Charles University in Prague
Faculty of Social sciences

BACHELOR THESIS

Jakub Stejskal

Socio-Demographic determinants of traffic accidents in the Czech Republic

Institute of economic studies

Supervisor of the bachelor thesis: Mgr. Roman Kalabiška
Study programme: Economics
Specialization: Economics and Finance

Prague 2021

I would like to thank my supervisor Mgr. Roman Kalabiška for guiding me through the process of writing the bachelor thesis. I would also like to thank my parents for supporting me during my studies.

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague on 27.07.2021 Jakub Stejskal

Název práce: Socio-Demografické determinanty dopravních nehod v České republice

Autor: Jakub Stejskal

Katedra: Institut ekonomických studií

Vedoucí bakalářské práce: Mgr. Roman Kalabiška

Abstrakt: Cílem této práce je analyzovat proměnné, které ovlivňují různé aspekty dopravních nehod v České republice jako je četnost či smrtnost. Dopravní nehody jsou velkým problémem. Jen v česku je jich každý rok přes 100 000. Hlavním přínosem této práce je nový přístup zkoumající specifické typy dopravních nehod jako jsou dopravní nehody způsobené nepřiměřenou rychlostí, představení nových proměnných, které by mohly částečně vysvětlovat rozdíly mezi nehodovostí napříč okresy a replikace již stávajících studií na území České republiky. Náš výzkum ukázal například, že na jednu stranu je procento sezdaných lidí v jednotlivém okrese negativně korelováno s počtem dopravních nehod, ale na druhou stranu je pozitivně korelováno s průměrným počtem smrtí na dopravní nehodu. Jedno z možných vysvětlení je, že sezdaní lidé řídí bezpečněji, ale v průměru jich jede v autě více, což v případě dopravní nehody vede k větší šanci, že někdo zemře.

Klíčová slova: dopravní nehoda, oběti dopravní nehody, Česká republika, řízení pod vlivem alkoholu

Title: Socio-Demographic determinants of traffic accidents in the Czech Republic

Author: Jakub Stejskal

Department: Institute of Economic Studies

Supervisor: Mgr. Roman Kalabiška

Abstract: This thesis aims to analyze variables influencing different aspects of traffic accidents in the Czech Republic, such as frequency or fatality. Traffic accidents are a huge issue as there are 100 000 of them in the Czech Republic every year alone. The main contribution of this thesis is a new approach looking only at specific types of traffic accidents, such as those caused by speeding, introducing new variables that might be related to traffic accidents, and replicating the results of existing research in the area of the Czech Republic. On the one hand, we found a negative correlation between the portion of wedded people in the area and the number of traffic accidents. On the other hand a positive correlation between portion of wedded people and the number of people killed per accident. One possible explanation for this is that wedded people drive safer, but on average, there are more people in their car, resulting in more deaths in case of traffic accidents.

Keywords: traffic accident, traffic fatality, drunk driving, the Czech Republic

Contents

List of Figures 2

List of Tables 3

1 Introduction 4
1.1 Road Traffic accidents . . . 4

2 Literature review 5
2.1 Intercountry studies . . . 5
2.2 Intra country studies . . . 6
2.3 Alcohol-related traffic accidents . . . 7

3 Data 8
3.1 Definition of traffic accidents . . . 8
3.2 Traffic accidents data . . . 8
3.2.1 Demographic data . . . 9
3.2.2 Infrastructure data . . . 9
3.3 Processing the data . . . 9
3.4 Variables . . . 10
3.4.1 Dependent variables . . . 10
3.4.2 Explanatory variables . . . 11
3.5 Descriptive analysis of the data . . . 13
3.5.1 Dependent variables . . . 13
3.5.2 Explanatory variables . . . 14

4 Methodology and models 15
4.1 Choosing the right method . . . 15
4.2 Assumptions . . . 15
4.2.1 Models . . . 18

5 Results 19
5.1 Frequency of the traffic accidents . . . 19
5.2 The severity of the traffic accidents . . . 21
5.3 Specific types of accidents . . . 25

6 Conclusion 29

Bibliography 31

A Tables 34

List of Figures

4.1 Correlation Matrix 17

A.1 Traffic accidents per inhabitant 34
A.2 Percentage of alcohol-related traffic accidents 39
A.3 Residuals vs fitted graph for model accidentsinhratio 40
A.4 Residuals vs fitted graph for model accidentsroadratio 41
A.5 Residuals vs fitted graph for model killedper10k 42
A.6 Residuals vs fitted graph for model carkills10k 43
A.7 Residuals vs fitted graph for model killedperacc 44
A.8 Residuals vs fitted graph for model killedperacc2 45
A.9 Residuals vs fitted graph for modelalco 46
A.10 Residuals vs fitted graph for modelpolice 47
A.11 Residuals vs fitted graph for modelspeed 48

List of Tables

3.1 Descriptive analysis of dependent variables 13
3.2 Descriptive analysis of the data 14

5.1 Frequency of the traffic accidents 20
5.2 Severity of the accidents 22
5.3 Fatalities per accident 24
5.4 Percentages of accidents 27

A.1 The p-values for the different tests 35
A.2 VIF model accinhratio 35
A.3 VIF model accroadratio 35
A.4 VIF model carkills10k 35
A.5 VIF model kills10kinh 35
A.6 VIF model killedperacc 1 35
A.7 VIF model killedperacc 2 36
A.8 VIF model alcoratio 36
A.9 VIF model polaccratio 36
A.10 VIF model speedratio 36
A.11 HCCM for modelalco 36
A.12 Fatalities during accidents not involving crash of 2 or more cars . 37
A.13 Fatality of single person per accident 38

Chapter 1

Introduction

1.1 Road Traffic accidents

The first recorded traffic accident involving a motor vehicle took place in Ireland in August 1869, resulting in the death of a 42-year-old woman (Byrne, 2008)

Nowadays, more than a million people die on the road every year worldwide, with other 20-50 million people suffering from non-fatal injuries. Nevertheless, that is not the only downside of traffic accidents. They also incur substantial economic costs on families of the victims via not only direct costs such as treatment costs, material damage, or administration costs but also through indirect costs (loss of productivity, life cost, pain, etc.), which altogether costs countries between 1% and 3% of their gross national product (WHO et al., 2018).

Even though the majority (up to 93% according to WHO et al. (2018)) of traffic accident fatalities occur in low and mid-income countries, there was 929 680 traffic accidents in the Czech Republic alone between years 2011 and 2020 which resulted in 5 879 deaths within 24 hours after the accident with estimated material damage over 56 billion CZK, with yearly damage steadily rising. (Policie ČR, 2021)

In the year 2019 alone, there were 107 572 reported road traffic accidents in the Czech Republic, 80% of which were caused by a driver of a motor vehicle, 15% by animals, 2.7% by non-motor vehicles, 1% by pedestrians, and 1.3% by others, such as technical defects or road conditions. 4.3% of accidents were caused by drivers of motor vehicles under the influence of alcohol, accounting for 9.1% of the deaths. For cyclists, the number of traffic accidents under the influence of alcohol was even higher, over a quarter of the accidents they caused. (Policie ČR, 2021)

Chapter 2

Literature review

In this section, we will look at already existing literature regarding traffic accidents, mainly the ones that focus on the economic aspect and socio-economic explanatory variables.

2.1 Intercountry studies

Some of the studies about traffic accidents try to look at the whole world or compare countries with each other. One of the most important discoveries in this field, pointed out by van Beeck et al. (2000), was that fatalities of road traffic accidents are rising with the GDP of a country until a certain point, then they start lowering again, creating an "inverted U shape", which is caused by an increase in income leading in an increase in a number of vehicles, thereby more traffic accidents and fatalities. This economic growth, however, from a certain point transforms instead into an increase in medical infrastructure, more road safety awareness, and safer cars resulting in fewer fatal accidents. The same cannot be said about the number of traffic accidents or injuries which does not decline after a certain income point (Bishai et al., 2006)

Bishai et al. (2006) take panel data from 41 countries using information about road crashes, the number of vehicles, kilometers of roadway, oil consumption, population, alcohol consumption, and GDP to investigate why this phenomenon of increase in the number of fatalities to some point followed by decrease exists. They run multivariate regression, Ordinary least squared (OLS), fixed effects (FE), and random effects (RE) models. This decrease is assumed to be caused by investing in improving drivers' behaviour, using different types of vehicles, or improving trauma care.

Paulozzi et al. (2007) then try to examine country development and traffic accidents correlation for different types of road users, such as pedestrians, bicyclists, motorcyclists, users of 4 wheeled vehicles, and others. They use cross-sectional regression analysis of 44 countries. They found that motor vehicle occupants account for more deaths at GNIs per capita over $10 000 than all the other groups combined, which is not true until that point. This is primarily because of the decline in pedestrian death rates for higher-income countries.

Grimm and Treibich (2010) use OLS and a negative binomial regression model to find socio-economic determinants of road traffic accidents in low and middle-income countries. Even though it is challenging to find sufficient data samples in low-income countries, they identified medical supply, alcohol abuse, population density, and urbanisation as significant variables for explaining traffic fatality rates.

2.2 Intra country studies

Studies looking at traffic accidents from a socio-economic perspective within countries are Traynor (2008) in Ohio, USA and La Torre et al. (2007) in Italy and Jones et al. (2008) in England and Wales.

Other studies focus on within-country cost analysis of specific aspects of the road traffic accidents such as Connelly and Supangan (2006) in Australia, Al-Masaeid et al. (1999) in Jordan, or García-Altés and Pérez (2007) in Spain.

Traynor (2008) examines the effect of variables such as income, population density, or the number of teen drivers in 88 counties of Ohio on accident fatality rate per 100 million vehicle miles driven in the years 1999-2003. He performs cross-county multivariate regression analysis using Robust estimates to allow for heteroscedastic errors and finds several significant variables. The number of teen drivers, having a highway in the county, alcohol-related deaths[1], and Amish population were all positively correlated. On the other hand, population density, income, being rural county, and college campus presence were all correlated negatively. We will be using a similar analysis as Traynor (2008) with performing multivariate analysis on counties[2] of one state with the difference that he uses data scaled by vehicle miles driven rather than by the number of inhabitants or traffic accidents.

La Torre et al. (2007) focus on the number of traffic accidents and deaths per 100 000 people and the death/injured ratio in regions of Italy. They use socio-demographics (employment, urbanization, income), the prevalence of alcohol use, the number of vehicles, and availability of medical help within regions as variables. Using multivariate linear regression models they found a significant correlation between the employment rate and fatality and accident rates. For the rate of traffic accidents, they also indicate the significance of urbanization. The only variable significantly correlated with the case-fatality ratio was the number of circulating vehicles. This is another study similar to ours with performing multivariate linear regression on territorial units of one country.

Jones et al. (2008) introduce new possible explanatory variables of traffic fatalities such as curvature of the roads, hilliness, land use, and the expected number of casualties. They use multilevel negative binomial models to identify which factors predict variations in the number of fatalities and injuries across counties.

[1] as a replacement for alcohol abusive behaviour

[2] districts in our case

Thanks to their geographical approach, they identified new significant variables of the average percentage of roads going through urban areas and curvature of the roads.

Connelly and Supangan (2006) calculate the economic burden on sub-national level in Australia. They found that the costs vary quite remarkably between the Australian jurisdictions. They calculated three types of costs. Human costs (medical care, quality of life, loss of labour, etc.), vehicle costs (repair, vehicle unavailability, towing), and general costs (police, travel delays, etc.). Human costs accounted for 55% of total costs, vehicle costs 27%, and general costs 16.5%. Altogether traffic accident costs exceed 2% of Australia GDP, ranging between 0.62 up to 3.63 between states.

Al-Masaeid et al. (1999) tried to estimate the overall cost of traffic accidents in Jordan using not only national statistics but also data from insurance companies, private hospitals, and medical centers. For calculating the costs, they used an extended framework of Andreassend (1985), which involves a cost of death, cost of different levels of injury seriousness, property damage, cost of police, and administration costs all accounted for different severity of traffic accidents.

García-Altés and Pérez (2007) focused on only one particular city, Barcelona. They point out that all the costs are calculated towards the hypothetical alternative of having no road traffic accidents. The costs were divided into direct (healthcare costs, police, property damage, etc.) and indirect costs using the human capital approach calculated through lost working days, mean wages, occupation rates, and mortality. They found out that crashes in the cities (e.g. Barcelona) are less severe and less lethal than in the non-urban areas, but the costs are high, with property damage accounting for around 80% of all costs.

2.3 Alcohol-related traffic accidents

Some of the studies mentioned above used alcohol-related behaviour, such as number of people with cirrhosis as a variable (Traynor (2008); Grimm and Treibich (2010)) and found a significant relationship between drinking behaviour and the number of crashes.
Hingson et al. (2002) conducted an analysis on an individual level using a national survey of 42 862 respondents on drinking behaviour, drunk driving, and alcohol-related traffic accidents. They found that even after adjusting for alcohol dependence, people who started drinking at a younger age (especially before 21) were much more likely to be in an alcohol-related traffic accident.

Eisenberg (2003) then evaluated different drunk driving policies across all states of the USA between the years 1982 and 2000 using a weighted least squared model with fixed effects (state fixed effects, year fixed effects). They found a negative correlation between having stricter laws (0.08 BAC[3] opposing to 0.1 BAC or higher) and the number of fatal traffic accidents.

[3] 1 BAC = 10 permille

Chapter 3

Data

In this part, we will explain what data we chose for our models, how we processed them, and we will perform a descriptive analysis of the data. We can divide our data into categories to the data about traffic accidents, the district itself, and the people in the districts.

3.1 Definition of traffic accidents

The road traffic accident is defined in Zak (2016-02-20), but for the purpose of the study, we just have to know that the traffic accident is a crash or collision which was caused by the operation of a vehicle in motion.

Meaning that a traffic accident is not something where a person is killed, injured or a vehicle is damaged as a result of an event unrelated to crash or collision with another vehicle, pedestrian, animal, or solid obstacle, for example during repair of a vehicle or loading and unloading of cargo.

3.2 Traffic accidents data

For the traffic accidents data, we used the database of the Czech police (Policie ČR, 2021). We decided to use inputs from 2017 to 2020 because since 2017 the data consist of an analysis of information about every single traffic accident, and before that, only summary reports were issued. First, we had to consolidate data about all the years and regions together and from the identification number of each accident add information about region and district (the files were organised by region of the police department working on each case, but especially near borders of regions police of the other region might get involved. So we divided the accidents accordingly to the place where they occurred rather than by the police department working on them. We summed up all the years rather than use it as panel data because of high variations across the years, especially in the smaller districts.

After consolidating all the data, we filtered the information by district: number of traffic accidents, number of deaths, number of severe injuries, the ratio of accidents involving alcohol, the ratio of accidents of the police, and the number of traffic accidents and deaths caused by cars.

The data involve only traffic accidents reported to the police. We believe that

this is not a massive problem as those unreported are usually minor as it is illegal not to report traffic accidents involving injury or high property damage.

3.2.1 Demographic data

We gathered demographic data of all the Czech republic districts to explain some of the differences in traffic accidents data. From the Czech statistical office, we got data regarding population density, percentage of males and females, percentage of people who live in villages up to 1000 inhabitants, towns between 1000 and 5000 inhabitants, towns between 5 and 50 thousand inhabitants and cities over 50 thousand. We also got unemployment rate, percentage of people fifteen years old or older with a college education (finished bachelor degree or higher), percentage of people in productive age (between 15 and 64), percentage of wedded people and average retirement pension. Since the data for household income are not available on a district level, we decided to use retirement pensions instead. This is a valid substitution as we found an 84% correlation between these two variables on a regional level.

We were limited by the availability of the data as some variables that other relevant studies (Traynor (2008); Grimm and Treibich (2010)) showed to be important such as alcohol abuse were not available on a district level for the Czech Republic

3.2.2 Infrastructure data

For the infrastructure data, we used a database from Road and motorway directorate (Ředitelství Silnic a Dálnic, 2021) to get the length of the roads by type and the ministry of Transport to get the number of cars. (Ministry of Transport, 2020a) We also got the number of hospital beds by region as a measurement of the availability of hospital care.

3.3 Processing the data

To have the data consistent, we corrected irregularities between data sets, such as standardizing the name of the districts (some districts were named variously across data sets, such as Frýdek Místek/Frýdek-Místek). We also created two new statistical units, Praha-venkov consisting of Praha-východ and Praha-západ and Plzeň-venkov consisting of Plzeň-jih and Plzeň-sever, because some of the sources used this division and it would not be possible to determine how much to attribute to which district. We also added data about Prague, which does not belong to any district, to have the data about the Czech republic complete.

In the data regarding traffic accidents, we had to omit accidents that took place on a highway, as they were investigated by highway police, and it was impossible to identify the district in which they took place. On top of that, we believe that highways are not representative of the district, as there is much higher traffic out of other regions.

3.4 Variables

When choosing variables for the model, we considered relevant literature, mainly by Traynor (2008) and van Beeck et al. (2000), but we were limited by availability of information on a district level. We also added some more variables that we thought might be related to traffic accidents such as wedded and religious ratio, because we believe that people who are married or religious might have different driving behaviour.

3.4.1 Dependent variables

Being limited by the fact that there is no data on vehicle miles traveled, we could not use the same approach as Traynor (2008), and we decided to use an approach similar to La Torre et al. (2007), scaling the numbers regarding traffic to the number of people living in the district. We also created models where we scaled the numbers by the length of road infrastructure in the respective district. We decided to divide our dependent variables into three categories.

Traffic accident frequency

The first category includes variables explaining overall traffic accident frequency.

Number of accidents per inhabitant (accidentsinhratio) - Yearly average number of traffic accidents occurring in the district per 10000 inhabitants (except highway accidents)[1]

Number of accidents per 100 km of road (accidentsroadratio) - Yearly average number of traffic accidents occurring in the district per 100 kilometres of first, second and third class roads combined.[2]

The severity of the accidents

The second category includes variables associated with the severity of the accidents.

The number of car fatalities per inhabitant (carkills10k) - Yearly average number of car traffic fatalities occurring in the district per 10 000 inhabitants. Fatality is when a person dies within 30 days of a traffic accident. In this variable, we focus only on accidents that involved cars.

The number of fatalities (killedper10k) - Yearly average number of people dying due to traffic accidents per 100 000 people. Unlike in the previous variable, all traffic accidents, such as car, motorcycle, truck accidents, are included.

[1] in years 2017-2020
[2] in years 2017-2020

The ratio of fatalities (killedperacc) - Average number of people dying per traffic accident.[3]

Accident percentages

The third category describes data about particular events in traffic accidents and their frequency.

Alcohol-related accidents (alcoratio) - Percentage of traffic accidents that were caused by a driver under the influence of alcohol.

Police accidents (polaccratio) - Percentage of traffic accidents that involved police vehicle.

Traffic accidents caused by speeding (speedratio) - Percentage of traffic accidents where the main cause was excessive driving speed.

3.4.2 Explanatory variables

For explaining the variety of dependent variables described above, we decided to use the variables below (we did not always use all of them. For example, we expect the impact of the number of hospital beds on the number of fatalities of traffic accidents, but not on the number of accidents). We decided to divide them into four categories.

Macroeconomic Factors

These variables explain the macroeconomic situation in the districts.

Pension - Average retirement pension in each district as of 31.12.2020

Unemployment rate (Unemployedper100k) - Number of unemployed people per 100 000 inhabitants as of 1.1.2021

Education

This category describes the proportion of the population older than 15 years and their highest achieved education. The data are from the 2011 census of the Czech republic, as this was the only dataset, which covered education at a district level.

Basic or no education (basnoneeduratio) - Percentage of people older than 15 years who completed none or basic education.

[3] Within 30 days of traffic accident

Secondary education without A-level exam - Percentage of people older than 15 years who completed secondary education without A-level exam. We will not be using this variable to avoid a problem with multicollinearity.

Secondary education with the A-level exam - Percentage of people older than 15 years who completed secondary education with an A-level exam.

College ratio - Percentage of people who have completed bachelor studies or higher.

Demographic data

Population density (popdensity) - Average number of inhabitants in the years 2017-2020 divided by area of the district[4]

Male ratio (mfratio) - Percentage of males in the population[5].

Young ratio - Percentage of people aged 18-23 in the population.

Young male ratio (youngmfratio) - Percentage of males in the age group of 18 to 23 as of 2020

Wedded ratio - Percentage of people who are married. This data was available only for 2011, as in 2011, there was a census in the Czech Republic.

Religious ratio - The percentage of people indicating that they are members of a church in a census. This data was available only for 2011, as in 2011, there was a census in the Czech Republic.

Village ratio (village) - Percentage of people living in municipalities up to one thousand inhabitants as of 2020.

Cities over fifty thousand inhabitants (cityover50) - Percentage of people living in municipalities bigger than fifty thousand inhabitants as of 2020.

Infrastructure

Number of cars per 100 000 inhabitants (carsperinh) - Number of motor vehicles of category M1 and M1G, which are defined by the European Union as "Vehicles designed and constructed for the carriage of passengers and comprising no more than eight seats in addition to the driver's seat, and having a maximum mass ("technically permissible maximum laden mass") not exceeding 3.5 tons" (European Union, 2021).

[4] As of 31.12.2019

[5] Average in years 2017 to 2020

First-class roads (ftclassinh) - Metres of first-class roads per inhabitant[6]

Second-class roads (ndclassinh) - Metres of second-class roads per inhabitant[6]

Third-class roads (rdclassinh) - Metres of third-class roads per inhabitant[6]

Number of hospital beds (hospitalbeds_1000) - Number of hospital beds per 1000 inhabitants as of 2020

3.5 Descriptive analysis of the data

In this section, we will look at our data from a descriptive perspective. We will be using a five-number summary and standard deviation

3.5.1 Dependent variables

Table 3.1: Descriptive analysis of dependent variables

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
accidentsinhratio	85.90	27.58	43.98	66.84	98.8	170.01
carkills10k	0.38	0.19	0.08	0.25	0.46	0.95
killedper10k	0.54	0.24	0.14	0.40	0.63	1.31
killedperacc	0.01	0.004	0.001	0.004	0.008	0.022
alcoratio	0.06	0.021	0.02	0.04	0.07	0.11
polaccratio	0.97	0.45	0.11	0.64	1.27	2.29
speedratio	0.16	0.05	0.03	0.12	0.18	0.30

In the summary of our dependent variables, we can see that all of them vary quite a lot among regions with the lowest number of accidents per 10,000 inhabitants per year being 44 in Znojmo and 170 in Louny district (see map A.1), between 0.14 traffic fatalities per 10,000 a year in Jablonec nad Nisou and 1.31 in Strakonice. The district with, on average least fatal traffic accidents is Prague, with just over 0.001 average fatalities per accident while traffic accidents in Písek district are twenty times more deadly.

11% of traffic accidents in the Tachov district involve alcohol, while in Prague, it is less than one in 50 accidents (see map A.2). Speeding as the cause of the accident has the highest portion in Prachatice with over 30% and lowest in Přerov with less than 3%. Meanwhile, police are involved in the least crashes in Prachatice with one in 10 000 crashes and highest in Sokolov with 66 crashes, which is over 2% of traffic accidents.

[6]as of 1.1.2021

3.5.2 Explanatory variables

In the table below, we can see all our independent variables. We have 75 observations for each, one for each district plus Prague. Interestingly enough, there is a district with no hospital beds (Tachov) and two with no first-class roads (Beroun and Sokolov). We can also see, that most of the districts does not have a city with over 50 000 inhabitants, as there are only sixteen, which do. We can see high extremes in population density and metres of third-class roads, which is caused by the capital city of Prague. As expected, there is no vast variance in the percentage of gender ratio, young people and the gender ratio of young people, but we believe that 1%, 14%, and 3% change respectively in population structure might have a significant influence on traffic accidents, as there are differences in age and gender driving behavior (Rhodes and Pivik, 2011)

Table 3.2: Descriptive analysis of the data

Variable	Average	Min	1st qtl	3rd qtl	Max	SD[a]
popdensity	199.44	37.03	75.42	153.75	2669.91	368.6
mfratio	49.43	48.44	49.17	49.68	50.55	0.42
Village	22.56	0	15.98	30.59	42.9	10.53
Cityover50	12.15	0	0	0	100	25.44
Pension	14309	13395	14100	14461	15376	330.9
Basnoneeduratio	19.32	10.53	18.08	20.64	25.57	2.5
Grad	30.56	25.65	28.75	31.57	56.88	3.7
Collegeratio	9.96	5.54	8.852	10.78	22.48	2.6
Weddedratio	42.2	35.26	40.98	43.72	45.68	2.38
Religiousratio	13.56	4.2	7.04	17.58	37.49	7.4
Unemploedper100k	2705	1257	2143	3094	5471	872.3
Young	5.5	4.69	5.53	5.72	6.10	0.29
Youngmfratio	51.52	49.94	51.09	52.03	52.92	0.67
carsperinh	0.4884	0.3696	0.449	0.5207	0.6279	0.048
ftclassinh	0.7169	0	0.4744	0.9601	1.7391	0.392
ndclassinh	1.9236	0.0225	1.0165	2.8543	5.0976	1.146
rdclassinh	5.05	0.00003	2.797	6.146	56.7488	6.426
hospitalbeds_1000	5.164	0	4.167	5.841	11.809	2.014

[a] Standard deviation

On average, there is almost seven times larger third-class road network than first class. There are also, on average more than twice as many roads of third-class than second-class. College ratio, wedded ratio, and religious ratio vary quite significantly across districts, with a range of 4% in Most up to 37.5% of people identifying as religious in Uherské-Hradiště, 10% difference between least and most wedded districts (Most and Pelhřimov respectively) and more than four times higher percentage of people with a college degree in the city of Prague than in Sokolov district. Sokolov district is also the one with the least amount of people whose highest achieved education is an A-level exam (25%), while Prague is the highest with 56%, followed by 34% in Brno.

Chapter 4

Methodology and models

In this section, we will introduce the models we will need to test our hypotheses, explain why we chose them, test the assumptions of the models and describe the methodology we followed.

4.1 Choosing the right method

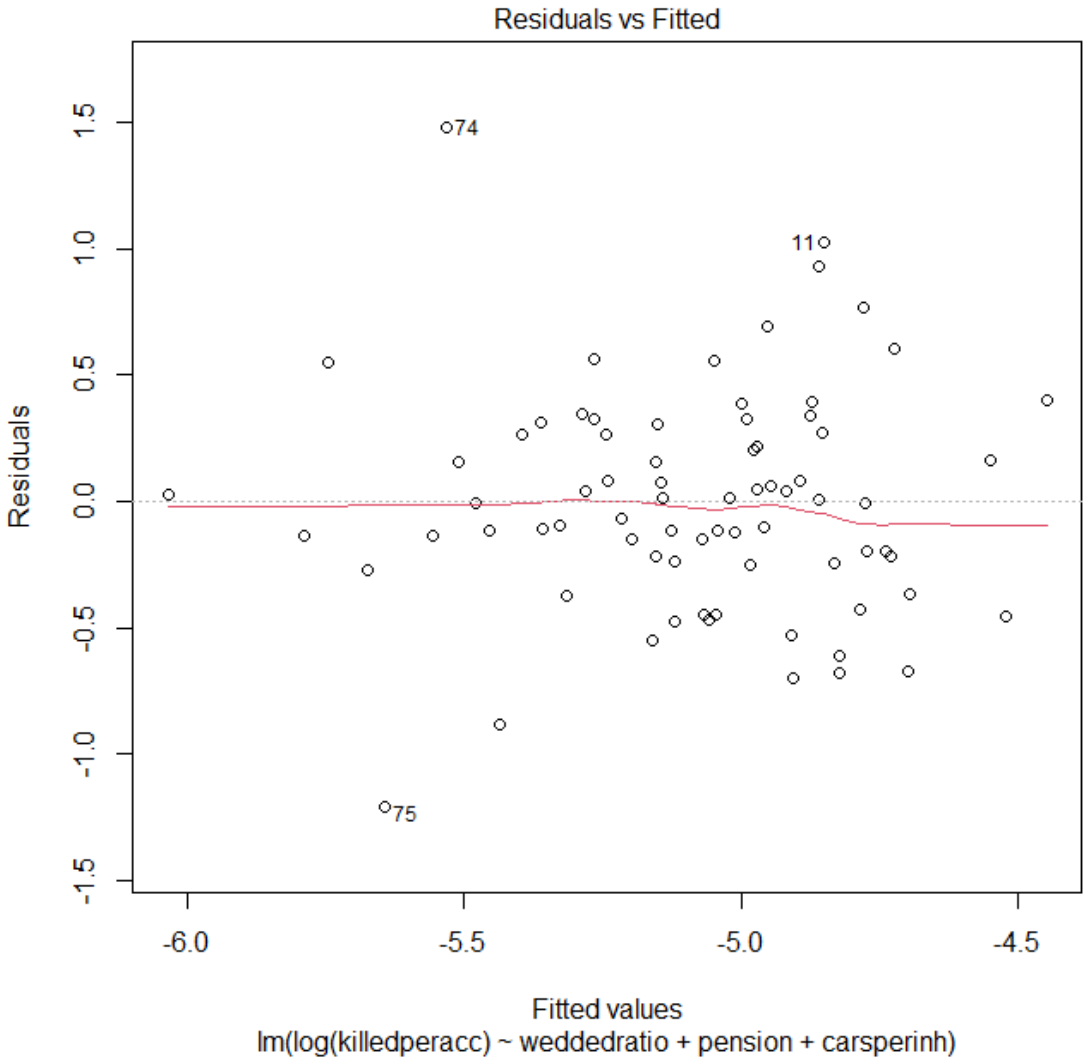
When choosing the method for the analysis, we based our selection on the characteristic of the data available. Similar to Traynor (2008) and La Torre et al. (2007), we used the average number of accidents per year rather than panel data. For this type of cross-county data, we will use multiple linear regression analysis. According to the Gauss-Markov theorem, the best linear unbiased estimate (BLUE) if certain assumptions are met is the ordinary least squared model (OLS). The assumptions are linearity of the parameters, random samples, non-collinearity, exogenity, and homoskedasticity, and we will test them later in our work. On top of that, we will test for normality of the errors, which might make OLS Maximum likelihood estimator (MLE).

4.2 Assumptions

To be sure that OLS is BLUE, we used R-studio to test the assumptions for all the models

Linearity

To test linearity, we use the Residuals vs. fitted plot. When the line approximately follows 0, we can assume linearity of the parameters.



A random sampling of the observations

In a truly random sample, every individual has the same chance of being picked up. Our dataset should consist of data about the whole population of the Czech republic, meaning that everyone has the same chance of being picked as an observation (100%). Problems with random sampling might occur in connection with data being incomplete if some group of people tends to underreport their data more than others. However, we do not assume this to have a significant effect.

Non-collinearity

To test for the presence of collinearity between independent variables, we created a correlation matrix.

When we evaluate all the considered independent variables together, we find a strong correlation between being **wedded** and **religious** (0.61), **wedded** and **unemployed** (-0.61), **village** and **second class road** (0.75), **college** and **city**

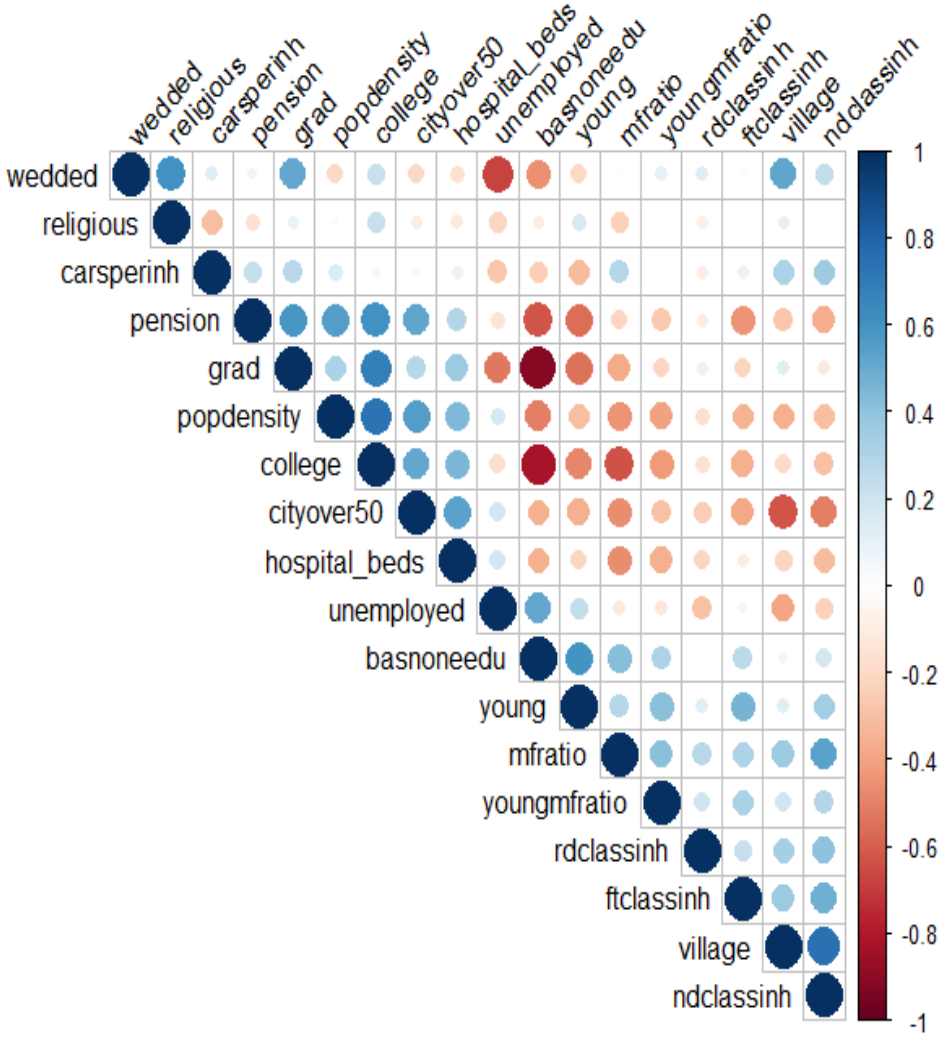


Figure 4.1: Correlation Matrix

over fifty thousand inhabitants (0.63) and also between different types of completed education (0.76 and -0.81 between **grad** and **Collegeratio** and **basnoneeduratio** respectively).

We then run the Variance Inflation Factor to test the multicollinearity further. Even though none of the values exceeds 10, which is often assumed as being the threshold for multicollinearity (O'brien, 2007), some of the values exceeded 5, which is sometimes defined as insufficient (Henseler et al., 2015). However, we will be using general to specific method for the models, which will lead to deleting some of the variables, so we will run a collinearity test for each model later.

Homoskedasticity

To test for homoskedasticity assumption, we run the Breusch-Pagan test (see table A.1) The null hypothesis for the test is "model is homoskedastic," and we will be using threshold of $p<0.05$ to reject the null hypothesis.

Normality

To test the normality of the errors, we will use the Shapiro-Wilk test (see table A.1) In the Shapiro-Wilk test, the null hypothesis is "The errors are normally distributed" and we, again, will be using the threshold of $p<0.05$ to reject the null hypothesis. If this assumption does not hold, we will still use OLS because it remains BLUE.

4.2.1 Models

Based on the relevant literature (Traynor, 2008) (La Torre et al., 2007) with adding some other variables we thought might have explanatory power, we came up with models in form of

$$dependentvariable = \beta_0 + \beta_1 * macroeconomicfactors + \beta_2 * demographicdata + \beta_3 * infrastructure + \varepsilon.$$

All the variables considered for the model are described above in subsection 3.4.2.

To finalize the models, we used a general to specific approach (GtS), first described by Hoover and Perez (1999) which consists of discarding insignificant variables one by one while keeping the model congruent until no variable can be discarded. All the specific models will be introduced in chapter 5.

Chapter 5

Results

In this chapter, we will present the results of the models and find a possible explanation why these variables, in particular, might be significant in predicting the results. We will also present formulas for every model which we used.

5.1 Frequency of the traffic accidents

We composed two models to predict the frequency of traffic accidents in each district. Accidents per inhabitant and accidents per road length.

Accidents per inhabitant

After following general to specific method of the model and arriving at the form

$$log(accidentsinhratio) = \beta_0 + \beta_1 * pension + \beta_2 * wedded + \beta_3 * ftclassinh + \varepsilon$$

we test the assumptions mentioned above. With residuals vs. fitted graph line no significantly different from the horizontal line around 0, we can assume linearity (see graphA.3). The Breusch-Pagan test confirmed homoskedasticity of the model with the p-value of 0.195, and normality is tested by the Shapiro-Wilk test with the p-value of 0.616 (see table A.1). No collinearity assumption was verified by the VIF not being higher than 1.26 (see table A.2). After testing all the assumptions, we can perform linear regression.

Accidents per road length

After following general to specific method of the model and arriving at the form

$$accidentsroadratio = \beta_0 + \beta_1 * popdensity + \beta_2 * pension + \beta_3 * wedded + \beta_4 * carsperinh + \varepsilon.$$

we test the assumptions mentioned above. As residuals vs. fitted graph line being slightly different from the horizontal line around 0, we cannot be sure about the assumption of linearity (see graph A.4). However, this is only a visual test, and we will perform analysis nevertheless. The Breusch-Pagan test confirmed the homoskedasticity of the model with the p-value of 0.14 (see table A.1). No

collinearity assumption was verified by the VIF not being higher than 1.5 (see table A.3). However, the Shapiro-Wilk test detected non-normality of the errors with the p-value 6.63×10^{-7} (see table A.1)

Table 5.1: Frequency of the traffic accidents

	Dependent variable:	
	log(accidentsinhratio)	log(accidentsroadratio)
	(1)	(2)
popdensity		0.002***
		(0.0002)
pension	0.0003***	0.001***
	(0.0001)	(0.0002)
weddedratio	−0.064***	−0.095***
	(0.012)	(0.023)
ftclassinh	0.181**	
	(0.084)	
carsperinh	0.585	−2.274**
	(0.640)	(1.109)
Constant	2.522*	1.354
	(1.484)	(2.658)
Observations	75	75
R^2	0.340	0.818
Adjusted R^2	0.303	0.808
Residual Std. Error (df = 70)	0.249	0.440
F Statistic (df = 4; 70)	9.024***	78.737***
Note:		*p<0.1; **p<0.05; ***p<0.01

In the first model, the R^2 for the estimated equation is 0.347, with the residuals standard error being 0.249. All our variables are significant thanks to GtS approach.

Pension, which is a replacement for missing variable income, has a positive effect on the number of traffic accidents. This finding is in line with a study from La Torre et al. (2007), even though their results for income were not statistically significant. Approximately 33 CZK increase in pension leads to a 1% increase in traffic accidents. This might be explained in several ways. People with growth in income tend to break more rules (Brown, 2012), which may lead to more traffic accidents. Another explanation might be that having more money, thus better car leads to riskier behaviour because people feel and are safer. Moreover, the Czech Republic uses a fixed level of fines, meaning that the fine represents a

smaller part of their wealth for wealthier people than for those with less money. Another significant variable with a positive effect is meters of first-class roads per inhabitant, with every meter per inhabitant resulting in a 20% increase in traffic accidents.

Variable with negative effect is the portion of wedded people with a marginal increase of 1% resulting in 6.5% fewer traffic accidents. This might be a result of lower risk-taking behavior by people with families because they either are more cautious or drive more with other members of a household, resulting in safer driving behavior or even by married people spending more time home, resulting in fewer kilometres on the road and thus fewer traffic accidents.

Even though the second model did not the satisfy normality assumption, it is still consistent and effective[1]. On top of significant variables in the first model the model showed a significant positive correlation with population density, which makes sense because the more people are in the area, the more people are using the roads and thus more people are involved in traffic accidents. What is surprising is the negative effect of cars per inhabitant. As the number of vehicles per inhabitant ranges from 37 to 63%, and there is less than 60% of people owning a driving license (Ministry of Transport, 2020b), and not everyone owning a driving license drives on a regular basis increase in the number of cars per inhabitant might mean more people with more than one car rather than more people with a car, which might be an indicator of more experienced drivers, resulting in fewer traffic accidents.

5.2 The severity of the traffic accidents

Traffic accident fatality

After following the GtS method we have come to the formula

$$killedper10k = \beta_0 + \beta_1 * popdensity + \beta_2 * carsperinh + \beta_3 * ftclassinh + \beta_4 * rdclassinh + \varepsilon.$$

and tested the assumptions. The linearity assumption holds, as seen from the residuals vs. fitted graph (see graph A.5), homoscedasticity is also satisfied with the p-value in the Breusch-Pagan test being 0.88 (see table A.1) and no multicollinearity was detected with VIF values not exceeding 1.2 (see table A.5). However, the Shapiro-Wilk test detected non-normality in the errors with the p-value of 0.003 (see table A.1).

Car fatalities per inhabitant

For the model involving only deaths caused by traffic accidents, we included one more significant variable, which is the percentage of people living in villages.

$$carkills10k = \beta_0 + \beta_1 * popdensity + \beta_2 * village + \beta_3 * carsperinh + \beta_4 * ftclassinh + \beta_5 * grad + \varepsilon$$

[1] If we assume that linearity holds

All the assumptions hold with residuals vs. fitted graph line being horizontal at 0 (see graph A.6), p-value in Shapiro-Wilk test being 0.165 and in Breusch-Pagan test 0.613 (see table A.1).

Table 5.2: Severity of the accidents

	Dependent variable:	
	log(killedper10k)	log(carkills10k)
	(1)	(2)
popdensity	−0.001***	−0.001***
	(0.0001)	(0.0001)
village		0.008*
		(0.005)
carsperinh	3.393***	3.718***
	(0.723)	(0.965)
ftclassinh	0.334***	0.253**
	(0.094)	(0.119)
rdclassinh	0.014***	
	(0.005)	
Constant	−2.571***	−3.145***
	(0.352)	(0.441)
Observations	75	75
R^2	0.573	0.526
Adjusted R^2	0.549	0.499
Residual Std. Error (df = 70)	0.290	0.361
F Statistic (df = 4; 70)	23.508***	19.403***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Both our models assign a significantly negative effect to population density, meaning that when population density increases by ten, the number of people dying from traffic accidents per 10 000 inhabitants will decrease by 1%. This relationship was firstly described by Baker et al. (1987) and some of the explanations might be higher severity of traffic accidents in less populated areas[2] and higher distance from trauma centers, leading to less effective treatment. (Clark, 2003)

A significant positive effect was associated with three variables. The number of cars per inhabitant, metres of first-class, and metres of third-class roads. When

[2]e.g. due to higher speed etc.

we account only for deaths caused by cars, the third class roads are replaced by the percentage of people living in villages. The change for one more car per inhabitant and *ftclassinh* seems high, but we have to remember that they range from 0.48 to 0.62 and 0.7 to 1.7, respectively. All these results are in line with the findings of La Torre et al. (2007). However, their results showed no significance. The number of cars per inhabitant and length of the road are also significant and positive in a study of Jones et al. (2008). *Carsperinh* remain significant even when we account only for deaths during traffic accidents involving only one car (see table A.12), so we believe that explanation for this is that with more cars per inhabitant there is higher traffic, thus more traffic accidents overall rather than more crashes of two or more cars because there are more cars.

Fatalities per accident

After GtS method implementation we ended up with model

$$killedperacc = \beta_0 + \beta_1 * wedded + \beta_2 * ndclassinh + \beta_3 * pension + \varepsilon.$$

When omitting variable *ndclassinh* variable *carsperinh* is significant. However, this model is heteroscedastic with Breusch-Pagan test p-value of 0.04 (see table A.1) and lower R^2 (0.32) than when we included both *carsperinh* and *ndclassinh* the R^2 improves to 0.42 with all assumptions of OLS met. (Linearity confirmed by fitted vs. residuals graph (see graph A.8), Breusch-Pagan test p-value 0.23, VIF values no higher than 1.6 (see table A.7)and Shapiro-Wilk test p-value 0.15(see table A.1)).

Table 5.3: Fatalities per accident

	Dependent variable:	
	log(killedperacc)	
	(1)	(2)
weddedratio	0.090***	0.070***
	(0.022)	(0.022)
carsperinh	2.549**	0.531
	(1.145)	(1.205)
pension	−0.001***	−0.0003*
	(0.0002)	(0.0002)
ndclassinh		0.193***
		(0.055)
Constant	−1.101	−4.183*
	(2.418)	(2.406)
Observations	75	75
R^2	0.320	0.423
Adjusted R^2	0.291	0.390
Residual Std. Error	0.455 (df = 71)	0.422 (df = 70)
F Statistic	11.140*** (df = 3; 71)	12.841*** (df = 4; 70)
Note:		*p<0.1; **p<0.05; ***p<0.01

A significant explanatory variable with a negative effect is *pension*, which might be result from owning safer cars with increased income. This connection was earlier described by many studies (van Beeck et al. (2000); Bishai et al. (2006)).
A significant positive effect was found for *weddedratio* and *ndclassinh*. One possible explanation for the correlation of *weddedratio* is that more people ride in cars of wedded people. Thus there is a higher likelihood of one or more people dying when involved in an accident. However, when accounting only for accidents where single person died, the variable *weddedratio* was still significant (see table A.13). We can argue that just one person per 100 accidents dies on average, so it is still likely that more people in the car will result in more deaths of just one person within the accident. On top of that, some studies, such as Durbin et al. (2015), argue that in some cases rear row might be deadlier than the front row during a crash, which might multiply the effect of more people in the car.

5.3 Specific types of accidents

For this section, we picked special types of traffic accidents and tried to find significant variables explaining differences in proportion between the districts.
We choose these three models

The proportion of alcohol-related accidents

To explain differences in drunk-crashing in different districts of the Czech Republic, we narrowed down the variables using GtS method resulting in model

$$alcoratio = \beta_0 + \beta_1 * village + \beta_2 * pension + \beta_3 * weddedratio + \beta_5 * ndclassinh + \varepsilon.$$

Assumption of linearity (see graph A.9), no-collinearity, and normality holds (see table A.1). However, the Breusch-Pagan test is shows heteroscedasticity in the results with the p-value of 0.001 (see table A.1). This makes our OLS results biased, thus making another model more optimal.
First, we will use the heteroscedasticity-corrected covariance matrices method (see table A.11) described by Ferrari and Cribari-Neto (2004), which we will then use in calculating robust standard error t-test to test our data.

The proportion of accidents involving a police vehicle

Model, which is trying to explain inter-district differences in the percentage of traffic accidents involving police. After using GtS model the resulting formula is

$$policeratio = \beta_0 + \beta_1 * village + \beta_2 * pension + \beta_3 * ftclassinh + \beta_4 * crimes10kinh + \varepsilon.$$

We can use the OLS model because all the assumptions hold. For linearity residuals vs. fitted graph is not significantly different from 0 (see graph A.10), Breusch-Pagan's test p-value to test heteroscedasticity is 0.556, Shapiro-Wilk test of normality p-value is 0.27 (see table A.1), and the highest VIF value is 2 (see table

A.9), meaning no multicollinearity is present.

The proportion of accidents that were caused by speeding

For the model of the percentage of accidents that were caused by speeding, the GtS method led to a formula

$$speedratio = \beta_0 + \beta_1 * young + \beta_2 * village + \beta_3 * ndclassinh + \beta_4 * ftclassinh + \varepsilon.$$

The assumptions were tested as in previous cases, and all of them hold with Breusch-Pagan test p-value being 0.16, Shapiro-Wilk test p-value 0.56 (see table A.1), VIF values not higher than 2 (see table A.10), and residuals vs. fitted graph line being approximately horizontal to 0 (see graph A.11).

Table 5.4: Percentages of accidents

	Dependent variable:		
	alcoratio	polaccratio	speedratio
	coefficient test	*OLS*	*OLS*
	(1)	(2)	(3)
young			-0.052^{***}
			(0.019)
village	-0.001^{**}	-0.017^{***}	-0.001^{**}
	(0.0004)	(0.006)	(0.001)
pension	-0.00002^{***}	-0.0004^{***}	
	(0.00001)	(0.0001)	
weddedratio	0.004^{***}		
	(0.001)		
ndclassinh	0.009^{**}		0.033^{***}
	(0.005)		(0.007)
ftclassinh		-0.308^{**}	0.032^{**}
		(0.122)	(0.015)
Crimes10kinh		0.001^{*}	
		(0.0003)	
Constant	0.215^{*}	7.098^{***}	0.390^{***}
	(0.110)	(2.062)	(0.102)
Observations		75	75
R^2		0.415	0.376
Adjusted R^2		0.381	0.341
Rsdl. Std. Error (df = 70)		0.354	0.042
F Statistic (df = 4; 70)		12.405^{***}	10.553^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

We can see from the model that percentage of villages and *pension* negatively affect the portion of traffic accidents caused under influence of alcohol. The former might be caused, for example, by more lively nightlife in the cities, which is associated with drinking alcohol, and the latter by lower alcohol abuse among wealthier people. Even though studies have shown that alcohol use is positively correlated with income, it is true only for moderate drinkers (Heien (1996); Cerdá et al. (2011)). The likelihood of alcohol abuse is correlated negatively. And we believe that the group with the highest risk of drinking and crashing under influence of alcohol are people with alcohol abusive behaviour.
On the other hand, the effect of percentage of wedded people and metres of second class roads per inhabitant is positive, indicating that there might be a higher likelihood of drunken drive among married people, which is surprising as studies show, that married people engage least in alcohol abusive behaviour (Power et al., 1999). The significance of meters of second-class roads might indicate that people under the influence of alcohol use this type of road the most.

Independent variables that were found to be significantly negatively correlated with *polaccratio* are *village*, possibly as a result of fewer police officers in smaller municipalities, *pension*, and *ftclassinh*.

As expected, the variable *Crimes10kinh* is positively correlated with *polaccratio*, which is most likely the result of police solving crimes and chasing criminals. R^2 of the model is 0.415.

The first variable that is statistically significantly negative in *speedratio* model is *young*, which firstly might lead to an idea that young people are less likely to be involved in a speeding accident crash. Studies have shown that inexperience, especially among young drivers who just got their driving license, plays an important role in the likelihood of being in a traffic accident (McCartt et al., 2009). We believe that inexperience might influence other causes of traffic accidents, such as the way of driving more than those caused by speeding. The second statistically significantly negative variable is *village*, while positive explanatory variables are *ndclassinh* and *ftclassinh*, which might be caused by longer roads leading into the faster ride, thus more speeding crashes. However, *rdclassinh* was not significant, maybe because it is hard to achieve high speed on such types of roads. R^2 of the model is 0.376.

Chapter 6

Conclusion

The focus of this study was to partly uncover which variables cause differences in traffic accidents across districts of the Czech republic. We used multiple socio-demographic explanatory variables such as population density, road length, or level of education to explain some of the differences in the three areas. The Number of accidents, the severity of the accidents, and specific types of accidents.

We used the summary of the traffic accident data from the years 2017-2020 from Policie ČR (2021) for better accuracy and used them alongside data from Český statistický úřad (2021) in OLS models to find which variables significantly correlate with the traffic accidents in different districts. For one model, the assumptions were not met, and we used a robust standard error test instead.

Some of the results confirmed findings from previous studies, such as that income[1] leads to less tragic traffic accidents[2]. However, we found that it also leads to more traffic accidents. One possible explanation is that people feel, and are usually safer in more expensive cars, thus drive more dangerously. Another interesting finding was that number of people being wedded is negatively correlated with the number of traffic accidents but positively correlated with the average number of fatalities per traffic accident. Possibly because married people drive safer but more people ride in a car of a married person.

In specific types of accidents, we were looking at accidents caused under the influence of alcohol, accidents of police vehicles, and accidents caused by speeding. We, surprisingly, found a negative correlation between the percentage of accidents caused by speeding and the portion of young people among the population. We also found a positive correlation between the percentage of people wedded and the percentage of alcohol-related traffic accidents. The police tend to crash less often in districts with a higher share of people living in villages, in poorer districts, and districts with a high length of first-class roads per inhabitant. On the other hand, unsurprisingly higher crime rate positively correlates with police traffic accidents.

There are several limitations to this study. One of them is omitted variable

[1]Income was not available and was substituted by pension, which has an 84% correlation with income

[2]The average number of people dying per traffic accident

bias, where we had to left out important variables that could play a significant role in the models because the data were not available at a district level. It can be, for example, the number of kilometres traveled by car in every district, road conditions, information about the driver such as sex, age, driving experience, and others. This limitation leads to lower R^2 in our models and, to some extent, might make our predictions not completely accurate as the omitted variable may shift the results. Another limitation that can lead to a bias is underreporting of traffic accidents. However, the traffic accidents which remain unreported are usually minor as it is illegal not to report traffic accidents involving injury or high property damage.

Bibliography

Zákon o provozu na pozemních komunikacích 361/2000 sb. *sbírka zákonů ČR*, 2016-02-20.

Hashem R Al-Masaeid, Adel A Al-Mashakbeh, and Abdalla M Qudah. Economic costs of traffic accidents in jordan. *Accident Analysis & Prevention*, 31(4): 347–357, 1999.

DC Andreassend. A framework for costing accidents and accident types. *Accident Analysis & Prevention*, 17(2):111–117, 1985.

Susan P Baker, RA Whitfield, and Brian O'neill. Geographic variations in mortality from motor vehicle crashes. *New England Journal of Medicine*, 316(22): 1384–1387, 1987.

David Bishai, Asma Quresh, Prashant James, and Abdul Ghaffar. National road casualties and economic development. *Health economics*, 15(1):65–81, 2006.

Eryn Brown. Wealthy, motivated by greed, are more likely to cheat, study finds. *The Independent*, 2012. URL https://www.latimes.com/nation/la-xpm-2012-feb-27-la-sci-0228-greed-20120228-story.html.

Gay Byrne. 'the first person killed by a car was irish. mary ward was killed in offaly in 1869 when she fell out of a steam-powered car'. *The Independent*, 2008. URL https://www.independent.ie/woman/celeb-news/the-first-person-killed-by-a-car-was-irish-mary-ward-was-killed-in-offaly-i html.

Magdalena Cerdá, Vicki D Johnson-Lawrence, and Sandro Galea. Lifetime income patterns and alcohol consumption: Investigating the association between long-and short-term income trajectories and drinking. *Social science & medicine*, 73(8):1178–1185, 2011.

David E Clark. Effect of population density on mortality after motor vehicle collisions. *Accident Analysis & Prevention*, 35(6):965–971, 2003.

Luke B Connelly and Richard Supangan. The economic costs of road traffic crashes: Australia, states and territories. *Accident Analysis & Prevention*, 38 (6):1087–1093, 2006.

Dennis R Durbin, Jessica S Jermakian, Michael J Kallan, Anne T McCartt, Kristy B Arbogast, Mark R Zonfrillo, and Rachel K Myers. Rear seat safety: variation in protection by occupant, crash and vehicle characteristics. *Accident Analysis & Prevention*, 80:185–192, 2015.

Daniel Eisenberg. Evaluating the effectiveness of policies related to drunk driving. *Journal of policy Analysis and management*, 22(2):249–274, 2003.

European Union. Vehicle categories, 2021. URL https://ec.europa.eu/growth/sectors/automotive/vehicle-categories_en.

Silvia Ferrari and Francisco Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7):799–815, 2004.

A García-Altés and Katherine Pérez. The economic cost of road traffic crashes in an urban setting. *Injury prevention*, 13(1):65–68, 2007.

Michael Grimm and Carole Treibich. Socio-economic determinants of road traffic accident fatalities in low and middle income countries. *ISS Working Paper Series/General Series*, 504:1–44, 2010.

Dale Heien. The relationship between alcohol consumption and earnings. *Journal of studies on alcohol*, 57(5):536–542, 1996.

Jörg Henseler, Christian M Ringle, and Marko Sarstedt. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, 43(1):115–135, 2015.

Ralph Hingson, Tim Heeren, Suzette Levenson, Amber Jamanka, and Robert Voas. Age of drinking onset, driving after drinking, and involvement in alcohol related motor-vehicle crashes. *Accident Analysis & Prevention*, 34(1):85–92, 2002.

Kevin D Hoover and Stephen J Perez. Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *The econometrics journal*, 2(2):167–191, 1999.

Andrew P Jones, Robin Haynes, V Kennedy, IM Harvey, T Jewell, and D Lea. Geographical variations in mortality and morbidity from road traffic accidents in england and wales. *Health & place*, 14(3):519–535, 2008.

Giuseppe La Torre, Ed Van Beeck, Gianluigi Quaranta, Alice Mannocci, and Walter Ricciardi. Determinants of within-country variation in traffic accident mortality in italy: a geographical analysis. *International journal of health geographics*, 6(1):1–8, 2007.

Anne T McCartt, Daniel R Mayhew, Keli A Braitman, Susan A Ferguson, and Herbert M Simpson. Effects of age and experience on young driver crashes: review of recent literature. *Traffic injury prevention*, 10(3):209–219, 2009.

Ministry of Transport. Number of vehicles, 2020a. URL https://www.mdcr.cz/Statistiky/Silnicni-doprava/Centralni-registr-vozidel.

Ministry of Transport. Number of vehicles, 2020b. URL https://www.mdcr.cz/Ministerstvo/Zadost-o-poskytnuti-informace-(1)/Poskytnute-informace/Pocet-ridicskych-prukazu.

Robert M O'brien. A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, 41(5):673–690, 2007.

Leonard J Paulozzi, George W Ryan, Victoria E Espitia-Hardeman, and Yongli Xi. Economic development's effect on road transport-related mortality among different types of road users: a cross-sectional international study. *Accident analysis & prevention*, 39(3):606–617, 2007.

Policie ČR. Statistika nehodovosti, 2021. URL https://www.policie.cz/clanek/statistika-nehodovosti.

Chris Power, Bryan Rodgers, and Steven Hope. Heavy alcohol consumption and marital status: disentangling the relationship in a national study of young adults. *Addiction*, 94(10):1477–1487, 1999.

Nancy Rhodes and Kelly Pivik. Age and gender differences in risky driving: The roles of positive affect and risk perception. *Accident Analysis & Prevention*, 43 (3):923–931, 2011.

Thomas L Traynor. Regional economic conditions and crash fatality rates–a cross-county analysis. *Journal of safety research*, 39(1):33–39, 2008.

Eduard F van Beeck, Gerard JJ Borsboom, and Johan P Mackenbach. Economic development and traffic accident mortality in the industrialized world, 1962–1990. *International journal of epidemiology*, 29(3):503–509, 2000.

WHO et al. Global status report on road safety 2018: Summary. Technical report, World Health Organization, 2018.

Český statistický úřad. Czso, 2021. URL https://www.czso.cz/csu/czso/statistiky.

Ředitelství Silnic a Dálnic. Délky a další data komunikací, 2021. URL https://www.rsd.cz/wps/portal/web/Silnice-a-dalnice/delky-a-dalsi-data-komunikaci.

Appendix A

Tables

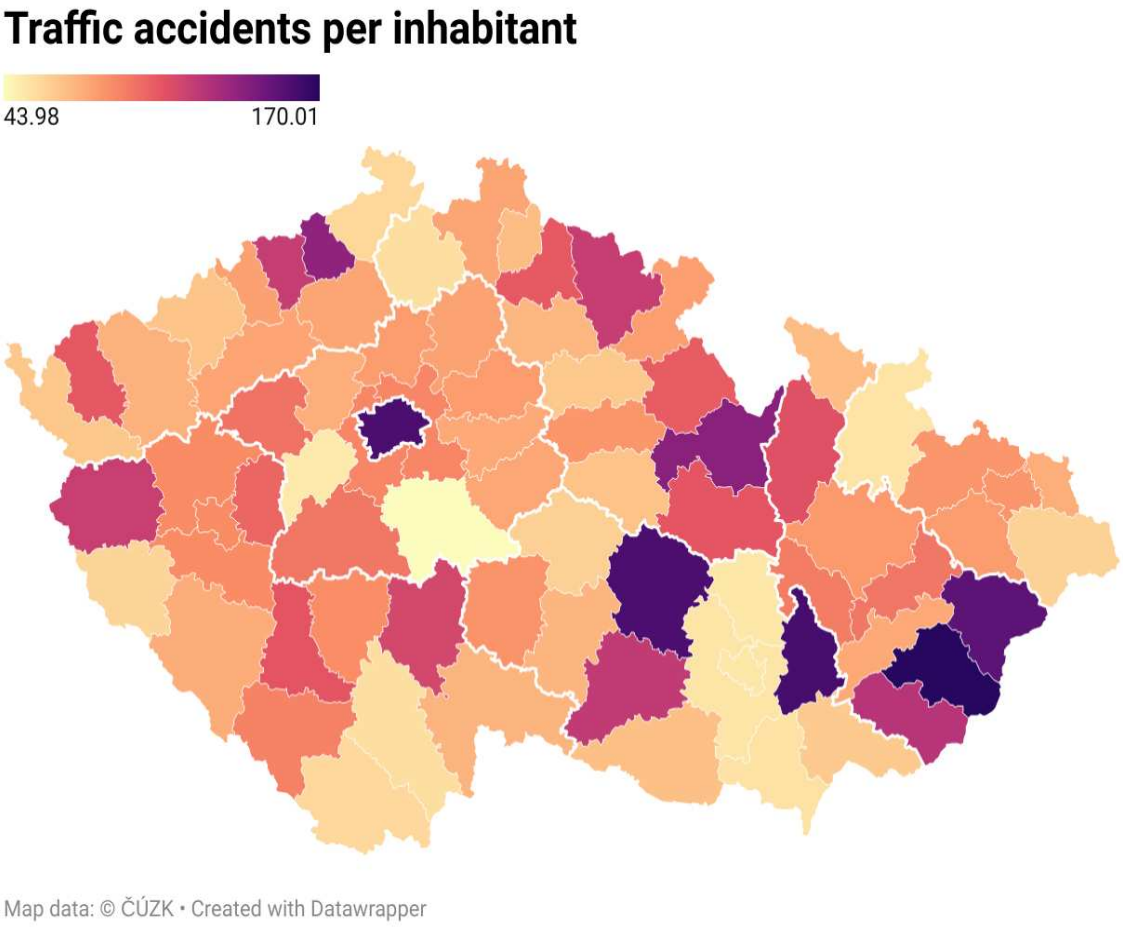


Figure A.1: Traffic accidents per inhabitant

Table A.1: The p-values for the different tests

Model	Shapiro-Wilk Test	Breusch-Pagan Test
accidentsinhratio	0.616	0.195
accidentsroadratio	6.63×10^{-7}***	0.144
killedper10k	0.003**	0.878
carkills10k	0.165	0.613
killedperacc	0.353	0.039*
killedperacc2	0.149	0.231
alcoratio	0.778	0.0001***
policeratio	0.266	0.556
speedratio	0.556	0.160
Note:		*p<0.05; **p<0.01; ***p<0.001

Table A.2: VIF model accinhratio

pension	weddedratio	ftclassinh
1.255	1.009	1.249

Table A.3: VIF model accroadratio

popdensity	pension	weddedratio	carsperinh
1.584	1.554	1.123	1.075

Table A.4: VIF model carkills10k

popdensity	village	carsperinh	ftclassinh
1.330	1.436	1.206	1.236

Table A.5: VIF model kills10kinh

popdensity	carsperinh	ftclassinh	rdclassinh
1.187	1.053	1.207	1.077

Table A.6: VIF model killedperacc 1

weddedratio	pension	carsperinh
1.019	1.058	1.072

Table A.7: VIF model killedperacc 2

weddedratio	carsperinh	pension	ndclassinh
1.100	1.381	1.429	1.621

Table A.8: VIF model alcoratio

village	pension	weddedratio	ndclassinh
3.203	1.206	1.564	2.505

Table A.9: VIF model polaccratio

village	pension	ftclassinh	Crimes10kinh
1.997	1.327	1.361	1.938

Table A.10: VIF model speedratio

young	village	ftclassinh	ndclassinh
1.358	2.347	1.517	2.709

Table A.11: HCCM for modelalco

	(Intercept)	village	pension	weddedratio	ndclassinh
(Intercept)	0.012	0.00001	−0	−0.00004	−0.0002
village	0.00001	0.00000	−0	−0.00000	−0.00000
pension	−0.00000	−0	0	−0	0
weddedratio	−0.00004	−0.00000	−0	0.00000	0.00000
ndclassinh	−0.0002	−0.00000	0	0.00000	0.00002

Table A.12: Fatalities during accidents not involving crash of 2 or more cars

	Dependent variable:
	log(nocrashdeathper10k)
popdensity	-0.0003^{***}
	(0.0001)
carsperinh	3.004^{***}
	(0.812)
ftclassinh	0.124
	(0.106)
rdclassinh	0.018^{***}
	(0.006)
Constant	-2.980^{***}
	(0.395)
Observations	75
R^2	0.358
Adjusted R^2	0.321
Residual Std. Error	0.326 (df = 70)
F Statistic	9.763^{***} (df = 4; 70)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.13: Fatality of single person per accident

	Dependent variable:
	log(solodeathperacc)
weddedratio	0.057^{**}
	(0.023)
carsperinh	2.781^{**}
	(1.272)
pension	-0.0002
	(0.0002)
ndclassinh	0.070
	(0.058)
Constant	-6.737^{***}
	(2.538)
Observations	75
R^2	0.252
Adjusted R^2	0.209
Residual Std. Error	0.445 (df = 70)
F Statistic	5.898^{***} (df = 4; 70)
Note:	*p<0.1; **p<0.05; ***p<0.01

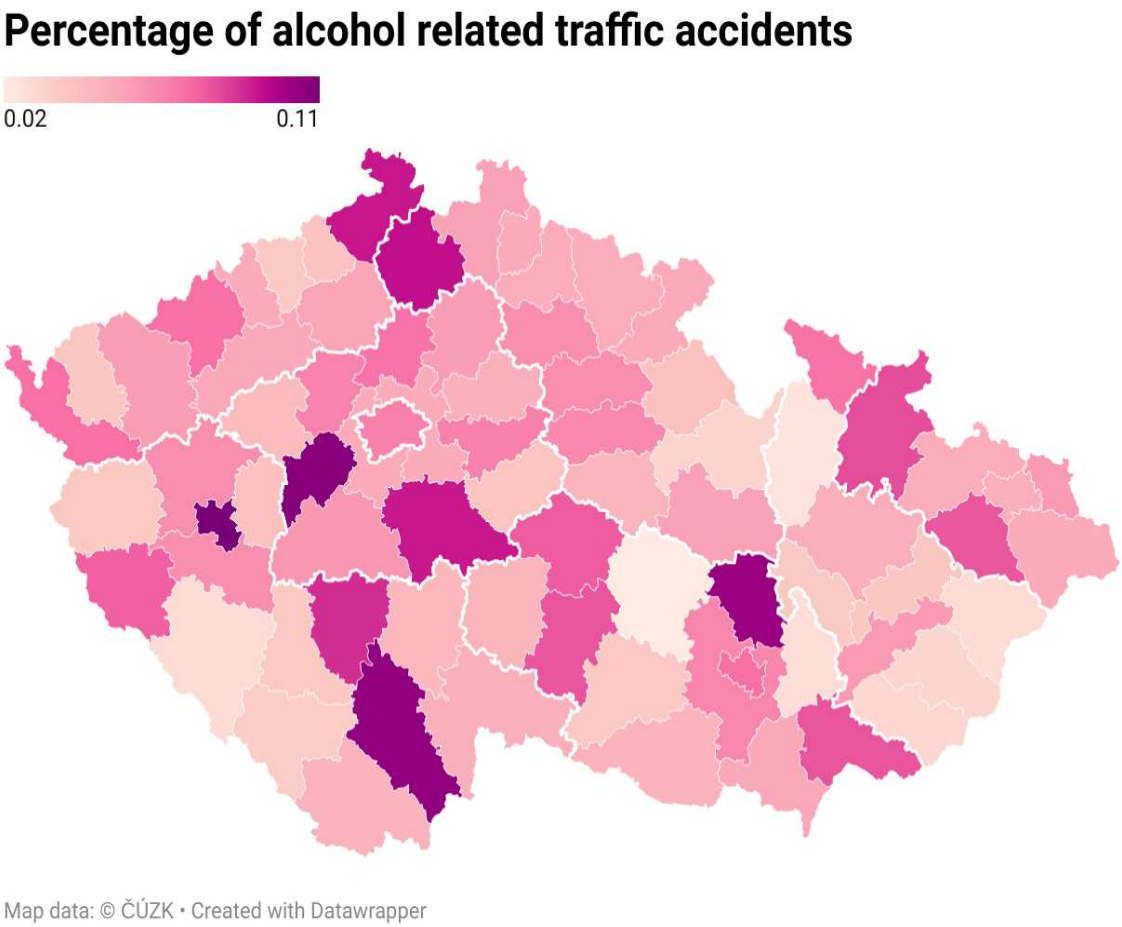


Figure A.2: Percentage of alcohol-related traffic accidents

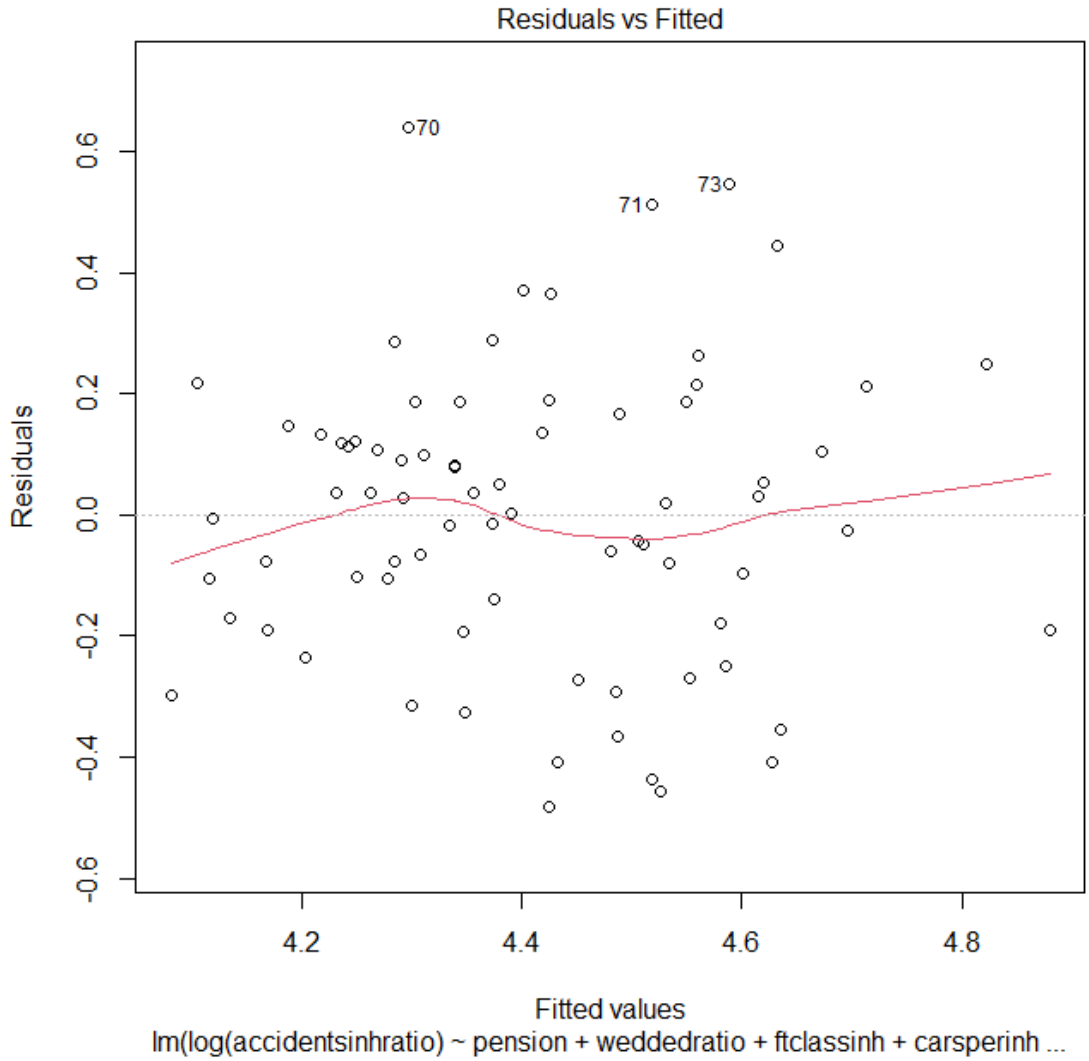


Figure A.3: Residuals vs fitted graph for model accidentsinhratio

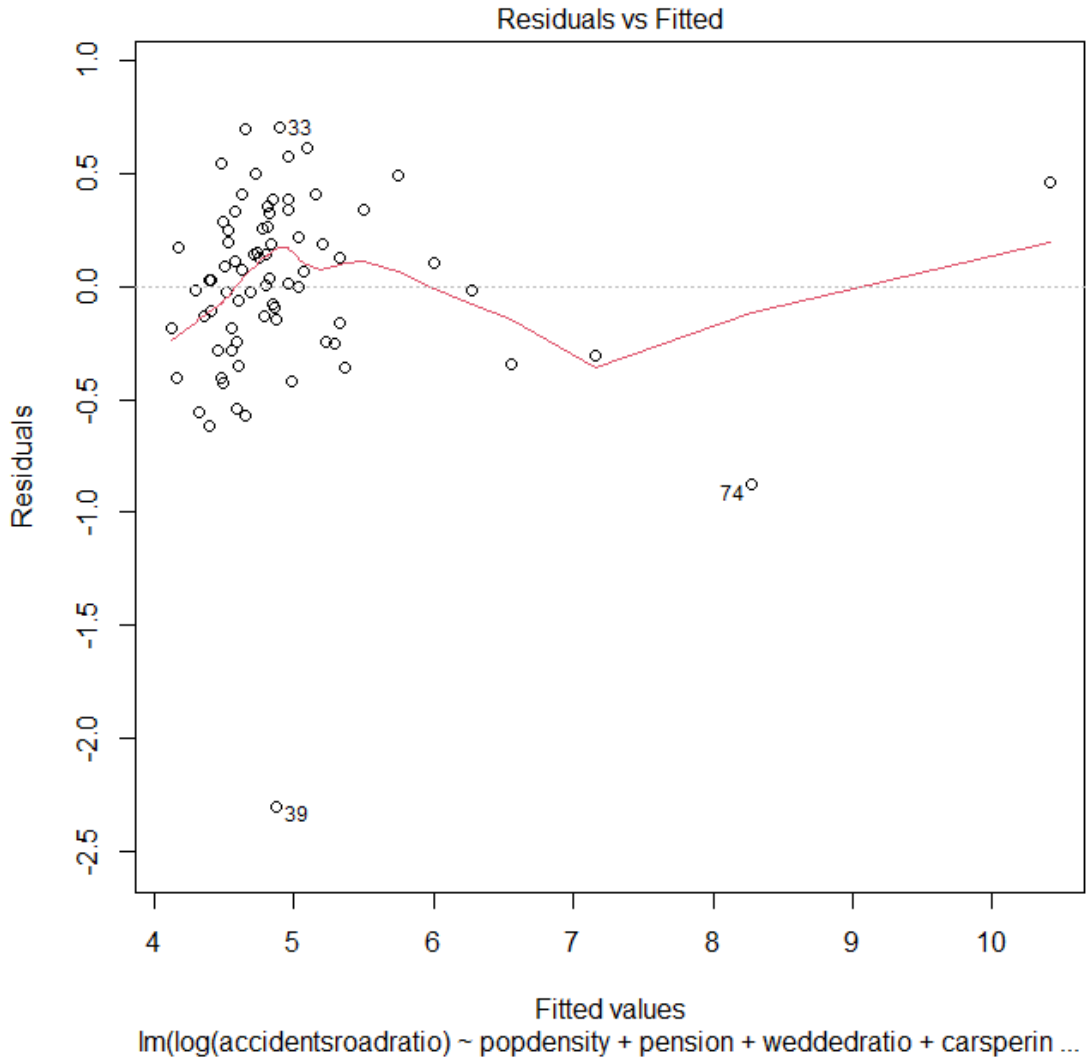


Figure A.4: Residuals vs fitted graph for model accidentsroadratio

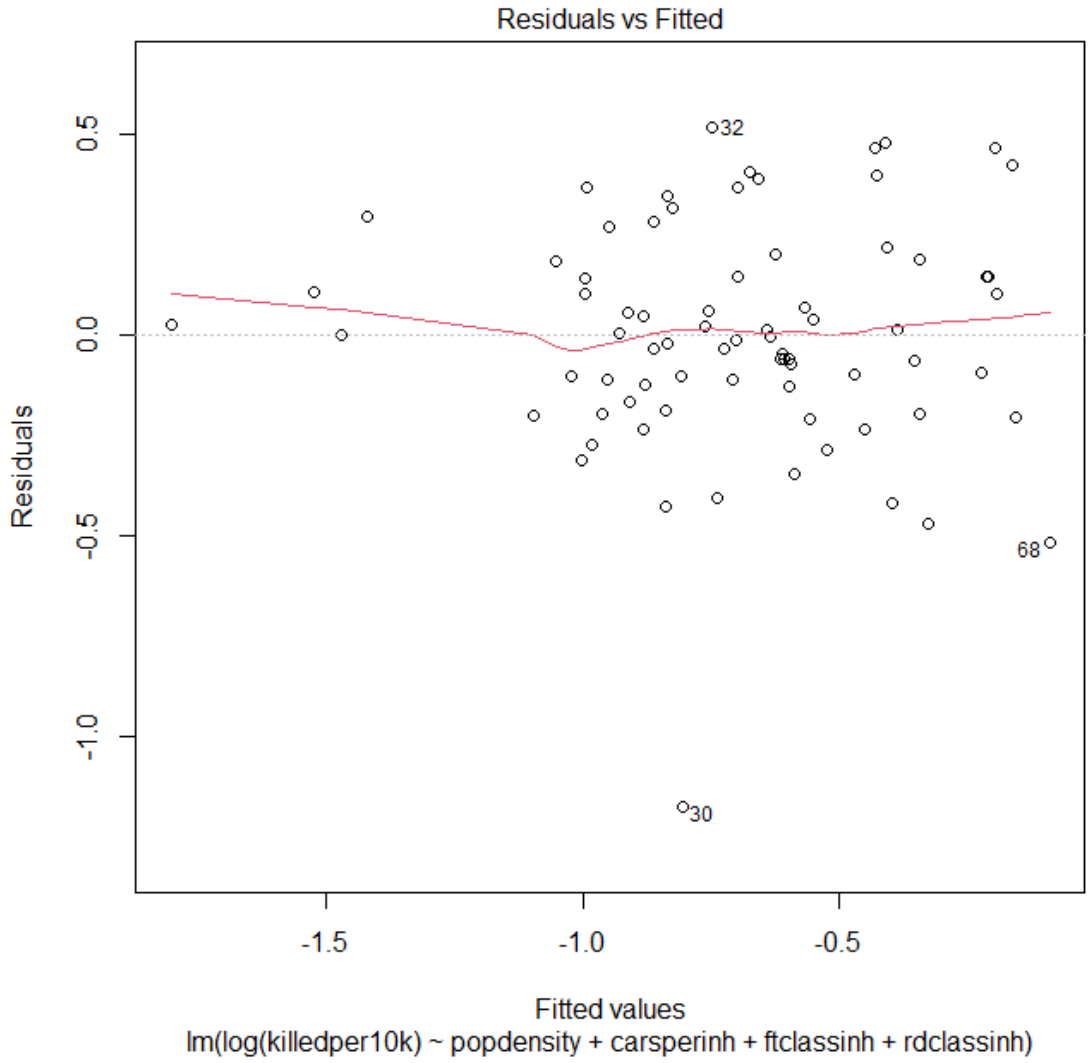


Figure A.5: Residuals vs fitted graph for model killedper10k

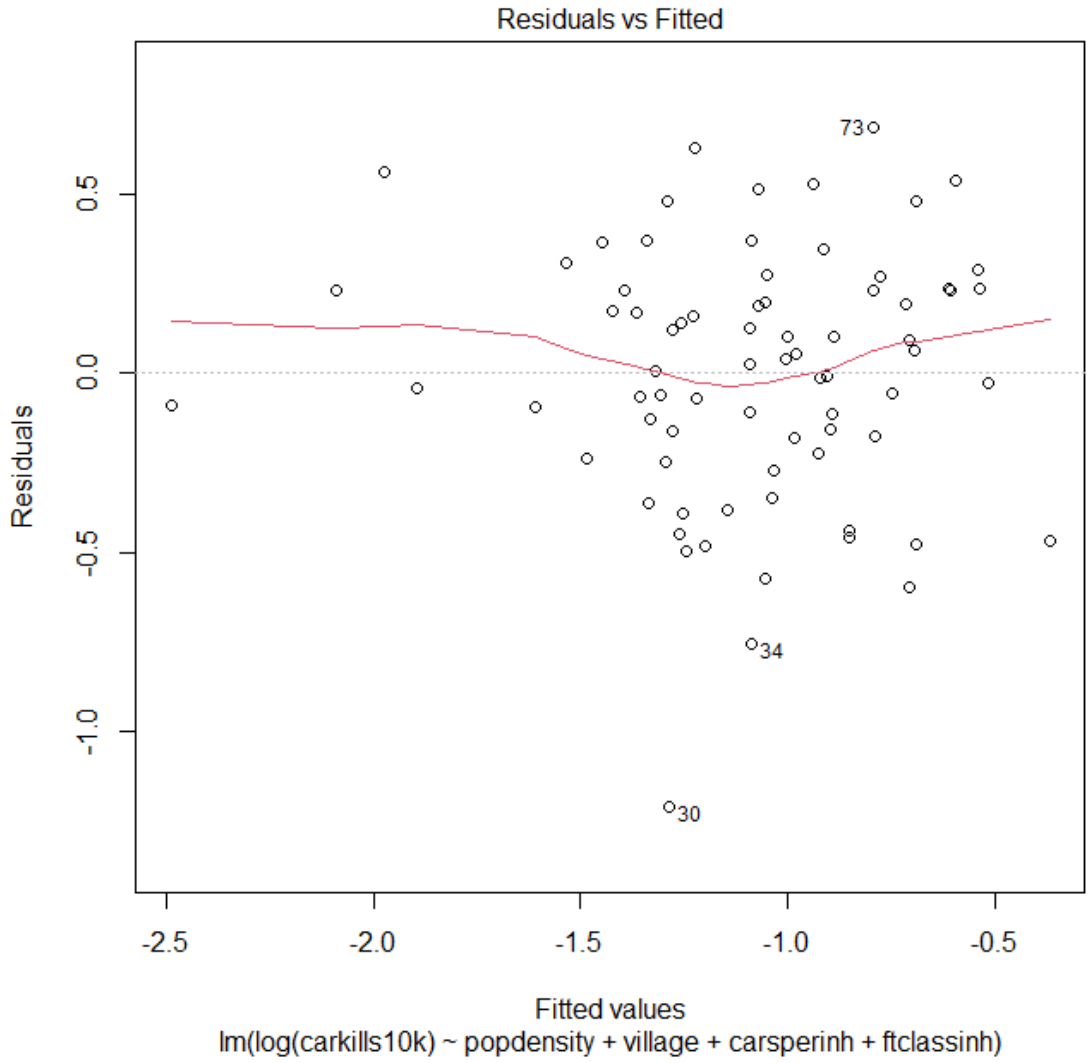


Figure A.6: Residuals vs fitted graph for model carkills10k

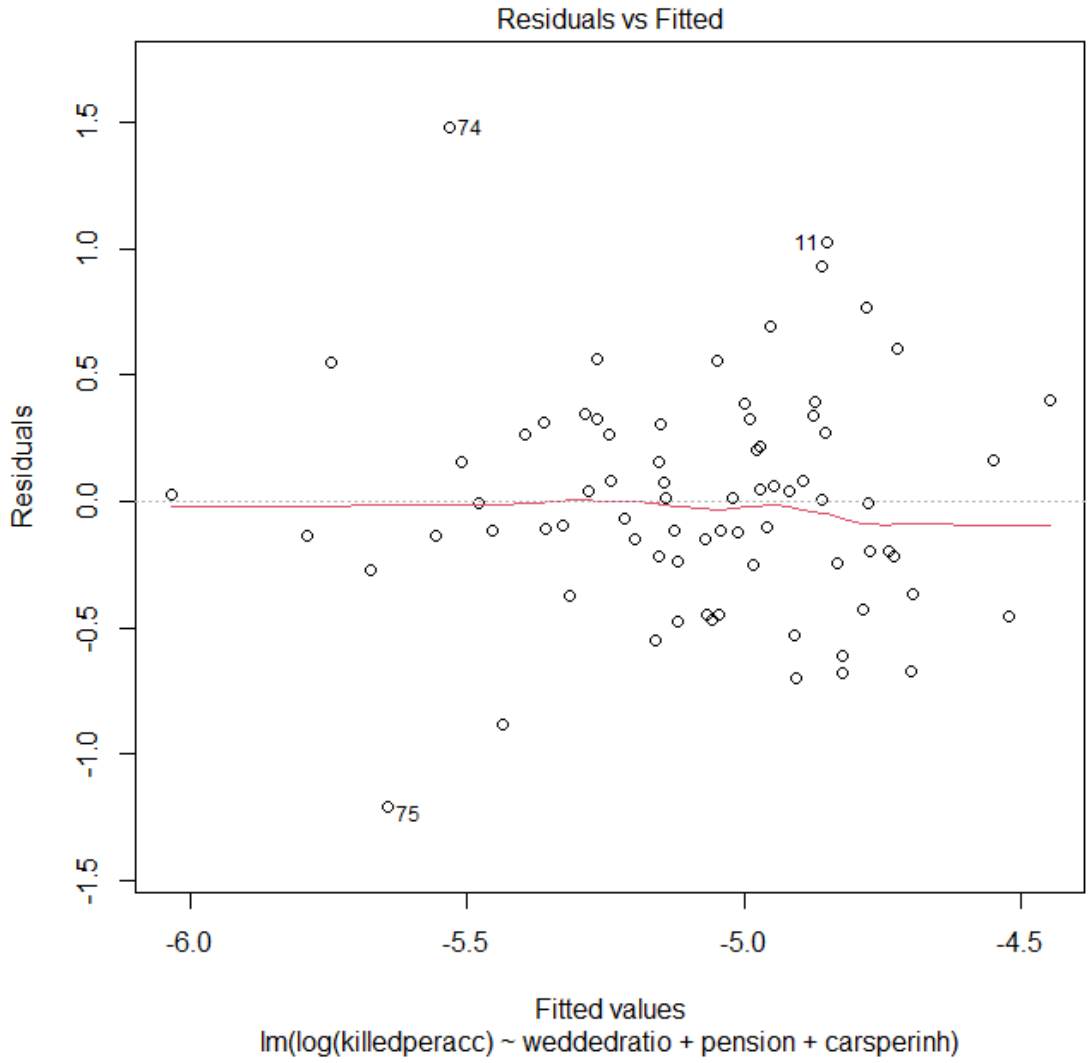


Figure A.7: Residuals vs fitted graph for model killedperacc

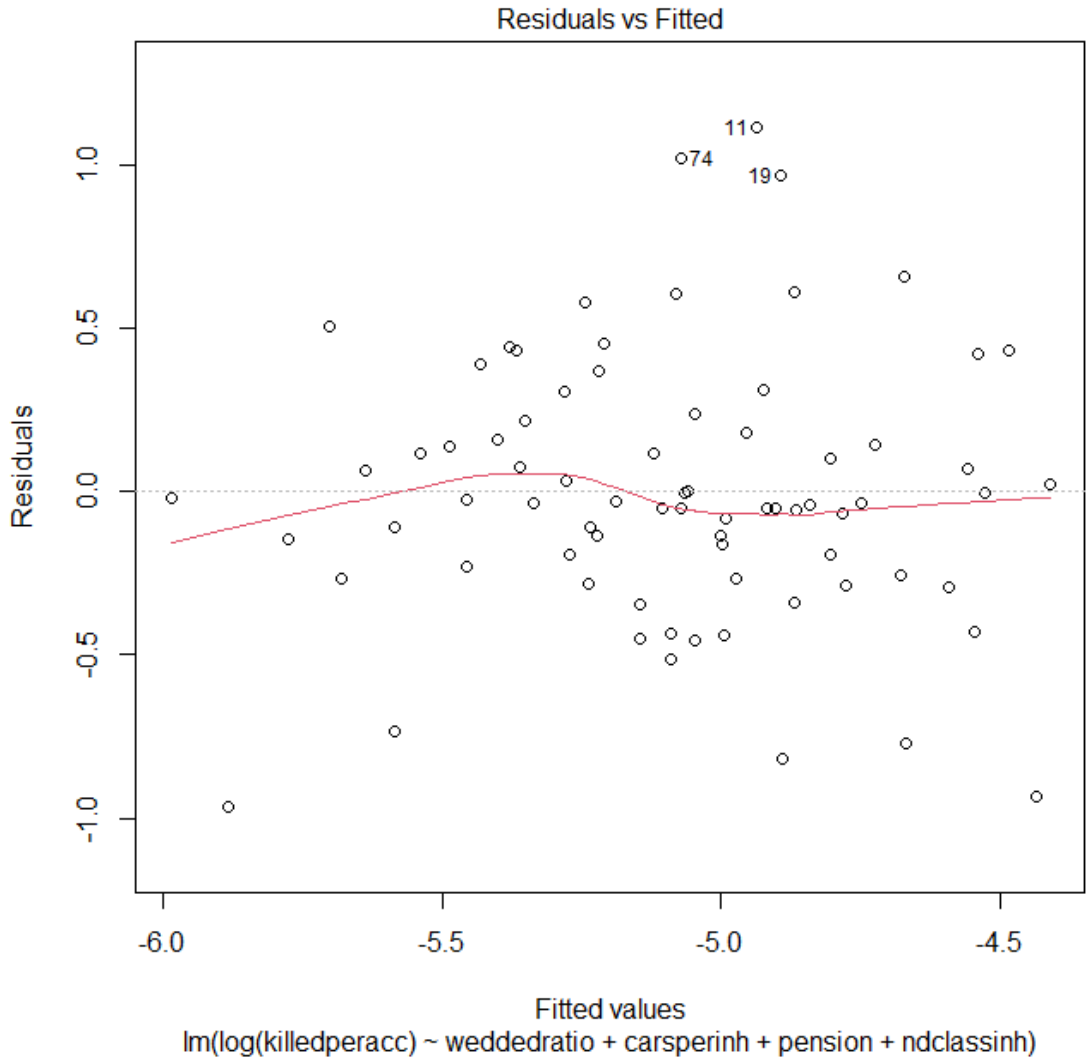


Figure A.8: Residuals vs fitted graph for model killedperacc2

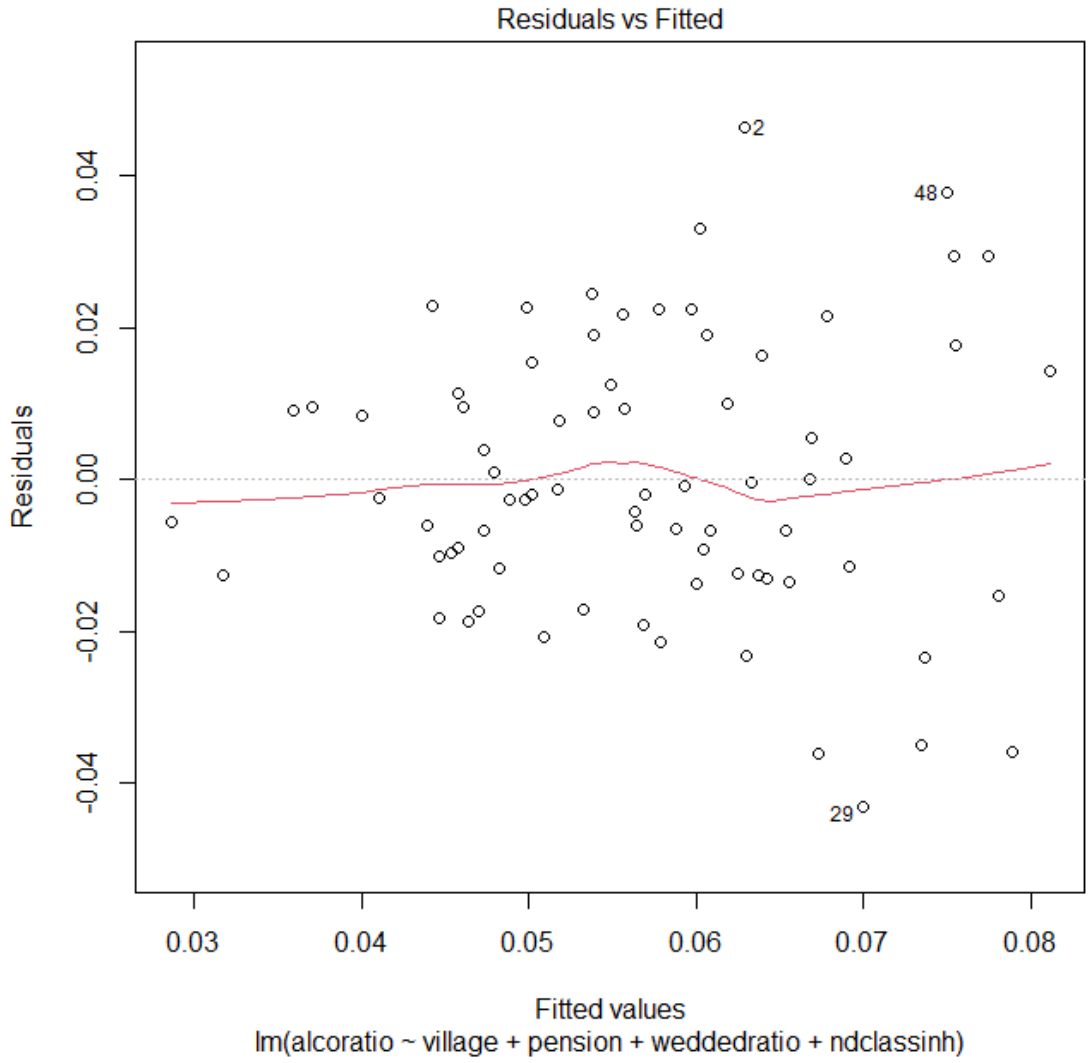


Figure A.9: Residuals vs fitted graph for modelalco

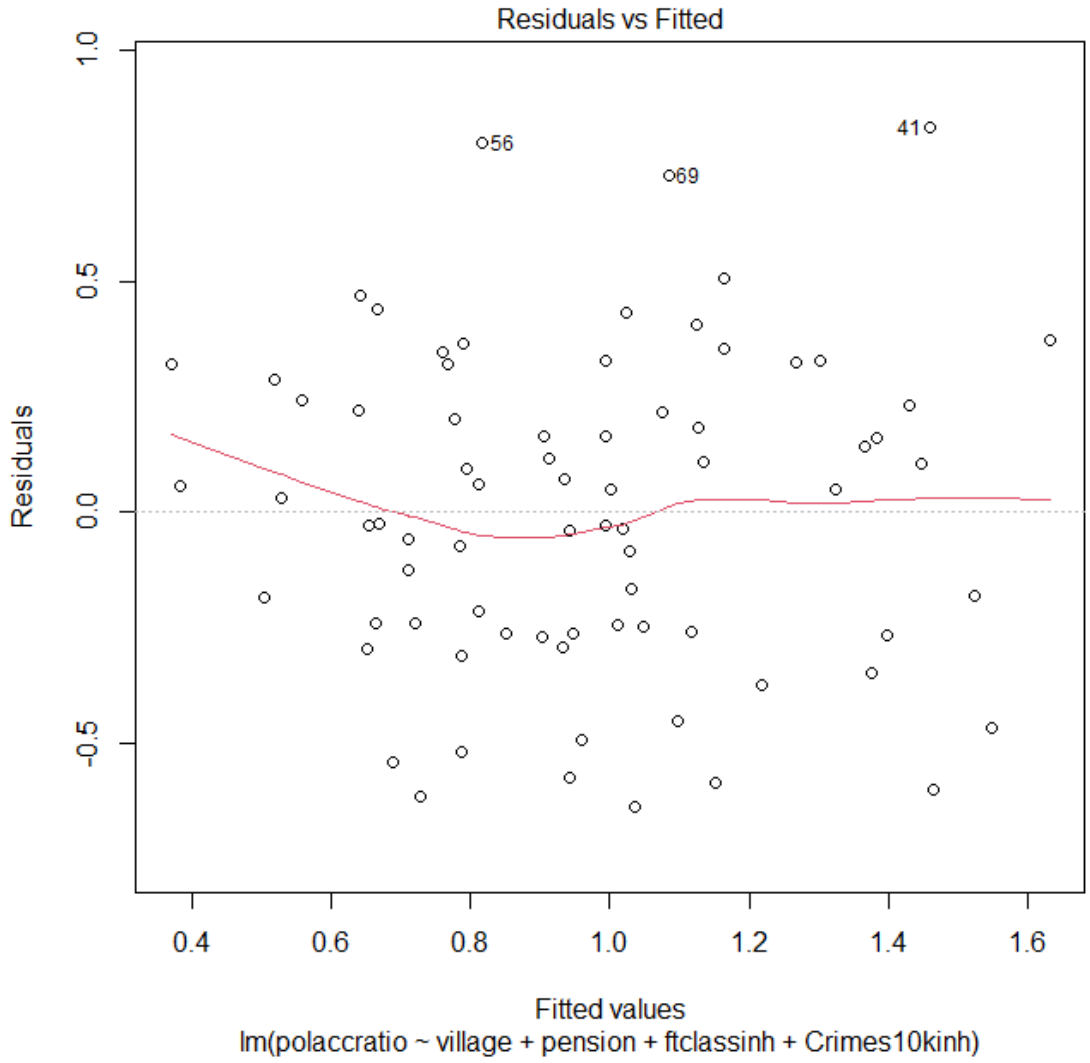


Figure A.10: Residuals vs fitted graph for modelpolice

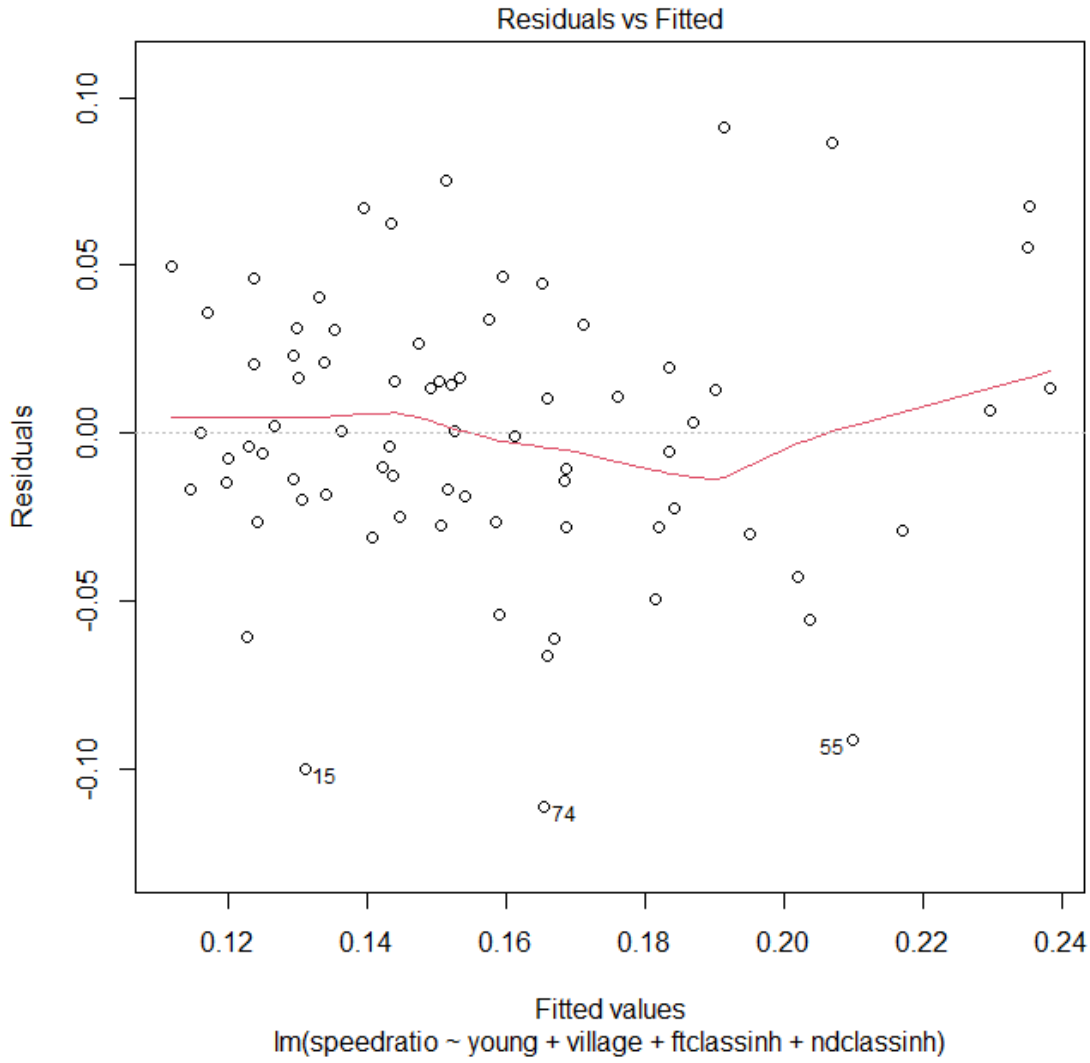


Figure A.11: Residuals vs fitted graph for modelspeed