

Article – Smart Energy

Forecasting the Electricity Consumption in a Higher Education Institution

Luiz Amilton Peplow^{1*}

<https://orcid.org/0000-0001-8124-1778>

Roberto C. Betini^{1*}

<https://orcid.org/0000-0003-1817-6330>

Thulio C. G. Pereira^{2*}

<https://orcid.org/0000-0002-2570-2944>

¹ Universidade Tecnológica Federal do Paraná - Department of Electrical Engineering, Curitiba, Paraná, Brazil; ² Universidade Tecnológica Federal do Paraná, Department of Management and Economics, Curitiba, Paraná, Brazil.

Received: 2018.11.05; Accepted: 2019.07.26.

* Correspondence: luizpeplow@utfpr.edu.br Tel.: +55-41-99261-2981(L.A.P.)

* Correspondence: betini@utfpr.edu.br (R.C.B.)

* Correspondence: thuliopereira@utfpr.edu.br (T. C. G. P.)

HIGHLIGHTS

- This paper presents a mathematical representation by Regression Analysis that enables the projection of electricity consumption according to the built area and population in Higher Education Institutions (HEI).

Abstract: The objective of this paper is to present a mathematical representation by Regression Analysis that enables the projection of electricity consumption according to the built area and population in Higher Education Institutions (HEI) and to define an Indicator that contemplates the most significant variable in consumption. of electric power. The Null Hypothesis -H₀ is that in a HEI the most appropriate indicator is the Kilo Watt Hour per square meter (kWh / m²) as proposed by the Ministry of Planning and Management - MP. The research universe is 2,368 HEI, identified in a report from the Ministry of Education (2015). As Sample and case study, data from the thirteen Campi of the Federal Technological University of Paraná (UTFPR) are used. As a computational tool we use the IBM SPSS Statistics Base Software for Windows version 23 from SPSS Inc .. For the considered Sample and research design, the conclusion is that the null hypothesis is rejected accepting that the most significant indicator is the kilo. Watt Time per user (kWh / user). This conclusion does not exclude the relationship between constructed area and Energy Consumption, but reveals that it is not as significant as the number of individuals in HEI for this sample.

Keywords: Indicators of electricity consumption, regression analysis, decision making.

INTRODUCTION

This paper presents the initial research phase that aims to define Indicators based on socioeconomic, technical and sustainability factors to assist in the decision making of the feasibility of Energy Efficiency projects for Higher Education Institutions - HEI.

Decision-making processes are characterized by complexity, as they involve data with varying precision, the existence of many agents and criteria to be considered, different levels of preference, both particular and the group of individuals in which it is inserted, with tangible and intangible factors that they depend on the level of abstraction and interaction that it is capable of processing, as well as the confusion between various objectives that can be proposed to understand and solve the problem.

Combine this with subjectivity, valuation, different weights between different individuals for the same criteria, etc. Martins defines the decision as a process of analysis and chooses from several available alternatives of the course of action to follow [1].

Electricity consumers include - HEI. Decision-makers are defined as Rectors, Pro-Rectors and / or Directors of Higher Education Institutions with the task of dividing previously defined resources and destined to specific costing items, which may be multicampi or monocampi, in which case they consider the "Division of courses" as cost centers.

The document available in [2] presents the percentage share of the commercial classification Education, expressed by Region of the Country. In the National context, Education represents 2.6% and is in seventh place in Energy Consumption, very close to other commercial classifications. This fact reveals its significance in the global picture.

Based on data from the Higher Education Census 2014 of the Ministry of Education (MEC) of the National Institute for Educational Studies and Research Anísio Teixeira - INEP (2015) [3], Brazil has 2,368 - IES, which offer almost 33,000 courses of University graduate.

Data presented by the Union of Maintainers of Higher Education - SEMESPO show growth in the number of HEIs in the last 13 years being 108.2% in private and 71% in public HEIs.

In 2012 there were 5.9 million students enrolled in face-to-face courses, and the following year, this total reached 6.2 million enrollments, an increase of 3.8%. Regarding technological courses, there was a growth of 2.2%, with 1.2% in private and 6.2% in public.

The data presented allow us to establish a scenario of constant modifications that may or may not interfere with the level of Electricity Consumption. Each incoming teaching unit represents kWh and KW in terms of Electricity Consumption and Demand.

Electricity consumption, combined with energy efficiency in HEI should be evaluated considering mainly three approaches:

- under the Public Management approach in which the provider of education maintenance is the Government, which is responsible for managing the public budget with rigor and austerity;
- the private sector which, in addition to the public sector aspects, must also consider the shareholder profitability factor;
- of the electric sector that needs to provide the generation and the physical means, considering quality, safety, efficiency and continuity requirements for the supply of electric energy.

The HEIs, in addition to being mere consumers, can be philanthropic consumers that provide services to society by improving the living conditions of individuals, generate technology and innovation that change the economic growth of the country with higher tax revenues and better conditions for personal progression, etc.

For [4], the managers should use useful tools such as energy bills, consumption monitoring and economic indicators. The expression of the indicator is of great practical application as it makes it possible to inform the public when a particular energy efficiency

measure is feasible. Energy Indicators are instruments that reflect the extent to which energy is used.

In the related literature the Energy Efficiency and HEI Indicators may be found in [4] that classifies consumer units using consumption indicators based on historical data on electricity consumption, in [5] that reports the experience of the Federal University of São João Del Rei (UFSJ) and contributions from [6]; [7]; [8] and [9].

Through a preliminary evaluation of the available literature, relationships are identified as: Consumption as a function of population; Consumption as a function of useful area; Consumption as a function of covered area. In general, the consumption and demand for electricity depend solely on the type and profile of use of the loads.

According to [10] by Ministerial Ordinance No. 23 of 10/09/2015, indicators are established for the monitoring of electricity consumption in the Federal Public Administration (FPA). The criteria were instituted to guide public entities on good management practices by the Secretariat of Logistics and Information Technology (SLIT) and the Federal Budget Secretariat (FBS) in Joint Ordinance 8, published in the Official Diary of the Union.

The criteria are calculated from the consumption data, the built area and the number of servers reported by the agencies in the Sustainable Esplanade Project System (SisPES). The methodology observed is below:

- Description: Electricity consumption by built area.
- Formula: Consumption (kWh) / Constructed Area (m²).

Under the planning approach, it is important to evaluate how investments in physical expansions and course offerings may or may not lead to an increase in built-up area and population with a proportional increase in electricity consumption. This may affect both the infrastructure project and the overall calculation of future expenditure. It is necessary to predict the impacts of these items on both consumption and demand and financial increases.

In the context of HEI, the research problem is the need to subsidize the Decision Maker, with information to assess the impact on Electricity Consumption due to the increase in built area (m²) or the number of students, or even, more assertively organize their cost centers that do not have effective measurements of electricity consumption.

In order to structure the appropriate statistical model and measure the results, the research question is described: What mathematical model can be defined to express the relationship between Electricity Consumption with built area and population of a HEI?

Considering as Dependent Variable the Energy Consumption and Independent Variables the constructed area and the population, it is presented as complementary question: Among the Independent Variables, which one is the most significant to define an indicator to express its relationship with the Consumption of Electricity?

The objective of this work is: To define an indicator to express the Electricity Consumption that subsidizes the Manager regarding the projection of the electric energy consumption as a function of the built area and the population in a HEI.

As justification, it is understood that: having the mathematical representation and the Indicator, in the expansion analysis of a HEI it is possible to evaluate other conditions that may go unnoticed. The increase in load may lead to the review of energy contracting and demand from the Concessionaire, a fact that is subject to limitations on its supply. In addition, increasing consumption and demand implies changes in the physical structure of existing electrical installations such as cabling, protective devices, selectivity studies, design adequacy, etc.

MATERIAL AND METHODS

Due to the characteristics of the defined analysis, this research is non-probabilistic and uses an Accessibility Sample. Following guidance from the Public Ministry, hypothesis H₀ (Null Hypothesis) is described: In an HEI the most appropriate indicator to be

associated with Consumption Electricity is the Kilo Watt Hour per square meter (kWh / m²) disregarding the occupancy rate of a classroom. The factor that directly influences the consumption of electricity is the number of laboratories and classrooms occupied / used, with lamps and equipment in operation, where the number of students does not affect this type of consumption. With one or ten students the electricity consumed is the same.

The research universe is 2,368 HEI, previously presented. Census data describe only built-up areas, facilities, equipment and population without contemplating Electricity Consumption.

To determine the numerical representation that establishes the relationships between different numerical variables and that even allow sensitivity tests to be performed, the appropriate statistical methods are the Correlational ANALYSIS and the Regression Model. The multiple regression equation identifies the best fit line based on the least squares method, which expresses a measure of the effect that variable X has on Y. The representation of the multiple regression is given by Equation 1.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad 01$$

Y – Dependent variable to be predicted;

X_i – Independent variable

β₁ - Constant

β₂ – Partial Regression Coefficients (Model Parameters)

ε – Error or disturbance

To obtain real quantitative data of the considered variables, it is used as Sample and case study of this research, the data of thirteen Campi of the Federal Technological University of Paraná (UTFPR) available in the Management Report (MR) of UTFPR - 2015, as compiled from Table 01.

Table 01: Compiled data from UTFPR MR.

Campus	Energy consumption average 4 years (kWh)	Campus Area (m ²)		Total population
		Covered building area	Area according to use	
AP	402785	8765	4587	913
CM	662009	19432	11884	1903
CP	794325	35611	15275	2469
CT	2973136	73569	46223	9038
DV	724261	19722	18894	1411
FB	383891	9895	5724	735
GP	228477	10681	7226	730
LD	477212	12206	6317	1505
MD	919251	21120	22786	2107
PB	1191373	35745	33023	2396
PG	611467	28806	25753	2884
SH	105785	3828	2190	441
TD	355285	10019	4937	1254

Source: Elaborated by the authors with base in the MR.

The data used are numerical and come from measurements and real physical records, fact that allows the use of Linear Regression Analysis.

DEVELOPMENT

Definition of the variables and significant data for the statistical study

- Dependent variable: Electricity Consumption in kWh, considering the Average Consumption from 2012 to 2015.
- Independent variable 01 - Area in m².

The Management Report has three areas:

- (i) The overall area of each campus.
- (ii) The area according to its use (Theoretical Classrooms, Laboratories, Pedagogical Support, Library, Sports Activities, Dental Health Care and Educational Production Unit).
- (iii) The covered built area: Independent Variable 01 includes the places where there is lighting, computers, pumps, etc. and open places such as courtyards and corridors that have lighting or other cargo such as drinking fountains, etc.

- Independent variable 02 - current population - refers to the total number of students enrolled, technical-administrative staff and teachers.

With the data and variables defined, the next step is to develop the regression analysis to find an equation that correlates the independent variables to determine the predicted value of the dependent variable.

In terms of statistical research it is necessary to define the acceptable error for the analysis to be developed. The accuracy considered is of 5% of actual value with 95% confidence. Working with $p = 0.05$ is acceptable. We could even work with a higher value, up to 10%, that is, $p = 0.1$, since the numbers that are evaluated are in the order of thousands for both area and students, reflecting values on the thousands scale with exponent 10^6 , of kWh, as can be seen in Table 01.

The steps indicated in a Regression Analysis were followed. To develop the calculations that provide the results required for this analysis, the IBM SPSS Statistics Base for Windows version 23 software from SPSS Inc. was used.

Normality Analysis

The analysis is made in 3 steps.

1. Check the Normality by the Shapiro-Wilk test, given that there are only 13 observations for each Variable (<30 observations). The result can be seen in Table 05, where the significance for all variables is less than 0.05 or the 95% confidence level established in the research design.

2. Check Outliers existence by mustache chart analysis. There was only one outlier that was evaluated as resulting from the existence of a data very different from the others, the Curitiba Campus that presents much larger data than the other 12 Campuses. See data from Table 01.

3. Kurtosis and Asymmetry greater in modulus than 1.5.

With all of the above conditions met, Regression Analysis is performed. Table 02 presents the result of the Normality tests.

Table 02: Test of Normality.

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	gl	Sig.	Statistic	gl	Sig.
Energy consumption (kWh)	0.258	13	0.018	0.698	13	0.001
Campus Area (m²)	0.217	13	0.096	0.799	13	0.007
Campus user	0.291	13	0.040	0.640	13	0.000

a - Lilliefors significance correlation.

Source: Prepared by the authors via [2].

Determination of the regression equation

With the aid of the software, the following values are calculated: Estimated Standard Error equal to 180501.82; the correlation coefficient R equal to 0.974; coefficient R² equal to 0.949 which indicates that there is a strong correlation between the considered variables. Table 03 presents the value of the ANOVA.

Tabela 03: ANOVA.

ANOVA ^a					
Model	Sum of Squares	gl	Medium square	F	Sig.
Regression	6.041 e ¹²	2	3.020 e ¹²	92.706	0
Residue	3.258 e ¹¹	10	3.258 e ¹⁰		
Total	6.367e ¹²	12			

a – Dependent variable: Energy consumption kWh.

b – Predictors: (Constant) Campus users, Campus Area (m²)

Source: Elaborated by the authors via reference [2].

The analysis of the ANOVA value with significance tending to zero indicates that for this sample there is a relationship between the variables. As positive, the Regression curve with Linear relation is ascending. Table 04 shows the values of the coefficients of the multiple regression analysis of the regression equation according to Equation 01.

Table 04: ANOVA Results.

Model	Coefficients				t	Sig.
	non-standardized coefficients		standardized coefficients			
	B	Standard error	Beta			
(Constant)	3590.442	86185.053			0.042	0.968
Campus Area (m²)	13.421	9.472	0.341		1.417	0.187
Campus user	212.28	79.442	0.643		2.672	0.023

Dependent variable: Energy consumption average.

Source: Elaborated by the authors via reference [2].

With the determination of the coefficients, the multiple regression equation for the considered sample is represented by Equation 02 as indicated by Equation 01:

$$y = 3,590.442 + 13.421 * x_1 + 212.280 * x_2 + \varepsilon$$

2

Y = Energy Consumption Value.

β_0 = constant = 3,590.442

β_1 = Coefficient of variable X1 = 13.421

β_2 = Coefficient of variable X2 = 212.280

ε = Error = 86.185

Sensitivity Analysis

If the value corresponding to the area according to use is considered as the effective area, a fact that decreases the area considered then it is questioned: Is there a significant change in the regression analysis coefficients to support the research hypothesis?

The data were entered into the IBM SPSS Statistics Base Software. Table 5 shows the coefficient values of the new regression analysis for area data according to use.

Another question: What will be the limit of the relationship with the built up area in which it may be feasible to consider the relationship between Energy Consumption and built up area?

Table 05: The Area regression coefficients according to use.

Model	Coefficients ^a			t	Sig.
	non-standardized coefficients		standardized coefficients		
	B	Standard error	Beta		
(Constant)	8526.904	75785.425		0.113	0.913
Campus Area (m²)	13.274	7.208	0.240	1.841	0.095
Campus user	251.916	43.097	0.763	5.845	0.000

a – Dependent variable: Energy consumption average.

Source: Elaborated by the authors via reference [2].

The results for the new regression tests considering the new areas are shown in Tables 06 and 07. These limits set quite different extremes in order to assess the influence of area variation.

Table 06 presents the first test that considers an area thirty percent smaller than the area according to use.

Table 06: Regression coefficients, first test.

Model	Coefficients ^a			t	Sig.
	non-standardized coefficients		standardized coefficients		
	B	Standard error	Beta		
(Constant)	8523.535	75,785.807	0.000	0.112	0.913
Campus Area (m²)	18.963	10.297	0.240	1.841	0.095
Campus user	251.915	43.097	0.763	5.845	0.000

a – Dependent variable: Energy consumption average.

Source: Elaborated by the authors via reference [2].

In Table 07 the second test is fifty percent larger than the Covered Built Area.

Table 07: Regression coefficients, second test

Model	Coefficients ^a		t	Sig.
	non-standardized coefficients	standardized coefficients		
	B	Standard error	Beta	
(Constant)	3,588.440	86,185.520	0.000	0.042 0.968
Campus Area (m²)	8.948	6.315	0.341	1.417 0.187
Campus user	212.280	79.442	0.643	2.672 0.023

a – Dependent variable: Energy consumption average.

Source: Elaborated by the authors via reference [2].

Based on the data presented from the regression analysis, it is possible to establish the analysis of results described in the following item.

RESULTS

In the first test performed, the correlation coefficient sign indicates the direction between Y and X. Since it is positive, the regression curve is upward and appropriate.

The software provides a graph of the Correlation Matrix which demonstrates that all variables have good correlation, with data close to the configuration of a line joining the points. There is small scatter with a scattered variable that corresponds to an Outlier.

The value of R equal to 0.974 and R² equal to 0.949 indicate that there is a strong correlation between the considered variables.

By analyzing the ANOVA value provided by the software, with significance tending to zero, it is possible to indicate that for this data set there is a relationship between the variables.

From the significance value of each independent variable, the following interpretations of the correlation with the independent variable Energy Consumption can be indicated:

- For the coefficient value of the independent variable number of users whose significance is 0.023, (less than 0.05), indicates a strong correlation.
- For the coefficient value of the independent variable built-up area whose significance is 0.187, (greater than 0.05), indicates that there is poor correlation.

In the analysis of the results considering as effective area the value corresponding to the area according to the use, from the results from the software calculations, the new interpretations of the correlation with the independent variable Energy Consumption are observed:

- The significance value for the number of users changed from 0.023 to 0.000, a fact that demonstrates an improvement in the significance level establishing a more significant relationship.
- The significance value for the built-up area changed from 0.187 to 0.095, which demonstrates an improvement in the Significance level, but even greater than 0.05.
- The values of the R Correlation coefficients, the Determination coefficient R², and the Standard Estimate Error express a better Regression curve for the new data set.

The value of the Regression equation constant changed most significantly from 3,590.442 to 8,526.904, more than double. The values of the coefficients of the independent variables changed little, being β_1 from 13.421 to 13.274 and β_2 from 212.280 to 251.916.

These values demonstrate that the reduction of the areas with the maintenance of the number of students establishes, as expected, a new relationship between the variables. The value of the constant in the determination of energy consumption increased significantly, indicating a higher consumption per individual, with a smaller increase in the ratio of energy consumption with the other variables.

Even with this analysis it is necessary to ascertain whether there is really any significance in the value of the area with respect to Energy Consumption. The Tables 06 and 07 show the results for the new regression tests for the new areas.

It is observed that for a smaller area value, considering a higher occupancy rate, the area coefficient increased to 18.963 with 0.095 significance and Correlation coefficients $R = 0.977$ and for a larger value considering a lower occupancy rate, the area coefficient increased. to 8.948 with significance 0.187 and correlation coefficients $R = 0.974$.

Once the analysis of the research results is established, following the text the main conclusions are described.

CONCLUSION

For the Database and boundary conditions presented for this study and from the results obtained it can be concluded that:

By linear regression analysis, it is possible to indicate with high degree of certainty that there is a correlation between the dependent variable Energy Consumption and the independent variables: built area and circulating population in a HEI.

From this relationship was defined the mathematical representation by the multivariate regression equation indicated in Equation 02 which expresses the relationship between the variables and their contributions on the Electricity Consumption.

Regarding the research hypothesis that suggested as the best condition to propose an Energy Consumption indicator the relationship between energy consumption and the built area:

- To answer the hypothesis, it is important to verify the results obtained from the sensitivity test that aims to verify if there is a significant change in the regression analysis coefficients when considering as the effective area the value corresponding to the area according to use, which value is much lower than previous one.

- Based on the results of the analyze, it was evidenced by the statistical tests of the database used in the research, that for HEI UTFPR this hypothesis is invalid.

From the analyzes performed, it can be concluded that the most significant indicator is the Kilo Watt Hour per user (kWh / user). This evidence is evidenced by the results of the Regression coefficients shown in Table 04, in which the only variable that presented Significance less than 0.05 was the number of users.

From this evidence, the complementary objective of Identifying an Indicator that contemplates the most significant Variable in Electricity Consumption to predict the increase of Electricity Consumption of an HEI is achieved.

This conclusion does not exclude the relationship between constructed area and Energy Consumption, but reveals that this is not as significant as the number of individuals in the HEI for the data set considered.

With the built-area sensitivity tests, it can be seen that even with variations more or less, the greatest significance is in the number of students, increasing as the area increases.

The increase in the significance of the built-up area warns of the need to increase the sample size and seek other types of universities with different areas and types and uses of loads, as the main feature of UTFPR is its courses in Technology that require larger areas with highest consumer equipment.

This type of research is the suggestion for future work, as the guidance established by the Ministry of Planning, Budget and Management indicates as an indicator kWh / m². If it is established that this Indicator is misleading as evidenced in this research, it will be motivation to inform the Ministry of the need to revise its guidance.

Acknowledgments: The authors thank UTFPR (Federal University of Technology - Paraná) for the infrastructure offered for the development of this research, and COPEL-DIS for the support and funding of the resources for this R&D project.

REFERENCES

1. MARTINS, G. A., Theóphilo, C. R..Metodologia da investigação científica para ciências sociais aplicadas, 2a.ed., São Paulo./Ed. Atlas. (2009)
2. Balanço Energético Nacional 2017: Ano base 2016, Empresa de Pesquisa Energética. – EPE, Rio de Janeiro: Available from: https://ben.epe.gov.br/downloads/Relatorio_Final_BEN_2015.pdf. Access in setembro/2018
3. Censo do INEP (2015) do Instituto Nacional de Estudos e Pesquisa Educacionais Anísio Teixeira – INEP. 2016. Available from <http://portal.inep.gov.br/basica-censo>, Access in set/2018
4. MORALES, Clayton. Indicadores de Consumo de Energia Elétrica: Classificação por Prioridades de Atuação na Universidade de São Paulo. São Paulo, 2007.
5. NEPOMUCENO, Erivelton. Eficiência Energética em Instituições de Ensino.São João Del Rei, 2005.Available from <https://www.iar.unicamp.br/lab/luz/Id/Arquitetural>. Access in set/2018.
6. PEPLOW, Luiz; BETINI, Roberto; GOBER, Cristiano. Avaliação do impacto no sistema energético a partir de projetos de eficiência energética em instituições de ensino tecnológico e superior. VIII Congresso Brasileiro de Planejamento Energético Gramado, 2016.
7. SAID EL, M. A., FAVAT, L. B., MORALES, C. Indicadores Energéticos e Ambientais: Ferramenta Importante na Gestão da Energia Elétrica, 2005. Anais do I CBEE-GEPEA, USP, 2005.
8. TAVARES, Flávio; MONTEIRO, Luciane. Indicadores de Eficiência Energética na Indústria de Fertilizantes de Amônia. 2011. *Sistemas & Gestão*, vol. 9, 2014. 216 – 223 p.
9. CUSTÓDIO, M. G. G. Eficiência energética em edifícios escolares.,106f. Dissertação apresentada à Universidade Técnica de Lisboa, como parte dos requisitos para a obtenção do Título de Mestre em Engenharia Mecânica, Lisboa, 2011.
10. Ministério do Planejamento, Orçamento e Gestão (MP) Portaria conjunta nº 8 – Portaria Ministerial nº 23. 2016. Available from http://www.planejamento.gov.br/secretarias/upload/Legislacao/Portarias/2015/150213_port_23.pdf Access in set/2018.



© 2018 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).