



**Two days training
program on
statistical methods
in Geography**

Detailed Report
on
Two Days Training Programme on Statistical Methods in Geography

DAY 1

The Department of Geography organized a Two Days Training Programme on Statistical Methods in Geography for the First Year B.A. (Hons.) Geography students on 13th and 14th November 2017 in the Seminar room. As per our scheduled Dr. Netrananda Sahu, Assistant Professor, Department of Geography, Delhi School of Economics, University of Delhi, Delhi-110007 was invited on 13th November 2017 to deliver lectures on the topic “Measures of Association and Correlation”. The topic was covered by the honourable guest faculty in two sessions. The first session was mainly focused on ‘Measures of Association’ and the second session was related to ‘Correlation’. During both the sessions students were given exercises to develop practical knowledge of the subject.

In his first session lecture Dr. Sahu described a measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location. He classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency that are most familiar with, but there are others, such as the median and the mode.

During his lecture Dr. Sahu said mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others. He explained that the mean (or average) is the most popular and well known measure of central tendency. It can be used with both discrete and continuous data, although its use is most often with continuous data. The mean is equal to the sum of all the values in the data set divided by the number of values in the data set. So, He further explained that if we have n values in a data set and they have values x_1, x_2, \dots, x_n , the sample mean, usually denoted by (pronounced \bar{x}), is:

$$\bar{x} = \frac{(x_1 + x_2 + \dots + x_n)}{n}$$

The mean is essentially a model of the data set. It is the value that is most common. According to him, the mean is not often one of the actual values that have been observed in the data

set. However, one of its important properties is that it minimizes error in the prediction of any one value in the data set. Moreover, Dr. Sahu explained that, it is the value that produces the lowest amount of error from all other values in the data set. An important property of the mean is that it includes every value in the data set as part of the calculation. In addition to this, the mean is the only measure of central tendency where the sum of the deviations of each value from the mean is always zero.

Furthermore, Dr. Sahu explained the concept of Median. He explained the median is the value that lies in the center of the distribution. When the data are ordered from least to greatest, the median is located in the middle of the list. He further stated that median can be found for both numbers and ranked categories. It is first necessary to order the values from least to greatest. If there is only one center value (there are an equal number of cases above and below), great, you've found the median. If there are two center values (this will happen when there is an odd number of cases), the median is found by taking the average of the two center values.

Dr. Sahu further demonstrated that mode is the value that occurs most frequently. It is found by determining the number or category that appears most often. If no value occurs more than once, there is no mode. If there are two values that occur most often, report both of them--this type of distribution is bimodal.

In the second session of his lecture he explained the concept of Correlation. He stated that, correlation is any of a broad class of statistical relationships involving dependence, though in common usage it most often refers to how close two variables are to having a linear relationship with each other. According to him, Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. He further explained the concept of Correlation by an example. He told that an electrical utility may produce less power on a mild day based on the correlation between electricity demand and weather. In this example, he said that there is a causal relationship, because extreme weather causes people to use more electricity for heating or cooling. However he also said that, in general, the presence of a correlation is not sufficient to infer the presence of a causal relationship (i.e., correlation does not imply causation).

Dr. Sahu gave various practical exercises on Mean, Median and Mode to the students and showed how Central Tendency is working in various conditions.



Figure 1: Glimpses of Day 1 photographs of Training Programme on Statistical Methods in Geography

DAY 2

On 14th November 2017 Dr. V. Ravi, Assistant Professor, Department of Statistics, Lady Shri Ram College for Women, University of Delhi deliver a talk on the topic “Sampling and Hypothesis Testing”. This topic was also covered in two sessions. In the first session the honourable guest faculty delivered lecture on Sampling while in the second session the focus was on ‘Hypothesis Testing’.

In the first of session of his lecture he talked about the concept and various types of Sampling. According to him, sampling is concerned with the selection of a subset of individuals from within a statistical population to estimate characteristics of the whole population. Two advantages of sampling are that the cost is lower and data collection is faster than measuring the entire population.

He further stated that sampling process comprises several stages such as defining the population of concern, specifying a sampling frame, a set of items or events possible to measure, specifying a sampling method for selecting items or events from the frame, determining the sample size, implementing the sampling plan, sampling and data collecting.

Dr. Ravi had given several examples to make better understanding on the concept of Sampling. For example, if anyone might have a list of information on 100 people out of 10,000 people. The can be use to make some assumptions about the entire population’s behavior. He further explained that unfortunately, it’s not quite that simple. According to him, the sample size must be optimal - not too large or too small. He said, once the sample size is decided a sound technique must be used for actually drawing the sample from the population.

He told that there are two main areas. The Probability Sampling and Non-probability sampling. According to him Probability Sampling uses randomization to select sample members. The probability of each member being chosen for the sample is known, although the odds do not have to be equal while Non-probability sampling uses non-random techniques (i.e. the judgment of the researcher). He explained that calculation can’t be done on odds of any particular item, person or thing being included in your sample.

He further explained the common used Sampling type such as Simple random sampling, Systematic sampling and Stratified sampling. In his opinion a simple random sample (SRS) of a given size, all such subsets of the frame are given an equal probability. Each element of the frame

thus has an equal probability of selection: the frame is not subdivided or partitioned. Furthermore, he stated that, any given pair of elements has the same chance of selection as any other such pair (and similarly for triples, and so on). This could minimize bias and simplifies analysis of results while Systematic sampling (also known as interval sampling) relies on arranging the study population according to some ordering scheme and then selecting elements at regular intervals through that ordered list. While explaining Stratified sampling he told that when the population embraces a number of distinct categories, the frame can be organized by these categories into separate "strata." Each stratum is then sampled as an independent sub-population, out of which individual elements can be randomly selected.

In the second session of his lecture he talked about Hypothesis Testing. According to him A statistical hypothesis, sometimes called confirmatory data analysis, is a hypothesis that is testable on the basis of observing a process that is modeled via a set of random variables. He further stated that statistical hypothesis testing is a method of statistical inference. He explained that commonly, two statistical data sets are compared, or a data set obtained by sampling is compared against a synthetic data set from an idealized model. According to him, a hypothesis is proposed for the statistical relationship between the two data sets, and this is compared as an alternative to an idealized null hypothesis that proposes no relationship between two data sets. He also said that the comparison is deemed statistically significant if the relationship between the data sets would be an unlikely realization of the null hypothesis according to a threshold probability—the significance level. He further explained the uses of Hypothesis Testing. He told that Hypothesis tests are used in determining what outcomes of a study would lead to a rejection of the null hypothesis for a pre-specified level of significance. According to him, the process of distinguishing between the null hypothesis and the alternative hypothesis is aided by identifying two conceptual types of errors (type 1 & type 2), and by specifying parametric limits on e.g. how much type 1 error will be permitted.

Both the faculties during the course of their lectures dwelt upon the subject exhaustively and everyone present in the Seminar room including our faculty members were benefited by their lectures. In overall the lecture was appreciated by all the students.



Day 2



Figure 2: Glimpses of Day 2 photographs of Training Programme on Statistical Methods in Geography