

10. Licenční studium

Statistické zpracování dat

3.3 Tvorba nelineárních regresních modelů v analýze dat

Vladimír Bajzík

Liberec, únor, 2007

Zadání

Vypracujte písemně s využitím paketu ADSTAT a vyřešte 3 příklady. Příklady postavte z dat vašeho pracoviště nebo naleznete v literatuře. Každý příklad bude mít strukturu dle příkladů v učebnici M.Meloun, J.Militký: Kompendium statistického zpracování experimentálních dat, Academia Praha 2002, t.zn. Nadpis příkladu, Zadání, Data, Program, Řešení, Output, Obrázky, Závěr.

Příklad 1.

Změna tepelného odporu koberců vlivem opotřebení.

Základní užitečnou vlastností koberců je její tepelná izolace, která může být např. charakterizována tepelným odporem R [Km^2W^{-1}] tj, odporem vůči vedení tepla. Během používání dochází k opotřebení koberců a tím i ke změně tepelných vlastností. Cílem tohoto příkladu je modelovat změny tepelného odporu pomocí simulovaného opotřebení. Pro simulaci opotřebení bylo použito zařízení tetrapod. Tepelný odpor byl měřen po 10, 20, 30, 40, 50, 100 a 200 tisících otáčkách (cyklech).

Použitý software

Adstat 2.0 – modul nelineární regrese

Data

Počet cyklů	Tepelný odpor
0	124,2
10000	101,2
20000	95,6
30000	90,1
40000	89,9
50000	88,8
100000	88,1
200000	86,6

Řešení

Navržené testované modely jsou

$$\text{Model A: } y = \beta_1 + \beta_2 e^{-\beta_3 x}$$

$$\text{Model B: } y = \beta_1 - \beta_2 x^{\beta_3}$$

$$\text{Model C: } y = \beta_1 - \frac{\beta_2}{1 + \beta_3 x}$$

Koeficienty β_i jsou odhadovány pomocí koeficientů p_i . Počáteční odhady jsou pro všechny modely shodné

$$p_1=100$$

$$p_2=30$$

$$p_3=0,0001$$

Model A

1) Odhady parametrů

parametr	Bodový Odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení	Poloviční délka konfidenčního intervalu spočtená z: délky poloos	Poloviční délka konfidenčního intervalu spočtená z: maxim
p ₁	88,04	0,667	-0,0196	-0,0222	±2,304	±2,687
p ₂	35,85	1,346	0,0262	0,0732	±5,465	±5,496
p ₃	8,99E-5	8,282E-6	4,575E-7	0,509	±2,449E-5	±3,340E-5

Relativní vychýlení u všech koeficientů je menší než 1%, koeficienty lze považovat za nevychýlené. Směrodatné odchylky jsou výrazně nižší než průměry, proto lze všechny koeficienty považovat za významné.

2) Mapa citlivosti funkce

Parametr	Relativní změna C _{JR} (-5%) [%]	Souhrnná citlivost	Relativní změna C _{JR} (+5%) [%]
p ₁	-1,699E-9	1	1,537E-9
p ₂	1,902	0,1498	-1,6775
p ₃	5,727	5,327	-5,295

Změna citlivosti u parametru p₁ je malá, u zbývajících dvou je výrazně vyšších. Parametry modelu nejsou příliš dobře podmíněny.

Model B

1) Odhady parametrů

parametr	Bodový Odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení	Poloviční délka konfidenčního intervalu spočtená z: délky poloos	Poloviční délka konfidenčního intervalu spočtená z: maxim
p ₁	124,32	2,797	0,001015	0,000816	±8,434	±0,0008164
p ₂	8,136	3,471	0,621	7,633	±13,134	±13,98
p ₃	0,1304	0,03609	0,000932	0,7145	±0,1245	±0,1454

Relativní vychýlení u koeficientu p₂ je větší než 1%, rovněž směrodatná odchylka je velká a konfidenční intervaly překrývají 0. Tudiž koeficient p₂ je nevýznamný.

2) Mapa citlivosti funkce

Parametr	Relativní změna C _{JR} (-5%) [%]	Souhrnná citlivost	Relativní změna C _{JR} (+5%) [%]
p ₁	7,37E-9	1	-2,572E-8
p ₂	-13,22	14,487	15,244
p ₃	-21,84	1,14E+10	27,317

Změna citlivosti u parametru p_1 je malá, u zbývajících dvou je výrazně vyšších. Parametry modelu nejsou příliš dobře podmíněny.

Model C

1) Odhady parametrů

parametr	Bodový Odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení	Poloviční délka konfidenčního intervalu spočtená z: délky poloos	Poloviční délka konfidenčního intervalu spočtená z: maxim
p_1	91,46	2,127	1,62E-11	1,77E-11	±7,867	±8,567
p_2	-32,72	6,015	2,6E-10	-7,95E-10	±24,20	±24,23
p_3	-741,49	5,626	-7,36E-19	9,92E-20	±22,67	±22,67

Relativní vychýlení u všech koeficientů je menší než 1%, koeficienty lze považovat za nevychýlené. Směrodatné odchylky jsou výrazně nižší než průměry, proto lze všechny koeficienty považovat za významné.

2) Mapa citlivosti funkce

Parametr	Relativní změna $C_{iR} (-5\%)$ [%]	Souhrnná citlivost	Relativní změna $C_{iR} (+5\%)$ [%]
p_1	-3,27E-09	1	2,96E-09
p_2	3,20E-08	0,125	1,36E-07
p_3	10,80	6,54E-18	-9,26

Změna citlivosti u parametru p_1 a p_2 je malá, u zbývajících je výrazně vyšší. Parametry modelu nejsou příliš dobře podmíněny.

3) Porovnání kvality jednotlivých modelů.

a) U žádného z modelů nebyly indikovány vlivné body.

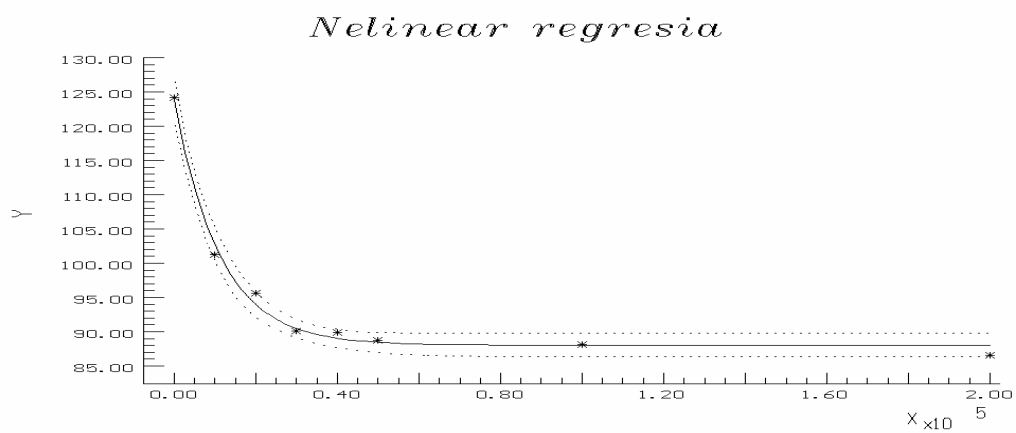
b) Analýza výsledků

Parametr	RSC	Regresní rabat D^2	s(e)	AIC	MEP
Model A	7,874	99,28	1,255	5,873	28,43
Model B	39,14	96,43	2,798	18,7	1,06E+4
Model C	158,28	85,55	5,626	29,88	1,69E+23

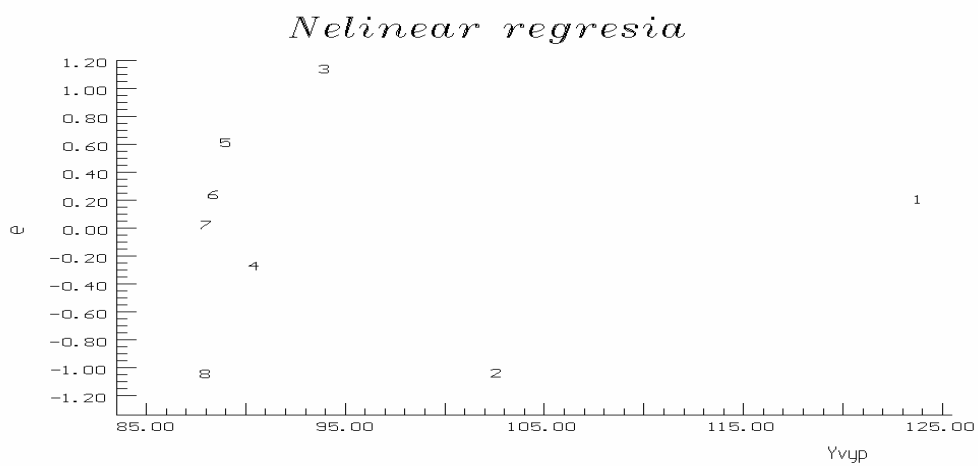
Porovnání výsledků analýzy ukazuje, že nejlepším modelem je model A.

Závěr

Pokles tepelného odporu koberců vlivem sešlapání, lze modelovat pomocí exponenciální funkce (model A). Průběh funkce a graf reziduí vs. predikce jsou na následujících grafech



Obr. 1.: Funkce: model A



Obr. 2.: Graf reziduí

Příklad 2

Měření průběhu pronikání iontu Ca^{++} přes plasmovou membránu. Data byla adaptována z knihy: Rawlings, John O.; Pantula, Sastry G.; Dickey, David A.: Applied Regression Analysis: A Research Tool, Second Edition. Springer.

Data

čas (min)	vápník (nmol/mt)
0,45	0,3875
1,3	1,1248
2,4	1,3999
4	2,7689
6,1	2,9524
8,05	3,4800
11,15	3,6488
13,15	4,6993
15	3,7264

Programy

Adstat 1.25 a 2.0 – moduly nelineární regrese

Řešení

Jako modely byly porovnávány 2 a 3 parametrické Weibullové růstové modely.

Model A:

$$y = \beta_1 \left(1 - e^{-\left(\frac{x}{\beta_2}\right)^{\beta_3}}\right)$$

Model B

$$y = \beta_1 \left(1 - e^{-\left(\frac{x}{\beta_2}\right)}\right)$$

Koeficienty β_i jsou odhadovány pomocí koeficientů p_i . Počáteční odhady jsou

$$p_1=3$$

$$p_2=3$$

$$p_3=1,5$$

Model A

2) Odhady parametrů

parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení	Poloviční délka konfidenčního intervalu spočtená z: délky poloos	Poloviční délka konfidenčního intervalu spočtená z: maxim
p_1	-5,01E+2	1,41E+5	-3,60E+7	7,19E+6	±5,32E+5	±5,32E+5
p_2	4,71E+4	2,23E+7	7,89E+9	1,68E+7	±8,43E+7	±8,43E+7
p_3	5,85E-1	4,58E-01	-6,93	-1,2E+03	±1,68	±1,73

Parametry jsou statisticky nevýznamné, směrodatné odchylky jsou výrazně větší než bodové odhady.

2) Mapa citlivosti funkce

parametr	Relativní změna C_{jR} (-5%) [%]	Souhrnná citlivost	Relativní změna C_{jR} (+5%) [%]
p_1	7,4096E+01	3,5442E-05	-4,2639E+01
p_2	5,7474E+01	1,3759E-09	-3,6868E+01
p_3	5,6285E+01	6,4705E+02	-3,6409E+01

Citlivost je velmi nízká.

Model B

2) Odhady parametrů

parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení	Poloviční délka konfidenčního intervalu spočtená z: délky poloos	Poloviční délka konfidenčního intervalu spočtená z: maxim
P_1	4,309	0,3422	0,0519	1,204	$\pm 0,935$	$\pm 1,204$
P_2	4,8	1,022	0,144	3,003	$\pm 3,143$	$\pm 3,146$

Směrodatná odchylka u obou parametrů je minimálně 3x menší než bodové odhady, což ukazuje že oba parametry jsou statisticky významné. Relativní odhad je větší než 1%, takže odhady jsou méně přesné.

2) Mapa citlivosti funkce

Parametr	Relativní změna C_{jR} (-5%) [%]	Souhrnná citlivost	Relativní změna C_{jR} (+5%) [%]
p_1	3,287	0,480	-3,165
p_2	-2,703	0,5383	2,573

Citlivost u obou parametrů je nízká.

3) Porovnání kvality jednotlivých výběrů

a) U žádného z modelů nebyly indikovány vlivné body.

b) Analýza výsledků

Parametr	RSC	Regresní rabat D^2	s(e)	AIC	MEP
Model A	81,15	-402,63	3,678	25,79	9,018
Model B	0,889	94,49	0,356	-16,83	0,209

Porovnání výsledků analýzy ukazuje, že nejlepším modelem je model B. Analýza výsledků modelu A a především regresní rabat naznačuje, že došlo k selhání během výpočtu.

Závěr

Vhodnější je 2 parametrické Weibullovo rozdělení.