Analysis of Household Electricity Consumption in Nonresident Rent Halls Using Linear Regression Analysis Model

Omondi S. Odhiambo¹, Samson W. Wanyonyi²*, Davis Mwenda Marangu¹, Irine Jemutai Nguli¹ and Mbuba Morris Mwiti¹

¹Department of Mathematics, Kibabii University, Box 1699-50200, Bungoma, Kenya. ²Department of Mathematics, University of Eldoret, Box 1125-30100 Eldoret, Kenya.

Authors’ contributions: This work was carried out in collaboration between all authors. OSO and SWW designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. DMM and IJN managed the analyses of the study. MMM managed the literature searches. All authors read and approved the final manuscript.

Article Information
DOI: 10.9734/AJPAS/2018/v2i328789

Original Research Article

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.

*Corresponding author: E-mail: samsonwangila@gmail.com;
Keywords: Power; ARIMA model; regression; root mean square error; correlation.

1 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 consumption

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 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 [17], 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 et al. [13], 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 [17] found that, contrary to Narayan et al. [13] 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 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 inform of a linear regression equation shown below;

$$y_i = a + bx_i + \varepsilon_i$$

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 $\varepsilon_i$ 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 $\sigma^2$) 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
\[ b = \frac{SS_{xy}}{SS_{xx}} \quad a = \bar{y} - b \bar{x} \]  

(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:

\[ SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \quad SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} \]  

(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:

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

(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 \((\alpha)\), 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

\[ b \pm t S_b \quad \text{where} \quad S_e = \sqrt{SS_{yy} - bSS_{xy}} \quad \text{and} \quad S_b = \frac{S_e}{\sqrt{SS_{xx}}} \quad \text{and} \ t \text{ is the tabulated value} \]  

(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.

*Fig. 1*  

*Fig. 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.

Haque [16] proposed some changes in data collection procedures for various versions while Babatunde and Shauibu [17] explained that population variable is significant in the short run and the long run. Due to resource and time constraint and so as to minimize errors, it was not possible to study the entire population. Mugenda and Mugenda [18] 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 [19].

\[ n = \frac{Z^2pqN}{e^2(N-1) + Z^2pq} \]  

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

2.7 Assumptions

This study is based on the assumption that in linear models the error terms \( \epsilon_i \) are:

i. Random and independent.
ii. Normally distributed with \( \text{E}(\epsilon_i) = 0 \) and \( \text{Var}(\epsilon_i) = \sigma^2 \) (homoscedasticity)
iii. Linearity

3. 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 1 gives the name of the variables entered and those removed in the study.

**Table 1. Variables entered**

<table>
<thead>
<tr>
<th>Variables Entered/Removeda</th>
<th>Model</th>
<th>Variables entered</th>
<th>Variables removed</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model</td>
<td>No household</td>
<td></td>
<td>Enter</td>
</tr>
</tbody>
</table>

a. Dependent variable: kilowatts  
b. All requested variables entered.

From the Table 1, 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 2 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.

**Table 2. The residual statistic**

<table>
<thead>
<tr>
<th>Residual Statisticsa</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted value</td>
<td>2.9350</td>
<td>33.1833</td>
<td>13.9634</td>
<td>8.50043</td>
<td>70</td>
</tr>
<tr>
<td>Residual</td>
<td>-8.62733</td>
<td>6.73782</td>
<td>.00000</td>
<td>3.22764</td>
<td>70</td>
</tr>
<tr>
<td>Std. predicted value</td>
<td>-1.297</td>
<td>2.261</td>
<td>.000</td>
<td>1.000</td>
<td>70</td>
</tr>
<tr>
<td>Std. residual</td>
<td>-2.654</td>
<td>2.072</td>
<td>.000</td>
<td>.993</td>
<td>70</td>
</tr>
</tbody>
</table>

a. Dependent variable: kilowatts

From Table 2, 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 3 outlines the nature of the correlation between the dependent and independent variables.

**Table 3. Correlation between the associated variables**

<table>
<thead>
<tr>
<th>Correlations</th>
<th>Kilowatts</th>
<th>No household</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>Kilowatts</td>
<td>.935</td>
</tr>
<tr>
<td></td>
<td>no household</td>
<td>1.000</td>
</tr>
<tr>
<td>Sig. (1-tailed)</td>
<td>Kilowatts</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>no household</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>Kilowatts</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>no household</td>
<td>70</td>
</tr>
</tbody>
</table>

Correlation describes the nature and strength of the relationship among different regression variables. Table 3 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 Fig. 3 shows the scatter plot and its significance discussed below.

![Plot of the variables](image)

Fig. 3. 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 Fig. 4 shows the plot of the Observed Cumulative probabilities and the Expected Cumulative Probability and its statistical significance discussed below.

![Normal P-P Plot of Regression Standardized Residual](image)

Fig. 4. P-P Plot of regression Standardized Residual

Fig. 4 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 4 shows the value of the regression coefficients and the results discussed below.

Table 4. Regression coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized coefficients</th>
<th>Standardized coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>1.620</td>
<td>.689</td>
<td>2.353</td>
<td>.022</td>
</tr>
<tr>
<td>No household</td>
<td>1.315</td>
<td>.061</td>
<td>.935</td>
<td>21.717</td>
</tr>
</tbody>
</table>

a. Dependent variable: kilowatts

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 4 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
\[
\hat{y} = 1.1620 + 1.135x_1
\]
\( \hat{y} \) (Electricity consumed) = 1.1620 + 1.135x (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.

3.1.7 Histogram on regression standardized residual against the frequencies

The Fig. 5 shows the Histogram on the Regression Standardized Residual against the Frequencies.

- Fig. 5. Histogram on the standardized residuals

The Fig. 5 displays the normal plot of the Regression Standardized Residuals. The bell shape of the curve indicates a Normal distribution of the data.
3.1.8 Coefficient of determination ($r^2$)

In the model summary Table 5, 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.

Table 5. The model summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R square</th>
<th>Adjusted R square</th>
<th>Std. error of the estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.935*</td>
<td>.874</td>
<td>.872</td>
<td>3.25129</td>
</tr>
</tbody>
</table>

*a. Predictors: (Constant), none household
b. Dependent Variable: kilowatts

3.1.9 Statistical significance

The Table 6 gives the analyses of variance as per the results.

Table 6. The ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4985.749</td>
<td>1</td>
<td>4985.749</td>
<td>471.650</td>
<td>.000*</td>
</tr>
<tr>
<td>Residual</td>
<td>718.819</td>
<td>68</td>
<td>10.571</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5704.568</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a. Dependent Variable: kilowatts
b. Predictors: (Constant), no household

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 ($\text{Sig} = 0.000$, this really since $0.00 < 0.05$).

4 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 × (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.

5 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.

Acknowledgement

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

Competing Interests

Authors have declared that no competing interests exist.

References


© 2018 Odhiambo et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar) http://www.sdiarticle3.com/review-history/45779