


# 1. Statistics

- Examine what statistics are.
- Examine the classification of data.
- Examine statistical software.

## 1.1 What is Statistics?

 Think	<p>The height of 10 students in our class was measured to see how tall they were. The measured height data were recorded on paper as follows:</p> <p style="text-align: center;">[Table 1.1] The height of 10 students in our class (unit cm)</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>165</td><td>163</td><td>170</td><td>175</td><td>155</td><td>167</td><td>157</td><td>153</td><td>161</td><td>162</td> </tr> </table>	165	163	170	175	155	167	157	153	161	162
165	163	170	175	155	167	157	153	161	162		
Explore	<p>– How can I find out where my height (165cm) is located among students in my class in descending order of height?</p>										

- Students' heights are called numerical **data**. In this example, the data was generated as many as the number of students. In our society, a large amount of data were generated and the number of data is expected to increase further with the development of computer technologies in the future.
- In order to understand the relative position of my height among students, arrange the data in descending order from taller to smaller as shown in [Table 1.2].

[Table 1.2] Height data in descending order (unit cm)

175	170	167	165	163	162	161	157	155	153
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

From this data, it is easy to tell that my height of 165cm is the 4th from the top.

- **Statistics** is the process of collecting, organizing, and summarizing data in this way and obtaining useful information by analysis. A country needs statistics such as the number of people, the number of households, and the amount of food in order to solve eating, clothing and sheltering of its people. Such statistics have been recorded since the time of Egyptian civilization 5,000 years ago.
- As the modern information society progresses, the size of data has grown tremendously, which is called Big Data. Typical examples of big data are as follows:
  - Google's search history data
  - SNS data from your smartphone
  - Records of access to and out of the Internet

- Telecommunication company's call log data
- In the information society of the future, the success or failure of an individual, company, or country can be different depending on how they make scientific decisions with reasonable statistical information on big data. There are many examples of statistics applied to reality by analyzing big data.
  - Predict the number of car models sold in the United States next month by surveying Google's search engine questions about car purchases.
  - An online shopping mall analyzed customers' web records to find out what products the member customers have and what products they were interested in, and then customized advertisements tailored to each customer, resulting in increased sales.
  - An oil exploration company increased the success rate of oil drilling by analyzing terabytes of geological data.
  - A university in the United States analyzed the information that students click on the system in an online class to monitor each student's learning performance, suggested class content for each level tailored to the student's level of understanding, and suggested subjects to take in the future for each student. As a result, the degree acquisition rate by major improved significantly.

## 1.2 Data Classification

☞ Think	If the gender of the 10 students in our class is investigated, they are 'male' or 'female', and if the height is measured, it is 160cm, etc., gender and height data have different attributes.
Explore	1) Is there an appropriate way to classify data? 2) Does the analysis method differ depending on the type of data?

- Data are values that observe or measure the attributes of objects of interest. The attribute of such an object or event is called a **variable**. For example, if the gender and height of an elementary school were measured, there were two variables (gender and height). Gender measurements are in the form as male, female, female, male, .... and height measurements are

- 160cm, 155cm, 158cm, 165cm, ...
- Variable such as 'gender' is classified into **qualitative data**, and variable such as height is classified into **quantitative data**. The reason for classifying data is that processing and analyzing methods are different depending on the type of data. Chapter 2 discusses visualization of qualitative data, and Chapter 3 discusses visualization of quantitative data.
  - Qualitative variable such as gender usually have several categories, which are also called a **categorical variable**. For a categorical variable, the frequency of each category is investigated, a frequency table is created, and a bar graph, pie graph, and band graph are drawn for analysis (Chapter 2). Table 1.3 is an example of a frequency table for gender in a class. It shows that there are 6 'male' and 4 'female' students.

[Table 1.3] Frequency table for gender in a class

Gender	Number of students
Male	6
Female	4

- Qualitative data such as words are analyzed by examining the frequency of each word to form a word cloud.
- For quantitative data, the mean and standard deviation are calculated and visualized using stem and leaf plot, histogram, and dot graph (Chapter 3).

### 1.3 『eStat』 Software

☞ Think	[Table 1.2] arranges the height data in descending order on a paper
Explore	Is there any way to use a computer? If yes, which software should I use?

- Small class data can be manually organized and analyzed, but if there are more data, it is impossible to analyze data without the help of a computer. In particular, **statistical software** with many professional analysis programs is essential for big data analysis. Statistical software such as SAS, SPSS, and R are now widely used for data analysis. However, these software are not easy for beginners to learn, and SAS and SPSS are commercial and are very expensive. In addition, these software are focused

- on data analysis and there are few modules required for statistical education.
- 『eStat』 is a free educational software designed to teach statistics easily from elementary school students to college students and general public. Given the data, you can draw a graph with just a click of the mouse, and experience a dynamic graph. 『eStat』 includes many examples from textbooks of each level, and because it is web-based, users can use it anytime, anywhere with a PC, tablet, or smartphone. 『eStat』 is provided free of charge and supports multiple languages, and currently 20 languages, including Korean and English are available.
  - 『eStat』 system works 100% well with Google's Chrome among many web browsers currently in use. In browsers such as MS Edge, 『eStat』 works, but some functions may not work.
  - This book is focused on modules in 『eStatH』 developed for elementary and middle school students in 『eStat』. Open the web browser Chrome, enter [www.estat.me](http://www.estat.me) in the address bar and click the icon 『eStatH』, then the screen as shown in <Figure 1.1> appears.

**eStatH - Middle / High School Statistics Education SW**

Middle School Statistics	High School Statistics
<a href="#">eBook Link</a> 	<a href="#">eBook Link</a> 
II Bar Graph - Pie Graph - Band Graph	I Permutation - Combination
II Line Graph	I Pascal Triangle - Binomial Theorem
II Word Cloud	II Statistical Probability
III Stem & Leaf Plot	II Addition Rule of Probability
III Histogram - Frequency Table	II Conditional Probability
III Frequency Polygon - Relative Freq Comparison	III Discrete Distribution
IV Dot Graph - Mean / Median	III Binomial Experiment
IV Scatterplot - Correlation Coefficient	III Binomial Distribution
IV Correlation Coefficient	III Law of Large Number
IV Correlation Coefficient - Regression Experiment	III Normal Distribution Comparison
	III Normal Distribution
	IV Random Number Generator
	IV Population vs Sample
	IV Dist of Sample Means
	IV Confidence Interval Simulation
	IV Population Mean Confidence Interval
	<b>Contact: <a href="mailto:jjlee@ssu.ac.kr">jjlee@ssu.ac.kr</a></b>

&lt;Figure 1.1&gt; 『eStatH』 menu