Please consider, when preparing the manuscript, the Instructions for the Authors at the end of each issue. Orders, exchange for other journals, manuscripts and all correspondences concerning off prints should be sent to the Executive Editors or to the editorial secretaries at the address:

Politehnica University Timisoara
Department of Mathematics
Sq Victoriei 2, 300 006, Timisoara, Romania
Tel.: +40-256-403099 Fax: +40-(0)256-403109

Politehnica University Timisoara
Department of Physical Fundamentals of Engineering,
Bd Vasile Parvan 2, 300223, Timisoara, Romania
Tel.: +40-256-403391 Fax: +40-(0)256-403392

Contents

SCIENTIFIC BULLETIN
OF
THE POLITEHNICA UNIVERSITY OF TIMISOARA

Vol.68(82)	MATHEMATICS–PHISYCS	2023

STUDY ON LONG-TERM CHANGES IN THE ELECTROMAGNETIC ENVIRONMENT USING DATA FROM A CONTINUOUS MONITORING SENSOR

Ileana-Roxana TUTELCĂ

Abstract

With the proliferation of base stations through mobile telephony development projects, as well as the diversification of telecommunications services, medical technology, and household appliances, there has been increasing concern among the population regarding the excessive use of electromagnetic fields. This concern is also due to the absence of direct perception of electromagnetic phenomena. The perception of the presence of a high-intensity electromagnetic field is indirect, through mechanical, thermal, optical, and acoustic effects. This paper aims to analyze the evolution of the electromagnetic field in the environment and consequently human exposure to it over a one-year period, from September 2022 to September 2023. The study is based on data collected from a non-ionizing radiation monitoring sensor in Timişoara. We used the Holt-Winters and ARIMA methods for analysis and prediction, and since the sensor includes frequency filters for separating received frequencies, we analyzed three frequencies and a wide band. The study shows a fluctuation over time of the electromagnetic field, without exceeding the reference level according to OMS 1193/2006.

Index Terms-electric, electromagnetic, frequency, OMS, prediction

1 Introduction

Wireless communication technology has experienced rapid development, and along with it, public concern regarding the perceived risk of exposure to electromagnetic fields has increased. As concerns about exposure to electromagnetic fields have grown, the first standards and regulations for protection against electromagnetic radiation have been developed and implemented. The international organization ICNIRP (International Commission on Non-Ionizing Radiation Protection)

has contributed to the establishment of these standards. In 1999, the European Union adopted "Recommendation 1999/519/EC on the limitation of exposure of the general public to electromagnetic fields (0 Hz to 300 GHz)."[1]. Thus, in Romania, this recommendation is transposed in the form of the "Order of the Minister of Public Health no. 1193/2006 for the approval of the Norms regarding the limitation of exposure of the general population to electromagnetic fields from 0 Hz to 300 GHz "[2], a document that is approved by (ICNIRP). Over time, research in this field has continued to advance, and standards and regulations have been updated and revised based on new discoveries and understandings regarding the impact of electromagnetic fields on health and the environment.

The assessment of human exposure is a constant concern and is encountered in specialized studies from various countries. A spatiotemporal model of radiofrequency field exposures has been evaluated in Chengdu [3], in China, to establish if there are sources whose emission causes exposure above the relevant levels provided in guidelines and standards (ICNIRP şi GB 8702-014). The analyzed data was collected from three categories of areas: commercial area, urban residential area, and rural residential area, using a system installed on a car for two years. No levels exceeding the established exposure limits were recorded.

Another study was conducted in a province of Turkey[4], For two years, taking into account the variations throughout the day. The study aimed to assess the population's exposure to radiofrequency fields in relation to the limits set by ICNIRP and ICTA (Turkey's Information and Communication Technologies Authority). Following the analysis of the data, it was found that during the day, the field levels were with $35,4\%$ higher compared to the levels during the night.

In Greece, there was a study [5] spanning a period of 20 years, regarding the long-term changes in the electromagnetic field, using data obtained from monitoring sensors and thus assessing human exposure to these fields. The study found that there is a fluctuation in the radiofrequency electromagnetic field.

Results of measurements in outdoor, urban environments are highlighted in the paper "Prediction of RFEMF Exposure by Outdoor Drive Test Measurements" [6], in which the artificial neural network (ANN) model was explored for spatial reconstruction, with real data, of exposure to radiofrequency electromagnetic fields. Measurements were taken on the electromagnetic field intensity in Paris, covering a distance of 65 km , in a wide band. With the collected data, two different models were constructed. An ANN1 model where the N nearest base stations were considered regardless of their operator, and in another model, ANN2, which contains base stations belonging to the same operator, grouped into blocks, each block containing $N = 3, 5, 7$ base stations. At the end of the study, it was found that the ANN1 model has better prediction quality as N increases (for N = 7), but the processing time increases, and with the ANN2 model, there is no improvement in prediction by increasing the number of base stations from 3 to 7 .

2 Experimental setup and data collection

As part of the electromagnetic field monitoring project in urban areas with heavy traffic, the data is made available to the public through the ANCOM platform (National Authority for Management and Regulation in Communications)[7], stations are installed throughout Romania that continuously monitor and transmit the data to the information platform every 24 hours.

The monitoring stations with fixed sensors are placed in public spaces, schools, university centers, public squares, in the vicinity of which there are multiple sources of radio emissions.

The equipment used for the continuous monitoring of the electromagnetic field is of the Narda AMB 8059 type and meets the ITU-T K. 83 requirements for continuous multi-band monitoring of the electromagnetic field [8]. The Narda AMB 8059, a remotely controlled module, is equipped with a triaxial isotropic probe that performs wideband measurements of the electric field (E) with a frequency range of 100 kHz to 7 GHz (a range that includes the bands used in Romania for 5G), and three frequency bands from the mobile telephony spectrum: 925 MHz - 960MHz, 1805MHz - 1880MHz, 2110MHz− 2170 MHz .

The monitoring equipment has low energy consumption, but to be operational 24/7 it can be powered by a solar panel or batteries, thus its energy autonomy can be extended up to 80 days. Additionally, it includes internal memory for storing collected information, which can be expanded. Recorded data is transmitted to the server/PC via an integrated modem. An internal GPS module allows for its localization.

The recorded values of the electric field intensity (RMS) are either peak values or maximum values recorded in the monitored frequency bands, or temporally averaged values over 6 minutes.

$$E_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_i)^2} \tag{1}$$

where n represents the total number of discrete values, and E_i represents the discrete values of the electric field intensity.

On the building of the Politehnica University of Timișoara (UPT), such a monitoring station is installed.

On the roof of the building, numerous mobile base stations are installed, but an important contribution also comes from the base stations and antennas located on the building of the West University of Timișoara (UVT), situated in the vicinity of the UPT building. Additionally,

Fig. 1: Screenshot of the map showing the sensor location [7]

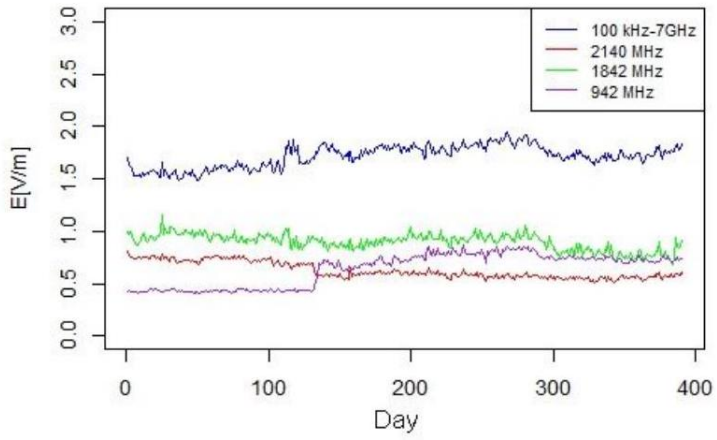
Fig. 2: Placement of the sensor on the UPT building[7] being a university area, the traffic on the mobile networks is high, so it is expected that the data recorded by the monitoring station would be higher compared to other less populated areas with lower antenna.

The data obtained from the sensor was checked to ensure there were no missing or outlier values. Additionally, we removed zero values recorded as average values that had corresponding non-zero peak values, and vice versa.

The values used in the study are daily average values of the data recorded by the sensor, in the wide band ($100\text{kHz} - 7\text{GHz}$) and for the three frequencies: 2140 MHz, 1842MHz, and 942 MHz , during the period from September 2022 to September 2023, as shown in Figure 3.

Fig. 3: Average values of the electric field intensity over a period of one year

In this graph, on the vertical axis, we represent the electric field intensity, measured in V/m, with values ranging approximately between 0.4 V/m and 2 V/m. On the horizontal axis, we represent the days of the year, from 0 to 400 . Each band/frequency is represented by a different color. The highest average values of the electric field, values approaching 2 V/m, are in the wide band. Moderate values, between 1.2 and 0.75 V/m, are at the frequency of 1842 MHz , while the other two have values below 1 V/m. This graph provides a perspective on how



the electric field intensity varies over the course of a year, for different frequencies.

3 Model and experimental validation

We used the simplified Holt-Winters method (simple exponential smoothing model), without trend and without seasonality, to adjust the level of the time series. The method uses the weighted sum of present and past terms, giving higher weights to present series terms and lower weights to past terms. Thus, by applying this method to the daily averaged values for the four frequency domains, we obtained predictions for the next value of the electric field.

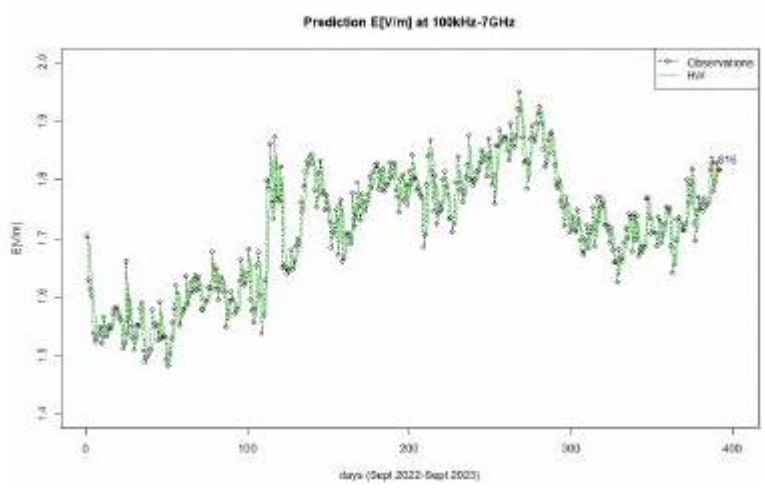


Fig. 4: Holt-Winters method E[V/m] in the 100kHz − 7

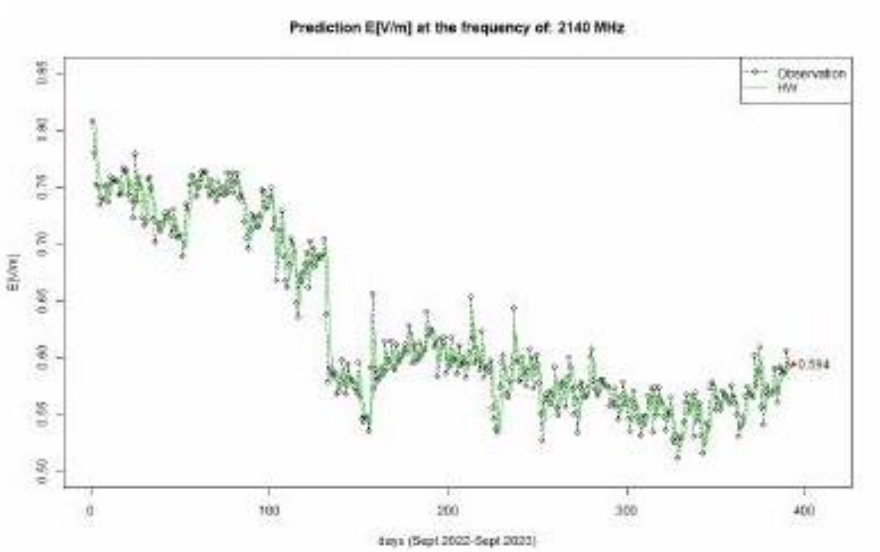


Fig. 5: Holt-Winters method $E[V/m]$ at the frequency of 2140 MHz GHz band

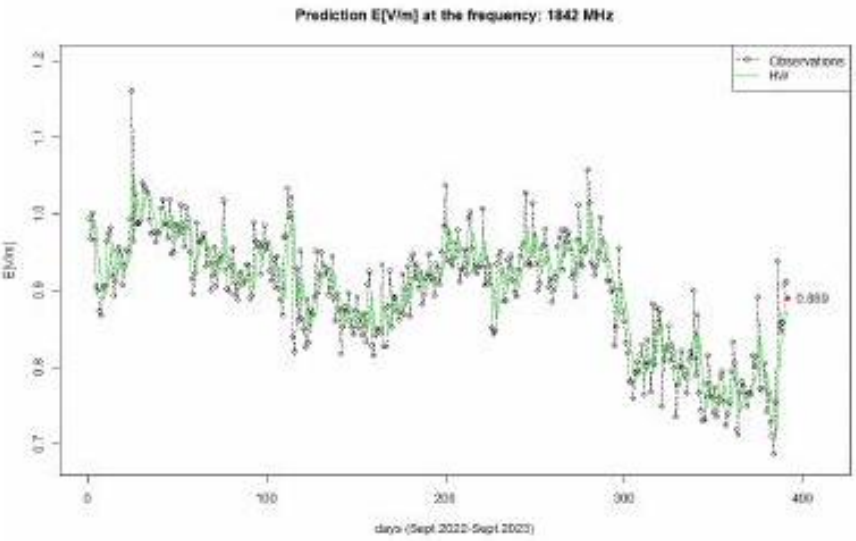


Fig. 6: Holt-Winters method E[V/m] at the frequency of 1842 MHz

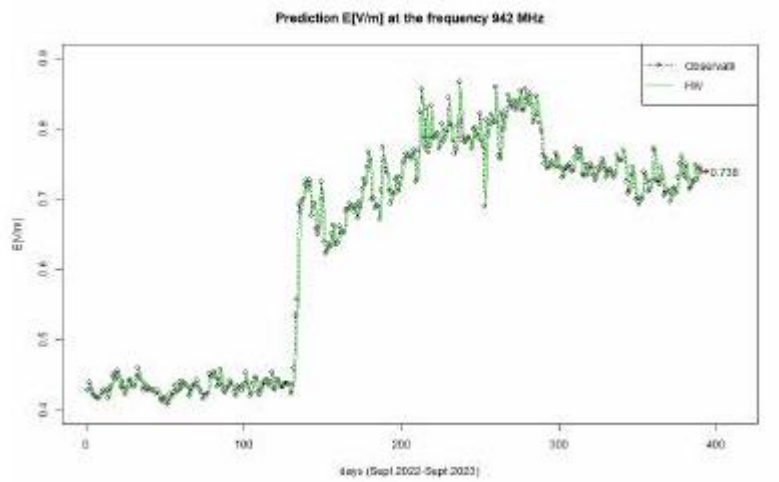


Fig. 7: Holt-Winters method E[V/m] at the frequency of 942 MHz

Another method we used in the time series analysis is ARIMA (p,d,q) (AutoRegressive Integrated Moving Average), composed of p - the number of autoregressive terms, d - the number of differencing needed to make the series stationary, and q - the number of moving average terms. We obtained the optimal parameters for ARIMA considering the AIC (Akaike Information Criterion) to identify the model that best balances complexity and data fitting.

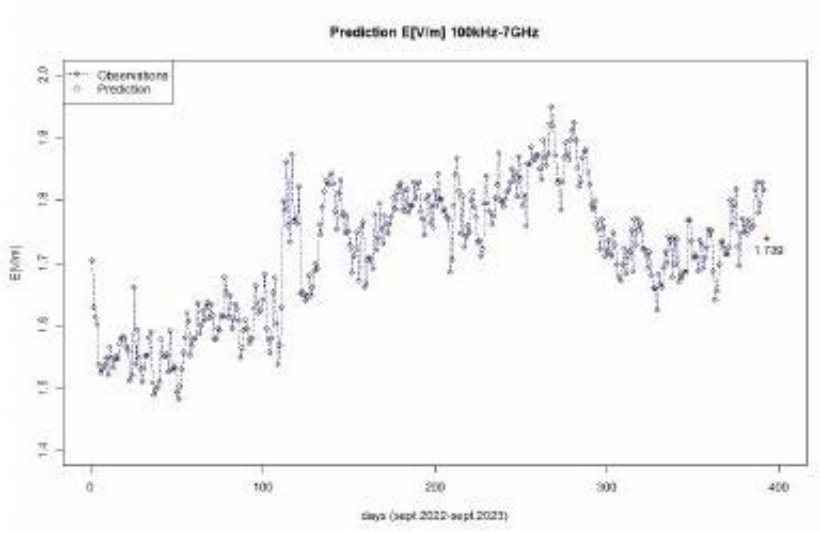


Fig. 8: The ARIMA $(3, 2, 0)$ method $E[V/m]$ in the 100kHz $-$ 7GHz band

Thus, for the data in the 100kHz−7GHz band, we obtained the ARIMA(3,2,0) model; for the 2140 MHz frequency, we obtained ARIMA$(1, 2, 0)$; for the 1842 MHz frequency, we obtained ARIMA(2,1,2); for the 942 MHz

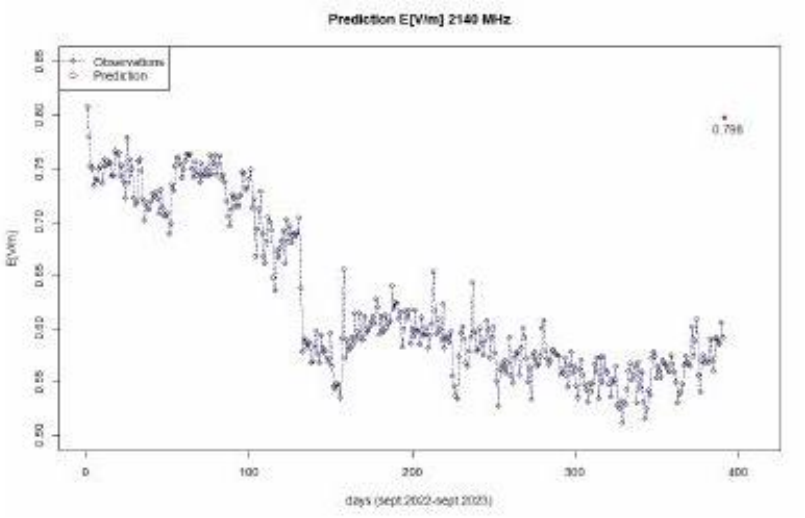


Fig. 9: The ARIMA $(1, 2, 0)$ method E[V/m] at the frequency of 2140 MHz

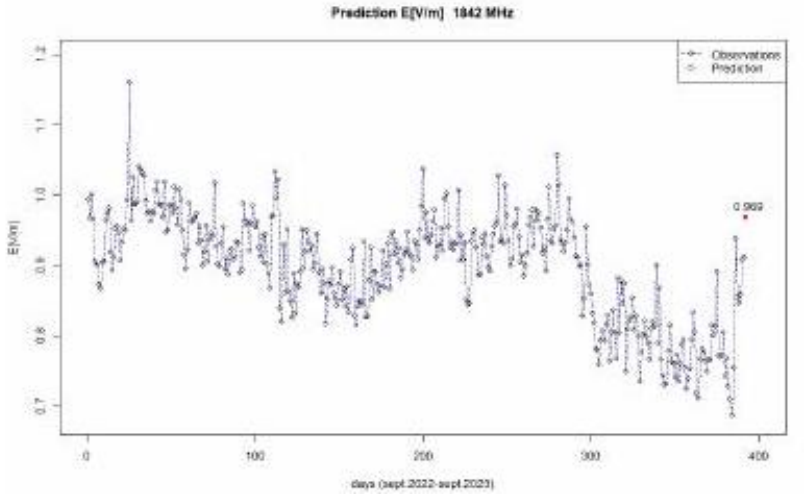


Fig. 10: The ARIMA $(2, 1, 2)$ method E[V/m] at the frequency of 1842 MHz

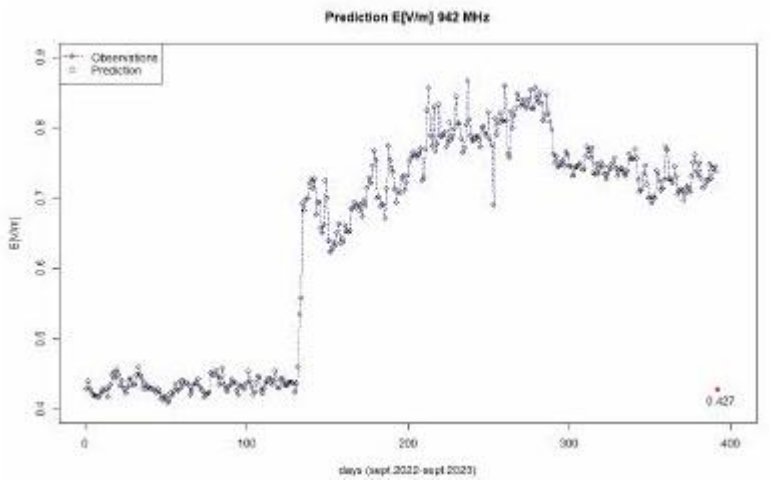


Fig. 11: The ARIMA($1, 2, 0$) method E[V/m] at the frequency of 942 MHz The values predicted by the two methods were compared with the last value in the series, kept for verification of the predictions, and we highlighted them in the graph 12 .

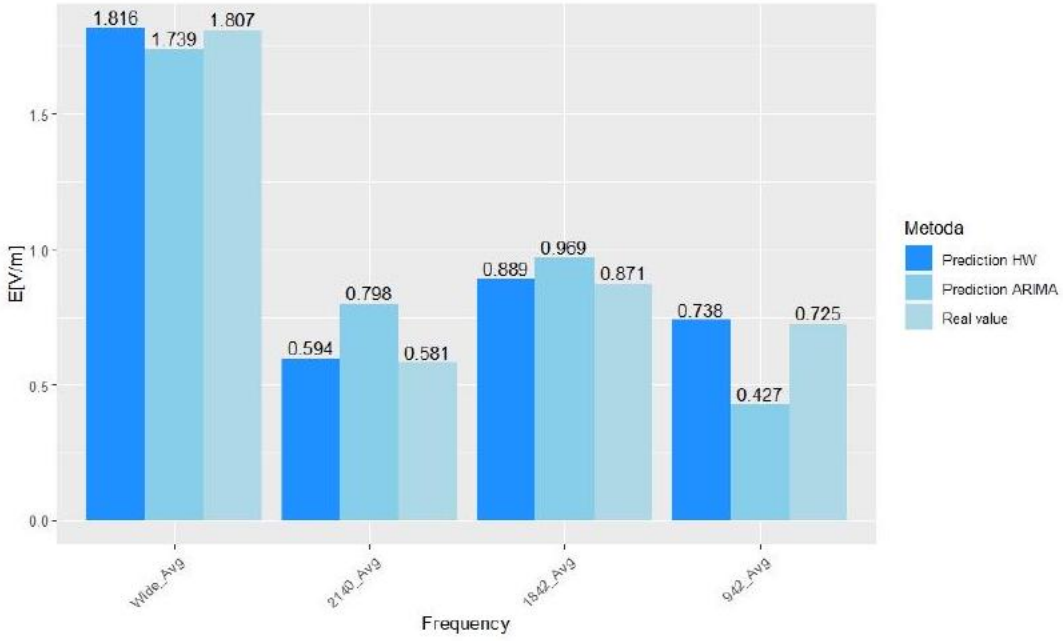


Fig. 12: Predicted values vs. actual value, for the average values of the electric field

A better prediction is observed using the exponential smoothing method. The smallest fluctuation in deviation from the actual value is achieved by the Holt-Winters method.

Following the use of the two methods, we calculated the deviations from the actual value for each frequency domain using the formula:

Deviation = Predicted value - Actual value

TABLE I: Prediction deviations

Frequency	deviation from the actual value	
	Holt-Winters	ARIMA
100kHz − 7GHz	0,009	$-0,068$
2140 MHz	0,013	0,217
1842 MHz	0,018	0,871
942 MHz	0,013	$-0,298$

In conclusion, the graph shows that both prediction methods can be useful, but the Holt-Winters method tends to be more accurate in this dataset. The ARIMA model has mixed performance and may require additional adjustments to improve prediction accuracy in some cases.

To track how the level of the electric field in the wide band has evolved, we analyzed the months of September 2022 and September 2023.

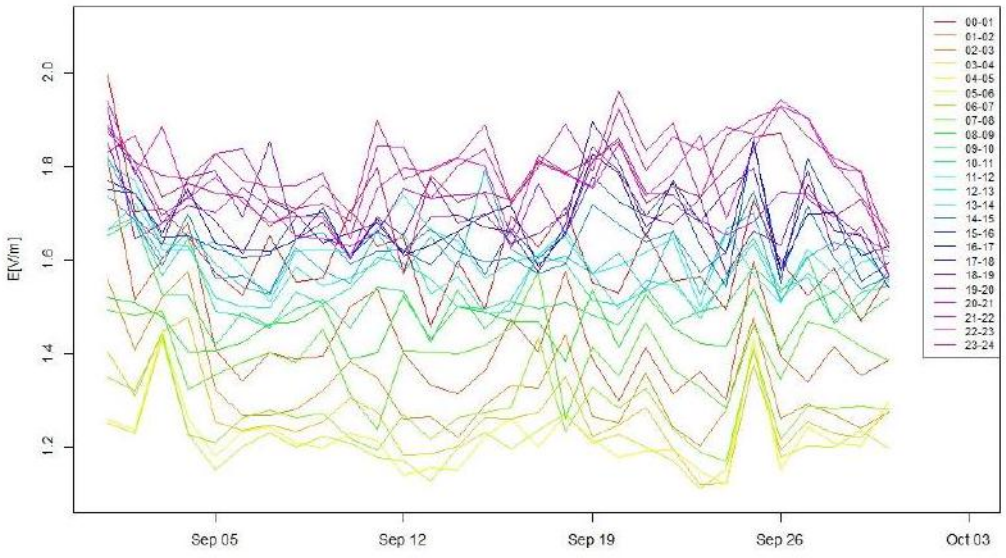


Fig. 13: The average values, per one-hour intervals, of the electric field intensity in the 100kHz − 7GHz band in September 2022

I averaged the electric field values over hourly intervals for the two months.

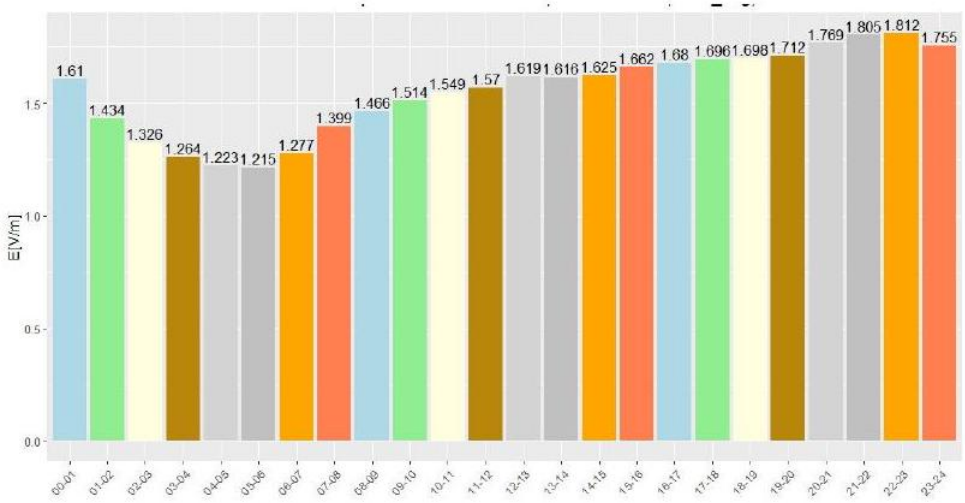


Fig. 14: The average values, per one-hour intervals, of the electric field intensity in the 100kHz-7GHz band in September 2022

We observe high values in the hourly interval $00-01$, which gradually decrease until around 05-06. This is followed by an approximately 25% increase until around noon, when the values remain relatively constant until 19. After this time, the values tend to increase until a peak around 23, then decrease slightly until 24 .

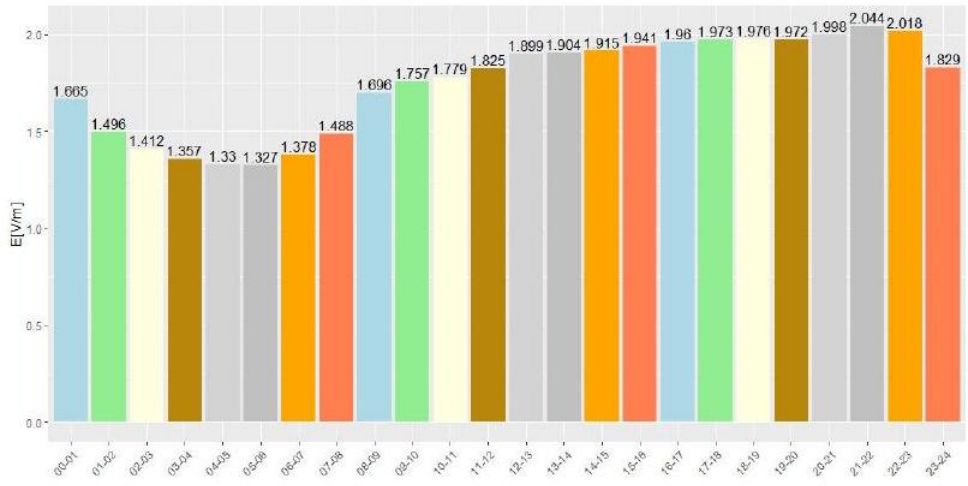


Fig. 15: The average values, per one-hour intervals, of the electric field intensity in the 100kHz-7GHz band in September 2023

For the month of September 2023, the same diurnalnocturnal pattern is observed, with lower values at night and higher values during the day and evening. The absolute difference between the maximum value, recorded in the 21-22 PM time interval, and the minimum value, recorded in the 5-6 AM time interval, is 0.717 [V/m] higher, and the percentage difference shows that the value 2.044 is approximately 54.03% higher than 1.327 .

By comparing the obtained average electric field values with the reference levels according to OMS 1193/2006, we obtained the graph below.

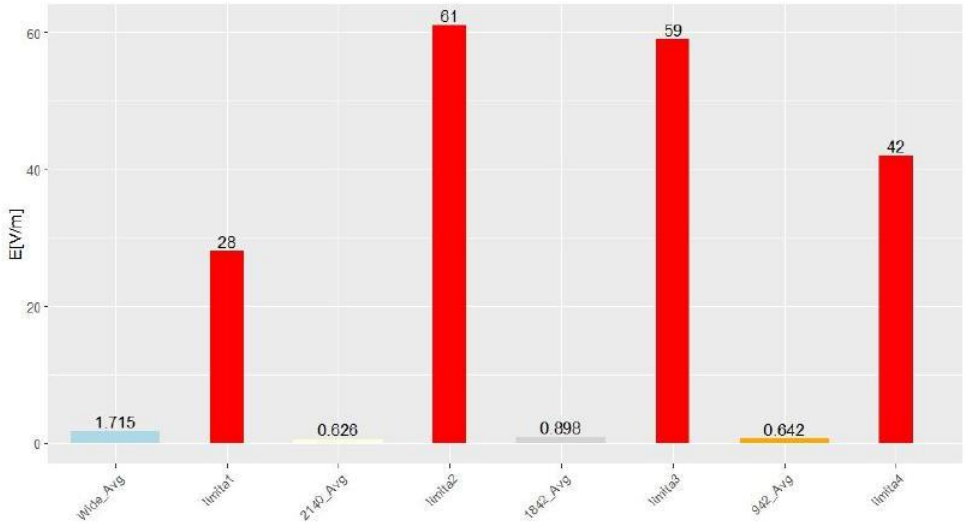


Fig. 16: Evaluation of the electric field and comparison with reference levels according to OMS 1193/2006

4 Conclusions

I conducted a statistical analysis of a database containing records related to the electric field level $E[V/m]$ (RMS values temporally averaged). The recordings were collected by the sensor installed on the UPT building. Values of the electric field were collected in a broad frequency range ($100\text{kHz} - 7\text{GHz}$) and on three specific frequencies ($2140\text{MHz}, 1842\text{MHz}, 942\text{MHz}$), primarily used by mobile phone operators.

From the graph showing the average electric field intensity values over a one-year period, I observed a decrease in the electric field level received at the frequency of 2140 MHz at the beginning of January 2023, which coincided with an increase in the level received at the frequency of 942 MHz . These increases/decreases in levels can be explained by reconfigurations resulting from the upgrading of base stations.

I performed a time series analysis using the HoltWinters and ARIMA methods to obtain a prediction of the last value in the dataset, which was then compared with the actual measured value. Using both methods, I obtained deviations ranging between $1\% - 1.2\%$ for the Holt-Winters method and between $0.59\% - 1.37\%$ for the ARIMA method. Both prediction methods can be useful, but the Holt-Winters method tends to be more accurate for this dataset. The ARIMA method has mixed performance here and may require further adjustments to improve prediction accuracy.

I calculated the hourly averages of the electric field, in the $100\text{kHz} - 7\text{GHz}$ band, for each hour of the day aggregated for the entire month (considering September 2022 and September 2023) to monitor the evolution of the field level. I found that there is a diurnal-nocturnal pattern, with lower values during the night, towards the morning, and higher values throughout the day and late evening. These differences reach up to approximately 54% higher late at night compared to the lowest values in the early morning.

At the end of the analysis, I compared the average electric field values from the period of September 2022 to September 2023 with the reference level according to WHO 1193/2006 and concluded that there are no exceedances. Moreover, the measured values represent 1 6% of the maximum level provided in the legislation. The obtained graph clearly illustrates the significant difference between the actual values and the legislative limits, emphasizing compliance with safety regulations.

Acknowledgment

Thank you for the guidance and support to Mr. Romeo Negrea and to Mr.Teodor Petrita.

References

[1] Council Recommendation et al. Limitation of exposure of the general public to electromagnetic fields (0 hz to 300 ghz). Official Journal of the European Communities, 199, 1999.

[2] The Government of Romania. Order of the minister of public health no. 1193/2006 approving the norms for limiting the exposure of the general population to electromagnetic fields from 0 hz to 300 ghz . Monitorul Oficial, 2006.

[3] Zhu G.; Gong X.; Luo R. Characterizing and mapping of exposure to radiofrequency electromagnetic fields (20-3000 mhz) in chengdu. Health Physics 2017, 112(3):266-275, 2017.

[4] Kurnaz C.; Mutlu M. Comprehensive radiofrequency electromagnetic field measurements and assessments: A city center example. Environ Monit Assess 2020, 192(6):334, 2020 .

[5] Athanasios Manassas, Christos Apostolidis, Serafeim Iakovidis, Dimitrios Babas, and Theodoros Samaras. A study of the long term changes in the electromagnetic environment using data from continuous monitoring sensors in greece. Scientific Reports, 13(1):13784, 2023.

[6] Shanshan Wang, Taghrid Mazloum, and Joe Wiart. Prediction of rf-emf exposure by outdoor drive test measurements. In Telecom, volume 3, pages 396406. MDPI, 2022.

[7] Monitor EMF. Measurements with emf sensors placed in fixed positions. http://www.monitor-emf.ro, 2024.

[8] International Telecommunication Union. Itu-t standards list for study group 5 and domain 40, 2024. Accessed: 2024-05-29.

Tutelcă Ileana-Roxana – MSc student, Department of Mathematics,
Politehnica University of Timişoara,
P-ta Victoriei 2, 300 006, Timişoara, ROMANIA
E-mail: ileana.tutelca@student.upt.ro

SCIENTIFIC BULLETIN
OF
THE POLITEHNICA UNIVERSITY OF TIMISOARA

Vol.68(82)	MATHEMATICS–PHISYCS	2023

A COMPARISON OF AMBIENT RADIOFREQUENCY ELECTROMAGNETIC FIELD (RF-EMF) LEVELS IN DIFFERENT OUTDOOR AREAS IN ROMANIA

Ioan Dorel TODOR

Abstract

The recent development of wireless communications in Romania has caused concerns among the population regarding how it can affect health. ANCOM has implemented a nationwide monitoring network of electromagnetic emissions with sensors located in over 100 localities that collect data on the ambient level of the electric field. Until now, there has not been an evaluation of this information at the national level and the present work pursues three objectives using data collected over a period of 12 months, from a sample of 44 localities from all historical regions.

The first objective is to establish a average values of the electric field level over a wide area in Romania and comparing it with the reference level in the public health legislation.

The second objective is to make some comparisons between the annual averages of the electric field level in four groups of 11 localities established on the population density criterion.

The third objective is the verification of a supposed increase in the electric field level with the implementation of 5G technology.

Using statistical evaluation methods, the results obtained show an average level of exposure for the 44 localities of 5.78 % of the maximum exposure limit established by OMS1163/2006, the statistical hypothesis regarding the homogeneity of the average values of the electric field for the four groups of localities cannot be rejected either. Multivariate analysis of broadband electric field level composition data cannot demonstrate a greater contribution of the level generated by the deployment of 5G technology. .

1 Introduction

The new wireless communication technologies have experienced an accelerated development in Romania as well as in the rest of the economically developed

countries, with a negative consequence and an increase in the level of electromagnetic pollution coming mainly from the base stations of mobile communication operators. In response to the population's concerns regarding how the limits imposed by the national legislation on exposure to electromagnetic fields are respected, ANCOM has implemented a network of sensors that measure these levels in real time and make them available to those interested on a public portal (emfmonitor .ro).A fear among an important part of the population has particularly induced the widespread use of mobile communications in more and more regions. Thus, a survey carried out in 2006 at the level of the European Union shows that over 60 % of the respondents believe that wireless communication systems would affect health to some extent [1] and through numerous recent studies an attempt has been made to clarify some aspects regarding these fears.

In Europe and beyond, studies were conducted that aimed to verify some hypotheses on how the noxiousness generated by new technologies have a negative influence on health. Thus, a 2014 study done following the International Agency for Research on Cancer's declaration of radiofrequency electromagnetic fields as possible causes of cancer [2], by researchers Michael Carlberg and Lennart Hardell, highlighted a direct relationship between the increased level of electromagnetic noxes and the survival rate of patients with Astrocytoma Grade IV [3]. A study carried out in Saudi Arabia verified the existence of a positive correlation between exposure to RF-EMF (Radiofrequency-Elecromagnetic Fields) and increased levels of glycated hemoglobin (HbA1c) in 250 volunteer subjects who were students at a university in Riyadh. The results of the study indicated an increased risk of diabetes in students who were exposed to electromagnetic noise compared to the control group [4].

Other studies have checked the possible changes over time in the levels of exposure of the population due to the technological changes in the field of wireless communications such as the one in Switzerland [5] which followed the evolution of the RF-EMF level compared between the years 2017 and 2021. The study was done in different areas representative for population density (Rural, suburban and urban) and in different micro environmental environments including in different means of public transport. The results obtained from the evaluation by statistical means of the data obtained from the measurements showed insignificant changes between the two years in terms of RF-EMF values.

Closer to us, a study from Serbia aimed to verify EMF levels near mobile communication stations and the results showed that on the one hand the general level of noise contains all the components generated by mobile communication stations, and on the other hand that the maximum of the electromagnetic field level measured is below the reference limit of the legislation in Serbia although the distance between the emission antenna and the measurement point was 60 m [6].

A 2020 study by researchers from the Gheorghe Asachi University in Iași, in

collaboration with ANCOM, formulated conclusions regarding the level of exposure to EMF at the level of the municipality of Iași [7] The data were taken from the 4 ANCOM sensors installed in populated areas from the city for the whole year 2020 and the average and maximum values were used for comparison with reference levels in Romaninan national legislation (OMS 1193/2006)

According to the data published by ANCOM in a synthesis of the electronic communications market from the first half of 2023 in Romania, a growth rate is maintained in the internet segment and data that express the trend of digital transformation of the data transmitted through these systems. From the same synthesis, a 4 % increase in mobile internet traffic is observed, most of it being done on 4G networks (94 %) but also a doubling of traffic on 5G networks [8]. Thus the legitimate question arises whether this increase is reflected in an increase in the level of electromagnetic pollution over a wide area and a scientifically based answer is desired to be made in the present work.

2 Materials and methodes

2.1 Equipment used for measurements

To measure the electric field level, specialized sensors for electric field level measurements like the one in Figure 1 were used. The Narda AMB module is con-

Figure 1: Narda AMB 8059/03 module and EB-4B-02 probe

nected to a specialized isotropic triaxial antenna for broadband measurement (100KHz-7GHz) which, together with the sensor's active filters, also allows measurement on three distinct bands (942MHz, 1842MHz and 2140MHz) used by mobile communications operators in Romania. With few exceptions, it can be said that most applications using wireless technology are covered, such as mobile communication networks, AM and UUS radio transmitters, DVB-T2 transmitters, Wi-Fi networks, mobile or fixed transmitters for point-to-point communications. ANCOM sensors are equipped with autonomous energy supply systems (solar panel) and GSM modem used to transmit data once a day to the server. I mention that I removed the data recorded during the transmission period from the database used in order not to influence the calculated average value. More details about the equipment used can be found in Table 1

Table 1: AMB 8059 Tehnical specification

AMB-8059 Multi-band EMF Area Monitor Tehnical specification	
Measurements units	Wm, kV/m, nT, VT, mT, %
Field measured	Total field, average and Peak (MAX)
Sampling	1 measurement every 1 s
Memorization interval	Programmable from 30 sec to 15 min
Memory	Over 128 MB
Max data storage capacity	Over 364d with 1 acquisition/min
Field strength alarm	Two programmable field strength thresholds
Clock	Real time internal clock
Sensor	Display of model and calibration date
Battery management	Every record includes Battery Voltage and Charge Current value
Temperature management	Every record includes Internal Temperature value
Humidity management	Every record includes Internal Humidity value
GPS coordinates	Programmable record

2.2 Location and measurement conditions (measured frequency bands

Location of the the sensors was done in compliance with the ITU-T K.83 recommendation regarding the monitoring of electromagnetic field levels. Sensors are usually placed on building roofs or higher locations to have an unobstructed line of sight to electromagnetic field sources, in more densely populated areas such as public buildings, university campuses, public squares, hospitals, transport stations. The measured values can be negatively influenced by the existence in the

immediate vicinity of the sensors of some metal structures or even by the own modem at the time when the data is transmitted to the server, so these aspects were taken into account at the time of data processing. An example of the location of a sensor is the one shown in Fig.2

Figure 2: Senzor EMF in Timisoara

Taking into account that from the point of view of the share of electromagnetic field sources in the exposure level of the population, the mobile communication systems are in the first place, also the ANCOM's sensor network was designed to monitor these systems.

The frequency bands allocated in Romania for this type of communications were auctioned and allocated to the five main operators as can be seen in Fig. 3

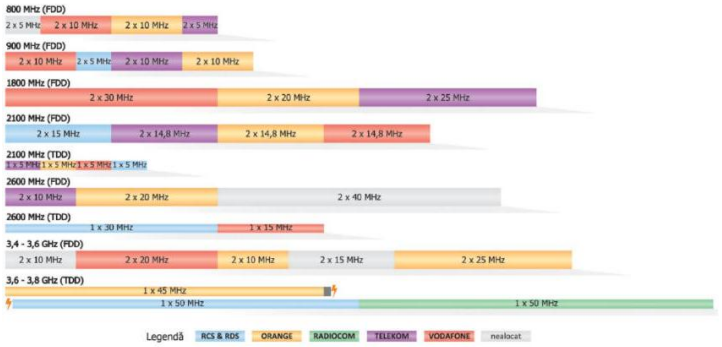


Figure 3: Mobile comunication frequency spectre used in România

Outside of these bands, radio transmitters transmitting in the medium-wave

(UM) and ultra-short-wave (UUS) bands and television transmitters transmitting in the ultra-high frequency (UHF) band have a smaller share. The values measured and recorded by the ANCOM sensors represent the total value of the electric field level measured on the three propagation planes with the formula:

$$E_{TRMS} = \sqrt{E_{XRMS}^2 + E_{YRMS}^2 + E_{ZRMS}^2} \tag{1}$$

and each of the three components are calculated in turn with the formula:

$$E_{RMS} = \sqrt{\frac{1}{T}\int_0^T E(t)^2, dt} \tag{2}$$

where E(t) is the electric field as a function of time, and T is the time period over which the integration is done.

In the context of exposure to electromagnetic radiation, RMS values of the electric field are worldwide used to assess exposure levels and ensure compliance with safety standards.

2.3 The locations of the sensors

In Romania, the EMF monitoring network includes a number of more than 70 localities where more than 140 sensors are located and used for monitor 24/7 the level of the electric field and the data collected by them is available on the monitor-emf.ro [9] platform. An interactive map with the location of the sensors on the national territory is presented in Figure 4

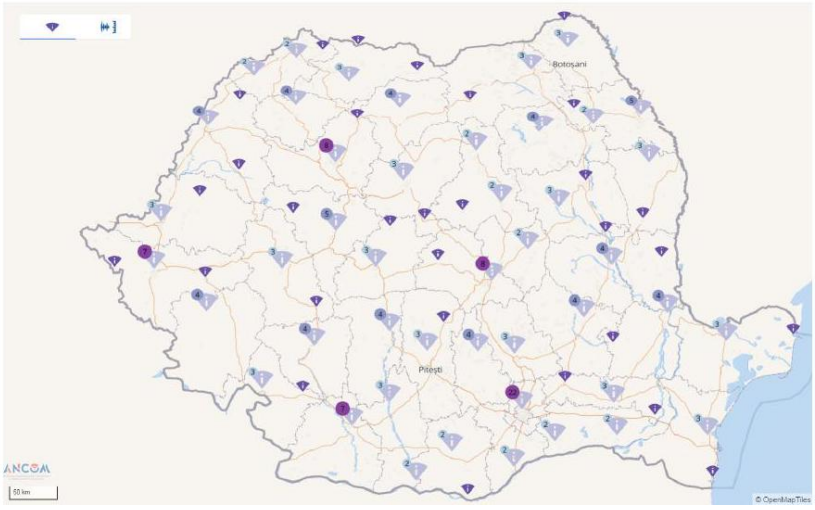


Figure 4: The locations of ANCOM sensors from Romanian cityes

On the web page accessible for free when accessing a location, the electric field values of the last 3 days are displayed in a window in the form of a graph as seen in the screenshot shown in Figure 5

Figure 5: Prezentarea informațiilor unui senzor din rețeaua EMF prin accesarea hărții interactive

For this study, we obtained with courtesy of ANCOM the electric field level value data from 44 localities, mainly county seats or municipalities in all historical provinces to have an overview of the RF-EMF level at the national scale. From the initial volume of data, we extracted average and peak electric field level values in broadband (100KHz-7GHz) and on the three bands assigned to mobile communications (925-960MHz, 1805-1880MHz, 2110-2170MHz). In addition to these data, we also retrieved information regarding the moments when the recording was made (timestamp) and the ambient temperature values from those moments. The data were initially processed in excel where we did a filtering of non-compliant values (values recorded at the time when the sensors were transmitting data to the server, or exaggerated values also possibly caused by recording/transmission errors).

In order to verify a supposed connection between the electric field values measured in the localities and the population density, we added data on the number of permanent residents in the respective localities with data taken from the romanian population census carried out in 2021 [10]. The 44 localities we then grouped in 4 categories of 11, each based on the population density criterion as can be seen in Table 2

Table 2: Grouping of localities according to the number of residents

EMF monitored localities table	
Cityes group names	Number of permanent residents
Municipii 1	10000-35000
Municipii 2	35100-72000
Reședințe 1	73000-183000
Reședințe 2	200000-1717000[a]

[a]Bucharest included

3 Statistical data processing and results

3.1 RF-EMF values measured at national level

The first data processing consisted in calculating the means of annual values for the CE level measured in broadband to visualize a possible correlation between it and the population density, a bar graph with the recorded values is shown in Figure 6.

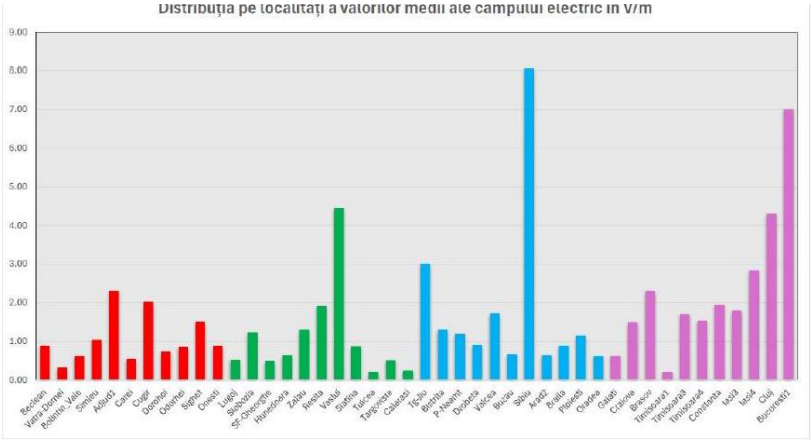


Figure 6: Recorded means CE values in 44 localities; Municipii 1 -Red, Municipii 2 - Green, Reședințe 1 - Blue, Reședințe 2 - Violet)

The values related to each locality represent the average annual value of the electric field and as can be seen they fell within the range of 0.20 - 8.05 [V/m].

The average annual EC value for the 44 localities was 1.53 [V/m] which represents a percentage of 5.56 % of the maximum exposure limit imposed by legislation [11].

In order to better observe the distribution of values in each group of localities, as well as the median values, quartiles and outliers, we created a boxplot type graph presented in Figure 7.

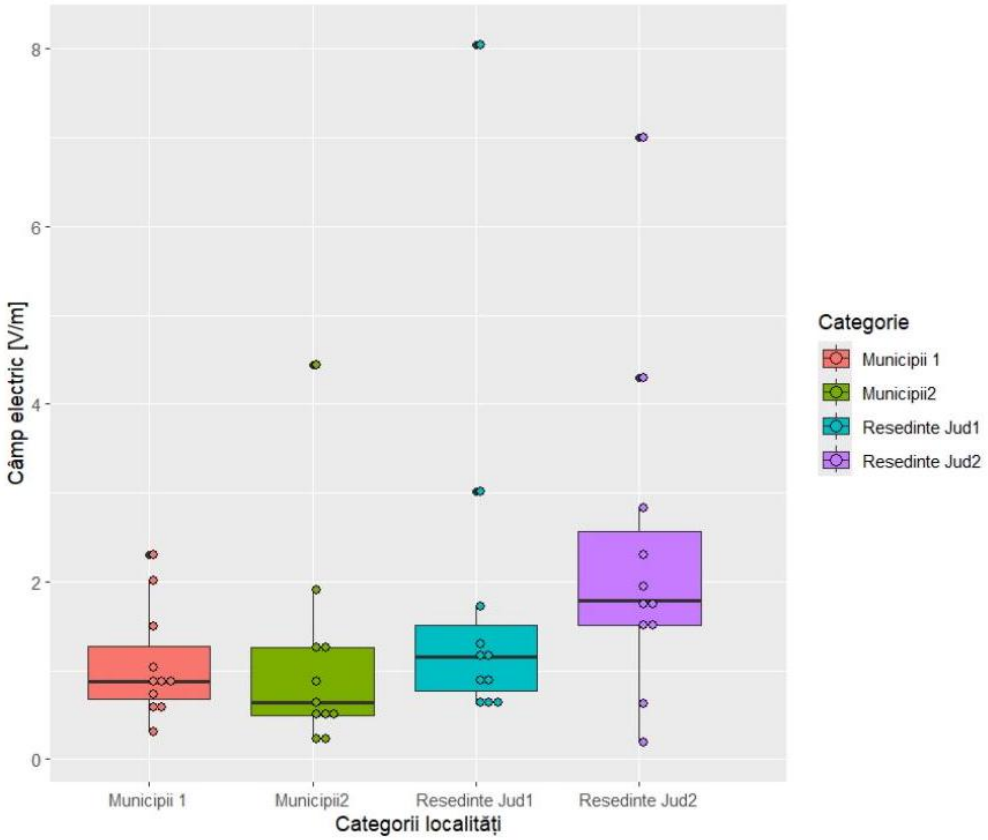


Figure 7: Boxplot of CE levels by locality groups

3.2 Hypothesis testing

In the statistical analysis by groups of localities, we started from the hypothesis of the existence of a significant difference between the means and variances for the 4 categories of localities, these data being included in the Table 3.

To test the statistical hypotheses, I decided to use the paired T-test on two groups of localities to verify the null hypothesis "There are no significant differences between the means of the two groups". I used the specialized R program

Table 3: Means and Variance by locality groups

	Mun 1	Mun 2	Res 1	Res2
AVG	1,06	1,12	1,83	2,34
VAR	0,06	0,28	0,27	0,41

for statistical studies, available in the free edition, and after running the test with the data presented above, I obtained the values from Table 4

Table 4: Results of paired T-tests for groups of localities

Tested groups	**test T**	**df**	**p -Val**
Mun1 vs Mun2	-0,15	14,93	0,88
Mun1 vs Res1	-1,13	11,63	0.28
Mun1 vs Res2	-2.13	12,15	0,05
Mun2 vs Res1	-0,94	15,69	0.36
Mun2 vs Res2	-1,80	17,07	0.09
Res1 vs Res2	-0,59	19,61	0.56

It can be seen that the p-values of the 6 tests are greater than the 0.05 level of significance and the p-value of 0.8832 from the first test suggests that there is an approximately 88.32 % probability that the observed difference is due to chance, in other words we can say we don't have enough evidence to reject the null hypothesis.

Another variant of statistical testing of the data from the 4 groups was by using the multiple paired T-test. After running this test using the Benjamini and Hochberg method, we obtained the results from Table 5.

Table 5: Multiple paired T-test results

-	Mun1	Mun2	Res1
Mun2	0,93	-	-
Res1	0,46	0,46	-
Res2	0,24	0,24	0,55

Table 5 contains only the p-values calculated for a significance level of 0.05 which apparently differ from the p-values calculated with the simple paired T-test but lead us to the same conclusion that we do not have enough evidence to reject the null hypothesis.

To deepen the statistical research, we also ran on the same data set one-way ANOVA tests as an alternative to the multiple paired T-test, Bartlett test to check the homogeneity of variances, the Shapiro-Wilk test to check the data distribution and a non-parametric test, the Kruskal-Wallis test to improve the result the ANOVA test.

The results of running these tests in R are shown in Tables 6, 7, 8 and 9.

Table 6: Rezults of ANOVA unidirecțional test means 4 groups

-	SumSq	MeanSq	Fvalue	**Pr(¿F)**
Variab	12,28	4,09	1,61	**0,20**
Resid	101,5	2,53	-	-

Table 7: Rezults of Bartlett test of variances for 4 groups

Bartlett test of homogenity of variances		
Date : Mun1, Mun2, Res1, Res2		
K-squared=**14.004**	df=**3**	p-value =**0.0029**

Table 8: Rezults of Shapiro-Wilk normality test

Shapiro-Wilk normality test	
Date : Reziduale test ANOVA	
W = **0.73848**	p-value = **1.701e-07**

Table 9: Rezults of Kruskal-Wallis rank sum test

Kruskal-Wallis rank sum test		
Date : groups means Mun1, Mun2, Res1, Res2		
K-W chi-sq = **6.98**	df = **3**	p-value = **0.073**

For the ANOVA test, as in the case of the T-tests, and in this case the null hypothesis states that there are no statistically significant differences between the means of the four groups of localities. A p-value of 0.202 obtained in this test convinces us that we cannot reject this hypothesis. To validate this model we used the graphical method using the Residuals vs Fitted and Q-Q graphs and when analyzing them we observed an inhomogeneity of the variance which led to the conclusion that the ANOVA test is not the most suitable for analyzing the data underlying this study.

By running the Bartlett test we obtained a result (p-value 0.0029) that confirms the inhomogeneity of the variance.

Next we checked the data for distribution using the Shapiro-Wilk test to check whether it follows a normal distribution or not. The result of this test confirmed beyond any doubt that the data does not follow a normal distribution (the p-value of the test was 5 orders of magnitude lower than the classical significance level)

The complementary, non-parametric test recommended to be used in cases where ANOVA is not suitable is the Kruskal-Wallis test, which verifies the homogeneity of the medians. After running this test, the obtained p-value confirms the fit of the null hypothesis of homogeneity of the medians.

3.3 Multivariate analysis

Another statistical approach to the data I made using multivariate analysis, trying to make an inference on the average value of CE according to 6 variables of 44 observations related to localities. We considered the response variable to be the average annual value of the electric field measured in broadband, and as predictor variables we used the average annual temperature, the number of inhabitants, the average annual value of the peak values of the electric field also measured in broadband as well as the values annual averages of the electric field measured in the three bands used in mobile communications. I also performed the analysis with the R program, the first step being the calculation of the correlation matrix between the variables and the displayed result is shown in Figure 8.

Although the data structure used in the paper is not complex, we also considered useful a principal components analysis (**PCA**) to verify the possibility of simplifying the analysis and even the possibility of visualizing the data in two or three dimensions. After running the PCA analysis in R we obtained the data on the proportion of variability of each principal component in the general regression model.

Thus, as can be seen in Figure 9, the first 4 components explain a proportion of 95 % of the variation of the regression model. Also, to verify the structure of each main component, we ran the PCA analysis in R and obtained the structure for Component 1 that explains the most variability of the regression model.

On the graph in Figure 10, you can see the weights of the predictor variables

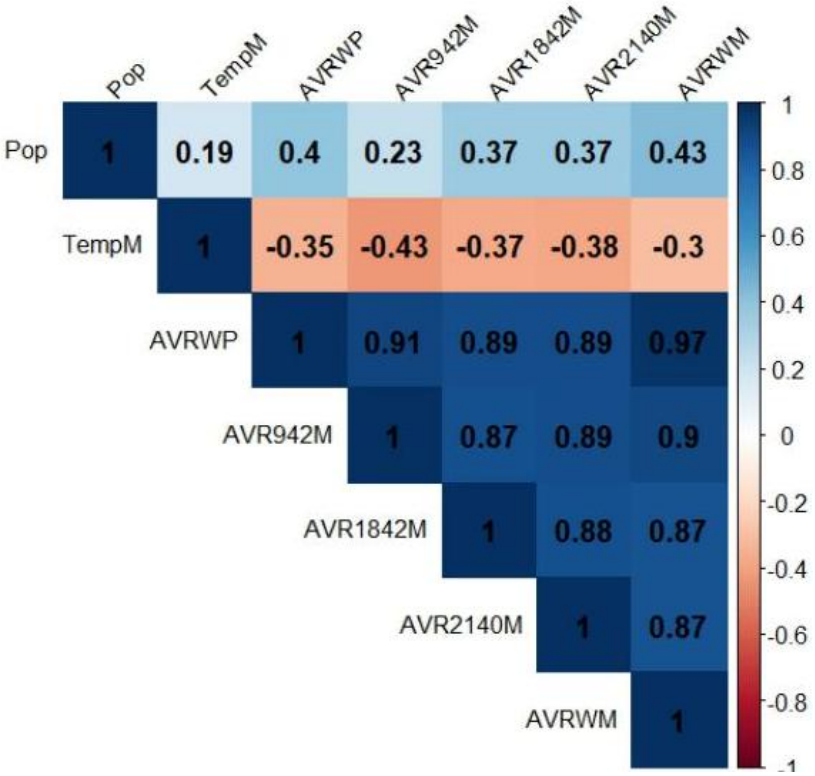


Figure 8: Spearman‘s corelation matrix

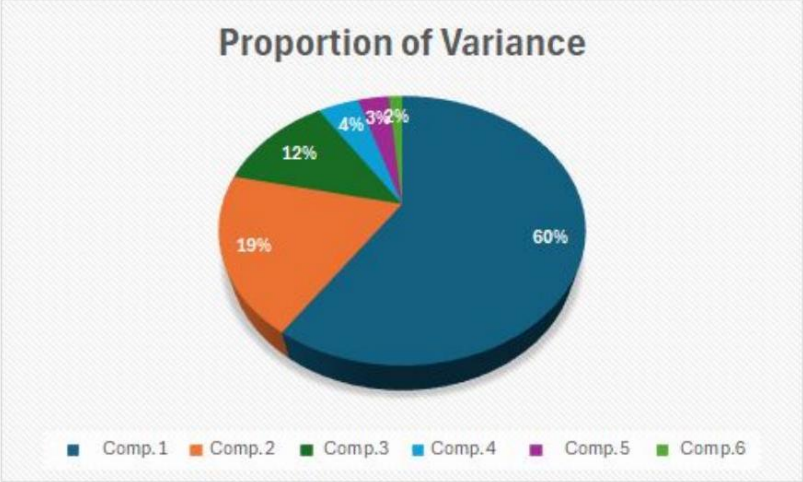


Figure 9: Proportion of Variance

within it.

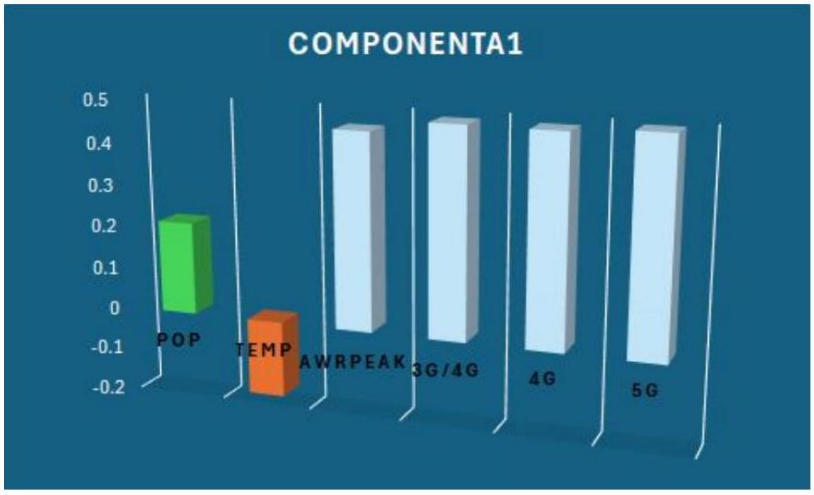


Figure 10: Component 1 structure of variables

The final step of the PCA analysis was to determine a multiple linear relationship between the principal components and the response variable **AVRWM**. The ***lm*** (linear model) function in R is used to fit linear regression models, and when applied to principal components, the goal is to use these components (which are linear combinations of the variables originals) as predictors in a regression model.

The result obtained after running this analysis in R confirmed the futility of using component 6 in building the model, the p-value of the T test related to it being higher than the level of significance, for which we removed it from the model and the result obtained was that of Table 10

Through the prism of comparing the p-values, a fit of the model is clearly observed, a confirmed situation and both the value close to the maximum for the adjusted R-squared coefficient and the p-value of the F-test.

4 Discussions

It can be seen on the graph in the figure 7 the average value of the electric field is slightly higher for the category of localities with a larger number of inhabitants (Residence Jud 2) and for the other groups the average values are quite close. There are two possible explanations for this situation, both of which are based on the characteristics of the sources that generate the electromagnetic field.

The first explanation is related to the fact that an increased population density is positively correlated with an increased number of mobile communications and this can only be achieved through a higher density of mobile phone base stations.

Table 10: Linear PCA model optimized for 5 components

Residuals:				
Min	**1Q**	**Median**	**3Q**	**Max**
-0.41706	-0.04321	0.00881	0.06202	0.51178
Coefficients:				
-	**Estimate**	**Std. Err**	**t value**	**Pr(¿—t—)**
(Intercept)	1.63179	0.02166	75.351	¡ 2e-16
Comp.1	0.71881	0.01147	62.683	¡ 2e-16
Comp.2	0.41357	0.02022	20.451	¡ 2e-16
Comp.3	0.34592	0.02514	13.762	2.4e-16
Comp.4	-1.20084	0.04234	-28.362	¡ 2e-16
Comp.5	-0.16913	0.04865	-3.476	0.0013
Residual standard error: 0.1436 on 38 degrees of freedom				
Multiple R-squared: 0.993–Adjusted R-squared: 0.992				
F-statistic: 1071 on 5 and 38 DF, p-value: ¡ 2.2e-16				

An increase in the number of mobile stations, on the other hand, automatically leads to an increase in electromagnetic noise within the radius of the respective localities. From the analysis of the localities, most of those with a higher level of the electric field have a high population density and are university centers (Bucharest, Iași, Brașov, Sibiu, Constanța, Timișoara).

The second explanation for a higher measured electric field level is related to the existence in or near the towns of high power radio/TV transmitters, which generate higher electric field levels. These transmitters were initially installed (‘50-‘70 years) outside the towns, but with the expansion of the residential areas they were included inside them and therefore the respective areas have a higher level of electromagnetic pollution (Timișoara, Sibiu, Bucharest, Iași)

The inconclusive results of the paired **T** and **ANOVA** tests were due to the fact that the distribution of the data was not suitable for the use of the first two. Finally, the verification of the normality of the data distribution with the **S-W** test and then of the homogeneity of the medians with the **K-W** test supported the hypothesis regarding the homogeneity of the average values of the electric field level in the case of the four groups of localities.

We considered useful the multivariate analysis using as response variable the average value of the electric field measured in broadband because it best reflects the level of exposure of the population to electromagnetic noxes, being taken as a reference in the case of the analysis of the level of exposure in accordance with the legislation in the field .

From the perspective of this analysis, the results of the calculation of the correlation matrix 8 lead us to the conclusion of the existence of a strong link between the response variable **AVRWM** and the predictor variable **AVRWP** and weaker links between the variables **AVR942M**, **AVR1842M**, **AVR2140M**

and **AVRWM**.

The situation can be explained by the fact that in the composition of the average value of the broadband electric field, the electric field values from the three bands allocated to mobile communications also have a significant weight. The correlation values between the predictor variables related to the electric field values can be explained by the fact that the emission levels of mobile communication base stations are relatively similar. We also observe a weak negative correlation between the variable temperature and the variables of the electric field values, explained by the fact that the propagation of electromagnetic waves is favored by low temperatures.

From Table 9 we can conclude that the first four components explain 95 % of the variability of the model and therefore they are sufficient for further analysis. Component 1 has the largest weight in explaining the variability of the linear model and within it, following the PCA analysis, we noticed that the variables containing average values of the peak electric field and on frequency bands have a similar but higher weight than the average temperature values or population density which comes as a confirmation of the interpretation of the correlation matrix results.

Although from the point of view of the results obtained after running the PC analysis in R, the model with 6 principal components is similar to the one with 5 components in practice simpler models are preferred for possible graphical processing or for the simplicity of the analysis. The efficiency of the PCA analysis can best be seen by comparing it with the model obtained from the linear analysis with the predictor variables in this latter case only two variables passed the T-test and could enter the model, a situation in which the model does not show confidence.

5 Summary and conclusions

Data were collected on the levels of exposure of the population to the electromagnetic field in the period September 2022-September 2023 from 44 localities in Romania spread over the entire national geographical area. Annual average values of the electric field level measured in broadband (100KHz-7GHz) and in three bands used by mobile communication operators (925-960MHz, 1805-1880MHz, 2110-2170MHz) were processed in order to use them for various analyzes statistics.

For the first part of the study, the data were divided into four groups selected on the basis of population density, and statistical tests were performed to verify the hypothesis that there are no statistically significant differences between the four means. The paired T, multiple paired T, ANOVA, Bartlett, Shapiro-Wilk, Kruskal-Wallis tests were run in R and their results confirm the homogeneity of

means, medians and the fact that the data do not have a normal distribution.

In the second part of the study, we did a multivariate analysis with 5 predictor variables containing observations from all 44 localities to obtain a model of variability of the average electric field level in broadband. Following the PCA analysis, we obtained a model that explains 99 % of this variability for the respective cities, the prediction error being very small.

From the point of view of the levels of exposure to electromagnetic pollution of the population at national level, the average value for the 44 localities represents a percentage of 5.78 % of the maximum limit allowed in the legislation (27.5 V/m), the minimum values and maximum exposure being recorded in Tulcea (0.73 %) respectively in Sibiu (29.27 %).

At the present moment, the implementation of 5G technology in Romania is in its early stages, being available only in large towns, the frequencies proposed for use being available on the ANCOM [12] page, but even so it can be stated based on the analysis in this study that no has a significant impact on the share of electromagnetic pollution in the respective localities, so the fears induced to the population through the mass media are unjustified.

As wireless technology develops by leaps and bounds, it is necessary to develop the electromagnetic nox monitoring systems and to continue studies in this field in the future in order to provide safety among the population regarding the correct use of new technologies without affecting the state of health .

Aknowlegements

Thanks to Mr. Conf.univ.dr. **Romeo Negrea** for the guidance and support offered in carrying out the study and Mr. ș.l.dr.eng. **Teodor Petrița** for the support provided in obtaining and processing the data taken from the ANCOM sensors.

References

[1] CE EF. Electromagnetic fields-report -november 2006. https: //ec.europa.eu /health/ ph-determinants /environment/EMF/ ebs272a en.pdf, 2023. Accesat: 2024-05-30.

[2] International Agency for Research on Cancer. Iarc monographs on the evaluation of carcinogenic risks to humans. http://monographs.iarc.fr/ENG/ Monographs/ vol102/ mono102.pdf [PMCfreearticle] [PubMed] [GoogleScholar], 2013. Accesat: 2024-05-21.

[3] Hardell L. Carlberg M. Decreased survival of glioma patients with astrocytoma grade iv (glioblastoma multiforme) associated with long-term use of mobile and cordless phones. Int. J. Environ. Res. Public Health, 2014.

[4] Sultan Ayoub Meo, Yazeed Alsubaie, Zaid Almubarak, Hisham Almutawa, Yazeed AlQasem, and Rana Muhammed Hasanato. Association of exposure to radio-frequency electromagnetic field radiation (rf-emfr) generated by mobile phone base stations with glycated hemoglobin (hba1c) and risk of type 2 diabetes mellitus. International Journal of Environmental Research and Public Health, 2015.

[5] Nicolas Loizeau, Marco Zahner, Johannes Schindler, Christa Stephan, Jürg Fröhlich, Markus Gugler, Toni Ziegler, and Martin Röösli. Comparison of ambient radiofrequency electromagnetic field (rf-emf) levels in outdoor areas and public transport in switzerland in 2014 and 2021. Environmental Research, 237:116921, 2023.

[6] Nikola Djuric, Dragan Kljajic, Teodora Gavrilov, Vidak Otasevic, and Snezana Djuric. The emf exposure monitoring in cellular networks by serbian emf ratel system. In 2022 IEEE International Symposium on Measurements and Networking, pages 1-6, 2022.

[7] Eduard Lunca and Alexandru Salceanu. An overview of rf-emf monitoring systems and associated monitoring data. In 2022 IEEE International Symposium on Measurements and Networking, 102016.

[8] ANCOM. Piața serviciilor de comunicații electronice din românia, raport de date statistice -semestrul i 2023. https://statistica. ancom.ro/sscpds/public/files/272-ro, 2023. Accesat: 202405-1.

[9] ANCOM. Măsurări cu senzori emf plasați în poziții fixe. https: // www. monitor-emf.ro/map, 2024. Accesat: 2024-05-20.

[10] INS. Recensământul populației 2021. https:// www.recensamant romania.ro/ rezultate-rpl-2021/ rezultate-definitive/, 2021. Accesat: 2024-03-10.

[11] MS. Ordin nr. 1193 din 29 septembrie 2006 pentru aprobarea normelor privind limitarea expunerii populaţiei generale la câmpuri electromagnetice de la 0 hz la 300 ghz. https://www.ancom.ro/uploads/links files/Odinul-1193-2006 norme.pdf, 2006. Accesat: 2024-05-1.

[12] ANCOM. Utilizarea serviciilor comerciale 5 g în românia. https: //info-centru.ancom.ro/, 2024. Accesat: 2024-06-3.

Todor Ioan Dorel – MSc student, Department of Mathematics,

Politehnica University of Timişoara,

P-ta Victoriei 2, 300 006, Timişoara, ROMANIA

E-mail: ioan.todor@student.upt.ro

SCIENTIFIC BULLETIN
OF
THE POLITEHNICA UNIVERSITY OF TIMISOARA

Vol.68(82)	MATHEMATICS–PHISYCS	2023

ANALYSIS OF STATISTICAL METHODS APPLIED IN THE STUDY OF CARDIAC PATHOLOGY

Miruna Cristina BĂTRÎNESCU and Dr. Cristina VĂCĂRESCU

Abstract

This article explores the application of correlation and regression statistical methods in cardiac pathology. The aim of this study is to identify and analyze the relationships between various cardiac parameters to improve the understanding and management of heart diseases.

The research utilizes a comprehensive dataset, including variables such as ejection fraction, age, QRS duration, end-diastolic volume, end-systolic volume, interventricular septum thickness, left ventricular end-diastolic diameter, systolic pulmonary artery pressure, and left atrial volume. The analysis focuses on descriptive statistics, correlation matrices, and regression models to highlight significant associations among these parameters.

Key findings reveal strong negative correlations between ejection fraction and both end-diastolic and end-systolic volumes, indicating that higher ejection fractions are associated with smaller volumes. Additionally, the study identifies moderate positive correlations between QRS duration and both interventricular septum thickness and left ventricular end-diastolic diameter. These relationships emphasize the interdependence between the heart's electrical and structural characteristics.

The research also employs graphical visualizations to facilitate the interpretation and communication of results, thereby contributing to a better understanding of the complexity of cardiac pathology and improving diagnostic and therapeutic approaches in this field.

Keywords and phrases:*cardiac pathology, statistical correlation, regression, ejection fraction*

1 Introduction

Cardiac pathology is a leading cause of morbidity and mortality globally, affecting millions and significantly impacting quality of life. Understanding the mechanisms and factors behind heart disease progression is crucial.

This study aims to explore the relationships between various cardiac parameters using correlation and regression statistical methods. By analyzing variables such as ejection fraction, QRS duration, and ventricular volumes, we seek to identify significant associations that can aid in the diagnosis and treatment of heart diseases.

The article first reviews the anatomy and physiology of the heart, congenital conditions, ischemic diseases, arrhythmias, and heart failure. It also covers diagnostic methods, evaluation strategies, and treatment and prevention options. The second part details the statistical methods used, including descriptive analysis and correlation and regression models.

Our goal is to provide insights into the interdependencies between cardiac parameters, enhancing clinical approaches and outcomes for patients. Rigorous statistical analysis highlights significant variable relationships, supporting clinical decisions and personalized therapeutic strategies.

2 Methods

2.1 Descriptive Statistical Methods Applied to the Dataset

We employed descriptive statistical methods to analyze the dataset, providing a clear overview of data distribution and characteristics.

Data Collection: Dr. Cristina Văcărescu collected data from 54 patients at the Institute of Cardiovascular Diseases in Timișoara. The patients were studied over several years to examine heart failure reduction through cardiac resynchronization therapy.

Data Analysis: The R software system was used for statistical analysis. The following methods were applied:

- Central Tendency Indicators: Mean, median, and mode were calculated to describe central tendencies.

- Dispersion Indicators: Standard deviation, quartiles, and deciles were measured to assess data spread. Correlation Analysis:

Pearson Correlation Coefficient: Used to measure the linear relationship between pairs of variables. The correlation matrix was computed to identify significant associations.

Regression Analysis:

Linear Regression: Applied to investigate the relationships between dependent and independent variables. The model included predictors such as QRS duration, end-diastolic volume, and end-systolic volume.

Multinomial Regression: Utilized to analyze risk factors associated with hypertension (HTN) considering variables such as sex, age, chronic kidney disease (CKD), and cardiomyopathy etiology.

2.2 Software Tools

The analysis was performed using the R statistical software. Specific packages used include: dplyr for data manipulation, ggplot2 for creating detailed and customizable data visualizations, readxl for reading Excel files.

Each step of the analysis was thoroughly documented, and graphical visualizations were generated to facilitate the interpretation of the results. The following sections will present these results along with the corresponding visualizations.

3 Results and Discussions

3.1 Central Tendency Indicators

The first step, presented in Figure 1, involved calculating the mean age of the patients using the mean function. The result was an average age of 62.34 years, displayed using the print function, indicating the average age value for the entire dataset.

```
> view(date)
> # varsta pacientilor
> date <- c(42, 33, 65, 62, 60, 71, 53, 63, 61, 62, 58, 58, 73, 66, 66, 59, 72, 60, 58, 65, 53, 80, 39, 60, 66, 61, 50, 84, 68, 7
4, 55, 63, 53, 52, 65, 40, 73, 74, 62, 78, 72, 60, 64, 59, 57, 58, 58, 75, 88, 69)
> date
 [1] 42 33 65 62 60 71 53 63 61 62 58 58 73 66 66 59 72 60 58 65 53 80 39 60 66 61 50 84 68 74 55 63 53 52 65 40 73 74 62 78
[41] 72 60 64 59 57 58 58 75 88 69
> # Calculul media varstei pacientilor --> este 62,34 ani
> media <- mean(date)
> print(media)
[1] 62.34
```

Figure 1: Calculating the mean age of the patients using the mean function

The next step, presented in Figure 1, was calculating the median age of the patients using the median function. The result was a median of 62 years, indicating the central value of the dataset when the values are arranged in ascending order. Next, Figure 2 illustrates the calculation of the mode of the patients' ages

```
> # Calculul medianei --> este 62
> mediana <- median(date)
> print(mediana)
[1] 62
```

Figure 2: Calculation of the patients' median age using the median function in the R system.

using the Mode function. The result indicated that the most frequent age among the patients is 58 years, appearing 5 times in the dataset, thus highlighting the most commonly encountered age within the dataset.

3.2 Dispersion Indicators for Numerical Variables

Standard deviation is a measure of data dispersion around the mean. For the age of patients, the standard deviation is approximately 10.89, which means that the age values are spread around the mean by this amount. The calculation of the standard deviation is presented in Figure 3.

```
> dev_standard <- sd(date)
> print(dev_standard)
[1] 10.89019
```

Figure 3: Calculation of the standard deviation for the patients' age in the R system

3.3 Dispersion Indicators for Ordinal Variables

Quartiles divide the data distribution into four equal parts, as shown in Figure 4. The first quartile (Q1) is 58, the median (Q2) is 62, and the third quartile (Q3) is 68.75. These values indicate the lower and upper bounds of the first 25%, 50% and 75% of the dataset, respectively. Deciles divide the data distribution into ten

```
> quartile <- quantile(date, probs = c(0.25, 0.5, 0.75))
> print(quartile)
  25%   50%   75%
58.00 62.00 68.75
```

Figure 4: Calculation of quartiles in the R system.

equal parts, as shown in Figure 5. These values mark the percentiles from 10% to 90%. For example, 10% of the patients have an age less than or equal to 51.8, 20% have an age less than or equal to 56.6, and so on.

```
> decile <- quantile(date, probs = seq(0.1, 0.9, by = 0.1))
> print(decile)
 10%  20%  30%  40%  50%  60%  70%  80%  90%
51.8 56.6 58.0 60.0 62.0 64.4 66.0 72.0 74.1
```

Figure 5: Calculation of deciles in the R system.

3.4 Indicators for Qualitative Variables

We began by defining the variable "NYHA", which contains the categories for which we wanted to calculate proportions and represent them in a chart.

We calculated the proportions of each category in the "NYHA" variable using the table function (Figure 5), which counts the occurrences of each category in

the dataset. We then divided this count by the total length of the vector to obtain the proportion.

For representing the data in a chart, we used the pie function. We included the main argument to add a title to the chart and col to specify colors for each category. We added the legend to the upper right side of the chart to indicate the correspondence between the colors and the categories in the variable using the legend function.

```
> NYHA <- c("III", "III", "III", "III", "III", "III", "III", "III", "III", "III", "III", "III", "III", "III", "III", "III", "III", "III", "II
I", "III", "II", "II", "II", "II", "II", "III", "II", "II", "II", "II", "II", "II", "II", "II", "III", "II", "II", "II", "II", "II", "III", "I
I", "II", "II", "II", "II", "II", "II", "III", "III", "II", "III", "III", "II")
> # Calculul proporțiilor pentru fiecare categorie
> proportii <- table(NYHA) / length(NYHA)
> # Reprezentarea proporțiilor sub formă de diagramă tip pie
> pie(proportii, main = "Proportia valorilor NYHA", col = rainbow(length(proportii)))
> legend("topright", names(proportii), fill = rainbow(length(proportii)))
```

Figure 6: Calculation of the Proportions for Each Category in the "NYHA" Variable Using the table Function.

Therefore, we successfully created a chart that visually represents the proportions of each category in the "NYHA" variable in a clear and efficient manner.

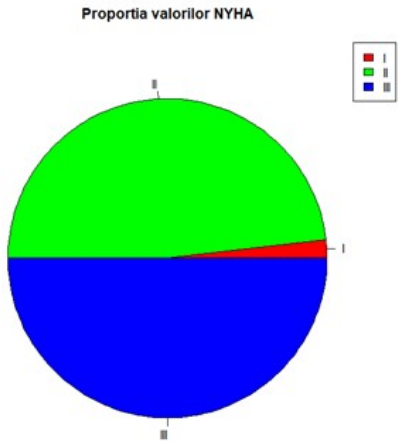


Figure 7: The proportion of NYHA values

From the chart, we observe that the majority of patients fall into the NYHA III category, represented by a large sector of the chart. This suggests that most patients in the dataset have moderate to severe heart failure symptoms according to the NYHA classification.

On the other hand, we see that there is only one patient in the NYHA I category, represented by a very small sector of the chart. This indicates that the

number of patients with mild heart failure symptoms is much smaller compared to those with moderate or severe symptoms.

Next, we chose to represent the distribution of mitral regurgitation grades of the patients recorded in the dataset through another chart. We began by defining the variable "RM" (Figure 8), which contains the mitral regurgitation grades for each patient in the dataset. Then, we used the table function to count the occurrences of each mitral regurgitation grade in the dataset. Subsequently, we used the barplot function to represent the frequency distribution as a bar chart. We specified the arguments main, xlab, and ylab to add a title and labels on the x and y axes, and the col argument to specify the color of the bars.

```
> frecventa <- table(RM)
> # Crearea barplot pentru variabila "RM"
> barplot(frecventa, main = "Distribuția Regurgitării Mitrale", xlab = "Clase RM", ylab = "Frecvență", col = "skyblue")
```

Figure 8: Calculating to find the distribution of Mitral Regurgitation grades in the R system

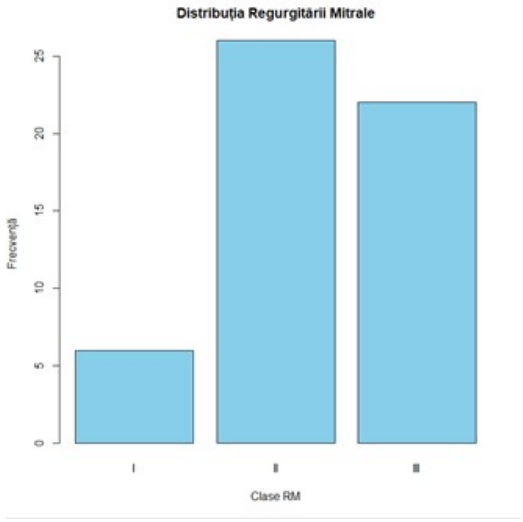


Figure 9: Distribution of Mitral Regurgitation

The chart presented in Figure 9 indicates that the majority of patients fall into grade 2 of mitral regurgitation, followed by grade 3, while the smallest number of patients are in grade 1.

We continued by creating a graph in R for tricuspid regurgitation (Figure 10). We entered the observed values for tricuspid regurgitation into a vector and calculated the occurrences of each category using the table function. Subsequently,

we created the graph using the barplot function, specifying the title, labels for the x and y axes, and the color of the bars.

```
> regurgitare_tricuspidiana <- c("I", "I", "I", "II", "I", "II", "II", "I", "II", "III", "I", "I", "I", "III", "I", "III", "II", "I", "I", "I",
 "II", "II", "I", "I", "II", "II", "I", "II", "I", "I", "II", "I", "II", "II", "II", "III", "II", "II", "II", "II", "III", "II", "I", "II",
 "I", "III", "II", "II", "III", "III", "I", "I", "II", "III")
> frecventa_regurgitare_tricuspidiana <- table(regurgitare_tricuspidiana)
> barplot(frecventa_regurgitare_tricuspidiana, main = "Distribuția Regurgitării Tricuspidiene", xlab = "Grad de regurgitare", ylab = "Număr de
 pacienți", col = "red")
```

Figure 10: Calculation for determining the distribution of tricuspid regurgitation grades in the R system.

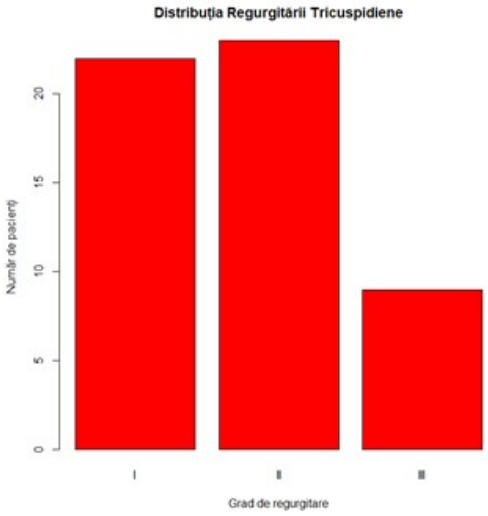


Figure 11: Distribution of Tricuspid Regurgitation in the R system.

From the analysis of the chart presented in Figure 11, it can be observed that the majority of patients fall into grade 2 of tricuspid regurgitation, suggesting that this severity grade is the most common in the dataset. Following this, patients in grade 1 are the next most frequent, with a smaller proportion of patients falling into grade 3. This distribution suggests that tricuspid regurgitation is more often encountered at a moderate severity level than at severe or very severe levels.

3.5

Distribution of Patients with Dilated Cardiomyopathy: Idiopathic vs. Ischemic Etiology

First, I read the data from the file using the read-excel function from the readxl package. I stored the data in the variable data. Then, I filtered the data to retain only the rows where the column "Etiologie CMD" has the values "idiopatica" or

"ischemica". I achieved this using the filter function from the dplyr package and stored the result in the variable filtered-data. The next step was to count each etiology type in filtered-data using the count function from dplyr. The result is stored in the variable etiologie-counts.

To add percentages to the chart, I used the mutate function from dplyr to create a new column percent in etiologie-counts. This column calculates the percentage of each etiology type out of the total number of patients by dividing the count of each etiology type by the total sum and multiplying the result by 100.

To create the chart, I used the ggplot2 package. I started with the ggplot function, specifying etiologie-counts as the dataset and defining aesthetics (aes) to use the number of patients (n) and the type of etiology for filling the chart. I used geom-bar to create a bar chart with a width of 1 and coord-polar to transform the bar into a pie chart.

I used theme-void to remove non-essential chart elements and theme to hide the legend title. I added percentage labels using geom-text, positioning them in the middle of each pie chart segment. The labels consist of percentages rounded to one decimal place.

I added a title to the chart using ggtitle. I specified the colors for each pie chart segment using scale-fill-manual with two colors: blue for "idiopatica" and red for "ischemica", as shown in Figure 12.

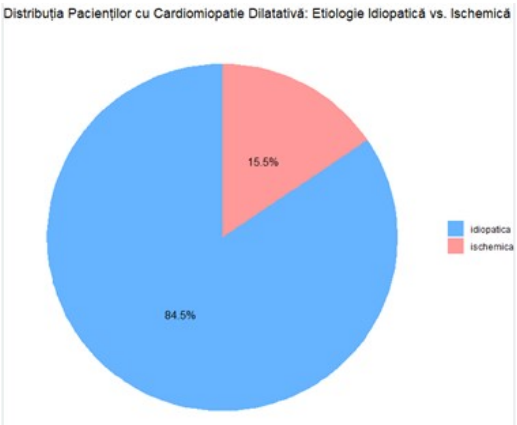


Figure 12: Distribution of Patients with Dilated Cardiomyopathy.

In the blue segment of Figure 12, representing idiopathic etiology, a percentage of 84.5% is observed. In the red segment, representing ischemic etiology, a percentage of 15.5% is observed.

3.6 Creating Pie Charts for Chronic Disease Analysis in R.

I commenced by installing and loading the necessary packages. These include "dplyr" for data manipulation, "ggplot2" for chart creation, and "magrittr" for utilizing the pipe operator. This was accomplished using the commands install. packages ("dplyr"), install.packages("ggplot2"), and install.packages ("magrittr"), followed by loading the packages with library(dplyr), library(ggplot2), and library(magrittr). After the installation and loading of these packages, I proceeded to import the data from the Excel file.

Subsequently, I ensured that the variables of interest were devoid of any missing values (Figure 13). This was achieved by replacing missing values with 0 for the variables "Diabet Zaharat", "Boală cronică de rinichi", and "HTA".

To represent the combinations of diseases for each patient, I introduced a new column in the dataset. This was done using the command data *Combination*, which amalgamates the information regarding each patient's diseases into a single column.

The following step involved substituting any missing values in the "Combination" column with "Niciuna". This step ensures that records with no associated diseases are appropriately marked (Figure 13) To exclude records devoid of any

```
> filtered_data <- data %>%
+   filter(Combination != "Niciuna" & Combination != "")
```

Figure 13: This step removes rows that do not contain any diseases

disease, specifically those with missing values, I filtered the dataset using the filtered-data command. This process eliminates rows that do not indicate any diseases (Figure 14).

To construct the pie charts, I employed the sophisticated capabilities of the ggplot2 package. The detailed implementation of this function is illustrated in Figure 14. The chart presented in Figure 15 illustrates the distribution of patients based on the presence of Diabetes Mellitus. This diagram is divided into two distinct sections, each representing the proportion of patients with and without Diabetes Mellitus.

The blue section, representing 62.5%, indicates the proportion of patients who have Diabetes Mellitus. This percentage suggests that a significant majority of the patients in the dataset are diagnosed with this condition. On the other hand, the red section, representing 37.5%, shows the proportion of patients who do not have Diabetes Mellitus. This indicates that a smaller portion of the patients in the dataset do not suffer from this disease. This pie chart serves as an effective tool for swiftly and clearly visualizing the distribution of patients based on the presence of Diabetes Mellitus.

Additionally, I employed the function depicted in Figure 15 to generate pie

```
> create_pie_chart <- function(variable, title) {
+   df <- filtered_data %>%
+     group_by(!!sym(variable)) %>%
+     summarise(count = n()) %>%
+     mutate(percentage = count / sum(count) * 100,
+            label = paste0(round(percentage, 1), "%"))
+
+   ggplot(df, aes(x="", y=percentage, fill=factor(!!sym(variable)))) +
+     geom_bar(stat="identity", width=1) +
+     coord_polar("y", start=0) +
+     theme_void() +
+     labs(title=title, fill=variable) +
+     geom_text(aes(label=label), position=position_stack(vjust=0.5))
+ }
> # Pie chart pentru Diabet Zaharat
> create_pie_chart("DZ", "Distribuția Diabet Zaharat")
```

Figure 14: To create the pie charts, I used the ggplot2 function.

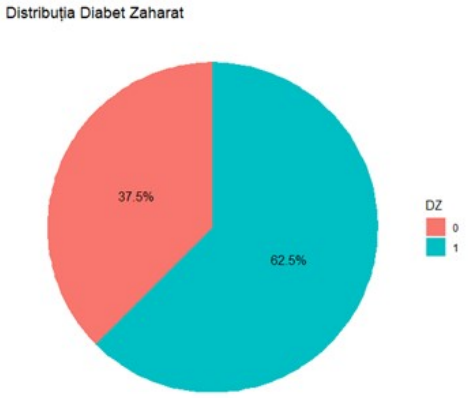


Figure 15: This chart is divided into two distinct sections, each representing the proportion of patients with and without Diabetes Mellitus.

charts for the variables Chronic Kidney Disease (CKD) and Hypertension (HTN). The pie chart presented in Figure 17 illustrates the distribution of patients based

```
> # Pie chart pentru Boală cronică de rinichi
> create_pie_chart("BCR", "Distribuția Boală cronică de rinichi")
```

Figure 16: Creating a pie chart for the variable BCR.

on the presence of Chronic Kidney Disease (CKD). This chart is divided into two distinct sections, each representing the proportion of patients with and without Chronic Kidney Disease. The blue section, representing 65%, indicates the proportion of patients with Chronic Kidney Disease. This percentage suggests that a significant majority of patients in the dataset are diagnosed with this condition.

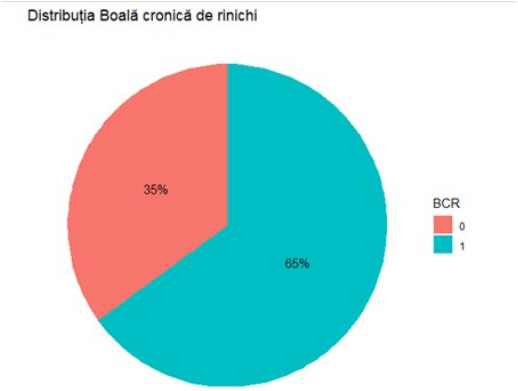


Figure 17: The chart is divided into two distinct sections, each representing the proportion of patients with and without BCR.

Conversely, the red section, representing 35%, shows the proportion of patients without Chronic Kidney Disease. This indicates that a considerable portion of patients in the dataset do not suffer from this disease.

This pie chart is useful for quickly and clearly visualizing how the patient set is divided based on the presence of Chronic Kidney Disease. The pie chart

```
> # Pie chart pentru HTA
> create_pie_chart("HTA", "Distribuția HTA")
```

Figure 18: Creating a pie chart for the variable HTA.

presented in Figure 19 illustrates the distribution of patients based on the presence of hypertension (HTN). This chart is divided into two distinct sections, each representing the proportion of patients with and without HTN.

The blue section, representing 42.5%, indicates the proportion of patients diagnosed with HTN. This percentage suggests that a significant portion of the patients in the dataset are diagnosed with this condition. Conversely, the red section, representing 57.5%, shows the proportion of patients without HTN. This indicates that the majority of the patients in the dataset do not suffer from hypertension.

This pie chart is an effective tool for swiftly and clearly visualizing the distribution of patients based on the presence of hypertension. Finally, for a pie chart representing multiple conditions, I defined a similar function to group and display the data from the "Combination" column. The steps are detailed in Figure 20. Subsequently, I generated the chart using the create-combination-pie-chart function. This sequence of commands produced a pie chart illustrating the distributions of Diabetes Mellitus, Chronic Kidney Disease, Hypertension, and their

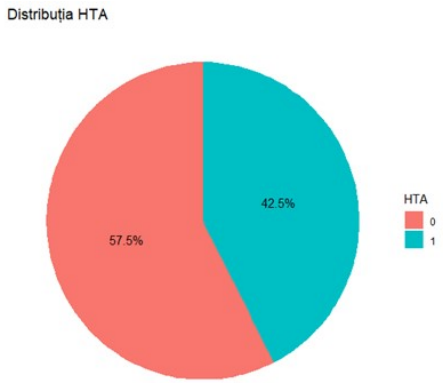


Figure 19: The distribution of patients with and without hypertension

```
> create_combination_pie_chart <- function() {
+   df <- filtered_data %>%
+     group_by(Combination) %>%
+     summarise(count = n()) %>%
+     mutate(percentage = count / sum(count) * 100,
+            label = paste0(round(percentage, 1), "%"))
+
+   ggplot(df, aes(x="", y=percentage, fill=Combination)) +
+     geom_bar(stat="identity", width=1) +
+     coord_polar("y", start=0) +
+     theme_void() +
+     labs(title="Distribuția Combinațiilor de Boli", fill="Combinații de Boli")
 +
+     geom_text(aes(label=label), position=position_stack(vjust=0.5))
+ }
> # Pie chart pentru combinațiile de boli
> create_combination_pie_chart()
```

Figure 20: Creating a chart for multiple conditions

combinations, while excluding patients without any conditions (Figure 21). The pie chart presented in Figure 21 illustrates the distribution of patients based on the multiple conditions they have.

- The red section indicates the proportion of patients with only Diabetes Mellitus (DM), representing 5% of the total patients.
- The orange section shows the proportion of patients with both Diabetes Mellitus (DM) and Hypertension (HTN), accounting for 10% of the total.
- The yellow section represents the proportion of patients with both Diabetes Mellitus (DM) and Chronic Kidney Disease (CKD), making up 12.5% of the total.
- The light green section indicates the proportion of patients with Diabetes Mellitus (DM), Chronic Kidney Disease (CKD), and Hypertension (HTN),

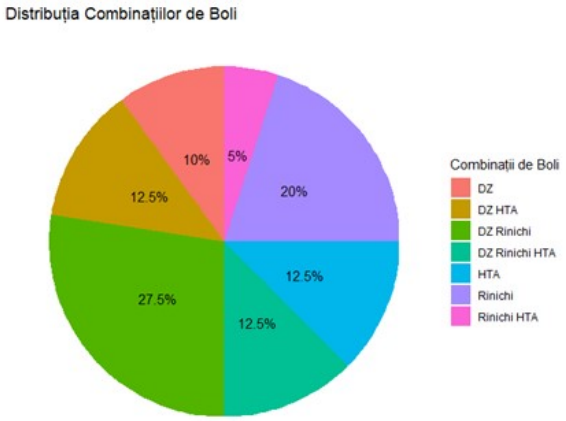


Figure 21: The pie chart based on various conditions

also representing 12.5% of the total.

- The green section shows the proportion of patients with only Hypertension (HTN), representing 27.5% of the total patients, which is the largest section, suggesting that hypertension is the most common condition among the studied patients.
- The light blue section indicates the proportion of patients with only Chronic Kidney Disease (CKD), representing 12.5% of the total.
- The purple section represents the proportion of patients with both Chronic Kidney Disease (CKD) and Hypertension (HTN), making up 20% of the total.

Hypertension (HTN) is the most prevalent condition, with 27.5% of patients having only HTN, indicating that this condition is very common. A significant number of patients exhibit combinations of diseases, such as DM and HTN (10%), DM and CKD (12.5%), and the complete combination of DM, CKD, and HTN (12.5%). Additionally, 20% of patients have both Chronic Kidney Disease (CKD) and Hypertension (HTN), suggesting a correlation between these two conditions. This chart provides a clear visualization of the various pathologies present among the studied patients.

3.7 Correlation and Regression of Variables

Correlation and linear regression for HTN, QRS Duration, EF, LVDD, and LVSD

I began by ensuring that all variables needed for correlation calculation are numeric. In the dataset, some variables may be read as factors or characters,

so I transformed them into numbers using the as.numeric function. Each line of code below performs this conversion for a specific variable. After converting the variables, I checked the structure and content of the data to identify any issues, such as missing values indicated by NA or values that were not correctly converted.

We used the summary function to obtain a brief description of each variable in the dataset: The next step was to select only the variables of interest for correla-

```
> date$HTA <- as.numeric(date$HTA)
> date$VTD <- as.numeric(date$VTD)
> date$VTS <- as.numeric(date$VTS)
> date$Durata.QRS <- as.numeric(date$Durata.QRS)
> date$FE. <- as.numeric(date$FE.)
> summary(date)
```

Figure 22: Transformation of variables HTA, QRS Duration, VTD, and VTS for numeric correlation calculation.

tion calculation. I did this by creating a vector called vars containing the names of the desired variables and then using this vector to extract the respective columns from the original dataset. The result is a new dataset Data containing only the selected variables. Then I calculated the correlation matrix using the cor function. I specified use = "complete.obs" to ensure that only complete observations, i.e., those without missing values, are used, and method = "pearson" to use the Pearson correlation method. This method measures the degree of linear association between variables: Finally, I displayed the correlation matrix using the print

```
> vars <- c("HTA", "VTD", "VTS", "Durata.QRS", "FE.")
> data <- date[, vars]
> cor_matrix <- cor(data, use = "complete.obs", method = "pearson")
> print(cor_matrix)
                HTA         VTD         VTS  Durata.QRS         FE.
HTA         1.00000000 -0.22485260 -0.04510154  0.01512290  0.13887000
VTD        -0.22485260  1.00000000  0.67785849  0.07875215 -0.59923268
VTS        -0.04510154  0.67785849  1.00000000  0.03162304 -0.41931053
Durata.QRS  0.01512290  0.07875215  0.03162304  1.00000000 -0.08391294
FE.         0.13887000 -0.59923268 -0.41931053 -0.08391294  1.00000000
```

Figure 23: Calculation of the Correlation Matrix and Calculation Results.

function. This shows the correlation coefficient values for all pairs of selected variables. The correlation matrix shows the Pearson correlation coefficients between the selected variable pairs. The Pearson correlation coefficient ranges from -1 to 1 and measures the degree of linear association between two variables: 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation, i.e., no linear association. I then proceeded to analyze the results for each pair of variables. Arterial Hypertension (HTA) and the other variables show the following correlations: between HTA and VTD, the correlation

coefficient is -0.22485260, suggesting a weak negative correlation between hypertension and right atrium volume. As VTD increases, HTA may slightly decrease, but this association is weak. Between HTA and VTS, the correlation coefficient is -0.04510154, suggesting almost no correlation between HTA and left atrium volume. Between HTA and QRS Duration, the correlation coefficient is 0.01512290, suggesting almost no correlation between HTA and QRS duration. Between HTA and FE, the correlation coefficient is 0.13887000, suggesting a very weak positive correlation between HTA and ejection fraction. VTD (Right Atrial Volume) and the other variables show the following correlations: between VTD and VTS, the correlation coefficient is 0.67785849, indicating a moderate-strong positive correlation between right atrial volume and left atrial volume. This suggests that as right atrial volume increases, left atrial volume tends to increase. Between VTD and QRS Duration, the correlation coefficient is 0.07875215, suggesting a very weak positive correlation between right atrial volume and QRS duration. Between VTD and FE, the correlation coefficient is -0.59923268, indicating a moderate-strong negative correlation between right atrial volume and ejection fraction. As right atrial volume increases, ejection fraction tends to decrease. VTS (Left Atrial Volume) and the other variables show the following correlations: between VTS and QRS Duration, the correlation coefficient is 0.03162304, suggesting almost no correlation between left atrial volume and QRS duration. Between VTS and FE, the correlation coefficient is -0.41931053, indicating a moderate negative correlation between left atrial volume and ejection fraction. As left atrial volume increases, ejection fraction tends to decrease. Regarding QRS duration and the other variables, the correlation coefficient between QRS Duration and FE is -0.08391294, suggesting a very weak negative correlation between QRS duration and ejection fraction. In conclusion, a moderately positive relationship was observed between right atrial volume (VTD) and left atrial volume (VTS). Moderate-negative relationships were also identified between right atrial volume (VTD) and ejection fraction (FE), as well as between left atrial volume (VTS) and ejection fraction (FE). The presented p-value matrix represents the results of a t-test for the dif-

```
> # Afișarea rezultatelor
> print("Matricea de p-value pentru testul t:")
[1] "Matricea de p-value pentru testul t:"
> print(result)
                HTA          VTD          VTS Durata.QRS       FE.
HTA              NA 1.933597e-01 3.655982e-01  0.9223837 0.5029072
VTD       0.1933597           NA 2.614586e-10  0.6113590 0.2090597
VTS       0.3655982 2.614586e-10           NA  0.8385288 0.6085367
Durata.QRS 0.9223837 6.113590e-01 8.385288e-01         NA 0.5881299
FE.       0.5029072 2.090597e-01 6.085367e-01  0.5881299        NA
```

Figure 24: It shows a matrix of p-values obtained from a t-test for multiple variables. Each p-value in the matrix represents the result of a t-test for the mean difference between pairs of variables.

ferences between multiple variables, namely HTA, VTD, VTS, Durata QRS, and FE. Each cell of the matrix, except those on the main diagonal, shows the p-value for the t-test between pairs of variables. The main diagonal contains "NA" values because it does not make sense to compare a variable with itself. Interpreting the p-values is essential for understanding the results of the t-test. A small p-value, usually below 0.05, indicates a statistically significant difference between the two compared variables, while a large p-value suggests that there is no statistically significant difference between them.

For the variables HTA and VTD, the p-value of 0.1933597 indicates that there is no statistically significant difference between them. Similarly, the p-value of 0.8385288 for the variables VTS and Durata.QRS suggests that there is no statistically significant difference between these variables. However, a notable exception is the pair of variables VTD and VTS, where the p-value of 2.614586e-10 indicates a statistically significant difference, being very small.

In conclusion, most of the p-values in the matrix are above the 0.05 threshold, suggesting that there are no statistically significant differences between the respective pairs of variables. The major exception is the pair VTD and VTS, where the extremely small p-value indicates a statistically significant difference. The results presented in Figure 25 are part of the summary of a linear regression

```
> # Crearea modelului de regresie liniară
> model <- lm(HTA ~ Durata_QRS + VTD + VTS, data = data_selected)
> # Afișarea rezumatului modelului de regresie
> summary(model)

Call:
lm(formula = HTA ~ Durata_QRS + VTD + VTS, data = data_selected)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5874 -0.3533 -0.2417  0.5528  0.8302

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4618950  0.5081742   0.909   0.3688
Durata_QRS   0.0008962  0.0036588   0.245   0.8078
VTD         -0.0020322  0.0011613  -1.750   0.0878 .
VTS          0.0013948  0.0014441   0.966   0.3399
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4786 on 40 degrees of freedom
  (1145 observations deleted due to missingness)
Multiple R-squared:  0.07326,   Adjusted R-squared:  0.00375
F-statistic: 1.054 on 3 and 40 DF,  p-value: 0.3794
```

Figure 25: Creation and Display of Linear Regression Model

model. In the first part of the result, the formula used in the regression model appears, where HTA is the dependent variable, and QRS Duration, VTD, and VTS are the independent variables. Residuals represent the differences between the observed values and the model-predicted values. These provide an idea of the error distribution: Min: -0.5874, 1Q (first quartile): -0.3533, Median: - 0.2417,

3Q (third quartile): 0.5528, Max: 0.8302. The coefficients present the estimated coefficients of the model and their statistical significance. The intercept has an estimated value of 0.4618950 with a standard error of 0.5081742 and a p-value of 0.3688, suggesting it is not statistically significant. QRS Duration has an estimated coefficient of 0.0008962, a standard error of 0.0036588, a t-value of 0.245, and a p-value of 0.8078, indicating it is not significant. VTD has an estimated coefficient of -0.0020322, a standard error of 0.0011613, a t-value of -1.750, and a p-value of 0.0878, nearly significant at the 10% level. VTS has an estimated coefficient of 0.0013948, a standard error of 0.0014441, a t-value of 0.966, and a p-value of 0.3399, indicating it is not significant. The p-value ($\Pr(> |t|)$) indicates the probability that the coefficient is different from zero simply due to chance. The common threshold for statistical significance is 0.05: p-values < 0.05 indicate statistical significance, while p-values ≥ 0.05 indicate lack of statistical significance. The significance codes show significance levels: '' p <0.001, '' p < 0.01, '' p < 0.05, '.' p < 0.1, and ' ' p ≥ 0.1. Conclusions show that none of the predictors analyzed are significant at the 5% level ($p < 0.05$). VTD has a p-value of 0.0878, suggesting it might be significant at a more relaxed level (10%). QRS Duration and VTS are not significant predictors of HTA. The scatterplots presented in Figure 26

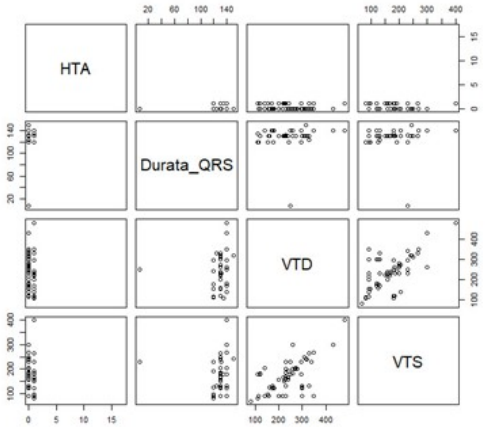


Figure 26: Scatterplots between two variables

show each cell of the matrix representing a scatter plot between two variables. For example, the cell in row 2, column 1 represents a scatter plot between HTA and QRS Duration. The main diagonal of the matrix contains histograms of each variable, showing the data distribution for that specific variable. The plots in the first row (excluding the histogram) show the relationship between HTA and the other variables. If the points are scattered without a clear pattern, it means there is no evident relationship between HTA and that variable. From the presented

graphs, there doesn't seem to be an evident relationship between HTA and the other variables (QRS Duration, VTD, VTS), as the points are scattered without a clear trend. The plots in row 2, i.e., columns 3 and 4, show the relationship between QRS Duration and the other two variables. Similarly, if the points are scattered without a clear pattern, there is no evident relationship between these variables. The plot in row 3, column 4 shows the relationship between VTD and VTS. It seems to present a certain positive trend, suggesting a positive correlation between VTD and VTS, i.e., as one increases, the other tends to increase. The lack of evident relationship, scatterplots between HTA and QRS Duration, VTD, and VTS do not show a clear pattern, suggesting there is no strong linear relationship between them. The relationship between VTD and VTS: there is a possible positive relationship between VTD and VTS, suggested by the tendency of the points to align on an increasing line. The plot in Figure 27 is a residuals

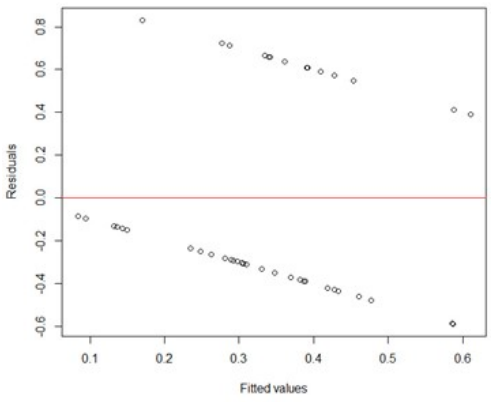


Figure 27: Residuals Plot against Fitted Values

versus fitted values plot and is used to assess the performance and adequacy of a linear regression model. This type of plot helps identify potential issues with the model, such as nonlinearity, heteroscedasticity (unequal variance of residuals), and the presence of influential points or outliers. The elements in the plot are the X-axis representing the fitted (predicted) model values for the dependent variable, HTA in this case, and the Y-axis representing the residuals, which are the differences between the observed values and the fitted values. The horizontal red line represents the zero residual line, where the model perfectly predicts the observed values. Interpretation of the Plot: The residuals should be randomly distributed around the horizontal zero line. If we observe a clear pattern or structure in the distribution of residuals, such as a curved shape or a cone, this may indicate issues in the model, such as nonlinearity or heteroscedasticity. Residual Variance:

The residuals should have constant variance (homoscedasticity). If the variance of the residuals systematically increases or decreases with the fitted values, this suggests heteroscedasticity, which can affect the validity of statistical inferences. Points that are far from the majority of the residuals may be influential points or outliers. These points can have a disproportionate effect on the model and may require further investigation. In the plot in Figure 26, it can be observed that the residuals are not completely randomly distributed around the zero line. There is a clear pattern, with residuals appearing to form a linear structure, suggesting that the model does not adequately capture the relationship between variables. This is a sign of nonlinearity, indicating that the relationship between the predictors (Durata-QRS, VTD, and VTS) and HTA is not linear and that a transformation of variables or the use of a more complex model may be necessary. Additionally, the residuals appear to have relatively constant variance, suggesting that there are no major issues of heteroscedasticity. Analysis of Risk Factors Associated with HTA Using Multinomial Regression I first checked if the nnet package is installed in R. If it was not installed, it needed to be automatically installed and then loaded to be used. I selected the data of interest from the dataset, retaining the HTA, Sex, Age, BCR, and CMD Etiology columns, and removed rows containing missing values to ensure the integrity of the analysis. Then, I transformed the Sex and CMD Etiology variables into factors to be correctly used in the regression model. I built a multinomial regression model for the dependent variable HTA, using the explanatory variables Sex, Age, BCR, and CMD Etiology. The results of this model were summarized to better understand the relationships between variables. The model in Figure 28 adjusted six parameters, including the intercept and co-

```
> library(nnet)
> model_multinom <- multinom(HTA ~ Sex + Varsta + BCR + Etiologie.CMD, data = sele
cted_data)
# weights:  6 (5 variable)
initial  value 40.202536
iter  10 value 31.337180
iter  20 value 31.301141
iter  30 value 31.297245
iter  40 value 31.297145
final  value 31.297092
converged
> summary(model_multinom)
Call:
multinom(formula = HTA ~ Sex + Varsta + BCR + Etiologie.CMD,
    data = selected_data)

Coefficients:
                          Values    Std. Err.
(Intercept)           -1.73487659 2.1084652693
SexM                   0.47886249 0.6474327977
Varsta                 0.01591882 0.0337293291
BCR                   -0.36771347 0.7255780956
Etiologie.CMDischemica -9.11117757 0.0003194576

Residual Deviance: 62.59418
AIC: 72.59418
```

Figure 28: Multinomial Regression and Its Results

efficients for the five explanatory variables. The initial value of the likelihood

function was 40.202536, and the model performed up to 40 iterations to optimize the likelihood function, reaching a final value of 31.297092. This indicates that the algorithm converged to a stable solution and stopped. The coefficient results of the model indicate that the intercept is -1.73487659, with a standard error of 2.1084652693, suggesting high variability in the estimation of this coefficient. The coefficient for SexM (males) is 0.47886249, indicating that being male increases the logit of the probability of having HTA compared to females, although standard errors of 0.6474327977 suggest uncertainty in this estimation. The coefficient for Age indicates that each additional year increases the logit of the probability of having HTA by 0.01591882, but standard errors of 0.0337293291 indicate that this effect is not significant. The coefficient for BCR (chronic kidney disease) suggests that the presence of BCR reduces the logit of the probability of having HTA by -0.36771347, but with a high degree of uncertainty, having a standard error of 0.7255780956. The very negative and small coefficient for CMD Etiology.ischemic (-9.11117757) suggests a strong association between this etiology and a reduced probability of HTA, and the extremely small standard errors (0.0003194576) indicate high precision in estimating this coefficient. The residual deviance of the model is 62.59418, measuring how well the model fits the data, and a lower value indicates a better fit. The AIC (Akaike Information Criterion) is 72.59418, which is a measure of model quality, penalizing its complexity. A lower AIC suggests a better model. In conclusion,the model converged to a stable solution, indicating that the calculated coefficients are reliable within the data and specified model. The coefficients for SexM and Age show a slight increase in the probability of having HTA with increasing age and being male, although these effects are not strongly supported by our data, having relatively large standard errors. The coefficient for CMD Etiology.ischemic shows a strong negative association with HTA, suggesting that patients with this etiology have a significantly lower probability of developing HTA.

The plot in Figure 29 illustrates the probability of developing hypertension (HTA) based on gender. The X-axis represents the Sex variable, which has two levels: female (F) and male (M). The Y-axis represents the probability of having HTA. For females, the estimated probability of having HTA is around 0.30, meaning that approximately 30% of females are likely to develop HTA according to our model. For males, the probability is higher, being around 0.40, meaning that 40% of males have a chance of developing HTA. The noticeable difference between the two points on the plot indicates that males have a higher probability of developing HTA compared to females. The trend line connecting the two points suggests a linear increase in the probability of having HTA from females to males. This underscores that male gender is associated with an increased risk of HTA within the multinomial regression model used. The plot simplifies the understanding of the influence of gender on the risk of HTA, highlighting that males have a higher risk compared to females, according to the analyzed data and model. This

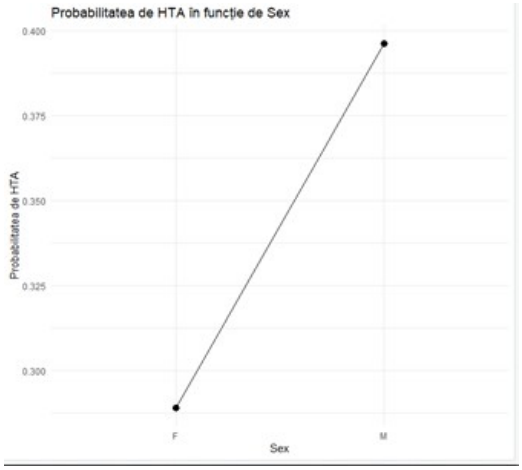


Figure 29: Plot for HTA Probability by Patient Gender.

suggests the importance of considering gender factor in risk assessment and HTA prevention strategies.

```
> ggplot(new_data_varsta, aes(x = Varsta, y = prob_HTA)) +
+   geom_point(size = 2) +
+   geom_line() +
+   labs(title = "Probabilitatea de HTA în funcție de vârstă",
+        x = "Vârstă",
+        y = "Probabilitatea de HTA") +
+   theme_minimal()
```

Figure 30: Creating the plot in the R system for HTA by Age.

The plot in Figure 31 illustrates the relationship between the probability of developing hypertension (HTA) and age. The X-axis represents age in years, and the Y-axis represents the probability of developing HTA. The title of the plot, "Probability of HTA by Age," clearly indicates that it shows how the probability of having HTA varies as a person ages. The X-axis presents numerical values from approximately 40 to 90 years old, indicating the age range studied. The Y-axis, on the other hand, shows numerical values from 0.20 to over 0.35, representing the probability of a person having HTA at a certain age. Each point on the plot represents the probability of HTA for a certain age. We observe that these points form an almost straight line, suggesting a linear relationship between age and the probability of HTA. As age increases, the probability of developing HTA also increases. For example, at the age of 40, the probability is approximately 20%, and at the age of 90, the probability increases to over 35%. This suggests that older individuals have a higher risk of developing HTA compared to younger individuals.

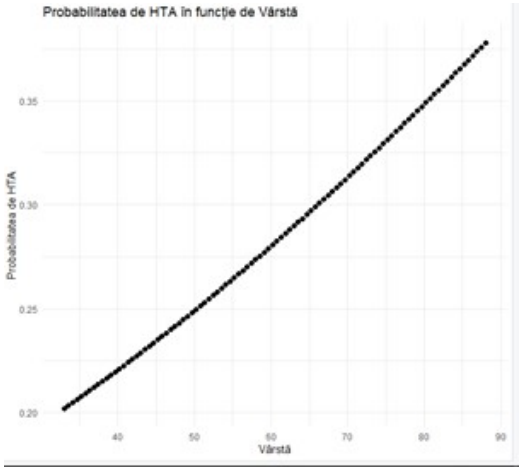


Figure 31: Plot for HTA Probability by Age.

```
> ggplot(new_data_bcr, aes(x = BCR, y = prob_HTA)) +
+   geom_point(size = 3) +
+   geom_line(aes(group = 1)) +
+   labs(title = "Probabilitatea de HTA în funcție de BCR",
+        x = "BCR",
+        y = "Probabilitatea de HTA") +
+   theme_minimal()
```

Figure 32: Creating the plot in the R system for HTA by Chronic Kidney Disease

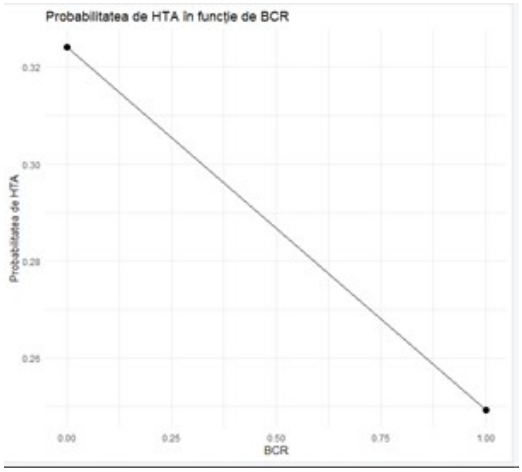


Figure 33: Plot for HTA Probability by Chronic Kidney Disease

The plot presented in Figure 33 illustrates the probability of developing hypertension (HTA) based on the presence of chronic kidney disease (CKD). The

X-axis represents the CKD variable, which has two values: 0 (absence of CKD) and 1 (presence of CKD). The Y-axis represents the probability of having HTA. For patients without CKD (value 0 on the X-axis), the estimated probability of having HTA is approximately 0.32 (or 32%). For patients with CKD (value 1 on the X-axis), the probability decreases to approximately 0.26 (or 26%). The line connecting the two points suggests an inversely proportional relationship between the presence of CKD and the probability of developing HTA. This means that, within the multinomial regression model used, patients with CKD have a lower probability of developing HTA compared to those without CKD. This negative relationship is surprising and could suggest a complex interaction between CKD and HTA in the analyzed data. Overall, the result emphasizes the importance of carefully examining all factors influencing the risk of HTA and the need for further analysis to fully understand these relationships.

```
> ggplot(new_data_etio, aes(x = Etiologie.CMD, y = prob_HTA)) +
+   geom_point(size = 3) +
+   geom_line(aes(group = 1)) +
+   labs(title = "Probabilitatea de HTA în funcție de Etiologie CMD",
+        x = "Etiologie CMD",
+        y = "Probabilitatea de HTA") +
+   theme_minimal()
```

Figure 34: Creating the plot in the R system for HTA by CMD Etiology

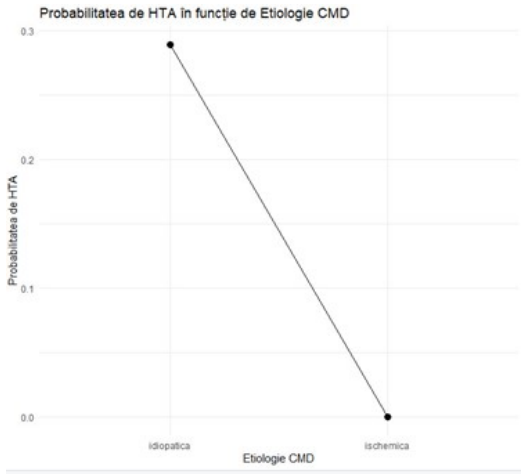


Figure 35: Plot for HTA Probability by Dilated Cardiomyopathy Etiology

The plot presented in Figure 35 illustrates the probability of developing hypertension (HTA) based on the etiology of dilated cardiomyopathy (CMD). The X-axis represents two types of etiology: idiopathic and ischemic. The Y-axis represents the probability of having HTA. For patients with idiopathic etiology, the

estimated probability of having HTA is approximately 30%. In the case of patients with ischemic etiology, the probability decreases significantly, approaching zero. The line connecting the two points indicates an inversely proportional relationship between CMD etiology and the probability of developing HTA. This suggests that patients with CMD of ischemic etiology have a much lower probability of developing HTA compared to those with CMD of idiopathic etiology. This strongly negative relationship reflected in the plot is consistent with the very negative coefficient for CMD Etiology.ischemic in the multinomial regression model, suggesting a strong and inverse association between ischemic etiology of CMD and the probability of developing HTA. These results underscore the importance of considering CMD etiology in assessing the risk of HTA and suggest that factors specific to each etiology type can have a significant impact on the risk of developing HTA.

Correlation between AV Sensed, AV Pace, HTA, METS, FE%

I started by checking that all variables needed for calculating the correlation are expressed as numeric values. In some cases, in the dataset, these may initially be represented as factors or characters, so I converted them into numeric values using the as.numeric function. Each line of code below is responsible for this conversion for a specific variable. After performing the necessary conversions, I checked the structure and content of the data to identify any anomalies, such as missing values or values that were not correctly converted. To obtain a summary description of each variable in the dataset, I called the summary function. The next step involved selecting only the relevant variables for calculating the correlation. This was achieved by creating a vector named vars, which contains the names of the desired variables, and then using this vector to extract the corresponding columns from the original dataset. As a result, I obtained a new dataset named data, which includes only the selected variables. Now, we can calculate the correlation matrix using the cor function. I specified use = "complete.obs" to ensure the use of only complete observations (without missing values) and method = "pearson" to apply the Pearson method of correlation, which measures the degree of linear association between variables. Finally, the correlation matrix is displayed using the print function, highlighting the correlation coefficients for all pairs of selected variables. In this section, we will examine the relationships between

```
> date$AV.sensed <- as.numeric(date$AV.sensed)
> date$AV.pace <- as.numeric(date$AV.pace)
> date$HTA <- as.numeric(date$HTA)
> date$METS <- as.numeric(date$METS)
> date$FE. <- as.numeric(date$FE.)
> summary(date)
```

Figure 36: Transformation of variables for numeric correlation calculation.

the AV.sensed (atrioventricular sensing) variable and other variables of interest: AV.pace (atrioventricular pacing), HTA (hypertension), METS (metabolic equivalents), and FE (ejection fraction). Between AV.sensed and AV.pace, the correla-

```
> vars <- c("AV.sensed", "AV.pace", "HTA", "METS", "FE.")
> data <- date[, vars]
> cor_matrix <- cor(data, use = "complete.obs", method = "pearson")
> print(cor_matrix)
           AV.sensed    AV.pace          HTA       METS          FE.
AV.sensed 1.0000000  0.8051803  0.25575208  0.4669951  0.24500545
AV.pace   0.8051803  1.0000000  0.14230296  0.5338741 -0.21710892
HTA       0.2557521  0.1423030  1.00000000  0.3430194 -0.06239852
METS      0.4669951  0.5338741  0.34301936  1.0000000 -0.14664341
FE.       0.2450055 -0.2171089 -0.06239852 -0.1466434  1.00000000
```

Figure 37: Calculating the Correlation Matrix and Results

tion coefficient is 0.8051803, indicating a strong positive correlation. This suggests that as atrioventricular sensing increases, atrioventricular pacing also increases. Between AV.sensed and HTA, the correlation coefficient is 0.25575208, suggesting a weak positive correlation between atrioventricular sensing and hypertension. Regarding the relationship between AV.sensed and METS, the correlation coefficient of 0.4669951 indicates a moderate positive correlation, suggesting that as atrioventricular sensing increases, the level of physical activity increases. Between AV.sensed and FE, the correlation coefficient is 0.24500545, indicating a weak positive correlation. Analyzing AV.pace and the other variables, between AV.pace and HTA, the correlation coefficient is 0.14230296, indicating a very weak positive correlation. Between AV.pace and METS, the correlation coefficient of 0.5338741 suggests a moderate positive correlation, indicating that as atrioventricular pacing increases, the level of physical activity increases. In contrast, the relationship between AV.pace and FE, with a correlation coefficient of -0.21710892, indicates a weak negative correlation, suggesting that as atrioventricular pacing increases, the ejection fraction tends to slightly decrease. In the relationship between HTA and the other variables, between HTA and METS, the correlation coefficient is 0.3430194, indicating a moderate positive correlation. This suggests that as hypertension increases, the level of physical activity increases, but this association is moderate. Between HTA and FE, the correlation coefficient of -0.06239852 indicates almost no correlation. For METS and the other variables, between METS and FE, the correlation coefficient is -0.14664341, indicating a very weak negative correlation. This suggests that as the level of physical activity increases, the ejection fraction tends to slightly decrease, but this association is very weak. In conclusion, all correlations with FE are quite weak, ranging between small negative and positive values, indicating that there is no strong linear association between ejection fraction and the other variables analyzed. Through correlation analysis, several types of relationships between the studied variables have been identified. Strong relationships are evident between atrioventricular sensing (AV.sensed) and

atrioventricular pacing (AV.pace), where there is a strong positive correlation. This suggests that as atrioventricular sensing activity increases, atrioventricular pacing also increases. Regarding moderate relationships, atrioventricular sensing (AV.sensed) and the level of physical activity (METS) exhibit a moderate positive correlation. Similarly, atrioventricular pacing (AV.pace) and the level of physical activity (METS) also have a moderate positive correlation, suggesting that both variables are associated with an increase in the level of physical activity. Additionally, hypertension (HTA) and the level of physical activity (METS) show a moderate positive correlation, indicating a relationship between increasing blood pressure and physical activity. The rest of the variable pairs show weak or almost non-existent correlations, suggesting that there is no significant linear association between them. This indicates that for these variable pairs, variations in one are not significantly associated with variations in the other. These results provide insight into how different patient characteristics are associated. For example, a strong correlation between AV.sensed and AV.pace indicates that these parameters are closely related, which could be relevant for understanding the functioning of medical devices or cardiac monitoring. Moderate and weak relationships may require further analysis to better understand the interactions and clinical implications. To create a correlation plot that visualizes the relationships between variables, I took the following steps: First, I installed the corrplot package. This package provides functions for creating the correlation plot. To install and load the package, I used the following commands: The data of interest are in a data frame named data, and the variables of interest are AV.sensed, AV.pace, HTA, METS, and FE. It is important that these variables are numeric to calculate the correlations. If the variables are not already numeric, they can be transformed using as.numeric. The next step involved calculating the correlation matrix for the variables of interest. For this, I used the cor function, where I specified the correlation method and how missing data are treated. Once the correlation ma-

```
> # Calcularea matricei de corelație
> cor_matrix <- cor(date[, c("AV.sensed", "AV.pace", "HTA", "METS", "FE.")], use = "complete.obs", method = "pearson")
> # Crearea graficului de corelație
> corrplot(cor_matrix, method = "circle")
```

Figure 38: Creating the correlation plot in the R system using the corrplot function for the variables AV Sensed, AV Pace, HTA, METS, FE%.

trix was obtained, I created the correlation plot using the corrplot function. I could choose different representation methods, such as circle, number, or color. The color and size of the circles represent the magnitude and direction of the correlation. Larger and more intense circles in color indicate stronger correlations. Blue color indicates positive correlations, while red color indicates negative correlations. If I chose the number method, the correlation coefficients are displayed directly on the plot, making it easier to interpret the exact values.

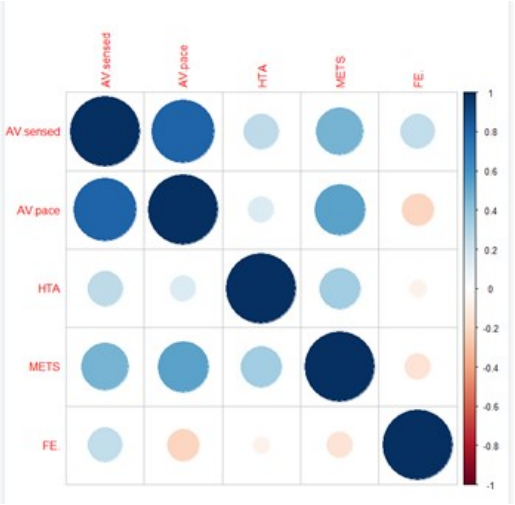


Figure 39: Correlation plot for the variables AV Sensed, AV Pace, HTA, METS, FE%.

```
> # Afișarea rezumatului modelului de regresie
> summary(model)

Call:
lm(formula = HTA ~ AV.sensed + AV.pace + METS + FE., data = data_selected)

Residuals:
     Min       1Q   Median       3Q      Max
-0.62856 -0.33749 -0.00417  0.39812  0.72066

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.900818   1.282487   0.702    0.498
AV.sensed    0.011949   0.009809   1.218    0.251
AV.pace     -0.010362   0.008989  -1.153    0.276
METS         0.083180   0.087894   0.946    0.366
FE.         -0.036573   0.038883  -0.941    0.369

Residual standard error: 0.5251 on 10 degrees of freedom
  (1174 observations deleted due to missingness)
Multiple R-squared:  0.2341,    Adjusted R-squared:  -0.07228
F-statistic: 0.7641 on 4 and 10 DF,  p-value: 0.5721
```

Figure 40: Displaying the regression model summary

The results presented in Figure 40 are part of the summary of a linear regression model.

In Figure 38, the formula used in the regression model is presented, where HTA is the dependent variable, and AV.sensed, AV.pace, METS, and FE are the independent variables. Residuals represent the differences between the observed values and the values predicted by the model. They provide an idea of the distribution of errors: Min: -0.62856, 1Q (first quartile): -0.33749, Median: -0.00417, 3Q (third quartile): 0.39812, Max: 0.72066. Coefficients present the estimated

coefficients of the model and their statistical significance. The intercept has an estimated value of 0.900818, but it is not statistically significant (p = 0.498). The coefficient for AV.sensed is 0.011949, but it is not statistically significant (p = 0.251). The coefficient for AV.pace is -0.010362, but it is not statistically significant (p = 0.276). The coefficient for METS is 0.083180, but it is not statistically significant (p = 0.366). The coefficient for FE is -0.036573, but it is not statistically significant (p = 0.369). Standard residuals show the standard error of the residuals, which is a measure of the variation of the residuals. The smaller the value, the better the model. In this case, the standard error of the residuals is 0.5251. R-squared represents the proportion of variance in the dependent variable (HTA) explained by the model. In this case, R-squared is 0.2341 (23.41%), which means the model explains 23.41% of the variance in HTA. Adjusted R-squared takes into account the number of predictors and the sample size and is -0.07228, indicating a poor model adjustment. The F-statistic is used to test the overall significance of the model. In this case, the F-value is 0.7641, and the p-value is 0.5721, suggesting that the model is not statistically significant (since p > 0.05). The conclusions show that none of the predictors analyzed are statistically significant at the 5% level (p < 0.05). The overall model is not statistically significant (p-value = 0.5721). R-squared and Adjusted R-squared suggest that the model poorly explains the variance in HTA. These results suggest that the variables AV.sensed, AV.pace, METS, and FE are not significant predictors for HTA in this dataset. It may be necessary to explore other variables or models to better understand the factors influencing HTA.

```
> print("Matricea de p-value pentru testul t:")
[1] "Matricea de p-value pentru testul t:"
> print(result)
           AV.sensed      AV.pace       HTA       METS       FE.
AV.sensed          NA 1.359167e-13 0.6551433 0.07925723 0.7539307
AV.pace  1.359167e-13           NA 0.2675836 0.04038120 0.5945990
HTA      6.551433e-01 2.675836e-01        NA 0.21069375 0.5029072
METS     7.925723e-02 4.038120e-02 0.2106937         NA 0.6020119
FE.      7.539307e-01 5.945990e-01 0.5029072 0.60201186        NA
```

Figure 41: The Figure shows a matrix of p-values obtained from a t-test for multiple variables: AV.sensed, AV.pace, HTA, METS, and FE. Each p-value in the matrix represents the result of a t-test for the mean difference between pairs of variables.

Interpreting the p-values is essential for understanding the results of the t-test. A small p-value ,usually below 0.05 indicates a statistically significant difference between the two compared variables, while a large p-value suggests that there is no statistically significant difference between them.

The results of the matrix are as follows:

- AV.sensed and AV.pace have a p-value of 1.359167e-13, indicating a statistically significant difference between them.

- AV.sensed and HTA have a p-value of 0.6551433, suggesting that there is no statistically significant difference between them.
- AV.sensed and METS have a p-value of 0.07925723, which is close to the 0.05 threshold but does not indicate a statistically significant difference.
- AV.sensed and FE have a p-value of 0.7539307, suggesting that there is no statistically significant difference between them.
- AV.pace and HTA have a p-value of 0.2675836, indicating that there is no statistically significant difference between them.
- AV.pace and METS have a p-value of 0.04038120, indicating a statistically significant difference between them.
- AV.pace and FE have a p-value of 0.5945990, suggesting that there is no statistically significant difference between them.
- HTA and METS have a p-value of 0.2106937, suggesting that there is no statistically significant difference between them.
- HTA and FE have a p-value of 0.5029072, suggesting that there is no statistically significant difference between them.
- METS and FE have a p-value of 0.6020119, suggesting that there is no statistically significant difference between them.

In conclusion, most pairs of variables do not show statistically significant differences, except for the pairs AV.sensed and AV.pace, and AV.pace and METS, where the p-values are very small, indicating statistically significant differences.

Each cell in the matrix represents a scatterplot between two variables. For example, the cell in the second row, first column represents a scatterplot between AV.sensed and AV.pace. On the main diagonal of the matrix are the histograms of each variable, showing the data distribution for that specific variable. The plots on the first row (except the histogram) show the relationship between AV.sensed and the other variables. The points are spread in a pattern that shows a positive correlation between AV.sensed and AV.pace, indicating that as one variable increases, the other tends to increase. The relationship between AV.sensed and the other variables (HTA, METS, FE) does not show a clear pattern, indicating a weak or nonexistent correlation. The plots in the second row (columns three, four, and five) show the relationship between AV.pace and the other variables. The points are scattered without a clear pattern, suggesting that there is no strong linear relationship between AV.pace and these variables. The plots in the third row (columns four and five) show the relationship between HTA and the other variables. The points are scattered without a clear pattern, indicating a weak or

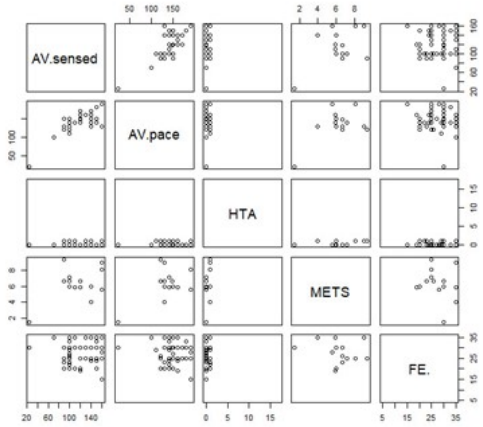


Figure 42: Each cell in the matrix represents a scatterplot between two variables.

nonexistent correlation between HTA and METS, as well as between HTA and FE. The plot in the fourth row, column five shows the relationship between METS and FE. The points are scattered randomly, suggesting that there is no strong linear relationship between METS and FE. In conclusion, scatterplots between most variables (AV.sensed, AV.pace, HTA, METS, FE) do not show a clear pattern, suggesting that there are no strong linear relationships between them. However, there is a positive correlation between AV.sensed and AV.pace, as suggested by the clear pattern in the scatterplot between these two variables.

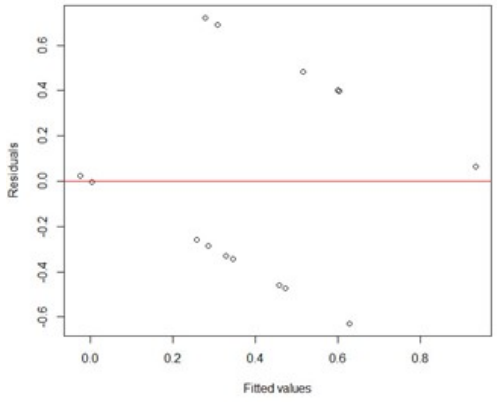


Figure 43: Plot of residuals versus fitted values

The plot I uploaded is a plot of residuals versus fitted values, and I use it to evaluate the performance and adequacy of the linear regression model. This type

of plot helps me identify possible issues with the model, such as non-linearity, heteroscedasticity (unequal variance of residuals), and the presence of influential points or outliers. On the X-axis are the fitted (estimated) values of the model for the dependent variable (HTA in this case), and on the Y-axis are the residuals, which are the differences between the observed values and the fitted values. The horizontal red line represents the zero residual line, where the model perfectly predicts the observed values. Residuals should be distributed randomly around the zero line. If I observe a clear pattern or structure in the distribution of residuals, such as a curved shape or a cone, this may indicate issues in the model, such as non-linearity or heteroscedasticity. Residuals should have constant variance. If the variance of residuals systematically increases or decreases with the fitted values, this suggests heteroscedasticity, which can affect the validity of statistical inferences. Points that are far from the majority of residuals may be influential points or outliers. These may have a disproportionate effect on the model and may require further investigation. In the plot I presented, I observe that the residuals are not completely randomly distributed around the zero line. There is a clear pattern, with residuals forming a linear structure, suggesting that the model does not adequately capture the relationship between variables. This is a sign of non-linearity, indicating that the relationship between my predictors (AV.sensed, AV.pace, METS, FE) and HTA is not linear and that a transformation of variables or the use of a more complex model may be necessary. Additionally, the residuals seem to have relatively constant variance, suggesting that there are no major issues with heteroscedasticity. However, the linearized pattern indicates that the current model is not adequate, and I should explore other functional forms or models to better capture the relationship between variables.

Correlation and Regression applied to the variables FE%, Age, QRS Duration, LVD, LVS, IVS, LV EDD, sPAP, and LA.V

The first step was to install and load the necessary packages for data manipulation and visualization. I installed the dplyr package, known for its data manipulation capabilities. I loaded the packages dplyr, readxl, and ggplot2 into the current R session. The readxl package is used for reading Excel files, while ggplot2 is used for creating detailed and customizable data visualizations. The next step was to read the ODS file containing the data needed for my analysis. I specified the path to the ODS file and used the read-excel function to read the data from the first sheet of the file. The data was then stored in the variable "data". To verify that the data was loaded correctly, I displayed the first few rows of the dataset using the head function. This step is important to confirm that the structure and content of the data align with expectations. Next, I used the attach function to attach the dataset, allowing me to access variables directly by their names, thereby simplifying the code and subsequent manipulations. To explore the dataset in more detail, I opened a viewing window using the View function

(Figure 42), providing an intuitive graphical interface for data examination.

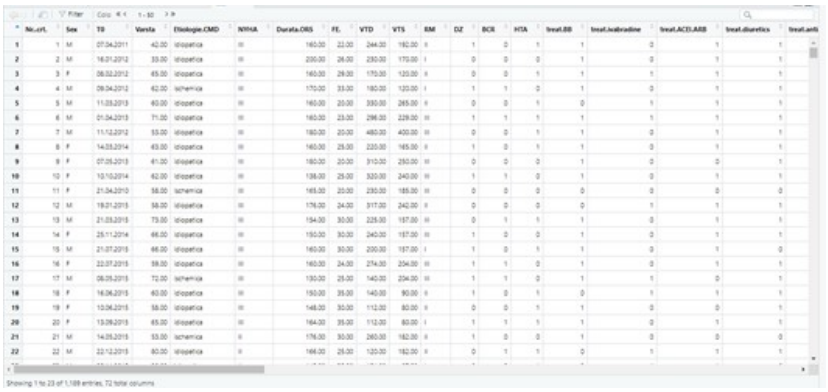


Figure 44: Displaying the dataset in the R system

I selected only the relevant columns from the dataset using the select function from dplyr. These columns include FE., Age, QRS Duration, VTD, VTS, IVS, LV.EDD, sPAP, and LA.V, which I consider important for my analysis. To convert categorical variables into factors, I used the mutate function, converting the variables Sex, Etiology.CMD, and NYHA into factors, facilitating their statistical analysis.

An important step was to remove rows with missing values from the dataset using the na.omit function. This ensured that subsequent analysis would not be affected by missing data. Finally, to verify the correctness of the data selection and preprocessing, I displayed the first few rows of the relevant-data dataset using the head function.

The result of these operations is a subset of data containing only the variables of interest and without rows with missing values. In this subset, the variables are: Ejection Fraction (FE.), Age (Varsta), QRS Duration (Durata.QRS), Left Ventricular Telediastolic Volume (VTD), Left Ventricular Telesystolic Volume (VTS), Interventricular Septum Thickness (IVS), Left Ventricular End-Diastolic Diameter (LV.EDD), Systolic Pulmonary Artery Pressure (sPAP), Left Atrial Volume (LA.V), Patient Sex (Sex), Dilated Cardiomyopathy Etiology (Etiologie.CMD), and NYHA Functional Classification (NYHA). This subset of data is now prepared for further statistical analyses and visualizations, ensuring that the data is clean and correctly preprocessed.

We began by selecting the relevant numeric variables using the select-if function from the dplyr package, thus ensuring that we are working only with numeric data. Subsequently, we calculated the correlation matrix between these variables using the cor function, which included only complete observations, thereby eliminating any missing values that could distort the results. To visualize the results, we used the print function to display the correlation matrix, indicating how the

```
> # Selectarea variabilelor relevante
> relevant_data <- date %>%
+   select(FE., Varsta, Durata.QRS, VTD, VTS, IVS, LV.EDD, sPAP, LA.V)
> # Convertirea variabilelor categorice în factori, dacă este cazul
> relevant_data <- relevant_data %>%
+   mutate(Sex = as.factor(Sex),
+          Etiologie.CMD = as.factor(Etiologie.CMD),
+          NYHA = as.factor(NYHA))
> # Eliminarea rândurilor cu valori lipsă
> relevant_data <- na.omit(relevant_data)
> # Verificarea variabilelor selectate
> head(relevant_data)
  FE. Varsta Durata.QRS VTD VTS IVS LV.EDD  sPAP LA.V Sex Etiologie.CMD NYHA
1  22     42        120 244 192 1.3    5.2    26   90   M    idiopatica  III
2  26     33        130 230 170 1.2    5.6    26   50   M    idiopatica  III
3  29     65        140 170 120 1.1    5.0    27   67   F    idiopatica  III
4  33     62        140 180 120 1.3    5.8    38  144   M     ischemica  III
5  20     60        130 330 265 1.3    7.0    38   80   M    idiopatica  III
6  23     71        120 296 229 1.4    7.0 56-60  131   M    idiopatica  III
```

Figure 45: The first 6 rows of the dataset were displayed using the head function

numeric variables correlate with each other. The values in the matrix range from -1 to 1, where 1 represents a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 suggests no correlation. The results of the correlation matrix highlighted several important relationships between variables. For example, the ejection fraction (FE.) showed a strong negative correlation with the left ventricular telediastolic volume (VTD) and left ventricular telesystolic volume (VTS), indicating that a higher ejection fraction is associated with smaller telediastolic and telesystolic volumes. Age showed weak correlations with the other variables, with the most notable being with the ejection fraction. The duration of the QRS complex had a moderate positive correlation with the thickness of the interventricular septum (IVS) and the left ventricular end-diastolic diameter (LV.EDD), suggesting that a longer duration of the QRS complex is associated with a greater thickness of the interventricular septum and a larger left ventricular end-diastolic diameter. The telediastolic and telesystolic volumes were strongly correlated with each other and had moderate correlations with the left ventricular end-diastolic diameter, indicating that larger volumes are associated with larger ventricular diameters. The thickness of the interventricular septum had a moderate positive correlation with the duration of the QRS complex and the left ventricular end-diastolic diameter. The left ventricular end-diastolic diameter showed moderate correlations with the duration of the QRS complex and the telediastolic volume, indicating a relationship between the dimensions of the left ventricle and these variables.

Moving on to the visual representation of the correlation. The presented graph is a visual correlation matrix showing the relationships between the numeric variables in the dataset. The horizontal and vertical axes display the variables FE, Age, QRS Duration, VTD, VTS, IVS, and LV.EDD. Each circle in the graph represents the correlation coefficient between two variables.

```
> # Calcularea matricei de corelație pentru variabilele numerice
> numeric_vars <- relevant_data %>% select_if(is.numeric)
> correlation_matrix <- cor(numeric_vars, use = "complete.obs")
> # Afișarea matricei de corelație
> print(correlation_matrix)
                FE.      Varsta  Durata.QRS         VTD         VTS         IVS      LV.EDD
FE.         1.00000000  0.17375776 -0.08391294 -0.59923268 -0.41931053  0.28935145 -0.40917269
Varsta      0.17375776  1.00000000 -0.03864959 -0.11107152  0.10734950 -0.04481518 -0.04001803
Durata.QRS -0.08391294 -0.03864959  1.00000000  0.07875215  0.03162304  0.56207591  0.67480485
VTD        -0.59923268 -0.11107152  0.07875215  1.00000000  0.67785849 -0.32538321  0.48657391
VTS        -0.41931053  0.10734950  0.03162304  0.67785849  1.00000000 -0.31730978  0.28674407
IVS         0.28935145 -0.04481518  0.56207591 -0.32538321 -0.31730978  1.00000000  0.23020844
LV.EDD     -0.40917269 -0.04001803  0.67480485  0.48657391  0.28674407  0.23020844  1.00000000
```

Figure 46: Visual representation of the correlation

Figure 47: Visualization of the correlation

The color of the circles indicates the direction of the correlation: red for positive correlations and blue for negative correlations. The redder the circle, the stronger the positive correlation, while a bluer circle indicates a stronger negative correlation. White circles indicate a very weak or nonexistent correlation. The size of the circles reflects the magnitude of the correlation; larger circles indicate a stronger correlation. From the graph, we observe that the ejection fraction (FE.) has a strong negative correlation with the left ventricular telediastolic volume (VTD) and left ventricular telesystolic volume (VTS). This means that a higher ejection fraction is associated with smaller telediastolic and telesystolic volumes. In contrast, the ejection fraction shows a moderate positive correlation with the thickness of the interventricular septum (IVS). Age (Age) has weak correlations with the other variables, with the most notable being a weak positive correlation with the ejection fraction. The duration of the QRS complex (QRS Duration) shows a strong positive correlation with the thickness of the interventricular septum and the left ventricular end-diastolic diameter (LV.EDD), suggesting that a longer duration of the QRS complex is associated with a greater thickness of the septum and a larger left ventricular end-diastolic diameter. The telediastolic volume (VTD) and telesystolic volume (VTS) are strongly correlated with each other and have moderate positive correlations with the left ventricular

end-diastolic diameter. The thickness of the interventricular septum (IVS) has a moderate positive correlation with the duration of the QRS complex and the left ventricular end-diastolic diameter. This graph is useful for understanding the relationships between the numeric variables in the dataset and for guiding further statistical analyses.

Next graph is a diagnostic set for a linear regression analysis. It consists of four sub-plots that help assess the quality and adequacy of the regression model. Here's a detailed explanation of each subplot: The "Residuals vs Fitted" plot

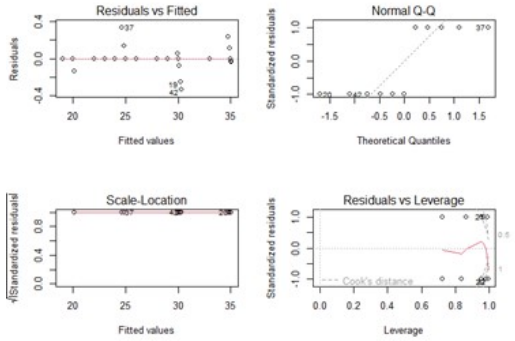


Figure 48: Linear Regression Analysis, consisting of four sub-plots that help assess the quality and adequacy of the regression model.

(Figure 48) shows the residuals (differences between observed and model-predicted values) on the vertical axis and the fitted values on the horizontal axis. Its purpose is to check linearity and homoscedasticity (constant variance of residuals). Residuals should be randomly distributed around the horizontal line at zero. The "Normal Q-Q" plot compares standardized residuals to the theoretical normal distribution (dotted line). Its purpose is to check whether residuals follow a normal distribution. Points should follow the straight diagonal line if residuals are normally distributed. Significant deviations from the line suggest non-compliance with the normality assumption. The "Scale-Location (Spread-Location)" plot shows the square root of standardized residuals on the vertical axis and fitted values on the horizontal axis. Its purpose is to check homoscedasticity. Points should be evenly distributed along the plot. The "Residuals vs Leverage" plot shows standardized residuals on the vertical axis and leverage on the horizontal axis. Its purpose is to identify influential observations. Observations with high leverage and large residuals have a significant impact on model fitting. The red line represents Cook's distance, a measure of each observation's influence. Points exceeding the Cook's line are considered influential and should be investigated. Conclusions: These diagnostic plots are essential for evaluating the quality and adequacy of the regression model. The "Residuals vs Fitted" plot shows that

residuals seem to be randomly distributed around the horizontal line, which is a good sign. The "Normal Q-Q" plot suggests that although most residuals follow the normal distribution, there are a few notable deviations. The "Scale-Location" plot indicates a relatively uniform distribution of points, suggesting homoscedasticity. The "Residuals vs Leverage" plot highlights some observations with high leverage, but they do not seem to significantly exceed the Cook's line, indicating that there are no excessively influential observations. Moving on to the graph representing the relationship between patients' age (Age) and ejection fraction (FE) using a linear regression model

```
Call:
lm(formula = FE. ~ Varsta, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-13.9356  -2.8449   0.3642   3.4977  11.5004

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.33625    3.49052   4.967 6.34e-06 ***
Varsta       0.15804    0.05527   2.859  0.00589 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.306 on 58 degrees of freedom
  (1129 observations deleted due to missingness)
Multiple R-squared:  0.1235,    Adjusted R-squared:  0.1084
F-statistic: 8.175 on 1 and 58 DF,  p-value: 0.005892
```

Figure 49: The results of a simple linear regression analysis, investigating the relationship between age (Varsta) and ejection fraction (FE).

The residual results show the differences between the observed values and the values estimated by the model. The minimum and maximum residual values are -13.9356 and 11.5004, respectively, while the first quartile, median, and third quartile values are -2.8449, 0.3642, and 3.4977, indicating the variability of the differences between the estimated and actual values. The intercept coefficient has a value of 17.33625, with a standard error of 3.49052, a t-statistic of 4.967, and a very small p-value of 6.34e-06, indicating high statistical significance. The coefficient for the age variable (Varsta) is 0.15804, with a standard error of 0.05527, a t-statistic of 2.859, and a p-value of 0.00589, showing that this coefficient is also statistically significant. The significance codes suggest different levels of statistical significance, where *** indicates a p-value less than 0.001, ** indicates a p-value less than 0.01, * indicates a p-value less than 0.05, and indicates a p-value less than 0.1. The very small p-values for both coefficients show that they are significantly different from zero.

The residual standard error is 5.306, calculated over 58 degrees of freedom, indicating the variability of the residuals and giving an idea of how well the model

fits. The R-squared statistic, with a value of 0.1235, indicates the proportion of the variability in the ejection fraction (FE) explained by age. An R-squared of 0.1235 means that approximately 12.35% of the variability in the ejection fraction is explained by age, while the adjusted R-squared is 0.1084.

The F-statistic has a value of 8.175 on 1 and 58 degrees of freedom, with a p-value of 0.005892, indicating that the model is significant overall. The black dots

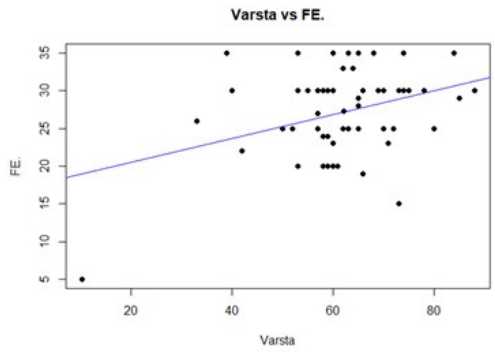


Figure 50: Observed values for each patient

in Figure 50 represent the observed values for each patient, with age on the X-axis and ejection fraction on the Y-axis. The blue line is the linear regression line, showing the general trend of the relationship between age and ejection fraction. Interpreting the regression coefficients shows that the intercept (17.33625) is the estimated value of the ejection fraction (FE) when age is zero. Although an age of zero is not realistic for patients, this coefficient provides a reference point for the regression line. The slope (0.15804) represents the rate of change of the ejection fraction per unit of age. For each year increase in age, the ejection fraction increases on average by 0.15804 units. The p-value for Age (0.00589) indicates that, since the p-value is less than 0.05, the coefficient for age is statistically significant. This means that there is a significant relationship between age and ejection fraction at a 95% confidence level. The Multiple R-squared (0.1235) indicates that approximately 12.35% of the variation in ejection fraction (FE) can be explained by variation in age. A lower R-squared may suggest that there are other important factors influencing the ejection fraction that are not included in this simple model. The overall interpretation of the graph and statistical results shows a positive relationship between age and ejection fraction, although this relationship is relatively weak (R^2 of 0.1235). Increasing age is associated with a slight increase in ejection fraction, but age explains only a small portion of the total variation in ejection fraction.

4 Conclusion

This study investigated the relationships between various cardiac parameters through the application of statistical correlation and regression methods, using data collected from 54 patients at the Timișoara Institute of Cardiovascular Diseases. The analysis aimed to improve the understanding of the mechanisms underlying heart failure and identify factors contributing to the optimization of treatment through cardiac resynchronization.

Significant Correlations: Correlation analysis revealed a strong negative correlation between ejection fraction (FE) and telediastolic (VTD) and telesystolic (VTS) volumes, indicating that patients with higher ejection fractions have smaller ventricular volumes. This relationship is crucial for understanding the pathophysiology of heart failure. Ventricular Volumes: The moderate positive correlation between telediastolic and telesystolic volumes suggests that ventricular dilatation is an important factor in the assessment and management of patients with heart failure.

AV Parameters and Physical Activity: Linear regression indicated a close link between AV.sensed and AV.pace, while physical activity measured by METS was not a significant predictor of ejection fraction. This underscores the importance of monitoring electrophysiological parameters in heart failure management.

Arterial Hypertension: Weak correlations between hypertension and other variables suggest a complexity of factors influencing this condition, highlighting the need for a more holistic approach in treating patients with both hypertension and heart failure. Scatterplots and Residual Plots: Analysis of residual plots highlighted the presence of influential points and anomalies, indicating the need for further investigation to improve the quality of the regression models used.

Clinical Implications The obtained results provide valuable insights for clinical practice. Understanding the relationships between different cardiac parameters can guide physicians in personalizing the treatment of patients with heart failure, thereby improving treatment efficiency and outcomes. Careful monitoring of ejection fraction, ventricular volumes, and electrophysiological parameters is essential for the effective management of these patients.

Limitations and Recommendations for Future Research While the study provided valuable information, there are limitations such as the relatively small sample size and potential unaccounted variables that could influence the results.

In conclusion, the use of descriptive statistical methods, correlations, and regressions provided a solid foundation for understanding the complexity of cardiac pathology. These methods allowed the highlighting of essential relationships between the studied variables, thus contributing to the improvement of clinical and therapeutic approaches in heart failure treatment.

References

[1] M. Gayaghan, "Cardiac Anatomy and Physiology: A Review," AORN Journal, vol. 67, no. 4, pp. 800-822, 1998.

[2] R. Sun, M. Liu, L. Lu, Y. Zheng și P. Zhang, Congenital Heart Disease: Causes, Diagnosis, Symptoms, and Treatments," Cell Biochemistry and Biophysics, vol. 72, pp. 857-860, 2015.

[3] J. A. C. Cardiol, Birth Prevalence of Congenital Heart Disease Worldwide: A Systematic Review and Meta-Analysis," JACC Journals, vol. 58, nr. 21, pp. 2241-2247, 2011.

[4] D. Choi, K.-C. Hwang, K.-Y. Lee și Y.-H. Kim, Ischemic heart diseases: Current treatments and future," Journal of Controlled Release, vol. 140, nr. 3, pp. 194-202, 2009.

[5] H. Yu, K. Lu, J. Zhu și J. Wang, Stem cell therapy for ischemic heart diseases," British Medical Bulletin, vol. 121, nr. 1, pp. 135-154, 2017.

[6] J. Bouchardy, J. Therrien, L. Pilote, R. Ionescu-Ittu, G. Martucci, N. Bottega și A. J. Marelli, Atrial Arrhythmias in Adults With Congenital Heart Disease," Circulation, vol. 120, nr. 17, pp. 1679-1686, 2009.

[7] D. Tomasoni, M. Adamo, C. M. Lombardi și M. Metra, Highlights in heart failure," ESC Heart Failure, vol. 6, nr. 6, pp. 1105-1127, 2020.

[8] G. Fent, J. Gosai și M. Purva, Teaching the interpretation of electrocardiograms: Which method is best?," Journal of Electrocardiology, vol. 48, nr. 2, pp. 190-193, 2015.

[9] M. Cikes, L. Tong, G. R. Sutherland și J. D'hooge, Ultrafast Cardiac Ultrasound Imaging: Technical Principles, Applications, and Clinical Benefits," JACC Journals, vol. 7, nr. 8, pp. 812-823, 2014.

[10] M. A. Chizner, The diagnosis of heart disease by clinical assessment alone," Disease-a-Month, vol. 48, nr. 1, pp. 7-98, 2002.

[11] P. P. Rossignol, A. F. Hernandez, S. D. Solomon și F. Zannad, Heart failure drug treatment," The Lancet, vol. 393, nr. 10175, p. 2019, 2019.

[12] T. Ide, K. Ohtani, T. Higo, M. Tanaka, Y. Kawasaki și H. Tsutsui, Ivabradine for the Treatment of Cardiovascular Diseases," Circulation Journal, vol. 83, nr. 2, pp. 252-260, 2019.

[13] N. Clappers, M. A. Brouwer și F. W. A. Verheugt, Antiplatelet treatment for coronary heart disease," BMJ Journals, vol. 93, nr. 2, pp. 258-265, 2007.

[14] S. Bangalore, F. H. Messerli, J. B. Kostis și C. J. Pepine, Cardiovascular Protection Using Beta-Blockers: A Critical Review of the Evidence," Journal of the American College of Cardiology, vol. 50, nr. 7, pp. 563-572, 2007.

[15] B. R. Chaitman, Ranolazine for the Treatment of Chronic Angina and Potential Use in Other Cardiovascular Conditions," Circulation, vol. 113, nr. 20, p. 2462–2472, 2006.

[16] S. Leeder și M. Gliksman, Prospects for preventing heart disease," National Center for Biotechnology Information, vol. 301, nr. 3, pp. 1004-1005, 1990.

[17] R. Estruch, E. Ros, J. Salas-Salvadó, M.-I. Covas, D. Corella, D. Pharm, F. Arós, E. Gómez-Gracia și V. Ruiz-Gutiérrez, Primary Prevention of Cardiovascular Disease with a Mediterranean Diet," The New England - Journal of Medicine, vol. 368, nr. 14, pp. 1279-1290, 2013.

[18] Y. Yamori, Predictive and Preventive Pathology of Cardiovascular Diseases," Pathology International, vol. 39, nr. 11, pp. 683-705, 1989.

[19] J. P. Veinot și R. Lemery, Innovations in cardiovascular pathology Anatomic and electrophysiologic determinants associated with ablation of atrial arrhythmias," Cardiovascular Pathology, vol. 14, nr. 4, pp. 204-213, 2005.

[20] D. S. Krantz1 și M. K. McCeney, Effects of Psychological and Social Factors on Organic Disease: A Critical Assessment of Research on Coronary Heart Disease," Annual Review of Psychology, vol. 53, pp. 341-369, 2002.

[21] T. G. N. PhD, Descriptive Statistics," în Topics in Biostatistics: Methods in Molecular Biology, Humana Press, 2007, p. 404.

[22] D. Chakrabarty, Model Describing Central Tendency of Data," International Journal of Advanced Research in Science, Engineering and Technology, vol. 8, nr. 9, 2021.

[23] Mishra, Prabhaker, Pandey, C. M., Singh, Uttam, Gupta, Anshul, Sahu, Chinmoy, Keshri și Amit, Descriptive Statistics and Normality Tests for Statistical Data," Annals of Cardiac Anaesthesia, vol. 22, nr. 1, pp. 67-72, 2019.

[24] D. Curran-Everett, Explorations in statistics: standard deviations and standard errors," American Physiological Society, vol. 32, nr. 3, pp. 203-208, 2008.

[25] D. G. Altman și J. M. Bland, Statistics Notes: Quartiles, quintiles, centiles, and other quantiles," The BMJ, pp. 309-996, 1994.

[26] Y. H. Chan, Biostatistics 104: Correlational Analysis," Basic Statistics For Doctors, vol. 44, nr. 12, pp. 614-619, 2003.

[27] M. MANOLACHE, E. C. D. A. C. MIRCIOIU, I. PRASACU și R. S. , BIOETHICS APPROACH OF BIOSTATISTICS IN CLINICAL TRIALS. AVOID THE USE OF EXCESSIVE OR INADEQUATE NUMBERS OF RESEARCH SUBJECTS," Journal of Science and Arts, vol. 42, nr. 1, pp. 239-246, 2018.

[28] S. L. Crawford, Correlation and Regression," AHA Journals, vol. 114, nr. 19, pp. 2083-2088, 2006.

[29] M. D. Sami L. Bahna, Statistics for Clinicians," American College of Allergy, Asthma Immunology, vol. 103, nr. 4, 2009.

[30] T. Amemiya, Chapter 6 Non-linear regression models," Handbook of Econometrics, vol. 1, pp. 333-389, 1983.

[31] H. J. Keselman, C. J. Huberty, L. M. Lix, S. Olejnik, R. A. Cribbie, B. Donahue, R. K. Kowalchuk, L. L. Lowman, M. D. Petoskey, J. C. Keselman și J.

R. Levin, Statistical Practices of Educational Researchers: An Analysis of their ANOVA, MANOVA, and ANCOVA Analyses," Sage Journals, vol. 68, nr. 3, 1997.

[32] A. G. Japp, A. Gulati, S. A. Cook, M. R. Cowie și S. K. Prasad, The Diagnosis and Evaluation of Dilated Cardiomyopathy," JACC Journals, vol. 67, nr. 25, pp. 2996-3010, 2016.

[33] C. Bredy, M. Ministeri, A. Kempny, R. Alonso-Gonzalez, L. Swan, A. Uebing, G.-P. Diller și K. D. Michael A. Gatzoulis, New York Heart Association (NYHA) classification in adults with congenital heart disease: relation to objective measures of exercise and outcome," European Heart Journal - Quality of Care and Clinical Outcomes, vol. 4, nr. 1, pp. 51-58, 2017.

[34] A. Brenyo și W. Zareba, Prognostic significance of QRS duration and morphology," Cardiology Journal, vol. 18, nr. 1, pp. 8-17, 2011.

[35] J. E. Sanderson, Heart failure with a normal ejection fraction," BMJ Journals, vol. 93, nr. 2, pp. 418-418, 2007.

[36] C. Luong, D. J. S. Thompson, M. Bennett, K. Gin, J. Jue, M. E. Barnes, P. Colley și T. S. M. Tsang, Right Atrial Volume Is Superior to Left Atrial Volume for Prediction of Atrial Fibrillation Recurrence After Direct Current Cardioversion," Canadian Journal of Cardiology, vol. 31, nr. 1, pp. 29-35, 2015.

[37] M. Enriquez-Sarano, C. W. Akins și A. Vahanian, Mitral regurgitation," The Lancet, vol. 373, nr. 9672, pp. 1382-1394, 2009.

[38] Y. C. Ng, P. Jacobs și J. Johnson, Productivity Losses Associated With Diabetes in the U.S," Diabetes Care, vol. 24, nr. 2, p. 257–261, 2001.

[39] P. Drawz și M. Rahman, Chronic Kidney Disease," Annals of Internal Medicine, vol. 162, nr. 11, 2015.

[40] J. E. Hall, J. P. Granger, J. M. d. Carmo, A. A. d. Silva, J. Dubinion, E. George, S. Hamza, J. Speed și M. E. Hall, The pathophysiology of hypertension," Comprehensive Physiology, vol. 2, nr. 4, pp. 322-912, 2012.

[41] P. E. Carson, Beta blocker treatment in heart failure," Progress in Cardiovascular Diseases, vol. 41, nr. 4, pp. 301-321, 1999.

[42] E. A. Bocchi și V. M. C. Salemi, Ivabradine for treatment of heart failure," Expert Opinion on Drug Safety, vol. 18, nr. 5, pp. 393-402, 2019.

[43] R. D. Toto, Treatment of Hypertension in Chronic Kidney Disease," Seminars in Nephrology, vol. 25, nr. 6, pp. 435-439, 2005.

[44] Blowey și D. L, Diuretics in the treatment of hypertension," Pediatric Nephrology, vol. 31, p. 2223–2233, 2016.

[45] J. A. Delyani, Anti-aldosterone therapy in the treatment of heart failure: new thoughts on an old hormone," Expert Opinion on Investigational Drugs, vol. 7, nr. 5, pp. 753-759, 1998.

[46] H.-I. Lu, M.-S. Tong, K.-H. Chen, F.-Y. Lee, J. Y. Chiang, S.-Y. Chung, P.-H. Sung și H.-K. Yip, Entresto therapy effectively protects heart and lung

against transverse aortic constriction induced cardiopulmonary syndrome injury in rat," National Library of Medicine, vol. 10, nr. 8, pp. 2290-2305, 2018.

[47] A. M. SCHER, M. I. RODRIGUEZ, J. LIIKANE și A. C. YOUNG, The Mechanism of Atrioventricular Conduction," Circulation Research, vol. 7, nr. 1, pp. 54-61, 1959.

[48] M. Cameli, G. E. Mandoli, E. Lisi, A. Ibrahim, E. Incampo, G. Buccoliero, C. Rizzo, F. Devito, M. M. Ciccone și S. Mondillo, Left atrial, ventricular and atrio-ventricular strain in patients with subclinical heart dysfunction," International Journal of Cardiovascular Imaging, vol. 35, pp. 249-258, 2018.

[49] S. Kaul, The interventricular septum in health and disease," American Heart Journal, vol. 112, nr. 3, pp. 568-581, 1986.

[50] M. Kukucka, A. Stepanenko, E. Potapov, T. Krabatsch, M. Redlin, A. Mladenow, H. Kuppe, R. Hetzer și H. Habazettl, Right-to-left ventricular end-diastolic diameter ratio and prediction of right ventricular failure with continuous-flow left ventricular assist devices," The Journal of Heart and Lung Transplantation, vol. 30, nr. 1, pp. 64-69, 2011.

[51] B. M. McQuillan, M. H. Picard, M. Leavitt și A. E. Weyman, Clinical Correlates and Reference Intervals for Pulmonary Artery Systolic Pressure Among Echocardiographically Normal Subjects," Circulation, vol. 104, nr. 23, pp. 2797-2802, 2001.

[52] S. J. J. D. W. M. Allison M Pritchett, R. J. Rodeheffer, K. R. Bailey și M. M. Redfield, Left atrial volume as an index ofleft atrial size: a population-based study," JACC Journals, vol. 41, nr. 6, pp. 1036-1043, 2003.

[53] M. Arsalan, T. Walther, R. L. Smith și P. A. Grayburn, Tricuspid regurgitation diagnosis and treatment," European Heart Journal, vol. 38, nr. 9, pp. 634-638, 2017.

[54] K. V., G. G., D. Oikonomou, C. Aggeli, C. Grassos, D. Papadopoulos, C. Thomopoulos, M. Marketou, K. Dimitriadis, K. Toutouzas, P. Nihoyannopoulos, C. Tsioufis și D. Tousoulis, Aortic Stenosis, Aortic Regurgitation and Arterial Hypertension," Current Vascular Pharmacology, vol. 17, nr. 2, pp. 180-190, 2019.

[55] D. S. Khoury, Reconstruction of endocardial potentials and activation sequences from intracavitary probe measurements: localization of stimulare sites and effects of myocardial structure.," Circulation, vol. 91, nr. 3, pp. 845-863, 1995.

[56] R. A. Bruce și J. R. McDonough., Stress testing in screening for cardiovascular disease," Bulletin of the New York Academy of Medicine, vol. 45, nr. 12, p. 1288, 1969.

[57] T. C. Edwards, B. Guest, A. Garner, K. Logishetty, A. D. Liddle și J. P. Cobb, The metabolic equivalent of task score," Bone Joint Res, vol. 11, nr. 5, pp. 317-326, 2022.

[58] J. Rankinen, P. Haataja, L.-P. Lyytikäinen, H. Huhtala, T. Lehtimäki,

M. Kähönen, M. Eskola, A. R. Pérez-Riera, A. Jula, H. Rissanen, K. Nikus și J. Hernesniemi, Long-term outcome of intraventricular conduction delays in the general population," Electrocardiogram, vol. 26, nr. 1, 2020.

[59] A. L. Aro, O. Anttonen, J. T. Tikkanen, M. J. Junttila, T. Kerola, H. A. Rissanen, A. Reunanen și H. V. Huikuri, Intraventricular Conduction Delay in a Standard 12-Lead Electrocardiogram as a Predictor of Mortality in the General Population," Circulation: Arrhythmia and Electrophysiology, vol. 4, nr. 5, pp. 704-710, 2011.

[60] O. Okafor, A. Zegard, P. v. Dam, B. Stegemann, T. Qiu, H. Marshall și F. Leyva, Changes in QRS Area and QRS Duration After Cardiac Resynchronization Therapy Predict Cardiac Mortality, Heart Failure Hospitalizations, and Ventricular Arrhythmias," Journal of the American Heart Association, vol. 8, nr. 21, 2019.

[61] R. J. Martis, U. R. Acharya și H. Adeli, Current methods in electrocardiogram characterization," Computers in Biology and Medicine, vol. 48, nr. 1, pp. 133-149, 2014.

[62] G. T. MA, Mechanisms of cardiac arrhythmias," Journal of Arrhythmia, vol. 32, nr. 2, pp. 75-81, 2016.

Bătrînescu Miruna Cristina – MSc student, Department of Mathematics,

Politehnica University of Timişoara,

P-ta Victoriei 2, 300 006, Timişoara, ROMANIA

E-mail: miruna.batrinescu@student.upt.ro

Cristina Văcărescu – Medical Doctor

Victor Babes University of Medicine and Pharmacy, Timisoara,,

P-ta Eftimie Murgu 2, 300 041, Timişoara, ROMANIA

E-mail: vacarescucristina@yahoo.com

INSTRUCTIONS FOR THE AUTHORS

The "Buletinul Ştiinţific al Universităţii Politehnica Timişoara" is a direct successor of the "Buletin scientifique de l'École Polytechnique de Timişoara" which was started in 1925 . Between 1982 - 1989 it was published as the "Lucrările Seminarului de Matematică şi Fizică ale Institutului Politehnic "Traian Vuia" din Timişoara".

Publication program: one volume per year, each series.

The "Mathematics-Physics" series of the "Buletinul Ştiinţific al Universităţii Politehnica Timişoara" publishes original papers in all areas of the pure and applied mathematics and physics.

1. The manuscript should be sent to the Editor, written in English (or French, or German, or Russian). The manuscript should be prepared in LATEX.

2. The length should preferably in an even number of pages, in Times New Roman (12 pt).

3. For the first page the authors must bear in view:

- an Abstract single spaced (at most 150 words) will be placed before the beginning of the text as: ABSTRACT. The problem of ...

- footnote with MSC (Mathematics Subjects Classification) or PACS (Physics Abstracts Classification System) Subject Classification Codes (10 pt)

4. Graphs. illustrations and tables should placed into the manuscript (send it as .eps or .pdf file), with the corresponding consecutively number and explanations under them.

5. The complete author's (authors') address(es) will be placed after References.

6. A Copyright Transfer Agreement is required together with the paper. By submitting a paper to this journal, authors certify that the manuscript has not been submitted to, nor is it under consideration for publication by another journal, conference proceedings, or similar publications.

7. There is no page charge and after registration a paper is sent to two independent referees. After acceptance and publication the author will receive 5 reprints free of charge. If a larger amount is required (fee of 2 USD per copy per article) this should be communicated to the Editor.

8. Manuscripts should be sent to:

For Mathematics

Dr. Liviu CĂDARIU
Politehnica University Timisoara
Department of Mathematics
Victoriei Square, №. 2
300006 - Timisoara, Romania
liviu.cadariu-brailoiu@upt.ro

For Physics

Dr. Duşan POPOV
Politehnica University Timisoara,
Department of Physical Fundamentals
B-dul. V. Parvan, №. 2
300223 - Timisoara, Romania
dusan.popov@upt.ro