Journal DOI: 10.36713/epra1013|SJIF Impact Factor (2021): 7.473 ISSN: 2347-4378
EPRA International Journal of Economics, Business and Management Studies (EBMS)
Volume: 9 | Issue: 11| November 2022 -Peer-reviewed Journal

A CASE STUDY ON APPROACHES TO DATA ANALYSIS

Saumya Singh[1], Aman Pushp[2], Bhakti Agarwal[3]
[1,2,3]PhD Scholar, Symbiosis Institute of Business Management, Pune;
Symbiosis International (Deemed) University, Pune, India.

ABSTRACT
This case study deals with the approaches to Data Analysis. Classification of data analysis is elaborate. It goes beyond the incorporation of classification based on sources and nature of the data to include various approaches to data analysis. Difference in approaches to these data analysis and their area of focus finds mention in this case let. This case study deals with data analysis classification and not just data classification broadly on several parameters and its usage in different analysis.
KEY WORDS : ***Data analysis, data classification***

1. INTRODUCTION

Data analysis is the conversion of data collected from relevant sources into meaningful information. Data can be analyzed or processed for its usage by stakeholders using different techniques or methods (Taherdoost, 2022). What comes prior to data analysis is data collection and what comes post data analysis is data presentation. What remains relevant and a pre-requisite for all the three afore mentioned stages is "Measurement of scales." The measurement scale of the data classifies data into four groups: 1) nominal; 2) ordinal; 3) interval; and 4) ratio.

However, the data classification is exhaustive and goes beyond measurement scale to be classified on several others parameters. Data can also be classified on the several ways. First classification is primary versus secondary. Primary data are collected for a certain purpose. Usually, questionnaire is used to collect the primary data. Another classification is of the nature of the data. The data can be discreate or continuous. The former is called categorial and the latter is called numerical data. When it comes to numerical data it can take up any arbitrary value whereas the same does not hold true for the categorical data. Categorical data can assume only a restricted set of probable values. While the usage of categorical data finds more frequency in solving real life problems related to medical sciences and weather forecasts, numerical representation of the categorical data remains a challenge and need of the hour (Zhang et al 2015).

In addition to the data classification, the data analysis can also be broadly classified into several categories. The contemporary world to data analysis broadly divides the data analysis into two approaches: machine learning and statistical learning. Machine learning does not bother about statistical significance, rather it uses the huge data to have more effective predictive ability. Means, if x is causing y or x1 is causing y, ML does not bother about it as long they get more effective predictive outcomes. On the contrary, statistical learning differentiate between significant and non-significant association (which includes knowing whether x is causing y or x1 or both).

Within the broad classification of statistical learning, a broad category of data analysis predominantly uses primary data and another secondary data. The former data set usually used for cluster analysis, factor analysis including SEM etc. Factor analysis being one such analysis which makes use of primary data, comes into picture when the number of variables is huge and derivation of factors of smaller sets are undertaken for simplification of the analysis (Taherdoost, 2017). The latter is usually used for econometric analysis including financial econometrics. Cluster analysis on the other hand groups the data based on criterion or characteristics making homogenous clusters all of which have maintained heterogeneity between two or more clusters (Taherdoost, 2022). Discriminant analysis as a method has differing aspects and implications in different fields. Studies incorporating multiple populations dealing with multi-variate research problems opt for discriminant analysis (Huberty, 1975).

2. LITERATURE REVIEW

In finance, qualitative studies are usually on the literature-based studies including systematic literature review and bibliometric studies (Pinto et al., 2019, Rastogi et al., 2022, Rastogi et al., 2021a, Patil and Rastogi, 2019). Primary survey-based studies are also witnessed under behavioral finance studies (Rastogi and Ragabiruntha, 2018, Goel and Rastogi, 2021b, Goel and Rastogi, 2021a, Rastogi et al., 2020). However, in finance and economics, most of the studies are using secondary data which includes time series (Rastogi et al., 2021b, Rastogi et al., 2021a, Sharma and Rastogi, 2020) and panel data (Sharma and Rastogi, 2022, Kanoujiya et al., 2022). Main problem encountered in usage of secondary data is the exploration of the sources and verification of the relevance of data sought. Primary data finds a lot of usage in social science studies and has own set of disadvantages (Hox et al 2005).

3. QUESTIONS

1. Describe all the measurement of scale used to collect the data.
2. Explain the categorical versus numerical data
3. Discuss the difference between factor analysis, cluster analysis, discriminant analysis, logistic regression (qualitative response models), structured equation modelling (SEM) and financial econometrics
4. Elaborate the difference between machine learning, statistical learning and data analytics

5. REFERENCES

1. *A. Goel & S. Rastogi, - Credit scoring of small and medium enterprises: a behavioural approach, || Journal of Entrepreneurship in Emerging Economies, (2021a).*
2. *A. Goel & S. Rastogi, - Understanding the impact of borrowers' behavioural and psychological traits on credit default: review and conceptual model, || Review of Behavioral Finance, (2021b).*
3. *J. Kanoujiya, S. Rastogi, & V.M. Bhimavarapu, - Competition and distress in banks in India: An application of panel data, || Cogent Economics & Finance, Vol. 10, pp. 2122177, (2022).*
4. *A. C. Patil & S. Rastogi, - Time-Varying Price–Volume Relationship and Adaptive Market Efficiency: A Survey of the Empirical Literature, || Journal of Risk and Financial Management, Vol. 12, pp. 1-18, (2019).*
5. *G. Pinto, S. Rastogi, S. Kadam & A. Sharma, - Bibliometric study on dividend policy, || Qualitative Research in Financial Markets, Vol. 12, pp. 72-95, (2019).*
6. *S. Rastogi, A. Doifode, J. Kanoujiya & S. P. Singh, - Volatility integration of gold and crude oil prices with the interest rates in India, || South Asian Journal of Business Studies, (2021a).*
7. *S. Rastogi, A. Goel & A. Doifode, - Open API in Indian banking and economic development of the poor: opportunities and challenges, || International Journal of Electronic Banking,Vol. 2, pp. 321-348, (2020).*
8. *S. Rastogi, & E. Ragabiruntha, - Financial inclusion and socioeconomic development: gaps and solution, || International Journal of Social Economics,Vol. 45, pp. 1122-1140, (2018).*
9. *S. Rastogi, A. Sharma, G. Pinto, & V. M. Bhimavarapu, - A literature review of risk, regulation, and profitability of banks using a scientometric study, || Future Business Journal,Vol. 8, pp. 1-17, (2022).*
10. *S. Rastogi, V. Tripathi, & S. Kuknor, - Informational role of futures volume for spot volatility, || Pacific Accounting Review, (2021b).*
11. *A. Sharma, & S. Rastogi, - SPOT VOLATILITY PREDICTION BY FUTURES AND OPTIONS: AN INDIAN SCENARIO, || International Journal of Modern Agriculture,Vol. 9, pp. 263-268, (2020).*
12. *A. Sharma, & S. Rastogi, - Voluntary Disclosures of Indian Microfinance Institutions, || Australasian Accounting, Business and Finance Journal,Vol. 16, pp. 108-130, (2022).*
13. *H. Taherdoost, Different types of data analysis; data analysis methods and techniques in research projects, || International Journal of Academic Research in Management, Vol. 9, 1-9, (2022).*
14. *H. Taherdoost, Understanding of e-service security dimensions and its effect on quality and intention to use, || Information & Computer Security, (2017).*
15. *K. Zhang, Q. Wang, Z.Chen, I. Marsic, V.Kumar, G.Jiang, & J. Zhang, From categorical to numerical: Multiple transitive distance learning and embedding. In Proceedings of the 2015 SIAM International Conference on Data Mining (pp. 46-54). Society for Industrial and Applied Mathematics, (June, 2015).*
16. *J.Hox, , & H. R. Boeije, Data collection, primary versus secondary, (2005).*
17. *C. J Huberty, Discriminant analysis, || Review of Educational Research, 45, 543-598, (1975).*