



Wstępna analiza korelacji danych dotyczących wypadków drogowych w Polsce

SŁAWOMIR STĘPIEŃ

Wojskowa Akademia Techniczna, Wydział Mechatroniki, Uzbrojenia i Lotnictwa,
ul. gen. S. Kaliskiego 2, 00-908 Warszawa, slawomir.stepien@wat.edu.pl

Streszczenie. W artykule przedstawiono wstępną analizę statystyczną danych dotyczących wypadków drogowych w Polsce, w tym liczby: wypadków, ofiar śmiertelnych wypadków, rannych w wypadkach oraz zatrzymanych nietrzeźwych kierowców. Analiza dotyczy danych dziennych z lat 2008-2021. Wyznaczono wartości średnie, mediany oraz przeprowadzono analizę korelacji, bazując na współczynnikach korelacji zupełnej, częściowej oraz wielorakiej. Wyniki przedstawiono dla całości danych oraz oddzielnie dla danych miesięcznych i rocznych.

Słowa kluczowe: inżynieria mechaniczna, analiza korelacji, bezpieczeństwo ruchu drogowego, wypadki drogowe

DOI: 10.5604/01.3001.0015.3839

1. Wprowadzenie

Bezpieczeństwo ruchu drogowego jest istotnym zagadnieniem, a szczególnie tragiczne wypadki wywołują powszechne zainteresowanie społeczne. Pojęcie „bezpieczeństwo” jest szerokie i może być analizowane z różnych punktów widzenia. Pojawiające się artykuły podejmują różnorodne zagadnienia. Ciekawą analizę przedstawiono w pracy [1], odnosząc się do Narodowego Programu Bezpieczeństwa Ruchu Drogowego i spełnienia wymagań Organizacji Narodów Zjednoczonych i Komisji Europejskiej w tym zakresie. Z kolei w pracy [2] podjęto problem wpływu zmian infrastruktury drogowej na bezpieczeństwo ruchu drogowego. Podobnie w pracy [3] omówiono tę problematykę w regionie świętokrzyskim. Artykuł [4] prezentuje wyniki badań w zakresie zarządzania czasem pracy kierowców zawodowych w odniesieniu

do bezpieczeństwa. Przykładem metod analizy bezpieczeństwa ruchu drogowego jest praca [5], w której zaprezentowano wskaźniki bezpieczeństwa w odniesieniu do przyjętego przedziału czasu.

Kolejnym przykładem, tym razem zastosowania metod probabilistycznych w analizach bezpieczeństwa ruchu drogowego, jest praca [6]. Autorzy podjęli temat zależności liczby wypadków od pory dnia oraz natężenia ruchu. Zastosowano wybrane modele matematyczne: Armax (model autoregresywny ze średnią ruchomą i zewnętrznym wejściem) oraz model Holta-Wintersa. Weryfikację hipotez statystycznych przeprowadzono z wykorzystaniem testu Kołmogorowa-Smirnowa oraz rangowego testu Kruskala-Wallisa. Stwierdzono, że pora dnia miała znaczny wpływ na występowanie wypadków drogowych, przy istnieniu również innych czynników. Potwierdzono możliwości prognostyczne stosowanych modeli. Przykładem analizy korelacji jest praca [7]. Autorka skorzystała z możliwości programu Statistica oraz danych zawartych w raporcie *Stan bezpieczeństwa ruchu drogowego oraz działania realizowane w tym zakresie w 2014 r.* Krajowej Rady Bezpieczeństwa Ruchu Drogowego. W analizie korelacji uwzględniono następujące dane dotyczące województwa: liczbę wypadków, liczbę ludności, długość dróg publicznych, liczbę zarejestrowanych samochodów oraz gęstość zaludnienia. Bazowano głównie na współczynniku korelacji zupełnej. Innym przykładem zastosowania programu Statistica jest artykuł [8]. Zespół autorów objął analizą dane z lat 2009-2013 dotyczące ofiar wypadków drogowych, ze szczególnym uwzględnieniem osób nietrzeźwych. W analizie uwzględniono wiek, płeć, wyjściową przyczynę zgonu, wynik badania sekcyjnego oraz stężenie alkoholu etylowego w materiale biologicznym. Za pomocą programu Statistica wykonano analizę podstawowych statystyk opisowych, analizę korelacji ze współczynnikiem korelacji zupełnej, testy t-Studenta, testy Manna-Whitneya, testy chi-kwadrat niezależności oraz jednoczynnikową analizę wariancji. W pracy [9] przedstawiono analizę korelacji bazującej na współczynniku korelacji zupełnej i współczynniku korelacji rang Spearmana oraz analizę regresji danych statystycznych obejmujących czteroletni okres od 2015 do 2018 r. Dane dotyczyły głównie liczby wypadków drogowych, ofiar śmiertelnych oraz rannych i pochodziły z wybranych regionów Federacji Rosyjskiej. Celem była identyfikacja głównych czynników wpływających na wypadkowość.

Ten artykuł również odnosi się do zagadnień analizy korelacji danych wypadków drogowych. Ta wstępna analiza dotyczy okresu od 1.12.2008 do 12.07.2021 r., więc bazuje na aktualnych danych. W ramach analizy korelacji wzięto pod uwagę następujące wielkości: liczbę wypadków drogowych, liczbę zabitych w tych wypadkach, liczbę rannych, liczbę zatrzymanych nietrzeźwych kierowców. Dane te pozyskano ze strony internetowej Policji [10]. Publikowane tam informacje pochodzą z raportów dziennych (od grudnia 2008 r.), przez co przytoczone liczby zabitych dotyczą bezpośrednich ofiar wypadków i nie obejmują osób zmarłych, w wyniku odniesionych

obrażeń, w późniejszym czasie. Takie ujęcie danych dobrze obrazuje bezpośrednie skutki wypadków. Oczywiście zdarzają się różnice w danych z raportów dziennych i raportów miesięcznych oraz rocznych — mogą wynikać z uwzględnienia w miesięcznych i rocznych raportach wspomnianych późniejszych ofiar oraz korekty ewentualnych błędów danych dziennych. Jednak błędy te nie powinny zaburzać w sposób istotny otrzymanych wyników. Wspomniane dane traktowane są jako zmienne losowe.

Warto zauważyć, że daneienne obejmują wszystkich uczestników ruchu drogowego, bez rozróżnienia miejsca, pory doby, uczestników (kierowca, pieszy, pasażer itp.) czy też pojazdu (motocykl, samochód osobowy, ciężarowy itp.). Takie informacje pojawiają się w raportach miesięcznych i rocznych, w których publikowane są również wartości stosowanych względnych wskaźników bezpieczeństwa. Niecelowe wydaje się powielanie publikowanych tam treści. Powiązanie danych, którymi dysponujemy, z innymi informacjami, np. liczbą zarejestrowanych pojazdów, jest również utrudnione, ze względu na dostęp do danych dziennych i błędów w bazie Centralnej Ewidencji Pojazdów i Kierowców CEPiK. Analiza korelacji obejmie wyznaczenie wartości współczynników korelacji zupełnej, częściowej i wielorakiej.

Celem publikacji jest wyznaczenie wartości współczynników korelacji przyjętych zmiennych losowych, bazując na w miarę aktualnych danych, a także określenie, czy występuje tendencja zmian tych wartości w czasie.

2. Aparat matematyczny

Na początku należy określić zmienne losowe: x — wypadki drogowe, y — zabici w wypadkach drogowych, z — ranni w wypadkach drogowych, w — zatrzymani nietrzeźwi kierujący. Przytoczone dalej zależności są powszechnie stosowane, dlatego pominięto powołania na źródło. Same obliczenia, z uwagi na dużą ilość danych i operacje na macierzach, przeprowadzono na bazie opracowanego samodzielnie, przeznaczonego do tego programu.

Oszacowanie wartości średnich:

$$\begin{aligned} x_{sr} &= \frac{1}{n} \sum_k x_k & y_{sr} &= \frac{1}{n} \sum_k y_k \\ z_{sr} &= \frac{1}{n} \sum_k z_k & w_{sr} &= \frac{1}{n} \sum_k w_k \end{aligned}, \quad (1)$$

gdzie: n — liczba uwzględnionych danych.

Oszacowanie wariancji:

$$\begin{aligned} S_x^2 &= \frac{1}{n} \sum_k x_k^2 - x_{sr}^2 & S_y^2 &= \frac{1}{n} \sum_k y_k^2 - y_{sr}^2 \\ S_z^2 &= \frac{1}{n} \sum_k z_k^2 - z_{sr}^2 & S_w^2 &= \frac{1}{n} \sum_k w_k^2 - w_{sr}^2 \end{aligned} \quad (2)$$

Oszacowanie kowariancji:

$$\begin{aligned} S_{xy} &= \frac{1}{n} \sum_k x_k y_k - x_{sr} y_{sr} & S_{xz} &= \frac{1}{n} \sum_k x_k z_k - x_{sr} z_{sr} & S_{xw} &= \frac{1}{n} \sum_k x_k w_k - x_{sr} w_{sr} \\ S_{yz} &= \frac{1}{n} \sum_k y_k z_k - y_{sr} z_{sr} & S_{yw} &= \frac{1}{n} \sum_k y_k w_k - y_{sr} w_{sr} & S_{zw} &= \frac{1}{n} \sum_k z_k w_k - z_{sr} w_{sr} \end{aligned} \quad (3)$$

Oszacowanie współczynników korelacji zupełnej:

$$\begin{aligned} r_{xy} &= \frac{S_{xy}}{\sqrt{S_x^2 S_y^2}} & r_{xz} &= \frac{S_{xz}}{\sqrt{S_x^2 S_z^2}} & r_{xw} &= \frac{S_{xw}}{\sqrt{S_x^2 S_w^2}} \\ r_{yz} &= \frac{S_{yz}}{\sqrt{S_y^2 S_z^2}} & r_{yw} &= \frac{S_{yw}}{\sqrt{S_y^2 S_w^2}} & r_{zw} &= \frac{S_{zw}}{\sqrt{S_z^2 S_w^2}} \end{aligned} \quad (4)$$

Uzyskane współczynniki korelacji można zestawić w macierzy korelacji C:

$$C = \begin{bmatrix} 1 & r_{xy} & r_{xz} & r_{xw} \\ r_{xy} & 1 & r_{yz} & r_{yw} \\ r_{xz} & r_{yz} & 1 & r_{zw} \\ r_{xw} & r_{yw} & r_{zw} & 1 \end{bmatrix} \quad (5)$$

Oszacowanie współczynników korelacji cząstkowej:

$$r_{i,j,1,\dots,(i-1),(i+1),\dots,(j-1),(j+1),\dots,k} = -\frac{C_{i,j}}{\sqrt{C_{i,i} C_{j,j}}}, \quad (6)$$

gdzie: $C_{i,j}$ jest dopełnieniem algebraicznym elementu $r_{i,j}$ macierzy korelacji C.

Oszacowanie współczynnika korelacji wielorakiej:

$$R_{i,1,\dots,(i-1),(i+1),\dots,k} = \sqrt{1 - \frac{|C|}{C_{i,i}}}, \quad (7)$$

gdzie: $|C|$ — wyznacznik macierzy C .

Dodatkowo wyznaczono medianę, wartości minimalne i maksymalne oraz wartości występujące najczęściej.

3. Uzyskane wyniki

Na początku zostaną zaprezentowane wyniki obliczeń podstawowych parametrów rozpatrywanych zmiennych losowych.

3.1. Wartości średnie i mediany rozpatrywanych zmiennych losowych

W tabeli 1 przedstawiono wartości średnie oraz mediany odpowiednich zmiennych dla całości danych, którymi dysponujemy (od 1 grudnia 2008 do 12 lipca 2021). Wartości średnie zostały podane w zaokrągleniu do drugiego miejsca po przecinku.

TABELA 1

Podstawowe parametry dla całości danych od 1.12.2008 do 12.07.2021

	x wypadki	y zabici	z ranni	w nietrzeźwi
Wartość średnia	84,80	7,61	102,51	291,63
Mediana	85	7	101	256

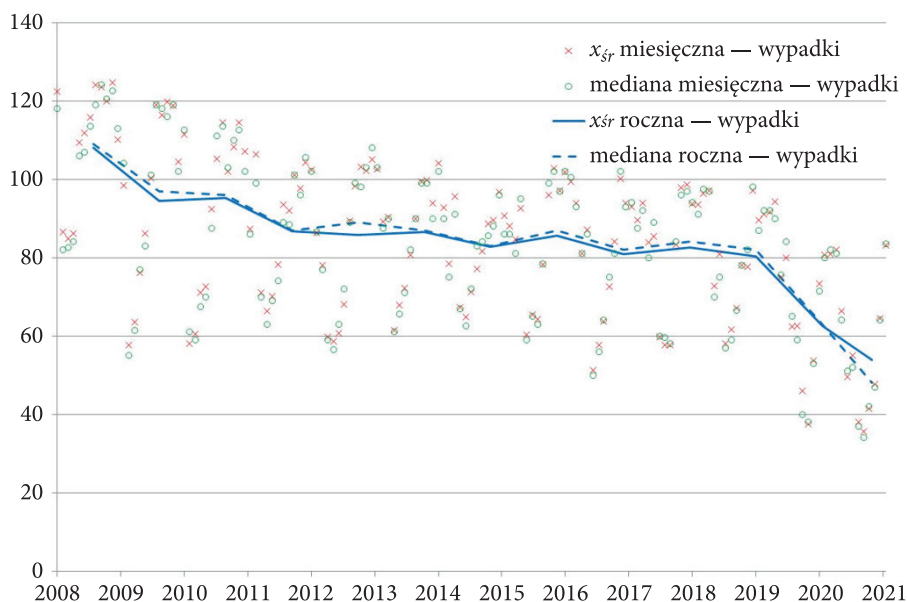
Zbliżone wartości mediany i wartości średniej dla „wypadków”, „zabitych” oraz „rannych” sugerują symetryczność odpowiednich rozkładów brzegowych, odmienna sytuacja występuje w przypadku „nietrzeźwych” — należy spodziewać się rozkładu niesymetrycznego.

Najczęściej: dziennie dochodziło do 85 wypadków, ginęło w nich sześć osób, a 101 odnosilo rany i zatrzymywano 161 nietrzeźwych kierowców. Minimalna liczba wypadków to 14 i doszło do nich 12 kwietnia 2020 r. Było 29 dni, w których nikt nie zginął (po jednym dniu w 2009, 2014, 2015, 2016, 2019; po dwa dni w 2010 i 2018; po trzy dni w 2012 i 2013; cztery dni w 2017; po pięć dni w 2020 i 2021). Nikt nie został ranny 4 stycznia 2021 r. Minimalna liczba zatrzymanych nietrzeźwych kierowców

to 57 — 6 listopada 2018. Z kolei maksymalne wartości kształtują się następująco: wypadki — 274 (5 października 2009), zabici — 30 (30 sierpnia 2009 i 12 października 2009), ranni — 365 (30 sierpnia 2009), nietrzeźwi — 1238 (5 października 2009). Jak widać, rok 2009 był najgorszym rokiem na polskich drogach, a 30 sierpnia 2009 to „czarny dzień”, w którym zginęło i zostało rannych najwięcej osób.

Uzyskane powyżej wyniki dotyczą całości danych, jednak pojawia się pytanie, jak kształtują się one w funkcji czasu. W celu uzyskania odpowiedzi przeprowadzono obliczenia uwzględniające dane pochodzące z poszczególnych miesięcy (parametry miesięczne) oraz z poszczególnych lat (parametry roczne). Rezultaty zostały zaprezentowane na rysunkach — rys. 1 (dotyczące liczby wypadków), rys. 2 (dotyczące liczby osób zabitych w wypadkach), rys. 3 (dotyczące liczby osób rannych w wypadkach), rys. 4 (dotyczące liczby zatrzymanych nietrzeźwych kierowców). Należy zwrócić uwagę na fakt, że skrajne roczniki, czyli 2008 oraz 2021, nie obejmują całego roku, jak również wyniki dotyczące lipca 2021 odnoszą się do pierwszej połowy tego miesiąca, czyli do 12.07.2021. Zdecydowano się je jednak zaprezentować, zdając sobie sprawę z mniejszych liczności odpowiednich prób.

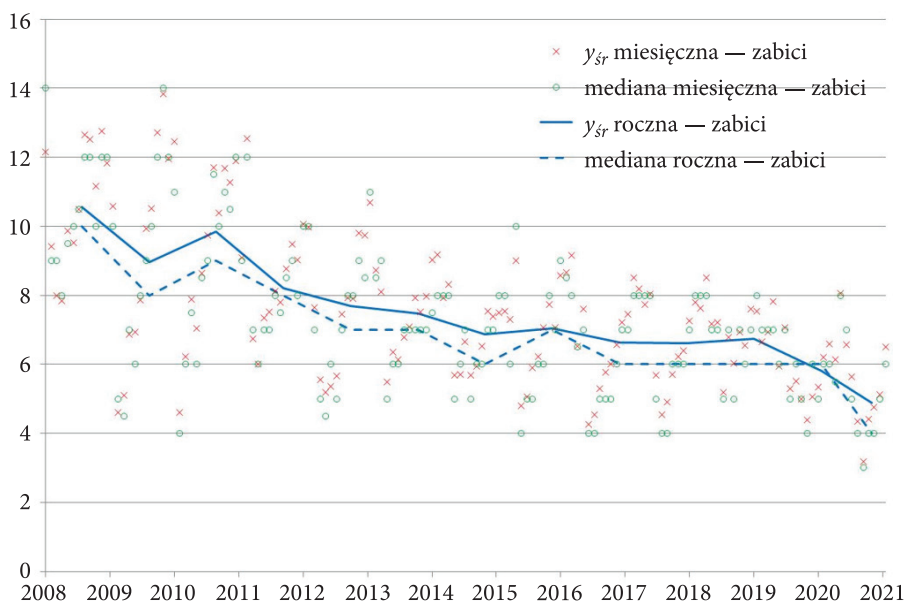
Na rysunkach 1-4 znacznikiem „x” koloru czerwonego zaznaczono odpowiednią miesięczną wartość średnią, znacznikiem „o” koloru zielonego medianę miesięczną odpowiedniej zmiennej losowej, natomiast liniami koloru niebieskiego — roczną wartość średnią (linia ciągła) oraz roczną medianę (linia przerywana).



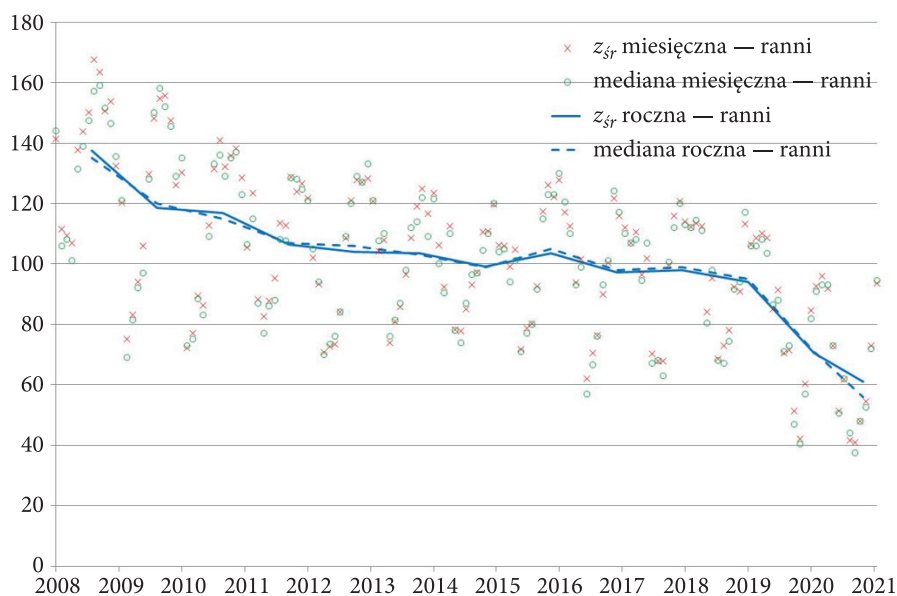
Rys. 1. Wartości średnie i mediany liczby wypadków w ujęciu miesięcznym i rocznym

Wyraźnie widać, zwłaszcza wyniki roczne jednoznacznie na to wskazują, że występuje pozytywny trend zmniejszenia wszystkich wartości średnich, czyli w długim okresie zmniejsza się liczba wypadków, zabitych, rannych i nietrzeźwych, pomimo sukcesywnego wzrostu natężenia ruchu. Z kolei wyniki miesięczne, o mniejszej liczności próby, bardziej odzwierciedlają charakter zmian źródłowych danych powiązanych z cyklem rocznym. Paradoksalnie w gorszych warunkach pogodowych — jesień, zima — dochodzi do mniejszej liczby wypadków i ich skutków, natomiast wartości maksymalne przypadają na miesiące letnie. Może to być spowodowane większą ostrożnością kierowców podczas złej pogody i brawurą przy dobrych warunkach jazdy. W przypadku kierowców nietrzeźwych wartości maksymalne występują latem, natomiast minimalne — zimą. Lato — okres urlopowy — sprzyja odpoczynkowi i może powodować utratę samokontroli.

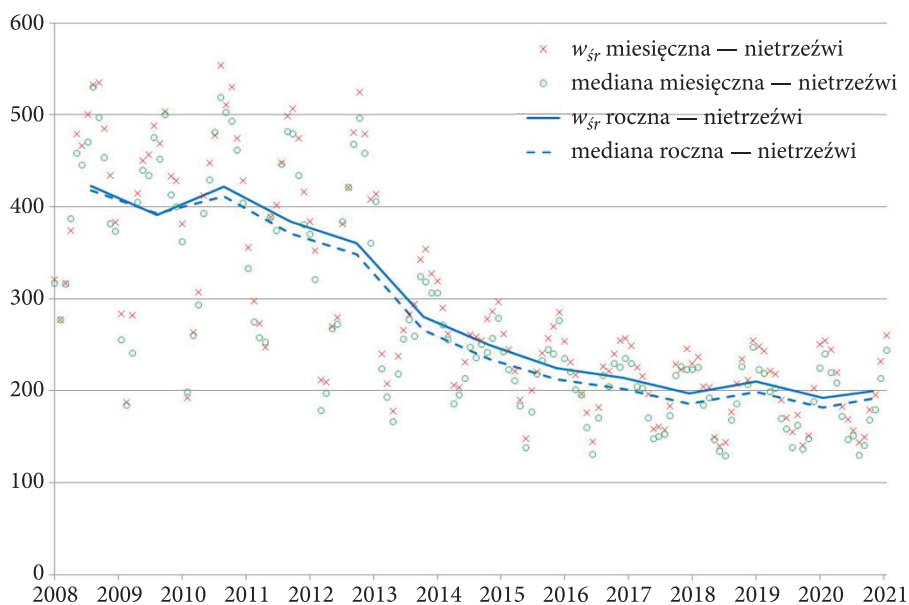
Dla liczby wypadków oraz liczby osób rannych można stwierdzić, zarówno dla danych miesięcznych, jak i rocznych, zbliżone wartości średnie i mediany. W tych przypadkach, zwłaszcza dla danych rocznych, mediana i wartość średnia praktycznie się pokrywają. Potwierdza to wcześniejszy wniosek o symetryczności odpowiednich rozkładów bazujący na danych całosciowych (tab. 1). Nieco większe różnice występują dla ofiar śmiertelnych oraz zatrzymanych nietrzeźwych kierowców, ale nie są zbyt znaczące, przez co można, odnosząc się do symetrii ewentualnych rozkładów, przyjąć wnioski wypływające z danych całosciowych.



Rys. 2. Wartości średnie i mediany liczby zabitych w wypadkach drogowych w ujęciu miesięcznym i rocznym



Rys. 3. Wartości średnie i mediany liczby rannych w wypadkach drogowych w ujęciu miesięcznym i rocznym



Rys. 4. Wartości średnie i mediany liczby zatrzymanych nietrzeźwych kierowców w ujęciu miesięcznym i rocznym

Warto zwrócić uwagę na końcowy odcinek rozpatrywanego okresu. Występuje tu większy spadek liczby wypadków, zabitych i rannych. Prawdopodobnie jest to związane z pandemią koronawirusa i zmniejszeniem mobilności społeczeństwa. Natomiast liczba zatrzymanych nietrzeźwych kierowców wskazuje na delikatny trend wzrostowy, co jest zjawiskiem niepokojącym. Oczywiście nie należy formułować zbyt daleko idących wniosków, ponieważ liczność próby dla roku 2021 jest zdecydowanie niższa niż w poprzednich latach, co mogło mieć wpływ na wyniki.

Oczywiście przedstawione wyniki stanowią wstępną analizę trendów czasowych i nie mogą być podstawą do pogłębionych wniosków.

Wyniki miesięczne odzwierciedlają sinusoidalny charakter zmian danych dziennych, przy czym występuje spadek zarówno wartości średnich, jak i amplitud wszystkich zmiennych. Jest to pozytywny trend.

3.2. Wyniki obliczeń współczynników korelacji zupełnej

Zaprezentowane dotychczas rezultaty tworzą pewien ogólny obraz zmian sytuacji na polskich drogach, jednak głównym celem artykułu jest wyznaczenie wartości współczynników korelacji przyjętych zmiennych losowych. Takie obliczenia zostały przeprowadzone. Podobnie jak wcześniej odniesiono się do całości danych oraz uwzględniono dane w ujęciu miesięcznym i rocznym. Wyniki dla całości danych zostały przedstawione w tabelach (tab. 2, 3, 4), przy czym zawarte tam wartości zostały zaokrąglone do drugiego miejsca po przecinku. Z uwagi na symetrię macierzy korelacyjnej zaprezentowano połowę macierzy.

TABELA 2

Współczynniki korelacji zupełnej dla całości danych od 1.12.2008 do 12.07.2021

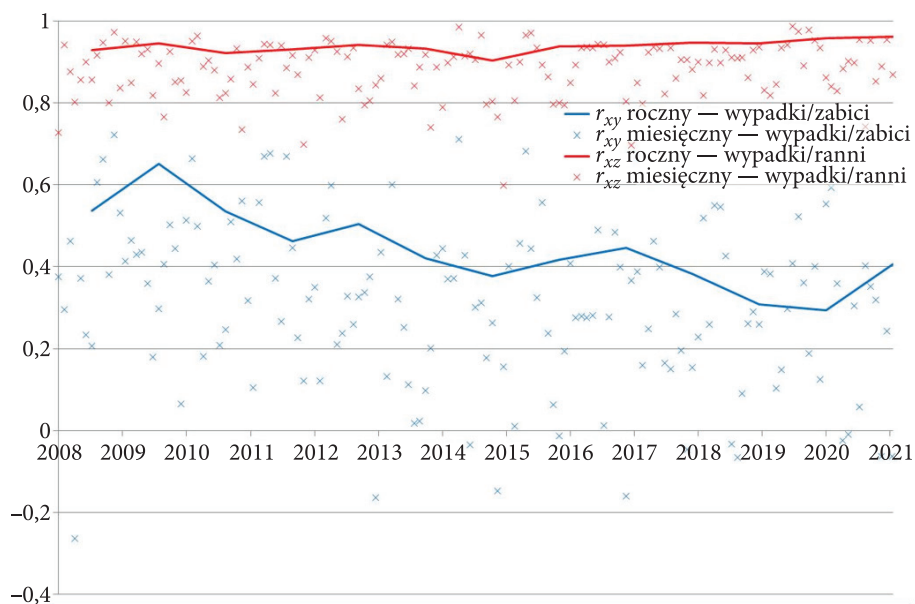
	<i>x</i> wypadki	<i>y</i> zabici	<i>z</i> ranni	<i>w</i> nietrzeźwi
<i>x</i> wypadki	1	0,55	0,95	0,48
<i>y</i> zabici		1	0,53	0,46
<i>z</i> ranni			1	0,56
<i>w</i> nietrzeźwi				1

Występuje niemal pełna dodatnia korelacja liczby wypadków z liczbą rannych oraz wysoka korelacja pozostałych zmiennych losowych. Wysokie wartości współczynnika korelacji zupełnej wskazują na liniową zależność zmiennych. Oczywiście w analizie korelacji mogą wystąpić dwa rodzaje błędów, brak korelacji przy istnieniu innej niż liniowa zależności zmiennych oraz wysoka korelacja przy generalnie braku ścisłej zależności liniowej. Sprawdzono, że w analizowanym zagadnieniu błędy te nie występują, jednak ze względu na ograniczenie objętościowe artykułu nie zostało to przedstawione.

Podobnie jak wcześniej dla wartości średnich postanowiono przedstawić zmianę w czasie odpowiednich współczynników korelacji, bazując na danych rocznych i danych miesięcznych.

Rezultaty dotyczące współczynników korelacji zupełnej przedstawione są na rysunkach 5-7.

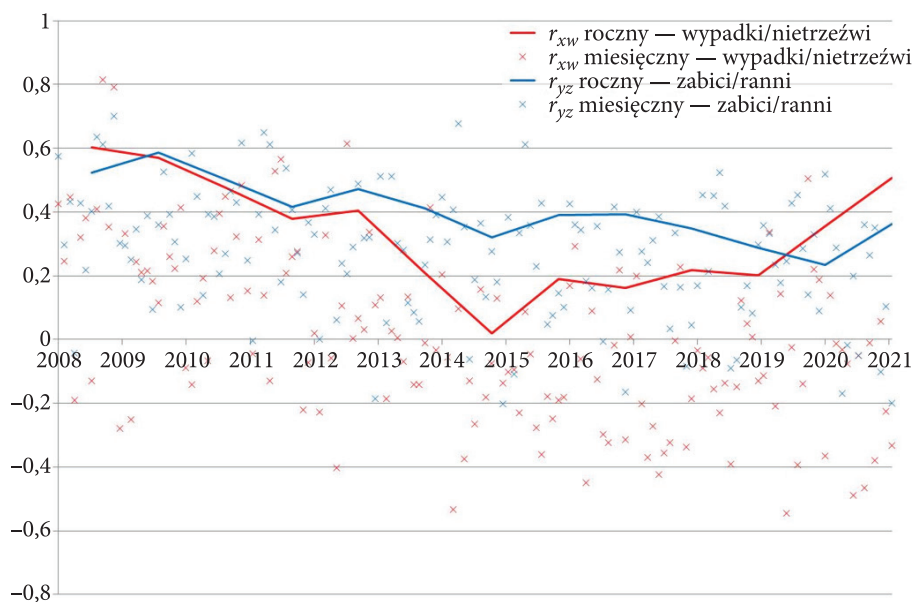
Biorąc pod uwagę liczności zbiorów danych uwzględnionych przy szacowaniu wartości odpowiednich współczynników korelacji, wiodącymi rezultatami są wyniki zaprezentowane powyżej, w następnej kolejności opracowane na podstawie danych rocznych, natomiast te uzyskane na podstawie danych miesięcznych należy traktować jako pewne ogólne przybliżenie. Podobnie jak poprzednio lata 2008 oraz 2021 obejmują próby o zdecydowanie mniejszej liczności niż inne lata, również dane z lipca 2021 mają mniejszą licznosc niż dane z innych miesięcy.



Rys. 5. Wartości współczynników korelacji zupełnej liczby wypadków i liczby zabitych oraz liczby wypadków i liczby rannych w ujęciu miesięcznym i rocznym

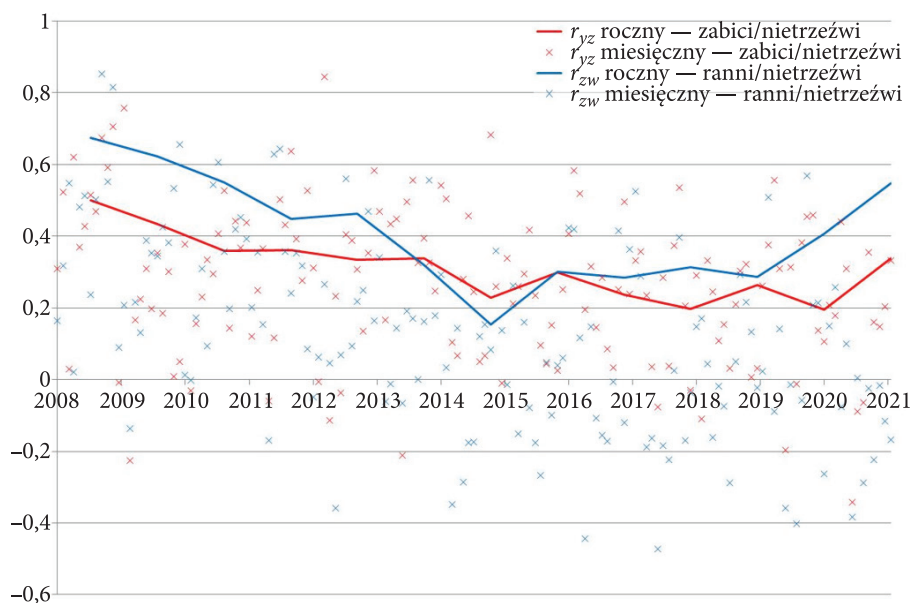
Wartości współczynników korelacji zupełnej (rys. 5) potwierdzają silną dodatnią zależność liniową w przypadku liczby rannych odniesionej do liczby wypadków, przy czym można przyjąć, bazując na „danych rocznych”, stałą i niezależną od czasu wartość tego współczynnika. Z kolei dla liczby ofiar śmiertelnych powiązanych z liczbą wypadków można zauważyć niższe wartości oraz pewien trend malejący. Mniejsze wartości korespondują z również mniejszą wartością współczynnika

korelacji dla danych całosciowych (tab. 2). Bezsprzecznie ofiary śmiertelne są skutkiem wypadków i istnieje tu bezpośredni związek. Zmiana w czasie współczynnika korelacji zupełnej sugeruje spadek zależności liniowej na rzecz innego powiązania tych zmiennych losowych. Niższe wartości mogą być spowodowane, ale nie muszą, pojawianiem się na drogach coraz nowocześniejszych samochodów, wyposażonych w układy bezpieczeństwa czynnego i biernego, przez co — zwłaszcza dla układów bezpieczeństwa biernego — zmieniają się skutki wypadków i w grupie ofiar następuje przesunięcie z grupy ofiar śmiertelnych do grupy rannych.



Rys. 6. Wartości współczynników korelacji zupełnej liczby wypadków i liczby nietrzeźwych kierujących oraz liczby zabitych i rannych w ujęciu miesięcznym i rocznym

Kolejny wykres (rys. 6) zawiera wartości współczynników korelacji zupełnej liczby zatrzymanych nietrzeźwych kierowców i liczby wypadków oraz liczby rannych w odniesieniu do liczby zabitych w wypadkach. W obu przypadkach (dane roczne) można zauważyć zarysowującą się pewną zmianę w czasie, jednak należy pamiętać, że w przypadku trzech i więcej zmiennych losowych korelacja zupełna nie daje pełnego obrazu. Podobnie jak poprzednio uzyskane wyniki odpowiadają wartościom w tabeli 2, wskazując na średnią korelację. Zmiana wartości opisuje zmianę „siły” zależności liniowej. Warto tu zwrócić uwagę na występowanie sporej grupy ujemnych wartości wyników miesięcznych, zwłaszcza dla korelacji wypadki – nietrzeźwi. Możliwa interpretacja tego faktu zostanie przedstawiona w dalszej części, przy analizie wyników korelacji cząstkowej.



Rys. 7. Wartości współczynników korelacji zupełnej liczby zabitych i nietrzeźwych kierujących oraz liczby rannych i nietrzeźwych kierujących w ujęciu miesięcznym i rocznym

Bardzo podobna sytuacja występuje dla korelacji liczby zabitych i nietrzeźwych kierujących oraz dla liczby rannych w wypadkach, również w powiązaniu z liczbą nietrzeźwych kierujących (rys. 7). Również tu istnieją ujemne wartości korelacji miesięcznej. Wartości korelacji rocznej wskazują na średnią korelację, potwierdzając przy tym wyniki zawarte w tabeli 2. Zmiana w czasie wartości współczynnika korelacji dotyczy głównie nietrzeźwych kierujących, co sugeruje inną niż liniową zależność tej zmiennej z pozostałymi.

3.3. Wyniki obliczeń współczynników korelacji cząstkowej

Oprócz wyznaczenia wartości współczynników korelacji zupełnej celowe wydaje się również określenie korelacji cząstkowej. Zaczerpnięta z Wikipedii anegdota mówi o istotnym statystycznie związku, na pewnym obszarze, pomiędzy liczebnością populacji bocianów a przyrostem naturalnym ludzi. Na tej podstawie można by wysnuć fałszywy wniosek, że to bociany przynoszą dzieci. Bardziej na poważnie: jeżeli mamy np. trzy zdarzenia A, B, C, przy czym zdarzenie A jest przyczyną zdarzeń B i C, natomiast skutki B i C nie mają ze sobą żadnego związku, to może wystąpić sytuacja, w której pojawi się wysoka (fałszywa) wartość współczynnika korelacji zupełnej zdarzeń B i C, powodowana wysoką korelacją par A — B oraz A — C.

Może to powodować błąd interpretacji wyników, zwłaszcza gdy przyczyna A pozostaje „poza kadrem”. Korelacja cząstkowa pozwala określić korelację pary zmiennych z danej grupy przy pominięciu wpływu pozostałych zmiennych.

TABELA 3

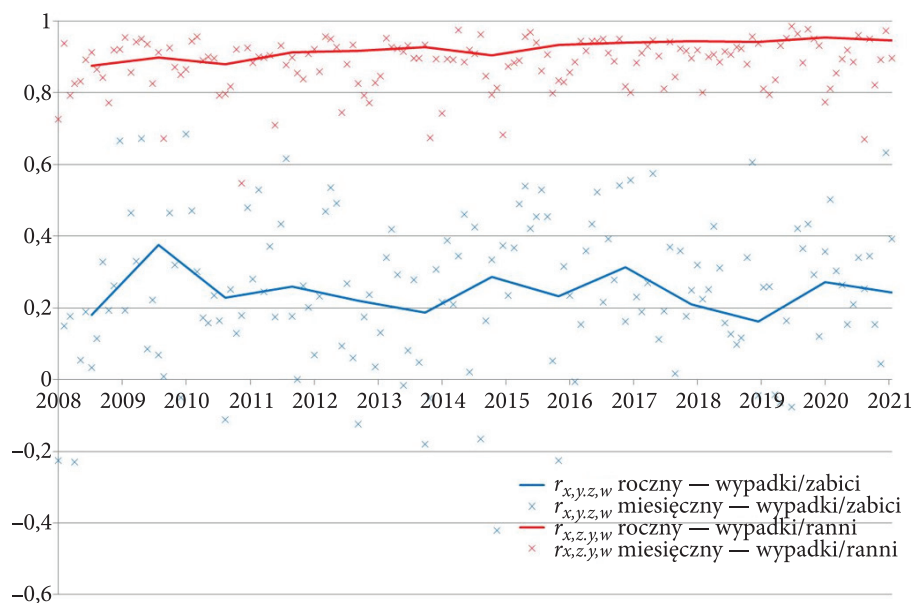
Współczynniki korelacji cząstkowej dla całości danych od 1.12.2008 do 12.07.2021

	x wypadki	y zabici	z ranni	w nietrzeźwi
x wypadki	1	0,22	0,92	-0,23
y zabici		1	-0,06	0,28
z ranni			1	0,37
w nietrzeźwi				1

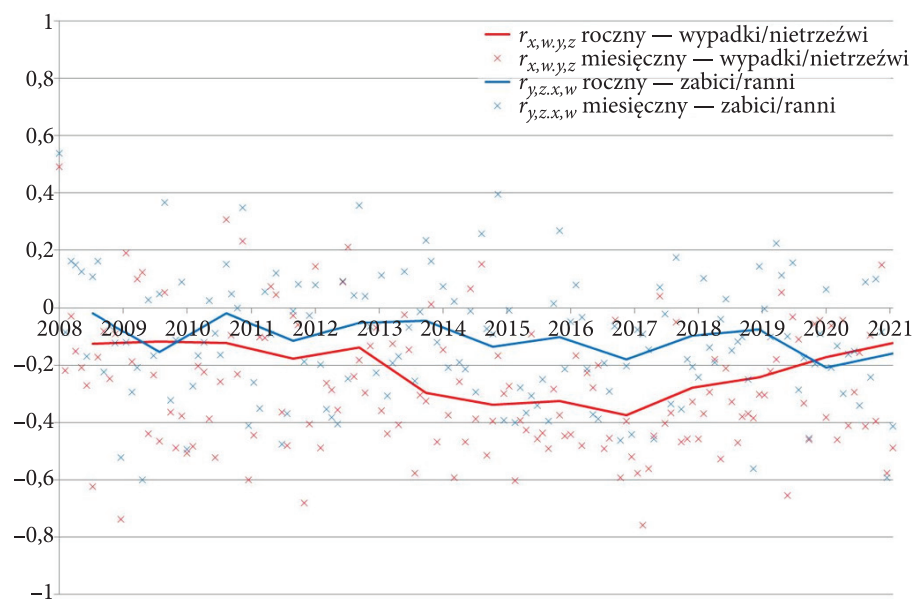
Współczynnik korelacji liczby wypadków z liczbą rannych ma nieco mniejszą wartość (niż współczynnik korelacji zupełnej), ale nadal występuje silny związek tych wielkości. Liczba rannych i liczba zabitych w wypadkach są praktycznie nieskorelowane. Z kolei liczba zatrzymanych nietrzeźwych kierowców w odniesieniu do liczby wypadków wskazuje na ujemną korelację, przy czym wartość bezwzględna współczynnika znacząco się zmniejszyła (w odniesieniu do współczynnika korelacji zupełnej). Zatrzymani nietrzeźwi kierowcy albo są sprawcami/uczestnikami wypadków, albo zostali zatrzymani w wyniku kontroli drogowej. Zwłaszcza w tym drugim przypadku Policja uniemożliwi kontynuację jazdy takiego kierowcy. Wskazuje to, że nietrzeźwość kierującego może być przyczyną wypadku oraz że działania prewencyjne Policji przynoszą pozytywny skutek (większa liczba zidentyfikowanych nietrzeźwych kierujących — mniejsza liczba wypadków). Z kolei stosunkowo mała wartość współczynnika korelacji może być spowodowana poruszaniem się po drogach znaczącej liczby niezidentyfikowanych nietrzeźwych kierowców. Oczywiście jest to jedna z wielu możliwych interpretacji tej sytuacji. Pozostałe współczynniki korelacji wskazują na mniejszą zależność zmiennych losowych niż w przypadku korelacji zupełnej.

Podobnie jak poprzednio, zachowując konwencję artykułu, na rysunkach 8-10 przedstawiono wartości współczynników korelacji cząstkowej w ujęciu miesięcznym i rocznym.

Dla korelacji wypadki - ranni (rys. 8) ogólny charakter zmienności współczynnika korelacji cząstkowej pozostaje podobny do współczynnika korelacji zupełnej. Natomiast współczynnik korelacji cząstkowej wypadki — zabici (rys. 8) przyjmuje mniejsze wartości i w przybliżeniu może być uznany za stały. Niestety podana poprzednio interpretacja (jedna z możliwych) jest zbyt optymistyczna, dynamika zmian zasobu pojazdów na rzecz nowocześniejszych rozwiązań nie jest tak duża, jak można by oczekiwać.



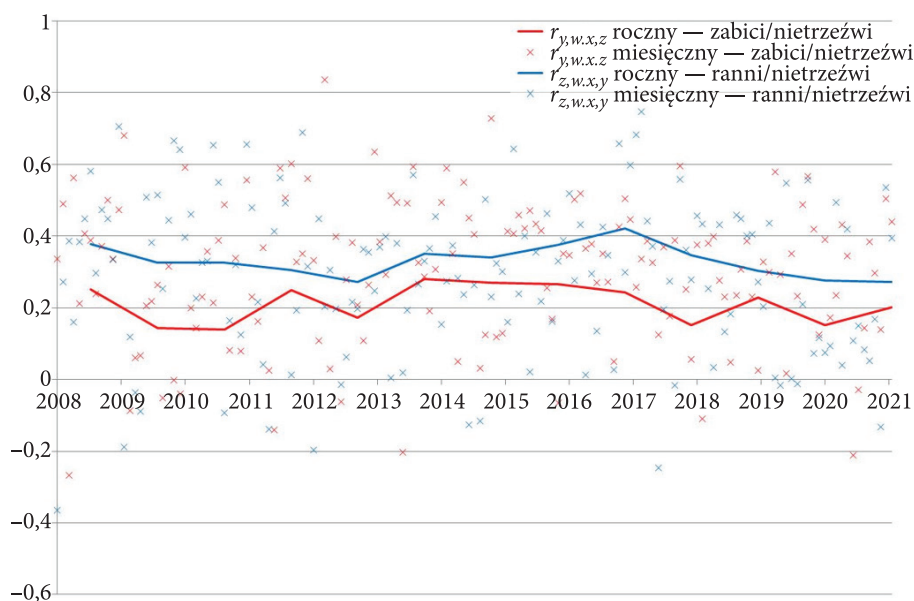
Rys. 8. Wartości współczynników korelacji cząstkowej liczby wypadków i zabitych oraz liczby wypadków i rannych w ujęciu miesięcznym i rocznym



Rys. 9. Wartości współczynników korelacji cząstkowej liczby wypadków i nietrzeźwych kierowców oraz liczby zabitych i rannych w ujęciu miesięcznym i rocznym

Korelacja cząstkowa wypadki — nietrzeźwi oraz zabici — ranni (rys. 9) przyjmuje ujemne wartości. Interpretacja ujemnej wartości współczynnika korelacji wypadki — nietrzeźwi została przedstawiona wcześniej. Korelacja zabici — ranni również jest ujemna, ale przyjmuje małe wartości (w sensie wartości bezwzględnej), sugerując słaby związek liniowy tych zmiennych losowych i nie daje podstaw do daleko idących wniosków. Oba współczynniki w przybliżeniu można uznać za stałe.

Pozostały jeszcze dwa współczynniki korelacji cząstkowej, których zmianę w czasie przedstawiono na rysunku 10. Oba współczynniki przyjmują średnie wartości, nieco mniejsze niż w przypadku korelacji zupełnej. Można je natomiast potraktować jak wielkości stałe, odmiennie niż w przypadku korelacji zupełnej.



Rys. 10. Wartości współczynników korelacji cząstkowej liczby zabitych i nietrzeźwych kierujących oraz liczby rannych i nietrzeźwych kierujących w ujęciu miesięcznym i rocznym

Podsumowując ten punkt, można stwierdzić, że wszystkie współczynniki korelacji można uznać za stałe, a określone trendy czasowe współczynników korelacji zupełnej spowodowane były wzajemnymi powiązaniami w rozpatrywanej grupie zmiennych losowych.

3.4. Wyniki obliczeń współczynników korelacji wielorakiej

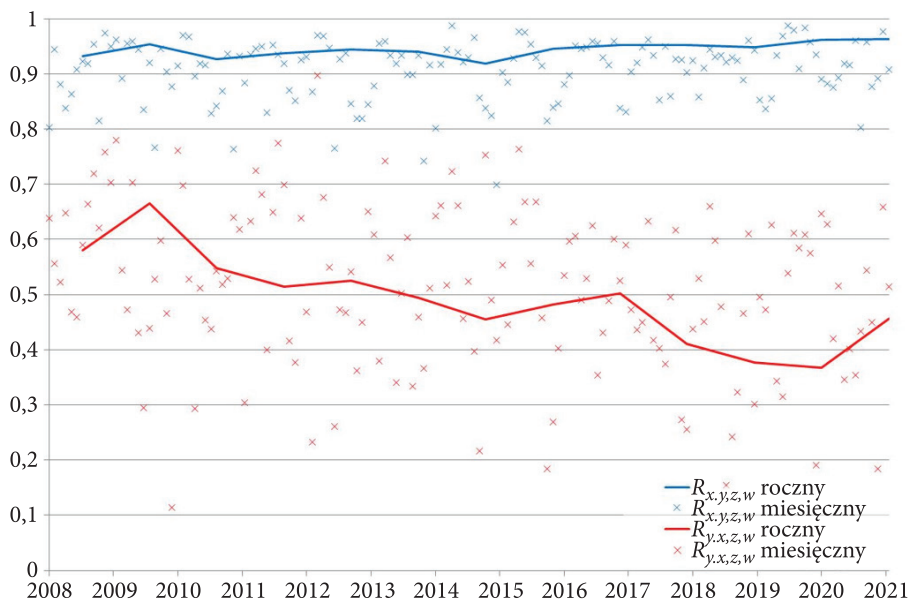
Pozostało jeszcze określenie wartości współczynników korelacji wielorakiej. Uzyskane wyniki prezentowane są w tabeli 4 oraz na rysunkach 11 i 12. Współczynnik korelacji wielorakiej zawsze przyjmuje wartości dodatnie, czyli wskazuje na „siłę” zależności danej zmiennej losowej od grupy pozostałych zmiennych, jednak nie określa „kierunku” — czy jest to zależność ujemna, czy też dodatnia.

TABELA 4

Współczynniki korelacji wielorakiej dla całości danych od 1.12.2008 do 12.07.2021

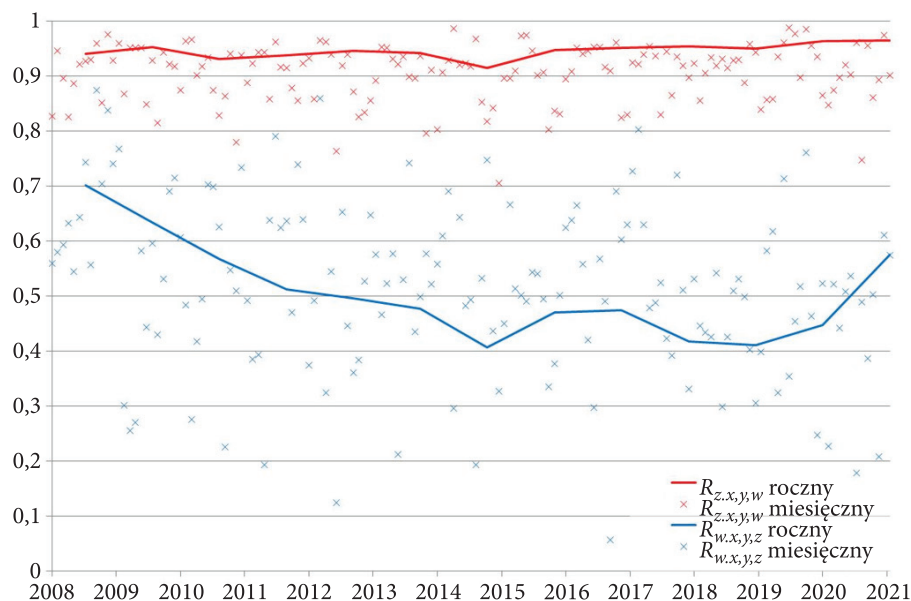
$R_{x,y,z,w}$ wypadki	0,95
$R_{y,x,z,w}$ zabici	0,60
$R_{z,x,y,w}$ ranni	0,95
$R_{w,x,y,z}$ nietrzeźwi	0,62

Współczynniki korelacji wielorakiej wskazują na silne zależności grupowe dla wypadków i liczby rannych, nieco słabsze, ale również w miarę wysokie dla pozostałych zmiennych.



Rys. 11. Wartości współczynników korelacji wielorakiej liczby wypadków oraz liczby zabitych z pozostałymi grupami zmiennych w ujęciu miesięcznym i rocznym

Ostatnimi wynikami są obliczone wartości współczynników korelacji wielorakiej prezentowane na rysunkach 11 i 12. Można zauważyć silne zależności grupowe dla „wypadków” i „rannych”. Słabszy związek występuje dla „zabitych” i „nietrzeźwych”. Spowodowane jest to wysokimi wartościami korelacji zupełnej i częściowej wypadki – ranni. Korelacja wieloraka może wskazywać na możliwość zgrupowania pewnych zmiennych, np. rannych i zabitych, jednak poprzednie wyniki nie dają takiej możliwości i wszystkie zmienne losowe należy traktować oddzielnie.



Rys. 12. Wartości współczynników korelacji wielorakiej liczby rannych oraz liczby nietrzeźwych kierujących z pozostałymi grupami zmiennych w ujęciu miesięcznym i rocznym

4. Podsumowanie

W toku prac wyznaczono wartości odpowiednich współczynników korelacji. Bazowano na aktualnych na dzień zgłoszenia artykułu danych — do 12 lipca 2021. Stwierdzono, że współczynniki korelacji można traktować jako niezależne od czasu, co oznacza, że sytuacja na polskich drogach podlega raczej ewolucji niż rewolucji. Biorąc pod uwagę dodatkowo wyznaczone wartości średnie i mediany, widać powolną poprawę bezpieczeństwa, czyli choć zmiany są ewolucyjne, zmierzają w pożądanym kierunku. Pocięszający jest spadek wszystkich wartości średnich, nawet przy jak się wydaje wzroście natężenia ruchu.

Najwyższa korelacja dotyczy pary wypadki — ranni. Czyli liczba rannych jako skutek jest w przybliżeniu liniowo zależna od liczby wypadków (przyczyna). Zmienne „ranni” i „zabici” można uznać za nieskorelowane, choć z pewnością są powiązane, jednak nie jest to zależność liniowa. Zmienna „nietrzeźwi” jest słabo skorelowana z pozostałymi zmiennymi, przy czym dla „wypadków” występuje ujemna zależność. Część nietrzeźwych kierujących (sprawców) identyfikowana jest już po wystąpieniu wypadku, jednak ujemna wartość wskazuje, że spora grupa jest identyfikowana w trakcie czynności profilaktycznych Policji i eliminowana z dalszej jazdy, co zapobiega wypadkom.

W przypadku wartości średnich i median stosunkowo małe liczności prób (miesięczne — około 30) dają wyniki niosące pewną informację, w skali roku można stwierdzić w przybliżeniu sinusoidalny charakter zmian zmiennych losowych. W przypadku korelacji pełniejszy obraz dają próby o większej liczności (roczne — około 365 i całociowe — 4584).

Jak się wydało, określony we wstępie cel został osiągnięty.

Źródło finansowania pracy — działalność statutowa Wojskowej Akademii Technicznej.

Artykuł wpłynął do redakcji 22.07.2021. Zatwierdzono do publikacji 13.08.2021.

Sławomir Stępień <https://orcid.org/0000-0002-5340-9980>.

LITERATURA

- [1] JAMROZ K., ŻUKOWSKA J., *Problemy i wyzwania w zakresie bezpieczeństwa ruchu drogowego w Nowej Dekadzie*, Transport Miejski i Regionalny, nr 10, 2020.
- [2] ORŁOWSKI Ł., WSZEBOROWSKI R., *Bezpieczeństwo w ruchu drogowym*, Systemy Logistyczne Wojsk, nr 53, 2020.
- [3] WRÓBEL D., *Wpływ rozwoju sieci dróg wojewódzkich na bezpieczeństwo ruchu*, Drogownictwo, vol. 75, nr 1, 2020.
- [4] SUKINNIK K., POWROŹNIK I., *Bezpieczeństwo w ruchu drogowym w Polsce a zarządzanie czasem pracy kierowców samochodów ciężarowych*, Zeszyty Naukowe Politechniki Częstochowskiej, nr 37, 2020.
- [5] GRZELAK M., BORUCKA A., *Ocena bezpieczeństwa ruchu drogowego (BRD) w Polsce*, Systemy Logistyczne Wojsk, nr 51, 2019.
- [6] BORUCKA A., KOZŁOWSKI E., OLESZCZUK P., ŚWIDERSKI A., *Predictive analysis of the impact of the time of day on road accidents in Poland*, Open Engineering, 11, 1, 2020, 142-150.
- [7] KULAWIECKA E., *Rachunek korelacji w naukach o bezpieczeństwie z wykorzystaniem programu Statistica*, Obronność – Zeszyty Naukowe Wydziału Zarządzania i Dowodzenia Akademii Sztuki Wojennej, 4, 20, 2016.
- [8] LASOTA D., STANISZEWSKA A., TARCHALSKA-KRYŃSKA B., MIROWSKA-GUZEL D., KRAJEWSKI P., *Nietrzeźwość a ryzyko zgonu w wypadku komunikacyjnym*, Journal of Education, Health and Sport, 5, 6, 2015, 419-432.

- [9] KAPITANOV V., MONINA O., SILYANOV V., CHUBUKOV A., *Probabilistic assessment of main factors determining the road traffic accident rate in regions of Russia*, XIV International Conference 2020 SPbGASU Organisation and safety of traffic in large cities, Elsevier, Transportation Research Procedia, 50, 2020, 218-225.
- [10] www.policja.pl, [dostęp: 12.07.2021].

S. STĘPIEŃ

Preliminary analysis of road accident data in Poland

Abstract. In the paper, preliminary statistical analysis of the data concerning road accidents in Poland are pre-sented. The number of accidents, fatalities, as well the number of persons injured in accidents with multitude of the detained drunk drivers are considered. The analysis covers daily data from the year 2008 to 2021. Mean and median values were determined and a comparative analysis based on the complementary, partial, and multiple correlation coefficients was carried out. The results are presented for the whole compiled data and separately for monthly and annual quantities.

Keywords: mechanical engineering, correlation analysis, road safety, road accidents

DOI: 10.5604/01.3001.0015.3839

