



statistics

Lecture two

Symbols and sampling techniques

First stage

Mathematical Symbols and Expressions

Certainly, there is a need to understand and comprehend how to perform analyses for the topics of statistics. This requires familiarity with some terminology, symbols, mathematical formulas, and some of their rules and possible uses. Some of these will be presented below.

1. Summation Symbol (Σ)

If we want to represent the result of adding a number of values of a random variable X , such as $(x_1, x_2, \dots, x_n)$, we can use a mathematical notation called the summation symbol.

This symbol is read as the sum, as follows:

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

It is read as: the sum of x_i where i takes values from 1 to n .

Example (1): If x_i represents the $x_1 = 13, x_2 = 11, x_3 = 14, x_4 = 12$ find: $\sum_{i=1}^4 x_i$

Sol/

$$\begin{aligned} \sum_{i=1}^4 x_i &= x_1 + x_2 + x_3 + x_4 \\ &= 13 + 11 + 14 + 12 \\ &= 50 \end{aligned}$$

• Rule 1

Since the constant quantity C does not have the index of change i , its summation is equal to C multiplied by n . In other words, the result of adding C n times is:

$$\sum_{i=1}^n c = c + c + \dots + c$$

(n times)

Example (2): What is the result of each of the following quantities?

1) $\sum_{i=1}^n 14$, 2) $\sum_{i=1}^7 14$, 3) $\sum_{i=1}^3 6$

1. $\sum_{i=1}^n 14 = 14n$
2. $\sum_{i=1}^7 14 = 7(14) = 98$
3. $\sum_{i=1}^3 6 = 3(6) = 18$

- **Rule 2:**

The summation symbol $\sum$ enters or distributes over the terms of an algebraic expression that represents the sum or the difference of two or more variables. That is:

$$\sum_{i=1}^n (x_i \mp y_i) = \sum_{i=1}^n x_i \mp \sum_{i=1}^n y_i$$

Example(3)/ let $x_i = 17, 15, 10, 7, 9$ and $y_i = 10, 9, 11, 8, 4$ find $\sum_{i=1}^5 (x_i \mp y_i)$

Sol/

$$\begin{aligned} \sum_{i=1}^5 (x_i \mp y_i) &= \sum_{i=1}^5 x_i \mp \sum_{i=1}^5 y_i \\ &= 17 + 15 + 10 + 7 + 9 \mp 10 + 9 + 11 + 8 + 4 \\ &= 58 \mp 42 \\ &= 58 + 42 = 100 \\ \text{or } &= 58 - 42 = 16 \end{aligned}$$

- **Rule 3**

If C is a constant quantity, then:

$$\sum_{i=1}^n Cx_i = C \sum_{i=1}^n x_i$$

Example(4)/ if the values of the variable x are as follows: $x_i = 17, 15, 10, 7, 9$ find $\sum_{i=1}^5 8x_i$

Sol/

$$\sum_{i=1}^5 8x_i = 8 \sum_{i=1}^5 x_i = 8(58) = 464$$

Example(5)/ Use the data from example(3) above and calculate the amount $\sum_{i=1}^5 \frac{x_i}{4}$?

Sol/

$$\sum_{i=1}^5 \frac{x_i}{4} = \frac{1}{4} \sum_{i=1}^5 x_i = \frac{1}{4}(58) = 14.5$$

2. Multiplier symbol π

If it is wanted to represent the product of a number of values of a random variable, let x it be $(x_1, x_2, \dots, x_n)$, in a mathematical form, the multiplication symbol π can be used as follows:

$$\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \dots \cdot x_n$$

Example(6)/ let $x_i = 4, 4, 3, 2, 6$ find 1) $\prod_{i=1}^n x_i$ 2) $(\prod_{i=1}^n x_i)^2$ 3) $\prod_{i=1}^n x_i^2$

Sol/

$$1) \prod_{i=1}^5 x_i = x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5$$

$$= 4 \cdot 4 \cdot 3 \cdot 2 \cdot 6 = 576$$

$$2) (\prod_{i=1}^n x_i)^2 = (x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5)^2$$

$$= (4 \cdot 4 \cdot 3 \cdot 2 \cdot 6)^2 = (576)^2 = 331776$$

$$3) \prod_{i=1}^n x_i^2 = x_1^2 \cdot x_2^2 \cdot x_3^2 \cdot x_4^2 \cdot x_5^2$$

$$= 4^2 \cdot 4^2 \cdot 3^2 \cdot 2^2 \cdot 6^2 = 331776$$

Probability Samples:

This refers to applying the principle of probability to every unit within the statistical population, determining whether or not it will be included in the selected sample. They include several types:

Simple Random Sample:

This refers to the process of selecting a random sample from the study population. This method ensures that every element of the population has an equal chance of being included in the sample, and the population must be homogeneous.

For example, if we were to study the reasons for smoking among females

Assuming the study population is homogeneous and finite, and that the number of individuals in the population is N , and assuming that this study requires selecting a sample of size n , then the probability of each individual appearing in the sample and the number of possible samples from this population is r , such that:

$$r = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Example 7/ A population of $N = 5$, $n = 2$ what is the number of possible samples?

$$\begin{aligned} r &= \binom{5}{2} = \frac{5!}{2!(5-2)!} \\ &= \frac{5 \cdot 4 \cdot 3!}{2 \cdot 1(3)!} = 10 \end{aligned}$$

Stratified Random Sampling:

Samples selected using this method are considered among the best and most accurate types of sampling in representing the population. We often observe that the elements of a statistical population are heterogeneous in terms of the characteristic relevant to the research topic. For example, if we were to study the academic level of students in the College of Administration and Economics, we would notice that this population is heterogeneous in terms of academic specialization. There is a group of students specializing in administration, another in economics, another in accounting, and another in statistics. In such a case, it is not possible to draw a simple random sample that represents this population. Instead, the population must be divided into four strata, each stratum being considered a homogeneous population in itself. From each stratum, a simple random sample is selected that is proportional to the size of that stratum of the population, such that the sum of the elements of these four

samples constitutes the required sample size for the study. Assuming that the study population consists of N elements, and that this population can be divided into L strata according to a homogeneity criterion suitable for that study, such that each stratum contains homogeneous elements, as follows:

1	2	3	...	L	strata sequence
N_1	N_2	N_3	...	N_L	Number of observations per stratum
$N = N_1 + N_2 + \dots + N_L$					where

The sample sizes to be selected from the layers are determined according to a method called the proportional distribution method, where we find: The sample sizes to be selected from the layers are determined according to a method called the proportional distribution method, where we find:

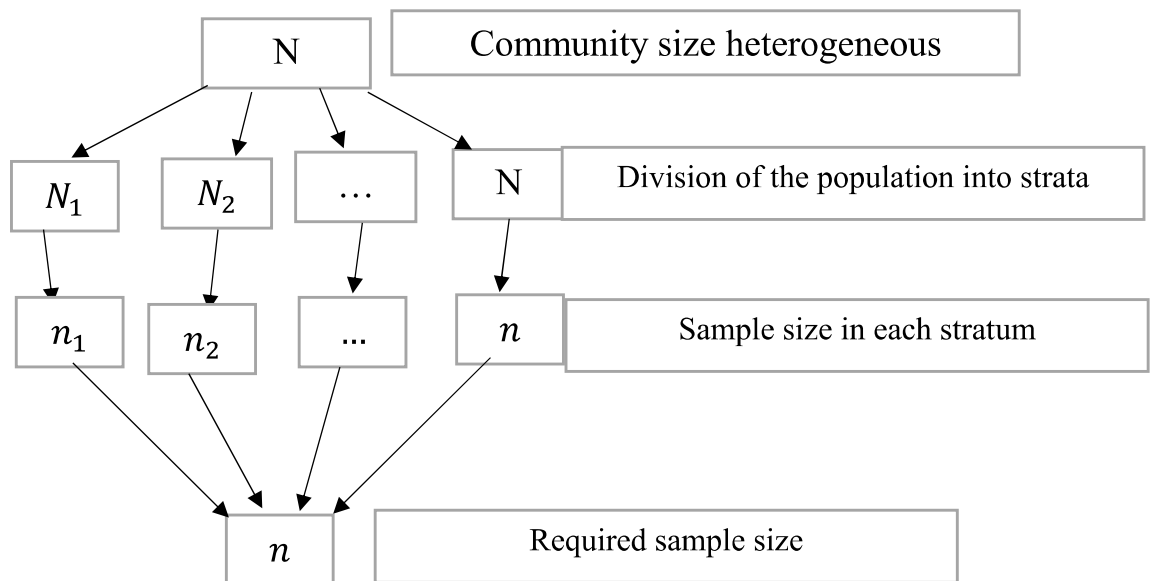
1) layer weight : $W_h = \frac{N_h}{N}$

where $h = 1, 2, \dots, L$, represents the layer sequence.

The required sample size to be selected from layer h is

2) $n_h = W_h \cdot n$

After knowing the number of items to be selected from each layer, a simple random sampling method can be used to select the sample items for each layer, which together constitute the required sample size. After identifying the sample items, data are taken from each of them.



Example 8/In a research project for students of the College of Administration and Economics, the required sample size was 110 students. The number of students in the Administration Department was 600, the Economics Department 800, the Statistics Department 500, and the Accounting Department 300. The task is to determine the appropriate sample sizes to be selected from the students of these four departments.

$$n = 110 \quad N_1 = 600 \quad N_2 = 800 \quad N_3 = 500 \quad N_4 = 300$$

$$\begin{aligned} N &= N_1 + N_2 + N_3 + N_4 \\ &= 600 + 800 + 500 + 300 \\ &= 2200 \end{aligned}$$

$$W_1 = \frac{N_1}{N} = \frac{600}{2200} = \frac{6}{22}$$

$$W_2 = \frac{N_2}{N} = \frac{800}{2200} = \frac{8}{22}$$

$$W_3 = \frac{N_3}{N} = \frac{500}{2200} = \frac{5}{22}$$

$$W_4 = \frac{N_4}{N} = \frac{300}{2200} = \frac{3}{22}$$

$$n_1 = W_1 \times n = \frac{6}{22} \times 110 = 30$$

$$n_2 = W_2 \times n = \frac{8}{22} \times 110 = 40$$

$$n_3 = W_3 \times n = \frac{5}{22} \times 110 = 25$$

$$n_4 = W_4 \times n = \frac{3}{22} \times 110 = 15$$

$$n = n_1 + n_2 + n_3 + n_4$$

$$n = 30 + 40 + 25 + 15$$

$$n = 110$$

Systematic Random Classification:

Assuming that the population of n individuals is ordered in a specific order, such as ascending, descending, or any other ranking criterion, and assuming that a particular study requires selecting a sample of n individuals, the ordered population is divided into a number of groups, each group containing $k = \frac{N}{n}$ one individual. Assuming that the population individuals are sequentially ordered, these groups will be:

First group	Second group	Last group
$1, 2, \dots, 3$	$k + 1, k + 2, \dots, 2k$	$(n - 1)k, \dots, nk$

In light of the above division and in order to determine the required sample items for the study, one item is randomly selected from the first group. After knowing the sequence of this item, the rest of the other items are automatically determined by adding the number K to the sequence of the first item to get the sequence of the second item, then to get the sequence of the first item, then to get the sequence of the third item, and so on. Thus, we ensure that each group has contributed one item to the total sample size.

Example 9/In an exam administered to students of a specific grade (a statistical population of $N=24$), whose names were arranged in descending order of their scores, it was necessary to gather the opinions of six of these students to determine the order of these students randomly.

$$N = 24 \quad n = 6$$

$$K = \frac{N}{n} = \frac{24}{6} = 4$$

The number of students in each group is:

$$\{1 \ 2 \ 3 \ 4\}, \{5 \ 6 \ 7 \ 8\}, \{9 \ 10 \ 11 \ 12\},$$

$\{13 \ 14 \ 15 \ 16\}, \{17 \ 18 \ 19 \ 20\}, \{21 \ 22 \ 23 \ 24\}$. This means there is a homogeneous population in each group, which contains four students randomly. The first student from the first group is chosen using simple random selection. Let's assume this student holds the sequences of the other sample groups by adding $K=4$.

$$\{3 \ 7 \ 11 \ 15 \ 19 \ 23\}$$

Multi-stage Sampling:

In this sampling method, the statistical population is divided into units called primary units. A random sample is then selected from these primary units as the first stage. Each selected primary unit is further divided into smaller units called secondary units, and a random sample is then selected from each of these secondary units as the second stage. The selected secondary units are then divided into even smaller units, and a random sample is selected from each of these as the third stage. This division and selection process continues in the same manner until the population is reduced to the individuals from whom the data is collected, thus forming the research sample.

Non-Random Samples:

A **non-random sample** refers to a group of elements selected from the study population in a way where the researcher has a role in choosing certain individuals over others, rather than selecting them randomly. This selection is usually based on considerations related to the nature of the study. Types of non-random samples include:

1. **Quota Sample**
2. **Purposive Sample**

1. Quota Sample:

A quota sample involves dividing the study population into several groups or strata based on certain criteria related to the nature of the study. Then, a number of individuals (a sample) is selected from each group in a personal (non-random) manner by the researcher. The total number of individuals selected from these groups forms the required sample size for the study.

Example:

If we conduct a survey about the public's opinions on television programs, the study population can be divided into **males and females**. Then a sample is selected from males and another from females, and the total number of individuals from these two groups forms the required sample size for the survey.

2. Purposive Sample:

This is a method of selecting a sample from the population intentionally, based on the belief that individuals with certain characteristics are the best representatives of the study population.

Example:

If we are studying ways to improve the sport of **football (soccer)**, it would be preferable to deliberately select a sample of specialists in football, such as coaches or experts, since this group has more experience and knowledge about this sport than others. ⚽

