

---

*Soňa Coss*

## MODELY S BINÁRNOU ZÁVISLE PREMENNOU<sup>1</sup>

### Úvod

Základnou štatistickou metódou pri skúmaní vplyvu vysvetľujúcich premenných na vysvetľovanú premennú je lineárna regresná analýza. Predpokladom jej použitia je, aby závisle premenná bola spojitá. V mnohých analýzach však táto podmienka nie je splnená a závisle premenná nadobúda diskkrétne hodnoty vo forme kategoriálnej premennej. Riešením je využitie skupiny modelov s diskrétnou kategoriálnou premennou, akými sú napr. binárne, ordinálne, multinomické modely. Základnými modelmi sú logitový a probitový model.

Modely s binárnou závisle premennou využívame v prípade, že kategoriálna závisle premenná nadobúda iba dve možné hodnoty, teda ide o binárnu premennú. Nezávislé premenné v modeli môžu byť spojité aj kategoriálne.

### 1 ZOVŠEOBECNENÝ LINEÁRNY MODEL

Zovšeobecnené lineárne modely (Generalized Linear Models - GLZ) predstavujú širokú skupinu modelov. Môžeme ich využiť pri analýze spojitých aj kategoriálnych nezávislých premenných a na posúdenie ich vplyvu na spojitú alebo diskrétnu závisle premennú. Patria sem modely lineárnej regresie a analýzy rozptylu pre spojitú závisle premennú, ako aj modely pre kategoriálnu závisle premennú. Modely s binárnou premennou sú teda špeciálnymi prípadmi zovšeobecného lineárneho modelu.

V každom zovšeobecnenom lineárnom modeli uvažujeme s tromi základnými zložkami modelu. Sú to náhodná zložka, deterministická (systematická) zložka a link (spojnica modelu, linková funkcia) [1].

Náhodná zložka identifikuje závisle premennú  $Y$  a robí úsudky o jej pravdepodobnostnom rozdelení. Deterministická zložka zahŕňa vysvetľujúce, nezávisle premenné  $X_1, X_2 \dots X_k$  a popisuje vzťah medzi nimi. Lineárnu kombináciu vysvetľujúcich premenných v zovšeobecnenom lineárnom modeli nazývame lineárny prediktor. Ak označíme očakávanú hodnotu  $Y$ , t.j. strednú hodnotu jej rozdelenia pravdepodobnosti

$$\mu = E(Y) \tag{1.1}$$

---

<sup>1</sup> VEGA: 1/0092/15 Moderné prístupy k navrhovaniu komplexných štatistických prieskumov, ktorého vedúcim je prof. Ing. Milan Terek, PhD.

vieme, že hodnota  $\mu$  sa mení v závislosti od úrovne vysvetľujúcich premenných. Deterministická zložka teda špecifikuje vysvetľujúce premenné, ktoré vystupujú ako lineárny prediktor na pravej strane modelovej rovnice v tvare:

$$\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1.2)$$

Link popisuje funkčný vzťah medzi náhodnou a deterministickou zložkou. Špecifikuje vzťah medzi očakávanou hodnotou  $\mu = E(Y)$  a vysvetľujúcimi premennými v lineárnom prediktore. Na popis tohto vzťahu používame monotónnu funkciu  $g(\mu)$ , ktorú nazývame link alebo linková funkcia.

$$g(\mu) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1.3)$$

Linková funkcia  $g(\cdot)$  teda zovšeobecňuje vzťah medzi závisle premennou a nezávislými premennými v modeli. Tento vzťah môže mať podobu buď lineárneho alebo nelineárneho efektu na závisle premennú. Najjednoduchšia forma linkovej funkcie je tzv. identity link (identický, zhodný link), ktorý má tvar:

$$g(\mu) = \mu \quad (1.4)$$

Pomocou tejto funkcie modelujeme priamo strednú hodnotu vysvetľovanej premennej a využívame ju pri lineárnej regresii spojitých premenných.

$$\mu = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1.5)$$

Nie vždy však lineárna regresia postačuje na adekvátne popísanie vzťahu medzi pozorovanými premennými. Najčastejšie to býva z dvoch hlavných príčin. Jednou z nich je rozdelenie pravdepodobnosti závisle premennej a druhou je nelineárny vzťah medzi prediktorom a závisle premennou [2].

Rozdelenie pravdepodobnosti skúmanej závisle premennej nemusí byť normálne. Ak pracujeme s kategoriálnou závisle premennou, rozdelenie pravdepodobnosti nie je dokonca ani spojité. V tomto prípade môže ísť o diskrétno binomické, multinomické, prípadne Poissonovo rozdelenie.

Druhým obmedzením je, že vzťah medzi prediktorom a závisle premennou nemusí mať nutne lineárnu podobu. Skôr sa stretáme s tým, že prediktory sú vzhľadom na závisle premennú v nelineárnom vzťahu. Použitie identity linku potom stráca svoje opodstatnenie a riešením je voľba inej vhodnej linkovej funkcie.

Zovšeobecný lineárny model je vhodný na použitie v oboch spomenutých prípadoch. Ak ide o závisle premennú s diskrétnym rozdelením pravdepodobnosti, a aj v prípade, že závisle premenná nie je v lineárnom vzťahu k prediktorom. Zovšeobecnenie

umožňuje, aby mala náhodná zložka aj iné pravdepodobnostné rozdelenie ako normálne a umožňuje aj modelovanie nejakej funkcie strednej hodnoty závisle premennej. Avšak je potrebné model vhodne transformovať pomocou linkovej funkcie.

## 2 MODEL Y S BINÁRNOU ZÁVISLE PREMENNOU

Pri modelovaní kategoriálnej závislej premennej sa snažíme odhadnúť model, pomocou ktorého by sme špecifikovali vzťah závisle premennej od jednej alebo množiny vysvetľujúcich premenných. Chceme zistiť efekt vysvetľujúcich premenných na kategoriálnu závisle premennú, ktorá nadobúda diskkrétne obmeny znaku. Vysvetľujúce premenné v modeli môžu byť spojené alebo kategoriálne.

Inými slovami, hľadáme vzťah medzi skupinou vysvetľujúcich premenných a pravdepodobnosťou nastania sledovaného javu. Zároveň predpokladáme, že táto pravdepodobnosť je lineárnou funkciou vlastností štatistickej jednotky, ktoré charakterizujú jednotlivé vysvetľujúce premenné v modeli.

Mnoho kategoriálnych závisle premenných nadobúda iba dve možné obmeny znaku. Ide napr. o rozhodnutie kúpiť, či nekúpiť daný produkt, status nezamestnaného alebo zamestnaného, volebné preferencie k danému, resp. inému kandidátovi, diagnostikovanie alebo neprítomnosť danej choroby a pod. Takéto premenné nazývame binárne (dichotomické, alternatívne) premenné.

Kategoriálne závisle premenné však môžu nadobúdať aj viac ako dve diskkrétne obmeny, napr. výber zo sortimentu rôznych produktov, sociálny status rodiny, počet plánovaných detí v rodine a pod. Vtedy ide o tzv. viackategoriálne, množné, teda multinomické znaky.

V ďalšom texte sa už obmedzíme na popis modelov pre binárnu závisle premennú. Ide o skupinu modelov, ktoré na modelovanie závisle premennej využívajú práve dichotomické znaky.

Označme binárnu závisle premennú  $Y$  a jej dve možné obmeny znaku 1 a 0. Hodnotu 1 za predpokladu, že daný jav nastal a hodnotu 0, ak daný jav nenastal. Táto binárna závisle premenná má alternatívne rozdelenie pravdepodobnosti a zvykne sa nazývať aj Bernoulliho premennou.

Jej rozdelenie je dané pravdepodobnosťou výskytu želanej udalosti, ak daný jav nastal:<sup>2</sup>

---

<sup>2</sup> Ak modelujeme pravdepodobnosť nastatia javu  $P(Y=1)$ , ide samozrejme o podmienenú pravdepodobnosť  $P(Y=1|x)$  nastatia javu  $Y$  za podmienky výskytu premennej  $X$ , teda že jav nastane pri danej úrovni  $x$ . V ďalšom texte sa obmedzíme len na výraz  $P(Y=1)$  a budeme predpokladať, že ide o podmienenú pravdepodobnosť.

$$P(Y = 1) = \pi \quad (1.6)$$

a pravdepodobnosťou, ak daný jav nenastal:

$$P(Y = 0) = 1 - \pi . \quad (1.7)$$

Toto rozdelenie pravdepodobnosti má strednú hodnotu:

$$E(Y) = 1 \cdot P(Y = 1) + 0 \cdot P(Y = 0) = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi \quad (1.8)$$

a rozptyl:

$$D(Y) = \pi(1 - \pi) \quad (1.9)$$

Pri  $n$  nezávislých pozorovaniach na binárnej závisle premennej s parametrom  $\pi$ , počet prípadov nastatia javu má binomické rozdelenie s parametrami  $n$  a  $\pi$ . Teda predpokladáme, že náhodná zložka v modeli má binomické rozdelenie.

Keďže hodnota pravdepodobnosti  $\pi$  sa môže meniť v závislosti od zmeny hodnoty  $x$  premennej  $X$ , preto použijeme na označenie tejto pravdepodobnosti symbol  $\pi(x)$ .

### 3 LINEÁRNY PRAVDEPODOBNOSTNÝ MODEL

Najjednoduchší spôsob modelovania vplyvu premených  $X_1, X_2 \dots X_k$  na očakávanú hodnotu premennej  $Y$  využíva formu lineárnej regresie v tvare rovnice<sup>3</sup>:

$$E(Y) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1.10)$$

Ak chceme modelovať pravdepodobnosť nastatia javu, môžeme zapísať<sup>4</sup>:

$$\pi(\mathbf{x}) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1.11)$$

Tento zovšeobecnený lineárny model s binomickou náhodnou zložkou a identickým linkom nazývame lineárny pravdepodobnostný model.

Predpokladáme v ňom, že pravdepodobnosť nastatia javu sa mení lineárne od úrovne vysvetľujúcich premenných  $X$ . Keďže v lineárnom pravdepodobnostnom modeli predikujeme pravdepodobnosť nastatia javu, interpretácia parametrov  $\beta$  sa v tomto prípade

<sup>3</sup> Ide o podmienenú strednú hodnotu závisle premennej  $E(Y|\mathbf{x})$  za podmienky výskytu pri danej úrovni hodnôt premenných  $X_1, X_2 \dots X_k$ .

<sup>4</sup> Kde  $\mathbf{x}$  je vektor hodnôt nezávisle premenných  $X_1, X_2 \dots X_k$ .

oproti lineárnemu regresnému modelu zmení. Parameter  $\beta_1$  predstavuje zmenu pravdepodobnosti pri jednotkovej zmene  $X_1$ , za predpokladu, že ostatné premenné zostanú nezmenené. Vyjadruje parciálny efekt pôsobenia premennej  $X_1$  na pravdepodobnosť výskytu javu, pod kontrolou ostatných premenných v zmysle *ceteris paribus*.

Ak chceme využiť metódu regresnej analýzy, musíme transformovať binárnu závisle premennú na premennú so spojitémi hodnotami, ktorá bude funkciou pravdepodobnosti výskytu udalosti.

Narazíme však na tri vážne obmedzenia [6].

1. Keďže lineárny pravdepodobnostný model predpokladá interpretáciu predikovaných hodnôt  $Y$  ako pravdepodobností, nastáva problém, ak predikované hodnoty ležia mimo interval  $(0,1)$ . Očakávaná hodnota závisle premennej totiž nie je ohraničená na tento interval, ale nadobúda hodnoty na celom obore reálnych čísiel. Model teda môže predikovať aj pravdepodobnosti  $\pi(\mathbf{x}) < 0$  alebo  $\pi(\mathbf{x}) > 1$ .
2. Rozptyl náhodnej zložky nie je konštantný. Keďže stredná hodnota náhodnej zložky je  $E(\varepsilon) = 0$  a jej rozptyl je  $D(\varepsilon) = \pi(\mathbf{x})[1 - \pi(\mathbf{x})]$ . Znamená to, že rozptyl náhodnej zložky závisí od  $\mathbf{x}$ . Náhodná zložka je teda heteroskedastická.
3. Rozdelenie náhodnej zložky nie je normálne, keďže analyzujeme binárnu závisle premennú.

Všetky uvedené obmedzenia lineárneho pravdepodobnostného modelu vedú k tomu, aby sme na modelovanie binárnej závisle premennej využili iné alternatívne modely.

#### 4 NELINEÁRNY PRAVDEPODOBNOSTNÝ MODEL

Riešením je transformovať pôvodný model pomocou vhodnej linkovej funkcie. Snažíme sa pritom dodržať dva základné predpoklady. Po prvé, aby sa hodnoty vysvetľujúcich premenných, ktoré môžu nadobúdať hodnoty cez celý obor reálnych čísiel, transformovali na pravdepodobnosti v intervale  $(0,1)$ . Po druhé, aby transformácia bola monotónna, teda zachovala vlastnosť, že nárast  $X$  je sprevádzaný nárastom (resp. poklesom) závislej premennej pre všetky hodnoty  $X$ .

Pravdepodobnosť závisle premennej teda transformujeme do podoby lineárnej funkcie nezávislých premenných pomocou linkovej funkcie. Pravdepodobnosť je potom vyjadrená ako nelineárny pravdepodobnostný model.

Riešenie sa ponúka vo voľbe funkcie  $g(\cdot)$ , ktorá transformuje nelineárny vzťah medzi prediktorom a pravdepodobnosťou nastatia javu. Pri tejto funkcii predpokladáme, že je rýdzo monotónna a existuje k nej inverzná funkcia  $g^{-1}(\cdot)$ .

Aldrich a Nelson [2] popisujú množstvo funkcií, ktoré vhodne transformujú pravdepodobnosti do nelineárnych pravdepodobnostných modelov. Najčastejšie sa však na tento účel využívajú distribučné funkcie logistického a normovaného normálneho rozdelenia. Distribučná funkcia logistického rozdelenia je základom pre logistickú regresiu. Pri probitovej regresii zase využívame distribučnú funkciu normovaného normálneho rozdelenia. Logitový a probitový model sú najznámejšie nelineárne pravdepodobnostné modely.

## 5 LOGISTICKÝ MODEL

Logistický regresný model je jedným z modelov, ktoré sa využívajú pri analýze binárnej závisle premennej. Modelovaná pravdepodobnosť nastatia javu je v nelineárnom vzťahu vzhľadom na sadu vysvetľujúcich premenných. Ako sme už spomenuli vyššie, cieľom nelineárnych pravdepodobnostných modelov je transformovať pravdepodobnosti výskytu javu pomocou linkovej funkcie tak, aby ich vzťah s vysvetľujúcimi premennými nadobudol lineárny charakter.

Funkcia, ktorú používame pri logitovej transformácii sa nazýva logit.

$$g(\pi(\mathbf{x})) = \text{logit}(\pi(\mathbf{x})) \quad (1.12)$$

Transformovať pravdepodobnosti nastatia javu do logitov znamená, vyjadriť ich vo forme logaritmu šanci:

$$\text{logit}(\pi(\mathbf{x})) = \ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) \quad (1.13)$$

Logitová funkcia zabezpečí transformáciu závisle premennej na spojitú premennú. Transformácia cez logaritmus šanci premení pravdepodobnosti výskytu javu, ktoré sa pohybujú iba v rozpätí intervalu  $(0, 1)$ , na logity, ktoré sa pohybujú sa na celom intervale  $(-\infty, \infty)$ .

Následne môžeme logistický regresný model vyjadriť:

$$\ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1.14)$$

$$\text{logit}(\pi(\mathbf{x})) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1.15)$$

V tomto regresnom modeli je logit podmienenej strednej hodnoty vyjadrený ako lineárna kombinácia vysvetľujúcich premenných. A teda rovnako ako v lineárnej regresii odhadujeme regresné koeficienty tohto modelu.

Lineárna kombinácia vysvetľujúcich premenných sa však nerovná priamo pravdepodobnosti, ale prirodzenému logaritmu šancí. Teda pravdepodobnosť výskytu javu je na základe takto definovaného modelu v nelineárnom vzťahu k vysvetľujúcim premenným. Aby sme mohli predikovať pravdepodobnosti nastatia javu pri zvolenej kombinácii premenných  $X$ , na vyjadrenie pravdepodobnosti použijeme spätnú transformáciu:

$$\pi(\mathbf{x}) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}} \quad (1.16)$$

## 6 PROBITOVÝ MODEL

Pri transformácii nelineárnych pravdepodobnostných modelov však môžeme využiť aj inú linkovú funkciu. Ak predpokladáme, že náhodná zložka v modeli má binomické rozdelenie, ďalšou takouto funkciou je probit.<sup>5</sup>

$$g(\pi(\mathbf{x})) = \text{probit}(\pi(\mathbf{x})) \quad (1.17)$$

Probitová linková funkcia je vtedy definovaná ako:

$$\text{probit}(\pi(\mathbf{x})) = \Phi^{-1}(\pi(\mathbf{x})) \quad (1.18)$$

Funkcia  $\Phi^{-1}$  je inverzná funkcia k distribučnej funkcii normovaného normálneho rozdelenia, v podstate 100π% kvantil normovaného normálneho rozdelenia.

Probitový regresný model získame, ak pravdepodobnosti nastatia javu podrobíme probitovej transformácii. Kým pôvodné pravdepodobnosti môžu nadobúdať hodnoty v intervale (0,1), hodnoty probitov sú už v rozpätí  $(-\infty, \infty)$ .

Takto transformované probity sú už spojitou premennou a môžu byť použité ako závisle premenná v lineárnom modeli.

Dostaneme tak vyjadrenie probitovej regresnej funkcie:

$$\Phi^{-1}(\pi(\mathbf{x})) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1.19)$$

$$\text{probit}(\pi(\mathbf{x})) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (1.20)$$

Ľavá strana rovnice predstavuje probit, čo je skratka pre probability unit, teda probitovo transformovanú hodnotu pravdepodobnosti. Vzťah medzi vysvetľujúcimi premennými a probitom má už lineárny charakter.

<sup>5</sup> V niektorých zdrojoch sa zvykne probitová linková funkcia nazývať aj normit.

Probitový regresný model teda umožňuje zistiť, ako sa mení probit v závislosti od hodnôt vysvetľujúcich premenných. Odhadom regresných koeficientov modelu môžeme posúdiť aký je vplyv jednotlivých vysvetľujúcich premenných na pravdepodobnosť nastatia javu.

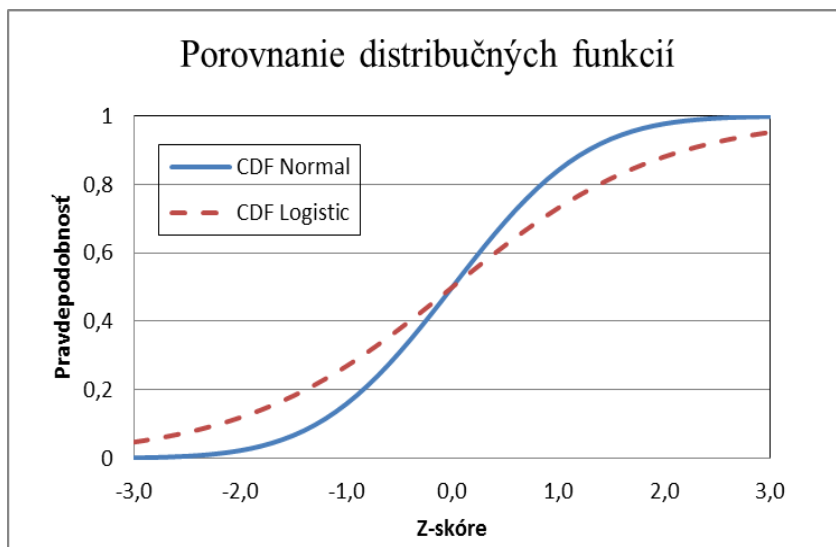
Druhou dôležitou vlastnosťou tohto modelu je, že aj v tomto prípade môžeme predikovať pravdepodobnosť nastatia udalosti pre zvolenú kombináciu hodnôt vysvetľujúcich premenných. Na tento účel si spätne transformujeme pravdepodobnosti podľa vzťahu:

$$\pi(\mathbf{x}) = \Phi(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k) \quad (1.21)$$

kde  $\Phi$  je distribučná funkcia normovaného normálneho rozdelenia.

## 7 DISTRIBUČNÉ FUNKCIE LOGISTICKÉHO A NORMOVANÉHO NORMÁLNEHO ROZDELENIA

Pri logistickej regresii predpokladáme, že sigmoidná krivka distribučnej funkcie logistického rozdelenia korešponduje s nelineárnym vzťahom medzi pravdepodobnosťou a vysvetľujúcimi premennými v modeli. Keďže krivka distribučnej funkcie normovaného normálneho rozdelenia, je tiež sigmoidnou krivkou s veľmi podobným priebehom hodnôt, usúdilo sa, že aj táto krivka je rovnako vhodná na transformáciu nelineárneho pravdepodobnostného modelu.



Z grafu veľmi dobre vidieť, že krivky distribučných funkcií logistického a normovaného normálneho rozdelenia sa veľmi podobajú. Distribučná funkcia



normovaného normálneho rozdelenia je jemne strmšia a teda klesá rýchlejšie. Logistická krivka má tzv. ťažšie konce, avšak rozdiely medzi krivkami sú veľmi malé.

Krivky majú však niekoľko spoločných vlastností. Obidve patria medzi S-krivky, sú monotónne rastúce a existujú k nim inverzné funkcie. Obe tieto krivky sú tiež symetrické okolo hodnoty pravdepodobnosti  $\pi(\mathbf{x}) = 0,5$ . Využitie týchto dvoch funkcií súvisí aj s tým, že nadobúdajú hodnoty na intervale  $(0,1)$ , rovnako ako pravdepodobnosti  $\pi(\mathbf{x})$  v modeli.

Keďže distribučná funkcia logistického aj normovaného normálneho rozdelenia sa veľmi podobajú, odhady parametrov z oboch modelov sa už líšia výraznejšie. Na odhad regresných koeficientov v oboch modeloch používame metódu maximálnej vierohodnosti. Rovnako, ako aj pri testovaní významnosti modelu ako celku.

Odhady parametrov získaných jednotlivými metódami nie je možné navzájom porovnávať. Logaritmické rozdelenie má rozptyl rovný  $\pi^2 / 3$  a pri normovanom normálnom rozdelení je rozptyl rovný 1. To znamená, že odhady z logitovho modelu musia byť pri porovnávaní s odhadmi probitového modelu následne prenasobené  $3^{1/2} / \pi$ .

## Záver

Probitový a logitový model sú alternatívnymi modelmi pri modelovaní s binárnou závisle premennou. Hoci cieľ skúmania vo forme predikcie pravdepodobnosti aj predpoklady pri vstupných premenných sú zhodné u oboch modelov, každá z metód využíva vlastný prístup pri transformácii premenných a odlišuje sa aj interpretáciou dosiahnutých výsledkov.

## Kľúčové slová

binárna závisle premenná, logit model, probit model

## Klasifikácia JEL

C35

## LITERATÚRA

- [1] AGRESTI, A. 1996. *An Introduction to Categorical Data Analysis*. New York : Wiley, 1996. 290 s. ISBN 0-471-11338-7.
- [2] ALDRICH, J. H. – NELSON F. D. 1984. *Linear Probability, Logit, and Probit models*. Iowa : Sage publications, 1984. 97 s. ISBN 0-8039-2133-0.
- [3] BOROOAH, V.K. 2002. *Logit and Probit. Ordered and Multinomial Models*. Iowa : Sage publications, 2002. 104 s. ISBN 0-7619-2242-3.

- [4] LIAO, T. F. 1994. *Interpreting Probability Models. Logit, Probit, and Other Generalized Linear Models*. Iowa : Sage publications, 1994. 88 s. ISBN 0-8039-4999-5.
- [5] PAMPEL, F. C. 2000. *Logistic Regression – a primer*. Iowa : Sage publications, 2000. 96 s. ISBN 0-7619-2010-2.
- [6] PINDYCK, R. S. – RUBINFELD D. L. 1981. *Econometric Models and Econometric Forecasts*. International student edition, 1981. ISBN 0-0791-3292-8.
- [7] SCOTT LONG, J. – FREEZE J. 2006. *Regression Models for Categorical Dependent Variables*. Iowa : Sage publications. ISBN 1-5971-8011-4.
- [8] Generalized linear model. <http://www.statsoft.com/textbook/>.
- [9] Analýza kategorizovaných dat v sociálnych viedách. <http://is.muni.cz/el/1423/jaro2006/SOC419/um>.

## RESUMÉ

Probitový a logitový model patria medzi najčastejšie používané zovšeobecnené lineárne modely, ktoré využívame v prípade binárnej závisle premennej. Hlavný rozdiel medzi logistickým a probitovým modelom predstavuje predpoklad o rozdelení náhodnej zložky. Obe linkové funkcie patria medzi krivky sigmoidného tvaru, tzv. S-krivky s hodnotami v intervale 0 až 1. V probitovom modeli, využívame ako linkovú funkciu, inverznú funkciu normovaného normálneho rozdelenia, ktorá transformuje lineárny prediktor na očakávané početnosti. V prípade logitovho modelu modelujeme logaritmus šancí nastatia javu ako lineárnu kombináciu prediktora. Logit je v tomto prípade kvantilovou funkciou logistického rozdelenia.

## SUMMARY

Probit and logit models are among the most used models of generalized linear models in the case of binary or dichotomous dependent variables. The difference between Logistic and Probit models lies in the assumption about the distribution of the errors. The logit and probit are both sigmoid functions with a domain between 0 and 1. In the probit model, the inverse standard normal distribution of the probability is modeled as a linear combination of the predictors. In the logit model the log odds of the outcome is modeled as a linear combination of the predictor variables and the logit is the quantile function of the logistic distribution.

## Kontakt

Ing. Soňa Coss, PhD., Katedra štatistiky, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave, Dolnozemska cesta 1b, 852 35 Bratislava, e-mail: [sona.coss@gmail.com](mailto:sona.coss@gmail.com)