

Nemlineáris regressziók alkalmazása gyakorlati példákban

Csanády Viktória
 NymE EMK Matematikai Intézet

A természetben előforduló különböző folyamatok vizsgálata során nyert, egy független és egy függőváltozós adatsorokra regressziós eljárással matematikai függvények illeszthetők, melyek meghatározzák a folyamatok törvényszerűségét. Az adatsorok által meghatározott pontok grafikus szemlélete alapján mód van megfelelő illesztendő függvény vagy függvények kiválasztására. A helyes döntést alapvetően a számítógépes regressziós eljárás végrehajtása során nyert 1-hez legközelebb álló korrelációs együttható (R) indokolhatja amellet, hogy a kiválasztott függvény számított paraméterei a valóságnak megfelelően értelmezhetők legyenek. A bemutatásra és elemzésre kerülő adatsorok a természetben előforduló folyamatok adatsorait modellezik a gyorsabb és egyszerűbb regressziós eljárások alkalmazása és értékelése érdekében.

1. Növekedési függvények

1.1. Telítési függvény (Awrami)

Az első adatsor az idő függvényében a fanövekedés értékeit vizsgálja. Az adatsor előzetes áttekintése vagy grafikus ábrázolása alapján könnyen megállapítható, hogy a függvény illesztéséhez telítési függvény alkalmazása a célszerű.

A matematikai alak: $y = a(1 - e^{-(bx)^c}) + d$.

A regressziós eljáráshoz szükséges kezdőértékeket, a nyert paraméter értékeket, a korrelációs együtthatót és az értelmezést az alábbi táblázat tartalmazza.

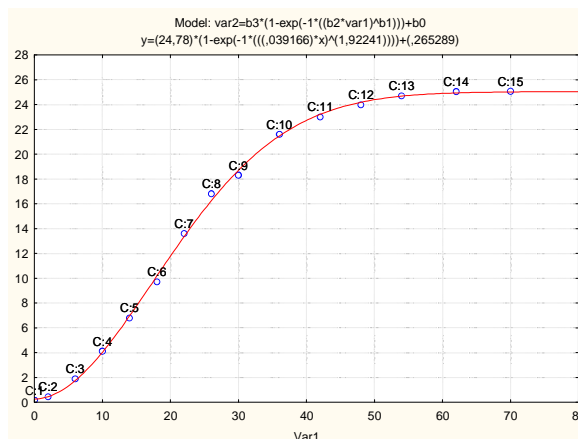
Kezdőértékek: $b_3=b_2=b_1=b_0=0,1$ (a programban alapbeállításként szereplő értékek, módosítást nem igényelnek) $var1=$ évek száma (év), $var2=$ famagasság (m).

Model: $var2=b_3*(1-\exp(-1*((b_2*var1)^{b_1}))+b_0$ (pelda1)				
Dep. var: Var2 Loss: (OBS-PRED)**2				
Final loss: ,929228643 R= ,99965 Variance explained: 99,929%				
N=15	b3	b2	b1	b0
Estimate	24,78002	0,039166	1,922410	0,265289

Értelmezés: $b_3+b_0=$ az elért legnagyobb (végső) famagasság (m), $b_0=$ a kezdő famagasság (m). Megadható az a $var1$ érték (x), melynél a határértéktől való eltérés 1%-os a $var2$ -re nézve.

Ez az alábbi képlettel számítható: $lnx = \frac{1}{b_1} \ln \left(\ln \frac{100}{1+\frac{b_0}{b_3}} \right) - lnb_2$.

Az illesztés grafikus reprezentációja:



Az előző függvény illesztése módosított kezdőértékekkel is végrehajtható, az eredmény nem változik, a függvény illesztése a kezdőértékekre kevésbé érzékeny. Nem ezt tapasztaljuk más típusú növekedési görbék estén. Az Awrami féle telítési függvényen kívül még számos hasonló növekedési görbe létezik. Az egyszerűbb nem rendelkezik inflexiós ponttal, az összetettebbek igen. Az alábbiakban a felhasznált példaszor alkalmazásával megadásra kerül illesztési eredményük, a paraméterek, a korrelációs együttható értéke, végül, de nem utolsón sorban a futtatásnál módosított kezdőértékek, melyek módosítása nélkül nem kapunk eredményt, vagy ha igen csak gyenge korrelációval. A paraméterek kezdőértékeinek kiokoskodása a függvény matematikai jellemzőinek ismeretében történhet, az adatszort figyelembevételével.

Minden egyes illesztésnél a var2 a famagasságot jelöli, var1 a független változó, az időt.

1.2. Bertalanffy növekedési függvénye

Matematikai alak: $y = a(1 - be^{-cx})$. Kezdőértékek: $b_2=b_1=b_0=0,1$.

	Model: var2=b2*(1-b1*exp(-1*(b0*var1))) (pelda1)		
	Dep. var: Var2 Loss: (OBS-PRED)**2		
	Final loss: 34,694807684 R= ,98669 Variance explained: 97,356%		
N=15	b2	b1	b0
Estimate	31,56415	1,069303	0,028769

1.3. Mitscherlich növekedési függvénye

Matematikai alak: $y = a(1 - e^{-bx})^c$. Kezdőértékek: $b_2=b_1=b_0=1$.

	Model: var2=b2*(1-Exp(-1*b1*var1))^b0 (pelda1)		
	Dep. var: Var2 Loss: (OBS-PRED)**2		
	Final loss: 2,393042626 R= ,99909 Variance explained: 99,818%		
N=15	b2	b1	b0
Estimate	25,85456	0,077052	3,136495

1.4. Richards növekedési függvénye

Matematikai alak: $y = a(1 - be^{-cx})^d$. Kezdőértékek: $b_3=10$ $b_2= -1$ $b_1=0,5$ $b_0=0,1$.

	Model: var2=b2*(1-Exp(b1*var1))^b0 (pelda1)		
	Dep. var: Var2 Loss: (OBS-PRED)**2		
	Final loss: 2,393042626 R= ,99909 Variance explained: 99,818%		
N=15	b2	b1	b0
Estimate	25,85457	-0,077052	3,136496

1.5. Chapman-Richard függvény

Matematikai alak: $y = a(1 - e^{bx})^c$. Kezdőértékek: $b_2=10$ $b_1= -1$ $b_0=0,5$.

Model: $var2=b_2*(1-Exp(b_1*var1))^b_0$ (pelda1)			
Dep. var: Var2 Loss: (OBS-PRED)**2			
Final loss: 2,393042626 R= ,99909 Variance explained: 99,818%			
N=15	b2	b1	b0
Estimate	25,85457	-0,077052	3,136496

1.6. Colin-Fokasz függvény

Matematikai alak: $y = a + \frac{(b-a)}{(1+ce^{-d(x-h)})^c}$.

Kezdőértékek: $b_0=10$ $b_1= -1$ $b_2=0,5$ $b_3=0,1$ $b_4=1$.

Model: $var2=b_0+(b_1-b_0)/(1+b_2*Exp(-1*b_3*(var1-b_4)))^{1/b_2}$ (pelda1)					
Dep. var: Var2 Loss: (OBS-PRED)**2					
Final loss: ,717894104 R= ,99973 Variance explained: 99,945%					
N=15	b0	b1	b2	b3	b4
Estimate	25,39397	-1,28865	1,733523	-0,155553	18,46158

A fenti példák azt illusztrálják, hogy bár számos telítési, nevezzük inkább növekedési folyamatot részében vagy teljességében (negatív értékekre is értelmezett függvények szerepeltek a felsoroltak között) leíró függvényeket ismerünk, ezek alkalmazása a gyakorlatban a számítógépes statisztikai programok használata esetén is gondot okoz. Az alkalmazott modell megválasztás esetén szem előtt tartandók az igények a paraméterek értelmezhetőségére, valamint a kezdőértékek megválasztásának, illetve kiválasztásának egyszerűségére, nem beszélve a modell alkalmazhatóságáról (az Awrami féle függvény akkor is alkalmazható, ha nincs inflexiós pont). Ezen kívül a modell értelmezési tartományának vizsgálata sem elhanyagolható, a már említett negatív független változók (példákban az idő) vonatkozásában, hiszen ez nem értelmezhető.

2. Rönkleltár

Egy faraktárban elhelyezett válogatással nyert rönkök leltárát elemzi az átmérő függvényében található darabszám szerint a feladat. Az adatsor egyszerű áttekintése alapján rögtön megállapítható, hogy megfelelően transzformált Gauss-görbe illesztése a célszerű.

A matematikai alak: $y = \frac{a}{e^{(b(x-c))^2}} + d$. Kezdőértékek: $b_3=b_2=b_0=1$ $b_1=40$.

A változók: $var1=$ fatörzsátmérő (cm), $var2=$ darabszám (db).

Model: $var2=b_3/exp((b_2*(var1-1*b_1))^2)+b_0$ (pelda2)				
Dep. var: Var2 Loss: (OBS-PRED)**2				
Final loss: 47,132016006 R= ,99631 Variance explained: 99,264%				
N=15	b3	b2	b1	b0
Estimate	62,00369	0,119303	40,13033	-3,50099

A $b_3+b_0=a$ legnagyobb darabszám, $b_1=a$ legnagyobb darabszámhoz tartozó átmérő.

3. Faanyagszárítás

A faanyag szárítási folyamata során nyert értékeket vizsgálja a feladat az idő függvényében. Az adatsor áttekintése vagy grafikus ábrázolása alapján eldönthető, hogy megfelelően transzformált tangens hiperbolikus görbe illesztése vezet helyes eredményre, és értelmezhető paraméterekhez.

A matematikai alak: $y = a \tanh(b(x - c)) + d$. Kezdőértékek: $b_3=b_2=b_0=1$ $b_1=15$.

A változók: $var1$ = az eltelt idő (óra), $var2$ = a nedvességtartalom (%).

Model: $var2=b_3*\tanh(b_2*(var1-1*b_1))+b_0$ (példa3)				
Dep. var: Var2 Loss: (OBS-PRED)**2				
Final loss: ,718110358 R= ,99976 Variance explained: 99,952%				
N=16	b3	b2	b1	b0
Estimate	-11,8755	0,185932	14,02402	14,92063

A b_0-b_3 = a kezdeti nedvességtartalom (%), b_0+b_3 = a végső nedvességtartalom (%).

4. Anyaglehülés

A negyedik feladat az idő függvényében történő anyag lehülés értékeit tartalmazza. Az adatsor egyszerű áttekintése vagy esetleges grafikus ábrázolása alapján itt megállapítható, hogy a függvényillesztéshez egy megfelelően transzformált exponenciális („negatív exponenciális”) görbe alkalmazása lehet a legmegfelelőbb.

A matematikai alak: $y = \frac{a}{e^{b(x-c)}} + d$. Kezdőértékek: $b_3=b_2=b_1=b_0=1$.

A változók: $var1$ = idő (min), $var2$ = hőmérséklet (C°).

Model: $var2=b_3/\exp(b_2*(var1-1*b_1))+b_0$ (Spreadsheet4)				
Dep. var: Var2 Loss: (OBS-PRED)**2				
Final loss: ,765464092 R= ,99984 Variance explained: 99,969%				
N=15	b3	b2	b1	b0
Estimate	14,92291	0,153119	7,191448	5,063373

A b_0 = a mért legalacsonyabb hőmérséklet (véghőmérséklet).

$b_3e^{(b_2b_1)} + b_0$ = a mért legmagasabb hőmérséklet (kezdőhőmérséklet).

5. Hangerő ingerérték

Az ötödik adatsor a hangerő függvényében észlelhető ingerértékek nyert elméleti adatait mutatja. A pontsor grafikus ábrázolása alapján logaritmikus függvény illesztése látszik legmegfelelőbbnek, ha az alkalmazott függvényt előzetesen megfelelően transzformáljuk, lehetővé téve az origóból való kiindulást a kezdő adatpár miatt.

A matematikai alak: $y = a \ln(b(x - c)) + d$. Kezdőértékek: $b_3=b_2=b_0=1$ $b_1=-1$.

A változók: $var1$ = hangerő (dB), $var2$ = ingerérték (i).

Model: $var2=b_3*\log(b_2*(var1-1*b_1))+b_0$ (Spreadsheet8)				
Dep. var: Var2 Loss: (OBS-PRED)**2				
Final loss: ,001401137 R= ,99999 Variance explained: 99,998%				
N=16	b3	b2	b1	b0
Estimate	1,738903	1,539166	-1,99920	-1,95678

A b_0 = a mért legalacsonyabb ingerérték, $b_3 \ln(b_2(100 + b_1)) - b_0$ = a legmagasabb mért ingerérték.

6. Lövedékpálya

A hatodik adatsor egy kilőtt lövedék útjának adatait mutatja. A pontsor értékeinek áttekintése, a gyakorlati ismeretek és elemzés alapján könnyen megállapítható, hogy a görbeillesztésre parabola, másodfokú hatvány függvény a megfelelő, a szükséges transzformálással.

A matematikai alak: $y = a(x - b)^2 + c$. Kezdőértékek: $b_2 = -0,001$, $b_1 = 80$, $b_0 = 0,1$.

A változók: var1 = a vízszintesen mért távolság (m), var2 = a lövedék magassága.

Model: $\text{var2} = b_2 * (\text{var1} - b_1)^2 + b_0$ (Spreadsheet12)			
Dep. var: Var2 Loss: (OBS-PRED)**2			
Final loss: ,000808408 R= 1,0000 Variance explained: 99,999%			
N=15	b2	b1	b0
Estimate	-0,000802	99,89942	7,999709

A b_0 = a lövedék legnagyobb magassága, b_1 = az a távolság ahol a lövedék legmagasabban van, $b_0 - b_2 * b_1^2$ = a lövedék kiindulási magassága.

7. Lázgörbe

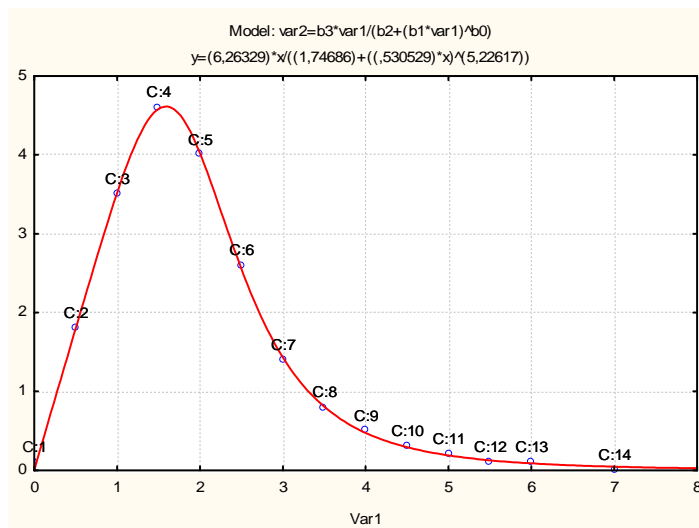
Az adatsor egy betegséggel együtt járó időbeli lázváltozás adatait mutatja (lázgörbe). Az adatsor elsődleges elemzése alapján transzformált Gauss-görbe alkalmazása látszik lehetségesnek. A grafikus ábrázolás azonban mutatja, hogy a görbe aszimmetrikus, hirtelen emelkedő és lassan csökkenő kellene, hogy legyen. Ezért a függvényillesztéshez egy speciálisan kialakított matematikai formulát szükséges alkalmazni.

A matematikai alak: $y = \frac{ax}{b + (cx)^d}$. Kezdőértékek: $b_3 = b_2 = b_1 = b_0 = 1$.

A változók: var1 = az idő (nap), var2 = hőmérséklet 36 C° felett.

Model: $\text{var2} = b_3 * \text{var1} / (b_2 + (b_1 * \text{var1})^{b_0})$ (lázgörbe4p)				
Dep. var: Var2 Loss: (OBS-PRED)**2				
Final loss: ,018167923 R= ,99973 Variance explained: 99,946%				
N=14	b3	b2	b1	b0
Estimate	6,263286	1,746860	0,530529	5,226167

A b_3/b_2 = a kezdő meredekség azaz, az egy nap alatti induló hőemelkedés értéke (betegség jellemző adat).



8. Napi levegő hőmérséklet

A nyolcadik adatsor egy 24 órás levegőhőmérséklet változás értékeit mutatja, éjfél-től-éjfélig. Az értékpárok elemzése és grafikus áttekintése itt jól mutatja, hogy megfelelően transzformált szinusz függvény alkalmazása a célszerű, ami a gyakorlati ismeretek alapján kézenfekvő.

A matematikai alak: $y = a \sin(b(x - c)) + d$. Kezdőértékek: $b_3=b_2=b_1=1$ $b_0=5$.

A változók: var1= idő (óra), var2= a hőmérséklet (C°).

	Model: var2=b3*Sin(b2*(var1-1*b1))+b0 (napihőingadozás)			
	Dep. var: Var2 Loss: (OBS-PRED)**2			
	Final loss: ,014721151 R= ,99996 Variance explained: 99,992%			
N=13	b3	b2	b1	b0
Estimate	5,199216	-0,261772	-3,00344	9,997983

A $b_0-b_3=$ a legalacsonyabb, $b_0+b_3=$ a legmagasabb hőmérséklet (C°) $b_1+6=$ a legalacsonyabb, $b_1+6+12=$ a legmagasabb hőmérséklet időpontja (óra).

9. Ötvözet vezetőképesség

A kilencedik adatsor egy olyan modellkísérlet adatait tartalmazza, ahol két fémből készült ötvözet vezetőképességének vizsgálata történt a százalékos összetétel függvényében. Az adatsor áttanulmányozása és grafikus elemzése alapján látható, hogy két határérték mutatkozik, azonban az ezek által meghatározott tartományon kívüli értékek is jelen vannak köztesen. Ez azt jelenti, hogy egyszerű klasszikus transzformált matematikai függvénnyel az illesztés nem látszik megoldhatónak. Ebből kiindulva, valamint a határértékek jelenléte miatt két különböző tangens hiperbolikus függvény megfelelően transzformált összege adhatja a jó regressziót.

A matematikai alak: $y = a \tanh(b(x - c)) + d \tanh(f(x - g)) + h$.

Kezdőértékek: $b_6=b_5=b_4=b_3=b_2=b_1=b_0=1$.

A változók: var1= a százalékos összetétel (%), var2= vezetőképesség (s/m 10^6).

Model: var2=b6*TanH(b5*(var1-1*b4))+b3*TanH(b2*(var1-1*b1))... (vezetőképesség)							
Dep. var: Var2 Loss: (OBS-PRED)**2							
Final loss: ,000998882 R= ,99992 Variance explained: 99,984%							
N=21	b6	b5	b4	b3	b2	b1	b0
Estimate	0,334742	0,084865	67,62294	-0,814370	0,081759	36,68642	2,097139

Ha $b_4 > b_1$ akkor $\frac{b_5 b_4 + b_2 b_1}{b_5 + b_2}$ = legkisebb vezetőképességhez tartozó %-os összetétel értéke.

var1=0 helyettesítéssel kiszámítható a 0%-hoz tartozó vezetőképesség értéke.

var1=100 helyettesítéssel kiszámítható a 100%-hoz tartozó vezetőképesség értéke.

10. Huzalfeszítés

A tizedik adatsor a huzal megnyújtás függvényében jelentkező feszítőerő adatpár sorát tartalmazza, azaz a huzalszakadás folyamatát jellemzi mért értékekkel. Az adatsor elemzése és grafikus áttanulmányozása alapján látható, hogy a kezdő és végső határérték egyforma (0), de a változás hirtelen mértékben aszimmetrikus. Ez azt jelenti, hogy klasszikus egyszerű transzformált matematikai függvénnyel az illesztés nem adódik megoldhatónak. Így következik, hogy bonyolultabb függvény kombináció használandó, jelen esetben is a két megfelelően transzformált tangens hiperbolikus függvény összege adhat várhatóan megbízhatóan jó regressziót 1-hez közeli korrelációs együtthatóval, és jól értelmezhető és értékelhető paraméterekkel.

A matematikai alak: $y = a \tanh(b(x - c)) + d \tanh(f(x - g)) + h$.

Kezdőértékek: $b_6=b_5=b_3=b_2=b_0=1$, $b_4=5$, $b_1=10$.

A változók: var1= a megnyújtás (mm), var2= a feszítő erő (N 10⁴).

Model: var2=b6*TanH(b5*(var1-1*b4))+b3*TanH(b2*(var1-1*b1))... (feszítőerő)							
Dep. var: Var2 Loss: (OBS-PRED)**2							
Final loss: ,000261920 R= 1,0000 Variance explained: 100,00%							
N=12	b6	b5	b4	b3	b2	b1	b0
Estimate	2,707871	0,591957	6,005157	-2,70603	4,378056	11,23117	-0,000973

Ha $b_4 > b_1$, akkor $z = \frac{b_5 b_4 + b_2 b_1}{b_5 + b_2}$ = a legnagyobb feszítő erőhöz tartozó megnyújtás (mm) értéke. $var2(z) = b_6 * \tanh(b_5 * (z - 1 * b_4)) + b_3 * \tanh(b_2 * (z - 1 * b_1)) + b_0 =$ a legnagyobb feszítő erő (a szakadást létrehozó erő).

11. Radioaktív sugárintenzitás

A tizenegyedik adatsor a radioaktív anyag idő függvényében észlelhető sugárintenzitásának értékeit tartalmazza. Az adott értékpár sorozat áttekintése alapján könnyen megállapítható, hogy egy negatív exponenciális függvény illesztése lehet a megfelelő. Mivel az ilyen jellegű vizsgálatoknál a felezési idő meghatározása is elemi követelmény, ezért a matematikai alak megfelelő transzformálása szükséges.

A matematikai alak: $y = \frac{a}{b(\frac{x}{c})}$ (vagy: $y = \frac{a}{2(\frac{x}{c})}$).

A változók: var1= az idő (hónap) var2= a sugárintenzitás (10⁶Bq).

Kezdőértékek: $b_2=1$, $b_1=b_0=2$ (módosított értékek).

	Model: $\text{var2} = b2/b1^{(\text{var1}/b0)}$ (felezési idő)		
	Dep. var: Var2 Loss: (OBS-PRED)**2		
	Final loss: ,002096773 R= ,99998 Variance explained: 99,996%		
N=11	b2	b1	b0
Estimate	7,993744	2,018864	2,931419

A megoldás akkor érvényes, ha $1,98 < b1 < 2,02$. Ekkor a felezési idő: $b0$, a kezdeti intenzitás $b2$.

12. Összefoglalás

A fentiekben bemutatásra került néhány növekedési görbe illesztése, melyek eredményeiből látható, hogy a számítógépes program használata sokszor igényel matematikai apparátust, itt konkrétan a kezdőértékek meghatározása. A további példák pedig azt mutatják, hogy a fizikailag értelmezhető paraméterű modell választása fontos, de ismerni kell hozzá a megfelelő függvénykompozíciót, annak additív esetleg multiplikatív esetét is.