

UNIVERZITA PARDUBICE

Fakulta chemicko-technologická

Katedra analytické chemie

Nám. Čs. Legií 565, 532 10 Pardubice

10. licenční studium chemometrie

STATISTICKÉ ZPRACOVÁNÍ DAT

Semestrální práce

KALIBRACE A LIMITY JEJÍ PŘESNOSTI

2004/2005

Vedoucí studia a odborný garant:
Prof. RNDr. Milan Meloun, DrSc.

Vyučující:
Prof. RNDr. Milan Meloun, DrSc.

Autor práce:
Ing. Zdeňka Dluhošová

OBSAH

ÚLOHA 1.: LINEÁRNÍ KALIBRACE.....	4
Zadání	4
Data	4
Užitý program	4
Řešení	4
1. Lineární regrese.....	4
1.1 Návrh modelu	4
1.2 Zvolená strategie regresní analýzy	5
1.3 Odhad parametrů a testy významnosti.....	5
1.4 Základní statistické charakteristiky	5
1.5 Regresní diagnostika.....	6
1.6 Testování regresního tripletu (data + model + metoda).....	7
2. Kalibrace	8
2.1 Parametry kalibrace	8
2.2 Kalibrační meze.....	8
2.3 Graf kalibrační přímky	9
2.4 Kalibrační tabulka	9
Závěr	9
ÚLOHA 2.: NELINEÁRNÍ KALIBRACE.....	10
Zadání	10
Data	10
Užitý program	10
Řešení	10
1. Lineární regrese.....	10
1.1 Návrh modelu	10
1.2 Zvolená strategie regresní analýzy	11
1.3 Odhad parametrů a testy významnosti.....	11
1.4 Základní statistické charakteristiky	11
1.5 Regresní diagnostika.....	12
1.6 Testování regresního tripletu (data + model + metoda).....	12
2. Kalibrace	13
2.1 Lineární spline	13
2.2 Kvadratický spline	13
2.3 Kubický spline.....	14
2.4 Kalibrační rovnice	14
2.4 Graf kalibrační modelu	14
2.5 Kalibrační tabulka	15
Závěr	15

10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT
Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

ÚLOHA 3.: ROZLIŠENÍ MEZI LINEÁRNÍ A NELINEÁRNÍ KALIBRACÍ	16
Zadání	16
Data	16
Užitý program	16
Řešení	16
1. Lineární regrese	16
1.1 Návrh modelu	16
1.2 Zvolená strategie regresní analýzy	17
1.3 Odhad parametrů a testy významnosti	17
1.4 Základní statistické charakteristiky	17
1.5 Regresní diagnostika	18
1.6 Testování regresního tripletu (data + model + metoda)	21
2. Kalibrace	22
2.1 Kvadratický spline	22
2.2 Kubický spline	23
2.3 Kalibrační rovnice	24
2.4 Graf kalibračního modelu	24
2.5 Kalibrační tabulka	24
Závěr	25
POUŽITÁ LITERATURA	25

Úloha 1.: Lineární kalibrace

Zadání

Pro stanovení 1-hydroxypyrenu v moči osob exponovaných polycyklickými aromatickými uhlovodíky bylo připraveno 12 kalibračních vzorků. Směsná moč neexponovaných osob byla spikována standardním roztokem 1-hydroxypyrenu ve 12 koncentračních úrovních. Kalibrační vzorky a reálné vzorky moče exponovaných pracovníků byly zpracovány stanoveným postupem a metodou HPLC byla stanovena plocha píku 1-hydroxypyrenu. Výsledky stanovení jsou uvedeny v tabulce. Vyčíslete bodový a intervalový odhad pro tři neznámé koncentrace a současně vyčíslete i limity přesnosti.

Data

Koncentrace 1-hydroxypyrenu v kalibračních a reálných vzorcích moče

Číslo vzorku	Koncentrace 1-hydroxypyrenu [$\mu\text{g}\cdot\text{l}^{-1}$]	Plocha píku
1K	0.10	428
2K	0.20	713
3K	0.50	1622
4K	0.75	2176
5K	1.00	2725
6K	2.00	5283
7K	4.00	10024
8K	6.00	14965
9K	8.00	19930
10K	10.00	25110
11K	12.00	29645
12K	14.00	34586
vzorek 1	?	2704
vzorek 2	?	19160
vzorek 3	?	15467

Užitý program

ADSTAT: Lineární regrese – Regresní diagnostika
Kalibrace - Přímka

Řešení

1. Lineární regrese

1.1 Návrh modelu

Použijeme lineární regresní model $y = \beta_0 + \beta_1 x$, kde x = koncentrace 1-hydroxypyrenu a y = plocha píku 1-hydroxypyrenu.

10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT

Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

1.2 Zvolená strategie regresní analýzy

Omezení:	$1 \cdot 10^{-34}$
Transformace:	Ne
Váhy:	Ne
Absolutní člen zahrnut:	Ano

Podmínky a kvantily pro statistické testy:

Hladina významnosti, α :	0.050
Počet bodů, n:	12
Počet parametrů, m:	2
Kvantil Studentova rozdělení $t(1-\alpha/2, n-m)$:	2.228
Kvantil rozdělení $\chi^2(1-\alpha, m)$:	5.991

1.3 Odhad parametrů a testy významnosti

Klasickou metodou nejmenších čtverců byly určeny odhady parametrů β_0 a β_1 . Studentův t-test ukázal, že úsek β_0 a směrnice β_1 jsou statisticky významné.

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$		
			t-kritérium	Hypotéza H_0 je	Hlad. význ.
B[0]	2.8886E+02	4.8052E+01	6.0114E+00	Zamítnuta	0.000
B[1]	2.4550E+03	7.0224E+00	3.4960E+02	Zamítnuta	0.000

1.4 Základní statistické charakteristiky

Vysoká hodnota vícenásobného korelačního koeficientu ukazuje, že navržený lineární regresní model je statisticky významný.

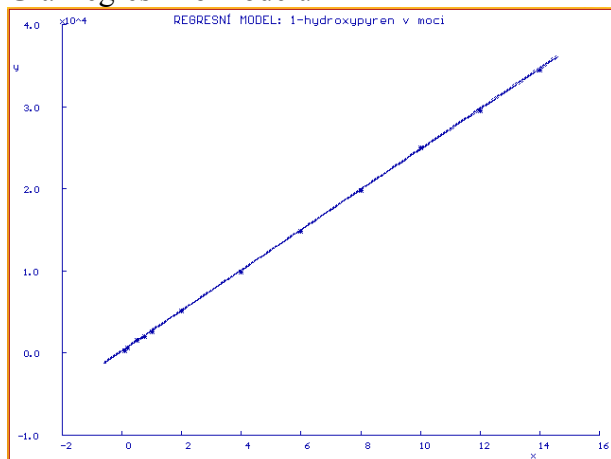
Koeficient determinace ukazuje, model vystihuje data v 99.99 %. Střední kvadratická chyba predikce MEP a Akaikeho informační kritérium AIC slouží pro optimalizaci modelu – za optimální se považuje model, pro který dosahuje MEP a AIC minimální hodnotu.

Vícenásobný korelační koeficient, R	9.9996E-01
Koeficient determinace, R^2	9.9992E-01
Predikovaný korelační koeficient, R_p^2	9.9994E-01
Střední kvadratická chyba predikce, MEP	1.7138E+04
Akaikeho informační kritérium, AIC	1.1604E+02

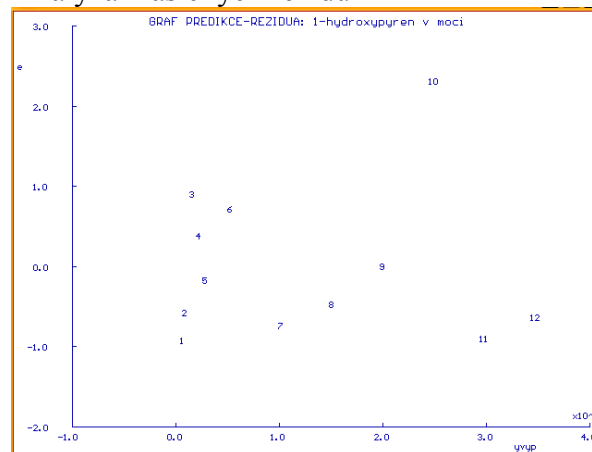
1.5 Regresní diagnostika

1.5.1 Klasická rezidua

Graf regresního modelu



Analýza klasických reziduí



Reziduální součet čtverců, RSC	1.3620E+05
Průměr absolutních hodnot reziduí, Me	8.4632E+01
Průměr relativních hodnot reziduí, Mer	3.9990E+00
Odhad reziduálního rozptylu, $s^2(e)$	1.3620E+04
Odhad směrodatné odchylky reziduí, $s(e)$	1.1670E+02
Odhad šikmosti reziduí, $g1(e)$	1.2404E+00
Odhad špičatosti reziduí, $g2(e)$	3.8369E+00

Grafická analýza klasických reziduí vzhledem k predikci ukazuje na nekonstantnost rozptylu. Odhad šikmosti a špičatosti indikují přibližně Gaussovo normální rozdělení.

1.5.2 Analýza ostatních reziduí

Odlehlé body: Jackknife rezidua – bod 10
 Extrémy: Diagonální prvky – bod 12
 Podezřelé body: Cookova vzdálenost – bod 10, 11, 12
 Atkinsonova vzdálenost – bod 10
 Vliv na predikci – bod 10
 Věrohodnostní vzdálenost – bod 10

1.5.3 Grafy vlivných bodů

Tyto grafy indikují přítomnost odlehlých hodnot a extrémů:

Graf predikovaných reziduí: vlivný bod – bod 10
 Pregibonův graf: středně vlivný bod – bod 10
 Williamsův graf: odlehlý bod - bod 10
 extrém – bod 12
 McCulloh-Meeterův graf: odlehlý bod – bod 10
 extrém - bod 12
 L-R graf: odlehlý bod – bod 10
 extrém – bod 11,12

1.5.4 Indexové grafy

Upozorňují na podezřelé body a kvantifikují velikost jejich vlivu.

Andrewsův indexový graf:	podezřelý bod – bod 10
Graf normalizovaných reziduí:	podezřelý bod – bod 10
Graf prvků H projekční matice:	extrém – bod 12

1.5.5 Rankitové grafy

Ukazují vedle normality rozdělení reziduí i vlivné body.

Graf normalizovaných reziduí:	odlehlý bod – bod 10
Andrewsův graf:	odlehlý bod – bod 10
Graf predikovaných reziduí:	odlehlý bod – bod 10
Graf Jackknife reziduí:	odlehlý bod – bod 10

Závěr regresní diagnostiky:

Bod 10 je odlehlý, je vhodné ho z výběru vyloučit a znovu určit regresní parametry. Na jeden neznámý parametr regresního modelu potřebujeme 5 bodů, tzn. pro přímku o 2 parametrech (úsek, směrnice) potřebujeme 10 bodů. Vzhledem k tomu, že kalibrace byla provedena z 12 kalibračních bodů, můžeme odlehlý bod č. 10 vyloučit.

1.6 Testování regresního tripletu (data + model + metoda)

Patří zde vyšetření splnění základních předpokladů metody nejmenších čtverců MNČ čili statistické testy předpokladů, za kterých by měla metoda vést k nejlepším nestranným lineárním odhadům regresních parametrů.

Testy předpokladů MNČ

Fischer-Snedocorův test významnosti regrese, F	1.2222E+05
Tabulkový kvantil, F(1-alpha, m-1, n-m)	4.9646E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	0.000
Scottovo kritérium multikolinearity, M	1.2945E-13
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, Sf	2.5227E+01
Tabulkový kvantil, Chi ² (1-alpha, 1)	3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti	0.000
Jarque-Berraův test normality reziduí, L(e)	3.4276E+00
Tabulkový kvantil, Chi ² (1-alpha, 2)	5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	0.180
Waldův test autokorelace, Wa	3.0125E-01
Tabulkový kvantil, Chi ² (1-alpha, 1)	3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	0.583

10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT
Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

Znaménkový test, Dt -2.0821E-01
 Tabulkový kvantil, N(1-alpha/2) 1.6449E+00
 Závěr: Rezidua nevykazují trend.
 Spočtená hladina významnosti 0.418

Závěr testů předpokladů MNČ

Předpoklady MNČ jsou splněny, kromě požadavku konstantního rozptylu reziduí.
 Heteroskedasticita je zřejmě způsobena vlivným bodem 10.

2. Kalibrace

2.1 Parametry kalibrace

Na základě závěru regresní diagnostiky vyloučíme z kalibrace bod 10.

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$		
			t-kritérium	Hypotéza H_0 je	Hlad. význ.
B [0]	2.9121E+02	2.9692E+01	9.8082E+00	Zamítnuta	0.000
B [1]	2.4489E+03	4.5820E+00	5.3445E+02	Zamítnuta	0.000

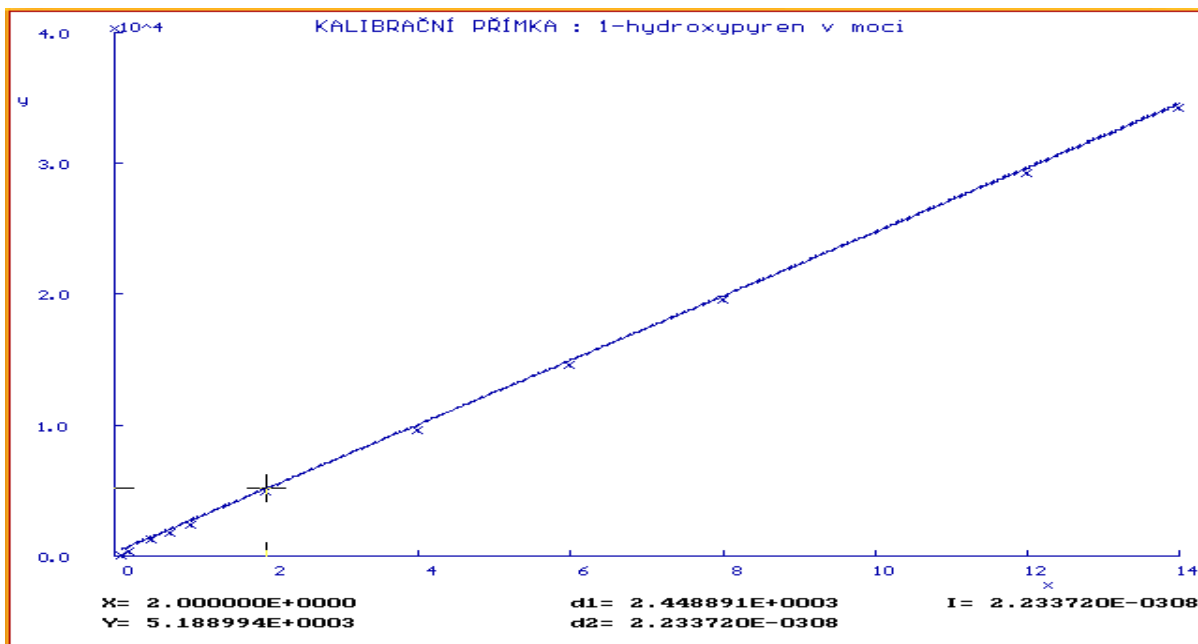
Reziduální součet čtverců, RSC	4.6782E+04
Průměr absolutních hodnot reziduí, Me	5.5716E+01
Průměr relativních hodnot reziduí, Mer	4.2760E+00
Odhad reziduálního rozptylu, $s^2(e)$	5.1980E+03
Odhad směrodatné odchylky reziduí, s(e)	7.2097E+01

Odhad směrodatné odchylky reziduí s(e), který indikuje těsnost proložení bodů přímkou, se snížil přibližně o 40 %. Došlo tedy k vylepšení lineárního regresního modelu.

2.2 Kalibrační meze

	Plocha píku	Koncentrace 1-hydroxypyrenu [$\mu\text{g.l}^{-1}$]
Kritická úroveň	y_c : 358	x_c : 0.027
Limita detekce	y_d : -----	x_d : -----
Mez stanovitelnosti	y_s : 780	x_s : 0.199

2.3 Graf kalibrační přímky



2.4 Kalibrační tabulka

Měřená hodnota - plocha píku	Přímý odhad koncentrace 1-hydroxypyrenu [$\mu\text{g}\cdot\text{l}^{-1}$]	Konfidenční interval	
		dolní mez	horní mez
2704	0.99	0.91	1.06
19160	7.71	7.63	7.78
15467	6.20	6.13	6.27

Závěr

Stanovení koncentrace 1-hydroxypyrenu v moči osob exponovaných PAU má pro zamýšlené použití dostatečnou citlivost, vypočtená mez stanovitelnosti odpovídá požadavku na zkušební metodu. Metoda je použitelná pro stanovení 1-hydroxypyrenu v reálných vzorcích moče v rozsahu koncentrací 0, 20 – 14.00 $\mu\text{g}\cdot\text{l}^{-1}$. Dále byly vypočteny bodové odhady pro tři neznámé koncentrace 1-hydroxypyrenu v moči exponovaných osob a příslušné intervalové odhady, u kterých můžeme s 95% jistotou tvrdit, že neznámá hodnota koncentrace 1-hydroxypyrenu leží právě v těchto intervalech. Výsledky stanovení koncentrací 1-hydroxypyrenu v moči exponovaných osob jsou uvedeny v kalibrační tabulce.

Úloha 2.: Nelineární kalibrace

Zadání

Pro stanovení o, p- DDE v pitných vodách bylo připraveno 12 kalibračních vzorků. Destilovaná demineralizovaná voda byla spikována standardním roztokem o, p - DDE ve 12 koncentračních úrovních. Kalibrační vzorky a reálné vzorky pitné vody byly zpracovány stanoveným postupem a metodou GC – ECD byly stanoveny výšky píku o, p - DDE. Výsledky stanovení jsou uvedeny v tabulce. Vyčíslete bodový a intervalový odhad pro tři neznámé koncentrace a současně vyčíslete i limity přesnosti.

Data

Koncentrace o, p - DDE v kalibračních a reálných vzorcích vody

Číslo vzorku	Koncentrace o, p - DDE [$\mu\text{g.l}^{-1}$]	Výška píku
1K	0.4	5.93
2K	1.0	10.65
3K	2.0	20.92
4K	4.0	42.30
5K	6.0	61.72
6K	8.0	79.68
7K	10.0	97.26
8K	15.0	136.64
9K	20.0	162.16
10K	30.0	214.82
11K	40.0	265.50
12K	50.0	308.64
vzorek 1	?	54.37
vzorek 2	?	173.20
vzorek 3	?	84.56

Užitý program

ADSTAT: Lineární regrese – Regresní diagnostika

Kalibrace – Lineární spline, kvadratický spline, kubický spline

Řešení

1. Lineární regrese

1.1 Návrh modelu

Použijeme lineární regresní model $y = \beta_0 + \beta_1 x$, kde x = koncentrace o, p - DDE a y = výška píku o, p – DDE.

10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT

Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

1.2 Zvolená strategie regresní analýzy

Omezení:	$1 \cdot 10^{-34}$
Transformace:	Ne
Váhy:	Ne
Absolutní člen zahrnut:	Ano

Podmínky a kvantily pro statistické testy:

Hladina významnosti, α :	0.050
Počet bodů, n:	12
Počet parametrů, m:	2
Kvantil Studentova rozdělení $t(1-\alpha/2, n-m)$:	2.228
Kvantil rozdělení $\chi^2(1-\alpha, m)$:	5.991

1.3 Odhad parametrů a testy významnosti

Klasickou metodou nejmenších čtverců byly určeny odhady parametrů β_0 a β_1 . Studentův t-test ukázal, že úsek β_0 a směrnice β_1 jsou statisticky významné.

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$		
			t-kritérium	Hypotéza H_0 je	Hlad. význ.
B[0]	2.0937E+01	6.4545E+00	3.2437E+00	Zamítnuta	0.009
B[1]	6.2789E+00	2.9243E-01	2.1471E+01	Zamítnuta	0.000

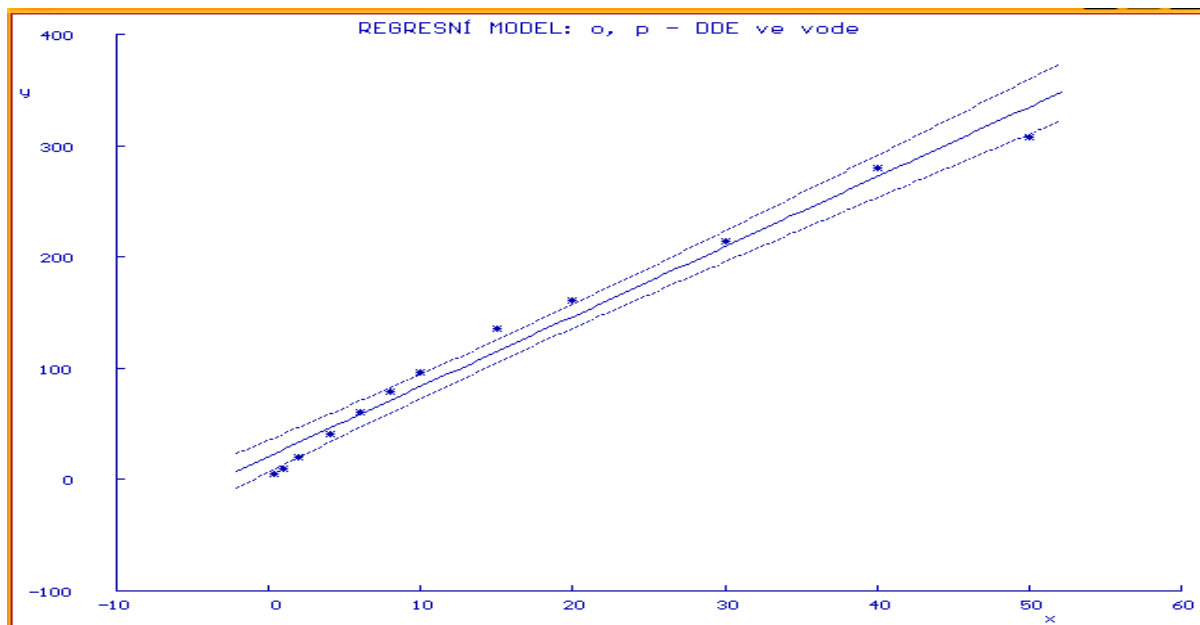
1.4 Základní statistické charakteristiky

Koeficient determinace ukazuje, model vystihuje data v 97.88 %. Střední kvadratická chyba predikce MEP a Akaikeho informační kritérium AIC slouží pro optimalizaci modelu – za optimální se považuje model, pro který dosahuje MEP a AIC minimální hodnotu.

Vícenásobný korelační koeficient, R	9.8933E-01
Koeficient determinace, R^2	9.7877E-01
Predikovaný korelační koeficient, R_p^2	9.7862E-01
Střední kvadratická chyba predikce, MEP	4.1897E+02
Akaikeho informační kritérium, AIC	6.8181E+01

1.5 Regresní diagnostika

Graf regresního modelu



Reziduální součet čtverců, RSC	2.5233E+03
Průměr absolutních hodnot reziduí, Me	1.2775E+01
Průměr relativních hodnot reziduí, Mer	4.9099E+01
Odhad reziduálního rozptylu, $s^2(e)$	2.5233E+02
Odhad směrodatné odchylky reziduí, $s(e)$	1.5885E+01
Odhad šikmosti reziduí, $g1(e)$	-3.3493E-01
Odhad špičatosti reziduí, $g2(e)$	1.8640E+00

Závěr regresní diagnostiky:

Z grafu regresního modelu je zřejmé, že proložení bodů kalibrace přímkou není optimální.

1.6 Testování regresního tripletu (data + model + metoda)

Testy předpokladů MNČ

Fischer-Snedecorův test významnosti regrese, F 4.6102E+02

Tabulkový kvantil, $F(1-\alpha, m-1, n-m)$ 4.9646E+00

Závěr: Navržený model je přijat jako významný.

Spočtená hladina významnosti 0.000

Scottovo kritérium multikolinearity, M -3.3307E-16

Závěr: Navržený model je korektní.

Cook-Weisbergův test heteroskedasticity, S_f 2.5046E+01

Tabulkový kvantil, $\chi^2(1-\alpha, 1)$ 3.8415E+00

Závěr: Rezidua vykazují heteroskedasticitu.

10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT

Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

Spočtená hladina významnosti	0.000
Jarque-Berraův test normality reziduí, L(e)	8.6963E-01
Tabulkový kvantil, $\chi^2(1-\alpha, 2)$	5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	0.647
Waldův test autokorelace, Wa	9.1675E+00
Tabulkový kvantil, $\chi^2(1-\alpha, 1)$	3.8415E+00
Závěr: Rezidua jsou autokorelována.	
Spočtená hladina významnosti	0.002
Znaménkový test, Dt	-2.0821E-01
Tabulkový kvantil, N(1-alpha/2)	1.6449E+00
Závěr: Rezidua vykazují trend.	
Spočtená hladina významnosti	0.019

Závěr testů předpokladů MNČ

Ze závěrů testování vyplývá, že rezidua regresního modelu vykazují autokorelaci a trend, Regresní model není vhodný pro lineární kalibraci.

2. Kalibrace

Další možností je použití spline pro danou kalibrační závislost. Postupně jsou vyhodnoceny lineární, kvadratický a kubický spline, tj. regrese s různými alternativami zadávání uzlových bodů. Jako výsledek hodnocení je srovnávána limita detekce koncentrace x_D a odhad směrodatné odchylky reziduí σ_e . Za nejlepší kalibrační model se považuje takový, který má nejnižší limitu detekce koncentrace x_D a odhad směrodatné odchylky reziduí σ_e při nejnižším počtu uzlových bodů.

2.1 Lineární spline

<i>Parametry</i>	<i>lineární spline</i>			
uzlové body	1	2	3	4
limit detekce x_D [$\mu\text{g.l}^{-1}$]	2.248	1.752	0.547	0.680
odhad směr. odchylky reziduí σ_e	8.671	6.596	1.966	2.297

2.2 Kvadratický spline

<i>Parametry</i>	<i>kvadratický spline</i>			
uzlové body	1	2	3	4
limit detekce x_D [$\mu\text{g.l}^{-1}$]	1.406	0.719	0.515	0.283
odhad směr. odchylky reziduí σ_e	5.539	2.834	1.713	0.762

10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT

Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

2.3 Kubický spline

<i>Parametry</i>	<i>kubický spline</i>			
uzlové body	1	2	3	4
limit detekce x_D [$\mu\text{g.l}^{-1}$]	0.843	0.458	0.434	0.390
odhad směr. odchylky reziduí σ_e	3.434	1.386	1.000	0.769

Z tabulek plyne použití kvadratického spline se čtyřmi uzlovými body jako nejlepšího kalibračního modelu.

2.4 Kalibrační rovnice

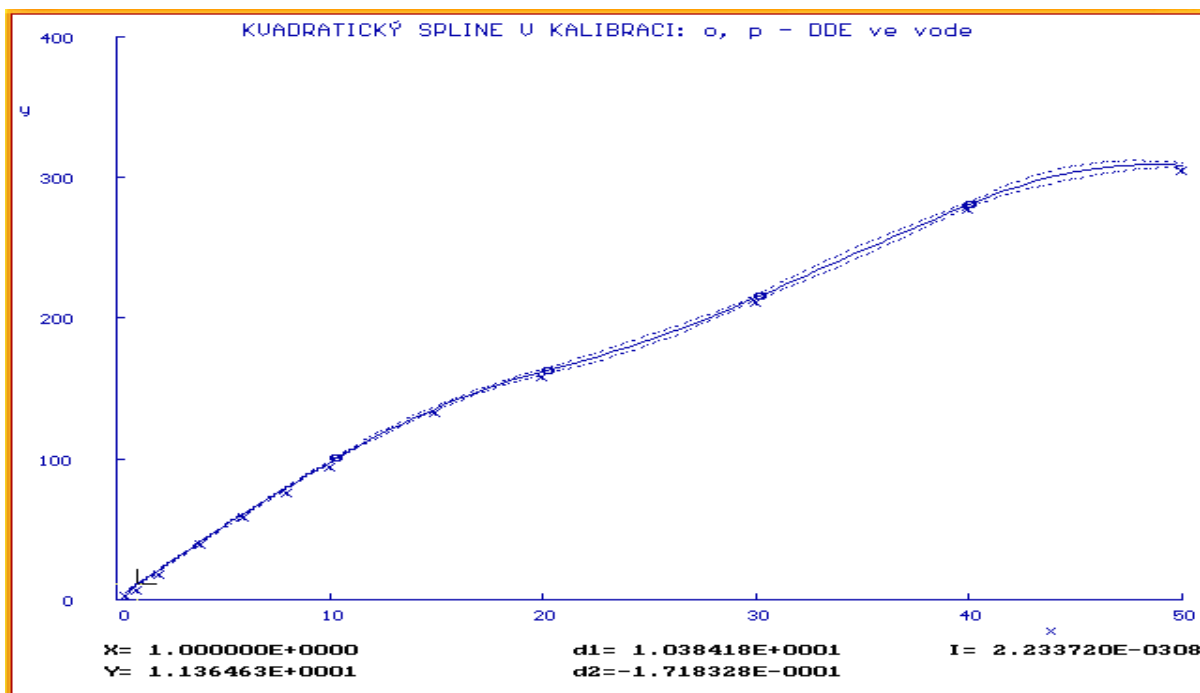
Kalibrační rovnice má tvar

$$Y = a \cdot x^2 + b \cdot x + c \quad \text{pro } k[i-1] < x \leq k[i]$$

Uzlové body a koeficienty kalibrační rovnice:

uzlový bod $k[i]$	a	b	c
10.32	-0.086	10.556	0.895
20.24	-0.250	13.950	-16.620
30.16	0.151	-2.309	147.920
40.08	-0.022	8.117	-9.306

2.4 Graf kalibrační modelu



10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT
Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

2.5 Kalibrační tabulka

Měřená hodnota - výšky píku o, p - DDE	Inverzní odhad koncentrace o, p - DDE [$\mu\text{g.l}^{-1}$]	Konfidenční interval koncentrace o, p - DDE [$\mu\text{g.l}^{-1}$]	
		dolní mez	horní mez
54.37	5.29	5.18	5.41
173.20	22.64	21.97	23.37
84.56	8.52	8.40	8.63

Závěr

Ze statistické analýzy různých kalibračních modelů se jeví jako nejlepší model kvadratický spline se čtyřmi uzlovými body. U tohoto modelu je limita detekce $x_D = 0.28 \mu\text{g.l}^{-1}$, což je dostatečně nízká hodnota k tomu, aby bylo možné tuto metodu použít pro stanovení stopových koncentrací o, p - DDE v pitných vodách. Dále byly vypočteny bodové odhady pro tři neznámé koncentrace o, p - DDE v pitných vodách a příslušné intervalové odhady, u kterých můžeme s 95% jistotou tvrdit, že neznámá hodnota koncentrace o, p - DDE leží právě v těchto intervalech. Výsledky stanovení koncentrací o, p - DDE v pitných vodách jsou uvedeny v kalibrační tabulce.

Úloha 3.: Rozlišení mezi lineární a nelineární kalibrací

Zadání

Pro stanovení hexachlorcyklobenzenu v potravinách bylo připraveno 12 kalibračních vzorků. Destilovaná demineralizovaná voda byla spikována standardním roztokem HCB ve 12 koncentračních úrovních. U kalibračních vzorků a zpracovaných vzorků rostlinných tuků byly metodou GC – ECD stanoveny výšky píku HCB. Výsledky stanovení jsou uvedeny v tabulce. Rozhodněte o počtu uzlových bodů kalibrační závislosti, typu splinové závislosti, vyčíslete bodový a intervalový odhad pro tři neznámé koncentrace a současně vyčíslete i limity přesnosti.

Data

Koncentrace HCB v kalibračních vzorcích a ve vzorcích rostlinných tuků

Číslo vzorku	Koncentrace o, p - DDE [$\mu\text{g.l}^{-1}$]	Výška píku
1K	0.5	10.69
2K	1.0	27.18
3K	2.0	55.49
4K	3.0	80.56
5K	4.0	119.09
6K	5.0	129.76
7K	6.0	148.84
8K	8.0	194.60
9K	10.0	228.48
10K	15.0	298.30
11K	20.0	354.16
12K	25.0	402.14
vzorek 1	?	
vzorek 2	?	
vzorek 3	?	

Užitý program

ADSTAT: Lineární regrese – Regresní diagnostika

Kalibrace – Přímka, lineární spline, kvadratický spline, kubický spline

Řešení

1. Lineární regrese

1.1 Návrh modelu

Použijeme lineární regresní model $y = \beta_0 + \beta_1 x$, kde x = koncentrace HCB a y = výška píku HCB .

10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT

Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

1.2 Zvolená strategie regresní analýzy

Omezení:	$1 \cdot 10^{-34}$
Transformace:	Ne
Váhy:	Ne
Absolutní člen zahrnut:	Ano

Podmínky a kvantily pro statistické testy:

Hladina významnosti, α :	0.050
Počet bodů, n:	12
Počet parametrů, m:	2
Kvantil Studentova rozdělení $t(1-\alpha/2, n-m)$:	2.228
Kvantil rozdělení $\chi^2(1-\alpha, m)$:	5.991

1.3 Odhad parametrů a testy významnosti

Klasickou metodou nejmenších čtverců byly určeny odhady parametrů β_0 a β_1 . Studentův t-test ukázal, že úsek β_0 a směrnice β_1 jsou statisticky významné.

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$		
			t-kritérium	Hypotéza H_0 je	Hlad. význ.
B[0]	3.8345E+01	1.0792E+01	3.5531E+00	Zamítnuta	0.005
B[1]	1.5971E+01	9.6358E-01	1.6575E+01	Zamítnuta	0.000

1.4 Základní statistické charakteristiky

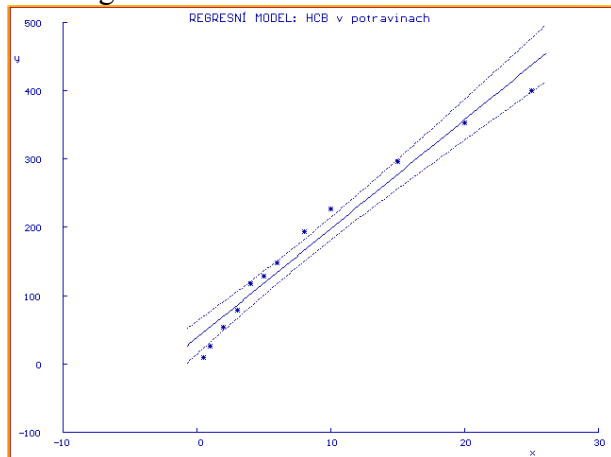
Koeficient determinace ukazuje, model vystihuje data v 96.5 %. Střední kvadratická chyba predikce MEP a Akaikeho informační kritérium AIC slouží pro optimalizaci modelu – za optimální se považuje model, pro který dosahuje MEP a AIC minimální hodnotu.

Vícenásobný korelační koeficient, R	9.8228E-01
Koeficient determinace, R^2	9.6488E-01
Predikovaný korelační koeficient, R_p^2	9.6716E-01
Střední kvadratická chyba predikce, MEP	9.6816E+02
Akaikeho informační kritérium, AIC	7.9191E+01

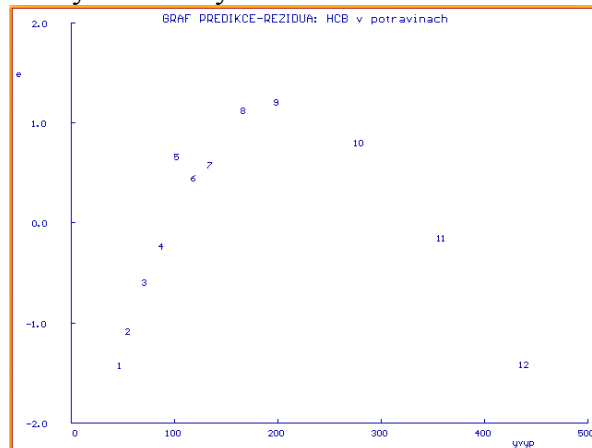
1.5 Regresní diagnostika

1.5.1 Analýza klasických reziduí

Graf regresního modelu



Analýza klasických reziduí



Reziduální součet čtverců, RSC	6.3159E+03
Průměr absolutních hodnot reziduí, Me	2.0396E+01
Průměr relativních hodnot reziduí, Mer	4.5378E+01
Odhad reziduálního rozptylu, $s^2(e)$	6.3159E+02
Odhad směrodatné odchylky reziduí, $s(e)$	2.5131E+01
Odhad šikmosti reziduí, $g1(e)$	-3.0503E-01
Odhad špičatosti reziduí, $g2(e)$	1.6917E+00

1.5.2 Analýza ostatních reziduí

Odlehlé body: Jackknife rezidua – bod 12
 Extrémy: Diagonální prvky – bod 12
 Podezřelé body: Cookova vzdálenost – bod 1, 2, 12
 Atkinsonova vzdálenost – bod 12
 Vliv na predikci – bod 12
 Věrohodnostní vzdálenost – bod 12

1.5.3 Grafy vlivných bodů

Tyto grafy indikují přítomnost odlehlých hodnot a extrémů:

Graf predikovaných reziduí: vlivný bod – bod 12
 Pregibonův graf: středně vlivný bod – bod 12
 Williamsův graf: odlehlý bod - bod 12
 extrém – bod 12
 McCulloh-Meeterův graf: odlehlý bod – bod 12
 extrém - bod 12
 L-R graf: odlehlý bod – bod 12
 extrém – bod 11, 12

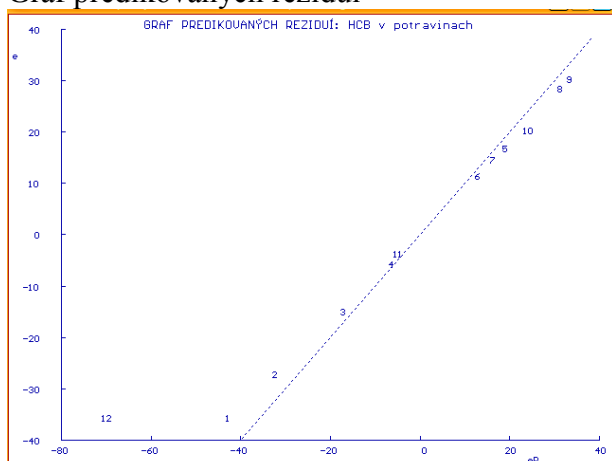
10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT

Kalibrace a limity její přesnosti

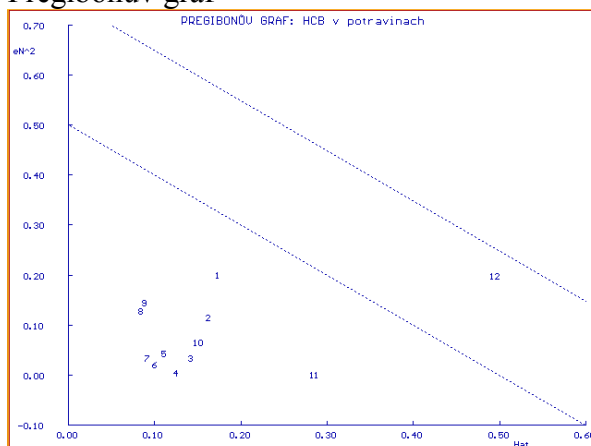
Semestrální práce

2004/2005

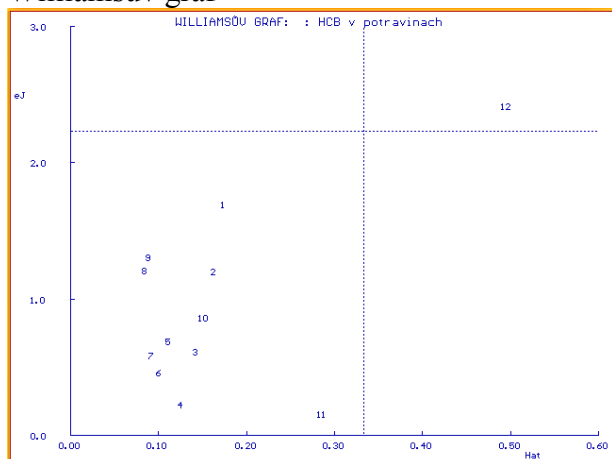
Graf predikovaných reziduí



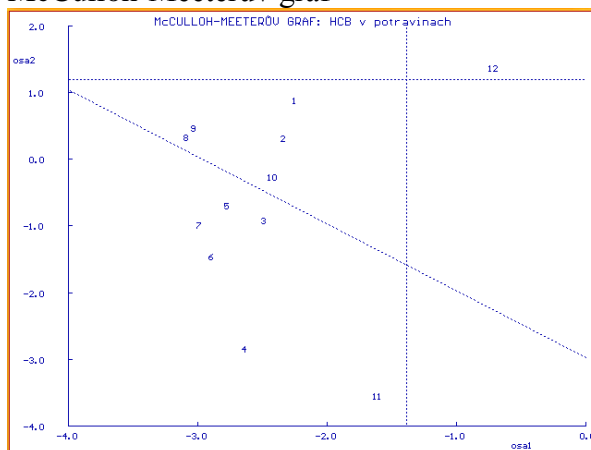
Pregibonův graf



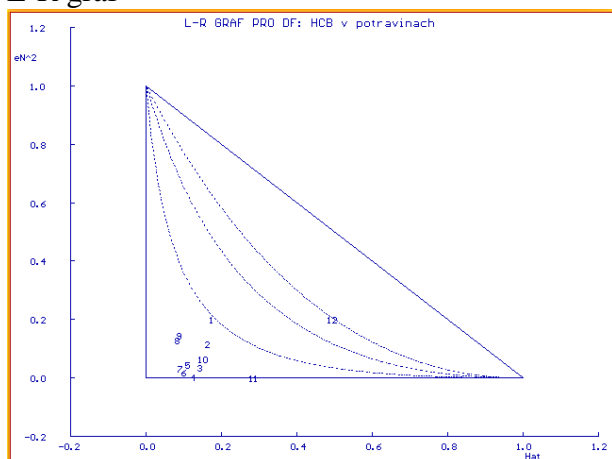
Williamsův graf



McCulloh-Meeterův graf



L-R graf



1.5.4 Indexové grafy

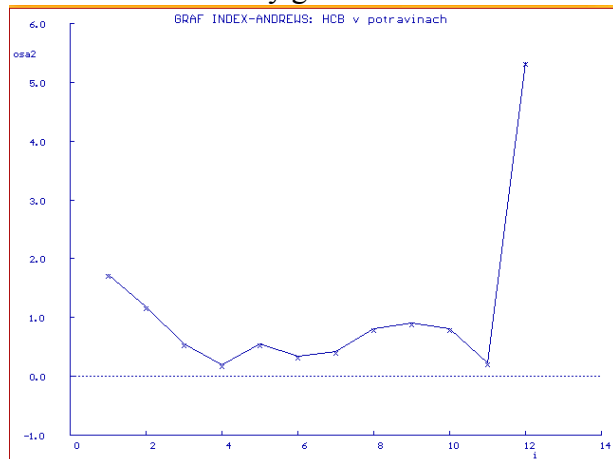
Upozorňují na podezřelé body a kvantifikují velikost jejich vlivu.

Andrewsův indexový graf: podezřelý bod – bod 12

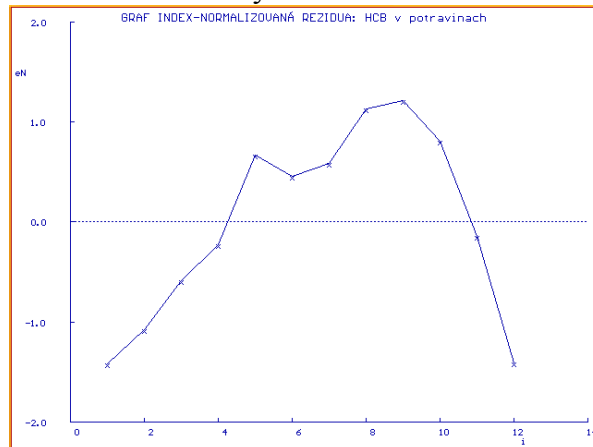
Graf normalizovaných reziduí: podezřelý bod – bod 12

Graf prvků H projekční matice: extrém – bod 12

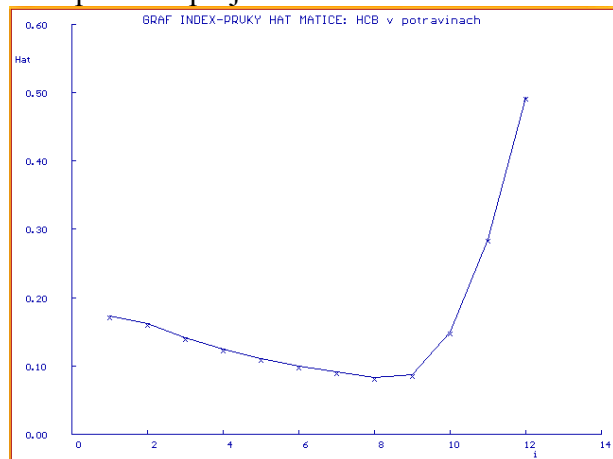
Andrewsův indexový graf



Graf normalizovaných reziduí



Graf prvků H projekční matice



1.5.5 Rankitové grafy

Ukazují vedle normality rozdělení reziduí i vlivné body.

Graf normalizovaných reziduí: odlehlý bod – bod 1

Andrewsův graf: odlehlý bod – bod 12

Graf predikovaných reziduí: odlehlý bod – bod 12

Graf Jackknife reziduí: odlehlý bod – bod 12

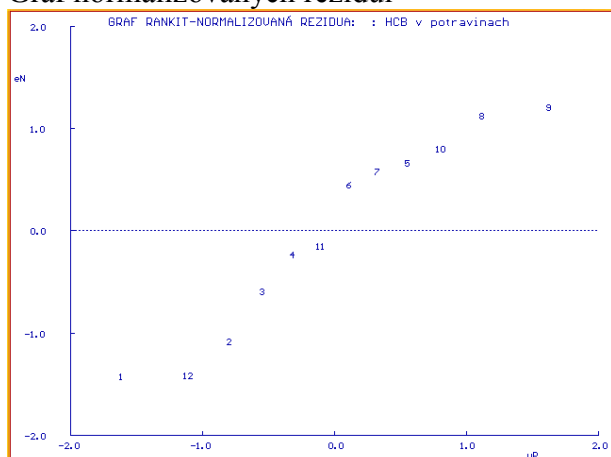
10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT

Kalibrace a limity její přesnosti

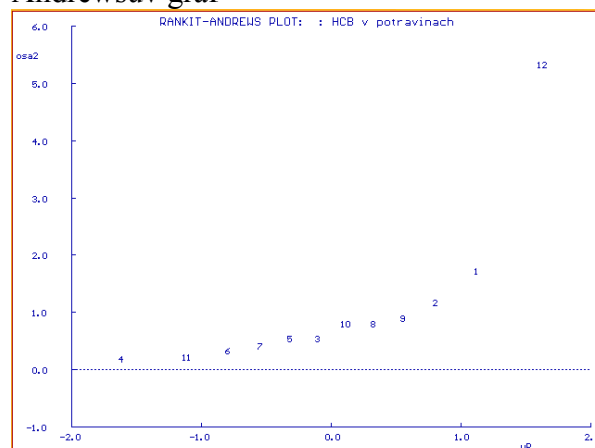
Semestrální práce

2004/2005

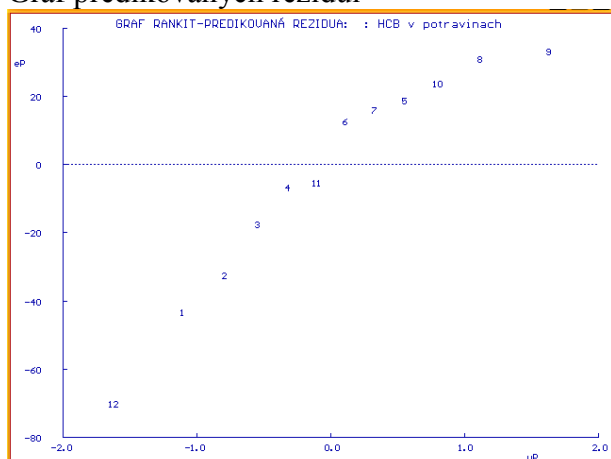
Graf normalizovaných reziduí



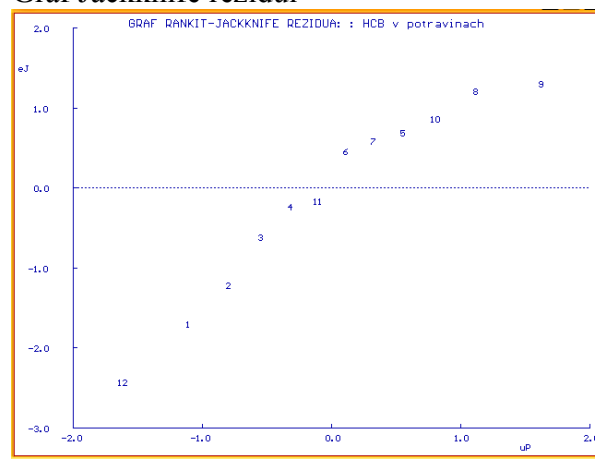
Andrewsův graf



Graf predikovaných reziduí



Graf Jackknife reziduí



Závěr regresní diagnostiky:

Bod 12 je odlehlý, při použití lineárního modelu by bylo nutné ho z výběru vyloučit. Jelikož ale z grafu regresního modelu je zřejmé, že proložení bodů kalibrace přímkou není optimální, ponecháme odlehlý bod č. 12 v souboru základních dat.

1.6 Testování regresního tripletu (data + model + metoda)

Testy předpokladů MNČ

Fischer-Snedocorův test významnosti regrese, F 2.7473E+02

Tabulkový kvantil, F(1-alpha, m-1, n-m) 4.9646E+00

Závěr: Navržený model je přijat jako významný.

Spočtená hladina významnosti 0.000

Scottovo kritérium multikolinearity, M 1.1102E-16

Závěr: Navržený model je korektní.

10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT
Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

Cook-Weisbergův test heteroskedasticity, Sf	2.5415E+01
Tabulkový kvantil, $\chi^2(1-\alpha, 1)$	3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti	0.000
Jarque-Berraův test normality reziduí, L(e)	1.0419E+00
Tabulkový kvantil, $\chi^2(1-\alpha, 2)$	5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	0.594
Waldův test autokorelace, Wa	1.3519E+01
Tabulkový kvantil, $\chi^2(1-\alpha, 1)$	3.8415E+00
Závěr: Rezidua jsou autokorelována.	
Spočtená hladina významnosti	0.000
Znaménkový test, Dt	-2.1194E+00
Tabulkový kvantil, $N(1-\alpha/2)$	1.6449E+00
Závěr: Rezidua vykazují trend.	
Spočtená hladina významnosti	0.017

Závěr testů předpokladů MNČ

Ze závěrů testování vyplývá, že rezidua regresního modelu vykazují autokorelaci a trend, Regresní model není vhodný pro lineární kalibraci.

2. Kalibrace

Postupně jsou vyhodnoceny kvadratický a kubický spline, tj. regrese s různými alternativami zadávání uzlových bodů. Vhodnost použitého modelu budeme posuzovat jednak pomocí grafické analýzy a dále srovnáváním tří parametrů - limita detekce koncentrace x_D , odhad směrodatné odchylky reziduí σ_e a průměr relativních hodnot reziduí Mer . Pro optimální model nabývají tyto parametry minimálních hodnot.

2.1 Kvadratický spline

<i>Parametry</i>	<i>kvadratický spline</i>	
	1	2
uzlové body	1	2
limit detekce x_D [$\mu\text{g.l}^{-1}$]	0.496	0.558
odhad směr. odchylky reziduí σ_e	4.683	4.785
průměr relativních hodnot reziduí Mer	3.166	2.531

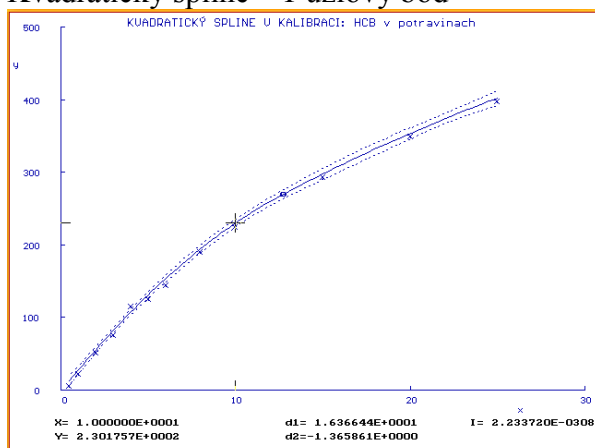
10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT

Kalibrace a limity její přesnosti

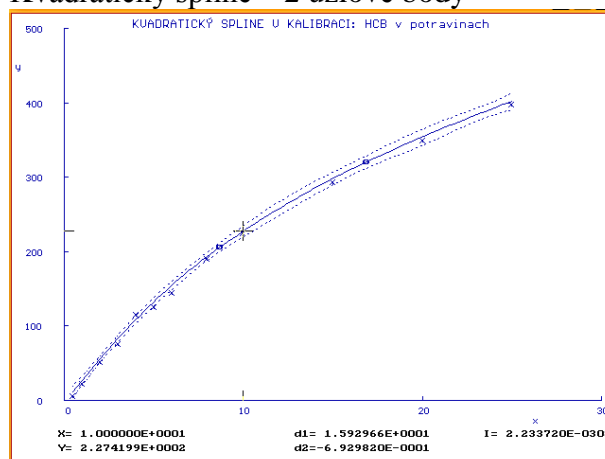
Semestrální práce

2004/2005

Kvadratický spline – 1 uzlový bod



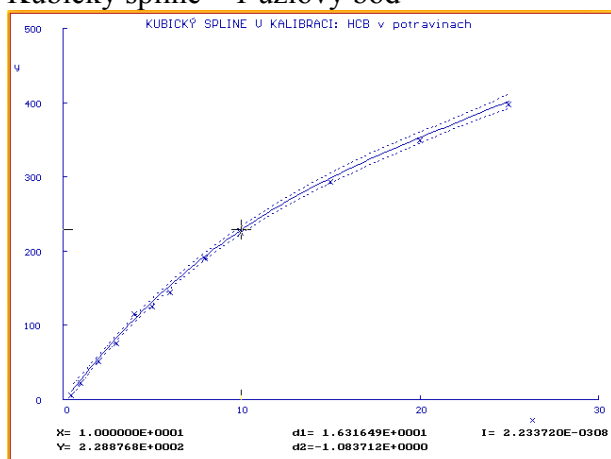
Kvadratický spline – 2 uzlové body



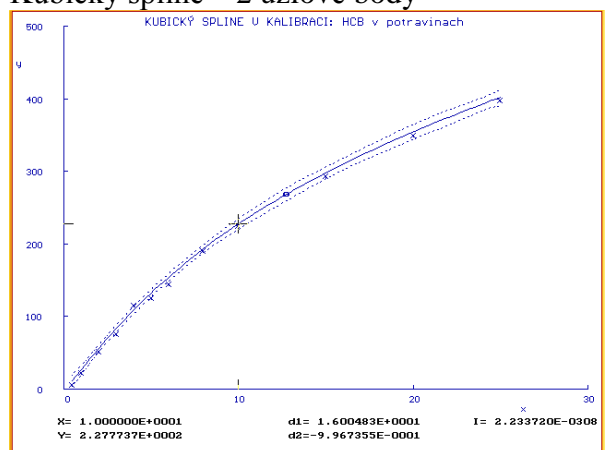
2.2 Kubický spline

<i>Parametry</i>	<i>kubický spline</i>	
uzlové body	1	2
limit detekce x_D [$\mu\text{g.l}^{-1}$]	0.563	0.665
odhad směr. odchylky reziduí σ_e	4.671	5.000
průměr relativních hodnot reziduí Mer	2.272	2.074

Kubický spline – 1 uzlový bod



Kubický spline – 2 uzlové body



Hodnoty sledovaných parametrů – limita detekce koncentrace x_D , odhad směrodatné odchylky reziduí σ_e a průměr relativních hodnot reziduí Mer – nabývají minimálních hodnot u kvadratického splinu s 1 uzlovým bodem.

10. LICENČNÍ STUDIUM CHEMOMETRIE: STATISTICKÉ ZPRACOVÁNÍ DAT

Kalibrace a limity její přesnosti

Semestrální práce

2004/2005

2.3 Kalibrační rovnice

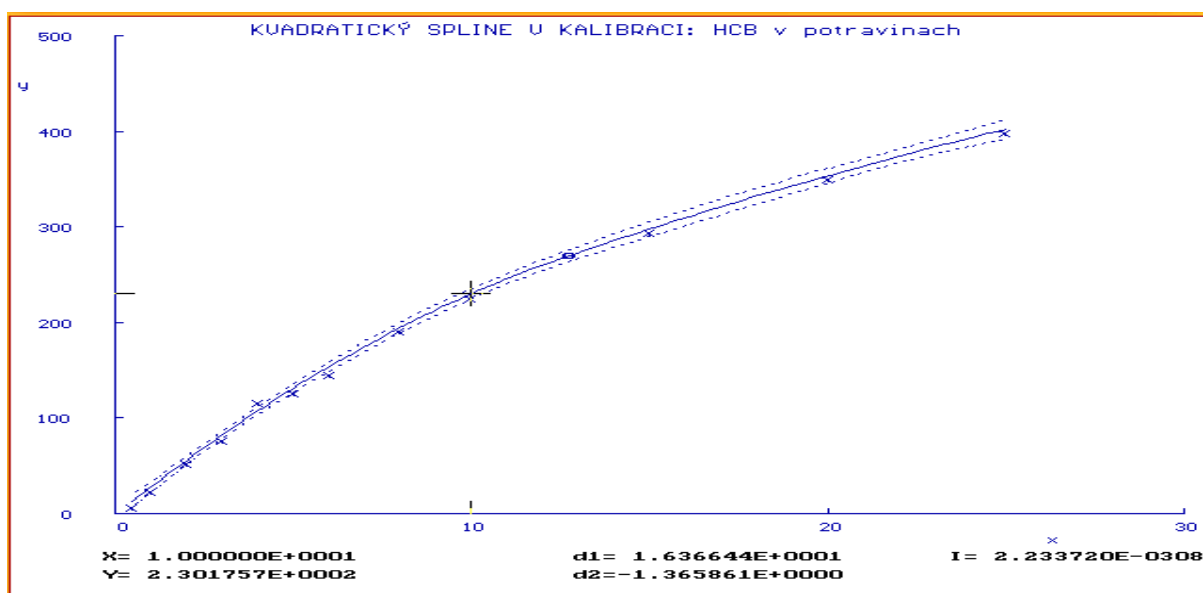
Kalibrační rovnice má tvar

$$Y = a \cdot x^2 + b \cdot x + c \quad \text{pro } k[i-1] < x \leq k[i]$$

Uzlové body a koeficienty kalibrační rovnice:

uzlový bod k[i]	a	b	c
12.75	-0.683	30.025	-1.781

2.4 Graf kalibrační modelu



2.5 Kalibrační tabulka

Měřená hodnota - výšky píku HCB	Inverzní odhad koncentrace HCB [$\mu\text{g}\cdot\text{l}^{-1}$]	Konfidenční interval koncentrace HCB [$\mu\text{g}\cdot\text{l}^{-1}$]	
		dolní mez	horní mez
46.87	1.68	1.50	1.86
111.53	4.17	4.01	4.34
256.48	1.18	11.34	12.15

Závěr

Ze statistické analýzy různých kalibračních modelů se jeví jako nejlepší model kvadratický spline s jedním uzlovým bodem. U tohoto modelu je limita detekce $x_D = 0.28 \mu\text{g.l}^{-1}$, což je dostatečně nízká hodnota k tomu, aby bylo možné tuto metodu použít pro stanovení stopových koncentrací HCB v potravinách. Dále byly vypočteny bodové odhady pro tři neznámé koncentrace HCB v rostlinných tucích a příslušné intervalové odhady, u kterých můžeme s 95% jistotou tvrdit, že neznámá hodnota koncentrace HCB leží právě v těchto intervalech. Výsledky stanovení koncentrací HCB v rostlinných olejích jsou uvedeny v kalibrační tabulce.

Použitá literatura

1. Meloun M., Militký J.: Kompendium statistického zpracování dat, Academia, Praha, 2002

Meloun M., Militký J.: Statistická analýza experimentálních dat, Academia, Praha, 2004