

Közzététel: 2021. május 12.

A tanulmány címe:

Diszkrét választási modellek becslése az R Apollo csomagjának használatával – látens osztályú modell

Szerzők:

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

DAJNOKI KRISZTINA, a Debreceni Egyetem egyetemi docense
E-mail: dajnoki.krisztina@econ.unideb.hu

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

DOI: <https://doi.org/10.20311/stat2021.5.hu0469>

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 részeit 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 99. évfolyam 5. számában megjelent, Czine Péter, Dajnoki Krisztina, Balogh Péter által írt, 'Diszkrét választási modellek becslése az R Apollo csomagjának használatával – látens osztályú 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 – Dajnoki Krisztina – Balogh Péter

Diszkrét választási modellek becslése az R Apollo csomagjának használatával – látens osztályú modell*

Estimating discrete choice models by using R Apollo package – The latent class model

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

DAJNOKI KRISZTINA, a Debreceni Egyetem
egyetemi docense
E-mail: dajnoki.krisztina@econ.unideb.hu

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

A tanulmány célja, hogy bemutassa a látens osztályú (latent class, LC) modell becslésének folyamatát és eredményeinek értelmezését az R Apollo csomag használatán keresztül. A szerzők egy mindenki számára elérhető adatbázis példáján keresztül prezentálják a modell felépítésének kritikus lépéseit és az eredmények értelmezését. Emellett kitérnek és ajánlásokat fogalmaznak meg a specifikációt érintő néhány kritikus kérdés vonatkozásában.

TÁRGYSZÓ: diszkrét választási kísérlet, látens osztályú modell, R Apollo csomag

The purpose of this study is to present the estimation process of the latent class (LC) model and the interpretation of its results, using the R Apollo software package. The critical steps of model building and the results are introduced through the example of a freely accessible database. In addition, some critical issues related to the specification are addressed and recommendations are made by the authors.

KEYWORD: discrete choice experiment, latent class 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 preferenciaértékelő eljárások közé sorolható diszkrét választási kísérlet (discrete choice experiment, DCE) számos kutatási terület gyakorta alkalmazott módszerévé vált az elmúlt évtizedek során, ilyenek például a marketing (*Wang–Ge–Ma* [2018]), a közlekedés- (*Bergantino–Capurso–Hess* [2020]), a környezet- (*Costa–Montemurro–Giuliani* [2019]) és az egészséggazdaságtan (*Liao–Ng–Cowling* [2020]). Ez egyrészt annak köszönhető, hogy a DCE mind a kinyilvánított (revealed preference, RP), mind pedig a feltárt preferencia (stated preference, SP) jellegű adatok kezelésére képes, akár szimultán módon is. Az előbbi megközelítés a fogyasztói viselkedés valódi piaci szituációban, míg utóbbi annak egy hipotetikus – a kutató által megtervezett – helyzetben történő elemzését célozza (*Train* [2009], *Hensher–Rose–Greene* [2015]). Másrészt pedig annak tulajdonítható, hogy a folyamatos fejlesztések hatására manapság már számos módszertani korlátozás hatékony kezelésére képes (*Hess–Daly* [2014]).

A diszkrét választási kísérletek elemzéséhez még jelenleg is gyakorta használt modelltípus a *McFadden* [1974] nevéhez köthető multinomiális logit (multinomial logit, MNL). A specifikációnak azonban – amellet, hogy számos előnyös tulajdonsággal rendelkezik – több korlátozása is van: az egyik legszélesebb körben ismert és több megközelítésből is kezelni próbált a homogén preferenciák feltételezése.

Az ízlésekben levő heterogenitást az elemzők többnyire folytonos vagy diszkrét eloszlások alkalmazásán keresztül kísérlik megragadni. Előbbi esetben a vizsgált tulajdonságok paramétervektoraira vonatkozóan lehetővé teszik, hogy azok egy folytonos eloszlás szerint változzanak, és annak bizonyos momentumait becsülik. Utóbbinál egymástól eltérő ízlésparaméterekkel rendelkező diszkrét számú osztályt különítenek el, melyeken belül a tagok preferenciái már homogének. Ez utóbbi megközelítés LC-modellezésként vált ismertté (*Louvière–Hensher–Swait* [2000], *Greene–Hensher* [2003]).

A tanulmány célja, hogy *Czine–Harangi–Rákos–Balogh* [2020] MNL-modellt leíró munkája szerinti felépítésben bemutassa az LC-modell becslésének folyamatát és az eredmények értelmezését az R Apollo csomag használatán keresztül. Ehhez egy mindenki által szabadon (az Apollo csomag weboldalán) hozzáférhető (`apollo_swissRouteChoiceData.csv`) adatbázist dolgozunk fel (*Hess–Palma* [2019a], [2019b]).

1. Látens osztályú modell becslése és eredményeinek értelmezése

Ebben a fejezetben az elemzéshez felhasznált mintaadatbázist, az LC-specifikáció felépítését, becslésének folyamatát és a kapott eredmények értelmezését mutatjuk be.

1.1. Az adatbázis

A felhasznált adatbázis 388 kitöltő különböző utazási alternatívákra vonatkozó preferenciáit vizsgálja. Az adatbázis táblázatának első oszlopában a választási szituációkban (minden egyes választási szituáció 2 lehetőséget tartalmaz) hozott döntések (a választott lehetőségek) kapnak helyet, ezeket az egyes alternatívákat jellemző tulajdonságok (*ttl* [utazási idő az első alternatíva esetében] – *ch2* [csomópontok száma a második alternatívánál]) követik. Az 1. táblázat az adatbázis első döntési helyzetét szemlélteti.

1. táblázat

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

Döntést befolyásoló tulajdonság	1. lehetőség	2. lehetőség
Time (mins) (utazási idő percben kifejezve)	58	50
Cost (CHF) (utazási költség svájci frankban kifejezve)	7	8
Headway (mins) (haladási idő percben kifejezve)	30	30
Interchanges (csomópontok száma)	1	0
Melyik utazási módot választaná? (X)		

Végül pedig a kitöltőket jellemző (*hh_inc_abs* [a kitöltő háztartásának éves jövedelme] – *leisure* [szabadidős utazó-e a válaszadó]), ún. esetspecifikus változók¹ szerepelnek az adatbázisban. További részletek a 2. táblázatban láthatók.

¹ 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>tt1/tt2</i> (utazási idő az első és második utazási lehetőségre vonatkozóan)	Az első és második alternatívára vonatkozó utazási idő, utazási költség, haladási idő és csomópontok száma az adott döntési helyzetben
<i>tc1/tc2</i> (utazási költség az első és második utazási lehetőségre vonatkozóan)	
<i>hw1/hw2</i> (haladási idő az első és második utazási lehetőségre vonatkozóan)	
<i>ch1/ch2</i> (csomópontok száma az első és második utazási lehetőségre vonatkozóan)	
<i>hh_inc_abs</i> (jövedelem)	A válaszadó háztartásának éves jövedelme (svájci frankban kifejezve)
<i>car_availability</i> (autó elérhetősége)	1, ha rendelkezésre áll 0, ha nem áll rendelkezésre
<i>commute</i> (ingázó utazó)	1, ha ingázó utazó a válaszadó 0, ha nem ingázó utazó a válaszadó
<i>shopping</i> (bevásárló utazó)	1, ha bevásárló utazó a válaszadó 0, ha nem bevásárló utazó a válaszadó
<i>business</i> (üzleti utazó)	1, ha üzleti utazó a válaszadó 0, ha nem üzleti utazó a válaszadó
<i>leisure</i> (szabadidős utazó)	1, ha szabadidős utazó a válaszadó 0, ha nem szabadidős utazó a válaszadó

1.2. A becslés folyamata

Ebben az alfejezetben az LC-modell felépítésének és becslésének jelentősebb lépéseit mutatjuk be. Kitérünk a kezdeti lépésekre, a modell alapbeállításaira, majd ismertetjük a paraméterek, az osztályszerkezetű tényezők és a hasznosságfüggvények meghatározásának módját.

Alapbeállítások és a modell paramétereinek meghatározása

Az LC-modell felépítése az R Apollo csomag esetében hasonló módon indul, mint azt *Czine–Harangi–Rákos–Balogh* [2020] az MNL-specifikációnál bemutatták.

A csomag telepítését követően az alapbeállítások megadása végezhető el, ahol a felhasználó meghatározhatja modelljének nevét, részletes leírását, továbbá szükséges az egyedi azonosítókat tartalmazó változót definiálnia. Ezután az adatbázis beolvasása következik, amelyre vonatkozóan különféle feltételek is szabhatók (ezt jelen esetben figyelmen kívül hagyjuk, részletekért lásd *Czine–Harangi–Rákos–Balogh* [2020]). A parancssort az 1. ábra szemlélteti.

1. ábra. A modell alapbeállításai és az adatok beolvasása
(Basic model settings and reading data)

```
rm(list=ls ( ) ) # A memória törlése
library(apollo) # Az Apollo csomag könyvtárának betöltése
apollo_initialise()
apollo_control <- list(
  modelName = "Apollo_example_18", # A modell neve
  ModelDescr = "Simple LC model on Swiss route choice data", # A modell leírása
  indivID = "ID", # Az egyedi azonosítókat (ID) tartalmazó változó meghatározása
  ncores = 2,
  noDiagnostics = TRUE
)
database <- read.csv("apollo_swissRouteChoiceData.csv",header=TRUE) # Az
adatbázis beolvasása
```

A következő lépés a paraméterek meghatározása az `apollo_beta` vektor használatával. Szembetűnő különbség az MNL-modellhez képest, hogy itt már ún. osztályspecifikus paramétereket kell definiálnunk. (Lásd a 2. ábrát.)

2. ábra. A paraméterek meghatározása
(Definition of the parameters)

```
apollo_beta <- c(asc_1 = 0, # A modell paramétereinek definiálása
  asc_2 = 0,
  beta_tt_a = 0,
  beta_tt_b = 0,
  beta_tc_a = 0,
  beta_tc_b = 0,
  beta_hw_a = 0,
  beta_hw_b = 0,
  beta_ch_a = 0,
  beta_ch_b = 0,
  delta_a = 0,
  delta_b = 0)

apollo_fixed <- c("asc_2","delta_b") # A kezdeti értékükön rögzített
paraméterek megadása
```

A 2. ábra szemlélteti, hogy jelen példa esetében alternatíváspecifikus konstansokat (`asc_1` és `asc_2`) és generikus (`tt_a` és `tt_b`, `tc_a` és `tc_b`, `hw_a` és `hw_b`, `ch_a` és `ch_b`) paramétereket szerepeltettünk a felsorolásban (a paraméterek típusairól

bővebben *Czine–Harangi–Rákos–Balogh* [2020] nyújt áttekintést). Ezentúl az ún. osztályallokációs egyenletek konstans paramétereit (**delta_a** és **delta_b**) is definiáltuk, melyekről a következő alfejezet szolgáltat magyarázatot. Végül pedig a becslési folyamat során kezdeti értékükön rögzíteni kívánt paramétereket határoztuk meg az **apollo_fixed()** függvénnyel.

Az osztályspecifikus tényezők és a hasznossági függvények meghatározása

Az LC-modell felépítésének következő lépésében a becslni kívánt osztályokra vonatkozóan néhány szempontot kell meghatározunk. (Lásd a 3. ábrát.) Fontos kiemelni azt, hogy ezen résztől kezdődően a közölt kódok nem futtathatók, működésükhöz azok függvénybe szervezése szükséges.

3. ábra. Az osztályspecifikus tényezők
(Class-specific factors)

```
lcpars <- list() # Az osztályspecifikus attribútumok csoportosítása
lcpars[["beta_tt"]] <- list(beta_tt_a, beta_tt_b)
lcpars[["beta_tc"]] <- list(beta_tc_a, beta_tc_b)
lcpars[["beta_hw"]] <- list(beta_hw_a, beta_hw_b)
lcpars[["beta_ch"]] <- list(beta_ch_a, beta_ch_b)

V<-list() # Az osztályallokációs egyenletek definiálása
V[["class_a"]] <- delta_a
V[["class_b"]] <- delta_b

mnl_settings <- list(
  alternatives = c(class_a=1, class_b=2), # A becslni kívánt osztályok
  meghatározása
  avail = 1,
  choiceVar = NA,
  V = V
)
```

Első lépés a 3. ábra alapján, az **lcpars** lista használatával, az egyazon tulajdonságra vonatkozó osztályspecifikus paraméterek meghatározása. Miután ezt elvégeztük, az osztályallokációs egyenletek definiálása következik. Mivel jelen példa egy kétosztályú LC-modell becslését illusztrálja, elegendő lenne csupán egy osztályra vonatkozóan felírni az egyenletet (miután egyet mindig fixen, referenciaosztályként kell kezelnünk), míg a másikat egyszerűen elhagyni (ebben az esetben a korábbiakban annak delta paraméterét sem kellene meghatározni). Azonban a könnyebb átláthatóság érdekében most a fixen rögzített osztály (jelen esetben ez a **class_b**) egyenletét és részeit is bemutatjuk. Az osztályallokációs egyenlet felépítése lineáris formulán alapul, a 3. ábra segítségével szemléltetett példában abban mindössze egy konstans tagot (delta) szerepeltettünk. Fontos említést tenni arról, hogy ez kibővíthető

különböző magyarázó változókkal is annak érdekében, hogy pontosabb képet kaphassunk az ízlések heterogenitásának forrásáról. Emellett szintén szükséges szót ejtenünk az `mnl_settings` lista használatáról, ahol mindössze az „alternatives” sort kell esetlegesen módosítanunk abban az esetben, ha nem csak kétosztályú LC-modellt szeretnénk becsülni. A 4. ábrán egy háromosztályú specifikációt mutatunk be, ahol két magyarázó változót is feltüntettünk az osztályallokációs egyenletben.

4. ábra. Paraméterek és osztályspecifikus tényezők háromosztályú esetben
(Parameters and class-specific factors in the case of three classes)

```
apollo_beta <- c(asc_1 = 0, # A modell paramétereinek definiálása
               asc_2 = 0,
               beta_tt_a = 0,
               beta_tt_b = 0,
               beta_tt_c = 0,
               .
               .
               .
               delta_a = 0,
               delta_b = 0,
               delta_c = 0,
               gamma_commute_a = 0,
               gamma_commute_b = 0,
               gamma_commute_c = 0,
               gamma_car_av_a = 0,
               gamma_car_av_b = 0,
               gamma_car_av_c = 0,

apollo_fixed <- c("asc_2", "delta_b", "gamma_commute_b", "gamma_car_av_b") # A
kezdeti értékükön rögzített paraméterek megadása

lcpars <- list() # Az osztályspecifikus attribútumok csoportosítása
lcpars[["beta_tt"]] <- list(beta_tt_a, beta_tt_b, beta_tt_c)
lcpars[["beta_tc"]] <- list(beta_tc_a, beta_tc_b, beta_tc_c)

V<-list() # Az osztályallokációs egyenletek definiálása
V[["class_a"]] <- delta_a + gamma_commute_a*commute + gamma_car_av_a*_car_av
V[["class_b"]] <- delta_b + gamma_commute_b*commute + gamma_car_av_b*_car_av
V[["class_c"]] <- delta_c + gamma_commute_c*commute + gamma_car_av_c*_car_av

mnl_settings <- list(
  alternatives = c(class_a=1, class_b=2, class_c=3), # A becsülni kívánt
  osztályok meghatározása
  avail = 1,
  choiceVar = NA,
  V = V
)
```

A 4. ábra azt mutatja, hogy az osztályok számának növelésekor mindössze az új paramétereinket kell megadnunk az `apollo_beta` vektor felsorolásában, majd

ezek közül az alternatívákhoz kapcsolódókat az **lcpars** listához, az osztályokra vonatkozókat pedig az osztályallokációs részhez.

A hasznosságfüggvények meghatározása előtt még meg kell adnunk a becslési kívánt osztályok számát. (Lásd az 5. ábrát.) Fontos említést tenni arról, hogy a jobb érthetőség kedvéért a továbbiakban a 2. és a 3. ábrán bemutatott kétosztályú esetet tárgyaljuk tovább.

5. ábra. Az osztályok számának és a hasznosságfüggvények meghatározása
(Determination of the number of classes and definition of the utility functions)

```
s=1
while(s<=2) {
    # A becslési kívánt osztályok számának meghatározása
    V[['alt1']] = asc_1 + beta_tt[[s]]*tt1 + beta_tc[[s]]*tc1 + beta_hw[[s]]*hw1 +
    beta_ch[[s]]*ch1
    # A hasznossági függvények meghatározása
    V[['alt2']] = asc_2 + beta_tt[[s]]*tt2 + beta_tc[[s]]*tc2 + beta_hw[[s]]*hw2 +
    beta_ch[[s]]*ch2
}
```

Az 5. ábra azt ismerteti, hogy a hasznosságfüggvényekben az **lcpars** lista felsorolásában definiált paramétereket (jelen példában alternatíváspecifikus [**asc_1**, **asc_2**] és generikus [**beta_tt**, **beta_tc**, **beta_hw** és **beta_ch**] paramétereket), illetve az adatbázisunk változóit (**tt1**, **tt2**, **tc1**, **tc2**, **hw1**, **hw2**, **ch1** és **ch2**) miként szükséges szerepeltetnünk.

1.3. Az eredmények értelmezése

Az alfejezetben a kétosztályú specifikáció becslésének eredményeit, a fizetési hajlandóság meghatározásának módját és a háromosztályú esetet mutatjuk be.

A kétosztályú LC-modell becslésének eredményei

Az LC-becslést követően először a teljes modellünkre vonatkozóan kapunk áttekintő adatokat a többi specifikációhoz hasonló módon. Itt a modell és a becslési folyamat alapvető jellemzői, továbbá a különféle információs kritériumok értékei (melyek a modell illeszkedését hivatottak számszerűsíteni) szerepelnek. Ez utóbbiakra különösen nagy hangsúly kerül az LC-specifikáció esetében, mivel a kutatók többnyire ezek alapján hozhatnak döntést arról, hogy meddig célszerű növelniük a becslési kívánt osztályok számát. Az áttekintő adatok részleteit a 3., míg az optimális osztályszámú modell kiválasztásakor figyelembe veendő szempontokat a 4. táblázat ismerteti.

3. táblázat

A kétosztályú modell áttekintő adatai
(Overview data of the two-class model)

Model name (A modell neve)	Apollo_example_18 (Apollo_példa_18)
Model description (A modell leírása)	Simple LC model on Swiss route choice data (Egyszerű LC-modell a svájci útvonalválasztási adatokra vonatkozóan)
Model run at (Becslés időpontja)	2021-01-24 15:20:32
Number of individuals (Egyének száma)	388
Number of observations (Megfigyelések száma)	3 492
LL(start) (LL kezdeti értéke)	-2 420,47
LL(0, whole model) (LL 0 érték a teljes modellre vonatkozóan)	-2 420,47
LL(final, whole model) (LL végső érték a teljes modellre vonatkozóan)	1 578,24
R -squared (R^2 érték)	0,35
Adjusted R -squared (Korrigált R^2 érték)	0,34
AIC (Akaiké-féle információs kritérium)	3 176,49
BIC (Bayesi információs kritérium)	3 238,07
LL(0, Class1) (LL 0 érték az első osztályra vonatkozóan)	-2 420,47
LL(final, Class1) (LL végső érték az első osztályra vonatkozóan)	-1 886,57
LL(0, Class2) (LL 0 érték a második osztályra vonatkozóan)	-2 420,47
LL(final, Class2) (LL végső érték a második osztályra vonatkozóan)	-1 955,64
Estimated parameters (Becsült paraméterek)	10
Time taken (Időtartam)	00:00:19.73
Iterations (Iterációk)	33

Megjegyzés. LL: log-likelihood.

4. táblázat

Információs kritériumok értékei a különböző osztályszámú modellek esetében
(Values of information criteria for models with different class numbers)

Információs kritérium	Kétosztályú modell	Háromosztályú modell	Négyosztályú modell	Ötosztályú modell	Hatososztályú modell
LL (végső)	-1 578,244	-1 496,831	-1 449,338	-1 428,224	-1 421,104
Pseudo R^2	0,348	0,3816	0,4012	0,4099	0,4129
AIC	3 176,49	3 023,66	2 938,68	2 906,45	2 902,21
BIC	3 238,07	3 116,04	3 061,84	3 060,40	3 086,95

A 4. táblázat eredményei alapján a log-likelihood (végső) AIC és BIC értéke egészen az ötosztályú esetig csökken, míg a pseudo R^2 nő; mindezek a modellek

illeszkedésének javulására utalnak. Megfigyelhető azonban, hogy a hatosztályú modellnél a BIC értéke már növekszik, így jelen specifikáció nem mutat egyértelmű javulást az előző esethez képest. Ebben a helyzetben a kutatóknak célszerű egyéb, további szempontokat (például a legalacsonyabb osztályvalószínűség értékét, a szignifikáns paraméterek számát) is figyelembe venniük annak érdekében, hogy a legjobb döntést tudják meghozni (Hess [2014]).

Az 5. táblázat a paraméterbecsléseket tartalmazza, ahol a sorokban a paraméterek megnevezései, míg az oszlopokban az együttható, a standard hiba, a t -érték és utóbbi kettő robusztus formái találhatóak. Ezen kívül a paraméterbecslésekre vonatkozó p -értéket Czine–Harangi–Rákos–Balogh [2020] szerint lehet előállítani.

5. táblázat

A kétosztályú modell paraméterbecsléseinek eredményei
(Estimation results of the two-class-model parameters)

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_1	-0,0094	0,0472	-0,20	0,0499	-0,19
asc_2	0,0000	n. a.	n. a.	n. a.	n. a.
beta_tt_a	-0,0489*	0,0073	-6,67	0,0105	-4,65
beta_tt_b	-0,1241*	0,0124	-9,98	0,0203	-6,12
beta_tc_a	-0,2445*	0,0353	-6,92	0,0681	-3,59
beta_tc_b	-0,1416*	0,0212	-6,68	0,0313	-4,53
beta_hw_a	-0,0327*	0,0036	-9,22	0,0054	-6,04
beta_hw_b	-0,0506*	0,0043	-11,88	0,0049	-10,32
beta_ch_a	-0,6628*	0,1081	-6,13	0,2100	-3,16
beta_ch_b	-2,0510*	0,1459	-14,05	0,2193	-9,35
delta_a	-0,3817*	0,2202	-1,73	0,3513	-1,09
delta_b	0,0000	n. a.	n. a.	n. a.	n. a.

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

Következtetéseink az 5. táblázat eredményei alapján:

- döntési szabályszerűség nem mutatkozott a választások során, azaz nem választották szignifikánsan kevesebbszer vagy többször egyik alternatívát sem a másikhoz képest (erre az **asc_1** nem szignifikáns értéke utal);
- az a osztályba (a b osztályhoz képest) szignifikánsan kevesebb válaszadó került (erre a **delta_a** negatív, szignifikáns értéke utal);

- mindkét osztályra vonatkozóan, minden szolgáltatásattribútum szignifikáns hatást képvisel;
- minden vizsgált szolgáltatásattribútumban bekövetkező növekedés negatívan hat a kitöltők hasznosságérzetére (mind az a , mind pedig a b osztály esetében);
- mindkét osztályt tekintve a ch (csomópontok száma) egy egységgel történő növekedése csökkenti leginkább a válaszadók hasznosságérzetét.

A paraméterbecsléseket követően az osztályvalószínűségi értékeket láthatjuk, melyek azt mutatják meg, hogy milyen valószínűséggel kerültek az egyes osztályokba a mintánk válaszadói. (Lásd a 6. táblázatot).

6. táblázat

A kétosztályú modell osztályvalószínűségi értékei
(Class probability values of the two-class model)

Osztály	Valószínűség
a	0,4057
b	0,5943

A 6. táblázat eredményei arra utalnak, hogy LC-modellünk b osztályába nagyobb valószínűséggel kerültek be mintánk válaszadói.

Fizetési hajlandóságra vonatkozó kalkulációk a látens osztályú modell esetében

Fizetési hajlandóságra (willingness to pay, WTP) vonatkozó kalkulációkat végezhetünk az LC-specifikáció esetében a *Czine–Harangi–Rákos–Balogh* [2020] által bemutatott `deltaMethod()` függvény alkalmazásával. (Lásd a 6. ábrát.)

6. ábra. A `delta_Method` alkalmazása LC-specifikáció esetében
(Application of the `delta_Method` for LC specification)

```
delta_Method_settings=list(operation="ratio",parName="beta_tt_a",
parName2="beta_tc_a") # A „deltaMethod” alkalmazása
apollo_deltaMethod(model, delta_Method_settings)

delta_Method_settings=list(operation="ratio",parName="beta_tt_b",
parName2="beta_tc_b")
apollo_deltaMethod(model, delta_Method_settings)
```

A 6. ábra példájában arra a kérdésre szeretnénk választ kapni, hogy az utazási idő változása hogyan hat az a és a b osztály válaszadóinak fizetési hajlandóságára. Az eredményeket a 7. táblázat ismerteti.

7. táblázat

A delta_Method alkalmazásának eredményei az LC-specifikáció esetén
(Results of the delta_Method applied for LC specification)

Paraméter	Value (Érték)	Robust standard error (Robusztus standard hiba)	Robust t -ratio (Robusztus t -érték)
Ratio of beta_tt_a and beta_tc_a (a beta_tt_a és beta_tc_a paraméterek hányadosa)	0,2001*	0,0284	7,06
Ratio of beta_tt_b and beta_tc_b (a beta_tt_b és beta_tc_b paraméterek hányadosa)	0,8764*	0,1629	5,38

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

A 7. táblázat eredményeit tekintve a **deltaMethod** outputablaka hasonlóképpen épül fel, mint a paraméterbecsléseké. (Lásd az 5. táblázatot.) Az oszlopokban a becsült együtthatók, a robusztus standard hiba és t -értékek, míg a sorokban a vizsgált (egymáshoz viszonyított) paraméterek megnevezései láthatók. Az eredmények alapján arra következtethetünk, hogy az a osztály válaszadói megközelítőleg 0,2001, míg a b osztály tagjai hozzávetőlegesen 0,8764 svájci frankkal fizetnének kevesebbet abban az esetben, ha az utazási idő egy perccel növekedne.

A háromosztályú modell becslésének eredményei

A következő lépésben a 4. ábrán szemléltetett kibővített modellünk becslésének eredményeit ismertetjük. Az áttekintő adatokat a 8. táblázat, a paraméterbecsléseket a 9. táblázat, míg az osztályvalószínűségi értékeket a 10. táblázat mutatja be.

Az áttekintő adatok alapján megállapítható, hogy több helyen is különbség mutatkozik a kétosztályú esethez képest. Elsőként az figyelhető meg, hogy az információs kritériumok értékei – log-likelihood (final), R -squared, AIC és BIC – egytől egyig javultak a kétosztályú (az osztályallokációs egyenletben további változókat nem szerepeltető) modellhez képest. A következő eltérés, hogy a log-likelihood értékek száma kettővel bővült – egy $LL(0)$ és egy $LL(\text{final})$ értékkel – az új osztály bevezetésének köszönhetően. Megállapítható, hogy a becsült paraméterek száma is nőtt, ami egyrészt az új osztály bevezetéséből, másrészt az osztályallokációs egyenlet kibővítéséből következik.

8. táblázat

A háromosztályú modell áttekintő adatai
(Overview data of the three-class model)

Model name (A modell neve)	Apollo_example_bővített_modell (Apollo_pelda_18_bővített_modell)
Model description (A modell leírása)	Simple LC model on Swiss route choice data (Egyszerű LC-modell a svájci útvonalválasztás adataira vonatkozóan)
Model run at (Becslés időpontja)	2021-02-03 13:10:23
Number of individuals (Egyének száma)	388
Number of observations (Megfigyelések száma)	3 492
LL(start) (LL kezdeti értéke)	-2 420,47
LL(0, whole model) (LL 0 érték a teljes modellre vonatkozóan)	-2 420,47
LL(final, whole model) (LL végső érték a teljes modellre vonatkozóan)	-1 485,29
R -squared (R^2 érték)	0,39
Adjusted R -squared (Korrigált R^2 érték)	0,38
AIC (Akaike-féle információs kritérium)	3008,57
BIC (Bayesi információs kritérium)	3 125,58
LL(0, Class1) (LL 0 érték az első osztályra vonatkozóan)	-2 420,47
LL(final, Class1) (LL végső érték az első osztályra vonatkozóan)	-5 823,25
LL(0, Class2) (LL 0 érték a második osztályra vonatkozóan)	-2 420,47
LL(final, Class2) (LL végső érték a második osztályra vonatkozóan)	-1 778,28
LL(0, Class3) (LL 0 érték a harmadik osztályra vonatkozóan)	-2 420,47
LL(final, Class3) (LL végső érték a harmadik osztályra vonatkozóan)	-2 723,45
Estimated parameters (Becsült paraméterek)	19
Time taken (Időtartam)	00:01:19,83
Iterations (Iterációk)	94

9. táblázat

A háromosztályú modell paraméterbecsléseinek eredményei
(Estimation results of the three-class-model parameters)

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_1	-0,0530	0,0533	-0,99	0,0814	-0,65
asc_2	0,0000	n. a.	n. a.	n. a.	n. a.
beta_tt_a	-0,2166*	0,0338	-6,41	0,0596	-3,64
beta_tt_b	-0,0495*	0,0077	-6,46	0,0221	-2,24
beta_tt_c	-0,2632*	0,0548	-4,80	0,2006	-1,31

(A táblázat folytatása a következő oldalon)

(Folytatás)

Becsült paraméter	Estimate (Becsült együtt- ható)	Standard error (Standard hiba)	<i>t</i> -ratio (<i>t</i> -érték)	Robust standard error (Robusztus standard hiba)	Robust <i>t</i> -ratio (Robusztus <i>t</i> -érték)
beta_tc_a	-1,7869*	0,2503	-7,14	0,4037	-4,43
beta_tc_b	-0,0523*	0,0146	-3,58	0,0233	-2,24
beta_tc_c	-0,8296*	0,1497	-5,54	0,4765	-1,74
beta_hw_a	-0,0520*	0,0106	-4,90	0,0197	-2,65
beta_hw_b	-0,0387*	0,0037	-10,53	0,0058	-6,67
beta_hw_c	-0,0636*	0,0072	-8,84	0,0120	-5,31
beta_ch_a	-1,5830*	0,2653	-5,97	0,5398	-2,93
beta_ch_b	-0,6780*	0,0923	-7,35	0,1885	-3,60
beta_ch_c	-3,4392*	0,3925	-8,76	0,8895	-3,87
delta_a	-0,6555*	0,2821	-2,32	0,3730	-1,76
delta_b	0,0000	n. a.	n. a.	n. a.	n. a.
delta_c	0,1568	0,2830	0,55	0,6662	0,24
gamma_commute_a	0,8859*	0,4876	1,82	0,6737	1,31
gamma_commute_b	0,0000	n. a.	n. a.	n. a.	n. a.
gamma_commute_c	0,4771	0,3766	1,27	0,4898	0,97
gamma_car_av_a	-0,9413*	0,4392	-2,14	0,5539	-1,70
gamma_car_av_b	0,0000	n. a.	n. a.	n. a.	n. a.
gamma_car_av_c	-0,6577*	0,3182	-2,07	0,4399	-1,50

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

Következtetéseink a 9. táblázat eredményei alapján:

- döntési szabályszerűség nem mutatkozott a választások során, azaz nem választották szignifikánsan kevesebbszer vagy többször egyik alternatívát sem a másikkal képest (erre az **asc_1** nem szignifikáns értéke utal);
 - az *a* osztályba (a *b* osztályhoz képest) szignifikánsan kevesebb válaszadó került (erre a **delta_a** negatív, szignifikáns értéke utal);
 - az *a* osztályba szignifikánsan több ingázó utazó, míg ugyancsak az *a* és a *c* osztályba kevesebb autóval rendelkező utazó került (a *b* osztályhoz képest) (erre a **commute_a** pozitív és a **gamma_car_av_a**, valamint **gamma_car_av_c** negatív, szignifikáns értékei utalnak);
 - mindhárom osztályra vonatkozóan, minden szolgáltatásattribútum szignifikáns hatást képvisel;

- minden vizsgált szolgáltatásattribútumban bekövetkező növekedés negatívan hat a kitöltők hasznosságérzetére (mindhárom osztály esetében);
- az a osztály esetében a tc (utazási költség), míg a b és c osztály válaszadóinál a ch (csomópontok száma) egy egységgel történő növekedése csökkenti leginkább a hasznosságérzetet.

10. táblázat

A háromosztályú modell osztályvalószínűségi értékei
(Class probability values of the three-class model)

Osztály	Valószínűség
a	0,1915
b	0,4013
c	0,4072

A 10. táblázat eredményei arra utalnak, hogy LC-modellünk b és c osztályába közel azonos, míg a osztályába lényegesen kisebb valószínűséggel kerültek mintánk válaszadói.

2. Összegzés

A tanulmányban az LC diszkrét választási modell becslését ismertettük az R Apollo csomag használatán keresztül, melynek előnyös tulajdonságai közé tartozik, hogy képes kezelni a preferenciákban levő heterogenitást, ami az MNL-specifikáció egyik hiányossága.

Első lépésben az adatbázist mutattuk be, amely bárki számára szabadon hozzáférhető az Apollo csomag weboldalán. Ezt követően a becslés folyamatára és az eredmények értelmezésére fektettük a hangsúlyt, ahol egy standard kétosztályú modell becslésén túl bemutattuk, hogy az osztályok számának növelése és az osztályallokációs egyenlet további változókkal történő kibővítése milyen hatásokkal jár együtt. Emellett ajánlásokat tettünk az optimális osztályszámú modell kiválasztására és a WTP-re vonatkozó kalkuláció módjára.

A tanulmányban bemutatott példák elvégzéséhez szükséges R parancsokat és azok eredményeit mellékletként elérhetővé tettük az érdeklődők számára (http://www.ksh.hu/statszemle_archive/all/2021/2021_05/06_czine_melleklet.doc).

Irodalom

- BERGANTINO, A. S. – CAPURSO, M. – HESS, S. [2020]: Modelling regional accessibility to airports using discrete choice models: An application to a system of regional airports. *Transportation Research Part A: Policy and Practice*. Vol. 132. February. pp. 855–871. <http://dx.doi.org/10.1016/j.tra.2019.12.012>
- COSTA, E. – MONTEMURRO, D. – GIULIANI, D. [2019]: Consumers' willingness to pay for green cars: A discrete choice analysis in Italy. *Environment, Development, and Sustainability*. Vol. 21. Issue 5. pp. 2425–2442. <http://dx.doi.org/10.1007/s10668-018-0141-z>
- CZINE P. – HARANGI-RÁKOS M. – BALOGH P. [2020]: Diszkrét választási modellek becslése az R Apollo csomagjának használatával – multinomiális logit modell. *Statistikai Szemle*. 98. évf. 11. sz. 1310–1323. old. <http://dx.doi.org/10.20311/stat2020.11.hu1310>
- GREENE, W. H. – HENSHER, D. A. [2003]: A latent class model for discrete choice analysis: Contrasts with mixed logit. *Transportation Research Part B: Methodological*. Vol. 37. Issue 8. pp. 681–698. [https://doi.org/10.1016/S0191-2615\(02\)00046-2](https://doi.org/10.1016/S0191-2615(02)00046-2)
- 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. (eds.) [2014]: *Handbook of Choice Modelling*. Edward Elgar Publishing. Cheltenham. <http://dx.doi.org/10.4337/9781781003152>
- HESS, S. [2014]: Latent class structures: Taste heterogeneity and beyond. In: Hess, S. – Daly A. (eds.): *Handbook of Choice Modelling*. Edward Elgar Publishing. Cheltenham. pp. 311–329. <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
- LIAO, Q. – NG, T. W. Y. – COWLING, B. J. [2020]: What influenza vaccination programmes are preferred by healthcare personnel? A discrete choice experiment. *Vaccine*. Vol. 38. Issue 29. pp. 4557–4563. <http://dx.doi.org/10.1016/j.vaccine.2020.05.012>
- LOUVIERE, J. – HENSHER, D. A. – SWAIT, J. [2000]: *Stated Choice Methods: Analysis and Application*. Cambridge University Press. Cambridge.
- MCFADDEN, D. [1974]: 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>
- WANG, J. – GE, J. – MA, Y. [2018]: Urban Chinese consumers' willingness to pay for pork with certified labels: A discrete choice experiment. *Sustainability*. Vol. 10. Issue 3. 603. pp. 1–14. <http://dx.doi.org/10.3390/su10030603>