대푯값이 모두 일치하지만, 자료가 비대칭적인 모양일 때는 차이가 있다.



대학수학능력시험의 표준점수 구하는 방법

시험 결과는 원점수 또는 표준점수로 제시될 수 있다. 원점수는 맞힌 문항에 부여된 배점을 합한 실제 점수이고, 표준점수는 원점수가 전체 응시자 중에서 어느 정도의 위치에 있는지 보여주는 상대적인 점수이다. 표준점수를 쓰는 이유는 시험의 난이도와 응시자 집단 간의 점수 차이를 조정하기 위한 것이다.

대학수학능력시험의 언어, 수리, 외국어 영역은 목표 평균과 목표 표준편차를 각각 100점과 20점으로 쓰고 있다. 이 영역에서 학생 개인의 원점수와 응시자 전체의 평균 점수와 표준편차를 알고 있다면, 표준 점수는 다음과 같이 구할 수 있다.

$$(\text{수능시험의 표준 점수})=(\text{목표 평균})+(\text{목표 표준편차})\times\frac{(\text{원점수})-(\text{영역별 응시자 전체의 평균 점수})}{\text{영역별 응시자 전체 점수의 표준편차}}$$

강물에 빠진 병사들

1920년대 중국의 내전 중에 벌어진 일이다. 병사들을 이끌고 적진을 향해 진격하고자 했던 한 장수가 눈앞에 큰 강을 만나게 되었다. 장수는 부하에게 강의 평균 수심이 얼마냐고 물었다. 부하는 평균 수심이 1.4미터라고 답했다. 장수는 평균 수심이 1.4미터이고 병사의 평균 키가 1.65미터이므로 걸어서 행군이 가능하다고 판단하고 진격을 명했다. 그러나 강의 가운데 수심은 병사의 키보다 훨씬 깊어서 모두 물에 빠져버렸

다.

병사들의 평균 키가 1.65미터라면 1.5미터 정도의 키 작은 병사는 1.4미터의 강을 건너는 것이 매우 힘들었을 것이다. 그러나 더 큰 문제는 강의 수심 1.4미터가 평균이라는 점이다. 평균이 1.4미터라면 어느 부분은 이보다 얕지만 어느 부분은 깊어서 평균 수심 이상이 된다. 그렇다면 수영을 못하는 병사는 강을 건널 수가 없다.

[출처: 통계의 미학 (최제호 지음, 동아시아)]

4. 선생님, 질문 있어요. - 확률과 통계

1. 초등학교에서 배운 '막대그래프'와 중학교에서 배우는 '히스토그램'에는 어떤 차이가 있는가?

2. '도수분포표에서의 평균'과 실제 자료의 '평균'은 항상 일치하는가? 왜 그런가?

3. (상대)도수분포표에서 상대도수는 보통 '소수'로 표현하는데, 이는 분수로 나타내는 것보다 소수로 나타내는 것이 자료를 비교하기에 편리하기 때문이다. 이때, 상대도수분포표에서 상대도수의 합이 1이 되지 않는 경우가 있을 수 있다. 왜 그런가?

4. '배반사건'과 '독립사건'의 차이점은?

5. $1^2+2^2+\cdots+n^2=\dfrac{n(n+1)(2n+1)}{n}$ 임을 조합을 이용하여 설명(증명)해 보아라.

6. ${}_nH_r={}_{n+r-1}C_r$ 임을 설명해 보아라.

7. $Y=2X+1$ 일 때, $P(Y)=?$