Learning through planning experiments

A case study in thermal comfort in hospitals

[2]Teixeira, I.M., [1]Graçoeiro, D.G., [2,3]Rodrigues, N.J., [2]Teixeira, S.F., [2]Rodrigues, C.S.
[1]Departamento de Produção e Sistemas;
[2] ALGORITMI;
[3] Metrics
Universidade do Minho
Guimarães, Portugal
ines.teixeira@dps.uminho.pt; a91747@alunos.uminho.pt; nelson.rodrigues@dps.uminho.pt; st@dps.uminho.pt; crodrigues@dps.uminho.pt

Abstract — **Learning engineering involves observing and understanding how systems or processes work, but also knowing the cause and effect relationships present. During the course, students learn that the process of engineering investigation requires conducting experiments to validate theories or hypotheses about what makes the system work. Each experimental run is a test. It is through active experimentation, the controlled intervention of a process that modifies the inputs and characteristics of the process, that it is possible to learn the effect of a set of different factors on a response variable. Based on experimental data collected to model thermal comfort in hospital waiting rooms, this paper aims to contribute to the knowledge of this phenomenon through the selection of the most appropriate statistical techniques to analyze the data. For engineering students this is an opportunity to bring theory to practice.**

Keywords - Experimental Planning; Experimental Methods; Thermal Comfort, Learning Statistics.

I. Introduction

Experimentation plays an important role in technology commercialization and product realization activities, which consist of new product design and formulation, manufacturing process development, and process improvement. The objective in many cases may be to develop a robust process, that is, a process affected minimally by external sources of variability. The purpose of experimental design is to maximize, for a given amount of resources, the chance that information-laden data will be generated and structured in such a way as to be conducive to extracting and communicating that information [1]. There are also many applications of designed experiments in a nonmanufacturing or non-product-development setting, such as marketing, service operations, general business operations, and research [2]. The planning of experiments is an important pillar in engineering and research as it allows studying in detail the effect of a set of different factors on a variable response of a process or system. According to Montgomery [2], a process or system has controllable or uncontrollable variables, and the objective of experimentation is one of the following:

1. Determining which variables are most influential on the response y;
2. Determining where to set the influential x's so that y is almost always near the desired nominal value;
3. Determining where to set the influential x's so that variability in y is small;
4. Determining where to set the influential x's so that the effects of the uncontrollable variables z1, z2, . . ., zq are minimized.

Statistics can be used to analyze the data and allow for objective practical results and conclusions, as well as suggestion of a course of action. Nevertheless, for an engineering student, the biggest challenge is to overcome the barriers of lack of knowledge in statistical techniques and the lack of specific training in their applicability [3]. According to these authors, to these educational barriers it can be added the negative image of statistics that invoke fear and resistance to its use. To overcome these barriers, it is increasingly recognized that learning integrated with a practical problem is a more effective pedagogical paradigm in engineering [4]. Funny et al [5] acknowledge the need for reflective thinking in the learning of statistics by engineering students. In turn, Acosta [6] refers to the need to introduce inductive thinking to improve the learning of statistics by engineering students.

Following several authors' suggestion of integration of statistical theory and practice with in-depth case studies, real-world situations or multidisciplinary activities [7]–[9] this paper presents a multidisciplinary study involving experimental engineering and statistical analysis.

In this sense, an academic study was carried out using an experimental procedure, which consisted in recording measurements of velocity, air temperature, humidity, and radiant temperature of a waiting room of a hospital in Portugal. These experimental data were used to give inputs to the numerical study of the room's thermal comfort. Then, the numerical results can be compared with those obtained experimentally using spreadsheets.

To determine thermal comfort it uses An index, Predicted Mean Vote (PMV), which is an index that predicts the average thermal sensation vote of a large group of people on a 7-point

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

scale, based on the thermal balance of the human body [10], [11].

The main goal of this study is to analyze the data collection of the study and verify how important or relevant the variables under study are for the determination of the thermal comfort study. In addition to analyzing the relevance of the variables, the study aims to provide recommendations to a future record of measurements to obtain reliable, valid results and, overall, the most accurate with reality.

II. Experimental Procedure

Case study

This study was used to validate, perform and recommend a future way to do data collection measurements of temperature, velocity, and relative humidity of the waiting room understudy, as can be seen in fig. 1.

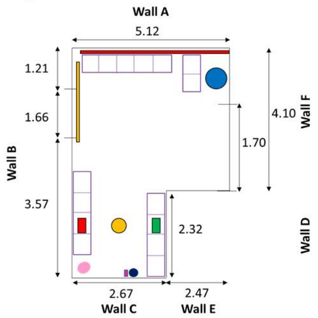

Figure 1. Schematic of the waiting room configuration to be modeled (dimensions are in meters)

The room has a headroom of 3 meters, one main door with 1.70 meters long and 2.10 meters high.

As for the number of chairs, this room contains 15 seats with 50 cm long and 65 cm in height. The room is equipped with some equipment, namely, television (hanging on wall C), fire extinguisher (hanging on wall E), side table, wastepaper basket and water dispenser. As for lighting, the room has 4 lamps, three of them rectangular, near the ceiling of wall B, and a circular one in the center of the ceiling of walls B, C, and D, as shown in yellow in fig. 1.

For the measurement study, three measuring devices were used, illustrated in fig. 2, namely, a USB data logger, a weather station, and a thermo-anemometer.

The USB data logger was used to collect relative humidity data, the weather station was used to determine the air velocity and temperature, as well as the radiant temperature, and, finally, the anemometer allows to measure velocity accurately, reliably, and quickly[10]–[12].

a) b) c)

Figure 2. Measurement Equipments, a) USB data logger, b) Equipment monitor Bruel and Kjasr 1213 and c) VelociCheck Thermo Anemometer - Model 8830

III. Sample Characterization

The present study was conducted using a random sample composed of twelve different positions referring to variables likely to affect the thermal comfort of a given patient in the waiting room of a hospital. Fig. 3 shows the point of measurements of velocity, humidity, air temperature, and radiant temperature that were taken in occupancy zones at different heights (0.1, 0.95, and 1.20 meters) coinciding with the height of the feet, chest, and head, of a seated person.

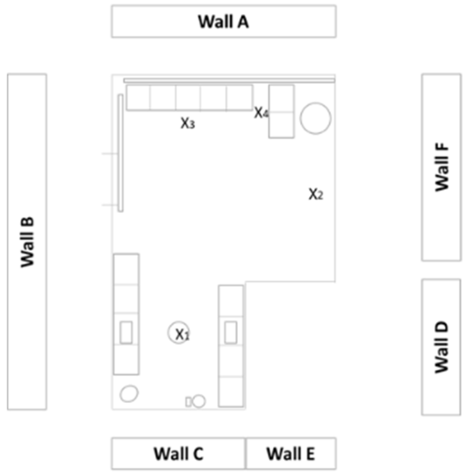

Figure 3. Point of measurement of humidity and air velocity, air temperature and radiant temperature at 0.1m, 0.95m, and 1.20m [11]

Table I compiles the obtained values of air temperature, air velocity, the mean radiant temperature, relative humidity, and PMV for the 4 study points, corresponding occupation zones, for the height 0.10 m, 0.95 m, and 1.20 m considered. These four points (X1 to X4) corresponding to occupation zones.

TABLE I. Registration Of The Experimental results

Position	v_{air} (m/s)	T_{air} (°C)	$\overline{T_r}$ (°C)	Hr (%)	PMV
X1 (0.1m)	0.06	23.3	23.7	61	-0.59
X1 (0.95m)	0.04	23.9	24.2	58.5	-0.5
X1 (1.20m)	0.05	23.9	24.3	58.0	-0.5
X2 (0.1m)	0.24	23.5	23.6	61.5	-1.4
X2 (0.95m)	0.33	23.9	23.8	58.8	-1.4
X2 (1.20m)	0.17	23.9	23.9	58.5	-1.3
X3 (0.1m)	0.20	23.7	24.1	60.5	-0.5
X3 (0.95m)	0.09	23.7	24.1	59.0	-0.5
X3 (1.20m)	0.17	23.8	23.9	58.5	-0.5
X4 (0.1m)	0.04	23.7	23.7	60.5	-0.5
X4 (0.95m)	0.09	23.7	24.1	59.5	-0.5
X4 (1.20m)	0.04	24.0	24.0	60.0	-0.5

$\overline{T_r}$ – Mean radiant temperature
Hr – Relative humidity

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

IV. Hypothesis and Statistic Method of Analysis

Hypothesis definition

In order to explore the existence of differences in the position of the room, the analysis formulated the following hypotheses:

- *Position in the room and Air temperature:*

$\boldsymbol{H1}$: There are significant differences in air temperature between the four different positions in the waiting room

- *Position in the room and Mean radiant temperature:*

$\boldsymbol{H2}$: There are significant differences in average radiant temperature between the four different positions in the waiting room.

- *Position in the room and Air velocity:*

$\boldsymbol{H3}$: There are significant differences in air velocity between the four different positions in the waiting room.

- *Position in the room and Humidity:*

$\boldsymbol{H4}$: There are significant differences in humidity between the four different positions in the waiting room.

- *Position in room and PMV:*

$\boldsymbol{H5}$: There are significant differences in PMV between the four different positions in the waiting room..

To explore the existence of differences in the height position measured, the analysis formulated the following hypotheses:

- *Height position and Air temperature:*

$\boldsymbol{H6}$: There are significant differences in air temperature between the three different height positions measured in the waiting room.

- *Height position and Mean radiant temperature:*

$\boldsymbol{H7}$: There are significant differences in mean radiant temperature between the three different height positions measured in the waiting room.

- *Height position and Air velocity:*

$\boldsymbol{H8}$: There are significant differences in air velocity between the three different height positions measured in the waiting room.

- *Height position and Humidity:*

$\boldsymbol{H9}$: There are significant differences in humidity between the three different height positions measured in the waiting room.

- *Height position and PMV:*

$\boldsymbol{H10}$: There are significant differences in PMV between the three different height positions measured in the waiting room.

In situations where the normality assumption is unjustified, the experimenter may use alternatives statistics procedures [2] such as the techniques called nonparametric methods. or distribution-free methods. As Kvam and Vidakovic [13] argument: "analysts limited to basic statistical methods can be trapped into making parametric assumptions about the data that are not apparent in the experiment or the data. In the case where the experimenter is not sure about the underlying distribution of the data, statistical techniques are needed which can be applied regardless of the true distribution of the data" (page 2).

Therefore, no tests for normality were performed due to the sample size, so only non-parametric tests were used, i.e., the assumption that the data follow a normal distribution are not required, only the assumption that the distribution of the underlying population is continuous.

Mann-Whitney test

The Mann-Whitney U test is the most popular of the two-independent-samples tests and it uses the rank of each case to test whether the groups are drawn from the same population. The first value in ascending order of the grouping field defines the first group and the second defines the second group. If the grouping field has more than two values, this test is not produced [14].

As for the Mann-Whitney test, it compares two independent samples and analyzes whether the distributions have identical means, the hypotheses being:

$H0$: The two-population means are identical. ($\mu1=\mu2$)

$H1$: The two means of the populations are not identical. ($\mu1\neq\mu2$)

Kruskal-Wallis test

The Kruskal-Wallis H test, an extension of the Mann-Whitney U test, is the nonparametric analog of one-way analysis of variance and detects differences in distribution location. The median test, which is a more general test, detects distributional differences in location and shape. The Kruskal-Wallis H test and the median test assume that there is no a priori ordering of the k populations from which the samples are drawn [14].

Regarding the Kruskal-Wallis test, this compares three or more independent samples (k) and analyzes if the distributions have identical means, and the hypotheses are:

$H0$: There are no significant differences between treatments. ($\mu1=\mu2=\cdots=\mu k$)

$H1$: Not all k distributions have identical means.

To perform the above non-parametric tests, with the help of IBM® SPSS®. This software platform offers advanced statistical analysis, a vast library of machine learning algorithms, text analysis, open-source extensibility, integration with big data, and seamless deployment into applications. Its ease of use, flexibility, and scalability make SPSS accessible to users of all skill levels. What is more, it is suitable for projects of all sizes and levels of complexity and can help to find new opportunities, improve efficiency and minimize risk [14], [15].

Fig. 4 shows all commands used to do the tests and the order that should be taken into account.

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

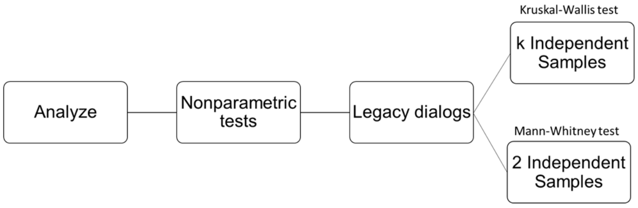

Figure 4. Menus choose to obtain tests for several independent samples

Regarding linear regression, this one obtains a model for the functional relationship of a variable Y (dependent variable) given the values of some explanatory variables X (independent variables). With this procedure the following equation is obtained:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta iXi + \varepsilon$$

In this, Y corresponds to the dependent variable, Xi to the i independent variables and ε to the error or deviation of the real value of Y about the straight line and can be obtained using the command sequence of fig. 5.

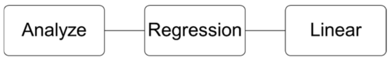

Figure 5. Menus choose linear regression

The Kruskal-Wallis test is used for more than two samples, this was selected for work purposes. The independent variables are "Air_temperature", "Radiant_temperature", "Air_velocity", "Humidity" and "PMV" divided by the group variable "Position" to test how significant each factor in the different room positions could be patient thermal comfort.

Multiple Linear Regression

A multiple linear regression was performed to find out the model that allows us to explain the individual value of PMV (dependent variable), based on all the other variables (independent variables). The goal with this model is to figure out which variables best explain the variable "PMV" and to evaluate if the estimating parameters are statistically significant, defining the hypotheses:

$$H_0: \beta i = 0$$
$$H_1: \beta i \neq 0$$

For multiple linear regression, a method was selected to enter the independent variable into SPSS. In this context, the most appropriate is the Stepwise Method which uses staged analysis to control variable entry and removal [14]. With this method, the program adds the independent variables incrementally, testing their statistical significance. The variables are sorted in descending order of significance, and any variable can be deleted if it is not considered significant.

V. Main Results

The results of Kruskal-Wallis tests, for the differences concerning position in the room are summarized in table II.

TABLE II. KRUSKAL-WALLIS TEST FOR POSITION IN THE WAITING ROOM

Hypothesis test for position dependency		
	Sig.	*Decision*
Air temperature (H1)	0.056	Validated ✓
Mean radiant temperature (H2)	0.344	Not validated ✗
Air velocity (H3)	0.025	Validated ✓
Humidity (H4)	0.750	Not validated ✗
PMV (H5)	0.025	Validated ✓

✓ Reject null hypothesis
✗ Not reject null hypothesis

Through table II it is possible to conclude the four positions in the waiting room have statistical differences concerning:

- air temperature (statistically significant for a significance level of 10%),
- air velocity (statistically significant for a significance level of 5%) and
- and PMV (statistically significant for a significance level of 5%).

On the other hand, both mean radiant temperature and humidity no longer vary significantly with the change of position in the room.

For the variables flagged with differences between the 4 positions, boxplots were analyzed to visually identify the differences.

Regarding fig. 6, air temperature by room positions, the temperature at position “2” is lower than the temperature at “3” and “4”, which in turn apparent to be equal. The Mann-Whitney tests allowed proving the existence of significant differences between position “2” and positions “3” and “4”. The air temperature at position “1” has no significant differences relative to position “2”, nor does it have significant differences relative to positions “3” and “4”. These results end up showing that the fact that there is a door near position 2 influences its temperature, and it is also possible to conclude that the outside temperature was lower than in the waiting room. On the other hand, the test results also end up meeting the other positions, since there is an air intake and outlet near position “1” and an air outlet near positions “3” and “4”.

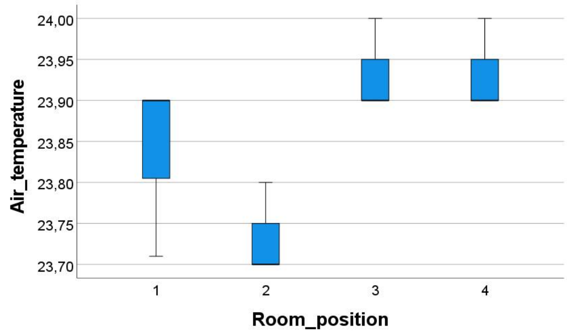

Figure 6. Boxplot of air temperature by room positions

Regarding fig. 7, air velocity by room positions, the air velocity at position “1” is lower than the velocity at position “3”, which in turn is lower than the velocity at position “2”. The

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

Mann-Whitney tests confirmed the existence of significant differences in position “2” and other positions. The air velocity at position “4” is between positions “1” and “3”, since there are no significant differences between these positions for a significance level of 10%. Following on from this, the data again points to the fact there are a door through which cooler air enters with some flow, as the velocity at position 2 is significantly higher than at the other positions.

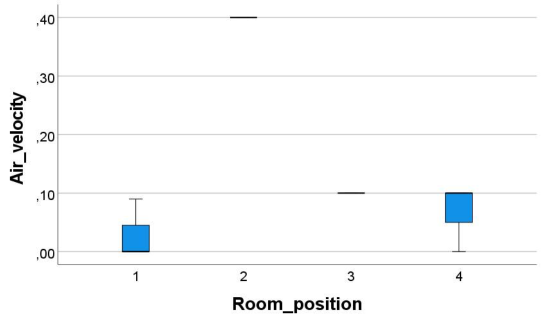

Figure 7. Boxplot of air velocity by room positions

Regarding fig. 8, PMV by room positions, the PMV in position “2” is lower than the PMV in positions “1”, “3”, and “4”. The Mann-Whitney tests confirmed the existence of significant differences in position “2” when compared with other positions. The fact that the air velocity is much higher and the air temperature is lower ends up making this position not so satisfactory for the user of the waiting room.

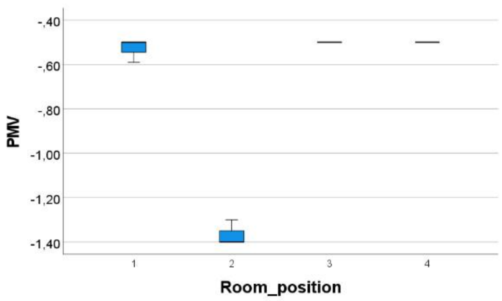

Figure 8. Boxplot of PMV by room positions

When performing the Kruskal-Wallis tests for height position (0,1 m, 0.95 m and 1.2 m), the test highlight differences only on variable “humidity”, with a significance level of 5%. Therefore, no significant differences were found in all other variables. The results are present in table III.

TABLE III. KRUSKAL-WALLIS TEST FOR HEIGHT DEPENDENCY

Hypothesis test for height dependency		
	Sig.	*Decision*
Air temperature (H6)	0.915	Not validated ✗
Mean radiant temperature (H7)	0.162	Not validated ✗
Air velocity (H8)	0.938	Not validated ✗
Humidity (H9)	0.02	Validated ✓
PMV (H10)	0.766	Not validated ✗

✓ Reject null hypothesis
✗ Not reject null hypothesis

Following the Mann-Whitey tests to compare the humidity between two height positions, results prove the existence of significant differences between the height position “0.1 meter” and the other two height positions. Therefore, the analysis concluded that the humidity near the surface is higher since the humidity for 0.1 meters was significantly higher than at the other heights.

In the logic of the exploratory approach used in this study, it was also decided to perform a linear regression in order to explain the behavior of thermal comfort (PMV) in the waiting room and which factors contribute to it.

Due the characteristics of positions “3” and “4” in room stand out in several variables differences, it was decided to create a dummy variable to include this information in the regression:

- Value 1 - Room positions “3 and 4”
- Value 0 – Room positions “1” and “2”

Since nominal variables cannot enter linear regressions, a dummy variable is an artificial variable created to represent an attribute with two or more distinct categories/levels [16].

The regression model considered the following variables:

Dependent variable – PMV
Independent variables:

- Room positions “3 and 4”
- Air temperature
- Mean radiant temperature
- Air velocity
- Humidity

Table IV summarizes the linear regression results using a stepwise regression procedure to fit the model. In stepwise regression, variables entered into the model one at a time, beginning with those that appear most important, until no variables remain that are reasonable candidates for entry [2].

TABLE IV. LINEAR REGRESSION EQUATIONS AND ITS COEFFICIENT OF DETERMINATION

Model	Variables	B	Sig	R^2
1	Constant Air velocity	-0.368 -2.386	0.000 0.000	0.92
2	Constant Air velocity Positions “3 and 4”	-0.487 -2.144 0.166	0.000 0.000 0.009	0.96

With the regression, the coefficient of determination (R2) indicates the proportion of the variance of the dependent variable that is explained by the regression model. The higher the R2 values, the more explanatory the model is. Thus, model 1 resulted in a R2 of 0.92 and model 2 in a R2 of 0.96, i.e. both models with a strong explanatory capacity.

The final model (model 2) to explain PMV is composed of the air velocity and the "Position 3_4". All the other variables considered in the experiment end up not being explanatory enough.

Therefore, it can be concluded that air velocity is the variable that most affects patient comfort, followed by the decision of the patient to sit in positions 3 and 4. In this model, the increase of one unit in air velocity (keeping the position constant) decreases thermal comfort in 2.144 values; when the room position is "Position_3_4" (keeping the air velocity constant) the thermal comfort increases 0.166 values.

VI. CONCLUSIONS

In summary, the results suggest that in the waiting room understudy there may be differences in the position in the room and the height position. Regarding the position in the room, there are differences in air temperature, air velocity, and PMV. Regarding the height position, there are differences in humidity.

Afterward, when testing an explanatory model of thermal comfort, the results suggest that in the registered situation thermal comfort is explained by two variables, air velocity and positions 3 and 4 in the room.

These results also need confirmation because the analyzed experimental data is a sample of only 12 cases. However, the statistical analysis allowed a better understanding of the variables involved in the assessment of thermal comfort and provided important clues for future data acquisition. In planning a new experiment, the following recommendations should be considered:

- New data collection to increase the sample and thus enable more statistically relevant conclusions;
- Maintain the reading location, i.e., the same waiting room;
- Reading the same variables in the room positions and height positions considered;
- Testing in a different situation (winter-summer) in order to ensure the variability of the variables excluded from the analysis (clothing and metabolism rate);
- Later, it will be seeking to replicate the statistical study to confirm the results obtained in this exploratory approach. It is thus expected to identify the explanatory variables in this specific problem of indoor thermal comfort and, in the future, to extend the study to different waiting rooms with similar or distinct characteristics.

As referred by Caballer-Tarazona and Coll-Serrano [9] "*it is useful for the learning process to propose practical activities that can connect theoretical concepts with real applications*". The statistical analysis of thermal comfort data was an opportunity to bring theory to practice and reinforced the need of a multidisciplinary activity that links different subjects in Engineering.

ACKNOWLEDGMENT

This research was supported by Portuguese Foundation for Science and Technology (FCT), within the Project Scope: UIDB/00319/2020 (ALGORITMI) and R&D Units Project Scope UIDP/04077/2020 (METRICS). Diogo Graçoeiro acknowledges the BII scholarship VCC-ALG-04/2021 attributed by FCT.

REFERENCES

[1] R. G. Easterling, *Fundamentals of Statistical Experimental Design and Analysis*, John Wiley., vol. 148. 2015.

[2] D. C. MONTGOMERY, *Design and Analysis of Experiments*. 2021.

[3] M. Tanco, E. Viles, L. Ilzarbe, and M. J. Alvarez, "Barriers faced by engineers when applying design of experiments," *TQM J.*, vol. 21, no. 6, pp. 565–575, 2009, doi: 10.1108/17542730910995846.

[4] T. N. Goh and S. W. Lam, "Problem-based learning approach to application of statistical experimentation," *Qual. Reliab. Eng. Int.*, vol. 26, no. 4, pp. 365–373, 2010, doi: 10.1002/qre.1089.

[5] R. A. Funny, M. A. Ghofur, W. Oktiningrum, and N. L. S. Nuraini, "Reflective thinking skills of engineering students in learning statistics," *J. Math. Educ.*, vol. 10, no. 3, pp. 445–458, 2019, doi: 10.22342/jme.10.3.9446.445-458.

[6] F. M. Aparicio Acosta, "Hints for the improvement of quality teaching in introductory engineering statistics courses," *Eur. J. Eng. Educ.*, vol. 25, no. 3, pp. 263–280, 2000, doi: 10.1080/030437900438694.

[7] D. Nolan and T. P. Speed, "Teaching statistics theory through applications," *Am. Stat.*, vol. 53, no. 4, pp. 370–375, 1999, doi: 10.1080/00031305.1999.10474492.

[8] R. Y. Mustafa, "The Challenge of Teaching Statistics to Non-Specialists," *J. Stat. Educ.*, vol. 4, no. 1, 1996, doi: 10.1080/10691898.1996.11910504.

[9] M. Caballer-Tarazona and V. Coll-Serrano, "The Raising Factor, That Great Unknown. A Guided Activity for Undergraduate Students," *J. Stat. Educ.*, vol. 28, no. 3, pp. 304–315, 2020, doi: 10.1080/10691898.2020.1832006.

[10] N. J. O. Rodrigues, R. F. Oliveira, S. F. C. F. Teixeira, A. S. Miguel, J. C. Teixeira, and J. S. Baptista, "Thermal comfort assessment of a surgical room through computational fluid dynamics using local PMV index," *Work*, vol. 51, no. 3, pp. 445–456, 2015, doi: 10.3233/WOR-141882.

[11] S. Teixeira, Inês, Rodrigues, Nelson, Teixeira, "Energy, Thermal Comfort and Pathologies –A Current Concern," in *Occupational and Environmental Safety and Health III*, vol. 406, P. M. Arezes, J. S. Baptista, P. Carneiro, J. Castelo Branco, N. Costa, J. Duarte, J. C. Guedes, R. B. Melo, A. S. Miguel, and G. Perestrelo, Eds. Cham: Springer International Publishing, 2022, pp. 273–281.

[12] N. Rodrigues, "Modelação computacional e avaliação experimental do conforto térmico ocupacional em salas de operação," p. 188, 2017 in portuguese.

[13] P. H. Kvam and B. Vidakovic, "Nonparametric Statistics with Applications to Science and Engineering," *Nonparametric Stat. with Appl. to Sci. Eng.*, pp. 1–429, 2007, doi: 10.1002/9780470168707.

[14] IBM and SPSS, "IBM SPSS statistics base 25," p. 341, 2017.

[15] I. B. M. S. Statistics and B. Guide, "IBM SPSS Statistics 24 Brief Guide."

[16] M. Lewis-Beck, A. Bryman, and T. Futing Liao, "Dummy Variable," *SAGE Encycl. Soc. Sci. Res. Methods*, 2012, doi: 10.4135/9781412950589.n259.