

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 VE VÝSTAVBĚ REGRESNÍHO MODELU

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)

PODSTATNÉ TESTY VÝZNAMNOSTI V KORELAČNÍ A REGRESNÍ ANALÝZE

- ◆ test významnosti korelačního koeficientu
- ◆ test významnosti modelu jako celku
- ◆ test významnosti jednotlivých regresních parametrů
- ◆ test shody lineárních regresních modelů

a mnoho dalších testů.....

TEST VÝZNAMNOSTI R

Test významnosti odpovídá, zda je korelace R mezi výběrovými proměnnými natolik silná, abychom ji mohli považovat za prokázanou a statisticky významnou i pro základní soubor ρ .

Pro párový R :
$$t_R = \frac{R \cdot \sqrt{n-2}}{\sqrt{1-R^2}}$$
 $t_{\alpha, n-2}$ n je počet hodnot výběru

Pro násobný R :
$$F_R = \frac{R^2(n-m)}{(1-R^2)(m-1)}$$
 $t_{\alpha, n-m}$ m je počet proměnných

Pro parciální R :
$$t_R = \frac{R \cdot \sqrt{n-k-2}}{\sqrt{1-R^2}}$$
 $t_{\alpha, n-k-2}$ k je počet „vyločených“ proměnných

TEST VÝZNAMNOSTI REGRESNÍHO MODELU V JEHO VÝSTAVBĚ

Co vlastně testujeme?

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_mx_m$$

1. Krok: Testujeme významnost odhadů jednotlivých parametrů:

když je testovaný odhad parametru statisticky nevýznamný, pak jeho příslušná proměnná x_j **nepřispívá** ke zpřesnění odhadu závisle proměnné y a tato proměnná x_j je pak v modelu zbytečná.

2. Krok: Testujeme model jako celek:

tj. zda příslušná kombinace všech nezávisle proměnných statisticky významně zpřesní odhad závisle proměnné y oproti použití pouhého průměru hodnot y .

1. krok: TEST VÝZNAMNOSTI REGRESNÍCH PARAMETRŮ

$\mathbf{H}_0: \beta_j = 0$, tj. j -tý regresní parametr je nevýznamný

$$t = \frac{b_j - \beta_j}{s_b} \quad \text{pro } \beta_j = 0 \quad \longrightarrow \quad t = \frac{b_j}{s_b}$$

Pokud platí, že $|t| > t_{\alpha/2; n-m}$, potom je j -tý regresní parametr statisticky významný a příslušná proměnná musí zůstat v modelu.

2. krok: HODNOCENÍ KVALITY LINEÁRNÍHO REGRESNÍHO MODELU

Střední kvadratická chyba predikce (MEP)

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

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

Akaikovo informační kritérium (AIC)

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

RSC reziduální součet čtverců
 m počet parametrů

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

Úlohy na TESTY STATISTICKÉ VÝZNAMNOSTI PARAMETRU V LINEÁRNÍM REGRESNÍM MODELU

Jak vysvětlit výstup programu a napsat 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 J6.24 Závislost tělesného tuku atletů-běžců na obsahu tuku ve stravě

Zadání: Postavte závislost mezi tělesným tukem lehkých atletů, kteří týdně trénují asi 12 hodin a zkonsumovaným tukem v jejich stravě x . U náhodného vzorku 18 běžců byl měřen tělesný podkožní tuk y [%] a sledován v závislosti na zkonsumovaném tuku ve stravě x [%].

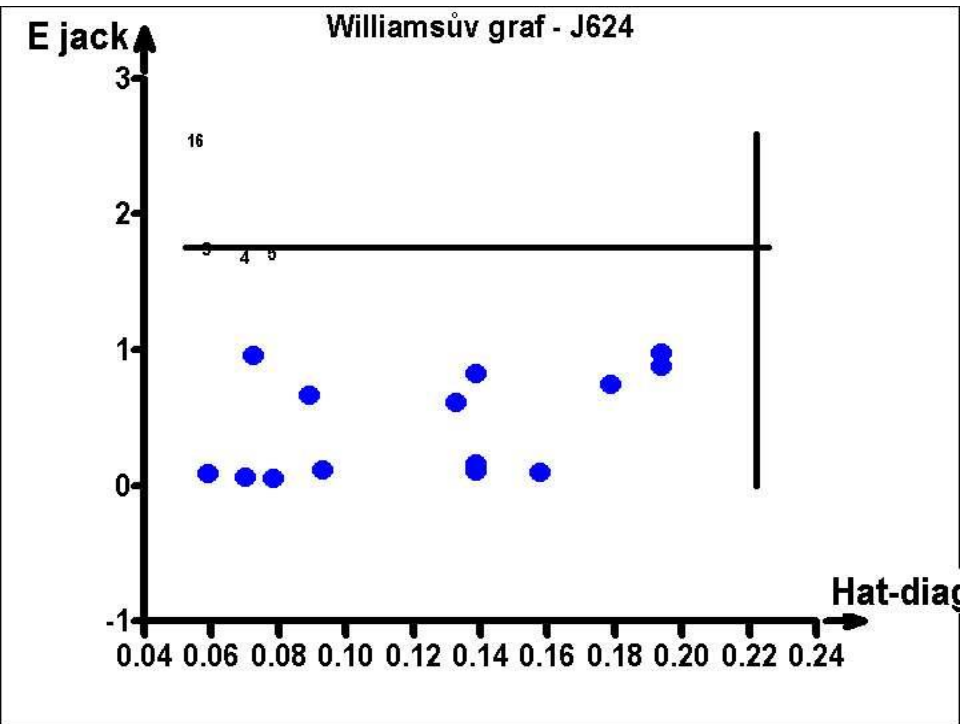
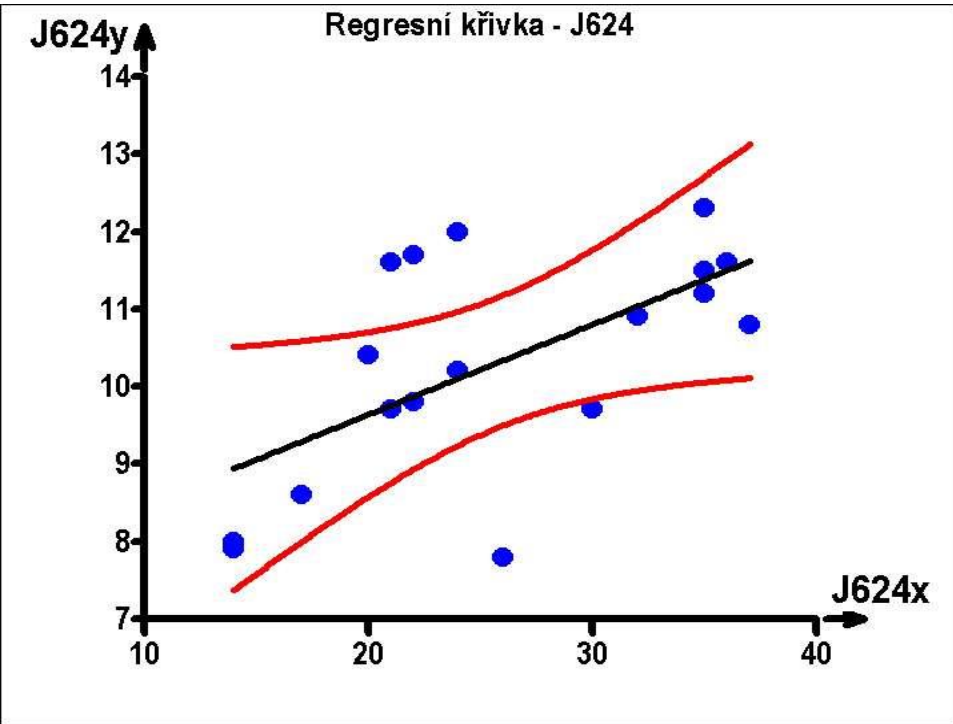
Úkoly:

Závislost lze popsat jednoduchým lineárním regresním modelem $y = \beta_0 + \beta_1 x$ a testujte, zda jsou oba parametry statisticky významné.

Data: Spotřebovaný tuk ve stravě x [%], tělesný podkožní tuk y [%]:

x	y
22	9.8
...	...
14	7.9

Software QC-EXPERT 3.1 (TriloByte)



Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	7.314069	1.005954	Významný	0.000002	5.181541	9.446597
J624x	0.11623	0.037399	Významný	0.006768	0.036947	0.195512

Statistické charakteristiky regrese

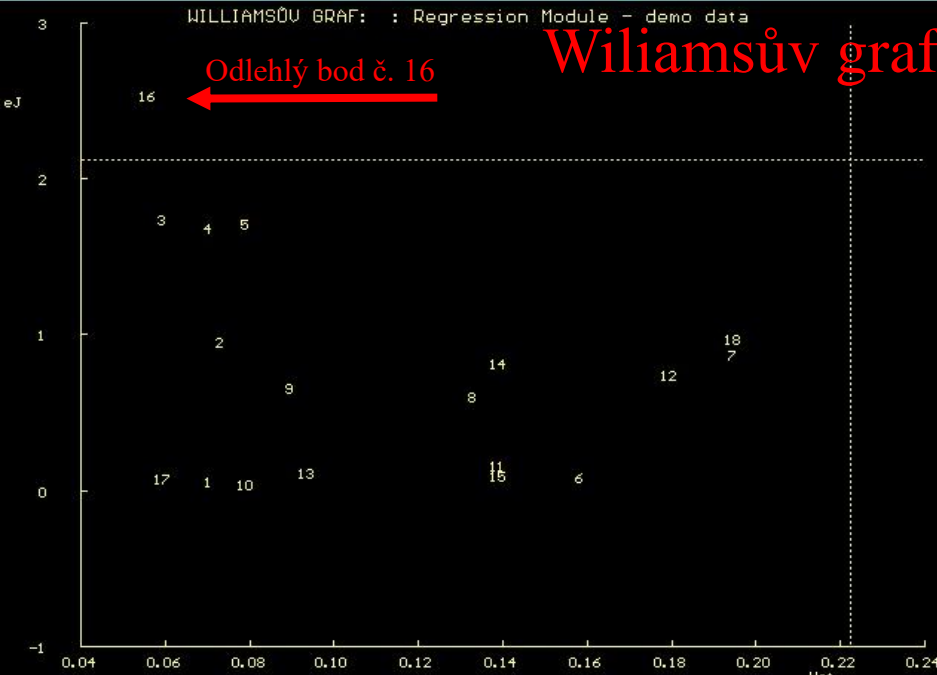
Vícenásobný korelační koeficient R :	0.6135
Koeficient determinace R ² :	0.3764
Predikovaný korelační koeficient Rp :	0.0629
Střední kvadratická chyba predikce MEP :	1.5092
Akaikeho informační kritérium :	8.1074

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	9.658629
Kvantil F (1-alfa, m-1, n-m) :	4.493998
Pravděpodobnost :	0.006768

Závěr : **Model je významný**



Kritika regresního modelu

(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]	7.3141E+00	1.0060E+00	7.2708E+00	Zamítnuta	0.000	
B[1]	1.1623E-01	3.7399E-02	3.1078E+00	Zamítnuta	0.007	

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 6.1354E-01
Koeficient determinace, R ²	: 3.7643E-01
Predikovaný korelační koeficient, R _p ²	: 5.0093E-01
Střední kvadratická chyba predikce, MEP	: 1.5092E+00
Akaikeho informační kritérium, AIC	: 8.1074E+00

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylka	Klasické reziduum	Relativní reziduum
i	y _{expl}	y _{ypf}	s(y _{ypf})	e _{il}	er _{il}
1	9.8000E+00	9.8711E+00	3.1476E-01	-7.1120E-02	-7.2571E-01
2	9.7000E+00	1.0801E+01	3.2063E-01	-1.1010E+00	-1.1350E+01
3	1.2000E+01	1.0104E+01	2.8848E-01	1.8964E+00	1.5804E+01
4	1.1700E+01	9.8711E+00	3.1476E-01	1.8289E+00	1.5631E+01

Napověda-F1 Řádek: 68 - 90 Celkem: 212 Délka: 11781

Těsnost proložení statistickou analýzou reziduí

4	1.1700E+01	9.8711E+00	3.1476E-01	1.8289E+00	1.5631E+01
5	1.1600E+01	9.7549E+00	3.3346E-01	1.8451E+00	1.5906E+01
6	1.1600E+01	1.1498E+01	4.7232E-01	1.0167E-01	8.7643E-01
7	8.0000E+00	8.9413E+00	5.2381E-01	-9.4128E-01	-1.1766E+01
8	8.6000E+00	9.2900E+00	4.3319E-01	-6.8997E-01	-8.0229E+00
9	1.0400E+01	9.6387E+00	3.5513E-01	7.6134E-01	7.3206E+00
10	9.7000E+00	9.7549E+00	3.3346E-01	-5.4891E-02	-5.6588E-01
11	1.1200E+01	1.1382E+01	4.4277E-01	-1.8210E-01	-1.6259E+00
12	1.0800E+01	1.1615E+01	5.0292E-01	-8.1456E-01	-7.5423E+00
13	1.0900E+01	1.1033E+01	3.6292E-01	-1.3342E-01	-1.2240E+00
14	1.2300E+01	1.1382E+01	4.4277E-01	9.1790E-01	7.4626E+00
15	1.1500E+01	1.1382E+01	4.4277E-01	1.1790E-01	1.0252E+00
16	7.8000E+00	1.0336E+01	2.8028E-01	-2.5360E+00	-3.2513E+01
17	1.0200E+01	1.0104E+01	2.8848E-01	9.6421E-02	9.4530E-01
18	7.9000E+00	8.9413E+00	5.2381E-01	-1.0413E+00	-1.3181E+01

Rezidualní součet čtverců, RSC	: 2.2614E+01
Průměr absolutních hodnot reziduí, Me	: 8.4063E-01
Průměr relativních reziduí, Mer	: 8.5271E+00
Odhad reziduálního rozptylu, s ² (e)	: 1.4134E+00
Odhad směrodatné odchylky reziduí, s(e)	: 1.1889E+00
Odhad špičatosti reziduí, g1(e)	: -2.9149E-03
Odhad špičatosti reziduí, g2(e)	: 1.8940E+00

Napověda-F1 Řádek: 90 - 112 Celkem: 212 Délka: 11781

Kritika regresní metody

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

Fisher-Snedecorův test významnosti regrese, F	: 9.6586E+00
Tabulkový kvantil, F(1-alpha, m-1, n-m)	: 4.4940E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.007

Scottovo kritérium multikolinearity, M	: 4.5981E-16
Závěr: Navržený model je korektní.	

Cook-Weisbergův test heteroskedasticity, Sf	: 6.8222E+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)	: 8.4319E-03
Tabulkový kvantil, Chi ² (1-alpha, 2)	: 5.9915E+00

Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.996

Waldův test autokorelace, Wa	: 8.6743E-01
Tabulkový kvantil, Chi ² (1-alpha, 1)	: 3.8415E+00

Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.352

Napověda-F1 Řádek: 115 - 137 Celkem: 212 Délka: 11781

Kritika regresního modelu

2.200000 E+01	9.800000
3.000000 E+01	9.700000
2.400000 E+01	1.200000 E+01
2.200000 E+01	1.170000 E+01
2.100000 E+01	1.160000 E+01
3.600000 E+01	1.160000 E+01
1.400000 E+01	8.000000
1.700000 E+01	8.600000
2.000000 E+01	1.040000 E+01
2.100000 E+01	9.700000
3.500000 E+01	1.120000 E+01
3.700000 E+01	1.080000 E+01
3.200000 E+01	1.090000 E+01
3.500000 E+01	1.230000 E+01
3.500000 E+01	1.150000 E+01
2.400000 E+01	1.020000 E+01
1.400000 E+01	7.900000

Odstranění odlehleho bodu

(3) ODHADY PARAMETRŮ A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylnka	Test H0: B _{ij} = 0 vs. HA: B _{ij} > 0	t-kriterium	hypoteza H0 je	Hlad. výz.
BI 01	7.4518E+00	8.7022E-01	8.5631E+00	Zamítnuta		0.000
BI 11	1.1667E-01	3.2290E-02	3.6132E+00	Zamítnuta		0.003

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 6.8216E-01
Koeficient determinace, R ²	: 4.6534E-01
Predikovaný korelační koeficient, R _p ²	: 5.6909E-01
Střední kvadratická chyba predikce, MEP	: 1.1756E+00
Akaikeho informační kritérium, AIC	: 2.7597E+00

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Predikovaná hodnota	Směrodatná odchylnka	Klasické reziduum	Relativní reziduum
i	yexpli	yvypil	s(yvypil)	elil	erli
1	9.8000E+00	1.0019E+01	2.7788E-01	-2.1861E-01	-2.2307E+00
2	9.7000E+00	1.0952E+01	2.8313E-01	-1.2520E+00	-1.2907E+01
3	1.2000E+01	1.0252E+01	2.5582E-01	1.7480E+00	1.4567E+01
4	1.1700E+01	1.0019E+01	2.7788E-01	1.6814E+00	1.4371E+01

Těsnost proložení statistickou analýzou reziduí

3	1.2000E+01	1.0252E+01	2.5582E-01	1.7480E+00	1.4567E+01
4	1.1700E+01	1.0019E+01	2.7788E-01	1.6814E+00	1.4371E+01
5	1.1600E+01	9.9019E+00	2.9366E-01	1.6981E+00	1.4639E+01
6	1.1600E+01	1.1652E+01	4.1226E-01	-5.2020E-02	-4.4845E-01
7	8.0000E+00	9.0852E+00	4.5578E-01	-1.0852E+00	-1.3565E+01
8	8.6000E+00	9.4352E+00	3.7836E-01	-8.3524E-01	-9.7121E+00
9	1.0400E+01	9.7853E+00	3.1199E-01	6.1474E-01	5.9110E+00
10	9.7000E+00	9.9019E+00	2.9366E-01	-2.0193E-01	-2.0818E+00
11	1.1200E+01	1.1535E+01	3.8701E-01	-3.3535E-01	-2.9942E+00
12	1.0800E+01	1.1769E+01	4.3843E-01	-9.6869E-01	-8.9694E+00
13	1.0900E+01	1.1185E+01	3.1899E-01	-2.8533E-01	-2.6177E+00
14	1.2300E+01	1.1535E+01	3.8701E-01	7.6465E-01	6.2167E+00
15	1.1500E+01	1.1535E+01	3.8701E-01	-3.5348E-02	-3.0737E-01
16	1.0200E+01	1.0252E+01	2.5582E-01	-5.1950E-02	-5.0932E-01
17	7.9000E+00	9.0852E+00	4.5578E-01	-1.1852E+00	-1.5003E+01

Rezidualní součet čtverců, RSC	: 1.5804E+01
Průměr absolutních hodnot reziduí, Me	: 7.6552E-01
Průměr relativních reziduí, Mer	: 7.4735E+00
Odhad reziduálního rozptylu, s ² (e)	: 1.0536E+00
Odhad směrodatné odchylnky reziduí, s(e)	: 1.0264E+00
Odhad šikmosti reziduí, g1(e)	: 6.0334E-01
Odhad špičatosti reziduí, g2(e)	: 2.2690E+00

Kritika regresní metody

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

Fisher-Snedocorřův test významnosti regrese, F	: 1.3055E+01
Tabulkový kvantil, F(1-alpha,m-1,n-m)	: 4.5431E+00
Závěr: Navržený model je přijat jako významný.	
Spočtená hladina významnosti	: 0.003
Scottovo kritérium multikolinearity, M	: 6.8034E-17
Závěr: Navržený model je korektní.	
Cook-Weisbergův test heteroskedasticity, Sf	: 5.4463E+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)	: 1.4099E+00
Tabulkový kvantil, Chi ² (1-alpha,2)	: 5.9915E+00
Závěr: Normalita je přijata.	
Spočtená hladina významnosti	: 0.494
Waldův test autokorelace, Wa	: 1.9355E+00
Tabulkový kvantil, Chi ² (1-alpha,1)	: 3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.164

Závěry a vysvětlení výstupu úlohy J624

Úkoly

(1) Závislost lze popsat lineárním regresním modelem $y = \beta_0 + \beta_1 x$.

Výsledky jsou po odstranění odlehlého bodu č. 16.

(2) Testujte, zda jsou oba parametry přímky statisticky významné.

Odpovědi, závěry a vysvětlení

```
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[j] = 0 vs. HA: B[j] <> 0
              odchylka      t-kriterium      hypoteza H0 je      Hlad. výz.
BI 01         7.4518E+00  8.7022E-01      8.5631E+00          Zamítnuta           0.000
BI 11         1.1667E-01  3.2290E-02      3.6132E+00          Zamítnuta           0.003

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:
Vícenásobný korelační koeficient, R                : 6.8216E-01
Koeficient determinace, R^2                         : 4.6534E-01
Predikovaný korelační koeficient, Rp^2             : 5.6909E-01
Střední kvadratická chyba predikce, MEP            : 1.1756E+00
Akaikeho informační kritérium, AIC                  : 2.7597E+00

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod      Meřená      Predikovaná      Směrodatná      Klasické      Relativní
         hodnota      hodnota          odchylka        reziduum      reziduum
i        yexpl[i]    yvyp[i]          s(yvyp[i])     e[i]          er[i]
1        9.8000E+00  1.0019E+01      2.7788E-01     -2.1861E-01  -2.2307E+00
2        9.7000E+00  1.0952E+01      2.8313E-01     -1.2520E+00  -1.2907E+01
3        1.2000E+01  1.0252E+01      2.5582E-01     1.7480E+00   1.4567E+01
4        1.1700E+01  1.0019E+01      2.7788E-01     1.6814E+00   1.4371E+01

Napověda-F1      Řádek: 68 - 90      Celkem: 208      Délka: 11456
```

Úsek b_0 je statisticky významný a proto přímka neprochází počátkem.

Směrnice b_1 je statisticky významná, a proto je rozličná od nuly a není rovnoběžná s x-ovou osou.

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 x ve stravě v gramech a obsah celkového cholesterolu y v mg na 100 ml krve. Byl navržen jednoduchý lineární regresní model $y = \beta_0 + \beta_1 x$.

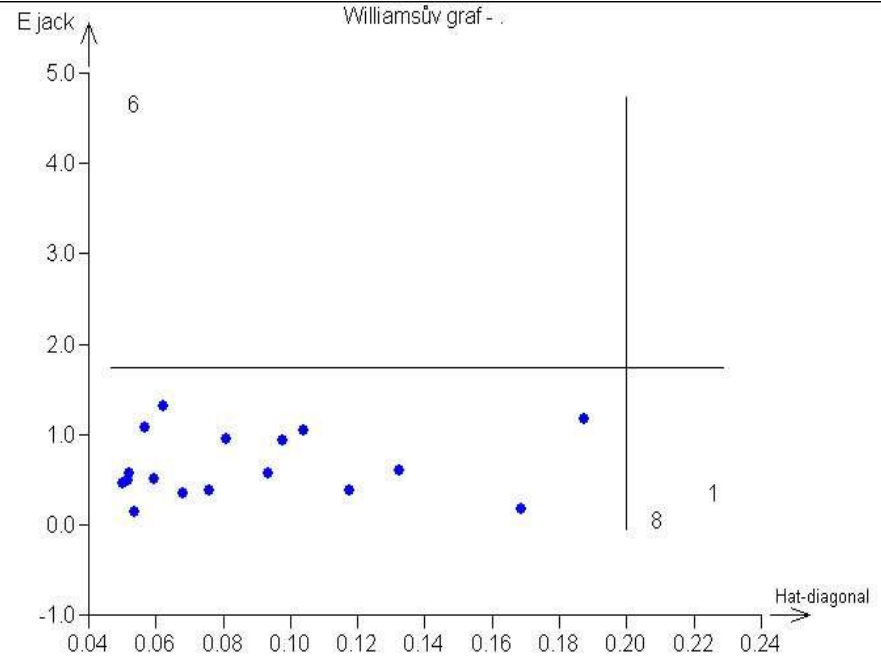
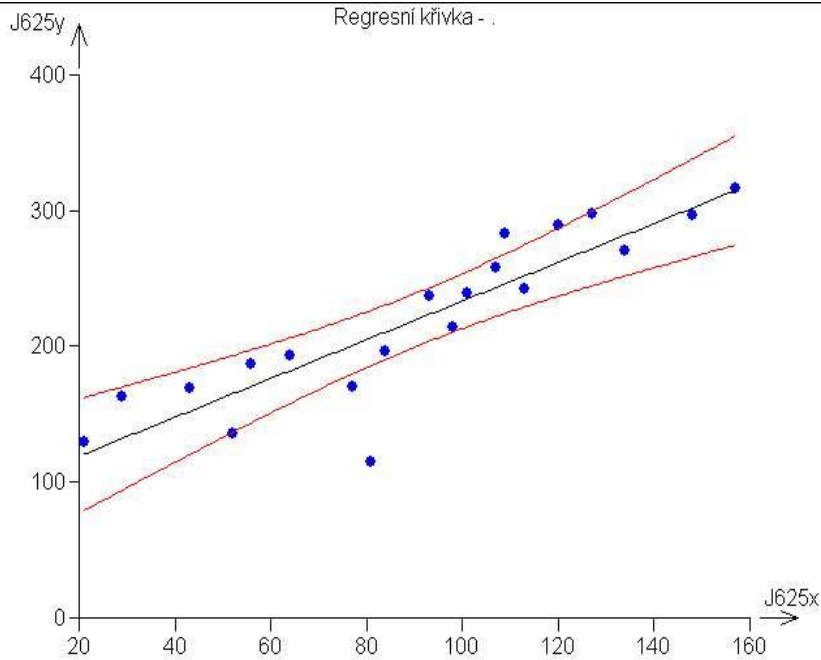
Úkoly:

- (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) Jaký je Pearsonův korelační koeficient 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

Software QC-EXPERT 3.1 (TriloByte)



Odhady parametrů

Proměnná	Odhad	Směr.odch.	Závěr
Abs	90.705314	17.420915	Významný
J625x	1.427174	0.177707	Významný

Pravděpodobnost	Spodní mez	Horní mez
0.000059	54.10533	127.305298
0.0	1.053825	1.800523

Statistické charakteristiky regrese

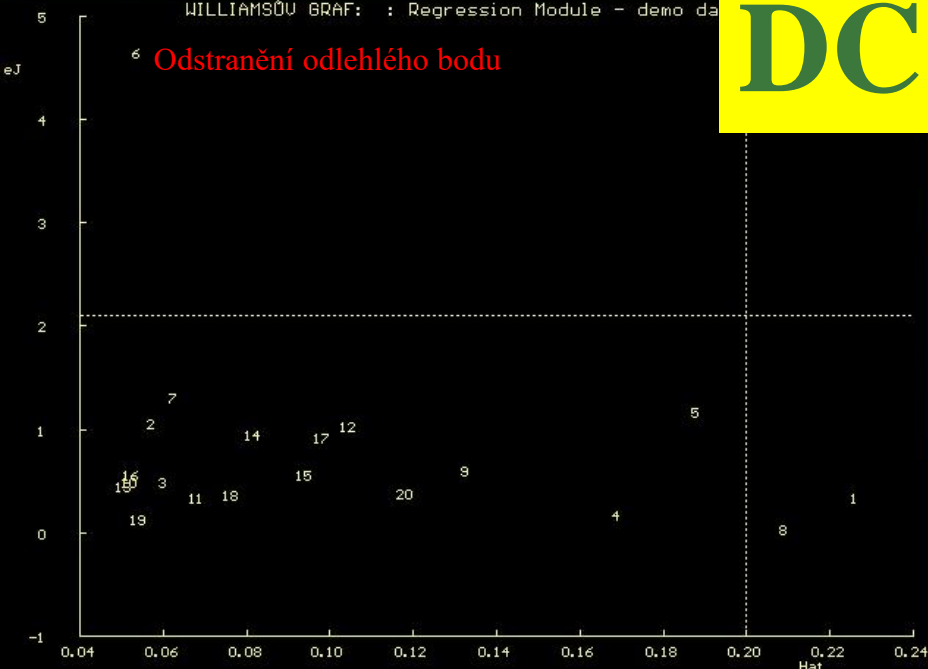
Vícenásobný korelační koeficient R :	0.884201
Koeficient determinace R ² :	0.781812
Predikovaný korelační koeficient Rp :	0.555095
Střední kvadratická chyba predikce MEP :	919.087734
Akaikeho informační kritérium :	137.35316

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

DC

6 Odstranění odlehlého bodu



Kritika regresního modelu

(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.
BI 01	9.0705E+01	1.7421E+01	Zamítnuta	5.2067E+00	Zamítnuta	0.000
BI 11	1.4272E+00	1.7771E-01	Zamítnuta	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, R _p 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	yvyp[i]	yvpil	s(yvyp[i])	elil	er[i]
1	1.3000E+02	1.2068E+02	1.4040E+01	9.3240E+00	7.1723E+00
2	1.7000E+02	2.0060E+02	7.0445E+00	-3.0598E+01	-1.7999E+01
3	2.5800E+02	2.4341E+02	7.2172E+00	1.4587E+01	5.6539E+00
4	2.9700E+02	3.0193E+02	1.2140E+01	-4.9271E+00	-1.6589E+00

Těsnost proložení statistickou analýzou reziduí

6	1.1500E+02	2.0631E+02	6.8314E+00	-9.1306E+01	-7.9397E+01
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
13	2.3700E+02	2.2343E+02	6.6230E+00	1.3567E+01	5.7247E+00
14	2.8900E+02	2.6197E+02	8.4148E+00	2.7034E+01	9.3543E+00
15	1.8700E+02	1.7063E+02	9.0400E+00	1.6373E+01	8.7556E+00
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-Snedecorův 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
Waldův test autokorelace, Wa	: 4.8191E+00
Tabulkový kvantil, Chi ² (1-alpha, 1)	: 3.8415E+00
Závěr: Rezidua jsou autokorelována.	
Spočtená hladina významnosti	: 0.028

Odstranění odlehčlého bodu

2.100000 E+01	1.300000 E+02
7.700000 E+01	1.700000 E+02
1.070000 E+02	2.580000 E+02
1.480000 E+02	2.970000 E+02
2.900000 E+01	1.630000 E+02
1.090000 E+02	2.830000 E+02
1.570000 E+02	3.160000 E+02
4.300000 E+01	1.690000 E+02
8.400000 E+01	1.960000 E+02
1.130000 E+02	2.420000 E+02
5.200000 E+01	1.360000 E+02
9.300000 E+01	2.370000 E+02
1.200000 E+02	2.890000 E+02
5.600000 E+01	1.870000 E+02
9.800000 E+01	2.140000 E+02
1.270000 E+02	2.980000 E+02
6.400000 E+01	1.930000 E+02
1.010000 E+02	2.390000 E+02
1.340000 E+02	2.710000 E+02

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	9.8595E+01	1.2013E+01	8.2073E+00	Zamítnuta	0.000	
BI 11	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, Rp^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	yexp[i]	yyp[i]	s(yyp[i])	ei[i]	er[i]
1	1.3000E+02	1.2786E+02	9.7080E+00	2.1447E+00	1.6498E+00
2	1.7000E+02	2.0588E+02	4.9415E+00	-3.5884E+01	-2.1108E+01
3	2.5800E+02	2.4768E+02	5.0118E+00	1.0315E+01	3.9982E+00
4	2.9700E+02	3.0481E+02	8.3108E+00	-7.8127E+00	-2.6305E+00

Těsnost proložení statistickou analýzou reziduí

5	1.6300E+02	1.3900E+02	8.8656E+00	2.3998E+01	1.4723E+01
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	2.3700E+02	2.2818E+02	4.6350E+00	8.8224E+00	3.7225E+00
13	2.8900E+02	2.6580E+02	5.8033E+00	2.3202E+01	8.0282E+00
14	1.8700E+02	1.7662E+02	6.3046E+00	1.0377E+01	5.5491E+00
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	2.7098E+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

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

Kritika regresní metody

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

Fisher-Snedecorův 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
Tabulkový kvantil, Chi^2(1-alpha, 1)	: 3.8415E+00
Závěr: Rezidua nejsou autokorelována.	
Spočtená hladina významnosti	: 0.932

Závěry a vysvětlení výstupu úlohy J625

Úkoly

Odpovědi, závěry a vysvětlení

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

(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.
β_0	9.8595E+01	1.2013E+01	8.2073E+00	8.2073E+00	Zamítnuta	0.000
β_1	1.3934E+00	1.2153E-01	1.1465E+01	1.1465E+01	Zamítnuta	0.000

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

Oba parametry, úsek β_0 a směrnice β_1 , jsou statisticky významné čili se liší od nuly.

(3) Jaký je Pearsonův korelační koeficient mezi celkovým cholesterolem v krvi y a denní spotřebou tuku x u sledovaných jedinců?

Původní data:

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

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

Po odstranění odlehlých bodů:

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

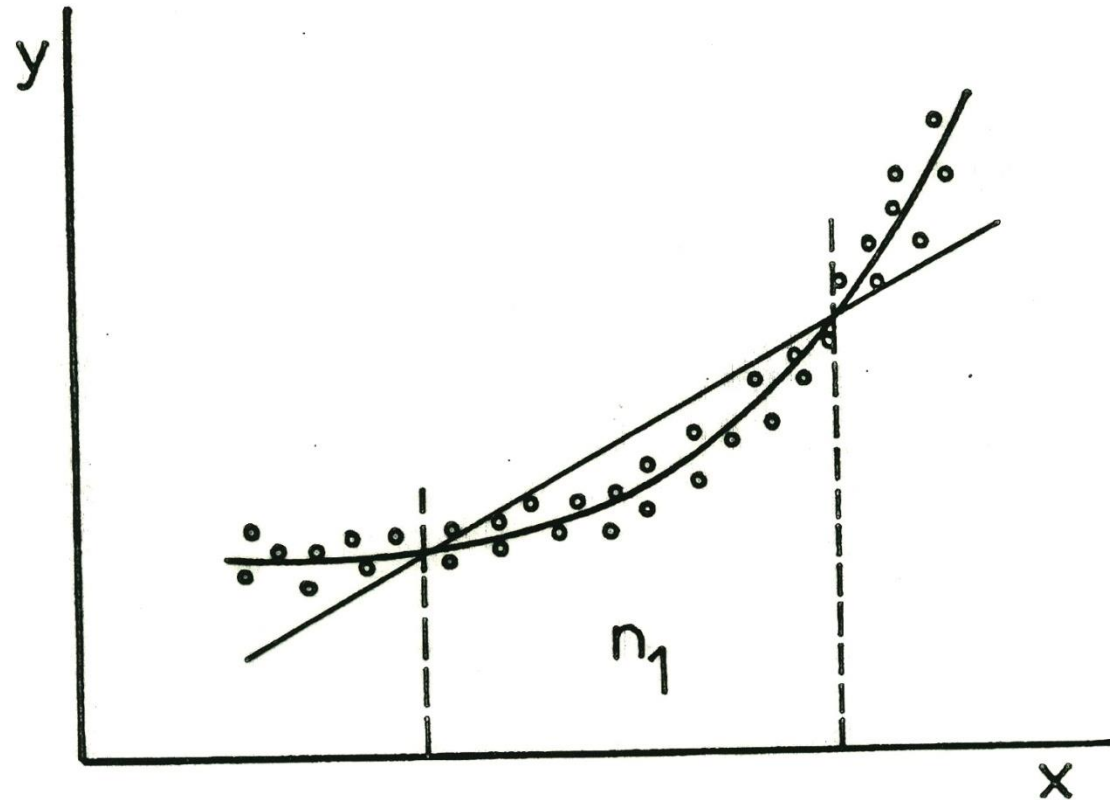
Vícenásobný korelační koeficient, R	: 9.4100E-01
Koeficient determinace, R^2	: 8.8548E-01
Předikovaný korelační koeficient, R_p^2	: 9.2796E-01
Střední kvadratická chyba predikce, MEP	: 4.4197E+02
Akaikeho informační kritérium, AIC	: 1.1607E+02

Testy vhodnosti lineárního modelu

1. Test správnosti lineárního modelu

$f(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{X} \boldsymbol{\beta}$ dle Uttsové:

H_0 : lineární model vs. H_A : nelineární model



RSC₁ regresí s využitím n₁ bodů,

RSC regresí s využitím všech n bodů.

Testační kritérium

$$F_U = \frac{(RSC - RSC_1)(n_1 - m)}{RSC_1(n - n_1)}$$

Uttsová: volit n₁ ≈ n/2 a body co nejbliže k těžišti

Test: Je-li $F_U < F_{1-\alpha}(n - n_1, n_1 - m)$, je H₀ přijata.

3. Linearita testem všech charakteristik

a) Střední kvadratická chyba predikce

$$\text{MEP} = \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \mathbf{b}_{(i)})^2$$

kde $\mathbf{b}_{(i)}$ je odhad, určený ze všech bodů kromě i -tého
Platí vztah

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

pro velké n jsou prvky $H_{ii} \approx 0$ a $\text{MEP} = \text{RSC}/n$.

b) Predikovaný koeficient determinace

$$\hat{R}_p^2 = 1 - \frac{n \text{ MEP}}{\sum_{i=1}^n y_i^2 - n\bar{y}}$$

c) Akaikovo informační kritérium

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

nejvhodnější model má AIC minimální

HODNOCENÍ KVALITY REGRESNÍHO MODELU

Střední kvadratická chyba predikce (MEP)

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

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

Akaikovo informační kritérium (AIC)

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

RSC reziduální součet čtverců
 m počet parametrů

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

Příklad 6.16 *Výběr ze tří polynomických regresních modelů*
Regresní analýzou dat vyšetřete, zda místo kvadratického modelu by lépe vyhovoval polynom třetího nebo pátého stupně.

Řešení:

a) Polynom 3. stupně:

MEP, \hat{R}_p^2 a AIC indikují polynom třetího stupně jako nejvhodnější

$$\hat{y}_p = 860.2 (\pm 85.17) - 5.057 (\pm 0.485) x \\ + 9.77 \cdot 10^{-3} (\pm 9.19 \cdot 10^{-4}) x^2 - 6.146 \cdot 10^{-6} (\pm 5.78 \cdot 10^{-7}) x^3$$

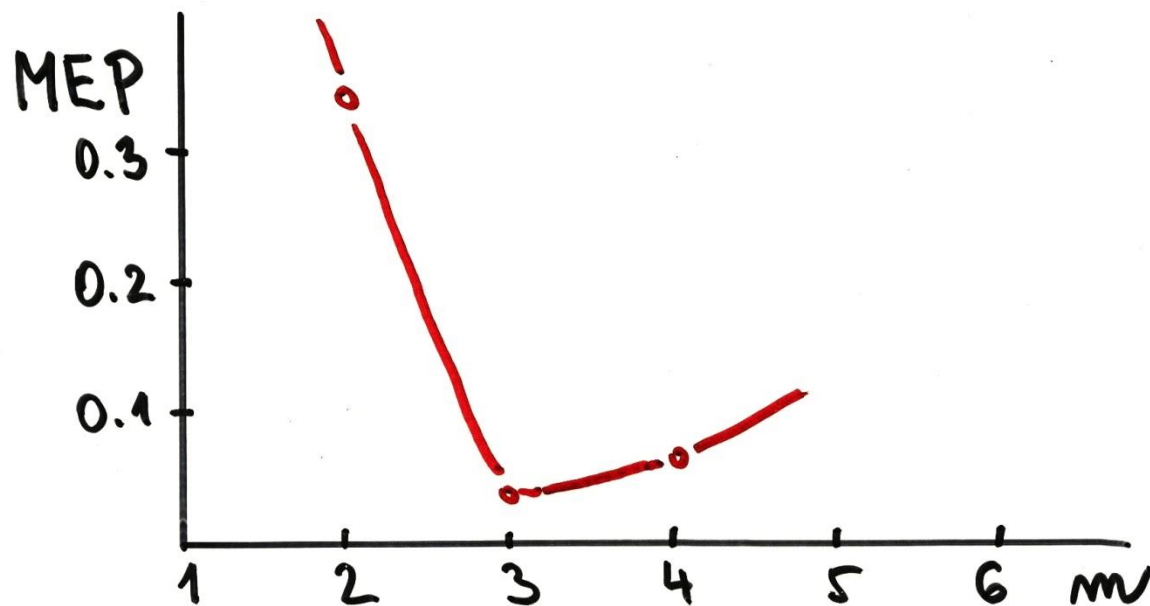
přičemž odhady všech tří parametrů vycházejí statisticky významné.

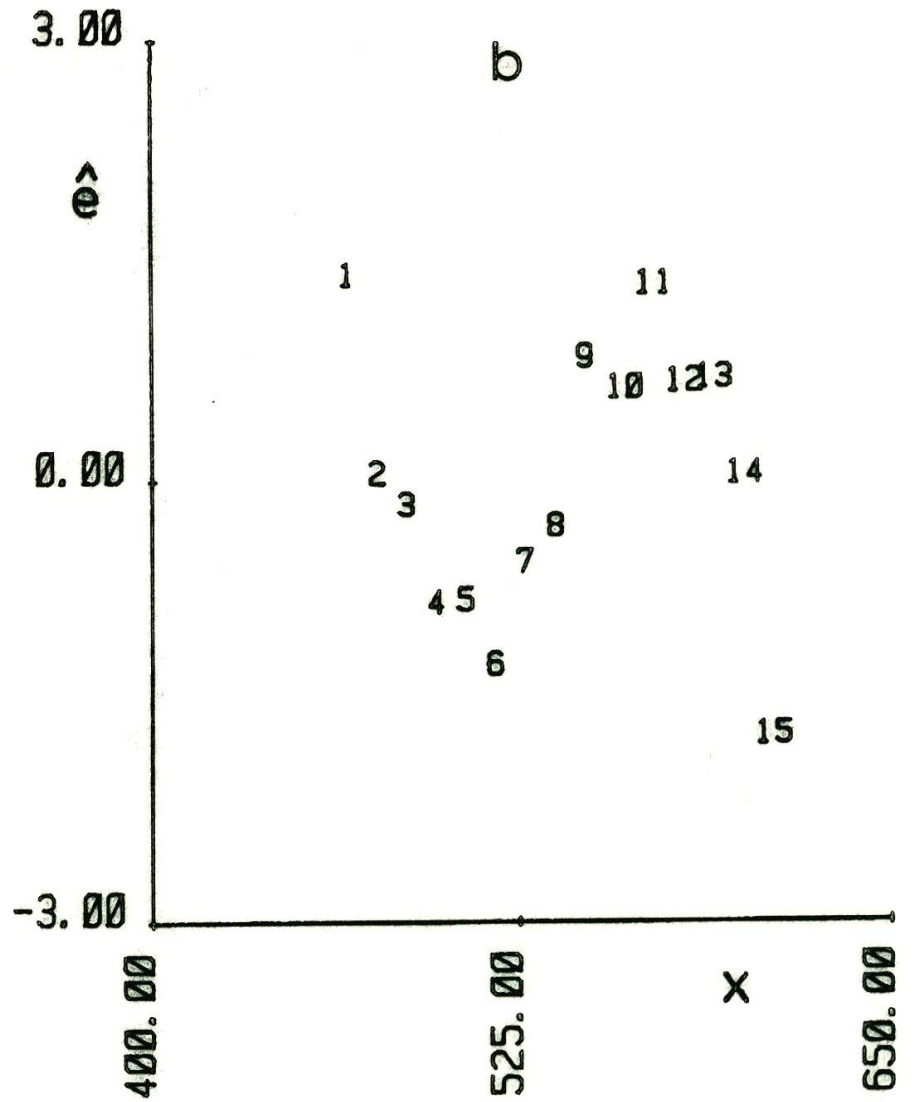
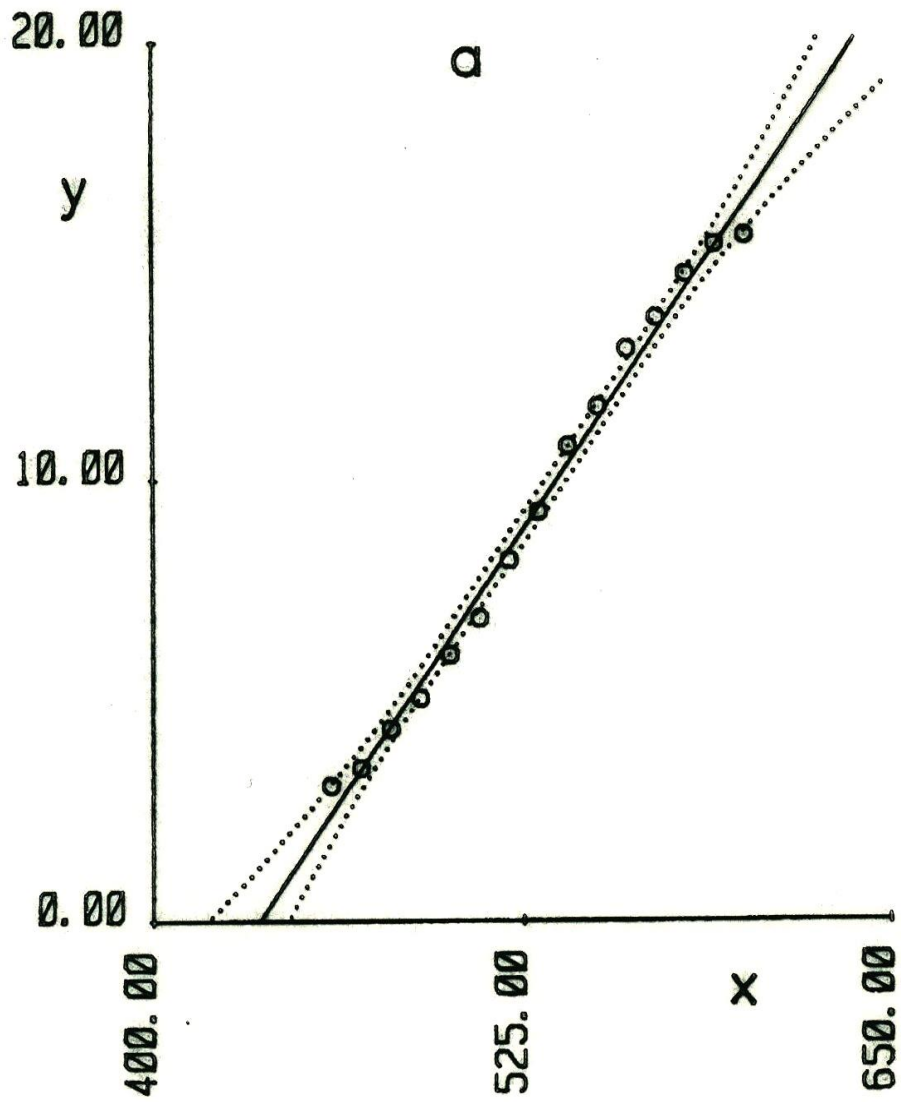
b) Polynom 5. stupně:

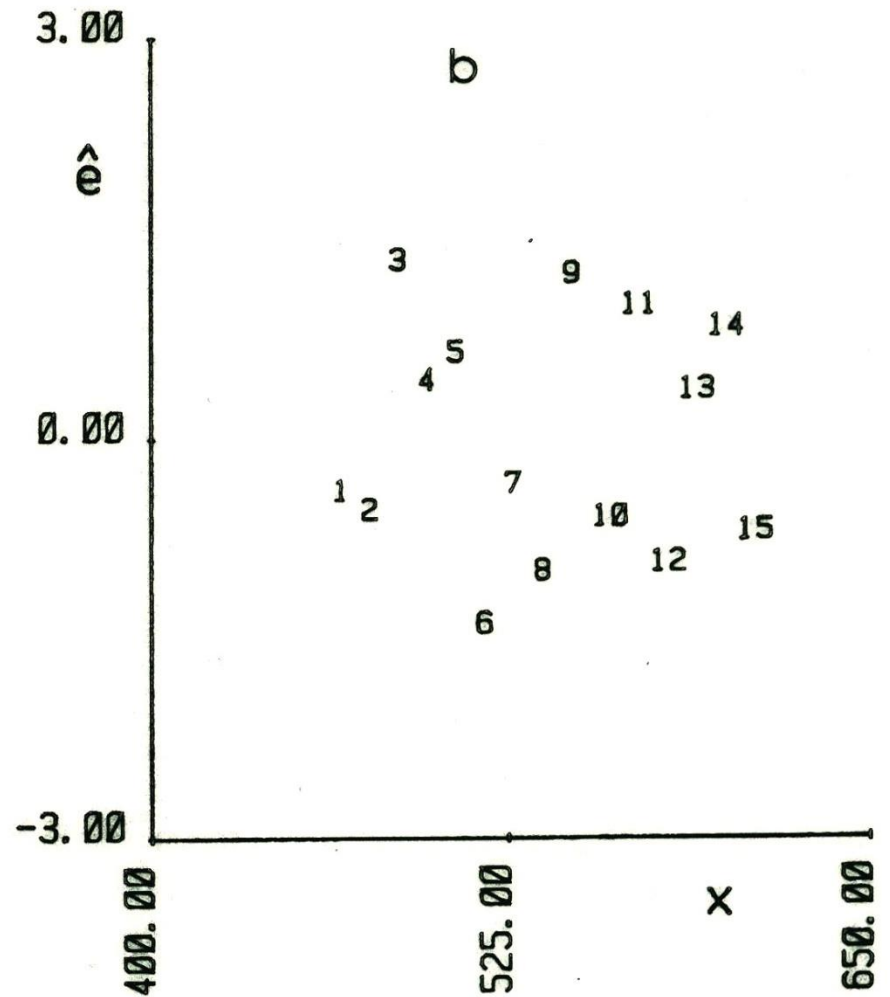
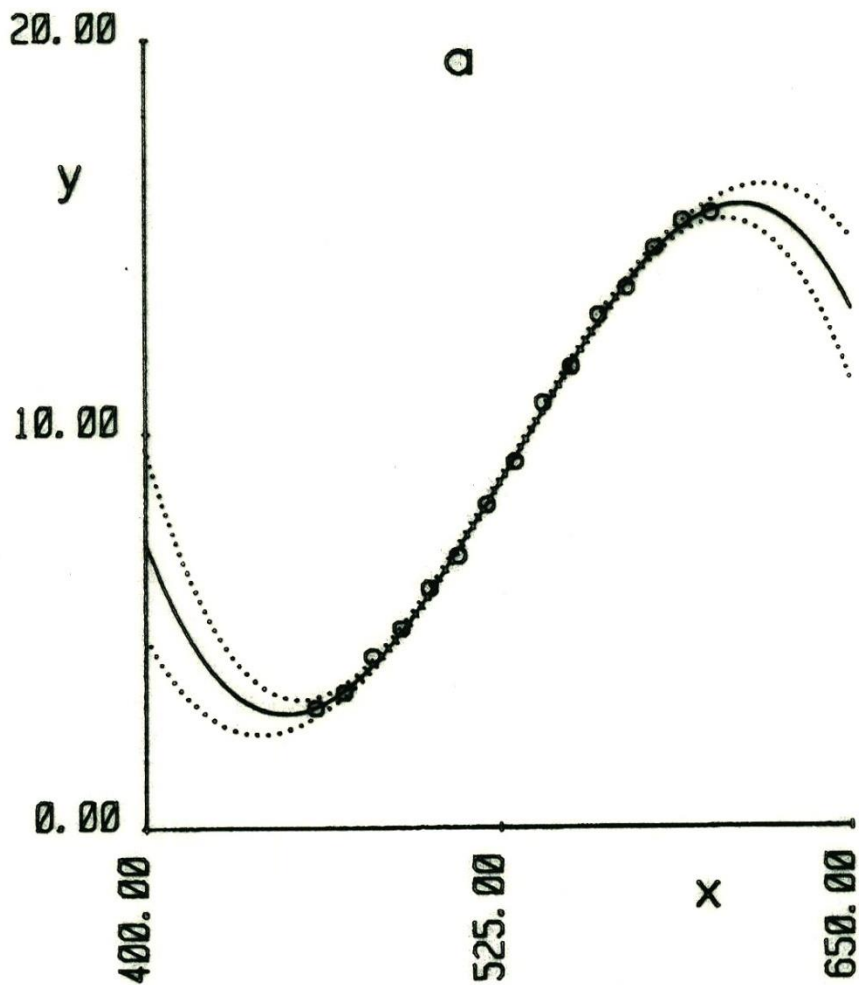
všechny parametry β kromě β_3 vycházejí statisticky nevýznamné, protože se zde projevuje multikolinearita.

Tabulka 6.3 Rozlišení stupně regresního polynomu statistikami MEP, \hat{R}_p^2 , \hat{R}^2 a AIC

Stupeň polynomu	MEP	\hat{R}_p^2	\hat{R}^2	AIC
$m = 2$	0.3502	0.9905	0.9915	-21.65
3	0.0283	0.9992	0.9992	-56.02
5	0.0613	0.9983	0.9997	-55.04

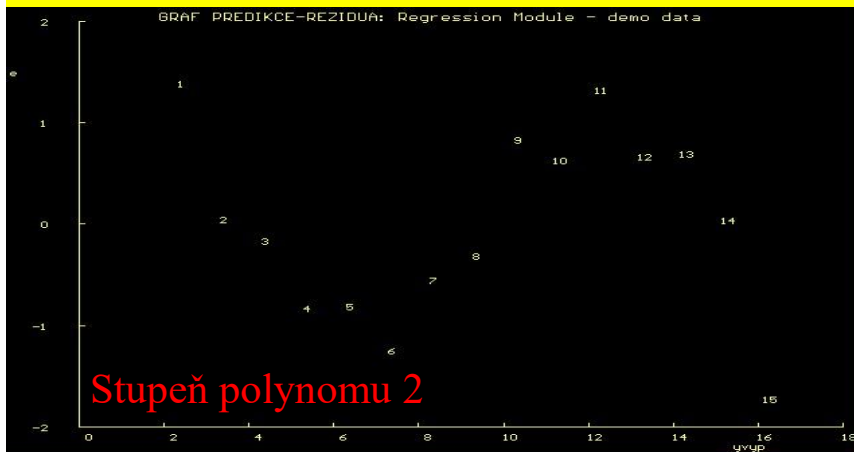






Závěr: Je patrné, že některé statistiky pro vystižení linearity modelu nebo vhodnosti specifikace modelu selhávají.

Vysvětlení výstupu úlohy P616 (data P609)



Stupeň polynomu 2

U Ý S L E D K Y

(3) ODHADY PARAMETRŮ A TESTY ÚZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylnka	Test $H_0: \beta_{ij} = 0$ vs. $H_A: \beta_{ij} \neq 0$	t-kritérium	hypoteza H_0 je	Hlad. úz.
BI 01	-4.3930E+01	1.9375E+01		-2.2673E+00	Zamítnuta	0.043
BI 11	1.0183E-01	7.3456E-02		1.3863E+00	Akceptována	0.191
BI 21	-2.5048E-06	6.9253E-05		-3.6170E-02	Akceptována	0.972

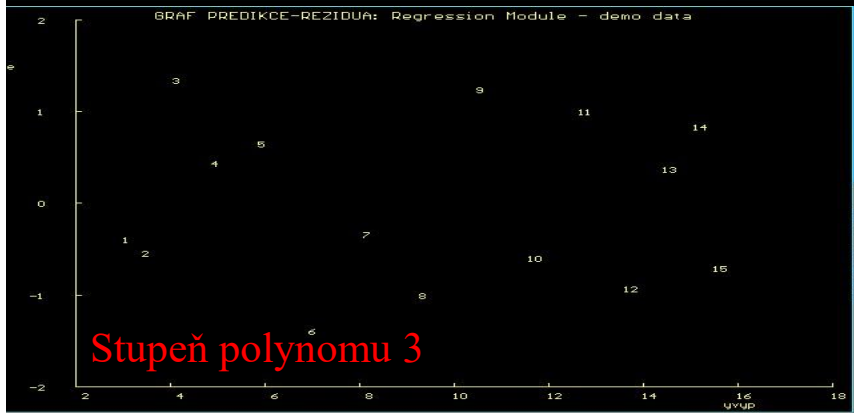
(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Učienásobný korelační koeficient, R	: 9.9572E-01
Koeficient determinace, R^2	: 9.9145E-01
Předikovaný korelační koeficient, R_p^2	: 9.9050E-01
Střední kvadratická chyba predikce, MEP	: 3.5022E-01
Ákaikeho informační kritérium, AIC	: -2.1651E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Předikovaná hodnota	Směrodatná odchylnka	Klasické reziduum	Relativní reziduum
i	yexpfil	yypfil	s(yypfil)	erfil	erfil
1	3.0000E+00	2.3832E+00	3.0322E-01	6.1676E-01	2.0559E+01
2	3.4000E+00	3.3783E+00	2.3030E-01	2.1723E-02	6.3890E-01

Napověda-F1 Rádek: 67 - 89 Celkem: 201 Délka: 11123



Stupeň polynomu 3

U Ý S L E D K Y

(3) ODHADY PARAMETRŮ A TESTY ÚZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylnka	Test $H_0: \beta_{ij} = 0$ vs. $H_A: \beta_{ij} \neq 0$	t-kritérium	hypoteza H_0 je	Hlad. úz.
BI 01	8.6023E+02	8.5176E+01		1.0099E+01	Zamítnuta	0.000
BI 11	-5.0571E+00	4.8532E-01		-1.0420E+01	Zamítnuta	0.000
BI 21	9.7701E-03	9.1857E-04		1.0636E+01	Zamítnuta	0.000
BI 31	-6.1463E-06	5.7756E-07		-1.0642E+01	Zamítnuta	0.000

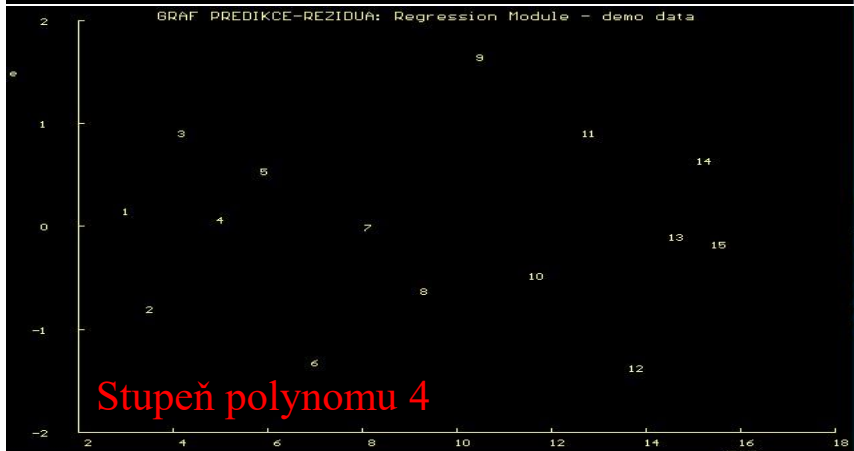
(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Učienásobný korelační koeficient, R	: 9.9962E-01
Koeficient determinace, R^2	: 9.9924E-01
Předikovaný korelační koeficient, R_p^2	: 9.9924E-01
Střední kvadratická chyba predikce, MEP	: 2.8291E-02
Ákaikeho informační kritérium, AIC	: -5.6017E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Předikovaná hodnota	Směrodatná odchylnka	Klasické reziduum	Relativní reziduum
i	yexpfil	yypfil	s(yypfil)	erfil	erfil
1	3.0000E+00	3.0544E+00	1.1339E-01	-5.4412E-02	-1.8137E+00

Napověda-F1 Rádek: 71 - 93 Celkem: 206 Délka: 11484



Stupeň polynomu 4

U Ý S L E D K Y

(3) ODHADY PARAMETRŮ A TESTY ÚZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylnka	Test $H_0: \beta_{ij} = 0$ vs. $H_A: \beta_{ij} \neq 0$	t-kritérium	hypoteza H_0 je	Hlad. úz.
BI 01	-8.1937E+02	1.1784E+03		-6.9530E-01	Akceptována	0.503
BI 11	7.7272E+00	8.9604E+00		8.6238E-01	Akceptována	0.409
BI 21	-2.6617E-02	2.5484E-02		-1.0445E+00	Akceptována	0.321
BI 31	3.9752E-05	3.2131E-05		1.2372E+00	Akceptována	0.244
BI 41	-2.1650E-08	1.5154E-08		-1.4287E+00	Akceptována	0.184

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Učienásobný korelační koeficient, R	: 9.9969E-01
Koeficient determinace, R^2	: 9.9937E-01
Předikovaný korelační koeficient, R_p^2	: 9.9939E-01
Střední kvadratická chyba predikce, MEP	: 2.2709E-02
Ákaikeho informační kritérium, AIC	: -5.6803E+01

(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod	Meřená hodnota	Předikovaná hodnota	Směrodatná odchylnka	Klasické reziduum	Relativní reziduum
i	yexpfil	yypfil	s(yypfil)	erfil	erfil
1	3.0000E+00	2.9801E+00	1.2021E-01	1.9894E-02	6.6314E-01

Napověda-F1 Rádek: 77 - 99 Celkem: 212 Délka: 11906