

Introduction to Statistics and Data Science using *eStat*

Chapter 6 Sampling Distribution and Estimation

6.1 Sampling Methods

Jung Jin Lee

Professor of Soongsil University, Korea

Visiting Professor of ADA University, Azerbaijan

6.1 Simple Random Sampling

- **Inferential statistics :**
 - ⇒ **Estimate population characteristics using samples**
- Some difference between population characteristics and sample characteristics.
- To reduce these differences, sampling methods have been developed.
 - ⇒ **Simple random sampling** : all elements of population have the same probability of being selected.

6.1 Simple Random Sampling

- Tools to ensure that each element of the population is selected equally.
 - ⇒ **random number table** : numbers from 0 to 9 without special regularity
 - ⇒ **computer random number generator** : uniform $[0,1]$ distribution
- **With replacement sampling** : Include an element extracted once again
- **Without replacement sampling** : does not include the extracted elements
⇒ in practice, almost all sampling is made without replacement.

6.1 Simple Random Sampling

[Ex 6.1.1] A class has 50 students. Select three of these students as a sample without replacement by using 『eStatU』.

〈Answer〉

- A student's list must first be made and serial numbers must be assigned from one to fifty.
- To extract students, select 'Uniform Random Number' in 『eStatU』. Enter (1 and 50) at the box of 'Uniform', check 'Integer', enter 10 at the box of 'Number of Data' and click [Execute].
- Since there is no overlap of 6, 13 and 45, and you will choose these student numbers as a sample.
- The generated random number may have the same numbers. If it is a sampling without replacement, discard the following same number.

Uniform Random Number

Menu

Uniform (,) Real Integer

Number of Data n = (≤ 100) Decimal Digit 0 ≤ ≤ 4

Execute

Table Save

id	Uniform(1 , 50) Random Number
1	6
2	13
3	45
4	17
5	27
6	43
7	19
8	34
9	4
10	30

6.1. 1 Sample Survey and Other Sampling Methods

- **Difference between the estimated value using the sample and the characteristic value of the population is called a **sampling error**.**
- **Each element of the population is called an **elementary unit** or an **observational unit****
- **what is selected to construct a sample is called a **sampling unit**.**
- **A table of the sampling units is called a **sampling frame**.**
- **Describing how to select samples using the sampling frame is called a **sampling design**.**

6.1. 1 Sample Survey and Other Sampling Methods

▪ **Steps for Sample Survey**

Step 1. The objective of the sample survey must be clearly established.

Step 2. Check general matters related to the sample survey.

Step 3. Do the sampling design which includes the most important things such as sampling unit, sample size, sampling method, etc. Also, the cost required for each step is calculated.

Step 4. Conduct the designed sample survey.

Step 5. Analyze and summarize the data obtained from the sample survey.

6.1. 1 Sample Survey and Other Sampling Methods

- **Stratified Sampling**

- **Divides the population into an appropriate number of **strata**.**
- **Selecting a sample of a fixed size from each stratum using simple random sampling method.**

6.1. 1 Sample Survey and Other Sampling Methods

▪ **Cluster Sampling**

- **When trying to investigate entertainment expenditures by household living in Seoul, if a household (**cluster**) is the sampling unit, all basic regional administrative units in Seoul are listed and then some regional units are selected by simple random sampling.**
- **In the selected regional units, some households are selected by simple random sampling and then survey all members in the household.**

6.1. 1 Sample Survey and Other Sampling Methods

▪ **System Sampling**

- **Suppose you want to get a quick estimate of total sales by examining 2% of a department store's sales slips, approximately 1000 slips each week. 20 sales slips need to be investigated.**
- **If the sales slips are filed and stored in sales order, one slip is selected between the first slip and the 50th slip and then continue selections from every 50 slips. For example, if it was the 7th slip, the next is the 57th, 107th, 157th, ... , 957th slip.**
- **This is called a systematic sampling. Here, 50 is called the **sampling interval**, and the starting point 7 is called the **random starting point**.**



Thank you