

APARATURA BADAWCZA I DYDAKTYCZNA

Modelowanie liczby wypadków drogowych z zastosowaniem pakietu LIMDEP

JOANNA KAMIŃSKA

UNIwersytet Przyrodniczy we Wrocławiu, Katedra Matematyki

STRESZCZENIE

Problem modelowania (przewidywania) liczby wypadków drogowych na poszczególnych odcinkach dróg (sekcjach) pojawił się w rozważaniach naukowców około 30 lat temu. Przez ponad ćwierć wieku do przewidywania liczby wypadków drogowych w zależności od różnych czynników ruchu drogowego stosowano bardzo różne modele oraz sposoby estymacji. W pracy opisano procedurę modelowania wypadków drogowych z zastosowaniem pakietu LIMDEP w zakresie regresji Poissona i jego podstawowych modyfikacji.

Accident prediction analysis using LIMDEP package

ABSTRACT

The relationship between traffic accidents and traffic conditions has been the subject of research for about 30 years, mainly in the last 20 years. Researchers have developed many different models, types, conditions and functional forms depending on available data, local conditions and purpose. Nowadays the models based on Poisson distribution or its modifications are most widely used. In the paper there have been presented accident analysis procedure using LIMDEP package.

1. WSTĘP

Pierwotnie zależność między czynnikami, a liczbą wypadków opisywano deterministycznym równaniem regresji liniowej. Jednak, jak wykazali Jovanis i Chang [1] zmienność natężenia ruchu w czasie oraz częste występowanie zerowej liczebności wypadków w sektorze sprawiają, że standardowa regresja liniowa może prowadzić do błędnych estymatorów parametrów, a nawet przewidywanych ujemnych wartości. W odpowiedzi na problem zastosowania regresji liniowej, rozpoczęto badania nad użyciem regresji Poissona [2], której zastosowanie jest możliwe dzięki temu, że dane dotyczące wypadków drogowych stanowią nieujemne całkowite wartości. W kolejnych latach pojawiały się rozszerzenia regresji Poissona, nowe modele oraz kolejne metody estymacji. Wśród najczęściej stosowanych modeli ekonometrycznych w przewidywaniu częstości wypadków drogowych w zależności od geometrii drogi oraz natężenia ruchu na określonych odcinkach, można wymienić [3]: klasyczny model regresji Poissona, model gamma-Poissona (NB – negative binomial), jednowymiarowy model lognormalny Poissona (PLN), wielowymiarowy model lognormalny Poissona (MVPLN), model Conway-Maxwell-Poisson oraz wiele innych.

2. MODELOWANIE

W procesie modelowania ekonometrycznego można wyróżnić następujące etapy: wybór modelu (poprzedzony badaniami zgodności rozkładów empirycznego z modelowanym), wybór zbioru zmiennych objaśniających, wybór metody estymacji, walidacja modelu.

Poniżej pokrótce omówiono poszczególne etapy dla modelowania częstości wypadków drogowych.

a) Wybór modelu

Model regresji Poissona zakłada, że dla niezależnych par obserwacji (n_i, x_i) , $(i=1, \dots, n)$, prawdopodobieństwo wystąpienia n_i wypadków w segmencie i wynosi $P(n_i)$, gdzie

$$P(n_i) = \frac{e^{-\lambda_i} \lambda_i^{n_i}}{n_i!} \quad (1)$$

λ_i jest parametrem Poissona równym wartości oczekiwanej liczby wypadków $E(n_i)$ w segmencie i . Regresja Poissona zakłada, że logarytm z parametru λ_i jest funkcją liniową zmiennych objaśniających.

$\lambda_i = \exp\left(\sum_j \beta_j X_j\right)$ Dla tak zdefiniowanego rozkładu zachodzi $\text{var}(Y_i) = \lambda_i = E(Y_i)$, co oznacza, że wariancja rozkładu powinna być równa jego wartości oczekiwanej. Dla danych empirycznych ten warunek zazwyczaj nie jest spełniony. Aby uniknąć tego ograniczenia wprowadzono rozszerzenie do funkcji definiującej parametr Poissona, w postaci dodania składnika losowego ε_i o zadanym rozkładzie. Zatem

$$\lambda_i = \exp\left(\sum_j \beta_j X_j + \varepsilon_i\right).$$

Jeśli e^{ε_i} ma rozkład gamma otrzymany model nazywany jest gamma Poissona - NB ([4, 5], jeśli natomiast

e^{ε_i} ma rozkład log-normalny ($\varepsilon_i \sim N(0, \sigma_u^2)$), model nazywany jest lognormalnym-Poissona - PLN [6, 7]. Funkcja gęstości prawdopodobieństwa w modelu NB jest postaci:

$$P(Y_i = n_i | \lambda_i, \alpha) = \frac{\Gamma\left(n_i + \frac{1}{\alpha}\right)}{n_i! \Gamma\left(\frac{1}{\alpha}\right)} \cdot \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}} \cdot \left(\frac{\lambda_i}{\frac{1}{\alpha} + \lambda_i}\right)^{\lambda_i} \cdot \left(\frac{1}{\alpha} + \lambda_i\right)^{-\lambda_i} \quad (2)$$

Parametr α nazywany jest parametrem dyspersji. Dla tego rozkładu wartość oczekiwana nadal równa jest parametrowi Poissona $\lambda_i = E(n_i)$, przy czym wariancja może być od niej różna $\text{var}(n_i) = \lambda_i + \alpha \cdot \lambda_i^2$. Dla modelu PLN wartość oczekiwana jest postaci $E(Y_i) = \lambda_i e^{0.5\sigma_u^2}$, a wariancja

$$\text{var}(Y_i) = E(Y_i) + E^2(Y_i) \left(e^{\sigma_u^2} - 1\right).$$

Rozkład Conway-Maxwell-Poisson (COM-Poisson) [8,9] jest uogólnieniem rozkładu Poissona z funkcją gęstości postaci

$$P(Y_i = n_i) = \frac{1}{Z(\lambda, \nu)} \cdot \frac{\lambda^{n_i}}{(n_i)!^\nu},$$

$$\text{gdzie } Z(\lambda, \nu) = \sum_{n=0}^{\infty} \frac{\lambda^n}{(n!)^\nu}.$$

λ jest również parametrem zbliżonym do wartości średniej, ν jest określany mianem parametru kształtu. Wyżej przedstawione modele zawierają jedną zmienną zależną opisującą ilość wypadków o zadanym nasileniu: wypadki śmiertelne, z bardzo poważnie rannymi, poważnie rannymi, lekko rannymi,

bez rannych – jedynie zniszczenie mienia. Możliwe jest jednoczesne modelowanie liczby wypadków dla wszystkich stopni nasilenia. Najczęściej stosowaną do tego celu metodą jest multivariate Poisson-lognormal regression MVPLN [10].

Wybór konkretnego modelu powinien być poprzedzony badaniem zgodności rozkładów empirycznych z modelowym. Po wybraniu modelu najbardziej odpowiadającego rozkładowi empirycznemu należy określić zbiór zmiennych objaśniających.

b) Zmienne objaśniające

Opisane wyżej modyfikacje regresji Poissona niezmiennie zakładają, że wartość oczekiwana jest funkcją eksponentialną względem zmiennych objaśniających. Najprostszy zbiór składa się z dostępnych danych empirycznych obejmujących najczęściej: średnie dzienne natężenie ruchu (annual average daily traffic – AADT) oraz charakterystyki drogi w obrębie sekcji. Wśród zmiennych opisujących rozważany segment drogi najczęściej występują: długość segmentu, minimalna liczba pasów ruchu, liczba skrzyżowań, obowiązujące ograniczenie prędkości, liczba mostów, liczba zakrętów na kilometr, kąt płaski zakrętu, charakterystyki nawierzchni oraz inne [11]. Lepsze dopasowanie modelu można otrzymać stosując dodatkowe zmienne. Najbardziej popularne jest wprowadzanie do zbioru zmiennych objaśniających logarytmu naturalnego z AADT. Rzadziej można spotkać wśród zmiennych niezależnych logarytm naturalny z liczby pasów na jezdni plus 1 [12] lub ograniczenie prędkości w kwadracie [13].

c) Metoda estymacji parametrów modelu

Najpowszechniej znana oraz stosowana ze względu na łatwość oraz szybkość obliczeń jest metoda maksimum prawdopodobieństwa (maksimum likelihood method). Maksimum funkcji znajduje się zazwyczaj metodą Newtona. Taki sposób estymacji stosują popularne pakiety ekonometryczne SPSS, LIMDEP. Obecnie coraz szersze zastosowanie znajduje metoda Bayesa z próbkowaniem metodą MCMC (Monte Carlo Markow Chain), którą stosuje m.in. software WinBUGS. Metoda Bayesa [14] umożliwia estymowanie parametrów bardzo złożonych modeli, które nie posiadają łatwo obliczalnej funkcji maksimum prawdopodobieństwa.

Spośród wymienionych powyżej najszerzej dostępny oraz najprostszy w obsłudze jest pakiet LIMDEP. Stosowanie tego narzędzia znaczenie podwyższa stopień zaawansowania kształcenia z zakresu modelowania matematycznego szczególnie w obszarze bezpieczeństwa drogowego.

d) weryfikacja modelu

Aby ocenić dopasowanie modelu do danych empirycznych stosuje się wiele różnych metod. Najczęściej do porównania dopasowania różnych modeli stosuje się miary dopasowania w postaci wskaźników. Do oceny stopnia dopasowania modelu nie wystarczy jeden wskaźnik [9]. Do poniższej analizy zastosowano 4 miary dopasowania opisane poniżej:

- Akaike's Information Criterion (AIC)

$$AIC = \frac{-2 \ln L(M_j) - 2k}{N} \quad (3)$$

gdzie $\ln L(M_j)$ jest wartością logarytmu funkcji prawdopodobieństwa (log-likelihood value) modelu j , k jest liczbą parametrów oraz N – liczbą obserwacji (tutaj $N = 1671$).

- Bayesian Information Criteria (BIC)

$$BIC = D(M_j) + k \cdot \ln N = -2 \ln L(M_j) + k \cdot \ln N \quad (4)$$

gdzie $D(M_j)$ jest dewiacją modelu M_j .

- Mean absolute deviation (MAD)

$$MAD = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (5)$$

gdzie \hat{y}_i jest przewidywaną liczbą wypadków w segmencie i , y_i jest obserwowaną liczbą wypadków w segmencie i .

- Mean squared prediction error (MSPE)

$$MSPE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (6)$$

Model jest lepiej dopasowany kiedy wartości wskaźników są mniejsze.

Inną metodą oceny stopnia dopasowania modelu jest zaproponowana przez Hauer i Bamfo [15] metoda CURE. Ta graficzna metoda jest coraz częściej stosowana w analizie bezpieczeństwa [9, 16, 17] innych obszarach nauki [18] do porównywania modeli niezależnie od liczby obserwacji. Ideą metody CURE jest wykres skumulowanych residuów dla jednej ze zmiennych objaśniających. Residua (e_i) przedstawiają różnicę między wartością obserwowaną (y_i) oraz przewidywaną (\hat{y}_i). Im bliżej residua oscylują wokół wartości zero tym lepsze jest dopasowanie modelu do danych.

2. DANE EMPIRYCZNE

Dane wykorzystane w niniejszej pracy, udostępnione autorowi dzięki uprzejmości Rune Elvik The Institute of Transport Economics in Oslo, zawierają informacje o 1671 sekcjach obejmujących odcinki dróg

krajowych na terenie jednostki administracyjnej Hordaland w Norwegii. Dane dotyczące liczby wypadków drogowych w sektorach pochodzą z okresu 1993-2000. Analiza została wykonana dla całkowitej liczby odnotowanych wypadków. Dane obejmują 10 zmiennych (AADT występuje dwukrotnie: wprost jako zmienna AADT dzielona przez 1000 oraz jako logarytm naturalny z AADT) opisujących charakterystyki geometryczne drogi, natężenie ruchu i dodatkowe informacje. W rozważanych segmentach odnotowano w badanym okresie 3175 wypadków, 113 (3,56%) z nich było śmiertelnych, w 790 (47%) segmentach nie odnotowano ani jednego wypadku w okresie 8 lat. Tabela 1 przedstawia podstawowe charakterystyki statystyczne rozważanych zmiennych.

Uwzględniając dostępne dane, po przetestowaniu wielu różnych kombinacji na podstawie wcześniejszych prac: [9, 12, 16, 17, 19-23] najbardziej odpowiednia do opisu częstości wypadków okazała się funkcja:

$$\lambda_i = \beta_8' AADT \cdot \beta_9' X_7 \cdot \exp\left(\beta_7 \frac{AADT}{1000} + \sum_{j=1}^6 \beta_j X_{ji} + \beta_0 + \varepsilon_i\right) \quad (7)$$

gdzie $\beta_i' = \exp(\beta_i)$ dla $i=8,9$.

Parametry funkcji (7) zostały oszacowane z zastosowaniem metody maksimum prawdopodobieństwa z dyskretnym algorytmem Newtona [24]. W niniejszej pracy proces szacowania wykonano z zastosowaniem pakietu ekonometrycznego LIMDEP. Walidacja modelu została przeprowadzona z zastosowaniem metody CURE [15].

Tabela 1. Charakterystyki statystyczne zmiennych

Nazwa zmiennej	Symbol	Średnia	Odchylenie stand.	Min.	Max.
Zmienna zależna					
Liczba wypadków		1,90	4,44	0	55
Zmienne niezależne					
Średnie dzienne natężenie ruchu /1000 (AADT/1000)	AADT/1000	2,96	4,80	0,04	53,15
Minimalna liczba psów ruchu	X_1	2,01	0,21	1	4
Droga krajowa (1 jest TAK, 0 jeśli NIE)	X_2	0,26	0,44	0	1
Ograniczenie prędkości 50 km/h (1=tak, 0=p.p.)	X_3	0,15	0,35	0	1
Ograniczenie prędkości 60 km/h (1=tak, 0=p.p.)	X_4	0,15	0,36	0	1
Ograniczenie prędkości 70 km/h (1=tak, 0=p.p.)	X_5	0,02	0,13	0	1
Długość segmentu w iloczynie z czasem pomiaru w latach (km/rok)	X_6	7,51	1,08	2,7	8
ln AADT	Ln(AADT)	7,29	1,15	3,74	10,88
Logarytm naturalny z liczby skrzyżowań + 1	Ln(X_7)	0,27	0,50	0	2,71
Liczba obserwacji		1671			

3. METODYKA

W celu dokonania wyboru modelu przeprowadzono testy chi kwadrat (χ^2) dla: klasycznego modelu Poissona (Poisson), modelu ujemnego dwumianowego - negative binomial model (NB) oraz lognormalnego modelu Poissona (PLN). Wyniki przedstawiono w Tabeli 2.

Na podstawie dokonanej analizy jako najlepiej oddający rozkład liczby wypadków wybrano model lognormalny Poissona.

Kolejnym etapem modelowania był wybór zbioru zmiennych objaśniających, a co za tym idzie, postaci

$$\text{analitycznej funkcji } \lambda_i = \exp\left(\sum_j \beta_j X_{ji} + \varepsilon_i\right).$$

4. WYNIKI

Wyniki estymacji parametrów modelu przedstawiono w Tabeli 3.

Na podstawie otrzymanych wartości współczynników oraz błędów standardowych estymacji można stwierdzić, że dla badanego obszaru zarówno liczba pasów ruchu jak i fakt, że sektor znajduje się na drodze krajowej nie mają statystycznie istotnego wpływu na liczbę wypadków w tym sektorze. Ze względu na dość złożoną postać funkcji prawdopodobieństwa (1) wystąpienia zadanej liczby wypadków oraz iloczynowo-wykładniczą postać parametru Poissona (7) nie jest możliwa interpretacja ilościowa wartości współczynników otrzymywanych w procesie

Tabela 2. Rozkład liczby sektorów dla liczby odnotowanych wypadków w trzech wariantach modeli

Liczba wypadków	Liczba sektorów			
	data	Poisson	NB	PLN
0	790	250	835	777
1	373	475	272	364
2	175	451	157	183
3	114	286	104	103
4	57	136	73	63
5	36	52	53	40
6	27	16	40	27
7	13	4	30	21
8	10	1	23	17
9	9	0	18	14
10	4	0	14	11
11	9	0	11	9
12	6	0	8	6
>12	48	0	32	31
χ^2		8764,3	233,7	32,5

modelowania. Możliwa jest jedynie ocena jakościowa wpływu poszczególnych czynników na liczbę wypadków. Na podstawie Tabeli 3 wyraźnie widoczny jest wpływ ograniczenia prędkości na liczbę wypadków. Największa ich liczba ma miejsce na terenie z obowiązującym ograniczeniem prędkości do 50 km/h. Wraz ze wzrostem prędkości w ograniczeniu maleje jego wpływ na liczbę wypadków. W każdym jednak przy-

Tabela 3. Wartości oszacowanych parametrów wraz z błędem oszacowania oraz miary dopasowania dla liczby wypadków

Nazwa zmiennej	współczynnik	Błąd stand.
Stała	-7,6930	0,2436
AA DT/1000	0,0083	0,0038
Minimalna liczba psów ruchu	0,0207	0,0315
Droga krajowa (1 jest TAK, 0 jeśli NIE)	0,0099	0,0283
Ograniczenie prędkości 50 km/h (1=tak, 0=p.p.)	0,9411	0,0353
Ograniczenie prędkości 60 km/h (1=tak, 0=p.p.)	0,4215	0,0419
Ograniczenie prędkości 70 km/h (1=tak, 0=p.p.)	0,3069	0,0613
Długość segmentu w iloczynie z czasem pomiaru	0,1221	0,0115
ln AA DT	0,8640	0,0289
Logarytm naturalny z liczby skrzyżowań + 1	0,2005	0,0178
Sigma	0,1461	0,0154
Miary dopasowania		
AIC	3,0	
BIC	5080,2	
MAD	1,288	
MSPE	6,369	

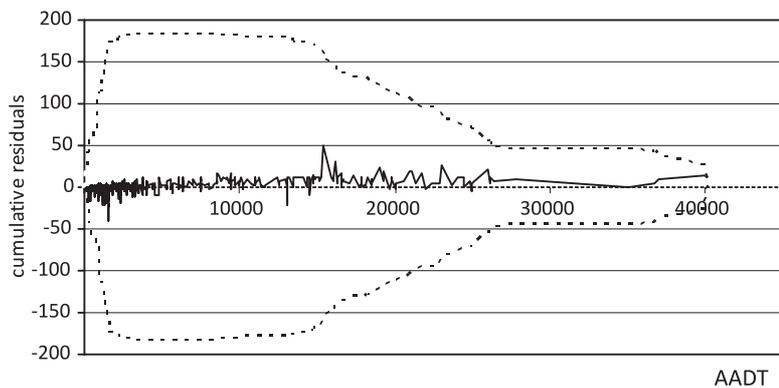
Uwaga: Italicami oznaczono parametry, które statystycznie nie różnią się od zera na poziomie istotności 95%

padku dodatnie wartości współczynnika oznaczają, że są to obszary zwiększonego ryzyka wystąpienia wypadku. Dodatnie są również współczynniki związane z natężeniem ruchu. Zwiększenie natężenia ruchu powoduje zwiększenie liczby wypadków. Podobnie liczba skrzyżowań jest czynnikiem, którego wzrost powoduje wzrost liczby wypadków. Długość odcinka oraz czas przez jaki zliczane są wypadki w sektorze w sposób oczywisty jest dodatnio skorelowana z ich liczbą.

Analiza dopasowania modelu zarówno na podstawie miar dopasowania (Tab. 3) jak i wykresu CURE pozwala pozytywnie ocenić model. Średnie odchylenie wartości szacowanych od empirycznych MAD=1,288 oznacza, że przeciętnie liczba wypadków otrzymana z modelu różni się od rzeczywistej o 1. Jest to dopuszczalna wartość. Wykres CURE (Rys. 1) wyraźnie pokazuje, że residua mieszczą się w całej rozpiętości natężenia Ruchu (AADT) w obszarze wyznaczonym przez dwukrotność odchylenia standardowego ([15]. Wartości residua oscylują wokół linii zerowej nie osiągając wartości większych niż 50. Można zatem stwierdzić, że model poprawnie opisuje zależność między liczbą wypadków, a czynnikami na nie wpływającymi.

5. PODSUMOWANIE I WNIOSKI

Modelowanie liczby wypadków drogowych jest bardzo ważną częścią badań nad bezpieczeństwem ogólnie pojętego transportu. Zastosowanie pakietu LIMDEP pozwala w sposób prosty estymować parametry różnych modeli w szczególności opartych na regresji Poissona wraz z modyfikacjami. W pracy, na podstawie danych pochodzących z Norwegii, wykazano, że do modelowania częstości wypadków drogowych model lognormalny Poissona jest zdecydowanie bardziej adekwatny niż podstawowy model Poissona oraz jego modyfikacja ujemna dwumianowa (NB). Analizy wpływu poszczególnych czynników środowiska (charakterystyka drogi, pogoda, oświetlenie) na ryzyko wystąpienia wypadku



Rysunek 1. Wykres kumulowanych residua CURE dla zmiennej objaśniającej AADT-średnie dzienne natężenie ruchu

w wielu krajach (np. skandynawskich) poprzedzają wprowadzenie zmian w regulacjach prawnych, czy projektowanie nowych połączeń drogowych. W Polsce badania takie są prowadzone w bardzo wąskim zakresie głównie z powodu braku odpowiedniej bazy danych.

LITERATURA

- [1] Jovanis, P.P., Chang, H.L., 1996. Modeling the relationship of accidents to Miles traveled. *Transportation Research Record* 1068, 42-51.
- [2] Joshua, S.C., Gerber, N.J., 1990. Estimating truck accident rate and involvements using linear and Poisson regression models. *Transportation Planning and Technology* 15(1), 41-58.
- [3] Lord, D., Mannering F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A* 44, 291-305.
- [4] Maycock, G., Hall, R.D., 1984. Accident at 4-Arm Roundabouts. TRRL Laboratory Report 1120, Transportation and Road Research Laboratory, Crowthorne, GB.
- [5] Hauer, E., Ng, J.C.N., Lovell, J., 1988. Estimation of safety at signalized intersections. *Transportation Research Record* 1185, 1-10.
- [6] Miaou, S.-P., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space-time modeling approach. *Journal of Transportation and Statistics* 6(1), 33-57.
- [7] Miaou S.-P., Blich, R.P., Lord, D., 2005. Developing median barrier installation guidelines: a benefit/cost analysis using Texas data. *Transportation Research Record* 1984, 3-19.
- [8] Conway, R.W., Maxwell, W.L., 1962. A queuing model with state dependent service rates. *Journal of Industrial Engineering* 12, 132-136.
- [9] Lord, D., Park, P.Y.-J., 2008. Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. *Accident Analysis and Prevention* 40, 1441-1457.
- [10] El-Basouyouny, K., Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. *Accident Analysis and Prevention* 41, 820-828.
- [11] Anastasopoulos, P.Ch., Tarko, A.P., Mannering, F.L., 2008. Tobit analysis of vehicle accident rates on interstate highways. *Accident Analysis and Prevention* 40, 768-775.
- [12] Elvik, R., 2008. The predictive validity of empirical Bayes estimates of road safety. *Accident Analysis and Prevention* 40, 1964-1969.
- [13] Ma, J., Kochelman, K.M., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis and Prevention* 40, 964-975.
- [14] Gilks, W.R., Richardson, S., Spiegelhalter, G.J., 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, UK.
- [15] Hauer, E., Bamfo, J., 1997. Two tools for finding what function links the dependent variable to the explanatory variables. In: *Proceedings of the ICTCT 1997 Conference*, Lund, Sweden.
- [16] Lord, D., Persaud, B.N., 2000. Accident prediction models with and without trend: application of the Generalized Estimating Equations (GEE) procedure. *Transportation Research Record* 1717, 102-108.
- [17] Wang, X., Abdel-Aty, M., 2007. Right-Angle Crash Occurrence at Signalized Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2019, 156-168.
- [18] Lin, D.Y., Wei, L.J., Ying, Z., 2002. Model-Checking Techniques Based on Cumulative Residuals. *Biometrics* 58, 1-12.
- [19] Karlaftis, M.G., Tarko, A.P., 1997. Heterogeneity considerations in accident modeling. *Accident Analysis and Prevention* 30, 425-433.
- [20] Abdel-Aty, M.A., Radwan A.E., 2000. Modeling traffic accident occurrence and involvement. *Accident Analysis and Prevention* 32, 633-642.

- [21] Anastasopoulos, P.Ch., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis and Prevention* 41, 153-159.
- [22] El-Basouyouny, K., Sayed, T., 2009a. Collision prediction models using multivariate Poisson-lognormal regression. *Accident Analysis and Prevention* 41, 820-828.
- [23] El-Basouyouny K., Sayed T., 2009b. Accident prediction models with random corridor parameters. *Accident Analysis and Prevention* 41, 1118-1123.
- [24] Winkelmann, R., 2008. *Econometric Analysis of Count Data*. Fifth edition. Springer, ISBN 978-3-540-77648-2.
- [25] Fridstrom, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., Thomsen, L.K., 1995. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis and Prevention* 27, 1-20.
- [26] Geibe, P., 2003. Accident Prediction models for urban roads. *Accident Analysis and Prevention* 35, 273-285.
- [27] Lord, D., Guikema, S., Geedipally, S.R., 2008. Application of the Conway-Maxwell-Poisson generalized linear model of analyzing motor vehicle crashes. *Accident Analysis and Prevention* 40(3), 1123-1134.