



Univerzita
Pardubice

**Fakulta chemicko technologická
Katedra analytické chemie
licenční studium Management systému jakosti**

Tvorba lineárních regresních modelů při analýze dat

Autor: Ing. Radek Růčka

Přednášející: Prof. RNDr. Milan Meloun, DrS

1. NAHRAZENÍ ROZTÍRACÍHO STROJKU PULVERISETTE 2 (P2) PŘI PŘÍPRAVĚ VZORKU V LABORATOŘI NOVÝM MODELEM P7 (porovnání dvou regresních přímek u jednoduchého lineárního regresního modelu)

ZADÁNÍ

Součástí kontroly kvality kalcinátu při výrobě titanové běloby je stanovení jeho optických vlastností, konkrétně barvivosti a podtónu v šedé pastě. Příprava šedé pasty se provádí na roztíracím strojku. V souvislosti se změnou přístrojového vybavení laboratoře byla provedena náhrada zastaralého roztíracího zařízení Pulverisette 2 novým modelem Pulverisette 7.

Na základě naměřených dat vyhodnoťte a porovnejte dosažené hodnoty barvivosti a podtónu rutilové modifikace titanové běloby v případě použití obou modelů roztíracího strojku. Porovnejte shodu obou modelů.

DATA

Tabulka 1.1: Stanovení podtónu a barvivosti

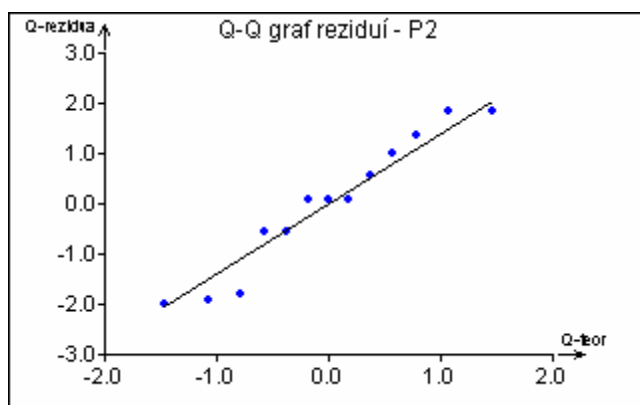
vzorek	P2-barvivot [X]	P2-podtón [Y]	P7-barvivot [X]	P7-podtón [Y]
1	1810	9,5	1825	10,5
2	1810	9,5	1825	10,0
3	1830	12,5	1840	13,0
4	1800	8,0	1815	8,5
5	1845	12,5	1850	12,5
6	1830	13,0	1845	13,0
7	1800	8,0	1820	8,5
8	1810	10,0	1820	10,0
9	1840	13,0	1845	12,5
10	1830	13,0	1845	12,5
11	1840	10,0	1845	11,0
12	1845	10,5	1845	10,5
13	1810	6,0	1820	6,5
14	1820	8,5	1835	9,0

Program: QC.Expert, modul Lineární regrese

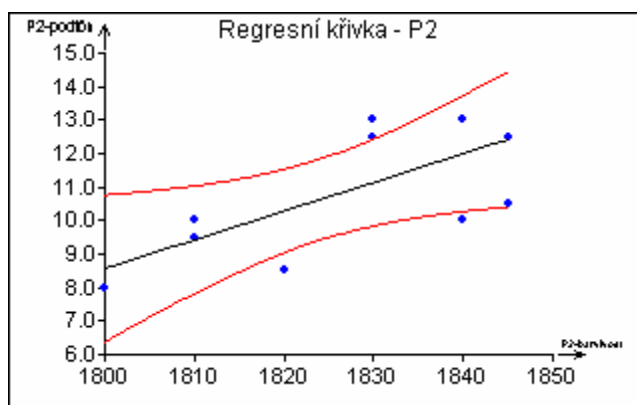
Název souboru s daty: data_porovnaní.txt

VYHODNOCENÍ

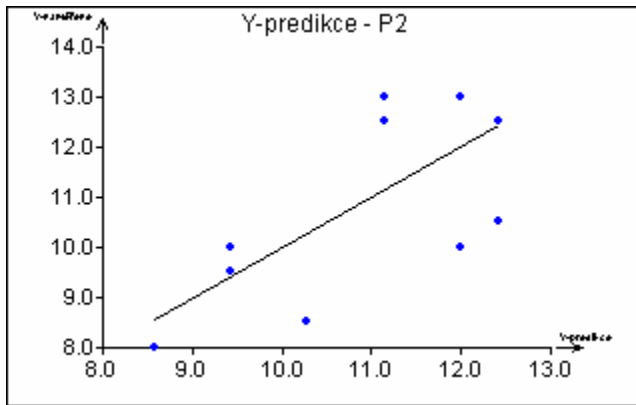
Lineární regresní model (Pulverisette 2)



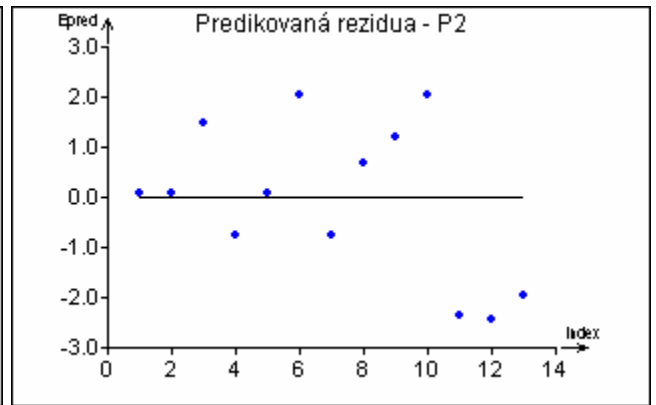
Obrázek 1.1: Q-Q graf pro posouzení normality reziduí. Přímka odpovídá normálnímu (Gaussovu) rozdělení reziduí, data přibližně normálnímu rozdělení odpovídají.



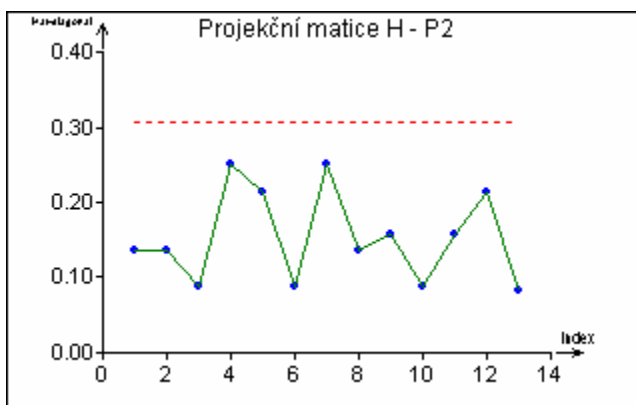
Obrázek 1.2: Graf představuje průběh regresního modelu. Červeně je vyznačen pás spolehlivosti modelu na zadané hladině významnosti.



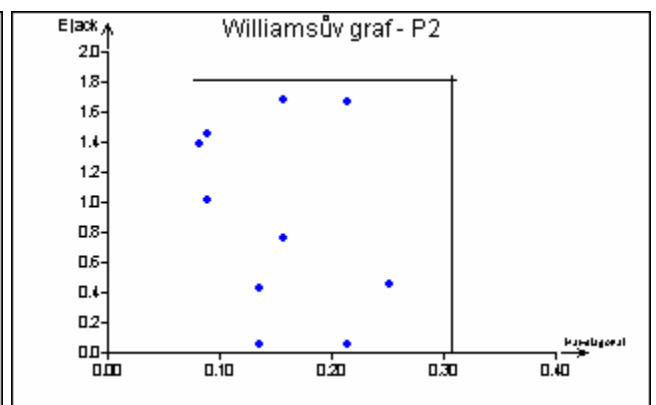
Obrázek 1.3: Graf vyjadřující těsnost proložení, svislé vzdálenost bodu od přímky odpovídá reziduu.



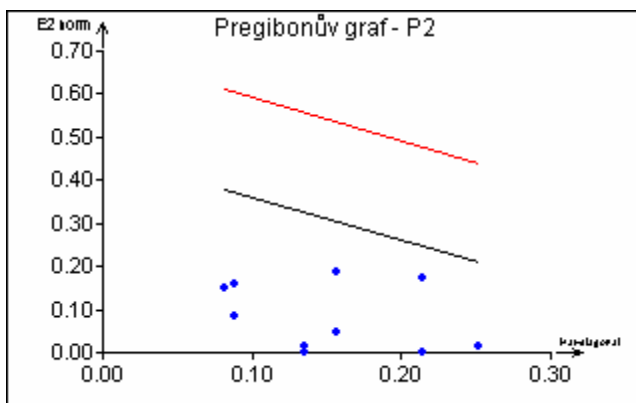
Obrázek 1.4: Predikovaná rezidua jsou velmi citlivým indikátorem vybočujících bodů.



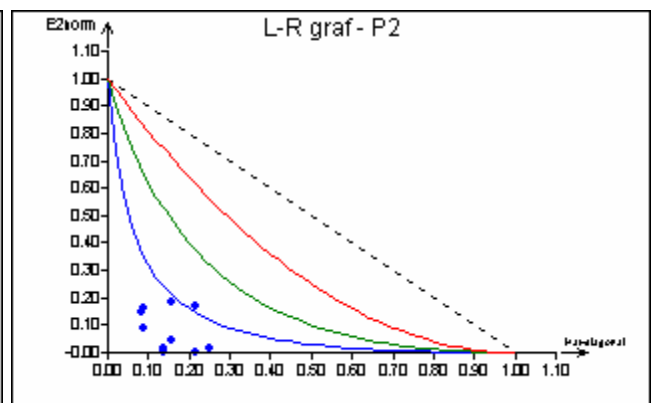
Obrázek 1.5: Body nad vodorovnou přímkou se považují za silně vlivné, v tomto případě nebyly indikovány.



Obrázek 1.6: Williamsův graf slouží k indikaci vlivných i vybočujících bodů. Vlivné, ani vybočující body nebyly nalezeny.



Obrázek 1.7: Graf pro společné posouzení vybočujících bodů a vlivných bodů. Body nad vyšší (červenou) přímkou za silně vlivné nebo vybočující.



Obrázek 1.8: Podle polohy bodů vůči třem křivkám lze data rozdělit na slabě vlivná, vlivná a silně vlivná. Opět nebyl identifikován podezřelý bod.

Regresní diagnostika (Pulverisette 2)

Hladina významnosti :	0,05
Kvantil $t(1-\alpha/2, n-m)$:	2,20098516009141
Kvantil $F(1-\alpha, m, n-m)$:	4,84433567494262
Absolutní člen :	Ano
Počet platných řádků :	13
Počet parametrů :	2
Metoda :	Nejmenší čtverce
Transformace :	Bez transformace

Základní analýza

Charakteristiky proměnných

Proměnná	Průměr	Směr.Odch.	Kor.vs.Y	Významnost
P2-barvivosť	1823,846154	16,47647896	0,7246624206	0,005078506929

Analýza rozptylu

Průměr Y :	10,61538462
Zdroj	Součet čtverců
Celková variabilita	45,57692308
Variabilita vysvětlená modelem	23,93406593
Reziduální variabilita	21,64285714
Hodnota kritéria F :	12,16450876
Kvantil $F(1-\alpha, m-1, n-m)$:	4,844335675
Pravděpodobnost :	0,005078506929
Závěr :	Model je významný

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
Abs	-145,714	44,824	Významný	0,0077
	Spodní mez	Horní mez		
	-244,371	-47,057		
P2-barvivosť	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
	0,0857	0,0245	Významný	0,00507
	Spodní mez	Horní mez		
	0,0316	0,139		

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,7246624206
Koeficient determinace R^2 :	0,5251356239
Predikovaný korelační koeficient R_p :	0,1325166388
Střední kvad. chyba predikce MEP :	2,229663925
Akaikeho informační kritérium :	10,62643954

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	12,16450876
Kvantil $F(1-\alpha, m-1, n-m)$:	4,844335675
Pravděpodobnost :	0,005078506929

Závěr : Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC : -0,3028292815

Závěr : Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 1,29659553

Kvantil $\chi^2(1-\alpha,1)$: 3,841458829

Pravděpodobnost : 0,2548360268

Závěr : Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 0,7164254669

Kvantil $\chi^2(1-\alpha,2)$: 5,991464547

Pravděpodobnost : 0,6989243746

Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA : 0,3932921394

Kvantil $\chi^2(1-\alpha,1)$: 3,841458829

Pravděpodobnost : 0,2548360268

Závěr : Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : 1,478547855

Kritické hodnoty DW 0 2

Závěr : Pozitivní autokorelace reziduí není prokázána.

Znaménkový test reziduí

Hodnota kritéria Sg : 0,4021861899

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

Pravděpodobnost : 0,687547005

Závěr : V reziduích není trend.

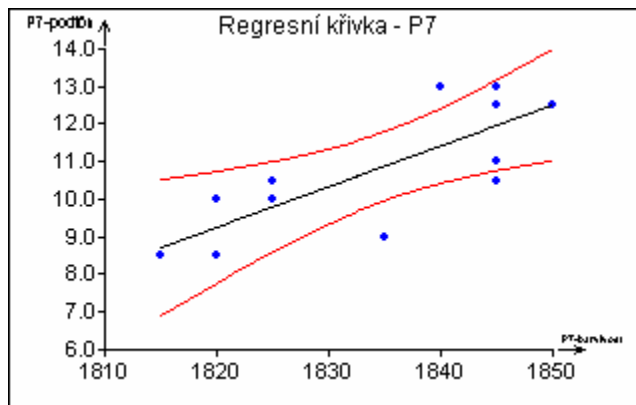
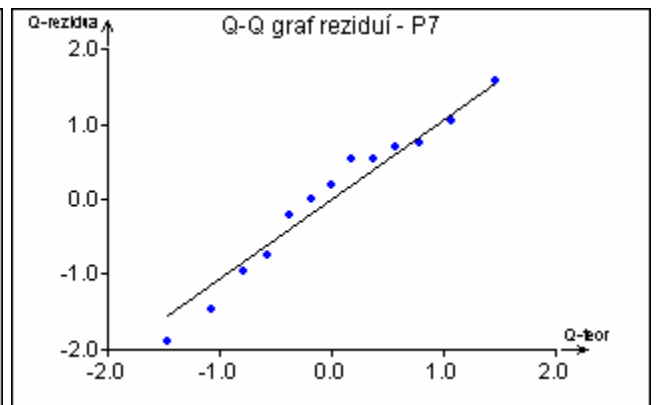
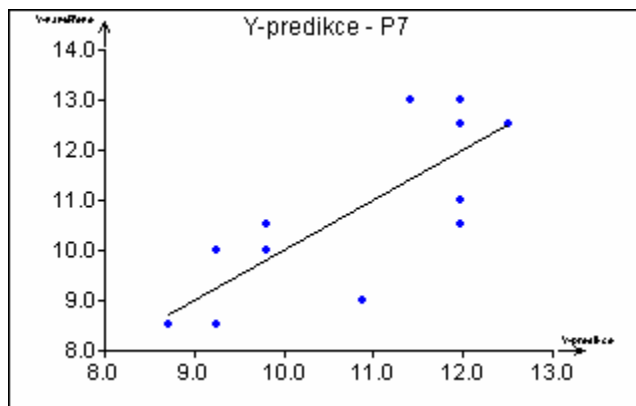
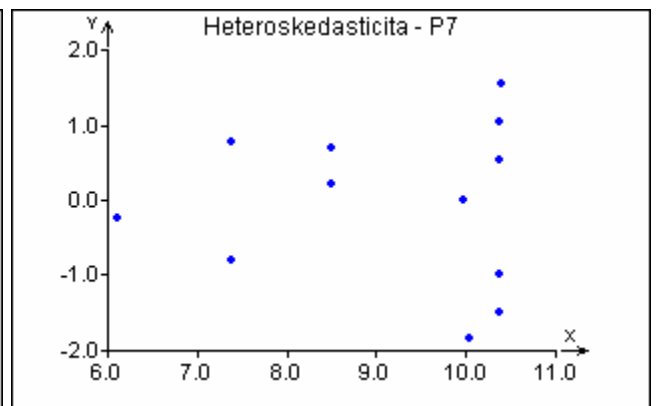
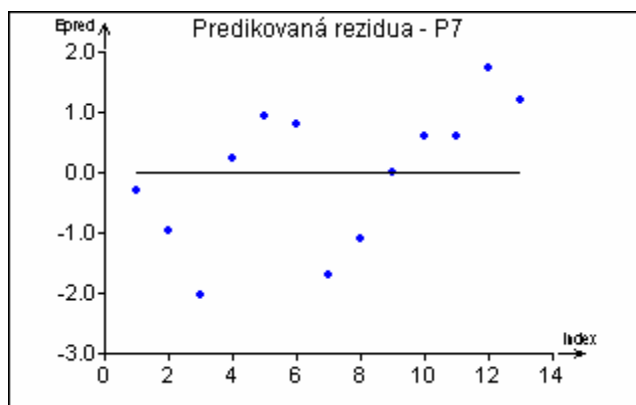
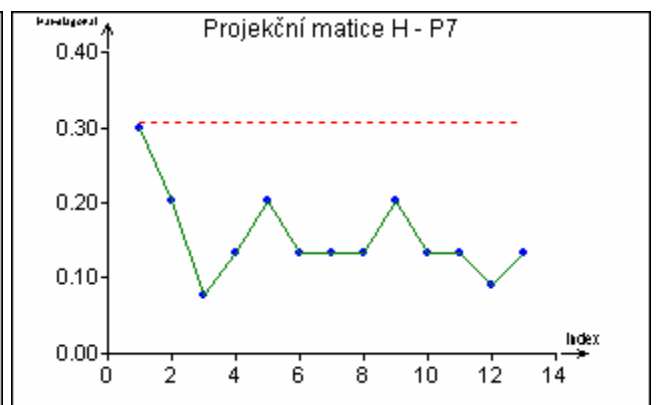
Závěr (Pulverisette 2)

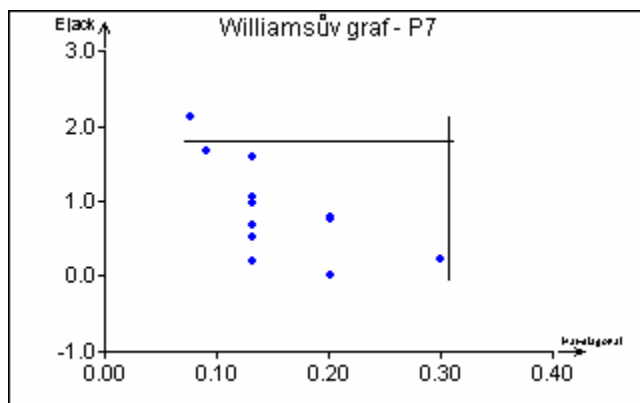
Na základě výsledků testování byl lineární model určen jako významný a korektní. Parametry regresní přímky byly určeny jako významné. Testem reziduí a také z jejich grafického zobrazení vyplývá, že je model v pořádku.

Rovnice regresní přímky

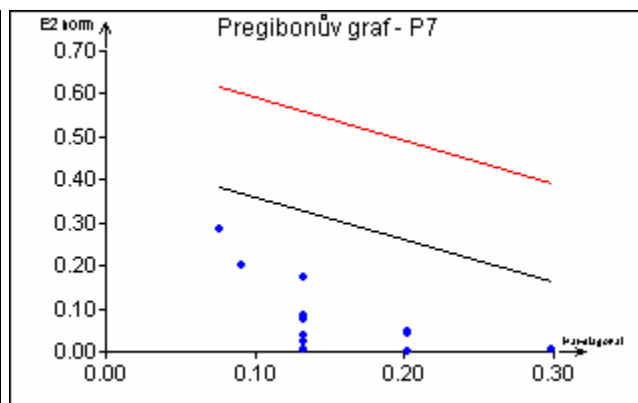
$$\text{“podtón“} = 0,0857 (\pm 0,0245) * \text{“barvivot“} - 145,714 (\pm 44,824)$$

$$Y = 0,0857 (\pm 0,245) * X - 1,579 (\pm 0.113)$$

Lineární regresní model (Pulverisette 7)**Obrázek 1.10:** Průběh regresního modelu.**Obrázek 1.11:** Data odpovídají normálnímu rozdělení.**Obrázek 1.12:** Těsnost proložení dat.**Obrázek 1.13:** Grafické posouzení nekonstantnosti rozptylu.**Obrázek 1.14:****Obrázek 1.15:** Není identifikován vlivný bod.



Obrázek 1.16: Jeden se nachází nad vodorovnou přímkou, mohlo by se jednat o vlivný bod.



Obrázek 1.17: Není identifikován vlivný bod.

Regresní diagnostika (Pulverisette 7)

Hladina významnosti :	0,05
Kvantil $t(1-\alpha/2, n-m)$:	2,20098516009141
Kvantil $F(1-\alpha, m, n-m)$:	4,84433567494262
Absolutní člen :	Ano
Počet platných řádků :	13
Počet parametrů :	2
Metoda :	Nejmenší čtverce
Transformace :	Bez transformace

Základní analýza

Charakteristiky proměnných

Proměnná	Průměr	Směr.Odch.	Kor.vs.Y	Významnost
P7-barvivost	1835	12,24744871	0,7931912432	0,00121481722

Analýza rozptylu

Průměr Y :	10,88461538
Zdroj	Součet čtverců
Celková variabilita	33,57692308
Variabilita vysvětlená modelem	21,125
Reziduální variabilita	12,45192308
Hodnota kritéria F :	18,66177606
Kvantil F (1-alfa, m-1, n-m) :	4,844335675
Pravděpodobnost :	0,00121481722
Závěr :	Model je významný

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
Abs	-187,907	46,018	Významný	0,00181
	Spodní mez	Horní mez		
	-289,192	-86,621		
P7-barvivost	Odhad	Směr.Odch.	Významný	Pravděpodobnost
	0,108	0,0251		
	Spodní mez	Horní mez		
	0,0531	0,1630		

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R : 0,7931912432
Koeficient determinace R^2 : 0,6291523482
Predikovaný korelační koeficient R_p : 0,2738844451
Střední kvad. chyba predikce MEP : 1,231137393
Akaikeho informační kritérium : 3,440034328

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 18,66177606
Kvantil F (1- α , m-1, n-m) : 4,844335675
Pravděpodobnost : 0,00121481722
Závěr : Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC : -0,3087844646
Závěr : Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0,2889638997
Kvantil $\chi^2(1-\alpha, 1)$: 3,841458829
Pravděpodobnost : 0,5908852539
Závěr : Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 0,7206230224
Kvantil $\chi^2(1-\alpha, 2)$: 5,991464547
Pravděpodobnost : 0,6974590259
Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA : 2,066719299
Kvantil $\chi^2(1-\alpha, 1)$: 3,841458829
Pravděpodobnost : 0,5908852539
Závěr : Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : 1,102574003
Kritické hodnoty DW 0 2
Závěr : Pozitivní autokorelace reziduí není prokázána.

Znaménkový test reziduí

Hodnota kritéria Sg : 1,72696102
Kvantil $N(1-\alpha/2)$: 1,959963999
Pravděpodobnost : 0,08417466421
Závěr : V reziduích není trend.

Závěr (Pulverisette 2)

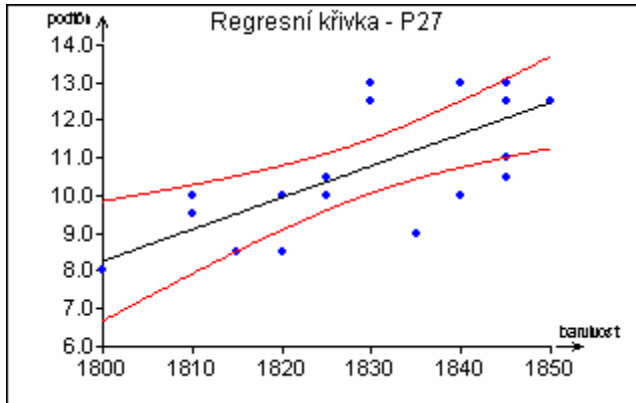
Na základě výsledků testování byl lineární model určen jako významný a korektní. Parametry regresní přímky byly určeny jako významné. Testem reziduí a také z jejich grafického zobrazení vyplývá, že je model v pořádku.

Rovnice regresní přímky

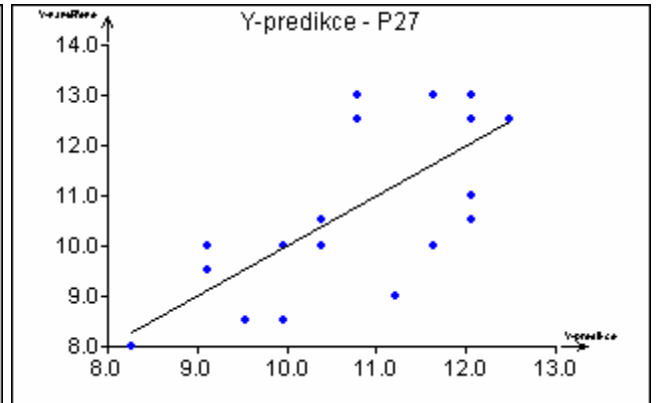
$$\text{“podtón“} = 0,108 (\pm 0,026) * \text{“barvivosť“} - 187,907 (\pm 46,018)$$

$$Y = 0,108 (\pm 0,026) * X - 187,907 (\pm 46,018)$$

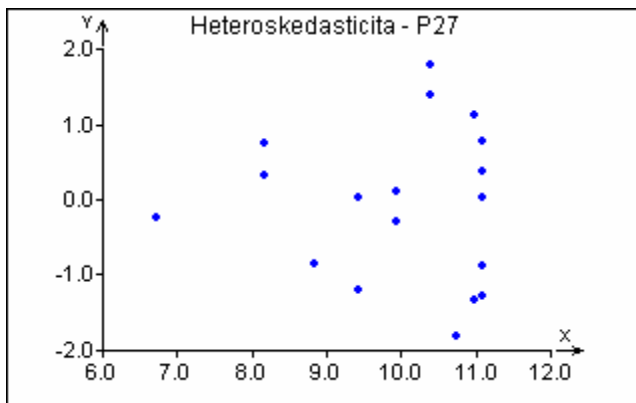
Lineární regresní model (společný model)



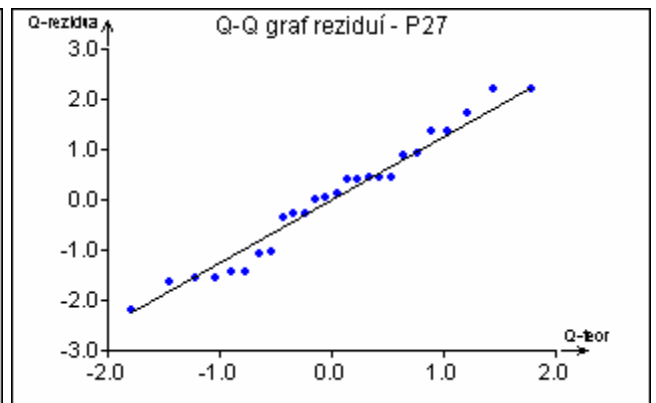
Obrázek 1.18: Podle rozložení bodů a proložení regresní přímky není jednoznačně zřetelný původ ve dvou zdrojích dat.



Obrázek 1.19: Není dosaženo příliš dobrého proložení.



Obrázek 1.20: Mírný příklon k heteroskedasticitě nebyl potvrzen testováním regr. tripletu.



Obrázek 1.21: Data vykazují normální rozložení.

Regresní diagnostika (společný model)

Hladina významnosti :	0,05
Kvantil $t(1-\alpha/2, n-m)$:	2,06389856105585
Kvantil $F(1-\alpha, m, n-m)$:	4,25967727032841
Absolutní člen :	Ano
Počet platných řádků :	26
Počet parametrů :	2
Metoda :	Nejmenší čtverce
Transformace :	Bez transformace

Základní analýza

Charakteristiky proměnných

Proměnná	Průměr	Směr.Odch.	Kor.vs.Y	Významnost
barvivot	1829,423077	15,31841526	0,722434609	3,078847921E-005

Analýza rozptylu

Průměr Y :	10,75
Zdroj	Součet čtverců
Celková variabilita	79,625
Variabilita vysvětlená modelem	41,55722423
Reziduální variabilita	38,06777577
Hodnota kritéria F :	26,19993843
Kvantil F (1-alfa, m-1, n-m) :	4,25967727
Pravděpodobnost :	3,078847921E-005
Závěr :	Model je významný

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
Abs	-143,226	30,082	Významný	7,6238E-005
	Spodní mez	Horní mez		
	-205,314	-81,138		
P7-barvivot	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
	0,0841	0,0164	Významný	3,079E-005
	Spodní mez	Horní mez		
	0,050	0,118		

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,722434609
Koeficient determinace R ² :	0,5219117642
Predikovaný korelační koeficient Rp :	0,2146290343
Střední kvad. chyba predikce MEP :	1,643702766
Akaikeho informační kritérium :	13,91306176

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	26,19993843
Kvantil F (1-alfa, m-1, n-m) :	4,25967727
Pravděpodobnost :	3,078847921E-005
Závěr :	Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC :	-0,3019645433
Závěr :	Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW :	0,6353058387
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,4254157504
Závěr :	Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 0,8303192111
 Kvantil $\chi^2(1-\alpha,2)$: 5,991464547
 Pravděpodobnost : 0,6602348951
 Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA : 9,169265878
 Kvantil $\chi^2(1-\alpha,1)$: 3,841458829
 Pravděpodobnost : 0,4254157504
 Závěr : Autokorelace je významná.

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : 0,8069639039
 Kritické hodnoty DW 1,22 1,55
 Závěr : Rezidua jsou pozitivně autokorelována!

Znaménkový test reziduí

Hodnota kritéria Sg : 2,131130336
 Kvantil $N(1-\alpha/2)$: 1,959963999
 Pravděpodobnost : 0,03307840619
 Závěr : V reziduích je trend!

Test shody obou modelů – ověření shody rozptylů

Hladina významnosti : 0,05
 Porovnávané sloupce : 2 4

Počet dat :	13	13
Průměr :	10,61538462	10,88461538
Směr. odchylka :	1,948865548	1,672745325
Rozptyl :	3,798076923	2,798076923

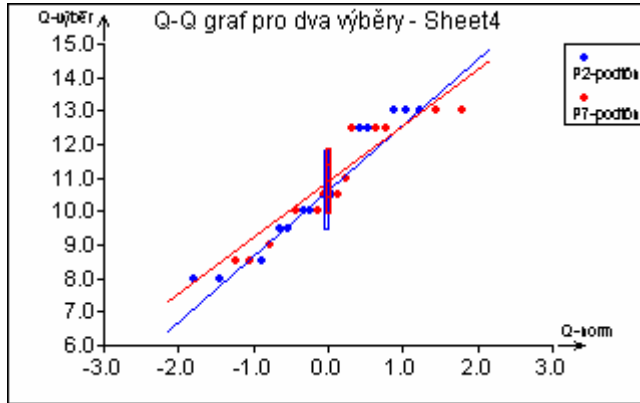
Korel. koef. $R(x,y)$: 0,1130658738

Test shody rozptylů

Poměr rozptylů : 1,357388316
 Počet stupňů volnosti : 12 12
 Kritická hodnota : 2,686637112
 Závěr : Rozptyly jsou SHODNÉ
 Pravděpodobnost : 0,3024493435

Robustní test shody rozptylů

Poměr rozptylů : 1,357388316
 Redukované stupně volnosti : 11 11
 Kritická hodnota : 2,81793047
 Závěr : Rozptyly jsou SHODNÉ
 Pravděpodobnost : 0,3105079109



Obrázek 1.22: Porovnání dvou výběrů. Data se berou jako jediný soubor, příslušnost k prvnímu nebo druhému výběru je rozlišena barvou. Přímký znázorňují polohu střední hodnoty, jejich směrnice odpovídají směrodatné odchylce

ZÁVĚR

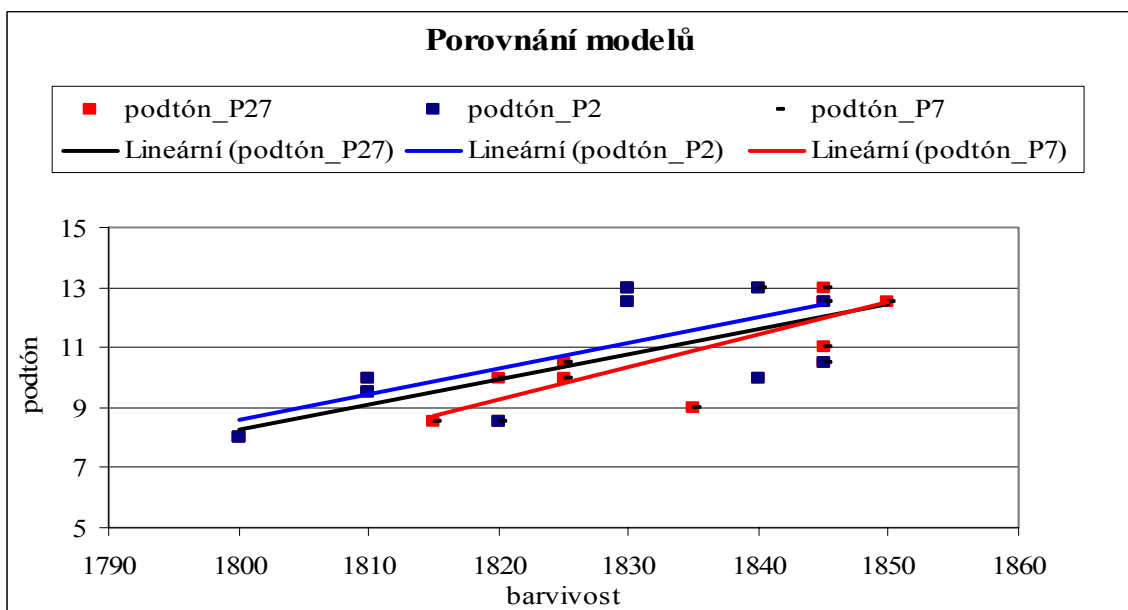
Byla prokázána shodnosti rozptylů (potvrzena homoskedasticita) u obou výběrů dat, proto lze použít Chowův test hypotézy $H_0: \beta_1 = \beta_2$ proti alternativní $H_A: \beta_1 \neq \beta_2$, založený na textačním kritériu :

$$F_c = (RSC - RSC_1 - RSC_2)(n - 2m) / (RSC_1 + RSC_2)(m)$$

$$F_c = (38,068 - 21,642 - 12,452)(26 - 2 \cdot 2) / (21,642 + 12,452)(2) = 0,162$$

$$F_{(2,22)} = 1,460$$

$F_c < F_{(2,22)} \rightarrow$ Je nutno hypotézu o shodě parametrů přijmout.



Obrázek 1.23: Grafické porovnání obou modelů a jejich lineárních trendů (proloženo metodou nejmenších čtverců – využito MS Excel)

Z porovnání obou modelů vyplývá, že se stanovení podtónu je pro oba modely srovnatelné.

STANOVENÍ ÚČINNOSTI MLETÍ NA MOKRÝCH MLÝNECH NETZSCH (Určení stupně polynomu metodou MNČ)

ZADÁNÍ

V souvislosti se zahájením provozu mokrých perlových mlýnů Netzsch byla realizovaná provozní zkouška, která měla za cíl určit vliv režimu provozování mlýna (otáček) na vlastnosti materiálu. Na základě laboratorních testů byl jako sledovaný parametr vybrána hodnota SFM 2, resp. rozdíl její hodnoty u vstupního a výstupního materiálu. Metoda SFM 2 je založena na spektrofotometrickém měření turbidity zředěné disperze oxidu titaničitého.

Úkolem je určit stupeň polynomu pro proměřený rozsah otáček mokrých mlýnů.

DATA

Tabulka 2.1: Závislost SFM2 na otáčkách mlýna Netzsch

otáčky mlýna	600	800	1000	1100	1200	1300	1400	1500
změna SFM2	5,26	6,47	9,82	11,26	12,54	15,09	17,44	19,40

Program: QCExpert 2.5: modul Lineární regrese

Název souboru s daty: data_polynom.txt

VYHODNOCENÍ

Pro nalezení optimálního stupně polynomu n byla použita metoda nejmenších čtverců (MNČ). Při výpočtu bude měněn stupeň polynomu a podle posouzení hodnot MEP (střední kvadratické chyby predikce) a AIC (Akaikeho informační kritérium) bude vybrán optimální stupeň. Testování bude probíhat na hladině významnosti $\alpha = 0.05$, pro $n = 1, 2, 3, 4, 5$.

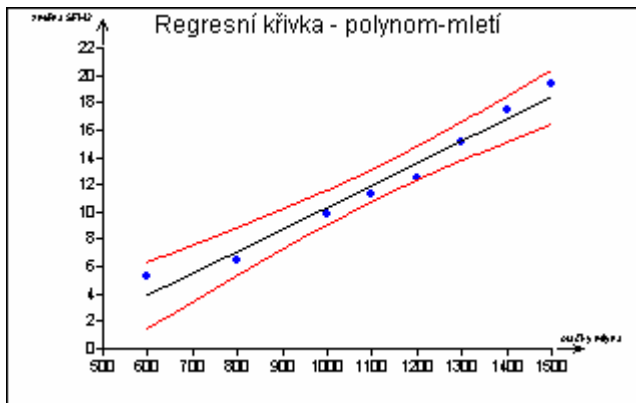
Nalezení optimálního stupně polynomu – návrh modelu

Pro nalezení optimálního stupně polynomu byla využita MNČ. Toto bylo provedeno na základě vyhodnocení hodnoty MEP (střední kvadratická chyba predikce) a AIC (Akaikeho informační kritérium).

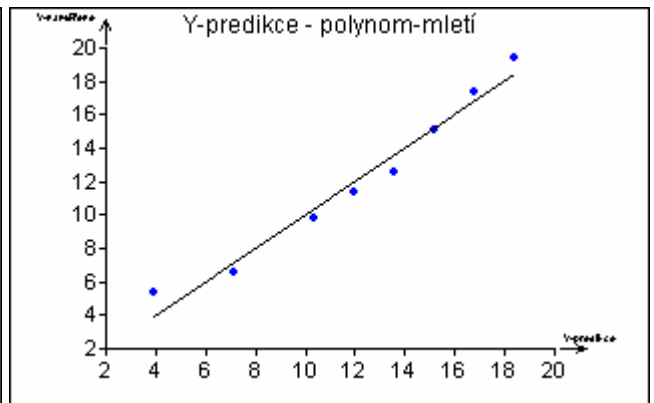
Tabulka 2.2: Statistické charakteristiky

Stupeň polynomu	1	2	3	4	5
Vícenásobný korelační koeficient R	0,9838	0,9982	0,9985	0,9985	0,9996
Koeficient determinace R^2 (D)	0,9679	0,9965	0,9969	0,9971	0,9993
Predikovaný korelační koeficient R_p	0,8336	0,9743	0,9465	0,0292	0,9137
MEP	1,8923	0,2813	0,5894	18,037	0,9591
AIC	1,1328	-14,535	-13,654	-12,050	-21,399

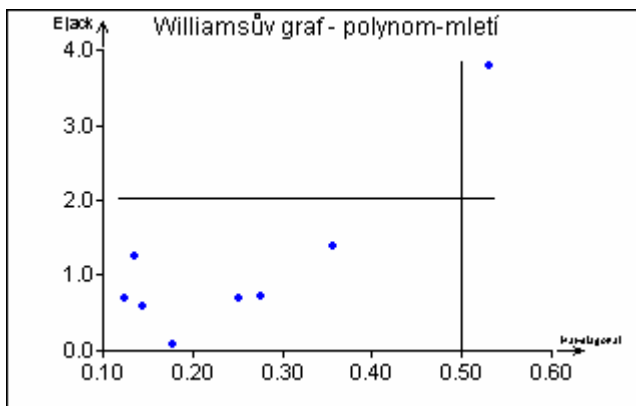
bez transformace



Obrázek 2.1: Graf ukazuje na těsnost proložení. Všechny body jsou uvnitř pásu spolehlivosti.

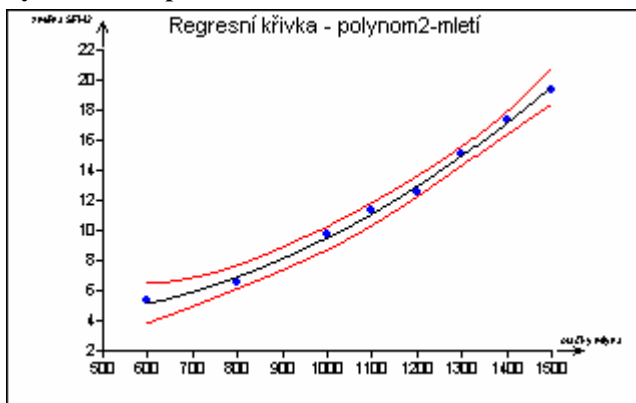


Obrázek 2.2: Graf vyjadřuje těsnost proložení. Svislé vzdálenosti bodu od přímky odpovídá reziduum.

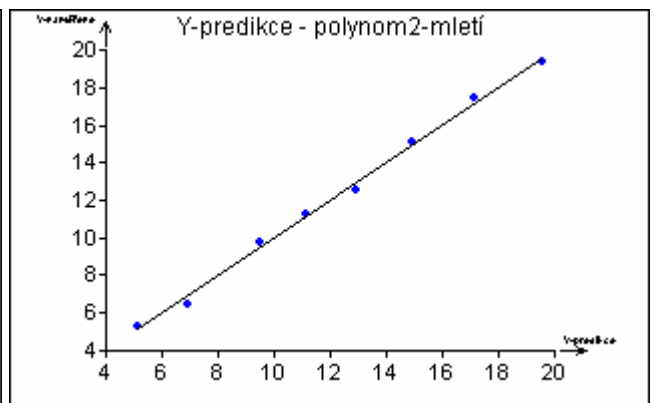


Obrázek 2.3: Graf identifikuje jeden bod, který je vlivným a současně vybočujícím.

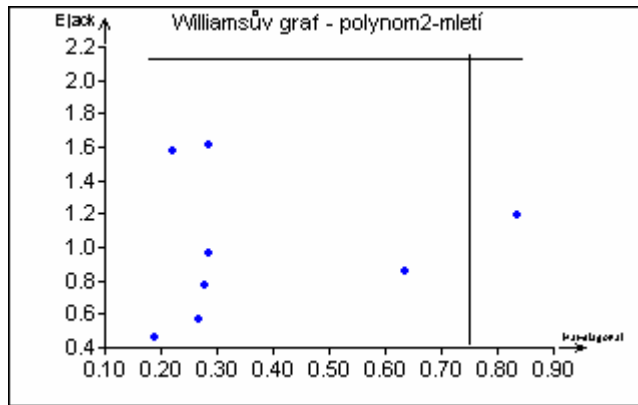
polynom 2. stupně



Obrázek 2.4: Graf ukazuje na těsnost proložení pro stupeň polynomu 2. Všechny body jsou uvnitř pásu spolehlivosti. V porovnání s předchozím případem je zřejmé zúžení pásu.

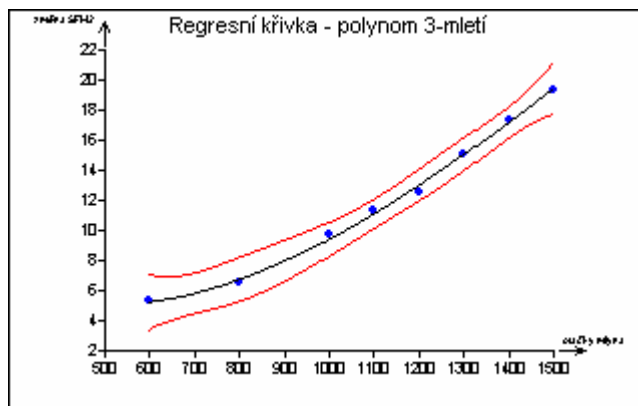


Obrázek 2.5: Graf ukazuje na těsnější proložení.

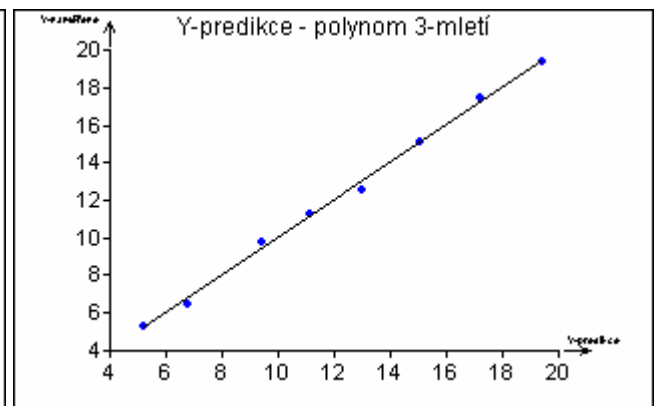


Obrázek 2.6: Graf identifikuje jeden vlivný bod.

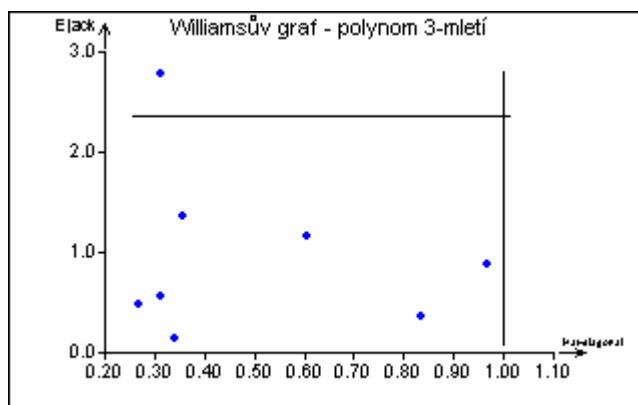
polynom 3. stupně



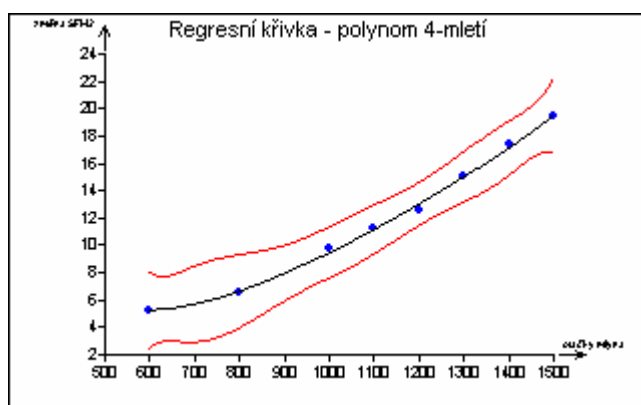
Obrázek 2.7: Graf ukazuje na těsnost proložení pro stupeň polynomu 3. V porovnání s předchozím případem došlo k rozšíření a deformaci pásu.



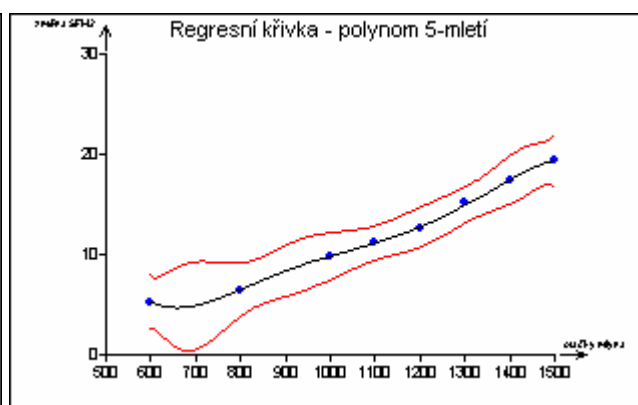
Obrázek 2.8: Nelze již rozlišit těsnost proložení oproti polynomu 2.



Obrázek 2.9: Graf identifikuje jeden vybočující bod.



Obrázek 2.10: Ukázka deformovaného pásu spolehlivosti pro stupeň polynomu 4.



Obrázek 2.11: Ukázka deformovaného pásu spolehlivosti pro stupeň polynomu 5.

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	P	Spodní mez	Horní mez
Abs	4,71318	1,7202	Významný	0,04080	0,29114	9,1352
otáčky mlýna	-0,00543	0,00341	Nevýznamný	0,17228	-0,01420	0,00334
otáčky mlýna ²	1,02322E-05	1,60760E-06	Významný	0,00141	6,09981E-06	1,43647E-05

Indikace multikolinearity

Proměnná	Vlas. čísla kor. m	Podmíněnost kappa	VI faktor	Vícenás. kor.
Abs	0,00815	1	1	0
otáčky mlýna	1	122,55247	61,52725	0,99184
otáčky mlýna ²	1,9918	244,10494	61,52725	0,99184

Podmíněnost kappa, její maximální hodnota $k_{\max} > 1000$ se považuje za indikaci silné multikolinearity. VI faktor, je faktor vzrůstu rozptylu v důsledku multikolinearity, hodnoty $VIF > 10$ se považují za indikaci silné multikolinearity.

Znamená to, že multikolinearita byla identifikována pouze VI faktorem. Pro určení stupně polynomu tak bude využito metody nejmenších čtverců.

SOUHRN A ZÁVĚR

Jako optimální stupeň polynomu byl metodou nejmenších čtverců určen optimální stupeň polynomu 2, kdy je MEP a AIC mají nejmenší hodnoty. Tento polynom vykázal i nejlepší těsnost proložení při posouzení grafických výstupů.

2. STANOVENÍ OBSAHU ZIRKONIA V SÍRANU ZIRKONIČITÉM METODOU RENTGENOVÉ SPEKTROMETRIE (validace nové analytické metody)

ZADÁNÍ:

Síran zirkoničitý je surovinou používanou při povrchové úpravě titanové běloby. Obsah zirkonia je v této surovině stanovován v laboratoři klasickou acidobazickou titrací hydrolyzou uvolněné kyseliny sírové. Použitím instrumentální energiově rentgenové analýzy (EDXRF) na přístroji Philips Minipal PW 4025 by bylo možné zkrátit dobu analýzy z 30 na 2 minuty.

Metodou lineární regrese má být ověřeno, zda obě metody poskytují srovnatelné výsledky a zda je možné nahradit klasickou analýzu instrumentální.

DATA

Tabulka 3.1: Stanovení obsahu ZrO_2

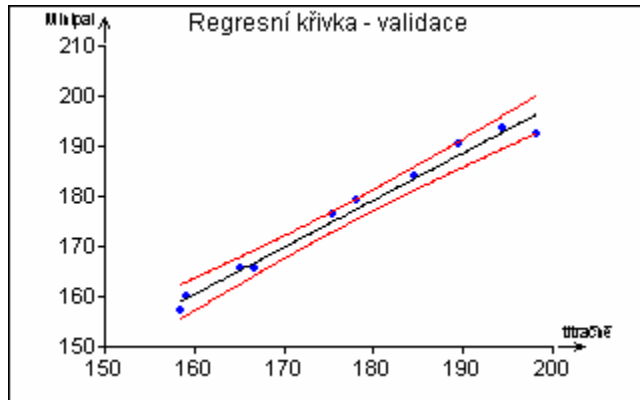
obsah ZrO_2 [g/l]	1	2	3	4	5	6	7	8	9	10
titračně [X]	158,5	175,4	165,1	184,5	159,2	166,8	178,2	189,6	194,4	198,3
Minipal [Y]	157,1	176,5	165,6	183,9	160,1	165,5	179,1	190,5	193,6	192,6

Program: QCExpert 2.5: modul Lineární regrese

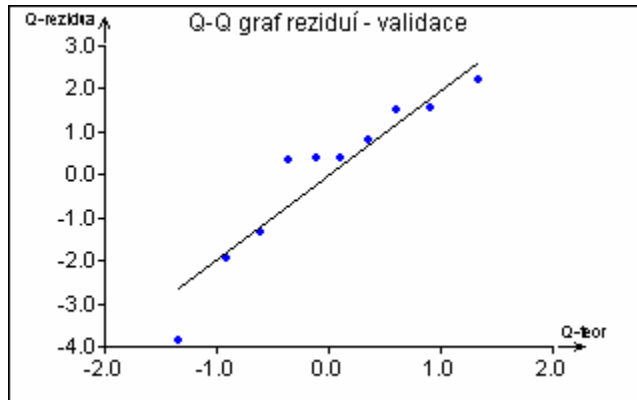
Název souboru s daty: data_validace.txt

VYHODNOCENÍ

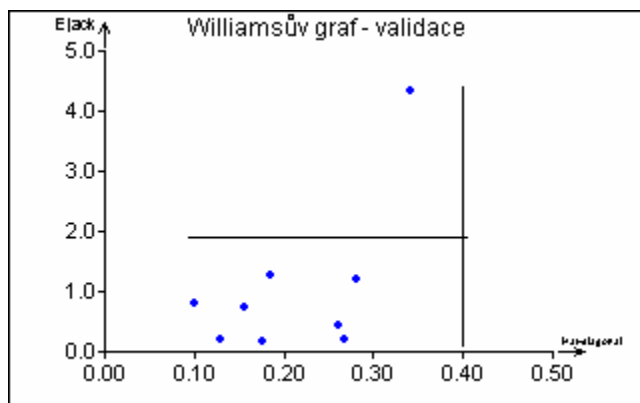
Lineární regresní model



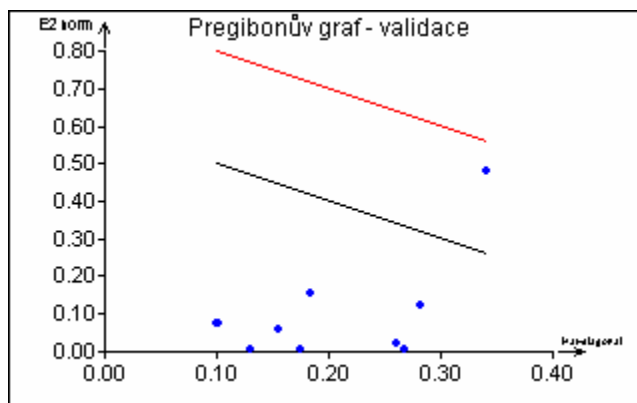
Obrázek 3.1: Graf představuje průběh regresního modelu. Všechny body jsou uvnitř nebo na hranici pásů.



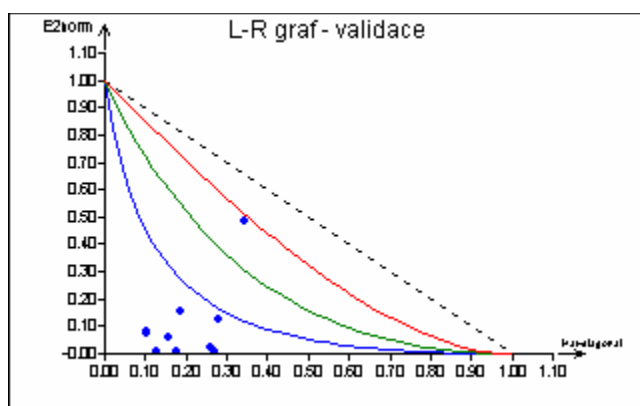
Obrázek 3.2: Graf slouží k posouzení normality rozdělení. Body jsou uloženy kolem přímky (odpovídá normálnímu rozdělení)



Obrázek 3.3: Jeden bod nad vodorovnou přímkou je silně vybočující.



Obrázek 3.4: Poskytuje obdobný výsledek jako graf Williamsův. Ukazuje na jeden vybočující bod.



Obrázek 3.5: Opět indikace jednoho vlivného bodu. Graf je vhodný pro menší počet dat.

Byl identifikován jeden silně vlivný bod. Vzhledem ke skutečnosti, že se jedná o malý počet dat, nebude z výpočtu regrese vypuštěn.

Regresní diagnostika

Hladina významnosti :	0,05
Kvantil $t(1-\alpha/2, n-m)$:	2,30600413519908
Kvantil $F(1-\alpha, m, n-m)$:	5,31765507155526
Absolutní člen :	Ano
Počet platných řádků :	10
Počet parametrů :	2
Metoda :	Nejmenší čtverce
Transformace :	Bez transformace

Základní analýza

Charakteristiky proměnných

Proměnná	Průměr	Směr. Odch.	Kor. vs. Y	Významnost
titračně	177	14,45375768	0,9907091361	3,223682166E-008

Analýza rozptylu

Průměr Y :	176,45		
Zdroj	Součet čtverců	Průměrný čtverec	Rozptyl
Celková variabilita	1694,245	169,4245	188,2494444
Variabilita vysvětlená modelem	1662,909248	166,2909248	184,7676942
Reziduální variabilita	31,33575183	3,133575183	3,481750204
Hodnota kritéria F :	424,5398054		
Kvantil F (1-alfa, m-1, n-m) :	5,317655072		
Pravděpodobnost :	3,22368216E-008		
Závěr :	Model je významný		

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
Abs	9,991676417	8,102995833	Nevýznamný	0,2525461712
	Spodní mez	Horní mez		
	-8,69386548	28,67721832		
Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost
titračně	0,9404425061	0,0456428802	Významný	3,223682166E-008
	Spodní mez	Horní mez		
	0,8351898356	1,045695177		

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,9907091361
Koeficient determinace R ² :	0,9815045924
Predikovaný korelační koeficient Rp :	0,9308521309
Střední kvadratická chyba predikce MEP :	5,962592989
Akaikeho informační kritérium :	15,42174584

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu	
Hodnota kritéria F :	424,5398054
Kvantil F (1-alfa, m-1, n-m) :	5,317655072
Pravděpodobnost :	3,22368216E-008
Závěr :	Model je významný

Scottovo kritérium multikolinearity	
Hodnota kritéria SC :	-0,001787557652
Závěr :	Model je korektní.

Cook-Weisbergův test heteroskedasticity	
Hodnota kritéria CW :	1,553556096
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,2126112318
Závěr :	Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality	
Hodnota kritéria JB :	1,354241937
Kvantil Chi ² (1-alfa,2) :	5,991464547
Pravděpodobnost :	0,5080776602
Závěr :	Rezidua mají normální rozdělení.

Waldův test autokorelace	
Hodnota kritéria WA :	0,06268654868

Kvantil $\chi^2(1-\alpha,1)$: 3,841458829
 Pravděpodobnost : 0,2126112318
 Závěr : Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : 1,575239253
 Kritické hodnoty DW 0 2
 Závěr : Pozitivní autokorelace reziduí není prokázána.

Znaménkový test reziduí

Hodnota kritéria Sg : 0,2454951265
 Kvantil $N(1-\alpha/2)$: 1,959963999
 Pravděpodobnost : 0,8060730812
 Závěr : V reziduích není trend.

SOUHRN A ZÁVĚR

Na základě výsledků testování byl lineární model určen jako významný a korektní s normálním rozdělením reziduí. Jako významný byl určen pouze parametr regresní přímky (titrační způsob), úsek byl určen jako nevýznamný.

Rovnice regresní přímky

$$\text{“Minipal“} = 0,9404 (\pm 0,0456) * \text{“titračně“}$$

$$Y = 0,9404 (\pm 0,0456) * X$$

Podstatou validace nové metody je lineární model $y = \beta_0 + \beta_1 x$ s nulovým úsekem $\beta_0 = 0$ a jednotkovou směrnici $\beta_1 = 1$.

Vztah pro interval spolehlivosti:

$$\beta_0 - t_{1-\alpha/2} \sqrt{D(\beta_0)} \leq \beta_0 \leq \beta_0 + t_{1-\alpha/2} \sqrt{D(\beta_0)}$$

Proměnná	Odhad	Směr.Odch.	Spodní mez	Horní mez
β_0	9,9916	8,1029	-8,6938	28,6772
β_1	0,9404	0,0456	0,8351	1,0456

Interval spolehlivosti pro úsek (β_0) obsahuje nulu, úsek tedy nelze považovat za významně odchylený od nuly. Interval spolehlivosti pro směrnici (β_1) obsahuje jedničku, směrnici tedy nelze považovat za významně odchylenou od jedničky. Z toho plyne, že obě metody dávají s 95% pravděpodobností shodné výsledky a lze tedy provést nahrazení klasické analýzy metodou instrumentální.

3. VLASTNOSTI VYSOCE POVRCHOVĚ UPRAVENÉHO RUTILOVÉHO DRUHU TITANOVÉ BĚLOBY (vícerozměrný lineární regresní model)

ZADÁNÍ:

Precheza Přerov má ve svém výrobním programu vysoce upravený rutilový typ RG18. Vyznačuje se vysokou anorganickou úpravou solemi hliníku a křemíku a úpravou organickými aditivy.

Problémem u těchto vysoce povrchově upravených materiálů je jejich vysoká vlhkost (obsah těkavých látek), která způsobuje aplikační potíže u některých odběratelů. Obsah těkavých látek je vedle dalších parametrů stanovován při výstupní kontrole produktu, některé z těchto parametrů byly použity pro nalezení modelu.

Metodou vícerozměrného lineárního regresního modelu má být nalezen vztah mezi obsahem těkavých látek a dalšími parametry.

DATA

Tabulka 4.1: Přehled kvalitativních atributů šarží RG18

Těkavé látky %	Měrná vodivost μS/cm	Spotřeba oleje g/100 g	Obsah Al ₂ O ₃ %	Obsah SiO ₂ %	Obsah C %	Obsah částic do 1 μm %	Měrný povrch m ² /g
Y	X1	X2	X3	X4	X5	X6	X7
0,6	120	23,6	3,662	2,027	0,306	98,2	22,9
0,5	130	23,7	3,710	2,017	0,275	99,4	24,2
0,6	120	24,8	3,813	2,018	0,289	98,4	23,6
0,6	110	23,5	3,699	1,993	0,256	98,1	24,7
0,6	120	23,6	3,606	1,956	0,302	98,6	24,0
0,6	120	24,5	3,803	2,018	0,239	98,6	25,3
0,5	120	24,3	3,666	1,961	0,204	98,5	24,7
0,6	120	23,9	3,599	1,981	0,279	98,6	24,1
0,6	120	23,6	3,505	1,914	0,288	98,3	22,6
0,5	120	23,7	3,567	1,940	0,331	97,5	23,8
0,5	120	23,7	3,567	1,940	0,331	97,5	23,8
0,5	120	23,6	3,605	1,982	0,218	97,0	25,0
0,5	110	23,6	3,632	2,031	0,212	98,8	24,7
0,5	130	23,4	3,647	2,105	0,248	98,2	23,7
0,6	130	23,2	3,600	2,027	0,260	99,4	22,2
0,5	130	23,1	3,374	1,544	0,281	98,1	23,1
0,5	120	24,3	3,700	1,901	0,212	98,2	24,4
0,6	130	24,1	3,732	1,903	0,199	99,2	25,3
0,7	110	23,3	3,596	2,036	0,230	98,3	23,6
0,7	90	23,8	3,676	2,037	0,213	98,9	23,2
0,8	100	23,4	3,594	1,853	0,235	97,7	23,1
0,3	100	23,5	3,636	2,104	0,215	99,6	24,1
0,3	100	23,6	3,475	2,054	0,234	98,4	22,3
0,6	110	23,6	3,644	2,103	0,194	98,9	22,9
0,5	130	23,7	3,618	2,114	0,215	99,7	22,5
0,5	120	23,5	3,607	2,064	0,260	99,1	23,6
0,7	120	23,6	3,519	2,138	0,210	97,9	22,1
0,6	120	23,6	3,460	2,124	0,156	98,4	22,6
0,5	120	23,7	3,520	2,133	0,201	97,7	22,3
0,5	120	23,8	3,616	2,133	0,236	97,5	23,5
0,5	170	23,0	3,537	2,046	0,338	96,3	23,2
0,5	140	23,6	3,420	2,053	0,281	97,7	22,2
0,5	130	23,4	3,450	2,065	0,277	98,4	23,3
0,6	130	23,3	3,503	2,104	0,280	98,4	23,0
0,5	130	22,7	3,629	2,188	0,231	97,0	23,7

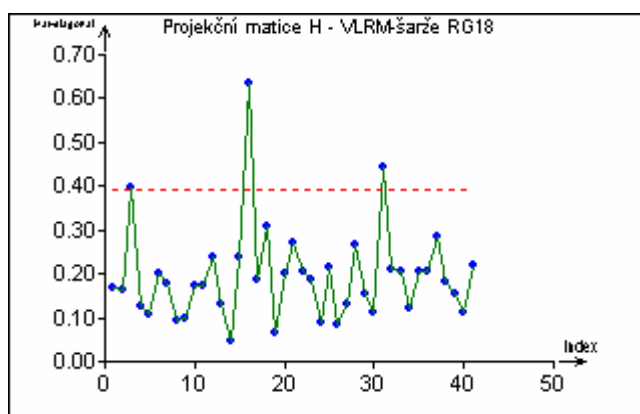
Těkavé látky %	Měrná vodivost $\mu\text{S/cm}$	Spotřeba oleje g/100 g	Obsah Al_2O_3 %	Obsah SiO_2 %	Obsah C %	Obsah částic do 1 μm %	Měrný povrch m^2/g
Y	X1	X2	X3	X4	X5	X6	X7
0,5	130	22,7	3,629	2,188	0,231	97,0	23,7
0,6	130	22,6	3,705	2,248	0,214	98,8	24,3
0,6	130	23,0	3,724	2,247	0,238	98,1	24,5
0,7	130	23,4	3,583	2,155	0,229	98,2	21,8
0,7	120	22,9	3,493	2,078	0,223	98,5	22,2
0,7	110	23,3	3,579	2,069	0,166	98,4	22,0

Program: QC Expert 2.5: modul Lineární regrese

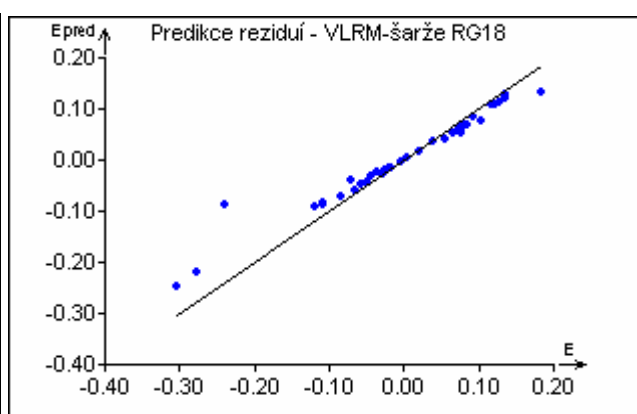
Název souboru s daty: data_vicerozmer.txt

VEHODNOCENÍ

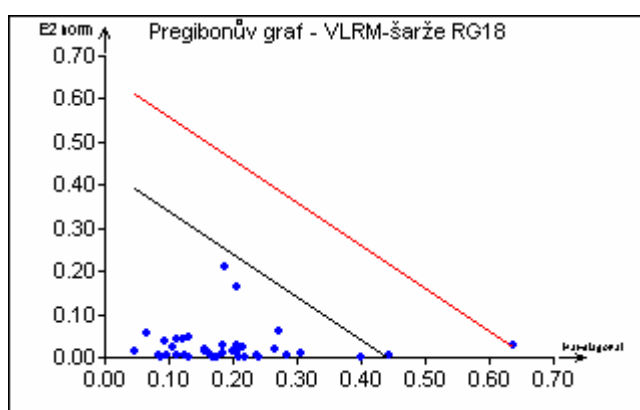
Lineární regresní model



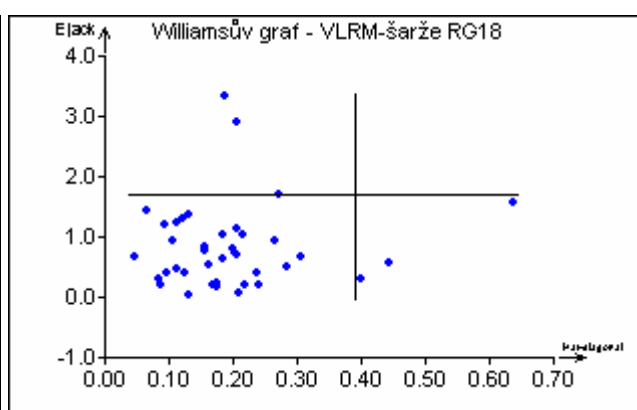
Obrázek 4.1: Body nad vodorovnou přímkou, tj. 16, 31 (popř. 3) mohou být považovány za silně vlivné.



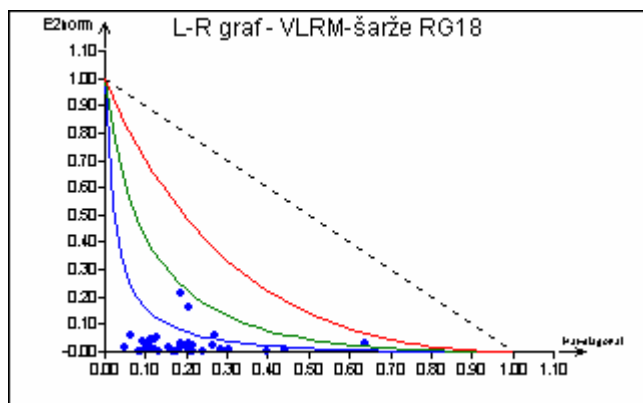
Obrázek 4.2: Výraznější odchylka od přímky indikuje vybočující hodnotu. V tomto případě se jedná o body 18, 22, 23.



Obrázek 4.3: Na tomto grafu je indikován jeden silně vlivný (16) a jeden vlivný bod (31).



Obrázek 4.4: Na tomto grafu jsou indikovány dva extrémní (16, 31) a dva vybočující body (22 a 23).



Obrázek 4.5: Indikovány tři podezřelé body (16, 22 a 23).

Na základě grafických diagnostik, numerické indikace vlivných bodů a za předpokladu dostatečného počtu dat bylo rozhodnuto o vypuštění bodu 16 (silně vybočující) a bodů 22, 23 (silně vlivné).

Regresní diagnostika

Hladina významnosti :	0,05
Kvantil $t(1-\alpha/2, n-m)$:	2,0422724563007
Kvantil $F(1-\alpha, m, n-m)$:	2,33434396484495
Absolutní člen :	Ano
Počet platných řádků :	38
Počet parametrů :	8
Metoda :	Nejmenší čtverce
Transformace :	Bez transformace

Základní analýza

Charakteristiky proměnných

Proměnná	Průměr	Směr.Odch.	Kor.vs.Y	Významnost
Měrná vodivost	122,3684211	12,3975173	-0,4751453535	0,0025739449
Spotřeba oleje	23,55263158	0,4665786004	-0,09371242059	0,5757396423
Obsah Al ₂ O ₃	3,610921053	0,09057813756	0,004472856131	0,9787376403
Obsah SiO ₂	2,049736842	0,09418001286	-0,06803650457	0,6848412235
Obsah C	0,2449210526	0,0439633641	-0,2667355072	0,1054853244
Obsah částic do 1 μm	98,24736842	0,7295821989	0,1954875121	0,2395146628
Měrný povrch	23,48157895	0,9655848985	-0,3368343787	0,0386507051

Indikace multikolinearity

Proměnná	Vlas. čísla kor. m.	Podmíněnost kappa	VI faktor	Vicenás. kor.
Abs	0,1883443676	1	1	0
Měrná vod.	0,4953993319	2,630284825	1,583741017	0,6071108897
Spotř. oleje	0,7606893721	4,038821982	1,857295778	0,6793988489
Obsah Al ₂ O ₃	1	5,30942344	2,397558851	0,7634848337
Obsah SiO ₂	1,213232903	6,441567211	2,261167611	0,7468270272
Obsah C	1,582094662	8,400010481	1,607880466	0,6148684573
Část. do 1 μm	0,3804066306	2,019739881	1,304267933	0,4829972788
Měrný povrch	2,379832733	12,6355397	2,135963435	0,7292648295

Analýza rozptylu

Průměr Y :	0,5763157895		
Zdroj	Součet čtverců	Průměrný čtverec	Rozptyl
Celková variabilita	0,2486842105	0,00654432133	0,0067211948
Variabilita vysvětlená modelem	0,1230844087	,003239063387	0,00332
Reziduální variabilita	0,1255998018	0,00330525794	0,003394
Hodnota kritéria F :	4,199884085		
Kvantil F (1-alfa, m-1, n-m) :	2,334343965		
Pravděpodobnost :	0,002476579037		
Závěr :	Model je významný		

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	P	Spodní mez	Horní mez
Abs	2,9962	1,7521	Nevýznamný	0,0975	-0,5821	6,5746
Měrná vod.	-0,0024	0,0010797	Významný	0,0326	-0,0046-	0,000213
Spotř. oleje	-0,0693	0,0310	Významný	0,0332	-0,1327	-0,0058
Obsah Al ₂ O ₃	0,4084	0,1818	Významný	0,0322	0,0370	0,7798
Obsah SiO ₂	-0,3306	0,1698	Nevýznamný	0,0609	-0,6775	0,0162
Obsah C	-0,2968	0,3068	Nevýznamný	0,3409	-0,9234	0,32972
Část. do 1 μm	0,0004	0,0166	Nevýznamný	0,9780	-0,0335	0,0344
Měrný povrch	-0,0536	0,0161	Významný	0,0022	-0,0865	-0,0208

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,7035215675
Koeficient determinace R ² :	0,4949425959
Predikovaný korelační koeficient R _p :	0,02674749076
Střední kvadratická chyba predikce MEP :	0,005474019819
Akaikeho informační kritérium :	-201,065149

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	4,199884085
Kvantil F (1-alfa, m-1, n-m) :	2,334343965
Pravděpodobnost :	0,002476579037
Závěr :	Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC :	-0,0699014629
Závěr :	Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW :	0,003737943846
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,9512487277
Závěr :	Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB :	2,934745465
Kvantil Chi ² (1-alfa,2) :	5,991464547
Pravděpodobnost :	0,2305303552
Závěr :	Rezidua mají normální rozdělení.

Waldův test autokorelace

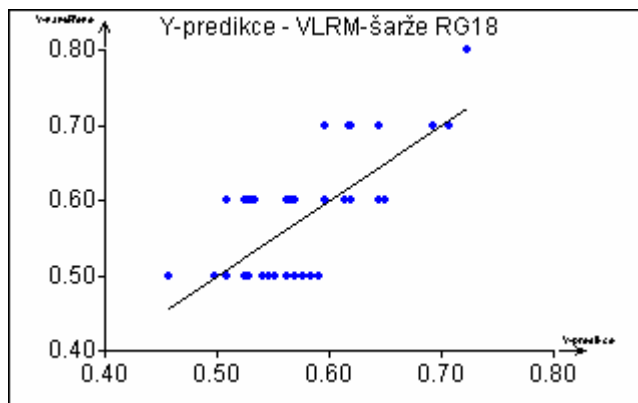
Hodnota kritéria WA : 1,14853448
 Kvantil $\chi^2(1-\alpha,1)$: 3,841458829
 Pravděpodobnost : 0,9512487277
 Závěr : Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

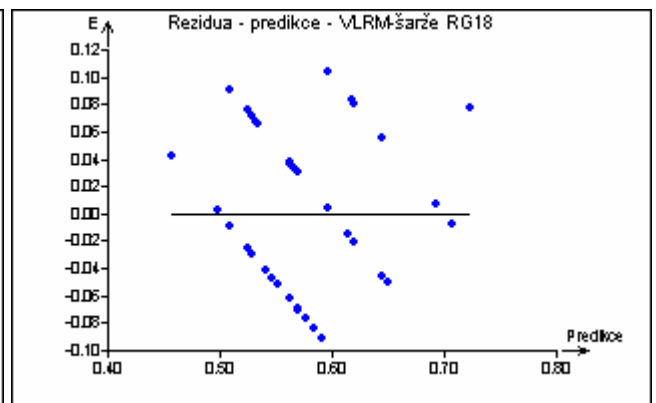
Hodnota kritéria DW : 1,573811782
 Kritické hodnoty DW 1,21 1,79
 Závěr : Pozitivní autokorelace reziduí není prokázána.

Znaménkový test reziduí

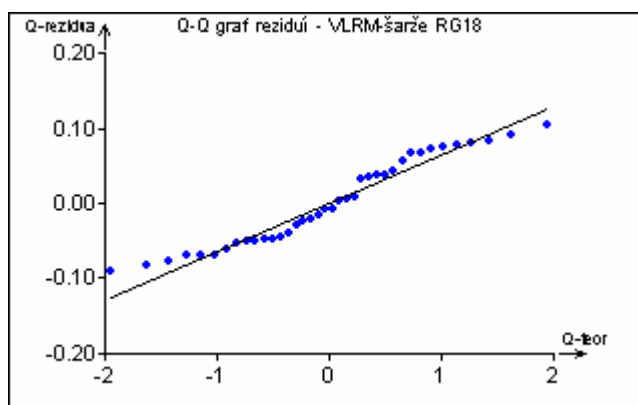
Hodnota kritéria Sg : 1,46699668
 Kvantil $N(1-\alpha/2)$: 1,959963999
 Pravděpodobnost : 0,1423769584
 Závěr : V reziduích není trend.

Grafická analýza reziduí

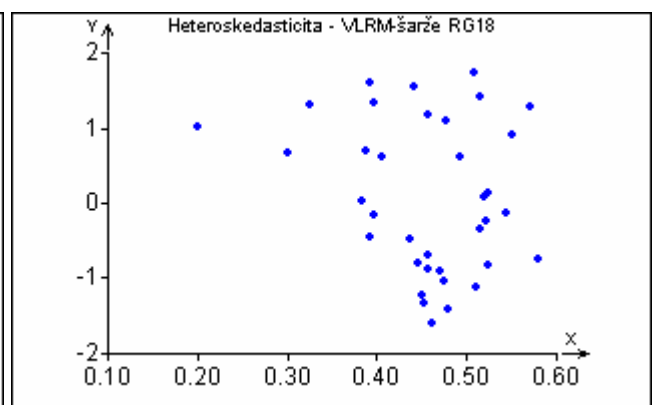
Obrázek 4.6: Graf vyjadřuje těsnost proložení.



Obrázek 4.7: Průměr reziduí (vodorovná přímka) je roven nule.



Obrázek 4.8: Q-Q graf slouží pro posouzení normality reziduí. Přímka odpovídá normálnímu rozdělení reziduí. Vyplývá z něj, že rezidua mají přibližně normální rozdělení.



Obrázek 4.9: S výjimkou jednoho bodu jsou ostatní rezidua v mraku, svědčí to o konstantním rozptylu.

SOUHRN A ZÁVĚR

Na základě výsledků testování byl lineární model určen jako významný a korektní, s normálním rozdělením dat. Nebyla zaznamenána multikolinearita.

Z regresní diagnostiky vyplynulo, že závisle proměnná Y (těkavé látky) závisí významně na nezávislé proměnných X1 (měrná vodivost), X2 (spotřeba oleje), X3 (obsah TiO₂) a X7 (měrný povrch). Z tohoto lze odvodit rovnici závislosti obsahu těkavých látek na uvedených proměnných X1, X2, X3 a X7. Ostatní parametry byly určeny jako nevýznamné, z čehož plyne, že jejich hodnota významně neovlivní výslednou vlhkost materiálu RG18.

Rovnice regresní závislosti (přímky)

“těkavé látky“ = 0,4084 (± 0,1818) * “obsah Al₂O₃“ - 0,0024 (± 0,0011) * “měrná vodivost“ - 0,0693 (± 0,0310) * “spotřeba oleje“ - 0,0536 (± 0,0161) * “měrný povrch“

$Y = 0,4084 (\pm 0,1818) * X3 - 0,0024 (\pm 0,0011) * X2 - 0,0693 (\pm 0,0310) * X2 - 0,0536 (\pm 0,0161) * X7$