

**Univerzita Pardubice
Chemicko-technologická fakulta
Katedra analytické chemie**

**12. licenční studium PYTHAGORAS
Statistické zpracování dat**

3.3 Tvorba nelineárních regresních modelů v analýze dat

Semestrální práce
2010

RNDr. Markéta Vaňková, Ph.D.
Endokrinologický ústav
Národní 8, 116 94 Praha 1

Zadání úlohy

Byla změřena kalibrační křivka pro STH (růstový hormon). Kalibrační křivka má sigmoidální průběh. Cílem úlohy je rozhodnout, zda je vhodnější model proložení sigmoidální křivka 3 parametrická nebo 4 parametrická.

Data

Koncentrace (mmol/l)	absorbance
0.1	1.161
1.2	1.208
2.6	1.303
7.1	1.939
13.1	2.943
25	3.32

Program

NCSS 2004 (a Adstat)

Řešení

V programu NCSS 2004 jsem nenašla parametr k rozlišení mezi modely – Akaiikovo informační kritérium (AIC) a k posouzení predikční schopnosti modelu – střední kvadratickou chybu predikce (MEP). Hodnotila jsem proto koeficient determinace a reziduální součet čtverců (RSC), které informují o přiblížení modelu experimentálním datům.

Parametry RSC, regresní rabat, AIC a MEP byly dodatečně hodnoceny i v programu Adstat.

1. SIGMOIDÁLNÍ KŘIVKA – 3 PARAMETRICKÁ

a) Minimization Phase Section

Itn No.	Error Sum Lambda	Lambda	A	B	C
0	11.4589	0.00004	1	1	1
1	0.5685374	0.000016	2.750439	1.974	4.560603
2	0.1095106	0.0000064	3.470339	4.926991	6.867038
3	0.0624544	2.56E-06	3.424296	4.69801	5.37776
4	5.909777E-02	1.024E-06	3.479328	4.96618	5.759183
5	5.906037E-02	4.096E-07	3.471792	4.928205	5.715791
6	0.0590592	1.6384E-07	3.473138	4.935338	5.722836
7	5.905917E-02	6.5536E-08	3.472914	4.934167	5.721624
8	5.905917E-02	2.62144E-08	3.472952	4.934367	5.721828

Convergence criterion met.

Ukazuje proces hledání algoritmu odhadu parametrů.

b) Model Estimation Section

Parameter Name	Parameter Estimate	Asymptotic Standard Error	Lower 95% C.L.	Upper 95% C.L.
A	3.472952	0.1892879	2.870553	4.075351
B	4.934367	0.9357085	1.956525	7.912209
C	5.721828	0.9287593	2.766101	8.677554

Model ABSORBANCE = A/(1+EXP((B-KONCENTRACE)/C))

R-Squared 0.986727

Iterations 8

Estimated Model

$(3.472952)/(1+\text{EXP}(((4.934367)-(KONCENTRACE))/(5.721828)))$

Odhad parametrů, směrodatných odchylek, červeně koeficient determinace (*100 = regresní rabat).

c) Analysis of Variance Table

Source	DF	Sum of Squares	Mean Square
Mean	1	23.49865	23.49865
Model	3	27.88931	9.296435
Model (Adjusted)	2	4.390659	2.195329
Error	3	5.905917E-02	1.968639E-02
Total (Adjusted)	5	4.449718	
Total	6	27.94836	

Červeně je zvýrazněn reziduální součet čtverců.

d) Asymptotic Correlation Matrix of Parameters

	A	B	C
A	1.000000	0.854933	0.710218
B	0.854933	1.000000	0.678250
C	0.710218	0.678250	1.000000

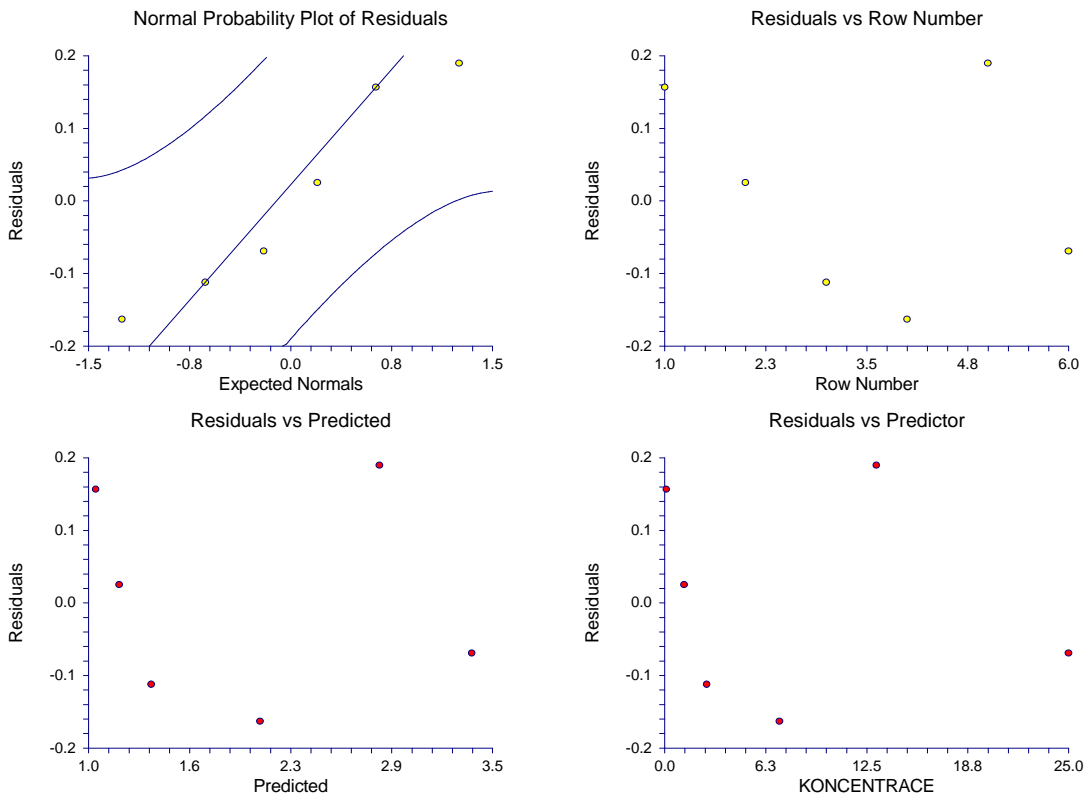
Korelační matice odhadů. Čím vyšší hodnoty, tím jsou odhady parametrů přesnější.

e) Predicted Values and Residuals Section

Row No.	Actual ABSORBANCE	Predicted ABSORBANCE	Lower ABSORBANCE	Upper ABSORBANCE	95.0% C.L.	95.0% C.L.
		Residual				
1	1.161	1.043636	0.5165802	1.570692	0.117364	
2	1.208	1.189111	0.6743443	1.703878	1.888896E-02	
3	1.303	1.387089	0.8827425	1.891435	-8.408883E-02	
4	1.939	2.061225	1.521633	2.600817	-0.1222248	
5	2.943	2.80076	2.237557	3.363964	0.1422397	
6	3.32	3.371828	2.748168	3.995488	-5.182773E-02	

Analýza reziduí a jejich predikovaných hodnot.

f) Plot Section



Grafické diagnostiky (v NCSS 2004) nejsou při malém počtu bodů doporučovány!!

g) Adstat

Reziduální součet čtverců RSC	0.059
Regresní rabat, D^2 (%)	98.673
Akaikeho informační kritérium AIC	-21.726
Střední kvadratická chyba predikce MEP	0.218

2. SIGMOIDÁLNÍ KŘIVKA – 4 PARAMETRICKÁ

a) Minimization Phase Section

Itn No.	Error Sum Lambda	Lambda	A	B	C	D
0	3.04822	0.00004	1	1	1	1
1	0.7797502	0.000016	4.615087	-2.267298	6.397882	-1.844245
2	0.3279935	0.000064	3.987476	2.081178	8.354099	-0.4053181
3	8.567064E-02	0.000256	3.890165	3.713575	6.013699	-0.414666
4	6.413639E-02	0.001024	3.595132	4.491834	6.001018	-0.1096292
5	4.130665E-02	0.0004096	2.988294	6.338586	4.971778	0.4369226
6	2.561019E-02	0.0016384	2.777275	6.960103	4.334159	0.6143579
7	5.118966E-03	6.5536E-04	2.380082	8.148273	3.356177	0.9636005
8	1.395417E-04	2.62144E-04	2.295516	8.375684	2.925484	1.03073
9	1.03597E-04	1.048576E-04	2.304099	8.341411	2.943629	1.022822
10	1.035822E-04	4.194304E-05	2.30422	8.340834	2.943745	1.022696
11	1.035822E-04	1.677722E-05	2.304219	8.340832	2.943741	1.022696

Convergence criterion met.

Ukazuje proces hledání algoritmu odhadu parametrů.

b) Model Estimation Section

Parameter Name	Parameter Estimate	Asymptotic Standard Error	Lower 95% C.L.	Upper 95% C.L.
A	2.304219	1.548135E-02	2.237608	2.37083
B	8.340832	5.345625E-02	8.110828	8.570836
C	2.943741	4.916644E-02	2.732195	3.155287
D	1.022696	1.130136E-02	0.9740703	1.071322

Model ABSORBANCE = A/(1+EXP((B-KONCENTRACE)/C))+D

R-Squared 0.999977

Iterations 11

Estimated Model

(2.304219)/(1+EXP(((8.340832)-(KONCENTRACE))/(2.943741)))+(1.022696)

Odhad parametrů, směrodatných odchylek, červeně koeficient determinace (*100 = regresní rabat).

c) Analysis of Variance Table

Source	DF	Sum of Squares	Mean Square
Mean	1	23.49865	23.49865
Model	4	27.94826	6.987065
Model (Adjusted)	3	4.449615	1.483205
Error	2	1.035822E-04	5.179109E-05
Total (Adjusted)	5	4.449718	
Total	6	27.94836	

Červeně je zvýrazněn reziduální součet čtverců.

d) Asymptotic Correlation Matrix of Parameters

	A	B	C	D
A	1.000000	-0.314623	0.859289	-0.884852
B	-0.314623	1.000000	-0.297885	0.628236
C	0.859289	-0.297885	1.000000	-0.830627
D	-0.884852	0.628236	-0.830627	1.000000

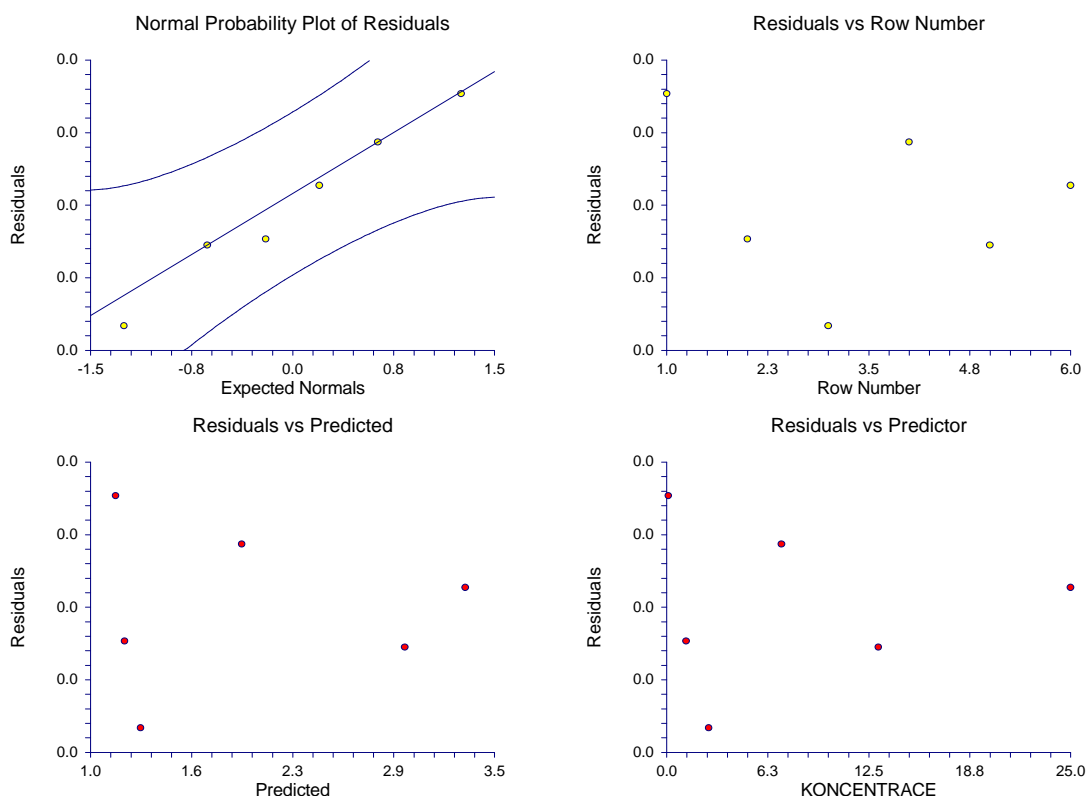
Korelační matice odhadů. Čím vyšší hodnoty, tím jsou odhady parametrů přesnější.

e) Predicted Values and Residuals Section

Row No.	Actual ABSORBANCE	Predicted ABSORBANCE	Lower ABSORBANCE	Upper ABSORBANCE	95.0% C.L.	95.0% C.L.
1	1.161	1.154853	1.116818	1.192888	6.146966E-03	
2	1.208	1.209865	1.173965	1.245765	-1.86493E-03	
3	1.303	1.309648	1.27394	1.345356	-6.648446E-03	
4	1.939	1.935522	1.893101	1.977943	3.477894E-03	
5	2.943	2.945199	2.901938	2.988461	-2.19948E-03	
6	3.32	3.318912	3.27525	3.362574	1.087997E-03	

Analýza reziduí a jejich predikovaných hodnot.

f) Plot Section



Grafické diagnostiky (v NCSS 2004) nejsou při malém počtu bodů doporučovány!!

g) Adstat

Reziduální součet čtverců RSC	0.0001
Regresní rabat, D^2 (%)	99.999
Akaikeho informační kritérium AIC	-57.801
Střední kvadratická chyba predikce MEP	0.0019

Závěr

Porovnáním parametrů AIC, MEP, RSC a koeficientu determinace (resp. regresního rabatu) se ukázalo, že sigmoidní model se 4 parametry je vhodnější než sigmoidní model se 3 parametry. Těsnější proložení je u sigmoidního modelu se 4 parametry, dále hodnoty RSC, AIC i MEP jsou nižší pro sigmoidní model se 4 parametry než pro model se 3 parametry.

Statistický software NCSS 2004 je oblíbený software na našem pracovišti. Úlohu jsem se proto rozhodla vyřešit právě tímto softwarem. Nicméně nelineární regrese, především hodnocení predikční schopnosti modelu a vhodnosti modelu, se začátečníkům v NCSS zpracovává obtížně.