TWMS J. App. and Eng. Math. V.15, N.3, 2025, pp. 501-510

NONPARAMETRIC STATISTICAL HYPOTHESIS TESTING IN SOFT SET THEORY

C. R. PARVATHY[1], A. SOFIA[1*], §

Abstract. Theories of uncertainty plays a vital role in decision making. Efforts are being made in combining statistical hypothesis testing methods with uncertainty theories having membership and nonmembership values. Soft set theory was developed as a generalization of fuzzy set theory to avoid having difficulties in assigning membership values. In this paper, an attempt is made to impose statistical hypothesis testing methods in soft set theory to handle decision making problems with linguistic data. For this purpose, non-normality of the data has been analyzed by using Shapiro Wilk test for normality following which Skillings Mack nonparametric test has been computed in soft data using chi-squared distribution and Monte Carlo method. To demonstrate this, significance difference in the sample data set of manpower positions (Radiographers, pharmacists, lab technician, nurses and specialty doctors) in the community health centers in Southern states of India from 2019 to 2022 has been computed.
Tools used: *R*

Keywords: Soft set, decision making, hypothesis testing, Skillings Mack test, Shapiro Wilk test.

AMS Subject Classification: 62G10, 62D10, 62-04

1. Introduction

Theories of uncertainty are important for studies involving decision making. To avoid difficulties in determining membership values, fuzzy set theory was generalised to create a soft set theory. Molodtsov [28] initiated the theory of soft sets. It was further investigated by Maji et al [25]. This theory has application in various fields, like decision making, game theory, operations, etc. There have been several works related to combination of soft set theory and other uncertainty tools in solving decision making problems. The relationship between soft set and the fuzzy set was studied by Alcantud[4] and the relationship between the soft, rough, and fuzzy sets was investigated by Feng et al [15],[16]. A review on soft set based parameter reduction and decision making was done by Sani Danjuma et al[13]. In [5] the authors presented a new class of generalized fuzzy soft sets, also defined

[1] Department of Mathematics, PSGR Krishnammal college for Women, Coimbatore, Tamil Nadu, India.
email: parvathytopo@gmail.com; ORCID:0000-0001-5601-053X.
email: sofiaarjunan11@gmail.com; ORCID:0000-0001-9113-4351.
* Corresponding author.
§ Manuscript received: September 30, 2023; accepted: February 03, 2024.
TWMS Journal of Applied and Engineering Mathematics, Vol.15, No.3;

arithmetic operators, and multi-criteria decision making strategy, and demonstrated a decision making problem to validate the defined method. In [26] the authors studied rough bipolar soft sets and applied them to decision making problems. In [20] the authors introduced a set called K_m^n rung picture fuzzy set and defined related operators to deal with unreliability in multi-attribute group decision making problems and compared with many other existing decision making operators.

In recent years, theories of uncertainty have been extended over two different universes. Sabir Hussain [18] generalized the concept of soft topological space over two different universes, defined a few important concepts like a binary soft open, closed set, interior, closure, and neighbourhood, and studied its properties. Furthermore, Sabir Hussain et al.[19] introduced binary soft mapping, defined and studied the properties of binary soft images, and inverse images, and applied it to a real-world problem dealing with medical diagnosis. As an extension, Benchalli et al.[10] defined and studied various separation axioms, binary soft topological, and hereditary properties. Zhiming Zhang [36] defined a rough set approach to an intuitionistic fuzzy soft set by using fuzzy relation, and threshold value pair, and validated this method with an example. Naime Demirtas and Orhan Dalkilic [14] proposed binary bipolar soft sets over two universal sets, defined operations like union, intersection, complement, AND, OR, etc., and studied their properties.

Kanwal and Shabir[21]proposed soft binary relations to approximate fuzzy sets, introduced a new approach to find upper and lower approximations, and also applied to semigroups and approximated fuzzy sub semigroups, ideals, and bi-ideals of semigroups. Generating new knowledge from existing knowledge is an important task in data mining. M. Z. Anwar et al.[7] proposed intuitionistic fuzzy ideals based on soft binary relations, they also approximated IF sub semigroups, ideals, and bi-ideals of semi-groups. Remya and Shalini[31] defined Pythagorean vague binary soft sets and various distance measures and also discussed higher dimension stage q-rung orthopair vague binary soft sets. M. Z. Anwar et al.[6] authors defined the multi granulation rough Set of intuitionistic fuzzy Sets and introduced approximation operators using foresets and aftersets, and studied its algebraic properties.

Statistical hypothesis testing is one of the major areas of statistical analysis and is being widely used in decision making problems. It involves obtaining decision about the assumptions of a population parameter using sample data obtained from the population. Parametric and non-parametric are two approaches in hypothesis testing. The use of non-parametric approach is a standard practice in problems which fail to satisfy the distributional assumption. There are many real life situations with imprecise data where classical statistical hypothesis methods cannot be used. Thus, many authors developed statistical hypothesis to analyze fuzzy and vague data [17],[22],[24],[27]. Considering interval valued data, Smarandache generalized the fuzzy logic in the sense of neutrosophic [32]. Further, neutrosophic statistics was defined as a generalization of classical statistics [33]. This test statistics was then applied by many authors in problems containing uncertainty and vague data[1]-[3],[11],[30] and also with decision-making analysis [23],[29],[34]. Homogeneity of variance, goodness of fit and Kolmogorov Smirnov test are introduced by Aslam for uncertain data using neutrosophic statistics [8],[9].

Though hypothesis testing has been extended over the theories of uncertainty, it is mainly extended to fuzzy sets, intuitionistic fuzzy sets, and neutrosophic sets. That is, an extension of hypothesis testing is associated only with the sets involving membership and non-membership functions. Soft set theory was developed as a generalization of the fuzzy set which has become one of the most important theories of uncertainty.

The main advantage of soft set theory is the presence of the parameter set.It is to be noted that the parameter set of soft set theory is different from the statistical parameters. Soft set theory is defined as a mapping from the parameter set to the power set of the universe set. In recent years, soft set theory has been extended over two dimensional universe sets. Let U, V be two different non-empty finite universe sets, K be the subset of the parameter set E. A pair (m, k) is called a soft binary relation over U, V , if (m, k) is a soft set over $U \times V$. Here, the parameter set is linguistic (it can be a word, sentence etc.,).

To the best of our knowledge, there is not much literatures involving a combination of soft set theory and hypothesis testing. Thus, in this paper, an attempt is made to combine statistical hypothesis with soft set theory. Soft set theory relies on approximations and does not impose restrictions on parameter choice, hence it is challenging to determine whether a parameter choice affects decision-making significantly.

To ensure the choice of parameters and to deal the problems with no membership values, hypothesis testing has been used in soft set theory which involves

(1) Formulation of hypothesis
(2) Conversion of sample to soft set and soft matrix.(Since, soft matrix is often binary or with ties, we use non-parametric method, in specific, Skillings Mack method).
(3) Computation
(4) Result

To analyse the non-normality of the data, the Shapiro-Wilk test for normality was used, and then the Monte Carlo method and chi-squared distribution were used to construct the Skillings Mack nonparametric test for soft data. In some of the problems where only binary data is involved, Cochran's Q nonparametric test can be used. An example is given where the decision maker compares the rating for a few specifications (parameters) of a mobile phone in six different e-commerce platforms. In addition, the sample data set of manpower positions (radiographers, pharmacists, lab technicians, nurses, and specialty doctors) at the community health centers in rural areas of Southern states of India from 2019 to 2022 has been computed to show the significant difference in this regard [35]. Results are computed in *R* using *Ski. Mack* package developed by Chukhrova et al [12].

The contents of this paper are organized as follows: In section 2 the statistical testing used in soft set theory is given in detail with examples. In section 4 an application is given using secondary data to validate the method described in this paper. Followed by the conclusion and future works in section 4.

2. Hypothesis testing in soft set theory

In this section hypothesis testing is merged with soft set theory to ensure if the choice of parametres makes a significant difference in decision making. An example is demonstrated to validate the proposed method.

Since the soft set is often linguistic, the mean and variance of a soft set need not be known. Thus, to perform the nonparametric test (Skillings Mack test) in soft set theory, the following assumptions has to be satisfied.

(1) Sufficient sample size
(2) Data/observed value should be at least an ordinal or continuous
(3) The samples need not be normally distributed

Let $U \times V = \{(u_i, v_j); i, j \in N\}$ where and are nonempty finite sets. Define a soft set $m(k_i)$ over $U \times V$ and Compute a soft matrix. Decide the treatments and blocks depending

on the decision making problem. In this paper, parameter sets are considered as groups and the elements of $U \times V$ are as blocks.

Let (u_{ik}, v_{jl}) be the observations of the soft set, remove any block having only one observation. Assign ranks to observations within each blocks from lowest to highest. In terms of ties, mean of the ranks are assigned. In case of any missing value in a block, $(k_i + 1)/2$ has to be assigned as a rank for missing value. In next step, arrange the table with ranks of the observations and find the sum of the ranks by r_{ij} . Decide the hypothesis assumptions of the problem, define null and alternate hypothesis. Further, the Skillings Mack test statistic for ranks without ties is computed in R , using the following formula,

$$T = A^{'}\mathcal{B}^{-}A$$

where, $A = \sum_{i=1}^{n}(12/k_i + 1)^{1/2}(r_{ij} - (k_i + 1)/2)$ for $j = 1, 2, .., k$ is the adjusted sum treatment, $A^{'}$ is the vector of k adjusted sum A and $\mathcal{B}^{-}$ is the inverse of covariance matrix $\mathcal{B}$.

The following algorithm is used to compute the desired result.

Algorithm:

(1) Decide the hypothesis of the problem.
(2) Construct the soft sets from the desired data using cartesian product over two different universes.
(3) Write the soft matrix from the constructed soft set.
(4) Frame null hypothesis and alternate hypothesis.
(5) Decide test statistic and calculate the p-value.
(6) Make a decision from the obtained p-value.

Example 2.1. *Skillings Mack test on Soft set theory with data available Assume an example where a person needs to buy a mobile phone say Oppo A74. He wants to cross check if the reviews are same in few websites before buying.*

Let $U = \{M_1\}$ be the mobile phone considered. He decides to check the details of the considered mobile phones in 6 different websites $V = \{A, F, P, D, G, M\}$ say amazon, flip kart, price baba, digit technology, gearwatch.in, my next phone. The parameters considered here are Battery life (e_1) , camera capacity (e_2) and design (e_3).

Here, the soft set taken over $U \times V$ is an absolute soft set. And the observations here are the ratings given for the parameter set in each website for mobile M_1.

Table 1 represents the observations. Here, the parameters e_1, e_2, e_3 are considered as

Table 1. Ratings

$U \times V/E$	e_1	e_2	e_3
(M_1, A)	*4.0*	*3.8*	*3.7*
(M_1, F)	*3.9*	*3.9*	*3.9*
(M_1, P)	*4.0*	*4.0*	*3.5*
(M_1, D)	*4.0*	*4.0*	*3.9*
(M_1, G)	*4.5*	*2.8*	*3.5*
(M_1, M)	*4.4*	*4.3*	*3.8*

groups and the elements of $U \times V$ be the blocks.

The null hypothesis H_0 is considered as there is no significant difference among the ratings in each website.

Alternate hypothesis H_1 is that there is significant difference among ratings.

Skillings Mack's test statistic at 0.05 significance level is computed using R software *and the output of the Skillings Mack test statistic is shown in table 2.*

Table 2

Result
Skillings-Mack Statistic = 5.333333, p-value = 0.069483
Note: the p-value is based on the chi-squared distribution with d.f = 2
Based on B=1000, Simulated p-value = 0.016000

Since, there exists ties in the ranks we consider the simulated p-value. From table 2, it is evident that the simulated p-value is less than 0.05. Thus, we reject null hypothesis concluding that the reviews are positive in each websites considered.

Example 2.2. *Since the soft set is linguistic, the observations need not be available for every problem. In such cases, binary values can be used in such a way that,*

$$(u_{ik}, v_{jl}) = \begin{cases} 1, \textit{if } (u_i, v_j) \in m(k_i) \\ 0, \textit{Otherwise} \end{cases} \tag{1}$$

where $m(k_i)$ is the considered soft set. Thus, problems with categorical data can be solved using Cochran's Q test.

3. Hypothesis testing of Soft set theory in decision making

In this paper, soft set theory has been considered over the soft binary relation of two different universes in such a way to make two different decisions simultaneously. Thus to apply this in real world, we consider a problem to check if there is a significant difference in manpower positions in the community health centers in rural areas from 2019-2022 in the southern states of India. For this purpose the data has been collected from yearly rural health statistics published by Ministry of Health and Family Welfare Statistics Division, Government of India.

This problem has been carried forward by considering the five southern states of India (Andhra Pradesh, Karnataka, Kerala, Tamil Nadu and Telangana) as a universe set U and the manpower positions (Radiographers, pharmacists, lab technicians, nursing staff, specialty doctors) as other universe set V. Let the parameter set be $E = 2019, 2020, 2021, 2022$.

Here, an absolute soft set is considered, the soft matrix with the observed data is given in table 3.

3.1. **Method.** Non-normality of the problem has been analyzed using Shapiro Wilk test for normality and histograms for each year. The p-values of Shapiro Wilk test statistic evaluated using R are 0.00685, 0.000407, 0.00121 and 0.0006464 for the years 2019, 2020, 2021 and 2022 respectively.

It is evident that the p-values obtained are less than 0.05. Also, from the figure 1, it is obvious, that the data fails to obey normal distribution, since, the curves are positively skewed.

Null hypothesis is assumed to be as there is no significant difference in the manpower positions in the community health centers from 2019-2022. Alternate hypothesis is that

TABLE 3

States,position/year	2019	2020	2021	2022
(AP, R)	40	56	48	47
(AP, P)	1016	1039	1205	1216
(AP, L)	817	913	1122	1197
(AP, N)	2861	3088	3770	3949
(AP, S)	2861	3088	3770	3949
(Ka, R)	167	122	114	114
(Ka, P)	1973	1749	2793	1641
(Ka, L)	1469	1725	1724	1823
(Ka, N)	3511	4741	4133	4858
(Ka, S)	3511	4741	4133	4858
(K, R)	0	12	12	13
(K, P)	1299	774	1166	1189
(K, L)	372	565	664	716
(K, N)	2427	1760	2826	2940
(K, S)	2427	1760	2826	2940
(TN, R)	92	144	212	220
(TN, P)	2043	1585	1572	1582
(TN, L)	1064	1165	1533	1570
(TN, N)	5888	7072	7811	7911
(TN, S)	5888	7072	7811	7911
(T, R)	44	44	44	24
(T, P)	442	442	442	553
(T, L)	647	647	647	526
(T, N)	2097	2076	2076	1583
(T, S)	2097	2076	2076	1583

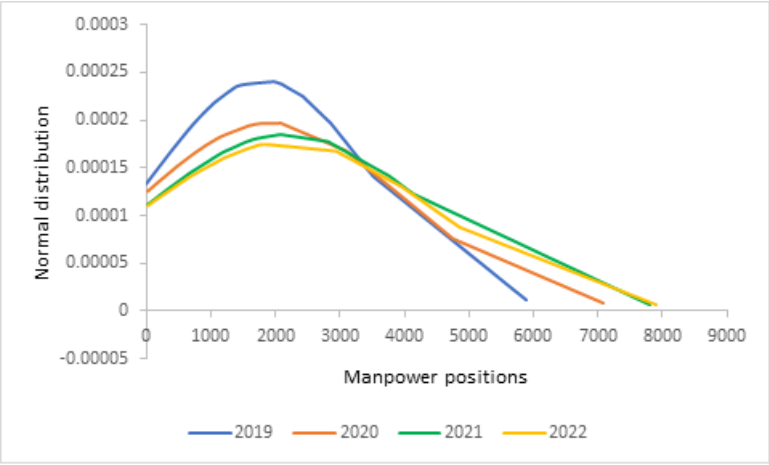

FIGURE 1

there is significant difference in manpower positions. Due to the presence of ties in the ranks, the p values are simulated for different values of B. Here, elements of $U \times V$ is

considered as blocks and the parameter set (years) as groups. Then, the observations are ranked within blocks and ranks are adjusted in the presence of ties. Adjusted sum treatment, covariance matrix, its inverse, and Skillings Mack test statistic is computed using *R software* for $B = 1000, 10000$ at 5% significance level.

3.2. **Result.** Skillings Mack test statistic and p value with respect to chi-squared distribution and simulated p-value with respect to Monte Carlo method computed using *Ski.Mack* in *R* are shown in table 4, 5.

Table 4

Result
Skillings-Mack Statistic = 10.668000, p-value = 0.013664
Note: the p-value is based on the chi-squared distribution with d.f = 3
Based on B=1000, Simulated p-value = 0.008000

Table 5

Result
Skillings-Mack Statistic = 10.668000, p-value = 0.013664
Note: the p-value is based on the chi-squared distribution with d.f = 3
Based on B=10000, Simulated p-value = 0.008700

As ranks are adjusted due to the presence of ties, we consider the simulated p-values computed using Monte Carlo method. It is clear that the simulated p-values 0.008 for B=1000 (shown in table 4) and 0.009 for B=10000 (shown in table 5) are less than the significance level 0.05. Hence, we reject null hypothesis. Thus, we conclude that there is significance difference in the manpower positions in the community health centers in Southern states of India.

Since the hypothesis result is significant, Durbin pairwise comparsion test was performed using Bonferroni p value adjustment method and the result is shown in table 6. It is evident that, p value 0.006 is less than 0.025, implying that there is a significant differenece betweeen 2019 and 2022.

Table 6

Result
Pairwise comparison using Durbin's all pairs test for a two-way balanced incomplete data
year 2019 2020 2021
2020 1.000
2021 0.265 1.000
2022 0.006 0.111 1.000
p value adjustment method: bonferroni

4. Conclusion

Through this work, testing of hypothesis has been extended over soft set theory. Skillings Mack nonparametric test has been implied to soft set theory. To demonstrate this, a sample data set containing manpower positions in the community health centers in southern states of India has been used validating that there is a significance difference in the manpower positions in the community health centers in Southern states of India from 2019 to 2022. Though there is no pairwise comparison is performed for Skillings Mack test, we try Durbin's pairwise comparison using bonferroni p value adjustment method. Further, many other statistical test can be studied over soft set theory using this technique and other statistical characteristics can be observed.

References

[1] Abdel-Basset, M., Nabeeh, N. A., El-Ghareeb, H. A., and Aboelfetouh, A., (2020), Utilising neutrosophic theory to solve transition difficulties to IoT-based enterprises, Enterprise Information System, 14(9-10), 1304-1324.

[2] Abdel-Basset, M., Chang, V., and Gamal, A., (2019), Evaluation of the green supply chain management practices: a novel neutrosophic approach, Computers in Industry, 108, 210-220.

[3] Abdel-Basset, M., Atef, A., and Smarandache, F., (2019), A hybrid neutrosophic multiple criteria group decision making approach for project selection, Cognitive Systems Research, 57, 216-227.

[4] Alcantud, J. C. R., (2016), Some formal relationships among soft sets, fuzzy sets and their extensions, International Journal of Approximate Reasoning, 68, 45-53.

[5] Al-Shami, T. M., Alcantud, J. C. R., and Mhemdi, A., (2023), New generalization of fuzzy soft sets: (a, b)-fuzzy soft sets, AIMS Mathematics, 8(2), 2995-3025.

[6] Anwar, M. Z., Bashir, S., Shabir, M., and Alharbi, M. G., (2021), Multigranulation roughness of intuitionistic fuzzy sets by soft relations and their applications in decision making, Mathematics, 9, 2587.

[7] Anwar, M. Z., Bashir, S., Aslam, M., and Shabir, M., (2022), Approximations of intuitionistic fuzzy ideals over dual spaces by soft binary relations, Journal of Function Spaces, Article ID 3996256, 17 pages.

[8] Aslam, M., (2021), A new goodness of fit test in the presence of uncertain parameters, Complex and Intelligent Systems, 7, 359-365.

[9] Aslam, M., (2019), Introducing Kolmogorov-Smirnov tests under uncertainty: an application to radioactive data, ACS Omega, 5(1), 914-917.

[10] Benchalli, S. S., Patil, P. G., Dodamani, A. S., and Pradeepkumar, J., (2017), On binary soft separation axioms in binary soft topological spaces, Global Journal of Pure and Applied Mathematics, 13(9), 5393-5412.

[11] Broumi, S., Talea, M., Bakali, A., Smarandache, F., Ullah, K., (2018), Bipolar neutrosophic minimum spanning tree, Smart Application and Data Analysis for Smart Cities (SADASC'18).

[12] Chukhrova, N., and Johannssen, A., (2021), Fuzzy hypothesis testing: systematic review and bibliography, Applied Soft Computing, 106, 107331.

[13] Danjuma, S., Herawan, T., Ismail, M.A., Abubakar, A.I., Zeki, A.M., and Chiroma, H., (2004), A review on soft set based parameter reduction and decision making, International Journal of General Systems, 33(5), 569-581.

[14] Demirtas, N., and Dalkilic, O., (2023), Binary bipolar soft sets, Boletim da Sociedade Paranaense de Matematica, 41, 1-12.

[15] Feng, F., Li, C., Davvaz, B., and Ali, M.I., (2010), Soft set combined with fuzzy sets and rough sets: a tentative approach, Soft Computing, 14(9), 899-911.

[16] Feng, F., Liu, X., Leoreanu-Fotea, V., and Jun, Y. B., (2011), Soft sets and soft rough sets, Information Sciences, 181(6), 1125-1137.

[17] Hecke, T. V., (2012), Power study of anova versus Kruskal-Wallis test, Journal of Statistics and Management Systems, 15, 241-247.

[18] Hussain, S., (2019), On some structures of binary soft topological spaces, Hacettepe Journal of Mathematics and Statistics, 48(3), 644-656.

[19] Hussain, S., and Alkhalifah, M. M. A., (2020), An application of binary soft mappings to the problem in medical expert systems, Journal of Applied Mathematics and Informatics, 38(5-6), 533-545.

[20] Ibrahim, H. Z., Al-shami, T. M., Arar, M., Hosny, M., (2023), k_m^n-rung picture fuzzy information in a modern approach to multi-attribute group decision-making, Complex and Intelligent Systems.
[21] Kanwal, R. S., and Shabir, M., (2019), Rough approximation of a fuzzy set in semigroups based on soft relations, Computational and Applied Mathematics, 38, 89.
[22] Kruskal, W. H., and Wallis, W. A., (1952), Use of ranks in one-criterion variance analysis, Journal of the American Statistical Association, 47, 583-621.
[23] Lou, Y., Yuen, S. Y., and Chen, G., (2018), Evolving benchmark functions using kruskal-wallis test, GECCO'18: Proceedings of the Genetic and Evolutionary Computation Conference Companion, ACM, 1337-1341.
[24] MacFarland, T. W., and Yates, J. M., (2016), Kruskal-Wallis H-Test for oneway analysis of variance (ANOVA) by ranks, Introduction to Nonparametric Statistics for the Biological Sciences Using R, 177-211.
[25] Maji, P. K., Biswas, R., and Roy, A. R., (2003), Soft set theory, Computers and Mathematics with Application, 45(4-5), 555-562.
[26] Malik, N., Shabir, M., Al - Shami, T. M., Gul, R., and Mhemdi, A., (2023), Medical decision - making techniques based on bipolar soft information, AIMS Mathematics, 8(8), 18185-18205.
[27] McKight, P. E., and Najab, J., (2010), Kruskal - Wallis test, The Corsini Encyclopedia of Psychology, 1, 1-10.
[28] Molodtsov, D., (1999), Soft set theory – first results, Computers and Mathematics with Applications, 37(4-5), 19-31.
[29] Muremi, L., and Bokoro, P., (2018), Assessing the significance of electro - thermal stress on varistor arresters using Kruskal - Wallis H - test, 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe, (EEEIC/I and CPS Europe), 1-4.
[30] Nabeeh, N. A., Abdel-Basset, M., El-Ghareeb, H. A., and Aboelfetouh, A., (2019), Neutrosophic multi-criteria decision making approach for iot-based enterprises, IEEE Access, 7, 59559-59574.
[31] Remya, P. B., Shalini, A. F., (2019), Pythagorean vague binary soft sets, The International Journal of Analytical and Experimental Model Analysis, 11(9), 845-855.
[32] Smarandache. F., (2003), Definition of neutrosophic logic - a generalization of the intuitionistic fuzzy logic, Proceedings of the Third Conference of the European Society for Fuzzy Logic and Technology, University of Applied Sciences, Zittau, Germany, EUSFLAT, 141-146.
[33] Smarandache, F., Khalid, H. E., and Essa, A. K., (2018), Neutrosophic logic: the revolutionary logic in science and philosophy-Proceedings of the National Symposium.
[34] Soltani, N., Safajou, F., Amouzeshi, Z., and Zameni, E., (2017), The relationship between body image and mental health of students in Birjand in 2016 academic year: a short report, Journal of Rafsanjan University of Medical Sciences, 16, 479-486.
[35] Zhang, Z., (2012), A rough set approach to intuitionistic fuzzy soft set based decision making, Applied Mathematical Modelling, 36, 4605-4633.
[36] Ministry of health and family welfare url https://main.mohfw.gov.in/

C. R. Parvathy received her Ph.D in Mathematics from Bharathiar University in 2018. She is currently working as an associate professor and Head in the department of mathematics (UG-SF), PSGR Krishnammal College for Women affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India. Her area of interest is Topology. She has published many research articles in the areas of ideal topology, soft topology, algebraic topology etc., in national and international journals.

A. Sofia is a research scholar in the department of mathematics, PSGR Krishnammal College for Women affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India. Her area of interest is Topology, soft set theory, rough set theory.