

UNIVERZITA PARDUBICE

Fakulta chemicko-technologická Katedra analytické chemie

Licenční studium chemometrie na téma

Statistické zpracování dat

Semestrální práce z 3. soustředění

Předmět: 2.1 Tvorba lineárních regresních modelů při analýze dat
2.2 Kalibrace a limity její přesnosti

Vedoucí licenčního studia: Prof. RNDr. Milan Meloun, DrSc.

Vypracoval: Ing. Radek Novotný

Licenční studium Statistické zpracování experimentálních dat.

Předmět: 2.1 Tvorba lineárních regresních modelů při analýze dat

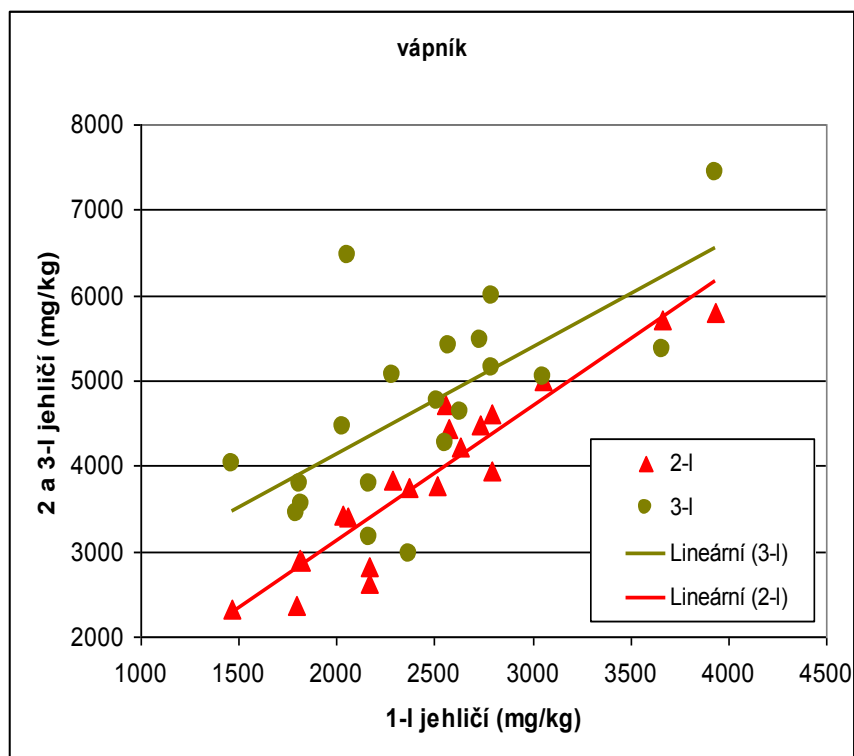
Přednášející: Prof. RNDr. Milan Meloun, DrSc.

Úloha 1. Porovnání dvou regresních přímek u jednoduchého lineárního regresního modelu

Zadání: Obsah vápníku v jehličí borovic stoupá s jejich stářím, přičemž vápník patří mezi živiny v asimilačním aparátu dřevin mezi málo pohyblivé a nedochází k žádným výraznějším přesunům mezi jednotlivými ročníky během vegetační sezóny ani během celého života jehlic. Je možné, že obsah vápníku se zvyšuje každý rok vždy jenom o tento roční přírůstek? Pokud ano, měly by mít regresní přímky vztahu obsahu vápníku ve starším jehličí k obsahu v ročníku nejmladším stejnou směrnici a rozdílný úsek větší o roční přírůstek obsahu vápníku.

Vstupní data:

vápník (mg/kg)		
1-l	2-l	3-l
2166	2815	3784
1798	2371	3446
3937	5791	7441
2035	3423	4469
1808	2898	3793
2793	3950	5158
2167	2627	3174
2570	4431	5420
2284	3844	5064
2791	4605	5985
1822	2877	3564
3664	5710	5376
2556	4711	4258
2633	4214	4632
2512	3776	4759
2056	3409	6478
1467	2329	4025
2367	3744	2970
3052	5001	5049
2734	4474	5473



Řešení:

1. Návrh modelu

Navržený regresní model přímky je $y = \beta_0 + \beta_1 x$. Bude testována nulová hypotéza $H_0: \beta_0 = 0, \beta_1 = 1$, tj. testování úseku a směrnice přímky.

2. Základní analýza dat

Název sloupce:	1-l	2-l	3-l
Průměr:	2460.6	3850	4715.9
Spodní mez:	2173.633687	3369.262424	4179.950658
Horní mez:	2747.566313	4330.737576	5251.849342
Rozptyl:	375961.5158	1055109.895	1311381.463

Název sloupce:	1-1	2-1	3-1
Směr. odchylka:	613.1570075	1027.185424	1145.15565
Šikmost:	0.763102097	0.2584966416	0.5173566674
Odchylka od 0:	Nevýznamná	Nevýznamná	Nevýznamná
Špičatost:	3.367835049	2.208890825	2.875958834
Odchylka od 3:	Nevýznamná	Nevýznamná	Nevýznamná
Polosuma:	2702	4060	5205.5
Modus:	2401.319048	3737.619048	4658.585714
Medián:	2439.5	3810	4695.5
IS spodní:	2077.485619	2968.503444	3850.265832
IS horní:	2801.514381	4651.496556	5540.734168
Med. směr. odchylka:	172.9623607	402.0482012	403.8339483
Medianový rozptyl:	29915.9782	161642.7561	163081.8578

Data jsou nezávislá, homogenita přijata, normalita přijata.

3a. Odhadování parametrů

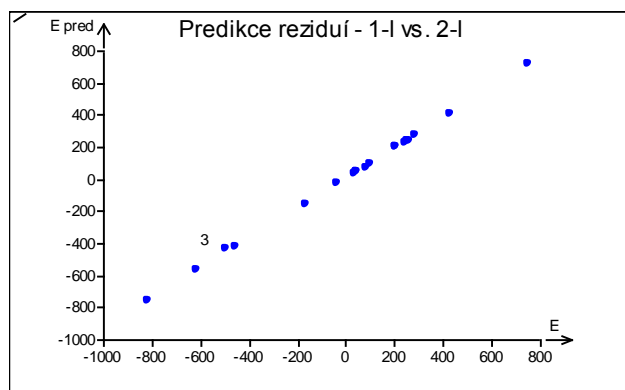
Proměnná	Odhad	Směr. Odch.	Závěr	Pravděpodobnost
Abs	-12.0186969	349.5224215	Nevýznamný	0.9729477559
1-1	1.569543484	0.1380345833	Významný	1.195219035E-009
Proměnná	Spodní mez	Horní mez		
Abs	-746.3380558	722.300662		
1-1	1.279543586	1.859543382		

Úsek směrnice (absolutní člen) je statisticky nevýznamný, tedy platí hypotéza $H_0: \beta_0=0$.
směrnice přímky je naopak statisticky významná.

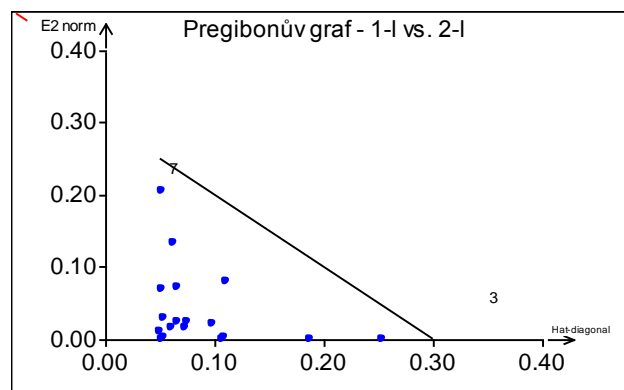
4a. Základní statistické charakteristiky

Vícenásobný korelační koeficient r :	0.9369063886
Koeficient determinace R^2 :	0.877793581
Predikovaný korelační koeficient R_p^2 :	0.7246816192
Střední kvadratická chyba predikce MEP :	149067.8027
Akaikeho informační kritérium AIC :	238.3163697

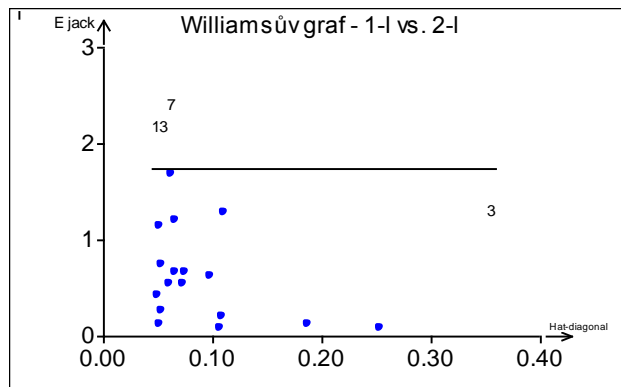
5a. Regresní diagnostika



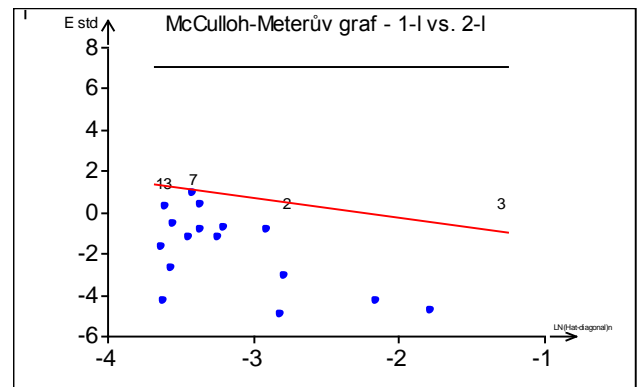
Obr. 1 – Graf predikce reziduí



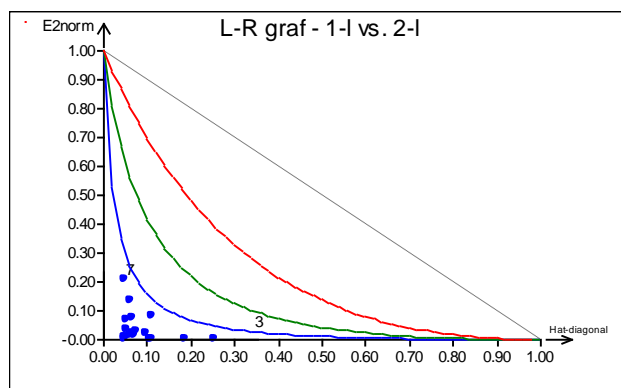
Obr. 2 – Pregibonův graf



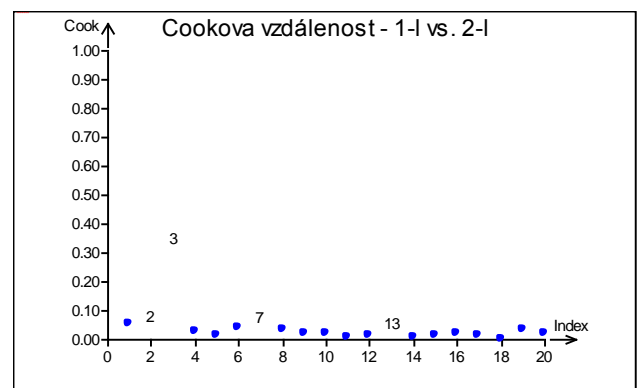
Obr. 3 – Williamsův graf



Obr. 4 – McCulloh-Meterův graf



Obr. 5 – L-R graf



Obr. 6 – Cookova vzdálenost

Z grafů na obr. 1-6 vyplývá, že nejvíce podezřelými body jsou bod č. 3 a č. 7. Tyto body budou odstraněny a bude proveden nový odhad parametrů regresní přímky a porovnány statistické rozhodující charakteristiky.

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F: 129.2917719

Kvantil F (1-alfa, m-1, n-m): 4.413873419

Pravděpodobnost: 1.195218969E-009

Závěr: Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC: 0.3333292688

Závěr: Model vykazuje multikolinearitu!

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW: 0.01215608537

Kvantil $\chi^2(1-\text{alfa}, 1)$: 3.841458829

Pravděpodobnost: 0.9122074309

Závěr: Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB: 0.5766130493

Kvantil $\chi^2(1-\text{alfa}, 2)$: 5.991464547

Pravděpodobnost: 0.7495318071

Závěr: Rezidua mají normální rozdělení.

Waldův test autokorelace

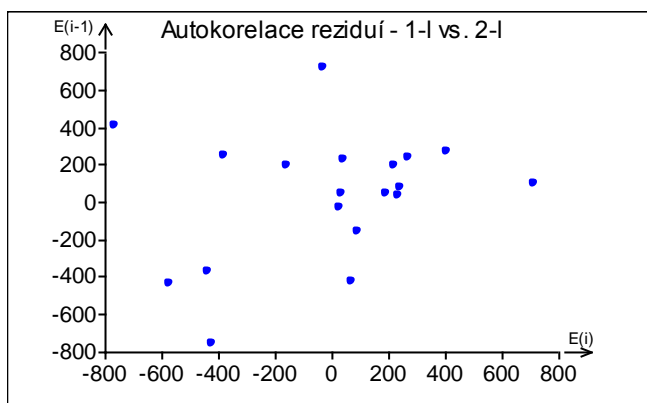
Hodnota kritéria WA: 1.008450358
 Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
 Pravděpodobnost: 0.3152743716
 Závěr: Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

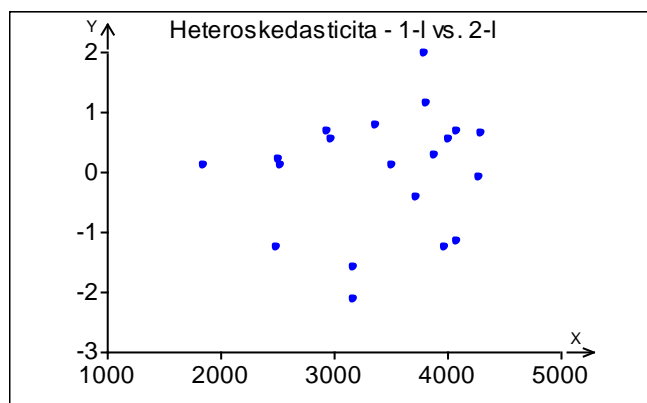
Hodnota kritéria DW: -1
 Kritické hodnoty DW 1.1 1.54
 Závěr: Pozitivní autokorelace reziduí není prokázána.

Znaménkový test reziduí

Hodnota kritéria Sg: 0.8123320297
 Kvantil $N(1-\alpha/2)$: 1.959963999
 Pravděpodobnost: 0.4166011364
 Závěr: V reziduích není trend.



Obr. 7 – Graf autokorelace reziduí



Obr. 8 – Graf heteroskedasticity

6a. Konstrukce zpřesněného modelu

Po odstranění podezřelých bodů (č. 3 a 7) byly vypočteny nové odhady parametrů přímky.

Proměnná	Odhad	Směr.odch.	Závěr	Pravděpodobnost
Abs	-151.6373965	355.9375808	Nevýznamný	0.6757672833
1-l	1.654251488	0.1452850437	Významný	4.385211527E-009
Proměnná	Spodní mez	Horní mez		
Abs	-906.1913602	602.9165673		
1-l	1.346260954	1.962242022		

Statistické charakteristiky (v závorce jsou původní hodnoty):

Vícenásobný korelační koeficient r :	0.9434750005	(0.9369063886)
Koeficient determinace R^2 :	0.8901450766	(0.877793581)
Predikovaný korelační koeficient R_p^2 :	0.7505408782	(0.7246816192)
Střední kvadratická chyba predikce MEP :	109567.7858	(149067.8027)
Akaikeho informační kritérium AIC :	209.3465607	(238.3163697)

Zpřesnění modelu bylo úspěšné – snížily se rozhodující kritéria MEP a AIC a zvýšily se parametry r , R^2 , předpoklady metody nejmenších čtverců jsou splněny.

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F: 129.6466356
 Kvantil F (1-alfa, m-1, n-m): 4.493998478
 Pravděpodobnost: 4.385211427E-009
 Závěr: Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC : 0.3327114362
 Závěr: Model vykazuje multikolinearitu!

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW: 0.005415995037
 Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
 Pravděpodobnost: 0.9413338877
 Závěr: Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB: 0.3479306089
 Kvantil $\chi^2(1-\alpha,2)$: 5.991464547
 Pravděpodobnost: 0.8403260527
 Závěr: Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA: 0.01336595637
 Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
 Pravděpodobnost: 0.9079606628
 Závěr: Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW: -1
 Kritické hodnoty DW 1.05 1.53
 Závěr: Rezidua nejsou autokorelována

Znaménkový test reziduí

Hodnota kritéria Sg: 0.8264597475
 Kvantil $N(1-\alpha/2)$: 1.959963999
 Pravděpodobnost 0.4085433428
 Závěr: V reziduích není trend.

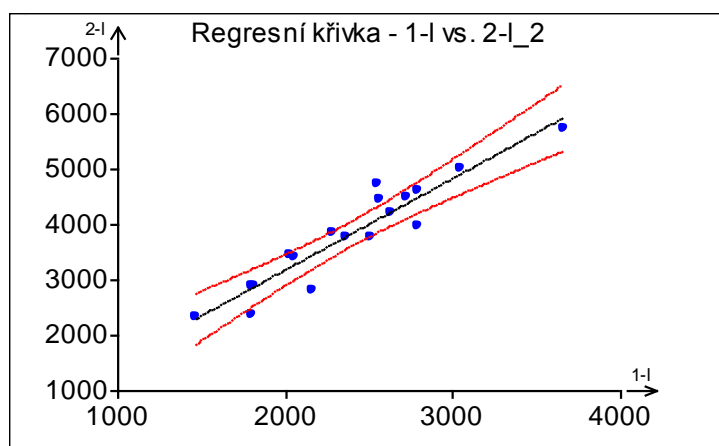
7a. Zhodnocení kvality modelu

Nalezený regresní model má tvar

$$y = -151.64 (355.94) + 1.654 (0.145)x$$

Intervalový odhad parametrů úseku (β_0) a směrnice (β_1) je (výstup z programu QCExpert):

Proměnná	Spodní mez	Horní mez
β_0	-906.1913602	602.9165673
β_1	1.346260954	1.962242022



Obr. 9 – Regresní křivka zpřesněného modelu

3b. Odhadování parametrů

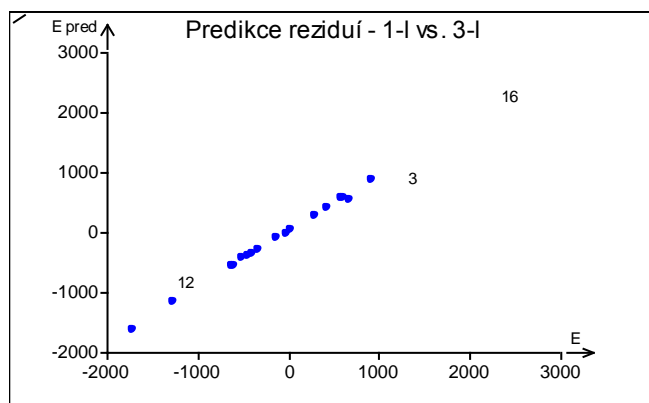
Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
Abs	1651.489346	830.6611751	Nevýznamný	0.06222005048
1-l	1.245391634	0.328047536	Významný	0.001322025073
Proměnná	Spodní mez	Horní mez		
Abs	-93.66502433	3396.643717		
1-l	0.5561893349	1.934593932		

Úsek směrnice (absolutní člen) je statisticky nevýznamný, tedy platí hypotéza $H_0: \beta_0=0$.
směrnice přímky je naopak statisticky významná.

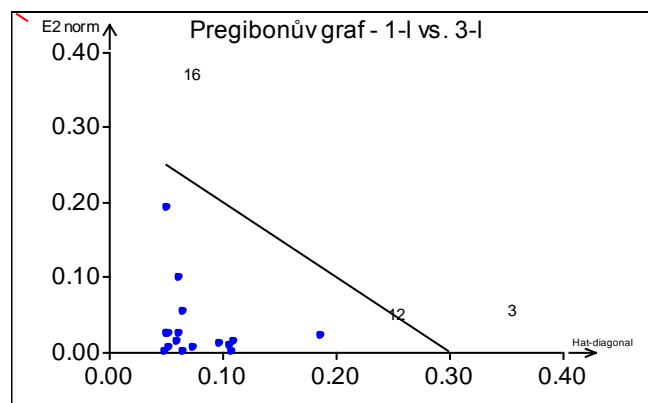
4b. Základní statistické charakteristiky

Vícenásobný korelační koeficient r :	0.666826913
Koeficient determinace R^2 :	0.4446581319
Predikovaný korelační koeficient R_p^2 :	0.09078021724
Střední kvadratická chyba predikce MEP :	870452.1617
Akaikeho informační kritérium AIC :	272.9425404

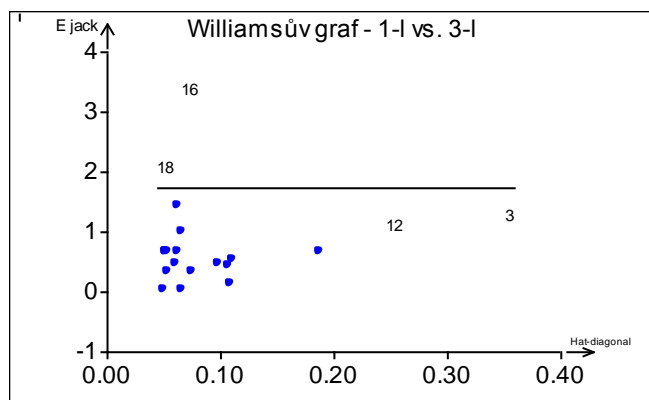
5b. Regresní diagnostika



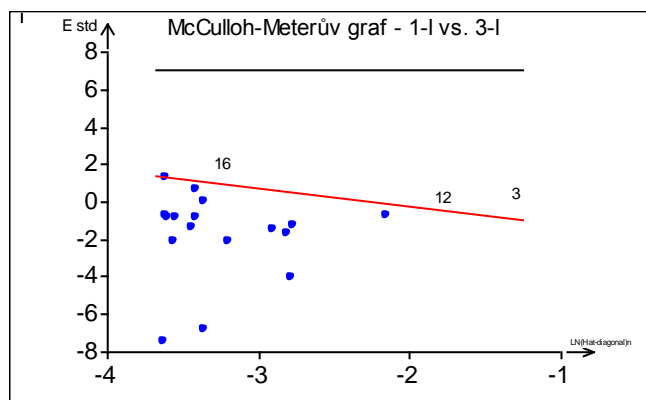
Obr. 10 – Graf predikce reziduí



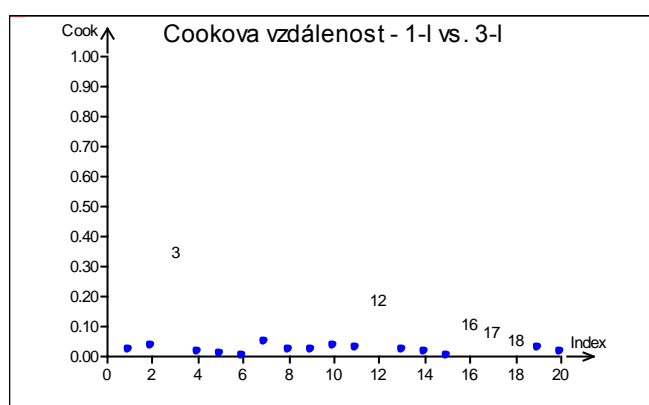
Obr. 11 – Pregibonův graf



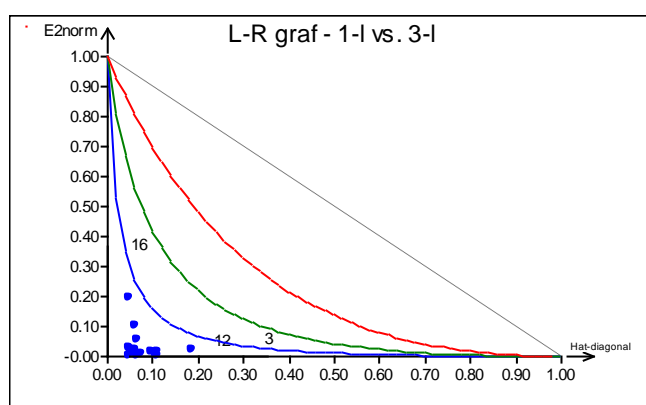
Obr. 12 – Williamsův graf



Obr. 13 – McCulloh-Meterův graf



Obr. 14 – Cookova vzdálenost



Obr. 15 – L-R graf

Na základě diagnostiky podle grafů na obr. 10-16 vyplývá, že nejvíce podezřelými body jsou body č. 3, 12 a 16. Tyto body budou odstraněny a bude proveden nový odhad parametrů regresní přímky a porovnány statistická rozhodující kritéria.

6b. Konstrukce zpřesněného modelu

Po odstranění podezřelých bodů (č. 3, 12 a 16) byly vypočteny nové odhady parametrů přímky.

Proměnná	Odhad	Směr.odch.	Závěr	Pravděpodobnost
Abs	1280.472265	916.7057109	Nevýznamný	0.1827924781
1-I	1.346352459	0.3876185245	Významný	0.003404407619
Proměnná	Spodní mez	Horní mez		
Abs	-673.4397061	3234.384236		
1-I	0.5201631305	2.172541787		

Statistické charakteristiky (v závorce jsou původní hodnoty):

Vícenásobný korelační koeficient r :	0.6676585684	(0.666826913)
Koeficient determinace R^2 :	0.4457679639	(0.4446581319)
Predikovaný korelační koeficient R_p^2 :	0.08988582829	(0.09078021724)
Střední kvadratická chyba predikce MEP :	510349.7052	(870452.1617)
Akaikeho informační kritérium AIC :	223.4544056	(272.9425404)

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F: 12.06447665
 Kvantil F (1-alfa, m-1, n-m): 4.543077165
 Pravděpodobnost: 0.003404407619
 Závěr: Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC: 0.265132937
 Závěr: Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW: 0.001381738543
 Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
 Pravděpodobnost: 0.9703480682
 Závěr: Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB: 0.918628237
 Kvantil $\chi^2(1-\alpha,2)$: 5.991464547
 Pravděpodobnost: 0.6317167798
 Závěr: Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA: 0.04622319144
 Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
 Pravděpodobnost: 0.8297705978
 Závěr: Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW: -1
 Kritické hodnoty DW 1.02 1.54
 Závěr: Rezidua nejsou autokorelována

Znaménkový test reziduí

Hodnota kritéria Sg: 0.4880460802
 Kvantil $N(1-\alpha/2)$: 1.959963999
 Pravděpodobnost: 0.6255172021
 Závěr: V reziduích není trend.

Zpřesnění modelu bylo úspěšné – snížily se rozhodující kritéria *MEP* a *AIC* a zvýšily se parametry r , R^2 , předpoklady metody nejmenších čtverců jsou splněny.

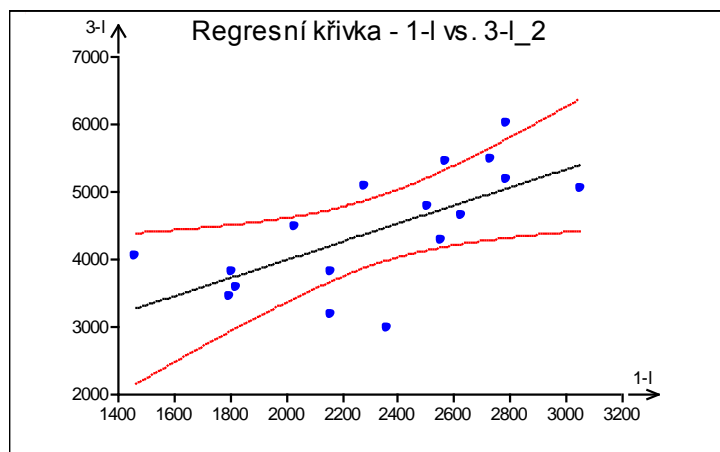
7b. Zhodnocení kvality modelu

Nalezený regresní model má tvar

$$y = 1280.47 (916.706) + 1.346 (0.387)x$$

Intervalový odhad parametrů úseku (β_0) a směrnice (β_1) je (výstup z programu QCExpert):

Proměnná	Spodní mez	Horní mez
β_0	-673.4397061	3234.384236
β_1	0.5201631305	2.172541787



Obr. 16 – Regresní křivka zpřesněného modelu

8. Porovnání regresních přímek

1-l vs. 2-l: $y = -151.64 (355.94) + 1.654 (0.145)x$
 Reziiduální součet čtverců: 1620935.516
 Reziiduální směr. odchylka: 318.2899146
 β_0 (-906.1913602; 602.9165673)
 β_1 (1.346260954; 1.962242022)

1-l vs. 3-l: $y = 1280.47 (916.706) + 1.346 (0.387)x$
 Reziiduální součet čtverců: 6867399.243
 Reziiduální směr. odchylka: 676.6288615
 β_0 (-673.4397061; 3234.384236)
 β_1 (0.5201631305; 2.172541787)

Úseky přímek se liší, tato odlišnost byla očekávána. Interval spolehlivosti směrnice u první přímky neobsahuje jedničku, interval spolehlivosti směrnice druhé přímky jedničku obsahuje. Během růstu jehlic tedy dochází k úbytku vápníku. Tento úbytek může být způsoben řadou vlivů, např. semenným rokem a tedy vyčerpáním živin ve prospěch plodů a semen nebo vyplavováním prvků z jehlic působením srážek.

Závěr: H_0 je zamítnuta, regresní přímky mají statisticky významně odlišné směrnice.

Licenční studium Statistické zpracování experimentálních dat.

Předmět: 2.1 Tvorba lineárních regresních modelů při analýze dat

Přednášející: Prof. RNDr. Milan Meloun, DrSc.

Úloha 2. Určení stupně polynomu

Zadání: Úkolem je určení optimálního stupně polynomu metodou nejmenších čtverců a metodou racionálních hodnot. Použitá jsou data z roku 1999 charakterizující změnu pH(KCl) s hloubkou v půdním profilu u Moldavy v Krušných horách.

Tab. 1 – Vstupní data

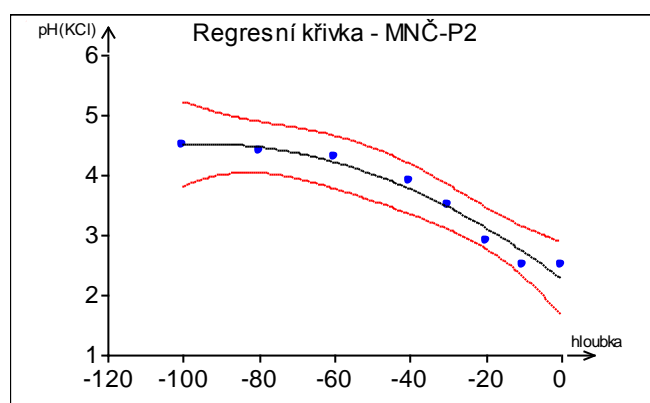
Horizont	pH(KCl)
H	2.5
0-10 cm	2.5
10-20 cm	2.9
20-30 cm	3.5
30-45 cm	3.9
45-60 cm	4.3
60-80 cm	4.4
80+ cm	4.5

A1. Návrh modelu

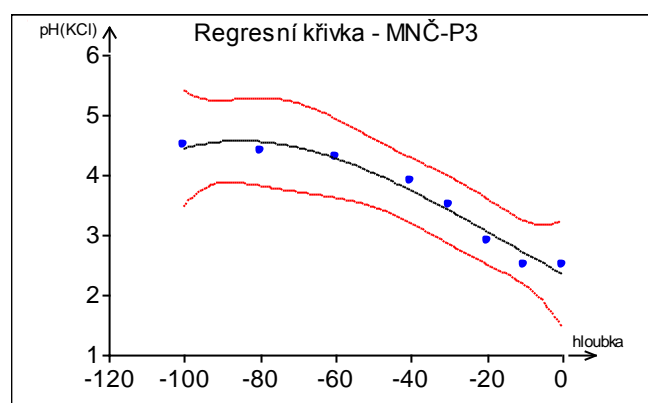
Hledám stupeň polynomu metodou nejmenších čtverců (MNČ). V tabulce č. 2 jsou uvedeny základní statistické charakteristiky pro různé stupně polynomu.

Tab. 2 – Statistické charakteristiky regrese pro různé stupně polynomu

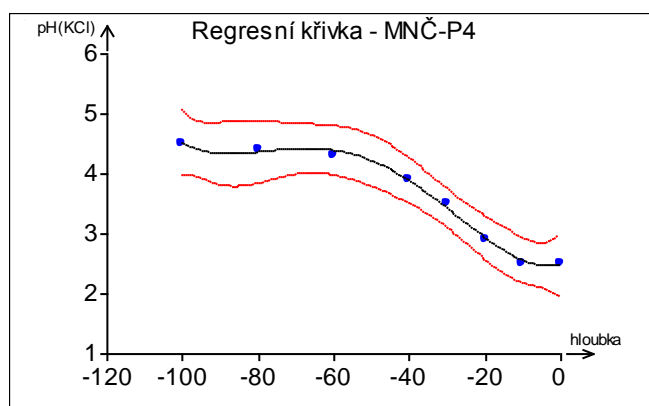
	P2	P3	P4	P5	P6
Vícenásobný korelační koeficient R:	0.9817456	0.9849367	0.9979174	0.9997856	0.9998036
Koeficient determinace R ² :	0.9638244	0.9701003	0.9958391	0.9995712	0.9996072
Predikovaný korelační koeficient Rp:	0.7804504	0.2003068	0.1155862	0.8746016	14909.00567
Střední kvadratická chyba predikce MEP:	0.0719631	0.3410476	0.4074594	0.0400030	75.9965215
Akaikeho informační kritérium:	-24.4135886	-23.9378759	-37.7147628	-53.8948818	-52.5972286



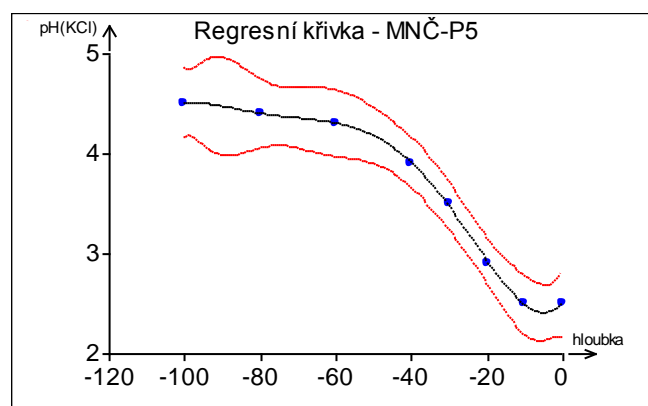
Obr. 1 – Polynom 2. stupně (MNČ)



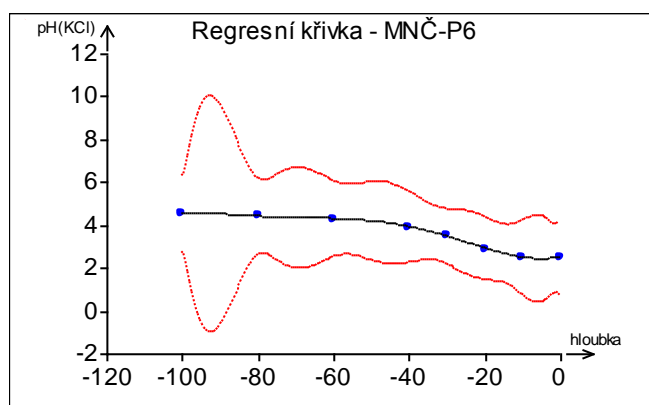
Obr. 2 – Polynom 3. stupně (MNČ)



Obr. 3 – Polynom 4. stupně (MNČ)



Obr. 4 – Polynom 5. stupně (MNČ)



Obr. 5 – Polynom 6. stupně (MNČ)

Po vyhodnocení statistických charakteristik regrese pro různé stupně polynomu a grafů těsnosti proložení regresních křivek se jako nejvhodnější model jeví $m=4$. Možné řešení je i varianta $m=5$, ale pro polynom čtvrtého stupně hovoří jednak to, že se jedná o jednodušší variantu a také červené pásy spolehlivosti o něco těsněji kopírují křivku proložení a netvoří „boule“ na koncích.

A2. Předběžná analýza dat

Proměnná	Průměr	Směr. odch.	Kor.vs.Y	Významnost
hloubka	-42.5	34.94894235	-0.9379999344	0.0005684596952
hloubka ²	2875	3612.18177	0.8240867822	0.01187692022
hloubka ³	-228500	357769.2792	-0.7316786105	0.03909795853
hloubka ⁴	19682500	35362837.94	0.6651889163	0.07184586825

Párové korelace (X_i, X_j)

hloubka - hloubka ²	-0.9624377387	0.0001287891595
hloubka - hloubka ³	0.9113904398	0.001625786657
hloubka - hloubka ⁴	-0.8660418296	0.005422025664
hloubka ² - hloubka ³	-0.9870230082	5.410352899E-006
hloubka ² - hloubka ⁴	0.9629076146	0.0001240604877
hloubka ³ - hloubka ⁴	-0.9934070117	7.129137212E-007

Indikace multikolinearity

Proměnná	Vlas. čísla kor. m.	Podmíněnost kappa	VIF faktor	Vícenás. kor.
Abs	4.683993197E-005	1	1	0
hloubka	1	21349.30513	221.8700904	0.997743884
hloubka ²	0.1524059946	3253.762082	5216.944716	0.999904153
hloubka ³	0.004739508893	101.1852215	12943.73924	0.999961370
hloubka ⁴	3.842807657	82041.27323	3184.565061	0.999842980

Pearsonův párový korelační koeficient ukazuje vysokou korelaci všech párů a velká korelace je indikována i mezi jednotlivými nezávisle proměnnými.

Maximální číslo podmíněnosti kappa je větší než 1000 a indikuje tak multikolinearitu stejně jako kritérium VIF, které dosahuje hodnot podstatně vyšších než 10.

A3. Odhadování parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
Abs	2.472097843	0.07980813676	Významný	7.392446805E-005
hloubka	0.01476606768	0.01333243738	Nevýznamný	0.348883669
hloubka ²	0.002704938029	0.0006255078013	Významný	0.0227942787
hloubka ³	4.503887656E-005	9.94766248E-006	Významný	0.02015638666
hloubka ⁴	2.150461829E-007	4.991972245E-008	Významný	0.02302829988

Parametry byly nalezeny klasickou MNC, parametry jsou statisticky významné kromě parametru x.

A4. Základní statistické charakteristiky

Pro zvolený model mají základní statistické charakteristiky tyto hodnoty (viz také tab. 2):

Vícenásobný korelační koeficient R:	0.9979173651
Koeficient determinace R ² :	0.9958390675
Predikovaný korelační koeficient Rp:	0.1155862014
Střední kvadratická chyba predikce MEP:	0.4074594026
Akaikeho informační kritérium:	-37.71476278

A5. Regresní diagnostika

Data

Indikace vlivných dat

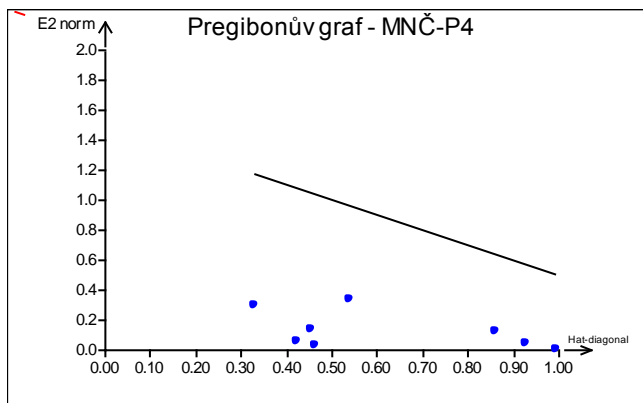
A. Analýza reziduí

Index	Standardní	Jackknife	Predikované
1	1.272762887	1.532185028	0.3976890946
2	-0.8524262593	-0.7995337589	-0.09564042508
3	-0.5235200545	-0.4484264589	-0.05715354181
4	1.156863214	1.269185044	0.1170782991
5	0.3726344369	0.3115502898	0.04206076492
6	-1.470740656	-2.273571538	-0.1795734703
7	1.59550463	3.347423364	0.3544386091
8	-1.626542737	-3.864193908	-1.707558254

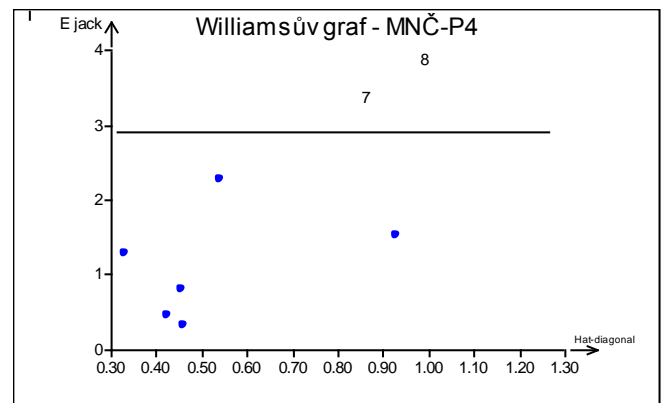
Index	Diag(Hii)	Diag(H*ii)	Cookova vzdál.
1	0.9298392696	0.9677243186	3.373582076*
2	0.4558519164	0.5876501195	-0.1428214693
3	0.4252661087	0.4777725034	-0.07747423278
4	0.3311978301	0.6295577224	0.1145781528
5	0.4623519167	0.4872372134	0.06408959745
6	0.540511862	0.8718147681	-0.3460166672
7	0.8611964575	0.9789775106	1.979838426*
8	0.993784639	0.9992658442	-52.01413665

Analýza ostatních reziduí neukazuje žádný odlehlý bod.

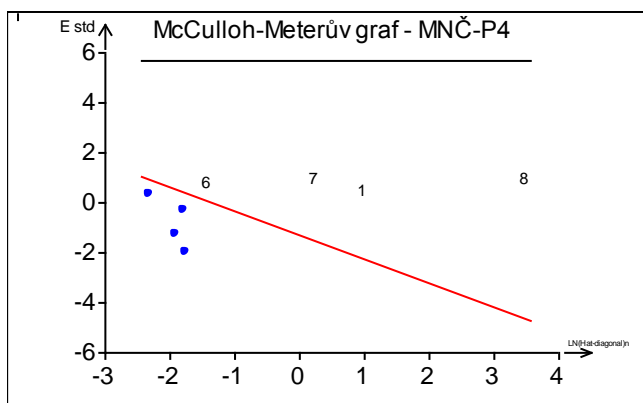
B. Grafy vlivných bodů



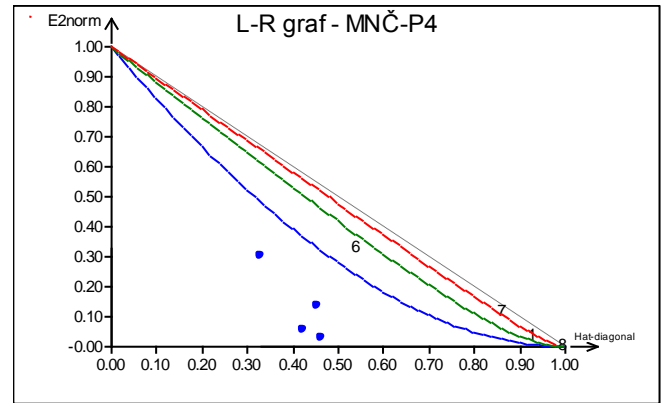
Obr. 6 – Pregibonův graf



Obr. 7 – Williamsův graf



Obr. 8 – McCulloh-Meterův graf



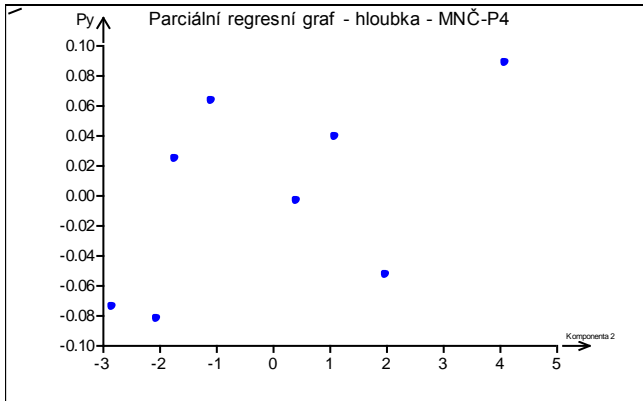
Obr. 9 – L-R graf

Žádný bod není většinou diagnostik prokázán za odlehlý. Body 7 a 8 jsou sice jako podezřelé zobrazeny na 2-3 grafech, ale jedná se spíše o extrémní než o odlehlé vlivné body.

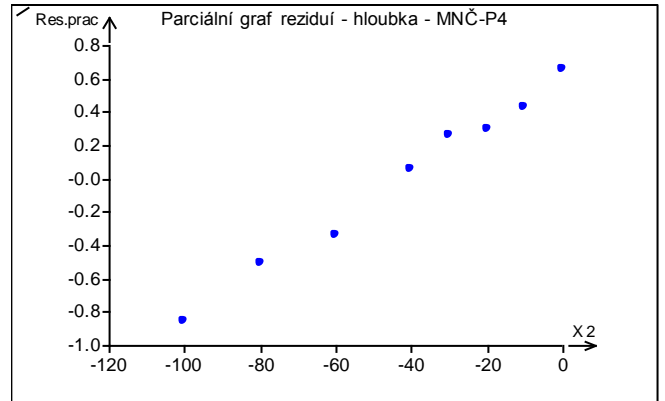
Model

Parciální regresní grafy a parciální reziduální grafy ukazují lineární závislost mezi všemi nezávisle proměnnými. Navržený model má tvar

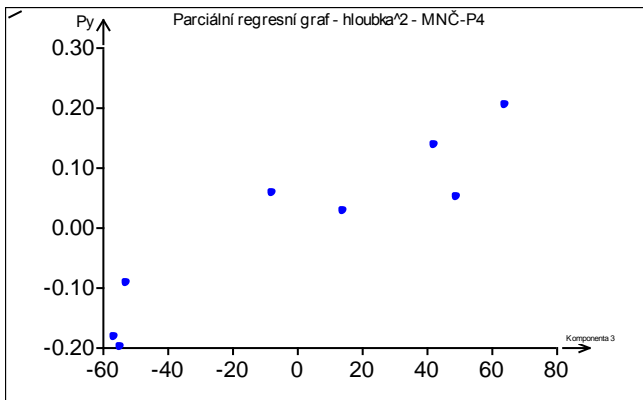
$$\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$$



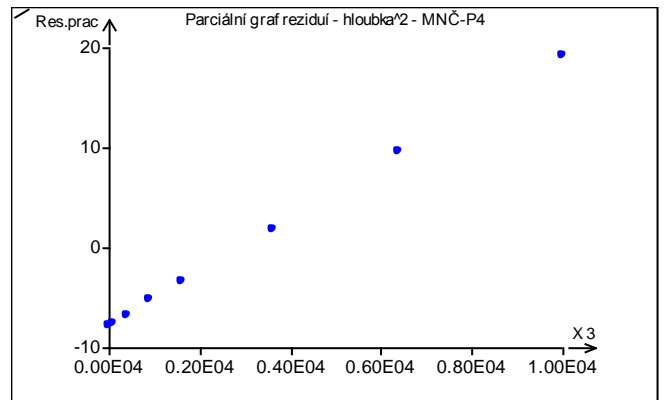
Obr. 11 – Parciální regresní graf – proměnná x



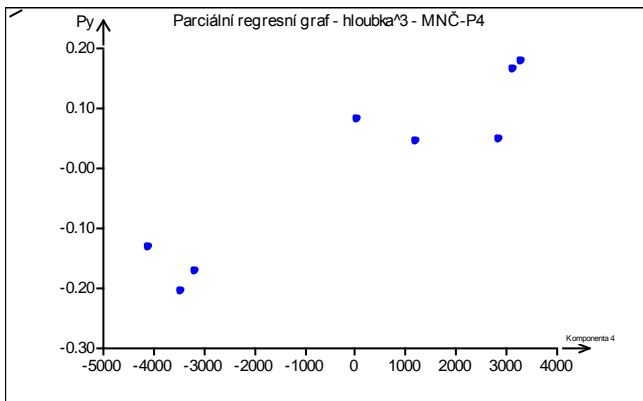
Obr. 12 – Parciální reziduální graf – proměnná x



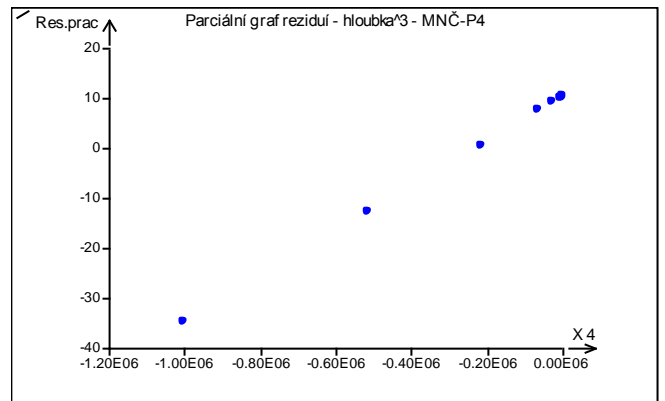
Obr. 13 – Parciální regresní graf – proměnná x^2



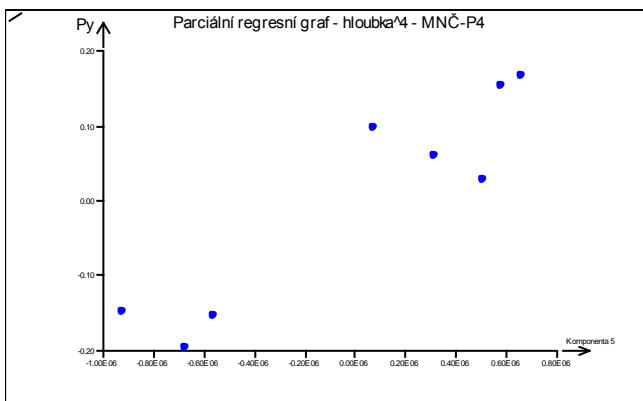
Obr. 14 – Parciální reziduální graf – proměnná x^2



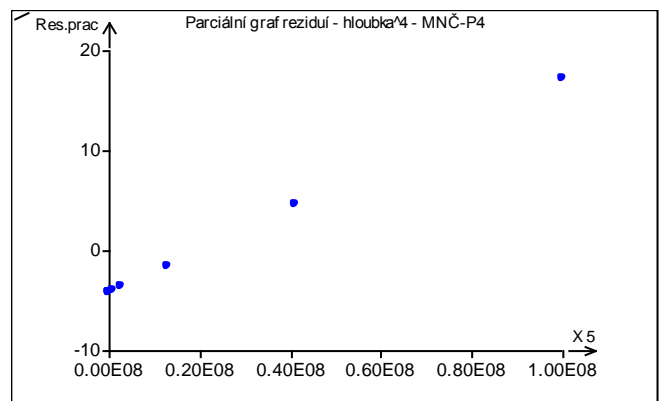
Obr. 15 – Parciální regresní graf – proměnná x^3



Obr. 16 – Parciální reziduální graf – proměnná x^3



Obr. 17 – Parciální regresní graf – proměnná x^4



Obr. 18 – Parciální reziduální graf – proměnná x^4

Metoda

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F: 179.4980569
Kvantil F (1-alfa, m-1, n-m): 9.117182253
Pravděpodobnost: 0.0006693301966
Závěr: Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC: -0.06314104508
Závěr: Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW: 0.01498976548
Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
Pravděpodobnost: 0.9025563442
Závěr: Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB: 0.4329862522
Kvantil $\chi^2(1-\alpha,2)$: 5.991464547
Pravděpodobnost: 0.8053380707
Závěr: Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA: 0.1737716165
Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
Pravděpodobnost: 0.676781377
Závěr: Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW: -1
Kritické hodnoty DW 0 2
Závěr: Negativní autokorelace reziduí není prokázána.

Znaménkový test reziduí

Hodnota kritéria Sg: 1.145643924
Kvantil $N(1-\alpha/2)$: 1.959963999
Pravděpodobnost: 0.2519425152
Závěr: V reziduích není trend.

Podle uvedených kritérií jsou splněny všechny předpoklady MNČ. Použití této metody by tedy v tomto případě mělo vést k nejlepším nestranným odhadům lineárních regresních parametrů.

A6. Zhodnocení kvality modelu

Nalezený model polynomické závislosti má tvar

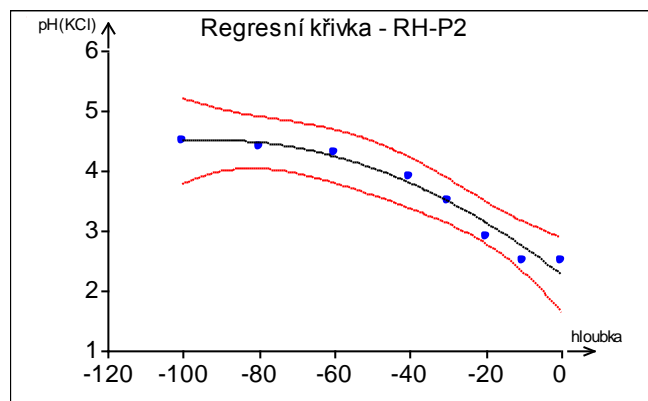
$$y = 2.472 (0.079) + 0.0027 (0.00062) x^2 + 4.504E-5 (9.95E-6) x^3 + 2.15E-7 (4.99E-8) x^4$$

B1. Návrh modelu

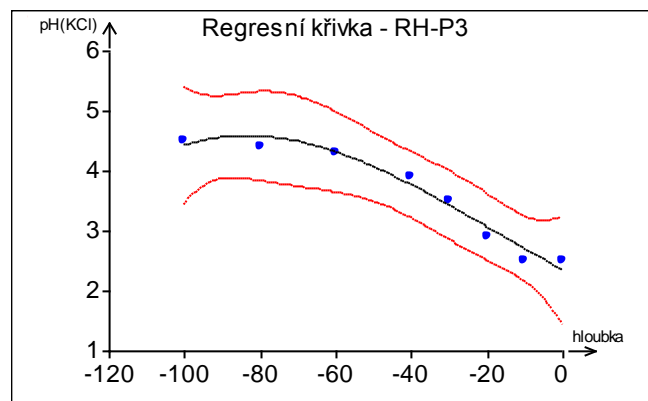
Hledám stupeň polynomu metodou racionálních hodnot (RH). V tabulce č. 2 jsou uvedeny základní statistické charakteristiky pro různé stupně polynomu.

Tab. 3 – Statistické charakteristiky regrese pro různé stupně polynomu

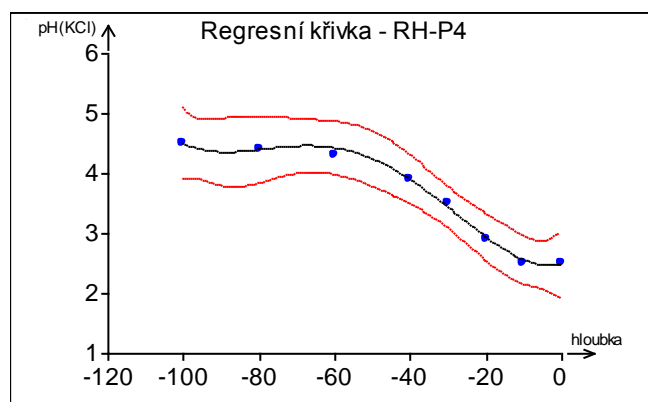
	RH-P2	RH-P3	RH-P4	RH-P5
Vícenásobný korelační koeficient R:	0.981377374	0.984555798	0.997591712	0.999502948
Koeficient determinace R ² :	0.96310155	0.969350119	0.995189224	0.999006143
Predikovaný korelační koeficient Rp:	0.778762429	0.001990502	0.076438131	17942.35312
Střední kvadratická chyba predikce MEP:	0.072553218	0.589800935	0.446663966	83.30996441
Akaikeho informační kritérium:	-24.2553202	-23.7396462	-36.55380969	-47.16997271



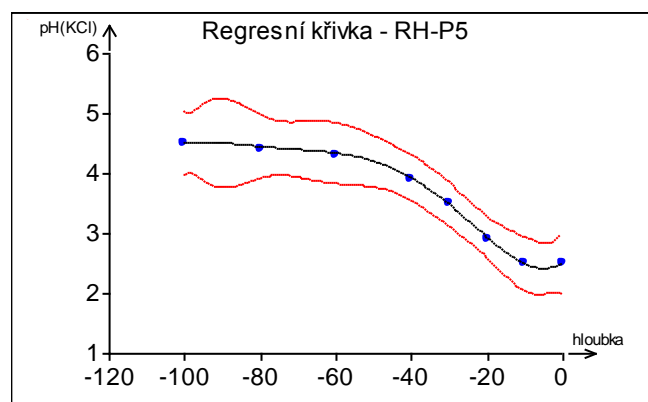
Obr. 19 – Polynom 2. stupně (RH)



Obr. 20 – Polynom 3. stupně (RH)



Obr. 21 – Polynom 4. stupně (RH)



Obr. 22 – Polynom 5. stupně (RH)

I zde byl po vyhodnocení statistických charakteristik regrese pro různé stupně polynomu a grafů těsnosti proložení regresních křivek jako nejvhodnější model zvolen polynom čtvrtého řádu. Možné řešení je opět i varianta $m=5$, ale pro polynom čtvrtého stupně hovoří jednak to, že se jedná o jednodušší variantu a také červené pásy spolehlivosti o něco těsněji kopírují křivku proložení.

B2. Předběžná analýza dat

Proměnná	Průměr	Směr. odch.	Kor.vs. Y	Významnost
hloubka	-42.5	34.94894235	-0.9379999344	0.0005684596952
hloubka ²	2875	3612.18177	0.8240867822	0.01187692022
hloubka ³	-228500	357769.2792	-0.7316786105	0.03909795853
hloubka ⁴	19682500	35362837.94	0.6651889163	0.07184586825

Párové korelace (Xi, Xj)

hloubka - hloubka ²	-0.9624377387	0.0001287891595
hloubka - hloubka ³	0.9113904398	0.001625786657
hloubka - hloubka ⁴	-0.8660418296	0.005422025664
hloubka ² - hloubka ³	-0.9870230082	5.410352899E-006
hloubka ² - hloubka ⁴	0.9629076146	0.0001240604877
hloubka ³ - hloubka ⁴	-0.9934070117	7.129137212E-007

Indikace multikolinearity

Proměnná	Vlas. čísla kor. m.	Podmíněnost kappa	VIF faktor	Vícenás. kor.
Abs	4.683993197E-005	1	1	0
hloubka	1	21349.30513	221.8700904	0.997743884
hloubka ²	0.1524059946	3253.762082	5216.944716	0.999904153
hloubka ³	0.004739508893	101.1852215	12943.73924	0.999961370
hloubka ⁴	3.842807657	82041.27323	3184.565061	0.999842980

Pearsonův párový korelační koeficient ukazuje vysokou korelaci všech párů a velká korelace je indikována i mezi jednotlivými nezávisle proměnnými.

Maximální číslo podmíněnosti kappa je větší než 1000 a indikuje tak multikolinearitu stejně jako kritérium VIF, které dosahuje hodnot podstatně vyšších než 10.

B3. Odhadování parametrů

Proměnná	Odhad	Směr. odch.	Závěr	Pravděpodobnost
Abs	2.472097843	0.08581424613	Významný	9.18483134E-005
hloubka	0.01476606745	0.01433579469	Nevýznamný	0.3787768747
hloubka ²	0.002704949086	0.0006725815511	Významný	0.02761362785
hloubka ³	4.460891715E-005	1.069650764E-005	Významný	0.02509705921
hloubka ⁴	2.105484329E-007	5.368121993E-008	Významný	0.02948225487

Parametry byly nalezeny metodou racionálních hodnotí, parametry jsou statisticky významné kromě nezávisle proměnné x .

B4. Základní statistické charakteristiky

Pro zvolený model mají základní statistické charakteristiky tyto hodnoty (viz také tab. 3):

Vícenásobný korelační koeficient R:	0.9975917122
Koeficient determinace R ² :	0.9951892243
Predikovaný korelační koeficient Rp:	0.07643813094
Střední kvadratická chyba predikce MEP:	0.4466639657
Akaikeho informační kritérium:	-36.55380969

B5. Zhodnocení kvality modelu

Nalezený model polynomické závislosti má tvar

$$y = 2.472 (0.086) + 0.0027 (0.00067) x^2 + 4.46E-5 (1.069E-5) x^3 + 2.105E-7 (5.37E-8) x^4$$

7. Porovnání obou metod

MNČ: $y = 2.472 (0.079) + 0.0027 (0.00062) x^2 + 4.504E-5 (9.95E-6) x^3 + 2.15E-7 (4.99E-8) x^4$

RH: $y = 2.472 (0.086) + 0.0027 (0.00067) x^2 + 4.46E-5 (1.069E-5) x^3 + 2.105E-7 (5.37E-8) x^4$

	MNČ-P4	RH-P4
Vícenásobný korelační koeficient R:	0.9979174	0.997591712
Koeficient determinace R ² :	0.9958391	0.995189224
Predikovaný korelační koeficient Rp:	0.1155862	0.076438131
Střední kvadratická chyba predikce MEP:	0.4074594	0.446663966
Akaikeho informační kritérium:	-37.7147628	-36.5538097

Licenční studium Statistické zpracování experimentálních dat.

Předmět: 2.1 Tvorba lineárních regresních modelů při analýze dat

Přednášející: Prof. RNDr. Milan Meloun, DrSc.

Úloha 3. Validace nové analytické metody

Zadání: Pro stanovení obsahu malondialdehydu je standardně využívána metoda spektrofotometrického stanovení v připraveném výluhu rostlinného materiálu. Porovnání bude provedeno s alternativní metodou, kdy se obsah MDA měří na Spekolu. Bude provedeno ověření výsledků měření získaných alternativní metodou.

Řešení:

1. Návrh modelu

Navržený regresní model přímky je $y = \beta_0 + \beta_1 x$. Bude testována nulová hypotéza $H_0: \beta_0 = 0, \beta_1 = 1$, tj. testování úseku a směrnice přímky.

2. Základní analýza dat

Název sloupce:	spefotom	spekol
Průměr:	6,456047059	6,379552941
Spodní mez:	6,20117194	6,120776811
Horní mez:	6,710922178	6,638329072
Rozptyl:	1,396287069	1,439356131
Směr. odchylka:	1,181645915	1,199731691
Šikmost:	0,007074057623	-0,1480511476
Odchylka od 0:	Nevýznamná	Nevýznamná
Špičatost:	2,610223341	2,572976196
Odchylka od 3:	Nevýznamná	Nevýznamná
Polosuma:	6,4575 6,054	
Modus:	6,633117373	6,546268673
Medián:	6,516	6,436
IS spodní:	6,064496145	6,012905376
IS horní:	6,967503855	6,859094624
Med. směr. odch.:	0,2270449867	0,21275901

Data jsou nezávislá, homogenita přijata, normalita přijata.

Charakteristiky proměnných

Proměnná	Průměr	Směr.odch.	Korelace vs.Y	Významnost
spefotom	6,456047059	1,181645915	0,9927960193	0

3. Odhadování parametrů

Proměnná	Odhad	Směr.odch.	Závěr	Pravděpodobnost
Abs	-0,1280865727	0,08762226366	Nevýznamný	0,1475708653
spefotom	1,007991338	0,01335289456	Významný	0
	Spodní mez	Horní mez		
Abs	-0,302363731	0,04619058557		
Spefotom	0,9814329679	1,034549708		

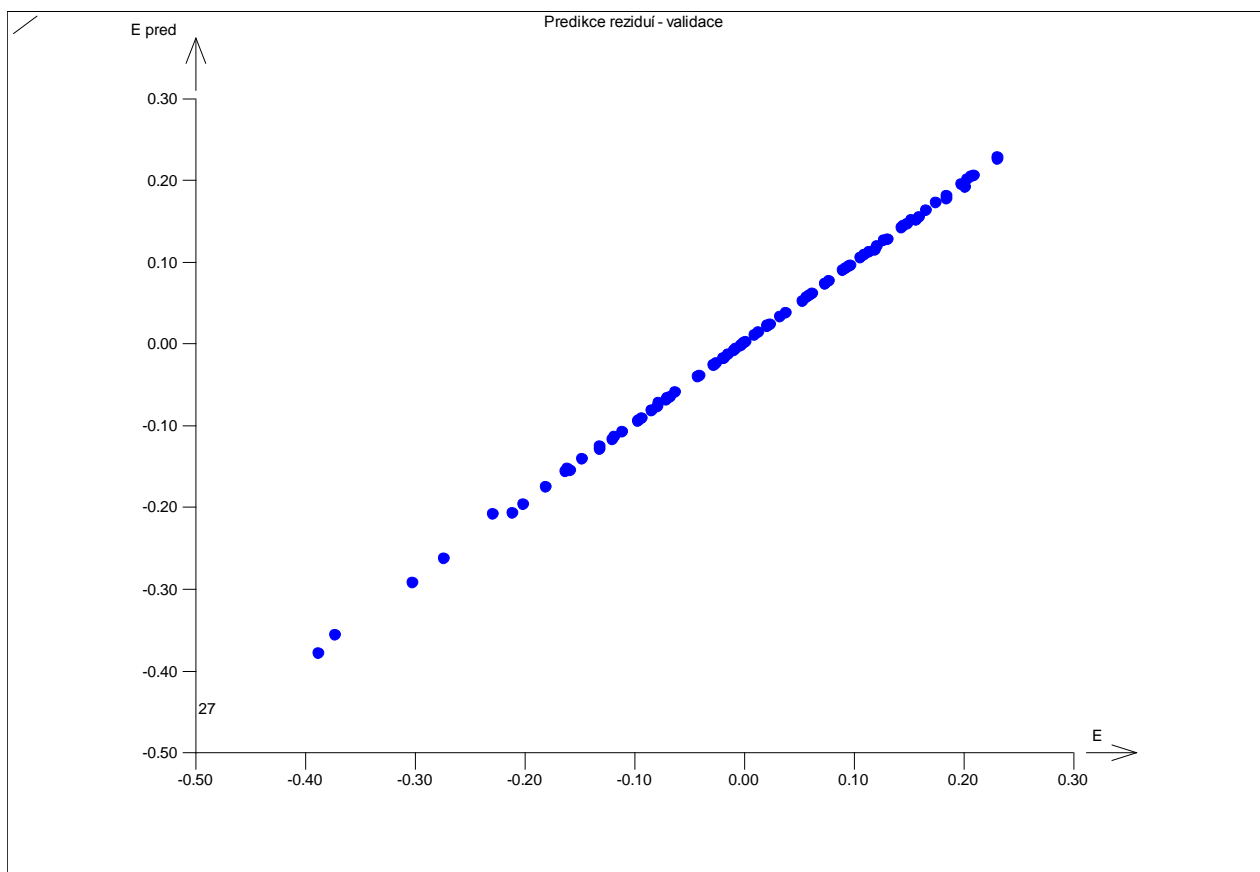
Odhady byly nalezeny metodou nejmenších čtverců. Absolutní člen (úsek) je statisticky nevýznamný, interval spolehlivosti úseku obsahuje nulu. Kvantil Studentova rozdělení $t_{0,95}(85-2) = 1,98896$, $\beta_0 = -0,1281$. Směrnice přímky je naproti tomu statisticky významná, interval spolehlivosti směrnice obsahuje jedničku.

4. Základní statistické charakteristiky

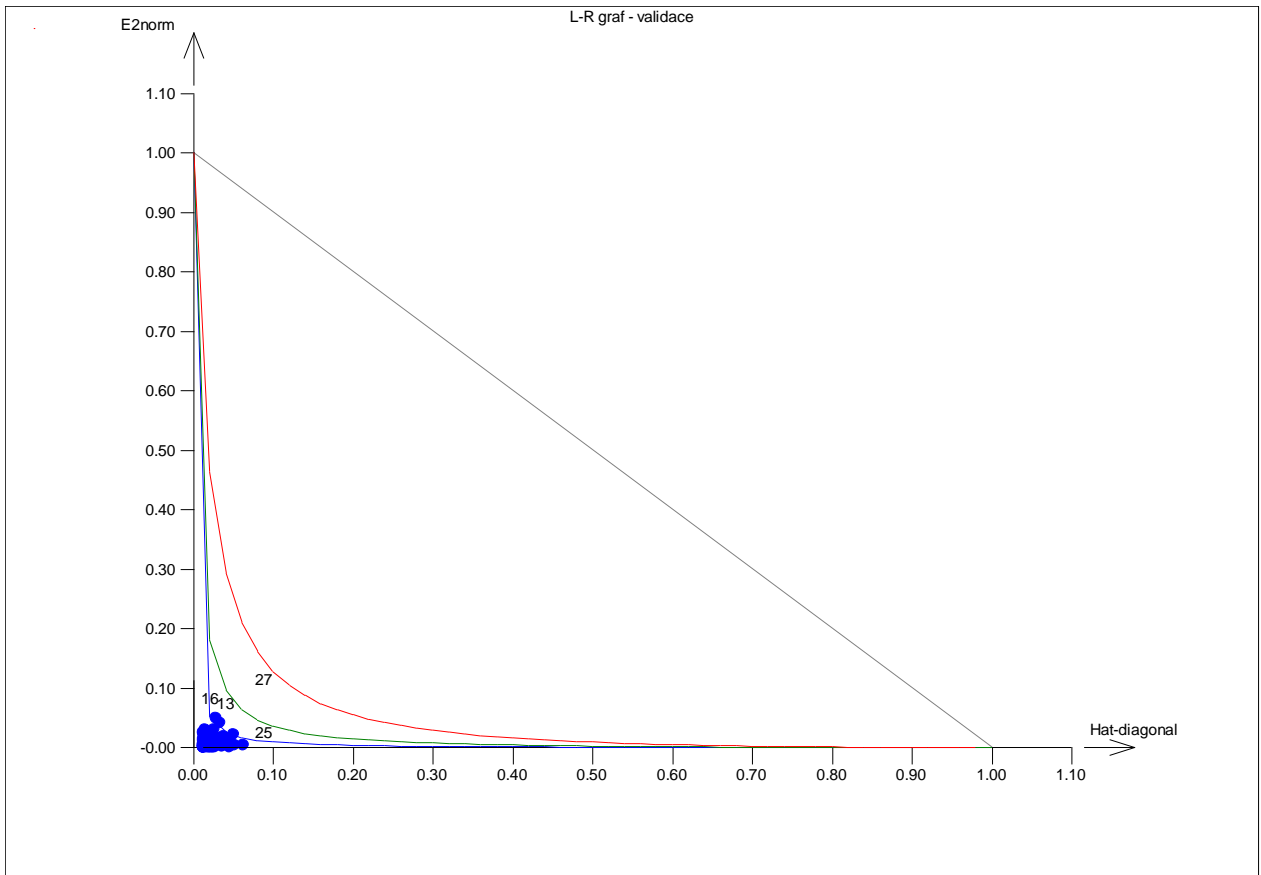
Vícenásobný korelační koeficient r :	0,9927960193
Koeficient determinace R^2 :	0,985643936
Predikovaný korelační koeficient R_p^2 :	0,9694195753
Střední kvadratická chyba predikce MEP :	0,02191800903
Akaikeho informační kritérium AIC :	-326,7538209

Párový korelační koeficient r ukazuje, že navržený lineární model je statisticky významný. Podle koeficientu determinace je vidět, že regresnímu modelu vyhovuje 98,56% bodů, to znamená, že body dobře korespondují s modelem přímky.

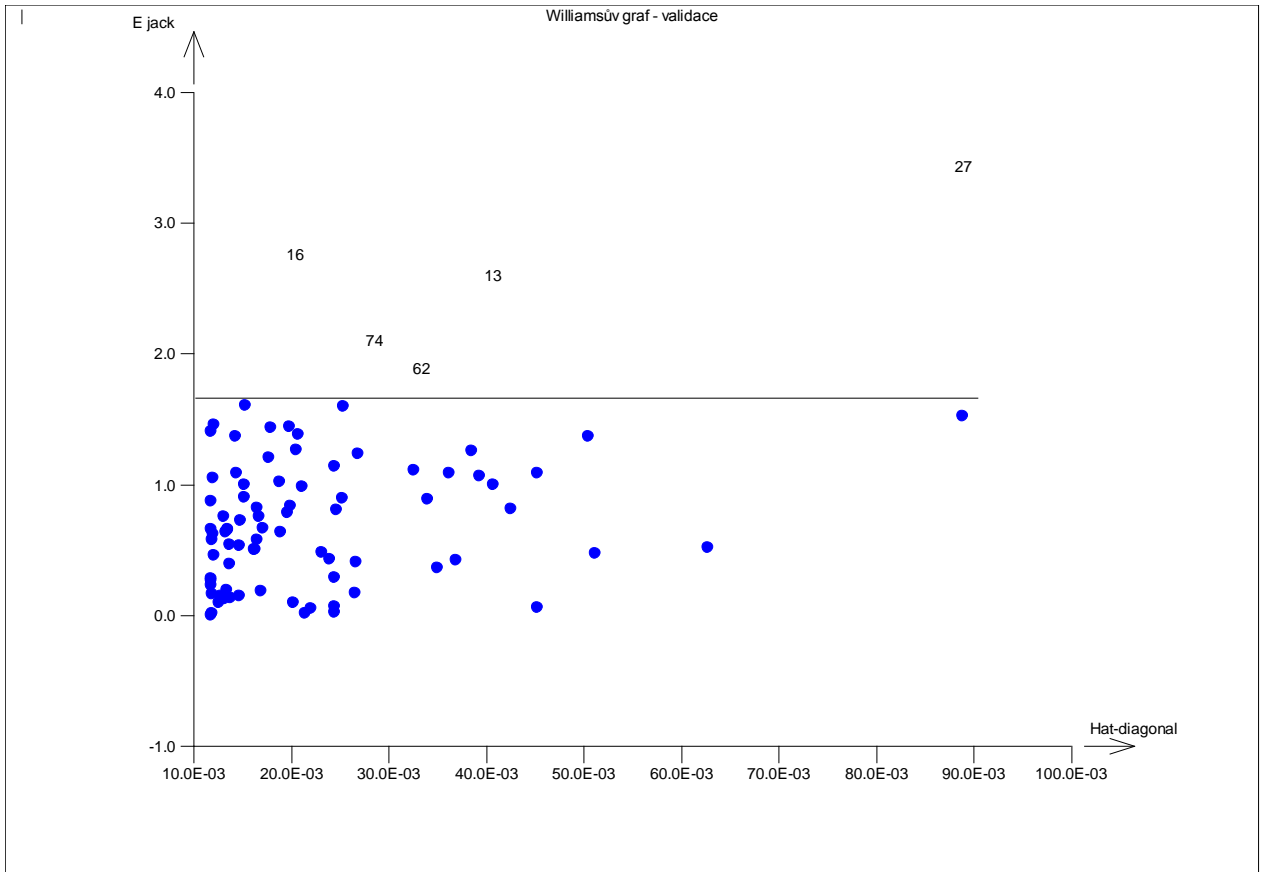
5. Regresní diagnostika



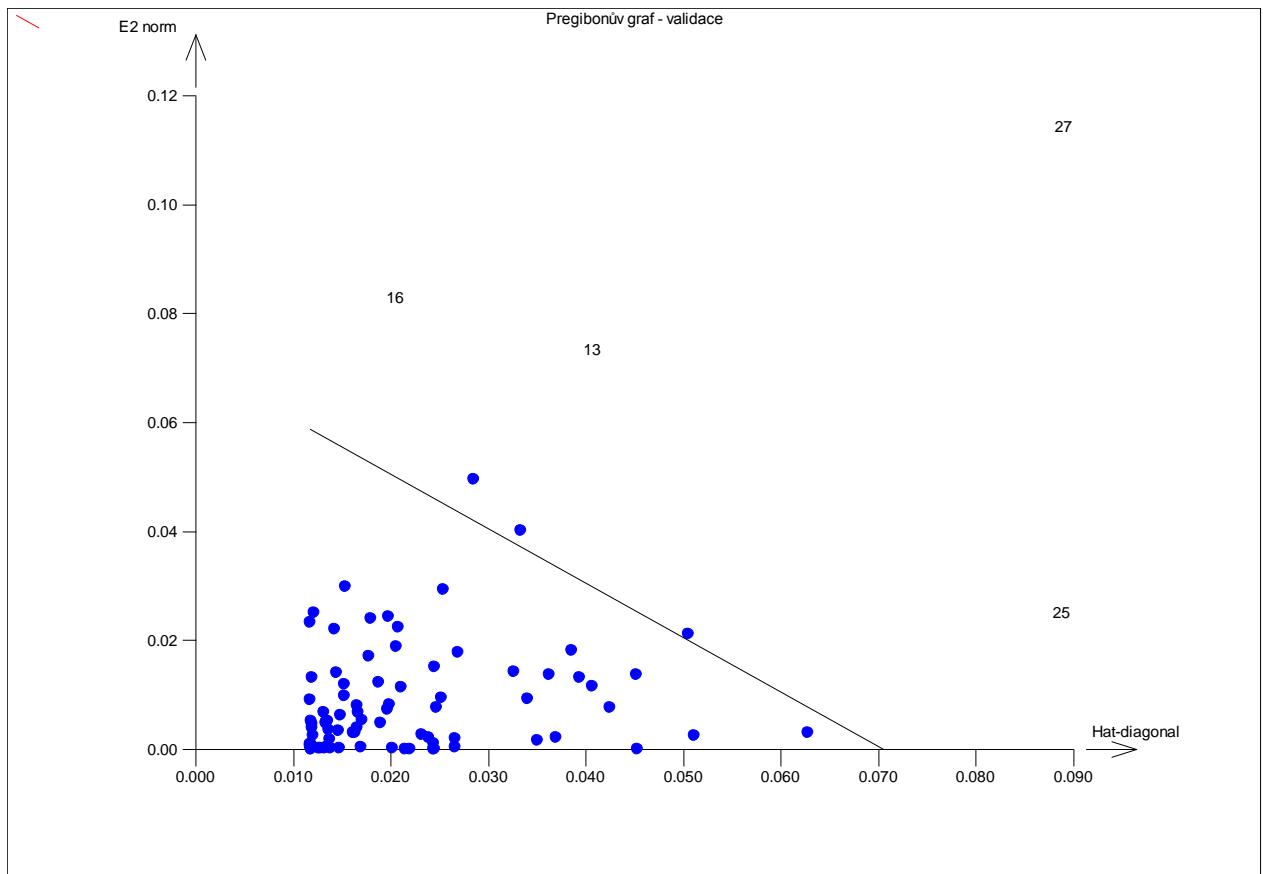
Obr. 1 – Graf predikce reziduí



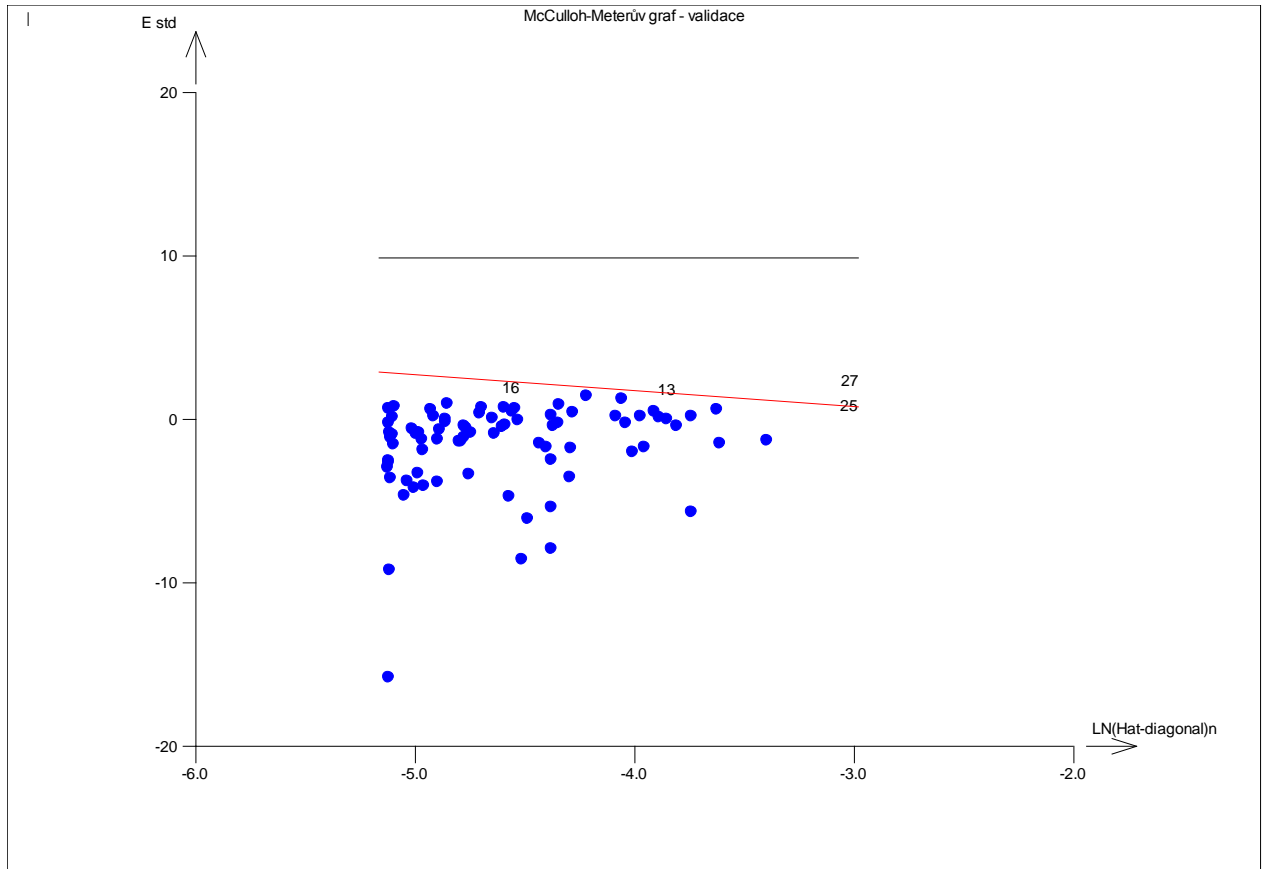
Obr. 2 – L-R graf



Obr. 3 – Williamsův graf



Obr. 4 – Pregibonův graf



Obr. 5 – McCulloh-Meterův graf

Na základě vyhodnocení grafů vlivných bodů byl za odlehlý označen bod číslo 27, který byl jako silně odlehlý indikován na čtyřech z pěti použitých grafů a také třemi metodami analýzy reziduí (Jackknife, Diag(Hii), Diag(H*ii)). Pro zpřesnění modelu bude tedy nutné tento bod odstranit.

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F: 5698.528976
Kvantil F (1-alfa, m-1, n-m): 3.955961007
Pravděpodobnost: 2.891525497E-078
Závěr: Model je významný

Scottovo kritérium multikolinearity (nemá u jednorozměrného regresního modelu smysl)

Hodnota kritéria SC: 0.3331666935
Závěr: Model vykazuje multikolinearitu!

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW: 0.1366966951
Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
Pravděpodobnost: 0.711587282
Závěr: Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB: 7.346306869
Kvantil $\chi^2(1-\alpha,2)$: 5.991464547
Pravděpodobnost: 0.0253962581
Závěr: Rezidua nemají normální rozdělení!

Waldův test autokorelace

Hodnota kritéria WA: 9.477668566
Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
Pravděpodobnost: 0.002079881191
Závěr: Autokorelace je významná

Durbin-Watsonův test autokorelace

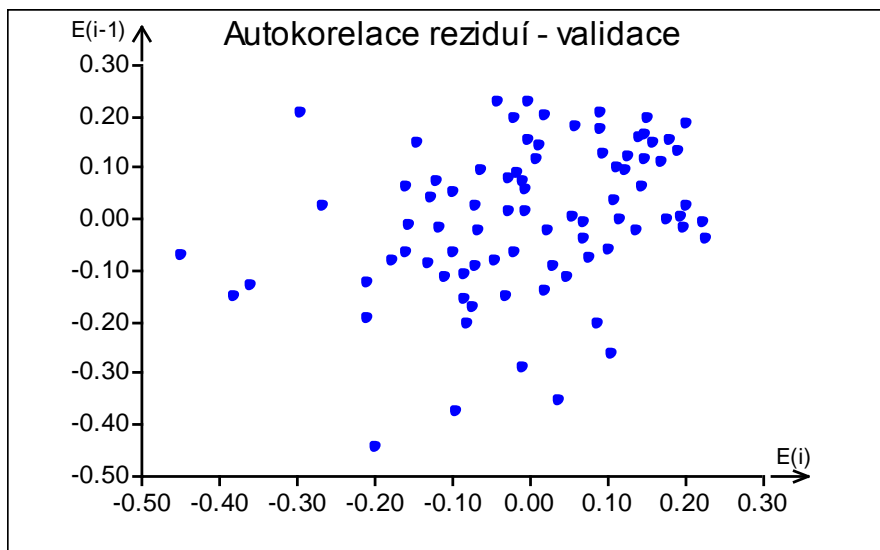
Hodnota kritéria DW: -1
Kritické hodnoty DW 1.6 1.7
Závěr: Rezidua jsou pozitivně autokorelována!

Znaménkový test reziduí

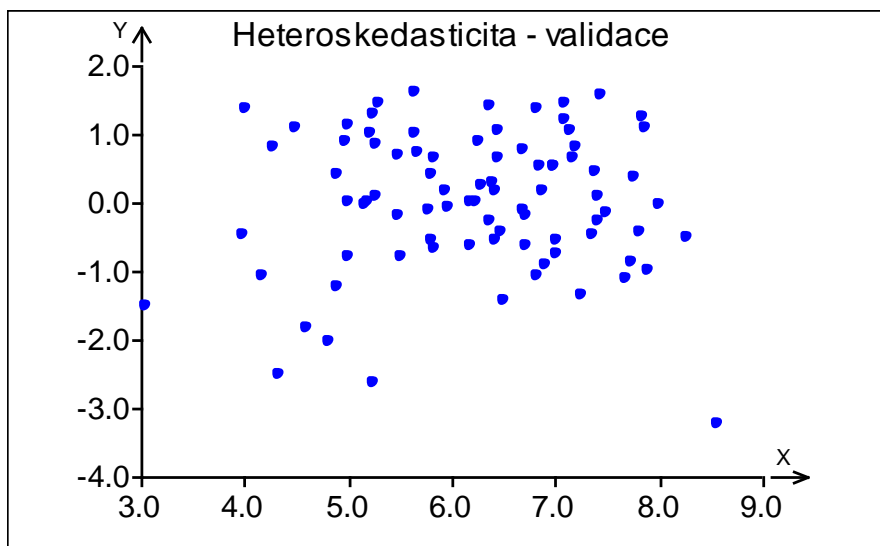
Hodnota kritéria Sg: 1.73650249
Kvantil $N(1-\alpha/2)$: 1.959963999
Pravděpodobnost: 0.0824750245
Závěr: V reziduích není trend.

Fisher-Snedecorův test významnosti modelu potvrzuje, že navržený model je přijat jako významný, Cook-Weisbergův test heteroskedasticity potvrzuje konstantnost rozptylu reziduí. Jarque-Berrův test normality ukazuje, že klasická rezidua nemají normální (Gaussovo) rozdělení. Waldův test autokorelace a Durbin-Watsonův test autokorelace naznačují, že

klasická rezidua jsou autokorelována. Autokorelaci mohou způsobovat také odlehlé body. Znaménkový test reziduí prokazuje dostatečné střídání znaménka klasických reziduí a rezidua tedy nevykazují žádný trend.



Obr. 6 – Graf autokorelace reziduí



Obr. 7 – Graf hereroskedasticity reziduí

6. Konstrukce zpřesněného modelu

Z dat byl odstraněn silně odlehlý bod č. 27 a byly určeny nové odhady parametrů zpřesněného modelu.

Odhad parametrů:

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
Abs	-0.31753668	0.08213662881	Významný	0.0002205875535
spefotom	1.037583062	0.01355411232	Významný	0
Proměnná	Spodní mez	Horní mez		
Abs	-0.4809325999	-0.15414076		
Spefom	1.010619614	1.064546509		

Statistické charakteristiky (v závorce jsou původní hodnoty):

Vícenásobný korelační koeficient r :	0.9918159595	(0,9927960193)
Koeficient determinace R^2 :	0.9836988975	(0,985643936)
Predikovaný korelační koeficient R_p^2 :	0.9714026313	(0,9694195753)
Střední kvadratická chyba predikce MEP :	0.01956863052	(0,02191800903)
Akaikeho informační kritérium AIC :	-316.0391195	(-326,7538209)

Testování regresního tripletu:

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F:	4948.334613
Kvantil F (1-alfa, m-1, n-m):	3.957388322
Pravděpodobnost:	4.447218328E-075
Závěr: Model je významný	

Scottovo kritérium multikolinearity

Hodnota kritéria SC:	0.2549911688
Závěr: Model je korektní.	

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW	2.241423811
Kvantil $\chi^2(1-\text{alfa},1)$:	3.841458829
Pravděpodobnost:	0.1343572131
Závěr: Rezidua vykazují homoskedasticitu.	

Jarque-Berrův test normality

Hodnota kritéria JB:	1.534865022
Kvantil $\chi^2(1-\text{alfa},2)$:	5.991464547
Pravděpodobnost:	0.4642033767
Závěr: Rezidua mají normální rozdělení.	

Waldův test autokorelace

Hodnota kritéria WA:	8.550461327
Kvantil $\chi^2(1-\text{alfa},1)$:	3.841458829
Pravděpodobnost:	0.003454346738
Závěr: Autokorelace je významná	

Durbin-Watsonův test autokorelace

Hodnota kritéria DW:	-1
Kritické hodnoty DW	1.59 1.69
Závěr: Rezidua jsou pozitivně autokorelována!	

Znaménkový test reziduí

Hodnota kritéria Sg:	2.020587177
Kvantil $N(1-\text{alfa}/2)$:	1.959963999
Pravděpodobnost:	0.04332251743
Závěr : V reziduích je trend!	

Po odstranění bodu č. 27 nedošlo ke zlepšení parametrů r , R^2 ani ke snížení rozhodujících kritérií MEP a AIC . Tento model nelze považovat za lepší (zpřesněný).

Z původních dat byly na základě grafů na obr. 1-5 odstraněny i další podezřelé body. Kromě bodu č. 27 to jsou body č. 25, 16 a 13. Pro soubor zbylých 81 bodů byl opět proveden odhad parametrů modelu.

Odhad parametrů:

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
Abs	-0.1316248623	0.08274277075	Nevýznamný	0.1156549642
spefotom	1.01047591	0.01334604831	Významný	0
Proměnná	Spodní mez	Horní mez		
Abs	-0.2963202278	0.03307050308		
Spefotom	0.9839112653	1.037040555		

Statistické charakteristiky (v závorce jsou původní hodnoty):

Vícenásobný korelační koeficient r :	0.9928693695	(0,9927960193)
Koeficient determinace R^2 :	0.9857895849	(0,985643936)
Predikovaný korelační koeficient R_p^2 :	0.9737259212	(0,9694195753)
Střední kvadratická chyba predikce MEP :	0.01577933197	(0,02191800903)
Akaikeho informační kritérium AIC :	-326.2490706	(-326,7538209)

Testování regresního tripletu:

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F:	5480.302766
Kvantil F (1-alfa, m-1, n-m):	3.961892039
Pravděpodobnost:	9.610589521E-075
Závěr: Model je významný	

Scottovo kritérium multikolinearity

Hodnota kritéria SC:	0.3129878955
Závěr : Model je korektní.	

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW:	1.008832316
Kvantil $\chi^2(1-\alpha,1)$:	3.841458829
Pravděpodobnost:	0.3151827424
Závěr: Rezidua vykazují homoskedasticitu.	

Jarque-Berrův test normality

Hodnota kritéria JB:	1.866133438
Kvantil $\chi^2(1-\alpha,2)$:	5.991464547
Pravděpodobnost:	0.3933455784
Závěr: Rezidua mají normální rozdělení.	

Waldův test autokorelace

Hodnota kritéria WA:	3.481482501
Kvantil $\chi^2(1-\alpha,1)$:	3.841458829
Pravděpodobnost:	0.06205911911
Závěr: Autokorelace je nevýznamná	

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : -1

Kritické hodnoty DW 1.59 1.69

Závěr: Rezidua jsou pozitivně autokorelována!

Znaménkový test reziduí

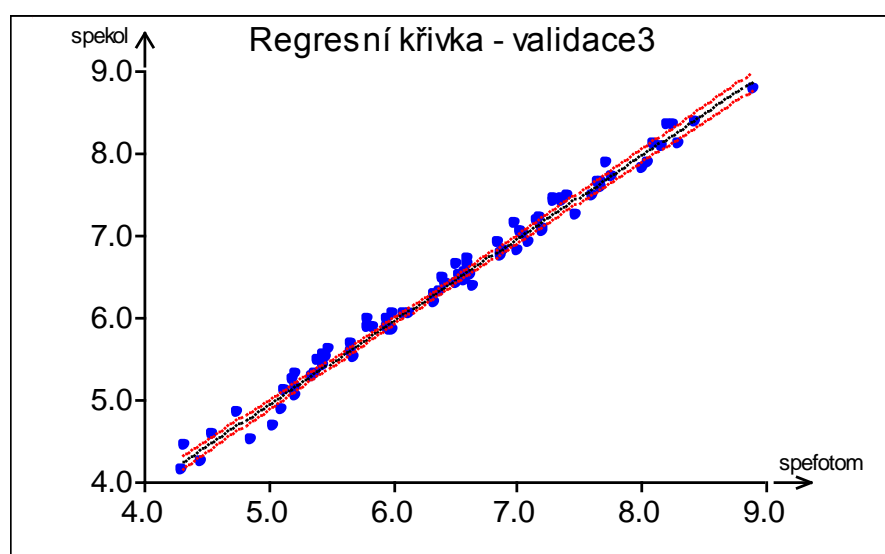
Hodnota kritéria Sg: 0.4459369696

Kvantil $N(1-\alpha/2)$: 1.959963999

Pravděpodobnost: 0.6556427764

Závěr: V reziduích není trend.

Zpřesnění modelu bylo úspěšné – došlo ke snížení rozhodujících kritérií *MEP* a *AIC* a k mírnému zlepšení parametrů r , R^2 , jsou splněny předpoklady metody nejmenších čtverců.



Obr. 8 – Regresní přímka zpřesněného modelu po odstranění podezřelých bodů

7. Zhodnocení kvality modelu

Nalezený regresní model má tvar

$$y = -0.1316 (0.0827) + 1.0105 (0.0133)x$$

Intervalový odhad parametrů úseku ($\text{abs. } \beta_0$) a směrnice (β_1) je (výstup z programu QCExpert):

Proměnná	Spodní mez	Horní mez
β_0	-0.2963202278	0.03307050308
β_1	0.9839112653	1.037040555

Interval spolehlivosti úseku regresní přímky obsahuje nulu, lze tedy tento úsek považovat za nulový. Interval spolehlivosti směrnice obsahuje jedničku, směrnici lze tedy považovat za jednotkovou.

Závěr: Úsek regresní přímky lze považovat za nulový, směrnice není významně odlišná od jedničky. Výsledky alternativní metody se proto statisticky významně neliší od výsledků stanovených metodou standardní.

Licenční studium Statistické zpracování experimentálních dat.

Předmět: 2.1 Tvorba lineárních regresních modelů při analýze dat

Přednášející: Prof. RNDr. Milan Meloun, DrSc.

Úloha 4. Vícerozměrný lineární regresní model

Zadání: Jak obsah základních živin v jehličí smrku v Jizerských horách ovlivňuje zdravotní stav porostů vyjádřený defoliací? Projeví se při nižším obsahu živin zhoršení zdravotního stavu (bude vyšší defoliace) porostů?

Řešení:

1. Návrh modelu

Navržený regresní model přímky je $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$.

y – defoliace

x_1 – obsah dusíku

x_2 – obsah fosforu

x_3 – obsah draslíku

x_4 – obsah vápníku

x_5 – obsah hořčíku

2. Základní analýza dat

	Průměr	Směr. odch.	Korelace vs. Y	Významnost
N	1.492747368	0.1687811102	-0.1472292942	0.1545077432
P	1382.125263	239.7119619	-0.1575649118	0.1272678599
K	6111.631579	919.600573	0.1770528172	0.08608138777
Ca	3905.231579	1342.885275	-0.2113828	0.03975093452
Mg	1039.9	246.9537018	-0.2349189726	0.02193079607

Párové korelace (X_i, X_j)

	Průměr	Směr. odch.
N - P	0.3443452151	0.0006332473665
N - K	0.1341415238	0.1949728897
N - Ca	0.09336695768	0.3681472395
N - Mg	0.1179667989	0.2548877645
P - K	0.02932483828	0.7778668018
P - Ca	0.2725286911	0.0075424687
P - Mg	0.2110138358	0.04010601727
K - Ca	-0.2322429196	0.02353000397
K - Mg	-0.2526873351	0.01349168867
Ca - Mg	0.7093912033	8.881784197E-016

Pearsonův párový korelační koeficient neukazuje u žádného prvku výraznou korelaci. Nejvyšších hodnot dosahuje u y vs. x_4 a y vs. x_5 , tedy defoliace vs. vápník a defoliace vs. hořčík, přičemž tato závislost je záporná.

Párové korelační koeficienty mezi dvojicemi proměnných ukazují nejsilnější korelaci mezi vápníkem a hořčíkem, lineární závislost existuje i mezi fosforem a ostatními prvky (kromě draslíku) a v případě draslík vs. vápník a draslík vs. hořčík je patrný antagonistický vztah.

3. Odhadování parametrů

Odhady byly nalezeny klasickou metodou nejmenších čtverců.

Prom.	Odhad	Směr. odch.	Závěr	Pravděpodobnost
Abs	44.35864715	10.41854446	Významný	5.106177893E-005
N	-6.271099445	5.673665299	Nevýznamný	0.2720088592
P	-0.002912818697	0.004084169516	Nevýznamný	0.4775895892
K	0.001445224096	0.001016512286	Nevýznamný	0.1585943849
Ca	-0.0003414768443	0.0009574817215	Nevýznamný	0.722205031
Mg	-0.004604249239	0.005163587437	Nevýznamný	0.374970787

Prom.	Spodní mez	Horní mez
Abs	23.65722123	65.06007306
N	-17.54455154	5.002352651
P	-0.01102797653	0.005202339135
K	-0.0005745641629	0.003465012356
Ca	-0.00224397263	0.001561018942
Mg	-0.01486418749	0.005655689012

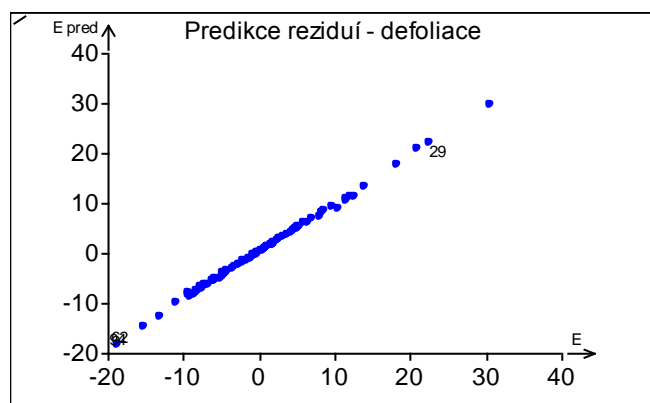
Test významnosti jednotlivých parametrů ukázal, že jediným významným parametrem na zvolené hladině významnosti 0.95 je absolutní člen. Vyloučení absolutního členu by v tomto případě nemělo biologické opodstatnění. Parametr β_0 charakterizuje „zbytkovou“ defoliaci bude-li obsah živin nízký nebo nulový. V takovém případě ovšem bude defoliace dosahovat maxima, tedy 100% nikoli nulové hodnoty.

4. Základní statistické charakteristiky

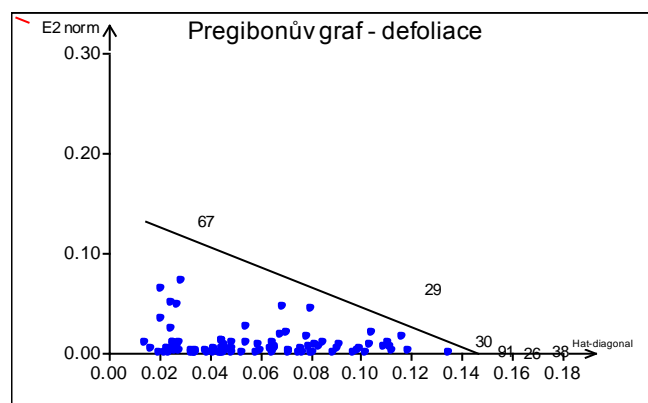
Vícenásobný korelační koeficient r :	0.3133247954
Koeficient determinace R^2 :	0.0981724274
Predikovaný korelační koeficient R_p^2 :	0.0002608755993
Střední kvadratická chyba predikce MEP :	78.09453366
Akaikeho informační kritérium AIC :	414.6637254

Vícenásobný korelační koeficient r ukazuje, že navržený lineární model není statisticky významný. Regresnímu modelu vyhovuje pouze necelých 10% bodů.

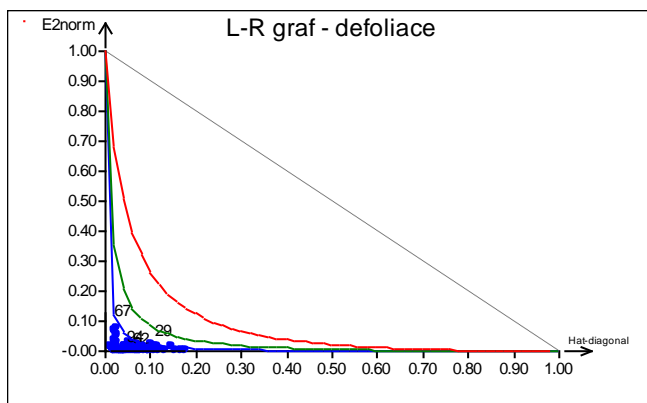
5. Regresní diagnostika



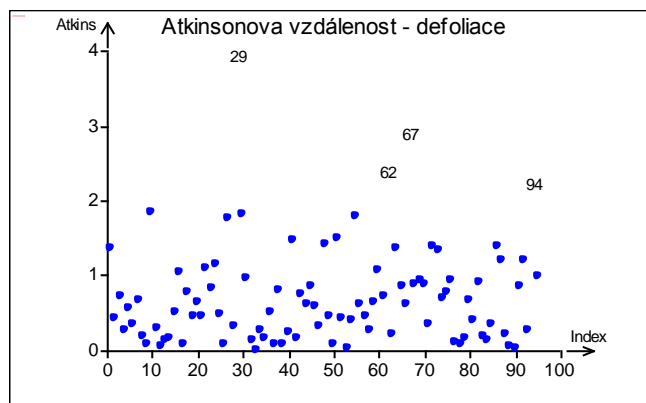
Obr. 1 – Graf predikce reziduí



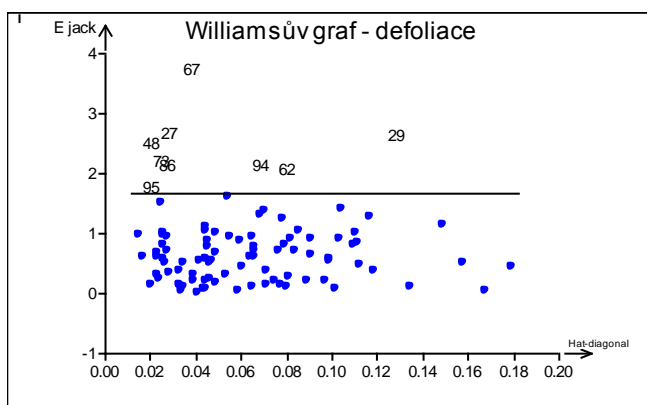
Obr. 2 – Pregibonův graf



Obr. 3 – L-R graf



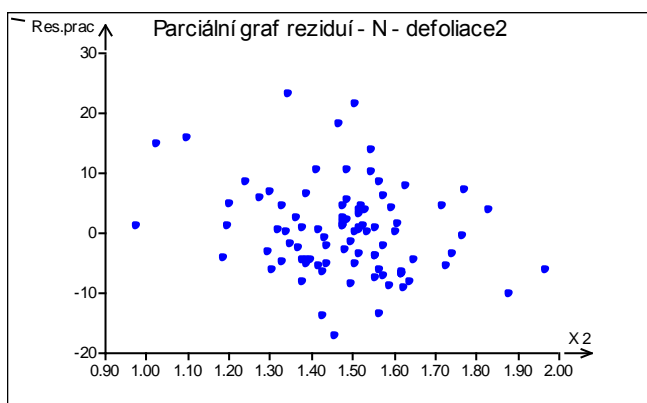
Obr. 4 – Atkinsonova vzdálenost



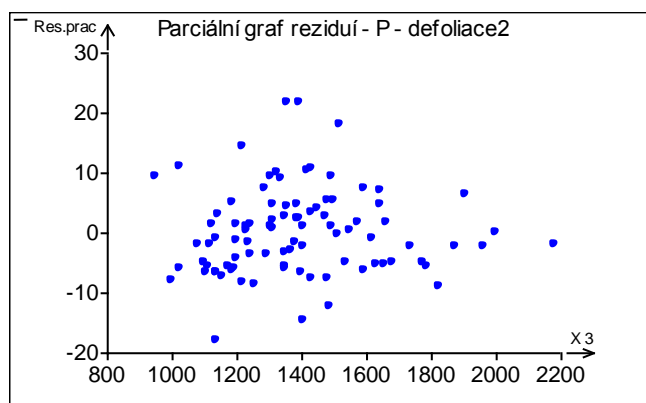
Obr. 5 – Williamsův graf

Na základě analýzy reziduí a vyhodnocení grafů vlivných bodů byly za podezřelé označeny body číslo 29, 62, 67, 94. Tyto body byly jako odlehlé na nejméně třech grafech. Pro zpřesnění modelu budou tyto body odstraněny.

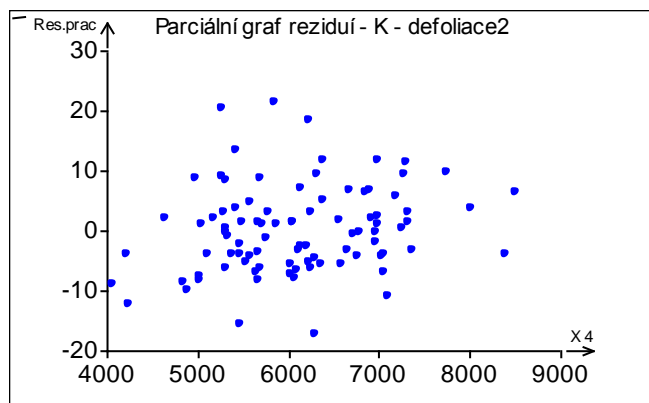
Při kritice modelu jsou sledovány zejména parciální regresní grafy a parciální reziduální grafy. Z grafů na obr. 6-10 vyplývá, že významnější lineární závislost existuje pouze v případě dusíku, přestože určitá závislost je vidět i v případě ostatních sledovaných prvků.



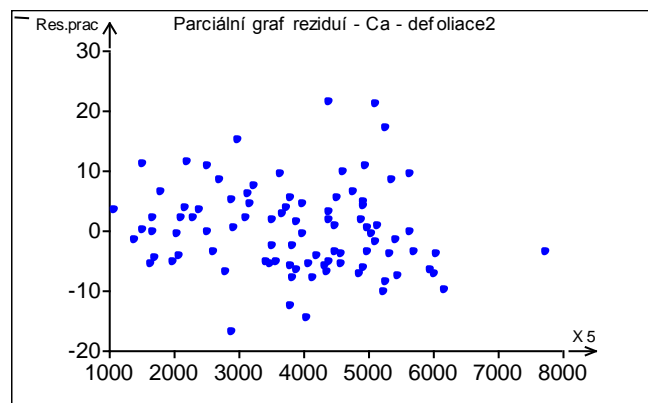
Obr. 6 – Parciální graf reziduí - dusík



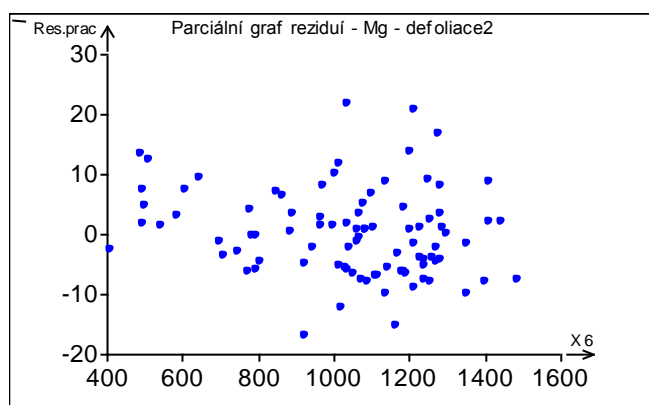
Obr. 7 – Parciální graf reziduí – fosfor



Obr. 8 – Parciální graf reziduí – draslík



Obr. 9 – Parciální graf reziduí – vápník



Obr. 10 – Parciální graf reziduí – hořčík

K testování modelu byla využita také metoda odstranění nezávisle proměnné z modelu a byl sledován vliv odstranění na rozhodující statistické parametry proložení přímky. Výsledky jsou shrnuty v tabulce č. 1. K největším změnám statistických charakteristik došlo při vypuštění proměnné x_1 z modelu, tedy v případě dusíku. Potvrzují se tak závěry grafických metod hodnocení významnosti lineární závislosti.

Tab. 1 – odstranění nezávisle proměnné z modelu

Vypuštěno	r	$R^2 (D)$	R_p^2	MEP	AIC
nic (celý model)	0.4233641	0.1792371	0.0072035	56.520375	369.24533
x1 (N)	0.3676417	0.1351604	0.0026094	58.607395	372.00553
x2 (P)	0.422496	0.1785029	0.0097627	55.659842	367.32671
x3 (K)	0.3856695	0.148741	0.0037678	57.971243	370.56522
x4 (Ca)	0.4163768	0.1733697	0.0083005	56.135355	367.89356
x5 (Mg)	0.4055177	0.1644446	0.0069855	56.600295	368.87082

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F: 1.937697694

Kvantil F (1-alfa, m-1, n-m): 2.316857516

Pravděpodobnost: 0.09592028975

Závěr: Model je nevýznamný

Scottovo kritérium multikolinearity

Hodnota kritéria SC: -0.324605946

Závěr: Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW: 0.3574717695

Kvantil $\chi^2(1-\alpha,1)$: 3.841458829

Pravděpodobnost: 0.5499137137

Závěr: Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB: 15.95264824

Kvantil $\chi^2(1-\alpha,2)$: 5.991464547

Pravděpodobnost: 0.000343499769

Závěr: Rezidua nemají normální rozdělení!

Waldův test autokorelace

Hodnota kritéria WA: 0.01478933538

Kvantil $\chi^2(1-\alpha,1)$: 3.841458829

Pravděpodobnost: 0.9032067756

Závěr: Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW: -1

Kritické hodnoty DW 1.56 1.78

Závěr: Rezidua nejsou autokorelována

Znaménkový test reziduí

Hodnota kritéria Sg: 0.8832332845

Kvantil $N(1-\alpha/2)$: 1.959963999

Pravděpodobnost: 0.3771102415

Závěr: V reziduích není trend.

Fisher-Snedecorův test významnosti modelu hodnotí navržený model jako nevýznamný, Cook-Weisbergův test heteroskedasticity potvrzuje konstantnost rozptylu reziduí. Jarque-Berrův test normality ukazuje, že klasická rezidua nemají normální rozdělení. Waldův a Durbin-Watsonův test autokorelace ukazují, že klasická rezidua nejsou autokorelována. Znaménkový test reziduí prokazuje dostatečné střídání znaménka klasických reziduí a rezidua tedy nevykazují žádný trend.

6. Konstrukce zpřesněného modelu

Z dat byly odstraněny podezřelé body číslo 29, 62, 67 a 94, absolutní člen byl zachován.

Prom.	Odhad	Směr. odch.	Závěr	Pravděpodobnost
Abs	47.05439978	9.205201917	Významný	1.94025721E-006
N	-11.04884984	5.171444321	Významný	0.03551234442
P	0.001019609298	0.003697441143	Nevýznamný	0.7834011065
K	0.001599669185	0.0009001332148	Nevýznamný	0.07912125774
Ca	-0.0006526221401	0.000837212766	Nevýznamný	0.4378398961
Mg	-0.005538955017	0.00447512635	Nevýznamný	0.2192278107

Prom.	Spodní mez	Horní mez
Abs	28.75199222	65.35680733
N	-21.33106662	-0.7666330597
P	-0.006331894267	0.008371112863
K	-0.0001900367982	0.003389375169
Ca	-0.002317225415	0.001011981134
Mg	-0.01443670512	0.003358795086

Statistické charakteristiky (v závorce jsou původní hodnoty):

Vícenásobný korelační koeficient r :	0.4233640789	(0.3133247954)
Koeficient determinace R^2 :	0.1792371433	(0.0981724274)
Predikovaný korelační koeficient R_p^2 :	0.007203478873	(0.0002608755993)
Střední kvadratická chyba predikce MEP :	56.52037517	(78.09453366)
Akaikeho informační kritérium AIC :	369.2453341	(414.6637254)

Testování regresního tripletu:

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F:	3.712438266
Kvantil F (1-alfa, m-1, n-m):	2.321812254
Pravděpodobnost:	0.004352935084
Závěr: Model je významný	

Scottovo kritérium multikolinearity

Hodnota kritéria SC:	-0.2364232595
Závěr: Model je korektní.	

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW:	0.5473999119
Kvantil $\chi^2(1-\alpha,1)$:	3.841458829
Pravděpodobnost:	0.4593820391
Závěr: Rezidua vykazují homoskedasticitu.	

Jarque-Berrův test normality

Hodnota kritéria JB:	9.07235502
Kvantil $\chi^2(1-\alpha,2)$:	5.991464547
Pravděpodobnost:	0.01071428362
Závěr: Rezidua nemají normální rozdělení!	

Waldův test autokorelace

Hodnota kritéria WA:	0.04618364402
Kvantil $\chi^2(1-\alpha,1)$:	3.841458829
Pravděpodobnost:	0.8298423207
Závěr: Autokorelace je nevýznamná	

Durbin-Watsonův test autokorelace

Hodnota kritéria DW:	-1
Kritické hodnoty DW	1.54 1.78
Závěr: Rezidua nejsou autokorelována	

Znaménkový test reziduí

Hodnota kritéria Sg: 0.4520033373

Kvantil $N(1-\alpha/2)$: 1.959963999

Pravděpodobnost: 0.6512665782

Závěr: V reziduích není trend.

Po odstranění čtyř podezřelých bodů došlo ke zlepšení parametrů r , R^2 a ke snížení rozhodujících kritérií MEP a AIC . Tento model již na základě analýzy rozptylu a Fisher-Snedecorova testu lze považovat za významný.

Zpřesnění modelu bylo tedy úspěšné – došlo ke snížení rozhodujících kritérií MEP a AIC a k mírnému zlepšení parametrů r , R^2 .

K dalšímu mírnému zlepšení statistických charakteristik dojde po odstranění parametrů x_2 a x_4 (tab. 2), které podle vypočtených hodnot (viz. např. tab. 1) mají v modelu nejmenší vliv.

Tab. 2 – odstranění nezávisle proměnných z modelu

Vypuštěno	r	D	Rp	MEP	AIC
nic (celý model)	0.4233641	0.1792371	0.0072035	56.520375	369.24533
x2-x5	0.2322686	0.0539487	0.0001527	60.999238	374.17303
x2, x4	0.4162544	0.1732677	0.0114545	55.1522	365.90479
x2, x3, x4	0.376013	0.1413858	0.0071062	56.555891	367.34812

Odhad parametrů regrese bude mít po odstranění x_2 a x_4 následující hodnoty:

Prom.	Odhad	Směr. odch.	Závěr	Pravděpodobnost
Abs	47.37559191	9.031716348	Významný	1.08363974E-006
N	-10.55223694	4.726194533	Významný	0.02813505277
K	0.001633626142	0.000891873135	Nevýznamný	0.07042137708
Mg	-0.007835132709	0.003212142465	Významný	0.01674994079
Prom.	Spodní mez	Horní mez		
Abs	29.4240777	65.32710612		
N	-19.94606034	-1.158413548		
K	-0.0001390682867	0.003406320572		
Mg	-0.01421961367	-0.001450651743		

Zpřesněný model má nyní tvar

$$y = 47.3756 (9.0317) - 10.5522 (4.7262)x_1 + 0.00163 (0.000892) x_3 - 0.00783 (0.00321) x_5$$

Jako nevýznamný byl v tomto zpřesněném modelu vyhodnocen parametr x_3 (draslík), který se podle výsledků v bodu 2 (základní analýza dat) chová vůči hořčíku antagonisticky.

7. Zhodnocení kvality modelu

Nalezený regresní model má tvar

$$y = 47.376 (9.032) - 10.552 (4.726)x_1 - 0.00783 (0.00321) x_5$$

Závěr: Defoliace mladých smrkových porostů v Jizerských horách je funkcí obsahu hořčíku a dusíku. S klesajícím obsahem těchto živin se zvyšuje defoliace.

Licenční studium Statistické zpracování experimentálních dat.

Předmět: 2.2 Kalibrace a limity její přesnosti

Přednášející: Prof. RNDr. Milan Meloun, DrSc.

Úloha 1. Lineární kalibrace

Zadání: Při stanovení fluoridů ve vodách metodou iontové chromatografie je proměřována kalibrační závislost plochy píku y na koncentraci x . Jedná se o lineární kalibrační model. Stanoveny byly také odezvy tří neznámých koncentrací, pro které bude vyčíslen bodový a intervalový odhad.

Vstupní data:

koncentrace	odezva	naměřené hodnoty
mg.l^{-1}	mV.s^{-1}	mV.s^{-1}
2.5	3427.3863	765.324
2	2733.0536	1680.0247
1	1363.444	2904.6831
0.75	1028.0197	
0.5	678.6232	

Řešení:

1. Návrh modelu

Pro tento typ kalibrace byl zvolen lineární model, který je pro tento případ nejvhodnější (viz výsledky a obr. 1 a 2).

Název úlohy: F-1 (stanovení fluoridů, lineární kalibrační model)

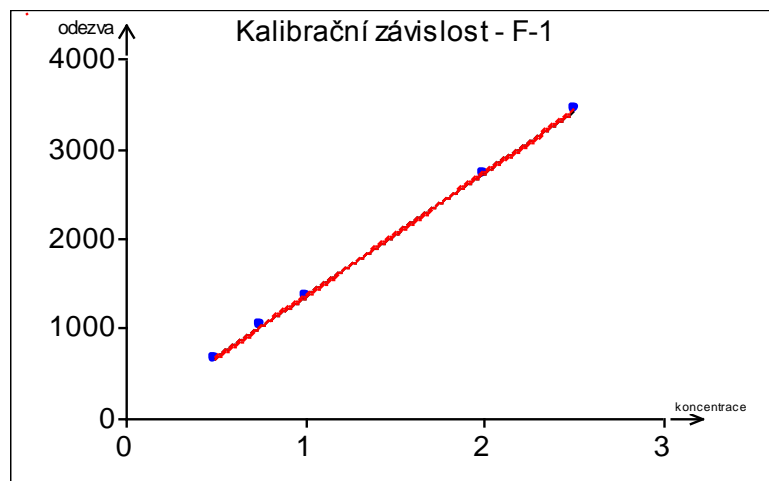
Počet dat: 5

Hladina významnosti: 0.05

Volba kalibračního modelu: Manuální

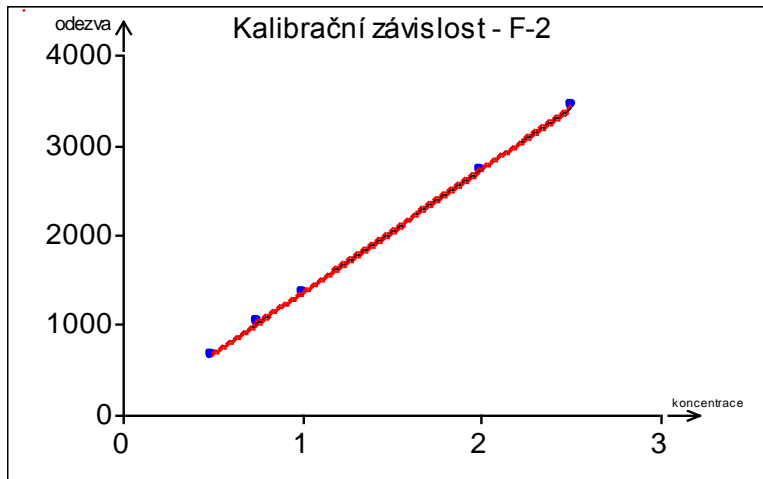
Použitý kalibrační model: Lineární

Vhodnost použitého modelu: Vyhovuje



Obr. 1 – Lineární kalibrační závislost

Název úlohy: F-2 (stanovení fluoridů, kvadratický kalibrační model)
 Počet dat: 5
 Hladina významnosti: 0.05
 Volba kalibračního modelu: Manuální
 Použitý kalibrační model: Kvadratický
 Vhodnost použitého modelu: Nevyhovuje



Obr. 2 – Kvadratická kalibrační závislost

2. Statistická analýza reziduí

Reziduální součet čtverců: 69.49089855

Průměrné absolutní reziduum: 3.458087627

Korelační koeficient: 0.9999937422

Číslo měření	Naměřené X	Naměřené Y	Vypočítané Y	Reziduum	Váha
1	2.5	3427.3863	3423.800942	3.585358051	1
2	2	2733.0536	2737.846341	-4.792741102	1
3	1	1363.444	1365.937139	-2.493139407	1
4	0.75	1028.0197	1022.959839	5.059861017	1
5	0.5	678.6232	679.9825386	-1.359338559	1

3. Kalibrační meze přesnosti

c..kritická

d..detekce

q..kvantifikace

	Yc	Yd	Yq
Metoda K*Sigma, ACS	3.460972378	12.89400704	22.32704171

	Xc	Xd	Xq
Metoda K*Sigma, ACS	0.006875844	0.013751689	0.020627534

	Yq(10sigma)	Xq(10sigma)
Metoda K*Sigma, ACS	42.15654963	0.02062753422

4. Kalibrační tabulka

Vzorek	Zpětný odhad	Spodní mez	Horní mez	Nepřímý odhad	Naměřené hodnoty
1	0.5622063	0.5533965	0.5709162	0	765.324
2	1.2289419	1.2227435	1.2351249	0	1680.0247
3	2.1216091	2.1129900	2.1303261	0	2904.6831

5. Závěr

Limita detekce je 12.894, nejnižší naměřená hodnota 765.324; všechny neznámé vzorky jsou nad limitou stanovení. Bodový a intervalový odhad neznámé koncentrace u tří měřených vzorků je uveden v kalibrační tabulce v bodu 4.

Parametry kalibračního modelu:

Parametr	Odhad	Sm. odchylka	Spodní mez	Horní mez
Abs.	-5.9720622	4.352371816	-19.82325189	7.879127317
X	1371.90920	2.802156122	1362.99149	1380.826913

Významnost absolutního členu

Hodnota	Spodní mez	Horní mez	Závěr
-5.972062288	-19.82325189	7.879127317	Nevýznamný

Validace směrnice

Hodnota	Spodní mez	Horní mez	Směrnice=1
1371.909202	1362.99149	1380.826913	Ne

Kalibrační model: $y = 1371.91 (2.802) x$

Licenční studium Statistické zpracování experimentálních dat.

Předmět: 2.2 Kalibrace a limity její přesnosti

Přednášející: Prof. RNDr. Milan Meloun, DrSc.

Úloha 2. Nelineární kalibrace

Zadání: Při stanovení obsahu dusičnanů ve vzorcích vod metodou iontové chromatografie je proměřována kalibrační závislost plochy píku y na koncentraci x . Jedná se o nelineární kalibrační model. Stanoveny byly také odezvy čtyř neznámých koncentrací, pro které bude vyčíslen bodový a intervalový odhad.

Vstupní data:

koncentrace	odezva	naměřené hodnoty
mg.l^{-1}	mV.s^{-1}	mV.s^{-1}
10	4272.439	698.5016
8	3332.1	1417.226
5	1983.3163	1739.62
2	765.375	3662.9914
1	375.9454	

Řešení:

1. Návrh modelu

Pro tento typ kalibrace byl zvolen nelineární (kvadratický) model, který je pro tento případ vhodnější než lineární (viz výsledky a obr. 1 a 2).

Název úlohy: NO₃-1 (stanovení dusičnanů, lineární kalibrační model)

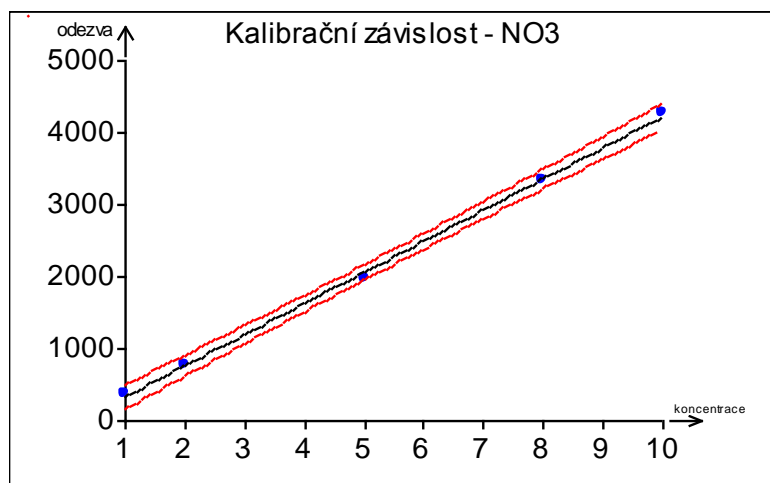
Počet dat: 5

Hladina významnosti: 0.05

Volba kalibračního modelu: Manuální

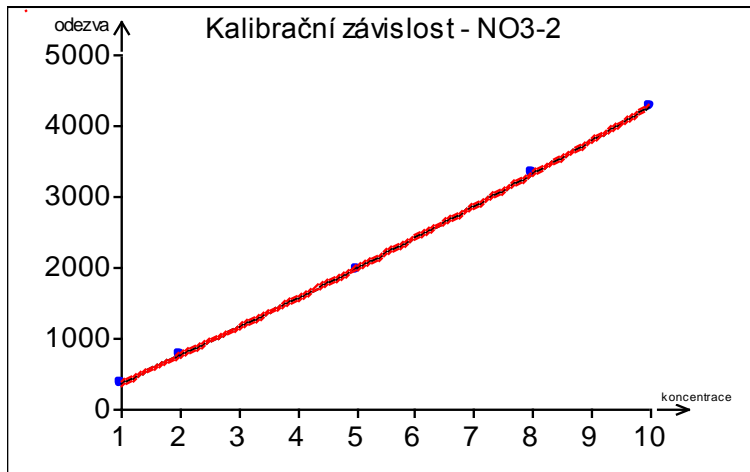
Použitý kalibrační model: Lineární

Vhodnost použitého modelu: Nevyhovuje



Obr. 1 – Lineární kalibrační závislost

Název úlohy: NO3-2 (stanovení dusičnanů, kvadratický kalibrační model)
 Počet dat: 5
 Hladina významnosti: 0.05
 Volba kalibračního modelu: Manuální
 Použitý kalibrační model: Kvadratický
 Vhodnost použitého modelu: Vyhovuje



Obr. 2 – Kvadratická kalibrační závislost

2. Statistická analýza reziduí

Reziduální součet čtverců: 109.1440239
 Průměrné absolutní reziduum: 4.119824391
 Korelační koeficient: 0.9999950363

Číslo měření	Naměřené X	Naměřené Y	Vypočítané Y	Reziduum	Váha
1	10	4272.439	4275.494974	-3.055973798	1
2	8	3332.1	3325.082604	7.017395994	1
3	5	1983.3163	1989.541343	-6.225043189	1
4	2	765.375	762.092835	3.282164984	1
5	1	375.9454	376.963944	-1.01854399	1

3. Kalibrační meze přesnosti

c..kritická

d..detekce

q..kvantifikace

	Yc	Yd	Yq
Metoda K*Sigma, ACS	18.32417575	32.80299268	47.28180961
	Xc	Xd	Xq
Metoda K*Sigma, ACS	0.039414215	0.078777739	0.1180907658
	Yq(10sigma)	Xq(10sigma)	
Metoda K*Sigma, ACS	77.71823067	0.1180907658	

4. Kalibrační tabulka

Vzorek	Zpětný odhad	Spodní mez	Horní mez	Nepřímý odhad	Naměřené hodnoty
1	1.837010441	1.772294053	1.899930768	0	698.5016
2	3.633967482	3.567716197	3.701452454	0	1417.226
3	4.410037113	4.338274407	4.482668114	0	1739.62
4	8.722743806	8.669211668	8.776772193	0	3662.9914

5. Závěr

Limita detekce je 32.803, nejnižší naměřená hodnota 698.5; všechny neznámé vzorky jsou nad limitou stanovení. Bodový a intervalový odhad neznámé koncentrace u čtyř měřených vzorků je uveden v kalibrační tabulce v bodu 4.

Parametry kalibračního modelu:

Parametr	Odhad	Sm. odchylka	Spodní mez	Horní mez
Abs.	3.84535881	9.540002972	-37.20196102	44.89267864
X	367.113432	4.679378434	346.9796919	387.2471726
X ²	6.00515292	0.4225568118	4.187037705	7.823268145

Významnost absolutního členu

Hodnota	Spodní mez	Horní mez	Závěr
3.845358814	-37.20196102	44.89267864	Nevýznamný

Validace směrnice

Hodnota	Spodní mez	Horní mez	Směrnice=1
367.1134323	346.9796919	387.2471726	Ne

Kalibrační model: $y = 367.113 (9.54) x + 6.005 (0.42) x^2$

Licenční studium Statistické zpracování experimentálních dat.

Předmět: 2.2 Kalibrace a limity její přesnosti

Přednášející: Prof. RNDr. Milan Meloun, DrSc.

Úloha 3. Rozlišení mezi lineární a nelineární kalibrací

Zadání: Při stanovení obsahu dusičnanů ve vzorcích vod metodou iontové chromatografie je proměřována kalibrační závislost plochy píku y na koncentraci x . Úkolem je nalézt kalibrační model. Stanoveny byly také odezvy čtyř neznámých koncentrací, pro které bude vyčíslen bodový a intervalový odhad.

Vstupní data:

koncentrace	odezva	naměřené hodnoty
mg.l ⁻¹	mV.s ⁻¹	mV.s ⁻¹
10	4272.439	698.5016
8	3332.1	1417.226
5	1983.3163	1739.62
2	765.375	3662.9914
1	375.9454	

Řešení: (ADSTAT)

1. Návrh modelu

A – lineární

Reziduální součet čtverců, RSC: 1.1131E+04

Průměr absolutních hodnot reziduí, Me: 3.9979E+01

Odhad reziduálního rozptylu, $s^2(e)$: 3.7103E+03

Odhad směrodatné odchylky reziduí, $s(e)$: 6.0912E+01

B – kvadratický spline

Konstantní uzlové intervaly

1 uzel

Reziduální součet čtverců, RSC: 6.5885E+01

Průměr absolutních hodnot reziduí, Me: 3.2369E+00

Odhad reziduálního rozptylu, $s^2(e)$: 6.5885E+01

Odhad směrodatné odchylky reziduí, $s(e)$: 8.1170E+00

2 uzly

Reziduální součet čtverců, RSC: 8.3089E-21

Průměr absolutních hodnot reziduí, Me: 3.0661E-11

Odhad reziduálního rozptylu, $s^2(e)$: 0.0000E+00

Odhad směrodatné odchylky reziduí, $s(e)$: 0.0000E+00

Příliš mnoho uzlů, statistická analýza vynechána.

Zvolen nelineární model s jedním uzlovým bodem.

2. Statistická analýza reziduí

Bod	Měřená hodnota	Predikovaná hodnota	Absolutní reziduum	Relativní reziduum
i	$y_{exp[i]}$	$y_{vyp[i]}$	$e_{[i]}$	$e_{r,[i]}$
1	3.7595E+02	3.7934E+02	3.3941E+00	8.9473E-01
2	7.6538E+02	7.5968E+02	-5.6913E+00	-7.4917E-01
3	1.9833E+03	1.9873E+03	3.9573E+00	1.9913E-01
4	3.3321E+03	3.3297E+03	-2.4010E+00	-7.2110E-02
5	4.2724E+03	4.2732E+03	7.4106E-01	1.7342E-02
	Reziduální součet čtverců, RSC:		6.5885E+01	
	Průměr absolutních hodnot reziduí, Me:		3.2369E+00	
	Průměr relativních reziduí, Mer[%]:		0.386	
	Odhad reziduálního rozptylu, $s^2(e)$:		6.5885E+01	
	Odhad směrodatné odchylky reziduí, $s(e)$:		8.1170E+00	

3. Analýza derivací a integrálů kalibračního grafu

Bod	Predikovaná hodnota	První derivace	Druhá derivace	Integrál
i	$Y_{vyp[i]}$	$der1_{[i]}$	$der2_{[i]}$	$int_{[i]}$
1	3.7934E+02	3.7313E+02	1.4426E+01	0.0000E+00
2	7.5968E+02	3.8756E+02	1.4426E+01	5.6831E+02
3	1.9873E+03	4.3084E+02	1.4426E+01	4.6563E+03
4	3.3297E+03	4.6211E+02	9.6261E+00	1.2609E+04
5	4.2732E+03	4.8137E+02	9.6261E+00	2.0206E+04

4. Kalibrační meze přesnosti

Kritická úroveň,	y_c : 2.141390E+02	x_c : 5.534033E-01
Limita detekce,	y_d : 3.201996E+02	x_d : 8.410151E-01

5. Kalibrační tabulka

Měřená hodnota	Inverzní odhad	Konfidenční interval	
		dolní mez	horní mez
$y_{exp[i]}$	$x_{vyp[i]}$	$LDx_{vyp[i]}$	$LHx_{vyp[i]}$
6.985016E+02	1.841668E+00	1.660214E+00	2.033772E+00
1.417226E+03	3.646196E+00	3.389558E+00	3.903961E+00
1.739620E+03	4.419538E+00	4.171206E+00	4.648356E+00
3.662991E+03	8.715896E+00	8.528717E+00	-----

6. Závěr

Model je kvadratický spline s jedním uzlovým bodem.

Limita detekce je 320.2, nejnižší naměřená hodnota 698.5; všechny neznámé vzorky jsou nad limitou stanovení. Bodový a intervalový odhad neznámé koncentrace u čtyř měřených vzorků je uveden v kalibrační tabulce v bodu 5.