UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
Katedra matematické analýzy a aplikací matematiky

BAKALÁŘSKÁ PRÁCE

Analýza dat týkajících se dopravních nehod na pozemních komunikacích se zaměřením na řidiče mopedů a motocyklů.

Vedoucí bakalářské práce: **RNDR. Tomáš Fürst, Ph.D.**
Vypracovala: **Irena Krobathová**
Studijní program: B1103 Aplikovaná matematika
Studijní obor: Aplikovaná statistika
Forma studia: Prezenční
Datum odevzdání: 2015

Bibliografická identifikace

Autor: Irena Krobathová

Název práce: Analýza dat týkajících se dopravních nehod na pozemních komunikacích se zaměřením na řidiče mopedů a motocyklů.

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: RNDR. Tomáš Fürst, Ph.D.

Rok obhajoby práce: 2015

Abstrakt: Cílem mé bakalářské práce je zkoumat data získané od Policie ČR, které se týkají nehodovosti mopedů, motocyklů do 50 ccm a motocyklů na silnicích České republiky v roce 2010.

Klíčová slova: Dopravní nehoda, počet nehod, motocykl, popisná statistika, kontingenční tabulka

Počet stran: 37

Počet příloh: 0

Jazyk: český

Bibliographical identification

Author: Irena Krobathová

Title: Analysis of traffic accidents data focused on motorcycles.

Type of thesis: Bachelor´s

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: RNDR. Tomáš Fürst, Ph.D.

The year of the presenation: 2015

Abstract: The aim of my bachelor´s thesis is to examine the data obtained from the Police of the Czech Republic concerning the accident mopeds, motorcycles up to 50 ccm and motorcycles on the roads in the Czech Republic in 2010.

Key words: Traffic accidents, number of accidents, motorcycle, descriptive statistics, pivot table

Number of pages: 37

Number of appendices: 0

Language: Czech

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracovala samostatně pod vedením pana Tomáše Fürsta a uvedla všechny použité zdroje.

Ve Zlíně dně 1. 4. 2014.

Irena Krobathová

Obsah

Úvod 7

1. Popis zkoumaných dat 8

   1.1. Popisná statistika 11

2. Teoretické pojmy 23

   2.1. Testování parametrických hypotéz 24

   2.2. Kontingenční tabulky 25

   2.3. Multinomické rozdělení 26

   2.4. Chi-kvadrát test nezávislosti 27

   2.5. Test homogenity 27

   2.6. Fisherův faktoriálový test 28

   2.7. Příklad s využitím softwaru R 28

3. Výzkum a analýza dat 31

   3.1. Druh pozemní komunikace 31

   3.2. Další testy nezávislosti pro viníky a účastníky nehod motocyklů 33

4. Závěr 36

5. Literatura 37

Poděkování

Chtěla bych především poděkovat vedoucímu mé práce Tomáši Fürstovi za ochotu, cenné rady a především trpělivost, bez níž bych tuto práci určitě neodevzdala. Dále bych chtěla poděkovat panu Pplk. Petru Svobodovi za poskytnutí dat. A ještě děkuji svým spolužákům za psychickou podporu.

Úvod

Cílem mé bakalářské práce je zkoumat data získané od Policie ČR, které se týkají nehodovosti mopedů, motocyklů do 50 ccm a motocyklů na našich silnicích v určitém roce a to v roce 2010.

Nehody motocyklistů bývají právem označovány za jedny z nejvíce život ohrožujících. Často se setkáváme v běžném životě se zprávami oznamujícími nehody motocyklistů s fatálními následky. Proto se motocyklisté často setkávají s nepochopením u svého okolí, které je převážně způsobeno strachem z neznámého, jelikož málokdo má dostatečné zkušenosti s řízením motocyklů. Dále díky mediálnímu obrazu, který byl vytvořen. Přesto by většina motocyklistů své stroje nevyměnila za jiné dopravní prostředky.

Počet usmrcených, těžce zraněných či lehce zraněných osob či celkový počet nehod patří mezi nejčastěji zmiňované ve sdělovacích prostředcích. I mě budou tyto počty zajímat. Patří totiž mezi to nejdůležitější, s čím budu dále pracovat, a na nichž můžu dále zkoumat různé souvislosti a předpoklady. Z toho důvodu si musím data upravit do požadovaného tvaru, rozdělit je a rozhodnout, která pro mě můžou mít dále nějaký význam. Na těchto datech potom vytvořím popisnou statistiku, která se inspiruje vlastní statistikou Policie ČR pro nehodovost všech motorových vozidel včetně mopedů, motocyklů do 50 ccm a motocyklů. Pro tuto popisnou statistiku budu využívat hlavně nehody způsobené řidiči motocyklů či nehody, kterých se řidiči motocyklů účastnili a to z toho důvodu, že počet dat u mopedů a motocyklů do 50 ccm není v mnoha případech dostatečný. Stejný problém budu mít i u testování hypotéz při analýze dat, tudíž se opět ve většině případů budu věnovat pouze řidičům silnějších motocyklů. Předtím ovšem uvedu teoretické předpoklady, abych mohla svou analýzu vypracovat. Tyto předpoklady se budou věnovat testovaní hypotéz a použití testu nezávislosti na data za použití programu R. Dále využití kontingenčních tabulek při této analýze a použití příslušné testovací statistiky a následnou interpretaci, zda data spolu souvisí či nikoliv. V závěru této práce budu interpretovat výsledky analýzy dat.

1. Popis zkoumaných dat

K výzkumu bylo potřeba získat data od Policie ČR, které mi zaslal pan Pplk. Petr Svoboda. Tato data byla rozdělená podle typu motorového vozidla na mopedy, motocykly do 50 ccm a motocykly. Mopedem se rozumí motorové vozidlo s objemem válců do 50 ccm, které je kombinací motocyklu a jízdního kola (má pedály). Příkladem může být Babetta. Motocyklem do 50 ccm je myšleno motorové vozidlo, které tvoří přechod mezi mopedem a ostatními silničními typy, jedná se převážně o skútry. Za motocykl lze považovat jakékoliv dvoukolové motorové vozidlo nad 50 ccm. Při těchto popisech a členění řidičských oprávnění bylo použito zdroje [2], [3].

Řízení mopedu nebo motocyklu do 50 ccm s maximální konstrukční rychlostí 45km/hod a výkonem motoru do 4 kW je automaticky možné pro člověka staršího 18 let, jenž vlastní řidičské oprávnění skupiny B. Řidičské oprávnění skupiny AM může získat člověk starší 15 let a povoluje řízení výše uvedeného typu motorového vozidla. Řidičské oprávnění skupiny A1 může získat člověk starší 16 let a opravňuje k řízení motocyklů o objemu válců nepřesahujících 125 ccm a o výkonu motoru nejvýše 11 kW s postranním vozíkem i bez něj. Řidičské oprávnění skupiny A2 opravňuje člověku staršímu 18 let řídit motocykly s postranním vozíkem i bez něj při maximálním výkonu motoru 35 kW. Řidičské oprávnění skupiny A lze získat při dovršení 24 let a opravňuje řídit motocykly s postranním vozíkem i bez něj a platí i na motorová vozidla skupin A1 a A2. Při dvouletém držení řidičského oprávnění skupiny A2 lze získat řidičské oprávnění skupiny A už ve 20 letech.

Dále jsou data rozdělena na viníky nehod a účastníky nehod pro každou skupinu zvlášť. U mopedu je 105 záznamů u viníků nehod a 55 u účastníků nehod. U motocyklu do 50 ccm se jedná o 132 záznamů u viníků nehod a 121 u účastníků nehod. U motocyklu je 1498 záznamů u viníků nehod a 1200 u účastníků nehod.

Pro všechny skupiny je vytvořeno členění, které jsem ponechala v původním stavu. Uvedu zde všechny proměnné, s kterými dále budu pracovat a následně na nich vytvořím popisnou statistiku.

Tab.1.1. Druh pozemní komunikace

	Druh pozemní komunikace
0	Dálnice
1	Silnice I. třídy
2	Silnice II. třídy
3	Silnice III. třídy
4	Uzel tzv. sledovaná křižovatka
5	Komunikace sledovaná
6	Komunikace místní
7	Komunikace účelová
8	Ostatní

V první tabulce (Tab.1.1.) je uvedeno členění do devíti různých kategorií podle toho, na jaké pozemní komunikaci se dopravní nehoda stala. V další tabulce (Tab.1.2.) rozdělujeme nehody podle toho, jestli se staly v obci nebo mimo obec.

Tab.1.2. Místo nehody

	Místo nehody
1	V obci
2	Mimo obec

Tab.1.3. Den v týdnu

	Den
0	Sobota
1	Neděle
2	Pondělí
3	Úterý
4	Středa

5	Čtvrtek
6	Pátek

Tab.1.3. rozděluje nehody do kategorií podle dne v týdnu.

Tab.1.4. Druh nehody

	Druh nehody
1	Srážka s jedoucím vozidlem
2	Srážka s odstaveným vozidlem
3	Srážka s pevnou překážkou
4	Srážka s chodcem
5	Srážka s lesní zvěří
6	Srážka s domácím zvířetem
7	Srážka s vlakem
8	Srážka s tramvají
9	Havárie
0	Jiná

Tato tabulka Tab.1.4. se zabývá druhem nehody, které se člení do 10 různých kategorií. Převažuje srážka s jedoucím vozidlem a havárie u všech druhů motocyklů.

Tab.1.5. Zavinění nehody

	Zavinění nehody
1	Řidič motorového vozidla
2	Řidič nemotorového vozidla
3	Chodec
4	Zvěř
5	Jiný účastník
6	Závada komunikace
7	Závada technická
0	Jiné

Tabulka (Tab.1.5.) je rozdělena na 8 kategorií podle toho, kdo zavinil dopravní nehodu. Nepřipadá do úvahy u účastníka, data se týkají pouze viníků nehod.

Tab.1.6. Alkohol

	Alkohol
1	Ano
2	Ne
0	Nezjištěno

Tabulka (Tab.1.6.) obsahuje data týkajících přítomnosti alkoholu v době nehody. V tomto případě je důležité si uvědomit, že ne všichni řidiči volají policii (například pokud se jedná o malou škodu), což má především vliv na data u mopedů a malých motocyklů do 50 ccm. Dále je možné dechovou zkoušku odmítnout, za což může být uložena sankce ve výši 25.000,- Kč až 50. 000,- Kč a zákaz řízení na dobu jednoho nebo dvou let. V České republice je nulová tolerance, přesto naměření hodnot nižších než 0,24 promile alkoholu není dostatečným důkazem, že by byl řidič pod vlivem alkoholu.

Následky nehody (stav 24 hodin od nehody)

Ty jsou rozděleny do tří podkategorií, které nás budou nejvíce zajímat a to na **Usmrceno osob**, **Těžce zraněno**, **Lehce zraněno**, jedná se o počty usmrcených osob, počty těžce zraněných osob a počty lehce zraněných osob.

Rok výroby vozidla

Konečná dvojice čísel roku, např. 99, což znamená rok 1999.

Rok narození

Poslední dvojčíslí roku. Stejně jako u předchozího příkladu.

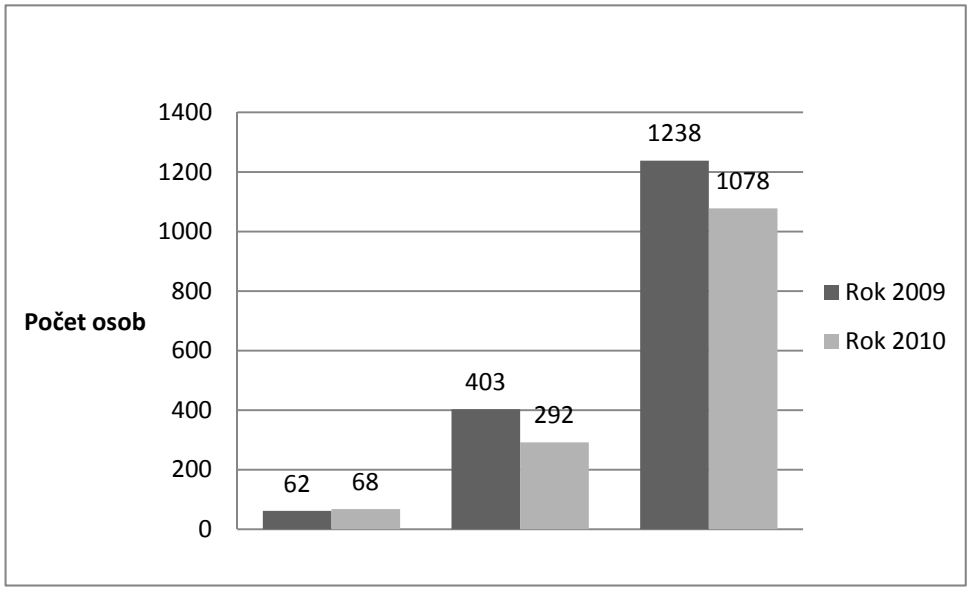
1.1. Popisná statistika

Policie ČR v roce 2010 šetřila celkem 1538 nehod z výše uvedených kategorií motorových vozidel. Ve všech následujících případech budu využívat zdroje [2]. Dále budu uvádět hodnoty pro mopedy, motocykly do 50 ccm a motocykly společně. Z toho bylo 68 lidí usmrceno, 292 lidí těžce zraněno a 1078 lidí lehce zraněno. Celková hmotná

škoda byla Policií ČR vyčíslena na 69, 4217 milionu Kč. Při 128 nehodách byl přítomen alkohol. Počet nehod v obci byl 868. Při porovnávání dat s rokem minulým, tedy rokem 2009, získáváme následující hodnoty:

Tab.1.1.1. Srovnání pro rok 2009 a 2010

	rok 2009	**rok 2010**	**rozdíl (pro rok 2010)**	**v procentech (pro rok 2010)**
počet dopravních nehod	1762	1538	-224	pokles o 12,7 %
počet usmrcených	62	68	6	nárůst o 9,7 %
počet těžce zraněných osob	403	292	-111	pokles o 27,5 %
počet lehce zraněných osob	1238	1078	-160	pokles o 12,9 %

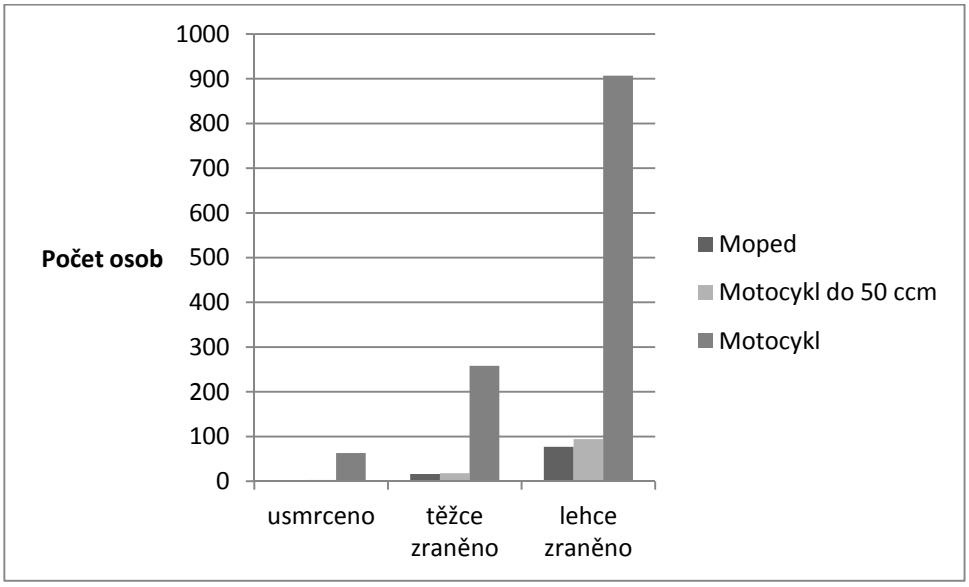


Obr.1.1.1. Srovnání roku 2009 a 2010, co se týká počtu usmrcených a zraněných osob

Z následující tabulky i grafu je jasně vidět, že v roce 2010 došlo ve všech případech až na počet usmrcených k poklesu. Největšího poklesu dosahuje počet těžce zraněných osob, jenž dosahuje poklesu o 27,5 %. Za zmínku ale stojí uvést to, že nejvíce dopravních nehod mají na svědomí řidiči motocyklů a to 1325, oproti mopedům s 99 dopravními nehodami a malým motocyklům do 55 ccm s 114 dopravními nehodami. Z toho vyplývá, že i počet usmrcených je výrazně vyšší u řidičů motocyklů než u řidičů slabších strojů.

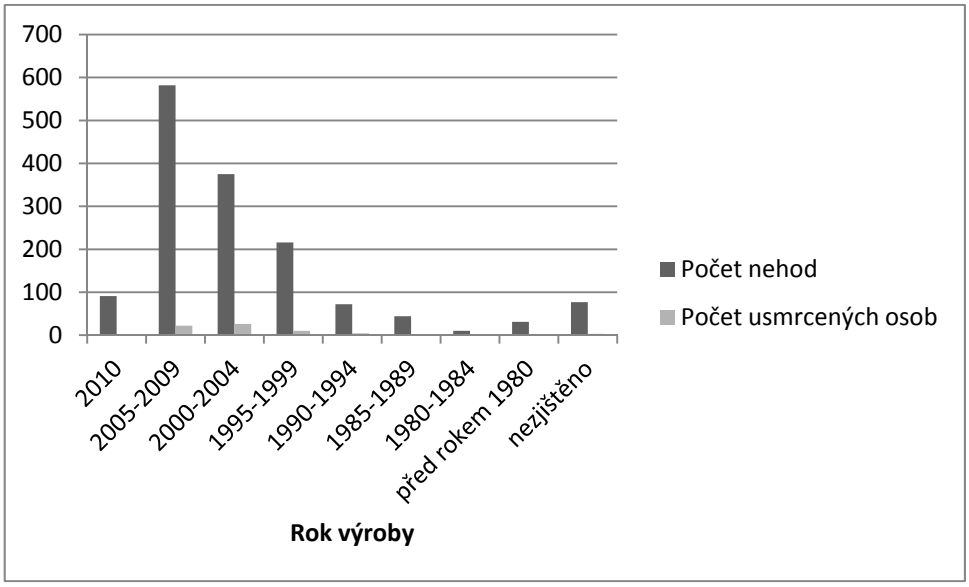
Pro Tab.1.1.1. a Obr.1.1.2. byly použity data, které shrnula Policie ČR pro roky 2009 a 2010, v dalších případech jsem byla odkázána na data s kterými jsem pracovala.

V grafu Obr.1.1.2 na další straně uvádím srovnání počtu usmrcených nebo zraněných podle typu motorového vozidla



Obr.1.1.2. Srovnání počtu usmrcených a zraněných pro daný typ motorového vozidla

Dále se zaměřím na viníky nehod, ale pouze na řidiče motocyklů. V grafu bude uveden přehled pro rok výroby a počet nehod zaviněných řidiči motocyklů, a dále počet usmrcených osob. Použiji proměnnou Rok výroby vozidla a Počet usmrcených osob.



Obr.1.1.3. Srovnání počtu nehod a počtu usmrcených osob podle roku výroby

Nejvyšší podíl na počtu usmrcených osob mají řidiči motocyklů vyrobených v rozmezí let 2000 až 2004, což činí 38%. Závažnost nehod (= počet usmrcených osob na 1000 nehod) v tomto rozmezí je 69,3. Druhou nejvíce zastoupenou skupinou bylo rozmezí let 2005 až 2009 s 33,3 %. Závažnost nehod v tomto rozmezí je 37,8. Závažnost nehod není ale jasným ukazatelem kvůli nízkému počtu zaviněných nehod. Přesto se ukazuje, že se zvyšuje postupně jak počet nehod, tak počet usmrcených osob.

Na těchto datech použiju f-test a uvedu příklad ze zdroje [4]. Teorie s dalšími příklady bude uvedena a podrobně popsána v následujících kapitolách. Žena z Velké Británie měla za úkol rozlišit, zda byl do šálku přidán nejprve čas a následně mléko či první mléko a až poté čaj. K testování jí bylo předloženo 8 šálků čaje, z nichž bylo do 4 přidáno první mléko a poté až čaj. Nulová hypotéza v tomto případě byla taková, že neexistuje ve skutečnosti žádný vztah mezi pořadím lití. Žena odhadovala, že zde bude existovat pozitivní asociace, což bude v tomto případě alternativa (poměr šancí bude větší než 1). Pro testování bylo použito programu R a následujícího příkazu

```
TeaTasting <-
matrix(c(3, 1, 1, 3),
       nrow = 2,
       dimnames = list(Guess = c("Milk", "Tea"),
                       Truth = c("Milk", "Tea")))
fisher.test(TeaTasting, alternative = "greater").
```

V tomto případě vyšla hodnota p-value 0,2429, což neznamená prokázání souvislosti mezi pořadím lití.

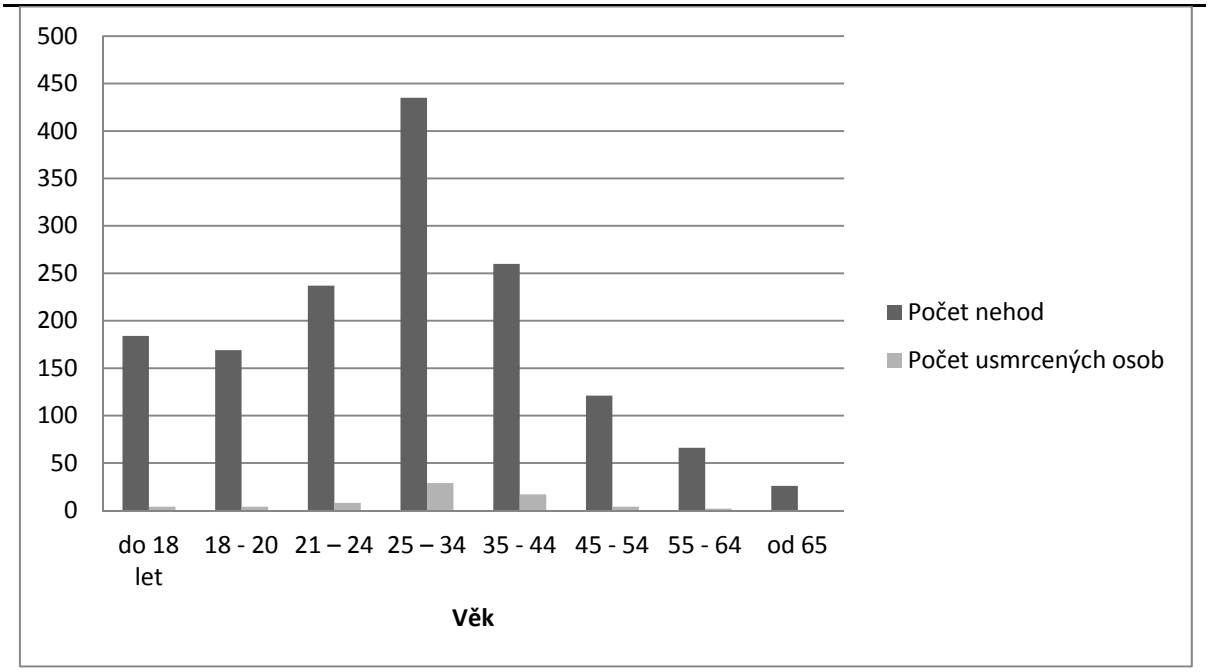
Podobně budu testovat data, které si uspořádám do tabulky v R a použiji na ně Fisherův test. Hodnota p-value vychází 0,2743, což nám ukazuje, že mezi těmito proměnnými nelze dokázat žádný vliv. Pro tato data vypadalo zadání v R následovně

```
t<-matrix(c(91,1,582,22,375,26,216,10,72,4,44,2,77,3),nrow=2)
fisher.test(t)
```

Opět se budu věnovat viníkům řídící motocykl, tentokrát se zaměřím na jejich věk a dalšími ukazateli bude jako v předchozím případě počet nehod a počet usmrcených osob, tedy proměnné Rok narození a Usmrceno osob. Toto dělení ročníků jsem převzala z prezentace Policie ČR.

Tab.1.1.2. Počet nehod a počet usmrcených osob podle věku řidiče motorového vozidla

Ročník	Počet nehod	Počet usmrcených osob	Podíl usmrcených osob (v %)
do 18 let	184	4	6
18 – 20	169	4	6
21 – 24	237	8	12
25 - 34	435	29	43
35 – 44	260	17	25
45 – 54	121	4	6
55 – 64	66	2	3
od 65 let	26	0	0

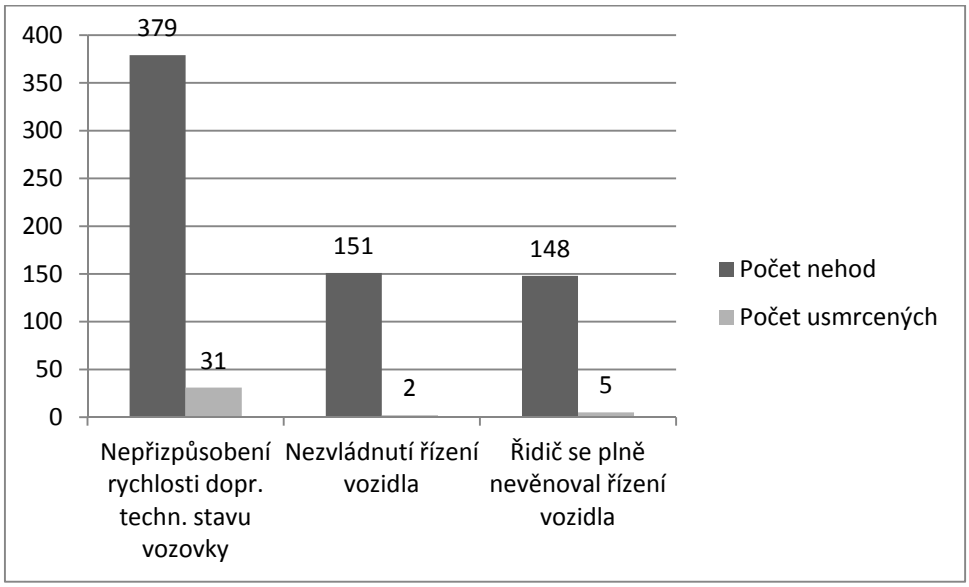


Obr.1.1.4. Počet nehod a počet usmrcených osob podle věku řidiče

Nejvíce se na nehodách podílejí řidiči mezi 25 – 34 let (29 %) a mají na svědomí i největší počet usmrcených osob a to celých 42,6 %. Přesto jde vidět, že většinu dopravních nehod způsobují mladší ročníky. Ale co se týká závažnosti, tak převahu má rozmezí 25 až 34 (66,6), dále ročníky 35 až 44 (65), 21 až 24 (33,7), 45 až 54 (33). Ovšem tyto vysoké hodnoty vznikají nízkým počtem dopravních nehod.

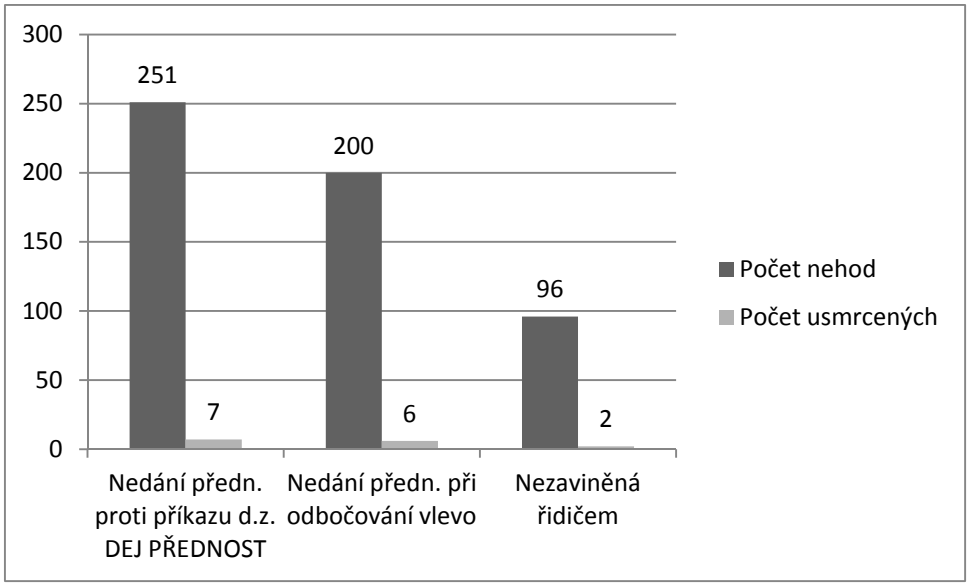
V tomto případě vychází hodnota kritického oboru W = <12,592; ∞) při 6 stupních volnosti (jelikož jsem byla nucena kontingenční tabulku upravit pro nevyhovující četnosti) a pro dané $\alpha = 0,05$. Potom hodnota testovací statistiky je rovna z = 11, 3458 a p-value je rovna 0,07826. Při této konstelaci nelze nulovou hypotézu H_0 zamítnout. Proto můžeme tvrdit, že mezi těmito proměnnými neexistuje závislost, tudíž že počet nemá žádnou souvislost s tím, kolik bude usmrcených osob.

V dalších grafech uvedu počet nehod řidičů motocyklů v souvislosti s hlavní příčinou nehody a s počty usmrcených. Nejprve pro viníky.



Obr.1.1.5. Hlavní příčiny nehody viníků

Následně pro účastníky nehod.



Obr.1.1.6. Hlavní příčiny nehod účastníků

U počtu nehod zaviněných řidičem motocyklu převažuje nepřizpůsobení rychlosti dopravnímu stavu vozovky a tvoří to 23,2 % všech zaviněných dopravních nehod řidičem motocyklu. Jde vidět jasný rozdíl oproti účastníkům dopravních nehod v zastoupení jednotlivých kategorií. V tomto případě převažuje nedání přednosti proti příkazu dej přednost v jízdě a tvoří to 20,3 % veškerých nehod, jichž se účastnili řidiči motocyklů. Co se týká počtu usmrcených, tak v případě viníků má na svědomí největší počet usmrcených nepřizpůsobení rychlosti dopravnímu stavu vozovky a to 31 lidí (51,7 % všech usmrcených), 5 usmrcených z důvodu, že se řidič plně nevěnoval řízení vozidla, dále 4 usmrcené při nepřizpůsobení rychlosti vlastního vozidla a nákladu a 4 usmrcené při překročení předepsané rychlosti stanové pravidly. Co se týká účastníků, tak počet 7 usmrcených bylo z důvodu nedání přednosti proti příkazu Dej přednost (18,4 % všech usmrcených), 6 usmrcených při nedání přednosti při odbočování vlevo, dále 4 usmrcení při nedání přednosti při vjíždění na silnici a taktéž 4 usmrcení při nedání přednosti proti příkazu Stůj dej přednost v jízdě.

I tyto data budeme testovat. Nejprve provedeme test pro viníky, kde kritický obor bude W = <5,99; ∞) při 2 stupních volnosti a $\alpha = 0,05$ vychází hodnota testovací statistiky z = 10,2811 a p-value je 0,005855, což znamená, že lze H_0 zamítnout a tím pádem můžeme konstatovat, že existuje souvislost mezi počtem nehod a počtem usmrcených osob.

Dále testujeme účastníky, kde kritický obor zůstává stejný při dané hladině významnosti $\alpha = 0,05$, kde hodnota testovací statistiky z = 0, 1979 a p-value je 0,9058, což

znamená, že nelze H_0 zamítnout, tudíž nelze určit, zda mezi proměnnými existuje souvislost. Rozdíl mezi viníkem a účastníkem je dán vysokým počtem usmrcených osob u viníka.

Dále se budu zabývat vlivem alkoholu ve srovnání s počtem nehod a počtem usmrcených u viníků nehod.

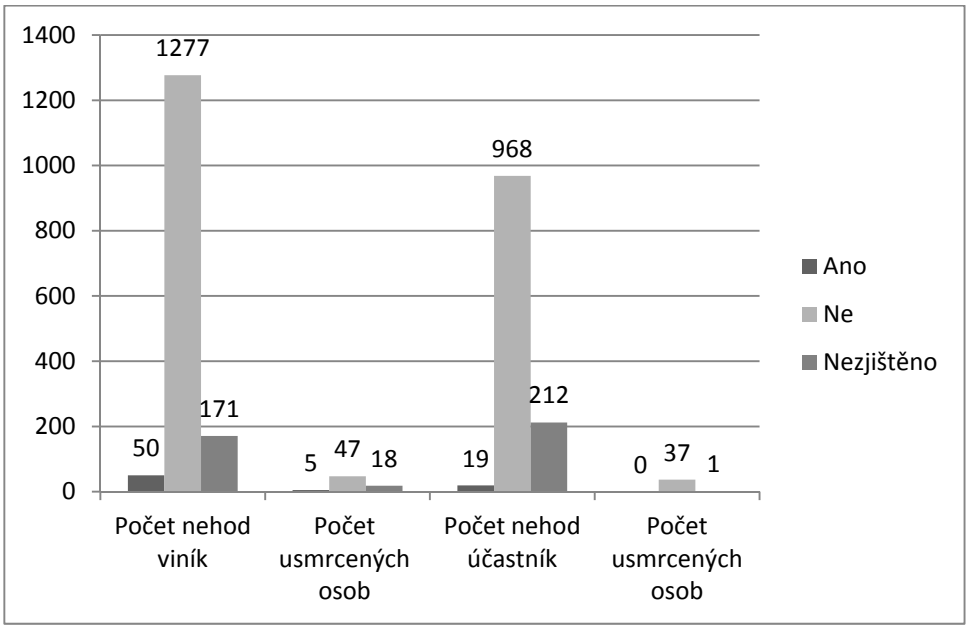
Tab.1.1.3. Vliv alkoholu pro viníky nehod

Alkohol	Počet nehod	Počet usmrcených
Ano	50	5
Ne	1277	47
Nezjištěno	171	18

Následně udělám tabulku i pro účastníky nehod.

Tab.1.1.4. Vliv alkoholu pro účastníky nehod

Alkohol	Počet nehod	Počet usmrcených
Ano	19	0
Ne	968	37
Nezjištěno	212	1



Obr.1.1.7. Srovnání počtu nehod a počtu usmrcených pro viníka a pro účastníka

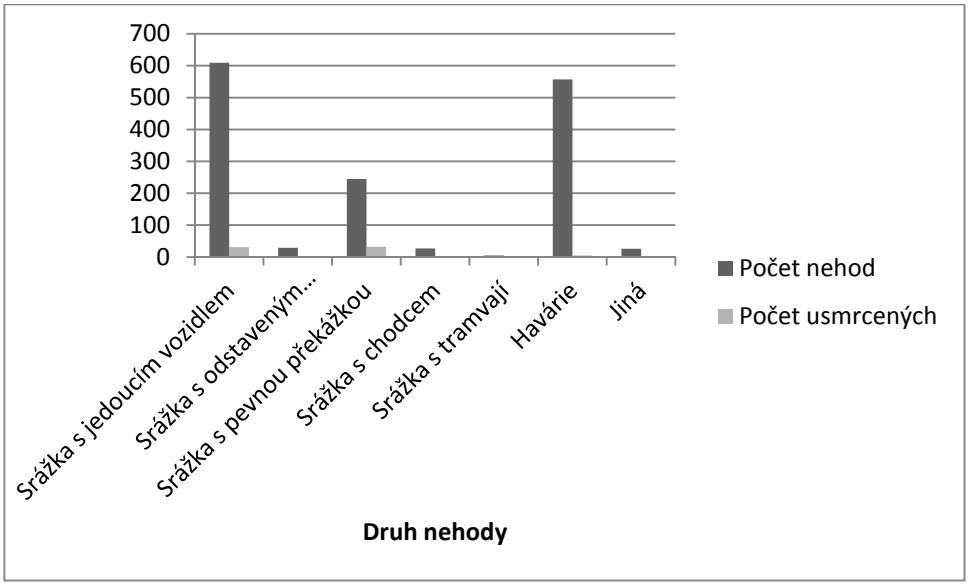
U viníků tedy Policie ČR eviduje celkem 50 nehod, za nimiž stála přítomnost alkoholu a to tvoří 3,3 % všech nehod způsobených viníky (řidiči motocyklů) dopravních nehod. Je ovšem nutné dodat, že pro tuto statistiku nebyla poskytnuta dostatečná data a jak jsem uváděla už v kapitole 1.1., tato data nebudou úplná a to z důvodu toho, že data jsou získána v tomto případě od Policie ČR.

I na těchto datech provedu testy. Vytvořím kontingenční tabulku a provedu test homogenity pomocí programu R a příkazu chisq.test. Program bude vypadat následovně pro Tab.1.1.3 a Tab.1.1.4.

```
a<-matrix(c(50,5,1277,47,171,18),nrow=2)
chisq.test(a)
a1<-matrix(c(19,0,968,37,212,1),nrow=2)
chisq.test(a1).
```

Nejprve si určíme kritický obor W = <5,991; ∞) při 2 stupních volnosti a hladině významnosti $\alpha = 0,05$. Potom pro Tab.1.1.3 vychází hodnota testovacího kritéria z = 6,7015 a p-value 0,0306, z čehož plyne, že lze H_0 na dané hladině významnosti zamítnout, tudíž existuje vztah mezi proměnnými. Pro účastníky tedy Tab.1.1.4 platí stejný kritický obor i hladina významnosti a hodnota testovacího kritéria z = 4,9837 a p-value = 0,02559, tudíž nelze H_0 zamítnout a mezi proměnnými proto nelze určit souvislost.

Další kategorií, na kterou se zaměřím, budou druhy nehod. Opět vytvořím graf pro druh nehody v souvislosti s počtem nehod a s počtem usmrcených. Budu to tvořit pouze pro viníka.



Obr.1.1.8. Počet nehody a počet usmrcených pro viníka nehody podle druhu nehody

Nejvíce nehod nastalo při srážce s jedoucím vozidlem a to 40,7 % všech dopravních nehod způsobených řidičem motocyklu, dále 557 nehod nastalo při havárii (37,2 % všech dopravních nehod způsobených řidičem motocyklu). Přitom nejvíce usmrcených si vyžádala srážka s pevnou překážkou a to 32 usmrcených osob, což tvoří 47,1 % všech usmrcených osob při dopravní nehodě způsobené řidičem motocyklu, dále 31 usmrcených osob při srážce s jedoucím vozidlem (45,6 %).

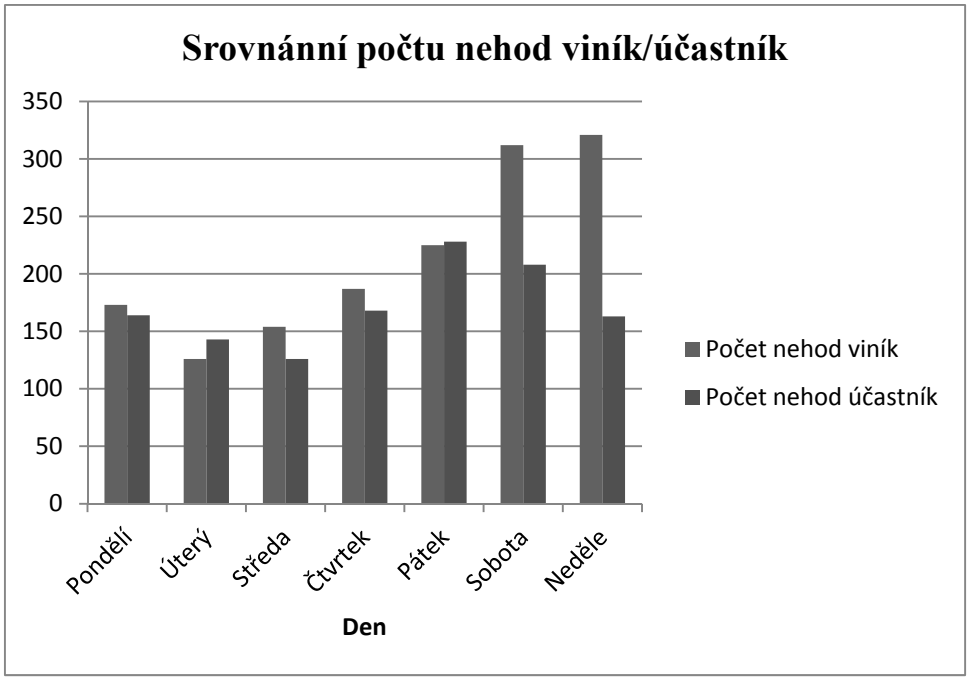
Další tabulku bude tvořit rozdělení počtu nehod a počtu usmrcených podle dne v týdnu. Tentokrát vytvořím tabulku jak pro viníky, tak účastníky v tomto pořadí.

Tab.1.1.5. Počet nehod a počet usmrcených podle dnu v týdnu pro viníky nehod

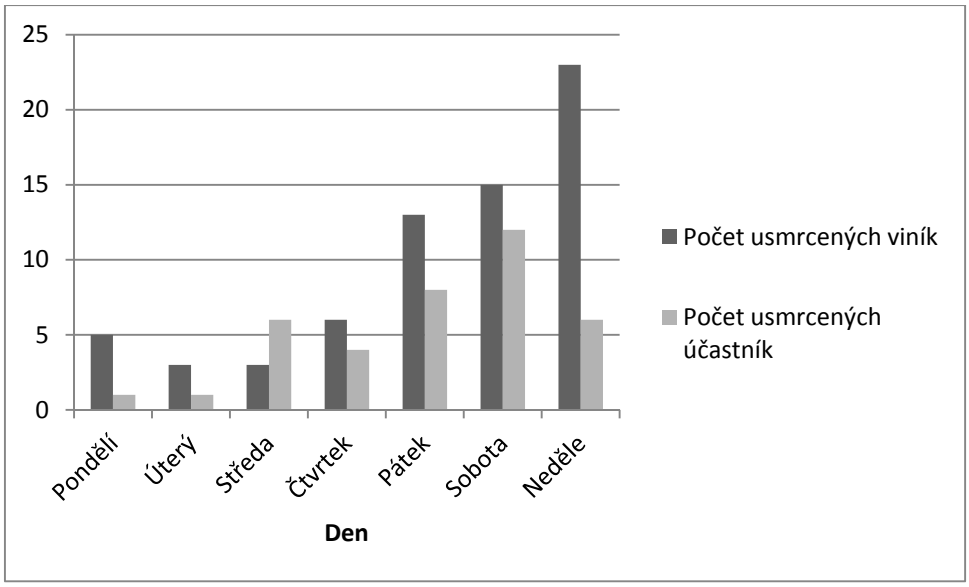
Den	Počet nehod	Počet usmrcených
Pondělí	173	5
Úterý	126	3
Středa	154	3
Čtvrtek	187	6
Pátek	225	13
Sobota	312	15
Neděle	321	23

Tab.1.1.6. Počet nehod a počet usmrcených pro účastníky nehod podle dne v týdnu

Den	Počet nehod	Počet usmrcených
Pondělí	164	1
Úterý	143	1
Středa	126	6
Čtvrtek	168	4
Pátek	228	8
Sobota	208	12
Neděle	163	6



Obr.1.1.9. Srovnání počtu nehod pro viníka a počtu nehod pro účastníka podle dne v týdnu



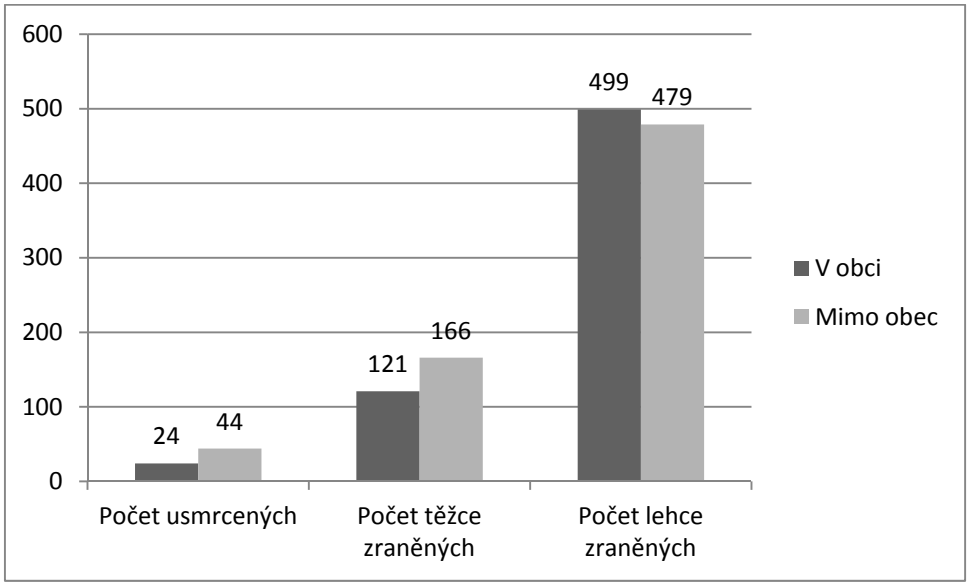
Obr.1.1.10. Srovnání počtu usmrcených pro viníka a počtu usmrcených pro účastníka podle dne v týdnu

Co se týká viníků nehod, tak nejvíce nehod připadá na víkend a pátek, to stejné platí pro počet usmrcených. Víkend společně s pátkem tvoří 57,3 % všech nehod a 75% všech usmrcených. Nejtragičtějším dnem je neděle.

Co se týká účastníků nehod, tak co se týče počtu nehod i počtu usmrcených vede pátek a sobota. Pondělí, čtvrtek i neděle jsou zhruba stejně vyrovnané, co se do počtu

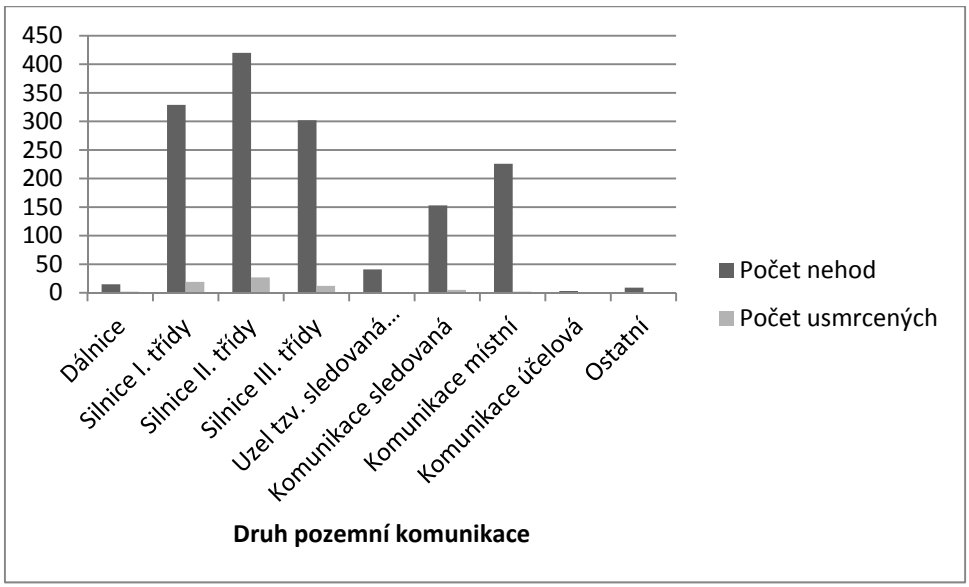
nehod týče, ovšem počet usmrcených je vyšší v neděli a ve středu. Zhruba polovinu počtu nehod tvoří pátek, sobota a neděle a vzhledem k počtu usmrcených se jedná o 68,4 %.

Rozdělím nehody podle místa a to na obec a mimo obec. K tomu v grafu uvedu počet usmrcených, těžce zraněných a lehce zraněných osob. Tento graf bude pouze pro viníky.



Obr.1.1.11. Počet usmrcených, zraněných osob podle místa nehody pro viníka nehody

Více nehod bylo zaznamenáno v obci, avšak více usmrcených náleží k nehodám mimo obec. Budu pokračovat grafem, který bude zaznamenávat druh komunikace, počet nehod a počet usmrcených.



Obr.1.1.12. Počet nehod a počet usmrcených podle druhu komunikace pro viníky

28 % všech nehod připadá na silnice II. třídy, 22 % všech nehod způsobených řidiči motocyklů připadá na silnice I. třídy. Nejvíce usmrcených (skoro celých 40 %) připadá na silnice II. třídy. Počet usmrcených na silnicích I., II., III. třídy tvoří 85 % všech usmrcených při nehodách.

2. Teoretické pojmy

Jelikož pracuji s diskrétními statistickými znaky, existuje nejjednodušší způsob jak otestovat vztahy mezi těmito náhodnými veličinami a tím jsou kontingenční tabulky. Budu se zabývat nezávislostí náhodných veličin a to pomocí Chí-kvadrát testu nezávislosti.

V následujících podkapitolách bude využito zdroje [1].

2.1. Testování parametrických hypotéz

Nechť zkoumaná náhodná veličina X má distribuční funkci patřící do známé třídy distribučních funkcí $\{F_X(x;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, tj. rozdělení pravděpodobností znaku X závisí na neznámém parametru, o kterém víme, že patří do parametrického prostoru $\Theta \subset R^k$. Na základě n nezávislých pozorování znaku X se lze domnívat, že parametr patří do nějaké vlastní podmnožiny množiny Θ, což znamená, že $\boldsymbol{\theta} \in \Theta_0 \subset \Theta$, kde $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$. Tvrzení, že $\boldsymbol{\theta} \in \Theta_0$ nazveme *nulovou hypotézou* (zapisujeme jako H_0), zatímco $\boldsymbol{\theta} \in \Theta_1$ nazveme *alternativní hypotézou* (zapisujeme jako H_A).

Postup, kterým na základě výsledků experimentu dospějeme k rozhodnutí o nulové hypotéze, nazýváme *test hypotézy*. Rozhodnutí o H_0 je následující:

a) H_0 se zamítá ve prospěch alternativy,

b) H_0 nelze zamítnout.

Chyby při rozhodování o H_0 jsou dvojího druhu, což uvedu v následující tabulce.

Tab.2.1.1. Stanovení nulové hypotézy

	H_0 je správná	**H_0 je chybná**
H_0 zamítneme	chyba I. druhu	správné rozhodnutí
H_0 nezamítneme	správné rozhodnutí	chyba II. druhu

Pokud máme vysloveny předpoklady o rozdělení pravděpodobností zkoumané náhodné veličiny X a jsou-li formulovány H_0 a H_A, zvolíme vhodnou výběrovou funkci $T=T(X_1,\ldots,X_n)$, které říkáme *testovací statistika*. Za předpokladu, že H_0 je správná,

musíme znát rozdělení pravděpodobností statistiky T. Hodnotu testovací statistiky t dostaneme při dosazení pozorované hodnoty **x** náhodného výběru.

Kritickým oborem nazveme množinu všech hodnot testového kritéria, kterou označíme $W \subset R^1$, při kterých budeme H_0 zamítat. Kritický obor W volíme tak, abychom omezili pravděpodobnost chyby I. druhu nějakým pevně zvoleným malým číslem α, $0 < \alpha < 1$, které budeme nazývat *hladina testu*. Platnou hypotézu budeme tedy zamítat nejvýše s pravděpodobností α. Zpravidla volíme $\alpha = 0{,}05$ nebo $\alpha = 0{,}01$. Já ve všech případech budu volit $\alpha = 0{,}05$.

Číslo $\sup_{\theta \in \Theta 0} P(T \in W)$ se nazývá *velikost testu*. Je to maximální pravděpodobnost zamítnutí H_0, je-li H_0 správná. Pokud je $\sup_{\theta \in \Theta 0} P(T \in W) \leq \alpha$, má test hladinu α, potom zamítáme nulovou hypotézu H_0.

Dalším důležitým pojmem je *p-hodnota*, kterou získáváme při práci se statistickým softwarem, a jedná se o nejmenší hladinu, při které bychom ještě hypotézu zamítli. Vyjadřuje pravděpodobnost spočítanou za platnosti H_0, že dostaneme právě hodnotu $t = T(\mathbf{x})$, anebo hodnotu ještě víc odporující testované hypotéze. Pokud je *p-hodnota* menší nebo rovna α, pak H_0 zamítáme na hladině testu α, pokud je *p-hodnota* větší než α, pak H_0 na dané hladině nelze zamítnout.

2.2. Kontingenční tabulky

Uvažujeme dvourozměrný náhodný vektor se složkami (náhodnými veličinami) X,Y, které budou nabývat hodnot *1,...r* a *1,...,s* s pravděpodobnostmi $p_{ij} = P(X = i, Y = j)$, kde $i = 1,...,r$ a $j = 1,...,s$. Označíme

$$p_{i.} = P(X = i) = \sum_{j=1}^{s} P(X = i, Y = j)$$

a

$$p_{.j} = P(Y = j) = \sum_{i=1}^{r} P(X = i, Y = j).$$

Dále označíme symbolem n_{ij} četnost jevu $(X = i, Y = j)$ při provedení dvourozměrného náhodného výběru $(X_1, Y_1),\ldots,(X_n, Y_n)$, příslušného náhodného vektoru (X, Y), a pro marginální četnosti zavedeme označení

$$n_{i.} = \sum_{j=1}^{s} n_{ij},\ n_{.j} = \sum_{i=1}^{r} n_{ij}.$$

Vše zapíšeme do *kontingenční tabulky* (Tab.2.2.1.), která bude vypadat následovně:

Tab.2.2.1. Kontingenční tabulka

X\Y	**1**	**2**	**3**	**…**	**s**	∑
1	n_{11}	n_{12}	n_{13}	…	n_{1s}	$n_{1.}$
2	n_{21}	n_{22}	n_{23}		n_{2s}	$n_{2.}$
⋮	⋮	⋮	⋮		⋮	⋮
r	n_{r1}	n_{r2}	n_{r3}	…	n_{rs}	$n_{r.}$
∑	$n_{.1}$	$n_{.2}$	$n_{.3}$	…	$n_{.s}$	n

Výběrové hodnoty $(x_1, y_1),\ldots,(x_n, y_n)$ roztřídíme do *rs* skupin určených *r* hodnotami X a *s* hodnotami Y. Potom platí

$$n = \sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{s} n_{.j} = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij}.$$

Protože n-krát nezávisle opakujeme pokus s *rs* možnými výsledky s pravděpodobnostmi $p_{11}, p_{12},\ldots, p_{rs}$, četnosti v kontingenční tabulce tak vyjadřují realizaci náhodného vektoru s multinomickým rozdělením s parametry $p_{11}, p_{12},\ldots, p_{rs}, n$.

2.3. Multinomické rozdělení

Uvažujeme *n* nezávislých pokusů, z nichž symbolem X_j označíme počet dílčích pokusů, které jsme vybrali v těchto n pokusech, j = 1,…,k. Takto vzniklý náhodný vektor $\boldsymbol{X} = (X_1,\ldots,X_k)'$ má *multinomické rozdělení s parametry* $p_1,\ldots,p_k,n$. Což lze zapsat jako

$$P(X_1 = x_1,\ldots,X_k = x_k) = \begin{cases} \frac{n!}{x_1!x_2!\ldots x_{k!}} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}, & x_1 + \cdots + x_k = n, \\ 0, jinak; \end{cases}$$

kde x_j jsou celá nezáporná čísla. Jednotlivé náhodné veličiny mají binomické rozdělení s parametry *n* a p_j, tj. $X_j \sim Bi(n, p_j)$.

Nechť náhodný vektor **X** má multinomické rozdělení s parametry $n, p_1,\ldots,p_k$. Potom náhodná veličina

$$\sum_{j=1}^{k} \frac{(X_j - np_j)^2}{np_j}$$

má při $n \to \infty$ asymptoticky rozdělení χ^2_{k-1}.

2.4. Chi-kvadrát test nezávislosti

Nejčastější úlohou je provedení testu hypotézy H_0, že veličiny X a Y jsou nezávislé. Tedy platí-li

$$p_{ij} = P(X = i, Y = j) = P(X = i) \,.\, P(Y = j) = p_{i.}p_{.j}, \forall\, 1 \le i \le r, \forall\, 1 \le j \le s.$$

Budou platit následující vztahy

$$p_{r.} = P(X = r) = 1 - \sum_{i=1}^{r-1} p_{i.}, \; p_{.s} = P(Y = s) = 1 - \sum_{j=1}^{s-1} p_{.j}.$$

Za předpokladu platnosti hypotézy H_0 si vystačíme s pravděpodobnostmi $p_{1.},\ldots, p_{r-1,.}, p_{.1},\ldots,p_{.,s-1}$, protože zbylé umíme dopočítat. Tyto neznámé parametry odhadneme metodou minimálního χ^2 a to následovně

$$\widehat{p_{i.}} = \frac{n_{i.}}{n}, \; \widehat{p_{.j}} = \frac{n_{.j}}{n}, \; i = 1,\ldots,r-1, j = 1,\ldots,s-1.$$

Odtud $\widehat{p_{r.}} = 1 - \sum_{i=1}^{r-1} \widehat{p_{i.}} = \frac{n_{r.}}{n}$, $\widehat{p_{.s}} = 1 - \sum_{j=1}^{s-1} \widehat{p_{.j}} = \frac{n_{.s}}{n}$ a následným dosazením do uvedených odhadů funkcí $p_{ij} = p_{i.}p_{.j}$ dostaneme testovou statistiku pro tento *test nezávislosti* ve tvaru

$$Z = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n}.$$

Za platnosti nulové hypotézy H_0 o nezávislosti veličin X, Y má statistika Z asymptoticky pro $n \to \infty$ rozdělení χ^2 a počet stupňů volnosti je roven $(r-1)(s-1)$. Hypotézu nezávislosti zamítáme, pokud $z \ge \chi^2_{(r-1)(s-1),1-\alpha}$, tedy realizuje-li se testová statistika Z v kritickém oboru $W = \langle \chi^2_{(r-1)(s-1),1-\alpha}, \infty)$. Musíme mít ovšem splněnu podmínku dostatečné četnosti, kde existuje pravidlo $n\widehat{p_{ij}} = \frac{n_{i.}n_{.j}}{n} \ge 5, \forall i,j$.

2.5. Test homogenity

Jsou-li řádkové součty $n_{i.}$ v kontingenční tabulce pevně zadány, lze její řádky pokládat za r výběrů z multinomického rozdělení s danými parametry $n_{1.},\ldots,n_{r.}$. Potom

testujeme hypotézu, že příslušná multinomická rozdělení mají stejné pravděpodobnosti, což můžeme zapsat následovně

$$H_0: p_{i1} = p_1, \ldots, p_{is} = p_s, \forall i = 1, \ldots, r,$$

kde $p_1, \ldots, p_s$ nejsou známé. Alternativou v tomto případě bude, že alespoň jedna z uvedených rovností nebude platit. Součty $n_{i.} = \sum_{j=1}^{s} n_{ij}$ jsou pevné, za platnosti nulové hypotézy má statistika Z asymptoticky pro $n \to \infty$ zase χ^2 rozdělení o $(r-1)(s-1)$ stupních volnosti.

2.6. Fisherův faktoriálový test

Tento test používáme ve statistice při testování v kontingenčních tabulkách, kde nejsou splněny podmínky četností. Proto využíváme Fisherův faktoriálový test při malých četnostech. Podmíněná pravděpodobnost p kontingenční tabulky

X/Y	**1**	**2**	∑
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
∑	$n_{.1}$	$n_{.2}$	n

s četnostmi n_{11}, n_{12}, n_{21}, n_{22} při zadaných marginálních četnostech $n_{1.}$, $n_{2.}$, $n_{.1}$, $n_{.2}$ je za předpokladu nezávislosti $p_{ij} = p_{i.}p_{.j}$, i,j = 1,2 rovna číslu

$$p = \frac{\binom{n_{1.}}{n_{11}}\binom{n_{2.}}{n_{21}}}{\binom{n}{n_{.1}}}.$$

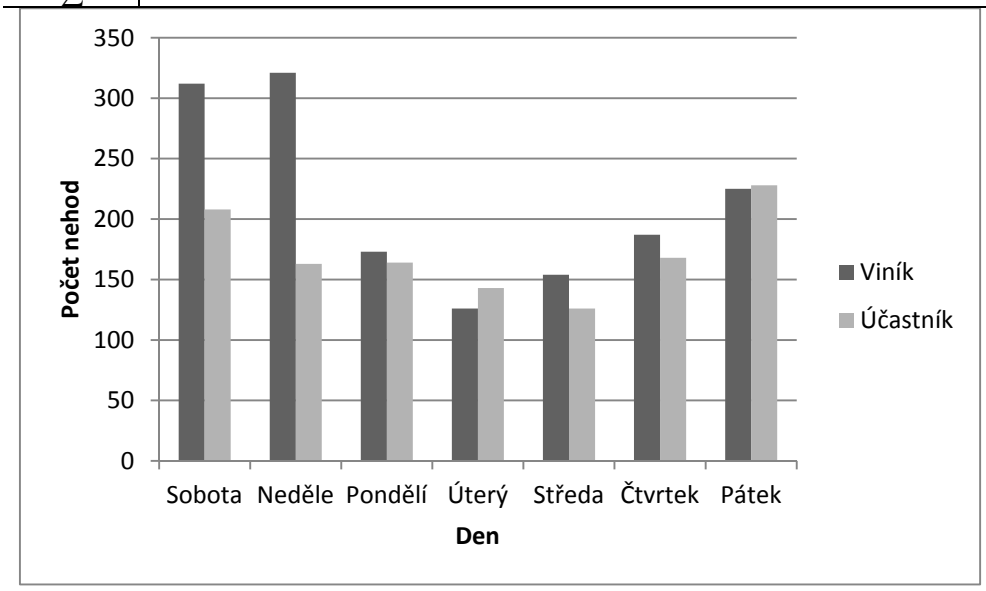
2.7. Příklad s využitím softwaru R

V této kapitole se budu věnovat spojení teorie uvedené výše se softwarem R a popíšu, jak software funguje na následujícím příkladu. Zabývala jsem se testem nezávislosti pro veličinu Den (Tab.1.3.). Zastoupení nehod v týdnu pro viníka vlevo, pro účastníka vpravo znázorňuje obrázek Obr.2.6.1.. Cílem bylo zjistit, zda má náhodná veličina X (viník=1, účastník=2) vliv na den v týdnu (náhodná veličina Y, sobota=1, neděle=2, pondělí=3, úterý=4, středa=5, čtvrtek=6, pátek=7). Nulová hypotéza je v tomto

případě, že tyto dvě veličiny jsou nezávislé. Alternativou bude, že náhodné veličiny budou závislé, tudíž spolu budou veličiny souviset, ovlivňovat se. Pokud budu nulovou hypotézu zamítat ve prospěch alternativy, bude to znamenat, že náhodné veličiny X a Y spolu souvisí, tudíž lze tvrdit, že náhodná veličina X bude mít vliv na den v týdnu. Kdežto v případě, že nulová hypotéza nelze zamítnout, můžeme tvrdit, že na hladině 5% hladině nulovou hypotézu o nezávislosti nelze zamítnout. Což budu zkoumat pomocí kontingenční tabulky (Tab.2.6.1.), kterou si pro tento příklad vytvořím.

Tab.2.6.1.

X/Y	1	2	3	4	5	6	7	∑
1	312	321	173	126	154	187	225	1498
2	208	163	164	143	126	168	228	1200
∑	520	484	337	269	280	355	453	2698



Obr. 2.6.1. Počet nehod podle dne v týdnu pro viníka a účastníka dopravní nehody

První na co se musím soustředit je, zda bude podmínka četnosti splněna, tedy $n\widehat{p_{ij}} = \frac{n_{i.}n_{.j}}{n} \geq 5$, $\forall i,j$, což například znamená $n\widehat{p_{11}} = \frac{n_{1.}n_{.1}}{n} = \frac{520.1428}{2698} = 288{,}72$ a z toho lze soudit, že četnosti budou splněny pro všechny zbývající hodnoty v tabulce. Tuto kontingenční tabulku zadám v softwaru pomocí tohoto příkazu

```
t <- matrix(c( 312,208,321,163,173,164,126,143,154,126,187,168,225,228),nrow=2).
```

Dosazení ruční do testovací statistiky Z by vypadalo následovně

$$z = \frac{\left(312 - \frac{520.1498}{2698}\right)^2}{\frac{520.1498}{2698}} + \ldots + \frac{\left(228 - \frac{453.1200}{2698}\right)^2}{\frac{453.1200}{2698}} = 45,\ 176,$$

kde je výsledek ovšem ovlivněn chybami při zaokrouhlování. A dále je potřeba určit ($r - 1$)($s - 1$) = (2 – 1)(7 – 1) = 6, proto výsledek realizace statistiky Z budeme srovnávat s hodnotou kvantilu $\chi^2_{6;0,95}$ = 12, 592. Kritický obor bude potom ve tvaru W = < 12, 592; ∞) a hodnota z testovací statistiky do kritického oboru patří, z toho důvodu lze nulovou hypotézu na hladině testu $\alpha = 0{,}05$ zamítnout.

Při použití softwaru R zadáme následující příkaz a získáme následující informace

chisq.test(t)

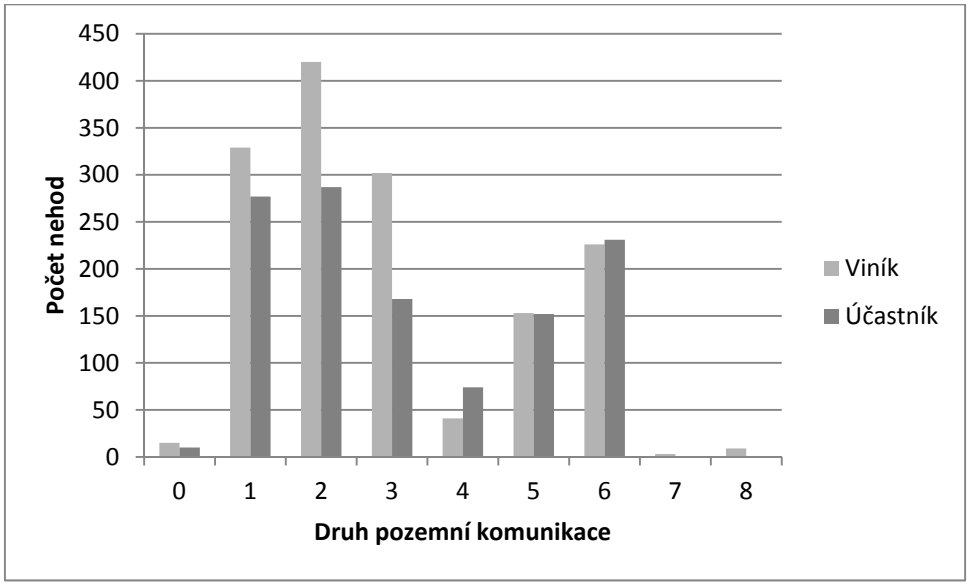
X-squared = 45.1662, df = 6, p-value = 4.337e-08,

kde lze vidět, že při ručním výpočtu jsem se nedopustila velkých chyb. Df udává stupně volnosti a p-hodnota, která nelze vypočítat ručně je zde uvedena a díky ní máme potvrzení toho, že nulovou hypotézu můžeme na dané hladině $\alpha = 0{,}05$ zamítnout, protože je splněno 4.337e-08 ≤ 0,05. Následný výsledek lze interpretovat tak, že je nutno hypotézu o nezávislosti na dané hladině zamítnout. Tudíž lze tvrdit, že náhodná veličina X souvisí se dnem v týdnu, v jakém se dopravní nehoda uskutečnila.

3. Výzkum a analýza dat

3.1. Druh pozemní komunikace

První se budu věnovat proměnné Druh pozemní komunikace, uvedené v tabulce Tab.1.1..Náhodná veličina X bude viník=1, účastník=2 a náhodná veličina Y bude 0=dálnice, 1=silnice I. třídy, 2=silnice II. třídy, 3=silnice III. třídy, 4=křižovatka, 5=sledovaná silnice, 6=místní, 7=účelová, 8=ostatní. Bude mě zajímat, zda spolu souvisí, jak viníci a účastníci nehod bourají na stejných druzích pozemních komunikací. Pomocí kontingenčních tabulek a testu nezávislosti rozhodnu o výsledku. Očekávala bych, že spolu proměnné souvisí (což můžu odhadovat i podle Obr.3.3.1.). První zde proto uvedu obrázek pro řidiče motocyklů, znovu rozdělené podle dopravní nehody na viníky a účastníky (Obr.3.3.1.).



Obr.3.3.1. Počet nehod podle druhu pozemní komunikace pro viníky a účastníky

Jde si všimnout, že rozložení se moc neliší, i když jsou viníci mnohem více zastoupeni než účastníci. V tabulce to bude vypadat následovně

Tab.3.3.1. Kontingenční tabulka pro viníky a účastníky podle druhu pozemní komunikace

P36	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	∑
Viník	15	329	420	302	41	153	226	3	9	1498
Účastník	10	277	287	168	74	152	231	0	1	1210
∑	25	606	707	470	115	305	457	3	10	2708

Jelikož poslední dva sloupce nesplňují předpoklady pro test nezávislosti, nebudou brány do úvahy. Budu tedy testovat nulovou hypotézu, že nehody (podle druhu pozemní komunikace) způsobené viníkem a účastníkem jsou nezávislé. Použiju statistiku Z k testu nezávislosti. Dostávám, že hodnota testové statistiky z = 48.0856 a kritický obor určím jako W = < 12,592; ∞) na hladině $\alpha = 0{,}05$ při 6 stupních volnosti, což znamená že H_0 lze na dané hladině zamítnout. Tudíž předpoklad byl správný a lze tvrdit, že druh pozemní komunikace souvisí s náhodnou veličinou X. P-hodnota je rovna 1.136e-08 ≤ 0,05, která samozřejmě dokládá, že nulovou hypotézu zamítáme.

Co se týká programování v R, tak příkazy vypadají následovně:

```
chisq.test(matrix(c(15,10,329,277,420,287,302,168,41,74,153,152,226,231),nrow=2))
```

Jelikož musím brát v úvahu i mopedy a motocykly do 50 ccm, srovnala jsem stejným způsobem následující dvojice (moped, motocykl do 50 ccm) pro viníky a pro účastníky, taky pro viníky samostatně a účastníky samostatně.

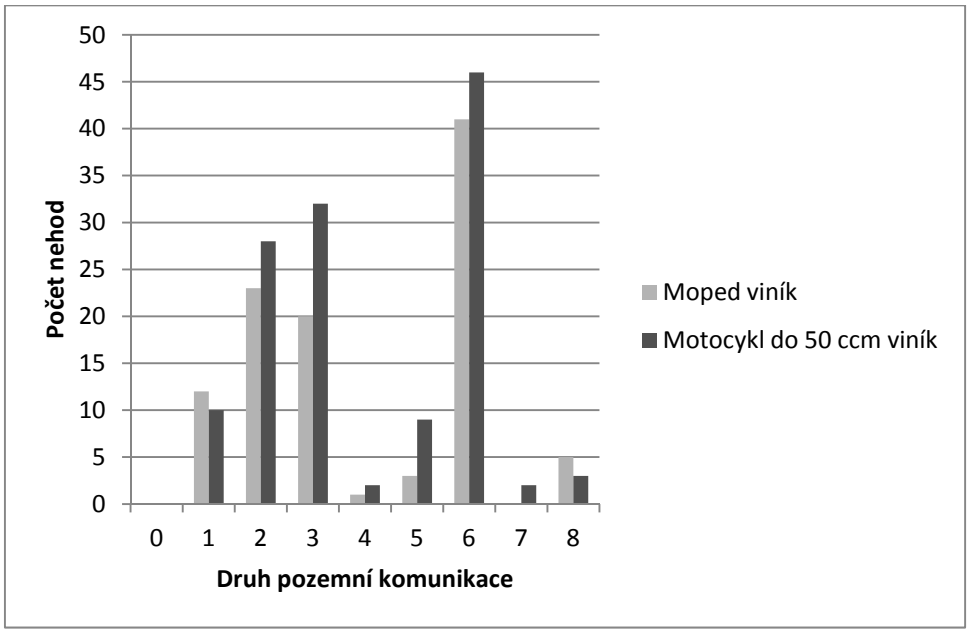
Tab.3.1.2.Viník moped, účastník moped

P36	**1**	**2**	**3**	**6**	∑
Viník	12	23	20	41	96
Účastník	7	16	10	18	51
∑	19	39	30	59	147

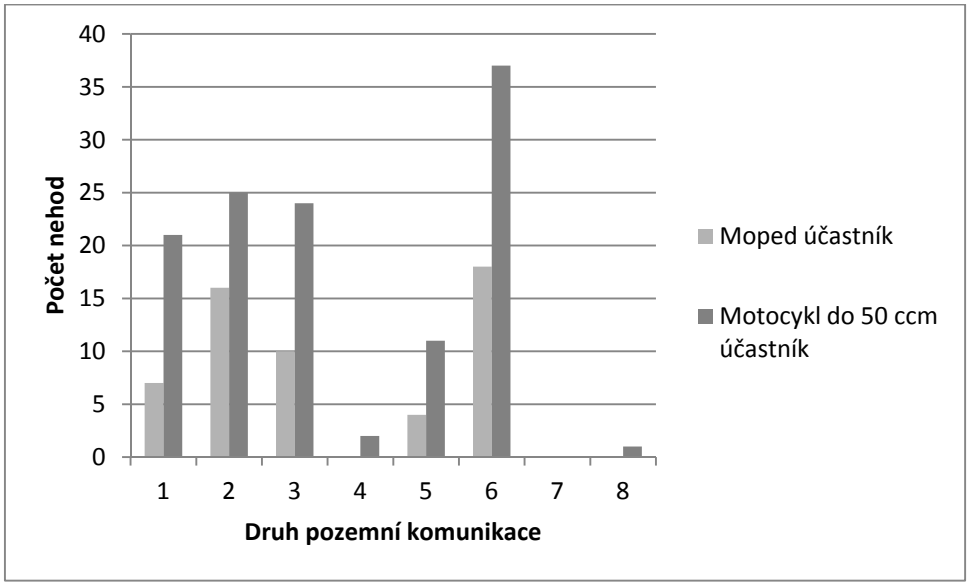
V tomto případě bylo nutné tabulku upravit pro nevyhovující předpoklady. V tomto případě je kritický obor W = < 7,815; ∞) při 3 stupních volnosti a $\alpha = 0{,}05$ vyšla hodnota testovací statistiky z = 1.2095 a p-hodnota je 0.7507. V tomto případě nelze H_0 zamítnout. Takže lze tvrdit, že mezi účastníkem nehod na mopedu a viníkem nehod na mopedu není souvislost, jsou nezávislé.

I v dalších případech jsem byla nucena tabulky upravit pro nevyhovující četnosti. Podle toho se kritický obor lišil, ale při všech zbývajících možnostech jsem H_0 nemohla zamítnout. Tudíž z toho mohu usuzovat, že mezi následujícími dvojicemi (motocykl do 50 ccm účastník i viník), (motocykl do 50 ccm účastník a viník moped), (motocykl do 50 ccm viník a moped účastník) můžu soudit, že mezi nimi není souvislost. Přitom se mohu jen domnívat, že je to způsobeno nízkými hodnotami u těchto uvedených dvojic, jelikož při

pohledu na grafy byly předpoklady naprosto odlišné. Proto v dalších příkladech budu brát do úvahy pouze motocykly.



Obr.3.1.2. Viníci (moped, motocykl do 50ccm)



Obr.3.1.3. Účastníci (moped, motocykl do 50 ccm)

3.2. Další testy nezávislosti pro viníky a účastníky nehod motocyklů

V této podkapitole se zaměřím na Počet usmrcených osob, Počet těžce zraněných osob a Počet lehce zraněných osob, kde bude nulová hypotéza H_0 opět taková, že náhodné veličiny jsou nezávislé. Opět zůstává náhodná veličina X (1=viník, 2= účastník) a

náhodnou veličinou Y bude počet osob, bud usmrcených, těžce zraněných či lehce zraněných. Zajímá mě, zda souvisí náhodná veličina X s náhodnou veličinou Y, tedy zda souvisí počet usmrcených (počet těžce zraněných, počet lehce zraněných) s tím, kdo způsobil či se účastnil nehody. Tentokrát pouze u motocyklů. Vše si uspořádám do přehledných kontingenčních tabulek (Tab.3.2.1.,Tab.3.2.2.,Tab.3.2.3.).

Tab.3.2.1. Kontingenční tabulka pro Počet usmrcených osob

Usmrceno osob	**0**	**1**	**2**	∑
Viník	1430	66	2	1498
Účastník	1162	36	2	1200
∑	2592	102	4	2698

V tomto případě opět musím vynechat poslední sloupec z důvodu nedodržení četností a získávám čtyřpolní tabulku. Určím si kritický obor $W = < 3{,}841; \infty)$ a hodnota testovací statistiky vyšla z = 3.2381 na hladině $\alpha = 0{,}05$ a p-hodnota je 0.07194. Z tohoto výsledku lze usoudit, že nelze H_0 na dané hladině zamítnout. Což vlastně znamená, že na 5% hladině nelze hypotézu H_0 o nezávislosti zamítnout a interpretaci může být, že počet usmrcených osob nesouvisí s člověkem, jenž byl u dopravní nehody přítomen.

Tab.3.2.2. Kontingenční tabulka pro Počet těžce zraněných osob

Těžce zraněno osob	**0**	**1**	**2**	**3**	∑
Viník	1211	265	21	1	1498
Účastník	961	218	17	4	1200
∑	2172	483	38	5	2698

V tomto je výsledkem kritický obor $W = < 5{,}991; \infty)$ s dvěma stupni volnostmi na hladině $\alpha = 0{,}05$, hodnota testovací statistiky z = 0.1283 a p-hodnota = 0.9379, což znamená, že na dané hladině nelze nulovou hypotézu zamítnout. Jinak řečeno, nelze říct, že by spolu počet těžce zraněných osob a člověk vyskytující se u dopravní nehody souviseli.

A posledním příkladem bude počet lehce zraněných, opět uspořádán do tabulky uvedené na další straně pro přehlednost.

Tab.3.2.3.Kontingenční tabulka pro Počet lehce zraněných osob

Lehce zraněno osob	**0**	**1**	**2**	**3**	**4**	∑
Viník	520	844	127	5	2	1498
Účastník	395	648	140	15	2	1200
∑	915	1492	267	20	4	2698

Tentokrát bude kritický obor W = < 7,815; ∞) s 3 stupni volnosti na hladině testu $\alpha = 0{,}05$, výsledkem bude hodnota testovací statistiky z = 15.6858 a p-hodnota = 0.001315, z čehož vyplývá, že v tomto případě nulovou hypotézu H_0 lze zamítnout, což znamená, že mezi viníkem společně s účastníkem dopravní nehody a počtem lehce zraněných osob existuje souvislost.

4. Závěr

Nejprve ze všeho jsem si data upravila v Excelu a rozdělila je na 6 souborů podle typu motorového vozidla na moped, motocykl do 50 ccm a motocykl a tyto tři kategorie jsem rozdělila ještě podle toho, zda byla nehoda způsobena (viník) či zda se jí dotyčný pouze účastnil (účastník). Dále jsem se rozhodla, které proměnné dále využiji a které nemají žádnou vypovídající hodnotu, jež jsem odstranila. Použila jsem i shrnutí Policie ČR, které srovnávalo rok 2010 s rokem předešlým, pro popisnou statistiku.

Nejdůležitějším ukazatelem pro mě představoval počet usmrcených, počet těžce zraněných, počet lehce zraněných a počet nehod. Většinou jsem tyto kategorie uváděla společně v souvislosti s jinými kategoriemi jako druh pozemní komunikace, rok výroby motorového vozidla, den v týdnu, věk a v převážném počtu pro řidiče motocyklů, kteří nehodu způsobili a pro řidiče motocyklů, kteří se nehod účastnili a to z toho důvodu, že tato data měly lepší vypovídající hodnoty než data pro mopedy a motocykly do 50 ccm, jichž bylo bohužel málo. Což se domnívám bylo způsobeno tím, že pro zavolání Policie ČR byla škoda způsobená na mopedu nebo motocyklu do 50 ccm nízká a získaná data pocházela především z nehod, kde byla způsobená větší škoda nebo zranění. Proto jsem se soustředila na řidiče motocyklů. Má očekávání byla taková, že mezi účastníkem a viníkem dopravní nehody musí existovat rozdíl. Z všeobecného mínění jsem vycházela z toho, že řidiči motocyklů (viníci) jsou převážně mladší řidiči novějších motocyklů, kteří o víkendu vyjedou za pěkného počasí na klikaté cesty, kde se snaží dostatečně projet svůj motocykl. Podobnou skladbu bych očekávala i u řidičů motocyklů (účastníků) s tím rozdílem, že příčina nehody bude zcela odlišná jako například přehlédnutí řidičem osobního vozidla jak v běžném provozu, tak při odbočování atd..

V neposlední řadě jsem se zabývala analýzou v softwaru R, kde jsem používala testy nezávislosti pro to, abych například zjistila, zda existuje souvislost mezi řidičem motocyklů a dnem v týdnu, kdy se nehoda stala. V tomto případě jsem došla k závěru, že zde skutečně souvislost existuje. Na rozdíl od toho pro řidiče mopedů či motocyklů do 50 ccm nelze uvést, že by mezi nimi a druhem pozemní komunikace existovala souvislost. Vysvětluji si to tím, že opravdu lidé používají většinou své motocykly pro radost a ne jako dopravní prostředek k dostání z místa A do místa B.

5. Literatura

[1] Hron, K., Kunderová, P.: Základy počtu pravděpodobnosti a metod matematické statistiky, Olomouc, 2013

[2] Pplk. Petr Sobotka, Ing.Josef Tesařík :Policejní zpráva v Pdf, O nehodovosti na pozemních komunikacích České republiky za rok 2010

[3] Wikipedia, https://cs.wikipedia.org/wiki/Hlavni_strana, zdroj informací pro klíčová slova: moped, řidičské oprávnění, motocykl, skútr, alkohol za volantem

[4] Agresti (1990, p. 61f; 2002, p. 91)