ISSN: 1231-4005
e-ISSN: 2354-0133
DOI: 10.5604/01.3001.0010.3140

USE OF REGRESSION ANALYSIS FOR COMPARATIVE EVALUATION OF ACCIDENT HAZARD IN POLAND, BY REGIONS

Leon Prochowski

Military University of Technology
Department of Mechanical Engineering
Kaliskiego Street 2, 00-908 Warsaw, Poland
tel.: +48 261 837 866, fax: +48 22 6837370
e-mail: leon.prochowski@wat.edu.pl

Mirosław Gidlewski

University of Technology and Humanities in Radom
Institute of Operation and Maintenance of Machines and Vehicles
Chrobrego Street 45, 26-600 Radom, Poland
tel.: +48 48 3617657
e-mail: miroslaw.gidlewski@uthrad.pl

Filip Dąbrowski

Automotive Industry Institute
Automotive Expertise and Training Centre
Jagiellonska Street 55, 03-301 Warsaw, Poland
tel.: +48 22 7777144
e-mail: f.dabrowski@pimot.eu

Abstract

Calculations carried out made it possible to characterize changes in the accident hazard, in numerical terms, based on the course of the regression line. An analysis of the distribution, by regions, of the accident hazard and, above all, the showing of changes in this hazard is an important contribution to the actions taken to improve the road transport system. The rate of decrease in the number of accidents undergoes changes, which are particularly conspicuous in some voivodships (provinces) against the background of the general process of decline in the nationwide number of accidents. A graph to characterize this hazard has been presented, plotted with taking into account the changes in the number of accidents, recorded in monthly steps for a period of 7 years. The rate of these changes was also evaluated, in medium-term and short-term intervals. The calculation results were used to ascertain whether a decline in the total number of accidents in Poland is accompanied by similar changes taking place in individual regions. The symptoms of increasing share of the number of accidents that occur in a few regions in the total number of accidents taking place in Poland may be a cause for alarm.

Keywords: *regression analysis, road accidents, accident hazard by regions*

1. Introduction

The accident hazard is predominantly caused by drivers' behaviour [1, 3] and a major improvement in the hazard level may only be achieved by improving the road transport system with taking into consideration the human safety criterion [10]. The knowledge of the distribution of this hazard over the state territory is an important factor in the policy of development of road

transport systems. The numerical data collected about road accidents were analysed to identify and evaluate the differences existing within this scope among individual regions of our country. In this study, a region is understood as the territory of a voivodship (province). To analyse the accident hazard, the monthly distribution of road accidents among the regions was taken into account. Details about the accidents, used as input data for the analysis, were obtained from the General Headquarter of Police.

World Health Organization's reports [2, 5] show a decline in the number of road accidents in many countries of the world; however, the rates of this decline in various countries are not only slow but also differ very much from each other. In Poland, the high number of road accidents is particularly obliging to identify the regions where the accident hazard indicators reach the highest values. In spite of a clearly visible improvement in the road traffic safety, Poland is still among the European countries where the level of this safety is low. The total annual cost of road accidents in Poland is estimated at over 30 billion zlotys [7, 9]. According to calculations carried out for various countries and summarized in report [5], the accident costs are at a level of 1.5-2.5% of the gross domestic product (GDP); in Poland, they amount to about 2% of the GDP.

The interrelations between the regional trends in the accident hazard with the trend in the numbers of accidents are identified with the use one of the known data exploration methods. The main analysis is based on the values of the relative and standardized indicators that significantly reduce the impact of the size of a specific region on the accident hazard measures taken into account. Additionally, the analysis carried out is focused on the evaluation of the rate of changes in the accident hazard, with only little attention being paid to the hazard indicator values. The research has been undertaken to obtain a comparative evaluation of the accident hazard in various regions of Poland, with a requirement that the impact of the region size (area, population size, total length of public roads) on the comparison results should be minimized. The following quantities are calculated:

– slope a_r (a_{10}, a_{14}, a_{15}, a_{16}) of the regression line, characterizing the course of the line, which is described by an equation:

$$\hat{y} = a_r x + b_r , \tag{1}$$

where subscript r defines the year interval, i.e. 2010-2016, 2014-2016, 2015-2016, and the year 2016, respectively, and x is the independent variable, whose values are laid off on the axes of abscissae in the graphs, e.g. in Fig. 1,

– slope a_{Pr} of the regression line of percentage changes in the number of accidents; the percentage changes were calculated in relation to the mean values determined for the corresponding months of the years 2010-2012, considered as the initial period, with the said mean values for specific months being taken as reference levels for the evaluation of changes in the accident hazard in the calendar months of consecutive years,

– coefficient of variation of the values under analysis around the regression line:

$$w_z = \frac{\delta}{\overline{w}} , \tag{2}$$

where:

δ – standard deviation of the values of differences $w = y_{ik} - \hat{y}_{ik}$ and, in turn:

$k = 1, \ldots, m$ indicates consecutive months in a year, with $m = 12$,

$i = 1, \ldots, N$ indicates consecutive years of the period under analysis,

y_{ik} – value of the quantity under analysis in the k^{th} month of the i^{th} year,

$\hat{y}_{ik}$ – value of the regression function, according to equation (1),

$\overline{w}$ – mean value of the differences as above.

Basically, the analysis covers the time history of changes in slope a_r of the regression line determined for the number of accidents, with the regression being calculated for the period

2010-2016 as a medium-term regression and for the period 2014-2016 as a short-term regression. The course of the regression lines is simultaneously treated as a graphic representation of the trend of changes in the values under analysis.

2. Trends in changes in the numbers of road accidents in individual regions

Figure 1 shows, in three parts, time histories of the monthly numbers of road accidents in the regions under analysis, in comparison with a time history of the total monthly number of road accidents in Poland. The courses of these curves reflect a general downward trend in the numbers of accidents in most of the regions as well as annual fluctuations in these figures, related to individual seasons.

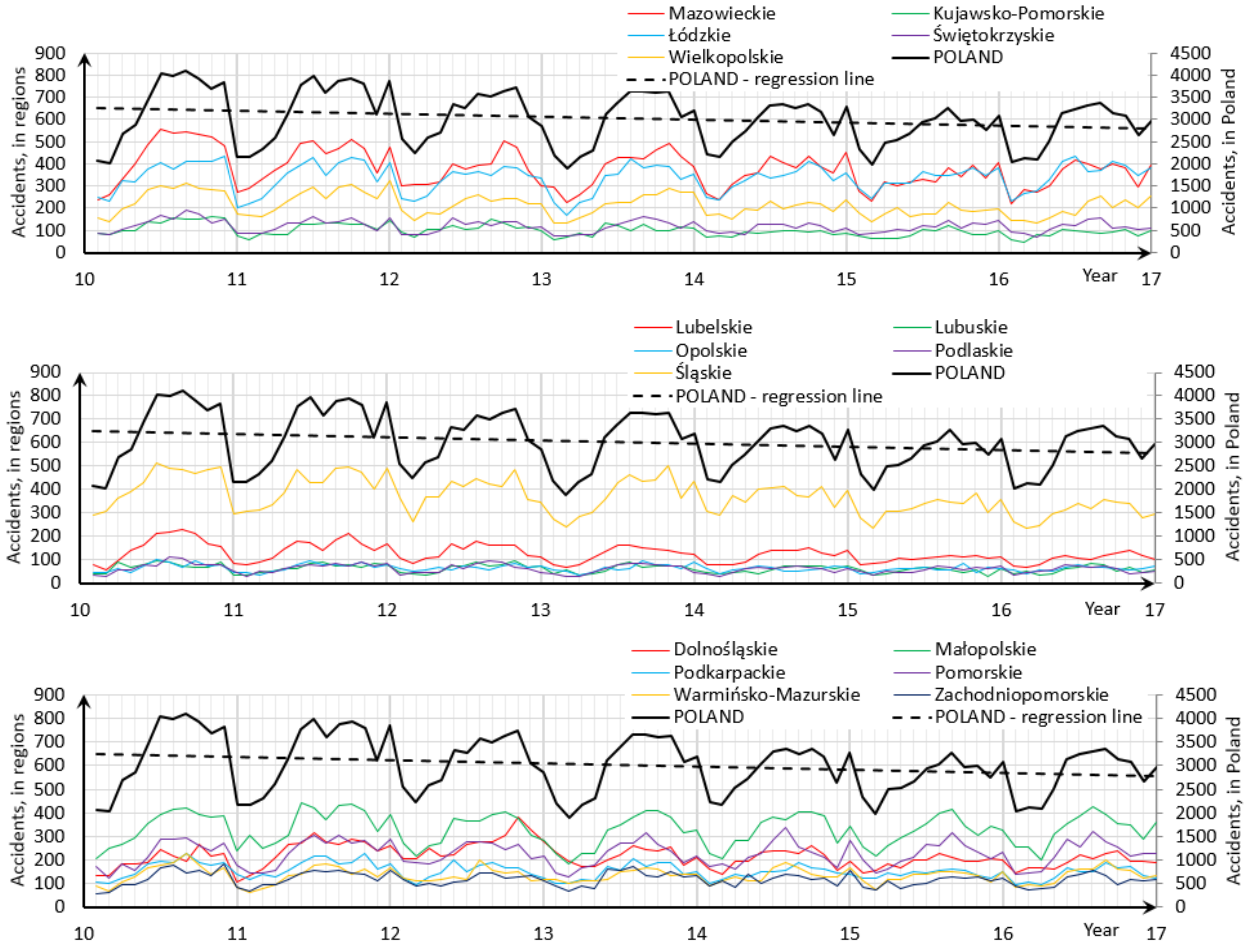

Fig. 1. Summary of time histories of the monthly numbers of road accidents in the years 2010-2016 in individual regions, juxtaposed with a time history of the total monthly number of road accidents in Poland

The numerical data shown in Fig. 1 were taken as a basis for determining regression lines (in the graphs, a regression line has only been plotted for the nationwide number of accidents). The slope of this line has been denoted by a_r, according to (1), and it shows the yearly average trend in the number of accidents (e.g. $a_r > 0$ would indicate an upward trend). The calculated values of slopes a_r of individual regression lines have been presented in the form of a bar graph in Fig. 2.

The graph shows four values of the slope of the regression line for each region:

– a_{10}, determined for the years 2010-2016, vary from −17.2 (for Śląskie Voivodship) to +2.0 (for Lodz Voivodship),
– a_{14}, determined for the years 2014-2016, vary from −20.2 (for Śląskie Voivodship) to +22.0 (for Lodz Voivodship),

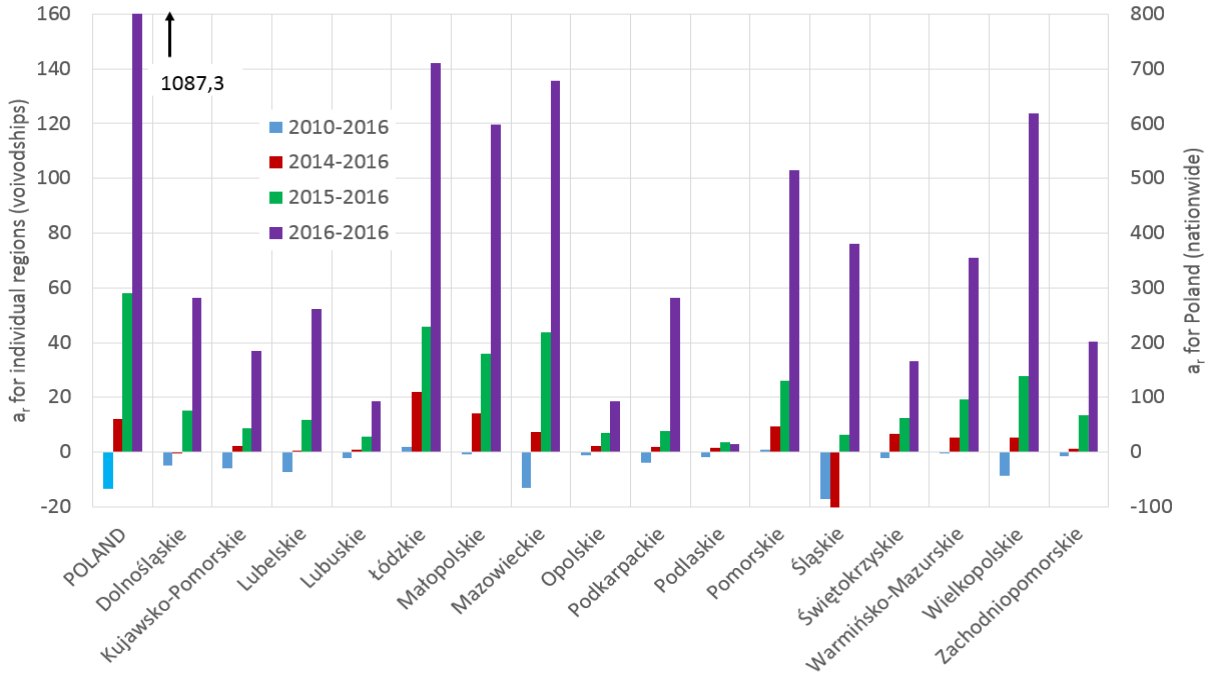

Fig. 2. Summary of the values of slopes a_r of the regression lines plotted for the monthly numbers of accidents that occurred in individual regions and in Poland as a whole

– a_{15}, determined for the years 2015-2016,
– a_{16}, determined for the year 2016.

The values of slopes of the regression lines plotted for the numbers of accidents (Fig. 2, Tab. 1) are a relative measure, which makes it possible to compare the general trend in changes in the monthly numbers of accidents that occurred in individual regions and in Poland as a whole. Simultaneously, they indicate the regions of Poland where the rate of changes in the number of accidents reached the highest values (see the slope values represented by the bars in Fig. 2) in the successive periods under analysis. The juxtaposition of the regression line slope values determined for the period 2010-2016 with the corresponding values for the recent years very clearly revealed the appearance of a strong upward trend in a few regions (Lodz, Małopolskie, Pomorskie, and Mazowieckie Voivodships). The nationwide medium-term trend is downward and such a many-year trend still holds. However, the nationwide trend represents the total changes, while the short-term regression lines show important changes, both in Poland as a whole and in individual regions. Substantial changes can already be seen in Fig. 2 and in Tab. 1 from 2014 on; this is confirmed by a comparison between the slope value of the medium-term regression line with those of the regression lines for the periods 2014-2016 and 2015-2016. As an example, the values of slope a_{10} for the above four regions ranged from −12.9 to 2.0 but slope a_{14} already took unfavourable positive values (from 7.5 to 22.0).

The coefficient of variation in the number of accidents around the regression line shows the concentration of calculation results around this line [11]. The calculated values $w_z < 0.30$ for the period 2010-2016 and $w_z < 0.24$ for the period 2014-2016 confirm that the set under analysis is homogeneous (Table 1) and its mean values (and the regression line) adequately characterize the properties of the analysed process of variation in the number of accidents.

The values given in Tab. 1 show a wide diversity in the accident hazard, taken by regions. In the four regions that predominantly contributed to the total number of accidents in Poland (over 10% each), i.e. Lodz, Małopolskie, Mazowieckie, and Śląskie Voivodships, the following extreme cases took place:

– Śląskie Voivodship, with the highest, i.e. very good, rate of decrease in the number of accidents ($a_{10} = -17.2$ for the years 2010-2016), and

– Lodz Voivodship, where an upward (undesirable) medium-term trend in the number of accidents held for the same years ($a_{10} = 2.0$) and the rate of growth in the number of accidents increased as much as eleven-fold (to $a_{14} = 22.0$) in the period 2014-2016.

Tab. 1. The main values that characterize changes in the slope of the regression line for the number of accidents in the periods 2010-2016 and 2014-2016

Region (Voivodship)	Percentage share of the region in the total number of accidents in Poland	Slope a_{10} of the regression line for the years 2010-2016	Slope a_{14} of the regression line for the years 2014-2016	Coefficient of variation around the regression line for the years 2010-2016	Coefficient of variation around the regression line for the years 2014-2016
Poland	100.00	−67.4	60.9	0.197	0.151
Dolnośląskie	7.16	−4.8	−0.4	0.224	0.154
Kujawsko-Pomorskie	3.32	−5.9	2.4	0.260	0.182
Lubelskie	4.20	−7.4	0.6	0.300	0.198
Lubuskie	2.07	−2.1	0.9	0.276	0.235
Lodz	11.21	2.0	22.0	0.187	0.150
Małopolskie	10.94	−0.9	14.0	0.206	0.192
Mazowieckie	12.52	−12.9	7.5	0.220	0.173
Opolskie	2.15	−1.1	2.3	0.227	0.170
Podkarpackie	5.06	−3.9	1.8	0.211	0.160
Podlaskie	2.07	−1.7	1.7	0.295	0.224
Pomorskie	7.54	0.9	9.4	0.231	0.229
Śląskie	12.31	−17.2	−20.2	0.198	0.144
Świętokrzyskie	3.90	−2.1	6.7	0.230	0.190
Warmińsko-Mazurskie	4.50	−0.3	5.4	0.240	0.222
Wielkopolskie	7.11	−8.5	5.3	0.229	0.163
Zachodniopomorskie	3.94	−1.6	1.4	0.244	0.202

Tab. 2. Values and characterization of the slopes of the regression lines for the number of accidents

Description \ Years	2010-2016	2014-2016	2015-2016	2016
Average value of regression line slope a_{SR} calculated for all the regions	−4.2	3.8	18.2	68.0
Number of the regions where a downward trend was observed, i.e. $a_r < 0$	14	2	0	0
Number of the regions where the downward trend was stronger (* the upward trend was weaker) than the average calculated for all the regions, i.e. $a_r < a_{SR}$	6	7*	10*	9*
Number of the regions where an upward trend was observed, i.e. $a_r > 0$	2	14	16	16
Sum of the slope values a_r calculated for all the regions	−67.4	60.9	291.1	1 087.3

The regression line slope values brought together and presented in Fig. 2 and Tab. 2 indicate the following:

– In the period 2010-2016, a downward medium-term trend prevailed in the number of accidents, which manifested itself in negative values of slope a_{10} in 14 regions and in a value of $a_{10} = -4.2$ for the whole territory of Poland.

– In the years 2014-2016, the short-term trend in changes in the number of accidents already reflected a process of growth in this number in 14 regions; from 2015 on, this process spread to all the regions in Poland.
– The growing trends in the number of accidents that appeared at the end of the period under analysis resulted in the fact that no region with $a_{15} < 0$ was then revealed and there were only 9-10 regions where the rate of growth in the number of accidents was lower than the nationwide average.

3. Percentage presentation of changes in the numbers of road accidents in individual regions

Changes in a specific quantity are often presented in percentage terms in relation to a specific value adopted as an initial or reference level. In this work, the percentage changes in the number of accidents were calculated in relation to the average value determined for the years 2010-2012. Such an approach was dictated by wide fluctuations observed in the number of accidents that happened in consecutive months in individual regions. Thanks to presenting these data as percentage changes relative to the data of the corresponding months of the said initial period, the seasonal fluctuations observed previously (Fig. 1) were minimized. In general, the percentage changes can now be clearly seen to be concentrated around the centre line and the regression line for the whole country (Fig. 3, the heavy solid and dashed line, respectively). Here, an extra subscript P has been added to the symbols introduced previously:

$$\hat{y}_P = a_{Pr}x + b_{Pr}, \tag{3}$$

where a_{Pr} is the slope of the regression line of percentage changes in the number of accidents.

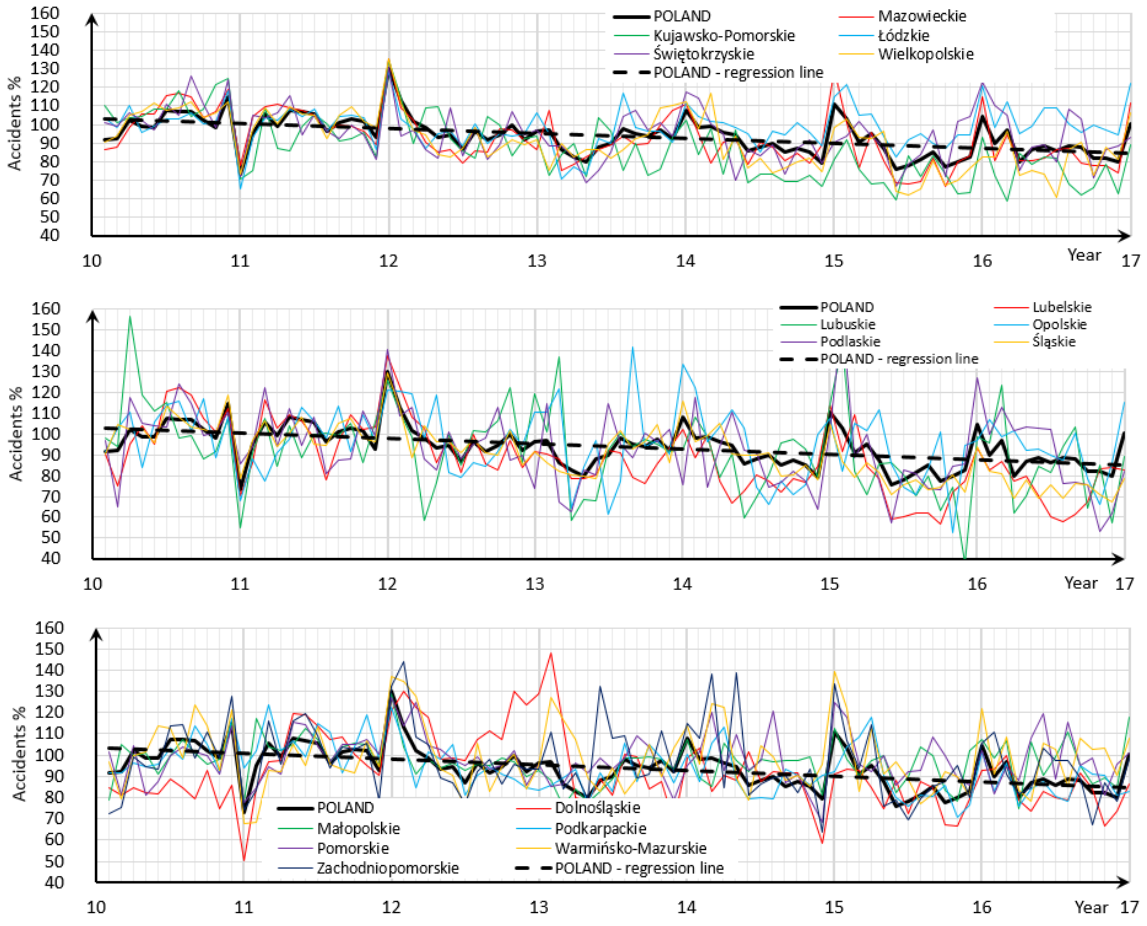

Fig. 3. Time histories of percentage changes in the number of accidents

The distribution of the a_{Pr} values, shown in Fig. 4, applies to the same years as shown in Fig. 2. However, the distribution of these values is completely different: now, negative values predominate, showing a downward trend in the number of accidents in relation to the 2010-2012 reference levels, which can be seen in the course of the regression line. Are these findings in contradiction with the calculation results presented in Fig. 2? No, they are not, because the regression line slope values in Fig. 2 directly show changes in the numbers of accidents in successive periods, while Fig. 4 shows changes in relation to the initial mean value. The course of the regression line in Fig. 3 provides information about the downward medium-term trend in the nationwide data. A similar downward medium-term trend in the a_{Pr} values can be seen in 15 of the 16 regions (see Tab. 3 and 4).

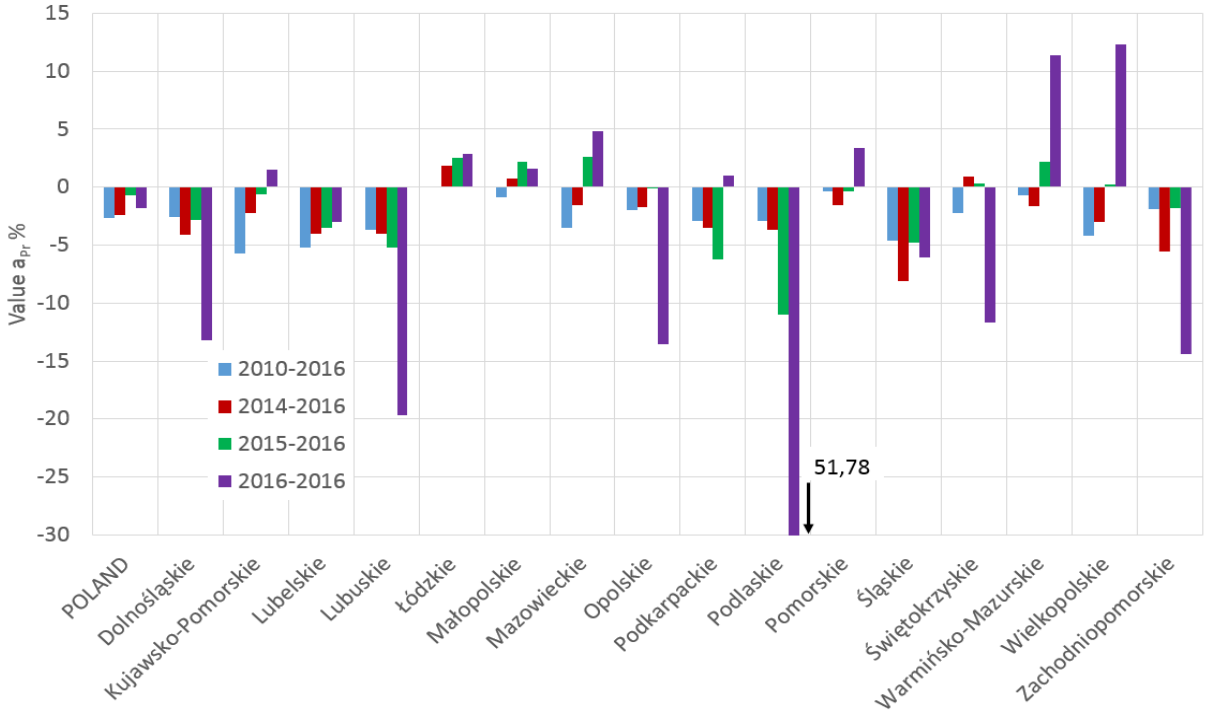

Fig. 4. Summary of regression line slope values a_{Pr} for the percentage presentation of accident numbers

Positive values of the regression line slope indicate an increase in the number of accidents in relation to the medium-term regression line. Simultaneously, the presence of positive a_{Pr} values means an alerting signal about the appearance of unfavourable upward trends in relation to the situation of 2010-2012 in individual regions.

The values presented in Tab. 3 and 4 show, in general and in details, the favourable and unfavourable trends (indicated by negative and positive values of the regression line slope, respectively) in the accident hazard in individual regions. They confirm a decline in the number of accidents in relation to the mean values of the period from 2010 to 2012, observed in 15 regions. However, the values of the regression line slopes for the years 2014-2016 indicate that the rate of decline in the number of accidents is slowing down, which is confirmed by a comparison between the values specified in the 3rd and 4th column of Tab. 3. These changes have been synthesized in Tab. 4.

Based on results of the calculations of percentage changes in the number of accidents in relation to the situation of 2010-2012, the following findings may be formulated:

– in the period from 2010 to 2016, a diminishing downward trend could be observed in the accident hazard, manifesting itself in the values of the regression line slopes increasing in many regions; for Poland as a whole, this slope grew, in average terms, from $a_{P10} = -2.64$ in the medium-term trend to $a_{P15} = -0.68$ in the years 2015-2016,

Tab. 3. The main values that characterize percentage changes in the slope of the regression line for the number of accidents in the periods 2010-2016 and 2014-2016

Region (Voivodship)	Percentage share of the region in the total number of accidents in Poland	Slope a_{P10} of the regression line for the years 2010-2016	Slope a_{P14} of the regression line for the years 2014-2016	Coefficient of variation around the regression line for the years 2010-2016	Coefficient of variation around the regression line for the years 2014-2016
Poland	100.00	−2.64	–2.43	0.107	0.093
Dolnośląskie	7.16	−2.58	−4.10	0.187	0.122
Kujawsko-Pomorskie	3.32	−5.71	−2.22	0.191	0.137
Lubelskie	4.20	−5.23	−4.00	0.192	0.174
Lubuskie	2.07	−3.71	−4.04	0.217	0.230
Lodz	11.21	0.01	1.89	0.108	0.096
Małopolskie	10.94	−0.85	0.71	0.104	0.102
Mazowieckie	12.52	−3.50	−1.52	0.153	0.151
Opolskie	2.15	−1.97	−1.69	0.177	0.172
Podkarpackie	5.06	−2.90	−3.54	0.126	0.123
Podlaskie	2.07	−2.89	−3.69	0.196	0.228
Pomorskie	7.54	−0.40	−1.54	0.119	0.131
Śląskie	12.31	−4.65	−8.10	0.144	0.120
Świętokrzyskie	3.90	−2.19	0.94	0.143	0.144
Warmińsko-Mazurskie	4.50	−0.72	−1.62	0.155	0.152
Wielkopolskie	7.11	−4.16	−3.02	0.159	0.156
Zachodniopomorskie	3.94	−1.88	−5.56	0.175	0.189

Tab. 4. Synthetic characterization of the values of slopes a_{Pr} of the regression lines for percentage changes in the numbers of accidents

Description / Years	2010-2016	2014-2016	2015-2016	2016
Average value of regression line slope a_{Pr} calculated for all the regions	−2.64	−2.43	−0.68	−1.84
Range of the a_{Pr} values in individual regions (min-max)	from −5.71 to +0.01	from −8.10 to +1.89	from −10.98 to +2.61	from −51.78 to +12.32
Number of the regions where a downward trend was observed, i.e. $a_{Pr} < 0$	15	13	10	8
Number of the regions where the downward trend was stronger than the average calculated for all the regions	8	8	7	8
Number of the regions where an upward trend was observed, i.e. $a_{Pr} > 0$	1	3	6	8

– a sharp downward trend favourably held in Dolnośląskie, Lubuskie, Podlaskie, and Zachodnio-pomorskie Voivodships; however, the share of these regions altogether in the total number of accidents in Poland was as low as 15%,
– in 2014-2016, unfavourable trends could be observed in 6 to 8 regions; this particularly applies to Lodz, Małopolskie, and Świętokrzyskie Voivodships, where 26% of all the accidents recorded in Poland occurred.

4. Problem of growing share of a few regions in the total number of accidents in Poland

The calculations carried out also showed that, apart from the general downward trend in the number of accidents in Poland, there were some regions whose shares in the nationwide number of accidents were growing. The calculation results presented in Table 5 show changes in the values of percentage shares of the numbers of accidents that occurred in individual regions in the nationwide number of accidents. The said shares were calculated for several periods. This revealed the history of changes in the values of these shares from the initial state recorded in the period 2010-2012 to the state observed in the most recent years. The results indicating an increase in this share recorded for four regions, i.e. Lodz, Małopolskie, Pomorskie, and Warmińsko-Mazurskie Voivodships cause alarm. On aggregate, about 36% of the nationwide number of accidents took place there. There were two regions where this share grew by 12% (Pomorskie Voivodship) and by 18% (Lodz Voivodship). Simultaneously, this share favourably declined, by about 15-17%, in Kujawsko-Pomorskie, Lubelskie, and Śląskie Voivodships in the years 2010-2016 (see Tab. 5).

Tab. 5. Percentage share of individual regions in the total number of accidents in Poland

Region	2010-2012	2014-2016	2015-2016	2016
Poland	100.0	100.0	100.0	100.0
Dolnośląskie	7.3	6.9	6.9	6.9
Kujawsko-Pomorskie	3.6	3.0	3.1	3.0
Lubelskie	4.5	3.9	3.8	3.8
Lubuskie	2.1	2.0	2.0	2.0
Lodz	10.7	12.0	12.3	12.5
Małopolskie	10.6	11.5	11.7	11.7
Mazowieckie	12.7	12.2	12.1	12.3
Opolskie	2.1	2.2	2.2	2.2
Podkarpackie	5.1	5.1	5.1	5.0
Podlaskie	2.1	2.0	2.1	2.0
Pomorskie	7.2	8.0	8.1	8.1
Śląskie	12.7	11.7	11.3	10.8
Świętokrzyskie	3.9	3.9	4.1	4.0
Warmińsko-Mazurskie	4.3	4.7	4.7	4.8
Wielkopolskie	7.3	6.8	6.8	6.9
Zachodniopomorskie	3.8	4.0	3.9	4.0

5. Recapitulation

The number of road accidents in Poland considerably varies, depending on region. To a significant extent, this arises from different sizes and locations of individual regions and from the actions taken to improve local traffic safety.

The calculation results presented herein highlight another aspect of this issue. In this study, indicators were used whose values only slightly depend on the size of individual regions. To eliminate the impact of the size of the regions under analysis, chiefly the rates of changes in the indicator values were considered and compared with each other, with the current values of the indicators being taken into account only occasionally.

In result of calculations thus arranged, the regions were identified where the indicators characterizing favourable and unfavourable changes in the accident hazard for the local population reached the extreme values. For the purposes of this assessment, chiefly the regression analysis

method was used. The calculated values of the regression line slopes clearly showed the rates of changes in the number of accidents in the years 2010-2016. The medium-term and short-term trends were discerned. The medium-term trend confirms a decrease in the number of accidents, with the data being presented both directly (the highest rate of decline recorded for Śląskie Voivodship) and in percentage terms. However, even the short-term trends determined for the period 2015-2016 showed an increase in the number of accidents in 14 of 16 regions. It is worth adding here that at a general downward trend in the number of accidents; at least four regions were identified where their share in the nationwide number of accidents grew in the years 2010-2016. In two regions, this growth reached a rate of 12% and 18%.

Acknowledgements

The authors would like to express their thanks to the General Headquarter of Police in Warsaw for offering open access to the datasets that made it possible for the authors to carry out the calculations presented in this study.

References

[1] Bąk-Gajda, D., *Dangerous drivers' behavior in road traffic – reasons, results, prevention*, Journal of KONES Powertrain and Transport, Vol. 15, No. 2, 2008.
[2] *Global status report on road safety*, World Health Organization, Switzerland, Geneva 2015.
[3] Guzek, M., Lozia, Z., Zdanowicz, P., Jurecki, R. S., Stańczyk, T. L., *Research on behaviour of drivers in accident situation conducted in car driver simulator*, Journal of KONES Powertrain and Transport, Vol. 16, No. 1, 2009.
[4] Owczarek, A. J., Kowalska, M., *Epidemiologia wypadków komunikacyjnych w Polsce (Epidemiology of traffic accidents in Poland)*, Problemy Higieny i Epidemiologii, 95 (1), 2014.
[5] Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., Colin, M., *World report on road traffic injury prevention*, World Health Organization, Geneva 2004.
[6] Prochowski, L., *Analysis of the impact of road category and day of the week on accident hazard in the road transport of goods in Poland*, Journal of KONES Powertrain and Transport, Vol. 20, No. 1, 2013.
[7] *Rachunek opłacalności działania na rzecz brd (Cost-effectiveness of actions taken to improve road traffic safety)*, World Bank Report for Poland, 2012, /www.worldbank.org/pl/road-safety-in-poland/.
[8] Spoerri, A., Egger, M., von Elm, E., *Mortality from road traffic accidents in Switzerland*, Accident Analysis and Prevention, 43, 2011.
[9] Urbanek, P., Walińska, E., *Ekonomia i nauki o zarządzaniu w warunkach integracji gospodarczej (Economy and theories o management in the conditions of economic integration)*, Vol. 9, Jaździk-Osmulska, A. (ed.): *Wycena kosztów wypadków i kolizji drogowych na sieci dróg w Polsce na koniec roku 2013 (Estimation of the costs of road accidents and collisions*. Wydawnictwo Uniwersytetu Łódzkiego, 2014.
[10] Woropay, M., Bojar, P., *Structuring of the safety evaluation model for the road transport system*, Journal of KONES Powertrain and Transport, Vol. 15, No. 3, 2008.
[11] Zaliaś, A., *Metody statystyczne (Statistical methods)*, PWE, Warszawa 2002.