JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 382-391
https://publishoa.com
ISSN: 1309-3452

The Interpretation of Statistical Data: Effect of Statistical fallacies

Dr. Umesh Kumar Bhayyalal Dubey,
School of Mathematics & Statistics, Faculty of Science,
MIT World Peace University, Paud Road, Kothrud Pune-411038, Maharashtra, India, Mobile: 7385302922,
Email: drumesh.dubey@gmail.com

Abstract

A statistician having collected and analyzed data has to draw inferences and explain their significance. In interpretation, the researcher goes beyond the descriptive data to extract meaning and insights from the data. Various methods are applied in the collection and analysis of statistical data. Statistics are not an end in themselves but they are a means to an end, the end being to draw certain conclusions from them. This has to be done very carefully, otherwise, misleading conclusions may be drawn and the whole purpose of doing research may get vitiated. A researcher or statistician besides the collection and analysis of data has to draw inferences and explain their significance. Through interpretation the meanings and implications of the study become clear. An analysis is not complete without interpretation, and interpretation cannot proceed without analysis. Both are, thus,interdependent. This paper attempts to articulate some basic steps and processes involved in the interpretation of Statistical data.

Keywords: Analysis, Bias, Fallacy, Generalization, Inference, Interpretation

1. Introduction

Analysis and interpretation are central steps in the research process. The purpose of analysis is to summarize the collected data, whereas interpretation is the search for the broader meaning of research findings.

The task of drawing conclusions or inferences and of explaining their significance after a careful analysis of selected data is known as interpretation. It is an inductive process, in which you make generalizations based on the connections and common aspects among the categories and patterns. Scientific interpretation seeks relationship between the data of a study and between the study findings and other scientific knowledge. Interpretation in a simple way means the translation of a statistical result into an intelligible description"

2. Review of Literature

Tamires Queiroz, Carlos Monteiro, Liliane Carvalho and Karen François [1] studied on the affective component in the interpretations of statistical data, which we call affective expressions. They point out the limitations of the empirical study, which need to be addressed in further investigations. A

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 382-391
https://publishoa.com
ISSN: 1309-3452

first limitation is the selection of participants. A second limitation is the selection of research tasks, which were comprised of themes associated with controversial issues. A third limitation is the process of categorization of protocol excerpts and the limitations for a quantitative analysis.

Herbert M. Kritzer [2] conclude from his studies that the lines between quantitative and qualitative social science are less clear than often presumed. Both types of analysis involve extensive interpretation, and tools of interpretation that have many fundamental similarities. He argued that in thinking about interpretation of quantitative data and the results of statistical analysis, one can identify three levels at which the interpretive process operates in assessing the meaning and adequacy of quantitative results. The first level-first order interpretation-draws upon the formal definition of the statistical measures being considered. The second order interpretation is that of recognizing how specific features of the data can influence statistical results in ways that are not closely tied to the substantive theory. Third order interpretation is the most complex and the least understood: connecting the statistical results to broader theoretical patterns.

Rebekah G, Ravindran V. [3] attempt to answer specific questions on human behaviour and response by collecting pertinent data. Large sets of data can be complex, and understanding what the data means requires advanced analytical tools. Statistics is a set of tools than can inform experts dealing with complex information. From their studies they conclude that Statistical analysis is a backbone of the research and unless the data is correctly entered and analysed with appropriate statistics the true essence behind the research findings will go unnoticed. Every parametric analysis has an alternative non parametric analysis. It is essential to check for assumptions and use appropriate statistic and arrive at right conclusions to enhance generalization of results.

Chittaranjan Andrade [4] point out that why 5% may be set as a reasonable cut-off for statistical significance, explains the correct interpretation of $P < 0.05$ and other values of P, examines arguments for and against the concept of statistical significance, and suggests other and better ways for analyzing data and for presenting, interpreting, and discussing the results. He recommended from his studies that all findings should be interpreted in the context of the study design, including the nature of the sample, the sample size, the reliability and validity of the instruments used, and the rigor with which the study was conducted.

3. Method and Discussion

3.1 Need of Interpretation

A researcher/ statistician is expected not only to collect and analyze the data but also to interpret the results of his/ her findings. Interpretation is essential for the simple reason that the usefulness and utility of research findings lie in proper interpretation. It is only through interpretation that the researcher can expose relations and patterns that underlie his findings. In case of hypothesis testing studies the researcher may arrive at generalizations. In case the researcher had no hypothesis to start with, he would try to explain his findings on the basis of some theory. It is only through

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 382-391
https://publishoa.com
ISSN: 1309-3452

interpretation that the researcher can appreciate why his findings are what they are, and can make others understand the real significance of his research findings. Interpretation is not a mechanical process. It calls for a critical examination of the results of one's analysis in the light of all the limitations of data gathering. For drawing conclusions you need a basis. Some of the common and important bases of interpretation are: relationships, ratios, rates and percentages, averages and other measures of comparison.

3.2 Essentials for Interpretation

Certain points should be kept in mind before proceeding to draw conclusions from statistics. It is essential that: a) the data are homogeneous: It is necessary to ascertain that the data are strictly comparable. We must be careful to compare the like with the like and not with the unlike. b) The data are adequate: Sometimes it happens that the data are incomplete or insufficient and it is neither possible to analyze them scientifically nor is it possible to draw any inference from them. Such data must be completed first. c) The data are suitable: Before considering the data for interpretation, the researcher must confirm the required degree of suitability of the data.
Inappropriate data are like no data. Hence, no conclusion is possible with unsuitable data. d) The data are properly classified and tabulated: Every care is to be taken as a pre-requisite, to base all types of interpretations on systematically classified and properly tabulated data and information. e) The data are scientifically analyzed: Before drawing conclusions, it is necessary to analyze the data by applying scientific methods. Wrong analysis can play havoc with even the most carefully collected data. If interpretation is based on uniform, accurate, adequate, suitable and scientifically analyzed data, there is every possibility of attaining a better and representative result. Thus, from the above considerations we may conclude that it is essential to have all the pre-requisites/pre-conditions of interpretation satisfied to arrive at better conclusions

3.3 Precautions in Interpretation

It is important to recognize that errors can be made in interpretation if proper precautions are not taken. The interpretation of data is a very difficult task and requires a high degree of skill, care, judgement and objectivity. In the absence of these, there is every likelihood of data being misused to prove things that are not true. The following precautions are required before interpreting the data. 1) The interpreter must be objective. 2) The interpreter must understand the problem in its proper perspective. 3) He / she must appreciate the relevance of various elements of the problem. 4) See that all relevant, adequate and accurate data are collected. 5) See that the data are properly classified and analyzed. 6) Find out whether the data are subject to limitations? If so what are they? 7) Guard against the sources of errors. 8) Do not make interpretations that go beyond the information / data. 9) Factual interpretation and personal interpretation should not be confused.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 382-391
https://publishoa.com
ISSN: 1309-3452

They should be kept apart. If these precautions are taken at the time of interpretation, reasonably good conclusions can be arrived at.

3.4 Concluding Remarks on Interpretation

The task of interpretation is not an easy job. It requires skill and dexterity on the part of the researcher. Interpretation is an art that one learns through practice and experience. The researcher may seek the guidance of experts for accomplishing the task of interpretation. The element of comparison is fundamental to all research interpretations. Comparison of one's findings with a criterion, or with results of other comparable investigations or with normal (ideal) conditions, or with existing theories or with the opinions of a panel of judges / experts forms an important aspect of interpretation.

The researcher must accomplish the task of interpretation only after considering all relevant factors affecting the problem to avoid false generalizations. He/she should not conclude without evidence. He/she should not draw hasty conclusions. He/she should take all possible precautions for proper interpretation of the data.

3.4 Conclusions and Generalizations

Results are direct observations summarized and integrated by the statistical analysis such as comparison of two groups of workers. Group 'A's' average wage is Rs.6,000 and that of group 'B' is Rs.7,000. A conclusion is an inference based on the data that group B workers are better paid than those of group 'A'. In everyday life, we often make generalizations. We believe that what is true of the observed instances will be true of the unobserved instances. Since, we have had a uniform experience, we expect that we shall have it even in the future. We are quite conscious of the fact that the observed instances do not constitute all the members of a class concerned. But we have a tendency to generalize. A generalization is a statement, the scope of which is wider than the available evidence. For example, A is a crow, it is black. B is a crow, it is black. C is a crow, it is also black. Therefore, it can be generalized that "all crows are black". Similarly, all swans are white. All rose plants possess thorns etc., the process by which such generalizations are made is known as induction by simple enumeration.

4. Methodology

4.1 Methods of Generalization

1) Logical Method

This method was first introduced by John Stuart Mill [5], who said that generalization should be based on logical processes. Mill thought that discovering causal connections is the fundamental task in generalization. If casual connections hold good, generalization can be done with confidence. Five

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 382-391
https://publishoa.com
ISSN: 1309-3452

methods of experimental enquiry have been given by Mill. These methods serve the purpose of discovering causal connections. These methods are as follows.

i) **The Method of Agreement**

This may be positive or negative. The method of agreement states that if two or more instances of a phenomenon under investigation have only one circumstance in common, the circumstance is the cause or the effect of the given phenomenon. For example, a person gets pain in his eyes whenever he roams in the sun. Negatively, when he is under the shade he does not have pain. Therefore, the cause for pain is roaming in the sun.

ii) **The Method of Difference**

This method is a combination of both positive and negative methods of agreement. In this method only two instances are required. The two instances resemble each other in every other respect, but differ in the absence or presence of the phenomenon observed. The circumstance in which alone the two instances differ, is the effect, or the cause. Let us take the example given by Mill. A man is shot, he is wounded and dies. Here the wound is the only differentiating circumstance between the man who is alive and the man who is dead. Hence, death is caused by the wound.

iii) **Joint Method of Agreement and Difference**

This is a combination of the method of agreement and the method of difference. According to this method, we require two sets of instances. This method can be stated like this: If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances in which it does not occur have nothing in common, save the absence of that circumstance, the circumstance in which alone the two sets of instances differ, is the effect or the cause.

For example:

A + B + C	Produce X
A + P + Q	Produce X
M + N + Non-A	Produce Non-X
G + H + Non-A	Produce Non-X

∴ A and X are causally connected

iv) **The Method of Residues**

This method is based on the principle of elimination. The statement of this method is that, subtract from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents. For example: A loaded lorry weighs 22 tons. The dead weight of the lorry is 2 ton. The weight of load = 22 – 2 = 20 tons.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 382-391
https://publishoa.com
ISSN: 1309-3452

v) **The Method of Concomitant Variation**

This method can be stated as "whatever phenomenon varies in any manner, whenever another phenomenon varies in some particular manner, is either the cause or the effect of that phenomenon or is connected with it. This method is quantitative in nature and needs statistical techniques for measurement. That is why it is also known as the method of quantitative induction, because we base our inference on quantitative change in the two factors and is applied as some form of correlation analysis

4.2 Statistical Method

Statistical method may be defined as "the collection, presentation, analysis and interpretation of numerical data". Thus statistical method involves four steps.

i) Collection of Data

The facts pertaining to the problem under study are to be collected either by survey method or by observation method or by experiment or from a library.

ii) Presentation of Data

The data collected as to be processed by classification, tabulation and then be presented in a clear manner.

iii) Analysis of Data

The processed data then should be properly analyzed with the help of statistical tools, such as measures of central tendency, measures of variation, measures of skewness, correlation, time series, index numbers etc.

iv) Interpretation of Data

The collected and analyzed data has to be interpreted. It involves explanation of facts and figures and drawing inferences and conclusions.

4.3 Statistical Fallacies

Interpretation of data, is a very difficult task and requires a high degree of care, objectivity, skill and judgement. In the absence of these things, it is likely that the data may be misused. In fact, experience shows that the largest number of mistakes are committed knowingly or unknowingly while interpreting statistical data which may lead to misinterpretation of data by most of the readers. Statistical fallacies may arise at any stage – in the collection, presentation, analysis and interpretation of data. The following are some of the (i) specific examples illustrating how statistics can be misinterpreted, (ii) Sources of errors leading to false generalizations, (iii) examples how fallacies arise in using statistical data and statistical methods.

i. Bias

Bias, whether it is conscious or unconscious, is very common in statistical work and it leads to false generalizations. It is found that wrong interpretations are made want only

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 382-391
https://publishoa.com
ISSN: 1309-3452

to prove their point. Sometimes deliberately statistical information is twisted as to grind one's own axe. For example, a business man may use statistics to prove the superiority of their firm over others by saying that our firm earned a profit of Rs.1,00,000 whereas firm 'X' earned only Rs.80,000 this year. On the face of it, it appears that firm 'X' has not performed well. But a little thinking reveals that many other variables have to be considered before drawing such a conclusion, such as what is the capital employed? If the capital employed is same, then the quality of product and so on. Unconscious bias is even more insidious. Perhaps, all statistical reports contain some unconscious bias, since the statistical results are interpreted by human beings after all. Each may look at things in terms of his own experience and his attitude towards the problem under study. People suffer from several inhibitions, prejudices, ideologies and hardened attitudes. They cannot help reflecting these in their interpretation of results. For example: A pessimist will see the future as being dark, where as an optimist may see it as being bright.

ii. Inconsistency in Definitions

Sometimes false conclusions are drawn because of failure to define properly the object being studied and hold that definition in mid for making comparisons. When the working capital of two firms is compared, net working capital of one must be compared with only net working capital of the other and not with gross working capital. Even within the organization, for facilitating comparison over a period of time it is necessary to keep the definition constant.

iii. Inappropriate Comparisons

Comparisons between two things cannot be made unless they are really alike. Unfortunately, this point is generally forgotten and comparisons are made between two dissimilar things, thereby, leading to fallacious conclusions. For example, the cost of living index of Pune is 160 (with base year 2019) and that of Bangalore is 165 (with base 2015). Therefore, Bangalore is a costlier city than Pune city. This conclusion is misleading as the base years of the Indices are different.

iv. Faulty Generalizations

Many a time people jump to conclusions or generalizations on the basis of either too small a sample or a sample that is not representative of the population. For example, if a foreigner came to Nagpur and his purse was stolen by a pick pocket and he comments that there is no safety and security for foreigners in India. This is not true as thousands of foreigners come to India. They are safe and secure. Sometimes the sample size may be adequate but not representative.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 382-391
https://publishoa.com
ISSN: 1309-3452

v. Drawing Wrong Inferences

Sometimes wrong inferences may be drawn from the data. For example, the population of a town has doubled in 10 years. From this it is interpreted that the birth rate in the town has doubled. Obviously, this is a wrong inference, as the population of the town can double in many ways (example: exodus from villages, migration from other places etc.) than doubling of birth rate only.

vi. Misuse of Statistical Tools

The various tools of analysis such as measures of central tendency, measures of variation, measures of correlation, ratios, percentages etc., are very often misused to present information in such a manner as to convince the public or to camouflage things. In a company there are 1,00,000 shares and 1,000 share holders. The company claims that their shares are well distributed as the average share holding is 100. But a close scrutiny reveals that 10 persons hold 90,000 shares where as 990 persons hold 10,000 shares, average being about 10. Similarly, range can be misused to exaggerate disparities. For example, in a factory the wages may range between Rs.10,000 to Rs.15,000 a month and the Manager gets Rs.40,000 a month. It is reported that the earnings of their employees range from Rs.10,000 to Rs.40,000.

vii. Failure to Comprehend the Data

Very often figures are interpreted without comprehending the total background of the data and it may lead to wrong conclusions. For example, see the following interpretations: – The death rate in the army is 10 per thousand, where as in the city of Pune it is 15 per thousand. Therefore, it is safer to be in the army than in the city. – Most of the patients who were admitted in the intensive care (IC) ward of a hospital died. Therefore, it is unsafe to be admitted to intensive care ward in that hospital.

6. Conclusion

Statistical methods and techniques are only tools. As such, they may be very often misused. Some people believe that “figures can prove anything.” Figures don’t lie but liers can figure”. Some people regard statistics as the worst type of lies. That is why it is said that “an ounce of truth can be produced from tons of statistics”. Mere quantitative results, or huge body of data, without any definite purpose, can never help to explain anything. The misuse of statistics may arise due to:
i)Analysis without any definite purpose. ii) Carelessness or bias in the collection and interpretation of data. iii) Deliberate cooking up of data. iv) Wrong definitions, inappropriate comparisons, inadequate data etc.

As a principle, statistics cannot prove anything, but they can be made to prove anything because statistics are like clay with which one can make God or the Devil. The fault lies not with statistics but with the person who is using statistics. The interpreter must carefully look into these points

before he sets about the task of interpretation. We may conclude with the words of Marshall who said "Statistical arguments are often misleading at first, but free discussion clears away statistical fallacies".

7. Acknowledgments

This research was supported by Pro Vice Chancellor of MIT WPU Dr. Milind Pande and Associate Dean Faculty of Science Dr. Shubhalaxmi Joshi and Head of School Dr. Prashant Malvadkar, MIT School of Mathematics & Statistics, Pune. Special thanks go to my colleagues Prof. Suvarna Ranade for her valuable contribution in providing helpful comments and advice on this manuscript.

References

[1] Tamires Queiroz, Carlos Monteiro, Liliane Carvalho and Karen François, 2017, "Interpretation of Statistical Data: The Importance of Affective Expressions," Statistics Education Research Journal, 16(1), 163-180, http://iase- web.org/Publications.php?p=SERJ© International Association for Statistical Education (IASE/ISI).
[2] Herbert M. Kritzer, 1996, "The Data Puzzle: The Nature of Interpretation in Quantitative Research," American Journal of Political Science Vol. 40, No. 1 (Feb., 1996), pp. 1-32 (32 pages) Published by: Midwest Political Science Association
[3] Rebekah G, Ravindran V., 2018, "Statistical analysis in nursing research," Indian J Cont Nsg Edn [serial online] 2018 [cited 2022 Jan 23]; 19:62-70. Available from : https://www.ijcne.org/txt.asp?2018/19/1/62/286497
[4] Chittaranjan Andrade, 2019, " The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives, Indian J Psychol Med.;41(3): 210-215. Doi:10.4103/IJPSYM.IJPSYM_193_19, PMCID: PMC6532382
[5] John Stuart, 1806-1873, "A System of Logic: Raciocinative and Inductive,".
[6] Achen, Christopher H., 1982, "Interpreting and Using Regression," Beverly Hills, CA: Sage Publications.
[7] Hanushek, Eric A., and John E. Jackson, 1977, "Statistical Methods for Social Scientists," New York: Academic Press.
[8] Huff, Darrell, 1954, "How To Lie with Statistics," New York: W.W. Norton.
[9] Zieffler, A., Garfield, J., Delmas, R., Le, L., Isaak, R., Bjornsdottir, A., & Park, J, 2011, Publishing in SERJ: an analysis of papers from 2002–2009. Statistics Education Research Journal, 10(2), 5–26. [Online: iase-web.org/Publications.php?p=SERJ_issues]
[10] Rumsey D, 2003, "Statistics for Dummies," 2003,New Jersey: Wiley Publishing Inc,.
[11] Swinscow TDV, 2022, "Statistics at square one," http://www.bmj.com//statsbk/ (accessed 2 January 2022)
[12] B.N.Gupta. Statistics. Sahitya Bhavan, Agra. S.P. Gupta. Statistical Methods, Sultan Chand & Sons,

New Delhi.
[13] B.N.Agarwal. Basic Statistics, Wiley Eastern Ltd. P.Saravanavel. Research Methodology, Kitab Mahal, Allahabad.
[14]C.R. Khothari. Research Methodology (Methods and Techniques), New Age International Pvt. Ltd, New Delhi.
[15] Umeshkumar Dubey, D.P. Kothari and G.K. Awari, Quantitative Techniques in Business Management and Finance: A Case-study Approach, Chapman and Hall/CRC, New York, 2016.