# PROBABILISTIC MODEL FOR ASSESSING ACCIDENT RATES

**Anastasia Shevtsova[1]\*, Alexander Novikov[2], Sergey Evtyukov[3], Alexey Marusin[3]**

[1] Belgorod State Technological University named after V.G. Shukhov, Belgorod, Russia
[2] Orel State University named after I.S. Turgenev, Orel, Russia
[3] Saint Petersburg State University of Architecture and Civil Engineering, Saint Petersburg, Russia
* shevcova-anastasiya @mail.ru

When working with accident rates, a specialist has to spend quite a lot of time to establish the main places of accidents, certain conditions in which they occurred, which is extremely necessary when determining measures aimed at reducing road accidents. As a result of the research, the authors processed a large amount of data - accident rates for 2015-2021, as a result of which certain dependencies were established between the considered indicators in relative data, which made it possible to develop a probabilistic model for calculating the necessary data with the ability to determine the required conditions. Based on the results obtained, the authors developed an algorithm, according to which procedures were determined when working with accident rates and an assessment of efficiency based on the calculation of the error was carried out. The results obtained allowed us to conclude that the developed probabilistic model and algorithm are effective, in view of the minimum error.

Keywords: accident rates, statistical databases, relative indicators, dependencies, probabilistic model, algorithm, efficiency

## 1 INTRODUCTION

For the effective implementation of measures aimed at improving the safety of the road network, it is extremely important to evaluate the primary data, which include accident rates. The public database in Russia [1] makes it possible to assess the accident rate for the country, as well as to determine the number of accidents outside urban and urban areas, as well as on a certain day of the week, but it is not possible to assess the causes and conditions in this case. In this case, a point analysis of accident rates is required, which involves working with each individual traffic accident. Such work is quite laborious and takes a lot of time. It should be noted that the accident rates in the statistical databases are presented for a long-term period, which in general allows us to assess the dynamics of changes in the indicators under consideration in the country and draw a conclusion about the effectiveness of the measures taken in the field of road safety. Over the years, a large number of studies in this area are based on these indicators [2-11], which consider various approaches to assessing both the number of incidents and the injured and injured in them. Within the framework of this article, a study was made of accident rates in Russia for the period 2015-2022. and a probabilistic model is proposed to evaluate the indicators under consideration. The main purpose of the study is to develop a model that allows to estimate accident rates with a certain accuracy. The main objectives of the study: 1. Perform an analysis of the number of accidents in Russia for the period 2015-2021; 2. Perform an analysis of the number of accidents in Russia in cities and beyond for the period 2015-2021; 3. Perform an analysis of the number of incidents by day of the week in Russia for the period 2015-2021; 4. Process the results obtained using a probabilistic model; 5. Develop an algorithm for processing the results obtained and verify it against the available accident rates for 2022.

## 2 METHODOLOGY

### 2.1 Analysis of accident rates

In the available accident rates presented in official sources, an analysis was made of the accident rates for the period 2015-2021, according to the following components:

- total number of traffic accidents (fig. 1);
- the number of incidents in the city (fig. 2 a);
- the number of incidents outside the urban area (fig. 2 b);
- the number of incidents depending on the day of the week (fig. 3).

During the period under review 2015-2021, 1,137,987 road accidents occurred in Russia. The current global trend aimed at striving for zero deaths on the roads, which is described in detail in the works [12-17], is reflected in legislative acts aimed at improving road safety, which also affects the accident rate, namely, reducing the death rate in accidents.
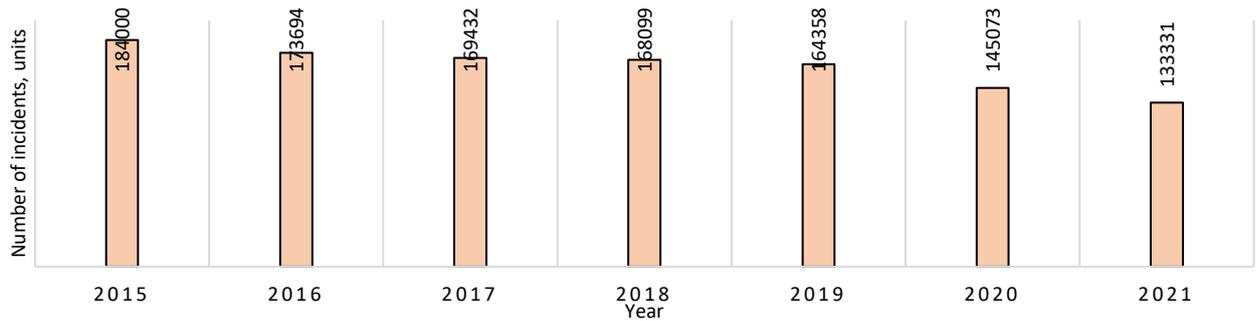
Fig. 1. Histogram of the number of road accidents in Russia for the period 2015-2021

In Russia, during the period under review, there is a systematic decrease in the indicator under consideration, which is explained by the implementation of a number of works aimed at improving road safety, which are being implemented today [18-21].

Another of the main indicators is the number of accidents in the city and beyond, these data also allow us to assess where the largest number of accidents occur, the analysis shows that in Russia, most accidents occur in settlements (fig. 2 a), fewer falls on areas outside settlements (fig. 2b). Despite a certain difference in the permitted speeds of movement, most accidents occur in cities. In Russia, during the same period, 853,183 accidents occurred in settlements, taking into account the fact that the total number was 1,137,987 (fig. 1), outside the settlements 284,803 (fig. 2 b).
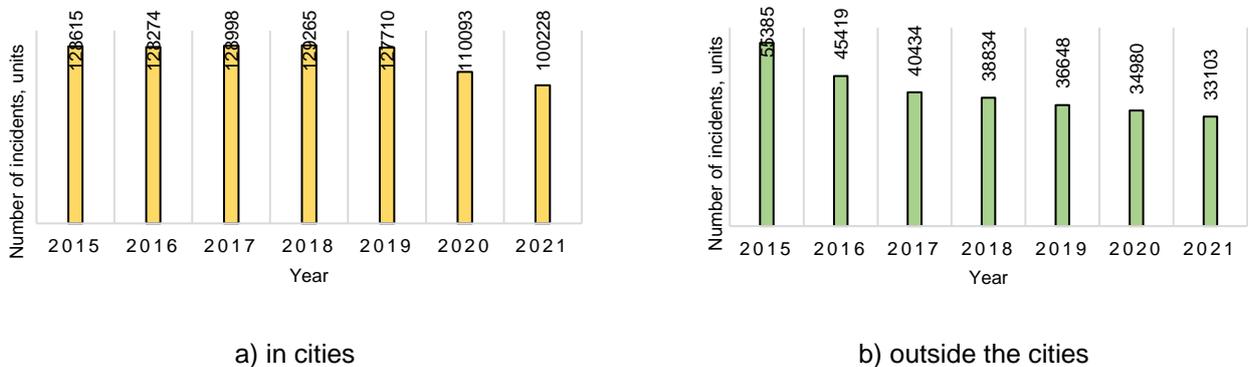


a) in cities



b) outside the cities

Fig. 2. Histogram of the number of road traffic accidents in cities and outside cities in Russia for the period 2015-2021

The analyzed database, in addition to the considered indicators, allows us to consider a large number of factors influencing the occurrence of an accident, for example, the day of the week, time, vehicle serviceability, etc. As part of this study, an analysis was made of the occurrence of incidents depending on the day of the week for the period 2015-2021. (Fig. 3).
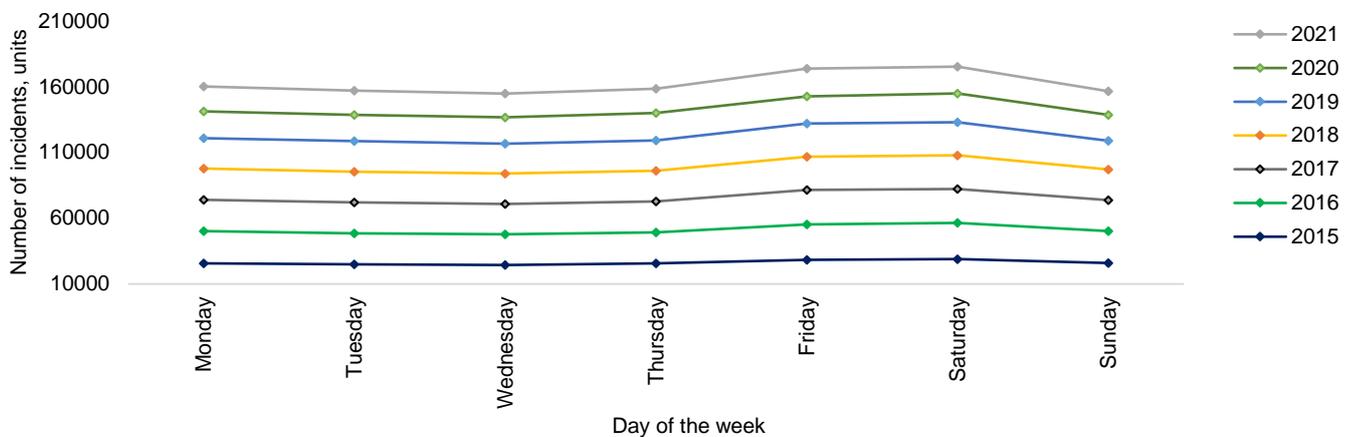


Fig. 3. Linear graphs of the distribution of the number of incidents depending on the day of the week in Russia for the period 2015-2021

The presented indicators make it possible to assess the change in the situation according to the data in the statistical database, but do not allow, for example, to assess the level of mortality in certain areas or to assess the causes and conditions affecting the occurrence of incidents. In this case, a methodology is required that will allow, based on the available data set, to assess the situation.

*Journal of Applied Engineering Science*

*Vol. 21, No. 3, 2023*

*www.engineeringscience.rs*

iipp
publishing

*Anastasia Shevtsova et al. - Probabilistic model for assessing accident rates*

## 2.2 Development of a probabilistic model for assessing accident rates

In order to be able to assess the conditions for the occurrence of accidents, the authors propose to develop a probabilistic model, which at the initial stage will allow the evaluation of the available primary data. It should be noted that the primary data - accident statistics data allow us to answer a number of questions regarding the effectiveness of measures taken to reduce these indicators and improve road safety. The performed analysis of accident rates allowed to establish two main types of conditions of occurrence of accidents - in cities and outside of it, which allows at this stage to classify such indicator as conditions of occurrence. For further assessment of accident rates, it is proposed to take as a basis a probabilistic model, which allows to translate the available data into relative indicators - probabilities. Despite the available variety of mathematical models, most of them are based on absolute values, but the performed analysis of the accident rate shows that the values change in a fairly wide range, which does not allow to apply the available methods and models. In this case, in order to assess the change in the distribution of the considered indicators in relative values, various probabilistic models can be used; in this study, a mathematical model based on the determination of the probability of occurrence of events in the statistical sense is developed. In this case, let us take P(A) as the probability of occurrence of accidents in certain conditions - urban and out-of-town combined equal to 1, then let $P(A_1)$ be the probability of occurrence of accidents in the city and $P(A_2)$ be the probability of occurrence of accidents outside the city, then given the available data set, the probabilities presented will have the values presented in Table 1.

Table 1. Statistical probability of occurrence of incidents in the city and outside it in Russia for the period 2015-2021

| Year / Probability | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|
| P(A) | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $P(A_1)$ | 0,70 | 0,74 | 0,76 | 0,77 | 0,78 | 0,76 | 0,75 |
| $P(A_2)$ | 0,30 | 0,26 | 0,24 | 0,23 | 0,22 | 0,24 | 0,25 |

According to the developed probabilistic model of accident rate estimation it is accepted that P(B) - probability of occurrence of accidents with the condition of days of week, then the probability of occurrence of accidents: on Monday - P(B1), Tuesday - P(B2), Wednesday - P(B3), Thursday - P(B4), Friday - P(B5), Saturday - P(B6), Sunday - P(B7), the performed calculations on determination of this type of probability are presented in Table 2.

Table 2. Statistical probability of occurrence of incidents by day of the week in Russia for the period 2015-2021

| Year / Probability | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|
| P(B) | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $P(B_1)$ | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 |
| $P(B_2)$ | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 |
| $P(B_3)$ | 0,13 | 0,13 | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 |
| $P(B_4)$ | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 | 0,14 |
| $P(B_5)$ | 0,15 | 0,16 | 0,15 | 0,15 | 0,15 | 0,14 | 0,16 |
| $P(B_6)$ | 0,16 | 0,16 | 0,15 | 0,15 | 0,15 | 0,15 | 0,15 |
| $P(B_7)$ | 0,14 | 0,14 | 0,14 | 0,14 | 0,13 | 0,13 | 0,14 |

The obtained probability values make it possible to calculate the probability of occurrence of incidents in the city on a certain day of the week P(Cn) and outside the city P(Dn), using the formulas:

$$P(C_n) = P(A_1) \cdot P(B_n), \qquad (1)$$

$$P(D_n) = P(A_2) \cdot P(B_n), \qquad (2)$$

where P(Cn) is the probability of occurrence of incidents in the city on a certain day of the week; n – day of the week, n = 1…7 (1 – Monday, 2 – Tuesday, 3 – Wednesday, 4 – Thursday, 5 – Friday, 6 – Saturday, 7 – Sunday); P(A1) - the probability of occurrence of incidents in the city; P(Bn) probability of occurrence of incidents on a certain day of the week; P(Dn) - the probability of occurrence of incidents outside the city on a certain day of the week.

Using formulas (1) and (2), the values of the corresponding probabilities were determined (Table 3, Table 4).

The performed calculations allow us to conclude that during the considered long-term period (7 years), the distribution of the probability of occurrence of incidents remains at the same level, so the probability of occurrence of incidents in the city is higher than 0.7, which indicates that more than 70% accidents happen in cities.

The "day of the week" indicator (Table 2) also allows us to conclude that during the week the probability of occurrence of incidents is approximately the same, moreover, the probability of occurrence of incidents from Monday to Thursday and on Sunday is the same and amounts to 14%, but on Friday and Saturday is slightly higher and is 15%.

Table 3. Statistical probability of occurrence of incidents in the city on a certain day of the week in Russia for the period 2015-2021

| Year<br>Probability | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|
| $P(C_1)$ | 0,10 | 0,10 | 0,11 | 0,11 | 0,11 | 0,11 | 0,11 |
| $P(C_2)$ | 0,09 | 0,10 | 0,11 | 0,11 | 0,11 | 0,10 | 0,10 |
| $P(C_3)$ | 0,09 | 0,09 | 0,10 | 0,11 | 0,11 | 0,11 | 0,10 |
| $P(C_4)$ | 0,10 | 0,09 | 0,11 | 0,11 | 0,11 | 0,11 | 0,10 |
| $P(C_5)$ | 0,11 | 0,11 | 0,12 | 0,12 | 0,12 | 0,11 | 0,12 |
| $P(C_6)$ | 0,11 | 0,11 | 0,12 | 0,12 | 0,12 | 0,11 | 0,11 |
| $P(C_7)$ | 0,10 | 0,10 | 0,11 | 0,11 | 0,10 | 0,10 | 0,10 |

Table 4. Statistical probability of occurrence of incidents outside the city on a certain day of the week in Russia for the period 2015-2021

| Year<br>Probability | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|
| $P(D_1)$ | 0,04 | 0,04 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 |
| $P(D_2)$ | 0,04 | 0,04 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 |
| $P(D_3)$ | 0,04 | 0,04 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 |
| $P(D_4)$ | 0,04 | 0,04 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 |
| $P(D_5)$ | 0,05 | 0,04 | 0,04 | 0,03 | 0,03 | 0,03 | 0,04 |
| $P(D_6)$ | 0,05 | 0,04 | 0,04 | 0,04 | 0,03 | 0,04 | 0,04 |
| $P(C_7)$ | 0,04 | 0,04 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 |

The obtained results of probabilities make it possible to develop a diagram of a probability relationship graph, according to which it is possible to estimate the probability of occurrence of incidents in various conditions on a certain day of the week (Table 3, Table 4, Fig. 4).
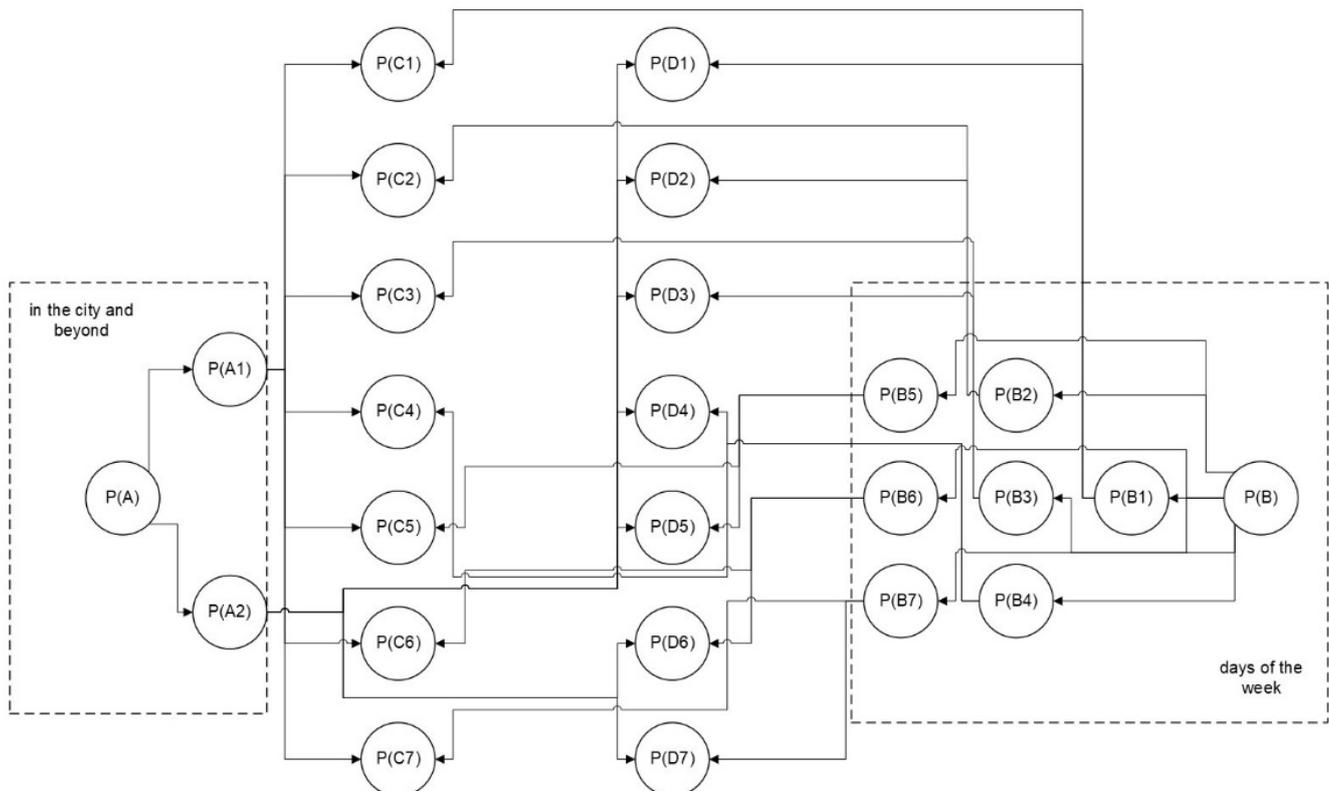


Fig. 4. Graph scheme for calculating the probability of occurrence of incidents in the city and outside it on a certain day of the week

The set of presented dependencies and their graphical representation is a probabilistic model that can be applied to determine the number of incidents using the algorithm (Fig. 5).

For a specialist, the established pattern, which has a linear character, is necessary, because allows you to significantly reduce the processing time of the results.

**Journal of Applied Engineering Science**

*Vol. 21, No. 3, 2023*

*www.engineeringscience.rs*

iipp
publishing

*Anastasia Shevtsova et al. - Probabilistic model for assessing accident rates*
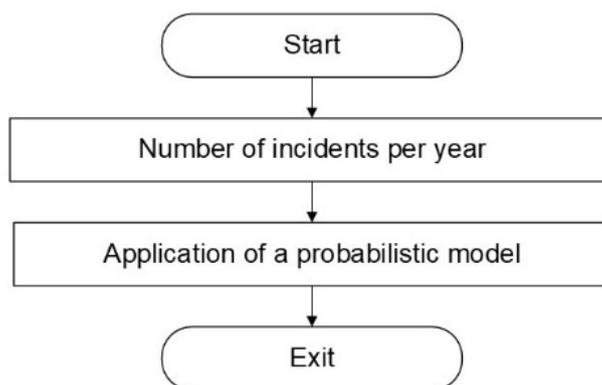
Fig. 5. Scheme of the algorithm for processing accident rates

The main requirement when applying the obtained probabilistic model is accuracy, for this purpose, as part of the study, using the available data for 2022 as an example, the accuracy of the results obtained was verified.

## 3 EXPERIMENTATION AND ASSESSMENT OF THE ACCURACY OF THE RESULTS

In the database used, indicators are given on the number of incidents on a certain day of the week for 2022, in order to evaluate the effectiveness of the probabilistic model, the accuracy of the developed model was assessed as part of the study. Table 5 presents the results of this procedure. Knowing the total number of accidents per year, which was 126,705, the N2 value was determined using:

$$N_2 = N \cdot \overline{P(B_n)}, \qquad (2)$$

$$\overline{P(B_n)} = {\sum_1^n P(B_n)}/{n}, \qquad (2)$$

where $N_2$ - is the calculated number of incidents using a probabilistic model (probability dependencies), units. ; N - is the total number of incidents obtained from statistical databases, units; $\overline{P(B_n)}$ - is the value of the average probability obtained from the result of calculations for the period under review, where n is the number of probabilities (n=7).

Table 5. The results of assessing the accuracy of the developed probabilistic model

| Index / Day of the week | Number of accidents in 2022 ($N_1$) | Probability $\overline{P(B)}$ | Received number of incidents for 2022 ($N_2$) | Absolute error | Relative error | rounded, % |
|---|---|---|---|---|---|---|
| Monday | 17573 | 0,141 | 17828 | 255 | 0,014 | 1,43% |
| Tuesday | 17462 | 0,138 | 17496 | 34 | 0,001 | 0,19% |
| Wednesday | 17662 | 0,136 | 17244 | 418 | 0,023 | 2,42% |
| Thursday | 17383 | 0,139 | 17662 | 279 | 0,01 | 1,57% |
| Friday | 19808 | 0,153 | 19341 | 467 | 0,023 | 2,41% |
| Saturday | 19595 | 0,154 | 19496 | 99 | 0,005 | 0,5% |
| Sunday | 17222 | 0,137 | 17417 | 195 | 0,011 | 1,11% |

The results obtained allow us to conclude that the developed probabilistic model gives sufficiently accurate indicators, thus, knowing the total number of incidents, it is possible to apply the obtained dependencies (1) and (2), converting the available indicators into relative values, to calculate the necessary data. As an example, within the framework of this study, it was determined that there is a certain relationship between the considered indicators in relative terms - the probability that can be used as a developed probabilistic model when processing accident rates, both in the region and the country as a whole. The maximum error obtained as a result of the calculations was 2.42%, the minimum 0.19%, which confirms the effectiveness of the developed probabilistic model.

The relative error is minimal and does not exceed 0.02, which also confirms the high degree of accuracy of the developed probabilistic model.

## 4 DISCUSSION

The issue of road safety for many years continues to be relevant, to reduce the number of deaths and injuries in accidents, it is necessary to apply a large number of measures, but in order to effectively apply them at the initial stage it is necessary to assess the causes and conditions of their occurrence. A large set of primary data is presented in official statistical sources - databases, but when working with them, there is no possibility of setting the logic of the research process, here it is necessary to develop a certain model with different variants of possible search queries,

the basis of such a model is a mathematical description that allows to ensure a certain accuracy when analyzing the formulated query. In the presented study, a probabilistic model of accidents occurrence on certain days of the week in urban and suburban environment is developed, which, taking into account changes in such indicators as the level of motorization and population size, allows to forecast changes in the considered indicators of accident rate. For specialists in the subject area under consideration, the results obtained and their application allows to effectively use a set of measures aimed at reducing accidents, for example, tightening control on certain road sections located in the city or outside it on certain, the most accident-prone days.

## 5 CONCLUSIONS

As a result of the research, the team of authors obtained the following results:

1. The analysis of absolute accident rates made it possible to determine that for the period under review - 2015-2021, there is a systematic decrease in accident rates;

2. The analysis of relative accident rates, with the calculation of the corresponding indicators of the probability of occurrence of events by statistical frequency, made it possible to establish a certain pattern in the distribution of the values under consideration, it was found that more than 70% of accidents occur in cities over a large number of years, the probability of occurrence of accidents on a certain day of the week is 14 %.

3. Certain probabilities made it possible to calculate the probabilities on a certain day of the week for cities - 3% and in the city 11%, and to develop a probabilistic model, the results obtained for the possibility of application were presented in the form of an algorithm that reflects the sequence of actions when working with statistical indicators.

4. Evaluation of the accuracy of the developed model and the application of the algorithm showed that its application is possible, due to the fact that the error is minimal and the relative value does not exceed 0.023, which is rounded to 2.42%.

## 6 ACKNOWLEDGMENT

## 7 REFERENCES

[1] Statistics of the State Road Safety Inspectorate (2023). Statistical data of the state traffic police. Available at: http://stat.gibdd.ru/ (date of information retrieval: January 5, 2023).

[2] Badie, M., Shaaban, F. (2021). Diagnosing the statistical rates of traffic accidents in Dohuk Governorate during the period (2011-2020). *Qubahan Academic Journal*, 1(2), 47–54. https://doi.org/10.48161/qaj.v1n2a42.

[3] Petrashenko, O., Yefymenko, R. (2021) Analysis of seasonal changes of accident level on the roads of state importance and their impact on road accidents statistics. *Roads and bridges*, vol.. 24,169–176.

[4] Kurakina, E., Evtiukov, S. (2020) Systemic indicators of road infrastructure at accident clusters. *Architecture and Engineering*, vol. 5 (1), 51-58.

[5] Savolainen, P., Mannering, F. (2011) The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accid. Anal. Prev*, vol. 43, 1666–1676.

[6] Kumar, S., Toshniwal, D. (2015) A data mining framework to analyze road accident data. *J. Big Data*, vol. 2, 2–26.

[7] Bin Islam, M., Kanitpong, K. (2008) Identification of factors in road accidents through in-depth accident analysis. *IATSS Res*, vol. 32, 58–67.

[8] Chubukov, A.; Kapitanov, V. (2017) Simulation of regional mortality rate in road accidents. *Trans. Res. Proc*, vol. 20, 112–124.

[9] Kapitanov, V., Monina, O. (2020) Probabilistic assessment of main factors determining the road traffic accident rate in regions of Russia. *Transportation Research Procedia*, vol. 50, 218-225.

[10] Evtiukov, S., Karelina, M. (2018). A method for multi-criteria evaluation of the complex safety characteristic of a road vehicle. *Transportation Research Procedia* vol. 36, 149–156. DOI: 10.1016/j.trpro.2018.12.057.

[11] Shevtsova, A.G. (2021) Mathematical analysis of certain road safety indicators in the Russian Federation. *The Russian Automobile and Highway Industry Journal*. Vol. 18(6), 700-711 https://doi.org/10.26518/2071-7296-2021-18-6-700-711

[12] Belin, M., Hansson, S. (2022). Vision Zero in Sweden: Streaming Through Problems, Politics, and Policies. *In: Edvardsson Björnberg. The Vision Zero Handbook*. Springer, Cham. https://doi.org/10.1007/978-3-030-23176-7_9-1

*Anastasia Shevtsova et al. - Probabilistic model for assessing accident rates*

[13] Belin, M., Per, T. (2012). Vision zero – A road safety policy innovation. *International Journal of Injury Control and Safety Promotion*, vol. 19(2), 171–179.;

[14] Andersson, R., Gell, T. (2022). Vision Zero on Fire Safety. In: Edvardsson Björnberg. *The Vision Zero Handbook*. Springer, Cham. https://doi.org/10.1007/978-3-030-23176-7_44-1

[15] Jamroz, K., Romanowska, A. (2022). Vision Zero in Poland. *The Vision Zero Handbook*. Springer, Cham. https://doi.org/10.1007/978-3-030-23176-7_14-1

[16] Nguyen, M., Novikov A. (2022) Vision Zero in the mining industry – a review of the application in Russia and Vietnam. *Economic relations*, vol. 12 (4), 769-782 doi: 10.18334/eo.12.4.116382.

[17] Shevtsova, A. (2021) Validity of Smeed's law in the context of the implementation of the Vision Zero program. *The world of transport and technological machines*. Vol. 4(75), 49-57, DOI 10.33979/2073-7432-2021-75-4-49-57.

[18] Shevtsova, A., Novikov, A. (2021) Development of an approach to determination of coupling qualities of road covering using weather-climate factor. *Journal of Applied Engineering Science vol.* 19 (1), 30-36 DOI 10.5937/jaes0-26642.;

[19] Novikov, A., Vasilieva, V. (2020) Development of approach to reduce number of accidents caused by drivers. *Transportation Research Procedia*, vol. 50, 491-498 DOI 10.1016/j.trpro.2020.10.090;

[20] Petrov, A., Evtiukov, S. (2019) Statistical Modelling of Orderliness of Regional Road Safety Provision Systems. *IOP Conference Series: Earth and Environmental Science:Current Problems and Solutions*, vol. 224 (1), 012033 DOI 10.1088/1755-1315/224/1/012033.

[21] Novikov, A., Zyryanov, V. (2018) Dynamic traffic re-routing as a method of reducing the congestion level of road network elements. *Journal of Applied Engineering Science* vol. 16 (1), 70-74 DOI 10.5937/jaes16-15289.