
Tóth Zoltán

docens,

Károly Róbert Főiskola

E-mail: zoli@karolyrobert.hu

Nemlineáris függvények illesztésének néhány kérdése

A nemlineáris regressziós és trendfüggvények illesztésekor számos esetben alkalmazzuk az adatok transzformációjának módszerét. Ekkor az összefüggésre és az adatokra egy alkalmas transzformációt alkalmazva, paramétereinket lineáris összefüggésből kell meghatározni, amelynek normál egyenletrendszerét már könnyedén, akár zsebszámológép használatával is meg tudjuk oldani.

A probléma ismert a szakirodalomban, azonban a szerzők, kiváltképp a tankönyvek szerzői nem fordítanak kellő figyelmet arra, hogy a linearizálásból adódó torzítás milyen mértékű és irányú, hanem többnyire megelégszenek azzal a megállapítással, hogy a torzítás mértéke általában olyan, hogy az összefüggés a gyakorlat számára még használható (például *Hunyadi-Vita* [2004]). A tapasztalat azt mutatja, hogy ezek az eltérések exponenciális vagy hatványfüggvények illesztése esetén valóban nem mindig jelentősek, de például hiperbolikus függvény illesztésekor nagyságrendileg nagyobb eltéréseket is tapasztalhatunk, sőt ismeretes, hogy bonyolultabb függvények (például a logisztikus függvény) esetében ezek a torzítások gyakran használhatatlanná teszik a linearizálással kapott eredményeket. Az eltérések adatainktól függően egyazon függvénytípus esetén is eltérőek lehetnek.

Ennek a tanulmánynak az a célja, hogy az említett transzformációkból adódó torzításokat elemezze, és ezáltal is felhívja a figyelmet azokra a problémákra, melyek felett kiváltképp a tapasztalatlan alkalmazók hajlamosak átsiklani. Ennek megfelelően a kérdés felvetése után először egy egyszerűsített feladaton, a két, illetve több megfigyelés (számérték) legkisebb négyzetekkel kapható közepének meghatározásakor mutatjuk be az oda-vissza alkalmazott transzformációk tulajdonságait, majd a nemlineáris görbeillesztés néhány gyakori esetét vizsgáljuk meg. A tanulmányt a következtetések összefoglalása, a felhasználók számára megfogalmazott javaslatok, valamint a megválaszolatlanul maradt kérdések zárják.

1. A legkisebb négyzetek módszerének tulajdonságai az adatok transzformációja esetén

Amennyiben nemlineáris regressziós függvényt illesztünk, gyakran transzformáljuk a függő vagy a független változót. Az alapötlet első pillantásra helyesnek tűnik, hiszen tökéletesen illeszkedő függvények esetén az oda-vissza transzformáció (ha létezik) helyben hagyja a függvényt. Ha például egy pontosan hiperbolikus függvény szerint alakuló adatsorra akarunk függvényt illeszteni, akkor a függő változó reciprokára illesztve a lineáris függvényt, majd ennek a lineáris függvénynek a reciprokát véve megkapjuk az eredeti adatokra illeszkedő hiperbolikus függvényt.

Ez a gondolat annyira egyszerű, hogy nem is érdemes tovább magyarázni. A probléma ott kezdődik, amikor az illeszkedés nem tökéletes, hiszen akkor az eredmények korántsem lesznek ilyen triviálisak. Annak érdekében, hogy a problémát érthetően megvilágítsuk, először egy egyszerűbb feladatot fogalmazzunk és oldunk meg: azt vizsgáljuk meg, hogy két adat M , illetve m között keresve a legkisebb négyzetes eltérést teljesítő x értéket hogyan változik x értéke, ha az adatokat transzformáljuk, majd a transzformált adatokra kapott értéket az inverz transzformációval visszatranszformáljuk. Az előzők alapján nyilvánvaló, hogy amennyiben $M = m$, akkor x is egyenlő lesz velük, más esetekben azonban nem ez a helyzet. Az itt következő kis elemzés előtanulmányának is tekinthető a görbeillesztés feladathoz, ugyanakkor önállóan is érdekes, mivel rámutat egyes statisztikai átlagok tulajdonságaira.

1.1. Alapeset (nincs transzformáció)

Két érték, M és m ($M \neq m$) között keresünk egy olyan x értéket, amelyre

$$f(x) = (M-x)^2 + (m-x)^2 \rightarrow \min . \quad /1/$$

A függvény láthatóan felfelé nyíló parabola, ott lesz minimális, ahol $f'(x) = 0$. Mivel $f'(x) = -2(M-x) - 2(m-x) = 0$, ezt átrendezve

$$M - x + m - x = 0, \text{ majd } x = \frac{M + m}{2}$$

kapható, tehát a négyzetes eltérés akkor lesz a legkisebb, ha x a két érték számtani közepe. Az eredmény általánosan ismert alapösszefüggés, és több adat esetén is igaz.

1.2. A lineáris transzformáció esete

Legyen $l(x) = a + b \cdot x$ lineáris transzformáció. Ekkor $l(M)$ és $l(m)$ értékek között keressünk egy olyan y értéket, amelyre

$$f(y) = (l(M) - y)^2 + (l(m) - y)^2 \rightarrow \min, \quad /2/$$

majd ezt az y értéket az l inverz transzformációjával visszaalakítva az a kérdés, hogy visszakapjuk-e az alapesetben meghatározott x értéket.

Ekkor $f(y) = (a + b \cdot M - y)^2 + (a + b \cdot m - y)^2$, ami ott lesz minimális, ahol $f'(y) = 0$. Ekkor az

$$f'(y) = -2(a + b \cdot M - y) - 2(a + b \cdot m - y) = 0$$

egyenlet megoldását keressük, amire azt kapjuk, hogy

$$a + b \cdot M - y + a + b \cdot m - y = 0, \text{ majd } y = a + b \frac{M + m}{2}, \text{ végül}$$

$$l^{-1}(y) = -\frac{a}{b} + \frac{1}{b}y = \frac{M + m}{2}.$$

Ez azt jelenti, hogy a transzformált adatokra meghatározva a legkisebb négyzetes eltérést adó értéket, majd ezen értékre végrehajtva a transzformáció inverzét az adatok számtani átlagát kapjuk. Tehát transzformációval meghatározva y értékét, valamint inverz transzformációval x értékét az alapesetnek megfelelő x értéket kapjuk.

1.3. A reciprokképzés esete

Legyenek adva a M és m értékek. Legyen a transzformáció a reciprokképzés, amelynek inverze önmaga. Keressük azt az y értéket, amelyre:

$$f(y) = \left(\frac{1}{M} - y\right)^2 + \left(\frac{1}{m} - y\right)^2 \rightarrow \min. \quad /3/$$

A szokásos módon ezúttal is az $f'(y) = 0$ egyenletet kell megoldani:

$$f'(y) = -2\left(\frac{1}{M} - y\right) - 2\left(\frac{1}{m} - y\right) = 0, \quad y = \frac{\frac{1}{M} + \frac{1}{m}}{2}, \quad \text{azaz } x = \frac{1}{y} = \frac{2}{\frac{1}{M} + \frac{1}{m}}.$$

Ez azt jelenti, hogy a transzformált adatokra meghatározva a legkisebb négyzetes eltérést adó értéket, majd ezen értékre végrehajtva a transzformáció inverzét az adatok harmonikus átlagát kapjuk. Ez azzal a következménnyel jár, hogy az így kapott x érték nem egyezik meg az alapesetben meghatározott számtani átlaggal, hanem annál kisebb. Tehát a transzformált adatokra meghatározva a legkisebb négyzetek elvét teljesítő y értéket, majd azt az inverz transzformációval visszaalakítva, az eredeti adatok között a legkisebb négyzetek elvét teljesítő számtani átlagnál kisebb értéket, a harmonikus átlagot kapjuk eredményként.

1.4. A logaritmusképzés esete

Ismét legyen adva M és m és legyen a transzformáció a természetes alapú logaritmus, melynek inverze: e^x . Keressük azt az y értéket, amelyre:

$$f(y) = (\ln M - y)^2 + (\ln m - y)^2 \rightarrow \min. \quad /4/$$

Tudjuk, ez akkor teljesül, ha $f'(y) = -2(\ln M - y) - 2(\ln m - y) = 0$, ahonnan $y = \frac{\ln M + \ln m}{2}$, majd $x = e^y = e^{\frac{\ln M + \ln m}{2}} = \sqrt{M \cdot m}$ adódik.

Ekkor tehát a transzformált (logaritmizált) adatokra meghatározva a legkisebb négyzetek elvét teljesítő y értéket, majd azt az inverz transzformációval visszatranszformálva, az eredeti adatok között a legkisebb négyzetek elvét teljesítő számtani átlagnál kisebb értéket, a mértani átlagot kapjuk.

1.5. A négyzetgyökvonás esete

Legyen adva továbbra is M és m , és a transzformáció a négyzetgyökvonás, melynek inverze a négyzetre emelés. Keressük azt az y értéket, amelyre:

$$f(y) = (\sqrt{M} - y)^2 + (\sqrt{m} - y)^2 \rightarrow \min. \quad /5/$$

Tudjuk, ez akkor teljesül, ha $f'(y) = -2(\sqrt{M} - y) - 2(\sqrt{m} - y) = 0$, azaz

$$y = \frac{\sqrt{M} + \sqrt{m}}{2}, \text{ és } x = y^2 = \frac{M + m + 2\sqrt{M \cdot m}}{4}.$$

Az így kapott (visszatranszformált) x érték nem az adatok számtani átlaga, mert:

$$x = \frac{M + m + 2\sqrt{M \cdot m}}{4} \leq \frac{M + m}{2}, \text{ hiszen } M + m + 2\sqrt{M \cdot m} \leq 2(M + m) \text{ és } \sqrt{M \cdot m} \leq \frac{M + m}{2} \text{ a számtani és mértani átlag közötti ismert relációból adódóan.}$$

Láthatjuk, hogy a transzformált adatokra meghatározva a legkisebb négyzetek elvét teljesítő y értéket, majd azt az inverz transzformációval visszatranszformálva, egy a számtani átlagnál kisebb x értéket kapunk. Vagyis a visszatranszformált érték nem teljesíti a legkisebb négyzetek elvét az eredeti adatok között, hanem annál kisebb.

1.6. A négyzetre emelés esete

Legyen adva M és m , és legyen a transzformáció most a négyzetre emelés, melynek inverze a négyzetgyökvonás. Keressük azt az y értéket, amelyre:

$$f(y) = (M^2 - y)^2 + (m^2 - y)^2 \rightarrow \min. \quad /6/$$

Tudjuk, ez akkor teljesül, ha $f'(y) = -2(M^2 - y) - 2(m^2 - y) = 0$, azaz

$$y = \frac{M^2 + m^2}{2}, \quad x = \sqrt{y} = \sqrt{\frac{M^2 + m^2}{2}}.$$

Az így kapott (visszatranszformált) x érték nem az adatok számtani átlaga, hanem annál nagyobb, mert:

$$\sqrt{\frac{M^2 + m^2}{2}} \geq \frac{M + m}{2}, \text{ hiszen}$$

$$\frac{M^2 + m^2}{2} \geq \frac{M^2 + m^2 + 2 \cdot M \cdot m}{4} \text{ és } \frac{M^2 + m^2}{2} \geq M \cdot m = \sqrt{M^2 \cdot m^2},$$

ami igaz az M^2 és m^2 számtani és mértani átlaga közötti ismert nagyságrendi összefüggés alapján. Ekkor tehát a transzformált adatokra meghatározva a legkisebb négyzetek elvét teljesítő y értéket, majd azt az inverz transzformációval visszatranszformálva, egy a számtani átlagnál nagyobb x értéket kapunk. Vagyis a visszatranszformált érték nem teljesíti a legkisebb négyzetek elvét az eredeti adatok között, hanem annál nagyobb.

1.7. Általános eset

Legyen $g(x)$ egy a vizsgált intervallumon monoton, folytonos és így invertálható függvény. Ekkor a $g(M)$ és $g(m)$ értékek között keresünk egy olyan y értéket, amelyre

$$f(y) = (g(M) - y)^2 + (g(m) - y)^2 \rightarrow \min. \quad /7/$$

$f(y)$ ott lesz minimális, ahol $f'(y) = -2(g(M) - y) - 2(g(m) - y) = 0$. Ekkor

$$y = \frac{g(M) + g(m)}{2}, \quad x = g^{-1}(y) = g^{-1}\left(\frac{g(M) + g(m)}{2}\right).$$

Ha $g^{-1}(y) = \frac{M + m}{2}$, akkor a legkisebb négyzetek elvének megfelelő x érték meghatározása és a g függvény által meghatározott transzformáció alkalmazási sorrendje felcserélhető. Ehhez a következő összefüggés kell, hogy teljesüljön:

$$g^{-1}\left(\frac{g(M) + g(m)}{2}\right) = \frac{M + m}{2}.$$

Az előző összefüggés pedig csak akkor teljesül, ha g inverze és így g is lineáris.

Összegezve megállapítható, hogy ha adatainkat transzformáljuk, csak a lineáris transzformáció hagyja változatlanul a két érték közötti legkisebb négyzetek elvét teljesítő értéket, azaz a számtani átlagot. Mivel a levezetések több adat esetén is ugyanúgy végigvihetők, megállapíthatjuk, hogy a transzformált adatok várható értékét meghatározva majd ezt az értéket az inverz transzformációval visszatranszformálva, csak lineáris transzformáció esetén kapjuk meg az eredeti adatok várható értékét.

Az eddig elmondottakhoz még egy megjegyzést kell fűznünk. Tudjuk, hogy ha g monoton, folytonos (nem lineáris) függvény és f folytonos, akkor amennyiben f -nek

létezik minimum helye, akkor $g(f)$ összetett függvénynek is létezik, és ez a két minimumhely egybeesik. Azonban esetünkben ezt a tulajdonságot nem alkalmazhatjuk, hiszen mi a következő típusú függvényeket akarjuk minimalizálni:

– az eredeti adatokra

$$F(a, b) = \sum_{i=1}^n (y_i - f(x_i))^2 \rightarrow \min ,$$

– a transzformált adatokra

$$G(a, b) = \sum_{i=1}^n (g(y_i) - g(f(x_i)))^2 \rightarrow \min .$$

Itt $G \neq g(F)$ és pontosan azt mutattuk meg az előző levezetésekben, hogy ez a minimum eltolódik a transzformációk hatására.

Korábban láttuk, hogy ha adataink illeszkednek, például egy hiperbolikus függvényre és azokat linearizáljuk, valamint a linearizált adatokra lineáris függvényt illesztünk, majd ezt az inverz transzformációval visszaalakítjuk, akkor visszkapjuk a kiindulási függvényt. Vagyis a transzformáció nem torzítja az eredményt. Ennek oka, hogy tökéletes illeszkedés esetén ezt a gondolatmenetet követve olyan esettel állunk szemben, ahol $M = m$, és így ekkor x értéke is illeszkedik $M = m$ értékre. Feladatainkban ilyen ideális eset nincs is, mert ekkor transzformáció nélkül is meg tudnánk találni a függvényt, illetve a paraméterek értékét. Viszont ha az adatok nem illeszkednek pontosan az eredeti vagy a transzformált függvényre, akkor a transzformáció eltérít minket a legjobban illeszkedő függvénytől.

A kétszeres transzformáció természetesen nem biztosítja automatikusan azt, hogy a kiinduló és a transzformációk után kapott értékek azonosak legyenek. Egy egyszerű példát említve illeszkedjenek az eredeti adatpárok egy egyenesre. A transzformáció legyen a reciprok képzése, és a transzformált adatokra (amelyek nyilván nem illeszkednek tökéletesen egy egyenesre) illesztünk lineáris függvényt, majd a transzformált adatokra kapott lineáris függvényt és annak értékeit alakítsuk vissza az inverz transzformációval. Az így kapott visszatranszformált adatok nyilvánvalóan nem egyeznek az eredetiekkel.

2. A hibatagok változása a transzformáció hatására

Vizsgáljuk meg a továbbiakban, hogy milyen mennyiségi összefüggés van az eredeti függvényre illesztett nemlineáris regressziós vagy trendfüggvény hibatagjai és a linearizált függvény hibatagjai között. Bevezetőül meg kívánjuk jegyezni, hogy

a transzformáció mindig az y függvényértékekre vonatkozik, illetve, hogy mindig additív hibagról fogunk beszélni. Az eredeti függvény hibáját az i -edik adat esetén ε_i , a transzformált függvény esetén ugyanezt δ_i jelöli.

Az itt következő elemzésekben meghatározzuk néhány alapvető nemlineáris esetben a hibatagok négyzetösszegét és keressük ezen összegek minimumhelyeit. Megmutatjuk, hogy ezen minimumhelyek az adattranszformáció alkalmazásával végzett megoldásokban és az eredeti adatokra vizsgálva nem esnek egybe. Egyben azt is igyekszünk megmutatni, hogy mi a viszony a kétféle hibatag között.

2.1. A hibatagok viszonya exponenciális függvény esetén

Ebben az esetben az illesztett függvény $y = a \cdot e^{bx}$ alakú, a transzformáció pedig, amivel ez linearizálható, a természetes alapú logaritmus. Megjegyezzük, hogy leggyakrabban az $a > 0$, $b > 0$ esettel találkozunk, így példánkban is ilyet mutattunk be.

Természetesen az itteni elemzés kiterjeszhető a függvény más parametrizálására is. Az eredeti adatokra, illetve az exponenciális függvényre vonatkozó additív hibatag legyen ε , a transzformált adatokra lineáris függvényt illesztünk és az itt szereplő additív hibatag legyen δ .

Az a és b paraméterek változtatásával eredeti célunk a $\sum \varepsilon_i^2 \rightarrow \min$ meghatározása. Ennek egyik útja volt az adatok transzformálása és $\sum \delta_i^2 \rightarrow \min$ meghatározása. A következőkben bemutatjuk, hogy ez a két minimum nem esik egybe, és meghatározzuk, hogy mekkora az eltérés az ε és δ értékek között. Legyen az eredeti adatokra illesztendő függvény alakja $y_i = a \cdot e^{b \cdot x_i} + \varepsilon_i$, amiből azonnal kapható az $y_i - \varepsilon_i = a \cdot e^{b \cdot x_i}$ forma. A transzformált adatokra illesztett lineáris függvény:

$$\ln y_i = \ln a + b \cdot x_i + \delta_i, \text{ ahonnan } \ln y_i - \delta_i = \ln a + b \cdot x_i, \text{ és } e^{\ln y_i - \delta_i} = \frac{y_i}{e^{\delta_i}} = a \cdot e^{b \cdot x_i}.$$

A két utóbbi kapott alak összevetése és némi átalakítás után a következő adódik:

$$y_i - \varepsilon_i = \frac{y_i}{e^{\delta_i}}; \frac{1}{e^{\delta_i}} = \frac{y_i - \varepsilon_i}{y_i}; \frac{1}{e^{\delta_i}} = e^{-\delta_i} = \left(1 - \frac{\varepsilon_i}{y_i}\right) \text{ és innen}$$

$$-\delta_i = \ln \left(1 - \frac{\varepsilon_i}{y_i}\right) = \ln \left(1 - \frac{1}{\frac{y_i}{\varepsilon_i}}\right) = \frac{\varepsilon_i}{y_i} \ln \left(1 - \frac{1}{\frac{y_i}{\varepsilon_i}}\right).$$

Mivel $\lim_{x \rightarrow \infty} \left(1 - \frac{1}{x}\right)^x = e^{-1}$ ezért nagy y_i és kicsi ε_i , tehát jó közelítés esetén eb-

ből következik, hogy $\ln \left(1 - \frac{1}{\frac{y_i}{\varepsilon_i}}\right)^{\frac{y_i}{\varepsilon_i}} \approx -1$, ami miatt

$$\delta_i \approx \frac{\varepsilon_i}{y_i}, \text{ azaz } \varepsilon_i \approx y_i \delta_i. \quad /8/$$

Látható, hogy $\sum \varepsilon_i^2$ és $\sum \delta_i^2$ minimuma nem feltétlenül egyezik meg. Pontosabban, mivel $\sum \frac{\varepsilon_i^2}{y_i^2} \approx \sum \delta_i^2$ és mi $\sum \delta_i^2 \rightarrow$ minimumát határozzuk meg, ezért az ez-

zel egyenlő $\sum \frac{\varepsilon_i^2}{y_i^2}$ összegben a nagy y értékekhez tartozó, ε_i hibák kisebb súllyal

szerepelnek. Ez azt jelenti, hogy a transzformációval kapott regressziós függvényünk nagy y értékekre nem lesz annyira pontos, vagyis a transzformációval valóban nem kapjuk meg a legjobban illeszkedő exponenciális függvényt. A legjobban illeszkedő függvény meghatározásához a $\sum \varepsilon_i^2 \rightarrow$ minimumot kellett volna meghatározni, ami viszont analitikusan nem kezelhető normálegenlet-rendszerhez vezet.

Alakítsuk most vissza regressziós függvényünk transzformált alakját hibataggal együtt:

$$\begin{aligned} \ln y_i &= \ln a + b \cdot x_i + \delta_i, \\ y_i &= a \cdot e^{b \cdot x_i} \cdot e^{\delta_i} = a \cdot e^{b \cdot x_i} \left(1 + \frac{P_i}{100}\right). \end{aligned}$$

Ekkor az látható, hogy az előző felírás esetén $e^{\delta_i} = 1 + \frac{P_i}{100}$ jelöléssel P_i azt mutatja meg, hogy hány százalék az eltérés az eredeti és számított y_i érték között, és mi ezt minimalizáljuk. Ez akár egy kritériuma is lehet a függvényillesztésnek (az eltérés százalékos összegének a minimalizálása), de jól látható, hogy ez nem felel meg a legkisebb négyzetek elvének.

Nézzük meg, hogy ez az eredmény hogyan látható a következő fiktív adatsoron:

1. példa

Legyen adva az alábbi 4 adatpár

Eredeti adatok				Transzformált adatok			$\ln\left(1 - \frac{\varepsilon_i}{y_i}\right)$
x_i	y_i	y_i számított (transzformáció)	ε_i	$\ln y_i$	$\ln y_i$ számított	δ_i	
0	2,0	1,379	0,621	0,693	0,321	0,372	-0,372
1	1,7	3,179	-1,479	0,531	1,156	-0,626	0,626
2	8,4	7,327	1,073	2,128	1,992	0,137	-0,137
3	19,0	16,889	2,111	2,944	2,827	0,118	-0,118

Ezen adatokat transzformálva és a transzformált adatokra alkalmazva a legkisebb négyzetek elvét (lineáris függvényt illesztve), majd a paramétereket visszatranszformálva $a = 1,379$, illetve $b = 0,8351$ értékeket kapunk, azaz a legjobb függvénynek a linearizálás módszerével az $y = 1,379 \cdot e^{0,8351x}$ függvény bizonyult. Ezen paraméterekkel kiszámítva a függvényértékeket, az előző ε_i , δ_i értékeket kapjuk, amelyek között megfigyelhetjük a levezetett összefüggést, miszerint $\varepsilon_i \approx y_i \delta_i$, illetve $-\delta_i = \ln\left(1 - \frac{\varepsilon_i}{y_i}\right)$. Ha ezután az y_i adatokra úgy illesztünk legjobban illeszkedő exponenciális függvényt, hogy nem végezzük el a linearizáló transzformációt, hanem az eredeti forma maradékainak négyzetösszegét minimalizáljuk (numerikus közelítéssel, iterációval), akkor a legjobban közelítő függvénynek az $y = 1,233 \cdot e^{0,914x}$ függvény bizonyul.

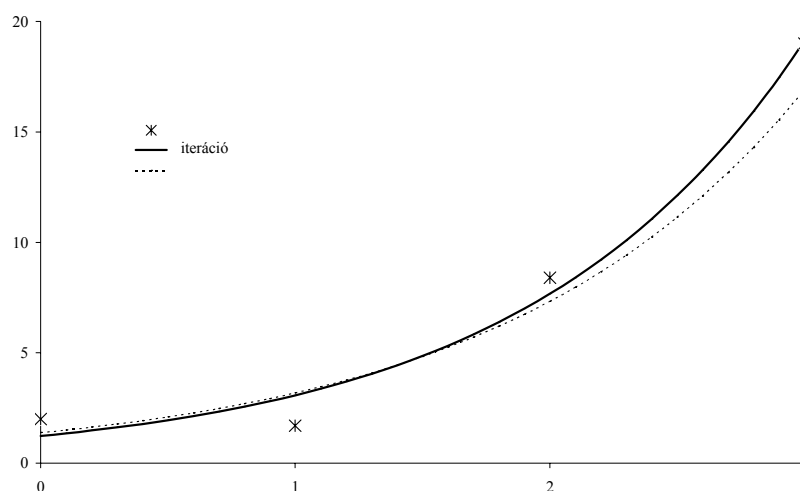
Az illeszkedések összehasonlítása érdekében kiszámítottuk az $I^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$ korrelációs indexet mind az iterációval meghatározott függvényre, mind a transzformációval meghatározott függvényre. Ekkor azt kaptuk, hogy:

$$- \text{transzformáció esetén: } y = 1,379 \cdot e^{0,8351x} \quad I^2 = 0,9584,$$

$$- \text{iteráció esetén: } y = 1,233 \cdot e^{0,914x} \quad I^2 = 0,9846.$$

Természetesen I^2 értékeiből is látszik, hogy a transzformált adatokra alkalmazva a legkisebb négyzetek elvét, majd az inverz transzformációval visszatranszformálva a paramétereket nem a legjobb regressziós függvényt kapjuk.

1. ábra. A transzformációval, illetve iterációval meghatározott exponenciális függvények



Az 1. ábrán is láthatók a levezetett összefüggések, miszerint a transzformációval meghatározott függvény „kisebb súllyal figyel” a nagyobb függvényértékek esetén elkövetett hibákra. Megjegyzendő, hogy a hatványfüggvény esetén hasonló nagyságrendű eltérést kapunk a transzformáció hatására.

Az itt tárgyalt exponenciális függvény egyik speciális esete az a konkáv függvény, amelyet leggyakrabban az $y = A(1 - ae^{bx})$, ($b < 0$) alakban szoktunk specifikálni, és amelyet egyszerű telítődési folyamatok leírására használhatunk. Az eddigiek alapján aligha meglepő az az állítás, miszerint a transzformált alakra alkalmazott legkisebb négyzetes illesztés ez esetben is más eredményt ad, mint az eredeti formára iterációs úton készített becslés.

Az eltéréseket illetően könnyen belátható, hogy a /8/-ban szereplő $\delta_i \approx \frac{\varepsilon_i}{y_i}$ össze-

függést ez esetben a $\delta_i \approx \frac{\varepsilon_i}{A - y_i}$ váltja fel, és a paraméterbecsléshez ennek négyzet-

összegét kell minimalizálni. Ennek következménye az, hogy a négyzetösszegben a kis y_i értékekhez tartozó ε_i hibák szerepelnek kisebb súllyal. Ez pedig azt jelenti, hogy a transzformációval kapott függvényünk a kis y értékekre lesz kevésbé pontos. Emlékeztetünk arra, hogy az eredetileg felírt (konvex) exponenciális függvény esetében ez fordítva volt. Ennek az eltérő eredménynek az egyszerű intuitív magyarázata az, hogy a telítődési szinthez közeledve egyre kisebb mozgástere van az illesztett görbének; egyre inkább, egyre kisebb ingadozásokkal simul a telítődési szinthez.

2.2. A hibatagok viszonya hiperbolikus regressziós függvény esetén

Ebben az esetben az illesztett függvény alakja: $y = \frac{1}{a + b \cdot x}$, a transzformáció a reciprokképzés, amelynek az inverze is a reciprokképzés lesz. Az eredeti adatokra és a hiperbolikus függvényre vonatkozó additív hibatag legyen ε , a transzformált adatokra lineáris függvényt illesztünk és az itt szereplő additív hibatag legyen δ .

Az a és b paraméterek változtatásával eredeti célunk a $\sum \varepsilon_i^2 \rightarrow \min$ meghatározása. Ennek egyik útja volt az adatok transzformálása és $\sum \delta_i^2 \rightarrow \min$ meghatározása. A következőkben bemutatjuk, hogy ez a két minimum nem esik egybe és meghatározzuk, hogy mekkora az eltérés az ε és δ értékek között.

Az eredeti adatokra illesztett függvény $y_i = \frac{1}{a + b \cdot x_i} + \varepsilon_i$, amit átalakítva

$$\frac{1}{y_i - \varepsilon_i} = a + b \cdot x_i \quad /9/$$

kapható. A transzformált alakból $\frac{1}{y_i} = a + b \cdot x_i + \delta_i$, majd

$$\frac{1}{y_i} - \delta_i = \frac{1 - y_i \cdot \delta_i}{y_i} = a + b \cdot x_i \quad /10/$$

adódik. A /9/ és /10/ jobb oldalának egyezősége miatt

$$\frac{1}{y_i - \varepsilon_i} = \frac{1 - y_i \cdot \delta_i}{y_i} \Rightarrow y_i = y_i - \varepsilon_i - y_i^2 \cdot \delta_i + y_i \cdot \delta_i \cdot \varepsilon_i,$$

amit átrendezve az $\varepsilon_i = \delta_i \cdot y_i^2 \cdot \left(\frac{\varepsilon_i}{y_i} - 1 \right)$ egyenlőséghez jutunk. Ekkor, ha $\left| \frac{\varepsilon_i}{y_i} \right|$ kicsi,

vagyis jó közelítés esetén $\frac{\varepsilon_i}{y_i} - 1 \approx -1$, így $\varepsilon_i \approx -\delta_i \cdot y_i^2$ azaz

$$\sum \varepsilon_i^2 \approx \sum \delta_i^2 \cdot y_i^4. \quad /11/$$

A korábbiakhoz hasonlóan itt is látható, hogy $\sum \varepsilon_i^2$ és $\sum \delta_i^2$ minimuma nem feltétlenül egyezik meg (sőt az egyezésnek kicsi az esélye), mivel $\sum \frac{\varepsilon_i^2}{y_i^4} \approx \sum \delta_i^2$ és mi $\sum \delta_i^2 \rightarrow$ minimumát határozzuk meg, ezért az ezzel egyenlő $\sum \frac{\varepsilon_i^2}{y_i^4}$ összegben a nagy y értékekhez tartozó ε_i hibák kisebb súllyal szerepelnek. Ez azt jelenti, hogy a transzformációval kapott regressziós függvényünk nagy y értékekre még kevésbé lesz pontos, mint exponenciális függvény esetén, vagyis a transzformációval továbbra sem kapjuk meg a legjobban illeszkedő hiperbolikus függvényt. A legjobban illeszkedő függvény meghatározásához a $\sum \varepsilon_i^2 \rightarrow$ minimumot kellett volna meghatározni. Nézzük meg, hogyan láthatók az eredmények egy egyszerű számpéldán.

2. példa

Tegyük fel, hogy az eredeti adatok körülbelül 10-20 százalékos hibával illeszkednek a normál hiperbolára

Eredeti adatok				Transzformált adatok			$\delta_i \cdot y_i^2 \left(\frac{\varepsilon_i}{y_i} - 1 \right)$
x_i	y_i	y_i számított	ε_i	$1/y_i$	$1/y_i$ szá- mított	δ_i	
0,1	12	2,5	9,5	0,083	0,4	-0,317	9,5
1	0,8	1,11	-0,31	1,25	0,902	0,349	-0,31
10	0,17	0,169	0,001	5,882	5,915	-0,033	0,001

Az értékeket transzformálva és a transzformált adatokra alkalmazva a legkisebb négyzetek elvét $a=0,3445$, illetve $b=0,557$ értékeket kapunk, azaz a legjobb függvénynek a linearizálás módszerével az $y = \frac{1}{0,3445 + 0,557 \cdot x}$ függvény bizonyul. Az ezen paraméterekkel kiszámított függvényértékeket és a hozzájuk tartozó ε_i , δ_i értékeket a 2. példa tartalmazza, amelyek között megfigyelhetjük a levezetett összefüggéseket, miszerint

$$\varepsilon_i \approx -\delta_i \cdot y_i^2, \text{ illetve } \varepsilon_i = \delta_i \cdot y_i^2 \cdot \left(\frac{\varepsilon_i}{y_i} - 1 \right).$$

Ha ezután az eredeti y_i adatokra iterációs módon illesztünk legjobban illeszkedő hiperbolikus függvényt az Excel-program segítségével, akkor a legjobban közelítő függvénynek az $y = \frac{1}{-0,045 + 1,28 \cdot x}$ függvény bizonyul.

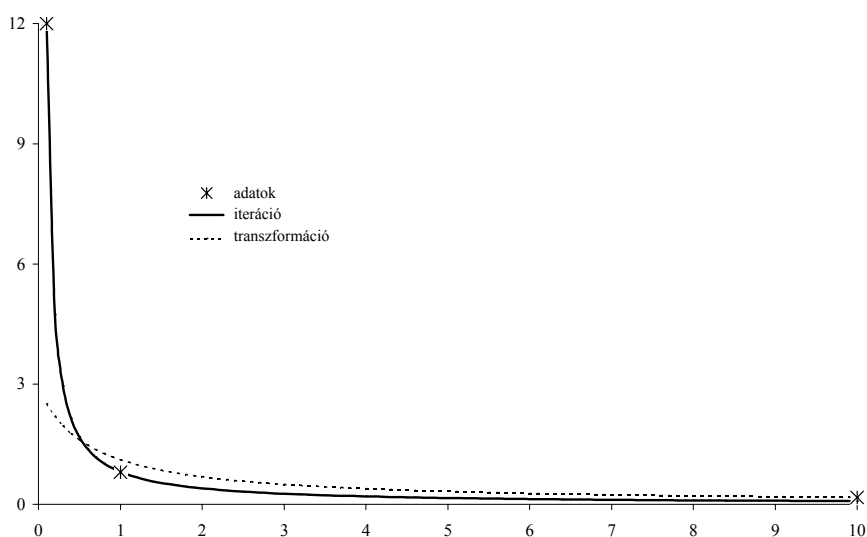
Az 1. példához hasonlóan itt is kiszámítottuk a korrelációs indexet mind az iterációval, mind pedig a transzformációval meghatározott függvényre:

$$\text{– transzformáció esetén: } y = \frac{1}{0,3445 + 0,557 \cdot x} \quad I^2 = -0,02 ,$$

$$\text{– iteráció esetén: } y = \frac{1}{-0,045 + 1,28 \cdot x} \quad I^2 = 0,9999 .$$

Annak felismerését, hogy nagy baj lehet az adatok transzformációjával történő regressziós függvényillesztéssel egy, a 2. példához hasonló feladat sugallta. Itt I^2 értéke negatív, ami azt jelenti, hogy a transzformációval meghatározott regressziós függvényünk rosszabb közelítő függvénye adatainknak, mint az \bar{y} konstans függvény. Ez alapján már várható volt, hogy a transzformációval meghatározott regressziós függvény ilyen esetben jelentősen eltér a legjobban illeszkedő regressziós függvénytől. Lehet olyan adatokat találni, ahol hasonló eredményt kapunk exponenciális függvény illesztése esetén is.

2. ábra. A transzformációval, illetve iterációval meghatározott hiperbolikus függvények



Az I^2 értéke alapján nyilvánvaló, hogy közelítőleg sem kaptuk meg eredményként a legjobb hiperbolikus regressziós függvényt. Az iterációs módszerrel pedig meg is határoztuk a legjobb hiperbolikus közelítő függvényt. A 2. példából az is látszik, hogy bár a δ_i értékek tűrhetőek, a nagy eredeti y érték esetén az $\varepsilon_i = 9,5$ tehát nagyon nagy, ahogy az előző levezetés alapján vártuk, hiszen $|\varepsilon| \approx |\delta| \cdot y^2$. Ezek az eredmények láthatók a 2. ábrán.

2.3. A hibatagok viszonya logisztikus regressziós függvény esetén

Ebben az esetben az illesztett függvény $y = \frac{A}{1 + a \cdot e^{b \cdot x}}$ alakú, az y érték transzformációja pedig összetett: A -val való osztás \rightarrow reciprok $\rightarrow -1 \rightarrow$ logaritmizálás.

Az eredeti és transzformált adatokra vonatkozó hibatagok jelölése ugyanaz mint eddig, és eredeti célunk itt is a $\sum \varepsilon_i^2 \rightarrow \min$ meghatározása. Ehelyett transzformáljuk az adatokat és a transzformáció utáni additív hibatagot minimalizáljuk, $\sum \delta_i^2 \rightarrow \min$. A következőkben itt is bemutatjuk, hogy ez a két minimum nem esik egybe és meghatározzuk, hogy mekkora az eltérés az ε és δ értékek között.

Az eredeti adatokra illesztett függvény $y_i = \frac{A}{1 + a \cdot e^{b \cdot x_i}} + \varepsilon_i$, átalakítva

$$\frac{A}{y_i - \varepsilon_i} - 1 = a \cdot e^{b \cdot x_i} \quad /12/$$

Az adatok transzformációja és a transzformált adatokra illesztett lineáris függvény leírása az alábbi: $\ln\left(\frac{A}{y_i} - 1\right) = \ln a + b \cdot x_i + \delta_i$ ezt átalakítva $\frac{A}{y_i} - 1 = a \cdot e^{b \cdot x_i} \cdot e^{\delta_i}$ majd

$$\left(\frac{A}{y_i} - 1\right) \cdot \frac{1}{e^{\delta_i}} = a \cdot e^{b \cdot x_i} \quad /13/$$

adódik. A /12/ és /13/ egyenlőségekből

$$\frac{A}{y_i - \varepsilon_i} - 1 = \left(\frac{A}{y_i} - 1\right) \cdot \frac{1}{e^{\delta_i}},$$

amit átrendezve

$$\frac{A - y_i + \varepsilon_i}{y_i - \varepsilon_i} = \frac{A - y_i}{y_i \cdot e^{\delta_i}}, \text{ majd innen}$$

$$y_i \cdot e^{\delta_i} = y_i - \frac{A \cdot \varepsilon_i}{A - y_i + \varepsilon_i}.$$

Tovább alakítva ezt

$$\frac{y_i}{A} (A - y_i + \varepsilon_i) \cdot (1 - e^{\delta_i}) = \varepsilon_i \text{ és mivel } \varepsilon_i \ll A - y_i, \text{ ezért}$$

$$(\varepsilon_i) \approx \left(\frac{y_i}{A} (A - y_i) \right) \cdot (1 - e^{\delta_i}), \text{ ahonnan}$$

$$(1 - e^{\delta_i})^2 \approx \frac{(\varepsilon_i)^2}{\left[-\frac{1}{A} \left(y_i - \frac{A}{2} \right)^2 + \frac{A}{4} \right]^2}.$$

Mivel δ_i^2 és $(1 - e^{\delta_i})^2$ minimuma egybeesik, $\sum \delta_i^2$ itt is ε_i^2 -k súlyozott összege és a súlyok láthatóan a szögletes zárójelben levő lefelé nyíló parabola értékeinek reciprokai. A legkisebb súly a nevező maximumához tartozik, ami azt jelenti, hogy az $y_i = \frac{A}{2}$ értékhez tartozó ε_i hiba szerepel a legkisebb súllyal míg 0-hoz illetve A -hoz közeli y_i értékek esetén az ε_i hiba nagyobb súllyal szerepel, ezért itt pontosabb lesz a transzformációval megalkotott regressziós függvény. Ez azt is jelenti, hogy a transzformációval kapott regressziós függvényünk $y_i = \frac{A}{2}$ értékekre lesz leginkább pontatlan.

3. példa

Illesszünk logisztikus regressziós függvényt az adatokra mind a transzformáció mind az iteráció módszerével. Legyen ebben a példában $A=100$. (A példára vonatkozó adatok a következő táblázatban találhatók.)

Eredeti adatok				Transzformált adatok			$\left[-\frac{1}{A} \left(y_i - \frac{A}{2} \right)^2 + \frac{A}{4} \right] \cdot (1 - e^{\delta_i})$
x_i	y_i	y_i számított (transzformáció)	ε_i	$\ln\left(\frac{A}{y_i} - 1\right)$	$\ln\left(\frac{A}{y_i} - 1\right)$ számított	δ_i	
0	3	4,407	-1,407	3,476	3,077	0,399	-1,427
1	20	13,949	6,051	1,386	1,820	-0,433	5,625
2	22	36,308	-14,310	1,265	0,562	0,704	-17,523
3	80	66,719	13,281	-1,386	-0,696	-0,691	7,981
4	95	87,577	7,423	-2,944	-1,953	-0,991	2,988
5	90	96,123	-6,123	-2,197	-3,211	1,013	-15,791

A kiindulási értékekre alkalmazva a leírt transzformációt és a transzformált adatokra lineáris függvényt illesztve $a = 21,693$, illetve $b = -1,258$ értékeket kaptunk, azaz a linearizálás módszerével adatainkra a legjobb logisztikus függvénynek az

$y = \frac{100}{1 + 21,693 \cdot e^{-1,258 \cdot x}}$ függvény bizonyult. Az adott paraméterértékeknek megfelelő

függvényre vonatkozó számított függvényértékek és a hozzájuk tartozó additív hibataragok szintén a táblázatban találhatóak. Azonban a kétszeres elhanyagolás miatt az

$\sum \frac{\varepsilon_i^2}{(A - y_i)^2} \approx \sum (1 - e^{\delta_i})^2 \approx \sum \delta_i^2$ összefüggésben, az ε_i értékeket csak kevésbé

pontosan követik az $(A - y_i) \cdot (1 - e^{\delta_i})$ értékek.

Ha ezután az eredeti adatokra iterációs módszerrel illesztünk legjobban illeszkedő

logisztikus függvényt, akkor ez a következő lesz $y = \frac{100}{1 + 139,66 \cdot e^{-2,028 \cdot x}}$.

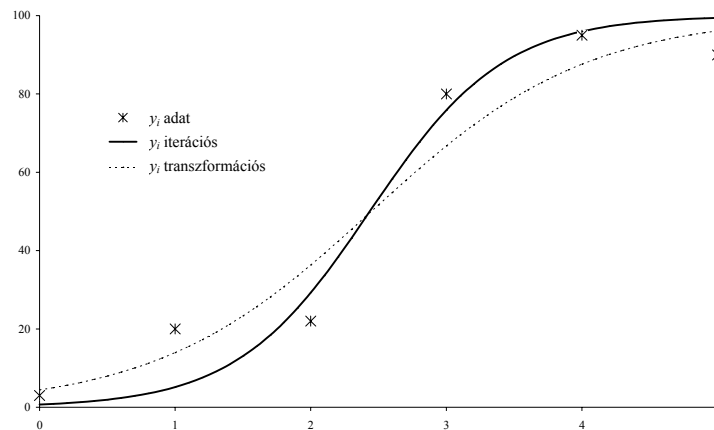
Mindkét függvény esetén kiszámítva a korrelációs indexet itt is érzékelhető az eltérés:

$$\text{– transzformáció esetén: } y = \frac{100}{1 + 21,693 \cdot e^{-1,258 \cdot x}} \quad I^2 = 0,939,$$

$$\text{– iteráció esetén: } y = \frac{100}{1 + 139,66 \cdot e^{-2,028 \cdot x}} \quad I^2 = 0,954.$$

A grafikonról látható, hogy jelentős eltérések vannak az iterációval meghatározott legjobban közelítő és a transzformációval meghatározott logisztikus függvényeink között. Az is látható a grafikonon, hogy a két középső érték esetén kapjuk a legnagyobb ε értéket.

3. ábra. A transzformációval, illetve iterációval meghatározott logisztikus függvények



3. Következtetések és nyitott kérdések

A korábbiakban bemutattuk azt, hogy a tankönyvek által javasolt linearizálás az esetek jó részében korántsem olyan ártatlan, mint ahogy azt olykor sugallják, hiszen ez a módszer a görbeillesztéskor nemritkán igencsak félrevezető eredményeket ad. Eredményeink nyilvánvalóan azt javasolják, hogy a linearizálás helyett (vagy inkább mellett) használjuk a nemlineáris legkisebb négyzetek módszerét, amely az eredeti nemlineáris feladatot oldja meg numerikus közelítéssel. Erre a gyorsan fejlődő számítástechnika egyre jobb lehetőségeket kínál.

Mint tudjuk, a gyakorlatilag bárki számára hozzáférhető Excel-programban lehetőségünk van a Solver-segédprogrammal szélsőérték meghatározására feltételek mellett és anélkül. A segédprogram használatakor meg kell adni, hogy melyik cella minimumát vagy maximumát keressük, valamint a szélsőérték típusát, és jelezni kell, hogy mely cellák iteratív változtatása mellett történjen a minimum meghatározása. Amennyiben vannak feltételek azokat is közölni kell.

Amikor regressziós függvények paramétereit kutatjuk, akkor a célcella minimumát keressük és a célcella tartalma a következő képlet és annak értéke:

$$\sum (y_i - y_{i, \text{számított}})^2 \rightarrow \min .$$

A módosuló cellák a paramétereket tartalmazzák és feltételek megadása nem szükséges. Amikor jelen írásban iterációs paraméter kereséséről beszélünk, mindig erre a módszerre gondolunk.

Az ezzel a módszerrel történő regressziós függvény meghatározásnál a paramétereknek kezdeti értéket kell ugyan adni, de elvben bármely függvénytípus esetén elkészíthető mechanikusan a paraméterek és a függvényértékek becslése. Amennyiben az eljárás valóban az abszolút minimumot eredményezi, a kapott paraméterek és függvényértékek torzítatlanok lesznek. Ugyanakkor felmerülhet az a kérdés, hogy a módszer minden függvénytípus és bármilyen kezdeti érték megadása esetén is konvergens-e, és ha igen a legkisebb négyzetekkel definiált függvény abszolút minimumát adja-e, vagy valamilyen lokális minimumot talál. Ezért érdemes azt a kérdést is felvetni, hogy bizonyos függvénytípusok esetében lehet-e analitikusan megoldható normálegyenleteket számolni. Végül megemlítjük, hogy amennyiben lehetőség van rá, érdemes egy feladatot több különböző, feltehetően eltérő megoldó algoritmusokat használó programcsomagok segítségével elvégezni. Az esetleg eltérő eredmények figyelmeztethetnek az említett hibákra.

Irodalom

- ALMÁSI A. – TÓTH Z. – TÖRCSVÁRI ZS. [2005]: *Have you ever seen a best exponential regression function?* Tagungsband. 9. Thüringisch – Ungarisches Symposium. Fachhochschule. Jena.
- HUNYADI L. – VITA L. [2004]: *Statisztika közgazdászoknak*. KSH. Budapest.
- MUNDRUCZÓ GY. [1981]: *Alkalmazott regressziószámítás*. Akadémiai Kiadó. Budapest.
- RAPPAL G. [2001]: *Üzleti statisztika Excellel*. KSH. Budapest.
- SZÜCS I. [2002]: *Alkalmazott statisztika*. Agroinform Kiadó. Budapest.