

PŘÍRODOVĚDECKÁ FAKULTA UNIVERZITY PALACKÉHO  
V OLOMOUCI

Katedra matematické analýzy a aplikací matematiky  
školní rok 2012/2013

DIPLOMOVÁ PRÁCE

Analýza dopravních nehod ve městě Zlín  
pomocí zobecněného lineárního modelu



Vedoucí diplomové práce:  
Mgr. Ondřej Vencálek, Ph.D.  
Rok odevzdání: 2012

Vypracovala:  
Bc. Monika Dírerová  
AME, II. ročník

## **Prohlášení**

Prohlašuji, že jsem vypracovala tuto diplomovou práci samostatně pod vedením Mgr. Ondřeje Vencálka, Ph.D., a že jsem uvedla v seznamu použité literatury všechny zdroje, se kterými jsem v této práci pracovala.

V Olomouci dne 10. prosince 2012

.....  
Dírerová Monika

## **Poděkování**

Ráda bych na tomto místě poděkovala vedoucímu diplomové práce Mgr. Ondřeji Vencálkovi, Ph.D. za spolupráci, ochotu a obzvlášť za čas věnovaný konzultacím této práce. Dále bych ráda poděkovala své rodině a přátelům, kteří mě podporovali po celou dobu studia.

# OBSAH

OBSAH .....	3
ABSTRAKT.....	5
ABSTRACT .....	6
SEZNAM SYMBOLŮ A ZKRATEK .....	7
ÚVOD .....	8
1 ZÁKON O SILNIČNÍM PROVOZU Č. 361/2000 SB.....	10
1.1 Základní pojmy dopravní problematiky, §2 .....	10
1.2 Podmínky pro řízení motorového vozidla .....	12
1.3 Dopravní nehoda, §47 .....	14
2 NEHODOVOST V ČR .....	16
2.1 ČR a země EU.....	19
2.2 Ministerstvo dopravy .....	20
3 MĚSTO ZLÍN .....	22
3.1 Historie a současnost .....	22
3.2 Doprava a silnice .....	23
4 CÍLE.....	25
5 ZÁKLADNÍ STATISTIKA .....	26
5.1 Základní pojmy .....	26
5.2 Diskrétní rozdělení.....	28
5.2.1 Poissonovo rozdělení.....	28
5.2.2 Multinomické rozdělení.....	29
5.2.3 Vztah mezi Poissonovým a multinomickým rozdělením .....	30
5.3 Kategoriální proměnná .....	32
5.4 Kontingenční tabulky.....	34
5.5 Klasické lineární regresní modely .....	35
5.5.1 Odhad parametrů a jejich vlastnosti .....	36
5.5.1.1 Metoda nejmenších čtverců (MNČ) .....	37
5.5.1.2 Metoda maximální věrohodnosti .....	37
5.5.1.3 Vlastnosti odhadu $\beta$ .....	38
6 ZOBECNĚNÉ LINEÁRNÍ MODEL Y .....	39

6.1	Složky zobecněných lineárních modelů .....	41
6.2	Poissonův loglineární model pro počty událostí.....	42
6.2.1	Odhad parametru $\beta$ .....	43
6.2.1.1	Newton-Raphsonova metoda.....	44
6.2.2	Kovarianční matice odhadu parametru $\beta$ .....	46
6.2.3	Hypotézy o parametru $\beta_j$ a intervaly spolehlivosti.....	46
6.2.4	Deviance .....	48
6.2.5	Akaikeho informační kritérium .....	50
7	UŽITÍ LOGLINEÁRNÍHO MODELU – ANALÝZA DOPRAVNÍ NEHODOVOSTI .....	51
7.1	Faktory charakterizující počet dopravních nehod.....	52
7.1.1	Dny v týdnu .....	52
7.1.2	Dny v týdnu a měsíc vzniku dopravní nehody .....	56
7.1.3	Dny v týdnu, měsíc a místo dopravní nehody .....	66
7.1.4	Dny v týdnu, měsíc, místo a druh dopravní nehody .....	69
7.1.5	Dny v týdnu, měsíc, místo, druh a zavinění dopravní nehody .....	75
7.1.6	Dny v týdnu, měsíc, místo, druh nehody, zavinění dopravní nehody a druh vozidla, které zavinilo nehodu.....	79
7.2	Faktory ovlivňující vznik dopravní nehody.....	83
7.2.1	Povětrnostní podmínky.....	83
7.2.2	Povětrnostní podmínky a alkohol .....	85
7.2.3	Povětrnostní podmínky, alkohol a stav povrchu vozovky.....	86
7.2.4	Povětrnostní podmínky, alkohol, stav povrchu vozovky a viditelnost .....	89
	ZÁVĚR.....	94
	POUŽITÁ LITERATURA.....	97
	SEZNAM TABULEK, OBRÁZKŮ A GRAFŮ.....	99
	PŘÍLOHY.....	102

## **ABSTRAKT**

Diplomová práce navazuje na bakalářskou práci, která se zabývala problematikou dopravní nehodovosti na hlavních přístupových komunikacích města Zlín.

V práci jsou uvedeny základní pojmy a údaje týkající se dopravní nehodovosti. Dále bude popsán současný stav nehodovosti v ČR, přístupové pozemní komunikace do Zlína a zpracovány jejich statistiky nehodovosti. Analýza nehodovosti bude provedena pomocí zobecněného lineárního modelu.

Cílem práce je najít nejvýznamnější faktory a vyloučit nejméně významné faktory, které přispěly ke vzniku dopravních nehod ve Zlíně v letech 2004 – 2011.

## **ABSTRACT**

The dissertation will follow the bachelor thesis, which addressed problems with traffic accidents on the main access roads of Zlín.

It will present the basic concepts and information relating to the traffic accident rates. Furthermore, it will provide a summary of the current road traffic accident rates in the Czech Republic, examine access roads to Zlín and analyse their traffic accident rates. The analysis of the data will be calculated using generalised linear model.

The aim of the dissertation is to identify the key factors that contributed to road traffic accidents between the years 2004 – 2011 and isolate those that had little significance.

## SEZNAM SYMBOLŮ A ZKRATEK

kW/kg	kilowatt na kilogram
km/h	kilometr za hodinu
g/l	gram na litr
I	silnice první třídy
II	silnice druhé třídy
III	silnice třetí třídy



# ÚVOD

Nehody jsou součástí našeho života. Setkáváme se s nimi dnes a denně, ačkoliv jsou zřizována nejrůznější opatření k jejich snížení. V posledních letech přibývá smrtelných nehod, což je způsobeno hektickým životním stylem. Člověk neustále někam spěchá, neodpočine si a to způsobuje nepozornost a únavu. Dopravní nehody však nejsou způsobeny pouze nepozorností lidí, mají na ně vliv i jiné faktory, kterými se bude zabývat tato diplomová práce. Budeme zkoumat, jak kombinace různých faktorů ovlivňují vznik dopravních nehod na přístupových cestách města Zlín. Naším úkolem bude najít nejvhodnější model popisující počty dopravních nehod.

V první kapitole budou popsány základní pojmy dopravní nehodovosti vybrané ze zákona o silničním provozu č. 361/2000 sb. Rovněž se zde dočtete, za jakých podmínek můžete řídit vozidlo, jaké jsou hlavní povinnosti řidiče a naopak jakých přestupků se řidič nesmí dopustit.

Druhá kapitola se týká České republiky. Můžete zde nalézt informace o nehodovosti v ČR v letech 1993 – 2011, což znamená, jak se nehodovost vyvíjela, jak změna zákona o nahlášení nehody policii ČR ovlivnila počet nehod, a které faktory způsobily nejvíce nehod. Dále jsme srovnali počet nehod v zemích EU s Českou republikou. Kapitulu zakončuje přehled nastupujících ministrů dopravy v období 1992 – 2011.

Jelikož se budeme zabývat nehodami ve městě Zlín, nezapomeneme o něm zmínit základní informace, čemuž se věnuje kapitola třetí. Naleznete zde pár hlavních událostí z historie ale i současnosti města Zlín. V závěru jsou uvedeny informace o dopravě a silnicích, kterých se práce týká.

Čtvrtá kapitola vytyčuje cíle práce, čímž jsme zakončili všeobecné informace o dopravě.

V páté kapitole se nachází přehled základní statistiky, která je potřebná pro další práci. Kapitola začíná základními pojmy a poté přecházíme k diskrétním rozdělením, která jsou pro tuto práci stěžejní. Budeme pracovat s kategoriálními daty, proto vysvětlujeme také tento pojem. Cílem práce je zjistit vztahy mezi faktory, proto dále popisujeme kontingenční tabulky. Konec kapitoly Základní statistika je věnován klasickým lineárním regresním modelům, což je přípravou pro následující kapitolu.

Z názvu práce jste zjistili, že budeme pracovat se zobecněným lineárním modelem, kterým se zabývá kapitola šestá. V úvodu využijeme pojmy z páté kapitoly. Dále se dozvíte, jak zobecněný lineární model vypadá, čímž můžeme přejít k objasnění a vlastnostem Poissonova loglineárního modelu. Pomocí kapitoly šest budeme řešit praktickou část, tedy analyzovat dopravní nehodovost na úsecích přístupových cest do obce Zlín.

Kapitola 7 bude rozdělena do dvou částí. V první části analyzujeme faktory charakterizující počet dopravních nehod a v druhé části budeme pracovat s faktory ovlivňujícími vznik dopravní nehody.

# 1 ZÁKON O SILNIČNÍM PROVOZU Č. 361/2000 SB.

Veškeré informace o dopravě získáme v zákoně č. 361/2000 Sb. ze dne 14. září 2000 o provozu na pozemních komunikacích a o změnách některých zákonů (zákon o silničním provozu). Zákon vysvětluje základní pojmy spojené s pozemními komunikacemi, vytyčuje povinnost účastníků silničního provozu a vymezuje právní předpisy, které souvisí s pohybem na pozemních komunikacích. Dále uvádí veškeré informace související s řídičským oprávněním a řídičským průkazem, registrem řidičů, bodovém hodnocení, atd.

První kapitola je zaměřena na některé části uvedeného zákona. Jejím účelem je objasnění pojmů, které se v práci vyskytnou [1].

## 1.1 Základní pojmy dopravní problematiky, §2

Nejprve nadefinujeme některé nejdůležitější pojmy, které jsou běžně používány, ovšem někdy špatně chápané, v lidském životě, a příkazy, které bývají často porušovány. Dané pojmy i příkazy budou rovněž použity v diplomové práci.

1) *Každého, kdo se přímým způsobem účastní provozu na pozemních komunikacích, označujeme za účastníka provozu na pozemních komunikacích.*

2) *Vlastník vozidla, jiná fyzická nebo právnická osoba zmocněná vlastníkem k provozování vozidla vlastním jménem je provozovatel vozidla.*

3) *Účastníka provozu na pozemních komunikacích, který řídí motorové nebo nemotorové vozidlo anebo tramvaj, nazýváme řidič. Je jím i jezdec na zvířeti.*

4) *Vozidlo je motorové vozidlo, nemotorové vozidlo nebo tramvaj. Motorovým vozidlem nazýváme nekolejové vozidlo poháněné vlastní pohonnou jednotkou a trolejbus. Nemotorové vozidlo je vozidlo pohybující se pomocí lidské nebo zvířecí síly, například jízdní kolo, ruční vozík nebo potahové vozidlo.*

5) *Chodec je i osoba, která tlačí nebo táhne sánky, dětský kočárek, vozík pro invalidy nebo ruční vozík o celkové šířce nepřevyšující 600 mm, pohybuje se na lyžích nebo kolečkových bruslích anebo pomocí ručního nebo motorového vozíku pro invalidy, vede jízdní kolo, motocykl o objemu válců do 50 cm<sup>3</sup>, psa a podobně.*

- 6) *Autobus, trolejbus nebo tramvaj řadíme mezi vozidla hromadné dopravy osob.*
- 7) *Každý účastník provozu se musí chovat tak, aby neohrozil jiného účastníka provozu na pozemních komunikacích a nevzniklo žádné nebezpečí. Rovněž nesmí nikoho omezit.*
- 8) *Stát znamená uvést vozidlo do klidu nad dobu dovolenou pro zastavení. Dát přednost v jízdě znamená povinnost řidiče nezahájit jízdu nebo jízdní úkon nebo v nich nepokračovat, jestliže by řidič, který má přednost v jízdě, musel náhle změnit směr nebo rychlost jízdy.*
- 9) *Silnice pro motorová vozidla je pozemní komunikace označená dopravní značkou "Silnice pro motorová vozidla".*
- 10) *Jízdní pruh označuje část vozovky dovolující jízdu vozidel jiných než dvoukolových (motocyklů) v jednom jízdním proudu za sebou. Část povrchu pozemní komunikace ležící mezi okrajem přilehlého jízdního pruhu a hranou koruny pozemní komunikace se nazývá krajnice. Skládá se zpravidla ze zpevněné a nezpevněné části.*
- 11) *Křižovatka je místo, kde se pozemní komunikace protínají nebo spojují. Vyústění polní nebo lesní cesty nebo jiné účelové pozemní komunikace na jinou pozemní komunikaci není křižovatka.*
- 12) *Místo, kde se úrovně kříží pozemní komunikace se železnicí, popřípadě s jinou dráhou ležící na samostatném tělese, nazýváme železniční přejezd. Je označen příslušnou dopravní značkou.*
- 13) *Přechod pro chodce je místo na pozemní komunikaci určené pro přecházení chodců, vyznačené příslušnou dopravní značkou.*
- 14) *Vše, co by mohlo ohrozit bezpečnost nebo plynulost provozu na pozemních komunikacích, se nazývá překážka provozu na pozemních komunikacích.*
- 15) *Jestliže účastníci provozu na pozemních komunikacích dostatečně zřetelně nerozeznají jiná vozidla, osoby, zvířata nebo předměty na pozemní komunikaci, pak mluvíme o snížené viditelnosti. Je to situace například od soumraku do svítání, za mlhy, sněžení, hustého deště nebo v tunelu.*
- 16) *Zařízení schválené podle zvláštního právního předpisu a určené k zajištění bezpečnosti přepravovaných osob nazýváme zádržný bezpečnostní systém. Je jím bezpečnostní pás nebo dětský zádržný systém (dále jen "dětská autosedačka"). Dětská autosedačka je vybavení schválené podle zvláštního právního předpisu určené k zajištění*

*bezpečnosti přepravovaných dětí, jejichž tělesná hmotnost nepřevyšuje 36 kg a tělesná výška nepřevyšuje 150 cm.*

## **1.2 Podmínky pro řízení motorového vozidla**

Ne každý člověk je způsobilý k řízení motorového vozidla. Pro řízení daného vozidla je nutná tělesná a duševní způsobilost občana (dále zdravotní způsobilost). Je posuzována lékařem, který vydává posudek o zdravotní způsobilosti. Poruchy chování způsobené závislostí na alkoholu nebo jiných psychoaktivních látkách jsou důvodem udělení zdravotní nezpůsobilosti k řízení. Lékařskou prohlídku musí podstupovat každá osoba, která má ve svém pokynu práce řízení vozidla. Dále se musí prohlídce podrobit držitel řidičského oprávnění nejdříve 6 měsíců před dovršením 60, 65 a 68 let věku a nejpozději v den dovršení stanoveného věku, po dovršení 68 let věku pak každé 2 roky.

Dále musí ovládat řízení vozidla a předpisy o provozu na pozemních komunikacích. Získání řidičského oprávnění zaručuje splnění uvedených požadavků, ve kterém je uvedena skupina, která udává, jaké vozidlo může vlastník průkazu řídit [viz. příloha 1]. V České republice je povoleno řídit motorové vozidlo osobě starší 18 let, která má na území České republiky trvalé bydliště, nebo zde alespoň 6 měsíců studuje. Vozidlo, které řidič užívá, musí splňovat technické podmínky stanovené zvláštním právním předpisem.

Existuje spousta nařízení, ve kterých je popsáno, co řidič, spolujezdec při jízdě smí, či nesmí. Jedním z nejčastějších porušení je telefonování při řízení. U řidiče vozidla hromadné obsluhy je rovněž zakázáno jíst, pít nebo se bavit s přepravovanými osobami v době řízení. Řidič i spolujezdec motocyklu nebo mopedu musí za jízdy použít přilbu, kterou má nasazenou a řádně připevněnou na hlavě. Rovněž spousta lidí překračuje povolenou rychlost. V obci smí řidič motorového vozidla o maximální přípustné hmotnosti nepřevyšující 3 500 kg nebo autobusu jet nejvýše rychlostí 50 km/h, mimo obec 90 km/h a na dálnici 130 km/h . Důležitou součástí jízdy jsou rozsvícená světla, přesněji rozsvícená obrysová světla a potkávací světla nebo světla pro denní svícení. Samozřejmostí je dodržování dopravních značek, světelných signálů tříbarevné soustavy, nebo pokynů policistů.

V období od 1. listopadu do 31. března je nutné mít na vozidle zimní pneumatiky, které musí mít hloubku dezénu hlavních dezénových drážek nebo zářezů nejméně 4 mm u vozidel s maximální přípustnou hmotností nepřevyšující 3 500 kg.

Spousta lidí řídí pod vlivem alkoholu nebo návykových látek, což je zakázáno. Pokud člověk požije alkohol, hrozí mu pokuta, odebrání bodů a zákaz řízení. Výše všech tří trestů závisí na výši promile alkoholu v krvi, což je uvedeno v následující tabulce.

TABULKA 1.2.1 Trest za řízení v opilosti [2], [9]

SITUACE	ZÁKON	POKUTA	BODY	ZÁKAZ ŘÍZENÍ
do 0,3 promile	přestupek	2 500 až 20 000Kč	0	6 měsíců až 1 rok
0,3 - 1 promile	přestupek	2 500 až 20 000Kč	7	6 měsíců až 1 rok
více než 1 promile	trestný čin a nepodmíněný trest až 1 rok vězení	25 000 až 50 000Kč	7	1 až 2 roky
odmítnutí vyšetření	-	25 000 až 50 000Kč	7	1 až 2 roky

Jak je vidět tak ať už má člověk jakékoliv množství alkoholu v krvi, odebírá se mu na určitou dobu řidičské oprávnění. Pozor, tabulka není zcela pravdivá. Experti označili hladinu 0,2 promile alkoholu v krvi za možnou i bez požití alkoholického nápoje. Proto policie ČR většinou neodebírá řidičský průkaz, pokud člověk při dechové zkoušce nadýchá 0,24 promile, kdy 4 setiny navíc jsou přidány kvůli kalibračním odchylkám měřících přístrojů. Přesnější informace se můžete dočíst v [8].

Lidé bývají často vymrštěni předním oknem ven z automobilu, což vede k smrtelným nehodám. Důvod je zřejmý. Lidé se odmítají, či zapomínají připoutat, což je porušení předpisů a jsou za to udělovány pokuty. Pásy nejsou jediným zádržným systémem. Pro děti, jejichž hmotnost nepřevyšuje 36 kg a tělesná výška nepřevyšuje 150 cm, jsou určeny autosedačky. Umístění dětské autosedačky je podrobně popsáno v zákoně o silničním provozu 361/2000 Sb., §6.

Pokud řidiče zastaví z nějakého důvodu policista, ať už se jedná o běžnou kontrolu, či řidič učinil přestupek, musí mu na vyzvání předložit řidičský průkaz, osvědčení o registraci vozidla, doklad prokazující pojištění odpovědnosti za škodu způsobenou provozem vozidla a doklad o zdravotní způsobilosti, pokud je potřeba.

### 1.3 Dopravní nehoda, §47

Pro vymezení pojmu dopravní nehoda je použit zákon o silničním provozu č. 361/2000 Sb., §47:

*(1) Dopravní nehoda je událost v provozu na pozemních komunikacích, například havárie nebo srážka, která se stala nebo byla započata na pozemní komunikaci a při níž dojde k usmrcení nebo zranění osoby nebo ke škodě na majetku v přímé souvislosti s provozem vozidla v pohybu.*

*(2) Řidič, který měl účast na dopravní nehodě, je povinen*

- a) neprodleně zastavit vozidlo,*
- b) zdržet se požití alkoholického nápoje a jiných návykových látek po nehodě po dobu, do kdy by to bylo na újmu zjištění, zda před jízdou nebo během jízdy požil alkoholický nápoj nebo návykovou látku, vždy však do doby příjezdu policisty v případě, že jsou účastníci nehody povinni ohlásit nehodu policistovi podle odstavců (4) a (5),*
- c) učinit opatření k zabránění vzniku škody osobám nebo věcem, pokud tak hrozí v důsledku dopravní nehody, a*
- d) spolupracovat při zjišťování skutkového stavu.*

*(3) Účastníci dopravní nehody jsou povinni*

- a) učinit vhodná opatření, aby nebyla ohrožena bezpečnost provozu na pozemních komunikacích v místě dopravní nehody; vyžadují-li to okolnosti, jsou oprávněni zastavovat jiná vozidla,*
- b) oznámit, v případech stanovených tímto zákonem, nehodu policii; došlo-li k zranění, poskytnout podle svých schopností první pomoc a k zraněné osobě přivolat zdravotnickou záchrannou službu,*
- c) označit místo dopravní nehody,*
- d) umožnit obnovení provozu na pozemních komunikacích, zejména provozu vozidel hromadné dopravy osob,*
- e) neprodleně ohlásit policii poškození pozemní komunikace, obecně prospěšného zařízení nebo životního prostředí, pokud k němu při dopravní nehodě došlo,*
- f) prokázat si na požádání navzájem svou totožnost a sdělit údaje o vozidle, které mělo účast na dopravní nehodě.*

*g) v případech, kdy nevznikne povinnost oznámit nehodu policii, sepsat společný záznam o dopravní nehodě, který podepíše a neprodleně předají pojistiteli; tento záznam musí obsahovat identifikaci místa a času dopravní nehody, jejích účastníků a vozidel, její příčiny, průběhu a následků.*

*(4) Dojde-li při dopravní nehodě k usmrcení nebo zranění osoby nebo k hmotné škodě převyšující zřejmě na některém ze zúčastněných vozidel včetně přepravovaných věcí nebo na jiných věcech částku 100 000 Kč, jsou účastníci dopravní nehody povinni*

- a) neprodleně ohlásit dopravní nehodu policistovi,*
- b) zdržet se jednání, které by bylo na újmu řádného vyšetření dopravní nehody, zejména přemístění vozidel; musí-li se však situace vzniklá dopravní nehodou změnit, zejména je-li to nutné k vyproštění nebo ošetření zraněné osoby nebo k obnovení provozu na pozemních komunikacích, především provozu vozidel hromadné dopravy osob, vyznačit situaci a stopy,*
- c) setrvat na místě dopravní nehody až do příchodu policisty nebo se na toto místo neprodleně vrátit po poskytnutí nebo přivolání pomoci nebo ohlášení dopravní nehody.*

*(5) Povinnost podle odstavce 4 platí i v případě, kdy při dopravní nehodě*

- a) dojde ke hmotné škodě na majetku třetí osoby, s výjimkou škody na vozidle, jehož řidič má účast na dopravní nehodě nebo škody na věci přepravované v tomto vozidle,*
- b) dojde k poškození nebo zničení součásti nebo příslušenství pozemní komunikace podle zákona o pozemních komunikacích (20a), nebo*
- c) účastníci dopravní nehody nemohou sami bez vynaložení nepřiměřeného úsilí zabezpečit obnovení plynulosti provozu na pozemních komunikacích.[8]*

Výše hmotné škody je jednou z podmínek pro nahlášení dopravní nehody policii ČR, ta se ale v průběhu let mění. Před 1. červencem 2006 byla částka pro nahlášení dopravní nehody ve výši 20 000 Kč, ale po tomto datu byla zvýšena na 50 000 Kč.

V nynější době částka vzrostla na 100 000 Kč, což bylo ustanoveno ke dni 1. ledna 2009. To se razantně projevilo v získaných datech. Počet nahlášených nehod v roce 2009 poklesl proti roku 2008 zhruba o ½, což bude názorně ukázáno v kapitole 2.



## 2 NEHODOVOST V ČR

Počet nehod je jedním z ukazatelů, pomocí kterého můžeme určit, jak je země bezpečná. V mé bakalářské práci se můžete dočíst, že Česká republika byla v 80. letech jednou z nejbezpečnějších, což je zobrazeno i v grafu.

Dopravní nehoda se nikdy neobejde bez následků. V nejlepším případě vznikne pouze hmotná škoda, bohužel tuto situaci často doplňuje zranění osob až jejich smrt. Od roku 1980 je zavedena doba, po kterou se sledují následky zranění, což napomáhá k objektivnějšímu posouzení závažnosti dopravních nehod. Doba sledování je 30 dnů, která je ještě rozdělena na dobu do 24 hodin po nehodě a od 24 hodin do 30 dnů po nehodě. Následující tabulka uvádí počet nehod, počet usmrcených osob a počet zraněných osob v období 1993 – 2011 [3].

TABULKA 2.1 Počet nehod a následky nehod v období 1993 – 2011

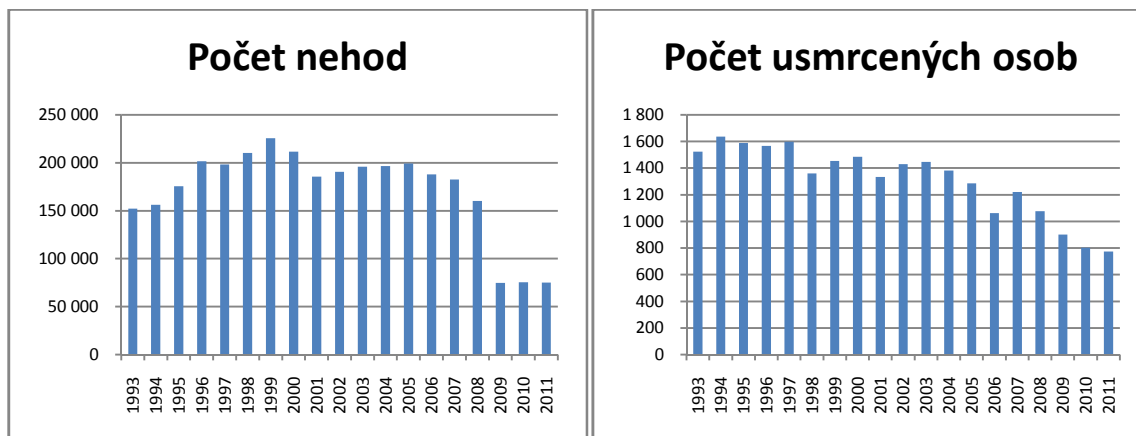
Rok	Počet nehod	Počet usmrcených osob			těžce zraněno	lehce zraněno	počet nehod/počet usmrcených osob celkem
		do 24 hodin po nehodě	od 24 hodin do 30 dnů po nehodě	Celkem			
1993	152 157	1 355	169	1 524	5 629	26 821	100
1994	156 242	1 473	164	1 637	6 232	29 590	96
1995	175 520	1 384	204	1 588	6 298	30 866	111
1996	201 697	1 386	182	1 568	6 621	31 296	129
1997	198 431	1 411	186	1 597	6 632	30 155	125
1998	210 138	1 204	156	1 360	6 152	29 225	155
1999	225 690	1 322	133	1 455	6 093	28 747	156
2000	211 516	1 336	150	1 486	5 525	27 063	143
2001	185 664	1 219	115	1 334	5 493	28 297	140
2002	190 718	1 314	117	1 431	5 492	29 013	134
2003	195 851	1 319	128	1 447	5 253	30 312	136
2004	196 470	1 215	167	1 382	4 878	29 543	143
2005	199 262	1 127	159	1 286	4 396	27 974	155
2006	187 965	956	107	1 063	3 990	24 231	177
2007	182 736	1 123	99	1 222	3 960	25 382	150
2008	160 376	992	84	1 076	3 809	24 776	149
2009	74 815	832	69	901	3 536	23 777	83
2010	75 522	753	49	802	2 823	21 610	95
2011	75 137	707	66	773	3 092	22 519	98

Následující grafy zobrazují počet nehod a počet usmrcených osob od roku 1993 do roku 2011. Levý graf se týká zmíněných počtů nehod, které charakterizuje sinusoida. Počátek, rok 1993, popisuje číslo 152 157, které se v dalších letech stále zvyšuje

až do roku 1999. Toto datum se stalo zlomovým, jelikož za celé období zde vzniklo nejvíce nehod (225 690). Následující dva roky byly příznivější a naopak do roku 2005 opět počet nehod začal stoupat. Zbytek sledovaného období se znovu obrátil k lepšímu, kdy počet nehod začal rok od roku klesat. Za zmínku stojí rok 2009. Jak již bylo řečeno, v roce 2009 byl změněn zákon, což se projevilo v počtu hlášených nehod. V grafu můžeme vypořadovat razantní pokles nehod o více jak polovinu oproti roku 2008.

Pravý graf se týká počtu usmrcených osob. Je vidět, že počet mrtvých osob při nehodě s malými výkyvy stále klesá. Zajímavostí je, že ačkoliv v roce 2009 počet nehod klesl téměř o polovinu, tak počet usmrcených osob nebyl nijak výraznější v daném čase. Velice pozitivní zprávou je fakt, že počet zemřelých osob byl v roce 2011 o polovinu nižší než počet mrtvých na začátku sledovaného období.

GRAF 2.1 Počet nehod a počet usmrcených osob v období 1993 - 2011



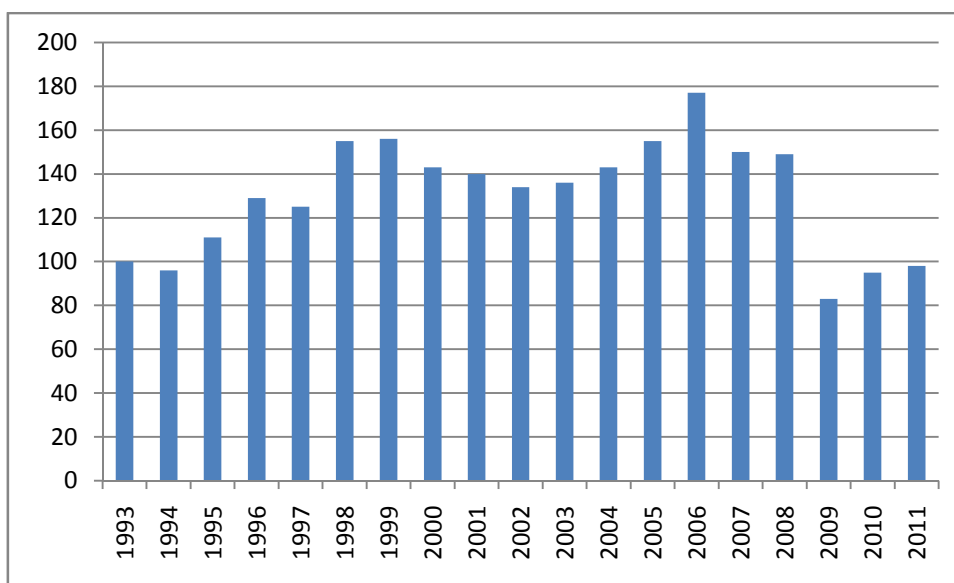
Z uvedených grafů je vidět, že počet nehod i počet usmrcených osob je v posledních letech nejnižší za celé období, což je velice dobrá zpráva. Pokud ale dáme do poměru počet nehod a počet usmrcených osob, získáme zajímavý výsledek, který nám zobrazuje poslední sloupec výchozí tabulky a názorně graf 2.2. Období 1995 – 2008 takto přepočtených dat bylo bezpečnější, jelikož každá více jak 100. nehoda se stala smrtelnou. V roce 2006 to byla dokonce každá 177. dopravní nehoda. Naopak v letech 1993, 1994, 2009, 2010 a 2011 se stala každá méně jak 100. nehoda tragickou. Rok 2009 byl z nich nejhorší, jelikož každá 83. nehoda se stala fatální. Poslední roky lze vysvětlit opět novou vyhláškou o nahlášení dopravních nehod policii ČR. Každá nehoda, při které došlo k jakémukoliv zranění, musí a musela být ohlášena. Pokud by nebyla změněna výše hmotné škody, nehod by vzniklo sice více, ale ukazatel daného poměru by byl pro nás příznivější.

**Příklad 2.1** Pokud by počet nehod pokračoval v původním trendu, předpokládejme, že by se stalo 150 000 dopravních nehod. Počet zemřelých osob by se z hlediska zákona nezměnil, tedy 802 mrtvých lidí.

*ŘEŠENÍ:* počet nehod/počet usmrcených osob = 150 000/802 = 188

Z výsledku je patrné, že pokud by nebyl změněn zákon a počet nehod by odpovídal našemu modelu, pak by se stala každá 188. nehoda smrtelnou. To je ale pouze hypotéza. Následující graf zobrazuje data, která odpovídají reálné situaci.

GRAF 2.2 Poměr počtu nehod a počtu usmrcených osob v období 1993 - 2011



Každá nehoda má svůj původ v něčem jiném. Jedna je zaviněna chybou řidiče jako například nepřiměřenou rychlostí, nesprávným způsobem jízdy, jiná je způsobena nepříznivým vlivem počasí nebo stavem vozovky. Nejvíce nehod je ovšem důsledkem zmíněné chyby řidiče, kdy tyto nehody jsou ve velké míře smrtelnými. Následující tabulka uvádí nejčastější příčiny fatálních nehod způsobených řidičem motorového vozidla v letech 2005 - 2011. V tabulce lze vidět, že nejvíce nehod způsobuje nepřiměřená rychlost jízdy, nesprávné předjíždění, nedání přednosti v jízdě a nesprávný způsob jízdy. Můžeme zde užít Paretova pravidla, kdy dané příčiny způsobují 80% smrtelných nehod. I v těchto případech je patrný pokles jak nehod tak i snížení počtu usmrcených osob v posledních letech.

TABULKA 2.2 Hlavní příčiny nehod řidičů motorových vozidel [3]

		Nepřiměřená rychlost jízdy	Nesprávné předjíždění	Nedání přednosti v jízdě	Nesprávný způsob jízdy	Celkem
2005	Počet nehod	31 066	4 274	33 152	115 975	184 467
	Počet usmrcených	481	71	142	321	1 015
2006	Počet nehod	25 892	3 732	31 376	113 152	174 152
	Počet usmrcených	420	35	107	293	855
2007	Počet nehod	25 019	3 421	32 179	107 014	167 633
	Počet usmrcených	492	67	121	312	992
2008	Počet nehod	23 187	2 975	28 625	92 551	147 338
	Počet usmrcených	432	69	137	275	913
2009	Počet nehod	15 348	1 654	12 241	37 977	67 220
	Počet usmrcených	368	31	104	252	755
2010	Počet nehod	14 633	1 543	12 060	39 219	67 455
	Počet usmrcených	279	37	114	245	675
2011	Počet nehod	13 426	1 458	11 539	39 666	66 089
	Počet usmrcených	284	29	107	232	652

## 2.1 ČR a země EU

Automobil používá v dnešní době každá rodina, což se netýká jen České republiky, ale samozřejmě celého světa. Podívejme se nyní, jaká je úmrtnost na silnicích v zemích Evropské unie. V následující tabulce je uveden počet usmrcených osob v letech 2001 – 2010, změna počtu usmrcených osob v roce 2001 oproti roku 2010 v procentech a nakonec průměrná roční změna v těchto letech v procentech. Země jsou seřazeny dle předposledního sloupce, tedy země s největší změnou počtu zemřelých osob v roce 2001 proti roku 2010 jsou na prvních místech. I v zemích EU se daný počet neustále snižuje. Největší pokles zaznamenalo Estonsko o 0,61% a nejmenší Rumunsko o 0,03%. Česká republika se umístila na 18. místě s poklesem usmrcených osob o 0,4%. Celkově počet mrtvých osob v zemích EU klesl o 11,38%, konkrétně v číslech je to z 54302 osob na 30937 osob.

TABULKA 2.1.1 Počet usmrcených osob v zemích EU v letech 2001 – 2010 [4]

Pořadí	Země	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	změna mezi 2001 a 2010 v %	průměrná roční změna v % (2001 a 2010)
1.	Estonsko	199	223	164	170	170	204	196	132	98	78	-0,61	-0,1
2.	Lotyšsko	558	559	532	516	442	407	419	316	254	218	-0,61	-0,1
3.	Litva	706	697	709	752	773	760	740	499	370	299	-0,58	-0,09
4.	Španělsko	5517	5347	5400	4749	4442	4104	3823	3100	2714	2479	-0,55	-0,09
5.	Lucembursko	70	62	53	50	47	43	46	35	48	32	-0,54	-0,08
6.	Švédsko	583	560	529	480	440	445	471	397	358	266	-0,54	-0,08
7.	Francie	8162	7655	6058	5530	5318	4709	4620	4275	4273	3992	-0,51	-0,08
8.	Slovinsko	278	269	242	274	258	262	293	214	171	138	-0,5	-0,07
9.	Irsko	412	376	337	377	400	365	338	280	239	212	-0,49	-0,07
10.	Německo	6977	6842	6613	5842	5361	5091	4949	4477	4152	3648	-0,48	-0,07
11.	Velká Británie	3598	3581	3658	3368	3336	3298	3059	2645	2337	1905	-0,47	-0,07
12.	Nizozemí	993	987	1028	804	750	730	709	677	644	537	-0,46	-0,07
13.	Belgie	1486	1306	1214	1162	1089	1069	1071	944	944	812	-0,45	-0,06
14.	Portugalsko	1670	1655	1542	1294	1247	969	974	885	840	845	-0,44	-0,06
15.	Itálie	7096	6980	6563	6122	5818	5669	5131	4731	4237	4090	-0,42	-0,06
16.	Rakousko	958	956	931	878	768	730	691	679	633	552	-0,42	-0,06
17.	Dánsko	431	463	432	369	331	306	406	406	303	255	-0,41	-0,06
18.	ČR	1333	1430	1447	1382	1286	1063	1221	1076	901	802	-0,4	-0,05
19.	Maďarsko	1239	1429	1326	1296	1278	1303	1232	996	822	740	-0,4	-0,06
20.	Slovensko	614	610	645	603	606	614	667	622	380	371	-0,4	-0,05
21.	Finsko	433	415	379	375	379	336	380	344	279	272	-0,37	-0,05
22.	Kypr	98	94	97	117	102	86	89	82	71	60	-0,39	-0,05
23.	Řecko	1880	1634	1605	1670	1658	1657	1612	1555	1456	1258	-0,33	-0,04
24.	Polsko	5534	5826	5642	5712	5444	5243	5583	5437	4572	3908	-0,29	-0,04
25.	Bulharsko	1011	959	960	943	957	1043	1006	1061	901	776	-0,23	-0,03
26.	Malta	16	16	16	13	17	11	14	15	21	15	-0,06	-0,01
27.	Rumunsko	2450	2411	2229	2442	2629	2587	2800	3061	2796	2377	-0,03	0
	EU	54302	53342	50351	47290	45346	43104	42540	38941	34814	30937	-11,38	-1,65

Evropská Unie se skládá z 27 členských států. Každý stát má svá pravidla pro řízení motorového vozidla. Liší se v mnoha ohledech. Jedním z nich je věk, při němž může člověk řídit motorové vozidlo, nebo věk tzv. dozorce, při jehož dohledu může řídit vozidlo osoba, která ještě nevlastní řidičské oprávnění. V některých zemích je také povolena určitá hladina alkoholu v krvi při řízení, což v ČR zatím zavedeno není.

## 2.2 Ministerstvo dopravy

Ministerstvo dopravy má na starosti veškerou problematiku týkající se dopravy jak silniční, tak drážní, letecké i vodní. V České republice se hlavně jedná o první tři způsoby

dopravy, pro diplomovou práci pouze o silniční dopravu. Ministerstvo se stará o bezpečnost na silnicích, legislativu, výstavbu nových silnic, dálnic, mostů, atd. V čele Silniční dopravy stojí ministr, kterým je v této době Pavel Dobeš, a generální ředitel David Čermák. Následující tabulka udává přehled ministrů dopravy, jejich období působení, jejich stranu a vládu, ve které byli zvoleni. Tabulka je zobrazuje ministry vládnoucí od roku 1992 až do dnešní doby.

TABULKA 2.2.1 Ministři dopravy [5], [6], [7]

MINISTR	OBDOBÍ	STRANA	VLÁDA
PhDr. Jan Stráský	2. 7. 1992 - 10. 10. 1995	ODS	první vláda Václava Klause
Ing. Vladimír Budínský	10. 10. 1995 - 4. 7. 1996	ODS	první vláda Václava Klause
Ing. Martin Říman	4. 7. 1996 - 2. 1. 1998	ODS	druhá vláda Václava Klause
Prof. Ing. Petr Moos	2. 1. 1998 - 17. 7. 1998	nestráník	vláda Josefa Tošovského
Doc. Ing. Antonín Peltrám	22. 7. 1998 - 26. 4. 2000	ČSSD	vláda Miloše Zemana
Ing. Jaromír Schling	26. 4. 2000 - 12. 7. 2002	ČSSD	vláda Miloše Zemana
Ing. Milan Šimonovský	15. 7. 2002 - 4. 8. 2004	KDU-ČSL	vláda Vladimíra Špidly
Ing. Milan Šimonovský	4. 8. 2004 - 25. 4. 2005	KDU-ČSL	vláda Stanislava Grosse
Ing. Milan Šimonovský	25. 4. 2005 - 16. 8. 2006	KDU-ČSL	vláda Jiřího Paroubka
Ing. Aleš Řebíček	4. 9. 2006 - 9. 1. 2007	ODS	první vláda Mirka Topolánka
Ing. Aleš Řebíček	9. 1. 2007 - 23. 1. 2009	ODS	druhá vláda Mirka Topolánka
Ing. Petr Bendl	23. 1. 2009 - 26. 3. 2009 - demise <sup>1</sup> 26. 3. 2009 - 8. 5. 2009 - ministr v demisi	ODS	druhá vláda Mirka Topolánka
Ing. Gustáv Slamečka	8. 5. 2009 - 25. 6. 2010 - demise 25. 6. 2010 - 13. 7. 2010 - ministr v demisi	nestráník	vláda Jana Fischera
JUDr. Vít Bárta	13. července 2010 – 21. dubna 2011	Věci veřejné	Vláda Petra Nečase
JUDr. Radek Šmerda, Ph.D.	21. dubna 2011 - 1. července 2011	nestráník	Vláda Petra Nečase
Mgr. Pavel Dobeš	ode dne 1. 7. 2011	Věci veřejné	Vláda Petra Nečase

<sup>1</sup> Demise - znamená vzdání se úřadu, podání demise je dokončeno odstoupením z úřadu

### 3 MĚSTO ZLÍN

Celá kapitola je čerpána z [10]. Obrázek 3.1 představuje znak a vlajku města Zlín.

OBRÁZEK 3.1 Znak a vlajka obce Zlín



#### 3.1 Historie a současnost

Zlín se nachází v jihovýchodní části Moravy a patří k jednomu z nejkrásnějších měst. To dokazuje fakt, že město bylo v roce 2009 označeno za Evropské kulturní dědictví. Městem protéká řeka Dřevnice, která je levostranným přítokem řeky Moravy. Zlín patří k menším městům, které obývá zhruba 85 000 lidí.

Název Zlín je vysvětlován mnoha způsoby. Jedním z nich je ten, že byl odvozen od zlých událostí, které kdysi město postihly. První spolehlivá zmínka o Zlíně pochází z roku 1322, kdy vládla královna Eliška Rejčka. V roce 1894 založili sourozenci (Anna, Antonín a Tomáš) Baťovi obuvnický podnik, který zajistil rozvoj města. Firma budovala ve městě obchodní domy, hotel, kino, velkou nemocnici, školní budovy, vědecké budovy, filmové ateliéry, tisíce nových bytů. Na výstavbě se podíleli významní architekti jako J. Kotěra, F. L. Gahura, V. Karfík, M. Lorenc. Zlín získal evropsky ojedinělý charakter funkcionalistického města. Někteří lidé znají město Zlín pod dřívějším názvem Gottwaldov, který lidé používali od 1. ledna 1949 do 31. prosince 1989. V tomto období vládli komunisté, ale po pádu komunistického režimu nastoupili demokraté, kteří vládnou dodnes.

Zlín se stal 14. 11. 2000, po podepsání zákona o zřízení Univerzity Tomáše Bati tehdejším prezidentem Václavem Havlem, univerzitním městem. Projekt byl připravován již v 90. letech, kdy byla zřízena Fakulta managementu a ekonomiky a Institut reklamní tvorby a marketingových komunikací. V současné době se univerzita skládá z šesti fakult: Fakulta technologická, Fakulta managementu a ekonomiky, Fakulta multimediálních

komunikací, Fakulta aplikované informatiky, Fakulta humanitních studií a Fakulta logistiky a krizového řízení.

V roce 2006 byl položen základní kámen Kongresového a univerzitního centra. Dohromady tedy tvoří multifunkční komplex, který byl dokončen v roce 2008. Projekt je výtvorem architektky Evy Jiříčné spolu s jejím týmem spolupracovníků z firmy AI Design.

### 3.2 Doprava a silnice

Prvním hromadným prostředkem ve Zlíně byl vlak. Železniční trať vznikla roku 1899. Musela být rozšířena o další prostředky, které vedly k novým obytným částem města mimo železniční trať. V roce 1928 zahájily činnost autobusy, později i trolejbusy. Letecká doprava ve Zlíně neexistuje, ale nejbližší civilní letiště se nachází v Otrokovicích, což je vzdáleno 10 km od Zlína. Ve večerních hodinách, když je hromadná doprava omezená, využívají lidé hodně taxislužby. Mají dvě základní stanoviště, na parkovišti před Flipem Zlín - náměstí Práce a na ulici Kvítková.

Do Zlína se člověk dostane jak po silnici první třídy, tak i po silnici druhé a třetí třídy. Na následujícím obrázku vidíte přesné přístupové cesty. V práci se budeme zabývat čtyřmi silnicemi I/49, II/490, II/497 a III/49016.

OBRÁZEK 3.2.1 Silnice města Zlína



Každá silnice někde začíná a někde končí, což udává směr staničení (kilometrůž). Používá se při stanovení místa opravy silnice, polohy dopravní nehod, atd. Pro silnice použité v práci platí následující směr staničení.



TABULKA 3.2.1 Směr staničení silnic

I/49	Otrokovice – Bratřejov
II/490	Fryšták – Polichno
II/497	Zlín – Šarovy
III/49016	Zlín – Racková

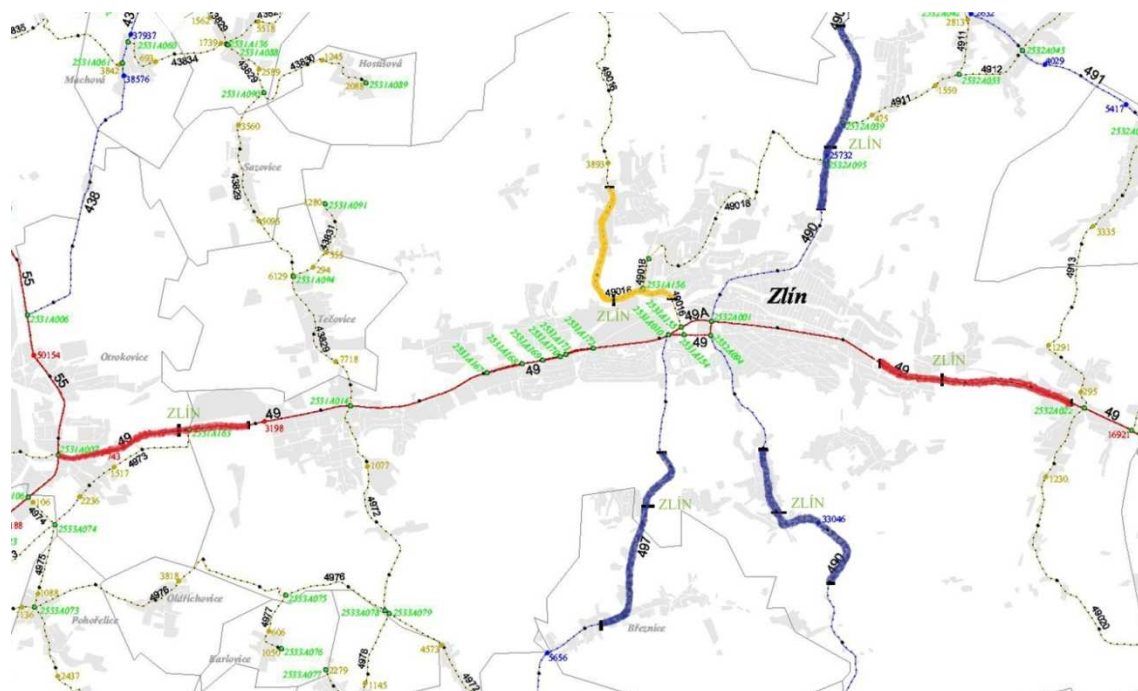
Kilometráž využijeme i v této práci. Jelikož zpracování dat by bylo velice zdouhavé, budeme testovat jen dopravní nehody na určitých úsecích zmíněných silnic. Budou to zhruba dva kilometry před obcí Zlín a jeden kilometr směrem do města. Silnice I/49 prochází celým Zlínem, proto budu brát dvě části z této silnice. Totéž se týká i silnice I/490, což je zřejmé z obrázku výše. V následující tabulce jsou uvedeny přibližné úseky, které se týkají práce.

TABULKA 3.2.2 Přibližné úseky silnic použité v práci

číslo silnice	třída silnice	km před obcí Zlín	začátek obce Zlín	km v obci Zlín
49	I.	0	1,9	2,9
49	I.	15,64	13,64	12,64
490	II.	23,73	25,73	26,73
490	II.	35,05	33,05	32,05
497	II.	4,74	2,74	1,74
49016	III.	3,44	1,44	0,44

Zvýrazněné části silnic na následujícím obrázku jsou přibližnou vizualizací dané tabulky.

OBRÁZEK 3.2.2 Přibližné úseky silnic použité v práci



## 4 CÍLE

Cílem práce je vytvořit modely závislosti počtu dopravních nehod na různých faktorech a faktorech danou nehodu charakterizujících. Těmito faktory jsou například dny v týdnu, měsíce, čas, místo nehody, druh nehody, druh vozidla, zavinění nehody, alkohol, hlavní příčina, stav vozovky, viditelnost a povětrnostní podmínky. Bude použit zobecněný lineární model, který je vhodný pro uvedený typ analýzy. Tento model umožní mimo jiné najít vztahy mezi jednotlivými faktory ovlivňujícími počet dopravních nehod a posouzení jejich významnosti.

Informace o dopravních nehodách jsem získala z *inspektorátu Policie České republiky Zlín* ve formě, která je uvedena v příloze 2 a to za období 2004 – 2011. Výpočet bude prováděn pomocí programu R.

## 5 ZÁKLADNÍ STATISTIKA

### 5.1 Základní pojmy

**Definice 5.1.1.** *Nechť je dán pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$ . Zobrazení  $X: \Omega \rightarrow \mathbb{R}^1$  nazveme náhodnou veličinou, jestliže pro každé  $x \in \mathbb{R}^1$  platí*

$$\{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{A}.$$

*Množinu  $M \subset \mathbb{R}^1$  všech hodnot náhodné veličiny  $X$  nazýváme obor hodnot náhodné veličiny  $X$ .*

**Definice 5.1.2.** *Nechť  $X$  je náhodná veličina definovaná na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ . Reálná funkce  $F_X$  definovaná na  $\mathbb{R}^1$  předpisem*

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}^1,$$

*se nazývá distribuční funkce náhodné veličiny  $X$ .*

**Definice 5.1.3.** *Rozdělení pravděpodobností náhodné veličiny  $X$  je množinová funkce  $P_X(B): \mathcal{B}_1 \rightarrow \mathbb{R}^1$  definovaná vztahem*

$$P_X(B) = P(X \in B), \quad B \in \mathcal{B}_1.$$

**Definice 5.1.4.** *Předpokládejme, že existuje konečná nebo nekonečná prostá posloupnost reálných čísel  $\{x_n\}$  taková, že*

$$\sum_n P(X = x_n) = 1.$$

*Označme  $p_n = P(X = x_n)$ . Posloupnost  $\{x_n\}$  hodnot, kterých nabývá náhodná veličina  $X$  a posloupnost  $\{p_n\}$  pravděpodobností, s nimiž náhodná veličina svých hodnot nabývá, určují tzv. diskrétní rozdělení pravděpodobností náhodné veličiny  $X$ . Náhodná veličina, která má diskrétní rozdělení pravděpodobností, se nazývá diskrétní resp. diskrétního typu. Distribuční funkce diskrétní náhodné veličiny je dána vztahem*

$$F_X(x) = \sum_{n: x_n \leq x} p_n, \quad \forall x \in \mathbb{R}^1.$$

*Tato funkce se nazývá diskrétní distribuční funkce.*

**Příklad 5.1.1.** Binomické rozdělení, Poissonovo rozdělení, alternativní rozdělení, hypergeometrické rozdělení, geometrické rozdělení

**Definice 5.1.5.** Řekneme, že náhodná veličina  $X$  má spojité rozdělení pravděpodobností (má rozdělení spojitěho typu, je spojitá), existuje-li nezáporná, borelovsky měřitelná funkce

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad \forall x \in \mathbb{R}^1.$$

Funkce  $f_X$  se nazývá hustota (rozdělení pravděpodobností) náhodné veličiny  $X$ .

**Příklad 5.1.2.** Rovnoměrné rozdělení, normální rozdělení, normální normované rozdělení, exponenciální rozdělení

**Definice 5.1.6.** Necht' je dán pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$  a skalární náhodná veličina  $X$ . Je-li náhodná veličina  $X$  diskrétní a má pravděpodobnostní funkci  $\pi(x)$ , nazýváme její střední hodnotou (vzhledem k pravděpodobnosti  $P$ ) číslo

$$E(X) = \sum_{x=-\infty}^{\infty} x \pi(x)$$

za předpokladu, že případná nekonečná řada vpravo absolutně konverguje.

Je-li náhodná veličina spojitá a má hustotu pravděpodobnosti  $\varphi(x)$ , nazýváme její střední hodnotou číslo

$$E(X) = \int_{-\infty}^{\infty} x \varphi(x) dx$$

za předpokladu, že nevlastní Riemannův integrál vpravo absolutně konverguje.

Číslo

$$\text{var}(X) = E([X - E(X)]^2) = E(X^2) - [E(X)]^2$$

nazýváme rozptylem náhodné veličiny  $X$  (vzhledem k pravděpodobnosti  $P$ ) za předpokladu, že obě střední hodnoty vpravo existují.

Číslo

$$\sqrt{\text{var}(X)}$$

nazýváme směrodatnou odchylkou náhodné veličiny  $X$ .

Uvedené pojmy jsou čerpány z literatury [12], [13], [14] Můžete zde najít i další pojmy a vlastnosti s nimi spojené.

## 5.2 Diskrétní rozdělení

Definice diskrétního rozdělení byla uvedena v kapitole 5.1 základní pojmy.

### 5.2.1 Poissonovo rozdělení

Poissonovo rozdělení patří mezi diskrétní rozdělení a za určitých předpokladů je aproximací binomického rozdělení. Nejprve tedy uvedeme tvar a vlastnosti binomického rozdělení, ze kterého bude následně odvozeno Poissonovo rozdělení potřebné k dalším výpočtům. Informace jsou získány z literatury [14].

Binomické rozdělení má náhodná veličina  $X$ , která nabývá hodnot  $k = 0, 1, \dots, n$  s pravděpodobnostmi

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

kde  $n \in N, p \in (0,1)$  jsou parametry tohoto rozdělení. Parametr  $n$  označuje rozsah náhodné veličiny  $X$  a parametr  $p$  je pravděpodobnost výskytu nějakého jistého náhodného jevu. Binomické rozdělení se označuje  $X \sim \text{Bi}(n, p)$ .

Pro binomické rozdělení platí vztahy

a) distribuční funkce

$$F(x) = \begin{cases} 0, & \text{pro } x < 0, \\ \sum_{k \leq x} \binom{n}{k} p^k (1-p)^{n-k}, & \text{pro } 0 \leq x < n, \\ 1, & \text{pro } x \geq n, \end{cases}$$

b) střední hodnota

$$E(X) = np,$$

c) rozptyl

$$\text{var}(X) = E(X^2) - [E(X)]^2 = np(1-p).$$

Odvození uvedených vztahů lze nalézt například v

Binomické rozdělení bude nyní použito pro odvození Poissonova rozdělení, které je jeho limitním případem pro  $n \rightarrow \infty, p_n \rightarrow 0, np_n = \lambda$ . Necht'  $k = 0, 1, \dots$  je libovolné pevné. Položíme  $p_n = \frac{\lambda}{n}$ . Pak

$$\begin{aligned}
\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} &= \lim_{n \rightarrow \infty} \frac{n!}{k! (n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\
&= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2) \dots (n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\
&= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left[ \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-k} \right] \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = \\
&= \frac{\lambda^k}{k!} e^{-\lambda}
\end{aligned}$$

Poissonovo rozdělení má náhodná veličina  $X$ , která nabývá hodnot  $k = 0, 1, \dots$  s pravděpodobnostmi

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0,$$

kde  $\lambda$  je parametr. Označuje se  $X \sim \text{Po}(\lambda)$ .

Pro distribuční funkci platí vztah

$$F(x) = \begin{cases} 0, & \text{pro } x < 0, \\ \sum_{k \leq x} \frac{\lambda^k}{k!} e^{-\lambda}, & \text{pro } x \geq 0. \end{cases}$$

Nyní odvodím vztah pro střední hodnotu Poissonova rozdělení.

$$E(X) = \sum_{k=0}^{\infty} k P(X = k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda.$$

Pro odvození rozptylu použijeme následující vztah:

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^2.$$

Nyní přejdeme k výpočtu samotného rozptylu.

$$\text{var}(X) = E(X^2) - [E(X)]^2 = E[X(X-1)] + E(X) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Vidíme tedy, že střední hodnota Poissonova rozdělení a rozptyl Poissonova rozdělení je shodný a to rovný  $\lambda$ .

Poissonovo rozdělení použijeme, jelikož pro počty dopravních nehod není žádný pevný horní limit  $n$  pro  $y$ .

## 5.2.2 Multinomické rozdělení

Uvažujme  $n$  nezávislých, identických pokusů, z nichž každý může dopadnout jedním z  $c$  ( $c > 2$ ) způsobů, kde  $c$  značí celkový počet tříd (kategorií).

Dále uvažujme  $i$ -tý pokus  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ic})$ . Necht' je pevně dáno  $\sum_j Y_{ij} = 1$ , kde

$$Y_{ij} = \begin{cases} 1 & \text{pokud výsledek } i - \text{ tého pokusu je z } j - \text{ té třídy,} \\ 0 & \text{jinak,} \end{cases}$$

pak stačí znát pouze  $(c - 1)$  členů vektoru  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ic})$  a zbytek dopočítat.

Předpokládejme, že výsledek každého pokusu můžeme zařadit do jedné z  $c$  možných tříd, a to s pravděpodobnostmi  $p_1 \in (0,1), \dots, p_c \in (0,1)$ , přičemž platí

$$p_1 + \dots + p_c = 1.$$

Necht'  $X_j = \sum_i Y_{ij}$  označují počet pokusů, které mají výsledek ve třídě  $j$ .

Vektor  $\mathbf{X} = (X_1, X_2, \dots, X_c)$  má multinomické rozdělení  $Mu(n, p_1, \dots, p_c)$  a jeho pravděpodobnostní funkce má tvar

$$P(\mathbf{X} = \mathbf{x}) = P(x_1, \dots, x_c) = \begin{cases} \left( \frac{n!}{x_1! \dots x_c!} \right) p_1^{x_1} \dots p_c^{x_c} & \text{pro } x_j \in \{0, 1, \dots, n\}, \text{ kde } j = 1, \dots, c, x_1 + \dots + x_c = n \\ 0 & \text{jinak} \end{cases}$$

Každé  $X_j$ , kde  $j = 1, \dots, c$  má binomické rozdělení, z čehož vyplývají následující vztahy.

Pro každé  $j \in \{1, \dots, c\}$  platí:

$$E(X_j) = np_j, \\ \text{var}(X_j) = np_j(1 - p_j).$$

Pro každé  $i \in \{1, \dots, c\}, j \in \{1, \dots, c\}, i < j$  platí:

$$\text{cov}(X_i, X_j) = -np_i p_j.$$

Uvedené vztahy a další můžete najít v literatuře [15] a [18].

### 5.2.3 Vztah mezi Poissonovým a multinomickým rozdělením

V literatuře [18] je uveden příklad zaměřený na počet zemřelých osob při nehodách v budoucím týdnu, který ilustruje vztah mezi Poissonovým a multinomickým rozdělením. Necht'  $y_1$  je počet lidí, kteří zemřou při automobilové nehodě,  $y_2$  je počet lidí, kteří zemřou při letecké nehodě a  $y_3$  je počet lidí, kteří zemřou při železniční nehodě. Každá z nezávislých náhodných veličin  $Y_1, Y_2, Y_3$  má Poissonovo rozdělení s parametry  $\mu_1, \mu_2, \mu_3$ . Celkový počet zemřelých je  $n = \sum Y_i$ , který je náhodný a má rovněž Poissonovo rozdělení s parametrem  $\sum \mu_i$ , což lze dokázat následujícím způsobem.

1) Předpokládejme dvě nezávislé náhodné veličiny s Poissonovým rozdělením, tedy  $Y_1 \sim \text{Po}(\mu_1)$  a  $Y_2 \sim \text{Po}(\mu_2)$ . Chceme zjistit, jaké rozdělení bude mít náhodná veličina  $Y_1 + Y_2$ . Pravděpodobnosti náhodných veličin  $Y_1, Y_2$  jsou

$$P(Y_1 = k) = \frac{\mu_1^k}{k!} e^{-\mu_1}, \quad P(Y_2 = k) = \frac{\mu_2^k}{k!} e^{-\mu_2}, \quad k \in \mathbb{N}_0.$$

Pravděpodobnost  $P(Y_1 + Y_2 = k)$  náhodné veličiny  $Y_1 + Y_2$  vypočítáme následujícím způsobem

$$\begin{aligned} &P(Y_1 = 0, Y_2 = k) + P(Y_1 = 1, Y_2 = k - 1) + \dots + P(Y_1 = k, Y_2 = 0) \\ &= \sum_{i=0}^k P(Y_1 = i, Y_2 = k - i) \end{aligned}$$

Díky nezávislosti náhodných veličin  $Y_1, Y_2$  můžeme dále psát

$$\begin{aligned} \sum_{i=0}^k P(Y_1 = i, Y_2 = k - i) &= \sum_{i=0}^k P(Y_1 = i)P(Y_2 = k - i) = \sum_{i=0}^k \frac{\mu_1^i}{i!} e^{-\mu_1} \frac{\mu_2^{k-i}}{(k-i)!} e^{-\mu_2} \\ &= e^{-(\mu_1 + \mu_2)} \frac{1}{k!} \sum_{i=0}^k k! \frac{\mu_1^i}{i!} \frac{\mu_2^{k-i}}{(k-i)!} = \frac{e^{-(\mu_1 + \mu_2)}}{k!} \sum_{i=0}^k \frac{k!}{i!(k-i)!} \mu_1^i \mu_2^{k-i} \\ &= \frac{e^{-(\mu_1 + \mu_2)}}{k!} \sum_{i=0}^k \binom{k}{i} \mu_1^i \mu_2^{k-i} = \frac{e^{-(\mu_1 + \mu_2)}}{k!} (\mu_1 + \mu_2)^k, \end{aligned}$$

což je pravděpodobnost náhodné veličiny  $Y_1 + Y_2$ , která má rozdělení  $Y_1 + Y_2 \sim \text{Po}(\mu_1 + \mu_2)$ .

2) Nyní budeme předpokládat dvě nezávislé náhodné veličiny  $Y_1 + Y_2 = Z$  a  $Y_3$  s Poissonovým rozdělením, tedy  $Z \sim \text{Po}(\mu_1 + \mu_2)$  a  $Y_3 \sim \text{Po}(\mu_3)$ . Chceme určit rozdělení náhodné veličiny  $Y_1 + Y_2 + Y_3$ , což získáme stejným postupem jako v případě 1). Místo náhodné veličiny  $Y_1$  uvažujeme náhodnou veličinu  $Z$  a místo náhodné veličiny  $Y_2$  uvažujeme náhodnou veličinu  $Y_3$ . Získáme

$$\sum_{i=0}^k P(Z = i, Y_3 = k - i) = \frac{e^{-(\mu_1 + \mu_2 + \mu_3)}}{k!} (\mu_1 + \mu_2 + \mu_3)^k,$$

což je pravděpodobnost náhodné veličiny  $Z + Y_3$ , která má Poissonovo rozdělení  $Y_1 + Y_2 + Y_3 \sim \text{Po}(\mu_1 + \mu_2 + \mu_3)$ .

Pokud tedy existují nezávislé náhodné veličiny  $Y_i$  s Poissonovým rozdělením, má i jejich součet  $\sum Y_i$  Poissonovo rozdělení s parametrem  $\sum \mu_i$ . ■



Pokud bychom ovšem zkoumali data z minulého týdne, kdy známe  $n$ , pak už by se vektor  $(Y_1, Y_2, Y_3)$  neskládal z nezávislých veličin a rozdělení by nebylo Poissonovo ale multinomické ( $Mu(n, \{p_i\})$ ).

Pro  $k$  nezávislých proměnných s Poissonovým rozdělením, s  $E(Y_i) = \mu_i$ , odvodíme multinomické rozdělení pomocí podmíněného rozdělení vzhledem k  $\sum Y_i = n$  zapsané

$$P\left((Y_1 = n_1, \dots, Y_k = n_k) \mid \sum Y_j = n\right)$$

Za předpokladu, že  $n_1 + n_2 + \dots + n_k = n$ , dále můžeme psát

$$\frac{P(Y_1 = n_1, \dots, Y_k = n_k)}{P(\sum Y_j = n)} = \frac{\prod_i [\exp(-\mu_i) \mu_i^{n_i} / n_i!]}{\exp(-\sum \mu_j) (\sum \mu_j)^n / n!} = \frac{n!}{\prod_i n_i!} \sum_i p_i^{n_i}$$

kde  $\{p_i = \mu_i / (\sum \mu_j)\}$ .

Multinomické rozdělení charakterizuje většinu kategoriálních dat, které mají obvykle stejné odhady parametrů jako odhady Poissonova rozdělení, díky podobnosti věrohodnostní funkce. Pro jednoduchost se v praktických výpočtech používá Poissonovo rozdělení, ve kterém není dán pevný rozsah hodnot, kterých nabývá náhodná veličina  $Y$ .

### 5.3 Kategoriální proměnná

V praxi se můžeme setkat s různými typy proměnných. Nejjednoduššími z nich jsou kvantitativní proměnné, které můžeme přímo měřit v zavedených, konstantních jednotkách, dostaneme tedy určitou hodnotu dané veličiny. Do této skupiny proměnných řadíme například měsíční příjem rodiny (v Kč), výšku člověka (v cm), tělesnou hmotnost člověka (v kg), čas uběhnutí 100 m (v sekundách), atd.

Existují ale i kvantitativní proměnné, které nelze přímo měřit. Patří mezi ně proměnné jako vzrušivost, spokojenost, atd. S takovými proměnnými si poradíme pomocí diskretních hodnot, což nazýváme diskretizací kvantitativní proměnné. Vytvoříme stupnici hodnot, která danou proměnnou charakterizuje (např. pracovitost vyjadřují čísla od jedné do pěti). Další proměnné, které nelze přímo měřit jsou kvalitativní proměnné, kterým přiřazujeme číselné hodnoty. Např. pokud pracujeme s proměnnou pohlaví, použijeme hodnoty nula pro muže a jedna pro ženy. Velká část proměnných, které se v této práci

vyskytují, má povahu kvalitativní proměnné. Například jsou to dny v týdnu, kterým přiřadíme hodnoty jedna až sedm.

Pro jednoduchost se proměnné, které nelze přímo měřit, ale lze je zařadit do třídy, nazývají kategoriální proměnné. Výše uvedené dělení kategoriálních proměnných není jediné.

Při výpočtech používáme některé proměnné jako vysvětlující, jiné jako vysvětlované. Chování vysvětlované (závislé) proměnné, je objasňováno pomocí vysvětlující (nezávislé) proměnné. V tradičních vícerozměrných metodách mohou kategoriální proměnné plnit pouze určitou úlohu. Například v diskriminační analýze vystupuje kategoriální proměnná jako vysvětlovaná proměnná. Naopak v této práci budeme používat lineární modely, ve kterých bude vystupovat kategoriální proměnná na místě vysvětlující proměnné. Vysvětlovanou proměnnou budou počty dopravních nehod a vysvětlujícími proměnnými budou faktory ovlivňující počet dopravních nehod, kterými jsou stav vozovky, viditelnost, povětrnostní podmínky a alkohol. Faktory, které danou nehodu charakterizují, jako je místo nehody, druh nehody, druh vozidla, zavinění nehody, dny v týdnu, měsíce, čas.

Dále mohou být data buď spojitá, nebo diskrétní v závislosti na charakteru hodnot, kterých sledované veličiny nabývají. Pokud sledované veličiny nabývají pouze konečně, nebo nejvýše spočetně mnoha hodnot, pak mluvíme o diskrétní kategoriální proměnné. Pro spojitou kategoriální proměnnou je naopak charakteristické, že nabývá všech hodnot z nějakého intervalu.

Posledním důležitým dělením je rozlišení nominální, ordinální a intervalové proměnné. Proměnné, jejichž přirozené uspořádání neexistuje, se nazývají nominální, což je např. místo nehody (na křižovatce, mimo křižovatku, uvnitř zóny 1-8 předm.kř., mimo zónu 11-19 a 22-28). Naopak, pokud proměnné lze uspořádat vzestupně či sestupně, mluvíme o proměnných ordinálních. Pokud známe vzdálenosti mezi třídami, jedná se o intervalovou proměnnou. Do této skupiny zařadíme čas (např. (00:0,08:00>, (08:00,16:00>, (16:00,00:00>).

Více informací můžete najít v literatuře [17] a [18].

## 5.4 Kontingenční tabulky

Již na základní škole žáci zkoumají vztahy mezi dvěma znaky a zapisují je pro přehlednost do tabulky. My budeme rovněž srovnávat dvě kategoriální proměnné s  $n$  opakováním experimentu, k čemuž nám pomůže kontingenční tabulka. Informace jsou získány z literatury [12].

Nechť  $X$  a  $Y$  označují dvě vysvětlující kategoriální proměnné, kdy  $X$  obsahuje  $I$  tříd a  $Y$  má  $J$  tříd. Označme

$$p_{ij} = P(X = i, Y = j), \quad p_{i.} = P(X = i) = \sum_{j=1}^J p_{ij}, \quad p_{.j} = P(Y = j) = \sum_{i=1}^I p_{ij}.$$

Platí

$$\sum_{i=1}^I \sum_{j=1}^J p_{ij} = \sum_{i=1}^I p_{i.} = \sum_{j=1}^J p_{.j} = 1.$$

Předpokládejme, že platí  $p_{ij} > 0$  pro všechny dvojice  $(i, j)$ . Dané pravděpodobnosti se pro přehlednost mohou zapsat do tabulky.

TABULKA 5.4.1 Pravděpodobnostní tabulka

X\Y	1	2	...	$J$	$\Sigma$
1	$p_{11}$	$p_{12}$	...	$p_{1J}$	$p_{1.}$
2	$p_{21}$	$p_{22}$	...	$p_{2J}$	$p_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$I$	$p_{I1}$	$p_{I2}$	...	$p_{IJ}$	$p_{I.}$
$\Sigma$	$p_{.1}$	$p_{.2}$	...	$p_{.J}$	1

Dále budeme pracovat s případy, kdy současně nastalo  $X = i$  a  $Y = j$ . Tento počet označíme  $n_{ij}$  a zapíšeme ho do kontingenční tabulky, ve které označíme

$$n_{i.} = \sum_{j=1}^J n_{ij}, \quad n_{.j} = \sum_{i=1}^I n_{ij}, \quad n = \sum_{i=1}^I \sum_{j=1}^J n_{ij}.$$

Platí

$$\sum_{i=1}^I \sum_{j=1}^J n_{ij} = \sum_{i=1}^I n_{i.} = \sum_{j=1}^J n_{.j} = n.$$

TABULKA 5.4.2 Kontingenční tabulka

X\Y	1	2	...	$J$	$\Sigma$
1	$n_{11}$	$n_{12}$	...	$n_{1J}$	$n_{1.}$
2	$n_{21}$	$n_{22}$	...	$n_{2J}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$I$	$n_{I1}$	$n_{I2}$	...	$n_{IJ}$	$n_{I.}$
$\Sigma$	$n_{.1}$	$n_{.2}$	...	$n_{.J}$	$n$

Jednotlivé buňky vyjadřují počet případů, ve kterých měl první znak hodnotu odpovídající příslušnému řádku a druhý znak hodnotu odpovídající příslušnému sloupci. Jsou to již zmíněné počty  $n_{ij}$ , které se nazývají simultánní četnosti, které mají sdružené multinomické rozdělení s parametrem  $n$  a s pravděpodobnostmi  $p_{ij}$ . Jak již bylo řečeno, kontingenční tabulky slouží ke srovnávání znaků.

Testujeme hypotézu, že znaky X a Y jsou nezávislé pomocí statistiky

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_i \cdot n_j}{n}\right)^2}{\frac{n_i \cdot n_j}{n}},$$

kteřá má za platnosti  $H_0$  asymptoticky  $\chi^2$  rozdělení o  $(I-1)(J-1)$  stupních volnosti. Hypotézu o nezávislosti znaků X a Y zamítáme, pokud platí  $\chi^2 \geq \chi_{(I-1)(J-1)}^2(\alpha)$ .

## 5.5 Klasické lineární regresní modely

Velmi často hledáme souvislost mezi zjištěnými hodnotami nějaké činnosti a faktory na tuto činnost působícími, hledáme tedy závislost výstupních veličin  $y_1, \dots, y_N$  na nastavované kombinaci vstupních proměnných  $x_1, \dots, x_k$ .

V našem případě závislou proměnnou  $y_i, i = 1, \dots, N$  představují počty dopravních nehod a nezávisle proměnnými  $x_j, j = 1, \dots, k$  jsou dny v týdnu, měsíce, čas, místo nehody, druh nehody, druh vozidla, zavinění nehody, alkohol, hlavní příčina, stav vozovky, viditelnost a povětrnostní podmínky. Danou situaci zapíšeme ve tvaru  $[y_i, x_{ij}]$ ,  $i = 1, \dots, N, j = 1, \dots, k$  a maticově

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nk} \end{pmatrix}.$$

Naším úkolem je danými hodnotami  $[y_i, x_{ij}]$  proložit lineární kombinaci funkcí  $y = \beta_1 f_1(\mathbf{x}) + \dots + \beta_k f_k(\mathbf{x})$ . Linearitou se zde rozumí závislost  $f(\mathbf{x})$  na koeficientech  $\beta_1, \dots, \beta_k$ . Předpokládáme, že veličiny  $x$  jsou nenáhodné, libovolně nastavitelné a veličina  $y$  je náhodná s určitým pravděpodobnostním rozdělením, konečnou střední hodnotou a konečným rozptylem. Podmíněná střední hodnota náhodné veličiny  $y$  v místě  $\mathbf{x}$  je

$$E(y|\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\beta}).$$

Předpokládáme aditivní tvar modelu ve tvaru

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix},$$

kde vektory  $\mathbf{y}$  a  $\boldsymbol{\varepsilon}$  mají rozměry  $N \times 1$ , vektor  $\boldsymbol{\beta}$  má rozměr  $k \times 1$  a matice  $\mathbf{X}$  je rozměru  $N \times k$ . Uvedenou rovnici můžeme rovněž vyjádřit maticově

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Pokud bychom brali v úvahu model s absolutním členem, měl by tvar

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix},$$

kde vektory  $\mathbf{y}$  a  $\boldsymbol{\varepsilon}$  mají rozměry  $N \times 1$ , vektor  $\boldsymbol{\beta}$  má rozměr  $(k + 1) \times 1$ ,  $(k + 1 = p)$  a matice  $\mathbf{X}$  je rozměru  $N \times p$ .

Klasický lineární regresní model musí splňovat několik podmínek:

1. Regresní parametry  $\boldsymbol{\beta}$  mohou nabývat libovolných hodnot.
2. Regresní model je lineární v parametrech, platí aditivní model měření  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ .
3. Matice nenáhodných, nastavovaných hodnot vysvětlujících proměnných  $\mathbf{X}$  má hodnost rovnou právě  $k$ , což znamená, že mezi  $\mathbf{x}_1, \dots, \mathbf{x}_k$  neexistuje funkční lineární závislost.
4. Náhodné chyby  $\varepsilon_i$  mají nulovou střední hodnotu  $E(\varepsilon_i) = 0$  a konstantní, stejný, konečný rozptyl  $\text{var}(\varepsilon_i) = \sigma^2$ . Neboli vektor náhodné složky  $\boldsymbol{\varepsilon}$  má  $N$ -rozměrné normální rozdělení  $N_N(\mathbf{0}_N, \sigma^2 \mathbf{I}_N)$ . Náhodné chyby  $\varepsilon_i$  jsou vzájemně nekorelované a platí  $\text{cov}(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) = 0$ . Pokud mají chyby navíc normální rozdělení, jsou nezávislé.
5. Pokud mají náhodné chyby  $\varepsilon_i$  normální rozdělení  $N(0, \sigma^2)$ , pak vektor  $\mathbf{y}$  má  $n$ -rozměrné normální rozdělení  $N_N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_N)$  se střední hodnotou  $\mathbf{X}\boldsymbol{\beta}$  a kovarianční maticí  $\sigma^2 \mathbf{I}_N$ .

### 5.5.1 Odhad parametrů a jejich vlastnosti

Parametry modelu lze odhadovat pomocí metody nejmenších čtverců, nebo pomocí maximalizace věrohodnostní funkce. Obě metody si nyní ukážeme.

### 5.5.1.1 Metoda nejmenších čtverců (MNC)

Pomocí metody nejmenších čtverců nalézáme takové řešení, ve kterém je součet druhých mocnin chyb nalezeného řešení minimální. Geometricky tedy minimalizujeme vzdálenost mezi vektorem  $\mathbf{y}$  a lineární kombinací  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ . Matematicky, odhad  $\hat{\boldsymbol{\beta}}$  vypočteme minimalizací kvadratické formy

$$\min(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Minimalizace kvadratické formy znamená, že ji položíme rovnou nule

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{o}_p$$

Levou stranu rovnice roznásobíme

$$\mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y} + \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{o}_p$$

Provedeme první derivaci podle  $\boldsymbol{\beta}$

$$-\mathbf{y}^T\mathbf{X} - \mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{o}_p$$

$$\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y},$$

řešením dané rovnice je odhad  $\hat{\boldsymbol{\beta}}$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

### 5.5.1.2 Metoda maximální věrohodnosti

Metoda maximální věrohodnosti se provádí maximalizací věrohodnostní funkce vzhledem k neznámým parametrům předpokládaného pravděpodobnostního rozdělení. Předpokládáme klasický lineární model a tedy  $\mathbf{y} \sim N_N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_N)$ . Vzhledem k nezávislosti  $N$  normálně rozdělených veličin  $y_i, i = 1, \dots, N$ , má věrohodnostní funkce tvar

$$l(\mathbf{y}, \mathbf{X}; \boldsymbol{\beta}, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left[-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right].$$

Nyní celou rovnici zlogaritmujeme

$$L(\mathbf{y}, \mathbf{X}; \boldsymbol{\beta}, \sigma^2) = \ln[l(\mathbf{y}, \mathbf{X}; \boldsymbol{\beta}, \sigma^2)] = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}$$

a nakonec maximalizujeme logaritmus věrohodnostní funkce vzhledem k  $\boldsymbol{\beta}$  a dojdeme ke stejnému výsledku jako při odhadu  $\boldsymbol{\beta}$  pomocí metody nejmenších čtverců

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

### 5.5.1.3 Vlastnosti odhadu $\beta$

1. Lineární odhad

Řekneme, že  $\hat{\beta}$  je lineárním odhadem vektoru  $\beta$ , existuje-li taková matice  $M = X^T X^{-1} X^T$  o rozměrech  $k \times N$ , že  $\beta = M y$ .

2. Nevychýlený odhad

$$E(\hat{\beta}) = E[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T X \beta = \beta, \quad \forall N, \forall \beta$$

3. Rozptyl

$$\begin{aligned} \text{var}(\hat{\beta}) &= \text{var}[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T \text{var}(y) X (X^T X)^{-1} \\ &= (X^T X)^{-1} X^T \sigma^2 I_N X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1} \end{aligned}$$

4. Pokud navíc  $e \sim N_N$ , pak

$$\hat{\beta} \sim N_N(\beta, \sigma^2 (X^T X)^{-1})$$

$$\hat{y} \sim N_N(X\beta, \sigma^2 P), \quad \text{kde } P = X(X^T X)^{-1} X^T \text{ je projekční matice.}$$

5. Nejlepší nestranný lineární odhad (NNLO)

Odhad  $\beta$  je NNLO, má-li nejmenší rozptyl mezi všemi lineárními nestrannými odhady tohoto parametru, ve smyslu  $\text{var}\tilde{\beta} - \text{var}\hat{\beta} \geq 0$

6. Eficientní odhad

**Definice 5.5.1** *Nechť  $f_n(y, \theta)$  je sdružená hustota náhodného výběru  $y = (Y_1, \dots, Y_n)^T$ , kde  $\theta \in \Theta \subseteq \mathbb{R}^k$  je parametr. Označ  $\psi(y, \theta) = \frac{d}{d\theta} \ln f_n(y, \theta)$ . Potom  $J(\theta) = E[\psi(y, \theta) \psi^T(y, \theta)]$  se nazývá Fisherova informační matice o parametru  $\theta$ .*

**Definice 5.5.2** *Nestranný odhad  $\hat{\theta}_n$  se nazývá eficientní, jestliže  $\text{var}\hat{\theta}_n = J^{-1}(\theta)$ , tj. má nejmenší rozptyl mezi všemi nestrannými odhady.*

## 6 ZOBECNĚNÉ LINEÁRNÍ MODELY

Je patrné, že podmínky a předpoklady klasického lineárního modelu a metody nejmenších čtverců, jsou často v regresních úlohách porušeny. V modelu se mohou vyskytovat náhodné vysvětlující proměnné, což souvisí s jejich vzájemnou závislostí, která se týká problematiky multikolinearity (není předmětem práce). Uvedená závislost je jeden z nejdůležitějších problémů regresních úloh, jelikož vede na špatné výsledky regresních odhadů. Dalším problémem lineárního regresního modelu shledáváme nestejné rozptyly  $\mathbf{Y}$  v podmíněných rozděleních, což nazýváme heteroskedasticita, a nakonec autokorelaci. Řešení nalzáme v zobecněných lineárních modelech (ZLM), ve kterých se vyskytují nové předpoklady a podmínky, kterými se vytváří nová omezení. ZLM obsahují regresní modely, obecné lineární modely, logistickou regresi a loglineární modely. Celá kapitola 6 je čerpána z literatury [11], [16] a [18].

Zobecněný lineární model chápeme jako klasický lineární model s uvedenými podmínkami v kapitole 5.5, ve kterém je změněna varianční matice náhodné složky  $\boldsymbol{\varepsilon}$  za předpokladu nenáhodnosti  $\mathbf{X}$ .

$$\text{var}(\boldsymbol{\varepsilon}) = \text{var}(\mathbf{y}) = \boldsymbol{\Omega} = \sigma^2 \mathbf{V}.$$

Předpokládá se, že rozptyly  $\text{var}(\boldsymbol{\varepsilon}_i)$  náhodných veličin  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$  nemusí být nutně stejné, ale jsou to nějaké kladné konstanty

$$\text{var}(\varepsilon_i) = \sigma_i^2, \quad i = 1, 2, \dots, N.$$

Dále mohou být některé nebo všechny dvojice veličin  $\varepsilon_i$  a  $\varepsilon_j$  korelované (lineárně závislé) a mimodiagonální prvky kovarianční matice  $\text{cov}(\boldsymbol{\varepsilon}) = \text{cov}(\mathbf{y})$  jsou nějaké nenulové konstanty

$$\text{cov}(\varepsilon_i, \varepsilon_j) = \sigma_{ij}, \quad i \neq j = 1, 2, \dots, N.$$

Kovarianční matice  $\boldsymbol{\Omega}$  má tři základní vlastnosti. V první řadě je symetrická, tedy platí  $\sigma_{ij} = \sigma_{ji}, \forall i, j$ . Jelikož musí platit nerovnost

$$|\sigma_{ij}| \leq \sqrt{\sigma_{ii}\sigma_{jj}},$$

je matice  $\boldsymbol{\Omega}$  pozitivně semidefinitní. V dalším textu budeme uvažovat přísnější předpoklad a to, že je matice  $\boldsymbol{\Omega}$  pozitivně definitní. Musí tedy platit

$$|\sigma_{ij}| < \sqrt{\sigma_{ii}\sigma_{jj}}.$$

A nakonec má na diagonále kladná čísla, což vyplývá z vlastností rozptylů za předpokladu jejich nenulovosti. Stejně vlastnosti má i matice  $\mathbf{V}$ . Pokud známe matici  $\boldsymbol{\Omega}$  můžeme určit



$V$  a odhadnout  $\sigma^2$ , ve většině případů je ale nutné zavést dodatečné předpoklady o rozptylech a kovariancích. Dále budeme brát v úvahu varianční matici ve tvaru  $\text{var}(\boldsymbol{\varepsilon}) = \text{var}(\mathbf{y}) = \boldsymbol{\Omega} = \sigma^2 \mathbf{V}$ , kde  $\mathbf{V}$  je známá matice a  $\sigma^2 > 0$ , která zabezpečuje normovaný součet diagonálních prvků matice  $\mathbf{V}$ , jejíž stopa je rovna  $N$ .

Platí-li ostatní předpoklady klasické metody nejmenších čtverců, můžeme odhadnout parametr  $\boldsymbol{\beta}$  z lineárního regresního modelu  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  minimalizací kritéria zobecněných nejmenších čtverců

$$\min (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Po úpravách daného výrazu dostaneme odhad  $\hat{\boldsymbol{\beta}}_z$ , který se rovněž nazývá Aitkenův odhad, ve tvaru

$$\hat{\boldsymbol{\beta}}_z = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}.$$

Jelikož matice  $\mathbf{V}$  je symetrická a pozitivně definitní, pak existuje matice  $\mathbf{P}$  typu  $N \times N$  taková, že platí  $\mathbf{PVP}^T = \mathbf{I}_N$  a  $\mathbf{P}^T \mathbf{P} = \mathbf{V}^{-1}$ . Vztah  $\mathbf{P}^T \mathbf{P} = \mathbf{V}^{-1}$  získáme pomocí vlastních čísel a ortonormálních vlastních vektorů, čemuž říkáme spektrální rozklad.

$$\mathbf{V}^{-1} = \mathbf{U}^T \boldsymbol{\Lambda} \mathbf{U} = \mathbf{U}^T \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\Lambda}^{\frac{1}{2}} \mathbf{U} = \mathbf{P}^T \mathbf{P},$$

kde  $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_N\}$  je diagonální matice vlastních čísel  $\lambda_1, \dots, \lambda_N > 0$ ,  $\mathbf{U}$  je matice ortonormálních vlastních vektorů a  $\mathbf{P}$  je regulární matice.

Nyní vynásobíme rovnici  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  maticí  $\mathbf{P}$  zleva, čímž dostáváme tvar

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}.$$

Pro  $\mathbf{P}\mathbf{y} = \mathbf{w}$ ,  $\mathbf{P}\mathbf{X} = \mathbf{W}$  a  $\mathbf{P}\boldsymbol{\varepsilon} = \mathbf{q}$  dostáváme transformovaný model ve tvaru

$$\mathbf{w} = \mathbf{W}\boldsymbol{\beta} + \mathbf{q},$$

který má všechny vlastnosti klasického lineárního modelu včetně podmínky na varianční matici

$$\text{var}(\mathbf{q}) = \mathbf{P}\text{var}(\boldsymbol{\varepsilon})\mathbf{P}^T = \mathbf{P}\sigma^2 \mathbf{V}\mathbf{P}^T = \sigma^2 \mathbf{PVP}^T = \sigma^2 \mathbf{I}_N.$$

Na model  $\mathbf{w} = \mathbf{W}\boldsymbol{\beta} + \mathbf{q}$  uijeme metodu nejmenších čtverců (MNČ), čímž dostaneme

$$\hat{\boldsymbol{\beta}} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{w} = (\mathbf{X}^T \mathbf{P}^T \mathbf{P} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}^T \mathbf{P} \mathbf{y} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} = \hat{\boldsymbol{\beta}}_z$$

Vidíme tedy, že zobecněný lineární model lze vhodnou transformací převést na klasický lineární model a tím použít pro výpočet odhadu  $\hat{\boldsymbol{\beta}}$  MNČ. Odhad parametru  $\boldsymbol{\beta}$  v zobecněném lineárním modelu má stejné vlastnosti jako odhad parametru  $\boldsymbol{\beta}$  v klasickém lineárním modelu, se změnou kovarianční matice odhadu parametrů  $\boldsymbol{\beta}$ , která má tvar

$$\begin{aligned} \text{var}(\hat{\beta}) &= \text{var}[(X^T V^{-1} X)^{-1} X^T V^{-1} y] = (X^T V^{-1} X)^{-1} X^T V^{-1} \text{var}(y) X V^{-1} (X^T V^{-1} X)^{-1} \\ &= (X^T V^{-1} X)^{-1} X^T V^{-1} \sigma^2 V X V^{-1} (X^T V^{-1} X)^{-1} = \sigma^2 (X^T V^{-1} X)^{-1}. \end{aligned}$$

## 6.1 Složky zobecněných lineárních modelů

Pomocí zobecněných lineárních modelů modelujeme funkce středních hodnot zkoumaných kategoriálních dat. Hlavním úkolem je vytvořit model, který nejlépe charakterizuje získaná data. Skládá se ze tří složek. První z nich je náhodná složka, reprezentována vysvětlovanou proměnnou  $\mathbf{Y}$  s nezávislými pozorováními  $(y_1, \dots, y_N)$  s hustotou pravděpodobnosti

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi) \right\},$$

kteřá se nazývá exponenciální rozptylová třída a  $\phi$  je parametr rozptylu. Pokud je  $\phi$  známé, můžeme funkci přepsat do tvaru

$$f(y_i; \theta_i) = a(\theta_i) b(y_i) \exp [y_i Q(\theta_i)],$$

kteřá má rozdělení patřící do exponenciální třídy. Hodnoty parametru  $\theta_i$  se mohou pro různá  $i = 1, \dots, N$  lišit, což závisí na velikostech vysvětlujících proměnných. Výraz  $Q(\theta_i)$  se nazývá přirozený parametr. Třída rozdělení, definovaná uvedenou hustotou, zahrnuje inverzní normální, gamma, Poissonovo a binomické rozdělení. V kapitole 6.2 bude popsáno Poissonovo rozdělení, které bude použito pro zkoumání dat této práce.

Druhá složka je systematická složka, která je charakterizována vztahem

$$\eta_i = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, N,$$

kde  $x_{ij}$  značí hodnotu prediktoru  $j$  ( $j = 0, 1, 2, \dots, k$ ) pro subjekt  $i$ . Uvedená lineární kombinace  $\eta_i$  vysvětlujících proměnných se nazývá lineární prediktor.

Poslední složka je transformační, neboli linková funkce, specifikující funkci střední hodnoty, kterou pokládá do rovnosti se systematickou složkou. Necht'  $\mu_i = E(Y_i)$ ,  $i = 1, \dots, N$ . Model propojuje  $\mu_i$  a  $\eta_i$  tak, že  $\eta_i = g(\mu_i)$ . Pak  $g$  propojuje  $E(Y_i)$  s vysvětlujícími proměnnými skrz rovnici

$$g(\mu_i) = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, N.$$

Linková funkce  $g(\mu) = \mu$ , se nazývá identická transformace, kde  $\eta_i = \mu_i$ , což je transformační funkce pro běžnou regresi s normálně rozděleným  $\mathbf{Y}$ . Transformační funkce,

kteřá mění střední hodnotu na přirozený parametr, se nazývá kanonická transformace, kde  $g(\mu_i) = Q(\theta_i)$  a  $Q(\theta_i) = \sum_j \beta_j x_{ij}$ .

Závěrem konstatujeme, že ZLM je lineární model pro transformovanou střední hodnotu vysvětlované proměnné, která má rozdělení v přirozené exponenciální třídě.

## 6.2 Poissonův loglineární model pro počty událostí

V kapitole 5.2.1 jsme zavedli Poissonovo rozdělení, které je vhodné použít pro modelování vysvětlovaných proměnných nabývajících nezáporných celočíselných hodnot. Pro naši práci s daty jsme vybrali Poissonovo rozdělení, jelikož vysvětlovanou proměnnou jsou počty dopravních nehod, které dané kritérium splňují.

Nechť  $Y$  značí počet a necht'  $\mu = E(Y)$ . Pravděpodobnostní funkce Poissonova rozdělení má tvar

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots$$

Funkci upravíme tak, abychom ji získali ve tvaru přirozené exponenciální formy

$$f(y; \mu) = e^{-\mu} \left(\frac{1}{y!}\right) e^{\log \mu^y} = \exp(-\mu) \left(\frac{1}{y!}\right) \exp(y \log \mu), \quad y = 0, 1, 2, \dots,$$

kde  $\theta = \mu$ ,  $a(\theta) = \exp(-\mu)$ ,  $b(y) = 1/y!$  a  $Q(\theta) = \log \mu$ . Přirozeným parametrem je  $\log \mu$ , z čehož plyne, že kanonickou transformační funkcí je logaritmus,  $\eta = \log \mu$ . Poissonův loglineární model zapíšeme ve tvaru

$$\log \mu_i = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, N,$$

kde lineárním prediktorem je lineární kombinace  $\sum_j \beta_j x_{ij}$ .

Maticový zápis

$$\log \mu = X\beta,$$

kde  $\mu = (\mu_1, \dots, \mu_N)^T$ ,  $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$  a  $X$  je matice vysvětlujících proměnných, tzv. dizajnová matice, o rozměru  $N \times p$ . Dále označme  $X\beta = \eta(x)$ , pak získáme tvar modelu

$$\log \mu = \eta(x)$$

Z rovnice vyjádříme střední hodnotu  $\mu$

$$\mu = \exp[\eta(x)].$$

Věnujme se nyní interpretaci parametrů. Pro jednoduchost budeme předpokládat pouze jedinou vysvětlující proměnnou  $x_1$ . Poissonův loglineární model bude mít tvar

$$\log \mu_i = \beta_0 + \beta_1 x_{i1}, \quad i = 1, \dots, N,$$

kde  $\beta_0$  je parametr jedné třídy vysvětlující proměnné a parametr  $\beta_1$  je parametr druhé třídy vysvětlující proměnné.

Odhad střední hodnoty bude mít tvar

$$\hat{\mu}_i = \exp(\beta_0 + \beta_1 x_{i1}) = e^{\beta_0} (e^{\beta_1})^{x_{i1}}, \quad i = 1, \dots, N,$$

kde  $e^{\beta_0}$  udává střední počet událostí při působení jedné třídy dané vysvětlující proměnné a  $e^{\beta_1}$  udává, kolikrát se vyskytlo více (méně při záporné hodnotě parametru  $\beta_1$ ) událostí pod vlivem první třídy oproti výskytu událostí pod vlivem druhé třídy dané vysvětlující proměnné.

### 6.2.1 Odhad parametru $\beta$

Jedním z možných způsobů, jak odhadnout parametry, je metoda maximální věrohodnosti, která byla zmíněna v kapitole 5.5.1. Předpokládejme model

$$\log \mu_i = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, N,$$

kde  $Y_i$  má Poissonovo rozdělení, tedy  $Y_i \sim \text{Po}(\mu_i)$ . Platí tedy

$$P(Y_i = y_i) = \frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i} = e^{-\mu_i + y_i \ln \mu_i - \ln(y_i!)}, \quad i = 1, \dots, N.$$

Sdružená pravděpodobnostní funkce má tvar

$$\prod_{i=1}^N P(Y = y_i) = \prod_{i=1}^N e^{-\mu_i + y_i \ln \mu_i - \ln(y_i!)}.$$

Víme, že

$$\ln \mu_i = \sum_j \beta_j x_{ij} \quad \text{tedy} \quad \mu_i = e^{\sum_j \beta_j x_{ij}}, \quad i = 1, \dots, N,$$

což dosadíme do sdružené pravděpodobnostní funkce, čímž získáme věrohodnostní funkci

$$l(\beta) = \prod_{i=1}^N e^{-e^{\sum_j \beta_j x_{ij}} + y_i \sum_j \beta_j x_{ij} - \ln(y_i!)}.$$

Pro snazší výpočet danou funkci zlogaritmuje a upravíme

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \ln l(\boldsymbol{\beta}) = \ln \left[ \prod_{i=1}^N e^{-e^{\sum_j \beta_j x_{ij}} + y_i \sum_j \beta_j x_{ij} - \ln(y_i!)} \right] \\
&= \sum_{i=1}^N \left( -e^{\sum_j \beta_j x_{ij}} + y_i \sum_j \beta_j x_{ij} - \ln(y_i!) \right) \\
&= \sum_{i=1}^N \left( -e^{\sum_j \beta_j x_{ij}} + y_i \sum_j \beta_j x_{ij} \right) - \sum_{i=1}^N \ln(y_i!).
\end{aligned}$$

Jelikož poslední člen funkce,  $-\sum_{i=1}^N \ln(y_i!)$ , neobsahuje hledaný parametr  $\boldsymbol{\beta}$ , můžeme tento člen zanedbat a získáme zjednodušenou funkci

$$L(\boldsymbol{\beta}) = \sum_{i=1}^N \left( y_i \sum_j \beta_j x_{ij} - e^{\sum_j \beta_j x_{ij}} \right) = \sum_{i=1}^N \left( y_i \sum_j \beta_j x_{ij} \right) - \sum_{i=1}^N e^{\sum_j \beta_j x_{ij}}.$$

Funkci budeme dále derivovat podle parametru  $\boldsymbol{\beta}$  a derivaci položíme rovnou nule

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^N y_i x_{ij} - \sum_{i=1}^N e^{\sum_j \beta_j x_{ij}} x_{ij} = 0.$$

S využitím vztahu  $\hat{\mu}_i = e^{\sum_j \beta_j x_{ij}}$  můžeme rovnici přepsat do tvaru

$$\sum_{i=1}^N y_i x_{ij} - \sum_{i=1}^N \hat{\mu}_i x_{ij} = 0.$$

Maticově

$$\mathbf{X}^T \mathbf{y} = \mathbf{X}^T \hat{\boldsymbol{\mu}},$$

kde  $\mathbf{X}$  je dizajnová matice typu  $N \times p$ .

Pro odhad  $\hat{\boldsymbol{\beta}}$  použijeme iterační Newton-Raphsonovu metodu, jelikož věrohodnostní rovnice jsou nelineární v parametru  $\boldsymbol{\beta}$ .

### 6.2.1.1 Newton-Raphsonova metoda

Newton-Raphsonova metoda je iterativní metoda, pomocí níž odhadujeme parametr  $\boldsymbol{\beta}$  a tedy i  $\boldsymbol{\mu}$ . Odhady v  $t$ -té iteraci označujeme  $\boldsymbol{\beta}^{(t)}, \boldsymbol{\mu}^{(t)}$ , kde  $t = 0, 1, \dots$  označuje pořadí iterace. Metoda spočívá v maximalizaci funkce  $L(\boldsymbol{\beta})$  uvedené v kapitole 6.2.1 s využitím gradientu a hessiánu dané funkce.

Gradient vypočítáme pomocí první derivace funkce  $L(\boldsymbol{\beta})$  podle parametru  $\boldsymbol{\beta}$ , který budeme značit  $\mathbf{u}(\boldsymbol{\beta}) = (u_0, u_1, \dots, u_k)$ , kde

$$u_j = \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^N y_i x_{ij} - \sum_{i=1}^N e^{\sum_j \beta_j x_{ij}} x_{ij} = \sum_{i=1}^N y_i x_{ij} - \sum_{i=1}^N \mu_i x_{ij},$$

pro  $j = 0, 1, \dots, k$ .

Hessián vypočítáme pomocí druhé derivace funkce  $L(\boldsymbol{\beta})$  podle parametru  $\boldsymbol{\beta}$ , který budeme značit  $\mathbf{H}(\boldsymbol{\beta}) = (h_{jl})$ , kde

$$h_{jl} = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^N x_{ij} \frac{\partial \mu_i}{\partial \beta_l} = - \sum_{i=1}^N x_{ij} \left\{ \frac{\partial [\exp(\sum_h \beta_h x_{ih})]}{\partial \beta_l} \right\} = - \sum_{i=1}^N \mu_i x_{ij} x_{il},$$

pro  $j = 0, 1, \dots, k, l = 0, 1, \dots, k$ .

Pro  $t$ -tou iteraci mají prvky gradientu a prvky hessiánu tvar

$$\mathbf{u}_j^{(t)} = \sum_{i=1}^N (y_i - \mu_i^{(t)}) x_{ij} \quad \text{a} \quad h_{jl}^{(t)} = - \sum_{i=1}^N \mu_i^{(t)} x_{ij} x_{il}.$$

Nyní můžeme přejít k výpočtu odhadu parametru  $\boldsymbol{\beta}$ . V  $t$ -té iteraci ( $t = 0, 1, \dots$ ) budeme funkci  $L(\boldsymbol{\beta})$  aproximovat v bodě  $\boldsymbol{\beta}^{(t)}$  s využitím Taylorova rozvoje, ve kterém zanedbáme derivace třetího a vyššího řádu

$$L(\boldsymbol{\beta}) \approx L(\boldsymbol{\beta}^{(t)}) + \mathbf{u}^{(t)T} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}) + \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})^T \mathbf{H}^{(t)} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}).$$

Funkci budeme derivovat podle parametru  $\boldsymbol{\beta}$  a dostaneme tvar

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \approx \mathbf{u}^{(t)} + \mathbf{H}^{(t)} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}),$$

ze kterého lze získat odhad  $\boldsymbol{\beta}^{(t+1)}$  v  $(t + 1)$ -ní iteraci

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\mathbf{H}^{(t)})^{-1} \mathbf{u}^{(t)}.$$

V kapitole 6.2.1.1 jsme odvodili vztah pro prvky hessiánu. Matici hessiánu v  $t$ -té iteraci lze tedy zapsat ve tvaru

$$\mathbf{H}^{(t)} = -\mathbf{X}^T \text{diag}(\boldsymbol{\mu}^{(t)}) \mathbf{X},$$

kde  $\text{diag}(\boldsymbol{\mu}^{(t)})$  je diagonální matice s diagonálními prvky  $\boldsymbol{\mu}^{(t)} = e^{\boldsymbol{\beta}^{(t)T} \mathbf{X}}$ .

Pak do vztahu pro odhad  $\boldsymbol{\beta}^{(t+1)}$  dosadíme hessián a gradient

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + (\mathbf{X}^T \text{diag}(\boldsymbol{\mu}^{(t)}) \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{y} - \boldsymbol{\mu}^{(t)}).$$

Počáteční volba  $\boldsymbol{\beta}^{(0)}$  ovlivňuje konvergenci, popř. nekonvergenci metody. Pokud je  $\boldsymbol{\beta}^{(0)}$  dobře zvoleno, pak s rostoucím  $t$   $\boldsymbol{\beta}^{(t)}, \boldsymbol{\mu}^{(t)}$  rychle konverguje k maximálně věrohodným odhadům  $\hat{\boldsymbol{\beta}}$  a  $\hat{\boldsymbol{\mu}}$ .

Pro velké  $t$  konvergence splňuje podmínku

$$\left| \beta_j^{(t-1)} - \hat{\beta}_j \right| \leq c \left| \beta_j^{(t)} - \hat{\beta}_j \right|^2, \quad \text{pro } \forall j \text{ a pro libovolné } c > 0.$$

a nazývá se kvadratická konvergence.

Hessián  $\mathbf{H}^{(t)}(\boldsymbol{\beta})$  konverguje k odhadu  $\hat{\mathbf{H}} = -\mathbf{X}^T \text{diag}(\hat{\boldsymbol{\mu}}) \mathbf{X}$ .

Iterační postup je ukončen ve chvíli, kdy se hodnoty parametru  $\boldsymbol{\beta}^{(t)}$  mění jen velmi málo.

### 6.2.2 Kovarianční matice odhadu parametru $\boldsymbol{\beta}$

Asymptotickou kovarianční matici získáme jako inverzi z informační matice, jejíž definici nalezneme v literatuře [12] definice 7.100. Informační matici jsme zmínili v definici 5.5.1. Prvky kovarianční matice jsou dány vztahem

$$E \left[ - \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_l} \right].$$

První i druhá derivace funkce  $L(\boldsymbol{\beta})$  byla odvozena v kapitole 6.2.1.1, po odvození jsme získali pro druhou derivaci tvar

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^N \mu_i x_{ij} x_{il}.$$

Jelikož log-lineární model je zobecněný lineární model užívající kanonický link, nezávisí druhá derivace (hessián) na pozorovaných datech  $y_i$ . Informační matice má tedy tvar

$$\mathbf{I} = \mathbf{X}^T \mathbf{W} \mathbf{X},$$

kde  $\mathbf{W} = \text{diag}(\boldsymbol{\mu})$  je diagonální matice.

Kovarianční matice odhadu parametru  $\hat{\boldsymbol{\beta}}$  má tvar

$$\text{cov}(\hat{\boldsymbol{\beta}}) = [\mathbf{X}^T \text{diag}(\boldsymbol{\mu}) \mathbf{X}]^{-1}.$$

Pokud do vzorce dosadíme odhad parametru  $\hat{\boldsymbol{\mu}}$  získáme odhad pro celou kovarianční matici

$$\widehat{\text{cov}}(\hat{\boldsymbol{\beta}}) = [\mathbf{X}^T \text{diag}(\hat{\boldsymbol{\mu}}) \mathbf{X}]^{-1}.$$

### 6.2.3 Hypotézy o parametru $\beta_j$ a intervaly spolehlivosti

V diplomové práci budeme chtít zjistit, jakou roli hrají jednotlivé parametry  $\beta_j$ . Budeme tedy testovat nulovou hypotézu  $H_0: \beta_j = \beta_j^{(0)}$ , kde  $\beta_j^{(0)} = 0$  proti alternativě  $H_A: \beta_j \neq 0$ . Předpokládejme model

$$\mathbf{Y} \sim N_N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{V}), \boldsymbol{\beta} \in \mathbb{R}^k, \sigma^2 \in (0, \infty), h(\mathbf{X}) = p < N,$$

kde  $h(\mathbf{X})$  je hodnota matice  $\mathbf{X}$ , která nám při splnění podmínky zaručuje větší počet pozorování, než je v modelu parametrů.

Uvedenou hypotézu můžeme testovat třemi způsoby. První z nich je použití Waldovy statistiky

$$z = \frac{(\hat{\beta}_j - \beta_j^{(0)})}{\sqrt{\widehat{\text{var}}(\hat{\beta})_{jj}}} = \frac{\hat{\beta}_j}{\sqrt{\widehat{\text{var}}(\hat{\beta})_{jj}}},$$

kde  $\sqrt{\widehat{\text{var}}(\hat{\beta})_{jj}} \neq 0$ . Pokud platí nulová hypotéza, má statistika  $z$  standardní normální rozdělení a  $z^2$  má pak  $\chi^2$  rozdělení o jednom stupni volnosti.

Interval spolehlivosti patřící k Waldově statistice  $z$  má tvar

$$\hat{\beta}_j \pm u_{1-\frac{\alpha}{2}} \sqrt{\widehat{\text{var}}(\hat{\beta})_{jj}},$$

kde  $u_{1-\frac{\alpha}{2}}$  je  $(1 - \frac{\alpha}{2})$ -kvantil z rozdělení  $N(0,1)$ .

Druhou metodou je test založený na poměru dvou věrohodnostních funkcí. Označme  $l_0$  maximum věrohodnostní funkce pro model za platnosti  $H_0$  a  $l_1$  je maximum věrohodnostní funkce bez ohledu na  $H_0$ . Pak platí

$$\Lambda = \frac{l_0}{l_1} \leq 1.$$

Testová statistika má obecný tvar

$$-2 \ln \Lambda = -2 \ln \frac{l_0}{l_1} = -2(L_0 - L_1),$$

kde  $L_0, L_1$  jsou maximální logaritmičké věrohodnostní funkce. Pokud budeme uvažovat Poissonovo rozdělení, platí vztahy

$$l_0 = e^{-\hat{\mu}_0 + y \ln \hat{\mu}_0 - \ln(y!)}, \text{ pak}$$

$$L_0 = \ln l_0 = \ln(e^{-\hat{\mu}_0 + y \ln \hat{\mu}_0 - \ln(y!)}) = -\hat{\mu}_0 + y \ln \hat{\mu}_0 - \ln(y!),$$

$$l_1 = e^{-\hat{\mu} + y \ln \hat{\mu} - \ln(y!)}, \text{ pak}$$

$$L_1 = \ln l_1 = \ln(e^{-\hat{\mu} + y \ln \hat{\mu} - \ln(y!)}) = -\hat{\mu} + y \ln \hat{\mu} - \ln(y!).$$

Po dosazení uvedených vztahů do obecného tvaru testové statistiky  $-2(L_0 - L_1)$  dostaneme testovou statistiku pro Poissonův parametr

$$-2(L_0 - L_1) = -2(y \ln \hat{\mu}_0 - \hat{\mu}_0 - (y \ln \hat{\mu} - \hat{\mu})) = 2 \left( y \ln \frac{\hat{\mu}}{\hat{\mu}_0} + \hat{\mu}_0 - \hat{\mu} \right),$$

kteřá má za platnosti  $H_0$  asymptoticky  $\chi^2$  rozdělení o jednom stupni volnosti.



Poslední způsob testování dané nulové hypotézy využívá skórové statistiky. Test je založen na sklonu a očekávaném zakřivení logaritmicke věrohodnostní funkce  $L(\boldsymbol{\beta})$  v bodě  $\beta_j^{(0)}$ . Označme

$$u(\beta_j^{(0)}) = \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j^{(0)}},$$

$$v(\beta_j^{(0)}) = -E \left[ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial (\beta_j^{(0)})^2} \right].$$

Pro skórovou statistiku pak platí

$$s = \frac{u(\beta_j^{(0)})}{\sqrt{v(\beta_j^{(0)})}},$$

kteřá má za předpokladu platnosti nulové hypotézy standardní normální rozdělení a  $s^2$  má pak  $\chi^2$  rozdělení o jednom stupni volnosti.

#### 6.2.4 Deviance

Pokud chceme zjistit, zda jsme vytvořili správný model pro popis našich dat, použijeme devianci. Při hledání vhodného modelu, začneme tím nejsložitějším. Uvažujme model, který obsahuje všechny proměnné a všechny jejich možné interakce. Model se nazývá saturevaný. Takový model ale nemusí dostatečně vyhladit uvažovaná data, proto ho budeme dále zjednodušovat.

Nechť  $\tilde{\boldsymbol{\mu}}$  je odhad střední hodnoty  $\boldsymbol{\mu}$  saturevaného modelu. Pak pro jednotlivá pozorování  $y_i$  platí  $\tilde{\mu}_i = y_i$ , pro  $\forall i = 1, \dots, N$ , maticově  $\tilde{\boldsymbol{\mu}} = \mathbf{y}$ . Maximální hodnotu logaritmu věrohodnostní funkce uvedeného saturevaného modelu značíme  $L(\mathbf{y}; \mathbf{y})$ . Maximální hodnotu logaritmu věrohodnostní funkce saturevaného modelu použijeme pro hodnocení jiného modelu, který už není saturevaný, což znamená, že má méně parametrů. Maximální hodnotu logaritmu věrohodnostní funkce nesaturevaného modelu budeme značit  $L(\hat{\boldsymbol{\mu}}; \mathbf{y})$ , kde  $\hat{\boldsymbol{\mu}}$  je odhad parametru  $\boldsymbol{\mu}$  nesaturevaného modelu. Deviance nesaturevaného modelu je dána obecně statistikou

$$-2 \ln \frac{\text{maximalizace v\text{e}rohodnostn\text{i} funkce nesaturovan\text{e}ho modelu}}{\text{maximalizace v\text{e}rohodnostn\text{i} funkce saturovan\text{e}ho modelu}} \\ = -2 \ln \frac{l(\hat{\boldsymbol{\mu}}; \mathbf{y})}{l(\mathbf{y}; \mathbf{y})} = -2[L(\hat{\boldsymbol{\mu}}; \mathbf{y}) - L(\mathbf{y}; \mathbf{y})].$$

Deviance je statistikou testu pom\text{e}ru v\text{e}rohodnost\text{i}, ve kter\text{e}m nulov\text{a} hypot\text{e}za p\text{r}edstavuje platnost nesaturovan\text{e}ho modelu a alternativa p\text{r}edstavuje platnost saturovan\text{e}ho modelu.

V\text{i}me, \text{z}e

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\},$$

Necht'  $L_i = \ln f(y_i; \theta_i, \phi)$  je logaritmus v\text{e}rohodnostn\text{i} funkce pro  $y_i$ , pak logaritmus v\text{e}rohodnostn\text{i} funkce pro  $\mathbf{y}$  m\text{a} tvar

$$L(\boldsymbol{\beta}) = \sum_{i=1}^N L_i = \sum_{i=1}^N \ln f(y_i; \theta_i, \phi) = \sum_{i=1}^N \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^N c(y_i, \phi).$$

Jeliko\text{z} poslední \text{c}len funkce nezávis\text{i} na parametru  $\boldsymbol{\beta}$ , kter\text{y} je skryt\text{y} v parametru  $\boldsymbol{\theta}$ , ve vztahu pro logaritmus v\text{e}rohodnostn\text{i} funkce nevystupuje

$$L(\boldsymbol{\beta}) = \sum_{i=1}^N \frac{y_i \theta_i - b(\theta_i)}{a(\phi)},$$

kter\text{y} dosad\text{i}me do vztahu pro devianci

$$-2[L(\hat{\boldsymbol{\mu}}; \mathbf{y}) - L(\mathbf{y}; \mathbf{y})] = 2 \sum_{i=1}^N \frac{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)}{a(\phi)} - 2 \sum_{i=1}^N \frac{y_i \hat{\theta}_i - b(\hat{\theta}_i)}{a(\phi)} \\ = 2 \sum_{i=1}^N \frac{y_i (\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)}{a(\phi)}.$$

Pokud m\text{a} z\text{a}visle prom\text{e}nn\text{a}  $\mathbf{Y}$  Poissonovo rozd\text{e}len\text{i}, plat\text{i} pro nesaturovan\text{y} a saturovan\text{y} model vztahy

$$\hat{\theta}_i = \log \hat{\mu}_i \text{ a } \tilde{\theta}_i = \log y_i, \\ b(\hat{\theta}_i) = \exp(\hat{\theta}_i) = \hat{\mu}_i \text{ a } b(\tilde{\theta}_i) = y_i, \\ a(\phi) = 1,$$

kter\text{e} nyn\text{i} dosad\text{i}me do obecn\text{e}ho tvaru deviance a z\text{i}sk\text{a}me devianci pro Poisson\text{u}v model

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \begin{cases} 2 \sum_{i=1}^N \left( \frac{y_i (\ln y_i - \ln \hat{\mu}_i) - y_i + \hat{\mu}_i}{1} \right) = 2 \sum_{i=1}^N \left[ y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) - y_i + \hat{\mu}_i \right] & \text{pro } \hat{\mu}_i > 0, y_i > 0, \\ 2\hat{\mu}_i & \text{pro } \hat{\mu}_i > 0, y_i = 0, \\ 0 & \text{pro } y_i = \hat{\mu}_i \end{cases}$$

Uveden\text{y} tvar m\text{u}zete nal\text{e}zt na internetov\text{y}ch str\text{a}nk\text{a}ch [19].

Pokud je počet pozorování  $N$  pevný, pak má deviance přibližně  $\chi^2$  rozdělení s  $N - p$  stupni volnosti, kde  $p$  je počet parametrů nesaturovaného modelu.

Uvažujme dva modely, první  $M_0$  s odhadem střední hodnoty  $\hat{\mu}_0$  a druhý  $M_1$  s odhadem střední hodnoty  $\hat{\mu}_1$ , kde  $M_0$  je podmodelem  $M_1$ . Je zřejmé, že v prvním modelu odhadujeme méně parametrů a proto maximalizace logaritmu věrohodnostní funkce není větší než maximum logaritmu věrohodnostní funkce pro druhý model. Pro deviance tomu je opačně

$$D(\mathbf{y}; \hat{\mu}_1) \leq D(\mathbf{y}; \hat{\mu}_0).$$

Pro testování nulové hypotézy, že platí podmodel  $M_0$  použijeme vztah

$$-2[L(\hat{\mu}_0; \mathbf{y}) - L(\hat{\mu}_1; \mathbf{y})] = D(\mathbf{y}; \hat{\mu}_0) - D(\mathbf{y}; \hat{\mu}_1)$$

který má asymptoticky  $\chi^2$  rozdělení s počtem stupňů volnosti rovným rozdílu počtu parametrů uvedených modelů.

Pomocí nárůstu deviance, předpokládáme-li na začátku nejsložitější model, postupně vyhodnocujeme, nakolik je změna počtu parametrů pro model významná.

### 6.2.5 Akaikeho informační kritérium

Jelikož deviance je obdobou indexu determinace užívaného v klasických lineárních modelech, vede výběr vhodného modelu na složitější model, což není naší prioritou. Chceme získat model s co nejméně parametry, které ovšem co nejdůvěhodněji popisují datový soubor. Byly proto navrženy nejrůznější modifikace deviance zohledňující počet parametrů nenasycených modelů. Jednou z nich je Akaikeho informační kritérium (AIC), které je dáno vztahem

$$\text{AIC} = -2[L(\hat{\mu}; \mathbf{y}) - p] = D(\mathbf{y}; \hat{\mu}) + 2p,$$

kde  $p$  je počet parametrů nesaturovaného modelu.

Při výběru jednoho z více modelů vybereme dle Akaikeho informačního kritéria ten, který má nejnižší hodnotu.

## 7 UŽITÍ LOGLINEÁRNÍHO MODELU – ANALÝZA DOPRAVNÍ NEHODOVOSTI

Problematika dopravní nehodovosti je věčné téma, proto budu v dalším textu analyzovat příčiny dopravních nehod v obci Zlín, ze které pocházím. Jak již bylo řečeno, nebudu zkoumat dopravní nehody na celých úsecích silnic města Zlín, ale pouze na přibližně vybraných částech uvedených v kapitole 3.2. Budu se snažit popsat vliv různých faktorů a charakteristik dopravních nehod na počet dopravních nehod. V příloze CD, soubor excel – vše můžete vidět zpracovaná data za období 2004 - 2011, která jsem získala z *Dopravního inspektorátu Policie České republiky Zlín* ve formě uvedené v příloze 2 a zpracované podle přílohy 3.

V bakalářské práci jsem se již uvedeným problémem zabývala. Analyzovala jsem počty dopravních nehod v letech 2004 – 2008 na daných úsecích silnic v souvislosti s některým faktorem za pomoci Kruskal-Wallisova testu. S využitím Paretova diagramu jsem určila, které hlavní příčiny mají největší vliv na vznik dopravních nehod. Rovněž jsem v bakalářské práci zavedla kontingenční tabulky, které jsem pro připomenutí zmínila i v teoretické části této práce. Budu se snažit objasnit souvislost mezi některými faktory.

Pro přehlednost jsem sečetla počet nehod vzhledem k jednotlivým faktorům, což můžete shlédnout v příloze CD, soubor excel – faktory. Ve většině případů způsobuje nejvíce dopravních nehod jedna třída každého faktoru. Faktory rozdělíme do dvou skupin, tedy na faktory charakterizující počet dopravních nehod a na faktory ovlivňující vznik dopravní nehody. V první skupině začneme testování počtu dopravních nehod v závislosti na dnech v týdnu. V dalších krocích budeme přidávat faktory a hledat model, který získaná data nejlépe popisuje. Stejným způsobem budeme pracovat i ve druhé skupině faktorů. K výpočtu použijeme software R, který už má nadefinovaný příkaz pro zobecněné lineární modely. Příkaz, pomocí kterého získáme odhady parametrů modelu, deviance, směrodatné odchylky, p-hodnoty odhanutých parametrů na základě hypotézy uvedené v kapitole 6.2.3, hodnotu Akaikeho informačního kritéria, atd., má tvar  $model < -glm(Y \sim x_1 + x_2 + \dots, family = "poisson")$ . Pokud budeme uvažovat interakce, model bude ve tvaru  $model < -glm(Y \sim x_1 * x_2 * \dots, family = "poisson")$ . Složka  $family = "poisson"$  určuje, že závisle proměnná má Poissonovo rozdělení a použijeme tedy log-lineární model.

## 7.1 Faktory charakterizující počet dopravních nehod

V následujících podkapitolách budeme postupně vytvářet model, který obsahuje faktory charakterizující počet dopravních nehod a jenž co nejlépe popisuje získaná data. Začneme jedním faktorem dny v týdnu a skončíme šesti faktory dny v týdnu, měsíc, místo, druh nehody, zavinění dopravní nehody a druh vozidla, které zavinilo nehodu.

### 7.1.1 Dny v týdnu

Pro jednoduchou ukázkou toho, jak odpovídající loglineární model hledáme, použijeme pouze faktor dny v týdnu. V následující tabulce je uveden počet nehod pro jednotlivé dny, které nejsou uspořádány, jak jsme zvyklí, jelikož software R faktory uspořádává podle abecedy.

TABULKA 7.1.1.1 Počet nehod v jednotlivých dnech

DEN	ČT	NE	PÁ	PO	SO	ST	ÚT
POČET NEHOD	149	58	145	155	82	140	130

Začneme situací, ve které předpokládáme, že počet nehod je různý pro každý den.

$$\ln[EY(den)] = \ln \mu_i = \beta_0 + \beta_1 I_{[den=NE]} + \beta_2 I_{[den=PÁ]} + \beta_3 I_{[den=PO]} + \beta_4 I_{[den=SO]} + \beta_5 I_{[den=ST]} + \beta_6 I_{[den=ÚT]}, \quad i = 1, \dots, 7.$$

První model nazýváme saturovaný, jelikož využijeme všechny možné parametry. Pomocí programu R jsme získali odhady parametrů, standardní odchylky, p-hodnoty daného modelu a odhad střední hodnoty. Ve všech tabulkách bude uveden pouze odhad středních hodnot ve tvaru  $\exp(\beta_0 + \beta_j)$ , kde  $j$  nabývá hodnot v závislosti na počtu parametrů. Tyto odhady budou přepočteny na jeden konkrétní den, což je zobrazeno v posledním sloupci tabulky. (víme, že rok má 52 týdnů a my zkoumáme data za 8 let, budeme tedy každý odhad  $\mu_i, i = 1, \dots, 7$  dělit číslem 416)

TABULKA 7.1.1.2 vlastnosti modelu 1

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepoččet odhadu	1 nehoda pro daný den za x týdnů
$\beta_0$	5,004	0,082	< 0,0001	149	0,358	$x = 2$ až 3
$\beta_1$	-0,944	0,155	< 0,0001	58	0,139	$x = 7$ až 8
$\beta_2$	<b>-0,027</b>	<b>0,117</b>	<b>0,816</b>	<b>145</b>	<b>0,349</b>	<b><math>x = 2</math> až 3</b>
$\beta_3$	<b>0,039</b>	<b>0,115</b>	<b>0,731</b>	<b>155</b>	<b>0,373</b>	<b><math>x = 2</math> až 3</b>
$\beta_4$	-0,597	0,138	< 0,0001	82	0,197	$x = 5$ až 6
$\beta_5$	<b>-0,062</b>	<b>0,118</b>	<b>0,597</b>	<b>140</b>	<b>0,337</b>	<b><math>x = 2</math> až 3</b>
$\beta_6$	<b>-0,136</b>	<b>0,12</b>	<b>0,256</b>	<b>130</b>	<b>0,313</b>	<b><math>x = 3</math> až 4</b>

Vidíme tedy, že odhad středních hodnot je identický s počtem dopravních nehod v jednotlivých dnech, což se projeví v hodnotě deviance, která by se měla rovnat nule. Vidíme, že některé parametry jsou nevýznamné. Testovali jsme hypotézu  $H_0: \beta_j = 0$ . Přesnou interpretaci hypotézy uvedeme pro parametr  $\beta_2$ . Odhad  $\mu_1 = \exp(\beta_0)$  je střední počet nehod ve čtvrtek a odhad  $\mu_3 = \exp(\beta_0 + \beta_2)$  je střední počet nehod v pátek. Testujeme tedy hypotézu  $H_0: \beta_2 = 0$ , což je ekvivalentní s  $H_0: EY(\text{čtvrtek}) = EY(\text{pátek})$ , tj. hypotézou, že střední počet nehod v pátek se významně neliší od středního počtu nehod ve čtvrtek. Hypotézu jsme provedli pro všechny parametry, čímž nám vyšlo, že střední počet dopravních nehod se významně neliší od čtvrtku v pondělí, úterý, středu ani pátek, proto tyto dny můžeme sloučit, což uděláme v dalším modelu.

Druhý model bude mít tvar

$$\ln[EY(\text{den})] = \ln \mu_i = \beta_0 + \beta_1 I_{[\text{den}=\text{SO}]} + \beta_2 I_{[\text{den}=\text{NE}]}, \quad i = 1, \dots, 7.$$

Opět využijeme program R a získáme obdobnou tabulku jako předchozí

TABULKA 7.1.1.3 Vlastnosti modelu 2

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepoččet odhadu	1 nehoda pro daný den za x týdnů
$\beta_0$	4,968	0,037	< 0,0001	143,8	0,346	$x = 2$ až 3
$\beta_1$	-0,562	0,117	< 0,0001	82	0,197	$x = 5$ až 6
$\beta_2$	-0,908	0,137	< 0,0001	58	0,139	$x = 7$ až 8

Zde představuje odhad  $\mu_i = \exp(\beta_0)$ ,  $i = 1, \dots, 5$  střední počet dopravních nehod v jeden všední den za všech osm let. Pro přesnost se stalo průměrně 0,346 nehody

v pondělí, totéž platí pro úterý, středu, čtvrtek i pátek. Nehoda tedy vznikla průměrně každé druhé (až třetí) pondělí, každé druhé (až třetí) úterý, každou druhou (až třetí) středu, každý druhý (až třetí) čtvrtek a každý druhý (až třetí) pátek. Pokud bychom nerozlišovali všední dny, řekli bychom, že se průměrně stala jedna až dvě nehody týdně ve všední den. Uvedenou hodnotu jsme získali po vydělení odhadu  $\mu_i$  číslem 416. Odhad  $\mu_6 = \exp(\beta_0 + \beta_1)$  je střední počet nehod v sobotu a odhad  $\mu_7 = \exp(\beta_0 + \beta_2)$  představuje střední počet nehod v neděli. V tomto modelu už se nevyskytují nevýznamné parametry, ale i přesto model zjednodušíme. K zjednodušení nás vede podobnost odhadů parametrů  $\beta_1$  a  $\beta_2$  ( $\hat{\beta}_1 = -0,562$ ,  $\hat{\beta}_2 = -0,908$ ). Zajímá nás, zda by nemohlo platit  $\beta_1 = \beta_2$ , čímž bychom získali jen jeden parametr, který dále značíme  $\beta_1$ .

Ve třetím modelu budeme předpokládat dva parametry a to pro všední den a pro víkend.

$$\ln[EY(den)] = \ln \mu_i = \beta_0 + \beta_1 I_{[den=SO|NE]}, \quad i = 1, \dots, 7.$$

Nyní bude tabulka obsahovat pouze dva řádky.

TABULKA 7.1.1.4 Vlastnosti modelu 3

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepočtený odhad	1 nehoda pro daný den za x týdnů
$\beta_0$	4,968	0,037	< 0,0001	143,8	0,346	$x = 2$ až $3$
$\beta_1$	-0,72	0,092	< 0,0001	70	0,168	$x = 5$ až $6$

Interpretace odhadu  $\mu_i = \exp(\beta_0)$ ,  $i = 1, \dots, 5$  je shodná s předchozím modelem, tedy že střední počet dopravních nehod v jeden všední den za všech osm let je 143,8 a parametr  $\mu_i$ ,  $i = 6, 7$  představuje střední počet nehod v sobotu nebo neděli. Po přepočtu jsme tedy zjistili, že vzniklo 0,168 nehody v sobotu, což platí i pro neděli. Nehoda se uskutečnila průměrně každou pátou (až šestou) sobotu a každou pátou (až šestou) nedělí.

Poslední, čtvrtý model, který v této situaci použijeme, bude předpokládat, že počet dopravních nehod byl ve všech dnech shodný, což znamená, že použijeme pouze jeden parametr, kterému říkáme absolutní člen. Vznik tohoto modelu nemá žádné odůvodnění, chtěli jsme pouze vědět, jaké bude mít hodnoty. Tvar modelu je

$$\ln[EY(den)] = \ln \mu_i = \beta_0, \quad i = 1, \dots, 7.$$

Tabulka vlastností modelu se zmenší na jeden řádek.

TABULKA 7.1.1.5 Vlastnosti modelu 4

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepoččet odhadu	1 nehoda pro daný den za $x$ týdnů
$\beta_0$	4,81	0,034	< 0,0001	122,714	0,295	$x = 3$ až 4

Odhad  $\mu_i = \exp(\beta_0)$ ,  $i = 1, \dots, 7$  představuje střední počet dopravních nehod v kterýkoliv den. Po přepočtu se tedy průměrně stalo 0,295 nehod ve všech jednotlivých dnech v týdnu. Platí, že se průměrně stala nehoda každé třetí (až čtvrté) pondělí, každé třetí (až čtvrté) úterý, každou třetí (až čtvrtou) středu, každý třetí (až čtvrtý) čtvrtek, každý třetí (až čtvrtý) pátek, každou třetí (až čtvrtou) sobotu a každou třetí (až čtvrtou) neděli.

Nyní přistoupíme k výběru nejvhodnějšího modelu, k čemuž nám pomohou charakteristiky v následující tabulce.

TABULKA 7.1.1.6 Modely a jejich charakteristiky

model	AIC	Deviance	Stupně volnosti	srovnání modelů	rozdíl deviancí	rozdíl stupňů volnosti	95%-ní kvantil chí kvadrát rozdělení	P-hodnota
1.	60,2	0	0					
2.	54,7	2,5	4	1. a 2.	2,5	4	9,5	0,642
3.	56,8	6,7	5	1. a 3.	6,7	5	11,1	0,248
4.	125,7	77,5	6	1. a 4.	77,5	6	12,6	0
				2. a 3.	4,1	1	3,8	0,042

V tabulce 7.1.1.6 vidíme různé charakteristiky, které jsme popsali v kapitole 6.2 a které nám pomohou vybrat nejvhodnější model charakterizující počet dopravních nehod v závislosti na dnech v týdnu. Po srovnání modelu 1 s modelem 2 jsme získali nerovnici  $2,5 < \chi_4^2(0,95)$ , kde  $\chi_4^2(0,95) = 9,5$ , což nám povoluje přejít k modelu 2. Po srovnání modelu 1 s modelem 3 jsme získali nerovnici  $6,7 < \chi_5^2(0,95)$ , kde  $\chi_5^2(0,95) = 11,1$ , což nám rovněž povoluje přejít k jednoduššímu modelu a to k modelu 3. Poslední srovnání modelu 1 s modelem 4 nám už nepovolí přejít k modelu 4, jelikož jsme získali nerovnici  $77,5 > \chi_6^2(0,95)$ , kde  $\chi_6^2(0,95) = 12,6$ . Máme tedy k dispozici dva modely, které lze považovat za vhodné. Dle hodnot deviancí a Akaikeho informačních kritérií modelů 2 a 3, zvolíme druhý model jako nejvhodnější, což jsme ověřili i srovnáním těchto dvou modelů.



Po srovnání modelů jsme získali nerovnici  $4,1 > \chi_1^2(0,95)$ , kde  $\chi_1^2(0,95) = 3,8$ , čímž jsme zamítli hypotézu o platnosti jednoduššího modelu.

Nejvhodnější model pro počet nehod v závislosti na dnech v týdnu má tedy tvar

$$\ln[EY(den)] = \ln \mu_i = \beta_0 + \beta_1 I_{[den=SO]} + \beta_2 I_{[den=NE]}, \quad i = 1, \dots, 7.$$

Model nám říká, že se významně neliší počet dopravních nehod v jednotlivých všedních dnech. V sobotu bylo ovšem dopravních nehod významně méně, přesně asi  $\exp(-\beta_1) = 1,75$  krát méně nehod než v některý všední den. A nakonec v neděli se stalo ještě méně nehod, tedy asi  $\exp(-\beta_2) = 2,48$  krát méně nehod než v některý všední den. V obou případech jsme pro lepší interpretaci použili převrácenou hodnotu odhadů. Znaménko nám totiž pouze říká, zda je počet nehod vyšší, nebo nižší.

## 7.1.2 Dny v týdnu a měsíc vzniku dopravní nehody

Počet dopravních nehod v závislosti na dnech v týdnu byl ukázán v tabulce 7.1.1.1. Nyní si ukážeme, kolik nehod se stalo v závislosti na měsíci, což dále budeme kombinovat s faktorem dny v týdnu.

TABULKA 7.1.2.1 Počet nehod v jednotlivých měsících

březen	červen	červenec	duben	květen	leden	listopad	prosinec	říjen	srpen	únor	září
72	76	61	71	68	83	88	86	85	71	56	42

Při vytváření modelů jsme postupovali stejným způsobem jako při vytváření modelů v kapitole 7.1.1. Jak v této kapitole, tak i v dalších kapitolách už nebudeme postup hledání nejvhodnějšího modelu nijak rozebírat, jen je shrneme do tabulky. Pro tabulku 7.1.2.1 jsme vytvořili sedm modelů, které můžete vidět v následující tabulce spolu s jejich deviancí, AIC a stupni volnosti.

TABULKA 7.1.2.2 Přehled modelů

	Model	deviance	AIC	stupně volnosti
1.	$\ln[EY(\text{měsíc})] = \ln \mu_i = \beta_0, \quad i = 1, \dots, 12$	30,5	105,6	11
2.	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{měsíc}=\text{červenec} \text{únor} \text{září}]}, \quad i = 1, \dots, 12$	9,9	87	10
3.	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{měsíc}=\text{únor} \text{září}]}, \quad i = 1, \dots, 12$	11,7	88,762	10
4.	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{měsíc}=\text{září}]} + \beta_2 I_{[\text{měsíc}=\text{únor}]}, \quad i = 1, \dots, 12$	9,7	88,755	9

5.	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{měsíc}=\text{červenec} \text{únor} \text{září}]} + \beta_2 I_{[\text{měsíc}=\text{březen} \text{červen} \text{duben} \text{květen} \text{srpen}]}, i = 1, \dots, 12$	4,4	83,5	9
6.	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{měsíc}=\text{březen} \text{červen} \text{duben} \text{květen} \text{srpen}]} + \beta_2 I_{[\text{měsíc}=\text{leden} \text{listopad} \text{prosinec} \text{říjen}]} + \beta_3 I_{[\text{měsíc}=\text{září}]}, i = 1, \dots, 12$	0,9	81,9	8
7.	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{měsíc}=\text{červen}]} + \beta_2 I_{[\text{měsíc}=\text{červenec}]} + \beta_3 I_{[\text{měsíc}=\text{duben}]} + \beta_4 I_{[\text{měsíc}=\text{květen}]} + \beta_5 I_{[\text{měsíc}=\text{leden}]} + \beta_6 I_{[\text{měsíc}=\text{listopad}]} + \beta_7 I_{[\text{měsíc}=\text{prosinec}]} + \beta_8 I_{[\text{měsíc}=\text{říjen}]} + \beta_9 I_{[\text{měsíc}=\text{srpen}]} + \beta_{10} I_{[\text{měsíc}=\text{únor}]} + \beta_{11} I_{[\text{měsíc}=\text{září}]}, i = 1, \dots, 12$	0	97,1	0

Z tabulky 7.1.2.2 jsme dle Akaikeho informačního kritéria vybrali model se čtyřmi parametry, tedy šestý model. Pomocí softwaru R jsme z modelu odhadli parametry  $\beta$ , které jsou uvedeny v následující tabulce spolu s dalšími vlastnostmi.

TABULKA 7.1.2.3 Parametry  $\beta_j, j = 0,1,2,3$

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepoččet odhadu
$\beta_0$	4,069	0,092	< 0,0001	58,5	7,312
$\beta_1$	<b>0,202</b>	<b>0,107</b>	<b>0,0578</b>	<b>71,6</b>	<b>8,95</b>
$\beta_2$	0,38	0,107	0,0004	85,5	10,688
$\beta_3$	<b>-0,331</b>	<b>0,18</b>	<b>0,0655</b>	<b>42</b>	<b>5,25</b>

Zde jsme odhady  $\mu_i$  přepočítávány pomocí čísla 8, které zastupuje osm let. Ačkoliv nám v modelu vyšli nevýznamné parametry  $\beta_1$  a  $\beta_3$  po jejich odstranění jsme získali horší model, proto považujeme model se čtyřmi parametry za nejvhodnější.

Model nám tedy říká, že počet dopravních nehod se významně neliší v měsících červenec a únor. V jednotlivých měsících březen, červen, duben, květen a srpen se stalo  $\exp(\beta_1) = 1,2$  krát více nehod než v jednom z měsíců červenec, či únor. V jednotlivých měsících leden, listopad, prosinec a říjen se uskutečnilo  $\exp(\beta_2) = 1,5$  krát více nehod než v jednom z měsíců červenec, či únor. A nakonec v září se naopak stalo  $\exp(-\beta_3) = 1,4$  krát méně nehod než v jednom z měsíců červenec, či únor.

Nyní přejdeme k analýze počtu dopravních nehod v závislosti na dnech v týdnu a na měsících. Začneme nejsložitějším modelem, což znamená saturovaným modelem s interakcemi o 84 parametrech (pracujeme se 7 dny a 12 měsíci, tedy  $7 \times 12 = 84$ ). V kapitole 7.1.1 jsme vytvořili model pro dny v týdnu o třech parametrech ve tvaru

$\ln \mu_i = \beta_0 + \beta_1 I_{[den=SO]} + \beta_2 I_{[den=NE]}$ . Rovněž jsme si nachystali model pro počet dopravních nehod v závislosti na měsících ve tvaru

$$\ln \mu_i = \beta_0 + \beta_1 I_{[m\acute{e}s\acute{ic}=b\breve{r}ezen|c\check{e}rven|duben|kv\acute{e}ten|srpen]} + \\ \beta_2 I_{[m\acute{e}s\acute{ic}=leden|listopad|prosinec|r\acute{ı}jen]} + \beta_3 I_{[m\acute{e}s\acute{ic}=z\acute{a}r\acute{ı}], \quad i = 1, \dots, 12.$$

Proto přejdeme na model s interakcemi o 12 parametrech  $\beta_j$ ,  $i = 0, 1, \dots, 11$ , ve kterém sloučíme uvedené modely. Takto získaný model bude mít tvar

$$\ln \mu_i = \beta_0 + \beta_1 I_{[den=SO]} + \beta_2 I_{[den=NE]} + \beta_3 I_{[m\acute{e}s\acute{ic}=b\breve{r}ezen|c\check{e}rven|duben|kv\acute{e}ten|srpen]} \\ + \beta_4 I_{[m\acute{e}s\acute{ic}=leden|listopad|prosinec|r\acute{ı}jen]} + \beta_5 I_{[m\acute{e}s\acute{ic}=z\acute{a}r\acute{ı}] \\ + \beta_6 I_{[den=SO, m\acute{e}s\acute{ic}=b\breve{r}ezen|c\check{e}rven|duben|kv\acute{e}ten|srpen]} \\ + \beta_7 I_{[den=NE, m\acute{e}s\acute{ic}=b\breve{r}ezen|c\check{e}rven|duben|kv\acute{e}ten|srpen]} \\ + \beta_8 I_{[den=SO, m\acute{e}s\acute{ic}=leden|listopad|prosinec|r\acute{ı}jen]} \\ + \beta_9 I_{[den=NE, m\acute{e}s\acute{ic}=leden|listopad|prosinec|r\acute{ı}jen]} + \beta_{10} I_{[den=SO, m\acute{e}s\acute{ic}=z\acute{a}r\acute{ı}] \\ + \beta_{11} I_{[den=NE, m\acute{e}s\acute{ic}=z\acute{a}r\acute{ı}], \quad i = 1, \dots, 84.$$

V modelu se vyskytují nevýznamné parametry spojené s některými jednoduchými interakcemi, proto jsme tyto interakce postupně odstranili. V modelu zůstanou dvě interakce, které jsou slabě nevýznamné a model bez jejich vypuštění má hodnotu AIC rovnou 420,19. Po odstranění obou interakcí získáme model bez interakcí o 6 parametrech s hodnotou AIC = 420,59, což je velmi nepatrný vzrůst oproti modelu s interakcemi. Model s interakcemi je hůře interpretovatelný a má více parametrů, což pro nás není žádoucí stav. Do modelu budeme ještě přidávat další faktory, proto nechceme hned na začátku složitý model. Z tohoto důvodu považujeme model bez interakcí za nejvhodnější a jeho tvar je

$$\ln \mu_i = \beta_0 + \beta_1 I_{[den=SO]} + \beta_2 I_{[den=NE]} + \beta_3 I_{[m\acute{e}s\acute{ic}=leden|listopad|prosinec|r\acute{ı}jen]} \\ + \beta_4 I_{[m\acute{e}s\acute{ic}=b\breve{r}ezen|c\check{e}rven|duben|kv\acute{e}ten|srpen]} + \beta_5 I_{[m\acute{e}s\acute{ic}=z\acute{a}r\acute{ı}], \quad i = 1, \dots, 84.$$

Pro ověření, zda jsme začali správným modelem, jsme zkoušeli model s interakcemi o 12 parametrech různě upravovat a zjednodušovat, což je uvedeno v příloze CD, soubor R, dnymesic.R, nebo v souboru excel, erko, kde je zaznamenáno i AIC, deviance a stupně volnosti. Ze všech modelů jsme i zde dle Akaikeho informačního kritéria vybrali stejný model bez interakcí s šesti parametry.

Hodnoty spojené s parametry  $\beta_j$ ,  $j = 0, 1, \dots, 5$  jsou uvedeny v tabulce 7.1.2.3. Odhad  $\mu_i$  je střední počet dopravních nehod, které vznikly za všech 8 let v daném dni a měsíci a poslední sloupec tabulky představuje odhadnutý počet dopravních nehod

v jednom roce, v konkrétním dni a měsíci. Každý odhad  $\mu_i$  budeme v tomto případě tedy dělit počtem let, tedy čísle 8. Ještě jednou zde upozorníme, že ve všech tabulkách týkajících se odhadu parametru  $\beta$  bude uveden pouze odhad středních hodnot ve tvaru  $\mu_i = \exp(\beta_0 + \beta_j)$ , kde  $j$  nabývá hodnot v závislosti na počtu parametrů.

TABULKA 7.1.2.4 Parametry  $\beta_j$ ,  $j = 0, 1, \dots, 5$

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepočet odhadu
$\beta_0$	2,282	0,094	< 0,0001	9,793	1,224
$\beta_1$	-0,562	0,117	< 0,0001	5,584	0,698
$\beta_2$	-0,908	0,137	< 0,0001	3,95	0,494
$\beta_3$	0,379	0,107	< 0,0001	14,313	1,789
<b><math>\beta_4</math></b>	<b>0,202</b>	<b>0,106</b>	<b>0,058</b>	<b>11,986</b>	<b>1,498</b>
<b><math>\beta_5</math></b>	<b>-0,331</b>	<b>0,18</b>	<b>0,065</b>	<b>7,031</b>	<b>0,879</b>

V modelu se vyskytují dva málo významné parametry  $\beta_4, \beta_5$ . Parametr  $\beta_5$  jsme zkusili z modelu vynechat, což vedlo k horšímu modelu s  $AIC = 422,15$ , proto zůstaneme u původního modelu s 6 parametry.

Odhad  $\mu_i = \exp(\beta_0) = 9,793$ ,  $i = 15; 17; 18; 20; 21; 71; 73; 74; 76$  a  $77$  určuje střední počet dopravních nehod v některý pracovní den v kombinaci s měsícem červenec nebo únor. Odhad  $\mu_i = \exp(\beta_0 + \beta_1) = 5,584$ ,  $i = 19; 75$  představuje střední počet dopravních nehod v sobotu, které se staly v červenci a střední počet dopravních nehod v sobotu, které se staly v únoru. Odhad  $\mu_i = \exp(\beta_0 + \beta_2) = 3,95$ ,  $i = 16; 72$  představuje střední počet dopravních nehod v neděli, které se staly v červenci a střední počet dopravních nehod v neděli, které se staly v únoru.

Odhad

$$\mu_i = \exp(\beta_0 + \beta_3) = 14,313, \quad i =$$

36; 38; 39; 41; 42; 43; 45; 46; 48; 49; 50; 52; 53; 55; 56; 57; 59; 60; 62 a 63

zastupuje střední počet dopravních nehod kombinace některého pracovního dne a některého měsíce leden, listopad, prosinec nebo říjen.

Odhad

$$\mu_i = \exp(\beta_0 + \beta_4) = 11,986, \quad i =$$

1; 3; 5; 6; 7; 8; 10; 11; 13; 14; 22; 24; 25; 27; 28; 29; 31; 32; 34; 35; 64; 66; 67; 69 a 70

zastupuje střední počet dopravních nehod kombinace některého pracovního dne a některého měsíce březen, červen, duben, květen nebo srpen.

Odhad  $\mu_i = \exp(\beta_0 + \beta_5) = 7,031$ ,  $i = 78; 80; 81; 83$  a  $84$  zastupuje střední počet dopravních nehod kombinace některého pracovního dne a měsíce září.

Odhad  $\mu_i = \exp(\beta_0 + \beta_1 + \beta_3) = 8,162$  představuje střední počet dopravních nehod, které se staly v lednové (popřípadě listopadové, prosincové nebo říjnové) soboty.

Odhad  $\mu_i = \exp(\beta_0 + \beta_1 + \beta_4) = 6,835$  představuje střední počet dopravních nehod, které se staly v březnové (popřípadě červnové, dubnové, květnové nebo srpnové) soboty.

Odhad  $\mu_i = \exp(\beta_0 + \beta_1 + \beta_5) = 4,009$  představuje střední počet dopravních nehod, které se staly v zářijových sobotách.

Odhad  $\mu_i = \exp(\beta_0 + \beta_2 + \beta_3) = 5,773$  představuje střední počet dopravních nehod, které se staly v lednové (popřípadě listopadové, prosincové nebo říjnové) neděle.

Odhad  $\mu_i = \exp(\beta_0 + \beta_2 + \beta_4) = 4,834$  představuje střední počet dopravních, které se staly v březnové (popřípadě červnové, dubnové, květnové nebo srpnové) neděle.

Odhad  $\mu_i = \exp(\beta_0 + \beta_2 + \beta_5) = 2,836$  představuje střední počet dopravních nehod v neděli, které se staly v září.

Jak již bylo zmíněno, jsou tyto odhady za všech osm let dohromady v daném dni a měsíci. Slovní interpretace může být zavádějící a člověk ji může špatně pochopit, proto odhadnutý počet nehod v závislosti na jednotlivých kombinacích zobrazíme v tabulce 7.1.2.5, které jsme počet nehod zaokrouhlili na celé číslo.

TABULKA 7.1.2.5 Přehled odhadů  $\mu_i, i = 1, \dots, 84$

	ČT	NE	PÁ	PO	SO	ST
březen	12	5	12	12	7	12
červen	12	5	12	12	7	12
červenec	10	4	10	10	6	10
duben	12	5	12	12	7	12
květen	12	5	12	12	7	12
leden	14	6	14	14	8	14
listopad	14	6	14	14	8	14
prosinec	14	6	14	14	8	14
říjen	14	6	14	14	8	14
srpen	12	5	12	12	7	12
únor	10	4	10	10	6	10
září	7	3	7	7	4	7

Nyní si ukážeme čtyři způsoby, jak můžeme zadat do softwaru R zdánlivě stejný model. Modely se liší zadáním jejich vstupu. Na tuto skutečnost si musíme dávat veliký

pozor, jelikož pak dostáváme jiné interpretace odhadů, o čemž se v tomto příkladu přesvědčíme. Budeme stále pracovat s modelem, kde vystupují jako vysvětlující proměnné dny v týdnu a měsíce. Nebudeme pracovat s nejlepším modelem, ale pro jednoduchost budeme uvažovat model se třemi parametry ve tvaru

$$\ln \mu_i = \beta_0 + \beta_1 I_{[den=SO|NE]} + \beta_2 I_{[měsíc=červenec|únor|září]},$$

ve kterém index  $i$  bude pro různé přístupy nabývat různých hodnot.

### PŘÍSTUP 1

Prvním způsobem jsme hledali nejvhodnější model, ve kterém jsme vycházeli z nejsložitější tabulky, která obsahovala veškeré kombinace měsíce a dní v týdnu. Saturovaný model obsahuje 84 parametrů.

TABULKA 7.1.2.6 Kontingenční tabulky

	březen	červen	červenec	duben	květen	leden	listopad	prosinec	říjen	srpen	únor	září
ČT	11	7	18	16	11	11	11	17	15	16	4	12
NE	6	5	8	3	2	4	5	6	6	4	6	3
PÁ	10	14	6	15	14	9	18	12	19	10	11	7
PO	10	15	10	12	17	21	20	10	17	11	9	3
SO	8	11	4	5	4	11	8	9	8	5	6	3
ST	11	13	7	10	10	13	13	17	12	14	11	9
ÚT	16	11	8	10	10	14	13	15	8	11	9	5

Pro daný model bude  $i = 1, \dots, 84$ . Odhad parametru  $\beta$  uvedeného modelu jsme získali pomocí softwaru R. Tyto odhady jsou uvedeny v tabulce 7.1.2.7, spolu s jejich dalšími vlastnostmi. Ze stejných důvodů jako v tabulce 7.1.2.4, budeme odhad  $\mu_i$  přepočítávat pomocí čísla 8, což je počet let.

TABULKA 7.1.2.7 Parametry  $\beta_j$ ,  $j = 0,1,2$

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepočet odhadu
$\beta_0$	2,567	0,041	< 0,0001	13,02	1,628
$\beta_1$	-0,72	0,092	< 0,0001	6,338	0,792
$\beta_2$	-0,384	0,088	< 0,0001	8,872	1,109

Odhad  $\mu_i = \exp(\beta_0) = 13,02$  zastupuje střední počet dopravních nehod kombinace některého pracovního dne spolu s některým měsícem březen, červen, duben, květen, leden, listopad, prosinec, říjen nebo srpen.

Odhad  $\mu_i = \exp(\beta_0 + \beta_1) = 6,338$  představuje střední počet dopravních nehod v některém z víkendových dnů, které se staly v některém z měsíců červenec, únor, nebo září.

Odhad  $\mu_i = \exp(\beta_0 + \beta_2) = 8,872$  představuje střední počet dopravních nehod v některém z měsíců červenec, únor, nebo září v kombinaci s některým pracovním dnem.

Střední počet dopravních nehod, které se staly v některém víkendovém dni v kombinaci s některým měsícem červenec, únor nebo září, je charakterizován odhadem  $\mu_i = \exp(\beta_0 + \beta_1 + \beta_2) = 4,319$ .

## PŘÍSTUP 2

Nyní bude mít vstupní tabulka rozměr  $2 \times 12$ , kde budeme rozlišovat víkend a pracovní dny oproti všem měsícům. Saturovaný model by tedy měl 24 parametrů.

TABULKA 7.1.2.8 Kontingenční tabulka

	březen	červen	červenec	duben	květen	leden	listopad	prosinec	říjen	srpen	únor	září
prac. den	58	60	49	63	62	68	75	71	71	62	44	36
víkend	14	16	12	8	6	15	13	15	14	9	12	6

Pro daný model bude  $i = 1, \dots, 24$ . Odhad parametru  $\beta$  uvedeného modelu jsme získali pomocí softwaru R. Tyto odhady jsou uvedeny v tabulce 7.1.2.9, spolu s jejich dalšími vlastnostmi. Odhady  $\mu_i$  jsou uvedeny za všech osm let dohromady v daném měsíci o víkendu nebo v pracovní den. V posledním sloupci jsou odhady opět přepočteny na jeden rok, jeden den v týdnu a měsíc.

TABULKA 7.1.2.9 Parametry  $\beta_j$ ,  $j = 0,1,2$

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepočet odhadu
$\beta_0$	4,176	0,041	< 0,0001	65,102	0,407
$\beta_1$	-1,636	0,092	< 0,0001	12,676	0,198
$\beta_2$	-0,384	0,088	< 0,0001	44,362	0,277

Odhad  $\mu_i = \exp(\beta_0) = 65,102$  zastupuje střední počet dopravních nehod kombinace všech pracovních dní dohromady a některého měsíce březen, červen, duben, květen, leden, listopad, prosinec, říjen nebo srpen. Jelikož jsme brali v úvahu všechny pracovní dny, je odhad přepočten pomocí čísla 160. Opět jsme předpokládali, že v daném měsíci se jeden pracovní den vyskytne čtyřikrát, ale těch dnů je v jednom týdnu pět a navíc ještě musíme započítat osm let ( $4 \times 5 \times 8 = 160$ ).

Odhad  $\mu_i = \exp(\beta_0 + \beta_1) = 12,676$  představuje střední počet dopravních nehod o víkendu, které se staly v některém z měsíců v červenec, únor, nebo září. Postup pro

přepočtení odhadu je shodný, s tím rozdílem, že uvažujeme pouze víkend, tedy dva dny, čímž získáme hodnotu  $64 (4 \times 2 \times 8 = 64)$ .

Odhad  $\mu_i = \exp(\beta_0 + \beta_2) = 44,362$  představuje střední počet dopravních nehod v některém z měsíců červenec, únor, nebo září v kombinaci se všemi pracovními dny dohromady. Přepočtení je provedeno pomocí hodnoty 160, ze stejných důvodů jako první odhad.

Střední počet dopravních nehod, které se staly o víkendu v kombinaci s některým z měsíců červenec, únor nebo září, je charakterizován odhadem  $\mu_i = \exp(\beta_0 + \beta_1 + \beta_2) = 8,638$ .

### PŘÍSTUP 3

V tomto případě vyměníme faktory. Budeme uvažovat červenec, únor a září oproti ostatním měsícům a jednotlivé dny v týdnu. Vznikne tak tabulka o rozměru  $2 \times 7$ . Saturovaný model by měl 14 parametrů.

TABULKA 7.1.2.10 Kontingenční tabulka

	ČT	NE	PÁ	PO	SO	ST	ÚT
ostatní	115	41	121	133	69	113	108
červenec, únor, září	34	17	24	22	13	27	22

Pro daný model bude  $i = 1, \dots, 14$ . Odhad parametru  $\beta$  uvedeného modelu jsme získali pomocí softwaru R. Tyto odhady jsou uvedeny v tabulce 7.1.2.11, spolu s jejich dalšími vlastnostmi. Odhady  $\mu_i$  jsou uvedeny za všech osm let dohromady v daném měsíci o víkendu nebo v pracovní den. V posledním sloupci jsou odhady opět přepočteny na jeden rok, jeden den v týdnu a měsíc.

TABULKA 7.1.2.11 Parametry  $\beta_j, j = 0, 1, 2$

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepočtení odhadu
$\beta_0$	4,764	0,041	< 0,0001	117,183	0,407
$\beta_1$	-1,482	0,088	< 0,0001	26,617	0,277
$\beta_2$	-0,72	0,092	< 0,0001	57,043	0,198

Odhad  $\mu_i = \exp(\beta_0) = 117,183$  zastupuje střední počet dopravních nehod v ostatních měsících dohromady v kombinaci s některým pracovním dnem. Předpokládejme,



že jeden den se v měsíci vyskytne čtyřikrát. Dále bereme v úvahu devět měsíců a osm let. Celková hodnota, pomocí které jsme přepočítali odhad  $\mu_i$ , je  $4 \times 9 \times 8 = 288$ .

Odhad  $\mu_i = \exp(\beta_0 + \beta_1) = 26,617$  představuje střední počet dopravních nehod měsíce červenec, únor a září dohromady v kombinaci s některým pracovním dnem. Zde uvažujeme stejně jako při přepočtu předchozího odhadu se změnou počtu měsíců. Místo devíti měsíců pracujeme pouze se třemi. Celková hodnota pro přepočet odhadu  $\mu_i$  je  $4 \times 3 \times 8 = 96$ .

Odhad  $\mu_i = \exp(\beta_0 + \beta_2) = 57,043$  představuje střední počet dopravních nehod v některém víkendovém dni v kombinaci s ostatními měsíci dohromady. K hodnotě pro přepočet odhadu  $\mu_i$  dojdeme stejnou úvahou jako v prvním případě. Dělili jsme číslem 288.

Střední počet dopravních nehod, které se staly v některý víkendový den v kombinaci s měsícem červenec, únor nebo září dohromady, je charakterizován odhadem  $\mu_i = \exp(\beta_0 + \beta_1 + \beta_2) = 12,957$ .

#### PŘÍSTUP 4

Poslední tabulka bude mít rozměr  $2 \times 2$ , což znamená, že faktor den rozdělíme na víkend a pracovní den a faktor měsíc zjednodušíme na skupinu obsahující červenec, únor a září a druhou skupinu, která obsahuje zbytek měsíců. Saturovaný model by měl 4 parametry.

TABULKA 7.1.2.12 Kontingenční tabulka

MĚSÍC\DEN	prac.den	Víkend
ostatní	590	110
červenec, únor, září	129	30

Pro daný model bude  $i = 1, \dots, 4$ . Odhad parametru  $\beta$  uvedeného modelu jsme získali pomocí softwaru R. Tyto odhady jsou uvedeny v tabulce 7.1.2.11, spolu s jejich dalšími vlastnostmi. Odhady  $\mu_i$  jsou uvedeny za všech osm let dohromady v daném měsíci o víkendu nebo v pracovní den. V posledním sloupci jsou odhady opět přepočteny na jeden rok, jeden den v týdnu a měsíc.

TABULKA 7.1.2.13 Parametry  $\beta_j$ ,  $j = 0,1,2$

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$	přepoččet odhadu
$\beta_0$	6,373	0,041	< 0,0001	585,914	0,407
$\beta_1$	-1,636	0,092	< 0,0001	114,086	0,198
$\beta_2$	-1,482	0,088	< 0,0001	133,086	0,277

Odhad  $\mu_i = \exp(\beta_0) = 585,914$  zastupuje střední počet dopravních nehod v ostatních měsících dohromady v kombinaci se všemi pracovními dny dohromady. Předpokládejme, že jeden den se v měsíci vyskytne čtyřikrát a těchto dnů je v jednom týdnu pět. Dále bereme v úvahu devět měsíců a osm let. Celková hodnota, pomocí které jsme přepočítali odhad  $\mu_i$ , je  $4 \times 5 \times 9 \times 8 = 1440$ .

Odhad  $\mu_i = \exp(\beta_0 + \beta_1) = 114,086$  představuje střední počet dopravních nehod o víkendu v kombinaci s ostatními měsíci dohromady. Opět předpokládejme, že jeden den se v měsíci vyskytne čtyřikrát a tyto dny jsou nyní v jednom týdnu pouze dva. Dále bereme v úvahu devět měsíců a osm let. Celková hodnota, pomocí které jsme přepočítali odhad  $\mu_i$ , je  $4 \times 2 \times 9 \times 8 = 576$ .

Odhad  $\mu_i = \exp(\beta_0 + \beta_2) = 133,086$  vyjadřuje střední počet dopravních nehod v měsíci červenci, únoru a září dohromady v kombinaci se všemi pracovními dny dohromady. Zde uvažujeme stejně jako při přepočtu prvního odhadu se změnou počtu měsíců. Místo devíti měsíců pracujeme pouze se třemi. Celková hodnota pro přepoččet odhadu  $\mu_i$  je  $4 \times 5 \times 3 \times 8 = 480$ .

Střední počet dopravních nehod, které se staly o víkendu v kombinaci s měsícem červenec, únor nebo září dohromady, je charakterizován odhadem  $\mu_i = \exp(\beta_0 + \beta_1 + \beta_2) = 25,914$ .

## ZÁVĚR

Hodně záleží na tabulce, která vstupuje do modelu. Musíme si dávat pozor na to, co nám odhad  $\mu_i$  říká. Viděli jsme, že po přepočtu odhadu  $\mu_i$  na jeden rok, konkrétní den a měsíc jsme v každém modelu dospěli ke stejným hodnotám, čímž jsme si ověřili, že jsme správně interpretovali v daném modelu odhady  $\mu_i$ . Ze všech čtyř modelů nám tedy vyšlo, že se průměrně stalo 0,407 nehody v některém pracovním dni v kombinaci s některým měsícem ze skupiny ostatní v jednom roce (například se průměrně stalo 0,407 nehody v jedno lednové pondělí). Za uvedených podmínek se průměrně uskutečnily dvě nebo tři

dopravní nehody v týdnu. V některý víkendový den v kombinaci s některým měsícem ze skupiny ostatní a jednom roce vzniklo pouze 0,198 nehody (například vzniklo v průměru 0,198 nehody v jednu lednovou sobotu). Můžeme tedy říct, že za daných podmínek se v jednom měsíci průměrně stala jedna nebo dvě dopravní nehody. V některý pracovní den v kombinaci s některým z měsíců červenec, únor, září se uskutečnilo 0,277 (například se průměrně stalo 0,221 nehody v jedno únorové pondělí), což znamená, že se za uvedených podmínek stala jedna nebo dvě nehody v týdnu. A nakonec v některém víkendovém dni v kombinaci s některým měsícem červenec, únor nebo září se stalo 0,54 nehody (například se průměrně stalo 0,54 nehody v jednu únorovou sobotu). Za uvedených podmínek se staly dvě nebo tři nehody za měsíc. Příkazy pomocí kterých jsme v softwaru R získali uvedené hodnoty, jsou uvedeny v příloze CD, soubor R, denmesicsrovnani.R.

### 7.1.3 Dny v týdnu, měsíc a místo dopravní nehody

Další faktor, který přidáme do modelu, je místo dopravní nehody. Tabulka 7.1.3.1 uvádí přehled počtu dopravních nehod v jednotlivých místech.

TABULKA 7.1.3.1 Počet nehod v závislosti na místě

mimo křižovatku	mimo zónu 11-19 a 22-26	na křižovatce	uvnitř zóny 1-8 předm.kř.
554	3	297	5

Je vidět, že v místě mimo zónu 11-19 a 22-26 a uvnitř zóny 1-8 předmětem křižovatky se stal zanedbatelný počet dopravních nehod. Proto počet nehod v místě mimo zónu 11-19 a 22-26 přičteme k počtu nehod v místě mimo křižovatku a počet nehod v místě uvnitř zóny 1-8 předmětem křižovatky přičteme k počtu nehod v místě na křižovatce. Touto úpravou nám vznikne tabulka o dvou faktorech a přejdeme rovnou k modelu, který se bude skládat z nezávislých proměnných dny v týdnu, měsíc a místo dopravní nehody.

Nejsložitější model s interakcemi se bude skládat ze 168 parametrů (7 dní, 12 měsíců a 2 možnosti místa). My dále využijeme model, kterým jsme začínali v kapitole 7.1.2 ve tvaru

$$\begin{aligned} \ln \mu_i = & \beta_0 + \beta_1 I_{[den=SO]} + \beta_2 I_{[den=NE]} + \beta_3 I_{[měsíc=březen|červen|duben|květen|srpen]} \\ & + \beta_4 I_{[měsíc=červenec,únor]} + \beta_5 I_{[měsíc=září]} \\ & + \beta_6 I_{[den=SO,měsíc=březen|červen|duben|květen|srpen]} \\ & + \beta_7 I_{[den=NE,měsíc=březen|červen|duben|květen|srpen]} \\ & + \beta_8 I_{[den=SO,měsíc=červenec,únor]} + \beta_9 I_{[den=NE,měsíc=červenec,únor]} \\ & + \beta_{10} I_{[den=SO,měsíc=září]} + \beta_{11} I_{[den=NE,měsíc=září]}, \quad i = 1, \dots, 84. \end{aligned}$$

a přidáme k němu faktor místo dopravní nehody. Do modelu můžeme jako další parametr přidat faktor na křižovatce nebo mimo křižovatku. Oba modely zde ukážeme.

Do prvního modelu přidáme faktor na křižovatce. Model s interakcemi bude mít 24 parametrů a hodnota AIC je rovna 750. V tomto modelu jsou veškeré interakce nevýznamné, proto je hned všechny odstraníme a získáme model s AIC = 732,35 ve tvaru

$$\begin{aligned} \ln \mu_i = & \beta_0 + \beta_1 I_{[místo=na\ křižovatce]} + \beta_2 I_{[den=SO]} + \beta_3 I_{[den=NE]} \\ & + \beta_4 I_{[měsíc=leden|listopad|prosinec|říjen]} \\ & + \beta_5 I_{[měsíc=březen|červen|duben|květen|srpen]} + \beta_6 I_{[měsíc=září]}, \quad i = 1, \dots, 168, \end{aligned}$$

jehož parametr  $\beta$  jsme odhadli pomocí softwaru R hodnoty. Tyto odhady jsou zobrazeny v tabulce 7.1.3.2 spolu s jejich dalšími vlastnostmi. Odhady jsou uvedeny za všech osm let dohromady v daném měsíci a v daném dni. V posledním jsou odhady  $\mu_i$  opět přepočteny na jeden rok, jeden den v týdnu, měsíc. Odhad  $\mu_i$  jsme dělili číslem 8.

TABULKA 7.1.3.2 Parametry  $\beta_j$ ,  $j = 0, 1, \dots, 6$

parametr	odhad	standartní odchylka	p-value	odhad	přepočet odhadu
$\beta_0$	1,848	0,097	< 0,0001	6,35	0,794
$\beta_1$	-0,612	0,071	< 0,0001	3,443	0,43
$\beta_2$	-0,562	0,117	< 0,0001	3,621	0,453
$\beta_3$	-0,908	0,137	< 0,0001	2,561	0,32
$\beta_4$	0,379	0,107	< 0,0001	9,281	1,16
<b><math>\beta_5</math></b>	<b>0,202</b>	<b>0,106</b>	<b>0,058</b>	<b>7,772</b>	<b>0,972</b>
<b><math>\beta_6</math></b>	<b>-0,331</b>	<b>0,18</b>	<b>0,065</b>	<b>4,559</b>	<b>0,57</b>

V modelu se vyskytují dva málo významné parametry  $\beta_5, \beta_6$ . Parametr  $\beta_5$  jsme zkusili z modelu vynechat, což vedlo k horšímu modelu s AIC = 733,9, proto zůstaneme u původního modelu se 7 parametry.

Odhad  $\mu_i = \exp(\beta_0) = 6,35$  určuje střední počet dopravních nehod v některý pracovní den v kombinaci s některým měsícem červenec nebo únor a v místě mimo

křižovatku. Odhad  $\mu_i = \exp(\beta_0 + \beta_1) = 3,443$  určuje střední počet dopravních nehod na křižovatce v některý pracovní den v kombinaci s některým měsícem červenec nebo únor. Odhad  $\mu_i = \exp(\beta_0 + \beta_2) = 3,621$  představuje střední počet dopravních nehod v sobotu, které se staly v červenci mimo křižovatku a střední počet dopravních nehod v sobotu, které se staly v únoru mimo křižovatku. Odhad  $\mu_i = \exp(\beta_0 + \beta_3) = 2,561$  představuje střední počet dopravních nehod v neděli, které se staly v červenci mimo křižovatku a střední počet dopravních nehod v neděli, které se staly v únoru mimo křižovatku. Odhad  $\mu_i = \exp(\beta_0 + \beta_4) = 9,281$  zastupuje střední počet dopravních nehod kombinace některého pracovního dne a některého měsíce leden, listopad, prosinec nebo říjen, které se staly mimo křižovatku. Odhad  $\mu_i = \exp(\beta_0 + \beta_5) = 7,772$  zastupuje střední počet dopravních nehod kombinace některého pracovního dne a některého měsíce březen, červen, duben, květen nebo srpen, které se staly mimo křižovatku. Odhad  $\mu_i = \exp(\beta_0 + \beta_6) = 4,559$  zastupuje střední počet dopravních nehod kombinace některého pracovního dne a měsíce září, které se staly mimo křižovatku. Zbytek odhadů v tabulce není uveden a nebudeme je zde ani zmiňovat, jelikož by to bylo zdlouhavé. Odhady se vytváří stejným způsobem jako v kapitole 7.1.2, podle toho, jakou kombinaci parametrů chceme zjistit.

Druhý model je shodný s předchozím až na změnu faktoru místo. Model s interakcemi má 24 parametrů a hodnota AIC je rovna 750 a po odstranění všech interakcí získáme opět model se 7 parametry a AIC = 732,4. Model má tvar

$$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{misto}=\text{mimo křižovatku}]} + \beta_2 I_{[\text{den}=\text{SO}]} + \beta_3 I_{[\text{den}=\text{NE}]} \\ + \beta_4 I_{[\text{měsíc}=\text{leden}|\text{listopad}|\text{prosinec}|\text{říjen}]} \\ + \beta_5 I_{[\text{měsíc}=\text{březen}|\text{červen}|\text{duben}|\text{květen}|\text{srpen}]} + \beta_6 I_{[\text{měsíc}=\text{září}]}, i = 1, \dots, 168$$

Opět zde uvedeme tabulku pro parametry  $\beta_j$ ,  $j = 0, 1, \dots, 6$  se stejnými vlastnostmi jako v prvním modelu.

TABULKA 7.1.3.3 Parametry  $\beta_j$ ,  $j = 0, 1, \dots, 6$

parametr	odhad	standartní odchylka	p-value	odhad	přepočtený odhad
$\beta_0$	1,236	0,105	< 0,0001	3,443	0,43
$\beta_1$	0,612	0,071	< 0,0001	6,35	0,794
$\beta_2$	-0,562	0,117	< 0,0001	1,963	0,245
$\beta_3$	-0,908	0,137	< 0,0001	1,389	0,174
$\beta_4$	0,379	0,107	< 0,0001	5,032	0,629
$\beta_5$	<b>0,202</b>	<b>0,106</b>	<b>0,058</b>	<b>4,214</b>	<b>0,527</b>
$\beta_6$	<b>-0,331</b>	<b>0,18</b>	<b>0,065</b>	<b>2,472</b>	<b>0,309</b>

V modelu se vyskytují dva málo významné parametry  $\beta_5, \beta_6$ . Parametr  $\beta_5$  jsme zkusili z modelu vynechat, což vedlo k horšímu modelu s  $AIC = 733,9$ , proto zůstaneme u původního modelu se 7 parametry.

Odhad  $\mu_i = \exp(\beta_0) = 3,443$  určuje střední počet dopravních nehod v některý pracovní den v kombinaci s měsícem červenec nebo únor a v místě na křižovatce. Odhad  $\mu_i = \exp(\beta_0 + \beta_1) = 6,35$  určuje střední počet dopravních nehod mimo křižovatku v některý pracovní den v kombinaci s některým z měsíců červenec nebo únor. Odhad  $\mu_i = \exp(\beta_0 + \beta_2) = 1,963$  představuje střední počet dopravních nehod v sobotu, které se staly v červenci na křižovatce a střední počet dopravních nehod v sobotu, které se staly v únoru na křižovatce. Odhad  $\mu_i = \exp(\beta_0 + \beta_3) = 1,389$  představuje střední počet dopravních nehod v neděli, které se staly v červenci na křižovatce a střední počet dopravních nehod v neděli, které se staly v únoru na křižovatce. Odhad  $\mu_i = \exp(\beta_0 + \beta_4) = 5,032$  zastupuje střední počet dopravních nehod kombinace některého pracovního dne a některého měsíce leden, listopad, prosinec nebo říjen, které se staly na křižovatce. Odhad  $\mu_i = \exp(\beta_0 + \beta_5) = 4,214$  zastupuje střední počet dopravních nehod kombinace některého pracovního dne a některého měsíce březen, červen, duben, květen nebo srpen, které se staly na křižovatce. Odhad  $\mu_i = \exp(\beta_0 + \beta_6) = 2,472$  zastupuje střední počet dopravních nehod kombinace některého pracovního dne a měsíce září, které se staly na křižovatce.

Vidíme, že oba modely jsou ekvivalentní a dávají stejné počty, nebo-li počty nehod se doplňují. Opět si tedy musíme dávat pozor na interpretaci odhadu  $\mu_i$ .

I v tomto případě jsme pro ověření faktory různým způsobem kombinovali, což je uvedeno v příloze CD, soubor R, denmesicmisto-vyber.R. Vybrali jsme rovněž modely bez interakcí se sedmi parametry.

#### **7.1.4 Dny v týdnu, měsíc, místo a druh dopravní nehody**

Tabulku pro druh dopravní nehody jsme zjednodušili na tvar uvedený v tabulce 7.1.4.1. Položku ostatní jsme vytvořili, protože faktory jiný druh nehody, srážku s domácím zvířetem, srážku s chodcem, srážku s vlakem a srážku s vozidlem zaparkovaným, odstaveným zastupovaly malý počet dopravních nehod.

TABULKA 7.1.4.1 Počet nehod v závislosti na druhu dopravní nehody

havárie	srážka s jedoucím nekolejovým vozidlem	srážka s lesní zvěří	srážka s pevnou překážkou	ostatní
47	642	52	59	59

Pro faktor druh nehody jsme vytvořili pět modelů uvedených v příloze CD, soubor R, druhneh.R a dle nejmenší hodnoty AIC = 37,5 jsme vybrali model se dvěma parametry ve tvaru

$$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]}, \quad i = 1, \dots, 5,$$

který využijeme pro model počtu dopravních nehod v závislosti na dni v týdnu, měsíci, místě a druhu nehody. Budeme pracovat s modely, které jsme vytvořili v kapitole 7.1.3. K těmto modelům přidáme faktor druh nehody.

Nejsložitější model s interakcemi má 840 parametrů (7 dnů, 12 měsíců, 2 možnosti místa a 5 možnosti druhu nehody), my ale rovnou přejdeme na model s interakcemi o 48 parametrech, jehož tvar i tvar ostatních modelů, které zjednodušují model o 48 parametrech, naleznete v příloze CD, soubor R, denmesicmistedruhneh.R. Ze všech modelů jsme opět získali modely, které využívají předchozích modelů.

Pro přehlednost tedy použijeme druhý model z kapitoly 7.1.3 a přidáme k němu do interakce parametr týkající se faktoru srážka s jedoucím nekolejovým vozidlem. Model o 48 parametrech tedy zjednodušíme na model o 14 parametrech, ve kterém se vyskytují pouze jednoduché interakce, a jehož Akaikeho informační kritérium nabývá hodnoty 1607,2. Model má nyní tvar

$$\begin{aligned} \ln \mu_i = & \beta_0 + \beta_1 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]} + \beta_2 I_{[\text{misto}=\text{mimo křižovatku}]} \\ & + \beta_3 I_{[\text{den}=\text{SO}]} + \beta_4 I_{[\text{den}=\text{NE}]} + \beta_5 I_{[\text{měsíc}=\text{březen}|\text{červen}|\text{duben}|\text{květen}|\text{srpen}]} \\ & + \beta_6 I_{[\text{měsíc}=\text{leden}|\text{listopad}|\text{prosinec}|\text{říjen}]} + \beta_7 I_{[\text{měsíc}=\text{září}]} \\ & + \beta_8 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}, \text{misto}=\text{mimo křižovatku}]} \\ & + \beta_9 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}, \text{den}=\text{SO}]} \\ & + \beta_{10} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}, \text{den}=\text{NE}]} \\ & + \beta_{11} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}, \text{měsíc}=\text{březen}|\text{červen}|\text{duben}|\text{květen}|\text{srpen}]} \\ & + \beta_{12} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}, \text{měsíc}=\text{leden}|\text{listopad}|\text{prosinec}|\text{říjen}]} \\ & + \beta_{13} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}, \text{měsíc}=\text{září}],} \quad i = 1, 2, \dots, 840, \end{aligned}$$

Parametry  $\beta_{11}$ ,  $\beta_{12}$ ,  $\beta_{13}$  jsou nevýznamné, proto je můžeme z modelu odstranit, čímž získáme model o 11 parametrech, ve kterém se vyskytují tři interakce. Hodnota

Akaikeho informačního kritéria v tomto modelu je  $AIC = 1603,5$ . Jak jsme již zmínili, v modelu se stále vyskytují interakce, a jelikož poslední interakce není příliš významná, tak ji odstraníme. Touto úpravou získáme sice horší model ( $AIC = 1606,4$ ), ale zbavili jsme se další interakce, které sledujeme v modelu za nežádoucí.

Nejvhodnější model s interakcemi o 10 parametrech má tvar

$$\ln \mu_i = \beta_0 + \beta_1 I_{[druhneh=srážka\ s\ jedoucím\ nekolejovým\ vozidlem]} + \beta_2 I_{[misto=mimo\ křižovatku]} \\ + \beta_3 I_{[den=SO]} + \beta_4 I_{[den=NE]} + \beta_5 I_{[měsíc=březen|červen|duben|květen|srpen]} \\ + \beta_6 I_{[měsíc=leden|listopad|prosinec|říjen]} + \beta_7 I_{[měsíc=září]} \\ + \beta_8 I_{[druhneh=srážka\ s\ jedoucím\ nekolejovým\ vozidlem, misto=mimo\ křižovatku]} \\ + \beta_9 I_{[druhneh=srážka\ s\ jedoucím\ nekolejovým\ vozidlem, den=SO]}, \quad i = 1, 2, \dots, 840,$$

jehož parametr  $\beta$  jsme odhadli pomocí softwaru R hodnoty. Tyto odhady jsou zobrazeny v tabulce 7.1.4.3 spolu s jejich dalšími vlastnostmi. Odhady jsou uvedeny za všech osm let dohromady v daném měsíci a v daném dni. V posledním jsou odhady  $\mu_i$  opět přepočteny na jeden rok, jeden den v týdnu, měsíc. Odhad  $\mu_i$  jsme dělili číslem 8.

TABULKA 7.1.4.3 Parametry  $\beta_j$ ,  $j = 0, 1, \dots, 9$

parametr	odhad	standartní odchylka	p-value	odhad	přepočet odhadu
$\beta_0$	-2,206	0,181	< 0,0001	0,11	0,014
$\beta_1$	3,316	0,171	< 0,0001	3,033	0,379
$\beta_2$	1,457	0,173	< 0,0001	0,473	0,059
<b><math>\beta_3</math></b>	<b>-0,068</b>	<b>0,192</b>	<b>0,724</b>	<b>0,103</b>	<b>0,013</b>
$\beta_4$	-0,908	0,137	< 0,0001	0,044	0,006
$\beta_5$	0,38	0,107	< 0,0001	0,161	0,020
<b><math>\beta_6</math></b>	<b>0,202</b>	<b>0,107</b>	<b>0,058</b>	<b>0,135</b>	<b>0,017</b>
<b><math>\beta_7</math></b>	<b>-0,331</b>	<b>0,18</b>	<b>0,065</b>	<b>0,079</b>	<b>0,01</b>
$\beta_8$	-1,079	0,191	< 0,0001	0,037	0,005
$\beta_9$	-0,717	0,242	0,003	0,054	0,007

Pokud bychom se rozhodli model ještě zjednodušit, odstraníme parametr  $\beta_7$ , který spadá pod faktor měsíc, jenž se nevyskytuje v interakcích. Po vypuštění tohoto parametru získáme model s 9 parametry, jehož Akaikeho informační kritérium se nepatrně zvýší na hodnotu 1607,9. Pokud bychom chtěli, mohli bychom tento model považovat za nejlepší, jelikož má nejméně parametrů a jednu z nejmenších hodnot AIC. My ale vybíráme model s nejmenší hodnotou AIC, proto zůstaneme u modelu s 10 parametry. V předchozích kapitolách jsme interpretovali odhady  $\mu_i$  a nyní budeme interpretovat pouze parametry  $\beta_j$ .



V tabulce můžeme vidět, že některé hodnoty odhadu parametru  $\beta_j$  jsou kladné a jiné záporné. Pro parametry  $\beta_1, \beta_2, \beta_5$  a  $\beta_6$  platí, že odhad  $\mu_i$  počtu dopravních nehod související s těmito parametry má větší hodnotu než odhad  $\mu_i$  počtu dopravních nehod související s parametrem  $\beta_0$ . Pro parametry  $\beta_3, \beta_4, \beta_7$  naopak platí, že odhad  $\mu_i$  počtu dopravních nehod související s těmito parametry má menší hodnotu než odhad  $\mu_i$  počtu dopravních nehod související s parametrem  $\beta_0$ .

Nyní přejdeme k parametru  $\beta_8$ , který se interpretuje dvěma způsoby. Oběma způsoby chceme dojít k závěru, co nám říká číslo  $\exp(\beta_8) \doteq 3$ . Parametr  $\beta_8$  zastupuje interakci mezi faktorem srážka s jedoucím nekolejovým vozidlem a místem mimo křižovatku. Následující tabulka uvádí přehled kombinací faktoru místo a druh nehody.

TABULKA 7.1.4.4 Kombinace faktoru místo a druh nehody

druh nehody	místo	$\mu_i$
jiný druh	na křižovatce	$\exp(\beta_0)$
srážka s jedoucím nekolejovým vozidlem	na křižovatce	$\exp(\beta_0 + \beta_1)$
jiný druh	mimo křižovatku	$\exp(\beta_0 + \beta_2)$
srážka s jedoucím nekolejovým vozidlem	mimo křižovatku	$\exp(\beta_0 + \beta_1 + \beta_2 + \beta_8)$

### ZPŮSOB 1

Nejprve dáme do poměru druhý a první řádek, tedy

$$\frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1) = \exp(3,316) = 27,54993 \doteq 28,$$

což slovně interpretujeme tak, že na 28 nehod při srážce s jedoucím nekolejovým vozidlem na křižovatce připadá jedna nehoda jiného druhu na křižovatce.

Dále dáme do poměru čtvrtý a třetí řádek, tedy

$$\frac{\exp(\beta_0 + \beta_1 + \beta_2 + \beta_8)}{\exp(\beta_0 + \beta_2)} = \exp(\beta_1 + \beta_8) = \exp(3,316 - 1,079) = 9,365194 \doteq 9,$$

což slovně interpretujeme tak, že na 9 nehod při srážce s jedoucím nekolejovým vozidlem mimo křižovatku připadá jedna nehoda jiného druhu mimo křižovatku.

Při tomto postupu nám číslo  $\exp(-\beta_8) \doteq 3$  říká, že hodnota 9 je zhruba 3 krát menší než hodnota 28. Tedy, že nehod při srážce s jedoucím nekolejovým vozidlem mimo křižovatku je zhruba třikrát méně než nehod při srážce s jedoucím nekolejovým vozidlem na křižovatce.

## ZPŮSOB 2

Nejprve dáme do poměru třetí a první řádek, tedy

$$\frac{\exp(\beta_0 + \beta_2)}{\exp(\beta_0)} = \exp(\beta_2) = \exp(1,457) = 4,293061 \doteq 4,$$

což slovně interpretujeme tak, že na jednu nehodu jiného druhu na křižovatce, připadají 4 nehody jiného druhu mimo křižovatku.

Dále dáme do poměru čtvrtý a druhý řádek, tedy

$$\frac{\exp(\beta_0 + \beta_1 + \beta_2 + \beta_8)}{\exp(\beta_0 + \beta_1)} = \exp(\beta_2 + \beta_8) = \exp(1,457 - 1,079) = 1,459363 \doteq 1,5,$$

což slovně interpretujeme tak, že na jednu nehodu, která se stala při srážce s jedoucím nekolejovým vozidlem na křižovatce, připadá 1,5 nehody při srážce s jedoucím nekolejovým vozidlem mimo křižovatku.

Při tomto postupu nám číslo  $\exp(-\beta_8) \doteq 3$  říká, že hodnota 1,5 je zhruba 3 krát menší, než hodnota 4. Tedy, že nehod, které se staly při srážce s jedoucím nekolejovým vozidlem mimo křižovatku, je zhruba třikrát méně než nehod jiného druhu mimo křižovatku.

Poslední parametr v tomto modelu je parametr  $\beta_9$ , který Opět zde uvedeme dva způsoby interpretace. Parametr  $\beta_9$  zastupuje interakci mezi faktorem srážka s jedoucím nekolejovým vozidlem a sobotou. Následující tabulka uvádí přehled kombinací faktoru den a druh nehody.

TABULKA 7.1.4.5 Kombinace faktoru místo a druh nehody

druh nehody	den	$\mu_i$
jiný druh	pracovní den	$\exp(\beta_0)$
srážka s jedoucím nekolejovým vozidlem	pracovní den	$\exp(\beta_0 + \beta_1)$
jiný druh	sobota	$\exp(\beta_0 + \beta_3)$
srážka s jedoucím nekolejovým vozidlem	sobota	$\exp(\beta_0 + \beta_1 + \beta_3 + \beta_9)$

## ZPŮSOB 1

Nejprve dáme do poměru druhý a první řádek stejně jako v předchozím způsobu, čímž nám vyjdou i stejné hodnoty, tedy

$$\frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1) = \exp(3,316) = 27,54993 \doteq 28.$$

Interpretace je taková, že na 28 nehod při srážce s jedoucím nekolejovým vozidlem v pracovní den připadá jedna nehoda jiného druhu v pracovní den.

Dále dáme do poměru opět čtvrtý a třetí řádek, tedy

$$\frac{\exp(\beta_0 + \beta_1 + \beta_3 + \beta_9)}{\exp(\beta_0 + \beta_3)} = \exp(\beta_1 + \beta_9) = \exp(3,316 - 0,717) = 13,45028 \doteq 13,5,$$

což slovně interpretujeme tak, že na 13,5 nehod při srážce s jedoucím nekolejovým vozidlem v sobotu připadá jedna nehoda jiného druhu v sobotu.

Při tomto postupu nám číslo  $\exp(-\beta_9) \doteq 2$  říká, že hodnota 13,5 je zhruba 2 krát menší než hodnota 28. Tedy, že nehod při srážce s jedoucím nekolejovým vozidlem v sobotu je zhruba dvakrát méně než nehod při srážce s jedoucím nekolejovým vozidlem v pracovní den.

## ZPŮSOB 2

Nejprve dáme do poměru třetí a první řádek, tedy

$$\frac{\exp(\beta_0 + \beta_3)}{\exp(\beta_0)} = \exp(\beta_3) = \exp(-0,068) = 0,9342605 \doteq 0,9,$$

což slovně interpretujeme tak, že na jednu nehodu jiného druhu v pracovní den, připadá 0,9 nehody jiného druhu v sobotu.

Dále dáme do poměru čtvrtý a druhý řádek, tedy

$$\frac{\exp(\beta_0 + \beta_1 + \beta_3 + \beta_9)}{\exp(\beta_0 + \beta_1)} = \exp(\beta_3 + \beta_9) = \exp(-0,068 - 1,079) = 0,4561197 \doteq 0,5,$$

což slovně interpretujeme tak, že na jednu nehodu, která se stala při srážce s jedoucím nekolejovým vozidlem v pracovní den, připadá 0,5 nehody při srážce s jedoucím nekolejovým vozidlem v sobotu.

Při tomto postupu nám číslo  $\exp(-\beta_9) \doteq 2$  říká, že hodnota 0,5 je zhruba 2 krát menší, než hodnota 0,9. Tedy, že nehod, které se staly při srážce s jedoucím nekolejovým vozidlem v sobotu, je zhruba 2 krát méně než nehod jiného druhu v sobotu.

### 7.1.5 Dny v týdnu, měsíc, místo, druh a zavinění dopravní nehody

Faktor zavinění dopravní nehody jsme opět upravili tak, aby tam nebyla odlehlá data. Faktor jiné zavinění v sobě zahrnuje i zavinění jiným účastníkem silničního provozu a zavinění závadnou komunikací.

TABULKA 7.1.5.1 Počet nehod v závislosti na zavinění nehody

chodcem	jiné zavinění	lesní zvěř, domácím zvířectvem	řidičem motorového vozidla	řidičem nemotorového vozidla	technickou závadou vozidla
14	16	53	753	13	10

Pro uvedená data jsme vybrali model se třemi parametry s hodnotou AIC = 39,409 ve tvaru

$$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{zavinění}=\text{lesní zvěř, domácím zvířectvem}]} + \beta_2 I_{[\text{zavinění}=\text{řidičem motorového vozidla}], \quad i = 1, \dots, 6,$$

který využijeme pro vytvoření modelu počtu dopravních nehod v závislosti na dni v týdnu, měsíci, místě, druhu nehody a zavinění dopravní nehody. Budeme pracovat s modely, které jsme vytvořili v kapitole 7.1.3. K těmto modelům přidáme faktor druh nehody.

Nejsložitější model s interakcemi má 5040 parametrů, což je mnohem více než máme k dispozici dat a s takovým model nemá smysl pracovat. Přecházíme rovnou k modelu s interakcemi o 144 parametrech, ve kterém se vyskytují nulové buňky vstupní tabulky. Využijeme tedy model z předchozí kapitoly ve tvaru

$$\begin{aligned} \ln \mu_i = & \beta_0 + \beta_1 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]} + \beta_2 I_{[\text{misto}=\text{mimo křižovatku}]} \\ & + \beta_3 I_{[\text{den}=\text{SO}]} + \beta_4 I_{[\text{den}=\text{NE}]} + \beta_5 I_{[\text{měsíc}=\text{leden}|\text{listopad}|\text{prosinec}|\text{říjen}]} \\ & + \beta_6 I_{[\text{měsíc}=\text{březen}|\text{červen}|\text{duben}|\text{květen}|\text{srpen}]} + \beta_7 I_{[\text{měsíc}=\text{září}]} \\ & + \beta_8 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, misto}=\text{mimo křižovatku}]} \\ & + \beta_9 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, den}=\text{SO}]} \quad i = 1, 2, \dots, 840, \end{aligned}$$

ke kterému přidáme do interakce faktor zavinění dopravní nehody. Tento model naleznete v příloze CD, soubor R, denmesicmisdruhnehzavineni1.R, ve kterém se vyskytují dvojné i jednoduché interakce. Je vidět, že všechny dvojné interakce jsou nevýznamné, proto je odstraníme a získáme model s jednoduchými interakcemi o 26 parametrech, jehož Akaikeho informační kritérium má hodnotu 2229,4. V modelu stále zbývají některé jednoduché interakce nevýznamné, proto je budeme postupně odstraňovat. Po odstranění sedmi jednoduchých interakcí získáme model s jednoduchými interakcemi o 19

parametrech, který má nejmenší hodnotu Akaikeho informačního kritéria,  $AIC = 2219,4$ . V modelu ovšem stále zůstávají nevýznamné parametry, proto budeme dále pokračovat v jejich odstraňování. Po odstranění všech nevýznamných parametrů získáme model se 14 parametry, jehož hodnota AIC vzroste na 2240,4. Nejlepší model bude mít tvar

$$\begin{aligned} \ln \mu_i = & \beta_0 + \beta_1 I_{[zavinění=lesní\ zvěř,domácím\ zvířectvem]} + \beta_2 I_{[misto=mimo\ křižovatku]} \\ & + \beta_3 I_{[zavinění=řidičem\ motorového\ vozidla]} \\ & + \beta_4 I_{[druhneh=srážka\ s\ jedoucím\ nekolejovým\ vozidlem]} + \beta_5 I_{[den=SO]} + \beta_6 I_{[den=NE]} \\ & + \beta_7 I_{[měsíc=leden|listopad|prosinec|říjen]} + \beta_8 I_{[měsíc=září]} \\ & + \beta_9 I_{[měsíc=březen|červen|duben|květen|srpen]} \\ & + \beta_{10} I_{[zavinění=lesní\ zvěř,domácím\ zvířectvem,misto=mimo\ křižovatku]} \\ & + \beta_{11} I_{[zavinění=řidičem\ motorového\ vozidla,druhneh=srážka\ s\ jedoucím\ nekolejovým\ vozidlem]} \\ & + \beta_{12} I_{[misto=mimo\ křižovatku,druhneh=srážka\ s\ jedoucím\ nekolejovým\ vozidlem]} \\ & + \beta_{13} I_{[druhneh=srážka\ s\ jedoucím\ nekolejovým\ vozidlem,den=SO]}, \quad i = 1, 2, \dots, 5040, \end{aligned}$$

jehož parametr  $\beta$  jsme odhadli pomocí softwaru R hodnoty. Tyto odhady jsou zobrazeny v tabulce 7.1.5.2 spolu s jejich dalšími vlastnostmi. Odhady jsou uvedeny za všech osm let dohromady v daném měsíci a v daném dni. V posledním jsou odhady  $\mu_i$  opět přepočteny na jeden rok, jeden den v týdnu, měsíc. Odhad  $\mu_i$  jsme dělili číslem 8.

TABULKA 7.1.5.2 Parametry  $\beta_j$ ,  $j = 0, 1, \dots, 14$

parametr	odhad	standartní odchylka	p-value	odhad	přepočet odhadu
$\beta_0$	-4,887	0,219	< 0,0001	0,008	0,0009
<b><math>\beta_1</math></b>	<b>-0,505</b>	<b>0,73</b>	<b>0,489</b>	<b>0,005</b>	<b>0,0006</b>
$\beta_2$	1,187	0,181	< 0,0001	0,025	0,0031
$\beta_3$	2,301	0,168	< 0,0001	0,075	0,0094
<b><math>\beta_4</math></b>	<b>-0,252</b>	<b>0,387</b>	<b>0,515</b>	<b>0,006</b>	<b>0,0007</b>
<b><math>\beta_5</math></b>	<b>-0,068</b>	<b>0,192</b>	<b>0,724</b>	<b>0,007</b>	<b>0,0009</b>
$\beta_6$	-0,908	0,137	< 0,0001	0,003	0,0004
$\beta_7$	0,38	0,107	< 0,0001	0,011	0,0014
<b><math>\beta_8</math></b>	<b>-0,331</b>	<b>0,18</b>	<b>0,065</b>	<b>0,005</b>	<b>0,0007</b>
<b><math>\beta_9</math></b>	<b>0,202</b>	<b>0,107</b>	<b>0,058</b>	<b>0,009</b>	<b>0,0012</b>
$\beta_{10}$	2,141	0,738	0,004	0,064	0,0080
$\beta_{11}$	3,937	0,363	< 0,0001	0,387	0,0484
$\beta_{12}$	-0,817	0,198	< 0,0001	0,003	0,0004
$\beta_{13}$	-0,717	0,242	0,003	0,004	0,0005

Nyní přejdeme k interpretaci jednoduchých interakcí. Nejprve se budeme zabývat parametrem  $\beta_{10}$  a opět zde uvedeme obě jeho interpretace. Parametr  $\beta_8$  zastupuje interakci mezi faktorem lesní zvěř, domácí zvířectvo a místem na křižovatce. Následující tabulka uvádí přehled kombinací faktoru místo a zavinění.

TABULKA 7.1.5.3 Kombinace faktoru zavinění a místo

zavinění	místo	$\mu_i$
jiné zavinění	na křižovatce	$\exp(\beta_0)$
lesní zvěř, domácím zvířectvem	na křižovatce	$\exp(\beta_0 + \beta_1)$
jiné zavinění	mimo křižovatku	$\exp(\beta_0 + \beta_2)$
lesní zvěř, domácím zvířectvem	mimo křižovatku	$\exp(\beta_0 + \beta_1 + \beta_2 + \beta_{10})$

#### ZPŮSOB 1

Nejprve dáme do poměru druhý a první řádek, tedy

$$\frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1) = \exp(-0,505) = 0,6035056 \doteq 0,6,$$

což slovně interpretujeme tak, že na 0,6 nehody zaviněné lesním zvířetem, nebo domácím zvířectvem na křižovatce připadá jedna nehoda jinak zaviněná (krom řidiče motorového vozidla) na křižovatce.

Dále dáme do poměru čtvrtý a třetí řádek, tedy

$$\frac{\exp(\beta_0 + \beta_1 + \beta_2 + \beta_{10})}{\exp(\beta_0 + \beta_2)} = \exp(\beta_1 + \beta_{10}) = \exp(-0,505 + 2,141) = 5,13459 \doteq 5,$$

což slovně interpretujeme tak, že na 5 nehod, které vznikly mimo křižovatku, když nehodu zavinilo lesní zvíře, nebo domácí zvířectvo, připadá jedna nehoda jinak zaviněná (krom řidiče motorového vozidla) mimo křižovatku.

Při tomto postupu nám číslo  $\exp(\beta_{10}) \doteq 8,5$  říká, že hodnota 5 je zhruba 8,5 krát větší než hodnota 0,6. Tedy, že nehod, které vznikly mimo křižovatku, když nehodu zavinilo lesní zvíře, nebo domácí zvířectvo je zhruba 8,5 krát více než nehod, které zavinilo lesní zvíře, nebo domácí zvířectvo na křižovatce.

#### ZPŮSOB 2

Nejprve dáme do poměru třetí a první řádek, tedy

$$\frac{\exp(\beta_0 + \beta_2)}{\exp(\beta_0)} = \exp(\beta_2) = \exp(1,187) = 3,277235 \doteq 3,$$

což slovně interpretujeme tak, že na 3 nehody, které byly jinak zaviněny (krom řidiče motorového vozidla) mimo křižovatku, připadá jedna nehoda, která byla jinak zaviněna (krom řidiče motorového vozidla) na křižovatce.

Dále dáme do poměru čtvrtý a druhý řádek, tedy

$$\frac{\exp(\beta_0 + \beta_1 + \beta_2 + \beta_{10})}{\exp(\beta_0 + \beta_1)} = \exp(\beta_2 + \beta_{10}) = \exp(1,187 + 2,141) = 27,88252 \doteq 28,$$

což slovně interpretujeme tak, že na jednu nehodu, kterou zavinilo lesní zvíře nebo domácí zvířectvo na křižovatce, připadá 28 nehod, které zavinilo lesní zvíře nebo domácí zvířectvo mimo křižovatku.

Při tomto postupu nám číslo  $\exp(\beta_{10}) \doteq 8,5$  říká, že hodnota 28 je zhruba 8,5 krát větší než hodnota 3. Tedy, že nehod, které zavinilo lesní zvíře nebo domácí zvířectvo mimo křižovatku, je zhruba třikrát více než nehod, které byly jinak zaviněny mimo křižovatku.

Interpretaci dalších parametrů už nebudeme rozebírat, napíšeme pouze výsledky obou způsobů interpretací jednotlivých parametrů do následující tabulky.

TABULKA 7.1.5.4 Vysvětlení interakcí

param.	Faktory interakce	vysvětlení
$\beta_{11}$	řidič motor. vozidla a srážka s jedoucím nekolej. vozidlem	Číslo $\exp(\beta_{11}) \doteq 51$ říká, že hodnota 512 je zhruba 51 krát větší než hodnota 10. Tedy, že nehod, které zavinil řidič motorového vozidla při srážce s jedoucím nekolejovým vozidlem, je zhruba 51 krát více než nehod, které zavinil řidič motorového vozidla při jiném druhu nehody.
		Číslo $\exp(\beta_{11}) \doteq 51$ říká, že hodnota 40 je zhruba 51 krát větší než hodnota 0,8. Tedy, že nehod, které zavinil řidič motorového vozidla při srážce s jedoucím nekolejovým vozidlem, je zhruba 51 krát více než nehod, které vznikly při srážce s jedoucím nekolejovým vozidlem při jiném zavinění (krom řidiče motorového vozidla).
$\beta_{12}$	srážka s jedoucím nekolej. vozidlem a mimo křižovatku	Číslo $\exp(-\beta_{12}) \doteq 2,3$ říká, že hodnota 0,34 je zhruba 2,3 krát menší než hodnota 0,78. Tedy, že nehod při srážce s jedoucím nekolejovým vozidlem mimo křižovatku je zhruba 2,3 krát méně než nehod při srážce s jedoucím nekolejovým vozidlem na křižovatce.
		Číslo $\exp(-\beta_{12}) \doteq 2,3$ říká, že hodnota 1,4 je zhruba 2,3 krát menší, než hodnota 3,3. Tedy, že nehod jiného druhu, které se staly mimo křižovatku, je zhruba 2,3 krát méně než nehod, které vznikly mimo křižovatku při srážce s jedoucím nekolejovým vozidlem

$\beta_{13}$	srážka s jedoucím nekolej. vozidlem a sobota	Číslo $\exp(-\beta_{13}) \doteq 2$ říká, že hodnota 0,4 je zhruba 2 krát menší než hodnota 0,8. Tedy, že nehod při srážce s jedoucím nekolejovým vozidlem v sobotu je zhruba dvakrát méně než nehod při srážce s jedoucím nekolejovým vozidlem v pracovní den.
		Číslo $\exp(-\beta_{13}) \doteq 2$ říká, že hodnota 0,46 je zhruba 2 krát menší, než hodnota 0,93. Tedy, že nehod, které se staly při srážce s jedoucím nekolejovým vozidlem v sobotu, je zhruba dvakrát méně než nehod jiného druhu v sobotu.

### 7.1.6 Dny v týdnu, měsíc, místo, druh nehody, zavinění dopravní nehody a druh vozidla, které zavinilo nehodu

Poslední faktor, který přidáme do našeho modelu, je druh vozidla, které zavinilo dopravní nehodu, který měl původně 12 tříd. Tento počet jsme snížili na 5 tříd, což uvádí tabulka 7.1.6.1.

TABULKA 7.1.6.1 Počet nehod v závislosti na druhu vozidla, které zavinilo nehodu

hromadná doprava	nákladní automobil	nezjištěno, řidič ujel	osobní automobil	prostředek pro 1 osobu
21	167	35	610	26

Pro daná data jsme vybrali model se třemi parametry a hodnotou  $AIC = 40,2$  ve tvaru

$\ln \mu_i = \beta_0 + \beta_1 I_{[druhvozza=osobní\ automobil]} + \beta_2 I_{[druhvozza=nákladní\ automobil]}, i = 1, \dots, 5$ , ačkoliv model se čtyřmi parametry měl nižší hodnotu AIC, ale čtvrtý parametr byl nevýznamný. Udělali jsme v tomto případě výjimku, jelikož potřebujeme co nejjednodušší model, aby model se všemi faktory nebyl příliš složitý.

Přecházíme k modelu, ve kterém se vyskytují všechny faktory. Nejsložitější model obsahuje 25 200 parametrů, což je obrovské množství a nemá smysl s ním pracovat.

Využijeme nejlepší model z kapitoly 7.1.5, ve kterém se vyskytli faktory dny v týdnu, měsíc, místo, druh nehody, zavinění dopravní nehody ve tvaru



$$\begin{aligned}
\ln \mu_i = & \beta_0 + \beta_1 I_{[\text{zavinění}=\text{lesní zvíři, domácí zvířectvem}]} + \beta_2 I_{[\text{misto}=\text{mimo křižovatku}]} \\
& + \beta_3 I_{[\text{zavinění}=\text{řidičem motorového vozidla}]} \\
& + \beta_4 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]} + \beta_5 I_{[\text{den}=\text{SO}]} + \beta_6 I_{[\text{den}=\text{NE}]} \\
& + \beta_7 I_{[\text{měsíc}=\text{leden}|listopad|prosinec|říjen}]} + \beta_8 I_{[\text{měsíc}=\text{září}]} \\
& + \beta_9 I_{[\text{měsíc}=\text{březen}|červen|duben|květen|srpen]} \\
& + \beta_{10} I_{[\text{zavinění}=\text{lesní zvíři, domácí zvířectvem, misto}=\text{mimo křižovatku}]} \\
& + \beta_{11} I_{[\text{zavinění}=\text{řidičem motorového vozidla, druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]} \\
& + \beta_{12} I_{[\text{misto}=\text{mimo křižovatku, druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]} \\
& + \beta_{13} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, den}=\text{SO}]}, \quad i = 1, 2, \dots, 5040,
\end{aligned}$$

ke kterému přidáme faktor druh vozidla, které zavinilo nehodu.

Model se 42 parametry obsahuje dvojné i jednoduché interakce, přičemž většina z nich je nevýznamných. Hodnota Akaikeho informačního kritéria je 3268,5. Začneme tedy postupně odstraňovat nevýznamné interakce, což už nebude tak jednoduché, jelikož některé dvojné interakce jsou významné, tak je musíme odstraňovat postupně, čímž se nakonec podaří odstranit všechny dvojné interakce. Model s nejnižší hodnotou Akaikeho informačního kritéria,  $AIC = 3255,5$ , má 24 parametrů a obsahuje ještě několik jednoduchých interakcí. Mezi těmito interakcemi jsou i parametry které nejsou velmi významné, proto je z modelu odstraníme. Získáme model s 21 parametry, ve kterém se vyskytuje 9 jednoduchých interakcí, a jehož Akaikeho informační kritérium nijak razantně nestoupne,  $AIC = 3265,8$ . Model má tvar

$$\begin{aligned}
\ln \mu_i = & \beta_0 + \beta_1 I_{[\text{druhvozza}=\text{nákladní automobil}]} \\
& + \beta_2 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]} + \beta_3 I_{[\text{den}=\text{NE}]} \\
& + \beta_4 I_{[\text{druhvozza}=\text{osobní automobil}]} + \beta_5 I_{[\text{zavinění}=\text{lesní zvěř, domácí zvířectvem}]} \\
& + \beta_6 I_{[\text{zavinění}=\text{řidičem motorového vozidla}]} + \beta_7 I_{[\text{misto}=\text{mimo křižovatku}]} + \beta_8 I_{[\text{den}=\text{SO}]} \\
& + \beta_9 I_{[\text{měsíc}=\text{březen}|červen|duben|květen|srpen]} + \beta_{10} I_{[\text{měsíc}=\text{září}]} \\
& + \beta_{11} I_{[\text{měsíc}=\text{leden}|listopad|prosinec|říjen]} \\
& + \beta_{12} I_{[\text{druhvozza}=\text{nákladní automobil, druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]} \\
& + \beta_{13} I_{[\text{druhvozza}=\text{nákladní automobil, den}=\text{NE}]} \\
& + \beta_{14} I_{[\text{druhvozza}=\text{osobní automobil, zavinění}=\text{lesní zvěř, domácí zvířectvem}]} \\
& + \beta_{15} I_{[\text{druhvozza}=\text{osobní automobil, zavinění}=\text{řidičem motorového vozidla}]} \\
& + \beta_{16} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, druhvozza}=\text{osobní automobil}]} \\
& + \beta_{17} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, misto}=\text{mimo křižovatku}]} \\
& + \beta_{18} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, den}=\text{SO}]} \\
& + \beta_{19} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, zavinění}=\text{řidičem motorového vozidla}]} \\
& + \beta_{20} I_{[\text{zavinění}=\text{lesní zvěř, domácí zvířectvem, misto}=\text{mimo křižovatku}]}, \quad i = 1, 2, \dots, 25200,
\end{aligned}$$

jehož parametr  $\beta$  jsme odhadli pomocí softwaru R hodnoty. Tyto odhady jsou zobrazeny v tabulce 7.1.6.2 spolu s jejich dalšími vlastnostmi. Odhady jsou uvedeny za všech osm let dohromady v daném měsíci a v daném dni. V posledním jsou odhady  $\mu_i$  opět přepočteny na jeden rok, jeden den v týdnu, měsíc. Odhad  $\mu_i$  jsme dělili číslem 8.

TABULKA 7.1.6.2 Parametry  $\beta_j$ ,  $j = 0, 1, \dots, 20$

parametr	odhad	standartní odchylka	p-value	odhad	přepočet odhadu
$\beta_0$	-6,83	0,254	< 0,0001	0,001	0,0001
$\beta_1$	0,78	0,241	0,001	0,002	0,0003
$\beta_2$	-1,179	0,436	0,007	0	0
$\beta_3$	-0,691	0,139	< 0,0001	0,001	0,0001
<b><math>\beta_4</math></b>	<b>0,602</b>	<b>0,327</b>	<b>0,065</b>	<b>0,002</b>	<b>0,0002</b>
$\beta_5$	-2,085	0,804	0,009	0	0
$\beta_6$	1,213	0,236	< 0,0001	0,004	0,0005
$\beta_7$	1,186	0,181	< 0,0001	0,004	0,0004
<b><math>\beta_8</math></b>	<b>-0,062</b>	<b>0,192</b>	<b>0,747</b>	<b>0,001</b>	<b>0,0001</b>
<b><math>\beta_9</math></b>	<b>0,202</b>	<b>0,106</b>	<b>0,058</b>	<b>0,001</b>	<b>0,0002</b>
<b><math>\beta_{10}</math></b>	<b>-0,331</b>	<b>0,18</b>	<b>0,065</b>	<b>0,001</b>	<b>0,0001</b>
$\beta_{11}$	0,379	0,107	< 0,0001	0,002	0,0002
$\beta_{12}$	1,664	0,301	< 0,0001	0,006	0,0007

$\beta_{13}$	-2,71	1,013	0,007	0	0
$\beta_{14}$	2,749	0,494	< 0,0001	0,017	0,0021
$\beta_{15}$	2,131	0,362	< 0,0001	0,009	0,0011
$\beta_{16}$	0,86	0,278	0,002	0,003	0,0003
$\beta_{17}$	-0,816	0,198	< 0,0001	0,000	0,0001
$\beta_{18}$	-0,724	0,242	0,003	0,001	0,0001
$\beta_{19}$	3,968	0,367	< 0,0001	0,057	0,0071
$\beta_{20}$	2,138	0,738	0,004	0,009	0,0011

V následující tabulce bude uvedena interpretace jednoduchých interakcí, které zůstaly v modelu. Budeme zde interpretovat pouze interakce, které se nevyskytly v kapitole 7.1.5. Rovněž se odkazujeme na tuto kapitolu v souvislosti s hodnotami, které se v tabulce vyskytnou, že jsme k nim došli stejným způsobem jako v kapitole 7.1.5.

TABULKA 7.1.6.3 Vysvětlení interakcí

param.	Faktory interakce	vysvětlení
$\beta_{12}$	srážka s jedoucím nekolej. vozidlem a nákladní automobil	Číslo $\exp(\beta_{12}) \doteq 5,3$ říká, že hodnota 12,5 je zhruba 5,3 krát větší než hodnota 2,2. Tedy, že nehod, které zavinil nákladní automobil při srážce s jedoucím nekolejovým vozidlem, je zhruba 5,3 krát více než nehod jiného druhu, které zavinil nákladní automobil.
		Číslo $\exp(\beta_{12}) \doteq 5,3$ říká, že hodnota 1,6 je zhruba 5,3 krát větší než hodnota 0,3. Tedy, že nehod, které zavinil nákladní automobil při srážce s jedoucím nekolejovým vozidlem, je zhruba 5,3 krát více než nehod, které zavinilo jiné vozidlo (krom osobního automobilu) při srážce s jedoucím nekolejovým vozidlem.
$\beta_{13}$	Nákladní automobil a neděle	Číslo $\exp(-\beta_{13}) \doteq 15$ říká, že hodnota 0,15 je zhruba 15 krát menší než hodnota 2,2. Tedy, že nehod, které zavinil nákladní automobil v neděli, je zhruba 2,3 krát méně než nehod, které zavinil nákladní automobil v pracovní den.
		Číslo $\exp(-\beta_{13}) \doteq 15$ říká, že hodnota 0,03 je zhruba 15 krát menší než hodnota 0,5. Tedy, že nehod, které zavinil nákladní automobil v neděli, je zhruba 15 krát méně než nehod, které způsobilo jiné vozidlo (krom řidiče motorového vozidla) v neděli.
$\beta_{14}$	Osobní automobil a lesní zvěř nebo domácí zvířectvo	Číslo $\exp(\beta_{14}) \doteq 15,6$ říká, že hodnota 28,5 je zhruba 15,6 krát větší než hodnota 1,8. Tedy, že nehod, které zavinil osobní automobil a lesní zvěř nebo domácí zvířectvo, je zhruba 15,6 krát více než nehod, které způsobil osobní automobil a jiné zavinění (krom řidiče motorového vozidla).
		Číslo $\exp(\beta_{14}) \doteq 15,6$ říká, že hodnota 1,94 je zhruba 15,6 krát větší než hodnota 0,12. Tedy, že nehod, které zavinil osobní automobil a lesní zvěř nebo domácí zvířectvo, je zhruba 15,6 krát více než nehod, které způsobilo jiné vozidlo (krom nákladního automobilu) a lesní zvěř nebo domácí zvířectvo.

$\beta_{15}$	Osobní automobil a řidič motor. vozidla	Číslo $\exp(\beta_{15}) \doteq 8,4$ říká, že hodnota 15,4 je zhruba 8,4 krát větší než hodnota 1,8. Tedy, že nehod, které zavinil osobní automobil a řidič motorového vozidla, je zhruba 8,4 krát více než nehod, které způsobil osobní automobil a jiné zavinění (krom lesní zvěře nebo domácí zvířectva).
		Číslo $\exp(\beta_{15}) \doteq 8,4$ říká, že hodnota 28,3 je zhruba 8,4 krát větší, než hodnota 3,4. Tedy, že nehod, které zavinil osobní automobil a řidič motorového vozidla, je zhruba 8,4 krát více než nehod, které způsobil řidič motorového vozidla a jiné vozidlo (krom nákladního automobilu).
$\beta_{16}$	Osobní automobil a srážka s jedoucím nekolej. vozidlem	Číslo $\exp(\beta_{16}) \doteq 2,4$ říká, že hodnota 0,73 je zhruba 2,4 krát větší než hodnota 0,3. Tedy, že nehod, které zavinil osobní automobil při srážce s jedoucím nekolejovým vozidlem, je zhruba 2,4 krát více než nehod při srážce s jedoucím nekolejovým vozidlem, což bylo jinak zaviněno (krom nákladního automobilu).
		Číslo $\exp(\beta_{16}) \doteq 2,4$ říká, že hodnota 4,3 je zhruba 2,4 krát větší než hodnota 1,8. Tedy, že nehod, které zavinil osobní automobil při srážce s jedoucím nekolejovým vozidlem, je zhruba 2,4 krát více než nehod jiného druhu, které způsobil osobní automobil.

## 7.2 Faktory ovlivňující vznik dopravní nehody

V následujících kapitolách budeme vytvářet model, který obsahuje faktory, díky nimž vznikly dopravní nehody. Začneme opět jedním faktorem a skončíme čtyřmi faktory.

### 7.2.1 Povětrnostní podmínky

V této sekci začneme faktorem povětrnostní podmínky, který prezentuje tabulka 7.2.1.1. Tabulka je již v upraveném tvaru, jelikož v původních hodnotách byla odlehlá data.

TABULKA 7.2.1.1 Počet nehod v závislosti na povětrnostních podmínkách

děšť	na počátku deště, slabý déšť	neztížené	ostatní	sněžení
45	24	708	24	58

Kolonka ostatní v sobě skrývá faktor jiné ztížení, mlha, nárazový vítr a stav, kdy se tvoří námraza, náledí. Pro faktor povětrnostní podmínky jsme vytvořili 5 modelů, které jsou uvedeny v tabulce 7.2.1.2 spolu s deviancí, AIC a stupni volnosti.

TABULKA 7.2.1.2 Přehled modelů

	model	deviance	AIC	stupně volnosti
1.	$\ln \mu_i = \beta_0 + \beta_1 I_{[povpodm=na\ po\cacute{a}tku\ de\c{s}\t\acute{e},\ slab\acute{y}\ de\c{s}\t\acute{e}]} + \beta_2 I_{[povpodm=nezt\acute{z}\acute{e}n\acute{e}]} + \beta_3 I_{[povpodm=ostatn\acute{i}]} + \beta_4 I_{[povpodm=sn\acute{e}\c{s}\t\acute{z}en\acute{i}]}, i = 1, \dots, 5$	0	40	0
2.	$\ln \mu_i = \beta_0 + \beta_1 I_{[povpodm=nezt\acute{z}\acute{e}n\acute{e}]} + \beta_2 I_{[povpodm=d\acute{e}\c{s}\t\acute{e}]} + \beta_3 I_{[povpodm=sn\acute{e}\c{s}\t\acute{z}en\acute{i}]}, i = 1, \dots, 5$	0	38	1
3.	$\ln \mu_i = \beta_0 + \beta_1 I_{[povpodm=nezt\acute{z}\acute{e}n\acute{e}]} + \beta_2 I_{[povpodm=d\acute{e}\c{s}\t\acute{e} povpodm=sn\acute{e}\c{s}\t\acute{z}en\acute{i}]}, i = 1, \dots, 5$	1,6	37,6	2
4.	$\ln \mu_i = \beta_0 + \beta_1 I_{[povpodm=nezt\acute{z}\acute{e}n\acute{e}]}, i = 1, \dots, 5$	22,1	56,1	3
5.	$\ln \mu_i = \beta_0, i = 1, \dots, 5$	1569,7	1601,7	4

Nejvhodnější model pro povětrnostní podmínky je třetí model se třemi parametry, pro který je hodnota AIC rovna 37,6. Pomocí softwaru R jsme rovněž vypočítali hodnoty spojené s parametry  $\beta_j, j = 0,1,2$ , což prezentuje tabulka 7.2.1.3.

TABULKA 7.2.1.3 Parametry  $\beta_j, j = 0,1,2$

parametr	Odhad $\beta_j$	standartní odchylka	p-value	Odhad $\mu_i$
$\beta_0$	3,178	0,144	< 0,0001	24,001
$\beta_1$	3,384	0,149	< 0,0001	708,04
$\beta_2$	0,764	0,175	< 0,0001	51,501

V tabulce můžete vidět, že parametry  $\beta_1$  a  $\beta_2$  mají kladné znaménko, což znamená, že odhady počtu dopravních nehod  $\mu_i$  související s těmito parametry jsou větší hodnoty než odhady  $\mu_i$  související s parametrem  $\beta_0$ . Interpretace odhadů počtu dopravních nehod  $\mu_i$  je zřejmá.

Odhad  $\mu_i = \exp(\beta_0) = 24,001$  zastupuje střední počet dopravních nehod, které vznikly na počátku deště, slabého deště, či za jiných povětrnostních podmínek. Odhad  $\mu_i = \exp(\beta_0 + \beta_1) = 708,04$  představuje střední počet dopravních nehod za neztížených povětrnostních podmínek. Za neztížených podmínek se tedy stalo nejvíce nehod, což ale neznamená, že by tato situace byla více nebezpečná než například mlha. Dnů, kdy jsou povětrnostní podmínky neztížené je mnohem více než dnů, kdy jsou povětrnostní podmínky ztížené. Počet těchto dnů ovšem neznáme. A nakonec odhad  $\mu_i = \exp(\beta_0 + \beta_2) = 51,501$  zastupuje střední počet dopravních nehod, které se staly za deště, nebo když sněžilo.

## 7.2.2 Povětrnostní podmínky a alkohol

Jednou z akcí policie po příjezdu k dopravní nehodě může být zjišťování, zda účastník dopravní nehody požil návykovou látku. Tabulka 7.2.2.1 prezentuje počet dopravních nehod v závislosti na požití alkoholu.

TABULKA 7.2.2.1 Počet nehod v závislosti na alkoholu

ano	ne	nezjišťováno
34	690	135

Tento faktor jsme nemuseli nijak testovat, jelikož tvar modelu je zřejmý již při pohledu na tabulku. Každá třída má svůj parametr a můžeme tedy rovnou přejít k vytváření modelů počtu dopravních nehod závislých na povětrnostních podmínkách a na alkoholu.

Saturovaný model je model s interakcemi o 15 parametrech, jehož hodnota AIC je rovna 94,6. Dále přecházíme k modelům s interakcemi o 9 parametrech, jejichž hodnota AIC je rovna 86,4. V těchto modelech jsou nevýznamné všechny dvojné interakce, proto jsme přešli na modely bez interakcí o 5 parametrech, jejichž Akaikeho informační kritérium nabývá hodnoty 83,6. Modely s 5 parametry jsou nejvhodnější, jelikož mají nejnížší hodnotu AIC. Tvar všech modelů můžete nalézt v příloze CD, soubor R, povpodmalkohol.R. Vybraný model má tvar

$$\ln \mu_i = \beta_0 + \beta_1 I_{[povpodm=neztížen\acute{e}]} + \beta_2 I_{[povpodm=d\acute{e}\acute{s}\acute{t}|povpodm=sn\acute{e}\acute{z}\acute{e}n\acute{i}]} + \beta_3 I_{[alkohol=ne]} + \beta_4 I_{[alkohol=nezji\acute{s}\acute{t}\acute{o}v\acute{a}no]}, \quad i = 1, \dots, 15,$$

jehož parametr  $\beta$  jsme odhadli pomocí softwaru R hodnoty. Tyto odhady jsou zobrazeny v tabulce 7.2.2.2 spolu s jejich dalšími vlastnostmi.

TABULKA 7.2.2.2 Parametry  $\beta_j$ ,  $j = 0, 1, \dots, 4$

parametr	odhad	standartní odchylka	p-value	Odhad $\mu_i$
$\beta_0$	-0,051	0,222	0,817	0,95
$\beta_1$	3,384	0,149	< 0,0001	28,023
$\beta_2$	0,764	0,175	< 0,0001	2,038
$\beta_3$	3,01	0,176	< 0,0001	19,278
$\beta_4$	1,379	0,192	< 0,0001	3,772

Krom odhadu parametru  $\beta_0$  mají ostatní hodnoty odhadu parametrů kladné znaménko, proto odhady  $\mu_i$  spojené s těmito parametry budou vyšší než odhady  $\mu_i$  spojené s parametrem  $\beta_0$ .

Odhad  $\mu_i = \exp(\beta_0)$  zastupuje střední počet dopravních nehod, které vznikly na počátku deště, slabého deště, či za jiných povětrnostních podmínek v kombinaci se stavem, kdy řidič požil alkohol. Odhad  $\mu_i = \exp(\beta_0 + \beta_1)$  představuje střední počet dopravních nehod, které se staly za neztížených povětrnostních podmínek v kombinaci se stavem, kdy řidič požil alkohol. Odhad  $\mu_i = \exp(\beta_0 + \beta_2)$  představuje střední počet dopravních nehod, které se staly za deště, nebo když sněžilo v kombinaci se stavem, kdy řidič požil alkohol. Odhad  $\mu_i = \exp(\beta_0 + \beta_3)$  zastupuje střední počet dopravních nehod, které vznikly na počátku deště, slabého deště, či za jiných povětrnostních podmínek v kombinaci se stavem, kdy řidič nepožil alkohol. Odhad  $\mu_i = \exp(\beta_0 + \beta_4)$  zastupuje střední počet dopravních nehod, které vznikly na počátku deště, slabého deště, či za jiných povětrnostních podmínek v kombinaci, kdy se přítomnost alkoholu nezjišťovala. Ostatní odhady naleznete ve zmíněném souboru.

### 7.2.3 Povětrnostní podmínky, alkohol a stav povrchu vozovky

Třetí faktor patřící do této skupiny je stav povrchu vozovky. Počet dopravních nehod v závislosti na tomto faktoru uvádí tabulka 7.2.3.1.

TABULKA 7.2.3.1 Počet nehod v závislosti na stavu povrchu vozovky

na vozovce je náledí, ujetý sníh - neposypané	na vozovce je náledí, ujetý sníh - posypané	povrch mokrý	povrch suchý, neznečištěný	souvislá sněhová vrstva, rozbředlý sníh
46	19	181	594	19

Pro faktor stav povrchu vozovky jsme vytvořili čtyři modely, které můžete vidět v příloze CD, soubor R, vozovka.R a dle nejmenší hodnoty AIC = 38,5 jsme vybrali model se čtyřmi parametry ve tvaru

$$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{vozovka}=\text{povrch suchý, neznečištěný}]} + \beta_2 I_{[\text{vozovka}=\text{povrch mokrý}]} + \beta_3 I_{[\text{vozovka}=\text{na vozovce je náledí, ujetý sníh - neposypané}]}, \quad i = 1, \dots, 5,$$

který využijeme pro model počtu dopravních nehod v závislosti na povětrnostních podmínkách, alkoholu a stavu povrchu vozovky. Budeme pracovat s modely, které jsme vytvořili v kapitole 7.2.2. K těmto modelům přidáme faktor stav povrchu vozovky.

Začali jsme saturevaným modelem, který má 75 parametrů a vyskytují se v něm dvojné a jednoduché interakce. Model má nejnižší hodnotu Akaikeho informačního

kritéria, AIC = 269,6. Přestože bychom dle AIC měli tento model považovat za nejvhodnější, neuděláme to, jelikož obsahuje velké množství parametrů a my chceme co nejjednodušší model. Přecházíme tedy k nejlepšímu modelu z kapitoly 7.2.2 ve tvaru

$$\ln \mu_i = \beta_0 + \beta_1 I_{[povpodm=neztížen\acute{e}]} + \beta_2 I_{[povpodm=d\acute{e}\text{š}\acute{t}|povpodm=sn\acute{e}\text{ž}\acute{e}\text{n}\acute{ı}]} + \beta_3 I_{[alkohol=ne]} + \beta_4 I_{[alkohol=nezji\text{š}\acute{t}ov\acute{a}no]}, \quad i = 1, \dots, 15$$

a přidáme k němu do interakce faktor stav povrchu vozovky. Takto vzniklý model má 20 parametrů a vyskytují se v něm jednoduché interakce. Některé z nich jsou nevýznamné, proto je začneme postupně odstraňovat. Po odstranění všech nevýznamných parametrů nám zůstane model o 14 parametrech, z nichž 6 se týká dvojných interakcí. Akaikeho informační kritérium v tomto modelu vzrostlo oproti saturovanému modelu na hodnotu AIC = 322,6. Ačkoliv vzrůst je poměrně vysoký, budeme model se 14 parametry považovat za nejvhodnější, jelikož jsme původní model zjednodušili o 61 parametrů. Konečný model této kapitoly má tvar

$$\begin{aligned} \ln \mu_i = & \beta_0 + \beta_1 I_{[vozovka=povrch\ such\acute{y},\ nezne\acute{c}i\text{š}\acute{t}\acute{e}\text{n}\acute{y}]} + \beta_2 I_{[povpodm=nezt\acute{ı}\text{ž}\acute{e}\text{n}\acute{e}]} \\ & + \beta_3 I_{[povpodm=d\acute{e}\text{š}\acute{t}|povpodm=sn\acute{e}\text{ž}\acute{e}\text{n}\acute{ı}]} + \beta_4 I_{[vozovka=povrch\ mokr\acute{y}]} \\ & + \beta_5 I_{[vozovka=na\ vozovce\ je\ n\acute{a}\text{ł}\acute{e}\text{d}\acute{ı},\ ujet\acute{y}\ sn\acute{ı}\acute{h} - neposypan\acute{e}]} + \beta_6 I_{[alkohol=nezji\text{š}\acute{t}ov\acute{a}no]} \\ & + \beta_7 I_{[alkohol=ne]} + \beta_8 I_{[vozovka=povrch\ such\acute{y},\ nezne\acute{c}i\text{š}\acute{t}\acute{e}\text{n}\acute{y},\ povpodm=nezt\acute{ı}\text{ž}\acute{e}\text{n}\acute{e}]} \\ & + \beta_9 I_{[vozovka=povrch\ such\acute{y},\ nezne\acute{c}i\text{š}\acute{t}\acute{e}\text{n}\acute{y},\ povpodm=d\acute{e}\text{š}\acute{t}|povpodm=sn\acute{e}\text{ž}\acute{e}\text{n}\acute{ı}]} \\ & + \beta_{10} I_{[povpodm=d\acute{e}\text{š}\acute{t}|povpodm=sn\acute{e}\text{ž}\acute{e}\text{n}\acute{ı},\ vozovka=povrch\ mokr\acute{y}]} \\ & + \beta_{11} I_{[povpodm=nezt\acute{ı}\text{ž}\acute{e}\text{n}\acute{e},\ vozovka=na\ vozovce\ je\ n\acute{a}\text{ł}\acute{e}\text{d}\acute{ı},\ ujet\acute{y}\ sn\acute{ı}\acute{h} - neposypan\acute{e}]} \\ & + \beta_{12} I_{[povpodm=d\acute{e}\text{š}\acute{t}|povpodm=sn\acute{e}\text{ž}\acute{e}\text{n}\acute{ı},\ vozovka=na\ vozovce\ je\ n\acute{a}\text{ł}\acute{e}\text{d}\acute{ı},\ ujet\acute{y}\ sn\acute{ı}\acute{h} - neposypan\acute{e}]} \\ & + \beta_{13} I_{[povpodm=d\acute{e}\text{š}\acute{t}|povpodm=sn\acute{e}\text{ž}\acute{e}\text{n}\acute{ı},\ alkohol=nezji\text{š}\acute{t}ov\acute{a}no]}, \quad i = 1, \dots, 75, \end{aligned}$$

jehož parametr  $\beta$  jsme odhadli pomocí softwaru R hodnoty. Tyto odhady jsou zobrazeny v tabulce 7.2.3.2 spolu s jejich dalšími vlastnostmi.

TABULKA 7.2.3.2 Parametry  $\beta_j$ ,  $j = 0, 1, \dots, 13$

parametr	odhad	standartní odchylka	p-value	odhad
$\beta_0$	-3,335	0,339	< 0,0001	0,036
<b><math>\beta_1</math></b>	<b>0,79</b>	<b>0,58</b>	<b>0,173</b>	<b>0,078</b>
$\beta_2$	1,919	0,201	< 0,0001	0,243
$\beta_3$	1,802	0,364	< 0,0001	0,216
$\beta_4$	2,749	0,266	< 0,0001	0,556
$\beta_5$	2,047	0,413	< 0,0001	0,276
$\beta_6$	1,436	0,192	< 0,0001	0,15



$\beta_7$	3,01	0,176	< 0,0001	0,723
$\beta_8$	3,767	0,541	< 0,0001	1,539
$\beta_9$	-3,188	1,176	0,007	0,001
$\beta_{10}$	-1,121	0,366	0,002	0,012
$\beta_{11}$	-1,408	0,473	0,003	0,009
$\beta_{12}$	-1,108	0,507	0,029	0,012
$\beta_{13}$	-2,184	1,016	0,032	0,004

Vidíme, že odhady parametrů  $\beta_1, \dots, \beta_7$  jsou kladné hodnoty, čímž víme, že počty dopravních nehod spojené s těmito parametry jsou větší než počet nehod spojený s parametrem  $\beta_0$ . Dále přejdeme k interpretaci interakcí, které se vyskytují v modelu. V následující tabulce bude uvedena interpretace jednoduchých interakcí, které zůstaly v modelu. K hodnotám, které se v tabulce vyskytují, jsme došli stejným způsobem jako v kapitole 7.1.5.

Tabulka 7.2.3.3 Vysvětlení interakcí

param.	Faktory interakce	vysvětlení
$\beta_8$	Povrch suchý, neznečištěný a neztížené pov. podm.	Číslo $\exp(\beta_8) \doteq 43,3$ říká, že hodnota 95,3 je zhruba 43,3 krát větší než hodnota 2,2. Tedy, že nehod, které vznikly za neztížených povětr. podmínek a při suchém, neznečištěném povrchu, je zhruba 43,3 krát více než nehod, které vznikly na suchém, neznečištěném povrchu, když začínalo pršet, či slabě přšelo nebo za ostatních pov. podm.
		Číslo $\exp(\beta_8) \doteq 43,3$ říká, že hodnota 294,7 je zhruba 43,3 krát větší než hodnota 6,8. Tedy, že nehod, které vznikly za neztížených povětr. podmínek a při suchém, neznečištěném povrchu, je zhruba 43,3 krát více než nehod, které vznikly za neztížených povětr. podmínek a pokud byl na vozovce náledí, ujetý sníh – posypané nebo souvislá sněhová vrstva, rozbředlý sníh.
$\beta_9$	Povrch suchý, neznečištěný a sněžení nebo déšť	Číslo $\exp(-\beta_9) \doteq 24,2$ říká, že hodnota 0,09 je zhruba 24,2 krát menší než hodnota 2,2. Tedy, že nehod, které vznikly za deště nebo sněžení při suchém, neznečištěném povrchu, je zhruba 24,2 krát méně než nehod, které vznikly na suchém, neznečištěném povrchu, když začínalo pršet, či slabě přšelo nebo za ostatních pov. podm.
		Číslo $\exp(-\beta_9) \doteq 24,2$ říká, že hodnota 6 je zhruba 24,2 krát menší než hodnota 0,25. Tedy, že nehod, které vznikly za deště nebo sněžení při suchém, neznečištěném povrchu, je zhruba 24,2 krát méně než nehod, které vznikly za deště nebo sněžení, pokud bylo na vozovce náledí, ujetý sníh – posypané nebo souvislá sněhová vrstva, rozbředlý sníh.

$\beta_{10}$	sněžení nebo déšť a povrch mokrý	Číslo $\exp(-\beta_{10}) \doteq 3$ říká, že hodnota 2 je zhruba 3 krát menší než hodnota 6. Tedy, že nehod, které vznikly za deště nebo sněžení při mokrému povrchu, je zhruba 3 krát méně než nehod, které vznikly za deště nebo sněžení, pokud bylo na vozovce náledí, ujetý sníh – posypané nebo souvislá sněhová vrstva, rozbředlý sníh.
		Číslo $\exp(-\beta_{10}) \doteq 3$ říká, že hodnota 5,1 je zhruba 3 krát menší než hodnota 15,6. Tedy, že nehod, které vznikly za deště nebo sněžení při mokrému povrchu, je zhruba 3 krát méně než nehod, které vznikly za mokrého povrchu a když začínalo pršet, či slabě přšelo nebo za ostatních pov. podm.
$\beta_{11}$	neztížené pov. podm. a na vozovce je náledí, ujetý sníh - neposypané	Číslo $\exp(-\beta_{11}) \doteq 4$ říká, že hodnota 1,7 je zhruba 4 krát menší než hodnota 6,8. Tedy, že nehod za neztížených povětr. podmínek, když bylo na vozovce náledí, ujetý sníh - neposypané, je zhruba 4 krát méně než nehod za neztížených povětr. podmínek a pokud bylo na vozovce náledí, ujetý sníh – posypané nebo souvislá sněhová vrstva, rozbředlý sníh.
		Číslo $\exp(-\beta_{11}) \doteq 4$ říká, že hodnota 1,9 je zhruba 4 krát menší než hodnota 7,7. Tedy, že nehod za neztížených povětr. podmínek, když bylo na vozovce náledí, ujetý sníh - neposypané, je zhruba 4 krát méně než nehod, které vznikly, když začínalo pršet, či slabě přšelo nebo za ostatních pov. podm. a na vozovce bylo náledí, ujetý sníh – neposypané.
$\beta_{12}$	sněžení nebo déšť a na vozovce je náledí, ujetý sníh - neposypané	Číslo $\exp(-\beta_{12}) \doteq 3$ říká, že hodnota 2 je zhruba 3 krát menší než hodnota 6. Tedy, že nehod za deště nebo sněžení, když bylo na vozovce náledí, ujetý sníh - neposypané, je zhruba 3 krát méně než nehod za deště nebo sněžení, pokud bylo na vozovce náledí, ujetý sníh – posypané nebo souvislá sněhová vrstva, rozbředlý sníh.
		Číslo $\exp(-\beta_{12}) \doteq 3$ říká, že hodnota 2,6 je zhruba 3 krát menší než hodnota 7,7. Tedy, že nehod za deště nebo sněžení, když bylo na vozovce náledí, ujetý sníh - neposypané, je zhruba 3 krát méně než nehod když začínalo pršet, či slabě přšelo nebo za ostatních pov. podm. a na vozovce bylo náledí, ujetý sníh – neposypané.
$\beta_{13}$	sněžení nebo déšť a alkohol nezjišťován	Číslo $\exp(-\beta_{13}) \doteq 8,9$ říká, že hodnota 0,7 je zhruba 8,9 krát menší než hodnota 6. Tedy, že nehod za deště nebo sněžení, když nebyl zjišťován alkohol, je zhruba 8,9 krát méně než nehod za deště nebo sněžení, když řidič požil alkohol.
		Číslo $\exp(-\beta_{13}) \doteq 8,9$ říká, že hodnota 0,5 je 8,9 krát menší než hodnota 4,2. Tedy, že nehod za deště nebo sněžení, když nebyl zjišťován alkohol, je zhruba 8,9 krát méně než nehod když začínalo pršet, či slabě přšelo nebo za ostatních pov. podm. a nebyl zjišťován alkohol.

#### 7.2.4 Povětrnostní podmínky, alkohol, stav povrchu vozovky a viditelnost

Poslední faktor, který patří do této skupiny, je viditelnost. Je to poslední faktor, který budeme analyzovat v této práci a počet nehod v závislosti na faktoru viditelnost je uvedený v tabulce 7.2.4.1.

TABULKA 7.2.4.1 Počet nehod v závislosti na viditelnosti

v noci - bez veřejného osvětlení, viditelnost nezhoršená (pov.podm.)	v noci - s veřejným osvětlením, viditelnost nezhoršená (pov.podm.)	v noci - zhoršená viditelnost (pov.podm.)	ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek	ve dne, zhoršená viditelnost
74	82	24	592	87

Pro faktor viditelnost jsme vytvořili pět modelů, které můžete vidět v příloze CD, soubor R, viditelnost.R, a dle nejmenší hodnoty AIC = 39,01 jsme vybrali model se třemi parametry ve tvaru

$$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{viditelnost}=\text{ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek}]} + \beta_2 I_{[\text{viditelnost}=\text{v noci} - \text{zhoršená viditelnost (pov.podm.)}]}, \quad i = 1, \dots, 5,$$

který využijeme pro model počtu dopravních nehod v závislosti na povětrnostních podmínkách, alkoholu, stavu povrchu vozovky a viditelnosti.

Nejprve jsme postupovali tak, že jsme spojili nejlepší modely jednotlivých faktorů, čímž vznikl model s interakcemi o 108 parametrech, ve kterém vystupuje spousta nevýznamných interakcí. Z modelu jsme odstranili všechny interakce až na většinu jednoduchých a dostali jsme model s 36 parametry a hodnotou Akaikeho informačního kritéria AIC = 723,1. Dále jsme postupně odstraňovali jednoduché interakce, až jsme došli k modelu, který je dle AIC = 700,6 nejvhodnější. Model s jednoduchými interakcemi obsahuje 22 parametrů, což není zrovna optimální počet, proto jsme pokračovali v odstraňování nejméně významných parametrů. Jakmile jsme odstranili všechny nevýznamné parametry, zůstal nám model s 16 parametry, kde AIC = 764,4. Všechny modely můžete nalézt v příloze CD, soubor R, povpodmalkoholvozovkaviditelnost.R.

Zkoušeli jsme hledat model i způsobem, který jsme uplatňovali v předchozích kapitolách. Využili jsme nejlepší model z kapitoly 7.2.3 a přidali k němu do interakce faktor viditelnost. Vznikl nám model s 42 parametry, ve kterém se vyskytly dvojné a jednoduché interakce. Dvojně interakce byly nevýznamné, proto jsme je hned všechny odstranili, čímž jsme získali model o 30 parametrech a AIC = 718,7. V modelu zůstaly nevýznamné některé jednoduché interakce, které jsme postupně oddělávali, až jsme dospěli dle AIC = 706,3 k nejlepšímu modelu o 20 parametrech. Tento model, ač má nejnižší hodnotu AIC, obsahuje stále hodně parametrů, proto budeme pokračovat v odstraňování málo významných parametrů. Po odstranění všech možných parametrů jsme získali model

s 15 parametry, kde AIC = 770,42. Následující tabulka je srovnání hodnot AIC a počtu parametrů obou postupů.

TABULKA 7.2.4.2 Srovnání modelů

počet parametrů	AIC postup 1	AIC postup 2
108	790,9	-
42	-	728,8
30	711,5	718,7
29	709,5	716,8
28	707,7	715,3
27	706	713,4
26	704,3	711,9
25	702,7	710,1
24	701,3	709
23	701,3	707,4
22	700,6	707
21	703,3	707
20	711,1	706,3
19	713,5	708,9
18	716,3	712,1
17	722,7	715,2
16	764,4	727,8
15	-	770,4

Naším cílem je mít model s co nejméně parametry, ale taky aby byla přijatelná hodnota Akaikeho informačního kritéria. Aby byly splněny obě podmínky, zvolili jsme druhý postup a jako nejvhodnější model jsme vybrali model s 16 parametry, kde AIC = 727,8. Konečný model pro faktory působící na dopravní nehodu má tvar

$$\begin{aligned}
 \ln \mu_i = & \beta_0 + \beta_1 I_{[\text{viditelnost}=\text{ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek}]} \\
 & + \beta_2 I_{[\text{povpodm}=\text{neztížený}]} + \beta_3 I_{[\text{alkohol}=\text{ne}]} \\
 & + \beta_4 I_{[\text{viditelnost}=\text{v noci} - \text{zhoršená viditelnost (pov.podm.)}]} + \beta_5 I_{[\text{vozovka}=\text{povrch suchý, neznečištěný}]} \\
 & + \beta_6 I_{[\text{vozovka}=\text{povrch mokrý}]} + \beta_7 I_{[\text{povpodm}=\text{děšť} | \text{povpodm}=\text{sněžení}]} + \beta_8 I_{[\text{alkohol}=\text{nezjišťováno}]} \\
 & + \beta_9 I_{[\text{vozovka}=\text{na vozovce je náledí, ujetý sníh} - \text{neposypané}]} \\
 & + \beta_{10} I_{[\text{viditelnost}=\text{ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek, povpodm}=\text{neztížený}]} \\
 & + \beta_{11} I_{[\text{viditelnost}=\text{ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek, alkohol}=\text{ne}]} \\
 & + \beta_{12} I_{[\text{viditelnost}=\text{v noci} - \text{zhoršená viditelnost (pov.podm.)}, \text{vozovka}=\text{povrch suchý, neznečištěný}]} \\
 & + \beta_{13} I_{[\text{povpodm}=\text{neztížený, viditelnost}=\text{v noci} - \text{zhoršená viditelnost (pov.podm.)}]} \\
 & + \beta_{14} I_{[\text{povpodm}=\text{neztížený, vozovka}=\text{povrch suchý, neznečištěný}]} \\
 & + \beta_{15} I_{[\text{povpodm}=\text{neztížený, vozovka}=\text{na vozovce je náledí, ujetý sníh} - \text{neposypané}], \quad i = 1, \dots, 375,
 \end{aligned}$$

jehož parametr  $\beta$  jsme odhadli pomocí softwaru R hodnoty. Tyto odhady jsou zobrazeny v tabulce 7.2.4.3 spolu s jejich dalšími vlastnostmi..

TABULKA 7.2.4.3 Parametry  $\beta_j$ ,  $j = 0, 1, \dots, 15$

parametr	odhad	standartní odchylka	p-value	odhad
$\beta_0$	-3,968	0,285	< 0,0001	0,019
<b><math>\beta_1</math></b>	<b>-0,427</b>	<b>0,231</b>	<b>0,065</b>	<b>0,012</b>
$\beta_2$	0,897	0,206	< 0,0001	0,046
$\beta_3$	2,331	0,201	< 0,0001	0,195
<b><math>\beta_4</math></b>	<b>-0,329</b>	<b>0,249</b>	<b>0,187</b>	<b>0,014</b>
$\beta_5$	-2,109	1,016	0,038	0,002
$\beta_6$	2,254	0,178	< 0,0001	0,18
$\beta_7$	0,764	0,175	< 0,0001	0,041
$\beta_8$	1,379	0,192	< 0,0001	0,075
$\beta_9$	1,328	0,242	< 0,0001	0,071
$\beta_{10}$	1,909	0,208	< 0,0001	0,128
$\beta_{11}$	1,101	0,177	< 0,0001	0,057
$\beta_{12}$	3,227	1,144	0,005	0,477
<b><math>\beta_{13}</math></b>	<b>-20,429</b>	<b>583,15</b>	<b>0,972</b>	<b>0</b>
$\beta_{14}$	6,24	1,011	< 0,0001	9,705
$\beta_{15}$	-1,272	0,382	0,001	0,005

V následující tabulce bude uvedena interpretace jednoduchých interakcí, které zůstaly v modelu. Budeme zde interpretovat pouze interakce, které se nevyskytly v kapitole 7.2.3. K hodnotám, které se v tabulce vyskytují, jsme došli stejným způsobem jako v kapitole 7.1.5. Vynecháme ještě interpretaci parametru  $\beta_{13}$ , jelikož má velmi nízké hodnoty. K hodnotám, které se v tabulce vyskytují, jsme došli stejným způsobem jako v kapitole 7.1.5.

TABULKA 7.2.4.4 Vysvětlení interakcí

param.	Faktory interakce	vysvětlení
$\beta_{10}$	Ve dne, viditelnost nezhoršená vlivem povětr. podm. a neztížené povětr. podm.	Číslo $\exp(\beta_{10}) \doteq 3$ říká, že hodnota 4,6 je zhruba 3 krát větší než hodnota 1,5. Tedy, že nehod, které vznikly ve dne, kdy viditelnost byla nezhoršená a povětr. podmínky byly neztížené, je zhruba 3 krát více než nehod, které vznikly ve dne, kdy viditelnost byla nezhoršená vlivem povětr. a začínalo pršet, či slabě přšelo nebo za ostatních pov. podm.
		Číslo $\exp(\beta_{10}) \doteq 3$ říká, že hodnota 16,5 je zhruba 3 krát větší než hodnota 2,5. Tedy, že nehod, které vznikly ve dne, kdy viditelnost byla nezhoršená a povětr. podmínky byly neztížené, je zhruba 3 krát více než nehod, které vznikly za neztížených povětr. podmínek v noci - bez veřejného osvětlení, kdy viditelnost byla nezhoršená (pov.podm.), v noci - s veřejným osvětlením, kdy viditelnost byla nezhoršená (pov.podm.) nebo ve dne za zhoršené viditelnosti.
$\beta_{11}$	Ve dne, viditelnost nezhoršená vlivem povětr. podm. a alkohol ne	Číslo $\exp(\beta_{11}) \doteq 3$ říká, že hodnota 4,6 je zhruba 3 krát větší než hodnota 1,5. Tedy, že nehod, které vznikly ve dne, kdy viditelnost byla nezhoršená a řidič nepožil alkohol, je zhruba 3 krát více než nehod, které vznikly ve dne, kdy viditelnost byla nezhoršená a řidič požil alkohol.
		Číslo $\exp(\beta_{11}) \doteq 3$ říká, že hodnota 31 je zhruba 3 krát větší než hodnota 10,3. Tedy, že nehod, které vznikly ve dne, kdy viditelnost byla nezhoršená a řidič nepožil alkohol, je zhruba 3 krát více než nehod, kdy řidič nepožil alkohol řídil v noci - bez veřejného osvětlení, kdy viditelnost byla nezhoršená (pov.podm.), v noci - s veřejným osvětlením, kdy viditelnost byla nezhoršená (pov.podm.) nebo ve dne za zhoršené viditelnosti.
$\beta_{12}$	V noci – zhoršená viditelnost (pov.podm.) a povrch suchý, neznečištěný	Číslo $\exp(\beta_{12}) \doteq 25,2$ říká, že hodnota 18,1 je zhruba 25,2 krát větší než hodnota 0,7. Tedy, že nehod, které vznikly v noci za zhoršené viditelnosti a když byl povrch suchý, neznečištěný, je zhruba 25,2 krát více než nehod, které vznikly v noci za zhoršené viditelnosti a na vozovce bylo náledí, ujetý sníh – posypané nebo souvislá sněhová vrstva, rozbředlý sníh.
		Číslo $\exp(\beta_{12}) \doteq 25,2$ říká, že 3,1 je zhruba 25,2 krát větší než hodnota 0,12. Tedy, že nehod, které vznikly v noci za zhoršené viditelnosti a když byl povrch suchý, neznečištěný, je zhruba 25,2 krát více než nehod, kdy byl povrch suchý, neznečištěný a řidič řídil v noci - bez veřejného osvětlení, kdy viditelnost byla nezhoršená (pov.podm.), v noci - s veřejným osvětlením, kdy viditelnost byla nezhoršená (pov.podm.) nebo ve dne za zhoršené viditelnosti.

## ZÁVĚR

V prvních třech kapitolách byly objasněny pojmy týkající se silničního provozu, popsána nehodovost v České republice a informace o městě Zlín. Kapitola pět se věnovala základní statistice. Dozvěděli jste se, jak se definují základní pojmy, jak vypadají některá diskrétní rozdělení důležitá pro práci, co jsou to kategoriální data a kontingenční tabulky a v závěru kapitoly pět jste našli informace o klasických lineárních regresních modelech. Poslední teoretická kapitola byla kapitola šest, ve které jsme se zabývali zobecněnými lineárními modely. Podrobně byl rozebrán Poissonův loglineární model, pomocí kterého jsme analyzovaly počty dopravních nehod v kapitole sedm.

Kapitola sedm uvádí praktické použití Poissonova loglineárního modelu, pomocí kterého jsme modelovali počty dopravních nehod v obci Zlín. Faktory, se kterými jsme pracovali, jsme podle jejich charakteru rozdělili do dvou skupin. První z nich jsme nazvali faktory charakterizující počet dopravních nehod, do kterých spadají dny v týdnu, měsíce, místo nehody, druh nehody, zavinění nehody a druh vozidla, které zavinilo nehodu. Do druhé skupiny patří faktory, které ovlivňují vznik dopravní nehody, což jsou povětrnostní podmínky, alkohol, stav povrchu vozovky a viditelnost. U obou dvou skupin jsme model vytvářeli postupně.

V první skupině jsme začali faktorem dny v týdnu a přidávali jsme do modelu postupně další faktory. Modely, které jsme pro jednotlivé faktory vybrali, uvádí následující tabulka.

TABULKA 1 Přehled modelů dle faktorů charakterizujících počet dopravních nehod

FAKTOR	MODEL
den	$\ln[EY(den)] = \ln \mu_i = \beta_0 + \beta_1 I_{[den=SO]} + \beta_2 I_{[den=NE]}, \quad i = 1, \dots, 7.$
měsíc	$\ln \mu_i = \beta_0 + \beta_1 I_{[měsíc=březen červen duben květen srpen]} + \beta_2 I_{[měsíc=leden listopad prosinec říjen]} + \beta_3 I_{[měsíc=září]}, \quad i = 1, \dots, 12$
místo nehody	$\ln \mu_i = \beta_0 + \beta_1 I_{[místo=mimo křižovatku]}, \quad i = 1, 2$
Druh nehody	$\ln \mu_i = \beta_0 + \beta_1 I_{[druhneh=srážka s jedoucím nekolejovým vozidlem]}, \quad i = 1, \dots, 5$
zavinění nehody	$\ln \mu_i = \beta_0 + \beta_1 I_{[zavinění=lesní zvěří, domácím zvířectvem]} + \beta_2 I_{[zavinění=řidičem motorového vozidla]}, \quad i = 1, \dots, 6$
Vozidlo, které zavinilo nehodu	$\ln \mu_i = \beta_0 + \beta_1 I_{[druhvozza=osobní automobil]} + \beta_2 I_{[druhvozza=nákladní automobil]}, \quad i = 1, \dots, 5$

Po přidání každého faktoru jsme vytvořili nový model a konečný model s 9 jednoduchými interakcemi obsahující všechny faktory měl tvar

$$\begin{aligned} \ln \mu_i = & \beta_0 + \beta_1 I_{[\text{druhvozza}=\text{nákladní automobil}]} + \beta_2 I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]} \\ & + \beta_3 I_{[\text{den}=\text{NE}]} + \beta_4 I_{[\text{druhvozza}=\text{osobní automobil}]} + \beta_5 I_{[\text{zavinění}=\text{lesní zvěři, domácím zvířectvem}]} \\ & + \beta_6 I_{[\text{zavinění}=\text{řidičem motorového vozidla}]} + \beta_7 I_{[\text{misto}=\text{mimo křižovatku}]} + \beta_8 I_{[\text{den}=\text{SO}]} \\ & + \beta_9 I_{[\text{měsíc}=\text{březen}|červen|duben|květen|srpen]} + \beta_{10} I_{[\text{měsíc}=\text{září}]} \\ & + \beta_{11} I_{[\text{měsíc}=\text{leden}|listopad|prosinec|říjen]} \\ & + \beta_{12} I_{[\text{druhvozza}=\text{nákladní automobil, druhneh}=\text{srážka s jedoucím nekolejovým vozidlem}]} \\ & + \beta_{13} I_{[\text{druhvozza}=\text{nákladní automobil, den}=\text{NE}]} \\ & + \beta_{14} I_{[\text{druhvozza}=\text{osobní automobil, zavinění}=\text{lesní zvěři, domácím zvířectvem}]} \\ & + \beta_{15} I_{[\text{druhvozza}=\text{osobní automobil, zavinění}=\text{řidičem motorového vozidla}]} \\ & + \beta_{16} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, druhvozza}=\text{osobní automobil}]} \\ & + \beta_{17} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, misto}=\text{mimo křižovatku}]} \\ & + \beta_{18} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, den}=\text{SO}]} \\ & + \beta_{19} I_{[\text{druhneh}=\text{srážka s jedoucím nekolejovým vozidlem, zavinění}=\text{řidičem motorového vozidla}]} \\ & + \beta_{20} I_{[\text{zavinění}=\text{lesní zvěři, domácím zvířectvem, misto}=\text{mimo křižovatku}], \quad i = 1, 2, \dots, 25200} \end{aligned}$$

Vysvětlení jednotlivých parametrů i interakcí naleznete v textu. Uvedu zde pouze jeden způsob interpretace parametru  $\beta_{13}$ , který mě zaujal. Dozvěděli jsme se, že nehod, které zavinil nákladní automobil v neděli, je zhruba 2,3 krát méně než nehod, které zavinil nákladní automobil v pracovní den, což odpovídá reálné situaci. Nákladní automobily jsou v neděli omezeny dobou, kdy smí jezdit.

Při práci s modelem obsahujícím faktor dny a měsíce jsme ukázali, že záleží na tabulce, která vstupuje do modelu. Musíme si dát pozor na interpretaci odhadů  $\mu_i$ . V kapitole 7.1.3 jsme do modelu přidali faktor místo nehody, který měl jen dvě kategorie. Ukázali jsme dva modely. První, když  $\beta_1$  souvisel s kategorií na křižovatce a druhý, když  $\beta_1$  souvisel s kategorií mimo křižovatku. Dozvěděli jsme se, že modely jsou ekvivalentní, počet nehod se musí doplňovat.

Stejným způsobem jsme postupovali i při práci faktory ovlivňujícími vznik dopravní nehody. Modely, které jsme pro jednotlivé faktory vybrali, uvádí následující tabulka.

TABULKA 2 Přehled modelů dle faktorů ovlivňujících počet dopravních nehod

FAKTOR	MODEL
povětrnostní podmínky	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{povpodm}=\text{neztíženě}]} + \beta_2 I_{[\text{povpodm}=\text{déšť} povpodm=\text{sněžení}]}, \quad i = 1, \dots, 5$
alkohol	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{alkohol}=\text{ne}]} + \beta_2 I_{[\text{alkohol}=\text{nezjišťováno}]}, \quad i = 1, 2, 3$



Stav povrchu vozovky	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{vozovka}=\text{povrch suchý, neznečištěný}]} + \beta_2 I_{[\text{vozovka}=\text{povrch mokrý}]} + \beta_3 I_{[\text{vozovka}=\text{na vozovce je náledí, ujetý sníh – neposypané}]}, \quad i = 1, \dots, 5$
viditelnost	$\ln \mu_i = \beta_0 + \beta_1 I_{[\text{viditelnost}=\text{ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek}]} + \beta_2 I_{[\text{viditelnost}=\text{v noci – zhoršená viditelnost (pov.podm.)}]}, \quad i = 1, \dots, 5$

Po přidání posledního faktoru viditelnost do modelu, jsme zkoušeli vytvářet modely dvěma způsoby. Nejlepší model s 6 jednoduchými interakcemi měl tvar

$$\begin{aligned} \ln \mu_i = & \beta_0 + \beta_1 I_{[\text{viditelnost}=\text{ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek}]} + \beta_2 I_{[\text{povpodm}=\text{neztížené}]} \\ & + \beta_3 I_{[\text{alkohol}=\text{ne}]} + \beta_4 I_{[\text{viditelnost}=\text{v noci – zhoršená viditelnost (pov.podm.)}]} \\ & + \beta_5 I_{[\text{vozovka}=\text{povrch suchý, neznečištěný}]} + \beta_6 I_{[\text{vozovka}=\text{povrch mokrý}]} \\ & + \beta_7 I_{[\text{povpodm}=\text{déšť}|\text{povpodm}=\text{sněžení}]} + \beta_8 I_{[\text{alkohol}=\text{nezjišťováno}]} \\ & + \beta_9 I_{[\text{vozovka}=\text{na vozovce je náledí, ujetý sníh – neposypané}]} \\ & + \beta_{10} I_{[\text{viditelnost}=\text{ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek, povpodm}=\text{neztížené}]} \\ & + \beta_{11} I_{[\text{viditelnost}=\text{ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek, alkohol}=\text{ne}]} \\ & + \beta_{12} I_{[\text{viditelnost}=\text{v noci – zhoršená viditelnost (pov.podm.)}, \text{vozovka}=\text{povrch suchý, neznečištěný}]} \\ & + \beta_{13} I_{[\text{povpodm}=\text{neztížené}, \text{viditelnost}=\text{v noci – zhoršená viditelnost (pov.podm.)}]} \\ & + \beta_{14} I_{[\text{povpodm}=\text{neztížené}, \text{vozovka}=\text{povrch suchý, neznečištěný}]} \\ & + \beta_{15} I_{[\text{povpodm}=\text{neztížené}, \text{vozovka}=\text{na vozovce je náledí, ujetý sníh – neposypané}]}, \quad i = 1, \dots, 375. \end{aligned}$$

Z tohoto modelu jsme se například dozvěděli, že nehod, které vznikly ve dne, kdy viditelnost byla nezhoršená a řidič nepožil alkohol, je zhruba 3 krát více než nehod, které vznikly ve dne, kdy viditelnost byla nezhoršená, a řidič požil alkohol. Toto zjištění může někoho trochu zaskočit, jelikož v médiích se dopravní nehody obvykle spojují s řízením pod vlivem alkoholu. Zlín je tedy město, ve kterém se stalo více nehod, aniž by řidič požil alkohol. Interpretace dalších interakcí naleznete v textu.

Diplomová práce pro mne osobně byla v mnoha ohledech přínosná. V úvodu práce jsem se zabývala pojmy souvisejícími s provozem na komunikacích, povinnostech řidiče, atd., a jelikož vlastním řidičské oprávnění, znalost některých informací využiji i v praxi. Praktická část práce byla velice zdouhavá, což se projevilo i v jejím rozsahu. Při řešení prvních tří faktorů jsem byla částečně zklamaná, jelikož výstup nepodal nijak zajímavou informaci, jakmile se ale v modelu začaly vyskytovat interakce, dospěli jsme občas i k zajímavým výsledkům. Jsem ráda, že jsem se naučila s metodou pracovat, protože při studiu jsem se setkala pouze se základními informacemi. Nakonec nesmím zapomenout zmínit software R, pomocí něhož jsem získala všechny výsledky a naučila se v něm model používat.

## POUŽITÁ LITERATURA

- [1] *POLICIE ČESKÉ REPUBLIKY: Zákon 361-2000 Sb. o provozu na pozemních komunikacích a o změnách některých zákonů.rtf* [online]. Dostupný z WWW: <http://www.policie.cz/clanek/vybrane-ceske-pravni-predpisy.aspx>. [citováno 6.6.2012]
- [2] *ALKOHOLIK: Trest za řízení v opilosti, pod vlivem alkoholu* [online]. Dostupný z WWW: [http://www.alkoholik.cz/zavislost/opily\\_ridic\\_tresty/trest\\_za\\_rizeni\\_v\\_opilosti\\_pod\\_vlivem\\_alkoholu.html](http://www.alkoholik.cz/zavislost/opily_ridic_tresty/trest_za_rizeni_v_opilosti_pod_vlivem_alkoholu.html). [citováno 11.6.2012]
- [3] Pracovníci ředitelství služby dopravní policie Policejního prezidia policie České republiky, *Přehled o nehodovosti na pozemních komunikacích v České republice za rok 2011*, Praha: ředitelství služby dopravní policie Policejního prezidia policie České republiky, duben 2012.
- [4] *EUROPA: EU road fatalities* [online]. Dostupný z WWW: [http://ec.europa.eu/transport/road\\_safety/pdf/observatory/trends\\_figures.pdf](http://ec.europa.eu/transport/road_safety/pdf/observatory/trends_figures.pdf) [citováno 14.6.2012]
- [5] *NASIPOLITICI* [online]. Dostupný z WWW: <http://www.nasipolitici.cz>. [citováno 6.6.2012]
- [6] *Členové vlády* [online]. Dostupné z WWW: <http://www.vlada.cz/cz/clenove-vlady/historie-minulych-vlad/prehled-vlad-cr/1993-2007-cr/vaclav-klaus-1/prehled-ministru-24631/> [citováno 6.6.2012]
- [7] *Politici* [online]. Dostupné z WWW: <http://www.denikpolitika.cz> [citováno 6.6.2012]
- [8] *Tolerance při dechové zkoušce: 0,2 promile může mít v krvi i abstinent, 27. ledna 2010, Autor: Roman Švidrnoch* [online]. Dostupný z WWW: [http://auto.idnes.cz/tolerance-pri-dechove-zkousce-0-2-promile-muze-mit-v-krvi-i-abstinent-1p7-/automoto.aspx?c=A100126\\_155545\\_automoto\\_fdv](http://auto.idnes.cz/tolerance-pri-dechove-zkousce-0-2-promile-muze-mit-v-krvi-i-abstinent-1p7-/automoto.aspx?c=A100126_155545_automoto_fdv) [citováno 24.10.2012]
- [9] *Bodové přestupky – pokuty, body* [online]. Dostupný z WWW: <http://www.12bodu.cz> [citováno 24.10.2012]
- [10] *O MĚSTĚ* [online]. Dostupné z WWW: <http://www.zlin.eu/> [citováno 24.10.2012]

- [11] MELOUN, Milan; MILITKÝ, Jiří, *Statistická analýza experimentálních dat*, 2. Vydání, Praha: Academia, 2004. ISBN: 80-200-1254-0
- [12] ANDĚL, Jiří, *Základy matematické statistik*, 1. vydání, Praha: Matfyzpress, 2005. ISBN: 80-86732-40-1
- [13] ANDĚL, Jiří, *Matematická statistika*, 1. vydání, Praha: SNTL/Alfa, 1978.
- [14] KUNDEROVÁ, Pavla, *Základy pravděpodobnosti a matematické statistiky*, 1. vydání, Univerzita Palackého v Olomouci, 2004. ISBN: 80-244-0813-9
- [15] BUDÍKOVÁ, Marie, MIKOLÁŠ, Štěpán, OSECKÝ, Pavel, *Teorie pravděpodobnosti a matematická statistika*, 3. vydání, Masarykova univerzita v Brně, 2007. ISBN: 80-210-3313-4
- [16] HEBÁK, Petr, HUSTOPECKÝ, Jiří, MALÁ, Iva, *Vícerozměrné statistické metody (2)*, 1. Vydání, Praha: Informatorium, 2005. ISBN: 80-7333-036-9
- [17] HEBÁK, Petr, HUSTOPECKÝ, Jiří, PECÁKOVÁ, Iva, PRŮŠA, Milan, ŘEZANKOVÁ, Hana, SVOBODOVÁ, Alžběta, VLACH, Petr, *Vícerozměrné statistické metody (3)*, 1. Vydání, Praha: Informatorium, 2005. ISBN: 80-7333-039-3
- [18] AGRESTI, Alan, *Categorical Data Analysis*, 2. edition, John Wiley & Sons, Inc., Hoboken, New Jersey, 2002. ISBN: 0-471-36093-7
- [19] *Deviance při Poissonově rozdělení dat* [online]. Dostupné z WWW: [http://publib.boulder.ibm.com/infocenter/spsstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Falg\\_genlog-poisson\\_residuals\\_deviance.htm](http://publib.boulder.ibm.com/infocenter/spsstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Falg_genlog-poisson_residuals_deviance.htm)  
[citováno 7.11.2012]
- [20] DÍREROVÁ, Monika, *Bakalářská práce – Analýza dopravní nehodovosti*, KMA, Přírodovědecká fakulta UP Olomouc, 2010

## SEZNAM TABULEK, OBRÁZKŮ A GRAFŮ

TABULKA 1.2.1	Trest za řízení v opilosti (str. 13)
TABULKA 2.1	Počet nehod a následky nehod v období 1993 – 2011 (str. 16)
TABULKA 2.2	Hlavní příčiny nehod řidičů motorových vozidel (str. 19)
TABULKA 2.1.3	Počet usmrcených osob v zemích EU v letech 2001 – 2010 (str. 20)
TABULKA 2.2.1	Ministři dopravy (str. 21)
TABULKA 3.2.1	Směr staničení silnic (str. 23)
TABULKA 3.2.2	Přibližné úseky silnic použité v práci (str. 24)
TABULKA 5.4.1	Pravděpodobnostní tabulka (str. 34)
TABULKA 5.4.2	Kontingenční tabulka (str. 34)
TABULKA 7.1.1.1	Počet nehod v jednotlivých dnech (str. 52)
TABULKA 7.1.1.2	Vlastnosti modelu 1 (str. 53)
TABULKA 7.1.1.3	Vlastnosti modelu 2 (str. 53)
TABULKA 7.1.1.4	Vlastnosti modelu 3 (str. 54)
TABULKA 7.1.1.5	Vlastnosti modelu 4 (str. 55)
TABULKA 7.1.1.6	Modely a jejich charakteristiky (str. 55)
TABULKA 7.1.2.1	Počet nehod v jednotlivých měsících (str. 56)
TABULKA 7.1.2.2	Přehled modelů (str. 56)
TABULKA 7.1.2.3	Parametry $\beta_j$ , $j = 0,1,2,3$ (str. 57)
TABULKA 7.1.2.4	Parametry $\beta_j$ , $j = 0,1, \dots, 5$ (str. 59)
TABULKA 7.1.2.5	Přehled odhadů $\mu_i$ , $i = 1, \dots, 84$ (str. 60)
TABULKA 7.1.2.6	Kontingenční tabulky (str. 61)
TABULKA 7.1.2.7	Parametry $\beta_j$ , $j = 0,1,2$ (str. 61)
TABULKA 7.1.2.8	Kontingenční tabulka (str. 62)
TABULKA 7.1.2.9	Parametry $\beta_j$ , $j = 0,1,2$ (str. 62)
TABULKA 7.1.2.10	Kontingenční tabulka (str. 63)
TABULKA 7.1.2.11	Parametry $\beta_j$ , $j = 0,1,2$ (str. 63)
TABULKA 7.1.2.12	Kontingenční tabulka (str. 64)
TABULKA 7.1.2.13	Parametry $\beta_j$ , $j = 0,1,2$ (str. 65)
TABULKA 7.1.3.1	Počet nehod v závislosti na místě (str. 66)
TABULKA 7.1.3.2	Parametry $\beta_j$ , $j = 0,1, \dots, 6$ (str. 67)

TABULKA 7.1.3.3	Parametry $\beta_j$ , $j = 0,1, \dots, 6$ (str. 68)
TABULKA 7.1.4.1	Počet nehod v závislosti na druhu dopravní nehody (str. 70)
TABULKA 7.1.4.3	Parametry $\beta_j$ , $j = 0,1, \dots, 9$ (str. 71)
TABULKA 7.1.4.4	Kombinace faktoru místo a druh nehody (str. 72)
TABULKA 7.1.4.5	Kombinace faktoru místo a druh nehody (str. 73)
TABULKA 7.1.5.1	Počet nehod v závislosti na zavinění nehody (str. 75)
TABULKA 7.1.5.2	Parametry $\beta_j$ , $j = 0,1, \dots, 14$ (str. 76)
TABULKA 7.1.5.3	Kombinace faktoru zavinění a místo (str.77)
TABULKA 7.1.5.4	Vysvětlení interakcí (str. 78)
TABULKA 7.1.6.1	Počet nehod v závislosti na druhu vozidla, které zavinilo nehodu (str. 79)
TABULKA 7.1.6.2	Parametry $\beta_j$ , $j = 0,1, \dots, 20$ (str. 81)
TABULKA 7.1.6.3	Vysvětlení interakcí (str. 82)
TABULKA 7.2.1.1	Počet nehod v závislosti na povětrnostních podmínkách (str.83)
TABULKA 7.2.1.2	Přehled modelů (str. 84)
TABULKA 7.2.1.3	Parametry $\beta_j$ , $j = 0,1,2$ (str. 84)
TABULKA 7.2.2.1	Počet nehod v závislosti na alkoholu (str. 85)
TABULKA 7.2.2.2	Parametry $\beta_j$ , $j = 0,1, \dots, 4$ (str. 85)
TABULKA 7.2.3.1	Počet nehod v závislosti na stavu povrchu vozovky (str. 86)
TABULKA 7.2.3.2	Parametry $\beta_j$ , $j = 0,1, \dots, 13$ (str. 87)
TABULKA 7.2.3.3	Vysvětlení interakcí (str. 88)
TABULKA 7.2.4.1	Počet nehod v závislosti na viditelnosti (str. 90)
TABULKA 7.2.4.2	Srovnání modelů (str. 91)
TABULKA 7.2.4.3	Parametry $\beta_j$ , $j = 0,1, \dots, 15$ (str. 92)
TABULKA 7.2.4.4	Vysvětlení interakcí (str. 93)
TABULKA 1	Přehled modelů dle faktorů charakterizujících počet dopravních nehod (str. 94)
TABULKA 2	Přehled modelů dle faktorů ovlivňujících počet dopravních nehod (str. 95)
GRAF 2.1	Počet nehod a počet usmrcených osob v období 1993 – 2011 (str. 17)
GRAF 2.2	Poměr počtu nehod a počtu usmrcených osob v období 1993 – 2011 (str. 18)

- OBRÁZEK 3.1      Znak a vlajka obce Zlín (str. 22)
- OBRÁZEK 3.2.1    Silnice města Zlín (str. 23)
- OBRÁZEK 3.2.2    Přibližné úseky silnic použité v práci (str. 24)

## **PŘÍLOHY**

- Příloha [1] Řidičské oprávnění
- Příloha [2] Přehled dopravních nehod
- Příloha [3] Formulář evidence nehod v silničním provozu
- Příloha [4] CD: Soubor Excel  
Soubor Erko

Příloha 1 Řidičské oprávnění [20]

Skupina	Oprávnění	Věk
A	k řízení motocyklů o výkonu do 25kW s poměrem výkon/hmotnost nepřesahujícím 0,16kW/kg nebo motocyklů s postranním vozíkem a s poměrem výkon/hmotnost nepřesahujícím 0,16kW/kg.	18 let
A	k řízení motocyklů o výkonu nad 25kW nebo s poměrem výkon/hmotnost přesahujícím 0,16kW/kg nebo motocyklů s postranním vozíkem a s poměrem výkon/hmotnost přesahujícím 0,16kW/kg.	21 let
A1	k řízení lehkých motocyklů o objemu válců nepřesahujícím $125\text{ cm}^3$ a o výkonu nejvýše 11kW.	16 let
B	a) k řízení motorových vozidel, s výjimkou vozidel uvedených v předchozích dvou skupinách a motocyklu o výkonu nad 25kW nebo s poměrem výkon/hmotnost přesahujícím 0,16kW/kg nebo motocyklu s postranním vozíkem a s poměrem výkon/hmotnost přesahujícím 0,16kW/kg, jejichž maximální přípustná hmotnost nepřevyšuje 3500kg a s nejvýše 8 místy k sezení, kromě místa řidiče; k tomuto motorovému vozidlu smí být připojeno přípojně vozidlo o maximální přípustné hmotnosti nepřevyšující 750kg, b) k řízení traktorů a pracovních strojů samojízdných, jejichž maximální přípustná hmotnost nepřevyšuje 3500kg, c) k řízení jízdních souprav složených z motorového vozidla podle písmene a) nebo b) a přípojně vozidla, pokud maximální přípustná hmotnost soupravy nepřevyšuje 3500kg a maximální přípustná hmotnost přípojně vozidla nepřevyšuje pohotovostní hmotnost 2) motorového vozidla.	18 let
B1	k řízení motorových tříkolových a čtyřkolových vozidel uvedených ve skupině B, jejichž maximální konstrukční rychlost převyšuje 45km/h nebo jsou poháněna spalovacím motorem o objemu válců převyšujícím $50\text{ cm}^3$ nebo jsou poháněna jakýmkoliv jiným zařízením srovnatelného výkonu. Pohotovostní hmotnost těchto vozidel nesmí být vyšší než 550kg; do pohotovostní hmotnosti vozidla elektricky poháněného se nezapočítává hmotnost akumulátorů.	17 let
C	k řízení motorových vozidel, s výjimkou vozidel uvedených ve skupině D a D1, jejichž maximální přípustná hmotnost převyšuje 3500kg; k tomuto motorovému vozidlu smí být připojeno přípojně vozidlo, jehož maximální přípustná hmotnost nepřevyšuje 750kg.	18 let
C1	k řízení motorových vozidel, s výjimkou vozidel uvedených ve skupině D a D1, jejichž maximální přípustná hmotnost převyšuje 3500kg, avšak nepřevyšuje 7500kg; k tomuto motorovému vozidlu smí být připojeno přípojně vozidlo, jehož maximální přípustná hmotnost nepřevyšuje 750kg.	18 let
D	k řízení motorových vozidel určených pro přepravu osob s více než 8 místy k sezení, kromě místa řidiče; k tomuto motorovému vozidlu smí být připojeno přípojně vozidlo, jehož maximální přípustná hmotnost nepřevyšuje 750kg.	21 let
D1	k řízení motorových vozidel určených pro přepravu osob s více než 8 místy k sezení, kromě místa řidiče, avšak ne s více než 16 místy k sezení, kromě místa řidiče; k tomuto motorovému vozidlu smí být připojeno přípojně vozidlo, jehož maximální přípustná hmotnost nepřevyšuje 750kg.	21 let
B + E	k řízení jízdních souprav složených z motorového vozidla uvedeného ve skupině B a přípojně vozidla, pokud nejde o jízdní soupravu podle skupiny B písm. c).	18 let



C + E	k řízení jízdních souprav složených z motorového vozidla uvedeného ve skupině C a přípojného vozidla, jehož maximální přípustná hmotnost převyšuje 750kg.	18 let
C1 + E	k řízení jízdních souprav složených z motorového vozidla uvedeného ve skupině C1 a přípojného vozidla, jehož maximální přípustná hmotnost převyšuje 750kg. Maximální přípustná hmotnost soupravy však nesmí převyšovat 12000kg a maximální přípustná hmotnost přípojného vozidla nesmí převyšovat pohotovostní hmotnost motorového vozidla.	18 let
D + E	k řízení jízdních souprav složených z motorového vozidla uvedeného ve skupině D a přípojného vozidla, jehož maximální přípustná hmotnost převyšuje 750kg.	21 let
D1 + E	k řízení jízdních souprav složených z motorového vozidla uvedeného ve skupině D1 a přípojného vozidla, jehož maximální přípustná hmotnost převyšuje 750kg a nejsou v něm přepravovány osoby. Maximální přípustná hmotnost soupravy však nesmí převyšovat 12000kg a maximální přípustná hmotnost přípojného vozidla nesmí převyšovat pohotovostní hmotnost motorového vozidla.	21 let
AM	k řízení mopedů a malých motocyklů s maximální konstrukční rychlostí 45km/h.	15 let
T	k řízení traktorů a pracovních strojů samojízdných; k motorovému vozidlu smí být připojeno přípojné vozidlo.	17 let



