FRI-2.203-2-TMS-02

ASSESMENT OF TIME IRREGULARITIES OF ROAD ACCIDENTS IN BULGARIA

Prof. Velizara Pentcheva, PhD
Department of Transport,
"Angel Kanchev" Univesity of Ruse
Phone: 082-888 240
E-mail: vpencheva@uni-ruse.bg

Assoc. Prof. Asen Asenov, PhD
Department of Transport,
"Angel Kanchev" Univesity of Ruse
Phone: 086-821 605
E-mail: asasenov@uni-ruse.bg

Eng. Stanimir Penev,
Department of Transport,
"Angel Kanchev" Univesity of Ruse
Phone: 086-821 605
E-mail: spenev@uni-ruse.bg

Abstract: *This document provides a time analysis of the problem (fatal road accident) for months of the year. The purpose of this study is to analyze recorded fatalities from road accidents for the period 2011 - 2018, to show that there is a correlation between the number of deaths and the days of the week and it is statistically significant, and to identify the characteristic points with the highest number of deaths during the year.*

Keywords: *assessment, time maps, temporal analysis, traffic accidents*

INTRODUCTION

Bulgaria's national road safety improvement strategy for 2011–20 is a political framework document that sets out the guidelines for implementing a policy to improve conditions for road users and reduce the number of victims of road accidents. For that matter the following indicators have been defined: road fatalities with a reduction target of 50% compared to 2010, or by 2020 the number should not go exceed 388; the number of serious injuries in road accidents has to be reduced by 20%, so that by 2020 their number should not exceed 6363. The prediction for 2018 for killed on the road was expected to be around 465 people, but the record revealed about 30% increase (611 people). Similar results appear for the number of injured on the road. The number of traffic accidents varies depending on several factors: the traffic volumes during weeks, months, day/night hours, weekends and holidays, weather conditions and other. Investigating the time irregularities of road accidents and the number of dead and injured is essential because it helps pointing out risky periods of time, planning and constructing control activities, setting up campaigns among drivers from different ages, as well as passengers and pedestrians. The quick change in the road conditions due to the weather or traffic density, leads to time irregularities of the road accidents and this survey will help provide statistical solutions.

Many studies view different types of traffic safety analysis for different time periods. Research experts apply time-based analysis to traffic accidents over time. A good example is the article [Lyubenov, D., Marinov, M., Kostadinov, S., and Gelkov, Zg., 2011], where the unit of observation point a period of time (hour, day, month and year). In the article [Atanasova, P., Lyubenov, D., Kostadinov, S., and Kirilov, F., 2017], parts 1 and 2 is shown another approach of analyss of the time irregularities of traffic accidents for the region of one of the five largest cities in the Republic of Bulgaria. Time analysis is an abstract interpretation and is based on time series, variables and the mathematical tools used to determine it. Additionally, in the article [Karacasua, M., Erb, A., Bilgiça, S., Barutb, H.], the chi-square test is used with the Statistical Package for Social Sciences (SPSS), to determine whether the days have a significant hour of occurrence of accidents on the road. The article [Pencheva, V., Tsekov, A., Georgiev, I., Kostadinov, S., 2018] assesses the regularity of mass urban passenger transport traffic in the conditions of the city of Ruse in Bulgaria. Various transport modes and a modelling of the interaction of different vehiclesis shown in [A. Sladkowski, 2019]. And in [Centeno, V., Georgiev, IR., Mihova, V. and Pavlov, V. 2019] is sh shown a good example of the Arima method.

An example of a common gap in statistical analyses in the public domain or those included in scientific developments – publications, dissertations, etc is the mathematical difference between the variables. The relative proportions in the sample population examined do not prove the significance (reliability) of the differences. When performing correlation analysis, the significance (reliability) of the correlation coefficients used is not examined, since the insignificant coefficients do not carry much content. Using specialised software product can greatly help facilitate a research. For this article is used a software for statistical processing and analysis of information – SPSS (Statistical Package for Social Sciences)

EXPOSITION

The modern motorization in Bulgaria, the poor condition of the infrastructure, as well as the incompetence of some drivers (mostly in the risky) situations, lead to an alarmingly high number of road accidents. The period 2011 – 2020 is crucial for the EU and for the country as a strategic task of lowering the numbers of victims of traffic accidents. In this regard, the article observes the conditions of the road safety situation for the period 2011 until the end of last year 2018.

Figure 1 indicates the number of deaths by years for the stated period.

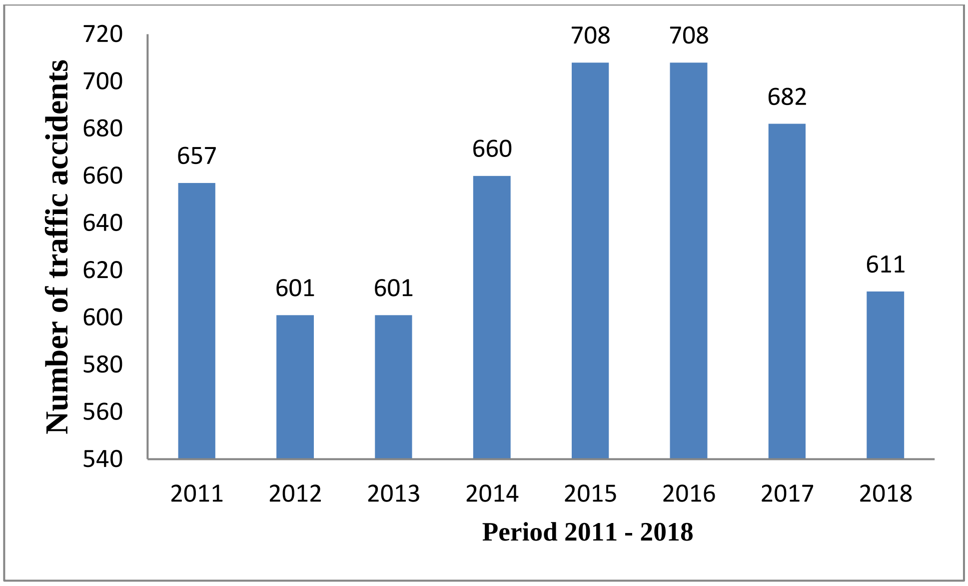

Fig. 1. Number of deaths by year for the period 2011 - 2020

The average annual death rate is 654 people, which is 87,3 per million inhabitants. This is significantly higher than the EU's average mortality rate, which is and average of 49 deaths per million inhabitants.

In table. 1., are the basic numerical characteristics of the deaths per month in an accident. The data obtained for the values of the different numerical characteristics in the table show that the average number of deaths per month for the last 8 years is 55 from the extracted population sample. The median is 53,5 which means that 50% of the cases the death toll is over 54 people, in the other 50% of the cases is below 53 people. The most common death toll during this period is

44 people. The standard deviation (random scatter) is approximately 13%. The variance is 172,6 which is $\sqrt{\text{Std. Deviation}}$. The Skeweness coefficient for the shape of the distribution has a positive asymmetry of 0,076. The Excess constitute a measure of the sharpness of the graph, in this case it is negative (-0,437). The smallest death toll per month is 25 people and the highest in a month is 86 people.

Table 1. Basic numerical characteristics of the number of deaths per month for the period (2011 - 2018)

Statistics		
Number of perished		
N	Valid	96
	Missing	0
Mean		55,0000
Median		53,5000
Mode		44,00
Std. Deviation		13,13733
Variance		172,589
Skewness		,076
Std. Error of Skewness		,246
Kurtosis		-,437
Std. Error of Kurtosis		,488
Minimum		25,00
Maximum		86,00

Time fluctuations are observed in the analysis of traffic data in the country. One of these time fluctuations is the number of accidents per month. The coefficient of irregularity for the period is:

$$\eta_{н} = \frac{Q_{max}}{Q_{average}} = \frac{86}{55} = 1,56,$$

Where Q_{max} is the highest number of deaths in an accident per month and, the Q_{cp} is an average number of deaths per month for the period 2011 – 2018.

The high coefficient of unevenness shows that there are significant differences in the number of deaths by month of the year. This is also confirmed by the coefficient of variation (measure of uniformity of the sample).

$$V = \frac{\text{Stand Dev}}{\text{Mean}} * 100\% = \frac{13,14}{55} * 100\% = 23,89\%$$

Therefore, the sample tends to be of average homogeneity.

It is necessary to determine whether there is a correlation between the death toll and the months of the year. The death toll is a quantative variable type and the months are a qualitive variable type. To determine the correlations depending we use the Pearson Correlation test, which is used in most cases to find a correlation between varriables. The results are shown in Table. 2.

Tab. 2. Standard Pearson Correlation between death toll and months

Correlations

		Number of perished	Month
Number of perished	Pearson Correlation	1	,509**
	Sig. (2-tailed)		,000
	N	96	96
Month	Pearson Correlation	,509**	1
	Sig. (2-tailed)	,000	
	N	96	96
**. Correlation is significant at the 0.01 level (2-tailed).			

From Table. 2. can be pointed that there is a correlation of (0.509) and we can assume that it is about medium to strong correlation. This coefficient is statistically significant – Sig. (2-tailed) $< \alpha = 0{,}05$.

To verify the result, are used the tests of Kendall's Tau-b and Spearman's Rho (Tab. 3). The test of Kendall's Tau-b correlation is used in most cases to determine the correlation qualitative and quantative variables. Spearman's Rho test is used to determine the correlation between qualitative and qualitative variables.

Tab. 3. Correlation of Kendall's Tau b and Spearman's rho between the number of perished and months

Correlations				
			Number of perished	Month
Kendall's tau_b	Number of perished	Correlation Coeff.	1,000	,370**
		Sig. (2-tailed)	.	,000
		N	96	96
	Month	Correlation Coeff.	,370**	1,000
		Sig. (2-tailed)	,000	.
		N	96	96
Spearman's rho	Number of perished	Correlation Coeff.	1,000	,534**
		Sig. (2-tailed)	.	,000
		N	96	96
	Month	Correlation Coeff.	,534**	1,000
		Sig. (2-tailed)	,000	.
		N	96	96
**. Correlation is significant at the 0.01 level (2-tailed).				

From Tab. 3. It we can conclude that in both cases there is a moderate to strong correlation and both factors are statistically significant.

The additional two tests confirm the Pearson Correlation test. We can assume that a link between the death toll and the month of the year exists and is statistically significant.

A histogram of one-dimensional distribution of the groups with the highest and lowest death tolls is presented in fig. 2. Data on the number of deaths by months of the year are combinet into six groups.

 Copyrights© 2020 ISBN 978-954-712-793-7 (Print)

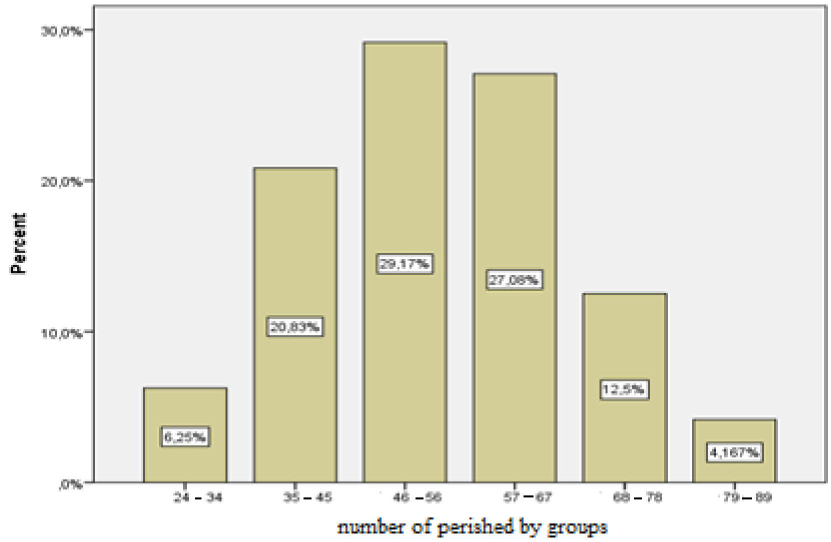

Fig. 2. Number of perished by groups in %

The group 79 to 89 holds high numbers of mortality with the lowest death toll (4%) for a month. The goup with the highest is

The group 46 ti 56 holds medium to high numbers of mortality with the highest death toll (29%) for a month, which is 1/3 of all deaths for the observed period.

Fig. 3. Shows the number of fatal occurrances on the road, distributed in six groups by months of the year. The highest number of deaths occurred in July and August (between 79 to 89 death cases). The lowest death occurrances were in January (between 24 to 34 cases).

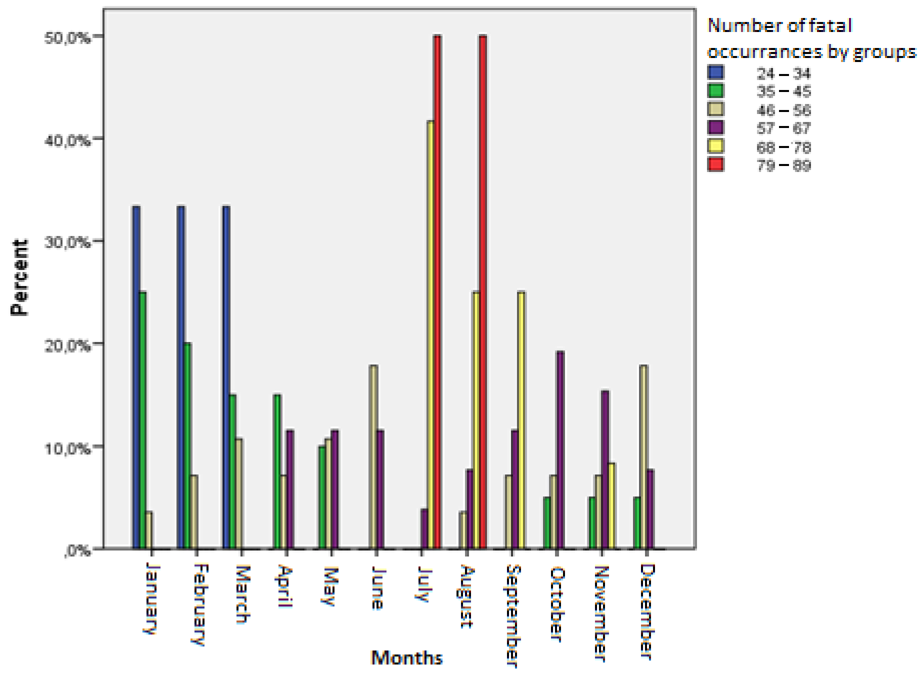

Fig. 3 Months, with the lowest and highest death occurrance rate

A cluster analysis is applied to establish the characteristics of the death toll during the year.

To determine the number and value of certain period applied is the multy cluster two – step analysis.

Figures 4 and 5 present the cluster analysis data, such as gravity, number of units in each cluster, and average death toll for each of the cluster centers.

- **two-cluster two-stage analysis**.

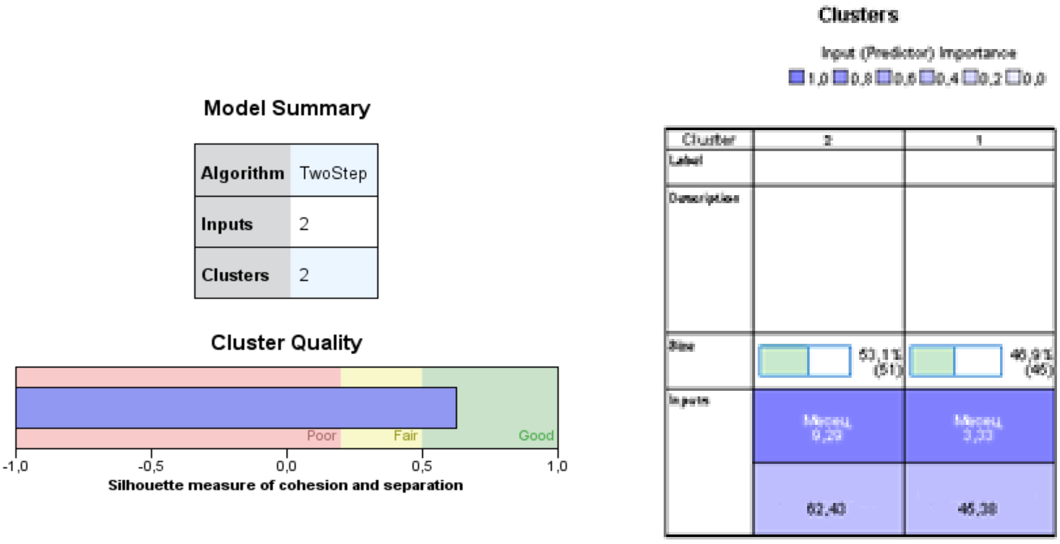

Fig. 4. two-cluster two-stage analysis

The two-cluster analysis is adequate. This is confirmed by a coefficient of contingency (fig. 4.) that is greater than 0,5. The second cluster (2) holds more gravity, as this point is approximately in the beginning of September (interpreted from month scale of 0 to 12) where the average number of fatalities is about 62 people, with 53% gravity. The other cluster point in the beginning of March holds the death toll of about 45 people, or a gravity of 47 %.

- **Three-cluster two – stage analysis**

In this analysis (fig. 5) the cluster center is in the beginning of September with an average number of perished about 63 people and gravity of 49%. The other two are almost equally heavy at 25-26% at the end of April and respectively the beginning of February. Together they hold gravity as almost as much as the third cluster center (3).

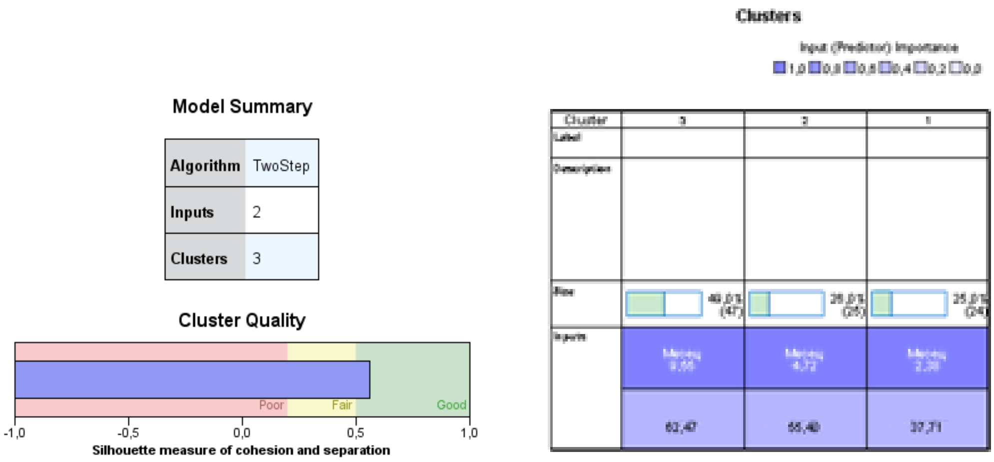

Fig. 5. Three-cluster two-step analysis

The cluster analysis confirms the conclusion that, at the beginning of the year, the number of deaths is lower in the spring it begins to increase and reaches its peak around the end of summer. The cluster analysis also confirms the claims made in fig. 2 and 3.

 Copyrights© 2020 ISBN 978-954-712-793-7 (Print)

CONCLUSION

1. Despite the decrease, the number of read fatalities in the country remains high. For the period 2011 – 2018, there are 654 perished, which is 87,3 per 1 million inhabitants, well above the European average for the same period (49 deaths per 1 million inhabitants).
2. The coefficient of irregularity by months of the year for the stated period is considered 1,56 and the inhomogeneity is 24%, which means that there are differences in the number of deaths in the different months of the year.
3. The data obtained for the values of the different numerical characteristics (tabl. 1.) show that the average number of deaths per month for the last 8 years is 55 from the extracted population sample. The median is 53,5 which means that 50% of the cases the death toll is over 54 people, in the other 50% of the cases is below 53 people. The most common death toll during this period is 44 people. The standard deviation (random scatter) is approximately 13%. The smallest death toll per month is 25 people and the highest in a month is 86 people.
4. The principle of Pearson together with other two criteria of Kendall's Tau-b and Spearman's Rho, confirm that between the two variables, number of the perished and month of the year exist a strong correlation and it is statistically significant.
5. The two-stage multiclustert analysis show that in the beginning of the year the average number of deaths is lower, in spring it begins to increase and reaches its maximum by the end of summer.

REFERENCES

A. Sładkowski, Modeling of the Interaction of the Different Vehicles and Various Transport Modes (Springer International Publishing, 2019

Anderson R., McLean A., Farmer M., Lee B., Brooks C. Vehicle travel speeds and the incidence of fatal pedestrian crashes.1997. Accid Anal Prev. Volume 29, Issue 5, P. 667-674

Centeno, V., Georgiev, I. R., Mihova, V., and Pavlov, V. 2019.Price forecasting and risk portfolio optimization AIP Conference Proceedings 2164, 060006 (2019); https://doi.org/10.1063/1.5130808

Georgiev, N., Damqnov, B., Veliova, V., 2015. Adaptation and application of statistical methods for traffic safety analysis and forecasting. Scientific Journal "Mechanics Transport Communications", Volume 13, Issue 3/1, paper № 1156.

Karacasua, M., Erb, A., Bilgiça, S., Barutb, H. 2011. Variations in Traffic Accidents on Seasonal, Monthly, Daily and Hourly Basis: Eskisehir Case, Scientific Journal "Procedia Social and Behavioral Sciences", Volume 20, p. 767 – 775

Kostadinov, S., D. Lyubenov, P. Atanasova, & F. Kirilov., 2017. Road Traffic Safety Analysis in Ruse Region for the Period 2012 – 2016 Part 2. Proceedings of International Conference Angel Kanchev University of Ruse and Union of Scientists. Ruse. Volume 56. Book 4, P. 122-127.

Komba, D. D. 2006. Risk Factors and Road Traffic Accidents in Tanzania: A Case Study of Kibaha District. Master Thesis, Department of Geography, Norwegian University of Science and Technology (NTNU), Trodheim.

Mitkov, A., Minkov, D., 1989. Statistical methods for evaluation and optimisation of agriculture technologies. Part 1, 2. (**Оригинално заглавие:** Митков, А., Минков, Д., 1989. Статистически методи за изследване и оптимизиране на селскостопанската техника. Част 1, 2)

Pavlov, V., Mihova, V., 2016. Applied statistics with SPSS. (***Оригинално заглавие:*** *Павлов, В., Михова, В., 2016. Приложна статистика с SPSS.*)

Pencheva, V, Tsekov, A., Georgiev, I., Kostadinov, S., 2018. Analysis and assessment of the regularity of mass urban passenger transport in the conditions of the city of Ruse, Transport Problems

Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment. Procedia - Social and Behavioral Sciences, 21, 317–325. doi:10.1016/j.sbspro.2011.07.020

Yue, S., P. Pilon, and G. Cavadias. 2002. Power of the MannKendall and Spearman's Rho Tests for Detecting Monotonic Trends in Hydrological Series. J. Hydrol. 259: 254-271.

Page of Central Trends and Dispersion Measures. https://www.btu.bg/statexcel/file3.html

 Copyrights© 2020 ISBN 978-954-712-793-7 (Print)