**UDC 656.13** **DOI 10.52928/2070-1616-2025-51-1-77-85**

# EVALUATION OF ANALYSES OF THE INFLUENCE OF VARIOUS FACTORS ON THE ACCIDENT RATE PEDESTRIAN ACCIDENTS AT PEDESTRIAN CROSSINGS THROUGH THE USE OF STATISTICA TOOLS

***PhD in Engineering D. KHODOSKIN, V. DAVHULEVICH, I. AFANASENKA***
***(Belarusian State University of Transport, Gomel)***

*Road traffic accidents (RTAs) with victims account for only about 5% (!) of the total number of accidents, but, especially at pedestrian crossings, are characterised by high severity and correspondingly high values of accident losses[1] [1; 2]. Thus, it is necessary to constantly assess the influence of the most significant factors and characteristics, such as traffic and pedestrian load, geometric parameters of the pedestrian crossing itself and others, most significantly affecting the accident rate, based on which to develop and implement the most effective measures. It is for this purpose that the article describes the mechanism for identifying such factors and characteristics, determined through a set of features provided by the modern version of the Statistica package.*

***Keywords:*** *accident rate with pedestrians, pedestrian crossing, share of reduction in the number of accidents, cluster analysis.*

**Introduction.** The main purpose of the publication is to establish a list of independent variables significantly affecting the values of accident rates and to develop proposals on the basis of this to reduce the number and severity of accidents involving pedestrians, as well as to establish specific values of the shares of reduction in the number of accidents using the tools of the *Statistica* package.

**Main part.** According to the analysis of the accident rate, the largest number of accidents occur due to violations of traffic rules by drivers of vehicles. But it should be noted that there are other very significant causes of accidents: violations of traffic rules by pedestrians, poor road conditions, technical malfunction of vehicles, etc.

The most common causes of road accidents committed by drivers according to the analysis of accident rate for 2021 and 2022 in Gomel are still: mismatch of speed with actual traffic conditions; lack of the right to drive a vehicle; driving on the oncoming traffic lane; driving under the influence of intoxicants; wrong choice of distance; violation of rules of passing a pedestrian crossing; operation of technically defective vehicles; driving at a prohibited traffic signal.

The most common causes of road accidents committed by pedestrians according to the analysis of accident rate for 2021 and 2022 in Gomel are: crossing at an undetermined place, walking along the carriageway, crossing in front of a closely following vehicle, drunkenness.

The research conducted allowed us to identify a number of factors affecting the total number of road accidents and the number and severity of casualties. Both temporal (month, hour) and systematic (weather conditions, etc.) factors were analysed. Of these, the ones that really affect the outcome were highlighted.

Quantitative analysis of accident statistics for the period under study allowed us to draw the following general conclusions:

– there is a certain dependence of the accident rate in the distribution of accidents by seasons (months): the number of accidents is higher in spring and summer than in winter, namely in April-May (in 2021 – 1,773 accidents, in 2022 – 1,976 accidents);

– the highest daily accident peaks are in the morning and evening hours, which is associated with travelling to work and school. The morning rush hour has the highest number of crashes, with 1,569 crashes in 2021 and 1,882 crashes in 2022. In the morning hours, traffic participants are characterised by motivation (rush to work, study), low level of attention and, of course, high intensity of traffic;

– the largest number of accidents is due to: collisions in the traffic flow when approaching the intersection and collisions of vehicles directly at the intersection;

– the accident rate is significantly influenced by such factors as weather conditions, the presence of structures (obstructing visibility), the condition of the road surface, the degree of illumination of the pedestrian crossing at night.

– in 2021 and 2022, about 40 and 55 per cent of accidents respectively are road traffic accidents with material damage.

Table 1 shows data on accidents with victims at the studied pedestrian crossings (more than 100) in Gomel for 2021. For 2021 in Gomel, out of the whole range of pedestrian crossings studied, accidents involving pedestrians occurred at only 7 crossings (all with injuries).

[1] Ходоскин Д.П. Снижение аварийности на подходах к регулируемым перекресткам путем управления движением попутных транспортных средств: дис. … канд. техн. наук: 05.22.10. – Минск, 2023. – 163 л.

Table 1. – Information on accidents with victims at pedestrian crossings for 2021

| Date & Time | Street | House | Number of injured persons |
|---|---|---|---|
| 22.09.2021 in 20:55 | Б. Tsarikova | 57a | 1 |
| 31.12.2021 in 07:50 | Vladimirova | 16b | 1 |
| 11.07.2021 in 13:42 | Lenina – Krasnoarmeyskaya | 8 | 1 |
| 04.02.2021 in 17:53 | Irininskaya | 25 | 1 |
| 15.03.2021 in 16:00 | Katunina | 20 | 1 |
| 23.06.2021 in 21:30 | Belogo – Kalennikov | 46a | 1 |
| 21.06.2021 in 22:05 | Sovetskaya – Fedyuninskogo | 248 | 1 |

*Note*: There were no road accidents with fatalities.

Directly, the data in Table 1 is presented on the city map (Figure 1).

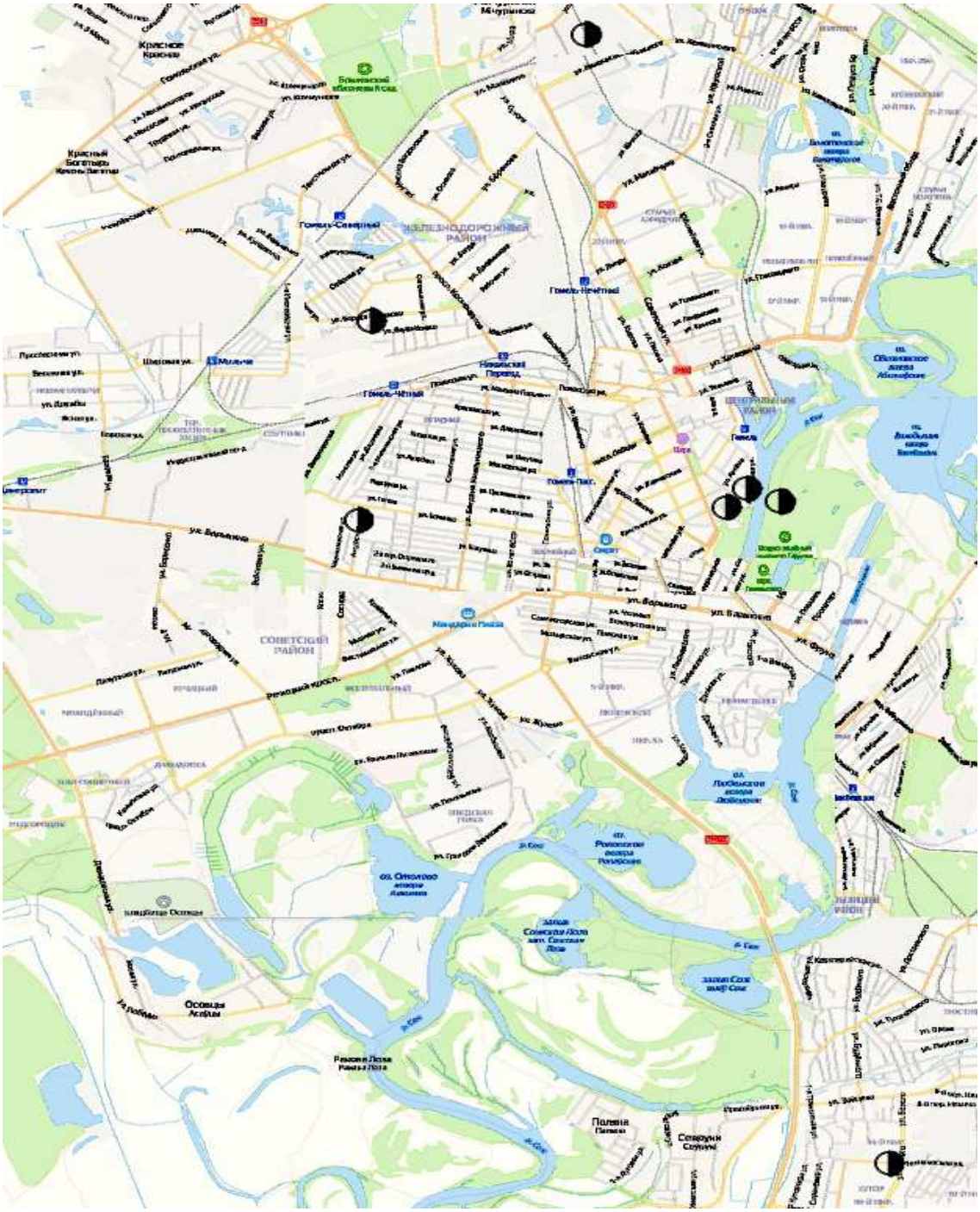

**Figure 1. – Map of topographic analysis of accident rate in Gomel for 2021**

Table 2 shows data on accidents with victims at the studied pedestrian crossings (more than 100) in Gomel for 2022.

Table 2. – Information on accidents with victims at pedestrian crossings for 2022

| Date & Time | Street | House | Number of injured persons |
|---|---|---|---|
| 24.12.2022 in 17:49 | Sovetskaya – Hataevicha | 63/4 | 1 |
| 24.08.2022 in 21:40 | Sovetskaya | 108 | 1 |
| 10.06.2022 in 23:00 | Barykin | 206 | 1 |
| 16.12.2022 in 21:20 | Mazurova – Golovatsky | 59/2 | 1 |
| 12.09.2022 in 07:50 | Pine | 10 | 1 |
| 22.10.2022 in 11:10 | Zhukova | 20 | 1 |
| 26.01.2022 in 07:10 | Ilyicha – Leningradskaya | – | 1 |

*Note*: There were no road accidents with fatalities.

Directly, the data in Table 2 is presented on the city map (Figure 2).


**Figure 2. – Map of topographic analysis of accident rate in Gomel for 2022**

In 2022, also in Gomel, accidents occurred at 7 pedestrian crossings (from the surveyed range). Also, all the road accidents are with injuries.

Topographical analysis of road accidents for the study period shows that for any settlement traditional areas of concentration of accidents on the street and road network are, first of all, intersections. Therefore, further improvement of traffic safety at intersections, especially in relation to unprotected road users, significantly reduces the overall level of road accidents, including their severity.

At the next stage, a universal integrated system for statistical analysis, data visualisation and development of user applications – *Statistica* – was used to identify specific parameters and characteristics that have a greater impact on the accident rate. *Statistica* is a modern package, in which all the latest computer and mathematical methods of statistical data analysis are implemented, that is the most dynamically developing statistical package and the world leader in the market of statistical software.

This system has the following generally recognised advantages: contains a full set of classical and advanced methods of data analysis; easy to master by a trained user; fully compatible with applications of the Windows operating system; is a tool for building applications in specific areas; *Statistica* data can be easily converted to various databases and spreadsheets; the package includes specially selected examples that allow you to systematically master the methods of analysis; supports most Internet formats; supports high quality graphics, allowing you to effectively visualise data and perform graphical analysis; contains a programming language that allows you to extend the system and run it from other Windows applications, etc.

In terms of its capabilities, the *Statistica* package allows you to: build various graphs: histograms, scatter plots, pie charts, build 3D and other graphs; calculate probability, mean and other basic parameters, plot various distributions using a probability calculator; build a Pareto chart; build a cause and effect diagram; build control charts; conduct cluster analysis; conduct non-linear estimation – regression analysis; conduct correlation analysis; calculate statistical charts; conduct time series analysis; organise the analysis using other statistical methods used in industry (and other areas) for data processing [3].

*Application of various methods of working with statistical data in Statistica package.*

The study identified 4 groups of initial data (totalling over 75 characteristics) describing each pedestrian crossing:

1) geometric parameters;
2) traffic and pedestrian load;
3) traffic conditions;
4) traffic light control parameters.

In general, the data included a wide range of parameters such as: crossing width, number and width of lanes, angle of the crossing to the edge of the carriageway, type of crossing, presence of an artificial bump, presence and width of a safety island, speed limit (on the approach to the crossing), traffic volume (by lane and by direction), pedestrian volume (by direction), presence of pedestrian signage, dimensions of side visibility triangles (from each side of the roadway), transparency of side visibility triangles (on each side of the crossing), height of the pedestrian crossing kerb, presence and length of pedestrian barriers installed on both sides, radius of carriageway edge roundings (on each side), distance from the end of the carriageway edge rounding to the edge of the crossing (on both sides), distance from the edge of the crossing to the edge of the nearest lighting support (on both sides), number of working traffic light control programmes, duration of the control cycle (of each of the envisaged programmes), duration of the pedestrian cycle in the traffic light control programme (pedestrian tact), duration of the green pedestrian signal, etc.

In this case, the dependent variable, directly showing the level of accidents at the pedestrian crossing, is the number of recorded accidents involving pedestrians, as well as the number of dead and injured pedestrians in accidents. The total sample size for the study period was more than 270 pedestrian crossings.

Given the heterogeneity and large amount of raw data, the following data mining techniques were applied to achieve the objective: neural network, support vector method, tree bousting, etc.

The factors shown in Table 3 were found to have a relevant impact on the number of road accidents.

Table 3. – The most relevant factors identified during the analysis

| Factors | Best predictors for continuous dependent variable: total road traffic accidents (initial data). Method for selecting relevant variables: feature screening | |
|---|---|---|
| | *F*-value | *p*-value |
| 13) width of the 4th lane, m | 4,773918 | 0,029784 |
| 24) traffic intensity – $N_{tr1}$ auto/hour | 2,707303 | 0,014461 |
| 28) traffic intensity (present value) – $N_{right}$, items/hour | 2,353892 | 0,041104 |
| 69) duration of the transition interval with the previous phase, sec | 2,344816 | 0,055183 |
| 22) traffic intensity – $N_{right}$, auto/hour | 2,281882 | 0,047039 |
| 30) traffic intensity (present value) – $N_{tr1}$, items/hour | 2,251799 | 0,038965 |

*Note*: in the column "Factors" the indicated number denotes the number of the parameter in order, which was adopted during the research.

Variables 22 and 28 are highly correlated with each other (correlation coefficient is 0,98), so to exclude the effect of multicollinearity we should leave variable 28 – traffic intensity – $N_{right}$, as it takes into account the traffic flow structure. Similarly with variables 24 and 30. Variable 30 – traffic intensity – $N_{tr1}$ should be left.

Variable 13 is the width of the 4th lane. According to observations, this variable takes 2 values: "0" – when there is no fourth lane and "3,5 m" – when there is a lane. Therefore, this variable was recoded into a categorical variable: it actually shows the presence/absence of the fourth lane.

Variable 69 is the duration of the transition interval with the previous traffic light cycle phase (for traffic light cycle phase No. 3), sec. For traffic light objects where there is no traffic light cycle phase No. 3 is set to "0". Therefore, the use of this variable for analysis is incorrect. It was excluded (others were similarly excluded). Thus, the remaining relevant variables are summarised in Table 4.

Table 4. – Final list of relevant factors

| Factors | Best predictors for continuous dependent variable: total road traffic accidents (initial data). Method for selecting relevant variables: feature screening | |
|---|---|---|
| | *F*-value | *p*-value |
| 13) width of the 4th lane, m | 4,773918 | 0,029784 |
| 28) traffic intensity (present value) – $N_{right}$, items/hour | 2,353892 | 0,041104 |
| 30) traffic intensity (present value) – $N_{tr1}$, items/hour | 2,251799 | 0,038965 |

Variables 30 and 28 characterise traffic intensities along different directions. They are difficult to manage because they depend on many factors. Therefore, the whole set of pedestrian crossings under study was divided into clusters according to variables 13, 28 and 30. The characteristics of the clusters are summarised in Table 5 and Figure 3.

Table 5. – Selected clusters

| Final classification – Clusters | 13) width of the 4th lane, m | 28) traffic intensity (present value) – $N_{right}$, items/hour | 30) traffic intensity (present value) – $N_{tr1}$, items/hour |
|---|---|---|---|
| 1 | 0 | 0–558 | 0–500 |
| 2 | 3,5 | 90–121,6 | 599,6–1600 |
| 3 | 0 | 624–1240 | 1045,2–1644 |
| 4 | 0 | 0–415,8 | 507,2–1468 |

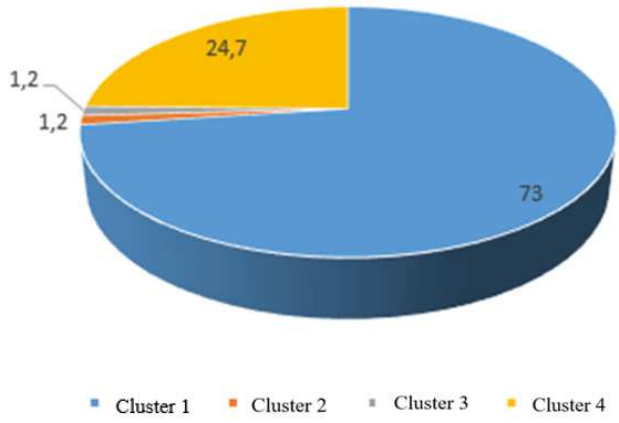


**Figure 3. – Distribution of clusters**

Based on the subject of the study and the obtained results of cluster distribution, cluster 4 is further adopted for consideration. This cluster includes 68 objects under study. Then the variable "Cluster" was added to the table with other data for analysis.

Figure 4 shows a diagram of the distribution of pedestrian crossing accidents for cluster 4.

Thus it is revealed that the most dangerous are 2 pedestrian crossings, which are presented in Table 6.

Further analyses similar to those described above were carried out separately for the cluster 4 under consideration and the relevant factors at this stage were considered, which are presented in Table 7.

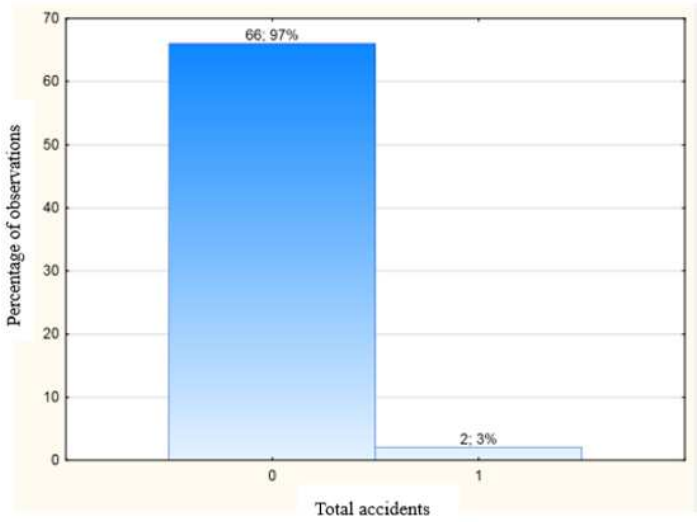


**Figure 4. – Diagram of the distribution of accidents at pedestrian crossings: "0" – no accidents, "1" – there are accidents**

Table 6. – Most dangerous pedestrian crossings

| Street, house | Name of intersection / street section |
|---|---|
| 206, Barykina str. | intersection |
| Vladimirova str. | street section, 16b, Vladimirova str. |

Table 7. – Relevant factors for cluster 4

| Factors | Best predictors for continuous dependent variable: total road traffic accidents (cluster 4). Method for selecting relevant variables: feature screening | |
|---|---|---|
| | *F*-value | *p*-value |
| 6) the presence of an artificial bump | 11,58677 | 0,001134 |
| 23) traffic intensity – $N_{left}$ auto/hour | 3,91904 | 0,003759 |
| 67) the duration of the transition interval with the previous traffic light cycle phase for work programme 1, sec | 3,40840 | 0,003935 |

The influence of factor 67 – the duration of the transition interval with the previous phase for work programme 1 – is presented in Figure 5.

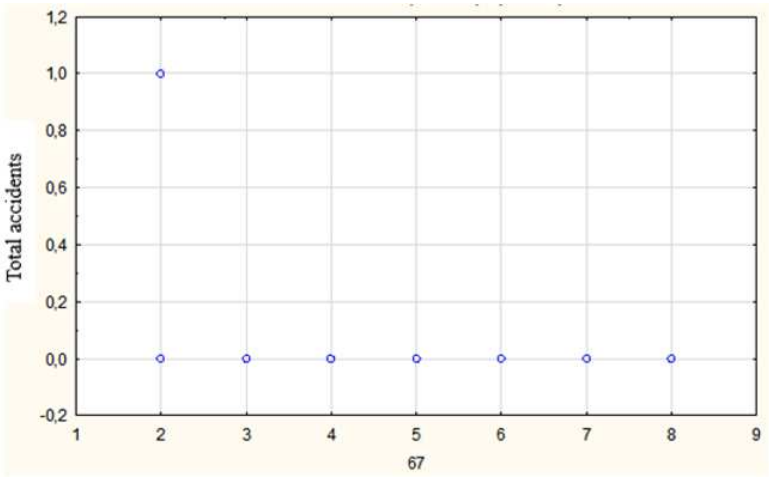


**Figure 5. – A diagram of the impact on the accident rate for the factor "duration of the transition interval with the previous traffic light cycle phase for work programme 1": x-axis - duration of the interval, sec**

The diagram shows that there should be at least 3 seconds of transition interval between switching off the green signal for vehicles in the previous traffic light cycle phase and switching on the green signal for pedestrians in the next traffic light cycle phase.

Further, for example, the influence on pedestrian crossings of cluster 4 of such factors as: the presence of traffic light regulation at the pedestrian crossing, limitation of speed limits to 50 and 40 km/h at the approach to them, the presence of a safety island at the pedestrian crossing and pedestrian barriers on the left, right or on the dividing strip is considered.

The impact on the accident rate of the factor "presence of traffic light regulation at the pedestrian crossing" is presented in Figure 6.

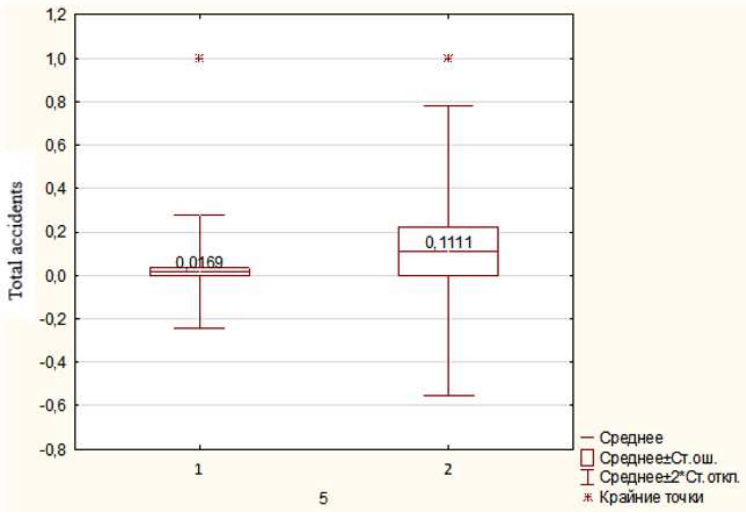


**Figure 6. – Influence diagram for the factor "presence of traffic light regulation at the pedestrian crossing": "1" – yes, "2" – no**

The diagram shows that the presence of traffic light regulation at pedestrian crossings in cluster 4 reduces pedestrian accidents by an order of magnitude (Δ=0,85, i.e. (0,1111 – 0,0169)/0,1111≈0,85).

The effect on the accident rate of the parameter "speed limit" for controlled intersections (pedestrian crossings) is presented in Figure 7.

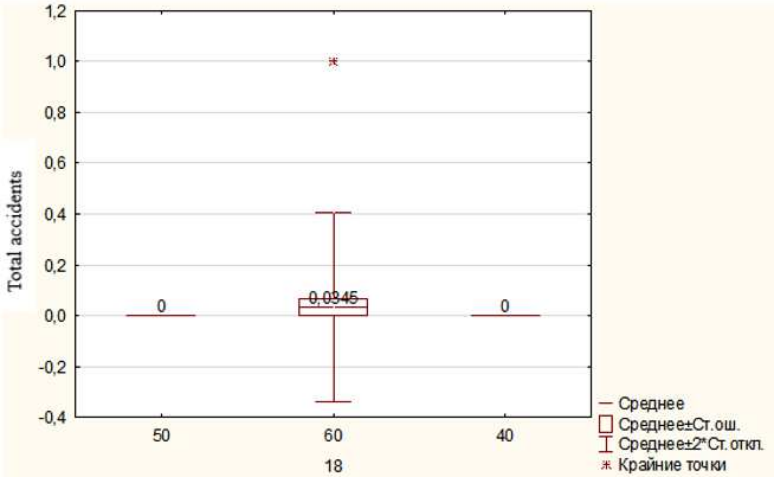


**Figure 7. – Diagram for the parameter "speed limit"**

The diagram shows that speed limitation from 60 km/h to 50 or 40 km/h on the approaches to controlled intersections reduces pedestrian accidents to 0 (Δ = 1).

The crash magnitude diagram for the parameter "presence of safety island" at regulated intersections (pedestrian crossings) is presented in Figure 8. The presence of a safety island reduces the number of pedestrian accidents at such regulated intersections to 0 (Δ = 1).

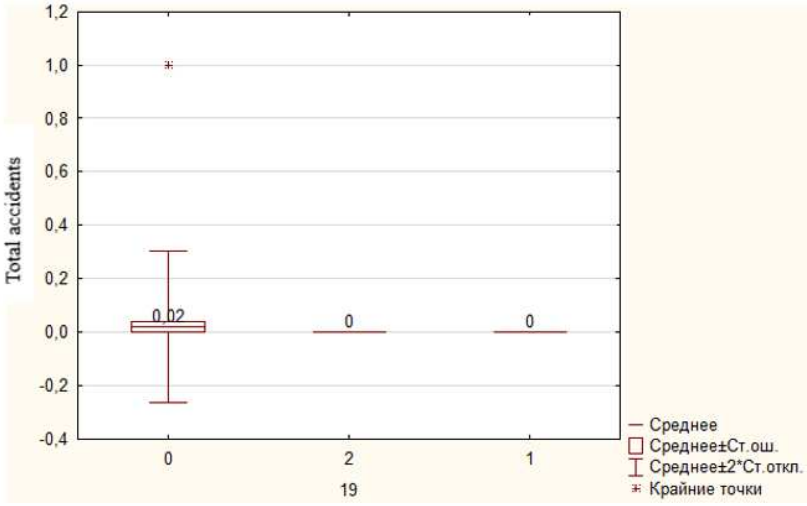


**Figure 8. – Diagram of the spread for the parameter "presence of a safety island": "0" – none, "1" – structurally allocated, "2" – gap in the dividing strip**

The crash magnitude diagram for the parameter "presence of pedestrian barriers (left, right or on the dividing lane)" at controlled intersections is presented in Figure 9.

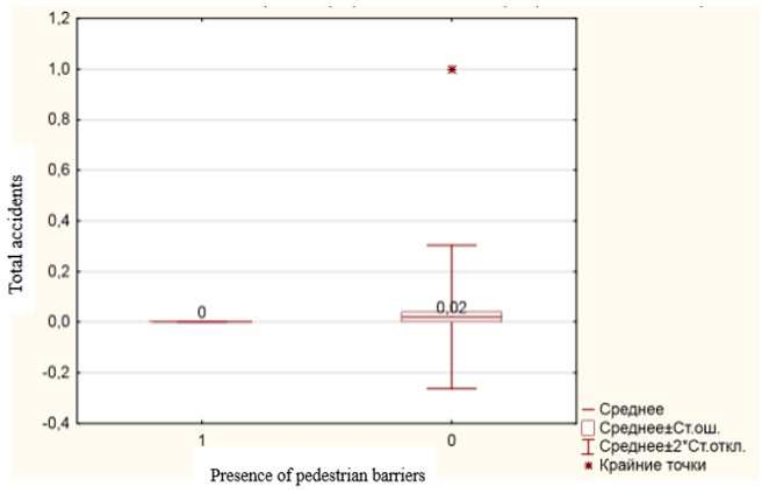


**Figure 9. – Diagram of the spread for the parameter "presence of pedestrian barriers (left, right or on the dividing lane)": "0" – none, "1" – yes**

As can be seen, the presence of pedestrian barriers reduces the number of accidents with pedestrians at such regulated intersections (pedestrian crossings) to 0 ($\Delta = 1$).

**Conclusion.** Analysing the capabilities of various statistical methods of the specialised *Statistica* package as applied to the processing of road traffic parameters we can draw the following conclusions:

1) the application of the topographic type of accident analysis and the methods of the research package showed almost identical results;

2) a specific list of independent variables that significantly influence the values of pedestrian accident rates at pedestrian crossings was established;

3) a specific list of the most accidental pedestrian crossings was established;

4) it seems possible to establish how any intervention affects the number of road traffic accidents (upwards or downwards);

5) it seems possible to establish a specific proportion of the reduction in road traffic accidents with the introduction of any intervention (included in the scope of the study);

6) a certain proportion of the reduction in the number of crashes (see point 5) can then be used in accident forecasting.

REFERENCES

1. Azemsha, S.A. & Shust, O.N. (2022). Razrabotka meropriyatii po snizheniyu kolichestva i tyazhesti dorozhno-transportnykh proisshestvii s uchastiem peshekhodov v g. Gomele [Development of Measures to Reduce the Numbers and Severity Road Accidents with Pedestrian Participation in Gomel]. *Vestnik Polotskogo gosudarstvennogo universiteta. Seriya V, Promyshlennost'. Prikladnye nauki [Herald of Polotsk State University. Series B, Industry. Applied Science]*, (10), 51–58. (In Russ., abstr. in Engl.).
2. Azemsha S.A., Starovoitov A.N. (2017). *Primenenie nauchnykh metodov v povyshenii bezopasnosti dorozhnogo dvizheniya.* – Gomel': BelGUT. (In Russ.).
3. Borovikov, V.P. (2013). *Populyarnoye vvedeniye v sovremennyy analiz dannykh v sisteme STATISTICA*. Moscow: Goryachaya liniya. Telekom. (In Russ.).

ЛИТЕРАТУРА

1. Аземша С.А., Шуст О.Н. Разработка мероприятий по снижению количества и тяжести дорожно-транспортных происшествий с участием пешеходов в г. Гомеле // Вестн. Полоц. гос. ун-та. Сер. В, Пром-сть. Приклад. науки. – 2022. – № 10. – С. 51–58.
2. Аземша С.А., Старовойтов А.Н. Применение научных методов в повышении безопасности дорожного движения. – Гомель: БелГУТ, 2017. – 191 с.
3. Боровиков В.П. Популярное введение в современный анализ данных в системе STATISTICA. – М.: Телеком, 2013. – 288 с.

*Поступила 20.11.2024*

## ОЦЕНКА АНАЛИЗА ВЛИЯНИЯ РАЗЛИЧНЫХ ФАКТОРОВ НА АВАРИЙНОСТЬ С ПЕШЕХОДАМИ НА ПЕШЕХОДНЫХ ПЕРЕХОДАХ ПОСРЕДСТВОМ ИСПОЛЬЗОВАНИЯ ИНСТРУМЕНТАРИЯ ПАКЕТА STATISTICA

***канд. техн. наук Д.П. ХОДОСКИН, О.А. ДОВГУЛЕВИЧ, И.С. АФАНАСЕНКО***
***(Белорусский государственный университет транспорта, Гомель)***

*Дорожно-транспортные происшествия (ДТП) с пострадавшими составляют порядка только лишь 5% от их общего числа, однако, особенно на пешеходных переходах, характеризуются высокой тяжестью и соответственно большими значениями аварийных потерь. В связи с этим необходимо в постоянном режиме производить оценку влияния наиболее значимых факторов и характеристик, таких как транспортно-пешеходная нагрузка, геометрические параметры самого пешеходного перехода и других, наиболее значимо влияющих на аварийность, основываясь на которой разрабатывать и внедрять наиболее эффективные мероприятия. Именно для этой цели в статье описан механизм выявления названных факторов и характеристик, определенных посредством комплекса возможностей, предоставляемых современной версией пакета Statistica.*

***Ключевые слова:*** *аварийность с пешеходами, пешеходный переход, доля снижения числа ДТП, кластерный анализ.*