Logit model pre výber spojenia v mestskej hromadnej doprave

doc. Ing. Ľudmila Jánošíková, PhD.[1], doc. Mgr. Jiří Slavík, CSc.[2], Ing. Michal Koháni, PhD.[3]

Abstrakt: Príspevok sa zaoberá modelovaním rozhodovania cestujúcich v mestskej hromadnej doprave (MHD) pri výbere dopravného spojenia. Analyzuje atribúty dopravného spojenia a ukazuje, že rozhodovanie možno vyjadriť pomocou multinomial logit modelu. Relatívny vplyv atribútov na preferenciu cestujúceho možno odhadnúť z údajov o označených cestovných lístkoch. Postup je ilustrovaný na príklade MHD v Žiline. Z prípadovej štúdie vyplýva, že cestujúci pri výbere dopravného spojenia kladú najväčší dôraz na dobu jazdy.

Kľúčové slová: model diskrétneho výberu, metóda maximálnej vierohodnosti, návrh liniek mestskej hromadnej dopravy

1 Úvod

V mestskej hromadnej doprave je pomerne častým javom, že cestujúci má niekoľko možností, ako sa dopraviť do cieľa svojej cesty. Môže si vybrať nielen spoj, ale často aj zastávku, na ktorej nastúpi, vystúpi, alebo prestúpi na inú linku. V mnohých prípadoch má k dispozícii niekoľko priamych liniek, ale pochopiteľne aj dopravné spojenie s prestupom z jedného spoja na iný. Od výberu spojenia závisí rozdelenie prúdov cestujúcich na úseky siete. Úseková intenzita prúdu cestujúcich udáva, koľko cestujúcich chce cestovať daným úsekom za jednotku času. Inými slovami vyjadruje dopyt po dopravnej službe. Ponukou dopravnej služby sa rozumie počet miest v dopravných prostriedkoch, ktoré prejdú úsekom za jednotku času. Pomer medzi ponukou a dopytom je jedným z dôležitých ukazovateľov pri posudzovaní kvality systému mestskej hromadnej dopravy. Vyhodnotením tohto pomeru na jednotlivých úsekoch dopravnej siete zistíme, či sú trasy a frekvencie liniek navrhnuté dobre, či zodpovedajú potrebám verejnosti. Ak tomu tak nie je, ak sú úseky, kde ponuka výrazne prevyšuje dopyt, alebo naopak, je nižšia než dopyt, treba pristúpiť k reorganizácii sústavy liniek [3, 4, 7].

Úsekové intenzity prúdov cestujúcich sa dajú určiť pomocou dopravného prieskumu. Táto metóda má však dve veľké nevýhody:

1. je drahá a časovo náročná,
2. výsledky prieskumu odrážajú len aktuálny stav, ale neumožňujú predpovedať, ako sa toky zmenia v dôsledku zmien v systéme MHD, napr. pri zmene tarify, cestovného poriadku, trás a frekvencií liniek.

Inou možnosťou, ako určiť úsekové záťaže, je použiť matematický model pre pridelenie prúdov cestujúcich (*passenger assignment*). Pridelenie prúdov cestujúcich znamená, že pre každú agregovanú požiadavku na prepravu z nejakej počiatočnej do cieľovej zastávky (prvok OD matice) určíme jej rozdelenie na jednotlivé prípustné cesty, resp. dopravné spojenia. Ak

[1] Žilinská univerzita v Žiline, Fakulta riadenia a informatiky, Katedra dopravných sietí, Univerzitná 1, 010 26 Žilina, Slovenská republika, tel.: (421-41) 513 42 00, fax: (421-41) 565 10 15, e-mail: ludmila.janosikova@fri.uniza.sk

[2] dtto, tel.: (421-41) 513 42 03, e-mail: jiri.slavik@fri.uniza.sk

[3] dtto, tel.: (421-41) 513 40 19, e-mail: michal.kohani@fri.uniza.sk

vieme, koľko cestujúcich si vyberie každú z možných ciest, potom úsekovú intenzitu vypočítame ako súčet počtu cestujúcich na tých cestách, ktoré prechádzajú daným úsekom.

Model pre pridelenie prúdov cestujúcich odráža citlivosť cestujúcich na tie atribúty dopravných spojení, ktoré ovplyvňujú voľbu cesty. Takými atribútmi sú napr. doba jazdy vo vozidle, interval medzi spojmi, počet prestupov atď. Atribúty dopravných spojení sú v modeli nezávislými premennými. Môžeme ich vyhodnotiť, ak poznáme trasy a frekvencie liniek a dobu jazdy na jednotlivých úsekoch. Atribúty sa násobia parametrami, ktoré vyjadrujú, do akej miery jednotlivé atribúty zavážia pri voľbe cesty. Aby sme mohli odhadnúť tieto parametre, potrebujeme poznať:

1. záujem o prepravu medzi každou dvojicou dopravných okrskov vyjadrenú v tvare OD matice,
2. správanie cestujúcich, t.j. pravidlá, podľa ktorých si vyberajú dopravné spojenie.

Aj keď model odhadneme na základe súčasného, existujúceho systému, umožňuje predpovedať budúcnosť, čiže dovoľuje vypočítať zmeny v úsekových intenzitách spôsobené zmenou atribútov.

Existuje viacero spôsobov, ako modelovať voľbu dopravného spojenia. Modely sa v zásade líšia tým, akú úroveň informovanosti prisudzujú cestujúcemu, čiže koľko toho cestujúci o spojení vie. V najjednoduchšom prípade sa predpokladá, že cestujúci o možnostiach spojenia medzi východiskovou a cieľovou zastávkou vie len to, ktoré linky medzi nimi chodia, nepozná cestovný poriadok ani cestovný čas, alebo mu na nich nezáleží. Po príchode na východiskovú zastávku použije prvý spoj, ktorý príde a smeruje k cieľovej zastávke. Iný model predpokladá, že cestujúci pozná cestovný čas a vždy použije najrýchlejšie spojenie. Obidva tieto modely sú však dosť vzdialené od skutočného správania cestujúcich. Pravidelne cestujúci, ktorí denne dochádzajú do práce alebo do školy, majú dobrý prehľad o dopravných spojeniach, poznajú cestovný poriadok aj dobu prepravy a prispôsobia im svoj príchod na zastávku aj voľbu dopravného spojenia. Aj u nepravidelne cestujúcich možno predpokladať dobrú informovanosť, ak sú cestovné poriadky zverejnené. Ak predpokladáme, že cestujúci pozná hodnoty jednotlivých atribútov dopravného spojenia (cestovný poriadok, dobu jazdy dopravného prostriedku, potrebu prestupu), môžeme sa oprávnene domnievať, že cestujúci si vyberie spojenie, ktoré mu najviac vyhovuje. Avšak nikdy nemôžeme predpovedať správanie cestujúcich s určitosťou. Vhodnejšie je modelovať správanie cestujúcich pomocou modelov diskrétneho výberu [1, 2, 5, 6, 8]. Účelom týchto modelov je analyzovať a predvídať rozhodovanie skupiny ľudí v situácii, kedy si musia vybrať jednu alternatívu z množiny navzájom sa vylučujúcich alternatív. Zo skupiny modelov diskrétneho výberu sme si zvolili tzv. multinomial logit model, ktorý je pomerne jednoduchý a dá sa použiť vtedy, keď alternatívy sú navzájom nezávislé.

2 Multinomial logit model

Multinomial logit model patrí medzi **pravdepodobnostné modely výberu**. To znamená, že vyjadruje pravdepodobnosť, s akou si užívateľ vyberie každú alternatívu, a nie jednoznačnú predikciu. Aby užívateľ (cestujúci) dokázal alternatívy navzájom porovnať, musí existovať funkcia obsahujúca atribúty alternatívy a charakteristiky užívateľa, ktorej hodnota udáva úžitok (*utility*) každej alternatívy pre daného užívateľa. V pravdepodobnostných modeloch je funkcia úžitku rozdelená na dva členy. Prvý člen reprezentuje tú časť úžitku, ktorú analytik (ten, kto zostavuje model) pozná, môže pozorovať a merať. Táto časť sa nazýva deterministickou alebo pozorovateľnou zložkou úžitku. Druhý člen predstavuje tú časť úžitku, ktorú analytik nepozná. Tento chybový člen vyjadruje skutočnosť, že analytik nie je schopný merať alebo špecifikovať všetky atribúty, ktoré ovplyvňujú hodnotu úžitku, ktorú má daná alternatíva pre užívateľa.

Formálne môžeme **funkciu úžitku** z alternatívy j pre užívateľa i vyjadriť ako

$$u_{ij} = v_{ij} + \varepsilon_{ij},$$

kde v_{ij} je deterministická (pozorovateľná) časť úžitku a ε_{ij} je chybový člen (náhodná premenná, ktorá reprezentuje tú časť úžitku, ktorú analytik nepozná).

Deterministický člen v_{ij} je funkciou premenných (atribútov alternatívy a charakteristík užívateľa) a parametrov, ktoré treba odhadnúť. Charakteristikami užívateľa sa rozumejú socioekonomické ukazovatele, ako sú vek, pohlavie, vzdelanie, príjem, ktoré ovplyvňujú jeho preferencie pri voľbe jednej z alternatív. Kvôli jednoduchosti sa v_{ij} najčastejšie formuluje ako aditívna funkcia. Pre modelovanie voľby dopravného spojenia budeme predpokladať, že symbol i neoznačuje len jedného užívateľa, ale skupinu užívateľov s rovnakými charakteristikami, ktorí majú rovnakú množinu prípustných alternatív (t.j. i je skupina ľudí, ktorí v danom časovom období cestujú z rovnakej východiskovej do rovnakej cieľovej zastávky). Potom deterministická zložka funkcie úžitku je len funkciou atribútov alternatív:

$$v_{ij} = b_1\, x_{ij1} + b_2\, x_{ij2} + \ldots + b_p\, x_{ijp},$$

kde $x_{ij1}, \ldots, x_{ijp}$ sú hodnoty atribútov popisujúcich alternatívu j, p je počet atribútov a $b_1, \ldots, b_p$ sú koeficienty, ktoré vyjadrujú smer vplyvu jednotlivých atribútov na úžitok a ich dôležitosť.

O náhodnej zložke funkcie úžitku ε_{ij} nevieme prakticky nič. Voľbou rozdelenia pravdepodobnosti tejto zložky dostávame rôzne modely. Ak zvolíme Gumbelovo rozdelenie pravdepodobnosti, dostávame tzv. **multinomial logit model**.

V nasledujúcom popise modelu budeme používať takéto označenie:

- I — množina všetkých skupín cestujúcich (t.j. množina všetkých dvojíc zastávok, medzi ktorými existuje viac ako jedno dopravné spojenie);
- A_i — množina alternatív prípustných pre skupinu i (t.j. množina možných dopravných spojení medzi i-tou dvojicou zastávok);
- n_i — počet cestujúcich v i-tej skupine;
- n_{ij} — počet tých cestujúcich v i-tej skupine, ktorí si vybrali j-tu alternatívu (zrejme platí $\sum_{j \in A_i} n_{ij} = n_i$);
- N — celkový počet cestujúcich, $N = \sum_{i \in I} n_i$;
- $P(i,j)$ — pravdepodobnosť, že si i-ta skupinu cestujúcich vyberie j-tú alternatívu.

V multinomial logit modeli je $P(i,j)$ definovaná takto:

$$P(i, j) = \frac{\exp(v_{ij})}{\sum_{t \in A_i} \exp(v_{it})} = \frac{\exp(b_1\, x_{ij1} + \ldots + b_p\, x_{ijp})}{\sum_{t \in A_i} \exp(b_1 x_{it1} + \ldots + b_p x_{itp})} \tag{1}$$

Pravdepodobnosti $P(i, j)$ však obsahujú neznáme koeficienty $b_1, \ldots, b_p$.. Tie môžeme odhadnúť **metódou maximálnej vierohodnosti** z hodnôt atribútov x_{ijk} pre $k = 1, 2, \ldots, p$ a zo známeho počtu n_{ij} tých cestujúcich, ktorí si vybrali alternatívu j. Za predpokladu, že voľba je nezávislá pre všetky i a j, nájdeme pravdepodobnosť, že medzi cestujúcimi sa vyskytol práve pozorovaný výber alternatív, takto:

Zostavíme funkciu vierohodnosti:

$$L(b_1, \ldots, b_p) = \prod_{i \in I} \frac{n_i!}{\prod_{j \in A_i} n_{ij}!} \prod_{j \in A_i} P(i, j)^{n_{ij}} \tag{2}$$

a hľadáme také hodnoty koeficientov $b_1, \ldots, b_p$, pre ktoré táto funkcia nadobúda maximum. (Lebo ak sa vyskytol práve tento výber, tak zrejme preto, že bol najpravdepodobnejší). Obvykle sa nehľadá maximum pôvodnej funkcie vierohodnosti, ale jej logaritmu ln $L(\boldsymbol{b})$, ktoré sa nachádza v rovnakých bodoch $b_1, ..., b_p$. Metóda je založená na tom, že v bodoch, kde leží maximum, sú parciálne derivácie podľa jednotlivých premenných $b_1, ..., b_p$ rovné nule. V tomto prípade to platí aj naopak, a teda stačí nájsť také $b_1, ..., b_p$, v ktorých sú parciálne derivácie rovné nule. Pretože ln $L(\boldsymbol{b})$ má tvar súčtu, ľahšie sa derivuje. Teda

$$\ln L(\boldsymbol{b}) = \sum_{i\in I} \ln\left(\frac{n_i!}{\prod_{j\in A_i} n_{ij}!}\right) + \sum_{i\in I}\sum_{j\in A_i} n_{ij}\ln(P(i,j)). \tag{3}$$

Potom

$$\frac{\partial \ln L(\boldsymbol{b})}{\partial b_k} = \sum_{i\in I} n_i \sum_{j\in A_i} x_{ijk}\left(\frac{n_{ij}}{n_i} - P(i,j\right) \text{pre}\, k = 1,2,\ldots,p. \tag{4}$$

Pre nájdenie koeficientov $b_1, ..., b_p$ teda vyriešime sústavu nelineárnych rovníc

$$\frac{\partial \ln L(\boldsymbol{b})}{\partial b_k} = \sum_{i\in I} n_i \sum_{j\in A_i} x_{ijk}\left(\frac{n_{ij}}{n_i} - P(i,j\right) = 0\, \text{pre}\, k = 1,2,\ldots,p. \tag{5}$$

Táto sústava sa dá riešiť Newtonovou metódou. Avšak riešenie je len odhadom neznámych koeficientov; označíme ich $\hat{b}_1,\ldots,\hat{b}_p$. Ako sme už povedali, veľkosť koeficientov b_k je vlastne váha, ktorú dávame hodnotám atribútov. Ak niektoré $b_k = 0$, znamená to, že k-ty atribút ničím neprispieva k úžitku, teda je nevýznamný. Ale vypočítané hodnoty $\hat{b}_1,\ldots,\hat{b}_p$ sú náhodné premenné a aj keď aktuálne vypočítaná hodnota je rôzna od nuly, nemožno vylúčiť, že príslušný atribút je nevýznamný. Preto je dôležitou súčasťou modelu štatistický test na nulovosť koeficientov b_k.

Budeme teda testovať štatistickú hypotézu:

$H_0 : b_k = 0$ proti alternatívnej hypotéze,

$H : b_k \neq 0$ pre $k = 1, 2, ..., p$.

Je to dvojstranný test. Zvolíme si hladinu významnosti testu α. Testovacie kritérium testu je

$$T = \frac{\hat{b}_k}{\sqrt{D\hat{b}_k}}\sqrt{N}\ , \tag{6}$$

kde $D\hat{b}_k$ je rozptyl odhadu $\hat{b}_k$. Ak platí hypotéza H_0, má testovacie kritérium T Studentovo rozdelenie s parametrom (N-1). Avšak rozptyl $D\hat{b}_k$ nepoznáme a opäť ho musíme odhadnúť. Odhadneme ho pomocou Rao-Cramérovej nerovnosti. Platí:

$$D\hat{b}_k \geq \frac{1}{-E\left(\frac{\partial^2 \ln L(\hat{\boldsymbol{b}})}{\partial \hat{b}_k^2}\right)} \tag{7}$$

Odhadneme teda tento rozptyl jeho dolným odhadom. Musíme ale najprv nájsť $\dfrac{\partial^2 \ln L(\hat{\boldsymbol{b}})}{\partial \hat{b}_k^2}$:

$$\frac{\partial^2 \ln L(\hat{\boldsymbol{b}})}{\partial \widehat{b}_k^2} = \frac{\partial}{\partial \widehat{b}_k} \sum_{i \in I} n_i \sum_{j \in A_i} x_{ijk} \left(\frac{n_{ij}}{n_i} - P(i, j \right) =$$
$$= -\sum_{i \in I} \sum_{j \in A_i} n_i x_{ijk} P(i,j) \left(x_{ijk} - \sum_{t \in A_i} P(i,t) x_{itk} \right) \text{pre}\, k = 1,2,\ldots, p \tag{8}$$

Strednú hodnotu z Rao-Cramérovej nerovnosti aproximujeme hodnotou tejto druhej derivácie a dostaneme:

$$D\widehat{b}_k = \frac{1}{-\left(\frac{\partial^2 \ln L(\hat{\boldsymbol{b}})}{\partial \widehat{b}_k^2} \right)} \quad \text{pre } k = 1, 2, ..., p. \tag{9}$$

Teraz už môžeme vypočítať T a vyhodnotíme test. Nájdeme kritickú hodnotu Studentovho rozdelenia s (N-1) stupňami voľnosti t_α^{N-1}. Platí pre ňu: $F\left(t_\alpha^{N-1}\right) = 1 - \alpha / 2$, kde F je distribučná funkcia Studentovho rozdelenia s parametrom (N-1). Ak $|T| \le t_\alpha^{N-1}$, hypotézu H_0 prijímame, ak $|T| < t_\alpha^{N-1}$, hypotézu H_0 zamietame a prijímame alternatívnu hypotézu, na hladine významnosti α. Prijatie hypotézy H_0 teda znamená, že príslušný koeficient b_k môže mať aj hodnotu nula, a teda príslušný atribút je nevýznamný.

Pri odhade parametrov modelu sa ešte obvykle udáva miera zhody modelu s pozorovanými údajmi. Jednou z možných mier je tzv. koeficient pomeru vierohodností (*likelihood ratio index*), ktorý je definovaný takto:

$$\rho^2 = 1 - \frac{\ln L(\hat{\boldsymbol{b}})}{\ln L(0)}, \tag{10}$$

kde $\ln L(\hat{\boldsymbol{b}})$ je hodnota logaritmickej funkcie vierohodnosti (3) v odhadnutých parametroch $\hat{b}_1, \ldots, \hat{b}_p$ a ln L(0) je jej hodnota, keď všetky parametre sú nulové. Ak odhadnuté parametre modelu nedávajú lepší výsledok než nulové parametre (t.j. ak odhadnutý model nie je lepší než žiadny model), potom $\ln L(\hat{\boldsymbol{b}}) = \ln L(0)$, a teda $\rho^2 = 0$. V opačnom krajnom prípade predpokladajme, že odhadnutý model je taký dobrý, že každú voľbu cestujúceho dokáže predpovedať s istotou. Potom funkcia vierohodnosti pre odhadnuté parametre musí byť 1, pretože pravdepodobnosť, že sa vyskytnú tie voľby, ktoré cestujúci urobili, je 1. A pretože logaritmus jednej je nula, $\ln L(\hat{\boldsymbol{b}}) = 0$ a $\rho^2 = 1$. Koeficient pomeru vierohodností teda nadobúda hodnoty od 0 do 1. Možno ho použiť na porovnanie dvoch modelov kalibrovaných s tými istými dátami, kedy sa dá povedať, že model s vyššou hodnotou koeficientu je lepší.

3 Experimenty

Ako prípadovú štúdiu pre modelovanie správania cestujúcich sme zvolili mestskú hromadnú dopravu v Žiline. Žilina je stredne veľké mesto v severozápadnej časti Slovenska, má 85 302 obyvateľov (údaj k 31.5.2009) a rozlohu 80 km^2.

MHD v Žiline prevádzkuje spoločnosťou Dopravný podnik mesta Žiliny, s.r.o. (ďalej len DPMŽ). Cez deň premáva 8 trolejbusových liniek a 10 autobusových liniek. V nočných hodinách obsluhuje územie mesta jedna nočná autobusová linka.

Pomocou modelu diskrétneho výberu popísaného v predchádzajúcej kapitole sme sa pokúsili zistiť, na základe čoho sa cestujúci rozhodujú pre jedno z niekoľkých možných dopravných spojení. Za atribúty dopravného spojenia sme považovali:

- dobu jazdy vo vozidle;
- čas chôdze pri prestupe alebo z výstupnej zastávky do cieľa cesty;
- počet prestupov (0 alebo 1);
- interval medzi spojmi linky (prvej v prípade spojenia s prestupom).

Počty cestujúcich n_{ij} sme odvodili z údajov, ktoré nám poskytol DPMŽ . Išlo o údaje zaznamenané v označovacích zariadeniach cestovných lístkov (papierových alebo elektronických), ktoré sú umiestnené vo vozidlách. Ak cestujúci použije elektronický cestovný lístok na čipovej karte, zariadenie zosníma identifikačné číslo čipovej karty a zaznamená údaje o spoji, nástupnej zastávke a čase, kedy bol lístok označený. Pomocou týchto údajov je potom možné sledovať všetky jazdy, ktoré cestujúci podnikol v daný deň. Za východiskovú zastávku jeho cesty považujeme nástupnú zastávku prvej jazdy. Hoci nepoznáme cieľ cesty, môžeme ho z údajov odvodiť. Ak rovnaká karta bola použitá viackrát za deň a medzi dvoma označeniami uplynula viac než hodina, potom sa môžeme domnievať, že cieľom cesty bola nástupná zastávka jazdy, ktorú cestujúci podnikol neskôr s dostatočným časovým odstupom. Čiže vieme určiť jednak celkový počet cestujúcich medzi dvoma zastávkami, ako aj to, koľko z nich použilo každú z možných alternatív dopravného spojenia. Z údajov za jeden pracovný týždeň (od 12. do 16.10.2009) sme na kalibráciu modelu vybrali tie dvojice zastávok, kde cestujúci použili aspoň dve spojenia. Pretože v priebehu dňa sa mení jednak charakteristika cestujúcich (intenzity prúdov, účel cesty, demografické ukazovatele), jednak trasy a frekvencie liniek, treba výpočet modelu urobiť zvlášť pre každú periódu, ktorá sa obvykle v dopravnej prevádzke sleduje. Zamerali sme sa na rannú špičku (od 6:00 do 8:00) a sedlo v priebehu dňa (8:00 – 14.00). V tab. 1 je štatistika sledovaných atribútov pre všetky alternatívy zahrnuté do modelu.

Vplyv jednotlivých atribútov na rozhodovanie cestujúcich vyjadrujú parametre b_1, b_2, b_3 a b_4, ktoré zodpovedajú zvoleným atribútom. Aby sme mohli určiť aj dôležitosť jednotlivých atribútov, treba ich upraviť tak, aby mali približne rovnaké hodnoty. Preto sme čas a interval normovali na hodnotu z intervalu $(0;1\rangle$ tak, že sme ich vydelili maximálnou hodnotou daného atribútu zo všetkých alternatív.

Tab. 2 a 3 udávajú odhady parametrov b_1, b_2, b_3 a b_4 pre raňajšiu špičku a dopoludňajšie sedlo. Pri každom parametri je v tabuľkách uvedená jeho smerodajná odchýlka a pomer odhadovaného parametra a smerodajnej odchýlky (tzv. *t-ratio*). Štatistický test v obidvoch skúmaných periódach vyhodnotil všetky parametre ako významné na hladine významnosti 0,05. To znamená, že všetky štyri atribúty majú vplyv na rozhodovanie cestujúcich. Všetky parametre majú podľa očakávania zápornú hodnotu, čiže čím je hodnota atribútov vyššia, tým je daná alternatíva menej atraktívna. Najväčšiu úlohu pri výbere dopravného spojenia má v oboch prípadoch čas. Poradie dôležitosti ostatných atribútov sa v jednotlivých periódach líši. Druhým

Tabuľka 1. Štatistika atribútov dopravných spojení

	Ranná špička (6:00 – 8:00)		Sedlo (8:00 – 14.00)	
	Stredná hodnota	Smerodajná odchýlka	Stredná hodnota	Smerodajná odchýlka
Čas jazdy [min]	15,37	8,35	15,33	8,33
Čas chôdze [min]	1,35	3,05	1,25	3,01
Počet prestupov	0,28	0,45	0,30	0,46
Interval medzi spojmi [min]	22,99	14,21	41,05	32,57

najvýznamnejším faktorom v rannej špičke je frekvencia spojov, zatiaľ čo v sedle má druhú najvyššiu váhu počet prestupov.

Väčšinu cestujúcich v rannej špičke tvoria zamestnaní ľudia a žiaci či študenti, ktorí cestujú do práce alebo do školy každý deň. Údaje za celý týždeň majú teda charakter nevyváženého panelu (obsahujú viac pozorovaní niektorých cestujúcich). Náhodnú (nepozorovateľnú) časť funkcie úžitku pre jednotlivé voľby urobené tým istým človekom nemožno považovať za nezávislú (pri výbere sa napr. prejavuje zvyk). Preto je vhodné použiť vzorkovanie, ktorým sa potlačí závislosť pre opakované pozorovania. Jednou z možných vzorkovacích metód je bootstrapová metóda.

Tabuľka 2. Odhad parametrov s normovanými hodnotami atribútov, 6:00 – 8:00

Číslo atribútu	Popis	Odhad parametra	Smerodajná odchýlka	*t-ratio*
1	Čas jazdy	-5.37	9.27E-2	-57,90
2	Čas chôdze	-2.64	4.97E-2	-53.14
3	Počet prestupov	-3.04	3.70E-2	-82.23
4	Interval medzi spojmi	-3.92	6.61E-2	-59.30
Súhrnná štatistika				
	Počet skupín cestujúcich $\|I\|$ =	513		
	Počet pozorovaní N =	24,722		
	$L(0)$ =	-18,862.51		
	$L(\hat{\boldsymbol{b}})$ =	-5,453.33		
	ρ^2 =	0.71		

Tabuľka 3. Odhad parametrov s normovanými hodnotami atribútov, 8:00 – 14:00

Číslo atribútu	Popis	Odhad parametra	Smerodajná odchýlka	*t-ratio*
1	Čas jazdy	-4.55	9.04E-2	-50.31
2	Čas chôdze	-3.54	7.15E-2	-49.45
3	Počet prestupov	-2.82	3.40E-2	-82.88
4	Interval medzi spojmi	-2.30	4.44E-2	-51.80
Súhrnná štatistika				
	Počet skupín cestujúcich $\|I\|$ =	598		
	Počet pozorovaní N =	23,808		
	$L(0)$ =	-18,692.74		
	$L(\hat{\boldsymbol{b}})$ =	-7,040.46		
	ρ^2 =	0.62		

Označme symbolom y_{it} pozorovanie *i*-teho človeka v čase *t* (v našom prípade pozorovaním je dopravné spojenie, ktoré *i*-ty človek v deň *t* použil). Vyvážený panel možno zapísať ako maticu

$$\boldsymbol{Y}(N,T)=\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & & & \\ y_{N1} & y_{N2} & \cdots & y_{NT} \end{pmatrix}$$

Vzorkovaním vo zvislom smere (výberom riadkov s opakovaním) vytvoríme z tejto matice novú maticu $\boldsymbol{Y}^*(N,T)$. Celkovo vytvoríme *K* vzoriek. Označme

$\hat{b}$ odhad parametra *b* (pre jednoduchosť bez indexu) zo základného súboru – matice ***Y***;

$\hat{b}(k)$ odhad parametra *b* z *k*-tej vzorky;

$\hat{b}_B$ bootstrapový odhad parametra *b*.

Potom presnejší odhad parametra *b* zo všetkých vzoriek vypočítame takto:

$$\hat{b}_B = K\hat{b} - \frac{K-1}{K}\sum_{k=1}^{K}\hat{b}(k) \tag{11}$$

Ak predpokladáme, že pravidelne cestujúci nastúpi aj vystúpi vždy na rovnakej zastávke, môžeme za riadky matice ***Y*** považovať OD dvojice zastávok (v súlade s logit modelom), a nie jednotlivých ľudí.

V tab. 4 je bootstrapový odhad parametrov modelu pre rannú dopravnú špičku. Vidíme, že v pôvodnom modeli boli podhodnotené atribúty, ktoré vieme z údajov o vykonaných jazdách presne určiť, a to čas jazdy vo vozidle, počet prestupov a interval medzi spojmi. Zároveň sú to atribúty, ktoré najviac ovplyvňujú dobu prepravy.

Ďalším poznatkom, ktorý vieme získať z odhadov parametrov, je ich vzájomný pomer, ktorý udáva prijateľnú výmenu (*trade-off*) medzi jednotlivými atribútmi (tab. 5 a 6). Pomer hovorí, aké množstvo jednej premennej by boli cestujúci ochotní vymeniť za jednotkové zlepšenie inej premennej. Napr. v rannej špičke by boli cestujúci ochotní cestovať o 33 minút dlhšie, keby nemuseli prestupovať. Inými slovami, zmena v systéme MHD, ktorá by eliminovala prestupy, ale nepredĺžila čas jazdy o viac ako 33 minút, by zvýšila atraktivitu systému. Pomer prestupu k ostatným atribútom je o jeden až dva rády vyšší než ostatné pomery v tabuľkách, čo potvrdzuje averziu cestujúcich voči prestupovaniu. V sedle je jeden prestup ekvivalentný približne 30 minútam jazdy vo vozidle a 18 minútam chôdze. V oboch sledovaných periódach sú cestujúci najviac ochotní obetovať interval medzi spojmi za zlepšenie ostatných atribútov.

Tabuľka 4. Pôvodný a bootstrapový odhad parametrov modelu pre rannú dopravnú špičku

	Pôvodný odhad	Bootstrapový odhad
Čas jazdy	-5.37	-6.33
Čas chôdze	-2.64	-2.15
Počet prestupov	-3.04	-3.61
Interval medzi spojmi	-3.92	-4.00

Tabuľka 5. Pomer parametrov pre absolútne hodnoty atribútov a bootstrapový odhad, 6:00 – 8:00

	Čas jazdy [min]	Čas chôdze [min]	Počet prestupov [-]	Interval medzi spojmi [min]
Čas jazdy [min]	N/A	0.64	0.03	1.61
Čas chôdze [min]	1.57	N/A	0.05	2.52
Počet prestupov [-]	33.23	21.19	N/A	53.36
Interval medzi spojmi [min]	0.62	0.40	0.02	N/A

Tabuľka 6. Pomer parametrov pre absolútne hodnoty atribútov, 8:00 – 14:00

	Čas jazdy [min]	Čas chôdze [min]	Počet prestupov [-]	Interval medzi spojmi [min]
Čas jazdy [min]	N/A	0.59	0.03	5.93
Čas chôdze [min]	1.70	N/A	0.06	10.06
Počet prestupov [-]	29.73	17.52	N/A	176.34
Interval medzi spojmi [min]	0.17	0.10	0.01	N/A

4 Záver

Článok popisuje multinomial logit model pre voľbu dopravného spojenia v mestskej hromadnej doprave. Ukázali sme, že model možno kalibrovať pomocou dát získaných z označovacích zariadení vo vozidlách MHD, čo je nepochybne lacnejší a presnejší spôsob než dotazníkový prieskum medzi cestujúcimi. V prípadovej štúdii sme odhadli parametre modelu pre mestskú hromadnú dopravu v Žiline, a to pre rannú špičku a dopoludňajšie sedlo. Najväčšiu úlohu pri výbere dopravného spojenia má v oboch prípadoch čas. Poradie dôležitosti ostatných atribútov sa v jednotlivých periódach líši. Druhým najvýznamnejším faktorom v rannej špičke je frekvencia liniek (resp. interval medzi spojmi), zatiaľ čo v sedle má druhú najvyššiu váhu počet prestupov.

Model môžeme použiť pre odhad správania cestujúcich pri zmene v sústave liniek, napr. pri zmene trás a frekvencií liniek alebo koordinovaní spojov na prestupných zastávkach. Výstupy modelu môžu slúžiť ako vstupné údaje pre optimalizáciu organizácie dopravy metódami matematického programovania a počítačovej simulácie.

Tento príspevok vznikol s podporou výskumných projektov VEGA 1/0361/10 „Optimálne navrhovanie verejných obslužných systémov v podmienkach neistoty“ a projektu APVV SK-CZ-0075-09 „Optimálne rozmiestňovanie obslužných stredísk pomocou IP-solvera“.

Literatúra

[1] BEN-AKIVA, M.E., BIERLAIRE, M. Discrete Choice Methods and Their Applications to Short Term Travel Decisions. In *Handbook of Transportation Science*, ed. Randolph W. Hall. Norwell : Kluwer Academic Publishers, 1999. pp. 5–34.

[2] BEN-AKIVA, M.E., LERMAN, S.R. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Ma, 1985.

[3] FAN, W., MACHEMEHL, R.B. *Optimal Transit Route Network Design Problem: Algorithms, Implementations, and Numerical Results*. Report No. SWUTC/04/167244-1. Center for Transportation Research, University of Texas at Austin, 2004.

[4] JÁNOŠÍKOVÁ, Ľ., BLATOŇ, M., TEICHMANN, D. Design of urban public transport lines as a multiple criteria optimisation problem. In *Proc. of the 16th International Conference on Urban Transport and the Environment – Urban Transport XVI*, eds. A. Pratelli & C.A. Brebbia. Southampton : WIT Press, 2010. pp. 137–146.

[5] KOPPELMAN, F.S., BHAT, C. *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*, 2006, online. http://www.civil.northwestern.edu/people/koppelman/PDFs/LM_Draft_060131Final-060630.pdf

[6] SLAVÍK, J. Discrete choice analysis. In *Studies of the Faculty of Management Science*, vol. 4, 1995. pp. 69–78.

[7] SUROVEC, P. *Provoz a ekonomika silniční dopravy I.* Ostrava : Vysoká škola báňská, 2000.

[8] TRAIN, K.E. *Discrete choice methods with simulation*. 2nd edition. Cambridge University Press, 2009.