

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Bc. Markéta Ondrušková

Stanovení sazeb pojistného v neživotním pojištění

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Mgr. Ing. Jakub Mertl

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2014

Ráda bych poděkovala vedoucímu diplomové práce Mgr. Ing. Jakubovi Mertlovi za trpělivost při konzultacích, za spoustu cenných rad a připomínek a za poskytnutí reálných dat. Dále děkuji RNDr. Martinovi Brandovi, Ph.D. za poskytnutí literatury. Můj dík patří také rodině, příteli a známým za podporu při psaní práce.

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 10. 4. 2014

Markéta Ondrušková

Název práce: Stanovení sazeb pojistného v neživotním pojištění

Autor: Bc. Markéta Ondrušková

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Mgr. Ing. Jakub Mertl

Abstrakt: Předmětem této diplomové práce je stanovit pojistné v pojištění odpovědnosti z provozu motorového vozidla. V práci jsou představeny jak přímé způsoby výpočtu pojistného, tak výpočty pojistného pomocí zvlášť spočtené frekvence škod a výše škod. Při stanovení pojistného je kladen důraz na využití zobecněných lineárních modelů, jejichž teorie je rovněž v práci popsána. Vyložené přístupy jsou aplikovány na reálná data a porovnány, následně je zvolen finální model pro výpočet pojistného. Nakonec je stanoveno výsledné pojistné se zahrnutím IBNR rezervy a nákladů.

Klíčová slova: zobecněný lineární model, frekvence škod, výše škod, nettopojistné, bruttopojistné, pojištění odpovědnosti z provozu motorového vozidla, bonus-malus.

Title: Pricing in non-life insurance

Author: Bc. Markéta Ondrušková

Department: Department of Probability and Mathematical Statistics

Supervisor: Mgr. Ing. Jakub Mertl

Abstract: The aim of this thesis is to determine the premium of motor third party liability insurance. Thesis explores both direct approach to premium calculation and calculation using separately calculated claims frequency and claims severity. Emphasis is put on the use of generalized linear models whose theory is also present. Described approaches to calculation are applied on real world data, compared with each other and then final model for premium calculation is selected. Premium is determined with the inclusion of IBNR reserves and expenses.

Keywords: generalized linear model, claims frequency, claims severity, net premium, gross premium, motor third party liability insurance, bonus-malus.

Obsah

Úvod	1
Historický úvod	2
1 Sazby v neživotním pojištění	5
1.1 Základní principy neživotního pojištění	5
1.2 Pojištění odpovědnosti za škodu způsobenou provozem motorového vozidla	7
1.2.1 Frekvence škod	8
1.2.2 Severita škod	10
1.2.3 Nettopojistné	12
1.2.4 Brutto pojistné	15
1.3 Kredibilita	16
2 Zobecněný lineární model	18
2.1 Základní prvky	18
2.2 Odhad parametrů	20
2.3 Kvalita modelu	21
2.4 Využití	22
3 Malé škody	25
3.1 Škodní frekvence	27
3.2 Škodní frekvence — GLM	30
3.3 Severita	38
3.4 Nettopojistné	39
4 Velké škody	43
5 Stanovení bruttopojistného	48
5.1 Bonus-malus	48
5.2 Brutto pojistné	51
Závěr	55
Seznam použité literatury	57
A Tabulky	58

Úvod

Neživotní pojištění v dnešní době zaujímá na českém trhu co do objemu předepsaného pojistného větší podíl než pojištění životní. V rámci neživotního pojištění tvoří smlouvy pojištění odpovědnosti z provozu vozidla více než třetinu uzavřených pojistek (za rok 2013 asi 6,87 mil. pojistek pojištění odpovědnosti z provozu vozidla). Toto pojištění je význačné v tom, že je povinně smluvní, takže každý provozovatel motorového vozidla má povinnost jej uzavřít. Na druhou stranu má ale možnost vybrat si pojistitele svého vozidla z více než deseti pojišťoven. Vysoká konkurence nutí pojišťovny poskytovat pojištění za co nejnižší ceny. Pojistné je však nutné stanovovat zodpovědně, aby byla pojišťovna schopná plnit své závazky i při velkém tlaku na cenu pojištění.

Cílem práce je popsat jak nejjednodušší přístupy ke stanovení pojistného v pojištění odpovědnosti z provozu motorového vozidla, tak přístupy, které využívají členění smluv do kategorií v rámci proměnných. Dále pak při výpočtu pojistného využít zobecněný lineární model a následně aplikovat různé přístupy k výpočtu pojistného na reálná data. Nakonec chceme zvolit model, který nejpřesněji pokryje nastalé škody a dopočítat pomocí IBNR rezervy a nákladů finální bruttopojistné.

V historickém úvodu se seznámíme s prvotní legislativní úpravou pojištění odpovědnosti z provozu vozidla, s první úpravou výše pojistného a s dalším vývojem až do konce 20. století.

První část první kapitoly věnujeme neživotnímu pojištění obecně. Od další části se již budeme soustředit na pojištění odpovědnosti z provozu vozidla. Vložíme různé přístupy k výpočtu frekvence škod, výše škod a nettopojistného, mj. se budeme soustředit na využití zobecněných lineárních modelů. Ty popisuje kapitola 2.

Další kapitoly se zabývají zpracováním reálných dat. Třetí kapitola se soustředí na malé škody, následující kapitola zpracovává velké škody. Poslední kapitola je věnována souhrnnému nettopojistnému, kalibraci modelu a výpočtu bruttopojistného.

Data jsou zpracována v programu Wolfram Mathematica 8.0, obrázky jsou vytvořeny také v tomto programu.

Historický úvod

Rozvoj automobilového průmyslu na přelomu 19. a 20. století s sebou postupně přinesl i otázku odpovědnosti za škody způsobené provozem motorového vozidla. V roce 1908 začal v zemích České a Moravskoslezské platit zákon č. 162/1908¹ o ručení za škody z provozování jízdních silostrojů (automobilů), který upravoval možnost (nikoliv povinnost) „pojištění proti ručební povinnosti za škody z provozování jízdních silostrojů“. Koncem roku 1932 vstoupil v platnost zákon č. 198/1932 Sb. o dopravě motorovými vozidly, který ukládal povinnost sjednat pojištění odpovědnosti z provozu motorového vozidla provozovatelům dopravy osob a nákladů za úplatu. Pojištění odpovědnosti za škody však nebylo povinné pro ostatní řidiče, což se změnilo během roku 1935, kdy začal platit zákon č. 81/1935 Sb. o jízdě motorovými vozidly. Tento zákon ukládal každému držiteli motorového vozidla, které bylo zapsáno v rejstříku československých motorových vozidel, povinnost pojistit se proti následkům odpovědnosti z provozu tohoto vozidla.

V roce 1948 došlo vyhláškou ministra financí (vyhláška 977/1948 Ú. 1. I) ke sloučení pojišťoven a vznikla tak jediná pojišťovna: Československá pojišťovna, národní podnik. Formu pojištění odpovědnosti za škody způsobené provozem vozidla upravil zákon č. 56/1950 Sb. a z doposud povinného pojištění se tak stalo pojištění zákonné. Sazby pojistného byly v roce 1951 upraveny vyhláškou č. 105/1951 Ú.l., pojistné bylo pro jednotlivé druhy motorových vozidel jednotné, tabulka 1 shrnuje sazby pro československá motorová vozidla.

Druh motorového vozidla	Celoroční pojistné
vozik pro nemocné, jízdní kolo s pomocným motorem	12 Kčs
motorové kolo, motocykl, tříkolka, motorové vozítko	48 Kčs
osobní automobil, sanitní, pohřební, dodávkový nebo kombinovaný vůz, autodrožka, hotelové auto	144 Kčs
nákladní automobil, zvláštní motorové vozidlo, motorový pojízdný stroj pracovní, tankový automobil, zvláštní poznávací značka	240 Kčs
autobus	321,60 Kčs
traktor	52,80 Kčs
Rozšíření platnosti pojištění na ostatní území	
Evropy a na území asijské části SSSR	4 Kčs
Pojistné z přívěsných vozů je zahrnuto v pojistném z motorového vozidla.	

Tabulka 1: Sazebník pro československá motorová vozidla z roku 1951.

¹Zákony čerpáme z internetového portálu www.epravo.cz.

O šest let později došlo vyhláškou ministerstva financí č. 38/1957 Ú. l. ke změnám v sazebníku: začaly se rozlišovat sazby pro objemové kategorie motorů vozidel (objemem motoru vozidla rozumíme zdvihový objem válců motoru vozidla). U motocyklů s obsahem válců do 100 ccm činilo celoroční pojistné 24 Kčs, u motocyklů s obsahem válců nad 100 ccm byla sazba 48 Kčs. Za osobní automobily s obsahem válců do 500 ccm (vozítka) bylo třeba uhradit 60 Kčs a za ostatní osobní automobily, sanitní, pohřební, dodávkové nebo kombinované vozy bylo roční pojistné ve výši 144 Kčs. Tyto sazby platily bez zásadních změn až do roku 1992, kdy nabyla účinnosti vyhláška č. 492/1991 Sb., ve které došlo ke změně rozdělení vozidel podle objemu motoru a navýšení sazeb pojistného. Pro ilustraci rozčlenění vozidel si uvedeme tabulku 2, která obsahuje sazby ročního pojistného pro vybraná tuzemská motorová vozidla.

Jednostopé motorové vozidlo a motorová tříkolka s hmotností do 400 kg, se zdvihovým objemem válců	
1. do 50 ccm včetně	60 Kčs
2. nad 50 ccm do 350 ccm včetně	120 Kčs
3. nad 350 ccm	168 Kčs
Osobní automobil do celkové hmotnosti 3 500 kg a motorová tříkolka s celkovou hmotností nad 400 kg, se zdvihovým objemem válců:	
1. do 2 000 ccm včetně nebo vozidla na elektrický pohon	288 Kčs
2. nad 2 000 ccm	384 Kčs

Tabulka 2: Sazebník pro vybraná tuzemská motorová vozidla z roku 1992.

Již po jednom roce došlo k navýšení sazeb pojistného, ale objemové kategorie zůstaly nezměněny. Ty byly změněny až v roce 1995, kdy nabyla účinnosti vyhláška ministerstva financí 246/1994 Sb. Osobní automobily do celkové hmotnosti 3 500 kg byly rozděleny do pěti kategorií podle zdvihového objemu válců následovně: do 1 000 ccm včetně, nad 1 000 ccm do 1 350 ccm včetně, nad 1 350 ccm do 1 850 ccm včetně, nad 1 850 ccm do 2 500 ccm včetně a nad 2 500 ccm.

Naposledy byly konkrétní sazby pojistného za zákonné pojištění odpovědnosti za škodu způsobenou provozem motorového vozidla stanoveny vyhláškou v roce 1999 (pro výše zmíněné kategorie osobních automobilů), tyto sazby pro vybraná tuzemská motorová vozidla jsou shrnuty v tabulce 3. V tomtéž roce byl vydán zákon 168/1999 Sb. o pojištění odpovědnosti z provozu vozidla, který sice upravil podmínky pojištění, ale již nestanovil konkrétní sazby pojistného. Do této doby bylo možné sjednat pojištění odpovědnosti za škodu způsobenou provozem motorového vozidla pouze u jedné pojišťovny (Česká pojišťovna a. s.), následně došlo k demonopolizaci trhu v této oblasti. Změnila se i forma pojištění: ze zákonného se stalo povinné smluvní pojištění. Po těchto změnách následovalo tříleté přechodné období částečné regulace a v roce 2003 se plně rozvinula konkurence v této oblasti trhu.

Dvoukolové motorové vozidlo a motorová tříkolka s celkovou hmotností do 400 kg se zdvihovým objemem válců	
1. do 50 ccm včetně	168,-
2. nad 50 ccm do 350 ccm včetně	420,-
3. nad 350 ccm do 500 ccm včetně	780,-
4. nad 500 ccm	1 188,-
Osobní automobil do celkové hmotnosti 3 500 kg včetně všech modifikací odvozených od osobního automobilu, motorová tříkolka s celkovou hmotností nad 400 kg; uvedená vozidla se zdvihovým objemem válců:	
1. do 1 000 ccm včetně nebo vozidla na elektrický pohon	912,-
2. nad 1 000 ccm do 1 350 ccm včetně	2 124,-
3. nad 1 350 ccm do 1 850 ccm včetně	2 952,-
4. nad 1 850 ccm do 2 500 ccm včetně	4 524,-
5. nad 2 500 ccm	7 128,-

Tabulka 3: Sazebník pro vybraná tuzemská motorová vozidla z roku 1999.

Závěrem tohoto historického úvodu uvedeme tabulku 4, v níž srovnáme průměrnou roční hrubou mzdou (2. sloupec) s ročním pojistným za pojištění odpovědnosti z provozu motorového vozidla (3. sloupec). Pro srovnání volíme motorové vozidlo o objemu 1 001–1 350 ccm, což byl do 90. let 20. století typický rozsah objemu rodinného automobilu. I v dnešní době (rok 2014) patří vozidla s objemem kolem 1 300 ccm mezi hojně prodávané.

V tabulce máme ve 4. sloupci uveden podíl ročního pojistného ku průměrné roční mzdě. Vidíme, že od roku 1960 do roku 1991 bylo pojistné stejně vysoké, zatímco průměrná roční mzda rostla, tudíž podíl se postupně snižoval. Mezi lety 1992–1999 dochází ke značnému růstu mezd, pojistné však roste rychleji, podíl se tedy zvyšuje. V roce 2000 došlo k demonopolizaci trhu, konkurenční prostředí stlačilo podíl až na hodnotu 0,67 % za rok 2013.

rok	průměrná roční mzda	roční pojistné	podíl (v %)
1960	15 636	144	0,92
1975	27 756	144	0,52
1991	45 504	144	0,32
1992	55 728	288	0,52
1996	117 900	804	0,68
1999	153 564	2 124	1,38
2013	301 536	2 021	0,67

Tabulka 4: Srovnání průměrné roční mzdy (data pochází z www.czso.cz) a ročního pojistného.

Kapitola 1

Sazby v neživotním pojištění

1.1 Základní principy neživotního pojištění

Při výpočtu sazeb pojistného se vychází ze statistických údajů zachycujících škodní vývoj, nazýváme je statistické podklady. Pojišťovna trvale pozoruje různé hodnoty a ukazatele, které mají vliv na výpočet pojistného - velmi zjednodušeně by se dalo říci, že sleduje, kolik bylo vyplaceno na pojistném plnění, jaké měla náklady a kolik bylo vybráno na pojistném. Následující text vychází z [2].

Jednotlivé pojistné smlouvy jsou shlukovány do *tarifních skupin* (tříd), kde v jedné skupině může být jednotná sazba pojistného. Podmínkou pro to, aby smlouvy mohly být v jedné tarifní skupině, je přibližně stejné pojištěné riziko. Jinými slovy: smlouvám v rámci jedné tarifní skupiny odpovídá určitá úroveň tarifních proměnných (rizikových faktorů). Například v pojištění odpovědnosti za škodu způsobenou provozem motorového vozidla může být tarifní proměnnou objem válců, typ vozidla, region, stáří vozidla apod.

V rámci jedné tarifní skupiny pak v daném roce sledujeme tyto statistické podklady:

- počet pojištění m ,
- počet pojistných událostí N ,
- celková pojistná částka (součet pojistných částek všech pojištění),
- celkové pojistné plnění (součet pojistných plnění u všech pojištění, obvykle zahrnujeme i rezervu na budoucí pojistná plnění z pojistných událostí vzniklých v daném roce),
- maximální škoda v daném roce
- celkové pojistné (součet přes všechna pojištění).

Na základě těchto podkladů dále počítáme následující statistické ukazatele:

- průměrné pojistné plnění: $PPP = \frac{\text{celkové pojistné plnění}}{m}$,
- průměrná pojistná částka: $PPČ = \frac{\text{celková pojistná částka}}{m}$,

- průměrná škoda: $P\check{S} = \frac{\text{celkové pojistné plnění}}{N}$,
- škodní frekvence: $\check{S}F = q_1 = \frac{N}{m}$,
- pojistná sazba: $PS = \frac{\text{celkové pojistné}}{\text{celková pojistná částka}}$,
- škodní sazba: $\check{S}S = \frac{\text{celkové pojistné plnění}}{\text{celkové pojistná částka}}$,
- škodní průběh: $\check{S}P = \frac{\text{celkové pojistné plnění}}{\text{celkové pojistné}}$,
- škodní stupeň: $\check{S}St = q_2 = \frac{P\check{S}}{PP\check{C}}$.

Při výpočtu ročního nettopojistného P předpokládáme, že každá z m pojistek v dané tarifní skupině má stejnou pojistnou částku S . Podle principu ekvivalence mezi příjmy a výdaji dostaneme:

$$m \cdot P = N \cdot P\check{S}.$$

Úpravou pak máme:

$$P = \frac{N}{m} \cdot \frac{P\check{S}}{PP\check{C}} \cdot PP\check{C}.$$

Využijeme zápis pomocí statistických ukazatelů q_1 a q_2 , průměrnou pojistnou částku $PP\check{C}$ nahradíme pojistnou částkou S (podle předchozího $PP\check{C} = S$) a máme vztah pro roční nettopojistné:

$$P = q_1 \cdot q_2 \cdot S.$$

Tento vzorec je možné též uvést ve tvaru pojistné sazby, tzn. ve tvaru ročního nettopojistného p na jednotkovou pojistnou částku:

$$p = q_1 \cdot q_2.$$

Výpočet nettopojistného je konstruován tak, že v průměru pokryje pojistná plnění. Navýšíme-li nettopojistné o nákladové složky, o složku sloužící k pokrytí náhlých nepříznivých výkyvů škodního průběhu, tzv. bezpečnostní přírážku, a o kalkulovaný zisk pojistitele dostáváme bruttopojistné. V případě, že uvažujeme pouze nettopojistné a bezpečnostní přírážku, můžeme mluvit o rizikovém pojistném. Toto rizikové pojistné (RP) můžeme zkonstruovat např. podle principu střední hodnoty:

$$RP = (1 + \lambda_1) \cdot P,$$

kde $\lambda_1 > 0$, nebo podle principu směrodatné odchylky:

$$RP = P + \lambda_2 s,$$

kde s je odhadnutá směrodatná odchylka související se statistickým odhadem nettopojistného. Principů pro stanovení pojistného je možné najít v literatuře více, uveďme ještě princip rozptylu:

$$RP = P + \lambda_3 s^2,$$

kde s^2 je odhadnutý rozptyl, který souvisí se statistickým odhadem nettopojistného.

Jak již bylo zmíněno výše, rozšířením rizikového pojistného o náklady a zisk dostaneme bruttopojistné, což můžeme zapsat takto:

$$B = RP + \textit{náklady}_n + (\textit{náklady}_z) \cdot B + z \cdot B, \quad (1.1)$$

kde B označuje bruttopojistné, $\textit{náklady}_n$ náklady nezávislé na výši pojistného, $\textit{náklady}_z$ (v procentech) náklady rostoucí s výší pojistného a konečně symbol z (v procentech) značí kalkulovaný zisk. Vztah pro bruttopojistné pak lze psát ve tvaru:

$$B = \frac{RP + \textit{náklady}_n}{1 - \textit{náklady}_z - z}. \quad (1.2)$$

1.2 Pojištění odpovědnosti za škodu způsobenou provozem motorového vozidla

Dále se v práci budeme zabývat pojištěním odpovědnosti za škodu způsobenou provozem motorového vozidla, proto se nyní zaměříme na teorii související s tímto konkrétním typem pojištění.

V pojištění odpovědnosti z provozu motorového vozidla nefiguruje pojistná částka, ale pracuje se zde s limity pojistného plnění. Tyto limity vyjadřují nejvyšší možné plnění pojistitele při pojistné události. Podle zákona č. 168/1999 Sb. musí minimální výše limitu při škodě na zdraví činit nejméně 35 000 000 Kč na každého zraněného nebo usmrčeného a minimální výše limitu při škodách na majetku musí činit 35 000 000 Kč bez ohledu na počet poškozených.

Tabulka 1.1 ilustruje, jaké výše limitů lze v dnešní době sjednat u českých pojišťoven. Vybrali jsme pojišťovny s největším objemem předepsaného smluvního pojistného za pojištění odpovědnosti z provozu vozidla. Limity pojistného plnění jsou vždy dva a jsou uvedeny ve tvaru limit pojistného plnění za škody na zdraví na každého zraněného či usmrčeného/limit pojistného plnění za škody na majetku¹.

Při stanovování nettopojistného v pojištění odpovědnosti za škodu způsobenou provozem motorového vozidla nás bude především zajímat frekvence škod a výše škod. Tyto dva základní ukazatele jsou často ovlivněny různými faktory neboli proměnnými. Typicky těmito proměnnými například jsou: věk pojistníka,

¹Zdrojem limitů jsou webové stránky jednotlivých pojišťoven. Informace o předepsaném smluvním pojistném jsou na internetových stránkách České asociace pojišťoven www.cap.cz.

pojišťovna	nabízené limity
Česká pojišťovna	35/35, 50/50, 100/100, 120/120, 150/150
Kooperativa	35/35, 70/70, 100/100
Česká podnikatelská pojišťovna	35/35, 50/50, 111/111
Allianz	35/35, 50/50, 150/150
ČSOB pojišťovna	44/35, 60/60, 100/100

Tabulka 1.1: Přehled nabízených limitů vybraných českých pojišťoven.

trvalé bydliště pojistníka (resp. počet obyvatel dané obce), bonusová třída, údaje o vozidle jako výkon motoru, hmotnost vozidla, objem motoru (resp. zdvihový objem válců motoru) či stáří vozidla. Jakmile odhadneme počet škod a jejich výši, které v následujícím období očekáváme, jejich zkombinováním získáme nettopojistné, které je třeba k pokrytí pojistného plnění.

V následujícím textu se seznámíme s různými přístupy stanovení škodní frekvence, výše škod a následně nettopojistného. Budeme používat označení m pro celkový počet zkoumaných smluv, N pro počet škod nastalých na těchto m smlouvách za jeden rok, expozici k -té smlouvy ($k = 1, \dots, m$) budeme značit ω_k . Expozicí rozumíme délku trvání pojistné smlouvy v letech — obvykle $\omega_k = 1$. Pro představu je v tabulce 1.2 uvedeno několik příkladů výpočtu expozice.

počátek pojištění	konec pojištění	expozice
1. 1. 2011	31. 12. 2011	1
1. 1. 2011	22. 8. 2011	234/365
1. 1. 2011	2. 2. 2011	33/365

Tabulka 1.2: Ukázka výpočtu expozice.

Dále budeme značit výši škody u k -té smlouvy jako z_k . Označení dalších veličin bude upřesněno tam, kde se bude vyskytovat.

1.2.1 Frekvence škod

Frekvence škod vyjadřuje počet škod nastalých na smlouvě za daný časový úsek, nadále pro nás bude tímto časovým úsekem jeden rok.

Začneme s nejjednodušším přístupem ke škodní frekvenci, kdy neuvažujeme žádné proměnné, ale bereme v úvahu souhrnně všechny smlouvy. Vztah pro výpočet frekvence ${}^I f$ je

$${}^I f = \frac{N}{\sum_{k=1}^m \omega_k}, \quad (1.3)$$

kde N je počet škod, m je počet všech pojištění (pojistných smluv) a ω_k značí expozici k -té smlouvy. Frekvence škod je v tomto případě pro všechny smlouvy, resp. klienty stejná.

Další přístup k výpočtu frekvence škod již bude využívat kategorizaci v rámci

proměnných (např. kategorie 18–29 let, 30–59 let a 60 a více let pro proměnnou věk). Uvažujme R kategoriálních proměnných ovlivňujících výpočet frekvence. Pro každou kategorii proměnné r ($r = 1, \dots, R$) spočteme frekvenci výskytu škod jako podíl počtu škod (N_{r,r_i}) a součtu expozic v dané kategorii, což můžeme zapsat takto:

$$f_{r,r_i} = \frac{N_{r,r_i}}{\sum_{k=1}^{|r,r_i|} \omega_{r,r_{ik}}}. \quad (1.4)$$

kde $|r, r_i|$ značí počet smluv v r_i -té kategorii r -té proměnné a $\omega_{r,r_{ik}}$ je expozice k -té smlouvy v r_i -té kategorii r -té proměnné.

U klienta s konkrétními kategoriemi proměnných spočteme frekvenci škod ${}^{II}f$ jako geometrický průměr frekvencí jednotlivých kategorií

$${}^{II}f = \sqrt[R]{\prod_{r=1}^R f_{r,r_i}}, \quad (1.5)$$

kde R je počet uvažovaných proměnných.

Frekvenci škod je také možné modelovat pomocí zobecněného lineárního modelu, kde u vysvětlované proměnné (frekvence) předpokládáme Poissonovo rozdělení. Zobecněným lineárním modelem se zabýváme v kapitole 2.

Při výpočtech v neživotním pojištění je možné využít i matematického modelování některých pojistných jevů.

Nyní uvedeme některé modely počtu pojistných událostí (počtu škod, počtu pojistných nároků), které vycházejí z rozdělení pravděpodobností diskrétní náhodné veličiny N , čerpáno z [1].

Poissonovo rozdělení, $N \sim \text{Po}(\vartheta)$:

$$\mathbf{P}(N = n) = \frac{\vartheta^n}{n!} e^{-\vartheta}, \quad n = 0, 1, 2, \dots, \quad (1.6)$$

kde $\vartheta > 0$ je jediný parametr tohoto rozdělení. Střední hodnota a rozptyl náhodné veličiny N :

$$\mathbf{E}N = \vartheta, \quad \text{var}N = \vartheta.$$

Použití tohoto modelu je vhodné při velkém počtu nezávislých pojistných smluv s malou pravděpodobností výskytu pojistné události.

Negativně binomické rozdělení, $N \sim \text{NB}(h, p)$:

$$\mathbf{P}(N = n) = \binom{h+n-1}{n} p^h (1-p)^n, \quad n = 0, 1, 2, \dots, \quad (1.7)$$

je rozdělení o dvou parametrech $h > 0$, $0 < p < 1$. Střední hodnota a rozptyl mají tvar:

$$\mathbf{E}N = \frac{h(1-p)}{p}, \quad \text{var}N = \frac{h(1-p)}{p^2}.$$

Při $h \in \mathbb{N}$ lze N interpretovat jako počet nezdarů před h -tým zdarem (resp. pojistnou událostí) s pravděpodobností zdaru p a pravěpodobností nezdaru $1 - p$.

Smíšené Poissonovo rozdělení: předpokládáme náhodnou veličinu θ s hustotou rozdělení pravděpodobnosti $u(x)$, $x \geq 0$ a dále předpokládáme, že podmíněné rozdělení počtu pojistných událostí N za podmínky $\theta = \vartheta$ je Poissonovo se střední hodnotou ϑ . Pravděpodobnostní rozdělení smíšeného Poissonova rozdělení (nepodmíněné rozdělení počtu pojistných událostí):

$$\mathbb{P}(N = n) = \int_0^\infty \frac{\vartheta^n}{n!} e^{-\vartheta} u(\vartheta) d\vartheta, \quad n = 0, 1, 2, \dots \quad (1.8)$$

Střední hodnota náhodné veličiny se složeným Poissonovým rozdělením je

$$\mathbb{E}N = \mathbb{E}\mathbb{E}(N|\theta) = \mathbb{E}\theta$$

a rozptyl je

$$\begin{aligned} \text{var}N &= \mathbb{E}\mathbb{E}[(N - \mathbb{E}N)^2|\theta] = \mathbb{E}\mathbb{E}[(N - \mathbb{E}(N|\theta) + \mathbb{E}(N|\theta) - \mathbb{E}N)^2|\theta] = \\ &= \mathbb{E}\mathbb{E}[(N - \mathbb{E}(N|\theta))^2|\theta] + \mathbb{E}[\mathbb{E}(N|\theta) - \mathbb{E}N]^2 = \\ &= \mathbb{E}\text{var}(N|\theta) + \text{var}\mathbb{E}(N|\theta) = \mathbb{E}\theta + \text{var}\theta. \end{aligned}$$

Rozdělení lze použít u heterogenních pojistných kmenů, kde u pojistných smluv s malým rizikem nabývá parametr ϑ malých hodnot a naopak (u pojistných smluv s velkým rizikem nabývá ϑ velkých hodnot).

1.2.2 Severita škod

V následující části se budeme zabývat tím, jaké výše dosahuje škoda. Opět začneme nejjednodušším přístupem — spočteme si průměrnou výši škody I_z přes všechny uvažované smlouvy

$$I_z = \frac{\sum_{k=1}^m z_k}{N}, \quad (1.9)$$

kde m je počet všech smluv, z_k značí výši škody na k -té smlouvě a N počet všech škod.

Nyní přistoupíme k otázce stanovení výše škody jiným způsobem, budeme uvažovat rozdělení smluv do kategorií. Mějme R proměnných a u každé proměnné uvažujeme určitý počet kategorií. Pro každou kategorii proměnné r ($r = 1, \dots, R$) spočteme průměrnou výši škody \bar{z}_{r,r_i} jako podíl součtu škod a počtu škod v dané kategorii, což zapíšeme takto:

$$\bar{z}_{r,r_i} = \frac{\sum_{k=1}^{|r,r_i|} z_{r,r_{ik}}}{N_{r,r_i}}, \quad (1.10)$$

kde r_i označuje i -tou kategorií r -té proměnné s počtem smluv $|r, r_i|$, $z_{r,r_{ik}}$ je výše k -té škody v i -té kategorii r -té proměnné a N_{r,r_i} značí počet škod v i -té kategorii r -té proměnné.

U klienta s konkrétními kategoriemi proměnných spočteme očekávanou škodu jako průměr výší škod jednotlivých kategorií

$$\Pi_z = \frac{\sum_{r=1}^R \bar{z}_{r,r_i}}{R}, \quad (1.11)$$

kde R je počet uvažovaných proměnných.

Výši škody je možné modelovat i pomocí zobecněného lineárního modelu, kde u vysvětlované proměnné (výše škody) typicky předpokládáme gama rozdělení. Zobecněný lineární model je vyložen v kapitole 2.

Při modelování výši škod lze také využít modely vycházející z pravděpodobnostního rozdělení spojitě náhodné veličiny X , která označuje výši škody. Rozdělení náhodné veličiny X popisujeme hustotou rozdělení pravděpodobnosti $f(x)$, čerpáme z [1], [3].

Exponenciální rozdělení, $X \sim \text{Exp}(\alpha)$:

$$f(x) = \alpha e^{-\alpha x}, \quad x \geq 0, \quad (1.12)$$

kde $\alpha > 0$. Střední hodnota a rozptyl náhodné veličiny X :

$$\mathbb{E}X = \frac{1}{\alpha}, \quad \text{var}X = \frac{1}{\alpha^2}.$$

Toto rozdělení může podcenit pravděpodobnost výskytu vysokých škod, protože pokles pravděpodobnosti $\mathbb{P}(X > x)$ k 0 pro $x \rightarrow \infty$ je v modelu příliš rychlý. Rozdělení může být využito k modelování doby mezi pojistnými událostmi.

Paretovo rozdělení:

$$f(x) = \alpha k^\alpha x^{-1-\alpha}, \quad x \geq k \quad (1.13)$$

je rozdělení s parametry: $\alpha > 0$, $k > 0$. Střední hodnota a rozptyl náhodné veličiny s tímto rozdělením:

$$\mathbb{E}X = \frac{k\alpha}{\alpha - 1} \text{ pro } \alpha > 1, \quad \text{var}X = \frac{k^2\alpha}{(\alpha - 1)^2(\alpha - 2)} \text{ pro } \alpha > 2.$$

Paretovo rozdělení typu II²:

$$f(x) = \alpha k^\alpha (k + x - \mu)^{-1-\alpha}, \quad x \geq \mu, \quad (1.14)$$

kde pro parametry platí: $\alpha > 0$, $k > 0$, $\mu > 0$. Střední hodnota a rozptyl:

$$\mathbb{E}X = \frac{k}{\alpha - 1} + \mu \text{ pro } \alpha > 1, \quad \text{var}X = \frac{k^2\alpha}{(\alpha - 1)^2(\alpha - 2)} \text{ pro } \alpha > 2.$$

²Zde čerpáme z dokumentace programu Wolfram Mathematica

Paretovo rozdělení lze použít v případech s odlehlými extrémními hodnotami škod.

Gama rozdělení, $X \sim \Gamma(\alpha, p)$:

$$f(x) = \frac{\alpha^p}{\Gamma(p)} x^{(p-1)} e^{-\alpha x}, \quad x \geq 0, \quad (1.15)$$

je rozdělení dvouparametrické, $\alpha > 0, p > 0$; $\Gamma(p)$ je hodnota funkce gama v bodě p . Střední hodnota a rozptyl:

$$\mathbb{E}X = \frac{p}{\alpha}, \quad \text{var}X = \frac{p}{\alpha^2}.$$

V případě, že p je celočíselné, je hustota gama rozdělení hustotou součtu p nezávislých stejně rozdělených náhodných veličin s hustotou $\alpha e^{-\alpha x}, x \geq 0$. Speciálně při hodnotě parametru $p = 1$ dostáváme exponenciální rozdělení ($X \sim \text{Exp}(\alpha)$).

Weibullovo rozdělení:

$$f(x) = b \alpha^b x^{b-1} \exp(-\alpha^b x^b), \quad x \geq 0, \quad (1.16)$$

kde $\alpha > 0$ a $b > 0$. Střední hodnota a rozptyl náhodné veličiny s Weibullovým rozdělením:

$$\mathbb{E}X = \frac{1}{\alpha} \Gamma\left(\frac{b+1}{b}\right), \quad \text{var}X = \frac{1}{\alpha^2} \left(\Gamma\left(\frac{b+2}{b}\right) - \Gamma^2\left(\frac{b+2}{b}\right) \right).$$

Pro $b = 1$ dostáváme exponenciální rozdělení.

Logaritmicko-normální rozdělení, $X \sim \text{LN}(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \quad x > 0, \quad (1.17)$$

je rozdělení o dvou parametrech $-\infty < \mu < \infty$ a $\sigma > 0$. Střední hodnota a rozptyl:

$$\mathbb{E}X = \exp\left(\mu + \frac{1}{2}\sigma^2\right), \quad \text{var}X = \exp(2\mu + \sigma^2) \cdot (\exp(\sigma^2) - 1).$$

U tohoto rozdělení platí, že logaritmus náhodné veličiny X má normální rozdělení se střední hodnotou μ a rozptylem σ^2 ($\log X \sim N(\mu, \sigma^2)$).

1.2.3 Nettopojistné

Nyní se budeme věnovat výpočtu ročního nettopojistného. Nejprve si představíme nejjednodušší výpočet, poté popíšeme postup stanovení nettopojistného pomocí proměnných a jejich kategorií a nakonec využijeme k výpočtu nettopojistného frekvenci škod a výši škod, kterými jsme se zatím zabývali odděleně.

Nejjednodušší způsob výpočtu nettopojistného je ten, kdy bez jakékoliv kategorizace sečteme všechny škody z_k (výše škody na k -té smlouvě) a vydělíme součtem expozičních ω_k (expoziční k -té smlouvy),

$$IP = \frac{\sum_{k=1}^m z_k}{\sum_{k=1}^m \omega_k}. \quad (1.18)$$

Výsledkem je jednotná sazba nettopojistného pro všechny smlouvy.

Druhý způsob výpočtu nettopojistného již využívá členění souboru smluv do kategorií v rámci vybraných proměnných. Cílem je získat nettopojistné jako součin *bazické hodnoty* a *faktorů*, kde každý faktor reprezentuje kategorii klienta v dané proměnné. Počet faktorů je tedy roven počtu uvažovaných proměnných.

Nettopojistné budeme počítat na základě R proměnných. Nejprve si spočteme nettopojistné pro každou kategorii všech uvažovaných proměnných. Toto nettopojistné je pouze dílčím výpočtem, od kterého se budou odvíjet výpočty dalších hodnot — faktorů a bazické hodnoty (viz dále). Nettopojistné pro r_i -tou kategorií proměnné r ($r = 1, \dots, R$) si spočteme takto:

$$P_{r,r_i} = \frac{\sum_{k=1}^{|r,r_i|} z_{r,r_{ik}}}{\sum_{k=1}^{|r,r_i|} \omega_{r,r_{ik}}}, \quad (1.19)$$

kde $|r, r_i|$ označuje počet smluv spadajících do r_i -té kategorie r -té proměnné, $z_{r,r_{ik}}$ je výše k -té škody v r_i -té kategorii a $\omega_{r,r_{ik}}$ značí expoziční k -té smlouvy v r_i -té kategorii.

Pro výpočet faktorů si v rámci r -té proměnné zvolíme referenční kategorii ref , ke které budeme vztahovat ostatní kategorie r -té proměnné. Kritériem pro stanovení referenční kategorie může například být největší součet expozičních, mělo by však být u všech proměnných stejné. Vztah pro faktor fac_{r,r_i} r_i -té kategorie r -té proměnné pak je:

$$fac_{r,r_i} = \frac{P_{r,r_i}}{P_{r,ref}}, \quad (1.20)$$

kde $P_{r,ref}$ je dílčí nettopojistné referenční kategorie.

Než se dostaneme ke vztahu pro výpočet bazické hodnoty, je potřeba si spočítat průměrný faktor r -té proměnné \overline{fac}_r . Ten spočteme jako vážený průměr faktorů jednotlivých kategorií r -té proměnné, kde jako váhy použijeme celkové expoziční jednotlivých kategorií:

$$\overline{fac}_r = \frac{\sum_{r_i} \left(fac_{r,r_i} \cdot \sum_{k=1}^{|r,r_i|} \omega_{r,r_{ik}} \right)}{\sum_{k=1}^m \omega_k}, \quad (1.21)$$

kde r_i probíhá přes všechny kategorie r -té proměnné. Vztah (1.21) můžeme dále upravit dosazením za fac_{r,r_i} podle (1.20), kde navíc využijeme vztah (1.19) pro P_{r,r_i} . Po zjednodušení pak máme:

$$\overline{fac}_r = \frac{1}{P_{r,ref}} \cdot \frac{\sum_{k=1}^m z_k}{\sum_{k=1}^m \omega_k}, \quad (1.22)$$

kde m je počet všech smluv.

Pro každou proměnnou máme tedy průměrný faktor. Nyní tyto průměrné faktory využijeme k výpočtu bazické hodnoty P_{base} :

$$P_{base} = \frac{1}{R} \cdot \frac{\sum_{k=1}^m z_k}{\prod_{r=1}^R \overline{fac}_r \sum_{k=1}^m \omega_k}, \quad (1.23)$$

kde R je počet uvažovaných proměnných. Bazická hodnota odpovídá jednotnému nettopojistnému vydělenému průměrnými faktory.

Nyní již máme všechny podklady ke stanovení nettopojistného IIP pro klienta s konkrétními kategoriemi R proměnných. Výpočet nettopojistného zapíšeme

$${}^IIP = P_{base} \cdot \prod_{r=1}^R fac_{r,r_i}. \quad (1.24)$$

Dále přejdeme ke stanovení nettopojistného pomocí frekvence a severity. Nettopojistné spočteme jako součin frekvence a severity. To vysvětlíme tím, že frekvence nám říká, kolik průměrně škod očekáváme u daného klienta a severity vyjadřuje průměrnou velikost jedné škody.

V případě nejjednoduššího přístupu k výpočtu frekvence, severity a nettopojistného dokonce platí rovnost

$${}^IP = {}^If \cdot {}^Iz, \quad (1.25)$$

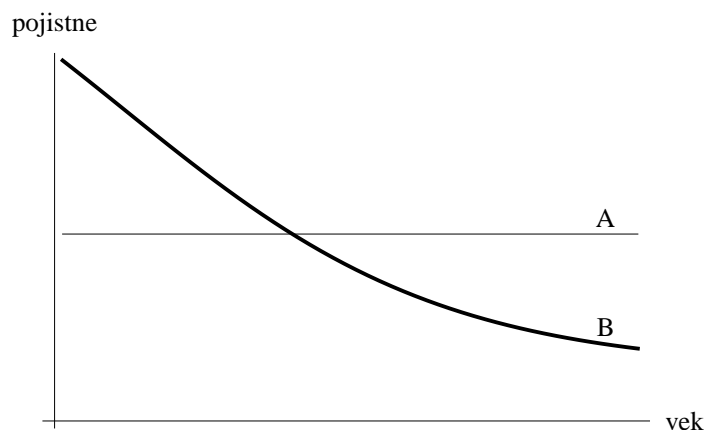
což je snadno vidět z dosazení podle (1.18), (1.3) a (1.9):

$$\frac{\sum_{k=1}^m z_k}{\sum_{k=1}^m \omega_k} = \frac{N}{\sum_{k=1}^m \omega_k} \cdot \frac{\sum_{k=1}^m z_k}{N}. \quad (1.26)$$

Jde tedy pořád o jednotnou sazbu pojistného a takový přístup k pojistnému je velmi nevýhodný.

Představme si situaci, kdy se pojišťovna A rozhodne použít tento přístup a bude od všech svých klientů vyžadovat stejné pojistné za pojištění odpovědnosti za škody způsobené provozem motorového vozidla. Dále uvažujme pojišťovnu

B, která využije byt' jen jediné proměnné související s různou frekvencí škod. Předpokládejme, že touto proměnnou je věk klientů. Pojišťovna B využije toho, že s rostoucím věkem klientů zpravidla roste řidičská zkušenost a opatrnost a tedy klesá škodní frekvence. Ve výši pojistného jsou starší klienti u pojišťovny B zvýhodněni, oproti tomu po mladších klientech je vyžadováno vyšší pojistné. Obrázek 1.1 ilustruje popisovanou skutečnost.



Obrázek 1.1: Jednotná sazba versus zvýhodnění určité skupiny klientů.

Předpokládejme, že podmínky obou pojišťoven jsou stejné, jediná odlišnost je ve výše popsaném pojistném. Dá se očekávat, že starší klienti pojišťovny A budou přecházet do pojišťovny B kvůli levnějšímu pojistnému, naopak mladší klienti pojišťovny B budou přecházet do pojišťovny A ze stejného důvodu. Pokud bude chování řidičů nadále zůstávat bez zásadních změn („mladší bourají častěji“), pojišťovna A nejspíše brzy nebude schopna plnit své závazky, protože nevybírá dostatečně vysoké pojistné.

Vhodnější než jednotné pojistné $^I P$ je přístup, kdy si smlouvy vhodným způsobem kategorizujeme a pro rizikovější kategorie stanovíme vyšší pojistné a naopak pro smlouvy z méně rizikových skupin stanovíme nižší pojistné.

V kapitolách 3 a 4 vyzkoušíme různé přístupy na konkrétních datech a porovnáme.

1.2.4 Bruttipojistné

Nettopojistné je stanovováno tak, aby v průměru pokrývalo pojistná plnění. Škody, které reálně nastanou se však mohou od očekávané hodnoty odchýlit, proto je nutné přidat k nettopojistnému složku, která by pokryla takovýto nepříznivý vývoj. Jak již bylo uvedeno v části 1.1 obecně pojednávající o neživotním pojištění, této složce říkáme bezpečnostní přírážka a nettopojistnému navýšenému o bezpečnostní přírážku říkáme rizikové pojistné. Dále se podíváme na jednotlivé náklady ve vztahu (1.1), které musí pojišťovna (pojistitel) započítat do výsledného bruttopojistného.

Začneme náklady nezávisujícími na výši pojistného — ty lze rozdělit na fixní a variabilní. U fixních nákladů nezáleží na počtu uzavřených smluv, ani na sjednané výši pojistného, k těmto nákladům zařadíme například nájem poboček a náklady

související s jejich provozem. Oproti tomu variabilní náklady souvisí s počtem uzavřených smluv, jde např. o náklady na tisk pojistek, poštovné za dopisy klientům apod.

Typickými náklady závisujícími na výši pojistného jsou provize. Dále je tímto nákladem příspěvek do fondu zábrany škod, který činí nejméně 3 % z ročního přijatého pojistného z pojištění odpovědnosti, tento příspěvek je zaveden zákonem č. 160/2013 Sb. s účinností od roku 2014.

Další položkou, která závisí na výši pojistného, je zisk pojišťovny.

Ve vzorci, který bude následovat, vyznačíme příspěvek do fondu zábrany škod zvlášť.

Vztah pro bruttopojistné B , tedy pro pojistné, které budeme skutečně požadovat od klienta můžeme zapsat

$$B = \frac{RP + náklady_{n,fix} + náklady_{n,var}}{1 - náklady_z - h - z}, \quad (1.27)$$

kde $náklady_{n,fix}$ a $náklady_{n,var}$ značí náklady fixní a variabilní nezávislé na výši pojistného, h je část pojistného symbolizující příspěvek do fondu zábrany škod, z označuje zisk jako část pojistného, jinak je značení stejné jako v (1.1).

1.3 Kredibilita

Teorie kredibility³ se zabývá konstrukcí pojistného na základě předchozích zkušeností jednotlivých tarifních skupin. Pokud jde o kombinaci pojistného zkonstruovaného na základě vlastních minulých údajů pojistitele ($P_{vlastní}$) a pojistného zkonstruovaného na základě globálních minulých údajů ($P_{globální}$; např. údaje za celý pojistný trh), pak mluvíme o tzv. americké kredibilitě. Kredibilitní pojistné pak má tvar:

$$P_{kredibilitní} = Z \cdot P_{vlastní} + (1 - Z) \cdot P_{globální}, \quad (1.28)$$

kde $0 \leq Z \leq 1$ je kredibilitní koeficient (násobitel), který udává váhu vlastních údajů o minulém škodním průběhu. Kredibilitní koeficient by měl růst s rostoucím objemem vlastních dat pojišťovny (pokud jsou ostatní podmínky neměnné) a na druhou stranu by měl kredibilitní koeficient klesat s rostoucím významem vedlejších dat (pokud jsou ostatní podmínky neměnné). Než si ukážeme, jak je možné určit vhodnou váhu Z , zavedeme pojem plné kredibility. Uvažujme odhad $\hat{\theta}$ parametru θ , kde parametrem θ může být některý z ukazatelů pro výpočet pojistného, např. škodní průběh, škodní frekvence, škodní stupeň apod. Plná kredibilita pro odhad parametru θ nastává, když chyba odhadu nepřekročí dané procento k z θ s pravděpodobností alespoň $1 - \alpha$:

$$P \left(|\theta - \hat{\theta}| \leq k \cdot \theta \right) \geq 1 - \alpha.$$

³V této části vycházíme z [2].

Používanou formulí pro stanovení kredibilitního koeficientu je

$$Z = \min \left(\sqrt{\frac{N}{n_f}}, 1 \right), \quad (1.29)$$

kde N je počet pojistných událostí jako rozsah vlastního datového souboru pojišťovny a n_f je počet pojistných událostí potřebných pro dosažení plné kredibility.

Kapitola 2

Zobecněný lineární model

Jak název napovídá, jde o rozšíření klasického lineárního modelu. V následujícím textu popíšeme zobecněný lineární model, odhad jeho parametrů modelu a kritéria pro posouzení kvality tohoto modelu. Čerpáme především z [4].

2.1 Základní prvky

Nejprve si stručně shrňme klasický lineární model. U něj předpokládáme vektor $\mathbf{Y}' = (Y_1, \dots, Y_n)$ nezávislých náhodných veličin s normálním rozdělením s konstantním rozptylem σ^2 a středními hodnotami $\boldsymbol{\mu}' = (\mu_1, \dots, \mu_n)$. Dále předpokládáme, že n -složkový vektor pozorování \mathbf{y} je realizací náhodného vektoru \mathbf{Y} a je vektorem vysvětlovaných proměnných.

Uvažujme dále matici vysvětlujících proměnných \mathbb{X} , kde rozměry této matice jsou $n \times p$, a vektor neznámých parametrů $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$. Systematickou část lineárního modelu pak lze zapsat ve tvaru

$$E(Y_i) = \mu_i = \sum_{j=1}^p x_{ij}\beta_j, \quad i = 1, \dots, n, \quad (2.1)$$

kde x_{ij} představuje hodnotu j -té proměnné i -tého pozorování. Hodnoty parametrů β_1, \dots, β_p , které obvykle nejsou známy, je nutné odhadnout z dat.

U náhodné složky modelu předpokládáme nezávislost a konstantní rozptyl chyb.

Aby bylo následně názorně vidět, v čem spočívá zobecnění lineárního modelu, zavedeme nyní specifikaci lineárního modelu o 3 složkách:

1. náhodnou složkou je vektor nezávislých náhodných veličin \mathbf{Y} , kde náhodná veličina Y_i má normální rozdělení se střední hodnotou μ_i , $i = 1, \dots, n$, a rozptylem σ^2 , vektorový zápis: $E\mathbf{Y} = \boldsymbol{\mu}$;

2. systematickou složkou je *lineární prediktor* $\boldsymbol{\eta}$ tvořený jednotlivými proměnnými $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$

$$\boldsymbol{\eta} = \sum_{j=1}^p \mathbf{x}_j\beta_j; \quad (2.2)$$

3. propojení náhodné a systematické složky:

$$\boldsymbol{\mu} = \boldsymbol{\eta}. \quad (2.3)$$

V uvedené charakterizaci lineárního modelu figuruje nový symbol $\boldsymbol{\eta}$ pro lineární prediktor, dále pak tvrdíme, že $\boldsymbol{\mu}$ a $\boldsymbol{\eta}$ jsou si rovny. Názorněji bychom však měli ve 3. bodě psát

$$\eta_i = g(\mu_i), \quad (2.4)$$

kde g označuje *linkovou funkci* (z anglického *link function*), která je v případě lineárního modelu funkcí identickou.

Nyní přejdeme k zobecnění lineárního modelu — to spočívá v rozšíření bodu 1. a 3. V 1. bodě u složek náhodného vektoru \mathbf{Y} předpokládáme pravděpodobnostní rozdělení z rodiny exponenciálních rozdělení. Ve 3. bodě pak může linkovou funkcí g být jakákoliv monotónní diferencovatelná funkce.

Podívejme se blíže na pravděpodobnostní rozdělení náhodné veličiny Y_i , to je z rodiny exponenciálních rozdělení. Do této rodiny řadíme např. tato rozdělení: normální, Poissonovo, alternativní, gama, exponenciální, geometrické a další. Pro jednoduchost budeme dále namísto Y_i psát pouze Y . Hustotu rozdělení pravděpodobností náhodné veličiny Y uvažujeme ve tvaru

$$f(y, \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}, \quad y \in \mathbb{R}, \quad (2.5)$$

kde a, b, c jsou známé funkce, θ je neznámý kanonický parametr a ϕ je neznámý disperzní parametr.

Pro výpočet střední hodnoty a rozptylu náhodné veličiny Y , jejíž rozdělení pochází z rodiny exponenciálních rozdělení, předpokládejme, že funkce b je dvakrát spojitě diferencovatelná. Parciální derivace $f(y, \theta, \phi)$ podle θ má tvar

$$\frac{\partial f(y, \theta, \phi)}{\partial \theta} = f(y, \theta, \phi) \frac{y - b'(\theta)}{a(\phi)}.$$

Zintegrováním parciální derivace $\frac{\partial f(y, \theta, \phi)}{\partial \theta}$ podle y za předpokladu, že lze zaměnit pořadí integrálu a derivace, dostáváme

$$0 = \frac{\mathbf{E}Y - b'(\theta)}{a(\phi)},$$

odsud máme střední hodnotu Y : $\mathbf{E}Y = b'(\theta)$.

Pro odvození rozptylu náhodné veličiny Y si spočteme druhou parciální derivaci $f(y, \theta, \phi)$ podle θ :

$$\frac{\partial^2 f(y, \theta, \phi)}{\partial \theta^2} = f(y, \theta, \phi) \left[\left(\frac{y - b'(\theta)}{a(\phi)} \right)^2 - \frac{b''(\theta)}{a(\phi)} \right],$$

tuto parciální derivaci zintegrujeme podle y , opět předpokládáme záměnnost integrálu a derivace, a máme

$$0 = \frac{\mathbf{E}(Y - \mathbf{E}Y)^2}{(a(\phi))^2} - \frac{b''(\theta)}{a(\phi)}.$$

Odsud již dostáváme vztah pro rozptyl $\text{var}Y = a(\phi) \cdot b''(\theta)$. Rozptyl můžeme zapsat pomocí *rozptylové funkce* $V(\mu) = b''[(b')^{-1}(\mu)]$ takto: $\text{var}Y = a(\phi) \cdot V(\mu)$.

Funkce $a(\phi)$ běžně bývá ve tvaru

$$a(\phi) = \frac{\phi}{w}, \quad (2.6)$$

kde ϕ je disperzní parametr, jenž je stejný pro všechna pozorování, a w je váha pozorování (jednotlivým pozorováním můžeme přiřadit různou důležitost, pokud to dává smysl).

Přejdeme k linkové funkci g — jak plyne z již uvedeného vztahu (2.4), linková funkce dává do souvislosti lineární prediktor a očekávanou hodnotu sledované náhodné veličiny. Jako příklad linkové funkce si uveďme identitu ($g(\mu) = \mu$), logaritmus ($g(\mu) = \log(\mu)$), logit ($g(\mu) = \log(\mu/(1-\mu))$) a probit ($g(\mu) = \Phi^{-1}(\mu)$, kde Φ je distribuční funkce standardního normálního rozdělení $N(0, 1)$). V případě, že pro linkovou funkci platí $g(\mu_i) = \theta$, jedná se o kanonickou linkovou funkci (kanonický link).

2.2 Odhad parametrů

Pro odhad neznámých parametrů β v lineárním prediktoru se typicky používá metoda maximální věrohodnosti. Parametry odhadujeme pomocí n nezávislých pozorování, tedy předpokládáme, že pro $i = 1, \dots, n$ máme pozorování x_{ij} ($j = 1, \dots, p$) a y_i , kde y_i je realizací náhodné veličiny Y_i , jejíž rozdělení pochází z rodiny exponenciálních rozdělení.

Věrohodnostní funkci jako funkci odhadovaného vektoru β mějme ve tvaru:

$$L(\beta) = \prod_{i=1}^n f(y_i, \theta_i, \phi), \quad (2.7)$$

kde $f(y_i, \theta_i, \phi)$ značí rozdělení náhodné veličiny Y_i (rozdělení viz (2.5)). Logaritmickou věrohodnostní funkci pak dostáváme ve tvaru

$$l(\beta) = \log(L(\beta)) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi). \quad (2.8)$$

Maximum funkce $l(\beta)$ vzhledem k β hledáme tak, že položíme parciální derivaci podle β_j rovnu nule, $j = 1, \dots, p$. Výpočet parciální derivace provádíme pomocí řetízkového pravidla a s postupným využitím vztahů $\theta_i = (b')^{-1}(\mu_i)$, $\mu_i = g^{-1}(\eta_i)$, $\eta_i = x_{ij}\beta_j$:

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi)} \cdot \frac{1}{V(\mu_i)} \frac{1}{g'(\mu_i)} \cdot x_{ij} = 0, \quad j = 1, \dots, p, \quad (2.9)$$

kde $a(\phi)$ uvažujeme ve tvaru ϕ/w_i .

Pro numerické řešení rovnic (2.9) je možné použít metodu iterativně převažovaných nejmenších čtverců — viz například [4] (část 2.5), nebo Newton-Raphsonův algoritmus — viz například [5] (část 3.2.3), případně jiné numerické metody.

2.3 Kvalita modelu

Při výstavbě modelu se snažíme co nejpřesněji vystihnout pozorovaná data ($\mathbf{y}' = (y_1, \dots, y_n)$), finální model nám pak dává odhady pozorovaných hodnot ($\hat{\boldsymbol{\mu}}' = (\hat{\mu}_1, \dots, \hat{\mu}_n)$). Tyto odhady se obecně neshodují s pozorovanými daty a tento nesoulad (chyba) nás zajímá.

Uvažujme dva extrémní případy: jako první si vezmeme tzv. „null model“ (nulový model), to je model, kdy neuvažujeme žádné proměnné, na kterých by závisela odhadovaná veličina. Jediným parametrem je zde střední hodnota μ pozorovaných hodnot vysvětlované proměnné. Jde o model, kde veškerá variabilita mezi vysvětlovanými proměnnými je na náhodné složce.

Druhým krajním případem máme na mysli tzv. „full model“ (saturovaný model, plný model). V této situaci naopak uvažujeme, že počet parametrů je roven počtu pozorování (n), takový model nepřináší žádné nové informace, pouze zopakuje vstupní data (odhady se shodují s pozorovanými hodnotami). Veškerá variabilita mezi vysvětlovanými proměnnými je na systematické složce.

Je třeba uvažovat model, který nebude ani příliš jednoduchý („null“), ani neinformativní („full“). Přesto nám plný model bude užitečný — poslouží k posouzení kvality uvažovaného modelu, u kterého předpokládáme $p < n$ parametrů. Logaritmičká věrohodnost (maximalizovaná přes $\boldsymbol{\beta}$ pro pevně daný parametr ϕ) dosahuje u plného modelu maximální možné hodnoty, této skutečnosti se využívá při srovnávání logaritmičké věrohodnosti modelu s p parametry.

Označme $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\hat{\boldsymbol{\mu}})$ odhad kanonického parametru modelem s p parametry a dále označme $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}(\mathbf{y})$ odhad kanonického parametru plným modelem, zde totiž máme: $\hat{\boldsymbol{\mu}} = \mathbf{y}$. Potom pro odchylku $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$ odhadů od pozorovaných hodnot nazývanou *deviance* platí:

$$\begin{aligned} \frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{\phi} &= \sum_{i=1}^n 2 \cdot \left[\frac{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)}{\frac{\phi}{w_i}} - \frac{y_i \hat{\theta}_i - b(\hat{\theta}_i)}{\frac{\phi}{w_i}} \right] = \\ &= \sum_{i=1}^n \frac{2w_i}{\phi} \left[y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right]. \end{aligned} \quad (2.10)$$

Poznamenejme, že

$$D^*(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{\phi} \quad (2.11)$$

je nazývána *škálovanou deviancí*.

Devianci jsme zavedli s cílem posoudit kvalitu modelu. Čím menší deviance u modelu dosáhneme, tím je model lepší (a tím je věrohodnost modelu blíže maximální dosažitelné věrohodnosti).

Dalším nástrojem pro posouzení kvality modelu je *Akaikeho informační kritérium* (AIC). Vychází z logaritmické věrohodnosti modelu, navíc je však do tohoto kritéria zahrnut i počet proměnných p . Vztah pro výpočet kritéria:

$$AIC = -2l + 2p, \quad (2.12)$$

kde l označuje logaritmickou věrohodnost vytvořeného modelu s p parametry. Při srovnávání více modelů pro pozorovaná data volíme model s co nejmenší hodnotou AIC.

Zmíníme si dále *McFaddenův index podílu věrohodností*, z anglického McFadden's likelihood-ratio index, zkráceně likelihood-ratio index (LRI). Jde o obecnější verzi koeficientu determinace R^2 známého z lineární regrese:

$$LRI = 1 - \frac{l}{l_0}, \quad (2.13)$$

kde l označuje logaritmickou věrohodnost vytvořeného modelu s p parametry a l_0 logaritmickou věrohodnost nulového modelu.

Poslední způsob sloužící k ohodnocení modelu, který si zde uvedeme¹, je *Pearsonův chí-kvadrát*, z anglického Pearson's Chi-Square (Pearson's χ^2 statistic). Hodnotu této statistiky vypočteme následovně:

$$Pearson\chi^2 = \sum_{i=1}^n r_{Pi}^2, \quad (2.14)$$

kde r_{Pi} značí *Pearsonovo reziduum* i -tého pozorování. Výpočet Pearsonova reziduua je následující:

$$r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}, \quad (2.15)$$

kde $\hat{\mu}_i$ je odhad i -tého pozorování y_i a V značí rozptylovou funkci.

2.4 Využití

V této části se zaměříme na to, které konkrétní zobecněné lineární modely nám mohou posloužit při stanovování sazeb pojistného (již bylo naznačeno v první kapitole).

Nejprve se podíváme na zobecněný lineární model vhodný pro modelování počtu pojistných událostí za jeden rok. Pro tento případ předpokládejme, že náhodná veličina Y (počet pojistných událostí za jeden rok neboli frekvence) má Poissonovo rozdělení s parametrem ϑ , pravděpodobnostní rozdělení viz (1.6). Nyní přepíšeme rozdělení pravděpodobností do tvaru (2.5):

¹Nástrojů pro ohodnocení kvality modelu existuje celá řada, shrnutí nabízí například J. W. Hardin, J. M. Hilbe v publikaci *Generalized Linear Models and Extensions*, Second Edition.

$$\begin{aligned}
f(y, \vartheta) &= \frac{\vartheta^y}{y!} e^{-\vartheta} = \exp \{y \log \vartheta - \log y! - \vartheta\} = \\
&= \exp \left\{ \frac{y \log \vartheta - \vartheta}{1} - \log y! \right\}.
\end{aligned} \tag{2.16}$$

Odsud dostáváme kanonický a disperzní parametr:

$$\theta = \log \vartheta, \quad \phi = 1.$$

Dále je vidět funkce b :

$$b(\theta) = e^\theta. \tag{2.17}$$

Střední hodnota a rozptyl náhodné veličiny Y je:

$$EY = b'(\theta) = e^\theta = \vartheta, \quad \text{var}Y = a(\phi) \cdot b''(\theta) = e^\theta = \vartheta.$$

Kanonickým linkem je v tomto případě logaritmus, protože jen pro něj zde platí $g(\mu) = \theta$.

Tento model využijeme v praktické části práce pro modelování počtu škod na smlouvě za jeden rok.

Dále se podíváme na zobecněný lineární model vhodný pro modelování výše škody. V tomto případě předpokládáme náhodnou veličinu Y (výše škody) s gama rozdělením s parametry α a p , viz (1.15). Hustotu rozdělení pravděpodobností náhodné veličiny Y vyjádříme ve tvaru (2.5):

$$\begin{aligned}
f(y, \alpha, p) &= \frac{\alpha^p}{\Gamma(p)} y^{(p-1)} e^{-\alpha y} = \\
&= \exp \{p \log \alpha - \log(\Gamma(p)) + (p-1) \log y - \alpha y\} = \\
&= \exp \left\{ \frac{y(-\frac{\alpha}{p}) - (-\log \frac{\alpha}{p})}{\frac{1}{p}} + p \log p + (p-1) \log y - \log(\Gamma(p)) \right\}.
\end{aligned} \tag{2.18}$$

Z tohoto zápisu vyčteme kanonický a disperzní parametr:

$$\theta = -\frac{\alpha}{p}, \quad \phi = \frac{1}{p}.$$

Funkce b je ve tvaru:

$$b(\theta) = -\log(-\theta). \tag{2.19}$$

Střední hodnotu a rozptyl náhodné veličiny Y lze zapsat takto:

$$EY = b'(\theta) = -\frac{1}{\theta} = \frac{p}{\alpha}, \quad \text{var}Y = a(\phi) \cdot b''(\theta) = \frac{1}{p} \frac{p^2}{\alpha^2} = \frac{p}{\alpha^2}.$$

V tomto modelu je kanonický link $g(\mu) = -\frac{1}{\mu}$.

Zobecněný lineární model, kde předpokládáme u vysvětlované náhodné veličiny gama rozdělení, využijeme v kapitole 3 k modelování výše škody.

Dva zobecněné lineární modely jsme si rozebrali podrobněji. Následující tabulka 2.1 shrnuje vlastnosti i některých dalších rozdělení z rodiny exponenciálních rozdělení.

rozdělení	$N(\mu, \sigma^2)$	$Po(\vartheta)$	$Alt(\mu)$	$\Gamma(\alpha, p)$
hustota	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$	$\frac{\vartheta^y e^{-\vartheta}}{y!}$	$\mu^y (1-\mu)^{1-y}$	$\frac{\alpha^p}{\Gamma(p)} y^{(p-1)} e^{-\alpha y}$
ϕ	σ^2	1	1	$\frac{1}{p}$
$b(\theta)$	$\theta^2/2$	e^θ	$\log(1+e^\theta)$	$-\log(-\theta)$
$\mu(\theta)$	θ	e^θ	$\frac{e^\theta}{1+e^\theta}$	$-\frac{1}{\theta}$
kanonický link	μ	$\log \mu$	$\log\left(\frac{\mu}{1-\mu}\right)$	$-\frac{1}{\mu}$
$V(\mu)$	1	μ	$\mu(1-\mu)$	μ^2

Tabulka 2.1: Charakteristiky některých rozdělení související se zobecněným lineárním modelem.

Kapitola 3

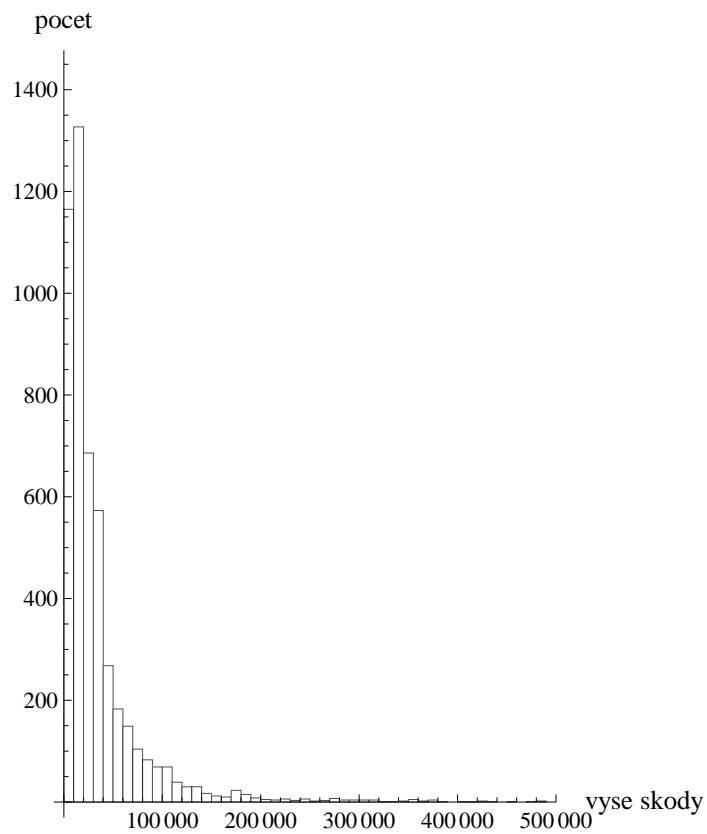
Malé škody

V následujících kapitolách si na konkrétních datech ukážeme dříve popsané různé přístupy k výpočtu pojistného, resp. frekvence škod a výše škod. Data pro zpracování poskytl vedoucí diplomové práce. Jde o reálná data z jedné pojišťovny, která jsou upravena tak, aby nebylo možné jednotlivé klienty/vozidla identifikovat (např. je z dat odstraněno jméno klienta, rodné číslo, registrační značka, apod.). V souboru smlouvy.txt (na přiloženém CD ve složce data) máme k dispozici údaje o 146 768 smlouvách, jsou to např. informace o počátku a konci pojištění, zda jde o fyzickou či právnickou osobu, stáří vozu, stáří pojistníka, trvalé bydliště pojistníka a počet obyvatel této obce, výkon motoru v kilowattech, objemová třída vozidla a další. Všechny smlouvy mají datum počátku pojištění 1. 1. 2011 a konec pojištění nejpozději 31. 12. 2011. Dále máme v souboru skody.txt (na přiloženém CD ve složce data) informace o 5 187 škodách, které nastaly během jednoho roku na smlouvách ze souboru smlouvy.txt. O každé škodě víme, na které smlouvě se stala a jak vysoké pojistné plnění bylo vyplaceno.

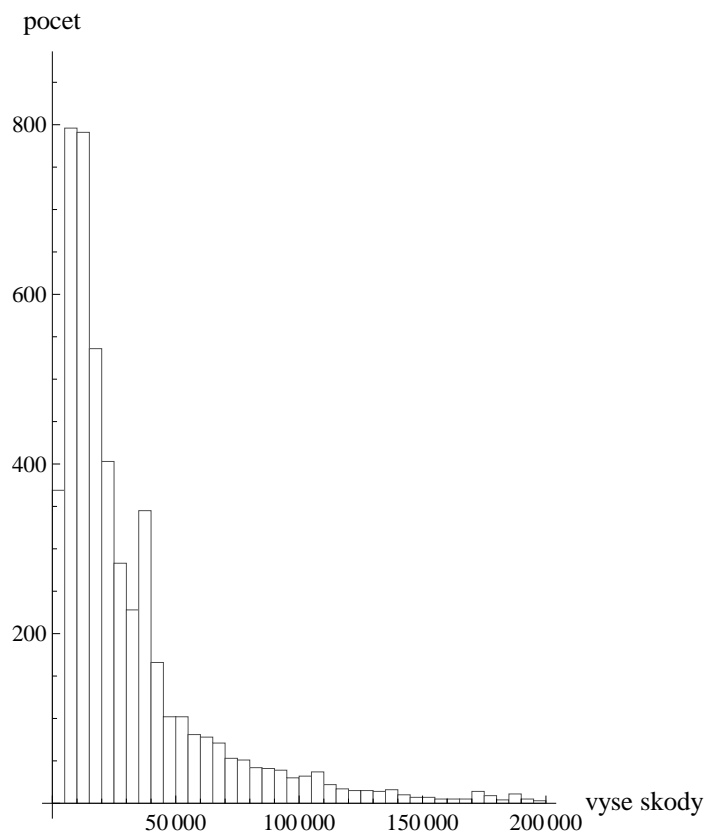
Při zpracování dat se soustředíme zvláště na tzv. malé a velké škody. Volíme tento přístup, protože se chceme vyhnout situaci, kdy jedna velká škoda ovlivní danou kategorii smluv. (Např. smlouvy spadající do určité kategorie mají největší počet škod v rozmezí 50 000–100 000 Kč, ale jedna smlouva v této kategorii má škodu za 10 000 000 Kč. Kvůli jedné extrémně vysoké škodě by byla celá kategorie považována za více rizikovou.) Nejdříve si tedy určíme hranici, podle které škodu přiřadíme buď k malým či velkým škodám. Navíc od hranice požadujeme, aby počet velkých škod pomocí ní určených byl dostatečně velký pro samostatnou analýzu.

V souboru skody.txt se vyskytují i pojistná plnění v nulové výši — tyto škody vyřadíme a nebudeme dále uvažovat, protože nemají vliv na stanovení nettopojistného, kdy se soustředíme právě na pokrytí pojistných plnění. Pravděpodobně jde o škody, které byly nahlášené, ale pojistné plnění je nulové, případně to můžou být chybně registrované škody. Naopak při výpočtu bruttopojistného bychom je měli do výpočtu zahrnout, protože pojišťovna vynaloží náklady na likvidaci těchto škod.

Po vyřazení nulových škod nám zůstává 4 978 škod. Nejmenší škoda byla ve výši 210 Kč, nejvyšší škoda dosáhla výše 13 856 463 Kč, průměrná škoda je za 47 983 Kč, medián má hodnotu 19 988 Kč. Podívejme se na rozložení škod do 500 000 Kč — to je znázorněno histogramem na obrázku 3.1. Nejvíce škod (4 607) je do výše 100 000 Kč (viz podrobnější histogram na obrázku 3.2), naopak škody



Obrázek 3.1: Histogram výší škod do 500 000 Kč.



Obrázek 3.2: Histogram výší škod do 200 000 Kč.

nad 200 000 Kč už nejsou tak časté (118 škod větších než 200 000 Kč). Hranici pro rozdělení malých a velkých škod proto zvolíme v intervalu (100 000, 200 000). Zvolme si hranici 150 000 Kč, počet škod v intervalu (100 000, 150 000] je 185, oproti tomu počet škod v intervalu (150 000, 200 000] je už jen 68. Přesný počet škod nad 150 000 Kč je 186, což je dostatečné množství pro zkoumání rozložení vysokých škod. Tato hranice splňuje výše uvedená kritéria a proto budeme dále považovat škody rovny nejvýše 150 000 Kč za malé a škody větší než 150 000 Kč budeme považovat za velké. V této kapitole se budeme věnovat malým škodám, o velkých škodách pojednává kapitola 4.

3.1 Škodní frekvence

Nejdříve se podíváme na nejjednodušší způsob výpočtu frekvence malých škod, resp. počtu malých škod na smlouvě za rok. Frekvenci vypočteme podle vztahu (1.3) pouze jako podíl celkového počtu malých škod a celkové expozice, pro daná data tedy máme:

$$I_f \doteq \frac{4\,792}{127\,271,39} \doteq 0,0377. \quad (3.1)$$

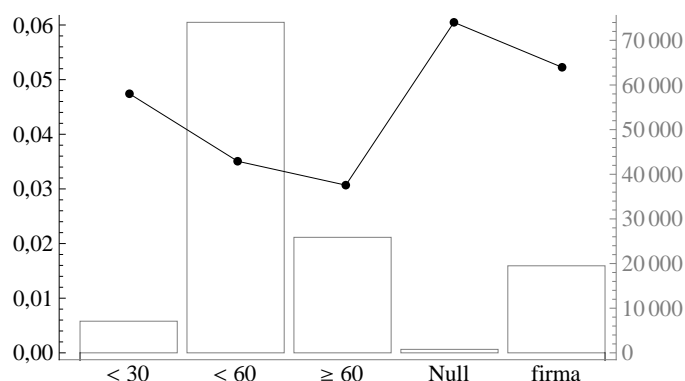
Nyní přistoupíme k výpočtu frekvence škod způsobem, který již využívá kategorizaci smluv podle proměnných (budeme využívat vztah pro frekvenci (1.5)). Budeme uvažovat proměnné dříve v praxi standardně využívané k výpočtu pojistného, resp. frekvence škod. Zaměříme se na proměnnou věk pojistníka (zkráceně budeme proměnnou označovat věk), objem motoru (zkráceně objem) a počet obyvatel obce, kde má pojistník trvalé bydliště (zkráceně město).

Nejprve si vhodným způsobem zvolíme kategorie proměnných. Pokud to bude možné, budeme jich pro jednoduchost a přehlednost volit raději menší počet. Navíc od každé kategorie očekáváme dostatečný počet pozorování.

Proměnnou věk si rozdělíme do tří kategorií podle věku klienta, zkusíme si zvolit kategorie do 29 let, od 30 do 59 let a nad 60 let (včetně šedesátiletých). Navíc však k těmto kategoriím přibudou ještě dvě skupiny klientů — jedna s označením firma a další s označením Null. Kategorie firma obsahuje smlouvy, které mají v původním souboru smlouvy.txt ve sloupci s názvem „druh_osoby“ hodnotu „IC“ (= IČ), jde o firemní vozidla, u nichž neznáme věk řidiče, resp. pojistníka. V kategorii Null jsou zařazeni klienti, kteří nejsou vedeni mezi firmami (může jít o chybu), ale ani nemají zaznamenán věk (resp. rok narození).

Na obrázku 3.3 máme znázorněný graf, kde na svislé ose vlevo máme stupnici pro frekvence škod a svislá osa vpravo je stupnicí pro expozice. Na vodorovné ose jsou jednotlivé kategorie proměnné věk, kde „< 30“ je označení pro věkovou kategorii do 29 let, „< 60“ reprezentuje věkovou kategorii 30–59 let, „≥ 60“ označuje věkovou kategorii nad 60 let, „Null“ a „firma“ postupně značí kategorii Null a firma. Podobným způsobem jsou značeny i další grafy (resp. vodorovné osy grafů) v práci.

Pro každou kategorii máme tečkou vyznačenu frekvenci škod (vypočteno podle (1.4)) a sloupcem celkovou expozici dané kategorie. Frekvence škod jednotlivých



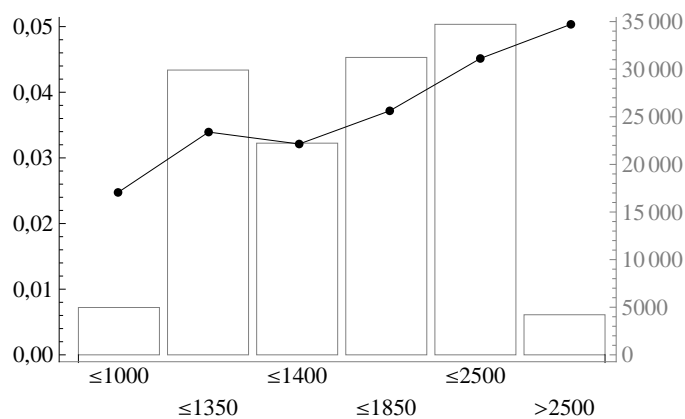
Obrázek 3.3: Frekvence (tečky) a expozice (sloupce) věku v jednotlivých kategoriích.

kategorií jsou navíc spojeny plnou čarou, aby byla naznačena změna oproti sousedním kategoriím. Z obrázku vidíme, že má smysl zavádět takovou kategorizaci, protože tak oddělíme rizikovější skupiny od těch méně rizikových. Vzhledem k poměrně malé expozici kategorie Null a blízkosti její frekvence škod k frekvenci škod kategorie firma, tyto dvě kategorie sloučíme a vzniklou kategorii označíme firma. V rámci proměnné věk budeme tedy dále uvažovat tyto čtyři kategorie: do 29 let, od 30 do 59 let, nad 60 let a firma.

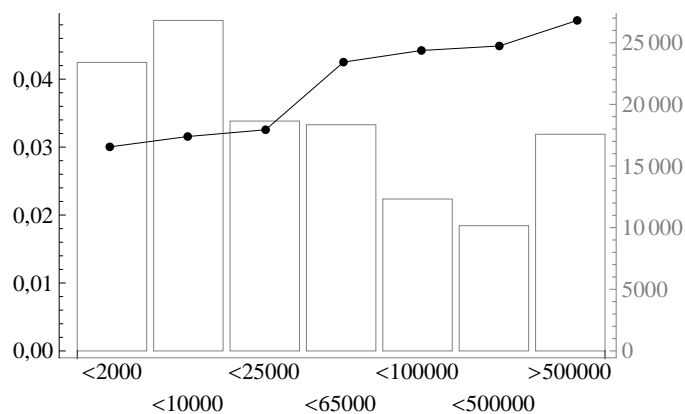
Dále si stanovíme vhodné kategorie pro proměnnou objem. Dříve se při stanovování pojistného (resp. frekvence škod) standardně využívalo těchto 5 kategorií: do 1 000 ccm, 1 001–1 350 ccm, 1 351–1 850 ccm, 1 851–2 500 ccm a nad 2 500 ccm. Vstupní soubor smlouvy.txt nám však nabízí tuto kategorizaci objemu: do 1 000 ccm, 1 001–1 350 ccm, 1 351–1 400 ccm, 1 401–1 850 ccm, 1 851–2 500 ccm a nad 2 500 ccm. Je vidět, že z kategorie 1 351–1 850 ccm se vyčlenila nová kategorie 1 351–1 400 ccm, protože se její frekvence vymyká rostoucímu trendu. Těchto 6 kategorií vidíme na obrázku 3.4, kde frekvence jsou zobrazeny pomocí teček (měřítko — svislá osa vlevo) a expozice pomocí sloupců (měřítko — svislá osa vpravo) pro jednotlivé kategorie (vodorovná osa). Tuto kategorizaci využijeme.

Zbývá si určit vhodné kategorie proměnné město. Zkusíme si nejdřív vytvořit více kategorií a následně sloučíme ty kategorie, které mají podobnou frekvenci škod. Zvolíme si tyto kategorie: počet obyvatel do 1 999, 2 000–9 999, 10 000–24 999, 25 000–64 999, 65 000–99 999, 100 000–499 999, nad 500 000, kde kategorie nad 500 000 obyvatel obsahuje pouze Prahu. Podíváme se na frekvence v jednotlivých kategoriích — ty nám reprezentuje spojnicový graf na obrázku 3.5, měřítkem frekvence je svislá osa vlevo. V grafu máme pomocí sloupců označeny expozice pro jednotlivé kategorie se stupnicí na svislé ose vpravo. Podle grafu si tedy sloučíme první 3 kategorie a čtvrtou až šestou kategorii, Prahu ponecháme odděleně. Takže pro proměnnou město máme kategorie do 24 999 obyvatel, 25 000–499 999 a nad 500 000 obyvatel.

Pro uvažované proměnné (věk, objem, město) máme určeny kategorie, nyní pro každou kategorii všech uvažovaných proměnných spočteme frekvenci škod podle



Obrázek 3.4: Frekvence (tečky) a expozice (sloupce) objemu v jednotlivých kategoriích.



Obrázek 3.5: Frekvence (tečky) a expozice (sloupce) města v jednotlivých kategoriích.

(1.4) (většinu z nich však již máme spočteno). Tyto frekvence spolu s počtem škod a expozicí pro jednotlivé kategorie vybraných proměnných shrnuje tabulka 3.1.

U klienta s konkrétními kategoriemi věku, objemu a města chceme znát odhad počtu škod, které na smlouvě za rok vzniknou. Tuto frekvenci (počet škod na smlouvě za rok) odhadneme podle (1.5) jako geometrický průměr frekvencí jednotlivých kategorií.

Kategorie věku	Počet škod	Součet expozicí	Frekvence
≤ 29	336	7 086,12	0,047 42
30–59	2 596	74 036,02	0,035 06
≥ 60	794	25 873,46	0,030 69
firma	1 066	20 275,80	0,052 58
Kategorie objemu			
$\leq 1\,000$ ccm	123	4 971,98	0,024 74
1 001–1 350 ccm	1 015	29 910,39	0,033 93
1 351–1 400 ccm	714	22 231,43	0,032 12
1 401–1 850 ccm	1 161	31 234,70	0,037 17
1 851–2 500 ccm	1 567	34 711,96	0,045 14
$\geq 2\,501$ ccm	212	4 210,93	0,050 35
Kategorie města			
$\leq 24\,999$	2 156	68 861,65	0,031 31
25 000–499 999	1 781	40 833,63	0,043 62
$\geq 500\,000$	855	17 576,11	0,048 65

Tabulka 3.1: Frekvence pro jednotlivé kategorie.

3.2 Škodní frekvence — GLM

Nyní se ve složitosti stanovení frekvence škod posuneme o kus dál a budeme se zabývat využitím zobecněného lineárního modelu (GLM). Odhadujeme počet škod, které vzniknou na jedné pojistné smlouvě za jeden rok. Pro vysvětlovanou proměnnou volíme z rodiny exponenciálních rozdělení Poissonovo, linkovou funkci volíme logaritmus (tedy kanonický link). Jako vysvětlující proměnné opět vybereme věk, objem a město, navíc však přidáme proměnnou osoba, která nese informaci o tom, zda je klientem žena, muž či právnická osoba.

Kategorií věku volíme více než v předchozí části, abychom lépe vystihli mladší rizikovější kategorie klientů. Máme tyto věkové kategorie: 18 let, 19 let, 20–21, 22–25, 26–29, 30–49, 50–64, 65 a více let a firma. Kategorie objemu využijeme ty, které máme k dispozici v původním souboru smlouvy.txt: objem do 1 000 ccm, 1 001–1 350 ccm, 1 351–1 400 ccm, 1 401–1 850 ccm, 1 851–2 500 ccm a nad 2500 ccm. Proměnnou město kategorizujeme stejně jako v předchozí části na město do 24 999 obyvatel, 25 000–499 999, nad 500 000; proměnná osoba již byla zmíněna — ta má tyto tři kategorie: žena, muž a firma. Do modelu tedy vstupují 4 kategoriální proměnné (věk \mathbf{x}_1 , osoba \mathbf{x}_2 , objem \mathbf{x}_3 , město \mathbf{x}_4), model si označíme „GLM 4“.

V datovém souboru se nacházejí smlouvy, kterým chybí některé údaje — tyto smlouvy vyřadíme, nebudou součástí modelu. Tento krok si můžeme dovolit, protože počet vyřazených smluv je vzhledem k celkovému počtu smluv zanedbatelný.

Zobecněný lineární model vytváříme v programu Wolfram Mathematica 8.0, kde využíváme zabudovanou funkci „GeneralizedLinearModelFit“. V modelu využijeme možnosti přidání vah k jednotlivým pozorováním — těmito váhami budou expozice u jednotlivých smluv.

V příloze A v tabulce A.1 najdeme výstup modelu ze softwaru. Odhady parametrů modelu jsou ve sloupci Estimate. Pro odhad střední hodnoty EY_i počtu

škod na i -té smlouvě za rok máme:

$$\hat{\mu}_i = g^{-1}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}), \quad (3.2)$$

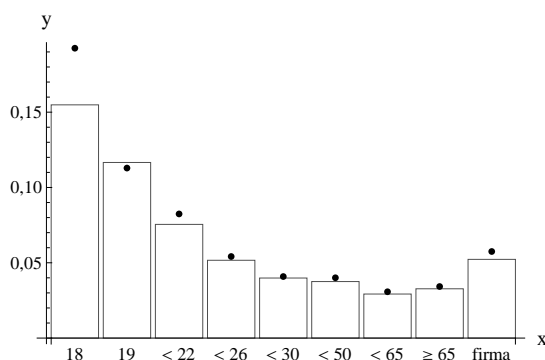
kde β_0 je pevně daný člen lineárního prediktoru (to je hodnota z prvního řádku tabulky, konkrétně v tomto modelu $-2,73281$). Parametr β_1 vybereme ten, který odpovídá kategorii klienta první proměnné, další parametry vybíráme analogicky. Uvedeme si příklad: pro čtyřicetiletého muže z Prahy s vozidlem o objemu nad 2 500 ccm dostáváme:

$$\begin{aligned} \hat{\mu}_i &= \exp\{-2,73281 + (-0,105991) + (-0,226512) + 0 + 0\} = \\ &= \exp\{-3,065313\} \doteq 0,04664. \end{aligned}$$

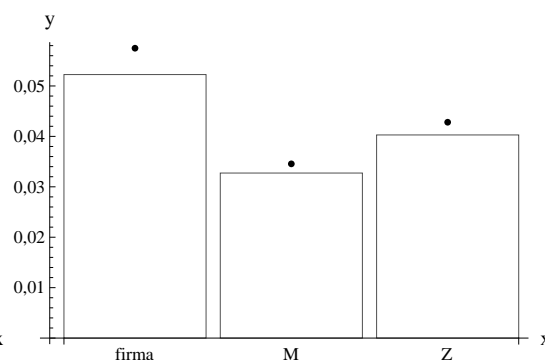
Máme tedy spočtený parametr Poissonova rozdělení, snadno už pak dopočítáme například pravděpodobnost, že daný klient nebude během jednoho roku bourat: $\exp\{-0,04664\} \doteq 0,95443$.

Dále využijeme model „GLM 4“ a pro každou smlouvu si spočteme predikci počtu škod. Pomocí těchto predikcí si spočteme v rámci jedné kategorie dané proměnné podíl součtu predikcí a součtu expozicí. Tento podíl označuje frekvenci odhadnutou modelem „GLM 4“. Porovnáme si tuto frekvenci s frekvencí vypočtenou z dat, což je podíl součtu pozorovaných počtů škod a součtu expozicí, vše v rámci jedné konkrétní kategorie.

Pomocí grafů na obrázcích 3.6, 3.7, 3.8, 3.9 si pro jednotlivé kategorie regresorů modelu „GLM 4“ porovnáme frekvence vypočtené z dat (sloupce) a frekvence odhadnuté modelem (tečky). Na všech obrázcích vidíme, že odhad frekvencí (frekvence jsou vynášeny na osu y) pro jednotlivé kategorie proměnných (kategorie proměnných jsou symbolicky označeny na ose x) se blíží skutečnosti.

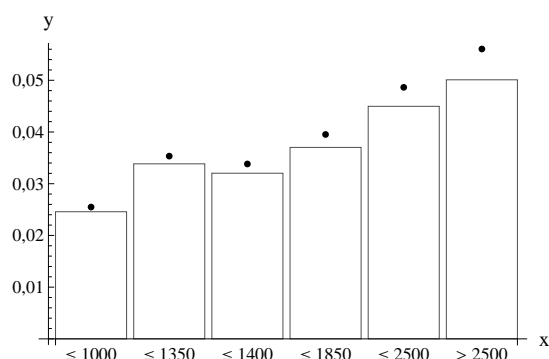


Obrázek 3.6: „GLM 4“, věk

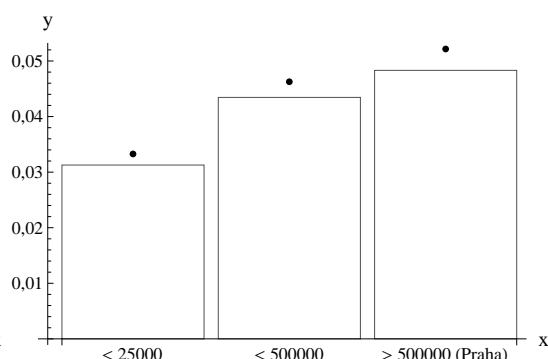


Obrázek 3.7: „GLM 4“, osoba

Zkusme si ale tento model pro proměnnou, která v modelu není zahrnuta. Pro vyzkoušení modelu „GLM 4“ zvolíme proměnné výkon motoru (zkráceně výkon) a hmotnost vozidla (zkráceně hmotnost) a pro každou z nich vytvoříme graf frekvencí (frekvence na ose y) pro jednotlivé kategorie (kategorie na ose x), kde sloupce značí frekvenci vypočtenou z dat a tečky odhadnutou frekvenci na základě modelu. Hodnoty proměnné výkon rozdělíme do těchto kategorií: do 44 kW, 45–49 kW, 50–54 kW, 55–59 kW, 60–64 kW, 65–69 kW, 70–74 kW, 75–79 kW, 80–84 kW, 85–94 kW, 95–104 kW, 105–124 kW, 125–139 kW, 140 kW a více.

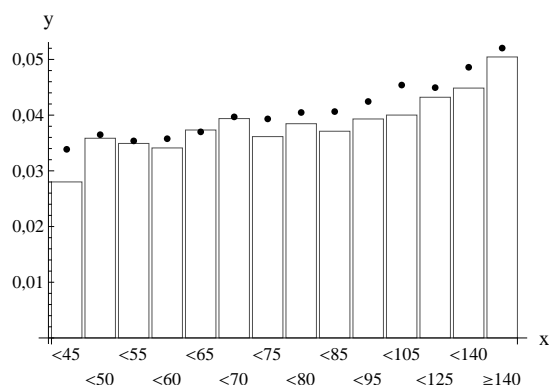


Obrázek 3.8: „GLM 4“, objem

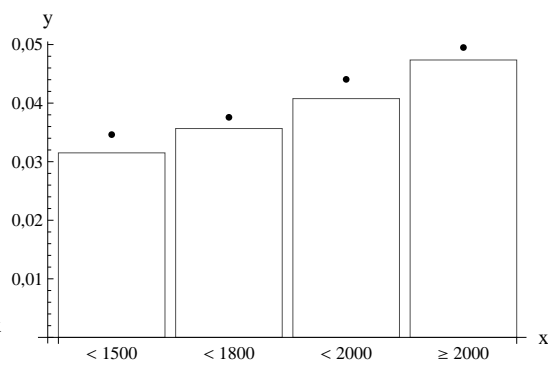


Obrázek 3.9: „GLM 4“, město

Pro hodnoty proměnné hmotnost zvolíme následující kategorizaci: do 1 499 kg, 1 500–1 799 kg, 1 800–1 999 kg, 2 000 kg a více. V grafech na obrázcích 3.10 a 3.11 vidíme, že proměnná výkon není modelem odhadována zcela přesně a u proměnné hmotnost to vypadá, že již je v modelu zahrnuta. Nicméně obě dvě zmiňované proměnné zařadíme do nově vytvořeného modelu.



Obrázek 3.10: „GLM 4“, výkon



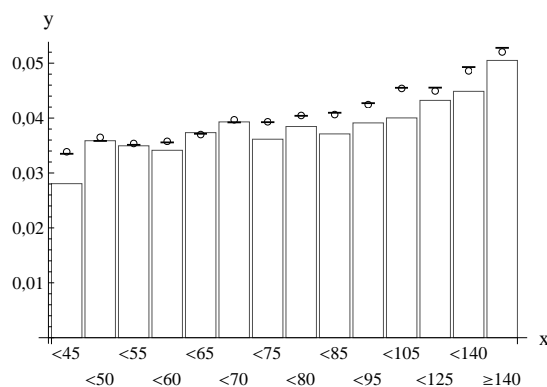
Obrázek 3.11: „GLM 4“, hmotnost

Nyní tedy vytvoříme model „GLM 6“ s těmito vysvětlujícími proměnnými: věk x_1 , osoba x_2 , objem x_3 , výkon x_4 , hmotnost x_5 a město x_6 . Stejně jako v předchozím modelu „GLM 4“ i v tomto jsou proměnné věk, osoba, objem a město kategoriální se stejným rozdělením do kategorií. Nově přidané regresory výkon a hmotnost vstupují do modelu jako spojité proměnné. Jinak se v modelu nic nemění (opět volíme Poissonovo rozdělení, kanonický link; expozice jako váhy pozorování).

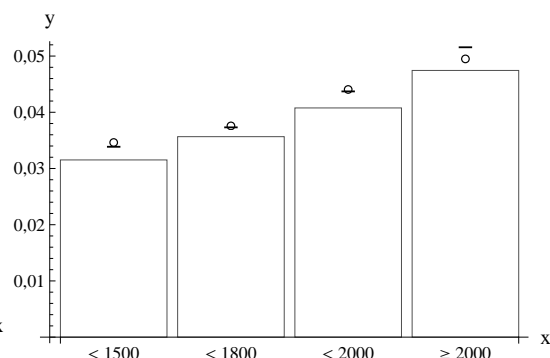
Odhady parametrů modelu jsou v příloze A v tabulce A.2.

V tomto modelu se mj. zaměříme na to, zda dostáváme lepší odhady frekvencí pro proměnné výkon a hmotnost. Na obrázku 3.12 máme frekvence pro proměnnou výkon a na obrázku 3.13 pro proměnnou hmotnost. Frekvence jsou znázorněny pro stejné kategorie proměnných jako v předchozím modelu, sloupce reprezentují frekvence vypočtené z dat, kroužky odhady frekvencí podle modelu „GLM 4“ se čtyřmi proměnnými a vodorovné značky jsou odhady frekvencí podle modelu „GLM 6“ se šesti proměnnými; na ose x leží vždy jednotlivé kategorie proměnné, na ose y jsou frekvence jednotlivých kategorií. Z obrázku není na první pohled zřejmé, zda je model se šesti proměnnými lepší pro odhadování frekvence

škod ve skupinách přidanych proměnných. Proto se zaměříme na změnu frekvencí v jednotlivých kategoriích oproti frekvenci v první kategorii.

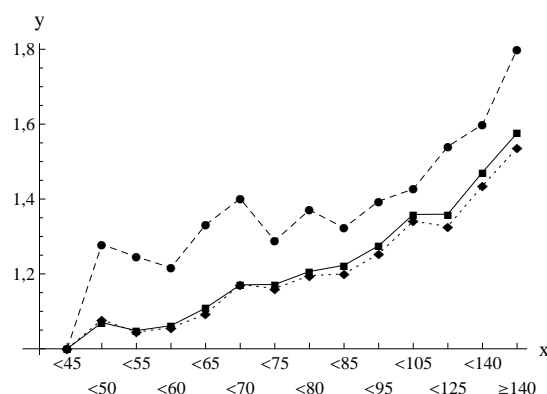


Obrázek 3.12: „GLM 6“, výkon

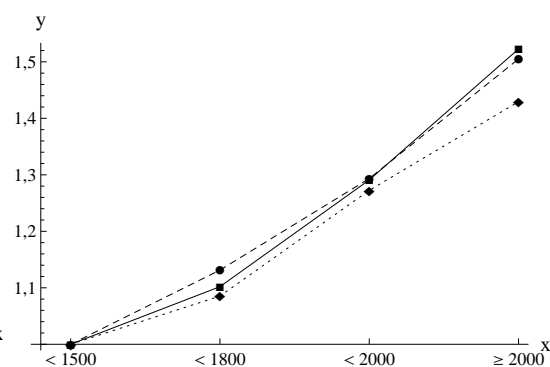


Obrázek 3.13: „GLM 6“, hmotnost

Na obrázcích 3.14 a 3.15 jsou grafy, kde na ose x jsou jednotlivé kategorie zkoumané proměnné a osa y udává poměr frekvence dané kategorie ku frekvenci první kategorie (první kategorie má teda vždy hodnotu 1). Grafy jsou v obrázku vždy zobrazeny tři — čárkovaně značíme spojnice hodnot pro skutečná data, plnou čarou spojnice hodnot vypočtených z modelu se šesti proměnnými („GLM 6“) a konečně tečkovaně spojujeme podíl frekvencí vypočtených z modelu se čtyřmi proměnnými („GLM 4“).



Obrázek 3.14: „GLM 6“, výkon — změna

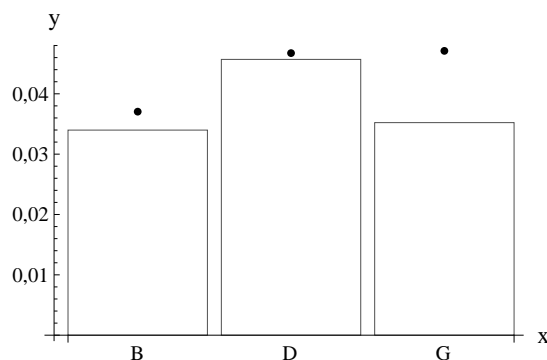


Obrázek 3.15: „GLM 6“, hmotnost — změna

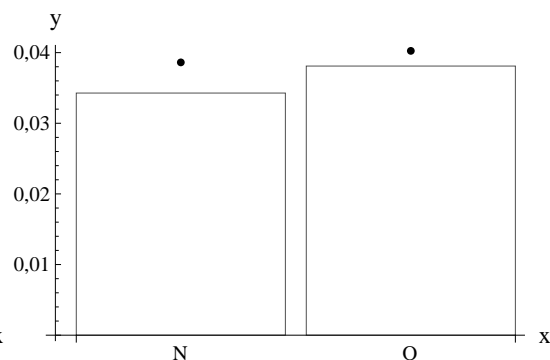
U obou proměnných vidíme, že skutečný nárůst frekvence v jednotlivých kategoriích oproti první kategorii je ve většině případů lépe popsán zobecněným lineárním modelem se šesti proměnnými („GLM 6“). V grafu je tato skutečnost lépe pozorovatelná pro kategorie výkonu od 75 kW a vyšší, pro skupiny hmotnosti od 1 800 kg a vyšší.

Nyní si vyzkoušíme model „GLM 6“ pro další proměnné, které nejsou zahrnuty v modelu jako regresory. Podíváme se na proměnnou palivo a nové vozidlo. Palivo je kategoriální proměnná se třemi skupinami — benzín (značka „B“), diesel (značka „D“) a plyn (značka „G“). Proměnná nové vozidlo má dvě kategorie: „N“ a „O“, kde „N“ (jako „nové“) označuje vozidla stará nejvýše jeden rok a „O“ (jako „ojeté“) symbolizuje vozidla stará dva roky a více.

V grafech si opět na osu x umístíme jednotlivé kategorie dané proměnné a na osu y zaneseme vypočtené frekvence; sloupečky značí skutečnou frekvenci podle datového souboru, tečky reprezentují frekvence odhadnuté modelem „GLM 6“. Graf pro palivo je na obrázku 3.16 a graf pro nové vozidlo je na obrázku 3.17. Vidíme, že odhady podle modelu se odchyľují od skutečných frekvencí v rámci kategorií vybraných proměnných.



Obrázek 3.16: „GLM 6“, palivo

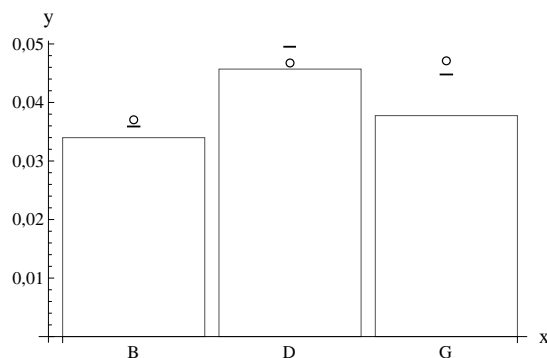


Obrázek 3.17: „GLM 6“, nové vozidlo

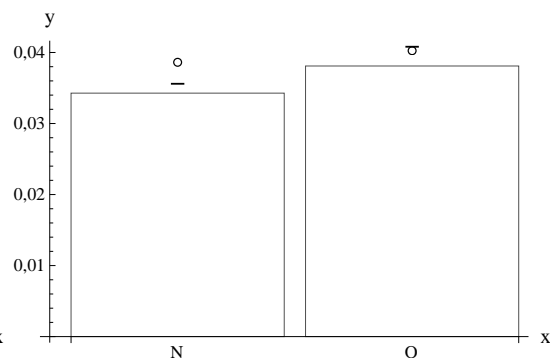
Mějme dále palivo a nové vozidlo jako další dva regresory v zobecněném lineárním modelu. Vytvoříme tedy další model („GLM 8“) s těmito vysvětlujícími proměnnými: věk x_1 , osoba x_2 , objem x_3 , výkon x_4 , hmotnost x_5 , palivo x_6 , město x_7 a nové vozidlo x_8 . Kromě spojitých proměnných výkon a hmotnost jsou všechny ostatní kategoriální, kde kategorie jsou pro příslušné regresory stejné jako v „GLM 6“. Nově přidané palivo a nové vozidlo mají výše zmíněné kategorie, tedy postupně B, D, G a N, O.

Odhady parametrů modelu jsou v příloze A v tabulce A.3.

Podíváme se, jak model „GLM 8“ odhaduje frekvence jednotlivých kategorií nově přidaných proměnných palivo a nové vozidlo a tyto odhady srovnáme s frekvencemi vypočtenými z dat a s odhady pomocí modelu „GLM 6“. Na obrázcích 3.18 a 3.19 máme na ose x kategorie daného regresoru a na ose y frekvence; sloupce značí hodnoty vypočtené z dat, kroužky zastupují odhady z „GLM 6“ a vodorovné značky symbolizují hodnoty odhadnuté pomocí „GLM 8“. Můžeme říci, že zkoumané frekvence jsou v obou případech lépe odhadovány modelem „GLM 8“.

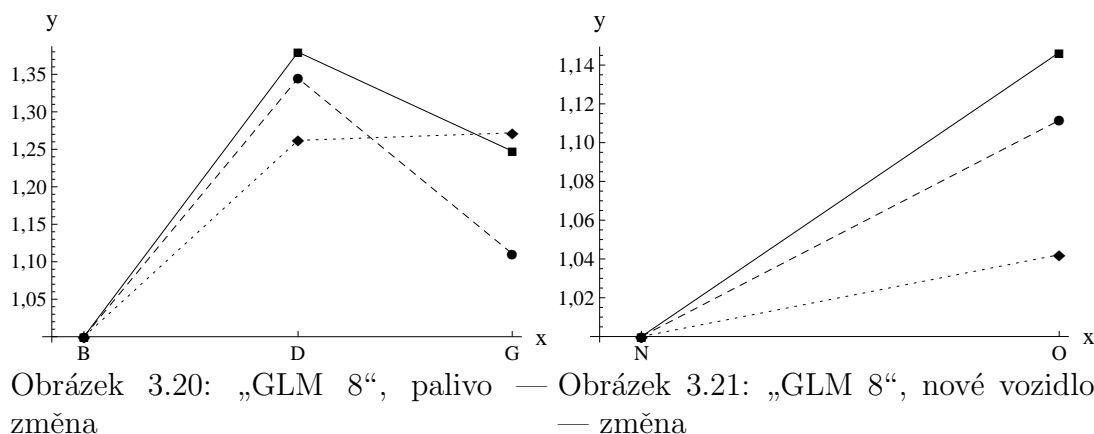


Obrázek 3.18: „GLM 8“, palivo



Obrázek 3.19: „GLM 8“, nové vozidlo

Zaměřme se na změny frekvencí v rámci jednotlivých kategorií u nově přidávaných proměnných — podíváme se na poměr frekvence dané kategorie ku frekvenci kategorie první. Na obrázku 3.20 vidíme pro jednotlivé kategorie regresoru palivo (osa x) změnu frekvence oproti frekvenci v první kategorii (změna frekvence na ose y), kde čárkovaně spojujeme hodnoty pro skutečná data, plnou čárou značíme spojnice hodnot vypočtených z modelu „GLM 8“ a tečkovaně spojnice hodnot vypočtených z modelu se šesti proměnnými („GLM 6“). Obrázek 3.21 bychom popsali analogicky pro proměnnou nové vozidlo.



Model „GLM 8“ ve zkoumaném ohledu v obou případech lépe vystihuje pozorovaná data než model „GLM 6“.

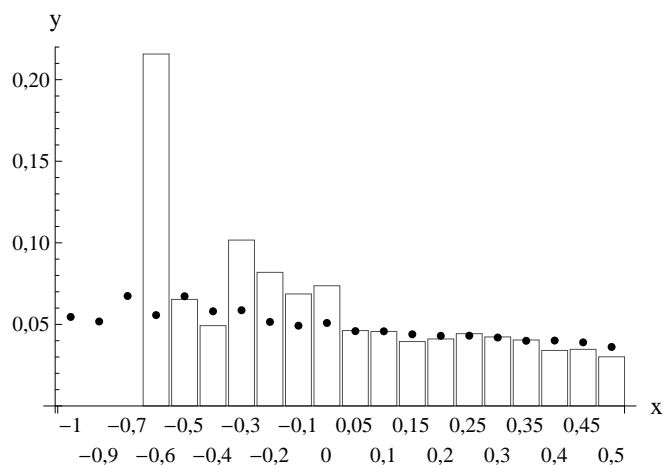
Opět si zkusíme modelem „GLM 8“ odhadnout frekvenci u dalších proměnných, které v modelu zatím nefigurují jako regresory. Zvolíme si proměnné bonus a historie. Bonus vyjadřuje, v jaké bonusové třídě se klient nachází a v datovém souboru nabývá těchto hodnot: -1 ; $-0,9$; $-0,7$; $-0,6$; $-0,5$; $-0,4$; $-0,3$; $-0,2$; $-0,1$; 0 ; $0,05$; $0,1$; $0,15$; $0,2$; $0,25$; $0,3$; $0,35$; $0,4$; $0,45$; $0,5$, kde např. hodnota -1 vyjadřuje, že klient má přírůstek k pojistnému ve výši 100%, hodnota 0 znamená, že klient nemá ani slevu ani přírůstek a např. hodnota $0,5$ říká, že klient má 50% slevu. Zavedeme si kategorie podle hodnoty bonus.

Na obrázku 3.22 máme na ose x jednotlivé kategorie proměnné bonus, na ose y jsou frekvence; pro každou kategorii nám sloupec značí skutečnou frekvenci v dané kategorii vypočtenou na základě datového souboru a tečka reprezentuje frekvenci stanovenou podle modelu „GLM 8“. V grafu vidíme, že především záporné hodnoty bonusu nejsou modelem „GLM 8“ dobře vystihnuty.

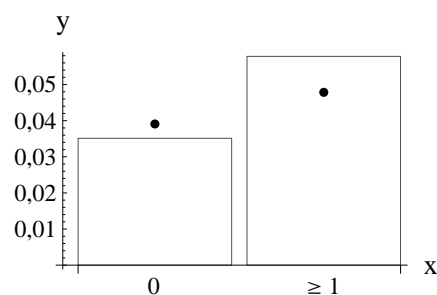
Proměnná historie nám říká, kolik vzniklo pojistných událostí na všech pojistkách klienta za posledních 5 let. V rámci této proměnné si vytvoříme kategorii pro 0 škod a kategorii pro 1 a více škod. Graf na obrázku 3.23 ukazuje, jak se liší frekvence jednotlivých kategorií vypočtené z dat (sloupce) a pomocí modelu „GLM 8“ (tečky).

Vytvoříme si další zobecněný lineární model, který bude mít stejných osm regresorů se stejnými kategoriemi jako model „GLM 8“, navíc však bude mít regresory bonus a historie. Tento nový model si označme „GLM 10“. Ještě si shrňme všech deset regresorů: bonus x_1 , výkon x_2 , hmotnost x_3 , palivo x_4 , objem x_5 , město x_6 , osoba x_7 , věk x_8 , nové vozidlo x_9 , historie x_{10} .

Odhady parametrů modelu jsou v příloze A v tabulce A.4.

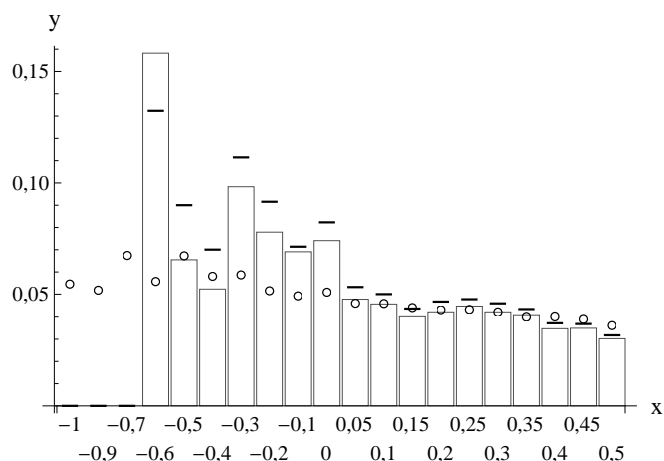


Obrázek 3.22: „GLM 8“, bonus

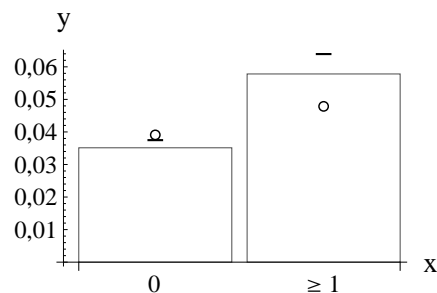


Obrázek 3.23: „GLM 8“, historie

Opět se podíváme, zda model „GLM 10“ odhaduje frekvence proměnných bonus a historie lépe než „GLM 8“ — to vidíme na grafech na obrázcích 3.24 a 3.25. Můžeme říci, že zahrnutím proměnných do modelu dostáváme lepší odhady. Je třeba však zdůraznit, že pro kategorie bonusu se zápornými hodnotami máme k dispozici velmi málo pozorování.



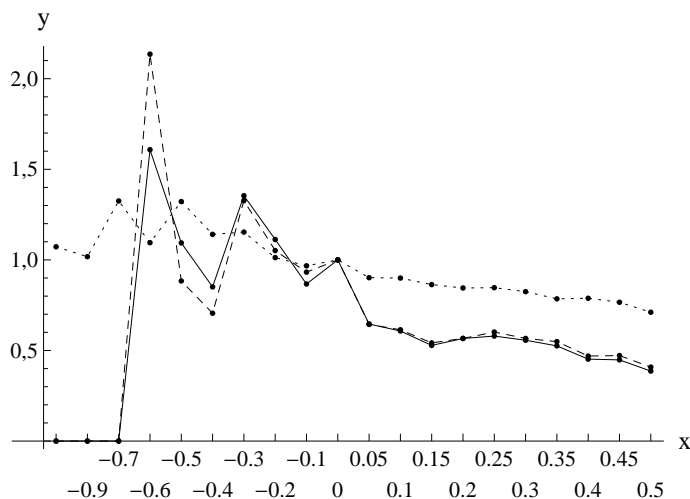
Obrázek 3.24: „GLM 10“, bonus



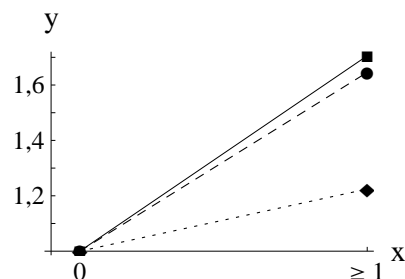
Obrázek 3.25: „GLM 10“, historie

U proměnné bonus porovnejme změny frekvencí v jednotlivých kategoriích oproti frekvenci v kategorii 0 (na rozdíl od předchozích případů to není první kategorie, ale na příslušném grafu leží přibližně uprostřed). Vytvoříme si grafy stejným způsobem jako v předchozích případech, tedy na osu x umístíme jednotlivé kategorie, na osu y podíl frekvence v dané kategorii a frekvence v kategorii 0. Spojnice podílů frekvencí jsou čárkované, plné a tečkované, postupně pro hodnoty vypočtené z pozorovaných dat, z modelu „GLM 10“ a z modelu „GLM 8“.

Podobně vytvoříme grafy pro proměnnou historie (zde také vztahujeme frekvence ku frekvenci v kategorii 0). Podle obrázků 3.26 a 3.27 lze říci, že změny frekvencí vypočtené podle modelu „GLM 10“ lépe popisují skutečnost než změny frekvencí vypočtené podle modelu „GLM 8“.



Obrázek 3.26: „GLM 10“, bonus — změna



Obrázek 3.27: „GLM 10“, historie — změna

Model „GLM 10“ je potřeba ještě mírně upravit, aby byl splněn pokyn k uplatňování směrnice Rady 2004/113/ES v pojišťovnictví. Je třeba, aby bylo dodrženo pravidlo stejného přístupu k oběma pohlavím, tudíž nahradíme v modelu „GLM 10“ regresor x_7 osobou proměnnou RCIC. Tato proměnná má pouze dvě kategorie: RC a IC, kde RC označuje smlouvu s rodným číslem (fyzickou osobu) a IC značí klienta s identifikačním číslem osoby (právníckou osobu). Tento pozměněný model označme „GLM“, odhady parametrů modelu jsou v příloze A v tabulce A.5.

V průběhu vytváření modelu jsme jednotlivé zobecněné lineární modely nesrovnávali jen na základě grafického zobrazení, ale srovnávali jsme je i pomocí statistických kritérií. Shrnutí těchto kritérií pro jednotlivé modely nalezneme v tabulce 3.2, kde AIC značí Akaikeho informační kritérium, LRI je McFaddenův index podílu věrohodností (likelihood-ratio index), $Pearson\chi^2$ označuje Pearsonovu chí kvadrát statistiku a D je deviance. S teoretickými výpočty jednotlivých statistik jsme se setkali v části 2.3.

Model	AIC	LRI	$Pearson\chi^2$	D
GLM 4	37948.4	0.0107959	142234	29397.5
GLM 6	37912.9	0.0111395	142072	29363.1
GLM 8	37890.0	0.0118176	141974	29334.2
GLM 10	36018.9	0.0153155	133686	27818.9
GLM	36040.8	0.0146740	133685	27842.8

Tabulka 3.2: Srovnání zobecněných lineárních modelů.

Po prozkoumání jednotlivých hodnot můžeme říci, že nejlépe vycházejí modely s deseti proměnnými, přičemž model „GLM“, který nerozlišuje mezi pohlavími, není tak dobrý jako model „GLM 10“. Přesto jako finální model musíme z legislativních důvodů zvolit statisticky horší model „GLM“.

3.3 Severita

V této části se budeme věnovat výši škody. Začneme nejjednodušším přístupem, spočteme si pro daná data průměrnou výši malé škody podle vztahu (1.9)

$$I_z = \frac{138\,001\,984}{4\,792} \doteq 28\,798,41 \text{ Kč.} \quad (3.3)$$

Nyní přistoupíme k otázce stanovení výše škody jiným způsobem. Rozdělíme si smlouvy podle zvolených proměnných do několika kategorií a v rámci kategorie spočteme výši průměrné škody. Proměnné a jejich kategorie zvolíme stejně jako při výpočtu frekvence v části 3.1. Máme tedy proměnné věk, objem a město, věkové kategorie jsou: do 29 let, od 30 do 59 let, nad 60 let a firma, pro objem máme kategorie: do 1 000 ccm, 1 001–1 350 ccm, 1 351–1 400 ccm, 1 401–1 850 ccm, 1 851–2 500 ccm a nad 2 500 ccm a pro město: do 24 999 obyvatel, 25 000–499 999 a nad 500 000 obyvatel.

Pro každou kategorii dané proměnné si spočteme průměrnou výši škody podle (1.10), což např. pro i -tou kategorií věku zapíšeme takto:

$$\bar{z}_{vek,vek_i} = \frac{\sum_{k=1}^{|vek,vek_i|} z_{vek,vek_{ik}}}{N_{vek,vek_i}}, \quad (3.4)$$

kde i -tá věková kategorie obsahuje celkem $|vek,vek_i|$ smluv, $z_{vek,vek_{ik}}$ je výše k -té škody v i -té věkové kategorii a N_{vek,vek_i} značí počet škod v i -té věkové kategorii.

Průměrné výše škod, součty škod a počty škod v jednotlivých kategoriích shrnuje tabulka 3.3. Pro konkrétního klienta pak očekávanou výši škody I_z dostáváme podle (1.11) jako průměr škod kategorií, do kterých daný klient spadá.

Kategorie věku	Součet škod	Počet škod	Průměrná výše škody
≤ 29	10 059 059	336	29 937,68
30–59	73 105 533	2 596	28 160,84
≥ 60	22 336 708	794	28 131,87
firma	32 500 684	1 066	30 488,45
Kategorie objemu			
$\leq 1\,000$ ccm	3 571 723	123	29 038,40
1 001–1 350 ccm	28 288 351	1 015	27 870,30
1 351–1 400 ccm	19 847 862	714	27 798,13
1 401–1 850 ccm	33 065 346	1 161	28 480,06
1 851–2 500 ccm	46 431 736	1 567	29 630,97
$\geq 2\,501$ ccm	6 796 966	212	32 061,16
Kategorie města			
$\leq 24\,999$	60 758 566	2 156	28 181,15
25 000–499 999	50 079 272	1 781	28 118,63
$\geq 500\,000$	27 164 146	855	31 770,93

Tabulka 3.3: Průměrné výše škod pro jednotlivé kategorie.

Podíváme se blíže na průměrné škody v jednotlivých kategoriích v tabulce 3.3. Nejprve se zaměříme na kategorie věku, zda je mezi průměrnými výšemi škod významný rozdíl. V programu Wolfram Mathematica 8.0 využijeme zabudované funkce „LocationTest“, ta zvolí vhodný test pro zadaná data (vstupem jsou výše škod dvou testovaných kategorií) a ověří, zda mediány nebo průměry¹ vstupních dat můžeme považovat za stejné. V rámci proměnné věk tak dospějeme k závěru, že mezi průměrnými výšemi škod v jednotlivých kategoriích není statistický rozdíl.

K podobnému závěru dojdeme i v případě proměnné obec. S jinými výsledky se však setkáme u proměnné město — zde totiž je významný rozdíl mezi řidiči z Prahy a mimo Prahu. Výpočet očekávané výše škody ${}^H z$ bychom tedy mohli zredukovat na pouhý výběr ze dvou možností: Praha/jiná obec, přičemž pro Prahu zůstává ${}^H z = 31\,770,93$ Kč a pro jiné obce pak máme ${}^H z = 28\,152,87$ Kč. Budeme-li dále uvažovat pro odhad výše škody pouze proměnnou město se dvěma kategoriemi (Praha/jiná obec), použijeme označení ${}^H z (P)$.

Pro očekávanou výši škody je možné vyvinout zobecněný lineární model. V tomto případě předpokládáme gama rozdělení u vysvětlované proměnné. Při výstavbě modelů však narážíme na situaci, kdy pomocí hodnotících kritérií (AIC, $Pearson\chi^2$ a další) těžce vybíráme nejvhodnější model. Konkrétní kritéria se totiž u jednotlivých modelů liší velmi málo (viz tabulka A.6 v příloze A). To můžeme vysvětlit tím, že se nedá snadno říci, kteří klienti (která kategorie jaké proměnné) zpravidla zapříčiňují vyšší škody.

Zde vidíme souvislost s jednoduchým přístupem ke stanovení očekávané výše škody — různé seskupování smluv nepřináší uspokojivé výsledky.

Pokud bychom chtěli zobecněný lineární model zde použít, vybrali bychom podle tabulky A.6 v příloze A model s těmito proměnnými: věk (\mathbf{x}_3), město (\mathbf{x}_2), palivo (\mathbf{x}_1). Odhady parametrů tohoto modelu jsou rovněž v příloze A v tabulce A.7. Pro další výpočet tento model označme severita GLM.

3.4 Nettopojistné

Tuto část začneme přímým výpočtem nettopojistného (nettopojistné ${}^I P$ a ${}^H P$ podle části 1.2.3). Dále zmíníme stanovení nettopojistného pomocí dílčích výsledků — frekvence a severity škod. Následně se budeme zabývat srovnáváním různých přístupů.

Nejdříve se podíváme na nejjednodušší způsob výpočtu nettopojistného podle (1.18). Pro daná data máme nettopojistné

$${}^I P \doteq \frac{138\,001\,984}{127\,271,39} \doteq 1\,084,31 \text{ Kč.} \quad (3.5)$$

Výsledkem jednoduchého výpočtu je jednotná sazba nettopojistného pro všechny smlouvy ve výši 1 084 Kč. Tato částka vyjadřuje průměrné zatížení na smlouvu z hlediska škodních nákladů.

¹V našich testovaných případech byl vždy Mathematicou zvolen Mann-Whitney test, který testuje shodnost mediánů.

Další způsob výpočtu nettopojistného již využívá členění souboru smluv do kategorií podle proměnných. S tímto postupem jsme se seznámili v části 1.2.3 a podle něj nyní budeme hledat nettopojistné ${}^I P$.

Proměnné a jejich kategorie volíme stejné jako v případě výpočtu frekvence ${}^I f$ a severity ${}^I z$. Začneme výpočtem dílčího nettopojistného, například pro i -tou kategorií věku podle (1.19) máme

$$P_{vek,vek_i} = \frac{\sum_{k=1}^{|vek,vek_i|} z_{vek,vek_{ik}}}{\sum_{k=1}^{|vek,vek_i|} \omega_{vek,vek_{ik}}}, \quad (3.6)$$

pokračujeme výběrem referenčních kategorií. Ty volíme podle největší expozice následovně: u proměnné věk to je kategorie 30–59 let, u proměnné objem kategorie 1 851–2 500 ccm a u proměnné město to je kategorie obcí do 24 999 obyvatel. Dále počítáme faktory u jednotlivých kategorií, například pro i -tou kategorií proměnné věk máme

$$fac_{vek,vek_i} = \frac{P_{vek,vek_i}}{P_{vek,<60}}, \quad (3.7)$$

kde $P_{vek,<60}$ označuje dílčí nettopojistné referenční kategorie věku, a to kategorie 30–59 let.

Podle vztahu (1.21) spočteme průměrné faktory a následně bazickou hodnotu, která v našem případě vychází následovně:

$$\begin{aligned} P_{base} &= \frac{1}{\overline{fac}_{vek} \cdot \overline{fac}_{objem} \cdot \overline{fac}_{mesto}} \cdot {}^I P = \\ &= \frac{1}{1,098\ 11 \cdot 0,810\ 623 \cdot 1,228\ 92} \cdot 1\ 084,31 \doteq 991,206. \end{aligned} \quad (3.8)$$

Tabulka 3.4 nám shrnuje dílčí výpočty (součty škod, součty expozic a nettopojistné) a faktory.

Konkrétnímu klientovi pak spočteme nettopojistné ${}^I P$ podle vztahu (1.24), kde bazická hodnota činí 991,206 Kč a faktory potřebné k výpočtu nalezneme v tabulce 3.4.

Výpočet nettopojistného pomocí zvlášť spočtené frekvence a výše škody spočívá pouze v jejich vynásobení.

Kategorie věku	Součet škod	Součet expozicí	Nettopojistné	Faktor
≤ 29	10 059 059	7 086,12	1 419,54	1,437 6
30–59	73 105 533	74 036,02	987,43	1,000 0
≥ 60	22 336 708	25 873,46	863,31	0,874 3
firma	32 500 684	20 275,80	1 602,93	1,623 3
Kategorie objemu				
$\leq 1\,000$ ccm	3 571 723	4 971,98	718,37	0,537 0
1 001–1 350 ccm	28 288 351	29 910,39	945,77	0,707 0
1 351–1 400 ccm	19 847 862	22 231,43	892,78	0,667 4
1 401–1 850 ccm	33 065 346	31 234,70	1 058,61	0,791 4
1 851–2 500 ccm	46 431 736	34 711,96	1 337,63	1,000 0
$\geq 2\,501$ ccm	6 796 966	4 210,93	1 614,13	1,206 7
Kategorie města				
$\leq 24\,999$	60 758 566	68 861,65	882,33	1,000 0
25 000–499 999	50 079 272	40 833,63	1 226,42	1,390 0
$\geq 500\,000$	27 164 146	17 576,11	1 545,52	1,751 6

Tabulka 3.4: Dílčí nettopojistné a faktory pro jednotlivé kategorie proměnných.

Nyní budeme porovnávat jednotlivé způsoby výpočtu nettopojistného — podíváme se na součty nettopojistného. Ty nám shrnuje tabulka 3.5, kde první sloupec nese informaci o způsobu výpočtu frekvence škod, druhý sloupec obsahuje způsob výpočtu výše škod a třetí sloupec je vyplněn pouze ve dvou případech, a to v těch, kdy byl použit přímý výpočet nettopojistného. Jinak je nettopojistné počítáno jako součin frekvence a severity.

frekvence	severita	nettopojistné	součet nettopojistného (v Kč)
-	-	I_P	149 920 324
I_f	I_z	-	149 920 324
-	-	II_P	152 786 696
I_f	II_z	-	149 176 163
I_f	II_z (P)	-	149 151 977
II_f	I_z	-	148 666 464
II_f	II_z	-	148 234 122
II_f	II_z (P)	-	148 190 119
GLM	I_z	-	139 792 207
GLM	II_z	-	139 799 599
GLM	II_z (P)	-	139 776 896
GLM	GLM	-	144 747 108
II_f	GLM	-	151 621 619
I_f	GLM	-	152 028 697

Tabulka 3.5: Porovnání celkového nettopojistného (v Kč) spočteného různými způsoby.

Pro jednoduchost a snadnou porovnatelnost počítáme nettopojistné u smluv, kde máme všechny potřebné údaje k dispozici. Součet všech malých škod na těchto smlouvách činí 129 962 726 Kč. Toto je částka, kterou chceme pokrýt celkovým nettopojistným. Hodnoty součtů odhadnutého nettopojistného u jednotlivých

smluv máme ve čtvrtém sloupci tabulky 3.5. Ve druhém sloupci tabulky vidíme dva způsoby výpočtu očekávané výše škody s horním levým indexem II (${}^{II}z$ a ${}^{II}z(P)$). Rozlišujeme tedy výpočet ${}^{II}z$ podle (1.11) s pomocí tabulky 3.3 a pod označením ${}^{II}z(P)$ máme na mysli stanovení výše škody pomocí jediné proměnné (město) se dvěma kategoriemi (Praha, jiná obec) — jak již bylo uvedeno v části 3.3.

Nejblíže je součtu malých škod u daných smluv hodnota 139 776 896 Kč. Tato hodnota podle tabulky 3.5 odpovídá tomu, že u všech smluv jsme frekvenci odhadli zobecněným lineárním modelem „GLM“ a výši škod jsme odhadli ${}^{II}z(P)$, odhadnutou frekvenci škod a výši škod jsme vynásobili a provedli součet přes uvažované smlouvy.

Pro další výpočty budeme uvažovat nettopojistné jako součin frekvence a severity, kde frekvenci odhadujeme zobecněným lineárním modelem „GLM“ a severitu odhadujeme jako ${}^{II}z(P)$.

Kapitola 4

Velké škody

V této kapitole se budeme věnovat velkým škodám, tedy těm škodám, které převyšují 150 000 Kč (tuto hranici jsme si definovali na začátku 3. kapitoly).

Začneme nahlédnutím na frekvenci a průměrnou výši velkých škod, provedeme jednoduché výpočty podle (1.3) a (1.9).

Označme N_V počet velkých škod, frekvenci velkých škod ${}^I f_V$ pak zapíšeme a spočítáme pro naše data následovně:

$${}^I f_V = \frac{N_V}{\sum_{k=1}^m \omega_k} = \frac{186}{127\,271,39} \doteq 0,001\,46. \quad (4.1)$$

Dále mějme Z_k velkou škodu na k -té smlouvě. Pro výši škody pak máme:

$${}^I z_V = \frac{\sum_{k=1}^m Z_k}{N_V} = \frac{100\,855\,181}{186} \doteq 542\,232,16 \text{ Kč}. \quad (4.2)$$

V tabulce 4.1 máme porovnání frekvence a výše škod pro malé a velké škody. Vidíme, že velké škody nenastávají tak často jako malé škody, zato výše velkých škod je několikanásobně vyšší.

	${}^I f$	${}^I z$
malé škody	0,037 7	28 798
velké škody	0,001 5	542 232

Tabulka 4.1: Průměrná frekvence a výše malých a velkých škod.

Nyní provedeme jednoduchý přímý výpočet nettopojistného ${}^I P_V$ pro velké škody podle (1.18):

$${}^I P_V = \frac{\sum_{k=1}^m Z_k}{\sum_{k=1}^m \omega_k} = \frac{100\,855\,181}{127\,271,39} \doteq 792,44 \text{ Kč} \quad (4.3)$$

I zde samozřejmě platí rovnost (1.26):

$${}^I P_V = {}^I f_V \cdot {}^I z_V. \quad (4.4)$$

Dále se zaměříme se na výše velkých škod. Pokusíme se popsat pozorované výše velkých škod pomocí vhodného rozdělení pravděpodobností, budeme uvažovat Paretovo rozdělení, gama rozdělení a logaritmicko-normální rozdělení. To rozdělení, které bude pozorovaná data vystihovat nejlépe, lze použít k modelování budoucích výší velkých škod. Nebudeme se podrobněji zabývat např. exponenciálním rozdělením, protože pravděpodobnost výskytu vysokých škod je značně podceněna (viz kapitola 1). Ze stejného důvodu se jako nedostačující ukázalo i Weibullovo rozdělení.

V programu Wolfram Mathematica 8.0 si pomocí zabudované funkce „EstimatedDistribution“ aplikované na pozorované výše velkých škod odhadneme parametry k , α a μ Paretova rozdělení typu II daného touto hustotou rozdělení pravděpodobností (jde o hustotu (1.14) z 1. kapitoly):

$$f(x) = \alpha k^\alpha (k + x - \mu)^{-1-\alpha}, \quad x \geq \mu.$$

Dostáváme tyto odhady parametrů:

$$\hat{k} = 126\,132,30; \quad \hat{\alpha} = 1,179\,337 \text{ a } \hat{\mu} = 150\,014. \quad (4.5)$$

Stejným způsobem si pomocí softwaru odhadneme parametry α a p gama rozdělení s hustotou rozdělení pravděpodobností (hustota (1.15) z 1. kapitoly):

$$f(x) = \frac{\alpha^p}{\Gamma(p)} x^{(p-1)} e^{-\alpha x}, \quad x \geq 0.$$

Odhady parametrů máme tyto:

$$\hat{\alpha} = 3,960\,24 \times 10^{-7} \text{ a } \hat{p} = 0,214\,737. \quad (4.6)$$

Postup zopakujeme pro odhad parametrů logaritmicko-normálního rozdělení s hustotou rozdělení pravděpodobností (jako (1.17) v 1. kapitole):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \quad x > 0.$$

Pro parametry μ a σ máme tyto odhady:

$$\hat{\mu} = 12,689\,398 \text{ a } \hat{\sigma} = 0,787\,748. \quad (4.7)$$

Máme tedy odhadnuta pravděpodobnostní rozdělení a nyní si otestujeme, které nejlépe popisuje pozorovaná data. Opět využijeme software Mathematica, tentokrát nám poslouží funkce „DistributionFitTest“, pomocí jíž zjistíme, zda data mají dané rozdělení. Nulovou hypotézou je vždy situace, kdy pozorované hodnoty považujeme za data pocházející z testovaného rozdělení, alternativní hypotézou je opak (data nepocházejí z testovaného rozdělení).

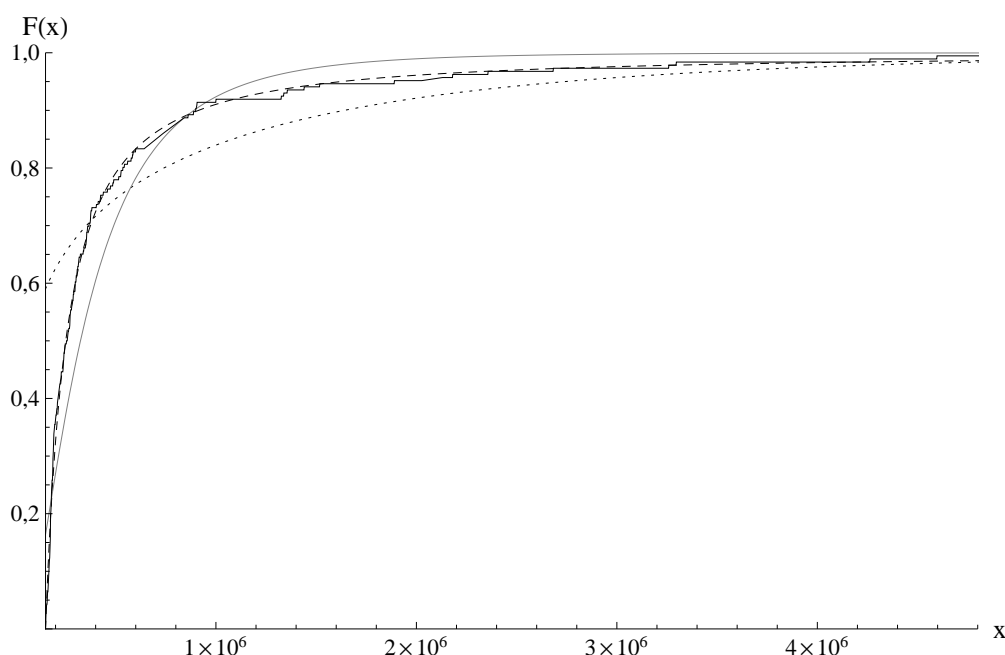
Výsledkem testů jsou tyto p-hodnoty:

rozdělení	Paretovo II	gama	log-normální
p-hodnota	0,753	$4,44 \times 10^{-16}$	$5,29 \times 10^{-5}$

Mathematica automaticky použila pro testování rozdělení Cramér-von Mises test. O datech lze tedy říci, že mají přibližně Paretovo rozdělení s distribuční funkcí

$$F(x) = 1 - \frac{1}{(1 + 7,928\ 18 \times 10^{-6}(x - 150\ 014))^{1,179\ 34}}, \quad x \geq 150\ 014. \quad (4.8)$$

Pro názornost si porovnejme uvažovaná rozdělení graficky. Na obrázku 4.1 jsou zobrazeny distribuční funkce Paretova rozdělení — vykresleno čárkovaně, gama rozdělení — tečkovaně a logaritmicke-normální rozdělení — znázorněno plnou šedou čarou. V grafu vidíme černou schodovitou křivku — to je empirická distribuční funkce vypočtená z pozorovaných dat. Je vidět, že nejbližší má empirická distribuční funkce k čárkované distribuční funkci (tedy k Paretovu rozdělení).



Obrázek 4.1: Distribuční funkce Paretova rozdělení (čárkovaně), gama rozdělení (tečkovaně), logaritmicke-normálního rozdělení (plná šedá) a empirická distribuční funkce (schodovitá černá).

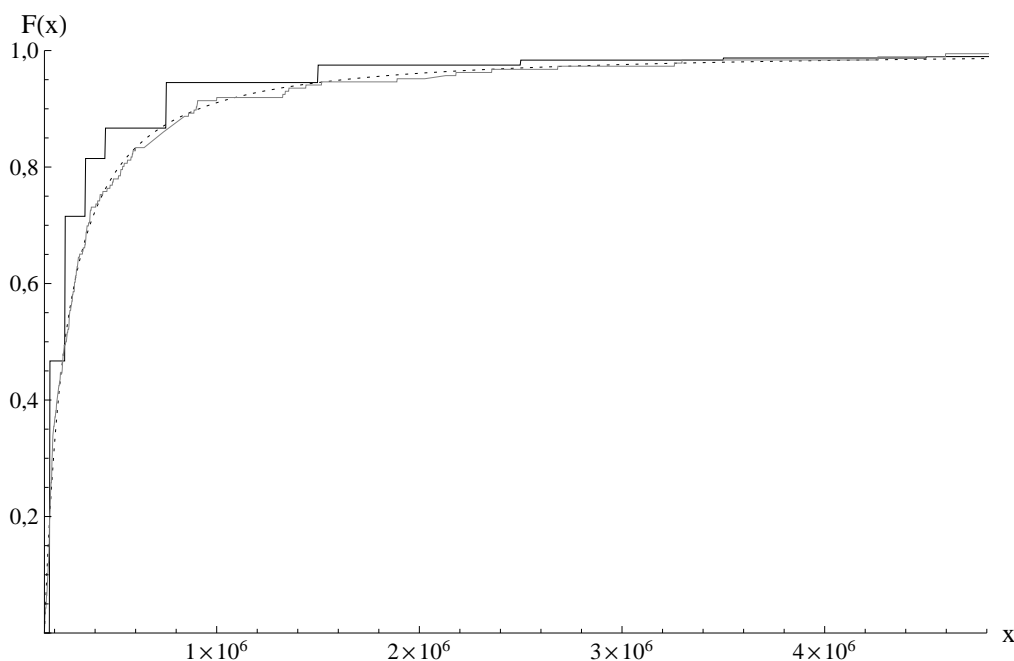
Porovnejme si rozdělení velkých škod našeho datového souboru s rozdělením (empirickým) velkých škod z českého trhu. Máme k dispozici data z celého českého trhu pojištění odpovědnosti z provozu motorového vozidla z let 2000–2008 (data¹ jsou v souboru LargeClaims_2008.xlsx na přiloženém CD ve složce data). Pro každý rok máme údaj o celkovém počtu nahlášených škod v určitých intervalech výší škod. Známe celkový počet škod pro dané roky a z dostupných dat snadno dopočteme počet škod nad 150 000 Kč (tedy počet těch škod, které považujeme za velké).

Pro velké škody si spočteme celkový počet škod v letech 2000–2008 v daných intervalech, dostáváme tak následující tabulku:

¹Data poskytl vedoucí diplomové práce.

interval výše škody (v Kč)	průměrná škoda	počet škod
150 001–200 000	175 000,5	43 132
200 001–300 000	250 000,5	22 931
300 001–400 000	350 000,5	9 164
400 001–500 000	450 000,5	4 822
500 001–1 000 000	750 000,5	7 219
1 000 001–2 000 000	1 500 000,5	2 768
⋮	⋮	⋮

Pomocí této tabulky a softwaru si odhadneme empirickou distribuční funkci rozdělení výše velké škody tržních dat. Jelikož nemáme k dispozici přesné výše škod, ale pouze intervaly, použijeme pro odhad distribuční funkce průměrné hodnoty intervalů. Empirickou distribuční funkci velkých škod tržních dat máme zakreslenou v obrázku 4.2 jako černou schodovitou křivku. Ve stejném obrázku je také zakreslena empirická distribuční funkce velkých škod našeho datového souboru (šedá křivka) a tečkovaně zde máme distribuční funkci Paretova rozdělení (4.8). Vidíme, že uvažované empirické distribuční funkce jsou si blízko, byť provedený Kolmogorov-Smirnovův test jednoznačně zamítá hypotézu shodnosti rozdělení (což není překvapivé vzhledem ke konstrukci empirické distribuční funkce tržních velkých škod).



Obrázek 4.2: Empirická distribuční funkce velkých škod tržních dat (černě), empirická distribuční funkce velkých škod našeho datového souboru (šedě) a distribuční funkce Paretova rozdělení (4.8).

Při odhadování výše velké škody se v tomto případě nebudeme spoléhat jen na vlastní data (máme na mysli data ze souborů smlouvy.txt a škody.txt), ale využijeme i tržní data, která máme k dispozici. Podíváme se na využití kredibility podle (1.28), přičemž namísto pojistného uvažujeme výši velké škody.

Vlastních pozorování s velkými škodami máme 186, jde o velké škody během jednoho roku. Průměrná výše velké škody je 542 232,16 Kč. V souboru Lar-

geClaims_2008.xlsx máme za 9 let celkem 92 342 velkých škod z českého trhu. Průměrná velká škoda za 9 let činí 503 199,24 Kč. Předpokládejme, že vlastní pozorování jsou reprezentativním vzorkem trhu. Pak stanovíme kredibilitní koeficient Z takto:

$$Z = \sqrt{\frac{186}{92\,342}} \doteq 0,044\,88. \quad (4.9)$$

Takto striktně stanovený kredibilitní koeficient příkládá velký význam tržním datům. V případě, že bychom měli k dispozici 92 342 vlastních pozorování velkých škod, kredibilitní faktor by byl roven jedné.

Očekávanou výši velké škody z_V s využitím tržních dat pak podle (1.28) zapíšeme:

$$z_V = 0,044\,88 \cdot 542\,232,16 + (1 - 0,044\,88) \cdot 503\,199,24 \doteq 504\,951 \text{ Kč} \quad (4.10)$$

Pro další výpočty budeme u všech smluv odhadovat výši velké škody stejně, a to jako z_V .

Vraťme se ještě k frekvenci velkých škod, resp. k počtu velkých škod a porovnejme si podíl velkých škod v našem datovém souboru a v tržních datech. V našem souboru dat máme celkem 4 978 škod, z toho je 186 velkých, což odpovídá 3,74 % všech škod. Tento údaj porovnáme s příslušnými procenty vypočtenými na základě dat z celého českého trhu pojištění odpovědnosti z provozu motorového vozidla z let 2000–2008. Tabulka 4.2 nám shrnuje počty škod, počty velkých škod a podíl velkých škod ku všem škodám v jednotlivých letech 2000–2008, v poslední řádce máme údaje o našich datech.

rok	počet všech škod	počet velkých škod	podíl velkých škod
2000	270 303	8 114	0,0300
2001	311 512	8 913	0,0286
2002	318 357	9 392	0,0295
2003	299 016	9 829	0,0328
2004	296 764	11 080	0,0373
2005	303 534	11 739	0,0387
2006	302 015	11 930	0,0395
2007	298 930	11 693	0,0391
2008	279 900	9 654	0,0345
data 2011	4 978	186	0,0374

Tabulka 4.2: Podíl velkých škod na českém trhu v letech 2000–2008 a podíl velkých škod ve zkoumaných datech.

Vidíme, že podíl velkých škod v námi zkoumaných datech se významně neliší od podílu velkých škod na trhu.

Frekvenci škod budeme pro další výpočty uvažovat pro všechny smlouvy stejnou — spočtenou nejjednodušším přístupem (viz vztah (4.1)). Vytváření kategorií v rámci proměnných v tomto případě činí výpočet složitějším bez znatelného přínosu.

Nettopojistné pro velké škody pak máme rovněž stejné pro všechny smlouvy:

$$P_V = {}^I f_V \cdot z_V = 0,001\,46 \cdot 504\,951 \doteq 738 \text{ Kč}. \quad (4.11)$$

Kapitola 5

Stanovení bruttopojistného

5.1 Bonus-malus

Než přistoupíme k závěrečné části práce, podíváme se na systém bonus-malus. Při stanovování výše pojistného má podle zákona č. 168/1999 Sb. pojistitel zohlednit předcházející škodný průběh pojištění odpovědnosti pojistníka. V případě bezeškodného průběhu pojištění se toto zohledňuje slevou na pojistném (tzv. *bonus*) a v případě výplaty pojistného plnění přírážkou k pojistnému (tzv. *malus*). Příklad modelu systému bonus-malus nalezneme například v [3] (část VI.4.).

Podle délky bezeškodné doby (v měsících) si pojistníky rozdělíme do tříd. Toto rozdělení je uvedeno v tabulce 5.1 — jde o příklad rozdělení do tříd. Pro jednoduchost uvažujeme 10 bonusových tříd (B 1, ..., B 10), pouze jednu malusovou třídu (M) a třída Z je základní, kde pojistník nemá ani slevu ani přírážku.

Bezeškodná doba v měsících	Třída	% z ročního pojistného
120 a více	B 10	50 %
108 až 119	B 9	55 %
96 až 107	B 8	60 %
84 až 95	B 7	65 %
72 až 83	B 6	70 %
60 až 71	B 5	75 %
48 až 59	B 4	80 %
36 až 47	B 3	85 %
24 až 35	B 2	90 %
12 až 23	B 1	95 %
0 až 11	Z	100 %
-1 měsíc a méně	M	110 %

Tabulka 5.1: Příklad rozdělení do tříd bonusu a malusu.

Bezeškodná doba může nabývat i záporných hodnot, protože při každé jedné škodě snížíme bezeškodnou dobu o 36 měsíců. Naopak bezeškodnou dobu je možné zvýšit při bezeškodném průběhu, uvažujme toto vyhodnocování jednou ročně, tedy navyšujeme o 12 měsíců.

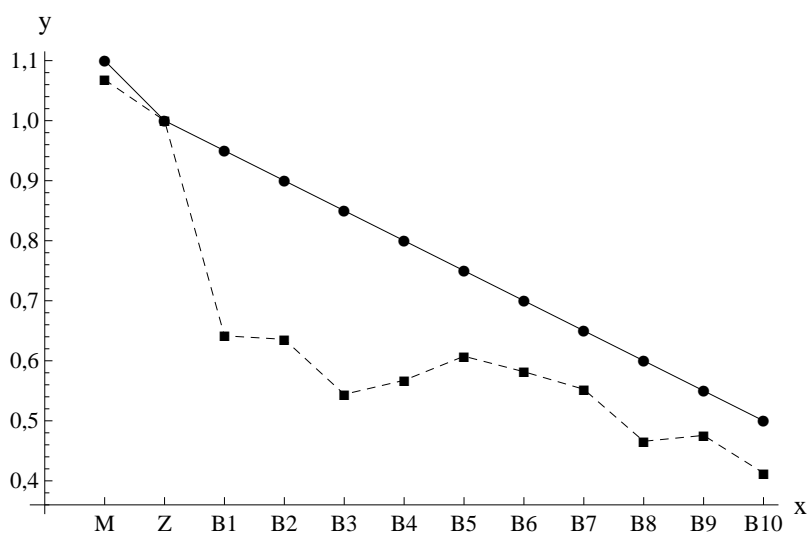
V dnešní době na českém trhu běžně nacházíme větší počet malusových tříd,

můžeme narazit i na větší počet bonusových tříd či dokonce na 60% bonus¹.

Podívejme se na bonusy, resp. malusy reálného souboru dat, který máme k dispozici. V souboru využijeme informaci o poskytnuté slevě na dané smlouvě (7. sloupec v původním souboru smlouvy.txt), kde např. hodnota 0,5 znamená 50% slevu a např. hodnota -1 znamená 100% přírůžku (jde o proměnnou bonus z části 3.2). Máme celkem 10 malusových tříd, 10 bonusových tříd a jednu základní, avšak kvůli malému (v jednom případě žádnému) počtu pozorování toto třídění celé nevyužijeme a za účelem dalšího zkoumání si smlouvy rozdělíme do tříd podle tabulky 5.1. Takže namísto deseti malusových tříd budeme mít pouze jednu.

Uvažujme situaci, kdy v základní třídě Z máme pojistné 10 000 Kč. Pokud bychom jednoduše aplikovali procentuální část základního pojistného z tabulky 5.1 na ostatní třídy, pak bychom měli pojistné 9 500 Kč pro třídu B 1, ..., 5 000 Kč pro třídu B 10 a 11 000 Kč pro třídu M. Zanesme si do grafu jak se mění poměr pojistného všech tříd ku základní třídě. V obrázku 5.1 to jsou body pospojované plnou čarou. Do stejného obrázku zakreslíme ještě jeden graf, dále vysvětlíme jaký.

Předpokládejme výpočet pojistného jako součin frekvence škod a výše škod. Pro jednoduchost mějme očekávanou výši škody u všech smluv stejnou a frekvenci škod počítejme pro jednotlivé třídy M, Z, B 1, ..., B 10 z reálných dat, které máme k dispozici. Pro každou třídu máme tedy skutečnou frekvenci škod (podíl počtu škod a celkové expozice v rámci dané skupiny) a dále si spočteme poměr frekvencí všech tříd ku frekvenci třídy základní. Stačí nám počítat poměr frekvencí, protože poměr dopočteného pojistného by byl stejný. Graf zaneseme do obrázku 5.1 a body spojíme čárkovaně.



Obrázek 5.1: Část základního pojistného placená v jednotlivých bonusových třídách, teoretické hodnoty spojeny plně, hodnoty z reálných dat spojeny čárkovaně.

V obrázku 5.1 vidíme rozdílné chování reálných dat a teoretických hodnot. Velmi výrazný skok je u reálných dat ze základní třídy do třídy B 1, což lze

¹Souhrn aktuálních informací z českého trhu je dostupný např. na www.abcpojisteni.cz nebo na www.epojisteni.cz.

interpretovat tak, že v základní třídě jsou začínající řidiči téměř bez zkušeností s řízením, proto jsou nehody v této kategorii časté, kdežto ve třídě B 1 již jsou „o rok“ zkušenější řidiči.

Tady bychom mohli říci, že systém bonus-malus podle tabulky 5.1 je nespravedlivý. To proto, že řidičům ve třídě B 1 by poskytoval slevu jenom 5 %, přestože řidiči v této třídě mají o třetinu nižší frekvenci škod než řidiči základní třídy.

Třída B 2 je na tom podobně jako třída B 1, co se týče frekvence škod. Co se týká slevy — ta by vzrostla na 10 %. V dalších třídách pozorujeme vylepšování křivky ve smyslu snižování frekvence škod (až na výjimky), což se dá očekávat s přibývajícím řidičskými zkušenostmi.

Dále uvažujme, že jednotlivé třídy tvoří stavy Markovského řetězce². Pro takovýto Markovský řetězec si můžeme spočítat matici pravděpodobností přechodu na základě souboru dat, který máme k dispozici. Každou ze smluv máme zařazenu v nějaké třídě podle tabulky 5.1, pro každou třídu si spočteme, resp. již máme spočtenou skutečnou frekvenci škod. Tato frekvence (udávající počet škod na smlouvě v dané třídě za jeden rok) bude parametrem Poissonova rozdělení počtu škod. Pomocí tohoto rozdělení a pomocí tabulky 5.1 spočteme pravděpodobnosti přechodu mezi jednotlivými třídami a sestavíme si matici pravděpodobností přechodu \mathbf{P} . Pro daná data to je matice

$$\mathbf{P} = \begin{pmatrix} 0,075 & 0,925 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,070 & 0 & 0,930 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,046 & 0 & 0 & 0,954 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,001 & 0,044 & 0 & 0 & 0,955 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,0008 & 0 & 0,038 & 0 & 0 & 0,961 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,0008 & 0 & 0 & 0,040 & 0 & 0 & 0,959 & 0 & 0 & 0 & 0 & 0 \\ 10^{-5} & 0,0009 & 0 & 0 & 0,042 & 0 & 0 & 0,957 & 0 & 0 & 0 & 0 \\ 10^{-5} & 0 & 0,0009 & 0 & 0 & 0,041 & 0 & 0 & 0,958 & 0 & 0 & 0 \\ 10^{-5} & 0 & 0 & 0,0008 & 0 & 0 & 0,039 & 0 & 0 & 0,960 & 0 & 0 \\ 10^{-8} & 10^{-6} & 0 & 0 & 0,0006 & 0 & 0 & 0,033 & 0 & 0 & 0,967 & 0 \\ 10^{-8} & 0 & 10^{-6} & 0 & 0 & 0,0006 & 0 & 0 & 0,034 & 0 & 0 & 0,966 \\ 10^{-8} & 0 & 10^{-6} & 0 & 0 & 0,0004 & 0 & 0 & 0,029 & 0 & 0 & 0,970 \end{pmatrix},$$

kde sloupce zleva doprava jsou řazeny stejně jako řádky odshora dolů, a to od třídy M, přes Z, B 1, ..., až po B 10. Pravděpodobnosti přechodu v matici \mathbf{P} jsou zaokrouhleny.

Nyní předpokládejme, že nám do daného souboru nikdo nový nepřijde, ani nám z něj nikdo neodejde. Tento předpoklad je sice nereálný, ale přesto jej uvažujeme za účelem výpočtu stacionárního rozdělení ($\boldsymbol{\pi}$). Naopak se ale můžeme podívat na věc z jiné strany a říci, že i kdyby konkrétní klient přešel k jinému pojistiteli, bonusová třída by mu měla zůstat, resp. by na základě svého vývoje počtu škod pokračoval dále ve výši bonusu/malusu. Takže ze systému by vlastně neodešel.

Stacionární rozdělení námi uvažovaného Markovského řetězce je:

$$\boldsymbol{\pi}' = (2,81 \times 10^{-6}; 7,54 \times 10^{-6}; 1,96 \times 10^{-5}; 6,91 \times 10^{-5}; 0,000159; 0,000678; 0,00181; 0,00268; 0,0300; 0,0289; 0,0279; 0,908).$$

Ze stacionárního rozdělení vidíme, že 90,8 % smluv by po čase skončilo v nejlepší bonusové třídě, takže bychom většině klientů měli poskytovat slevu 50 %.

²Potřebná teorie k Markovským řetězcům je v publikaci Z. Práškové a P. Lachouta Základy náhodných procesů.

Předpokládejme, že počáteční rozdělení klientů v jednotlivých třídách je $\pi'_0 = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$, tedy uvažujeme situaci, kdy 100 % klientů je v základní třídě Z. Mějme matici pravděpodobností přechodu \mathbf{P} jako výše. Budeme sledovat vývoj rozdělení klientů do jednotlivých tříd během několika let. Do tabulky 5.2 jsme pro vybrané roky zaznačili, jaké procento klientů je ve 3 nejméně atraktivních bonusových třídách (M 1, Z, B 1) a ve 3 třídách s nejvyšší slevou (B 8, B 9, B 10). V posledním sloupci tabulky je spočtena průměrná sleva poskytovaná v daném roce.

Rok	M 1	Z	B 1	B 8	B 9	B 10	průměrná sleva v %
0	0	100	0	0	0	0	0
1	7,035	0	92,965	0	0	0	3,945
2	4,779	6,508	0	0	0	0	8,394
3	0,909	8,348	6,050	0	0	0	12,916
4	0,997	0,840	10,992	0	0	0	17,307
5	0,710	1,178	0,781	0	0	0	21,691
8	0,136	0,242	0,223	68,778	0	0	34,815
10	0,040	0,050	0,396	4,476	0	64,212	43,708
12	0,010	0,057	0,068	2,381	21,803	64,636	46,390
15	0,003	0,013	0,017	2,837	7,343	84,446	48,366
18	0,001	0,004	0,006	2,890	3,820	89,315	48,847
20	0,001	0,001	0,003	3,123	2,854	90,370	48,939
23	10^{-6}	0,001	0,002	2,938	2,806	90,683	48,980
25	10^{-6}	0,001	0,002	2,890	2,791	90,753	48,988
30	10^{-6}	0,001	0,002	2,886	2,792	90,773	48,992

Tabulka 5.2: Vývoj rozdělení klientů do tříd a průměrné slevy.

Z tabulky 5.2 vidíme, že při daném počátečním rozdělení a dané matici pravděpodobností přechodu můžeme již po deseti letech očekávat většinu klientů v nejlepší bonusové třídě. Proto je nutné nastavit základní výši pojistného tak, aby po poskytnutí slev byla pojišťovna schopná dostát svým závazkům.

5.2 Bruttopojistné

Zatím jsme prováděli pouze dílčí výpočty vedoucí ke stanovení pojistného, resp. bruttopojistného, které budeme požadovat po klientovi. Nyní se budeme zabývat kalibrací modelu a následně výpočtem finálního bruttopojistného.

Doposud jsme pracovali zvláště s malými a velkými škodami, potřebujeme však znát celkové nettopojistné za všechny škody dohromady. Pro pokrytí malých škod jsme zvolili výpočet nettopojistného jako součin frekvence odhadnuté modelem „GLM“ a severity spočtené podle ${}^H_Z(\mathbf{P})$. V dosavadních výpočtech nettopojistného jsme neuvažovali ty smlouvy, u kterých jsme neměli všechny potřebné údaje. Těmto smlouvám nyní přiřadíme jednotné nettopojistné ${}^I P = 1\,084,31$ Kč.

Pro pokrytí velkých škod jsme zvolili nettopojistné $P_V = 737,96$ Kč. Pro každou smlouvu máme nyní nettopojistné za malé i velké škody — tyto dvě hodnoty sečteme a získáme tak pro každou smlouvu celkové nettopojistné $P_i, i = 1, \dots, m$.

Souhrnné nettopojistné sečtené přes všechny smlouvy činí

$$P_S = \sum_{i=1}^m P_i \doteq 257\,307\,544 \text{ Kč.} \quad (5.1)$$

Je třeba však podotknout, že máme souhrnné nettopojistné P_S stanoveno tak, aby pokrylo škody nastalé během roku 2011, nahlášené a uhrazené do roku 2012. Musíme ale zajistit i úhradu škod vzniklých v roce 2011, ale doposud nahlášených. Vytváříme tzv. *IBNR rezervu* (z anglického *incurred but not reported*) na pojištění plnění za škody, které nastaly, ale doposud nebyly nahlášený. Zahrnutím této rezervy do výpočtu nettopojistného by pojistitel měl být schopen plnit všechny své závazky plynoucí ze škodního roku 2011.

Výpočet IBNR rezervy vychází z předchozí zkušenosti pojistitele s vývojem škod. Využijeme kumulativní vývojový trojúhelník³, kde hodnota $X_{j,s}$ vyjadřuje celkový úhrn pojištění plnění za škody, které vznikly v roce j a byly uhrazeny do roku $j + s$, v našem případě $j = 2005, \dots, 2012$ a $s = 0, \dots, 7$.

Kumulativní trojúhelník je naznačen v tabulce 5.3, přičemž zvýrazněná hodnota (238 857 165 Kč, pole $X_{2011,1}$) představuje součet všech škod z našeho souboru dat. Metodou chain-ladder doplníme trojúhelník na čtverec. Část tohoto čtverce vidíme v tabulce 5.4.

vznik/vývoj	0	1	2	...	7
2005	117 535 923	150 796 700	190 467 908	...	216 662 866
2006	143 043 846	185 957 000	230 586 680	...	
⋮	⋮	⋮	⋮		
2010	181 158 376	248 505 889	304 147 303		
2011	183 736 281	238 857 165			
2012	209 250 595				

Tabulka 5.3: Kumulativní vývojový trojúhelník, hodnoty v Kč.

vznik/vývoj	0	1	2	...	7
2005	117 535 923	150 796 700	190 467 908	...	216 662 866
2006	143 043 846	185 957 000	230 586 680	...	263 683 552
⋮	⋮	⋮	⋮		⋮
2010	181 158 376	248 505 889	304 147 303	...	348 474 697
2011	183 736 281	238 857 165	296 120 530	...	339 278 076
2012	209 250 595	273 144 719	338 628 146	...	387 980 887

Tabulka 5.4: Kumulativní vývojový trojúhelník doplněný na čtverec, hodnoty v Kč.

Předpokládáme, že vývoj škod není po sedmi letech ukončen. Provedeme tedy úpravu hodnoty $\hat{X}_{2011,7}$ a tím dostaneme odhad $\hat{X}_{2011,\infty}$. Úprava spočívá v přičtení

³Po teoretické stránce vycházíme ze zdroje [3] (část III.1), zde nalezneme také metody k nalezení odhadu rezervy. Potřebná data do kumulativního vývojového trojúhelníku poskytl vedoucí diplomové práce.

3% rizikové přírážky⁴, pro konkrétní hodnoty dostáváme

$$\widehat{X}_{2011,\infty} = 1,03 \cdot \widehat{X}_{2011,7} \doteq 1,03 \cdot 339\,278\,076 \doteq 349\,456\,419 \text{ Kč.} \quad (5.2)$$

Výše IBNR rezervy tedy je

$$\text{IBNR} = \widehat{X}_{2011,\infty} - X_{2011,1} = 349\,456\,419 - 238\,857\,165 = 110\,599\,254 \text{ Kč.} \quad (5.3)$$

Nyní provedeme kalibraci modelu tak, aby souhrnné nettopojistné bylo rovno součtu všech škod na smlouvách z našeho datového souboru navýšeného o IBNR rezervu. Najdeme tedy konstantu kal tak, aby platilo

$$\text{IBNR} + \sum_{k=1}^m (z_k + Z_k) = P_S \cdot kal. \quad (5.4)$$

Hledaná konstanta je

$$kal = \frac{\text{IBNR} + \sum_{k=1}^m (z_k + Z_k)}{P_S} = \frac{110\,599\,254 + 238\,857\,165}{257\,307\,544} \doteq 1,358. \quad (5.5)$$

Touto konstantou kal přenásobíme nettopojistné na všech smlouvách, takže pro i -tou smlouvu dostáváme nettopojistné P_i^* spočtené na základě kalibrovaného modelu

$$P_i^* = P_i \cdot kal. \quad (5.6)$$

Sečtením P_i^* ($i = 1, \dots, m$) přes všechny smlouvy pak dostáváme nové souhrnné nettopojistné, označíme jej P_S^* . Mezi P_S a P_S^* platí vztah

$$P_S^* = P_S \cdot kal. \quad (5.7)$$

Při výpočtu IBNR rezervy jsme již zohlednili 3% rizikovou přírážku. Další rizikovou přírážku již nebudeme započítávat a za rizikové pojistné budeme považovat P_S^* .

Ke stanovení bruttopojistného již zbývá jenom zahrnout k rizikovému pojistnému P_S^* nákladové složky a zisk. Předpokládejme, že fixní a variabilní náklady nezávislé na pojistném celkem činí 70 mil. Kč, provize uvažujme 15 %, zisk 2 % a příspěvek do fondu zábrany škod 3 % bruttopojistného. Podle vztahu (1.28) spočteme bruttopojistné

$$B_S = \frac{P_S^* + \text{náklady}_n}{1 - \text{náklady}_z - h - z} = \frac{349\,456\,419 + 70\,000\,000}{1 - 0,15 - 0,02 - 0,03} = 524\,320\,523 \text{ Kč.} \quad (5.8)$$

Vztah mezi bruttopojistným B_S a rizikovým pojistným P_S^* lze pomocí předchozího vztahu vyjádřit takto: $B_S = P_S^* \cdot 1,50039$. Pokud označíme bruttopojistné na i -té smlouvě B_i , pak jej spočteme

$$B_i = P_i^* \cdot 1,50039. \quad (5.9)$$

⁴Tato riziková přírážka bývá nazývána *tail faktor*.

Toto je tedy finální bruttopojistné, které budeme požadovat po klientovi.

Na českém trhu je zvykem využívat systém bonus-malus, příklad je uveden v tabulce 5.1. Klient tedy platí pojistné po odečtení slevy/přičtení přírážky.

Pokud bychom chtěli klientovi ukázat cenu před slevou a po slevě, provedli bychom nejprve navýšení bruttopojistného na $B_i/(1 - sleva)$ a o tomto bychom prohlásili, že je pojistné před slevou. Pojistné po slevě (tedy pojistné požadované k úhradě) je B_i .

Závěr

Cílem práce bylo popsat různé způsoby výpočtu pojistného v pojištění odpovědnosti z provozu motorového vozidla, při výpočtu pojistného využít zobecněný lineární model a pak na reálných datech použít vhodné metody a stanovit pojistné.

Nejprve jsme se seznámili s obecnými principy v neživotním pojištění, dále jsme se již soustředili na pojištění odpovědnosti z provozu motorového vozidla. Zvláště jsme se zabývali metodami pro odhad frekvence škod a výše škody. Pro frekvenci škod jsme uvedli nejjednodušší výpočet, který nebere v úvahu žádné proměnné (I_f) a dále přístup, který již zohledňuje kategorizaci smluv v rámci proměnných (II_f). Důležitým přístupem k modelaci frekvence škod je zobecněný lineární model, kterému je věnována 2. kapitola.

Dále jsme se zabývali odhadem výše škody. Podobně jako u frekvence jsme uvedli nejjednodušší způsob výpočtu bez zahrnutí proměnných (I_z) a dále způsob, kdy se využívá kategorií v rámci proměnných (II_z). Také zde je možné využít zobecněný lineární model.

V další části jsme se věnovali stanovení nettopojistného, nejprve jsme uvedli postup bez využití výpočtu frekvence a výše škody. Pak jsme uvedli, že nettopojistné lze spočítat jako součin frekvence a výše škody. Na příkladu jsme vysvětlili nevhodnost jednotného nettopojistného.

Uvedli jsme vztah pro výpočet bruttopojistného. V závěru první kapitoly jsme zmínili americkou kredibilitu.

Ve druhé kapitole jsme stručně shrnuli klasický lineární model, na jehož základě jsme popsali zobecněný lineární model. Dále jsme uvedli typicky používanou metodu pro odhad parametrů modelu, a to metodu maximální věrohodnosti. V následující části jsme zmínili nástroje vhodné pro posouzení kvality zobecněného lineárního modelu. V závěru jsme uvedli dva konkrétní zobecněné lineární modely, kde jeden lze využít při modelování frekvence škod a druhý při modelování výše škody.

V praktické části (kapitola 3–5) jsme aplikovali výše popsané metody na reálná data, přičemž jsme se zabývali odděleně malými a velkými škodami.

U malých škod jsme počítali jednak frekvenci I_f a II_f a jednak jsme se zabývali nalezením vhodného zobecněného lineárního modelu. Statisticky nejpřesnější byl zobecněný lineární model s deseti proměnnými, kde jedna z proměnných nabývala hodnot „žena“, „muž“ a „firma“, což je v rozporu s pokynem k uplatňování směrnice Rady 2004/113/ES. Pozměnili jsme proto tuto proměnnou tak, aby rozlišovala pouze mezi fyzickou osobou a firmou. Tento upravený model pro výpočet frekvence jsme označili „GLM“.

Výši škody jsme rovněž počítali jako I_z a II_z , ale zjistili jsme, že kromě jednoho případu mezi kategoriemi zvolených proměnných není statisticky významný rozdíl

ve výši škod. Zavedli jsme proto ${}^{II}z$ (P) s jedinou proměnnou. Dále jsme vytvořili zobecněné lineární modely, které se však od sebe hodnotícími kritérii lišily velmi málo.

Poté jsme srovnali různé přístupy k výpočtu nettopojistného (tabulka 3.5) a zjistili jsme, že nejbližší k součtu malých škod je souhrnné nettopojistné spočtené pomocí frekvence „GLM“ a severity ${}^{II}z$ (P).

U velkých škod jsme frekvenci škod odhadli jako ${}^I f_V$. Dále jsme velké škody srovnávali s tržními daty za období 2000–2008. Dospěli jsme k tomu, že naše data lze považovat za reprezentativní vzorek dat tržních. Tohoto jsme využili při výpočtu očekávané výše škody — aplikovali jsme zde kredibilitní vztah.

V poslední kapitole jsme se nejprve zabývali systémem bonus-malus. Jak je tento systém na českém trhu používán (typicky za každý rok bez nehody sleva vzroste o 5 procentních bodů), jsme neshledali spravedlivým (zejména pro první bonusové třídy). V závěru této kapitoly jsme se věnovali stanovení bruttopojistného, přičemž jsme zohlednili IBNR rezervu.

Seznam použité literatury

- [1] CIPRA, T. (2006). *Finanční a pojistné vzorce*. První vydání. Grada, Praha. ISBN 80-247-1633-X.
- [2] CIPRA, T. (1999). *Pojistná matematika: teorie a praxe*. První vydání. Ekopress, Praha. ISBN 80-861-1917-3.
- [3] MANDL, P. a MAZUROVÁ, L. (1999). *Matematické základy neživotního pojištění*. První vydání. Matfyzpress, Praha. ISBN 80-858-6342-1.
- [4] MCCULLAGH, P. a NELDER, J. A. (1989). *Generalized linear models*. Second edition. Chapman and Hall, Boca Raton. ISBN 04-123-1760-5.
- [5] OHLSSON, E. a JOHANSSON, B. (2010). *Non-life insurance pricing with generalized linear models*. Springer, New York. ISBN 978-3-642-10790-0.

Příloha A

Tabulky

GLM 4

	Estimate	Standard Error	z-Statistic	P-Value
1	-2.73281	0.07947	-34.3879	3.82573×10^{-259}
$x_1[18]$	1.37262	0.386232	3.55388	0.000379598
$x_1[19]$	0.876975	0.31475	2.78626	0.00533195
$x_1[< 22]$	0.559807	0.196456	2.84953	0.00437841
$x_1[< 26]$	0.156743	0.113553	1.38035	0.167479
$x_1[< 30]$	-0.123231	0.0909861	-1.35439	0.175612
$x_1[< 50]$	-0.105991	0.0481453	-2.20148	0.0277023
$x_1[< 70]$	-0.297424	0.0543697	-5.4704	4.49016×10^{-8}
$x_1[\geq 70]$	-0.15496	0.0688502	-2.25068	0.0244061
$x_2[M]$	-0.226512	0.0374047	-6.05572	1.39795×10^{-9}
$x_3[\leq 1000]$	-0.574908	0.121199	-4.74351	2.10049×10^{-6}
$x_3[\leq 1350]$	-0.254589	0.0815674	-3.12121	0.00180111
$x_3[\leq 1400]$	-0.284732	0.0836791	-3.40266	0.000667326
$x_3[\leq 1850]$	-0.179831	0.0796737	-2.25709	0.0240025
$x_3[\leq 2500]$	-0.0128788	0.0775553	-0.16606	0.86811
$x_4[< 25000]$	-0.369963	0.0429203	-8.61975	6.70994×10^{-18}
$x_4[< 500000]$	-0.0629084	0.0439507	-1.43134	0.152333

Tabulka A.1: Výstup „GLM 4“, sloupec Estimate označuje odhady jednotlivých parametrů modelu.

GLM 6

	Estimate	Standard Error	z-Statistic	P-Value
1	-3.13991	0.162858	-19.2801	7.89386×10^{-83}
$x_1[18]$	1.41324	0.386432	3.65715	0.000255037
$x_1[19]$	0.911443	0.314923	2.89417	0.00380158
$x_1[< 22]$	0.597744	0.196779	3.03764	0.00238436
$x_1[< 26]$	0.190102	0.113964	1.66808	0.0952992
$x_1[< 30]$	-0.0946718	0.0913828	-1.03599	0.300206
$x_1[< 50]$	-0.0851376	0.0486841	-1.74878	0.0803297
$x_1[< 70]$	-0.273133	0.0549236	-4.97297	6.59348×10^{-7}
$x_1[\geq 70]$	-0.128547	0.0693403	-1.85385	0.06376
$x_2[M]$	-0.229507	0.0374266	-6.1322	8.66738×10^{-10}
$x_3[\leq 1000]$	-0.394592	0.149544	-2.63863	0.00832427
$x_3[\leq 1350]$	-0.100047	0.114673	-0.872457	0.382959
$x_3[\leq 1400]$	-0.153459	0.110524	-1.38846	0.164996
$x_3[\leq 1850]$	-0.0712807	0.101397	-0.702984	0.482066
$x_3[\leq 2500]$	0.0425857	0.0921757	0.462006	0.644077
x_4	0.000160883	0.000589999	0.272683	0.785097
x_5	0.000153256	0.000038634	3.96686	0.0000728244
$x_6[< 25000]$	-0.365105	0.0430803	-8.47499	2.35093×10^{-17}
$x_6[< 500000]$	-0.0592418	0.0440484	-1.34493	0.178649

Tabulka A.2: Výstup „GLM 6“, sloupec Estimate označuje odhady jednotlivých parametrů modelu.

GLM 8

	Estimate	Standard Error	z-Statistic	P-Value
1	-3.30586	0.271646	-12.1698	4.50449×10^{-34}
$x_1[18]$	1.40656	0.386506	3.63915	0.000273535
$x_1[19]$	0.916037	0.314987	2.90817	0.00363548
$x_1[< 22]$	0.591861	0.196937	3.00533	0.00265289
$x_1[< 26]$	0.180665	0.114212	1.58184	0.113686
$x_1[< 30]$	-0.108632	0.0916457	-1.18535	0.23588
$x_1[< 50]$	-0.0952574	0.0490313	-1.94279	0.052042
$x_1[< 70]$	-0.275224	0.055063	-4.99834	5.78263×10^{-7}
$x_1[\geq 70]$	-0.125161	0.0694575	-1.80197	0.0715494
$x_2[M]$	-0.23099	0.0374292	-6.17137	6.77×10^{-10}
$x_3[\leq 1000]$	-0.153063	0.15656	-0.977665	0.32824
$x_3[\leq 1350]$	0.134998	0.123238	1.09542	0.273332
$x_3[\leq 1400]$	0.0725802	0.118877	0.610548	0.541499
$x_3[\leq 1850]$	0.0940016	0.106064	0.886268	0.375473
$x_3[\leq 2500]$	0.0745898	0.0923357	0.807811	0.419199
x_4	0.00117666	0.000601427	1.95645	0.0504126
x_5	0.000144396	0.0000411326	3.5105	0.00044726
$x_6[B]$	-0.0882171	0.215389	-0.409571	0.682121
$x_6[D]$	0.113352	0.215128	0.526904	0.59826
$x_7[< 25000]$	-0.370286	0.0430916	-8.59299	8.47351×10^{-18}
$x_7[< 500000]$	-0.0614742	0.0440487	-1.3956	0.162836
$x_8[N]$	-0.118634	0.045831	-2.5885	0.00963937

Tabulka A.3: Výstup „GLM 8“, sloupec Estimate označuje odhady jednotlivých parametrů modelu.

GLM 10

	Estimate	Standard Error	z-Statistic	P-Value
1	-3.29791	0.30516	-10.8072	3.18379×10^{-27}
$x_1[-1]$	-8.50774	104.19	-0.0816564	0.93492
$x_1[-0.9]$	-8.7442	209.888	-0.0416613	0.966769
$x_1[-0.7]$	-8.57584	110.787	-0.0774087	0.938298
$x_1[-0.6]$	0.801728	0.953458	0.840864	0.400424
$x_1[-0.5]$	0.281826	0.726881	0.387719	0.698224
$x_1[-0.4]$	0.178403	0.733783	0.243128	0.807906
$x_1[-0.3]$	0.596505	0.293931	2.0294	0.0424172
$x_1[-0.2]$	0.526306	0.219444	2.39836	0.0164688
$x_1[-0.1]$	0.335099	0.197683	1.69514	0.0900497
$x_1[0]$	0.626248	0.0800562	7.82261	5.17377×10^{-15}
$x_1[0.05]$	0.268611	0.081534	3.29447	0.000986091
$x_1[0.1]$	0.22462	0.0732056	3.06834	0.00215253
$x_1[0.15]$	0.123422	0.0749998	1.64563	0.0998406
$x_1[0.2]$	0.212041	0.0662988	3.19826	0.00138258
$x_1[0.25]$	0.236209	0.064545	3.6596	0.00025261
$x_1[0.3]$	0.208213	0.0618149	3.36832	0.000756267
$x_1[0.35]$	0.209354	0.0583618	3.58718	0.000334275
$x_1[0.4]$	0.0622261	0.0659413	0.943658	0.345344
$x_1[0.45]$	0.0781796	0.0588898	1.32756	0.184324
x_2	0.00139165	0.000607812	2.2896	0.0220443
x_3	0.000117677	0.0000454264	2.59049	0.00958404
$x_4[B]$	-0.0902379	0.244828	-0.368577	0.712443
$x_4[D]$	0.108906	0.244284	0.445816	0.65573
$x_5[\leq 1000]$	-0.138221	0.158956	-0.869552	0.384545
$x_5[\leq 1350]$	0.120965	0.12581	0.96149	0.336306
$x_5[\leq 1400]$	0.085122	0.121167	0.702519	0.482356
$x_5[\leq 1850]$	0.0930765	0.10809	0.861104	0.389181
$x_5[\leq 2500]$	0.0824136	0.0938506	0.878136	0.37987
$x_6[< 25000]$	-0.319868	0.0445019	-7.18773	6.58765×10^{-13}
$x_6[< 500000]$	-0.0382818	0.0452444	-0.846112	0.39749
$x_7[IC]$	0.0496166	0.0720062	0.689061	0.490785
$x_7[M]$	-0.188691	0.0385302	-4.89722	9.71997×10^{-7}
$x_8[18]$	1.23046	0.390466	3.15127	0.00162563
$x_8[19]$	0.744513	0.318877	2.3348	0.019554
$x_8[< 22]$	0.52988	0.200798	2.63886	0.00831841
$x_8[< 26]$	0.162927	0.122345	1.3317	0.182958
$x_8[< 30]$	-0.0764813	0.099278	-0.770375	0.441077
$x_8[< 50]$	-0.0184393	0.0579338	-0.318283	0.75027
$x_8[< 70]$	-0.161397	0.0606044	-2.66313	0.00774186
$x_9[N]$	-0.0791361	0.0476935	-1.65926	0.097063
$x_{10}[0]$	-0.288234	0.0442754	-6.51003	7.51354×10^{-11}

Tabulka A.4: Výstup „GLM 10“, sloupec Estimate označuje odhady jednotlivých parametrů modelu.

GLM

	Estimate	Standard Error	z-Statistic	P-Value
1	-3.46334	0.303854	-11.398	4.27634×10^{-30}
$x_1[-1]$	-8.49452	104.287	-0.0814533	0.935081
$x_1[-0.9]$	-8.75751	209.888	-0.0417247	0.966718
$x_1[-0.7]$	-8.52021	110.636	-0.0770113	0.938615
$x_1[-0.6]$	0.797467	0.953504	0.836354	0.402956
$x_1[-0.5]$	0.321625	0.72684	0.442498	0.658129
$x_1[-0.4]$	0.226306	0.73368	0.308454	0.757737
$x_1[-0.3]$	0.632533	0.293831	2.15271	0.0313416
$x_1[-0.2]$	0.548336	0.219401	2.49925	0.0124458
$x_1[-0.1]$	0.362774	0.19759	1.836	0.0663577
$x_1[0]$	0.655398	0.0798595	8.20689	2.26999×10^{-16}
$x_1[0.05]$	0.29482	0.0813693	3.62323	0.000290945
$x_1[0.1]$	0.251898	0.0730002	3.45065	0.00055924
$x_1[0.15]$	0.144385	0.074877	1.9283	0.0538176
$x_1[0.2]$	0.231218	0.0661797	3.49379	0.00047622
$x_1[0.25]$	0.252077	0.0644719	3.90988	0.0000923436
$x_1[0.3]$	0.222181	0.0617565	3.59769	0.000321058
$x_1[0.35]$	0.217114	0.0583451	3.72121	0.000198271
$x_1[0.4]$	0.0713025	0.0659201	1.08165	0.279408
$x_1[0.45]$	0.0826568	0.0588861	1.40367	0.160417
x_2	0.00139379	0.000608821	2.28932	0.0220608
x_3	0.000112846	0.0000460271	2.45174	0.0142168
$x_4[B]$	-0.0881079	0.24484	-0.359859	0.718953
$x_4[D]$	0.110767	0.244292	0.453421	0.650245
$x_5[\leq 1000]$	-0.0975876	0.158884	-0.614208	0.539078
$x_5[\leq 1350]$	0.145584	0.125798	1.15728	0.247158
$x_5[\leq 1400]$	0.0994879	0.121199	0.820863	0.411724
$x_5[\leq 1850]$	0.09635	0.108128	0.891075	0.372889
$x_5[\leq 2500]$	0.0813614	0.0938881	0.866578	0.386173
$x_6[< 25000]$	-0.328546	0.0444911	-7.38453	1.52997×10^{-13}
$x_6[< 500000]$	-0.0444105	0.045238	-0.981709	0.326243
$x_7[IC]$	0.209129	0.064481	3.24327	0.00118168
$x_8[18]$	1.27375	0.390453	3.26225	0.00110533
$x_8[19]$	0.74658	0.318872	2.34131	0.019216
$x_8[< 22]$	0.550396	0.200776	2.74134	0.00611895
$x_8[< 26]$	0.194334	0.122208	1.5902	0.11179
$x_8[< 30]$	-0.0445881	0.0990825	-0.450009	0.652704
$x_8[< 50]$	0.0110438	0.0576258	0.191647	0.848019
$x_8[< 70]$	-0.139837	0.0604427	-2.31355	0.0206927
$x_9[N]$	-0.0755384	0.047709	-1.58331	0.11335
$x_{10}[0]$	-0.286155	0.0442708	-6.46375	1.02142×10^{-10}

Tabulka A.5: Výstup „GLM“, sloupec Estimate označuje odhady jednotlivých parametrů modelu.

Model — proměnné	AIC	LRI	<i>Pearson</i> χ^2	D
věk, měs	98566.0	0.573469	3972.20	3770.77
věk, měs, pal	98564.2	0.573493	3971.32	3765.35
věk, měs, pal, hmot	98565.9	0.573495	3969.74	3765.23
věk, měs, pal, hmot, nové	98567.2	0.573498	3967.03	3764.93
„10“	98602.4	0.573544	3954.16	3753.45

Tabulka A.6: Zobecněné lineární modely pro výši škody. Vysvětlení proměnných v prvním sloupci tabulky: věk = proměnná věk, měs = město, pal = palivo, nové = proměnná nové vozidlo. Proměnné byly použity jako u zobecněného lineárního modelu pro frekvenci škod. Model označený „10“ obsahuje stejné proměnné jako model „GLM“ při modelování frekvence škod.

	Estimate	Standard Error	z-Statistic	P-Value
1	10.1131	0.23497	43.0399	1.04063×10^{-404}
x_1 [B]	0.333358	0.232859	1.43159	0.152262
x_1 [D]	0.393619	0.23333	1.68697	0.0916099
x_2 [< 25000]	-0.150453	0.041185	-3.65311	0.000259085
x_2 [< 500000]	-0.138891	0.0421995	-3.2913	0.00099726
x_3 [18]	-0.166729	0.428479	-0.389118	0.697189
x_3 [19]	0.538954	0.257829	2.09036	0.0365857
x_3 [< 22]	0.390377	0.177506	2.19924	0.0278611
x_3 [< 26]	0.0453108	0.103472	0.437905	0.661456
x_3 [< 30]	-0.210701	0.0811941	-2.59503	0.00945831
x_3 [< 50]	-0.0490619	0.0397872	-1.23311	0.217536
x_3 [< 70]	-0.105489	0.045341	-2.32656	0.0199886
x_3 [\geq 70]	-0.0562253	0.0582496	-0.965248	0.33442

Tabulka A.7: Výstup pro zobecněný lineární model výše škody, sloupec Estimate označuje odhady jednotlivých parametrů modelu.