

# Statisztikai Szemle

Közzététel: 2020. május 12.

A tanulmány címe:

**Diszkrét választási modellek bemutatása, különös tekintettel a latent class elemzésre**

Szerzők:

CZINE PÉTER, a Debreceni Egyetem PhD-hallgatója

E-mail: [czine.peter@econ.unideb.hu](mailto:czine.peter@econ.unideb.hu)

BALOGH PÉTER, a Debreceni Egyetem

egyetemi tanára

E-mail: [balogh.peter@econ.unideb.hu](mailto:balogh.peter@econ.unideb.hu)

DOI: <https://doi.org/10.20311/stat2020.5.hu0400>

**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 (Szt.) 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 Szt. 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 5. számában megjelent, **Czine Péter, Balogh Péter** által írt, **'Diszkrét választási modellek bemutatása, különös tekintettel a latent class elemzésre'** 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 – Balogh Péter

## **Diszkrét választási modellek bemutatása, különös tekintettel a latent class elemzésre\***

### **Presentation of discrete choice models, with special reference to latent class analysis**

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

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

A tanulmány célja, hogy egy margarinofogyasztásra vonatkozó preferenciákat vizsgáló kísérlet adatbázisát példaként felhasználva különböző, a hazai szakirodalomban még kevésbé tárgyalt diszkrét választási modelleket mutasson be, majd összevesse azok eredményeit. A becült modellek között a már elterjedtebbnek számító multinomiális és random paraméterű logit mellett egy „félparametrikus” megoldás, a latent class és annak egy kiterjesztése, a random paraméterű latent class modellek is szerepelnek. Az eredmények alapján megállapítható, hogy a modellek paraméterre és fizetési hajlandóságra vonatkozó becslései között jelentős eltérések vannak.

TÁRGYSZÓ: diszkrét választási kísérlet, latent class modell, random paraméterű latent class modell

The aim of the study is to present some discrete choice models that are less discussed in the Hungarian literature, using the database of an experiment on margarine consumption preferences as an example, and then compare their results. The estimated models include, in addition to the multinomial and random parameter logit models, a „semi-parametric” solution, the latent class model and its extension, the random parameter latent class models. Based on the results, it can be concluded that there are significant differences between the parameter and willingness-to-pay estimates of the models.

KEYWORD: discrete choice experiment, latent class model, random parameter latent class model

\* A tanulmány az EFOP 3.6.3-VEKOP-16-2017-00007 „»Tehetségből fiatal kutató« – A kutatói életpályát támogató tevékenységek a felsőoktatásban” című projekt támogatásával készült.

A preferenciaértékelő eljárások közé tartozó diszkrét választási kísérletek elemzésére számos modell terjedt el az elmúlt évek során. A módszer iránt mutatott növekvő érdeklődést mi sem bizonyítja jobban, mint az, hogy folyóiratot (*Journal of Choice Modelling* néven) is létrehozta a területhez tartozó cikkek bemutatására. A lap alapszemlélete, hogy a választási modellezéshez kapcsolódó módszertani újításokat bemutassa.<sup>1</sup> Fontos említést tenni arról, hogy a módszer használata Magyarországon még kevésbé elterjedtnek tekinthető. A hazai cikkek közül *Baji* [2012] egy elméleti áttekintést nyújt, melyben a hagyományos MNL- (multinomial logit – multinomiális logit) és az NL- (nested logit – beágyazott logit) modellek bővebben, míg az RPL- (random parameter logit – random paraméterű logit) és LC- (latent class – látens osztály) modellek csak érintőlegesen kerülnek szóba. *Brandtmüller* [2009] tanulmányában már empirikus kutatási eredmények is találhatóak. A szerző az orvosok betegválasztási preferenciáit vizsgálta a módszertan alkalmazásával. Becsült modellspecifikációi között a multinomiális logiton kívül, random paraméteres megoldás is megtalálható. Eredményei alapján azt a megállapítást tette, hogy az RPL szignifikánsan jobb illeszkedést mutat, mint a fix paraméterekkel rendelkező MNL-modell, tehát az orvosok preferenciájában heterogenitás tapasztalható.

A modellek többsége a véletlen hasznosság elméletén alapul, azaz az egyének hasznosságmaximalizáló viselkedését feltételezi. Legelsőnek a *McFadden* [1973] nevéhez köthető MNL tekinthető, amely könnyed alkalmazhatósága miatt manapság is igen elterjedt. Számos korlátja (melyek között elsőként az irreleváns alternatívák<sup>2</sup> függetlensége feltételezést szokás említeni) miatt azonban egyre gyakrabban felmerülő kérdés az, hogy mely modellek a leginkább alkalmasak a helyettesítésére. Az RPL-modell, amely képes az egyéni preferenciák heterogenitásának megragadására, valamint rugalmas variancia-kovariancia struktúrát engedélyez a hibatagokra vonatkozóan, meglehetősen vonzó lehetőséget nyújt, viszont számos nehezen megválaszolható kérdést is hordoz magában. Valahol a kettő között (MNL és RPL) helyezhető el az LC-modell, amely ugyan nem ragadja meg egyénspecifikusan a preferenciákat, egymástól eltérő csoportokat képez, melyekre különböző paramétervektorokat határoz meg (*Greene–Hensher* [2003]). A csoportokon belül pedig a hagyományos MNL-modell alapján történik a becslés (*Shen* [2009]). Egy újfajta kiterjesztés,

<sup>1</sup> <https://www.journals.elsevier.com/journal-of-choice-modelling>

<sup>2</sup> Az „alternatíva” a továbbiakban a nemzetközi szakirodalomból átvett „alternative” kifejezéssel egyezik meg.

az RLC- (random parameter latent class – random paraméterű látens csoport) modell, amely a két utóbb említett alternatívát kombinálja, és mind a csoportokon belül, mind pedig azok között lehetővé teszi a heterogenitást (*Greene–Hensher* [2013]).

A modellek elemzésére napjainkban már számos program nyújt lehetőséget. Ezek közé lehet sorolni a STATA, Nlogit, R: Apollo szoftvereket és szoftverbővítvényeket, melyek mindegyikének megvan a sajátos tulajdonsága. Példaként említve, amíg a STATA és az Nlogit adatstruktúrája az ún. „long data” formát követi, addig az Apollo bővítvény a „wide” jellegűt preferálja (*Hensher–Rose–Greene* [2005], *Hess–Palma* [2019a]). További differenciaként van jelen a választható modellspecifikációk széleskörűsége (*Lancsar–Fiebig–Hole* [2017]). A STATA nem teszi lehetővé látens osztályú és további haladóbb szintű modellek becslését, míg például az Apolloval a nem véletlen hasznosságon alapuló (random regret minimisation, decision field theory), random paraméterekkel rendelkező latent class és hibrid modelleket is elemezhetünk (*Hess–Palma* [2019b], *StataCorp LLC* [2019]).

Az említettek tükrében jelen tanulmány célja, hogy egy margarinfogyasztásra vonatkozó diszkrét választási kísérlet adatbázisát példaként felhasználva, a korábban hivatkozott hazai művekben bemutatottak mellett két újabb (LC- és RLC-) modellt ismertessen, majd összevesse azok eredményeit. Mindezt az egyik leginkább elterjedt ingyenes R: Apollo szoftvercsomag alkalmazásával.

## 1. Anyag és módszer

A következő fejezetben négy diszkrét választási modellt (multinomiális logit, random paraméterű logit, latent class, random paraméterű latent class) mutatunk be. Ismertetjük a jellemzőiket (különös hangsúlyt fektetve erősségeikre, gyengeségeikre és alkalmazási problémáikra), majd a kutatás során felhasznált adatokat is.

### 1.1. Multinomiális logit modell

A diszkrét választási kísérletek eredményeinek elemzésére leggyakrabban alkalmazott MNL-modell *McFadden* [1973] nevéhez köthető. A véletlen hasznosság elméletén alapul, ami azt jelenti, hogy az egyének a számukra legnagyobb hasznosságértékkal rendelkező lehetőséget választják a rendelkezésre álló döntési halmaz elemei közül. Ebben az esetben a teljes hasznosság /1/ egy szisztematikus és egy nem megfigyelhető részre /2/ bontható fel.

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

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

ahol  $\beta$  a  $k$  attribútumhoz tartozó együttható értéke,  $x$  a megfigyelt változó,  $n$  pedig a válaszadót,  $i$  az alternatívát,  $k$  az attribútumot,<sup>3</sup>  $V_{n,i}$  a szisztematikus részt és  $\varepsilon_{n,i}$  a nem megfigyelhető részt fejezi ki.

A modell esetében adott alternatíva választási valószínűsége ( $n$ -edik személy,  $i$ -edik alternatívára vonatkoztatva) a /3/-ból következik.

$$Prob_{n,i} = \frac{\exp \sum_{k=1}^K \beta_k X_{n,i,k}}{\sum_{i=1}^I \exp \sum_{k=1}^K \beta_k X_{n,i,k}} \quad /3/$$

Fontos említést tenni arról, hogy két jelentős korlátozó feltétele is van a modellnek, melyek közül az egyik, hogy homogén preferenciák vonatkoznak minden egyes válaszadóra. Ebből azt valószínűsíthetnénk, hogy azonos jellemzőkkel rendelkező személyek ugyanolyan preferenciákat mutatnak bizonyos terméktulajdonságokra vonatkozóan. A másik hátránya pedig az irreleváns alternatívák függetlenségének feltételezése, melyből az következik, hogy a döntési helyzet alternatívái között nincs korreláció (*Balogh* [2017]).

## 1.2. Random paraméterű logit modell

Az RPL-modell több nagy előnnyel is rendelkezik versenytársaihoz képest. Ezek közül az egyik, hogy képes a preferenciaheterogenitás megragadására. Ezt úgy valósítja meg, hogy a  $\beta$ -kat előre meghatározott eloszlás szerint engedi változni a válaszadók között, majd ezek paramétereit (várható érték, szórás) becsüli (*Fosgerau–Bierlaire* [2007]). A másik, hogy feloldja az irreleváns alternatívák függetlenségének feltételezését úgy, hogy a hasznosság nem megfigyelhető részére, azaz a hibatagra vonatkozóan rugalmas variancia-kovariancia struktúrát engedélyez (*Train* [2003]). Említést kell tenni arról, hogy az RPL-ben a  $\beta$ -k eloszlása a kutató által szabadon választható (*Balogh et al.* [2016]). Ez lehet akár normál, lognormál, egyenletes és számos további eloszlástípus. A becslések végrehajtásához véletlenszerű mintavételi eljárás szükséges, melyet általában „Halton-húzással” szoktak lefolytatni (*Chang–Lusk* [2011]).

<sup>3</sup> Termék/szolgáltatás tulajdonsága.

Az  $n$ -edik döntéshozó,  $i$ -edik alternatívára vonatkozó hasznosságának szisztematikus részét a /4/ adja meg (Hensher–Rose–Greene [2008]).

$$V_{n,i} = (\bar{\beta} + \eta_n) X_{n,i}, \quad /4/$$

ahol  $\bar{\beta}$  az átlagtagot,  $\eta_n$  pedig a személytől függő eltérést szimbolizálja.

Az  $n$ -edik személy  $i$ -edik alternatívára vonatkozó választási valószínűsége a  $t$ -edik döntési szituációban az /5/ alapján határozható meg (Shen [2009], Hole–Kolstad [2011]).

$$Prob_{n,i,t} = \frac{\exp(\alpha + \beta X_{n,i,t} + \varphi F_{n,i,t})}{\sum_{i=1}^I \exp(\alpha + \beta X_{n,i,t} + \varphi F_{n,i,t})}, \quad /5/$$

ahol  $\alpha$  az alternatíváspecifikus konstans értéket,  $\beta$  a random paraméter vektorát,  $\varphi$  a fix paraméter vektorát,  $X_{n,i,t}$  és  $F_{n,i,t}$  pedig az egyénspecifikus karakterisztikák és alternatíváspecifikus attribútumok változóit jelölik random ( $X$ ) és fix ( $F$ ) paraméterek esetén.

A random paraméterű modellnek több alternatívája is megtalálható a szakirodalomban, melyek egy része a preferencia heterogenitás mellett, a skála heterogenitást is figyelembe veszi. Ezekről nyújt áttekintést Greene–Hensher [2010] és Fiebig *et al.* [2010].

### 1.3. Latent class modell

Az LC-modellt egyfajta „félparametrikus” megoldásnak is nevezik a szakirodalomban, az MNL és RPL között elhelyezve (Greene–Hensher [2003]). Alapfeltevése, hogy az egyének viselkedése a megfigyelhető tulajdonságok és a látens heterogenitás függvénye, ez utóbbi pedig az elemzők által nem megfigyelhető faktorokból tevődik össze (Kamakura–Russell [1989]). Az LC ezt a változatosságot célozza megragadni diszkrét paraméterek becslésén keresztül. Feltételezései szerint a személyek bizonyos számú ( $Q$ ), egymástól elkülönülő osztályokba sorolhatók. Ezek heterogének és különböző (a csoport tagjaira vonatkozó)  $\beta$  paraméterekkel rendelkeznek (Boxall–Adamowicz [2002]). Az LC-modell előnyös tulajdonsága az RPL-modellekkel szemben, hogy nem szükséges feltételezéseket tenni a heterogenitás formájára vonatkozóan (Savolainen [2016]). Az  $i$ -edik alternatíva választásának valószínűsége az  $n$ -edik személyre vonatkozóan, aki a  $q$ -adik osztályhoz tartozik, a /6/ alapján írható fel (Wedel–Kamakura [1999], Morey–Thacher–Breffle [2006]).

$$Prob_{i,n|q} = \frac{\exp(\beta_q X_{i,n})}{\sum_{i=1}^{I_n} \exp(\beta_q X_{i,n})} \quad q=1, \dots, Q, \quad /6/$$

ahol  $\beta_q$  az  $X_{i,n}$  változóhoz tartozó paramétervektor.

Jól látható, hogy a választás valószínűsége /6/ az MNL-modellhez hasonló specifikáció szerint épül fel, viszont az LC-modell arra is előrejelzést tesz, mekkora eséllyel kerülnek az egyének a különböző osztályokba /7/. Ezzel kiegészítve /6/-t, a választás valószínűsége a /8/-ből adódik.

$$H_{n,q} = \frac{\exp(z_n \theta_q)}{\sum_{q=1}^Q \exp(z_n \theta_q)} \quad q=1, \dots, Q, \quad \theta_0 = 0, \quad /7/$$

ahol  $z_n$  a megfigyelhető jellemzők halmazát jelöli, a  $Q$ -adik paramétervektor pedig nullára van normalizálva.

$$Prob_{i,n} = \sum_{q=1}^Q Prob_{i,n|q} H_{n,q} \quad /8/$$

A modell korlátai közé tartozik, hogy az osztályok számának pontos meghatározása még sok kérdést hordoz magában. Többnyire az AIC (Akaike information criterion – Akaike-féle információs kritérium), illetve a CAIC (consistent AIC – konzisztens AIC) szerint szoktak erről döntést hozni, melyeket a log-likelihood érték alapján számolják. Ezek összefüggéseit a /9/ és a /10/ mutatja be (Bozdogan [1987], Ashok–Dillon–Yuan [2002]).

$$AIC = -2 \left[ LL(\hat{\beta}) - Q * K_Q (Q-1) K_C \right] \quad /9/$$

$$CAIC = -2LL(\hat{\beta}) - [Q * K_Q + (Q-1) K_C - 1] [\ln(2N) + 1], \quad /10/$$

ahol  $LL(\hat{\beta})$  a becsült paraméterekre vonatkozó log-likelihood,  $K_Q$  az elemek száma az osztályspecifikus választási modellek hasznossági funkciójában,  $K_C$  a paraméterek teljes száma az osztályozási modellben,  $N$  a mintában levő megfigyelések száma, a  $Q$  értéke pedig minimalizálja az AIC-t és CAIC-t.

Az információs kritériumok a mintán kívüli predikciós hiba értékét becsülik, ezáltal szolgáltatva információt arról, hogy az adott modell mennyire alkalmas az adathalmaz vizsgálatára. Az osztályszámok növelésének hatása az AIC és CAIC

értékére, iránymutatást jelenthet a megfelelő modellek felépítéséhez (*Cavanaugh–Neath* [2018]).

#### 1.4. Random paraméterű latent class modell

Az LC-modell kiterjesztése az RLC-modell, kombinálja az LC- és RPL-modell tulajdonságait (*Bujosa–Riera–Hicks* [2010]). Sajátossága abban rejlik, hogy lehetővé teszi a heterogenitás meglétét mind a csoportokon belül, mind pedig azok között. A csoportokon belüli heterogenitás ebben az esetben a /11/ és /12/ szerint alakul (*Greene–Hensher* [2013]).

$$\beta_{n|q} = \beta_q + w_{n|q} \quad /11/$$

$$w_{n|q} \sim E[w_{n|q} | X] = 0, \quad Var[w_{n|q} | X] = \sum q, \quad /12/$$

ahol  $X$  azt jelzi, hogy  $w_{n|q}$  nincs korrelációban a mintában szereplő adatok egyikével sem, és  $w_n$  a random (véletlen) vektort jelöli, míg  $q$  az adott csoportot (*Hensher–Rose–Greene* [2015]).

Az egyének hozzájárulása a log-likelihood értékéhez a csoporton belüli, majd pedig a csoportok közötti heterogenitás integrációján keresztül valósul meg.

A modellre vonatkozóan az  $n$ -edik személy  $I$  alternatívák között történő feltételes választásának valószínűségét a /13/ szemlélteti.

$$f[y_{n,t} | (\beta_q + w_n), X_{n,t}] = \frac{\exp\left[\sum_{i=1}^I y_{n,t,i} (\beta_q + w_n) X_{n,t,i}\right]}{\sum_{i=1}^I \exp\left[\sum_{i=1}^I y_{n,t,i} (\beta_q + w_n) X_{n,t,i}\right]}, \quad i=1, \dots, I, \quad /13/$$

ahol  $y_{n,t}$  az eredményvektorokat ( $y_{n,t,i} = 1$  amennyiben az  $i$  alternatívát választottuk, más esetben 0),  $X_{n,t,i}$  pedig az  $i$  alternatívára vonatkozó attribútumok vektorát jelöli az  $n$ -edik döntéshozó számára, a  $t$ -edik döntési szituációban (*Bujosa–Riera–Hicks* [2010]).

#### 1.5. Diszkrét választási modellek összehasonlítása

Az 1. táblázatban látható, hogy az ismertett modellek között jelentős különbségek vannak. Kivétel nélkül rendelkeznek előnyökkel, hátrányokkal, korlátozások-



kal és nehezen megválaszolható kérdésekkel. A szakirodalomban ezek közül legtöbb figyelmet kapó tényezőket az 1. táblázat mutatja be.

1. táblázat

*Diszkrét választási modellek előnyei, hátrányai és az alkalmazás problémái*  
(Advantages, disadvantages and application problems of discrete choice models)

Modell	Előnyök	Hátrányok, az alkalmazás problémái
MNL	<ul style="list-style-type: none"> <li>– Egyszerű becslés</li> <li>– Könnyedén értelmezhető eredmények</li> </ul>	<ul style="list-style-type: none"> <li>– Homogén preferenciák</li> <li>– IIA-feltevés</li> </ul>
RPL	<ul style="list-style-type: none"> <li>– Heterogenitás megragadása (egyének között)</li> <li>– IIA elkerülése</li> </ul>	<ul style="list-style-type: none"> <li>– Megfelelő eloszlás megválasztása</li> <li>– Random és fix paraméterek kijelölése</li> <li>– Véletlen minta generálása</li> </ul>
LC	<ul style="list-style-type: none"> <li>– Heterogenitás megragadása (csoportok között)</li> <li>– IIA elkerülése</li> </ul>	<ul style="list-style-type: none"> <li>– Csoportok ideális száma</li> <li>– Csoporton belüli homogén preferenciák feltételezése</li> </ul>
RLC	<ul style="list-style-type: none"> <li>– Heterogenitás megragadása (egyének és csoportok között)</li> <li>– IIA elkerülése</li> </ul>	<ul style="list-style-type: none"> <li>– Megfelelő eloszlás megválasztása</li> <li>– Random és fix paraméterek kijelölése</li> <li>– Véletlen minta generálása</li> <li>– Csoportok ideális száma</li> </ul>

*Megjegyzés.* Itt és a további táblázatokban: MNL (multinomial logit): multinomiális logit; RPL (random parameter logit): random paraméterű logit; LC (latent class): látens csoport; RLC (random parameter latent class): random paraméterű látens csoport; IIA (independence of irrelevant alternatives): irreleváns alternatívák függetlensége.

*Forrás:* Hess–Daly [2014].

Habár nyilvánvaló módon a választott modell komplexitásának növelésével egyidejűleg egyre jobban illeszkedő modellspecifikációt érhetünk el, az 1. táblázatban látható, hogy ehhez számos nehezen megválaszolható kérdésre kell helyes választ adnunk. A random paraméterekkel rendelkező modellek esetén az eloszlás és a véletlenminta-generálás módjának kiválasztása képviselnek kulcsfontosságú tényezőket, míg az LC-modelleknél a csoportok ideális számának meghatározása.

## 1.6. A minta bemutatása

A modellek összehasonlításához egy korábbi diszkrét választási kísérlet adatbázisát használtuk fel. A kutatás a Debreceni Egyetem Gazdaságtudományi Karának hallgatói körében zajlott, és a margarinra vonatkozó fogyasztói preferenciákat vizsgálta, unlabelled<sup>4</sup> formában. A minta 261 kitöltőt tartalmaz. Fontos említést tenni

<sup>4</sup> A felkínált alternatívák mindössze a tulajdonságok szintjeiben különböztek egymástól. Megnevezésük alternatíva 1, alternatíva 2 és alternatíva 3.

arról, hogy a mintavétel nem reprezentatív módon történt. A felmérés elsődleges célja a diszkrét választási kísérlet gyakorlatban történő alkalmazhatóságának vizsgálata volt, ezért az elemzés alapján levont következtetések sem általánosíthatók teljes mértékben. A kísérletbe bevont attribútumokat és azok szintjeit a 2. táblázat szemlélteti.

2. táblázat

*Attribútumok, azok szintjei és kódolásuk  
a diszkrét választási kísérletben*  
(Attributes, their levels and coding in the discrete choice experiment)

Attribútum	Attribútumszint	Kódolás
Ár (Ft)	350	Folytonos változó
	450	
	550	
Zsírtartalom(%)	< 31	1
	31–50	2
	50 <	3
Sótartalom (%)	< 0,51	1
	0,51–0,8	2
	0,8 <	3
Napraforgóolaj-tartalom	Tartalmaz	1
	Nem tartalmaz	0

*Forrás: Czine–Szakály–Balogh [2019].*

A kérdőív összeállítása során nem a full factorial elrendezést<sup>5</sup> alkalmaztuk, mivel az így kapott döntési helyzetek számát túl nagyra ítéltük meg. Helyette D-efficient elrendezést<sup>6</sup> választottunk, és a szituációk számának csökkentését az Ngene 1.2 szoftver segítségével végeztük el (*ChoiceMetrics* [2018]). Ezen kívül az R programban is van olyan csomag, amely képes különböző kísérleti elrendezések előállítására (például <https://cran.r-project.org/web/packages/AlgDesign/index.html>; <https://cran.rproject.org/web/packages/choiceDes/index.html>). Azért döntöttünk mégis az Ngene 1.2 program mellett, mivel a szerzőpáros közül *Balogh Péter* több diszkrét választási modellezéssel kapcsolatos nemzetközi kurzuson (Londonban, Sydney-ben) is részt vett már, melyek mindegyikén a külföldi neves előadók (*Stephane Hess, Michiel Bliemer, Thijs Dekker, William Greene, David Hensher, Andrew Collins*) ezt tartották leginkább alkalmasnak. A végső kérdőívben 8 darab döntési helyzet szerepelt, ezek mindegyike 3 eltérő termékből történő választási lehetőséget tartalmazott.

<sup>5</sup> Full factorial elrendezés során minden lehetséges kombináció bekerül a kérdőívbe.

<sup>6</sup> D-efficient elrendezés során úgy csökken a lehetséges kombinációk száma, hogy az elrendezési hibák (D-error) száma a lehető legalacsonyabb legyen.

A válaszadók ezek közül választottak, minden szituációban egyet. Szükséges említést tenni arról, hogy a kitöltők abban az esetben is döntöttek, amennyiben jelenleg nem fogyasztanak margarint (ekkor aszerint kellett választaniuk, melyik termék állna a legközelebb ahhoz, hogy megvásárolják). Erre a 3. táblázat mutat példát.

3. táblázat

*Példa a döntési szituációra*  
(An example of a decision situation)

Attribútum	Termék 1	Termék 2	Termék 3
Ár (Ft)	450	350	550
Zsírtartalom (%)	50	< 31	< 31
Sótartalom (%)	< 0,51	0,51–0,8	0,51–0,8
Napraforgóolaj-tartalom	Tartalmaz	Tartalmaz	Nem tartalmaz
Az Ön választása (X):			

*Forrás: Czine–Szakály–Balogh [2019].*

A kérdőívben szerepeltek még a termék vásárlásával és fogyasztásával kapcsolatos kérdések, valamint rákérdeztünk a kitöltők szociodemográfiai jellegű adataira is. Utóbbi eredményeit a 4. táblázat szemlélteti.

4. táblázat

*A válaszadók szociodemográfiai jellegű adatai (százalék)*  
(Sociodemographic data of respondents [percentage])

Változó	Minta (N = 261)
Nem	
Férfi	42,9
Nő	55,2
Nem válaszolt	1,9
Életkor (átlag)	21,4
Lakhely	
Község	11,1
Kisváros	16,5
Középváros	19,5
Nagyváros	52,5
Nem válaszolt	0,4

*(A táblázat folytatása a következő oldalon)*

(Folytatás)

Változó	Minta (N = 261)
Legmagasabb iskolai végzettség	
Érettségi	56,3
Érettségi és további képzés	42,9
Nem válaszolt	0,8
Havi nettó jövedelem a háztartásban (1 főre jutó)	
< 150 000 Ft	26,4
150 000–250 000 Ft	34,9
250 000–350 000 Ft	19,2
350 000 Ft <	18,0
Nem válaszolt	1,5
Családi állapot	
Egyedül élő	83,5
Élettárs/Házas	16,5

Forrás: Czine et al. [2019].

A kutatásról további részletek Czine–Szakály–Balogh [2019] tanulmányában találhatóak.

## 2. Eredmények

Az 5. táblázatban látható MNL-modell mellett – ahol a paraméterek mindegyike fixként szerepel – található a RP-modell becslései is. Utóbbi esetében a véletlenszerű paraméterek kiválasztása Hensher–Rose–Greene [2005] ajánlásai szerint történt. Először minden paraméter (az ár kivételével) véletlenként szerepelt, majd a végső specifikációban (amely az 5. táblázatban látható) már csak a szignifikáns szórással rendelkezők (közepes zsír-, magas zsír-, közepes só- és napraforgóolaj-tartalom) kaptak helyet randomként. Fontos említést tenni arról, hogy az árat minden esetben fixként kezeltük annak érdekében, hogy a későbbiek során pontbecsléseket tudjunk tenni a WTP-re (willingness to pay – fizetési hajlandóság). A véletlen paraméterek mindegyikét normál eloszlás mellett becsültük, 500 Halton-húzás alkalmazása mellett. Ugyancsak az 5. táblázatban láthatók a két osztállyal rendelkező LC-modell paraméterbecslései is.

5. táblázat

*Az MNL-, az RPL- és a kétosztályú LC-modellek becslései*  
(Estimates of MNL, RPL and two-class LC models)

Attribútum	MNL-modell	RPL-modell	LC-modell (kétosztályú)	
			<i>a</i>	<i>b</i>
			osztály	
ASC_alt2	0,206 (3,85)	0,217 (3,67)	0,178 (3,05)	
ASC_alt3	-0,246 (4,11)	-0,392 (-5,76)	-0,359 (-5,22)	
Ár	-0,002 (-5,44)	-0,003 (-6,99)	-0,001 (-1,05)	-0,003 (-6,35)
Közepes zsírtartalom	-0,236 (-2,42)	-0,341 (-2,90)	-1,059 (-4,42)	-0,089 (-0,69)
Magas zsírtartalom	-0,666 (-9,83)	-0,703 (-7,58)	-1,189 (-7,04)	-0,382 (-4,20)
Közepes sórtartalom	-0,240 (-3,36)	-0,286 (-3,31)	-1,366 (-5,87)	0,151 (1,48)
Magas sórtartalom	-0,416 (5,77)	-0,501 (-6,31)	-1,193 (-6,17)	-0,229 (-2,30)
Napraforgóolaj-tartalom	-0,045 (-0,83)	-0,062 (-0,62)	-1,453 (-7,66)	0,646 (6,55)
SD Közepes zsírtartalom		0,643 (3,18)		
SD Magas zsírtartalom		0,632 (4,85)		
SD Közepes sórtartalom		0,601 (4,59)		
SD Napraforgóolaj-tartalom		1,285 (12,34)		
Delta			-0,723 (-3,67)	
Osztályvalószínűségi érték			0,33	0,67
Megfigyelések	2 088			
Pseudo $R^2$	0,063	0,105	0,109	
Log-likelihood	-2 148,946	-2 052,081	-2 043,342	
AIC	4 313,89	4 128,16	4 116,68	

*Megjegyzés.* Itt, valamint a 7. táblázatban: ASC (alternative-specific constant): alternatíváspecifikus konstans; ASC\_alt1, alacsony zsír- és sórtartalom, nem tartalmaz napraforgóolajat, a *b* osztályra vonatkozó delta változók a bázisszintet jelentik a becslések során; delta az LC-modell osztályaira vonatkozó konstans értéket reprezentálja; a *t*-értékek a modellek esetében zárójelben szerepelnek; SD (standard deviation): szórás (a random paraméterekre).

*Forrás:* Saját szerkesztés a modellbecslések alapján.

Az 5. táblázatból jól látható, hogy a modellspecifikációk között több lényegi különbség is van. A pseudo  $R^2$  értéke az RPL- és LC-modellek esetében jelentősen nagyobb, mint az MNL-nél. Paraméterek tekintetében megfigyelhető, hogy az MNL- és RPL-modellek  $\beta$  értékei egészen közel vannak egymáshoz, míg az LC-modell osztályaira becsültek már nagyobb eltéréseket is mutatnak. Szembetűnő differenciák láthatók a paraméterek szignifikanciaértékeiben is. Míg az MNL- és RPL-modellek esetében a napraforgóolaj-tartalom nem tekinthető jelentős tényezőnek, addig az LC-modell mindkét osztályára vonatkozóan szignifikánsan szempontot képvisel ezen termékattribútum.

Az 5. táblázatban bemutatott modellekre vonatkozó WTP-becsléseket a 6. táblázat szemlélteti.

6. táblázat

A WTP-becslések eredményei az MNL-, az RPL- és a kétosztályú LC-modellek esetében  
(Results of WTP-estimates for MNL, RPL and two-class LC models)

Attribútum	MNL-modell	RPL-modell	LC-modell (kétosztályú)	
			<i>a</i>	<i>b</i>
			osztály	
Közepes zsírtartalom	-124,17*	-125,16* (236,26*)	-1 016,55	-28,92
Magas zsírtartalom	-351,17*	-258,22* (232,02*)	-1 141,33	-124,04*
Közepes sótartalom	-126,43*	-104,84* (220,54*)	-1 310,86	49,02
Magas sótartalom	-219,36*	-183,76*	-1 145,09	-74,13*
Napraforgóolaj-tartalom	-23,62	-22,84 (471,89*)	-1 394,74	209,39*

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

Megjegyzés. A szórásértékek az RPL-modell esetében zárójelben szerepelnek.

Forrás: Saját szerkesztés a modellbecslések alapján.

A 6. táblázatból jól kivehető, hogy az RPL-modell WTP-becslései (a közepes zsírtartalom kivételével) alacsonyabb értékeket mutatnak az MNL-hez képest. Látható továbbá, hogy az LC-modell osztályai közül az elsőben (*a*) rendkívül magas értékek születtek, viszont ezek egyike sem tekinthető szignifikánsnak. A második osztályban (*b*) már fellelhető három szignifikáns érték is. Ezek közül a magas zsír- és sótartalom kisebb értéket mutat a másik két modellspecifikáció becsléseihez képest,

míg a napraforgóolaj-tartalom lényegesen magasabbat és ellentétes előjelűt (viszont ez a szempont a másik két modell esetében nem tekinthető szignifikánsnak).

A 7. táblázatban további (három és négy osztályra vonatkozó) LC-modell becslései láthatók.

7. táblázat

Három- és négyosztályú LC-modellek becslései  
(Estimates of three- and four-class LC models)

Attribútum	LC-modell (háromosztályú)			LC-modell (négyosztályú)			
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
	osztály						
ASC_alt2	0,221 (3,55)			0,206 (3,25)			
ASC_alt3	-0,387 (-5,02)			-0,417 (-5,07)			
Ár	-0,002 (-1,50)	-0,020 (-3,67)	<0,001 (-0,02)	-0,002 (-1,63)	-0,020 (-3,56)	-0,001 (-1,08)	0,003 (1,78)
Közepes zsírtartalom	-1,012 (-3,85)	0,283 (0,48)	-0,193 (-1,27)	0,038 (0,11)	-0,163 (-0,18)	-1,243 (-4,42)	-0,282 (-0,89)
Magas zsírtartalom	-1,136 (-6,18)	-0,740 (-2,07)	-0,475 (-4,10)	-0,063 (-0,26)	-1,025 (-2,90)	-1,372 (-6,51)	-1,267 (-4,04)
Közepes sótartalom	-1,365 (-5,34)	0,003 (0,01)	-0,068 (-0,54)	-0,267 (-1,32)	0,110 (0,25)	-1,543 (-5,13)	0,037 (0,13)
Magas sótartalom	-1,165 (-5,49)	-0,196 (-0,46)	-0,290 (-2,60)	-0,044 (-0,18)	-0,468 (-0,85)	-1,455 (-6,29)	-0,569 (-2,26)
Napraforgóolaj-tartalom	-1,557 (-6,99)	-0,056 (-0,12)	0,669 (5,69)	-0,008 (-0,04)	0,446 (0,55)	-1,643 (-7,08)	1,189 (4,67)
Delta	0,603 (2,11)		1,076 (3,54)	0,666 (1,44)		0,515 (1,49)	0,336 (0,80)
Osztályvalószínűségi érték	0,32	0,17	0,51	0,32	0,17	0,28	0,23
Megfigyelések	2 088						
Pseudo $R^2$	0,136			0,144			
Log-likelihood	-1982,82			-1962,968			
AIC	4009,64			3983,94			
Likelihood-ratio-teszt	121,04* (2- 3 osztály modell)			39,7* (3- 4 osztály modell)			

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

Forrás: Saját szerkesztés a modellbecslések alapján.

Az 5. és a 7. táblázat pseudo  $R^2$  értékeiből megállapítható, hogy az osztályok számának növelésével egyidejűleg növekszik a modellek magyarázóereje. Amíg az 5. táblázatban látható kétosztályos modell esetében megközelítőleg 11 százalékot, addig a négy osztályt szerepeltetőnél már 14 százalékot mutat az érték. A likelihood-ratio-tesztek eredményei ugyanerre a következtetésre vezetnek, mivel szignifikáns javulás tapasztalható mind a kettőről hámosztályúra, mind pedig a háromról négyosztályúra növelt modellek között. Jól látható viszont az, hogy szignifikáns paraméterek terén a három- és négyosztályos esetek közül mindössze az előbbinél látható olyan attribútumszint (magas zsírtartalom), amely minden osztályra vonatkozóan jelentős szempontot képvisel. Mivel az AIC-érték további (például ötosztályos) esetekben már nem mutatott jelentősebb csökkenést, továbbá 10 százalék alatti osztályvalószínűségi érték is megjelent, a becslések során megállapítható, hogy a négy csoporttal rendelkező modell tekinthető leginkább alkalmasnak a kísérletbe bevont minta vizsgálatára (a csak konstans [delta] tagokat tartalmazó LC-modellek közül).

A 7. táblázatban bemutatott három- és négyosztályú LC-modellek WTP-re vonatkozó becslését a 8. táblázat szemlélteti.

8. táblázat

*A WTP-becslések eredményei a három- és négyosztályú LC-modellek esetében*  
(Results of WTP-estimates for three- and four-class LC models)

Attribútum	LC-modell (hámosztályú)			LC-modell (négyosztályú)			
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
	osztály						
Közepes zsírtartalom	-634,05	14,40	-12 236,11	16,98	-8,18	-904,44	91,60
Magas zsírtartalom	-712,09	-37,56*	-30 174,47	-28,34	-51,43	-998,16	411,32
Közepes sótartalom	-855,16	0,13	-4 310,17	-120,57	5,53	-1 122,53	-12,12
Magas sótartalom	-730,19	-9,96	-18 452,97	-20,00	-23,50	-1 058,46	184,66
Napraforgóolaj-tartalom	-976,09	-2,85	42 531,27	-3,78	22,40	-1 195,31	-386,13

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

Forrás: Saját szerkesztés a modellbecslések alapján.

A 8. táblázat becsléseiből azt a következtetést vonhatjuk le, hogy az osztályok számának növelésével egyidejűleg jelentősen lecsökkent (a négyosztályú modell esetében nem is található) a szignifikáns WTP-k száma. A hámosztályú modellben egyedül a *b* osztályra vonatkozó, magas zsírtartalom attribútumszint tekinthető szignifikánsnak. Ez a WTP 38 Ft-hoz közeli értéket mutat, amiből arra következtethetünk, hogy ezen osztály tagjai megközelítőleg ekkora összeggel fizetnének keveseb-



bet a magas zsírtartalmú (50 százalék feletti) margarinért, a bázisszintet jelentő alacsony (31 százalék alattival) szemben.

Annak érdekében, hogy a csoportok heterogenitásának forrását megtaláljuk, a 9. táblázatban levő modellekbe két szociodemográfiai jellegű változót (lakhely és nem) is beemeltünk. Továbbá a még pontosabb becslés eléréséhez az egyik modellben már random paramétereket is szerepeltettünk.

Következtetésünk – a 9. táblázat eredményeit összevetve a 7. táblázatban látottakkal –, hogy a háromosztályú modellek esetében nem hozott jelentős mértékű javulást a két szociodemográfiai jellegű változó beemelése a pseudo  $R^2$  mutató értékében, ám a négyosztályú modellnél ugyanez – kiegészítve a random paraméterekkel – már nagyobb hatást gyakorolt a magyarázóerőre. Emellett, összevetve a 9. táblázatban látható két modellt, a likelihood-ratio-teszt értéke is azt mutatja, hogy a random paraméterű modell szignifikánsan jobban magyaráz.

9. táblázat

*Szociodemográfiai változókat tartalmazó háromosztályú  
LC- és négyosztályú RLC-modellek becslései*  
(Estimates of three-class LC and four-class RLC models  
with sociodemographic variables)

Attribútum	LC-modell			RLC-modell			
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
	osztály						
ASC_alt2	0,224 (3,54)			0,268 (3,72)			
ASC_alt3	-0,381 (-4,92)			-0,568 (-6,24)			
Ár	-0,002 (-1,67)	-0,021 (-2,81)	-0,0001 (-0,09)	-0,024 (-4,32)	-0,022 (-2,61)	0,010 (0,00)	-0,0003 (-0,41)
Közepes zsírtartalom	-1,064 (-3,93)	0,194 (0,32)	-0,203 (-1,38)	0,704 (0,47)	1,834 (2,30)	-4,596 (0,00)	-0,242 (-1,32)
Magas zsírtartalom	-1,185 (-6,32)	-0,790 (-1,96)	-0,469 (-4,11)	-0,882 (-2,13)	-0,264 (-0,92)	-4,805 (-3,84)	-0,497 (-2,87)
Közepes sótartalom	-1,432 (-5,41)	-0,155 (-0,28)	-0,073 (-0,59)	0,111 (0,19)	-3,067 (-3,11)	-3,861 (-1,68)	-0,040 (-0,33)
Magas sótartalom	-1,208 (-5,69)	-0,305 (-0,68)	-0,294 (-2,72)	0,018 (0,02)	-0,414 (-1,30)	-3,784 (-3,46)	-0,221 (-1,97)
Napraforgóolaj-tartalom	-1,645 (-7,41)	-0,078 (-0,13)	0,633 (5,88)	-0,142 (-0,14)	-6,001 (-2,97)	0,173 (0,00)	0,528 (3,57)
SD közepes zsírtartalom				0,488 (0,60)	1,463 (1,63)	1,538 (0,43)	0,492 (1,15)

(A táblázat folytatása a következő oldalon)

(Folytatás)

Attribútum	LC-modell			RLC-modell			
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
	osztály						
SD magas zsírtartalom				1,104 (1,48)	0,086 (0,11)	0,928 (0,81)	0,689 (3,56)
SD közepes sőtartalom				1,367 (2,17)	0,027 (0,05)	2,386 (2,39)	0,051 (0,12)
SD napraforgóolaj- tartalom				0,893 (2,01)	1,471 (2,08)	0,863 (1,26)	0,971 (6,41)
Közép-/nagyváros	1,183 (2,31)		0,415 (0,97)	-3,346 (-4,71)		-0,746 (-1,48)	-2,548 (-3,02)
Nő	0,089 (0,20)		-0,467 (-1,11)	-1,001 (-1,16)		-0,714 (-0,77)	-1,327 (-1,48)
Delta	-0,353 (-0,59)		1,129 (2,26)	4,144 (1,00)		0,858 (4,36)	5,710 (2,30)
SD delta				5,120 (0,00)		0,380 (0,91)	3,803 (1,49)
Osztályvalószínűségi érték	0,30	0,17	0,53	0,17	0,18	0,10	0,55
Megfigyelések	2 088						
Pseudo $R^2$	0,138			0,163			
Log-likelihood	-1 978,024			-1 921,064			
AIC	4 008,05			3 950,13			
Likelihood-ratio-teszt	113,92*						

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

*Megjegyzés.* ASC (alternative-specific constant): alternatíváspecifikus konstans; ASC\_alt1, alacsony zsír- és sőtartalom, nem tartalmaz napraforgóolajat, a *b* osztályra vonatkozó delta, község/kisváros és a férfiak változók jelentik a bázisszintet a becslések során; delta az LC-modell osztályaira vonatkozó konstans értéket reprezentálja; a *t*-értékek a modellek esetében zárójelben szerepelnek; SD (standard deviation): szórás (a random paraméterekre).

*Forrás:* Saját szerkesztés a modellbecslések alapján.

Az 9. táblázatban bemutatott szociodemográfiai változókat és a random paramétereket is tartalmazó LC-modellek WTP-re vonatkozó becsléseit a 10. táblázat szemlélteti.

10. táblázat

*A WTP-becslések eredményei a szociodemográfiai változókat tartalmazó LC és RLC modellek esetében*  
(Results of WTP-estimates for LC and RLC models with sociodemographic variables)

Attribútum	LC-modell			RLC-modell			
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
	osztály						
Közepes zsírtartalom	-577,09	9,25	-3 069,17	29,38 (20,38)	84,02 (67,04)	469,68 (157,20)	-705,82 (1 433,27)
Magas zsírtartalom	-643,17	-37,60*	-7 102,37	-36,82 (46,09)	-12,09 (3,92)	490,95 (94,82)	-1448,52 (2 006,10)
Közepes sótartalom	-777,22	-7,40	-1 104,20	4,62 (57,11)	-140,54* (1,24)	394,53 (243,84)	-117,08 (147,18)
Magas sótartalom	-655,34	-14,49	-4 456,43	0,74	-18,96	386,64	-643,82
Napraforgóolaj-tartalom	-892,73	-3,71	9 587,24	-5,92 (37,30)	-274,96* (67,42)	-17,72 (88,14)	1 537,47 (2 828,60)

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

Megjegyzés. A szórásértékek az RLC-modell esetében zárójelben szerepelnek.

Forrás: Saját szerkesztés a modellbecslések alapján.

Arra következtetünk a 10. táblázat eredményeiből, hogy a háromosztályú modell esetében a szociodemográfiai változók beemelése nem hozott jelentősebb változásokat a WTP-re vonatkozóan. Jól látható, hogy a 8. táblázatban bemutatotthoz hasonlóan, mindössze a *b* osztály magas zsírtartalomra vonatkozó WTP-érték tekinthető szignifikánsnak. A random paramétereket is tartalmazó modell esetében már két érték (a *b* osztályra vonatkozó só- és napraforgóolaj-tartalom) is szignifikáns. Az eredmények alapján azt mondhatjuk, hogy a háromosztályú modell esetében a *b* osztály tagjai megközelítőleg 38 Ft-tal fizetnének kevesebbet a magas (50 százaléknál feletti), szemben a bázisszintet képviselő alacsony (31 százaléknál alatti) zsírtartalmú margarinért. Emellett a négyosztályú, random paraméteres modell esetében ugyanezen osztály tagjai hozzávetőleg 141 Ft-tal fizetnének kevesebbet a közepes (0,51–0,8%), szemben az alacsony (0,51 százaléknál alatti) sótartalmú margarinért; továbbá körülbelül 275 Ft-tal fizetnének kevesebbet a napraforgóolajat tartalmazó termékért, az azt nélkülözővel szemben.

### 3. Összegzés

Jelen tanulmányban két, a hazai szakirodalomban még kevésbé tárgyalt diszkrét választási modell (LC, RLC) bemutatását, majd azok eredményeinek összevetését tűztük ki célul. Kutatásunkhoz egy margarinfogasztásra vonatkozó preferenciákat vizsgáló kísérlet adatait használtuk fel példaként.

A kapott eredményekből levonhatjuk azt a következtetést (ezeket fenntartással kezelve, kizárólag jelen mintára vonatkoztatva), hogy az MNL-hez képest mind az RPL-, mind pedig az LC-modellek becslései jobbnak bizonyulnak. Fontos említést tenni arról, hogy a paraméterértékek és a szignifikanciaértékek tekintetében az RPL eredményei közelebb esnek az MNL- és távolabb az LC-modellek becsléseitől. Az LC-modellek eredményeiből azt a konzekvenciát vonhatjuk le, hogy az osztályok számának növelésével egyidejűleg javul a modellek magyarázóereje egy bizonyos szintig. A vizsgálatba bevont mintára vonatkozóan a leginkább hatékonynak a négyosztályú modell tekinthető, amit a pseudo  $R^2$  értékek, valamint a likelihood-ratio tesztek eredményei is igazolnak. Fontos említést tenni arról, hogy az osztályok számának növelése, a szignifikáns paraméterek csökkenésével jár együtt. A háromosztályú LC-modell szociodemográfiai változókkal történő bővítése jelentősebb mértékű változást nem mutatott. A random paraméterek alkalmazása a négyosztályú modellbe viszont már számottevő javulást eredményezett. A WTP-re vonatkozó becslésekből megállapítható, hogy az RPL-modell értékei kisebb értékeket mutatnak az MNL-hez viszonyítva. Továbbá az LC-modellek esetében az osztályok számának növelésével egyidejűleg csökken a szignifikáns WTP-értékek száma. A kapott eredmények alapján azt mondhatjuk, hogy a megfelelően felépített LC-modellek ígéretes lehetőséget nyújthatnak a diszkrét választási kísérletek elemzéséhez.

### Irodalom

- ASHOK, K. – DILLON, W. R. – YUAN, S. [2002]: Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research*. Vol. 39. Issue 1. pp. 31–46. <http://dx.doi.org/10.1509/jmkr.39.1.31.18937>
- BAJI P. [2012]: A diszkrét választás módszere. *Statistikai Szemle*. 90. évf. 10. sz. 943–963. old.
- BALOGH, P. – BÉKÉSI, D. – GORTON, M. – POPP, J. – LENGYEL, P. [2016]: Consumer willingness to pay for traditional food products. *Food Policy*. Vol. 61. pp. 176–184. <http://dx.doi.org/10.1016/j.foodpol.2016.03.005>
- BALOGH P. [2017]: *A sertéshús-előállítás és -fogyasztás gazdasági elemzése*. MTA doktori értekezés. Debrecen. [http://real-d.mtak.hu/1045/6/dc\\_1399\\_17\\_doktori\\_mu.pdf](http://real-d.mtak.hu/1045/6/dc_1399_17_doktori_mu.pdf)
- BOXALL, P. C. – ADAMOWICZ, W. L. [2002]: Understanding heterogeneous preferences in random utility models: A latent class approach. *Environmental and Resource Economics*. Vol. 23. No. 4. pp. 421–446.

- BOZDOGAN, H. [1987]: Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*. Vol. 52. No. 3. pp. 345–370. <http://dx.doi.org/10.1007/BF02294361>
- BRANDTMÜLLER Á. [2009]: Diszkrét választási kísérlet magyar háziorvosok körében. *Statisztikai Szemle*. 87. évf. 12. sz. 1153–1174. old.
- BUJOSA, A. – RIERA, A. – HICKS, R. L. [2010]: Combining discrete and continuous representations of preference heterogeneity: a latent class approach. *Environmental and Resource Economics*. Vol. 47. pp. 477–493. <http://dx.doi.org/10.1007/s10640-010-9389-y>
- CAVANAUGH, J. E. – NEATH, A. A. [2019]: The Akaike information criterion: background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*. Vol. 11. Issue 3. pp. e1460. <http://dx.doi.org/10.1002/wics.1460>
- CHANG, J. B. – LUSK, J. L. [2011]: Mixed logit models: accuracy and software choice. *Journal of Applied Econometrics*. Vol. 26. No. 1. pp. 167–172. <http://dx.doi.org/10.1002/jae.1201>
- CHOICEMETRICS [2018]: *Ngene 1.2 User Manual & Reference Guide*. ChoiceMetrics Pty Ltd. Sydney.
- CZINE P. – SZAKÁLY Z. – BALOGH P. [2019]: Margarinnal kapcsolatos preferenciák vizsgálata egyetemista fogyasztók körében. *Táplálkozásmarketing*. 6. évf. 2. sz. 3–12. old. <http://dx.doi.org/10.20494/TM/6/2/1>
- FIEBIG, D. G. – KEANE, M. P. – LOUVIERE, J. – WASI, N. [2010]: The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science*. Vol. 29. Issue 3. pp. 393–421. <http://dx.doi.org/10.1287/mksc.1090.0508>
- FOSGERAU, M. – BIERLAIRE, M. [2007]: A practical test for the choice of mixing distribution in discrete choice models. *Transportation Research Part B: Methodological*. Vol. 41. Issue 7. pp. 784–794. <http://dx.doi.org/10.1016/j.trb.2007.01.002>
- GREENE, W. H. – HENSHER, D. A. [2003]: A latent class model for discrete choice analysis: contrast with mixed logit. *Transportation Research Part B: Methodological*. Vol. 37. Issue 8. pp. 681–698. [http://dx.doi.org/10.1016/S0191-2615\(02\)00046-2](http://dx.doi.org/10.1016/S0191-2615(02)00046-2)
- GREENE, W. H. – HENSHER, D. A. [2010]: Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation*. Vol. 37. Issue 3. pp. 413–428. <http://dx.doi.org/10.1007/s11116-010-9259-z>
- GREENE, W. H. – HENSHER, D. A. [2013]: Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics*. Vol. 45. Issue 14. pp. 1897–1902. <http://dx.doi.org/10.1080/00036846.2011.650325>
- 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. [2008]: Combining RP and SP data: biases in using the nested logit “trick” – contrasts with flexible mixed logit incorporating panel and scale effects. *Journal of Transport Geography*. Vol. 16. Issue 2. pp. 126–133. <http://dx.doi.org/10.1016/j.jtrangeo.2007.07.001>
- 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. Vol. 32. *Journal of Choice Modelling*. <http://dx.doi.org/10.1016/j.joem.2019.100170>
- HESS, S. – PALMA, D. [2019b]: *Apollo version 0.0.6, user manual*. [www.ApolloChoiceModelling.com](http://www.ApolloChoiceModelling.com)
- HOLE, A. R. – KOLSTAD, J. R. [2011]: Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. *Empirical Economics*. Vol. 42. Issue 2. pp. 445–469. <http://dx.doi.org/10.1007/s00181-011-0500-1>
- KAMAKURA, W. A. – RUSSELL, G. J. [1989]: A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*. Vol. 26. No. 4. pp. 379–390. <http://dx.doi.org/10.2307/3172759>
- LANCSAR, E. – FIEBIG, D. G. – HOLE, A. R. [2017]: Discrete choice experiments: a guide to model specification, estimation and software. *PharmacoEconomics*. Vol. 35. No. 7. pp. 697–716. <http://dx.doi.org/10.1007/s40273-017-0506-4>
- McFADDEN, D. [1973]: Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (ed.): *Frontiers in Econometrics*. Academic Press. New York. pp. 105–142.
- MOREY, E. – THACHER, J. – BREFFLE, W. [2006]: Using angler characteristics and attitudinal data to identify environmental preference classes: a latent-class model. *Environmental and Resource Economics*. Vol. 34. No. 1. pp. 91–115. <http://dx.doi.org/10.1007/s10640-005-3794-7>
- SAVOLAINEN, P. T. [2016]: Examining driver behavior at the onset of yellow in a traffic simulator environment: comparisons between random parameters and latent class logit models. *Accident Analysis and Prevention*. Vol. 96. pp. 300–307. <http://dx.doi.org/10.1016/j.aap.2016.01.006>
- SHEN, J. [2009]: Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics*. Vol. 41. Issue 22. pp. 2915–2924. <http://dx.doi.org/10.1080/00036840801964633>
- STATA CORP LLC [2019]: *Stata Choice Models Reference, Manual Release 16*. Stata Press Publication. Texas.
- TRAIN, K. E. [2003]: *Discrete Choice Methods with Simulation*. Cambridge University Press. Cambridge. <http://dx.doi.org/10.1017/CBO9780511753930>
- WEDEL, M. – KAMAKURA, W. A. [1999]: *Market Segmentation: Concepts and Methodological Foundations*. Second Edition. Springer-Verlag. New York.