

UNIVERZITA PARDUBICE

Fakulta chemicko - technologická
Katedra analytické chemie

Licenční studium chemometrie:
Počítačové zpracování dat při kontrole a řízení jakosti

SEMESTRÁLNÍ PRÁCE

Předmět: Tvorba nelineárních regresních modelů
v analýze dat
Přednášející: Prof. J. Militký

Příklad: Hledání adekvátního empirického modelu pro daná data (S8.01).

1. Zadání:

Rozhodněte, který z předložených modelů nejlépe odpovídá následujícím datům:

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

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

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

2. Data:

Vstupní data jsou v souboru S801.dat:

x	1	2	3	4	5	6	7	8	9	10	11
y	1.64	1.5	1.4	1.34	1.3	1.26	1.23	1.21	1.19	1.18	1.16

3. Program:

Adstat 1.25: Nelineární regrese.

4. Output z PC:

Soubory: S801a.out, S801b.out, S801b1.out, S801c.out.

5. Obrázky:

V příloze:

- Graf regresní křivky, model A
- Graf reziduí vs. predikce, model A
- Graf reziduí vs. x, model A
- Graf regresní křivky, model B
- Graf reziduí vs. predikce, model B
- Graf reziduí vs. x, model B
- Graf regresní křivky, model B1
- Graf reziduí vs. predikce, model B1
- Graf reziduí vs. x, model B1
- Graf regresní křivky, model C
- Graf reziduí vs. predikce, model C
- Graf reziduí vs. x, model C

6. Řešení:

Model A:

Pro zadaná data byla použita regresní funkce ve tvaru $y = \beta_1 + \beta_2 \cdot e^{-\beta_3 \cdot x}$ s počátečními hodnotami parametrů $\beta_1 = 0$, $\beta_2 = 1$, $\beta_3 = 1$. Odhady parametrů modelu jsou uvedeny v tabulce č.1.

Tabulka 1. Odhady parametrů modelu A

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení[%]
β_1	1.14810	$7.705 \cdot 10^{-3}$	$-3.070 \cdot 10^{-4}$	$-2.674 \cdot 10^{-2}$
β_2	0.65801	$1.246 \cdot 10^{-2}$	$5.811 \cdot 10^{-4}$	$8.831 \cdot 10^{-2}$
β_3	0.30498	$1.507 \cdot 10^{-2}$	$1.539 \cdot 10^{-4}$	$5.046 \cdot 10^{-2}$

Z hodnot jejich směrodatných odchylek a jejich intervalových odhadů vyplývá, že všechny parametry modelu A jsou statisticky významné. Z regresního grafu je vidět poměrně dobré proložení experimentálních bodů navrženou křivkou, ale z grafů reziduí je patrné, že rezidua mají trend ve tvaru ležatého S. Kvalitu regresního modelu potvrzuje i hodnota regresního rabatu D^2 99.78 %. Hodnoty koeficientů korelační matice svědčí o korelaci mezi parametry β_1 a β_3 (0.885). Z mapy citlivostní funkce je zřejmé, že odezva na změnu parametru β_2 v blízkosti minima $U(b)$ je zřetelně asymetrická. Hodnoty RSC, AIC a MEP jsou uvedeny v tabulce č. 5.

Model B:

Druhým modelem byla funkce ve tvaru $y = \beta_1 + \beta_2 \cdot x^{\beta_3}$ s počátečními hodnotami parametrů $\beta_1 = 0$, $\beta_2 = 1$, $\beta_3 = 1$. Odhady parametrů modelu jsou uvedeny v tabulce č. 2.

Tabulka 2. Odhady parametrů modelu B

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení[%]
β_1	-0.09161	0.40826	$-1.142 \cdot 10^{-1}$	$1.246 \cdot 10^{+2}$
β_2	1.73660	0.40475	$1.142 \cdot 10^{-1}$	6.574
β_3	-0.13780	0.03839	$2.367 \cdot 10^{-5}$	$-1.718 \cdot 10^{-2}$

Grafy regresní křivky i reziduí jsou velmi podobné jako u modelu A. Z hodnot směrodatných odchylek parametrů a intervalových odhadů vyplývá, že parametr β_1 není statisticky významný. Také hodnoty koeficientů korelační matice (-0.999) svědčí o silné korelaci mezi všemi parametry regresního modelu, model je přeuren. Proto upravíme regresní model B na tvar $y = \beta_2 \cdot x^{\beta_3}$ (model B₁). Počáteční hodnoty regresních parametrů β_i byly zvoleny rovny 1. Nově vypočtené parametry modelu B₁ jsou statisticky významné a jsou uvedeny v tabulce č.3.

Tabulka 3. Odhady parametrů modelu B₁

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení[%]
β_2	1.64590	$4.956 \cdot 10^{-3}$	$2.183 \cdot 10^{-6}$	$1.326 \cdot 10^{-4}$
β_3	-0.14698	$1.854 \cdot 10^{-3}$	$8.108 \cdot 10^{-7}$	$-5.516 \cdot 10^{-4}$

Hodnoty RSC, AIC a MEP vypočtené pro oba modely (B i B1) jsou uvedeny v tabulce č. 5. U modelu B1 sice došlo k nepatrnému zhoršení RSC a D^2 , ale AIC a MEP se vzhledem k modelu B zlepšily. Odezva na změnu parametrů v blízkosti minima $U(b)$ je poměrně symetrická.

Model C:

Rovnice třetího modelu má tvar $y = \beta_1 + \frac{\beta_2}{1 + \beta_3 \cdot x}$, počáteční hodnoty parametrů

byly zvoleny $\beta_1 = 0$, $\beta_2 = 1$, $\beta_3 = 1$. Odhady parametrů modelu jsou statisticky významné a jsou uvedeny v tabulce č. 4.

Tabulka 4. Odhady parametrů modelu C

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení[%]
β_1	0.99752	$6.818 \cdot 10^{-3}$	$-1.179 \cdot 10^{-4}$	$-1.182 \cdot 10^{-2}$
β_2	0.90600	$8.773 \cdot 10^{-3}$	$3.954 \cdot 10^{-4}$	$4.364 \cdot 10^{-2}$
β_3	0.40846	$1.887 \cdot 10^{-2}$	$4.726 \cdot 10^{-4}$	$1.157 \cdot 10^{-1}$

Regresní křivka velmi dobře vyhovuje pro daná data, navíc je patrné zúžení konfidenčního pásu vzhledem k oběma předchozím modelům. Rezidua už nemají trend, jejich průběh má již charakter šumu. Hodnoty koeficientů korelační matice (0.93) svědčí o významné korelaci mezi parametry β_1 a β_3 . Odezva na změnu parametrů v blízkosti minima $U(b)$ je pro β_1 a β_2 nesymetrická. Hodnoty RSC, AIC a MEP jsou uvedeny v tabulce č. 5.

Tabulka 5. Porovnání kvality regresních modelů

Model:	A	B	B1	C
Reziduální součet čtverců, RSC	$4.967 \cdot 10^{-4}$	$3.497 \cdot 10^{-4}$	$3.523 \cdot 10^{-4}$	$9.234 \cdot 10^{-5}$
Regresní rabat, D^2 [%]	$9.978 \cdot 10^1$	$9.985 \cdot 10^1$	$9.985 \cdot 10^1$	$9.996 \cdot 10^1$
Akaikeho informační kritérium, AIC	$-1.041 \cdot 10^2$	$-1.079 \cdot 10^2$	$-1.098 \cdot 10^2$	$-1.226 \cdot 10^2$
MEP	$1.884 \cdot 10^{-4}$	$3.363 \cdot 10^{-4}$	$6.641 \cdot 10^{-5}$	$2.055 \cdot 10^{-5}$
Průměr abs. hodnot reziduí, MA	$5.670 \cdot 10^{-3}$	$4.436 \cdot 10^{-3}$	$4.381 \cdot 10^{-3}$	$2.363 \cdot 10^{-3}$
Průměr relativních hodnot reziduí, MR	$4.301 \cdot 10^{-1}$	$3.345 \cdot 10^{-1}$	$3.284 \cdot 10^{-1}$	$1.814 \cdot 10^{-1}$
Odhad reziduálního rozptylu, $s^2(e)$	$6.209 \cdot 10^{-5}$	$4.371 \cdot 10^{-5}$	$3.914 \cdot 10^{-5}$	$1.154 \cdot 10^{-5}$
Odhad rez. směrodatné odchylky, $s(e)$	$7.880 \cdot 10^{-3}$	$6.611 \cdot 10^{-3}$	$6.256 \cdot 10^{-3}$	$3.397 \cdot 10^{-3}$
Odhad šikmosti reziduí, $g_1(e)$	$-5.258 \cdot 10^{-1}$	1.076	1.103	$3.946 \cdot 10^{-1}$
Odhad špičatosti reziduí, $g_2(e)$	2.000	3.257	3.488	2.028

7. Závěr:

Porovnáním hodnot RSC, D² AIC a MEP uvedených v tabulce č. 5. zjistíme, že pro daná data nejlépe vyhovuje model C vyjádřený rovnicí:

$$y = 0.99752 + \frac{0.906}{1 + 0.40846 \cdot x}.$$

Název souboru: NLREG
Adresář: E:\Pom\nELIN
Šablona: C:\OFFICE\WORD20\SKOL_UK.DOT
Název: Semestrální práce
Předmět: Nelineární regrese
Autor: Ing. Miloslav Korhoň
Klíčová slova: nelineární regrese
Komentáře: hotovo
Datum vytvoření: 14.09.96 12:16
Číslo revize: 34
Poslední uložení: 10.06.97 20:51
Uložil: Ing. Miloslav Korhoň
Celková doba úprav: 843 min.
Poslední tisk: 15.09.00 08:36
Jako poslední úplný tisk
 Počet stránek: 5
 Počet slov: 971 (přibližně)
 Počet znaků: 5 537 (přibližně)