

EÖTVÖS LORÁND TUDOMÁNYEGYETEM  
BUDAPESTI CORVINUS EGYETEM

---

Horvat Anna

**BIZTOSÍTÁSI TERMÉKEK SZIMULÁCIÓS  
MODELLEZÉSE**

Szakdolgozat

MSc Biztosítási és Pénzügyi Matematika - Aktuárius Szakirány

Témavezető:

Vékás Péter

Operációkutatás és Aktuáriustudományok Tanszék



Budapest, 2019

# Tartalomjegyzék

<b>Bevezetés</b>	<b>3</b>
<b>1. Elméleti háttér</b>	<b>5</b>
1.1. Ritkán bekövetkező események modellezése logisztikus regresszióval . . . . .	5
1.1.1. Logit modell . . . . .	6
1.1.2. Mintavételezés . . . . .	8
1.1.3. Korrekciók . . . . .	9
1.2. Korrelált rendezett és diszkrét valószínűségi változók szimulációja ismert peremeloszlások mellett . . . . .	11
1.2.1. Ordinális változók korrelációja . . . . .	12
1.2.2. Az eljárás . . . . .	13
1.2.3. Kiterjesztések . . . . .	14
<b>2. Alkalmazás - biztosítási állomány tervezése</b>	<b>16</b>
2.1. Adatbázis bemutatása . . . . .	16
2.1.1. Ügyfél személyét jellemző változók . . . . .	18
2.1.2. A biztosítási kötvényt jellemző adatok . . . . .	21
2.1.3. A biztosított gépjárművet jellemző adatok . . . . .	25
2.1.4. A változók és törlések kapcsolatának előzetes vizsgálata . . . . .	27
2.2. Állomány mozgás előrejelzése . . . . .	28
2.2.1. Évfordulók szimulációja . . . . .	28
<b>Összegzés</b>	<b>41</b>

# Bevezetés

A biztosítók és pénzügyintézetek sok szempontból különböznek a más szférába eső vállalatoktól. A biztosítók tevékenységükből fakadóan másképpen működnek és másképpen vannak szabályozva (Szolvencia I-II, Bit., IFRS 17), hiszen működésükre halmozottan hatással van a véletlen. Ilyen véletlenszerű események a károk bekövetkezései és a kifizetések mértéke, de ugyanúgy a hozamok értéke és az infláció is. Ezen sztochasztikus feléltelezéseket fontos a biztosítóknak beépíteniük modelljeikbe, hogy megfelelően árazzák termékeiket, és olyan kockázatokat vállaljanak, melyeket tudnak teljesíteni, és emellett még némi profitra is szert tegyenek.

A szakdolgozatban egy KGFB állomány alakulásának előrejelzésével foglalkoztam, sztochasztikus és matematikai, statisztikai eszközök segítségével. A célom az volt, hogy minden évforduló végére a felmondott szerződések darabszámát előre lehessen jelezni a tapasztalatok alapján. Egy biztosító számára egy ilyen kalkuláció kiemelkedően fontos lehet, hiszen ez alapján lehetőség nyílik a következő év stragégiai tervezhetőségére, díjak kalkulációjára és tartalékolásra is.

A két legfontosabb feladat a kockázatban lévő szerződések törlésének előrejelzése, valamint az állomány új szerződéseinek szimulációja volt. Az ezekhez kapcsolódó matematikai háttérrel foglaltam össze az 1. fejezetben.

A törlések előrejelzéséhez logisztikus regressziót alkalmaztam, ami kézenfekvőnek tűnt, tekintve, hogy ismert volt szerződésenként, hogy élő-e még a szerződő biztosítási kötvénye az évforduló után. A modellezések során nehézséget okozott, hogy alulreprezentáltak voltak a törlésre vonatkozó megfigyelések. A probléma megoldására King és Zeng korrekcióit [7] alkalmaztam, ennek áttekintését mutatom be az 1.1 alfejezetben.

Az újonnan érkező szerződések szimulációja során összefüggő diszkrét változók segítségével új mintákat generáltam, melynek matematikai hátterét az 1.2 alfejezetben fejtem ki bővebben. A Barbiero és Ferrari [3] által kidolgozott eljárást mutatja be a fejezet, mely elsősorban ordinális változók szimulációjával foglalkozik, de általánosan bármely diszkrét változóra alkalmazható.

A 2. fejezetben egy valós biztosító adatán az 1. fejezetben bemutatott eszközök segítségével készítettem 3 éves előrejelzést az állomány alakulásáról. A fejezet elején az adatbázis felépítése, majd az évfordulók szimulációja kerül bemutatásra. A szimulációk R programozási környezetben készültek.

# 1. fejezet

## Elméleti háttér

### 1.1. Ritkán bekövetkező események modellezése logisztikus regresszióval

A logisztikus regresszió alkalmazása során a megfigyeléseinket előre meghatározott, egymást kölcsönösen kizáró osztályok valamelyikébe soroljuk be a magyarázó változókból származó ismeretek alapján. Az  $Y$  eredményváltozó tehát nem folytonos, hanem egy  $0 - 1$  értékű bináris változó, ami kifejezheti, hogy egy személy rákos beteg-e (1) vagy nem (0), csődkockázati szempontól kockázatos-e egy hitelt felvevő ügyfél (1) vagy nem (0), vagy mint a szakdolgozat 2. fejezetében is látható, hogy az ügyfél felmondta-e évfordulóra a tárgyalt biztosítónál a szerződését (1) vagy sem (0). A logisztikus regresszió az egyik leggyakrabban alkalmazott klasszifikációs módszer, előnyei közé tartozik többek között, hogy nem túlságosan számításigényes, illetve jól, könnyen értelmezhető, és egyszerűsége mellett is hatékony módszer. A paraméterek becslése jellemzően maximum likelihood becsléssel történik, amely nagy minta esetén konzisztens és teljesül a minimum variancia tulajdonság is, azonban kis minta nagyság vagy kiegyensúlyozatlan minták esetén, tehát ahol az 1 vagy 0 megfigyelések aránya alacsony, a megoldás már magas mintavételi varianciával rendelkezik és torzított is. Fontos továbbá a megfelelő mintavételi stratégia megtalálása is ritka jelenségek elemzése során. A gyakorlatban az adatok gyűjtésekor fennállhat az a

kockázat, hogy az adatok nem tartalmaznak 1-es egyedeket, így gyakran a lehető legtöbb megfigyelést összegyűjtik az elemzők, amely az adat minőségének hanyatlásával, továbbá kevesebb információ tartalommal járhat. Léteznek azonban ezeknél hatásosabb stratégiák is, amelyek kedvező kompromisszumot biztosítanak a jobb minőségű változók és a több megfigyelés gyűjtése között. Alapvetően elmondható, hogy az adat információtartalma elsősorban az 1-ekben található, és így érdemes például a függő változó szerinti mintavételezést alkalmazni. Torzítást okozhat azonban az, hogy ha a mintánkban és a sokaságban az 1-es egyedek aránya lényegesen különböznek, mely probléma megoldásához további korrekciók szükségesek. A fejezetben bemutatott King-Zeng [7] korrekciók összekötik a ritka jelenségek elemzésének ezen nehézségeit, és közös megoldást kínálnak.

### 1.1.1. Logit modell

A logisztikus regressziós modellnek több ekvivalens megfogalmazása is létezik. Az első megközelítéshez ([4],[5],[7],[8]) tekintsük a független  $Y_1, Y_2, \dots, Y_n$  változók sorozatát, ahol  $Y_i$  vagy az 1 vagy a 0 értéket veheti fel minden  $i = 1, 2, \dots, n$  esetén. Jelölje  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$  a magyarázó változókat, ahol  $\mathbf{x}_k$  a  $k$ -adik magyarázó változó  $n$  elemű oszlopvektorral. Jelölje  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)$  a megfelelő együtthatókat, amelyek a logisztikus regresszió paramétereire, ahol  $\beta_0$  egy konstans skalár. Jelölje  $p_{\mathbf{x}}$  a  $P(Y = 1 \mid \mathbf{x})$  feltételes valószínűséget, azaz az 1-es csoportba tartozás valószínűségét  $\mathbf{x}$  magyarázó változók mellett. Az odds-arányból, azaz az esemény bekövetkezési valószínűségének és be nem következés valószínűségének hányadosából kifejezhető a  $p_{\mathbf{x}} = \frac{p_{\mathbf{x}}/(1-p_{\mathbf{x}})}{1+p_{\mathbf{x}}/(1-p_{\mathbf{x}})} = \frac{odds_{\mathbf{x}}}{1+odds_{\mathbf{x}}}$ . Ezután alkalmazzuk a logit-transzformációt: a  $p_{\mathbf{x}}$  logitja az odds-arány logaritmus, azaz  $\ln(odds_{\mathbf{x}}) = \text{logit}(p_{\mathbf{x}})$ . Ezt modellezzük lineárisan:  $\text{logit}(p_{\mathbf{x}}) = \mathbf{x}^T \boldsymbol{\beta}$ , amelyből következik, hogy

$$p_{\mathbf{x}} = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}^T \boldsymbol{\beta})} = \frac{1}{1 + \exp(-\mathbf{x}^T \boldsymbol{\beta})} \quad (1.1)$$

Egy  $y_1, y_2, \dots, y_n$  független, véletlen mintán becsüljük a  $\boldsymbol{\beta}$ -t, és ezután ezt felhasználva jelezük előre  $Y$  értékét úgy, hogy meghatározunk egy alkalmas küszöbértéket, amelyet jelöljünk  $c$ -vel. Ha ez az előre meghatározott  $c$  érték nagyobb, mint a  $p_{\mathbf{x}}$  feltételes valószínűség, akkor az előrejelzett  $\hat{Y}$  értéke 0, különben 1.

A ritka jelenségek problémáinak vizsgálatához tekintünk a látens változón alapuló megközelítést ([4],[5],[7]), amely ekvivalens az előbb bemutatott származtatási móddal. Tekintsük  $Y^*$  látens eredményváltozót, amely egy folytonos, nem megfigyelhető változó, mint például egy egyén egészsége. Az  $Y^*$  várható értékét  $\mathbf{x}$  feltétel mellett jelölje  $\eta_x$ . Ezen megközelítés során logisztikus eloszlás szerint fogunk döntést hozni. A logisztikus eloszlás sűrűségfüggvénye:

$$\text{Logistic}(Y^* | \eta_x) = \frac{\exp(-(Y^* - \eta_x))}{(1 + \exp(-(Y^* - \eta_x)))^2},$$

ahol az  $\eta_x$  az eredeti modell magyarázó változói által lineáris regresszióval kifejezhető, azaz  $\eta_x = \mathbf{x}^\top \boldsymbol{\beta}$ .

Tekintsük a következő diszkrétizálást: ha az  $Y^* > 0$ , akkor az eredeti modell bináris változója 1 lesz, amennyiben pedig  $Y^* \leq 0$ , az eredeti modell bináris változója 0 lesz. Ekkor

$$P(Y^* > 0) = \int_0^\infty \text{Logistic}(Y^* | \eta_x) dY^* = \frac{1}{1 + \exp(-\eta_x)} = \frac{\exp(\eta_x)}{1 + \exp(\eta_x)},$$

ami megegyezik a fentebb bemutatott (1.1) kifejezéssel.

A modell paraméterbecslése történhet maximum likelihood elv alapján vagy a legkisebb négyzetek elvének segítségével is. Jelen esetben az előbbit alkalmazzuk egy  $y_1, y_2, \dots, y_n$  független mintát véve. Ekkor a minta együttes likelihood függvénye a következőképpen írható fel:

$$L = P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \prod_{\{y_i=1\}} p_i \prod_{\{y_i=0\}} (1 - p_i) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i},$$

amelyből a log-likelihood a logit modell behelyettesítésével [7]:

$$\ln L = \sum_{\{y_i=1\}} \ln(p_i) + \sum_{\{y_i=0\}} \ln(1 - p_i) = - \sum_{i=1}^n \ln(1 + \exp((1 - 2y_i)\mathbf{x}_i^\top \boldsymbol{\beta}))$$

A  $\boldsymbol{\beta}$  szerinti maximalizálási feladat megoldását jelöljük  $\hat{\boldsymbol{\beta}}$ -val. A becslt paraméterek aszimptotikus varianca-kovarianca mátrixa a következő:

$$\mathbf{Var}(\hat{\boldsymbol{\beta}}) = \left[ \sum_{i=1}^n p_i(1 - p_i)\mathbf{x}_i\mathbf{x}_i^\top \right]^{-1} \quad (1.2)$$

Az 1.2 egyúttal a Fisher-féle információs mátrix inverze is. A variancia-kovariancia mátrixot vizsgálva látható, hogy ritka jelenségek vizsgálatakor az 1 esetek több információt hordoznak a 0 eseteknél, ugyanis annak ellenére, hogy  $p_i$  becslt értéke kicsi lesz, a  $p_i$  a bekövetkező jelenségek (azaz  $Y_i = 1$ ) mellett magasabb értékkel rendelkezik, mint nem bekövetkező jelenségek (azaz  $Y_i = 0$ ) esetén (ami még így is csak 0.5-höz közeli, hiszen a bekövetkezési valószínűségek ritka jelenségek esetén alacsonyak). Ebből az következik, hogy 1-esek esetén a  $p_i(1 - p_i)$  magasabb lesz, az információs mátrix értéke nagyobb és a becslt paraméterek varianciája kisebb lesz, a standard hiba csökken újabb 1 esetek bevonásával [7]. Ennek megfelelő mintavételi stratégiákat illetve azok korrekcióit tekintsük át továbbiakban.

### 1.1.2. Mintavételezés

A minták két csoportját különböztetjük meg rétegzésük alapján: az exogén minta a magyarázó változón keresztül rétegzett, az endogén minta pedig a függő változón keresztül rétegzett minta [4]. Ritka jelenségek vizsgálata során célszerű az utóbbi stratégiát alkalmazni, ugyanis ahogy korábban a variancia-kovariancia mátrix vizsgálatából is láttuk, további 1-es esetek mintába való választása csökkenti a variációt és növeli az információt, így érdemes minél több vagy az összes elérhető 1-es megfigyelést csatolni a mintához. Ekkor felmerül a kérdés, hogy a 0 esetek közül mennyit válasszunk be a mintába: általánosságban elmondható, hogy a 0 és 1 megfigyelések kb. azonos aránya optimális a mintában [7]. Endogén (vagy más néven eset-kontroll) mintavételezés során szükségeszerű odafigyelni a következőkre:

- a megfigyelések független kiválasztása szükséges mind az  $Y = 1$  és  $Y = 0$  esetekben
- a két (függő változó szerinti) mintavételezés során  $\mathbf{x}$  mentén ugyanazon szempontok szerint kell kiválasztani a megfigyeléseket, a magyarázó változókat ugyanazon módon szabad válogatni a két mintán

Diszkrét döntési modellek esetén exogén mintavételezés esetén nincs szükség a becslési módszer korrigálására, míg endogén minta esetén igen. Logit modellek esetén elegendő a



konstans tag becslését módosítani, a többi paraméter maximum likelihood becslése konzisztens lesz [4]. A valószínűségek precíz számításának szempontjából fontos a regressziós paraméterek pontos becslése, amelynek egyik lehetséges módja a Prior korrekció.

A Prior korrekció [10], [13] a maximum likelihood becslést korigálja felhasználva azt az előzetes információt, hogy milyen arányban fordulnak elő az 1-es egyedek a sokaságban ( $\tau$ ) és a mintában ( $\bar{y}$ ). Ezen információk tudatában  $\beta_0$  becslése a következőképpen módosul:

$$\hat{\beta}_0 - \ln \left[ \left( \frac{1 - \tau}{\tau} \right) \left( \frac{\bar{y}}{1 - \bar{y}} \right) \right] = \hat{\beta}_0 - \left[ \ln \left( \frac{\bar{y}}{1 - \bar{y}} \right) - \ln \left( \frac{\tau}{1 - \tau} \right) \right], \quad (1.3)$$

amely így már konzisztens becslése  $\beta_0$ -nak. Az 1.3 alapján amennyiben magyarázó változók nincsenek bevonva a modellbe, akkor az odds hányados a sokasági odds aránnyal egyezik meg [5]. A módszer előnye, hogy könnyen alkalmazható.

A súlyozott exogén mintán alapuló maximum likelihood becslés [10] egy másik módszer a sokaságbeli és mintabeli 1-es egyedek arányának kompenzálására. A módszer alkalmazása során a következő súlyozott loglikelihood függvényt maximalizáljuk:

$$\ln L_w(\boldsymbol{\beta} | \mathbf{y}) = w_1 \sum_{\{Y_i=1\}} \ln(p_i) + w_0 \sum_{\{Y_i=0\}} \ln(1 - p_i) \quad (1.4)$$

$$= - \sum_{i=1}^n w_i \ln(1 + \exp((1 - 2y_i)\mathbf{x}_i^\top \boldsymbol{\beta})), \quad (1.5)$$

ahol  $w_i = w_1 Y_i + w_0(1 - Y_i)$ , feltéve, hogy  $w_1 = \frac{\tau}{\bar{y}}$  és  $w_0 = \frac{1 - \tau}{1 - \bar{y}}$ .

Mivel a gyakorlatban a modellek téves specifikációja gyakori probléma, ezért előnyösebb lehet a súlyozás módszerét alkalmazni a prior korrekcióval szemben, mert az ebben az esetben robusztusabb [15]. A módszer egyik hátránya, hogy a standard hiba számítás jelentősen torzított, másrészt szükség van a ritka jelenségek miatti korigálására is. A [7] szerinti korrekciók erre a két problémára kínálnak megoldást, amelyek alkalmazásával a módszer előnyösebb lehet az 1-es egyedek sokasági arányának ismerete mellett.

### 1.1.3. Korrekciók

A  $\boldsymbol{\beta}$  maximum likelihood becslése alapvetően torzított bármely kiegyensúlyozott véges mintában, amelynek mértéke a mintaméret növekedésével csökken, és kb. 200-as minta-

méret fölött a mértéke közel jelentéktelenné válik [12]. A torzítás mértéke a következő módon kifejezhető [7]:

$$B(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \boldsymbol{\xi}, \quad (1.6)$$

ahol  $\xi_i = 0.5Q_{ii}[(1 + w_1)\hat{p}_i - w_1]$ ,  $\mathbf{Q} = \mathbf{X}(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top$ ,  $\mathbf{W} = \text{diag}\{\hat{p}_i(1 - \hat{p}_i)w_i\}$ .

Ahhoz tehát, hogy a  $\hat{\boldsymbol{\beta}}$  közel torzítatlan legyen, ki kell vonnunk belőle a torzítás mértékét, azaz

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - B(\hat{\boldsymbol{\beta}})$$

A módszer egyúttal a varianciát is csökkenti, ugyanis [11] alapján a  $V(\tilde{\boldsymbol{\beta}})$  és  $V(\hat{\boldsymbol{\beta}})$  között a következő összefüggés írható fel:

$$V(\tilde{\boldsymbol{\beta}}) = \left(\frac{n}{n+k}\right)^2 V(\hat{\boldsymbol{\beta}}),$$

és mivel  $\left(\frac{n}{n+k}\right)^2$  1-nél kisebb szám, így  $V(\tilde{\boldsymbol{\beta}}) < V(\hat{\boldsymbol{\beta}})$  is teljesül.

Logit modellek esetén elsősorban az esély hányados illetve a bekövetkezés valószínűség megfelelő becslésén van a hangsúly, amihez szükség van a paraméterek pontos becslésére, azonban torzítatlan paraméterek esetén is torzítottak lesznek ezek az értékek, ugyanis amennyiben az 1-es egyedek alulreprezentáltak, akkor a bekövetkezés valószínűsége alulbecsült lesz. Ezt szemlélteti, hogy az  $Y^*$  látens változó tekintetében a sokaság eloszlását leíró logisztikus sűrűségfüggvény szórása nagyobb, mint a mintavétel során eredményként kapott torzítatlan  $\tilde{\boldsymbol{\beta}}$ -k által generált eloszlás szórása, azaz  $P(Y = 1 | \tilde{\boldsymbol{\beta}})$  alulbecsült lesz.

A ritka jelenségek korrigálásához [7] tekintsük  $p_i$ -t bayes-i szemléletben várható értéként, azaz  $p_0 = P(Y_0 = 1) = E\left(\frac{1}{1 + \exp(-\mathbf{x}_0^\top \boldsymbol{\beta})}\right)$ . Ennek kiszámításához vegyük  $\frac{1}{1 + \exp(-\mathbf{x}_0^\top \boldsymbol{\beta})}$  a  $\tilde{\boldsymbol{\beta}}$  körüli Taylor sorfejtését, amelynek vége a várható értékét és felhasználva, hogy a várható torzítás  $E(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) = 0$ , adódik, hogy

$$P(Y_i = 1) \approx \tilde{p}_i + C_i, \quad (1.7)$$

ahol a korrekciós faktor

$$C_i = (0.5 - \tilde{p}_i)\tilde{p}_i(1 - \tilde{p}_i)\mathbf{x}_0^\top V(\tilde{\boldsymbol{\beta}})\mathbf{x}_0.$$

A korrekciós faktor formulájából látható, hogy ha a  $\tilde{\beta}$  szórása 0, a  $C_i$  is 0, és amennyiben nő a szórás, úgy fog a korrekciós faktor is nőni. Az is megfigyelhető, hogy  $\tilde{p}_i$  önmagában alulbecsli a ritka jelenségek valószínűségét, hiszen ha  $\tilde{p}_i < 0.5$ , ami jellemző ritka valószínűségek esetén, akkor a  $C_i$  növeli a becsült valószínűséget. A becslés nem torzítatlan, de az átlagos négyzetes hibája alacsonyabb  $p_i$  más közelítéseinél. A gyakorlatban a becslés előnyösebb  $\hat{p}_i$  maximum likelihood közelítésnél, leszámítva amennyiben a variancia mátrix 0 vagy a megfigyelések közel fele 1, ekkor az előnyök nem tudnak érvényesülni [7].

## 1.2. Korrelált rendezett és diszkrét valószínűségi változók szimulációja ismert peremeloszlások mellett

Számos kutatás keretein belül gyűjtött adat gyakran tartalmaz ordinális változókat. Tipikusan ilyenek például a társadalmi tudományokban gyakran alkalmazott Likert-skálás kérdésekre adott válaszok. A biztosítási állományt reprezentáló adatbázis is tartalmaz ordinális és diszkrét változókat, melyeket az állomány alakulásának előrejelzéséhez szükséges volt szimulálni. A biztosítandó személygépjármű paraméterei, mint a motor hengerűrtartalma vagy teljesítménye diszkrét változók, ahogy az értékesítési csatorna és a Bonus Malus fokozat is. Ezek a változók jellemzően korrelálnak egymással. Ilyen változók szimulációjára ismert peremeloszlások mellett készített eljárást Barbiero és Ferrari [3], melyet R-ben ([16] csomag) implementáltak is. Az eljárás többváltozós diszkrét véletlen változóból generál mintát előre meghatározott korrelációs mátrix és peremeloszlások alapján, ahol a perem eloszlások együttes eloszlását Gauss kopula írja le. Az eljárás először előállítja a Gauss kopulát, mellyel a kívánt összefüggési struktúra elérhetővé válik a célt zott diszkrét valószínűségi változókon, míg a következő lépésben a célváltozók segítségével mintát generál. A módszer bármely véges tartó mellett képes véletlen diszkrét változóból szimulálni, mind Pearson- és Spearman-féle korrelációval is, sőt az eljárást kiterjesztették nem korlátos diszkrét változók esetére is [1].

### 1.2.1. Ordinális változók korrelációja

A Barbiero és Ferrari [3] által ismerttetett szimulációs módszer a Pearson-féle együtthatót alkalmazza elsősorban, de az eljárás működik Spearman-féle  $\rho$ -val is. Alapvetően ordinális, diszkrét változók korrelációját Spearman-féle együtthatóval előnyösebb jellemezni, hiszen az a lineáris kapcsolaton túl bármilyen monoton kapcsolatot képes leírni, azonban mivel ordinális változók korrelációjának számítására a szakirodalomban nagyobb hangsúlyt kap a Pearson-féle együttható, a módszer is erre koncentrál. Az eljárás Spearman-féle korrelációra vonatkozó kiterjesztését és a két együttható kapcsolatát a fejezet végén mutatom be.

A Pearson-féle együttható nem mindig korlátos  $-1$  és  $1$  között diszkrét változók vizsgálata során ismert peremeloszlások mellett, így szükséges kiszámítani az együttható minimumát ( $\rho_m$ ) és maximumát ( $\rho_M$ ). Legyenek  $(X_1, X_2)$  ordinális vagy diszkrét változók ismert peremeloszlások mellett,  $F_1(x_1)$  és  $F_2(x_2)$  az  $X_1$  és  $X_2$  eloszlásfüggvényei, és  $F(x_1, x_2)$  az együttes eloszlásfüggvény. Ekkor ismert, hogy  $F_m(x_1, x_2) \leq F(x_1, x_2) \leq F_M(x_1, x_2)$ , ahol  $F_m(x_1, x_2) = \max\{0, F_1(x_1) + F_2(x_2) - 1\}$  és  $F_M(x_1, x_2) = \min\{F_1(x_1), F_2(x_2)\}$ . A korrelációs együttható értéke az  $F(x_1, x_2) = F_m(x_1, x_2)$  esetén minimális, és az  $F(x_1, x_2) = F_M(x_1, x_2)$  esetén pedig maximális [3].

Ordinális (vagy diszkrét) változók folytonos változókból történő generálása során szükséges a folytonos változók közti korrelációs együttható (jelölje  $\rho^C$ ) és a megfelelő diszkrétizált változók közti korrelációs együttható (jelölje  $\rho^O$ ) kapcsolatát vizsgálni és az eljárás során figyelembe venni. A két együttható közti kapcsolatra néhány speciális esetet leszámítva nem található zárt formában lévő kifejezése. Ferrari és Barbiero [3] megvizsgálták standard normális eloszlású változók korrelációs együtthatója ( $\rho^C$ ) és a belőlük diszkrétizálással képzett egyenletes eloszlású változók (egyenlő valószínűségeket véve diszkrétizálva), illetve szimmetrikus, de nem egyenletes eloszlású változók (egyenlő hosszúságú intervallumokkal diszkrétizálva) korrelációs együtthatója ( $\rho^O$ ) közti kapcsolatot (a kategóriák száma a két változó esetén egyszerűsítés céljából megegyezik). Az elemzés eredménye, hogy bármely diszkrétizálási módszert választva  $\rho^C$  nagyobb lesz, mint  $\rho^O$ , de közelednek egymáshoz, ha a diszkrétizált változók kategóriáinak számát növeljük, vagy ha  $\rho^C$  értéke

növekszik.

### 1.2.2. Az eljárás

Az [3] által bemutatott módszer a következő. Legyen  $\mathbf{Z}$  egy  $m$ -dimenziós standard normális eloszlású valószínűségi változó  $\mathbf{R}^C$  korrelációs mátrixszal. Az első szakaszban teljesül, hogy  $\mathbf{R}^C = \mathbf{R}^{O*}$ , ahol  $\mathbf{R}^{O*}$  a célzott korrelációs mátrix. Az eljárás célja az, hogy az eredeti  $\mathbf{Z}$  valószínűségi változót kategorikussá transzformálja, melyet  $\mathbf{X}$  jelöli a későbbiekben. A  $\mathbf{Z}$  változó minden egyes komponense a megfelelő kvantilisek,  $r_{i1} < r_{i2} < \dots < r_{il} < \dots < r_{i,k_i-1}$  mentén felosztásra kerül. Ezek a kvantilisek fogják reprezentálni az ordinális csoportokat az új  $\mathbf{X}$  valószínűségi változóban. Matematikai felírása a következő:

$$\begin{aligned} \text{ha } Z_i < r_{i1} &\rightarrow X_i = 1, \\ \text{ha } r_{i1} \leq Z_i < r_{i2} &\rightarrow X_i = 2, \\ &\vdots \\ \text{ha } r_{i(k_i-2)} \leq Z_i < r_{i(k_i-1)} &\rightarrow X_i = k_{i-1}, \\ \text{ha } r_{i(k_i-1)} \leq Z_i &\rightarrow X_i = k_i. \end{aligned}$$

Az így kapott  $\mathbf{X}$  egyes  $X_i$  elemei különböző számú kategóriával és különböző peremeloszlásokkal rendelkeznek (azaz az egyedet leíró változónként különböző dimenziójú és eloszlású vektor változók állnak elő), a választott  $k_i$  indexnek és  $F_{il}$  értékeinek megfelelően.

Az  $X$  vektor együttes eloszlásához tartozó  $\mathbf{R}^O$  érzékelhetően különbözhet a választott  $\mathbf{R}^C = \mathbf{R}^{O*}$ -tól a fenti eljárás miatt. A kívánt korrelációs mátrix eléréséhez egy köztes  $\mathbf{R}^{C*}$  folytonos mátrixot határozunk meg, melyet iteratív módon transzformációs lépéseken keresztül közelítjük az  $\mathbf{R}^{O*}$  mátrixhoz.

#### Algoritmus

- A célzott korrelációs mátrix legyen egyenlő az eredeti folytonos korrelációs mátrixszal, azaz  $\mathbf{R}^{C(0)} = \mathbf{R}^{O*}$ . A  $\mathbf{Z}$  valószínűségi változóból transzformált diszkrét  $\mathbf{X}^{(1)}$  valószínűségi változót a fent említett eljárás segítségével előállítjuk.

- Az  $\mathbf{R}^{O(1)}$  kiszámítása  $\mathbf{X}^{(1)}$  valószínűségi változóból.
- Az iterációt addig folytatjuk, amíg  $\max |\rho_{ij}^{O(t)} - \rho_{ij}^{O*}| > \epsilon$   $i \neq j$ ,  $t \leq t_{max}$  esetén, ahol  $\epsilon$  a maximális megengedett hiba a végső véletlen változó és a cél korrelációs mátrix megfelelő elemei között. Az iterációk számát legfeljebb  $t_{max}$ -ban rögzítjük. Ezek mellett a következő lépéseket végezzük el:

- A többváltozós normális eloszlás korrelációs mátrixának minden elemét felülírjuk a következőképpen:

$$\rho_{ij}^{C(t)} = \rho_{ij}^{C(t-1)} f_{ij}(t), \quad i = 1, 2, \dots, m, j > i, \quad (1.8)$$

ahol  $f_{ij}(t) = \frac{\rho_{ij}^{O*}}{\rho_{ij}^{O(t)}}$ ,  $\forall i \neq j$ . Az  $f_{ij}(t)$  valójában nem más, mint a diszkretizálást leíró korrekciós együttható.

- Ha  $\mathbf{R}^{C(t)}$  az adott lépésben már nem korrelációs mátrix, akkor megkeressük a hozzá legközelebbi pozitív definit mátrixot.
- Az  $\mathbf{R}^{C(t)}$  alkalmazásával előállítható az  $\mathbf{X}^{(t+1)}$  ordinális véletlen változó a diszkretizálási algoritmus segítségével.
- Kiszámítjuk  $\mathbf{R}^{O(t+1)}$ -t.

Az iteráció legvégén kapott folytonos  $\mathbf{R}^{C*}$  korrelációs mátrixot alkalmazva már a kívánt peremeloszlásokkal és összefüggési struktúrával rendelkező (diszkretizálási transzformáció után) minta generálható.

### 1.2.3. Kiterjesztések

#### Kiterjesztés bármely véges és végtelen tartójú diszkrét eloszlású valószínűségi változóra

A bemutatott eljárás ordinális valószínűségi változókra készült azzal a megkötéssel, hogy minden egyes peremeloszlás véges tartóval rendelkezik. Könnyen látható, hogy bármely véges tartójú diszkrét valószínűségi változóra működik a módszer azzal a módosítással, hogy  $Z_i$  diszkretizálása során annak értékeit egész számok helyett a véges tartó rendezett

értékei reprezentálják [2]. A végtelen tartójú diszkrét valószínűségi változóra vonatkozó kiterjesztésre Ferrari és Barbiero [1],[2] alapján a következő módszer alkalmazható: a valószínűségi változó tartóját meg kell vágni, nem korlátos változóból korlátosat generálni úgy, hogy a tartó felső korlátja legyen  $F^{-1}(1 - \epsilon)$  kvantilis, ahol az  $\epsilon$  a lehető legkisebb legyen. Ezután a lépés után folytatható a fent bemutatott eljárás. A végtelen tartójú diszkrét valószínűségi változóra vonatkozó kiterjesztésre Ferrari és Barbiero [1],[2] alapján a következő módszer alkalmazható: a valószínűségi változó tartóját meg kell vágni, nem korlátos változóból korlátosat generálni úgy, hogy a tartó felső korlátja legyen  $F^{-1}(1 - \epsilon)$  kvantilis, ahol az  $\epsilon$  a lehető legkisebb legyen. Ezután a lépés után folytatható a fent bemutatott eljárás.

### **Kiterjesztés Spearman korrelációra**

A Spearman korrelációs együttható ( $\rho_s$ ) egy Pearson-féle korreláció a rangszámok között, azaz ha az intervallum vagy arányskálán mért értékeket azok rangszámaival helyettesítjük, vagy az ordinális skálájú változókat  $1, 2, \dots, n$ -nel kódoljuk, akkor a belőlük számolt Spearman és Pearson korrelációk megegyeznek. Amennyiben az  $X_1$  és  $X_2$  véletlen változókat tekintjük  $F_1(x_1)$  és  $F_2(x_2)$  eloszlásfüggvényekkel, akkor a  $\rho_s$  az  $F_1(X_1)$  és  $F_2(X_2)$  közti Pearson korrelációs együtthatója, azaz  $\rho_s = \text{cor}(F_1(X_1), F_2(X_2))$  [2],[3]. A Pearson-féle és Spearman-féle  $\rho$  együtthatók megegyeznek meghatározott körülmények mellett, és valós adaton hajlamosak nagyon közeli értékeket feltételezni. A kétváltozós normális eloszlás esetén pontosan ismert a kapcsolatuk, amely a következőképpen leírható:

$$\rho_s = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right),$$

és így csaknem azonos értéket ad a két index [14].

## 2. fejezet

# Alkalmazás - biztosítási állomány tervezése

### 2.1. Adatbázis bemutatása

Szakdolgozatomban egy biztosító kötelező gépjármű felelősség biztosítási állományának alakulását jelzem előre statisztikai módszerekkel. Ez két feladatot foglal magába: egyrészt a szerződések törlésének modellezését, másrészt új szerzések szimulációját. Az adat egy valós biztosító 2011-es állományát mutatja be, mely szerződésenként tartalmazza azt, hogy az évfordulón váltott-e biztosítót az adott ügyfél, vagy a tárgyalt biztosítónál maradt a következő biztosítási időszakban is. A törlések modellezése erre az információra épült. A nyers adat körülbelül 770 000 rekordot és több mint 60 attribútumot tartalmaz, ahol minden egyes rekord az állomány egy-egy szerződőjéről, valamint annak biztosított személygépjárműjéről szolgáltat információt. Ebben a fejezetben az adatbázis szerkezetét, főbb változóit mutatom be.

Munkám első lépése az adatok elemzése és feltérképezése volt. Az adatbázis tartalmazza a biztosító KGFB állományának évforduló napján a kötvények aktuális állapotát. Az adatbázis változóit alapvetően 3 kategóriába lehet sorolni:

1. a szerződőt jellemző tulajdonságok, pl. az ügyfél kora, neme, lakóhelye stb.



2. a biztosított személygépjárműt jellemző paraméterek, pl. a személygépjármű teljesítménye, hengerűrtartalma, kora stb.
3. a biztosítási kötvény technikai adatai, pl. kockázat kezdet éve, évfordulós mozgás stb.

A változókat meg lehet különböztetni aszerint is, hogy azok időtől függenek-e vagy nem. Egyes változók, mint a gépjármű teljesítménye, a szerződő irányítószáma nem változnak az évek múlásával, viszont a szerződő kora, a szerződés kora, a Bonus Malus besorolás előre lépnek (vagy károkozás esetén vissza).

A modellezéshez az eredeti adatbázist több lépésben szükséges volt egyszerűsíteni, transzformálni vagy kipótolni. A biztosítási szerződések adatainak rögzítésekor előfordulhatnak pontatlanságok, amelyek kiugró vagy hiányzó értéként is megjelenhetnek, és ezek modellezés esetén jelentősen torzíthatják az eredményeket. A kérdéses rekordok kiszűrésére a változókat vizuális elemzéseknek vetettem alá, valamint változónként hiányzó értékeket kerestem. A változók értékkészletét - amennyiben lehetséges, minimális információvesztéssel - célszerű aggregálni az eredmények könnyebb értelmezhetőségének kedvéért, illetve számítástechnikai szempontból a memóriaigény csökkentésének érdekében. Így pl. a kategorikus változók esetén csoportokat képeztem, vagy némelyik numerikus változóból kategorikusát készítettem. A következőkben tekintsük át a legfőbb változókat.

### 2.1.1. Ügyfél személyét jellemző változók

Az adatbázis a szerződő legfontosabb adatait tartalmazta, de olyan formában, amely alapján nem lehet az ügyfél személyére következtetni.

Változó	Leírás
Szerződés azonosító	Kategorikus
Területi csoport	Kategorikus
Szerződő neme	Kategorikus
Szerződő kora	Kategorikus
Jogosítvány kora	Kategorikus

2.1. táblázat. Szerződő személyét jellemző változók.

Az alapadat tartalmazta az ügyfelek irányítószámait, amelyeket a könnyebb elemezhetőség és értelmezhetőség érdekében magasabb aggregáltsági szintű kategorikus változóvá transzformáltam. Az új változó az irányítószámhoz tartozó települések jogállása alapján került kialakításra:

Változó	Leírás	Eloszlás
1	Megyeszékhely, megyei jogú város	20.39%
2	Fővárosi kerület	16.80%
3	Megyei jogú város	2.77%
4	Város	29.73%
5	Nagyközség	4.46%
6	Község	25.85%

2.2. táblázat. Területi csoportok és eloszlásuk.

A területi csoportok eloszlása a 2.2 táblázatban láthatók. Az eloszlások alapján látható, hogy a szerződők kb. 2/3-a városokban él.

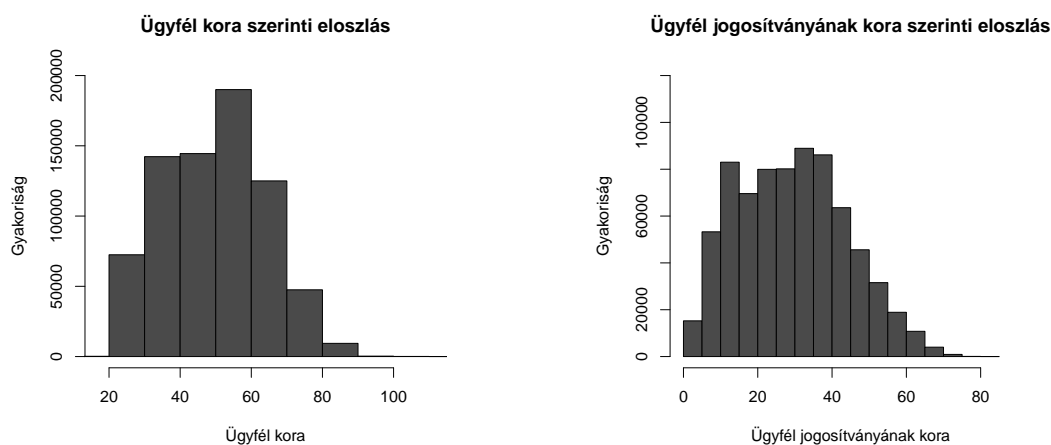
A férfiak és nők aránya az állományban  $2/3 - 1/3$ , nem természetes személyek elvételre találhatók.

	Férfi	Nő	Nem természetes személy
Eloszlás	63.96%	30.89%	5.15%

2.3. táblázat. Szerződő neme szerinti eloszlás.

A szerződő kora valamint a jogosítvány kora változók a szerződés évforduló utáni állapotára vonatkozik. Mindkét attribútum az eredeti adatbázisban numerikus változóként szerepelt, melyeket kategorikus változókká transzformáltam. Ennek oka egyrészt a könnyebb értelmezhetőség volt, másrészt gondot okozott az is, hogy a nem természetes személy szerződők esetén nincs értelme életkorról és jogosítvány szerzésről beszélni, és így az ő esetükben a kor nem értelmezhető numerikus változóként, ami az elemzéseket jelentősen megnehezíti. Ezért mindkét esetben a korokat egyenlő gyakoriságú csoportokra osztottam a decilisek mentén, így képezvén kategorikus változót belőlük, ahol a nem természetes személyek külön kategóriát alkotnak.

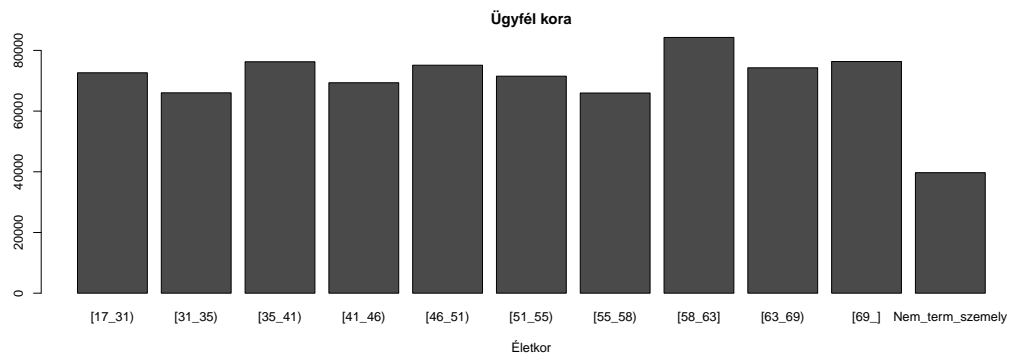
A nyers adatot is szükséges volt korrigálni ezen változók esetében. Ezekben az esetekben a hiányzó adatokat a fent maradó adatok átlagával pótoltam, nagyságrendileg az adatok kevesebb, mint 1%-t érintette.



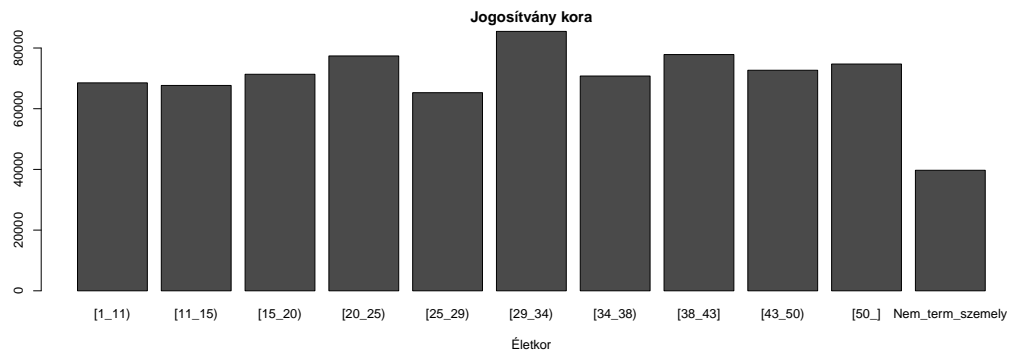
(a) Életkor szerinti megoszlás.

(b) Jogosítvány kora szerinti megoszlás.

2.1. ábra. Ügyfél korának és jogosítvány korának megoszlásai természetes személy szerződők esetén.



2.2. ábra. Életkor szerinti megoszlás kategorizálva.



2.3. ábra. Jogosítvány kora szerinti megoszlás kategorizálva.

### 2.1.2. A biztosítási kötvényt jellemző adatok

A szerződésekhez kapcsolódó legfontosabb változókat a 2.4 táblázat foglalja össze.

Változó	Leírás
Státusz	Indikátor
Szerződés kora	Numerikus
Szerzési csatorna	Kategorikus
Ügyfélszegmens	Kategorikus
Bonus Malus évforduló előtt	Ordinális
Bonus Malus évforduló után	Ordinális
Kár darab kifizetés tárgyalt biztosítási évben	Numerikus
Kármentes évek száma	Dátum
Díjfizetés módja	Kategorikus
Díjfizetés gyakorisága	Kategorikus
Casco együttkötés-e	Indikátor
Üzembentartó és tulajdonos személye megegyezik-e	Indikátor
Konkurens biztosítók ajánlott állománydíja	Numerikus
Tárgyalt biztosító évforduló előtti állománydíja	Numerikus
Tárgyalt biztosító évforduló után ajánlott díja	Numerikus
Tárgyalt biztosító ajánlott díja hányadik legolcsóbb	Ordinális
Tárgyalt biztosító által kínált indexmérték	Numerikus
Biztosítók közti egyéb relatív viszonyt kifejező változók	Numerikus

2.4. táblázat. Biztosítási szerződést jellemző változók.

A törlés modellezéshez egyik legfontosabb változó a státusz indikátor változó, amely 1-es értéket ad, ha évfodulón az adott szerződés törlődött, és 0-t, ha a tárgyalt biztosítónál maradt a szerződő a következő biztosítási időszakban is. Az adatban a töröltek aránya lényegesen kisebb a nem töröltekénél, és az adat ezen kiegyensúlyozatlansága miatt több nehézség is felmerült a törlés-modellezésénél.

<b>Törölt</b>	<b>Nem törölt</b>
74.88%	25.12%

2.5. táblázat. Szerződések biztosítási évfordulóján történtő mozgásának eloszlása.

A kockázat kezdet éve változóból kalkuláltam az "Élő év" nevű változót, amely az ügyfél évfordulóig kockázatban töltött idejét reprezentálja.

<b>Kockázatban töltött évek</b>	
átlag	3.80
tapasztalati szórás	4.15
0%-os kvantilis	1
25%-os kvantilis	1
50%-os kvantilis	2
75%-os kvantilis	5
100%-os kvantilis	20

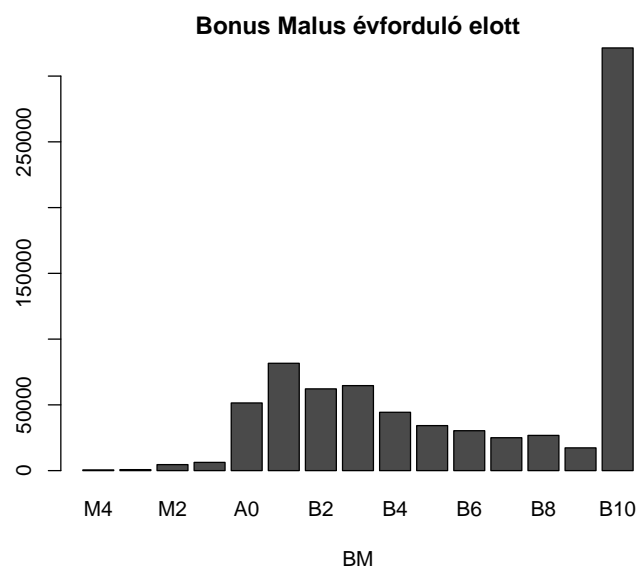
2.6. táblázat. Kockázatban töltött évek leíró statisztikái.

Értékesítési csatornák esetén szintén szükség volt magasabb szintű aggregációra, ahol a következő szempontok alapján csoportosítottam a csatornákat: egyrészt érdemes megkülönböztetni a brókeri és saját függő ügynöki csatornákat, hiszen az általuk szerzett állományok minősége is jelentősen eltérhet. A brókerek közös jellemzője, hogy az ügyfelek érdekeit tartják szem előtt, az ő megbízásukból dolgoznak, míg a biztosító függő ügynökei a biztosítóval állnak szerződésben és az ő termékeit kínálják az ügyfeleknek. Másrészt célszerű lehet különválasztani az internetes biztosítás közvetítőket az "offline" csatornáktól, hiszen szintén eltérő lehet a két típus ügyfélköre.

Szerzési csatorna	Eloszlás
Offline bróker	19.70%
Online bróker	14.09%
Offline függő ügynök	56.32%
Online függő csatornák	3.18%
Egyéb	6.70%

2.7. táblázat. Szerzési csatornák eloszlása.

Az állomány túlnyomórészt B10-es bonus malus fokozatú ügyfeleket tartalmaz, továbbá magas arányban láthatóak az A0-ás ügyfelek.



2.4. ábra. Biztosítási évforduló előtti Bonus Malus.

Az adat tartalmazta a tárgyalt biztosító, valamint a piac többi szereplője által kínált biztosítási díját az ügyfeleknek. Ezekből kalkuláltam több relatív viszonyt kifejező változót. Ilyen változó pl.: a tárgyalt biztosító díja és a piacon a szerződő által választható legkedvezőbb díj hányadosa, illetve a tárgyalt biztosító által kínált díj és az előző évi díjának hányadosa, vagy a tárgyalt biztosító által kínált díj piacon való rangszámát

kedvezőség szempontjából. A KGFB szerződések esetén a verseny erősebb, mint bármely más ágazatban, itt kiemelten számít, hogy egy biztosító mennyire tud kedvező díjat kínálni.

<b>Tárgyalt biztosító díja/Piacon elérhető legkedvezőbb díj</b>	
átlag	1.27
tapasztalati szórás	0.30
0%-os kvantilis	1
25%-os kvantilis	1.00
50%-os kvantilis	1.18
75%-os kvantilis	1.41
100%-os kvantilis	4.98

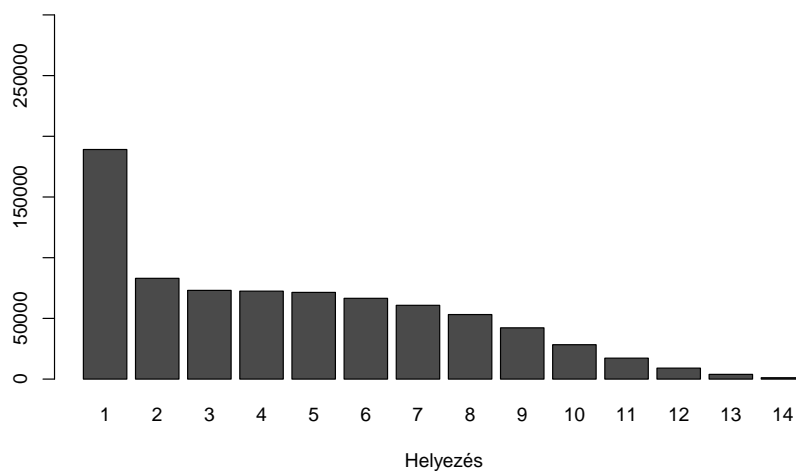
2.8. táblázat. A tárgyalt és piacon lévő többi konkurens biztosító által kínált díjakból kalkulált változó leíró statisztikái.

<b>Tárgyalt biztosító által kínált indexmérték</b>	
átlag	0.68
tapasztalati szórás	0.18
0%-os kvantilis	0.14
25%-os kvantilis	0.56
50%-os kvantilis	0.67
75%-os kvantilis	0.78
100%-os kvantilis	15.41

2.9. táblázat. A tárgyalt biztosító által indexált díjak mértékének leíró statisztikái.



Tárgyalt biztosító helyezései



2.5. ábra. A tárgyalt biztosító helyezéseinek eloszlása a kínált díj kedvezősége szerint. Látható, hogy az első négy helyezés az állomány több mint felét kiteszi.

Helyezés	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
Eloszlás	24.52%	10.76%	9.47%	9.39%	9.26%	8.63%	7.88%	6.89%	5.47%	3.67%	2.27%	1.17%	0.5%	0.15%

2.10. táblázat. A tárgyalt biztosító helyezéseinek eloszlása a kínált díj kedvezősége szerint.

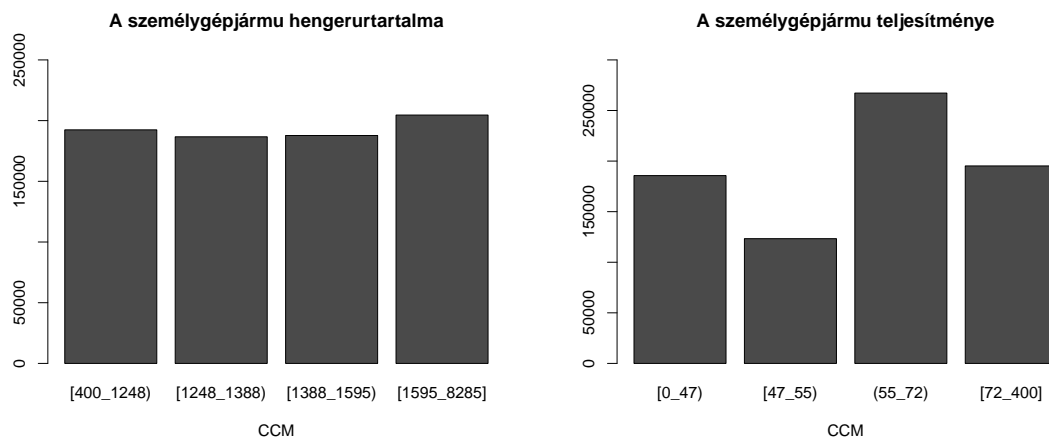
### 2.1.3. A biztosított gépjárművet jellemző adatok

Változó	Leírás
Hengerűrtartalom	Kategorikus
Teljesítmény	Kategorikus
Gyártmány	Kategorikus
Jármű kora	Numerikus
Üzemanyag	Kategorikus

2.11. táblázat. A járművet jellemző változók.

A 2.11 táblázat összefoglalja a járművekkel kapcsolatos technikai információkat. A teljesítmény és hengerűrtartalom változóknál a korábbiakhoz hasonlóan előfordultak hiányzó

értékek és outlierek, esetükben szintén alkalmaztam a korábban említett korrekciókat. Az egyenlő gyakorisági osztályközöket esetükben a kvartilisek alapján képeztem.



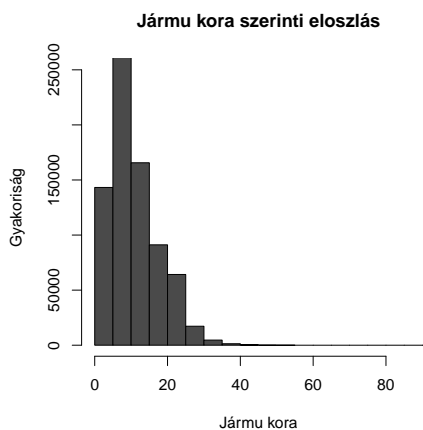
(a) Hengerűrtalom.

(b) Teljesítmény.

2.6. ábra. A személygépjárművek hengerűrtartalmának és teljesítményének eloszlása.

A gépjármű kora változó alapján látható, hogy a biztosított autók átlagosan 11 évesek, és kis számban még veterán korú járművek is megtalálhatók az állományban. Esetükben szintén a decilisek mentén kategorizáltam.

Változó	Leírás
átlag	11.21
szórás	6.47
0%-os kvantilis	1
25%-os kvantilis	6
50%-os kvantilis	10
75%-os kvantilis	15
100%-os kvantilis	111



2.12. táblázat. Jármű korának leíró statisztikai.

2.13. táblázat. Jármű kora empirikus eloszlása.

### 2.1.4. A változók és törlések kapcsolatának előzetes vizsgálata

Az előzetes elemzések során fontos a törlést, mint eseményt, a szerződőt leíró jellemzőinek függvényében látni. Nem célom az adatbázis összes változójával bemutatni az eredményeket, néhány érdekesebb esetet emelnék ki.

A 2.14 táblázat eredményei alapján elmondható, hogy a függetlenség a státusz és ezen változók között elvethető, és legerősebb összefüggést a törlés/nem törlés kimenetelével a szerzési csatorna mutat, de emellett a szerződés kora és jogosítvány kora is jelentősebb korrelációt jelez.

	$\chi^2$ p-érték	Cramer V
Területi csoport	$< 2.2 \cdot 10^{-16}$	0.11
Díjfizetés módja	$< 2.2 \cdot 10^{-16}$	0.11
Díjfizetés gyakorisága	$< 2.2 \cdot 10^{-16}$	0.12
Szerződő neme	$< 2.2 \cdot 10^{-16}$	0.02
Szerződő kora	$< 2.2 \cdot 10^{-16}$	0.21
Jogosítvány kora	$< 2.2 \cdot 10^{-16}$	0.22
Szerzési csatorna	$< 2.2 \cdot 10^{-16}$	0.36
Évforduló előtti Bonus Malus	$< 2.2 \cdot 10^{-16}$	0.19
Jármű kora	$< 2.2 \cdot 10^{-16}$	0.09

2.14. táblázat. Az évfordulós mozgás és kategorikus adatok közötti korreláció.

A numerikus adatok és státusz kapcsolatának vizsgálatára varianciaanalízist alkalmaztam, melyek esetében a függetlenség szintén elvethető.

	ANOVA p-érték
Tárgyalt biztosító által kínált díj/Piacon kínált legolcsóbb díj	0.00397
Tárgyalt biztosító indexmértéke	$< 2.2 \cdot 10^{-16}$
Kár darab kifizetés	$< 2.2 \cdot 10^{-16}$

2.15. táblázat. Variancia analízis eredményei.

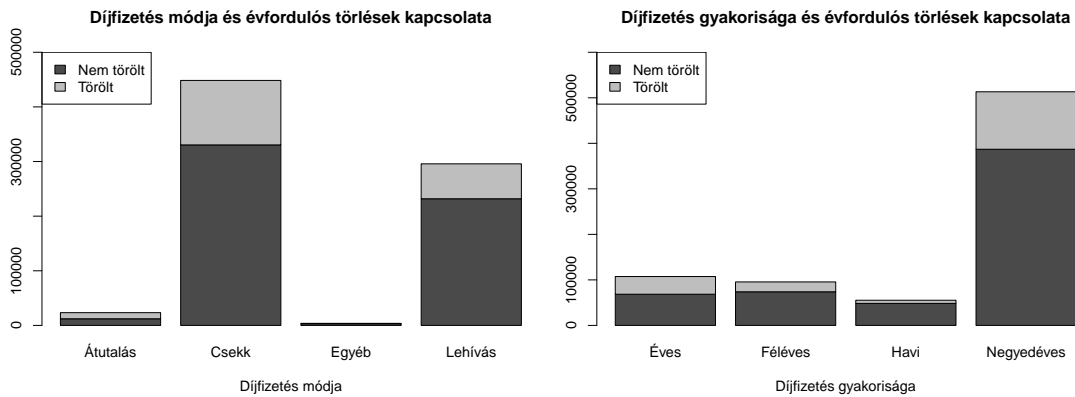
## 2.2. Állomány mozgás előrejelzése

A következőkben az egyes évekre előrejelzett eredményeket fogom bemutatni. Az állományra 3 éves előrejelzést készítettem, az első lépés a kezdeti állapot alapján törlés modellezés volt, amelyet a nem törölt állományon alkalmaztam az 1. év végén. Ezt követően az alapadatban a változók között fellelhető korrelációk vizsgálata után szimuláltam új szerződések kötését a tárgyalat biztosítóhoz.

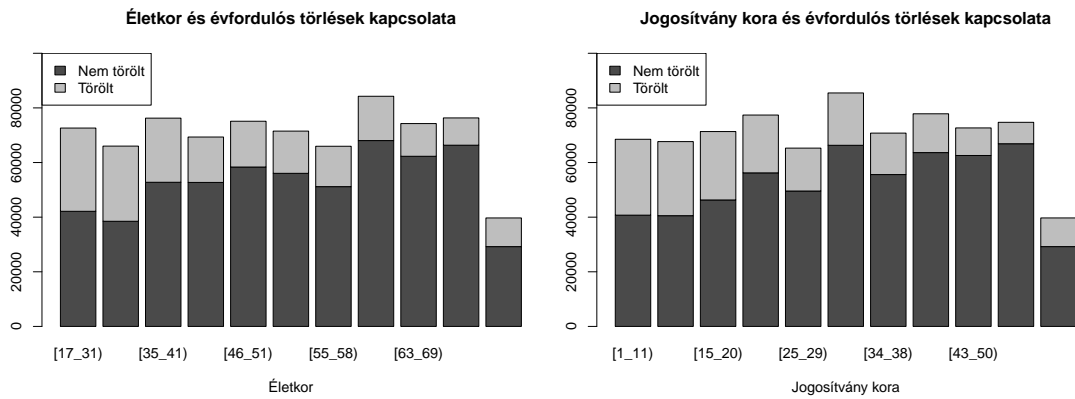
### 2.2.1. Évfordulók szimulációja

A törölt események adatbázisban való alulreprezentáltsága miatt a hagyományos logit modellezés helyett az 1.1 fejezetben bemutatott, ritka események modellezésére készült korigált logit modellt alkalmaztam.

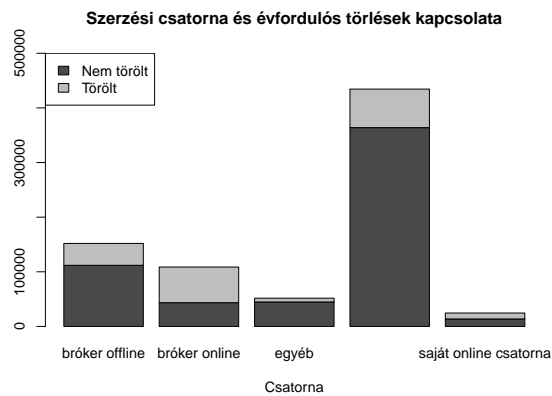
Az első lépés az adatbázis szelekciója volt. Ezt kétféle módszerrel tettem meg, így két modell készült, melyek közül a jobban teljesítőt alkalmaztam a későbbiekben. A két módszerben közös volt, hogy az összes 1-es státuszú megfigyelést kiválasztottam, hiszen annál több információt tartalmaz a modell a törlésekről, minél nagyobb számú megfigyelés van róluk. A különbség az évfordulón nem törölt megfigyelések kiválasztásában volt. A tárgyalat biztosítónál maradt megfigyelésekből a biztosítási szerződést törölt szerződőkkel azonos mennyiségűt választottam, hiszen minden további 0-s ezen túl már a törlések információ tartalmából venne el. Az első modellben az élő szerződések megfigyeléseiből vizatevés nélküli mintavételezéssel választottam ki a kellő mennyiséget, így készítve az almintát, amin a modellt építettem. A másik módszer az volt, hogy a magyarázó változók alapján szegmentálva szelektáltam 0-s almintát. Ehhez megvizsgáltam az egyes magyarázó változók kapcsolatát a törlésekkel, és amennyiben valamely kategóriánál magasabb volt a töröltek aránya, annak mentén generáltam almintát. A logit modelleket a súlyozás módszerével készítettem R-ben a [18] csomag `Relogit` függvény alkalmazásával.



(a) Díjfizetés módja és törlések kapcsolata. (b) Díjfizetés gyakorisága és törlések kapcsolata.



(c) Ügyfél kora és törlések kapcsolata. (d) Jogosítvány kora és törlések kapcsolata.



(e) Értékesítési csatorna és törlések kapcsolata.

2.7. ábra. Egyes változók megoszlása törlés-nem törlés szerint megbontva.

Ezek alapján a 2. modell 0 státuszú rekordjait az online bróker értékesítési csator-

nán keresztül szerzett, 17 – 35 életév közötti, negyedévente csekkel fizető, 1 – 20 éves jogosítvánnyal rendelkező ügyfelek közül választottam ki véletlen módon.

A modellek elkészítése után alkalmaztam őket a teljes állományon. Megvizsgáltam az eredményeket 3 valószínűségi küszöbérték mellett: a kiindulási érték a standard 0.5 volt, azonban mivel elég magas másodfajú- és viszonylag jó elsőfajú hibával rendelkeztek a modellek, csökkentettem a küszöbértéket, tekintve, hogy a törölt szerződések pontosabb besorolása volt a fő cél. Végül 0.3-as küszöbértéket választottam, annál alacsonyabb küszöbértékek mellett az elsőfajú hiba is magas volt. Az egyes küszöbértékek melletti konfúziós mátrixok alább láthatók.

	<b>1. modell</b>		<b>2. modell</b>	
	<b>Megfigyelt értékek</b>			
<b>Előrejelzett értékek</b>	Nem törölt	Törölt	Nem törölt	Törölt
Nem törölt	538 955	128 438	537 307	135 081
Törölt	38 640	65 347	40 288	58 704

2.16. táblázat. A modellek konfúziós mátrixai 0.5 küszöbérték mellett. Az 1. modell elsőfajú hibájának mértéke 6.69%, a másodfajú hibája viszont magas, 66.28%. A 2. modell hasonló értékeket mutat, alig magasabb elsőfajú hibával (6.98%) és másodfajú hibával (69.7%).

	<b>1. modell</b>		<b>2. modell</b>	
	<b>Megfigyelt értékek</b>			
<b>Előrejelzett értékek</b>	Nem törölt	Törölt	Nem törölt	Törölt
Nem törölt	515 677	107 680	496 779	107 579
Törölt	61 918	86 105	80 816	86 206

2.17. táblázat. A modellek konfúziós mátrixai 0.4 küszöbérték mellett. Az 1. modell elsőfajú hibája 10.71%, másodfajú hibája 55.57%. A 2. modell elsőfajú hibája: 13.99%, másodfajú hibája: 55.51%.

	<b>1. modell</b>		<b>2. modell</b>	
	<b>Megfigyelt értékek</b>			
<b>Előrejelzett értékek</b>	Nem törölt	Törölt	Nem törölt	Törölt
Nem törölt	470 136	82 758	427 089	79 139
Törölt	107 459	111 027	150 506	114 646

2.18. táblázat. A modellek konfúziós mátrixai 0.3 küszöbérték mellett. Az 1. modell esetén az elsőfajú hiba mértéke 18.6%, a másodfajú hiba 42.71%. A 2. modell esetén 26.06% az elsőfajú hiba, 40.84% a másodfajú hiba.

A két modell közül az 1. modellt választottam a törlések modellezésére, mert bár enyhén rosszabbul találja el a törölt szerződéseket, de az élő szerződések esetén lényegesen jobban teljesít. A logit modellek teljesítményét tipikusan az AUC számok mutatja meg, melyek alapján mindkét modell elfogadható.

	<b>1. modell</b>	<b>2. modell</b>
AUC	0.7571363	0.7264153

2.19. táblázat. A modellek AUC (Area Under Curve) értékei 0.3 küszöbérték mellett.

A következő lépés a kiinduló állomány új szerződéseinek szimulációja volt, melyeket a [16] csomag segítségével hoztam létre R programozási környezetben. Mivel az új szerződésekről, valamint a piacon történő átkötésekről nem áll rendelkezésre információ, így a MABISZ által publikált kampány adatokból [9] indultam ki: ez alapján 2011 január 1-jén  $\approx 1\,000\,000$  db, 2012-ben  $\approx 700\,000$  db és 2013-ra már csak  $\approx 400\,000$ -en váltottak biztosítót. Ez a csökkenés az új törvényi előírásoknak volt köszönhető, hiszen ezekben az években az ügyfelek már nem csak január 1-jei évfordulóra köthettek KGFB-t. A szakdolgozatban évközi KGFB állományt figyelmen kívül hagytam, csak a január 1-es állományt szimuláltam.

A szimulációk összefüggési struktúráját az alap adat biztosította. A korrelált valószínűségi változók szimulációjának előfeltétele, hogy a korrelációs mátrix pozitív szemidefinit.

Azonban, ha numerikus változókon kívül kategorikus vagy ordinális változók korrelációját is tartalmazza a mátrix, nem minden esetben teljesül az előbbi feltétel. Ez gyakori probléma például a pénzügyben részvények korrelációjának vizsgálatakor is. A problémával [6] cikk foglalkozott részletesebben, és az abban bemutatott algoritmus R-beli implementációját (`nearPD`, [19] csomag) alkalmaztam a mátrixok korrigálására. Az algoritmus a közelítendő (tipikusan valamely korrelációs vagy kovariancia) mátrixhoz leghasonlóbb pozitív szemidefinit mátrixot állítja elő a legkisebb négyzetek elve alapján.

Az alap adatban tehát ismertek voltak a felmondott szerződések. Az új szerzés előállításához 1 000 000 db szerződést szimuláltam, melyek a piacon átkötő szerződéseket reprezentáltak. Az adatállomány méretéből, illetve a piacon felmondott szerződések darabszámából 20%-os részesedést tételeztem fel, így 200 000 szerződést választottam ki véletlenszerűen, amelyben az eredeti adatban található díjak szerinti helyezések eloszlását megtartottam.

Az alap adat kárain Poisson regressziót alkalmazva modelleztem a károk darabszámát, mellyel az aktuális év kárdarabszámát jeleztem előre. A Poisson regresszió a GLM speciális esete log-link függvényvel és Poisson eloszlású kimenettel [11]. A modell változóit stepwise módszerrel választottam ki, melyek között szerepelt a Bonus Malus, a gépjármű hengerűrtartalma, teljesítménye, hajtás típusa és kora, az értékesítési csatorna, melyen keresztül érkezett, a díjfizetés jellemzői, ügyfél kora, továbbá azok az információk, hogy Cascoval együtt kötötték-e a KGFB szerződést, és hogy az üzembentartó és a tulajdonos személye megegyezik-e. A kárdarabszám modellt minden évre újra építettem, az aktuális évet megelőző év előrejelzett kárait bekövetkezettként figyelembe véve.

Az első évfordulóhoz szükséges volt léptetni a különböző korokat reprezentáló változókat, a modellezett kárdarabszámok ismeretében a Bonus Malust és a kármentes évek számát, továbbá a kockázatban töltött évek számát. A két adat összefésülésével elkészült az első évfordulós állomány, melyen több scenáriót is készítettem.



## Első évforduló

Érdemes megvizsgálni, hogy az évfordulós törlés mennyire érzékeny a konkurencia díjak növekedésére/csökkenésére, illetve a tárgyalt biztosító díjváltozásaira.

	Törlések száma	Törlések aránya
Bázis	140 857	18.11%
Konkurencia 10%-os díjemelés	137 317	17.66%
Konkurencia 10%-os díj csökkentés	145 754	18.74%
Konkurencia 20%-os díjemelés	134 847	17.34%
Konkurencia 20%-os díj csökkentés	152 883	19.66%
Tárgyalt biztosító 10%-os díjemelés	145 287	19.84%
Tárgyalt biztosító 10%-os díj csökkentés	134 226	18.74%
Tárgyalt biztosító 20%-os díjemelés	150 195	19.31%
Tárgyalt biztosító 20%-os díj csökkentés	133 955	17.22%

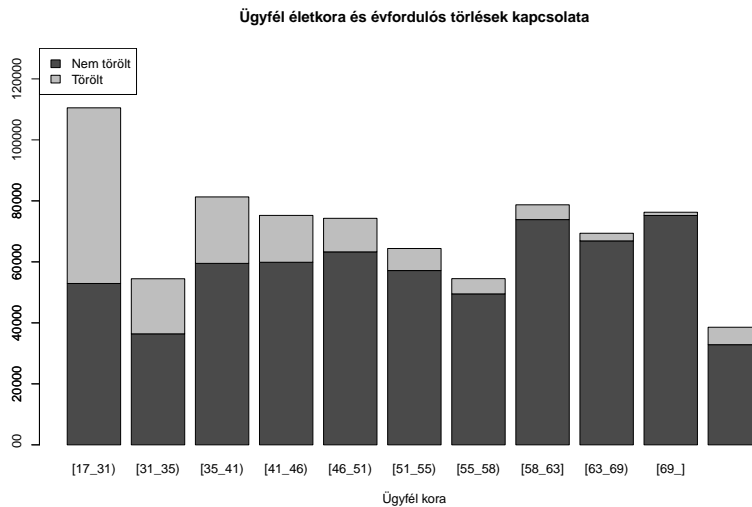
2.20. táblázat. Az első évforduló előrejelzett törlési arányai különböző scenáriók mellett. A bázis scenárió a nyers törléseket mutatja be, amely csak az egyes paraméterek léptetéséből fakadó változásokat tartalmazza, és feltételezi, hogy a piaci szereplők nem módosítanak a díjakon.

A 2.20 táblázat szemlélteti, hogy a kiinduló állományon épített törlési modell előrejelzései alapján hogyan alakulnak a felmondások a 777 595 db kockázatban lévő KGFB szerződés állományán. A bázis scenáriótól a többi kimenetel az egyes biztosítók piaci viselkedése alapján tér el. A kimenetek szempontjából lényegi különbséget okozott a tárgyalt biztosító és a konkurencia díjai egymáshoz viszonyított pozícionálódása. Azt az egyszerűsítő feltevést alkalmaztam, hogy a konkurencia díjváltozása együtt mozog, tehát mindegyik szereplő ugyanolyan arányban emel vagy csökkent díjat. A várakozásoknak megfelelően a törlések érzékenyek a tarifa módosításokra: ha díjat csökkent a konkurencia, többen mondják fel a szerződésüket a tárgyalt biztosítónál, fordított esetben pedig éppen az ellenkezője teljesült. Az ügyfelek a modell szerint másképpen reagálnak arra,

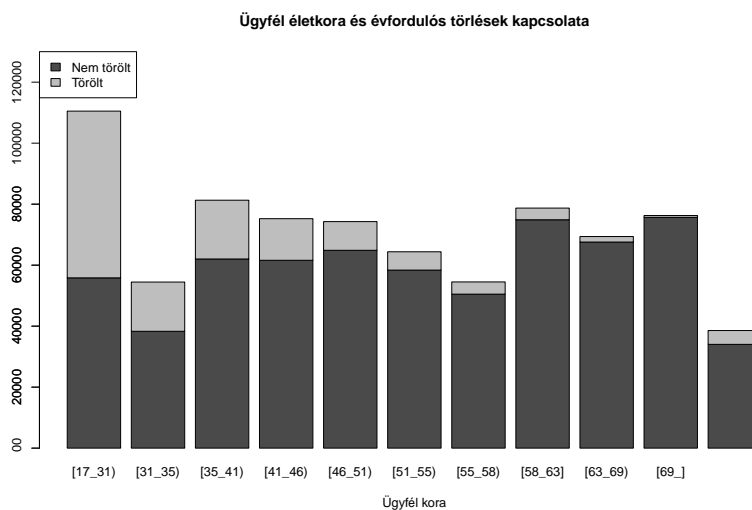
ha a tárgyalt biztosító emeli, vagy ha a konkurencia csökkent díjat. Az ügyfelek kevésbé érzékenyek a konkurencia díjcsökkentésére, aminek oka lehet például márkahűség, vagy adott esetben némileg alacsonyabb konkurencia által kínált díj nem éri meg számukra a felmondással járó fáradozásokat.

A következő két évet két irányba ágaztattam el. Megvizsgáltam, hogy hogyan alakul az állomány, ha a tárgyalt biztosító minden évben 20%-kal emeli a díját, illetve ha ugyanazzal a mértékkel csökkenti. A szimulált új állományon kopást feltételeztem, de a [9] által bemutatottnál enyhébbet állítottam be: 2. évben 180 000 db új szerződést választttam ki a 700 000-ból, 3. évben 140 000 db-ot a 400 000-ból. A törlési modelleket minden évben minden scenárióra újraépítettem.

Az első scenárió (20% díjemelés a tárgyalt biztosító részéről) alapján a szerződést felmondók aránya 19.31%. Alapvetően elmondható, hogy elsősorban fiatalok mondták fel a KGFB szerződéseiket. Ez részben betudható annak, hogy a fiatalabb korosztályban maga a Bonus Malus fokozat is magasabb a korból fakadóan, emiatt magasabb díjat is kaphatnak. Másfelől ez a korosztály sokkal inkább fogékonyabb az online összehasonlító oldalak használatára, ami még jobban élesíti a piacon a versenyt. Ez látható a törlések értékesítési csatorna szerinti megbontásából is. Szintén megfigyelhető magasabb arányú törlés az újonnan szerződést kötöttek körében, melyeknek jelentős részét a fiatalok teszik ki. Ezekre az ügyfélcsoportokra jellemző, hogy a díjfizetés módja többnyire csekkel és/vagy lehívással történik. A többi adatra vonatkozó ábra megtalálható a függelékben.



2.8. ábra. Életkor és törlések kapcsolata első scenárió esetén.



2.9. ábra. Életkor és törlések kapcsolata második scenárió esetén.

### Második évforduló

A második szimulált évfordulón a 20%-os díjemelés scenárió szerint 807 400 db szerződés volt, ebből a töröltek száma a 2.21 táblázatban látható. A 20%-os csökkentés esetén már 823 640 db szerződés adatán vizsgálhatjuk a törléseket. Mindkét scenárió esetén csökkent a törlések aránya az első évfordulóhoz képest 4 – 5%-kal, egymáshoz viszonyítva pedig

megmaradt az alacsonyabb törlési arány a díjcsökkentés javára.

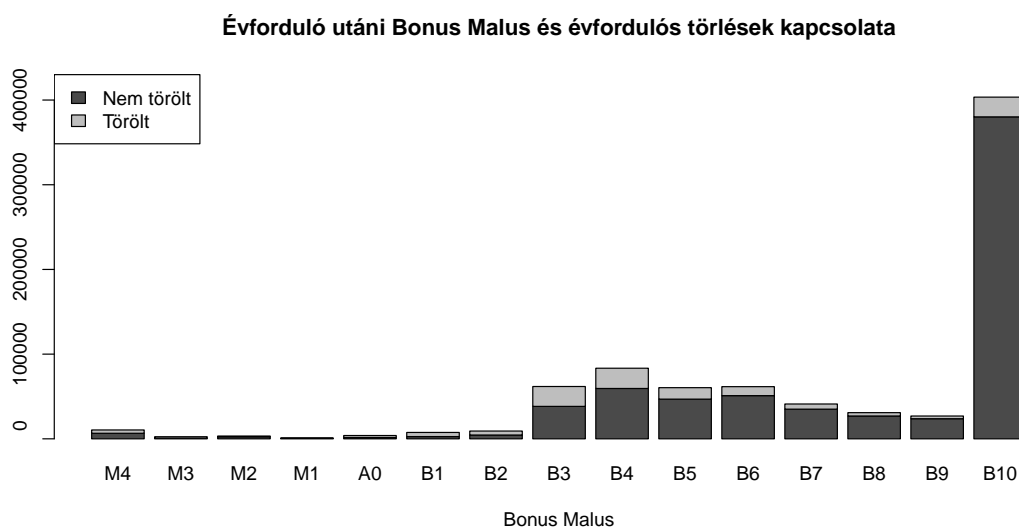
Az alapvető tendencia a változók mentén megmaradtak, legfeljebb arányaikban térnek el egymástól. A kockázatban töltött évek száma szerint a 20%-os díjcsökkentés esetén főleg az 1 éves szerződésekre korlátozódnak a törlések, míg a díjemeléssel szemben a 2 – 3 éves kötvényekkel rendelkezők is töröltek. Az első scenárióban némileg magasabb törlést figyelhetünk meg a B10-es kategóriákban is, tehát a díjemelésre az idősebb korosztály is érzékeny. Díjmelés esetén az online csatornák állományán is jobban töröltek, továbbá a díjmelésre továbbra is a fiatalabb korosztály érzékenyebb.

	<b>Törlések száma</b>	<b>Törlések aránya</b>
Bázis	113 838	14.10%
Konkurencia 10%-os díjmelés	110 522	13.69%
Konkurencia 10%-os díj csökkentés	120 210	14.89%
Konkurencia 20%-os díjmelés	108 186	13.40%
Konkurencia 20%-os díj csökkentés	131 223	16.25%
Tárgyalt biztosító 10%-os díjmelés	119 421	14.79%
Tárgyalt biztosító 10%-os díj csökkentés	110 229	13.65%
Tárgyalt biztosító 20%-os díjmelés	126 912	15.72%
Tárgyalt biztosító 20%-os díj csökkentés	107 105	13.27%

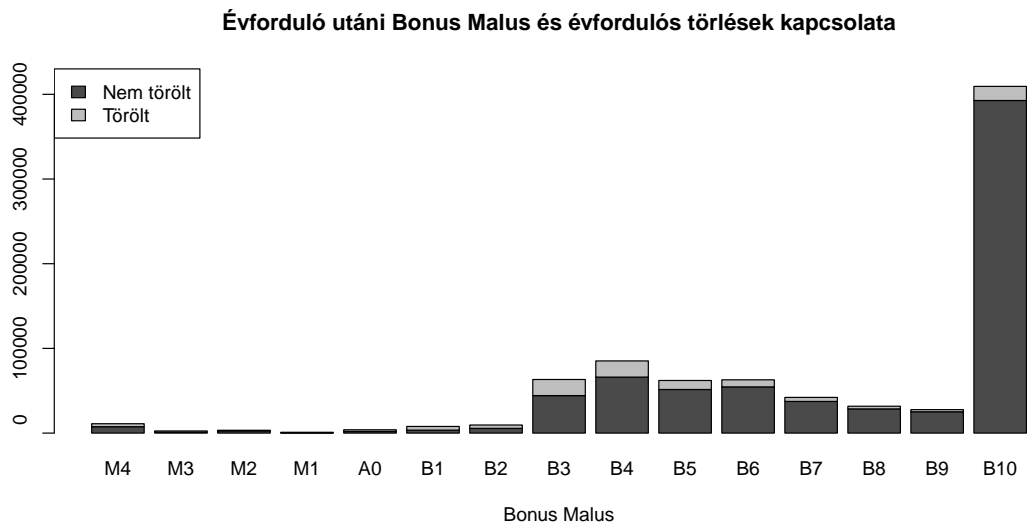
2.21. táblázat. Első scenárió – második évforduló törlések aránya.

	Törlések száma	Törlések aránya
Bázis	107 640	13.05%
Konkurencia 10%-os díjemelés	104 412	13.51%
Konkurencia 10%-os díj csökkentés	111 788	12.64%
Konkurencia 20%-os díjemelés	101 930	12.38%
Konkurencia 20%-os díj csökkentés	117 941	14.32%
Tárgyalt biztosító 10%-os díjemelés	111 298	13.51%
Tárgyalt biztosító 10%-os díj csökkentés	104 110	12.64%
Tárgyalt biztosító 20%-os díjemelés	115 632	14.04%
Tárgyalt biztosító 20%-os díjcsökkenés	100 785	12.24%

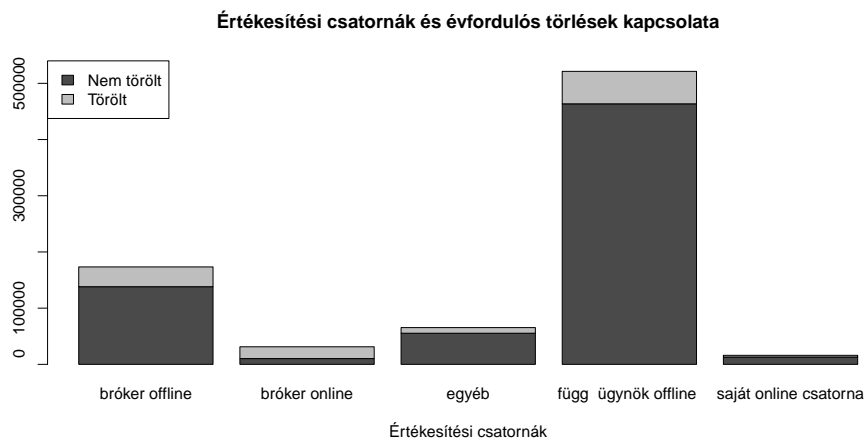
2.22. táblázat. Második scenárió – második évforduló törlések aránya.



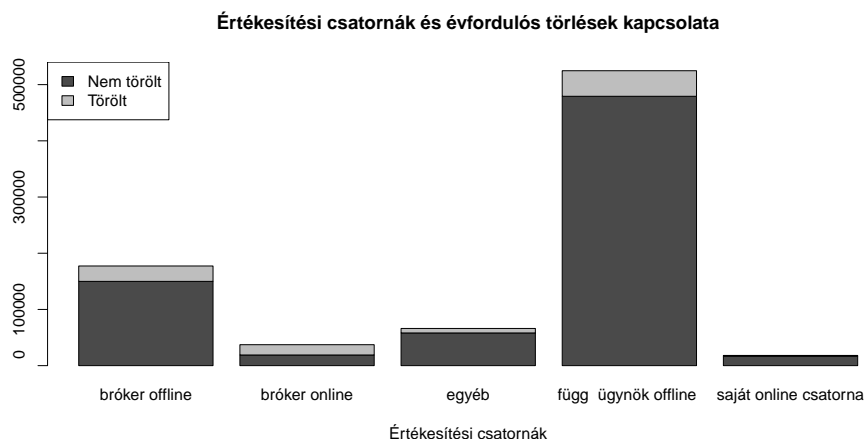
2.10. ábra. Bonus Malus és törlések kapcsolata.



2.11. ábra. Bonus Malus és törlések kapcsolata.



2.12. ábra. Értékesítési csatornák és törlések kapcsolata.



2.13. ábra. Értékesítési csatornák és törlések kapcsolata.

### Harmadik évforduló

A harmadik szimulált évfordulón már jelentős eltérés volt az állomány darabszámában. Első scenárió esetén 830 488 darab szerződést figyelhattünk meg, a második scenárió esetén 862 855 darabot. A 3. évfordulóra az M3 és B3 besorolások közé eső ügyfelek száma jelentősen lecsökkent, azaz a károkozók és fiatal vezetők töröltek a legmagasabb arányban. A harmadik évre az online brókeri csatorna állománya teljesen elkopott.

	Törlések száma	Törlések aránya
Bázis	102 382	12.48%
Konkurencia 10%-os díjemelés	99 583	12.14%
Konkurencia 10%-os díj csökkentés	103 499	12.61%
Konkurencia 20%-os díjemelés	96 833	11.80%
Konkurencia 20%-os díj csökkentés	104 690	12.76%
Tárgyalt biztosító 10%-os díjemelés	103 419	12.60%
Tárgyalt biztosító 10%-os díj csökkentés	99 260	12.10%
Tárgyalt biztosító 20%-os díjemelés	104 243	12.70%
Tárgyalt biztosító 20%-os díj csökkentés	95 525	11.64%

2.23. táblázat. Első scenárió – harmadik évfordulós törlések aránya.

	<b>Törlések száma</b>	<b>Törlések aránya</b>
Bázis	96 508	11.18%
Konkurencia 10%-os díjemelés	93 279	10.81%
Konkurencia 10%-os díj csökkentés	97 951	11.35%
Konkurencia 20%-os díjemelés	90 363	10.47%
Konkurencia 20%-os díj csökkentés	100 137	11.61%
Tárgyalt biztosító 10%-os díjemelés	97 797	11.33%
Tárgyalt biztosító 10%-os díj csökkentés	92 926	10.77%
Tárgyalt biztosító 20%-os díjemelés	99 294	11.51%
Tárgyalt biztosító 20%-os díj csökkentés	88 9875	10.10%

2.24. táblázat. Második scenárió – harmadik évfordulós törlések aránya.

A várakozásnak megfelelően a 3 éven keresztül tartó folyamatos 20%-os díjcsökkentés jelentősen kisebb arányú törlést eredményezett, viszont az emelés és csökkentés mértéke nem szegmentálta jelentősen az ügyfelek körét, jórészt az ugyanolyan jellemzőkkel rendelkező biztosítottak mondták fel a szerződéseiket: fiatal, alacsony Bonus Malus fokozattal rendelkező, nem hosszú ideje a biztosítónál szerződésben álló ügyfelek kötöttek át. A módszer ezért természetesen nem elegendő összetettebb állománytisztításra, viszont érzékenység vizsgálatra megfelelő, amennyiben piaci részesedést szeretne egy biztosító növelni, és előrejelezni állományának változását.



# Összegzés

Szakedolgozatomban egy biztosító 2010. január 1-jei KGFB állományának valós adata alapján készítettem annak jövőbeli mozgásáról előrejelzéseket, szimulációs módszerek alkalmazásával. A biztosítási kötvények felmondását modelleztem, és újonnan kötött szerződéseket szimuláltam, az 1. fejezetben bemutatott elméleti összefoglalók eredményei alapján. A törlési modellek készítéséhez a törölt események alulreprezentáltsága miatt a klasszikus logit modellek helyett a ritkán bekövetkező jelenségek adatára speciális, korrigált logisztikus regressziót alkalmaztam. A törlés modellt minden évben újraépítettem az aktuális adaton. A károk darabszámának modellezéséhez Poisson regressziót alkalmaztam, minden évre külön kalkulálva. Az új szerződések szimulációja során összefüggő diszkrét véletlen változókat generáltam az alapadatban található összefüggési struktúra alapján.

Az állomány mozgásának előrejelzését 3 évfordulóra készítettem el, évközi KGFB állományt figyelmen kívül hagyva. Két alapesetet futtattam végig: a tárgyalt biztosító 3 éven keresztül minden évfordulóján 20%-kal magasabb tarifát hirdet az előző évinél, valamint ennek ellenkezőjeként, mind a 3 évben 20%-kal díjat csökkent. Az alapeseteken kívül minden évre megvizsgáltam egyéb kimeneteleket is, így a konkurencia díjak csökkenésével/emelkedésével járó törlések arányát is, valamint a tárgyalt biztosító egyéb mértékben történő díjváltoztatását is.

Az érzékenységi vizsgálat eredménye szerint mindkét scenárió esetén első sorban a fiatal, alacsony Bonus Malus fokozatú, viszonylag rövid ideje kockázatban álló, online brókeri csatornán érkező ügyfelek törölődtek. Ennek oka lehet, hogy a fiatalabb generáció alacsonyabb Bonus Malus besorolás mellett jelentősen magasabb díjat fizethet, így érzékenyebb a piac változásaira, valamint tudatosabbak is lehetnek, jobban kihasználják az internet

adta piac-összehasonlítási lehetőségeket. A várakozásoknak megfelelően díjemelések esetén magasabb törlési arányok figyelhetők meg.

# Irodalomjegyzék

## Cikkek:

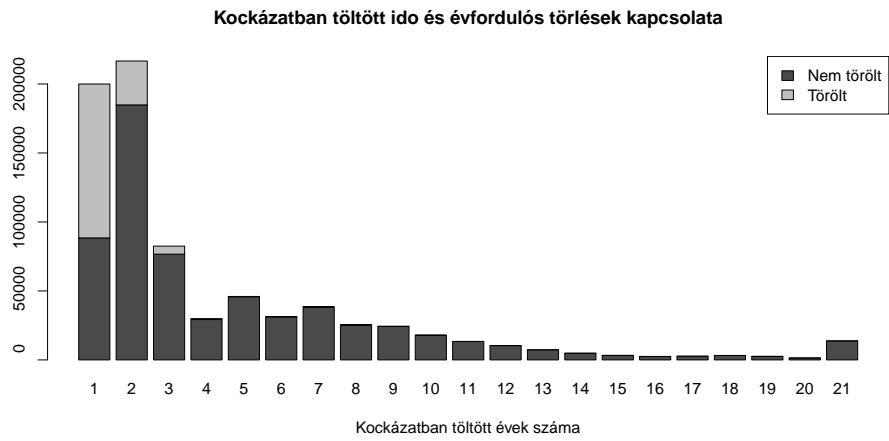
- [1] Barbiero, A., and Ferrari, P. A.: *Simulation of correlated Poisson variables*. Appl. Stochastic Models Bus. Ind., 31: 669-680., <https://doi.org/10.1002/asmb.2072> (2015)
- [2] Barbiero, A., and Ferrari, P.: *Simulating Correlated Ordinal and Discrete Variables with Assigned Marginal Distributions*, Springer Science+Business Media New York, V.B. Melas et al. (eds.), Topics in Statistical Simulation, Springer Proceedings in Mathematics & Statistics 114, DOI 10.1007/978-1-4939-2104-1\_4 (2014)
- [3] Pier Alda Ferrari & Alessandro Barbiero: *Simulating Ordinal Data*, Multivariate Behavioral Research, 47:4, 566-589. old, DOI: 10.1080/00273171.2012.692630 (2012)
- [4] Fülöp P.: *A bináris logit modellek használatának és tesztelésének eszközei*. Statisztikai Szemle. 80. évf. 3. sz. 261-278. old. (2002)
- [5] Hajdu O.: *A csődesemény logit-regressziójának kismintás problémái*. Statisztikai Szemle. 82. évf. 4. sz. 392-422. old. (2004)
- [6] Nicholas J. Higham: *Computing the nearest correlation matrix – a problem from finance* (2002)
- [7] Gari King, Langche Zeng: *Logistic Regression in Rare Events Data* (2001)
- [8] Kovács Erzsébet: *Többváltozós adatelemzés* (2014)
- [9] <https://mabisz.hu/kevesebben-valtottak-a-tavalyi-kampanyban-biztositot/>

- [10] Manski, Charles F., and Steven R. Lerman.: *The Estimation of Choice Probabilities from Choice Based Samples*. *Econometrica* 45(8):1977-1988. (1977)
- [11] McCullagh, P. and Nelder, J.A.: *Generalized Linear Models*. 2nd Edition, Chapman and Hall, London., <http://dx.doi.org/10.1007/978-1-4899-3242-6> (1989)
- [12] Schaefer, Robert L.: *Bias Correction in Maximum Likelihood Logistic Regression*. *Statistics in Medicine* 2:71-78. (1983)
- [13] Prentice, R. L., and R. Pyke: *Logistic Disease Incidence Models and Case-Control Studies*. *Biometrika* 66:403-411. (1979)
- [14] Vékás Péter: *Összefüggő biztosítási kockázatok modellezése (műhelytanulmány)*, Budapesti Corvinus Egyetem. URL: [http://unipub.lib.uni-corvinus.hu/2093/1/VekasPeter\\_Osszefuggo\\_wp.pdf](http://unipub.lib.uni-corvinus.hu/2093/1/VekasPeter_Osszefuggo_wp.pdf) (2012)
- [15] Xie, Yu, and Charles F. Manski.: *The Logit Model and Response-Based Samples*. *Sociological Methods and Research* 17(3):283-302. (1989)

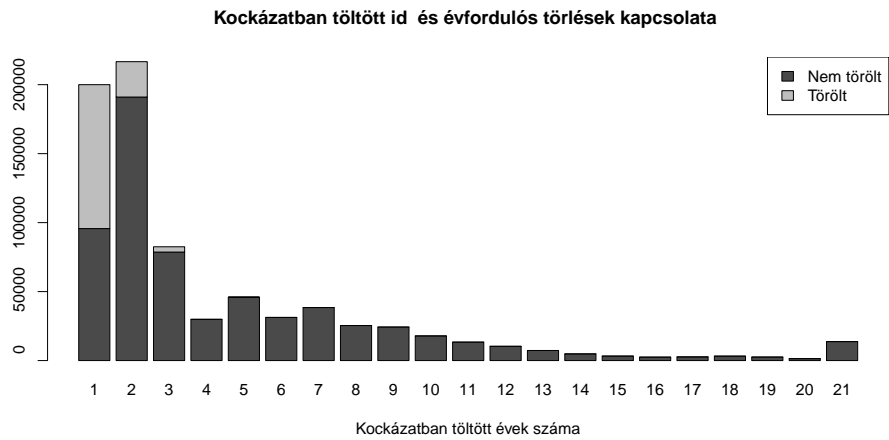
**R-csomagok:**

- [16] Alessandro Barbiero, Pier Alda Ferrari, *GenOrd*: Simulation of Discrete Random Variables with Given Correlation Matrix and Marginal Distributions
- [17] G. Grothendieck, *squidf*: Manipulate R Data Frames Using SQL
- [18] James Honaker, *Zelig-relogit-class*: Rare Events Logistic Regression For Dichotomous Dependent Variables
- [19] Thomas Kvalnes, *nearPD*: Find Nearest Positive Definite Matrix

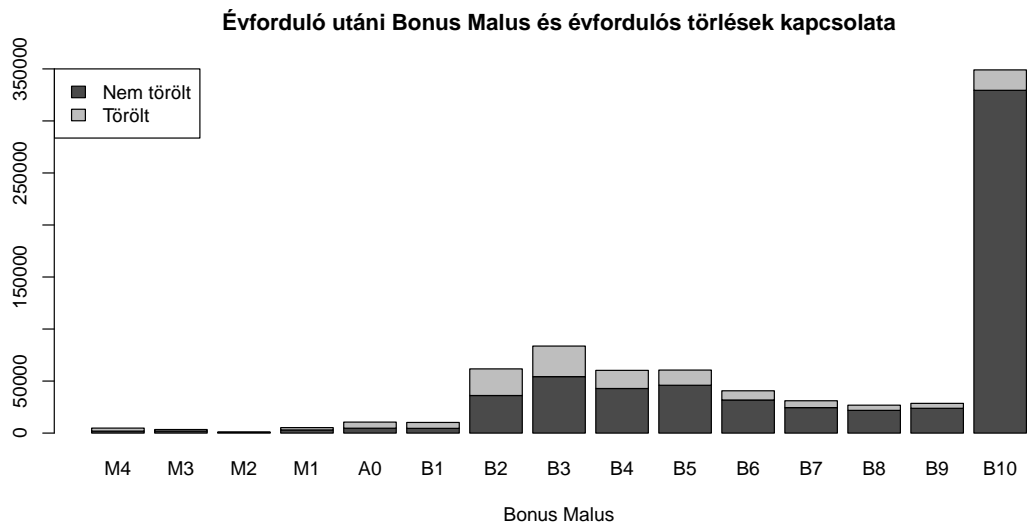
# Függelék



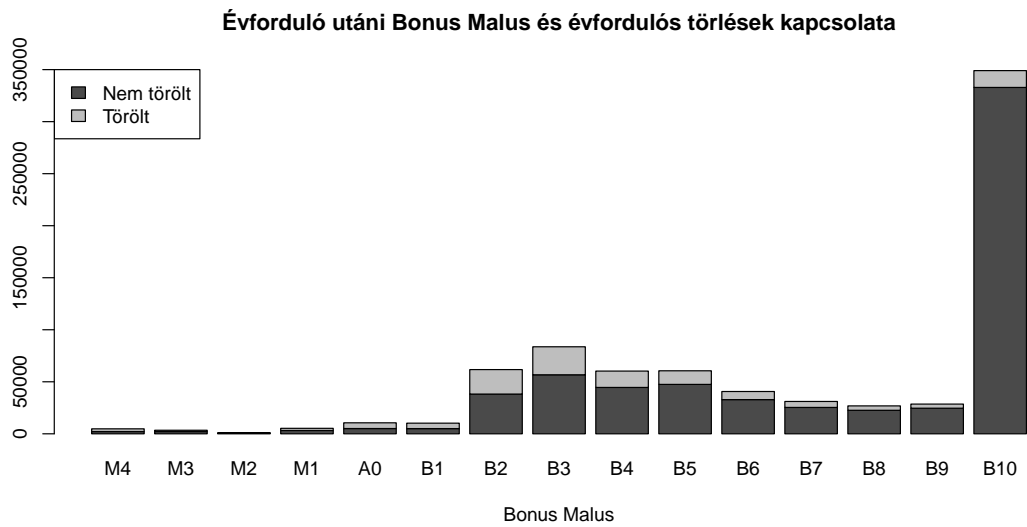
2.14. ábra. Kockázatban töltött évek és törlések kapcsolata – első évforduló, első szcenárió.



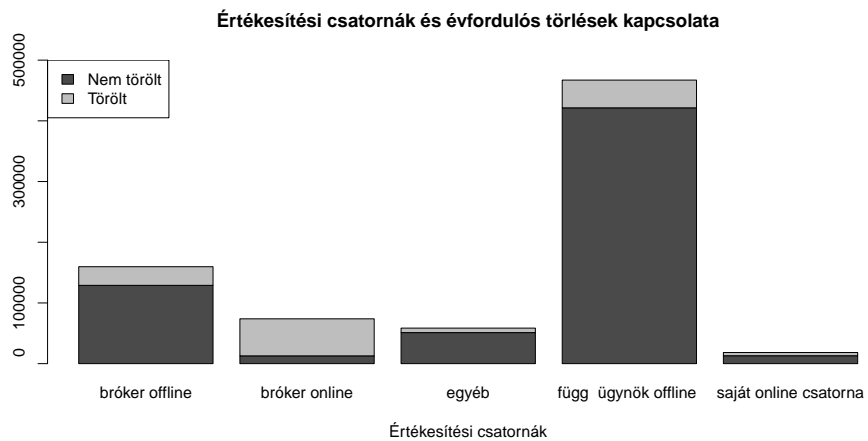
2.15. ábra. Kockázatban töltött évek és törlések kapcsolata – első évforduló, második szcenárió.



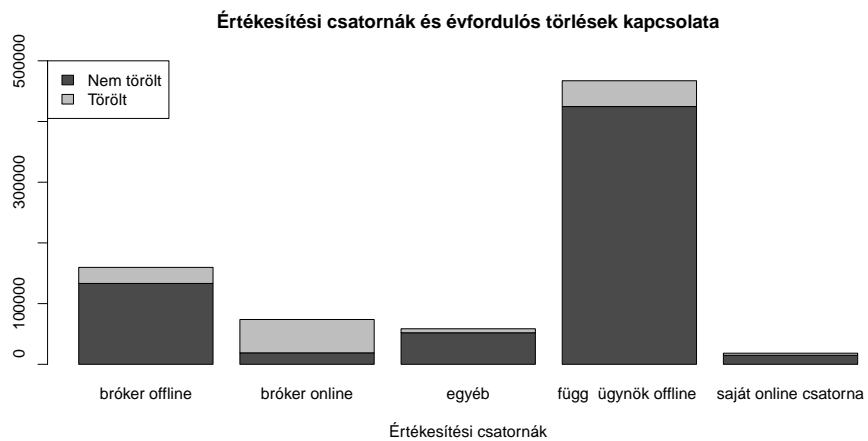
2.16. ábra. Bonus Malus és törlések kapcsolata – első évforduló, első scenárió.



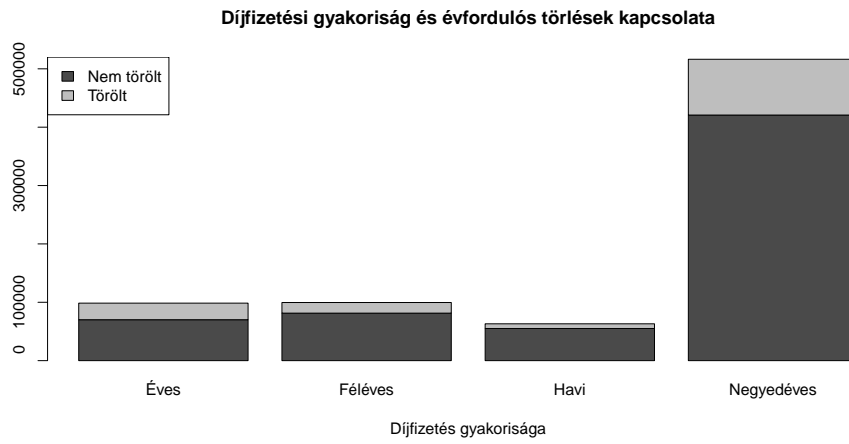
2.17. ábra. Bonus Malus és törlések kapcsolata – első évforduló, második scenárió.



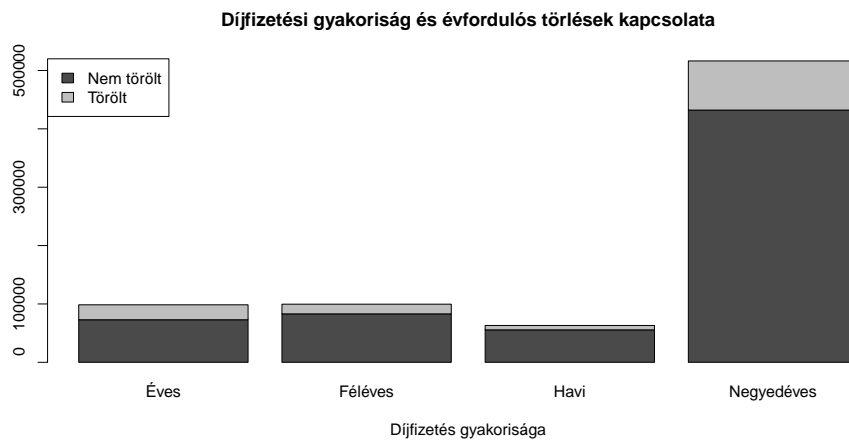
2.18. ábra. Értékesítési csatornák és törlések kapcsolata – első évforduló, első scenárió.



2.19. ábra. Értékesítési csatornák és törlések kapcsolata – első évforduló, második scenárió.

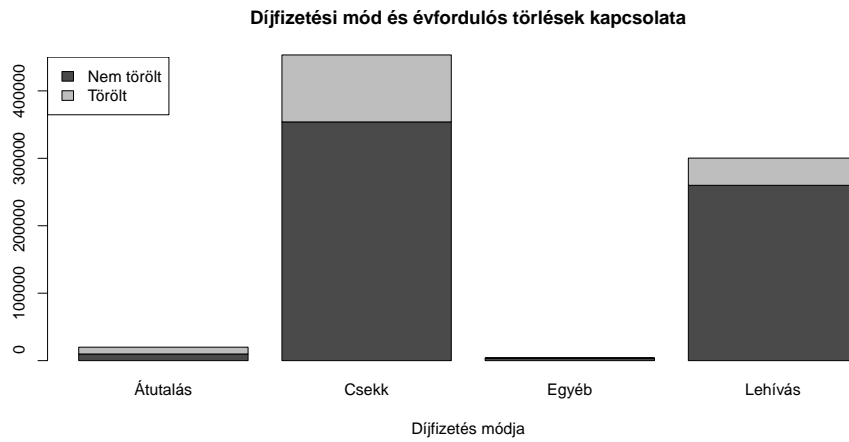


2.20. ábra. Díjfizetési gyakoriság és törlések kapcsolata – első évforduló, első scenárió.

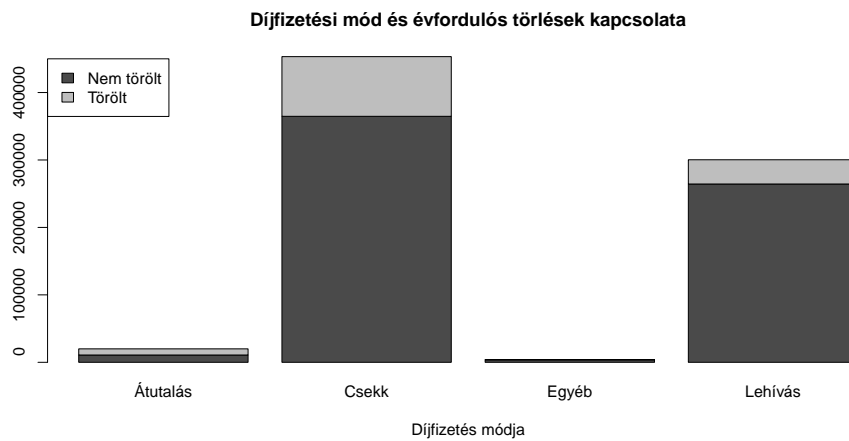


2.21. ábra. Díjfizetési gyakoriság és törlések kapcsolata – első évforduló, második scenárió.

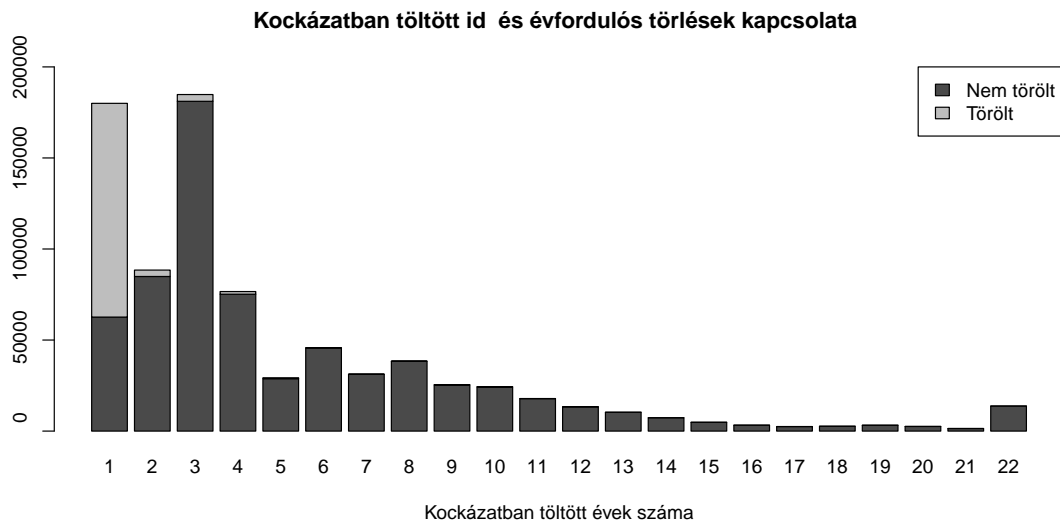




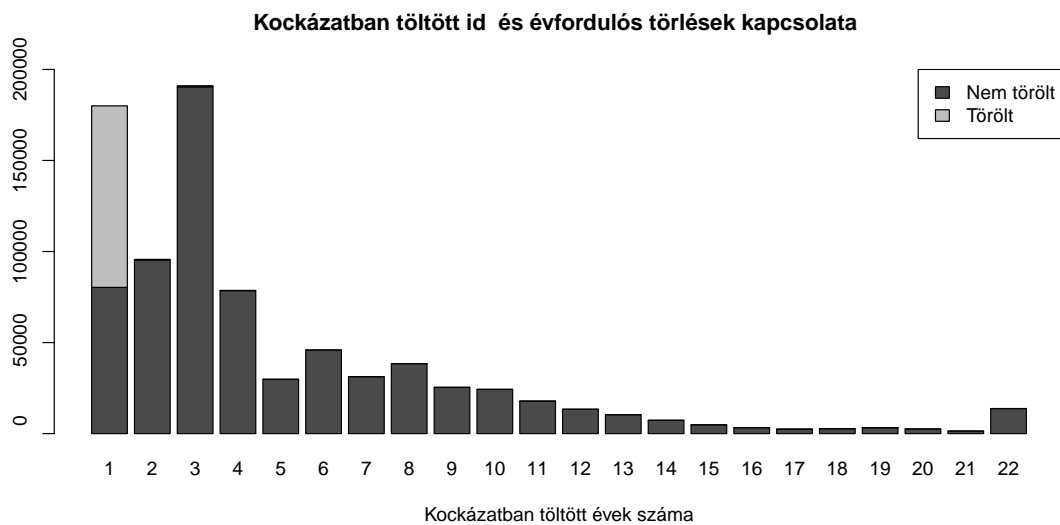
2.22. ábra. Díjfizetés módja és törlések kapcsolata – első évforduló, első scenárió.



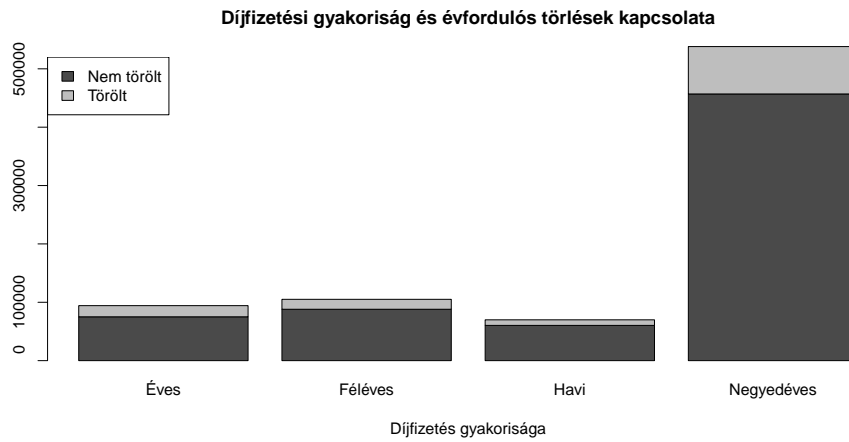
2.23. ábra. Díjfizetés módja és törlések kapcsolata – első évforduló, második scenárió.



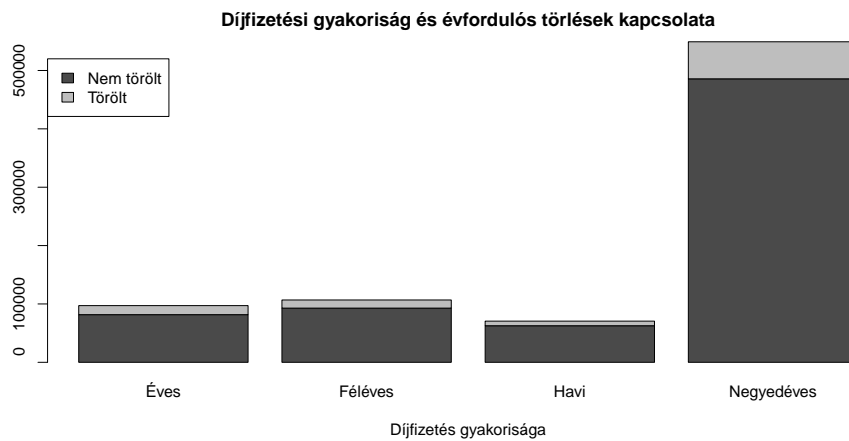
2.24. ábra. Kockázatban töltött évek és törlések kapcsolata – második évforduló, első szcenárió.



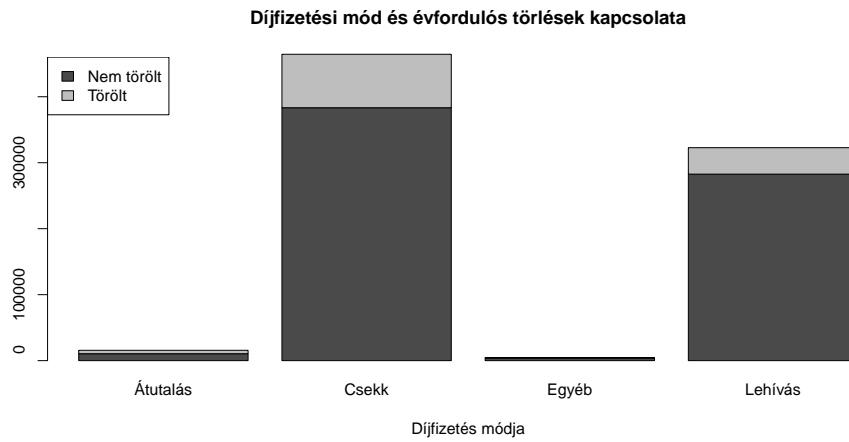
2.25. ábra. Kockázatban töltött évek és törlések kapcsolata – második évforduló, második szcenárió.



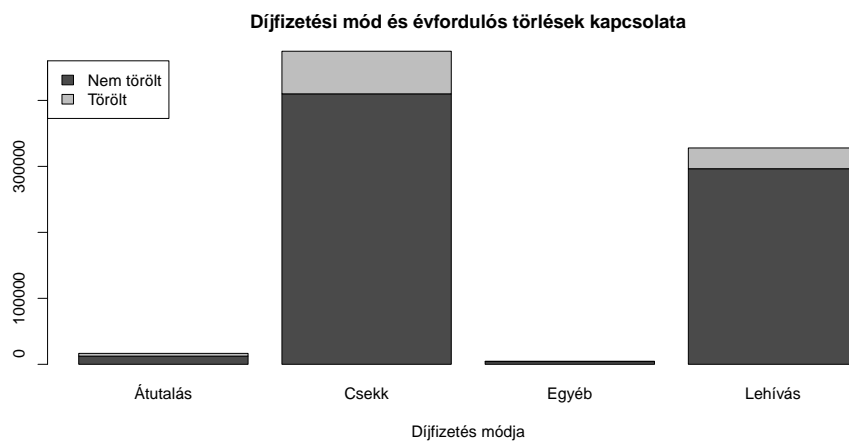
2.26. ábra. Díjfizetési gyakoriság és törlések kapcsolata – második évforduló, első scenárió.



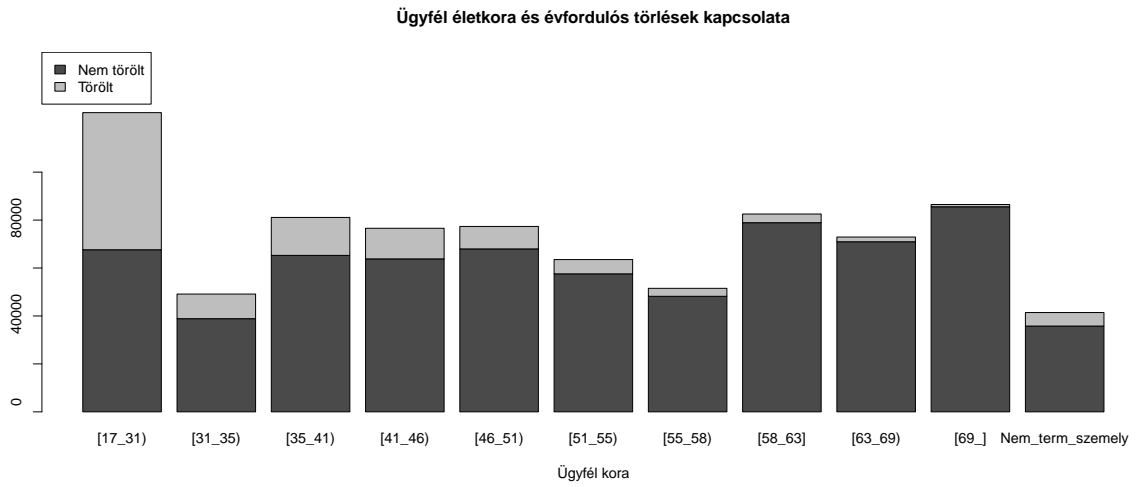
2.27. ábra. Díjfizetési gyakoriság és törlések kapcsolata – második évforduló, második scenárió.



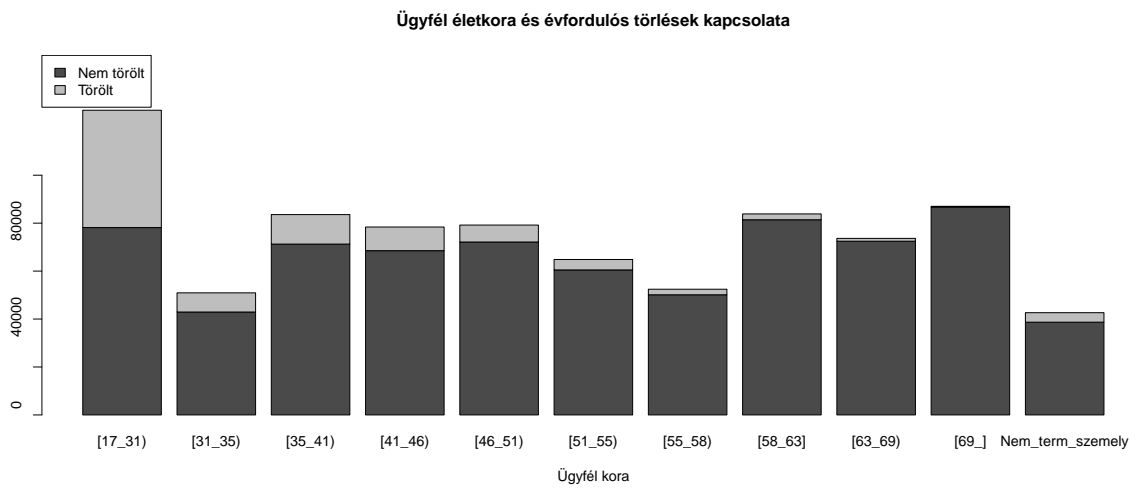
2.28. ábra. Díjfizetés módja és törlések kapcsolata – második évforduló, első scenárió.



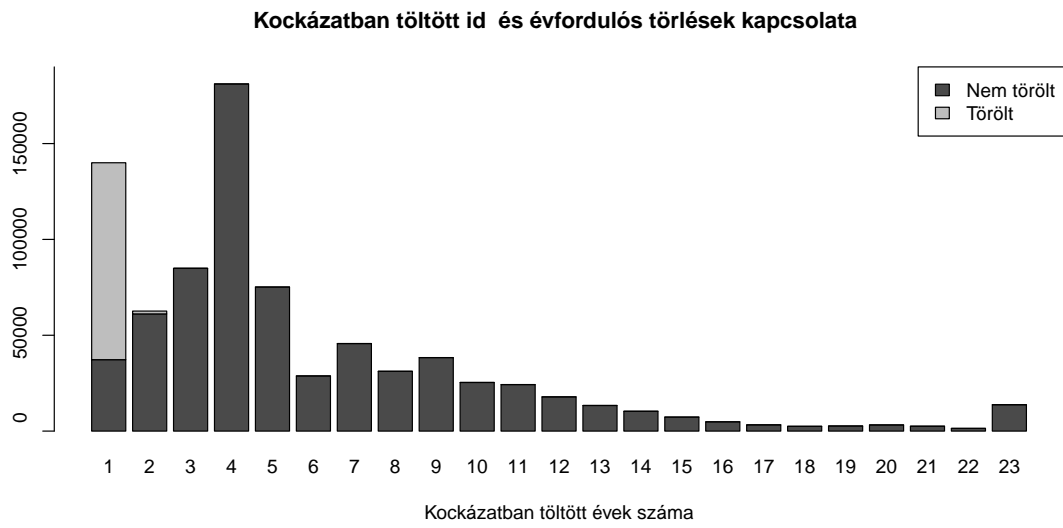
2.29. ábra. Díjfizetés módja és törlések kapcsolata – második évforduló, második scenárió.



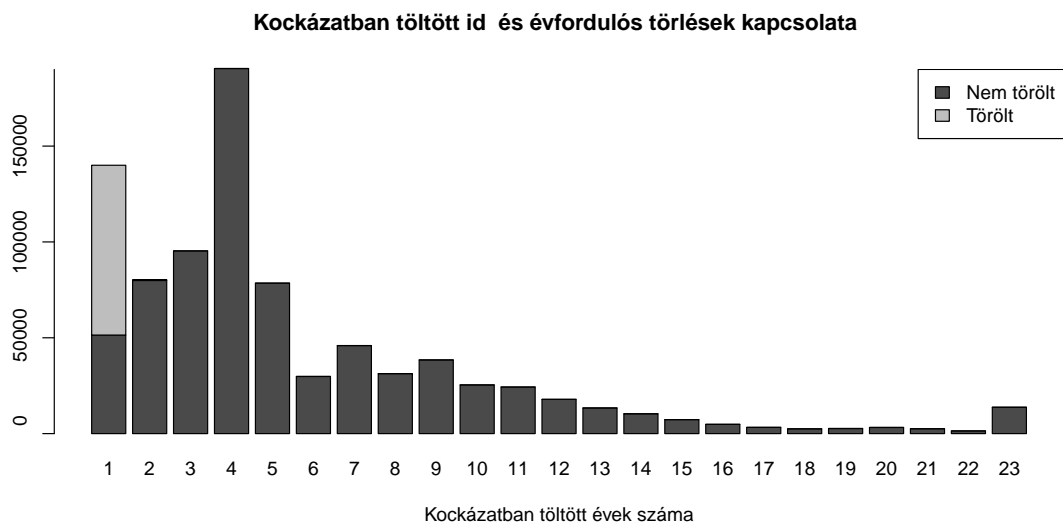
2.30. ábra. Életkor és törlések kapcsolata – második évforduló, első scenárió.



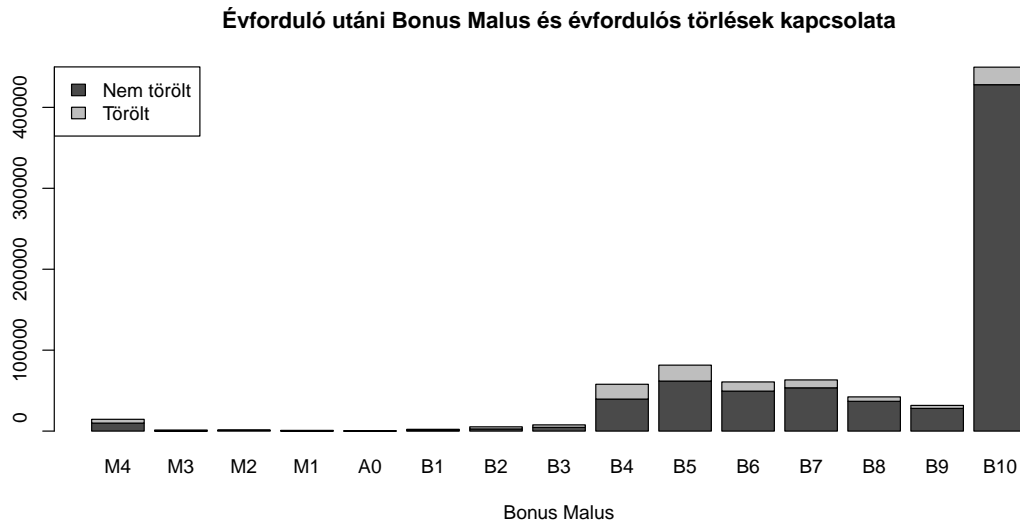
2.31. ábra. Életkor és törlések kapcsolata – második évforduló, második scenárió.



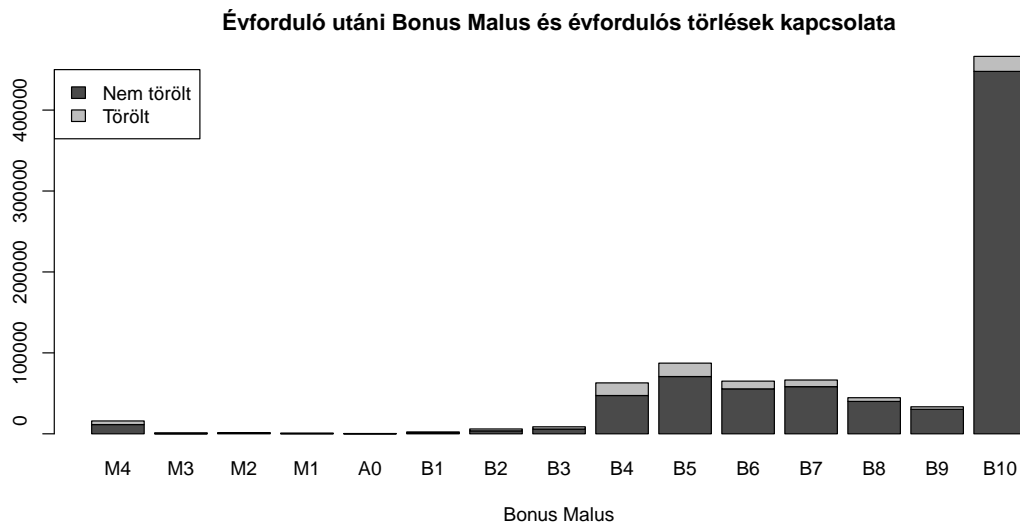
2.32. ábra. Kockázatban töltött évek és törlések kapcsolata – harmadik évforduló, első szcenárió.



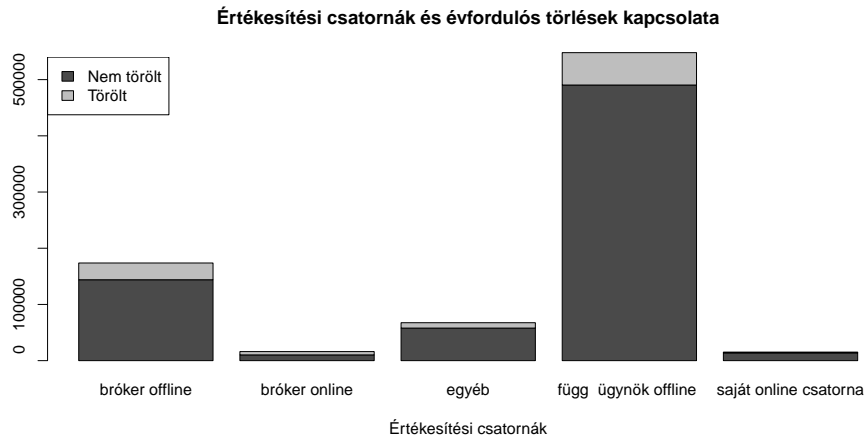
2.33. ábra. Kockázatban töltött évek és törlések kapcsolata – harmadik évforduló, második szcenárió..



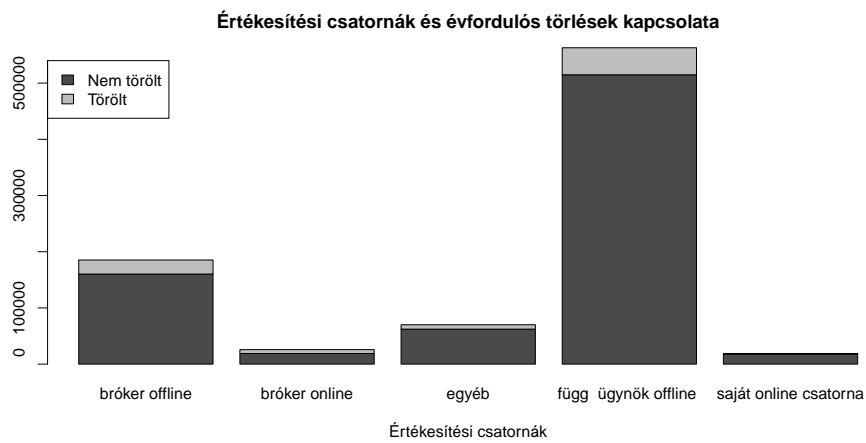
2.34. ábra. Bonus Malus és törlések kapcsolata – harmadik évforduló, első scenárió.



2.35. ábra. Bonus Malus és törlések kapcsolata – harmadik évforduló, második scenárió.

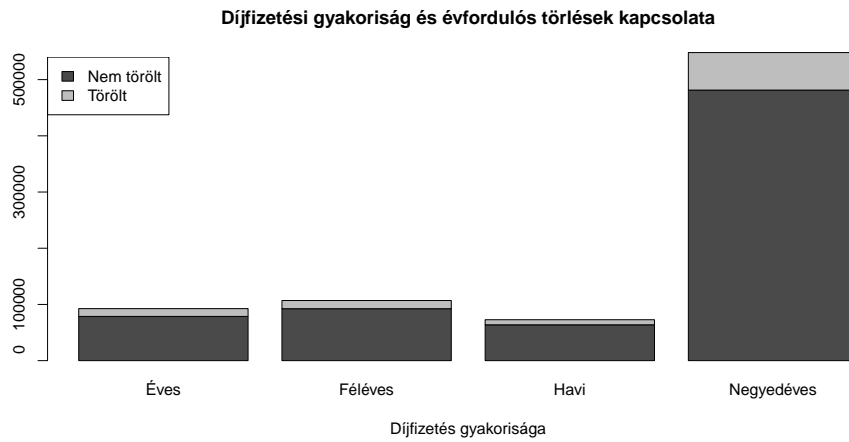


2.36. ábra. Értékesítési csatornák és törlések kapcsolata – harmadik évforduló, első szcenárió.

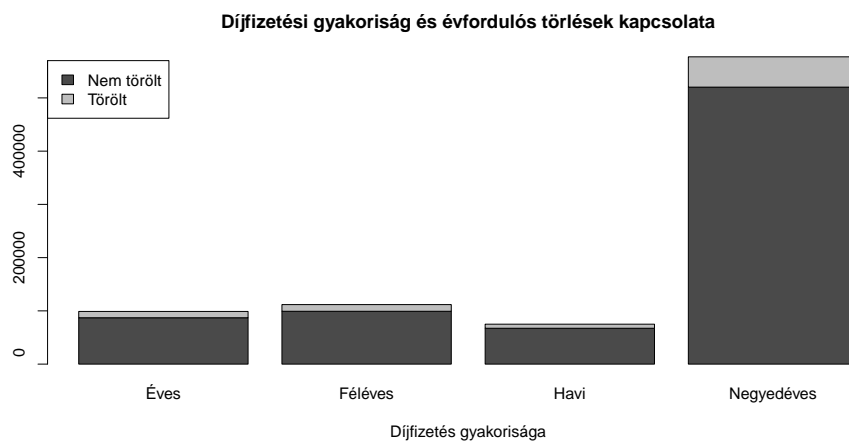


2.37. ábra. Értékesítési csatornák és törlések kapcsolata – harmadik évforduló, második szcenárió.

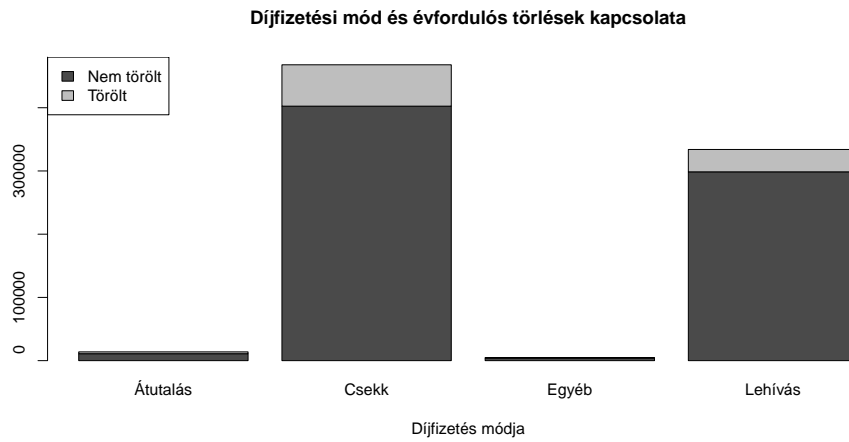




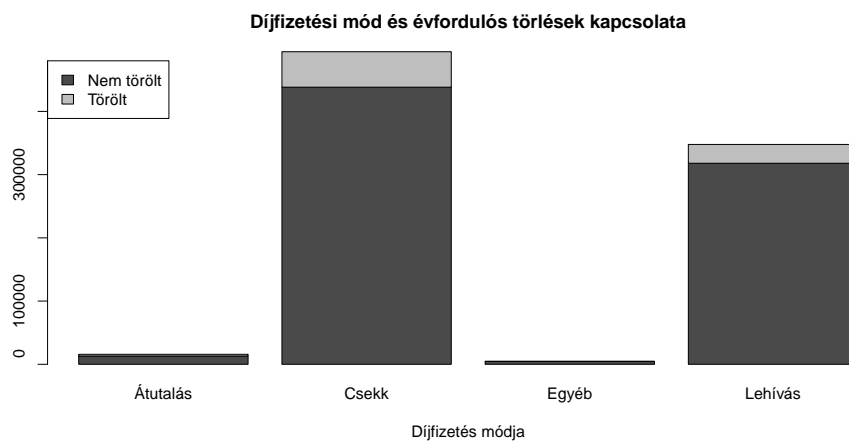
2.38. ábra. Díjfizetési gyakoriság és törlések kapcsolata – harmadik évforduló, első szcenárió.



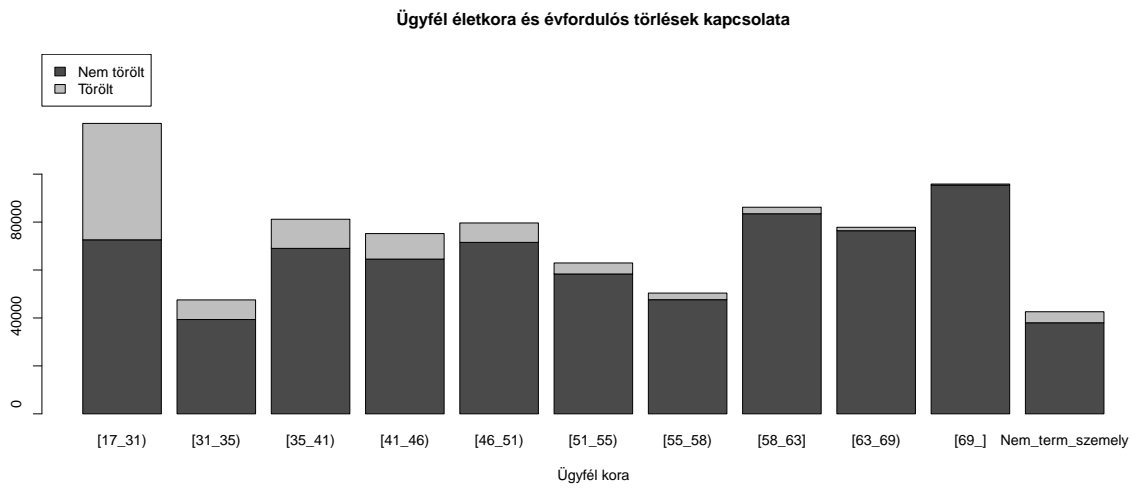
2.39. ábra. Díjfizetési gyakoriság és törlések kapcsolata – harmadik évforduló, második szcenárió.



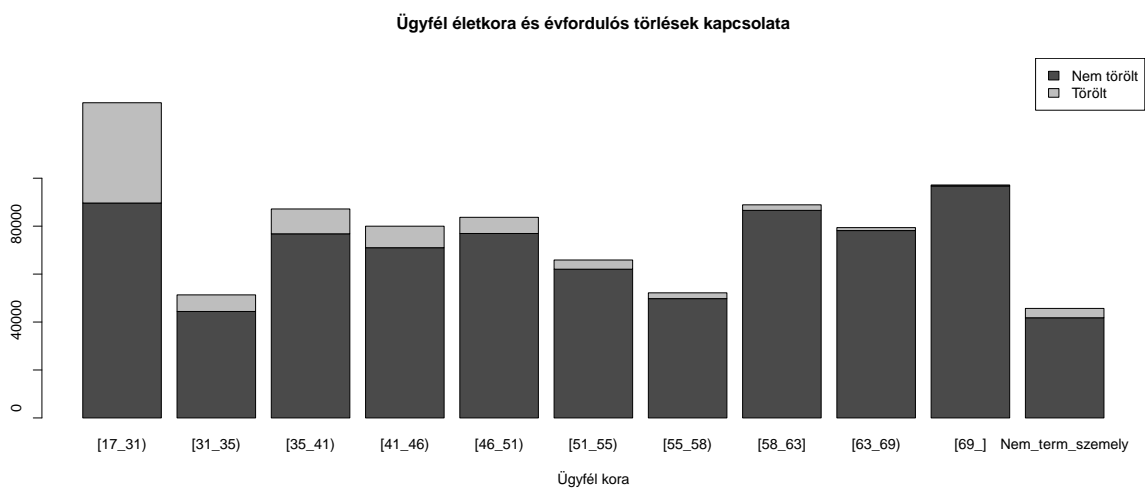
2.40. ábra. Díjfizetés módja és törlések kapcsolata – harmadik évforduló, első scenárió.



2.41. ábra. Díjfizetés módja és törlések kapcsolata – harmadik évforduló, második scenárió.



2.42. ábra. Életkor és törlések kapcsolata – harmadik évforduló, első scenárió.



2.43. ábra. Életkor és törlések kapcsolata – harmadik évforduló, második scenárió.