

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

REGRESNÍ DIAGNOSTIKA

Vyšetřuje **regresní triplet**, což představuje

- ◆ **Kritiku dat** (zkoumá kvalitu dat pro navržený model)
- ◆ **Kritiku modelu** (zkoumá kvalitu modelu pro daná data)
- ◆ **Kritiku metody odhadu** (prověřuje splnění všech předpokladů požadovaných metodou MNČ)

1. KRITIKA DAT

Výskyt *vlivných bodů* (VB):

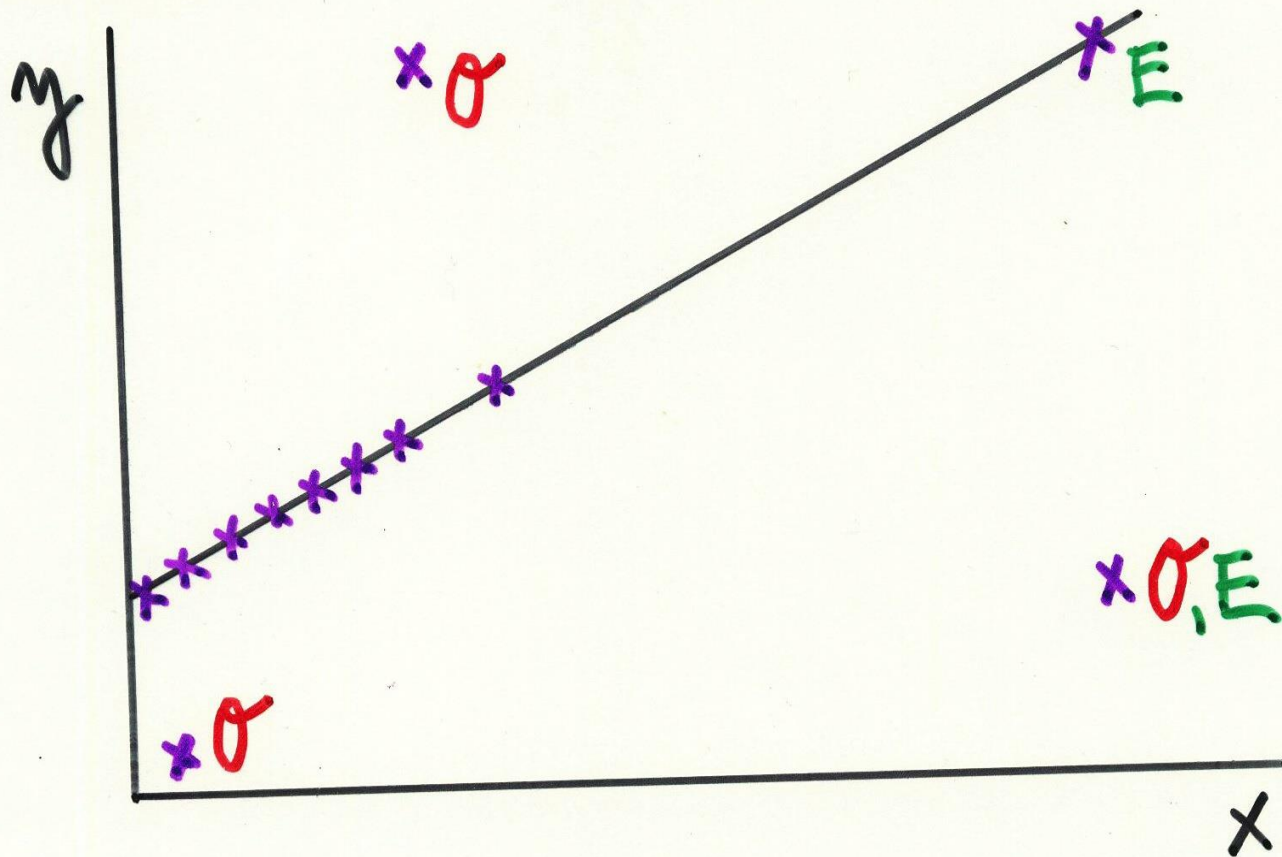
- zdrojem řady problémů,
- zkreslení odhadů a růst rozptylů odhadů parametrů.

Dělení *vlivných bodů* dle charakteru:

1. Hrubé chyby: vybočující pozorování, ~~extrémy~~.
2. Body s vysokým vlivem (tzv. golden points): rozšiřují predikční schopnosti modelu.
3. Zdánlivě vlivné body: důsledek nesprávného regresního modelu.

Dělení vlivných bodů dle výskytu:

1. vybočující pozorování (outliers, O): na y se liší,
2. extrémy (high leverage points, E): liší se na ose x



Statistická analýza reziduí

1. Klasická rezidua: $\hat{e}_i = y_i - \mathbf{x}_i \mathbf{b}$,

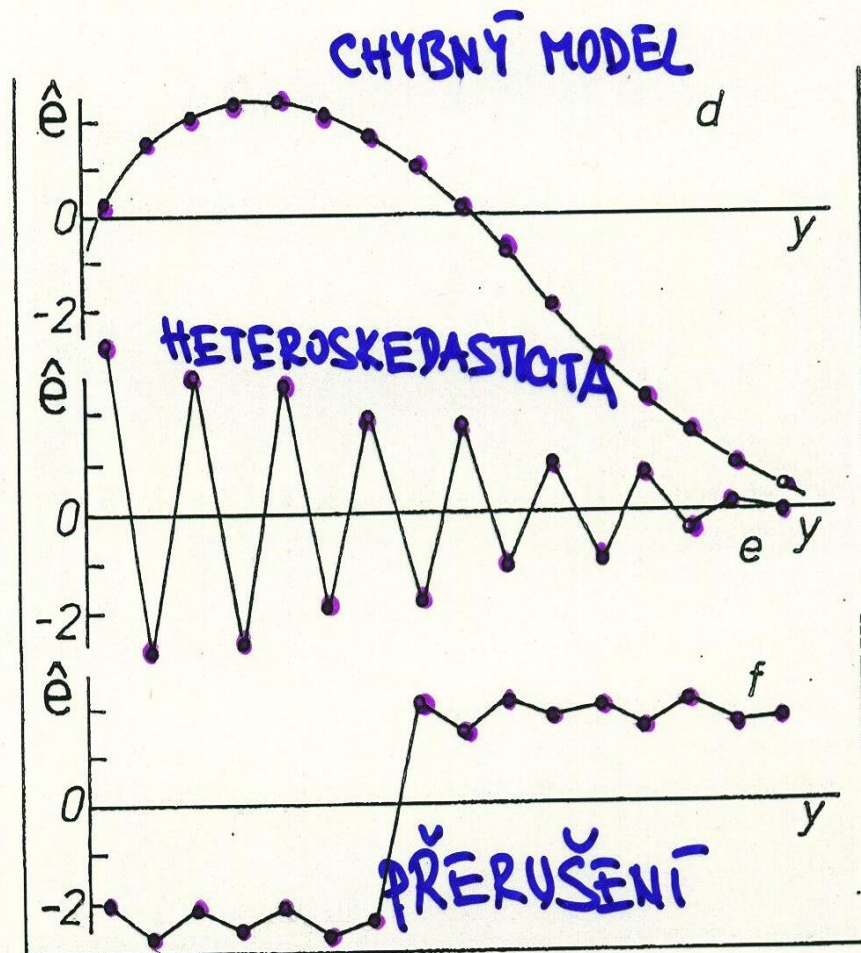
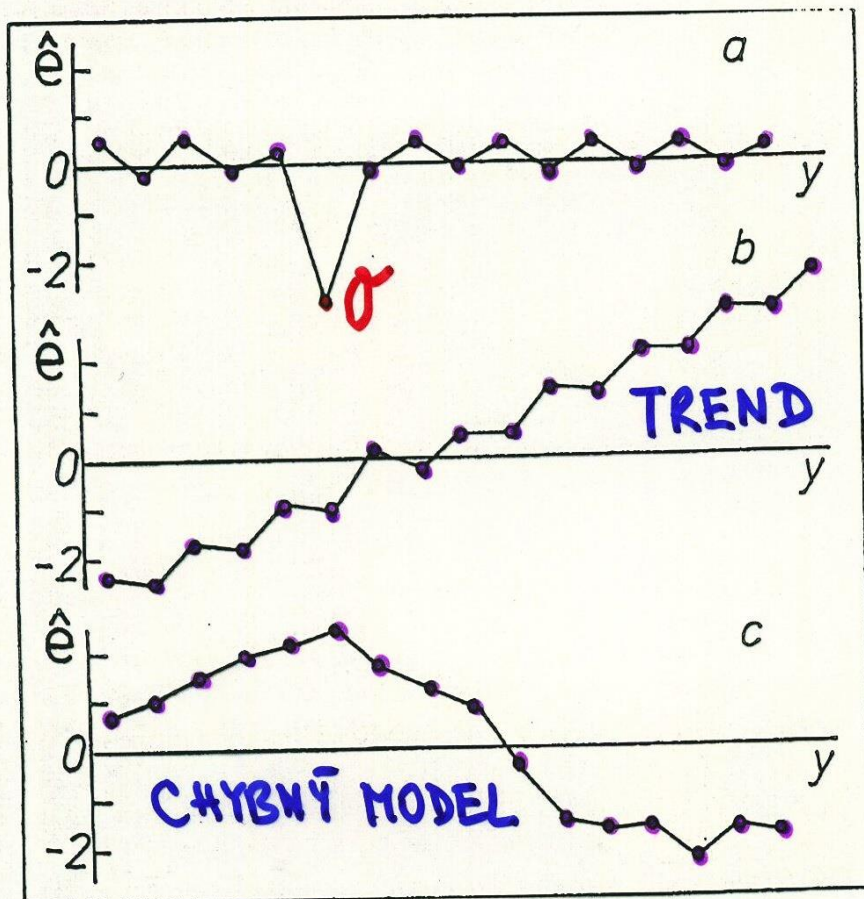
Nesprávné představy o reziduích:

1. Rozdělení reziduí je stejné jako rozdělení chyb,
2. Vlastnosti reziduí jsou shodné s vlastnostmi chyb,
2. Čím je reziduum \hat{e}_i větší, tím je bod vlivnější, a tím spíše by se měl z dat vyloučit.

$$\hat{\mathbf{e}} = \mathbf{P} \mathbf{y} = \mathbf{P} (\mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{P} \boldsymbol{\varepsilon} = (\mathbf{E} - \mathbf{H}) \boldsymbol{\varepsilon}$$

Statistická analýza klasických reziduí

1. Grafická analýza



2. Numerická analýza:

- (a) **Střední hodnota reziduí** $E(\hat{\epsilon}) = 0$,
- (b) **Průměrné reziduuum** $|\bar{\epsilon}| = \frac{1}{n} \sum |\hat{\epsilon}_i| \approx \epsilon$,
- (c) **Směrodatná odchylka střední hodnoty reziduí** $s(\hat{\epsilon}) \approx \epsilon$,
- (d) **Koeficient šikmosti souboru reziduí** $g_1(\hat{\epsilon}) \approx 0$,
- (e) **Koeficient špičatosti souboru reziduí** $g_2(\hat{\epsilon}) \approx 3$,
- (f) **Pearsonův χ^2 - test dobré shody:** χ^2_{exp} vs. $\chi^2_{\text{crit.}}$,
- (g) **Hamiltonův R-faktor relativní těsnosti:** $R \approx 0.5 \%$.

1. Vzorová úloha J625 na výstavbu jednorozměrného lineárního regresního modelu

Jak napsat a vysvětlit úlohu v semestrální práci je v
Kompendiu, str. 583

Návod k sestavení a napsání semestrální práce k zadané úloze najdete ve cvičebnici *M. Meloun, J. Militký: Kompendium statistického zpracování dat, Karolinum Praha 1992, ISBN 978-80-246-2196-8*, ve vzorovém příkladu na str. 583 – 593 a obecný výklad outputu v lineární regresi na str. 582 - 583. V ADSTATu používejte na přenos výstupů a obrázků **PrintScreen** a upravte pak negativ na černobílý pozitiv ve MS Windows.

Úloha J6.25 Závislost celkového cholesterolu v krvi na denní spotřebě tuku

Zadání: U vzorku 20 Američanů byla v analýze krve sledována denní spotřeba tuku ve stravě x v gramech a obsah celkového cholesterolu y v mg na 100 ml krve. Byl navržen jednoduchý lineární regresní model $y = \beta_0 + \beta_1 x$.

Úkoly: Ukažte

- (1) Dokažte platnost navrženého regresního modelu a existenci vlivných bodů.
- (2) Testujte statistickou významnost obou parametrů, úseku β_0 a směrnice β_1 .
- (3) Sestrojte 95%ní interval spolehlivosti úseku β_0 a vysvětlete fakt, že $\beta_0 = 0$.
- (4) Sestrojte 95 % oboustranný interval spolehlivosti směrnice β_1 .
- (5) Nalezněte 95 % interval spolehlivosti celkového cholesterolu u lidí, kteří denně spotřebují 50 g tuku.
- (6) Jaká je korelace mezi celkovým cholesterolem v krvi y a denní spotřebou tuku x u sledovaných jedinců?

Data: Denní spotřeba tuku x [g], obsah celkového cholesterolu v krvi y [mg/100 ml]:

x	y
21	130
...	...
134	271

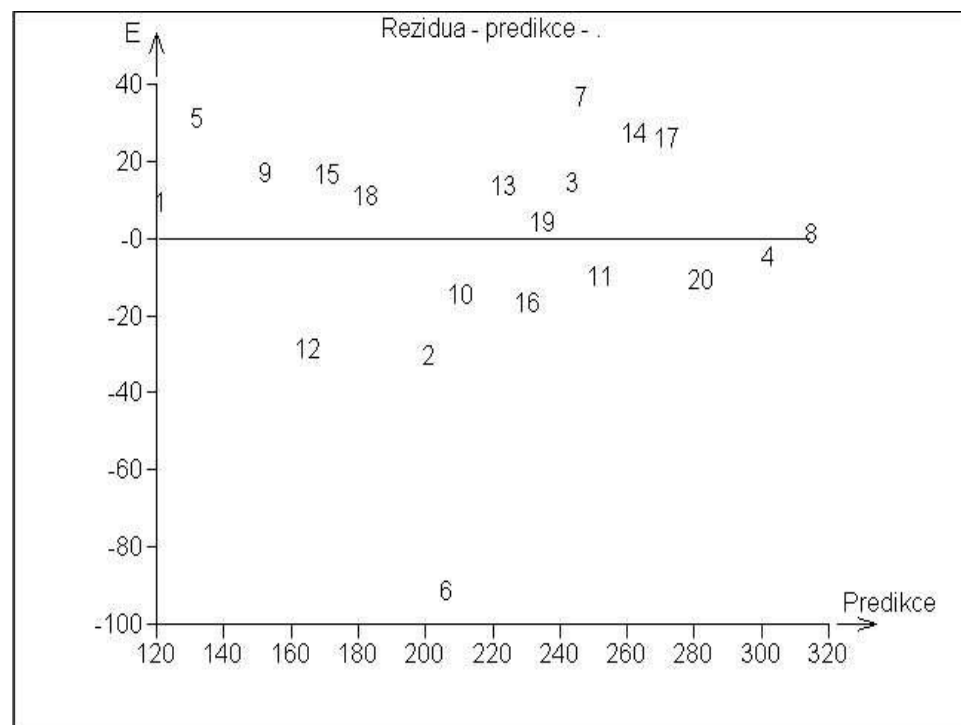
Reziduum

9.324031
-30.597716
14.587063
-4.927073
30.906639
-91.306412
36.732715
1.228361
16.926202
-14.587934
-9.975981
-28.918364
13.5675
27.0338
16.372939
-16.568371
26.043582
10.955547
4.150107
-10.946636

Reziduum [%Y]

7.172332
-17.998656
5.6539
-1.658947
18.961128
-79.39688
12.979758
0.388722
10.015504
-7.442823
-4.122306
-21.263503
5.724683
9.354256
8.755583
-7.742229
8.739457
5.676449
1.736447
-4.039349

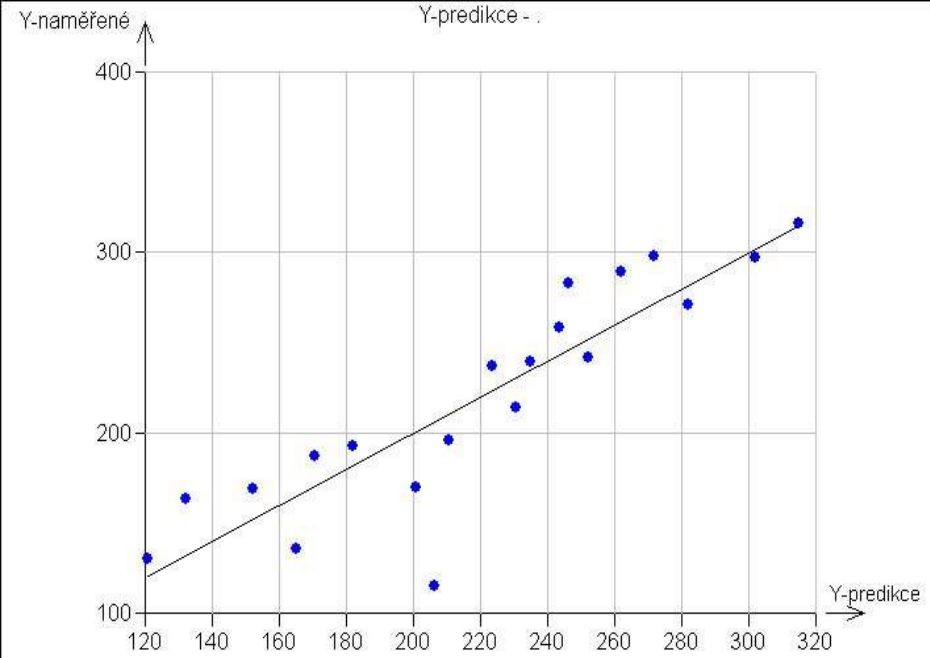
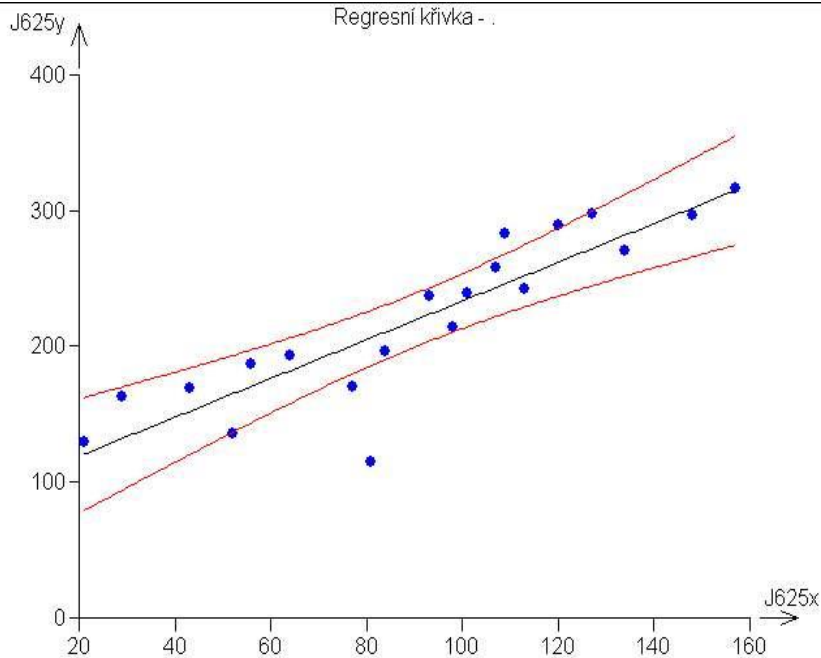
Grafická analýza klasických reziduí:



Numerická analýza klasických reziduí:

Reziduální součet čtverců: 15731.026573
Průměr absolutních reziduí: 20.782849
Reziduální směr. odchylka: 29.562576
Reziduální rozptyl: 873.945921
Šikmost reziduí: 2.450042
Špičatost reziduí: 6.078201

Software QC-EXPERT 3.1 (TriloByte)



Odhady parametrů

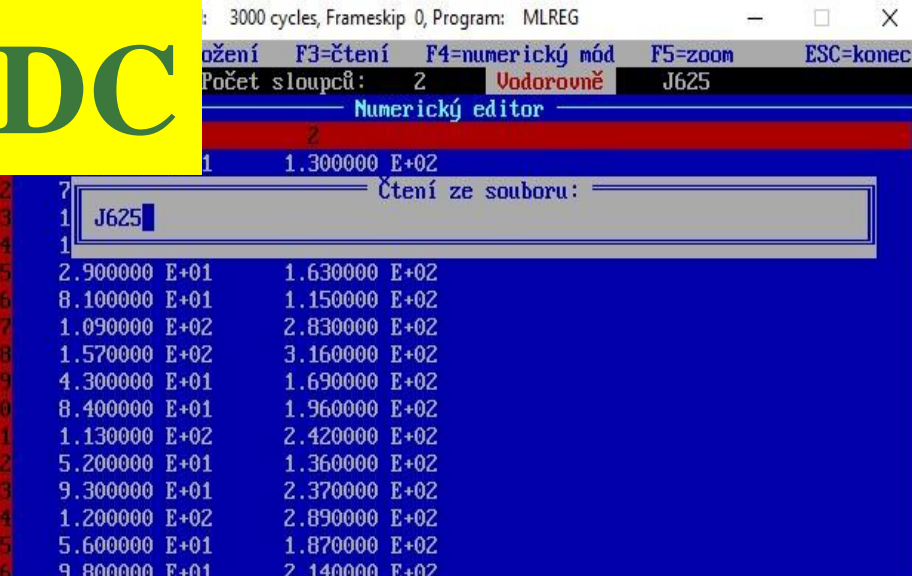
Proměnná	Odhad	Směr.odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	90.705314	17.420915	Významný	0.000059	54.10533	127.305298
J625x	1.427174	0.177707	Významný	0.0	1.053825	1.800523

Statistické charakteristiky regrese

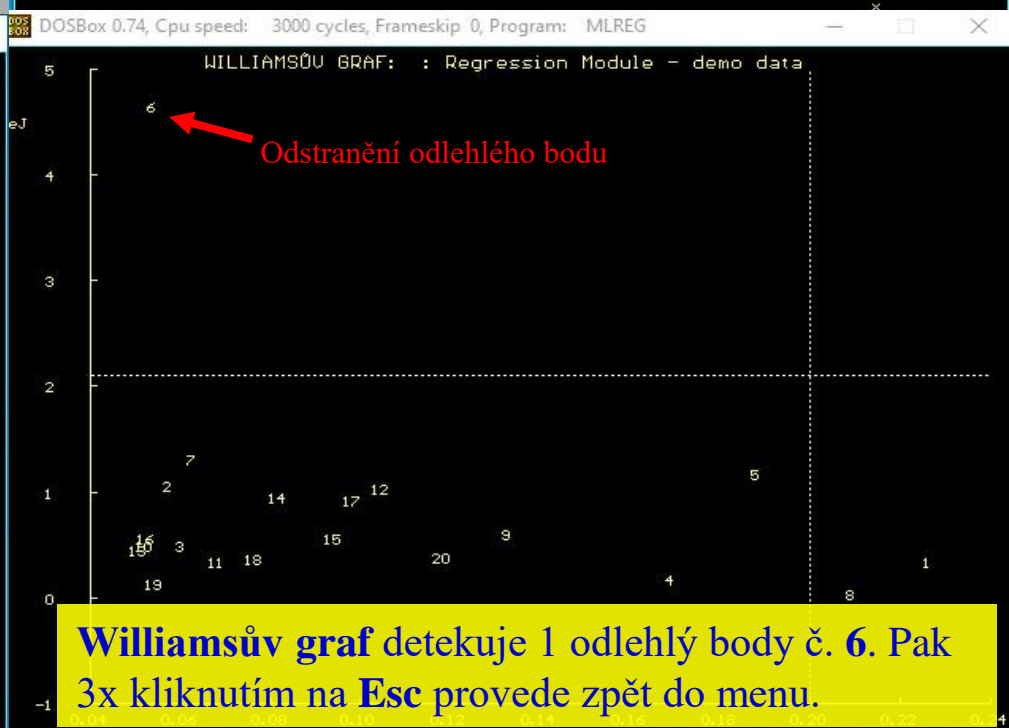
Vícenásobný korelační koeficient R :	0.884201
Koeficient determinace R ² :	0.781812
Predikovaný korelační koeficient Rp :	0.555095
Střední kvadratická chyba predikce MEP :	919.087734
Akaikeho informační kritérium :	137.35316

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu	
Hodnota kritéria F :	64.497725
Kvantil F (1-alfa, m-1, n-m) :	4.413873
Pravděpodobnost :	0.0
Závěr :	Model je významný



Otevřít **Data**, pak stisk **F3**, pak smazat *demo.dat*, napsat **J625** a ukončit **Esc a A**.



2. Normovaná rezidua $\hat{\sigma}_N$

$$\hat{e}_{Ni} = \hat{e}_i / \hat{\sigma}$$

normálně rozdělené veličiny $\hat{e}_{Ni} \sim N(0, 1)$.

Diagnostika: pravidlo 3σ , tj. rezidua větší než $\pm 3\hat{\sigma}$ indikují vybočující.

3. Standardizovaná rezidua $\hat{\sigma}_S$

$$\hat{e}_{Si} = \frac{\hat{e}_i}{\hat{\sigma} \sqrt{1 - H_{ii}}}$$

Mají konstantní rozptyl. Maximální hodnota \hat{e}_{Si} je ohraničena velikostí $\sqrt{n - m}$.

Diagnostika: k indikaci heteroskedasticity.

4. Jackknife rezidua $\hat{\sigma}_J$ ("plně studentizovaná")
užijeme místo $\hat{\sigma}$ odhadu směrodatné odchylky $\hat{\sigma}_{(-i)}$,

$$\hat{e}_{Ji} = \hat{e}_{Si} \sqrt{\frac{n - m - 1}{n - m - \hat{e}_{Si}^2}} = \sqrt{n - m} \cotg \Theta_i$$

mají Student. rozdělení s $(n - m - 1)$ stupni volnosti.
Diagnostika: k identifikaci vybočujících bodů
(outliers).

5. Predikovaná rezidua $\hat{\sigma}_P$

$$\hat{e}_{Pi} = y_i - x_i \mathbf{b}_{(i)} = \frac{\hat{e}_i}{1 - H_{ii}}$$

kde $\mathbf{b}_{(i)}$ jsou MNČ odhady ze všech bodů kromě i -tého.

Diagnostika: indikace vybočujících hodnot (outliers).

6. Rekurzivní rezidua $\hat{\sigma}_R$

Dopředná rekurzivní rezidua jsou definována vztahy

$$\hat{e}_{Ri} = 0, \quad i = 1, \dots, m$$

$$\hat{e}_{Ri} = \frac{y_i - \mathbf{x}_i \mathbf{b}_{i-1}}{\sqrt{1 + \mathbf{x}_i (\mathbf{X}_{i-1}^T \mathbf{X}_{i-1})^{-1} \mathbf{x}_i^T}} \quad i = m + 1, \dots, n$$

kde \mathbf{b}_{i-1} jsou odhady získané z prvních $(i - 1)$ bodů.

Diagnostika: umožňují identifikovat autokorelaci či nestabilitu modelu, např. v čase.

Rezidua k diagnostice:

1. K detekci heteroskedasticity: standard. rezidua $\hat{\epsilon}_{Si}$.
2. K detekci vybočujících bodů: Jackknife rezidua $\hat{\epsilon}_{Ji}$,
nebo predikovaná rezidua $\hat{\epsilon}_{Pi}$.
3. K detekci autokorelace: rekurzivní rezidua $\hat{\epsilon}_{Ri}$.

A. Analýza rozličných druhů reziduí číselně v tabulce

Index	Standardní	Jackknife	Predikované	Diag(Hii)	Diag(H*ii)	Cookova vzdál.
1.0	0.358396	0.349548	12.039487	0.225546	0.231072	0.052188
2.0	-1.065714	-1.069995	-32.439711	0.056782	0.116296	-0.032078
3.0	0.508826	0.498085	15.511562	0.059601	0.073127	0.016124
4.0	-0.18279	-0.177805	-5.926527	0.168641	0.170184	-0.018539
5.0	1.159883	1.171846	38.041795	0.187561	0.248283	0.133886
6.0	-3.174503	-4.650163	-96.457219	0.0534	0.583363	-0.089541
7.0	1.283018	1.308118	39.164901	0.062101	0.147874	0.042476
8.0	0.046714	0.045401	1.552602	0.208837	0.208933	0.006165
9.0	0.614627	0.603678	19.505109	0.132217	0.150429	0.046823
10.0	-0.506712	-0.495985	-15.381984	0.051622	0.06515	-0.013791
11.0	-0.349541	-0.340852	-10.703492	0.067969	0.074296	-0.012745
12.0	-1.03349	-1.035566	-32.279234	0.104119	0.157279	-0.060056
13.0	0.470912	0.460489	14.284453	0.050191	0.061893	0.012442
14.0	0.95392	0.951404	29.417222	0.081021	0.127479	0.042051
15.0	0.581705	0.570705	18.061898	0.093509	0.110551	0.030003
16.0	-0.575594	-0.564597	-17.475813	0.051926	0.069376	-0.015763
17.0	0.92739	0.923596	28.860813	0.097614	0.140731	0.05016
18.0	0.385478	0.376173	11.853575	0.07576	0.08339	0.015799
19.0	0.144322	0.140337	4.386234	0.053834	0.054928	0.004106
20.0	-0.394223	-0.384781	-12.407616	0.117749	0.125366	-0.026307

QC-EXPERT 3.1 (TriloByte)

Jackknife rezid.

Predikované rezid.

0.349548

12.039487

-1.069995

-32.439711

0.498085

15.511562

-0.177805

-5.926527

1.171846

38.041795

-4.650163*outlier

-96.457219*outlier

1.308118

39.164901

0.045401

1.552602

0.603678

19.505109

-0.495985

-15.381984

-0.340852

-10.703492

-1.035566

-32.279234

0.460489

14.284453

0.951404

29.417222

0.570705

18.061898

-0.564597

-17.475813

0.923596

28.860813

0.376173

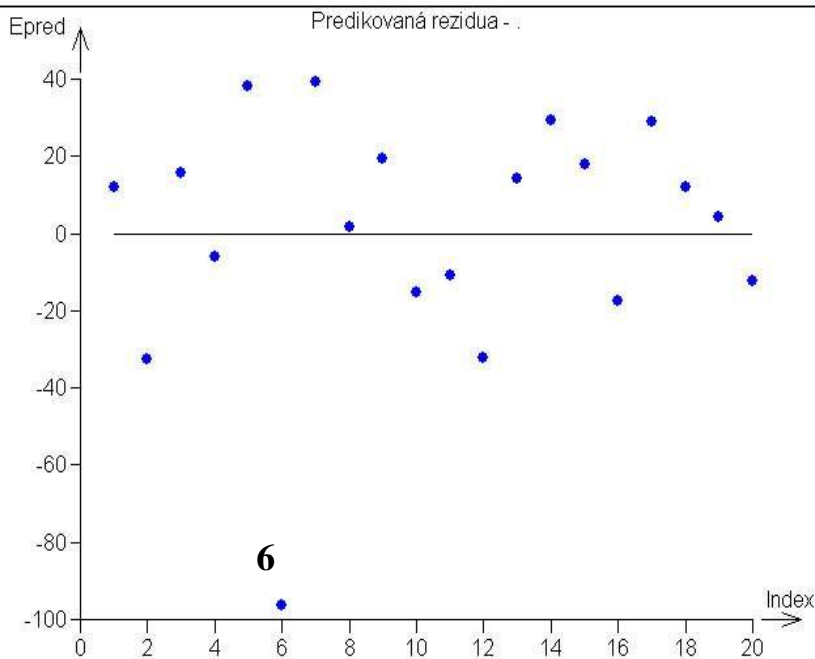
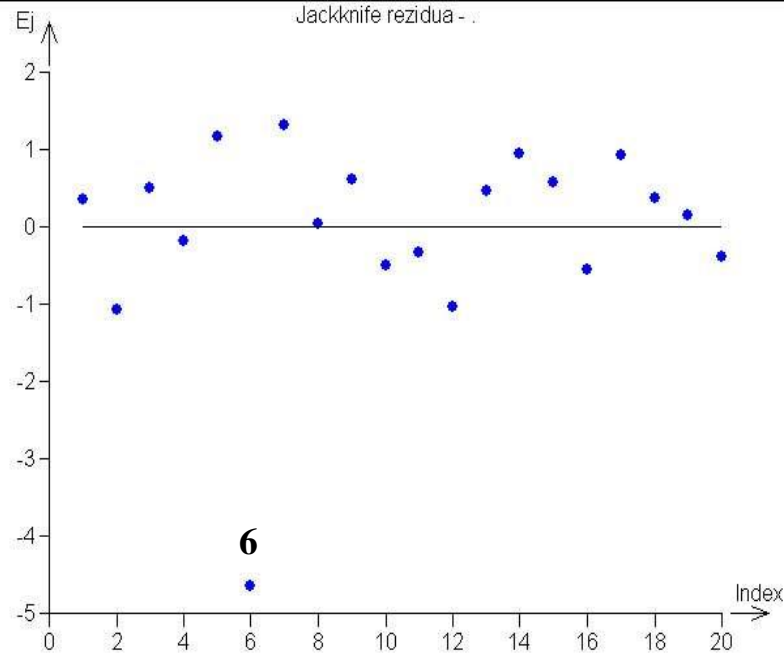
11.853575

0.140337

4.386234

-0.384781

-12.407616

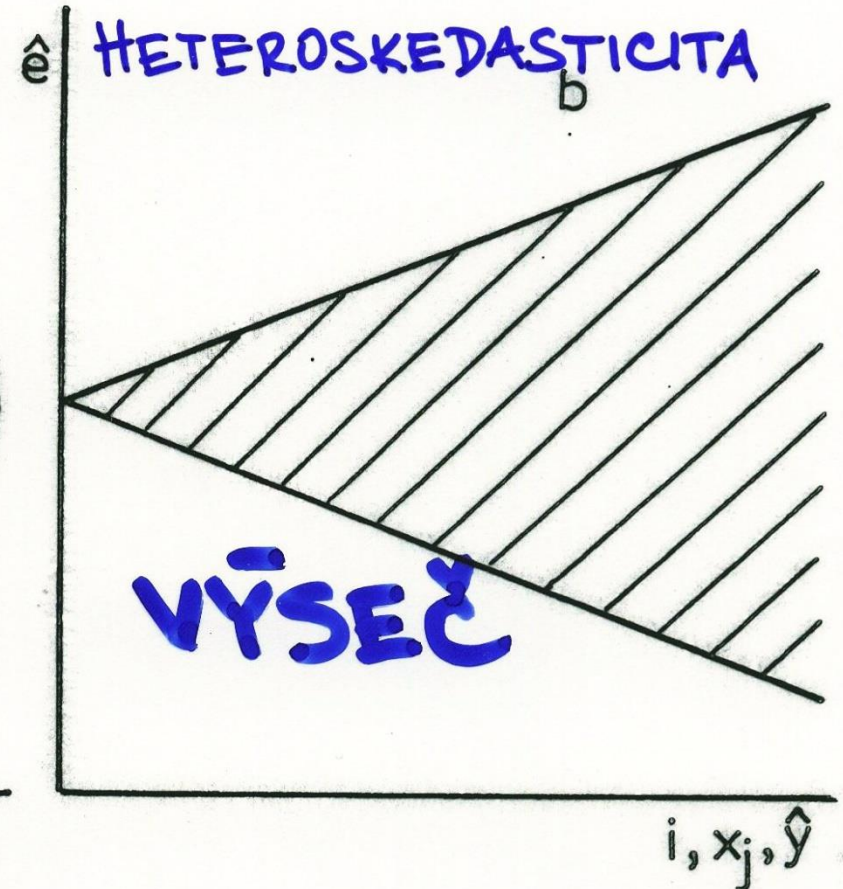
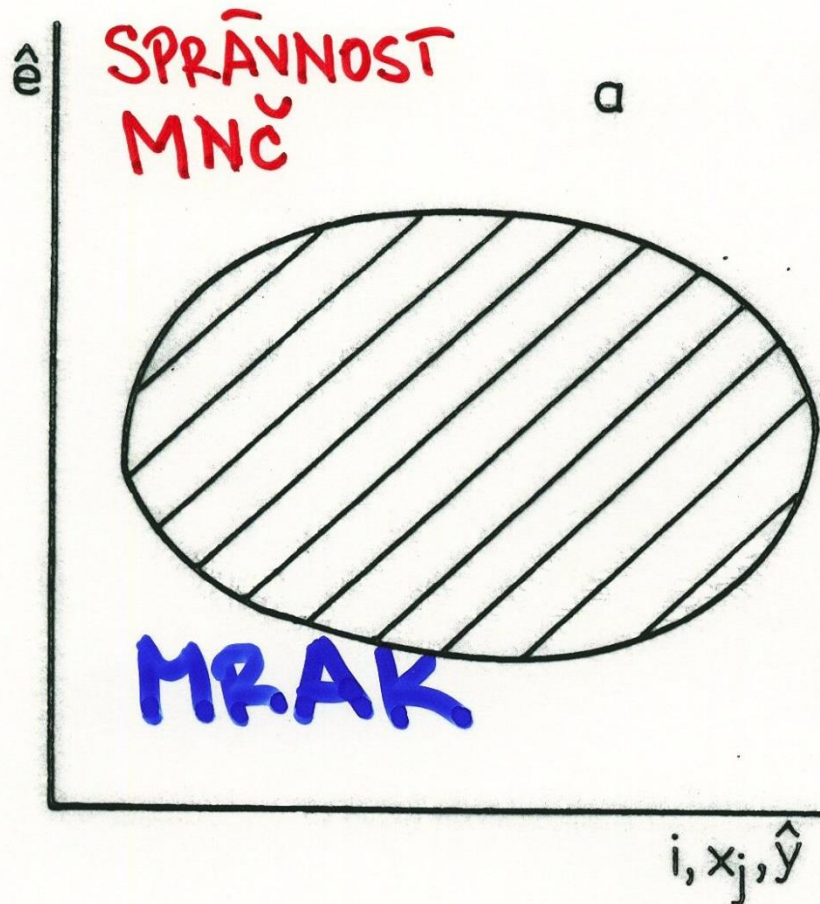


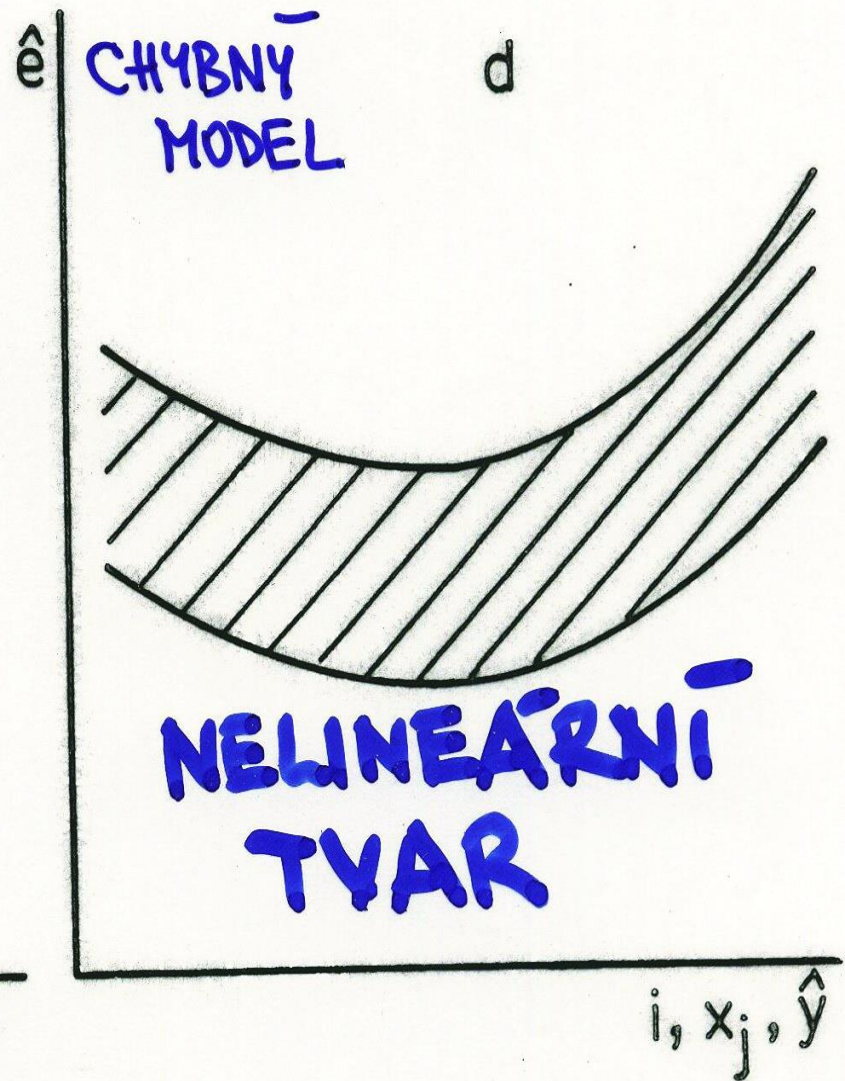
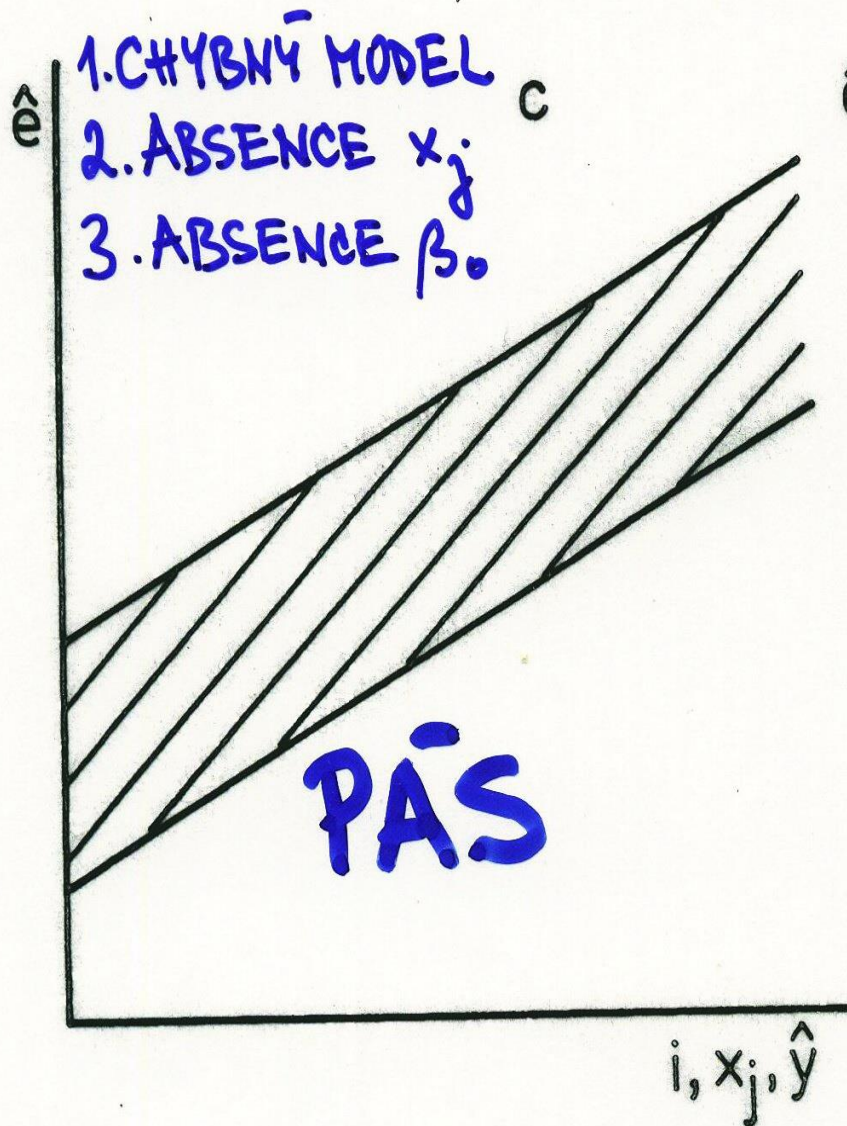
Obrazce v diagnostických grafech

Typ I: Graf reziduí \hat{e}_i proti indexu i .

Typ II: Graf reziduí \hat{e}_i proti proměnné x_j .

Typ III: Graf reziduí \hat{e}_i proti predikci \hat{y}_i .



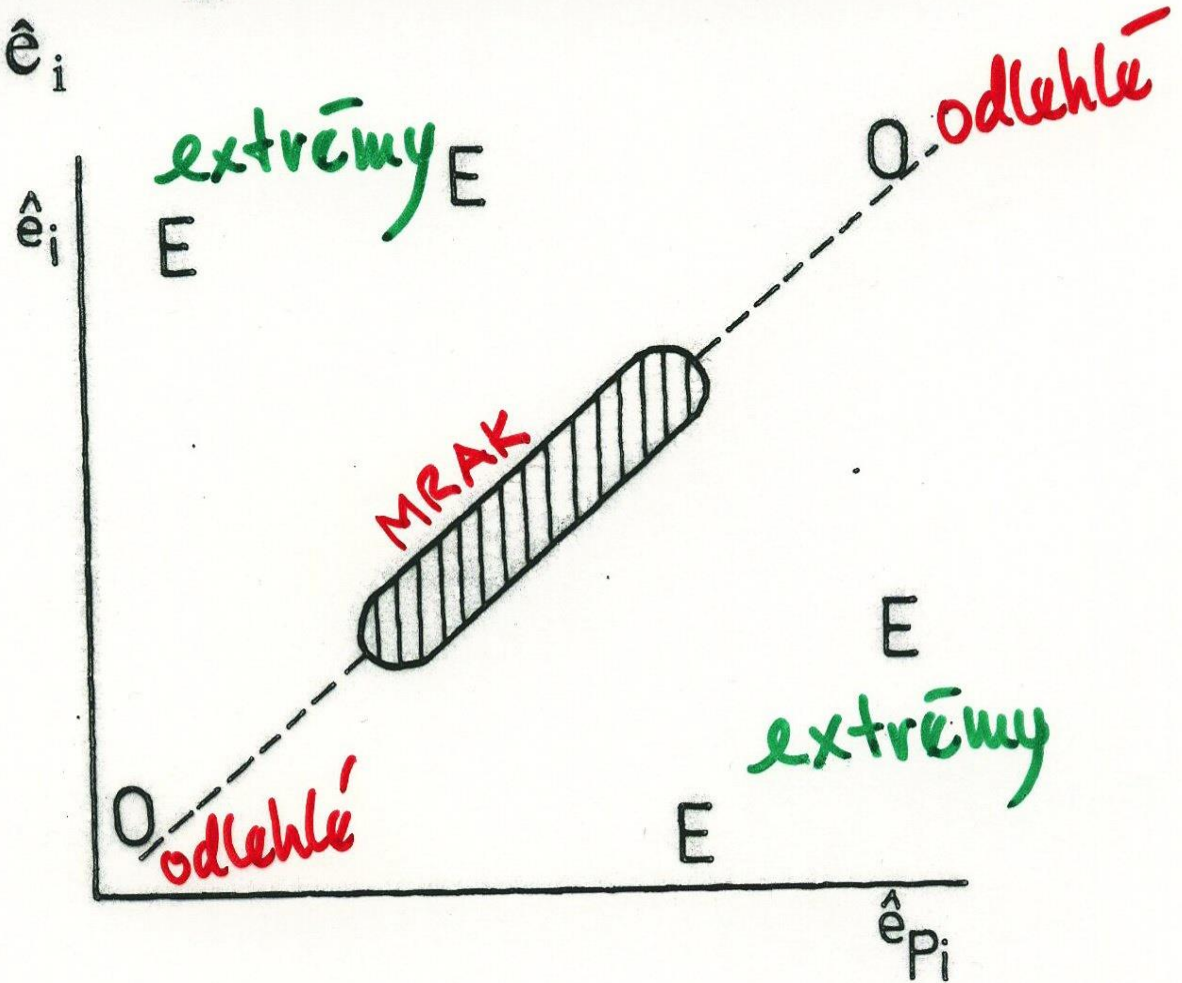


(a) tvar mraku, (b) tvar výseče, (c) tvar pásu a (d) nelineární tvar

Grafy identifikace vlivných bodů

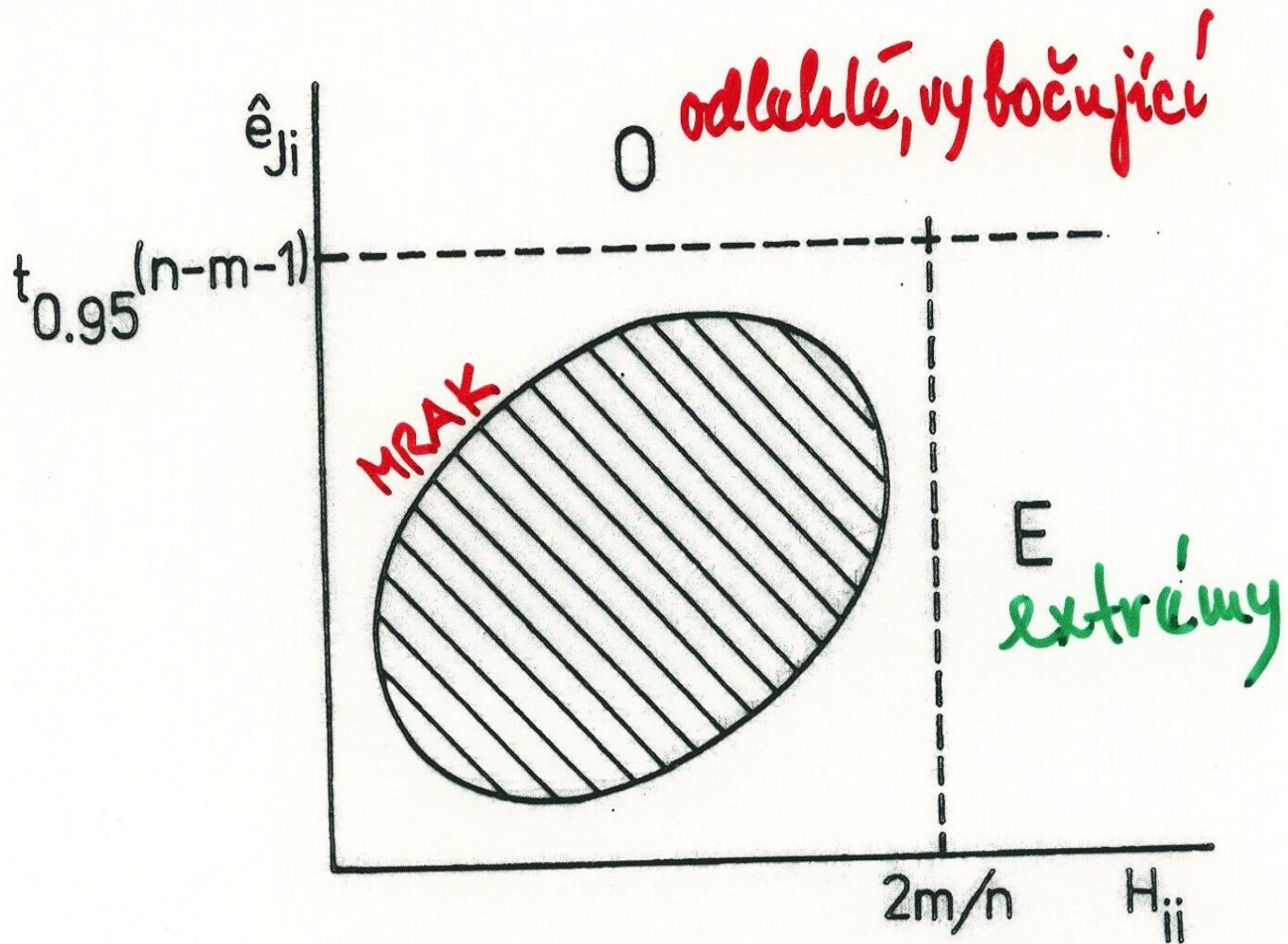
1. Graf predikovaných reziduí (GPR),

osa x: \hat{e}_{Pi} , osa y: \hat{e}_i

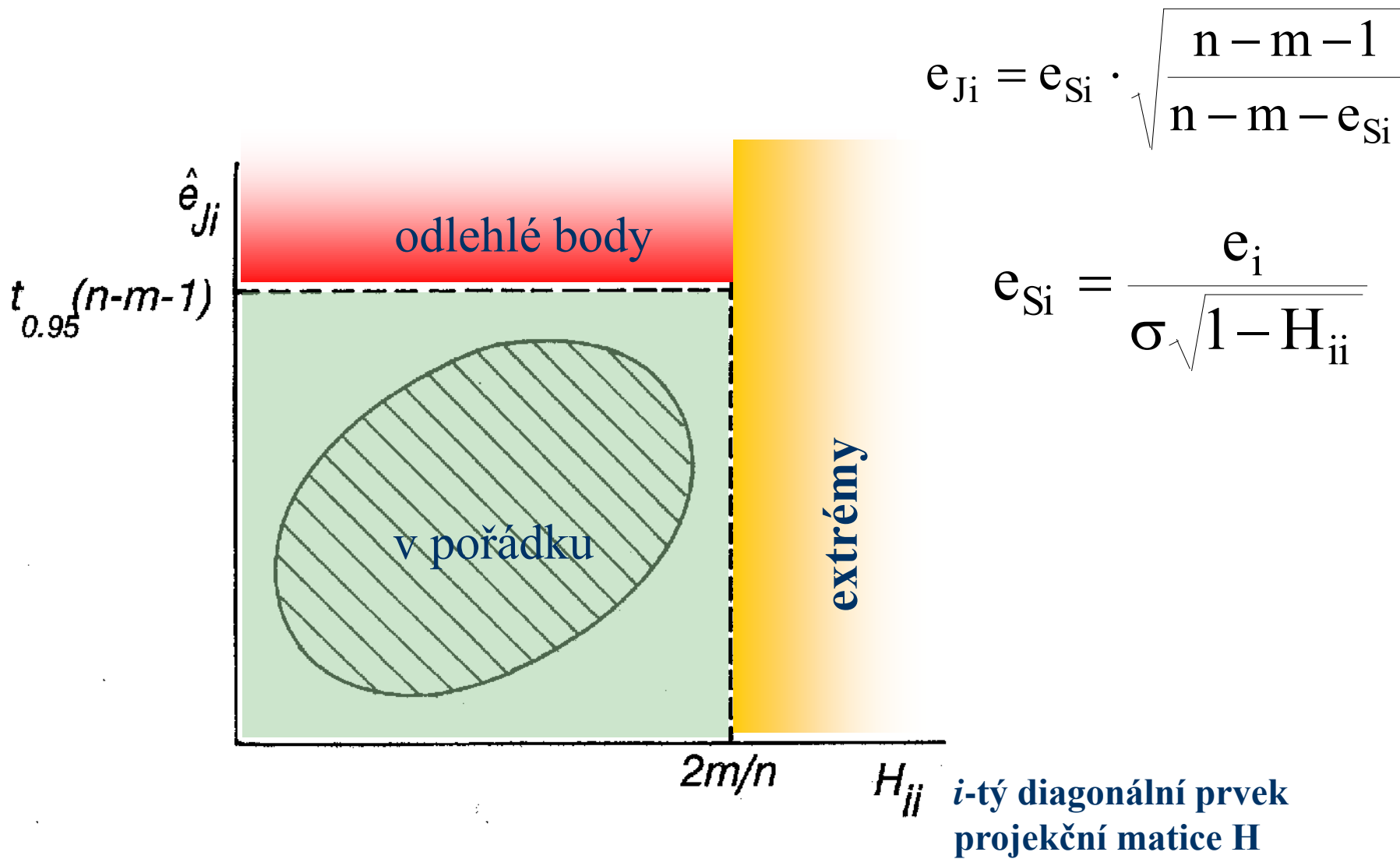


2. Williamsův graf (WG),

osa x: prvky H_{ii} , osa y: \hat{e}_{ji}



Williamsův graf vlivných bodů



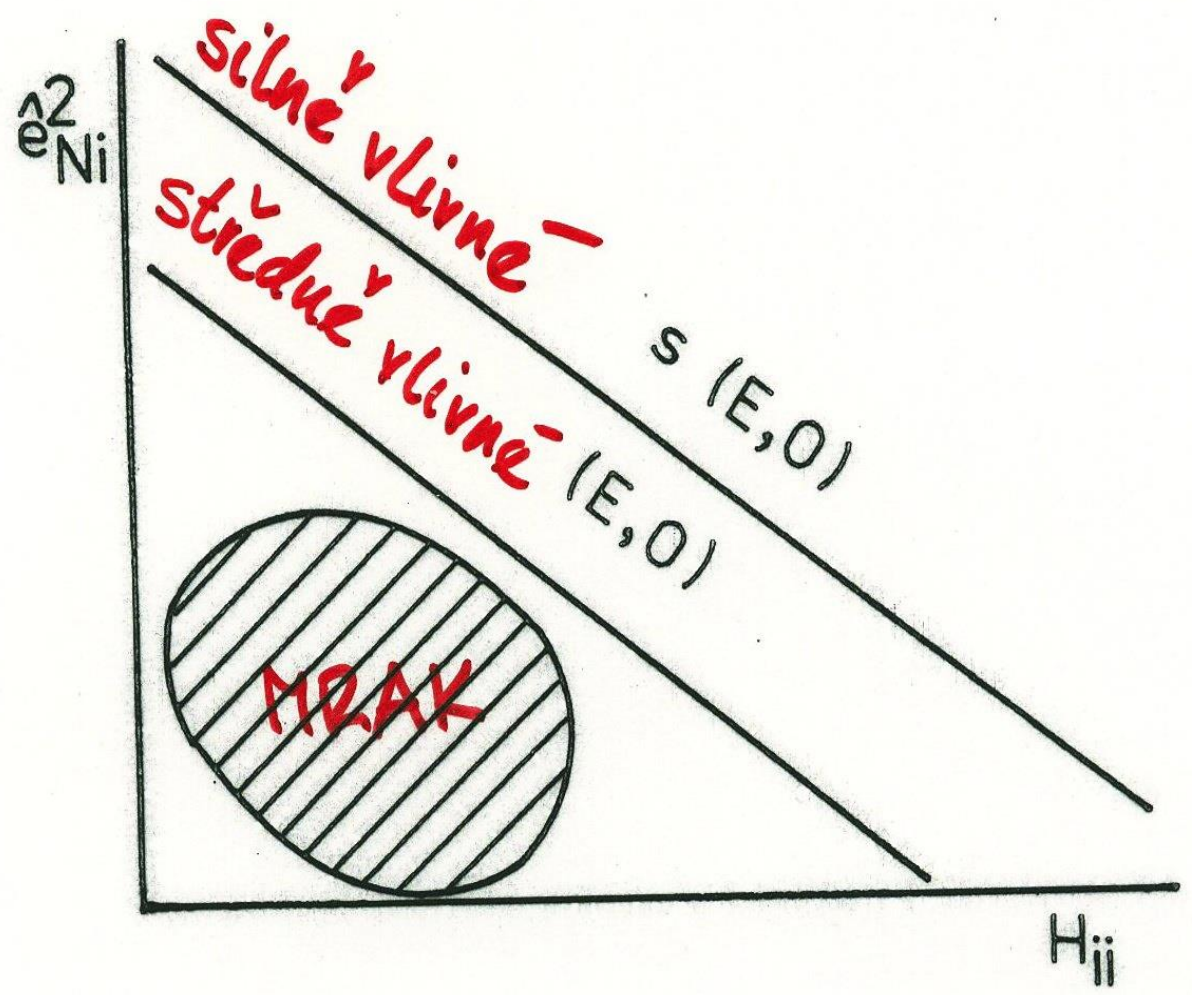
$$e_{Ji} = e_{Si} \cdot \sqrt{\frac{n - m - 1}{n - m - e_{Si}}}$$

$$e_{Si} = \frac{e_i}{\sigma \sqrt{1 - H_{ii}}}$$

H_{ii} *i*-tý diagonální prvek projekční matice H

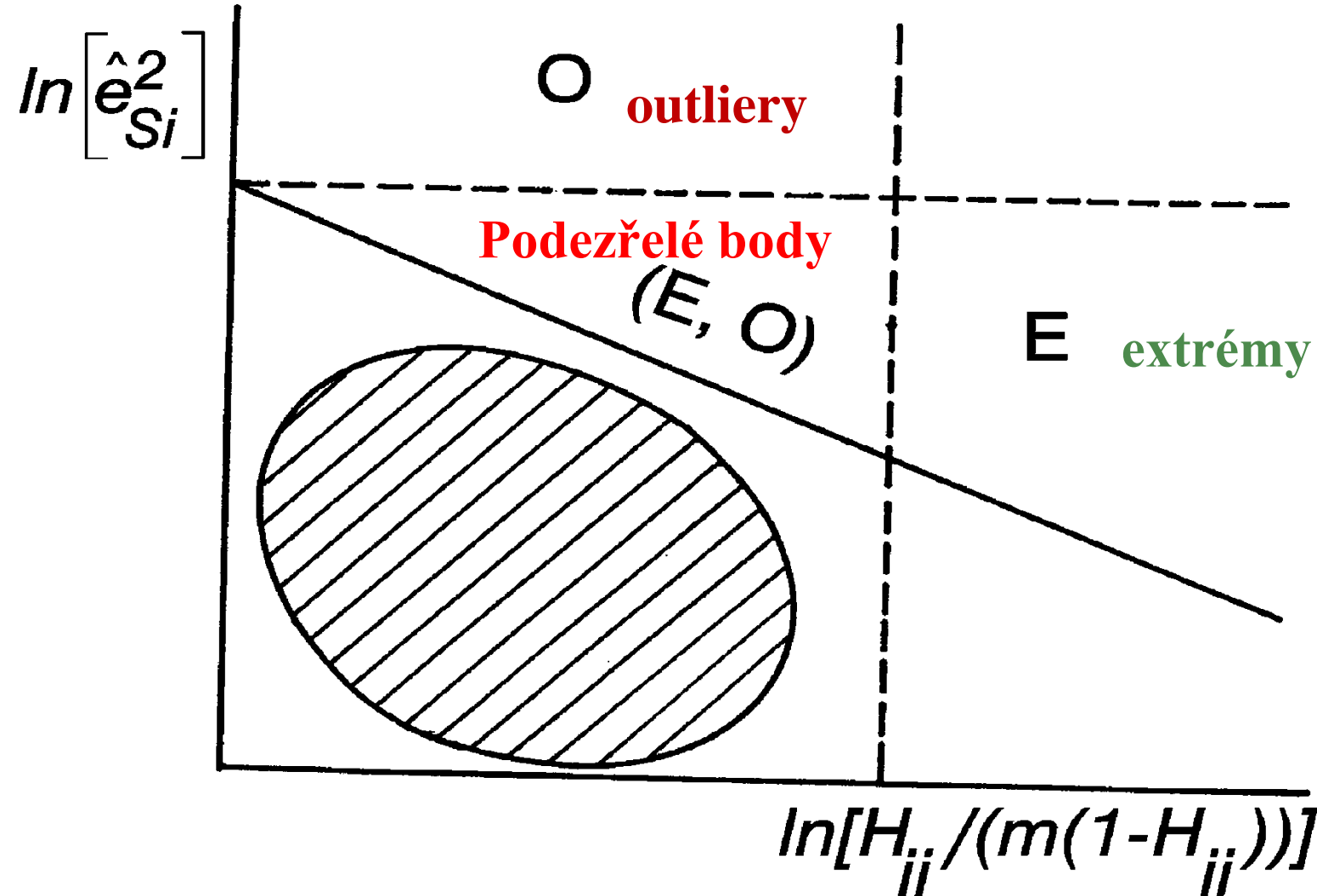
3. Pregibonův graf (PG),

osa x: prvky H_{ii} , osa y: \hat{e}_{Ni}^2

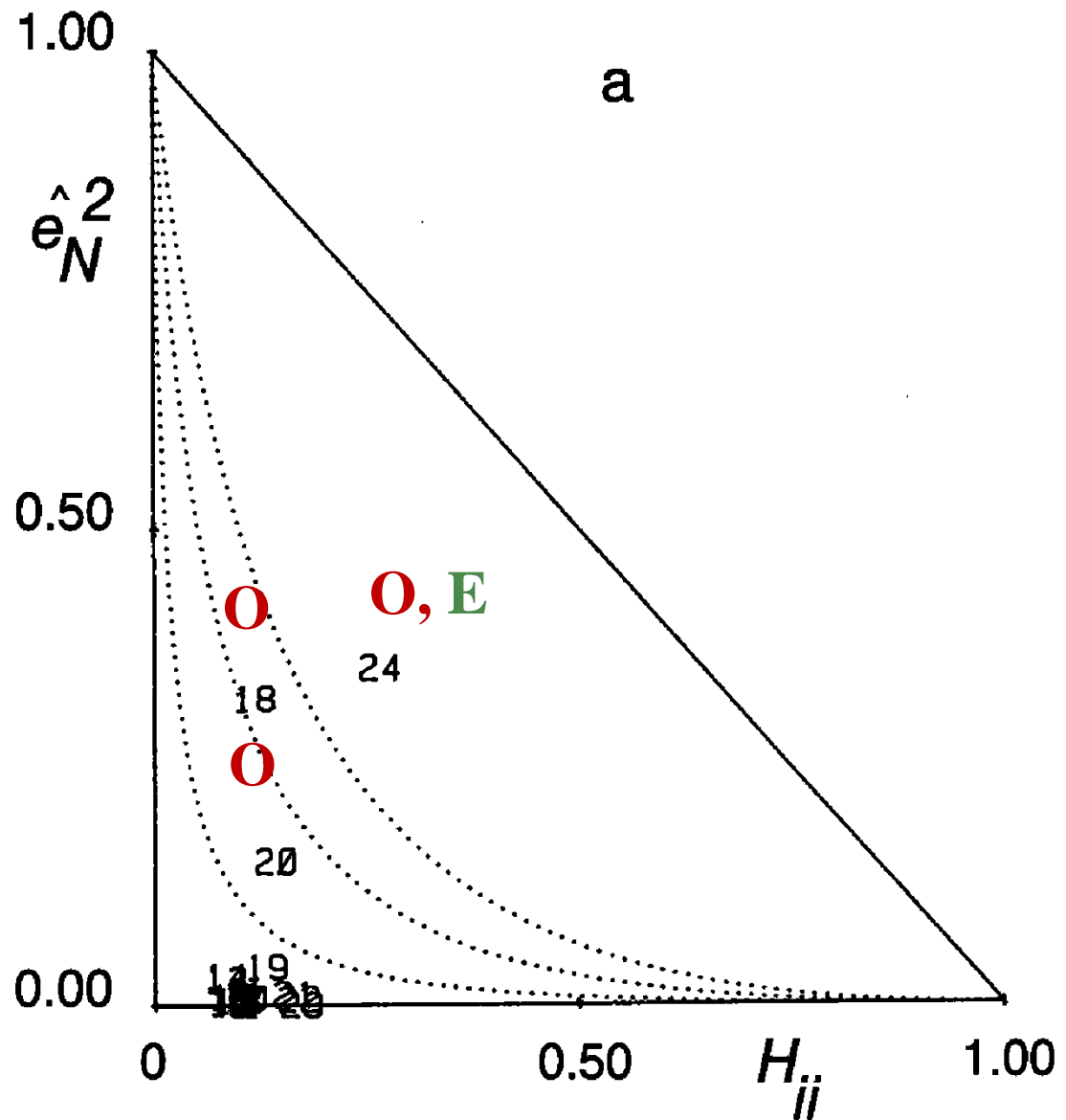


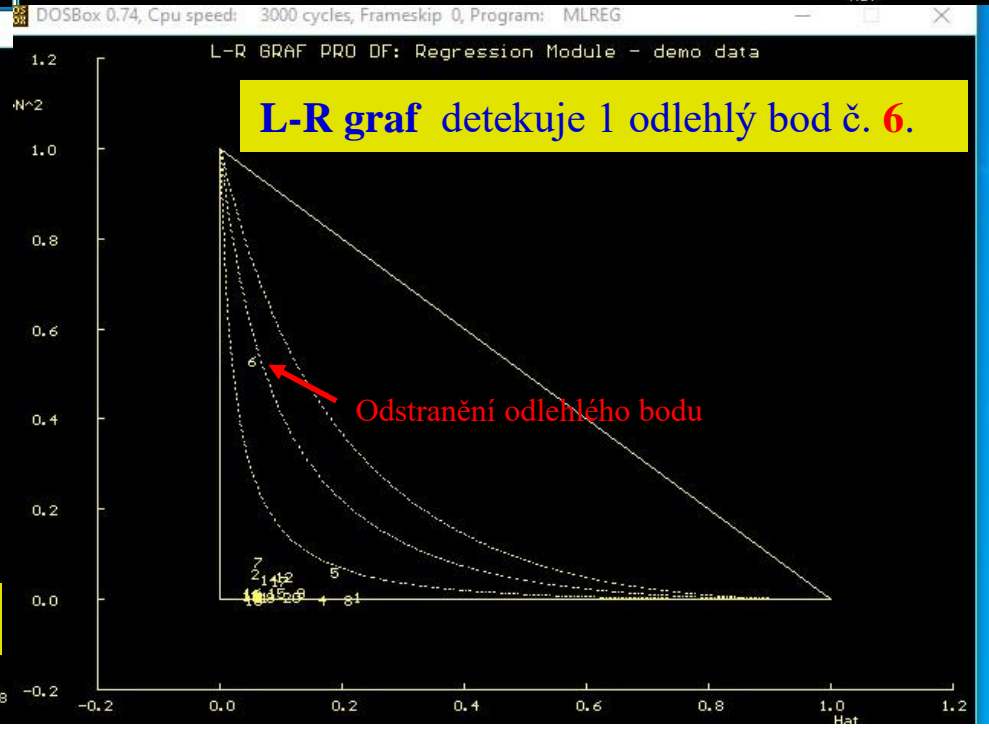
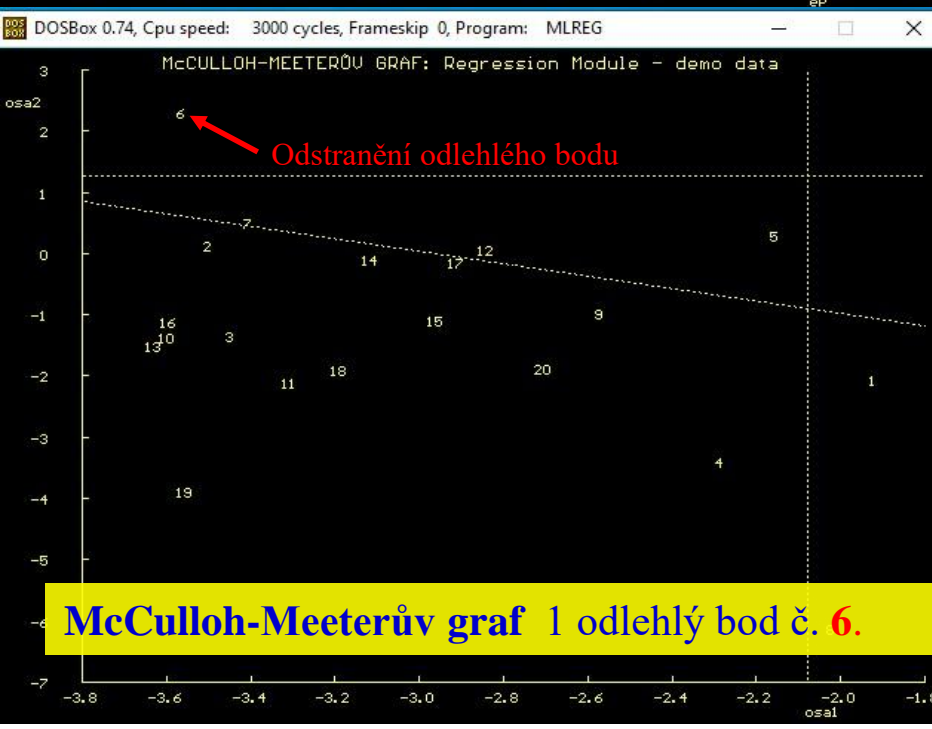
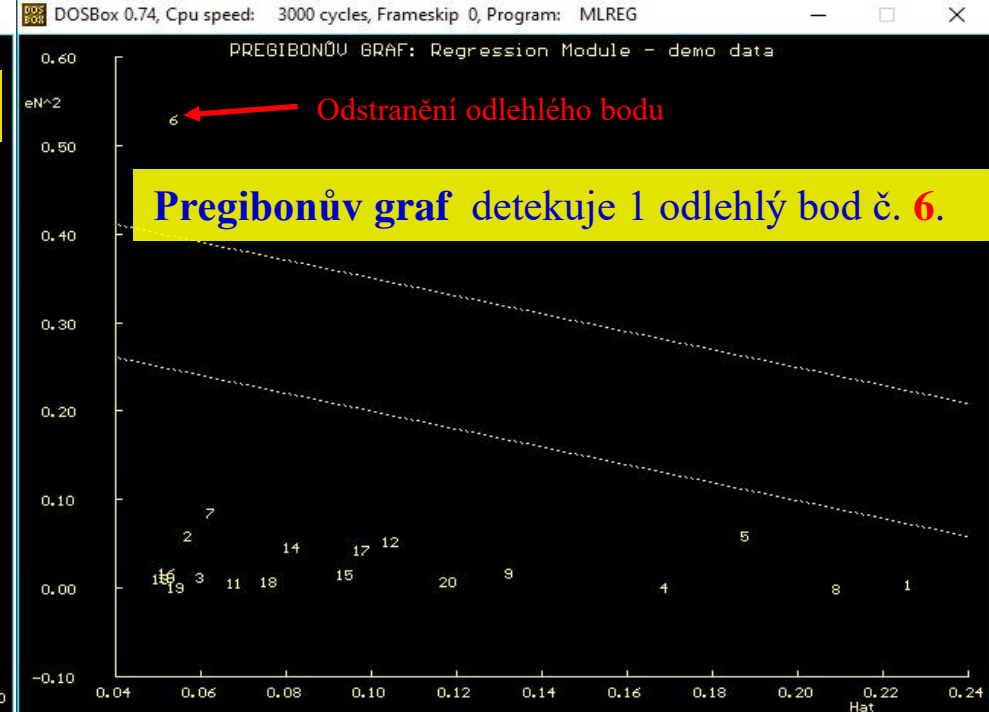
4. McCullohův-Meeterův graf (MMMG)

osa x: $\ln [H_{ii} / (m(1 - H_{ii}))]$, osa y: $\ln e_{S,i}^2$



5. L-R graf (LR)





Kritika regresního modelu

(3) ODHADY PARAMETRU A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B[j] = 0 vs. HA: B[j] <> 0	t-kriterium	hypoteza H0 je	Hlad. výz.
BI 0]	9.0705E+01	1.7421E+01	5.2067E+00	Zamítnuta		0.000
BI 1]	1.4272E+00	1.7771E-01	8.0310E+00	Zamítnuta		0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 8.8420E-01
Koeficient determinace, R ²	: 7.8181E-01
Predikovaný korelační koeficient, Rp ²	: 8.6316E-01
Střední kvadratická chyba predikce, MEP	: 9.1909E+02
Akaikeho informační kritérium, AIC	: 1.3735E+02

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	yexp[i]	yvyp[i]	s(yvyp[i])	e[i]	er[i]
1	1.3000E+02	1.2068E+02	1.4040E+01	9.3240E+00	7.1223E+00
4	2.9700E+02	3.0195E+02	1.2140E+01	-4.9271E+00	-1.6589E+00

V bloku Výsledky doľů na bloky (3) a (4).

Těsnost proložení statistickou analýzou reziduí

7	2.8300E+02	2.4627E+02	7.3670E+00	3.6733E+01	1.2980E+01
8	3.1600E+02	3.1477E+02	1.3510E+01	1.2284E+00	3.8872E-01
9	1.6900E+02	1.5207E+02	1.0749E+01	1.6926E+01	1.0016E+01
10	1.9600E+02	2.1059E+02	6.7168E+00	-1.4588E+01	-7.4428E+00
11	2.4200E+02	2.5198E+02	7.7072E+00	-9.9760E+00	-4.1223E+00
12	1.3600E+02	1.6492E+02	9.5391E+00	-2.8918E+01	-2.1264E+01

Nyní blok (5) Analýza klasických reziduí

16	2.1400E+02	2.3057E+02	6.7365E+00	-1.6568E+01	-7.7422E+00
17	2.9800E+02	2.7196E+02	9.2363E+00	2.6044E+01	8.7395E+00
18	1.9300E+02	1.8204E+02	8.1370E+00	1.0956E+01	5.6764E+00
19	2.3900E+02	2.3485E+02	6.8591E+00	4.1501E+00	1.7364E+00
20	2.7100E+02	2.8195E+02	1.0144E+01	-1.0947E+01	-4.0393E+00

Rezidualní součet čtverců, RSC	: 1.5731E+04
Průměr absolutních hodnot reziduí, Me	: 2.0783E+01
Průměr relativních reziduí, Mer	: 1.1941E+01
Odhad reziduálního rozptylu, s ² (e)	: 8.7395E+02
Odhad směrodatné odchylky reziduí, s(e)	: 2.9563E+01
Odhad šikmosti reziduí, g1(e)	: -1.5653E+00
Odhad špičatosti reziduí, g2(e)	: 6.0782E+00

Kritika regresní metody

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

Fisher-Snedocorův test významnosti regrese, F	: 6.4498E+01
Tabulkový kvantil, F(1-alpha, m-1, n-m)	: 4.4139E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.000
Scottovo kritérium multikolinearity, M	: -2.2034E-16
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, Sf	: 1.0939E+01
Tabulkový kvantil, Chi ² (1-alpha, 1)	: 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti	: 0.001
Jarque-Berraův test normality reziduí, L(e)	: 1.6063E+01
Tabulkový kvantil, Chi ² (1-alpha, 2)	: 5.9915E+00
Závěr: Normalita není přijata.	
Spočtená hladina významnosti	: 0.000
Hladina testování významnosti, H	: 4.8191E+00
Tabulkový kvantil, Chi ² (1-alpha, 2)	: 3.8415E+00
Závěr: Rezidua jsou autororelovaná.	
Spočtená hladina významnosti	: 0.028

Táhněte doľů na blok (6)
Testování regresního tripletu.

Tabulka indikace vlivných bodů

(7) INDIKACE 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	3.5840E-01	3.4955E-01	1.2039E+01	2.2555E-01*
2	-1.0657E+00	-1.0700E+00	-3.2440E+01	5.6782E-02
3	5.0883E-01	4.9809E-01	1.5512E+01	5.9601E-02
4	-1.8279E-01	-1.7781E-01	-5.9265E+00	1.6864E-01
5	1.1599E+00	1.1718E+00	3.8042E+01	1.8756E-01
6	-3.1745E+00	-4.6502E+00*	-9.6457E+01	5.3400E-02
7	1.2830E+00	1.3081E+00	3.9165E+01	6.2101E-02
8	4.6714E-02	4.5401E-02	1.5526E+00	2.0884E-01*
9	6.1463E-01	6.0368E-01	1.9505E+01	1.3222E-01
10	-5.0671E-01	-4.9599E-01	-1.5382E+01	5.1622E-02
11	-3.4954E-01	-3.4085E-01	-1.0703E+01	6.7969E-02
12	-1.0335E+00	-1.0356E+00	-3.2279E+01	1.0412E-01
13	4.7091E-01	4.6049E-01	1.4284E+01	5.0191E-02
14	9.5382E-01	9.5440E-01	2.9447E+01	8.1021E-02
15	2.8861E+01	2.8861E+01	2.8861E+01	9.3509E-02
16	2.8861E+01	2.8861E+01	2.8861E+01	5.1926E-02
17	2.8861E+01	2.8861E+01	2.8861E+01	9.7614E-02

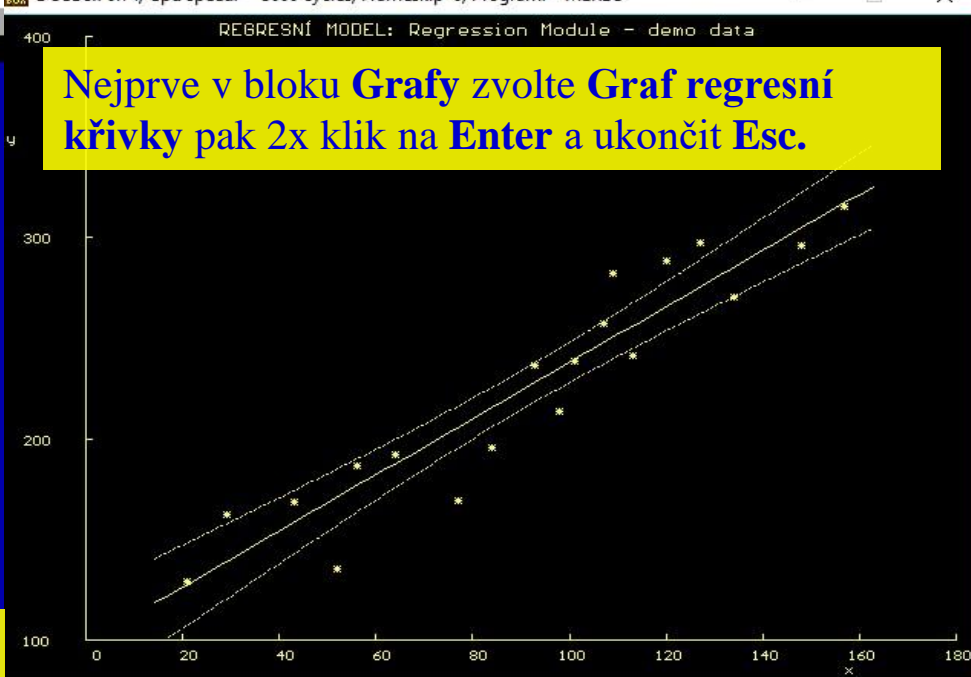
Tabulka (7) Indikace vlivných bodů (= odlehlých a extrémů) značeno *.

F1=náponověda F2=uložení F3=čtení F4=numerický mód F5=zoom ESC=konec

Rádek	6	Sloupec	5	Insert	NUM	J625
2.	1.00000	E+01	1.30000	E+02		
7.	7.00000	E+01	1.70000	E+02		
1.	070000	E+02	2.58000	E+02		
1.	480000	E+02	2.97000	E+02		
2.	900000	E+01	1.63000	E+02		
1.	090000	E+02	2.83000	E+02		
1.	570000	E+02	3.16000	E+02		
4.	300000	E+01	1.69000	E+02		
8.	400000	E+01	1.96000	E+02		
1.	130000	E+02	2.42000	E+02		
5.	200000	E+01	1.36000	E+02		
9.	300000	E+01	2.37000	E+02		
1.	200000	E+02	2.89000	E+02		
5.	600000	E+01	1.87000	E+02		
9.	800000	E+01	2.14000	E+02		
1.	270000	E+02	2.98000	E+02		
6.	400000	E+01	1.93000	E+02		
1.	010000	E+02	2.39000	E+02		
1.	340000	E+02	2.71000	E+02		

Odstranění odlehlého bodu

Odstraňte odlehlý bod č. 6 a opakujte celou regresi.



Williamsův graf detekuje žádný odlehlý bod. Pak ukončit klikem Esc zpět do menu.



McCulloh-Meeterův graf detekuje žádný odlehlý bod. Pak ukončit klikem Esc zpět do menu.

Kritika regresního modelu

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: $ B_j = 0$ vs. H1: $ B_j > 0$	t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	9.8595E+01	1.2013E+01	8.2073E+00	Zamítnuta		0.000
B[1]	1.3934E+00	1.2153E-01	1.1465E+01	Zamítnuta		0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.4100E-01
Koeficient determinace, R ²	: 8.8548E-01
Predikovaný korelační koeficient, R _p ²	: 9.2796E-01
Střední kvadratická chyba predikce, MEP	: 4.4197E+02
Akaikeho informační kritérium, AIC	: 1.1607E+02

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	y _{exp[i]}	y _{yp[i]}	s(y _{yp[i]})	e _{il}	er _{il}
1	1.3000E+02	1.2786E+02	9.7080E+00	2.1447E+00	1.6498E+00
2	1.7000E+02	2.0589E+02	9.4153E+00	-2.5884E+03	-2.1400E+01
3	2.1000E+02	2.0589E+02	9.4153E+00	5.8847E+00	4.6835E+00
4	2.5000E+02	3.1011E+02	8.5108E+00	-7.8121E+00	-2.8393E+00
5	2.9000E+02	3.1011E+02	8.5108E+00	-7.8121E+00	-2.8393E+00

V bloku Výsledky dole na bloky (3) a (4).

Napověda-F1 Řádek: 68 - 90 Celkem: 216 Délka: 12106

Těsnost proložení statistickou analýzou reziduí

6	2.8300E+02	2.5047E+02	5.1098E+00	3.2529E+01	1.1494E+01
7	3.1600E+02	3.1735E+02	9.2393E+00	-1.3530E+00	-4.2815E-01
8	1.6900E+02	1.5851E+02	7.4676E+00	1.0491E+01	6.2075E+00
9	1.9600E+02	2.1564E+02	4.7121E+00	-1.9637E+01	-1.0019E+01
10	2.4200E+02	2.5604E+02	5.3337E+00	-1.4045E+01	-5.8037E+00
11	1.3600E+02	1.7105E+02	6.6441E+00	-3.5050E+01	-2.5772E+01
12	2.8700E+02	2.2618E+02	4.6850E+00	3.8224E+00	3.7225E+00
13	3.3700E+02	3.3218E+02	7.3218E+00	1.1493E+01	1.1493E+01
14	3.8700E+02	3.3218E+02	7.3218E+00	1.1493E+01	1.1493E+01
15	2.1400E+02	2.3514E+02	4.7029E+00	-2.1144E+01	-9.8806E+00
16	2.9800E+02	2.7555E+02	6.3525E+00	2.2448E+01	7.5329E+00
17	1.9300E+02	1.8777E+02	5.6896E+00	5.2300E+00	5.7098E+00
18	2.3900E+02	2.3932E+02	4.7803E+00	-3.2452E-01	-1.3578E-01
19	2.7100E+02	2.8531E+02	6.9627E+00	-1.4306E+01	-5.2788E+00

Nyní na (5) Analýza klasických reziduí

Rezidualní součet čtverců, RSC	: 6.9239E+03
Průměr absolutních hodnot reziduí, Me	: 1.5743E+01
Průměr relativních reziduí, Mer	: 7.7195E+00
Odhad reziduálního rozptylu, s ² (e)	: 4.0729E+02
Odhad směrodatné odchylky reziduí, s(e)	: 2.0181E+01
Odhad šikmosti reziduí, g1(e)	: -2.7401E-01
Odhad špičatosti reziduí, g2(e)	: 2.2342E+00

Napověda-F1 Řádek: 92 - 114 Celkem: 216 Délka: 12106

Kritika regresní metody

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

Fisher-Snedocorův test významnosti regrese, F	: 1.3145E+02
Tabulkový kvantil, F(1-alpha, m-1, n-m)	: 4.4513E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.000

Scottovo kritérium multikolinearity, M	: 1.0809E-16
Závěr: Navržený model je korektní.	

Cook-Weisbergův test heteroskedasticity, Sf	: 5.8472E+00
Tabulkový kvantil, Chi ² (1-alpha, 1)	: 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti	: 0.016

Jarque-Berraův test normality reziduí, L(e)	: 7.0209E-01
Tabulkový kvantil, Chi ² (1-alpha, 2)	: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.704

Waldův test autokorelace, Wa	: 7.3135E-03
Spočtená hladina významnosti	: 3.8415E+00

Táhněte na blok (6) Testování regresního tripletu.

Napověda-F1 Řádek: 117 - 103 Celkem: 216 Délka: 12106

Tabulka indikace vlivných bodů

(7) INDIKACE VLIVNÝCH BODŮ:
(* indikuje odlehlý nebo vlivný bod)

Bod	Standardizované reziduum	Jackknife reziduum	Predikované reziduum	Diagonální prvky
i	eS _{il}	eJ _{il}	eP _{il}	H _{i, i}
1	1.2122E-01	1.1765E-01	2.7904E+00	2.3140E-01*
2	-1.8339E+00	-1.9864E+00	-3.8172E+01	5.9955E-02
3	5.2766E-01	5.1615E-01	1.0993E+01	6.1673E-02
4	-4.2482E-01	-4.1434E-01	-9.4082E+00	1.6959E-01
5	1.3237E+00	1.3559E+00	2.9736E+01	1.9298E-01
6	1.6661E+00	1.7671E+00	3.4757E+01	6.4108E-02
7	-7.5407E-02	-7.3168E-02	-1.7117E+00	2.0959E-01
8	5.5953E-01	5.4790E-01	1.2155E+01	1.3692E-01
9	-1.0007E+00	-1.0007E+00	-2.0770E+01	5.4517E-02
10	-7.2159E-01	-7.1102E-01	-1.5100E+01	6.9849E-02
11	-1.8393E+00	-1.9937E+00	-3.9310E+01	1.0839E-01
12	4.4916E-01	4.3836E-01	9.3137E+00	5.2748E-02
13	1.2004E+00	1.2172E+00	2.5293E+01	8.2689E-02
14	5.3127E-01	5.2970E-01	1.1493E+01	9.7592E-02
15	2.1400E+02	2.3514E+02	4.7029E+00	5.4303E-02
16	2.9800E+02	2.7555E+02	6.3525E+00	9.9082E-02
17	1.9300E+02	1.8777E+02	5.6896E+00	7.9482E-02

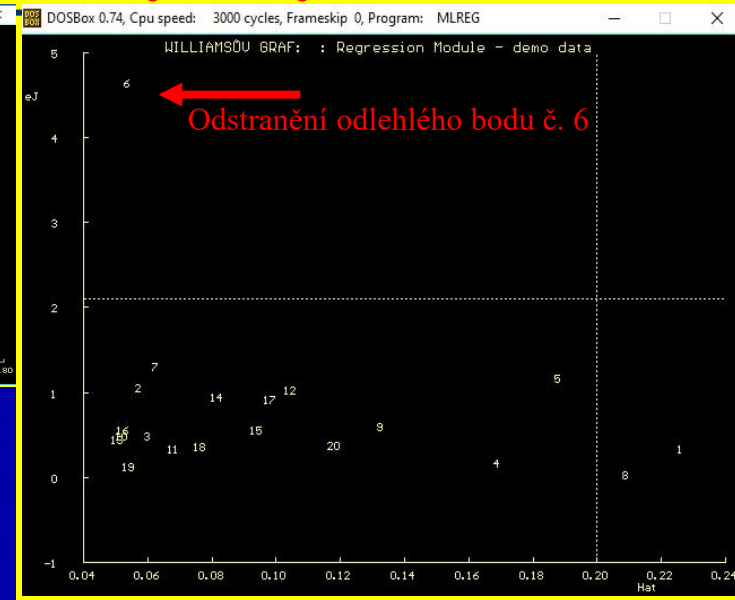
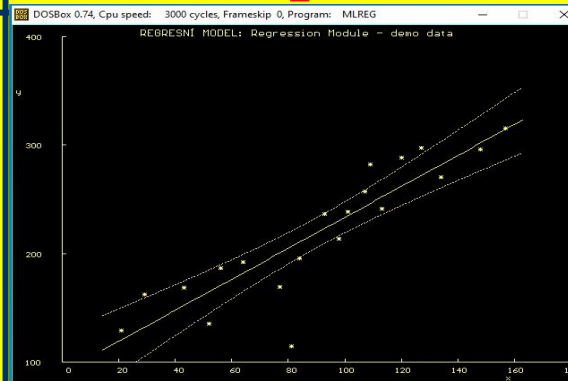
Tabulka (7) Indikace vlivných bodů (= odlehlých a extrémů) značeno *.

Napověda-F1 Řádek: 117 - 103 Celkem: 216 Délka: 12106

Závěry a vysvětlení výstupu úlohy J625:

Úkoly

(1) Dokažte platnost navrženého regresního modelu a existenci vlivných bodů.



Rezidualní součet čtverců, RSC : 6.9239E+03
 Průměr absolutních hodnot reziduí, Me : 1.5743E+01
 Průměr relativních reziduí, Mer : 7.7195E+00
 Odhad reziduálního rozptylu, $s^2(e)$: 4.0729E+02
 Odhad směrodatné odchylky reziduí, $s(e)$: 2.0181E+01
 Odhad šikmosti reziduí, $g1(e)$: -2.7401E-01
 Odhad špičatosti reziduí, $g2(e)$: 2.2342E+00

(2) Testujte statistickou významnost obou parametrů, úseku β_0 a směrnice β_1 .

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
BI 01	9.8595E+01	1.2013E+01	8.2073E+00	8.2073E+00	Zamítnuta	0.000
BI 11	1.3934E+00	1.2153E-01	1.1465E+01	1.1465E+01	Zamítnuta	0.000

(3) Sestrojte 95%ní interval spolehlivosti β_0 a β_1 .

Není v ADSTATu.

(4) Jaká je korelace mezi celkovým cholesterolem v krvi y a denní spotřebou tuku x u sledovaných jedinců?

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 9.4100E-01
 Koeficient determinace, R^2 : 8.8548E-01
 Predikovaný korelační koeficient, R_p^2 : 9.2796E-01
 Střední kvadratická chyba predikce, MEP : 4.4197E+02
 Akaikeho informační kritérium, AIC : 1.1607E+02

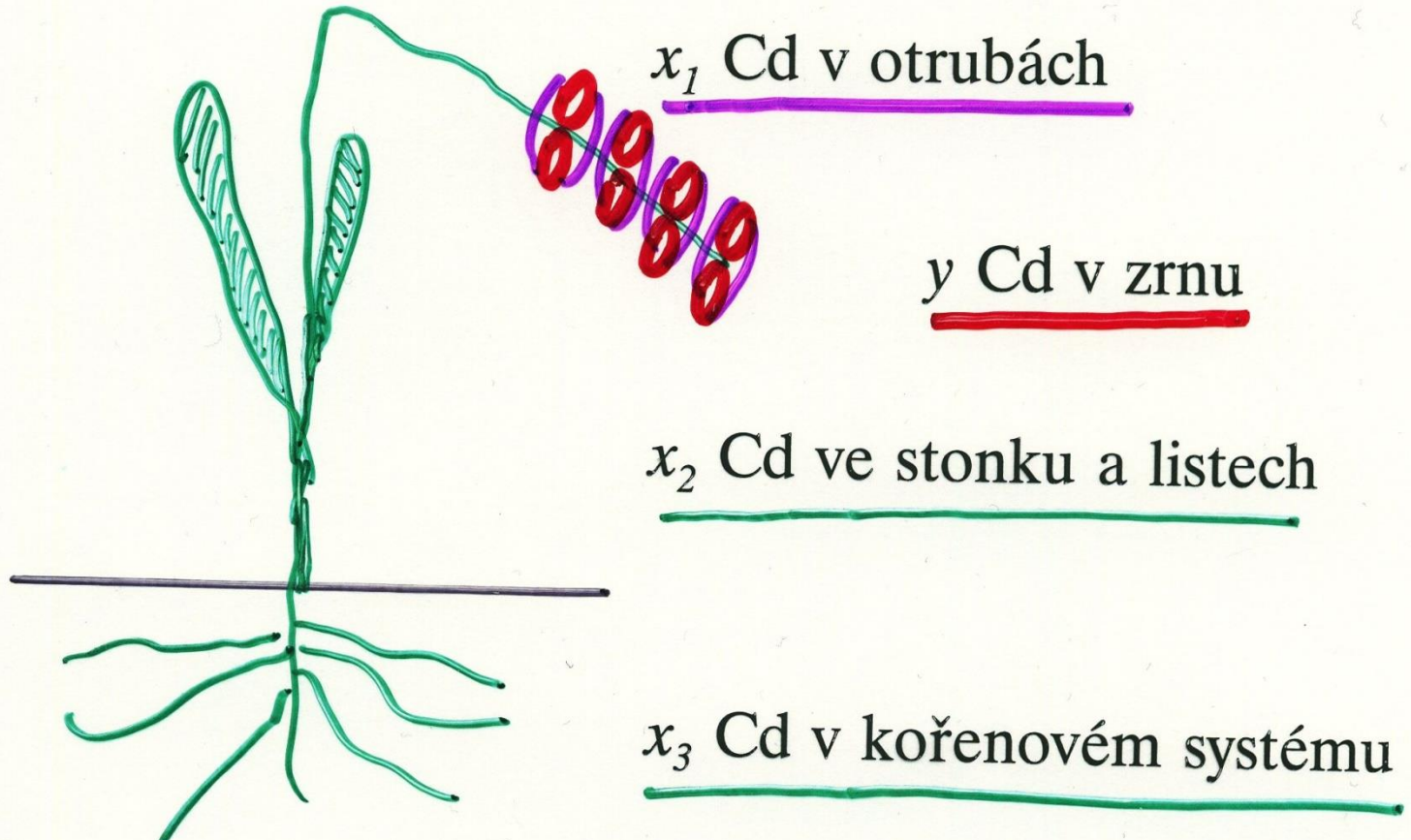
2. Vzorová úloha M619 na výstavbu vícerozměrného lineárního regresního modelu

Jak napsat a vysvětlit úlohu vsemestrální práci je
v *Kompendiu*, str. 692

Návod k sestavení a napsání semestrální práce k zadané úloze najdete ve cvičebnici *M. Meloun, J. Militký: Kompendium statistického zpracování dat, Karolinum Praha 1992, ISBN 978-80-246-2196-8*, ve vzorovém příkladu na str. 692 – 705 a obecný výklad outputu v lineární regresi na str. 582 - 583. V ADSTATu používejte na přenos výstupů a obrázků **PrintScreen** a upravte pak negativ na černobílý pozitiv ve MS Windows.

Úloha M619. Vliv tří parametrů na obsah kadmia v potravinářské pšenici

Obsah kadmia v zrně 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.



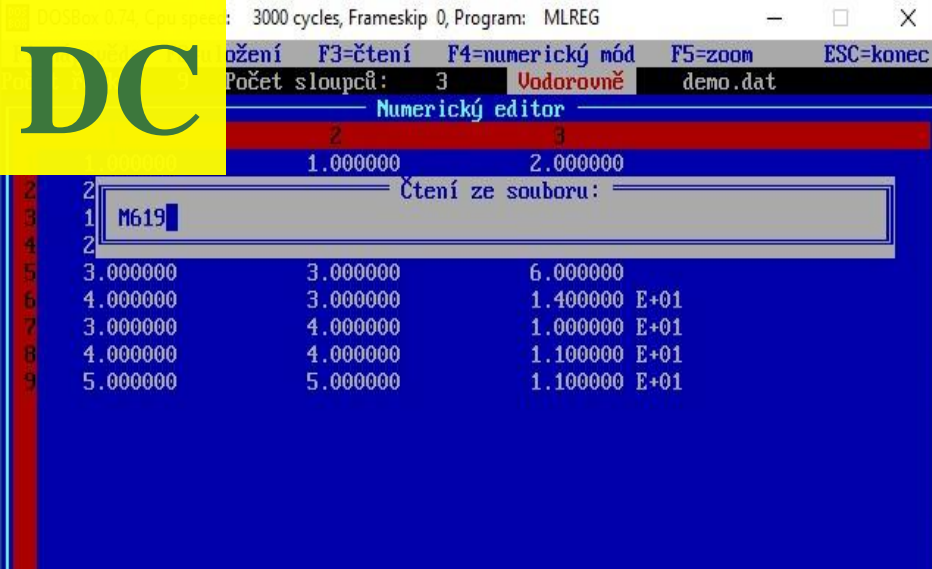
Výstavba lineárního regresního modelu:

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

Kritika dat

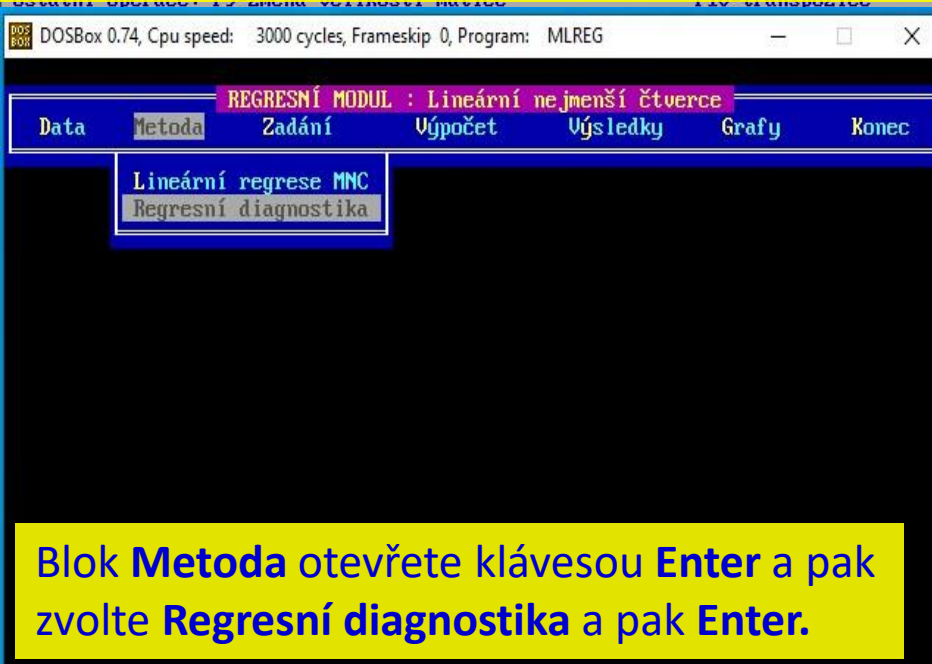
Kritika modelu

Kritika metody



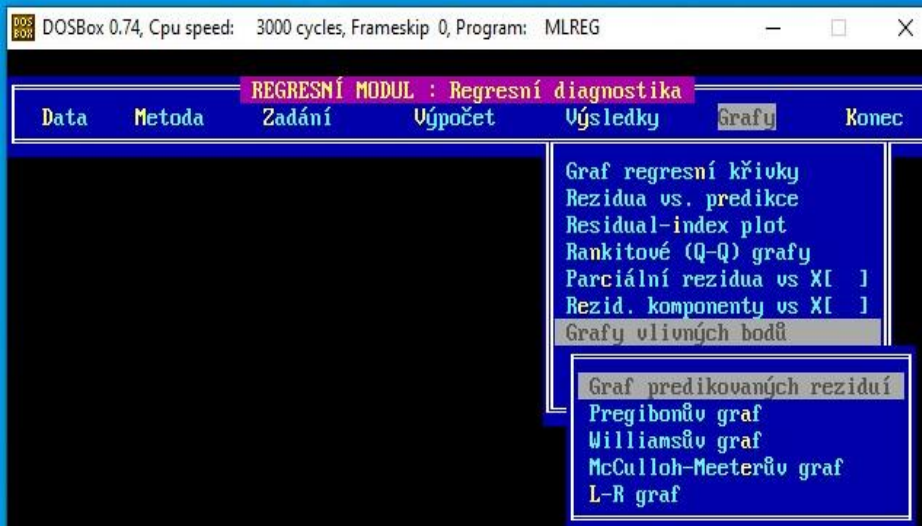
Otevřít **Data**, pak stiskněte **F3** a smažte zápis *demo.dat* a napište **M619** a ukončete **Esc a A**.

Otevřít **Data** a pak **F5** a ukončit **Esc a A**.

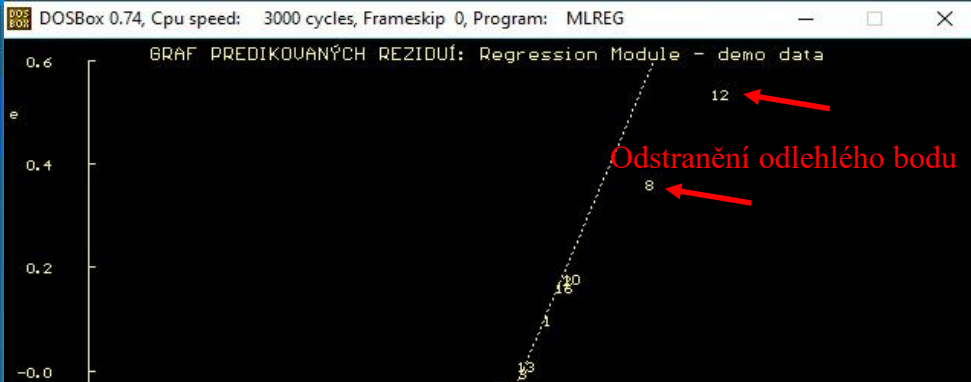


Blok **Metoda** otevřete klávesou **Enter** a pak zvolte **Regresní diagnostika** a pak **Enter**.

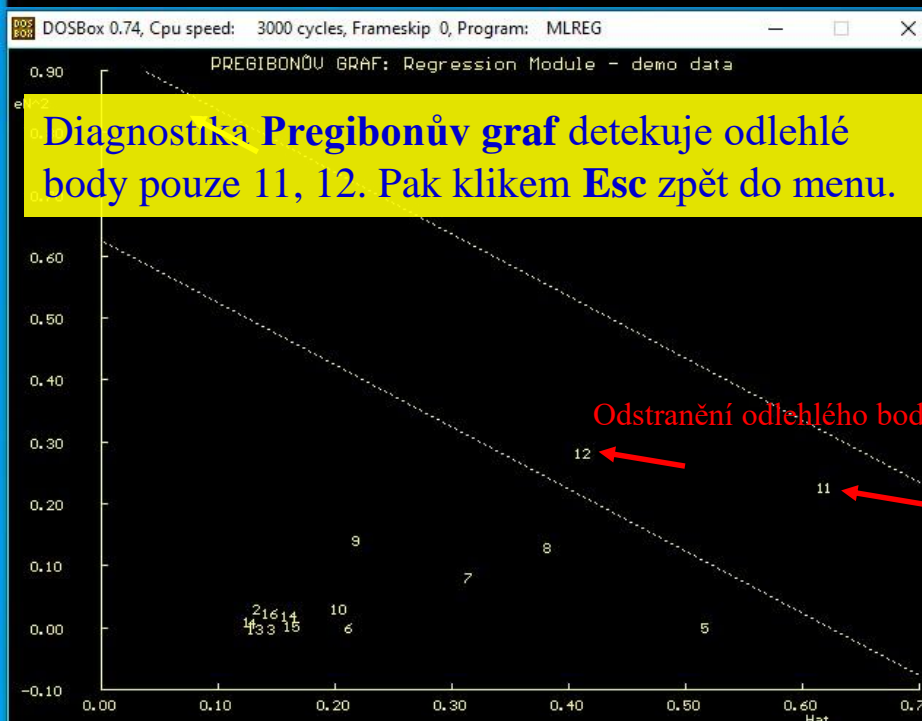
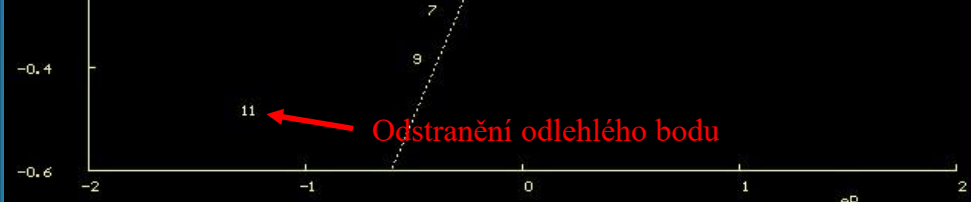
V **Podmínky** nechte vše předvolené defaultní a dejte **7x Enter**.



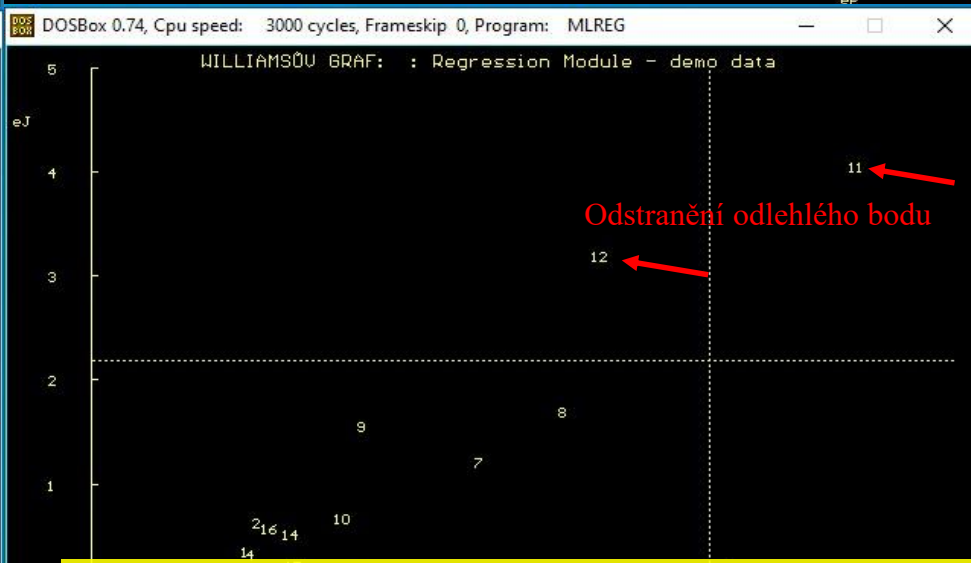
Běžte do bloku **Grafy** a zvolte **Grafy vlivných bodů** a otevřete kliknutím na **Enter**.



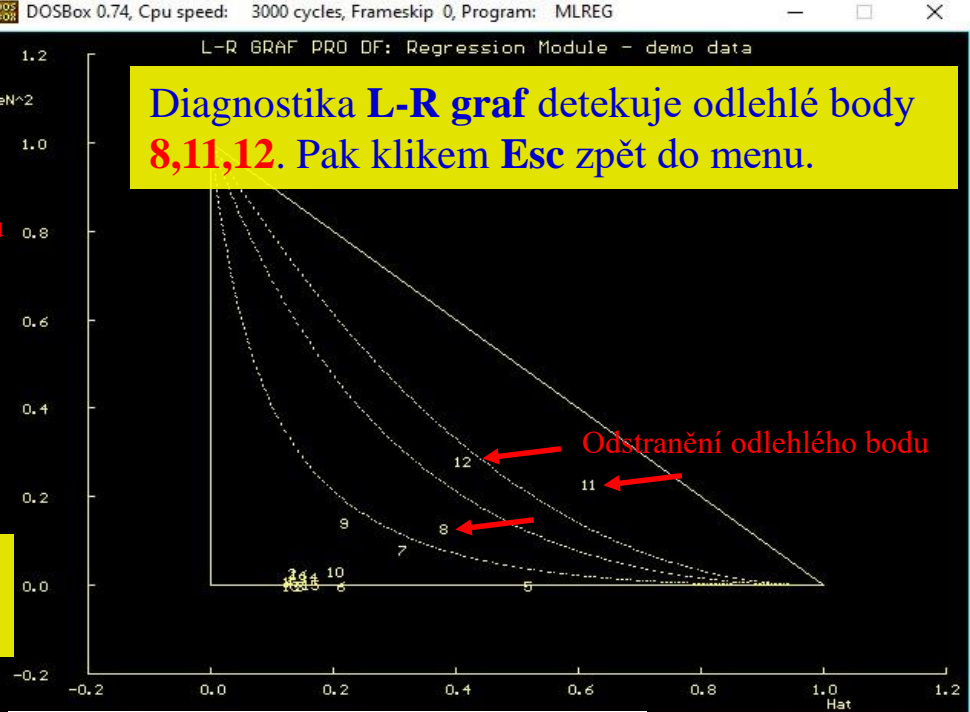
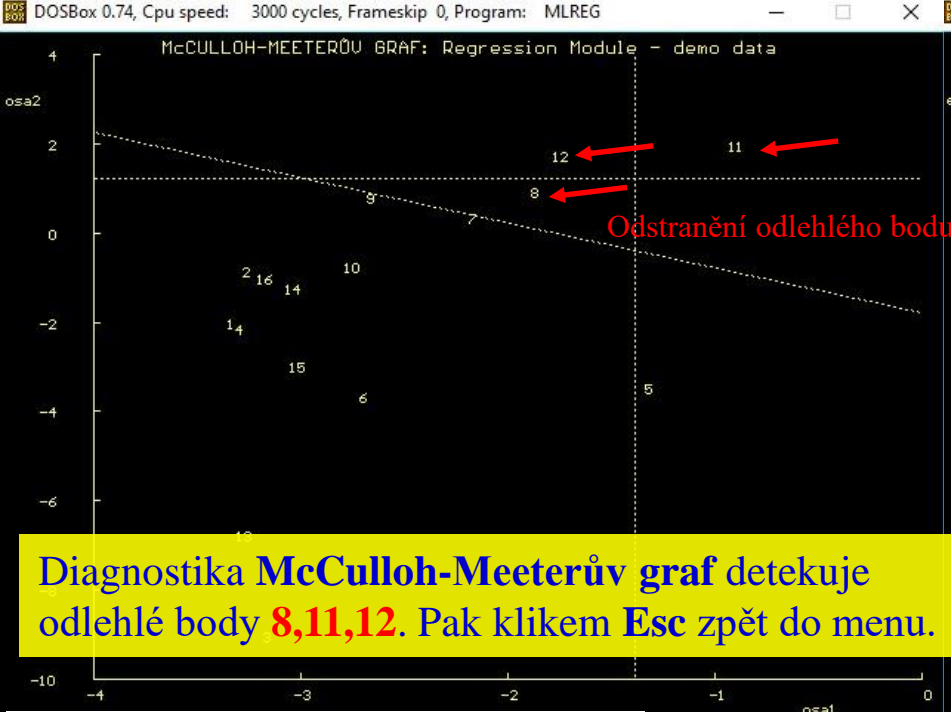
Diagnostika **Graf predikovaných reziduí** detekuje odlehlé body 8, 11, 12. Pak klikem **Esc** zpět do menu.



Diagnostika **Pregibonův graf** detekuje odlehlé body pouze 11, 12. Pak klikem **Esc** zpět do menu.



Diagnostika **Williamsův graf** detekuje odlehlé body pouze 11, 12. Pak klikem **Esc** zpět do menu.



Tabulka indikace vlivných bodů

(7) INDIKACE VLIVNÝCH BODŮ:
(* indikuje odlehlý nebo vlivný bod)

Bod	Standardizované reziduum eS[i],i	Jackknife reziduum eJ[i],i	Predikované reziduum eP[i],i	Diagonální prvky H[i],i
1	3.6578E-01	3.5217E-01	1.1343E-01	1.2352E-01
2	6.5601E-01	6.3966E-01	2.0453E-01	1.3298E-01
3	-1.0877E-02	-1.0414E-02	-3.4146E-03	1.4474E-01
4	-3.4147E-01	-3.2854E-01	-1.0620E-01	1.2863E-01
5	-1.7553E-01	-1.6827E-01	-7.3310E-02	5.1683E-01*
6	-1.5911E-01	-1.5249E-01	-5.2019E-02	2.1152E-01
7	-1.1925E+00	-1.2161E+00	-4.1777E-01	3.1326E-01
8	1.5859E+00	1.7078E+00	5.8594E+00*	5.8594E+00*
9	-1.4821E+00	-1.5700E+00	-4.8658E-01	2.1800E-01
10	6.9110E-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	3.3932E-02	3.2489E-02	1.0568E-02	1.3115E-01
14	-5.4162E-01	-5.2502E-01	-1.7463E-01	1.6062E-01
15	-2.2469E-01	-2.1709E-01	-7.8065E-02	1.4474E-01
16	5.9952E-01	5.8392E-01	1.8065E-01	1.3298E-01

Blok (7) Indikace vlivných bodů přináší diagnostiky vlivných bodů postupně ve 3 tabulkách v číselné formě.

Tabulka indikace vlivných bodů

Bod	Zobecněné diag. prvky Hm[i],i	Cookova vzdálenost D[i],i	Atkinsonova vzdálenost A[i],i	Vliv na predikci DF[i],i
1	1.3329E-01	4.7137E-03	2.2899E-01	1.3221E-01
2	1.6407E-01	1.6501E-02	4.3390E-01	2.5051E-01
3	1.4475E-01	5.0059E-06	7.4206E-03	-4.2843E-03
4	1.3709E-01	4.3030E-03	2.1863E-01	-1.2622E-01
5	5.1807E-01	8.2390E-03	3.0143E-01	-1.7403E-01
6	2.1318E-01	1.6978E-03	1.3680E-01	-7.8983E-02
7	3.9465E-01	1.6218E-01	1.4226E+00	-8.2134E-01
8	5.1049E-01	3.8649E-01*	2.3192E+00*	1.3390E+00*
9	3.6115E-01	1.5310E-01	1.4358E+00	8.2894E-01
10	2.3414E-01	3.0299E-02	5.8916E-01	3.4015E-01
11	8.4709E-01*	2.9145E+00*	8.9433E+00*	-5.1634E+00*
12	6.9419E-01*	1.0062E+00*	4.6176E+00*	2.6660E+00*
13	1.3124E-01	4.3450E-05	2.1863E-02	1.2623E-02
14	1.8114E-01	1.4034E-02	3.9779E-01	-2.2966E-01
15	1.6657E-01	2.4587E-03	1.6480E-01	-9.5150E-02
16	1.6888E-01	1.5020E-02	4.1270E-01	2.3827E-01

Blok (7) Indikace vlivných bodů přináší diagnostiky vlivných bodů postupně ve 3 tabulkách v číselné formě.

Tabulka indikace vlivných bodů

Bod	U ě r o h o d n o s t n í v z d á l e n o s t i		
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	2.6698E-05	3.2606E-02	3.2606E-02
4	2.2933E-02	2.3569E-02	4.5295E-02
5	4.3881E-02	3.0095E-02	7.1396E-02
6	9.0523E-03	3.0538E-02	3.9045E-02
7	8.4238E-01	3.0992E-02	9.5090E-01
8	1.9389E+00	2.4682E-01	2.6917E+00
9	7.9637E-01	1.5878E-01	1.0958E+00
10	1.6079E-01	4.5424E-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

Kritika regresního modelu

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B _{ij} = 0 vs. H _A : B _{ij} <> 0	t-kriterium	hypoteza H0 je	Hlad. výz.
B _{1 0}	-7.2665E-02	1.3791E-01	-5.2692E-01		Akceptována	0.608
B _{1 1}	-6.8505E-01	1.9165E-01	-3.5746E+00		Zamítnuta	0.004
B _{1 2}	8.9619E-01	1.6072E-01	5.5761E+00		Zamítnuta	0.000
B _{1 3}	8.3769E-01	1.3322E-01	6.2879E+00		Zamítnuta	0.000

(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
Ákaikeho informační kritérium, AIC	:-3.6180E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	uexp[i]	uump[i]	s(uump[i])	e[i]	er[i]

V bloku Výsledky táhnete dolů na bloky (3) a (4).

Těsnost proložení statistickou analýzou reziduí

3	2.1000E+00	2.1029E+00	1.1045E-01	-2.9204E-03	-1.3907E-01
4	2.1000E+00	2.1925E+00	1.0412E-01	-9.2540E-02	-4.4066E+00
5	8.1000E+00	8.1354E+00	2.0871E-01	-3.5421E-02	-4.3730E-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	8.6000E+00	8.8905E+00	1.2555E-01	3.8050E-01	3.8636E+00

Táhněte na blok (5) Analýza klasických reziduí

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	9.1822E-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)	: 9.2355E-02
Odhad špičatosti reziduí, g ₂ (e)	: 2.8755E+00

Kritika regresní metody

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

Fisher-Snedocorův test významnosti regrese, F	: 1.4050E+03
Tabulkový kvantil, F(1-alpha, m-1, n-m)	: 3.4903E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.000
Scottovo kritérium multikolinearity, M	: 9.6119E-01
Závěr: Navržený model není korektní.	

Cook-Weisbergův test heteroskedasticity, Sf	: 1.9926E+01
Tabulkový kvantil, Chi ² (1-alpha, 1)	: 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti	: 0.000

Jarque-Berraův test normality reziduí, L(e)	: 3.3085E-02
Tabulkový kvantil, Chi ² (1-alpha, 2)	: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.984

Táhněte na blok (6) Testování regresního tripletu.

Tabulkový kvantil, Chi ² (1-alpha, 1)	: 1.0320E+01
Spočtená hladina významnosti	: 3.8415E+00
Spočtená hladina významnosti	: 0.001

Výstavbu regresního modelu uspořádejte do tohoto přehledu kroků

A, B, C a D.

LINEÁRNÍ REGRESE

Regresní diagnostika

Název: m619

V S T U P

(1) ZVOLENÁ STRATEGIE REGRESNÍ ANALÝZY:

Omezení, P	: 1.0000E-34
Transformace	: Ne
Váhy	: Ne
Absolutní člen zahrnut	: Ano

(2) PODMÍNKY A KVANTILY PRO STATISTICKÉ TESTY:

Hladina významnosti, alfa	: 0.050
Počet bodů, n	: 16
Počet parametrů, m	: 3
Kvantil Studentova rozdělení $t(1-\alpha/2, n-m)$: 2.179
Kvantil rozd. Chí-kvadrát $\text{Chi-square}(1-\alpha, m)$: 9.488
Jméno výstupního souboru	: M619.txt

V Ý S T U P

(1) PŘEDBĚŽNÁ STATISTICKÁ ANALÝZA:

Proměnná	Průměr	Směrodatná odchylka	Párový korelační koeficient	Spočtená hladina výz.
y	6.0125E+00	4.8734E+00	1.0000	-----
x1 OTRUBY	4.8937E+00	3.5692E+00	0.9837	0.000
x2 LISTY	5.7812E+00	4.5296E+00	0.9935	0.000
x3 KOŘEN	5.0812E+00	3.8782E+00	0.9948	0.000

Párové korelační koeficienty mezi dvojicemi vysvětlujících proměnných

x1 versus x2 :	9.9344E-01	Spočtená hladina výz.	0.000
x1 versus x3 :	9.8693E-01	Spočtená hladina výz.	0.000
x2 versus x3 :	9.8847E-01	Spočtená hladina výz.	0.000

(2) INDIKACE MULTIKOLINEARITY:

Č [j]	Vlastní čísla korel. matice l[j]	Čísla podmíněnosti K[j]	Variance inflation factor VIF[j]	Vícenás.korel. koef pro X[j]
1	6.4568E-03	4.6141E+02	8.3272E+01	0.9940
2	1.4307E-02	2.0823E+02	9.4324E+01	0.9947
3	2.9792E+00	1.0000E+00	4.7508E+01	0.9894

Maximální číslo podmíněnosti K : 4.6141E+02

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

(VIF[j] > 10 indikuje silnou multikolaritu)

Kritika dat: původní data obsahují vlivné body 8, 11, 12

Kritika modelu $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$: technika MEP, AIC, $s(e)$) našla, že absolutní člen β_0 statisticky nevýznamný

Odhad parametru	Směrodatná odchylka	Student t-test	Parameter je	Hladina významnosti
-----------------	---------------------	----------------	--------------	---------------------

$t_{0.95}(16-4) = 2.18$

A. Původní data a původní model vedou k závěru

$$y = -0.07(0.14) - 0.69(0.19)x_1 + 0.90(0.16)x_2 + 0.84(0.13)x_3$$

(LS: $t_{1-0.05/2}(16-4) = 2.179$, $D = 99.72\%$, $MEP = 0.21014$, $AIC = -36.18$, $s(e) = 0.290$)

β_0	-0.073	0.138	-0.5269	Nevýznamný	0.608
β_1	-0.685	0.192	-3.5746	Významný	0.004
β_2	0.896	0.161	5.5761	Významný	0.000
β_3	0.838	0.133	6.2879	Významný	0.000

B. Data bez 8, 11, 12 a původní model vedou k závěru

$$y = -0.02(0.10) - 0.80(0.50)x_1 + 0.92(0.34)x_2 + 0.91(0.15)x_3$$

(LS: $t_{1-0.05/2}(13-4) = 2.262$, $D = 99.83\%$, $MEP = 0.06118$, $AIC = -41.58$, $s(e) = 0.178$)

β_0	-0.017	0.103	-0.1637	Nevýznamný	0.874
β_1	-0.802	0.504	-1.5919	Nevýznamný	0.146
β_2	0.920	0.338	2.7193	Významný	0.024
β_3	0.906	0.152	5.9698	Významný	0.000

3000 cycles, Frameskip 0, Program: MLREG

ožení F3=čtení F4=numerický mód F5=zoom ESC=konec

oupec 5 Insert NUM M619

1.500000	1.500000	1.600000
1.600000	1.300000	1.600000
1.900000	2.200000	2.100000
2.000000	2.200000	2.100000
6.600000	7.100000	8.100001
7.100000	8.200000	7.900000
7.800000	9.100000	8.400000
8.400000	9.600000	9.600000
8.600000	1.000000 E+01	1.080000 E+01
1.300000	1.300000	1.300000
1.100000	1.300000	1.200000
1.300000	1.600000	1.500000
1.500000	1.600000	1.500000

Odstranění odlehlého bodu

Odstranění odlehlého bodu

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

REGRESNÍ MODUL : Lineární nejmenší čtverce

Data Metoda **Zadání** Výpočet Výsledky Grafy Konec

Podmínky
Volby

PODMÍNKY VÝPOČTU

Název: Regression Module - demo data

Počáteční podmínky:

Omezení : 1.000000E-34 Absolutní člen : **Ne**

Transformace : Ne

Uáhy : Ne

Hladina význam.: 0.050

V Podmínky nechte vše předvolené a dejte 7x Enter.
Do návěští Absolutní člen dejte mezeríkem Ne.

Opakujte nyní celou regresi po odstranění odlehlých bodů **8, 11, 12** v bloku **Data**, která otevřete klikem na **Enter** a pak **F4** a **F5**.

Kritika regresního modelu

(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.
BI 01	0.0000E+00					
BI 11	-8.5450E-01	3.7149E-01	-2.3002E+00	Zamítnuta		0.044
BI 21	9.5419E-01	2.5085E-01	3.8038E+00	Zamítnuta		0.003
BI 31	9.1554E-01	1.3218E-01	6.9263E+00	Zamítnuta		0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 9.9917E-01

Koeficient determinace, R² : 9.9834E-01

Predikovaný korelační koeficient, Rp² : 9.9808E-01

Střední kvadratická chyba predikce, MEP : 5.1006E-02

Akaikeho informační kritérium, AIC :-4.3554E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
1	1.6000E+00	1.5229E+00	5.5129E-02	7.7150E-02	4.8219E+00
2	1.6000E+00	1.4352E+00	5.0385E-02	1.6484E-01	1.0302E+01
3	2.1000E+00	2.1182E+00	8.7234E-02	-1.8157E-02	-8.6464E-01
4	2.1000E+00	2.2136E+00	6.3945E-02	-1.1358E-01	-5.4084E+00
5	8.1000E+00	8.0932E+00	1.2963E-01	6.8229E-03	8.4234E-02
6	7.8000E+00	7.8000E+00	8.5221E-02	1.0001E-01	1.2658E+00
7	7.8000E+00	7.8000E+00	8.5221E-02	1.0001E-01	1.2658E+00
8	7.8000E+00	7.8000E+00	8.5221E-02	1.0001E-01	1.2658E+00
9	1.0800E+01	1.0525E+01	1.2270E-01	2.7536E-01	2.5496E+00
10	1.3000E+00	1.3198E+00	4.7778E-02	-1.9804E-02	-1.5234E+00
11	1.2000E+00	1.3991E+00	2.2917E-02	-1.9915E-01	-1.6596E+01
12	1.5000E+00	1.6061E+00	3.1690E-02	-1.0606E-01	-7.0706E+00
13	1.5000E+00	1.3436E+00	6.1910E-02	1.5639E-01	1.0426E+01

Rezidualní součet čtverců, RSC : 2.8758E-01

Průměr absolutních hodnot reziduí, Me : 1.2398E-01

Průměr relativních reziduí, Mer : 4.9992E+00

Odhad reziduálního rozptylu, s²(e) : 2.8739E-02

Odhad směrodatné odchylky reziduí, s(e) : 1.6952E-01

Odhad šikmosti reziduí, g1(e) : 1.1648E-01

Odhad špičatosti reziduí, g2(e) : 2.1464E+00

V bloku **Výsledky** táhněte dolů na bloky (3) a (4).

Těsnost proložení statistickou analýzou reziduí

Nyní na blok (5) Analýza klasických reziduí

Napověda-F1 Řádek: 77 - 99 Celken: 203 Délka: 11182

Napověda-F1 Řádek: 98 - 120 Celken: 203 Délka: 11182

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

$$y = -1.18(0.38)x_1 + 1.25(0.24)x_2 + 0.92(0.15)x_3$$

(LS: $t_{1-0.05/2}(14-3) = 2.201$, $D = 99.80\%$, $MEP = 0.06078$, $AIC = -43.47$, $s(e) = 0.193$)

β_0	0.000	---	---	---	---
β_1	-1.181	0.383	-3.0854	Významný	0.010
β_2	1.245	0.236	5.2751	Významný	0.000
β_3	0.917	0.151	6.091	Významný	0.000

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

Regresní diagnostiky vlivných bodů v grafech lze uvést také číselně v tabulce, kde **tučně červeně jsou diagnostiky odlehlých bodů** a **tučně zeleně jsou diagnostiky extrémů**.

i	y_i	$y_{P,i}$	$s(y_i)$	e_i	$e_{rel,i}$	$e_{S,i}$	$e_{J,i}$	$e_{P,i}$	H_{ii}	$Hm_{i,i}$	D_i	A_i	AP_i	DF_i	$LD(b)_i$	$LD(s^2)_i$	$LD(b, s^2)_i$
1	1.60	1.50	0.10	0.10	6.21	0.37	0.35	0.11	0.12	0.13	0.00	0.23	0.87	0.13	0.03	0.02	0.05
2	1.60	1.42	0.11	0.18	11.08	0.66	0.64	0.20	0.13	0.16	0.02	0.43	0.84	0.25	0.09	0.01	0.09
3	2.10	2.10	0.11	0.00	-0.14	-0.01	-0.01	0.00	0.14	0.14	0.00	0.01	0.86	0.00	0.00	0.03	0.03
4	2.10	2.19	0.10	-0.09	-4.41	-0.34	-0.33	-0.11	0.13	0.14	0.00	0.22	0.86	-0.13	0.02	0.02	0.05
5	8.10	8.14	0.21	-0.04	-0.44	-0.18	-0.17	-0.07	0.52	0.52	0.01	0.30	0.48	-0.17	0.04	0.03	0.07
6	7.90	7.94	0.13	-0.04	-0.52	-0.16	-0.15	-0.05	0.21	0.21	0.00	0.14	0.79	-0.08	0.01	0.03	0.04
7	8.40	8.69	0.16	-0.29	-3.42	-1.19	-1.22	-0.42	0.31	0.39	0.16	1.42	0.61	-0.82	0.84	0.03	0.95
8	10.30	9.94	0.18	0.36	3.52	1.59	1.71	0.59	0.38	0.51	0.39	2.32	0.49	1.34	1.94	0.25	2.69
9	9.60	9.98	0.14	-0.38	-3.96	-1.48	-1.57	-0.49	0.22	0.36	0.15	1.44	0.64	-0.83	0.80	0.16	1.10
10	10.80	10.62	0.13	0.18	1.66	0.69	0.68	0.22	0.20	0.23	0.03	0.59	0.77	0.34	0.16	0.00	0.16
11	13.10	13.58	0.23	-0.48	-3.67	-2.68	-4.06	-1.26	0.62	0.85	2.91	8.94	0.15	-5.16	10.86	7.83	44.18
12	15.10	14.57	0.19	0.54	3.55	2.40	3.19	0.91	0.41	0.69	1.01	4.62	0.31	2.67	4.63	3.44	13.13
13	1.30	1.29	0.11	0.01	0.71	0.03	0.03	0.01	0.13	0.13	0.00	0.02	0.87	0.01	0.00	0.03	0.03
14	1.20	1.34	0.12	-0.14	-12.01	-0.54	-0.53	-0.17	0.16	0.18	0.01	0.40	0.82	-0.23	0.07	0.01	0.08
15	1.50	1.56	0.12	-0.06	-3.98	-0.22	-0.22	-0.07	0.16	0.17	0.00	0.16	0.83	-0.10	0.01	0.03	0.04
16	1.50	1.34	0.11	0.16	10.74	0.60	0.58	0.19	0.14	0.17	0.02	0.41	0.83	0.24	0.08	0.01	10.09

Klasické reziduum
 Relativní reziduum
 Standardizované reziduum
 Jack-knife reziduum
 Predikované reziduum
 Diagon. prvky H projekce
 Modif. diagon. prvky H projekce
 Cookova vzdálenost
 Atkinsonova vzdálenost
 Andrews-Pregibonova vzdálenost
 Belseyho DF statistika
 Věrohodnostní vzdálenost
 Věrohodnostní vzdálenost
 Věrohodnostní vzdálenost

Závisle proměnná
 Predikce
 Směrodatná odchylka predikce

V Ý S L E D K Y

(7) INDIKACE VLIVNÝCH BODŮ:

(* indikuje odlehlý nebo vlivný bod)

Bod	Standardizované reziduum eS[i]	Jackknife reziduum eJ[i]	Predikované reziduum eP[i]	Diagonální prvky H[i, i]
1	3.6578E-01	3.5217E-01	1.1343E-01	1.2352E-01
2	6.5601E-01	6.3966E-01	2.0453E-01	1.3298E-01
3	-1.0877E-02	-1.0414E-02	-3.4146E-03	1.4474E-01
4	-3.4147E-01	-3.2854E-01	-1.0620E-01	1.2863E-01
5	-1.7553E-01	-1.6827E-01	-7.3310E-02	5.1683E-01*
6	-1.5911E-01	-1.5249E-01	-5.2019E-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	6.9110E-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	3.3932E-02	3.2489E-02	1.0568E-02	1.3115E-01

Táhněte na blok (7) Indikace vlivných bodů (hvězdičkou *),
kde jsou diagnostiky v číselné formě. Ukončete výstup Esc.

V Y S L E D K Y

Bod	$LD(b)[i]$	$LD(s^2)[i]$	$LD(b,s^2)[i]$
1	2.5120E-02	2.2351E-02	4.6185E-02
2	8.7766E-02	6.2215E-03	9.1797E-02
3	2.6698E-05	3.2606E-02	3.2632E-02
4	2.2933E-02	2.3569E-02	4.5295E-02
5	4.3881E-02	3.0095E-02	7.1396E-02
6	9.0523E-03	3.0538E-02	3.9045E-02
7	8.4238E-01	3.0992E-02	9.5090E-01
8	1.9389E+00	2.4682E-01	2.6917E+00
9	7.9637E-01	1.5878E-01	1.0958E+00
10	1.6079E-01	4.5424E-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

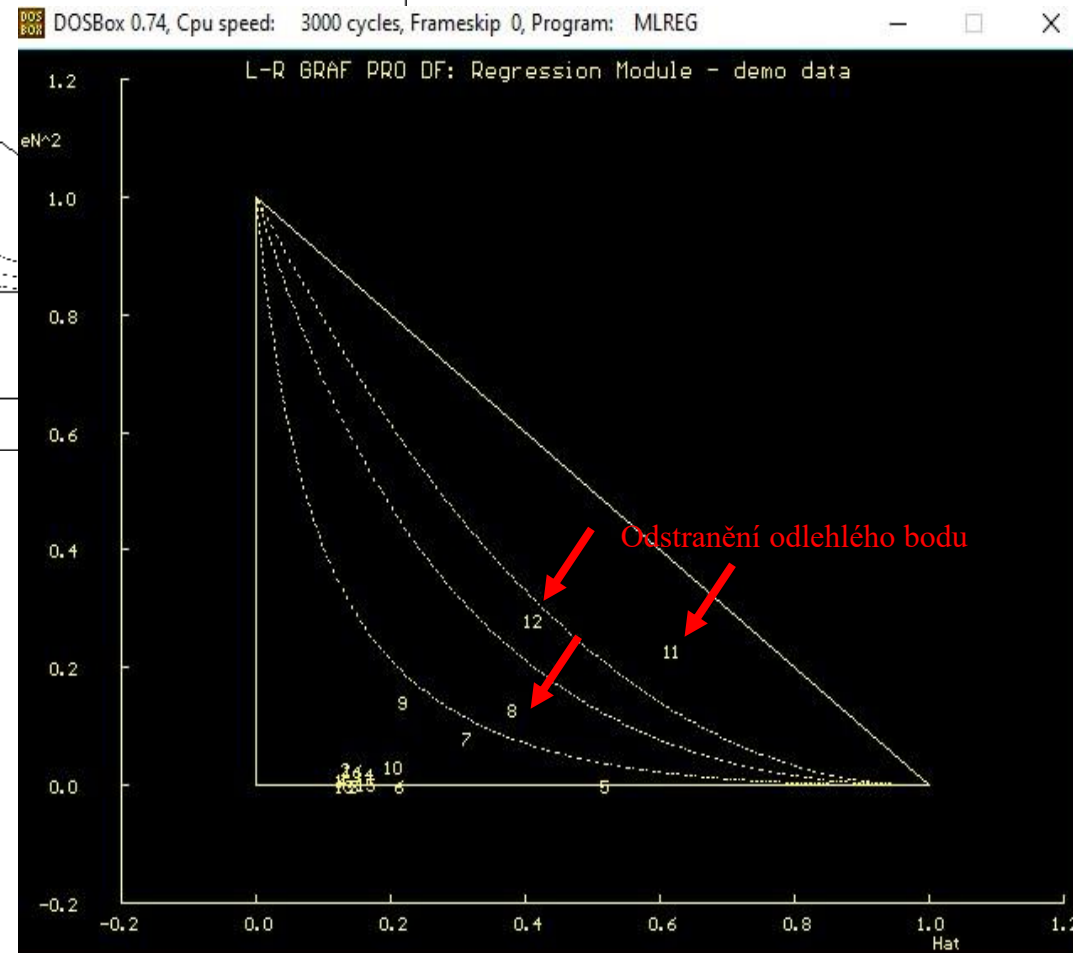
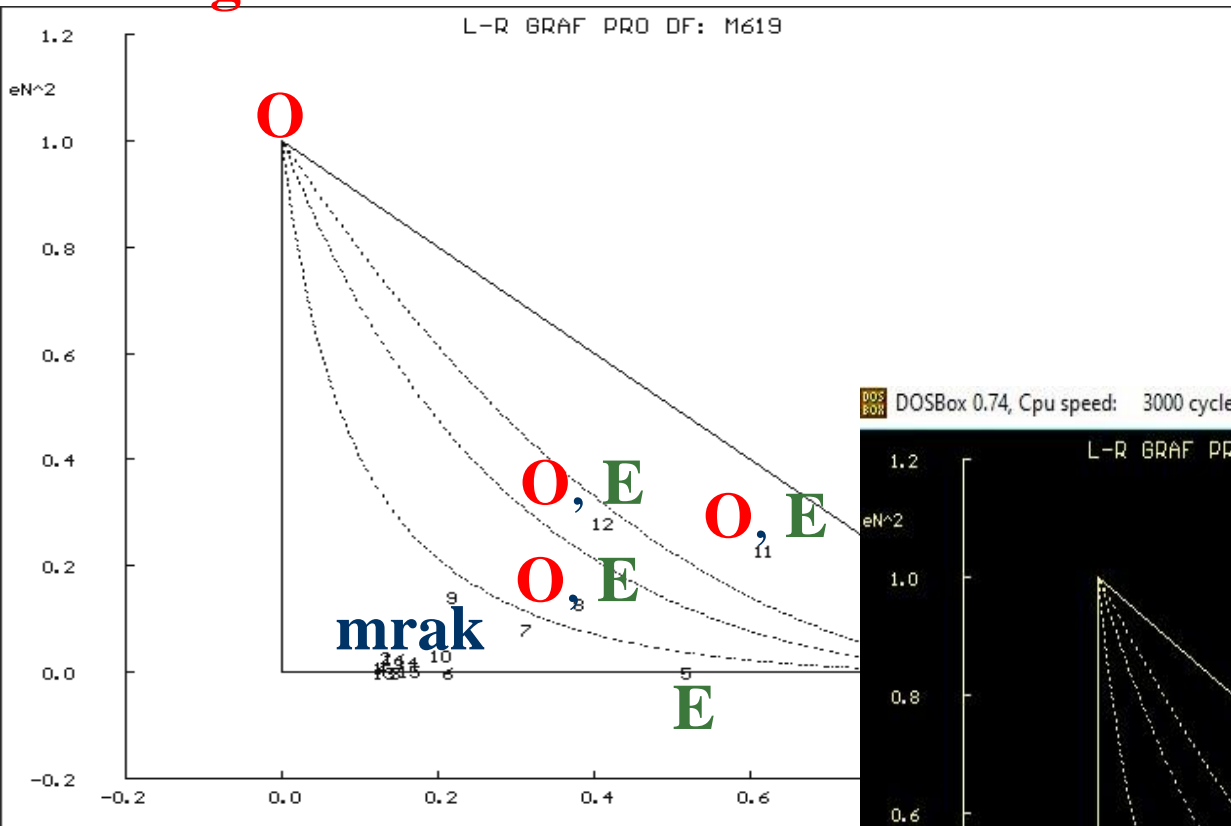
Táhněte dále dolů v bloku (7) Indikace vlivných bodů (hvězdičkou *), kde jsou další diagnostiky v číselné formě. Ukončete výstup Esc.

V Ý S L E D K Y

Bod	Zobecněné diag. prvky	Cookova uzdálenost	Atkinsonova uzdálenost	Vliv na predikci
i	Hm[i, i]	D[i]	A[i]	DF[i]
1	1.3329E-01	4.7137E-03	2.2899E-01	1.3221E-01
2	1.6407E-01	1.6501E-02	4.3390E-01	2.5051E-01
3	1.4475E-01	5.0059E-06	7.4206E-03	-4.2843E-03
4	1.3709E-01	4.3030E-03	2.1863E-01	-1.2622E-01
5	5.1807E-01	8.2390E-03	3.0143E-01	-1.7403E-01
6	2.1318E-01	1.6978E-03	1.3680E-01	-7.8983E-02
7	3.9465E-01	1.6218E-01	1.4226E+00	-8.2134E-01
8	5.1049E-01	3.8649E-01*	2.3192E+00*	1.3390E+00*
9	3.6115E-01	1.5310E-01	1.4358E+00	-8.2894E-01
10	2.3414E-01	3.0299E-02	5.8916E-01	3.4015E-01
11	8.4709E-01*	2.9145E+00*	8.9433E+00*	-5.1634E+00*
12	6.9419E-01*	1.0062E+00*	4.6176E+00*	2.6660E+00*
13	1.3124E-01	4.3450E-05	2.1863E-02	1.2623E-02
14	1.8114E-01	1.4034E-02	3.9779E-01	-2.2966E-01
15	1.6657E-01	2.4587E-03	1.6480E-01	-9.5150E-02
16	1.6888E-01	1.5020E-02	4.1270E-01	2.3827E-01

Táhněte dále dolů v bloku (7) Indikace vlivných bodů (hvězdičkou *), kde jsou další diagnostiky v číselné formě. Ukončete výstup Esc.

9. L-R graf:



A D S T A T - Hlavní menu statistického systému
(c) TriloByte 1990

Jednorozměrná data
Vícerozměrná data
Kalibrace
Lineární regrese
Nelineární regrese
Vyhlazování
Pravděpodobnostní modely
Růstové křivky
Chemometrie
Analýza rozptylu
Konec práce

Ukončete regresní analýzu kliknutím na **K** a **A**.

Úlohy na výstavbu vícerozměrného lineárního regresního modelu

Kritika dat

Jak napsat a vysvětlit úlohu v semestrální práci je
v *Kompendiu*, str. 692

Návod k sestavení a napsání semestrální práce k zadané úloze najdete ve cvičebnici *M. Meloun, J. Militký: Kompendium statistického zpracování dat, Karolinum Praha 1992, ISBN 978-80-246-2196-8*, ve vzorovém příkladu na str. 692 – 705 a obecný výklad outputu v lineární regresi na str. 582 - 583. V ADSTATu použijte na přenos výstupů a obrázků **PrintScreen** a upravte pak negativ na černobílý pozitiv ve MS Windows.

Úloha M6.09 *Závislost obsahu cholesterolu na věku a váze pacienta*

Zadání: Výběr 25 pacientů, nemocných s hyperliproteinemií byl vyšetřován na hladinu lipidů v plasmě totálního cholesterolu y s přihlédnutím k váze x_1 a věku x_2 pacienta.

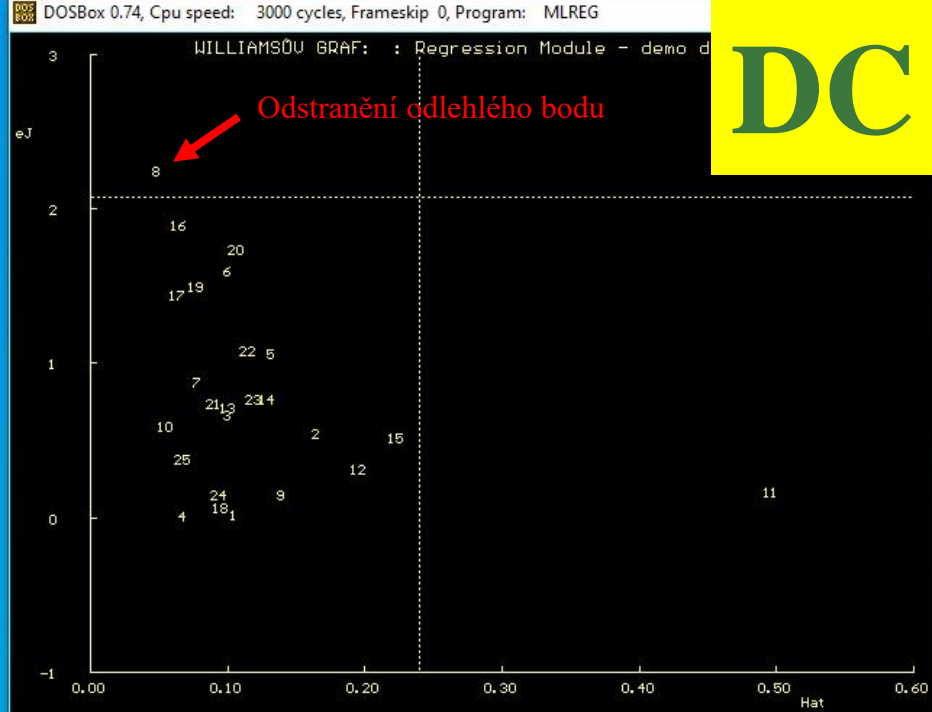
Úkoly: (1) Navrhněte lineární regresní model a testujte statistickou významnost jednotlivých regresních parametrů.

(2) Vyšetřete regresní triplet a soustřed'te se na odhalení vlivných bodů.

(3) Jaké závěry můžete učinit z parciálních regresních a parciálních reziduálních grafů?

Data: Váha x_1 [kg] a věk pacienta x_2 [roky], obsah cholesterolu y [mg/100 ml]:

x_1	x_2	y
84	46	354
...
63	30	244



Kritika regresního modelu

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
BI 01	7.7983E+01	5.2430E+01	1.4874E+00	1.4874E+00	akceptována	0.151
BI 11	4.1736E-01	7.2878E-01	5.7269E-01	5.7269E-01	akceptována	0.573
BI 21	5.2166E+00	7.5724E-01	6.8889E+00	6.8889E+00	Zamítnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 8.3997E-01
 Koeficient determinace, R^2 : 7.0555E-01
 Predikovaný korelační koeficient, R_p^2 : 8.0455E-01
 Střední kvadratická chyba predikce, MEP : 2.0510E+03
 Akaikeho informační kritérium, AIC : 1.9214E+02

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	yexpli	yyppli	s(yyppli)	erli	erli
1	3.5400E+02	3.5300E+02	1.4202E+01	9.9587E-01	2.8132E-01
2	1.9000E+02	2.1278E+02	1.7868E+01	-2.2782E+01	-1.1990E+01
3	4.0500E+02	3.7637E+02	1.3882E+01	2.8626E+01	7.0682E+00

Napověda-F1 Řádek: 80 - 102 Celkem: 253 Délka: 14773

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

F1=náповěda F2=uložení F3=čtení F4=numerický mód F5=zoom ESC=konec

Řádek	8	Sloupec	5	Insert	NUM	M609.txt
8.400000	E+01	4.600000	E+01		3.540000	E+02
7.300000	E+01	2.000000	E+01		1.900000	E+02
6.500000	E+01	5.200000	E+01		4.050000	E+02
7.000000	E+01	3.000000	E+01		2.630000	E+02
7.600000	E+01	5.700000	E+01		4.510000	E+02
6.900000	E+01	2.500000	E+01		3.020000	E+02
6.300000	E+01	2.800000	E+01		2.880000	E+02
7.900000	E+01	5.700000	E+01		4.020000	E+02
7.500000	E+01	4.400000	E+01		3.650000	E+02
2.700000	E+01	2.400000	E+01		2.090000	E+02
8.900000	E+01	3.100000	E+01		2.900000	E+02
6.500000	E+01	5.200000	E+01		3.460000	E+02
5.700000	E+01	2.300000	E+01		2.540000	E+02
5.900000	E+01	6.000000	E+01		3.950000	E+02
6.900000	E+01	4.800000	E+01		4.340000	E+02
6.000000	E+01	3.400000	E+01		2.200000	E+02
7.900000	E+01	5.100000	E+01		3.740000	E+02
7.500000	E+01	5.000000	E+01		3.080000	E+02
8.200000	E+01	3.400000	E+01		2.100000	E+02
5.900000	E+01	4.600000	E+01		3.200000	E+02
6.700000	E+01	2.300000	E+01		1.810000	E+02
8.500000	E+01	3.700000	E+01		2.740000	E+02

Odstranění odlehlého bodu

Kritika regresního modelu

(2) INDIKACE MULTIKOLINEARITY:

Č	Vlastní čísla	Čísla podmí- něnosti	Variance inflation factor	Vícenás. korel. koef pro X _l
[j]	korel. matice	KL[j]	VIF[j]	
1	4.8069E-02	4.0607E+01	1.0000E+00	0.0000
2	1.9519E+00	1.0000E+00	1.0000E+00	0.0000

Maximální číslo podmíněnosti K : 4.0607E+01
 (KL[j], K > 1000 indikuje silnou multikolaritu)
 (VIF[j] > 10 indikuje silnou multikolaritu)

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
BI 01	0.0000E+00					
BI 11	1.1914E+00	4.0085E-01	2.9722E+00	2.9722E+00	Zamítnuta	0.007
BI 21	5.6990E+00	6.7969E-01	8.3847E+00	8.3847E+00	Zamítnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 8.5027E-01
 Koeficient determinace, R^2 : 7.2296E-01
 Predikovaný korelační koeficient, R_p^2 : 8.2267E-01
 Střední kvadratická chyba predikce, MEP : 1.8805E+03
 Akaikeho informační kritérium, AIC : 1.8116E+02

Napověda-F1 Řádek: 71 - 93 Celkem: 249 Délka: 14437

Závěry a vysvětlení výstupu úlohy M609:

Úkoly

(1) Navrhněte lineární regresní model a testujte statistickou významnost jednotlivých regresních parametrů.

(2) Vyšetřete regresní triplet a soustřeďte se na odhalení vlivných bodů.

(3) Jaké závěry můžete učinit z parciálních regresních a parciálních reziduálních grafů?

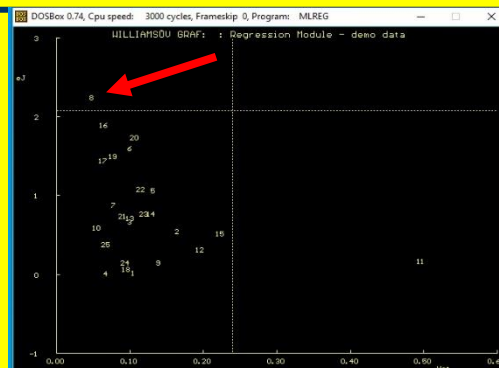
Odpovědi, závěry a vysvětlení

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

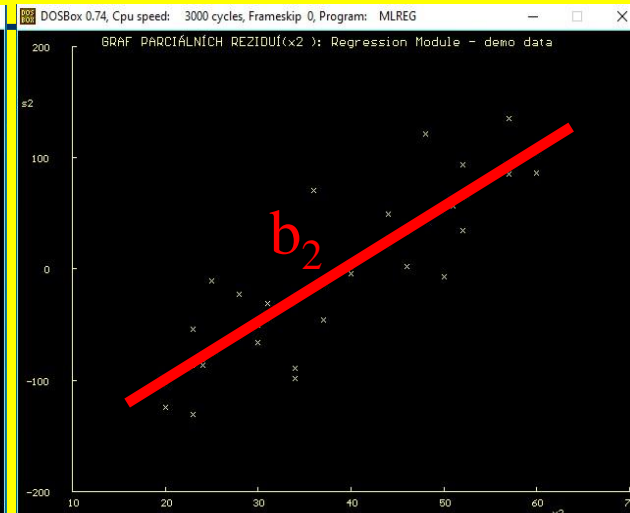
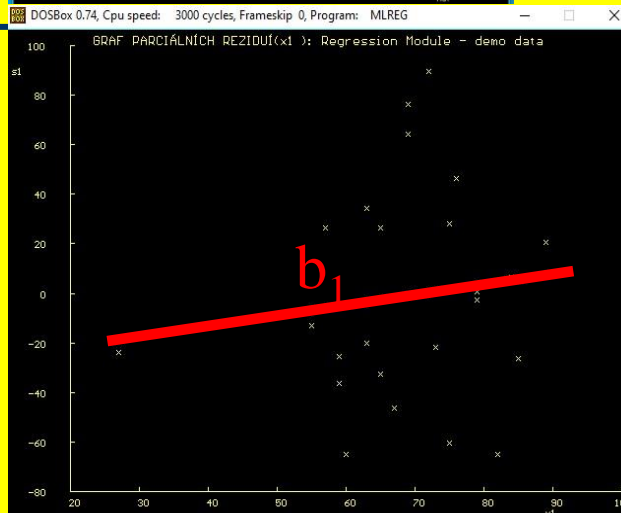
Parametr	Odhad	Směrodatná odchylna	Test $H_0: \beta_{[j]} = 0$ vs. $H_A: \beta_{[j]} \neq 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
$\beta_{[0]}$	0.0000E+00					
$\beta_{[1]}$	1.1914E+00	4.0085E-01		2.9722E+00	Zamítnuta	0.007
$\beta_{[2]}$	5.6990E+00	6.7969E-01		8.3847E+00	Zamítnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 8.5027E-01
Koeficient determinace, R^2	: 7.2296E-01
Predikovaný korelační koeficient, R_p^2	: 8.2267E-01
Střední kvadratická chyba predikce, MEP	: 1.8805E+03
Akaikeho informační kritérium, AIC	: 1.8116E+02



V datech je 1 odlehlý bod č. 8



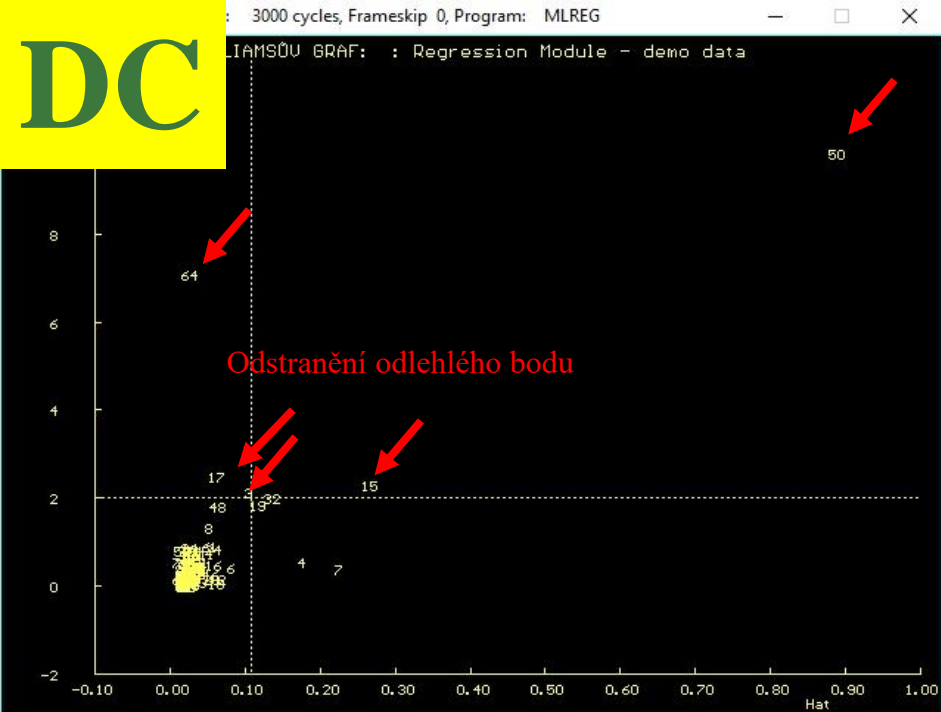
Úloha M6.15 *Vliv obsahu PCB v kongenerech na PCB v mateřském mléce*

Zadání: V tuku mateřského mléka 74 matek byl stanoven obsah polychlorova-ných bifenylyů, a to celkový obsah suma PCB y , a jednak jako obsah PCB jednotlivých kongenerů x_1 , x_2 , x_3 .

- Úkoly:** (1) Vyšetřete lineární regresní model mezi sumou PCB y a obsahem kongenerů PCB#138 x_1 , PCB#153 x_2 a PCB#180 x_3
(2) Posud'te predikční schopnost regresního modelu a spolehlivost predikce celkového obsahu PCB na základě znalosti tří kongenerů.
(3) Vyšetřete regresní triplet a testujte statistickou významnost jednotlivých kongenerů.

Data: Obsah PCB#138 x_1 [mg/kg], obsah PCB#153 x_2 [mg/kg], obsah PCB#180 x_3 [mg/kg], suma PCB y [mg/kg].:

x_1	x_2	x_3	y
0.184	0.218	0.228	1.737
...
0.107	0.186	0.081	1.031



Kritika regresního modelu

(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.
BI 01	2.8006E-01	7.5768E-02	3.6963E+00	Zamítnuta	0.000	0.000
BI 11	2.1801E+00	6.0059E-01	3.6300E+00	Zamítnuta	0.001	0.001
BI 21	-1.7419E+00	7.5235E-01	-2.3153E+00	Zamítnuta	0.024	0.024
BI 31	7.1782E+00	3.8305E-01	1.8740E+01	Zamítnuta	0.000	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 9.5870E-01
 Koeficient determinace, R² : 9.1911E-01
 Predikovaný korelační koeficient, Rp² : 7.0076E-01
 Střední kvadratická chyba predikce, MEP : 6.1223E-01
 Akaikeho informační kritérium, AIC :-1.6441E+02

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	yexp[i]	yvyp[i]	s(yvyp[i])	e[i]	er[i]
1	1.7370E+00	1.9381E+00	4.4975E-02	-2.0108E-01	-1.1576E+01
2	2.0900E+00	2.0398E+00	8.4177E-02	5.0193E-02	2.4016E+00

Napověda-F1 Řádek: 149 - 171 Celkem: 519 Délka: 33365

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

F1=nápověda F2=uložení F3=čtení F4=numerický mód F5=zoom ESC=konec

Řádek 65	Sloupec 2	Insert	NUM	M615.txt
8.100000 E-02	1.220000 E-01	4.600000 E-02	6.870000 E-01	
7.700000 E-02	1.000000 E-01	7.900000 E-02	7.060000 E-01	
8.400000 E-02	1.220000 E-01	8.500000 E-02	8.030000 E-01	
2.200000 E-02	1.510000 E-01	7.500000 E-02	6.840000 E-01	
1.070000 E-01	1.710000 E-01	1.300000 E-01	1.125000	
3.280000 E-01	4.650000 E-01	3.270000 E-01	3.089000	
1.000000 E-01	1.160000 E-01	1.000000 E-01	8.710000 E-01	
1.050000 E-01	1.610000 E-01	1.280000 E-01	1.087000	
1.480000 E-01	1.790000 E-01	1.370000 E-01	1.280000	
7.200000 E-02	9.900000 E-02	5.800000 E-02	6.320000 E-01	
1.150000 E-01	1.300000 E-01	1.010000 E-01	9.540000 E-01	
1.180000 E-01	1.430000 E-01	1.170000 E-01	1.042000	
2.600000 E-01	3.510000 E-01	3.510000 E-01	2.653000	
1.820000 E-01	2.120000 E-01	1.940000 E-01	1.622000	
2.940000 E-01	3.320000 E-01	2.540000 E-01	2.427000	
2.970000 E-01	3.740000 E-01	3.820000 E-01	2.904000	
2.020000 E-01	2.430000 E-01	2.370000 E-01	1.880000	
3.150000 E-01	3.370000 E-01	3.860000 E-01	2.863000	
1.660000 E-01	1.920000 E-01	1.680000 E-01	1.451000	
5.900000 E-02	7.600000 E-02	5.000000 E-02	5.100000 E-01	
1.980000 E-01	2.530000 E-01	2.330000 E-01	1.886000	

Odstranění odlehlého bodu

Odstranění odlehlého bodu

Kritika regresního modelu

(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.
BI 01	-3.0851E-04	6.2036E-04	-4.9731E-01	Akceptována	0.621	0.000
BI 11	2.7553E+00	4.7862E-03	5.7567E+02	Zamítnuta	0.000	0.000
BI 21	2.7542E+00	7.5621E-03	3.6421E+02	Zamítnuta	0.000	0.000
BI 31	2.7647E+00	7.2723E-03	3.8018E+02	Zamítnuta	0.000	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 1.0000E+00
 Koeficient determinace, R² : 1.0000E+00
 Predikovaný korelační koeficient, Rp² : 1.0000E+00
 Střední kvadratická chyba predikce, MEP : 5.1421E-06
 Akaikeho informační kritérium, AIC :-8.4911E+02

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	yexp[i]	yvyp[i]	s(yvyp[i])	e[i]	er[i]
1	1.7370E+00	1.7374E+00	4.5111E-04	-4.3179E-04	-2.4858E-02
2	2.0900E+00	2.0899E+00	6.3208E-04	9.6880E-05	4.6354E-03

Napověda-F1 Řádek: 143 - 165 Celkem: 497 Délka: 31897

Závěry a vysvětlení výstupu úlohy M615:

Úkoly

Odpovědi, závěry a vysvětlení

(1) Vyšetřete lineární regresní model mezi sumou PCB y a obsahem kongenerů PCB#138 x_1 , PCB#153 x_2 a PCB#180 x_3

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
BI 01	-3.0851E-04	6.2036E-04		-4.9731E-01	Akceptována	0.621
BI 11	2.7553E+00	4.7862E-03		5.7567E+02	Zamítnuta	0.000
BI 21	2.7542E+00	7.5621E-03		3.6421E+02	Zamítnuta	0.000
BI 31	2.7647E+00	7.2723E-03		3.8018E+02	Zamítnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 1.0000E+00
Koeficient determinace, R^2	: 1.0000E+00
Predikovaný korelační koeficient, R_p^2	: 1.0000E+00
Střední kvadratická chyba predikce, MEP	: 5.1421E-06
Akaikeho informační kritérium, AIC	:-8.4911E+02

(2) Posuďte predikční schopnost regresního modelu a spolehlivost predikce celkového obsahu PCB na základě znalosti tří kongenerů.

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
BI 01	0.0000E+00					
BI 11	1.3705E+00	6.0704E-01		2.2576E+00	Zamítnuta	0.027
BI 21	-4.4057E-01	7.2175E-01		-6.1042E-01	Akceptována	0.544
BI 31	7.3934E+00	4.1097E-01		1.7990E+01	Zamítnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.5043E-01
Koeficient determinace, R^2	: 9.0333E-01
Predikovaný korelační koeficient, R_p^2	: 6.9106E-01
Střední kvadratická chyba predikce, MEP	: 6.2847E-01
Akaikeho informační kritérium, AIC	:-1.5621E+02

(3) Vyšetřete regresní triplet a testujte statistickou významnost jednotlivých kongenerů.

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
BI 01	0.0000E+00					
BI 11	1.0323E+00	2.4718E-01		4.1763E+00	Zamítnuta	0.000
BI 21	7.2123E+00	2.8311E-01		2.5475E+01	Zamítnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.5017E-01
Koeficient determinace, R^2	: 9.0282E-01
Predikovaný korelační koeficient, R_p^2	: 8.2392E-01
Střední kvadratická chyba predikce, MEP	: 3.8633E-01
Akaikeho informační kritérium, AIC	: 1.5736E+02

Úloha M6.49 *Závislost vlhkosti vzduchu na měřených klimatických charakteristikách*

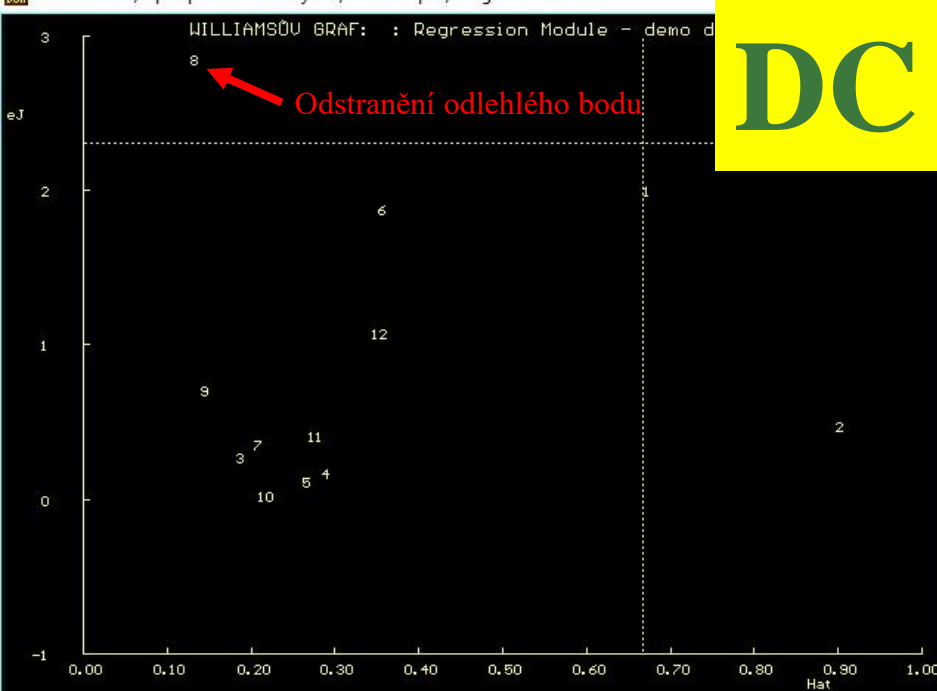
Zadání: Automatická klimatická stanice NOEL zaznamenává v lesním porostu kontinuálně zvolené charakteristiky. Ve výšce 2 m nad zemí jsou umístěná čidla měřící teplotu vzduchu [$^{\circ}\text{C}$], směr větru [stupně], rychlost větru [$\text{m}\cdot\text{s}^{-1}$] a vlhkost vzduchu [%].

Úkoly:

- (1) Navrhněte lineární regresní model, který by vystihoval závislost vlhkosti vzduchu [%] na ostatních měřených charakteristikách.
- (2) Vysvětlete všech sedm předpokladů metody nejmenších čtverců a řešení regresního tripletu.
- (3) Která kritéria jsou nejvhodnější při hledání lineárního regresního modelu?

Data: Počet $n = 12$, teplota vzduchu x_1 [$^{\circ}\text{C}$], směr větru x_2 [stupně], rychlost větru x_3 [$\text{m}\cdot\text{s}^{-1}$], vlhkost vzduchu y [%].

x_1	x_2	x_3	y
13.7	90.5	0.3	89
...
6.3	0.6	0.04	96



Kritika regresního modelu

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] <> 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
BI 01	1.1692E+02	4.7901E+00		2.4410E+01	Zamítnuta	0.000
BI 11	-2.8957E+00	4.0824E-01		-7.0932E+00	Zamítnuta	0.000
BI 21	-1.9111E-03	2.9857E-02		-6.4010E-02	Akceptována	0.951
BI 31	2.5642E+01	2.5071E+01		1.0228E+00	Akceptována	0.336

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 9.3147E-01
 Koeficient determinace, R^2 : 8.6764E-01
 Predikovaný korelační koeficient, R_p^2 : 7.8127E-01
 Střední kvadratická chyba predikce, MEP : 3.6658E+01
 Akaikeho informační kritérium, AIC : 3.8265E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	yexp[i]	yvyp[i]	s(yvyp[i])	e[i]	er[i]
1	8.9000E+01	8.4772E+01	3.5421E+00	4.2284E+00	4.7510E+00
2	8.2700E+01	8.3384E+01	4.1019E+00	-6.8354E-01	-8.2653E-01

Napověda-F1 Řádek: 77 - 99 Celkem: 199 Délka: 10829

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

F1=nápoředa F2=uloření F3=čtení F4=numerický mód F5=zoom ESC=konec

Řádek	Sloupec	1
1	3	1.370000 E+01
1	4	8.900000 E+01
1	5	1.490000 E+01
1	6	8.270000 E+01
1	7	1.540000 E+01
1	8	7.710000 E+01
1	9	1.570000 E+01
1	10	7.180000 E+01
1	11	1.540000 E+01
1	12	7.290000 E+01
1	13	1.560000 E+01
1	14	6.990000 E+01
1	15	1.450000 E+01
1	16	7.450000 E+01
1	17	9.800000
1	18	9.250000 E+01
1	19	8.100003
1	20	9.440000 E+01
1	21	7.200000
1	22	9.550000 E+01
1	23	6.300000
1	24	9.600000 E+01

Odstranění odlehlého bodu

Kritika regresního modelu

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] <> 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
BI 01	1.1464E+02	4.9868E+00		2.2988E+01	Zamítnuta	0.000
BI 11	-2.5235E+00	3.8586E-01		-6.5401E+00	Zamítnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 9.0894E-01
 Koeficient determinace, R^2 : 8.2616E-01
 Predikovaný korelační koeficient, R_p^2 : 8.7356E-01
 Střední kvadratická chyba predikce, MEP : 2.3402E+01
 Akaikeho informační kritérium, AIC : 3.5277E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	yexp[i]	yvyp[i]	s(yvyp[i])	e[i]	er[i]
1	8.9000E+01	8.0065E+01	1.4672E+00	8.9347E+00	1.0039E+01
2	8.2700E+01	7.7037E+01	1.6808E+00	5.6630E+00	6.8476E+00
3	7.7100E+01	7.5775E+01	1.7977E+00	1.3247E+00	1.7182E+00
4	7.1800E+01	7.5018E+01	1.8739E+00	-3.2182E+00	-4.4822E+00

Napověda-F1 Řádek: 64 - 86 Celkem: 180 Délka: 9468

Závěry a vysvětlení výstupu úlohy M649:

Úkoly

Odpovědi, závěry a vysvětlení

(1) Navrhněte lineární regresní model, který by vystihoval závislost vlhkosti vzduchu [%] na ostatních měřených charakteristikách.

```
(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:
Parametr      Odhad      Směrodatná      Test H0: B[i,j] = 0 vs. HA: B[i,j] <> 0
              odchylka      t-kriterium      hypoteza H0 je      Hlad. úzj.
BI 01         1.1692E+02  4.7901E+00      2.4410E+01          Zamítnuta          0.000
BI 11         -2.8957E+00  4.0824E-01      -7.0932E+00         Zamítnuta          0.000
BI 21         -1.9111E-03  2.9857E-02      -6.4010E-02         Akceptována       0.951
BI 31         2.5642E+01  2.5071E+01      1.0228E+00          Akceptována       0.336

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:
Vícenásobný korelační koeficient, R          : 9.3147E-01
Koeficient determinace, R^2                 : 8.6764E-01
Predikovaný korelační koeficient, Rp^2      : 7.8127E-01
Střední kvadratická chyba predikce, MEP     : 3.6658E+01
Akaikeho informační kritérium, AIC          : 3.8265E+01
```

(2) Vysvětlete všech sedm předpokladů metody nejmenších čtverců a řešení regresního tripletu.

```
(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:
Parametr      Odhad      Směrodatná      Test H0: B[i,j] = 0 vs. HA: B[i,j] <> 0
              odchylka      t-kriterium      hypoteza H0 je      Hlad. úzj.
BI 01         1.1464E+02  4.9868E+00      2.2988E+01          Zamítnuta          0.000
BI 11         -2.5235E+00  3.8586E-01      -6.5401E+00         Zamítnuta          0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:
Vícenásobný korelační koeficient, R          : 9.0894E-01
Koeficient determinace, R^2                 : 8.2616E-01
Predikovaný korelační koeficient, Rp^2      : 8.7356E-01
Střední kvadratická chyba predikce, MEP     : 2.3402E+01
Akaikeho informační kritérium, AIC          : 3.5277E+01
```

(3) Která kritéria jsou nejvhodnější při hledání lineárního regresního modelu?

```
DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG
U V S L E D K Y
Odhad směrodatné odchylky reziduí, s(e)      : 3.4100E-01
Odhad šikmosti reziduí, g1(e)               : -2.5209E+00
Odhad špičatosti reziduí, g2(e)            : 1.3995E+01

(6) TESTOVÁNÍ REGRESNÍHO TRIPLETU (DATA + MODEL + METODA):
Fisher-Snedcorův test významnosti regrese, F : 6.6889E+02
Tabulkový kvantil, F(1-alpha, m-1, n-m)     : 3.9739E+00
Závěr: Navržený model je přijat jako významný.
Spočtená hladina významnosti                 : 0.000

Scottovo kritérium multikolinearity, M       : 3.3498E-01
Závěr: Navržený model není korektní.

Cook-Weisbergův test heteroskedasticity, Sf   : 1.2463E+03
Tabulkový kvantil, Chi^2(1-alpha, 1)       : 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.
Spočtená hladina významnosti                 : 0.000

Jarque-Berraův test normality reziduí, L(e)  : 4.5112E+02
Tabulkový kvantil, Chi^2(1-alpha, 2)       : 5.9915E+00
Závěr: Normalita není přijata.
Spočtená hladina významnosti                 : 0.000
Napověda-F1   Řádek: 229 - 251                Celkem: 499   Délka: 31851
```

R, D, MEP, AIC