USE OF STATISTICAL TOOLS IN RESEARCH
– WITH SPECIAL REFERENCE OF GEOGRAPHY

Priyanka Bisht
Research Scholar, Department of Geography, Jai Narain Vyas University, Jodhpur

ABSTRACT

The paper is attempted to describe the role of statistics in research especially in geography. To study the different aspect of geography the statistical tools has been used in different time- period. Not only recent scholar but the geographer of ancient time practised the statistical tool in their study. Now a day the practise of statistics in any discipline has increased, especially in field of research whether the research is scientific or social science. This created a buzz among the researcher to use statistical tools in study which promote the blind fold use of statistics in research. The paper depict the some important tools and test which could be apply in study with the attempt of explanation of the important issues keep in mind during applying the tools.

There is introduced some important tools and test which could be apply in the study. Commonly used tool is measure of central tendency which provide the average of data. Along with the some formulas of measure of dispersion, correlation etc. are given. The major problem rise during the research is the testing of hypothesis so there is provide the condition at which these test should be apply along with their necessary criteria. Along with this a short description of the misleading of concepts of statistics are given which should always try to avoid in practice during the using statistical tools. So the paper gives an overall view over the practising of statistics in field of geography with the methods to use it in the research of social-science.

Key words: Research, Geography, Tools, Statistics, Hypothesis , Techniques etc.

INTRODUCTION

Now days, use of statistical tools become the integrated part of research. Different forms of Statistics are employed in Research to prove the hypothesis. The use of Statistical tools start with the collection of data and carry with the presentation of data and analysis of data and end with the interpretation of data by employing different test for hypothesis or other conveying tools of statistics.

“Statistics is a science which deal with the method of collection, classifying, presenting, comparing and interpreting the numerical data to throw light on enquiry” (Seligmen). “Statistics as the Science of estimates and Probabilities”(Prof. Boddington ). “Statistics is the science of collection, presentation, analysis and interpretation of numerical data for logical analysis” (Croxton and Cowden). In this way, the main object to employing the statistics to present the data in systematic manner so the inference could be drawn.

STUDY AREA?

Geography is the subject which is both descriptive and scientific in nature. As descriptive nature of Geography emphasise the theoretical elaboration of cause behind the phenomena. Whereas scientific nature support the use of statistical to prove the hypothesis in Research. To keep this mind, the paper investigates the use of statistics tool in geography from ancient to till date in different branches with the discussion over the use of statistics in different theories elaborated by different scholars.

OBJECTIVE

Use of statistical tools and techniques in the research has increased very fast. Introduction of easy handled software reduce the complication and time consume in the calculation. This promoted the researchers to use the statistical tools in the research to make it more authentic and qualitative. So aim of study is to discuss at what extent the use of statistics tools is appropriate along with the discussion over the necessity of statistics in Research.

USE OF STATISTICS IN GEOGRAPHY

Practise of statistics in geography is a not new it has been used by scholars to introduce the new theory and model in different branches e.g. Agricultural Geography, Settlement Geography, Urban Geography, Population Geography.

AGRICULTURE GEOGRAPHY

In field of agriculture, Statistics tools are used to calculate irrigation intensity, cropping intensity, crop combination etc. Various Scholars such as Weaver, Rafiullah, Nelson and Harris Used mean and Standard Deviation to calculate the Crop combination.

POPULATION GEOGRAPHY

Study of demography and its variable is always be the centre of study for scholars. Different tools are used to calculate the different factors such as Literacy Rate, Sex- Ratio, and Density etc. Projected population is calculated by different method e.g. Graphic Method, Straight line method, Quadratic Parabolic method etc.
SETTELEMENT GEOGRAPHY

This branch deals with spatial distribution of the concentration and dispersion of settlement. Different methods such as Village Density Pattern, Lorenze Curve, Gini’s Concentration Coefficient, Nearest Neighbour Analysis method are Quantitative method to analysis the spatial distribution. In Quantitative method, Robert Hammonand and Petrick Mcullagh introduced Rn scale to analysis the distribution pattern.

ECONOMIC GEOGRAPHY

Various Scholars used the statistical tools to show the spatial distribution of economic activities to measure the concentration and dispersion in an area. To measure the concentration or dispersion of activities different methods are used such as Iornze curve, Gini’s coefficient, Index of Grid Concentration and Concentration index of kant , Index of Simple dispersion, Location quotient etc.

INDUSTRIAL GEOGRAPHY

Industry is the milestone of the economy of any country so it becomes important to know their degree of concentration or dispersion which could help to boost the development of the industry. Weaver used least square method in which standard deviation is used to measure the concentration which he developed to know the crop combination, now it is started to use to know the industrial combination. Similarly Rafiullah used standard deviation, Nelson and Harris used standard Deviation and mean to calculate the concentration of industry and their combination.

In this way, statistics always has been considered by scholar to introduce the model in their respective field.

ROLE OF STATISTICS IN RESEARCH GEOGRAPHY

Geography is always considered as the study of earth-surface. But later Man is to be considered to study with the earth as Earth is considered home of the man. So the Geography includes the study of phenomena occurring over the earth- surface (physical Geography), Man in respect to the earth (Human Geography) , Study of tradition (social Geography), life supporting activities such agriculture, industrial and economical activities etc. Research in geography comprises the multi-dimensional data which is to organise and draw an inference from it.

There is follow a pattern in research to propound a theory or to prove the hypothesis. First formulize a concept which could be in question form or statement form. Then organize data to support the concept which give tentative idea to the solution which help in formation of hypothesis. All research work revolves around the hypothesis to collect data, fact to formulate law which provide the description of an observed phenomena and explanation of phenomena formulate the theory.

Statistics tools are used for systematic arrangement of data and its treatment. To get the information from the collected the statistical tools such as mean, mode, median, standard deviation etc. are used. Relation between two variables is measured by correlation. Similarly there is many method for testing of hypothesis such chi-square test, F test, T test etc. which support in proving of statement. So in this way the statistics provide base to research by testing the hypothesis.

MAJOR STATISTICS TECHNIQUES AND TOOL

MEASURE OF CENTRAL TENDENCY

It is the measure of central value of a data set which provide over all information about the set of data not denote to single value. The value of data set revolve around the central value that is denote by mean, median or mode.

MEAN- It is known as average. It is calculated by sum of value of all variables divided by number of variables.

$$X^- = \frac{\sum x}{N}$$

Where $X^- = \text{Mean}$
$\sum x = \text{Sum of value of All variables}$
$N = \text{Number of variables}$

MEDIAN- It is a positional mean. First data is arranged either ascending or descending order then value of mid-term is median.

$$\text{Median}= \frac{\text{Size of N+1th item}}{2}$$

MODE – It is frequency based. Mode is the most occurring value in a data set.

MEASURE OF DISPERSION

STANDARD DEVIATION – It measure the deviation of each and every item from the mean. It is denoted by sigma $\sigma$ (a greek letter).

$$\Sigma = \frac{\sqrt{\sum(x-x^-)^2}}{N}$$

Where $x^- = \text{Mean}$
$x= \text{variables}$
N = Number of variables

CORRELATION ANALYSIS
“Correlation analysis deals with the association between two or more variables”- Simpson & Kafka
It shows whether two variable are interrelated or not.

KARL PEARSON’S COEFFICIENT OF CORRELATION- Karl pearson introduced the formula to computing the correlation coefficient (r)

\[ r = \frac{\sum(X-X^-)(Y-Y^-)}{N\sigma_x\sigma_y} \]

\[ X^- = \text{Mean of } x \text{ series} \]
\[ Y^- = \text{Mean of } y \text{ series} \]
\[ N= \text{Number of Variables} \]
\[ \sigma_x = \text{Standard Deviation of } x \text{ series} \]
\[ \sigma_y = \text{Standard Deviation of } y \text{ series} \]

SPEARMAN’S CORRELATION COEFFICIENT – The coefficient of correlation is calculated on the basis of ranking.

\[ R = 1 - \frac{6\sum D^2}{N(N^2-1)} \]

Where R denotes rank coefficient of correlation
D denotes the difference of rank between paired items in two series
N denotes number of observation

TESTING OF HYPOTHESIS
Hypothesis is the tentative solution of research problem which give direction to the research. Give an idea about the collection of data such as what data is to collect. Testing of hypothesis follow 5 steps:

1. FORMULATION OF HYPOTHESIS- To proceed the research there is need to formulate two hypothesis.
   a. Null Hypothesis – Null hypothesis says that there is no difference between the mean of two groups. Null Hypothesis is denoted by Ho.
   b. Alternate Hypothesis- Alternate hypothesis is the opposite of the Null Hypothesis which states there is difference between the mean of two groups. It is denoted by Ha.

2. Fix the Suitable significance level- It is level of acceptance or rejection of null hypothesis. If it is said that hypothesis is accepted at the 5% level that means there is risk of rejecting a hypothesis in 5 out of 100 occasion. The risk of rejection could be decrease by taking the confidence level 1% that means there is risk of rejection of hypothesis in 1 out of 100 occasion.

3. Applying the Test criteria- Construction of hypothesis testing procedure is to test criterion. There is need to involves of an appropriated probability distribution for the particular test.

4. Computation and decision making- It includes testing statistics and the standard error of the testing statistics. And finally draw the inference that which hypothesis is true.

TYPES OF ERRORS
There are two types of error:
1. Type I Error (α) - Null Hypothesis is true but test reject it
2. Type II Error(β) - Null Hypothesis is false but test accept it.

PARAMETRIC TEST
Parametric test are applied when data is normally distributed. The prerequisite to apply on data:

1. Means of sample group are normally distributed.
2. Variances of samples and their corresponding population are equal.

If the data is normally distributed then parametric tests are applied. The commonly used test are Student’s t test, Analysis of Variance (ANOVA).

STUDENT’S t TEST- This t test was introduced by W.S. Gosset (1876- 1937) who adopted the pen name ‘student’ and published finding under this name. The t-distribution is used when

\[ t = \frac{X^- - \mu}{S.E} \times \sqrt{n} \]

Where \( S = \sqrt{\frac{\sum(X-X^-)^2}{n-1}} \)

- Statistics for different conditions
  1. t-statistics when the mean of sample is significantly differ from the population mean
Here, µ = Mean of Population
X = Mean of Sample
S.E. = Standard error of mean

2. When population mean is estimated by two independent samples significantly different

\[ t = \frac{X_1 - X_2}{S.E._{x_1-x_2}} \]

\(X_1 - X_2\) = Difference between the mean of samples
S.E. \(x_1-x_2\) = Difference between the standard deviation of samples

3. When population mean is estimated by dependent sample but different significantly

\[ t = \frac{d}{S.E._{d}} \]

d = Difference between mean of the samples
S.E. \(d\) = Difference between the standard deviation for samples.

F- TEST
F-test is named after the great statisticians R.A. Fisher. The main object of test is to find out that the two samples are drawn from normal population having same variance.

\[ F = S_1^2 / S_2^2 \]

Where \(S_1 > S_2\)

NON-PARAMETRIC TEST
Non-Parametric test are applied when the data is not normally distributed or not satisfy the prerequisite condition of parametric test. Some usually adopted non-parametric test are Sign test, Wilcoxon’s signed rank test etc.

MISLEADING CONCEPT OF STATISTICS IN RESEARCH

Statistics provide arrangement of data in systematic manner. The use of statistics is not easily handled as it is necessary to in depth knowledge of the condition of applying the test which is usually found lack. Before applying the statistics the following things keep in mind:

- SELECTION OF SAMPLE- accuracy in drawing inference could be done with the selection of sample by non-random sampling. Random sampling reduces the biasness and gets the equal chance of selection.
- INCORRECT ANALYSIS CHOICE- Statistical tools should be applied according to the nature of data. It is to keep in mind that statistical tools should be use at limited extent.
- INAPPROPRIATE TEST- In order to draw right inference, there is need to apply right test for testing the hypothesis.
- CONCLUDING CAUSATION- It is keep to precision to test the correlation between two factors as one factor affect the other one. The cause behind the correlation should be elaborated precisely. For example:
- DISCARDING UNFAVOURABLE DATA- There is collected a bunch of enormous data out of which some data are irrelevant to study. Before applying any model there is need to sort the data so misconception and misunderstanding could be reduced.
- MISLEADING GRAPHS- Graph is the way to represent the data in pictorial form. There is need to apply right choice of graph according to the available data. For example- If the contribution of different sector in economy is shown so the pie is one of the most suitable graph.
- OVER GENERALISATION- The inference drawn from one population could not be apply to another as the both have different variance. So there is need to aware during the inference drawn applying from particular population is applying to the universe.

CONCLUSION

From above discussion, it is clear that the statistics is very important tool in research but there is need to keep precision while applying it. If any relation could be explain easily with theoretical explanation so avoid the use of statistics. Avoid to use the statistical tools in excess and inappropriately.
REFERENCE


Cabonovic Katarina & Sokolovska Valentine. Use of statistical methods in sociology. Challenges for Analysis of the Economy, the Business and Social Progress. 879-892.

REFERENCE


Cabonovic, Katarina & Sokolovska, Valentine. Use of statistical methods in sociology. Challenges for Analysis of the Economy, the Business and Social Progress. 879-892.


