

6. Lineární regresní modely

- 6.1 Jednoduchá regrese a validace
- 6.2 Testy hypotéz v lineární regresi
- 6.3 Kritika dat v regresním tripletu
- 6.4 Multikolinearita a polynomy
- 6.5 Kritika modelu v regresním tripletu**
- 6.6 Kritika metody v regresním tripletu
- 6.7 Lineární a nelineární kalibrace
- 7. Korelační modely

1

FORMULACE REGRESNÍHO MODELU

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

závisle proměnná nezávisle proměnná regresní parametry náhodná chyba

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Kritika modelu

Vyšetření významnosti parametrů

Metoda:

- Parciální regresní grafy
- Parciální reziduální grafy
- Studentův t -test významnosti parametru β

TYPY REGRESNÍHO MODELU

Regresní model předpokládá, že **nezávislá proměnná** (proměnné) je **nenáhodná** (tj. pevně určena, např. experimentátorem) a **závislá proměnná** je **náhodná** (měřená).

Tento předpoklad nebývá v praxi splněn (často jsou obě nebo všechny veličiny naměřené, potom nazýváme tento model **korelačním**).

Rozeznáváme:

- ◆ regresní modely **lineární** – mají **lineární postavení parametrů**
- ◆ regresní modely **nelineární** – mají **nelineární postavení parametrů**

PODSTATA REGRESNÍ ANALÝZY

Podstatou řešení regrese je:

- ◆ Stanovit **nejlepší regresní model** (čili určit matematickou rovnici, která bude popisovat závislost y na x),
- ◆ Stanovit **parametry modelu** (tj. stanovit nejlepší odhady parametrů β),
- ◆ Stanovit **statistickou významnost modelu** (určit, zda nalezený model přispěje ke zpřesnění odhadu závisle proměnné oproti použití pouhého průměru),
- ◆ Výsledky dané modelem **interpretovat** z hlediska zadání.

KRITÉRIA PRO HLEDÁNÍ A ROZLIŠENÍ NEJLEPŠÍHO REGRESNÍHO MODELU

Střední kvadratická chyba predikce (MEP)

$$\text{MEP} = \frac{1}{n} \sum_{i=1}^n \frac{e_i^2}{(1 - H_{ii})^2}$$

e_i^2 čtverec reziduí modelu
 H_{ii} i -tý diagonální prvek projekční matice \mathbf{H}

Akaikovo informační kritérium (AIC)

$$\text{AIC} = n \cdot \ln \left(\frac{\text{RSC}}{n} \right) + 2m$$

RSC reziduální součet čtverců
 m počet parametrů

Pravidlo: Čím je AIC (nebo MEP) menší, tím je model vhodnější.

URČENÍ VHODNÉHO MODELU

- 1) Najít řadu modelů, které svými vlastnostmi vyhovují řešenému problému (např. rozličné formy růstové funkce),
- 2) Potom najít takový model, který nejlépe vyhovuje naměřeným datům.

Je nutné dbát, aby byla modelována

skutečná příčinná závislost!

Kvalita navrženého regresního modelu

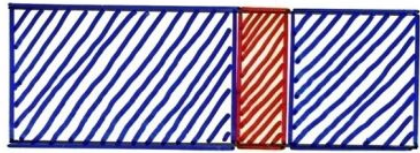
1. **Jediná proměnná x :** rozptylový graf y na x .
Více proměnných x : rozptylové grafy mohou *mylně indikovat* nelinearitu.
2. K posouzení vztahu y a x_j :
 - a) parciální regresní grafy,
 - b) parciální reziduální grafy.

Grafy vyjadřují **závislost mezi vysvětlovanou proměnnou** (vektorem \mathbf{y}) a **jednou vysvětlující proměnnou x_j** , při statisticky neměnném vlivu ostatních vysvětlujících proměnných, které tvoří matici $\mathbf{X}_{(j)}$. Jde o grafickou obdobu parciálního korelačního koeficientu u korelačních modelů.

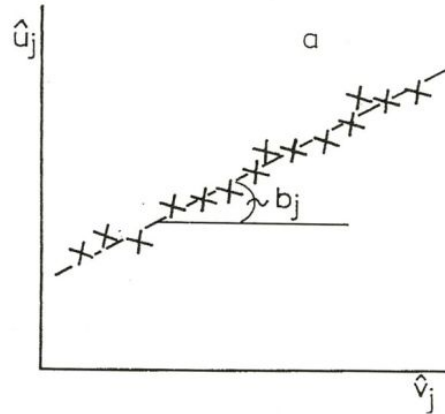
Belseyho parciální regresní grafy

(partial regression leverage plots)

$$\hat{y} = X_{(j)}\beta^* + x_j \cdot c + \varepsilon$$



$X_{(j)}$ jsou konstanty



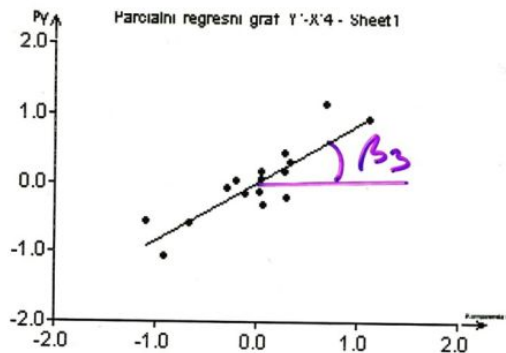
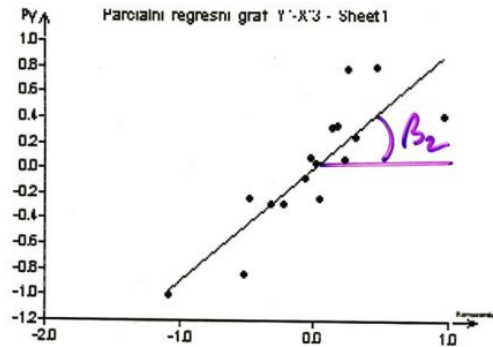
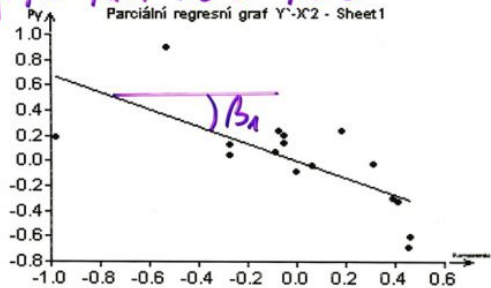
umožňují:

1. Posouzení kvality navrženého regresního modelu,
2. Indikují přítomnost vlivných bodů,
3. Nesplnění předpokladů klasické MNČ,
4. Vyjadřují závislost mezi y a zvolenou proměnnou x_j při statisticky neměnném vlivu ostatních $X_{(j)}$

Ukázky parciálních regresních grafů u vícenásobného regresního modelu pro $m = 3$

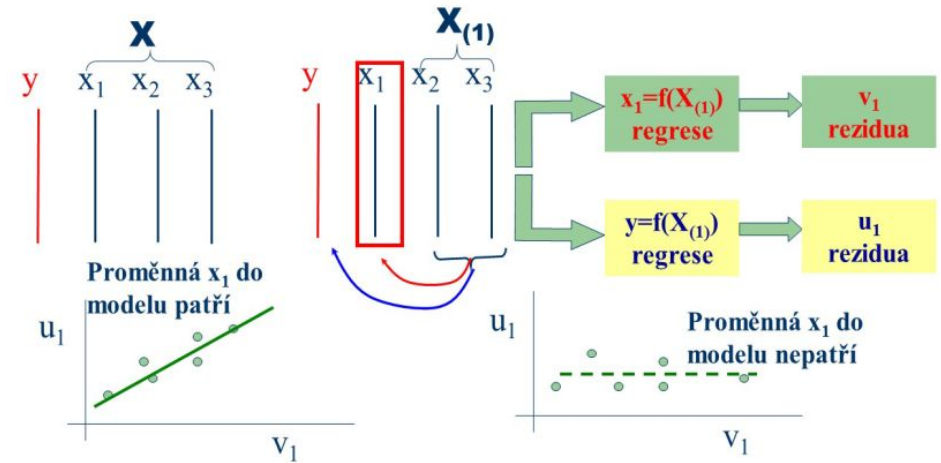
Kritika modelu: parciální regresní graf

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$



Grafický výklad parciálního regresního grafu

Příklad: Zajímá nás, zda všechny proměnné x_{1-3} jsou v modelu oprávněně. Postup je vysvětlen pro proměnnou x_1 .



- ◆ Pokud body parciálního regresního grafu leží na přímce s nulovým úsekem, existuje lineární závislost mezi y a x_j .
- ◆ Směrnice přímky proložená body parciálního regresního grafu číselně odpovídá regresnímu koeficientu b_j původního regresního modelu.
- ◆ Korelační koeficient mezi u_j a v_j odpovídá parciálnímu korelačnímu koeficientu.
- ◆ Rezidua regresní přímky mezi u_j a v_j odpovídají reziduíům původního modelu.

1. vzorová úloha na výstavbu lineárního regresního modelu pomocí parciálního regresního grafu

M619

13

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI: ADSTAT 1.25

Parametr	Odhad	Směrodatná odchylna	Test H0: B[j] = 0 vs. HA: B[j] <> 0	t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	-7.2666E-02	1.3791E-01	-5.2692E-01	Akceptována	0.608	
B[1]	-6.8505E-01	1.9165E-01	-3.5746E+00	Zamítnuta	0.004	
B[2]	8.9619E-01	1.6072E-01	5.5761E+00	Zamítnuta	0.000	
B[3]	8.3769E-01	1.3322E-01	6.2879E+00	Zamítnuta	0.000	

Odhady parametrů

Proměnná	Odhad	Sm. od.	Závěr	Pravděp.	Spodní	Horní
Abs	-0.073	0.138	Nevýzn.	0.608	-0.373	0.228
M619x1	-0.685	0.192	Význ.	0.004	-1.103	-0.267
M619x2	0.896	0.161	Význ.	0.000	0.546	1.246
M619x3	0.838	0.133	Význ.	0.000	0.547	1.128

NCSS2007

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	-7.266543E-02	0.137906	-0.5269	0.607852	Accept Ho	0.077473
M619x1	-0.6850502	0.1916463	-3.5746	0.003820	Reject Ho	0.906070
M619x2	0.8961921	0.1607203	5.5761	0.000121	Reject Ho	0.999127
M619x3	0.8376914	0.1332221	6.2879	0.000040	Reject Ho	0.999919
R-Squared	0.997161					

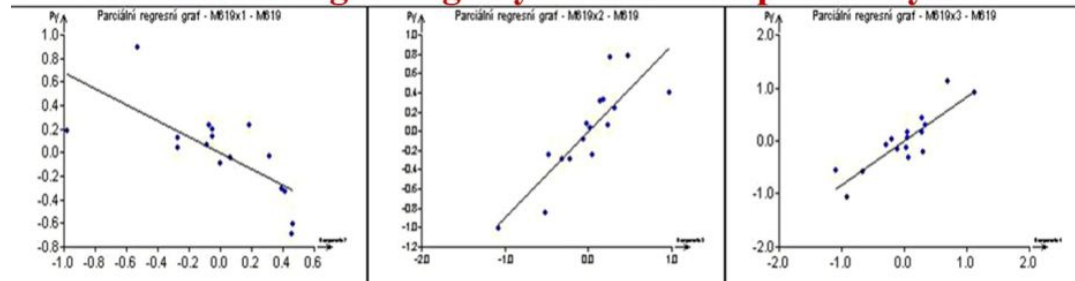
Model: -7.266543E-02 - 0.6850502*M619x1 + 0.8961921*M619x2 + 0.8376914*M619x3

Úloha M619. Vliv tří parametrů na obsah kadmia v potravinářské pšenici
Obsah kadmia v zrna y [mg/l] v závislosti na obsahu kadmia v otrubách x_1 [mg/l], ve stonku s listy x_2 [mg/l] a v kořenovém systému x_3 [mg/l].
Vyšetřete regresní triplet (data, model, metoda) a nalezněte lineární regresní model.

Výstavba lineárního regresního modelu:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Parciální regresní grafy tří nezávisle proměnných



ADSTAT 1.25: statistická kritéria věrohodnosti regr. modelu

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

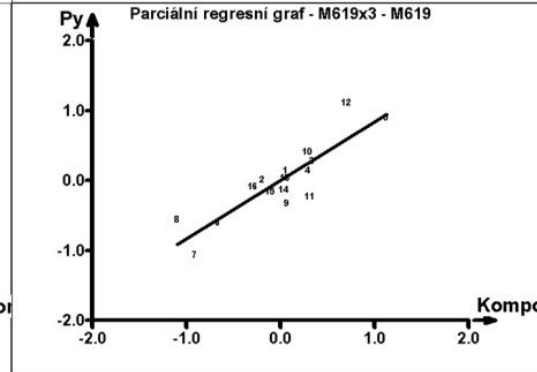
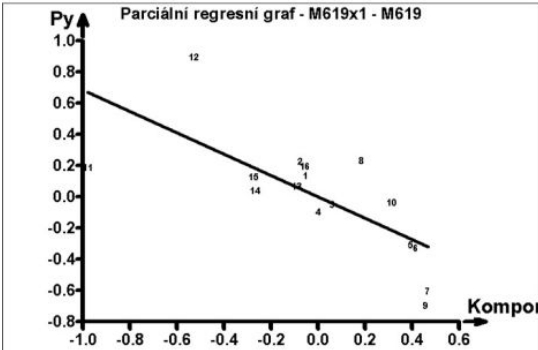
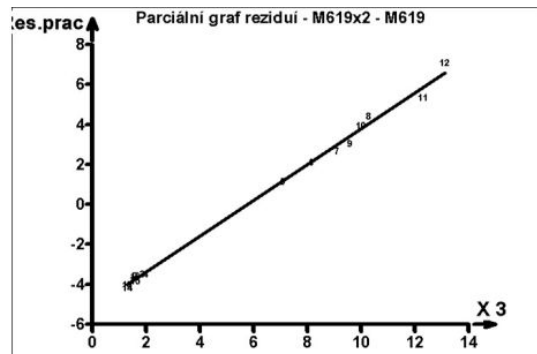
Vícenásobný korelační koeficient, R	: 9.9858E-01
Koeficient determinace, R ²	: 9.9716E-01
Predikovaný korelační koeficient, R _p ²	: 9.9527E-01
Střední kvadratická chyba predikce, MEP	: 2.1014E-01
Akaikeho informační kritérium, AIC	: -3.6180E+01

QC-EXPERT 3.1

Statistické charakteristiky:

Vícenásobný korelační koeficient F	0.999
Koeficient determinace R ²	0.997
Predikovaný korelační koeficient F	0.991
Střední kvadratická chyba predikce	0.210
Akaikeho informační kritérium :	-36.180

Parciální regresní grafy pro tři nezávisle proměnné mají vesměs nenulovou směrnici.



Závěr: Nejlepší lineární regresní model musí vždy obsahovat kritéria věrohodnosti

D. Data bez 8, 11, 12 a opravený model vedou k závěru

$$y = -0.86(0.37)x_1 + 0.95(0.25)x_2 + 0.92(0.13)x_3$$

(LS: $t_{1-0.05|2}(13-3) = 2.228$, $D = 99.83\%$, $MEP = 0.05101$, $AIC = -43.55$, $s(e) = 0.170$)

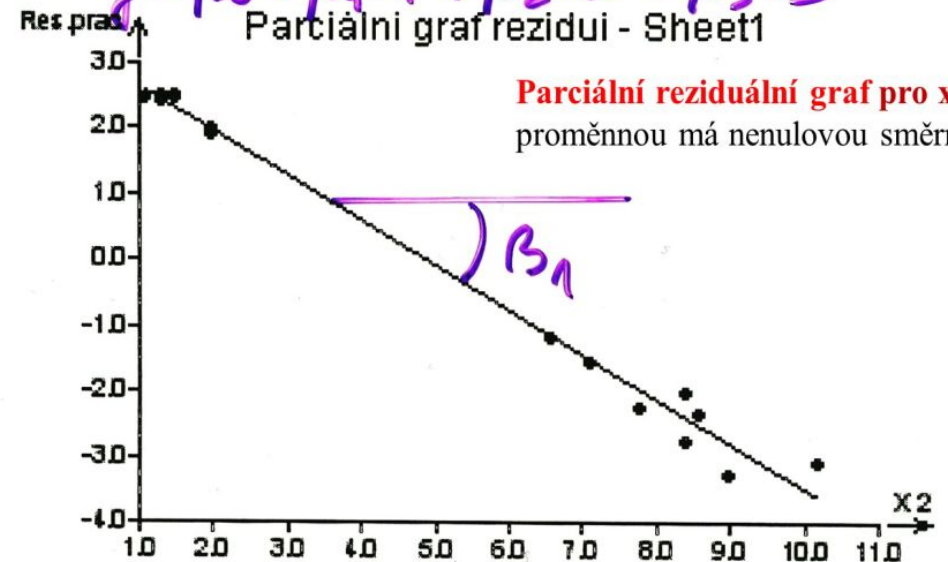
β_0	0.000	---	---	---	---
β_1	-0.855	0.372	-2.3002	Významný	0.044
β_2	0.954	0.251	3.8038	Významný	0.003
β_3	0.916	0.132	6.9263	Významný	0.000

Model: $y = -0.86(0.37)x_1 + 0.95(0.25)x_2 + 0.92(0.13)x_3$

18

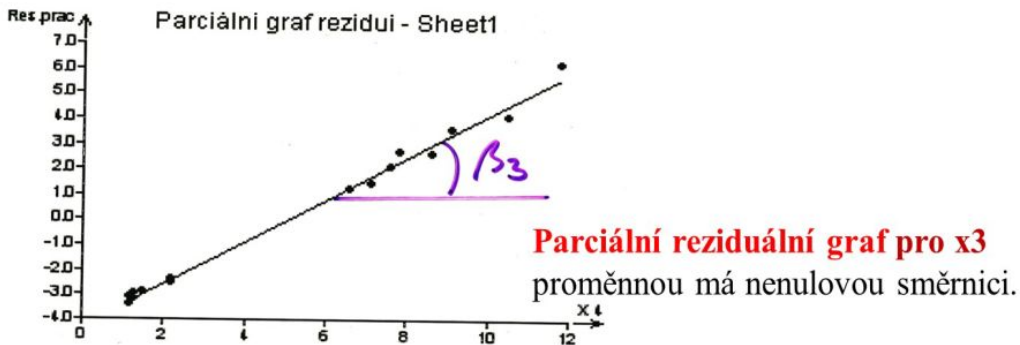
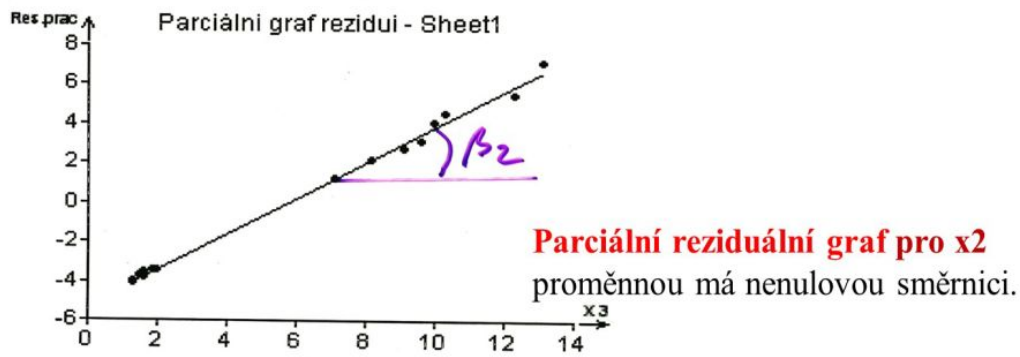
Kritika modelu: parciální reziduální graf

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$



1. vzorová úloha
na výstavbu
lineárního regresního modelu
pomocí parciálního reziduálního grafu

M619



Závěr: Nalezený nejlepší lineární regresní model musí obsahovat kritéria věrohodnosti

D. Data bez 8, 11, 12 a opravený model vedou k závěru

$$y = -0.86(0.37)x_1 + 0.95(0.25)x_2 + 0.92(0.13)x_3$$

(LS: $t_{1-0.05/2}(13-3) = 2.228$, $D = 99.83\%$, $MEP = 0.05101$, $AIC = -43.55$, $s(e) = 0.170$)

β_0	0.000	---	---	---	---
β_1	-0.855	0.372	-2.3002	Významný	0.044
β_2	0.954	0.251	3.8038	Významný	0.003
β_3	0.916	0.132	6.9263	Významný	0.000

Model: $y = -0.86(0.37)x_1 + 0.95(0.25)x_2 + 0.92(0.13)x_3$

22

1. Kvalita nalezených odhadů parametrů

a) Podle **intervalů spolehlivosti** (čím menší interval spolehlivosti, tím lépe)

$$\beta_j = b_j \pm \sqrt{C_{mm} \cdot m \cdot s^2 \cdot F_{1-\alpha; m; n-m}}$$

b) Podle **rozptylů parametrů**, kde pro kvalitní odhad musí platit

$$2 \cdot \sqrt{D(b_j)} < |b_j|$$

(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]	-7.2666E-02	1.3791E-01	-5.2692E-01	Akceptována	0.608
B[1]	-6.8505E-01	1.9165E-01	-3.5746E+00	Zamítnuta	0.004
B[2]	8.9619E-01	1.6072E-01	5.5761E+00	Zamítnuta	0.000
B[3]	8.3769E-01	1.3322E-01	6.2879E+00	Zamítnuta	0.000

POSTUP VÝSTAVBY REGRESNÍHO MODELU

2. Kvalita dosažené těsnosti proložení

- Podle **reziduálního rozptylu $s(y)$** .
- Podle **regresního rabatu D** (= koeficient determinace v %: čím více se blíží 100 %, tím lepší je proložení).

3. Vhodnost navrženého modelu

Akaikovo informační kritérium AIC (čím je menší nebo zápornější, tím vhodnější je navržený model).

Střední kvadratická chyba predikce MEP (čím je MEP menší, tím je predikční schopnost navrženého modelu lepší).

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.9858E-01
Koeficient determinace, R^2	: 9.9716E-01
Predikovaný korelační koeficient, R_p^2	: 9.9527E-01
Střední kvadratická chyba predikce, MEP	: 2.1014E-01
Akaikovo informační kritérium, AIC	:-3.6180E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	$y_{exp}[i]$	$y_{vyp}[i]$	$s(y_{vyp}[i])$	$e[i]$	$er[i]$
1	1.6000E+00	1.5006E+00	1.0203E-01	9.9415E-02	6.2135E+00
2	1.6000E+00	1.4227E+00	1.0587E-01	1.7733E-01	1.1083E+01
3	2.1000E+00	2.1029E+00	1.1045E-01	-2.9205E-03	-1.3907E-01
4	2.1000E+00	2.1925E+00	1.0412E-01	-9.2540E-02	-4.4067E+00
5	8.1000E+00	8.1354E+00	2.0871E-01	-3.5421E-02	-4.3729E-01
6	7.9000E+00	7.9410E+00	1.3352E-01	-4.1016E-02	-5.1919E-01
7	8.4000E+00	8.6869E+00	1.6249E-01	-2.8690E-01	-3.4155E+00
8	1.0300E+01	9.9377E+00	1.7913E-01	3.6232E-01	3.5176E+00
9	9.6000E+00	9.9805E+00	1.3555E-01	-3.8050E-01	-3.9636E+00
10	1.0800E+01	1.0621E+01	1.3061E-01	1.7918E-01	1.6591E+00
11	1.3100E+01	1.3581E+01	2.2832E-01	-4.8080E-01	-3.6703E+00
12	1.5100E+01	1.4565E+01	1.8609E-01	5.3530E-01	3.5450E+00
13	1.3000E+00	1.2908E+00	1.0514E-01	9.1821E-03	7.0632E-01
14	1.2000E+00	1.3441E+00	1.1635E-01	-1.4406E-01	-1.2005E+01
15	1.5000E+00	1.5597E+00	1.1723E-01	-5.9675E-02	-3.9784E+00
16	1.5000E+00	1.3389E+00	1.0987E-01	1.6110E-01	1.0740E+01
Rezidualní součet čtverců, RSC :					1.0114E+00
Průměr absolutních hodnot reziduí, Me :					1.9048E-01
Průměr relativních reziduí, Mer :					4.3750E+00
Odhad reziduálního rozptylu, $s^2(e)$:					8.4282E-02
Odhad směrodatné odchylky reziduí, $s(e)$:					2.9031E-01
Odhad šikmosti reziduí, $g_1(e)$:					9.2356E-02
Odhad špičatosti reziduí, $g_2(e)$:					2.8755E+00

4. Predikční schopnost modelu

Střední kvadratická chyba predikce MEP (čím je MEP menší, tím je predikční schopnost navrženého modelu lepší).

5. Kvalita experimentálních dat

- Na základě analýzy rozličných druhů **reziduí**.
- Na základě **Indikace vlivných bodů**

(Jackknife rezidua, standardizovaná rezidua, normovaná rezidua, predikovaná rezidua, rekurzivní rezidua,

Cookova vzdálenost, diagonální prvky projekční matice a věrohodnostní vzdálenosti).

(7) INDIKACE VLIVNÝCH BODŮ:

Bod	Standard. reziduum	Jackknife reziduum	Predikované reziduum	Diagonální prvky
i	$eS[i]$	$eJ[i]$	$eP[i]$	$H[i,i]$
1	3.6577E-01	3.5217E-01	1.1343E-01	1.2352E-01
2	6.5601E-01	6.3966E-01	2.0453E-01	1.3298E-01
3	-1.0878E-02	-1.0415E-02	-3.4148E-03	1.4474E-01
4	-3.4147E-01	-3.2854E-01	-1.0620E-01	1.2863E-01
5	-1.7552E-01	-1.6827E-01	-7.3309E-02	5.1683E-01*
6	-1.5911E-01	-1.5250E-01	-5.2020E-02	2.1152E-01
7	-1.1925E+00	-1.2161E+00	-4.1777E-01	3.1326E-01
8	1.5859E+00	1.7078E+00	5.8504E-01	3.8069E-01
9	-1.4821E+00	-1.5700E+00	-4.8658E-01	2.1800E-01
10	6.9109E-01	6.7525E-01	2.2465E-01	2.0240E-01
11	-2.6814E+00	-4.0550E+00*	-1.2604E+00	6.1852E-01*
12	2.4023E+00	3.1923E+00*	9.0863E-01	4.1087E-01
13	3.3931E-02	3.2488E-02	1.0568E-02	1.3115E-01
14	-5.4162E-01	-5.2502E-01	-1.7163E-01	1.6062E-01
15	-2.2469E-01	-2.1557E-01	-7.1301E-02	1.6305E-01
16	5.9952E-01	5.8279E-01	1.8803E-01	1.4322E-01

Bod

Věrohodnostní vzdálenosti

i	LD(b)[i]	LD(s ²)[i]	LD(b,s ²)[i]
1	2.5120E-02	2.2351E-02	4.6185E-02
2	8.7766E-02	6.2215E-03	9.1797E-02
3	2.6701E-05	3.2606E-02	3.2632E-02
4	2.2933E-02	2.3569E-02	4.5295E-02
5	4.3879E-02	3.0095E-02	7.1395E-02
6	9.0525E-03	3.0538E-02	3.9045E-02
7	8.4237E-01	3.0991E-02	9.5089E-01
8	1.9389E+00	2.4682E-01	2.6917E+00
9	7.9637E-01	1.5878E-01	1.0958E+00
10	1.6078E-01	4.5425E-03	1.6232E-01
11	1.0861E+01*	7.8276E+00*	4.4183E+01*
12	4.6276E+00	3.4386E+00	1.3130E+01*
13	2.3173E-04	3.2520E-02	3.2738E-02
14	7.4671E-02	1.2499E-02	8.4425E-02
15	1.3108E-02	2.8534E-02	4.0880E-02
16	7.9906E-02	9.2132E-03	8.6631E-02

Indikace vlivných dat

A. Analýza reziduí

Index	Standard.	Jackknife	Predik.	Diag(Hii)	Diag(H ² ii)	Cook. vzd.
1.000	0.366	0.352	0.113	0.124	0.133	0.013
2.000	0.656	0.640	0.205	0.133	0.164	0.025
3.000	-0.011	-0.010	-0.003	0.145	0.145	0.000
4.000	-0.341	-0.329	-0.106	0.129	0.137	-0.013
5.000	-0.176	-0.168	-0.073	0.517	0.518	-0.047
6.000	-0.159	-0.152	-0.052	0.212	0.213	-0.011
7.000	-1.193	-1.216	-0.418	0.313	0.395	-0.136
8.000	1.586	1.708	0.585	0.381	0.510	0.244
9.000	-1.482	-1.570	-0.487	0.218	0.361	-0.103
10.000	0.691	0.675	0.225	0.202	0.234	0.044
11.000	-2.681	-4.055	-1.260	0.619	0.847	-1.087
12.000	2.402	3.192	0.909	0.411	0.694	0.419
13.000	0.034	0.032	0.011	0.131	0.131	0.001
14.000	-0.542	-0.525	-0.172	0.161	0.181	-0.026
15.000	-0.225	-0.216	-0.071	0.163	0.167	-0.011
16.000	0.600	0.583	0.188	0.143	0.169	0.025

B. Analýza vlivu

Index	Atkinson. vzd.	Andrew.-P _i Vliv	na Y ^A	Vliv LD(b)	Vliv LD(s)	Vliv LD(b,s)
1.000	0.229	0.867	0.132	0.025	0.022	0.046
2.000	0.434	0.836	0.251	0.088	0.006	0.092
3.000	0.007	0.855	-0.004	0.000	0.033	0.033
4.000	0.219	0.863	-0.126	0.023	0.024	0.045
5.000	0.301	0.482	-0.174	0.044	0.030	0.071
6.000	0.137	0.787	-0.079	0.009	0.031	0.039
7.000	1.423	0.605	-0.821	0.842	0.031	0.951
8.000	2.319	0.490	-1.339	1.939	0.247	2.692
9.000	1.436	0.639	-0.829	0.796	0.159	1.096
10.000	0.589	0.766	0.340	0.161	0.005	0.162
11.000	6.943	0.153	-5.163	16.861	1.822	44.183
12.000	4.618	0.306	2.666	4.528	3.439	13.130
13.000	0.022	0.869	0.013	0.000	0.033	0.033
14.000	0.398	0.819	-0.230	0.075	0.012	0.084
15.000	0.165	0.833	-0.095	0.013	0.029	0.041
16.000	0.413	0.831	0.238	0.080	0.009	0.087

6. Testy regresního tripletu

(Data + Model + Metoda):

- 6.1 Fisher-Snedecorův test celkové regrese,
- 6.2 Scottovo kritérium multikolinearity,
- 6.3 Cook-Weisbergův test heteroskedasticity,
- 6.4 Jarque-Berrův test normality reziduí,
- 6.5 Waldův test autokorelace,
- 6.6 Znaménkový test reziduí.

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

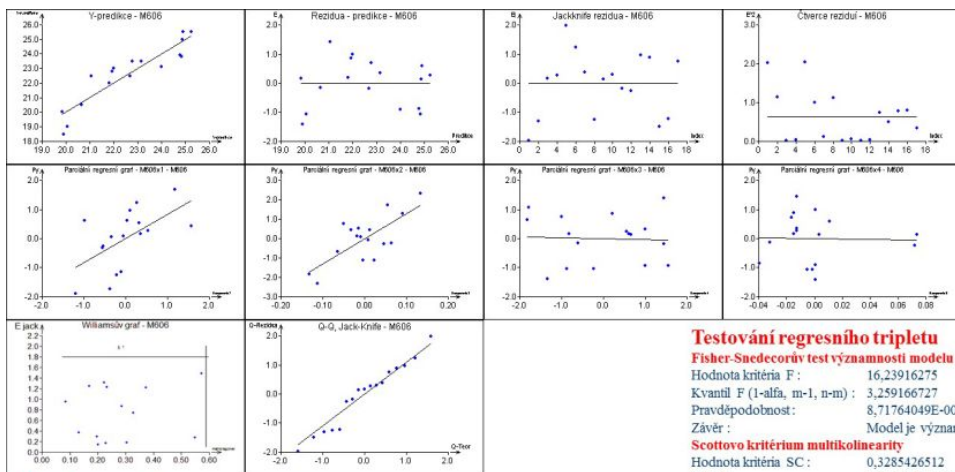
Fisher-Snedecorův test významnosti regrese, F:	1.4050E+03
Tabulkový kvantil, F(1-alpha,m-1,n-m):	3.4903E+00
<i>Závěr:</i> Navržený model je významný.	
Spočtená hladina významnosti:	0.000
Scottovo kritérium multikolinearity, M:	9.6106E-01
<i>Závěr:</i> Navržený model není korektní.	
Cook-Weisbergův test heteroskedasticity, Sf:	5.3739E+00
Tabulkový kvantil, Chi ² (1-alpha,1):	3.8415E+00
<i>Závěr:</i> Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti:	0.020
Jarque-Berraův test normality reziduí, L(e):	3.3085E-02
Tabulkový kvantil, Chi ² (1-alpha,2):	5.9915E+00
<i>Závěr:</i> Normalita je prokázána.	
Spočtená hladina významnosti:	0.984
Waldův test autokorelace, Wa:	1.0320E+01
Tabulkový kvantil, Chi ² (1-alpha,1):	3.8415E+00
<i>Závěr:</i> Rezidua jsou autokorelována.	
Spočtená hladina významnosti:	0.001
Znaménkový test, Dt:	-1.9739E-01
Tabulkový kvantil, N(1-alpha/2):	1.6449E+00
<i>Závěr:</i> Rezidua nevykazují trend.	
Spočtená hladina významnosti:	0.422

Úlohy na výstavbu lineárního regresního modelu

Kritika modelu

Software QC-EXPERT 3.1, ADSTAT 1.25

33



Odhady parametrů

Prom	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-88,98757912	19,28775187	Významný	0,0005966607643	-131,0119804	-46,96317788
M606x1	0,8279111161	0,3387506886	Významný	0,03093003933	0,08983676969	1,565985462
M606x2	12,88100861	3,474346922	Významný	0,002994662667	5,311056959	20,45096026
M606x3	-0,033083552	0,2070700617	Nevýznamný	0,8757201295	-0,4842504594	0,4180833545
M606x4	-0,657626405	7,850213402	Nevýznamný	0,9346191575	-17,76177208	16,44651927

Statistické charakteristiky regrese

Vicenasobný korelační koeficient R :	0,9187317663
Koeficient determinace R ² :	0,8440680585
Předikovaný korelační koeficient Rp :	0,4509785476
Střední kvadratická chyba predikce MEP :	1,334873393
Akaikého informační kritérium :	2,245782063
Reziduální součet čtverců :	10,77342956
Průměr absolutních reziduí :	0,6637999433
Reziduální směr. odchylka :	0,9475155919
Reziduální rozptyl :	0,897785797
Šikmost reziduí :	0,02974295803
Špičatost reziduí :	2,03648524

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	16,23916275
Kvantil F (1-alfa, m-1, n-m) :	3,259166727
Pravděpodobnost :	8,71764049E-005
Závěr :	Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC :	0,3285426512
Závěr :	Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW :	0,4297173633
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,5121276663
Závěr :	Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB :	0,7418605389
Kvantil Chi ² (1-alfa,2) :	5,991464547
Pravděpodobnost :	0,6900920604
Závěr :	Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA :	0,789038755
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,3743909432
Závěr :	Autokorelace je nevýznamná

Úloha M6.06 Vliv čtyř faktorů na koncentraci amoniakálního dusíku

Zadáni: Je vyšetřován vliv teploty x_1 , pH x_2 , koncentrace celkového dusíku x_3 a koncentrace rozpuštěného kyslíku x_4 na koncentraci amoniakálního dusíku y v odtoku z dosazovací nádrže.

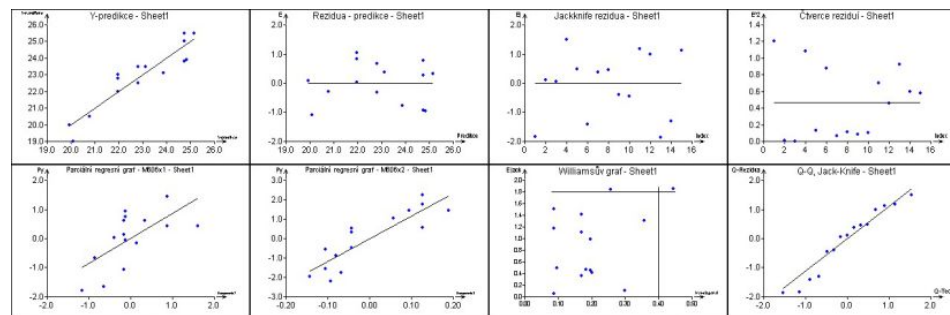
Úkoly:

- (1) Postavte vícerozměrný lineární regresní model a vyšetřete regresní triplet.
- (2) Pomocí parciálních regresních grafů a parciálních reziduálních grafů vyšetřete statistickou významnost jednotlivých faktorů.
- (3) Jsou v datech vlivné body? Je nutné odstranit vybočující hodnoty?
- (4) K čemu slouží znaménkový test navrženého regresního modelu?
- (5) Jak řešíme úlohu v případě porušení předpokladů MNČ, a to především při nalezené heteroskedasticitě v datech, autokorelaci a nenormalitě chyb?

Data: Teplota x_1 , pH x_2 , koncentrace celkového dusíku x_3 , koncentrace rozpuštěného kyslíku x_4 , koncentrace amoniakálního dusíku y :

x_1	x_2	x_3	x_4	y
21	7.2	35	0.1	18.5
...
22.5	7.5	38.5	0.1	25.5

34



Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-82,62448798	12,65926065	Významný	2,823491952E-005	-110,2066475	-55,04232845
M606x1	0,8624507307	0,2864832789	Významný	0,01085334768	0,2382572871	1,486644174
M606x2	11,72768375	1,882604404	Významný	4,387394308E-005	7,625841116	15,82952637

Statistické charakteristiky regrese

Vicenasobný korelační koeficient R :	0,9293375138
Koeficient determinace R ² :	0,8636682145
Předikovaný korelační koeficient Rp :	0,5744221714
Střední kvadratická chyba predikce MEP :	0,8210092009
Akaikého informační kritérium :	-5,571805689

Reziduální součet čtverců :	6,935107041
Průměr absolutních reziduí :	0,5846417807
Reziduální směr. odchylka :	0,7602141716
Reziduální rozptyl :	0,5779255867
Šikmost reziduí :	0,04465331303
Špičatost reziduí :	1,77621562

Testování regresního tripletu

Hodnota kritéria F :	38,01027961
Kvantil F (1-alfa, m-1, n-m) :	3,885293835
Pravděpodobnost :	6,42070528E-006
Závěr :	Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC :	0,1152181523
Závěr :	Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW :	0,0007023910311
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,9788564113
Závěr :	Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB :	1,046889618
Kvantil Chi ² (1-alfa,2) :	5,991464547
Pravděpodobnost :	0,5924760618
Závěr :	Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA :	0,0301759228
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,8629916618
Závěr :	Autokorelace je nevýznamná

35

36

Úloha M6.13 Vliv šesti parametrů na výťažek destilace cyklohexanolu

Zadání: Při studiu destilační kolony byly proměřovány jednotlivé fyzikálně-chemické veličiny, ovlivňující výťažek destilace. Pomocí lineárního regresního modelu diskutujte vliv dále v datech uvedených šesti sledovaných veličin x_1 až x_6 na koncentraci cyklohexanolu y .

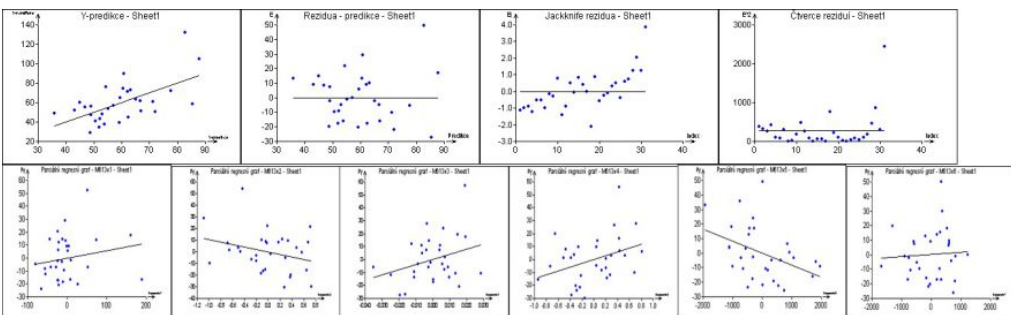
Úkoly:

- (1) Testujte statistickou významnost jednotlivých regresních parametrů.
- (2) Jsou v datech vlivné body?
- (3) Vysvětlete všech sedm předpokladů MNČ a řešení regresním tripletem.
- (4) Která kritéria jsou nejvhodnější při hledání lineárního regresního modelu?
- (5) Které z následujících kritérií je nejvýhodnější: střední kvadratická chyba predikce MEP, Akaiikovo informační kritérium AIC a predikovaný koeficient determinace R^2_p .

Data: Koncetrace cyklohexanolu v surovině x_1 [ppm], teplota na hlavě kolony x_2 [°C], tlak na hlavě kolony x_3 [atm], teplota na patě kolony x_4 [°C], reflux x_5 [kg/h], odtah x_6 [kg/h], koncetrace výsledného cyklohexanolu v produktu y [ppm]:

x_1	x_2	x_3	x_4	x_5	x_6	y
80	155	0.53	165	10000	9000	29
...
1760	156	0.52	166	9200	8000	327

37



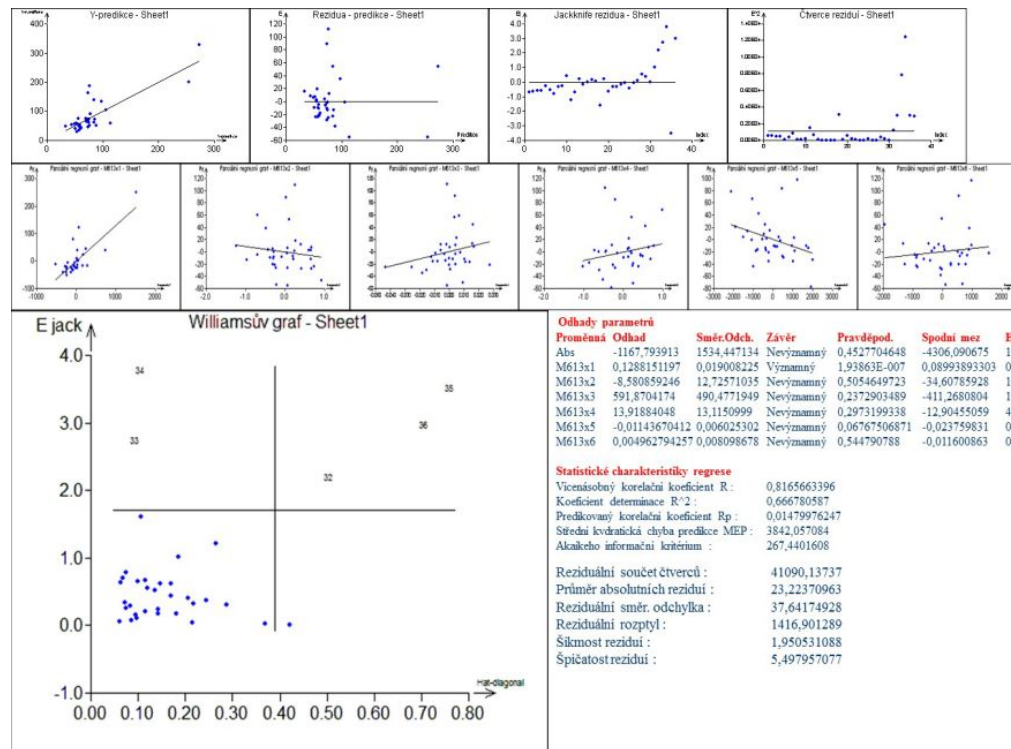
Proměnná	Odhad	Směr.Odch.	Závěr	Pravděp.	Spodní mez	Horní mez
Abs	-856,2810978	813,7530106	Nevýznamný	0,3031593342	-2535,784765	823,2225698
M613a1	0,05647144576	0,06013711173	Nevýznamný	0,3027750838	-0,06764545261	0,1805883441
M613a2	-10,64506019	7,144869871	Nevýznamný	0,149259051	-25,39188065	4,100674264
M613a3	402,4811795	251,653262	Nevýznamný	0,12285965	-116,9596268	921,8579408
M613a4	14,58406483	7,707453661	Nevýznamný	0,07058853925	-1,323337693	30,49146735
M613a5	-0,008191937301	0,003752121024	Významný	0,03902651542	-0,01593593468	-0,000447940319
M613a6	0,001416103344	0,005057927103	Nevýznamný	0,7634777952	-0,009022945127	0,01185515181

Statistické charakteristiky regrese
 Vícenásobný korelační koeficient R : 0,5937268088
 Koeficient determinace R² : 0,3525103735
 Predikovaný korelační koeficient Rp : 0,09219080283
 Střední kvadratická chyba predikce MEP : 556,6401512
 Akaiikovo informační kritérium : 188,2855383

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděp.	Spodní mez	Horní mez
Abs	-810,5994832	631,3077229	Nevýznamný	0,2059300162	-2103,899626	481,1206592
M613a1	0,08185267186	0,0617484583	Nevýznamný	0,1989837035	-0,046490375	0,2101957197
M613a2	-9,503955839	5,79638685	Nevýznamný	0,1159766152	-21,55820218	2,550206503
M613a3	311,5899268	210,6388682	Nevýznamný	0,1539138032	-126,4575797	749,6374352
M613a4	12,53007956	6,246202983	Významný	0,0419742732	0,5391041696	36,52185497
M613a5	-0,007050508875	0,002873653914	Významný	0,0129404371	-0,013781099	-0,00182891
M613a6	0,000667046302	0,004002935859	Nevýznamný	0,8692480634	-0,007657520	0,008991612

Statistické charakteristiky regrese
 Vícenásobný korelační koeficient R : 0,6164290997
 Koeficient determinace R² : 0,3799848349
 Predikovaný korelační koeficient Rp : 0,036931694
 Střední kvadratická chyba predikce MEP : 286,1171281
 Akaiikovo informační kritérium : 154,072977

39



Proměnná	Odhad	Směr.Odch.	Závěr	Pravděp.	Spodní mez	Horní mez
Abs	-1167,793913	1534,447134	Nevýznamný	0,4527704648	-4306,090675	1970,502849
M613a1	0,1288151197	0,019008225	Významný	1,93863E-007	0,0899389303	0,1676913064
M613a2	-8,580859246	12,72571035	Nevýznamný	0,5054649723	-34,60785928	17,4614079
M613a3	591,8704174	490,4771949	Nevýznamný	0,2372903489	-411,2680804	1595,008915
M613a4	13,91884048	13,1150999	Nevýznamný	0,2973199338	-12,90455059	40,74223155
M613a5	-0,01143670412	0,006025302	Nevýznamný	0,06767596871	-0,023759831	0,000886422
M613a6	0,004962794257	0,008098678	Nevýznamný	0,544790788	-0,011600863	0,021526451

Statistické charakteristiky regrese
 Vícenásobný korelační koeficient R : 0,8165663396
 Koeficient determinace R² : 0,666780587
 Predikovaný korelační koeficient Rp : 0,01479976247
 Střední kvadratická chyba predikce MEP : 3842,057084
 Akaiikovo informační kritérium : 267,4401608

Reziduální součet čtverců : 41090,13737
 Průměr absolutních reziduí : 23,22370963
 Reziduální směr. odchylka : 37,64174928
 Reziduální rozptyl : 1416,901289
 Šikmost reziduí : 1,950531088
 Špičatost reziduí : 5,497957077

40