

Univerzita Pardubice

**Fakulta chemicko-technologická
Katedra analytickej chémie**

**“ Interaktívna štatistická analýza dát“
(Licenčné štúdium)**

4. semestrálna práca

Tvorba lineárnych regresných modelov

Obsah

Úloha 1 Porovnanie dvoch regresných priamok	3
1.1 Zadanie.....	3
1.2 Dáta.....	3
1.3 Riešenie.....	3
1.3.1 Kritika dát, modelu a metódy pre L-homocysteín – rok 2013.....	3
1.3.2 Kritika dát, modelu a metódy pre L-homocysteín – rok 2015.....	10
1.3.3 Kritika dát, modelu a metódy pre L-metionín – rok 2013.....	17
1.3.4 Kritika dát, modelu a metódy pre L-metionín – rok 2015.....	23
1.3.5 Porovnanie dvoch regresných kriviek.....	30
1.3.5.1 Test zhodnosti rozptylov.....	30
1.3.5.2 Chowov test zhody dvoch lineárnych modelov.....	31
Úloha 2 Určenie stupňa polynómu.....	32
2.1 Zadanie.....	32
2.2 Dáta.....	32
2.3 Návrh modelu.....	33
Úloha 3 Viacrozmerný lineárny regresný model.....	38
3.1 Zadanie.....	38
3.2 Dáta.....	39
3.3 Riešenie.....	39
Úloha 4 Validácia novej analytickej metódy.....	44
4.1 Zadanie.....	44
4.2 Dáta.....	44
4.3 Riešenie.....	45

1. úloha: Porovnanie dvoch regresných priamok pri jednoduchom lineárnom regresnom modeli

1.1 Zadanie

Kontrola chemickej čistoty rádiofarmaka [11C]-metionínu, injekcia (11C-MET) metódou ionexovej kvapalinovej chromatografie spočíva v stanovení koncentrácie L-homocysteínu (LHC), jedného z prekursorov výroby spomínaného rádiofarmaka, neaktívneho D,L-metionínu (MET) a L-homocysteín tiolaktónu (TLN). Porovnajzte regresné priamky pre LHC a MET získané na ionexovej kolóne v septembri 2013 a v januári 2015 a zhodnoťte či je medzi nimi výrazný rozdiel v selektivite oboch analytov a či sa výrazne zmenila životnosť kolóny. Vyšetrite regresný triplet pre oba analyty, nájdite prípadné vybočujúce body a či sú splnené požiadavky pre použitie metódy najmenších štvorcov, MNŠ.

1.2 Dáta

Tab. 1: Dáta namerané v septembri 2013 (darina4A.txt, DARINA4A, darina4B.txt, DARINA4B darina_4_sem_praca.xls)

c(LHC), ug/ml	A(LHC), (μV.s)	c(MET), ug/ml	A(MET), (μV.s)
25,00	19680,91	25,00	53904,50
50,00	38513,51	50,00	108222,64
75,00	56606,80	75,00	160284,04
100,00	76803,96	100,00	220341,64
150,00	115759,29	150,00	329783,53
200,00	155005,80	200,00	432808,30

Tab. 2: Dáta namerané v januári 2015 (darina4C.txt, DARINA4D, darina4D.txt, DARINA4D, darina_4_sem_praca.xls)

c(LHC), ug/ml	A(LHC), (μV.s)	c(MET), ug/ml	A(MET), (μV.s)
25,00	19744,26	25,00	52143,48
50,00	38138,19	50,00	101960,28
75,00	54821,30	75,00	152615,75
100,00	72730,95	100,00	205199,21
150,00	112490,10	150,00	306630,64
200,00	146208,08	200,00	411605,19

1.3 Riešenie

Regresná diagnostika pozostáva z pomôcok a postupov pre interaktívnu analýzu dát, modelu a metódy, ktoré predstavujú tzv. regresný triplet. Regresná diagnostika dát sa skladá z tabuliek rôznych druhov reziduí a grafických diagnostík. Pri riešení sa použil predovšetkým program QC Expert (viď nižšie uvedené výstupy) a pre porovnanie aj program ADSTAT.

1.3.1 Kritika dát, modelu a metódy pre L-homocysteín (LHC) – rok 2013

Testovanie úseku a smernice

Pre vyššie uvedené zadanie bol navrhnutý model v tvare $y = \beta_0 + \beta_1 \cdot x$. Previedla sa predbežná diagnostika dát pomocou programu QC Expert ver.2.5.

Tab 3: Odhady parametrov pre model LHC_2013 s absolútnym členom.

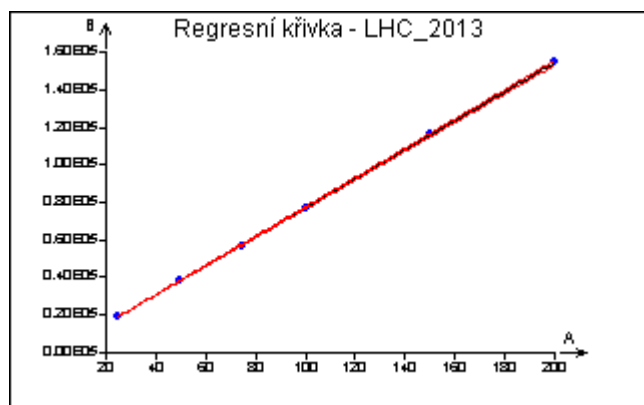
Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_0	-451,843843	568,2374367	Nevýznamný	0,4710332877	-2029,5239	1125,8362
β_1	775,1355318	4,883079365	Významný	9,447152172 E-009	761,5779	788,6931

Úsek β_0 (absolútny člen) je štatisticky nevýznamný (nulový), testom vyhodnotený interval spoľahlivosti zahŕňa 0. Smernica β_1 je štatisticky významná, nie je jednotková.

V ďalšom testovaní rezíduí sa pokračuje bez **absolútneho člena**, keďže je **štatisticky nevýznamný**.

Tab 4: Odhady parametrov pre model LHC_2013 bez absolútneho člena.

Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	771,7988388	2,403668011	Významný	5,560440997 E-012	765,6200134	777,9776641



Graf 1: Regresná krivka pre LHC nameraná v septembri 2013.

Konfidenčný pás (vymedzený červenými krivkami) je veľmi úzky, dáta vykazujú silnú linearitu.

Analýza klasických rezíduí

Index	Y naměřené	Y vypočítané	Směr. odch. Y	Reziduum	Reziduum [%Y]
1	19680,91	19294,97097	60,09170027	385,9380308	1,960976654
2	38513,51	38589,94194	120,1834005	-76,42793846	-0,1984444693
3	56606,80	57884,91291	180,2751008	-1278,117908	-2,257887781
4	76803,96	77179,88388	240,3668011	-375,9268769	-0,4894629022
5	115759,29	115769,8258	360,5502016	-10,53981538	-0,009104941598
6	155005,80	154359,7678	480,7336021	646,0272462	0,4167761897

Reziduální součet čtverců : 2347158,087

Průměr absolutních reziduí : 462,1629692

Reziduální směr. odchylka : 685,150799

Reziduální rozptyl : 469431,6173

Šikmost reziduí : 1,527504496

Šířka reziduí : 3,142017178

Analýza rezíduí

Index	Standardní	Jackknife	Predikované	Diag(Hii)	Diag(H*ii)	Cook. vzdál.
1	0,5655	0,5228	388,9302	0,00769	0,07115	0,00438
2	-0,1133	-0,1015	-78,8597	0,03077	0,03326	-0,00356
3	-1,9336	-3,4434	-1373,1812	0,06923	0,7652	-0,1438
4	-0,5859	-0,5430	-428,6881	0,1231	0,1833	-0,0822
5	-0,01809	-0,0162	-14,5762	0,2769	0,2770	-0,00693
6	1,3233	1,4684	1272,4776	0,4923	0,6701	1,2832

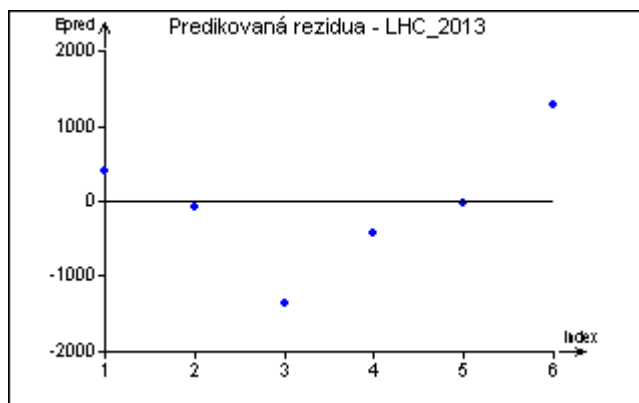
Hodnoty sú zaokrúhlené na 4, resp. 5 desatinných miest podľa matematických pravidiel.

Analýza vplyvu

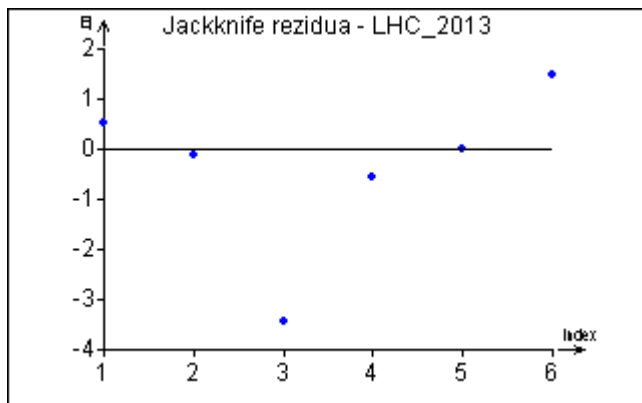
Index	Atkinson. vzdál.	Andrews-Pregibon st.	Vliv na Y^{\wedge}	Vliv na param. LD(b)	Vliv na rozptyl LD(s)	Celk. vliv LD(b,s)
1	0,1029	0,9288	0,0460	0,00297	0,0390	0,0417
2	0,0404	0,9667	-0,0181	0,00049	0,0914	0,0918
3	2,0999	0,2348	-0,9391	0,3248	6,6516	7,7540
4	0,4549	0,8167	-0,2034	0,0575	0,0358	0,0875
5	0,0224	0,7230	-0,0100	0,0002	0,0939	0,0940
6	3,2332	0,3299	1,4459	1,7543	0,2022	2,8156

Testovanie vplyvných bodov pomocou diagnostických grafov – LHC_2013

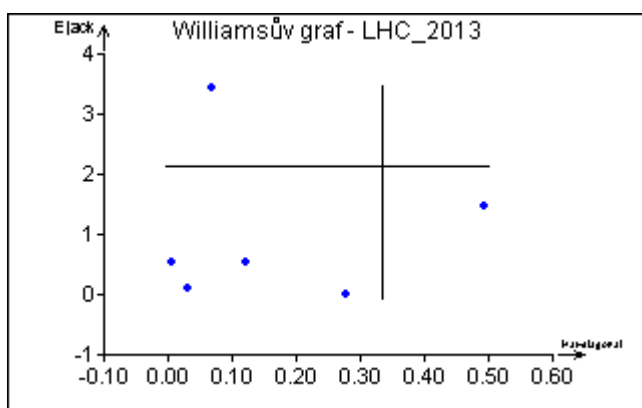
Grafy vplyvných bodov indikujú prítomnosť vplyvných bodov a extrémov. K najviac používaným patria nasledujúce grafy.



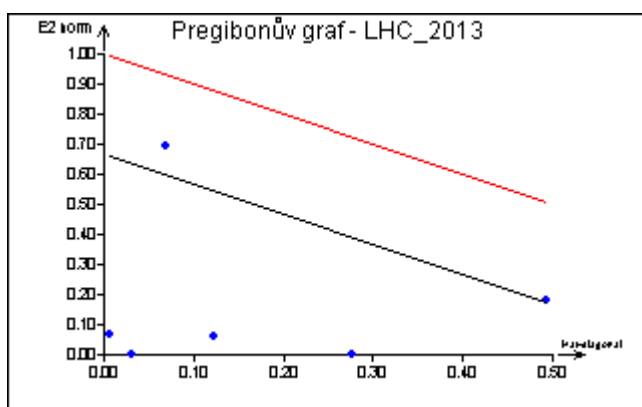
Graf 2: Graf predikovaných rezíduí poukazuje na body 3 a 6 ako odľahlé, keďže sú vzdialené od ostatných bodov.



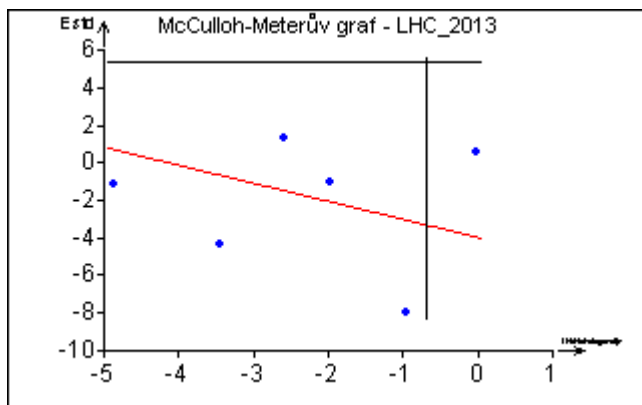
Graf 3: Graf Jackknife reziduí identifikuje bod 3 ako podozrivý z odľahlosti vzhľadom na jeho vzdialenosť od ostatných bodov.



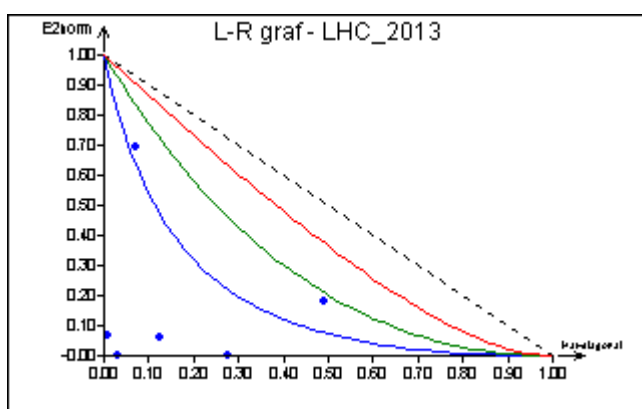
Graf 4: Wiliamsonov graf indikuje bod č. 3 ako odľahlý bod – nachádza sa nad vodorovnou čiarou. V datach sa nachádza jeden extrém (za zvislou čiarou) – bod 6. Ostatné dáta vykazujú mrak.



Graf 5: Pregibonov graf vykazuje jeden stredne vplyvný bod 3 – nachádza sa v oblasti medzi čiernou a červenou priamkou a žiadne silno vplyvné body (nachádzajú sa nad červenou priamkou).



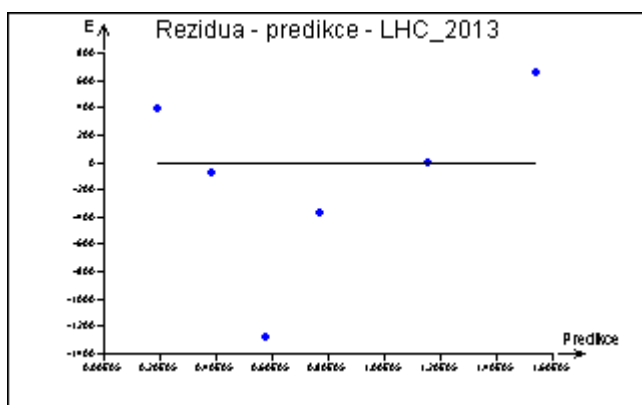
Graf 6: McCullohov- Meeterov graf predikuje body 3, 4 ako podozrivé body, keďže ležia v oblasti ohraničenej všetkými troma líniami. Bod 6 je však indikovaný ako extrém (nachádza sa za zvislou líniou).



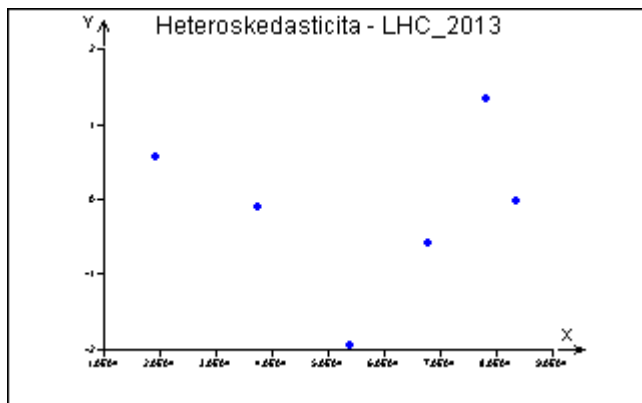
Graf 7: L-R graf poukazuje na body 3,6 ako podozrivé body.

Kritika modelu a metódy (LHC_2013)

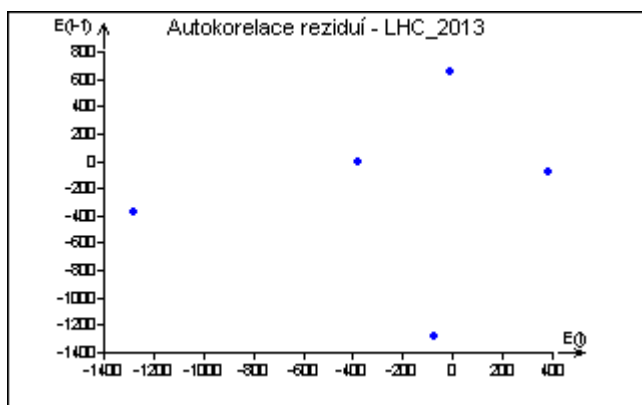
Diagnostické grafy



Graf 8: Rezidua – predikce. Body v grafe kmitajú po oboch stranách vodorovnej priamky, čo signalizuje, že navrhnutý model je správny.



Graf 9: Heteroskedasticita (nekonštantný rozptyl). Ak body tvoria výseč alebo trojuholník, ide o heteroskedasticitu. Body naznačujú pravdepodobne homoskedasticitu.



Graf 10: Autokorelácia reziduí. Body tvoria mrak, majú nenulovú smernicu.

Štatistické testy

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 27197,62553
 Kvantil F (1-alfa, m-1, n-m) : 6,607890974
 Pravdepodobnost : 1,555268283E-010

Závěr : Model je významný

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0,0111147001
 Kvantil $\chi^2(1-\text{alfa}, 1)$: 3,841458829
 Pravdepodobnost : 0,9160375654

Závěr : Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 1,532546715
 Kvantil $\chi^2(1-\text{alfa}, 2)$: 5,991464547
 Pravdepodobnost : 0,4647417715

Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA : 0,3430117373
 Kvantil $\chi^2(1-\text{alfa}, 1)$: 3,841458829
 Pravdepodobnost : 0,5580959275

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 : 0,1767766953

Kvantil $N(1-\alpha/2)$: 1,959963999

Pravděpodobnost : 0,8596837952

Závěr : V reziduích není trend.

Štatistické charakteristiky regresie

V tab. 5 sú zhrnuté niektoré regresné charakteristiky pôvodného modelu s absolútnym členom, bez absolútného člena a s vylúčením bodu 6, ktorý sa javí ako najviac vplyvný.

Tab. 5: Zhrnutie riešenia regresných modelov pre LHC_2013

	Pôvodný model s abs. členom	Bez abs. člena	Bez abs. člena a bez bodu 6
RSC	2026779,447	2347158,087	1525102,687
S²(e)	506694,8617	469431,6173	381275,6718
R	0,9999206384	0,9999080929	0,9998608243
R²	0,999841283	0,9998161942	0,9997216679
R_p	0,9991936854	0,9993976827	0,9990087494
MEP	858210,8445	641051,0822	543284,0675
AIC	80,38119486	79,26173816	65,14067195
Tvar modelu	y= - 451,8438(568,2374)+775,13 55 (4,8831)x	y= 771,7988 (2,4037)x	y= 768,6666 (3,0402)x

Reziduálny súčet štvorcov, RSC:

optimálna hodnota je čo najnižšia

Reziduálny rozptyl, S²(e):

optimálna hodnota je čo najnižšia

Vícenásobný korelačný koeficient R :

optimálna hodnota je čo najbližšie 1

Koeficient determinácie R² :

optimálna hodnota je čo najbližšie 1

Predikovaný korelačný koeficient R_p :

optimálna hodnota je čo najbližšie 1

Stredná kvadratická chyba predikcie MEP :

optimálna hodnota je čo najbližšie 0

Akaikeho informační kritérium :

optimálna hodnota je čo najzápornejšia (najnižšia)

Absolútna hodnota párového korelačného koeficientu R poukazuje, že navrhnutý lineárny regresný model je štatisticky významný. Vysoká hodnota koeficienta determinácie R² predstavuje % variability vysvetlené modelom. Predikčná schopnosť modelu je vysvetlená predikovaným korelačným koeficientom R_p. Stredná kvadratická chyba predikcie MEP a Akaikeho informačné kritérium, AIC sa používajú na rozlíšenie medzi viacerými navrhnutými modelmi. Za optimálny sa považuje ten model, pre ktorý MEP a AIC dosahuje minimálnu hodnotu.

Konštrukcia spresneného modelu

Na základe analýzy vplyvných bodov a grafických znázornení bol bod č. 6 vynechaný a vypočítali sa nové parametre modelu (Tab. 6).

Tab 6: Lineárny regresný model pre LHC_2013 bez absolútneho člena a bez bodu 6.

Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	768,6666	3,04023	Významný	1,4682 E-009	760,2255	777,1076

Záver: Pri vyšetrení navrhnutého lineárneho regresného modelu pre stanovenie LHC_2013 a prevedení

regresného tripletu boli prevedené nasledujúce kroky:

- absolútny člen β_0 je štatistický nevýznamný na hladine významnosti $\alpha=0,05$, utápa sa vo svojom šume, o tento člen sa regresný model upravil;
- na základe viacerých analýz rezíduí a grafov bol vylúčený bod 6 a je považovaný ako vysoko podozrivý bod až extrém a pre porovnanie bol vypočítaný regresný model;
- model po vynechaní podozrivého bodu 6 dosahuje o niečo lepšie hodnoty RSC, $S^2(e)$, MEP a AIC, ale hodnoty R, R^2 a R_p majú nižšie hodnoty v porovnaní s pôvodným modelom;
- pri porovnaní pôvodného modelu s absolútnym členom a bez absolútneho člena (viď tab. 5), pôvodný model dosahuje lepšie hodnoty RSC, R, R^2 a R_p , hodnoty AIC sú takmer porovnateľné, väčší rozdiel je v hodnotách MEP;
- vzhľadom na malý počet bodov a aj napriek nevýraznému zlepšeniu hodnôt MEP, AIC a niektorých reziduálnych štvorcov sa navrhuje model bez absolútneho člena a bez vylúčenia bodu 6 v tvare

$$y = 771,7988 (2,4037).x$$

- použitie metódy najmenších štvorcov má svoje opodstatnenie.

1.3.2 Kritika dát, modelu a metódy pre L-homocysteín (LHC) – rok 2015

Odhady parametrov

Tab. 7: Odhady parametrov pre navrhnutý model $y = \beta_0 + \beta_1 \cdot x$.

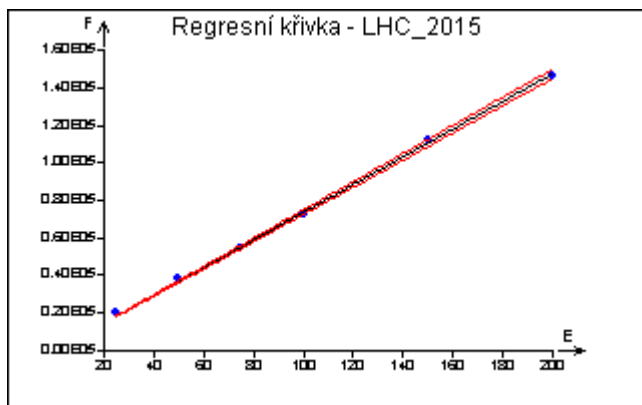
Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_0	1141,9043	1106,0438	Nevýznamný	0,3602	-1928,9657	4212,7743
β_1	728,8024	9,5047	Významný	1,7337 E-007	702,4133	755,1916

Úsek β_0 (absolútny člen) je štatisticky nevýznamný (nulový), testom vyhodnotený interval spoľahlivosti zahŕňa 0. Smernica β_1 je štatisticky významná, nie je jednotková.

V ďalšom testovaní rezíduí a vplyvných bodov je **absolútny člen** vynechaný, keďže je **štatisticky nevýznamný**.

Tab. 8: Odhady parametrov pre model bez absolútneho člena.

Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	737,2349477	4,892684216	Významný	2,442335223E-010	724,6579025	749,8119929



Graf 11: Regresná krivka pre LHC nameraná v januári 2015.

Konfidenčný pás (vymedzený červenými krivkami) je úzky, nepatrne sa rozširuje v hornej oblasti, dáta vykazujú linearitu.

Analýza klasických reziduí

Index	Y naměřené	Y vypočítané	Směr. odch. Y	Reziduum	Reziduum [%Y]
1	19744,26	18430,87369	122,3171054	1313,386308	6,651990542
2	38138,19	36861,74738	244,6342108	1276,442615	3,346888291
3	54821,3	55292,62108	366,9513162	-471,3210769	-0,8597407886
4	72730,95	73723,49477	489,2684216	-992,5447692	-1,364680056
5	112490,1	110585,2422	733,9026324	1904,857846	1,69335599
6	146208,08	147446,9895	978,5368432	-1238,909538	-0,8473605142

Reziduální součet čtverců : 9724958,279

Průměr absolutních reziduí : 1199,577026

Reziduální směr. odchylka : 1394,629577

Reziduální rozptyl : 1944991,656

Šikmost reziduí : 0,446484865

Špičatost reziduí : 1,406623836

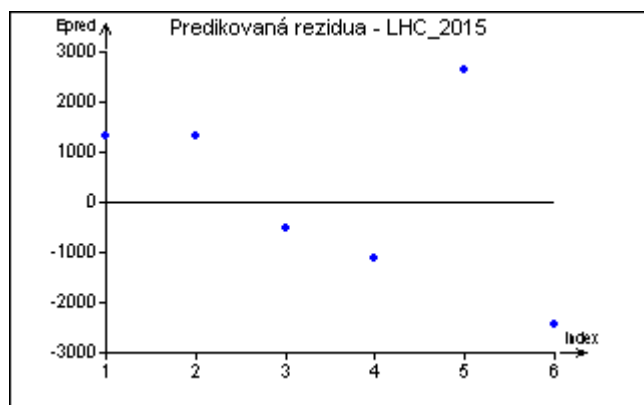
Analýza reziduí

Index	Standardní	Jackknife	Predikované	Diag(Hii)	Diag(H*ii)	Cook. vzdál.
1	0,945388759	0,9330786669	1323,567597	0,0076923077	0,1850692736	0,00733
2	0,92967002	0,9142892061	1316,964603	0,0307692308	0,1983078159	0,02951
3	-0,35029745	-0,317232455	-506,378017	0,0692307692	0,09207339243	-0,02606
4	-0,75999456	-0,722788216	-1131,84929	0,1230769231	0,224377627	-0,1067
5	1,606243744	2,065073263	2634,377872	0,2769230769	0,650033512	0,6152
6	-1,24675303	-1,34331408	-2440,27636	0,4923076923	0,6501383791	-1,2090

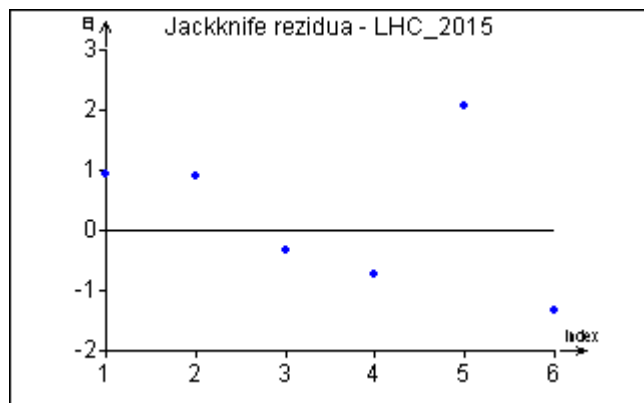
Analýza vplyvu

Index	Atkinson. vzdál.	Andrews- Pregibon st.	Vliv na Y^{\wedge}	Vliv na param. LD(b)	Vliv na rozptyl LD(s)	Celk. vliv LD(b,s)
1	0,1836996831	0,8149307264	0,08215299579	0,008308290672	0,0006433595297	0,0091
2	0,3642615388	0,8016921841	0,1629027125	0,03283518446	0,0001672114992	0,0333
3	0,19346	0,9079266076	-0,08651794	0,01094249269	0,07063791396	0,0800
4	0,605485338	0,775622373	-0,270781275	0,09649843504	0,01043679949	0,1021
5	2,857641797	0,349966488	1,277976263	1,082009879	1,070519559	3,1120
6	2,957880255	0,3498616209	-1,322804264	1,580911329	0,1155180596	2,3028

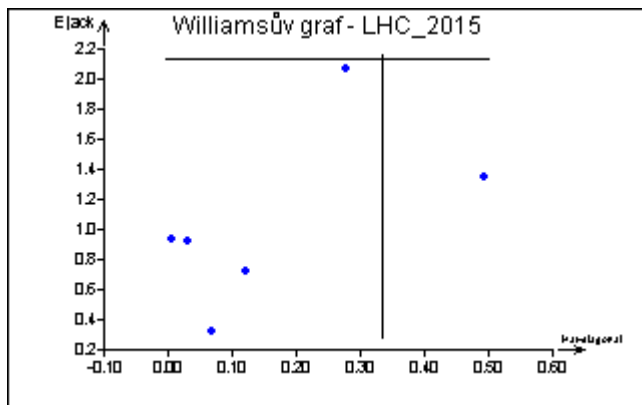
Testovanie vplyvných bodov pomocou diagnostických grafov – LHC_2015



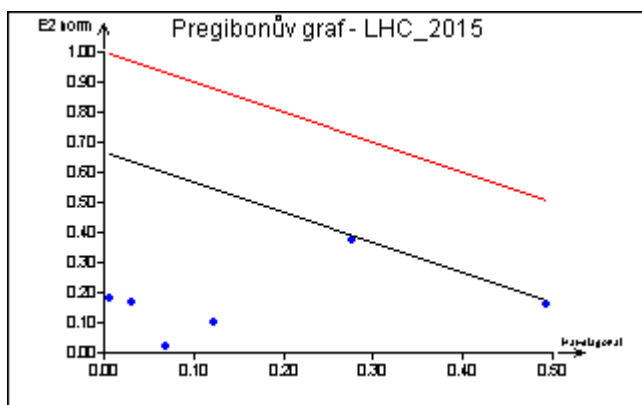
Graf 12: Graf predikovaných rezíduí jednoznačne nepoukazuje odľahlý bod.



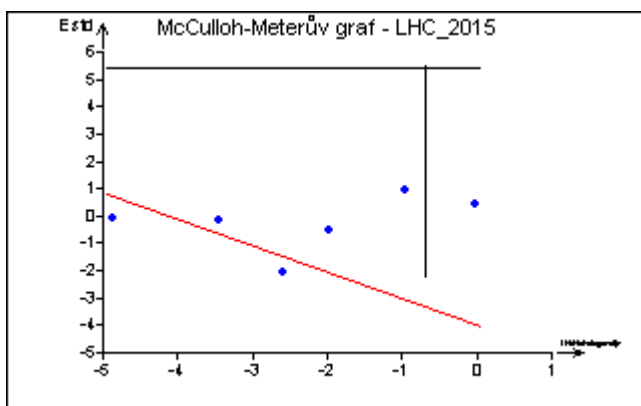
Graf 13: Graf Jackknife rezíduí neindikuje žiadny odľahlý bod.



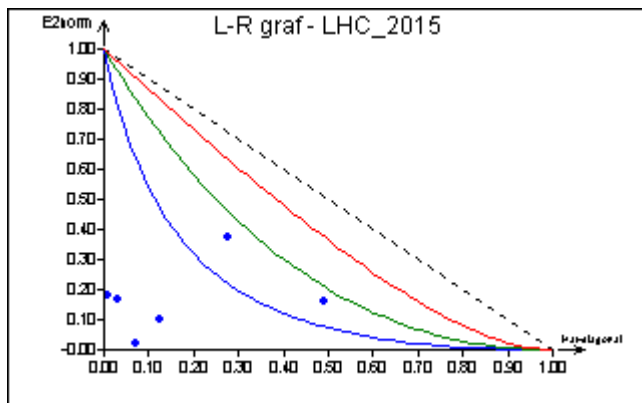
Graf 14: Wiliamsonov graf indikuje bod č. 6 ako extrém – nachádza sa za zvislou čiarou. Ostatné data vykazujú mrak. V dátach sa nenachádza žiaden odľahlý bod (nad vodorovnou čiarou).



Graf 15: Pregibonov graf navykazuje žiadne stredne vplyvné body (oblasť medzi čiernou a červenou priamkou) ani silno vplyvné body (nachádzajú sa nad červenou priamkou).



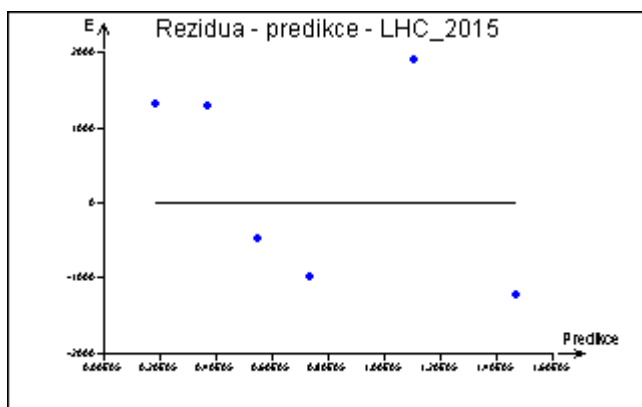
Graf 16: McCullohov- Meeterov graf indikuje body 2,4,5 ako podozrivé body, keďže ležia v oblasti ohraničenej všetkými troma líniami. Bod 6 je indikovaný ako extrém (nachádza sa za zvislou líniou).



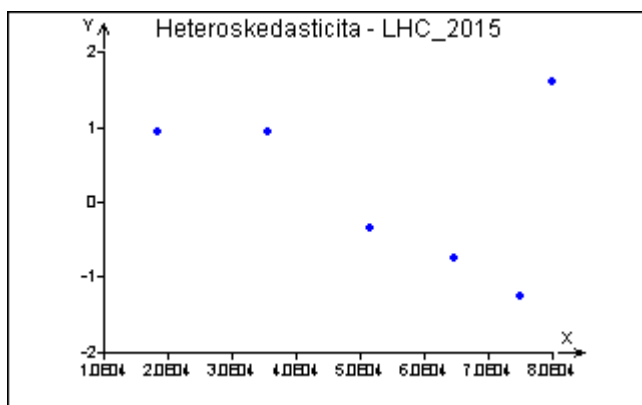
Graf 17: L-R graf poukazuje na body 5 a 6 ako na podozrivé.

Kritika modelu a metódy (LHC_2015)

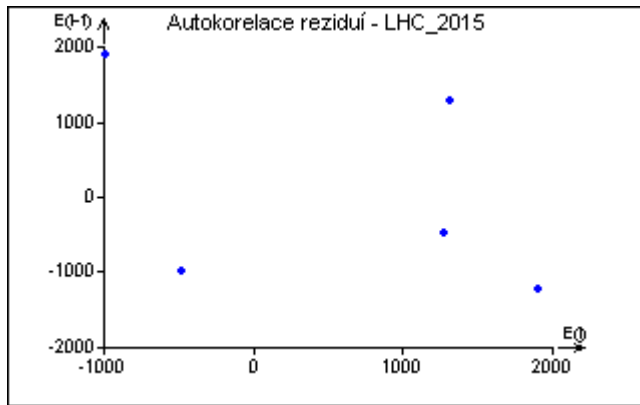
Diagnostické grafy



Graf 18: Rezidua – predikce. Body v grafe kmitajú po oboch stranách vodorovnej priamky, čo signalizuje, že navrhnutý model je správny.



Graf 19: Heteroskedasticita (nekonštantný rozptyl). Ak body tvoria výseč alebo trojuholník, ide o heteroskedasticitu. Body naznačujú trojuholník, heteroskedasticita je pravdepodobná.



Graf 20: Autokorelácia reziduí. Body tvoria mrak, majú nenulovú smernicu.

Štatistické testy

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 5802,058041
 Kvantil F (1-alfa, m-1, n-m) : 6,607890974
 Pravděpodobnost : 7,388342849E-009

Závěr : Model je významný

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0,0816555375
 Kvantil $\chi^2(1-\text{alfa},1)$: 3,841458829
 Pravděpodobnost : 0,775066289

Závěr : Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 1,081196765
 Kvantil $\chi^2(1-\text{alfa},2)$: 5,991464547
 Pravděpodobnost : 0,5823996504

Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA : 0,5043210894
 Kvantil $\chi^2(1-\text{alfa},1)$: 3,841458829
 Pravděpodobnost : 0,4776076013

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 : 0,4564354646
 Kvantil $N(1-\text{alfa}/2)$: 1,959963999
 Pravděpodobnost : 0,6480768681

Závěr : V reziduích není trend.

Štatistické charakteristiky regresie

Tab. 9: Zhrnutie vybraných štatistických charakteristík navrhovaných modelov.

	Bez abs. členu	Bez abs. členu a bez bodu 5	Pôvodný model s abs. členom
RSC	9724958,279	4706842,919	7678766,939
S²(e)	1944991,656	1176710,73	1919691,735
R	0,9995693967	0,9997527337	0,9996600137
R²	0,9991389788	0,9995055286	0,999320143
R_p	0,996829585	0,9988017903	0,9970562132
MEP	2986437,331	1140910,38	2772802,838
AIC	87,79068013	70,77545018	88,37326041
Tvar modelu	$y = 737,2349(4,8927)x$	$y = 732,3715(4,754)x$	$y = 1141(1106) + 728,8024(9,5047)x$

Reziduálny súčet štvorcov, RSC: optimálna hodnota je čo najnižšia
 Reziduálny rozptyl, S²(e): optimálna hodnota je čo najnižšia
 Vícenásobný korelačný koeficient R : optimálna hodnota je čo najbližšie 1
 Koeficient determinácie R² : optimálna hodnota je čo najbližšie 1
 Predikovaný korelačný koeficient R_p : optimálna hodnota je čo najbližšie 1
 Stredná kvadratická chyba predikcie MEP : optimálna hodnota je čo najbližšie 0
 Akaikeho informačné kritérium : optimálna hodnota je čo najzápornejšia (najnižšia)

Absolútna hodnota párového korelačného koeficientu R poukazuje, že navrhnutý lineárny regresný model je štatisticky významný. Vysoká hodnota koeficienta determinácie R² predstavuje % variability vysvetlené modelom. Predikčná schopnosť modelu je vysvetlená predikovaným korelačným koeficientom R_p. Stredná kvadratická chyba predikcie MEP a Akaikeho informačné kritérium, AIC sa používajú na rozlíšenie medzi viacerými navrhnutými modelmi. Za optimálny sa považuje ten model, pre ktorý MEP a AIC dosahuje minimálnu hodnotu.

Konštrukcia spresneného modelu

Na základe analýzy vplyvných bodov a grafických znázornení bol bod č. 5 vynechaný a vypočítali sa nové parametre modelu (tab. 10).

Tab. 10: Odhad parametrov pre model bez absolútneho členu a bez bodu č. 5.

Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	732,3715	4,4754	Významný	8,364522E-009	719,9458	744,79716

Záver: Pri vyšetrení navrhnutého lineárneho regresného modelu pre stanovenie LHC-rok 2015 a prevedení regresného tripletu boli prevedené nasledujúce kroky:

- absolútny člen β_0 je štatisticky nevýznamný na hladine významnosti $\alpha=0,05$, o tento člen sa regresný model upravil;
- na základe viacerých analýz reziduí a grafov bol vylúčený bod 5 a je považovaný ako vysoko podozrivý bod až extrém oveľa viac ako bod 6;

- c) model po vynechaní podozrivého bodu 5 dosahuje v niektorých štatistických hodnotách výrazne lepšie hodnoty RSC, $S^2(e)$, MEP a AIC, ale aj hodnôt R, R^2 a R_p v porovnaní s pôvodným modelom;
- d) aj napriek malému počtu bodov sa navrhuje model bez bodu 5 v tvare

$$y = 732,3715(4,4754).x$$
- e) použitie metódy najmenších štvorcov má svoje opodstatnenie.

1.3.3 Kritika dát, modelu a metódy pre D,L-metionín (MET) – rok 2013

Testovanie úseku a smernice

Pre vyššie uvedené zadanie bol navrhnutý model v tvare $y = \beta_0 + \beta_1 \cdot x$. Previedla sa predbežná diagnostika dát pomocou programu QC Expert ver.2.5 postupne pre MET pre rok 2013 a 2015. Pre porovnanie sa dáta vyhodnotili aj pomocou programu ADSTAT, v práci nie sú ukázané.

Odhady parametrov

Tab. 11: Odhady parametrov pre navrhovaný model $y = \beta_0 + \beta_1 \cdot x$.

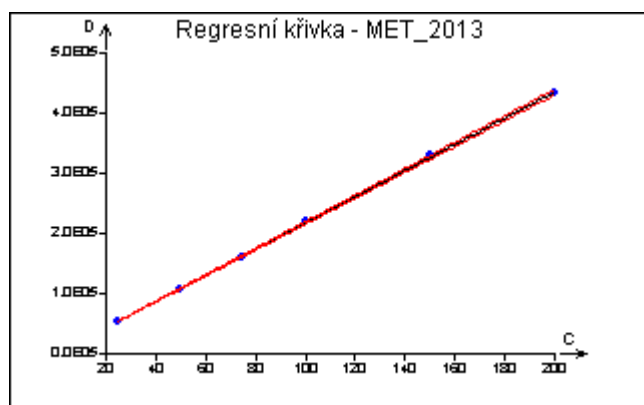
Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_0	-366,9647	2311,3120	Nevýznamný	0,8815	-6784,1956	6050,2662
β_1	2179,2441	19,8620	Významný	4,1379E-008	2124,0984	2234,3898

Úsek β_0 je štatisticky nevýznamný, testom vyhodnotený interval spoľahlivosti zahŕňa 0. Smernica β_1 je štatisticky významná, nie je jednotková.

V ďalšom testovaní rezíduí a vplyvných bodov sa vynechá **absolútny člen**, keďže je **štatisticky nevýznamný**.

Tab. 12: Odhady parametrov pre model bez absolútneho člena.

Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	2176,534174	9,113798847	Významný	2,442823721 E-011	2153,106408	2199,96194



Graf 21: Regresná krivka zostrojená z dát pre MET nameraných v septembri 2013. Konfidenčný pás (vymedzený červenými krivkami) je veľmi úzky, dáta vykazujú linearitu.

Analýza klasických reziduí

Index	Y naměřené	Y vypočítané	Směr. odch. Y	Reziduum	Reziduum [%Y]
1	53904,5	54413,3544	227,8450	-508,8573462	-0,9439979491
2	108222,64	108826,7087	455,6899	-604,06669	-0,5581703432
3	160284,04	163240,063	683,5349	-2956,027038	-1,844242953
4	220341,64	217653,4174	911,37988	2688,225615	1,220026128
5	329783,53	326480,1261	1367,0698	3303,407923	1,00168977
6	432808,3	435306,8348	1822,7598	-2498,534769	-0,5772843934

Reziiduální součet čtverců : 33743665,08
 Průměr absolutních reziduí : 2093,186564
 Reziiduální směr. odchylka : 2597,832369
 Reziiduální rozptyl : 6748733,016
 Šikmost reziduí : 0,02928985246
 Špičatost reziduí : 1,511448728

Analýza reziduí

Index	Standardní	Jackknife	Predikované	Diag(Hii)	Diag(H*ii)	Cook. vzdál.
1	-0,1966	-0,1766	-512,7988	0,00769	0,0154	-0,0015
2	-0,2362	-0,2124	-623,2451	0,03077	0,0416	-0,0075
3	-1,17944	-1,2417	-3175,8919	0,0692	0,3282	-0,0877
4	1,1050	1,1369	3065,5179	0,1231	0,3372	0,1551
5	1,4954	1,7990	4568,5389	0,2769	0,6003	0,5727
6	-1,3498	-1,5144	-4921,3533	0,4923	0,6773	-1,3089

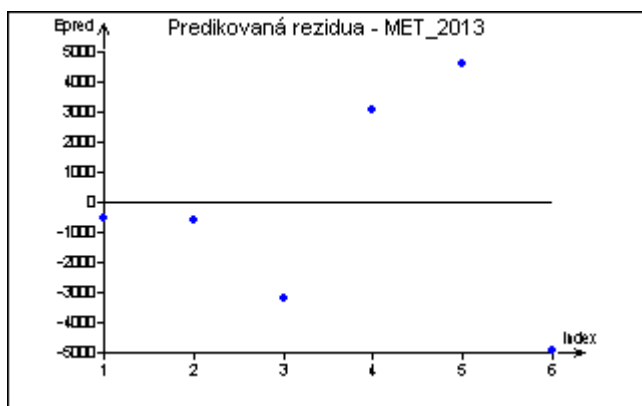
Hodnoty sú zaokrúhlené na 4 desatinné miesta podľa matematických pravidiel

Analýza vplyvu

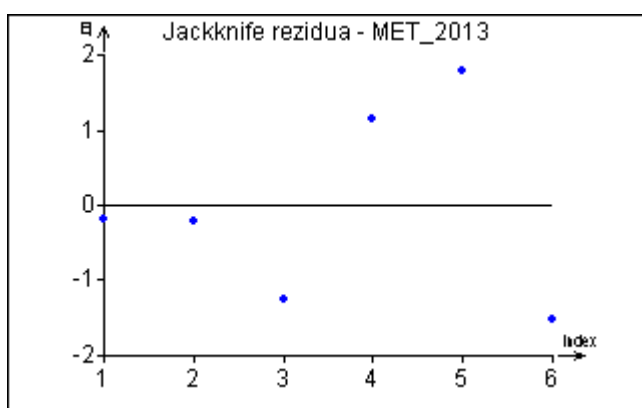
Index	Atkinson. vzdál.	Andrews-Pregibon st.	Vliv na Y [^]	Vliv na param. LD(b)	Vliv na rozptyl LD(s)	Celk. vliv LD(b,s)
1	0,03476	0,9846	-0,0155	0,00036	0,0863	0,0866
2	0,08464	0,9584	-0,0379	0,0021	0,0830	0,0848
3	0,7572	0,6718	-0,3386	0,1229	0,0650	0,2084
4	0,9524	0,6628	0,4259	0,2022	0,0296	0,2563
5	2,4895	0,3997	1,1133	0,9486	0,5825	2,1319
6	3,3345	0,3227	-1,4912	1,8156	0,2414	3,0211

Hodnoty sú zaokrúhlené na 4 desatinné miesta podľa matematických pravidiel

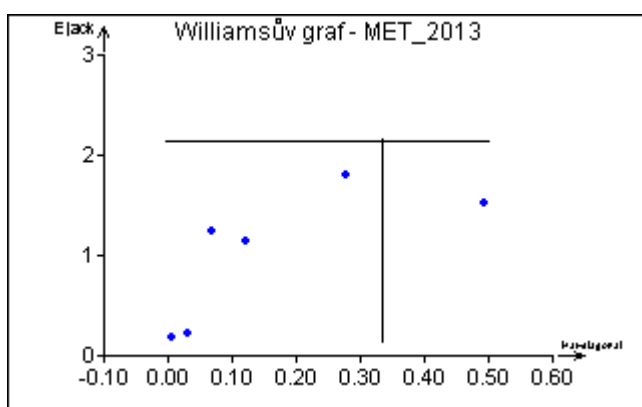
Testovanie vplyvných bodov pomocou diagnostických grafov – MET_2013



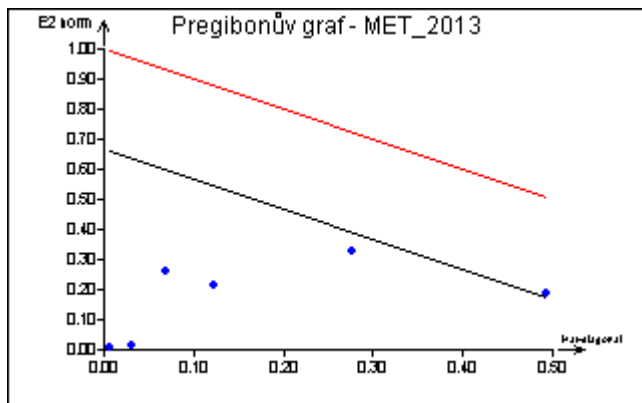
Graf 22: Graf predikovaných rezíduí poukazuje na bod 6 značne vzdialený od ostatných bodov



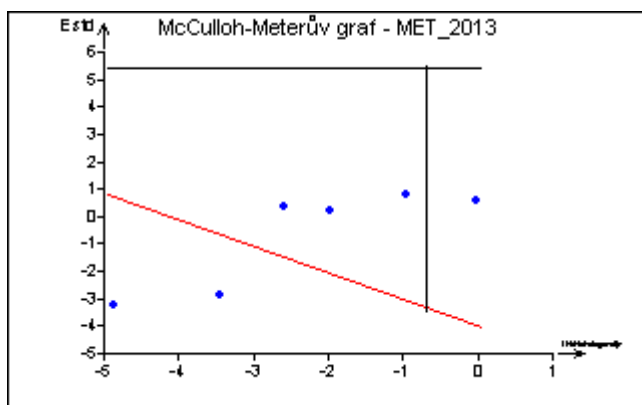
Graf 23: Graf Jackknife rezíduí neindikuje výrazne odľahlý bod – bod 6.



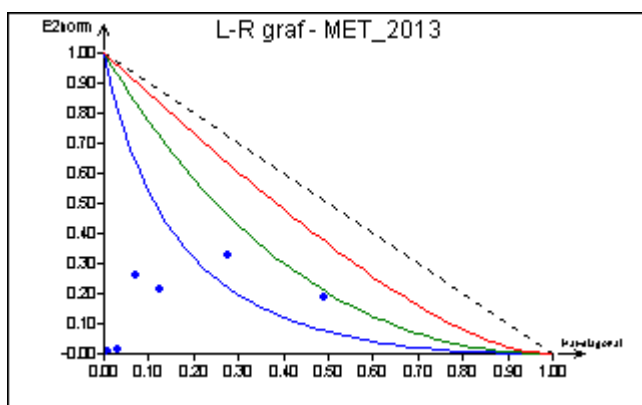
Graf 24: Dáta vo Wiliamsonovom grafe vykazujú mrak pod vodorovnou čiarou (nad ňou sa nachádzajú odľahlé body. V dátach sa nachádza extrém - bod 6 (za zvislou čiarou).



Graf 25: Pregibonov graf vykazuje stredne vplyvný bod 6 (oblasť medzi čiernou a červenou priamkou). Silno vplyvné body (nachádzajú sa nad červenou priamkou) - nie sú indikované.



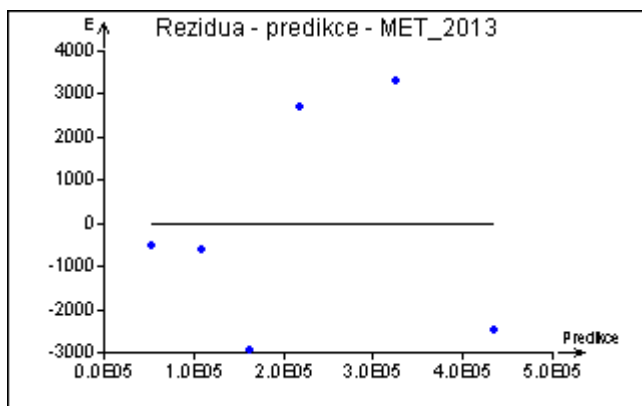
Graf 26: McCullohov- Meeterov graf indikuje body 3,4,5 ako podozrivé body, keďže ležia v oblasti ohraničenej všetkými tromi líniami. Žiaden bod však nebol indikovaný ako výlučne odľahlý (nachádzajúci sa nad vodorovnou líniou), ako extrém (nachádzajúci sa za zvislou líniou) sa ukazuje bod č.6.



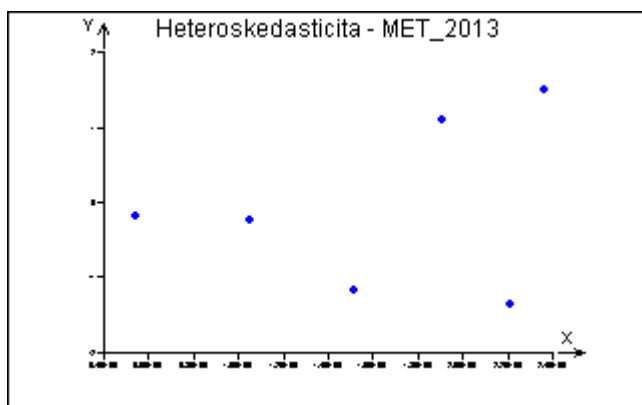
Graf 27: L-R graf poukazuje na body 5 a 6 ako na podozrivé.

Kritika modelu a metódy

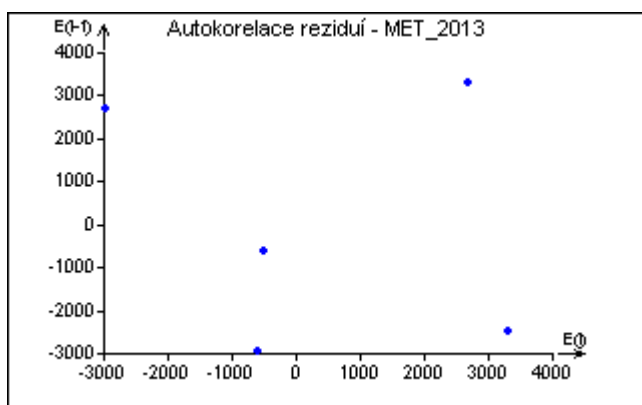
Diagnostické grafy



Graf 28: Rezidua – predikce. Body 3 a 6 sú vzdialené od ostatných 4 bodov, ktoré sú rozložené po oboch stranách priamky.



Graf 29: Heteroskedasticita (nekonštantnosť rozptylu). Pravdepodobne ide o homoskedasticitu.



Graf 30: Autokorelácia reziduí. Autokorelácia je nevýznamná nakoľko dáta vykazujú mrak a majú nenulovú smernicu.

Štatistické testy

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 14953,6607

Kvantil F (1-alfa, m-1, n-m) : 6,607890974

Pravdepodobnosť : 6,936237181E-010

Závěr : Model je významný

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0,432771309
Kvantil $\chi^2(1-\alpha,1)$: 3,841458829
Pravděpodobnost : 0,5106322329

Závěr : Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 0,5832360745
Kvantil $\chi^2(1-\alpha,2)$: 5,991464547
Pravděpodobnost : 0,7470538282

Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA : 0,1474989979
Kvantil $\chi^2(1-\alpha,1)$: 3,841458829
Pravděpodobnost : 0,7009369549

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 : 0,1767766953
Kvantil $N(1-\alpha/2)$: 1,959963999
Pravděpodobnost : 0,8596837952

Závěr : V reziduích není trend.

Štatistické charakteristiky regresie

Tab. 13: Zhrnutie niektorých štatistických charakteristík pre navrhované modely.

	Bez abs. členu	Bez abs. členu a bez bodu 6	Pôvodný model s abs. členom
RSC	33743665,08	21447485,09	33532347.61
S²(e)	6748733,016	5361871,273	8383086.901
R	0,9998328588	0,9997635188	0,9998339056
R²	0,9996657455	0,9995270935	0,9996678387
R_p	0,9987081902	0,9983754584	0,9980445114
MEP	10871077,8	7370695,376	164588925,6
AIC	95,25526272	78,35840019	97,21756997
Tvar modelu	$y = 2176,5342(9,1138)x$	$y = 2188,648(11,401)x$	$y = -366,9647(2311,3120) + 2179,2441(19,8620)x$

Reziduálny súčet štvorcov, RSC: optimálna hodnota je čo najnižšia
Reziduálny rozptyl, S²(e): optimálna hodnota je čo najnižšia
Vícenásobný korelačný koeficient R : optimálna hodnota je čo najbližšie 1
Koeficient determinace R² : optimálna hodnota je čo najbližšie 1
Predikovaný korelačný koeficient R_p : optimálna hodnota je čo najbližšie 1
Stredná kvadratická chyba predikcie MEP : optimálna hodnota je čo najbližšie 0

Akaikeho informační kritérium : optimálna hodnota je čo najzápornejšia (najnižšia)

Absolútna hodnota párového korelačného koeficientu R poukazuje, že navrhnutý lineárny regresný model je štatisticky významný. Vysoká hodnota koeficienta determinácie R^2 predstavuje % variability vysvetlené modelom. Predikčná schopnosť modelu je vysvetlená predikovaným korelačným koeficientom R_p . Stredná kvadratická chyba predikcie MEP a Akaikeho informačné kritérium, AIC sa používajú na rozlíšenie medzi viacerými navrhnutými modelmi. Za optimálny sa považuje ten model, pre ktorý MEP a AIC dosahuje minimálnu hodnotu.

Konštrukcia spresneného modelu

Na základe analýzy vplyvných bodov a grafických znázornení bol bod č. 6 vynechaný a vypočítali sa nové parametre modelu (tab. 14).

Tab. 14: Odhad parametrov pre model bez absolútneho člena a s vylúčením bodu 6.

Paramete r	Odhad	Odhad. smer. odchýlky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	2188,648	11,401	Významný	4,417266997 E-009	2156,994	2220,3028

Záver: Pri vyšetrení navrhnutého lineárneho regresného modelu pre stanovenie MET pre rok 2013 a prevedení regresnej diagnostiky boli prevedené nasledujúce kroky:

- použitie metódy najmenších štvorcov má svoje opodstatnenie;
- absolútny člen β_0 je štatisticky nevýznamný na hladine významnosti $\alpha=0,05$, o tento člen sa regresný model upravil;
- na základe viacerých analýz reziduí a grafov nebol vylúčený žiaden bod; ako z tab. 13 vyplýva, štatistické hodnoty dosahujú horšie hodnoty;
- model bez absolútneho člena je najpriateľnejší vzhľadom na všetky štatistické charakteristiky a má tvar

$$y=2176,5342(9,1138).x$$

1.3.4 Kritika dát, modelu a metódy pre D,L-metionín (MET) – rok 2015

Odhady parametrov

Tab. 15: Odhady parametrov pre navrhovaný model $y=\beta_0+ \beta_1*x$.

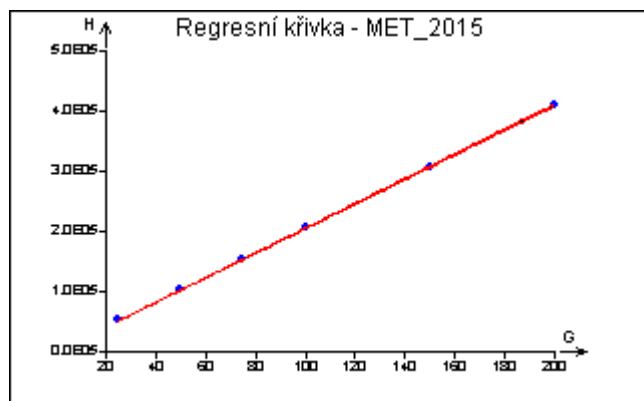
Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_0	-470,158	912,0743808	Nevýznamný	0,6334	-3002,483	2062,166
β_1	2054,959	7,837800364	Významný	1,2696 E-009	2033,198	2076,720

Úsek β_0 je štatisticky nevýznamný, testom vyhodnotený interval spoľahlivosti zahŕňa 0. Smernica β_1 je štatisticky významná, nie je jednotková.

V ďalšom testovaní reziduí a vplyvných bodov sa nepracuje s **absolútnym členom**, keďže je **štatisticky nevýznamný** (tab. 16).

Tab. 16: Odhady parametrov pre model bez absolútneho člena.

Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	2051,487228	3,702313993	Významný	3,632649737E-013	2041,970127	2061,004329



Graf 31: Regresná krivka pre MET nameraná v septembri 2015. Konfidenčný pás (vymedzený červenými krivkami) je veľmi úzky, data vykazujú linearitu.

Analýza klasických reziduí

Index	Y naměřené	Y vypočítané	Směr. odch. Y	Reziduum	Reziduum [%Y]
1	52143,48	51287,18069	92,55784982	856,2993077	1,642198234
2	101960,28	102574,3614	185,1156996	-614,0813846	-0,6022751062
3	152615,75	153861,5421	277,6735495	-1245,792077	-0,8162932574
4	205199,21	205148,7228	370,2313993	50,48723077	0,02460400835
5	306630,64	307723,0842	555,3470989	-1092,444154	-0,3562736437
6	411605,19	410297,4455	740,4627986	1307,744462	0,3177181662

Reziduální součet čtverců : 5568521,117
 Průměr absolutních reziduí : 861,1414359
 Reziduální směr. odchylka : 1055,321858
 Reziduální rozptyl : 1113704,223
 Šikmost reziduí : 0,01268872381
 Špičatost reziduí : 1,439146739

Analýza reziduí

Index	Standardní	Jackknife	Predikované	Diag(H _{ii})	Diag(H* _{ii})	Cook. vzdál.
1	0,81455	0,7823	862,9373	0,00769	0,1394	0,0063
2	-0,5911	-0,5482	-633,5760	0,03077	0,0985	-0,0188
3	-1,2236	-1,3076	-1338,4543	0,06923	0,3479	-0,0910
4	0,05109	0,0457	57,5736	0,1231	0,1235	0,00717
5	-1,2174	-1,2981	-1510,827	0,2769	0,4912	-0,4662
6	1,7392	2,4748	2575,8603	0,4923	0,7994	1,6865

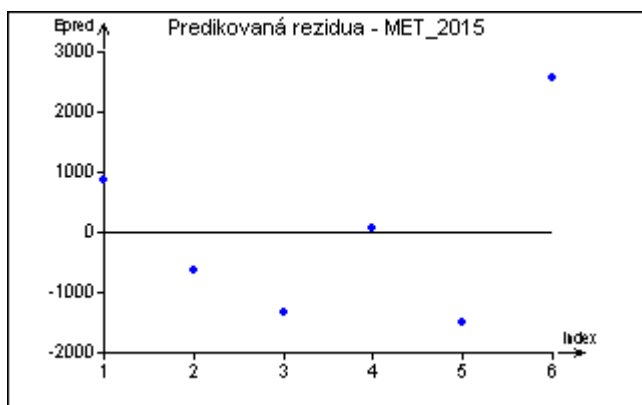
Hodnoty sú zaokrúhlené na 4 desatinné miesta podľa matematických pravidiel

Analýza vplyvu

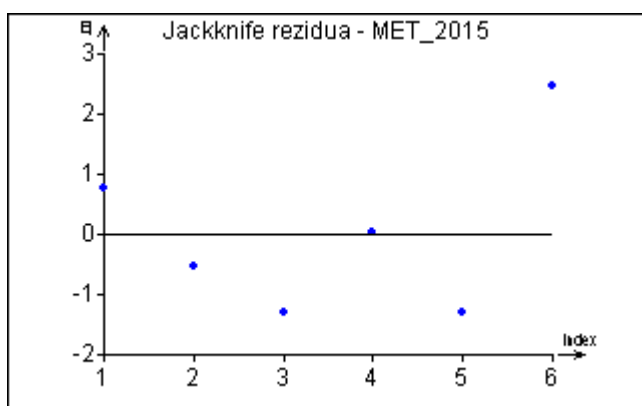
Index	Atkinson. vzdál.	Andrews- Pregibon st.	Vliv na \hat{Y}	Vliv na param. LD(b)	Vliv na rozptyl LD(s)	Celk. vliv LD(b,s)
1	0,1540	0,8606	0,0689	0,0062	0,0047	0,0107
2	0,2184	0,9015	-0,0977	0,0133	0,0349	0,0469
3	0,7974	0,6521	-0,3566	0,1322	0,0958	0,2548
4	0,0383	0,8765	0,01712	0,0004	0,0934	0,0938
5	1,7963	0,5088	-0,8033	0,6451	0,0910	0,8976
6	5,4494	0,2006	2,4370	2,7696	2,1778	9,6018

Hodnoty sú zaokrúhlené na 4 desatinné miesta podľa matematických pravidiel

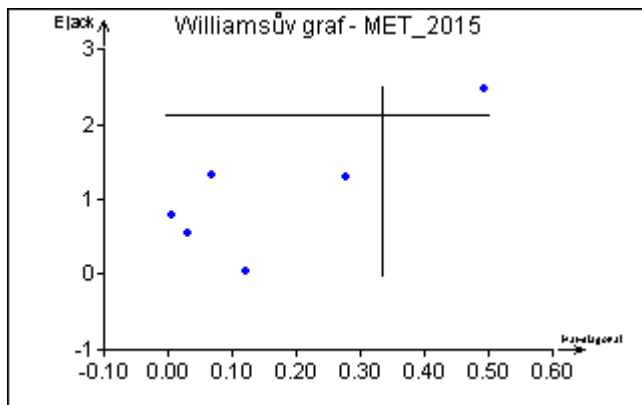
Testovanie vplyvných bodov pomocou diagnostických grafov – MET 2015



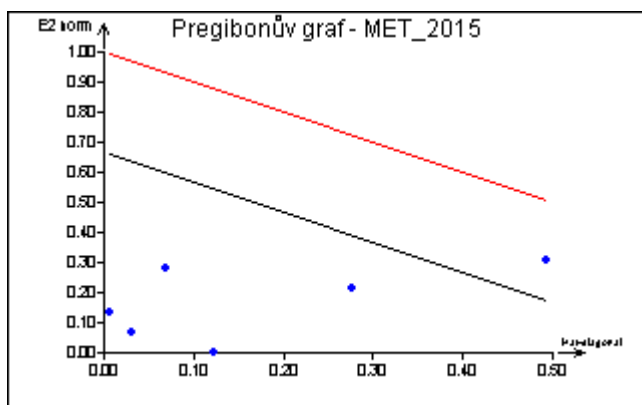
Graf 32: Graf predikovaných rezíduí poukazuje na bod 6 značne vzdialený od ostatných bodov



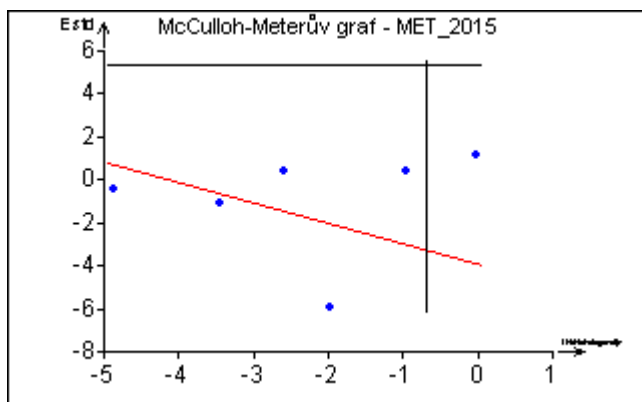
Graf 33: Graf Jackknife rezíduí neindikuje výrazne odľahlý bod – bod 6.



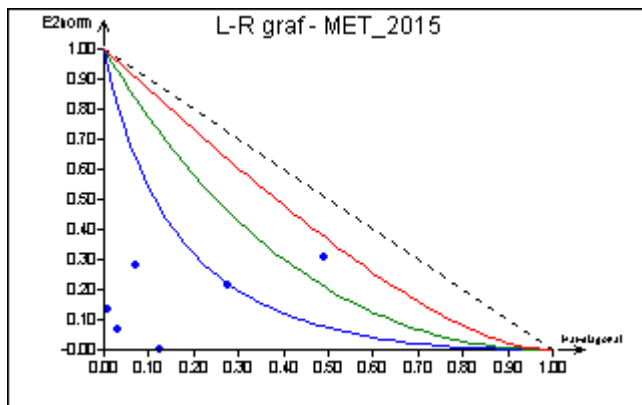
Graf 34: Dáta vo Wiliamsonovom grafe vykazujú mrak pod vodorovnou čiarou (nad ňou sa nachádzajú odľahlé body. V dátach sa nachádza extrém - bod 6 (za zvislou čiarou).



Graf 35: Pregibonov graf vykazuje stredne vplyvný bod 6 (oblasť medzi čiernou a červenou priamkou). Silno vplyvné body (nachádzajú sa nad červenou priamkou) nie sú indikované.



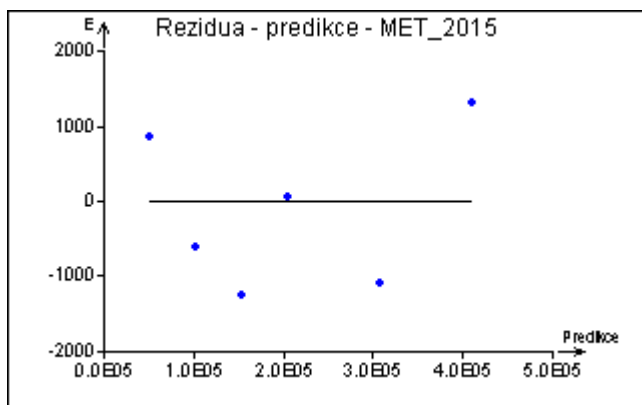
Graf 36: McCullohov- Meeterov graf indikuje body 3,5 ako podozrivé body keďže ležia v oblasti ohraničenej všetkými tromi líniami. Žiaden bod však nebol indikovaný ako výlučne odľahlý (nachádzajúci sa nad vodorovnou líniou), ako extrém (nachádzajúci sa za zvislou líniou) sa ukazuje bod č.6.



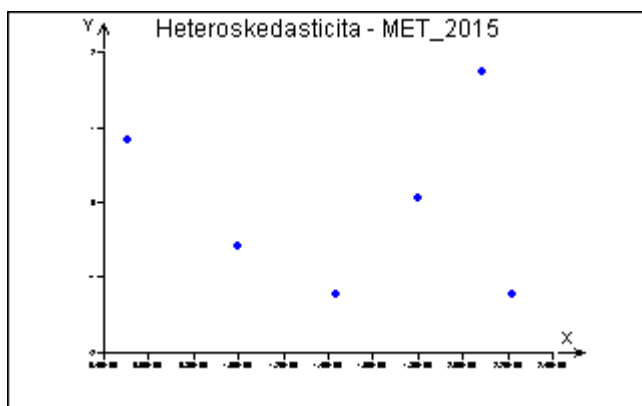
Graf 37: L-R graf poukazuje na body 5 a 6 ako na podozrivé.

Kritika modelu a metódy

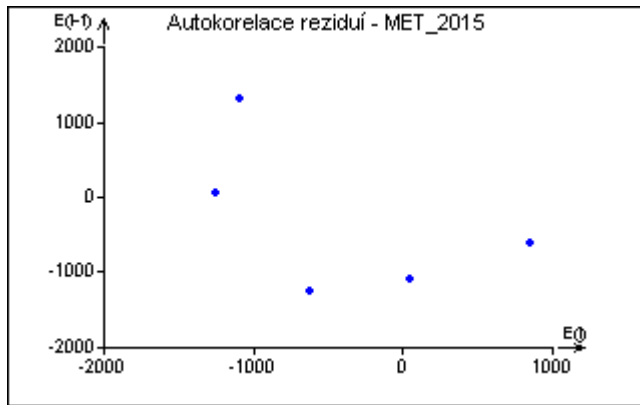
Diagnostické grafy



Graf 38: Rezidua – predikce. Body sú rozložené po oboch stranách priamky.



Graf 39: Heteroskedasticita (nekonštantnosť rozptylu). Pravdepodobne ide o homoskedasticitu. Body tvoria mrak.



Graf 40: Autokorelácia reziduí. Autokorelácia je nevýznamná nakoľko dáta vykazujú mrak a majú nenulovú smernicu.

Štatistické testy

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 80573,78797
 Kvantil F (1-alfa, m-1, n-m) : 6,607890974
 Pravděpodobnost : 1,029820801E-011

Závěr : Model je významný

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0,2658294259
 Kvantil $\chi^2(1-\alpha,1)$: 3,841458829
 Pravděpodobnost : 0,60614325

Závěr : Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 0,6217544495
 Kvantil $\chi^2(1-\alpha,2)$: 5,991464547
 Pravděpodobnost : 0,7328038405

Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA : 0,3500981429
 Kvantil $\chi^2(1-\alpha,1)$: 3,841458829
 Pravděpodobnost : 0,554057579

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 : 1,369306394
 Kvantil $N(1-\alpha/2)$: 1,959963999
 Pravděpodobnost : 0,1709035202

Závěr : V reziduích není trend.

Tab. 17: Zhrnutie niektorých štatistických charakteristík pre navrhované modely.

	Bez abs. člena	Bez abs. člena a bez bodu 6	Pôvodný model s abs. členom
RSC	5568521,117	2199954,072	5221644,736
S²(e)	1113704,223	549988,5179	1305411,184
R	0,999968974	0,9999714516	0,9999709067
R²	0,9999379489	0,9999429041	0,9999418142
R_p	0,9997357343	0,9998510132	0,9995960599
MEP	1976418,085	574080,5242	3021135,221
AIC	84,44528359	66,97254565	86,05938115
Tvar modelu	$y = 2051,487 (3,702)x$	$y = 2045,147 (3,651)x$	$y = -470,158(912,074) + 2054,959 (7,838)x$

Reziduálny súčet štvorcov, RSC: optimálna hodnota je čo najnižšia
 Reziduálny rozptyl, S²(e): optimálna hodnota je čo najnižšia
 Víčenásobný korelačný koeficient R : optimálna hodnota je čo najbližšie 1
 Koeficient determinácie R² : optimálna hodnota je čo najbližšie 1
 Predikovaný korelačný koeficient R_p : optimálna hodnota je čo najbližšie 1
 Stredná kvadratická chyba predikcie MEP : optimálna hodnota je čo najbližšie 0
 Akaikeho informačný kritérium : optimálna hodnota je čo najzápornejšia (najnižšia)

Absolútna hodnota párového korelačného koeficientu R poukazuje, že navrhnutý lineárny regresný model je štatisticky významný. Vysoká hodnota koeficienta determinácie R² predstavuje % variability vysvetlené modelom. Predikčná schopnosť modelu je vysvetlená predikovaným korelačným koeficientom R_p. Stredná kvadratická chyba predikcie MEP a Akaikeho informačný kritérium, AIC sa používajú na rozlíšenie medzi viacerými navrhnutými modelmi. Za optimálny sa považuje ten model, pre ktorý MEP a AIC dosahuje minimálnu hodnotu.

Konštrukcia spresneného modelu

Na základe analýzy vplyvných bodov a grafických znázornení bol bod č. 6 vynechaný a vypočítali sa nové parametre modelu (tab. 18).

Tab. 18: Odhady parametrov pre model bez absolútneho člena a bez bodu 6.

Parameter	Odhad	Odhad. smer. odchýlky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	2045,146 648	3,651445601	Významný	6,096811944 E-011	2035,00861	2055,284687

Záver: Pri vyšetrení navrhnutého lineárneho regresného modelu pre stanovenie MET pre rok 2015 a prevedení regresnej diagnostiky boli prevedené nasledujúce kroky:

- použitie metódy najmenších štvorcov má svoje opodstatnenie;
- absolútny člen β_0 je štatistický nevýznamný na hladine významnosti $\alpha=0,05$, o tento člen sa regresný model upravil;
- na základe viacerých analýz reziduí a grafov bol vylúčený bod 6; v prevažnej väčšine testov bol indikovaný ako extrém;
- z vyššie uvedenej tabuľky vyplýva, že všetky štatistické charakteristiky po vylúčení bodu 6 dosahujú výrazne lepšie hodnoty (tab. 17);
- výsledný model bez absolútneho člena a bez bodu 6 je najpriateľnejší vzhľadom na všetky štatistické charakteristiky a má tvar

$$y = 2045,147(3,651).x$$

1.3.5 Porovnanie regresných kriviek

Úlohou je porovnanie regresných kriviek LHC_2013 a LHC_2015 a podobne MET_2013 a MET_2015. Najskôr sa pristúpi k testovaniu rozptylov.

1.3.5.1 Test zhodnosti rozptylov

Keďže ide o priamkové závislosti, pre testovanie sa využije Fischer-Snedecorov test.

Testujeme hypotézu $H_0: \sigma^2_A(e) = \sigma^2_B(e)$ vs. alternatívna hypotéza $H_A: \sigma^2_A(e) \neq \sigma^2_B(e)$
Do výpočtu sa vzali upresnené modely bez absolútneho člena, prípadne s vylúčenými bodmi ako je uvedené v nasledujúcich prehľadných tab. 19-22.

Tab. 19: LHC_2013 (bez absolútneho člena)

Parameter	Odhad	Odhad. smer. odchýlky	Dolná medza	Horná medza	Rezid. rozptyl
β_1	771,7988	2,4037	765,6200	777,9777	469431,6173

Tab. 20: LHC_2015 (bez absolútneho člena a bez bodu 5)

Parameter	Odhad	Odhad. smer. odchýlky	Dolná medza	Horná medza	Rezid. rozptyl
β_1	732,3715	4,4754	719,9458	744,7972	1176710,73

Regresná priamka s vyššou hodnotou reziduálneho rozptylu, **A.....LHC_2015**

Regresná priamka s nižšou hodnotou reziduálneho rozptylu, **B.....LHC_2013**

Testačné kritérium $F = \sigma^2_A(e) / \sigma^2_B(e)$ sa porovná s kvantilom F-rozdelenia $F_{0,95}(v_A, v_B)$, pričom $v_i = n_i - m_i$ počet stupňov voľnosti (n_i - počet riadkov dát, m_i - počet parametrov, pre priamkovú závislosť maximálne 2).

$$F_{A,B} = 1176710,73 / 469431,6173 = 2,506671$$

$$F_{krit}(0,05; v_A, v_B) = F_{krit}(0,05; 5-2, 6-2) = F_{krit}(0,05; 3, 4) = 6,591$$

Hypotéza H_0 sa **prijíma**, rozptyly sú zhodné, pretože $F_{A,B} < F_{krit}(0,05; 3, 4)$

Tab. 21: MET_2013 (bez absolútneho člena)

Parameter	Odhad	Odhad. smer. odchýlky	Dolná medza	Horná medza	Rezid. rozptyl
β_1	2176,5342	9,1138	2153,1064	2199,9619	6748733,016

Tab. 22: MET_2015 (bez absolútneho člena a bez bodu 6)

Parameter	Odhad	Odhad. smer. odchýlky	Dolná medza	Horná medza	Rezid. Rozptyl
β_1	2045,147	3,651	2035,007	2055,285	549988,5179

Regresná priamka s vyššou hodnotou reziduálneho rozptylu, **A.....MET_2013**

Regresná priamka s nižšou hodnotou reziduálneho rozptylu, **B.....MET_2015**

Testačné kritérium $F = \sigma_A^2(e) / \sigma_B^2(e)$ sa porovnáva s kvantilom F-rozdelenia $F_{0,95}(v_A, v_B)$, pričom $v_i = n_i - m_i$ počet stupňov voľnosti (n_i - počet riadkov dát, m_i - počet parametrov, pre priamkovú závislosť maximálne 2).

$$F_{A,B} = 6748733,016 / 549988,5179 = 12,2707$$

$$F_{\text{krit}}(0,05; v_A, v_B) = F_{\text{krit}}(0,05; 6-2, 5-2) = F_{\text{krit}}(0,05; 4, 3) = 9,117$$

Hypotéza **H₀** sa **zamieta**, rozptyly nie sú zhodné, pretože $F_{A,B} > F_{\text{krit}}(0,05; 4, 3)$.

1.3.5.2 Chowov test zhody dvoch lineárnych modelov

Tab. 23: Zhrnutie dát pre výpočet zlúčeného modelu pre obidva analyty po spresnení modelu a vylúčení podozrivých bodov (modro označené)

LHC_2013		LHC_2015	
c(LHC), ug/ml	A(LHC), (μV.s)	c(LHC), ug/ml	A(LHC), (μV.s)
25,00	19680,91	25,00	19744,26
50,00	38513,51	50,00	38138,19
75,00	56606,80	75,00	54821,30
100,00	76803,96	100,00	72730,95
150,00	115759,29	150,00	112490,10
200,00	155005,80	200,00	146208,08

MET_2013		MET_2015	
c(MET), ug/ml	A(MET), (μV.s)	c(MET), ug/ml	A(MET), (μV.s)
25,00	53904,50	25,00	52143,48
50,00	108222,64	50,00	101960,28
75,00	160284,04	75,00	152615,75
100,00	220341,64	100,00	205199,21
150,00	329783,53	150,00	306630,64
200,00	432808,30	200,00	411605,19

Testujeme hypotézu **H₀: β_A = β_B** vs. alternatívna hypotéza **H_A: β_A ≠ β_B**

Testačné kritérium: $F_C = (RSC - RSC_A - RSC_B) * (n - 2m) / (RSC_A + RSC_B) * m$

kde

RSC je reziduálny súčet štvorcov pre zlúčený model

RSC_A je reziduálny súčet štvorcov pre model A

RSC_B je reziduálny súčet štvorcov pre model B

n = n_A + n_B počet riadkov zlúčeného modelu

m počet parametrov (pre priamkovú závislosť 2)

Testačné kritérium sa porovnáva s kvantilom F-rozdelenia a môžu nastať dva prípady:

1. Reziduálne rozptyly sa rovnajú $\sigma_A^2(e) = \sigma_B^2(e) \rightarrow F_{0,05}(m, n-2m)$

2. Reziduálne rozptyly sa nerovnajú $\sigma_A^2(e) \neq \sigma_B^2(e) \rightarrow F_{0,05}(m, r)$

$$r = \frac{[(n_A - m)\sigma_A^2(e) + (n_B - m)\sigma_B^2(e)]}{(n_A - m)\sigma_A^4(e) + (n_B - m)\sigma_B^4(e)}$$

Tab. 24: Zhrnutie štatistických charakteristík a výpočtov pre Chowov test (dáta uložené v darina_4_sem_praca.xls)

	LHC_2013 a LHC_2015	MET_2013 a MET_2015
RSC	60056769,62	508245309,3
RSC_A	4706842,919	33743665,08
RSC_B	2347158,087	2199954,072
σ^2_A	1176710,73	6748733,016
σ^2_B	469431,6173	549988,5179
N	11	11
M	2	2
F_C	26,299	45,990
R	-	4,482 (4)
F_{0,05(2, 7)}	4,737	-
F_{0,05(2, 4)}	-	6,944

Záver:

Regresné priamky pre LHC namerané v septembri 2013 a januári 2015 nie sú zhodné, pretože $F_C > F_{0,05}(2, 7)$ a rovnako nie sú zhodné na základe Chowovho testu ani priamky pre MET namerané v tom istom časovom období, keďže platí $F_C > F_{0,05}(2, 4)$. Tieto výsledky z praktického hľadiska indikujú, že kolóna neposkytuje dostatočne prijateľné výsledky v skúmanom časovom horizonte a je potrebné ju vymeniť za novú v kratšom časovom rozsahu (napr. po 12 mesiacoch používania, kedy je pravidelná výrobná odstávka a metódy sa validujú).

2. úloha: Určenie stupňa polynómu

2.1 Zadanie

Vyhľadajte stupeň polynómu závislosti signálu plynovej chromatografie, GC - (plochy píku) od koncentrácie acetónu (ACT), ktorý je jedným zo sledovaných zostatkových rozpúšťadiel pri kontrole kvality 18F-rádiofarmák. Pri chromatografickej metóde sa použil roztok metanolu (MeOH) ako vnútorný štandard. Na získanie kalibračnej závislosti sa používajú pomery $c(\text{ACT})/c(\text{MeOH})$ ku pomeru plôch píkov $A(\text{ACT})/A(\text{MeOH})$. Zhodnoťte použitie metódy najmenších štvorcov (MNS), resp. metódy racionálnych hodností (RH).

2.2 Dáta (polynom.txt, polynom1.txt, polynom, darina_4_sem_praca.xls)

Tab. 25: Dáta k úlohe č.2.

c(ACT)/c(MeOH)	A(ACT)/A(MeOH)
0,0053	0,003
0,0067	0,007
0,0134	0,013
0,0200	0,021
0,0401	0,039
0,0668	0,063
0,1336	0,115

2.3 Návrh modelu

K určení optimálneho stupňa polynómu metódou MNŠ sa sledujú štatistické charakteristiky pre rôzne stupne polynómu ako je uvedené v tab. 26. Za najvhodnejší sa pokladá taký polynóm, ktorý má minimálnu hodnotu MEP, príp. AIC, zatiaľčo koeficient determinácie, R^2 má maximálnu hodnotu.

Tab. 26: Súhrnná tabuľka štatistických charakteristík pre rôzne stupne polynómu (výstup z QC Expert).

SP	ABS	RSC	S ² (e)	R	R ²	R _p	MEP	AIC	MUL	PAR
1	nie	6,0081 E-005	1,0014 E-005	0,9969	0,9938	0,9539	3,2160 E-005	-79,66	nie	1/0
	áno	2,1943 E-005	4,3886 E-006	0,9989	0,9977	0,9760	1,6617 E-005	-84,71	áno	2/0
2	nie	1,0265 E-005	2,0531 E-006	0,9995	0,9989	0,9970	2,0665 E-006	-90,03	áno	2/0
	áno	5,6195 E-006	1,4049 E-006	0,9997	0,9994	0,9753	1,7157 E-005	-92,25	áno	3/1
3	nie	1,0187 E-005	2,5467 E-006	0,9995	0,9989	0,9804	1,3547 E-005	-88,08	áno	3/2
	áno	3,9715 E-006	1,3238 E-006	0,9998	0,9996	0,4718	0,00043	-92,68	áno	4/3
4	nie	1,0169 E-005	3,3897 E-006	0,9995	0,9989	369,1586	0,0279	-86,09	áno	4/3
	áno	2,7543 E-006	1,3772 E-006	0,99986	0,9997	19386,62	0,1933	-93,24	áno	5/5

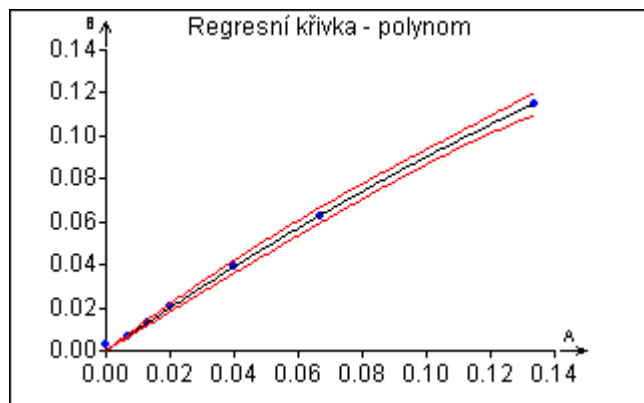
Tab. 27: Súhrnná tabuľka štatistických charakteristík pre rôzne stupne polynómu (výstup z ADSTAT).

SP	ABS	RSC	S ² (e)	R	R ²	R _p	MEP	AIC	MUL	PAR
1	nie	5,3977 E-005	7,6824 E-006	0,9972	0,9944	0,9889	3,0546 E-005	-81,52	nie	1/0
	áno	3,8956 E-005	7,7913 E-006	0,99798	0,99596	0,9891	2,9833 E-005	-80,69	nie	2/1
2	nie	7,0443 E-006	1,3242 E-006	0,9996	0,9993	0,9995	1,3170 E-006	-93,10	áno	2/0
	áno	5,4793 E-006	1,3698 E-006	0,9997	0,9994	0,9942	1,5961 E-005	-92,42	áno	3/1
3	nie	6,9796 E-006	1,6698 E-006	0,9996	0,9993	0,9657	9,2884 E-005	-91,04	áno	3/2
	áno	4,2399 E-006	1,4133 E-006	0,99978	0,99956	0,5819	9,1194 E-04	-92,22	áno	4/3
4	nie	6,5320 E-006	2,1314 E-006	0,99966	0,9993	0,0000	1,2904 E-02	-89,34	áno	4/3
	áno	3,0633 E-006	1,5317 E-006	0,99984	0,99968	0,9961	1,0803 E-05	-92,49	áno	5/5

SP – stupeň polynómu, ABS – absolútny člen, MUL – multikolinearita, PAR – počet parametrov/ parametre vyhodnotené ako nevýznamné

Pri porovnaní výstupov z QC Expert a z ADSTAT sú patrné rozdiely vo výstupných hodnotách sledovaných charakteristík. V tab. 26 dokonca možno pozorovať až nezmyselné hodnoty R_p , či pomerne veľké rozdiely v MEP hodnotách pri polynóme $m=4$ s abs. členom. Rozhodujúce kritérium MEP je najbližšie k 0, hodnota AIC je najzápornejšia a hodnota R_p je najbližšie k 1, ak stupeň polynómu = 2 bez absolútneho člena. Vybraný model sa ďalej testoval programom ADSTAT a QC Expert.

Testovanie vybraného modelu – polynóm 2. stupňa bez absolútneho člena (ADSTAT Data Test 1.20)



Graf 41: Regresná krivka (polynóm 2. stupňa) - stanovenie acetónu.

(1) PŘEDBĚŽNÁ STATISTICKÁ ANALÝZA:

Proměnná	Průměr	Směrodatná odchylka	Párový korelační koeficient	Spočtená hladina výz.
y	3.7286E-02	4.0107E-02	1.0000	----
x1	4.0843E-02	4.6359E-02	0.9980	0.000
x2	3.5102E-03	6.5195E-03	0.9506	0.001

(2) INDIKACE MULTIKOLINEARITY:

[j]	Vlastní čísla korel. matice I[j]	Čísla podmí- něnosti K[j]	Variance inflation factor VIF[j]	Vícenás.korel. koef pro X[j]
1	4.7581E-02	4.1033E+01	5.6489E+00	0.9072
2	1.9524E+00	1.0000E+00	8.0439E+00	0.9358

Maximální číslo podmíněnosti K : 4.1033E+01

(K[j], K > 1000 indikuje silnou multikolaritu)

(VIF[j] > 10 indikuje silnou multikolaritu)

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B[j] = 0 vs. HA: B[j] <> 0	t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	0.0000E+00	-----	-----	-----	-----	----
B[1]	1.0235E+00	2.4843E-02	4.1198E+01	Zamítnuta		0.000
B[2]	-1.2167E+00	2.1080E-01	-5.7717E+00	Zamítnuta		0.002

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.9963E-01
Koeficient determinace, R ²	: 9.9927E-01
Predikovaný korelační koeficient, R _p ²	: 9.9952E-01
Střední kvadratická chyba predikce, MEP	: 1.3170E-06
Akaikeho informační kritérium, AIC	:-9.3098E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	yexp[i]	yvyp[i]	s(yvyp[i])	e[i]	er[i]
1	3.0000E-03	5.3903E-03	1.2604E-04	-2.3903E-03	-7.9677E+01
2	7.0000E-03	6.8028E-03	1.5746E-04	1.9723E-04	2.8176E+00
3	1.3000E-02	1.3496E-02	2.9707E-04	-4.9631E-04	-3.8178E+00
4	2.1000E-02	1.9983E-02	4.1735E-04	1.0169E-03	4.8422E+00
5	3.9000E-02	3.9086E-02	6.8125E-04	-8.5542E-05	-2.1934E-01
6	6.3000E-02	6.2940E-02	8.1569E-04	5.9954E-05	9.5165E-02
7	1.1500E-01	1.1502E-01	1.1769E-03	-2.1811E-05	-1.8966E-02

Rezidualní součet čtverců, RSC	: 7.0443E-06
Průměr absolutních hodnot reziduí, Me	: 6.0972E-04
Průměr relativních reziduí, Mer	: 1.3070E+01
Odhad reziduálního rozptylu, $s^2(e)$: 1.3243E-06
Odhad směrodatné, odchylky reziduí, $s(e)$: 1.1508E-03
Odhad šikmosti reziduí, $g1(e)$:-1.2012E+00
Odhad špičatosti reziduí, $g2(e)$: 3.7916E+00

(6) TESTOVÁNÍ REGRESNÍHO TRIPLETU (DATA + MODEL + METODA):

Fisher-Snedocorův test významnosti regrese, F	: 6.8456E+03
Tabulkový kvantil, $F(1-\alpha, m-1, n-m)$: 6.6079E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.000

Scottovo kritérium multikolinearity, M	: 7.7557E-01
Závěr: Navržený model není korektní.	

Cook-Weisbergův test heteroskedasticity, Sf	: 6.5268E+00
Tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti	: 0.011

Jarque-Berraův test normality reziduí, L(e)	: 1.8662E+00
Tabulkový kvantil, $\chi^2(1-\alpha, 2)$: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.393

Waldův test autokorelace, Wa	: 1.9771E-01
Tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.657

Znamkový test, Dt	: 1.7584E+00
Tabulkový kvantil, $N(1-\alpha/2)$: 1.6449E+00
Závěr: Rezidua vykazují trend.	
Spočtená hladina významnosti	: 0.039

Testovanie vybraného modelu – polynóm 2. stupňa bez absolútneho člena (QC Expert)

Tab. 28: Odhady parametrov pre polynóm 2. stupňa bez abs. člena.

Parameter	Odhad	Odhad. smer. odchylky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	1,023490839	0,024843319	Významný	1,589258365 E-007	0,95962905	1,0873526
β_2	-1,21668516	0,21080151	Významný	0,00219456	-1,7585677	- 0,6748026

Tab. 29: Indikácia multikolinearity

Parameter	Č. podmienenosti K	VIF
β_1	1	15,63162422
β_2	60,50997068	15,63162422

Kappa, K → ak $K > 1000$ ide o silnú multikolinearitu

VIF (Variance Inflation Factor) → ak $VIF > 10$, ide o silnú multikolinearitu

Testovanie regresného tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 6845,561998

Kvantil F (1-alfa, m-1, n-m) : 6,607890974

Pravdepodobnosť : 4,887666879E-009

Záver : Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC : 0,9903145098

Záver : Model je nekorektní!

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 1,134643321

Kvantil $\chi^2(1-\text{alfa}, 1)$: 3,841458829

Pravdepodobnosť : 0,286787073

Záver : Rezidua vykazujú homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 4,689287541

Kvantil $\chi^2(1-\text{alfa}, 2)$: 5,991464547

Pravdepodobnosť : 0,0958813518

Záver : Rezidua majú normální rozdění.

Waldův test autokorelace

Hodnota kritéria WA : 0,1361077433

Kvantil $\chi^2(1-\text{alfa}, 1)$: 3,841458829

Pravdepodobnosť : 0,7121815195

Záver : Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : -1

Kritické hodnoty DW 0

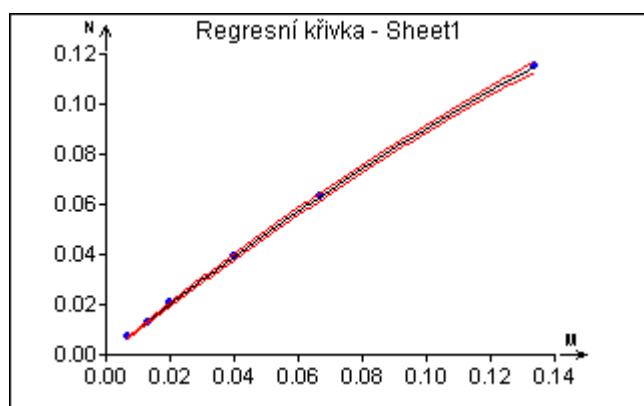
Záver : Pozitívni autokorelace reziduů není prokázána.

Znaménkový test reziduí
Hodnota kritéria Sg : 2,607257969
Kvantil $N(1-\alpha/2)$: 1,959963999
Pravděpodobnost : 0,00912705693
Závěr : V reziduách je trend!

Model je nekorektný (hodnota Scottovho kritéria je vysoká), navyše v reziduách je trend. Program Adstat indikuje heteroskedasticitu a nižšie hodnoty K a VIF ako softvér QC Expert. Vo výbere je indikovaný podozrivý bod č. 1. Vylúčenie bodu 1 zlepšuje viaceré štatistické charakteristiky (nižšie uvedené), navyše sa odstráni trend v reziduách, a heteroskedasticita. Multikolinearita však ostáva (potvrdené obidvoma softvérm).

Tab. 30: Odhady parametrov pre polynóm 2. stupňa bez abs. člena po vylúčení bodu 1.

Parameter	Odhad	Odhad. smer. odchýlky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
β_1	1,02886363	0,0118394187	Významný	1,05113434 E-007	0,995992134	1,061735126
β_2	-1,2599064	0,10040298	Významný	0,000232	-1,5386697	-0,9811430



Graf 42: Regresná krivka zvolený model – polynóm 2. stupňa po vylúčení bodu 1.

Tab. 31: Indikácia multikolinearity

Parameter	Č. podmienosti K	VIF
β_1	1	17,32866762
β_2	67,2998116	17,32866762

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R : 0,9999235814
Koeficient determinace R^2 : 0,9998471686
Predikovaný korelační koeficient R_p : 0,9986802414
Střední kvadratická chyba predikce MEP : 9,10934111E-007
Akaikeho informační kritérium : -88,23108357

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu
Hodnota kritéria F : 26168,6393

Kvantil F (1-alfa, m-1, n-m) : 7,708647422

Pravdepodobnost : 8,759480224E-009

Záver : Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC : 0,9880373672

Záver : Model je nekorektní!

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0,5347920095

Kvantil $\chi^2(1-\text{alfa},1)$: 3,841458829

Pravdepodobnost : 0,4645991373

Záver : Rezidua vykazujú homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 1,108768651

Kvantil $\chi^2(1-\text{alfa},2)$: 5,991464547

Pravdepodobnost : 0,5744258117

Záver : Rezidua majú normální rozdění.

Waldův test autokorelace

Hodnota kritéria WA : 2,091834472

Kvantil $\chi^2(1-\text{alfa},1)$: 3,841458829

Pravdepodobnost : 0,1480881539

Záver : Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : -1

Kritické hodnoty DW 0

Záver : Negativní autokorelace reziduí není prokázána.

Znaménkový test reziduí

Hodnota kritéria Sg : 1,369306394

Kvantil $N(1-\text{alfa}/2)$: 1,959963999

Pravdepodobnost : 0,1709035202

Záver : V reziduích není trend.

Záver

Optimální stupeň polynómu je 2 bez absolutného člena, keďže MEP má minimálnu hodnotu, AIC je záporné a R_p je najbližšie k 1. Po vylúčení bodu 1 sa v reziduích odstraňuje trend a heteroskedasticita. Model vykazuje multikolinearitu a je nekorektný – hodnota Scottovho kritéria je vysoká ($> 0,8$).

Vzhľadom k tomu, že odhady parametrov β sú nenulové – nulové hypotézy sú zamietnuté - nie je nutné použiť metódu racionálnych hodností (RH), ktorou sa optimalizuje filter obmedzenia. Výsledný model pre výpočet koncentrácie acetónu má tvar $y=0,8869(\pm 0,0191)*x$.

3. úloha: Viacrozmerný lineárny regresný model

3.1 Zadanie

U pacientov s diabetes mellitus II. typu sa sledovala hladina glykovaného hemoglobínu (HbA1c), jeho hladinu môžu ovplyvniť faktory ako je vek pacienta, hladina cholesterolu (CHOL), plazmatická hladina glukózy (GLU), hladiny lipoproteínov – nízkoenzitný (LDL) a vysokoenzitný (HDL) ako aj hladina triglyceridov (TGL). Navrhните viacrozmerý lineárny regresný model a rozhodnite či uvedené

faktory ovplyvňujú hladinu HbA1c, ktorá sa považuje za najdôležitejší parameter kompenzácie diabetu. Vyšetrite štatistickú významnosť jednotlivých parametrov, vyšetrite regresný triplet, popíšte najdôležitejšie regresné a parciálne reziduálne grafy, sú v dátach vplyvné body? Ako sa pristúpi k úlohe, ak sú porušené predpoklady MNS, ak sa prejavuje heteroskedasticita, nenormalita dát či autokorelácia.

3.2 Dáta (darina_4_sem_praca.xls, darina_4_sem_praca.vts)

Tab. 32: Dáta pre riešenie viacrozmerného lineárneho regresného modelu.

	y	x1	x2	x3	x4	x5	x6
Poradie	HbA1c (%)	Vek, roky	c(CHOL), mmol/l	c(GLU), mmol/l	c(HDL), mmol/l	c(LDL), mmol/l	c(TGL), mmol/l
1	4,200	62	4,2000	7,1000	0,6500	3,1500	1,7300
2	6,900	51	3,8700	9,6600	0,7900	2,1600	2,4200
3	4,300	63	4,0100	9,3400	0,9300	2,6200	1,5400
4	4,100	67	2,8200	6,9500	0,8900	1,3900	1,4700
5	8,400	65	5,3600	9,8500	1,0500	3,8600	1,3900
6	4,200	76	4,5600	7,7600	1,2500	3,1000	1,1700
7	8,600	63	4,0000	8,8800	0,6700	2,4800	2,3000
8	8,100	61	3,2800	10,0600	0,9400	1,7100	2,2600
9	4,500	64	6,5400	9,4400	1,2400	5,6000	1,8800
10	4,300	62	4,4900	6,7800	1,0900	3,1900	0,6200
11	4,400	57	5,4800	8,0200	0,9600	2,8200	5,1600
12	4,500	77	3,1300	5,7900	1,1100	1,7000	1,1400
13	5,500	71	3,6000	7,0200	0,7900	2,4000	1,1400
14	8,400	73	4,7000	10,5800	1,3300	2,7800	1,7200
15	6,400	77	4,2300	9,1100	1,2400	2,6400	1,0800
16	4,800	76	4,3700	7,0000	1,6800	2,3800	0,9200
17	8,200	64	5,2800	13,1500	1,2600	3,4700	2,480
18	4,900	75	3,8500	7,7100	0,8100	2,3400	1,9000
19	4,500	66	5,7000	7,1000	1,4400	3,4400	1,6300
20	8,800	58	3,0500	9,4000	1,2100	1,5400	0,5900
21	6,700	68	3,8700	11,4700	1,3700	2,3700	1,0700
22	6,500	61	3,2600	5,7400	0,9700	1,4600	2,2800
23	5,400	54	5,4400	7,3200	1,2600	3,3200	1,7500
24	4,800	60	4,3700	6,8100	1,1100	2,6100	1,2700
25	5,600	49	2,9900	8,1700	1,1800	1,5400	1,3100

3.3 Riešenie

Návrh modelu:

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 + \beta_4 * x_4 + \beta_5 * x_5 + \beta_6 * x_6$$

Hladina významnosti : 0,05
 Kvantil t(1-alfa/2,n-m) : 2,10092204024043
 Kvantil F(1-alfa,m,n-m) : 2,6613045229279
 Absolutní člen : Ano
 Počet platných riadkov : 25
 Počet parametrov : 7
 Metoda : Najmenší čtverce
 Sloupce pro výpočet : G
 Abs
 A - vek
 B - CHOL

C - GLU
D - HDL
E - LDL
F - TGL
Transformace : Bez transformace

Základná analýza

Charakteristiky premenných

Promenná	Průměr	Směr.Odch.	Kor.vs.Y	Významnost
A	64,8	7,942501705	-0,1497220797	0,4750253999
B	4,258	0,9616435584	-0,1170612819	0,5773479213
C	8,4084	1,797219334	0,6540423005	0,0003603294343
D	1,0888	0,250637853	-0,01222024812	0,9537685048
E	2,6428	0,9250248645	-0,1491240707	0,476811165
F	1,6888	0,8989406729	0,03016442157	0,8861853432

Analýza rozptylu

Průměr Y : 5,88

Zdroj	Součet čtverců	Průměrný čtverec	Rozptyl
Celková variabilita	66,76	2,6704	2,781666667
Variabilita vysvětlená modelem	39,74756797	1,589902719	1,656148665
Reziduální variabilita	27,01243203	1,080497281	1,125518001
Hodnota kritéria F :	4,414363867		
Kvantil F (1-alfa, m-1, n-m) :	2,661304523		
Pravděpodobnost :	0,006485207108		

Závěr : **Model je významný**

Odhady parametrov

Tab. 33: Odhady parametrov z programu QC Expert ver. 2.5.

Parameter	Odhad	Odhad. smer. odchýlky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
b0	1,9918956	2,912734	Nevýznamný	0,50277	-4,1275	8,111323508
Vek (roky)	-0,0096775	0,0344179	Nevýznamný	0,78178	-0,08199	0,062631767
CHOL, (mmol/l)	2,0073261	1,3207042	Nevýznamný	0,14591	-0,7674	4,78202271
GLU, (mmol/l)	0,76799129	0,1553406	Významný	0,000105	0,44163	1,094349687
HDL, (mmol/l)	-2,6649088	1,814792	Nevýznamný	0,159244	-6,4776	1,147827058

LDL, (mmol/l)	-2,319505	1,200823	Nevýznamný	0,069309	-4,8423	0,203329772
TGL, (mmol/l)	-0,8633493	0,548329	Nevýznamný	0,13278	-2,0153	0,288647733

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R : 0,7716087011
 Koeficient determinace R² : 0,5953799876
 Predikovaný korelační koeficient R_p : 0,02520926166
 Střední kvadratická chyba predikce MEP : 2,24640925
 Akaikeho informační kritérium : 15,93553452

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu
 Hodnota kritéria F : 4,414363867
 Kvantil F (1-alfa, m-1, n-m) : 2,661304523
 Pravděpodobnost : 0,006485207108
Závěr : Model je významný

Scottovo kritérium multikolinearity
 Hodnota kritéria SC : -0,07158358281
Závěr : Model je korektní.

Cook-Weisbergův test heteroskedasticity
 Hodnota kritéria CW : 0,4190782705
 Kvantil Chi²(1-alfa,1) : 3,841458829
 Pravděpodobnost : 0,517397328
Závěr : Rezidua vykazují homoskedasticitu.

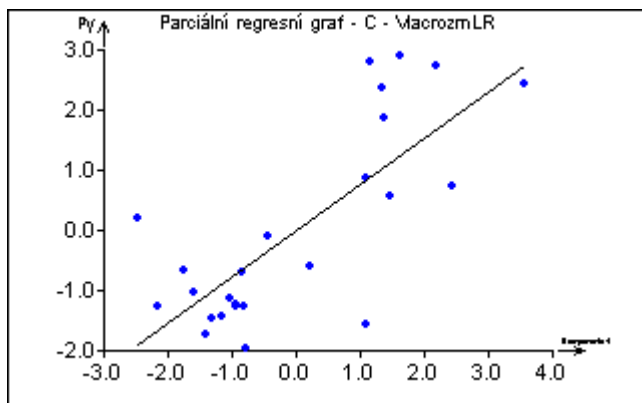
Jarque-Berrův test normality
 Hodnota kritéria JB : 0,1280153602
 Kvantil Chi²(1-alfa,2) : 5,991464547
 Pravděpodobnost : 0,9379977956
Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace
 Hodnota kritéria WA : 0,02852194182
 Kvantil Chi²(1-alfa,1) : 3,841458829
 Pravděpodobnost : 0,865887567
Závěr : Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace
 Hodnota kritéria DW : -1
 Kritické hodnoty DW 0,95
Závěr : Rezidua nejsou autokorelována

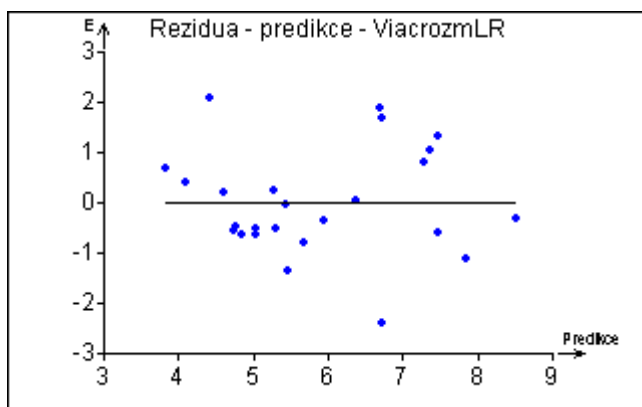
Znaménkový test reziduí
 Hodnota kritéria S_g : 0,7550025901
 Kvantil N(1-alfa/2) : 1,959963999
 Pravděpodobnost : 0,4502474281
Závěr : V reziduích není trend.

Grafy

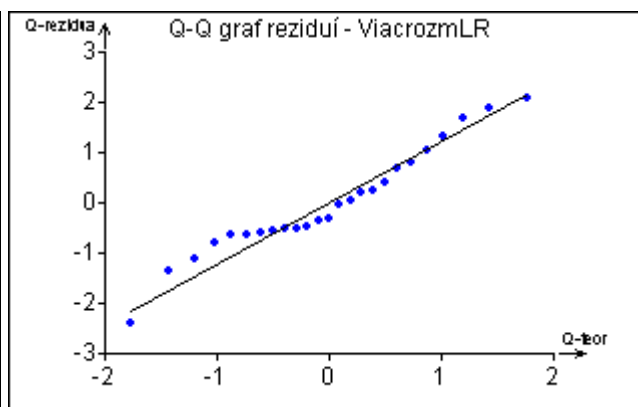


Graf 43: Parciální regresní graf vyjadruje závislosť HbA1c od glukózy. Naznačený lineárny trend je potvrdený významnosťou koeficienta β_3 .

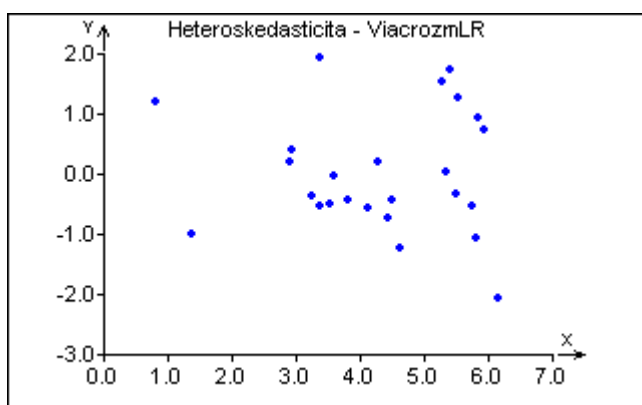
Reziduá



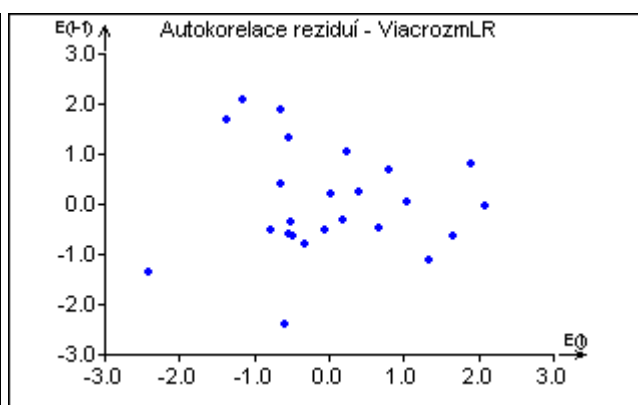
Graf 44: Graf reziduí



Graf 45: Graf normality reziduí



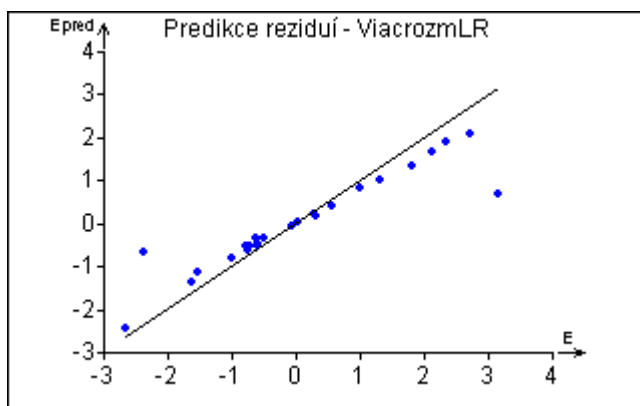
Graf 46: Graf homoskedasticity



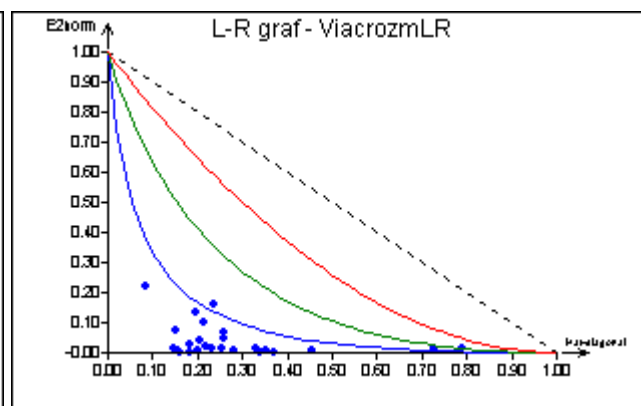
Graf 47: Graf autokorelácie reziduí

Vyššie uvedené testy a grafy 44 – 47 potvrdzujú, že sú splnené všetky predpoklady na použitie metódy najmenších štvorcov. Zvolený model je významný, z hľadiska multikolinearity je korektný. Reziduá náhodne oscilujú okolo nulovej línie. Dáta vykazujú lineárnu závislosť, je patrná normalita. Dáta nie sú autokorelované, nie je medzi nimi trend.

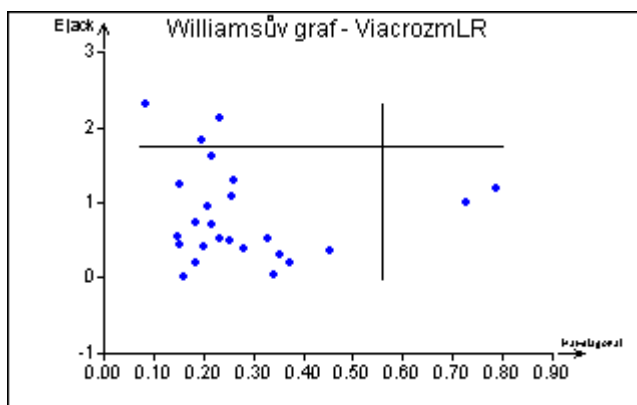
Grafy vplyvných bodov



Graf 48: Graf predikovaných reziduí.

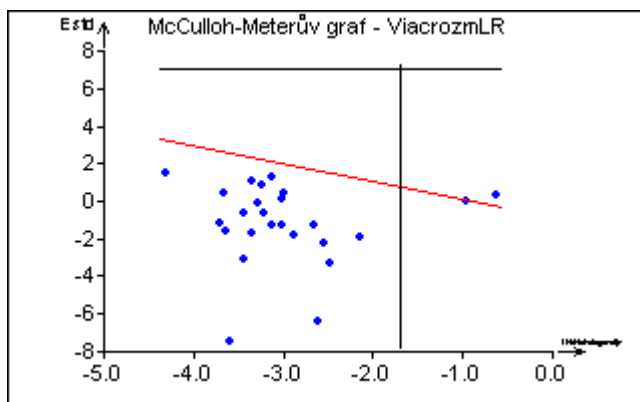
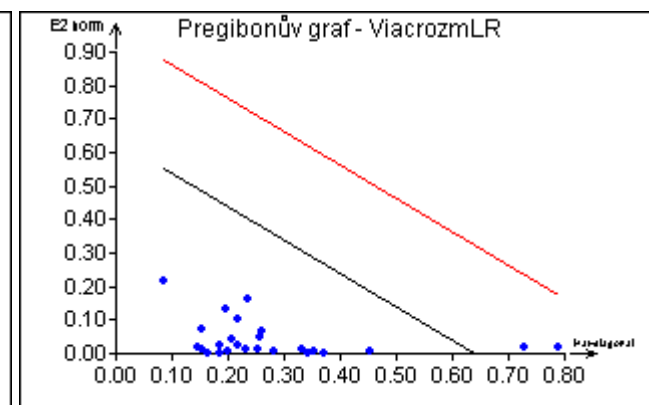


Graf 49: L-R graf – body 9 a 11 extrémny.



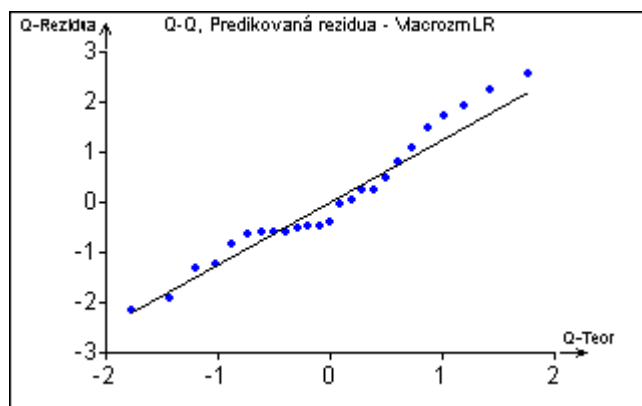
Graf 50: Williamsonov graf signalizuje odľahlé body 3,7 a 22, extrémny body 9 a 11.

Graf 51: Pregibonov graf indikuje body 9 a 11 ako podozrivé.

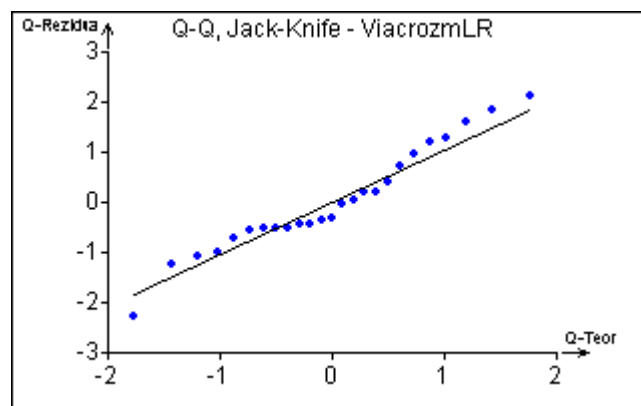


Graf 52: McCulloh-Meeterov graf ukazuje body 9 a 11 ako extrémny.

Rankitové grafy



Graf 53: Graf predikovaných reziduí



Graf 54: Graf Jack-Knife reziduí.

Grafy 53 a 54 potvrdzujú normalitu výberu. Výrazne odchýlky od normality – silne odľahlé body nie sú potvrdené.

Záver: Aplikáciou viacrozmernej lineárnej regresie na sledovanie hladiny glykovaného hemoglobínu (HbA1c) na vzorke 25 pacientov s diabetes mellitus II. typu od ich veku, hladiny cholesterolu (CHOL), plazmatickej hladiny glukózy (GLU), hladín lipoproteínov – nízkozdenzitného (LDL) a vysokozdenzitného (HDL) a hladiny triglyceridov (TGL). Zistilo sa, že jediným faktorom, od ktorého závisí hladina HbA1c je hladina plazmatickej glukózy. Vplyv ostatných sledovaných faktorov je nevýznamný. Ani jeden z dvoch extrémov – bodu 9 a 11 – potvrdené viacerými analýzami a grafmi vplyvných bodov- neboli vylúčené.

Skúmaný jav popisuje model $y = 0,768 (0,155)x$.

4. úloha: Validácia novej analytickej metódy

4.1 Zadanie

Porovnajme namerané hodnoty etanolu ako reziduálneho rozpúšťadla pri výrobe rádiofarmák metódou plynovej chromatografie. Porovnajme starú metódu bez použitia vnútorného štandardu a novú metódu, ktorá využíva vnútorný štandard - metanol. Na základe štatistického testovania regresných parametrov určte či nová metóda vedie ku správnym výsledkom, určte systematickú chybu metódy.

4.2 Dáta (darina_4_sem_praca.xls/validacia, darina_4_sem_praca.vts, validacia.txt)

Tab. 34: Hodnoty pre riešenie zadania č.4

c(EtOH) bez IS - x	c(EtOH) s IS - (y)
0,0140	0,0289
0,0498	0,0527
0,1062	0,1031
0,1698	0,1562
0,3042	0,2919
0,4783	0,4895
1,0129	1,0129

Hodnoty koncentrácie sú v g/l.

4.3 Riešenie

Pre zadanie úlohy bol navrhnutý model $y = \beta_0 + \beta_1 x$ a bude sa testovať hypotéza $H_0: \beta_0 = 0$ a $\beta_1 = 1$ voči alternatívnej hypotéze $H_A: \beta_0 \neq 0$ a $\beta_1 \neq 1$.

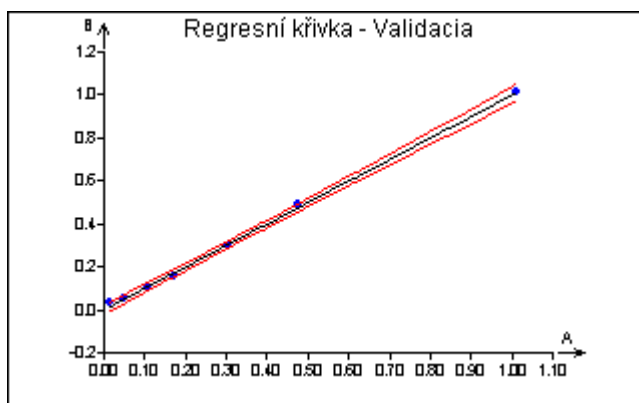
Kritika dát

Hladina významnosti : 0,05
 Kvantil $t(1-\alpha/2, n-m)$: 2,57058183561494
 Kvantil $F(1-\alpha, m, n-m)$: 6,6078909735935
 Absolutní člen : Ano
 Počet platných rádků : 7
 Počet parametrov : 2
 Metoda : Najmenší čtverce
 Sloupce pro výpočet : B
 Abs
 A
 Transformace : Bez transformace

Tab. 35: Odhady parametrov – abs. člena a smernice.

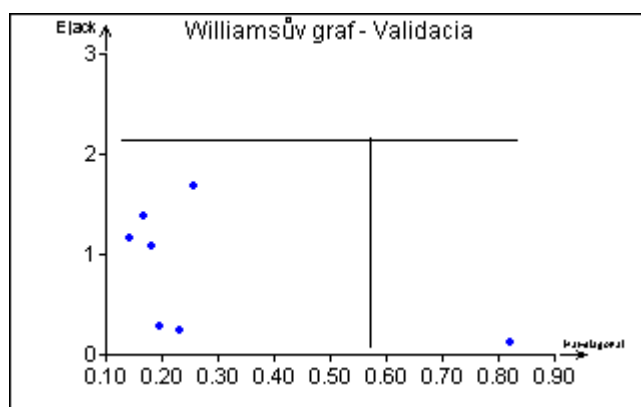
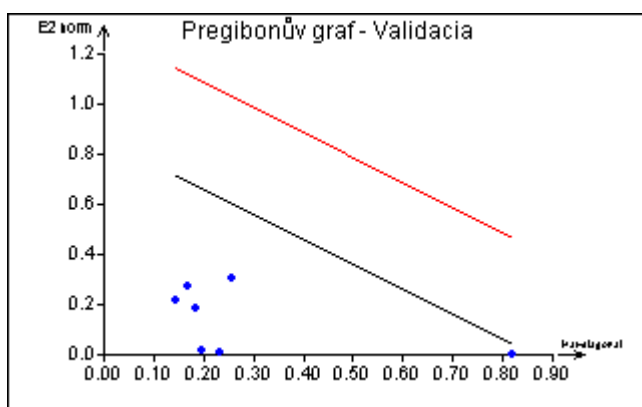
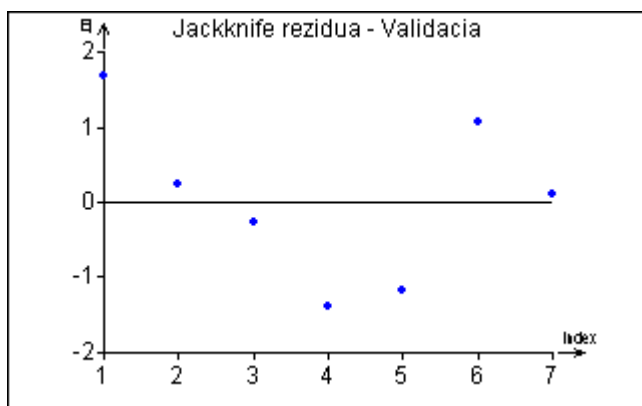
Parameter	Odhad	Odhad. smer. odchýlky	Záver	Pravdepodobnosť	Dolná medza	Horná medza
abs. člen	0,0002767	0,0061394	Nevýznamný	0,96579	-0,015505	0,016058
smernica	0,99909	0,01378	Významný	9,44428E-009	0,96368	1,034508

Absolútny člen lineárneho regresného modelu je vyhodnotený ako nevýznamný, keďže v intervale spoľahlivosti sa nachádza 0, resp. p-hodnota testu je vyššia ako zadaná hladina významnosti 0,05. Hodnota smernice je vyhodnotená ako významná – interval spoľahlivosti obsahuje 1, resp. p-hodnota testu je nižšia ako zadaná hladina významnosti 0,05. **Hypotéza H_0 o nulovosti absolútneho člena a jednotkovej smernici na hladine významnosti 0,05 je prijatá.**



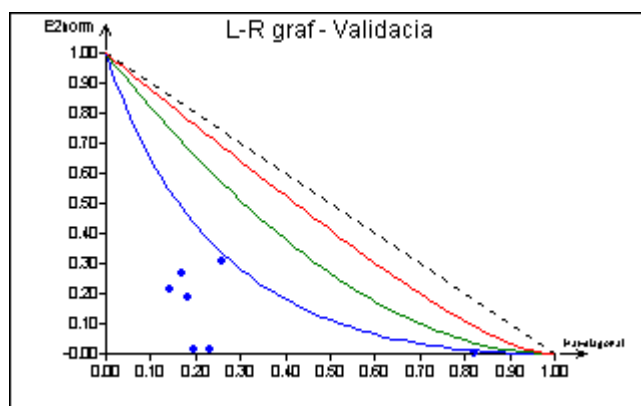
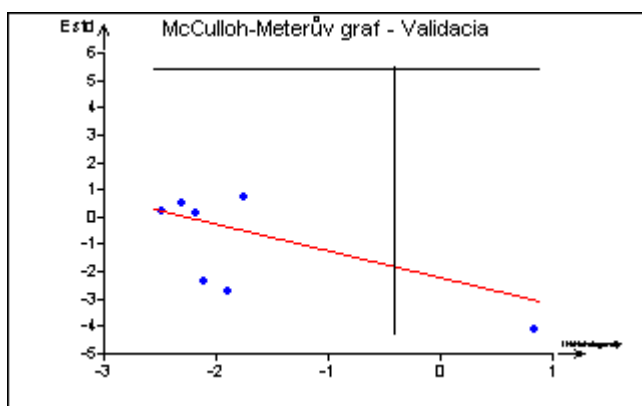
Graf 55: Regresná krivka modelu – predpokladá lineárnu závislosť, v hornej časti sa mierne rozširuje.

Testovanie vplyvných bodov



Graf 56: Bod 7 sa javí ako podozrivý, extrém.

Graf 57: Williamsonov graf indikuje bod 7 ako



Graf 58: McCulloh-Meeterov graf poukazuje na bod 7 ako extrém.

Graf 59: V L-R garfe sa opäť javí bod 7 ako vysoko podozrivý.

Bod 7 bol v aspoň troch grafoch vyhodnotený ako podozrivý až extrém. Jeho prípadné vylúčenie sa bude testovať na vplyv štatistických charakteristík.

Kritika modelu a metódy

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 5258,96478

Kvantil F (1-alfa, m-1, n-m) : 6,607890974
 Pravdepodobnost : 9,444279003E-009
Záver : **Model je významný**

Scottovo kritérium multikolinearity
 Hodnota kritéria SC : 0,3333331617
Záver : **Model vykazuje multikolinearitu!**

Cook-Weisbergův test heteroskedasticity
 Hodnota kritéria CW : 0,165060262
 Kvantil $\chi^2(1-\text{alfa},1)$: 3,841458829
 Pravdepodobnost : 0,6845397508
Záver : **Rezidua vykazujú homoskedasticitu.**

Jarque-Berrův test normality
 Hodnota kritéria JB : 0,47552429
 Kvantil $\chi^2(1-\text{alfa},2)$: 5,991464547
 Pravdepodobnost : 0,7883901913
Záver : **Rezidua majú normálne rozdelenie.**

Waldův test autokorelace
 Hodnota kritéria WA : 0,1256183208
 Kvantil $\chi^2(1-\text{alfa},1)$: 3,841458829
 Pravdepodobnost : 0,7230190908
Záver : **Autokorelace je nevýznamná**

Durbin-Watsonův test autokorelace
 Hodnota kritéria DW : -1
 Kritické hodnoty DW 0
Záver : **Pozitívna autokorelace reziduí není prokázána.**

Znaménkový test reziduí
 Hodnota kritéria Sg : 0,7882407814
 Kvantil $N(1-\text{alfa}/2)$: 1,959963999
 Pravdepodobnost : 0,430555879
Záver : **V reziduích není trend.**

Model je významný, dáta majú normálne rozdelenie a vykazujú homoskedasticitu, v reziduích nie je trend ani autokorelácia. Vykazujú však multikolinearitu (Scottovo kritérium mierne presahuje hodnotu 0,33).

Analýza reziduí

Index	Standardní	Jackknife	Predikované	Diag(H _{ii})	Diag(H* _{ii})	Cook. vzdál.
1	1,43439	1,67239	0,01971	0,25750	0,56304	0,248730
2	0,25698	0,23138	0,00347	0,23103	0,24119	0,038604
3	-0,309019	-0,2791	-0,004082	0,19637	0,21172	-0,03775
4	-1,270184	-1,3804	-0,016486	0,16761	0,43620	-0,12788
5	-1,122009	-1,16018	-0,014351	0,14286	0,35867	-0,09350
6	1,0614096	1,078615	0,0139095	0,183497	0,36747	0,11927
7	0,12822	0,11487	0,0035900	0,82113	0,82172	0,29430

Hodnoty sú zaokrúhlené na 4, resp. 5 desatinných miest podľa matematických pravidiel.

Analýza vplyvu

Index	Atkinson. vzdál.	Andrews- Pregibon st.	Vliv na Y [^]	Vliv na param. LD(b)	Vliv na rozptyl LD(s)	Celk. vliv LD(b,s)
1	1,55723	0,43696	0,98488	0,93382	0,56318	2,018157
2	0,20053	0,75881	0,12683	0,027723	0,06629	0,09042
3	0,21812	0,78828	-0,13795	0,03259	0,06089	0,08944
4	0,97942	0,56380	-0,61944	0,44065	0,21020	0,78576
5	0,74890	0,64133	-0,47365	0,28775	0,06768	0,40419
6	0,80848	0,63253	0,51133	0,34578	0,03705	0,42924
7	0,38916	0,17828	0,24612	0,10487	0,07579	0,16666

Hodnoty sú zaokrúhlené na 4, resp. 5 desatinných miest podľa matematických pravidiel

V tab. 36 sa porovnali modely s vylúčením a bez vylúčenia bodu 7. Vylúčenie bodu 7 nezlepšuje uvedené štatistické charakteristiky, práve naopak. Z vyššie uvedených analýz rezíduí a vplyvu nie je dostatok argumentov na jeho vylúčenie.

Tab 36: Porovnanie regresných modelov s a bez vylúčenia bodu 7.

	Bez vylúčenia bodu 7	S vylúčením bodu 7
RSC	0,0007011120373	0,0006988066924
S²(e)	0,0001402224075	0,0001747016731
R	0,99952496	0,9977216088
R²	0,9990501456	0,9954484087
R_p	0,997018022	0,961558513
MEP	0,0001573370222	0,0004966471537
AIC	-60,46127105	-50,34737523
Tvar modelu	y = 0,0002767 +0,99909 (0,01378)x	y = 0,0008131+0,99565 (0,03366)x

Zhodnotenie kvality modelu

Navrhovaný model má tvar $y = 0,0002767(0,0061394) + 0,99909(0,01378)x$, pričom v zátvorke sú uvedené odhady smerodajnej odchýlky parametrov. Pre intervalový odhad úseku β_0 a smernice β_1 platí:

$$b_0 - t_{(1-\alpha/2, n-m)} \cdot \sqrt{D(b_0)} \leq \beta_0 \leq b_0 + t_{(1-\alpha/2, n-m)} \cdot \sqrt{D(b_0)}$$

po dosadení

$$0,0002767 - 2,571 \cdot 0,0061394 \leq \beta_0 \leq 0,0002767 + 2,571 \cdot 0,0061394$$

vyjde

$$- 0,015508 \leq \beta_0 \leq 0,016061$$

Interval spoľahlivosti úseku regresnej priamky zahŕňa 0, takže tento úsek možno považovať za nulový.

Analogicky vypočítame intervalový odhad smernice β_1 :

$$0,99909 - 2,571 \cdot 0,01378 \leq \beta_1 \leq 0,99909 + 2,571 \cdot 0,01378$$

a po vypočítaní dostaneme

$$0,96366 \leq \beta_1 \leq 1,03452$$

Interval spoľahlivosti smernice regresnej priamky zahŕňa 1, takže smernicu možno považovať za jednotkovú.

Záver

Záverom možno konštatovať, že úsek regresnej priamky sa považuje za nulový a smernica nie je významne odlišná od 1. Výsledky novej metódy s použitím vnútorného štandardu sa štatisticky významne nelíšia od metódy stanovenia etanolu bez použitia vnútorného štandardu.

Príloha

CD so súbormi patriacimi k tejto semestrálnej práci:

- ◀ 4_semestralna_praca_Darina_Tothova.doc
- ◀ 4_semestralna_praca_Darina_Tothova.pdf
- ◀ darina_4_sem_praca.xls, darina_4_sem_praca.vts
- ◀ darina4A.txt, darina4B.txt, darina4C.txt, darina4D.txt
- ◀ polynom.txt, polynom1.txt
- ◀ validacia.txt
- ◀ súbory pre ADSTAT