A 60 éves ötös lottó tapasztalatai, tévhitek és tények*

Szőke Szilvia,

a Debreceni Egyetem egyetemi adjunktusa

E-mail: szoke.szilvia@econ.unideb.hu

Huzsvai László,

a Debreceni Egyetem habilitált egyetemi docense

E-mail: huzsvai.laszlo@econ.unideb.hu

A második világháború után újraszervezett ötös lottó 2017 márciusában volt hatvanéves. Ebben a népszerű szerencsejátékban természetesen mindenki nyerni szeretne, ezért sok helyen – főként az interneten – különböző, minden tudományosságot mellőző taktikákat találhatunk. Ezek a tanácsok tele vannak pontatlansággal és megbízhatatlansággal, ami az elégtelen megfigyelésből, a reprezentativitás hiányából fakad. Sokan úgy gondolják, hogy az eddig húzott számok alapján előre jelezhetők a későbbi nyertes számok. Amikor bizonyos lottószámokat hónapok óta nem sorsoltak ki, akkor tévesen azt hiszik, hogy a következő héten megnövekszik az előfordulásuk valószínűsége. Más megközelítésben, a már kihúzott számokra hiszik azt, hogy hosszabb ideig nem fognak előfordulni. Mások a nyeremények maximalizálásához adnak tanácsokat. Szerintük, akik elkerülik az alacsonyabb értékeket – mert azok születésnapokkal esnek egybe – vagy a szerencseszámokat (3, 7 stb.), maximalizálni tudják a nyereményeiket.

A szerzők jelen munkájukban a Szerencsejáték Zrt. adatai alapján, egzakt matematikai statisztikai módszerekkel tisztázzák az ötös lottó körül kialakul néhány tévhitet, legendát, és bizonyítják a tényeket.

TÁRGYSZÓ:
Lottó.
Valószínűségszámítás.
Hipergeometrikus eloszlás.

DOI: 10.20311/stat2017.10.hu0976

* A szerzők ezúton fejezik ki köszönetüket a Szerencsejáték Zrt. felé a készségesen megküldött adatokért.

A második világháború után újraszervezett ötös lottó első húzását 1957. március 7-én tartották meg. A szerencsejátékok népszerűségét bizonyító adat szerint „a felnőtt magyar lakosság valamivel több, mint fele (54 százaléka, körülbelül 4,3 millió fő) rendszeresen (legalább 2-3 havonta egyszer) játszik a Szerencsjáték Zrt. valamelyik játékával” (*Kutatópont* [2011]). A magyarországi szerencsejátékokat és ezen belül a lottó játékot a *Statisztikai Szemle* hasábjain is elemezték már: *Tessényi–Kazár* [2012] a szerencsejáték-vásárlási szokásokat, míg *Tessényi–Kovács* [2011] a szerencsejáték-függés kérdését járták körül.

Mindenki szeretne nyerni, azonban a nyerés valószínűsége nagyon kicsi. Ezért különböző, minden tudományosságot mellőző taktikákat ajánlanak. Ilyen tanács például az „egyetlen szelvény, egyetlen mező”; „a számok legyenek minél szélsőségesebbek, olyanok, amiket remélhetőleg senki más nem játszik meg”. Jó tanács, de honnan tudhatnánk, hogy mik a mások által preferált számok?

Olyan számokat kell választani, amelyeket ha kihúznak, mások valószínűleg ritkábban játszanak meg. Több internetes felületen olvashatjuk, hogy kerülni kell a születésnapoknak megfelelő számokat, mert a 31 alattiak kihúzásakor általában sok részre osztódik a főnyeremény, hiszen sokan játszanak a hónapok 1–12 közötti és a napok 1–31 közötti számaival. Magyarországon az egyik legkedveltebb kombináció az 1, 2, 3, 4, 5 lehetett egy ideig, főleg a „Valami Amerika” című film óta.

Az is érdekes, hogy egyesek számsorozatokat játszanak meg, de előfordul sokszor az előző heti nyerőszámok megtétele is. Vajon a hatvan év alatt ezt a taktikát alkalmazva mennyit nyertünk volna? Mennyire ismétlődnek a korábban kihúzott számok? *Fegyverneki Sándor* a Miskolci Egyetem tanszékvezetője szerint „annak a valószínűsége, hogy újra kihúzzák ugyanazt az öt számot a lottón, ami egyszer már telitalálatnak bizonyult, egyre növekszik”, olvashatjuk a *Heti Világgazdaság*nak adott interjújában (*Szegő* [2011]). Véleménye szerint: „a jelenlegi, alig több mint 3 ezer összes húzás alapján mindössze 10 százalék a valószínűsége annak, hogy egy korábban már nyertes kombináció ismét előfordul, de körülbelül 7 800 húzás után ez az esély már 50 százalékra nő. Sőt, ha 31 800 húzást végeznének el, akkor már 99,9 százalék fölé szökik a valószínűsége annak, hogy egy korábbi telitalálat ismét előfordul”.

A média megemlíti, hogy Magyarországon az eddigi legjobb egyezés két nyerőszámsorozat között 2003-ban következett be. Akkor egy sorsoláson a hetvenes évekbeli telitalálat négy számát is kihúzták újra. Tényleg csak egyszer fordult elő ilyen a lottó történetében?

A nagyon kicsi nyerési valószínűség miatt egyes emberekben felmerül, hogy „biztosan csalnak, mert ennyi pénzt nem fizetnek ki, valaki ezt zsebre teszi, aki közel van a tűzhöz”. Mások ezt ostobaságnak tartják. Ugyanis, ha egy üzlet jól megy, akkor nincs szükség csalásra, sőt, a fedhetetlenség a legnagyobb érdek.

Jelen dolgozatunkban az ezekhez hasonló néhány „legenda” vizsgálatára vállalkoztunk egzakt matematikai módszerekkel, a Szerencsejáték Zrt. adatai alapján. Az elméleti eloszlások meghatározása után kiszámítottuk a valószínűségeket és a gyakoriságokat, valamint összehasonlítottuk ez eddigi sorsolások tényadataival.

A dolgozat megírása során gondoltunk a laikus érdeklődőkre és a valószínűségszámítással ismerkedő kezdőkre, ezért a magyarázatokat olyan részletesen próbáltuk leírni, hogy ők is követni tudják a gondolatmenetet. Azonban mindez a szakmai alaposság rovására mehet, amit próbáltunk elkerülni, és reméljük a témában jártasabb olvasóknak is tudtunk új, értékes információkat nyújtani.

Amikor bizonyos lottószámokat hónapok óta nem sorsoltak ki, akkor tévesen azt hiszik, hogy a következő héten megnövekszik az előfordulásuk valószínűsége. Más megközelítésben, a már kihúzott számokról hiszik azt, hogy hosszabb ideig nem fognak előfordulni. A nemzetközi szakirodalomban a „szerencsejátékosok tévedése” (gambler’s fallacy), más néven a „Monte-Carlo-tévedés” néven ismert ez a jelenség. Sok kutató szentelt ennek figyelmet, vizsgálták lélektani, viselkedési, illetve gazdasági szempontok alapján is. *Militana–Wolfson–Cleaveland* [2010] számítógépes játékkal önkénteseken végzett kísérleteik alapján úgy gondolják, hogy ez nem velünk született, ösztönös tulajdonságunk, azonban a tanulási folyamat pontos magyarázatának megadására a kevés kísérleti adat miatt nem vállalkoztak.

Suetens–Tyran [2012] a dán állami lottóval kapcsolatosan nemek szerint tanulmányozták a fogadók hiedelmeit. Megállapították, hogy többnyire a férfiakat befolyásolják a választásban az előzőleg kihúzott számok, a nők inkább ragaszkodnak a kedvenc számaikhoz.

Lien–Yuan [2015] a lottózással kapcsolatos tévhiteket vizsgálták, az interneten található tanácsok igazságának próbáltak utánajárni, illetve, hogy a játékosok mennyire fogadják meg ezeket a javaslatokat. A vizsgálatokat olyan lottójáték adatain végezték, ahol 33 számból hatot kellett eltalálni a telitalálathoz. A szerzők a játszott számok között távolságfüggvényt definiálva hasonlították össze a húzott és a megjátszott számokat. Tapasztalataik alapján az internetes tanácsokat nagyon sokan megfogadják, és túlságosan egyenletes távolságra adják meg tippjeiket. Összehasonlítva a sorsolt számok közötti távolságokkal azt kapták, hogy azok kevésbé egyenletesen helyezkednek el.

Simon [1999] szerint az Egyesült Királyságban eladott állami lottó 10 százalékában a legnépszerűbb számok 1 százalékát játsszák meg, azaz a legnépszerűbb számokat tízszer gyakrabban játsszák meg a véletlen számválasztáshoz képest. Ez lényegesen csökkenti a nyeremény várható összegét. Számításai szerint a népszerű számok megjátszásával a várható hozam a maximális nyeremény 18 százalékára is csökkenhet, míg kevésbé népszerű számokkal ez 87 százalék.

Xu–Harvey [2014] online sportfogadás adatai alapján a nyerőszéria hiedelmet vizsgálták, vagyis azt, hogy a szerencsejátékosok hisznek abban, hogy egy véletlen nyeremény után nagyobb az esélyük az újabb nyerésre. A vizsgált adatok alátámasztották ezt az elképzelést, de a kutatók az okokat abban látták, hogy nyerés után kevésbé kockázatos fogadásokat kötnek a játékosok, míg vesztés után éppen ellenkező tendenciát mutat a viselkedésük, és nem a szerencse áll melléjük.

1. Adatbázis és módszer

Az ötös lottó eddigi húzási eredményei a Szerencsejáték Zrt. honlapjáról származnak. A https://bet.szerencsejatek.hu/jatekok/otoslotto/sorsolasok helyről mindenki letöltheti az adatokat, és szabadon felhasználhatja őket. Jelen cikkben a feldolgozott adatok 1957. év 10. hetétől 2017. év 6. hétig terjedő időszakot ölelik fel, összesen 3 128 hét húzási eredményeit tartalmazzák. Ez a mintanagyság már alkalmas arra, hogy megbízható statisztikai elemzéseket végezzünk.

Sajnos a honlapon található adatbázis erősen hiányos, mivel a nyereményjegyzéket csak 1998-tól tartalmazza napjainkig. A Szerencsejáték Zrt. ügyfélszolgálatának írásos megkeresése után a hiányzó adatokat készségesen megküldték kizárólag tudományos elemzés céljára. Az utólagosan beszerzett információk azonban a heti rendszeres húzások mellett további extra, jubileumi, televíziós stb. húzások eredményeit is tartalmazták. Összeses 102 ilyen sorsolási adat volt az adatbázisban. Ezeket azonban nem használtuk fel, csak a rendszeres sorsolások 3 128 hetének nyerőszámait.

A modellszámításokat, az ábrákat és a statisztikai elemzéseket az R 3.3.2 változatával végeztük, grafikus környezetnek az RStudio 1.0.136 verzióját használtuk. A programmal közel tízmillió elemet tartalmazó mátrixműveletek is gyorsan elvégezhetők egy nyolcmagos processzoron.

A vizsgálatok eredményeinek statisztikai igazolására χ^2-próbát és Kolmogorov–Smirnov-tesztet alkalmaztunk. Vannak esetek, amikor mindkét eljárás megfelelő, mi ilyenkor az alkalmasabb próbát választottuk az adott probléma megválaszolására.

1.1. A χ^2-próba és a Kolmogorov–Smirnov-teszt alkalmazásáról

A nemparaméteres χ^2-próba diszkrét eloszlásokkal kapcsolatos kérdések megválaszolására alkalmas, melyről a különböző statisztikai szakkönyvekből részletesen tájékozódhatunk. Alkalmazhatjuk homogenitás- (két valószínűségi változó egyforma eloszlású-e vagy sem), illeszkedés- (a minta eloszlását hasonlítjuk egy elméleti eloszláshoz) és függetlenségvizsgálat (két tulajdonság szerint osztályozott mintaelemek függetlenek-e) esetén. A lottózással kapcsolatos számításaink során illeszkedésvizsgálatokat végeztünk.

1.2. Tiszta illeszkedésvizsgálat

A χ^2-próba struktúrája:

$$\chi^2 = \sum_{i=1}^{r} \frac{\left(megfigyelt\ gyakoriságok - várt\ érték\right)^2}{várt\ érték},$$

$$\chi^2 = \sum_{i=1}^{r} \frac{\left(k_i - np_i\right)^2}{np_i},$$

ahol k_i az i-edik osztály megfigyelt gyakorisága, p_i az i-edik osztályba kerülés valószínűsége, n a minta elemszáma.

A megfigyelt gyakoriságok, ha messze vannak a várttól, akkor χ^2 értéke nagy lesz, ha azonban közel, akkor kicsi. Így a χ^2 megad egy értéket a megfigyelt és a várt gyakoriságok távolságának mérésére. Ez a statisztika a kiindulási hipotézis (nullhipotézis: H_0) fennállása esetén $r-1$ szabadsági fokú χ^2-eloszlást követ (*Vincze–Varbanova* [1993]). A döntést a számított és a χ^2-eloszlás táblázatbeli értékek összehasonlításával hozhatjuk meg.

A χ^2-próba hátránya, hogy kis mintánál nem alkalmazható, valamint a legkisebb p_i valószínűségek meghatározásához legalább az $np_i > 5$ feltételnek teljesülnie kell, de még jobb, ha az $np_i > 10$ (*Hunyadi–Mundruczó–Vita* [1996]).

A Kolmogorov–Smirnov-teszt az eloszlások homogenitásvizsgálatára alkalmas. Ez a teszt a χ^2-próbával szemben kis elemszámú minták esetén is jól használható, a statisztikai próba ereje ilyenkor nagyobb. Alkalmazható folytonos és diszkrét eloszlások esetén is. Létezik egymintás és kétmintás változata is. Az egymintás változatot egy adott nevezetes eloszlás tesztelésére használjuk, ilyenkor a mintából becsüljük az eloszlás paramétereit, és az elméleti, valamint tapasztalati kumulált gyakoriságokat hasonlítjuk össze, azaz a kumulatív eloszlásfüggvényeket. A kétmintás változat két minta kumulatív gyakoriságát hasonlítja össze. A H_0: a két minta kumulatív gyakorisága nem tér el szignifikánsan. Természetesen a tévedés lehetősége ennél a tesztnél is fennáll, aminek a valószínűségét, a gyakorlati tapasztalatok alapján, 5 százalékban maximalizáltuk. Amennyiben a számítások során ez a valószínűség (*p*-érték) 5 százaléknál kisebb lesz, akkor a H_0-t elutasítjuk, és nem tudjuk megerősíteni a gyakoriságok azonosságát. Abban az esetben, ha 5 százaléknál nagyobb valószínűséget kapunk, akkor a H_0-t igazoljuk, azaz a két minta gyakorisága azonosnak tekinthető. Annak, hogy mennyivel nagyobb ez a valószínűség, mint 5 százalék, nincs jelentősége, nem lehet azt mondani, hogy a 90 százalékos valószínűség jobban igazolja a H_0-t, mint a 80 százalékos. Ezek valójában nem a H_0 igaz valószínűségei, hanem feltételes valószínűségek. Feltételezzük, hogy a H_0 igaz, és megnézzük, hogy a mintáink milyen valószínűséggel fordulhatnak elő ebben az esetben. Gyakorlatilag ez egy közvetett bizonyítás. A Kolmogorov–Smirnov-teszt statisztikája a megfigyelt és teoretikus kumulált eloszlásfüggvények közötti legnagyobb abszolút különbségből számítható. Ezt az értéket a megfigyelések négyzetgyökével kell szorozni.

A teszt hátrányaként szokták megemlíteni, hogy érzékeny a helyzeti különbségekre és az eloszlások alakjára. A helyzeti különbség azt jelenti, hogy a két eloszlás hol helyezkedik el a skálán. A Kolmogorov–Smirnov-teszt akkor is különbözőnek mutatja a két eloszlást, ha az alakjuk megegyezik, de egymástól távol helyezkednek el. Ezek szerint két eloszlás akkor különbözik, ha az alakjuk vagy az elhelyezkedésük különbözik, illetve mindkettő. Amennyiben a két eloszlás helyzeti különbsége nem érdekel bennünket, toljuk el a skálát az origóra, aminek a legegyszerűbb módja az adatok standardizálása (ettől az eloszlások alakja semmit sem változik).

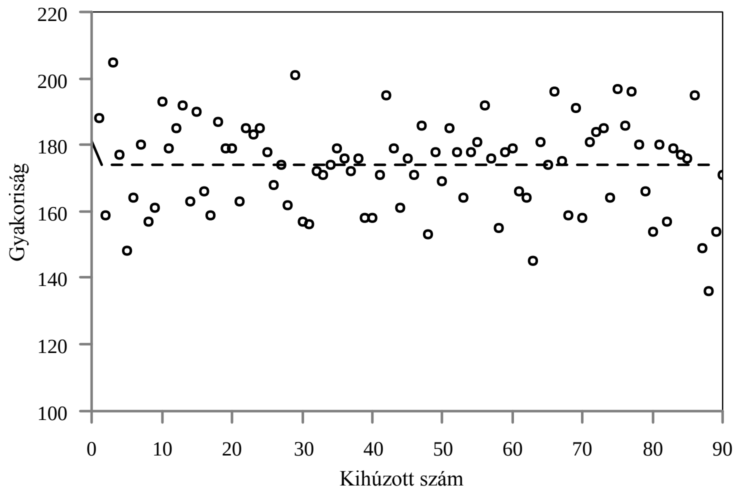
2. Eredmények

Ebben a fejezetben az előzőkben felvetett kérdések részletes vizsgálata olvasható. A modellezett elméleti eloszlásokat összevetettük a tapasztalati, empirikus értékekkel.

2.1. A számok húzásának véletlenszerűsége

Első lépésben megvizsgáltuk az eddig kihúzott számok empirikus gyakoriságát, melyet az 1. ábra mutat.

1. ábra. A kihúzott számok empirikus gyakorisága 3 128 játékhét alapján



A vízszintes szaggatott vonal az empirikus gyakoriságok átlagát mutatja, amely 173,7778 volt. Ezután meghatároztuk egy adott szám kihúzásának elméleti és empirikus valószínűségét. Egy adott szám kihúzásának elméleti valószínűsége a visszatevés nélküli mintavételezés képletét felhasználva:

$$P(\textit{egy adott szám húzása}) = \frac{\binom{1}{1}\cdot\binom{89}{4}}{\binom{90}{5}} = \frac{2\,441\,626}{43\,949\,268} = 0{,}05556\,.$$

Az empirikus valószínűség:

$$P = \frac{173{,}7778}{3128} = 0{,}05556\,.$$

A két érték tökéletes egyezést mutat, tehát a tapasztalati gyakoriság hűen követi az elméletit. Ennek ismeretében a húzások során semmilyen manipuláció nem tételezhető fel. Ezt megerősíti az elméleti és a tapasztalati várható érték, valamint szórás összehasonlítása. Az elméleti értékek könnyen számíthatók, hiszen a számok 1-től 90-ig egyenletes eloszlásúak, ezeket a tapasztalati értékekkel az 1. táblázatban vetettük össze.

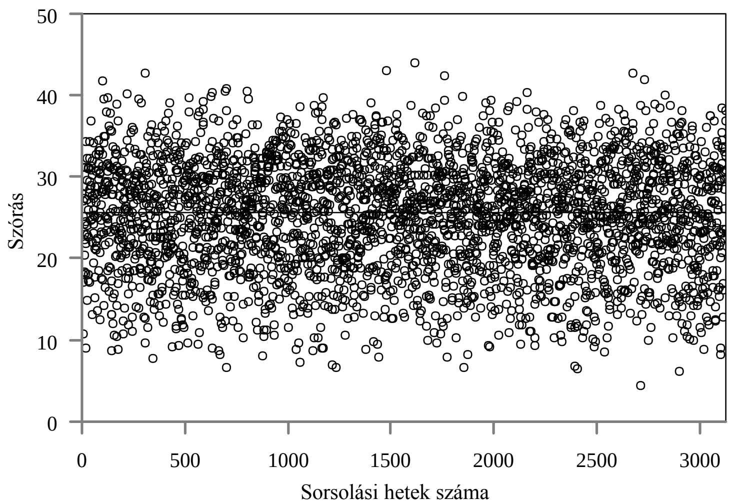
1. táblázat

A lottószámokat jellemző leíró statisztikai mutatók

Mutató	Elméleti	Tapasztalati
Várható érték	45,50	45,33
Szórás	25,98	25,93

A szabályos, véletlenszerű sorsolást bizonyítja a homoszkedaszticitás vizsgálata is. A heti szórások alakulását mutatja a 2. ábra. A vízszintes vonal az elméleti szórás értéke.

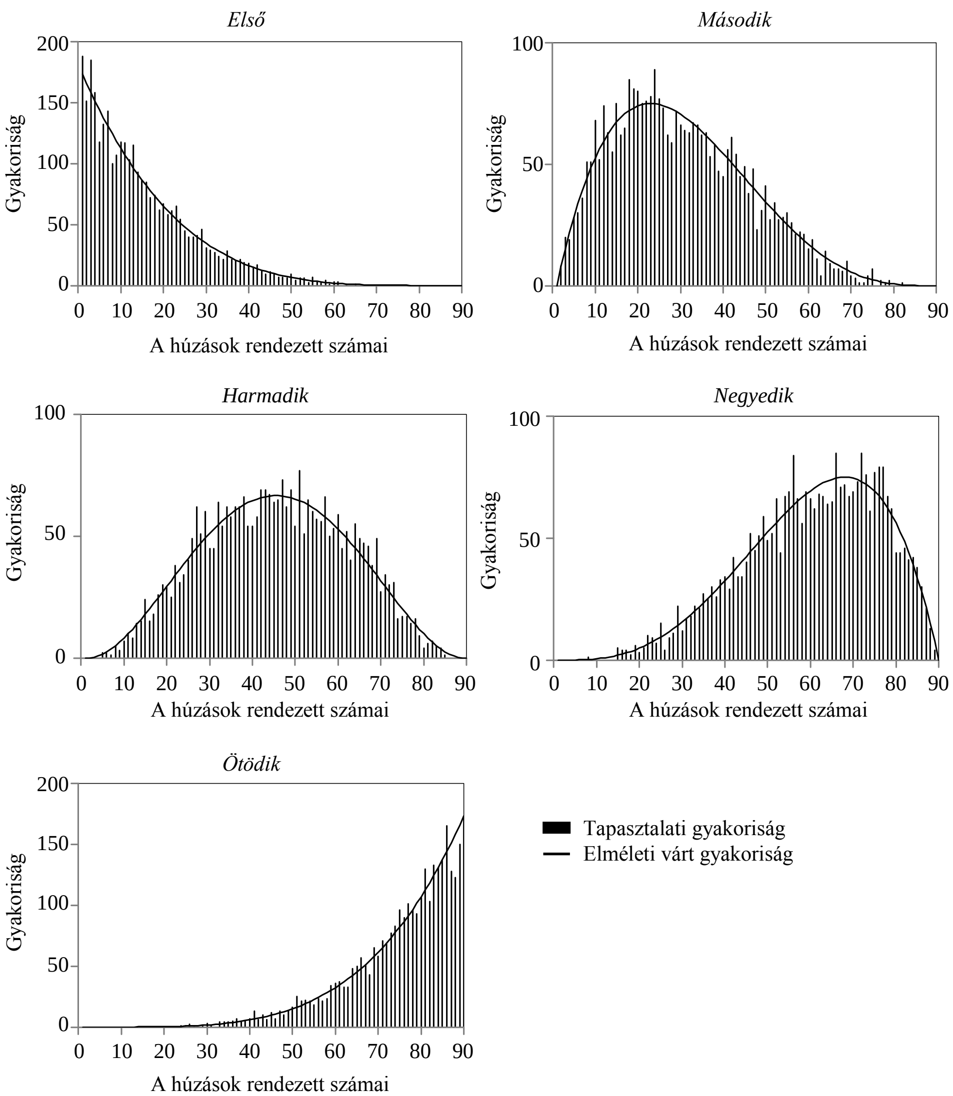
2. ábra. A hetente kisorsolt számok szórása



A szórások egyenletesen, szimmetrikusan helyezkednek el az elméleti szórás körül, szinte tökéletesen homoszkedasztikusak, nem tapasztalható semmilyen torzulás.

Ezek után meghatároztuk annak a gyakoriságát és valószínűségét, hogy a kisorsolt és rendezett számoknál egy adott szám előfordul az első, második, harmadik, negyedik és ötödik helyen. A folyamatos vonal az elméleti eloszlásokat mutatja. (Lásd a 3. ábrát.). Az öt kihúzott számból a legkisebb szám (az első) csak 1-től 86-ig terjedhet. A második szám csak 2 és 87, a harmadik 3 és 88, a negyedik 4 és 89 és végül az ötödik, azaz a legnagyobb szám 5 és 90 között fordulhat elő. Ez logikai úton könnyen belátható.

3. ábra. A kihúzott, rendezett számok tapasztalati gyakorisága és az elméletileg várt gyakoriságok



A polihipergeometrikus eloszlás a hipergeometrikus eloszlás (visszatevés nélküli mintavételezés) általánosításának tekinthető. Legyen a sokaság elemszáma N és a minta elemszám n. A független, visszatevés nélküli mintavételnél, amennyiben adott r-féle különböző kategória, ahol egy-egy kategórián belül rendre S_1, S_2, ..., S_r elem van, akkor annak a valószínűsége, hogy az 1. kategóriába kerülők közül k_1-et, a 2. kategóriába k_2-t, …, az r. kategóriába k_r-t tartalmaz az n elemű minta (*Lukács* [1996]):

$$p = \frac{\binom{S_1}{k_1} \cdot \binom{S_2}{k_2} \cdot \ldots \cdot \binom{S_r}{k_r}}{\binom{N}{n}},$$

ahol $S_1 + S_2, \ldots, + S_r = N$ és $k_1 + k_2, \ldots, + k_r = n$.

Azokat a valószínűségeket, melyek esetén egy adott szám az első, a második, …, az ötödik helyen szerepel a sorsolás eredményeinek rendezése után, a polihipergeometrikus eloszlás adja meg. Itt a képletben látható három kategóriának megfeleltethető:

1. az adott számtól kisebb sorszámú (i) és kisebb számértékű (m) szám választása,

2. az adott sorszámú (i) adott szám (m) választása,

3. az adott számtól nagyobb sorszámú (i) és nagyobb számértékű (m) szám választása.

Jelölje m_i, hogy egy húzásnál i-edik számként éppen m-et sorsoltak:

$$p_{m_i} = \frac{\binom{m_i - 1}{i - 1} \cdot \binom{1}{1} \cdot \binom{90 - m_i}{5 - i}}{\binom{90}{5}} \qquad i = 1, 2, 3, 4, 5.$$

Példánk részletesen:

$$p_{m_1} = \frac{\binom{m_1 - 1}{0} \cdot \binom{1}{1} \cdot \binom{90 - m_1}{4}}{\binom{90}{5}} = \frac{\binom{1}{1} \cdot \binom{90 - m_1}{4}}{\binom{90}{5}} \qquad m_1 \neq 87, 88, 89, 90,$$

$$p_{m_2} = \frac{\binom{m_2 - 1}{1} \cdot \binom{1}{1} \cdot \binom{90 - m_2}{3}}{\binom{90}{5}} \qquad m_2 \neq 1, 88, 89, 90,$$

$$p_{m_3} = \frac{\binom{m_3 - 1}{2} \cdot \binom{1}{1} \cdot \binom{90 - m_3}{2}}{\binom{90}{5}} \qquad m_3 \neq 1, 2, 89, 90\,,$$

$$p_{m_4} = \frac{\binom{m_4 - 1}{3} \cdot \binom{1}{1} \cdot \binom{90 - m_4}{1}}{\binom{90}{5}} \qquad m_4 \neq 1, 2, 3, 90\,,$$

$$p_{m_5} = \frac{\binom{m_5 - 1}{4} \cdot \binom{1}{1} \cdot \binom{90 - m_5}{0}}{\binom{90}{5}} = \frac{\binom{m_5 - 1}{4} \cdot \binom{1}{1}}{\binom{90}{5}} \qquad m_5 \neq 1, 2, 3, 4\,.$$

Ezután kétmintás Kolmogorov–Smirnov-tesztet végeztünk az elméleti és empirikus eloszlás egyezésére. A 2. táblázat alapján megállapítottuk, hogy az elméleti és a tapasztalati eloszlások mind az öt szám esetében megegyeznek. A két minta eloszlásának összehasonlítására χ^2-próba is alkalmazható lett volna. A Kolmogorov–Smirnov-tesztet megfelelőbbnek tartottuk, hiszen egyes osztályokban a gyakoriságok öt alattiak voltak.

2. táblázat

A Kolmogorov–Smirnov-teszt eredménye, a tapasztalati és az elméleti gyakoriságok összehasonlítása a sorsolt, rendezett számok esetén

Eset	*D*-érték	*p*-érték
Az első szám helyén álló számok	0,0476	0,9999
A második szám helyén álló számok	0,1026	0,8065
A harmadik szám helyén álló számok	0,0741	0,9794
A negyedik szám helyén álló számok	0,0921	0,9039
Az ötödik szám helyén álló számok	0,0615	0,9997

A leíró statisztikai mutatók, a tapasztalt és az elméleti valószínűségek alapján megállapítottuk, hogy az eddigi sorsolások valóban véletlenszerűen történtek, nyomát sem találtuk manipulációnak.

2.2. Az előző heti nyerőszámok megtétele

Vajon milyen eredményre vezetett volna, ha mindig az előző héten kihúzott számokat játsszuk meg? A 3 128 húzást elemezve a kép nem nagyon biztató, mivel egyetlen egy háromtalálatos mellé 91 darab kéttalálatos nyereményt zsebelhettünk volna be. A 3 128 darab szelvény ára többszörösen meghaladja a nyeremények összegét.

Háromtalálatos nyeremény az 1987. év 2. hetében lett volna, amikor 11, 41, 61, 68, 76 számokat húzták. Az előző heti nyertes számok: 11, 57, 58, 68, 76.

Érdekes, hogy ebben az évben az első három héten a legkisebb szám a 11 volt. Mi annak a valószínűsége, hogy háromszor egymásután a 11 legyen a legkisebb szám? 0,03418717 a valószínűsége, hogy a 11 a legkisebb, és háromszor egymásután 0,00003996 valószínűséggel következik be ez az esemény (0,03418717 · 0,03418717 · · 0,03418717 = 0,00003996). A nagyon kicsi valószínűség ellenére mégis előfordult.

2.3. A nyerőszámok ismétlődése

Érdekes megvizsgálni, hogy a lottó történetének hatvan éve alatt mennyire ismétlődnek a kihúzott számok. Kihúzták már ugyanazt az öt számot? Elemzésünk alapján ilyen még nem fordult elő, viszont négy egyforma számot már több alkalommal is kisorsoltak. A média csak egyetlen egy esetről számolt be, de nem egyszer, hanem 42-szer fordult már elő ilyen. Sőt, 2006. év 45. hetében öt számból négy megegyezett az 1999. év 36. hetében és az 1999. 41. hetében kihúzottal. A 3. táblázat mutatja a négyes találatok ismétlődéseit.

3. táblázat

Négy szám ismétlődései

Év, hét	A sorsolt számok					Év, hét	A sorsolt számok				
2016. 32.	13	14	21	34	84	2011. 9.	1	56	69	75	76
1973. 4.	14	16	21	34	84	1961. 27.	39	56	69	75	76
2013. 32.	42	44	49	70	78	2010. 41.	2	8	38	44	50
1968. 20.	44	49	70	72	78	1981. 27.	8	35	38	44	50
2013. 8.	2	3	47	61	67	2010. 39.	13	19	31	53	73
2007. 4.	2	32	47	61	67	2010. 33.	13	19	26	53	73
2012. 49.	20	31	34	69	84	2008. 36.	22	23	31	42	73
1964. 4.	31	34	56	69	84	2003. 12.	23	31	42	73	85

(A táblázat folytatása a következő oldalon.)

(Folytatás.)

Év, hét	A sorsolt számok					Év, hét	A sorsolt számok				
2007. 36.	2	4	17	25	41	1995. 21.	3	12	22	29	36
1999. 12.	2	4	17	25	61	1976. 43.	3	12	20	29	36
2006. 45.	25	42	48	73	87	1995. 19.	16	45	51	52	90
1999. 41.	42	48	57	73	87	1992. 37.	16	45	52	87	90
2006. 45.	25	42	48	73	87	1992. 2.	26	33	38	43	53
1999. 36.	25	42	48	54	87	1957. 14.	26	33	38	47	53
2006. 37.	19	21	42	73	87	1988. 20.	11	24	37	52	54
1982. 8.	21	42	54	73	87	1987. 28.	11	24	37	52	61
2006. 23.	1	29	34	73	88	1987. 47.	7	8	10	22	69
1993. 50.	1	34	64	73	88	1968. 14.	7	8	10	15	69
2005. 10.	2	16	36	37	72	1987. 29.	3	9	46	64	74
1994. 48.	2	26	36	37	72	1962. 34.	3	9	55	64	74
2004. 22.	1	24	49	67	71	1981. 52.	12	23	42	56	72
1957. 11.	1	49	64	67	71	1977. 31.	23	42	56	72	77
2003. 42.	13	40	55	56	76	1981. 8.	4	10	47	55	77
1970. 16.	11	13	40	55	56	1975. 47.	4	10	28	47	77
2003. 41.	32	45	55	70	78	1980. 52.	28	45	49	55	70
1962. 25.	32	45	49	55	70	1962. 25.	32	45	49	55	70
2003. 28.	4	56	74	77	89	1978. 32.	9	23	27	33	88
1984. 46.	4	74	77	78	89	1968. 17.	23	27	33	75	88
2001. 38.	19	52	63	72	81	1977. 28.	10	18	21	60	76
1976. 16.	19	32	52	72	81	1962. 16.	10	18	60	68	76
2000. 27.	28	34	39	40	81	1977. 4.	13	66	71	75	77
1964. 49.	28	33	34	40	81	1966. 12.	13	29	66	71	77
1998. 43.	2	10	49	66	73	1973. 43.	3	38	46	59	83
1964. 7.	2	10	49	66	84	1966. 2.	3	38	59	65	83
1998. 42.	58	75	81	82	86	1971. 28.	2	3	20	62	66
1987. 38.	29	75	81	82	86	1969. 3.	2	3	62	63	66
1997. 28.	11	23	24	31	64	1970. 51.	11	23	37	39	64
1989. 41.	23	24	31	49	64	1964. 52.	11	23	39	50	64
1996. 39.	29	40	51	65	66	1967. 26.	21	35	48	52	60
1993. 4.	29	41	51	65	66	1963. 29.	21	35	36	48	60
1995. 27.	2	17	28	77	79	1961. 18.	1	15	33	35	49
1988. 49.	2	28	36	77	79	1958. 41.	1	15	35	46	49

A 42 darab négyes találat elég soknak tűnik a hatvan év alatt, ezért érdemes meghatározni, hogy mi az elméleti valószínűsége ennek az eseménynek. Annak a valószínűsége, hogy négyesünk lesz $9,67 \cdot 10^{-6}$. A 3 128 játékhétig összesen 4 890 628 szelvényt kellett volna kitölteni, ami megegyezik a lehetséges összehasonlítások számával $\binom{3128}{2} = 4\,890\,628$. A valószínűség szorozva az összehasonlítások számával adja az elméleti ismétlődések számát, amely 47 darab. Kiszámítottuk a többi ismétlődés elméleti gyakoriságát is, amely a 4. táblázatban látható.

4. táblázat

Az elméleti (kerekített értékek)
és a tapasztalati számismétlődések gyakorisága 3 128 játékhét esetén

Ismételten sorsolt számok	Elméleti gyakoriság	Tapasztalati gyakoriság
5	0 (0,1113)	0
4	47	42
3	3 973	4 000
2	109 910	109 648
1	1 126 580	1 128 083

Az elméleti és a tapasztalati gyakoriságok nagyon közel estek egymáshoz. Az illeszkedésvizsgálatot χ^2-próbával végeztük, összehasonlítottuk az elméleti és tapasztalati gyakoriságokat. A szignifikanciaszintet 5 százalékon rögzítettük. Amennyiben a számított *p*-érték nagyobb, mint 0,05, akkor elfogadjuk a két gyakoriság azonosságát. Az illeszkedésvizsgálat alapján a két gyakoriság megegyezett (*p*-érték: 0,79).

Egy gondolatkísérlet keretében vizsgáljuk meg, hogy milyen találatokat értünk volna el, ha mindig a már előzőleg kihúzott összes számot játszottuk volna meg 60 éven keresztül. Ehhez nagyon sok szelvényre lett volna szükségünk, pontosan $\binom{3128}{2} = \frac{3128 \cdot 3127}{2} = 4\,890\,628$-ra. Ez mai áron számolva több, mint 1,2 milliárd forint. Ennyi befektetés után lett volna 109 648 darab 2 találatos, 4 000 darab háromtalálatos és 42 darab négytalálatos nyeremény. A 2017. 10. hét nyereményjegyzéke alapján mindössze 337 385 650 forintot nyertünk volna, pedig ez a hét az átlagosnál sokkal jobban fizetett. Sajnos, ez sem kifizetődő.

2.4. A születésnapi taktika

Mint a bevezetőben említettük, állítólag sokan születésnapi dátumokat játszanak meg a lottón, ezért ha maximalizálni akarjuk a nyereményünket, nem érdemes a 31 alatti számokat választani. Tényleg nem érdemes 31 alatti számokat játszani?

A Szerencsejáték Zrt. adatbázisa alapján erre a kérdésre nagyon bizonytalan válasz adható, mert ebből az adatbázisból nem tudjuk meg, hogy hetente hány szelvénnyel játszottak, és milyenek voltak a megjátszott számok. Ráadásul a vásárolt szelvények száma hétről hétre változik. Sok szelvényt vesznek, amikor halmozódik a nyereményalap, mert ilyenkor a rendszertelenül játszók is szeretnének sok pénzt nyerni. Ennek ellenére megvizsgáltuk, hogy a heti átlagos találatok száma megegyezik-e a 31 és alatti, valamint az e feletti nyertes számok esetén. Tehát az adatbázist két részre osztottuk, az egyik felében a „legkisebb szám nagyobb, mint 31” esetek szerepeltek, a másik felében a maradék lehetőségek. Mint várható volt az egyik rész sokkal több adatot tartalmazott, mint a másik. A 31 feletti csoport alul, a másik pedig felül reprezentáltnak bizonyult.

5. táblázat

Az átlagos találatszám a nyereményjegyzék alapján, amikor a legkisebb szám 31 és az alatt, illetve felett van

Találatok száma	31 és az alatti a legkisebb szám n = 2 784	31 feletti a legkisebb szám n = 344
5	0,152	0,108
4	64,934	43,029
3	5 388,121	4 063,422
2	147 923,000	126 298,200

Az 5. táblázat tényleg azt sugallja, ha csak 31 feletti számokat játszunk, akkor kisebb az átlagos találatszám, azonban a különbség nem jelentős. Az aránytalan mintanagyság miatt ezzel az állítással óvatosan kell bánni.

2.5. A páros és páratlan számok taktikája

Számít-e, hogy az öt számon belül hogyan oszlanak el a páros és páratlan számok? A médiában erre is találunk ajánlást, azonban érdemes ezt is egzakt módon vizsgálni. A 90 szám között 45 páros és 45 páratlan szám található. Tehát 50 százalék annak a valószínűsége, hogy az első kihúzott szám páros lesz és ugyanennyi a páratlan számnak is. Annak a valószínűsége, hogy az első két szám páros lesz azonban már nem $0{,}5^2$, hiszen ez csak független valószínűségi eseményeknél igaz, hanem ennél kisebb, mivel a maradék 89 számból már csak 44 páros. A harmadik és további páros számokra is ezt a gondolatmenetet kell követni. Valójában ez egy hipergeometrikus eloszlás, ahol a „kedvező” esetek száma 45, és annak a valószínűségét keressük, ha ötöt kihúzunk, pontosan k számú lesz páros. Ezeket a valószínűségeket az összes lehetséges esetre a 6. táblázat mutatja.

Jelölje k, hogy hány páros szám van a kihúzott számok között:

$$p_k = \frac{\binom{45}{k} \cdot \binom{45}{5-k}}{\binom{90}{5}}.$$

6. táblázat

Egy sorsolásban szereplő k számú páros szám valószínűségei

Páros szám (darab)	Páratlan szám (darab)	Valószínűség	Elméleti gyakoriság (kerekítve)	Tapasztalati gyakoriság
5	0	0,02780	87	100
4	1	0,15256	477	436
3	2	0,31964	1 000	983
2	3	0,31964	1 000	1 026
1	4	0,15256	477	491
0	5	0,02780	87	92

Az elméleti és a tapasztalati gyakoriságokra elvégzett illeszkedésvizsgálat a gyakoriságok azonosságát erősítette meg (p-érték = 0,61). A homogenitás vizsgálatot itt is χ^2-próbával végeztük. A legjobb taktika tehát, ha kettő vagy három páros számot játszunk, mivel ennek a legnagyobb a valószínűsége. Ez egy praktikus tanács, amivel tényleg lehet növelni a nyerési esélyünket.

2.6. Hogyan válasszunk számokat? A számok elhelyezkedésének egyenletessége

A nyerési esély növelésére egyesek azt tanácsolják, hogy a lehető legegyenletesebben válasszuk ki az öt számot. Véleményük szerint lehetőleg minden ötödből (pentilisből) legyen egy-egy szám. Meghatároztuk az öt tartományt: az első 1-től 18-ig, az utolsó 73-tól 90-ig terjed. Ezután a 3 128 húzást besoroltuk aszerint, hogy hány tartományból (pentilisből) kerültek ki a nyertes számok. A pentilisek száma így egytől ötig terjedhetett.

7. táblázat

Az egyes ötödökbe esés valószínűségei

A pentilis, ahonnan a sorsolt számok kikerültek	Valószínűség	Elméleti gyakoriság (kerekítve)	Tapasztalati gyakoriság
Egy	0,00098	3	1
Kettő	0,08188	256	258
Három	0,46809	1 464	1 499
Négy	0,40606	1 270	1 267
Öt	0,04299	135	103
Összesen	*1,00000*	*3 128*	*3 128*

A valószínűségeket a következő képletekkel számoltuk ki. Amikor azt feltételezzük, hogy a sorsolás összes számát egy pentilisből húzták ki.

$$p_1 = \frac{\binom{18}{5}\cdot\binom{5}{1}}{\binom{90}{5}} = 0,00098$$

A valószínűségeket a klasszikus valószínűségszámítás szabályai alapján kaptuk, a kedvező esetek számát osztottuk az összes eset számával. A számlálóban álló szorzat azt fejezi ki, hogy az öt pentilisből választunk ki egy pentilist és a 18 számot tartalmazó pentilisből pedig öt számot. A nevezőben mind az öt képletben ugyanaz a szám szerepel: a kilencven számból ennyi módon választhatunk ki öt számot.

A „két pentilisből lett sorsolva az összes szám” eset képlete a következőkben látható. Itt az egyik pentilisből négy számot vagy a másikból egyet választunk, vagy három, illetve két szám kerül ki a két pentilisből. A további szorzótényezők magyarázata: ki kell választani azt a két pentilist, amiből a számok származnak, és az adott két pentilisből is kétféleképpen kerülhetnek ki a számok: 1, 4 vagy 4, 1; illetve 2, 3 és 3, 2 sorrendben.

$$p_2 = \frac{\binom{18}{4}\cdot\binom{18}{1}\cdot\binom{5}{2}\cdot 2+\binom{18}{3}\cdot\binom{18}{2}\cdot\binom{5}{2}\cdot 2}{\binom{90}{5}} = 0,08188\,.$$

A „három pentilisből lett sorsolva az összes szám” esetében lehet, hogy az egyes pentilisekből 3 + 1 + 1 szám kerül ki, de lehet, hogy 2 + 2 + 1 módon választottunk. Tehát meg kell határozni, hogy az öt pentilisből melyik három adja a számokat, majd hárommal még meg kell szorozni, hiszen a 3 + 1 + 1 és a 2 + 2 + 1 esetben is háromféleképpen kerülhetnek ki a számok az adott pentilisekből.

$$p_3 = \frac{\binom{18}{3}\cdot\binom{18}{1}\cdot\binom{18}{1}\cdot\binom{5}{3}\cdot 3 + \binom{18}{2}\cdot\binom{18}{2}\cdot\binom{18}{1}\cdot\binom{5}{3}\cdot 3}{\binom{90}{5}} = 0,46809$$

A „négy pentilisből lett sorsolva az összes szám” esetben egyszerűsödik a formula, hiszen itt csak 2 + 1 + 1 + 1 lehet az egyes pentilisekből származó számok mennyisége, de az négyféleképpen valósulhat meg, hiszen a kiválasztott pentilisek bármelyike lehet az, amelyik két számot adott.

$$p_4 = \frac{\binom{18}{2}\cdot\binom{18}{1}\cdot\binom{18}{1}\cdot\binom{18}{1}\cdot\binom{5}{4}\cdot 4}{\binom{90}{5}} = 0,40606$$

Végül az az eset, amikor mind az öt szám külön pentilisből való:

$$p_5 = \frac{\binom{18}{1}\cdot\binom{18}{1}\cdot\binom{18}{1}\cdot\binom{18}{1}\cdot\binom{18}{1}}{\binom{90}{5}} = 0,04299\,.$$

A 7. táblázat azt mutatja, hogy a nyerés valószínűsége akkor a legkisebb, amikor az öt szám egy pentilisből való. A második legkisebb valószínűséget arra kaptuk, amikor a húzott számok egyenletesen oszlanak el, mind az öt szám a saját ötödébe kerül. A legnagyobb valószínűség szerint a nyertes számok három vagy négy pentilisből származnak. Az említett tanácsnak tehát pont az ellenkezője igaz, nem növeli a nyerési esélyeket, ha ötödönként választunk egy-egy számot.

3. Összefoglalás

Mindenki szeretne nyerni, de a nyerés valószínűsége nagyon kicsi. Vajon milyen praktikus tanácsot lehetne adni az esélyek növelésére? Amióta a fogadások elektronikus úton történnek, számtalanszor felreppen a gyanú, hogy a fogadások beérkezés után, a megjátszott számok ismeretében, olyan számokat sorsolnak ki, amelyeket senki sem játszott meg. Ez a vélekedés főként a hosszantartó telitalált nélküli időszakokban szokott felerősödni. Tényleg manipulálják a számokat, vagy ezek az események a lottó sajátosságaiból fakadó normális viselkedés?

Jelen munkánkban 3 128 hét sorsolási adatai alapján összehasonlítottuk a kilencven számból ötöt játékelméleti eloszlásait és értékeit a tényleges húzási eredményekkel. Ezentúl néhány nyerési esélyt növelő taktikát is megvizsgáltunk. Egyáltalán létezik ilyen, vagy mindegyik csak a mítosz kategóriába tartozik?

Az 1–90 számok egyenkénti előfordulási gyakorisága, a várható érték és a szórás tökéletesen megegyezett az elméleti értékekkel. A nagyság szerint sorba rendezett öt szám polihipergeometrikus eloszlást mutatott. Az elméleti eloszlás és az empirikus gyakoriság ebben az esetben is statisztikailag igazoltan megegyezett. Az elvégzett vizsgálatok alapján semmi nem utal manipulációra, a hatvan év húzási adatai alapján a sorsolások tisztaságához nem fér kétség.

A korábban már kisorsolt számok újrajátszásának taktikája nem vezet meggazdagodáshoz. Az előző heti nyerőszámok megjátszása egyetlen három találatot és 91 darab két találatot eredményezett volna. A nyerőszámok ismétlődésének elemzése érdekes eredményre vezetett. Ugyanazt az öt számot még nem húzták ki kétszer, de négy egyforma számot már 42 alkalommal sorsoltak ki. A média tévesen csak egyetlen egy ilyen eseményről tudósított. A számok ismétlődése megegyezett az elméletileg előre jelzett értékekkel. Ha mindig megjátszanánk a korábban már kihúzott számokat – a nagyon sok szelvény ellenére – ez is veszteséges taktika lenne.

A „ne játszunk születésnapi számokat” tanácsot nem tudtuk egyértelműen igazolni. Ennek az egyik oka az adatbázis sajátosságából adódik, a másik pedig a két mintanagyság aránytalanságából. Némi tendencia felfedezhető, de ez statisztikailag nem igazolható.

A páros és a páratlan számok taktikája viszont ténylegesen növeli a nyerési valószínűséget. A legnagyobb esély a nyerésre a 2 vagy 3 páros szám megjátszásakor következik be. A legkisebb valószínűség az öt páros vagy öt páratlan szám ikszelésekor adódik. A két valószínűség között 11,5-szeres a különbség. Amennyiben egy szelvénnyel játszunk, a nyerési valószínűség $2,275351 \cdot 10^{-8}$, ha ezt megtízszerezzük, még akkor is nagyon kicsi számot kapunk. Igaz, hogy ez a taktika növeli a valószínűséget, de a nyerési esély így is nagyon kicsi.

A számok egyenletes elhelyezkedésének vizsgálatakor azt számoltuk ki, hogy milyen valószínűségekkel esnek a számok egy, kettő, …, öt pentilisbe. Ez alapján azt állíthatjuk, hogy legnagyobb valószínűséggel a sorsolt számok három, illetve négy pentilisbe esnek. Nem érdemes tehát egy-egy ötödből egy-egy számot választani.

Az évfordulót ünneplő játékkal kapcsolatos számításainkkal sajnos nem sikerült csodamódszert biztosítani a garantált nyereményhez, de talán néhány tévhittel leszámoltunk.

Irodalom

HUNYADI L. – MUNDRUNCZÓ GY. – VITA L. [1996]: *Statisztika*. Aula Kiadó. Budapest.

KUTATÓPONT [2011]: *Szerencsejáték – személyes megkérdezésen alapuló tracking kutatás*. Budapest.

LIEN, J. W. – YUAN, J. [2015]: The cross-sectional "gambler's fallacy": set representativeness in lottery number choices. *Journal of Economic Behavior and Organization.* Vol. 109. January. pp. 163–172. http://dx.doi.org/10.1016/j.jebo.2014.10.011

LUKÁCS O. [1987]: *Matematikai statisztika*. Műszaki Könyvkiadó. Budapest.

MILITANA, E. – WOLFSON, E. – CLEAVELAND, J. M. [2010]: An effect of inter-trial duration on the gambler's fallacy choice bias. *Behavioural Processes*. Vol. 84. Issue 10. pp. 455–459. http://dx.doi.org/10.1016/j.beproc.2010.02.010

R CORE TEAM [2016]. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/.

SIMON, J. [1999]: An analysis of the distribution of combinations chosen by UK national lottery players. *Journal of Risk an Uncertainty*. Vol. 17. Issue 3. pp. 243–276.

SUETENS, S. – TYRAN, J. R. [2012]: The gambler's fallacy and gender. *Journal of Economic Behavior and Organization.* Vol. 83. Issue 1. pp. 118–124. http://dx.doi.org/10.1016/j.jebo.2011.06.017

SZEGŐ I. M. [2011]: Ha már nyerünk, hogyan legyünk milliárdosok a lottón? *Heti Világgazdaság.* Április 2. http://hvg.hu/tudomany/20110720_lotto_milliardos_lotto

TESSÉNYI J. – KAZÁR K. [2012]: Szerencsejáték-vásárlási szokások vizsgálata „prediktív analitika" segítségével. *Statisztikai Szemle*. 90. évf. 7–8. sz. 677–695. old.

TESSÉNYI J. – KOVÁCS P. [2011]: Szerencsejáték és bűnözés. *Statisztikai Szemle*. 89. évf. 4. sz. 399–419. old.

VINCZE I. – VARBANOVA M. [1993]: *Nemparaméteres matematikai statisztika*. *Elmélet és alkalmazások*. Akadémiai Kiadó. Budapest.

XU, J. – HARVEY, N. [2014]: Carry on winning: the gambler's fallacy creates hot hand effects in online gambling. *Cognition.* Vol. 131. Issue 2. pp. 173–180. http://dx.doi.org/10.1016/j.cognition.2014.01.002

Summary

The Hungarian lottery, which was re-organized after World War II, turned sixty in March 2017. In Hungary just as in other parts of the world, everyone would like to win the lottery, and the Internet is full of advice on how to do so (all of these tips, however, are inaccurate, unreliable and with-

out any scientific basis). Many people think that the winning numbers are predictable. Some of them believe that if a number is drawn more frequently than "normal" during a certain period, it will be drawn less frequently in the future; while others think just the opposite. There are also believes about maximizing the amount of the potential lottery prize by avoiding smaller numbers (that refer to birthdays or the so-called lucky numbers [3, 7,..]).

The paper is based on the data of the Szerencsejáték Private Limited Company operating the Hungarian lottery, and uses mathematical-statistical methods to separate fact from fiction.