

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

Kritika metody

Ověření předpokladů metody nejmenších čtverců

Metoda:

Vyšetření heteroskedasticity

Vyšetření autokorelace

Vyšetření multikolinearity

Vyšetření normality náhodných chyb

Vyšetření omezení parametrů

Vyšetření trendů reziduí

ZÁVAŽNÉ PŘEDPOKLADY MNČ

- kdy je určitě nutné korigovat klasickou MNČ

- 1) **Regresní parametry β** mohou teoreticky nabývat libovolných hodnot.
- 2) Regresní model je **lineární v parametrech**.
- 3) Jednotlivé nezávislé proměnné jsou skutečně **vzájemně nezávislé**, tedy mezi nimi nedochází k tzv. **multikolinearitě**.
- 4) **Podmíněný rozptyl $D(y/x) = \sigma^2$** je konstantní (tzv. podmínka homoskedasticity).
- 5) **Náhodné chyby** mají nulovou střední hodnotu $E(\varepsilon_i) = 0$, mají konečný rozptyl $E(\varepsilon_i^2) = \sigma^2$ a jsou nekorelované.

REGRESNÍ DIAGNOSTIKA

testy vybraných předpokladů klasické MNČ

Multikolinearita: VIF diagnostika indikuje

Heteroskedasticita: testy heteroskedasticity (např. Cook-Weisberg)

Autokorelace reziduí: test významnosti autokorelačního koeficientu

Normalita reziduí: testy normality

Úloha P638

na výstavbu lineárního regresního modelu regresním tripletem

Test heteroskedasticity

Jak vysvětlit výsledky úlohy a napsat 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.

Příklad 6.38 *Test autokorelace u kinetických dat*

Byla sledována kinetika inverze sacharidu v 1M HCl při 30 °C. Lze očekávat, že mezi chybami ε_i existuje autokorelace jako důsledek postupu odebrání vzorků. Určete graficky významnost autokorelace v datech.

Data: x je čas [min]; y je desetinásobek logaritmu podílu zbylého nezreagovaného sacharidu v reakční směsi.

x	y
0	1.000
10	0.954
20	0.895
30	0.843
40	0.791
50	0.735
60	0.685
70	0.628
80	0.581

Ověření předpokladů MNČ

1. Hetero/homoskedasticita (nekonstantnost rozptylu)

Rozptyl veličiny y_i v itém bodě je popsán

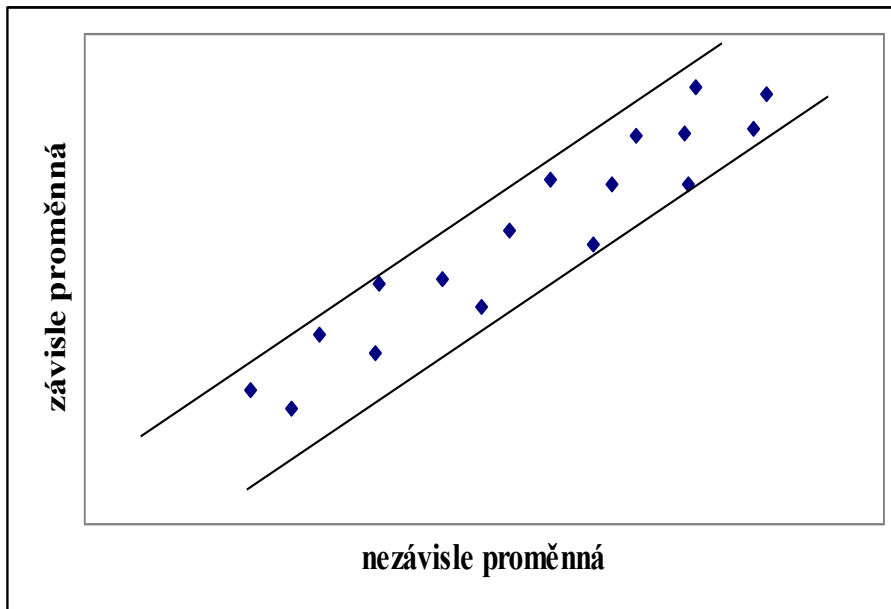
$$\sigma_i^2 = \sigma^2 \exp(\lambda \mathbf{x}_i \boldsymbol{\beta}),$$

kde \mathbf{x}_i je i -tý řádek matice \mathbf{X} .

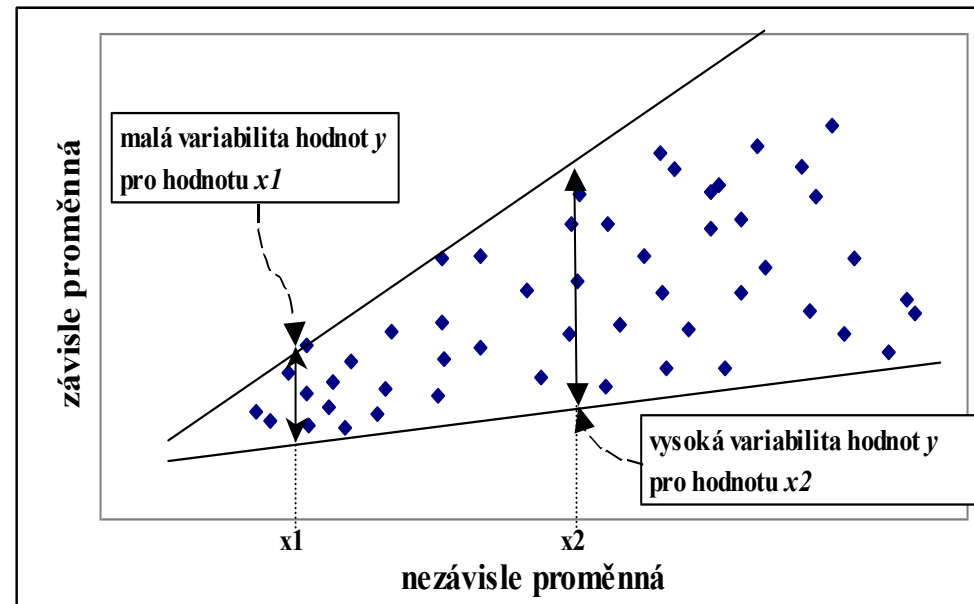
HOMOSKEDASTICITA vs. HETEROSKEDASTICITA

Grafické vysvětlení principu

Homoskedasticita znamená, že hodnoty závisle proměnné y mají pro všechny hodnoty nezávisle proměnné x **konstantní rozptyl** (variabilitu, proměnlivost).



Homoskedasticita



Heteroskedasticita

Cookův - Weisbergův test homoskedasticity

Vycházíme z předpokladu, že rozptyl naměřené hodnoty y_i je určitou funkcí proměnné x_i , β (např. exponenciální funkcí).

$$S_f = \frac{\left[\sum_{i=1}^n (y'_i - \bar{y}')^2 e_i^2 \right]^2}{2 \cdot \sigma^4 \sum_{i=1}^n (y'_i - \bar{y}')^2}$$

Test: Pokud v datech není heteroskedasticita potom platí, že $S_f < \chi^2(1)$.

Test: ověření $H_0: \lambda = 0$ (homoskedasticita).

Cook-Weisbergovo testační kritérium

$$S_f = \frac{\left[\sum_{i=1}^n (\hat{y}_i - \hat{y}_p) \hat{e}_i^2 \right]^2}{2 \hat{\sigma}^4 \sum_{i=1}^n (\hat{y}_i - \hat{y}_p)^2}$$

kde

$$\hat{y}_p = \frac{1}{n} \sum_{i=1}^n \hat{y}_i$$

Testování:

1. Je-li $S_f < \chi^2(1)$, H_0 (homoskedasticita) je přijata.

2. Pro homoskedasticitu tvoří diagnostický graf

$$\hat{e}_{Si}^2 \text{ na } (1 - H_{ii}) \hat{y}_i$$

náhodný mrak bodů. Pro heteroskedasticitu vznikne
typický klínový obrazec

HETEROSKEDASTICITA

Vhodné řešení v lineární regresi při nesplnění jednoho předpokladu klasické MNČ

Nejčastějším způsobem je užití metody **vážených nejmenších čtverců**,

kdy se účelová funkce čili suma čtverců reziduí násobí

vhodně zvolenými váhami

$$U(\mathbf{b}) = \sum_{i=1}^n \left[y_i \mathbf{V}_{ii} - \sum_{j=1}^m \mathbf{V}_{ii} x_{ij} b_j \right]^2$$

V běžných případech je možné jako váhy volit

hodnoty $1/y_i$ nebo $1/y_i^2$.

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 1 Sloupec 5 Insert NUM P638

Textový editor

```
P638x P638y
0.00000 1.00000
10.00000
20.00000
30.00000
40.00000 0.79100
50.00000 0.73500
60.00000 0.68500
70.00000 0.62800
80.00000 0.58100
```

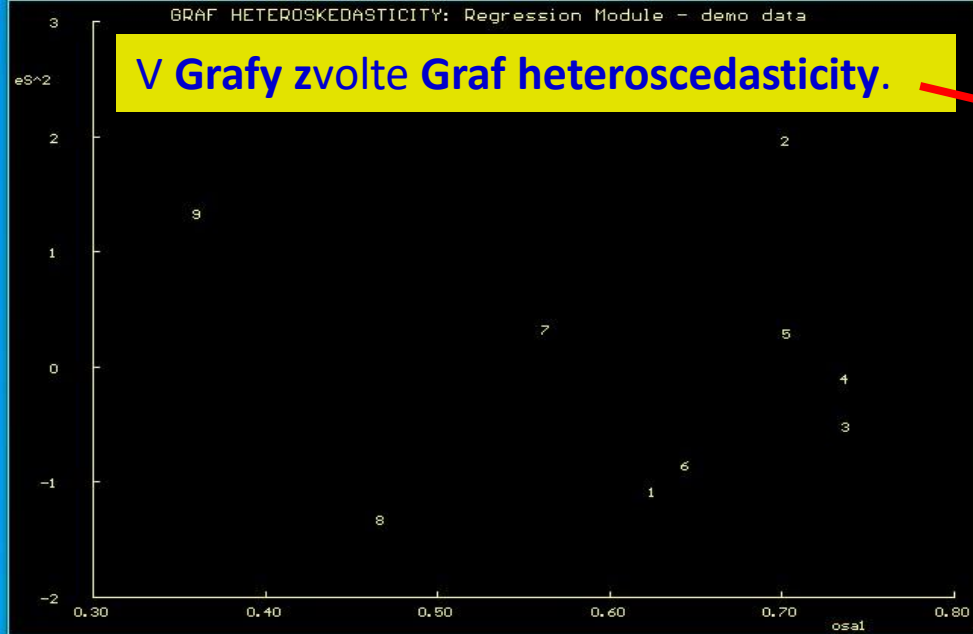
Čtení ze souboru: P638

Otevřít Data, pak stisk F3, pak smazat demo.dat, napsat P638 a ukončit Esc a A.



V Grafy zvolte Residual-Index plot. Vložte síť mezníkem. Pak Esc zpět.

Uhledávání chyb: F6=první F7=předchozí F8=následující F9=poslední



V Grafy zvolte Graf heteroscedasticity.

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

U V Ý S L E D K Y

Scottovo kritérium multikolinearity, M : -1.5384E-13
Závěr: Navržený model je korektní.

Cook-Weisbergův test heteroskedasticity, Sf : 1.3713E+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) : 6.5511E-01
Tabulkový kvantil, Chi^2(1-alpha,2) : 5.9915E+00
Závěr: Normalita je přijata.
Spočtená hladina významnosti : 0.721

Waldův test autokorelace, Wa : 1.1740E+01
Tabulkový kvantil, Chi^2(1-alpha,1) : 3.8415E+00
Závěr: Rezidua jsou autokorelována.
Spočtená hladina významnosti : 0.001

Znamékový test, Dt : 6.1791E-01
Tabulkový kvantil, N(1-alpha/2) : 1.6449E+00
Závěr: Rezidua nevykazují trend.
Spočtená hladina významnosti : 0.268

Napověda-F1 Řádek: 103 - 125 Celken: 164 Délka: 8452

Úloha P638

na výstavbu lineárního regresního modelu regresním tripletem

Test autokorelace

Jak vysvětlit výsledky úlohy a napsat 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 používejte na přenos výstupů a obrázků **PrintScreen** a upravte pak negativ na černobílý pozitiv ve MS Windows.

2. Autokorelace

Data časových řad mají chyby ε_i vzájemně korelované.

Nejčastější je případ autokorelace prvního řádu

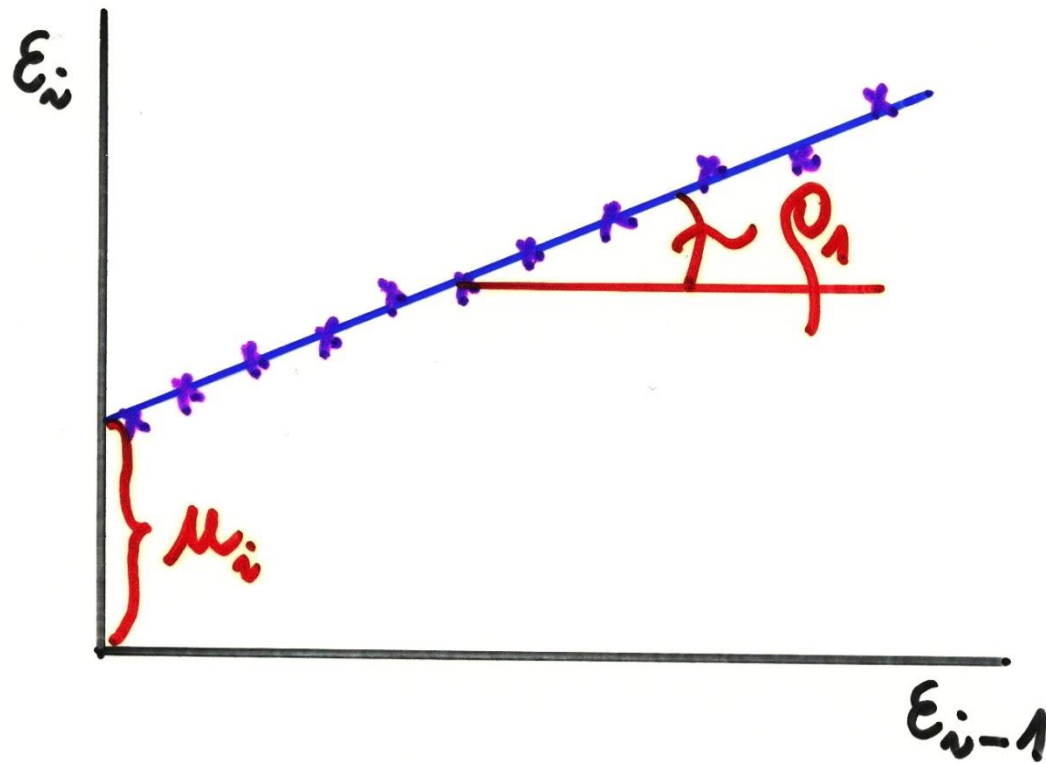
$$\varepsilon_i = \rho_1 \varepsilon_{i-1} + u_i$$

kde $u_i \sim N(0, \sigma^2)$.

- a) Pro $\rho_1 = 1$ případ kumulativních chyb, který se v chemii vyskytuje často.
- b) Pro $\rho_1 \leq 1$ jde o autokorelační koeficient 1. řádu.

Test:

1. Grafická indikace autokorelace:



2. Waldův test pro ρ_1 : $H_0: \rho_1 = 0$ vs. $H_A: \rho_1 \neq 0$

Je-li Waldovo kritérium $W_a = \frac{n \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} < \chi^2(1)$, H_0 je přijata.

Příklad 6.38 *Test autokorelace u kinetických dat*

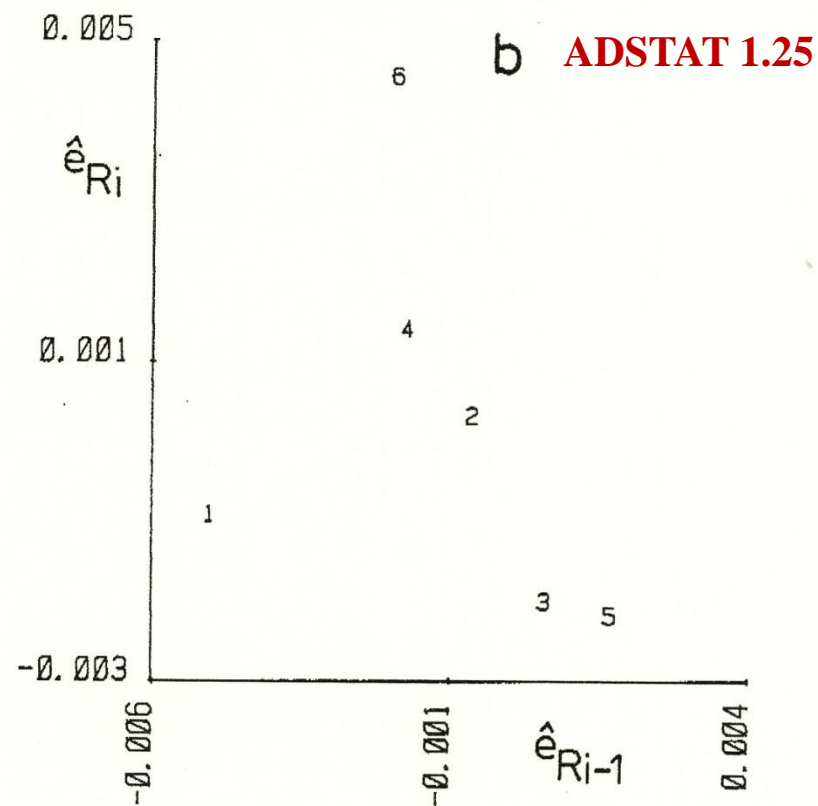
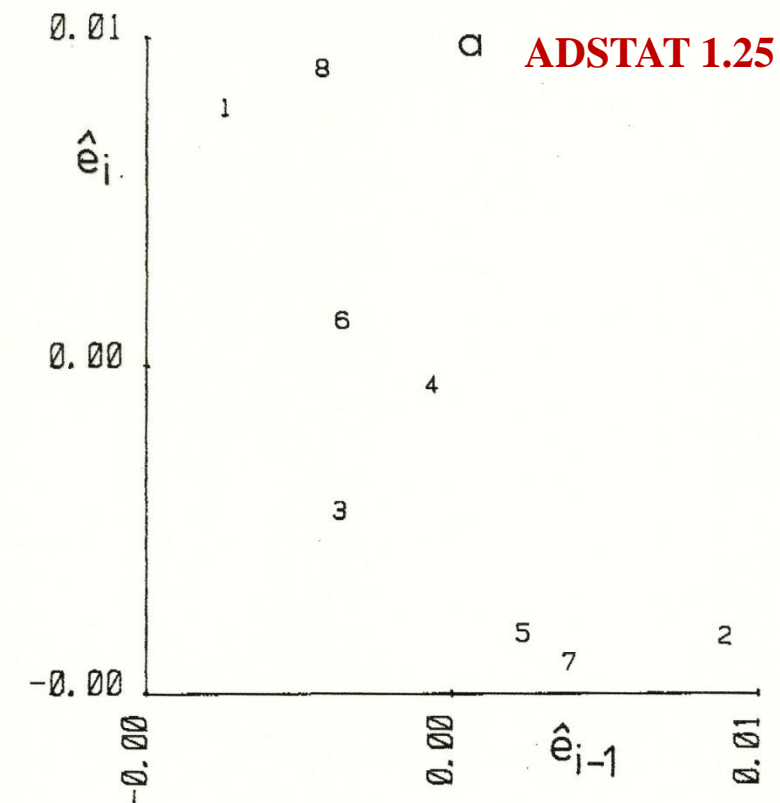
Byla sledována kinetika inverze sacharidu v 1M HCl při 30 °C. Lze očekávat, že mezi chybami ε_i existuje autokorelace jako důsledek postupu odebrání vzorků. Určete graficky významnost autokorelace v datech.

Data: x je čas [min]; y je desetinásobek logaritmu podílu zbylého nezreagovaného sacharidu v reakční směsi.

x	y
0	1.000
10	0.954
20	0.895
30	0.843
40	0.791
50	0.735
60	0.685
70	0.628
80	0.581

Řešení:

1. Graf závislosti \hat{e}_i na \hat{e}_{i-1} a graf závislosti \hat{e}_{Ri} na \hat{e}_{Ri-1} ukazují výraznou negativní autokorelaci.
2. Waldovo testační kritérium autokorelace $W_a = 9.424$ je větší než $\chi_{0.95}^2(2) = 5.99$, což indikuje autokorelaci reziduí.



Závěr: Grafy umožňují odhadnout znaménko a velikost autokorelačního koeficientu 1. řádu ρ_1 .

```

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG
F1=nápoředa F2=uloženi F3=čteni F4=numerický mód F5=zoom ESC=konec
Řádek 1 Sloupec 5 Insert NUM P638
Textový editor
P638x P638y
0.00000 1.00000
10.00000
20.00000 P638
30.00000
40.00000 0.79100
50.00000 0.73500
60.00000 0.68500
70.00000 0.62800
80.00000 0.58100
Čtení ze souboru:

```

Otevřít Data, pak stisk F3, pak smazat demo.dat, napsat P638 a ukončit Esc a A.

```

Uyhledávání chyb: F6=první F7=předchozí F8=následující F9=poslední
DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG

```

```

U Ý S L E D K Y
(6) TESTOVÁNÍ REGRESNÍHO TRIPLETU (DATA + MODEL + METODA):
Fisher-Snedocorův test významnosti regrese, F : 2.2084E+04
Tabulkový kvantil, F(1-alpha,m-1,n-m) : 5.5914E+00
Závěr: Navržený model je přijat jako významný.
Spočtená hladina významnosti : 0.000

Scottovo kritérium multikolinearity, M : -7.5529E-14
Závěr: Navržený model je korektní.

Cook-Weisbergův test heteroskedasticity, Sf : 1.0050E+01
Tabulkový kvantil, Chi^2(1-alpha,1) : 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.
Spočtená hladina významnosti : 0.002

Jarque-Berraův test normality reziduí, L(e) : 6.8980E-01
Tabulkový kvantil, Chi^2(1-alpha,2) : 5.9915E+00
Závěr: Normalita je přijata.
Spočtená hladina významnosti : 0.708

Waldův test autokorelace, Wa : 9.4237E+00
Tabulkový kvantil, Chi^2(1-alpha,1) : 3.8415E+00
Závěr: Rezidua jsou autokorelována.
Spočtená hladina významnosti : 0.002
Napověda-F1 Řádek: 98 - 120 Celkem: 168 Délka: 8777

```

```

DOSBox 0.74, Cpu speed: 3000 cycles, Frameskip 0, Program: MLREG
U Ý S L E D K Y
(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:
Parametr Odhad Směrodatná Test H0: B[i,j] = 0 vs. HA: B[i,j] <> 0
odchylka t-kritérium hypoteza H0 je Hlad. výz.
BI 01 1.0024E+00 1.6990E-03 5.8996E+02 Zamítnuta 0.000
BI 11 -5.3033E-03 3.5687E-05 -1.4861E+02 Zamítnuta 0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:
Vícenásobný korelační koeficient, R : 9.9984E-01
Koeficient determinace, R^2 : 9.9968E-01
Predikovaný korelační koeficient, Rp^2 : 9.9969E-01
Střední kvadratická chyba predikce, MEP : 1.1646E-05
Akaikeho informační kritérium, AIC : -1.0430E+02

(5) ANALÝZA KLASICKÝCH REZIDUÍ:
Bod Meřená Predikovaná Směrodatná Klasické Relativní
hodnota hodnota odchylka reziduum reziduum
i yexp[i] yvyp[i] s(yvyp[i]) e[i] er[i]
1 1.0000E+00 1.0024E+00 1.6990E-03 -2.3556E-03 -2.3556E-01
2 9.5400E-01 9.4932E-01 1.4125E-03 4.6778E-03 4.9033E-01
3 8.9500E-01 8.9629E-01 1.1655E-03 -1.2889E-03 -1.4401E-01
4 8.4300E-01 8.4326E-01 9.8812E-04 -2.5556E-04 -3.0315E-02
Napověda-F1 Řádek: 60 - 82 Celkem: 168 Délka: 8777

```



Úloha P638

na výstavbu lineárního regresního modelu regresním tripletem

Test normality

Jak vysvětlit výsledky úlohy a napsat 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 používejte na přenos výstupů a obrázků **PrintScreen** a upravte pak negativ na černobílý pozitiv ve MS Windows.

Normalita chyby

1. **Rankitový (Q-Q) graf:** $\hat{e}_{(i)}$ nebo \hat{e}_{R_i} na u_{P_i}
pro $P_i = i / (n + 1)$

2. **Test normality:** H_0 : normalita vs. H_A : nenormalita
Jarque-Berrova testační statistika

$$L(\hat{e}) = n \left[\frac{\hat{u}_3^2}{6 \hat{u}_2^3} + \frac{\frac{\hat{u}_4}{\hat{u}_2^2} - 3}{24} \right] + n \left[\frac{3 \hat{u}_1^2}{2 \hat{u}_2} - \frac{\hat{u}_3 \hat{u}_1}{\hat{u}_2^2} \right]$$

kde \hat{u}_j je j -tý výběrový moment reziduí

$$\hat{u}_j = \frac{\sum_{i=1}^n \hat{e}_i^j}{n}$$

Test: je-li $L(\hat{\epsilon}) > \chi_{1-\alpha}^2(2) = 5.99$, je H_0 (normalita)
zamítnuta.

Např. pro lineární modely s absolutním členem je $\hat{u}_1 = 0$ a $L(\hat{\epsilon})$ je

$$L(\hat{\epsilon}) = n \left(\frac{\hat{g}_1}{6} + \frac{(\hat{g}_2 - 3)^2}{24} \right)$$

kde

$$\hat{g}_1 = \frac{\hat{u}_3^2}{\hat{u}_2^3} \qquad \hat{g}_2 = \frac{\hat{u}_4}{\hat{u}_2^2}$$

(test není vhodný pro malé výběry)

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

F1=náponěda F2=uložení F3=čtení F4=numericcký mód F5=zoom ESC=konec

Řádek 1 Sloupec 5 Insert NUM P638

Textový editor

```
P638x P638y
0.00000 1.00000
10.00000
20.00000
30.00000
40.00000 0.79100
50.00000 0.73500
60.00000 0.68500
70.00000 0.62800
80.00000 0.58100
```

Čtení ze souboru: P638



Otevřít Data, pak stisk F3, pak smazat *demo.dat*, napsat *P638* a ukončit *Esc a A*.

Zvolte graf regresní křivky a Esc zpět.



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

U Ý S L E D K Y

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

Fisher-Snedocorův test významnosti regrese, F : 2.2084E+04
 Tabulkový kvantil, F(1-alpha,m-1,n-m) : 5.5914E+00
 Závěr: Navržený model je přijat jako významný.
 Spočtená hladina významnosti : 0.000

Scottovo kritérium multikolinearity, M : -7.5529E-14
 Závěr: Navržený model je korektní.

Cook-Weisbergův test heteroskedasticity, Sf : 1.0050E+01
 Tabulkový kvantil, Chi^2(1-alpha,1) : 3.8415E+00
 Závěr: Rezidua vykazují heteroskedasticitu.
 Spočtená hladina významnosti : 0.002

Jarque-Berraův test normality reziduí, L(e) : 6.8980E-01
 Tabulkový kvantil, Chi^2(1-alpha,2) : 5.9915E+00
 Závěr: Normalita je přijata.
 Spočtená hladina významnosti : 0.708

Waldův test autokorelace, Wa : 9.4237E+00
 Tabulkový kvantil, Chi^2(1-alpha,1) : 3.8415E+00
 Závěr: Rezidua jsou autokorelována.
 Spočtená hladina významnosti : 0.002

Náponěda-F1 Řádek: 98 - 120 Celkem: 168 Délka: 8777

Zvolte Rankitové grafy a pak Normalizovaná rezidua. Vložte síť mezníkem. Pak Esc zpět.

Úloha P638

na výstavbu lineárního regresního modelu regresním tripletem

Test vhodnosti modelu

Jak vysvětlit výsledky úlohy a napsat 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.

Znaménkový test vhodnosti modelu

nenáhodnost reziduí lze testovat znaménkovým testem.

Postup:

1. Určuje se počet sekvencí n_U , kde mají rezidua stejná znaménka.
(např. pro rezidua -1, -1, 1, -1, 1, 2, 1 je počet sekvencí roven $\hat{n}_U = 4$)
2. Stanoví se počet reziduí kladných (n_+) a záporných (n_-).
3. Teoretický počet sekvencí n_t a jeho rozptyl D_t je

$$n_t = 1 + \frac{2 n_+ n_-}{n_+ + n_-} \approx 1 + \frac{n}{2}$$

$$D_t = \frac{2 n_+ n_- (2 n_+ n_- - n_+ - n_-)}{(n_+ + n_-)^2 (n_+ + n_- - 1)} \approx \frac{n}{4}$$

Pro $n_+ > 10$ a $n_- > 10$ lze využít náhodné veličiny

$$U = \frac{n_u - n_t + C}{\sqrt{D_t}}$$

s normovaným normálním rozdělením $N(0, 1)$.

Testování: Je-li $n_U < (n_t - C 1.96 \sqrt{D_t})$, pak je v reziduích trend a model je nesprávně navržen. (Konstanta $C = 0.5$ je korekcí na spojitost).

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 1 Sloupec 5 Insert NUM P638

Textový editor

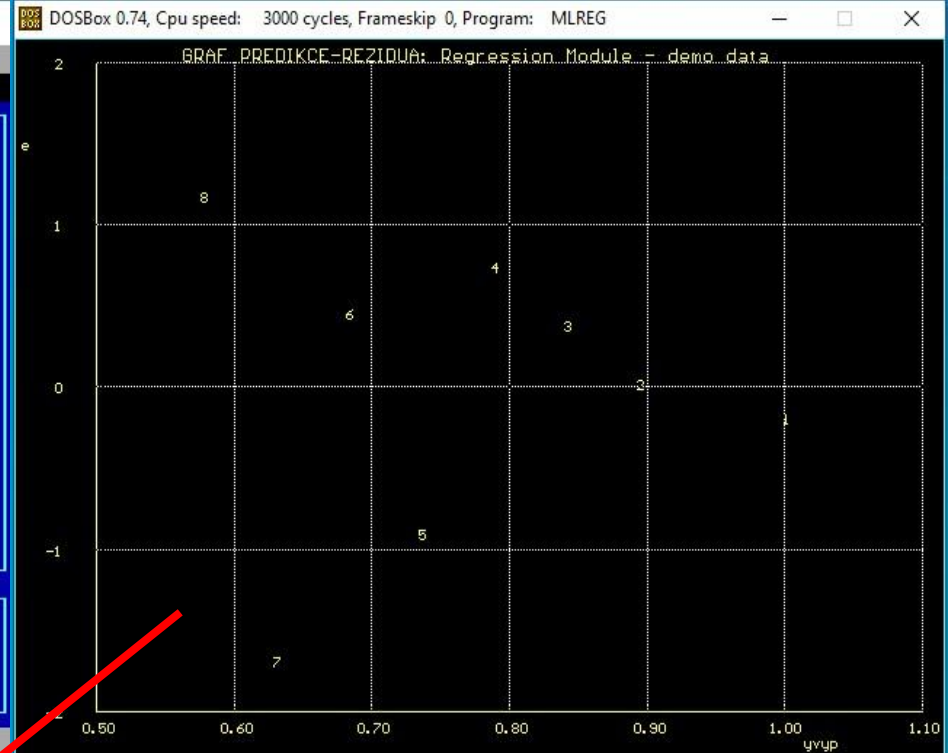
P638x P638y

0.00000	1.00000
10.00000	
20.00000	
30.00000	
40.00000	0.79100
50.00000	0.73500
60.00000	0.68500
70.00000	0.62800
80.00000	0.58100

Čtení ze souboru: P638

Hlášení

Uyhledávání chyb: F6=první F7=předchozí F8=následující F9=poslední



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

V Ý S L E D K Y

Scottovo kritérium multikolinearity, M : -1.5384E-13
Závěr: Navržený model je korektní.

Cook-Weisbergův test heteroskedasticity, Sf : 1.3713E+01
Tabulkový kvantil, Chi²(1-alpha,1) : 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu

Spočtená hladina významnosti : 0.000

Jarque-Berraův test normality reziduí, L(e) : 6.5511E-01
Tabulkový kvantil, Chi²(1-alpha,2) : 5.9915E+00
Závěr: Normalita je přijata.

Spočtená hladina významnosti : 0.721

Waldův test autokorelace, Wa : 1.1740E+01
Tabulkový kvantil, Chi²(1-alpha,1) : 3.8415E+00
Závěr: Rezidua jsou autokorelována.

Spočtená hladina významnosti : 0.001

Znamékový test, Dt : 6.1791E-01
Tabulkový kvantil, N(1-alpha/2) : 1.6449E+00
Závěr: Rezidua nevykazují trend.

Spočtená hladina významnosti : 0.268

Napověda-F1 Řádek: 103 - 125 Celkem: 164 Délka: 8452

Závěry a vysvětlení výstupu úlohy P638

Úkoly

Odpovědi, závěry a vysvětlení

(1) Test heteroskedasticity.

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

V Ý S L E D K Y

Scottovo kritérium multikolinearity, M : -1.5384E-13
Závěr: Navržený model je korektní.

(2) Test normality reziduí.

Cook-Weisbergův test heteroskedasticity, Sf : 1.3713E+01
Tabulkový kvantil, $\text{Chi}^2(1-\alpha,1)$: 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.
Spočtená hladina významnosti : 0.000

(3) Test autokorelace.

Jarque-Berraův test normality reziduí, L(e) : 6.5511E-01
Tabulkový kvantil, $\text{Chi}^2(1-\alpha,2)$: 5.9915E+00
Závěr: Normalita je přijata.
Spočtená hladina významnosti : 0.721

(4) Test vhodnosti modelu.

Waldův test autokorelace, Wa : 1.1740E+01
Tabulkový kvantil, $\text{Chi}^2(1-\alpha,1)$: 3.8415E+00
Závěr: Rezidua jsou autokorelována.
Spočtená hladina významnosti : 0.001

Znaménkový test, Dt : 6.1791E-01
Tabulkový kvantil, $N(1-\alpha/2)$: 1.6449E+00
Závěr: Rezidua nevykazují trend.
Spočtená hladina významnosti : 0.268

Úloha M619

na výstavbu vícerozměrného lineárního regresního modelu regresním tripletem

Kritika metody

Jak vysvětlit výsledky regresního modelu v úloze a napsat 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.

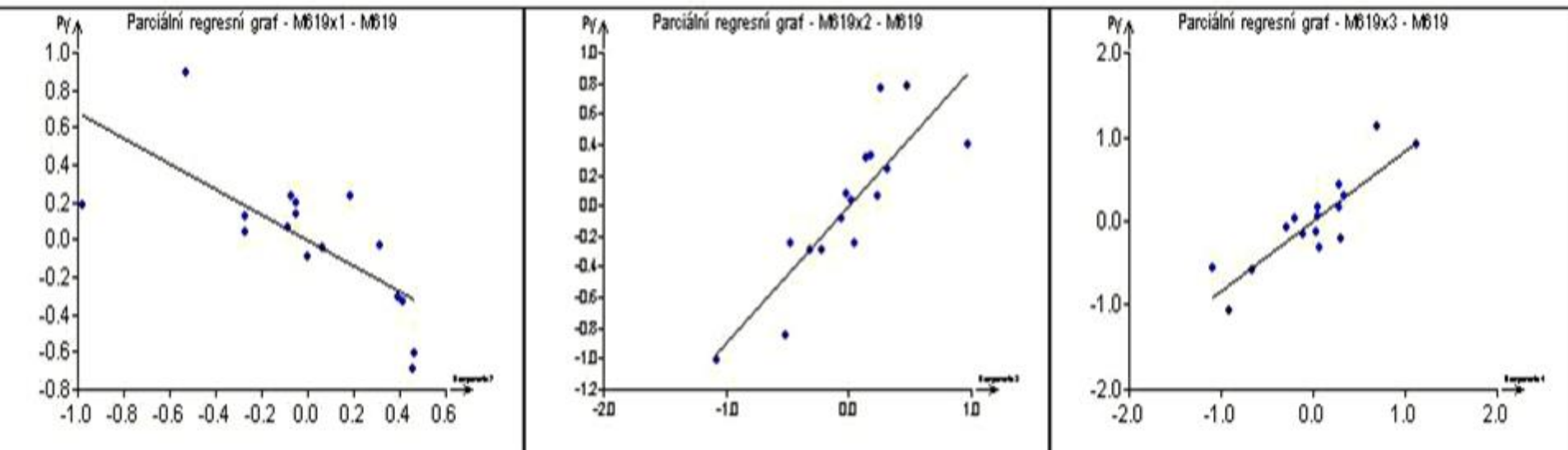
Úloha M619. Vliv tří parametrů na obsah kadmia v potravinářské pšenici

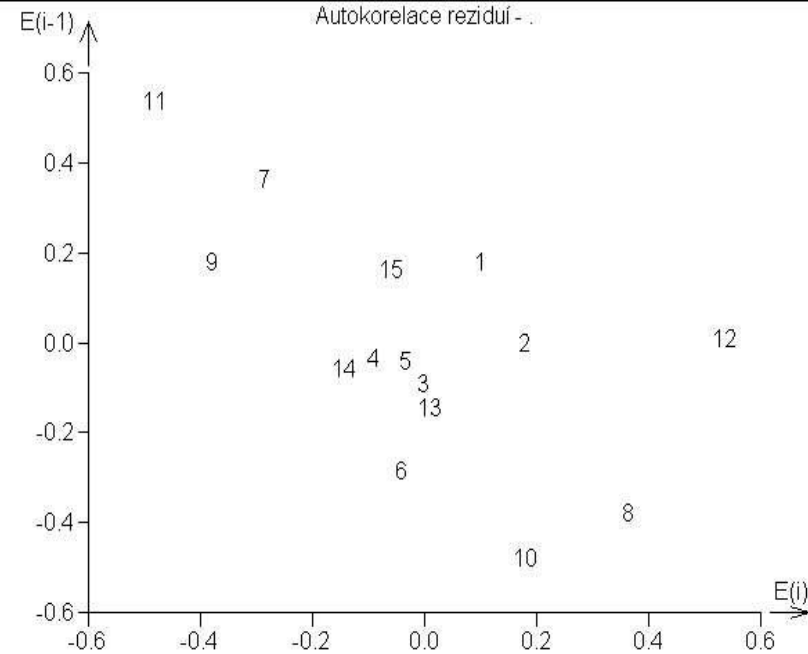
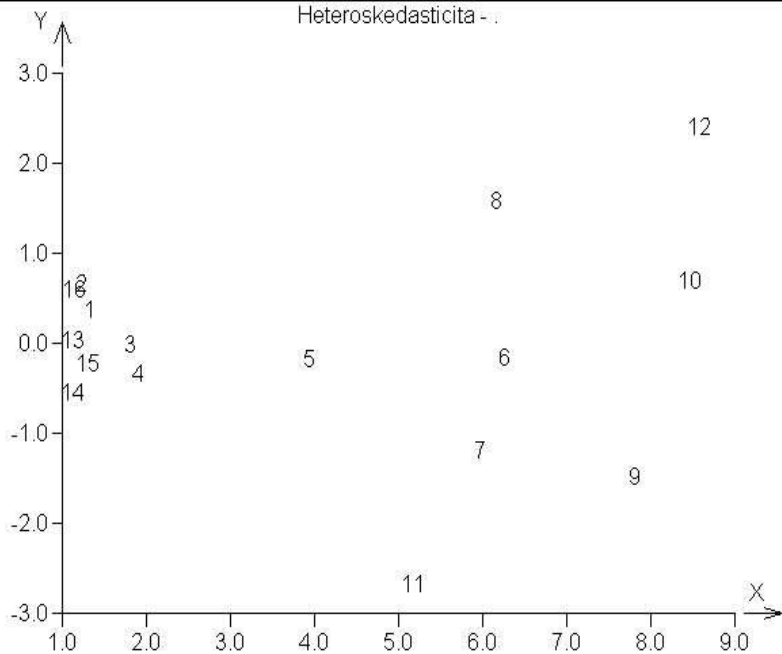
Obsah kadmia v zrně y [mg/l] v závislosti na obsahu kadmia v otrubách x_1 [mg/l], ve stonku s listy x_2 [mg/l] a v kořenovém systému x_3 [mg/l]. Vyšetřete regresní triplet (data, model, metoda) a nalezněte lineární regresní model.

Výstavba lineárního regresního modelu:

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

Parciální regresní grafy tří nezávisle proměnných





Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 5.373857
 Kvantil $\chi^2(1-\alpha,1)$: 3.841459
 Pravděpodobnost : 0.020441

Závěr: Rezidua vykazují heteroskedasticitu!

Waldův test autokorelace

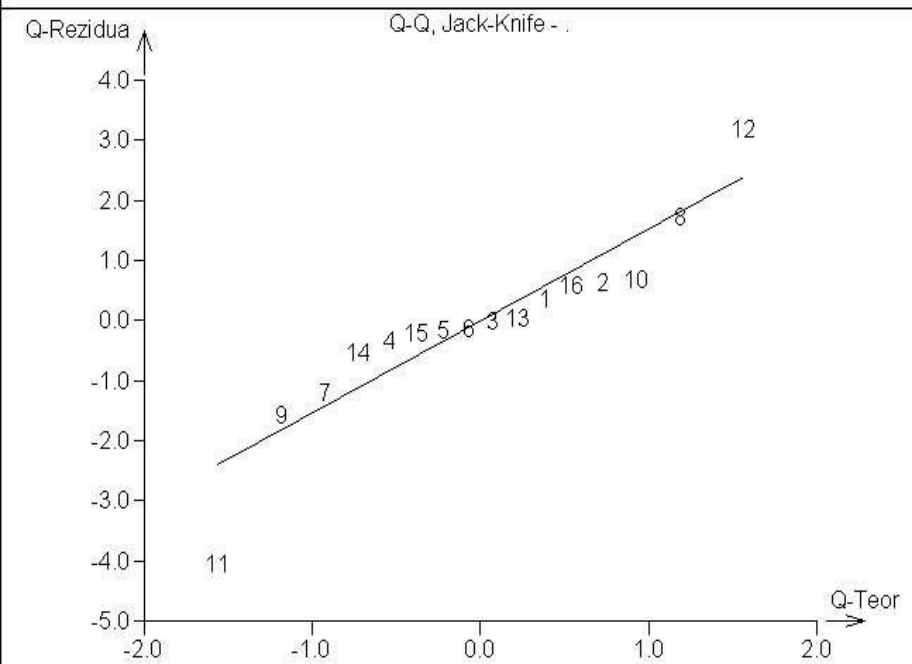
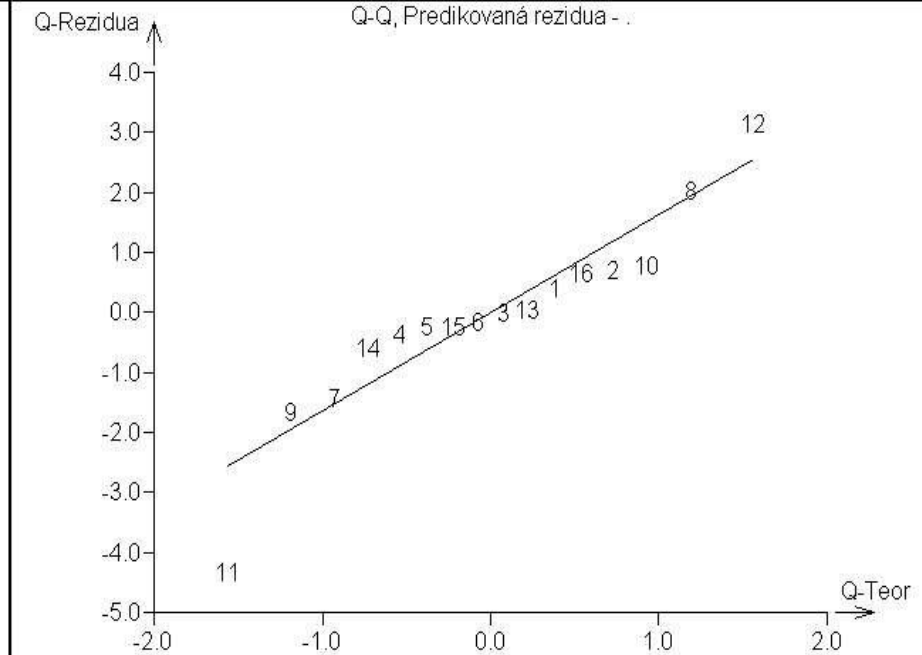
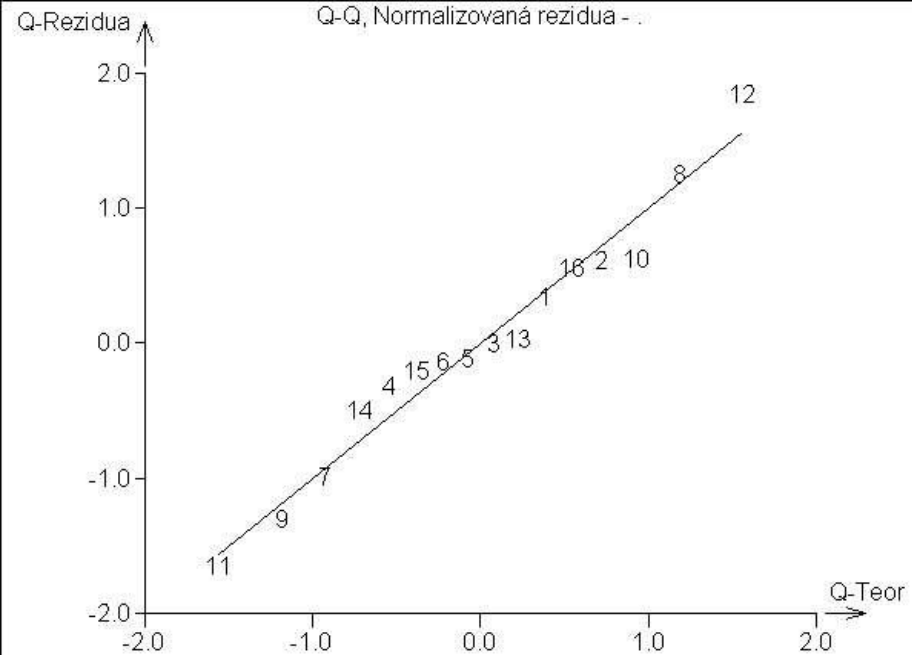
Hodnota kritéria WA : 5.003828
 Kvantil $\chi^2(1-\alpha,1)$: 3.841459
 Pravděpodobnost : 0.025291

Závěr: Autokorelace je významná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW : -1.0
 Kritické hodnoty DW 0.74 1.93

Závěr: Negativní autokorelace reziduí není prokázána.



Jarque-Berrův test normality

Hodnota kritéria JB : 0.033085

Kvantil $\chi^2(1-\alpha, 2)$: 5.991465

Pravděpodobnost : 0.983593

Závěr: Rezidua mají normální rozdělení.

Vyšetření předpokladů u MNČ při testování regresního tripletu

Fisher-Snedecorův test významnosti modelu, Hodnota kritéria F: 1404.982774

Kvantil F (1-alfa, m-1, n-m) : 3.490295, Pravděpodobnost: 0.0, **Závěr: Model je významný**

Scottovo kritérium multikolinearity, Hodnota kritéria SC: 0.970655, **Závěr: Model je nekorektní!**

Cook-Weisbergův test heteroskedasticity, Hodnota kritéria CW: 5.373857,

Kvantil $\chi^2(1-\alpha, 1)$: 3.841459, Pravděpodobnost: 0.020441,

Závěr: Rezidua vykazují heteroskedasticitu!

Jarque-Berrův test normality, Hodnota kritéria JB: 0.033085, Kvantil $\chi^2(1-\alpha, 2)$: 5.991465,

Pravděpodobnost: 0.983593, **Závěr: Rezidua mají normální rozdělení.**

Waldův test autokorelace, Hodnota kritéria WA: 5.003828,

Kvantil $\chi^2(1-\alpha, 1)$: 3.841459, Pravděpodobnost: 0.025291, **Závěr: Autokorelace je významná**

Durbin-Watsonův test autokorelace, Hodnota kritéria DW: -1.0, Kritické hodnoty DW 0.74, 1.93,

Závěr: Negativní autokorelace reziduí není prokázána.

Znaménkový test reziduí, Hodnota kritéria Sg: 0.328976, Kvantil $N(1-\alpha/2)$: 1.959964,

Pravděpodobnost: 0.742174, **Závěr: V reziduích není trend.**

Úlohy na výstavbu vícerozměrného lineárního regresního modelu regresním tripletem **Kritika metody**

Jak vysvětlit výsledky úlohy a napsat 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 používejte na přenos výstupů a obrázků **PrintScreen** a upravte pak negativ na černobílý pozitiv ve MS Windows.

Úloha M6.40 *Závislost koncentrace kreatininu na hmotnosti, věku a výšce mužů – předpoklady MNČ*

Zadání: Zjistěte, zda existuje závislost celkové koncentrace kreatininu y u zdravých mužů na jejich hmotnosti x_1 , věku x_2 a výšce x_3 . Výběr je tvořen muži, u kterých nebylo prokázáno renální onemocnění.

Úkoly:

- (1) Postavte lineární regresní model a vyšetřete regresní triplet.
- (2) Vyšetřete statistickou významnost jednotlivých regresních parametrů.
- (3) Užijte také parciální regresní grafy a komentujte fyzikální význam jednotlivých regresních parametrů.
- (4) Vyšetřete porušení předpokladů MNČ.

Data: Věk x_1 [roky], hmotnost x_2 [kg], výška x_3 [cm], koncentrace kreatininu y [$\mu\text{mol/l}$] :

x_1	x_2	x_3	y
20	70	180	66
...
75	84	179	119

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

F1=nápořevda F2=ulořeni F3=řtení F4=numerický mód F5=zoom ESC=konec

Řádek 1 Sloupec 5 Insert NUM M640.txt

Textový editor

M640x1 M640x2 M640x3 M640y

20.00000 70.00000 180.00000 66.00000

24.00000 Čtení ze souboru: M640.txt

26.00000

30.00000

33.00000 85.00000 190.00000 74.00000

35.00000 82.00000 172.00000 73.00000

38.00000 63.00000 168.00000 77.00000

40.00000 91.00000 180.00000 80.00000

42.00000 87.00000 178.00000 79.00000

45.00000 88.00000 183.00000 85.00000

47.00000 96.00000 188.00000 88.00000

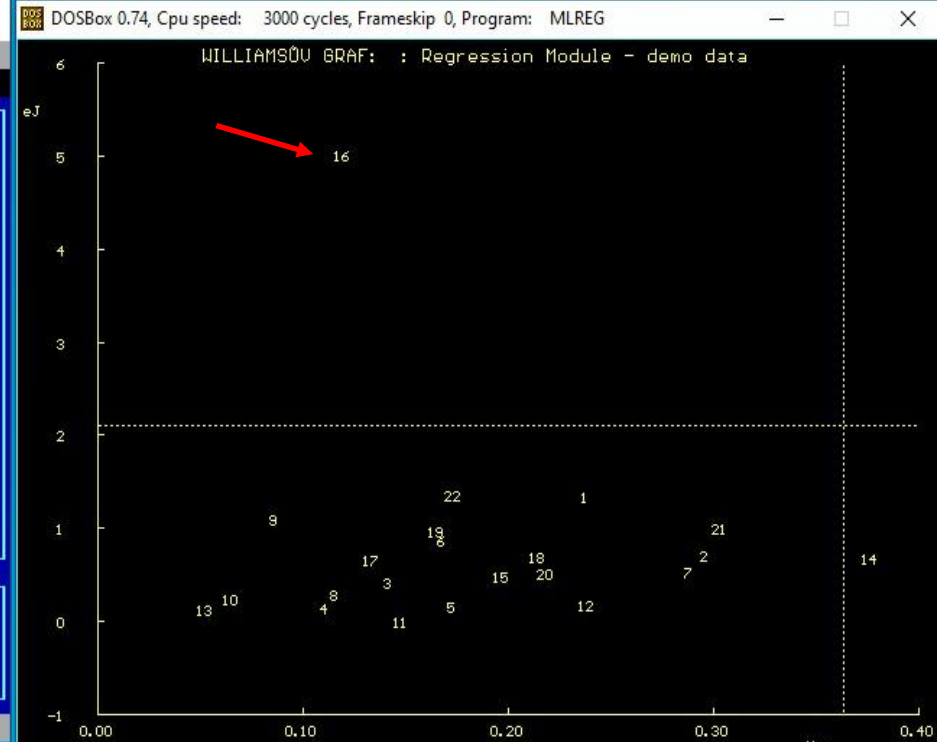
50.00000 92.00000 172.00000 91.00000

52.00000 81.00000 179.00000 93.00000

54.00000 102.00000 198.00000 97.00000

Hlášení

Vyhledávání chyb: F6=první F7=předchozí F8=následující F9=poslední



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

F1=nápořevda F2=ulořeni F3=řtení F4=numerický mód F5=zoom ESC=konec

Řádek 16 Sloupec 5 Insert NUM M640.txt

2.000000 E+01	7.000000 E+01	1.800000 E+02	6.600000 E+01
2.400000 E+01	6.800000 E+01	1.850000 E+02	6.800000 E+01
2.600000 E+01	8.000000 E+01	1.780000 E+02	6.900000 E+01
3.000000 E+01	7.900000 E+01	1.820000 E+02	7.200000 E+01
3.300000 E+01	8.500000 E+01	1.900000 E+02	7.400000 E+01
3.500000 E+01	8.200000 E+01	1.720000 E+02	7.300000 E+01
3.800000 E+01	6.300000 E+01	1.680000 E+02	7.700000 E+01
4.000000 E+01	9.100000 E+01	1.800000 E+02	8.000000 E+01
4.200000 E+01	8.700000 E+01	1.780000 E+02	7.900000 E+01
4.500000 E+01	8.800000 E+01	1.830000 E+02	8.500000 E+01
4.700000 E+01	9.600000 E+01	1.880000 E+02	8.800000 E+01
5.000000 E+01	9.200000 E+01	1.720000 E+02	9.100000 E+01
5.200000 E+01	8.100000 E+01	1.790000 E+02	9.300000 E+01
5.400000 E+01	1.020000 E+02	1.980000 E+02	9.700000 E+01
5.700000 E+01	9.900000 E+01	1.810000 E+02	9.900000 E+01
6.300000 E+01	8.900000 E+01	1.880000 E+02	1.010000 E+02
6.500000 E+01	8.300000 E+01	1.700000 E+02	1.070000 E+02
6.700000 E+01	7.600000 E+01	1.740000 E+02	1.100000 E+02
7.000000 E+01	7.200000 E+01	1.780000 E+02	1.080000 E+02
7.200000 E+01	7.500000 E+01	1.870000 E+02	1.150000 E+02
7.500000 E+01	8.400000 E+01	1.790000 E+02	1.190000 E+02

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

V Ý S L E D K Y

(VIP[j] > 10 indikuje silnou multikolinearitu)

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

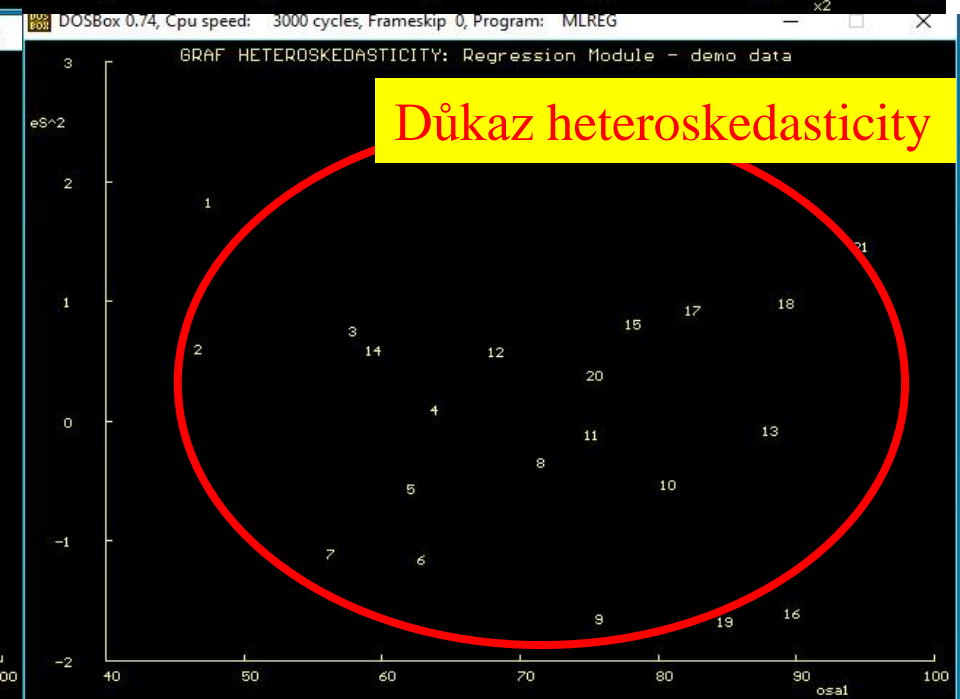
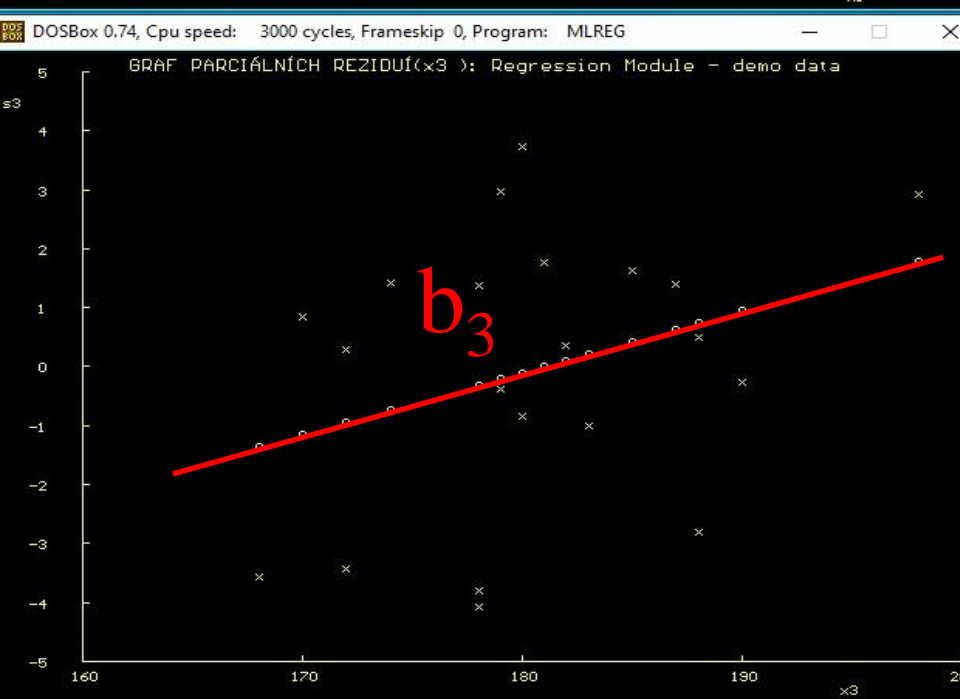
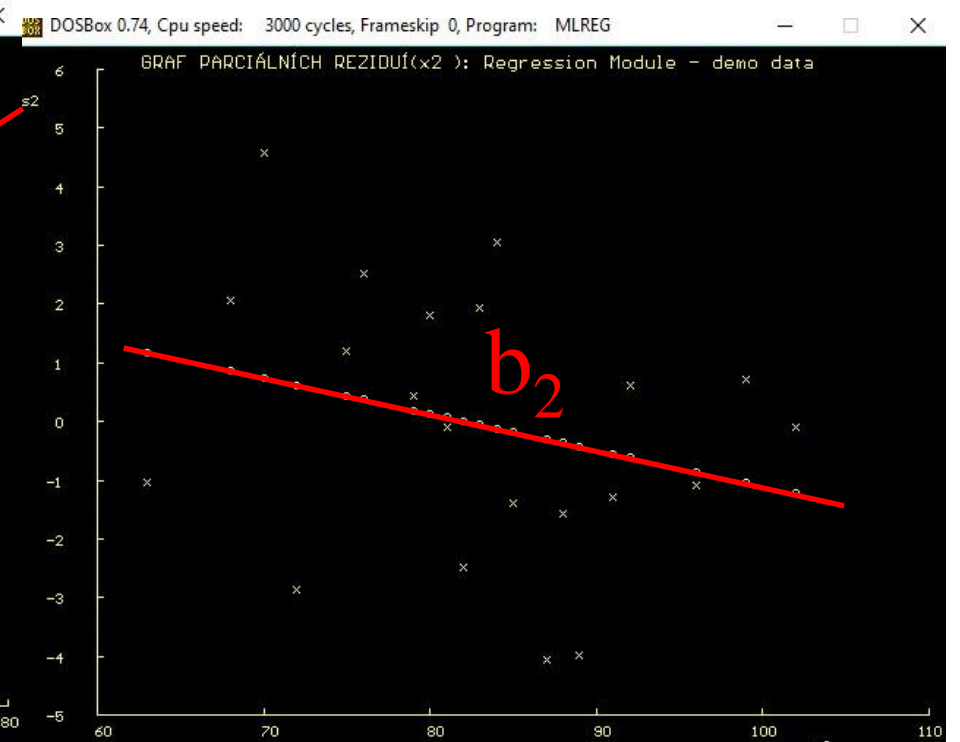
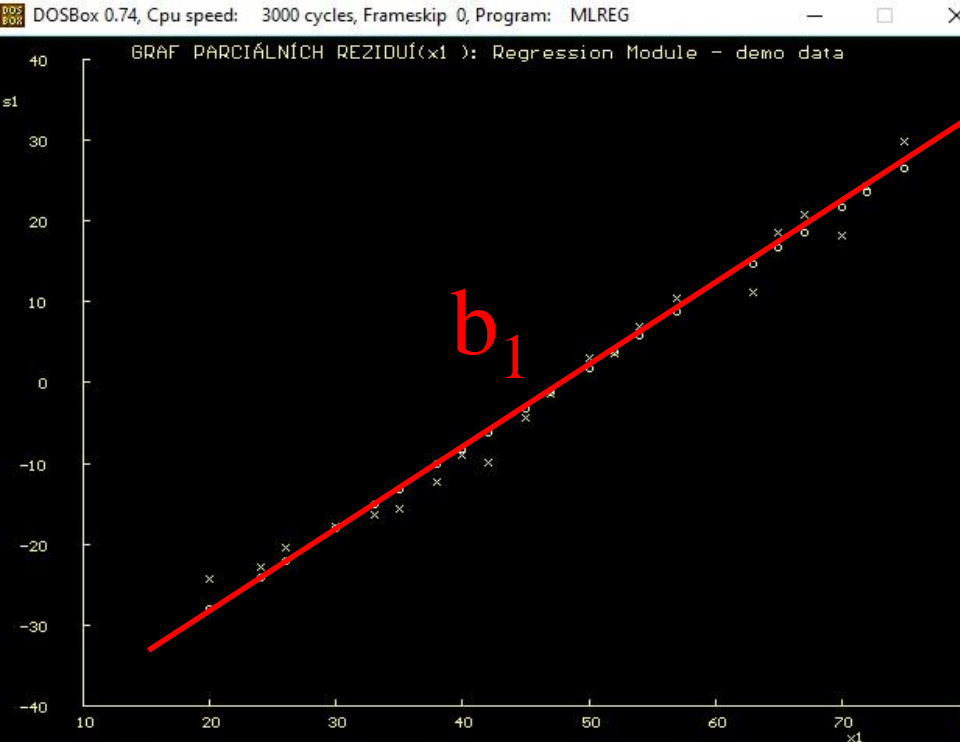
Parametr	Odhad	Směrodatná odchylka	Test H0: B[j] = 0 vs. HA: B[j] <> 0	t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	2.7696E+01	1.3577E+01	2.0399E+00	Akceptována		0.057
B[1]	9.9309E-01	3.3158E-02	2.9950E+01	Zamítnuta		0.000
B[2]	-6.0992E-02	6.0445E-02	-1.0090E+00	Akceptována		0.327
B[3]	1.0490E-01	8.2529E-02	1.2711E+00	Akceptována		0.221

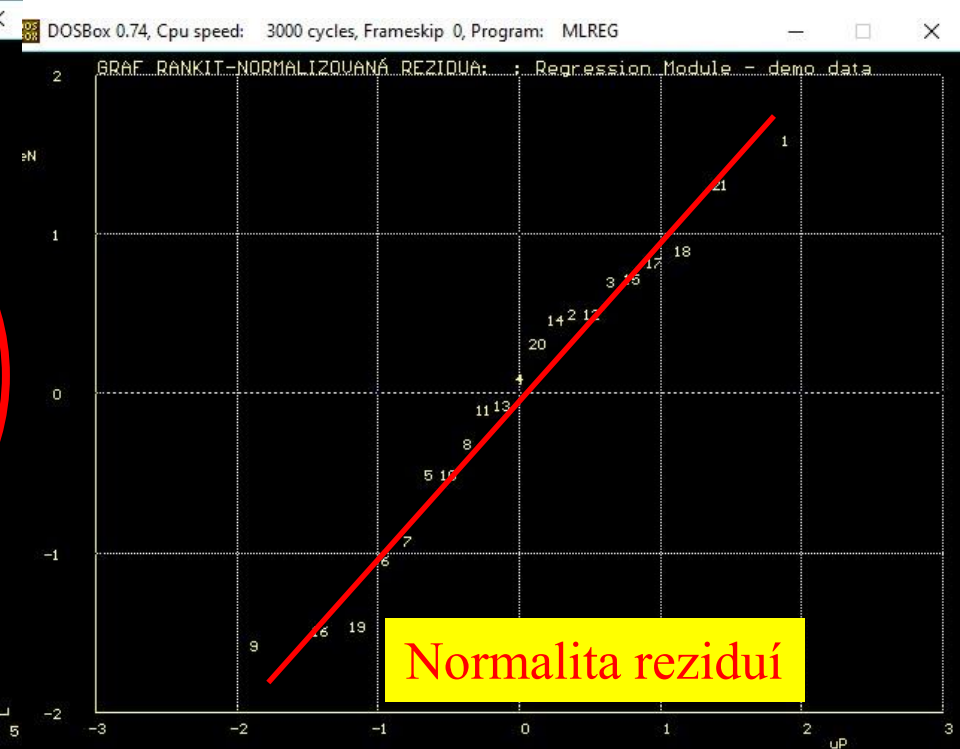
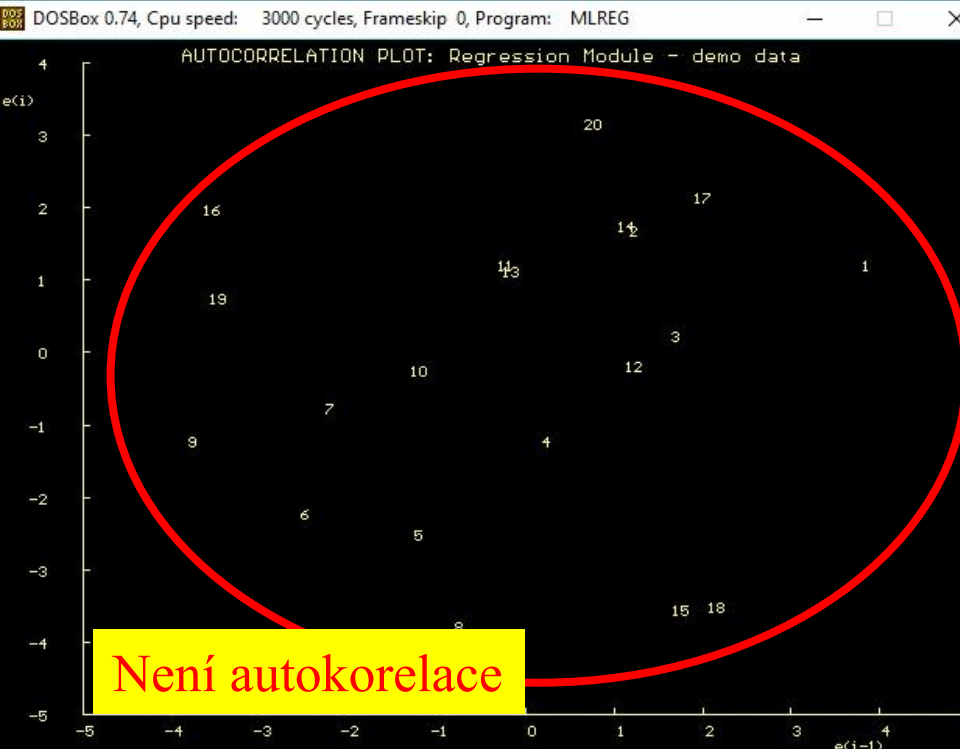
(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.9109E-01
Koeficient determinace, R ²	: 9.8225E-01
Predikovaný korelační koeficient, R _p ²	: 9.8621E-01
Střední kvadratická chyba predikce, MEP	: 7.1890E+00
Akaikeho informační kritérium, AIC	: 4.0310E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	y _{expi1}	y _{yppi1}	s(y _{yppi1})	e _{il}	er _{il}
Napověda-F1 Řádek: 87 - 109					
				Celkem: 247	Délka: 14118





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

U Ý S L E D K Y

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

Fisher-Snedcorův test významnosti regrese, F	: 3.1364E+02
Tabulkový kvantil, F(1-alpha, m-1, n-m)	: 3.1968E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.000
Scottovo kritérium multikolinearity, M	: 2.2422E-02
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, Sf	: 4.8039E+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)	: 9.1451E-01
Tabulkový kvantil, Chi^2(1-alpha, 2)	: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.633
Waldův test autokorelace, Wa	: 2.3552E-01
Tabulkový kvantil, Chi^2(1-alpha, 1)	: 3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.627

Napověda-F1 Řádek: 141 - 163 Celkem: 247 Délka: 14118

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

F1=nápořveřa F2=ulořženř F3=řtenř F4=numerickř mřd F5=zoom ESC=konec

Přet řádkř: 21 Přet sloupcř: 4 Uodorovně M640.txt

Numerickř editor

	1	2	3	4
1	2.000000 E+01	7.000000 E+01	1.800000 E+02	6.600000 E+01
Transformace hodnot v bloku (zadej funkci s parametrem x)				
x:=0				
5	3.300000 E+01	8.500000 E+01	1.900000 E+02	7.400000 E+01
6	3.500000 E+01	8.200000 E+01	1.720000 E+02	7.300000 E+01
7	3.800000 E+01	6.300000 E+01	1.680000 E+02	7.700000 E+01
8	4.000000 E+01	9.100000 E+01	1.800000 E+02	8.000000 E+01
9	4.200000 E+01	8.700000 E+01	1.780000 E+02	7.900000 E+01
10	4.500000 E+01	8.800000 E+01	1.830000 E+02	8.500000 E+01
11	4.700000 E+01	9.600000 E+01	1.880000 E+02	8.800000 E+01
12	5.000000 E+01	9.200000 E+01	1.720000 E+02	9.100000 E+01
13	5.200000 E+01	8.100000 E+01	1.790000 E+02	9.300000 E+01
14	5.400000 E+01	1.020000 E+02	1.980000 E+02	9.700000 E+01
15	5.700000 E+01	9.900000 E+01	1.810000 E+02	9.900000 E+01
16	6.300000 E+01	8.900000 E+01	1.880000 E+02	1.010000 E+02

značení bloku: Z=začátek K=konec B=buňka X=zrušení/obnovenř
 S=sloupec R=řádek N=nová poloha MEZERA=změna směru
 blokové operace: F6=kopřování F7=zaplňení F8=transformace
 ostatní operace: F9=změna velikosti matice F10=transpozice

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

F1=nápořveřa F2=ulořženř F3=řtenř F4=numerickř mřd F5=zoom ESC=konec

Řádek 1 Sloupec 43 Insert NUM M640.txt

2.000000 E+01	0	0	6.600000 E+01
2.400000 E+01	0	0	6.800000 E+01
2.600000 E+01	0	0	6.900000 E+01
3.000000 E+01	0	0	7.200000 E+01
3.300000 E+01	0	0	7.400000 E+01
3.500000 E+01	0	0	7.300000 E+01
3.800000 E+01	0	0	7.700000 E+01
4.000000 E+01	0	0	8.000000 E+01
4.200000 E+01	0	0	7.900000 E+01
4.500000 E+01	0	0	8.500000 E+01
4.700000 E+01	0	0	8.800000 E+01
5.000000 E+01	0	0	9.100000 E+01
5.200000 E+01	0	0	9.300000 E+01
5.400000 E+01	0	0	9.700000 E+01
5.700000 E+01	0	0	1.010000 E+02
6.300000 E+01	0	0	1.070000 E+02
6.500000 E+01	0	0	1.100000 E+02
6.700000 E+01	0	0	1.180000 E+02
7.000000 E+01	0	0	1.080000 E+02
7.200000 E+01	0	0	1.150000 E+02
7.500000 E+01	0	0	1.190000 E+02

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

F1=nápořveřa F2=ulořženř F3=řtenř F4=numerickř mřd F5=zoom ESC=konec

Řádek 21 Sloupec 24 Insert NUM M640.txt

2.000000 E+01	6.600000 E+01
2.400000 E+01	6.800000 E+01
2.600000 E+01	6.900000 E+01
3.000000 E+01	7.200000 E+01
3.300000 E+01	7.400000 E+01
3.500000 E+01	7.300000 E+01
3.800000 E+01	7.700000 E+01
4.000000 E+01	8.000000 E+01
4.200000 E+01	7.900000 E+01
4.500000 E+01	8.500000 E+01
4.700000 E+01	8.800000 E+01
5.000000 E+01	9.100000 E+01
5.200000 E+01	9.300000 E+01
5.400000 E+01	9.700000 E+01
5.700000 E+01	9.900000 E+01
6.300000 E+01	1.010000 E+02
6.500000 E+01	1.070000 E+02
6.700000 E+01	1.100000 E+02
7.000000 E+01	1.080000 E+02
7.200000 E+01	1.150000 E+02
7.500000 E+01	1.190000 E+02

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

U Ÿ S L E D K Y

(3) ODHADY PARAMETRŮ A TESTY ŸZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B[j] = 0 vs. H _A : B[j] <> 0	t-kriterium	hypoteza H0 je	Hlad. Ÿz.
B[0]	0.0000E+00					
B[1]	1.7706E+00	6.0659E-02		2.9190E+01	Zamřtnuta	0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

- Vřcenásobnř korelační koeficient, R : 5.3240E-01
- Koeficient determinace, R² : 2.8345E-01
- Predikovanř korelační koeficient, R_p² : 4.8459E-01
- Střednř kvadratická chyba predikce, MEP : 2.0083E+02
- Akaikeho informační kritérium, AIC : 1.0974E+02

(5) ANALÝZA KLASICKÝCH REZIDUŮ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativnř reziduum
i	y _{expli}	y _{yyp} [i]	s(y _{yyp} [i])	e[i]	er[i]
1	6.6000E+01	3.5413E+01	1.2132E+00	3.0587E+01	4.6345E+01
2	6.8000E+01	4.2495E+01	1.4558E+00	2.5505E+01	3.7507E+01
3	6.9000E+01	4.6036E+01	1.5771E+00	2.2964E+01	3.3281E+01
4	7.2000E+01	5.3119E+01	1.8198E+00	1.8881E+01	2.6224E+01

Napověřa-F1 Řádek: 72 - 94 Celkem: 228 Dělka: 12790

V Ý S L E D K Y

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylna	Test H0: B[j] = 0 vs. HA: B[j] <> 0	t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	4.1971E+01	1.6183E+00	2.5934E+01	Zamítnuta		0.000
B[1]	9.8470E-01	3.2012E-02	3.0760E+01	Zamítnuta		0.000

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.9011E-01
Koeficient determinace, R ²	: 9.8031E-01
Predikovaný korelační koeficient, R _p ²	: 9.8723E-01
Střední kvadratická chyba predikce, MEP	: 6.6624E+00
Ákaikeho informační kritérium, AIC	: 3.8488E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylna	Klasické reziduum	Relativní reziduum
i	yexp[i]	yvyp[i]	s(yvyp[i])	e[i]	er[i]
1	6.6000E+01	6.1664E+01	1.0331E+00	4.3356E+00	6.5690E+00
2	6.8000E+01	6.5603E+01	9.2478E-01	2.3968E+00	3.5247E+00
3	6.9000E+01	6.7573E+01	8.7265E-01	1.4274E+00	2.0687E+00
4	7.2000E+01	7.1511E+01	7.7377E-01	4.8861E-01	6.7862E-01

Napověda-F1 Řádek: 72 - 94 Celkem: 228 Délka: 12792

V Ý S L E D K Y

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

Fisher-Snedcorův test významnosti regrese, F	: 9.4616E+02
Tabulkový kvantil, F(1-alpha, m-1, n-m)	: 4.3807E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.000

Scottovo kritérium multikolinearity, M	: -9.0116E-16
Závěr: Navržený model je korektní.	

Cook-Weisbergův test heteroskedasticity, Sf	: 9.7246E+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)	: 6.4837E-01
Tabulkový kvantil, Chi ² (1-alpha, 2)	: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.723

Waldův test autokorelace, Wa	: 2.3916E+00
Tabulkový kvantil, Chi ² (1-alpha, 1)	: 3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.122

Napověda-F1 Řádek: 122 - 144 Celkem: 228 Délka: 12792

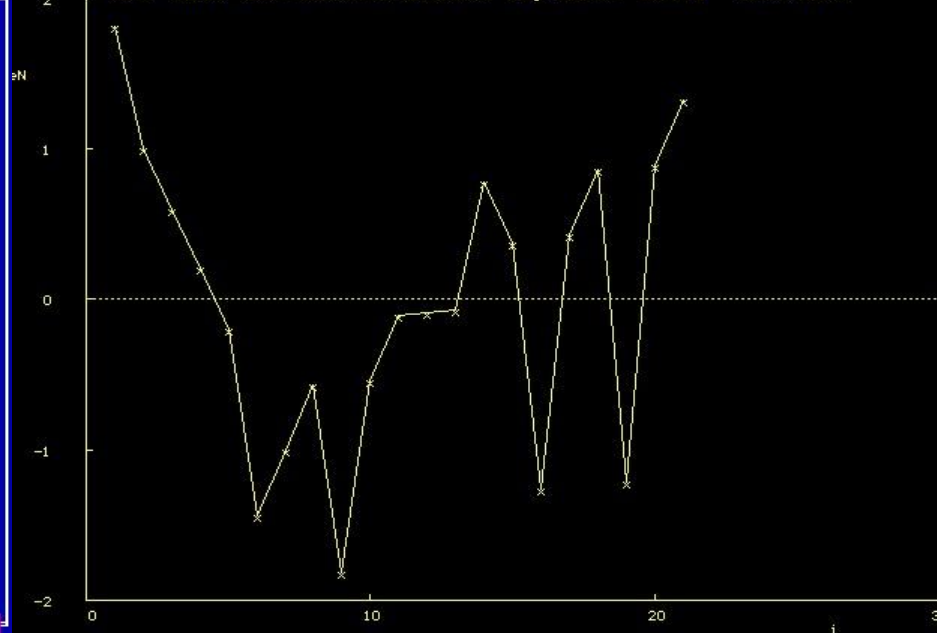
V Ý S L E D K Y

7	7.7000E+01	7.9389E+01	6.0952E-01	-2.3890E+00	-3.1025E+00
8	8.0000E+01	8.1358E+01	5.7897E-01	-1.3583E+00	-1.6979E+00
9	7.9000E+01	8.3328E+01	5.5416E-01	-4.3277E+00	-5.4781E+00
10	8.5000E+01	8.6282E+01	5.2944E-01	-1.2818E+00	-1.5080E+00
11	8.8000E+01	8.8251E+01	5.2220E-01	-2.5121E-01	-2.8547E-01
12	9.1000E+01	9.1205E+01	5.2597E-01	-2.0530E-01	-2.2560E-01
13	9.3000E+01	9.3175E+01	5.3808E-01	-1.7469E-01	-1.8784E-01
14	9.7000E+01	9.5144E+01	5.5732E-01	1.8559E+00	1.9133E+00
15	9.9000E+01	9.8098E+01	5.9800E-01	9.0183E-01	9.1094E-01
16	1.0100E+02	1.0401E+02	7.1199E-01	-3.0063E+00	-2.9766E+00
17	1.0700E+02	1.0598E+02	7.5704E-01	1.0243E+00	9.5726E-01
18	1.1000E+02	1.0795E+02	8.0466E-01	2.0549E+00	1.8681E+00
19	1.0800E+02	1.1090E+02	8.8000E-01	-2.8992E+00	-2.6845E+00
20	1.1500E+02	1.1287E+02	9.3235E-01	2.1314E+00	1.8534E+00
21	1.1900E+02	1.1582E+02	1.0134E+00	3.1773E+00	2.6700E+00

Reziduální součet čtverců, RSC	: 1.0850E+02
Průměr absolutních hodnot reziduí, Me	: 1.8851E+00
Průměr relativních reziduí, Mer	: 2.2140E+00
Odhad reziduálního rozptylu, s ² (e)	: 5.7108E+00
Odhad směrodatné odchylny reziduí, s(e)	: 2.3897E+00
Odhad šikmosti reziduí, g1(e)	: -1.4018E-01
Odhad špičatosti reziduí, g2(e)	: 2.1861E+00

Napověda-F1 Řádek: 97 - 119 Celkem: 228 Délka: 12792

GRAF INDEX-NORMALIZOVANÁ REZIDUA: Regression Module - demo data



Závěry a vysvětlení výstupu úlohy M640

Úkoly

Odpovědi, závěry a vysvětlení

(1) Postavte lineární regresní model a vyšetřete regresní triplet. Vyšetřete statistickou významnost všech regresních parametrů.

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] <> 0$	t-kritérium	hypoteza H_0 je	Hlad. výz.
BI 01	2.7696E+01	1.3577E+01		2.0399E+00	Akceptována	0.057
BI 11	9.9309E-01	3.3158E-02		2.9950E+01	Zamítnuta	0.000
BI 21	-6.0992E-02	6.0445E-02		-1.0090E+00	Akceptována	0.327
BI 31	1.0490E-01	8.2529E-02		1.2711E+00	Akceptována	0.221

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Učicenasobný korelační koeficient, R	: 9.9109E-01
Koeficient determinace, R^2	: 9.8225E-01
Predikovaný korelační koeficient, R_p^2	: 9.8621E-01
Střední kvadratická chyba predikce, MEP	: 7.1890E+00
Akaikeho informační kritérium, AIC	: 4.0310E+01

(2) Užijte také parciální regresní grafy a komentujte fyzikální význam jednotlivých regresních parametrů.

Výše uvedené parciální reziduální grafy b_1 , b_2 , b_3 ukazují sklon směrnice v souladu se statistickou významností každého parametru b .

(3) Vyšetřete porušení předpokladů MNČ.

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

Fisher-Snedocorův test významnosti regrese, F	: 3.1364E+02
Tabulkový kvantil, $F(1-\alpha, m-1, n-m)$: 3.1968E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.000
Scottovo kritérium multikolinearity, M	: 2.2422E-02
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, Sf	: 4.8039E+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)	: 9.1451E-01
Tabulkový kvantil, $\chi^2(1-\alpha, 2)$: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.633
Waldův test autokorelace, Wa	: 2.3552E-01
Tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.627

Úloha M6.54 *Vliv glukosy, proteinů, albuminů a globulinů na přírůstek hmotnosti kuřat – předpoklady MNČ*

Zadání: Před začátkem studie byl stanoven přírůstek váhy a biochemické parametry testovaných kuřat. Zjistěte jestli glukosa, celkové množství proteinů, obsah albuminů a globulinů má vliv na přírůstek hmotnosti kuřat.

Úkoly:

- (1) Vyšetřením regresního tripletu nalezněte nejlepší model.
- (2) Využijte regresní diagnostiku a parciální reziduální grafy ke statistické významnosti jednotlivých parametrů.

Data: Obsah glukosy x_1 [mmol.L⁻¹], celkový obsah proteinů x_2 [g.L⁻¹], obsah albuminů x_3 [g.L⁻¹], obsah globulinů x_4 [g.L⁻¹], přírůstek váhy v kg za den y [kg]:

x_1	x_2	x_3	x_4	y
13.1	33	28	5	0.016
...
13.6	29	25	4	0.079

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 1 Sloupec 5 Insert NUM M654.txt

Textový editor

M654x1	M654x2	M654x3	M654x4	M654y
13.1	33	28	5	0.016
11.9				
12.6				
11.3				
10.4	48	21	27	0.028
11.9	33	26	7	0.036
12.3	29	24	5	0.037
12.5	30	26	4	0.038
13.3	28	26	3	0.038
12.5	29	25	4	0.04
11.7	29	24	5	0.042
11.5	33	27	6	0.043
11.3	30	22	8	0.043
12.3	33	25	8	0.044

Čtení ze souboru: M654.txt

Hlášení

Vyhledávání chyb: F6=první F7=předchozí F8=následující F9=poslední

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

U Ý S L Ě D K Y

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylna	Test H0: $B[j] = 0$ vs. $H_A: B[j] <> 0$	t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	2.1120E-02	4.4206E-02	4.7776E-01	Akceptována	0.635	
B[1]	4.1002E-03	2.2602E-03	1.8141E+00	Akceptována	0.077	
B[2]	8.4721E-03	4.5779E-03	1.8507E+00	Akceptována	0.071	
B[3]	-9.3009E-03	4.7140E-03	-1.9730E+00	Akceptována	0.055	
B[4]	-9.2695E-03	4.5674E-03	-2.0295E+00	Zamítnuta	0.049	

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 4.9680E-01
 Koeficient determinace, R² : 2.4681E-01
 Predikovaný korelační koeficient, Rp² : 1.4720E-01
 Střední kvadratická chyba predikce, MEP : 1.6386E-04
 Akaikeho informační kritérium, AIC : -4.2094E+02

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylna	Klasické reziduum	Relativní reziduum
i	yexp[i]	yvyp[i]	s(yvyp[i])	er[i]	er[i]
1	1.6000E-02	4.7638E-02	4.4401E-03	-3.1638E-02	-1.9774E+02

Napověda-F1 Řádek: 134 - 156 Celkem: 401 Délka: 25091

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

U Ý S L Ě D K Y

34	5.5000E-02	5.1742E-02	2.2876E-03	3.2579E-03	5.9235E+00
35	5.7000E-02	5.3765E-02	2.9754E-03	3.2354E-03	5.6761E+00
36	5.9000E-02	4.7223E-02	2.3230E-03	1.1777E-02	1.9961E+01
37	5.9000E-02	6.1074E-02	5.2159E-03	-2.0743E-03	-3.5157E+00
38	5.9000E-02	4.7619E-02	2.5673E-03	1.1381E-02	1.9289E+01
39	6.0000E-02	5.4479E-02	5.2552E-03	5.5211E-03	9.2019E+00
40	6.0000E-02	4.6844E-02	1.8941E-03	1.3156E-02	2.1926E+01
41	6.1000E-02	5.0931E-02	1.9287E-03	1.0069E-02	1.6507E+01
42	6.2000E-02	4.7237E-02	3.2605E-03	1.4763E-02	2.3811E+01
43	6.3000E-02	5.0079E-02	1.9975E-03	1.2921E-02	2.0509E+01
44	6.4000E-02	4.9683E-02	2.4610E-03	1.4317E-02	2.2370E+01
45	6.5000E-02	5.2958E-02	2.2037E-03	1.2042E-02	1.8526E+01
46	6.9000E-02	5.7932E-02	4.7967E-03	1.1068E-02	1.6040E+01
47	7.1000E-02	5.6470E-02	4.9655E-03	1.4530E-02	2.0465E+01
48	7.9000E-02	5.2972E-02	2.7495E-03	2.6028E-02	3.2947E+01

Rezidualní součet čtverců, RSC : 6.0554E-03
 Průměr absolutních hodnot reziduí, Me : 8.3454E-03
 Průměr relativních reziduí, Mer : 2.2204E+01
 Odhad reziduálního rozptylu, s²(e) : 1.4082E-04
 Odhad směrodatné odchylny reziduí, s(e) : 1.1867E-02
 Odhad šikmosti reziduí, g1(e) : -5.4361E-01
 Odhad špičatosti reziduí, g2(e) : 3.6544E+00

Napověda-F1 Řádek: 189 - 211 Celkem: 401 Délka: 25091

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

U Ý S L Ě D K Y

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

Fisher-Snedecorův test významnosti regrese, F : 3.5226E+00
 Tabulkový kvantil, F(1-alpha,m-1,n-m) : 2.5888E+00
 Závěr: Navržený model je přijat jako významný.
 Spočtená hladina významnosti : 0.014

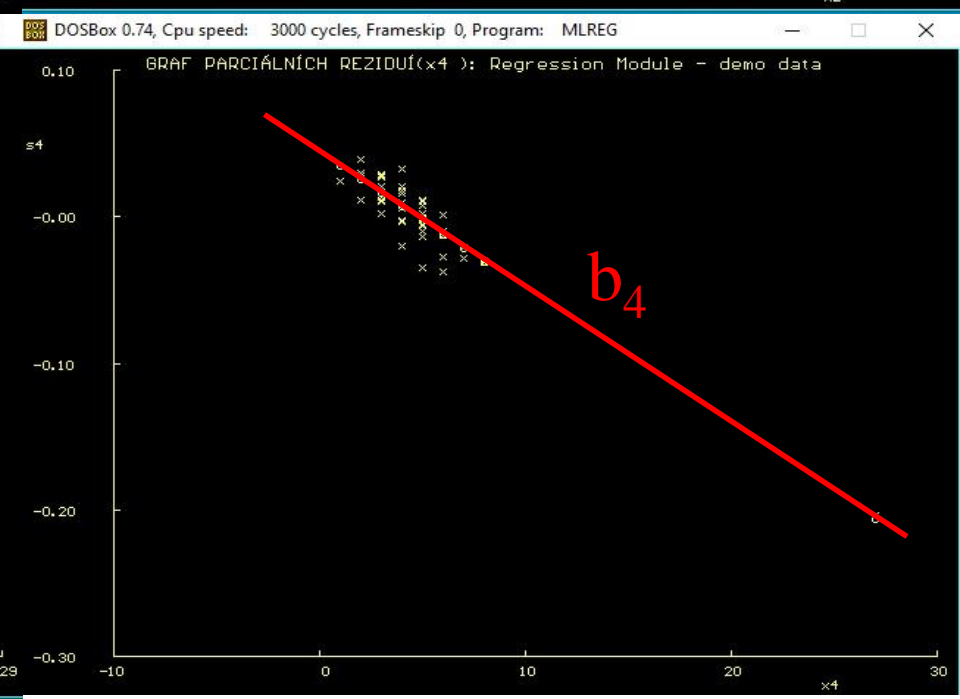
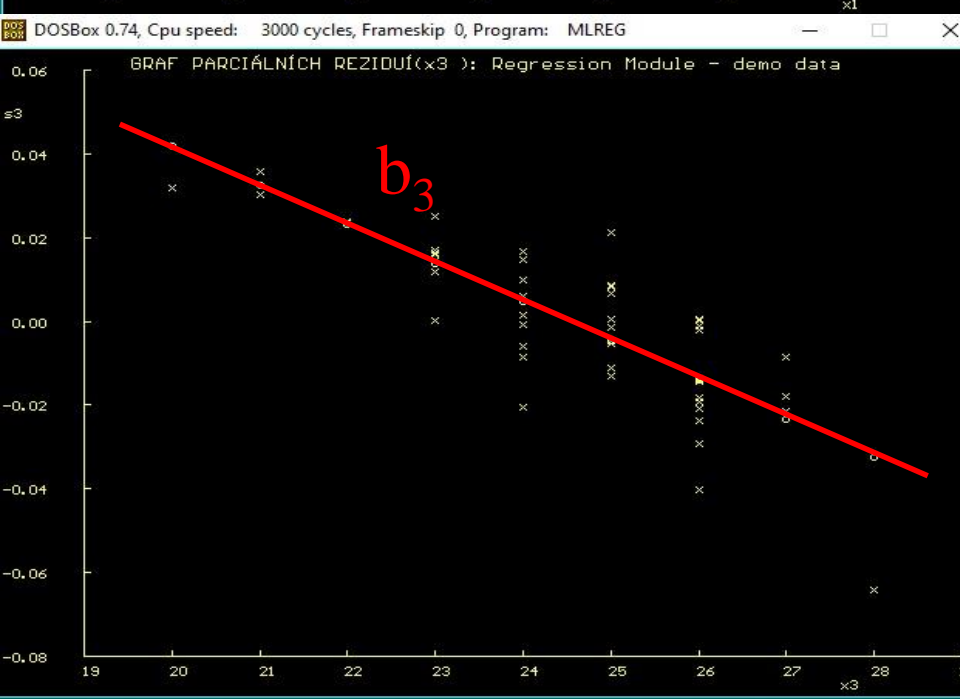
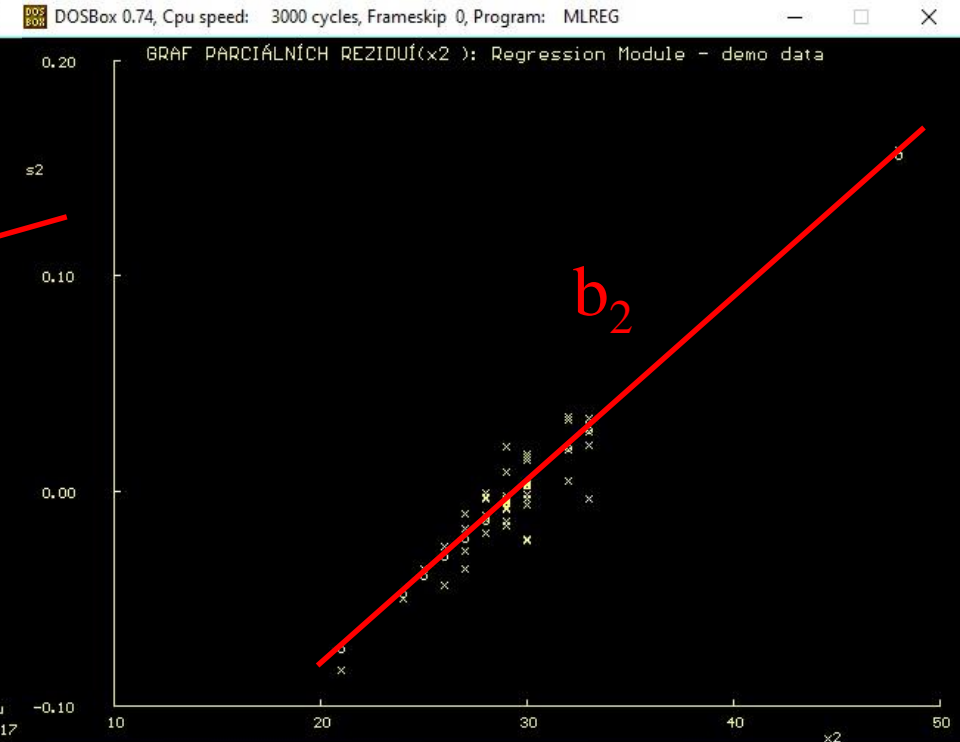
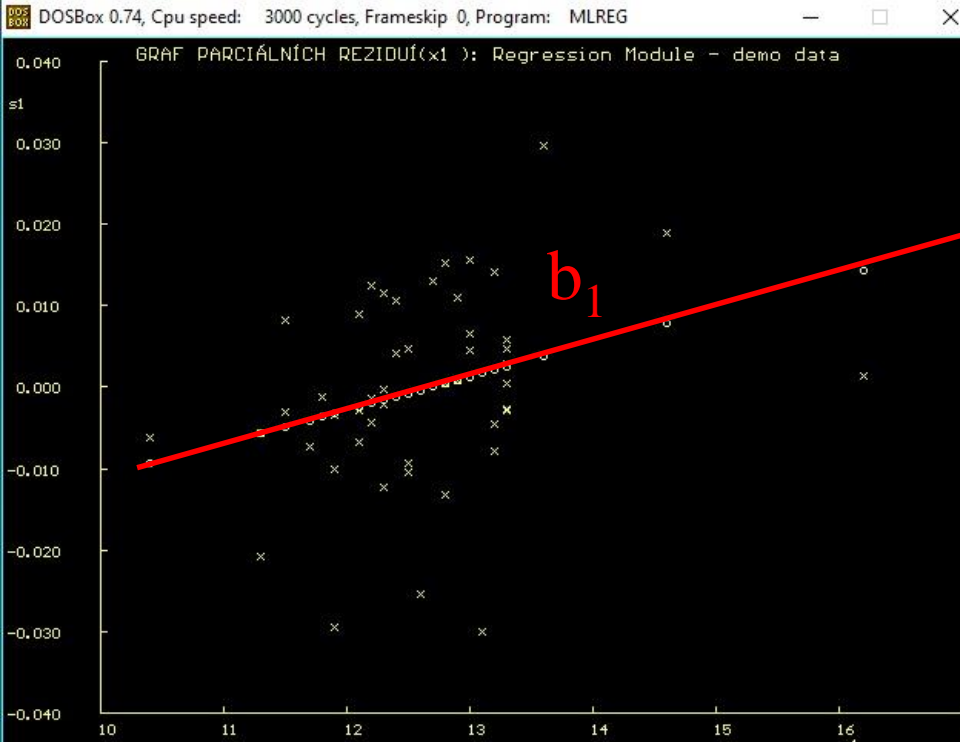
Scottovo kritérium multikolinearity, M : -2.2108E-02
 Závěr: Navržený model je korektní.

Cook-Weisbergův test heteroskedasticity, Sf : 9.7893E+00
 Tabulkový kvantil, Chi²(1-alpha,1) : 3.8415E+00
 Závěr: Rezidua vykazují heteroskedasticitu.
 Spočtená hladina významnosti : 0.002

Jarque-Berraův test normality reziduí, L(e) : 3.2205E+00
 Tabulkový kvantil, Chi²(1-alpha,2) : 5.9915E+00
 Závěr: Normalita je přijata.
 Spočtená hladina významnosti : 0.200

Waldův test autokorelace, Wa : 5.7514E+01
 Tabulkový kvantil, Chi²(1-alpha,1) : 3.8415E+00
 Závěr: Rezidua jsou autokorelována.
 Spočtená hladina významnosti : 0.000

Napověda-F1 Řádek: 214 - 236 Celkem: 401 Délka: 25091



ČEKÁM ...

REGRESNÍ MODUL : Regresní diagnostika

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

Váhy : Ne

Hladina význam.: 0.050

Zadání podmínek pro výpočet

VÝSLEDKY

25	5.5000E-02	4.3537E-02	2.5079E-03	1.1463E-02	2.0842E+01
26	5.5000E-02	5.5604E-02	1.4553E-03	-6.0442E-04	-1.0989E+00
27	5.7000E-02	5.7568E-02	2.5147E-03	-5.6772E-04	-9.9600E-01
28	5.9000E-02	5.1124E-02	1.6585E-03	7.8765E-03	1.3350E+01
29	5.9000E-02	6.3913E-02	3.9338E-03	-4.9129E-03	-8.3270E+00
30	5.9000E-02	5.1376E-02	1.8346E-03	7.6241E-03	1.2922E+01
31	6.0000E-02	5.9602E-02	3.2884E-03	3.9783E-04	6.6305E-01
32	6.0000E-02	4.9137E-02	1.3445E-03	1.0863E-02	1.8105E+01
33	6.1000E-02	5.3765E-02	1.4417E-03	7.2349E-03	1.1860E+01
34	6.2000E-02	5.1357E-02	1.6744E-03	1.0643E-02	1.7166E+01
35	6.3000E-02	5.4293E-02	1.7079E-03	8.7070E-03	1.3821E+01
36	6.4000E-02	5.4041E-02	1.4068E-03	9.9595E-03	1.5562E+01
37	6.5000E-02	5.6829E-02	1.9441E-03	8.1708E-03	1.2571E+01
38	6.9000E-02	6.0530E-02	3.5140E-03	8.4702E-03	1.2276E+01
39	7.1000E-02	6.3066E-02	3.8121E-03	7.9343E-03	1.1175E+01

Rezidualní součet čtverců, RSC : 1.9372E-03

Průměr absolutních hodnot reziduí, Me : 5.7338E-03

Průměr relativních reziduí, Mer : 1.1534E+01

Odhad reziduálního rozptylu, s²(e) : 5.5349E-05

Odhad směrodatné odchylky reziduí, s(e) : 7.4397E-03

Odhad šikmosti reziduí, g1(e) : -1.1683E-01

Odhad špičatosti reziduí, g2(e) : 2.0546E+00

Napověda-F1 Řádek: 166 - 188 Celkem: 351 Délka: 21660

VÝSLEDKY

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test H0: B _j = 0 vs. H _A : B _j <> 0	t-kriterium	hypoteza H0 je	Hlad. výz.
B[0]	0.0000E+00					
B[1]	4.8617E-03	1.1251E-03	4.3213E+00		Zamítnuta	0.000
B[2]	9.7034E-03	3.1061E-03	3.1239E+00		Zamítnuta	0.004
B[3]	-9.8087E-03	3.2581E-03	-3.0106E+00		Zamítnuta	0.005
B[4]	-1.1309E-02	3.1866E-03	-3.5489E+00		Zamítnuta	0.001

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R : 6.2757E-01

Koeficient determinace, R² : 3.9384E-01

Predikovaný korelační koeficient, R_p² : 5.0079E-01

Střední kvadratická chyba predikce, MEP : 6.1395E-05

Akaikeho informační kritérium, AIC : -3.7849E+02

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	yexp[i]	yyp[i]	s(yyp[i])	e[i]	er[i]
1	3.6000E-02	4.3876E-02	2.6201E-03	-7.8761E-03	-2.1878E+01

Napověda-F1 Řádek: 120 - 142 Celkem: 351 Délka: 21660

VÝSLEDKY

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

Fisher-Snedocorův test významnosti regrese, F : 7.5802E+00

Tabulkový kvantil, F(1-alpha,m-1,n-m) : 2.8742E+00

Závěr: Navržený model je přijat jako významný.

Spočtená hladina významnosti : 0.000

Scottovo kritérium multikolinearity, M : -2.4586E-01

Závěr: Navržený model je korektní.

Cook-Weisbergův test heteroskedasticity, Sf : 4.7860E+01

Tabulkový kvantil, Chi²(1-alpha,1) : 3.8415E+00

Závěr: Rezidua vykazují heteroskedasticitu.

Spočtená hladina významnosti : 0.000

Jarque-Berraův test normality reziduí, L(e) : 1.5412E+00

Tabulkový kvantil, Chi²(1-alpha,2) : 5.9915E+00

Závěr: Normalita je přijata.

Spočtená hladina významnosti : 0.463

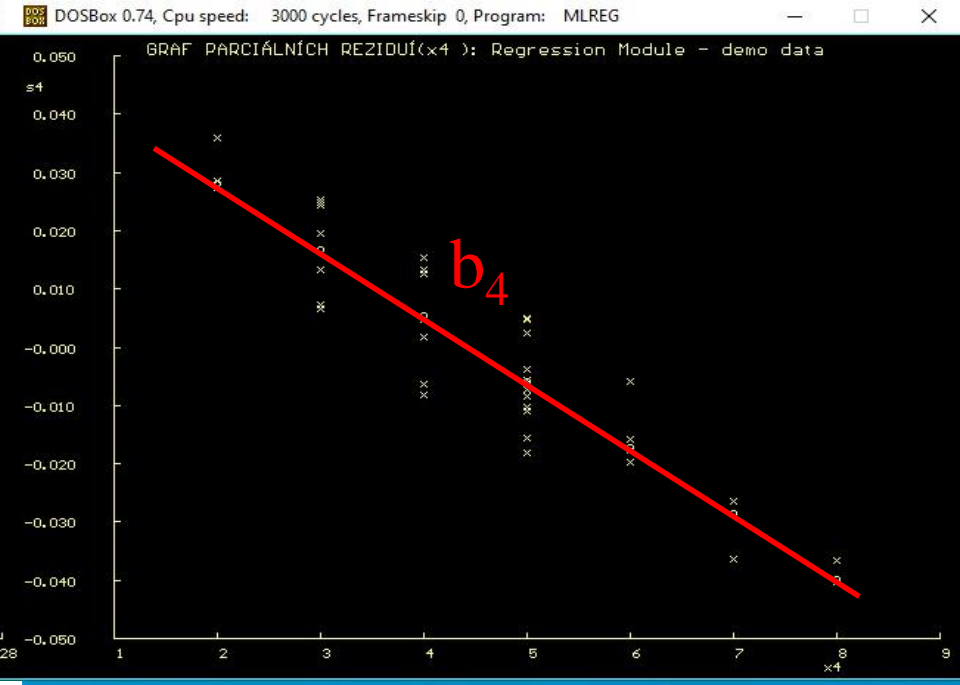
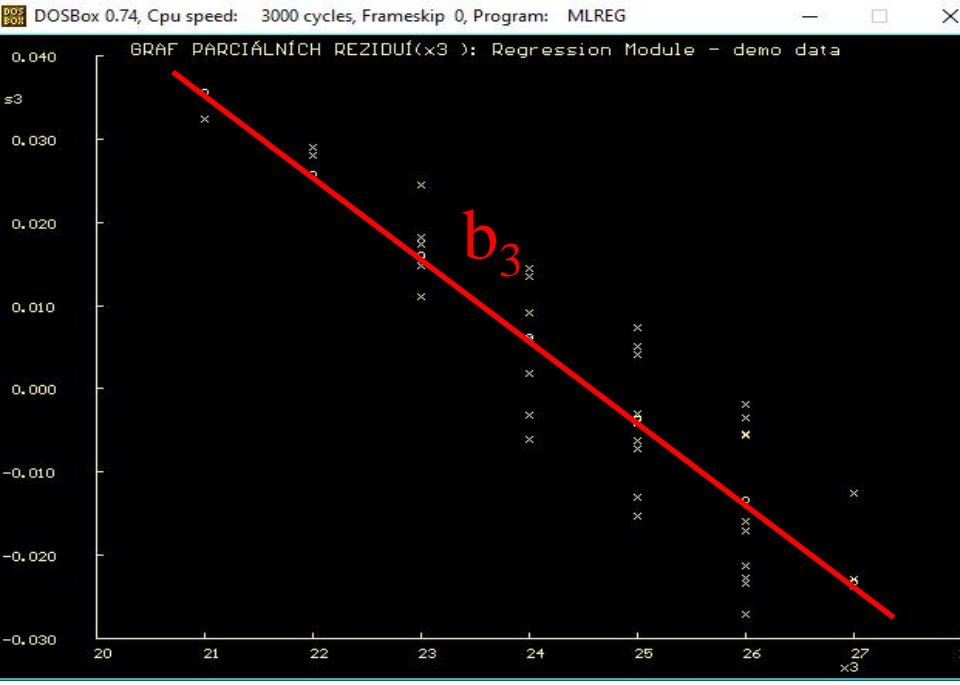
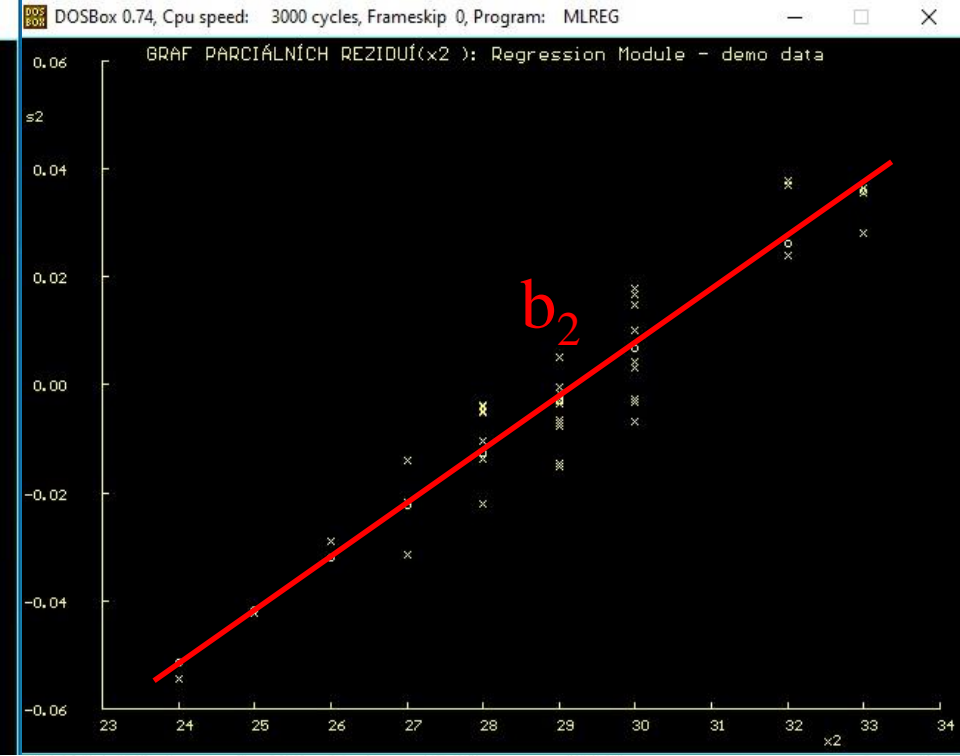
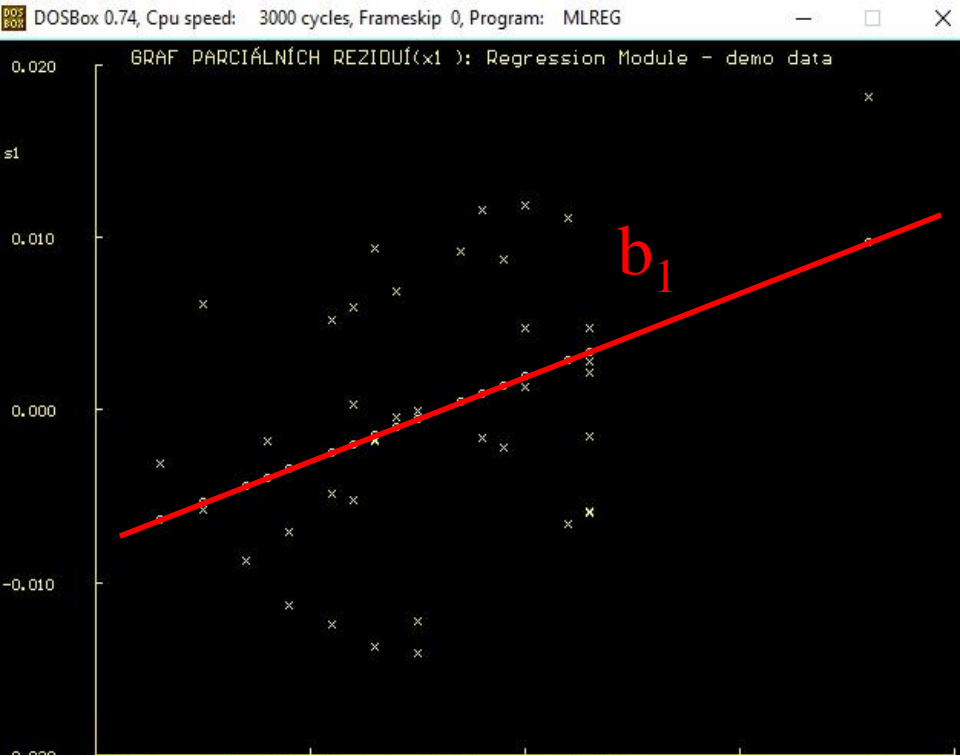
Waldův test autokorelace, Wa : 2.4670E+01

Tabulkový kvantil, Chi²(1-alpha,1) : 3.8415E+00

Závěr: Rezidua jsou autokorelována.

Spočtená hladina významnosti : 0.000

Napověda-F1 Řádek: 191 - 213 Celkem: 351 Délka: 21660



Závěry a vysvětlení výstupu úlohy M654

Úkoly

Odpovědi, závěry a vysvětlení

(1) Vyšetřením regresního tripletu nalezněte nejlepší model.

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: \beta_j = 0$ vs. $H_A: \beta_j \neq 0$	t-kriterium	hypoteza H_0 je	Hlad. výz.
BI 01	2.1120E-02	4.4206E-02		4.7776E-01	Akceptována	0.635
BI 11	4.1002E-03	2.2602E-03		1.8141E+00	Akceptována	0.077
BI 21	8.4721E-03	4.5779E-03		1.8507E+00	Akceptována	0.071
BI 31	-9.3009E-03	4.7140E-03		-1.9730E+00	Akceptována	0.055
BI 41	-9.2695E-03	4.5674E-03		-2.0295E+00	Zamítnuta	0.049

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 4.9680E-01
Koeficient determinace, R^2	: 2.4681E-01
Predikovaný korelační koeficient, R_p^2	: 1.4720E-01
Střední kvadratická chyba predikce, MEP	: 1.6386E-04
Akaikeho informační kritérium, AIC	:-4.2094E+02

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

Fisher-Snedocorův test významnosti regrese, F	: 3.5226E+00
Tabulkový kvantil, $F(1-\alpha, m-1, n-m)$: 2.5888E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.014
Scottovo kritérium multikolinearity, M	:-2.2108E-02
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, S _f	: 9.7893E+00
Tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.	
Spočtená hladina významnosti	: 0.002
Jarque-Berraův test normality reziduí, L(e)	: 3.2205E+00
Tabulkový kvantil, $\chi^2(1-\alpha, 2)$: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.200
Waldův test autokorelace, W _a	: 5.7514E+01
Tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3.8415E+00
Závěr: Rezidua jsou autokorelována.	
Spočtená hladina významnosti	: 0.000

(2) Využijte regresní diagnostiku a parciální reziduální grafy ke statistické významnosti jednotlivých parametrů.

Výše uvedené parciální reziduální grafy b_1, b_2, b_3, b_4 ukazují sklon směrnice v souladu se statistickou významností každého parametru b .