Role of location of household and its socio-economic status on energy consumption dynamics in rural Nepal: a categorical data analysis

Jyoti U. Devkota
Department of Mathematics, School of Science, Kathmandu University, Nepal

Article Info

ABSTRACT

This study is based on data collected from two sample surveys. They are namely survey of 300 households of national grid energy users and 400 households of biogas users. It was conducted in three different rural settings of Nepal. The responses to questions were classified into multiple choice options. This generated categorical data and reduced ambiguity and confusion between interviewer and interviewee. Such data were classified into ordinal scale and modelled. As the dependent variable had more than two categories, polytomous and not dichotomous models are developed and fitted. Ten different hypotheses assessing and measuring the energy consumption dynamics are tested. Values of parameters of these model and odds ratio are used in quantifying the impact of change with respect to energy consumption. The variables considered were namely time spent in the collection of firewood, type of house, amount of firewood saved, time saved, employer and school located within 15 min distance. Such data based studies are very crucial for country like Nepal which lacks a strong backbone of accurate and regularly updated official records. They can be generalized to other counties of Asia and Africa. Such results provide guidelines to policy makers and planners regarding formulation of realistic energy policies for such countries.

Corresponding Author:

Jyoti U. Devkota,
Department of Mathematics,
Kathmandu University,
Dhulikhel 45200, Kavre, Nepal.
Email: drjdevkota@ku.edu.np

1. INTRODUCTION

In developed countries immense energy sources are needed to keep pace with the rapid development. Whereas in developing countries marked with long hours of power outages, need for continuous electricity is important for social and economic development. There is a rising cost of exploitation of ever depleting fossil fuel reserves. This is associated through pollution with an adverse impact on the ambient environment. So there is a growing urgency to switch over to alternative and renewable sources of energy.

Total energy consumption of Nepal in the year 2008/09 was about 9.3 million tons of oil equivalents (401 million GJ), out of which 87 percent were derived from traditional resources, 12 percent from commercial sources and less than 1 percent from the alternative sources [1]. Here traditional energy resources include fuel wood from forests and tree resources, agricultural residues coming from agricultural crops and animal dung in the dry form. Energy resources primarily coming from coal, grid electricity and petroleum products are termed as commercial whereas biogas, solar power, wind and micro level hydropower are categorized into the alternative energy resources in Nepal.

Nepal has an agriculture based economy where cattle are kept for their draft power and milk. Biogas generates energy from the dung produced by farm animals and also from agricultural wastes. Thus biogas as an alternative energy source is very suitable for Nepal. The potential of producing biogas is about 1.9 million plants out of which 57% in terrain, 43% in hills and mountains. Biogas plants were promoted by Government of Nepal in agriculture year 1974/75 as a part of special program. This program installed 250 plants in different part of Nepal under the supervision of governmental and non governmental agencies. As a part of alternative energy year 2009/2010 government of Nepal planned to install 100,000 biogas plants in 70 districts [1]. There were 201775 biogas plants in 2009, installed in various districts of Nepal [1, pg 60]. According to 2011 census, 131,596 households use biogas for cooking, where 14.5% households are in urban areas and 85.5% are in rural areas. This is 2.52% of the total household. Households depending on wood/firewood for cooking are 63.99%. The use of Kerosene for cooking is in 1.02% of the households, whereas LPG gas is used in 21.03% for the same purpose. Kerosene is used in 18.28% households for cooking purposes. Similarly households using electricity as a source of lighting is 67.2% [2].

Evidence based studies are very important for many countries as they lack the backbone of strong and good quality official data. Special statistical techniques have to be applied and developed for such countries with limited and scarce data. Several factors playing a critical role in energy consumption dynamics need to be identified and quantified. Here, there are several intangible advantages of energy use especially in rural areas. These outweigh directly attributable advantages and benefits.

In this paper the impact of various factors governing the energy consumption dynamics of rural households is quantified. Categorical data are analysed and polytomous models are developed. The work is novel as unlike other papers the use of statistics is not superficial. Statistical methods are used in generation of categorical data and its in-depth analysis. This approach is unique to energy research problems. This method is useful in situations where we have lack of accurate measurement instruments. It can also be useful in situations where due to lack of awareness, exact answers cannot be furnished, but a multiple choice option can be chosen correctly. Here dependent variables are classified in more than two categories and are hence polytomous and not dichotomous. We interpret the parameters of these models. We also calculate odds ratio and odds in favour for quantification of impact. Categorizing data into different groups which can be later reduced to ordinal data has reduced the chances of error due to ambiguous response. Thus the dynamics of change of variables related to energy consumption of 700 households such as time spent in the collection of firewood, type of house, amount of firewood saved, time saved, employer and school located within 15 min distance are minutely analyzed. The data used here is based on sample survey of 300 households of normal users which are national grid energy users and 400 households of biogas users.

Devkota [3] has discussed and developed several statistical methods for countries with limited and scarce data demographic data. This approach has wide applicability in various interdisciplinary fields. Saleh et al. [4] implemented statistical optimization of parameters and conditions for reduction of consumed chemicals and reagents in experimental works. Kanak et al. [5] applied statistics and artificial intelligence and developed models with aim to reduce the risk of incidence of blood related anemia. Statistical analysis was used in evaluation of corrosion resistant steel bars in sustainable building construction by Imam et al. [6]. Rui et al. [7] have analysed causes, characteristics and consequences of tunnel fire accidents in China using statistical methods. Pacheco et al. [8] analyzed the compressive strength of three Portuguese cement brand using probabilistic modelling. Bhattacharyya [9] studied the access of energy to India's poor. Whereas an overview of energy consumption pattern by available data and the analysis of some relevant aspects of energy policy in rural China are presented by Zhang et al. [10]. Petrides and Furnham [11] analyzed several dimensions human's emotional intelligence with exploratory factor analysis. Similarly Garcia et al. [12] used multivariate statistics is used to estimate the theoretical, technical and economic potentials of biomasses for bioenergy production. They used Hierarchial cluster agglomerates and Principle components. Sun et al. [13] used stochastic processes to select from large number of scenarios for representing the variability of operating points, a suitable scenario. This was used in transmission network expansion planning. Xu et al. [14] used multivariate statistical regression model to accurately predict the output power of photovoltaic grid-connected power generation. This section is followed by section 2 on theoretical background and hypothesis titled research methods which is followed by section 3 on results and discussion. This is followed by a section 4 titled conclusion.

2.   RESEARCH METHOD

Categorical data is the only means of getting accurate data in various studies. Categorizing data into several groups reduces the chances of having ambiguous and erroneous data. This is especially true for countries with limited data. Due to lack of awareness among various stakeholders and a very limited (often unreliable) database, categorical data minimize the ambiguity of response between interviewee and interviewer. These data are also useful when the measurement instruments of data collection are not very

precise and accurate. Identification of all the possible response to a question and then putting them into several categories reduces the chances of generating faulty data due to uncertainty and vagueness. These data can be classified in such a way that the resulting data can be reduced to put on ordinal scale. Due to large sample size the ordinal data can be treated as continuous data by central limit theorem. As the dependent variable takes more than two values, polytomous models instead dichotomous models are suitable.

2.1. Measures

The probability distributions of attributes I and J can be cross tabulated into contingency tables in the following manner as shown in Table 1. This cross tabulation gives a better overview of the data. Special case of this I×J contingency table is a 2×2 contingency table, which is given in Table 2.

Table 1. Contingency table of order I×J

j i	1	2	3	……	….	J	Total
1	π_{11}	π_{12}	π_{13}	……	……	π_{1I}	$\pi_{1.}$
2	π_{21}	π_{22}	π_{23}	……	……	π_{2I}	$\pi_{2.}$
…..	….			…..			…..
I	π_{J1}	π_{J2}	π_{13}	……	……	π_{JI}	$\pi_{J.}$
Total	$\pi_{.1}$	$\pi_{.2}$	$\pi_{.3}$	……..	………	$\pi_{.I}$	

Table 2. Contingency table of order 2×2

i	1	2	Total
1	π_{11}	π_{12}	$\pi_{1.}$
2	π_{21}	π_{22}	$\pi_{2.}$
Total	$\pi_{.1}$	$\pi_{.2}$	

Odds ratio can be used in quantifying the impact of a technique. This is done by calculating the ratio of two conditional probabilities of different response values under the same condition. When the response is dichotomous then the odds are [15]:

$$\frac{\pi_{1|j}}{\pi_{2|j}} = \frac{\pi_{1|j}}{1-\pi_{1|j}} = \frac{\pi_{1j}}{\pi_{2j}}$$

so in the 2×2 table given above for j=1,

$$\frac{\pi_{1|1}}{\pi_{2|1}} = \frac{\pi_{1|1}}{1-\pi_{1|1}} = \frac{\pi_{11}}{\pi_{21}}$$

for j=2,

$$\frac{\pi_{1|2}}{\pi_{2|2}} = \frac{\pi_{1|2}}{1-\pi_{1|2}} = \frac{\pi_{12}}{\pi_{22}}$$

also $\pi_{1|j} = \frac{\pi_{1j}}{\pi_{j.}}$, j=1, 2.

Based on product law of compound events;

$$\pi_{1|1} = \mu'\alpha_1'$$

$$\pi_{2|1} = \mu'(1 - \alpha_1')$$

$$\ln\left(\frac{\pi_{1|1}}{\pi_{2|1}}\right) = \ln\left(\frac{\pi_{1|1}}{1-\pi_{1|1}}\right) = \ln\left(\frac{\mu'\alpha_1'}{1-\mu'\alpha_1'}\right) = \mu + \alpha_1$$

$$\ln\left(\frac{\pi_{1|j}}{\pi_{2|j}}\right) = \ln\left(\frac{\pi_{1|j}}{1-\pi_{1|j}}\right) = \mu + \alpha_j \quad j = 1, 2$$

or,

$$\ln\left(\frac{\pi_{1|j}}{\sqrt{\pi_{2|j}\pi_{1|j}}}\right) = \mu + \alpha_j \dots (1)$$

or,

$$\log\left(\frac{\pi_{1|j}}{\dot{\pi}_j}\right) = \mu + \alpha_j$$

$$\dot{\pi}_j = (\prod_i \pi_{i|j})^{\frac{1}{2}}, j = 1, 2$$

as,

$$\ln(\sqrt{\frac{\pi_{1|j}}{\pi_{2|j}}}) = \mu + \alpha_j$$

or,

$$\frac{1}{2}\ln(\frac{\pi_{1|j}}{\pi_{2|j}}) = \mu + \alpha_j$$

$$\ln\left(\frac{\pi_{1|j}}{1-\pi_{1|j}}\right) = \mu + \alpha_j$$

$$\frac{\pi_{1|j}}{1-\pi_{1|j}} = e^\mu e^{\alpha_j}$$

$$\pi_{1|j} = \frac{e^{\mu+\alpha_j}}{1+e^{\mu+\alpha_j}}$$

$$\pi_{2|j} = 1 - \pi_{1|j} = \frac{1}{1+e^{\mu+\alpha_j}}$$

When the response variable is not dichotomous that is it has more than two options, then it is called polytomous. So polytomous response models have more than two categories, say more than two categories in row or column or both. This can be explained by multinomial probability density function. Here:

$$\log(\dot{\pi}_j) = \frac{1}{I}\sum_i \log(\pi_{i|j})$$

where i = 1, 2, …I ; j = 1, 2, …J.

$\dot{\pi}_j$ is the value of the probability obtained in this way for the jth category of the explanatory variables. This value is called the geometric mean. Back-transformation, eliminating the logarithm, shows its definition:

$$\dot{\pi}_j = (\prod_i \pi_{i|j})^{\frac{1}{I}}$$

So, polytomous logistic models with one explanatory variable can be written as:

$$\log\left(\frac{\pi_{i|j}}{\dot{\pi}_j}\right) = \mu_i + \alpha_{ij} \quad i = 1, 2, …I ; j = 1, 2, …J \tag{1}$$

Here there is one such equation for each category j of explanatory variable, as well as for each category of response i. This is a set of $I \times J$ linear equations to describe how the multinomial is changing in different categories of the explanatory variable. Because the response categories are compared with a mean following constraints are imposed on the parameters: $\sum_i \mu_i = 0$, $\sum_i \alpha_{ij} = 0 \ \forall j$ and $\sum_j \alpha_{ij} = 0 \ \forall i$. The estimates may be obtained by solving I sets of J equations. Polytomous model given in (1) satisfy properties that are given below by (2) and (3):

$$\ln\left(\frac{\pi_{i|j}}{\dot{\pi}_j}\right) = \mu_i + \alpha_{ij}$$

$$\dot{\pi}_j = (\prod_i \pi_{i|j})^{\frac{1}{I}} = (\pi_{1|j}\pi_{2|j}………\pi_{I|j})^{\frac{1}{I}}, j = 1, …J$$

also,

$$(\frac{\pi_{1|j}}{\dot{\pi}_j})\left(\frac{\pi_{2|j}}{\dot{\pi}_j}\right)\left(\frac{\pi_{3|j}}{\dot{\pi}_j}\right)…..\left(\frac{\pi_{I|j}}{\dot{\pi}_j}\right) = 1 \tag{2}$$

or,

$$\ln\left(\frac{\pi_{1|j}}{\acute{\pi}_j}\right) + \ln\left(\frac{\pi_{2|j}}{\acute{\pi}_j}\right) + \ln\left(\frac{\pi_{3|j}}{\acute{\pi}_j}\right) + \cdots .. + \ln\left(\frac{\pi_{I|j}}{\acute{\pi}_j}\right) = 0$$

LHS of (2);

$$\frac{\pi_{1|j}\pi_{2|j}\pi_{3|j}\cdots\cdots\cdots\cdots\pi_{I|j}}{\acute{\pi}_j{}^I} = \frac{\pi_{1|j}\pi_{2|j}\pi_{3|j}\cdots\cdots\cdots\cdots\pi_{I|j}}{\left(\pi_{1|j}\pi_{2|j}\pi_{3|j}\cdots\cdots\cdots\cdots\pi_{I|j}\right)^{\frac{I}{I}}} = 1$$

also,

$$\ln\left(\frac{\pi_{i|j}}{\acute{\pi}_j}\right) - \ln\left(\frac{\pi_{k|j}}{\acute{\pi}_j}\right) = \ln\left(\frac{\pi_{i|j}}{\pi_{k|j}}\right) \tag{3}$$

as, LHS;

$$\ln\left(\frac{\pi_{i|j}}{\acute{\pi}_j}\right) - \ln\left(\frac{\pi_{k|j}}{\acute{\pi}_j}\right) = \ln\left(\frac{\frac{\pi_{i|j}}{\acute{\pi}_j}}{\frac{\pi_{k|j}}{\acute{\pi}_j}}\right) = \ln\left(\frac{\pi_{i|j}}{\pi_{k|j}}\right) = R.H.S$$

Hypothesis, there is dependence between the location and socio-economic status of the house on energy consumption dynamics.

Hypothesis 1   : For household living within 15 minutes to the school the dependence on firewood is less (normal users)

Hypothesis 2   : For household living within 15 minutes to the employer the dependence on firewood is less (normal users)

Hypothesis 3   : Low socio economic status (indicated by the type of house) implies more time spent on the collection of firewood (normal users)

Hypothesis 4   : Low socio economic status (indicated by the type of house) implies more kilograms of firewood consumed (normal users)

Hypothesis 5   : Low socio economic status (indicated by the type of house) implies less litters of kerosene consumed (normal users)

Hypothesis 6   : Low socio economic status of biogas owners (indicated by the type of house) implies more time spent on the collection of firewood before the installation of plant

Hypothesis 7   : Low socio economic status of biogas owners (indicated by the type of house) implies more time spent on the collection of firewood after the installation of plant

Hypothesis 8   : Low socioeconomic status (indicated by type of house) implies more time saved after construction of biogas plant

Hypothesis 9   : Low socioeconomic status (indicated by type of house) implies more firewood saved after a switch over to biogas plant for cooking

Hypothesis 10 : More time spent on collection of firewood before the construction of biogas plant implies "relatively" more time spent in after the construction of biogas plant

3. RESULTS AND DISCUSSION

3.1. Data

The primary data collected for this study are from 700 households. They are obtained from two sample surveys of 300 and 400 households of normal energy users and biogas consumers respectively. These are households inhabiting in different regions of Nepal. In these two surveys the basic set of questions was the same, only some questions to be asked from these two different categories of respondents were modified. Pre-test of questionnaire and training to the interviewer ensured the quality of collected data. The possible response was provided as a multiple choice option with answers classified on an ordinal scale. The details of variables are given in Table 3. This Table 3 shows that variables analyzed here are categorical data classified on ordinal scale. As the sample size is large this ordinal data can be treated as a continuous data by using the central limit theorem.

3.2. Result

The responses to questions in the questionnaire were structured. All the possible answers to each question were properly worked out during the pre-test. It was mentioned as a multiple choice option. This resulted in a categorical data that could be classified on ordinal scale. For example as we see from Table 3, the 'time spent in the collection of firewood' for biogas users is classified into six categories. According to the amount of time devoted for the collection of firewood, these categories are labelled as 0, 1, 2, 3, 4, 5. Here 0 stands for no time spent whereas 5 denote maximum time of 1-2 hours spent. These numbers are on ordinal scale as these values signify the amount of time spent. As seen from Table 3, this holds true for national grid energy users. Same is true for the attribute 'type of house' labelled as 1, 2, 3 and 4 for both biogas users and grid energy users. Here household with concrete house with label 1 is highest in the socioeconomic status whereas the mud house labelled 4 ranks lowest in the socioeconomic status. For the attributes 'amount of firewood saved and amount of time saved', the values labels are 0, 1, 2 and 3, 1, 2, 3 and 4 respectively. These numbers are also classified on ordinal scale going from smallest to highest as the amount of firewood saved and amount of time saved increases. Similarly, 'distance from employer' and 'distance from school' are also classified on ordinal scale of 1, 2 proportional to the closeness from these places.

Table 3. Details of categorical data

User/Sample size	Variable name	Values on ordinal scale	Frequency
Biogas	Time spent in the collection of firewood before		
(400 household)	No time spent	0	33
	<15 minutes	1	15
	15-30 minutes	2	21
	30-45 minutes	3	18
	45-60 minutes	4	53
	1 hour-2 hours	5	260
	Time spent in the collection of firewood after		
	No time spent	0	153
	<15 minutes	1	137
	15-30 minutes	2	67
	30-45 minutes	3	24
	45-60 minutes	4	7
	1 hour-2 hours	5	12
	Type of house		
	Concrete	1	166
	Tile/asbestos	2	17
	Modern light roof	3	77
	Mud house	4	140
	Amount of firewood saved after biogas		
	Up to 30 Kg	1	39
	30-50 Kg	2	115
	Above 50 Kg	3	246
	Time saved		
	No change	0	13
	Less than 60 minutes	1	132
	1 hour to 3 hours	2	195
	3-5 hours	3	23
	More than 5 hours	4	37
Normal	Time for the collection of firewood		
(300 household)	Not applicable	0	13
	<15 minutes	1	29
	15-30 minutes	2	176
	30-45 minutes	3	78
	45 min-1 hour	4	04
	Type of house		
	Concrete house	1	4
	Tiled/asbestos	2	89
	Modern light roof	3	185
	Mud house	4	22
	Employer within 15 minutes		
	Yes	1	17
	No	2	283
	School within 15 minutes		
	Yes	1	86
	No	2	214

Then the attributes mentioned in ten hypotheses of section 2 are tested for independence using chi square test of independence of attributes. Table 4 shows these ten hypotheses under the assumption of a true null hypothesis. These results of the hypothesis testing using chi square test of goodness of fit are shown in Table 4. G^2 is a maximum likelihood estimate of the degree of association between these variables. It is useful when the cell frequency is 0. As seen from Table 4, both chi square and G^2 give same results. The results are highly significant for eight out of ten hypotheses shown in Table 4. As seen from Table 4 for national grid energy users there is a great degree of dependence between times spent in the collection of firewood and distance of the household from the school. Similarly there is dependence between time spent in the collection of firewood and distance of the household from the employer. The dependence between socio-economic status (indicated by type of house) and time taken to collect the firewood is highly significant. As seen from Table 4, the dependence between socio-economic status and amount of firewood used is also highly significant.

Table 4. Results of hypothesis testing at 5% level of significance

User	Null hypothesis	Chi square	G^2	P value (chi Square)	Conclusion
National grid	There is no dependence between the location of school and time taken to collect firewood	19.545	21.311	0.031	There is a dependence
	There is no dependence between the location of employer and time taken to collect firewood	10.611	9.021	0.001	There is a dependence
	There is no dependence between socioeconomic status (indicated by type of house) and time take to collect firewood	56.320	53.607	0.000	There is a dependence
	There is no dependence between socioeconomic status and kilograms of firewood consumed per month	30.823	19.359	0.005	There is dependence
	There is no dependence between socioeconomic status and litters of Kerosene consumed per month	15.343	11.878	0.08	There is no dependence
Biogas	There is no dependence between socioeconomic status and time take to collect firewood before the plant construction	57.566	49.425	0.00	There is a dependence
	There is no dependence between socioeconomic status and time take to collect firewood after the plant construction	20.174	20.947	0.165	There is a no dependence
	There is no dependence between socioeconomic status and amount of time saved after the construction of biogas plant	29.759	35.478	0.003	There is a dependence
	There is no dependence between socioeconomic status and amount of time saved after the construction of biogas plant	29.788	27.504	0.00	There is a dependence
	There is no dependence between the time take to collect the firewood before the plant and time taken to collect the firewood after the plant	128.374	139.804	0.00	There is a dependence

Hypothesis 1: For household living within 15 minutes to the school the dependence on firewood is less (normal grid users)

The method of computation of parameters for polytomous model suited to data given in Table 5 is explained in detail under hypothesis 1. The values of the parameters obtained from this model under hypothesis 1 are given in Table 6. Table 7 gives the values of all parameters obtained by fitting model to the cross tabulated data. Here positive estimates indicate over representation and negative estimates indicate under representation of the variables. The strength of these relationships is further described in terms of parameters of models fitted to the data obtained from the contingency tables. Interpretation of these parameters in terms of odds ratio is provided in Table 8.

Table 5. Cross tabulation of how much time is spent in the collection of firewood and is their school within 15 minutes from their home (grid users)

		How much time do they spend on fuel wood collection					Total
		0	1	2	3	4	
Is their school within 15 minutes	No	12	26	110	64	2	214
Walking distance from their home	Yes	1	3	66	14	2	86
Total		13	29	176	78	4	300

Now the data in Table 5 is modelled using the following methodology:

$$\ln\left(\frac{\pi_{i|j}}{\hat{\pi}_j}\right) = \mu_i + \alpha_{ij}$$

$$\log(\hat{\pi}_{i|j}) - \frac{1}{I}\sum_i \log(\hat{\pi}_{i|j}) = \log(n_{ij}) - \frac{1}{I}\sum_i \log(n_{ij}) = \mu_i + \alpha_{ij}$$

From Table 5 we get the following results;

$$\frac{\ln(12)+\ln(26)+\ln(110)+\ln(64)+\ln(2)}{5} = 3.059103$$

$$\frac{\ln(1)+\ln(3)+\ln(66)+\ln(14)+\ln(2)}{5} = 1.724094$$

$$\ln(12) - 3.059103 = \hat{\mu}_1 + \alpha_{11}$$

$$\ln(1) - 1.724094 = \hat{\mu}_1 + \alpha_{21}$$

$$\hat{\mu}_1 + \alpha_{11} = -0.5742$$

$$\hat{\mu}_1 + \alpha_{21} = -1.72409$$

$$\alpha_{21} = -\alpha_{11}$$

$$\hat{\mu}_1 = -1.1492, \alpha_{11} = -0.57495, \alpha_{21} = 0.57495$$

$$\ln(26) - 3.059103 = \hat{\mu}_2 + \alpha_{12}$$

$$\ln(3) - 1.724094 = \hat{\mu}_2 + \alpha_{22}$$

$$\hat{\mu}_2 + \alpha_{12} = 0.198994$$

$$\hat{\mu}_2 + \alpha_{22} = -0.62548$$

$$\alpha_{22} = -\alpha_{12}$$

$$\hat{\mu}_2 = -0.2133, \alpha_{12} = 0.4123, \alpha_{22} = -0.4133$$

Similarly
$\hat{\mu}_3 = 2.0535, \alpha_{13} = 0.4121, \alpha_{23} = -0.4121$
$\hat{\mu}_4 = 1.0074, \alpha_{14} = -0.0924, \alpha_{24} = 0.0924$
$\hat{\mu}_5 = -1.6985, \alpha_{15} = 0.6675, \alpha_{25} = -0.6675$

The values of the parameters obtained from cross tabulation data provided Table 5 is written in Table 6. From Table 6 we get that, the average odds in favour of no time spent in the collection of firewood versus 45 min-1 hour spent is $e^{-1.1492-(-1.6985)} = 1.732$. However for those respondents who have school within 15 minutes is $e^{-0.57495-0.6675} = e^{-1.2425} = 0.28867$ times this average whereas those who do not have schools within 15 minutes $e^{+0.57495+0.6675} = e^{1.2425} = 3.46$ times this average. So those who have schools more than 15 minutes away are 3.46 times (the average) more likely to spend 0 time in the collection of firewood. So the odds of a household not having school less than 15 min and spending no time in the collection of firewood versus spending 45 min-1 hour is $\frac{12/214}{2/214} = 6 = 1.732 * 3.46$.

Table 6. Parameters of the model

Parameter	Time spend on fuel wood collection				
	1	2	3	4	5
$\hat{\mu}$	−1.1492	−0.2133	2.0535	1.0074	−1.6985
α_1	−0.57495	0.4123	0.4121	-0.0924	0.6675
α_2	0.57495	−0.4123	-0.4121	0.0924	−0.6675

However the odds in favour of person having school less than 15 minutes from home and spending no time in the collection of firewood versus spending 45 min-1 hour is 0.5. The odds of not having school within 15 minutes and spending no time in the collection of firewood 12 times more than that of having school with 15 minutes. So the data here does not validate hypothesis 1. The details of parameters obtained fitting polytomous models are given in Table 7. The impact of interrelationship between these variables given in Table 7 is summarized in terms of odds ratio in Table 8.

Table 7. Details of polytomous models

User/Sample size	Adjective	Variable	Time taken (minutes)				
			No time	<15	15-30	30-45	45-60
Normal (300 households)	Time taken to collect firewood versus school within 15 min of home	School					
		μ_i (average)	-1.1492	-0.2133	2.0535	1.0074	-1.6985
		α_{i1} (yes)	-0.57495	0.4123	0.4121	-0.0924	0.6675
		α_{i2} (no)	0.57495	-0.4123	-0.4121	0.0924	-0.6675
		α_{i1} (concrete house	0.5922	0.7949	-1.8184	-0.5754	1.0068
		α_{i2} (tiled/asbestos)	-0.5002	0.1474	0.657	0.8741	-1.1812
		α_{i3} (modern light roof)	0.4367	-1.4867	0.5912	1.2204	0.1114
		α_{i4} (mud house)	0.3417	0.5445	0.5701	-1.5191	0.06301
	Time taken to collect firewood versus employer within 15 min of employer	Employer					
		μ_i (average)	-0.2925	0.001395	1.7588	0.1788	-1.6466
		α_{i1} (yes)	0.6918	0.00755	-0.0602	-0.8781	0.2541
		α_{i2} (no)	-0.6918	-0.00755	0.0602	0.8781	-0.2541
	Time taken to collect firewood versus type of house	Type of house					
		μ_i (average)	-0.3960	0.4999	1.3215	0.0785	-1.5037
		α_{i1} (concrete house)	0.5922	0.7949	-1.8184	-0.5754	1.0068
		α_{i2} (tiled/asbestos)	-0.5002	0.1474	0.657	0.8741	-1.1812
		α_{i3}(modern light roof)	-0.4367	-1.4867	0.5912	1.2204	0.1114
		α_{i4} (mud house)	0.3417	0.5445	0.5701	-1.5191	0.06301

| | | | Firewood used (kg/month) | | | | |
			0	100	200	300	
	Type of house versus amount of firewood used	Type of house					
		μ_i (average)	-1.5663	-1.0466	2.2235	0.389	
		α_{i1} (concrete house)	0.8733	1.0466	-1.53	-0.389	
		α_{i2} (tiled/asbestos)	-0.3207	-.14744	0.179	0.239	
		α_{i3}(modern light roof)	-0.6897	-1.2096	1.356	0.533	
		α_{i4} (mud house)	0.1373	0.31045	-0.0156	-0.432	

| | | | Kerosene used (liters/month) | | | | |
			0	1	2	3	
	Type of house versus amount of kerosene used	Type of house					
		μ_i (average)	2.03	-0.386	-0.28	1.364	
		α_{i1} (concrete house)	0.471	-0.1341	-0.24	0.844	
		α_{i2} (tiled/asbestos)	0.43	0.1374	0.619	-1.187	
		α_{i3}(modern light roof)	0.077	-0.506	+0.237	-0.819	
		α_{i4} (mud house)	-0.036	0.51	-0.616	1.61	

| | | | Time taken (minutes) | | | | |
			No time	<15	15-30	30-45	45-60
Biogas (400 households)	Type of house versus time spent in the collection of firewood before biogas installation	Type of house					
		μ_i (average)	-0.1115	-0.8868	-0.7948	-0.4581	0.6576
		α_{i1} (concrete house)	0.4212	0.3978	-0.09942	-0.63101	-0.51035
		α_{i2} (tiled/asbestos)	0.1896	0.2718	-0.51333	0.5237	1.030026
		α_{i3} (modern light roof)	-0.169	0.0954	-0.40203	0.39731	-0.35033
		α_{i4} (mud house)	-0.4417	-0.765	1.01477	-0.29001	-0.16935
	Type of house versus time spent in the collection of firewood after biogas installation	Type of house					
		μ_i (average)	1.619992	1.239025	0.4277	-0.6344	-1.4128
		α_{i1} (concrete house)	-0.31207	-0.0346	-0.0118	0.5395	-0.697
		α_{i2} (tiled/asbestos)	0.680932	-0.4044	-0.6919	-0.3228	0.4556
		α_{i3}(modern light roof)	-0.1788	0.3076	0.3567	-0.1216	0.2513
		α_{i4} (mud house)	-0.1899	0.1315	0.3468	-0.0952	-0.0099

| | | | Time saved | | | | |
			30 min	1-3 h	3-5 h	>5 h	no time
	Type of house versus time saved after biogas installation	Type of house					
		μ_i (average)	0.821712	1.74324	-0.68261	-0.73179	-1.15056
		α_{i1} (concrete house)	-0.15683	-0.56319	0.083797	0.70352	-0.06729
		α_{i2} (tiled/asbestos)	-1.08606	0.700459	0.418257	-0.22571	0.19306
		α_{i3}(modern light roof)	0.842625	0.315002	-0.3272	-0.97117	0.140748
		α_{i4} (mud house)	0.400269	-0.45227	-0.17485	0.493365	-0.26652

| | | | Amount of firewood saved in (kg/month) | | | | |
			<=30 kg	30-50 kg	>50 kg		
	Type of house versus amount of firewood	Type of house					
		μ_i (average)	-1.0805	0.39897	0.68160		
		α_{i1} (concrete house)	-0.0997	-0.2355	0.3352		
		α_{i2} (tiled/asbestos)	-0.627	1.225	-0.598		
		α_{i3}(modern light roof)	0.479	-0.541	0.0610		
		α_{i4} (mud house)	0.248	-0.449	0.202		

Table 7. Details of polytomous models (*continue*)

User/Sample size	Adjective	Variable	Time taken (minutes)					
			No time	<15	15-30	30-45	45-60	>60
Biogas (400 households)	Time spent after versus time spent in the collection of firewood before biogas installation	Time spent after biogas installation						
		μi (average)	-0.58287	-0.94764	-0.51339	-0.46169	0.286038	2.21955
		αi1 (no time)	1.378904	-0.93047	0.021568	-0.72327	0.753621	-0.50035
		αi2 (<15 min)	-0.74252	1.494059	0.892749	-0.45823	-1.61142	0.425357
		αi3 (15-30 min)	-1.36678	-1.00201	-0.04997	1.507774	0.536899	0.374093
		αi4 (30-45 min)	-0.4461	-0.08132	-0.51558	0.125876	0.476757	0.440365
		αi5 (45-60 min)	-0.06233	0.302441	-0.13181	-0.18351	0.860521	-0.78531
		αi6 (>60 min)	1.238825	0.217304	-0.21695	-0.26864	-1.01638	0.045844

Table 8. Impacts of factors governing energy consumption in terms of odds ratio

Hypo.	User	Factors	Odds ratio	Conclusion
1	Normal grid energy (300 households)	Distance of the households from the School Versus Time taken to collect Firewood	12 times more in favor of households with schools far way for <15 min of firewood collection per day than 45-60 min per day.	Odds in favor tilted towards households far away from school for spending less time in firewood collection
2		Distance of the households from the Employer Versus Time taken to collect Firewood	3 time more in favor of households with employers close to the household for no times spent in the collection of firewood than 30-45 min spent per day.	Odds in favor tilted towards households close to the employer for spending less time in firewood collection
3		Socioeconomic status (indicated by type of house) Versus Time spent in the collection of firewood	14 time more in favor of households with mud houses than households with concrete houses for spending 15-30 minutes in collection of firewood than no time spent	Odds in favor tilted towards households living in mud houses - Indicator of energy poverty
4		Socioeconomic status (indicated by type of house) Versus Kilograms of firewood consumed per month	1.69 time more in favor of people with mud houses than households with concrete houses for spending 200 kg of firewood per month than 300 kg of firewood	Odds in favor tilted towards households living in mud houses- Indicator of energy poverty
5		Socioeconomic status (indicated by type of house) Versus liters of Kerosene consumed per month	2.59 time more in favor of people with mud houses than households with concrete houses for spending 0 liters of Kerosene per month than 2 liters of Kerosene	Odds in favor tilted towards households living in mud houses- Indicator of energy poverty
6		Socioeconomic status (indicated by type of house) Versus time taken to collect the firewood per day before biogas plant	3 times more in favor of people with mud houses than households with concrete houses for spending more than 60 minutes in the collection of firewood than no time in the collection of firewood	Odds in favor tilted towards households living in mud houses- this is indicator of energy poverty
7		Socioeconomic status (indicated by type of house) Versus time taken to collect the firewood per day after biogas plant	More than 2 times for people with mud houses than people with concrete houses for spending more than 60 minutes in the collection of firewood than no time in the collection of firewood	Odds in favor tilted towards households living in mud houses- this is indicator of reduction in energy poverty
8		Time saved from firewood collection Versus Socioeconomic status (indicated by type of house) after the construction of biogas plant	1.55 times more for people in concrete houses than people in mud houses	Odds in favor tilted towards people in concrete houses. This indicated the benefit of biogas in terms of time saved per day is substantial not only in low socioeconomic groups but also in high socioeconomic groups.
9	Biogas (400 households)	Amount of firewood saved from biogas plant Versus Socioeconomic status (indicated by type of house) after the construction of biogas plant	1.67 times more for people in concrete houses than people in mud houses.	Odds in favor tilted towards people in concrete houses. This indicated the benefit of biogas in terms of firewood saved per day is substantial not only in low socioeconomic groups but also in high socioeconomic groups.
10		Time spent in the collection of firewood before Versus time taken to collect the firewood after biogas construction	More than 1.28 times for households with no time in the collection of firewood after than less than 15 minutes after to spend more than 60 minutes before the plant than no time before the plant.	Odds in favor tilted towards households spending no time in the collection of firewood after the plant-this is indicator of immensity of positive impact in terms of time saved after biogas construction.

4.   CONCLUSION

Benefits of use of energy are not only direct but permeate in an intangible manner to several sectors. The factors related to cause and effect of energy use is studied here using statistics. The results of this paper try to fill the knowledge gap by using statistical methods in quantifying the impact of various factors related to energy consumption. The results obtained here are based on two sample surveys of 400 households of biogas consumers and 300 household of normal energy users. The time and energy spent in the collection of firewood has been minutely analyzed in connection with other variables. The categorical data used here are summarized in Table 3. Ten different hypotheses related to energy consumption pattern of normal users and biogas users are tested with the help of this data. In eight of these ten hypotheses the null hypothesis of no dependence is rejected. It is found that location of school, location of employer and socioeconomic status plays a critical role in the energy consumption dynamics for both types of users. Energy poverty is assessed with the help of socioeconomic poverty with these hypotheses. The use of odds ratio has been used in quantifying the impact. The summary of the impact in terms of odds ratio is provided in Table 8. It is seen that the positive effects of biogas are substantial and immensity of impact of this positive effect is shown here with the help of odds ratio. The results obtained here can be generalized to other countries like Nepal existing in Asia and Africa.

REFERENCES

[1]   WECS (Water and Energy Commission Secretariat), "Energy sector synopsis report 2010," 2010. [Online]. Available at: https://www.wecs.gov.np/uploaded/snyopsis.pdf. [Accessed: 5 May 2020].
[2]   CBS, "National report on national population and housing census 2011," *National Planning Commission Sectariat*, Central Bureau of Statistics, Kathmandu, Nepal, vol. 1, pp. 25-29, 2012.
[3]   Jyoti U. Devkota, "Mortality and fertility models for countries with limited data-results based on demographic data of Nepal, India and Germany," *Lambert Academic Publishing, Saarbruecken*, Germany, 2012.
[4]   Tawfik A. Saleh, Salawu O. Adio, Mohammad Asif, and H. Dafalla, "Statistical analysis of phenols adsorption on diethylenetriamine-modified activated carbon," *Journal of Cleaner Production*, vol. 182, pp. 960-968, May 2018.
[5]   Kanak Meena, Devendra K. Tayal, Vaidehi Gupta, and Aiman Fatima, "Using classification techniques for statistical analysis of anemia," *Artificial Intelligence in Medicine*, vol. 94, pp. 138-152, March 2019.
[6]   I. Taji, S. Ghorbani, J. de Brito, V. W. Y. Tam, S. Sharifi, A. Davoodi, and M. Tavakkolizadeh, "Application of statistical analysis to evaluate the corrosion resistance of steel rebars embedded in concrete with marble granite waste dust," *Journal of Cleaner Production*, vol. 210, pp. 837-846, 2019.
[7]   Rui Ren, Hui Zhou, Zhao Hu, Siyue He, and Xiuling Wang, "Statistical analysis of fire accidents in Chinese highway tunnels 2000-2016," *Tunneling and Underground Space Technology Energy*, vol. 83, pp. 452-460, January 2019.
[8]   João Nuno Pacheco, Jorge de Brito, Carlos Chastre, and Luís Evangelista, "Statistical analysis of ready mixed Portuguese concrete production," *Construction and Building Materials*, vol. 209, pp. 283-294, June 2019.
[9]   Subhes C. Bhattacharyya, "Energy access problem of the poor in India: Is rural electrification a remedy?", *Energy Policy*, vol. 34, no. 18, pp. 3387-3397, December 2006.
[10] Lixiao Zhang, Zhifeng Yang, Bin Chen, and Guoqian Chen, "Rural energy in China: Pattern and policy," *Renewable Energy,* vol. 34, no. 12, pp. 2813-2823, December 2009.
[11] K. V. Petrides and Adrian Furnham, "On the dimensional structure of emotional intelligence," *Personality and Individual Differences*, vol. 29, no. 2, pp. 313-320, August 2000.
[12] D. P. Garcia, J. C. Caraschi, G. Ventorim, F. H. A. Vieira, T P. Protasio, "Assessment of plant biomass for pellet production using multivariate statistics (PCA and HCA)," *Renewable Energy*, vol. 139, pp.796-805, August 2019.
[13] M. Sun, F. Teng, I. Konstantelos, and G. Strbac, "An objective-based scenario selection method for transmission network expansion planning with multivariate stochasticity in load and renewable energy sources," *Energy*, vol. 145, pp. 871-885, February 2018.
[14] Ji Xu, Hong Zhou, and Yanjun Fang, "Prediction of grid-connected photovoltaic power generation based multivariate statistics," *Nonlinear Optics*, vol. 51, pp. 105-119, 2019.
[15] James K. Lindsey, "Introduction to applied statistics a modeling approach," *Oxford University Press*, New York, 1994.

BIOGRAPHY OF AUTHOR

Jyoti U. Devkota is a Professor of Mathematical Statistics in Kathmandu University, Nepal. Her research interests are interdisciplinary applications of Statistics to problems from Energy and Environment. She focuses on development of new methodologies and application of existing methodologies to data generation, correction, prediction and analysis. She obtained her PhD degree from Department of Computer Sciences and Mathematics, University of Osnabrueck, Germany, supported by DAAD fellowship. She completed her Masters and Bachelors degree in Mathematical Statistics from Lady Shriram College, University of Delhi, New Delhi, India.