

POLITECHNIKA WROCŁAWSKA  
WYDZIAŁ INFORMATYKI I TELEKOMUNIKACJI



---

# Wpływ różnych czynników na liczbę wypadków na drogach

---

Sprawozdanie z laboratorium

AUTOR

**Paweł Wieszczezyński**

nr albumu: **260414**

kierunek: **Informatyka Stosowana**

*13 czerwiec 2022*

### Streszczenie

W przeprowadzonej analizie zbadano wpływ różnych czynników na ilość wypadków (zdarzenie mające związek z ruchem pojazdów na drogach publicznych, w wyniku którego nastąpiła śmierć lub uszkodzenie ciała osób) na drogach w Polsce. Dane pobrano z Głównego Urzędu Statystycznego, a dokładniej z Banku Danych Lokalnych. W analizie wykorzystano model liniowy, oraz model kwadratowy regresji. W wyniku analizy otrzymano 5 wykresów, ukazujących wpływ każdego z wybranych przeze mnie czynników na liczbę wypadków na drogach.

## 1 Wstęp – sformułowanie problemu

Autor potrzebuje ocenić, jakie czynniki realnie wpływają na liczbę wypadków na drogach. Dzięki tej analizie, możliwa będzie zmiana niektórych czynników, tak aby w rezultacie zmniejszyć liczbę wypadków.

## 2 Opis danych

Wielkość datasetu to 17 wierszy. Są to poszczególne dane z lat 2004 - 2020. Analizie poddane zostały następujące czynniki:

1. cena 0,5l wódki czystej 40 procent [zł]
2. wydatki budżetów województw w dziale transport i łączność [tys. zł]
3. emisja zanieczyszczeń pyłowych z zakładów szczególnie uciążliwych [ton rocznie]
4. cena kursu samochodowego kat. "B" [zł]
5. liczba osób w wieku produkcyjnym (mężczyźni 18-64 lat, kobiety 18-59 lat)

## 3 Opis rozwiązania

Z Banku Danych Lokalnych eksportowano potrzebne dane do plików MS Excel, następnie dane zostały umieszczone w Data Frame z biblioteki Pandas. Następnie dla każdego z czynników stworzono wykresy pokazujące regresję liniową i kwadratową. Użyte zostały LinearRegression z biblioteki sklearn oraz model oparty na optimize.curve\_fit pokazany na laboratoriach. Użyto również metod mean\_square\_error oraz mean\_absolute\_error z biblioteki sklearn, w celu obliczenia błędów regresji.

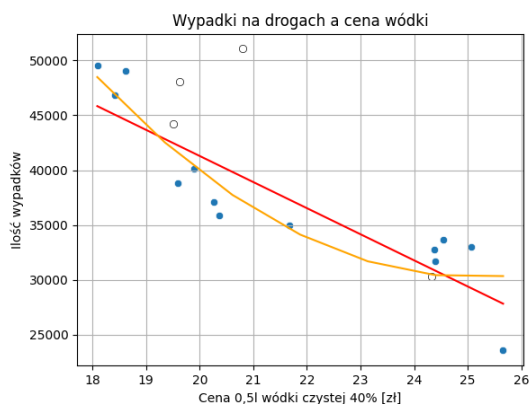
## 4 Rezultaty obliczeń

### 4.1 Plan badań

Zbiór danych zostanie podzielony na dwie części: treningową i testową w stosunku 80:20.

## 4.2 Wyniki obliczeń

### 4.2.1 Cena wódki



Rysunek 1: Wypadki na drogach a cena wódki

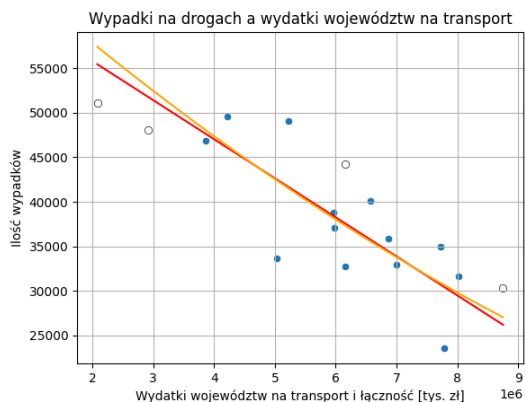
```

Ceny wódki
MAE_lin: 4998.7163536839835
MAE_custom: 5777.470783091216
MSE_lin: 43697366.48490093
MSE_custom: 61176752.70317642
  
```

Rysunek 2: MAE i MSE cen wódki

Jak można zauważyć na wykresie, wraz ze wzrostem cen wódki, znacznie maleje ilość wypadków na drogach (stromy wykres liniowy). Analizując regresję kwadratową, można również zauważyć, że ilość wypadków spada bardzo szybko, kiedy ceny wódki rosną z niskich wartości. Kiedy wódka jest droga, wzrost ceny nie powoduje już tak dużego zmniejszenia wypadków na drogach. Jeśli chodzi o Mean Absolute i Mean Square Error, to regresja liniowa wykazuje mniejsze błędy od regresji kwadratowej, co oznacza, że trafniej określa ona rzeczywistą zależność między ceną wódki a liczbą wypadków. Przewidywana przez regresję wartość średnio różni się o około 5400 wypadków.

## 4.2.2 Wydatki województw



Rysunek 3: Wypadki na drogach a wydatki województw na transport

```

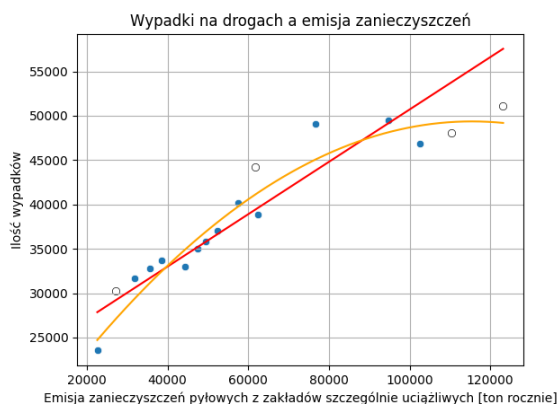
Wydatki województw
MAE_lin: 4696.033143906203
MAE_custom: 5314.538331160424
MSE_lin: 23332956.972158946
MSE_custom: 30219417.433760177

```

Rysunek 4: MAE i MSE wydatków województw

Ponownie, zwiększenie badanego czynnika znacząco redukuje ilość wypadków na drogach. Co ciekawe, regresja liniowa, jak i kwadratowa wyglądają praktycznie tak samo, co oznacza, że już model liniowy, relatywnie dobrze opisuje zależność pomiędzy wypadkami a wydatkami województw. Podobnie jest również w przypadku błędów, funkcja liniowa lepiej określa rzeczywistą zależność między wydatkami województw na transport a liczbą wypadków. Warto jednak zauważyć, że różnica pomiędzy błędami średniokwadratowymi jest dużo mniejsza, niż w przypadku cen wódki, co oznacza, że większość błędów jest do siebie podobna i nie ma błędów znacznie odbiegających od reszty. W przypadku cen wódki, prawdopodobnie jest trochę błędów małych i trochę błędów bardzo dużych, przez co MAE wyglądają podobnie jak tutaj, ale MSE są już dużo większe. Przewidywana przez regresję wartość średnio różni się o około 5000 wypadków.

## 4.2.3 Emisja zanieczyszczeń



Rysunek 5: Wypadki na drogach a emisja zanieczyszczeń

```

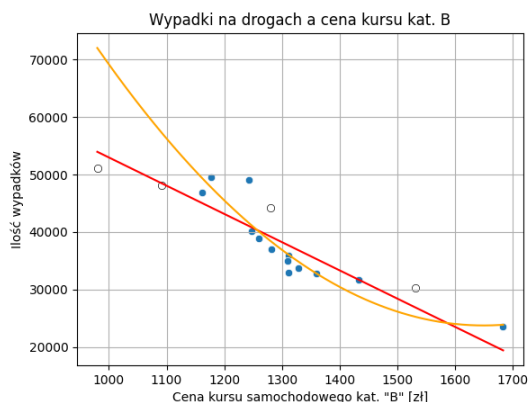
Emisja zanieczyszczeń
MAE_lin: 4514.925573805559
MAE_custom: 2347.526056382412
MSE_lin: 24635425.05685594
MSE_custom: 6232882.4962938065

```

Rysunek 6: MAE i MSE emisji zanieczyszczeń

W tym przypadku zależność jest odwrotna. Zwiększenie emisji zanieczyszczeń pyłowych w zakładach szczególnie uciążliwych zwiększa ilość wypadków na Polskich drogach. Dzięki regresji kwadratowej, można zauważyć, że najszybciej liczba wypadków rośnie, kiedy emisja zanieczyszczeń jest niska. Kiedy emisja jest wysoka, dodatkowe jej zwiększenie nie wpływa już tak mocno na ilość wypadków. W tym przypadku regresja kwadratowa jest zdecydowanie lepsza od regresji liniowej. Trafniej ocenia ona zależność, gdyż jej MAE jest prawie dwa razy mniejsze, a MSE prawie 4 razy mniejsze. Przewidywana przez regresję kwadratową wartość, średnio różni się o 2350 wypadków.

## 4.2.4 Cena kursu kat. "B"



Rysunek 7: Wypadki na drogach a cena kursu kat. "B"

```

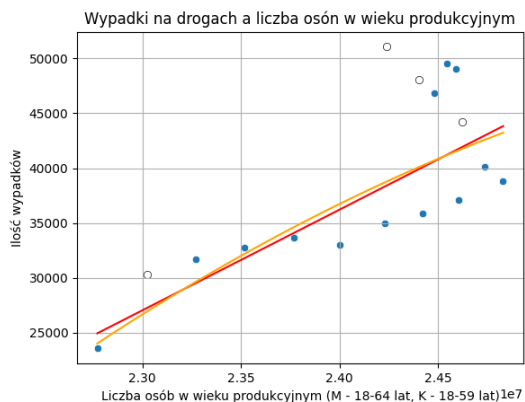
Cena kursu kat. "B"
MAE_lin: 2908.84156063551
MAE_custom: 10218.558491269694
MSE_lin: 11170509.676308025
MSE_custom: 145302491.5730199

```

Rysunek 8: MAE i MSE cen kursu kat. "B"

Jak można zauważyć na wykresie, wraz ze wzrostem cen kursu kat. "B", zmniejsza się ilość wypadków na drogach. Wykres regresji kwadratowej znacznie różni się od regresji liniowej i pokazuje że zmiany bliżej początku wykresu mają większe znaczenie niż zmiany przy wyższych wartościach. Błędy regresji kwadratowej są jednak znacznie większe niż błędy regresji liniowej, co oznacza, że dużo gorzej określa ona zależność pomiędzy ceną kursu kat. "B" a liczbą wypadków. Przewidywana przez regresję liniową wartość, średnio różni się o około 3000 wypadków.

## 4.2.5 Liczba osób w wieku produkcyjnym



Rysunek 9: Wypadki na drogach a liczba osób w wieku produkcyjnym

```

Liczba osób w wieku produkcyjnym
MAE_lin: 6533.705614448401
MAE_custom: 6515.861694476713
MSE_lin: 60407475.03937385
MSE_custom: 58049660.47963487

```

Rysunek 10: MAE i MSE liczby osób w wieku produkcyjnym

Jak pokazuje ostatni wykres, wraz ze wzrostem liczby osób w wieku produkcyjnym, rośnie również liczba wypadków. Ponownie, wykres regresji kwadratowej jest niemalże taki sam jak regresji liniowej. Zarówno MAE, jak i MSE w tym przypadku są bardzo do siebie zbliżone, na co wpływ ma na pewno bardzo duże podobieństwo pomiędzy liniami regresji obu metod. Minimalnie lepsza okazuje się regresja kwadratowa. Przewidywana przez regresję wartość średnio różni się o około 6500 wypadków.

## 5 Wnioski

Pokazane na wykresach zależności pokazują, że liczbę wypadków można skojarzyć z wieloma czynnikami. Zazwyczaj największe zmiany zachodzą na początku wykresu, przy relatywnie niskich wartościach. Analizowane zależności są również zgodne z logiką, i można określić ich prawdopodobną przyczynę.

Jeśli chodzi o cenę wódki, to jej wzrost może znacząco wpływać na zmniejszenie liczby wypadków, przez to, że w Polsce jest duży problem z kierowaniem pod wpływem alkoholu. Wyższe ceny sprawiają, że mniej wódki jest kupowane, a co za tym idzie, być może mniej osób kieruje po pijaku.

W przypadku liczby osób w wieku produkcyjnym liczba ta bezpośrednio przekłada się na liczbę samochodów na drogach. Więcej ludzi dojeżdżających do pracy — więcej samochodów, więcej wypadków — więcej wypadków.

Wydatki województw na transport i łączność bezpośrednio wpływają na infrastrukturę drogową w Polsce, co przekłada się na bezpieczeństwo i mniejsze szanse na zdarzenie się wypadku.

Zwiększenie cen kursu kat. "B" prawdopodobnie zmniejsza ilość kierowców na drogach. Część osób może nie być skłonna zapłacić większej ceny za kurs, a w efekcie mogą oni podjąć decyzję o niezdawaniu prawa jazdy. Być może również większe ceny kursu przekładają się na lepszą jakość nauczania, co zwiększa umiejętności kierowców na drogach. Oba te czynniki zmniejszają ilość wypadków na drogach.

Negatywnie z kolei wpływa zwiększenie emisji zanieczyszczeń pyłowych. Również to jest prawdopodobne, gdyż ilość zanieczyszczeń pyłowych może ograniczać widoczność, co zmniejsza bezpieczeństwo na drogach i może zwiększyć szanse na wypadek.

Przedstawiona analiza może pozytywnie wpłynąć na zmniejszenie ilości wypadków na drogach. Przykładowo zarządy województw mogą przeznaczać więcej pieniędzy na transport i łączność, gdyż lepsza infrastruktura może zwiększyć bezpieczeństwo na drogach. Analogicznie, kiedy podniesione zostaną ceny wódki, jest bardzo duża szansa na to, że wraz z konsumpcją wódki, zmniejszy się również ilość wypadków. Model ten jest jednak dużym uproszczeniem i w rzeczywistości na liczbę wypadków drogowych ma wpływ jeszcze wiele innych czynników.

## A Dodatek

Kod źródłowy został umieszczony w repozytorium na github.com.  
[https://github.com/PawelWieszczeczynski/MSiD\\_Projekt](https://github.com/PawelWieszczeczynski/MSiD_Projekt).