

통계적 소양을 위한 교수 학습자료

2016. 12.



통계청
통계교육원

제 1 장 통계적 소양을 위한 교수 학습자료 개발 방향	1
제 1 절 개발 방향	3
제 2 절 교수 학습자료 내용	6
제 2 장 자료 수집	7
제 1 절 개발 의도	9
제 2 절 지도를 위한 배경	14
제 3 절 교수 학습자료 해설	17
제 3 장 그래프	29
제 1 절 개발 의도	31
제 2 절 지도를 위한 배경	32
제 3 절 교수 학습자료 해설	36
제 4 장 대푯값	55
제 1 절 개발 의도	57
제 2 절 지도를 위한 배경	59
제 3 절 교수 학습자료 해설	60
부록	75
부록1. 자료 수집 학생 활동지	77
부록2. 그래프 학생 활동지	87
부록3. 대푯값 학생 활동지	101

제 1 장 통계적 소양을 위한 교수 학습자료 개발 방향

제 1 장 통계적 소양을 위한 교수 학습자료 개발 방향

제 1 절 개발 방향

- 일상생활에서 접하는 실제 통계 정보를 해석, 이해 및 비판적으로 평가하고 의사결정 및 의사소통에 활용하는 **통계적 소양 교육** 강조

현대 정보 사회에서는 일상생활과 정치, 경제, 과학 등 각 분야에서 일어나는 각종 불확실한 현상에 대한 합리적인 판단 도구로서 통계적 방법이 광범위하게 사용되고 있다. 어떻게 정보가 처리되며 유용한 지식으로 전이되는 가에 대한 이해가 필수적이다.

통계 자료 수집, 통계적 요약, 통계적 추론은 다른 형태의 수학적 분석보다도 일상생활에서 빈번하게 나타난다. 따라서 올바른 예측과 의사결정을 위해 자료 분석에 사용되는 개념과 그 과정을 이해하는 등의 통계적 소양이 현대 사회생활을 영위하는데 필수적으로 요구되고 있다.

통계 교육은 통계치의 계산보다 실제적인 자료를 수집하여 이를 표현하고 해석하는 경험을 통해 통계의 기본적인 원리를 이해하도록 함으로써 자료에 기반한 논리적인 추론 및 비판 능력을 개발해야 한다. 특히 최근 통계 교육에서 통계적 소양의 중요성이 주목을 받고 논의되고 있다.

통계적 소양에 대한 정의는 다양하지만, 학생들이 학교를 졸업한 후에 일상생활, 직장 생활에서 통계의 역할을 이해하고 자료의 소비자, 생산자, 의사소통자로서 능력을 갖추어야 한다.

학생들은 정보와 자료를 처리하는 능력에 관한 확률과 통계의 기본적인 지식을 습득하고 다양한 맥락에서 접하는 통계 정보나 자료와 관련된 주장 또는 통계적 현상들을 해석하고 비판적으로 평가할 수 있어야 한다. 또한 통계 정보에 대해 토론하고 의사소통하는 능력을 갖추어야 한다.

통계적 소양을 위한 통계교육의 목적을 다음과 같이 제시할 수 있다.

1. 문제설정, 자료수집, 자료분석, 결과해석이라는 4단계의 통계적 문제해결 과정을 경험하도록 한다.
2. 통계적 문제해결 과정의 각 단계와 관련된 핵심 이슈를 이해하도록 한다.
3. 자료 수집, 표현, 분석, 해석하는 과정을 통해 올바른 예측과 합리적인 의사결정

과정에서 통계적 도구를 활용함으로써 그 역할과 가치를 이해하게 한다.

4. 제시된 정보 및 자료의 근원, 요약된 정보를 비판적으로 평가하고 의사소통 할 수 있다.

- 2015 개정 수학과 교육과정에서도 강조하고 있는 통계 정보를 해석하고 비판적으로 평가하는 능력 개발

통계적 소양 중 다양한 맥락에서 접하는 통계 정보나 자료와 관련된 주장 또는 확률 통계적 현상들을 해석하고 비판적으로 평가하는 능력과 통계 정보에 기초하여 토론하고 의사소통하는 능력을 개발하고자 한다. 적절한 정당화의 과정이 없는 주장에 대한 비판적인 사고 능력이 필요하다.

정보화 시대의 시민에게 제공되는 정보와 자료들은 일반적으로 여러 가지 목적이 있기 때문에 제시되는 자료에 대해 해석하고, 비판적으로 평가할 수 있는 어느 정도 수준의 통계적 소양이 요구된다.

대중매체는 통계적 정보의 사슬에서 중간자 역할을 하기 때문에 통계 자료를 활용하는 사람들은 그들에게 제공된 정보가 얼마나 잘 만들어 졌는지 그리고 대중매체에 의해 얼마나 잘 요약되고 해석되었는지 검토해야만 한다.

제시되는 통계적 주장의 합리성을 조사하고 그것에 대해 의심해야만 한다. 자료의 타당성과 제시된 정보나 결론이 기초하는 증거의 본성이나 신뢰성에 대하여 관심을 가져야 하며, 이로부터 나오는 결론에 대한 대안적인 해석을 주의 깊게 고려해야 한다.

통계적 정보 사슬의 중요한 구성원으로서 학생들은 통계적 탐구로부터 해석한 결과에 포함되는 것을 배우고, 대중매체나 교실 동료들로부터의 보고서 등에서 자료와 요약된 통계를 참조하는 논쟁에 관하여 비판적이고 반성적인 질문을 해야 할 필요가 있다.

- 수업에서 실제로 사용 가능한 활동지 개발

- 2015 개정 수학과 교육과정에서 추구하는 교과 핵심 역량(정보처리, 의사소통, 문제해결, 추론, 태도 및 실천) 함양에 기여함과 동시에 확률과 통계 영역의 교수-학습 상의 유의점을 반영

<2015 개정 교육과정 중 확률과 통계 영역에서 교수·학습 자료 개발과 관련된 교수-학습 상의 유의점>

<초등학교>

- ◆ 꺾은선그래프를 그릴 때 변화의 경향이 잘 드러날 수 있도록 눈금의 크기를 적절히 선택하게 한다.
- ◆ 간단한 그림그래프, 막대그래프, 꺾은선그래프의 특성을 비교하여 자료의 특성에 맞는 그래프로 나타내게 한다.
- ◆ 막대그래프, 꺾은선그래프, 그림그래프, 띠그래프, 원그래프의 특성을 비교하여 목적에 맞는 그래프로 나타내게 한다.
- ◆ 자료와 가능성 영역의 문제 상황에서 문제 해결 전략 비교하기, 주어진 문제에서 필요 없는 정보나 부족한 정보 찾기, 조건을 바꾸어 새로운 문제 만들기, 문제 해결 과정의 타당성 검토하기 등을 통하여 문제 해결 능력을 기르게 한다.
- ◆ 평균을 구하는 것뿐만 아니라 평균이 사용된 상황에서 그 의미를 파악하는지 평가한다.

<중학교>

- ◆ 다양한 상황에서 자료를 수집하고, 수집한 자료가 적절한지 판단하게 한 후, 자신의 판단 근거를 설명해 보게 한다.
- ◆ 눈금 등을 부적절하게 사용하여 자료를 부정확하게 나타낸 표나 그래프에서 오류를 찾는 활동을 하게 한다.
- ◆ 자료의 특성에 따라 적절한 대푯값을 선택하여 구해보고, 각 대푯값이 어떤 상황에서 유용하게 사용될 수 있는지 토론하게 한다.

<고등학교의 확률과 통계>

- ◆ 실제적인 예를 통하여 표본조사의 필요성을 알게 하고, 올바른 표본추출이 모집단의 성질을 예측하는 기본조건임을 이해하게 한다.
- ◆ 표본평균은 추출한 표본에 따라 다른 값을 가질 수 있는 확률변수임을 알게 한다.

제 2 절 교수 학습자료 내용

- 교육과정에 제시되었으나 교과서에 구현하기에 제한적인 내용을 선정하고, 통계 정보를 비판적으로 평가하고 판단할 수 있는 기회를 제공하여 통계적 소양의 개발에 기여
- 3가지 주제 선정
 - 자료수집: 통계 교육의 가장 기본이 되는 내용이면서 현재 교육과정에서 누락되어 있는 대표적인 내용요소로서 고등학교에서 다룸. 특히 통계 포스터 활동과 관련된 과정에서 활동 초기에 필요한 기본 교육 내용으로 필요함
 - 그래프: 그래프로 제시되는 통계적 정보를 실제로 해석하고 비판적으로 검토하는 과정을 경험함. 또한 자료를 요약, 표현하는 여러 가지 방법(표와 다양한 그래프)을 적절하게 표현하고 사용하고 해석하는 방법
 - 대푯값 : 현재 중학교 3학년에서 다루는 내용으로 다양한 대푯값의 필요성 인식과 현재 중학교 1학년 교육과정에서는 삭제된 도수분포표에서의 평균과 관련된 교수-학습
- 초등학교 5~6학년에서 고등학교 단계까지 선택적으로 수업에서 실제로 활용 가능

제 2 장 자료 수집

제 2 장 자료 수집

제 1 절 개발 의도

가. 활동지1: 설문조사하기

활동지1의 첫 번째 활동 ‘자료는 어떻게 수집해야 할까?’는 통계적 문제해결 과정 중 ‘자료 수집’ 단계에 주목할 수 있도록 개발되었다. 자료 수집 단계에서 목표모집단과 조사모집단의 불일치로 인해 편의표본(biased sample)이 추출될 경우, 이를 기반으로 도출된 통계적 결론은 실제와 매우 동떨어진 것이 되어버린다. 이러한 통계 왜곡 현상을 일으키는 요소에 주목하고, 통계적으로 올바른 자료 수집 방법에 대한 이해를 돕기 위한 의도로 활동지의 각 문항들이 개발되었다. 특히, 통계 소비자(consumer)로서의 비판적 자세와 통계적 지식, 그리고 통계 생산자(producer)로서의 역량을 아우르는, 자료 수집과 관련된 통계적 소양을 지도하는 것이 본 활동지의 목적이다.

첫 번째 문항은 동일한 문제의식으로 시작한 두 설문조사의 상이한 결과를 제시하는 것으로 시작한다. 이는 우연 맥락에서 동일한 자료 수집에도 다른 결과를 얻게 됨으로써 불확실성을 인정하는 통계의 특징을, 수학과 대비하여 인식할 수 있는 기회를 제공할 것으로 기대한다. 그리고 단순히 우연으로 설명하기에는 조사 결과의 차이가 많이 큰 경우를 보여줌으로써, 본 활동지의 주제가 ‘자료 수집’임을 학생들이 주목할 수 있게끔 하는 것이 도입 문항의 개발 의도이다.

두 번째 문항은 학생들이 자료 수집의 ‘방법’에 주목하여 표본추출에 대한 자신의 직관적인 이해를 최대한 드러내도록 설계되었다. 모집단을 대표하는 표본을 얻기 위해서는 표본 자체의 성질보다 표본을 추출하는 방법을 고려해야 한다. 구체적으로, 자료를 어떻게 수집하느냐에 따라 결과가 한쪽으로 치우치게 될 가능성이 있음을 학생들이 인지하는 한편, 적절한 자료 수집 방법에 대해 학생들이 가지고 있는 이해를 다양하게 드러내어 이후의 논의를 위한 발판을 제공한다.

세 번째 문항은 임의표본추출 상황을 제시하여 적절한 자료 수집 방법에 대한 학생

들의 다양한 논의를 수렴하는 역할을 할 것으로 기대된다. 학교수학 수준에서 통계적으로 적절한 표본추출은 모집단의 각 원소가 표본으로 추출될 가능성이 모두 같은 표본추출이며 이는 확률로 설명된다. 교사들은 이 문항을 활용하여 학생들이 ‘통계적으로 적절한 자료 수집 방법’에 대해 이해하도록 할 수 있다.

위에서의 논의를 바탕으로 학생들이 적절한 자료 수집 방법을 직접 생각해내는 통계 생산자의 역할을 수행하도록 설계된 것이 네 번째 문항이다. 통계 소비자로서의 소양을 넘어서 생산자로서 자료 수집 단계에서 고려해야 할 다양한 요소들을 논의하는 기회를 학생들에게 제공할 수 있을 것으로 기대된다.

학교수학에서는 통계적 추정까지만 다루고 있고 통계적 검정을 다루고 있지 않기 때문에, 통계 포스터 제작과 같이 학교에서 활용할 수 있는 통계적 탐구 과정에서 학생들이 도출한 통계적 결과를 검증하기 어렵다는 약점이 있다. 따라서 통계 생산자로서의 비판적 소양에 대한 교육이 선행될 필요가 있다는 점에서 본 활동지의 마무리 문항은 중요할 것으로 예상된다.

활동지1의 두 번째 활동 ‘설문조사를 믿는 방법’은 대중매체 등에서 익숙하게 찾아볼 수 있는 설문조사 결과를 해석하는 데 필요한 통계적 소양을 중심으로 개발되었다. 구체적으로, 통계 결과를 해석하는 과정에서 결과를 도출하기까지의 통계적 과정, 특히 ‘자료 수집’ 단계에 주목하게 함으로써, 자료 수집에 대한 비판적 의문을 가지고 평가하는 능력을 개발하기 위한 의도로 활동지의 각 문항들이 설계되었다.

첫 번째 문항은 설문조사 결과만을 제시한 채로 자료 수집 과정을 추측하게 하는 질문을 제기하는 것으로 시작한다. 대중매체를 통해 보도되는 많은 설문조사들이 조사 방법에 대한 충분한 설명 없이 결과만을 보도하는 경우가 많다. 따라서 이러한 통계 조사 결과들을 해석할 때는 조사 과정, 특히 자료 수집 단계에 대해 고려할 필요가 있음을, 도입 문항의 발문을 통해 인식케 할 수 있다.

두 번째 문항은 설문조사 결과를 해석할 때 학생들이 수집된 자료의 특성에 주목할 수 있도록 의도되었다. 통계학에서는 한정된 자료를 기반으로 결론을 도출하기 때문에, 그 자료에 대한 새로운 정보는 결론에 영향을 준다. 대중매체에서는 설문조사 과정에서 결론을 왜곡할 소지가 있는 자료 수집 과정의 문제를 은폐하기 위해 수집된 자료에

대한 정보를 감추거나 선별적으로 제공하는, 통계적 윤리부정행위를 일삼기도 한다. 이를 바탕으로 학생들은 자료에 대한 정보의 인지가 결론에 영향을 주는 통계학의 귀납적 성질을 이해하는 한편, 통계 조사 결과를 해석할 때 무엇을 고려해야 하는지를 생각해볼 수 있다.

세 번째 문항은 설문조사 결과를 자료 수집의 관점에서 비판적으로 평가하는 통계적 소양의 개발을 목적으로 설계된 것이다. 통계 조사 결과를 신뢰하기 위해 필요한 정보에 대한 논의는 결국 통계적으로 적절한 자료 수집에 대한 논의와 연결되면서, 궁극적으로는 자료 수집에 대한 올바른 이해가 통계적 소양의 핵심 요소임을 드러내게 된다. 통계학에서 자료 수집의 중요성을 일깨우면서 통계 조사 결과를 뒷받침할 수 있는 요소들을 이해하는 것은, 이후에 다룰 통계 생산자로서의 역량과 연결된다.

위에서의 논의를 바탕으로 학생들이 적절한 자료 수집을 스스로 생각해보게끔 설계된 것이 네 번째 문항이다. 자료 수집은 그 자체가 독립적인 단계가 아니라 문제 인식, 자료 분석, 결과 해석 등 통계적 조사 과정의 모든 단계와 연계되는 통계적 행위이다. 따라서 통계 포스터와 같이 통계적 결과를 산출할 때 생산자로서 자신의 결과를 정당화하는 의사소통으로서의 소양을 기르는 데 이러한 활동은 매우 중요하다.

나. 활동지2: 실험하기, 관찰하기, 측정하기

활동지2의 첫 번째 활동 ‘일조량에 따른 콩의 성장 비교’는 학생들이 직접 또는 간접적으로 조건에 따라 콩의 성장 비교 실험을 의도, 설계, 계획, 평가해보는 맥락을 제공하기 위해 개발되었다. 구체적으로 이 활동지의 계열은 실험의도, 실험 및 관찰 계획, 실험 및 관찰계획에 대한 질문 요소 순으로 개발·제시되었다. 이 하위계열을 기준으로 활동지의 개발의도를 상세하게 기술한다.

먼저 실험의 의도는 실험의 동기와 목표를 말한다. 이 활동지에서 실험의 의도는 서로 다른 빛의 조건에 따라 콩의 성장을 비교하기 위한 것이다. 실제로 학생들이 콩의 성장 비교 실험을 계획할 때 콩의 성장에 영향을 미치는 여러 조건들 중 어떤 조건을 선택할지 결정해야 하고 이후에 이를 준거로 실험이 설계된다. 실험의 의도는 이후에 전개될 실험 및 관찰 계획을 평가하는 기준이 되기 때문에 이를 명확히 하는 것은 때

우 중요하다.

다음은 실험 및 관찰 계획은 실험의도에 따라 실험을 설계하고 설계된 실험에 대해 관찰 계획을 세우는 단계이다. 실험을 설계할 때 동일한 조건(통제변인)과 서로 다른 조건(실험변인)을 명확하게 설정할 필요가 있다. 또, 설계된 실험에 대해 언제, 무엇을, 어떻게 관찰하고 자료를 수집할지에 대해 세부적인 계획을 세워야한다. 이 활동지에서 실험을 설계하고 관찰계획을 세우는 구체적인 과정을 안내하기 위해 실험 및 관찰 계획 단계와 관찰기록지를 제시하고 있다.

마지막으로 실험 및 관찰계획에 대한 질문은 상기 실험의도를 기준으로 설계한 실험 및 관찰에 대해 분석하고 평가해보는 활동을 이끌어내기 위해 개발되었다.

첫 번째 질문(질문1)은 설계된 실험에서 조건에 대해 분석해보는 활동을 통해 학생들이 실험설계상황에서 조건 설정에 민감해질 수 있도록 하기 위해 개발되었다. 구체적으로 작성된 실험 및 관찰계획에서 ‘같이 한 조건’과 ‘다르게 한 조건’을 식별해보게 하고 ‘같은지 다른지 알 수 없는 조건’도 찾아보게 하였다. 이때 학생들의 수준과 능력을 고려하여 실험에서 고려할 수 있는 다양한 조건을 보기로 제시하였다. 그래서 학생들의 수준과 능력이 높은 경우에는 보기를 제시하지 않을 수도 있다.

두 번째 질문(질문2)은 실험 및 관찰계획 단계에서 변이성(variability)¹⁾과 전체적인 조망(global views)을 고려해볼 수 있도록 하기 위해 개발되었다.

세 번째 질문(질문3)은 작성된 실험 및 관찰 계획이 실험의도에 적합한지를 평가해보고 이에 대한 근거를 마련하거나 실험 및 관찰 계획을 수정해볼 수 있도록 하기 위해 개발되었다.

활동지2의 두 번째 활동 ‘개미를 유인하는 화학물질 실험’은 어떤 화학물질이 개미를 유인하는지를 조사하기 위해 설계한 실험 및 측정 계획이 명시되어 있다. 위의 맥락은 학생들이 실험상황으로부터 실험을 설계하고 측정방법을 계획하는 것에 민감해질 수 있도록 할 것이다. 만약 실제 수업에서 이 실험을 직접 구현하고자 하는 경우 ‘어떤 화학물질’은 실생활에서 쉽게 구할 수 있는 것으로 대체할 수 있다. 예를 들어, 설탕물, 소금물, 식초, 베이킹파우더를 희석한 물 등 다양할 수 있다.

1) 영어 ‘variability’는 국내외 통계교육연구공동체에서 통계적 사고의 본질적인 요소로 강조되는 개념이다. 통상 우리나라에서 영어 ‘variability’는 ‘변이성’(우정호, 2004)이라고 한역하여 사용되고 있다.

첫 번째 문항은 학생들이 실험 및 측정방법에 대해 제대로 이해하고 있는지 확인하기 위해 개발되었다. 또 가상적으로 측정된 시침의 방향값을 구체적인 실험상황으로 옮겨 표시할 수 있는지를 확인하고자 하였다. 만약 수업 중 학생들이 이 과제에 대해 어려움을 보이는 경우 과제에서 제시된 시침의 방향값을 표시하는 방법을 교사가 시범적으로 보여준 이후 학생들이 직접 스스로 해볼 수 있도록 지도할 수 있다.

두 번째 문항은 실험 및 측정계획에 따라 실험을 수행한 결과로부터 결론을 내리는 것이 적절한지에 대해 학생들이 생각해보게 하는 것을 의도하였다. 일반적으로 실험상황에서 실험군과 대조군의 개념처럼 실험군의 결과로부터 신뢰할 수 있는 결론을 이끌어내기 위해서는 대조군의 실험 결과도 뒷받침될 필요가 있다.

하지만 초·중학교 수준의 학생들에게 실험군과 대조군 개념을 형식적으로 지도하는 것은 어려울 수 있으므로 이 개념에 대해 직관적으로 이해할 수 있도록 지도할 필요가 있다. 학생들의 수준과 능력이 높은 경우, 이 문항에서 수연이의 문제제기를 도입하기 전에 학생들 스스로 상원이의 실험결과로부터 내린 결론에 대해 평가해볼 수 있게 한다. 만약 상원의의 결론이 적절하다는 의견으로 몰리는 경우, 상원의의 결론이 부적절한 경우라고 가정하고 이에 대한 근거를 찾아보게 할 수도 있다.

한편, 이 과제를 해결하는데 어려움을 겪는 학생들에게 교사는 적절한 발문을 통해 논리적이고 비판적인 사고를 촉진시켜나가야 한다. 가령, 위의 과제에서 교사는 적절한 발문을 통해 상원이가 기본적으로 가정하고 있는 사실(화학물질이 칠해지지 않은 상태에서 실험을 진행했을 때 개미는 어느 한쪽 방향에 쏠리지 않고 모든 방향에 고르게 퍼진다)이 무엇인지를 학생들이 스스로 파악할 수 있도록 해야 한다.

다. 활동지3: 환경보호 프로젝트 참가하기

세 번째 활동지 ‘환경보호 프로젝트 참가하기’는 학생들이 환경보호관련 프로젝트를 수행하기 위해 구체적인 주제설정과 자료 수집을 계획해보는 것을 의도하고 개발되었다. 특히 학생들이 주제설정과 자료수집의 계획서 작성에 동기를 유발하고 그 내용과 필수요소를 안내하기 위해 ‘환경보호 프로젝트 참가 공고문’ 형식을 사용하였다. 또, 이 형식은 학생들의 프로젝트 계획서 작성활동에 평가기준으로도 사용될 수 있다. 교사는 학생들이 프로젝트 계획서 작성에 몰두할 수 있도록 동기를 충분히 유발시켜야 하고

계획서 작성 요령 및 작성된 계획서에 대한 평가방법을 상세하게 안내해야 한다. 이를 위해 자료수집 관련 교수학습 자료해설서의 ‘프로젝트 참가계획서 예시’를 참고할 수 있다.

제 2 절 지도를 위한 배경

통계교육의 목표는 학생들 스스로 통계적으로 사고할 수 있는 힘을 배양하는 것이다. 이때 ‘통계적으로 사고한다’는 것은 우리가 살고 있는 세계 도처에 존재하는 변이성(variability)을 인식하고, 기술하고, 양화하고, 설명해나가는 것을 의미한다. 특히 자료수집을 위해 설문조사 또는 실험을 설계하고 계획하는 단계에서 자료의 변이성을 미리 예측하고 고려하는 것은 매우 중요하다. 그래서 이하에는 지도를 위한 배경으로 다양한 종류의 변이성과 그 예시를 간략하게 소개한다²⁾.

가. 측정 변이성

동일한 사람 또는 사물을 대상으로 반복해서 측정할 경우 그 측정결과는 다를 수 있다. 가령, 어떤 한 학생이 30cm자를 이용하여 책상의 가로를 반복해서 측정하는 경우 그 측정값들 사이에 차이가 있을 수 있다. 또는 한 학급의 학생들이 한 번씩 30cm 자를 이용하여 책상의 가로를 측정하는 경우 그 측정값들 또한 다를 수 있다.

이와 같이 하나의 대상을 한 사람이 여러 번 측정하는 경우나 여러 사람이 한 번씩 측정하는 경우 모두 그 측정값들의 결과는 일정하지 않고 다를 것임을 예상할 수 있다. 이와 같이 측정상황에서 발생할 수 있는 변화, 차이, 다양성, 다름을 ‘측정에 의한 변이성’이라고 한다.

한편, 측정하려는 대상이 변하기 때문에 측정에 의한 변화가 발생할 수 있다. 예를 들어, 어떤 사람의 체온을 재는 경우 아무리 정확한 측정 도구를 사용하더라도 그 사람의 체온의 측정값은 서로 다를 것이다.

나. 고유 변이성

‘고유 변이성’은 체계에 고유하게 존재하는 변이성을 말한다. 가령, 한 학급에 있는

2) 본고에서 요약된 변이성의 종류와 예시는 GAISE(2005)보고서를 참조함.

학생들의 키를 측정할 때 그 측정값들은 서로 다를 것이다. 학생들의 키가 서로 다른 것은 일부 측정도구 때문에 발생할 수도 있지만 대부분은 사람의 키가 갖는 고유한 특성 때문이다. 사람들의 신체는 자연적으로 서로 다르고, 사람은 서로 다른 태도와 능력을 가지고 있으며, 서로 다른 의견이나 감정을 가지고 있다.

우리가 이런 자연적으로 발생한 사람의 속성(키, 몸무게, 태도, 의견, 감정 등)을 측정할 때 측정치들 간에 차이를 예상하는 것은 자연스럽다. 같은 종자의 콩을 동일한 환경에서 키우더라도 콩의 길이는 다를 수 있다. 그 이유는 콩의 종자가 서로 다르고 그 종자들마다 고유의 성장가능성을 잠재하고 있기 때문이다.

다. 유도 변이성

한 화분에 같은 종류의 콩 종자 5개를 심어 햇빛이 잘 드는 곳에서 키우고, 다른 화분에 동일한 종류의 콩 종자 5개를 심어 그늘에서 키운다고 했을 때 콩의 성장에서 발생하는 차이는 콩의 자연발생에 의한 변이성 때문일 수도 있고 빛의 조건이 달라서 이기 때문일 수도 있다. 이와 관련된 변이성은 햇빛이 콩의 성장에 미치는 영향을 알아보기 위해 의도적으로 유도하면서 발생한 것이다. 한 화분에는 비료를 주고 다른 화분에는 다른 종류의 비료를 주어 키웠을 때 콩의 성장의 차이는 비료의 차이 때문에 생겼을 것이다.

이와 같이 조건을 달리하여 설계한 실험은 다양한 요인들의 영향, 효과, 관계를 결정할 수 있게 해준다. 자연발생에 의한 변이성과 의도적으로 요인을 달리하여 다른 요인에 의해 발생하는 변이성을 비교하는 이 기본적인 아이디어가 현대통계학에서 가장 핵심이다. 의학을 연구하는 과학자들은 이 아이디어를 이용하여 어떤 약이 효과가 있고 안전하며 다른 약은 효과가 없거나 해로운 부작용이 있음을 밝힌다. 농업과 관련된 식물을 연구하는 과학자들은 이 아이디어를 이용하여 어떤 식물이 잘 자랄 수 있는 특정 환경을 조사하고 특정 비료가 다른 비료보다 더 효과적이며 특정 화학물질이 다른 물질보다 인체에 무해하고 병충해를 막는데 더 효과적이라는 주장을 세울 수 있다.

라. 표집 변이성

특정 후보를 지지하는 유권자들의 비율을 알아보기 위해 일부 유권자들을 대상으로 여론조사를 실시하여 이를 통해 추정하는 것은 통계에서 일반적이다. 이때 한 표본에

서 얻은 통계치가 다른 한 표본에서 얻은 통계치와 완전히 같지는 않을 것이다. 즉, 동일한 크기의 표본을 반복해서 추출할 때 표본마다 통계치는 서로 다를 것이다. 이와 같이 추출된 표본들 사이에 발생하는 차이를 ‘표집 변이성’이라고 한다.

만약 한 표본에서 구한 비율을 이용하여 모집단의 비율을 0.6이라고 추정하고 다른 한 표본에서 구한 비율을 이용하여 모집단의 비율을 0.4라고 추정하게 된다면 표본으로부터 얻은 모집단의 정보를 신뢰하기 힘들 것이다. 이런 결과가 가급적 발생하지 않도록 하기 위해서는 표본을 전략적으로 추출할 필요가 있다. 이를 위해 표본은 모집단을 잘 반영하도록 추출³⁾해야 하고 표본의 크기는 충분히 커야한다. 이와 같은 방법으로 뽑힌 표본으로부터 모집단을 추정할 때 표본들 사이에 차이를 수용하지 못할 만큼 큰 경우가 발생할 가능성은 거의 희박하다고 주장할 수 있다.

마. 우연 변이성

여론조사를 위해 표본을 추출하거나 실험을 설계하는 경우 무작위 아이디어는 두 가지 용도로 사용될 수 있다. 먼저 하나는 표본을 뽑을 때 무작위로 뽑는다. 다른 하나는 실험을 설계할 때 한 집단의 사람들을 무작위로 할당한다.

무작위 선택과 무작위 할당은 편향을 제거하지만 그 결과에 있어서 우연에 의한 차이를 발생시킬 수 있다. 이러한 차이를 ‘우연에 의한 변이성’이라고 하고, 이 우연에 의한 변이성은 확률모형으로 설명할 수 있다. 확률은 무작위로 뽑아서 나온 특정 표본의 결과에 대해서는 관심이 없지만 표본을 반복하여 추출했을 때 표본들 간에 차이에 대해서는 관심이 많다. 실험에서는 특정 실험 결과보다는 동일한 실험을 계속 반복할 경우에 생기는 변이성이 중요하다.

<참고문헌>

우정호(2004) 학교수학의 교육적 기초. 서울: 서울대학교출판부.

GAISE(2005). *Guideline for assessment and instruction in statistics education (GAISE) Report: A Curriculum framework for PreK-12 statistics education*. The American Statistical Association (ASA).

3) 모집단을 잘 반영하는 표본을 추출하는 대표적 기법으로 ‘임의추출법’이 있다. 만약 모집단이 너무 방대해 임의추출법을 적용하기 어려울 경우 대안적으로 층화추출이나 무선할당기법 등을 사용한다.

제 3 절 교수 학습자료 해설



활동지 1

설문조사하기



Activity 1 자료는 어떻게 수집해야 할까?

대전중학교 학생회장에 출마한 철구는 교내 매점에 대한 공약을 만들기 위해 학생들의 매점 이용 실태를 확인하고 싶어 합니다. 그래서 학생들이 매점을 하루에 몇 번 이용하는지 조사하기로 하였고, 친구인 상수와 민희에게 조사를 부탁하기로 하였습니다.

대전중학교에는 각 학급당 20명씩 학년당 5개 반이 있어, 전교생은 총 300명입니다. 상수와 민희는 모두 각각 30명의 학생을 대상으로 조사하기로 했습니다. 상수와 민희가 조사한 결과는 아래 표와 같습니다.

이용 횟수	0	1	2	3	4회 이상	합계
학생 수	6	11	7	4	2	30

[상수의 조사 결과]

이용 횟수	0	1	2	3	4회 이상	합계
학생 수	0	5	10	9	6	30

[민희의 조사 결과]

설문조사를 어떻게 했는지 묻는 질문에 상수와 민희는 다음과 같이 이야기했습니다.

- 상수 : 내 휴대전화 주소록에 있는 친구 30명을 대상으로 조사했어.
- 민희 : 교내 매점 앞에 조사 부스를 설치해서 선착순 30명을 대상으로 조사했어.

(1) 상수와 민희의 조사 결과가 다른 이유는 무엇인지 이야기해 봅시다.

- 우연과 변화가능성에 주목
 - 응답한 학생들이 다르기 때문이다.
 - 똑같은 조사를 해도 응답하는 학생들이 달라질 수 있다.
 (우연을 인정하더라도 허용할 수 있는 변화가능성의 범주에서 벗어난 결과임을 통해 문제 제기 필요)
- 표집 방법에 주목
 - 조사 방법이 달라지면 결과가 달라진다.

(2) 상수와 민희의 조사 방법은 적절한 것일까요? 자신의 생각을 이야기해봅시다.

[상수] ○ 표집 환경이 편의를 유발하는 변수로 작용 - 조사에 참여한 학생들은 모두 상수와 아는 사람들이다. - 상수의 친구들은 대부분 남자일 가능성이 있다. - 상수의 친구들은 상수와 매점 이용 횟수가 비슷할 것이다. - 상수가 일부러 매점 이용 횟수가 많은 혹은 적은 친구들만을 골라서 조사할 수 있다. ○ 표본조사의 불확실성 - 모든 학생들을 대상으로 조사하지 않았다. - 모든 학생들이 상수의 친구가 아니다. - 조사 학생 수가 너무 적다.	[민희] ○ 표집 환경이 편의를 유발하는 변수로 작용 - 매점을 이용하는 학생들만 조사에 참여했을 것이다. - 자신의 의견을 표현하기를 희망하는 학생들만 조사에 참여했을 것이다. ○ 표본조사의 불확실성 - 모든 학생들을 대상으로 조사하지 않았다. - 모든 학생들이 매점을 이용하는 것은 아니다. - 조사 학생 수가 너무 적다.
--	---

민식이의 다음과 같이 새로운 설문조사 방법을 제안했습니다.

○ 민식 : 전교생의 이름이 각각 적힌 종이를 상자에 넣어서 그 중 30개를 뽑아 그 학생들을 조사하자.

(3) 민식이의 조사 방법이 상수, 민희의 방법과 비교했을 때 더 적절한지 이야기해봅시다.

○ 반응 1: 적절하다 - 임의로 학생들을 조사하였다. - 모든 학생들이 조사에 참여하게 될 가능성이 같다. - 모든 학생들이 공평하게 조사에 참여할 수 있다. - 민식이의 의도가 개입되지 않는다.	○ 반응 2: 적절하지 않다 - 모든 학생들을 대상으로 조사하지 않았다. - 조사 학생 수가 너무 적다. - 조사 방법이 너무 복잡하고 시간이 오래 걸린다. - 전교생의 이름을 다 적는 종이를 만들기가 어렵다. - 뽑힌 학생들이 조사에 응하지 않을 수 있다.
--	---

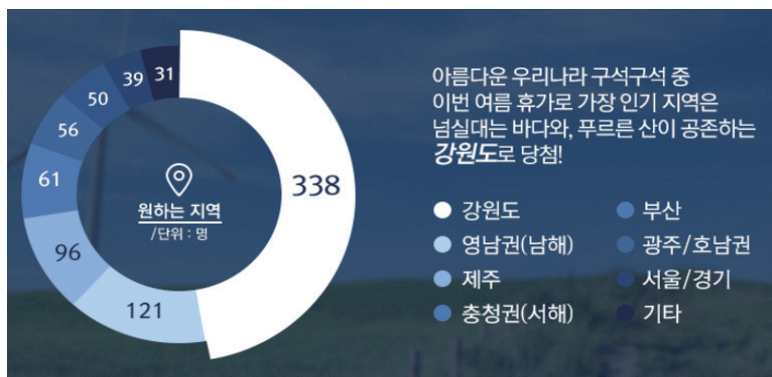
(4) 만약 본인이 직접 조사를 한다면 어떻게 할 것인지 생각해보고 자신의 조사가 왜 타당한지 이야기해봅시다.

○ 전수조사 시행 - 전교생 모두에게 설문지를 돌린다.	○ 임의표집 시행 - 난수표나 컴퓨터를 이용한다.
○ 상수와 민희의 방법을 수정(표집 환경의 변화를 통한 편의 제어) - 조사 부스를 매점 앞이 아닌 다른 곳으로 옮긴다. - 여러 사람의 휴대전화 주소록에서 조사할 학생들을 선정한다.	○ 층화표집 혹은 군집표집 시행 - 각 학년별로 동일한 숫자만큼의 학생을 대상으로 조사한다. - 동일한 수의 남학생과 여학생을 대상으로 조사한다. - 각 반에서 똑같은 인원을 선발하여 조사한다. - 특정 번호가 출석번호인 학생들을 선발하여 조사한다.

Activity 2 설문조사를 믿는 방법

많은 사람들이 여름휴가를 보내기 위해 다른 지역으로 여행을 갑니다. 사람들이 여름휴가를 어디서 어떻게 보내는지에 대한 조사는 기업에게 숙박, 관광, 교통수단 등 상품 개발과 마케팅 전략을 세우는 데 활용되는 중요한 정보입니다. 정부에서는 여가와 관련한 다양한 정책을 입안하는 데 조사 결과를 활용할 수 있으며, 여행객 또한 자신의 여행 계획을 세우는 데 참고할 수 있습니다.

한 기업에서는 지난 2016년 7월에 약 800명을 대상으로 설문조사를 실시하여 아래와 같이 블로그를 작성하였습니다.



(1) 희선이는 블로그를 보고 다음과 같이 의문이 생겼습니다.

○ 희선 : 이 기업에서는 이런 정보를 어떻게 얻을 수 있었을까?

이 기업에서는 조사에 참여할 사람들을 어떻게 모집하였을지 논의해봅시다.

- 표집자에 의한 편의가 발생하는 경우
 - 블로그나 홈페이지, SNS에서 설문조사를 했을 것이다.
 - 기업 고객이나 사원을 대상으로 설문조사를 했을 것이다.
- 표집자에 의한 편의가 발생하지 않는 경우
 - 사람이 많은 곳에서 부스를 설치해 설문조사를 했을 것이다.
 - 임의로 전화걸기를 하여 설문조사를 했을 것이다.
 - 남녀를 나누어 조사했을 것이다.

(2) 현욱이는 블로그를 보고 다음과 같이 말했습니다.

○ 현욱 : 이 설문조사에서는 조사에 참여한 사람들의 연령대, 성비, 거주 지역과 같은 정보가 없어.

현욱이가 말한 이 사실이 위 그래프를 해석하는 데 어떤 영향을 미칠 수 있을지 이야기해 봅시다.

반응 1: 영향을 미친다.

- 특정 연령대, 성, 거주자들을 대상으로 편중된 조사를 했을 가능성을 생각해야 한다.
- 조사한 기업이 의도적으로 원하는 대상만을 조사했을 수도 있다.
- 자세한 정보가 없기 때문에 그래프의 결과를 믿기 어렵게 된다.

반응 2: 영향을 미치지 않는다.

- 모든 요소들을 다 고려하여 조사를 할 수는 없기 때문에 해석에 영향을 주지 않는다.
- 참여자를 임의로 모집했다면 해석에 영향을 주지 않는다.

(3) 이 조사가 적절하게 이루어졌다고 확인하기 위해서는 어떤 정보가 더 필요한지 이야기해 봅시다.

○ 층화표집을 위한 정보

- 참여자의 연령대 비율, 성비, 거주 지역 비율, 직업군 비율 등이 필요하다.
- 참여자 거주 지역에 대한 정보는 다른 정보보다 조사 내용에 영향을 미칠 가능성이 높아 필요하다.

○ 임의표집을 위한 정보

- 참여자를 어떻게 모집했는지에 대한 정보가 필요하다.
- 참여자들을 임의로 모집했다는 사실을 보여줘야 한다.

(4) 만약 본인이 직접 조사를 한다면 누구를 대상으로 어떻게 조사를 할 것인지 생각해봅시다.

임의표집, 층화표집에 해당하는 다양한 아이디어가 나올 수 있다. 혹은, 표본의 수를 늘리는 방법에 대해 이야기하는 학생이 나올 수도 있으나, 이 경우에는 적절한 안내(예를 들어 아래 읽기자료)를 통해 표집 '방법'에 주목할 수 있도록 유도할 필요가 있다. 통계의 불확실성을 이유로 전수조사를 주장하는 학생도 있을 수 있으나, 경제적·현실적인 문제를 인식시켜 임의표집이나 층화표집이 차선택임을 이해할 수 있도록 유도한다. 이 과정에서 통계에 대한 맹목적 불신이 일어나지 않도록 표본조사의 유용성을 일깨워주는 일화를 활용하는 방안도 고려해볼 수 있다.



활동지 2

실험하기, 관찰하기, 측정하기



Activity 1 일조량에 따른 콩의 성장 실험

상원이는 일조량에 따라 콩의 성장을 비교하기 위해 다음과 같은 관찰 계획을 세웠다.

- 가. 동일한 크기의 화분 A, B 각각에 일정한 크기로 싹이 난 5개의 콩을 동일한 간격으로 심는다.
- 나. 두 화분을 모두 실내에서 키우되 화분A는 햇빛이 잘 드는 창가 옆에 놓고, 화분B는 마분지로 그늘을 만들어 키운다.
- 다. 50일 동안 5일 간격으로 같은 시간에 각각의 화분에 심은 5개의 콩의 키를 자를 이용하여 밀리미터 단위까지 측정한다.
- 라. 측정한 콩의 줄기의 길이는 관찰기록지에 기록한다.
- 마. 콩의 줄기의 길이를 관찰하는 기간 동안 두 화분에 주는 물의 양은 동일하게 유지한다.

화분 종류	콩 번호	관찰기록지										관찰 중 특이사항
		5일	10일	15일	20일	25일	30일	35일	40일	45일	50일	
A	1											
	2											
	3											
	4											
	5											
B	1											
	2											
	3											
	4											
	5											

(1) 상원이의 관찰계획에서 같게 한 조건과 다르게 한 조건을 아래의 <보기>에서 고르시오.

<보기>

- ㉠ 빛의 양
- ㉡ 물의 양
- ㉢ 물의 온도
- ㉣ 화분의 크기
- ㉤ 토양의 종류
- ㉥ 콩의 종자
- ㉦ 싹의 크기
- ㉧ 측정간격

① 같게 한 조건

㉡, ㉢, ㉥, ㉦, ㉧

② 다르게 한 조건

㉞

③ 같은지 다른지 알 수 없는 조건

㉟, ㊱

(2) 상원이의 관찰 계획을 읽고 민수와 혜원이 다음과 같은 문제를 제기하였다.

- 민수: 두 화분 각각에 콩을 하나씩만 심어서 크기를 비교하면 될 텐데 화분 당 5개씩 심은 이유가 뭐야?
- 혜원: 서로 다른 조건에서 자란 콩의 크기에 대해 최종적으로 측정할 자료(50일째)만 있으면 되지 않나? 5일째부터 45일째까지 측정할 필요가 있나?

다음 물음에 답하여라.

① 민수의 물음에 상원이가 어떻게 대답했는지 생각해보고 친구들과 이야기해봅시다.

콩이 동일한 조건에서 자라더라도 키는 제각각일 수 있기 때문에 서로 다른 조건에서 자란 콩의 크기를 하나씩 비교하는 것은 신뢰하기 힘들다. 한 화분에서 자란 한 개의 콩이 조건에 상관없이 너무 크게 자라거나 작게 자랄 수 있기(혹은 콩의 싹이 죽어 자라지 않는 경우도 발생할 수 있음) 때문에 각 화분에 여러 개의 콩을 심어 자란 콩의 길이를 통계적으로 비교하면 더욱 신뢰할 수 있는 결과를 얻을 수 있다.

② 혜원의 물음에 상원이가 어떻게 대답했는지 생각해보고 친구들과 이야기해봅시다.

서로 다른 조건에서 다 자란 콩의 크기만을 측정하면 콩의 성장 과정을 전체적으로 비교할 수 없다. 또 서로 다른 조건에서 자란 콩의 한 살이를 비교하면 더욱 세부적인 콩의 성장 과정을 비교할 수 있다.

(3) 상원이는 콩의 성장에 대한 관찰을 통해 다음과 같은 주장을 세우고자 한다.

콩은 일조량이 많을수록 더 잘 자랄 것이다.

다음 물음에 답하여라.

① 상원이의 관찰 계획은 위와 같은 주장을 뒷받침하기에 적절한 방법인지 이야기해봅시다.

- 반응1: 적절하다.
- 반응2: 적절하지 않다.

- ② 위의 물음①에 대해 적절하다고 판단한 경우, 그 이유에 대해 친구들과 이야기해보십시오.

상원이는 빛의 양(햇빛을 받고 자란 콩, 그늘에서 자란 콩)은 달리고 그 외 대부분의 조건(물의 양, 화분의 크기, 콩의 종자, 싹의 크기, 측정 간격)은 동일하게 유지하여 콩의 성장에 대한 관찰계획을 세웠기 때문이다.

- ③ 위의 물음①에 대해 부적절하다고 판단한 경우, 상원이의 관찰 계획에서 수정되어야 할 것들에 대해 친구들과 이야기해보십시오.

- 반응1: 상원이의 관찰 계획에는 같게 한 조건과 다르게 한 조건 이외에도 같은지 다른지 알 수 없는 조건(물의 온도, 토양의 종류)이 포함되어있다. 화분에 주는 물의 온도나 토양의 종류는 콩의 성장에 영향을 미칠 수 있으므로 상원이의 관찰 계획은 부적절하다. 그래서 상원이의 관찰 계획에서 같은지 다른지 알 수 없는 두 조건을 동일하게 유지한다는 조건을 추가할 필요가 있다.
- 반응2: 상원이는 일조량에 따른 콩의 성장을 확인하고자 하였다. 그런데 일조량에 대해 그늘에서 키운 경우와 햇빛을 받고 자란 경우로만 나누어 실험을 계획하였다. 일조량에 따른 콩의 성장에 대한 주장을 세우기 위해서는 햇빛을 받는 시간에 변화를 주어 즉 햇빛을 받는 시간이 0시간, 3시간, 6시간, 9시간인 경우와 같이 일조량에서 차이를 두어 실험을 계획한 후 이에 대한 실험을 실시해야한다.

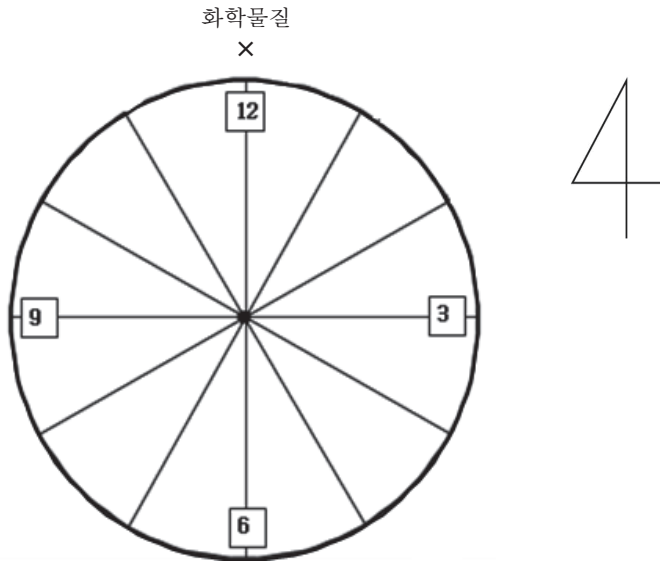


Activity 2 개미를 유인하는 화학물질 실험

상원이는 어떤 화학물질의 냄새가 개미를 유인할 수 있는지를 실험해보기 위해 다음과 같은 실험을 계획하였다.

가. 페트리 접시에 개미 30마리를 준비한다.

나. 아래와 같이 반경 10cm의 원 둘레에 시계눈금을 그려두고 12시 방향의 눈금 1cm 뒤에 그 화학물질을 칠한다.



다. 원의 중심의 위치에서 개미를 한 마리씩 출발시켜 원 둘레에 도착한 위치의 시침 방향 값을 측정한다.(단, 시침은 10분 단위로 작성)

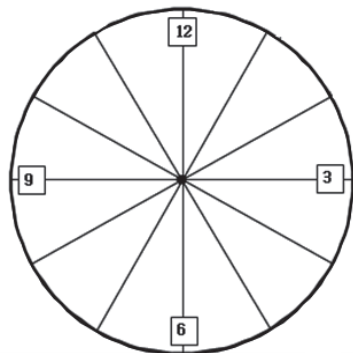
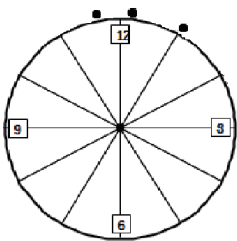
라. 30마리의 개미에 대해 다.의 과정을 독립적으로 반복 실시한다.

(1) 다음은 상원이가 처음 세 마리의 개미에 대해 실험을 실시하여 측정한 시침의 방향값이다.

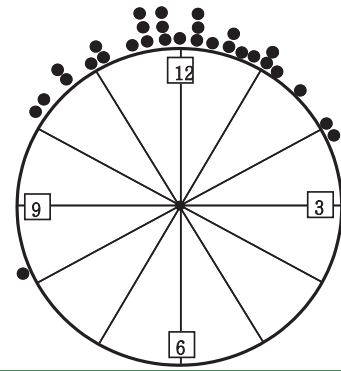
1시 0분 11시 40분 12시 10분

세 시침의 방향 값에 대응되는 위치를 오른쪽 그림에 표시해봅시다.

[해답]



- (2) 오른쪽 그림은 상원이가 30마리의 개미에 대해 실험을 실시하여 각각의 개미가 도착한 위치와 횟수를 점으로 표시한 것이다. 점은 그 위치에 도착한 개미의 횟수만큼 표시된다. 상원이는 대부분의 개미가 12시 방향에서 가까운 위치에 도착한다는 사실을 통해서 '그 화학물질은 개미를 유인 한다'고 결론 내렸다. 상원이의 결론에 대해 수연이는 다음과 같은 문제를 제기하였다.



○ 수연: 개미가 12시 방향으로 움직이는 것이 화학물질 때문이 아니라 원래 그 방향으로 이동하는 경향이 있다면 이런 결과가 나올 수 있을 것 같아.

수연이가 제기한 문제를 해결하기 위해 상원이가 적절히 대처할 수 있는 방법에 대해 생각해보고 친구들과 이야기해봅시다.

- 반응1: 상원이는 위의 그림을 통해 화학물질이 개미를 유인한다는 결론을 내렸다. 그림에서 볼 수 있듯이 대부분의 개미가 12시 방향으로 이동했기 때문이다. 이 결론은 화학물질이 칠해지지 않은 상태에서 실험을 진행했을 때 어느 한쪽 방향에 쏠리지 않고 모든 방향에 고르게 퍼질 것이라는 가정이 깔려있다. 이 가정이 사실인지를 확인하기 위해 화학물질이 칠해지지 않은 상태에서 실험을 진행하여 그 결과를 확인해본다.
- 반응2: 수연이의 문제를 해결하기 위해 12시 방향에 화학물질을 칠해서 실험한 결과와 다른 방향(예를 들어, 6시)에 화학물질을 칠하여 실험한 결과를 각각 비교해본다.



활동지 3

환경보호 프로젝트 참가하기



Activity 1 환경보호 프로젝트 계획하기

○○환경단체에서 ‘되살리자! 우리 세계!’라는 슬로건을 내걸고 환경 되살리기 운동을 실시하고 있다. 이 단체에서는 이 운동의 일환으로 중학생들을 대상으로 환경보호 프로젝트 대회를 개최하고자 한다. 이 대회에 참가하기 위해서 학생들은 직접 환경보호와 관련된 주제를 설정하고 어떤 자료를 어떻게 수집할지 계획서를 작성하여 제출해야 한다. 아래의 ‘환경보호 프로젝트 참가 공고문’을 참고하여 환경보호 프로젝트 계획서를 작성해보자.

환경보호 프로젝트 참가 공고문

1. 환경보호와 관련된 주제를 설정한다.
2. 선정된 주제와 관련된 자료수집 계획을 작성한다.
3. 자료수집계획은 다음을 반드시 포함하여 구체적으로 작성한다.
 - 수집 대상
 - 수집 방법
 - 측정 방법 및 단위
 - 측정 도구
 - 측정기록지 양식
4. 계획서 양식은 자유롭다.



<출처: 네이버 어린이 백과>



♣ 프로젝트 참가 계획서 예시

1. 주제설정: 교실에서 버리는 쓰레기양
2. 자료수집계획
 - 수집 대상: 각 학급에서 하루에 버리는 종류별(종이, 캔, 플라스틱, 기타) 쓰레기양
 - 수집 방법: 각 학급에 배치된 분리수거함을 이용하여 하루에 배출하는 쓰레기양을 측정하여 기록
 - 측정 방법 및 단위: 종이(버려진 종이의 무게(g) 재기), 캔(버려진 캔의 무게(g) 재기), 플라스틱(버려진 플라스틱 무게(g) 재기)
 - 측정 도구: 저울
 - 측정기록지 양식

종류 학급	종이(g)	캔(g)	플라스틱(g)	기타	합계
1-1					
⋮					
합계					

제 3 장 그래프

제 3 장 그래프

제 1 절 개발 의도

통계자료를 제시하는 방법에는 그래프, 표, 자료를 해석해서 글로 작성하는 방법 등이 있다. 이러한 자료를 요약하고 표현하는 방법 중 어떤 것을 이용할 것인지는 전달 목적과 대상, 전달하고자 하는 자료의 특성에 따라 달라진다. 전체적인 경향을 전달하거나 강조하고자 한다면 그래프를, 상세한 정보를 쉽게 찾아보게 하거나 정밀한 내용이나 복잡한 정보를 전달하고 싶다면 표를, 그리고 이러한 정보 전달에 익숙하지 않은 사람이거나 정보의 해석이 정확해야 한다고 본다면 글을 이용하는 것이 좋다. 이러한 방법은 상호 보완적으로 사용될 수 있다. 그래프나 표의 경우 글의 도움 없이 쉽게 해석이 되는 경우도 있지만 글의 도움을 받아야 의미를 정확하게 전달할 수 있는 경우도 있기 때문이다.

수학교육과정에서는 이러한 방법이 그 쓰임이나 목적에 따라 상호 보완적인 역할을 할 수 있도록 구성되어 있다. 특히 중등교육과정에서는 평균, 중앙값, 최빈값, 평균, 분산과 같은 대푯값을 통한 자료의 해석에만 치우쳐 표와 그래프로 나타내고 이를 해석하는 학습이 부족한 경향이 있다.

이 장에서는 그래프의 효과, 목적에 맞는 그래프의 사용, 그래프 해석을 할 때 주의해야 할 점 등 그래프로 나타낼 때 주의해야 할 점을 생각할 수 있도록 하였다.

첫 번째, ‘그래프의 마술’ 활동에서는 나이팅게일의 ‘로즈다이어그램’을 통해 하나의 그래프가 가지는 위력을 느낄 수 있다. ‘Activity 1, 그래프의 착시와 선택’에서는 착시 효과를 불러오는 원인에 대해 생각해 보고 목적에 알맞은 그래프는 어떤 것인지 생각해 볼 수 있도록 하였다. ‘Activity 2, 효과적인 그래프’에서는 그래프에서 변수명, 측정 단위, 집단 명과 같은 범례를 추가하면서 새로 찾아낼 수 있는 정보를 생각해 볼 수 있도록 하였다. 이 두 활동을 통해 목적에 맞는 그래프를 효과적으로 작성하는 방법을 알게 될 것이다.

두 번째, ‘스마트폰으로부터 학생을 지켜라’는 수집한 자료를 표로 정리하고 이를 자료의 특성에 맞는 그래프를 이용하여 설명할 수 있도록 하였다. 교과서에서의 표는 항상 그 계급의 크기가 고정되어 있으며 계급의 도수를 구하는 활동에 국한되어 있다. 하지만 본 교수학습 자료에서 제시하는 활동에서는 계급의 크기를 조정할 때 생기는 현상을 살펴봄으로써 표에서 계급 크기의 중요성을 인식할 수 있을 것이다. 또한

자료의 특성에 맞는 가장 적합한 그래프를 찾는 과정을 통해 그래프의 다양한 종류와 쓰임에 대해 알게 될 것이다.

마지막으로 ‘도전! Treasure Hunt’에서는 집단의 특성을 추가할 경우 자료의 해석이 달라진다는 것을 알 수 있도록 구성하였다. ‘Activity 1, 성별에 따른 자료 분석하기(1)’에서 학생을 성별로 구분하면 남학생이 22명, 여학생이 18명이다. 스마트폰의 주된 사용 목적에 대해 남학생과 여학생이 선택한 인원수는 9명으로 동일하다. 같은 9명이지만 남학생과 여학생의 전체인원이 다르기 때문에 다른 의미를 가질 수 있다는 의문을 가질 수 있으며 이는 ‘상대도수, 비율’의 개념을 유추하는 힌트가 된다. ‘Activity 2, 성별에 따른 자료 분석하기(2)’에서는 집단에 따른 변량을 분포를 효과적으로 비교하기 위한 방법을 생각할 수 있도록 하였다.

제 2 절 지도를 위한 배경

자료의 형태와 목적에 따라 여러 가지 그래프로 나타낼 수 있다. 자료는 수치화 가능 여부에 따라 양적 자료와 질적 자료로 구분된다.

양적 자료는 수치로 나타낼 수 있는 자료로 이산형 자료와 연속형 자료로 구분된다. 이산형 자료는 사람 수, 점수 등과 같이 셀 수 있으나 연속형 자료는 키, 몸무게, 길이 등과 같이 셀 수 없다.

질적 자료는 수치로 나타낼 수 없는 자료로 범주형 자료와 순위형 자료로 구분된다. 범주형 자료는 성별, 혈액형, 좋아하는 계절 등과 순위를 구분할 수 없다. 순위형 자료는 최종학력(예를 들면 중졸, 고졸, 대졸, 대졸이상), 선호도(예를 들면 매우불만, 불만, 보통, 만족, 매우만족) 와 같은 자료로 순위를 구분할 수 있으며 이로 인해 특정 통계 분석에서는 양적 자료로 취급하기도 한다.

통계에서는 정리하고자 하는 자료의 특성과 자료를 정리하는 목적에 따라 여러 가지 형태의 그래프로 나타낼 수 있으며 초·중등학교에 나오는 그래프를 살펴보면 다음과 같다.⁴⁾

가. 선그래프

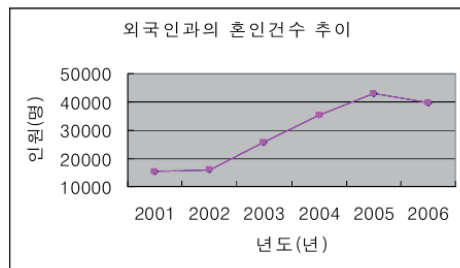
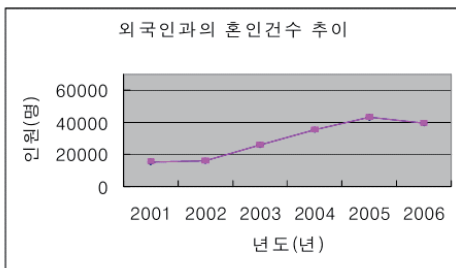
선그래프는 특정 변량이 시간이나 선형 함수관계의 변화에 따라 어떻게 변화하여 가는지를 살펴보기에 좋은 그래프이다. 예를 들면, 대선 후보들의 시기별 지지율 변화, 월별 태양이 지는 평균 시간과 같은 그래프를 들 수 있다. 선그래프는 하나의 변량을 몇몇 시점에 관찰하여 나온 값들을 점으로 찍고 그것들을 선으로 연결하여 그리는 것

4) 이선애(2007), 통계 그래프의 지도에 대한 고찰과 학생들의 통계 그래프 이해 능력 분석, 서울대학교 석사 논문

으로, 관찰하지 않은 중간시기에 변량의 관측값이 어떠한지를 예측할 수 있는 장점이 있다.

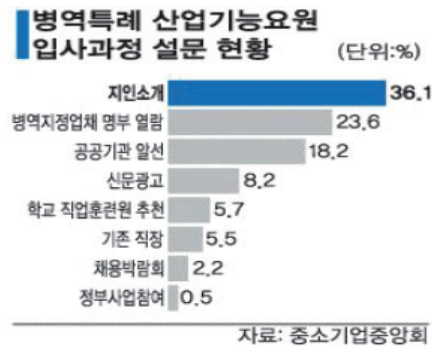


위와 같은 선그래프의 지도 시 주의할 점은 같은 그래프라도 아래 두 그래프와 같이 세로축 눈금 척도를 달리하여 그림을 그릴 경우에 변량의 변화율이 다르게 보인다는 점이다. 따라서 상황과 목적에 따라 적당한 눈금 척도를 사용하고, 그래프를 읽을 때에도 세로축의 눈금 간격이 어떤 수를 상징하는지 잘 관찰할 필요가 있다.



나. 막대그래프

막대그래프는 기둥그래프라고도 하며, 막대를 가로로 그리느냐 세로로 그리느냐에 따라 수평 막대그래프와 수직 막대그래프가 있다. 막대그래프는 질적 자료를 정리하는데 많이 사용한다. 그리고 가족의 형제자매의 수와 같이 양적 자료이지만 자료가 이산적이면서 변량의 수가 많지 않은 경우에도 사용한다. 이러한 막대그래프는 도수가 막대 길이에 비례하여 이산적인 자료를 비교하는데 유용하다. 막대그래프를 그릴 때 유의할 점은 막대의 너비를 같게 그리고 막대의 길이가 도수에 정비례하도록 그리는 것이다. 그리고 막대가 등간격으로 서로 떨어져 있도록 그린다. 이러한 막대그래프는 도수의 크기를 정확히 나타낼 수 있고, 자료들의 상대적인 크기나 전체적인 분포를 살펴보기에 좋다.



막대그래프 역시 선그래프처럼 물결선을 사용하여 척도의 크기를 변화하면 자료의 크기가 상대적으로 다르게 보일 수 있다는 점에 주의할 필요가 있다.

다. 줄기와 잎 그림

줄기와 잎 그림은 자료의 분포를 보고자 할 때 주로 사용하는 그래프이다. 줄기와 잎 그림은 아래의 그림과 같이 하나의 자료를 두 영역(줄기영역과 잎영역)으로 나누어서 줄기를 기준으로 사전식 배열처럼 나열하는 그래프이다.

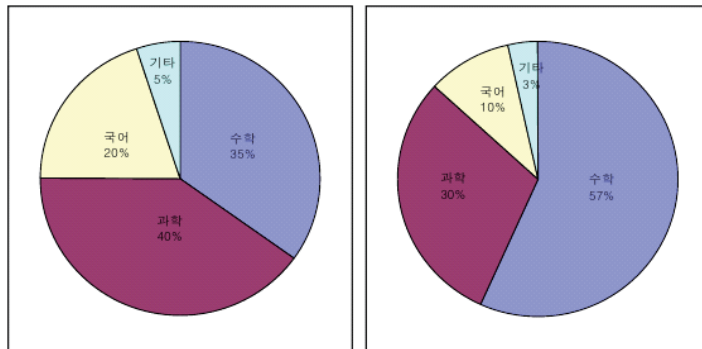
줄기	잎
1	124
2	013
3	122358
4	2399
5	7
6	23
7	2

위 그래프에서 첫째행의 “1|124”는 11살, 12살, 14살의 자료가 있음을 상징한다. 이 그래프는 자료 전체의 형태를 빨리 파악할 수 있고 자료값을 그대로 유지하고 특정한 값이 얼마나 자주 나타나는지를 쉽게 알아볼 수 있는 장점이 있으나, 대략적으로 알 수 있는 자료나 자료의 크기가 클 때는 부적절하다. 줄기와 잎 그림에서 줄기를 몇 개로 선택하느냐에 따라 분포가 달라지기 때문에 적절한 줄기를 결정하는 것이 중요하다.

라. 비율그래프

비율그래프는 대통령 선거에서 각 후보의 지지율과 같이 질적 자료에 많이 사용한다. 대통령 선거에서는 A후보의 당선여부는 투표자 중에서 몇 %의 사람이 그 후보를 지지하였는지가 중요하다. 이와 같이 비율그래프는 전체에 대한 부분의 비율을 보고자 하는 경우에 사용한다. 비율그래프의 대표적인 예는 우리가 일반적으로 알고 있는 원그래프와 띠그래프이다 원그래프는 아래의 그림처럼 전체에 대한 부분의 비율만큼 원을 부채꼴 모양으로 나누어 나타낸 그래프이고, 막대그래프는 전체에 대한 부분의 비율만큼 막대를 분해하여 나타낸 그래프이다. 줄기와 옆 그림은 자료의 분포를 보고자 할 때 주로 사용하는 그래프이다.

구분	수학	과학	국어	기타	계
A학교	84	96	48	12	240
B학교	68	36	12	4	120



위의 그림을 보면 A학교 학생들은 과학을 좋아하는 학생이 많고, B학교 학생들은 수학을 좋아하는 학생이 많다는 사실을 알 수 있다. 원그래프의 지도 시 학생들이 전체에 대한 부분의 비율을 구하고 원을 비율에 맞게 나누어 그릴 수 있도록 해야 하고, 비율의 크기가 도수의 절대적인 크기가 아님을 인지시킬 필요가 있다. 예를 들면 위의 두 원그래프에서 수학을 좋아하는 학생의 비율을 보면 B학교가 많지만, 수학을 좋아하는 학생 수는 A학교가 84명으로 B학교의 68명보다 많다.

제 3 절 교수 학습자료 해설

그래프



활동지 1

그래프의 마술

최근 ○○중학교에서는 스마트폰을 사용하며 계단을 내려가던 K군과 스마트폰을 사용하며 계단을 올라가던 J양이 부딪쳐 계단 아래로 굴러 떨어지는 사고가 발생하였습니다. K군은 손목뼈에 금이 가서 깁스를 하게 되었고 J양은 타박상을 입었습니다.

이 사건을 취재하던 신문부에서는 이 사건과 함께 ‘청소년들의 스마트폰 이용실태’와 ‘○○중학교 학생들의 스마트폰 중독 실태’를 조사하여 다음 달 교내 신문에 특집 기사로 실기로 하였습니다.

‘○○중학교 학생들의 스마트폰 중독 실태’는 1, 2, 3학년 각반에 3명(10번, 20번, 30번)의 학생을 정한 후 한국인터넷진흥원에서 만든 ‘스마트폰 중독 자가진단 검사지’를 이용하여 스마트폰 중독실태를 조사하였습니다.

○○중학교 신문부 학생들이 특집 기사를 효과적으로 표현하기 위해 편집자 회의 를 하는 과정을 살펴보도록 합시다.



○진혁 : 조사한 자료를 좀 더 알기 쉽게 설명할 방법이 없을까?

○성욱 : 숫자나 표로 나타내는 것보단 그래프를 이용해서 나타내면 좀 더 쉽게 전달할 수 있을 것 같아.

○대준 : 그래프? 그래프로 나타내면 학 와 닿아서 좋긴 한데 잘못 받아들일 수도 있지 않을까?

○성욱 : 자료의 특성과 목적에 맞는 그래프를 사용한다면 괜찮은 것 같은데

○대준 : 그래프로 나타낼 때 또 신경써야할 건 없을까?

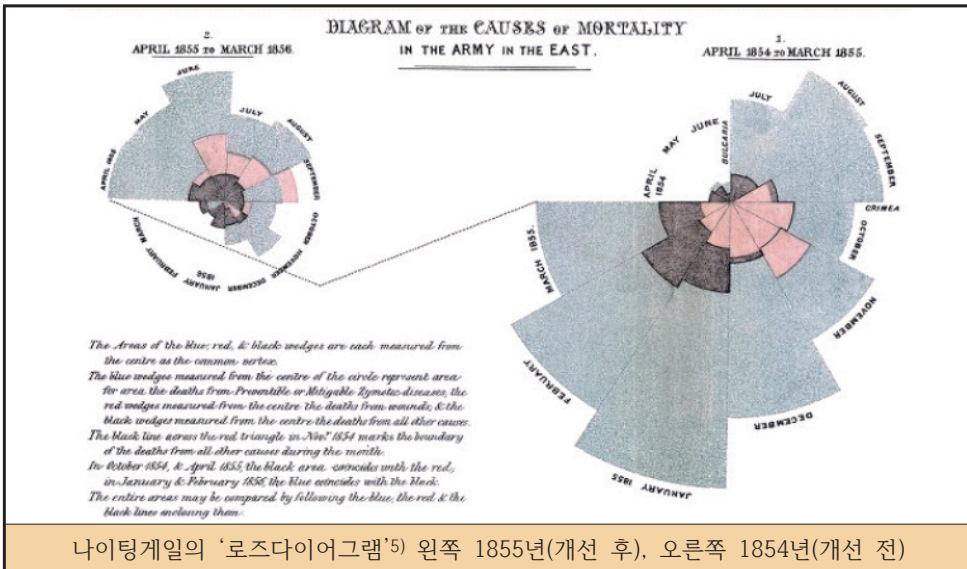
○진혁 : 그래프에 눈금이라든지 무엇을 나타낸 것인지... 이런 걸 잘 표시해야 할 것 같아.

○대준 : 그렇잖아. 그래프에서 1명 단위인지 1000명 단위인지 이런 거 때문에 잘못 이해했던 적이 있어. 어디 참고할 만한 좋은 그래프는 없을까?

○성욱 : ‘로즈 다이어그램’이란 걸 본 적이 있어. 나이팅게일이 만든 새로운 종류의 그래프인데 크림전쟁에서 죽어가는 군인들의 수를 월별, 죽은 원인별로 정리해서 나타낸 거야. 이 그래프를 통해 막상 전쟁에서 죽는 사람들이 전쟁이 아닌 위생이 안 좋아 죽어간다는 걸 알렸고 그 결과 많은 군인들의 생명을 지킬 수 있었어. 한 번 살펴볼까?

♣ 나이팅게일의 ‘로즈다이어그램’

‘백문(百聞)이 불여일견(一見)’은 100번 듣는 것 보다 한 번 보는 것이 좋다는 뜻입니다. 자료를 분석 하는 경우 한 장의 그래프가 100문장의 설명보다 좋을 때도 있습니다. 다음을 통해 그래프의 위력을 알아보도록 합시다.



나이팅게일의 ‘로즈다이어그램’⁵⁾ 왼쪽 1855년(개선 후), 오른쪽 1854년(개선 전)

1854년 러시아와 연합군 간에 ‘크림전쟁’이 일어났습니다. 많은 영국 군인들이 부사와 질병으로 죽었습니다. 나이팅게일은 크림전쟁에서 간호장교로 활동하였습니다. 나이팅게일은 먼저 통계 작성기준을 세워 기록체계를 통일하였습니다.

오른쪽 그래프는 1854년에 사망한 군인의 수를 나타낸 것이고 왼쪽은 1855년에 사망한 군인의 수를 나타낸 것으로 부채꼴의 넓이는 사망자의 수를 나타냅니다.

또, 그래프에 달을 표시하고 사망원인을 색깔로 구분하였습니다. 파란색(가장 바깥쪽)은 질병, 빨간색(가운데)은 부상, 검은색(가장 안쪽)은 기타 이유로 죽은 군인의 수를 표시하였습니다.

기준에 따라 입원, 부상, 질병, 사망 등의 내역을 매일 상세히 작성하고 이를 토대로 영국 정부에 야전병원의 위생을 개선해야 한다고 주장하였습니다.

개선 사업을 시작한지 육 개월이 지나자 42%에 달하던 환자의 사망률이 2%까지 떨어졌습니다. 당시 사람들은 질병의 원인이 무엇이며, 깨끗해지면 왜 사망자가 줄어드는지 알지 못 했습니다. 하지만 나이팅게일이 제시한 통계는 깨끗한 위생이 사람을 살린다는 증거가 되었습니다.

○ 대준 : 그런데 사망자가 수의 차이를 잘 모르겠어. 부채꼴의 넓이가 사망자수라고 했는데 반지름의 길이를 사망자 수로 하면 사망자 수의 차이가 잘 나타나겠는데...

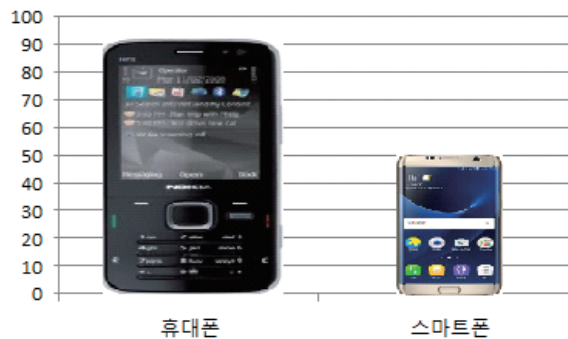
5) <http://blog.daum.net/mathkorea/15811174>

Activity 1 그래프의 착시와 선택

♣ 정보통신정책연구원(KISDI)에서 공개한 '2016한국미디어패널조사'결과입니다.

년도	2011년	2012년	2013년	2014년	2015년	2016년
휴대폰 보급률(%)	89.2	91.2	91.8	92.4	93.2	94.0
스마트폰 보급률(%)	21.6	49.3	66.2	73.4	77.6	81.6

2012년 휴대폰과 스마트폰의 보급률은 각각 91.2%와 49.3%로 휴대폰의 보급률이 스마트폰의 보급률의 1.85배입니다. 이것을 그림그래프로 나타내면 다음과 같습니다.

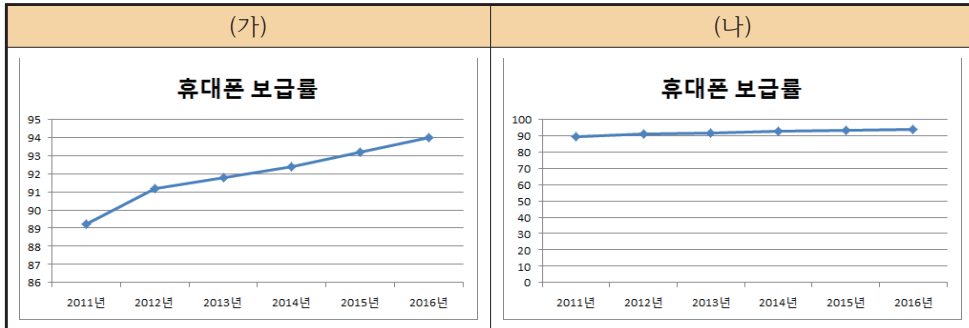


○ 대준 : 막대 대신에 그림으로 나타내니까 눈에 더 잘 들어오는 걸^^ 그런데 휴대폰 보급률이 스마트폰 보급률의 약 2배인데.. 왜 훨씬 더 차이가 나는 것으로 느껴질까?

(1) 대준이가 실제보다 보급률의 차이를 더 크게 느끼는 이유는 무엇인가요?

휴대폰의 보급률이 스마트폰의 보급률보다 1.85배 높다는 것을 나타내기 위해서는 막대의 폭이 같고 높이만 1.85배 높아야 하는데 폭도 1.85배 늘려서 실제로는 3.4배 정도 높은 것처럼 느껴진다.

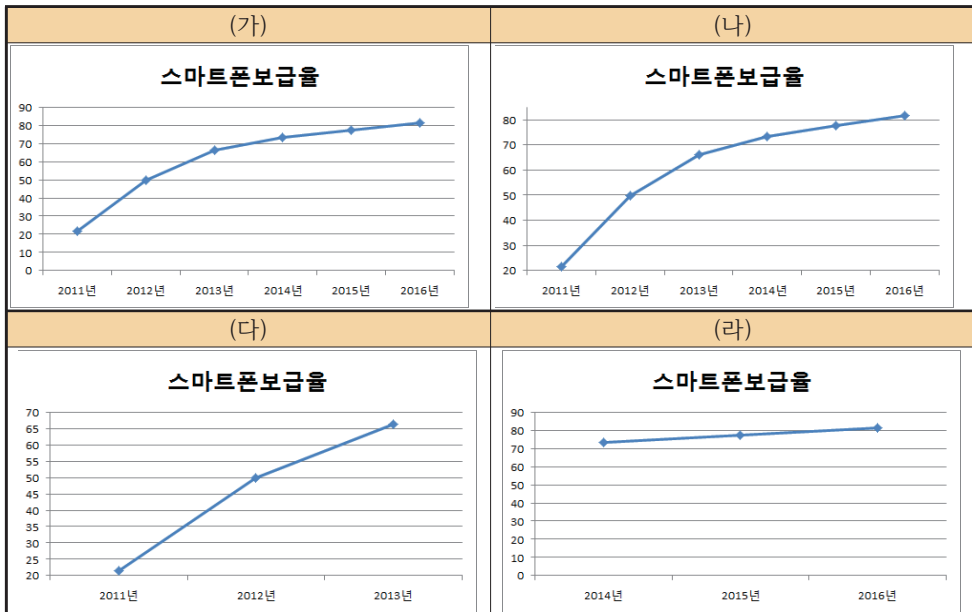
대준 : 정말 그래프의 착시를 조심해야겠는걸. 또 그래프를 볼 때 조심해야 할 건 없니?
 성욱 : 그럼 다음 그래프를 한 번 볼까? 년도에 따른 휴대폰보급률을 나타낸 꺾은선그래프야.
 대준 : 같은 자료를 나타낸 그래프인데 두 그래프에서 느껴지는 보급률이 전혀 다르게 보이네.
 성욱 : 그런데 왜 이런 차이가 생기는 걸까?



(2) 대준이가 두 그래프에서 휴대폰 보급률을 다르게 느끼는 이유는 무엇인가요?

(나)는 세로축을 0에서 100으로 나타내었지만 (가)는 세로축을 86에서 95로 나타내어 (가)가 더 급격히 증가하는 것처럼 느껴진다.

○ 성욱 : (가)에서 (라)는 년도에 따른 스마트폰 보급률을 나타낸 그래프야. 우리가 특집 기사를 어떤 방향으로 쓸지 결정하느냐에 따라 사용하는 그래프의 종류가 달라지지. 대준아~! 우리의 방향에 따라 어떤 그래프가 좋을지 생각해 볼까?

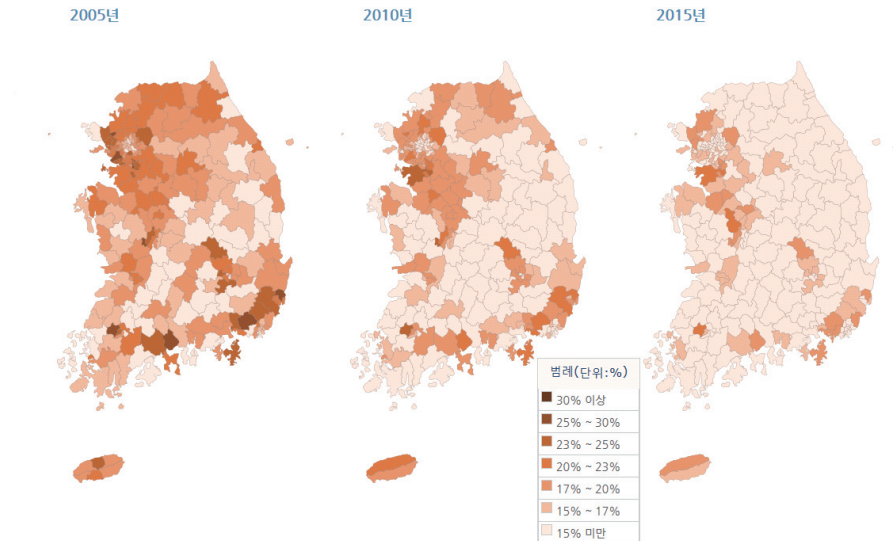


(3) 기사에의 방향에 따라 적합한 그래프를 쓰고 적합한 이유를 써봅시다.

기사의 방향	적합한 그래프	이유
자료의 일부를 이용하여 스마트폰의 인기가 주춤한다는 것을 설명하고 싶을 때	(라)	최근 3년간의 자료만 사용하여 스마트폰 보급률이 어느 정도 높아진 이후의 상황만 나타내어 상대적으로 스마트폰 보급률이 천천히 증가하는 것으로 보인다.
자료의 일부를 이용하여 스마트폰의 보급률이 급격히 증가한다는 것을 설명하고 싶을 때	(다)	처음 3년간의 자료만 사용하여 스마트폰 보급률이 가장 낮을 때를 시작점으로 잡아 상대적으로 스마트폰 보급률이 급격히 증가하는 것으로 보인다.
전체 자료를 이용하여 스마트폰 보급률이 천천히 변하는 것을 설명하고 싶을 때	(가)	조사된 전체 기간의 자료를 나타내었으며 세로축을 0에서 90까지로 나타내어 상대적으로 천천히 증가하는 것처럼 보인다.
전체 자료를 이용하여 스마트폰 보급률이 급격히 변하는 것을 설명하고 싶을 때	(나)	조사된 전체 기간의 자료를 나타내었으며 세로축을 20에서 80까지로 나타내어 상대적으로 급격히 증가하는 것처럼 보인다.

♣ 읽을거리 : 시간에 따른 변화를 나타내는 그래프의 예

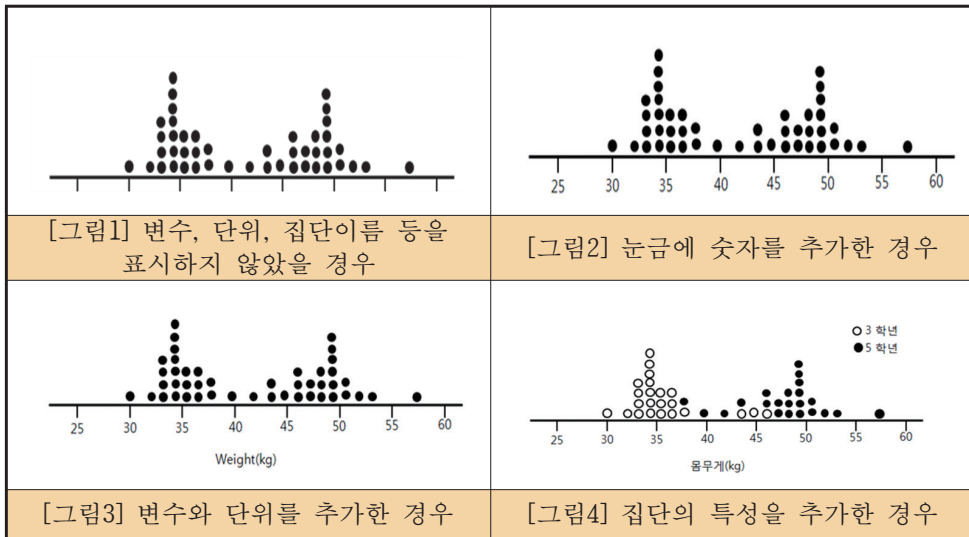
다음은 통계청에서 조사한 ‘15세 미만 인구의 비율’을 지도그래프를 이용하여 나타낸 것입니다. 흔히 시간에 따른 변화를 나타낼 때는 꺾은선그래프를 사용해야 한다는 생각이 있는데 가장 편하고 효율적인 방법이지 한 가지 방법밖에 없다는 것은 아닙니다.



Activity 2 효과적인 그래프

그래프를 잘 보면 자료 속에 숨은 정보를 효과적으로 알 수 있습니다. 하지만 일반적인 그래프에서도 변수, 단위, 집단이름 등을 표현하느냐 하지 않느냐에 따라 자료를 설명할 수 있는 정도가 달라집니다.

다음 점도표를 통해 변수, 단위, 집단이름 등 표현의 필요성을 살펴보도록 합시다.



○ 진혁 : [그림1]은 범례를 하나도 표시하지 않은 점도표야. 이 점도표에서는 자료가 두 곳을 중심으로 모여 있다는 것을 알 수 있어. 대준아! 여러 가지 표현을 추가하면서 더 알 수 있는 것을 찾아볼까?

(1) 대준이가 [그림1]의 점도표에서 범례를 하나씩 추가함으로써 새롭게 알 수 있는 정보는 무엇일까요?

구분	새로 알 수 있는 정보
[그림2] 눈금에 숫자를 추가하는 경우	눈금을 추가함으로써 35와 50을 중심으로 자료가 모여 있다는 것을 알 수 있다
[그림3] 변수명과 단위를 추가하는 경우	weight를 가로축에 표시함으로써 무게를 kg으로 나타낸다는 것을 알게 되었고 자료의 중심이 35kg과 50kg이라는 것을 알 수 있었다.
[그림4] 집단의 특성을 추가하는 경우	학년을 추가함으로써 3학년은 35kg, 5학년은 50kg을 중심으로 모여 있다는 것을 알게 되었다. 그리고 학년별로 구분할 때 자료의 중심은 1개이다.



활동지 2

스마트폰으로부터 학생을 지켜라~!



- 진혁 : 그럼 이제 우리 학교 학생들의 ‘스마트폰 중독 실태’를 정리해 볼까? 그래프로 정리하면 좋겠지
- 대준 : 성별, 학년, 스마트폰 사용 시간 등 여러 가지를 조사했는데 어떤 그래프를 사용해야 할지 모르겠네. 일단 표로 정리해야 그래프로 나타낼 수 있으니깐 표로 정리해 보자.

♣ 다음은 ○○ 중학교 신문부원들이 조사한 ‘스마트폰 중독 실태’를 정리한 표입니다.

구분	도수(명)
남자	22
여자	18
합계	40

<응답자 성별 빈도표>

구분	도수(명)
1학년	15
2학년	14
3학년	11
합계	40

<응답자 학년 빈도표>

구분	도수(명)
없다	2
있다	38
합계	40

<스마트폰 소유여부 빈도표>

스마트폰 사용 시간	도수(명)
~ 60분 미만	13
60분 이상 ~ 120분 미만	9
120분 이상 ~ 180분 미만	11
180분 이상 ~ 240분 미만	1
240분 이상 ~ 300분 미만	1
300분 이상 ~ 360분 미만	3
360분 이상 ~	2
합계	40

<스마트폰 사용시간 도수분포표>

구분	도수(명)
동영상 시청	8
게임	12
SNS	14
정보검색	5
뉴스검색	1
합계	40

<스마트폰 주사용목적 빈도표>

자가진단점수	도수(명)
10점 이상 ~ 20점 미만	7
20점 이상 ~ 30점 미만	9
30점 이상 ~ 40점 미만	12
40점 이상 ~ 50점 미만	9
50점 이상 ~ 60점 미만	3
합계	40

<자가진단점수 도수분포표>

구분	도수(명)
일반 사용군	30
잠재적 위험군	6
고위험군	4
합계	40

<스마트폰 사용자군 빈도표>

- 대준 : 수집한 자료를 표로 정리하니깐 한 눈에 알아볼 수 있는 걸
- 성욱 : 그렇지, 표나 그래프로 나타내면 자료의 특징을 한 눈에 알아볼 수 있지. 대신 표가 그래프보다는 더 정확하게 받아들이 수 있어.
- 진혁 : 그런데 표도 그래프처럼 축의 크기 같은 걸 확인해야 할까?

- (1) 스마트폰 사용 시간을 계급의 크기가 5시간인 도수분포표로 정리하여 봅시다. 계급의 크기가 1시간일 때와는 어떤 점이 달라지나요? 또, 계급의 크기가 10분 이라면 어떤 현상이 일어날지 생각해 볼까요?

스마트폰 사용 시간	도수(명)
~ 5시간 미만	
5시간 이상~ 10시간 미만	
합계	

<스마트폰 사용시간 도수분포표>

스마트폰 사용 시간의 계급의 크기를 5시간으로 한다면 5시간 미만, 5시간 이상의 두 구간으로 나누어지며 도수가 35명, 5명으로 나타나 두 집단 간에는 차이가 크게 나타나지만 1시간처럼 구체적인 분포는 알 수 없습니다. 계급의 크기를 10분으로 한다면 38개의 계급이 필요하며 전체 도수가 40명이기 때문에 어떤 곳에 집중되어 있는지를 정확히 구분할 수 없습니다.

- 대준 : 그런데 표는 자료의 특성에 따라 정리할 수 없을까? 나이팅게일의 그래프에서는 사망 원인에 따라 색깔을 구분했잖아
- 성욱 : 표에서도 구분할 수 있어. 스마트폰의 주된 사용 목적을 남학생과 여학생으로 구분하여 정리해볼까?
- 진혁 : 스마트폰 주된 사용 목적을 성별에 따라 정리해도 앞서 정리한 표와 같은 결과가 나올까?

- (2) 스마트폰의 주된 사용 목적을 성별에 따라 정리한 것입니다. 이전 자료에서 스마트폰의 주된 사용 목적을 성별 구분 없이 나타낸 자료의 결과와 같나요?

구분	도수(명)	
	남학생	여학생
동영상 시청	5	3
게임	9	3
SNS	5	9
정보검색	2	3
뉴스검색	1	0
합계	22	18

성별을 구분하지 않았을 때는 SNS, 게임, 동영상 시청, 자료검색, 뉴스검색의 순으로 나타납니다. 그런데 성별을 구분하여 정리한다면 여학생의 자료에서는 뉴스검색이 없다는 것을 알 수 있습니다. 또, 남학생은 게임이 가장 많으며 여학생은 SNS가 가장 많이 선택한 것을 알 수 있습니다. 이를 통해 성별에 따른 스마트폰의 주된 사용 목적에 차이가 있음을 추측할 수 있습니다.

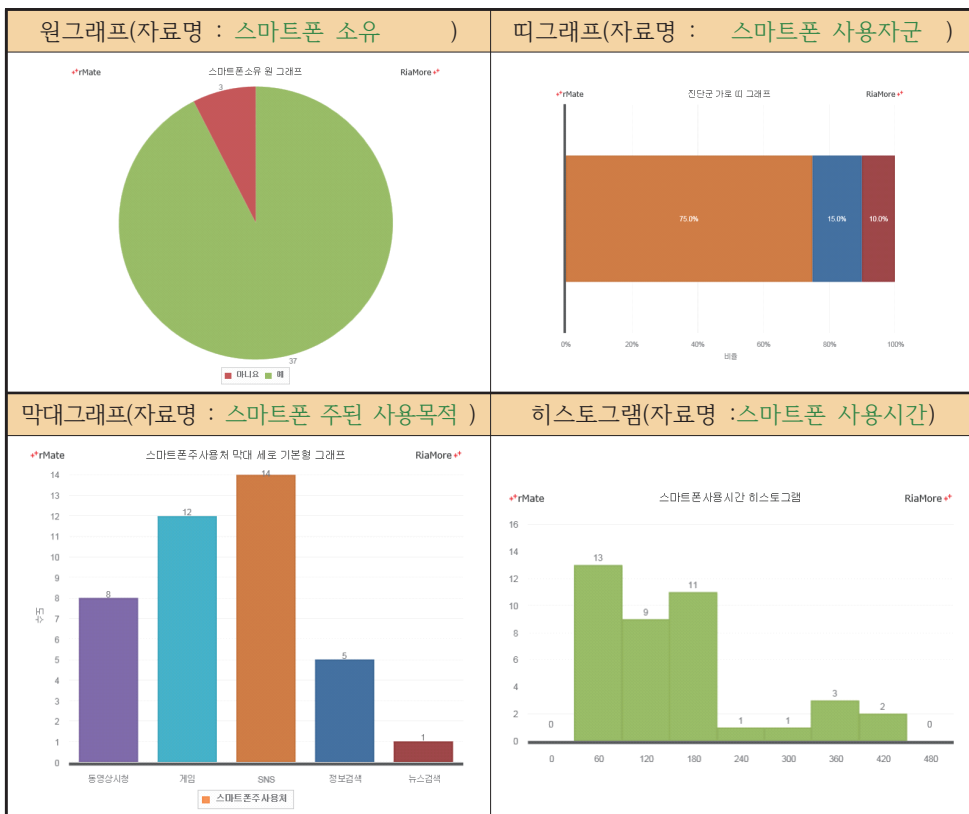


Activity 1 정리한 표를 그래프로 나타내기

- (1) ○○ 중학교 신문부원들이 정리한 표를 원그래프, 띠그래프, 막대그래프, 히스토그램 중 어떤 것을 사용하여 나타내는 것이 가장 좋을지 생각해보고 그 이유를 적어봅시다.

구분	가장 적합한 그래프	이유
스마트폰 소유	원그래프	예, 아니오의 2가지 특성을 표현하는 명목형 자료는 원그래프가 가장 좋습니다.
스마트폰 사용시간	히스토그램	연속형자료는 막대그래프가 아닌 히스토그램으로 표현합니다.
스마트폰 주사용목적	막대그래프	세부항목이 4개 이상인 경우의 범주형자료는 막대그래프를 이용하는 것이 좋습니다.
스마트폰 사용자군	원그래프, 띠그래프	사용군은 명목형자료로 원그래프나 띠그래프 중 어떤 것으로 나타내어도 무방합니다.

- (2) 앞에서 선택한 그래프를 이용하여 설문의 결과를 나타내어 봅시다.



통계적 소양을 위한 교수 학습자료

다음은 가장 많이 사용되는 그래프와 특징에 대한 설명입니다. 이를 바탕으로 앞에서 만든 그래프가 자료를 나타내는데 적절하였는지 토의해 보도록 합니다.

♣ 참고자료 : 자료의 종류	
자료는 크게 측정형자료와 명목형자료로 구분할 수 있습니다. 측정형자료는 연속형자료와 이산형자료로 구분됩니다. 키와 몸무게와 같은 자료는 연속형자료이고 가족수, 다녀온 횟수 등과 같은 자료는 이산형자료입니다. 범주형 자료는 설문지에서 가장 많이 보는 매우불만, 불만, 보통, 만족, 매우만족 등의 순위형자료와 혈액형, 성별과 같은 명목형자료로 나뉩니다. 자료 분석에서는 명목형자료는 질적변수, 나머지 자료는 양적변수라고 하기도 합니다.	
원그래프	띠그래프
원그래프는 원의 크기에 따라 각 범주의 비율을 나타냅니다. 토마토가 가장 큰 비율을 차지하고 있습니다.	띠그래프는 수직 막대를 사용하여 범주별 비율을 나타냅니다. 0%에서 100%까지의 구간을 표시하고 있습니다.
원을 변숫값의 도수에 비례하여 여러 조각으로 나누어 나타낸 그래프. 범주형 자료의 분석에 적합	직사각형을 변숫값의 도수에 비례하여 여러 조각으로 나누어 나타낸 그래프. 범주형 자료의 분석에 적합함. 원그래프보다 객관적 비교가 가능
막대그래프	히스토그램
막대그래프는 수직 막대의 높이를 사용하여 범주별 도수를 나타냅니다. 토마토의 도수가 가장 높습니다.	히스토그램은 연속형 자료의 분포를 나타내며, 막대의 폭과 높이를 사용하여 빈도와 밀도를 표현합니다.
범주형 자료의 도수분포를 한 눈에 비교할 수 있도록 각 자료의 값의 도수(또는 상대도수)를 같은 폭의 막대로 나타낸 것. 범주형 자료의 분석에 적합. 연속형 자료는 히스토그램으로 표현	키나 몸무게와 같은 연속형 자료는 값의 종류가 너무 많아 점그래프로 나타낼 수 없으므로 몇 개의 구간으로 나누어 (막대 사이의 간격이 없는) 막대그래프의 형태로 그린 그림
꺾은선그래프	상관도
꺾은선그래프는 시간에 따른 변숫값의 변화를 선으로 연결하여 나타냅니다. 남녀성비의 추이를 보여줍니다.	상관도는 두 연속형 변수 간의 관계를 점으로 나타내어 그 패턴을 분석하는 그래프입니다. 키와 몸무게의 양의 상관관계를 보여줍니다.
시간변수를 가로축, 분석변수를 세로축으로 하여 모든 분석변수의 값을 시간에 따라 선으로 이어주는 그래프	한 변수를 가로축(원인), 다른 변수를 세로축(결과)으로 하여 각각의 관찰값을 좌표값으로 하여 나타낸 그래프. 두 개의 연속형 자료의 관계를 분석하는데 효과적임.



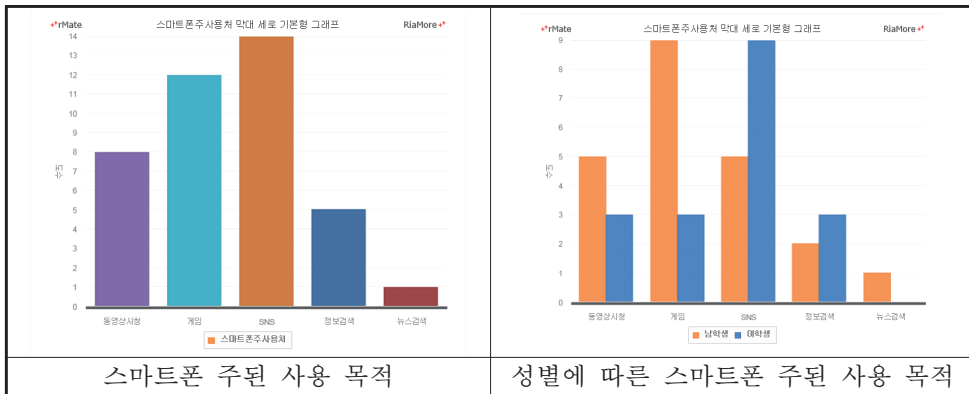
활동지 3

도전! Treasure Hunt(보물 사냥꾼)

○○중학교 학생회에서 조사한 자료를 표와 그래프를 이용하여 나타내어 보았습니다. 그런데 앞에서 살펴본 내용 이외에 더 찾아낼 수 있는 정보가 있을까요? 이번 활동에서는 수집한 자료 속에 숨어 있는 보물(정보)을 찾아보도록 합니다.



Activity 1 성별에 따른 자료 분석하기(1)



성별에 따라 스마트폰의 주된 사용 목적을 정리하면 남학생은 게임을 가장 많이 하고 동영상상시청과 SNS는 공동 2등, 그 다음이 정보검색, 뉴스검색이란 것을 알 수 있습니다. 반면 여학생은 SNS가 가장 많이 나타났으며 동영상상시청과 게임, 정보검색이 공동 2등을 한 것을 알 수 있습니다. 그리고 남학생과는 달리 뉴스검색은 아무도 선택하지 않았습니다.

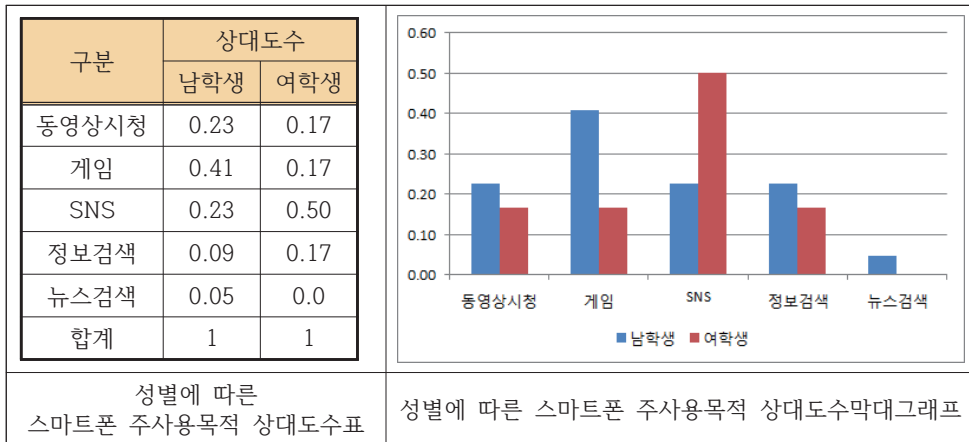
이렇게 성별에 따라 스마트폰의 주된 사용 목적을 정리해보니 전체와는 다른 해석이 가능합니다. 학년, 스마트폰 소유 여부, 스마트폰 사용자군에 따라 스마트폰의 주된 사용 목적을 정리해 본다면 또 다른 정보를 얻을 수 있을 것입니다.

- 대준 : 성별로 구분하니깐 스마트폰의 주된 사용 목적이 달라지는 걸 알 수 있었어. 그리고 남학생은 게임이 9명, 여학생은 SNS가 9명으로 가장 높게 나타났어. 가장 많이 응답한 9명으로 같아
- 진혁 : 9명으로 같긴 하지만 그 의미는 다르지 않을까?

(1) 남학생과 여학생에서 9명의 의미를 생각해 봅시다. 둘은 같은 의미일까요?

스마트폰의 주된 사용 목적이라는 질문에 남학생은 게임에 9명, 여학생은 SNS에 9명이 대답하고 있습니다. 하지만 남학생은 22명, 여학생은 18명으로 도수의 합계가 다릅니다. 같은 9명이더라도 남학생은 41%, 여학생은 50%에 해당하는 값입니다.

성별에 따른 스마트폰의 주된 사용처의 상대도수와 상대도수를 이용한 막대그래프를 나타내 보았습니다.



(2) 상대도수를 이용하여 성별에 따른 스마트폰의 주된 사용 목적을 설명하여 봅시다.

◦ 남학생

남학생은 스마트폰의 주된 사용 목적을 묻는 문항에 게임을 9명이 선택하여 가장 많이 대답하였고 그 다음으로는 동영상 시청, SNS이 같은 인원이 5명이 선택하였으며 정보검색(2명), 뉴스검색(1명)의 순으로 나타났다.

◦ 여학생

여학생은 스마트폰의 주된 사용 목적을 묻는 문항에 9명이 SNS를 선택하였으며 동영상시청, 게임, 정보검색인 모두 3명으로 나타났다. 남학생과 달리 뉴스검색을 선택한 학생은 없었다.

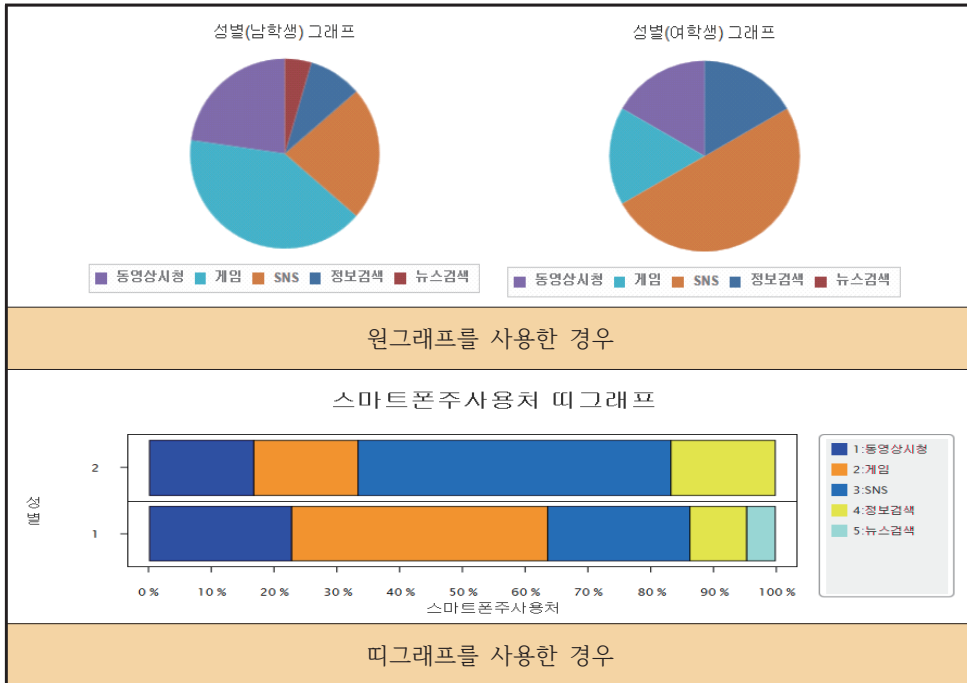
◦ 대준 : 그림 도수와 상대도수는 각각 어떤 경우에 사용하는 게 좋을까?

(3) 도수 혹은 상대도수를 사용하느냐에 따라 자료의 해석이 달라집니다.

상대도수는 어떤 경우에 사용하는 것이 좋을까요?

상대도수는 앞의 사례와 같이 조사한 집단별로 전체 인원이 다를 경우 사용하는 것이 좋다. 예를 들어 중국의 인구가 15억명, 대한민국은 5천만명이라고 할 때 백만장자의 비율을 각각 2%와 5%라고 한다면 중국은 백만장자가 3천만명, 대한민국은 250만명이 된다. 백만장자의 수를 따지면 중국이 대한민국의 12배가 되지만 자국 내 인구 중 백만장자의 비율은 대한민국이 중국의 2.5배이다. 이렇게 집단의 도수 차이가 클수록 상대도수를 사용하는 것이 더 정확한 비교를 할 수 있다.

남녀에 따른 스마트폰의 주사용처를 원그래프와 띠그래프를 이용하여 나타낸 것입니다.

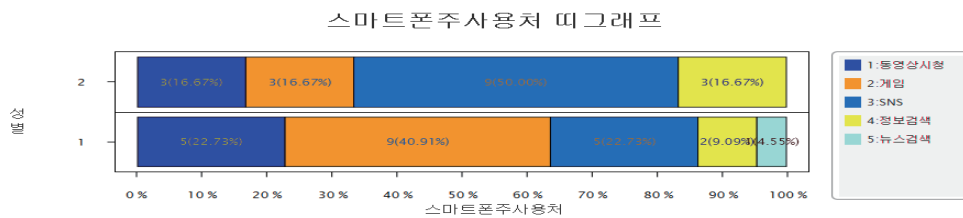


- (4) 성별에 따른 스마트폰 주된 사용목적은 어떤 그래프를 사용하는 것이 좋을까요?
그 이유는 무엇인가요?

띠그래프

원그래프는 비율을 예쁘게 나타낼 수 있지만 비율이 비슷할 때 육안으로 정확하게 구분할 수 없는 단점을 가지고 있다. 넓이가 같은 경우에도 어떤 색깔을 칠하느냐에 따라 다르게 받아들일 수도 있다. 또, 집단의 수가 많을수록 정확하게 받아들이기가 어렵다.

- (5) 다음과 같이 상대도수를 그래프에 나타내면 그래프를 해석하는데 도움이 될까요?

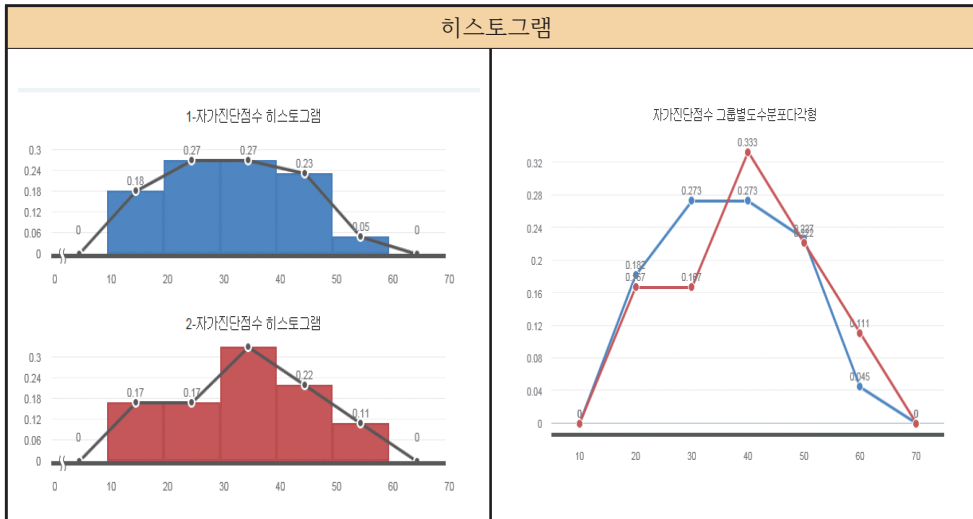


띠그래프를 이용하면 원그래프보다 정확한 비교가 가능하다. 여기에 도수 혹은 상대도수까지 나타낸다면 시각적으로 정확히 구분하기 힘든 경우에도 정확한 비교를 할 수 있다.

Activity 2 성별에 따른 자료 분석하기(2)

- (1) 앞선 활동을 참고하여 성별에 따른 자가진단점수를 가장 잘 설명할 수 있는 그래프가 무엇인지 생각해보고, 다음의 그래프 중 하나를 사용하여 나타내어 볼까요?

- 가장 적합한 그래프 : 상대도수히스토그램 (상대도수의 분포다각형 모양)
○ 이유 : 성별에 따른 인원수가 다르기 때문에 도수보다는 상대도수를 사용하는 것이 적합하며 자가진단점수는 변량(연속형 변수)이기 때문에 히스토그램을 사용하는 것이 적합하다.



Tip

- ♣ 막대그래프와 히스토그램은 같은 양식을 사용하여 정리할 수 있도록 만들었습니다. 또 이 양식은 상관도(산점도)를 나타낼 때도 사용할 수 있습니다.
♣ 성별에 따라 두 개의 그래프로 나타내어야 하므로 학생 활동 전에 막대그래프 혹은 히스토그램은 이름을 바꾸어 사용하도록 안내합니다.

○ 대준 : 남학생과 여학생의 자가진단점수는 어떤 차이가 있다고 기사로 정리하면 좋을까?

- (2) 정리된 그래프를 통해 알 수 있는 남학생과 여학생의 자가진단점수의 특성을 적어볼까요?

전체적으로 볼 때 여학생이 남학생보다 자가진단점수가 높은 것을 알 수 있다. 남학생은 10점 이상 30점 미만, 40점 이상~50점 미만에서 여학생보다 비율이 높게 나타났으며, 30점 이상 40점 미만, 50점 이상 60미만 구간에서는 여학생이 남학생의 비율보다 높게 나타났다.

♣ 참고자료 : 설문지 양식 및 자료입력 양식, 스마트폰 사용 서약서

○○중학교 스마트폰 사용 실태 조사

학생회에서는 우리 학교 학생들의 스마트폰 사용 실태와 스마트폰 중독 실태를 조사하고자 합니다. 조사 결과는 학생회에서 자료 분석에만 사용되며 선생님이나 외부로 유출되지 않습니다. 학생 여러분의 솔직한 대답 부탁드립니다.

○○중학교 학생회

- 당신의 성별은? ① 남 ② 여
- 당신의 학년은? ① 1학년 ② 2학년 ③ 3학년
- 당신은 스마트폰을 가지고 있나요? ① 아니요 ② 예
- 하루 평균 스마트폰을 사용하는 시간은 몇 분인가요? ()시간 ()분
- 스마트폰을 이용해서 가장 많이 하는 것은 무엇인가요?
 ① 동영상 시청(인강 제외) ② 게임 ③ SNS ④ 정보검색(인강 포함) ⑤ 뉴스검색



- ♣ 매 학기마다 한 번씩 의무적으로 인터넷 및 스마트폰 사용실태 조사를 실시하도록 되어 있습니다. 형식적인 조사를 할 수도 있지만 이 자료를 이용하여 **계기교육 및 교과 수업에 활용**할 수 있다면 일석이조일 것입니다.
- ♣ 스마트폰 중독 자가진단 검사를 실시할 때는 **자가진단 계산 방법 및 진단 군에 대한 부분은 제외**하고 조사하도록 합니다.
- ♣ 활동을 통해 만들어진 표와 그래프를 이용하여 **신문기사, 통계포스터** 등을 만들 수 있습니다.

스마트폰 중독 자가진단 검사지

번호	항목	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
1	스마트폰의 지나친 사용으로 학교 성적이 떨어진다				
2	가족이나 친구들과 함께 있는 것보다 스마트폰을 사용하고 있는 것이 더 즐겁다				
3	스마트폰을 사용할 수 없게 된다면 견디기 힘들 것이다				
4	스마트폰 사용시간을 줄이려고 해보았지만 실패한다				
5	스마트폰 사용으로 계획한 일(공부, 숙제 또는 학원 수강 등)을 하기 어렵다				
6	스마트폰을 사용하지 못하면 온 세상을 잃은 것 같은 생각이 든다				
7	스마트폰이 없으면 안전부절못하고 초조해진다				
8	스마트폰 사용시간을 스스로 조절할 수 있다				
9	수시로 스마트폰을 사용하다가 지적받은 적이 있다				
10	스마트폰이 없어도 불안하지 않다				
11	스마트폰을 사용할 때 그만해야지 하고 생각은 하면서도 계속한다				
12	스마트폰을 너무 자주 또는 오래한다고 가족이나 친구들과로부터 불평을 들은 적이 있다				
13	스마트폰 사용이 지금 하고 있는 공부에 방해가 되지 않는다				
14	스마트폰을 사용할 수 없을 때 패닉 상태에 빠진다				
15	스마트폰 사용에 많은 시간을 보내는 것이 습관화되었다				

스마트폰 자가 진단 점수 계산

- 전혀 그렇지 않다 : 1점, 그렇지 않다 : 2점, 그렇다 : 3점, 매우 그렇다 4점
- 단 8번, 10번, 13번은 역계산
전혀 그렇지 않다 : 4점, 그렇지 않다 : 3점, 그렇다 : 2점, 매우 그렇다 1점
- 일반 사용자군(1) : 총점 41점 이하, 잠재적 위험 사용자군(2) : 총점 42~44점, 고위험 사용자군(3) : 45점 이상

45명의 대상 학생 중 40명이 조사에 응해 주었으며 그 결과는 다음과 같습니다.

구분	성별	학년	스마트폰소유	사용시간	주된목적	자가진단점수	사용군
1	1	1	1	10	4	15	1
2	1	1	2	30	1	19	1
3	1	1	2	190	2	42	2
4	1	1	2	30	2	24	1
5	1	1	2	150	1	34	1
6	1	1	2	20	3	24	1
7	1	1	2	40	1	21	1
8	1	1	2	70	2	51	3
9	1	1	2	20	2	15	1
10	1	2	1	10	3	15	1
11	1	2	2	140	1	44	2
12	1	2	2	130	2	34	1
13	1	2	2	110	2	28	1
14	1	2	2	140	3	39	1
15	1	2	2	30	4	24	1
16	1	2	2	60	3	26	1
17	1	2	2	120	1	35	1
18	1	2	2	110	3	40	1
19	1	3	2	130	2	42	2
20	1	3	2	140	5	41	1
21	1	3	2	110	2	34	1
22	1	3	2	80	2	35	1
23	2	1	1	50	3	26	1
24	2	1	2	70	3	34	1
25	2	1	2	300	3	43	3
26	2	1	2	30	4	19	1
27	2	1	2	100	2	25	1
28	2	1	2	50	1	17	1
29	2	2	2	50	1	33	1
30	2	2	2	130	2	42	2
31	2	2	2	110	3	35	1
32	2	2	2	340	3	44	2
33	2	2	2	10	4	16	1
34	2	3	2	370	3	55	3
35	2	3	2	150	2	28	1
36	2	3	2	170	4	38	1
37	2	3	2	160	1	32	1
38	2	3	2	360	3	54	3
39	2	3	2	310	3	42	2
40	2	3	2	260	3	39	1



Tip

- ♣ 학생들에게 직접 조사한 자료를 사용하면 학생들이 수업에 더 잘 집중할 수 있습니다.
- ♣ 자료의 입력은 번호순이 아니라 무작위로 입력하셔야 개인 정보가 보호됩니다.

행복기록장 [스마트폰(인터넷) 사용 서약서]

※ 우리 가족은 스마트폰(인터넷)을 효율적으로 사용하고, 절제할 수 있는 습관을 기르며 행복한 생활을 하기 위해 일주일 동안 다음과 같이 약속을 지키도록 하겠습니다. 가족이 함께 사용시간을 정하고, 그에 따라 준수여부를 기록한 후 담임선생님께 제출하세요.

〈 자녀의 약속 〉

준 수 사 항	준 수 여 부 (○ 잘함 , △ 보통 , × 노력)				
	월	화	수	목	금
1. 나는 스마트폰을 잠시 내려놓고 하루에 ()시간 이상 가족과 대화를 하도록 하겠습니다.					
2. 나는 스마트폰(인터넷) 사용시간을 하루에 ()시간 만큼만 사용하겠습니다.					
3. 나는 식사시간 및 잠들기 1시간 전에는 스마트폰(인터넷)을 사용하지 않겠습니다.					
4. 나는 스마트폰(인터넷)을 가급적 정보검색 및 교육용으로 활용할 수 있도록 노력하겠습니다.					
5. 나는 훌륭한 사이버 시민이므로 SNS, 게임, 웹툰 등 다양한 콘텐츠를 통하여 다른 사람에게 상처를 주지 않겠습니다.					

〈 부모의 약속 〉

준 수 사 항	준수여부 (○ 잘함 , △ 보통 , × 노력)				
	월	화	수	목	금
1. 나는 스마트폰을 잠시 내려놓고 하루에 ()시간 이상 가족과 대화를 하도록 하겠습니다.					
2. 나는 스마트폰(인터넷) 사용시간을 하루에 ()시간 만큼만 사용하겠습니다.					
3. 나는 식사시간 및 잠들기 1시간 전에는 스마트폰(인터넷)을 사용하지 않겠습니다.					
4. 나는 스마트폰(인터넷)을 가급적 정보검색 및 교육용으로 활용할 수 있도록 노력하겠습니다.					
5. 나는 우리 아이가 약속한 시간동안 스마트폰(인터넷)을 이용할 때에는 잔소리를 하지 않겠습니다.					

※ 사용시간은 자녀와 부모가 함께 협의하여 결정합니다.

자녀대표 _____ (인) 부모대표 _____ (인)

제 4 장 대푯값

제 4 장 대푯값

제 1 절 개발 의도

일반적으로 ‘산술평균’을 의미하는 평균은 방대한 데이터를 단 하나의 간략한 수로 표현하기 때문에 일상적으로 많이 사용된다. 그러나 일상생활에서 사용하는 “평균적으로”라는 표현은 때로 또 다른 대푯값들 중 하나인 중앙값, 최빈값과 구분되지 않고 사용되기도 한다.

중앙값은 자료를 크기순으로 나열하였을 때 가운데에 위치하는 값을 말하며, 최빈값은 자료의 값 중에서 가장 많이 나타나는 값을 말한다. 중학교 학생들의 수준에서 이러한 다양한 대푯값들의 용어와 의미를 명확하게 이해하고 적절하게 사용할 수 있도록 이번 교수-학습 자료를 설계하였다.

첫 번째, ‘통계에도 오류가 있다?’ 활동을 통해 학생들은 자료를 대표하는 하나의 수로서의 평균을 구할 수 있다. 그러나 두 가지 서로 다른 자료집합에서의 평균을 비교하고 대조함으로써 평균이 설명할 수 없는 자료의 분포를 눈여겨보게 될 기회를 가질 수 있다. 이 때 학생들의 활동에서 주목해야 할 점은 평균을 계산하는 것뿐만 아니라 평균이 자료를 대표하는 값으로 적절한지 자료와 함께 비교함으로써 해석해 내는 것이다. 그러므로 평균을 계산할 때도 공학 도구를 활용하는 것을 권장한다.

예를 들어 교사는 학생들이 이지통계를 활용하여 두 자료 집합의 히스토그램을 그리도록 하고 이를 비교하도록 안내할 수 있다. 또한 교사는 히스토그램을 충분히 이해한 상위 수준의 학생들에게 두 자료집합의 상자그림을 제시하여 평균과 자료집합의 분포를 시각적으로 비교해 보는 기회를 제공할 수도 있다. 여기에 더해 Activity2에서 주어진 자료를 대표하는 대푯값을 말하고 설명하게 함으로써 주어진 자료에 따라 적절한 대푯값이 달라질 수 있으며, 서로 다른 대푯값들을 함께 이용하여 자료를 설명할 수도 있음을 안내할 수 있다.

두 번째, ‘균형점으로써의 평균’ 활동을 통해 학생들은 평균의 또 다른 의미에 주목할 수 있다. 특히 평균을 계산하는 것과 히스토그램과 같은 자료의 그래프를 연결함으로써 균형점으로써의 평균의 의미에 주목할 수 있다.

앞선 활동에서 자료의 분포를 주목하게 하는 것이 주목적이라면 이번 활동에서는 분포의 범위, 최댓값, 최솟값, 극단값(자료들 중 지나치게 크거나 작은 값) 등을 함께 살펴보게 함으로써 표준편차와 같은 이후 학습하게 될 활동의 기초를 마련할 수 있다.

이 때 도수분포표에서 계급값을 이용하여 평균을 계산하는 것보다는 자료의 균형점으로서 평균의 아이디어에 좀 더 주목할 수 있도록 자료를 히스토그램으로 나타낸 다음 이를 오려낸 후 연필 위에 올려보게 하는 활동을 구상하였다. 학생들은 연필 위에 히스토그램을 올려 균형을 잡는 활동으로 균형점으로써의 평균의 의미를 직관적으로 이해할 수 있다. 특히 대칭적인 그래프로 표현되는 자료와 비대칭적인 그래프로 표현되는 자료에서 평균의 위치를 서로 비교해 봄으로써 분포를 나타내는 척도로써 중앙값과의 차이를 눈여겨 볼 수도 있다.

제 2 절 지도를 위한 배경

가. 평균

일반적으로 전체 자료의 값을 모두 더한 다음 전체 자료의 사례 수로 나눈 값인 산술평균(arithmetic mean)을 의미한다. 산술평균은 무게중심과 같은 평형점(balance point)의 의미를 갖는다. 즉, 평균을 중심으로 개별 자료 값으로부터 편차점수를 구하면, 그 편차점수의 합은 '0'이 된다. 평균은 산술평균과 동의어로 사용되기도 하지만 기하평균(geometric mean)이나 조화평균(harmonic mean)을 포함하는 넓은 의미로 사용되기도 한다⁶⁾.

나. 중앙값

자료를 크기 순서대로 배열했을 때, 중앙에 위치하게 되는 값이다. 중앙값을 기준으로 자료의 반은 중앙값보다 큰 값을 갖고, 나머지 반은 중앙값보다 작은 값을 갖는다. 자료가 (4, 5, 7, 8, 10)으로 구성되어 있다면, 7보다 작은 값이 2개, 7보다 큰 값이 2개 있으므로 7이 중앙값이 된다. 자료가 (1, 3, 5, 7, 8, 9)와 같이 짝수로 구성되어 있는 경우, 가운데 있는 두 값인 5와 7의 평균으로 중앙값이 결정된다. 자료에 중복된 값이 있는 경우 중앙값 계산은 복잡해지며, 이 경우는 누적 백분율이 50%인 점을 계산하는 공식을 활용하여 중앙값을 산출할 수 있다.⁷⁾

다. 최빈값

자료의 분포 중에서 가장 빈번하게 관찰되는 최다도수를 갖는 자료 값이다. 자료가 (1, 2, 3, 3, 5, 8)로 구성되어 있다면, 3이 두 번 관찰되었고 다른 값들은 한 번씩 관찰되었으므로, 3이 최빈값이 된다. 최빈값은 한 개 이상 존재할 수 있는데, 최빈값이 두 개 존재할 경우를 이봉분포(bimodal distribution), 세 개 이상 있을 경우를 다봉분포(multi-modal distribution)라고 한다⁸⁾.

6) [네이버 지식백과] 평균 [mean] (교육평가용어사전, 2004. 5. 31., 학지사)

7) [네이버 지식백과] 중앙값 [median] (교육평가용어사전, 2004. 5. 31., 학지사)

8) [네이버 지식백과] 최빈값 [mode] (교육평가용어사전, 2004. 5. 31., 학지사)

제 3 절 교수 학습자료 해설

대푯값



활 동 지 1

통계에도 오류가 있다?

♣ 어느 나라의 1인당 국민총생산⁹⁾이 높으면 사람들은 그 나라에 사는 전체 국민의 소득 수준이 높다고 생각할 것입니다. 그러나 그 나라에 억만장자가 몇 명 있다면 국민들의 실제 소득은 그리 높지 않은데도 1인당 국민총생산이 높게 나타나게 될 것입니다.

통계에서는 평균이 갖는 이러한 한계를 개선하기 위하여 대푯값으로 중앙값이나 최빈값을 사용합니다. 중앙값은 자료를 크기가 작은 값부터 크기순으로 늘어놓았을 때 가장 중앙에 위치하는 값입니다. 만약 자료의 개수가 홀수라면 중앙값은 가운데 위치하는 값이고, 자료의 개수가 짝수라면 중앙값은 가운데 위치하는 두 값의 평균이 됩니다. 또한 최빈값은 통계 자료에서 가장 빈번하게 나오는 값으로 두 개 이상이 될 수도 있습니다.

우리는 시험 성적부터 국민총생산을 살펴보는 경우까지 다양한 통계 자료를 해석할 때 평균을 사용하고 있습니다. 생활 속에서 자료의 대푯값으로 평균을 사용할 때 적절하지 않았던 경우를 설명해 보세요.¹⁰⁾

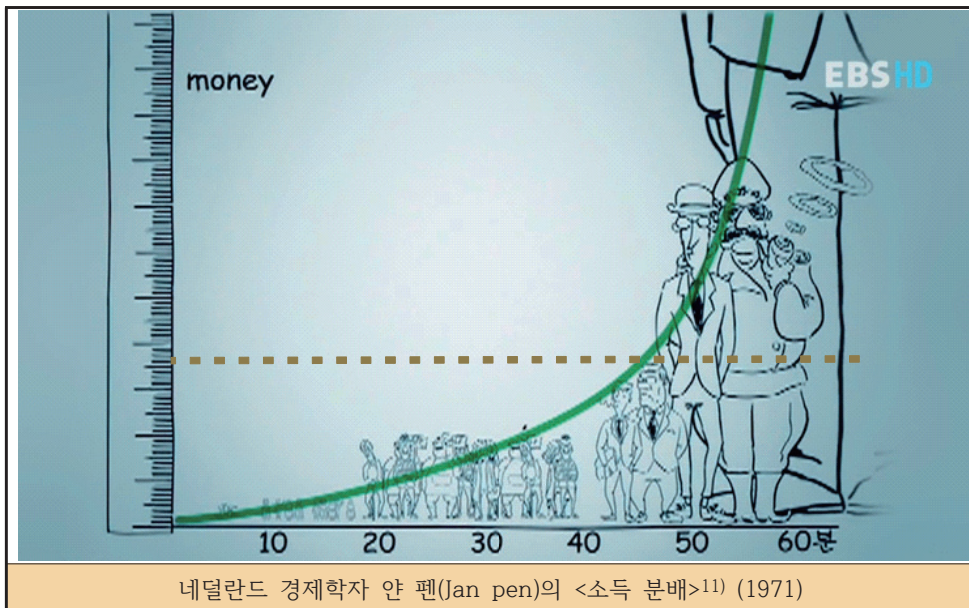
- 저수지나 강의 평균 수심만 가지고 판단하면 위험할 수 있다.
- 중학교 1학년 친구들의 키나 몸무게를 비교해 보면 평균 키와 몸무게보다 크거나 작은 친구들이 많이 있다.

다양한 상황에서 얻어진 자료는 통계적인 처리 방식에 따라 서로 다른 해석을 낼 수 있습니다. 특히 실생활에서 얻어지는 통계 자료는 위의 세 가지 대푯값 중 한 가지만 선택함으로써 부적절한 결과를 유도할 수 있습니다. 이에 통계 자료를 다룰 때는 비판적인 관점에서 오류가 없는지 살펴보려는 태도가 중요합니다.

9) 국민총생산(Gross National Product) : 일정 기간 동안 한 나라의 국민이 국내외에서 새롭게 생산한 재화와 용역의 부가 가치 또는 최종재의 값을 화폐 단위로 합산한 것.

10) 안재찬. 명품수학(2005)

네덜란드 경제학자 얀 펜(Jan pen)이 1971년도에 <소득분배>라는 책을 통하여 난쟁이의 행렬을 설명했습니다. 이는 소득의 순서대로 사람을 나열하면 어떤 형태가 될 것인가에 대한 질문의 답변이었습니다. 난쟁이의 행렬은 60분간 행진할 때 소득이 작은 사람부터 소득이 큰 사람으로 행진하며, 소득의 크기만큼 키가 정해진다는 것입니다. 즉 키가 작을수록 소득이 작고, 키가 클수록 소득이 크다는 뜻입니다. 처음에 행진이 시작되면 땅 속에 머리를 박고 행진을 시작하는데, 이들은 마이너스의 소득을 가진 사람들(채무자)입니다. 이후 1미터가 안 되는 사람들이 20분 가까이 행진하게 됩니다. 이 때 한 가지 발견할 수 있는 것은 여성들이 남성들보다 앞에 위치한다는 것입니다.



그리고 30분이 지나도 평균키의 사람들은 나타나지 않다가 48분이 되어서야 평균 키의 사람들이 나타나게 됩니다. 즉, 100명 중 80명 쯤 나타나는 사람이 평균 소득을 가진 사람입니다. 이후 6분을 남겨놓고 2미터 이상의 사람들이 나타납니다. 마지막 몇 초는 100미터에 달하는 거인들이 나타나고, 제일 마지막에는 석유왕 폴 게티가 나타납니다. 석유왕 폴 게티는 “만일 당신이 당신의 재산을 계산할 수 있다면 당신은 진짜 부자는 아닙니다.”라는 말로 유명합니다.¹²⁾ 얀 펜의 난쟁이의 행렬 역시 일상생활에서 흔히 사용하는 평균의 한계를 보여줍니다.

실제로 얀 펜의 난쟁이의 행렬에서 30분에 나타나는 사람들의 소득이 중앙값이므로 평균보다는 전체 소득을 대표하는 값으로 적절할 수도 있습니다. 또한 25분에 나타나는 값이 전체 소득 중 가장 많이 나타나는 최빈값이므로 이것이 전체 소득의 대푯값으로 더 적절하다고 볼 수도 있습니다.

11) EBS 동영상 (2010. 9. 13.)

12) <http://blog.naver.com/xyzkxx/10094279994>



Activity 1 평균의 합성

♣ 다음 표는 2014년 A지역 공무원 37명의 연간 봉급 총액과 2011년 LPGA에서 활동하는 한국계 골프선수 41명이 각자 획득한 상금의 액수입니다.

25	27	37	33	25	25
28	27	37	33	25	35
28	30	38	33	27	33
32	32	33	38	27	41
33	33	36	38	30	40
37	33	28	38	33	30
38					

[2014' A지역 공무원 37명의 연봉] (단위 : 백만원)

44	256	53	5	161	477
109	8	23	18	721	912
325	149	18	415	30	619
1357	886	22	72	16	5
245	165	408	852	77	28
87	262	94	365	21	82
182	350	42	42	628	

[2011' LPGA 한국계 골프선수 41명의 연간 상금 액수] (단위 : 천달러)

(1) 위의 자료를 보고 고은이는 다음과 같이 분석했습니다.

○ 고은 : A지역 공무원 연봉의 평균은 32(백만원)으로 대푯값으로 적절해.

여러분은 이 의견에 대해 어떻게 생각하나요?

- 32(백만원)을 기준으로 대부분의 연봉이 그 차이가 적으므로 적절하다.
- 히스토그램을 참고하면 32 가까이에 대부분 분포되어 있다.

(2) 고은이가 분석한 것을 보고 덕화는 다음과 같이 분석했습니다.

○ 덕화 : 골프선수들의 연간 상금 획득액 평균은 259(천달러)로 대푯값으로 적절해.

여러분은 이 의견에 대해 어떻게 생각하나요?

- 자료에서 평균 259와 비슷한 상금을 받은 선수는 약 8명 정도 뿐이므로 대푯값으로 적절하지 않다.
- 히스토그램을 참고하면 지나치게 큰 상금을 받은 선수들이 있어서 평균이 높아졌기 때문에 적절하지 않다.

(3) 공유는 다음과 같이 분석했습니다. () 안에 적절한 이유를 써 넣어 보세요.

○ 공유 : 내 생각엔 공무원 연봉과 달리 골프선수의 상금은 중앙값이 자료의 대푯값으로 적절한 것 같은데?
(극단값(지나치게 크거나 작은 값)의 영향을 받지 않기) 때문이야.

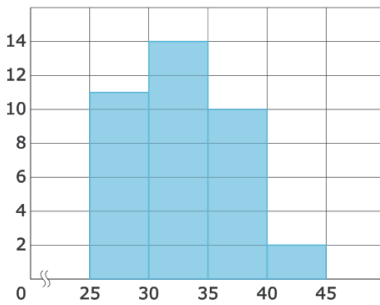
(4) 위 활동을 통해서 알게 된 점이나 느낀 점에는 어떤 것이 있는지 써 보세요.

- 대푯값을 여러 개 배우는 이유를 알게 되었다.
- 평균이 지닌 단점을 알게 되었다.
- 중앙값이 언제 유용한지 알게 되었다.

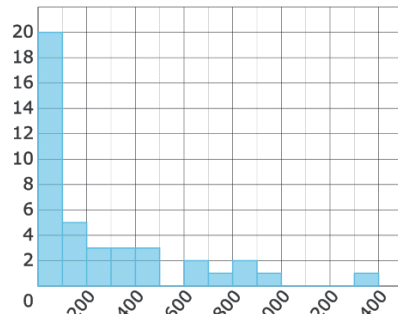


이지통계13)를 활용하여 주어진 자료를 다음과 같이 히스토그램으로 나타내어 주어진 과제에 답할 수도 있습니다.

2014년 A지역 공무원 37명의 연간
봉급 총액(백만원) 히스토그램

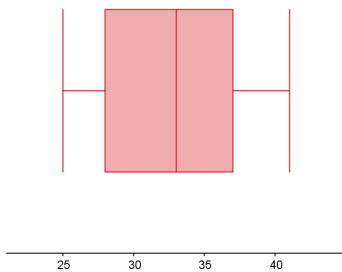


2011년 LPGA에서 활동하는 한국계 골프선수
41명이 각자 획득한 상금의 액수(천달러)
히스토그램

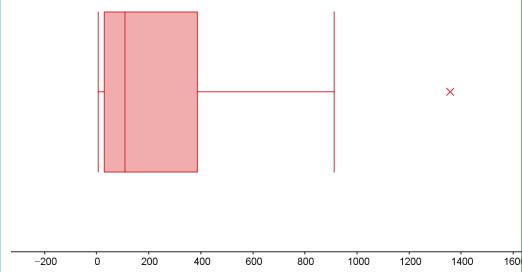


위와 같은 히스토그램을 이용하여 과제를 해결한 학생들에게는 지오지브라 프로그램을 통해 주어진 자료를 다음과 같이 상자그림으로 나타내어 비교하도록 안내할 수도 있습니다.

2014년 A지역 공무원 37명의 연간
봉급 총액(백만원) 상자 그림



2011년 LPGA에서 활동하는 한국계 골프선수
41명이 각자 획득한 상금의 액수(천달러)
상자그림



상자그림이란 자료들을 크기가 작은 값부터 순서대로 나열하여 4개의 그룹으로 나누고 이것을 꼬리와 상자로 나타낸 그래프입니다. 이 때 4개의 그룹을 5개의 수로 표시하여 분포를 설명할 수 있습니다. 두 그래프 중 LPGA 선수의 상금이 지나치게 큰 값이 X로 표시되어 있습니다. 이와 같이 자료들보다 지나치게 크거나 작은 값을 극단값이라고 하며, 자료에서 제외하고 분포를 분석하기도 합니다.



Activity 2 상황에 적절한 대푯값 고르기

♣ 청소년기에는 학업으로 인해 취미 생활을 갖기가 어렵고 스마트폰이나 컴퓨터에 접근하는 것이 쉬워 과도하게 이를 사용하는 경우가 있습니다. 하지만 청소년기에 과도한 컴퓨터의 사용은 학업에도 영향을 미칠 뿐만 아니라 신체, 정서, 인지 발달에 전반적인 영향을 미친다고 합니다. 이에 무엇보다 규칙적인 운동을 하는 것이 스트레스 해소에도 도움이 될 뿐만 아니라 오랜 시간 앉아서 공부하는 청소년기에 성장판을 자극하는 기회가 되기도 합니다. 청소년기에는 호르몬 분비가 활발하기 때문에 저녁 7시부터 9시까지 가벼운 걷기나 맨손 체조와 같은 운동을 지속적으로 하는 것이 중요합니다.

소은이는 우리 반 친구들의 운동 시간이 얼마나 되는지를 알아보기 위하여 한 달에 1시간 이상 운동하는 날이 며칠인지 조사하였습니다. 8명의 친구들이 운동한 시간은 다음과 같습니다.

2	3	7	10	12	26	30	30
---	---	---	----	----	----	----	----

8명 중 한 명인 도빈이는 한 달에 12일을 1시간 이상 운동합니다. 다음은 도빈이가 다른 7명의 친구들에 비해 운동을 자주하는 편인지에 대한 학생들의 대화입니다.

- 덕화 : 8명의 평균 운동 시간이 15일인데 이에 비하면 도빈이는 적게 운동하는 편이야.
- 선형 : 최빈값이 30일인데 12일은 턱없이 작잖아. 도빈이는 다른 학생들보다 적게 운동하는 것 같은데...

위 대화에 대한 자신의 의견을 말해 보세요.

- 한 달에 1시간 이상 운동하는 날이 2일에서 30일까지로 흩어져 있는 정도가 크므로 이러한 경우에는 자료의 대푯값으로 중앙값인 11일을 택하는 것이 적절하다. 따라서 도빈이는 중앙값에 비추어 볼 때 약간 많이(적당히) 운동한다고 볼 수 있다.

13) <http://www.ebsmath.co.kr/easyTong>



활동지 2

균형점으로써의 평균

♣ 그림과 같이 양팔 저울에 무게가 같은 추를 왼쪽에 2개, 오른쪽에 4개 걸었더니 균형을 이루었습니다. 왜 이런 일이 일어날까요?



평균은 자료 값을 모두 더하여 자료의 전체 개수로 나눈 값, 즉 자료의 공평한 몫으로서의 의미를 가지고 있습니다. 또 한 가지는 수치 자료의 중심으로서의 의미를 가지고 있습니다. 자료 분포의 균형점으로써의 평균의 의미를 말합니다.

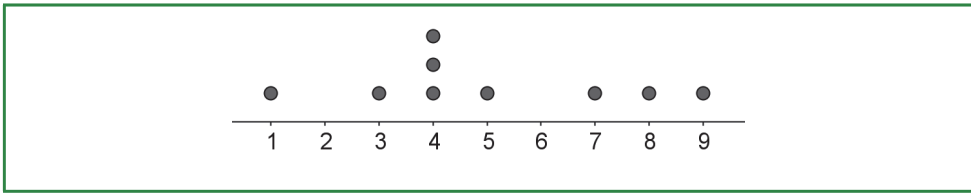
9명의 학생들에게 다음과 같이 질문했습니다.

집에서 애완동물을 몇 마리 키우나요?

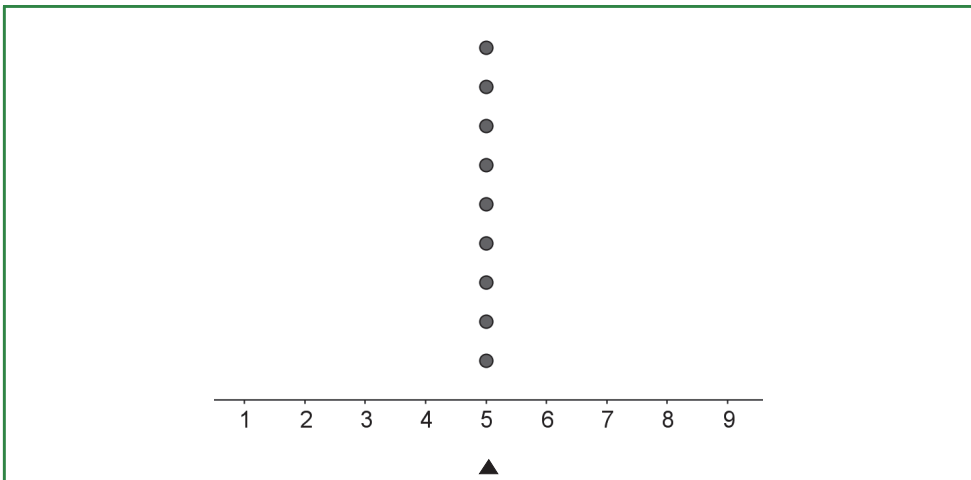
조사 결과 학생들은 다음과 같이 애완동물을 키우고 있다고 답했습니다.

1 3 4 4 4 5 7 8 9

학생들이 가지고 있는 애완동물의 수를 점도표로 나타내면 다음과 같습니다.

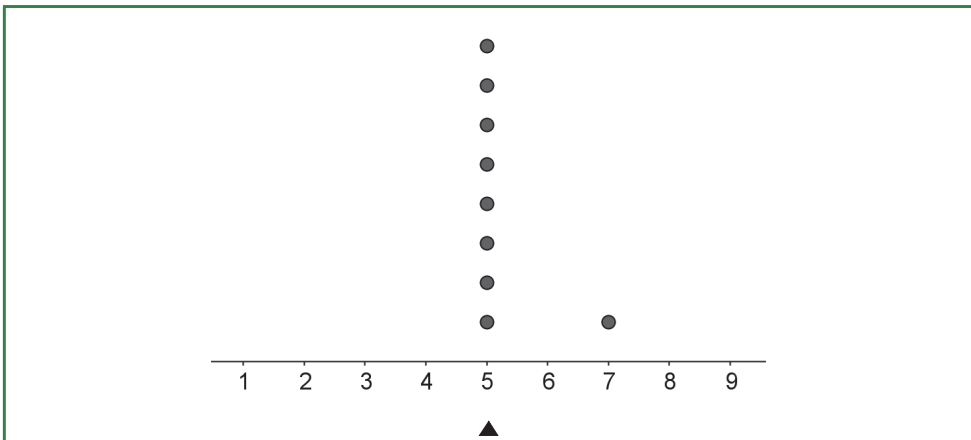


애완동물의 수를 모두 더하면 45마리가 됩니다. 45마리의 애완동물들을 9명의 학생들에게 골고루 나눠준다면 5마리씩 가지게 됩니다. 즉 애완동물 수의 평균은 5입니다. 9명의 학생들이 각각 5마리씩 가진 것을 점도표로 나타내면 다음과 같습니다.



만약 위의 점도표를 시소 위에 올려둔다면 중심이 5일 때 시소는 균형을 이룰 것입니다. 즉 이 점도표에서 균형점은 5입니다. 그렇다면 원래의 애완동물의 수로 나타낸 점도표의 균형점은 얼마일까요?

아래 그림처럼 5 위에 올려져있는 점 하나를 떼어 7위에 올리면 어떤 일이날까요?

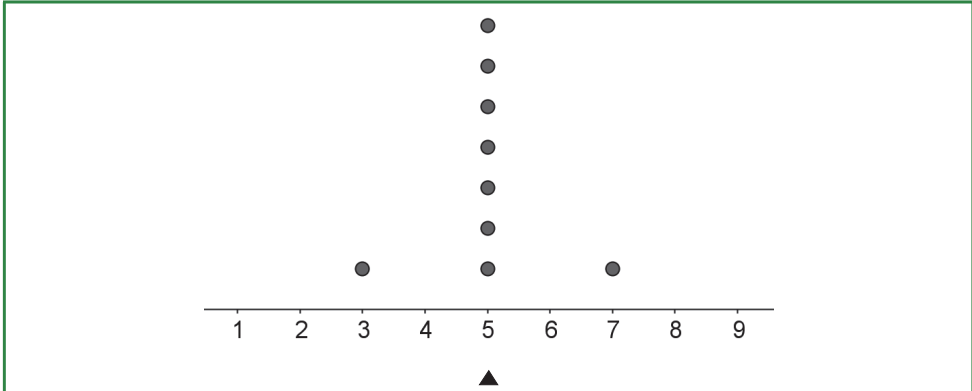


중심이 그대로 5라면 시소는 분명히 오른쪽으로 기울어 질 것입니다. 중심을

통계적 소양을 위한 교수 학습자료

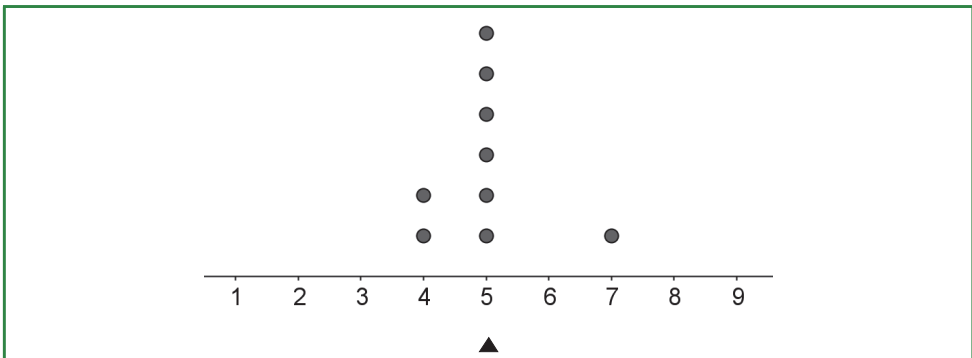
그대로 두고 시소의 균형을 맞추려면 5위에 있는 점들을 어떻게 해야 할까요?

시소 위에서 7은 5보다 2칸 오른쪽(큰 쪽)에 있기 때문에 5위에 있는 점 하나를 5보다 2칸 왼쪽(작은 쪽)인 3에 옮기는 것이 한 가지 해결 방법입니다. 다음 그림과 같이 나타낼 수 있습니다.



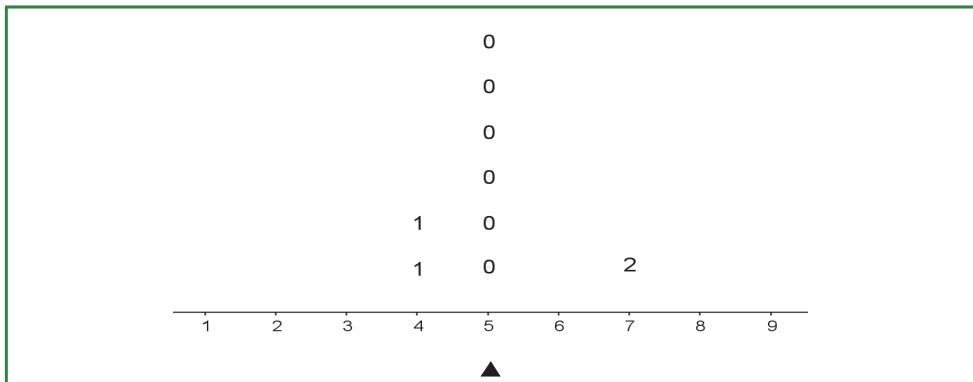
이제 시소는 중심 5에서 다시 균형을 이루게 됩니다. 그렇다면 다른 방법은 없을까요?

5위의 점을 한 개가 아니라 두 개 옮기려면 어디로 옮겨야 할까요? 또 다른 방법은 아래 점도표처럼 점 두 개를 4 위에 올리는 것입니다.



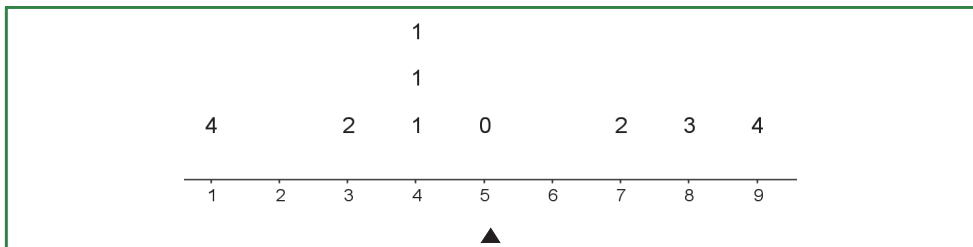
시소는 다시 중심 5에서 균형을 이루었습니다. 즉 시소의 균형점은 5입니다.

이때 각각의 점들을 중심 5에서 떨어진 거리로 바꾸면 다음과 같은 점도표가 됩니다.



중심인 5보다 왼쪽(작은 쪽)에 있는 두 수의 합(떨어진 거리의 합)은 5보다 오른쪽(큰 쪽)에 있는 한 개의 수와 같습니다. 그렇기 때문에 시소의 균형점은 5입니다.

이제 학생들이 키우는 애완동물의 수를 중심 5에서 떨어진 거리로 표시하면 아래와 같은 점도표로 나타낼 수 있습니다.



중심 5의 왼쪽(작은 쪽)의 값들의 떨어진 거리들의 합과 오른쪽(큰 쪽)의 값들의 떨어진 거리들의 합이 모두 9로 같습니다. 이러한 이유로 평균 5가 자료의 균형점이 됩니다.¹⁴⁾

14) Guidelines for Assessment and Instruction in Statistics Education(GAISE) Report(2005). A PRE-K-12 Curriculum framework.



Activity 1 평균의 의미 1

♣ 오른쪽 표는 사과 30개의 무게(g)를 적은 표입니다.

232	203	308	216	239	266
242	259	265	225	286	251
257	277	257	229	244	260
278	263	248	242	240	294
208	312	210	295	271	220

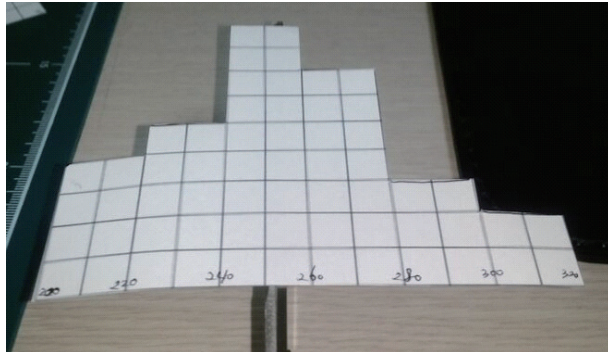
(1) 자료의 평균을 구해 보세요.
(소숫점 아래 첫째 자리에서 반올림하여 자연수로 나타내세요.)

사과의 무게(g)

평균 253

(2) 아래 도수분포표를 완성하여 [활동지 1]의 모눈종이에 위 자료의 히스토그램을 그리고 오려 보세요.

무게 (g)	도수(개)
200 ^{이상} ~ 220 ^{미만}	4
220 ~ 240	5
240 ~ 260	9
260 ~ 280	7
280 ~ 300	3
300 ~ 320	2
합계	30



(3) 연필이나 자 위에 오른쪽 사진과 같이 올려놓았을 때 떨어지지 않는 균형점을 찾아 표시해 보세요. 균형점으로 추측한 평균과 실제 평균은 어떻게 다른지 설명해 보세요.

균형점으로 추측한 평균은 약 253g이고 도수분포표의 평균과 매우 비슷하다. 사과의 무게가 253g을 기준으로 고르게 분포되어 있으므로 대푯값으로 적절하다.





Activity 2 평균의 의미 2

♣ 다음은 2011년 한 해동안 우리나라 20개 국립공원의 입장객 수를 조사하여 나타낸 도수분포표입니다. 이 도수분포표를 이용하여 구한 평균은 ?(만명)입니다.

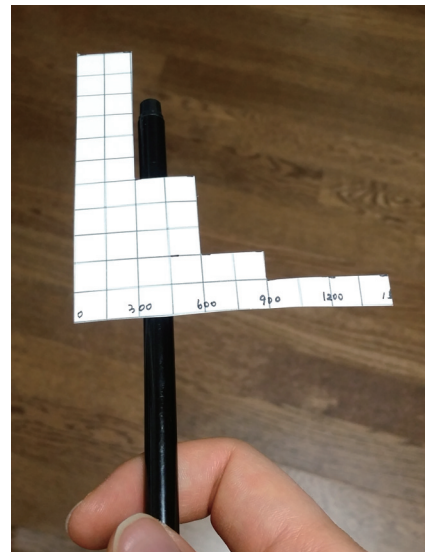
입장객 수(만명)	도수(개)
0 ^{이상} ~ 300 ^{미만}	11
300 ~ 600	5
600 ~ 900	2
900 ~ 1200	1
1200 ~ 1500	1
합계	20

(1) 모눈종이에 히스토그램을 그리고 균형점을 예상해서 표시해 보세요.



Tip

1. 학생들이 활동을 통해 균형점을 추측한 후에 교사가 도수분포표로 구한 평균(390만 명)을 알려준다.
2. 2015 수학과 개정 교육과정부터 도수분포표에서 평균을 구하지 않도록 지도함을 유의한다.

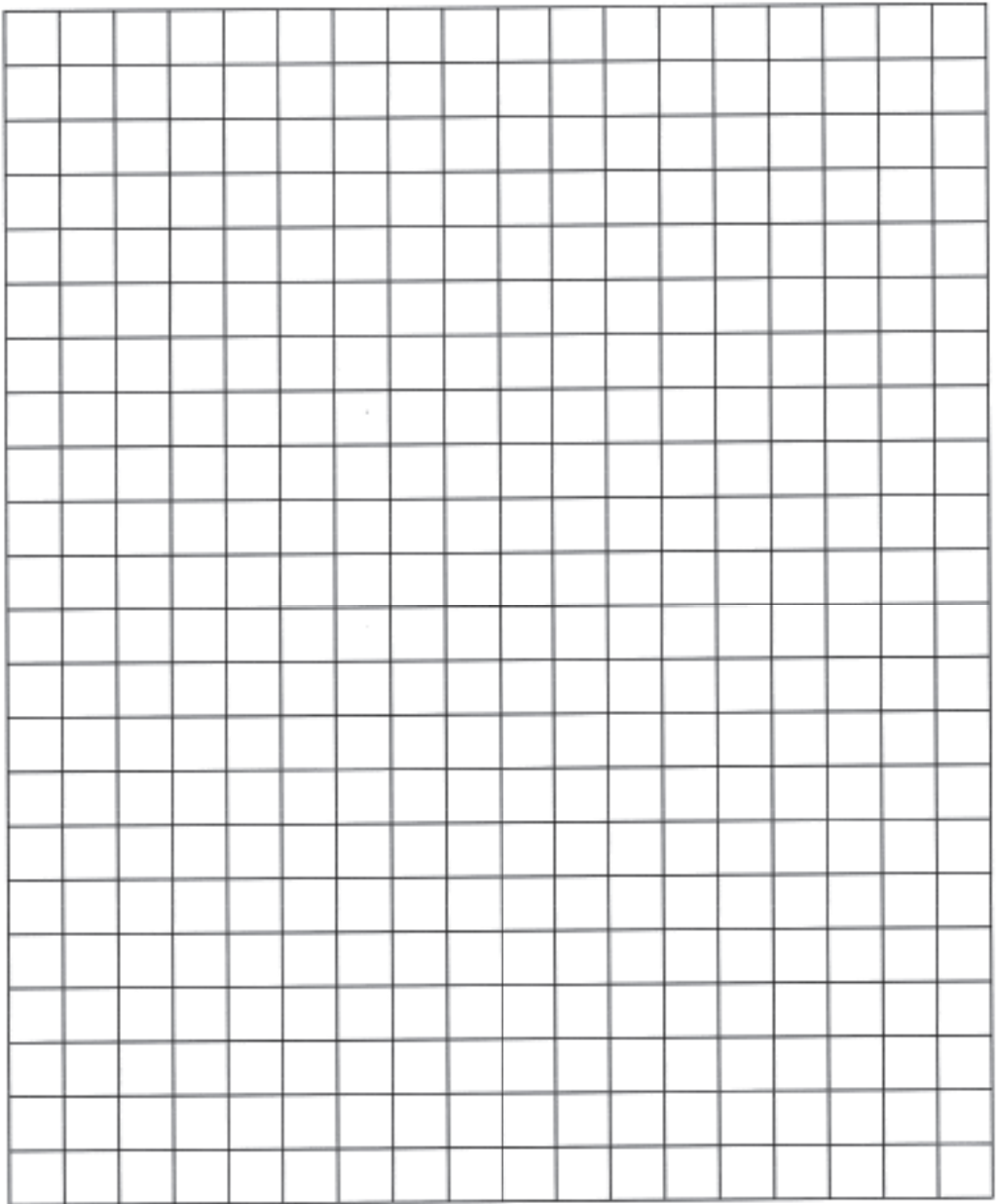


(2) 히스토그램을 오려 연필 위에 올렸을 때 떨어지지 않는 균형점과 예상한 균형점을 비교해 보세요.

(3) 균형점으로 추측한 평균은 입장객 수를 대표하는 대푯값으로 적절한가요? 이유를 적어 보세요.

- 균형점으로 추측한 평균은 약 410(만명)이고 도수분포표의 평균 390(만명)과 비슷하다. 자료가 고르게 분포되어 있지 않은 경우에도 균형점은 평균과 비슷함을 확인할 수 있다. 다만 대부분의 국립공원 입장객수가 0~300(만명)의 계급에 해당되는 것을 감안할 때 대푯값으로 적절하지 않다. 이 경우에는 최빈계급 또는 중앙값이 포함된 계급의 계급값을 대푯값으로 선택하는 것이 적절할 수 있다.

♣ 아래 모눈종이에 히스토그램을 그리세요. 자르기 전에 계급의 범위를 적으세요.



부록

부록1. 자료 수집 학생 활동지

자료수집



활동지 1

설문조사하기



Activity 1 자료는 어떻게 수집해야 할까?

대전중학교 학생회장에 출마한 철구는 교내 매점에 대한 공약을 만들기 위해 학생들의 매점 이용 실태를 확인하고 싶어 합니다. 그래서 학생들이 매점을 하루에 몇 번 이용하는지 조사하기로 하였고, 친구인 상수와 민희에게 조사를 부탁하기로 하였습니다.

대전중학교에는 각 학급당 20명씩 학년당 5개 반이 있어, 전교생은 총 300명입니다. 상수와 민희는 각각 30명의 학생을 대상으로 조사하기로 했습니다. 상수와 민희가 조사한 결과는 아래 표와 같습니다.

이용 횟수	0	1	2	3	4회 이상	합계
학생 수	6	11	7	4	2	30

[상수의 조사 결과]

이용 횟수	0	1	2	3	4회 이상	합계
학생 수	0	5	10	9	6	30

[민희의 조사 결과]

설문조사를 어떻게 했는지 묻는 질문에 상수와 민희는 다음과 같이 이야기했습니다.

- 상수 : 내 휴대전화 주소록에 있는 친구 30명을 대상으로 조사했어.
- 민희 : 교내 매점 앞에 조사 부스를 설치해서 선착순 30명을 대상으로 조사했어.

(1) 상수와 민희의 조사 결과가 다른 이유는 무엇인지 이야기해 봅시다.

(2) 상수와 민희의 조사 방법은 적절한 것일까요? 자신의 생각을 이야기해봅시다.

민식은 다음과 같이 새로운 설문조사 방법을 제안했습니다.

○ 민식 : 전교생의 이름이 각각 적힌 종이를 상자에 넣어서 그 중 30개를 뽑아 그 학생들을 조사하자.

(3) 민식의 조사 방법은 상수, 민희의 방법과 비교했을 때 더 적절한지 이야기해 봅시다.

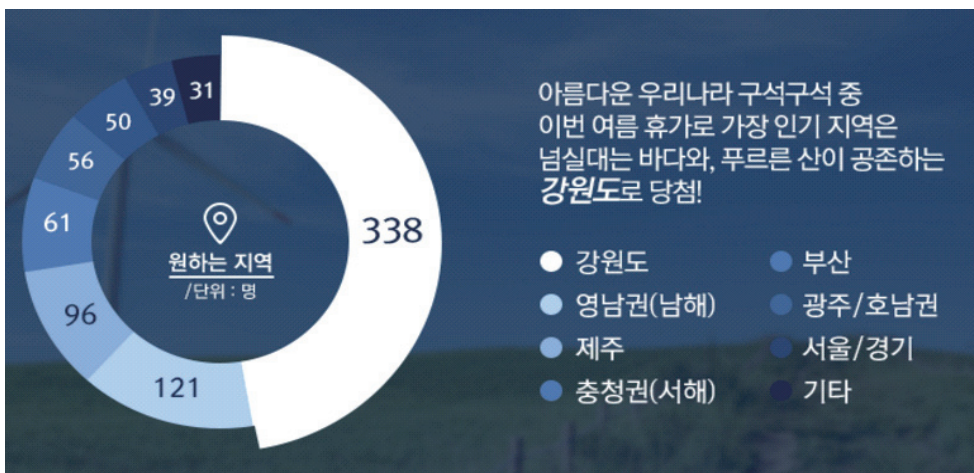
(4) 만약 본인이 직접 조사를 한다면 어떻게 할 것인지 생각해보고 자신의 조사가 왜 타당한지 이야기해봅시다.



Activity 2 설문조사를 믿는 방법

많은 사람들이 여름휴가를 보내기 위해 다른 지역으로 여행을 갑니다. 사람들이 여름휴가를 어디서 어떻게 보내는지에 대한 조사는 기업에게 숙박, 관광, 교통수단 등 상품 개발과 마케팅 전략을 세우는 데 활용되는 중요한 정보입니다. 정부에서는 여가와 관련한 다양한 정책을 입안하는 데 조사 결과를 활용할 수 있으며, 여행객 또한 자신의 여행 계획을 세우는 데 참고할 수 있습니다.

한 기업에서는 지난 2016년 7월에 약 800명을 대상으로 설문조사를 실시하여 아래와 같이 블로그를 작성하였습니다.



(1) 희선이는 블로그를 보고 다음과 같이 의문이 생겼습니다.

○희선 : 이 기업에서는 이런 정보를 어떻게 얻을 수 있었을까?

이 기업에서는 조사에 참여할 사람들을 어떻게 모집하였을지 논의해봅시다.

(2) 현욱이는 블로그를 보고 다음과 같이 말했습니다.

○ 현욱 : 이 설문조사에서는 조사에 참여한 사람들의 연령대, 성비, 거주 지역과 같은 정보가 없어.

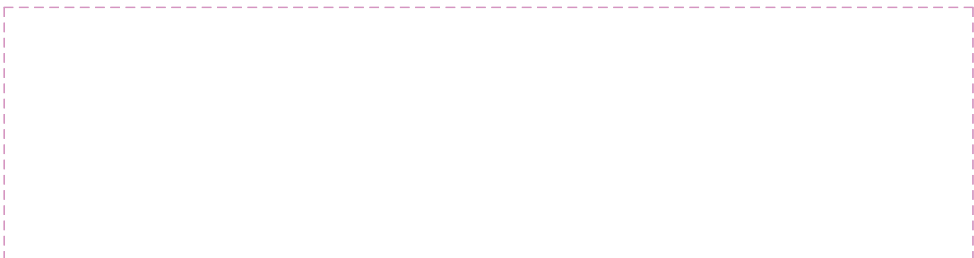
현욱이가 말한 이 사실이 위 그래프를 해석하는 데 어떤 영향을 미칠 수 있을지 이야기해 봅시다.



(3) 이 조사가 적절하게 이루어졌다고 확인하기 위해서는 어떤 정보가 더 필요한지 이야기해 봅시다.



(4) 만약 본인이 직접 조사를 한다면 누구를 대상으로 어떻게 조사를 할 것인지 생각해봅시다.





활동지 2

실험하기, 관찰하기, 측정하기



Activity 1 일조량에 따른 콩의 성장 실험

상원이는 일조량에 따라 콩의 성장을 비교하기 위해 다음과 같은 관찰 계획을 세웠다.

- 가. 동일한 크기의 화분 A, B 각각에 일정한 크기로 썬 5개의 콩을 동일한 간격으로 심는다.
- 나. 두 화분을 모두 실내에서 키우되 화분A는 햇빛이 잘 드는 창가 옆에 놓고, 화분B는 마분지로 그늘을 만들어 키운다.
- 다. 50일 동안 5일 간격으로 같은 시간에 각각의 화분에 심은 5개의 콩의 키를 자를 이용하여 밀리미터 단위까지 측정한다.
- 라. 측정한 콩의 줄기의 길이는 관찰기록지에 기록한다.
- 마. 콩의 줄기의 길이를 관찰하는 기간 동안 두 화분에 주는 물의 양은 동일하게 유지한다.

화분 종류	콩 번호	관찰기록지										관찰 중 특이사항
		5일	10일	15일	20일	25일	30일	35일	40일	45일	50일	
A	1											
	2											
	3											
	4											
	5											
B	1											
	2											
	3											
	4											
	5											

(1) 상원이의 관찰계획에서 같게 한 조건과 다르게 한 조건을 아래의 <보기>에서 고르시오.

<보기>

- | | | | |
|----------|---------|---------|----------|
| ㉠ 빛의 양 | ㉡ 물의 양 | ㉢ 물의 온도 | ㉤ 화분의 크기 |
| ㉥ 토양의 종류 | ㉦ 콩의 종자 | ㉧ 싹의 크기 | ㉨ 측정간격 |

① 같게 한 조건

② 다르게 한 조건

③ 같은지 다른지 알 수 없는 조건

(2) 상원이의 관찰 계획을 읽고 민수와 혜원이가 다음과 같은 문제를 제기하였다.

- 민수: 두 화분 각각에 콩을 하나씩만 심어서 크기를 비교하면 될 텐데 화분 당 5개씩 심은 이유가 뭐야?
- 혜원: 서로 다른 조건에서 자란 콩의 크기에 대해 최종적으로 측정할 자료(50일째)만 있으면 되지 않나? 5일째부터 45일째까지 측정할 필요가 있나?

다음 물음에 답하시오.

① 민수의 물음에 상원이가 어떻게 대답했는지 생각해보고 친구들과 이야기해 봅시다.

② 혜원의 물음에 상원이가 어떻게 대답했는지 생각해보고 친구들과 이야기해 봅시다.

(3) 상원이는 콩의 성장에 대한 관찰을 통해 다음과 같은 주장을 세우고자 한다.

콩은 일조량이 많을수록 더 잘 자랄 것이다.

다음 물음에 답하시오.

① 상원이의 관찰 계획은 위와 같은 주장을 뒷받침하기에 적절한 방법인지 이야기해봅시다.

② 위의 물음①에 대해 적절하다고 판단한 경우, 그 이유에 대해 친구들과 이야기해봅시다.

③ 위의 물음①에 대해 부적절하다고 판단한 경우, 상원이의 관찰 계획에서 수정되어야 할 것들에 대해 친구들과 이야기해봅시다.

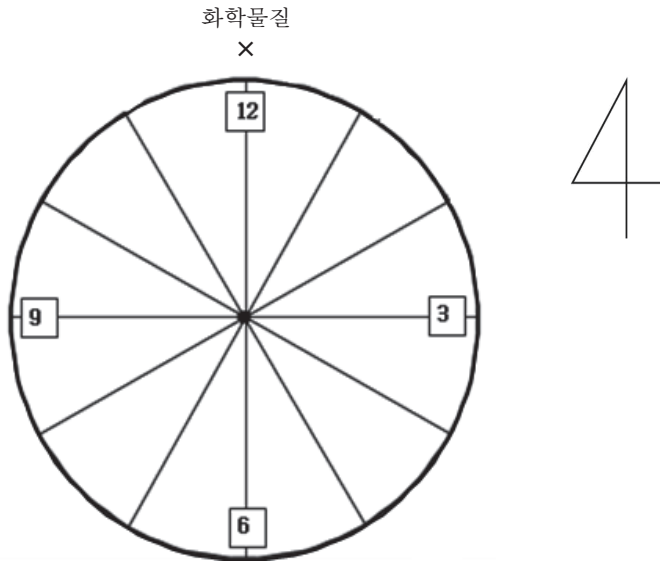


Activity 2 개미를 유인하는 화학물질 실험

상원이는 어떤 화학물질의 냄새가 개미를 유인할 수 있는지를 실험해보기 위해 다음과 같은 실험을 계획하였다.

가. 페트리 접시에 개미 30마리를 준비한다.

나. 아래와 같이 반경 10cm의 원 둘레에 시계눈금을 그려두고 12시 방향의 눈금 1cm 뒤에 그 화학물질을 칠한다.



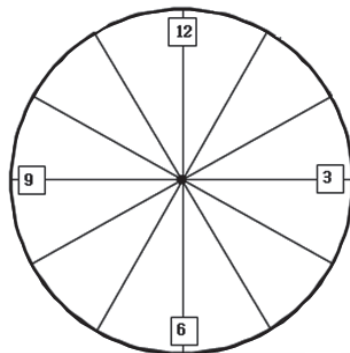
다. 원의 중심의 위치에서 개미를 한 마리씩 출발시켜 원 둘레에 도착한 위치의 시침 방향 값을 측정한다.(단, 시침은 10분 단위로 작성)

라. 30마리의 개미에 대해 다.의 과정을 독립적으로 반복 실시한다.

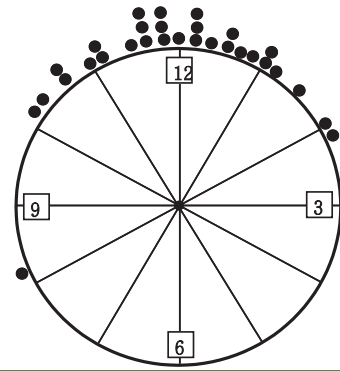
(1) 다음은 상원이가 처음 세 마리의 개미에 대해 실험을 실시하여 측정한 시침의 방향값이다.

1시 0분	11시 40분	12시 10분
-------	---------	---------

세 시침의 방향 값에 대응되는 위치를 오른쪽 그림에 표시해봅시다.



- (2) 오른쪽 그림은 상원이가 30마리의 개미에 대해 실험을 실시하여 각각의 개미가 도착한 위치와 횟수를 점으로 표시한 것이다. 점은 그 위치에 도착한 개미의 횟수만큼 표시된다. 상원이는 대부분의 개미가 12시 방향에서 가까운 위치에 도착한다는 사실을 통해서 '그 화학물질은 개미를 유인 한다'고 결론 내렸다. 상원이의 결론에 대해 수연이는 다음과 같은 문제를 제기하였다.



○수연: 개미가 12시 방향으로 움직이는 것이 화학물질 때문이 아니라 원래 그 방향으로 이동하는 경향이 있다면 이런 결과가 나올 수 있을 것 같아.

수연이가 제기한 문제를 해결하기 위해 상원이가 적절히 대처할 수 있는 방법에 대해 생각해보고 친구들과 이야기해봅시다.



활동지 3

환경보호 프로젝트 참가하기



Activity 1 환경보호 프로젝트 계획하기

○○환경단체에서 ‘되살리자! 우리 세계!’라는 슬로건을 내걸고 환경 되살리기 운동을 실시하고 있다. 이 단체에서는 이 운동의 일환으로 중학생들을 대상으로 환경보호 프로젝트 대회를 개최하고자 한다. 이 대회에 참가하기 위해서 학생들은 직접 환경보호와 관련된 주제를 설정하고 어떤 자료를 어떻게 수집할지 계획서를 작성하여 제출해야 한다. 아래의 ‘환경보호 프로젝트 참가 공고문’을 참고하여 환경보호 프로젝트 계획서를 작성해봅시다.

환경보호 프로젝트 참가 공고문

1. 환경보호와 관련된 주제를 설정한다.
2. 선정된 주제와 관련된 자료수집 계획을 작성한다.
3. 자료수집계획은 다음을 반드시 포함하여 구체적으로 작성한다.
 - 수집 대상
 - 수집 방법
 - 측정 방법 및 단위
 - 측정 도구
 - 측정기록지 양식
4. 계획서 양식은 자유롭다.



<출처: 네이버 어린이 백과>

부록2. 그래프 학생 활동지

그래프



활동지 1

그래프의 마술

최근 ○○중학교에서는 스마트폰을 사용하며 계단을 내려가던 K군과 스마트폰을 사용하며 계단을 올라가던 J양이 부딪쳐 계단 아래로 굴러 떨어지는 사고가 발생하였습니다. K군은 손목뼈에 금이 가서 깁스를 하게 되었고 J양은 타박상을 입었습니다.

이 사건을 취재하던 신문부에서는 이 사건과 함께 이 '청소년들의 스마트폰 이용실태'와 '○○중학교 학생들의 스마트폰 중독 실태'를 조사하여 다음 달 교내 신문에 특집 기사로 실기로 하였습니다.

'○○중학교 학생들의 스마트폰 중독 실태'는 1, 2, 3학년 각반에 3명(10번, 20번, 30번)의 학생을 정한 후 한국인터넷진흥원에서 만든 '스마트폰 중독 자가진단 검사지'를 이용하여 스마트폰 중독실태를 조사하였습니다.

○○중학교 신문부 학생들이 특집 기사를 효과적으로 표현하기 위해 편집자 회의를 하는 과정을 살펴보도록 합시다.



○ 진혁 : 조사한 자료를 좀 더 알기 쉽게 설명할 방법이 없을까?

○ 성욱 : 숫자나 표로 나타내는 것보단 그래프를 이용해서 나타내면 좀 더 쉽게 전달할 수 있을 것 같아.

○ 대준 : 그래프? 그래프로 나타내면 학 와 닿아서 좋긴 한데 잘못 받아들일 수도 있지 않을까?

○ 성욱 : 자료의 특성과 목적에 맞는 그래프를 사용한다면 괜찮은 것 같은데

○ 대준 : 그래프로 나타낼 때 또 신경써야할 건 없을까?

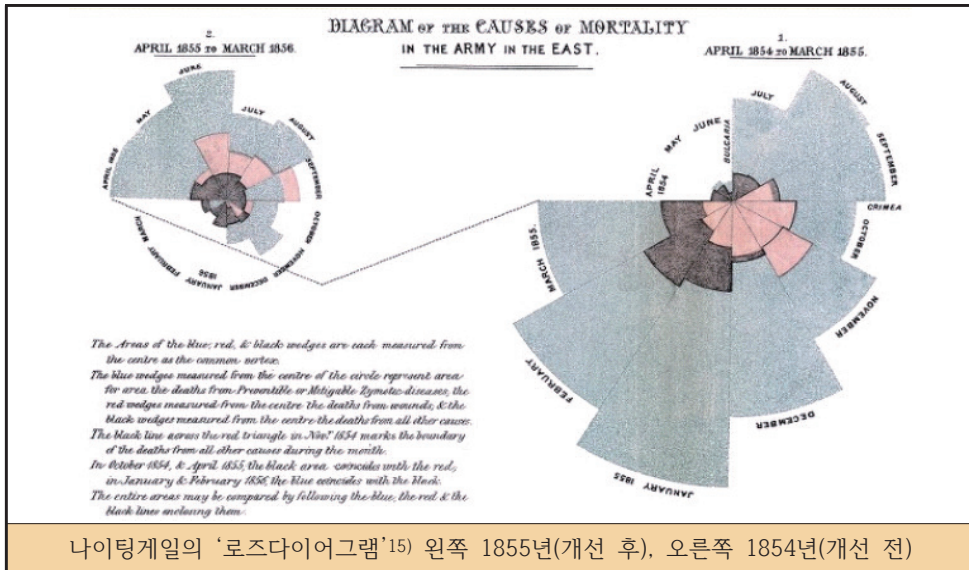
○ 진혁 : 그래프에 눈금이라든지 무엇을 나타낸 것인지... 이런 걸 잘 표시해야 할 것 같아.

○ 대준 : 그렇겠다. 그래프에서 1명 단위인지 1000명 단위인지 이런 거 때문에 잘못 이해했던 적이 있어. 어디 참고할 만한 좋은 그래프는 없을까?

○ 성욱 : '로즈 다이어그램'이란 걸 본 적이 있어. 나이팅게일이 만든 새로운 종류의 그래프인데 크림전쟁에서 죽어가는 군인들의 수를 월별, 죽은 원인별로 정리해서 나타낸 거야. 이 그래프를 통해 막상 전쟁에서 죽는 사람들이 전쟁이 아닌 위생이 안 좋아 죽어간다는 걸 알렸고 그 결과 많은 군인들의 생명을 지킬 수 있었어. 한 번 살펴볼까?

♣ 나이팅게일의 '로즈다이어그램'

'백문(百聞)이 불여일견(一見)'은 100번 듣는 것 보다 한 번 보다는 것이 좋다는 뜻입니다. 자료를 분석을 하는 경우 한 장의 그래프가 100문장의 설명보다 좋을 때도 있습니다. 다음을 통해 그래프의 위력을 알아보도록 합시다.



1854년 러시아와 연합군 간에 '크림전쟁'이 일어났습니다. 많은 영국 군인들이 부상과 질병으로 죽었습니다. 나이팅게일은 크림전쟁에서 간호장교로 활동하였습니다. 나이팅게일은 먼저 통계 작성기준을 세워 기록체계를 통일하였습니다.

오른쪽 그래프는 1854년에 사망한 군인의 수를 나타낸 것이고 왼쪽은 1855년에 사망한 군인의 수를 나타낸 것으로 부채꼴의 넓이는 사망자의 수를 나타냅니다.

또, 그래프에 달을 표시하고 사망원인을 색깔로 구분하였습니다. 파란색(가장 바깥쪽)은 질병, 빨간색(가운데)은 부상, 검은색(가장 안쪽)은 기타 이유로 죽은 군인의 수를 표시하였습니다.

기준에 따라 입원, 부상, 질병, 사망 등의 내역을 매일 상세히 작성하고 이를 토대로 영국 정부에 야전병원의 위생을 개선해야 한다고 주장하였습니다.

개선 사업을 시작한지 육 개월이 지나자 42%에 달하던 환자의 사망률이 2%까지 떨어졌습니다. 당시 사람들은 질병의 원인이 무엇이며, 깨끗해지면 왜 사망자가 줄어드는지 알지 못 했습니다. 하지만 나이팅게일이 제시한 통계는 깨끗한 위생이 사람을 살린다는 증거가 되었습니다.

○ 대준 : 그런데 사망자가 수의 차이를 잘 모르겠어. 부채꼴의 넓이가 사망자수라고 했는데 반지름의 길이를 사망자 수로 하면 사망자 수의 차이가 잘 나타나겠는데...

15) <http://blog.daum.net/mathkorea/15811174>

Activity 1 그래프의 착시와 선택

♣ 정보통신정책연구원(KISDI)에서 공개한 '2016한국미디어패널조사' 결과입니다.

년도	2011년	2012년	2013년	2014년	2015년	2016년
휴대폰 보급률(%)	89.2	91.2	91.8	92.4	93.2	94.0
스마트폰 보급률(%)	21.6	49.3	66.2	73.4	77.6	81.6

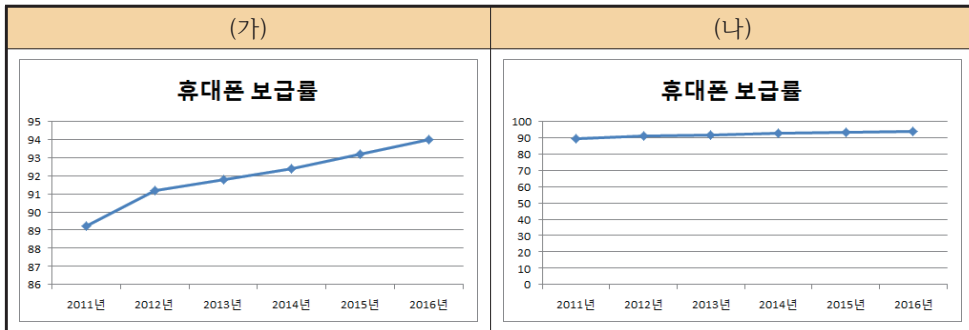
2012년 휴대폰과 스마트폰의 보급률은 각각 91.2%와 49.3%로 휴대폰의 보급률이 스마트폰의 보급률의 1.85배입니다. 이것을 그림그래프로 나타내면 다음과 같습니다.



○ 대준 : 막대 대신에 그림으로 나타내니깐 눈에 더 잘 들어오는 걸^^ 그런데 휴대폰 보급률이 스마트폰 보급률의 약 2배인데.. 왜 훨씬 더 차이가 나는 것으로 느껴질까?

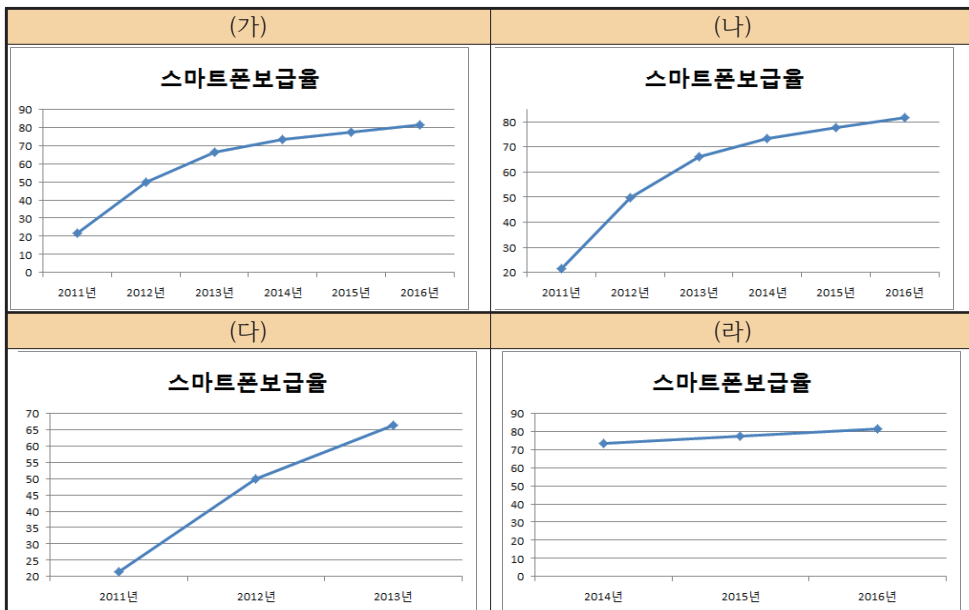
(1) 대준이가 실제보다 보급률의 차이를 더 크게 느끼는 이유는 무엇인가요?

대준 : 정말 그래프의 착시를 조심해야겠는걸. 또 그래프를 볼 때 조심해야 할 건 없니?
 성욱 : 그럼 다음 그래프를 한 번 볼까? 년도에 따른 휴대폰보급률을 나타낸 꺾은선그래프야.
 대준 : 같은 자료를 나타낸 그래프인데 두 그래프에서 느껴지는 보급률이 전혀 다르게 보이네.
 성욱 : 그런데 왜 이런 차이가 생기는 걸까?



(2) 대준이가 두 그래프에서 휴대폰 보급률을 다르게 느끼는 이유는 무엇인가요?

○ 성욱 : (가)에서 (라)는 년도에 따른 스마트폰 보급률을 나타낸 그래프야. 우리가 특집 기사를 어떤 방향으로 쓰지 결정하느냐에 따라 사용하는 그래프의 종류가 달라지지. 대준아~! 우리의 방향에 따라 어떤 그래프가 좋을지 생각해 볼까?

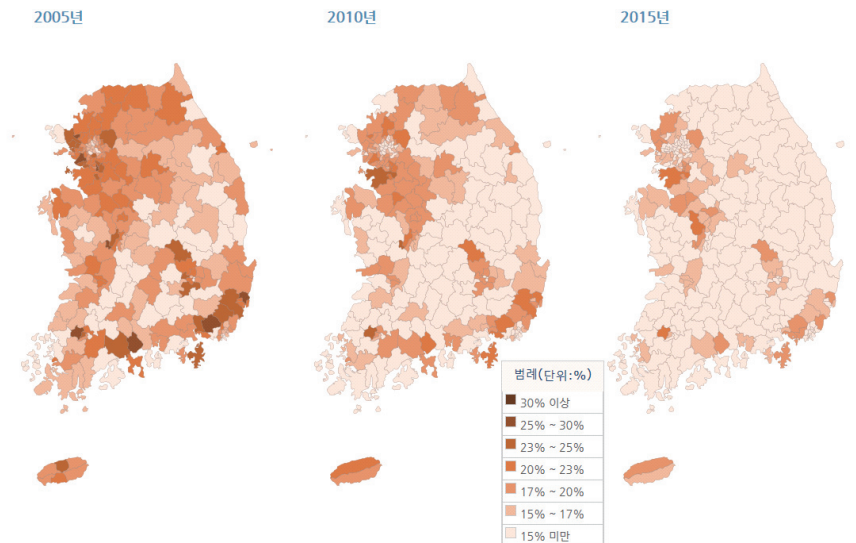


(3) 기사의 방향에 따라 적합한 그래프를 쓰고 적합한 이유를 써봅시다.

기사의 방향	적합한 그래프	이유
자료의 일부를 이용하여 스마트폰의 인기가 주춤한다는 것을 설명하고 싶을 때		
자료의 일부를 이용하여 스마트폰의 보급률이 급격히 증가한다는 것을 설명하고 싶을 때		
전체 자료를 이용하여 스마트폰 보급률이 천천히 변하는 것을 설명하고 싶을 때		
전체 자료를 이용하여 스마트폰 보급률이 급격히 변하는 것을 설명하고 싶을 때		

♣ 읽을거리 : 시간에 따른 변화를 나타내는 그래프의 예

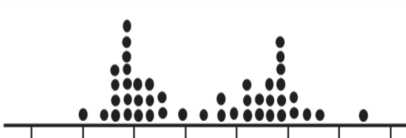
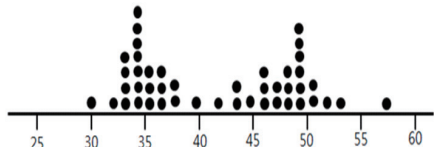
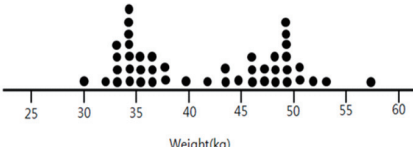
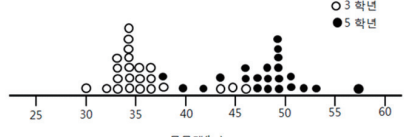
다음은 통계청에서 조사한 '15세 미만 인구의 비율'을 지도그래프를 이용하여 나타낸 것입니다. 흔히 시간에 따른 변화를 나타낼 때는 꺾은선그래프를 사용해야 한다는 생각이 있는데 가장 편하고 효율적인 방법이지 한 가지 방법밖에 없다는 것은 아닙니다.



Activity 2 효과적인 그래프

그래프를 잘 보면 자료 속에 숨은 정보를 효과적으로 알 수 있습니다. 하지만 일반적인 그래프에서도 변수, 단위, 집단이름 등을 표현하느냐 하지 않느냐에 따라 자료를 설명할 수 있는 정도가 달라집니다.

다음 점도표를 통해 변수, 단위, 집단이름 등 표현의 필요성을 살펴해보도록 합시다.

	
[그림1] 변수, 단위, 집단이름 등을 표시하지 않았을 경우	[그림2] 눈금에 숫자를 추가한 경우
	
[그림3] 변수와 단위를 추가한 경우	[그림4] 집단의 특성을 추가한 경우

○ 진혁 : [그림1]은 범례를 하나도 표시하지 않은 점도표야. 이 점도표에서는 자료가 두 곳을 중심으로 모여 있다는 것을 알 수 있어. 대준아! 여러 가지 표현을 추가하면서 더 알 수 있는 것을 찾아볼까?

(1) 대준이가 [그림1]의 점도표에서 범례를 하나씩 추가함으로써 새롭게 알 수 있는 정보는 무엇일까요?

구분	새로 알 수 있는 정보
[그림2] 눈금에 숫자를 추가하는 경우	
[그림3] 변수명과 단위를 추가하는 경우	
[그림4] 집단의 특성을 추가하는 경우	



활동지 2

스마트폰으로부터 학생을 지켜라~!



- 진혁 : 그럼 이제 우리 학교 학생들의 스마트폰 중독 실태'를 정리해 볼까? 그래프로 정리하면 좋겠지
- 대준 : 성별, 학년, 스마트폰 사용 시간 등 여러 가지를 조사했는데 어떤 그래프를 사용해야 할지 모르겠네
- 대준 : 일단 표로 정리해야 그래프로 나타낼 수 있으니깐 표로 정리해보자.

♣ 다음은 ○○ 중학교 신문부원들이 조사한 '스마트폰 중독 실태'를 정리한 표입니다.

구분	도수(명)
남자	22
여자	18
합계	40

<응답자 성별 빈도표>

구분	도수(명)
1학년	15
2학년	14
3학년	11
합계	40

<응답자 학년 빈도표>

구분	도수(명)
없다	2
있다	38
합계	40

<스마트폰 소유여부 빈도표>

스마트폰 사용 시간	도수(명)
~ 60분 미만	13
60분 이상 ~ 120분 미만	9
120분 이상 ~ 180분 미만	11
180분 이상 ~ 240분 미만	1
240분 이상 ~ 300분 미만	1
300분 이상 ~ 360분 미만	3
360분 이상 ~	2
합계	40

<스마트폰 사용시간 도수분포표>

구분	도수(명)
동영상시청	8
게임	12
SNS	14
정보검색	5
뉴스검색	1
합계	40

<스마트폰 주사용목적 빈도표>

자가진단점수	도수(명)
10점 이상 ~ 20점 미만	7
20점 이상 ~ 30점 미만	9
30점 이상 ~ 40점 미만	12
40점 이상 ~ 50점 미만	9
50점 이상 ~ 60점 미만	3
합계	40

<자가진단점수 도수분포표>

구분	도수(명)
일반 사용군	30
잠재적 위험군	6
고위험군	4
합계	40

<스마트폰 사용자군 빈도표>

- 대준 : 수집한 자료를 표로 정리하니깐 한 눈에 알아볼 수 있는 걸
- 성욱 : 그렇지. 표나 그래프로 나타내면 자료의 특징을 한 눈에 알아볼 수 있지.
대신 표가 그래프보다는 더 정확하게 받아들일 수 있어.
- 진혁 : 그런데 표도 그래프처럼 축의 크기 같은 걸 확인해야 할까?

- (1) 스마트폰 사용 시간을 계급의 크기가 5시간인 도수분포표로 정리하여 봅시다. 계급의 크기가 1시간일 때와는 어떤 점이 달라지나요? 또, 계급의 크기가 10분 이라면 어떤 현상이 일어날지 생각해 볼까요?

스마트폰 사용 시간	도수(명)
~ 5시간 미만	
5시간 이상~ 10시간 미만	
합계	

<스마트폰 사용시간 도수분포표>

- 대준 : 그런데 표는 자료의 특성에 따라 정리할 수 없을까? 나이팅게일의 그래프에서는 사망 원인에 따라 색깔을 구분했잖아
- 성욱 : 표에서도 구분할 수 있어. 스마트폰의 주된 사용 목적을 남학생과 여학생으로 구분하여 정리해볼까?
- 진혁 : 스마트폰 주된 사용 목적을 성별에 따라 정리해도 앞서 정리한 표와 같은 결과가 나올까?

- (2) 스마트폰의 주된 사용 목적을 성별에 따라 정리한 것입니다. 이전 자료에서 스마트폰의 주된 사용 목적을 성별 구분 없이 나타낸 자료의 결과와 같나요?

구분	도수(명)	
	남학생	여학생
동영상시청	5	3
게임	9	3
SNS	5	9
정보검색	2	3
뉴스검색	1	0
합계	22	18



Activity 1 정리한 표를 그래프로 나타내기

- (1) ○○ 중학교 신문부원들이 정리한 표를 원그래프, 띠그래프, 막대그래프, 히스토그램 중 어떤 것을 사용하여 나타내는 것이 가장 좋을지 생각해보고, 그 이유를 적어봅시다.

구분	가장 적합한 그래프	이유
스마트폰 소유		
스마트폰 사용시간		
스마트폰 주사용목적		
스마트폰 사용자군		

- (2) 앞에서 선택한 그래프를 이용하여 설문의 결과를 나타내어 봅시다.

원그래프(자료명 :)	띠그래프(자료명 :)
막대그래프(자료명 :)	히스토그램(자료명 :)

통계적 소양을 위한 교수 학습자료

다음은 가장 많이 사용되는 그래프와 특징에 대한 설명입니다. 이를 바탕으로 앞에서 만든 그래프가 자료를 나타내는데 적절하였는지 토의해 보도록 합니다.

♣ 참고자료 : 자료의 종류	
자료는 크게 측정형자료와 명목형자료로 구분할 수 있습니다. 측정형자료는 연속형자료와 이산형자료로 구분됩니다. 키와 몸무게와 같은 자료는 연속형자료이고 가족수, 다녀온 횟수 등과 같은 자료는 이산형자료입니다. 범주형 자료는 설문지에서 가장 많이 보는 매우불만, 불만, 보통, 만족, 매우만족 등의 순위형자료와 혈액형, 성별과 같은 명목형자료로 나뉩니다. 자료 분석에서는 명목형자료는 질적변수, 나머지 자료는 양적변수라고 하기도 합니다.	
원그래프	띠그래프
원그래프는 원의 크기에 따라 각 범주의 비율을 나타냅니다. 토마토가 가장 큰 비율을 차지하고 있습니다.	띠그래프는 수직축을 따라 각 범주의 비율을 나타내는 띠를 그립니다. 0%부터 100%까지의 구간을 나타냅니다.
원을 변숫값의 도수에 비례하여 여러 조각으로 나누어 나타낸 그래프. 범주형 자료의 분석에 적합	직사각형을 변숫값의 도수에 비례하여 여러 조각으로 나누어 나타낸 그래프. 범주형 자료의 분석에 적합함. 원그래프보다 객관적 비교가 가능
막대그래프	히스토그램
막대그래프는 수직축을 따라 각 범주의 도수를 나타내는 막대를 그립니다. 토마토가 가장 높은 도수를 가지고 있습니다.	히스토그램은 연속형 자료를 구간별로 나누어 각 구간의 빈도를 나타냅니다. 몸무게 히스토그램은 35kg부터 85kg까지의 구간을 나타냅니다.
범주형 자료의 도수분포를 한 눈에 비교할 수 있도록 각 자료의 값의 도수(또는 상대도수)를 같은 폭의 막대로 나타낸 것. 범주형 자료의 분석에 적합. 연속형 자료는 히스토그램으로 표현	키나 몸무게와 같은 연속형 자료는 값의 종류가 너무 많아 점그래프로 나타낼 수 없으므로 몇 개의 구간으로 나누어 (막대 사이의 간격이 없는) 막대그래프의 형태로 그린 그림
꺾은선그래프	상관도
꺾은선그래프는 시간변수를 가로축, 분석변수를 세로축으로 하여 모든 분석변수의 값을 시간에 따라 선으로 이어주는 그래프입니다. 남녀성비 꺾은선 그래프는 1990년부터 2014년까지의 성비 변화를 나타냅니다.	상관도는 두 연속형 변수 간의 관계를 분석하는 데 효과적입니다. 키와 몸무게 상관도는 키가 증가함에 따라 몸무게도 증가하는 양의 상관관계를 보여줍니다.
시간변수를 가로축, 분석변수를 세로축으로 하여 모든 분석변수의 값을 시간에 따라 선으로 이어주는 그래프	한 변수를 가로축(원인), 다른 변수를 세로축(결과)으로 하여 각각의 관찰값을 좌표값으로 하여 나타낸 그래프. 두 개의 연속형 자료의 관계를 분석하는데 효과적임.



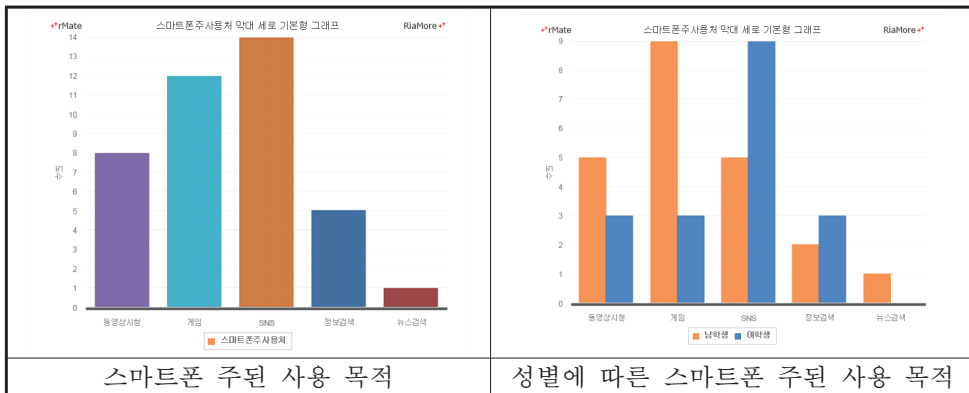
활동지 3

도전! Treasure Hunt(보물 사냥꾼)

○○중학교 학생회에서 조사한 자료를 표와 그래프를 이용하여 나타내어 보았습니다. 그런데 앞에서 살펴본 내용 이외에 더 찾아낼 수 있는 정보가 없을까요? 이번 활동에서는 수집한 자료 속에 숨어 있는 보물(정보)을 찾아보도록 합니다.



Activity 1 성별에 따른 자료 분석하기(1)



성별에 따라 스마트폰의 주된 사용 목적을 정리하면 남학생은 게임을 가장 많이 하고 동영상시청과 SNS는 공동 2등, 그 다음이 정보검색, 뉴스검색이란 것을 알 수 있습니다. 반면 여학생은 SNS가 가장 많이 나타났으며 동영상시청과 게임, 정보검색이 공동 2등을 한 것을 알 수 있습니다. 그리고 남학생과는 달리 뉴스검색은 아무도 선택하지 않았습니다.

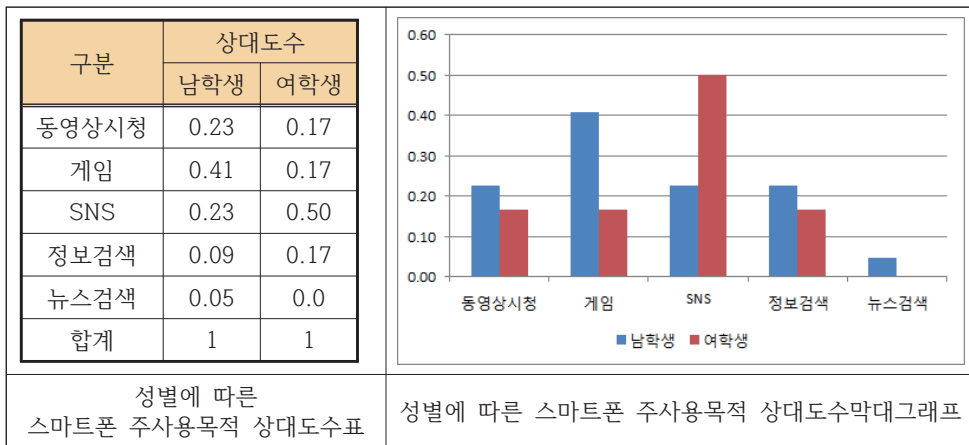
이렇게 성별에 따라 스마트폰의 주된 사용 목적을 정리해보니 전체와는 다른 해석이 가능합니다. 학년, 스마트폰 소유 여부, 스마트폰 사용자군에 따라 스마트폰의 주된 사용 목적을 정리해 본다면 또 다른 정보를 얻을 수 있을 것입니다.

- 대준 : 성별로 구분하니까 스마트폰의 주된 사용 목적이 달라지는 걸 알 수 있었어. 그리고 남학생은 게임이 9명, 여학생은 SNS가 9명으로 가장 높게 나타났어. 가장 많이 응답한 9명으로 같아
- 진혁 : 9명으로 같긴 하지만 그 의미는 다르지 않을까?

(1) 남학생과 여학생에서 9명의 의미를 생각해 봅시다. 둘은 같은 의미일까요?

통계적 소양을 위한 교수 학습자료

성별에 따른 스마트폰의 주된 사용처의 상대도수와 상대도수를 이용한 막대그래프를 나타내 보았습니다.



(2) 상대도수를 이용하여 성별에 따른 스마트폰의 주된 사용 목적을 설명하여 봅시다.

◦ 남학생

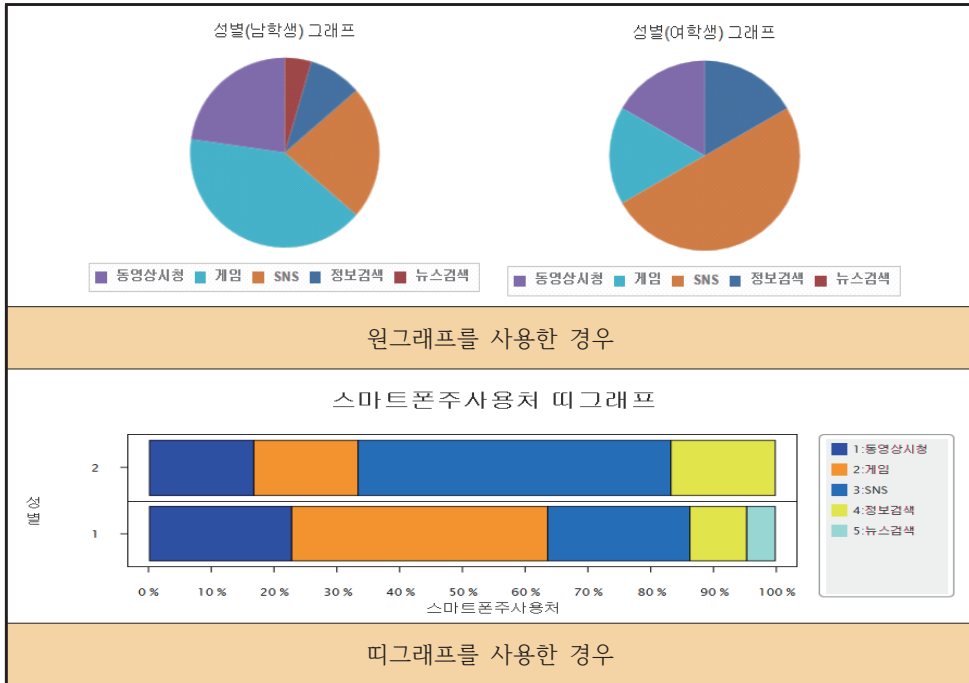
◦ 여학생

◦ 대준 : 그럼 도수와 상대도수는 각각 어떤 경우에 사용하는 게 좋을까?

(3) 도수 혹은 상대도수를 사용하느냐에 따라 자료의 해석이 달라집니다.

상대도수는 어떤 경우에 사용하는 것이 좋을까요?

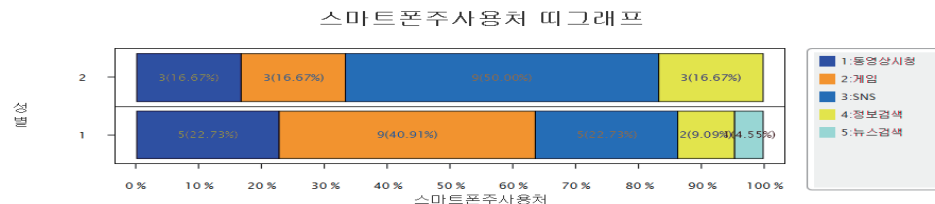
남녀에 따른 스마트폰의 주사용처를 원그래프와 띠그래프를 이용하여 나타낸 것입니다.



○대준 : 이 둘 중에 어떤 그래프로 표현하는 것이 더 좋을까?

(4) 성별에 따른 스마트폰 주된 사용목적은 어떤 그래프를 사용하는 것이 좋을까요?
그 이유는 무엇인가요?

(5) 다음과 같이 상대도수를 그래프에 나타내면 그래프를 해석하는데 도움이 될까요?





Activity 2 성별에 따른 자료 분석하기(2)

- (1) 앞선 활동을 참고하여 성별에 따른 자가진단점수를 가장 잘 설명할 수 있는 그래프가 무엇인지 생각해보고, 다음의 그래프 중 하나를 사용하여 나타내어 볼까요?

- 가장 적합한 그래프 :
○ 이유 :

원그래프	띠그래프
막대그래프 또는 히스토그램	

- 대준 : 남학생과 여학생의 자가진단점수는 어떤 차이가 있다고 기사로 정리하면 좋을까?

- (2) 정리된 그래프를 통해 알 수 있는 남학생과 여학생의 자가진단점수의 특성을 적어볼까요?

부록3. 대푯값 학생 활동지



활동지 1

통계에도 오류가 있다?

♣ 어느 나라의 1인당 국민총생산¹⁶⁾이 높으면 사람들은 그 나라에 사는 전체 국민의 소득 수준이 높다고 생각할 것입니다. 그러나 그 나라에 억만장자가 몇 명 있다면 국민들의 실제 소득은 그리 높지 않은데도 1인당 국민총생산이 높게 나타나게 될 것입니다.

통계에서는 평균이 갖는 이러한 한계를 개선하기 위하여 대푯값으로 중앙값이나 최빈값을 사용합니다. 중앙값은 자료를 크기가 작은 값부터 크기순으로 늘어놓았을 때 가장 중앙에 위치하는 값입니다. 만약 자료의 개수가 홀수라면 중앙값은 가운데 위치하는 값이고, 자료의 개수가 짝수라면 중앙값은 가운데 위치하는 두 값의 평균이 됩니다. 또한 최빈값은 통계 자료에서 가장 빈번하게 나오는 값으로 두 개 이상이 될 수도 있습니다.

우리는 시험 성적부터 국민총생산을 살펴보는 경우까지 다양한 통계 자료를 해석할 때 평균을 사용하고 있습니다. 생활 속에서 자료의 대푯값으로 평균을 사용할 때 적절하지 않았던 경우를 설명해 보세요.¹⁷⁾

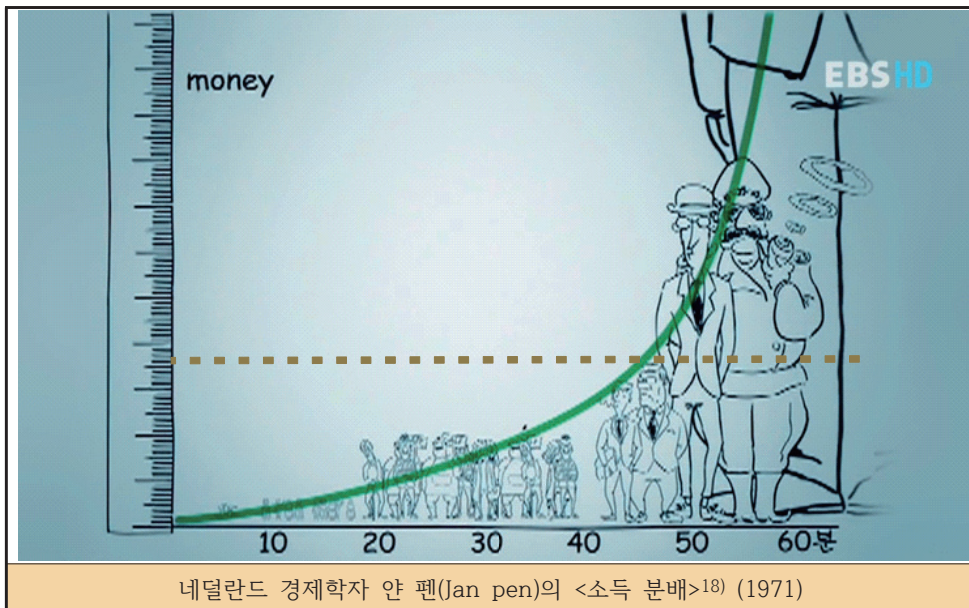
다양한 상황에서 얻어진 자료는 통계적인 처리 방식에 따라 서로 다른 해석을 낼 수 있습니다. 특히 실생활에서 얻어지는 통계 자료는 위의 세 가지 대푯값 중 한 가지만 선택함으로써 부적절한 결과를 유도할 수 있습니다. 이에 통계 자료를 다룰 때는 비판적인 관점에서 오류가 없는지 살펴보려는 태도가 중요합니다.

16) 국민총생산(Gross National Product) : 일정 기간 동안 한 나라의 국민이 국내외에서 새롭게 생산한 재화와 용역의 부가 가치 또는 최종재의 값을 화폐 단위로 합산한 것.

17) 안재찬. 명품수학(2005)

통계적 소양을 위한 교수 학습자료

네덜란드 경제학자 얀 펜(Jan pen)이 1971년도에 <소득분배>라는 책을 통하여 난쟁이의 행렬을 설명했습니다. 이는 소득의 순서대로 사람을 나열하면 어떤 형태가 될 것인가에 대한 질문의 답변이었습니다. 난쟁이의 행렬은 60분간 행진할 때 소득이 작은 사람부터 소득이 큰 사람으로 행진하며, 소득의 크기만큼 키가 정해진다는 것입니다. 즉 키가 작을수록 소득이 작고, 키가 클수록 소득이 크다는 뜻입니다. 처음에 행진이 시작되면 땅 속에 머리를 박고 행진을 시작하는데, 이들은 마이너스의 소득을 가진 사람들(채무자)입니다. 이후 1미터가 안 되는 사람들이 20분 가까이 행진하게 됩니다. 이 때 한 가지 발견할 수 있는 것은 여성들이 남성들보다 앞에 위치한다는 것입니다.



그리고 30분이 지나도 평균키의 사람들은 나타나지 않다가 48분이 되어서야 평균 키의 사람들이 나타나게 됩니다. 즉, 100명 중 80명 쯤 나타나는 사람이 평균 소득을 가진 사람입니다. 이후 6분을 남겨놓고 2미터 이상의 사람들이 나타납니다. 마지막 몇 초는 100미터에 달하는 거인들이 나타나고, 제일 마지막에는 석유왕 폴 게티가 나타납니다. 석유왕 폴 게티는 “만일 당신이 당신의 재산을 계산할 수 있다면 당신은 진짜 부자는 아닙니다.”라는 말로 유명합니다.¹⁹⁾ 얀 펜의 난쟁이의 행렬 역시 일상생활에서 흔히 사용하는 평균의 한계를 보여줍니다.

실제로 얀 펜의 난쟁이의 행렬에서 30분에 나타나는 사람들의 소득이 중앙값이므로 평균보다는 전체 소득을 대표하는 값으로 적절할 수도 있습니다. 또한 25분에 나타나는 값이 전체 소득 중 가장 많이 나타나는 최빈값이므로 이것이 전체 소득의 대푯값으로 더 적절하다고 볼 수도 있습니다.

18) EBS 동영상 (2010. 9. 13.)

19) <http://blog.naver.com/xyzkxx/10094279994>



Activity 1 평균의 함정

♣ 다음 표는 2014년 A지역 공무원 37명의 연간 봉급 총액과 2011년 LPGA에서 활동하는 한국계 골프선수 41명이 각자 획득한 상금의 액수입니다.

25	27	37	33	25	25
28	27	37	33	25	35
28	30	38	33	27	33
32	32	33	38	27	41
33	33	36	38	30	40
37	33	28	38	33	30
38					

[2014' A지역 공무원 37명의 연봉] (단위 : 백만원)

44	256	53	5	161	477
109	8	23	18	721	912
325	149	18	415	30	619
1357	886	22	72	16	5
245	165	408	852	77	28
87	262	94	365	21	82
182	350	42	42	628	

[2011' LPGA 한국계 골프선수 41명의 연간 상금 액수] (단위 : 천달러)

(1) 위의 자료를 보고 고은이는 다음과 같이 분석했습니다.

○ 고은 : A지역 공무원 연봉의 평균은 32(백만원)으로 대푯값으로 적절해.

여러분은 이 의견에 대해 어떻게 생각하나요?

(2) 고은이가 분석한 것을 보고 덕화는 다음과 같이 분석했습니다.

○ 덕화 : 골프선수들의 연간 상금 획득액 평균은 259(천달러)로 대푯값으로 적절해.

여러분은 이 의견에 대해 어떻게 생각하나요?

(3) 공유는 다음과 같이 분석했습니다. () 안에 적절한 이유를 써 넣어 보세요.

○ 공평 : 내 생각엔 공무원 연봉과 달리 공프선수의 상금은 중앙값이 자료의 대푯값으로 적절한 것 같은데?
() 때문이야.

(4) 위 활동을 통해서 알게 된 점이나 느낀 점에는 어떤 것이 있는지 써 보세요.



Activity 2 상황에 적절한 대풏값 고르기

♣ 청소년기에는 학업으로 인해 취미 생활을 갖기가 어렵고 스마트폰이나 컴퓨터에 접근하는 것이 쉬워 과도하게 이를 사용하는 경우가 있습니다. 하지만 청소년기에 과도한 컴퓨터의 사용은 학업에도 영향을 미칠 뿐만 아니라 신체, 정서, 인지 발달에 전반적인 영향을 미친다고 합니다. 이에 무엇보다 규칙적인 운동을 하는 것이 스트레스 해소에도 도움이 될 뿐만 아니라 오랜 시간 앉아서 공부하는 청소년기에 성장판을 자극하는 기회가 되기도 합니다. 청소년기에는 호르몬 분비가 활발하기 때문에 저녁 7시부터 9시까지 가벼운 걷거나 맨손 체조와 같은 운동을 지속적으로 하는 것이 중요합니다.

소은이는 우리 반 친구들의 운동 시간이 얼마나 되는지를 알아보기 위하여 한 달에 1시간 이상 운동하는 날이 며칠인지 조사하였습니다. 8명의 친구들이 운동한 시간은 다음과 같습니다.

2	3	7	10	12	26	30	30
---	---	---	----	----	----	----	----

8명 중 한 명인 도빈이는 한 달에 12일을 1시간 이상 운동합니다. 다음은 도빈이가 다른 7명의 친구들에 비해 운동을 자주하는 편인지에 대한 학생들의 대화입니다.

- 덕화 : 8명의 평균 운동 시간이 15일인데 이에 비하면 도빈이는 적게 운동하는 편이야.
- 선형 : 최빈값이 30일인데 12일은 턱없이 작잖아. 도빈이는 다른 학생들보다 적게 운동하는 것 같은데...

위 대화에 대한 자신의 의견을 말해 보세요.



활동지 2

균형점으로써의 평균

♣ 그림과 같이 양팔 저울에 무게가 같은 추를 왼쪽에 2개, 오른쪽에 4개 걸었더니 균형을 이루었습니다. 왜 이런 일이 일어날까요?



평균은 자료 값을 모두 더하여 자료의 전체 개수로 나눈 값, 즉 자료의 공평한 몫으로서의 의미를 가지고 있습니다. 또 한 가지는 수치 자료의 중심으로서의 의미를 가지고 있습니다. 자료 분포의 균형점으로써의 평균의 의미를 말합니다.

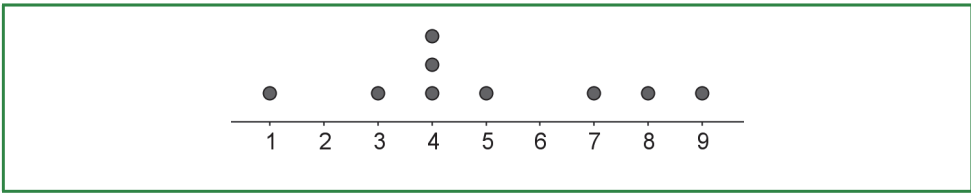
9명의 학생들에게 다음과 같이 질문했습니다.

집에서 애완동물을 몇 마리 키우나요?

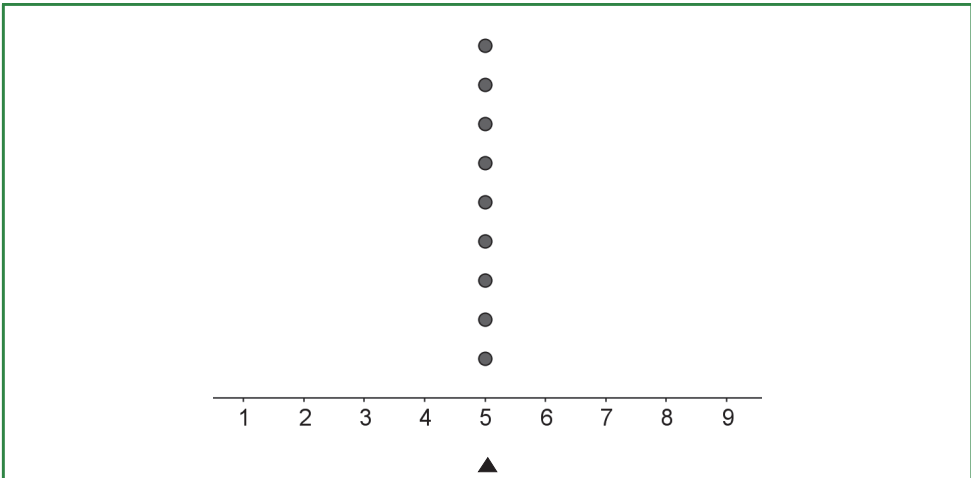
조사 결과 학생들은 다음과 같이 애완동물을 키우고 있다고 답했습니다.

1 3 4 4 4 5 7 8 9

학생들이 가지고 있는 애완동물의 수를 점도표로 나타내면 다음과 같습니다.

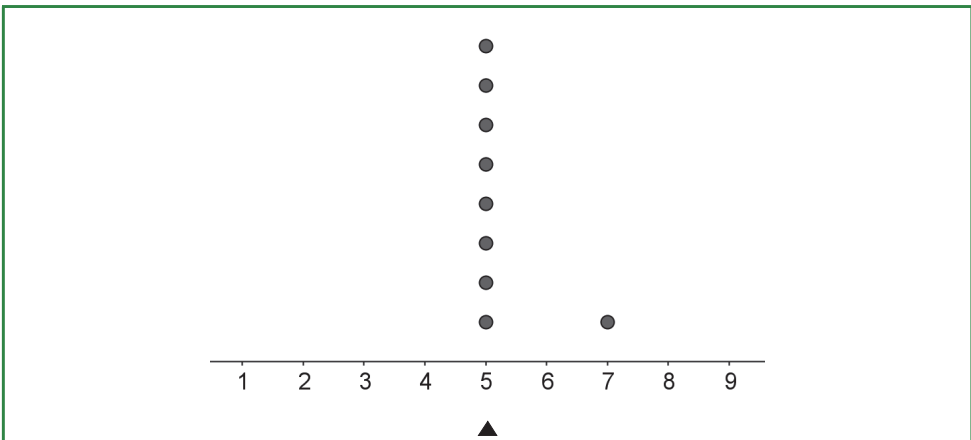


애완동물의 수를 모두 더하면 45마리가 됩니다. 45마리의 애완동물들을 9명의 학생들에게 골고루 나눠준다면 5마리씩 가지게 됩니다. 즉 애완동물 수의 평균은 5입니다. 9명의 학생들이 각각 5마리씩 가진 것을 점도표로 나타내면 다음과 같습니다.



만약 위의 점도표를 시소 위에 올려둔다면 중심이 5일 때 시소는 균형을 이룰 것입니다. 즉 이 점도표에서 균형점은 5입니다. 그렇다면 원래의 애완동물의 수로 나타낸 점도표의 균형점은 얼마일까요?

아래 그림처럼 5 위에 올려져있는 점 하나를 떼어 7위에 올리면 어떤 일이날까요?

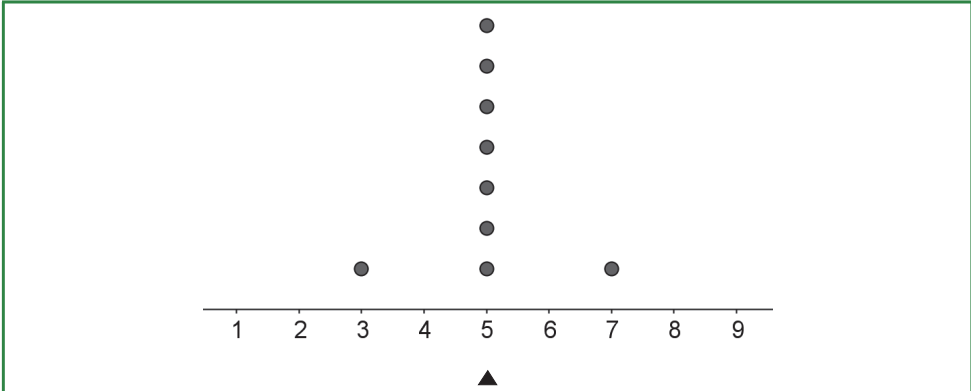


중심이 그대로 5라면 시소는 분명히 오른쪽으로 기울어 질 것입니다. 중심을

통계적 소양을 위한 교수 학습자료

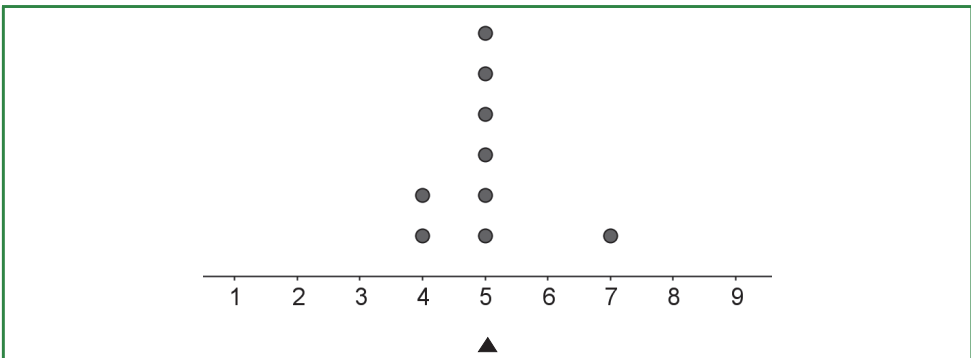
그대로 두고 시소의 균형을 맞추려면 5위에 있는 점들을 어떻게 해야 할까요?

시소 위에서 7은 5보다 2칸 오른쪽(큰 쪽)에 있기 때문에 5위에 있는 점 하나를 5보다 2칸 왼쪽(작은 쪽)인 3에 옮기는 것이 한 가지 해결 방법입니다. 다음 그림과 같이 나타낼 수 있습니다.



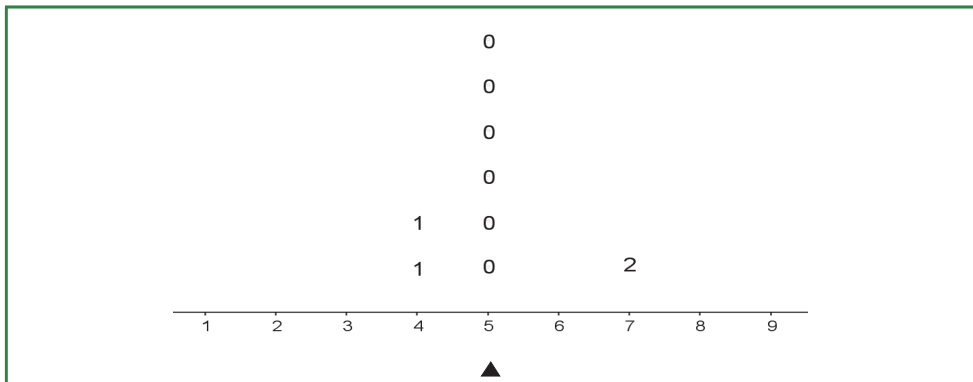
이제 시소는 중심 5에서 다시 균형을 이루게 됩니다. 그렇다면 다른 방법은 없을까요?

5위의 점을 한 개가 아니라 두 개 옮기려면 어디로 옮겨야 할까요? 또 다른 방법은 아래 점도표처럼 점 두 개를 4 위에 올리는 것입니다.



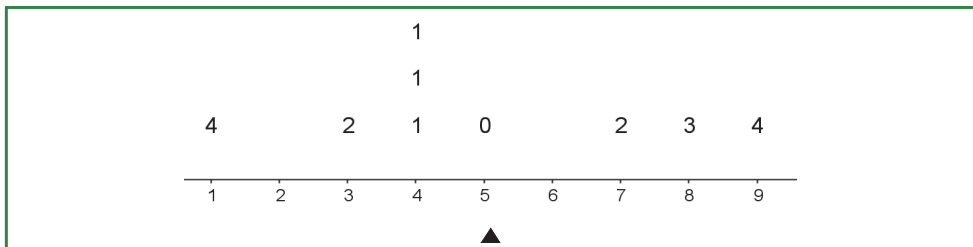
시소는 다시 중심 5에서 균형을 이루었습니다. 즉 시소의 균형점은 5입니다.

이때 각각의 점들을 중심 5에서 떨어진 거리로 바꾸면 다음과 같은 점도표가 됩니다.



중심인 5보다 왼쪽(작은 쪽)에 있는 두 수의 합(떨어진 거리의 합)은 5보다 오른쪽(큰 쪽)에 있는 한 개의 수와 같습니다. 그렇기 때문에 시소의 균형점은 5입니다.

이제 학생들이 키우는 애완동물의 수를 중심 5에서 떨어진 거리로 표시하면 아래와 같은 점도표로 나타낼 수 있습니다.



중심 5의 왼쪽(작은 쪽)의 값들의 떨어진 거리들의 합과 오른쪽(큰 쪽)의 값들의 떨어진 거리들의 합이 모두 9로 같습니다. 이러한 이유로 평균 5가 자료의 균형점이 됩니다.²⁰⁾

20) Guidelines for Assessment and Instruction in Statistics Education(GAISE) Report(2005). A PRE-K-12 Curriculum framework.

Activity 1 평균의 의미 1

♣ 오른쪽 표는 사과 30개의 무게(g)를 적은 표입니다.

232	203	308	216	239	266
242	259	265	225	286	251
257	277	257	229	244	260
278	263	248	242	240	294
208	312	210	295	271	220

(1) 자료의 평균을 구해 보세요.
(소숫점 아래 첫째 자리에서 반올림하여 자연수로 나타내세요.)

사과의 무게(g)

평균

(2) 아래 도수분포표를 완성하여 [활동지 1]의 모눈종이에 위 자료의 히스토그램을 그리고 오려 보세요.

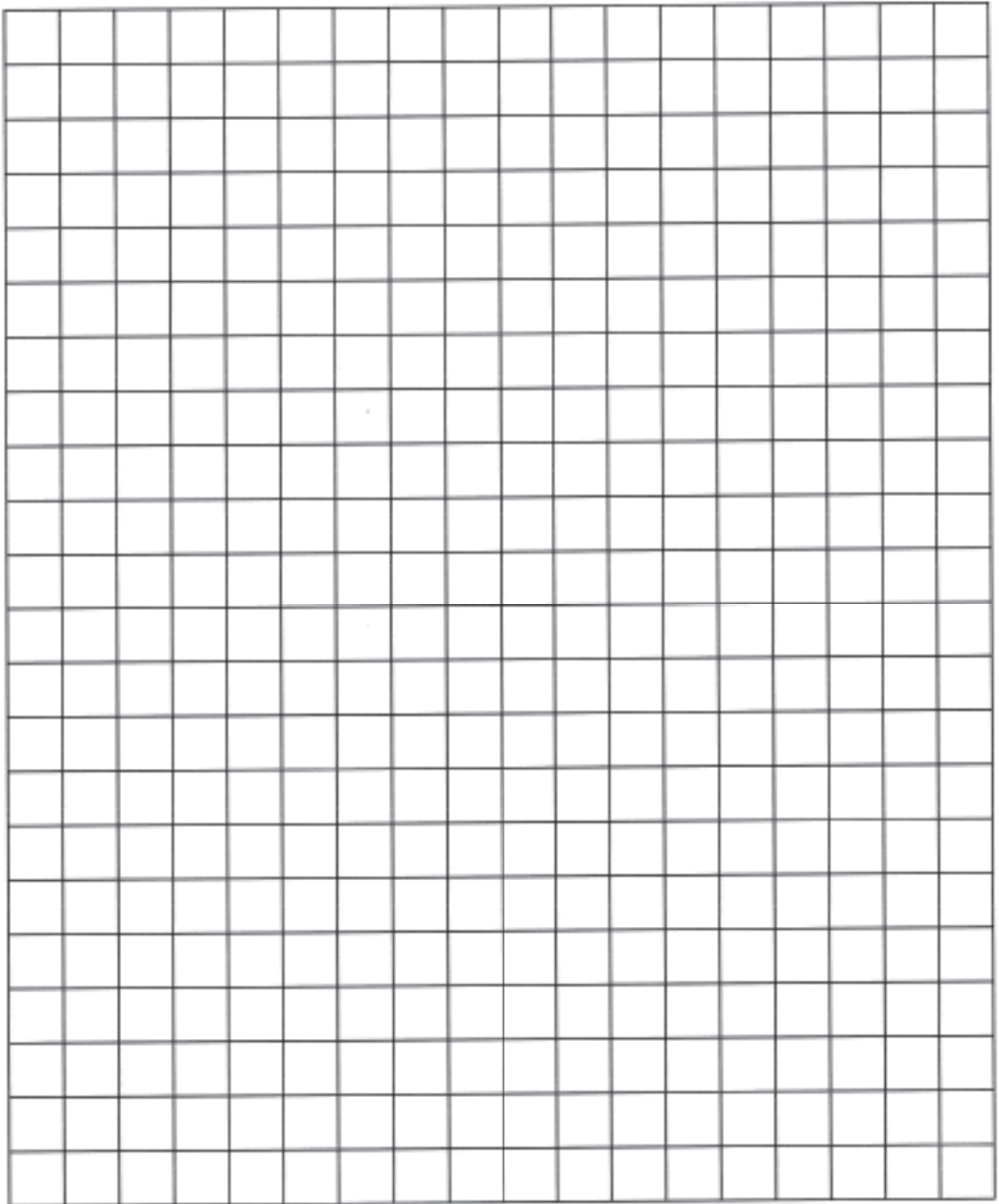
무게 (g)	도수(개)
200 ^{이상} ~ 220 ^{미만}	4
220 ~ 240	5
합계	30

<(3)의 실험이 끝난 후 여기에 히스토그램을 붙이세요>

(3) 연필이나 자 위에 오른쪽 사진과 같이 올려놓았을 때 떨어지지 않는 균형점을 찾아 표시해 보세요. 균형점으로 추측한 평균과 실제 평균은 어떻게 다른지 설명해 보세요.



♣ 아래 모눈종이에 히스토그램을 그리세요. 자르기 전에 계급의 범위를 적으세요.



Activity 2 평균의 의미 2

♣ 다음은 2011년 한 해동안 우리나라 20개 국립공원의 입장객 수를 조사하여 나타낸 도수분포표입니다. 이 도수분포표를 이용하여 구한 평균은 ?(만명)입니다.

입장객 수(만명)	도수(개)
0 ^{이상} ~ 300 ^{미만}	11
300 ~ 600	5
600 ~ 900	2
900 ~ 1200	1
1200 ~ 1500	1
합계	20

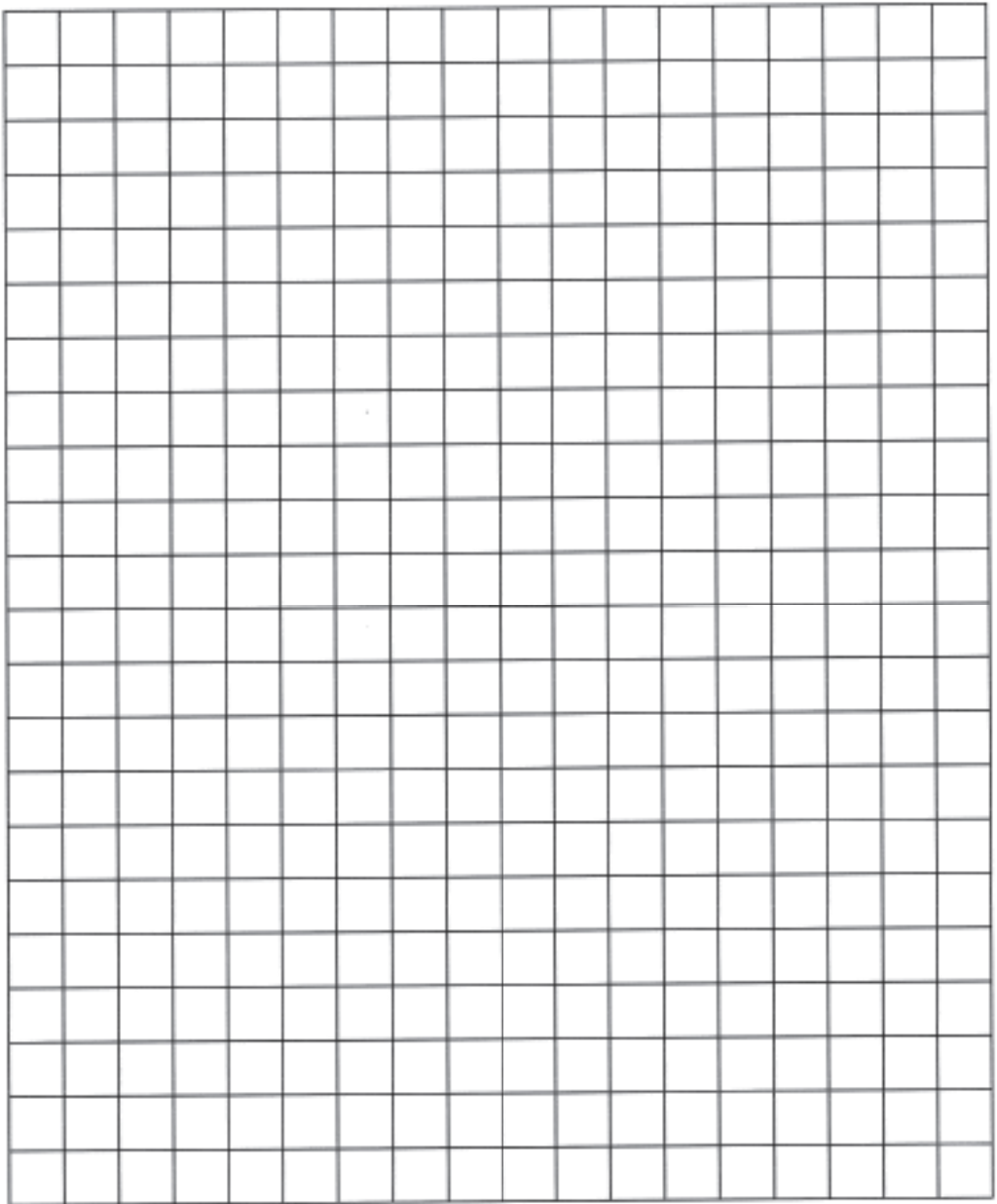
(1) 모눈종이에 히스토그램을 그리고 균형점을 예상해서 표시해 보세요.

<(2)의 실험이 끝난 후 여기에 히스토그램을 붙이세요>

(2) 히스토그램을 오려 연필 위에 올렸을 때 떨어지지 않는 균형점과 예상한 균형점을 비교해 보세요.

- (3) 균형점으로 추측한 평균은 입장객 수를 대표하는 대푯값으로 적절한가요?
이유를 적어 보세요.

♣ 아래 모눈종이에 히스토그램을 그리세요. 자르기 전에 계급의 범위를 적으세요.



연구 및 자료 개발 | 목원대학교 강현영 교수
전주교육대학교 고은성 교수
전남대학교 신보미 교수
영남중학교 정승호 교사
장대중학교 지영명 교사
지산중학교 류경민 교사
전주동신초등학교 이자미
서울대학교 수학교육과 탁병주 박사과정

자문위원 | 교육부 융합교육지원팀 안종선 연구사
부산교육대학교 이동환 교수
한림대학교 심송용 교수
신탄진초등학교 김정원 교사
영선중학교 고득환 교사
경희고등학교 홍창섭 교사

감수 | 충남대학교 이석훈 교수
상명대학교 김화경 교수
부산교육대학교 이동환 교수

통계적 소양을 위한 교수 학습자료

발행일 | 2016년 12월 25일

발행인 | 통계교육원장

발행처 | 통계교육원

(35220) 대전광역시 서구 한밭대로 713(월평동) 통계센터 5층 통계교육원

기획 | 류제정 · 이승구

©2016, 통계교육원

본 내용에 교사 및 학생들에게 자유 배포를 목적으로 제작되었으므로 무단 도용 및 상업적 활용을 금합니다.