

ANALYSIS OF HOUSEHOLD ELECTRICITY CONSUMPTION IN NONRESIDENT RENT HALLS USING LINEAR REGRESSION ANALYSIS MODEL

ABSTRACT

This paper is based on electricity consumption pattern in rental houses around Kibabii University (KU) situated in Western region of Kenya. Because of unexpected blackout faced by nonresident students at the time they need electricity most for their studies, this work intends to find out the directive measure to curb this crisis. Since the usage of electricity showed high relationship to the number of households sharing a common meter, Regression analysis prove to be the most effective method to model a solution to this problem. SPSS was used to analyze the data obtained. The results showed the consistency in linear trend of usage of electrical power on a monthly basis among students, it is observed also that the rate of consumption of power among nonresident students of KU is affected by the number of households sharing the meter. The consequence of this study is that with the correct data in place one is able to know the amount of power in kilowatt-hours (KWh) needed for consumption throughout the semester and plan effectively so that power loss is not experienced. The results will be so useful to the KPLC (Kenya Power and Lighting Company) and KU fraternity for planning purposes.

Keywords: Power, ARIMA Model, Regression, Root Mean Square Error, Correlation

1.0 INTRODUCTION AND BACKGROUND

Energy is a key component in day to day operations of every sector. It makes all matter to move from one point to another. However, it has to be generated, distributed and then used. At every point there are huge costs involved and as such the need to be as efficient as possible to get the maximum output. In Kenya electric power is generated by Kenya Electricity Generating Company Ltd (**KenGen**), Emergency Power Producers (EPP) Independent Power Producers (IPPs) and transmitted by (**KPLC**). This power is transmitted country wide to the consumers. The Kenya's demand for power is 1656MW during peak time, and with the country generating 1610MW.

The Kenyan Government has accelerated electrification in the country through various initiatives which includes enhanced funding of the Rural Electrification Program, new connections' financing revolving fund (**Stima loan**) by KPLC. This is aimed at availing credit to potential customers who cannot afford to pay upfront the cost of new connections, transformer maximization where customers within a distance of 600m from a distribution transformer pay uniform connection charges. Line maximization initiative by KPLC Company has created a reinforcement fund for the extension of MV lines and installation of transformers within a reasonable distance to cover existing and potential customers.

This has followed a quick response from the country inhabitant having applied the usage of electricity ranging from industrial, commercial and residential purposes. For example, to run life support machines, to operate electric trains, also in homes for heating, cooling and lighting purposes among others. (**K.P.L.C.**) has been using a postpaid method where customers use electricity and pay their bill later after use. My interest of study has begun with the introduction of pre-pay method (tokens) where one pays and then amount of token corresponding to the cost given in return.

Fellow nonresident students have undergone through hard time by use of these tokens. Most of them at times stay in darkness simply because the amount of token they paid for was finished, this often happened when they are not prepared and without even a single cent to pay for the token. This is a hindering factor towards success academic line since studying at night proves to be impossible without light. This has propelled me to undertake a research on regression analysis to study the rate at which nonresident students consume electricity.

1.1 OVERVIEW OF ELECTRICITY COMSUPTION

Electricity demand in Kenya has been growing rapidly over the last five years, with the latest recorded demand values almost outpacing the projected demand. The core business of Kenya Power Lighting Company is to ensure distribution, transmission and retail of electricity supplied in bulk from the Kenya Electricity Generating Company Limited (KenGen), Independent Power Producers (IPPs), Tanzania Electric Supply Company Limited (TANESCO) and Uganda Electricity Transmission Company Limited (UETCL) [1].

Theoretically, electricity demand is a function of electricity price, number of electricity appliances, consumer incomes and number of households sharing a meter. Selection of the proper methodology and computer program for the electric energy demand forecasting is an extremely important task. A lot of studies over the world have been done in to establishment the correlation between electricity demand and other variables:

In a report on Kenya's energy demand, supply and policy strategy for households, small scale industries and service establishments (KEDSPHS), the Government of Kenya established that at household level, electricity is used for lighting by 99% of the households, entertainment (e.g. television, radios) by 90%, ironing of cloths by 69%, refrigeration by 35%, heating water by 26%, domestic cooking by 24%, home businesses by 16% and house heating by 9%. It further established that the national average household per capita consumption is 694 KWh/year with rural areas using 544 KWh/year and urban 844 KWh/year. Higher income urban households consumed the greatest amount of electricity (1,352 KWh/year) and the low income least (606 KWh). The middle-income class in urban areas consumed 931 KWh/year according to a research done by the Government of Kenya [2].

In the Government report for Well-Being in Kenya, KNBS [2008] the majority of poor households have no access to electricity. Among the rural poor, only 0.7% was connected to electricity compared to 22.4% urban poor. Among the non-poor, 6.5% of households in the rural areas are connected to electricity compared to 62.0% non-poor in the urban areas. Fuel based (paraffin) lighting is the most preferred source of energy for poor and non-poor in rural areas. Further, 0.1% of the poor use electricity as the source of cooking fuel and 1.7% non-poor use electricity for cooking. The report further establishes that the expenditure for electricity for poor households was 5.2% of the total expenditure compared to 20% expenditure on electricity for the non-poor [3].

Houri and Korfali [4] did a study on residential energy usage trend in Lebanon urban areas and it showed that months and seasons have a great effect on consumption of energy. The study also indicated Correlations for the consumption of energy with income, apartment area, and number of residents.

Zheng et al [5] also carried out a study on the issue of characteristics of energy consumption in households of China from 1980 to 2009 with VAR model. The Autoregressive Integrated Moving Average (ARIMA) and BVAR forecasting models were used. The results indicated that both of the models are good in predicting the sustained growth of household energy consumption (HEC) trends.

Saab and other colleagues [6] did a study on the forecasting method on the electric energy consumption in every month in Lebanon. ARIMA and AR (1) modeling methods were used together with a high pass filter. AR (1) is seen to be the best method for forecasting this particular energy data.

Filippini and Pachauri [7] did a regression study about electricity consumption on household expenditure, the mean prices of power and some other fuels and a set of other geographic and socio economic variables. The flexibility used to project electricity demand is based on future income and price scenarios.

Zachariadis et al, [8] did a study on the empirical analysis of consumption of electricity in Cyprus. Results show that elasticities of electricity in long-term used above unity for the income, in the order of -0.3 up-to -0.4 for the prices.

Bianco et al [9] carried out a study that analyzed and forecasted nonresident electricity usage in Romania. Short run non-residential GDP and elasticities of the prices are found to be approximate at 0.136 and 0.0752, respectively, and the long run GDP and elasticities of prices are 0.496 and 0.274 respectively. The two models are then seen to give similar results.

Al-Ghandoor, A., et al [10] performed a study on consumption of electricity and its associated GHG emissions of the industrial sector in Jordan. The study based on the multivariate linear regression in identifying main drivers affecting the consumption of electricity. The results showed that the industrial outputs of production and utilization of capital are the most vital variables affecting the demand of power and that multivariate linear regression technique can be utilized sufficiently in simulating the industrial power consumption associated with very high coefficient of determination (CV).

Payne and James [11] did a survey on consumption of electricity growth. Results of the survey showed that 31.15% of the respondents supported the hypothesis of neutrality, 27.87% on hypothesis on conservation, 22.95% on the hypothesis of growth and finally 18.03% on the hypothesis of feedback.

Yuan et al, [12] studied on consumption of electricity and the economic growth in china. The paper was aimed at establishing the relationship between consumption of electricity and the actual GDP. The results indicated that causality is actually related to electricity consumption.

Narayan et al, [13] also performed a study on the consumption of electricity in the G7 countries. The results showed that the long-run residential factor affecting demand of electricity is the price elasticity and income inelasticity.

Kavousian et al, [14] did a study on the determinants of residential consumption of electricity using a smart meter data. Results showed that location, weather and floor area are among the very important determinants of consumption of electricity in residential areas. Babatunde and shaibu (2009), in examining the residential electricity demand in Nigeria, used annual data from 1970 to 2006 for income, price of electricity, price of substitute and population. As in the case of Narayan and Smyth (2005), income was found to be very significant in long-run. The price of substitute and population were also very significant in the determination of the demand for electricity, but Babatunde and Shaibu (2009) found that, contrary to Narayan and Smyth's findings, the price of electricity was insignificant.

The main mission of Kenya Power Lighting Company is to supply people with power for good and better lives with the vision of providing world-class power that delights its' customers. Despite all these efforts, power cut-off has been observed as a critical issue in most areas with the outcomes being terrible. This makes the company lose a lot of profit the customers also suffering severely. Due to this, a study had to be done about the rate at which customers (residents living around KU) consume power.

1.2 Justification of the study

The findings of this research will assist nonresident students to plan effectively as their semester begins so that they don't miss power. This will also be useful for the university in such a way that through this prediction, the administration will predict amount of electricity they need to operate their machines. Kenya Power Lighting Company will benefit in such a way that the data obtained will be used to calculate the amount of energy needed to supply their customers without gap. Hence the study is of great significance.

2.0 RESEARCH METHODOLOGY

This chapter explains the methods used to formulate the extraction and analysis of data from source population. It gave a clear picture that lead to the conclusion of the true rate of consumption of electrical energy among households. The outcome of this research intended to find solution to prevailing inefficiency in power supply using a mathematical model.

This research uses linear regression to model the rate of household electricity consumptions of rental houses around KU. Performance evaluation of this model is carried out using the Root Mean Square Error (RMSE) and coefficient of determination (r^2).

2.1 Linear Regression analysis

When two variables (x , y) have a linear relationship, the scatter plot tends to cluster around a least squares line. With this linear trend, conclusions are drawn based on the two variables under study. The line gives a mathematical model in form of a linear regression equation shown below;

$$y_i = a + bx_i + \varepsilon_i \quad (2.1.1)$$

Where y_i is the dependent variable, x_i is the independent variable, a and b are the intercept and the regression coefficient respectively while ε_i is called the i^{th} error term.

With the stronger assumption of error terms being normally distributed (having a mean equal to zero;

$\mu = 0$ and a constant variance σ^2) and random and independent, the estimate of the intercept and regression coefficients can be obtained by $y_i = a + bx_i$ where a and b are

$$b = \frac{SS_{xy}}{SS_{xx}} \quad a = \bar{y} - b\bar{x} \quad (2.1.2)$$

Where SS_{xx} and SS_{xy} are the variance of x and covariance between (x, y) respectively, these values can also be calculated from the formulae

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \quad S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} \quad (2.1.3)$$

2.2 Goodness of fit statistic

This is a quantity that measures how best a model gives a clear explanation for some given set of data. The Correlation coefficient (r) measures the extent of strength of linear relationship between some two variables say x and y . r has some limitations, Coefficient of Determination (r^2) is therefore used and it shows the percentage of variations in the dependent variable which is explained by independent variables. Both r and r^2 will be computed as follows;

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} \quad \text{and} \quad r^2 = \frac{bSS_{xy}}{SS_{yy}} \quad \text{where} \quad SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} \quad (3.3.2.1)$$

When the value of r approaches -1, it indicates that the variables x and y have a strong correlation and are negatively related. But if it approaches +1, then it shows a strong positive relationship.

2.3 The Standard Deviation Error

S_e is the standard error and can be used to obtain the confidence interval for b given the level of significance (α), the area in each tail for t distribution given by ($\alpha/2$) with ($n-2$) degrees of freedom, the confidence interval for b will be

$$b \pm t S_b \quad \text{where; } S_e = \sqrt{\frac{SS_{yy} - bSS_{xy}}{n-2}} \quad \text{and} \quad S_b = \frac{S_e}{\sqrt{SS_{xx}}} \quad \text{and } t \text{ is the tabulated value} \quad (3.3.3.1)$$

2.4 Population of the Study

Population is the complete enumeration of individuals with common features under the study according to Mugenda and Mugenda [15]. The study population consisted of the nonresident students of KU using electricity in their hostels who majorly share a common smart meter to pay for their tokens.

2.5 Data Collection

Primary source of data was used; the readings from a smart meter were recorded one month for all the households selected for the experiment. Most households follow a monthly routine. This imply that monthly energy-consumption data is classic for research. Although the residence of the building and the heating patterns might vary throughout the month, the patterns are usually fairly consistent from one month to the next



Figure 3.1



Figure 3.2

2.6 Sampling procedures

The Cluster sampling technique was used in this study. KU area as a population under study has numerous households and may take time to prepare the list of all these houses. It will therefore be subdivided into finite number of distinct and identifiable units, called regions (i.e. Tuuti, Booster, Sewage and Dominion). These regions formed part of our clusters.

Due to resource and time constraint and so as to minimize errors, it was not possible to study the entire population. Mugenda and Mugenda [19] argue that the sample size is an important and useful feature of the study whose purpose is making statistical inference about the population from the sample which best represents the target population. Therefore, a manageable sample was selected. This was obtained using a formula recommended by Creswell [18]. Z is 1.96 at 5% level of significance. N is the population of

study and p is the proportion that we estimate to have the desired characteristics. In our case, we have set $P = 0.95$ since the study population is comprised of students and hence power is a key factor that is very vital. The small value of q is for catering the few who cannot stay in rent halls with power. Fixing the sample error (e) at 5%. during the pre-visit the population size (N) was approximated to be 1000 households, the sample was then calculated as [18]

$$n = \frac{Z^2 pqN}{e^2(N-1) + Z^2 pq} \quad (2.6.1)$$

$$= \frac{1.96^2 \times 0.95 \times 0.05 \times 1000}{0.05^2(1000-1) + 1.96^2 \times 0.95 \times 0.05} = 68.0887 \approx 70 \text{ households}$$

2.7 Assumptions

This study is based on the assumption that in linear models the error terms ε_i are:

- i. Random and independent.
- ii. Normally distributed with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. (homoscedasticity)
- iii. Linearity

3.0 DATA ANALYSIS AND INTERPRETATIONS

This section presents the analysis and interpretations of the data collected. The research was conducted on a sample of 70 households within the KU surrounding, who happens to be the residents of houses that use electricity to which observations were used to gather data in this residential area.

3.1 Data analysis

The collected data is used to obtain the linear regression model which shows the relationship between the independent variable number of household and the dependent variable, the amount of power consumed in kilowatts per hour. The regression model is of the form;

$y_i = a + bx_i$ Where y_i is the amount of electricity consumed in kilowatts per hour, a is Intercept (constant), b is regression coefficient due to number of household and x_i is the number of households using electricity

The discussion below is based on the results obtained from the analysis as displayed in the respective tables.

3.1.1 Variables Used in the Study

The table below gives the name of the variables entered and those removed in the study.

Variables Entered/Removed ^a			
Model	Variables Entered	Variables Removed	Method
1	nohousehold ^b	.	Enter

a. Dependent Variable: kilowatts

b. All requested variables entered.

Table 3.0 Variables Entered

From the table above, it is clear that number of households was an independent variable while kilowatts are the Dependent Variable.

3.1.2 The Residual Statistic Summary

The table below shows a summary on the minimum, maximum, mean, standard deviation and the total number of the predicted value, Residual, Std. predicted Value and the standard residuals of the variables involved.

Residuals Statistics ^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2.9350	33.1833	13.9634	8.50043	70
Residual	-8.62733	6.73782	.00000	3.22764	70
Std. Predicted Value	-1.297	2.261	.000	1.000	70
Std. Residual	-2.654	2.072	.000	.993	70

a. Dependent Variable: kilowatts

Table 3.1: The Residual Statistic

From able 3.1 above, it is clear that we have used a sample size of 70 which is quiet enough to give a conclusion of the population.

3.1.3 Correlation Between Variables

The table below outlines the nature of the correlation between the dependent and independent variables.

Correlations		kilowatts	no household
Pearson Correlation	Kilowatts	1.000	.935
	no household	.935	1.000
Sig. (1-tailed)	Kilowatts	.	.000
	no household	.000	.
N	Kilowatts	70	70
	no household	70	70

Table 3.2 Correlation Between the associated Variables

Correlation describes the nature and strength of the relationship among different regression variables. Table 3.2 above shows that there exists a positive correlation between the two variables, i.e. dependent and independent. The correlation coefficient for the number of households is 0.935 which shows that there is a very strong positive correlation between the two variables which is interpreted statistically as; electricity consumption increases as the number of household's increase.

3.1.4 Plot of the Number of Households against Kilowatts

The figure below shows the scatter plot and its significance discussed below.

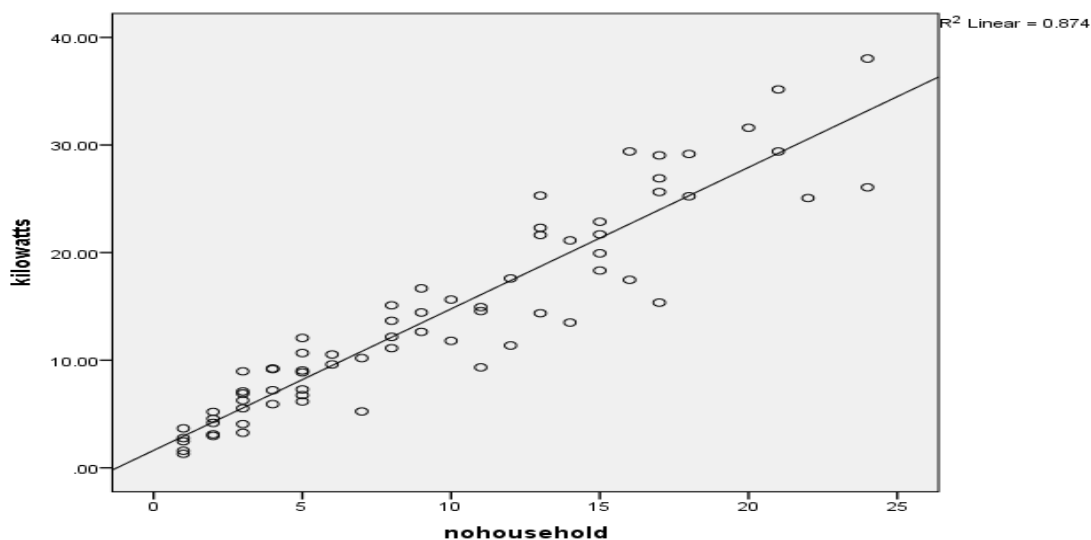


Fig 3.4: Plot of the Variables

Since no point is seen far away from the regression line, then the linearity assumption is satisfied.

3.1.5 Normal P-P Plot of regression Standardized Residual

The figure below shows the plot of the Observed Cumulative probabilities and the Expected Cumulative Probability and its statistical significance discussed below.

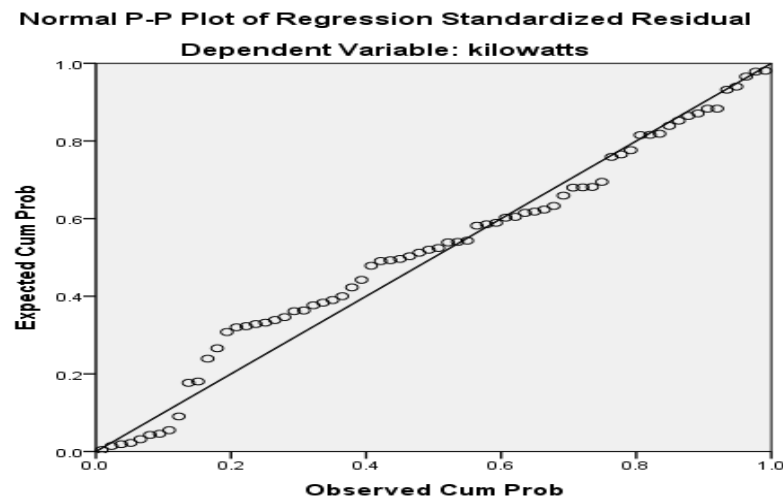


Figure 3.5 P-P Plot of regression Standardized Residual

Figure 3.5 above checks normality which shows that almost all the points lie at a realistic straight diagonal line (from left to top right) which suggest that there is no major deviation from normality.

3.1.6 Regression Coefficients

The table below shows the value of the regression coefficients and the results discussed below.

Coefficients ^a					
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	1.620	.689		2.353	.022
nohousehold	1.315	.061	.935	21.717	.000

a. Dependent Variable: kilowatts

Table 3.3.0: Regression Coefficients

Linear regression model was used and the independent variables have been entered into the equation accordingly. The coefficient of regression model is given in the table 3.3.0 above of coefficients that is in the column of unstandardized coefficient labeled B.

Coefficient for number of household (***b***) = **1.135**

Constant (***a***) = **1.1620**

Therefore, the regression model becomes

$$\hat{y} = 1.1620 + 1.135x_i$$

$$\hat{y} \text{ (Electricity consumed)} = 1.1620 + 1.135x_i \text{ (number of household)}$$

This equation is the most important in our study. It shall be used to predict or even calculate the power consumed in the households.

3.2.1 Histogram on Regression Standardized residual against the frequencies

The table below shows the Histogram on the Regression Standardized Residual against the Frequencies.

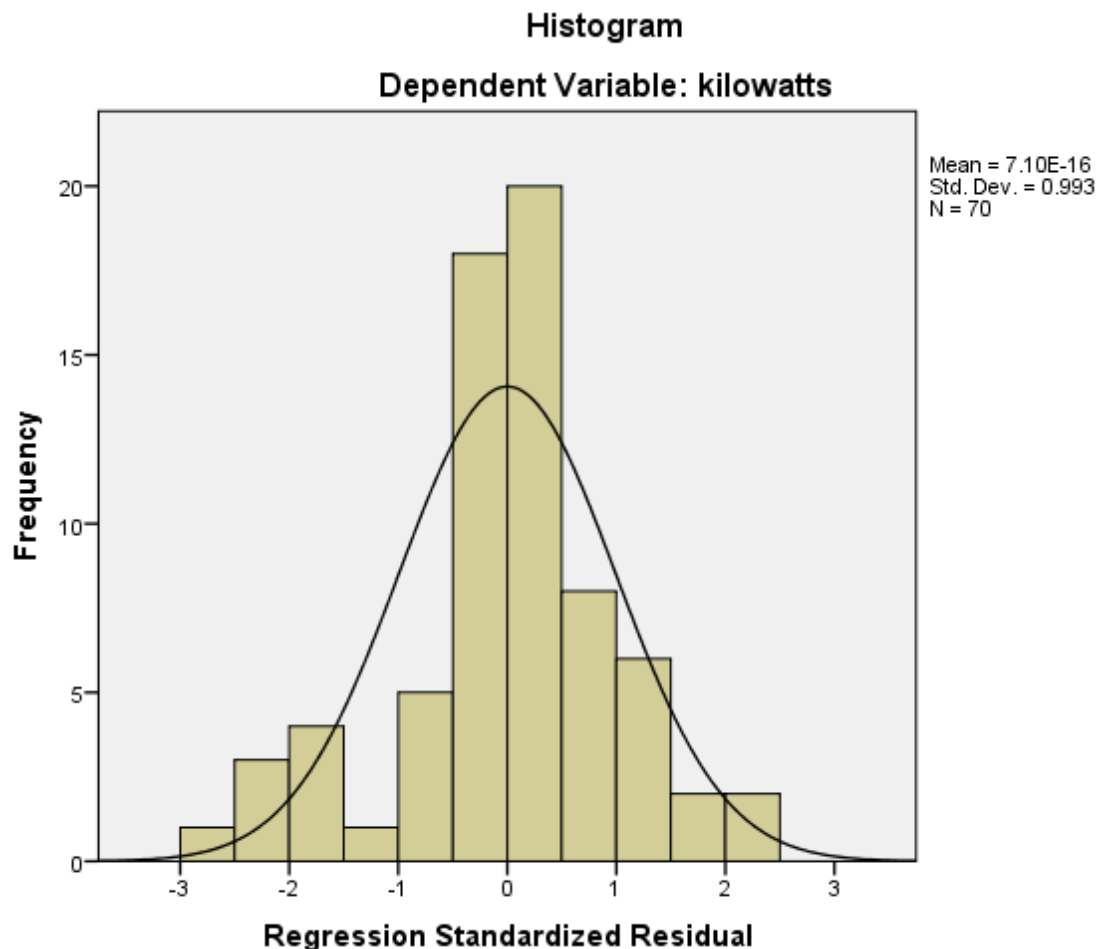


Fig 3.3.1 Histogram on the Standardized Residuals

The figure above displays the normal plot of the Regression Standardized Residuals. The bell shape of the curve indicates a Normal distribution of the data.

3.3.4 Coefficient of Determination (r^2)

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.935 ^a	.874	.872	3.25129

a. Predictors: (Constant), none household

b. Dependent Variable: kilowatts

Table 3.3.2: The Model Summary

In the model summary table, the values given under the heading R square tells how much of the variance in the dependent variable (electricity consumed in kilowatts per hour) is explained by the model (number of households). With this value equivalent to 0.874 *and* expressed as a percentage 87.4% shows that the number of households explains 87.4% of the electricity consumed in kilowatts per hour. Adjusted r square shows the correct square of the big population value for the population which is 0.872

3.2.2 Statistical significance

The table below gives the analyses of variance as per the results.

ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	4985.749	1	4985.749	471.650	.000 ^b
Residual	718.819	68	10.571		
Total	5704.568	69			

a. Dependent Variable: kilowatts

b. Predictors: (Constant), no household

Table 3.3.3: The ANOVA.

To assess the statistically significant, the result the researcher assessed the table labeled ANOVA. This tests the null hypothesis that multiple R in the population equals to zero. Hence the regression model reaches the statistical significance (*Sig* = 0 .000, this really since $0.00 < 0.05$).

3.2 CONCLUSION

The purpose of this study was to investigate the best prediction model for measuring the amount of electricity consumed using different number households within KU area. Data of the study was collected from 70 households. The Normality assumption is checked using normal probability plot which showed that almost all the points lie at a realistic straight diagonal line (from left to top right) which suggest that there is no major deviation from normality. Also, the Linearity assumption is checked where almost we have seen that all points lye on a straight line in a normal probabilistic plot. Also in the scatter plot the points are roughly rectangular distributed where almost all the points lie on the center line. There is no presence of outliers in this data. This is because in the scatter plot the points are seen almost at the same cluster.

It was proved that the best fit regression model was;

Electricity (estimated unit) = $1.1620 + 1.135 \times (\text{number of household})$

The equation above shall be used to predict or even calculate the power consumed. It is discovered that the electricity amounts consumed by KU students varied within different households. The regression model drawn from the data collected was fitted to predict the future energy consumption pattern. It was shown that number of households affected amount of power used. Regression model is hence the best model to analyze electricity consumption.

3.3 RECOMMENDATION

Students should adopt this model to predict the amount of electricity they consume. The university should also use this model to predict energy they consume. Bodies producing electricity should also adopt a method to help them identify amount of power their consumers need so that the demand doesn't outweigh the amount of power produced. Finally, further studies should be done on electricity consumption using other methods like Econometrics, Time series regression and others.

3.4 COMPETING INTEREST

The authors declare that there are no competing interests exist.

ACKNOWLEDGMENT

I thank all the authors for having made this research project successful. May God bless you all.

REFERENCES

- [1]. Government of Kenya, Ministry of Energy. *Study on Kenya's Energy Demand, Supply and Policy Strategy for Households, Small Scale Industries and Service Establishments*. 2002.
- [2]. Kenya Institute for Public Policy Research and Analysis (KIPPRA). *A comprehensive study and analysis on energy consumption patterns in Kenya*. 2010.
- [3]. KENYA POWER AND LIGHTING COMPANY LTD, (2017). ***Annual Report and Financial Statements***, Nairobi
- [4]. HOURI, A. and IBRAHIM-KORFALI, S. *Residential Energy Consumption Patterns: the case of Lebanon*. Lebanese American University, Beirut. 2004
- [5]. Zheng, Xinye, et al. "Characteristics of residential energy consumption in China: Findings from a household survey." *Energy Policy* 75 (2014): 126-135.
- [6]. Saab, Samer, Elie Badr, and George Nasr. "Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon." *Energy* 26.1 (2001): 1-14.
- [7]. Filippini, M., & Pachauri, S. (2004). Elasticities of electricity demand in urban Indian households. *Energy policy*, 32(3), 429-436.
- [8]. Zachariadis, Theodoros, and Nicoletta Pashourtidou. "An empirical analysis of electricity consumption in Cyprus." *Energy Economics* 29.2 (2007): 183-198.
- [9]. Bianco, Vincenzo, et al. "Analysis and forecasting of nonresidential electricity consumption in Romania." *Applied Energy* 87.11 (2010): 3584-3590.
- [10]. Al-Ghandoor, A., et al. "Electricity consumption and associated GHG emissions of the Jordanian industrial sector: empirical analysis and future projection." *Energy Policy* 36.1 (2008): 258-267.
- [11]. Payne, James E. "A survey of the electricity consumption-growth literature." *Applied energy* 87.3 (2010): 723-731.
- [12]. Yuan, Jiahai, et al. "Electricity consumption and economic growth in China: cointegration and co-feature analysis." *Energy Economics* 29.6 (2007): 1179-1191.
- [13]. Narayan, Paresh Kumar, Russell Smyth, and Arti Prasad. "Electricity consumption in G7 countries: A panel cointegration analysis of residential demand elasticities." *Energy policy* 35.9 (2007): 4485-4494.
- [14]. Kavousian, Amir, Ram Rajagopal, and Martin Fischer. "Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior." *Energy* 55 (2013): 184-194.

- [15]. Mugenda, D. M., and D. Mugenda. "Research Methods; Quantitative and Qualitative Research." (1999).
- [16]. Haque, M.H. (2005). Income elasticity and economic development Methods and applications. Springer, The Netherlands.
- [17]. Babatunde, A. and SHAUIBU, I. (2009) The demand for Residential electricity in Nigeria: A Bound Testing approach [online], available: www.africametrics.org/documents/conference09 [accessed 11/03/2010].
- [18]. Creswell, J.W. (2013). Research Design: Qualitative, Quantitative and Mixed Methods Approaches. (4th Ed.). Lincoln, NE: SAGE Publications, Inc.
- [19]. Mugenda, O.M., & Mugenda, A.G. (2003). Research Methods: Quantitative and Qualitative Approaches (2nd ed ed.). Nairobi: Acts Press.