Nelineární regrese - linearizace

Často se setkáme s případem, že teoretická závislost mezi dvěma proměnnými není lineární. V tomto případě je někdy řešením problému linearizace, tedy transformace závislosti do lineárního tvaru.

Nejčastěji se linearizace provádí logaritmováním a substitucí.

Logaritmus kladného reálného čísla *x* při základu *a* je takové reálné číslo

$$y = \log_a x$$

pro které platí

$a^y = x$.

Dále budeme uvažovat pouze přirozené logaritmy se základem *e* = 2,71828182845904523536028747135… (Eulerovo číslo)

Při tom budeme používat značení $\log_e x = \ln x$.

Nelineární regrese - linearizace

Pravidla pro počítání s logaritmy:

$\ln e^x = x$

$\ln (a.b) = \ln a + \ln b$

$\ln (a/b) = \ln a - \ln b$

$\ln a^n = n.\ln a$

$$\ln \sqrt[n]{x} = \frac{1}{n} \ln x$$

Např.:

$$\ln\left(\frac{4}{3}\pi r^3\right) = \ln\left(\frac{4}{3}\pi\right) + 3\ln r = \ln 4 + \ln \pi - \ln 3 + 3\ln r$$

Nelineární regrese - linearizace

Příklady linearizovatelných funkcí

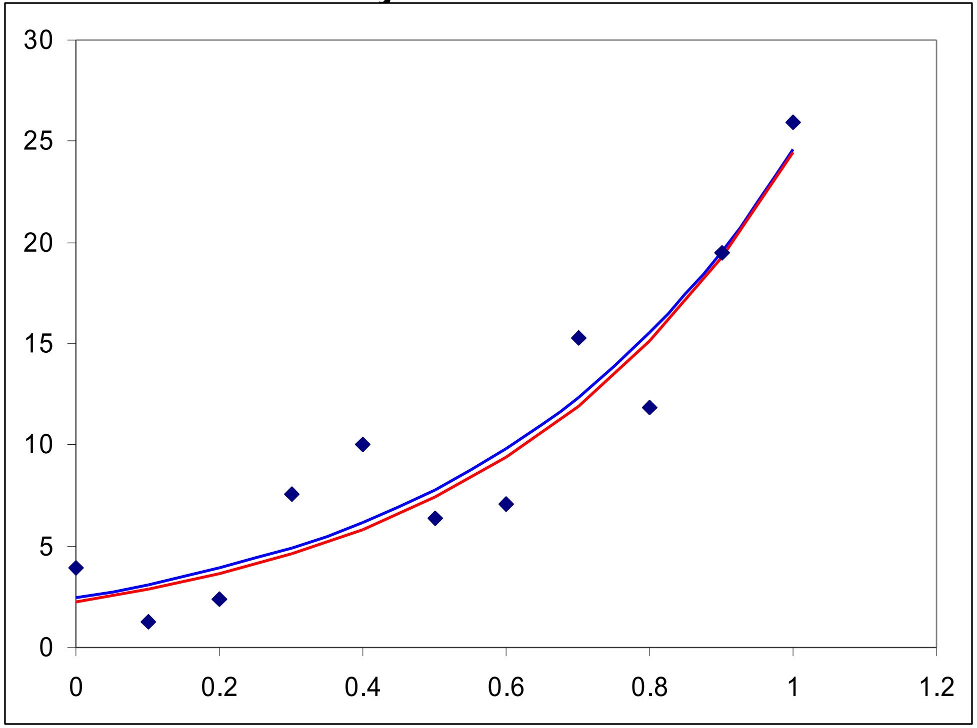
název funkce substituce x substituce y linearizace

název	funkce	subst.x	subst. y	
mocninná	$z = a.t^b$	$x = \ln t$	$y = \ln z$	$\ln z = \ln a + b.\ln t$
logaritmická	$z = a + \ln t$	$x = \ln t$	$y = z$	$z = a + \ln t$
exponenciální	$z = a.\exp(b.t)$	$x = t$	$y = \ln z$	$\ln z = \ln a + b.t$
hyperbolická	$z = a + b/t$	$x = 1/t$	$y = z$	$z = a + b.x$

Nelineární regrese - linearizace

Problémy linearizace

- velmi omezená množina funkcí
- logaritmování mění velikost odchylek, výsledná funkce neminimalizuje sumu čtverců odchylek



Linearizace logaritmováním

$$y=(2{,}25\, e^{2{,}39\, x})$$

Metoda nejmenších čtverců

$$y=(2{,}30\, e^{2{,}47\, x})$$

Nelineární regrese - linearizace

Příklad: Závislost vodivosti polovodiče na teplotě je dána vztahem

$$G = G_0 . \exp\left(-\frac{E_g}{2kT}\right)$$

kde E_g je šířka pásu zakázaných energií, k je Boltzmannova konstanta a T absolutní teplota. Z experimentálních dat určete hodnotu E_g.

- Zlogaritmujeme vztah:

$$\ln G = \ln\left(G_0 e^{-\frac{E_g}{2kT}}\right) = \ln G_0 - \frac{E_g}{2kT}$$

- Označíme-li $y = \ln G$, $x = 1/T$, $a = \ln G_0$ a $b = E_g/2k$, dostaneme rovnici přímky $y = a + b.x$.

Nelineární regrese - linearizace

	A	B	C	D
1	T [K]	G[Ω⁻¹]	1/T	ln G
2	293	0.34	0.0034	-1.09
3	313	0.64	0.0032	-0.44
4	333	3.61	0.0030	1.28
5	353	10.4	0.0028	2.34
6	373	27.8	0.0027	3.32
7				
8	**Linregrese:**			
9	-6322	20.21		
10	540	1.641		
11	0.98	0.313		
12				

- Ve sloupcích A a B jsou naměřené hodnoty teploty a vodivosti, ve sloupcích C a D pak 1/*T* a ln *G*.
- Odhadneme parametry přímky pomocí funkce =LINREGRESE(D2:D6;C2:C6;1;1)
- Parametr a není důležitý, závisí na rozměrech vzorku.
- Parametr a jeho směrodatná odchylka je v buňkách A9 a A10 b = -(63 ± 6).10^2
- Podle substituce $E_g = 2.k.b$.
- E_g = (1,74 ± 0,14).10^{-19} J = = (1,09 ± 0,09) eV.

Nelineární regrese - =LINREGRESE()

Nelineární regresi lze provádět i pomocí funkce =LINREGRESE() v případě, že prokládaná funkce je lineární kombinací jednoduchých funkcí - typicky polynomiální regrese.
Např. funkce $y = a_2x^2 + a_1x + b$.
Vytvoříme novou proměnnou $x_2 = x^2$ a budeme prokládat lineární funkci dvou proměnných.
Podobný postup lze po užít i u funkcí typu:
$y = a_1x + a_2/x + b$,
nebo i u vícenásobné regrese $y = a_1x_1 + a_2x_2 + a_3x_1x_2 + b$.

Nelineární regrese - =LINREGRESE()

Odpor kovového vodiče závisí na teplotě podle vztahu
$R = R_0 + A.t + \mathrm{B}.t^2$.
Aproximujme touto funkcí teplotní závislost odporu platinového vodiče

H2 {=LINREGRESE(F2:F9;D2:E9;1;1)}

	A	B	C	D	E	F	G	H	I	J
1	teplota [°C]	odpor [mΩ]		t	t^2	R				
2	20	131.9		20	400	131.9		-4.43E-04	0.50317	121.641
3	40	140.5		40	1600	140.5		7.1E-05	0.01308	0.51327
4	60	150.2		60	3600	150.2		0.999776	0.3679	#N/A
5	80	159.3		80	6400	159.3		11143.25	5	#N/A
6	100	167.4		100	10000	167.4		3016.482	0.67675	#N/A
7	120	176.0		120	14400	176.0				
8	140	183.2		140	19600	183.2				
9	160	190.8		160	25600	190.8				

Nelineární regrese - Řešitel

Funkce Řešitel v Excelu umožňuje nastavit velikost několika proměnných tak, aby výsledek byl roven nějaké hodnotě, případně byl extremální.
Můžeme tedy hledat parametry modelové funkce tak, aby suma čtverců odchylek funkčních hodnot od experimentálních dat byl minimální.

Nelineární regrese - Řešitel

Pokusme se proložit experimentálními daty (sloupce A a B) funkci

$$y = \sqrt{a + bx^2}$$

Do sloupce C vložíme funkci =ODMOCNINA(G1+G2*A2*A2), kde G1 je parametr *a* a G2 je *b*. Ve sloupci D je čtverec rozdílu mezi buňkami ve sloupci B a C, buňka D12 obsahuje součet sloupce D.

	A	B	C	D	E	F	G	H	I	J	K
1	x	y	y_{mod}	CO		a	88.28654				
2	1	9.73	9.50	0.049098		b	2.04777				
3	2	9.83	9.82	5.74E-05							
4	3	10.12	10.33	0.042468							
5	4	10.75	11.00	0.062139							
6	5	11.80	11.81	0.000164							
7	6	12.62	12.73	0.010897							
8	7	14.33	13.73	0.355871							
9	8	14.63	14.81	0.032521							
10	9	16.08	15.94	0.019341							
11	10	16.91	17.12	0.043453							
12			suma	0.61601							
13											
14											
15											
16											
17