SciForce

International Journal of Computer Science and Data Engineering

Journal homepage: www.sciforce.org
ISSN : 3066-6813

Open Access Research Article

Assessing Normality in Healthcare Expenditure Data: A Shapiro-Wilk Test Approach In Pythons

Rakesh Mittapally*

**Business Intelligence Architect / AI and ML Engineer, USA*

Abstract

Healthcare financing remains a critical challenge in India, where 58.7% of all medical costs are paid for out of pocket, pushing families towards catastrophic expenses and poverty. This study examines the factors that contribute to catastrophic health costs in Indian households, focusing on aspects such as family structure, socioeconomic status, and insurance coverage, which are less explored in existing studies. Primary data were collected from 1,018 respondents using a structured questionnaire during July-September 2025. Due to the non-normal distribution of the data, non-parametric statistical methods such as Mann-Whitney U tests, Kruskal-Wallis tests, and Spearman's correlation were employed [21].Despite moderate insurance coverage adequacy (mean=2.73), a significant financial burden persists. Out-of-pocket spending on medical expenses (mean=2.71) and families primarily relying on personal savings (mean=3.28) are observed. Lifestyle changes are common (mean=3.45), however, accessing treatment for chronic diseases remains challenging (mean=3.18).Significant gaps exist in India's healthcare financing system. Comprehensive policy interventions, including increased public funding, strengthened insurance schemes, and reduction in drug prices, are essential to achieve universal health coverage and prevent catastrophic expenditures.

Keywords: Catastrophic health expenditure, Direct cash transfers, Health insurance coverage, Health care financing, Non-parametric analysis, Socio-economic determinants, Universal health coverage.

Introduction

Healthcare financing management is a critical challenge in contemporary healthcare systems, particularly in developing countries where the burden of medical expenses threatens the financial stability of families. In most high-income developed countries, insurance systems are in place to ensure that healthcare is provided fairly and equitably. affordably [1]. This heavy reliance on out-of-pocket payments creates a precarious situation where families have to choose between seeking necessary medical care and maintaining their economic stability. Catastrophic health expenditure this occurs when the direct out-of-pocket spending on healthcare for a family exceeds a certain percentage their budget, forcing them to reduce expenses for other essential needs expenses. This threshold is typically set between 10% and 40% of a family's total expenditure [2]. However, the situation is quite different in countries such as India, where public health spending is only 1.25% of the GDP. As a result, 58.7% of total healthcare expenditure is paid directly by individuals out of their own pockets [3]. Healthcare has emerged as a major category of public expenditure, requiring substantial government investment. Its increasing share in budgets worldwide underscores its growing importance in the global economy [4]. The financial consequences of illness extend beyond immediate medical costs, particularly affecting vulnerable populations such as families with elderly members or those suffering from chronic diseases, who face the danger of being impoverished as a result of medical costs [5].

The global discussion on health equity has gained momentum since the World Health Organization's Alma-Ata Declaration. It has become a crucial policy objective, spurring concerted efforts to reduce inequalities. However, persistent concerns remain regarding access to and quality of healthcare in nations with poor and moderate incomes [6]. Across Asia, significant public health challenges arise from disparities in healthcare access, costs, and service delivery. Reliance on out-of-pocket spending pushes many families into catastrophic expenditures and poverty, with India being a prominent country in this worrying trend [7]. Regional disparities in health financing reveal stark differences in public commitment to health systems. In the period analysed, public funding contributed only in certain areas, slightly more than a quarter of all health spending.

This contrasts sharply with East Asia and the Pacific, Europe, and Central Asia, where public funds cover approximately three-quarters of health spending [8]. Sub-Saharan Africa even has a greater proportion of public funding than some developing countries, highlighting the severity of the funding gap in countries like India. Social determinants significantly impact access to healthcare and the financial burden [9]. In India, the

Received date: December 07, 2025 **Accepted date**: December 15, 2025; **Published date**: December 23, 2025

***Corresponding Author**: Mittapally. R, Business Intelligence Architect/AI and ML Engineer, USA , E- mail: mittapallyrakesh@gmail.com

Citation: Mittapally. R (2025). Assessing Normality in Healthcare Expenditure Data: A Shapiro-Wilk Test Approach In Python. International Journal of Computer Science and Data Engineering, 2(4), 1–8 doi: https://dx.doi.org/10.55124/csdb.v2i4.272

caste system creates hierarchies that deepen existing inequalities. Belonging to a lower caste exacerbates the hardships of poverty and worsens health disparities. This is evident in the higher rates of malnutrition among lower-caste and tribal children [10]. Policy interventions have attempted to address these challenges through various insurance plans. In the previous twenty years, Indian governments have launched publicly financed health programs, such as the Universal Health Insurance Program of 2003, which was restructured in 2008 and became the Rashtriya Swasthya Bima Joana [11]. International examples, such as Mexico's PI financing scheme, demonstrate alternative approaches designed to guarantee fast, effective, and high-quality healthcare services while mitigating out-of-pocket expenses for uninsured families [12]. Gender disparities persist, as disaggregated data show that women often face disadvantages due to higher healthcare needs at certain life stages and lower financial capacity to pay for treatment [13]. Integrating multiple data sources improves accuracy and establishes robust foundations for healthcare expenditure estimates [14]. Methodological approaches to studying healthcare expenditures have advanced significantly.

The analysis of healthcare expenditures relies on comprehensive national datasets, such as the Medical Expenditure Panel Survey, which gathers comprehensive information on healthcare use and expenditures, examining cost distributions and their associations with age, socioeconomic status, insurance coverage, and financial resources [15]. Current research indicates a lack of studies that combine theoretical and empirical analyses to examine the factors contributing to high medical costs and the affordability of healthcare, with studies often failing to consider how family structure impacts high medical expenses [9]. Although public health expenditure forecasting has emerged as an important tool, patient and diagnostic data often remain unanalysed and underutilized, losing its significant potential to shed light on patient healthcare costs [16]. Additionally, there is limited evidence on how beneficiaries' health and financial security are impacted by health insurance in developing countries [18].To analyse the determinants of catastrophic health expenditures in Indian families are family structure, socio-economic status, and caste insurance coverage.

Methodology

This study utilized primary data obtained directly from participants through a specially designed research instrument. This data source was cross-sectional in nature, collecting information from diverse groups of people at a specific point in time. The primary nature of the data ensured that the information collected was specifically tailored to address the research objectives related to individual medical expenses and healthcare costs across various demographic segments.

Data collection was conducted over a three-month period in July, August, and September 2025, using a self-administered structured questionnaire distributed via Google Forms. The questionnaire underwent rigorous development before final use, including extensive literature review, pre-testing for clarity and reliability, and validation for face validity. The research instrument included both closed-ended and multiple-choice questions to facilitate standardized quantitative analysis. Distribution was strategically implemented through multiple digital channels, including social media platforms such as Facebook, WhatsApp, X (formerly Twitter), and LinkedIn, as well as other online communication networks. This multi-platform distribution strategy was deliberately chosen to maximize geographical reach, ensure cost-effectiveness, and achieve representation across diverse social, educational, and professional groups. The study successfully enrolled 1,018 participants who voluntarily provided anonymous responses after being informed about the study's objectives and confidentiality assurances through an introductory note. No personal identifiers were collected to maintain privacy and data security throughout the collection process.

This study employed a comprehensive variable system encompassing both independent and dependent variables. The independent variables recorded socio-demographic characteristics such as age (categorized as 18–29 years old, 30–44 years old, 45–59 years old, and older than 60), gender (male, female, or prefer not to say), educational qualification (primary school, high school, degrees (bachelor's, master's, and doctoral), marital status (widowed, separated, married, and single), and annual income level (below ₹2,00,000, ₹2,00,000–₹4,00,000, ₹4,00,001–₹8,00,000, ₹8,00,001–₹15,00,000, and above ₹15,00,000), as well as occupational status [22]. The dependent variables, focusing on medical expenses and healthcare spending patterns, included costs for consultations, medications, hospitalization, diagnostic tests, and preventive care services. These outcome measures were selected based on their relevance to understanding the economic burden of healthcare and individual financial preparedness in healthcare management, influenced by income level, lifestyle factors, access to healthcare facilities, and health management awareness.

The analytical framework was designed using Python-based statistical methods suitable for Shapiro-Wilk normality tests verify that the data is not normally distributed (all $p < 0.001$). The analytical framework followed a systematic workflow beginning with data import and cleaning procedures, followed by comprehensive descriptive statistics to characterize the sample distribution. Considering the violation of normality assumptions, the analysis utilized non-parametric inference methods, including Mann-Whitney U tests for comparing two independent groups, Kruskal-Wallis tests for comparing three or more groups across categorical variables, and to investigate relationships between continuous or ordinal variables, use Spearman's rank correlation coefficient. Statistical computations were implemented using the scipy.stats module, while data visualization through the matplotlib and seaborn libraries generated detailed graphical representations of categorical health variables, cost patterns, and demographic distributions. The complete analytical pipeline integrated normality testing, non-parametric hypothesis testing, correlation analysis, and multidimensional visualization to provide robust insights into healthcare spending patterns across diverse demographic segments.

Citation: Mittapally. R (2025). Assessing Normality in Healthcare Expenditure Data: A Shapiro-Wilk Test Approach In Python. International Journal of Computer Science and Data Engineering, 2(4), 1–8 doi: https://dx.doi.org/10.55124/csdb.v2i4.272

Results and Discussion

Table 1. Descriptive statistics of health insurance, medical expenses, and health management variables (N = 1018)					
Variable	N	Mean	Std. Deviation	Minimum	Maximum
Last blood pressure check	1009	2.762141	1.275888	1	5
Last cholesterol check	1009	2.712587	1.36155	1	5
Type of health insurance	1009	3.966303	1.553559	1	6
Insurance coverage adequacy	1009	2.73439	0.946389	1	4
Out-of-pocket medical expenditure (last 12 months)	1009	2.709613	0.93643	1	4
Percentage of monthly income spent on medical expenses	1009	2.405352	0.955981	1	4
What would reduce medical expense burden the most	1009	2.766105	1.062903	1	4
Financial support relied on for medical expenses	1009	3.27552	1.339793	1	5
Lifestyle changes tried to manage health	1009	3.450942	1.048346	1	5
Biggest challenge in managing cancer and chronic conditions	1009	3.178394	1.22551	1	5

Table 1 descriptive statistics reveal key patterns in healthcare access and burden for this cohort (N=1018). On average, respondents reported relatively recent blood pressure and cholesterol tests (mean ~2.76 and ~2.71 on a 1-5 regency scale). Health insurance type varied widely (mean 3.97, SD 1.55, range 1-6), while perceived coverage adequacy was moderate (mean 2.73). Out-of-pocket spending burden was significant, with medical expenses consuming a moderate percentage of income (mean ~2.41). Respondents most frequently identified formal solutions (e.g., lower medication costs) as the primary way to reduce costs (mean 2.77) and heavily relied on personal/family savings for medical financing (mean 3.28). Lifestyle modifications were commonly attempted for health management (mean 3.45), but navigating care complexities remained a significant challenge (mean 3.18). The data collectively highlights the intersections between insurance structures, financial strain, and proactive health management efforts.

Table 2. Normality assessment of health insurance, medical expenses, and healthcare management variables (Shapiro–Wilk test, N = 1018)						
Variable	N	Skewness	Kurtosis	Shapiro–Wilk Statistic	Sig. (p-value)	Normality Decision
Last blood pressure check	1009	-0.03321	-1.11899	0.889021	2.68E-26	Not Normal
Last cholesterol check	1009	0.172702	-1.24549	0.884625	9.59E-27	Not Normal
Type of health insurance	1009	-0.34413	-0.81755	0.909115	4.59E-24	Not Normal
Insurance coverage adequacy	1009	-0.19403	-0.91686	0.872824	6.99E-28	Not Normal
Out-of-pocket medical expenditure (last 12 months)	1009	-0.33634	-0.7382	0.868044	2.55E-28	Not Normal
Percentage of monthly income spent on medical expenses	1009	0.131277	-0.91666	0.878967	2.67E-27	Not Normal
What would reduce medical expense burden the most	1009	-0.26285	-1.20458	0.852669	1.2E-29	Not Normal
Financial support relied on for medical expenses	1009	-0.15711	-1.12879	0.890158	3.52E-26	Not Normal
Lifestyle changes tried to manage health	1009	-0.75655	-0.11612	0.848179	5.15E-30	Not Normal
Biggest challenge in managing cancer and chronic conditions	1009	-0.25234	-0.97932	0.901644	6.18E-25	Not Normal

The normality assessment Table 2 demonstrates unequivocally that every measured variable violates the assumption of normal distribution; this is a crucial consideration for subsequent statistical analyses. With a sample size of 1018, the Shapiro-Wilk test results for each variable, including health insurance details, medical expenses, and healthcare management factors, are highly significant (p-values range from 5.15E-30 to 4.59E-24). These p-values are far below the conventional alpha level of 0.05, leading to the conclusion of "non-normality" for each input. The accompanying skewness and kurtosis statistics further describe the specific non-normal shapes of each distribution. Consequently, it becomes necessary to use non-parametric statistical tests for any further inferential analysis involving these variables.

Table 3. Default results and recommended statistical tests for health, insurance, and medical expense variables		
Variable	Normality Decision	Recommended Test
Last blood pressure check	Not Normal	Non-Parametric
Last cholesterol check	Not Normal	Non-Parametric
Type of health insurance	Not Normal	Non-Parametric
Insurance coverage adequacy	Not Normal	Non-Parametric

Citation: Mittapally. R (2025). Assessing Normality in Healthcare Expenditure Data: A Shapiro-Wilk Test Approach In Python. International Journal of Computer Science and Data Engineering, 2(4), 1–8 doi: https://dx.doi.org/10.55124/csdb.v2i4.272

Out-of-pocket medical expenditure (last 12 months)	Not Normal	Non-Parametric
Percentage of monthly income spent on medical expenses	Not Normal	Non-Parametric
What would reduce medical expense burden the most	Not Normal	Non-Parametric
Financial support relied on for medical expenses	Not Normal	Non-Parametric
Lifestyle changes tried to manage health	Not Normal	Non-Parametric
Biggest challenge in managing cancer and chronic conditions	Not Normal	Non-Parametric

Table 3 presents the recommended statistical approach based on the prior assessment of normality. Since each variable related to health examinations, insurance, medical expenses, and health management was definitively determined to be “non-normal” (as described in Table 2), the corresponding recommendation for each is consistently “non-parametric”. This consistent finding across all ten variables provides a clear methodological directive. Therefore, the researcher should abandon parametric tests such as t-tests or ANOVA, which require normally distributed data. Instead, to ensure valid and reliable inferences from this dataset, the analyses should utilize robust non-parametric alternatives such as the Mann-Whitney U test, the Kruskal-Wallis test, or Spearman’s rank correlation.

Figure 1: (a) Box plot showing respondents’ opinions regarding the adequacy of health insurance coverage for medical expenses; (b) Box plot illustrating the distribution of different types of health insurance held by insured respondents.

Figure 1 illustrates respondents’ opinions regarding health insurance coverage and type. In panel (a), the box plot indicates moderate opinions on the adequacy of insurance, with responses clustered around median values, suggesting partial satisfaction but also significant variability. Panel (b) shows a wide distribution across insurance types, reflecting the diversity in insurance plans among insured respondents and varying levels of access and coverage benefits.

Citation: Mittapally. R (2025). Assessing Normality in Healthcare Expenditure Data: A Shapiro-Wilk Test Approach In Python. International Journal of Computer Science and Data Engineering, 2(4), 1–8 doi: https://dx.doi.org/10.55124/csdb.v2i4.272

Figure 2: (a) Respondents' opinions regarding the sufficiency of health insurance coverage for medical costs, and (b) the allocation of health insurance types among insured respondents.

Figure 2 illustrates the respondents' perspectives on health insurance plans and policy types. Most respondents indicate having moderate to adequate coverage for medical expenses, suggesting partial satisfaction with existing insurance plans, while a small proportion reported insufficient coverage. Additionally, the distribution of insurance types reveals a heavy reliance on a few dominant plans, highlighting a lack of diversity in insurance options and potential disparities in healthcare financial protection.

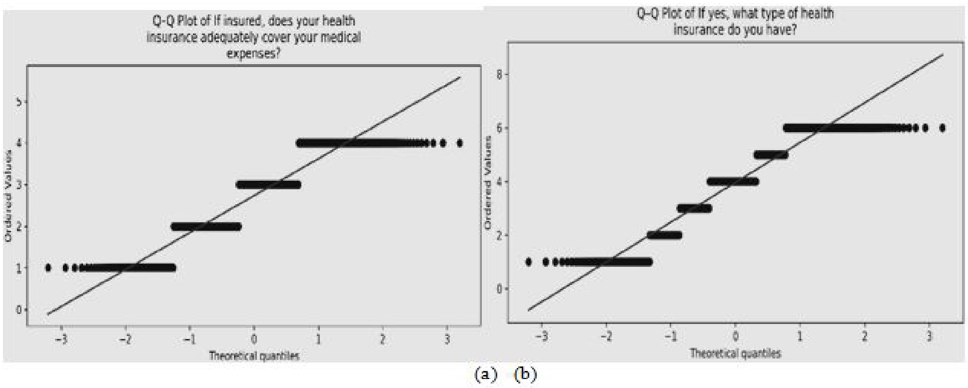

Figure 3: Q–Q plots illustrating (a) respondents' opinions on whether their health insurance adequately covers medical expenses, and (b) the distribution of the types of health insurance received by the respondents..

Citation: Mittapally. R (2025). Assessing Normality in Healthcare Expenditure Data: A Shapiro-Wilk Test Approach In Python. International Journal of Computer Science and Data Engineering, 2(4), 1–8 doi: https://dx.doi.org/10.55124/csdb.v2i4.272

Figure 3 shows that both Q-Q plots exhibit significant deviations from the reference line, indicating that a normal distribution is not exactly followed by the data. The staircase-like patterns reflect the categorical nature of the responses. Overall, the respondents' opinions regarding the adequacy of insurance coverage and the types of insurance received show a skewed distribution, illustrating the variability and clustering within specific response categories.

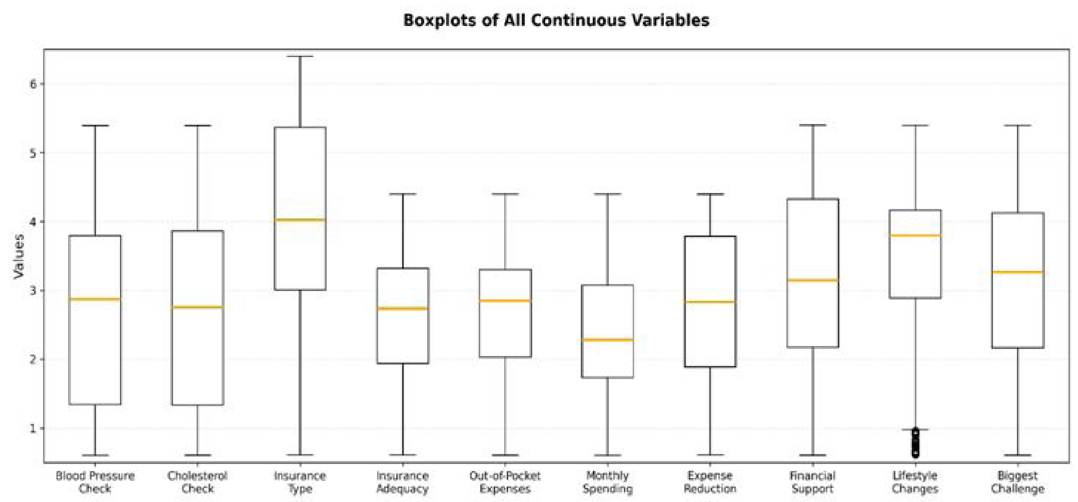

Figure 4: Box plots showing the distribution and variability of ongoing financial, insurance, and health factors among the respondents.

Figure 4 shows significant differences in continuous economical, insurance, and health factors. The differences in median values and inter quartile ranges indicate varying experiences among the respondents; while the presence of outlier values in some variables points to extreme responses. Overall, these box plots highlight the diversity in health behaviours, insurance-related perceptions, and financial circumstances among the respondents.

Conclusion

Based on the research findings, this study reveals the complex interplay between health financing, insurance coverage and household out-of-pocket spending in India. The examination of 1,018 respondents demonstrates that despite moderate insurance coverage, a significant financial burden persists, with medical expenses consuming a substantial portion of household income. A non-parametric statistical approach confirmed that healthcare spending patterns are heavily influenced by socio-demographic factors, insurance adequacy, and access to preventive care. The study found that respondents primarily rely on personal and family savings for medical expenses, indicating inadequate protection against catastrophic health expenditures. Although participants showed proactive engagement through lifestyle modifications for health management, they continue to face considerable challenges in navigating healthcare issues, particularly for chronic conditions. These findings underscore critical gaps in India's healthcare financing system, where low public health spending pushes families into vulnerable financial situations.

The moderate perceived adequacy of insurance coverage, coupled with limited diversity in insurance types, suggests that existing schemes do not adequately protect families from medical financial risks. These results align with broader concerns about health inequities in developing countries, where out-of-pocket payments remain the dominant financing mechanism. Addressing these challenges requires comprehensive policy interventions, including expanding public health financing, strengthening insurance schemes, reducing drug costs, and improving insurance adequacy. Future research should examine the long-term effects of medical costs on poverty in households and evaluate the effectiveness of government insurance initiatives in preventing catastrophic expenditures. Ultimately, achieving universal health coverage requires concerted efforts to reduce financial barriers, improve insurance coverage, and guarantee fair access to high-quality medical care for all socio-economic groups in India.

References

1. Gaddam, Rohin, and K. Ramachandra Rao. "Incidence, inequality and determinants of catastrophic health expenditure in India." Journal of Health Management 25, no. 1 (2023): 30-39.

2. Akhtar, A., Nadeem Ahmad, and Indrani Roy Chowdhury. "Socio-economic inequality in catastrophic health expenditure among households in India: a decomposition analysis." Indian Economic Review 55, no. 2 (2020): 339-369.

3. Perikala. K (2024). Architecting Retail-Scale Product Knowledge Graph Systems. International Journal of Artificial intelligence and Machine Learning, 2(3), 1-6. doi: https://doi.org/10.55124/jaim.v2i3.292

4. Sriram, Shyamkumar, Veenapani Rajeev Verma, Pavan Kumar Gollapalli, and MuayadAlbadrani. "Decomposing the inequalities in the catastrophic health expenditures on the hospitalization in India:

empirical evidence from national sample survey data." Frontiers in Public Health 12 (2024): 1329447.

5. Ud Din, Mohammad Azhar, MuzffarHussain Dar, and ShaukatHaseen. "Inter-state disparities in government health expenditure in India: a study of national rural health mission." International Journal of Health Governance 28, no. 1 (2023): 82-94.
6. Nagababu Kandula. (2025). Evaluating Research Methods for DMR Reporting Using the Balanced Scorecard Approach. International Journal of Computer Engineering and Technology (IJCET), 16(3), 330–352.
7. Ahlin, Tanja, Mark Nichter, and Gopukrishnan Pillai. "Health insurance in India: what do we know and why is ethnographic research needed." Anthropology & medicine 23, no. 1 (2016): 102-124.
8. Dwivedi, Rinshu, and Jalandhar Pradhan. "Does equity in healthcare spending exist among Indian states? Explaining regional variations from national sample survey data." International journal for equity in health 16, no. 1 (2017): 15.
9. Brinda, Ethel Mary, Paul Kowal, JørnAttermann, and Ulrika Enemark. "Health service use, out-of-pocket payments and catastrophic health expenditure among older people in India: The WHO Study on global AGEing and adult health (SAGE)." J Epidemiol Community Health 69, no. 5 (2015): 489-494.
10. Mohanty, Ranjan Kumar, and Deepak Kumar Behera. "Heterogeneity in health funding and disparities in health outcome: a comparison between high focus and non-high focus states in India." Cost Effectiveness and Resource Allocation 21, no. 1 (2023): 44.
11. Divya Soundarapandian. (2024). Predictive Analytics for Customer Loyalty: A Machine Learning Approach to Loyalty Engineering in Digital Banking. International Journal of Information Technology and Management Information Systems (IJITMIS), 15(1), 54–71.
12. Dwivedi, Rinshu, and Jalandhar Pradhan. "Does affordability matter? Examining the trends and patterns in health care expenditure in India." Health Services Management Research 33, no. 4 (2020): 207-218.
13. RamPrakash, Rajalakshmi, and Lakshmi Lingam. "Why is women's utilization of a publicly funded health insurance low?: a qualitative study in Tamil Nadu, India." BMC Public Health 21, no. 1 (2021): 350.
14. Sharma, Santosh Kumar, Jaison Joseph, and DevakiNambiar. "Assessing inequalities in publicly funded health insurance scheme coverage and out-of-pocket expenditure for hospitalization: findings from a household survey in Kerala." International Journal for Equity in Health 22, no. 1 (2023): 197.
15. Perikala. K (2024). Large-Scale Architecture for Retail Platforms Using Cloud-Native Big Data Systems. International Journal of Computer Science and Data Engineering, 1(3), 1–7 doi: https://dx.doi.org/10.55124/csdb.v1i3.268
16. Martínez-García, M., J. Vargas-Barrón, F. Bañuelos-Téllez, H. González-Pacheco, C. Fresno, E. Hernández-Lemus, M. A. Martínez-Ríos, and M. Vallejo. "Public insurance program impact on catastrophic health expenditure on acute myocardial infarction." Public Health 158 (2018): 47-54.
17. Nagababu. K. (2025). Enterprise-Level Emphasis Operational Analytics: Predicting Healthcare Encounter Resolution Success Using Machine Learning.International Journal of Robotics and Machine Learning Technologies, 1(2), 1-6, doi: https://doi.org/10.55124/ijrml.v1i2.242
18. Sharma, Santosh Kumar, DevakiNambiar, HariSankar, Jaison Joseph, Surya Surendran, and Gloria Benny. "Gender-specific inequalities in coverage of publicly funded health insurance schemes in southern states of India: evidence from national family health surveys." BMC Public Health 23, no. 1 (2023): 2414.
19. Bernard, D., Cowan, C. A., Selden, T. M., Cai, L., Catlin, A., &Heffler, S. (2012). Reconciling Medical Expenditure Estimates from the MEPS and NHEA, 2007. 2(4). https://doi.org/10.5600/MMRR.002.04.A09
20. Miller, Jeffrey D., Talia Foster, Luke Boulanger, Meredith Chace, Mason W. Russell, Jeno P. Marton, and Joseph Menzin. "Direct costs of COPD in the US: an analysis of Medical Expenditure Panel Survey (MEPS) data." COPD: Journal of Chronic Obstructive Pulmonary Disease 2, no. 3 (2005): 311-318.
21. Kulkarni, Mukund, Dhammadeep D. Mesharam, BhagyeshPatil, Rahul More, Mirdual Sharma, and PravinPatange. "Medical insurance cost prediction using machine learning." 2022 International Journal for Research in Applied Science & Engineering Technology 10 (2022).
22. Gurubasannavar. S D, Sunku. R, Dandasi. V V. (2025). Selecting an Extract, Transform, and Load (ETL) Software Solution: A Comprehensive Evaluation and Comparison . International Journal of Cloud Computing and Supply Chain Management, 1(3), 1-7. doi: https://doi.org/10.55124/ijccscm.v1i3.249
23. Tang, Yao, Tianran Chen, Yuan Zhao, and FarhadTaghizadeh-Hesary. "The impact of the long-term care insurance on the medical expenses and health status in China." Frontiers in Public Health 10 (2022): 847822.
24. Sood, Neeraj, EranBendavid, ArnabMukherji, Zachary Wagner, SomilNagpal, and Patrick Mullen. "Government health insurance for people below poverty line in India: quasi-experimental evaluation of insurance and health outcomes." Bmj 349 (2014).
25. Divya Soundarapandian (2024). Machine Learning-Driven Reliability Engineering for E-Commerce Sites: A Study on Browsing, Cart, and Checkout Phases. Journal of Business Management and Entrepreneurship, 3(1), 1-9. doi: https://doi.org/10.55124/jbme.v3i1.239
26. Perikala. K (2025). Architecting MCP-Based Platforms for Enterprise-Scale Agentic Generative AI. Journal of Business Intelligence and Data Analytics, 2(3), 1-8. https://doi.org/10.55124/jbid.v2i3.264
27. Sriram, Shyamkumar, and M. Mahmud Khan. "Effect of health insurance program for the poor on out-of-pocket inpatient care cost in India: evidence from a nationally representative cross-sectional survey." BMC Health Services Research 20, no. 1 (2020): 839.
28. Kiran Kumar Mandula Samuel, Venkata Pavan Kumar Aka. (2023). AI for Special Education Closing Gaps in Inclusive Learning Using MOORA Method. International Journal of Computer Engineering and Technology (IJCET), 14(1), 249-267.
29. Blackburn, Justin, and Shinae Choi. "Patterns and factors associated

with medical expenses and health insurance premium payments." Choi, S., & Blackburn, J.(2018). Patterns and factors associated with medical expenses and health insurance premium payments. Financial Counseling and Planning 29, no. 1 (2018): 6-18.

30. Lala, Yohanes Brian Caesaryano, Raymond SunardiOetama, and Kimberly Lvina. "The Effect of Video Games Towards the Students' Academic Performance." IJNMT (International Journal of New Media Technology) 11, no. 2 (2024): 49-55.

31. Venkata Pavan Kumar Aka and Kiran Kumar Mandula Samuel. (2023). Predictive Modeling for Brownfield Migration from SAP ECC 6.0 To S4HANA A Machine Learning Approach for Effort Evaluation Under the SAP Rise Program. International Journal of Computer Engineering and Technology (IJCET), 14(1), 232-248. DOI: https://doi.org/10.34218/IJCET_14_01_018

32. Ziegenfuss, Jeanette Y., Elanadora U. Sour, Erica J. Roelofs, Jennifer M. Vesely, Karen L. Margolis, and Stephanie A. Hooker. "A randomized study comparing patient portal and email communications for trial recruitment." Clinical Trials 22, no. 5 (2025): 597-606.

33. Suresh Deepak Gurubasannavar, Varun Venkatesh Dandasi, Raghavendra Sunku. (2025). Enhancing Smart Home Automation with AI And Topsis-Based Decision Making. International Journal of Information Technology and Management Information Systems (IJITMIS), 16(6), 1-22.

34. Divya Soundarapandian. (2024) Reliability Analysis of Data Science Workflow Components Using SPSS A Correlation-Based Study. International Journal of Computer Science and Data Engineering, 1(2), 1–7. doi: https://dx.doi.org/10.55124/csdb.v1i2.266

Citation: Mittapally. R (2025). Assessing Normality in Healthcare Expenditure Data: A Shapiro-Wilk Test Approach In Python. International Journal of Computer Science and Data Engineering, 2(4), 1–8 doi: https://dx.doi.org/10.55124/csdb.v2i4.272