

## Influence of meteorological conditions on road accidents. A model for observations with excess zeros

Indexed by:



Anna Borucka<sup>a,\*</sup>, Dariusz Pyza<sup>b</sup>

<sup>a</sup>Military University of Technology, Faculty of Security, Logistics and Management, ul. gen. Sylwestra Kaliskiego 2, 00–908 Warsaw, Poland

<sup>b</sup>Warsaw University of Technology, Faculty of Transport, pl. Politechniki 1, 00-661 Warsaw, Poland

### Highlights

- Determining the influence of meteorological conditions on the number of accidents.
- More accurate analysis of accidents by limiting the area (territory).
- Solving the problem of excess zero observations in the accident data set.
- Applying Negative Binomial Hurdle Model in the research on road accidents.

### Abstract

Road accidents are one of the basic road safety determinants. Most research covers large territorial areas. The results of such research do not take into account the differences between individual regions, which often leads to incorrect results and their interpretation. What makes it difficult to conduct analyses in a narrow territorial area is the small number of observations. The narrowing of the research area means that the number of accidents in time units is often very low. There are many zero observations in the data sets, which may affect the reliability of the research results. Such data are usually aggregated, which leads to information loss. The authors have therefore applied a model that addresses such problems. They proposed a method that does not require data aggregation and allows for the analysis of sets with an excess of zero observations. The presented model can be implemented in different territorial areas.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

### Introduction

Road accidents are one of the basic sources of data for road safety analysis [1, 5, 10]. However, it is very difficult to find their causes as there are numerous factors that affect them [25, 26]. The basic road safety analyses carried out in most countries concern the general trends in the number of accidents and casualties in relation to data characterizing a given area. However, determination of the exact causes of accidents requires much more advanced methods. The studies presented in the literature are of varied nature [10, 12]. Some of them are limited to evaluation of the impact of single variables such as driver drowsiness [22], driving speed before the incident [27], traffic jams [19, 29], driver's gender [2], driving under the influence of alcohol or drugs [16], etc. In other publications, many variables are analyzed simultaneously. Singh [23], for example, evaluates the impact of inexperience and lack of skills characteristic of young drivers, while in the group of older drivers he emphasizes impairment of sight, cognitive functions and motor skills. Ashraf et al. [4] also take into account many different elements, considering, among other things, driver's gender, experience, time of incident, observance of traffic rules [20].

There are a lot of publications on road accidents. All of them analyze a limited number of factors, as it is not possible to take into account all variables that affect the number of such incidents. Moreover, not all of them are identifiable or measurable, and some data are difficult to obtain. These include, for example, detailed weather data, which in publicly available form concern only average measurement values for larger administrative areas and sometimes the whole country. Such aggregated values are useless, as meteorological conditions may vary dramatically among distant regions.

Another problem in the analysis of road accidents is the availability of information in this area. In many countries, no accurate records are compiled for areas smaller than the whole country [6, 7], or the available information is not complete [9, 28], so that only country-wide analyses are possible. Examples include the research conducted in Saudi Arabia [3], South Korea [4], India [21] or Poland [8, 11, 24].

The results of such research, however, do not take into account the differences between individual regions, which may occur even within a single country/region. They may result (e.g. when comparing small towns and large agglomerations) from different lifestyles, traffic volumes at different times of the day, different numbers of traffic users, the condition of road infrastructure, and even driver experience or driving culture. Analyses carried out within different areas allow to

(\*) Corresponding author.

E-mail addresses: A. Borucka - [anna.borucka@wat.edu.pl](mailto:anna.borucka@wat.edu.pl), D. Pyza - [dariusz.pyza@pw.edu.pl](mailto:dariusz.pyza@pw.edu.pl)

compare them, find similarities concerning factors conducive to accidents, as well as elements improving road safety, which, when bringing the expected results in one region, can be implemented elsewhere. Systematic research, conducted in parallel in different locations, is therefore desirable. However, in addition to the data availability, the nature thereof poses a significant obstacle in this respect. A significant narrowing of the research area results in a very small number of accidents per time unit and a large number of zero observations in the data sets, which may affect the reliability of research results. Such data are therefore often aggregated prior to analyses [6], which in turn may lead to a significant loss of information.

This paper is part of the analysis of issues related to the trend of continuous improvement of road safety, carried out through monitoring of hazard levels and permanent evaluation of factors that shape it. The authors adopted a research hypothesis stating that meteorological factors significantly influence the number of road accidents. Due to the high variability of weather conditions in relation to geographical location, only the city of Warsaw was analyzed. As a result, in addition to the main research objective, i.e. to indicate meteorological factors that significantly influence the number of accidents, there was an additional objective to present the possibility of mathematical analysis of a set of data with excess zeros thus eliminating the necessity of measurements aggregation and the related loss of information. Moreover, factors related to the time of the incident were also taken into account, i.e.: time, day of the week and month.

The research was conducted using data (including meteorological data) on road accidents by hour that occurred in 2018 and 2019 in Warsaw. Data on accidents were obtained from the Polish Road Safety Observatory (operating at the Motor Transport Institute in Warsaw), while meteorological data were made available by the Warsaw-Okęcie Airport.

The article consists of an introduction, methodological and practical parts and a summary. The introduction presents the research objective and justifies the necessity to conduct it. The methodological part presents the applied analysis methods dedicated to the empirical data gathered. In the practical part, the research sample and the mathematical model of road accidents are characterized in detail. The whole article ends with the summary of the research carried out and the final conclusions.

## 1. Methodology

The numerator variable represents a category whose possible values are non-negative integers. Linear regression is the most common way of studying the influence of independent factors on the explained variable [17], but using the classic model with the endogenous variable being the numerator variable can lead to serious cognitive errors, especially when the expected value of the variable is not large.

Poisson regression is a popular approach to modeling count data [18, 29]. It is assumed that the distribution of observations is consistent with Poisson distribution with the mean depending on the predictors. The problem arises if the empirical data show deviations from the assumptions of this model. In many applications, for example, an excessive dispersion occurs and the assumption of equality of the expected value and variance of distribution is not fulfilled. Therefore, other models are adopted in place of Poisson regression that take into account two types of zeros, i.e., “true zeros” and “excess zeros”, estimating two equations, one for the counting model and one for the excess zeros. The most commonly used are the zero-inflated model and the hurdle model [1, 13, 15, 30, 31].

The article includes an estimation of parameters of four models: the Zero inflated Poisson model (ZIP), the Zero inflated negative binomial model (ZINB), the Poisson hurdle model (PLH), and the Negative binomial hurdle model (NBLH). Using Akaike’s criterion, the selection of the best one was made. A method allowing to simplify the expanded model was then presented and a negligible loss of information that was associated with this was shown.

## 2. Research sample

The presented research is based on the data on road accidents that occurred in the years 2018-2019 in the Polish capital city – Warsaw, archived on an hourly basis. The research sample consisted of 17,250 observations. The narrowed area of research strongly influenced the number of events recorded in each hour. The maximum number of accidents in the analyzed period was as low as 4, and the average value was 0.11. Other descriptive statistics are presented in Table 1.

Table 1. Basic descriptive statistics of the Warsaw road accidents variable

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0	0	0.114	0	4

The reason behind such results of descriptive statistics is that the vast majority of observations are zeros. There are as many as 15,723 of them in the whole set, which represents more than 89% of the measurements. The remaining numbers are presented in Table 2.

Table 2. Number of individual observations in the data set

Number of accidents	0	1	2	3	4
Number of observations	15,723	1,618	162	16	1

The distribution of the data gathered, sorted in ascending order, is shown in Fig. 1.

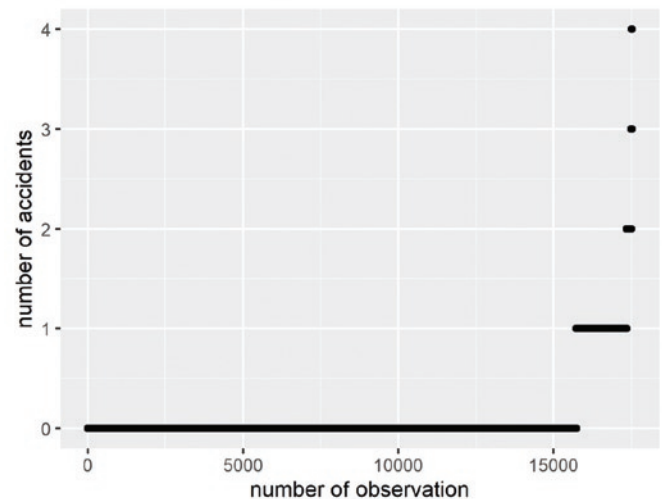


Fig. 1. The observations gathered, sorted in ascending order (by number of accidents)

The form of the dependent variable dictated the use of mathematical models that are dedicated to data with excess zeros. Since, according to the assumed research hypothesis, the research objective was to analyze the influence of meteorological factors on the number of accidents, additional information was collected for each hour describing the weather conditions prevailing then. Detailed data concerning Warsaw were obtained from the Warsaw-Okęcie Airport, from Meteorological Aerodrome Reports. It is a coded weather report format used in aeronautical meteorology and weather forecasting. It contains information about ambient temperature, dew point temperature, pressure, wind speed and direction, precipitation, cloud cover, cloud base height, visibility. It may also contain other important annotations, concerning for example the condition of runways.

The set of factors used in the study contained information on visibility, wind speed, pressure, temperature, precipitation, type of clouds, mist.

Table 3. Fixed effects/basic descriptive statistics

Quantitative variables	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Wind direction [°]	100	160	250	278.2	310.0	901.0
Wind speed [KT]	0	4	6.000	6.729	9.000	77.000
Visibility [m]	180	9,999	9,999	9,340	10,000	10,004
Pressure [hPa]	987	1,012	1,012	1,012	1,012	1,038
Temperature [°C]	-16.00	3.00	11.00	10.88	18.00	36.00
Qualitative variables	Category – number of observations in the set					
Mist	No mist – 15,707	BR – 1,252	FG – 561	where <b>BR</b> - Mist (brume) (visibility 1,000-5,000 m) , <b>FG</b> - Fog (thickness from ground to above 2m, visibility below 1,000 m)		
Clouds	No clouds – 8,030	BKN – 4,451	FEW – 2,271	NSC – 776	OCV – 343	SCT – 1649
	<b>NSC</b> - no significant clouds. <b>FEW</b> - 1-2 octas covered (12.5-25%), <b>SCT</b> - 3-4 octas covered (37.5%-50%), <b>BKN</b> - 5-7 octas covered (62.5%-87.5%), <b>OVC</b> - 8 octas coverage (100%)					
Precipitation	RA - 2865	SN - 300	No precipitation - 14,355	<b>RA</b> – Rain, <b>SN</b> – Snow		

The original data set adopted for the study contained 7 variables that could occur in the fixed effect category and that were used for preliminary model construction, while their descriptive statistics are presented in Table 3. Additionally, the variables resulting from the date – calendar, i.e. month, day of the week and time of the incident, were included.

### 3. Mathematical model of road accidents

The parameters of four models were estimated: Zero inflated Poisson (ZIP) model, Zero inflated negative binomial (ZIMB) model, Poisson logit hurdle (PLH) model, Negative binomial logit hurdle (NBLH) model. Since some of the variables had no significant impact on the number of accidents, the following variables were used for the final estimation of model parameters: clouds, precipitation, mist, temperature, month, week, hour. For the models constructed in this way the value of the AIC information criterion was calculated and on its basis the best of them was selected, which turned out to be the negative binomial hurdle model, for which the AIC value was the lowest (Table 4).

Table 4. Values of AIC information criterion of individual models

model	AIC
Zero inflated Poisson model	11,892
Zero inflated negative binomial model	11,894
Poisson hurdle model	11,806
Negative binomial hurdle model	<b>11,766</b>

Thus, the number of accidents can be presented as a two-part model (see Appendix 1 for estimated parameter values). First of all, it is a logit model, which is designed to model the probability of values  $y_i = 0$ . The second part concerns positive values and is modeled as a variable with negative binomial distribution, taking into account selected predictors. The resulting model can help us to determine which conditions are conducive to road accidents. The model is interpreted as two separate processes. First of all, it is a process that generates zero numbers for road accidents. The constructed model indicates that the probability of no incident is significantly influenced by cloud and fog variables, which

increase this probability. Among the individual categories, overcast turned out to be significant, which is probably due to the increased caution of drivers during such unfavorable weather conditions, as well as cloudless sky and no mist, which in turn increase visibility and facilitate safe driving. The days of the week (Sunday and Tuesday) also proved to be significant, as they increase the probability of accidents. The second part of the model is a process that generates the number of road accidents, taking into account the occurrence of at least one accident. The stimulants in this case are overcast (OVC) and temperature, as well as the following hours: 4:00 a.m. and from 6:00 a.m. to 10:00 p.m. The destimulants are: no precipitation, the months of July, August and November, and the following days of the week: Tuesday and Wednesday.

Not all the factors for individual predictors in groups are statistically significant. Moreover, the model is extensive, due to a large number of independent variables. It was therefore analyzed whether it would be possible to combine variables in individual groups in order to simplify the model.

### 4. Simplified model construction

#### 4.1. Analysis of qualitative variables

To simplify the model, an analysis was made of the possibility of combining the variables in each group. For this purpose, the Kruskal-Wallis test was used to see if there were differences between the variables in the group and then the Wilcoxon rank sum test was used to determine which variables in the group were significantly different [14]. Tests were conducted for each group of variables.

Analysis of individual categories of the cloud group using the Kruskal-Wallis test showed that there are significant differences between at least two categories. The Kruskal-Wallis test statistics are  $T = 22.433$  and  $p - value = 4.33 * 10^{-4}$ . This is confirmed by the interaction plot presented in Figure 2. If the influence of each category in the group was the same, the lines in the plot would be parallel.

In order to find the categories that are significantly different from each other, Wilcoxon rank sum test was used, the results of which are presented in Table 5.

Based on the Wilcoxon rank sum test results, three groups were distinguished. The first one includes cloudless sky and NSC, FEW, SCT clouds. Consistency within the group was again confirmed by the Kruskal-Wallis test ( $T = 7.059$ ,  $p - value = 0.07$ ). In the second group there were only clouds of BKN type, while in the third – of OCV

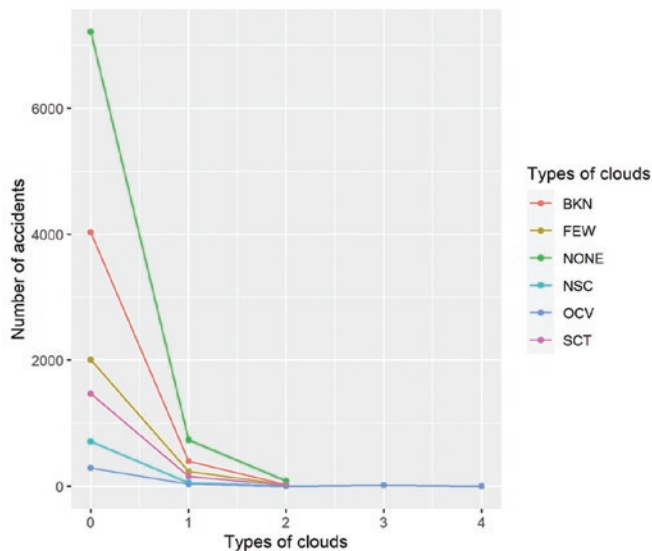


Fig. 2. Interaction plot of individual categories in the cloud group

Table 5. Pairwise comparisons using Wilcoxon rank sum test for the cloud variable

	BKN	FEW	NONE	NSC	OCV
FEW	0.020				
NONE	0.222	0.119			
NSC	0.415	0.038	0.171		
OCV	0.002	0.043	0.004	0.002	
SCT	0.171	0.470	0.496	0.123	0.020

Table 6. Pairwise comparisons using Wilcoxon rank sum test for the month variable

	January	February	March	April	May	June	July	August	September	October	November
February	0.038										
March	0.543	0.161									
April	0.222	0.001	0.060								
May	0.106	0.001	0.020	0.731							
June	0.166	0.001	0.038	0.892	0.852						
July	0.199	0.466	0.556	0.010	0.003	0.006					
August	0.852	0.062	0.710	0.157	0.062	0.108	0.307				
September	0.166	0.001	0.038	0.892	0.852	0.997	0.006	0.108			
October	0.463	0.004	0.147	0.719	0.466	0.592	0.035	0.307	0.592		
November	0.045	0.915	0.189	0.001	0.001	0.001	0.529	0.079	0.001	0.004	
December	0.189	0.478	0.543	0.009	0.003	0.006	0.543	0.295	0.006	0.033	0.543

Table 7. Pairwise comparisons using Wilcoxon rank sum test for the day of the week variable

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Tuesday	0.042					
Wednesday	0.221	0.519				
Thursday	0.642	0.001	0.013			
Friday	0.537	0.006	0.045	0.642		
Saturday	0.235	0.001	0.008	0.814	0.537	
Sunday	0.070	0.000	0.000	0.519	0.279	0.001

type. The analysis of the precipitation variable showed no significant differences in individual categories. Kruskal-Wallis T test statistics = 4.416 and p-value = 0.11.

Next, the variables related to time, i.e. month, day of the week and time of the incident were analyzed. The results of the Kruskal-Wallis test for the month variable indicate significant differences in the group ( $T=58.079$ ,  $p\text{-value}=2.110^{-08}$ ). Based on the Wilcoxon test results (Table 6) three groups of similar months were distinguished.

The following were distinguished:

- Group 1, which included April, May, June, September and October. Consistency within the group was confirmed again by the Kruskal-Wallis test ( $T=1.027$ ,  $p\text{-value}=0.906$ ).
- Group 2, which included the months of January, February, March, July, August, November and December ( $T=11.564$ ,  $p\text{-value}=0.0724$ ).

The analysis of individual days of the week also revealed the existence of significantly different groups ( $T=61.524$ ,  $p\text{-value}=2.210^{-11}$ ), which were created on the basis of the Wilcoxon test results (Table 7).

Two groups of days of the week were created: Group 1, which included Monday, Thursday, Friday, Saturday, Sunday (consistency within the group was confirmed by the Kruskal-Wallis test,  $T=5.096$ ,  $p\text{-value}=0.278$ ) and group 2, which included Tuesday and Wednesday ( $T=0.723$ ,  $p\text{-value}=0.395$ ). The last variable studied was the time of the incident, for which the zero hypothesis of equal distribution in groups was also rejected ( $T=723.01$ ,  $p\text{-value}<2.2\cdot 10^{-16}$ ).

Based on the Wilcoxon test, the following groups were created:

- Group 1: 10:00 p.m., 06:00 a.m., 11:00 p.m., 09:00 p.m. ( $T=9.733$ ,  $p\text{-value}=0.021$ ).
- Group 2: 00:00, 01:00 a.m., 02:00 a.m., 03:00 a.m., 04:00 a.m., 05:00 a.m. ( $T=11.25$ ,  $p\text{-value}=0.0467$ ).
- Group 3: 07:00 a.m., 08:00 a.m., 09:00 a.m., 10:00 a.m., 11:00 a.m., 12:00, 01:00 p.m., 02:00 p.m., 03:00 p.m., 06:00 p.m., 07:00 p.m., 08:00 p.m. ( $T=22.136$ ,  $p\text{-value}=0.024$ ).
- Group 4: 04:00 p.m., 05:00 p.m. ( $T=0.002$ ,  $p\text{-value}=0.969$ ).

When it comes to the time of the incident variable, the consistency within groups was confirmed by the Kruskal-Wallis test at the significance level of  $\alpha = 0.01$ , and therefore, a chi-squared test was also performed, which is also used to compare the distributions in groups. Consistency was confirmed at the significance level of  $\alpha = 0.05$  (Table 8).

Table 8. Chi-squared test results for each group of days of the week

	Pearson's Chi-squared test	p-value
Group 1	12.415	0.053
Group 2	16.357	0.090
Group 3	59.573	0.059
Group 4	0.145	0.986

#### 4.2. Estimation of simplified model parameters

Grouping of variables allowed to construct a simplified Negative binomial hurdle model. Estimates of parameters of the first and second part of the model are presented in Table 9.

The constructed model is simpler and thus more transparent. The influence of individual variables is obviously the same as in the extended model. The AIC criterion is 11,897, compared to the AIC = 11,766 obtained for the model before grouping, which means a slight loss of quality in the context of significant model simplification. The adjustment of the proposed model to the empirical data is presented in Fig 3.

#### 5. Summary

In order to address the problems presented in the introduction, the article proposes a mathematical model allowing to estimate the number of road accidents, including a correction for random effect (i.e. resistant to excess zeros in the data set) and eliminating the problem of excessive dispersion by applying the binomial negative distribution. The application of such a model to traffic accidents is virtually non-existent/unnoticeable in the literature. This is because traffic accident data are usually aggregated to lower frequency data or such events are considered for large areas. While this provides a sufficient number of observations for analysis, it is associated with significant data loss or even obtaining a model that is inadequate for individual component areas. Therefore, this article proposes a model that solves these problems while providing a reliable assessment of

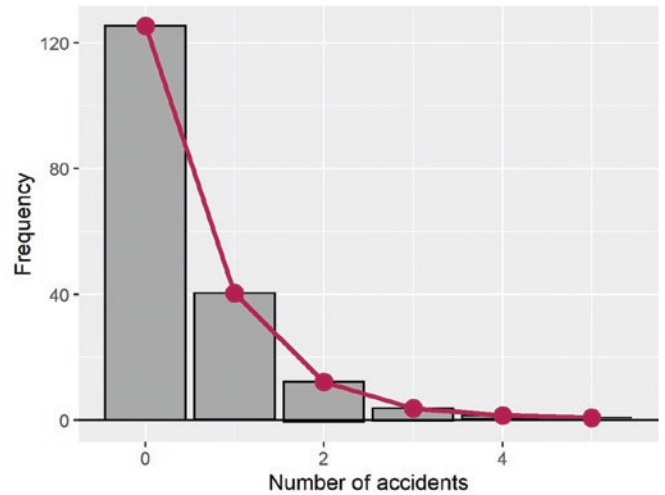


Fig. 3. Adjustment of the Negative binomial hurdle model (red line) to empirical data

the factors affecting accidents for a narrow area and a high frequency of observations. In this study, meteorological factors were the main focus, however other variables that were not used in this case, e.g., terrain characteristics, traffic conditions, vehicle type, etc., can also be studied in this way.

The authors focused on meteorological factors because they are often considered the cause of accidents, and there are few studies that support this. Some of the numerous variables being analyzed turned out not to significantly influence road hazard occurrence. Temperature, precipitation, type of cloud coverage and mist turned out to be significant. Moreover, the impact of variables related to the date of the event, i.e. calendar month, day of the week and time of the accident, was also significant.

The presented study shows that selected weather factors influence the number of accidents. This may be due to their impact on the condition of road traffic users and is an important part of further work in this area. Furthermore, the results obtained prompt us to consider other factors not taken into account here, such as traffic volume, which can be correlated with weather conditions (cloudy, rainy days may be conducive to vehicle use) as well as the date of the incident (peak traffic hours, varying traffic volumes depending on the day of the week or month). The above assumptions will be the subject of further research / investigations by the authors.

Table 9. Estimates of the parameters of the simplified Negative binomial hurdle model

	Group	Factor	First part of the model				Second part of the model			
			Estimate	Std. Error	z value	Pr(> z )	Estimate	Std. Error	z value	Pr(> z )
Parameters		Intercept	-4.482	0.617	-7.261	0.000	-2.942	0.130	-22.672	< 2e-16
	Clouds	gr I	0.337	0.244	1.383	0.167	-0.094	0.067	-1.395	0.163
		OCV	2.256	0.269	8.385	<2e-16	0.612	0.165	3.714	0.000
	Fog	FG	-12.756	0.590	-30.73	<2e-16	-0.165	0.220	-0.748	0.454
		No fog	1.039	0.455	2.284	0.022	0.037	0.110	0.333	0.739
	Temperature		0.009	0.009	0.934	0.351	0.009	0.003	2.668	0.008
	Month	Group I	0.487	0.162	3.004	0.003	0.312	0.056	5.525	0.000
	Day of the week	Group 2	0.256	0.143	1.794	0.073	0.085	0.056	1.530	0.126
	Hour	Group II	-0.817	1.076	-0.759	0.448	-1.252	0.137	-9.149	<2e-16
		Group III	1.036	0.416	2.494	0.013	0.902	0.084	10.772	<2e-16
Group IV		1.326	0.430	3.081	0.002	1.258	0.103	12.270	<2e-16	

## Appendix 1. Estimates of parameters of the Negative binomial hurdle model

NBLH		First part of the model				Second part of the model			
Parameter		Estimate	Std, Error	z-value	Pr(> z )	Estimate	Std, error	z value	Pr(> z )
(Intercept)		-7.096	6.877	-1.032	0.302	-3.446	0.302	-11.417	<210 <sup>-16</sup>
Clouds	FEW	0.285	0.299	0.952	0.341	-0.071	0.091	-0.785	0.432
	None	0.249	0.264	0.941	0.347	-0.113	0.077	-1.474	0.140
	NSC	1.299	0.397	3.271	0.001	0.199	0.157	1.272	0.204
	OCV	2.146	0.289	7.427	1.1110 <sup>-13</sup>	0.701	0.170	4.127	3.6710 <sup>-5</sup>
	SCT	0.457	0.319	1.432	0.152	-0.112	0.101	-1.112	0.266
Precipitation	SN	-0.694	1.022	-0.679	0.497	-0.236	0.236	-1.003	0.316
	no	-0.247	0.180	-1.371	0.170	-0.147	0.075	-1.970	0.049
Fog	FG	-10.960	85.375	-0.128	0.898	-0.234	0.230	-1.021	0.307
	None	1.152	0.468	2.464	0.014	0.064	0.113	0.564	0.573
Temperature		0.019	0.017	1.106	0.269	0.013	0.006	2.232	2.232
Month	January	-0.177	0.560	-0.316	0.752	0.084	0.182	0.462	0.644
	February	-0.194	0.623	-0.311	0.756	-0.256	0.187	-1.369	0.171
	March	-0.160	0.514	-0.311	0.756	-0.055	0.163	-0.339	0.735
	April	0.366	0.280	1.310	0.190	0.110	0.131	0.842	0.400
	May	0.157	0.296	0.530	0.596	0.074	0.120	0.617	0.537
	July	-0.599	0.364	-1.644	0.100	-0.409	0.126	-3.250	0.001
	August	-0.288	0.336	-0.857	0.391	-0.249	0.121	-2.052	0.040
	September	0.209	0.294	0.711	0.477	0.096	0.123	0.780	0.435
	October	0.163	0.352	0.462	0.644	0.067	0.137	0.488	0.626
	November	-0.102	0.459	-0.223	0.824	-0.403	0.168	-2.399	0.017
	December	0.266	0.482	0.551	0.582	-0.178	0.173	-1.024	0.306
Day of the week	Monday	-0.242	0.242	-1.003	0.316	-0.116	0.094	-1.231	0.218
	Tuesday	-0.651	0.339	-1.922	0.055	-0.368	0.099	-3.708	0.000
	Wednesday	-0.217	0.261	-0.831	0.406	-0.282	0.097	-2.918	0.004
	Friday	-0.116	0.225	-0.515	0.607	-0.054	0.093	-0.580	0.562
	Saturday	-0.235	0.227	-1.032	0.302	0.015	0.092	0.162	0.871
	Sunday	-0.632	0.264	-2.390	0.017	0.087	0.091	0.958	0.338
Hour	1:00 AM	-4.167	67.952	-0.061	0.951	-0.565	0.383	-1.476	0.140
	2:00 AM	4.117	6.917	0.595	0.552	-0.307	0.357	-0.860	0.390
	3:00 AM	-3.119	52.172	-0.060	0.952	-0.645	0.395	-1.634	0.102
	4:00 AM	-2.548	47.382	-0.054	0.957	-1.168	0.472	-2.477	0.013
	5:00 AM	-6.395	160.937	-0.040	0.968	0.039	0.325	0.118	0.906
	6:00 AM	3.639	6.870	0.530	0.596	0.948	0.277	3.423	0.001
	7:00 AM	4.478	6.853	0.654	0.513	1.577	0.260	6.060	1.3610 <sup>-9</sup>
	8:00 AM	4.405	6.851	0.643	0.520	1.931	0.255	7.573	3.6410 <sup>-14</sup>
	9:00 AM	4.297	6.852	0.627	0.531	1.584	0.261	6.069	1.2910 <sup>-9</sup>
	10:00 AM	4.156	6.854	0.606	0.544	1.724	0.259	6.655	2.8310 <sup>-11</sup>
	11:00 AM	3.549	6.859	0.517	0.605	1.828	0.258	7.089	1.3510 <sup>-12</sup>
	12:00	4.640	6.851	0.677	0.498	1.814	0.258	7.022	2.1910 <sup>-12</sup>
	1:00 PM	3.808	6.857	0.555	0.579	1.886	0.258	7.326	2.3710 <sup>-13</sup>
	2:00 PM	4.180	6.854	0.610	0.542	1.724	0.260	6.644	3.0610 <sup>-11</sup>
	3:00 PM	4.130	6.853	0.603	0.547	1.783	0.258	6.904	5.0710 <sup>-12</sup>
	4:00 PM	4.423	6.850	0.646	0.519	2.146	0.253	8.475	<210 <sup>-16</sup>
	5:00 PM	4.530	6.850	0.661	0.508	2.135	0.253	8.450	<210 <sup>-16</sup>
	6:00 PM	4.297	6.850	0.627	0.531	2.034	0.253	8.027	9.9810 <sup>-16</sup>
	7:00 PM	4.256	6.853	0.621	0.535	1.871	0.255	7.327	2.3510 <sup>-13</sup>
	8:00 PM	3.533	6.860	0.515	0.607	1.561	0.261	5.991	2.08*10 <sup>-9</sup>
9:00 PM	2.561	6.919	0.370	0.711	1.134	0.271	4.186	2.8310 <sup>-5</sup>	
10:00 PM	3.423	6.882	0.497	0.619	0.908	0.278	3.262	0.001	
11:00 PM	-8.839	401.660	-0.022	0.982	0.429	0.300	1.429	0.153	

## References

1. Aga MA, Woldeamanuel BT, Tadesse M. Statistical modeling of number of human deaths per road traffic accident in Oromia region, Ethiopia. *PLoS one* 2020; 16(5) e0251492, <https://doi.org/10.1371/journal.pone.0251492>.
2. Al-Balbissi AH. Role of gender in road accidents. *Traffic Injury Prevention* 2003; 4(1): 64-73, <https://doi.org/10.1080/15389580309857>.
3. Ansari S, Akhdar F, Mandoorah M, Moutaery K. Causes and effects of road traffic accidents in Saudi Arabia. *Public health* 2000; 114(1): 37-39, <https://doi.org/10.1038/sj.ph.1900610>.
4. Ashraf I, Hur S, Shafiq M, Park Y. Catastrophic factors involved in road accidents: Underlying causes and descriptive analysis. *PLoS one* 2019; 14(10), e0223473, <https://doi.org/10.1371/journal.pone.0223473>.
5. Borucka A, Kozłowski E, Oleszczuk P, Świdorski A. Predictive analysis of the impact of the time of day on road accidents in Poland. *Open Engineering* 2020; 11(1): 142-150, <https://doi.org/10.1515/eng-2021-0017>.
6. Edwards JB. The temporal distribution of road accidents in adverse weather. *Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling* 1999; 6(1): 59-68, <https://doi.org/10.1017/S1350482799001139>.
7. Elvik R, Mysen A. Incomplete accident reporting: meta-analysis of studies made in 13 countries. *Transportation Research Record* 1999; 1665(1), 133-140, <https://doi.org/10.3141/1665-18>.
8. Frej D, Ludwinek K. Analysis of road accidents in 2002-2019 on the example of Poland. *The Archives of Automotive Engineering - Archiwum Motoryzacji* 2020; 89(3): 5-18.
9. Golias J, Yannis G. Dealing with lack of exposure data in road accident analysis. In 12th 557 International Conference: Traffic Safety on Three Continents 2001, September; 558.
10. Guzek M, Lozia Z. Computing Methods in the Analysis of Road Accident Reconstruction Uncertainty. *Archives of Computational Methods in Engineering* 2020; 28: 2459-2476, <https://doi.org/10.1007/s11831-020-09462-w>.
11. Jacyna M, Merkisz J. Proecological approach to modelling traffic organization in national transport system. *Archives of Transport* 2014; 30(2): 31-41, <https://doi.org/10.5604/08669546.1146975>.
12. Jamroz K, Budzyński M, Romanowska A, Żukowska J, Oskarski J, Kustra W. Experiences and challenges in fatality reduction on polish roads. *Sustainability* 2019; 11(4): 959, <https://doi.org/10.3390/su11040959>.
13. Kasin JA, Papastathopoulos I. A spatial Poisson hurdle model with application to wildfires. *arXiv preprint arXiv:2007.00137*, 2020.
14. Kozłowski E, Mazurkiewicz D, Żabiński T, Prucnal S, Sep J. Machining sensor data management for operation-level predictive model. *Expert Systems with Applications* 2020; 159, 113600, <https://doi.org/10.1016/j.eswa.2020.113600>.
15. Lukusa MT, Phoa FKH. A Horvitz-type estimation on incomplete traffic accident data analyzed via a zero-inflated Poisson model. *Accident Analysis & Prevention* 2020; 134, 105235, <https://doi.org/10.1016/j.aap.2019.07.011>.
16. Martin JL, Gadegbeku B, Wu D, Viallon V, Laumon B. Cannabis, alcohol and fatal road accidents. *PLoS one* 2017; 12(11), e0187320, <https://doi.org/10.1371/journal.pone.0187320>.
17. Mitkow S, Świdorski A. Regression Model In Road Transport Services. *Business Logistics in Modern Management* 2019; 19: 263-275
18. Pourhassan MR, Raissi S, Hafezalkotob A. A simulation approach on reliability assessment of complex system subject to stochastic degradation and random shock. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2020; 22 (2): 370-379, <https://doi.org/10.17531/ein.2020.2.20>.
19. Pyza D, Jachimowski R. Modelling of Parcels' Transport System. *Proceedings of the International Scientific Conference Transport Means* 2015; 22-23.
20. Pyza D, Jacyna-Golda I, Gołda P, Gołbiowski P. Alternative Fuels and Their Impact on Reducing Pollution of the Natural Environment. *Annual Set The Environment Protection* 2018; 20: 819-836.
21. Rolison JJ, Regev S, Moutari S, Feeney A. What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records. *Accident Analysis & Prevention* 2018; 115: 11-24, <https://doi.org/10.1016/j.aap.2018.02.025>.
22. Sagberg F. Road accidents caused by drivers falling asleep. *Accident Analysis & Prevention* 1999; 31(6): 639-649, [https://doi.org/10.1016/S0001-4575\(99\)00023-8](https://doi.org/10.1016/S0001-4575(99)00023-8).
23. Singh SK. Road traffic accidents in India: issues and challenges. *Transportation Research Procedia* 2017; 25: 4708-4719, <https://doi.org/10.1016/j.trpro.2017.05.484>.
24. Świdorski A, Borucka A, Skoczynski P. Characteristics and Assessment of the Road Safety Level in Poland with Multiple Regression Model, *Transport Means - Proceedings of the International Conference* 2018; Part I : 92-97.
25. Świdorski A. Decision-Making Problems in the Standardization and Certification of Transport Services, *Wydawnictwa Komunikacji i Łączności*, 2019.
26. Świdorski A. Modeling of transport processes in terms of seasonality of transport. *Scientific Journal of Polish Naval Academy* 2019; 216(1): 103-116, <https://doi.org/10.2478/sjpn-2019-0008>.
27. Taylor MC, Lynam DA, Baruya A. The effects of drivers' speed on the frequency of road accidents. *Crowthorne: Transport Research Laboratory*, 2000.
28. Trépanier M, Leroux MH, de Marcellis-Warin N. Cross-analysis of hazmat road accidents using multiple databases. *Accident Analysis & Prevention*, 2009; 41(6): 1192-1198, <https://doi.org/10.1016/j.aap.2008.05.010>.
29. Wang C, Quddus MA, Ison SG. Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. *Accident Analysis & Prevention*, 2009; 41(4): 798-808, <https://doi.org/10.1016/j.aap.2009.04.002>.
30. Zhang X, Yi N. Fast zero-inflated negative binomial mixed modeling approach for analyzing longitudinal metagenomics data. *Bioinformatics* 2020; 36(8): 2345-2351, <https://doi.org/10.1093/bioinformatics/btz973>.
31. Zuur AF. Zero inflated models and generalized linear mixed models with R, 2012: 574.50182 Z8.