

UNIVERZITA PARDUBICE  
FAKULTA CHEMICKO-TECHNOLOGICKÁ



SEMESTRÁLNÍ PRÁCE

**Tvorba lineárních regresních modelů při analýze dat**

## 1. Úvod

Tato semestrální práce je vypracováním otázek z předmětu *Tvorba lineárních regresních modelů při analýze dat*.

## 2. Zadané úkoly

**Úloha 1. Porovnání dvou regresních přímek u jednoduchého lineárního regresního modelu** (včetně testování úseku a směrnice, s vyšetřením vlivných bodů a jejich event. odstraněním, posouzením míry spolehlivosti navrženého modelu). Test shodnosti dvou (nebo i více) přímek, test jejich paralelity a společného úseku.

**Úloha 2. Určení stupně polynomu** metodou MNČ a RH křivkové závislosti (porovnání obou metod vede k odstranění multikolinearity, testování statistické významnosti nalezených parametrů, vyšetření regresního tripletu metodou regresní diagnostiky, zdůvodnění a výklad všech užitých diagnostik a statistik).

**Úloha 3. Validizace nové analytické metody** (vyšetřením regresního tripletu testujte a diskutujte statistickou významnost jednotlivých parametrů v modelu stejně jako i jejich fyzikální smysl, zdůvodnění a výklad všech užitých diagnostik a statistik).

**Úloha 4. Vícerozměrný lineární regresní model** o alespoň 4 proměnných (vyšetřením regresního tripletu nalezněte nejlepší model, využijte regresní diagnostiku a pomocí parciálních regresních a parciálních reziduálních grafů diskutujte významnost jednotlivých parametrů v modelu stejně jako i jejich fyzikální smysl).

## 3. Vypracované otázky

### 3.1. Porovnání dvou regresních přímek

#### 3.1.1. Zadání

Obsah tuhé fáze v suspenzi (sušina) se orientačně určuje stanovením měrné hmotnosti suspenze. Stanovení sušiny bylo provedeno dvěma různými postupy. Určete zda mezi měrnou hmotností a sušinou existuje stejná závislost pro oba způsoby stanovení sušiny.

#### Data:

$\rho_T$ [kg/m <sup>3</sup> ]	1240	1180	1140	1120	1080	1060	1040
sušina1 [%]	26,37	21,23	17,9	15,27	11,12	8,43	6,98
sušina2 [%]	26,03	20,28	17,8	14,8	10,36	8,07	6,79

Software: QC-Expert 2.5

#### 3.1.2. Model. sušina1 vs. $\rho_T$

##### 3.1.2.1. Návrh modelu

Nejprve provedeme určení modelu sušina1 (závisle proměnná) vs.  $\rho_T$  (nezávisle proměnná) a provedeme kritiku dat a ověření předpokladů metody nejmenších čtverců.

**Tabulka 1: Parametry modelu sušina1 vs.  $\rho_T$** 

Parametr	Odhad	Směrodatná Odchylka	t-kriterium	Pravděpodobnost
úsek	-96,445	3,737	-25,8032	1,63E-06
směrnice	0,099544	0,00332	29,95495	7,78E-07

**Tabulka 2: Statistické charakteristiky regrese**

Vícenásobný korelační koeficient, R	0,99723
Koeficient determinace, $R^2$	0,99446
Predikovaný korelační koeficient, $R_p^2$	0,99257
Střední kvadratická chyba predikce, MEP	0,63116
Akaikeho informační kritérium, AIC	-6,1019

Jak hodnota úseku tak i hodnota směrnice vychází statisticky významně odlišné od nuly (hodnota kritického kvantilu Studentova rozdělení ( $t_{krit. [0,5;5]}=2,57$ ) je menší než hodnota testovacího kritéria-primární hypotéza o  $\beta_i=0$  je zamítnuta). Hodnota  $R^2$  je 0,997, což znamená, že cca 99,4% bodu leží na přímce určené metodou nejmenších čtverců.

### 3.1.2.2. Kritika dat

V hodnotách Jackknife reziduí (Tabulka 3) není indikována přítomnost odlehlých bodů. Z hodnot Cookovi a Atkinsonovi vzdálenosti je patrná přítomnost podezřelých bodů 1, 3 a 6.

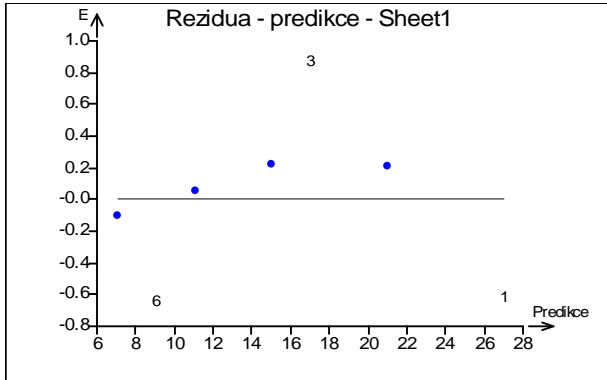
**Tabulka 3: Analýza reziduí**

Bod	Standardizované reziduum	Jackknife reziduum	Predikované reziduum	Diagonální prvky
i	$eS[i]$	$eJ[i]$	$eP[i]$	$H[i,i]$
1	-1,7057	-2,3592	-1,553	6,0115E-01*
2	0,42868	0,39067	0,285	0,25191
3	1,6341	2,1412	1,0208	0,15267
4	0,42428	0,38651	0,26356	0,14313
5	0,11228	0,10055	0,072374	0,2042
6	-1,3101	-1,4459	-0,88463	0,27481
7	-0,22089	-0,19854	-0,1603	0,37214
Bod	Zobecněné diag. prvky	Cookova vzdálenost	Atkinsonova vzdálenost	Vliv na predikci
i	$H_m[i,i]$	$D[i]$	$A[i]$	$DF[i]$
1	0,83322	2,1924*	4,5795E*	-2,8963*
2	0,2794	0,03094	0,35844	0,2267
3	0,60519	0,24057E*	1,4371	0,90889
4	0,17398	0,015034	0,24977	0,15797
5	0,2062	0,0016173	0,080534	0,050934
6	0,52374	0,32519E*	1,4074	-0,89009
7	0,37826	0,01446	0,24168	-0,15285

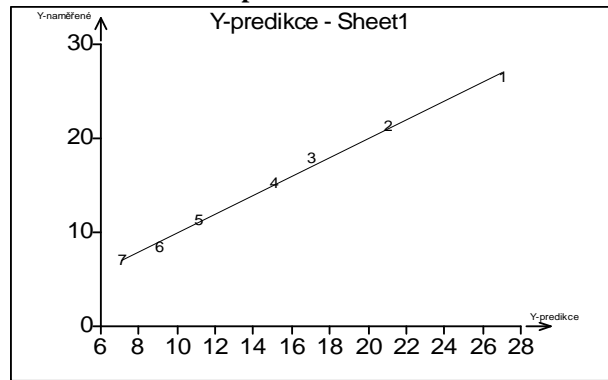
Z grafu predikovaných reziduí (obr.1) je patrná přítomnost odlehlých bodů 1 a 6 a odlehlého bodu č.3. Ve Williamsově grafu (obr.3) jsou body 1 a 3 silně vybočující.

V L-R grafu je bod 1 vyhodnocen jako podezřelý. Rovněž graf na obrázku č.7 detekuje bod č.1 jako silně vlivný.

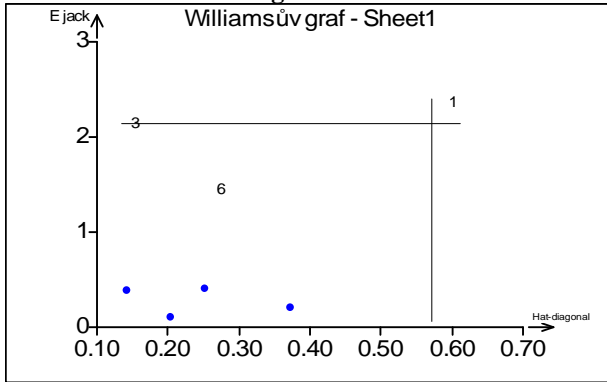
**Obrázek 1: Graf rezidua vs. Predikce**



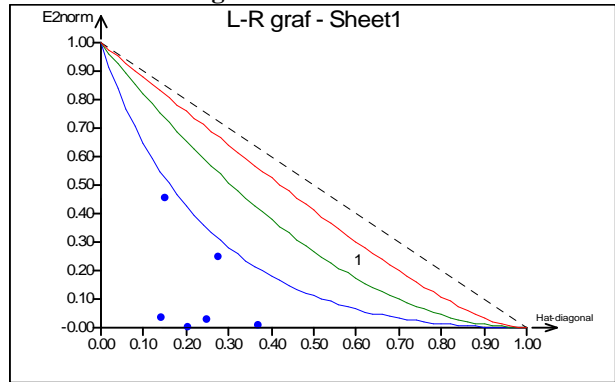
**Obrázek 2: Graf Y-predikce**



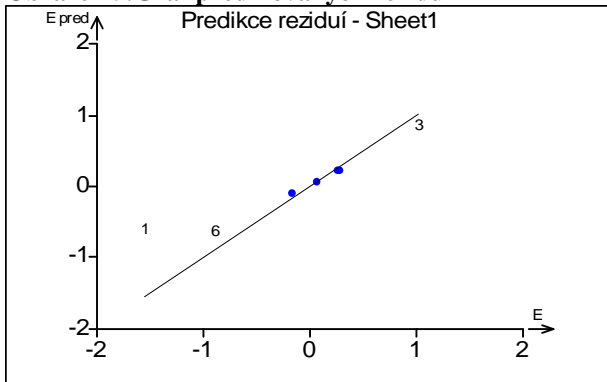
**Obrázek 3: Williamsův graf**



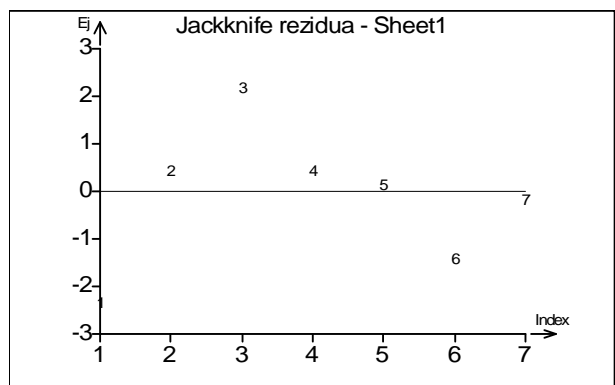
**Obrázek 4: L-R graf**



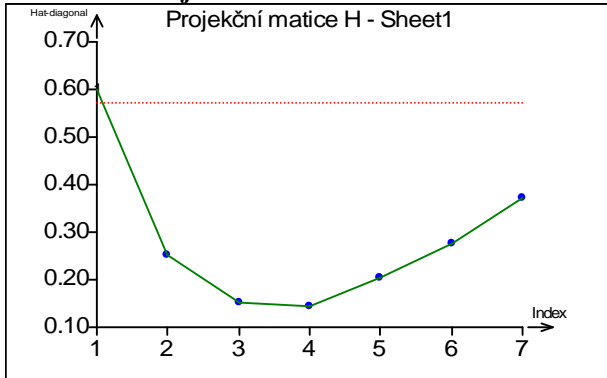
**Obrázek 5: Graf předikovaných reziduí**



**Obrázek 6: Jack-Knife rezidua**



**Obrázek 7: Projekční matice**



Z provedených grafických diagnostik lze usoudit, že body 1 a 3 jsou vlivné, ale k jejich vypuštění není dostatek argumentů.

### 3.1.2.3. Ověření předpokladů metody nejmenších čtverců

Použití metody nejmenších čtverců je podmíněno splněním určitých kritérií, které prověřují níže uvedená statistické testy .

#### **Fisher-Snedecorův test významnosti modelu**

Hodnota kritéria F : 897,2990568

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

Pravděpodobnost : 7,776587037E-007

**Závěr : Model je významný**

#### **Scottovo kritérium multikolinearity**

Hodnota kritéria SC : 0,06894399561

**Závěr : Model je korektní.**

#### **Cook-Weisbergův test heteroskedasticity**

Hodnota kritéria CW : 0,1700989488

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

Pravděpodobnost : 0,6800238837

**Závěr : Rezidua vykazují homoskedasticitu.**

#### **Jarque-Berrův test normality**

Hodnota kritéria JB : 0,210946455

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

Pravděpodobnost : 0,8998985651

**Závěr : Rezidua mají normální rozdělení.**

#### **Waldův test autokorelace**

Hodnota kritéria WA : 0,1510229186

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

Pravděpodobnost : 0,6975597311

**Závěr : Autokorelace je nevýznamná**

#### **Znaménkový test reziduí**

Hodnota kritéria Sg : 0,7882407814

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

Pravděpodobnost : 0,430555879

**Závěr : V reziduích není trend.**

### 3.1.2.4. Závěr analýzy modelu sušina1 vs. $\rho_T$

Předpoklady pro použití metody nejmenších čtverců jsou splněny. Výběr obsahuje podezřelé body, ale k jejich vypuštění není dostatek argumentů. Výsledný tvar zpřesněného modelu je  $y = -96,445(\pm 3,737) + 0,09954(\pm 0,0033) \cdot x$  . V závorce je uvedena směrodatná odchylka.

### 3.1.3. Model sušina2 vs. $\rho_T$

#### 3.1.3.1. Návrh modelu

**Tabulka 4: Parametry modelu sušina2 vs.  $\rho_T$**

Parametr	Odhad	Směrodatná odchylka	t-kriterium	Hypotéza $H_0$ je	Hlad. významnosti
úsek	-96,015	4,3458	-22,094	Zamítnuta	0.000
směrnice	0,098758	0,0038637	25,56	Zamítnuta	0.000

**Tabulka 5: Statistické charakteristiky regrese**

Vícenásobný korelační koeficient, R	0,9962
Koeficient determinace, $R^2$	0,9924
Predikovaný korelační koeficient, $R_p^2$	0,99287
Střední kvadratická chyba predikce, MEP	0,59748
Akaikeho informační kritérium, AIC	-3,9917

Jak úsek tak i směrnice vychází statisticky významně odlišné od nuly ( $H_0: \beta_i = 0$  zam.). Vysoká hodnota  $R^2$  svědčí o tom, že přímka určená metodou nejmenších čtverců dobře vystihuje závislost.

#### 3.1.3.2. Kritika dat

Jackknife rezidua označují bod č.3 jako odlehlý. Body 1 a 6 jsou vyhodnoceny jako podezřelé.

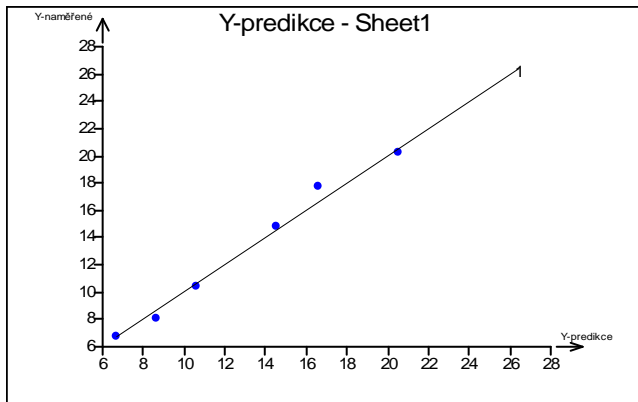
**Tabulka 6: Analýza reziduí**

Bod	Standardizované reziduum	Jackknife reziduum	Predikované reziduum	Diagonální prvky
i	eS[i]	eJ[i]	eP[i]	H[i,i]
1	-0,98159	-0,97714	-1,0391	0,60115*
2	-0,41332	-0,37616	-0,31949	0,25191
3	2,0007	4,0072*	1,4532	0,15267
4	0,33359	0,30174	0,24094	0,14313
5	-0,47491	-0,43468	-0,35592	0,2042
6	-1,0505	-1,0643	-0,82474	0,27481
7	0,18322	0,16443	0,15459	0,37214
Bod	Zobecněné diag. prvky	Cookova vzdálenost	Atkinsonova vzdálenost	Vliv na predikci
i	Hm[i,i]	D[i]	A[i]	DF[i]
1	0,67801	0,72609*	1,8967*	-1,1996*
2	0,27747	0,028762	0,34514	-0,21828
3	0,83102	3,6062*	2,6894*	1,7009*
4	0,1622	0,0092939	0,19499	0,12332
5	0,24009	0,028936	0,34815	-0,22019
6	0,43486	2,0909*	1,036	-0,6552
7	0,37635	0,0099479	0,20015	0,12659

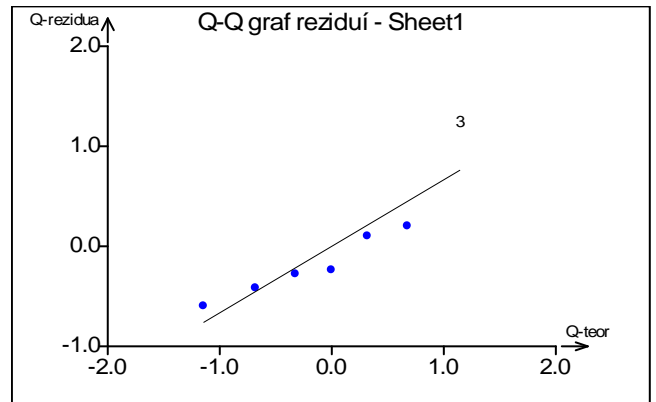
Z grafu předikovaných reziduí je patrná přítomnost odlehlého bodu č.1. Z Williamsova grafu je patrná přítomnost odlehlého bodu č.3. V Pregibonově grafu leží bod č.3 na hranici vlivné oblasti. Z obr. 13 je vidět že bod č.1 má silný vliv na regresi. V L-R grafu jsou

vyhodnoceny body 1 a 3 jako slabě vlivné. Z Q-Q grafu na obr. 9 je patrné, že bod č.3 narušuje mírně normalitu reziduí.

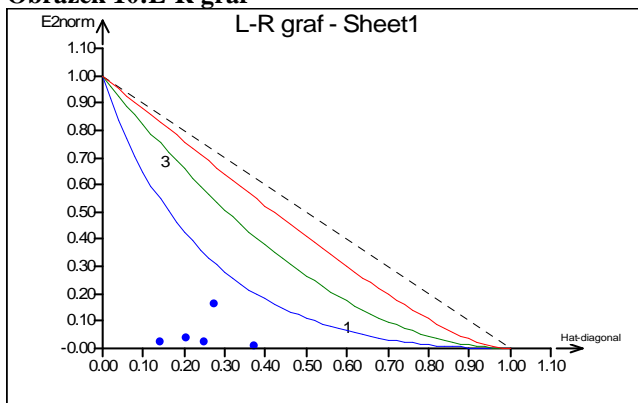
**Obrázek 8: Graf Y-predikce**



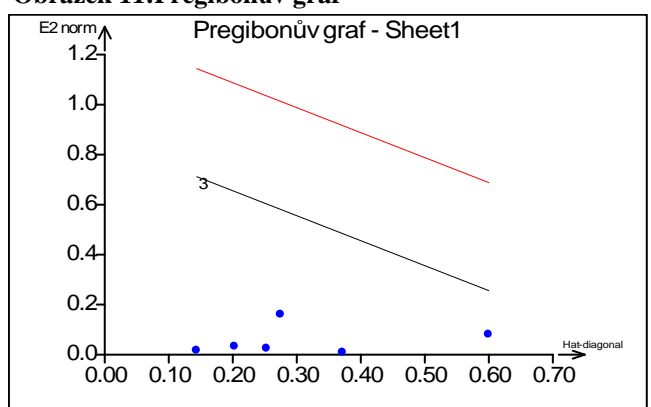
**Obrázek 9: Q-Q graf reziduí**



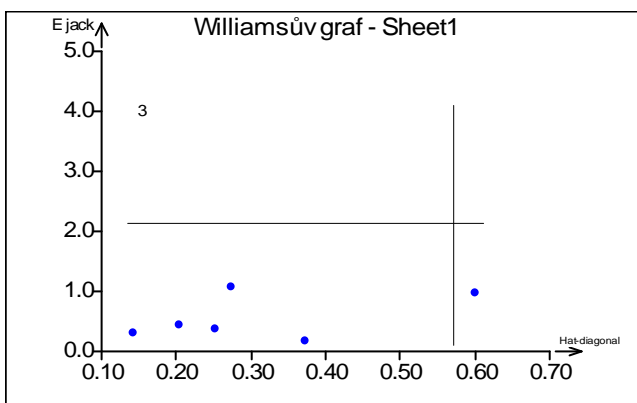
**Obrázek 10: L-R graf**



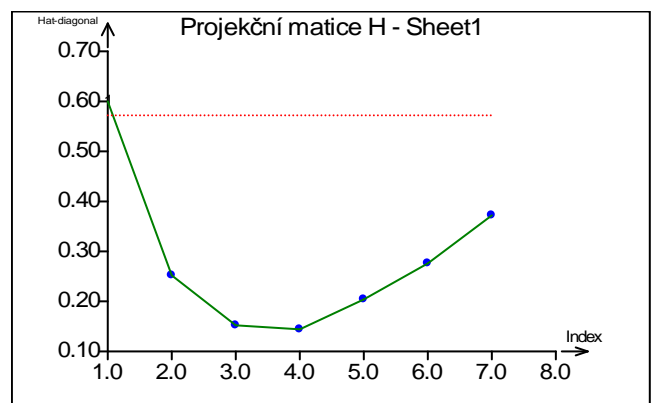
**Obrázek 11: Pregibonův graf**



**Obrázek 12: Williamsův graf**



**Obrázek 13: Graf projekční matice**



Body č. 1 a 3 jsou zjevně vybočující, ale k jejich vypuštění není dostatek argumentů

### 3.1.3.3. Ověření předpokladů metody nejmenších čtverců

#### **Fisher-Snedecorův test významnosti modelu**

Hodnota kritéria F : 653,3263582

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

Pravděpodobnost : 1,711521725E-006

**Závěr : Model je významný**

#### **Scottovo kritérium multikolinearity**

Hodnota kritéria SC : 0,06747350455

**Závěr : Model je korektní.**

#### **Cook-Weisbergův test heteroskedasticity**

Hodnota kritéria CW : 0,04250454411

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

Pravděpodobnost : 0,8366610094

**Závěr : Rezidua vykazují homoskedasticitu.**

#### **Jarque-Berrův test normality**

Hodnota kritéria JB : 1,829900281

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

Pravděpodobnost : 0,4005365962

**Závěr : Rezidua mají normální rozdělení.**

#### **Waldův test autokorelace**

Hodnota kritéria WA : 0,01219625317

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

Pravděpodobnost : 0,9120630901

**Závěr : Autokorelace je nevýznamná**

#### **Znaménkový test reziduí**

Hodnota kritéria Sg : 0,06063390626

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

Pravděpodobnost : 0,9516507699

**Závěr : V reziduích není trend.**

Z výsledků výše uvedených testů plyne, že jsou splněny všechny předpoklady užití metody nejmenších čtverců.

### 3.1.3.4. Závěr analýzy modelu sušina<sub>2</sub> vs. $\rho_T$

Jsou splněny podmínky použití metody nejmenších čtverců. Výběr obsahuje odlehlé body, ale k jejich vypuštění není dostatek argumentů. Konečný tvar modelu je tedy:  $y = -96,015(\pm 4,354) + 0,0987(\pm 0,00386) \cdot x$ . V závorce jsou uvedeny hodnoty směrodatné odchylky příslušného parametru.



### 3.1.4. Porovnání dvou lineárních závislostí

Nyní provedeme vlastní porovnání zkoumaných přímek. V první fázi provedeme test shody rozptylů. A to pomocí klasického F-testu. Testovací kritérium má následující tvar:

$$F = \max\left(\frac{\sigma_1^2}{\sigma_2^2}\right), \text{ kde } \sigma_1^2, \text{ resp. } \sigma_2^2 \text{ jsou hodnoty reziduálního rozptylu obou výběrů pro první}$$

testovaný výběr a druhý testovaný výběr. Hodnotu testovací statistiky porovnáváme s hodnotou F rozdělení  $F(0,05;6;5)$ .

$$F_{\text{test}} = \frac{0,447}{0,3306} = 1,3521 \quad F(0,05;7;7) = 3,787$$

Protože je hodnota F testu menší než je hodnota kritického kvantilu F rozdělení, je podmínka shodnosti rozptylů splněna.

Nyní můžeme provést *Chowův test shodnosti přímek*. V tomto případě porovnáváme hodnotu testovacího kritéria  $F_{\text{CH}}$  s hodnotou kvantilu F rozdělení  $F(0,05;2;7)$ . V případě, že je hodnota testovacího kritéria  $F_{\text{CH}}$  menší než hodnota kritického kvantilu F rozdělení, hypotézu o shodnosti přímek na hladině významnosti 0,05 přijímáme.

$$F_{\text{CH}} = \frac{(RSC - RSC_1 - RSC_2)(n - 2m)}{(RSC_1 + RSC_2)m} = 0,748, \quad F(0,05;2;5) = 5,7861$$

RSC1 - je reziduální součet čtverců odchylek prvního výběru

RSC2 - je reziduální součet čtverců odchylek druhého výběru

n - je celkový počet pro obě přímky

m - je počet parametrů regresního modelu, v tomto případě  $m=2$

Hodnota testovacího kritéria je menší než hodnota kritického kvantilu F rozdělení, takže lze akceptovat nulovou hypotézu a obě lineární závislosti považovat za shodné.

## 3.2. Validace nové analytické metody

### 3.2.1. Zadání

Bylo provedeno stanovení sušiny (obsahu tuhé fáze v hm.%) v suspenzi pigmentového oxidu titaničitého dvěma různými způsoby. Prověřte pomocí regrese, zda oba způsoby stanovení poskytují shodné výsledky.

**Software:** Qc-Expert 2.15

**Data:**

sušina1 [%]	26,37	21,23	20,84	17,9	15,27	11,12	8,43	6,98
sušina2 [%]	26,03	20,28	20,99	17,8	14,8	10,36	8,07	6,79

### 3.2.2. Návrh modelu

Pomocí metody nejmenších čtverců určíme parametry modelu  $y = \beta \cdot x + \beta_0$ , kde  $y$  je sušina2 a  $x$  je sušina1. V ideálním případě bychom měli získat přímku s nulovým úsekem a jednotkovou směrnici.

V tabulce č.7 je proveden odhad parametrů modelu a je testována hypotéza  $H_0$  o statistické nevýznamnosti parametrů regrese ( $\beta=0$ ). Hodnota kritického kvantilu Studentova

rozdělení  $t(0,05; 6)$  je 2,447 a vzhledem k tomu, že hodnota testovacího kritéria je v případě úseku menší než je hodnota kritického kvantilu můžeme primární hypotézu akceptovat (úsek se statisticky významně neodlišuje od nuly).

**Tabulka 7: parametry modelu**

Parametr	Odhad	Směrodatná odchylka	t-kriterium	Hypotéza $H_0$ je	Hlad. významnosti
$\beta_0$	-0,40636	0,36493	-1,1135	Akceptována	0.308
$\beta$	1,0018	0,021169	47,323	Zamítnuta	0.000

**Tabulka 8: Statistické charakteristiky regrese**

Vícenásobný korelační koeficient, R	0,99866
Koeficient determinace, $R^2$	0,99733
Predikovaný korelační koeficient, $R_p^{\wedge 2}$	0,99786
Střední kvadratická chyba predikce, MEP	0,17458
Akaikeho informační kritérium, AIC	-13,716

Hodnoty uvedené v Tabulce č.8 (vysoká hodnota  $R^2$ , nízká hodnota střední kvadratické chyby predikce a Akaikeho kritéria) svědčí o dobré shodě obou metod stanovení sušiny. Hodnota  $R^2$  je 0,99733, což znamená, že cca 99,73% bodů leží na přímce určené metodou nejmenších čtverců.

### 3.2.3. Kritika dat

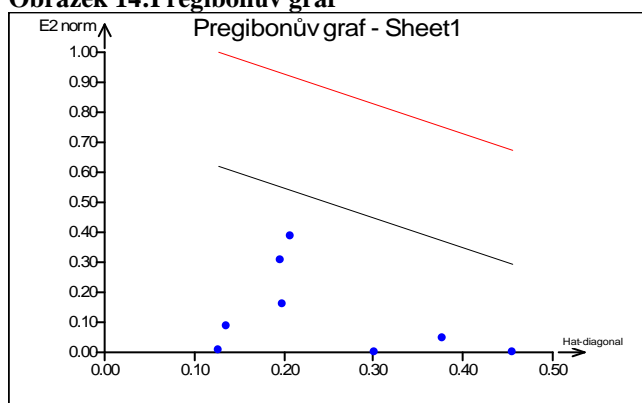
V další fázi je nutné posoudit, zda data neobsahují odlehlé hodnoty, které by mohli vést ke zkreslení modelu při návrhu regresního vztahu.

**Tabulka 9: Reziduální odchylky modelu**

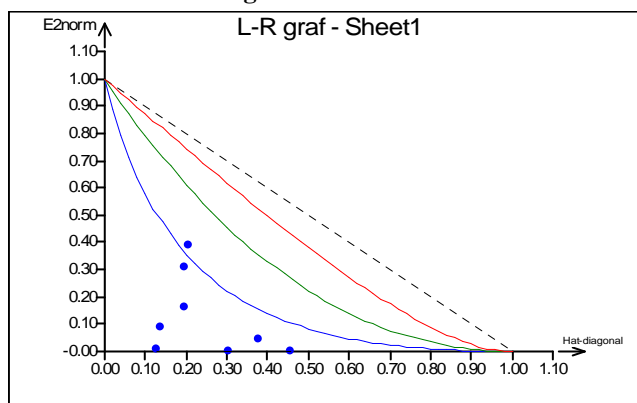
Bod	Standardizované reziduum	Jackknife reziduum	Predikované reziduum	Diagonální prvky
i	eS[i]	eJ[i]	eP[i]	H[i,i]
1	0,066892	0,061087	0,034571	0,45483
2	-1,7141	-2,1905	-0,73528	0,20862
3	1,5168	1,7634	0,64575	0,19657
4	0,77275	0,74338	0,31722	0,13591
5	-0,25562	-0,23463	-0,10438	0,12672
6	-1,094	-1,1162	-0,4664	0,19882
7	0,097783	0,089335	0,044668	0,30217
8	0,67624	0,64228	0,32677	0,37636

V Tabulce č.9 není žádná hodnota označena hvězdičkou, což znamená že Jackknife rezidua neindikují přítomnost odlehlých hodnota není indikována ani přítomnost podezřelých bodů. To potvrzují i grafy na obrázcích 14 až 16. Výjimku představuje pouze Williamsův graf uvedený na obr.č.17, který indikuje bod č.2 jako odlehlý.

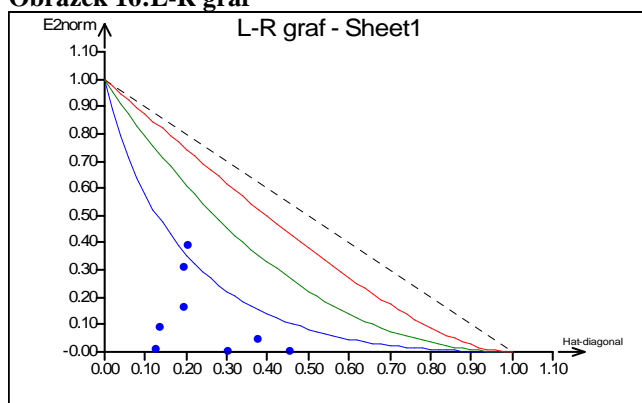
Obrázek 14: Pregibonův graf



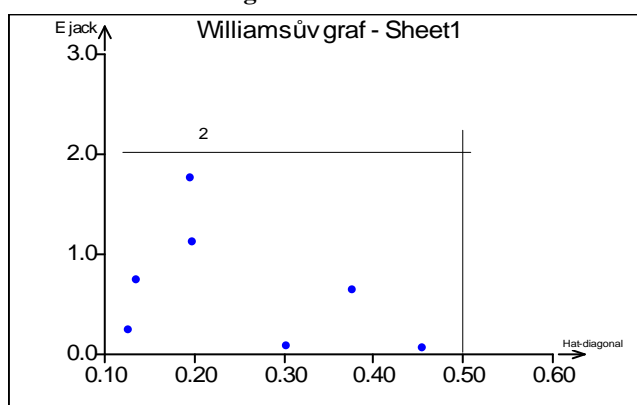
Obrázek 15: Williamsův graf



Obrázek 16: L-R graf



Obrázek 17: Williamsův graf



### 3.2.4. Ověření předpokladů metody nejmenších čtverců

Použití metody nejmenších čtverců je podmíněno splněním určitých kriterií. Mezi ně patří normalita reziduí, homoskedasticita a v reziduích nesmí být trend. Důležitým hodnotícím prvkem je i Scottovo kriterium multikolinearity. To nepatří mezi základní předpoklady použití metody nejmenších čtverců, ale vysoké hodnoty multikolinearity indikují přeuročenost modelu a způsobují nárůst hodnot směrodatných odchylek parametrů modelu.

**Fisher-Snedocorův test významnosti regrese, F** 2239,5

Tabulkový kvantil,  $F(1-\alpha, m-1, n-m)$  5,9874

**Závěr: Navržený model je přijat jako významný.**

Spočtená hladina významnosti 0

**Scottovo kriterium multikolinearity, M** 8,6299E-15

**Závěr: Navržený model je korektní.**

**Cook-Weisbergův test heteroskedasticity, Sf** 3,6479

Tabulkový kvantil,  $\chi^2(1-\alpha, 1)$  3,8415

**Závěr: Rezidua vykazují homoskedasticitu.**

Spočtená hladina významnosti 0,056

**Jarque-Berraův test normality reziduí, L(e)** 0,29646

Tabulkový kvantil,  $\chi^2(1-\alpha, 2)$  5,9915

**Závěr: Normalita je přijata.**

Spočtená hladina významnosti	0,862
------------------------------	-------

**Waldův test autokorelace, Wa**

0,33515

Tabulkový kvantil,  $\text{Chi}^2(1 - \alpha, 1)$ 

3,8415

**Závěr: Rezidua nejsou autokorelována.**

Spočtená hladina významnosti

0,563

Z výsledků výše uvedených testů je patrné, že jsou splněny všechny předpoklady použití metody nejmenších čtverců.

**3.2.5. Závěry provedené validace**

Navržený model je korektní a rovněž je korektní použití metody nejmenších čtverců. V případě úseku je 95% interval spolehlivosti úseku (-1,299;0,4866) a interval spolehlivosti směrnice je (0,95;1,05). Interval spolehlivosti úseku obsahuje nulu a tedy úsek je statisticky nevýznamný a interval spolehlivosti směrnice obsahuje jedničku a tedy se statisticky významně neodlišuje od jedničky. Obě použité metody vedou ke stejným výsledkům.

**3.3. Vícerozměrný lineární regresní model****3.3.1. Zadání**

Z oxidu železitého vyráběného v PRECHEZE je podíl nad 0,045 mm vytříděn pomocí hydroseparace. Je to z toho důvodu, že hrubé podíly oxidu železitého jsou nevhodné pro použití v průmyslu nátěrových hmot. Z důvodu snížení ztrát je snaha vést režim kalcinace takovým způsobem, aby množství vznikajících hrubých podílů bylo co nejmenší. Byl vysloven předpoklad, že na produkci hrubých podílů mají vliv: granulometrie vstupního materiálu (monohydrátu, vyjádřeno jako poměr hmotnosti vytříděného monohydrátu ke hmotnosti vyrobeného pigmentu-MS30/ŽČ), chemické složení ilmenitu (za významný je považován vliv Mn a Mg) a požadovaný odstín produktu (vyjádřený pomocí sytosti C\*). Provedte regresní diagnostiku a diskutujte význam a fyzikální smysl jednotlivých parametrů.

**Software:** Qc-Expert 2.15**Data:**

Měsíc	> 0,045 mm (kg Fe/t ŽČ)	C* TP303	MS30/ŽČ	MgO (hm. %)	MnO (hm.%)
leden	33,9	17,92	0,82	0,465	0,550
únor	22,86	18,22	0,72	0,430	0,550
březen	10,79	18,37	0,87	0,46	0,520
květen	67,87	17,74	0,69	0,335	0,548
červen	65,61	17,69	0,75	0,381	0,542
červenec	44,83	17,57	0,74	0,411	0,617
srpen	60,57	17,74	0,72	0,431	0,599
říjen	16,2	18,2	0,85	0,416	0,562
listopad	22,37	17,78	0,87	0,426	0,572
Prosinec	21,5	18,02	0,96	0,467	0,565

**Poznámka:** Monohydrát je před vstupem do kalcinace tříděn – vytříděný podsítný podíl je využit jako přísada do krmných směsí MONOSAL 30 Čím je větší hodnota poměru MS30/ŽČ

(ŽČ vyrobený pigment), tím je větší množství vytríděného monohydrátu. Tento poměr charakterizuje granulometrii monohydrátu vstupujícího do kalcinace.

Navržený model má následující tvar:

$$y = c + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

<b>y</b>	měrný výskyt podílu nad 0,045 mm vztažený na jednotku hmotnosti pigmentu
<b>x<sub>1</sub></b>	sytost vyrobeného pigmentu (odmocnina sumy čtverců par. a*, b* v sys. CIE L*a*b*)
<b>x<sub>2</sub></b>	poměr MS30/ŽČ
<b>x<sub>3</sub></b>	obsah MgO v ilmenitu vyjádřený v hmotnostních procentech
<b>x<sub>4</sub></b>	obsah MnO v ilmenitu vyjádřený v hmotnostních procentech
<b>c</b>	konstanta

### 3.3.2. Předběžná analýza dat

Tabulka 10: Párové korelační koeficienty

Proměnná	Průměr	Směrodatná odchylka	Párový korelační koeficient	Spočtená hladina významnosti
y	36,65	21,527	1	
x <sub>1</sub>	17,92	0,2667	-0,7830	0,007
x <sub>2</sub>	0,799	0,0877	-0,7350	0,015
x <sub>3</sub>	0,422	0,0406	-0,7142	0,020
x <sub>4</sub>	0,563	0,0283	0,2575	0,473

Tabulka 11: Párové korelační koeficienty mezi dvojicemi nezávisle proměnných

Parametr	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>
x <sub>1</sub>	1			
x <sub>2</sub>	0,448357	1		
x <sub>3</sub>	0,505701	0,67312	1	
x <sub>4</sub>	-0,62709	-0,20888	-0,0233	1

Z hodnot párových korelačních koeficientů je patrná vysoká korelace mezi y a x<sub>1</sub>, x<sub>2</sub> a x<sub>3</sub>. Významná korelace je mezi nezávisle proměnnými x<sub>1</sub> vs. x<sub>4</sub> a x<sub>2</sub> vs. x<sub>3</sub>.

### 3.3.3. Odhady parametrů a testy významnosti

Tabulka 12: Odhady parametrů modelu a test významnosti parametrů

H<sub>0</sub>: B<sub>i</sub>=0 vs. H<sub>a</sub>: B<sub>i</sub><>0

Parametr	Odhad	Směrodatná odchylka	t-kriterium	hypotéza H <sub>0</sub> je	Hlad. výz.
konstanta	1364,6	445,09	3,0658	Zamítnuta	0,028
B[ 1]	-61,423	21,993	-2,7929	Zamítnuta	0,038
B[ 2]	-104,04	56,624	-1,8373	Akceptována	0,126
B[ 3]	-27,029	140,09	-0,19295	Akceptována	0,855
B[ 4]	-235,27	182,5	-1,2892	Akceptována	0,254

Testujeme hypotézu  $H_0$ .  $B_i$  je rovno nule. Vzhledem k tomu, že tabulkový kvantil  $t(0,05, 5)$  je 2,57 a hodnoty testovacího kritéria pro  $B_2$ ,  $B_3$  a  $B_4$  jsou menší, je primární hypotéza pro tyto parametry akceptována (příslušné směrnice a úsek se statisticky významně neodlišují od nuly). Úsek a směrnice  $b_1$  (parametr sytost pigmentu) jsou významné parametry.

### 3.3.4. Statistické charakteristiky

**Tabulka 13: Statistické charakteristiky regrese**

Vícenásobný korelační koeficient, R	0,929
Koeficient determinace, $R^2$	0,86437
Predikovaný korelační koeficient, $R_p^2$	0,74439
Střední kvadratická chyba predikce, MEP	185,96
Akaikeho informační kritérium, AIC	50,354

Koeficient determinace má hodnotu 0,8644- cca 86,4% hodnot vyhovuje zvolenému regresnímu modelu. Lze tedy očekávat, že zvolený model je statisticky významný.

### 3.3.5. Charakteristiky reziduí

**Tabulka 14: Charakteristiky reziduí**

Rezidualní součet čtverců, RSC	565,67
Průměr absolutních hodnot reziduí, $M_e$	6,0217
Průměr relativních reziduí, $M_{er}$	19,806
Odhad reziduálního rozptylu, $s^2(e)$	113,13
Odhad směrodatné odchylky reziduí, $s(e)$	10,636
Odhad šikmosti reziduí, $g_1(e)$	0,0051896
Odhad špičatosti reziduí, $g_2(e)$	2,3533

Z hodnot šikmosti a špičatosti lze usuzovat, že bude splněn jeden z předpokladů použití metody nejmenších čtverců – normalita.

### 3.3.6. Testování regresního tripletu

Použití metody nejmenších čtverců je podmíněno splněním určitých podmínek. Jejich prověření je provedeno v následujícím bloku testů a doloženo pomocí níže uvedených grafů..

#### **Fisher-Snedecorův test významnosti modelu**

Hodnota kritéria F : 7,977428113

Kvantil F (1-alfa, m-1, n-m) :5,192167773

Pravděpodobnost : 0,0213522664

**Závěr : Model je významný**

#### **Scottovo kritérium multikolinearity**

Hodnota kritéria SC : 0,2768522487

**Závěr :Model je korektní.**

#### **Cook-Weisbergův test heteroskedasticity**

Hodnota kritéria CW : 0,002991181157

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

Pravděpodobnost : 0,9563840884

**Závěr : Rezidua vykazují homoskedasticitu.**

**Jarque-Berrův test normality**

Hodnota kritéria JB : 0,179070802

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

Pravděpodobnost : 0,9143558954

**Závěr : Rezidua mají normální rozdělení.**

**Waldův test autokorelace**

Hodnota kritéria WA : 0,4734917748

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

Pravděpodobnost : 0,4913851596

**Závěr : Autokorelace je nevýznamná**

**Znaménkový test reziduí**

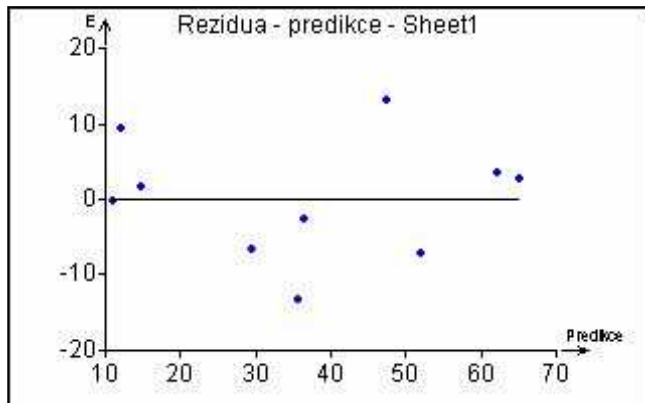
Hodnota kritéria Sg : 0,3354101966

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

Pravděpodobnost : 0,7373156772

**Závěr : V reziduích není trend.**

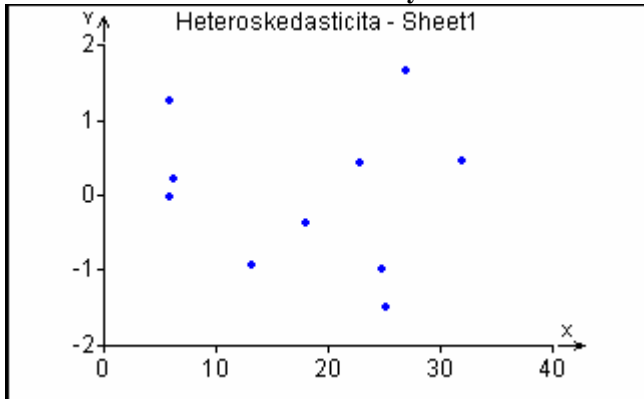
Obrázek 18: Graf reziduí



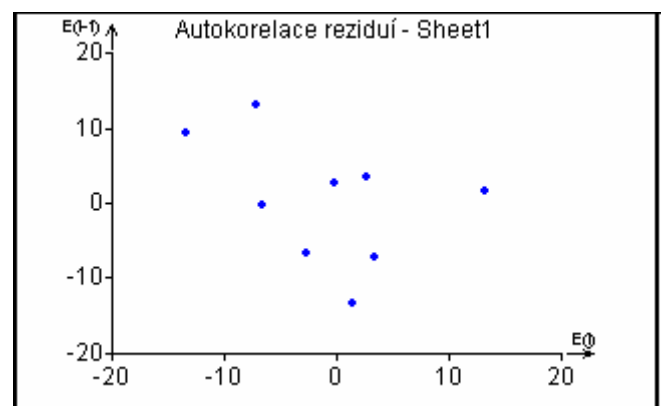
Obrázek 19: Graf normality reziduí



Obrázek 20: Graf homoskedasticity



Obrázek 21: Autokorelace reziduí



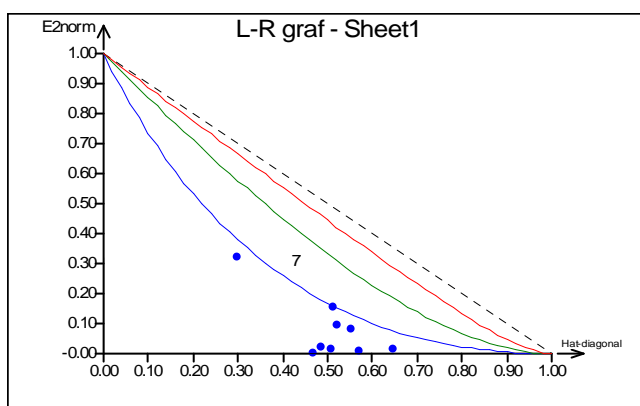
Z výše uvedených testů je patrné, že jsou splněny všechny podmínky použití metody nejmenších čtverců. Zvolený model je statisticky významný, ale z hlediska multikolinearity není

příliš korektní. To je důsledek přeučení modelu. Tento problém odstraníme vypuštěním nevýznamných parametrů. Na Obr.18 rezidua náhodně oscilují kolem nulové hodnoty. Obdobně tomu je i u grafu homoskedasticity (Obr.20) a autokorelace reziduí (Obr.21). Z Q-Q grafu je dobře patrná normalita reziduí (body vykazují hladkou lineární závislost, Obr.19).

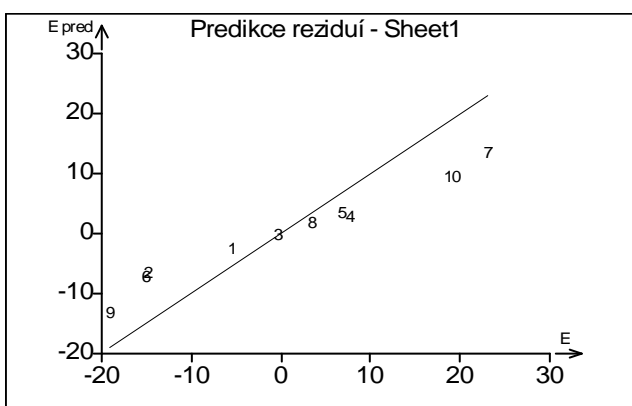
### 3.3.7. Identifikace odlehlých bodů

. Přítomnost podezřelých bodů je patrná z grafu Predikce reziduí a z Willimasova grafu (viz. obr.23 a 25). Nicméně ostatní grafy (Obr.22 a Obr.24) přítomnost odlehlých bodů neindikují. V tabulce č.1 (Jackknife) není rovněž zaznačen žádný odlehlý bod. Ve výběru jsou podezřelé body (bod č.7 a 10, viz. Cookova a Atkinsonova vzdálenost a parametr DF v tab.17), ale k jejich vyloučení není dostatek argumentů.

Obrázek 22:L-R graf



Obrázek 23:Graf predikce reziduí

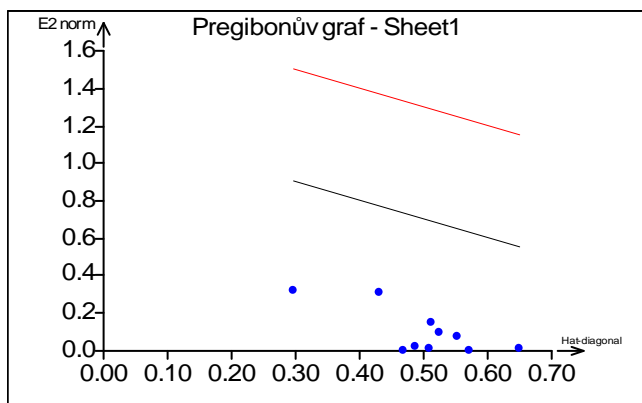


Tabulka 15: Identifikace vlivných bodů (\* indikuje odlehlý nebo vlivný bod)

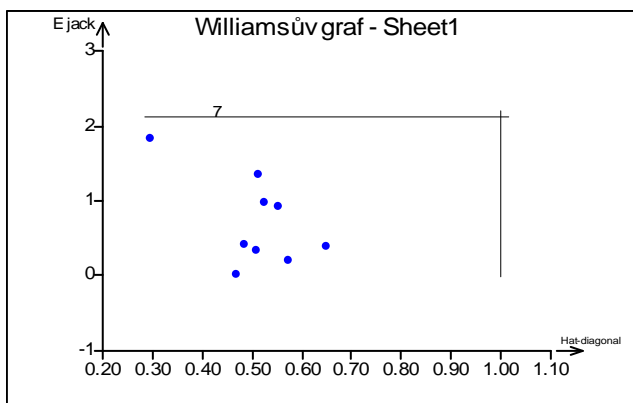
Bod	Standardizované reziduum	Jackknife reziduum	Predikované reziduum	Diagonální prvky
i	eS[i]	eJ[i]	eP[i]	H[i,i]
1	-0,34715	-0,31431	-5,2371	0,50292
2	-0,94574	-0,9335	-15,032	0,55216
3	-0,018975	-0,016972	-0,2765	0,46724
4	0,41642	0,37909	7,517	0,65281
5	0,46254	0,42285	6,8337	0,48169
6	-0,97879	-0,97369	-15,117	0,52571
7	1,6438	2,1686	23,175	0,43085
8	0,2207	0,19837	3,6144	0,5782
9	-1,5087	-1,8283	-19,138	0,29686
10	1,2433	1,3379	18,922	0,51156



Obrázek 24: Pregibonův graf



Obrázek 25: Williamsův graf



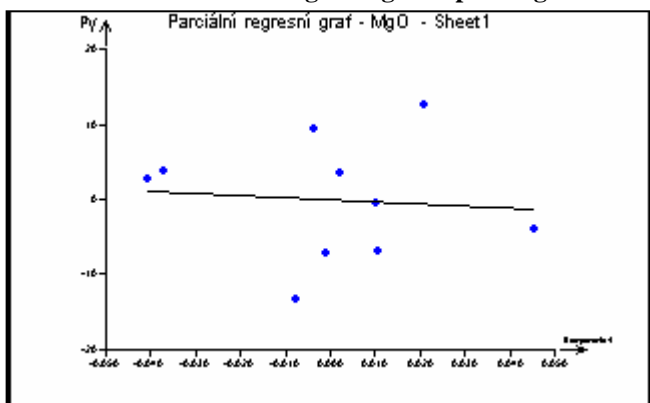
Tabulka 16: Identifikace odlehlých bodů

Bod	Zobecněné diag. prvky	Cookova vzdálenost	Atkinsonova vzdálenost	Vliv na predikci
i	Hm[i,i]	D[i]	A[i]	DF[i]
1	0,5149	0,024385	0,31615	-0,31615
2	0,63227	0,22056	1,0365	-1,0365
3	0,46728	0,00006315	0,015894	-0,01589
4	0,66485	0,06521	0,51982	0,51982
5	0,50387	0,039766	0,40765	0,40765
6	0,61658	0,21238	1,0251	-1,0251
7	0,73841	0,40908*	1,8868*	1,8868*
8	0,58231	0,013354	0,23225	0,23225
9	0,61697	0,19221	1,188	-1,188
10	0,66256	0,32379*	1,3692	1,3692

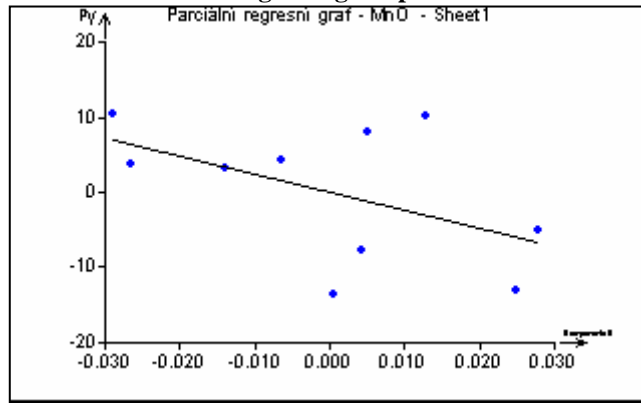
### 3.3.8. Upravený model

Z předchozího textu je patrné, že je nutné vypustit parametry 3 a 4, které se jeví jako nevýznamné (viz. Obr.26 a Obr.27). Hodnoty parametrů modelu jsou po úpravě následující (viz. Tab.17):

Obrázek 26: Parciální regresní graf – par. MgO



Obrázek 27: Parciální regresní graf- par.MnO



**Tabulka 17: Odhady parametrů upraveného modelu a testy významnosti**

Parametr	Odhad	Směrodatná odchylka	t-kriterium	hypotéza $H_0$	Hlad. výz.
B[ 0]	951,72	260,65	3,6513	Zamítnuta	0,008
B[ 1]	-45,794	15,353	-2,9827	Zamítnuta	0,02
B[ 2]	-117,9	46,678	-2,5259	Zamítnuta	0,039

Regresní parametry upraveného regresního modelu jsou následující:

**Tabulka 18: Parametry původního a upraveného regresního modelu**

Model	Původní	Upravený
Vícenásobný korelační koeficient, R	0,92971	0,89305
Koeficient determinace, $R^2$	0,86437	0,79754
Predikovaný korelační koeficient, $R_p^2$	0,74439	0,73313
Střední kvadratická chyba predikce, MEP	185,96	192,9
Akaikeho informační kritérium, AIC	50,354	50,36

Z výsledků níže uvedených testů je patrné, že jsou také všechny splněny podmínky použití metody nejmenších čtverců.

**Fisher-Snedecorův test významnosti modelu**

Hodnota kritéria F : 13,78777218

Kvantil F (1-alfa, m-1, n-m) : 4,737414126

Pravděpodobnost : 0,003733804267

**Závěr : Model je významný**

**Scottovo kritérium multikolinearity**

Hodnota kritéria SC : 0,182285338

**Závěr : Model je korektní.**

**Cook-Weisbergův test heteroskedasticity**

Hodnota kritéria CW : 0,6662945904

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

Pravděpodobnost : 0,4143464721

**Závěr : Rezidua vykazují homoskedasticitu.**

**Jarque-Berrův test normality**

Hodnota kritéria JB : 0,8498263975

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

Pravděpodobnost : 0,6538265356

**Závěr : Rezidua mají normální rozdělení.**

**Waldův test autokorelace**

Hodnota kritéria WA : 0,5191533213

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

Pravděpodobnost : 0,4712030746

**Závěr : Autokorelace je nevýznamná**

### Znaménkový test reziduí

Hodnota kritéria Sg : 0,3354101966

Kvantil N(1-alfa/2) : 1,959963999

Pravděpodobnost : 0,7373156772

**Závěr :V reziduích není trend.**

### 3.3.9. Závěr a vyhodnocení

Výsledný model má tedy tvar  $y=951,72(\pm 260,65)-45,794(15,35).x_1-117,9(\pm 46,68).x_2$ . Kde  $x_1$  je sytost C\* pigmentu a  $x_2$  je poměr vytrádkového monohydrátu k celkově vyrobenému pigmentu. Čím je odstín pigmentu světlejší a čím je větší množství vytrádkového pigmentu, tím je nižší tvorba částic nad 0,045 mm v kalcinaci (y).

### 3.4. Určení stupně polynomu

Byla sledována rychlost dehydratace sádrovce ( $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ ) při teplotě 100°C jako závislost úbytku hmotnosti na čase. Proložte tuto závislost vhodným typem polynomu.

#### Data:

čas (min)	0	2	4	6	10	11	14
$\Delta m$ (%)	0	0,75	2,31	4,8	9,45	10,31	13,1

čas (min)	24	28	31	48
$\Delta m$ (%)	15,68	15,8	15,8	15,98

#### 3.4.1. Návrh modelu

Při určení stupně polynomu budeme postupovat tak, že zkusmo vygenerujeme pro stupeň polynomu 2, 4, 5 a 6 statistické charakteristiky regrese (AIC,  $R_p^2$ , MEP a  $R^2$ ). Pro optimální stupeň polynomu vychází hodnota parametru AIC (Akaikeho informační kritérium) a MEP (střední kvadratická chyba predikce) nejmenší. Pomocí těchto ukazatelů provedeme návrh modelu.

**Tabulka 19: Určení stupně polynomu**

Stupeň polynomu	2	3	4	5	6	7
Vícenásobný korelační koeficient, R	0,984	0,992	0,997	1,000	0,999915	0,999925
Koeficient determinace, $R^2$	0,968	0,984	0,994	1,000	0,999831	0,99985
Predikovaný korelační koeficient, $R_p$	0,451	0,039	182,731	5,517	317,2103	0,796943
Střední kvadratická chyba predikce, MEP	12,358	45,076	546,074	125,964	707,5368	4,035374
Akaikeho informační kritérium, AIC	7,951	2,478	-5,565	-41,488	-41,6159	-40,9765

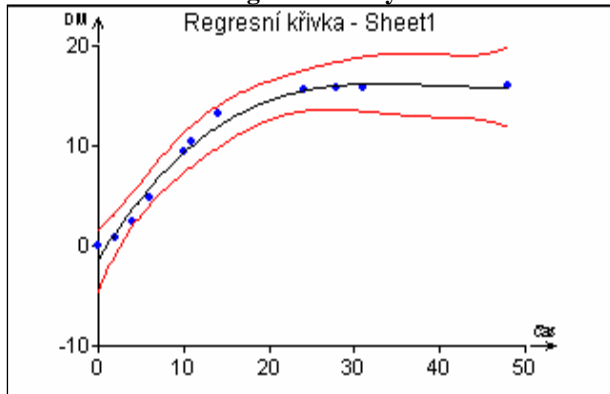
Z tab.19 a Obr.29 a Obr.30 je patrné, že nejvýhodnější bude proložit uvedenou závislost polynomem 3.stupně.

**Tabulka 20: Odhady parametrů (s absolutním členem)**

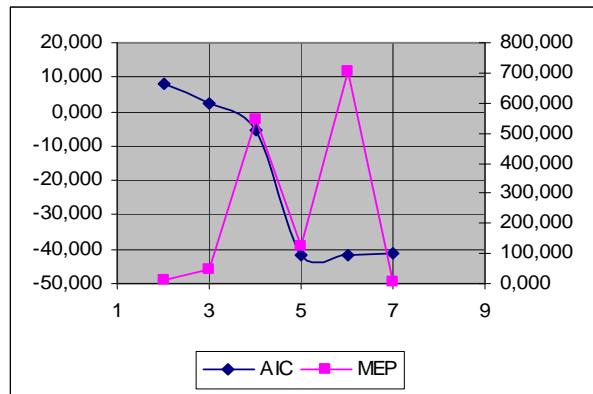
Proměnná	Odhad	Směr.odchylka	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-1,4466	0,733242	Nevýznamný	0,089111	-3,18045	0,287237
čas	1,408948	0,166127	Významný	6,27E-05	1,016119	1,801777
čas ^2	-0,03673	0,008691	Významný	0,003907	-0,05728	-0,01618
čas ^3	0,00031	0,000119	Významný	0,034951	2,91E-05	0,000592

Součástí tab.20 je testování významnosti parametrů modelu. Absolutní člen vychází jako nevýznamný (hodnota Pravděpodobnosti je větší než 0,05 a 95% interval spolehlivosti zahrnuje nulu).

**Obrázek 28: Graf regresní křivky**



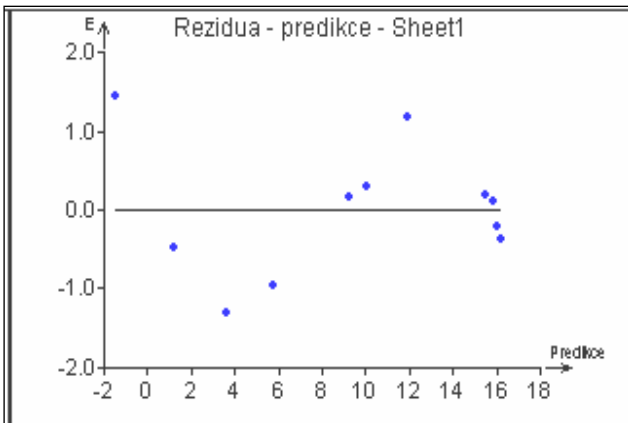
**Obrázek 29: Graf AIC a MEP**



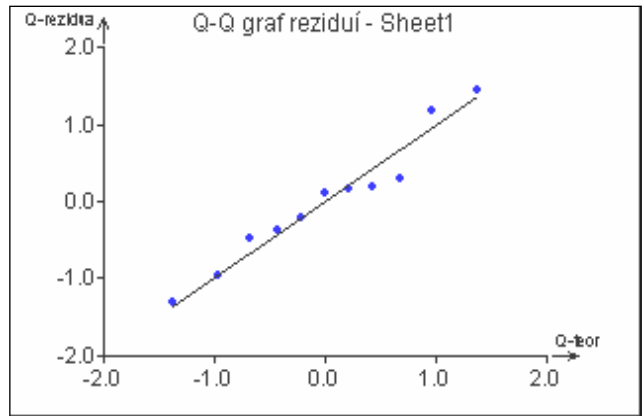
### 3.4.2. Kritika dat

Z Obr. 31 je patrné, že bude splněna podmínka normality reziduí. Body na Obr.30 (graf rezidua vs. predikce) lze proložit křivkou z čehož vyplývá, že dojde patrně k problému s multikolinearitou.

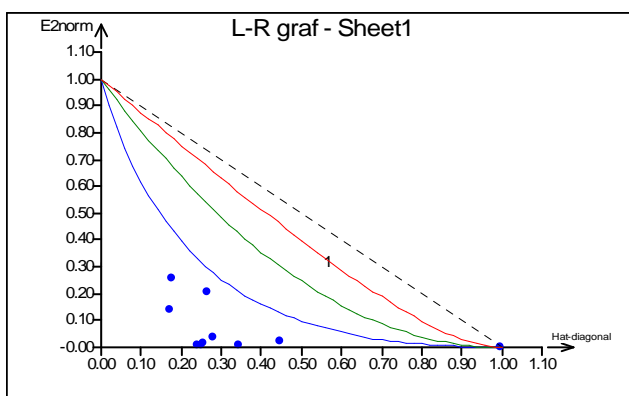
**Obrázek 30: Graf rezidua vs. predikce**



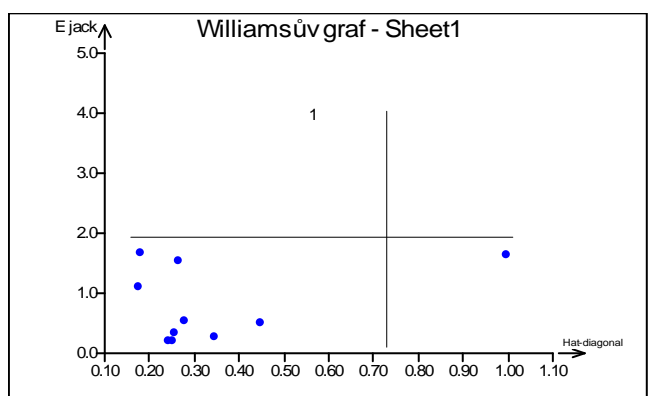
**Obrázek 31: Q-Q graf reziduí**



**Obrázek 32: L-R graf**



**Obrázek 33: Williamsův graf**



Z Williamsova grafu (obr. 33) je patrná přítomnost odlehlého bodu s číslem 1. Tento bod leží v L-R grafu na obr. 34 těsně na hranici, za kterou jsou body považovány za odlehlé.

### 3.4.3. Testování regresního modelu

Použití metody nejmenších čtverců je podmíněno splnění určitých kritérií. Jejich a výčet a ověření je uveden v následujícím textu.

#### **Fisher-Snedecorův test významnosti modelu**

Hodnota kritéria F : 142,6608807

Kvantil F (1-alfa, m-1, n-m) : 4,3468314

Pravděpodobnost : 1,223144941E-006

**Závěr : Model je významný**

#### **Scottovo kritérium multikolinearity**

Hodnota kritéria SC : 0,700536177

**Závěr : Model vykazuje multikolinearitu!**

#### **Cook-Weisbergův test heteroskedasticity**

Hodnota kritéria CW : 1,485961043

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

Pravděpodobnost : 0,2228441858

**Závěr : Rezidua vykazují homoskedasticitu.**

#### **Jarque-Berrův test normality**

Hodnota kritéria JB : 0,2385088578

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

Pravděpodobnost : 0,8875819455

**Závěr : Rezidua mají normální rozdělení.**

#### **Waldův test autokorelace**

Hodnota kritéria WA : 0,3309817873

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

Pravděpodobnost : 0,5650815472

**Závěr : Autokorelace je nevýznamná**

#### **Durbin-Watsonův test autokorelace**

Hodnota kritéria DW : -1

Kritické hodnoty DW0

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

#### **Znaménkový test reziduí**

Hodnota kritéria Sg : 0,6123724357

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

Pravděpodobnost : 0,5402913746

**Závěr : V reziduích není trend.**

Z výše uvedených statistických charakteristik modelu je patrné, že prakticky všechny důležité parametry jsou v případě tohoto modelu splněny. Výjimkou je pouze vysoká hodnota Scottova kritéria. Hodnota tohoto ukazatele, ale nedosahuje kritické hodnoty, která by vyžadovala zamítnutí modelu.

### 3.4.4. Upravený model

Absolutní člen vyšel v Tab.22 jako nevýznamný. Model s absolutním členem vykazuje rovněž vysokou hodnotu Scottova kritéria. Z těchto důvodů provedeme vypočet parametrů regrese pro model bez absolutního členu a následně provedeme testování regresního tripletu.

**Tabulka 21: Odhad parametrů modelu (bez absolutního členu)**

Proměnná	Odhad	Směr.odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
čas	1,143277	0,113519	Významný	8,05E-06	0,881501	1,405053
čas ^2	-0,02498	0,007387	Významný	0,009616	-0,04202	-0,00795
čas ^3	0,000167	0,00011	Nevýznamný	0,166754	-8,64E-05	0,000421

Z tab.21 je patrné, že parametru čas ^3 se jeví jako nevýznamný.

**Tabulka 22: Statistické parametry regrese (bez absolutního členu)**

Vícenásobný korelační koeficient R :	0,9874
Koeficient determinace R <sup>2</sup> :	0,974959
Předikovaný korelační koeficient Rp :	1,134662
Střední kvadratická chyba predikce MEP :	77,6809
Akaikeho informační kritérium :	5,341414

ABS-absolutní člen

### 3.4.5. Testování modelu (bez absolutního členu)

#### Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F : 155,7396911

Kvantil F (1-alfa, m-1, n-m) : 4,458970107

Pravděpodobnost : 3,931774528E-007

**Závěr : Model je významný**

#### Scottovo kritérium multikolinearity

Hodnota kritéria SC : 0,915444722

**Závěr : Model je nekorektní!**

#### Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0,7277003236

Kvantil Chi<sup>2</sup>(1-alfa,1) : 3,841458829

Pravděpodobnost : 0,3936296202

**Závěr : Rezidua vykazují homoskedasticitu.**

#### Jarque-Berrův test normality

Hodnota kritéria JB : 1,193808738

Kvantil Chi<sup>2</sup>(1-alfa,2) : 5,991464547

Pravděpodobnost : 0,5505131867

**Závěr : Rezidua mají normální rozdělení.**

**Waldův test autokorelace**

Hodnota kritéria WA : 3,02030673  
Kvantil  $\chi^2(1-\alpha,1)$  : 3,841458829  
Pravděpodobnost : 0,08222791247

**Závěr : Autokorelace je nevýznamná**

**Durbin-Watsonův test autokorelace**

Hodnota kritéria DW : -1  
Kritické hodnoty DW0

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

**Znaménkový test reziduí**

Hodnota kritéria Sg : 0,6123724357  
Kvantil  $N(1-\alpha/2)$  : 1,959963999  
Pravděpodobnost : 0,5402913746

**Závěr : V reziduích není trend.**

Použití polynomu 3.stupně bez absolutního členu vedlo k velkému zhoršení Scottova kritéria a model je z tohoto hlediska nekorektní.

**3.4.6. Závěr**

Volbou polynomu 3. stupně s absolutním členem bylo dosaženo nejlepších hodnot statistických ukazatelů regrese. Model je zatížen multikolinearitou, ale hodnota Scottova kritéria není vyšší než 0,8. Z toho důvodu nebyla použita metoda Racionálních hodnotí. Ve výběru je indikována přítomnost podezřelého bodu (č.11), ale k jeho vypuštění není dostatek argumentů. Absolutní člen vychází jako statisticky nevýznamný. Jeho vypuštění vedlo k vysoké hodnotě multikolinearity a proto byl zachován. Tvar zvoleného modelu je následující (v závorce je uvedena směrodatná odchylka, t je čas a y je pokles hmotnosti v hmotnostních procentech):

$$y = -1,45(\pm 0,733) + 1,4089(\pm 0,166) \cdot t - 0,0367(\pm 0,0087) \cdot t^2 + 0,00031(\pm 0,000119) \cdot t^3$$

## Obsah:

1. Úvod.....	2
2. Zadané úkoly .....	2
3. Vypracované otázky .....	2
3.1. Porovnání dvou regresních přímek.....	2
3.1.1. Zadání .....	2
3.1.2. Model sušina1 vs. $\rho_T$ .....	2
3.1.2.1. Návrh modelu.....	2
3.1.2.2. Kritika dat .....	3
3.1.2.3. Ověření předpokladů metody nejmenších čtverců.....	5
3.1.2.4. Závěr analýzy modelu sušina1 vs. $\rho_T$ .....	5
3.1.3. Model sušina2 vs. $\rho_T$ .....	6
3.1.3.1. Návrh modelu .....	6
3.1.3.2. Kritika dat .....	6
3.1.3.3. Ověření předpokladů metody nejmenších čtverců.....	8
3.1.3.4. Závěr analýzy modelu sušina2 vs. $\rho_T$ .....	8
3.1.4. Porovnání dvou lineárních závislostí .....	9
3.2. Validace nové analytické metody .....	9
3.2.1. Zadání .....	9
3.2.2. Návrh modelu.....	9
3.2.3. Kritika dat .....	10
3.2.4. Ověření předpokladů metody nejmenších čtverců.....	11
3.2.5. Závěry provedené validace .....	12
3.3. Vícerozměrný lineární regresní model .....	12
3.3.1. Zadání .....	12
3.3.2. Předběžná analýza dat .....	13
3.3.3. Odhady parametrů a testy významnosti.....	13
3.3.4. Statistické charakteristiky .....	14
3.3.5. Charakteristiky reziduí .....	14
3.3.6. Testování regresního tripletu.....	14
3.3.7. Identifikace odlehlých bodů.....	16
3.3.8. Upravený model .....	17
3.3.9. Závěr a vyhodnocení.....	19
3.4. Určení stupně polynomu .....	19
3.4.1. Návrh modelu.....	19
3.4.2. Kritika dat .....	20
3.4.3. Testování regresního modelu.....	21
3.4.4. Upravený model .....	22
3.4.5. Testování modelu (bez absolutního členu) .....	22
3.4.6. Závěr .....	23