

1. 통계

- 통계가 무엇인지 살펴본다.
- 자료의 구분에 대해 살펴본다.
- 통계 소프트웨어에 대해 살펴본다.

1.1 통계란?

<p>👁️ 생각열기</p>	<p>우리 학급 10명 학생의 신장은 어느 정도나 되는지 알아보기 위해 자를 이용하여 측정하였다. 측정된 신장 자료는 종이에 기록하였다.</p> <p style="text-align: center;">[표 1.1] 우리 학급 학생들의 신장 자료 (단위 cm)</p> <div style="border: 1px solid black; padding: 5px; text-align: center;"> 165 163 170 175 155 167 157 153 161 162 </div>
<p>탐구</p>	<p>우리 반 학생중에서 내장 (165cm)이 키가 큰 순으로 몇 번째에 있는지 알아보려면 어떻게 할까?</p>

- 학생들의 신장을 숫자로 된 **자료**라 부른다. 이 예에서는 학생의 수만큼 자료가 생기는데 우리 사회에서는 많은 양의 자료가 생성되어 왔다. 컴퓨터의 발전과 더불어 자료의 수는 앞으로 더욱 증가될 예정이다.
- 내 신장의 상대 위치를 파악하기 위해서는 이 자료를 신장이 큰 사람부터 내림차 순으로 [표 1.2]와 같이 정리한다.

[표 1.2] 신장이 큰 순으로 정리한 학생들의 신장 자료 (단위 cm)

175 170 167 165 163 162 161 157 155 153

- 이 자료를 보면 쉽게 내 신장이 위에서 4번째 학생임을 알 수 있다.
- 이렇게 자료를 수집해 정리, 요약한 후 유용한 정보를 얻어내는 것을 **통계**라고 한다. 한 나라는 국민들의 의식주를 해결하기 위해 인구수, 가구수, 식량 등의 통계가 필요하다. 이와 같은 통계는 5천년 전 이집트 문명시대부터 기록이 남아있다.
 - 현대 정보화 사회가 더욱 진행되면서 자료의 크기가 엄청나게 커졌는데 이를 **빅데이터**라 부른다. 빅데이터의 대표적인 예는 다음과 같다.
 - 구글의 검색기록 자료
 - 스마트폰의 네이버나 카톡 자료
 - 인터넷에 접속하고 나간 기록 자료
 - 통신회사의 통화기록 자료
 - 미래의 정보화 사회에서는 어떻게 빅데이터에 대한 통계를 잘하여 과학적인 의사결정을 하느냐에 따라 한 개인, 회사, 국가의 성패가 달라질 수 있다. 빅데이터를 분석하여 현실에 응용하는 통계의 활용 예는 많이 있다.
 - 구글의 검색 엔진에 자동차 구입에 관한 질문을 조사하여 다음 달 미국서

판매되는 자동차 모델의 수를 예측하였다.

- 한 온라인 쇼핑몰은 고객의 웹 기록을 분석하여, 회원 고객이 어떤 취향을 가지고 어떤 제품에 관심이 있는지 파악하여 고객 개개인에 맞는 맞춤형 광고를 하여 매출이 증가하였다.
- 한 원유 탐사회사에서 테라바이트 규모의 지질학 자료를 분석해 원유 시추의 성공률을 높였다.
- 미국의 한 대학에서 온라인 수업에서 학생들이 시스템에 클릭하는 정보를 분석하여 학생 개개인의 학습 성과를 모니터링하고 학생의 이해도에 맞춘 수준별 수업 내용을 제안하고, 향후 수강할 과목 등을 학생별로 제안하였다. 이 결과 전공별 학위 취득률이 많이 향상되었다.

1.2 자료의 구분

☞ 생각열기	우리 학급 10명 학생의 성별을 조사하면 ‘남’ ‘여’ 이고, 신장을 조사하면 160cm 등이다. 성별과 신장 자료는 속성이 다르다.
탐구	1) 성별과 신장은 속성이 다른데 자료를 구분하는 적당한 방법이 있을까? 2) 자료의 종류에 따라 분석 방법이 다를까?

- 자료는 관심의 대상이 되는 사물의 속성을 관찰하거나 측정된 값들이다. 이러한 사물이나 사건의 속성을 **변수** 또는 **변량**이라고 한다. 예를 들어, 어느 초등학교 한 학급의 성별과 신장을 측정하였다면 여기에는 두개의 변량(성별, 신장)이 있다. 성별에 대한 측정값은 ‘남’, ‘여’, ‘여’, ‘남’, ... 과 같은 형태이고, 신장에 대한 측정값은 160cm, 155cm, 158cm, 165cm, ... 와 같은 형태일 것이다.
- ‘성별’과 같은 변량의 자료를 **질적 자료**, 신장과 같은 변량의 자료를 **양적 자료**로 구분한다. 자료를 구분하는 이유는 자료의 종류의 따라 처리하는 방법과 분석 방법이 다르기 때문이다. 2장은 질적 자료의 분석에 사용하는 그래프에 대해 살펴보고, 3장은 양적자료의 분석에 사용하는 그래프를 살펴본다.
- 성별과 같은 질적 변량은 대개 몇 개의 범주 형태를 갖는데 이를 **범주형 자료**라고도 한다. 범주형 자료는 각 범주별로 도수를 조사하여 도수분포표를 만들고 막대그래프, 원그래프, 띠그래프를 만들어 그린다(2장). 표 1.3은 한 학급 성별에 대한 도수분포표의 예이다. ‘남’이 6명이고 ‘여’가 4명임을 보여준다.

표 1.3 한 학급의 성별 도수분포표

성별	학생수
남	6
여	4

- 단어와 같은 질적 자료는 각 단어에 대한 빈도수를 조사하여 단어구름으로 만들어 분석한다.
- 양적 자료는 평균 및 표준편차를 계산하고 줄기와 잎 그림, 히스토그램, 점그래프 등을 이용하여 시각화 한다(3장).

1.3 『eStat』 소프트웨어

☞ 생각열기	[표 1.1]의 신장 자료를 측정해 종이에 적고 이를 내림차순으로 정리하였다.
탐구	컴퓨터를 이용하는 방법은 없을까? 있다면 어떤 소프트웨어를 사용하여야 하나?

- 작은 학급 자료라면 수작업으로도 자료를 정리해 분석할 수 있으나 자료가 많아지면 컴퓨터의 도움이 없이는 자료의 분석이 불가능하다. 특히 빅데이터 분석을 위해서는 전문적인 분석 프로그램을 많이 가지고 있는 **통계 소프트웨어**가 반드시 필요하다. 현재 자료 분석을 위해서는 SAS, SPSS, R과 같은 통계 소프트웨어가 많이 사용되고 있다. 하지만 이들 소프트웨어는 초보자가 배우기는 쉽지 않고, SAS와 SPSS는 상업용이어서 엄청난 고가이다. 그리고 이러한 소프트웨어는 자료 분석의 핵심이고 통계 교육에 필요한 모듈의 기능은 거의 없다.
- 『eStat』은 통계를 초등생부터 대학 및 일반인까지 쉽게 교육하기 위하여 만든 교육용 소프트웨어이다. 자료가 주어지면 단지 마우스 클릭만으로 그래프를 그릴 수 있고, 동적인 그래프를 경험할 수 있다. 『eStat』은 각급 교과서에 있는 많은 예를 포함하고 있으며, 웹 기반이어서 사용자들은 언제 어디서나 PC, 태블릿, 또는 스마트폰으로 이용할 수 있다. 『eStat』은 무료로 서비스하고 있고 다국적 언어를 지원하며 현재 한국어, 영어를 포함한 20개 언어가 가능하다.
- 『eStat』 시스템은 현재 통용되는 많은 웹브라우저 중에서 구글(Google)사의 크롬(Chrome)에서 100% 잘 작동한다. MS 엣지(Edge)와 같은 브라우저에서도 『eStat』이 작동은 되나 일부 기능이 안 될 수 있다.
- 이 책은 『eStat』에서 초등 및 중학생을 대상으로 개발된 『eStatH』에 있는 몇 가지 모듈에 대해 설명한다. 웹브라우저 크롬을 열어 주소창에 www.estat.me 를 입력한 후 아이콘 『eStatH』를 클릭하면 <그림 1.1>과 같은 화면이 나타난다.



eStatH 중 / 고 통계교육 SW

- II 막대그래프 - 원그래프 - 띠그래프
- II 단어 구름
- III 줄기와 잎 그림
- III 히스토그램 - 도수분포표
- III 도수분포다각형 - 상대도수 비교
- IV 점그래프 - 평균 / 중앙값
- IV 산점도 - 상관계수
- IV 상관계수
- IV 상관계수 - 회귀선 실험
- V 순열 - 조합
- V 대수의 법칙 시뮬레이션
- V 이항분포실험
- V 이항분포
- V 정규분포실험
- V 정규분포
- VI 모집단과 표본 시뮬레이션
- VI 표본평균의 표집분포 시뮬레이션
- VI 모평균 신뢰구간 시뮬레이션
- VI 신뢰구간



전자책 링크
[eStat 통계](#)

이메일: jjlee@ssu.ac.kr © eStat.org, Korea



<그림 1.1> 『eStatH』 메뉴의 구성