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

Ověření předpokladů MNČ

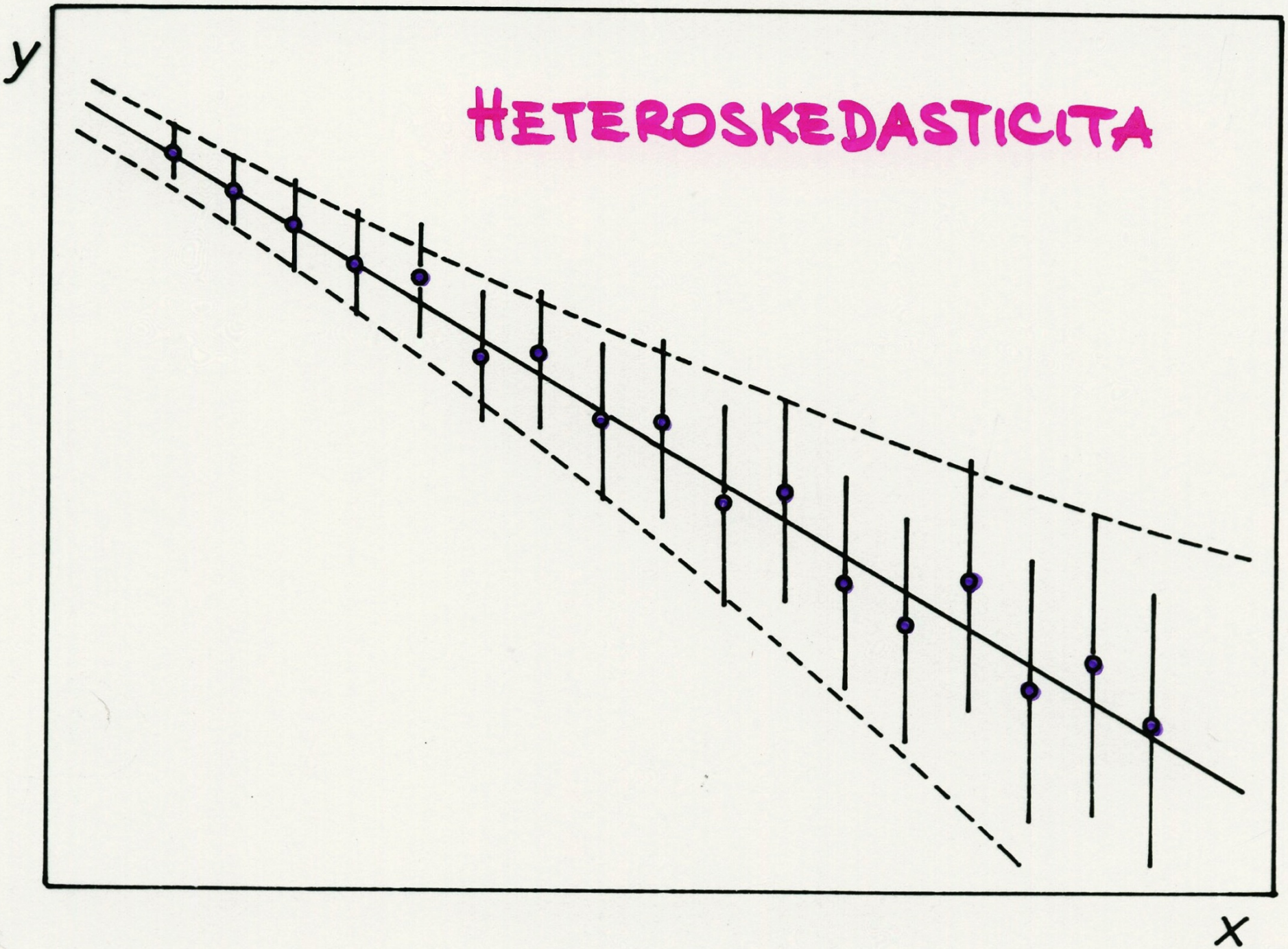
1. Hetero/homoskedasticita

(nekonstantnost rozptylu)

Rozptyl veličiny y_i v itém bodě je popsán

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

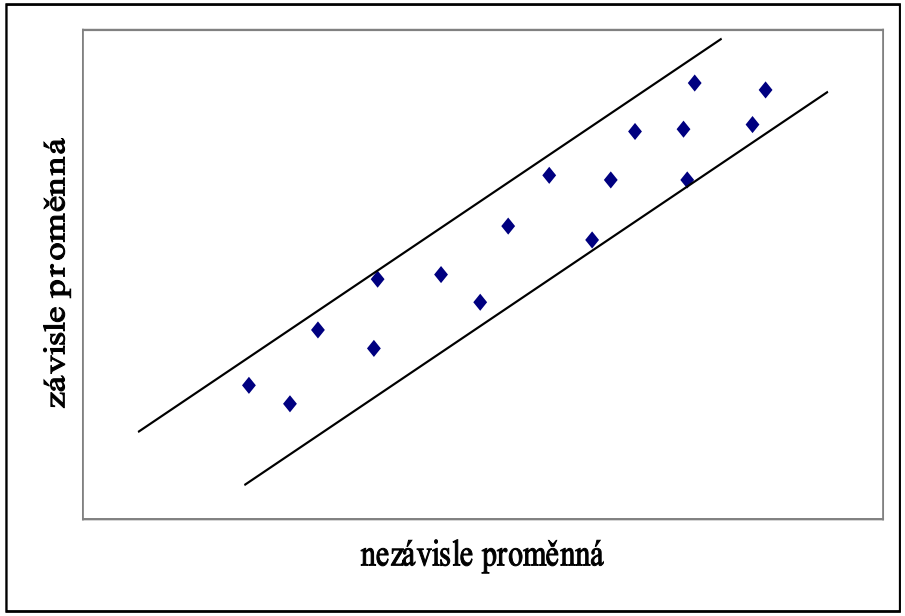
kde $\mathbf{x}_i$ je i-tý řádek matice $\mathbf{X}$.


HETEROSKEDASTICITA
y
x

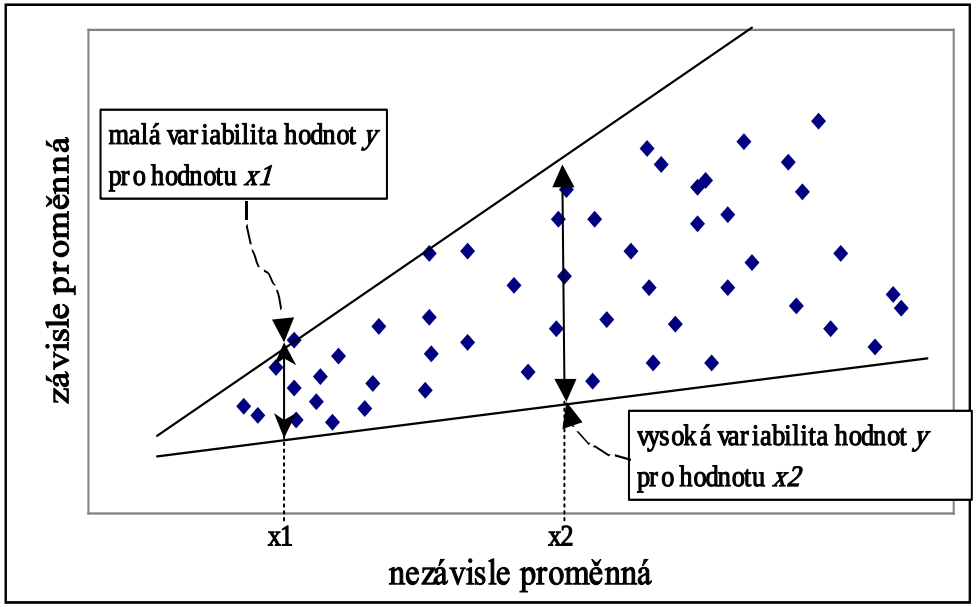
HOMOSKEDASTICITA *vs.* ETEROSKEDASTICITA

Grafické vysvětlení principu

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



Homoskedasticita



Heteroskedasticita

Testování homoskedasticity testem trendu reziduí

$$D = \sum_{i=1}^{n} \left[R\left(\left|\hat{e}_i\right|\right) - i \right]^2$$

$$\rho_s = 1 - \frac{6}{n^3 - n} \cdot D$$

Testujeme významnost Spearmanova korelačního koeficientu ρ_s

$$t_R = \frac{\rho_s \cdot \sqrt{n-2}}{\sqrt{1-\rho_s^{\,2}}}$$

Cookův - Weisbergův test homoskedasticity

Vycházíme z předpokladu, že rozptyl naměřené hodnoty y_i je určitou funkcí proměnné $\boldsymbol{x_i}\,\boldsymbol{\beta}$ (např. exponenciální funkcí).

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

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

1. vzorová úloha
na výstavbu
lineárního regresního modelu

P651

Test heteroskedasticity

Příklad P6.51

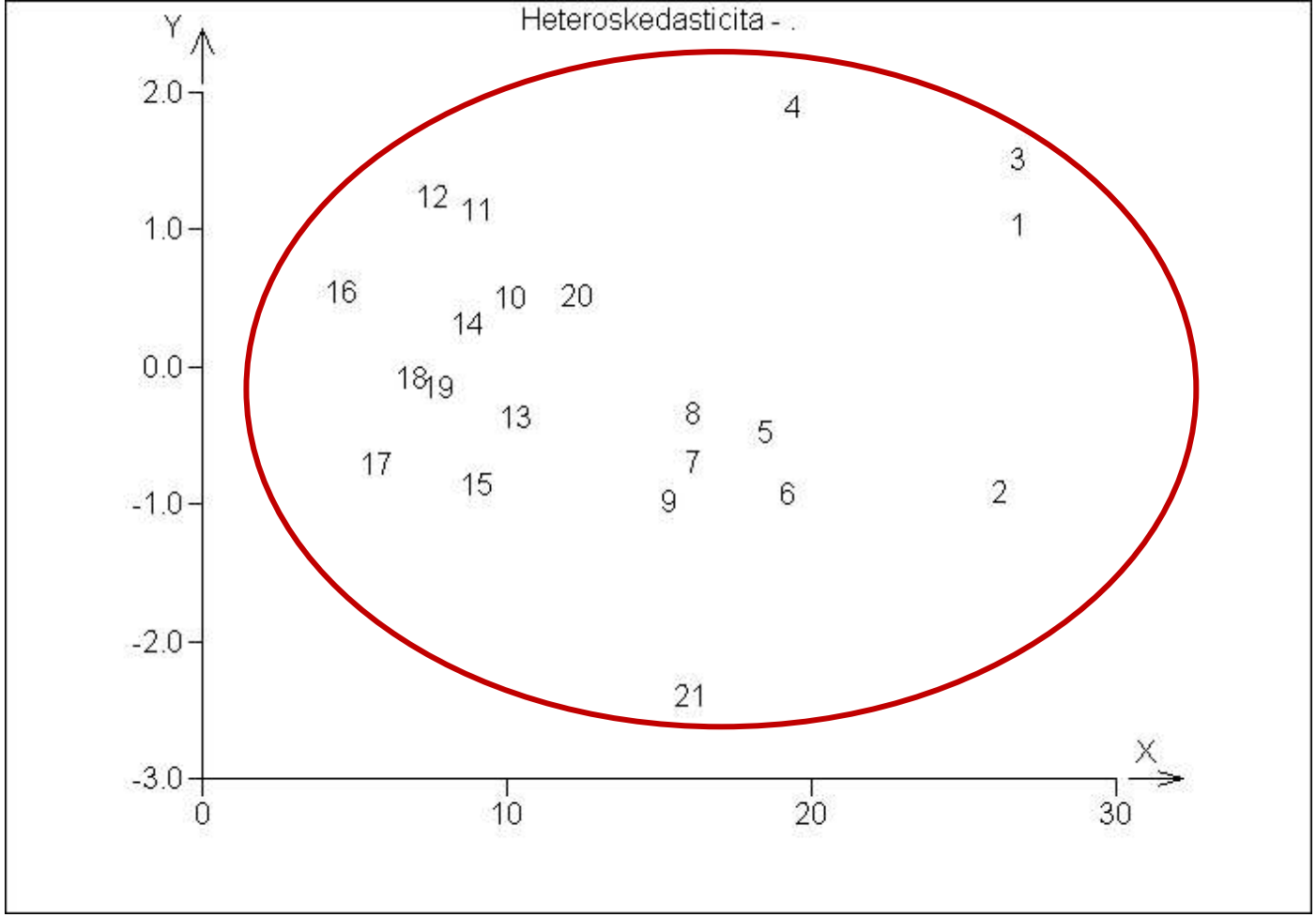
Procentní ztráta amoniaku v aparatuře při výrobě kyseliny dusičné

Byla sledována procentní ztráta amoniaku při oxidační reakci amoniaku na kyselinu dusičnou po dobu 21 dní. Určete regresní model mezi relativní ztrátou amoniaku (y) a třemi vysvětlujícími nezávisle proměnnými: průtokem vzduchu (x_1), teplotou chladící vody (x_2) a koncentrací kyseliny dusičné (x_3). Vyšetřete regresní triplet při užití metody nejmenších čtverců (MNČ) a L_1-aproximace.

Model: $y = \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4$

Data: n = 21 points

Day	y (NH_3)	x_1 (vzd.)	x_2 (temperat.)	x_3 (HNO_3)
1	42	80	27	89
2	37	80	27	88
3	37	75	25	90
4	28	62	24	87
5	18	62	22	87
6	18	62	23	87
7	19	62	24	93
8	20	62	24	93
9	15	58	23	87
10	14	58	18	80
11	14	58	18	89
12	13	58	17	88
13	11	58	18	82
14	12	58	19	93
15	8	58	18	89
16	7	50	18	86
17	8	50	19	72
18	8	50	19	79
19	9	50	20	80
20	15	56	20	82
21	15	70	20	91



Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 1.059618

Kvantil Chi^2(1-alfa,1) : 3.841459

Pravděpodobnost : 0.303302

Závěr : Rezidua vykazují homoskedasticitu.

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

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

V běžných případech je možné jako váhy volit

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

2. Autokorelace

Data časových řad mají chyby ε_i vzájemně korelované.

Nejčastější je případ autokorelace prvního řádu

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

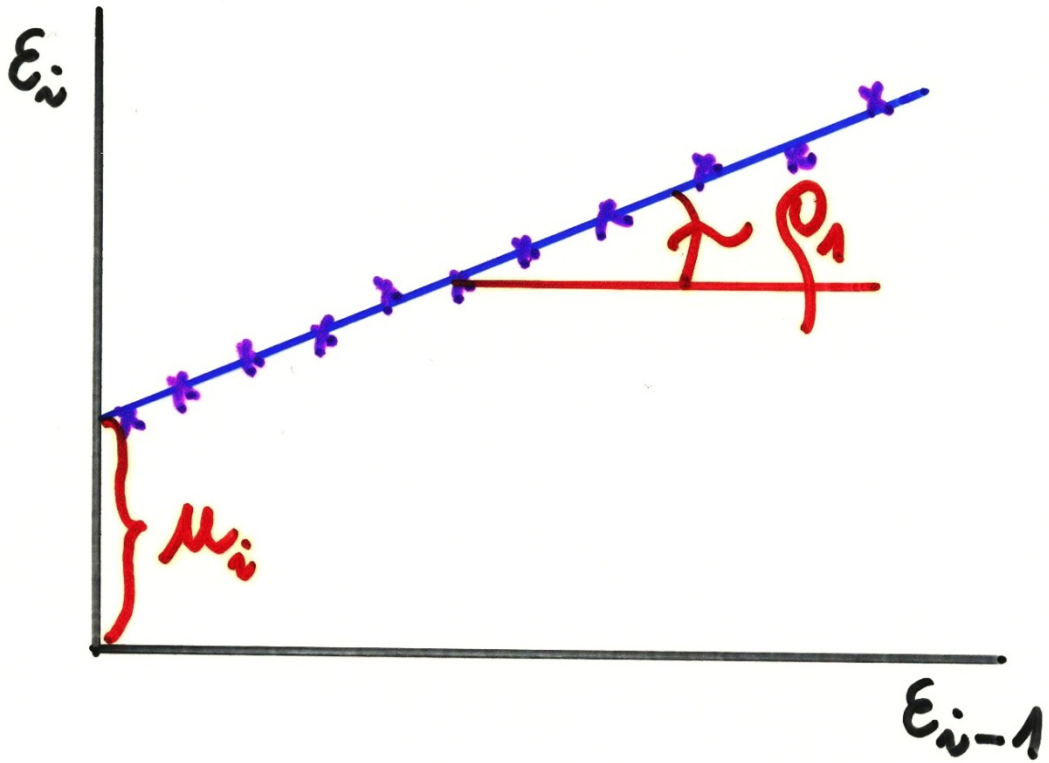
kde $u_i \sim N(0, \sigma^2)$.

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 $W_a = \dfrac{n\,\hat{\rho}_1^2}{1 - \hat{\rho}_1^2} < \chi^2(1)$, H_0 je přijata.

2. vzorová úloha
na výstavbu
lineárního regresního modelu

P638

Test autokorelace

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 odebírání vzorků. Určete graficky významnost autokorelace v datech.

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

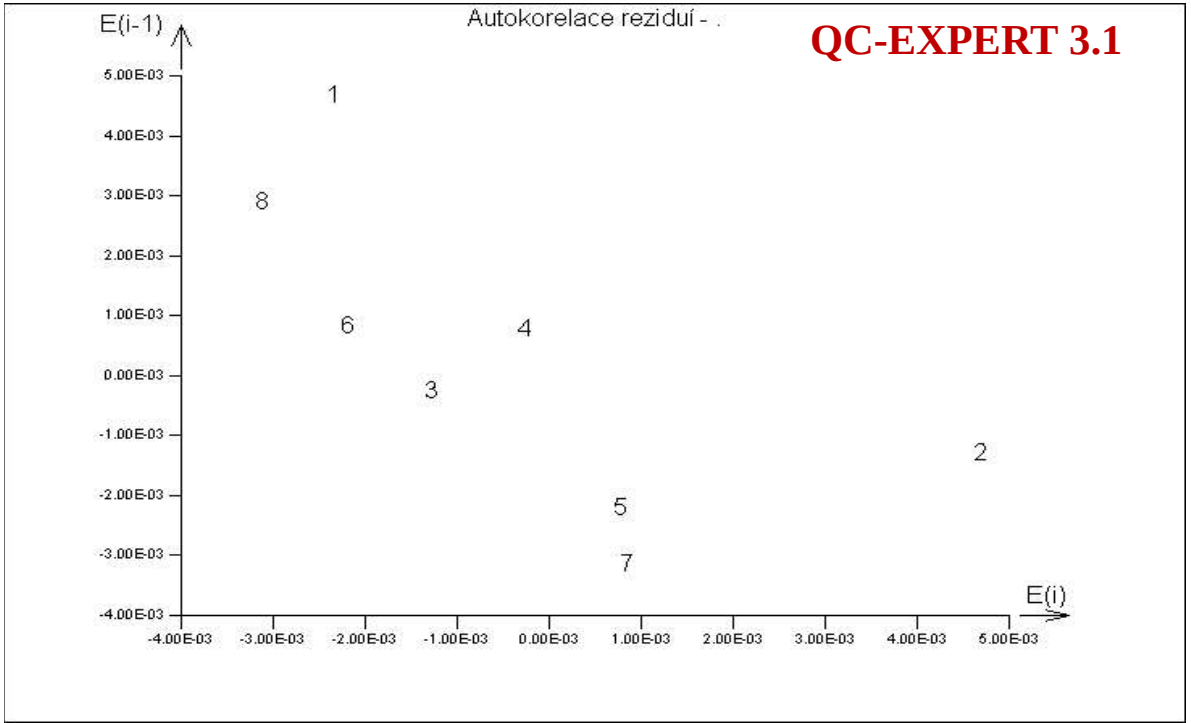
x	*y*
0	1.000
10	0.954
20	0.895
30	0.843
40	0.791
50	0.735
60	0.685
70	0.628
80	0.581

Waldův test autokorelace

Hodnota kritéria WA : 3.454228
Kvantil Chi^2(1-alfa,1) : 3.841459
Pravděpodobnost : 0.06309
Závěr : Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

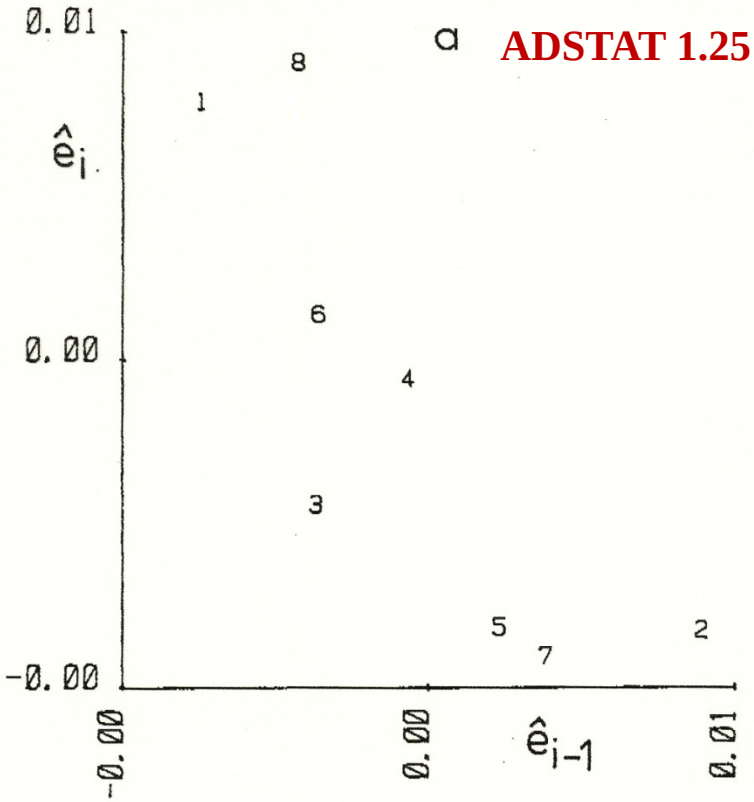
Hodnota kritéria DW : -1.0
Kritické hodnoty DW 0.0 2.0
Závěr : Negativní autokorelace reziduí není prokázána.

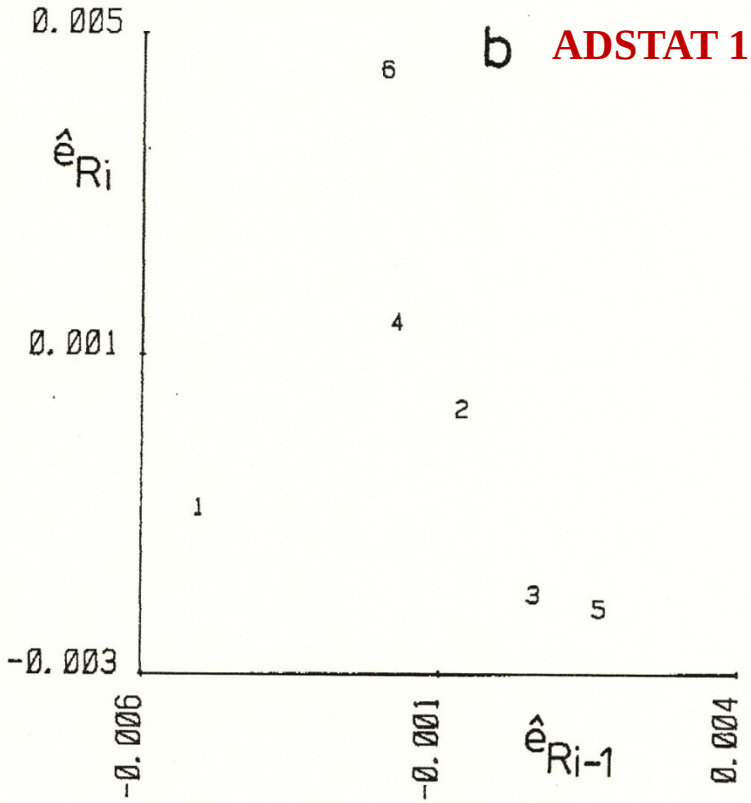


Řešení:

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.

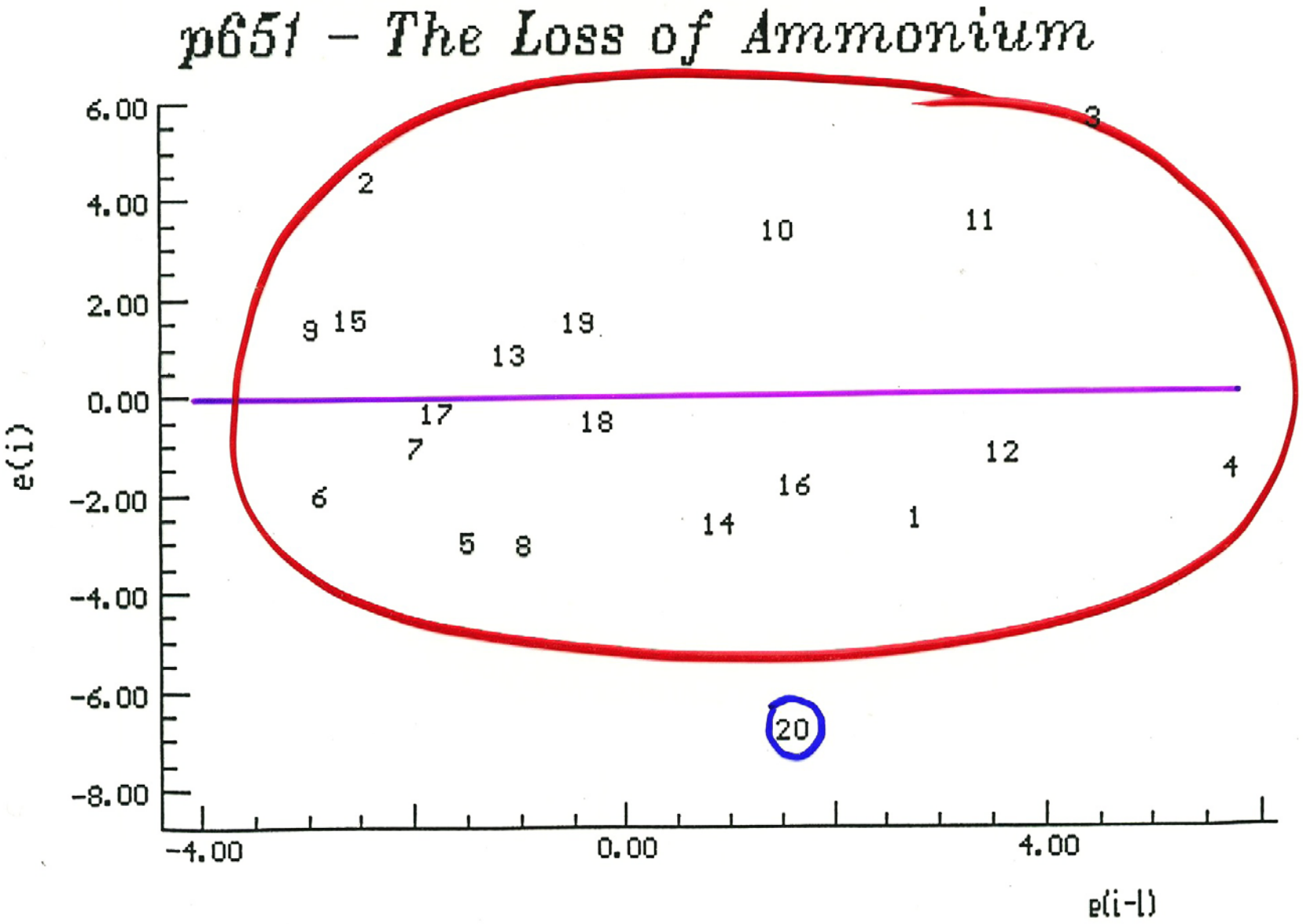
2. Waldovo testační kritérium autokorelace $W_a = 9.424$ je větší než $\chi^2_{0.95}(2) = 5.99$, což indikuje autokorelaci reziduí.





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

Příklad P651 Autocorrelation plot **ADSTAT 1.25**



Normalita chyb

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

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

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

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 $L(\hat{e}) > \chi^2_{1-\alpha}(2) = 5.99$, je H_0 (normalita) zamítnuta.

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

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

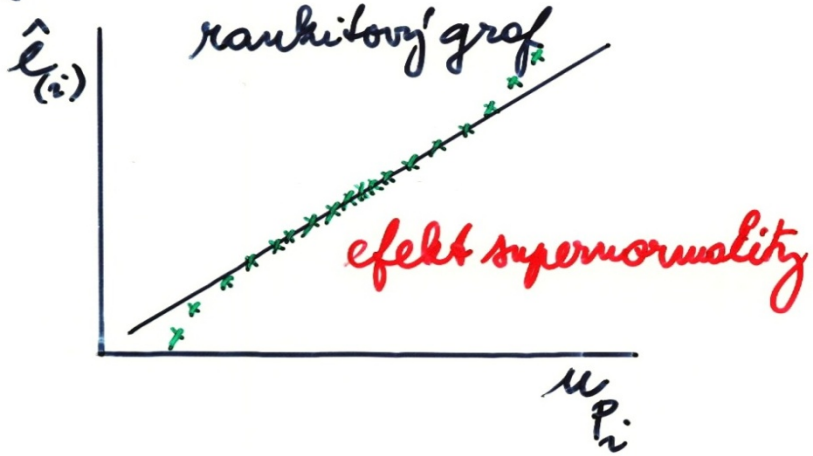
kde

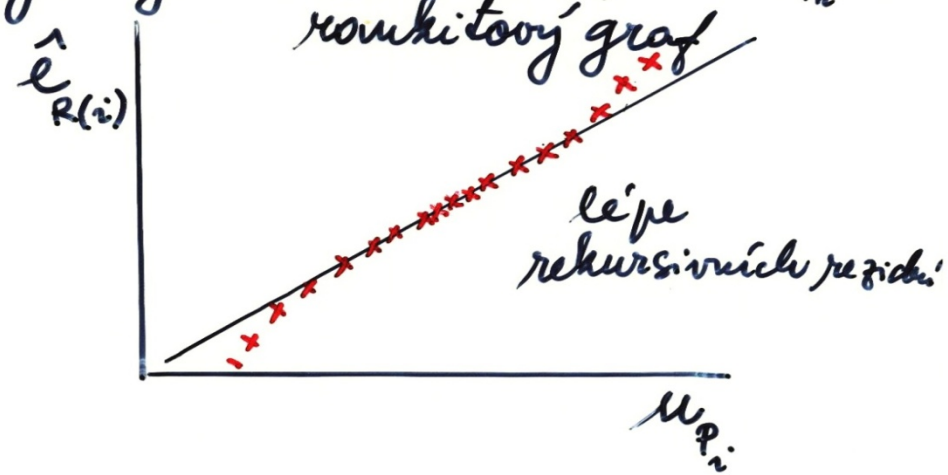
$$\hat{g}_1 = \frac{\hat{u}_3^{\,2}}{\hat{u}_2^{\,3}} \qquad \hat{g}_2 = \frac{\hat{u}_4}{\hat{u}_2^{\,2}}$$

(test není vhodný pro malé výběry)

NORMALITA CHYB

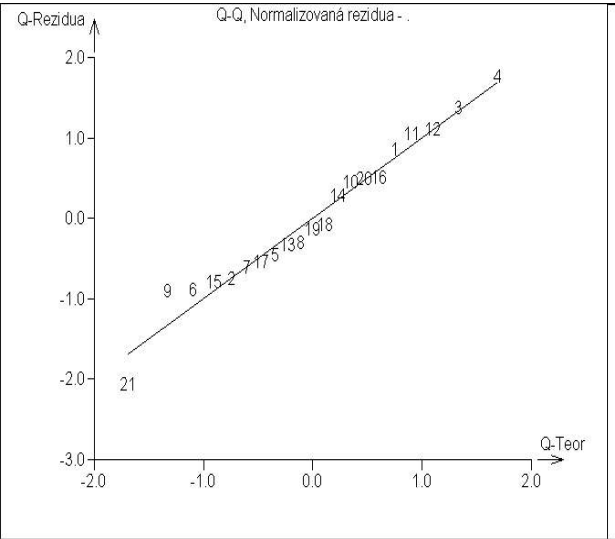
grafická indikace normality chyb: $\hat{e}_{(i)}$ nebo $\hat{e}_{R(i)}$ vs. u_{P_i}

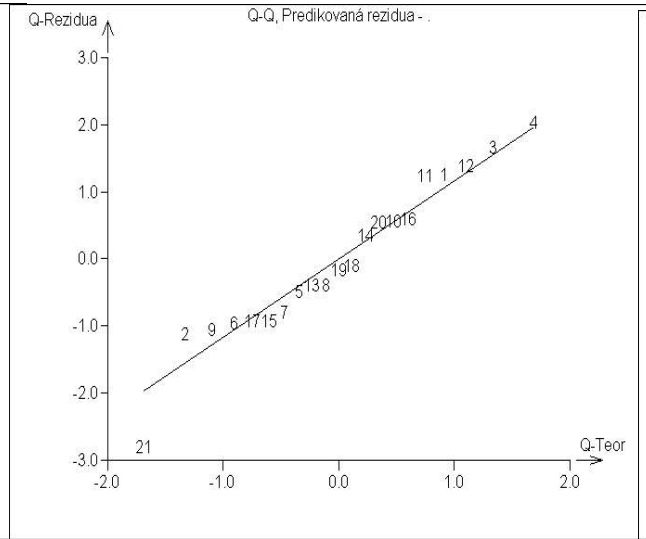


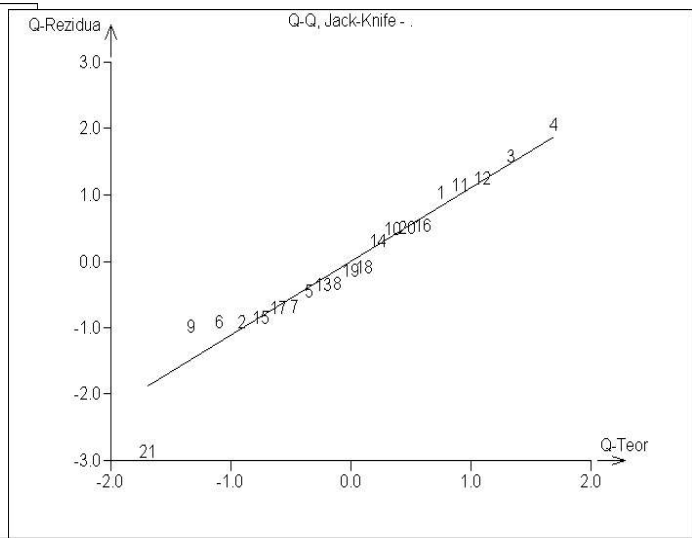


Příklad P6.51

Jarque-Berrův test normality, Hodnota kritéria JB: 0.080308
Kvantil Chi^2(1-alfa,2): 5.991465, Pravděpodobnost: 0.960641,
Závěr: Rezidua mají normální rozdělení.







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).

Příklad P638 *Test autokorelace u kinetických dat*

Znaménkový test reziduí

Hodnota kritéria Sg :	2.208853
Kvantil N(1-alfa/2) :	1.959964
Pravděpodobnost :	0.027185

Závěr : V reziduích je trend!

Vyšetření předpokladů u MNČ při testování regresního tripletu

Fisher-Snedecorův test významnosti modelu, Hodnota kritéria F : 59.57416
Kvantil F (1-alfa, m-1, n-m) : 3.196777, Pravděpodobnost : 0.0, Závěr : Model je významný

Scottovo kritérium multikolinearity, Hodnota kritéria SC : 0.626359
Závěr : Model vykazuje multikolinearitu!

Cook-Weisbergův test heteroskedasticity, Hodnota kritéria CW : 1.059618
Kvantil Chi^2(1-alfa,1): 3.841459, Pravděpodobnost:0.303302, Závěr: Rezidua ukazují homoskedasticitu.

Jarque-Berrův test normality, Hodnota kritéria JB : 0.080308
Kvantil Chi^2(1-alfa,2): 5.991465, Pravděpodobnost: 0.960641, Závěr Rezidua mají normální rozdělení.

Waldův test autokorelace, Hodnota kritéria WA: 0.001722
Kvantil Chi^2(1-alfa,1): 3.841459, Pravděpodobnost: 0.966903, Závěr: Autokorelace je nevýznamná.

Durbin-Watsonův test autokorelace, Hodnota kritéria DW: -1.0,
Kritické hodnoty DW 0.93 1.81, Závěr : Pozitivní autokorelace reziduí není prokázána.

Znaménkový test reziduí: Hodnota kritéria Sg: 0.555662,
Kvantil N(1-alfa/2): 1.959964, Pravděpodobnost : 0.578442, Závěr: V reziduích není trend.

3. vzorová úloha

na hledání metody výstavby lineárního regresního modelu

M619

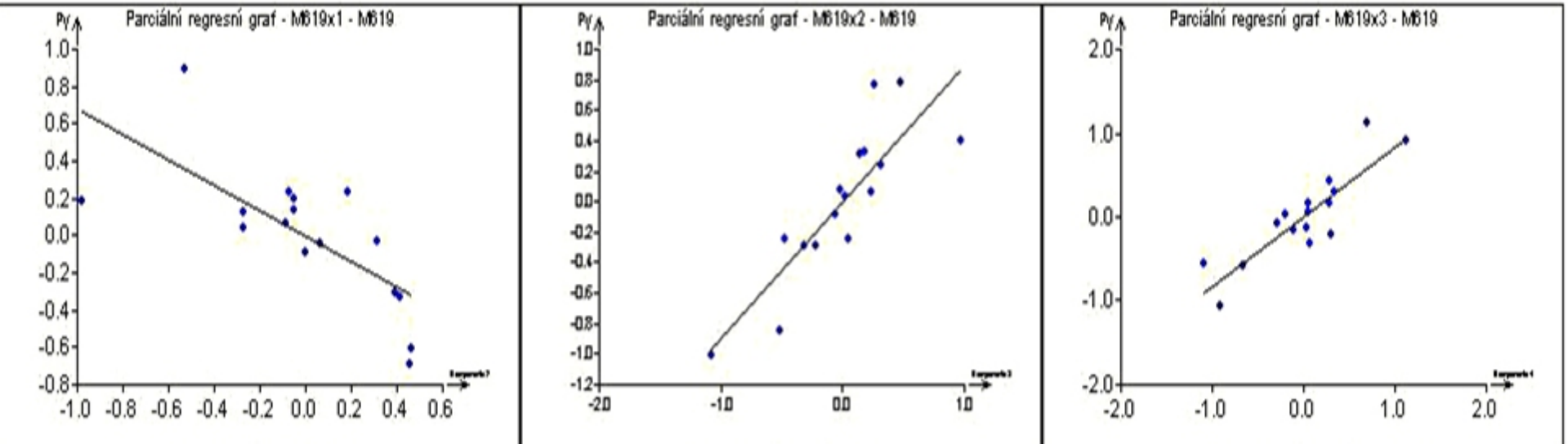
Úloha M619. ***Vliv tří parametrů na obsah kadmia v potravinářské pšenici***

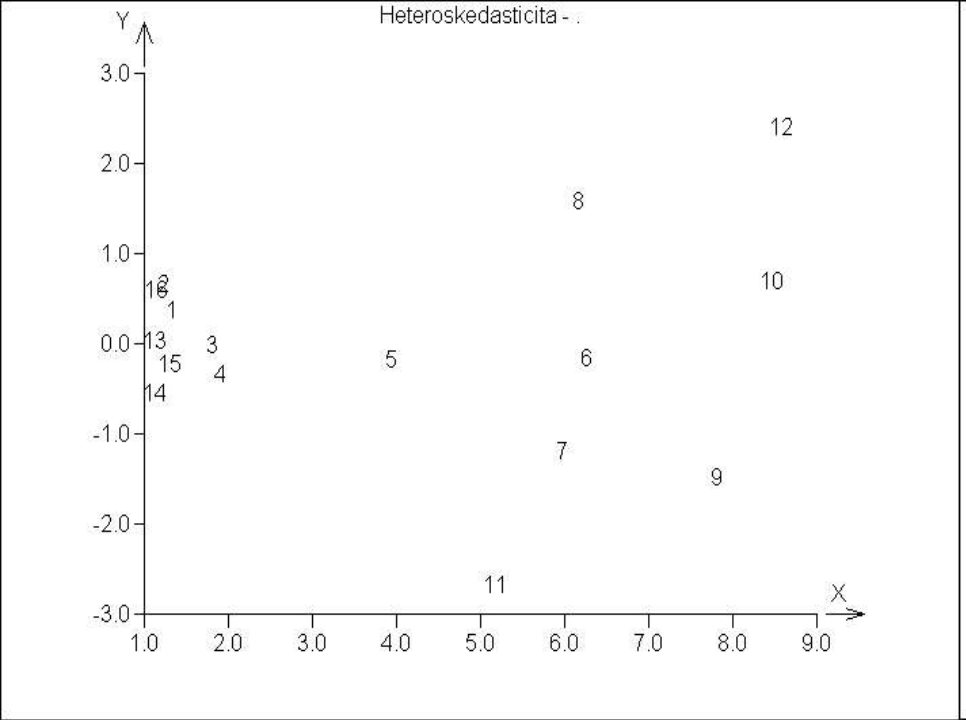
Obsah kadmia v zrnu y [mg/l] v závislosti na obsahu kadmia v otrubách x_1 [mg/l], ve stonku s listy x_2 [mg/l] a v kořenovém systému x_3 [mg/l]. Vyšetřete regresní triplet (data, model, metoda) a nalezněte linearní regresní model.

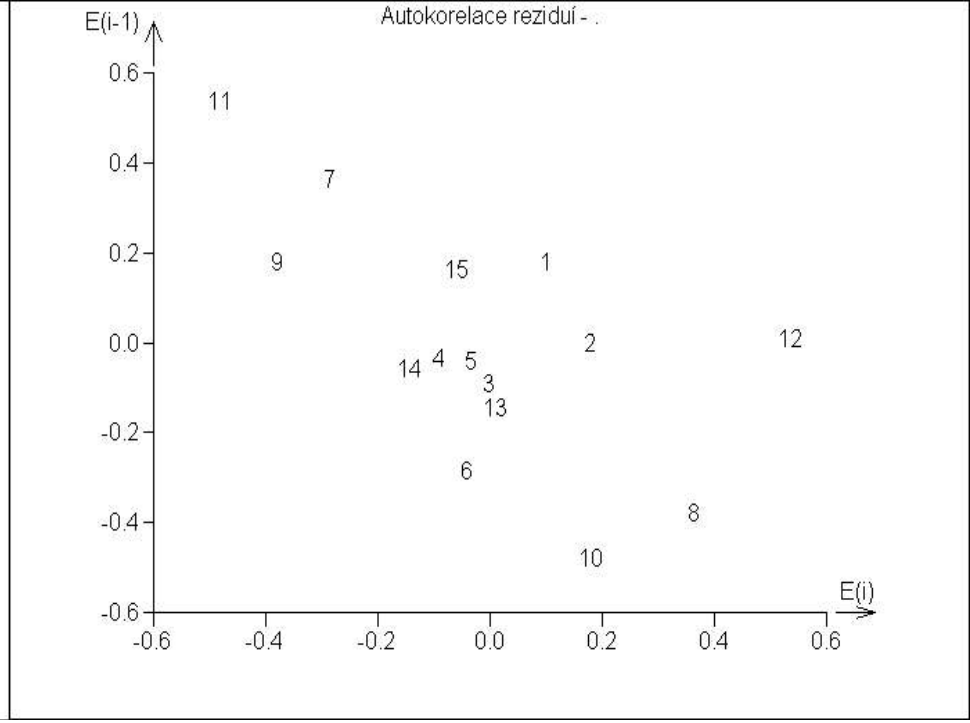
Výstavba lineárního regresního modelu:

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

Parciální regresní grafy tří nezávisle proměnných







Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 5.373857

Kvantil Chi^2(1-alfa,1) : 3.841459

Pravděpodobnost : 0.020441

Závěr: Rezidua vykazují heteroskedasticitu!

Waldův test autokorelace

Hodnota kritéria WA : 5.003828

Kvantil Chi^2(1-alfa,1) : 3.841459

Pravděpodobnost : 0.025291

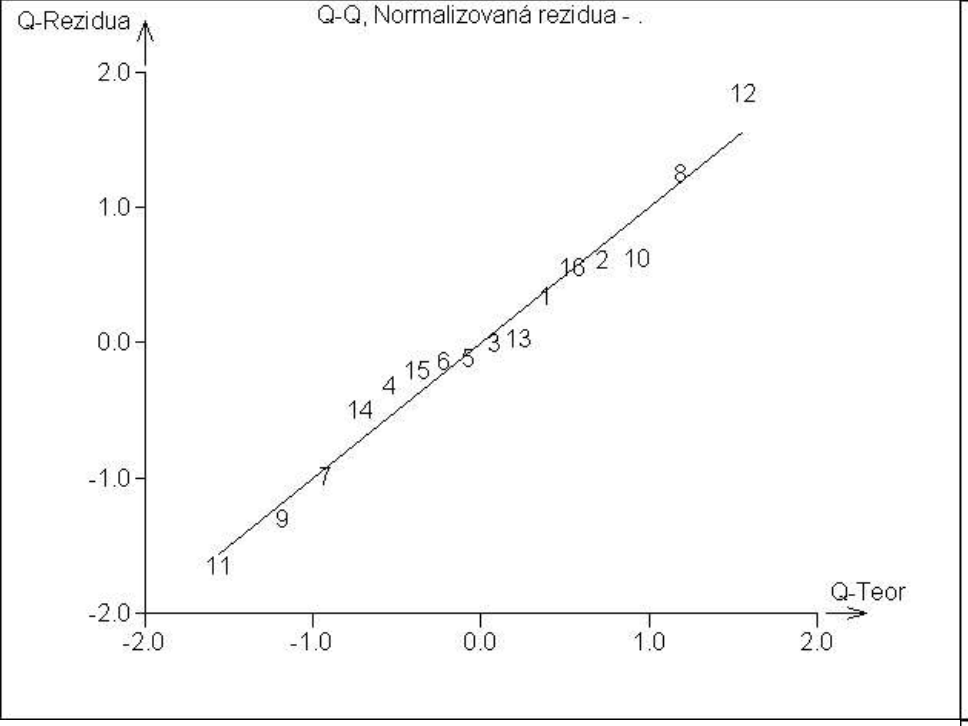
Závěr: Autokorelace je významná

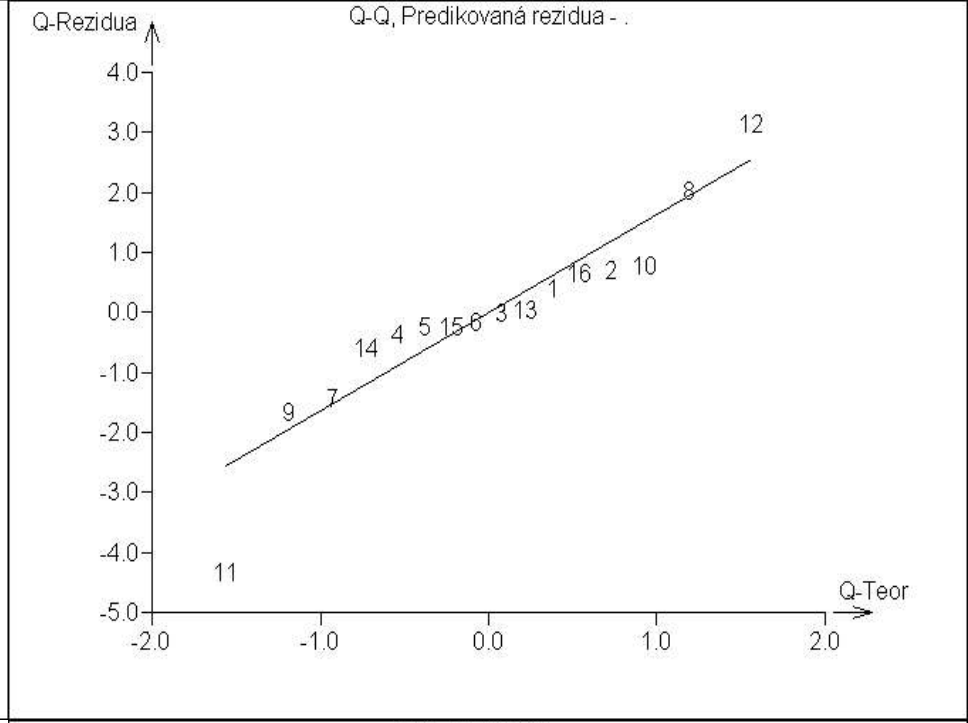
Durbin-Watsonův test autokorelace

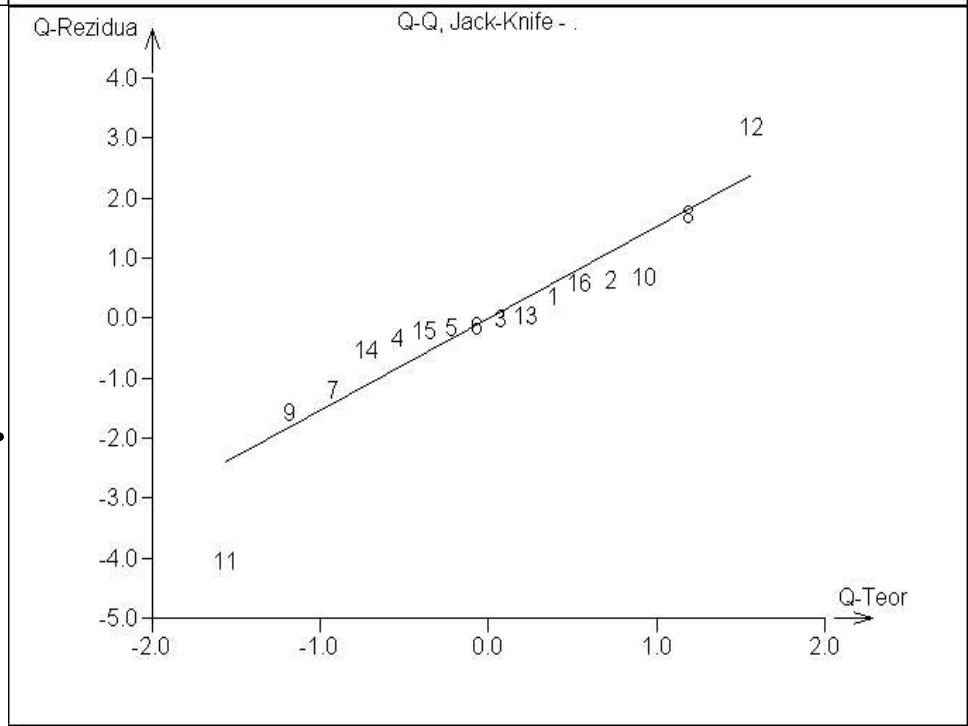
Hodnota kritéria DW : -1.0

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-alfa,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-alfa,1): 3.841459, Pravděpodobnost: 0.020441,
Závěr: Rezidua vykazují heteroskedasticitu!

Jarque-Berrův test normality, Hodnota kritéria JB: 0.033085, Kvantil Chi^2(1-alfa,2) : 5.991465,
Pravděpodobnost: 0.983593, **Závěr: Rezidua mají normální rozdělení.**

Waldův test autokorelace, Hodnota kritéria WA: 5.003828,
Kvantil Chi^2(1-alfa,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-alfa/2): 1.959964,
Pravděpodobnost: 0.742174, **Závěr: V reziduích není trend.**

Úlohy na výstavbu lineárního regresního modelu

Kritika metody

Software QC-EXPERT 3.1, ADSTAT 1.25

Úloha M6.04 ***Vliv čtyř faktorů stanovení hořčíku na absorbanci AAS – předpoklady MNČ***

Zadání: Při stanovení hořčíku ve vodě metodou AAS byla sledována závislost absorbance y při $\lambda = 202.5$ nm na výšce hořáku x_1, nasávaném množství vzorku x_2, průtoku vzduchu x_3 a průtoku acetylenu x_4. Předpokládejme platnost vícerozměrného lineárního regresního modelu ve tvaru $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$.

Úkoly: (1) Vyšetřete regresní triplet.

(2) Pomocí parciálních regresních a u parciálních reziduálních grafů diskutujte významnost jednotlivých regresních parametrů modelu.

(3) Proveďte Cookův-Weisbergův test heteroskedasticitu v datech.

(4) Proveďte autokorelační test.

(6) Proveďte Jarque-Berrův test normality v datech.

Data: Výška hořáku x_1, množství vzorku x_2, průtok vzduchu x_3, průtok acetylenu x_4, měřená absorbance y:

x_1	x_2	x_3	x_4	y
10.4	4.4	12.3	37	0.25
...	...	...	...	...
9.6	3.9	11.2	31.9	0.275

Úloha M6.21 ***Vliv pěti parametrů na obsah manganu ve vodárenské vodě – předpoklady MNČ***

Zadání: Nalezněte závislost mezi obsahem rozpuštěného manganu y a pěti parametry: teplotou vody x_1, obsahem kyslíku x_2, pH x_3, průtokem vody vypouštěné x_4 a časem x_5.

Úkoly:

(1) Jsou všechny regresní parametry statisticky významné?

(2) Vyšetřete regresní triplet a odhalte vlivné body.

(3) Vyšetření porušení předpokladů MNČ.

Data: Teplota vody x_1 [℃], obsah kyslíku x_2 [mg/l], pH x_3, průtok vody vypouštěné dnovými výpusťmi x_4 [%] a časem x_5 [týden], obsah rozpuštěného manganu y [μg/l]:

x_1	x_2	x_3	x_4	x_5	y
3.4	9.9	7.7	16.26	17	10
...	...	...	...		...
4.9	5	6.9	6.55	49	410

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 [μmol/l] :

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

Úloha M6.45 ***Vliv čtyř parametrů na koncentraci TSH v kontrolním séru – předpoklady MNČ***

Zadání: Souprava na stanovení TSH je dlouhodobě sledována i pomocí interních kontrolních sér. Data obsahují kromě koncentrace TSH v séru také šarže složek použitých v jednotlivých analýzách a datum provedení analýzy.

Úkoly:

(1) Vysvětlete všech sedm předpokladů MNČ a regresní triplet.

(2)Jak se řeší úloha v případě porušení předpokladů MNČ, a to především při heteroskedasticitě, autokorelaci a nenormalitě chyb?

Data: Šarže radioindikátoru x_1, šarže potahovaných zkumavek x_2, šarže kalibrátorů x_3, datum analýzy (formát ymmdd) x_4, koncentrace TSH v séru (mIU/L) y: :

x_1	x_2	x_3	x_4	y
313	1142	1057	10904	0.72
...	...	...	...	...
330	1151	1060	2050	0.7

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:

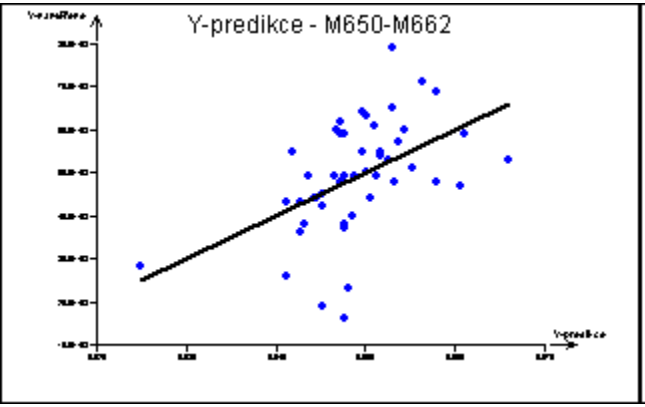
(1) Vyšetřením regresního tripletu nalezněte nejlepší model.

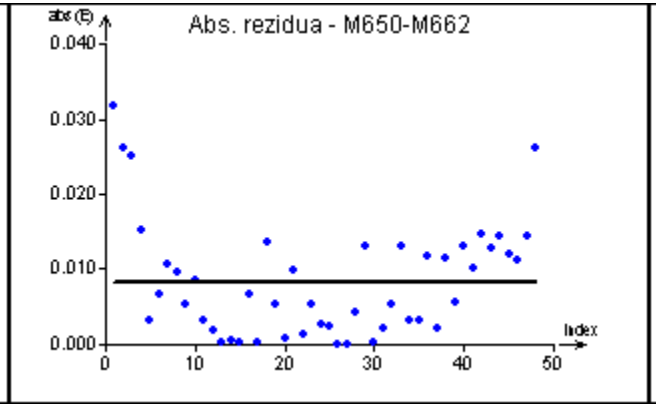
(2) Využijte regresní diagnostiku a parciální regresních a parciální reziduální grafy ke statistické významnosti jednotlivých parametrů.

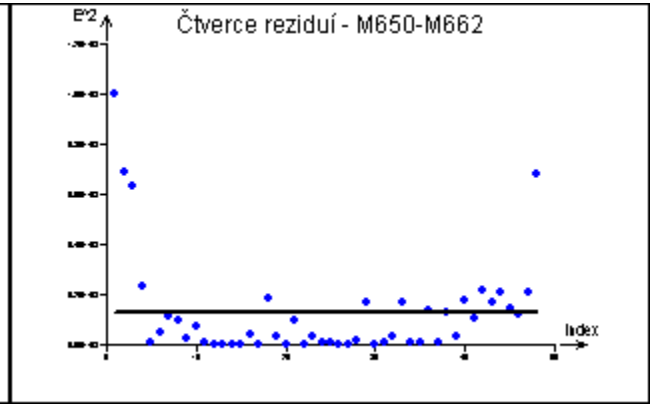
Data: Obsah glukosy x_1 [mmol.L^{-1}], celkový obsah proteinů x_2 [g.L^{-1}], obsah albuminů x_3 [g.L^{-1}], obsah globulinů x_4 [g.L^{-1}], přírůstek váhy v kg za den y [kg]:

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

KRITIKA DAT: statistická analýza reziduí a predikční schopnost modelu







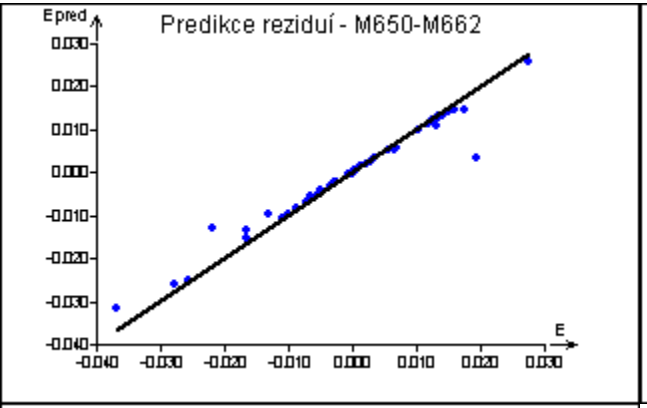
Statistické charakteristiky regrese

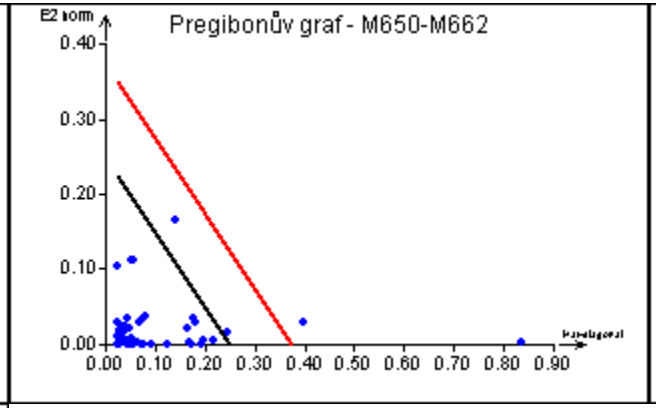
Vícenásobný korelační koeficient R :	0,496798809
Koeficient determinace R^2 :	0,2468090566
Predikovaný korelační koeficient Rp :	0,000469498496
Střední kvdratická chyba predikce MEP :	0,0001638638304
Akaikeho informační kritérium :	-420,9442483

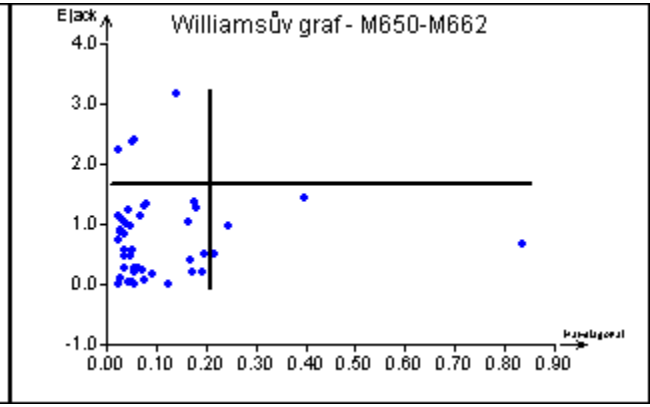
Statistická analýza reziduí:

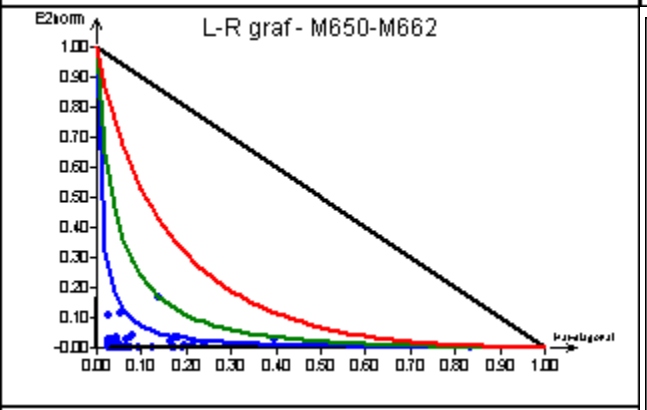
Reziduální součet čtverců :	0,006055404121
Průměr absolutních reziduí :	0,008345371079
Reziduální směr. odchylka :	0,01186690152
Reziduální rozptyl :	0,0001408233517
Šikmost reziduí :	0,2955113844
Špičatost reziduí :	3,654372016

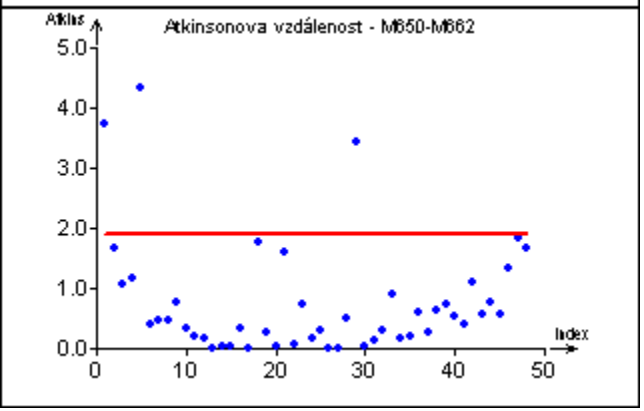
KRITIKA DAT: Analýza vlivných bodů (outlierů a extrémů)

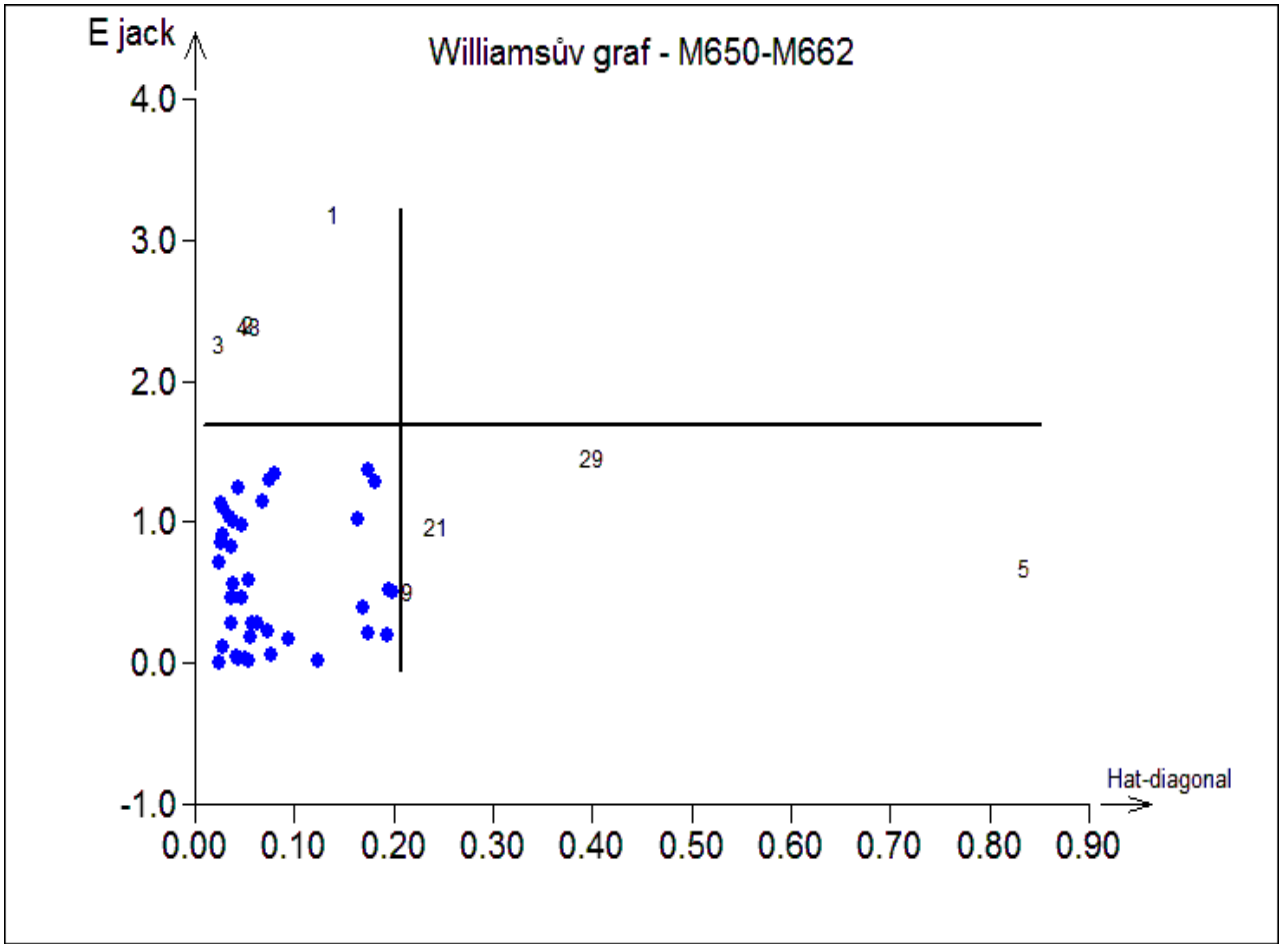




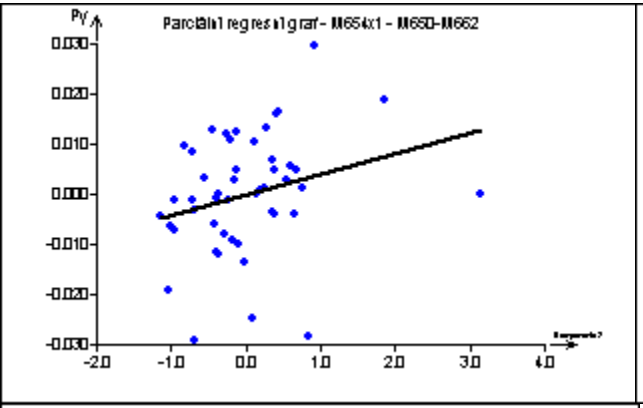


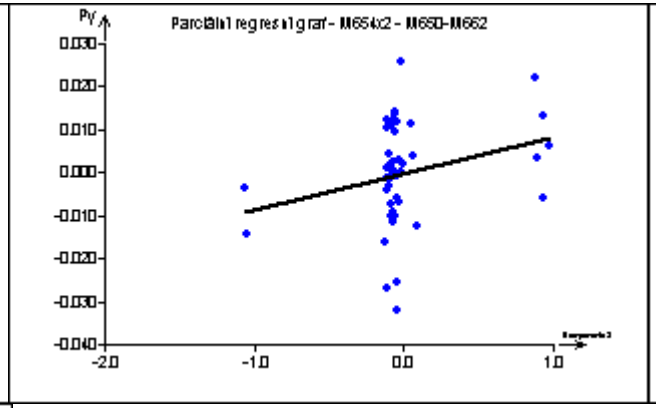


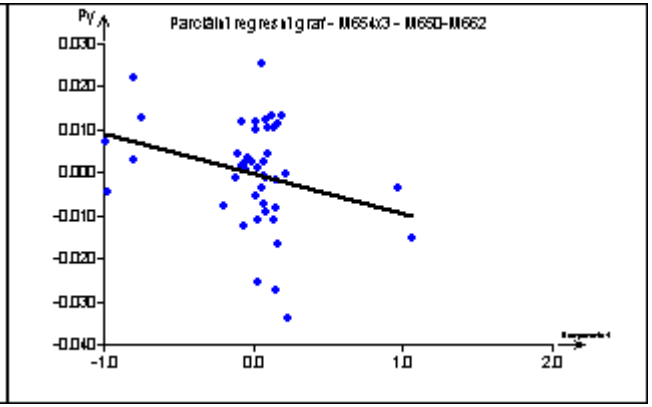


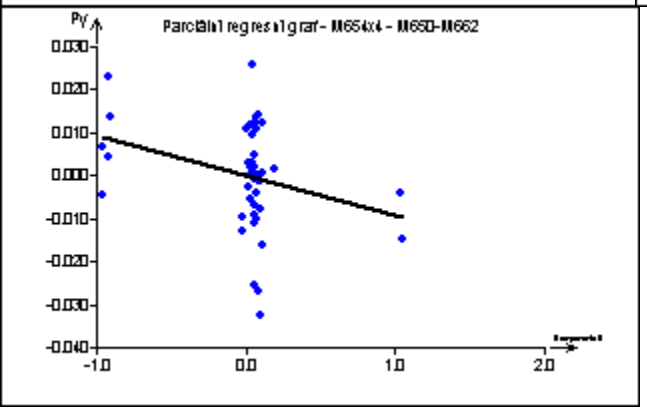


KRITIKA MODELU: Parciální regresní grafy ověří statistickou významnost jednotlivých parametrů dle směrnice přímky.





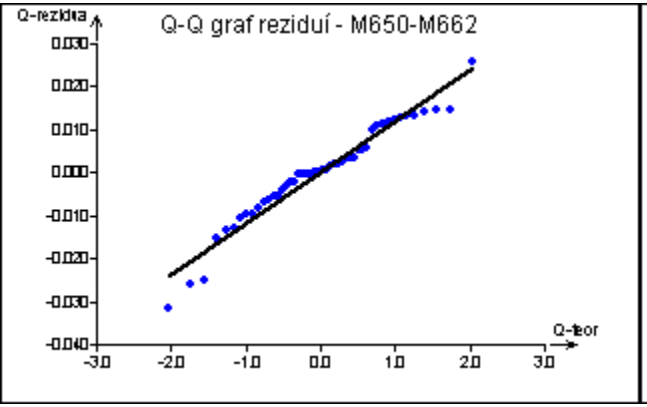


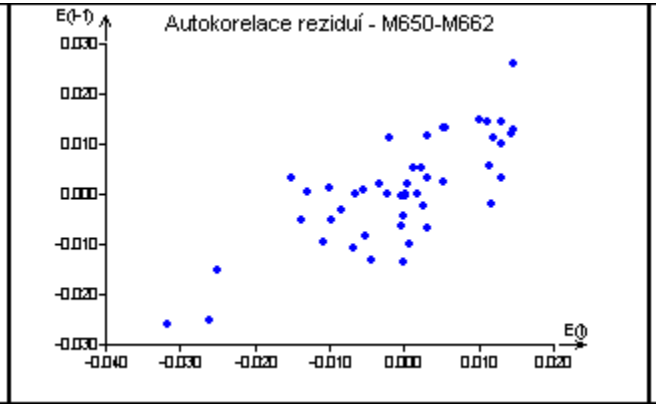


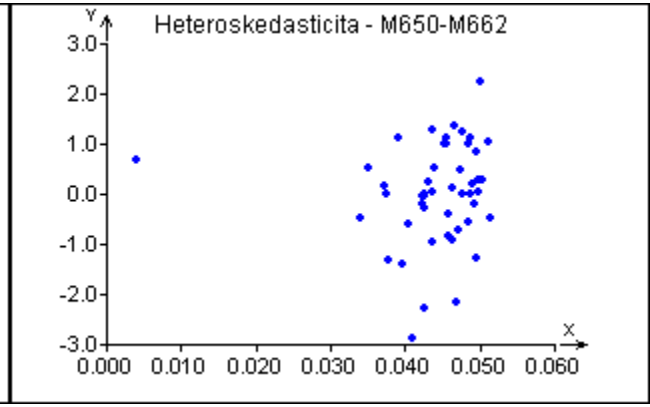
Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	0,02111964025	0,04420587936	Nevýznamný	0,6352434054	-0,068030011	0,1102692923
M654x1	0,004100194902	0,00226021707	Nevýznamný	0,07664830618	-0,000457967	0,0086583570
M654x2	0,008472057513	0,00457785591	Nevýznamný	0,07109597378	-0,0007600687	0,0177041838
M654x3	-0,00930086559	0,00471395446	Nevýznamný	0,05494014486	-0,0188074607	0,0002057296
M654x4	-0,00926954516	0,004567428878	Významný	0,04862658629	-0,0184806436	-5,84469E-05

KRITIKA METODY: Grafická analýza







Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	3,522609216
Kvantil F (1-alfa, m-1, n-m) :	2,588836146
Pravděpodobnost :	0,01423116878
Závěr :	Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC :	0,08158861571
Závěr :	Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW :	0,0819917845
Kvantil Chi^2(1-alfa,1) :	3,841458829
Pravděpodobnost :	0,7746161346
Závěr :	Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB :	3,220496546
Kvantil Chi^2(1-alfa,2) :	5,991464547
Pravděpodobnost :	0,1998379936
Závěr :	Rezidua mají normální rozdělení..

Waldův test autokorelace

Hodnota kritéria WA :	26,98531573
Kvantil Chi^2(1-alfa,1) :	3,841458829
Pravděpodobnost :	2,050069886E-007
Závěr :	Autokorelace je významná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW :	-1	
Kritické hodnoty DW	1,29	1,78
Závěr :	Rezidua jsou pozitivně autokorelována!	

Znaménkový test reziduí

Hodnota kritéria Sg :	3,036772907
Kvantil N(1-alfa/2) :	1,959963999
Pravděpodobnost :	0,002391255353
Závěr :	V reziduích je trend!

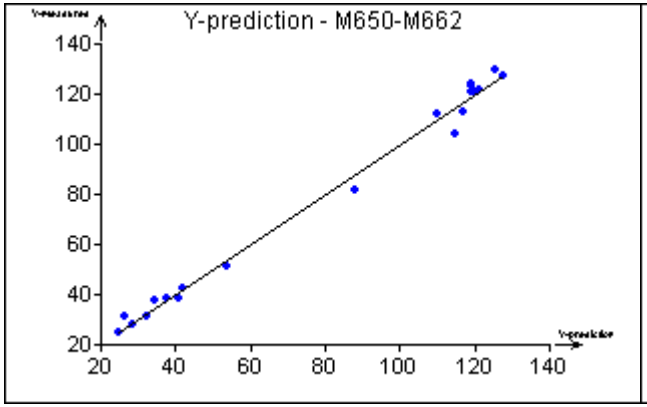
Úloha M6.53 *Vliv teploty vzduchu, koncentrace NO_2 a UV záření na koncentraci ozónu v ovzduší*

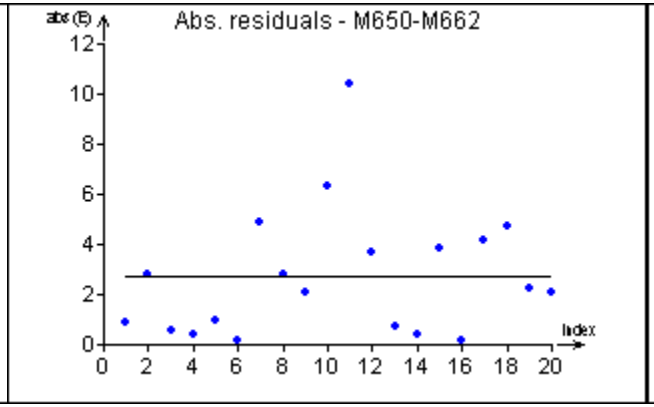
Zadání: Je vyšetřován vliv teploty vzduchu, koncentrace NO_2 a UV záření na koncentraci ozónu v ovzduší. Vyšetřením regresního tripletu proveďte kritiku dat a nalezněte nejlepší regresní model. Využijte regresní diagnostiku a pomocí Studentova *t*-testu vyšetřete statistickou významnost jednotlivých parametrů v modelu.

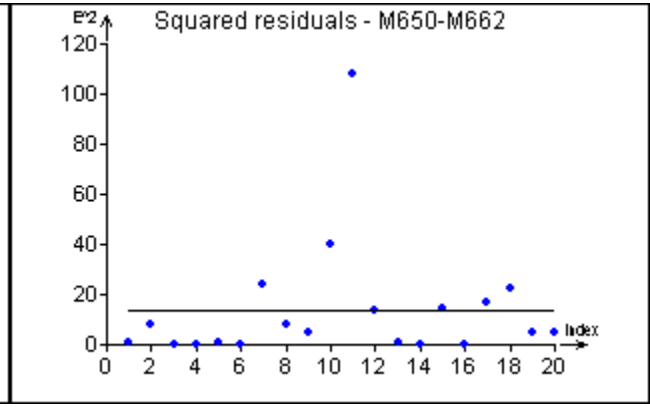
Data: Vliv teploty vzduchu x_1 [**EC**], koncentrace NO_2 [μg/m^3] x_2 a UV záření x_3 [W/m^2] na koncentraci ozónu v ovzduší y [μg/m^3]:

x1	x2	x3	y
18.96	32.42	0.22	31.3
…	…	…	…
28.61	10.97	0.67	112.2

KRITIKA DAT: statistická analýza reziduí a predikční schopnost modelu







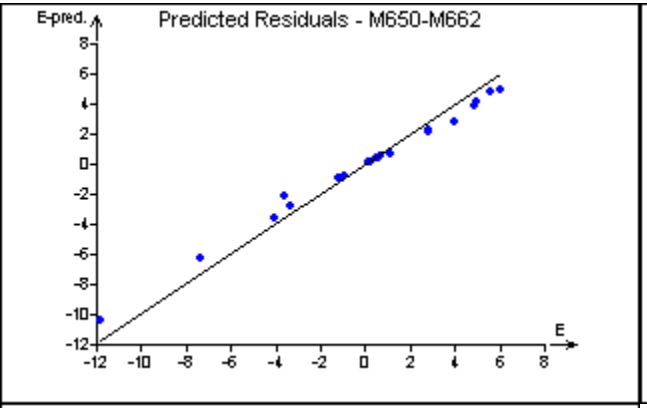
Statistické charakteristiky regrese

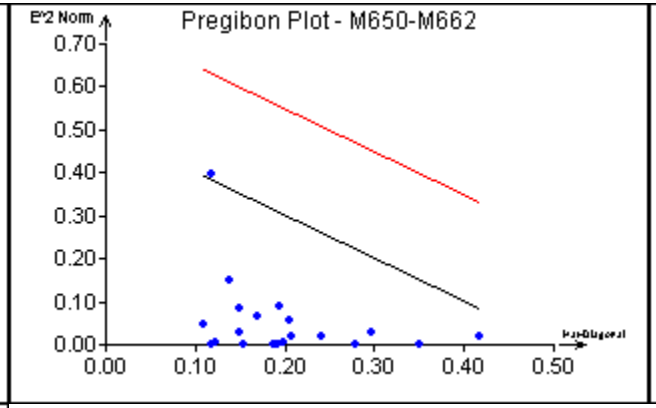
Vícenásobný korelační koeficient R :	0,9960355375
Koeficient determinace R^2 :	0,9920867919
Predikovaný korelační koeficient Rp :	0,9776877371
Střední kvdratická chyba predikce MEP :	19,30425947
Akaikeho informační kritérium :	60,22488937

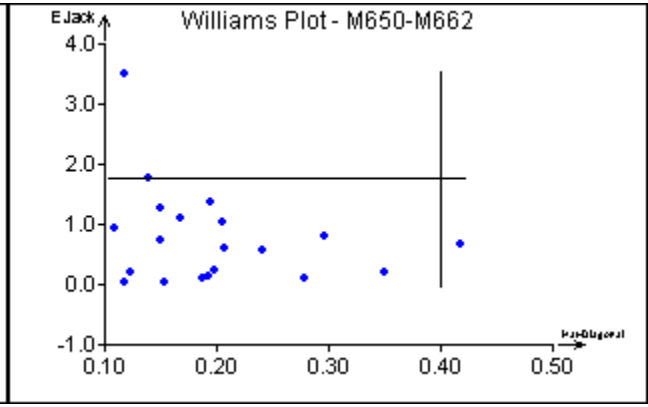
Statistická analýza reziduí:

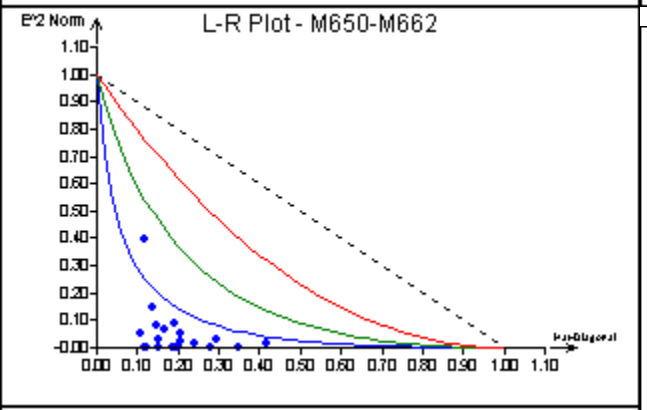
Reziduální součet čtverců :	272,3196995
Průměr absolutních reziduí :	2,708940615
Reziduální směr. odchylka :	4,125527993
Reziduální rozptyl :	17,01998122
Šikmost reziduí :	1,130417017
Špičatost reziduí :	4,123790626

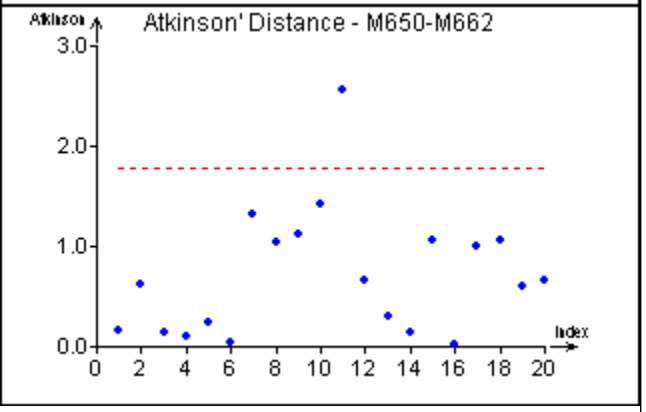
KRITIKA DAT: Analýza vlivných bodů (outlierů a extrémů)

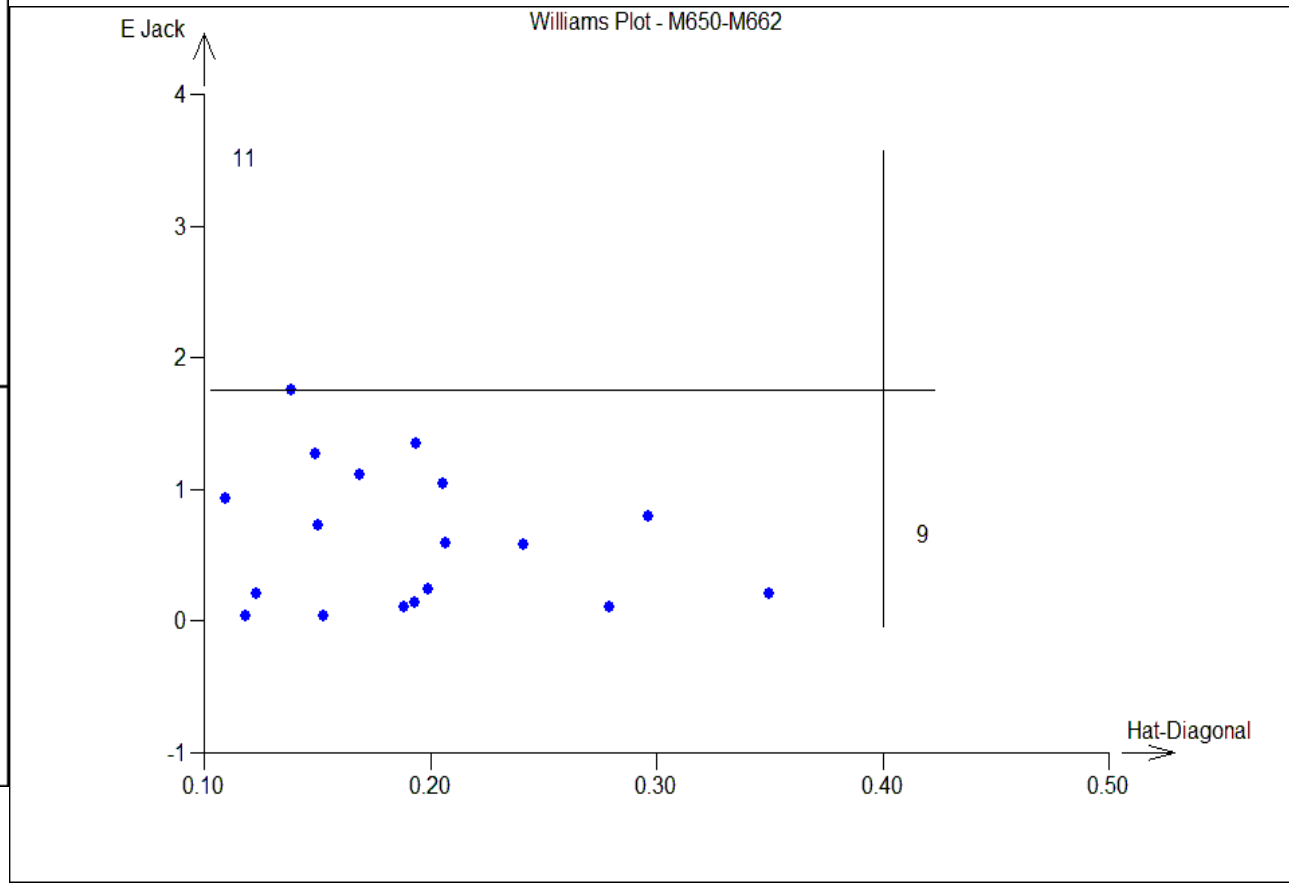




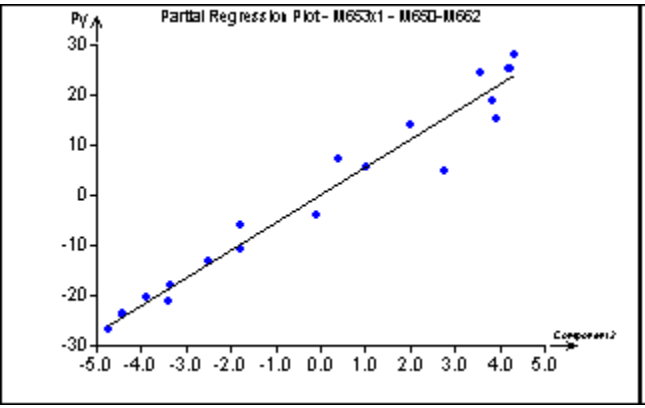


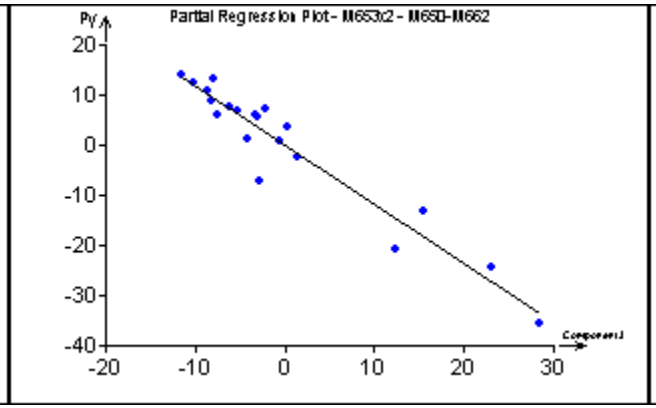


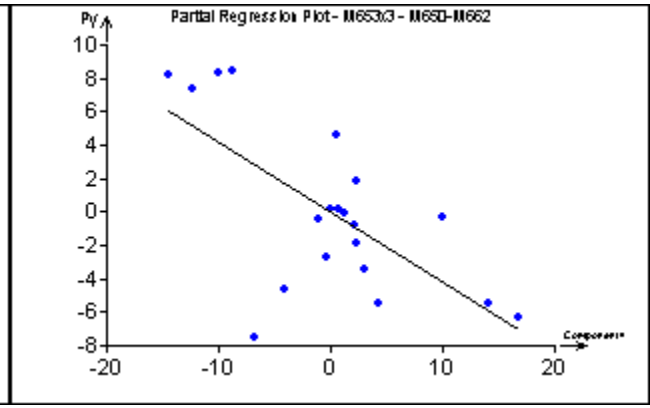




KRITIKA MODELU: Parciální regresní grafy ověří statistickou významnost jednotlivých parametrů dle směrnice přímky.







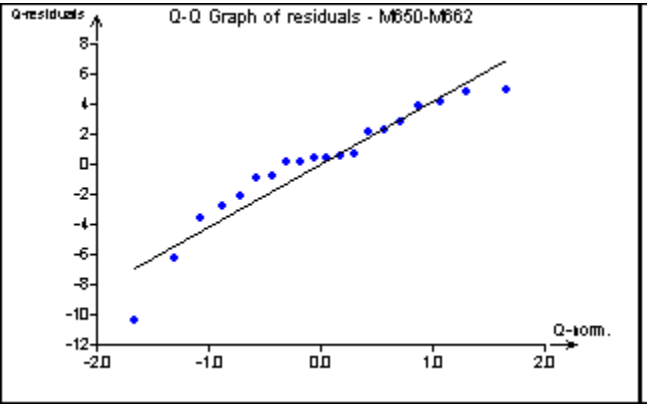
Odhady parametrů

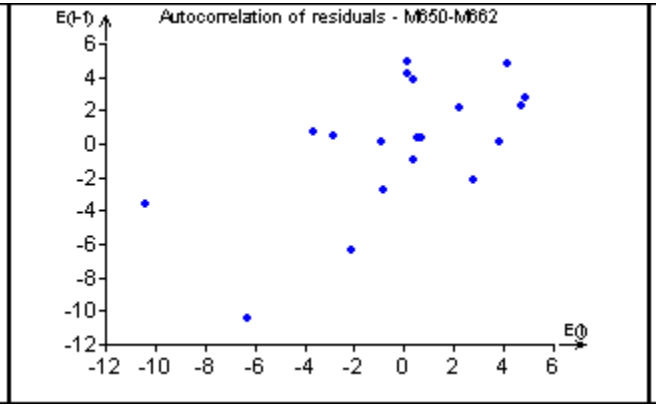
Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-33,65413317	7,551254979	Významný	0,0003975745162	-49,66207862	-17,64618773
M653x1	5,484202459	0,2769631171	Významný	1,116218229E-012	4,897066879	6,071338038
M653x2	-1,174953315	0,08515766747	Významný	2,657285503E-010	-1,355479505	-0,9944271244
M653x3	-0,4170599852	0,1176032595	Významný	0,00268726349	-0,6663677583	-0,1677522122

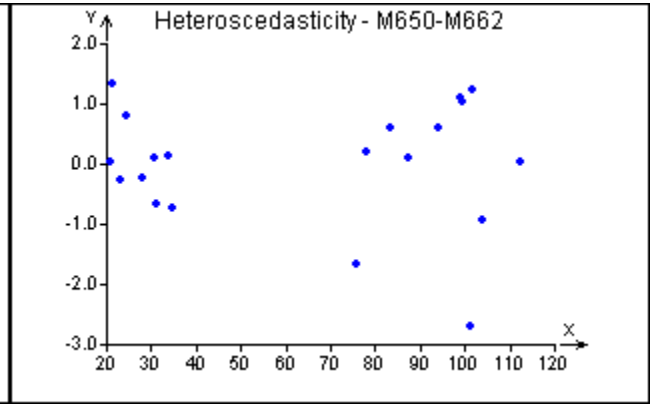
Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0,9960355375
Koeficient determinace R^2 :	0,9920867919
Predikovaný korelační koeficient Rp :	0,9776877371
Střední kvdratická chyba predikce MEP :	19,30425947
Akaikeho informační kritérium :	60,22488937

KRITIKA METODY: Grafická analýza







Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu

Hodnota kritéria F :	668,6453187
Kvantil F (1-alfa, m-1, n-m) :	3,238871517
Pravděpodobnost :	5,114879406E-017
Závěr :	Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC :	0,626144312
Závěr :	Model vykazuje multikolinearitu!

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW :	1,247794863
Kvantil Chi^2(1-alfa,1) :	3,841458829
Pravděpodobnost :	0,2639740648
Závěr :	Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB :	4,820477866
Kvantil Chi^2(1-alfa,2) :	5,991464547
Pravděpodobnost :	0,0897938373
Závěr :	Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA :	6,652281048
Kvantil Chi^2(1-alfa,1) :	3,841458829
Pravděpodobnost :	0,009902896597
Závěr :	Autokorelace je významná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW :	-1	
Kritické hodnoty DW	0,9	1,83
Závěr :	Rezidua jsou pozitivně autokorelována!	

Znaménkový test reziduí

Hodnota kritéria Sg :	1,827747067
Kvantil N(1-alfa/2) :	1,959963999
Pravděpodobnost :	0,06758752075
Závěr :	V reziduích není trend.