

8

NELINEÁRNÍ REGRESNÍ MODEL Y

Vzorová úloha 8.1 *Odhad tří parametrů rozšířeného Debyeova-Hückelova vztahu*

Na vzorové **Úloze C8.08** ukážeme postup analýzy nelineárního regresního modelu se zadáním: Stanovte termodynamickou disociační konstantu pK_a^T (parametr β_1), efektivní průměr iontů \tilde{a} (parametr β_2) a vysolovací konstantu C (parametr β_3) závislosti smíšené disociační konstanty y na iontové síle x podle rozšířeného Debyeova-Hückelova vztahu⁶⁵ pro vybrané sulfonftaleiny. Mají-li oba ionty L^{Z-1} a HL^Z zhruba stejnou velikost \tilde{a} [10^{10} m] a je-li celkový vysolovací koeficient $C = C_{HL}^Z - C_L^{Z-1}$, lze formulovat Debyeův-Hückelův vztah tvarem

$$y = \beta_1 + \frac{(1 + 2Z) A \sqrt{x}}{(1 + B \beta_2 \sqrt{x})} \beta_3 x^Z,$$

kde $A = 0.5112 \text{ mol}^{-1/2} \cdot \text{l}^{1/2} \cdot \text{K}^{3/2}$, $B = 0.3291 \text{ mol}^{-1/2} \cdot \text{m}^{-1} \cdot \text{l}^{1/2} \cdot \text{K}^{1/2}$ jsou hodnoty konstant pro vodné roztoky a 25 °C. Předpokládejte aditivní model měření a normalitu chyb závisle proměnné y , zatímco nezávisle proměnná x je zatížena podstatně menší experimentální chybou.

Data: Bromkrezolová zeleň: $Z = -1$, $\{x, y\}$.

0.010	4.901	0.022	4.871	0.040	4.834	0.060	4.808	0.116	4.765	0.232	4.709
0.392	4.691	0.594	4.677	0.923	4.664	1.330	4.662	2.050	4.686	3.720	4.785

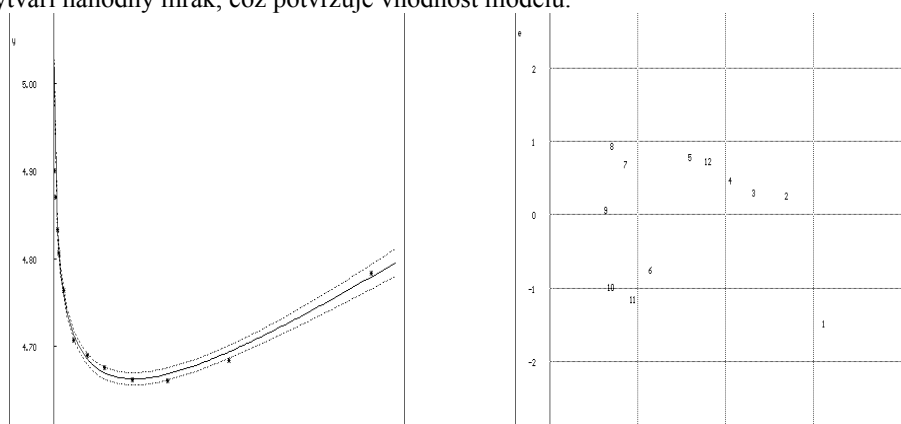
1. Návrh modelu: Označme parametr pK_a^T jako první parametr β_1 s jeho odhadem b_1 , \tilde{a} jako druhý parametr β_2 s jeho odhadem b_2 a konečně C třetí parametr β_3 se odhadem b_3 . Pro nulté přiblížení odhadovaných parametrů je voleno $b_1^{(0)} = 1.0$, $b_2^{(0)} = 1.0$, $b_3^{(0)} = 1.0$.

2. Odhadování parametrů: Jsou nalezeny bodové a intervalové odhady b_j s pol šířkami 95%ních intervalů spolehlivosti, vychýleními h_j a relativními vychýleními $h_{R,j}$.

Bodové odhady parametrů:				
Parametr	Bodový	Směrodatná	Absolutní	Relativní
	odhad b_j	odchylka $s(b_j)$	vychýlení h_j	vychýlení $h_{R,j}$ [%]
β_1	5.0336E+00	4.2468E-03	-1.9583E-05	-3.8905E-04
β_2	7.6010E+00	2.1432E-01	4.2364E-03	5.5735E-02

β_3	6.8337E-02	3.4380E-03	1.7995E-05	2.6332E-02
Intervalové odhady parametrů:				
Parametr	Bodový	Poloviční délka intervalu spolehlivosti spočtená z		
β_j	odhad b_j	délky poloos	maxim	
β_1	5.0336E+00	+1.1915E-02	+1.4456E-02	
β_2	7.6010E+00	+7.2957E-01	+7.2957E-01	
β_3	6.8337E-02	+9.9546E-03	+1.1703E-02	

3. Graf regresní křivky: Danými experimentálními body a 95 % pásy spolehlivosti je zde zobrazen na obr. 8-1a. Graf analýzy klasických reziduí na obr. 8-1b ukazuje, že rezidua vytváří náhodný mrak, což potvrzuje vhodnost modelu.



Obr. 8-1 Vlevo: rozptylový graf, vpravo: analýza klasických reziduí, *ADSTAT*.

4. Korelační matice parametrů: Ukazuje, že korelace mezi parametry je výrazná.

	b_1	b_2	b_3
$x[1, i]$	1.00000	-0.82415	0.52400
$x[2, i]$	-0.82415	1.00000	-0.85058
$x[3, i]$	0.52400	-0.85060	1.00000

5. Základní statistické charakteristiky: Jsou to hodnoty regresního rabatu 100 D [%], Akaikeho informačního kritéria AIC a střední kvadratické chyby predikce MEP k rozlišování mezi modely, obecné statistické momenty reziduí m_1 , m_2 , m_3 , a m_4 , odhady vnitřní křivosti Γ^T a vnější křivosti Γ^N (vysvětlení na str. 617 v cit.¹⁰⁰). Bylo dosaženo výtečného proložení, regresní rabat 99.49% ukazuje, že vysoké procento bodů vyhovuje navrženému modelu Debyeovy-Hückelovy závislosti.

Regresní rabat D [%]	: 99.487
Akaikeho informační kritérium AIC	: -117.59
Střední kvadratická chyba predikce MEP	: 1.5516E-04
První statistický moment residuí m_1	: -1.9709E-07
Třetí statistický moment residuí m_3	: 1.1306E-07
Čtvrtý statistický moment residuí m_4	: 2.0362E-09
Parametr Γ^T	: 8.4501E-02
Parametr Γ^N	: 8.6108E+02

6. Regresní diagnostika: Obsahuje pomůcky pro kritiku dat, kritiku modelu a kritiku metody.

A. Analýza klasických reziduí: *Směrodatná odchylka reziduí* dosahuje hodnoty stejné velikosti, jako je odhad náhodných chyb (šumu) proměnné y , tj. $\sqrt{p}K_{\alpha,i} \cdot 0.01$. Rozdělení reziduí je mírně asymetrické, sešikmené k nižším hodnotám, protože *odhad šikmosti* dosahuje záporné hodnoty. Rozdělení se blíží rovnoměrnému, protože *odhad špičatosti* je blízký hodnotě 1.80. S ohledem na malý počet dat nelze z výběrové šikmosti a špičatosti usuzovat na nenormalitu.

Bod	Měřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Vychýlení	Klasické reziduum
i	y_i	$y_{vyp,i}$	$s(y_{vyp,i})$	h_i	e_i
1	4.9010	4.9115	3.7122E-03	-9.6193E-06	-1.0524E-02
2	4.8710	4.8691	3.3094E-03	-3.8063E-06	1.9192E-03
3	4.8340	4.8318	2.9072E-03	1.0885E-06	2.2400E-03
4	4.8080	4.8046	2.6222E-03	4.1321E-06	3.3797E-03
5	4.7650	4.7593	2.2938E-03	7.4433E-06	5.6887E-03
6	4.7090	4.7142	2.3949E-03	7.2323E-06	-5.2232E-03
7	4.6910	4.6860	2.7365E-03	4.1517E-06	5.0140E-03
8	4.6770	4.6703	2.9978E-03	3.8001E-07	6.7445E-03
9	4.6640	4.6635	3.1121E-03	-3.8612E-06	5.2314E-04
10	4.6620	4.6689	3.0511E-03	-6.3855E-06	-6.9445E-03
11	4.6860	4.6941	3.1750E-03	-6.2046E-06	-8.1348E-03
12	4.7850	4.7797	6.1717E-03	5.4489E-06	5.3175E-03
Reziduální součet čtverců <i>RSC</i>				: 4.0410E-04	
Směrodatná odchylka reziduí $s(e)$: 5.8030E-03	
Odhad šikmosti g_1				: -0.579	
Odhad špičatosti g_2				: 1.796	
Hamiltonův <i>R-faktor</i> [%]				: 0.122	

B. Tabulka vlivných bodů: Obsahuje základní charakteristiky k odhalení vlivných bodů: Jackknife rezidua $\hat{e}_{j,i}$, Cookovy vzdálenosti D_i , diagonální prvky H_{ii} projekční matice, normalizovanou vzdálenost *FDA* a věrohodnostní vzdálenosti *LDA*.

Bod	Jackknife reziduum	Cookova vzdálenost	Diagonální prvky	Normalizovaná vzdálenost	Věrohodnostní vzdálenost
i	$\hat{e}_{j,i}$	D_i	H_{ii}	<i>FDA</i>	<i>LDA</i>
1	-2.2875E+00	5.2536E-01	3.0692E-01	1.2144E-01	2.6567E-01
2	3.1245E-01	1.1668E-02	2.4392E-01	3.8138E-03	7.8886E-03
3	3.5252E-01	1.0641E-02	1.8824E-01	4.6230E-03	7.9056E-03
4	5.2560E-01	1.8108E-02	1.5314E-01	9.3211E-03	8.6827E-03
5	8.9337E-01	3.6124E-02	1.1718E-01	2.1943E-02	1.4799E-02
6	-8.1923E-01	3.4006E-02	1.2774E-01	1.9626E-02	1.3011E-02
7	8.0345E-01	4.4839E-02	1.6679E-01	2.1813E-02	1.4450E-02
8	1.1447E+00	1.0565E-01	2.0015E-01	4.4082E-02	3.4816E-02
9	8.3151E-02	7.1250E-04	2.1571E-01	2.7009E-04	7.3859E-03
10	-1.1908E+00	1.1815E-01	2.0734E-01	4.5560E-02	3.7563E-02

11	-1.4634E+00	1.8341E-01	2.2452E-01	6.5850E-02	7.1201E-02
12	2.6179E+00	7.7430E+00	8.4834E-01	4.1027E-03	2.7646E-03

7. Mapa citlivostní funkce: Citlivostní funkce vyjadřuje změnu regresního modelu při změně parametru o $\pm 5\%$. Hodnoty v tabulce ukazují, že parametr b_1 a b_3 jsou dobře podmíněny v modelu, jejich změna způsobí změnu účelové funkce 8 až 9 řádů. Druhý parametr b_2 je ve srovnání s předchozími parametry méně citlivý, hůře podmíněný v modelu, změna je podstatně menší.

Parametr	Relativní změna	Souhrnná citlivost	Relativní změna
j	$C_{jR}(-5\%), [\%]$	$C_j, [\%]$	$C_{jR}(+5\%), [\%]$
1	8.0486E-09	1.0000	-7.2821E-09
2	1.5937E+01	1.0301E-03	-1.3244E+01
3	-3.3799E-07	1.7701	9.5080E-08

8. Predikční schopnost modelu: Pro $n = 12$ bude $M_1 = 1$ až 6, $M_2 = 7$ až 12, $RSC(M_1) = 2.8534E-05$, $RSC(M_2) = 5.3549E-05$, $U(\mathbf{b}) = 40.410E-05$, a proto $K = 4.923$. Jelikož se K neblíží k jedné, je predikční schopnost modelu slabší.

9. Souhlas s požadavky fyzikálního smyslu: První parametr představuje *termodynamickou disociační konstantu* $pK_a^T = 5.034$, která má fyzikální smysl a dále pak třetí parametr *vysolovací konstantu* $C = 0.068$, jež má rovněž fyzikální smysl. Druhý parametr představuje *efektivní průměr iontu* $\bar{a} = 7.6 \times 10^{-10}$ m, což je v soulase s hodnotami Kielandových tabulek.