ANALYSIS OF TRAFFIC ACCIDENTS AND THE DEPLOYMENT OF THE FIRE RESCUE SERVICE IN THE CZECH REPUBLIC

[a]KOLISKOVA PETRA, [b]NEUBAUER JIRI

[ab]University of Defence, Kounicova 65, 66210 Brno
email: [a]petra.koliskova@unob.cz, [b]jiri.neubauer@unob.cz

This paper was supported from the research project Conduct of Land Operations, LANDOPS, Ministry of the Defence of the Czech Republic.

Abstract: Traffic accidents remain a topical safety issue. The aim of this article is to use the statistical apparatus to describe a random variable, the daily number of accidents, and to model it. The data on traffic accidents for the period 2012–2021, including time and location information, came from the database of the Fire Rescue Service of the Czech Republic. The study utilised an analysis of variance for statistical analysis with a count variable, following a Poisson distribution. One-factor and two-factor analyses are used to describe the dependence of the daily number of accidents involving the deployment of the fire rescue service on the day of the week, and month. The use of generalised linear model for count data to analyse the number of traffic accidents is unique in the Czech Republic.

Keywords: Traffic accident, fire rescue service, generalised linear models, Poisson distribution.

1 Introduction

Passenger vehicle transport is an integral part of the modern world. Although modern vehicles are equipped with all sorts of technologies and are, compared to the past, manufactured with a much greater regard for safety, they are becoming more accessible and more prevalent on the roads, resulting in denser traffic, as well as more powerful, contributing to an increasing number of accidents. The deployment of the fire rescue service is required after traffic accidents for the purposes of car extrication, the cleanup of leaked operating fluids, the need to carry out fire protection measures, or to restore smooth traffic flow. In the past 10 years, there have been 198,773 such accidents in the 14 regions of the Czech Republic. The focus of the study presented in this article is traffic accidents that involved the deployment of the fire rescue service in the Czech Republic in the period 2012-2021. The records of accidents come from the database of the Fire Rescue Service of the Czech Republic. The accidents are identified by geographical location and the time of occurrence. The time of occurrence is of particular interest in terms of the extent to which the month and day of the week influence the occurrence of a traffic accident. The aim of the presented study is to analyse the accident rate involving the deployment of the fire rescue service in the Czech Republic, to estimate the probability of the daily number of accidents, and to model this random variable using an appropriate regression function.

There are many articles from both the Czech Republic and abroad that are devoted to the topic of traffic accidents. Several studies deal with the analysis of traffic accidents registered in the database of the Police of the Czech Republic (Kvet, 2022; Senk, 2012). Senk (2012) introduces the possibility of using accident prediction models to identify dangerous places on roads. These models are based on a generalised linear model, namely negative binomial regression. Traffic accidents in a selected area of the Slovak Republic are discussed in Harantová (2019). The aim was to analyse the probability distribution of the number of traffic accidents, the number of injuries, and the number of deaths in a certain area using various methods. In terms of causes, the vehicle, the driver and the road are examined as factors. Bačkalić (2013) presents a time analysis of traffic accidents, focusing on the creation of time maps. These maps can be compiled for individual months, days, hours, and for different road users – motorists, cyclists, pedestrians, etc.

The topic of fire rescue service deployment is currently also being researched by experts from the University Burgundy Franche-Comté in France. As the basis for their research, the authors are using data on the deployment of the fire rescue service in the Doubs department of France for the period 2012-2017. They added up to 747 statistical characters to the data obtained from the fire rescue service's database, e.g. meteorological data, astrological data, calendar data and traffic events (Guyeux, 2020). Cerna (2019) mentions the long short-term memory method (LSTM) from the field of artificial neural networks. This research is further extended and developed in Cerna (2020), in which the method is compared to another method – extreme gradient boosting (XGBoost). The aim was to make predictions for each year based on data from previous years. The predictions for the years in which natural disasters occurred, and which therefore saw an unexpected increase in rescue operations, were less accurate. Regarding the method of comparison, the results proved comparable, with the XGBoost approach proving better for extreme event recognition during natural disasters. However, the authors also concluded that none of the neural network designs used were ideal due to the limits of the numbers of neurons and hidden layers, and that it is therefore important to continue the search for another approach in order to create the best neural network for the description of fire rescue service deployment. Mallouhy (2022) chooses a different approach. The aim of the study was to predict the number of fire rescue service deployments through the application of the exponential smoothing method. Three types of exponential smoothing were used on a specific dataset, namely simple, double (Holt) and triple (Holt-Winters) smoothing. The results of the calculations showed that the Holt-Winters method results in the fewest prediction errors and is therefore the most suitable of the three methods. This conclusion corresponds to the seasonal nature of the researched data, which appears to reflect hourly or weekly periodic changes in human activities that affect the need for fire rescue service deployment.

Using neural network models is a common approach in the field of traffic accident research. Zheng (2020) deals with the precise determination of traffic rush hours based on the deep learning method and LSTM, which are compared to the traditional time series approach models, namely ARIMA (autoregressive integrated moving average) and BPNN (backpropagation neural network). The conclusion of this comparison is that the predictions made via the LSTM method are the most accurate in comparison to real measured data. Time series models are used in Quddus (2008), which introduces integer autoregressive (INAR) Poisson models that feature Poisson regression properties, but describe time series. Although the INAR model was found to be better than the ARIMA model at describing serial correlations, it has its limits in terms of the seasonality and heterogeneity of the data.

There are a number of variations of the Poisson regression model depending on the properties. The Poisson regression approach is discussed in Njå (2021). The authors study accidents involving fire in Norwegian road tunnels. Two Poisson regression models were used to find the key factors influencing fires in tunnels. The authors compare their results with those in articles with a similar focus and in conclusion suggest the use of other methods, namely Bayesian Poisson regression. Lord (2005) deals with the comparison of Poisson regression, Poisson-gamma and zero-inflated models on motor vehicle accident data. They note that the Poisson and negative binomial models serve as statistical models for accident processes, with the Poisson model showing better results for almost homogeneous conditions, while the negative binomial model is better suited to describe other conditions.

Among the factors influencing the occurrence of traffic accidents is undoubtedly the influence of weather. Gao (2016) analyses traffic accidents in the Chinese city of Shantou and examines the impact of meteorological parameters on accident rates using time series methods, correlation analysis and multiple linear regression analysis. The results show that road traffic injuries correlate positively with temperature and duration of sunshine, and correlate negatively with wind speed.

2 Data and methods

The initial dataset contained records of traffic accidents investigated by fire rescue service units from 1 January 2012 to 31 December 2021. During this period 198,773 such accidents occurred in the Czech Republic. Each accident was described with the precise time and location data. This article focuses on analysing the dependence of the daily number of traffic accidents on such factors as *day of the week* and *month*. An overview of the basic statistics regarding the daily number of traffic accidents across the Czech Republic, including the median and upper and lower quartiles, is shown graphically in the boxplots in Figures 1–2 and described in Tables 1–2.

Figure 1: Boxplots of the daily number of accidents for day of the week

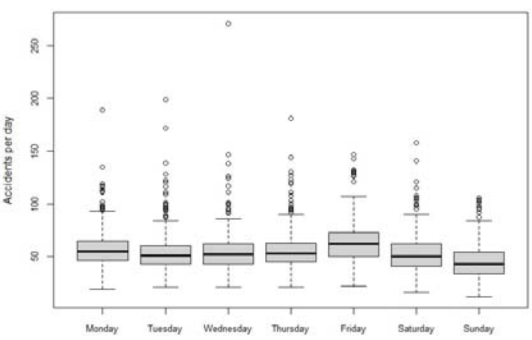
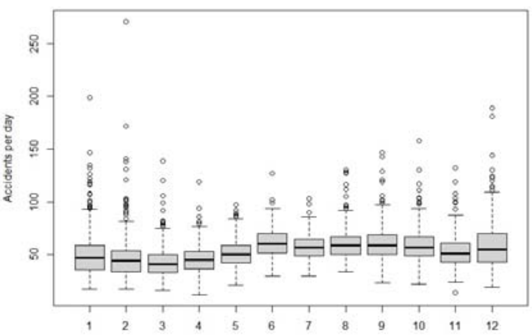


Figure 3: Boxplots of the daily number of accidents in months



Tab. 1: Characteristics of the daily number of accidents for day of the week

Day	*n*	*Mean*	*Median*	*Standard deviation*	*Min*	*Max*
Monday	522	56.54	55	17.85	19	189
Tuesday	522	53.42	51	18.44	21	199
Wednesday	522	54.05	52	19.06	21	271
Thursday	522	55.56	53	17.82	21	181
Friday	522	63.31	63	18.84	22	147
Saturday	521	52.86	50	17.34	16	158
Sunday	522	45.15	43	15.54	12	106

Tab. 2: Characteristics of the daily number of accidents in months

Month	*n*	*Mean*	*Median*	*Standard deviation*	*Min*	*Max*
January	310	52.44	47	23.78	17	199
February	283	48.63	44	25.01	17	271
March	310	42.82	41	15.58	16	139
April	300	45.42	45	13.45	12	119
May	310	51.47	50	13.08	21	97
June	300	61.24	60	13.90	30	127
July	310	58.03	57	11.80	30	103
August	310	60.20	59	14.80	34	130
September	300	60.98	59	16.80	23	147
October	310	58.98	57	16.59	22	158
November	300	53.59	51	16.05	14	132
December	310	58.78	55	23.76	19	189

The discussion now turns to the chosen modelling methods and a brief summary of the theory. The modelled variable is the average number of accidents per day in the Czech Republic. Information on the geographical location of the accident and the exact time of the accident were available as part of the data regarding the accidents. The aim of the study presented in this article is to verify the dependence of the daily number of accidents on the day of the week and month. The analysis of variance ANOVA for count data (Poisson ditribution) used for the dependency analysis in this article is a method based on a generalised linear model, for more details see Dobson (2008).

ANOVA for count data is an example of generalised linear model, whereby the Poisson distribution of the explanatory random variable is assumed. The estimate of the mean value of this variable is then expressed using a logarithmic link function. To determine whether the variable in question is subject to the Poisson distribution, the Kolmogorov-Smirnov, Cramer-von Mises or Anderson-Darling tests can be used. For the Poisson probability distribution of a random variable, the mean value and the variance parameter are identical and correspond to the parameter λ, i.e. the average number of observed events per unit of time. If the dispersion parameter happens to be φ-times the mean, the model is said to exhibit overdispersion and the calculations need to take into account the estimate of the dispersion parameter φ, which can be done in the computational setting quasi-poisson. Explanatory variables, referred to as factors, have only a small number of variations by which their values can be classified into groups. Proving the dependence of an explanatory variable on factors consists in demonstrating different variations of this variable in different groups created by sorting the values of the explanatory variable by factor. In doing so, the Poisson probability distribution of the explanatory random variable is assumed.

Parametric ANOVA can be used if several assumptions are met: 1) the error components of the explanatory variable have a normal distribution; 2) the mean error is zero; 3) the errors are independent; 4) the errors have the same variance within groups. To verify the normality of the explanatory variable, the Shapiro, Lilliefors and Anderson-Darling tests can be used. The ANOVA method then determines whether the mean values of the explanatory random variable differ between groups. Parameters that characterise the mean of the assumed probability distribution of the explanatory variable can be estimated from empirical data in advance using the method of maximum likelihood estimation.

When the assumptions of normality for parametric analysis of variance are not met, one of the options is to assess the dependence of variables by means of nonparametric analysis, such as the Kruskal-Wallis test, see for example Devore (2012), or to choose a generalised linear model. The result shows whether the explanatory random variable depends on the given factor. Conducting Nemenyi's All-Pairs Rank Comparison Test, a multiple comparison test, further shows which sorted groups differ. Finally, it is possible to calculate from the coefficients of the linear model, how the mean of the groups changes with respect to a specified reference value, e.g. the mean. Compared to the linear model, the conditions of the generalised linear model are more general, which allows its use in cases where the explanatory variable is not normally distributed, or when it is a non-linear function of parameters, or when error terms are correlated or heteroscedastic, see Dobson (2008).

3 Results

At the beginning of the model creation, tests of the normal distribution of the daily number of accidents throughout the Czech Republic were performed, namely Shapiro, Lilliefors and Anderson-Darling tests. Test p-values were studied for all months of the period in question. Most of the values were less than the chosen significance level of 0.05, implying that the normality hypothesis could be rejected. Since this implies a different than normal distribution, a Kruskal-Wallis test, a nonparametric alternative to one-factor analysis of variance, was conducted for each factor, *day of the week* and *month*.

A sample of data for which the tests did not reject the Poisson probability distribution were chosen for analysis. The result would be similar even for another small part of the studied dataset. The dataset was therefore modelled using generalised linear model for count data (Poisson distribution). In the computational programme, the natural logarithm was chosen as the link function. The calculations were performed in R, version 4.3.1.

3.1 One-factor analyses

According to the Kruskal-Wallis test, the p-value is much smaller than 0.05 for each of the factors *day of the week* and *month* (p-value < 0.001), so the null hypothesis of equality of the mean values between the groups of explanatory variables, sorted by the factors *day of the week* and *month*, can be ruled out. This means that there is always at least one group which is statistically different from the others. Based on this test, it can be said that the daily number of accidents depends on the factors monitored. Nemenyi's multiple comparison tests, in which p-values less than 0.05 are significant, show that when sorted by the factor *day of the week*, Friday and Sunday differ statistically significantly from the other days (see Table 3). When sorted by the factor *month*, several statistically significant different groups emerge, with the group of summer months differing the most from the others (see Table 4).

Tab. 3: Multiple comparison test for days of the week (p-values)

	Mon.	*Tues.*	*Wed.*	*Thur.*	*Fri.*	*Sat.*
Tuesday	0.004	–	–	–	–	–
Wednesday	0.112	0.944	–	–	–	–
Thursday	0.850	0.205	0.832	–	–	–
Friday	<0.001	<0.001	<0.001	<0.001	–	–
Saturday	0.002	1.000	0.894	0.144	<0.001	–
Sunday	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Tab. 4: Multiple comparison test for months (p-values)

Month	*1*	*2*	*3*	*4*	*5*	*6*
2	*0.148*	–	–	–	–	–
3	*<0.001*	0.173	–	–	–	–
4	*0.027*	1.000	0.457	–	–	–
5	*0.906*	0.001	<0.001	<0.001	–	–
6	*<0.001*	<0.001	<0.001	<0.001	<0.001	–
7	*<0.001*	<0.001	<0.001	<0.001	<0.001	0.690
8	*<0.001*	<0.001	<0.001	<0.001	<0.001	0.995
9	*<0.001*	<0.001	<0.001	<0.001	<0.001	0.999
10	*<0.001*	<0.001	<0.001	<0.001	<0.001	0.427
11	*0.259*	<0.001	<0.001	<0.001	0.997	<0.001
12	*<0.001*	<0.001	<0.001	<0.001	0.015	0.001

Month	*7*	*8*	*9*	*10*	*11*
2	–	–	–	–	–
3	–	–	–	–	–
4	–	–	–	–	–
5	–	–	–	–	–
6	–	–	–	–	–
7	–	–	–	–	–
8	*0.999*	–	–	–	–
9	*0.996*	1.000	–	–	–
10	*1.000*	0.982	0.954	–	–
11	*<0.001*	<0.001	<0.001	0.001	–
12	0.471	0.059	0.036	0.734	0.302

The model of one-factor analysis of variance for the day of the week shows that the average number of accidents involving fire rescue service deployment per day in the Czech Republic is 54.18 (see Table 5). The boxplots in Figure 1 and the ANOVA model show that on Fridays, 63.31 accidents on average occur, which represents 1.17 times the average. Mondays, with 1.04 times the average, and Thursdays, with 1.03 times the average, are both also above the average number of accidents. Wednesdays do not differ statistically significantly from the average. The fewest average accidents (45.15) occurred on Sundays.

Tab. 5: GLM – one-factor – parameter estimates for day

Factor	*Estimate*	*p-value*		*Times*	*Mean*
Intercept	3.9923	<0.001	***		54.18
Monday	0.0427	<0.001	***	1.04	56.54
Tuesday	–0.0141	0.011	**	0.99	53.42
Wednesday	–0.0024	0.662		1.00	54.05
Thursday	0.0251	<0.001	***	1.03	55.56
Friday	0.1558	<0.001	***	1.17	63.31
Satruday	–0.0246	<0.001	***	0.98	52.86
Sunday	–0.1823	<0.001	***	0.83	45.15

Analysis for the factor *month* shows significant differences in all sorted groups (see Table 6). In January, February, March, April and May, the number of accidents decreases; in June, July, August, September and October, the number of accidents is above average; and in November and December, the number of accidents corresponds the most to the annual average of the daily number of accidents involving the deployment of the fire rescue service.

Tab. 6: GLM – one-factor – parameter estimates for month

Factor	*Estimate*	*p-value*		*Times*	*Mean*
Intercept	3.9895	<0.001	***		54.03
January	–0.0298	<0.001	***	0.97	52.44
February	–0.1052	<0.001	***	0.90	48.63
March	–0.2326	<0.001	***	0.79	42.82
April	–0.1736	<0.001	***	0.84	45.42
May	–0.0486	<0.001	***	0.95	51.47
June	0.1253	<0.001	***	1.13	61.23
July	0.0714	<0.001	***	1.07	58.03
August	0.1082	<0.001	***	1.11	60.20
September	0.1211	<0.001	***	1.13	60.98
October	0.0877	<0.001	***	1.09	58.98
November	–0.0082	0.277		0.99	53.59
December	0.0843	<0.001	***	1.09	58.78

The parameters estimated according to the one-factor model did not yield any surprising results, they just correspond to the statistical characteristics shown in Tables 1–2.

3.2 Multifactor analyses

A comparison was subsequently made for two-factor additive (without interactions) and multiplicative (with interactions) analysis with the factors *day of the week* and *month* using a generalised linear model (Poisson distribution). The additive two-factor analysis model shows the difference between the averages for each day of the week and month and the reference value (unweighted overall average), assuming the influence of both factors on the daily number of traffic accidents. Based on the p-values, the model appears statistically significant; the day of the week and the month significantly affect the value of the daily number of accidents involving the deployment of the fire rescue service. The differences are shown in Table 7, where the first column contains the groups sorted by factor, and the second column represents the calculated estimates, i.e. the difference of the logarithms of the estimated means. The estimated logarithmic value is accompanied by the p-value and statistical significance marker in the adjacent columns. The “Times” column is a multiple of the overall average, with the last column showing the daily average number of traffic accidents involving the deployment of the fire rescue service.

Tab. 7: Generalised linear model – parameter estimates (the model without interactions)

Factor	*Estimate*	*p-value*		*Times*
Intercept	3.9852	<0.001	***	
day1	0.0427	<0.001	***	1.04
day2	–0.0144	0.009	**	0.99
day3	–0.0026	0.637		1.00
day4	0.0253	<0.001	***	1.03
day5	0.1562	<0.001	***	1.17
day6	–0.0248	<0.001	***	0.98
day7	–0.1823	<0.001	***	0.83
month1	–0.0303	<0.001	***	0.97
month2	–0.1052	<0.001	***	0.90
month3	–0.2325	<0.001	***	0.79
month4	–0.1733	<0.001	***	0.84
month5	–0.0490	<0.001	***	0.95
month6	0.1255	<0.001	***	1.13
month7	0.0713	<0.001	***	1.07
month8	0.1076	<0.001	***	1.11
month9	0.1223	<0.001	***	1.13
month10	0.0872	<0.001	***	1.09
month11	–0.0088	0.242		0.99
month12	0.0853	<0.001	***	1.09

Table 8 displays the results of the analysis of variance for the model without interactions, indicating significant statistical values for both factors.

Tab. 8: ANOVA without interactions

	DF	*Deviance*	*Resid. DF*	*Resid. Dev*	*p-value*	
null			3652	21211		
day	6	1685.7	3646	19526	<0.001	***
month	11	2515.5	3635	17010	<0.001	***

Predicted values of daily number of traffic accidents for the model without interactions are summarized in Table 9.

Tab. 9: Predicted values of the daily number of accidents (model without interactions)

Month	*Mon.*	*Tues.*	*Wed.*	*Thur.*	*Fri.*	*Sat.*	*Sun.*
1	54.47	51.44	52.06	53.53	61.01	50.91	43.49
2	50.53	47.73	48.30	49.66	56.61	47.24	40.35
3	44.50	42.03	42.53	43.73	49.84	41.59	35.53
4	47.21	44.59	45.12	46.40	52.88	44.13	37.70
5	53.46	50.49	51.09	52.54	59.88	49.97	42.69
6	63.65	60.12	60.83	62.55	71.30	59.49	50.83
7	60.29	56.94	57.62	59.25	67.54	56.36	48.14
8	62.52	59.05	59.75	61.44	70.03	58.44	49.92
9	63.44	59.92	60.64	62.35	71.07	59.30	50.66
10	61.26	57.86	58.54	60.20	68.62	57.26	48.91
11	55.65	52.56	53.18	54.69	62.34	52.02	44.44
12	61.14	57.75	58.43	60.09	68.49	57.15	48.82

In the case of the multiplicative model. the interaction of the factors *day of the week* and *month* are also taken into account. Row d1 in Table 10, for example, indicates the result in the case that the average of the daily number of accidents on the first day is increased compared to the overall average, i.e. the Monday average. Row m1 represents the daily average for January. The line of interactions d1:m1 reveals that the logarithmic value of the difference from the overall mean is 0.025 greater on Mondays in January. If this interaction is included in the overall average, the change in d1 compared to the overall average, and the change in m1 compared to the overall average, it appears that Mondays in January represent a risk of 55.44 accidents per day across the Czech Republic on average. It may seem odd that although Tuesdays are below average and January is below average, the average for Tuesdays in January is above average. However, it should be noted that there are outliers present in the dataset that can cause this and that the Tuesdays of other months equalise the average. Likewise, other days in January equalise the average of January.

Tab. 10: Generalised linear model – parameter estimates (the model with interactions)

Factor	*Estimate*	*p-value*		*Factor*	*Estimate*	*p-value*	
	3.9835	<0.001	***	*d5:m5*	-0.0278	0.113	
d1	0.0442	<0.001	***	*d6:m5*	-0.0184	0.333	
d2	-0.0092	0.099	.	*d7:m5*	0.0268	0.181	
d3	0.0007	0.898		*d1:m6*	-0.0012	0.943	
d4	0.0269	<0.001	***	*d2:m6*	-0.0257	0.146	
d5	0.1579	<0.001	***	*d3:m6*	-0.0828	<0.001	***
d6	-0.0303	<0.001	***	*d4:m6*	-0.0265	0.132	
d7	-0.1903	<0.001	***	*d5:m6*	0.0248	0.126	
m1	-0.0375	<0.001	***	*d6:m6*	0.0543	0.002	**
m2	-0.1102	<0.001	***	*d7:m6*	0.0571	0.002	**
m3	-0.2327	<0.001	***	*d1:m7*	-0.0279	0.108	
m4	-0.1748	<0.001	***	*d2:m7*	-0.0641	<0.001	***
m5	-0.0467	<0.001	***	*d3:m7*	-0.0077	0.661	
m6	0.1265	<0.001	***	*d4:m7*	0.0104	0.550	
m7	0.0751	<0.001	***	*d5:m7*	-0.0365	0.029	*
m8	0.1116	<0.001	***	*d6:m7*	0.0583	0.001	***
m9	0.1221	<0.001	***	*d7:m7*	0.0676	<0.001	***
m10	0.0909	<0.001	***	*d1:m8*	-0.0341	0.048	*
m11	-0.0084	0.266		*d2:m8*	-0.0413	0.021	*
m12	0.0841	<0.001	***	*d3:m8*	-0.0237	0.175	
d1:m1	0.0250	0.170		*d4:m8*	-0.0589	<0.001	***
d2:m1	0.1101	<0.001	***	*d5:m8*	-0.0238	0.144	
d3:m1	0.0686	<0.001	***	*d6:m8*	0.0544	0.002	**
d4:m1	0.0019	0.916		*d7:m8*	0.1274	<0.001	***
d5:m1	0.0253	0.139		*d1:m9*	0.0216	0.203	
d6:m1	-0.0781	<0.001	***	*d2:m9*	-0.1070	<0.001	***
d7:m1	-0.1529	<0.001	***	*d3:m9*	-0.0544	0.002	**
d1:m2	-0.0192	0.328		*d4:m9*	-0.0660	<0.001	***
d2:m2	0.0489	0.013	*	*d5:m9*	0.0637	<0.001	***
d3:m2	0.1652	<0.001	***	*d6:m9*	0.0442	0.011	*
d4:m2	-0.0511	0.012	*	*d7:m9*	0.0978	<0.001	***
d5:m2	0.0357	0.054	.	*d1:m10*	-0.0493	0.005	**
d6:m2	-0.1088	<0.001	***	*d2:m10*	-0.0245	0.163	
d7:m2	-0.0707	0.002	**	*d3:m10*	0.0129	0.456	
d1:m3	-0.0124	0.536		*d4:m10*	0.0092	0.592	
d2:m3	0.1003	<0.001	***	*d5:m10*	-0.0451	0.007	**
d3:m3	0.0070	0.733		*d6:m10*	0.0391	0.025	*
d4:m3	-0.0044	0.825		*d7:m10*	0.0577	0.002	**
d5:m3	-0.0052	0.785		*d1:m11*	0.0507	0.005	**
d6:m3	-0.0372	0.072	.	*d2:m11*	0.0017	0.927	
d7:m3	-0.0480	0.031	*	*d3:m11*	-0.0236	0.210	
d1:m4	0.0044	0.820		*d4:m11*	0.0362	0.046	*
d2:m4	0.0261	0.190		*d5:m11*	-0.0311	0.074	.
d3:m4	-0.0080	0.692		*d6:m11*	0.0033	0.860	
d4:m4	0.0628	0.001	**	*d7:m11*	-0.0372	0.068	.
d5:m4	0.0113	0.549		*d1:m12*	0.0245	0.148	
d6:m4	-0.0368	0.076	.	*d2:m12*	-0.0260	0.141	
d7:m4	-0.0600	0.007	**	*d3:m12*	-0.0217	0.220	
d1:m5	0.0178	0.332		*d4:m12*	0.0547	0.001	**
d2:m5	0.0015	0.935		*d5:m12*	0.0085	0.606	
d3:m5	-0.0317	0.090	.	*d6:m12*	0.0258	0.144	
d4:m5	0.0318	0.078	.	*d7:m12*	-0.0658	0.001	***

Table 11 displays the results of the analysis of variance for the model with interactions. The values indicate significant statistical results for both factors and the interactions. Thus, for the final description of the number of accidents, we use a model with interactions.

Tab. 11: ANOVA with interactions

	DF	*Deviance*	*Resid. DF*	*Resid. Dev.*	*p-value*	
null			3652	21211		
day	6	1685.7	3646	19526	<0.001	***
month	11	2515.5	3635	17010	<0.001	***
day:month	66	539.1	3569	16471	<0.001	***

Predicted values of daily number of traffic accidents for the multifactor model with interactions are summarized in Table 12.

Tab. 12: Predicted values of the daily number of accidents (model with interactions)

Month	*Mon.*	*Tues.*	*Wed.*	*Thur.*	*Fri.*	*Sat.*	*Sun.*
1	55.44	57.22	55.44	53.24	62.13	46.42	36.70
2	49.32	50.05	56.78	46.95	58.38	41.85	37.05
3	43.93	46.61	42.88	43.52	49.58	39.78	33.53
4	47.34	45.86	44.77	49.33	53.40	42.17	35.12
5	54.53	50.86	49.69	54.36	58.38	48.82	43.52
6	63.63	58.86	56.14	60.98	73.16	62.43	53.35
7	58.84	53.80	57.49	60.09	65.36	59.53	51.20
8	60.66	57.09	58.68	58.16	68.67	61.51	56.39
9	64.81	54.02	57.51	58.36	75.74	61.53	55.32
10	58.52	56.87	59.62	60.98	65.84	59.34	51.51
11	58.56	52.86	52.05	56.72	60.45	51.84	42.42
12	62.58	56.40	57.20	63.39	69.00	58.16	45.22

From the predicted numbers of accidents shown in Table 12, it can be seen that the highest values are reached on Fridays, with the fewest accidents on Sundays. The influence of the month of the year is also significant. The results show that the lowest number of accidents corresponds to the months of March and April, while a higher number of accidents can be expected in the summer and winter months. These results are consistent with the results obtained using univariate analyses of variance. However, the advantage of the two-factor model is that it also describes the effect of the relationship (interaction) between the factors (day of the week and month), so the estimates are more accurate.

4 Discussion

The results obtained by analysis reflect the basic statistical description of the dataset. The obtained averages of the daily number of traffic accidents in days and months correspond to the traffic accident statistics shown in Figures 1–2 in the introduction. Outliers have not yet been removed from or replaced in the dataset. which is also reflected in the resulting averages.

In comparison to other articles on traffic accidents in the Czech Republic. such as Kvet (2022) and Harantová (2019), the fundamental difference lies in the source information included in the datasets and therefore in the focus of the analyses. Traffic accident records from the databases of the Police of the Czech Republic usually include a broad classification of traffic accidents, e.g. in terms of the culpability of the driver, type of vehicle, characteristics of the place of the accident, type of road, type of intersection, etc. The present analysis of the dependence of the occurrence of traffic accidents on time is important for creating traffic flow free of accidents and rush hour traffic jams. It is advisable to supplement the time analysis by location data, e.g. geographical coordinates of the accident site or at least a more precise administrative unit under which the accident falls. Inspiration can be found in Senk (2012) and Bačkalić (2013), which use the geographic information system (GIS) for the purposes of geographic statistics. There is also an opportunity for analyses using multifactor models with both time and location factors.

Abroad, modern analyses of traffic accidents are often carried out using neural networks: Cerna (2019), Cerna (2020), Guyeux (2020) and Zheng (2020). The more traditional method of time series analysis can be found in Mallouhy (2022). However, the probability distribution of a random variable which is not distributed normally complicates matters. An elaboration on the application of time series for random variables with Poisson distribution can be found in Quddus (2008). Available literary sources indicate that the incorporation of advanced statistical methods into the analysis of traffic accidents in the Czech Republic is not common. The application of generalised linear model for count data (Poisson distribution) to traffic accident data in the Czech Republic is an addition in this area and represents a contribution to research on the issue. The aim of further research will be to expand the field of dependent variables and seek and explore other dependencies on various factors, such as weather phenomena mentioned in Guyeux (2020), Gao (2016) and other available records.

5 Conclussion

The objective of this study is to analyse the number of traffic accidents that required the departure of fire rescue service units. No changes in content have been made. The study investigates the relationship between the number of accidents and the days of the week and months of the year.

The analysis of the daily number of traffic accidents involving the deployment of the fire rescue service was carried out using data on traffic accidents that occurred in the Czech Republic in the period 2012–2021. The data was obtained from the database of the Fire Rescue Service and processed. The influence of the factors *day of the week*. and *month* was investigated. The probability distribution of the explanatory random variable was tested. The hypothesis of normality was generally rejected. Good match tests did not rule out Poisson distribution. Subsequent modelling was carried out using Poisson analysis of variance.

Firstly, one-factor analyses were performed for the individual factors separately. Tests revealed that all the factors examined are statistically significant. The factor *day of the week* showed that in terms of accidents. Mondays, Thursdays and Fridays are above average, while on Sundays the risk of a traffic accident involving the deployment of the fire rescue service was below average. The development of the risk of traffic accidents during the year, based on the collected data sample, shows that from November to May the risk is below average, while from June to October the risk is above average, with a total annual average of 54 traffic accidents per day involving the deployment of the fire rescue service throughout the country.

Secondly, a two-factor analysis was performed for the factors *day of the week* and *month* in two forms. namely without interactions and with interactions. These two forms were tested and compared. The conclusion of this article is that the two-factor model with interactions seems to be a suitable model for analysing the dependence of the daily number of accidents involving the deployment of the fire rescue service according to the day of the week and month. From the results obtained, it can be summarised that the highest number of accidents occurs on Fridays and the lowest number of accidents on Sundays. In terms of the season (month), fewer accidents can be expected in March and April, and more accidents in the summer and winter months.

Literature:

1. Bačkalić, S.: *Temporal analysis of the traffic accidents occurrence in province of Vojvodina*. Transport Problems. 2013. 8. 87-93.
2. Cerna Ñahuis, S., Guyeux, C., H. Arcolezi, H., Couturier, R., Royer, G., Lotufo, A.: *Long Short-Term Memory for Predicting Firemen Interventions*. 2019. 1132-1137. 10.1109/CoDIT.2019.882067
3. Cerna Ñahuis, S., Guyeux, C., H. Arcolezi, H., Couturier, R., Royer, G.: *A Comparison of LSTM and XGBoost for Predicting Firemen Interventions*. 2020. 10.1007/978-3-030-45691-7_39.
4. Devore, J. L.: *Probability and Statistics for Engineering and the Sciences*. 8th ed. Boston: Brooks/Cole. 2012.
5. Dobson, A.: *An Introduction to Generalized Linear Models*. London: Chapman & Hall. 2008.
6. Guyeux, C. et al.: *Firemen Prediction by Using Neural Networks: A Real Case Study*. In: Bi. Y., Bhatia. R., Kapoor. S. (eds) Intelligent Systems and Applications. IntelliSys 2019. Advances in Intelligent Systems and Computing. vol 1037. Springer. Cham. 2020. https://doi.org/10.1007/978-3-030-29516-5_42
7. Gao, J., Chen, X., Woodward, A. et al.: *The association between meteorological factors and road traffic injuries: a case analysis from Shantou city. China*. Sci Rep 6. 37300. 2016. https://doi.org/10.1038/srep37300
8. Harantová, V., Kubíková, S., Rumanovský, L.: *Traffic Accident Occurrence. Its Prediction and Causes*. In: Mikulski. J.

(eds) Development of Transport by Telematics. TST 2019. Communications in Computer and Information Science. vol 1049. Springer. Cham. 2019. https://doi.org/10.1007/978-3-030-27547-1_10
9. Kvet, M., Konštiaková, L., Dupák, B., Dunaj, J.: *Police Force - Traffic Accidents*. 31st 2022 Conference of Open Innovations Association (FRUCT)}. Helsinki. Finland. 2022. pp. 138-144. doi: 10.23919/FRUCT54823.2022.9770910.
10. Lord, D., Washington, S., Ivan, J.: *Poisson. Poisson-gamma and zero inflated regression models of motor vehicle crashes: Balancing statistical fit and theory*. Accident; analysis and prevention. 2005. 37. 35-46. 10.1016/j.aap.2004.02.004.
11. Mallouhy, R.E., Guyeux, C., Jaoude, C.A., Makhoul, A.: *Forecasting the Number of Firemen Interventions Using Exponential Smoothing Methods: A~Case Study*. In: Barolli. L., Hussain. F., Enokido. T. (eds) Advanced Information Networking and Applications. AINA 2022. Lecture Notes in Networks and Sysms. vol 449. Springer. Cham. 2022. https://doi.org/10.1007/978-3-030-99584-3_50
12. Njå, Å., Kvaløy, J. T., Njå, O.: *Modelling fire occurrences in heavy goods vehicles in road tunnels*. Fire Safety Journal. Volume 127. 2021. 103508. ISSN 0379-7112. https://doi.org/10.1016/j.firesaf.2021.103508
13. Quddus, M. A.: *Time series count data models: An empirical application to traffic accidents*. Accident Analysis & Prevention. Volume 40. Issue 5. 2008. p. 1732-1741. ISSN 0001-4575. https://doi.org/10.1016/j.aap.2008.06.011.
14. Senk, P., Ambros, J., Pokorný, P., Striegler, R.: *Use of Accident Prediction Models in Identifying Hazardous Road Locations*. Transactions on Transport Sciences. 2012. 5(4). 223-232. doi: 10.2478/v10158-012-0025-0.
15. Zheng, J., Huang, M.: *Traffic Flow Forecast Through Time Series Analysis Based on Deep Learning*. 2020. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2990738.

Primary Paper Section: B

Secondary Paper Section: BB