Statistical Interpretation of Carbon Dioxide Emission in the Photocopying Process

Jelena Kiurski[1*], Ivana Oros[1], Vesna Kecić[1], Snežana Aksentijević[2]

[1]University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia
[2]Business Technical College, Trg Svetog Save 34, 31000 Užice, Serbia
e-mail: kiurski@uns.ac.rs

Abstract
The concentration levels of carbon dioxide, monitored in indoor of three photocopying shops in Novi Sad, Serbia, were subjected to two-way ANOVA, in order to investigate statistically significant emission, due to the different sampling points and time intervals. Obtained results pointed out that only selection of time interval significantly affects the emission of carbon dioxide. In addition, the least significant difference test indicated that second time interval has the greatest influence on the CO_2 emission, whereas the CO_2 concentrations of the second time interval were subjected to further cluster analysis. Hierarchical clustering grouped three photocopying shops into four clusters with similar CO_2 concentration levels.

Introduction
Carbon dioxide (CO_2) is a natural constituent of the atmosphere, but in increased concentration it can be harmful to human health. Carbon dioxide can be present in occupational environment as a result of people's respiration process, as a product of combustion and as a component of soil gas. Also, it can be produced during the photocopying process, when photocopier or laser printer toner is heated in an inadequate air supply. Further, paper and electricity consumption, as well as human activity during the photocopying process can indirectly contribute to the formation of CO_2 [1]. The importance of controlled carbon dioxide emission is reflected through global warming potential, considering carbon dioxide as a principal greenhouse gas. Since 36% of CO_2 emission is attributable to manufacturing industries, controlled emission is considered as a major requirement and a principal part of environmental maintenance [2].
The present study aims to examine a carbon dioxide emission in three photocopying shops. A two way analysis of variance (ANOVA) with *post hoc* test, as primary statistical methods, was applied in order to investigate the statistically significant differences of pollutant emission. The obtained results served as a basis for further application of cluster analysis, in order to investigate the similarities/dissimilarities between analyzed photocopying shops, in terms of carbon dioxide emission.

Materials and methods
Sampling method. Five day measurements of carbon dioxide were carried out in three photocopying shops located in Novi Sad, Serbia. Air samples were collected and analyzed by using an instrument Aeroqual Series 200 (Aeroqual Limited, New Zealand). Three sampling points A, B and C (A and B – near photocopier machines; C – near the door) were selected based on the CO_2 emission sources. Also, three time intervals were chosen during the day: at the beginning of the working time - from 8 to 10 a.m., during the maximum productivity time - from 13 to 15 p.m. and the end of the working time - from 16 to 18 p.m. [3].
Two way ANOVA with post hoc test. The two-way ANOVA is statistical method that examines the influence of two different categorical independent variables on one continuous dependent variable. Obtained results of ANOVA test are interpreted by *F*-value for each factor, which is compared to $F_{critical}$, obtained from the table of limit values of *F* distribution

for a certain degrees of freedom. Statistically significant difference exists between observed groups if the value of parameter F is higher than $F_{critical}$, meaning that independent variable has effect on the observed dependent variable [4]. However, the calculated F value using the ANOVA test does not give an answer to the question whether a statistically significant difference occurs between the mean values of all groups or only between particular groups. Therefore, it is necessary to test the differences between arithmetic means of samples and to determine the correctness of certain alternative hypothesis, which is performed using various *post hoc* tests. The most commonly used *post hoc* test is the least significant differences (LSD) test that compare differences between the absolute values of the examined groups $\left|\overline{x_i} - \overline{x_{i+1}}\right|$ with a critical value, LSD. The difference between two samples is significant if the difference between two sample means is larger than LSD value, and vice versa [4].

Cluster analysis. Clustering is one of the most widely used multivariate techniques for exploratory data analysis. The purpose of cluster analysis is to maximize between-group variance and minimize within-group variance by grouping the data objects, based only on information found in the data. In this way, a relationship between the objects is described. Cluster analysis organizes natural groups within the data in such way that each element in the group is similar to each other as possible. At the same time, the groups are dissimilar to other groups [5]. If plotted, geometrically, the objects within the clusters will be close together, while the distance between clusters will be farther apart. If the similarity or homogeneity within a group is greater, the clustering will be better, or more distinct. Cluster analysis is not a statistical technique and the results obtained are justified according to their value in interpreting data and indicating patterns [6].

All the data in ANOVA test and cluster analysis were analyzed using Microsoft Excel 2007 and XLSTAT 2015.1.01.

Results and discussion

Carbon dioxide emission in the indoor environment may vary depending on the selection of time interval and sampling point, which are identified as factor A and factor B in the present experiment. In order to determine the statistically significant differences in carbon dioxide emissions due to the defined factors, two-way ANOVA without replication was applied on the average CO_2 concentrations in three photocopying shops (Table 1), as primary statistical method in the interpretation of the experimental results. The results of two-way ANOVA are presented in Table 2.

Table 1. Average carbon dioxide concentration (ppm)

Photocopying shop 1	1st time interval	2nd time interval	3rd time interval
Sampling point A	1075.16	1186.68	1041.72
Sampling point B	1100.20	1247.88	1053.12
Sampling point C	1097.16	1202.76	1034.96
Photocopying shop 2	1st time interval	2nd time interval	3rd time interval
Sampling point A	911.16	960.84	913.60
Sampling point B	896.72	941.36	913.64
Sampling point C	884.60	940.48	920.80
Photocopying shop 3	1st time interval	2nd time interval	3rd time interval
Sampling point A	874.56	846.28	856.28
Sampling point B	875.96	848.72	864.20
Sampling point C	880.44	844.52	876.56

Table 2. Results of two-way ANOVA without replication

	Photocopying shop 1	Photocopying shop 2	Photocopying shop 3
Source of Variation	***F-value***		
Sampling point	3.69	1.77	1.45
Time interval	101.69	22.21	20.73

The obtained *F*-values were compared to the $F_{critical}$ value of 6.94 for significance level α = 0.05 and the degrees of freedom d_f = 2 and 4. Based on the obtained results, *F*-values for factor A - time interval (101.69; 22.21; 20.73) in all photocopying shops were significantly higher than $F_{critical}$, in contrast to factor B (sampling point). On that basis, it can be concluded that only selection of time interval significantly affects the emission of carbon dioxide, as opposed to the selection of sampling points. In order to compare statistical differences between time intervals, the least significant difference test is applied. The results are presented in Table 3, where x_1, x_2 and x_3 represent the average CO_2 concentration in first, second and third time interval, respectively.

Table 3. Results of Least significant difference test

	Photocopying shop 1	Photocopying shop 2	Photocopying shop 3
$\lvert\overline{x_1}-\overline{x_2}\rvert$	**121.60**	**50.07**	**30.48**
$\lvert\overline{x_1}-\overline{x_3}\rvert$	47.57	18.52	11.30
$\lvert\overline{x_2}-\overline{x_3}\rvert$	**169.17**	**31.55**	**19.18**
LSD[a]	62.38	21.06	13.26

[a]least significant difference value

Obtained results of LSD test (Table 3) pointed out that highly significant difference exists between the first and second, as well as between the second and third time interval (bold values), indicating that the second time interval is the most significant in terms of carbon dioxide emission. Based on the obtained results, only the average CO_2 concentrations from second time interval are used for further mathematical processing, applying cluster analysis.

Cluster analysis included a hierarchical clustering using the Euclidean distance as a measure of the similarities/disimilarities, as well as Ward's methods of connecting objects. Hierarchical clustering was performed with the aim to group photocopying shops per sampling day, based on the determined carbon dioxide emissions. The result of clustering is presented in a form of dendrogram (Figure 1) which provides a visual representation of analyzed data set grouping (photocopying shop_sampling day).

Four clusters are observed on dendogram: cluster 1 (1_1, 2_3, 2_5 and 3_3), cluster 2 (1_2, 1_4, 1_5, 2_1, 2_2 and 3_2), cluster 3 (1_3) and cluster 4 (2_4, 3_1, 3_4 and 3_5). In all formed clusters each member is most similar to its adjacent member. It can be observed that cluster 3 coresspondes to the photocopying shop 1_third sampling day, where was the the highest CO_2 emission during the photocopying process. The remaining clusters 1, 2 and 4 grouped the photocopying shops per sampling day based on the similar carbon dioxide emissions.

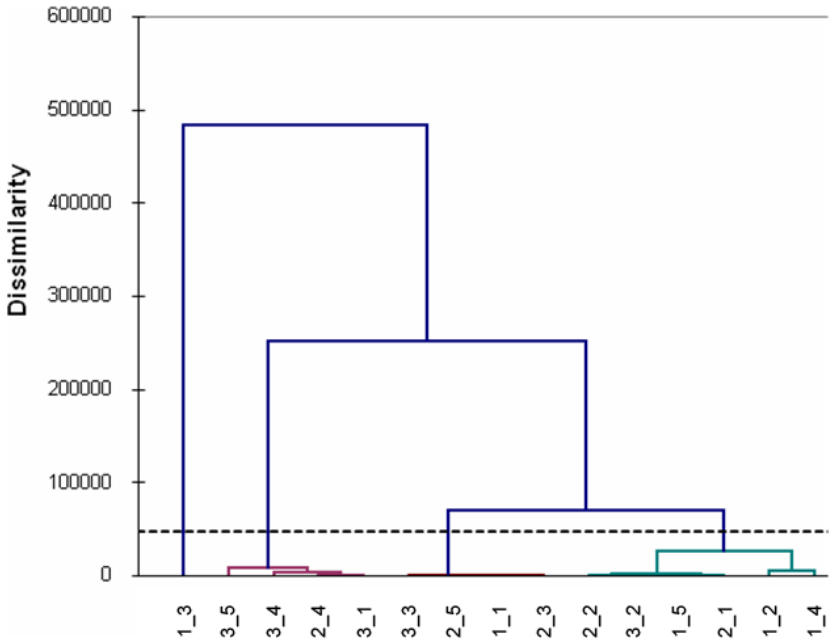

Figure 1. Clustering of photocopying shops according to the carbon dioxide emission

Conclusions

A two way ANOVA with least significant difference test was performed in order to determine the statistically significant emission of carbon dioxide during the photocopying process, due to the different sampling points and time intervals. Given that the results pointed out that second time interval had the greatest influence on carbon dioxide emissions, only the average CO_2 concentrations from this time interval were subjected to cluster analysis. Hierarchical clustering grouped three photocopying shops into four clusters based on similar carbon dioxide emission. Cluster 3 corresponds to photocopying shop 1_third sampling day with the highest CO_2 emission, which is in a compliance with the obtained results of CO_2 measurement.

Acknowledgement

The authors acknowledge the financial support of the Ministry of Education, Science and Technological Development of the Republic of Serbia, in the frame of Project applied under No. TR 34014.

References

[1] M. Santamouris, A. Synnefa, M. Asssimakopoulos, I. Livada, K. Pavlou, M. Papaglas, D. Kolokotsa, V. Assimakopoulos, Energ. Buildings. 40 (2008) 1833.
[2] X. Wang, Z. Duan, L. Wua, D. Yang, J. Clean. Prod. 103 (2015) 705.
[3] J. Kiurski, S. Aksentijević, I. Oros, V. Kecić, 7th International Scientific Conference "Science and Higher Education in Function of Sustainable Development-SED2014", Užice, Serbia, 2014, pp. 16.
[4] M. Hadži Stojković, in: Statistics, Descriptive and statistical analysis, Faculty of Economy, Subotica, University of Novi Sad, 2006. pp. 75. in Serbian
[5] Yu. Pepelyshev, Ts. Tsogtsaikhan, Ann. Nucl. Energy. 83 (2015) 50.
[6] D.J. Kern, K.E. Culley, Comput. in Hum. Behav. 49 (2015) 313.