IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A case study on inhibitors of data analytics using 'kruskal-wallis test'

Alakh Bhatnagar, Ph.D. Research Scholar, Dayalbagh Educational Institute, Deemed University, Agra, Uttar Pradesh, India

Prof. Shailini Nigam, Dayalbagh Educational Institute, Deemed University, Agra, Uttar Pradesh, India

Abstract

This case study is of an apparel manufacturing cum export unit in Uttar Pradesh, India. The unit has been running several initiatives to move from a judgement-driven to a data-driven organization, with an aim to transform digitally and become more competitive. The management and employees were asked to rank five inhibitors of leveraging data analytics based on their assessment of the impact of those inhibitors. Kruskal-Wallis statistical test revealed that not all inhibitors have the same impact.

Keywords: Kruskal-Wallis, Non-parametric, Inhibitors, Analytics, Manufacturing, Exports

I. INTRODUCTION

Businesses must be adaptable to satisfy the ever-changing demands of the business environment, which is becoming increasingly digital and data-driven. The company in this case study is a clothing manufacturer cum exporter based in Uttar Pradesh, India. As the number of companies operating in this field increases, the business landscape is also getting competitive. The company acquires raw materials, processes the raw materials, manufactures a variety of garments, and exports most of its finished goods to international markets (with a majority to European countries). The promoter founded the company as a small-scale operation, and today it has over 250 people (full-time and contracted). Their success and ability to remain in business depend on their ability to build and/or modify operational processes in response to changing global-business situations. The company has identified several initiatives to leverage data analytics in their decision-making process. However, several challenges, like data quality, data integration and insufficient data on critical events, keep the company and management from effectively leveraging their data for decision-making. As a result, it is becoming increasingly challenging for the organization to understand if all inhibitors have the same impact.

II. LITERATURE REVIEW

The following literature was studied to see the applications and discussions on Kruskal-Wallis Test:

Year	Author(s)	Applications/Usage of the Kruskal-Wallis Test
2021	Naseri and Younesi	In a study of root rots in bean crops
2021	Nabi and El-adaway	For risks affecting modular construction projects
2021	Kanetaki et al.	For student performance in online education
2021	Lubrano et al.	For the respiratory function in kids wearing face masks
2021	Hertina, Hidayat and Saudi	For share portfolio performance
2016	Santos et al.	For analysis of an oil and gas multinational company
2016	Biswas and Bhattacharyya	To identify the appropriate date of transplanting for rice yield
2014	Ostertagova, Ostertag, and Kováč	Discussed methodology and applications of the Kruskal-Wallis Test
2013	Kasmer, Liu, Roberts & Valadao	To study runners under competition conditions

III. RESEARCH METHODOLOGY

1. Selection of variables for study: Seven inhibitors of leveraging analytics were identified from a broader literature review on barriers of data analytics in business. Inputs were taken from experts and management of the organization, and finally, two inhibitors were dropped from the study
    - The selected variables are T1: Data Integration; T2: Unclear KPIs (Key Performance Indicators); T3: Insufficient Data; T4: Infosec (Information Security), and T5: Data Quality
2. Survey tool: The final list of five inhibitors was converted to a web-based survey. The questionnaire was based on a five-point Likert scale, requesting the respondents to score them basis the 'severity in inhibiting effective data analytics' in their organization. The five-point scale was 'Very Low', 'Low', 'Medium', 'High' and 'Very High.
3. Statistical tool: The responses were consolidated into an excel file, and then, the R Statistical computing was leveraged to perform data analysis.
4. Statistical test: The statistical package 'ggstatsplot' (Patil, 2021) was leveraged to perform the Kruskal-Wallis test.

IV. PERFORMING THE KRUSKAL-WALLIS HYPOTHESIS TEST

The Kruskal–Wallis test (1952) is a nonparametric one-way ANOVA technique. The Kruskal-Wallis test is frequently employed because to its adaptability and application in many sectors, including engineering, manufacturing, healthcare, biology, psychology, education (Ostertagova, Ostertag, and Kováč, 2014) and sports. For ordinal or non-normally distributed data (Likert scale in this case), the Kruskal-Wallis test can be very helpful. The Kruskal-Wallis test can assist in answering two questions: (1) Is there at least one inhibitor with an impact distinct from the other inhibitors? (2) Between which inhibitors is this distinction most pronounced? The steps deployed in the Kruskal-Wallis test are as follows: First, the survey data were prepared with the grouping variable (inhibitors) on the X-axis and the respondents' scores on the Y-axis (1 to 5 for 'Very Low' to 'Very High'). Since the data are not normally distributed or ordinal, the function 'ggbetweenstats' executed the nonparametric statistical method Kruskal-Wallis. The function produces visual results, as shown in Figure 1.0, where the results are printed on the top left corner.

V. CONCLUSION AND RECOMMENDATIONS

Let us review the results from the Kruskal-Wallis test:

1. Chi-square and degrees of freedom were historically used to calculate P-value, but here R Statistical Programming has already calculated the P-value discussed below, and hence we have not used them
2. Since P-value is very small (<0.001), it provides a "solid evidence" to reject the null hypothesis (Ho), which states that "scores/impact of all inhibitors are similar", and the alternative hypothesis (Ha), can be accepted, which states that "scores/impact of at least one inhibitor differs."

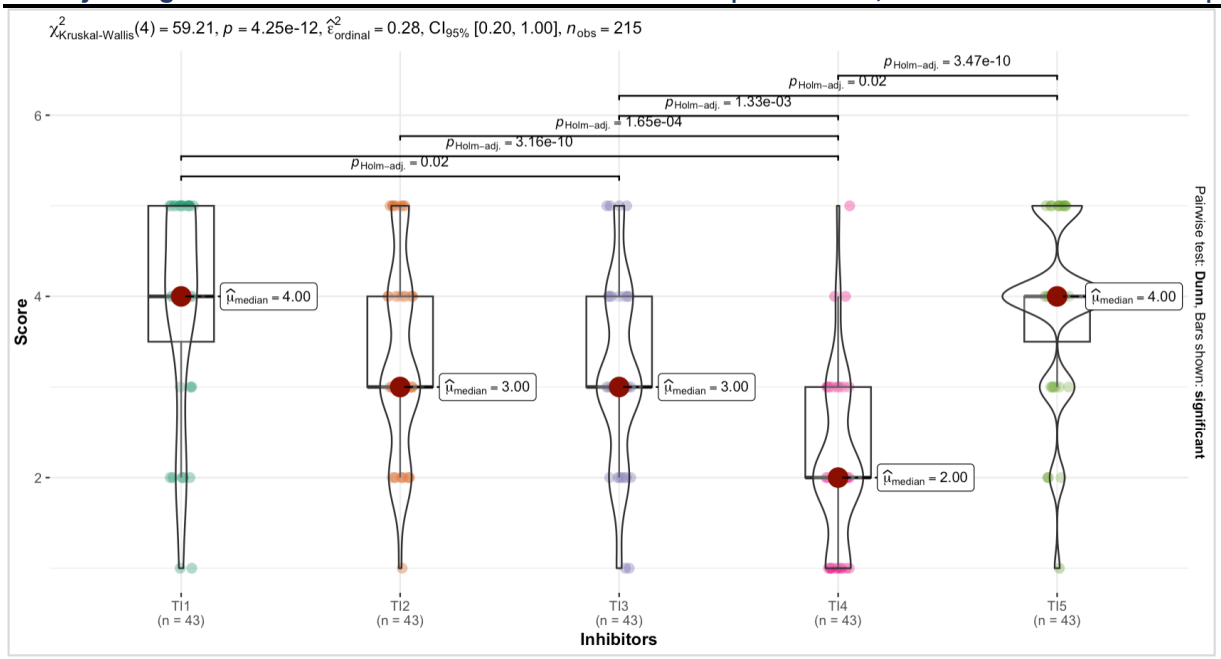

Figure 1.0: Result of Kruskal-Wallis test on five inhibitors of using data analytics

However, there are two problems with the P-value, which we should discuss: (1) a small P-value only indicates that a difference between inhibitors exists for sure, and the difference is not by chance, but the P-value does not tell how significant the difference is. The function 'ggbetweenstats' provides 'partial epsilon squared' ($\varepsilon^{\wedge 2}$) with 95% confidence intervals as the measure of the 'effect size' for the Kruskal-Wallis test. For inhibitors, the effect size of 0.28 indicates that the effect of inhibitors on the respondent score was 'large' (Field, 2013), where 'ES' means 'effect size'.

- ES < 0.02 : Very small
- 0.02 <= ES < 0.13 : Small
- 0.13 <= ES < 0.26 : Medium
- ES >= 0.26 : Large

The second issue with the P-value is that it indicates that a difference exists across groups (inhibitors) but does not specify between which groups. Therefore, it is necessary to compare each pair of inhibitors. The 'ggbetweenstats' function also executes pair-wise tests for significant Kruskal-Wallis, displays P-values, and corrects these values for multiple comparisons. The two-group instance of the Kruskal-Wallis test is equivalent to the Wilcoxon rank sum test since they both compare rank sums of groups but not medians, although medians are frequently displayed. Even if the groups' medians are similar, there may be significant statistical differences between them.

VI. CONCLUSION

We used the Kruskal-Wallis test to confirm that not all inhibitors have an equal impact when preventing the efficient use of data analytics within the company. Moreover, the statistical findings suggest that the disparity between inhibitors is statistically "substantial". The relative relevance of each inhibitor can be seen by comparing it to all the others through pair-wise comparisons. In examining the p-values of pair-wise tests, we find that the biggest difference is between T1 (Data Integration) and T4 (Infosec) and also between T4 (Infosec) and T5 (Data Quality). For the given organization, the proposed study and results can augment this organization's efforts to shift from adhoc, gut-level decisions to a more systematic and data-driven. The manufacturing unit's management can use the analysis findings to focus more on inhibitors T1 and T5 and devote more resources and time to managing them. This process can be repeated by including new inhibitors and removing the ones with low priority.

References

1. Abdul Nabi, M., & El-adaway, I. H. (2021). Understanding the key risks affecting cost and schedule performance of modular construction projects. *Journal of Management in Engineering, 37(4), 04021023.*
2. Biswas, R., & Bhattacharyya, B. (2016). Kruskal-Wallis Statistic to Identify the Appropriate Date of Transplanting for Rice Yield in Nadia District of West Bengal. *Environment & Ecology, 34(2A), 643-646.*
3. Field, A (2013). Discovering statistics using IBM SPSS Statistics. *Fourth Edition. Sage: London*.
4. Hertina, D., Hidayat, M. B. H., & Saudi, M. H. (2021). Share Portfolio Performance Analysis Using Sharpe, Trey nor and Jensen Methods with the Geographical Perspective of Indonesia Stock Exchange. *Review of International Geographical Education Online, 11(3), 55-61*.
5. Kanetaki, Z., Stergiou, C., Bekas, G., Troussas, C., & Sgouropoulou, C. (2021). The impact of different learning approaches based on MS Teams and Moodle on students' performance in an online mechanical CAD module. *Glob. J. Eng. Educ, 23, 185-190.*
6. Kasmer, M. E., Liu, X. C., Roberts, K. G., & Valadao, J. M. (2013). Foot-strike pattern and performance in a marathon. *International journal of sports physiology and performance, 8(3), 286-292.*
7. Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association, 47(260), 583-621.*
8. Lubrano, R., Bloise, S., Testa, A., Marcellino, A., Dilillo, A., Mallardo, S., ... & Ventriglia, F. (2021). Assessment of respiratory function in infants and young children wearing face masks during the COVID-19 pandemic. *JAMA network open, 4(3), e210414-e210414.*
9. Naseri B. and Younesi H. (2021). Beneficial microbes in biocontrol of root rots in bean crops: A meta-analysis (1990–2020). *Physiological and Molecular Plant Pathology, Volume 116, December 2021, 101712.*
10. Ostertagova, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis test. In *Applied Mechanics and Materials (Vol. 611, pp. 115-120). Trans Tech Publications Ltd.*
11. Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. *Journal of Open Source Software, 6(61), 3167, doi:10.21105/joss.03167.*
12. Santos, M., Vianna, D. V., Dias, F. D. C., & Reis, M. F. (2016). Demand Seasonality Analysis of High Valued-Added Items: a Kruskal-Wallis Test Application in an Oil and Gas Multinational Company. *International Journal of Basic & Applied Sciences IJBAS-IJENS Vol:16 No:05.*
13. Xia, Y. (2020). Correlation and association analyses in microbiome study integrating multiomics in health and disease. *Progress in Molecular Biology and Translational Science, 171, 309-491.*