

Statisztikai Szemle

Közzététel: 2020. november 10.

A tanulmány címe:

Diszkrét választási modellek becslése az R Apollo csomagjának használatával – multinomiális logit modell

Szerzők:

CZINE PÉTER, a Debreceni Egyetem PhD-hallgatója
E-mail: czine.peter@econ.unideb.hu

HARANGI-RÁKOS MÓNIKA, a Debreceni Egyetem egyetemi docense
E-mail: rakos.monika@econ.unideb.hu

BALOGH PÉTER, a Debreceni Egyetem egyetemi tanára
E-mail: balogh.peter@econ.unideb.hu

DOI: <https://doi.org/10.20311/stat2020.11.hu1310>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) Statisztikai Szemle c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Sztj.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Sztj. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c.) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:

„*Forrás: Statisztikai Szemle c. folyóirat 98. évfolyam 11. számában megjelent, **Czine Péter, Harangi-Rákos Mónika és Balogh Péter** által írt, 'Diszkrét választási modellek becslése az R Apollo csomagjának használatával – multinomiális logit modell' című tanulmány (link csatolása)*”

7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem esnek szükségképpen egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

Czine Péter – Harangi-Rákos Mónika – Balogh Péter

Diszkrét választási modellek becslése az R Apollo csomagjának használatával – multinomiális logit modell*

Estimating discrete choice models by using R Apollo package – multinomial logit model

CZINE PÉTER, a Debreceni Egyetem PhD-hallgatója
E-mail: czine.peter@econ.unideb.hu

HARANGI-RÁKOS MÓNIKA, a Debreceni Egyetem
egyetemi docense
E-mail: rakos.monika@econ.unideb.hu

BALOGH PÉTER, a Debreceni Egyetem
egyetemi tanára
E-mail: balogh.peter@econ.unideb.hu

A cikk célja, hogy bemutassa egy preferenciaértékelő módszer (diszkrét választási kísérlet [discrete choice experiments, DCE]) egyik leggyakrabban alkalmazott modelljének (multinomiális logit, MNL) becslési folyamatát és eredményeinek értelmezését az R Apollo csomag alkalmazásán keresztül. A szerzők egy mindenki számára hozzáférhető adatbázist használnak a tanulmányban példaként, amely a különböző utazási alternatívákra vonatkozó preferenciákat vizsgálja. Az adatbázis részletes bemutatását követően modelljük felépítésének lépéseit ismertetik, és végül az eredményeket értelmezik.

TÁRGYSZÓ: diszkrét választási kísérlet, multinomiális logit modell, az R program Apollo csomagja

The purpose of this study is to present the estimation process and interpretation of the results by one of the most commonly used model (multinomial logit, MNL) of a preference evaluation method (discrete choice experiment, DCE) through the use of the R Apollo software package. The study uses a publicly accessible database as an example that examines preferences for different travel alternatives. After a detailed presentation of the database, the steps of the model-building process are carried out and the results are interpreted, thus providing guidance to future users of the MNL model through the use of this R package.

KEYWORD: discrete choice experiment, multinomial logit model, R Apollo package

* A tanulmány az Innovációs és Technológiai Minisztérium ÚNKP-20-3 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

A fogyasztók preferenciáinak vizsgálata számos kutatási terület érdekes témája. A kérdések túlnyomó többsége arra irányul, hogy milyen szabályszerűség áll a döntéshozatal mögött, illetve hogyan tudnánk minél hatékonyabb előrejelző becsléseket adni a folyamatot jellemző magas szintű variabilitás mellett (*Hensher–Rose–Greene* [2005]).

Az elmúlt évek során számos, ún. „preferenciaértékelő eljárás” terjedt el, amelyek a már említett kérdésekre kísérelnek meg válaszokat adni. Ezeknek két fő irányvonala különíthető el: a feltárt preferencia (stated preference, SP) és a kinyilvánított preferencia (revealed preference, RP) módszerei. A kinyilvánított eljárások során valós piaci helyzetekben figyelhető meg az egyének viselkedése, míg a feltárt módszerek esetében egy hipotetikus helyzet elemzése zajlik annak érdekében, hogy a piacon jelenleg még el nem érhető alternatívák¹ is értékelhetők legyenek (*Train* [2003]). Vannak azonban olyan módszerek, amelyek mindkét megközelítésből származó adatok elemzésére alkalmasak. Idesorolható a DCE, amely az egyének hasznosságmaximalizáló viselkedését feltételezi, a termékek/szolgáltatások hasznosságát azok tulajdonságaiból eredőnek tekinti, diszkrét választási helyzetet feltételez, és a fogyasztói hasznosságot egy szisztematikus (a termék/szolgáltatás tulajdonságaiból ered, a kutató által megfigyelhető) és egy véletlen (a kutató által nem megfigyelhető) részre bontja fel (*Hensher–Rose–Greene* [2015], *Lancaster* [1966]).

A DCE folyamata többnyire szakirodalmi tájékozódással, előzetes kérdőíves felméréssel, esetleg valamilyen kvalitatív eljárással (legtöbbször fókuszcsoportos interjúval) indul, melynek célja, hogy megtudjuk, mely tulajdonságok súlya a legnagyobb a fogyasztói döntések során a vizsgálni kívánt termékkel/szolgáltatással kapcsolatban. Miután meghatározzuk a tulajdonságokat és azok szintjeit, szoftveres támogatással (például Ngene 1.2 program) összeállítjuk a döntési kártyákat (a tulajdonságszintekből kombinált alternatívákat) és a választási helyzeteket (annak érdekében, hogy a döntési helyzetek a kísérleti elrendezéssel kapcsolatosan elvárt minden követelménynek megfeleljenek) (*ChoiceMetrics* [2018]). A kísérletből nyert adatok feldolgozása számos modellspecifikáción keresztül történhet. A modellek becsléséhez már több szoftver (például Nlogit, STATA, R Apollo) is rendelkezésre áll. Ezek mindegyikének megvan a maga sajátos tulajdonsága, előnye és hátránya. Az R program Apollo nevű csomagja számos diszkrét választási modell alkalmazását teszi elérhetővé a felhasználók számára (a véletlen hasznosság elméletén alapuló modelle-

¹ Az „alternatíva” a továbbiakban a nemzetközi szakirodalomból átvett „alternative” kifejezéssel egyezik meg.

ken túl például a nem véletlen hasznosságon alapulót [random regret minimisation]) (*Hess–Daly*, [2014], *Hess–Palma* [2019a], *Hess–Palma* [2019b]). A modellbecslésekből származó eredmények számos, a gyakorlatban is hasznosítható információt hordoznak magukban. Elemzésükön keresztül többek között megtudhatjuk, hogy a döntéshozók (a vizsgált termékkel/szolgáltatással kapcsolatosan) mely tulajdonságok meglétét ítélik a legfontosabbnak (egy ételmiszer esetében a márkajelzés, egy utazási lehetőségre vonatkozóan az utazási idő, illetve egy egészségügyi kezelésnél annak kockázata), ezek közül melyek hatnak pozitívan (például a termék márkajelzése), illetve negatívan (az utazási idő) a hasznosságérzetükre. Így a következő kérdések megválaszolására nyílik lehetőség: hogyan tervezzenek új termékeket/szolgáltatásokat, milyen változtatásokat hajtsanak végre a már piacon levők tulajdonságstruktúráiban, vagy milyen árazási döntéseket hozzanak.

A *McFadden* nevéhez köthető MNL specifikáció, a DCE alapmodelljének tekinthető. Előnyös tulajdonságai közé tartozik, hogy viszonylag könnyedén becsülhető, és eredményeinek interpretálása sem hordoz magában annyi problematikát, mint a komplexebb modellek esetében. Említést kell viszont tenni arról, hogy több hátránya is van, melyek közül a homogén preferenciák és az irreleváns alternatívák függetlenségének feltételezését szokás említeni elsődlegesen (*McFadden* [1973]).

Jelen írás célja, hogy az említettek tükrében bemutassa az MNL-modell becslésének folyamatát és eredményeinek értelmezését az R Apollo csomag alkalmazásán keresztül. Ehhez egy mintaadatbázist (`apollo_modeChoiceData`) használunk fel, amely szabadon hozzáférhető az Apollo csomag weboldalán (<http://www.apollochoicemodelling.com/examples.html>).

1. Multinomiális logit modell becslése és eredményeinek értelmezése

A következő fejezetben a modellbecsléshez használt adatbázist és a becslés folyamatának lépéseit mutatjuk be, valamint értelmezzük az eredményeket.

1.1. Az adatbázis

A mintaként szolgáló adatbázis 500 fő utazási² alternatívákra vonatkozó preferenciáit vizsgálja, melynek első oszlopában a kitöltők egyedi azonosítói

² Fontos említést tenni arról, hogy a DCE meglehetősen sok területen alkalmazható, használatán keresztül a termékek/szolgáltatások széles köre értékelhető.

(identification, ID) található, ezek a döntési helyzetek számához igazodóan ismétlődnek.³ Ezt követően a döntési szituációk típusáról szerezhetünk információt (RP vagy SP), majd azt láthatjuk, hogy mely lehetőségek érhetők el a különböző választási helyzetek során. Az adatbázisban szereplő első választási helyzetet az 1. táblázat szemlélteti.

1. táblázat

Az adatbázis első választási helyzetének bemutatása
(First choice situation in the database)

Döntést befolyásoló tulajdonság	Air (repülő)	Rail (vonat)
Time (mins) (utazási idő percben kifejezve)	50	140
Cost (£) (utazási költség fontban kifejezve)	80	55
Access (mins) (hozzáférhetőségi idő percben kifejezve)	55	5
Service (szolgáltatás)	–	–
Melyik utazási módot választaná? (X)		

Az adatbázis további oszlopai (time_car [utazási idő az autó utazási lehetőség-nél] – service_rail [szolgáltatás a vonat utazási lehetőség esetén]) az egyes utazási lehetőségek jellemzőit tartalmazzák a döntési helyzetekre vonatkozóan. Ezután a válaszadók különböző karakterisztikáit (esetspecifikus változók⁴) tartalmazó oszlopok következnek, melyek között jelen példa esetében a kitöltő neve, utazói típusa (üzleti utazó, vagy sem) és jövedelme szerepel. Végül pedig a döntési szituációkban választott lehetőség jelölését találjuk. Az adatbázis további részleteit az 2. táblázat tartalmazza.

³ Jelen példa esetében minden válaszadó 16 darab döntési helyzetben választott, így az összes megfigyelés száma $500 \times 16 = 8\,000$ darab.

⁴ Minden döntési helyzetben ugyanazt az értéket veszik fel egy adott válaszadóra vonatkozóan.

2. táblázat

Az adatbázis változóinak bemutatása
(Variables in the database)

Változó	Érték
<i>ID</i> (egyedi azonosító)	1–500
<i>RP</i> (kinyilvánított preferencia)	1, ha a döntési helyzet RP 0, ha a döntési helyzet SP
<i>av_car/av_bus/av_air/av_rail</i> (av: availability) (az autó/busz/repülő/vonat utazási lehetőségek elérhetősége)	1, ha elérhető a car/bus/air/rail alternatíva a döntési helyzetben 0, ha nem érhető el a car/bus/air/rail alternatíva a döntési helyzetben
<i>time_car/time_bus/time_air/time_rail</i> (utazási idő az autó/busz/repülő/vonat utazási lehetőségre vonatkozóan) <i>cost_car/cost_bus/cost_air/cost_rail</i> (utazási költség az autó/busz/repülő/vonat utazási lehetőségre vonatkozóan)	A car/bus/air/rail alternatívára vonatkozó utazási idő és költség értéke a döntési helyzetben
<i>access_bus/access_air/access_rail</i> (hozzáférhetőségi idő a busz/repülő/vonat utazási lehetőségre vonatkozóan)	A bus/air/rail alternatívára vonatkozó hozzáférhetőségi idő a döntési helyzetben
<i>service_air/service_rail</i> (szolgáltatás a repülő/vonat utazási lehetőségre vonatkozóan)	Az air/rail alternatívára vonatkozóan elérhető szolgáltatás a döntési helyzetben 0, ha nincs elérhető szolgáltatás 1, ha fapados 2, ha van internetelérhetőség 3, ha van étkezési lehetőség
<i>female</i> (női válaszadó)	1, ha nő a válaszadó 0, ha nem nő a válaszadó
<i>business</i> (üzleti utazó)	1, ha üzleti utazó a válaszadó 0, ha nem üzleti utazó a válaszadó
<i>income</i> (jövedelem)	A válaszadó jövedelme
<i>choice</i> (választás)	1, ha autó 2, ha busz 3, ha repülő 4, ha vonat a kiválasztott utazási lehetőség

1.2. A becslés folyamata

Alapbeállítások és az adatok beolvasása

Az Apollo bővítmény telepítéséhez mindössze az `install.packages("apollo")` parancs futtatása szükséges (internetelés mellett), az R program környezetében.

A modellek parancsora mindig az alapbeállításokkal indul. (Lásd az 1. ábrát.) Idesorolható a memória törlése, az Apollo csomag könyvtárának betöltése, a modell nevének meghatározása (`modelName`), részletes leírása (`ModelDescr`) és az egyedi azonosítókat tartalmazó változó definiálása (`indivID`).

1. ábra. A modell alapbeállításai
(Basic model settings)

```
rm(list=ls () ) # A memória törlése
library(apollo) # Az Apollo csomag könyvtárának betöltése
apollo_initialise()
apollo_control <- list(
  modelName = "Apollo_example_1", # A modell neve
  ModelDescr = "Simple MNL model on mode choice RP data", # A modell leírása
  indivID = "ID" # Az egyedi azonosítókat (ID) tartalmazó változó meghatározása
)
```

A következő lépésben a használni kívánt adatbázis beolvasása következik (`read.csv()`), majd leválogathatjuk a szükséges sorokat (szűrés) és új változókat is létrehozhatunk az adatbázisban (szükséges említést tenni arról, hogy a beolvasni kívánt adatbázis formátuma nem csak csv kiterjesztésű lehet). Erre a 2. ábra mutat példát, ahol mindössze az RP-jellegű adatokat vettük számításba, és egy átlagos jövedelemváltozót is létrehoztunk.

2. ábra. Az adatok beolvasása
(Reading the data)

```
database <- read.csv("apollo_modeChoiceData.csv",header=TRUE) # Az adatbázis
# beolvasása
database <- subset(database, database$RP==1) # RP sorok kiválasztása
database$mean_income <- mean(database$income) # Új változó létrehozása (átlagos
# jövedelem)
```

A választások elemzése

A választások elemzése kiegészítő funkció használatán keresztül megtudhatjuk, hogy van-e szignifikáns eltérés⁵ az egyes alternatívákat választók és nem választók csoportjai között, bizonyos változók tekintetében (fontos kiemelni, hogy a példában szemléltetettekén túl, számos, a kutató számára érdekelt további változó is beemelhető a vizsgálatba). A funkció használatához meg kell adnunk az alternatíváinkat (**alternatives**), elérhetőségüket (**avail**) a döntési helyzetekben és a választást jelölő (**choiceVar**), valamint a magyarázó (**explanators**) változókat. (Lásd a 3. ábrát.)

3. ábra. A választások elemzése
(Analysis of choices)

```
choiceAnalysis_settings <- list(
  alternatives = c(car=1, bus=2, air=3, rail=4), # A választások elemzése
  avail = list(car=database$av_car, bus=database$av_bus, air=database$av_air,
    rail=database$av_rail), # Alternatívák meghatározása
  # Az adott alternatíva elérhetőségét
  jelölő változó meghatározása
  choiceVar = database$choice, # A választást jelölő változó meghatározása
  explanators = database[,c("female","business","income")] # A magyarázó változó/k
  meghatározása
)
apollo_choiceAnalysis(choiceAnalysis_settings, apollo_control, database)
```

A modell paramétereinek meghatározása

A következő lépésben először a modell paramétereinek nevét, majd becslésük kezdeti értékét⁶ határozzuk meg. Az **apollo_fixed()** nevezetű függvény használatával kijelölhetjük azon paramétereket, melyeket a kezdeti értékükön szeretnénk rögzíteni a becslési folyamat során (fontos viszont említést tenni arról, hogy ebben az esetben is szükséges a paraméter feltüntetése az **apollo_beta()** felsorolásában). Amennyiben már rendelkezésünkre áll egy korábbi, a mintánkra vonatkozó (például MNL-) modellbecslés, annak értékei beolvashatók (**apollo_readBeta()**)⁷ azzal a céllal, hogy a log-likelihood becslés hamarabb megoldást találjon/elérje a megfelelő konvergálási pontot. (Lásd a 4. ábrát.)

A paraméterek három típusát különböztethetjük meg a modellbecslések során. *Generikusnak*/általánosnak nevezzük azokat a paramétereket, melyek minden egyes

⁵ A választások elemzése funkció független kétmintás *t*-próbát használ ahhoz, hogy megállapítsa, van-e szignifikáns eltérés az adott alternatívát választó és nem választó csoportok között.

⁶ MNL-modell esetén ez legtöbbször 0 lesz.

⁷ A 4. ábrán látható példa esetében szereplő **overwriteFixed=FALSE** kiegészítés arra utal, hogy a kezdeti értéken rögzített paramétereket nem módosítjuk a beolvasott, korábbi becslés szerint.

alternatívára vonatkozóan ugyanazt az értéket veszik fel. *Alternatíváspecifikusként* definiálhatjuk őket akkor, ha mindössze egy alternatívára vonatkozóan jelennek meg. Végül pedig elkülöníthetők még az ún. *alternatíváspecifikus konstans* paraméterek, melyek az adott alternatíva hasznosságának azt a részét reprezentálják, amely nem annak vizsgált tulajdonságaiból ered.

4. ábra. A paraméterek meghatározása
(Definition of parameters)

```
apollo_beta <- c(asc_car = 0, # A modell paramétereinek definiálása
                b_tt_car = 0,
                b_access = 0,
                b_cost = 0)

apollo_fixed <- c("asc_car") # A kezdeti értékükön rögzített paraméterek
                 megadása

apollo_beta <- apollo_readBeta(apollo_beta, apollo_fixed, "Apollo_example_0",
                               overwriteFixed=FALSE) # Korábbi becslés paramétereinek beolvasása
```

Megjegyzés. Az *asc* megnevezés a különböző utazási alternatívákra (car, bus, air, rail) vonatkozó alternatíváspecifikus konstans paramétereket, *b_tt* (car, bus, air, rail) a különböző utazási alternatívákra vonatkozó alternatíváspecifikus paramétereket, *b_access* és *b_cost* pedig a generikus paramétereket jelölik. Az *asc_car* alternatíváspecifikus konstans paraméter a bázisszintet képviselte a becslések során, így kezdeti értékén (nullán) rögzítettük.

A hasznosságfüggvény formája

A modellalkotás következő lényeges pontja a hasznosságfüggvény megfelelő felépítése és formájának megválasztása. Mindenekelőtt fontos kitérni arra, hogy a diszkrét választási kísérlet gyakorlatában az egyének hasznossága egy megfigyelhető/szisztematikus és egy nem megfigyelhető/véletlen részből tevődik össze. (Lásd az /1/ egyenletet.) Előbbi pedig a karakterisztikaelméletnek megfelelően a termékek/szolgáltatások bizonyos tulajdonságaiból (azok szintjeiből) ered. (Lásd a /2/ egyenletet.) Meg kell még említenünk, hogy a hasznosságfüggvények formája leggyakrabban az ún. linearitás a paraméterekben és attribútumokban (linear in the parameters, linear in the attributes, LPLA) elven alapul (Lancaster [1966], McFadden [1973], Hensher–Rose–Greene [2005]).

$$U_{n,i} = V_{n,i} + \varepsilon_{n,i}, \quad /1/$$

$$V_{n,i} = \sum_{k=1}^K \beta_k X_{n,i,k}, \quad /2/$$

ahol, n a válaszadót, i az alternatívát, $V_{n,i}$ a szisztematikus részt, $\varepsilon_{n,i}$ a nem megfigyelhető részt, k a tulajdonságot, β a k tulajdonsághoz tartozó együttható értékét és X a megfigyelt változót fejezi ki.

Az előzetes céloknak megfelelően irányulhat a becslés a hasznossági együtthatókra (preference-space estimation – preferenciatér-becslés) (lásd az 5. ábrát) vagy az egyes tulajdonságszintekre vonatkozó fizetési hajlandóság közvetlen módon történő számszerűsítésére (willingness to pay-space estimation – fizetési hajlandóság-tér-becslés) (lásd a 6. ábrát).

5. ábra. Hasznosságfüggvények a preferenciatérben
(Utility functions in the preference space)

```
V[['car']] = asc_car + b_tt_car * time_car + b_cost * cost_car
# A hasznossági függvények meghatározása
V[['bus']] = asc_bus + b_tt_bus * time_bus + b_access * access_bus + b_cost * cost_bus
```

Az 5. ábra alapján jól látható, hogy a preferenciatér hasznossági függvényeinek részei között a korábbiak során meghatározott paraméterek (ezek között jelen példában szerepel alternatíváspecifikus konstans [például **asc_car**], alternatíváspecifikus [például **b_tt_car**] és generikus [például **b_cost**] paraméter is) és az adatbázis változói szerepelnek.

6. ábra. Hasznosságfüggvények a fizetési hajlandóságtérben
(Utility functions in the willingness-to-pay space)

```
V[['car']] = asc_car + b_cost * (b_tt_car * time_car + cost_car)
V[['bus']] = asc_bus + b_cost * (b_tt_bus * time_bus + b_access * access_bus + cost_bus)
```

A 6. ábra hasznosságfüggvényének formája mindössze abban módosult, hogy az árra vonatkozó paramétert kiemeltük, így a különböző tulajdonságszintekre (például **b_tt_car**, **b_tt_bus**, **b_access**) becsült együtthatók már a fizetési hajlandóságot reprezentálják.

Fontos kitérni arra, hogy amennyiben preferenciatérben szeretnénk becsülni, viszont szükségünk van a fizetési hajlandóság értékekre is, nem szükséges transzformálnunk a hasznosságfüggvényt és új becslést végeznünk. Erre az esetre az ún. „**deltaMethod()**” függvény nyújt kiváló alternatívát. (Lásd a 7. ábrát.)

7. ábra. A *deltaMethod* funkció
(DeltaMethod function)

```
deltaMethod_settings<-list(operation="ratio", parName1="b_tt_car",
  parName2="b_cost") # A „deltaMethod” alkalmazása
apollo_deltaMethod(model, deltaMethod_settings)
```

A 7. ábra példájában arra a kérdésre szeretnénk választ kapni, hogy a car utazási lehetőségre vonatkozóan a travel time növekedése/csökkenése hogyan hat a válaszadók fizetési hajlandóságára (például mekkora összeggel fizetnének többet a válaszadók abban az esetben, ha az utazási idő 5 perccel csökkenne).

1.3. Az eredmények értelmezése

A választások elemzése funkció eredményei

A választások elemzése funkció futtatásának eredményei egy különálló, csv kiterjesztésű fájlba (choice analysis) mentődnek. (Lásd a 3. táblázatot.)

3. táblázat

A választások elemzése funkció eredményei
(Results of the analysis of choices function)

Utazási alternatíva megnevezése	Mean for income if chosen (Átlagos jövedelem, amennyiben választja)	Mean for income if not chosen (Átlagos jövedelem, amennyiben nem választja)	<i>t</i> -test for difference (<i>t</i> -próba az eltérés vizsgálatára)
	font/év		
Car	53 317,87	51 278,14	-1,87
Bus	47 824,65	52 430,30	3,18*
Air	52 568,40	50 945,80	-1,37
Rail	51 452,97	51 948,94	0,47

* A szignifikáns értéket jelöli: $p < 0,05$.

Az eredmények egy-egy sort (az utazási alternatívákra vonatkozóan) és három oszlopot tartalmaznak. Az első oszlop az adott lehetőséget választó, míg a második a nem választó csoportra vonatkozóan adja meg az elemzés alá vont változó (jelen esetben jövedelem) átlagos értékét. A harmadik oszlopban a független kétmintás *t*-próbára vonatkozó tesztstatisztika-érték szerepel. A 3. táblázat példája alapján jól látható, hogy 5 százalékos szint mellett mindössze a bus utazási alternatívát választók és nem választók csoportjai között figyelhető meg szignifikáns eltérés a jövedelemre vonatkozóan.

A multinomiális logit modell becslésének eredményei

A modellbecslés futtatását követően először a teljes modellünkre vonatkozóan kapunk áttekintő adatokat (lásd a 4. táblázatot): többek között a neve, részletes leírása, becslésének időpontja, a mintaelemszám, a megfigyelések száma, a modellünk illeszkedéséhez közvetlenül és közvetetten kapcsolódó mutatók (a log-likelihood funkció kezdeti értéke [LL(start)], a csak konstans értékeket tartalmazó modell log-likelihood értéke [LL(0)], a becsült modellünk végső log-likelihood értéke [LL(final)], az r^2 [rho-square], a korrigált r^2 [adjusted rho-square], Akaike-féle információs kritérium [Akaike information criterion, AIC] és a Bayesi információs kritérium [Bayesian information criterion, BIC], a becsült paraméterek száma, a modellünk futtatásának időtartama, valamint a log-likelihood funkció iterációinak száma) szerepel.

4. táblázat

A modell áttekintő adatai
(Overview data of the model)

Model name (A modell neve)	<i>Apollo_example_1</i> (Apollo_pelda_1)
Model description (A modell leírása)	Simple MNL model on mode choice RP data (Egyszerű MNL-modell az utazási választás RP-adataira vonatkozóan)
Model run at (Becslés időpontja)	2020-09-04 14:54:35
Number of individuals (Egyének száma)	500
Number of observations (Megfigyelések száma)	1 000
LL(start) (LL kezdeti érték)	-1 170,86
LL(0) (LL 0 érték)	-1 170,86
LL(final) (LL végső érték)	-1 025,756
Rho-square (R^2 -érték)	0,1239
Adjusted rho-square (Korrigált R^2 -érték)	0,1162
AIC (Akaike-féle információs kritérium)	2 069,51
BIC (Bayesi információs kritérium)	2 113,68
Estimated parameters (Becsült paraméterek)	9
Time taken (Időtartam)	00:00:5.57
Iterations (Iterációk)	15

Ezután a paraméterbecslések eredményei következnek. A sorok a különböző paraméterek megnevezéseit, míg az oszlopok azok hasznossági együtthatóit, standard hibáit, t -értékeit és utóbbi kettő robusztus formáit tartalmazzák. (Lásd az 5. táblázatot.) Szükséges említést tenni arról, hogy amennyiben a paraméterekre vonatkozó p -értékeket is látni szeretnénk, az egyszerű **apollo_modelOutput (model)**

(az eredmények kiírása az R program környezetében) és az `apollo_saveOutput` (`model`) (az eredmények mentése) parancsokat a 8. ábra szerint szükséges kiegészítenünk.

8. ábra. Az eredmények kiírása és mentése p -értékekkel együtt
(Printing and saving the results with p -values)

```
modelOutput_settings <- list() # Az eredmények kiírása
modelOutput_settings$printPVal=TRUE
apollo_modelOutput(model,modelOutput_settings)

saveOutput_settings <- list() # Az eredmények mentése
saveOutput_settings$printPVal=TRUE
apollo_saveOutput(model,modelOutput_settings)
```

5. táblázat

A paraméterbecslések eredményei
(Results of the parameter estimates)

Becsült paraméter	Estimate (Becsült együttható)	Standard error (Standard hiba)	t -ratio (t -érték)	Robust standard error (Robusztus standard hiba)	Robust t -ratio (Robusztus t -érték)
asc_car	0,0000	n. a.	n. a.	n. a.	n. a.
asc_bus	0,4747	1,0187	0,47	0,9896	0,48
asc_air	1,6291*	0,8274	1,97	0,8167	1,99
asc_rail	0,9445	0,8036	1,18	0,7944	1,19
b_tt_car	-0,0036*	0,0016	-2,35	0,0016	-2,33
b_tt_bus	-0,0088*	0,0026	-3,37	0,0026	-3,38
b_tt_air	-0,0207*	0,0066	-3,13	0,0064	-3,22
b_tt_rail	-0,0112*	0,0044	-2,56	0,0045	-2,52
b_access	-0,0115	0,0064	-1,78	0,0063	-1,82
b_cost	-0,0339*	0,0033	-10,30	0,0032	-10,69

* A szignifikáns értéket jelöli: $p < 0,05$.

Következtetéseink az 5. táblázat eredményei alapján:

– a car utazási alternatívához viszonyítva (amely a bázisszintet képviselte a becslések alatt, így kezdeti értékén [nullán] rögzítettük) minden más alternatíva jobban preferált volt a választások során (ez a következtetés az asc_bus, asc_air és asc_rail paraméterek pozitív együtthatói alapján vonható le), viszont ezen következtetések közül mindössze az asc_air [asc_repülő] tekinthető 5 százalékos szint mellett szignifikánsnak;

– a travel time növekedésével egyidejűleg szignifikánsan csökken a fogyasztói hasznosságérzet minden utazási lehetőségre vonatkozóan;

– az access (amely generikus paraméter és a *car* alternatíva esetében nem szerepel) növekedésével egyidejűleg csökken a fogyasztói hasznosságérzet, viszont ez a következtetés már nem tekinthető 5 százalékos szint mellett szignifikánsnak;

– a cost növekedésével (amely a hozzáférhetőségi időhöz hasonlóan generikus paraméterként szerepel) egyidejűleg szignifikánsan csökken a fogyasztói hasznosságérzet.

Végül pedig egy áttekintő táblázatot kapunk a választásokra vonatkozóan. (Lásd a 6. táblázatot.)

6. táblázat

A választások jellemzése
(Characterization of choices)

Szempont	Car	Bus	Air	Rail
Összesen elérhető (db)	778	902	752	874
Kiválasztott (db)	332	126	215	327
Kiválasztott az összes RP-megfigyelésre vonatkoztatva (%)	33,20	12,60	21,50	32,70
Kiválasztott az elérhető helyzetekre vonatkoztatva (%)	42,67	13,97	28,59	37,41

A 6. táblázatból látható, hogy az első sorban a döntési helyzetekben elérhető különböző alternatívák száma szerepel, míg a másodikban az, hogy az egyes lehetőségeket hányszor választották. A harmadik sorban az összes megfigyelés számához viszonyítva láthatjuk az egyes alternatívák választásának százalékos megoszlását (például a *car* esetében: $332/1000 * 100 = 33,20$ százalék, ahol az 1000-es osztó az egyének [500 fő] és az RP-döntési helyzetek/válaszadó [2 darab] szorzatából adódik). Az utolsó sorban pedig az elérhető helyzetek számához viszonyítva kapunk információt az egyes lehetőségek választásának arányáról (például a *car* esetében: $332/778 * 100 = 42,67$ százalék).

2. Összegzés

A tanulmányban az egyik leggyakrabban alkalmazott diszkrét választási modellet, a multinomiális logit becslést mutattuk be az R Apollo csomag alkalmazásán keresztül.

Ismertettük a példaként felhasznált adatbázist, amely bárki számára szabadon elérhető az Apollo csomag weboldalán. A becslés folyamata során kiemeltük az alapbeállításokat és az adatbázis beolvasását, majd részletes áttekintést adtunk a paraméterekről. Ezt követte a hasznosságfüggvények formáinak és részeinek bemutatása és a modellbecslésből kapott eredmények részletezése: az alapadatok, a paraméterek becslései és a választások áttekintése.

A jövőbeni felhasználók számára mellékletként elérhetővé tettük a becslési folyamat elvégzéséhez szükséges R parancssort és a futtatásból származó eredményeket (<http://search.ksh.hu/#/year/2020?c=s#11>).

Irodalom

- CHOICEMETRICS [2018]: *Ngene 1.2 User Manual & Reference Guide*. ChoiceMetrics Pty Ltd. Sydney. <http://www.choice-metrics.com/>
- HENSHER, D. A. – ROSE, J. M. – GREENE, W. H. [2005]: *Applied Choice Analysis*. Cambridge University Press. Cambridge. <http://dx.doi.org/10.1017/CBO9780511610356>
- HENSHER, D. A. – ROSE, J. M. – GREENE, W. H. [2015]: *Applied Choice Analysis*. Cambridge University Press. Cambridge. <http://dx.doi.org/10.1017/CBO9781316136232>
- HESS, S. – DALY, A. [2014]: *Handbook of Choice Modelling*. Edward Elgar Publishing. Cheltenham. <http://dx.doi.org/10.4337/9781781003152>
- HESS, S. – PALMA, D. [2019a]: Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*. Vol. 32. September. pp. 1–26. <http://dx.doi.org/10.1016/j.jocm.2019.100170>
- HESS, S. – PALMA, D. [2019b]: *Apollo version 0.0.6, user manual*. www.ApolloChoiceModelling.com
- LANCASTER, K. J. [1966]: A new approach to consumer theory. *Journal of Political Economy*. Vol. 74. No. 2. pp. 132–157. <https://doi.org/10.1086/259131>.
- MCFADDEN, D. [1973]: Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (ed.): *Frontiers in Econometrics*. Academic Press. New York. pp. 105–142.
- TRAIN, K. E. [2003]: *Discrete Choice Methods with Simulation*. Cambridge University Press. Cambridge. <http://dx.doi.org/10.1017/CBO9780511753930>