Application possibilities of the logistic regression in the insurance sector

Možnost aplikace metody logistické regrese v oblasti pojišťovnictví

Martina Borovcová [1]

Abstract

Paper is focused on application possibilities of the logistic regression in the insurance sector. There are generally defined the areas of the possible application of the logistic regression, there is in detail the area of the non-life property insurance analyzed, more precisely accident insurance. By applying logistic regression are key factors defined with influence on the occurrence probability of insured accident, where influences of binary, categorial and continuous variables are analyzed. Impacts of selected factors are within one dimensional analysis quantified, as well.

Key words

Regression analysis, Logistic Regression, Generalized Linear Models, Binary variables, Categorial variables, Continuous variables,

JEL Classification: C20, G22

1 Úvod

S ohledem na skutečnost, že regresní modely patří mezi nejčastěji využívané přístupy k analýze dat nejrůzmější povahy, je reálné i jejich využití v oblasti pojišťovnictví. Nejčastěji se nám při vyslovení slova regrese vybaví regrese lineární, méně často nelineární nebo logistická, i když právě logistická regrese je již nejméně tři desetiletí standardní metodou v západoevropské a americké vědě včetně společenské.

Cílem příspěvku je proto nalézt možnost využití logistické regrese v oblasti pojišťovnictví. Na konkrétním příkladu pak pomocí logistické regrese zhodnotit, zda vybrané faktory jsou určujícími pro vznik pojistné události a tento vztah dále kvantifikovat. V článku je nejprve definována a vysvětlena regresní analýza a stručně jsou popsány regresní modely (klasický lineární model, obecný lineární model). Následně je zmíněna podstata logistické regrese, včetně odhadu koeficientů metodou maximální věrohodnosti a poté je provedena analýza vlivu pohlaví a lokality na vznik pojistné události ve smyslu havárie motorového vozidla.

2 Regresní analýza

Statistické metody, pomocí nichž odhadujeme hodnotu určité náhodné veličiny na základě znalosti veličin jiných, označujeme jako regresní analýzu. Přitom náhodná veličina, jejíž

[1] Ing. Martina Borovcová, Ph.D., VŠB-TU Ostrava, Ekonomická fakulta, katedra financí, Sokolská třída 33, 702 21 Ostrava, martina.borovcova@vsb.cz.

hodnota je odhadována, může být označena také jako závisle proměnná, cílová proměnná, proměnná vysvětlovaná nebo také odezva, regresand. Naproti tomu veličina, jejíž znalost již máme, je nezávisle proměnná, proměnná vysvětlující, regresor. Ne vždy je použita nezávisle proměnná jediná. Často vystupuje regresorů několik, přičemž může jít o další veličiny, nebo funkce menšího počtu veličin.

Modelování vztahů mezi vysvětlující a vysvětlovanou proměnnou patří mezi základní aktivity, se kterými je možné se setkat ve statistice. Obvyklý je předpoklad, že závisle proměnná je náhodnou veličinou s normálním rozdělením. Pro odvození modelu je pak zpravidla použita metoda nejmenších čtverců.

Je-li však závisle proměnná znakem binárním, nikoli spojitým statistickým znakem, může nastat problém. V takovém případě by k odhadu parametrů bylo použití regresní analýzy s odhadem regresních koeficientů prostřednictvím metody nejmenších čtverců problematické.

Podstatou řešení regrese je:

- stanovení nejlepšího regresního modelu, spočívající v určení matematické rovnice, která bude popisovat závislost *y* na *x*,
- stanovení parametrů modelu, související se stanovením nejlepších odhadů parametrů *β*,
- stanovení statistické významnosti modelu, související s určením, zda nalezený model přispěje ke zpřesnění odhadu závisle proměnné oproti použití pouhého průměru,
- interpretace výsledků zjištěných modelem z hlediska zadání.

3 Regresní model

Regresní model předpokládá, že nezávislá proměnná je nenáhodná, pevně daná či určená, a závislá proměnná je náhodná, tzn. měřená. V praxi však tento předpoklad nebývá vždy splněn, a často jsou obě či všechny veličiny naměřené. V takovém případě jde o model korelační.

Mezi regresními modely jsou rozlišovány regresní modely lineární, které mají lineární postavení parametrů a jejichž grafem je zpravidla přímka, parabola či hyperbola a regresní modely nelineární, jejichž grafem mohou být například některé typy neuronových sítí.

Je-li závisle proměnná *Y* skalár nebo vektor z nějakého lineárního prostoru, bývá v takovém případě úloha regrese obvykle formulována jako úloha hledání podmíněné střední hodnoty jakožto funkce nezávisle proměnných. Nejčastěji se tato funkce předpokládá v nějakém obecném tvaru závislém na neznámých regresních parametrech, regresních koeficientech, a tyto koeficienty se poté odhadují na základě pozorovaných dat. Nejčastějším případem je lineární regresní funkce. Druhou základní možností je, že závisle proměnná *Y* je diskrétní a není tudíž definována její střední hodnota. Regresní analýza v této situaci je označována jako diskriminační analýza a jejím úkolem je hledat podmíněné pravděpodobnosti toho, že zkoumaný objekt patří do jednotlivých tříd. Typickými metodami používanými pro řešení úloh tohoto typu jsou Pearsonova lineární diskriminační analýza, logistická regrese a metody z nich odvozené.

3.1 Klasický lineární model

Lineární regresní modely mají lineární postavení parametrů. Klasickým lineárním modelem je chápán model ve tvaru

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

kde **Y** je vektor *n* hodnot vysvětlované proměnné, **X** je matice hodnot vysvětlujících proměnných o rozměrech $n \cdot (k+1)$, **β** je vektor *p* neznámých parametrů $(p=k+1)$ a **ε** je vektor *n* hodnot náhodné složky.

Přitom platí podmínky klasického lineárního modelu.

1. Střední hodnota náhodné složky je nulová. Náhodná složka tedy nepůsobí systematickým způsobem na hodnoty vysvětlované proměnné *Y*.

$$E(\varepsilon_i) = 0 \text{ pro každé } i=1,2,\ldots,n.$$

2. Rozptyl náhodné složky je konstantní. Variabilita náhodné složky nezávisí na hodnotách vysvětlujících proměnných a tudíž i podmíněná variabilita vysvětlované proměnné nezávisí na hodnotách vysvětlujících proměnných a je rovna neznámé kladné konstatně σ^2.

$$E(\varepsilon_i) = \sigma^2 \text{ pro každé } i=1,2,\ldots,n.$$

3. Kovariance náhodné složky je nulová. Hodnoty náhodné složky jsou nekorelované a z toho vyplývá i nekorelovanost různých dvojic pozorování vysvětlované proměnné *Y*.

$$Cov(\varepsilon_i \varepsilon_j) = 0 \text{ pro každé } i \neq j = 1,2,\ldots,n.$$

4. **X** je nestochastická (nenáhodná) matice. Znamená to tedy, že vysvětlující proměnné jsou nenáhodné.
5. Matice **X** má plnou hodnost, $h(\mathbf{X})=k+1\leq n$, přičemž *n* je počet pozorování. Podmínka vyžaduje, aby mezi vysvětlujícími proměnnými nebyla funkční lineární závislost, tedy v matici **X** nesmí existovat lineárně závislé sloupce. Počet vysvětlujících proměnných nesmí být větší než počet pozorování. V praxi by měl být počet pozorování výrazně větší než počet vysvětlujících proměnných.
6. ε_i mají normální rozdělení pro každé $i=1,2,\ldots, n$. Z této podmínky vyplývá normalita i pro vysvětlovanou proměnnou *Y*. Náhodný vektor **Y** má potom n-rozměrné normální rozdělení s vektorem středních hodnot $\mathbf{X\beta}$ a kovarianční maticí $\sigma^2\mathbf{I}_n$.
7. Na vektor **β** nejsou kladeny žádné omezující podmínky. Parametry $\beta_j, j=1,2,\ldots,k$ tak mohou nabývat libovolných hodnot.

Matematická formulace regresního modelu je znázorněna ve schématu 1.

Schéma 1: Regresní model

$$\underset{\mathbf{Y}}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}} = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nk} \end{bmatrix}}_{\mathbf{X}} \cdot \underset{\boldsymbol{\beta}}{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_k \end{bmatrix}} + \underset{\boldsymbol{\varepsilon}}{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

3.2 Obecný lineární model (GLM)

Výše uvedené podmínky či předpoklady 1. – 7. představují omezení, která v řadě aplikací nejsou splněna. Když v obecném lineárním modelu nahradíme tyto podmínky podmínkami obecnějšími, dospějeme ke zobecnění lineárního modelu.

Zobecněný lineární model poskytuje obecný rámec pro vytváření jednotné třídy modelů, které pracují se spojitými i kategorizovanými nezávislými proměnnými. Tato třída obsahuje regresní modely, obecné lineární modely, logistickou regresi pro binární závislou proměnnou a log-lineární modely. Zobecněný lineární model navíc dovoluje pomocí funkcí spojení vytvářet modely s vlastním rozdělením chyby modelu.

Ve zobecněném lineárním modelu je tedy cílem nalézt model, který zmenšuje celkovou devianci (úměrnou rozdílu logaritmů věrohodnostních funkcí mezi úplným modelem a nulovým modelem s jedním parametrem). Takový model může být vytvářen i postupně, mohou být do modelu zařazovány ty regresory, které nejvíce snižují devianci vzhledem k aktuálnímu modelu se zařazenými k parametry. Regresory ve zobecněném lineárním modelu mohou být i kvalitativní (faktory) a regresory mohou být i interakce (součiny) původních regresorů, takže pomocí zobecněného modelu je možné odhadovat parametry i složitých modelů analýzy rozptylu.

4 Logistická regrese

Cílem analýzy, která využívá metodu regrese, je nalézt co nejlepší, nejúspornější a současně věcně smysluplný model, který popíše vztah mezi závislou proměnnou a skupinou nezávislých proměnných. Je-li vysvětlovaná proměnná spojitá, obracíme se k regresi lineární, není-li spojitá, pak k regresi logistické. Metoda logistické regrese není omezená jen na případ, kdy vysvětlovaná proměnná je binární. I když pro tuto situaci byla logistická regrese původně vyvinutě a je interpretačně, ale i jinak nejsnazší. Existují však metody a také programy, které pracují s případy, kdy kategorizovaná závislá proměnná není binární, a dokáží respektovat požadavek, aby ji považovaly za ordinální.

4.1 Možnost využití logistické regrese v oblasti pojišťovnictví

S ohledem na výše uvedenou podstatu logistické regrese je šíře jejího využití v oblasti pojišťovnictví evidentní. Ať už se jedná o měření solventnosti pojišťoven, hospodaření pojišťoven, hodnocení úrovně pojistného trhu, odhad výše technických rezerv, stanovení pojistného v jednotlivých odvětvích pojištění a další, ve všech případech je možná snaha o vytvoření modelu a nalezení vztahu mezi konkrétními závislými a nezávislými proměnnými.

V navazujícím textu bude analyzován vliv pohlaví a lokality na vznik pojistné události, ve smyslu havárie motorového vozidla.

4.2 Formulace modelu

Předpokládejme, že máme binární veličinu Y_i charakterizující vznik nebo nevzniknutí pojistného události i-tého pojistníka, tedy

$$Y_i = \begin{cases} 1 \text{ pro vznik pojistné události,} \\ 0 \text{ pro "nevznik" pojistné události,} \end{cases} \text{pro } i = 1,\dots,n,$$

kde n je počet pojistníků. Každý tento pojistník je charakteristický vektorem $\mathbf{x}_i = (1, x_{1i}, x_{2i},\dots,x_{ki})$ obsahující k prvků, Strišš, Valečková, Valecký (2010, str. 206-207).

Pravděpodobnost vzniku pojistné události i-tého pojistníka $P_i = P(Y_i=1)$ na základě jeho charakteristického vektoru $\mathbf{x}_i$ lze vyjádřit funkcí $F(\boldsymbol{\beta};\mathbf{x}_i)$, jenž je monotonně rostoucí

$F'(\boldsymbol{\beta};\mathbf{x}_i)\geq 0$ a má definiční obor $(-\infty,\infty)$ a obor hodnot (0,1). Platí tedy, že $F(-\infty)=0$ a $F(+\infty)=1$ a funkci pravděpodobnosti odpovědi lze psát jako

$$P_i=F(\boldsymbol{\beta};\mathbf{x}_i), \tag{2}$$

kde $\boldsymbol{\beta}$ je vektor parametrů $(\beta_0, \beta_1,\dots, \beta_k)$.

Tyto vlastnosti jsou splněny kumulativní distribuční funkcí logistického rozdělení ve tvaru

$$P_i = P(Y_i = 1) = F(\boldsymbol{\beta};\mathbf{x}_i) = \frac{e^{\beta' x_i}}{1+e^{\beta' x_i}}, \tag{3}$$

která je zároveň funkcí pravděpodobnosti vzniku pojistné události. Pravděpodobnost negativní varianty, nevzniknutí pojistné události, lze pak vyjádřit ve tvaru

$$1-P_i = P(Y_i = 0) = 1-F(\boldsymbol{\beta};\mathbf{x}_i) = \frac{1}{1+e^{\beta' x_i}}. \tag{4}$$

Definujme dále podíl pravděpodobnosti vzniku a nevzniknutí pojistné události známé také jako šance (odds) ve tvaru

$$\frac{\pi}{1-\pi} = \frac{P(Y_i = 1)}{P(Y_i = 0)} = e^{\beta' x_i}, \tag{5}$$

a dále tzv. logitovou transformaci (log-odds, logit) vztahu (5)

$$\ln\left[\frac{\pi}{1-\pi}\right] = \boldsymbol{\beta}'\mathbf{x}_i = g(\mathbf{x}_i). \tag{6}$$

4.3 Odhad parametrů modelu

K odhadu neznámých parametrů $\boldsymbol{\beta}$ je nejčastěji používána metoda maximální věrohodnosti. Tato metoda spočívá v nalezení věrohodnostní funkce l ( ), která je posléze maximalizována. Mějme pravděpodobnost kladné odpovědi i-tého respondent charakteristického vektorem $\mathbf{x}_i$, tedy

$$P(Y_i = 1|\mathbf{x}_i) = \pi(\mathbf{x}_i), \tag{7}$$

a dále pravděpodobnost negativní varianty, nevzniknutí pojistné události

$$P(Y_i = 0|\mathbf{x}_i) = 1 - P(Y_i = 1|\mathbf{x}_i) = 1-\pi(\mathbf{x}_i). \tag{8}$$

Sdružená pravděpodobnost kladných a záporných variant vzniku pojistné události lze poté vyjádřit ve tvaru

$$P(Y_i|\mathbf{x}_i) = \pi(\mathbf{x}_i)^{Y_i}\left[1-\pi(\mathbf{x}_i)\right]^{(1-Y_i)}. \tag{9}$$

Jsou-li jednotlivá pozorování nezávislá, pak věrohodnotstní funkce je určena jako součin sdružených pravděpodobností pro všechny pojistníky, tedy

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{N} \pi(\mathbf{x}_i)^{Y_i} [1 - \pi(\mathbf{x}_i)]^{(1-Y_i)}. \tag{10}$$

Odhad parametrů metodou maximální věrohodnoti je získán maximalizací logaritmu rovnice (10) ve tvaru

$$L(\boldsymbol{\beta}) = \ln l(\boldsymbol{\beta}) = \sum_{i=1}^{N} Y_i \cdot \ln(\pi(\mathbf{x}_i)) + (1 - Y_i) \cdot \ln(1 - \pi(\mathbf{x}_i)) \tag{11}$$

za podmínek

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = 0 \text{ pro } j = 1,\dots,k. \tag{12}$$

Odhad metodou maximální věrohodnoti bývá prováděn pomocí iteračních algoritmů, přičemž nejčastěji je používána Newton-Raphsonova metoda. Princip této metody spočívá v aproximaci logaritmu věrohodnoctví funkce v okolí počátečního odhadu pomocí prvních tří členů Taylorova rozvoje, viz Pecáková (2007), přičemž počáteční odhad lze získat například metodou nejmenších čtverců.

4.3.1 Ověření vlivu pohlaví na vznik pojistné události

V této části příspěvku jsou analyzovány vybrané faktory, pohlaví a lokalita, a zjištěna jejich statistická významnost na dané hladině spolehlivosti. Nejprve je ověřeno, zda faktor pohlaví ovlivňuje vznik pojistné události a poté je analyzován tentýž důsledek vlivu lokality.

Analýza je provedena pomocí dat získaných z pojistného kmene konkrétní pojišťovny a skládá se z datového vzorku 54 824 smluv. Smlouvy, tvořící pojistný kmen, jsou uzavřeny na produkt neživotního pojištění, pojištění majetku (vozu) ve smyslu havarijního pojištění. Jsou použita data za rok 2009, přičemž časová expozice smlouvy je jeden rok.

Obr. 1: Odhad parametru logistického modelu (pohlaví pojistníka)

note: gender dropped due to collinearity
Iteration 0: log likelihood = -11131.721
Iteration 1: log likelihood = -11106.985
Iteration 2: log likelihood = -11106.84
Iteration 3: log likelihood = -11106.84

Logistic regression Number of obs = 54824
LR chi2(1) = 49.76
Prob > chi2 = 0.0000
Log likelihood = -11106.84 Pseudo R2 = 0.0022

claim	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
gender	.283731	.0397847	7.13	0.000	.2057545	.3617075
_cons	-3.01228	.0245697	-122.60	0.000	-3.060436	-2.964125

Dle získaných výsledků lze říci, že koeficient je statisticky významný na hladině spolehlivosti 95%, což dokazují výsledky ve čtvrtém hodnotovém sloupci Tab.1. Je tedy možné tvrdit, že pohlaví pojistníka je významným faktorem ovlivňujícím vznik pojistné události.

V dalším kroku je analyzováno, zda lokalita pojistníka má vliv na vznik pojistné události. Opět je provedena logistická regrese, kde vysvětlovanou veličinou je vznik pojistné události a vysvětlující veličinou je lokalita.

4.3.2 Ověření vlivu lokality na vznik pojistné události

V této části příspěvku je provedena analýza vlivu lokality pojistníka na vznik pojistné události pomocí logistické regrese.

Obr. 2: Odhad parametrů logistického modelu (lokalita pojistníka)

Iteration 0: log likelihood = -11131.721
Iteration 1: log likelihood = -11085.942
Iteration 2: log likelihood = -11085.435
Iteration 3: log likelihood = -11085.435

Logistic regression — Number of obs = 54824
LR chi2(13) = 92.57
Prob > chi2 = 0.0000
Log likelihood = -11085.435 — Pseudo R2 = 0.0042

claim	Coef.	Std. Err.	z	P>|z	[95% Conf. Interval]	
kraj1 - Jihočeský	.034873	.0872912	0.40	0.690	-.1362146	.2059605
kraj2- Královéhradecký	-.4078399	.1245666	-3.27	0.001	-.6519859	-.163694
kraj3 - Pardubický	-.5590701	.1320296	-4.23	0.000	-.8178433	-.3002968
kraj4 - Ústecký	.0192431	.0820108	0.23	0.814	-.1414951	.1799812
kraj5 - Jihomoravský	-.3137378	.0826583	-3.80	0.000	-.4757451	-.1517305
kraj6 - Liberecký	-.2879822	.1123663	-2.56	0.010	-.5082161	-.0677483
kraj7 - Plzeňský	-.263994	.1153009	-2.29	0.022	-.4899795	-.0380084
kraj8 - Zlínský	-.3832842	.1150247	-3.33	0.001	-.6087285	-.1578399
kraj9 - Karlovarský	.0366602	.1149661	0.32	0.750	-.1886693	.2619896
kraj10 - Moravskoslezský	-.4235861	.0745511	-5.68	0.000	-.5697036	-.2774686
kraj12 - Vysočina	-.223489	.1088716	-2.05	0.040	-.4368735	-.0101045
kraj13 - Olomoucký	-.5095669	.1113598	-4.58	0.000	-.7278282	-.2913056
kraj14 - Středočeský	-.1033222	.0582132	-1.77	0.076	-.2174179	.0107736
_cons	-2.747702	.0351344	-78.21	0.000	-2.816564	-2.67884

Z výsledků je opět zřejmé, že koeficient, jako takový, je statisticky významný na hladině spolehlivosti 95 %. Protože je posuzována kategoriální veličina, bylo nutné z této jedné proměnné vytvořit nové kontrastní proměnné, které určitým způsobem korespondují s původními kategoriemi. Za lokalitu je v tomto případě považován kraj. Do analýzy je zahrnuto 14 krajů, přičemž referenční kategorií, která byla vynechána, byl zvolen kraj Praha.

Obr.3: Test významnosti rozdělení krajů

test (kraj1 kraj2 kraj3 kraj4 kraj5 kraj6 kraj7 kraj8 kraj9 kraj10 kraj12 kraj13 kraj14)

(1) kraj1 = 0
(2) kraj2 = 0
(3) kraj3 = 0
(4) kraj4 = 0
(5) kraj5 = 0
(6) kraj6 = 0
(7) kraj7 = 0
(8) kraj8 = 0
(9) kraj9 = 0
(10) kraj10 = 0
(11) kraj12 = 0
(12) kraj13 = 0
(13) kraj14 = 0

chi2(13) = 89.17
Prob > chi2 = 0.0000

Z výsledků provedeného testu je patrná statistická významnost modelu pro vstupní proměnnou kraj.

Konkrétně je možné tvrdit, viz čtvrtý hodnotový sloupec Tab. 2, že koeficient Královéhradecký kraj (0,001), Pardubický kraj (0,000), Jihomoravský kraj (0,000), Liberecký kraj (0,010), Plzeňský kraj (0,022), Zlínský kraj (0,001), Moravskoslezský kraj (0,000), Vysočina (0,040) a Olomoucký kraj (0,000) jsou statisticky významné na hladině spolehlivosti 95 %. Zatímco koeficienty Jihočeský kraj (0,690), Ústecký kraj (0,814), Karlovarský kraj (0,750) a Středočeský kraj (0,076) jsou na hladině spolehlivosti 95 % statisticky nevýznamné.

5 Závěr

Cílem příspěvku bylo nalézt možnosti využití logistické regrese v oblasti pojišťovnictví, především však potom na konkrétním příkladu pomocí logistické regrese zhodnotit, zda vybrané faktory jsou určujícími pro vznik pojistné události a tento vztah dále kvantifikovat. V článku byla nejprve definována a vysvětlena regresní analýza a stručně byly popsány regresní modely (klasický lineární model, obecný lineární model). Následně byla zmíněna podstata logistické regrese, včetně odhadu koeficientů metodou maximální věrohodnosti a poté byla provedena analýza vlivu pohlaví a lokality na vznik pojistné události ve smyslu havárie motorového vozidla.

Z výše uvedených výsledků vyplývá, že vznik pojistné události je determinován jak pohlavím pojistníka, tak lokalitou.

References

[1] HARDIN, J. W., HILBE J. M., 2007. *Generalized Linear Models and Extensions*. Texas: Stata Press.

[2] HOSMER, D.W., LEMESHOW, S., 2000. *Applied Logistic Regression.* New Jersey: John Wiley & Sons.

[3] PECÁKOVÁ, I., 2007. Logistická regrese s vícekategoriální vysvětlovanou proměnnou. *Acta Oeconomica Pragensia*, roč. 15, č. 1, pp. 86-96.

[4] STRIŠŠ, J., VALEČKOVÁ, J., VALECKÝ, J., 2010. Aplikace logistické regrese v měření spokojenosti zákazníků. *Rozvoj marketingu v teórii a praxi*, Žilinská univerzita v Žilině, pp. 205-210.

[5] ŠIMURDA, M., 2008. Zobecněný lineární model (GLM). Dostupné na: http://www.actuaria.cz/upload/GLM_SMM_MFF_web.pdf.

[6] WEISBERG, S., 2005. *Applied Linear Regression.* New Jersey: John Wiley & Sons.

[7] ZMEŠKAL, Z., 2004. *Finanční modely*. Praha: EKOPRESS.