

Univerzita Pardubice

Fakulta chemicko-technologická
Katedra analytické chemie

Licenční studium
Statistické zpracování dat

Semestrální práce
Tvorba lineárních regresních modelů

Obsah

1	POROVNÁNÍ REGRESNÍCH KŘIVEK	3
1.1	Křemičitany – 410nm.....	3
1.1.1	Kritika dat.....	3
1.1.2	Kritika modelu a metody	7
1.1.3	Opravený model (bez bodu 6).....	10
1.1.4	Závěr regresní diagnostiky pro křemičitany – 410nm.....	10
1.2	Křemičitany – 690nm.....	11
1.2.1	Kritika dat.....	11
1.2.2	Kritika modelu a metody	15
1.2.3	Opravený mode (bez bodu 5)	17
1.2.4	Závěr regresní diagnostiky pro křemičitany – 690nm.....	18
1.3	Porovnání dvou regresních přímek	18
1.3.1	Test shodnosti rozptylů	18
1.3.2	Chowův test shody dvou lineárních modelů	19
1.3.3	Závěr.....	20
2	URČENÍ STUPNĚ POLYNOMU.....	21
2.1	Metoda nejmenších čtverců.....	21
2.2	Metoda racionálních hodnotí	23
2.3	Závěr.....	24
3	VALIDACE NOVÉ ANALYTICKÉ METODY.....	25
3.1	Kritika dat.....	25
3.2	Kritika modelu a metody	28
3.3	Opravený model (bez bodu 9).....	29
3.4	Závěr.....	30
4	VÍCEROZMĚRNÝ LINEÁRNÍ REGRESNÍ MODEL	31
4.1	Kritika dat.....	31
4.2	Kritika modelu a metody	35
4.3	Opravený model (bez abs. členu a proměnných TZL, CO a NH ₃)	36
4.4	Závěr.....	37
	LITERATURA A SOFTWARE.....	37

1 Porovnání regresních křivek

Porovnání regresních křivek závislosti absorbance na koncentraci křemičitanů (mg/l SiO_2) ve vzorcích vody. Vzorky křemičitanů jsou stanovovány fotometricky dvěma rozdílnými postupy:

1. vzorek vybarven na žluto (H_2SO_4 , $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}$, kyselina citronová), absorbance stanovována při vlnové délce 410 nm, šířka kyvety 5cm
2. vzorek vybarven na modro (H_2SO_4 , $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}$, kyselina citronová, kyselina askorbová), absorbance stanovována při vlnové délce 690 nm, šířka kyvety 1cm

Předpokládáme přímkovou závislost absorbance na koncentraci.

Data

	c(SiO ₂)	A (410nm, 5cm)
1	0	0.042
2	1	0.115
3	5	0.348
4	10	0.594
5	15	0.810
6	20	0.924

	c(SiO ₂)	A (690nm, 1cm)
1	0	0.012
2	2	0.207
3	4	0.402
4	6	0.597
5	8	0.804
6	10	0.986

1.1 Křemičitany – 410nm

1.1.1 Kritika dat

Testování úseku a směrnice

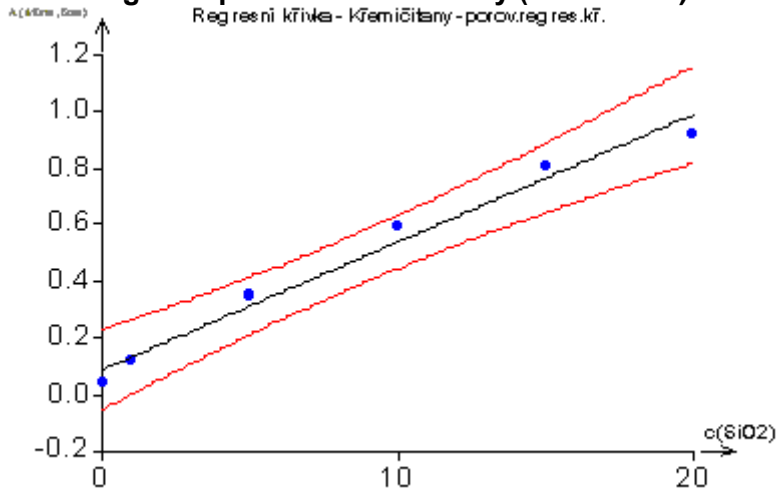
Odhady parametrů

Parametr	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
abs. člen	0.088140682	0.035903056	Nevýznamný	0.070069978	-0.011542182	0.187823547
směrnice	0.045179528	0.003209128	Významný	0.000147729	0.036269559	0.054089496

Hladina významnosti $\alpha = 0,05$ (95% jistota)

Absolutní člen lineárního regresního modelu je na základě testu intervalem spolehlivosti vyhodnocen jako nevýznamný (nulový), protože interval vymezený spodní a horní mezí obsahuje 0 resp. p-hodnota testu je vyšší než zadaná hladina významnosti 0,05. Hodnota směrnice je vyhodnocena jako významná, protože interval spolehlivosti neobsahuje 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05. Zároveň interval spolehlivosti neobsahuje ani 1, to znamená, že směrnice není jednotková.

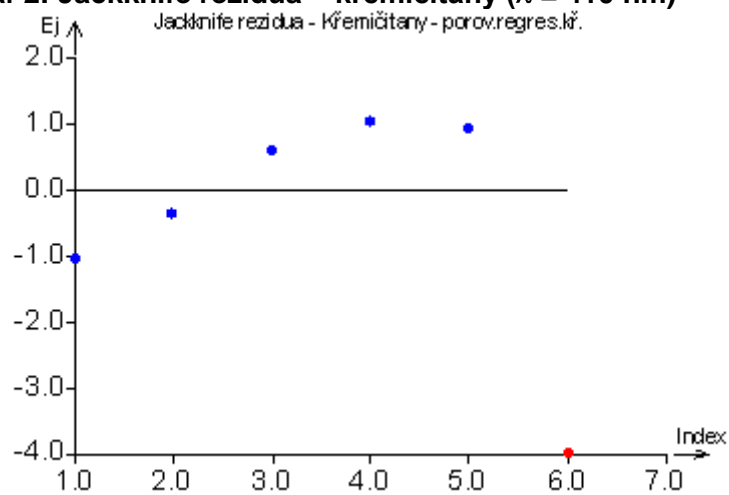
Graf 1: Regresní přímka – křemičitany ($\lambda = 410 \text{ nm}$)



Konfidenční pás (vymezený okrajovými křivkami) je poměrně široký a je patrný mírný nelineární trend v bodech. Bylo by vhodné testovat i jiný model.

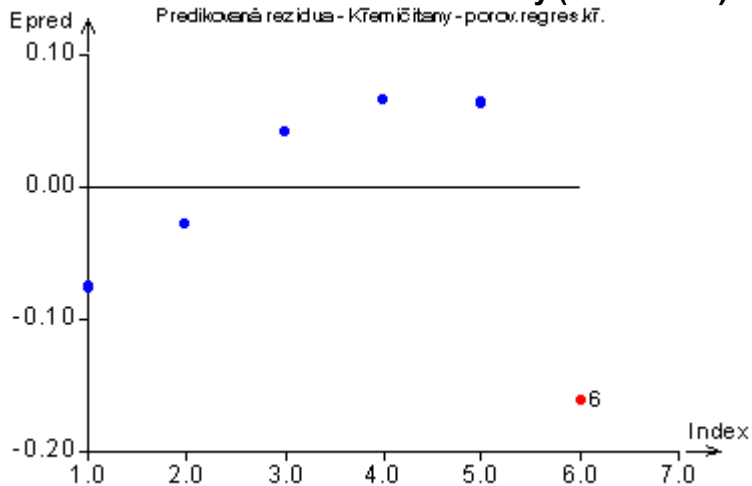
Testování vlivných bodů diagnostickými grafy

Graf 2: Jackknife rezidua – křemičitany ($\lambda = 410 \text{ nm}$)



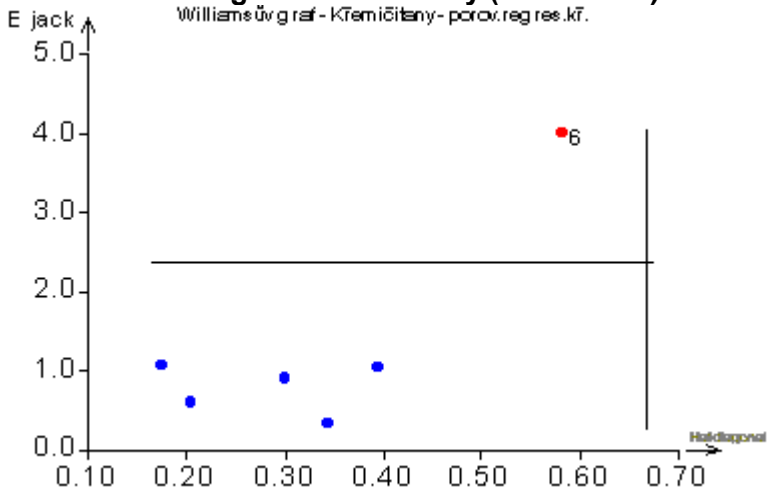
V grafu Jackknife reziduí je bod 6 identifikován jako podezřelý z odlehlosti. (vzdálen od ostatních)

Graf 3: Predikovaná rezidua – křemičitany ($\lambda = 410 \text{ nm}$)



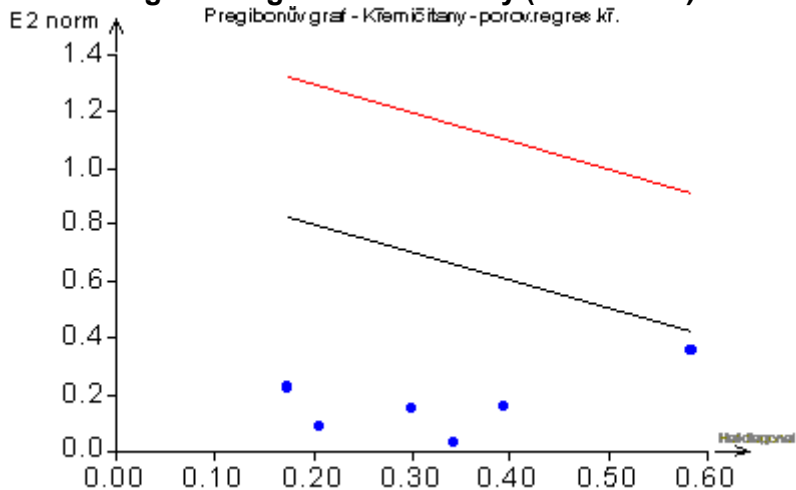
V grafu predikovaných reziduí je bod 6 identifikován jako podezřelý z odlehlosti. (vzdálen od ostatních)

Graf 4: Williamsův graf – křemičitany ($\lambda = 410 \text{ nm}$)



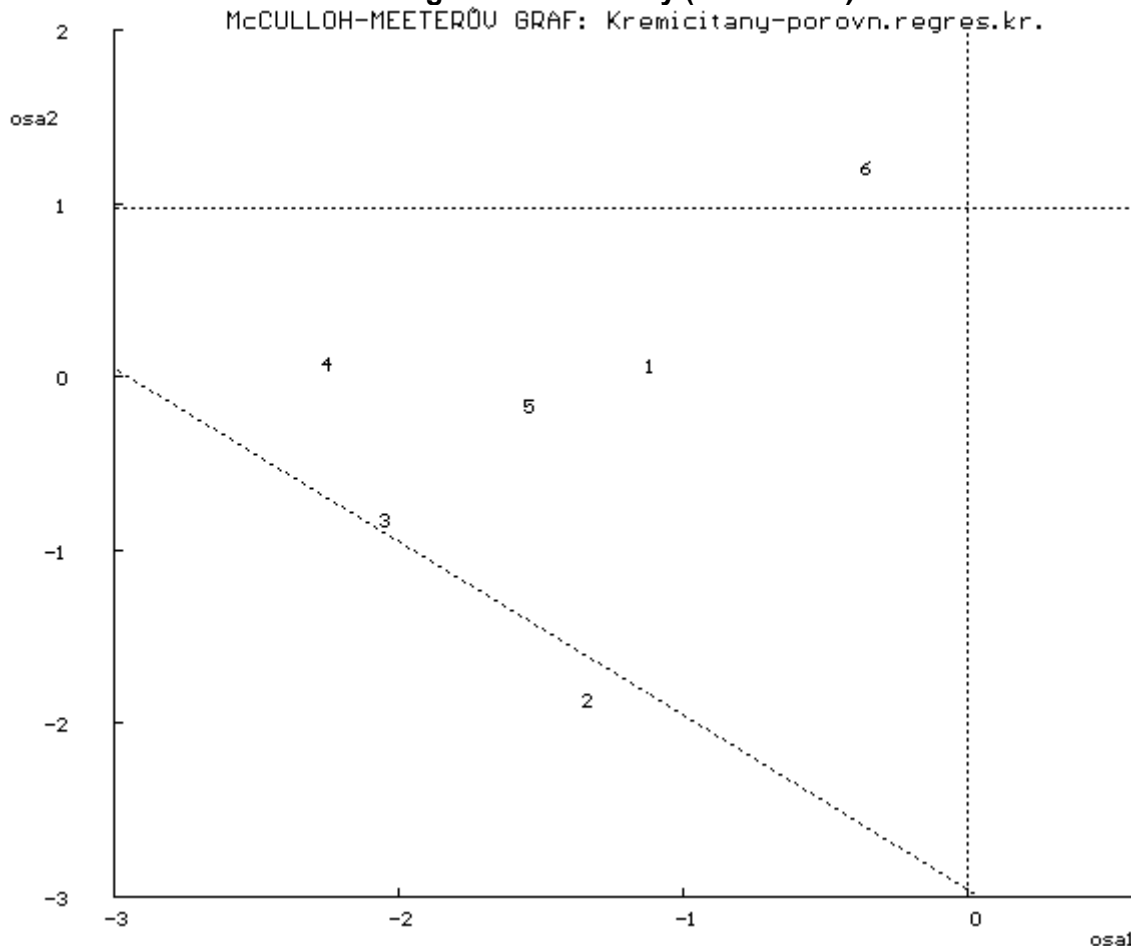
Ve Williamsově grafu je bod 6 vyhodnocen jako odlehlý, protože se nachází nad vodorovnou linií. Žádný z bodů nebyl vyhodnocen jako extrém (za svislou linií).

Graf 5: Pregibonův graf – křemičitany ($\lambda = 410 \text{ nm}$)



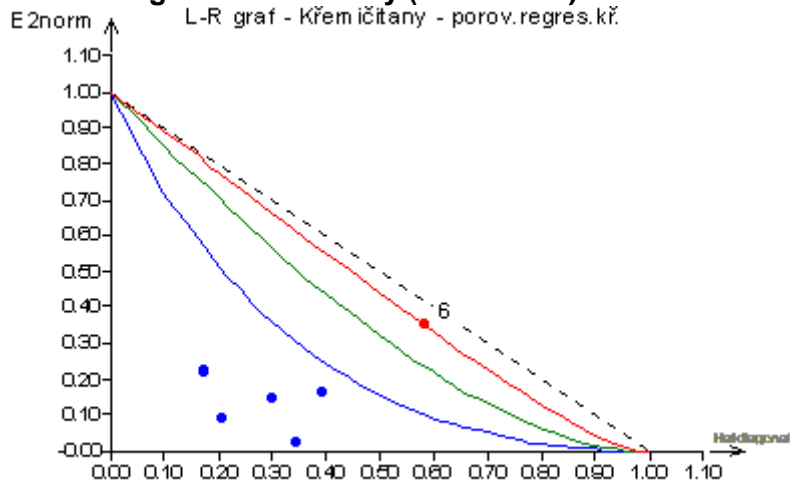
V Pregibonově grafu nebyl žádný z bodů vyhodnocen jako vlivný. Žádný bod se nenachází nad dolní přímkou (středně vlivné body) ani nad horní přímkou (silně vlivné body).

Graf 6: McCullohův-Meeterův graf – křemičitany ($\lambda = 410 \text{ nm}$)



V McCulloh-Meeterově grafu je bod 6 vyhodnocen jako odlehlý, protože se nachází nad vodorovnou linií. Žádný bod není vyhodnocen jako extrém, protože neleží za svíslou linií. Body 1, 3, 4 a 5 jsou podezřelé, protože leží v oblasti ohraničené všemi třemi liniemi.

Graf 7: L-R graf – křemičitany ($\lambda = 410 \text{ nm}$)



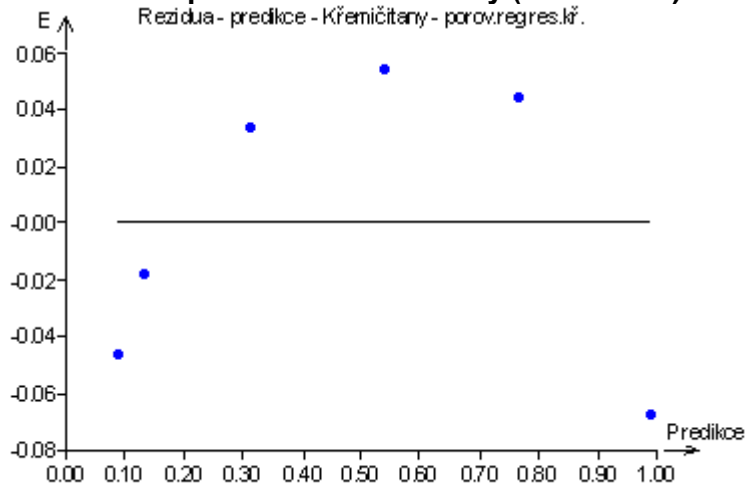
Podle vyhodnocení L-R grafu je bod 6 odlehlý (nachází se nad křivkami oddělujícími oblast odlehlých bodů)

Závěr testování vlnných bodů.: Bod 6 je odlehlý, protože byl v alespoň třech diagnostických grafech vlnných bodů tak vyhodnocen.

1.1.2 Kritika modelu a metody

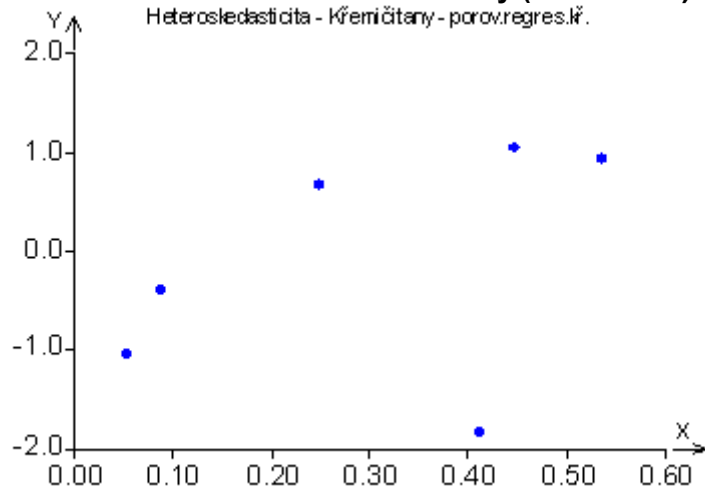
Diagnostické grafy

Graf 8: Rezidua - predikce – křemičitany ($\lambda = 410 \text{ nm}$)



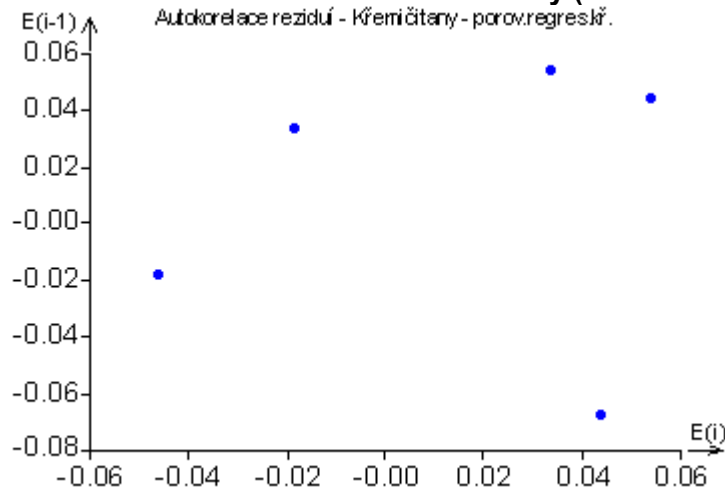
Body v grafu nekmitají v okolí vodorovné přímky se střídajícím se znaménkem. Tvoří spíše tvar paraboly, což poukazuje na nevhodný model.

Graf 9: Heteroskedasticita – křemičitany ($\lambda = 410 \text{ nm}$)



Pokud body v grafu tvoří výšeč nebo trojúhelník, svědčí to o heteroskedasticitě (nekonstantním rozptylu) v datech. V bodech sice lze spatřit výšeč, nicméně při tak nízkém počtu bodů to není příliš průkazné.

Graf 10: Autokorelace reziduí – křemičitany ($\lambda = 410 \text{ nm}$)



Mrak bodů s nenulovou směrnicí ukazuje na autokorelaci (nestabilitu v čase) reziduí. Zde je směrnice spíše kladná, což znamená pozitivní autokorelaci. Autokorelace reziduí ještě nemusí znamenat autokorelaci chyb.

Statistické testy

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 198.202543
Kvantil F (1-alfa, m-1, n-m) : 7.708647422
Pravděpodobnost : 0.0001477285592
Závěr : **Model je významný**

Scottovo kritérium multikolinearity

Hodnota kritéria SC : 0.3199544932
Závěr : **Model je korektní**

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0.3005310764
Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
Pravděpodobnost : 0.583549675
Závěr : **Rezidua vykazují homoskedasticitu.**

Jarque-Berrův test normality

Hodnota kritéria JB : 0.7019413218
Kvantil $\chi^2(1-\alpha,2)$: 5.991464547
Pravděpodobnost : 0.7040044084
Závěr : **Rezidua mají normální rozdělení.**

Waldův test autokorelace

Hodnota kritéria WA : 0.04794176918
Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
Pravděpodobnost : 0.8266842532
Závěr : **Autokorelace je nevýznamná**

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : -1
Kritické hodnoty DW : 0
Závěr : **Pozitivní autokorelace reziduí není prokázána.**

Znaménkový test reziduí

Hodnota kritéria Sg : 0.4564354646
Kvantil $N(1-\alpha/2)$: 1.959963999
Pravděpodobnost : 0.6480768681
Závěr : **V reziduích není trend.**

Statistické charakteristiky regrese

	bez abs. členu	s abs. členem	s abs. čl. a bez bodu 6
RSC	0.03278552597	0.01307910026	0.002072518892
$S^2(e)$	0.003269775066	0.006557105193	0.0006908396306
R	0.9748907003	0.990059521	0.9975069183
R^2	0.9504118776	0.9802178551	0.995020052
R_p	0.791919784	0.8739451511	0.9610235622
MEP	0.01213236571	0.007179024237	0.001638214925
AIC	-29.25716567	-32.77099515	-34.94214234

RSC Reziduální součet čtverců (RSC): optimální je hodnota co nejnižší
 $S^2(e)$ Reziduální rozptyl: optimální je hodnota co nejnižší
R Vícenásobný korelační koeficient: optimální je hodnota co nejbližší 1
 R^2 Koeficient determinace: optimální je hodnota co nejbližší 1
 R_p Predikovaný korelační koeficient: optimální je hodnota co nejbližší 1
MEP Střední kvadratická chyba predikce: optimální je hodnota co nejbližší 0
AIC Akaikeho informační kritérium: optimální je hodnota co nejzápornější (tedy nejnižší)

1.1.3 Opravený model (bez bodu 6)

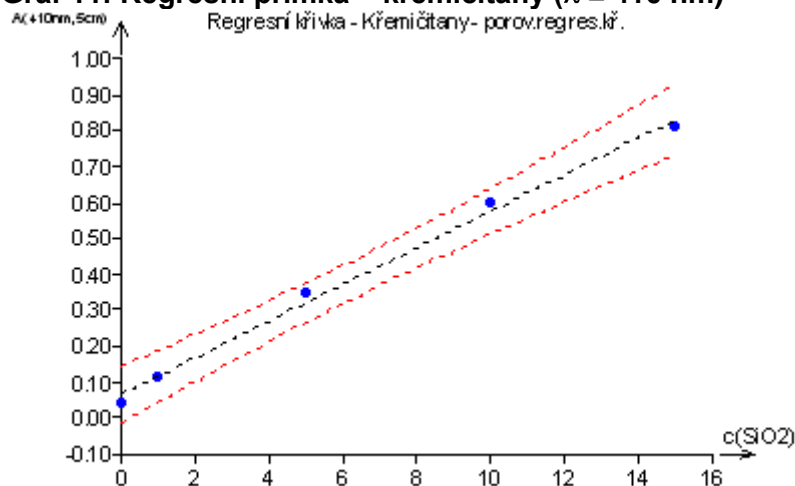
Odhady parametrů

Parametr	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
abs. člen	0.06519395466	0.01747559328	Významný	0.03356161106	0.009578817382	0.1208090919
směrnice	0.05106549118	0.002085755381	Významný	0.0001493743922	0.04442768668	0.05770329569

Hladina významnosti $\alpha = 0,05$ (95% jistota)

Absolutní člen lineárního regresního modelu je na základě testu intervalem spolehlivosti vyhodnocen jako významný (nenulový), protože interval vymezený spodní a horní mezí neobsahuje 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05. Hodnota směrnice je vyhodnocena jako významná, protože interval spolehlivosti neobsahuje 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05. Zároveň interval spolehlivosti neobsahuje ani 1, to znamená, že směrnice není jednotková.

Graf 11: Regresní přímka – křemičitany ($\lambda = 410 \text{ nm}$)



Konfidenční pás (vymezený okrajovými křivkami) je užší než u původního modelu. Proložení je těsnější.

1.1.4 Závěr regresní diagnostiky pro křemičitany – 410nm

Testování regrese je poznamenáno nízkým počtem dat. To zejména stěžuje interpretaci diagnostických grafů, které jsou jinak považovány za lepší nástroj než méně přísné statistické testy. Bod 6 byl vyhodnocen jako odlehlý a z modelu odstraněn. Z provedeného testování vyplývá, že navržený lineární model a použití metody nejmenších čtverců lze považovat za korektní.

Navržený model je $A(410\text{nm}) = 0.06519395466 + 0.05106549118 c(\text{SiO}_2)$

1.2 Křemičitany – 690nm

1.2.1 Kritika dat

Testování úseku a směrnice

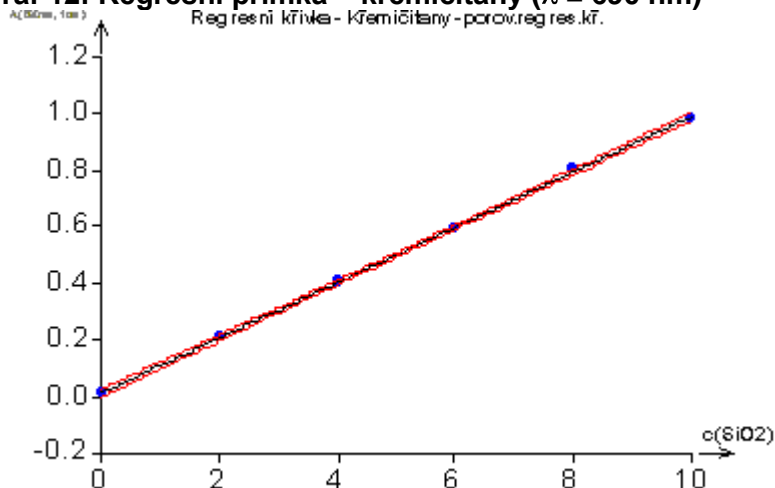
Odhady parametrů

Parametr	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
abs. člen	0.011619048	0.003814372	Významný	0.038172708	0.001028653	0.022209442
směrnice	0.097942857	0.000629923	Významný	1.03E-08	0.096193911	0.099691803

Hladina významnosti $\alpha = 0,05$ (95% jistota)

Absolutní člen lineárního regresního modelu je na základě testu intervalem spolehlivosti vyhodnocen jako významný (nenulový), protože interval vymezený spodní a horní mezí neobsahuje 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05. Hodnota směrnice je vyhodnocena jako významná, protože interval spolehlivosti neobsahuje 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05. Zároveň interval spolehlivosti neobsahuje ani 1, to znamená, že směrnice není jednotková.

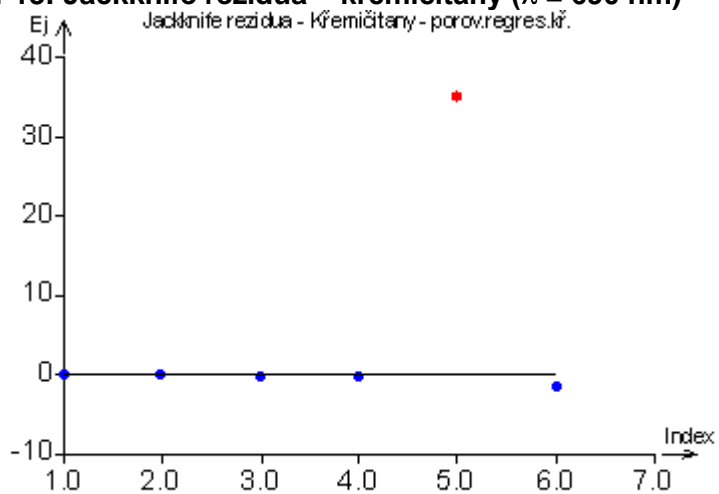
Graf 12: Regresní přímka – křemičitany ($\lambda = 690$ nm)



Konfidenční pás je úzký. Přímka těsně prokládá body. Dá se předpokládat oprávněnost lineárního modelu.

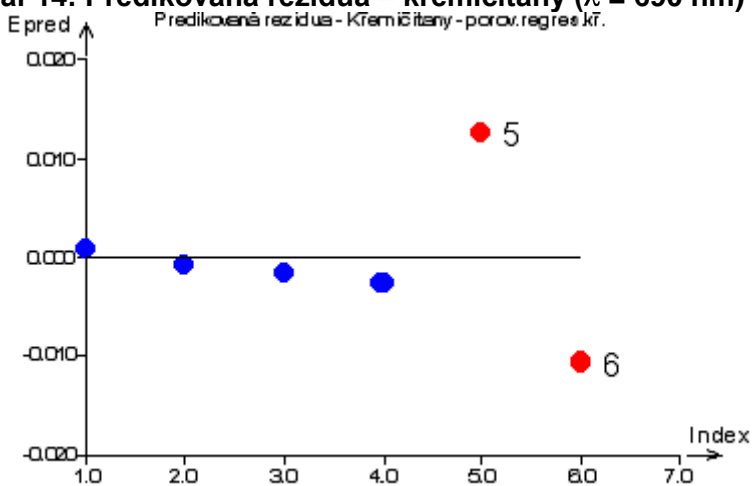
Testování vlivných bodů diagnostickými grafy

Graf 13: Jackknife rezidua – křemičitany ($\lambda = 690 \text{ nm}$)



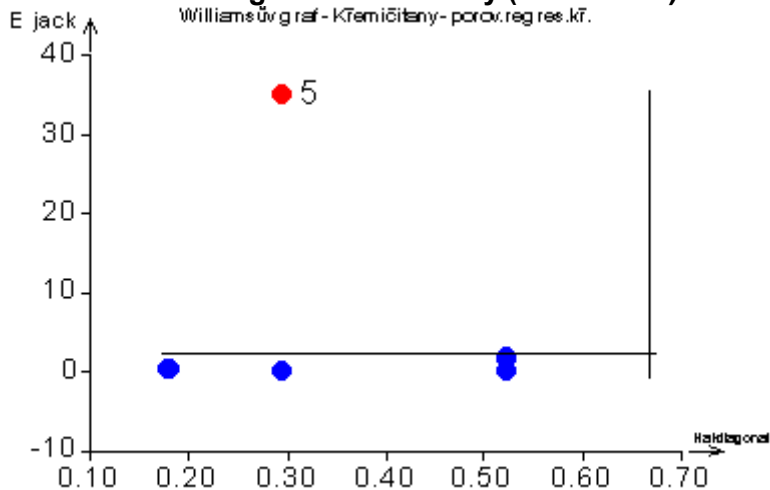
V grafu Jackknife reziduí je bod 5 identifikován jako podezřelý z odlehlosti. (vzdálen od ostatních)

Graf 14: Predikovaná rezidua – křemičitany ($\lambda = 690 \text{ nm}$)



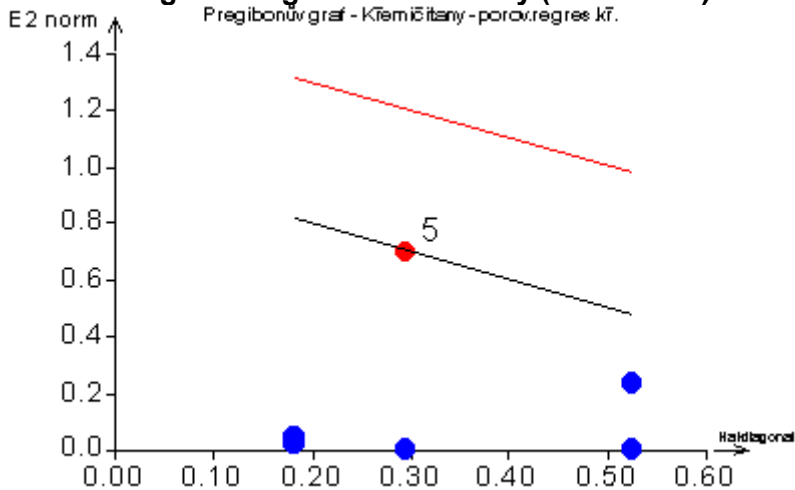
Z grafu predikovaných reziduí se jako ,podezřelé z odlehlosti jeví body 5 a 6.

Graf 15: Williamsův graf – křemičitany ($\lambda = 690$ nm)



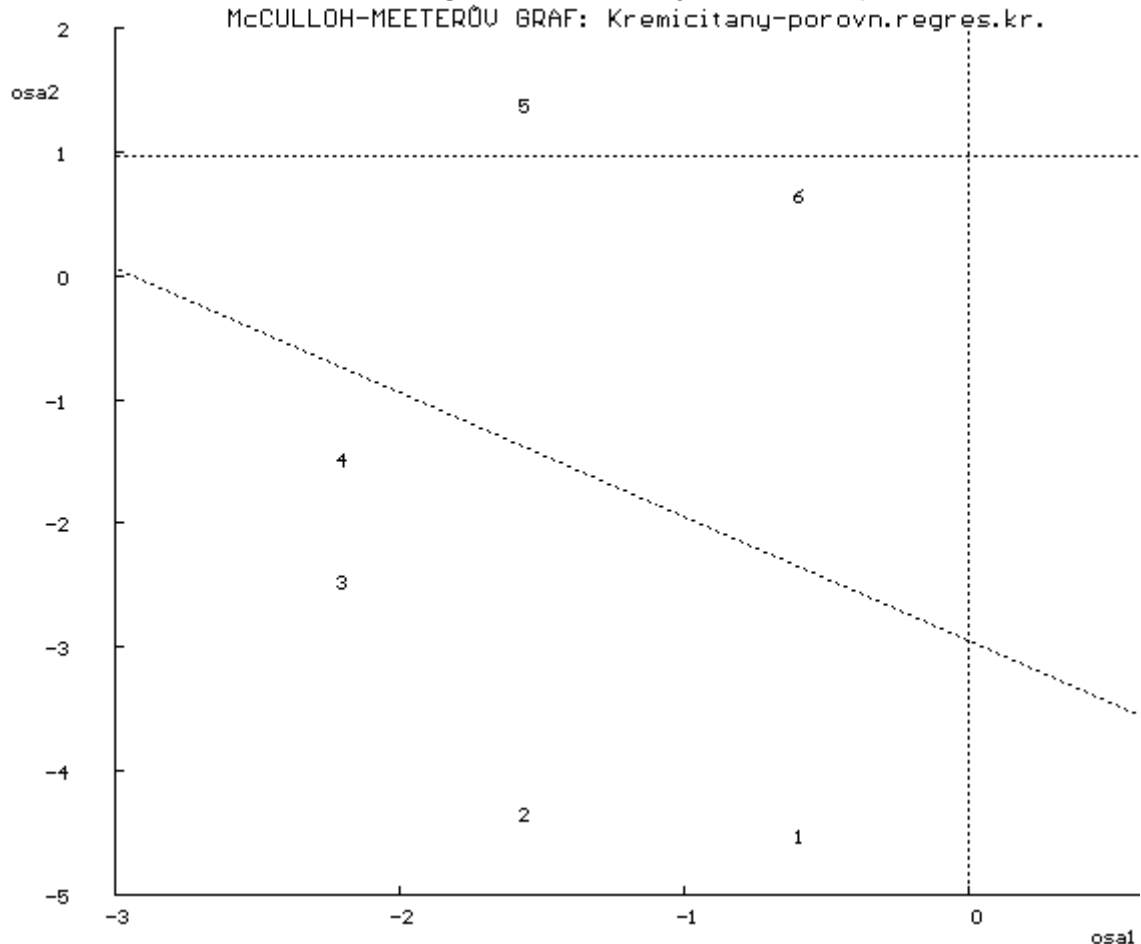
Ve Williamsově grafu je bod 5 vyhodnocen jako odlehlý, protože se nachází nad vodorovnou linií. Žádný z bodů nebyl vyhodnocen jako extrém (za svislou linií).

Graf 16: Pregibonův graf – křemičitany ($\lambda = 690$ nm)



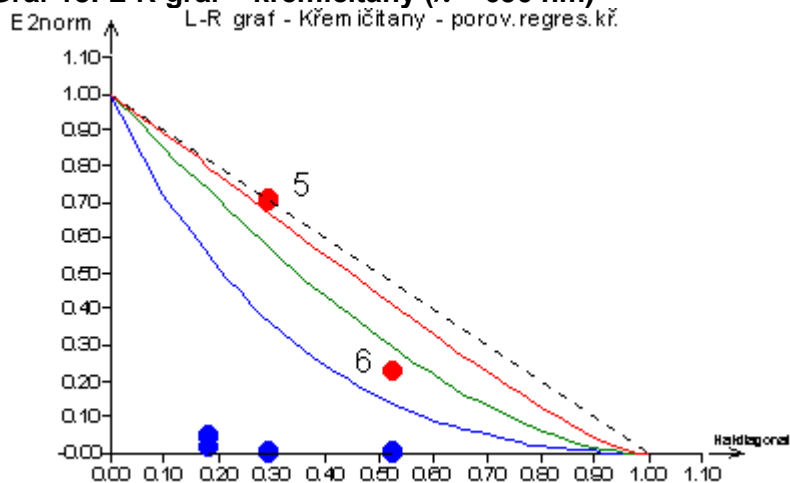
V Pregibonově grafu není žádný z bodů za hranicí přímkou oddělující oblast vlivných bodů, nicméně bod 5 se přece jen odlišuje od ostatních.

Graf 17: McCullohův-Meeterův graf – křemičitany ($\lambda = 690 \text{ nm}$)



Bod 5 je nad vodorovnou linií a tudíž odlehlý. Bod 6 je v oblasti mezi liniemi a tudíž podezřelý.

Graf 18: L-R graf – křemičitany ($\lambda = 690 \text{ nm}$)



Podle vyhodnocení L-R grafu je bod 5 odlehlý (nachází se nad křivkami oddělujícími oblast odlehlých bodů) a bod 6 podezřelý.

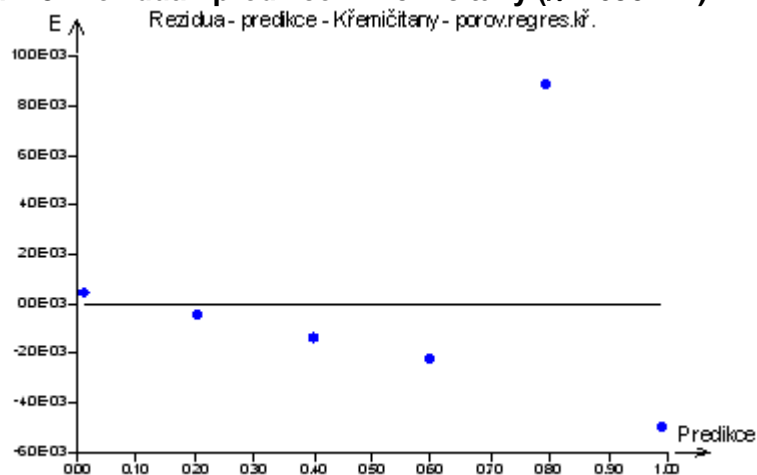
Závěr testování vlnných bodů.: Bod 5 je odlehlý, protože byl v alespoň třech diagnostických grafech vlnných bodů tak vyhodnocen. Bod 6 je podezřelý vlnný bod,

nicméně nemůže být odstraněn, protože není splněna podmínka počtu alespoň 5 bodů na jeden neznámý parametr.

1.2.2 Kritika modelu a metody

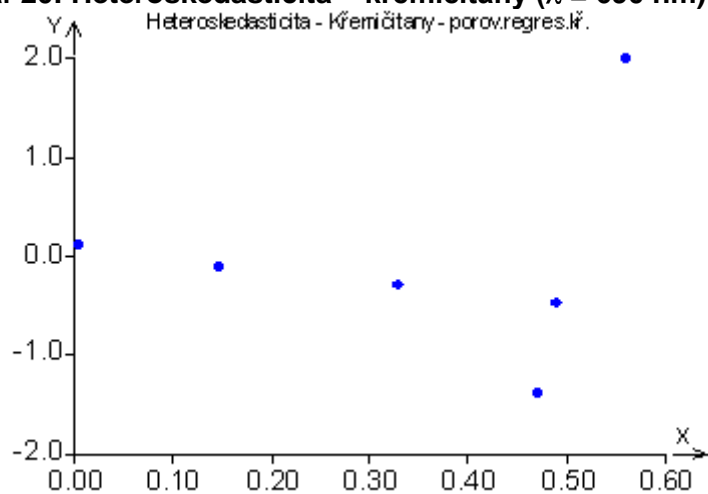
Diagnostické grafy

Graf 19: Rezidua - predikce – křemičitany ($\lambda = 690$ nm)



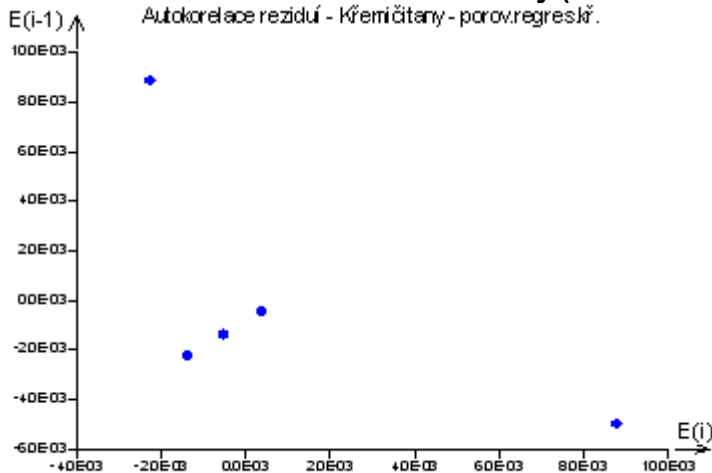
Body v grafu tvoří výšeč, což poukazuje na možnou heteroskedasticitu.

Graf 20: Heteroskedasticita – křemičitany ($\lambda = 690$ nm)



Pokud body v grafu tvoří výšeč nebo trojúhelník, svědčí to o heteroskedasticitě (nekonstantním rozptylu) v datech. V bodech lze spatřit trojúhelník.

Graf 21: Autokorelace reziduí– křemičitany ($\lambda = 690 \text{ nm}$)



Mrak bodů s nenulovou směrnici ukazuje na autokorelaci (nestabilitu v čase) reziduí. Směrnice je patrně rostoucí = pozitivní atokorelace.

Statistické testy

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 24175.24567
Kvantil F (1-alfa, m-1, n-m) : 7.708647422
Pravděpodobnost : 1.026336336E-008
Závěr : **Model je významný**

Scottovo kritérium multikolinearity

Hodnota kritéria SC : 0.3331627699
Závěr : **Model vykazuje multikolinearitu!**

Kritérium sice ukazuje na multikolinearitu. Nicméně oba parametry regresního modelu byly vyhodnoceny jako významné. Metodu nejmenších čtverců (MNC) lze použít. Kdyby parametry byly nulové, bylo by nutné použít Metodu racionálních hodnotí (RH).

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 1.223210766
Kvantil $\chi^2(1-\text{alfa}, 1)$: 3.841458829
Pravděpodobnost : 0.2687314571
Závěr : **Rezidua vykazují homoskedasticitu.**

Jarque-Berrův test normality

Hodnota kritéria JB : 1.331840201
Kvantil $\chi^2(1-\text{alfa}, 2)$: 5.991464547
Pravděpodobnost : 0.5138005619
Závěr : **Rezidua mají normální rozdělení.**

Waldův test autokorelace

Hodnota kritéria WA : 1.437342236
Kvantil $\chi^2(1-\text{alfa}, 1)$: 3.841458829
Pravděpodobnost : 0.2305699091
Závěr : **Autokorelace je nevýznamná**

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : -1

Kritické hodnoty DW : 0 a 2

Závěr : **Negativní autokorelace reziduí není prokázána.**

Znaménkový test reziduí

Hodnota kritéria Sg : 0.8838834765

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

Pravděpodobnost : 0.3767591178

Závěr : **V reziduích není trend.**

Statistické charakteristiky regrese

	s bodem 5	bez bodu 5
RSC	0.0001111047619	2.702702703E-007
$S^2(e)$	2.777619048E-005	9.009009009E-008
R	0.999917281	0.9999997594
R^2	0.9998345689	0.9999995188
Rp	0.9991622352	0.999994775
MEP	4.689724165E-005	2.934775059E-007
AIC	-61.38077882	-79.66640645

RSC	Reziduální součet čtverců (RSC):	optimální je hodnota co nejnižší
$S^2(e)$	Reziduální rozptyl:	optimální je hodnota co nejnižší
R	Vícenásobný korelační koeficient:	optimální je hodnota co nejbližší 1
R^2	Koeficient determinace:	optimální je hodnota co nejbližší 1
Rp	Predikovaný korelační koeficient:	optimální je hodnota co nejbližší 1
MEP	Střední kvadratická chyba predikce:	optimální je hodnota co nejbližší 0
AIC	Akaikeho informační kritérium:	optimální je hodnota co nejjápnější (tedy nejnižší)

1.2.3 Opravený mode (bez bodu 5)

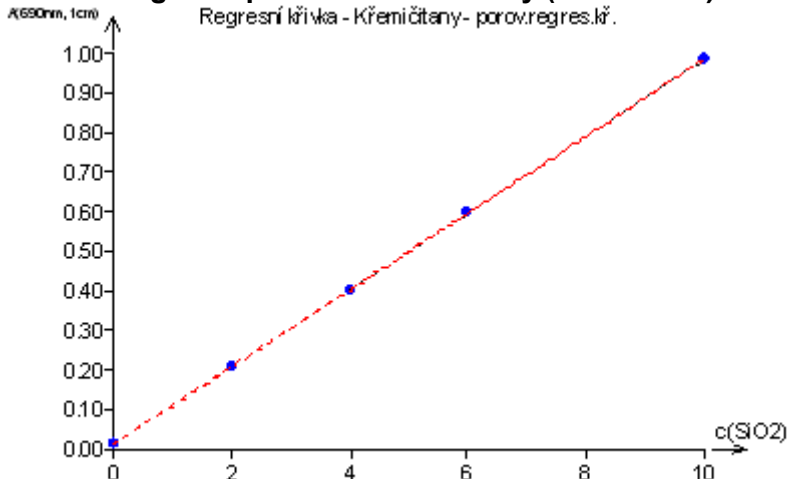
Odhady parametrů

Parametr	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
abs. člen	0.01221621622	0.0002178988581	Významný	1.250052578E-005	0.0115227648	0.01290966763
směrnice	0.09740540541	3.901015332E-005	Významný	1.416622375E-010	0.09728125769	0.09752955312

Hladina významnosti $\alpha = 0,05$ (95% jistota)

Absolutní člen lineárního regresního modelu je na základě testu intervalem spolehlivosti vyhodnocen jako významný (nenulový), protože interval vymezený spodní a horní mezí neobsahuje 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05. Hodnota směrnice je vyhodnocena jako významná, protože interval spolehlivosti neobsahuje 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05. Zároveň interval spolehlivosti neobsahuje ani 1, to znamená, že směrnice není jednotková.

Graf 22: Regresní přímka – křemičitany ($\lambda = 690 \text{ nm}$)



1.2.4 Závěr regresní diagnostiky pro křemičitany – 690nm

Testování regrese je poznamenáno nízkým počtem dat. To zejména stěžuje interpretaci diagnostických grafů, které jsou jinak považovány za lepší nástroj než méně přísné statistické testy. Bod 5 byl vyloučen na základě vyhodnocení grafů vlivných bodů a porovnání hodnot statistických charakteristik regrese pro model s a bez bodu 5. Z provedeného testování vyplývá, že navržený lineární model a použití metody nejmenších čtverců lze považovat za korektní.

Navržený model je $A(690\text{nm}) = 0.01221621622 + 0.09740540541 c(\text{SiO}_2)$

1.3 Porovnání dvou regresních přímek

1.3.1 Test shodnosti rozptylů

Pro přímkové závislosti postačí **Fischer-Snedecorův test**.

Testujeme hypotézu $H_0: \hat{\sigma}_A^2(e) = \hat{\sigma}_B^2(e)$ proti alternativní $H_A: \hat{\sigma}_A^2(e) \neq \hat{\sigma}_B^2(e)$

Regresní přímka **A**: s **vyšší** hodnotou reziduálního rozptylu.....Křemičitany 410 nm

Regresní přímka **B**: s **nižší** hodnotou reziduálního rozptylu.....Křemičitany 690 nm

Testační kritérium: $F_2 = \frac{\hat{\sigma}_A^2(e)}{\hat{\sigma}_B^2(e)}$ Porovnáváme s kvantilem F-rozdělení: $F_{0.95}(v_A, v_B)$

$v_i = n_i - m_i$ počet stupňů volnosti

n_i počet řádků dat

m_i počet parametrů (pro přímkovou závislost maximálně 2)

$$F_2 = \frac{0.0006908396306}{9.009009009\text{E-}008} = 7668.3199$$

$$F_{0.95}(v_A, v_B) = F_{0.95}(5 - 2, 5 - 2) = F_{0.95}(3, 3) = 9.27662815314474$$

Protože je $F_2 > F_{0.95}(3, 3)$ je hypotéza o rovnosti rozptylů zamítnuta = **rozptyly jsou různé**

1.3.2 Chowův test shody dvou lineárních modelů

Sloučený model:
$$\begin{bmatrix} y_A \\ y_B \end{bmatrix} = \begin{bmatrix} X_A \\ X_B \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_A \\ \varepsilon_B \end{bmatrix}$$

Data

(po odstranění bodu 6 a 5 z původních modelů)

		c(SiO2)	A (absorbance)
410 nm	1	0	0.042
	2	1	0.115
	3	5	0.348
	4	10	0.594
	5	15	0.810
690 nm	1	0	0.012
	2	2	0.207
	3	4	0.402
	4	6	0.597
	6	10	0.986

Testujeme nulovou hypotézu $H_0: \beta_A = \beta_B$ proti alternativní hypotéze $H_A: \beta_A \neq \beta_B$

Testační kritérium:
$$F_C = \frac{(RSC - RSC_A - RSC_B)(n - 2m)}{(RSC_A + RSC_B)(m)}$$

RSC Reziduální součet čtverců pro sloučený model = 0.1672934985

RSC_A Reziduální součet čtverců pro model A = 0.002072518892

RSC_B Reziduální součet čtverců pro model B = 2.702702703E-007

$n = n_A + n_B$ počet řádek sloučeného modelu (součet pro modely A a B)

$n = 5 + 5 = 10$

m počet parametrů (pro přímkovou závislost 2)

$$F_C = \frac{(0.1672934985 - 0.002072518892 - 2.702702703E-007)(10 - 4)}{(0.002072518892 + 2.702702703E-007)(2)}$$

$$F_C = 239.128097075873$$

Testační kritérium porovnáme s kvantilem F-rozdělení a rozeznáváme dva případy:

1. Pokud se reziduální rozptyly modelů rovnají: $\hat{\sigma}_A^2(e) = \hat{\sigma}_B^2(e) \Rightarrow F_{0.95}(m, n - 2m)$

2. Pokud se reziduální rozptyly modelů nerovnájí: $\hat{\sigma}_A^2(e) \neq \hat{\sigma}_B^2(e) \Rightarrow F_{0.95}(m, r)$

$$r = \frac{[(n_A - m)\hat{\sigma}_A^2(e) + (n_B - m)\hat{\sigma}_B^2(e)]^2}{(n_A - m)\hat{\sigma}_A^4(e) + (n_B - m)\hat{\sigma}_B^4(e)}$$

$$r = \frac{(5 - 2)0.0006908396306 + (5 - 2)9.009009009E-008^2}{(5 - 2)4.77259395207545E-07 + (5 - 2)8.11622433242432E-15}$$

$$r = 3.0007824399577 \square 3$$

$$F_{0,95}(m, r) = F_{0,95}(2, 3) = 9.55209449592114$$

Protože $F_C > F_{0,95}(2, 3)$ zamítáme hypotézu H_0 pro 95% pravděpodobnost.

1.3.3 Závěr

Nalezené lineární modely pro stanovení křemičitanů fotometrií při vlnové délce 410 nm a 690 nm nejsou shodné.

2 Určení stupně polynomu

Hledání stupně polynomu závislosti absorpance na koncentraci křemičitanů (mg/l SiO₂) ve vzorcích vody. Stanovováno fotometricky při vlnové délce 410nm (viz. Příklad Porovnání regresních křivek)

Data

	c(SiO ₂)	A (410nm, 5cm)
1	0	0.042
2	1	0.115
3	5	0.348
4	10	0.594
5	15	0.810
6	20	0.924

2.1 Metoda nejmenších čtverců

Statistické charakteristiky regrese

SP	ABS	RSC	S ² (e)	R	R ²	R _p	MEP	AIC	MUL	PAR
1	ne	3.279E-02	3.270E-03	9.749E-01	9.504E-01	7.919E-01	1.213E-02	-29.26	ne	1 / 0
	ano	1.308E-02	6.557E-03	9.901E-01	9.802E-01	8.739E-01	7.179E-03	-32.77	ne	2 / 0
2	ne	3.746E-03	9.366E-04	9.972E-01	9.943E-01	9.862E-01	7.634E-04	-40.27	ano	2 / 0
	ano	3.719E-04	1.240E-04	9.997E-01	9.994E-01	9.895E-01	5.790E-04	-52.13	ano	3 / 0
3	ne	3.575E-03	1.192E-03	9.973E-01	9.946E-01	8.182E-01	1.052E-02	-38.55	ano	4 / 2
	ano	2.067E-04	5.427E-03	9.998E-01	9.997E-01	9.582E-01	2.328E-03	-53.66	ano	3 / 2
4	ne	2.280E-03	1.140E-03	9.983E-01	9.966E-01	3.049E-02	1.294E-01	-39.25	ano	5 / 3
	ano	2.240E-06	2.240E-06	1.000E+00	1.000E+00	9.873E-01	7.030E-04	-78.80	ano	4 / 4
5	ne	1.764E-03	1.764E-03	9.987E-01	9.973E-01	2.441E-02	9.298E-02	-38.79	ano	5 / 5

SP Stupeň polynomu

ABS Absolutní člen

MUL Multikolinearita

PAR Počet parametrů / počet parametrů vyhodnocených jako nevýznamné

RSC Reziduální součet čtverců (RSC):

optimální je hodnota co nejnižší

S²(e) Reziduální rozptyl:

optimální je hodnota co nejnižší

R Vícenásobný korelační koeficient:

optimální je hodnota co nejbližší 1

R² Koeficient determinace:

optimální je hodnota co nejbližší 1

R_p Predikovaný korelační koeficient:

optimální je hodnota co nejbližší 1

MEP Střední kvadratická chyba predikce:

optimální je hodnota co nejbližší 0

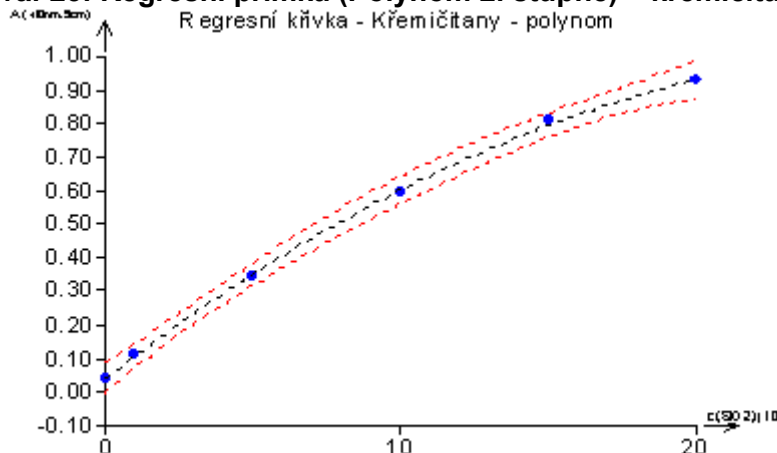
AIC Akaikeho informační kritérium:

optimální je hodnota co nejzápornější (tedy nejnižší)

Rozhodčím kritériem pro výběr nejvhodnějšího modelu je MEP!!! Přestože ostatní kritéria kvality regrese mají optimální hodnotu pro polynom čtvrtého stupně, jako nejlepší model byl na základě minimální hodnoty MEP vybrán **polynom druhého stupně s absolutním členem.**

Vybraný model

Graf 23: Regresní přímka (Polynom 2. stupně) – křemičitany ($\lambda = 410 \text{ nm}$)



Odhady parametrů

Par.	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
β_0	0.04320428092	0.00828042053	Významný	0.01369002126	0.0168522872	0.06955627464
β_1	0.06760496083	0.002301341045	Významný	8.663018457E-005	0.06028106652	0.07492885513
β_2	-0.001163886402	0.0001149535878	Významný	0.002052398218	-0.001529720023	-0.0007980527809

Hladina významnosti $\alpha = 0,05$ (95% jistota)

Indikace multikolinearity

Parametr	Číslo podmíněnosti Kappa	VIF (Variance Inflation Factor)
β_0	1	1
β_1	26.6211329	13.56535185
β_2	52.2422658	13.56535185

Kappa Pokud je > 1000 , pak jde o silnou multikolinearitu

VIF Pokud > 10 , pak jde o silnou multikolinearitu

Statistické testy

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 2665.357814

Kvantil F (1-alfa, m-1, n-m) : 9.552094496

Pravděpodobnost : 1.333942459E-005

Závěr : **Model je významný**

Scottovo kritérium multikolinearity

Hodnota kritéria SC : 0.7791234762

Závěr : **Model vykazuje multikolinearitu!**

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0.3246177248

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

Pravděpodobnost : 0.5688456874

Závěr : **Rezidua vykazují homoskedasticitu.**

Jarque-Berrův test normality

Hodnota kritéria JB : 0.7305533879
Kvantil $\chi^2(1-\alpha,2)$: 5.991464547
Pravděpodobnost : 0.6940045974
Závěr : **Rezidua mají normální rozdělení.**

Waldův test autokorelace

Hodnota kritéria WA : 0.7390366835
Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
Pravděpodobnost : 0.3899696376
Závěr : **Autokorelace je nevýznamná**

Durbin-Watsonův test autokorelace

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

Znaménkový test reziduí

Hodnota kritéria Sg : 1.944543648
Kvantil $N(1-\alpha/2)$: 1.959963999
Pravděpodobnost : 0.05182992722
Závěr : **V reziduích není trend.**

Závěr k MNČ: Scottovo kritérium a stejně tak hodnoty faktoru VIF (hodnoty jsou vyšší než 10) ukazují na multikolinearitu. Z ohledem na to, že žádný parametr nebyl vyhodnocen jako nevýznamný, náhrada MNČ (Metoda nejmenších čtverců) metodou RH (Racionálních hodnot) není nutná.

2.2 Metoda racionálních hodnotí

Metoda RH se používá v případě, kdy model vykazuje multikolinearitu a ne všechny parametry jsou vyhodnoceny jako významné. Podle Silenova pravidlo má platit $\beta_i > 3s_i$ (každý parametr má být větší než trojnásobek jeho směrodatné odchylky), jinak je parametr statisticky nevýznamný. Parametr se „utápí ve svém šumu“. Podle metody RH se nastaví omezení, čímž se pak počítá pouze s dostatečně velkými čísly.

Hledání optimální hodnoty omezení pro určený stupeň polynomu

SP	ABS	omezení	Rp ²	MEP	AIC
2	ano	1.E-38	9.9737E-01	5.7903E-04	-5.2132E+01
2	ano	1.E-04	9.9737E-01	5.7903E-04	-5.2132E+01
2	ano	1.E-03	9.9737E-01	5.7903E-04	-5.2132E+01
2	ano	1.E-02	9.9737E-01	5.7903E-04	-5.2132E+01
2	ano	1.E-01	9.9014E-01	2.1613E-03	-4.8256E+01

SP Stupeň polynomu

ABS Absolutní člen

Rp² Predikovaný korelační koeficient: optimální je hodnota co nejbližší 1

MEP Střední kvadratická chyba predikce: optimální je hodnota co nejbližší 0

AIC Akaikeho informační kritérium: optimální je hodnota co nejzápornější (tedy nejnižší)

Závěr k metodě RH: Z hodnot vybraných nejdůležitějších regresních charakteristik je zřejmé, že nebylo nutné zavádět omezení na vlastní čísla. Minimálních hodnot kritérií bylo dosaženo již při minimální, použitým softwarem povolené, hodnotě omezení 10^{-38} . Tím je potvrzen závěr o korektnosti použití MNČ.

2.3 Závěr

Na základě provedené regresní analýzy byla potvrzena oprávněnost použití metody nejmenších čtverců pro vyhledání optimálního stupně polynomu. Jako nejvhodnější byl vybrán polynom druhého stupně s absolutním členem.

$$Y = 0.04320428_{(0.008280425)} + 0.067604960_{(0.002301341)} X - 0.001163886_{(0.000114954)} X^2$$

3 Validace nové analytické metody

Porovnání nově navržené metody stanovení křemičitanů fotometricky ve vzorcích vody vybarvených modře (dodatečným přídatkem kyseliny askorbové) při vlnové délce 690nm se zavedenou metodou, kde je u žlutě vybarveného vzorku měřena absorbance při vlnové délce 410nm.

Data

	c(SiO ₂) při 410nm	c(SiO ₂) při 690nm
1	0.1	0.0
2	1.0	1.0
3	1.8	1.9
4	3.5	3.9
5	4.6	4.8
6	5.4	5.9
7	7.5	8.0
8	9.9	9.9
9	15.8	14.6

Návrh modelu $y = \beta_0 + \beta_1 x$ a budeme testovat hypotézu $H_0: \beta_0 = 0$ a $\beta_1 = 1$ proti alternativní hypotéze $H_A: \beta_0 \neq 0$ a $\beta_1 \neq 1$.

3.1 Kritika dat

Testování úseku a směrnice

Odhady parametrů

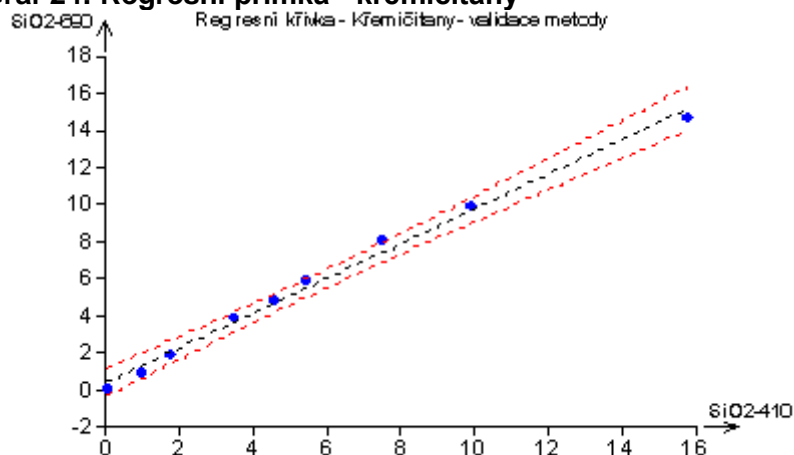
Parametr	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
abs. člen	0.3857863448	0.2289858033	Nevýznamný	0.1359046573	-0.1556790389	0.9272517285
směrnice	0.9380629616	0.03166273943	Významný	1.286048157E-008	0.8631924801	1.012933443

Hladina významnosti $\alpha = 0,05$ (95% jistota)

Absolutní člen lineárního regresního modelu je na základě testu intervalem spolehlivosti vyhodnocen jako nevýznamný (nulový), protože interval vymezený spodní a horní mezí obsahuje 0 resp. p-hodnota testu je vyšší než zadaná hladina významnosti 0,05. Hodnota směrnice je vyhodnocena jako významná, protože interval spolehlivosti neobsahuje 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05. Interval spolehlivosti však obsahuje 1, to znamená, že směrnice je jednotková.

Hypotéza H_0 o nulovosti absolutního členu a jednotkové směrnici je na hladině významnosti 0,05 přijata.

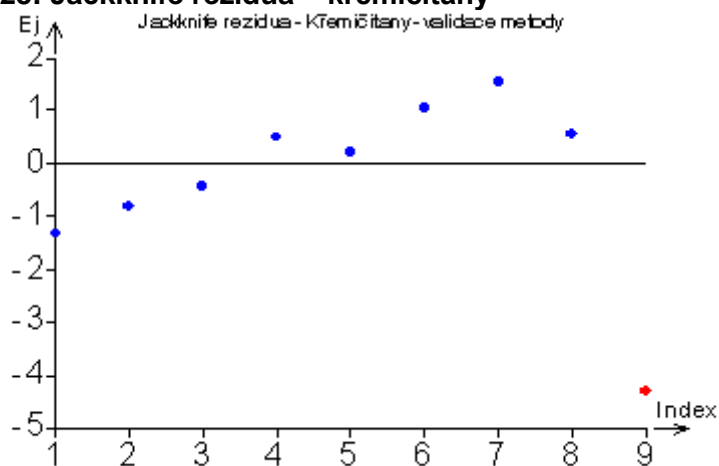
Graf 24: Regresní přímka - křemičitany



Přímka neprokládá body zcela optimálně. Bude nutné prověřit vlivné body a předpoklady pro metodu nejmenších čtverců a model případně opravit.

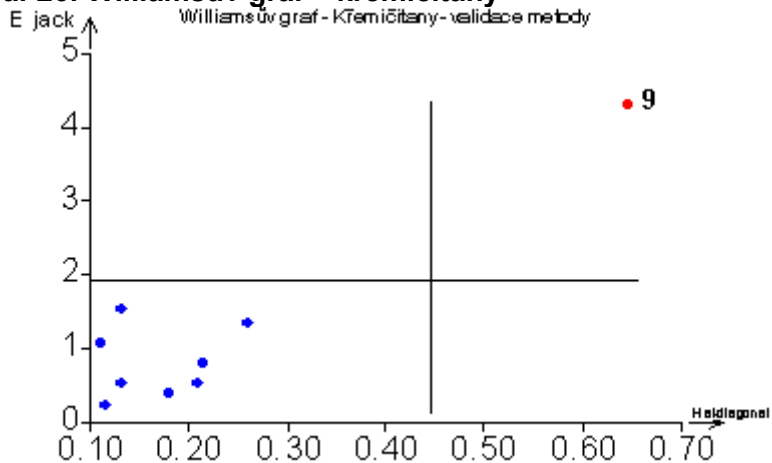
Testování vlivných bodů diagnostickými grafy

Graf 25: Jackknife rezidua – křemičitany



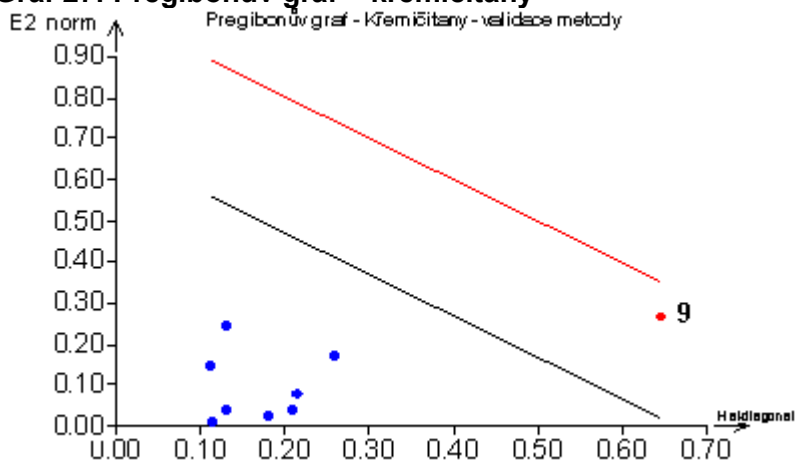
V grafu Jackknife reziduí je bod 9 identifikován jako podezřelý z odlehlosti. (vzdálen od ostatních)

Graf 26: Williamsův graf – křemičitany



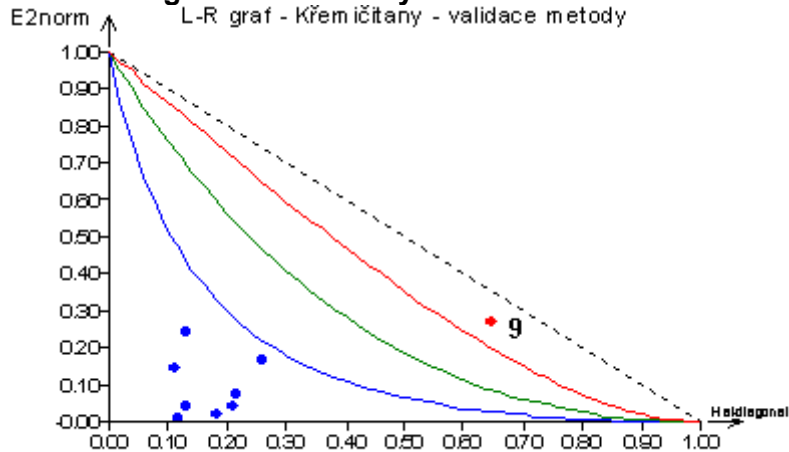
Ve Williamsově grafu je bod 9 vyhodnocen jako odlehlý a zároveň extrém, protože se nachází nad vodorovnou linií a vpravo od svislé linie.

Graf 27: Pregibonův graf – křemičitany



V Pregibonově grafu byl bod 9 vyhodnocen jako středně vlivný, protože se nachází nad dolní linií.

Graf 28: L-R graf – křemičitany



Podle vyhodnocení L-R grafu je bod 9 odlehlý (nachází se nad křivkami oddělujícími oblast odlehlých bodů).

Závěr testování vlivných bodů.: Bod 9 je odlehlý, protože byl v alespoň třech diagnostických grafech vlivných bodů tak vyhodnocen.

3.2 Kritika modelu a metody

Statistické testy

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 877.7422502
Kvantil F (1-alfa, m-1, n-m) : 5.591447851
Pravděpodobnost : 1.286048151E-008
Závěr : **Model je významný**

Scottovo kritérium multikolinearity

Hodnota kritéria SC : 0.3318976483
Závěr : **Model vykazuje multikolinearitu!**

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0.4777281899
Kvantil $\chi^2(1-\alpha, 1)$: 3.841458829
Pravděpodobnost : 0.4894531593
Závěr : **Rezidua vykazují homoskedasticitu.**

Jarque-Berrův test normality

Hodnota kritéria JB : 0.6698718528
Kvantil $\chi^2(1-\alpha, 2)$: 5.991464547
Pravděpodobnost : 0.7153839221
Závěr : **Rezidua mají normální rozdělení.**

Waldův test autokorelace

Hodnota kritéria WA : 0.9627251977
Kvantil $\chi^2(1-\alpha, 1)$: 3.841458829
Pravděpodobnost : 0.3265012199
Závěr : **Autokorelace je nevýznamná**

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : -1

Kritické hodnoty DW 0

Závěr : **Pozitivní autokorelace reziduí není prokázána.**

Znaménkový test reziduí

Hodnota kritéria Sg : 1.405633826

Kvantil N(1-alfa/2) : 1.959963999

Pravděpodobnost : 0.1598328863

Závěr : **V reziduích není trend.**

Scottovo kritérium sice ukazuje na multikolinearitu, nicméně nevýznamnost absolutního členu v lineárním přímkovém modelu je požadována z ohledem na validaci nové metody a směrnice je významná (nenulová).

Statistické charakteristiky regrese

	s bodem 9	bez bodu 9
RSC	1.385076338	0.3392858512
S ² (e)	0.1978680482	0.05654764186
R	0.9960361897	0.9979548796
R ²	0.9920880912	0.9959139417
Rp	0.9496343404	0.9774582638
MEP	0.4961684107	0.117651362
AIC	-12.8432239	-21.2828308

RSC	Reziduální součet čtverců (RSC):	optimální je hodnota co nejnižší
S ² (e)	Reziduální rozptyl:	optimální je hodnota co nejnižší
R	Vícenásobný korelační koeficient:	optimální je hodnota co nejbližší 1
R ²	Koeficient determinace:	optimální je hodnota co nejbližší 1
Rp	Predikovaný korelační koeficient:	optimální je hodnota co nejbližší 1
MEP	Střední kvadratická chyba predikce:	optimální je hodnota co nejbližší 0
AIC	Akaikeho informační kritérium:	optimální je hodnota co nejzápornější (tedy nejnižší)

3.3 Opravený model (bez bodu 9)

Odhady parametrů

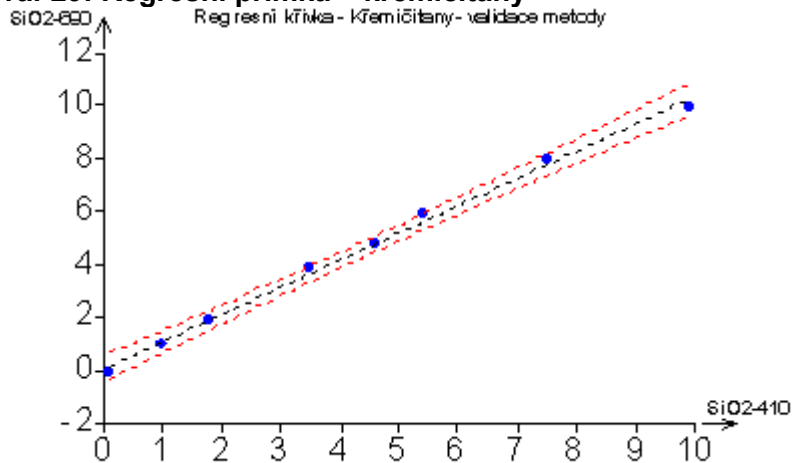
Parametr	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
abs. člen	0.08233152347	0.1412945614	Nevýznamný	0.5813207563	-0.2634038134	0.4280668603
směrnice	1.027850527	0.02687792336	Významný	2.135159294E-008	0.9620826178	1.093618436

Hladina významnosti $\alpha = 0,05$ (95% jistota)

Absolutní člen lineárního regresního modelu je na základě testu intervalem spolehlivosti vyhodnocen jako nevýznamný (nulový), protože interval vymezený spodní a horní mezí obsahuje 0 resp. p-hodnota testu je vyšší než zadaná hladina významnosti 0,05. Hodnota směrnice je vyhodnocena jako významná, protože interval spolehlivosti neobsahuje 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05. Interval spolehlivosti však obsahuje 1, to znamená, že směrnice je jednotková.

Hypotéza H₀ o nulovosti absolutního členu a jednotkové směrnici je na hladině významnosti 0,05 přijata.

Graf 29: Regresní přímka – křemičitany



3.4 Závěr

Regresní analýza potvrdila, že fotometrická metoda stanovení křemičitanů při vlnové délce 690nm nám dává stejné výsledky jako původní metoda (vlnová délka 410nm) po testování na hladině významnosti 95%.

4 Vícerozměrný lineární regresní model

V Českém hydrometeorologickém ústavu jsou průběžně shromažďována data z monitoringu ovzduší. Množství znečišťujících látek jsou pak uveřejněna v rámci Emisní bilance České republiky podle krajů. Celková roční bilance pak zahrnuje emise těchto základních polutantů: NO_x – oxidy dusíku, SO_2 – oxidy síry (zejména oxid siřičitý), TZL – tuhé znečišťující látky, CO – oxid uhelnatý, VOC - těkavé organické látky a NH_3 – amoniak. Budeme sledovat vztah mezi množstvím oxidů dusíku a ostatními znečišťujícími látkami s využitím vícerozměrného lineárního modelu.

Data

Kraj	NO_x [t/rok]	TZL [t/rok]	SO_2 [t/rok]	CO [t/rok]	VOC [t/rok]	NH_3 [t/rok]
Hl.m. Praha	19533.0	3925.7	3104	33927.0	14196.0	246.5
Jihočeský	18521.0	7072.9	12097.2	35311.0	14762.8	8445.3
Jihomoravský	21352.6	5328.2	4655.1	37673.9	19643.7	9207.1
Karlovarský	12795.9	2283.2	17322.1	11618.7	5780.7	1299.0
Královéhradecký	12479.2	4231.4	9472.1	27655.3	11989.4	5627.7
Liberecký	8169.4	2774.5	4307	18890.7	7901.3	935.9
Moravskoslezský	38399.8	9156.0	29144.7	176437.8	19819.2	4696.2
Olomoucký	15382.6	4391.6	7493.6	27111.7	12188.9	5281.0
Pardubický	21579.6	4286.0	16859.7	25237.1	11023.9	5175.3
Plzeňský	16481.7	5447.7	12094.6	30351.7	12239.4	6247.6
Středočeský	40091.8	11517.1	26154.2	65029.5	27801.9	8517.3
Ústecký	76788.6	6312	71963.3	37151.3	17907.0	2628.0
Vysočina	12900.1	5627.1	4205.9	25563.2	11211.0	7477.4
Zlínský	11839.8	3352.4	9114.4	18834.4	11199.3	3378.5
Celkem	326128.4	75494.9	227922.7	570569.2	197546.9	69162.8

Navržený model: $\text{NO}_x = \beta_0 + \beta_1(\text{TZL}) + \beta_2(\text{SO}_2) + \beta_3(\text{CO}) + \beta_4(\text{VOC}) + \beta_5(\text{NH}_3)$

Budeme testovat významnost jednotlivých parametrů.

4.1 Kritika dat

Testování parametrů

Odhady parametrů

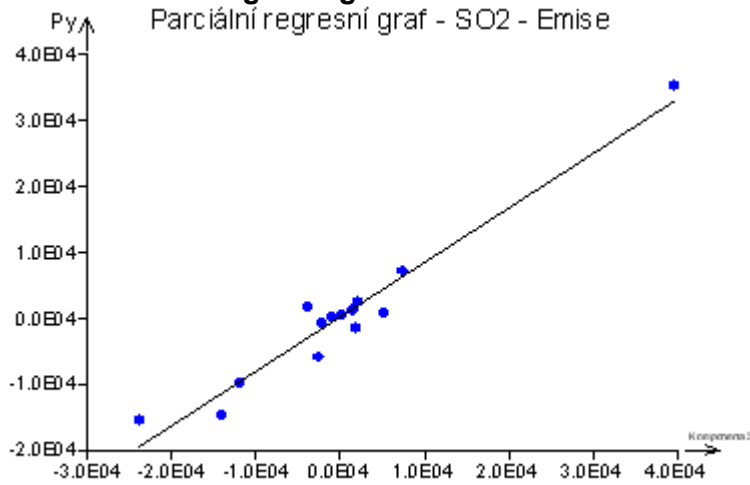
Par.	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
β_0	-3255.598733	2713.404138	Nevýznamný	0.2645329176	-9512.719896	3001.522429
β_1	-1.251362776	1.176996882	Nevýznamný	0.3187355748	-3.965522454	1.462796902
β_2	0.8348549014	0.06920737814	Významný	2.059863406E-006	0.6752624012	0.9944474015
β_3	0.006150701929	0.03607828849	Nevýznamný	0.8688637273	-0.07704598051	0.08934738437
β_4	1.459369148	0.3984322889	Významný	0.006375585158	0.5405826425	2.378155654
β_5	-0.2267717024	0.5238819625	Nevýznamný	0.6765438891	-1.434845674	0.9813022695

Hladina významnosti $\alpha = 0,05$ (95% jistota)

Absolutní člen β_0 lineárního regresního modelu je na základě testu intervalem spolehlivosti vyhodnocen jako nevýznamný (nulový), protože interval vymezený spodní a horní mezí obsahuje 0 resp. p-hodnota testu je vyšší než zadaná hladina významnosti 0,05. Parametry

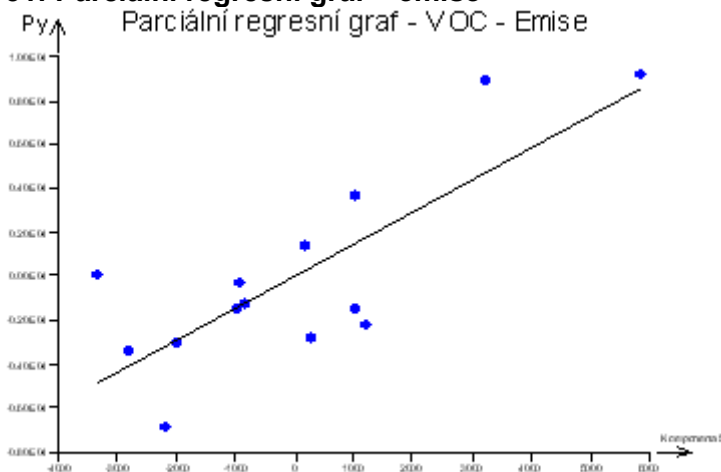
β_1 , β_3 a β_5 jsou vyhodnoceny jako nevýznamné, protože jejich intervaly spolehlivosti neobsahují 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05.

Graf 30: Parciální regresní graf – emise



Graf vyjadřuje závislost množství NO_x na SO_2 . Lineární trend v bodech značí silnou korelaci, což je potvrzeno i významností parametru β_2 .

Graf 31: Parciální regresní graf – emise



Graf vyjadřuje závislost množství NO_x na VOC. Lineární trend v bodech značí středně silnou korelaci, což je potvrzeno i významností parametru β_4 .

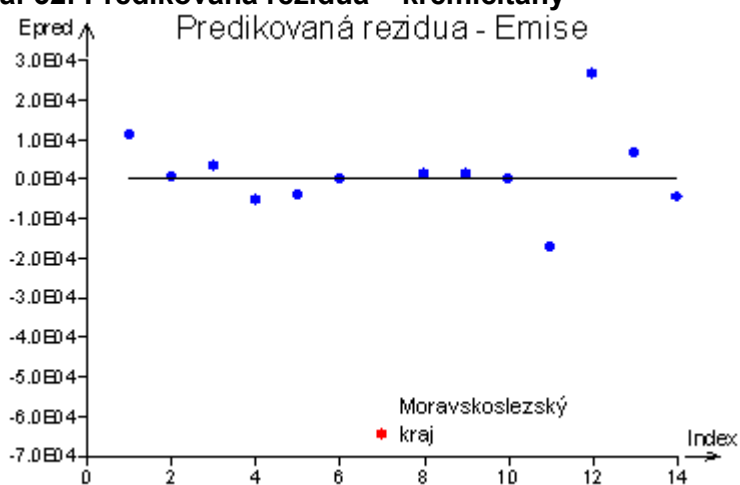
Párové korelace mezi proměnnými

Abs	TZL	0
Abs	SO ₂	0
Abs	CO	0
Abs	VOC	0
Abs	NH ₃	0
TZL	SO ₂	0.3916821502
TZL	CO	0.6669052358
TZL	VOC	0.89679613
TZL	NH ₃	0.5830438618
SO ₂	CO	0.2879015868
SO ₂	VOC	0.3897495436
SO ₂	NH ₃	0.1281216474
CO	VOC	0.5602468731
CO	NH ₃	0.1413266396
VOC	NH ₃	0.5206222326

Mezi proměnnými TZL a VOC je silná korelace, Mezi proměnnými TZL a CO, TZL a NH₃, CO a VOC, VOC a NH₃ jsou středně silné korelace.

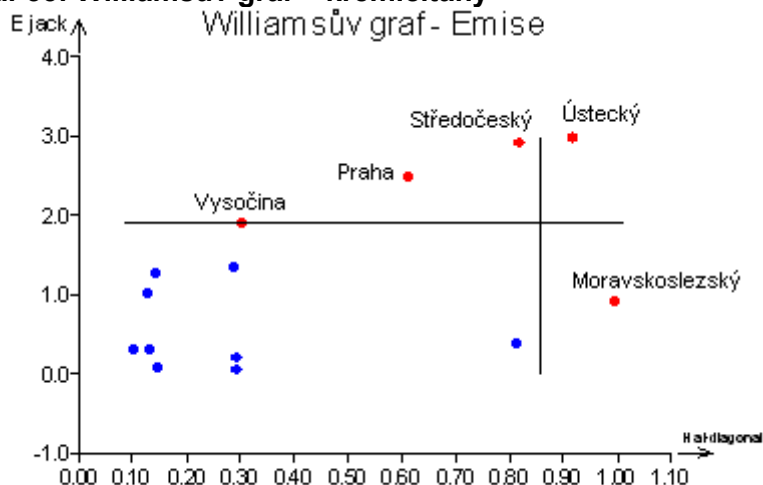
Testování vlivných bodů diagnostickými grafy

Graf 32: Predikovaná rezidua – křemičitany



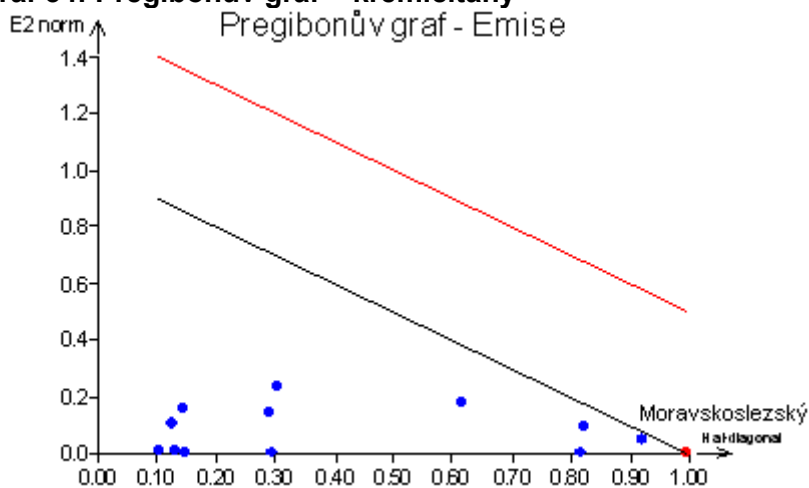
V grafu Predikovaných reziduí je bod náležící Moravskoslezskému kraji podezřelý z odlehlosti (vzdálen od ostatních)

Graf 33: Williamsův graf – křemičitany



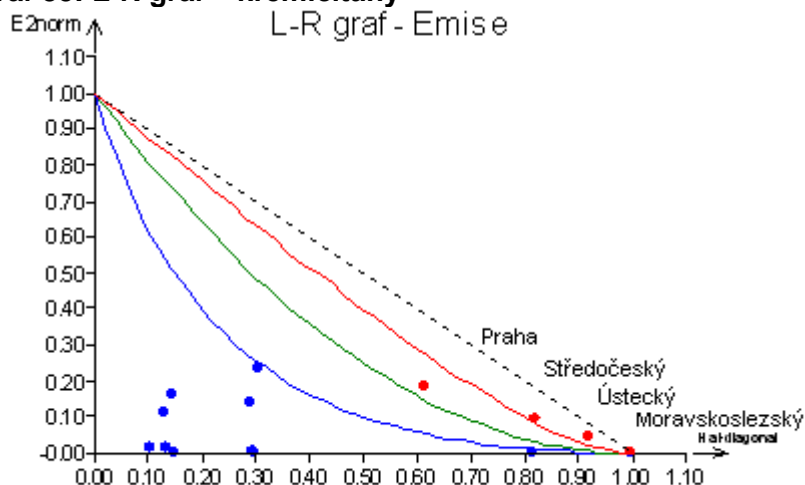
Ve Williamsově grafu jsou jako odlehlé body vyhodnoceny kraje Středočeský, Vysočina a Praha. Ústecký kraj je odlehlý a zároveň extrém, protože se nachází nad vodorovnou linií a vpravo od svislé linie. Moravskoslezský kraj je extrém.

Graf 34: Pregibonův graf – křemičitany



V Pregibonově grafu byl Moravskoslezský kraj vyhodnocen jako středně vlivný bod, protože se nachází nad dolní linií.

Graf 35: L-R graf – křemičitany



Podle vyhodnocení L-R grafu jsou kraje Moravskoslezský, Ústecký, Středočeský a Praha vyhodnoceny jako odlehlé body (nachází se nad křivkami oddělujícími oblast odlehlých bodů).

Závěr testování vlivných bodů.: Kraje Moravskoslezský, Ústecký, Středočeský a Praha se jeví jako odlehlé body. Nicméně je nebudeme odstraňovat, protože z podstaty dat je nemožné nevzít v úvahu informace o emisích z těchto krajů.

4.2 Kritika modelu a metody

Statistické testy

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 65.9487202
Kvantil F (1-alfa, m-1, n-m) : 3.687498666
Pravděpodobnost : 2.76008948E-006
Závěr : **Model je významný**

Scottovo kritérium multikolinearity

Hodnota kritéria SC : 0.4197460952
Závěr : **Model vykazuje multikolinearitu!**

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0.1244181986
Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
Pravděpodobnost : 0.7242911367
Závěr : **Rezidua vykazují homoskedasticitu.**

Jarque-Berrův test normality

Hodnota kritéria JB : 0.3816180382
Kvantil $\chi^2(1-\alpha,2)$: 5.991464547
Pravděpodobnost : 0.8262903788
Závěr : **Rezidua mají normální rozdělení.**

Waldův test autokorelace

Hodnota kritéria WA : 0.0002010380115
Kvantil $\chi^2(1-\alpha,1)$: 3.841458829
Pravděpodobnost : 0.9886873435
Závěr : **Autokorelace je nevýznamná**

Durbin-Watsonův test autokorelace

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

Znaménkový test reziduí

Hodnota kritéria Sg : 0.8345229604
Kvantil $N(1-\alpha/2)$: 1.959963999
Pravděpodobnost : 0.4039863545
Závěr : **V reziduích není trend.**

Statistické charakteristiky regrese

	Původní model	Bez abs.členu a proměn. TZL, CO a NH ₃
RSC	99952948.42	152816878
S ² (e)	12494118.55	12734739.83
R	0.9880857214	0.981725943
R ²	0.9763133928	0.9637858271
Rp	0.08721828627	0.7324446935
MEP	390430984.6	43455038.07
AIC	232.936139	230.8797098

RSC Reziduální součet čtverců (RSC): optimální je hodnota co nejnižší
S²(e) Reziduální rozptyl: optimální je hodnota co nejnižší
R Vícenásobný korelační koeficient: optimální je hodnota co nejbližší 1
R² Koeficient determinace: optimální je hodnota co nejbližší 1
Rp Predikovaný korelační koeficient: optimální je hodnota co nejbližší 1
MEP Střední kvadratická chyba predikce: optimální je hodnota co nejbližší 0
AIC Akaikeho informační kritérium: optimální je hodnota co nejzápornější (tedy nejnižší)

4.3 Opravený model (bez abs. členu a proměnných TZL, CO a NH₃)

Odhady parametrů

Parametr	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
β_2	0.8438306969	0.05996361073	Významný	8.046381561E-009	0.7131812125	0.9744801812
β_4	0.7076475216	0.09409584044	Významný	7.03811463E-006	0.5026302972	0.912664746

Hladina významnosti $\alpha = 0,05$ (95% jistota)

$$NO_x = \beta_2(SO_2) + \beta_4(VOC)$$

Parametry β_2 a β_4 jsou vyhodnoceny jako významné, protože jejich intervaly spolehlivosti neobsahují 0 resp. p-hodnota testu je nižší než zadaná hladina významnosti 0,05.

4.4 Závěr

Vícerozměrnou lineární regresní analýzou bylo zjištěno, že množství emisí oxidů dusíku je v korelaci s emisemi oxidu siřičitého a těkavých organických látek. Všechny tyto polutanty jsou typické pro průmyslovější oblasti. Emise NH_3 jsou odlišné tím, že se na nich navíc podílí i zemědělská výroba, takže vyšší emise NH_3 mohou mít i oblasti méně průmyslové. Koncentraci emisí tuhých znečišťujících látek a oxidu uhelnatého zřejmě ovlivňují ještě další faktory.

Literatura a software

- [1] Meloun M., Militký J.: Kompendium statistického zpracování dat. Academia, Praha 2002.
- [2] Meloun M., Militký J.: Statistické zpracování dat. East publishing, Praha 1998.
- [3] QCExpert 2.7 (<http://www.trilobyte.cz/>)
- [4] Adstat 1.25 (<http://www.trilobyte.cz/>)