

**Keywords:** road safety; logistic regression; mortality

**Anna BORUCKA, Malgorzata GRZELAK\***

Military University of Technology, Faculty of Security, Logistics and Management  
Kaliskiego 2, 00-908 Warsaw, Poland

**Andrzej ŚWIDERSKI**

Motor Transport Institute  
Jagiellońska 80, 03-301 Warsaw

\*Corresponding author. E-mail: [malgorzata.grzelak@wat.edu.pl](mailto:malgorzata.grzelak@wat.edu.pl)

## **APPLICATION OF THE LOGISTIC REGRESSION MODEL TO ASSESS THE RISK OF DEATH IN ROAD TRAFFIC ACCIDENTS IN THE MAZOWIECKIE VOIVODESHIP**

**Summary.** Mortality caused by road accidents is a significant problem for most countries, including Poland, where approximately 2,900 people die each year, and another 37,359 are injured. Research in this area has been conducted on a large scale. One of the most important elements is the evaluation of factors leading to fatalities in road accidents, which is also the goal of this article. The analysis was based on data on road accidents from the Mazowieckie Voivodeship, which is characterized by one of the highest mortality rates gathered for the period 2016-2018. Owing to the dichotomous form of the studied variable, logistic regression was used. Estimated model parameters and calculated odds ratios allowed to assess the effect of selected factors on road traffic mortality rate. As significant, the type of the perpetrator and the traffic participant, sex and age of the victim, road lighting, and the driver's experience were selected. It was assessed that pedestrians are the group most exposed to death in a road accident, both as perpetrators and victims. It was also pointed out that the risk of death for women is 1.8 times higher than men. In addition, the effect of driving experience is also important, and the risk of death is 0.64 times lower for drivers with longer practice. It was also assessed that with each subsequent year of life, the risk of death in a road accident increased by 2%. Furthermore, according to incident site lighting, the study demonstrated that the risk of death was greatest when driving at night on an unlit road. The results obtained may support public safety and law enforcement authorities in carrying out preventive actions and also can be helpful in shaping the overall strategy on road safety.

### **1. INTRODUCTION**

Road safety is an important and topical issue. Statistics show that more than 1.3 million people die worldwide every year owing to traffic incidents. It is estimated that traffic accidents are in eighth place among the main causes of human mortality, and in the age group of 5 to 26 years, in the first place.

Therefore, it is important to take measures to minimize and counteract their negative effects. This is facilitated by the identification and assessment of factors that affect the risk of being involved in a traffic incident. The use of mathematical models allows the estimation of the strength and direction of this effect, and as a result can be a support in making decisions on improving the level of road safety.

This article features an analysis of road accidents in the Mazowieckie Voivodeship, which has the highest mortality rate among all voivodeships (regions) in Poland. Figure 1 shows the number of fatalities in individual provinces in Poland in 2016-2018. Every year, an average of 473 people die in

this area, which is 16% of all nationwide fatalities, and means that every sixth victim loses their life in the Mazowieckie Voivodeship.

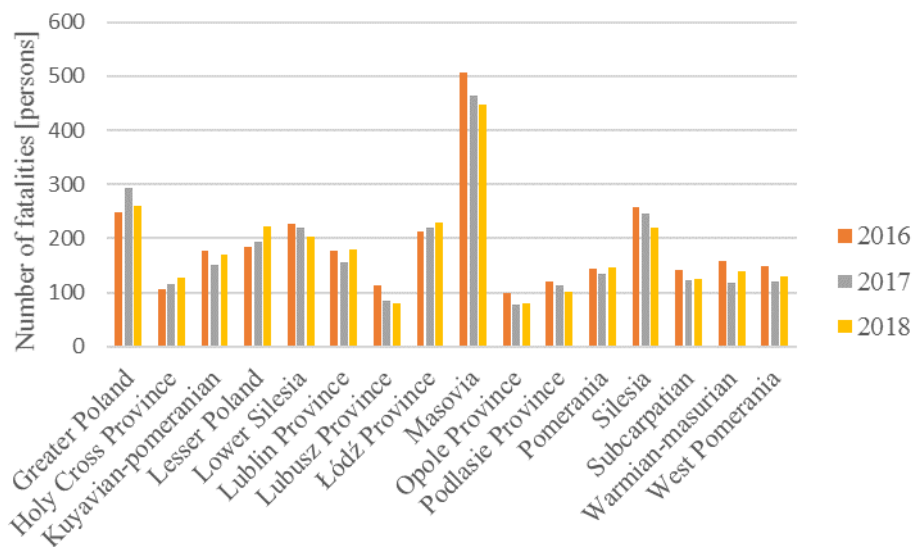


Fig. 1. Graph of number of fatalities in Poland by voivodeships

The article applied logistic regression. The collected observations and preliminary analyses allowed to isolate the main causes conducive to deaths in road accidents, and on this basis, model predictors were selected. These included qualitative variables characterizing accident participants and quantitative variables describing atmospheric conditions prevailing at that time. The estimated parameters of the regression model made it possible to examine relationships. The quality of the proposed model was verified based on an analysis of significance of parameters and an evaluation of goodness-of-fit using the ROC curve (Receiver Operating Characteristic Curve) and the AUC (Area Under the ROC Curve).

## 2. LITERATURE REVIEW – PROBLEM STATEMENT

Road safety has been subject to many scientific studies and publications. Most often, quantitative analyses are carried out, aimed at estimating the parameters of mathematical models describing the phenomenon under consideration and predicting its future development.

The frequency of occurrence of road incidents and their numbers are investigated [2, 9]. There are also assessments of the number of traffic accidents causing fatalities [8, 11] or causing injuries [13, 34]. Analyses regarding the number of fatalities as well as injuries are less popular [5, 17].

Widely used in this area are generalized linear regression models with adopted probability distributions, including Poisson distribution, gamma-Poisson distribution – negative binomial distribution (NB) [1, 10], binomial distribution 21, Poisson lognormal (PL) [26], and distributions with an excess of zero counts, i.e., zero-inflated Poisson (ZIP) and zero-inflated negative Binomial (ZINB) 18. In addition, solutions based on complex econometric models are used, among others on the Bayes method, a Monte Carlo simulation method based on Markov chains 14, as well as one using time series analysis methods 11. The main purpose of qualitative analyses is to present a descriptive characteristic of the research objects, assess the severity of road accidents broken down into individual categories of victims [24], as well as analyze the concentration of incidents and spatial differentiation of the level of road safety [25]. Among the studies available in the literature, the most frequently mentioned as explained variables are as follows:

- parameters related to vehicle traffic: average daily traffic volume 12, transport performance [6], and the participation of heavy goods vehicles 15;

- characteristics of the driver: sex 1, age, nationality, and driving license 5, as well as driving under the influence of alcohol 29;
- vehicle characteristics 19;
- incident location [7];
- road surroundings 24;
- type of infrastructure: urban [7, 32] and non-urban road sections [2, 3];
- road network elements: tunnels, motorway junctions, road junctions, intersections 16, or long road sections [20];
- road surface [7];
- lighting and visibility 24;
- atmospheric conditions [17, 34]; and
- time of occurrence of the event 24.

A review of the literature also revealed the use of the logistic regression function 5. The available studies include research on the severity of road accidents depending on the location of the accident 28 or the causes of the incident (excessive speed, inadequate distance between vehicles, red light violations 5 etc.). Other authors consider the age of the perpetrator and other participants of the incident to be significant, as well as the time when the traffic accident occurred 31. This is demonstrated, for example, by Shakaya and Marsani by analyzing the death rate of accidents in the Kathmandu Valley. Based on an analysis of events in Shijiazhuang, Ma et al. indicated the significant effect of variables related to the cross-section of the road, place of the accident, type of the road, density of intersections, and lighting conditions 28. However, Ahmed as the main reason indicated the speed and type of the vehicle and the location of the incident 4.

The analysis of the literature also showed that research conducted for large, demographically diverse areas does not always fully reflect the characteristics of occurring phenomena 24. Therefore, this article assesses one territorially distinguished region in which the number of accidents was the highest and which is characterized by significant population migration resulting from professional and educational factors. It was found that the risk of death as a result of traffic accidents is significantly influenced by factors related to the type of the perpetrator and the participant of the incident, urbanization of the place where it occurred, the manner the road is lit, as well as the sex of the victim and the driver's experience.

### 3. LOGISTIC REGRESSION

The analysis and assessment of the mortality rate of road accidents in the Mazowieckie Voivodeship was carried out using the logistic regression model. It allows mathematical recording of the effect of several variables  $X_1, X_2, \dots, X_n$  on the dichotomous variable  $Y$ , which takes the value 1 in the event of death as a result of a traffic incident, and 0, when participants are injured or do not take any injuries. The link function of the model is the logistic function, which takes the following form (1) 23:

$$f(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}} \quad , \quad (1)$$

where:  $e$  – Euler's number, and  $x$  – value of explanatory variable  $X$ .

The logistic regression model is described by the following equation:

$$P(Y = 1|x_1, x_2, \dots, x_k) = \frac{e^{\beta_0 + \sum_{i=1}^k \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^k \beta_i x_i}} \quad , \quad (2)$$

where:  $\beta_i \ i=0, \dots, k$  – regression coefficients, and  $x_1, x_2, \dots, x_k$  – independent variables that can be measurable or qualitative.

The model is based on the method of expressing probability through odds, which is the ratio of the probability of occurrence of an incident and the probability that it will not happen. For a given incident A, the aforementioned definition takes the following form:

$$Odds(A) = \frac{P(A)}{P(nieA)} = \frac{P(A)}{1-P(A)} \quad . \quad (3)$$

The most commonly used measure of association is the odds ratio OR, which, for two compared groups A and B, is defined as the odds of occurrence of A to the odds of occurrence of B. The odds ratio is recorded as follows:

$$OR_{AxB} = \frac{S(A)}{S(B)} = \frac{P(A)}{1-P(A)} : \frac{P(B)}{1-P(B)}. \quad (4)$$

Regression coefficients  $\beta_i$ , where  $i=0, \dots, k$  are estimated based on the maximum likelihood estimation method (MLE), which maximizes the likelihood function or its square. The likelihood function takes the following form 33:

$$L(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n p(x_i; \theta_1, \theta_2, \dots, \theta_k) \text{ for discrete variables,} \quad (5)$$

$$L(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_k) \text{ for continuous variables,} \quad (6)$$

where:  $\theta_1, \theta_2, \dots, \theta_k$  – unknown distribution parameters;  $x_1, x_2, \dots, x_n$  – the value of the variable X observed in the random sample;  $p(x)$  – distribution determined by the probability function;  $f(x)$  – distribution determined by the density function, and  $L$  – likelihood function.

Assuming that all observations are independent of each other, the probability of observing the entire data set (with the model parameters set) is equal to the product of the probabilities (likelihood functions) of individual samples. The estimates of parameters are those values for which the likelihood is the highest (the higher, the better the fit of the model to the data).

The evaluation of the model and the quality of the prediction is carried out by following:

- applying statistics used in the evaluation of diagnostic tests, i.e. accuracy (ACC), sensitivity (SE), and specificity (SP);
- performing the Hosmer-Lemeshow test; and
- plotting the ROC curve and analyzing the AUC.

The application of statistics used in the evaluation of diagnostic tests requires the determination of the value called the cut-off point. This is the value of the examined variable that optimally divides the group into two parts, the first in which the phenomenon occurs and the second in which it does not. The cut-off point is therefore the determined value of  $\pi_0$  from the range  $[0,1]$ , for which the following is assumed:

$$\hat{P}(Y = 1|x) > \pi_0 \quad . \quad (7)$$

To evaluate the quality of classification, two measures are most often used simultaneously: sensitivity and specificity (defined as difference  $1 - SP$ ). To evaluate, the obtained pairs of values  $1 - SP$  and SE are marked on a plane, where the horizontal axis represents the values of  $1 - SP$ , whereas the vertical axis the values of SE. The combination of the obtained points creates the so-called ROC curve that allows an overall assessment of the predictive quality of the model. It shows all possible cut-off points and statistics related to them. In addition, it does not depend on the adopted scale and allows easy reading of sensitivity and specificity values.

If there is a continuous independent variable in the logistic regression model, many different cutoff points can be tested to distinguish occurrences from non-occurrences, depending on the predicted probability. It is important to find a point that has high values of both sensitivity and specificity. This is the optimal cutoff point, which is determined using the Youden index (J). It takes the following form:

$$J = \text{sensitivity} + \text{specificity} - 1 \quad . \quad (8)$$

The optimal cutoff point corresponds to the case where the J value reaches its maximum.

#### 4. MODEL OF LOGISTIC REGRESSION USED FOR TESTING ROAD TRAFFIC MORTALITY RATE IN THE MAZOWIECKIE VOIVODESHIP

This article analyzes road accidents recorded in 2016-2018 in the Mazowieckie Voivodeship. The focus was on the victims of the aforementioned incidents as well as the categories of injuries sustained; the group included persons who died, were seriously and slightly injured or did not sustain injuries. The dependent variable was the mortality index with a binomial distribution, for which  $I$  indicates the loss of life as a result of a road accident. Model building began by selecting factors that

may affect the explained variable. Factors were selected that determine the atmospheric conditions at the time of the incident, i.e. air temperature, wind speed, and total precipitation, as well as parameters characterizing the participants and the place of the incident, i.e., the type of the perpetrator and the traffic participant, sex and age of the victim, the effect of narcotics, area of the incident, road lighting, or the driver’s experience. In the first step, the effect of the aforementioned variables on the dependent variable was evaluated. In the group of qualitative variables, the analysis was carried out based on contingency tables and by gauging association strength using the independence test  $\chi^2$ . In the first step, in the group of quantitative variables, basic descriptive statistics were calculated. The obtained values of skewness and kurtosis indicate the lack of compliance of the distribution of variables with the normal distribution (absolute values, respectively, greater than 1 for skewness and from 2 for kurtosis) (Table 1). The aforementioned results are confirmed by the results of the Kolmogorov-Smirnov test (Fig. 2, Tab. 2). Therefore, the assessment of the effect of quantitative variables on the fatality of road accidents was based on the homogeneity of variance analysis based on the non-parametric Kruskal-Wallis test 22.

Table 1

Descriptive statistics for the independent variables

Variable	Mean	Median	Variance	Standard deviation	Skewness	Kurtosis
air temperature	11.4	13.0	68.1	8.3	-1.3	-2.7
wind speed	2.4	2.3	1.0	1.0	1.7	2.7
total precipitation	1.7	0.0	19.0	4.4	4.5	26.2
participant's age	39.7	38.0	316.8	17.8	1.5	-3.3

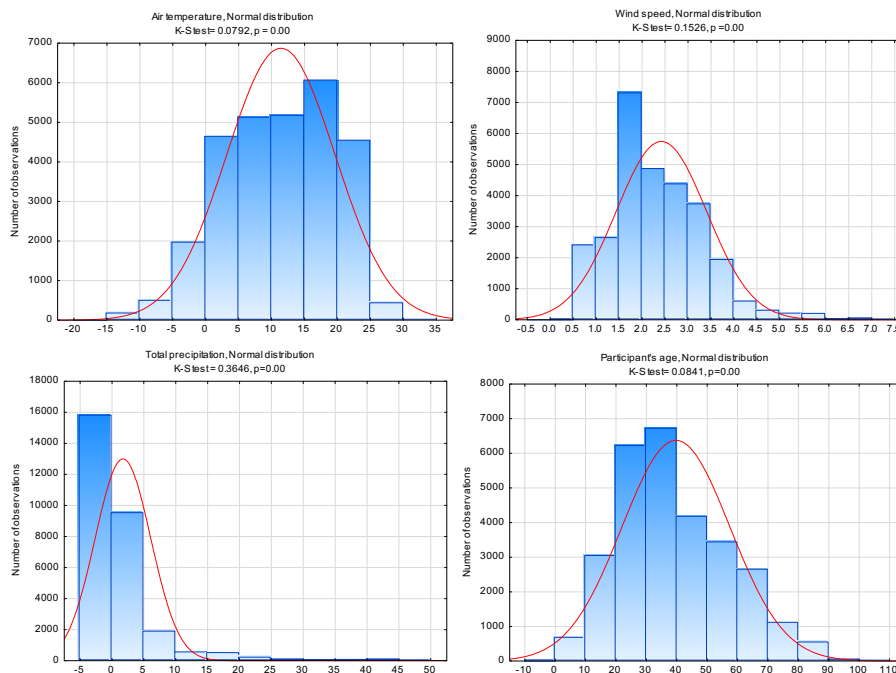


Fig. 2. Histograms of the observed distribution of variables

The performed Kruskal-Wallis test, for which the null hypothesis implies that the median from all samples do not differ from each other, showed that only the effect of *the participant's age* variables on the examined explained variable is statistically significant. The results of the conducted test are shown in Table 3.

In the next step, the significance of the effect of the selected qualitative variables on traffic accidents mortality was evaluated. For this purpose, contingency tables and the independence test  $\chi^2$  were used. The results are presented in Table 4.

Table 2

Kolmogorov-Smirnov test results for normality of distribution

Variable	Value of test statistic	p-value
air temperature	0.0792	0.00
wind speed	0.1526	0.00
total precipitation	0.3646	0.00
participant's age	0.0841	0.00

Table 3

Kruskal-Wallis test results

Variable	Value of test statistic	df	p-value
air temperature	27.84	42	0.95
wind speed	2.27	4	0.69
total precipitation	0.91	3	0.82
participant's age	358.02	91	0.00

Table 4

Independence (Pearson) test results  $\chi^2$ 

Variable	Value of test statistic	df	p-value
type of perpetrator	842.09	1	0.00
type of incident participant	533.76	2	0.00
urbanisation of the area	399.91	1	0.00
incident site lighting	504.85	3	0.00
participant's gender	33.09	1	0.00
driver's experience	299.84	4	0.00
effect of narcotics	0.003	1	0.96

The analysis showed that only the variable *effect of narcotics* does not significantly affect the dependent variable studied. Confirmation of the aforementioned are the results of the independence test  $\chi^2$ , according to which, for all variables (excluding *effect of narcotics*) the null hypothesis should be rejected and an alternative hypothesis should be adopted claiming that road traffic mortality depends on the qualitative variable being studied.

The use of logistic regression model requires meeting the assumption of linearity of independent variables. Quantitative predictors were evaluated. The significance of effect was found only for *the participant's age* variable. A graphical evaluation of the scatter plot of this variable and the corresponding value of the logarithm of the odds was carried out (Fig. 3) 33. In addition, a linearity test was performed. The calculated  $p$  value indicates that there are no grounds for rejecting the null hypothesis on the linear effect of the studied variable on the analyzed phenomenon.

In the next stage, the construction of the multivariate model was started. Variables whose effect on mortality owing to traffic accidents was significant were used. For qualitative variables, parametrization was carried out with an adopted reference level (Table 4). When building the model, cross-evaluation was selected as a way of its validation. Estimated values of logistic regression parameters are presented in Table 5.

All estimated model parameters proved to be statistically significant. For all variables, the calculated Wald test value and the corresponding  $p$  value are smaller than the assumed significance level  $\alpha = 0.05$ . On this basis, it can be concluded that the listed factors have a significant effect on the

risk of death as a result of a traffic incident. The equation of the logistic regression model is as follows (9):

$$P(\text{death} = 1|X) = \frac{e^a}{1 + e^a} \quad (9)$$

where:  $a = -2.01 + 0.02 \cdot \text{participant's age} - 1.17 \cdot \text{type of perpetrator.dr} + 0.55 \cdot \text{type of participant.p} + 0.08 \cdot \text{type of participant.dr} - 1.14 \cdot \text{built-up area} - 0.42 \cdot \text{lighting day light} + 0.31 \cdot \text{lighting night unlit road} + 0.59 \cdot \text{gender male} + 0.65 \cdot \text{experience no data} - 0.44 \cdot \text{experience high}$ .

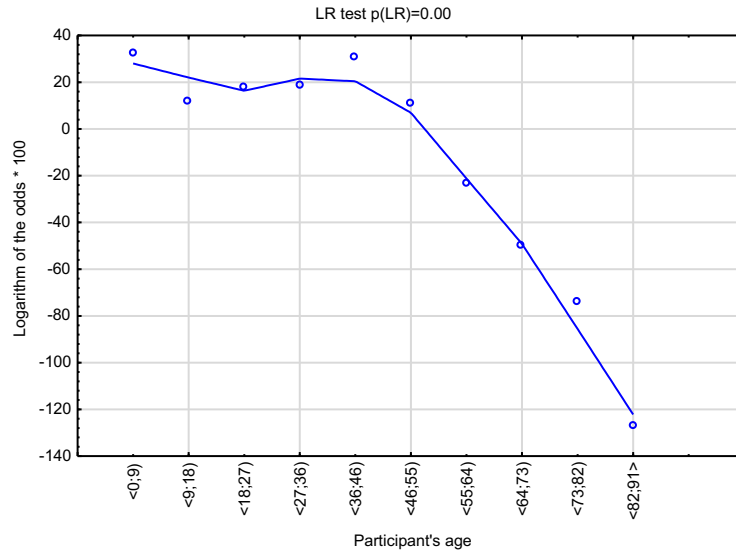


Fig. 3. A scatter plot of the logarithm of the odds as a function the predictor *Participant's age*

Table 5

Estimated parameters of the model of logistic regression

Effect	Modelled probability: road traffic mortality Distribution: binomial, link function: LOGIT				
	Effect level	Parameter	Standard error of estimation	Wald test statistics	p
intercept		-2.01	0.21	92.87	0.00
participant's age		0.02	0.00	113.05	0.00
type of perpetrator	driver (dr)	-1.17	0.12	88.23	0.00
type of participant	pedestrian (p)	0.55	0.12	20.73	0.00
type of participant	driver (dr)	0.08	0.12	0.52	0.00
area of incident	built-up	-1.14	0.08	220.99	0.00
lighting	daylight	-0.42	0.12	11.56	0.00
lighting	night - unlit road	0.31	0.14	5.03	0.02
gender	male	0.59	0.08	52.78	0.00
experience	no data	0.65	0.18	13.10	0.00
experience	high	-0.44	0.19	5.46	0.02

The aforementioned equation can take the equivalent form of odds (10):

$$P \frac{P(\text{death}=1|X)}{1-P(\text{death}=1|X)} = e^a \quad (10)$$

where:  $a = -2.01 + 0.02 \cdot \text{participant's age} - 1.17 \cdot \text{type of perpetrator.dr} + 0.55 \cdot \text{type of participant.p} + 0.08 \cdot \text{type of participant.dr} - 1.14 \cdot \text{built-up area} - 0.42 \cdot \text{lighting day light} + 0.31 \cdot \text{lighting night unlit road} + 0.59 \cdot \text{gender male} + 0.65 \cdot \text{experience no data} - 0.44 \cdot \text{experience high}$ .

The next step examined the goodness of fit of the model based on a comparison of predicted and observed values. For this purpose, the Hosmer-Lemeshow (HS) test was used. The HS test is based on a comparison of the number of occurrences of the modelled class with its actual implementation. The null hypothesis assumes that the observed and expected number of occurrences do not differ from each other. The calculated  $p$  value of the HS test is 0.64 and does not provide grounds for rejecting the null hypothesis, indicating that the model is well fitted to the data.

Then, the calculation of the odds ratios of death in a traffic accident was made, taking into account the effect of individual factors. The sign at the calculated parameter values indicates whether the analyzed factor increases or decreases the odds of the occurrence of the studied phenomenon.

Table 6

## Estimated odds ratios

Effect	Modelled probability: road traffic mortality Distribution: binomial, link function: LOGIT				
	Effect level	Odds ratio	95.00% CI	-95.00% CI	p
participant's age		1.02	1.02	1.02	0.00
type of perpetrator	driver	0.31	0.24	0.40	0.00
type of participant	pedestrian	1.74	1.37	2.21	0.00
type of participant	driver	1.09	0.86	1.37	0.00
area of incident	built-up	0.32	0.28	0.37	0.00
lighting	daylight	0.65	0.51	0.84	0.00
lighting	night - unlit road	1.37	1.04	1.80	0.02
Gender	male	1.80	1.53	2.10	0.00
Experience	no data	1.92	1.35	2.73	0.00
Experience	high	0.64	0.44	0.93	0.02

The calculated odds ratio values (Table 6) indicate that each of the model parameters significantly affects the risk of death in a road accident. In relation to the age of the incident participant, the odds ratio indicates that with each subsequent year of life the risk of death as a result of a traffic incident increases by 2%. Considering the type of the perpetrator, it is estimated that the driver's odds of dying are 0.31 times smaller in relation to the pedestrian. An analysis of the odds ratio of the type of the participant indicates that in relation to the passenger, the risk of the pedestrian's death is 1.74 times higher, whereas in the driver's case, it is 1.09 times higher. The odds of losing one's life in a traffic accident that takes place in a built-up area are 0.32 times smaller than in an undeveloped area. Considering the type of road lighting, it should be noted that in relation to driving after dark or at dawn, the odds of losing life owing to a traffic incident are 0.65 times smaller, whereas at night on an unlit road, they are 1.37 times higher. In addition, studies indicate that the risk of men dying in traffic accidents is 1.80 times higher. In turn, the experience of incident participants shows that drivers who have extensive experience are 0.64 times less likely to die compared with persons with a shorter possession of a driving license.

The developed model can be used to predict the risk of death as a result of a traffic incident; therefore, the quality of prediction was evaluated. For this purpose, the ROC curve was plotted; the most important evaluation parameter was AUC, i.e. the area under the ROC curve. It takes values from 0 to 1. The greater the area under the curve, the greater the discriminatory performance of the model; for its evaluation, Kleinbaum and Klein proposed the following classification (Table 7) 33.

As a method of model validation, cross-evaluation was used, for which the sample was split into 10 subgroups containing the training and test sets 26. The study was then conducted in accordance with



the LOO (leave one out) strategy, which assumes that all observations are used to estimate the model, with the exception of one observation, which is used to calculate the prediction error. In cross-validation, the global evaluation of prediction is an average of errors from individual sets 33. Owing to the use of cross-validation for the model, ROC curves were plotted separately for the training and test sample sets and the AUC fields were calculated (Figs. 4 and 5). In addition, the cutoff point and the matrix of classification of actual observations and those predicted by the model were determined (Table 8 and 9).

Table 7

AUC classification

AUC value	Evaluation
0.9<AUC<1.0	Excellent discrimination
0.8<AUC<0.9	Good discrimination
0.7<AUC<0.8	Fair discrimination
0.6<AUC<0.7	Poor discrimination
0.5<AUC<0.6	Failed discrimination

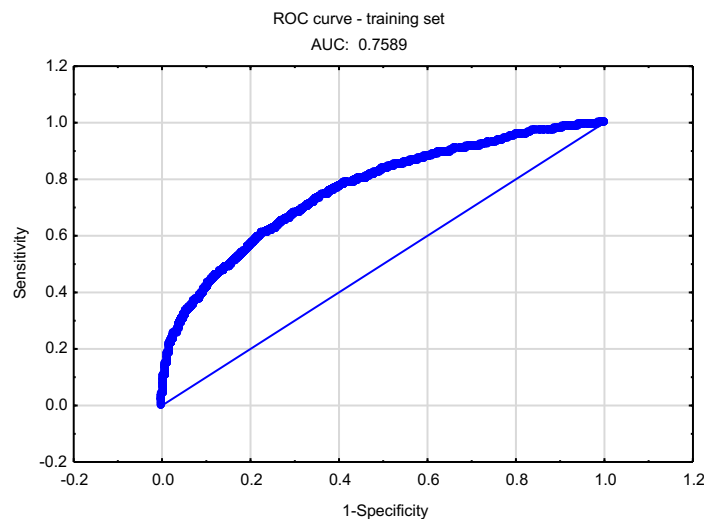


Fig. 4. ROC curve for the training set

Table 8

Matrix of classification of actual observations and the model's predictions

	Predicted: 1	Predicted: 0	Correct Percent
Observed:1	692	339	67.11
Observed:0	5570	13348	70.05

The calculated AUC values for individual sets are 0.7589 for the training set and 0.7567 for the test set, respectively, which shows their fair discrimination. This proves both sufficient adjustment to the data and sufficient quality of the model in the event of receiving new data. In addition, the calculated AUC values are similar (the difference between them does not exceed 0.05), which allows the model to be considered correct.

Table 9

Cutoff point of the evaluated model

Observation No.	Cut-off point	True-positive	True-negative	False-positive	False-negative	SE	1-SP
2241	0.049	692	13348	5570	339	0.67	0.29

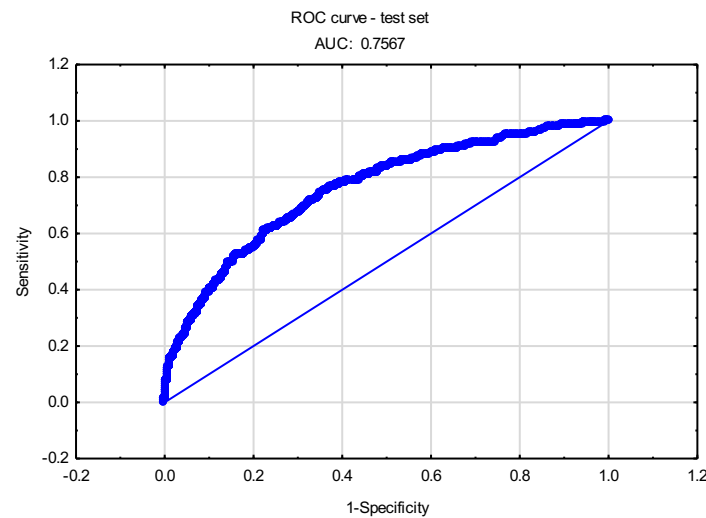


Fig. 5. ROC curve for the test set

For the proposed cutoff point, the sensitivity is at 0.67, whereas the specificity is at 0.29. A total of 14,040 cases (692 true-positive and 13,348 true-negative) have been correctly classified, whereas 5,909 cases (5,570 false-positive and 339 false-negative) have been classified incorrectly. For such a cutoff point, the value of the accuracy indicator assessing the prediction efficiency is 70.37%, allowing the model to be considered satisfactory. In addition, the percentage share of correct indications for death as a result of a traffic incident is 67.11%, which demonstrates that the model has good predictive properties.

## 5. CONCLUSIONS

The study presented in the article was conducted on the basis of data on accidents registered only in the Mazowieckie Voivodeship. The scope was adopted on the account of demographic and social diversity as well as differences related to the quality of the available road infrastructure, which could affect the quality of the model. The territorial limitation of the area enabled an accurate evaluation of the incidents occurring there.

The aim of the article was to develop a model for estimating the level of risk of death as a result of traffic incidents and evaluating the factors that affect them. Reference was made not to the number of accidents but to their consequences, which made it possible to clearly highlight the scale of the threat. Analysis concentrating solely on the frequency of such situations does not fully reflect their consequences, as many of them do not lead to fatalities, which may obscure the proper assessment of the problem. The developed model can be applied in processes related to road safety management.

Among the significant factors, among others, age was distinguished; it was assessed that with each subsequent year of life, the risk of death in a road accident increased by 2%. It was also pointed out that pedestrians are the group most exposed to death, both as perpetrators and victims. In addition, the significance of the effect of time of day on the phenomenon under study was emphasized; the study demonstrated that the risk of death was greatest when driving at night on an unlit road. Furthermore, according to the conducted study, men were the sex most exposed to the negative effects of road incidents. The risk of death for women is 1.8 times higher. The effect of driving experience is also important, and the risk of death is 0.64 times lower for drivers with longer practice.

The research results available in the literature indicate that various variables affect the mortality rate in road accidents. The results obtained in this study show a similarity to the analyses carried out in the Kathmandu region, where the effect of the victim's age was also significant (the risk of death increases by 2.3% with each subsequent year of life) and road lighting (driving at night after unlit

roads). On the contrary, studies conducted in Iraq and Saudi Arabia have shown that mortality is influenced by other factors, such as driving speed, type of car, or the scene of the accident.

The proposed model can be applied in processes related to road safety management. It indicates the factors that most significantly increase the risk of death in an accident. Thus, it can support decision-making processes of traffic participants concerning their route planning. In addition, it may support public safety and law enforcement authorities in carrying out preventive actions, including preventive police actions, aimed at reducing the number of road accidents, as well as social campaigns aimed at increasing awareness of the factors contributing to the tragic effects of road accidents.

## Acknowledgment

The article has been presented on 13th International BRD GAMBIT 2020 Conference. Co-funded by the Science Excellence programme of the Ministry of Science and Higher Education.

## References

1. Abdel-Aty, M. & Radwan, A.E. Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*. 2000. No. 32. P. 633-42.
2. Ackaah, W. & Salifu, M. Crash prediction model for two-lane rural highways in the Ashanti region of Ghana. *IATSS Research*. 2011. No. 35. P. 34-40.
3. Agüero-Valverde, J. Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. *Accident Analysis & Prevention*. 2013. No. 50. P 289-97.
4. Ahmed, L. A. Using logistic regression in determining the effective variables in traffic accidents. *Applied Mathematical Sciences*. 2017. Vol. 11. No. 42. P. 2047-2058.
5. Al-ghamdi, A.S. Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention*. 2002. No. 34. P. 729-741.
6. Anastasopoulos, P. Ch. & Mannering, F. L. A note on modeling vehicle accident frequencies with random-parameters. *Accident Analysis and Prevention*. 2009. No. 41. P. 153-159.
7. Anastasopoulos, P.C. & Mannering, F. & Shankar V.N. & Haddock J.E. A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Analysis and Prevention*. 2012. No. 45. P. 628-633.
8. Bijleveld, F.D. The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. *Accident Analysis and Prevention*. 2005. No. 37. P. 591-600.
9. Cafiso, S. & Di Graziano, A. & Di Silvestro, G. & La Cava, G. & Persaud, B. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. *Accident Analysis and Prevention*. 2010. No. 42. P. 1072-1079.
10. Caliendo, C. & Guida, M. & Parisi, A. A crash-prediction model for multilane roads, *Accident Analysis and Prevention*. 2007. No. 39. P. 657-670.
11. Chudy-Laskowska, K. & Pisula, T. Prognoza liczby wypadków w Polsce. *Logistyka*. 2014. No. 6. P. 2710-2722.
12. Council, F.M. & Harwood, D.W. & Hauer E. & Hughes, W.E. & Vogt, A. prediction of the expected safety performance of rural two-lane highways. Washington: Federal Highway Administration. 2000. 194 p.
13. Deublein, M. & Schubert, M. & Adey, B.T. & Köhler, J. & Faber, M.H. Prediction of road accidents: A Bayesian hierarchical approach. *Accident Analysis and Prevention*. 2013. No. 51. P. 274-291.
14. El-Basyounym, K. & Sayed, T. Accident prediction models with random corridor parameters. *Accident Analysis and Prevention*. 2009. No. 41. P. 1118-1123.

15. Fernandes, A. & Neves, J. An approach to accidents modeling based on compounds road environments. *Accident Analysis and Prevention*. 2013. No. 53. P. 39-45.
16. Guo, F. & Wang, X. & Abdel-Aty, M. Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis and Prevention*. 2010. No. 42. P. 84-92.
17. Ivan, J.N. & Garder, P.E. & Deng, Z. & Zhang, C. *The effect of segment characteristics on the severity of head-on crashes on two-lane rural highways*. Connecticut: University of Connecticut, University of Maine. 2006. 69 p.
18. Ivan, J.N. & Lord, D. & Washington, S.P. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention*. 2005. No. 37. P. 35-46.
19. Ivan, J.N. & Qin, X. & Ravishanker, N. Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis and Prevention*. 2004. No. 36. P. 183-191.
20. Iyınam, A. & Iyınam, S. & Ergun, M. Analysis of relationship between highway safety and road geometric design elements: Turkish case. In: *Traffic Forum*. Dresden, 2003.
21. Karacasu, M. & Ergül, B. & Altın Yavuz, A. Estimating the causes of traffic accidents using logistic regression and discriminant analysis. *International Journal of Injury Control and Safety Promotion*. 2014. No. 21. Vol. 4. P. 305-313.
22. Khaniyev, T. & Baskir, M.B. & Gokpinar, F. & Mirzayev, F. Statistical distributions and reliability functions with type-2 fuzzy parameters. *Eksplatacja i Niezawodność – Maintenance and Reliability*. 2019. No. 21. Vol. 2. P. 268-274.
23. Kozłowski, E. & Borucka, A. & Świdorski, A. Application of the logistic regression for determining transition probability matrix of operating states in the transport systems. *Eksplatacja i Niezawodność – Maintenance and Reliability*. 2020. No. 22. Vol. 2. P. 192-200.
24. Lee, J. & Mannering, F. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. *Accident Analysis and Prevention*. 2002. No. 34. P. 149-161.
25. Lemp, J.D. & Kockelman, K.M. & Unnikrishnan, A. Analysis of large truck crash severity using heteroskedastic ordered probit models. *Accident Analysis and Prevention*. 2011. No. 43. P. 370-380.
26. Li, X. & Huang, H-Z. & Li, F. & Ren, L. Remaining useful life prediction model of the space station. *Eksplatacja i Niezawodność – Maintenance and Reliability*. 2019. No. 21. Vol. 3. P. 501-510.
27. Ma, J. & Kockelman, K. M. & Damien, P. A multivariate Poisson-longnormal regression model for prediction of crash count by severity, using Bayesian methods. *Accident Analysis and Prevention*. 2008. No. 40. P. 964-975.
28. Ma, Z. & Shao, C. & Ma, S. & Yue, H. Analysis of the Logistic Model for Accident Severity on Urban Road Environment, In: *IV'09 IEEE Intelligent Vehicles Symposium*. Xi'an, 2009.
29. Mannering, F. & Venkataraman, S. & Woodrow, B. Statistical analysis of accident rural freeways. *Accident Analysis and Prevention*. 1996. No. 28. P. 391-401.
30. Migut, G. *Kreator regresji logistycznej*. Kraków: Statsoft Polska. 2011. 13 p.
31. Shakaya, R. & Marsani, A. Using logistic regression to estimate the influence of crash. *Accident Analysis and Prevention*. 2002. No. 34. Vol. 6. P. 729-741.
32. Shankar, V.N. & Ulfarsson, G.F. & Pendyala, R.M. & Nebergall, M.B. Modeling crashes involving pedestrians and motorized traffic. *Safety Science*. 2003. No. 41. P. 627-640.
33. Stanisław, A. *Modele regresji logistycznej. Zastosowanie w medycynie, naukach przyrodniczych i społecznych*. [In Polish: *Logistic regression models. Application in medicine, natural and social sciences*]. Kraków: StatSoft Polska. 2016. 708 p.
34. Świdorski, A. & Borucka, A. Mathematical Analysis of Factors Affecting the Road Safety in Selected Polish Region. In: *Proceedings of the 22nd International Scientific Conference "Transport Means" part II*. Kaunas, 2018.