Klasifikacija rizika u osiguranju i bonus-malus sustav

Katić, Dunja

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku
Sveučilišni diplomski studij
Financijska matematika i statistika

Dunja Katić

Klasifikacija rizika u osiguranju i bonus-malus sustav

Diplomski rad

Osijek, 2022.

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku
Sveučilišni diplomski studij
Financijska matematika i statistika

Dunja Katić

Klasifikacija rizika u osiguranju i bonus-malus sustav

Diplomski rad

Mentor: doc. dr. sc. Danijel Grahovac

Osijek, 2022.

Sadržaj

1 Uvod

Svaka odrasla osoba vrlo dobro poznaje pojam osiguranja. Kada kažemo osiguranje, mnogi će pomisliti na zdravstveno osiguranje, osiguranje automobila ili neke druge imovine te će tu riječ povezati s iznosom koji je nužno platiti osiguravajućem društvu kako bismo od njega u slučaju nekakvog manje poželjnog događaja dobili određenu naknadu. Jasno je kako osiguranik priželjkuje da će iznos koji mora platiti biti što manji a da se istovremeno osiguravajuće društvo mora osigurati da će taj iznos biti dovoljan za pokrivanje eventualnih šteta u budućnosti. Dakle, cilj svakog osiguravajućeg društva je što bolje procijeniti klijenta te mu dodijeliti odgovarajući rizik na osnovu kojega će mu se izračunati odgovarajuća premija. U ovom radu objasnit ćemo tijek klasifikacije osiguranika koji započinje prikupljanjem osnovnih podataka poput dobi, spola i ostalih karakteristika ovisno o kojoj vrsti osiguranja se radi i koje karakteristike su uistinu bitne za to osiguranje. Na temelju tih karakteristika se pomoću jednostavnih generaliziranih linearnih modela procjenjuje rizičnost klijenta i dodjeljuje mu se određena a priori premija. Posebnu važnost pridodat ćemo osiguranju motornih vozila. Kod takve vrste osiguranja je a priori premiju potrebno dodatno korigirati i to se vrši putem bonus-malus premijskih stupnjeva. Njima se smanjuje a priori premija osiguranika koji tijekom protekle godine nisu imali prijavljenih šteta, odnosno povećava a priori premija osiguranika koji su tijekom protekle godine prijavili jednu ili više šteta. Bitna karakteristika bonus-malus premijskog sustava je ta da je on Markovljev lanac u diskretnom vremenu te kao takav ima određena svojstva korisna za osiguravajuće društvo. Stacionarna distribucija Markovljevog lanca omogućuje uvid u kretanje klijenata po premijskim stupnjevima tijekom vremena na temelju kojih osiguravajuće društvo ovisno o svojim ciljevima i potrebama može korigirati premiju po stupnjevima i tako upravljati svojim prihodima.

2 Generalizirani linearni model

Generalizirane linearne modele (GLM) možemo promatrati kao proširenje linearnih regresijskih modela zbog čega ćemo najprije navesti nekoliko ključnih obilježja linearnih regresijskih modela. Definicije i primjeri u ovom poglavlju prate [3] i [5].

Linearni regresijski model je model čije je očekivanje linearna funkcija parametara, tj. očekivanje zavisne varijable Y možemo modelirati pomoću poznatih vrijednosti $x_0, x_1, \ldots, x_k$ nezavisnih varijabli $X_0, \ldots, X_k$ u obliku

$$E(Y) = \beta_0 x_0 + \beta_1 x_1 + \cdots + \beta_k x_k, \tag{1}$$

gdje su $\beta_1, \ldots, \beta_k$ nepoznati parametri koje je potrebno procijeniti. Parametre procjenjujemo metodom najmanjih kvadrata. Dodatna pretpostavka je da varijanca ne ovisi o $\mathbf{x}$, tj. $Var(Y) = \sigma^2$. Tada linearni regresijski model možemo zapisati na sljedeći način

$$Y = \beta_0 x_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon,$$

gdje je ϵ nemjerljiva slučajna varijabla greške za koju pretpostavljamo da vrijedi $E[\epsilon] = 0$ i $Var(\epsilon) = \sigma^2$. Ukoliko u vektoru $\mathbf{x} = (x_0, \ldots, x_k)$ prvi element x_0 postavimo na 1, u jednadžbu uvodimo konstantni član. Tada model poprima sljedeći oblik

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon.$$

Ovisno o broju prediktora, razlikujemo jednostavni i složeni regresijski model. Ukoliko je k=1 i primjerice $\mathbf{x} = (1, x_1)$ dobivamo jednostavni regresijski model sljedećeg oblika

$$Y = \beta_0 + \beta x_1 + \epsilon,$$

gdje je $E[\epsilon] = 0$ i $Var(\epsilon) = \sigma^2$.

Ukoliko slučajnu varijablu Y želimo opisati na temelju $x_{1i}, \ldots, x_{ik}$, $i = 1, \ldots, n, k \in \mathbb{N}$, dobivamo složeni regresijski model sljedećeg oblika

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i,$$

gdje je $E[\epsilon_i] = 0$ i $Var(\epsilon_i) = \sigma^2$ za svaki $i = 1, \ldots, n$.

U aktuarskoj matematici česte su situacije u kojima navedene pretpostavke nisu zadovoljene zbog čega je nužno generalizirati linearni regresijski model. U tu svrhu ćemo najprije definirati eksponencijalnu familiju distribucija.

2.1 Eksponencijalna familija distribucija

Definicija 1. *Neka je Y slučajna varijabla čija vjerojatnosna distribucija ovisi o parametru θ. Kažemo da distribucija slučajne varijable Y pripada eksponencijalnoj familiji distribucija ukoliko se njena funkcija gustoće može zapisati u sljedećem obliku:*

$$f(y;\theta) = s(y)t(\theta)e^{a(y)b(\theta)}, \tag{2}$$

gdje su a, b, s i t poznate funkcije.

Definiramo li $s(y)$ kao $s(y) = e^{d(y)}$ i $t(\theta)$ kao $t(\theta) = e^{c(\theta)}$, tada funkcija (2) poprima sljedeći oblik:

$$\begin{aligned} f(y;\theta) &= s(y)t(\theta)e^{a(y)b(\theta)} \\ &= e^{d(y)}e^{c(\theta)}e^{a(y)b(\theta)} \\ &= \exp[a(y)b(\theta) + c(\theta) + d(y)]. \end{aligned} \tag{3}$$

Ukoliko je $a(y) = y$, odnosno ukoliko je funkcija a identiteta, kažemo da distribucija ima kanonsku (standardnu) formu i tada $b(\theta)$ nazivamo prirodnim parametrom distribucije. Ako uz parametar θ postoje i drugi parametri, njih smatramo poznatima te ih ne procjenjujemo. Takve parametre nazivamo nusparametrima.
Eksponencijalnoj familiji distribucija pripadaju binomna, Bernoullijeva, Poissonova, normalna te gama distribucija.

Primjer 1. *Neka je Y slučajna varijabla s binomnom distribucijom s parametrima n i p. Funkcija gustoće slučajne varijable Y dana je s*

$$h(y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y \in \{0, 1, \ldots, n\}.$$

Želimo li pokazati da binomna distribucija pripada eksponencijalnoj klasi distribucija, moramo moći pripadnu funkciju gustoće zapisati u formi (3).

$$\begin{aligned} \binom{n}{y} p^y (1-p)^{n-y} &= \exp[\ln \binom{n}{y} p^y (1-p)^{n-y}] \\ &= \exp[\ln \binom{n}{y} + y + (n-y)\ln(1-p)] \\ &= \exp[\ln \frac{n!}{y!(n-y)!} + y + n\ln(1-p) - y\ln(1-p)] \\ &= \exp[\ln \frac{n!}{y!(n-y)!} + y(-\ln(1-p)) + n\ln(1-p)] \\ &= \exp[y \ln \frac{p}{1-p} + n\ln(1-p) + \ln \frac{n!}{y!(n-p)!}] \end{aligned}$$

Slijedi da je:

$$a(y) = y, \quad b(p) = \ln \frac{p}{1-p}, \quad c(p) = n\ln(1-p), \quad d(y) = \ln \frac{n!}{y!(n-p)!}.$$

Dakle, binomna distribucija pripada eksponencijalnoj klasi distribucija.

Za $n = 1$ i $y \in \{0, 1\}$, jednadžba se svodi na funkciju gustoće iz Bernoullijeve distribucije i ima sljedeći oblik

$$\exp[y \ln \frac{p}{1-p} + \ln(1-p) + \ln \frac{n!}{y!(1-p)!}].$$

Vrijedi:

$$a(y) = y, \quad b(p) = \ln \frac{p}{1-p}, \quad c(p) = n \ln(1-p), \quad d(y) = 0.$$

Dakle, Bernoullijeva distribucija također pripada eksponencijalnoj klasi distribucija.

Primjer 2. *Neka je Y slučajna varijabla s Poissonovom distribucijom s parametrom $\lambda > 0$. Funkcija gustoće slučajne varijable Y je dana s*

$$p_i = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i \in \{0, 1, \dots\}.$$

Pripadnu funckiju gustoće možemo zapisati kao

$$\begin{aligned} e^{-\lambda} \frac{\lambda^i}{i!} &= \exp[-\lambda + \ln \lambda^i - \ln i!] \\ &= \exp[i \ln \lambda - \lambda - \ln i!], \end{aligned}$$

odakle slijedi:

$$a(i) = e^i, \quad b(\lambda) = \ln \lambda, \quad c(\lambda) = -\lambda, \quad d(i) = -\ln i!.$$

Zaključujemo kako Poissonova distribucija pripada eksponencijalnoj klasi distribucija.

2.2 Pretpostavke generaliziranog linearnog modela

Generalizirani linearni model definiran je u terminima međusobno nezavisnih slučajnih varijabli $Y_1, \dots, Y_n$ čije distribucije moraju zadovoljavati sljedeće zahtjeve:

1) Distribucija svakog Y_i pripada eksponencijalnoj familiji distribucija i može se parametrizirati jednim parametrom θ_i, tj.

$$f_{Y_i}(y; \theta_i) = exp[y b_i(\theta_i) + c_i(\theta_i) + d_i(y)], \quad i = 1, \dots, n.$$

2) Distribucije svih Y_i istog su tipa, tj. b_i, c_i i d_i ne ovise o indeksu i zbog čega ćemo ih označiti sa b, c i d. Zajednička funkcija gustoće slučajnih varijabli $Y_1, \dots, Y_n$ tada ima sljedeći oblik

$$f(y_1, \dots, y_n; \theta_1, \dots, \theta_n) = exp[\sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i)].$$

Skup parametara $\{\theta_1, \dots, \theta_n\}$ nije nam od direktnog interesa za model zbog čega ćemo ga zamijeniti manjih skupom parametara $\{\beta_1, \dots, \beta_k\}$. Pretpostavimo da je $E[Y_i] = \mu_i$, pri čemu je μ_i funkcija od θ_i. Za generalizirani linearni model postoji transformacija od μ_i takva da je

$$\begin{aligned} g(\mu_i) &= \mathbf{x}_i^T \boldsymbol{\beta} \\ &= \begin{bmatrix} x_{i1}, \dots, x_{ik} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \\ &= x_{i1}\beta_1 + \dots + x_{ik}\beta_k, \end{aligned}$$

gdje je g monotona i diferencijabilna funkcija koja se naziva link funkcija ili funkcija veze. Vektor $\mathbf{x}_i^T$ je $i-$ti redak matrice $\mathbf{X}$ koju zovemo matricom dizajna i sljedećeg je oblika

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nk} \end{bmatrix}.$$

Najčešće korištene link funkcije ovisno o distribucijama prikazane su u Tablici 1.

Distribucija	Link funkcija
Normalna	$g(\mu) = \mu$
Poissonova	$g(\mu) = \log(\mu)$
Binomna	$g(\mu) = \log(\frac{\mu}{1-\mu})$
Gama	$g(\mu) = \frac{1}{\mu}$

Tablica 1: Link funkcija po distribucijama

Navedene link funkcije dobro rade za svaku od gornjih distribucija, ali nije nužno koristiti točno njih. Primjerice, možemo koristiti identitetu kao link funkciju za Poissonovu distribuciju, logaritamsku funkciju kao link funkciju za gama distribuciju, itd. Pri odabiru funkcije veze nužno je razmotriti njene posljedice izbora na moguće vrijednosti od μ. Tako kod Poissonove distribucije μ mora biti pozitivan. Ako upotrijebimo logaritamsku funkciju veze za Poissonovu distribuciju, vrijedi da je μ eksponencijalna funkcija i sigurno je pozitivna. Međutim, uzmemo li identitetu kao link funkciju za Poissonovu distribuciju, to neće biti tako.

2.3 Procjena koeficijenata metodom maksimalne vjerodostojnosti

Koeficijenti u generaliziranom linearnom modelu obično se procjenjuju upotrebom procjene maksimalne vjerodostojnosti (Maximum Likelihood Method). Riječ je o popularnoj statističkoj metodi koja je primjenjiva na najveći dio teorijskih distribucija. Otkrio ju je engleski statističar A.Fisher (1890.-1962.). Da bi objasnili ovu metodu prvo ćemo definirati funkciju vjerodostojnosti.

Definicija 2. *Za danu realizaciju $x = (x_1, \dots, x_n)$ slučajnog uzorka $(X_1, \dots, X_n)$ iz gustoće $f(\boldsymbol{x}; \boldsymbol{\theta})$ je funkcija vjerodostojnosti*

$$L(\boldsymbol{\theta}; \boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\theta}),$$

gustoća za fiksni $\boldsymbol{x}$ razmatrana kao funkcija parametra.

Dakle, funkcija vjerodostojnosti definirana je na sljedeći način:

$$L_{\mathbf{X}}(\boldsymbol{\theta}) = \begin{cases} \prod_{i=1}^{n} P_{\boldsymbol{\theta}}\{X_i = x_i\}, & X_i \text{ diskretna slučajna varijabla} \\ \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}), & X_i \text{ neprekidna slučajna varijabla.} \end{cases} \tag{4}$$

Metoda maksimalne vjerodostojnosti temelji se na traženju maksimuma funkcije vjerodostojnosti. Bitno je znati da umjesto traženja maksimuma funkcije vjerodostojnosti, radi jednostavnosti možemo promatrati ekvivalentan problem, tj. možemo tražiti maksimum logaritma funkcije vjerodostojnosti.

Primjer 3. *Iz (2) slijedi da je funkcija vjerodostojnosti za jednostavan slučajan uzorak iz Poissonove distribucije dana sljedećim izrazom*

$$L_{\boldsymbol{X}}(\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.$$

Tada je:

$$\begin{aligned} l_{\boldsymbol{X}}(\lambda) = \ln L_{\boldsymbol{X}}(\lambda) &= -n\lambda + \ln\Big(\frac{1}{\prod_{i=1}^{n} x_i!}\Big) + \ln(\lambda^{\sum_{i=1}^{n} x_i}) \\ &= -n\lambda + \ln\Big(\frac{1}{\prod_{i=1}^{n} x_i!}\Big) + \sum_{i=1}^{n} x_i \ln\lambda. \end{aligned} \tag{5}$$

U cilju maksimizacije funkcije $l_{\boldsymbol{X}}(\lambda)$, najprije deriviramo izraz (5)

$$l'_X(\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n.$$

Izjednačimo li ga s nulom dobivamo

$$\frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0,$$

odakle slijedi:

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Budući da je

$$l''_\lambda = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i \leq 0,$$

funkcija vjerodostojnosti postiže maksimum u $\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X_n}$.

Dakle, $\overline{X_n}$ je procjenitelj za parametar λ iz Poissonove distribucije dobiven metodom maksimalne vjerodostojnosti.

2.4 Devijanca i odabir modela

Odluka koji model koristiti obično započinje razmatranjem devijance za niz modela. Devijancu D definiramo u terminima maksimalne vjerodostojnosti funkcije vjerodostojnosti $l(\boldsymbol{\beta})$ i modela s maksimalnim brojem parametara:

$$D = 2[l(\boldsymbol{\beta_{max}})) - l(\boldsymbol{\beta})],$$

gdje je $l(\boldsymbol{\beta_{max}})$ maksimalna vrijednost prirodnog logaritma funkcije vjerodostojnosti za model s najvećim brojem parametara. Dakle, ukoliko želimo usporediti dva modela možemo promatrati njihove devijance. Bolji model je onaj koji ima manju devijancu. Međutim, trebamo biti oprezni jer veći modeli uvijek imaju manju devijancu u odnosu na manje modele. Stoga, ukoliko odluku donosimo samo na temelju devijance uvijek ćemo kao najbolji model dobiti onaj koji uključuje sve promatrane nezavisne varijable. Zato je uz devijancu nužno promatrati i broj stupnjeva slobode nekog modela.

Iz tog razloga ćemo u cilju usporedbe dva modela koristiti i Anova test za generalizirane linearne modele. On u svojoj pozadini promatra i devijancu i broj stupnjeva slobode. Model bez nezavisnih varijabli nazivamo nul modelom. Dodajući k nezavisnih varijabli u taj model možemo istražiti doprinose li one kvaliteti modela.

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$H_1 : \beta_1 \neq 0, \beta_2 \neq 0 \ldots, \beta_k \neq 0$$

Navedene hipoteze mogu se interpretirati na sljedeći način:
H_0 : Nezavisna varijabla X_i, $\forall i = 1 \ldots, k$ nije od značaja za zavisnu varijablu Y.
H_1 : Nezavisna varijabla X_i, $\imath = 1 \ldots, k$ je od značaja za zavisnu varijablu Y.

Za donošenje odluke koji model odabrati, promatrat ćemo i Akaike informacijski kriterij (AIC). Prema njemu, najbolji model je onaj koji ima najnižu vrijednost AIC-a. Međutim, treba biti oprezan jer povećanjem broja parametara u modelu raste vrijednost AIC-a što je jasno iz same formule

$$AIC = -2l + 2k,$$

gdje je l maksimum funkcije vjerodostojnosti promatranog modela, a k broj procijenjenih parametara modela.

2.5 Primjena generaliziranih linearnih modela u aktuarstvu

Bitna razlika između linearnih i generaliziranih linearnih modela je u tome što generalizirani linearni modeli dozvoljavaju da distribucija zavisne varijable ne bude normalna. To je izrazito važno u aktuarskom radu gdje podaci vrlo često nemaju normalnu distribuciju. Tako se primjerice Poissonova distribucija koristi za modeliranje intenziteta smrtnosti, a binomna distribucija za početnu stopu smrtnosti. Kod neživotnog osiguranja, Poissonova distribucija se često koristi za modeliranje frekvencije šteta, a gama distribucija za veličinu šteta. Svrha analize podataka obično se sastoji u odlučivanju koje varijable su važni prediktori promatranog rizika, te zatim u kvantificiranju relacije između prediktora i rizika kako bi se odredile odgovarajuće premije. Stoga generalizirani linearni modeli predstavljaju glavni alat u modernoj aktuarskoj znanosti.

U nastavku ovoga rada ilustrirat ćemo primjer primjene generaliziranih linearnih modela u zdravstvenom osiguranju koristeći funkciju glm() iz paketa stats besplatnog softverskog okruženja R. U softverskim paketima postoje automatizirane procedure odabira najboljeg modela zasnovane na AIC-u i sličnim kriterijima. Aproksimacijske metode za generiranje „najboljeg“ modela koje ćemo primjenjivati u ovome radu su selekcijska procedura unaprijed (engl. Forward selection) i selekcijska procedura unatrag (engl. Backward elimination). Selekcijska procedura unaprijed započinje nul modelom u koji u svakom sljedećem koraku uključuje dodatnu varijablu, uspoređuje modele i proces nastavlja dok ne pronađe najbolji model. S druge strane, selekcijska procedura unatrag započinje maksimalnim modelom iz kojega u svakom sljedećem koraku eliminira varijable, uspoređuje modele međusobno i proces nastavlja dok ne pronađe najbolji među njima. Modele dobivene na taj način uspoređivat ćemo na temelju Anova testa i Akaike informacijskog kriterija. Na kraju ćemo interpretirati dobivene rezultate.

2.5.1 Primjena generaliziranih linearnih modela: zdravstveno osiguranje

U ovom ćemo potpoglavlju pokazati primjenu generaliziranih linearnih modela u zdravstvenom osiguranju. Na raspolaganju imamo 1339 klijenata, pri čemu je za svakog klijenta zabilježena dob, spol, bmi (indeks tjelesne mase), broj djece, regija iz koje dolazi te je li klijent pušač ili nije. Varijable dob, bmi i brojdjece su numeričke. Preostale varijable su kategorijalne te su njihovi opisi prikazani u tablici 2.

Varijabla	Opis varijable
spol	0-žensko, 1-muško
regija	stambeno područje osiguranika u SAD-u (0=sjeveroistok, 1=sjeverozapad, 2=jugoistok, 3=jugozapad)
pušenje	0=nepušač, 1=pušač

Tablica 2: Opis nezavisnih kategorijalnih varijabli

Cilj nam je izgraditi model u kojem je zavisna varijabla osiguravajućizahtjev. Budući da se američki sustav osiguranja razlikuje od onoga u Republici Hrvatskoj, potrebno je dodatno objasniti značenje zavisne varijable osiguravajućizahtjev. Osiguravajući zahtjev je službeni zahtjev osiguravajućem društvu kojim se traži isplata i pokrivanje bolničkih troškova na temelju uvjeta police osiguranja. Osiguravajuće društvo provjerava valjanost zahtjeva i zatim u slučaju odobrenja isplaćuje iznos osiguraniku ili stranci (u ime osiguranika). Dakle, zavisna varijabla osiguravajućizahtjev dijeli klijente na one koji su u promatranom periodu podnijeli zahtjev za pokrivanjem troškova (oznaka 1) i one koji nisu (oznaka 0). Logično je za pretpostaviti kako ona ima Bernoullijevu distribuciju. Budući da je Bernoullijeva distribucija poseban slučaj binomne distribucije, pri modeliranju ćemo koristiti generalizirani linearni model s funkcijom veze za binomnu distribuciju. Dakle, želimo pronaći model koji će dobro opisati o čemu ovisi hoće li klijent izvršiti zahtjev ili neće. Najprije trebamo promotriti odnos nezavisnih varijabli sa zavisnom.

	zahtjev iz osiguranja-ne	zahtjev iz osiguranja-da
žensko	285	377
muško	270	406

Tablica 3: Frekvencije varijable spol s obzirom na varijablu osiguravajućizahtjev

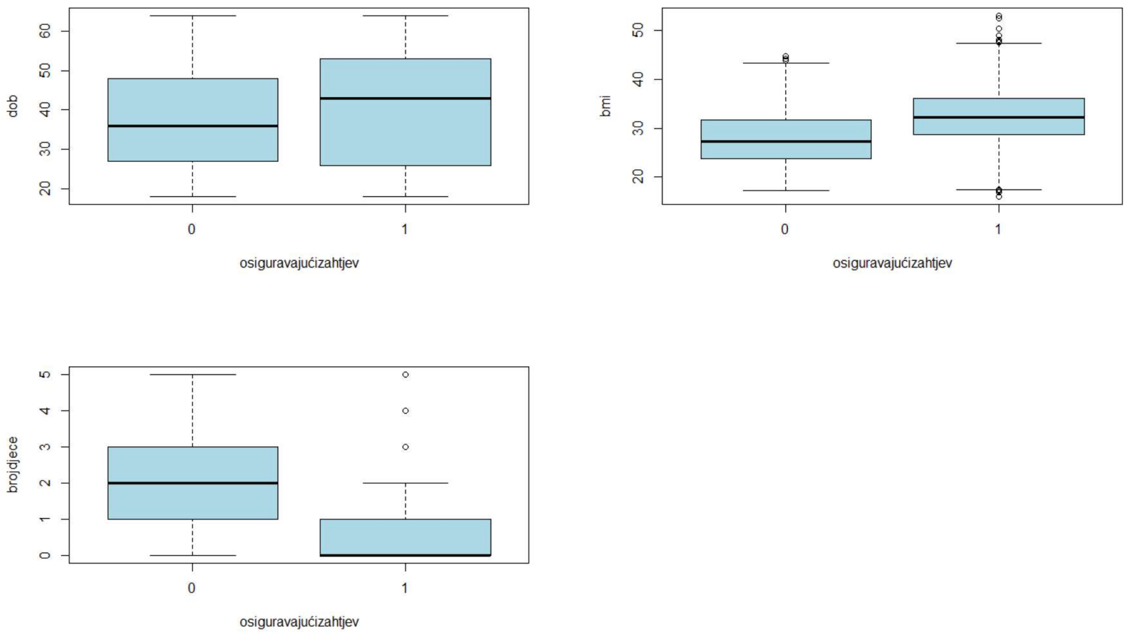
	zahtjev iz osiguranja-ne	zahtjev iz osiguranja-da
sjeveroistok	132	192
sjeverozapad	162	163
jugoistok	119	245
jugozapad	142	183

Tablica 4: Frekvencije varijable regija s obzirom na varijablu osiguravajućizahtjev

	zahtjev iz osiguranja-ne	zahtjev iz osiguranja-da
nepušač	530	534
pušač	25	249

Tablica 5: Frekvencije varijable pušenje s obzirom na varijablu osiguravajućizahtjev

Iz Tablice (3) možemo zaključiti kako je približno 60% klijenata i muškog i ženskog spola podnijelo zahtjev za naplatom iz osiguranja, tj. približno 40% klijenata i muškog i ženskog spola nije podnijelo zahtjev za naplatom iz osiguranja. Prema tome pretpostavljamo kako varijabla spol neće biti od interesa u našem modelu. Iz Tablice (5) jasno je kako baza podataka sadrži veći broj nepušača u odnosu na broj pušača. Također, približno je jednak broj nepušača koji su podnijeli i koji nisu podnijeli zahtjev. Međutim, postoji veliko odstupanje u broju pušača koji jesu i koji nisu podnijeli zahjev iz osiguranja. Tako je približno 91% ukupnih pušača podnijelo zahtjev iz osiguranja zbog čega pretpostavljamo kako će varijabla pušač biti od interesa pri modeliranju. Iz tablice (4) primjećujemo kako je unutar svake regije veći broj klijenata koji su podnijeli zahtjev od onih koji nisu. Najveće razlike u broju klijenata koji jesu i koji nisu podnijeli zahtjev su na jugoistoku SAD-a, a najmanje na sjeverozapadu. Preostaje nam još promotriti odnose nezavisnih varijabli dom, bmi i brojdjece sa zavisnom varijablom osiguravajućizahtjev. Odnosi su prikazani u Tablici 6.



Tablica 6: Boxplot nezavisnih numeričkih varijabli

Na grafičkom prikazu uočavamo kako osobe koje su podnijele zahtjev u prosjeku imaju veći broj godina i veći indeks tjelesne mase u odnosu na one koji nisu podnijeli zahtjev. Također, osobe koje nisu podnijele zahtjev u prosjeku imaju veći broj djece od osoba koje su isti podnijele. Pretpostavljamo kako će sve tri navedene numeričke varijable biti od interesa za naš model.

Na temelju prethodno provedene analize, kreirat ćemo model s varijablama koje su se pokazale značajnima. To su dob, bmi, brojdjece, pušenje i regija. Bitno je naglasiti kako u takvom modelu nije prisutan problem multikolinearnosti. Nazvat ćemo ga model 1. Selekcijskim procedurama tražimo model s najmanjom AIC vrijednosti. Tako dobiven model koristi nezavisne varijable dob, bmi, brojdjece i pušenje. Nazovimo ga model 2.

	Resid.Df	Resid.dev	Df	Deviance	p(>Chi)
Model2	1333	181.48			
Model1	1330	180.59	3	0.89567	0.08593

Tablica 7: Usporedba modela 1 i modela 2

Anova usporedba ovih dvaju modela dana je u Tablici 7. Rezultati pokazuju ne značajnu razliku (p-vrijednost=0.08593) zbog čega ćemo u nastavku promatrati model 2. Procijenjeni parametri modela 2 dani su u tablici 8.

	Parametar	St.greška	z vrijednost	p-vrijednost
(Intercept)	-0.3821365	0.0572815	-6.671	3.71e-11
dob	0.0034659	0.0007234	4.791	1.84e-06
bmi	0.0305112	0.0016648	18.327	$<$ 2e-16
brojedjece	-0.1721806	0.0083790	-20.549	$<$ 2e-16
pušenje1	0.4121140	0.0250060	16.481	$<$ 2e-16

Tablica 8: Procjena parametara modela 2

Sada konačni model možemo zapisati u sljedećem obliku:

$$\log\left(\frac{\mu}{1-\mu}\right) = -0.3821365 + 0.0034659 \cdot dob + 0.0305112 \cdot bmi$$
$$-0.1721806 \cdot brojdjece + 0.41211405 \cdot I_{\{pušenje=1\}}.$$

Dakle, vrijednost procijenjenog parametra uz varijablu dob iznosi približno 0.0035. To znači da ako se dob klijenta poveća za 1, logaritam šanse da klijent podnese zahtjev povećava se za 0.0035. Odnosno, jediničnim povećanjem dobi klijenta povećava se šansa da će klijent podnijeti zahtjev u odnosu na šansu da isti neće podnijeti za 1.0035. Vrijednost procijenjenog parametra uz varijablu bmi iznosi približno 0.0305. To znači da se jediničnim povećanjem indeksa tjelesne mase povećava šansa da će klijent podnijeti zahtjev u odnosu na šansu da ga neće podnijeti za 1.031. Vrijednost procijenjenog parametra uz varijablu broj djece iznosi približno -0.1722 što znači da se povećanjem broja djece za jedan, smanjuju šanse da će klijent podnijeti zahtjev u odnosu da neće za 0.8418. Baznu kategoriju za varijablu pušenje čine klijenti koji su nepušači. Prema tome, šansa da zahtjev podnesu klijenti koji su pušači povećava se za 1.9392 u odnosu na šansu da zahtjev podnesu klijenti koji su nepušači.

3 Bonus-malus sustav određivanja premije osiguranja

U ovome poglavlju upoznat ćemo se sa bonus-malus sustavom određivanja premije koji se obično koristi u osiguranju motornih vozila. Njime se osiguranike svrstava u homogene skupine ovisno o broju podnesenih zahtjeva tijekom proteklih godina te se na taj način kažnjavaju rizičniji i nagrađuju manje rizični vozači. Ovo poglavlje prati [1] i [6].

3.1 Osiguranje od automobilske odgovornosti

Motorna i priključna vozila smiju sudjelovati u prometu jedino ukoliko su registrirana i ako imaju važeću prometnu dozvolu. Registrirati se mogu samo ona vozila za koja se na tehničkom pregledu utvrdi da su ispravna. Za registrirano vozilo izdaju se registarske pločice i prometna dozvola. Prometnu dozvolu potrebno je produžiti svake godine. Ukoliko vlasnik vozila odluči da neće koristiti to vozilo nakon isteka valjanosti prometne dozvole, dužan je odjaviti to vozilo u zakonskom roku od 15 dana od dana isteka prometne dozvole. Kako bi stanica za tehnički pregled registrirala vozilo osim što ono mora biti ispravno, vlasnik vozila dužan je priložiti policu osiguranja od auto odgovornosti i podmiriti potrebne naknade. Osiguranje od auto odgovornosti propisano je zakonom i pokriva štetu koju u slučaju prometne nesreće vozač nanese drugim osobama ili njihovim stvarima. Njime se pokrivaju štete nastale na području Republike Hrvatske te na teritoriju zemalja članica Sustava zelene karte osiguranja. Ono vrijedi jednu osigurateljnu godinu, odnosno 365 dana od dana sklapanja ugovora. Osiguravajuće društvo je dužno ugovaratelju osiguranja uz policu osiguranja od automobilske odgovornosti uručiti i Europsko izvješće o nesreći. Za vrijeme upotrebe vozila u prometu vozač je dužan imati to izvješće i predati ga na zahtjev ovlaštene osobe u prometu. U slučaju prometne nesreće sudionici moraju ispuniti, potpisati i međusobno razmijeniti Europsko izvješće o nesreći.

Uz policu obveznog osiguranja od automobilske odgovornosti moguće je ugovoriti i kasko osiguranje koje pokriva štetu na vlastitom vozilu. Dakle, u slučaju prometne nezgode obvezno osiguranje od automobilske odgovornosti pokriva štetu nastalu na automobilu druge osobe dok kasko osiguranje pokriva štetu na našem automobilu (bili mi krivi ili ne). Izuzetak je jedino vožnja pod utjecajem alkohola ili opijata.

Cijena osiguranja razlikuje se od osiguranika do osiguranika, vozila do vozila te od osiguravatelja do osiguravatelja. Osiguranik je pri sklapanju ugovora dužan priložiti svoj osobni identifikacijski broj (OIB), datum rođenja, registarsko područje vozila te snagu motora vozila. Budući da postoji više osoba istog imena i prezimena, OIB služi kako bi se utvrdio točan korisnik vozila. Na temelju datuma rođenja utvrđuje se broj godina vozača. Vozači mlađi od 25 godina te vozači stariji od 70 godina smatraju se rizičnijima i njihovo je osiguranje skuplje u odnosu na osiguranje klijenata iz ostalih dobnih skupina. Registarsko područje osiguranika je bitno jer različita registarska područja imaju različite faktore rizika. Povećanjem faktora rizika raste premija osiguranja. U Republici Hrvatskoj gradovi s najvećom premijom su Grad Zagreb i Krapina. Vozilo koje ima više kilovata može razviti veću brzinu, a time i uzrokovati veću štetu. To je razlog zbog kojeg je cijena osiguranja veća za vozila veće snage.

Osim navedenoga, postoji još faktora koji utječu na cijenu osiguranja kao što su kategorija vozila, svrha vozila, vlasništvo nad vozilom, itd. Kako bi privukli nove klijente i zadržali postojeće, osiguravatelji svojim klijentima nude mnogobrojne pogodnosti. Različita osiguravajuća društva imaju različite pogodnosti, a ono što im je zajedničko je da pri izračunavanju premije osiguranja najveći utjecaj ima vozačka povijest osiguranika. Točnije, sva osiguravajuća društva pri izračunu premije osiguranja od automobilske odgovornosti koriste bonus-malus sustav određivanja premije osiguranja.

3.2 Bonus-malus premijski stupnjevi

Bonus-malus sustav određivanja premije svrstava osiguranike na temelju njihove vozačke povijesti u homogene skupine. Tako će se osiguraniku koji tijekom proteklih godina nije imao niti jedan osigurani slučaj dodijeliti bonus, odnosno on će steći pravo na niži iznos premije osiguranja od osnovnog (100%). S druge strane, osiguraniku koji je u prethodnom razdoblju imao jedan ili više takvih slučajeva dodijelit će se malus, odnosno on će morati platiti višu premiju osiguranja od osnovne (100%).

Homogene skupine predstavljaju premijske stupnjeve. Premijski stupnjevi nisu jednaki u svim državama stoga ćemo se najprije upoznati sa premijskim stupnjevima u Hrvatskoj, a zatim i u Nizozemskoj i Maleziji.

3.2.1 Bonus-malus premijski stupnjevi u Republici Hrvatskoj

PREMIJSKI STUPNJEVI																	
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.
% od premije osnovnog stupnja																	
50	55	60	65	70	75	80	85	90	100	115	130	150	170	190	210	230	250

Tablica 9: Premijski stupnjevi bonus-malus sustava osiguranja u Republici Hrvatskoj

Premijski stupnjevi u Tablici 9 odnose se na osiguranja od automobilske odgovornosti vlasnika svih vrsta vozila (osobna, teretna, priključna vozila, autobusi, traktori, motocikli, mopedi i radni strojevi) u Republici Hrvatskoj. Osiguraniku koji ima sklopljen ugovor o osiguranju od automobilske odgovornosti visina premije se utvrđuje ovisno o prijavljenim štetama u proteklom razdoblju. Smatra se da je šteta prijavljena ukoliko je njome utvrđena obveza osiguravatelja. Šteta nije prijavljena ako je do dana sklapanja ugovora za sljedeće osigurateljno razdoblje u potpunosti regresirana.

Kada osiguranik sklopi osiguranje po prvi puta, plaća premiju desetog (osnovnog) stupnja. Ako je vozilo bilo osigurano najmanje pola godine i ako u tom razdoblju nije bilo prijavljenih šteta, osiguraniku se za sljedeću godinu odobrava jedan premijski stupanj niže, najviše do prvog stupnja. Za svaku prijavljenu štetu u razdoblju promatranja, osiguranik se pomiče za 3 premijska stupnja više u sljedećoj osiguravateljnoj godini, a najviše do 18. stupnja. Razdoblje promatranja je prethodna kalendarska godina, a primjenjuje se na osiguranja koja počinju 1.2. tekuće kalendarske godine do 31.1. sljedeće kalendarske godine.

3.2.2 Bonus-malus premijski stupnjevi u Nizozemskoj

PREMIJSKI STUPNJEVI													
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14
% od premije osnovnog stupnja													
120	100	90	80	70	60	55	50	45	40	37.5	35	32.5	30

Tablica 10: Premijski stupnjevi bonus-malus sustava osiguranja u Nizozemskoj

U Nizozemskoj osiguranik koji ugovara osiguranje od automobilske odgovornosti po prvi puta, plaća osnovnu premiju. Osnovna je premija u ovom slučaju premija stupnja 2. Ukoliko osiguranik ne prijavi štetu tokom jedne kalendarske godine, u idućoj godini plaća premiju prvog sljedećeg višeg stupnja. Premija se prve četiri godine umanjuje za 10% u odnosu na prethodnu godinu, zatim naredne četiri za 5% u odnosu na prethodnu godinu te preostale četiri godine umanjuje se za 2.5% u odnosu na prethodnu godinu. Najviši stupanj koji osiguranik može postići ukoliko iz godine u godinu ne prijavljuje štetu je stupanj 14 unutar kojeg premija iznosi 30% premije osnovnog stupnja.

Prijelaz među stupnjevima ukoliko osiguranik tijekom prve godine prijavi jednu štetu prikazan je u Tablici 11.

2 →	1 →	1 →	1 →	1 →	2 →	3 →	4 →	5 →	6 →	7 →	7 →	8 →	8 →	9

Tablica 11: Premijski stupnjevi ukoliko osiguranik prijavi štetu tijekom prve godine

Prijelaz među stupnjevima ukoliko osiguranik tijekom prve godine prijavi dvije štete prikazan je u Tablici 12.

2 →	1 →	1 →	1 →	1 →	1 →	1 →	1 →	1 →	2 →	3 →	3 →	4 →	4 →	5

Tablica 12: Premijski stupnjevi ukoliko osiguranik prijavi dvije štete tijekom prve godine

Ukoliko osiguranik tijekom prve osiguravateljne godine prijavi tri ili više šteta, zavijek gubi mogućnost prijelaza iz stupnja u stupanj i uvijek plaća premiju stupnja 1.

3.2.3 Bonus-malus premijski stupnjevi u Maleziji

PREMIJSKI STUPNJEVI					
0.	1.	2.	3.	4.	5.
% od premije osnovnog stupnja					
100	75	70	61.23	55	45

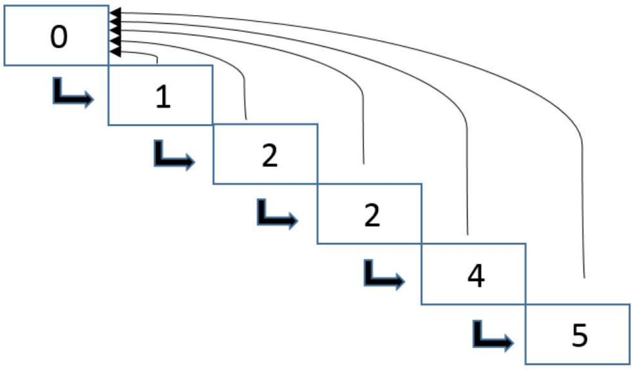
Tablica 13: Premijski stupnjevi bonus-malus sustava osiguranja u Maleziji

Malezijski prijelaz između stupnjeva vrlo je jednostavan. Ukoliko osiguranik ugovara osiguranje od automobilske odgovornosti po prvi puta, plaća osnovnu premiju. Dokle god ne prijavljuje štetu, osiguranik u sljedećoj osiguravateljnoj godini prelazi u viši premijski stupanj. Ukoliko prijavi štetu, kažnjava ga se na način da se u idućoj osiguravateljnoj godini vraća u nulti stupanj, odnosno plaća osnovnu premiju.

3.3 Dupliciranje, zaštita i prijenos bonusa

Neke osiguravajuće kuće omogućuju svojim korisnicima da zarađeni bonus preslikaju na prvu policu užeg člana obitelji, a da se pritom njihov bonus ne smanji. Primjerice, ako otac ima bonus od 40% može ga preslikati na prvu policu osiguranja svoga sina. Tada će i otac i sin imati bonus od 40%. U osiguravateljskoj terminologiji to se naziva dupliciranjem bonusa.

Osiguravajuće kuće svojim klijentima nude i osiguranje zaštite bonusa. To je osiguranje koje daje pravo osiguraniku na zadržavanje stečenog bonusa ukoliko tokom godine prijavi



Slika 1: Dijagram prijelaza između premijskih stupnjeva u Maleziji

jednu štetu. Dakle, ukoliko osiguranik prijavi štetu, njegov bonus bi iduće godine trebao biti umanjen za 3 premijska stupnja. Međutim, ako osiguranik ima ugovoreno osiguranje zaštite bonusa njegov bonus će u idućoj godini ostati nepromijenjen.

Bonus-malus sustav vezan je uz vozačku povijest, a ne za konkretno vozilo koje se osigurava. To znači da ukoliko kupimo novi automobil i za njega ugovaramo policu osiguranja od automobilske odgovornosti, bonus sa starog vozila prenosi se na novo. U osiguravateljskoj terminologiji to se naziva prijenosom bonusa.

3.4 Generalizirani linearni modeli i bonus-malus sustav određivanja premije auto osiguranja

Generalizirani linearni modeli se u osiguranju koriste kao a priori klasifikacija rizika. Procjena rizičnosti osobe koja po prvi puta želi osigurati svoje vozilo određuje se na temelju karakteristika poput spola, dobi, vrijednosti automobila, težine automobila, itd. Karakteristike se razlikuju ovisno o državi i osiguravajućem društvu. Klijenti se na temelju njih svrstavaju u homogene grupe te im se dodjeljuje faktor rizika i određena premija. Jasno je da početni faktori nisu najbolji pokazatelji je li neka osoba manje ili više rizična. U analizu je potrebno uključiti faktore poput brzine refleksa, poznavanja prometnih pravila, agresivnosti za volanom i druge koje je teško izračunati. Pretpostavlja se da se ti faktori mogu kvalitetno uvesti u analizu kroz broj prijavljenih šteta tijekom godine. Zato osiguravajuća društva nakon prve godine trajanja ugovora pristupaju individualizaciji rizika putem bonus-malus sustava. Tada iznos premije koju plaća ugovaratelj ne ovisi samo o faktorima rizika već i o povijesti šteta. U nastavku ovoga poglavlja, prikazat ćemo primjenu generaliziranih linearnih modela i bonus-malus sustava u auto osiguranju.

Cilj nam je modelirati broj zahtjeva, odnosno broj prijavljenih šteta u promatranom razdoblju na temelju spola i zanimanja vozača te veličini automobila i prostoru na kojem se automobil koristi. U tu svrhu koristimo simuliranu bazu podataka. Četiri nezavisne varijable koje ćemo koristiti pri izgradnji i njihovi opisi prikazani su u Tablici 14.
Osiguranike jednakih obilježja svrstavamo u iste grupe. Postoje $2 \cdot 3 \cdot 3 \cdot 3 = 54$ moguće kombinacije zbog čega razlikujemo 54 grupe. Prvih 27 grupa čine klijenti ženskog spola, a preostalih 27 grupa čine klijenti muškog spola. Klijenti u prvih 9 grupa automobil voze u obalnim mjestima, u drugih 9 u mjestima koja nisu ni obalna ni veliki gradovi te u trećih 9 grupa klijenti automobil voze isključivo u velikim gradovima. Na isti način nastavljamo pridruživati regije klijentima u preostalim grupama. Poslove kojima se klijenti bave rasporedili

Varijabla	Opis varijable
sex	spol (1=žena, 2=muškarac)
region	regija u kojoj se automobil koristi (1=obala, 2=drugo, 3=veliki grad)
type of car	vrsta automobila obzirom na veličinu (1=mali, 2=srednji, 3=veliki)
job class	vrsta posla kojim se vlasnik automobila bavi (1=državni službenik/aktuar/..., 2=između, 3=profesionalni vozač)

Tablica 14: Opis nezavisnih varijabli

smo u 3 kategorije te smo klijentima u prvoj grupi pridružili posao iz kategorije 1, u drugoj grupi iz kategorije 2 i u trećoj grupi iz kategorije 3. Zatim smo klijentima u četvrtoj grupi pridružili posao iz kategorije 1, u petoj grupi iz kategorije 2 i šestoj grupi iz kategorije 3. Na isti način smo pridružili vrste poslova i preostalim klijentima u sljedećim grupama. Baza podataka još sadrži informaciju o veličini automobila klijenta. Razlikujemo male, srednje i velike automobile. Klijenti iz prve tri grupe voze male automobile, iz druge tri voze automobile srednje veličine i iz sjedeće tri voze velike automobile. Na isti način nastavili smo popunjavati podatke o veličini automobila za klijente u ostalim grupama. Uvid u konačnu bazu podataka daje nam sljedeća tablica:

N_i	sex	region	type	job
1	1	1	1	1
8	1	1	1	2
10	1	1	1	3
8	1	1	2	1
5	1	1	2	2
11	1	1	2	3
14	1	1	3	1
12	1	1	3	2
11	1	1	3	3
10	1	2	1	1
$\vdots$				
16	2	2	3	3
16	2	3	1	1
13	2	3	1	2
14	2	3	1	3
8	2	3	2	1
19	2	3	2	2
20	2	3	2	3
9	2	3	3	1
23	2	3	3	2
27	2	3	3	3

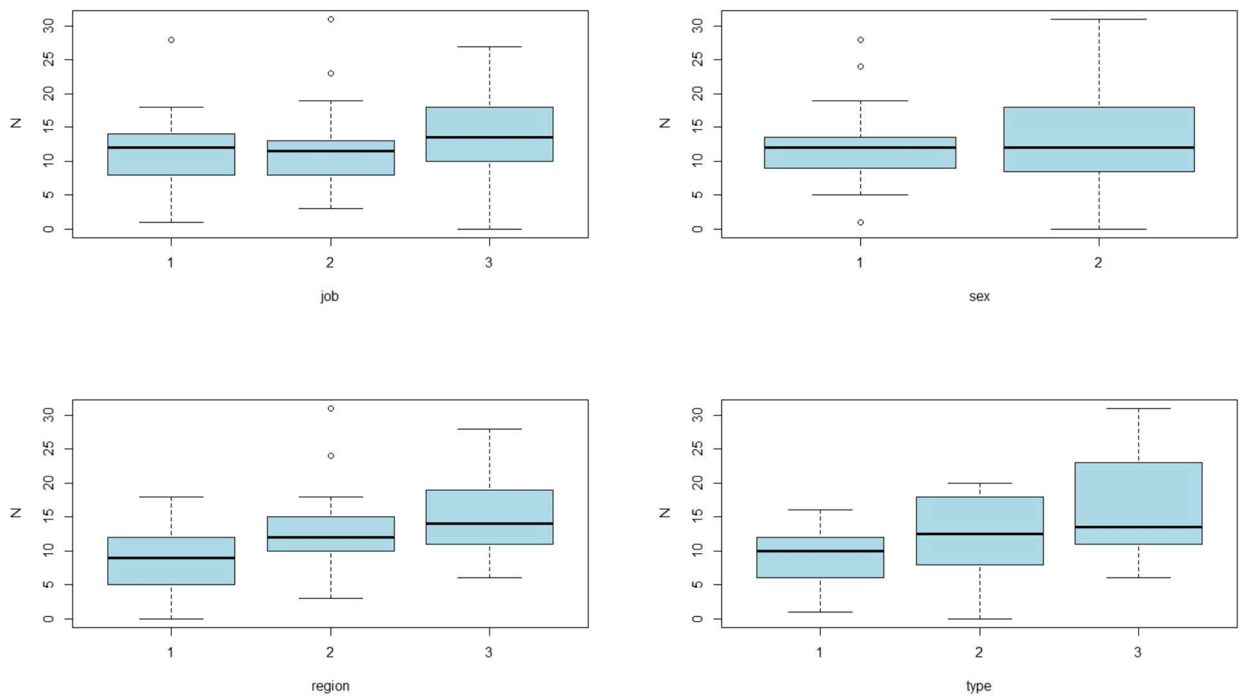
Ukoliko sa N_i, $i \in \{1, \ldots, 54\}$ označimo zavisnu varijablu kojom modeliramo broj zahtjeva i-te grupe, logično je za pretpostaviti kako ona ima Poissonovu distribuciju. Stoga ćemo pri modeliranju koristiti generalizirani linearni model s logaritamskom funkcijom veze. Pripadni model možemo zapisati kao:

$$log\mu = \beta_0 + \beta_1 \cdot I_{\{sex=2\}} + \sum_{i=2}^{3} \beta_i I_{\{region=i\}} + \sum_{i=2}^{3} \beta_{i+2} I_{\{type=i\}} + \sum_{i=2}^{3} \beta_{i+4} I_{\{job=i\}},$$

odnosno kao:

$$\mu = \exp(\beta_0 + \beta_1 \cdot I_{\{sex=2\}} + \sum_{i=2}^{3} \beta_i I_{\{region=i\}} + \sum_{i=2}^{3} \beta_{i+2} I_{\{type=i\}} + \sum_{i=2}^{3} \beta_{i+4} I_{\{job=i\}}).$$

Pogledajmo najprije odnose nezavisnih varijabli sa zavisnom varijablom prikazane u Tablici 15.



Tablica 15: Boxplot nezavisnih varijabli

Na grafičkom prikazu vidimo kako grupe u kojima se voze veliki automobili u prosjeku imaju veći broj prijavljenih šteta u odnosu na grupe u kojima se voze mali i srednji automobili. Također je uočljivo da postoji razlika u broju prijavljenih šteta ovisno o regiji u kojoj se automobil vozi. Tako je broj prijavljenih šteta u grupama u kojima se automobili voze u velikim gradovima u prosjeku veći od broja prijavljenih šteta u svim drugim grupama. Ženske i muške grupe u prosjeku imaju približno jednak broj prijavljenih šteta zbog čega pretpostavljamo kako varijabla sex neće imati veliku ulogu pri izgradnji modela. Također, grupe s klijentima koji su profesionalni vozači imaju u prosjeku veći broj prijavljenih šteta od svih ostalih grupa što je i logično za pretpostaviti budući da profesionalni vozači provode najviše vremena na cestama.

Za odabir modela koristit ćemo se selekcijskim procedurama. Selekcijska procedura unaprijed sugerira nam da je najbolji model puni model, odnosno model sa svim navedenih nezavisnim varijablama. Nazovimo ga model 1. S druge strane, selekcijska procedura unazad nam sugerira model koji uključuje samo varijable region i type. Nazovimo taj model modelom 2. Ta dva modela ulaze u daljnju analizu.

	Resid.Df	Resid.dev	Df	Deviance	p(>Chi)
Model2	49	119.56			
Model1	46	114.39	3	5.1997	0.1577

Tablica 16: Usporedba modela 1 i modela 2

Anova usporedba modela 1 i modela 2 dana je u Tablici (16). Rezultati pokazuju ne značajnu razliku (p-vrijednost=0.1577) zbog čega se odlučujemo promatrati model 2, tj. model dobiven selekcijskom procedurom unazad. Procijenjeni parametri tog modela dani su u Tablici 17.

	Parametar	St.greška	z vrijednost	p-vrijednost
(Intercept)	1.88670	0.10371	18.192	< 2e-16
region2	0.40962	0.10197	4.017	5.90e-05
region3	0.53795	0.09949	5.407	6.41e-08
type2	0.25728	0.10304	2.497	0.0125
type3	0.55876	0.09702	5.759	8.44e-09

Tablica 17: Procjena parametara modela 2

Model 2 sada možemo zapisati kao:

$$\mu = \exp(1.88670 + 0.40962 \cdot I_{\{region=2\}} + 0.53795 I_{\{region=3\}} + 0.25728 \cdot I_{\{type=2\}} + 0.55876 \cdot I_{\{type=3\}}).$$

Uzeti ćemo eksponencijalnu vrijednost parametra. Pozitivna vrijednost implicira veći broj prijavljenih šteta, dok vrijednosti bliže nuli predstavljaju neutralni efekat budući je $e^0 = 1$. Ako je eksponencijalna vrijednost veća od 1 to znači da se parametar Poissonove distribucije povećava jer ga množimo s nečim većim od 1. U suprotnom se smanjuje.

Na temelju ovoga modela, zaključujemo kako se povećanjem veličine automobila povećava i broj prijavljenih šteta. Osim veličine automobila na broj prijavljenih šteta u našem primjeru utječe i područje na kojem se automobil vozi. Tako se povećanje mjesta u kojem se automobil vozi reflektira povećanjem broja prijavljenih šteta. Zaključujemo kako bi osiguravajuća kuća na osnovu ovoga modela naplatila veću premiju velikim automobilima u odnosu na premiju srednjim i malim automobilima. Isto tako, premija bi bila veća za automobile koji se voze u velikim gradovima u odnosu na ostala mjesta i gradove.

Dakle, kada novi klijent ulazi u osiguranje najprije se pomoću generaliziranih linearnih modela procijeni broj zahtjeva koje bi on mogao podnijeti. Veći broj zahtjeva predstavlja rizičnijeg klijenta dok manji broj zahtjeva predstavlja manje rizičnog klijenta. Premija se dobije sumiranjem iznosa dobivenog na temelju procjene broja zahtjeva uvećanog za određeni iznos kako bi osiguravajuće društvo osiguralo pozitivno poslovanje sa naknadama poput eko testa, naknade za ceste, naknade za okoliš i poreza na cestovna motorna vozila. Od iznimne je važnosti što točnije odrediti premiju osiguranja. U slučaju da smo manje rizičnim klijentima dodijelili premiju puno veću nego ih pripada, oni će prijeći u povoljnije osiguravajuće kuće. S druge strane ako rizičnim klijentima dodijelimo niske premije privući ćemo veliki broj rizičnih klijenata koji na kraju mogu dovesti do negativnog poslovanja.

Nakon prve godine trajanja ugovora osiguravajuće društvo na osnovu prijavljenih šteta može zaključiti je li dobro procijenilo klijenta te ukoliko nije ima mogućnost korigirati njegovu premiju kroz bonus-malus premijske stupnjeve. Bonus-malus premijski stupnjevi objašnjeni su u potpoglavlju 3.2.

3.5 Glad za bonusom

Glad za bonusom (engl. hunger for bonus) je strategija osiguranika koji u cilju zadržavanja postojeće premije odlučuje samofinancirati nastalu štetu umjesto da ju prijavi osiguranju koje bi ju financiralo. Ova strategija je česta kod osiguranika u višim premijskim stupnjevima. Njima je u slučaju malih šteta isplativije pokriti štetu vlastitim novcem nego ju prijaviti osiguranju i izgubiti trenutni bonus. Glad za bonusom je razlog zbog kojega ne možemo vjerovati da su osiguranici u višim premijskim stupnjevima manje rizični. Dobro osmišljen bonus-malus sustav mora uzeti u obzir glad za bonusom.

3.6 Nedostatak bonus-malus sustava određivanja premije osiguranja

A priori klasifikacijom rizika se klijentima pri ugovaranju police osiguranja na osnovu temeljnih karakteristika dodjeljuje određena premija. Ideja bonus-malus sustava je ovisno o broju podnesenih zahtjeva tokom godine klijentu što bolje prilagoditi odgovarajuću a priori premiju kroz premijske stupnjeve i njihove odgovarajuće postotke od premije osnovnog stupnja. Ta ideja je dobra jedino ukoliko je premija koju klijent plaća po vrijednosti jako blizu premiji koju bi on uistinu trebao plaćati. S druge strane, osiguranicima koji su a priori klasifikacijom rizika svrstani u rizičniju kategoriju, dodijeljena je i viša premija. Ukoliko je nekom klijentu dodijeljena premija viša od one koja mu stvarno pripada, a postotci od premijskog osnovnog stupnja su jednaki za sve osiguranike, on će zauvijek plaćati premiju višu od one koja ga pripada bez obzira što će se ona tokom godina smanjivati ukoliko on ne prijavi zahtjev. Dakle, klijenti kojima je a priori dodijeljena viša premija će uvijek plaćati veću premiju od klijenata kojima je dodijeljena niža a priori premija. Problem nastaje zbog toga što su svim klijentima dodijeljeni jednaki postotci od premije osnovnog stupnja bez obzira na a priori premiju. Prijedlog rješavanju ovakvog problema je uvođenje većih postotaka od premijskog osnovnog stupnja za klijente sa većom a priori premijom ali to nije praksa.

4 Bonus-malus sustav i Markovljevi lanci

U ovome ćemo se poglavlju prisjetiti osnovnih pojmova vezanih uz homogene Markovljeve lance u diskretnom vremenu i pokazati kako je bonus-malus sustav homogen Markovljev lanac u diskretnom vremenu. Teoremi i definicije u ovom poglavlju prate [1], [4] i [8].

4.1 Markovljevi lanci u diskretnom vremenu

Definicija 3. *Neka je S diskretan skup stanja. Slučajni proces $X = (X_n, n \in \mathbb{N}_0)$ na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ s vrijednostima u skupu S je Markovljev lanac ako jednakost*

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \tag{6}$$

vrijedi za svaki $n \in \mathbb{N}_0$ i za sve $i_0, \dots, i_{n-1}, i, j \in S$ za koje su uvjetne vjerojatnosti u (6) dobro definirane.

Svojstvo (6) zove se Markovljevo svojstvo te ga možemo interpretirati na način da je vjerojatnosno ponašanje Markovljevog lanca u neposrednoj budućnosti uvjetno na sadašnjost i prošlost jednako ponašanju tog lanca u neposrednoj budućnosti uvjetno samo na sadašnjost. Vjerojatnost prijelaza iz jednog u drugo stanje Markovljevog lanca zadajemo funkcijom prijelaznih vjerojatnosti. Za stanja $i, j \in S$ i trenutke $s, t \in \mathbb{N}_0$ t.d. je $s < t$ funkciju prijelaznih vjerojatnosti definiramo pravilom

$$p(i, s; t, j) = P(X_t = j, X_s = i).$$

Za konkretna stanja $i, j \in S$ i trenutke $s, t \in \mathbb{N}_0$, $s < t$ vrijednost funkcije $p(i, s; t, j)$ je vjerojatnost da se Markovljev lanac koji se u trenutku s nalazio u stanju i, u trenutku t nađe u stanju j. Za stanja $i, j \in S$ i $n \in \mathbb{N}_0$ definiramo funkciju

$$p(i, n; n+1, j) = P(X_{n+1} = 1 | X_n = i)$$

koju nazivamo funkcijom 1-koračnih prijelaznih vjerojatnosti. Ukoliko funkcija 1-koračnih prijelaznih vjerojatnosti Markovljevog lanca X ovisi samo o n, tj. ako za proizvoljne $n, m \in \mathbb{N}_0$, $n \neq m$ vrijedi da je

$$p(i, n; n+1, j) = p(i, m; m+1, j),$$

kažemo da se radi o (vremenski) homogenom Markovljevom lancu. U tom slučaju funkcija 1-koračnih prijelaznih vjerojatnosti ovisi samo o stanju iz kojega lanac polazi i o stanju u kojega on dolazi te uvodimo oznaku

$$p(i, n; n+1, j) = p_{ij}.$$

Za homogeni Markovljev lanac 1-koračne prijelazne vjerojatnosti p_{ij}, $i, j \in S$ možemo zapisati u formi matrice koja se zove matrica 1-koračnih prijelaznih vjerojatnosti i označava se sa Π. Za Markovljev lanac sa konačnim skupom stanja $S = \{1, 2, \dots n\}, n \in \mathbb{N}$ matrica Π je sljedećeg oblika:

$$\Pi = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2n} & \cdots \\ p_{31} & p_{32} & \cdots & p_{3n} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} & \cdots \end{bmatrix}.$$

Važna svojstva matrice Π su:

i) $p_{ij} \geq 0$, za sve $i, j \in S$

ii) za svaki $i \in S$ je

$$\sum_{j \in S} p_{ij} = 1$$

tj. suma 1-koračnih prijelaznih vjerojatnosti u istom retku matrice Π je 1.

Matrica koja zadovoljava prethodno navedena svojstva zove se stohastička matrica. Vektor vjerojatnosti $\lambda = (\lambda_i, i \in S)$ pri čemu je $\lambda_i = P(X_0 = i)$ a X_0 slučajna varijabla kojom modeliramo početno stanje Markovljevog lanca $(X_n, n \in \mathbb{N}_0)$, naziva se početnom distribucijom tog lanca. Markovljev lanac je u potpunosti određen početnom distribucijom λ i matricom 1-koračnih prijelaznih vjerojatnosti Π. Za takav Markovljev lanac kažemo da je (λ, Π)-Markovljev lanac.

4.2 Bonus-malus sustav određivanja premije kao Markovljev lanac u diskretnom vremenu

Budući je za određivanje premijskog stupnja osiguranika u narednoj godini potrebno znati trenutni premijski stupanj osiguranika te broj prijavljenih šteta tijekom godine, Bonus-malus sustav može se predstaviti kao Markovljev lanac u diskretnom vremenu. U nastavku ovoga rada promatrat ćemo bonus-malus sustav s aspekta Markovljevih lanaca te ćemo sve ilustrirati primjerima.

4.2.1 Pravilo prijelaza iz jednog premijskog stupnja u drugi

Vjerojatnost prijelaza iz jednog premijskog stupnja u drugi ovisi o broju prijavljenih zahtjeva tokom godine. Za k prijavljenih zahtjeva definiramo funkciju prijelaza na sljedeći način

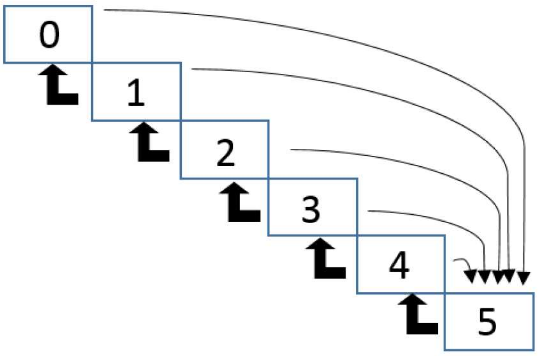
$$t_{ij}(k) = \begin{cases} 1, & \text{ukoliko osiguranik prijeđe iz i-tog u j-ti premijski stupanj} \\ 0, & \text{inače.} \end{cases}$$

Rasporedimo li $t_{ij}(k)$-ove u matricu $T(k)$ dobivamo matricu sljedećeg oblika:

$$T(k) = \begin{bmatrix} t_{00}(k) & t_{01}(k) & \dots & t_{0s}(k) \\ t_{10}(k) & t_{11}(k) & \dots & t_{1s}(k) \\ \vdots & \vdots & \ddots & \vdots \\ t_{s0}(k) & t_{s1}(k) & \dots & t_{ss}(k) \end{bmatrix}.$$

Suma elemenata svakog retka te matrice iznosi 1.

Primjer 4. *U ovome ćemo poglavlju u cilju ilustracije bonus-malus sustava kao Markovljevog lanca koristiti -1/Top Scale premijski sustav. U takvom sustavu je nulti stupanj najpoželjniji, s posljednji peti najmanje poželjan. Pravilo prijelaza iz jednog u drugi premijski stupanj prikazano je na Slici 2.*
Klijent koji tijekom godine ne prijavi niti jedan zahtjev u narednoj godini prelazi u prvi sljedeći niži premijski stupanj. Ukoliko prijavi jedan ili više zahtjeva, direktno prelazi u posljednji peti premijski stupanj. Matrica $T(k)$ -1/Top Scale premijskog sustava ovisno o broju zahtjeva k poprima sljedeći oblik:



Slika 2: Dijagram prijelaza između premijskih stupnjeva -1/Top Scale premijskog sustava

$$T(0) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad T(1) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad T(k) = T(1) \quad \forall k \geq 2.$$

4.2.2 Vjerojatnost prijelaza iz jednog premijskog stupnja u drugi i matrica prijelaznih vjerojatnosti

Pretpostavimo da su $N_1, N_2, \ldots$ nezavisne slučajne varijable s Poissonovom distribucijom s parametrom σ kojim modeliramo broj zahtjeva tokom godina. Također, pretpostavimo kako su $L_1(\sigma), L_2(\sigma), \ldots$ slučajne varijable kojima modeliramo premijske stupnjeve klijenata te neka njihova distribucija ovisi o σ. Tada se vjerojatnost prijelaza iz stupnja l_1 u stupanj l_2 klijenta s godišnjim očekivanim brojem zahtjeva σ definira kao

$$p_{l_1 l_2}(\sigma) = P[L_{k+1}(\sigma) = l_2 | L_k(\sigma) = l_1], \quad l_1, l_2 \in \{0, 1, \ldots, s\}$$

i vrijedi:

$$p_{l_1 l_2}(\sigma) \geq 0 \quad \text{za sve} \quad l_1 \quad i \quad l_2, \quad \sum_{l_2=0}^{s} p_{l_1 l_2}(\sigma) = 1.$$

Iskoristimo li definiciju funkcije t_{ij}, činjenicu da su N_{k+1} i $L_k(\sigma)$ nezavisne i da su $N_1, N_2, \ldots$ slučajne varijable s Poissonovom distribucijom, vrijedi:

$$\begin{aligned} p_{l_1 l_2}(\sigma) &= \sum_{n=0}^{\infty} P(L_{k+1}(\sigma) = l_2 | N_{k+1} = n, L_k(\sigma) = l_1) P(N_{k+1} = n | L_k(\sigma) = l_1) \\ &= \sum_{n=0}^{\infty} P(L_{k+1}(\sigma) = l_2 | N_{k+1} = n, L_k(\sigma) = l_1) P(N_{k+1} = n) \\ &= \sum_{n=0}^{\infty} \frac{\sigma^n}{n!} \exp(-\sigma) t_{l_1 l_2}(n). \end{aligned}$$

Prijelazne vjerojatnosti omogućuju aktuarima da izračunaju vjerojatnost prijelaza klijenta iz jednog u bilo koji drugi premijski stupanj. Ako s l_i označimo premijski stupanj u trenutku $i = 1, \ldots, n$, vjerojatnost prijelaza klijenta s prosječnim godišnjim brojem zahtjeva σ iz stupnja l_0 u stupanj l_n preko stupnjeva $l_1, \ldots, l_{n-1}$ računa se kao

$$P[L_1(\sigma) = l_1, \ldots, L_n(\sigma) = l_n | L_0(\sigma) = l_0] = p_{l_0 l_1}(\sigma) \cdots p_{l_{n-1} l_n}(\sigma).$$

Točnije, kako bi izračunali vjerojatnost prijelaza u bilo koji drugi premijski stupanj potrebno je poznavati samo trenutni premijski stupanj. Odnosno vrijedi

$$P[L_n(\sigma) = l_n | L_{n-1}(\sigma) = l_{n-1}, \ldots, L_0(\sigma) = l_0] = p_{l_{n-1} l_n}(\sigma),$$

za $P[L_{n-1}(\sigma) = l_{n-1}, \ldots, L_0(\sigma) = l_0] > 0$.

Primjer 5. *Matrica 1-koračnih prijelaznih vjerojatnosti -1/Top Scale premijskog sustava ima sljedeći oblik:*

$$P(\sigma) = \begin{bmatrix} \exp(-\sigma) & 0 & 0 & 0 & 0 & 1 - \exp(-\sigma) \\ \exp(-\sigma) & 0 & 0 & 0 & 0 & 1 - \exp(-\sigma) \\ 0 & \exp(-\sigma) & 0 & 0 & 0 & 1 - \exp(-\sigma) \\ 0 & 0 & \exp(-\sigma) & 0 & 0 & 1 - \exp(-\sigma) \\ 0 & 0 & 0 & \exp(-\sigma) & 0 & 1 - \exp(-\sigma) \\ 0 & 0 & 0 & 0 & \exp(-\sigma) & 1 - \exp(-\sigma) \end{bmatrix}.$$

4.2.3 Višekoračne prijelazne vjerojatnosti

Kako i sam naziv kaže, višekoračne prijelazne vjerojatnosti su vjerojatnosti prijelaza iz stanja i u stanje j ali u više koraka. Za n koraka vrijedi:

$$p_{ij}^{(n)}(\sigma) = P[L_{k+n}(\sigma) = j | L_k(\sigma) = i] \quad \text{za svaki k.}$$

$p_{ij}^{(n)}$ je vjerojatnost prijelaza iz stanja i u stanje j u n koraka. Kako osiguranik u jednoj godini samo jednom mijenja svoj premijski stupanj, odnosno u jednoj godini napravi jedan korak u Markovljevom lancu, broj koraka ekvivalentan je broju godina te $p_{ij}^{(n)}$ predstavlja vjerojatnost prijelaza iz stanja i u stanje j za n godina. Tada je matrica višekoračnih prijelaznih vjerojatnosti Markovljevog lanca dana s:

$$P^{(n)}(\sigma) = \begin{bmatrix} p_{00}^{(n)}(\sigma) & p_{01}^{(n)}(\sigma) & \ldots & p_{0s}^{(n)}(\sigma) \\ p_{10}^{(n)}(\sigma) & p_{11}^{(n)}(\sigma) & \ldots & p_{1s}^{(n)}(\sigma) \\ \vdots & \vdots & \vdots & \vdots \\ p_{s0}^{(n)}(\sigma) & p_{s1}^{(n)}(\sigma) & \ldots & p_{ss}^{(n)}(\sigma) \end{bmatrix}.$$

Matrica višekoračnih prijelaznih vjerojatnosti je također stohastička matrica i dobiva se potenciranjem matrice 1-koračnih prijelatnih vjerojatnosti. Dakle, u mnogim problemima ključno je znati odrediti n-tu potenciju matrice 1-koračnih prijelaznih vjerojatnosti $P(\sigma)$. Tom problemu možemo pristupiti na nekoliko načina. Ponekad $P^{(n)}(\sigma)$ možemo odrediti induktivno izračunavajući prvih nekoliko potencija matrice $P(\sigma)$ iz kojih tada zaključujemo kako izgledaju elementi matrice $P^{(n)}(\sigma)$. Također, matricu $P^{(n)}(\sigma)$ možemo odrediti pomoću dijagonalizacije matrice $P(\sigma)$ ali se time nećemo baviti u ovome radu.

4.2.4 Dugoročno ponašanje Markovljevog lanca i stacionarna distribucija

Prirodno je zapitati se kako će izgledati bonus-malus sustav s odmakom u budućnosti. Intuitivno pretpostavljamo kako će se bonus-malus sustav nakon određenog vremena stabilizirati, odnosno da će se svaki klijent nakon određenog vremena naći u nekom premijskom stupnju u kojemu će ostati do daljnjega. Kako bismo opravdali prethodno navedeno, pogledajmo sljedeći primjer.

Primjer 6. *Promotrimo matricu prijelaznih vjerojatnosti -1/Top Scale premijskog sustava kroz vrijeme za* $\sigma = 0.1$.

$$P(0.1) = \begin{bmatrix} 0.904837 & 0 & 0 & 0 & 0 & 0.095163 \\ 0.904837 & 0 & 0 & 0 & 0 & 0.095163 \\ 0 & 0.904837 & 0 & 0 & 0 & 0.095163 \\ 0 & 0 & 0.904837 & 0 & 0 & 0.095163 \\ 0 & 0 & 0 & 0.904837 & 0 & 0.095163 \\ 0 & 0 & 0 & 0 & 0.904837 & 0.095163 \end{bmatrix},$$

$$P^{(2)}(0.1) = \begin{bmatrix} 0.818731 & 0 & 0 & 0 & 0.086107 & 0.095163 \\ 0.818731 & 0 & 0 & 0 & 0.086107 & 0.095163 \\ 0.818731 & 0 & 0 & 0 & 0.086107 & 0.095163 \\ 0 & 0.818731 & 0 & 0 & 0.086107 & 0.095163 \\ 0 & 0 & 0.818731 & 0 & 0.086107 & 0.095163 \\ 0 & 0 & 0 & 0.818731 & 0.086107 & 0.095163 \end{bmatrix},$$

$$P^{(3)}(0.1) = \begin{bmatrix} 0.740818 & 0 & 0 & 0.077913 & 0.086107 & 0.095163 \\ 0.740818 & 0 & 0 & 0.077913 & 0.086107 & 0.095163 \\ 0.740818 & 0 & 0 & 0.077913 & 0.086107 & 0.095163 \\ 0.740818 & 0 & 0 & 0.077913 & 0.086107 & 0.095163 \\ 0 & 0.740818 & 0 & 0.077913 & 0.086107 & 0.095163 \\ 0 & 0 & 0.740818 & 0.077913 & 0.086107 & 0.095163 \end{bmatrix},$$

$$P^{(4)}(0.1) = \begin{bmatrix} 0.67032 & 0.00000 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.67032 & 0.00000 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.67032 & 0.00000 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.67032 & 0.00000 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.67032 & 0.00000 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.00000 & 0.67032 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \end{bmatrix},$$

$$P^{(5)}(0.1) = \begin{bmatrix} 0.606531 & 0.063789 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.606531 & 0.063789 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.606531 & 0.063789 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.606531 & 0.063789 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.606531 & 0.063789 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \\ 0.606531 & 0.063789 & 0.070498 & 0.077913 & 0.086107 & 0.095163 \end{bmatrix},$$

$$P^{(k)}(0.1) = P^{(5)}(0.1) \quad \forall k \geq 6.$$

Primijetimo kako su vjerojatnosti nakon n godina jednake po stupcima odnosno da svaki osiguranik bez obzira u kojem se premijskom stupnju tada nalazio ima jednaku vjerojatnost prelaska u nulti, prvi, drugi i ostale stupnjeve. Vektor redak te matrice

$$\pi = (0.606531, 0.063789, 0.070498, 0.077913, 0.086107, 0.095163)$$

naziva se stacionarnom distribucijom -1/Top Scale premijskog sustava. Dakle, 60.6531% osiguranika će zauzeti stupanj 0, 6.3789% stupanj 1, itd.

Dakle, n-ta potencija $P^{(n)}(\sigma)$ matrice 1-koračnih prijelaznih vjerojatnosti $P(\sigma)$ konvergira ka matrici $\Pi(\sigma)$ sa svim jednakim stupcima $\pi^T(\sigma)$:

$$\lim_{n\to\infty} P^{(n)}(\sigma) = \Pi(\sigma) = \begin{bmatrix} \pi^T(\sigma) \\ \pi^T(\sigma) \\ \vdots \\ \pi^T(\sigma) \end{bmatrix}.$$

Vektor redak $\pi(\sigma)$ matrice $\Pi(\sigma)$ naziva se stacionarnom distribucijom, a cijela matrica $\Pi(\sigma)$ stacionarnom matricom. Stacionarna distribucija, odnosno uvid u raspodjelu klijenata po premijskim stupnjevima u budućnosti omogućuje osiguravajućem društvu upravljanje vlastitim prihodima. Navedeno ćemo opravdati sljedećim primjerom.

Primjer 7. *Pretpostavimo kako osiguravajuće društvo za izračun premije koristi -1/Top Scale premijski sustav sa postotcima od premije osnovnog stupnja Malezijskog bonus-malus sustava osiguranja iz 3.2.3. Koristeći se rezultatima iz primjera 6, iznos premije kada je postignuto stacionarno stanje za osiguranike sa očekivanim godišnjim brojem zahtjeva $\sigma = 0.1$ izračunavamo na sljedeći način:*

$$\begin{aligned} &= \sum_j \textit{a priori premija } \cdot \pi_j \cdot (\textit{postotak od premije u } j\textit{-tom premijskom stupnju}/100\ ) \\ &= m\,[0.606531 + 0.063789 \cdot 0.75 + \cdots + 0.095163 \cdot 0.45] \\ &= 0.8419m. \end{aligned}$$

Rezultat pokazuje da se konačna premija dugoročno smanjuje s m na $0.8419m$. Dakle, osiguravajuće je društvo na početku procijenilo da za klijente u ovoj klasi treba posjedovati iznos m za pokrivanje mogućih šteta. Izračunali smo da će iznos koji će osiguranje imati od tih klijenata u stacionarnom stanju iznositi $0.8419m$. S financijskog stajališta, prikupljena premija neće biti dovoljna kako bismo njome pokrili očekivani trošak štete za te klijente, odnosno postoji mogućnost kako prikupljeni novac neće biti dovoljan te osiguravajuće društvo može bilježiti minuse. Kako bi se izbjegao takav scenarij, potrebno je povećati početnu premiju te se na taj način osigurati od mogućih gubitaka. Ovaj rezultat ne iznenađuje budući da u Malezijskom bonus-malus sustavu osiguranja ne postoje malus stupnjevi. Dakle, kako bi bonus-malus sustav bio u ravnoteži nužno je postojanje i bonus i malus stupnjeva jer bi tek tada postojala mogućnost da očekivana premija u stacionarnoj distribuciji bude jednaka m.

Osim što možemo izračunati iznos premije u stacionarnom stanju, možemo izračunavati i kretanje prosječne premije nakon n godina. Za izračun je potrebno poznavati vjerojatnost $p_{ij}^{(n)}$ odnosno (i,j)-ti element matrice $P^{(n)}$.

Primjer 8. *Promotrimo premiju kroz 30 godina dobivenu primjenom bonus-malus premijskog sustava Republike Hrvatske uz pretpostavku da a priori premija iznosi m, a vjerojatnost podnošenja k zahtjeva je* $p_k = \frac{e^{-0.1}(0.1)^k}{k!}$*, za* $k = 0, 1, 2, \ldots$.

Matrica 1-koračnih prijelaznih vjerojatnosti je:

$$
\begin{bmatrix}
0.9048 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0 & 0.0002 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0 & 0.0002 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0 & 0.0002 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0 & 0.0002 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0 & 0.0002 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0 & 0.0002 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0 & 0.0002 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0 & 0.0002 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0 & 0.0002 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0 & 0.0002 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0.0002 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0.0905 & 0 & 0 & 0.0045 & 0.0002 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0 & 0.0047 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0905 & 0.0047 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0 & 0.0952 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0 & 0.0952 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0 & 0.0952 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9048 & 0.0952
\end{bmatrix}.
$$

Prosječna premija za klijente za koje je $\lambda = 0.1$ *nakon implementacije bonus-malus premijskih stupnjeva iznosi:*

$$
\begin{aligned}
&= \sum_j \textit{a priori premija } \cdot (\textit{prosječan broj osiguranika u } j\textit{-tom premijskom stupnju}) \\
&\cdot (1 - \textit{postotak od premije u } j\textit{-tom premijskom stupnju/100 }) \\
&= 100 \cdot \left[\frac{\sum_i p_{i1}}{18} \cdot (1-0.5) + \frac{\sum_i p_{i2}}{18} \cdot (1-0.55) + \cdots + \frac{\sum_i p_{i18}}{18} \cdot (1-2.5) \right] \\
&\approx 94.
\end{aligned}
$$

Istim postupkom možemo izračunati i premiju u n-toj godini za $n = 1, 2, \ldots, 30$. *Za izračun ćemo koristiti R kao u [1] te dobivamo sljedeće iznose premija:*

94,88, 82, 77, 72, 68, 64, 60, 57, 55, 54, 52, 51, 50, 49, 49, 48, 48, 48, 47, 47, 47, 47, 47, 47, 47, 47, 46, 46, 46.

Glavni cilj bonus-malus sustava je korigirati pogreške napravljene a priori klasifikacijom rizika i na taj način točno odijeliti dobre od loših klijenata. Vrijeme potrebno za postizanje navedenoga je zapravo vrijeme potrebno za postizanje stacionarne distribucije. Jasno nam je da želimo da je ono što kraće. Ukoliko nas u n-toj godini zanima koliko smo daleko od stacionarne distribucije, to možemo izračunati na sljedeći način:

$$\sum_j \mid prosjek(p_{ij}^{(n)}) - \pi_j \mid .$$

Manja udaljenost implicira skorije dostizanje stacionarne distribucije odnosno bržu konvergenciju u stacionarnu distribuciju.

Primjer 9. *Koristeći pretpostavke i rezultate iz primjera 8 odredimo brzinu konvergencije u stacionarnu distribuciju bonus-malus premijskog sustava Republike Hrvatske.*

Najprije je potrebno odrediti stacionarnu distribuciju. Koristeći R kao u [1] dobivamo:

$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \\ \pi_6 \\ \pi_7 \\ \pi_8 \\ \pi_9 \\ \pi_{10} \\ \pi_{11} \\ \pi_{12} \\ \pi_{13} \\ \pi_{14} \\ \pi_{15} \\ \pi_{16} \\ \pi_{17} \\ \pi_{18} \end{bmatrix} = \begin{bmatrix} 0.6684 \\ 0.0703 \\ 0.0777 \\ 0.0859 \\ 0.0281 \\ 0.0240 \\ 0.0188 \\ 0.0088 \\ 0.0066 \\ 0.0045 \\ 0.0025 \\ 0.0017 \\ 0.0011 \\ 0.0007 \\ 0.0003 \\ 0.0003 \\ 0.0002 \\ 0.0001 \end{bmatrix}.$$

Tada je udaljenost od stacionarne distribucije po premijskim stupnjevima jednaka:

$$\left| \frac{\sum_i p_{i1}}{18} - \pi_1 \right| = 0.5679$$

$$\left| \frac{\sum_i p_{i2}}{18} - \pi_2 \right| = 0.02$$

$$\vdots$$

$$\left| \frac{\sum_i p_{i18}}{18} - \pi_{18} \right| = 0.0216,$$

te je konačna udaljenost od stacionarne distribucije jednaka:

$$\sum_j \left| \sum_i \frac{p_{ij}}{18} - \pi_j \right| = 1.2920.$$

Udaljenost od stacinarne distribucije kroz 10 godina iznosi redom:

$$1.2920, 1.2230, 1.1550, 1.0890, 1.0260, 0.9619, 0.8984, 0.8359, 0.7738, 0.7117.$$

Primijetimo kako bonus-malus premijski sustav Republike Hrvatske zahtijeva mnogo vremena za postizanje stacionarnog stanja. Općenito, složenijim sustavima je potrebno puno vremena za postizanje stacionarnosti što svakako predstavlja nedostatak jer je glavni cilj svakog takvog sustava u što kraćem vremenskom periodu odijeliti dobre od loših klijenata.

Literatura

[1] Actuarial Community, *Loss Data Analytics*, 2020.

[2] M. Benšić, *Predavanja za kolegij Statistika*, https://www.mathos.unios.hr/images/homepages/mirta/statistika/sve1.pdf

[3] M. Benšić, *Predavanja za kolegij Multivarijatna analiza*

[4] M. Denuit, X. Marechal, S. Pitrebois, J. F. Walhin *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, Wiley-Interscience, 2007.

[5] A. J. Dobson, A. G. Barnett, An introduction to generalized linear models, CRC Press, Boca Raton, 2018.

[6] R. Kaas, M. Goovaerts, J. Dhaene, M. Denuit, *Modern Actuarial Risk Theory: Using R*, Springer, 2008.

[7] A. Katrien, J. Crevecoeur, *Risk Modelling in Insurance : a collection of computer labs in R*, 2019.

[8] N. Šuvak, *Predavanja za kolegij Slučajni procesi*

Sažetak

Cilj svakog osiguravajućeg društva je na temelju određenih karakteristika što bolje procijeniti klijenta, dodijeliti mu određeni rizik i adekvatnu premiju. U tu svrhu, aktuari najprije koriste generalizirane linearne modele, neovisno o kojem se obliku osiguranja radi. U auto-osiguranju se uz generalizirane linearne modele koristi i bonus-malus sustav kojim se dodijeljena premija dodatno prilagođava klijentu. Primjena bonus-malus sustava je nužna jer u praksi nije moguće odrediti karakteristike vozača poput brzine refleksa ili primjerice agresivnosti u vožnji. Smatra se kako je adekvatna zamjena za te karakteristike broj šteta koje klijent prijavi osiguravajućem društvu tijekom jedne godine. Bonus-malus sustav se temelji upravo na tome te se njime klijente nagrađuje ili kažnjava ovisno o broju prijavljenih šteta tijekom protekle godine. Osim što su bonus-malus premijski stupnjevi laki za primjenu, važnost bonus-malus sustava je u tome što je on Markovljev lanac u diskretnom vremenu. Stacionarna distribucija Markovljevog lanca osiguravajućem društvu daje uvid u raspodjelu klijenata po premijskim stupnjevima u budućnosti te mu omogućuje izračunavanje prosječne premije kroz godine. Tako osiguravajuće društvo može znati kakve prihode očekivati te obzirom na to ima mogućnost dodatno prilagoditi premije po klasama.

Ključne riječi: generalizirani linearni modeli, auto-osiguranje, bonus-malus sustav određivanja premije osiguranja, Markovljevi lanci, stacionarna distribucija

Abstract

The goal of every insurance company is to assess a client as best as possible according to certain characteristics and then assign them a certain risk factor and adequate insurance premium. In order to achieve this, actuaries primarily use generalized linear models regardless of the insurance type. In car insurance companies, bonus-malus system of insurance is often used along with generalized linear models so as to ensure even further adjustment of the premium for each client. The application of the bonus-malus system has shown itself as necessary as actuaries are more often than not unable to ascertain drivers' individual characteristics, e.g. reaction time and style of driving (aggressiveness, for example). Number of traffic accidents reported to the insurance company during the one-year period has proven to be an adequate substitute for the information on drivers' individual characteristics. Bonus-malus system of insurance is founded on those premises, which means that the clients' premiums are adjusted according to their previous claims during the year. As well as being very simple in its application, bonus-malus system is also important as it is a discrete-time Markov chain. Stationary distribution of Markov chain provides insurance companies with insight into the distribution of clients according to different premium degrees in the future, which allows them to calculate average premiums during the course of several years. In that way, an insurance company can estimate its revenue and adjust the premiums by classes, according to those estimations.

Key words: generalized linear models, car insurance, bonus-malus system of insurance, Markov chain, stationary distribution

Životopis

Rođena sam 17.9.1996. godine u Slavonskom Brodu gdje sam završila osnovnu školu "Dragutin Tadijanović". Nakon osnovne škole upisujem opću gimnaziju "Matija Mesić" u Slavonskom Brodu koju završavam 2015. godine. Preddiplomski studij matematike na Odjelu za matematiku upisujem iste te godine i 2019. godine stječem naziv prvostupnice matematike uz završni rad Ekstremi realne diferencijabilne funkcije jedne i više realnih varijabli pod mentorstvom prof.dr.sc. Kristiana Sabe. U jesem 2019. godine upisujem diplomski studij matematike, smjer Financijska matematika i statistika na istoimenom fakultetu. U veljači 2022. godine završavam Pedagoško-psihološko-didaktičko-metodičku izobrazbu na Filozofskom fakultetu u Osijeku te sam trenutno zaposlena kao profesor matematike u gimnaziji Matija Mesić u Slavonskom Brodu.