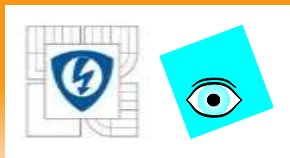


Statistická analýza jednorozměrných dat

Prof. RNDr. Milan Meloun, DrSc.
Univerzita Pardubice, Pardubice

31.ledna 2011

Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky.



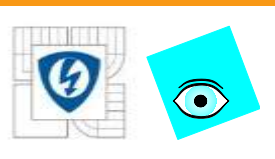
Kapitola 6.5

REGRESNÍ MODEL

31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Regresní model

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

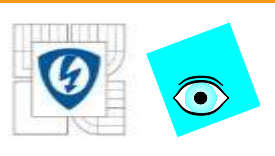
závisle
náhodná
proměnná
chyba

nezávisle proměnná

regresní
parametry

a zapsáno vektorově

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



Hodnocení kvality regresního modelu

Střední kvadratická chyba predikce (MEP)

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

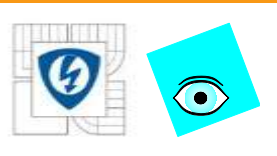
e_i^2 čtverec reziduí modelu
 H_{ii} i -tý diagonální prvek
projekční matice H

Akaikovo informační kritérium (AIC)

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

RSC reziduální součet čtverců
 m počet parametrů

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



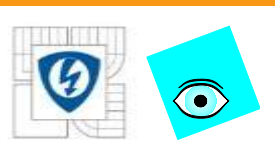
Regresní diagnostika – kvalita modelu

1) Graf reziduí

2) Parciální regresní grafy

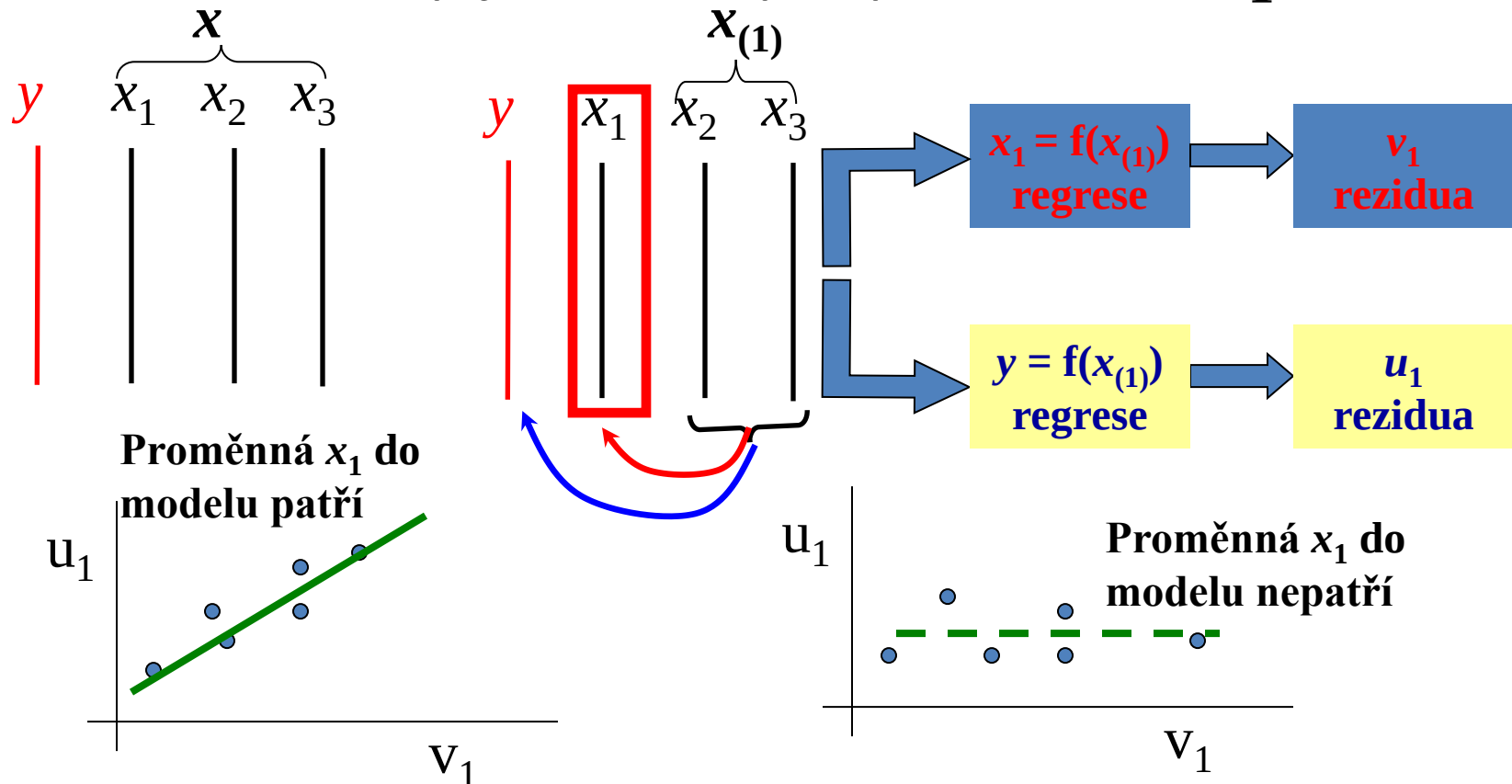
Vyjadřuje závislost **mezi vysvětlovanou proměnnou** (vektorem **y**) a jednou **vysvětlující proměnnou x_j** při statisticky neměnném vlivu ostatních vysvětlujících proměnných, které tvoří matici $X_{(j)}$ (kdy je vynechána j -tá proměnná).

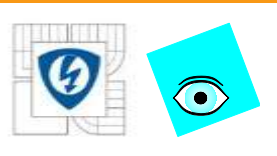
Jde zde o určitou grafickou obdobu parciálního korelačního koeficientu u korelačních modelů.



Regresní diagnostika – kvalita modelu

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





Hodnocení regresního modelu

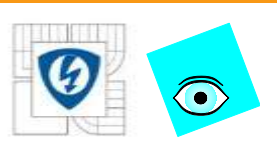
1. Kvalita nalezených odhadů parametrů

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

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

b) Podle rozptylů parametrů, kde by pro statisticky významný odhad mělo platit pravidlo Sillénovy podmínky

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



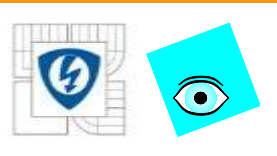
Hodnocení regresního modelu

2. Kvalita dosažené těsnosti proložení

- a) podle reziduálního rozptylu,
- b) podle regresního rabatu, což je v procentech vyjádřený koeficient. determinace (čím více se blíží 100%, tím lepší je proložení modelem).

3. Vhodnost navrženého modelu

- a) Akaikovo informační kritérium (AIC): čím je AIC menší, tím je model vhodnější.
- b) Střední kvadratická chyba predikce (MEP): čím je MEP menší, tím je predikční schopnost modelu lepší.



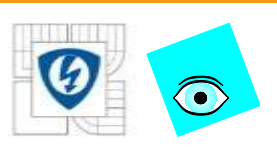
Hodnocení regresního modelu

4. Predikční schopnost modelu

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

5. Kvalita experimentálních dat

- a) na základě analýzy reziduí,
- b) na základě analýzy vlivných bodů (podle Jackknife reziduí, Cookovy vzdálenosti, diagonální prvky projekční matice a věrohodnostní vzdálenost).



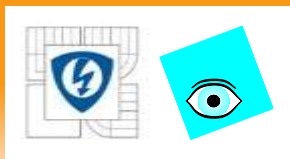
Kvalita navrženého regresního modelu

1. Jediná proměnná x : rozptylový graf y na x .

Více proměnných x : rozptylové grafy mohou *mylně indikovat* nelinearitu.

2. K posouzení vztahu y a x_j :

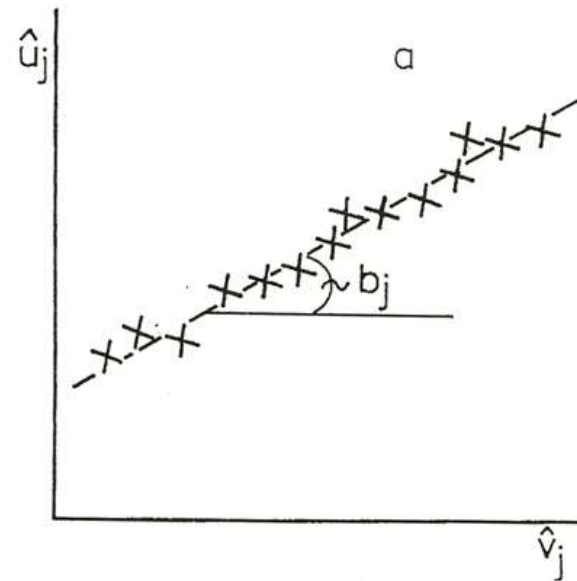
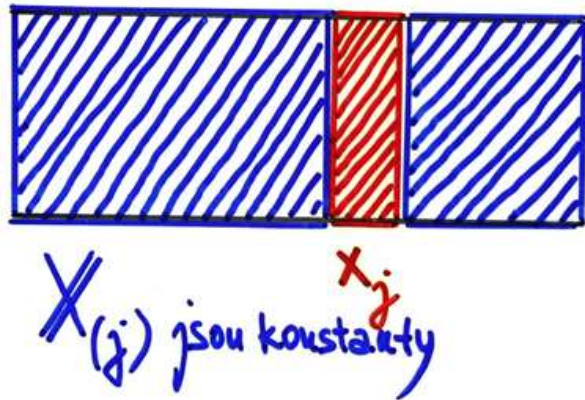
- a) parciální regresní grafy,
- b) parciální reziduální grafy.



Belseyho parciální regresní grafy

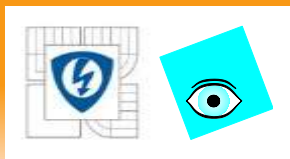
Partial Regression Leverage Plots

$$y = X_j \beta^* + x_j c + \varepsilon$$



Umožňují:

1. posouzení kvality navrženého regresního modelu,
2. indikují přítomnost vlivných bodů,
3. nesplnění předpokladů klasické MNČ,
4. vyjadřují závislost mezi y a zvolenou proměnnou x_j při statisticky neměnném vlivu ostatních x_j .



Belseyho parciální regresní grafy

Partial Regression Leverage Plots

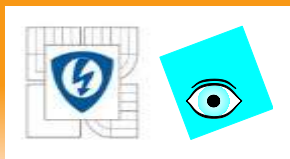
Vychází se z regresního modelu

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

kde β^* má rozměr $(m - 1) \times 1$, c je regresní parametr příslušející j -té proměnné.

Projekcí do prostoru kolmého na prostor sloupců matice $X_{(j)}$ bude

$P_{(j)} y$	=	$P_{(j)} X_{(j)}$	+	$P_{(j)} x_j c$	+	$P_{(j)} \varepsilon$
$\hat{u}_j$		$= 0$		$\hat{v}_j$		
vektor reziduí proměnné y na proměnných, které tvoří sloupce $X_{(j)}$				vektor reziduí proměnné y na proměnných, které tvoří sloupce $X_{(j)}$		



Belseyho parciální regresní grafy

Partial Regression Leverage Plots

Střední hodnota $E(\hat{u}_j)$ je

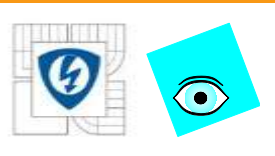
$$E(\hat{u}_j) = cE(\hat{v}_j)$$

a závislost $\hat{u}_j$ na $\hat{v}_j$ tvoří **parciální regresní graf** (u správného modelu jde o závislost s nulovým úsekem).

Odhad směrnice z MNČ bude

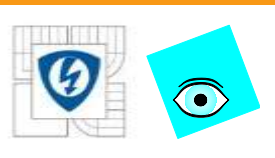
$$\hat{c} = \frac{\hat{u}_j^T \hat{u}_j}{\hat{v}_j^T \hat{v}_j} = \frac{x_j^T P_{(j)} y}{x_j^T P_{(j)} x_j}$$

platí $\hat{e} = \hat{u}_j - \hat{v}_j \hat{c}$ a ukazuje na vztah mezi $\hat{e}$ s parciálními rezidui $\hat{u}_j$ a $\hat{v}_j$.



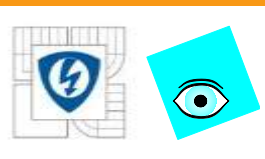
Vlastnosti

- a) směrnice $\hat{c}$ je stejná s b_j v neděleném modelu a úsek je roven nule, když je navržený model správný,
- b) korelační koeficient mezi $\hat{u}_j$ a $\hat{v}_j$ odpovídá parciálnímu korelačnímu koeficientu $\hat{R}_{jx_j(x)}$,
- c) rezidua regresní přímky v parciálním regresním grafu jsou shodná s rezidui $\hat{e}_i$ pro nedělený model
- d) v grafu jsou zvýrazněny vlivné body a některá porušení předpokladů MNČ (heteroskedasticita)



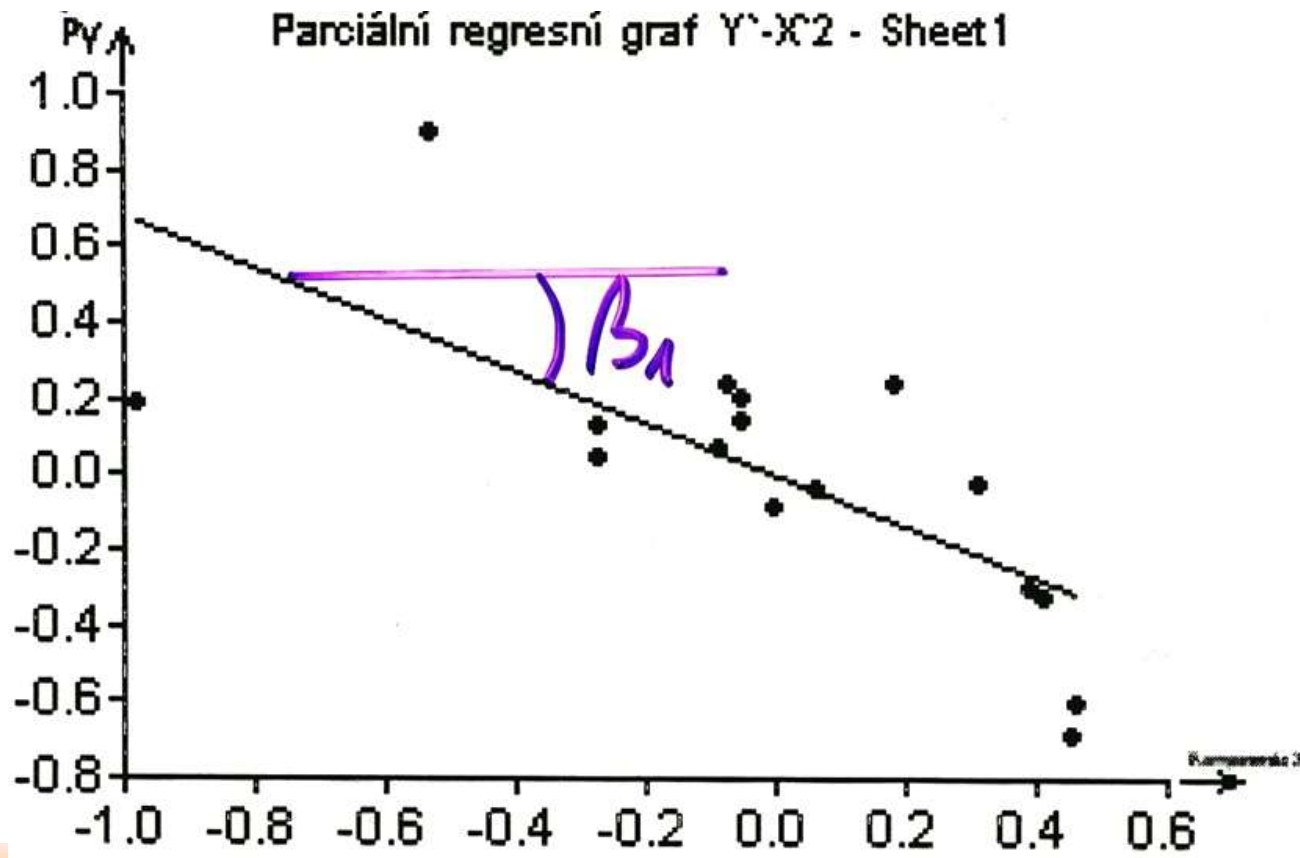
Nevýhody

- a) x-ové souřadnice v $\hat{v}_j$ parciálních regresních grafech nejsou v původním měřítku proměnné x_j ,
- b) jsou-li proměnné sloupců matice X silně korelované, nemusí parciální regresní grafy indikovat správně nelinearitu, a tím i vhodnost navrženého modelu.



Kritika modelu – parciální regresní graf

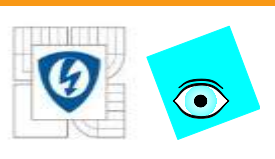
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$



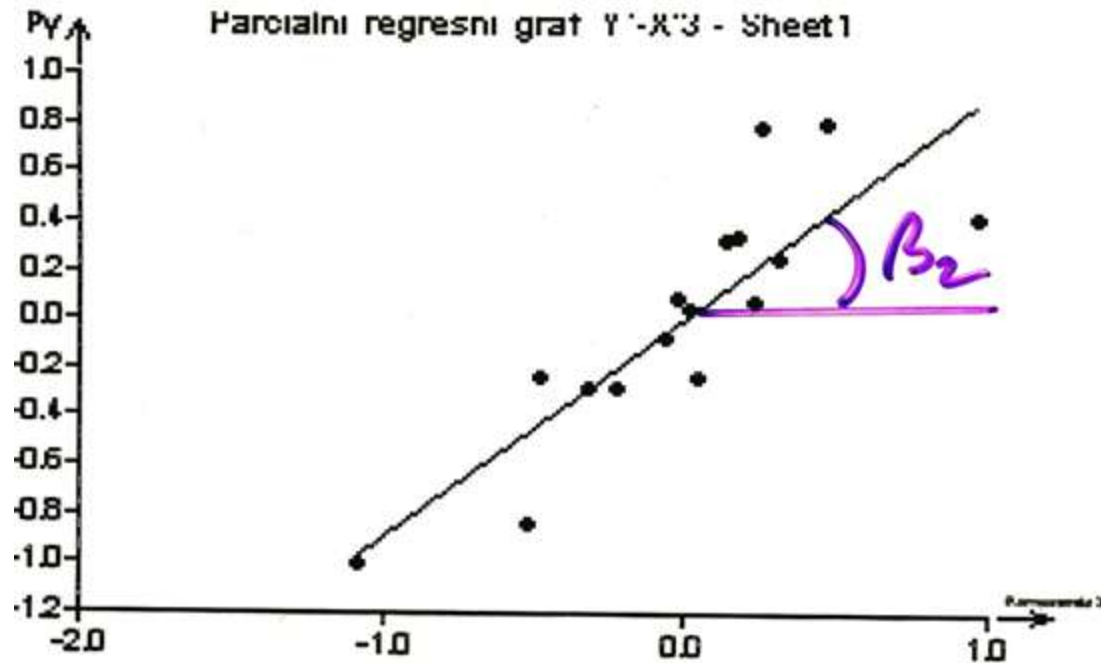
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





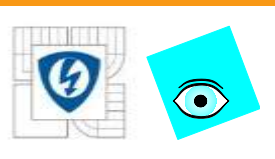
Parciální regresní graf



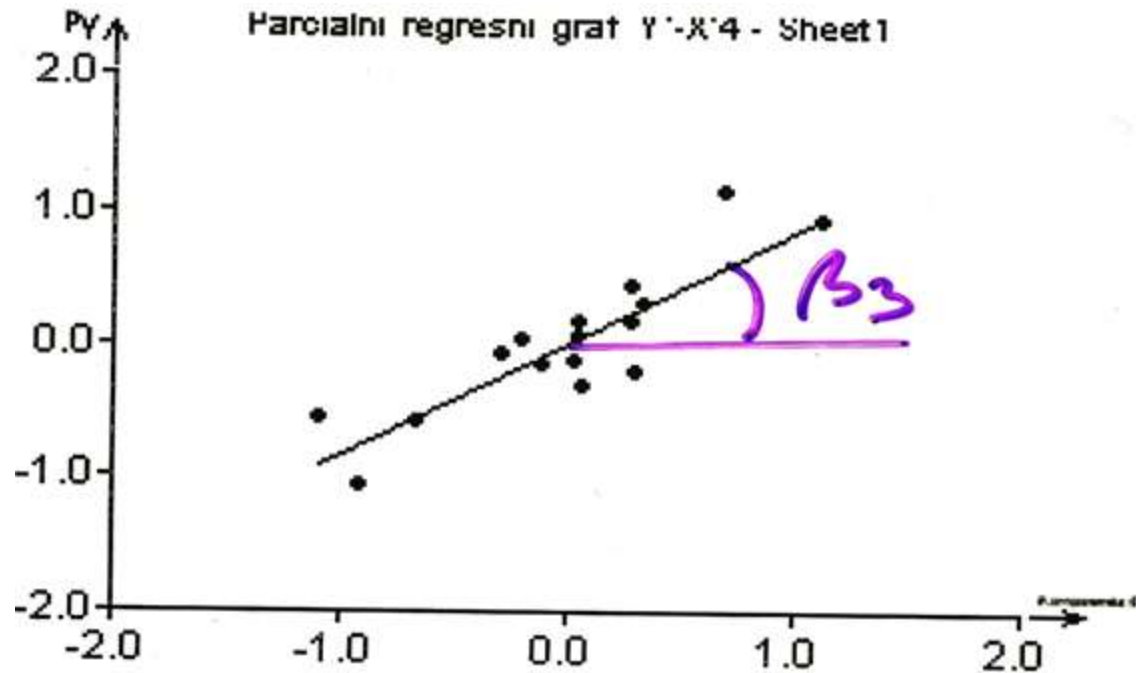
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





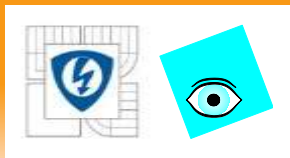
Parciální regresní graf



31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



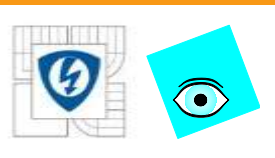


Procentní ztráta amoniaku v aparatuře při výrobě kyseliny dusičné

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 chladicí vody x_2 a koncentrací kyseliny dusičné x_3 . Vyšetřete regresní triplet při užití metody nejmenších čtverců MNČ a L_1 – aproximace.

Model:

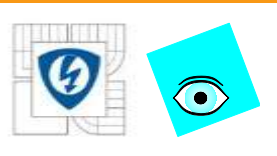
$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4$$



Procentní ztráta amoniaku v aparatuře při výrobě kyseliny dusičné

Data: $n = 21$ bodů

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



Procentní ztráta amoniaku v aparatuře při výrobě kyseliny dusičné

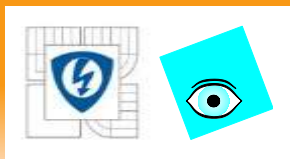
Program: VLR, ILWLS

Řešení:

$$\text{MNČ: } y = -39.92 + 0.7156x_1 + 1.295x_2 - 0.1521x_3$$

L_1 – aproximace:

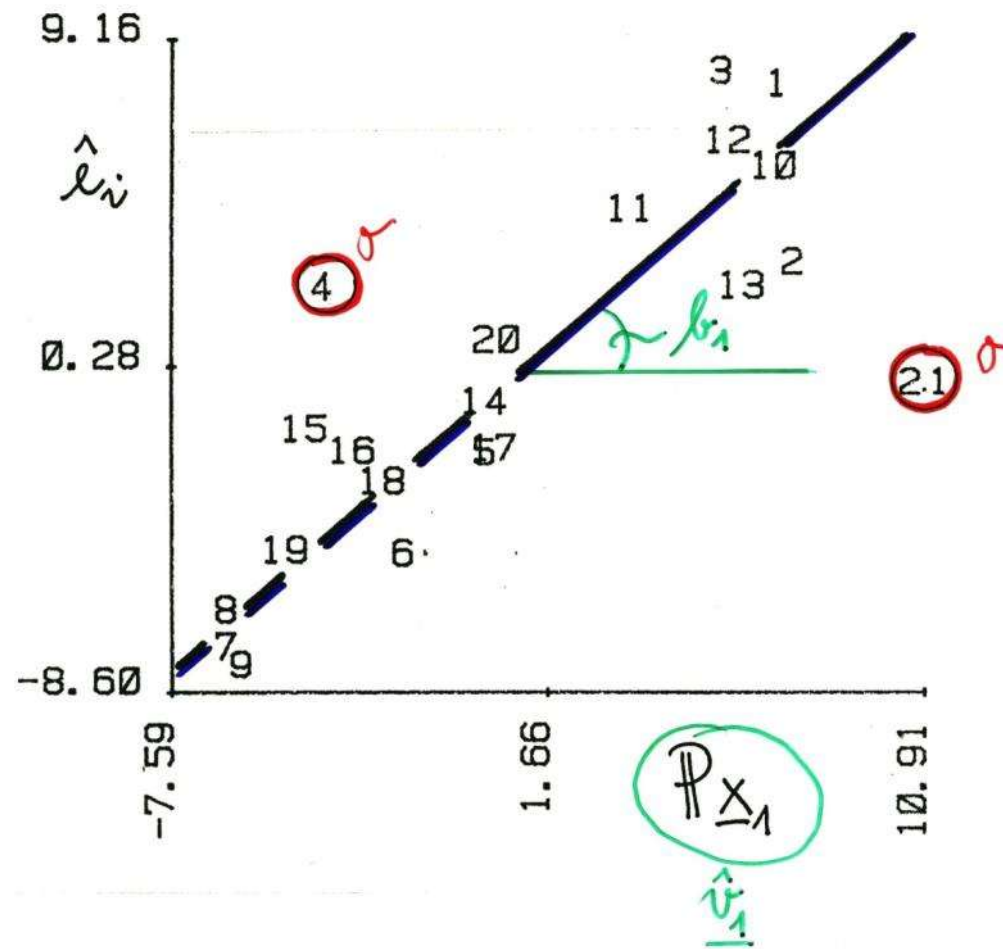
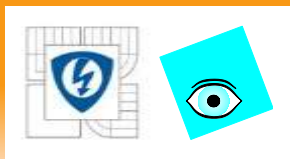
$$y = -39.65 + 0.83x_1 + 0.581x_2 - 0.0621x_3$$

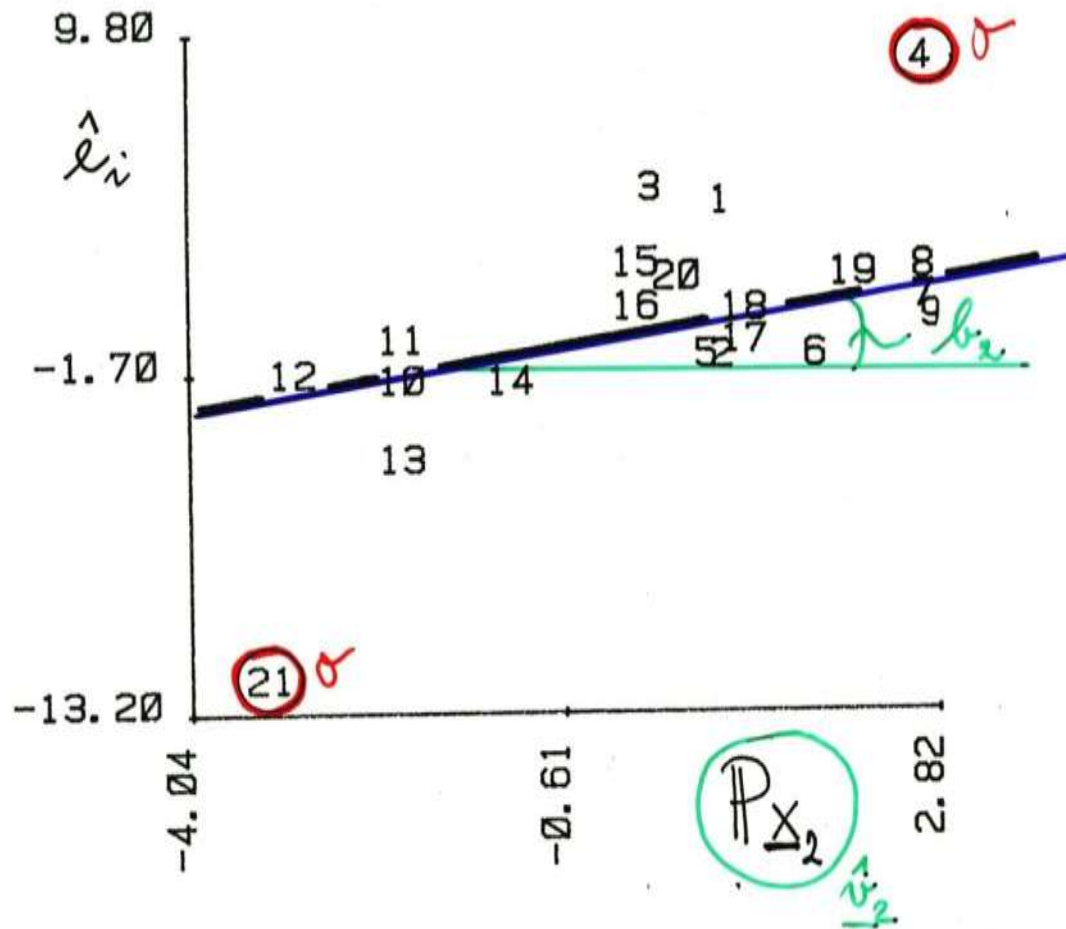
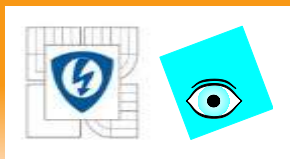


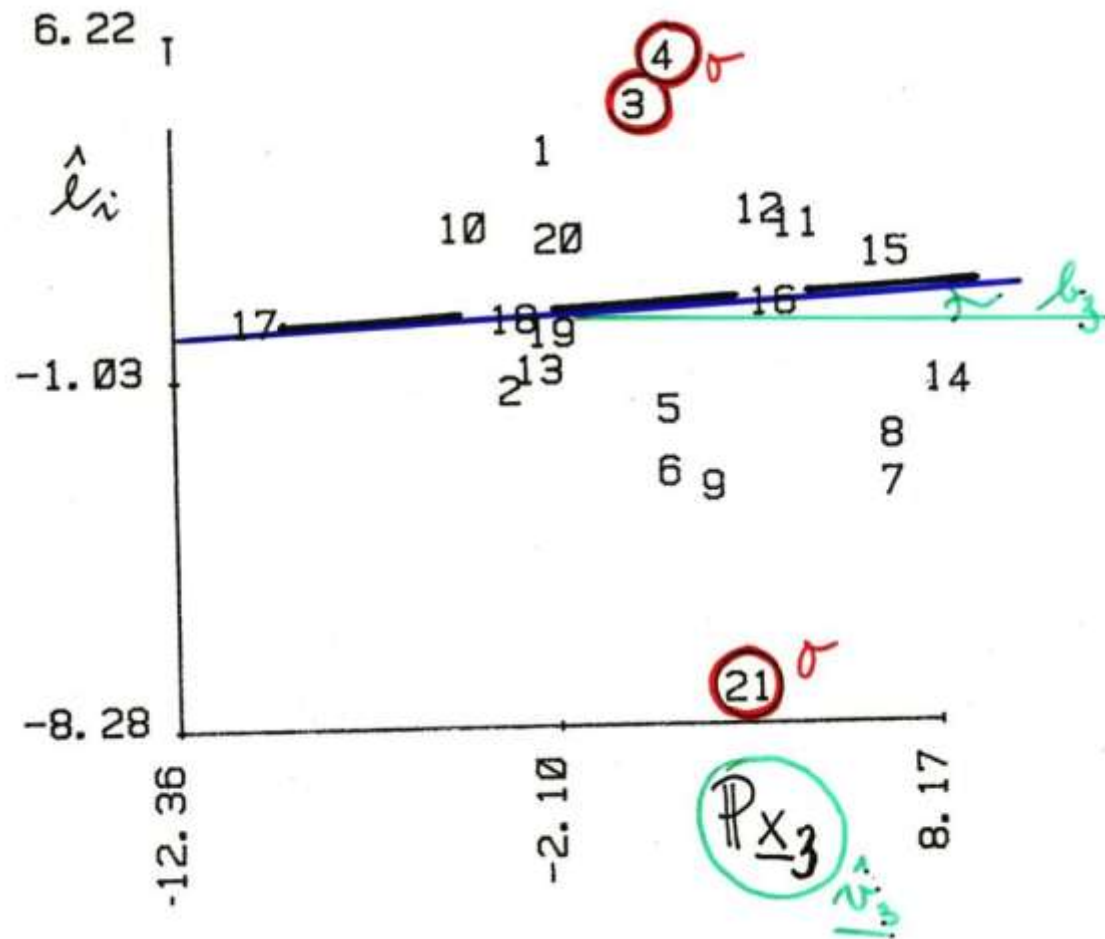
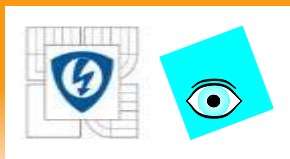
Procentní ztráta amoniaku v aparatuře při výrobě kyseliny dusičné

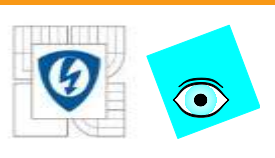
Ověření kvality navrženého regresního modelu

- (1) X-Y Scatterův graf: y na x_j
- (2) Parciální regresní graf (GAV1) pro k -tou proměnnou: určení kvality modelu
 - pro k -tou proměnnou: $z = x_k$,
 - původní model $y = X_{(k)} \cdot b + \varepsilon$, kde $X_{(k)}$ je matice X bez k -tého sloupce,
 - GAV1 graf interpretuje: $\hat{e}$ vs. Px_k
 - GAV1 graf indikuje: částečnou závislost mezi y a x_k
 - (a) nelinearitu nebo linearitu (sklon je b_k), (b) outliers, gross errors a (c) heteroskedasticitu.



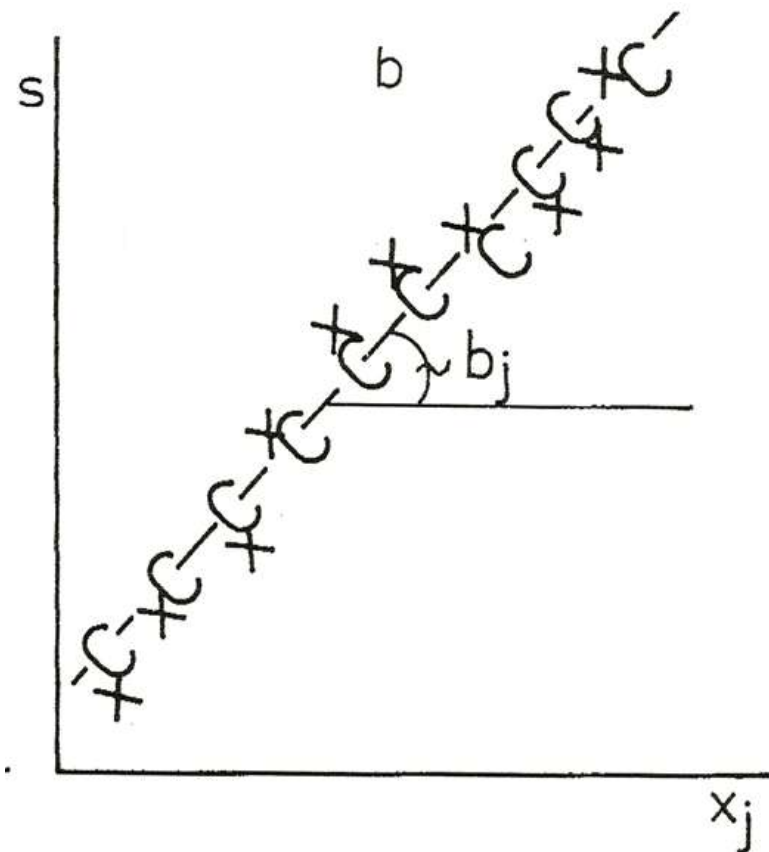


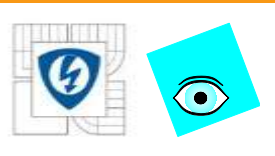




Parciální reziduální grafy

(grafy „komponenta + reziduum“)





Parciální reziduální grafy

Rovnici $\hat{e} = \hat{u}_j + \hat{v}_j \cdot b_j$ přepíšeme do tvaru

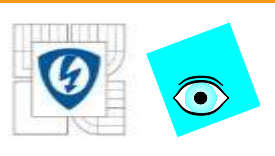
$$\hat{u}_j = \hat{e} + b_j(E - H_{(j)})x_j$$

a vyjádříme jako závislost veličiny $\hat{e} + b_j(E - H_{(j)})x_j$ na $(E - H_{(j)})x_j$.

Parciální reziduální graf je analogií parc. regr. grafu při volbě $H_{(j)} = 0$.

Jedná se o závislost parciálních reziduí s na proměnné x_j .

$$s = \hat{e} + bx_j = y - \sum_{k \neq j}^m x_k b_k$$



Parciální reziduální grafy

Pokud model obsahuje absolutní člen, je možno použít modifikovaná parciální rezidua

$$s_i^* = \hat{e}_i + (x_{ij} - \bar{x}_j)b_j + \bar{y}$$

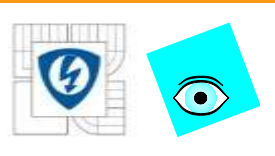
kde $\bar{x}_j$ a $\bar{y}$ jsou aritmetické průměry veličin x_j a y .

V grafu se znázorňuje deterministická komponenta

$$c_{ij} = (x_{ij} - \bar{x}_j)b_j \quad i = 1, \dots, n$$

která se značí písmenem „C“ a parciální reziduum

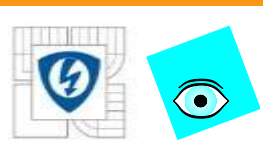
$s_i = c_{ij} + \hat{e}_i$, $i = 1, \dots, n$, které se označuje křížkem „+“.



Parciální reziduální grafy

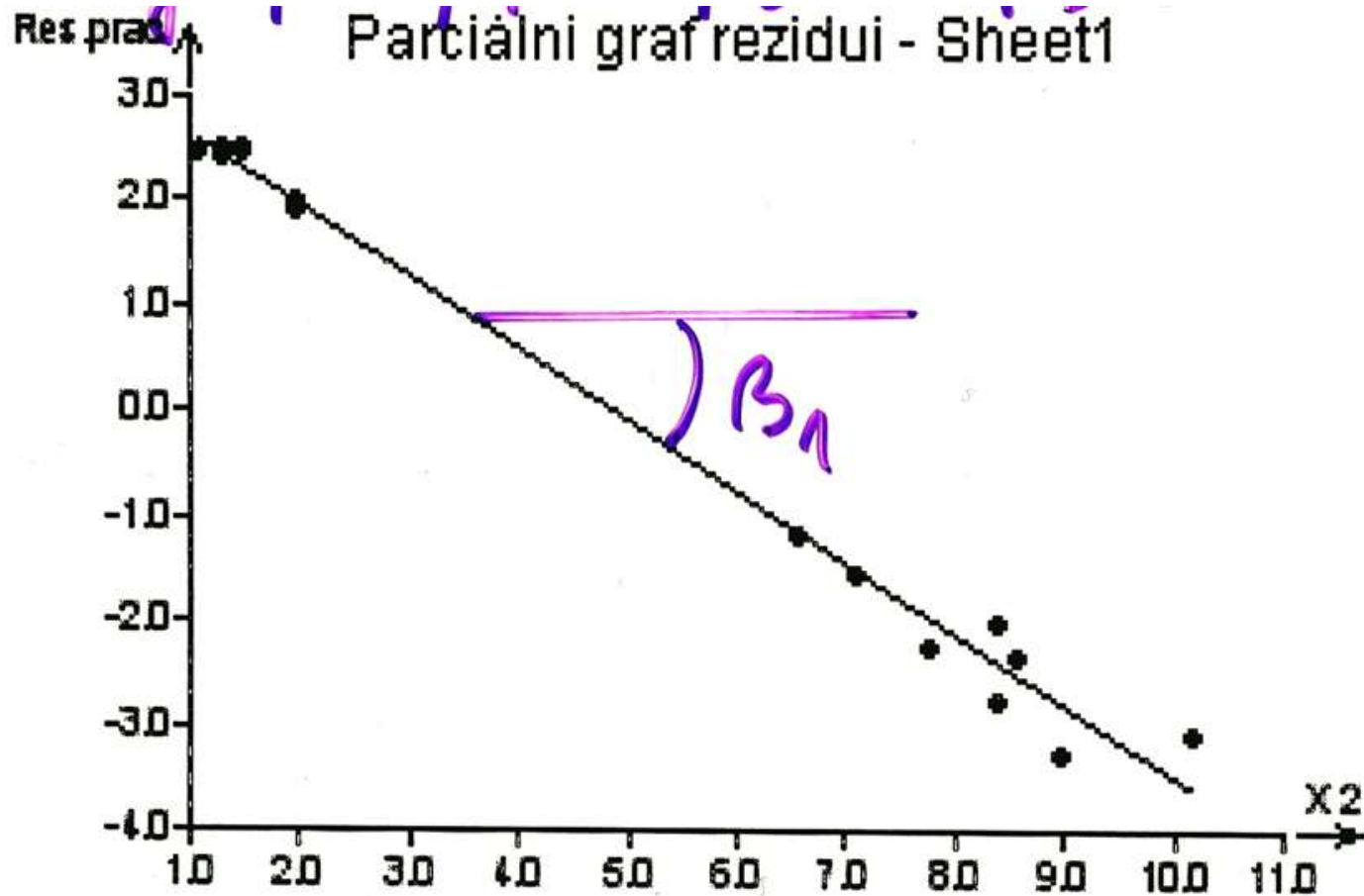
Vlastnosti:

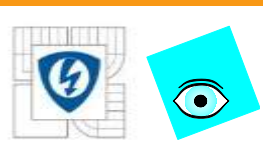
- a) směrnice závislosti s na x_j je rovna b_j a úsek je nulový. Lineární závislost ukazuje na vhodnost navržené x_j v modelu,
- b) rezidua přímky jsou přímo rezidua $\hat{e}_i$ pro nedělený model,
- c) pokud je úhel mezi x_j a některými sloupci matice $X_{(j)}$ malý (multikolinearita), ukazuje parciální reziduální graf nesprávně malý rozptyl kolem regresní přímky $b_j x_j$ a dochází i k potlačení efektu vlivných bodů.



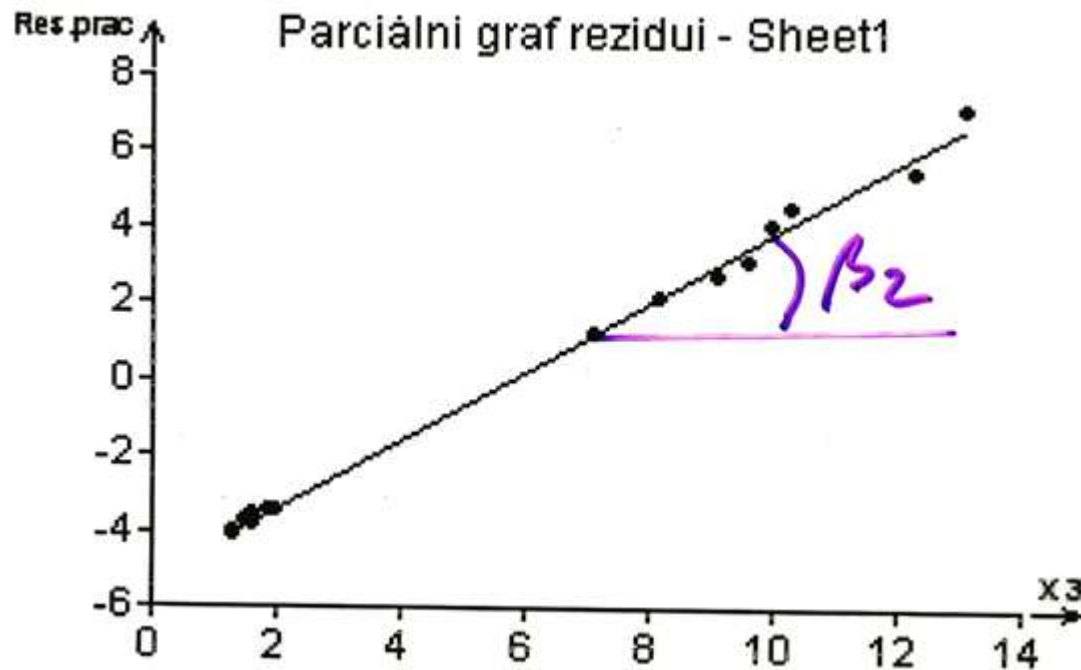
Kritika modelu: parciální reziduální graf

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





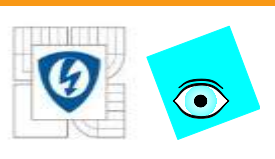
Parciální graf reziduí



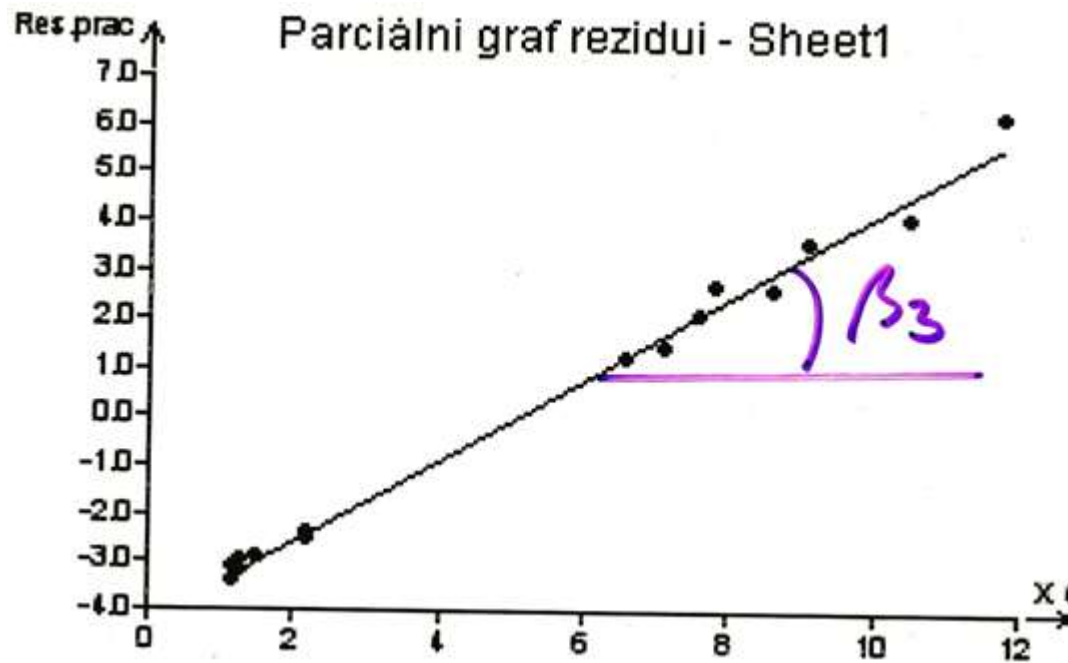
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





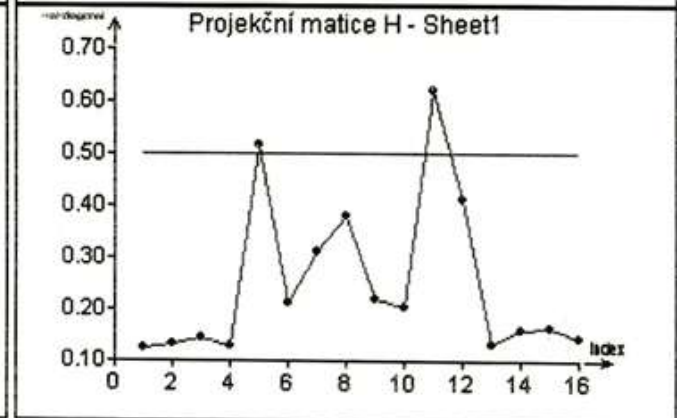
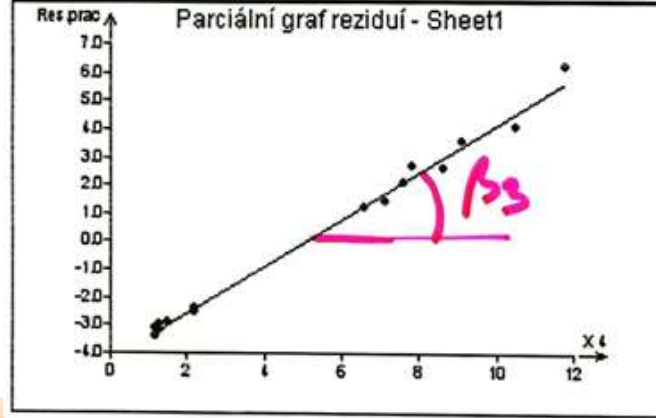
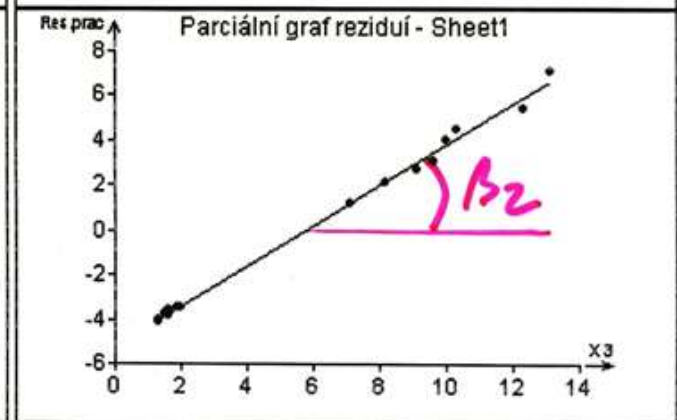
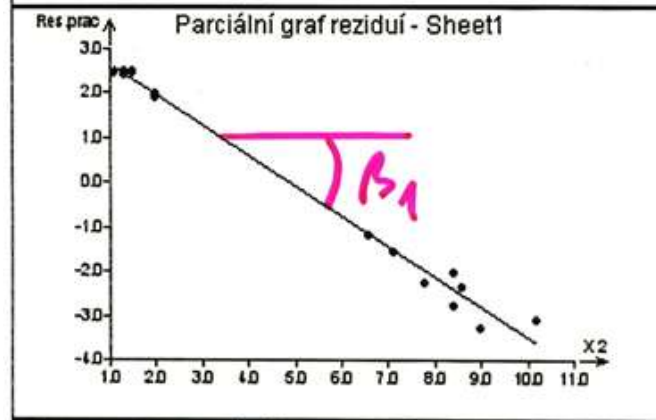
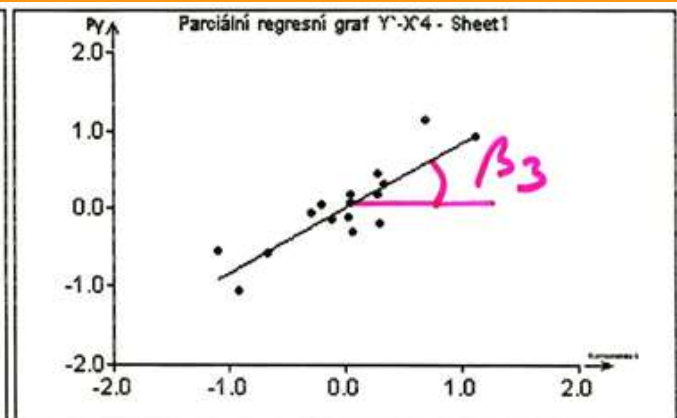
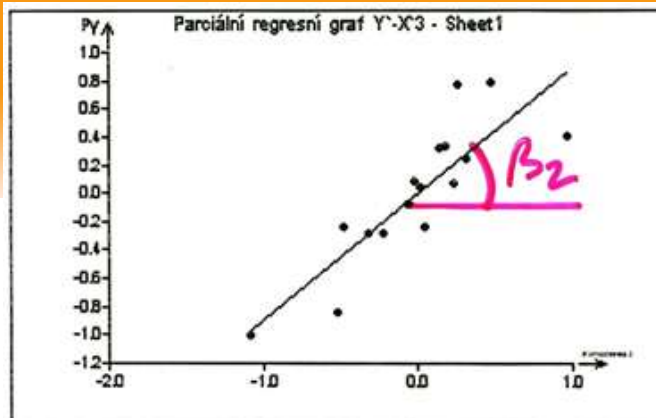
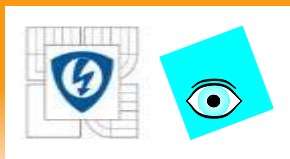
Parciální graf reziduí



31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

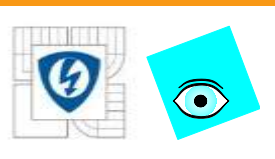




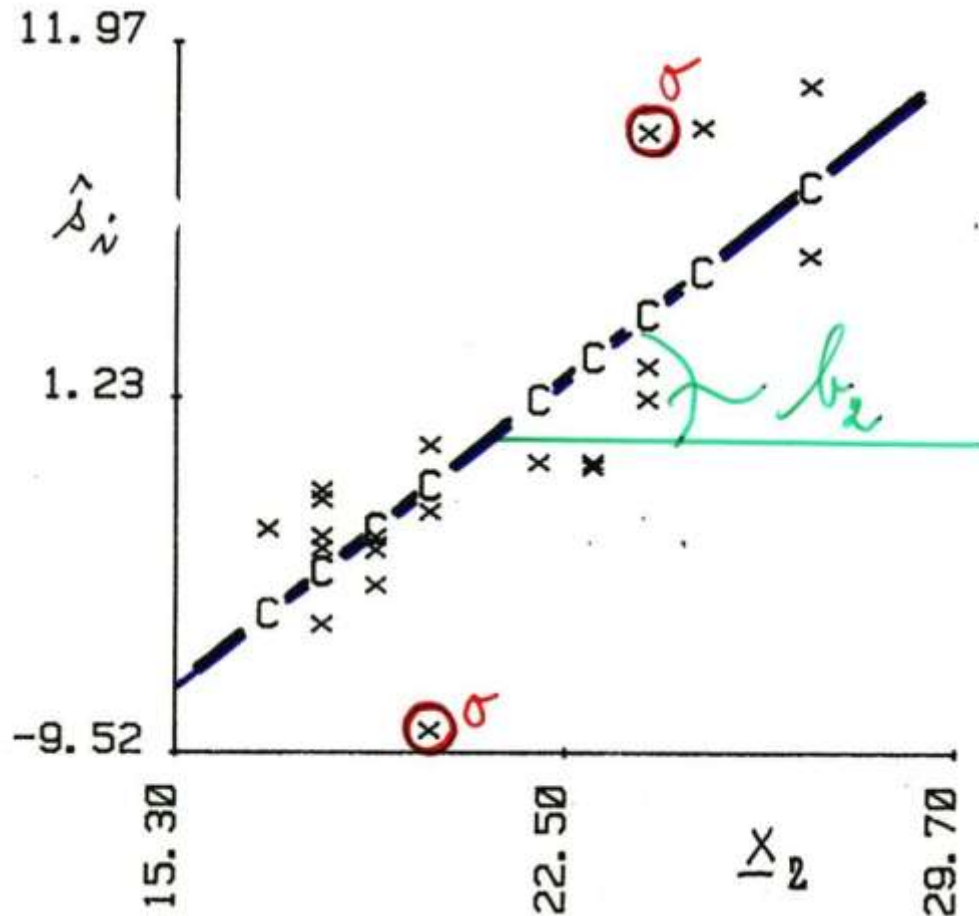
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





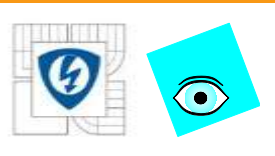
Parciální reziduální grafy



31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Parciální reziduální graf

$$\hat{u}_j = \hat{e} + b_j(E - H_{(j)})x_j$$
$$s = \hat{e} + b \cdot x_j$$

3. Parciální reziduální graf GAV2 pro k-tou proměnnou
pro k-tou proměnnou: $z = x_k$

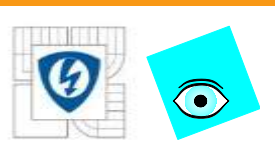
původní model: $\hat{s}_i = \hat{e}_i + (x_{ik} - \bar{x}_k) \cdot b_k$

GAV2 graf zobrazuje $\hat{s}$ na x_k

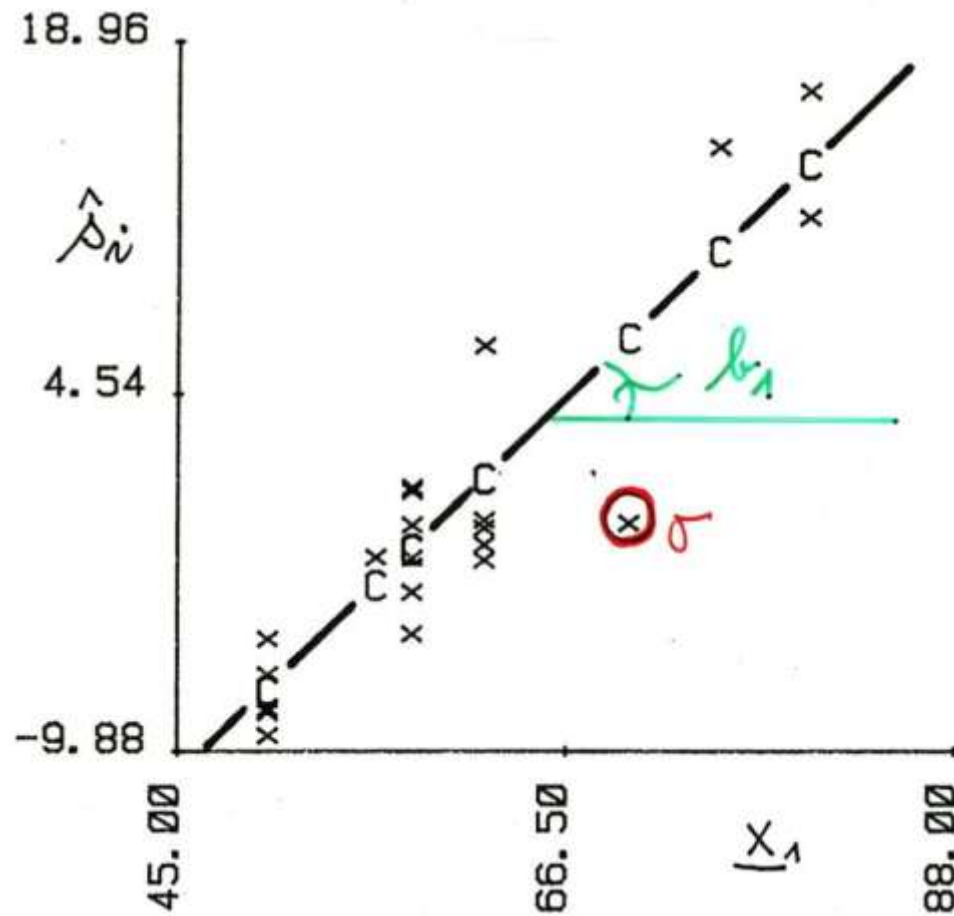
Vlastnosti:

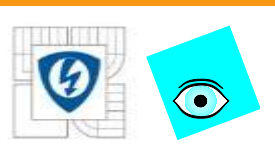
(1) sklon $s = f(x_j)$ je b_j

(2) případ multikolinearity: malý úhel mezi x_j a x_k .

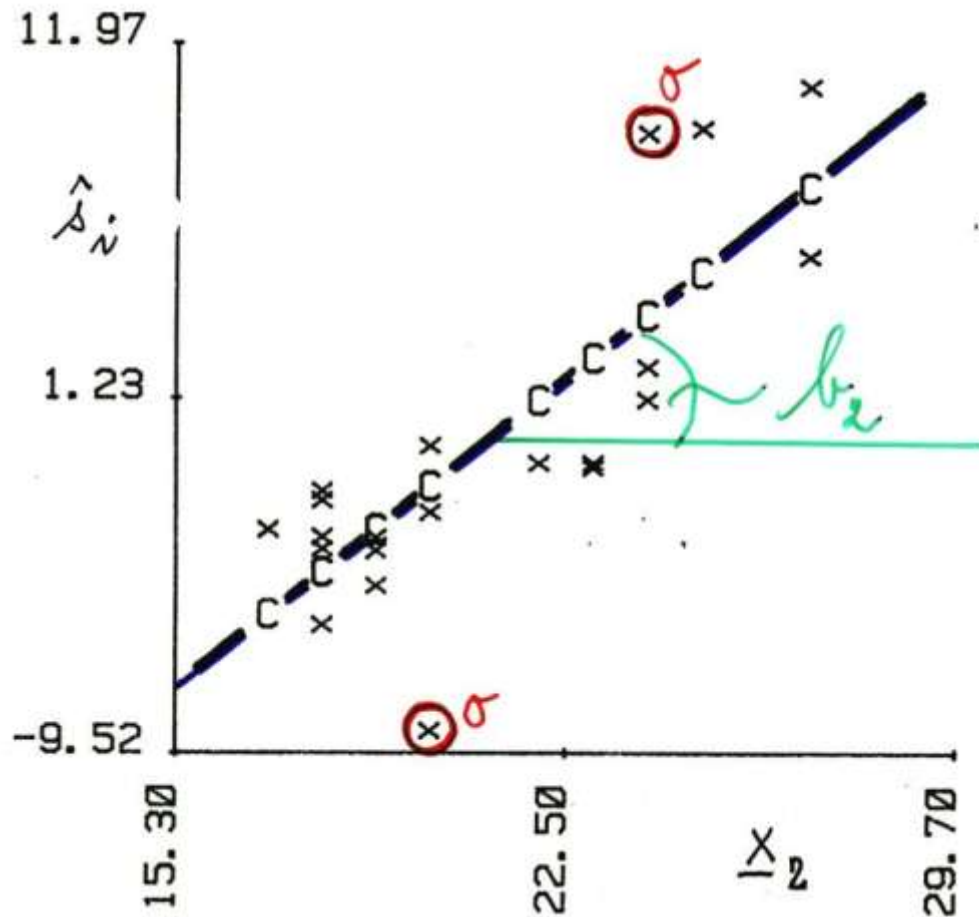


Parciální reziduální grafy





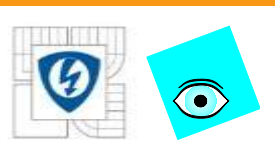
Parciální reziduální grafy



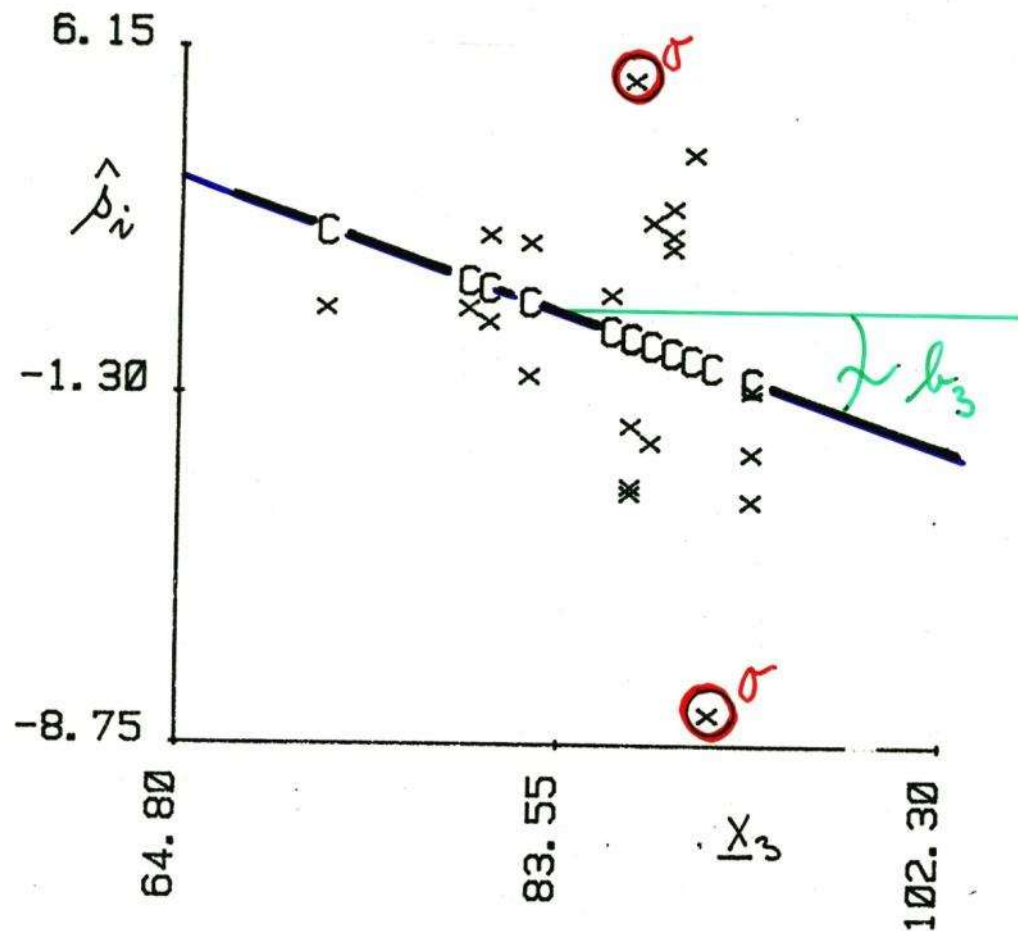
31.1.2011

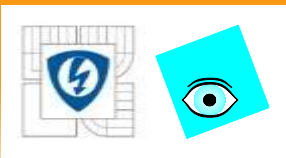
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Parciální reziduální grafy





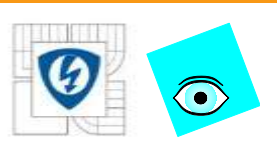
Kapitola 6.6

KRITIKA METODY

31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

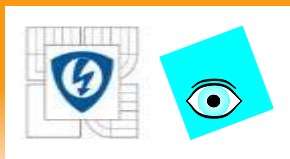




Ověření předpokladů metody nejmenších čtverců

Metoda:

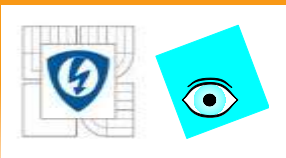
- vyšetření heteroskedasticity,
- vyšetření autokorelace,
- vyšetření multikolinearity,
- vyšetření normality náhodných chyb,
- vyšetření omezení parametrů,
- vyšetření trendů reziduí.



Závažné předpoklady MNČ

- kdy je 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** $D\left(\frac{y}{x}\right) = \sigma^2$ 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

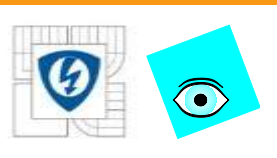
- testy vybraných předpokladů klasické MNČ

Multikolinearita: VIF diagnostika indikuje

Heteroskedasticita: testy heteroskedasticity (např.
Cook-Weisberg)

Autokorelace reziduí: test významnosti autokorelačního
koeficientu

Normalita reziduí: testy normality

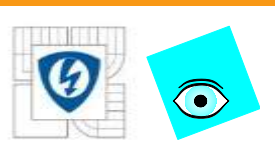


1. Hetero/homoskedasticita (nekonstattnost rozptylu)

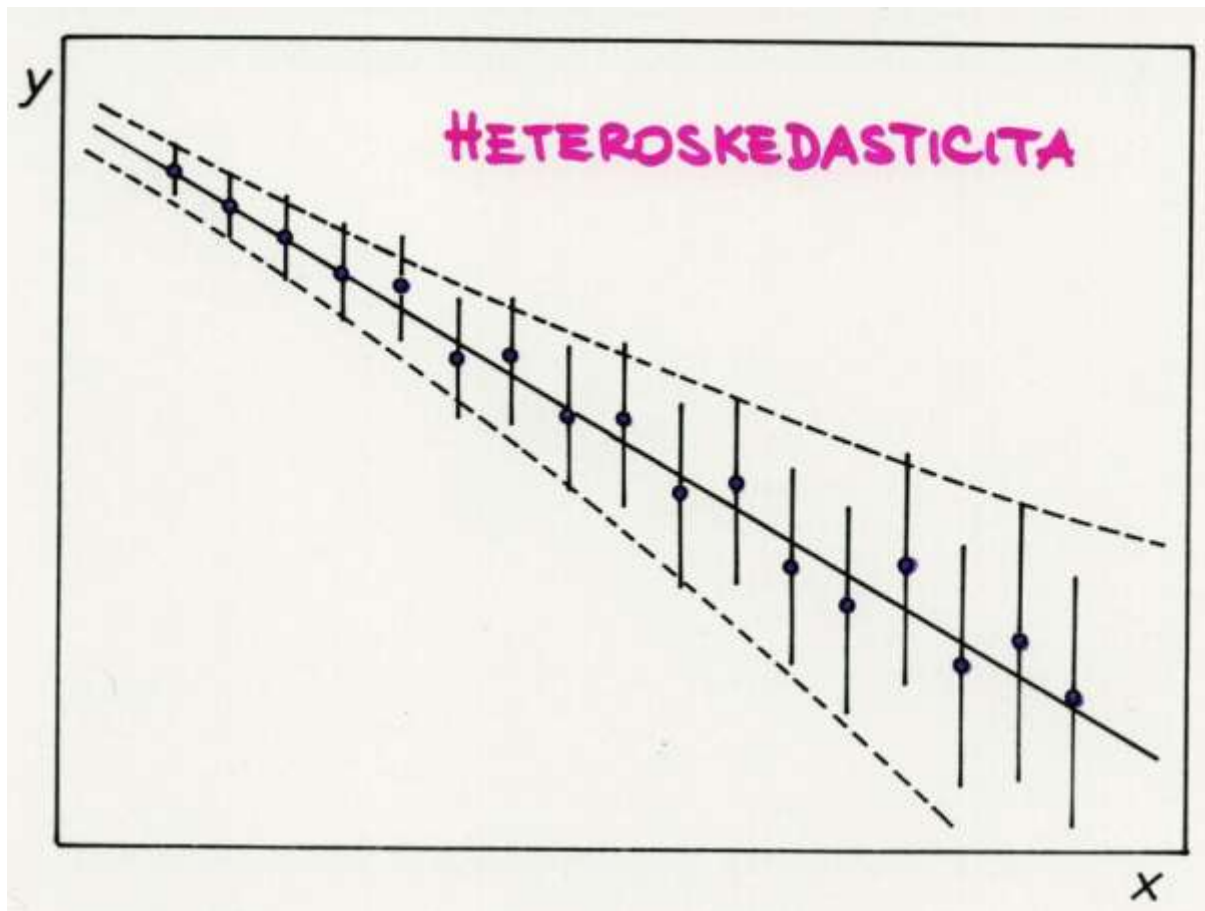
Rozptyl veličiny y_i v i -tém bodě je popsán

$$\sigma_i^2 = \sigma^2 \exp(\lambda x_i \beta)$$

kde x_i je i -tý řádek matice X .

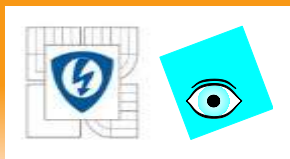


Heteroskedasticita



31.1.2011

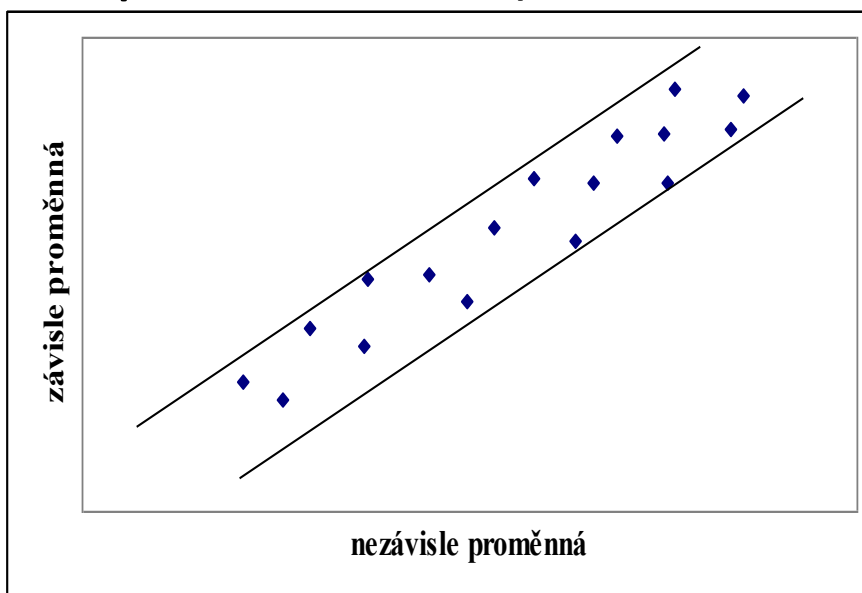
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



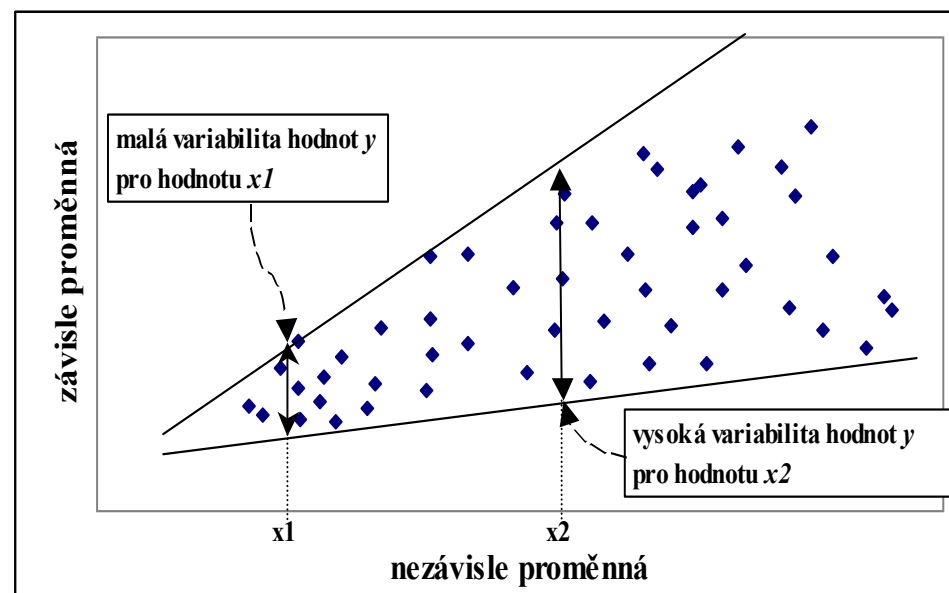
Homoskedasticita vs. heteroskedasticita

- grafické vysvětlení principu

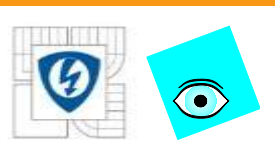
Homoskedasticita znamená, že hodnoty závisle proměnné y mají pro všechny hodnoty nezávisle proměnné x **konstantní rozptyl** (variabilitu, proměnlivost).



Homoskedasticita

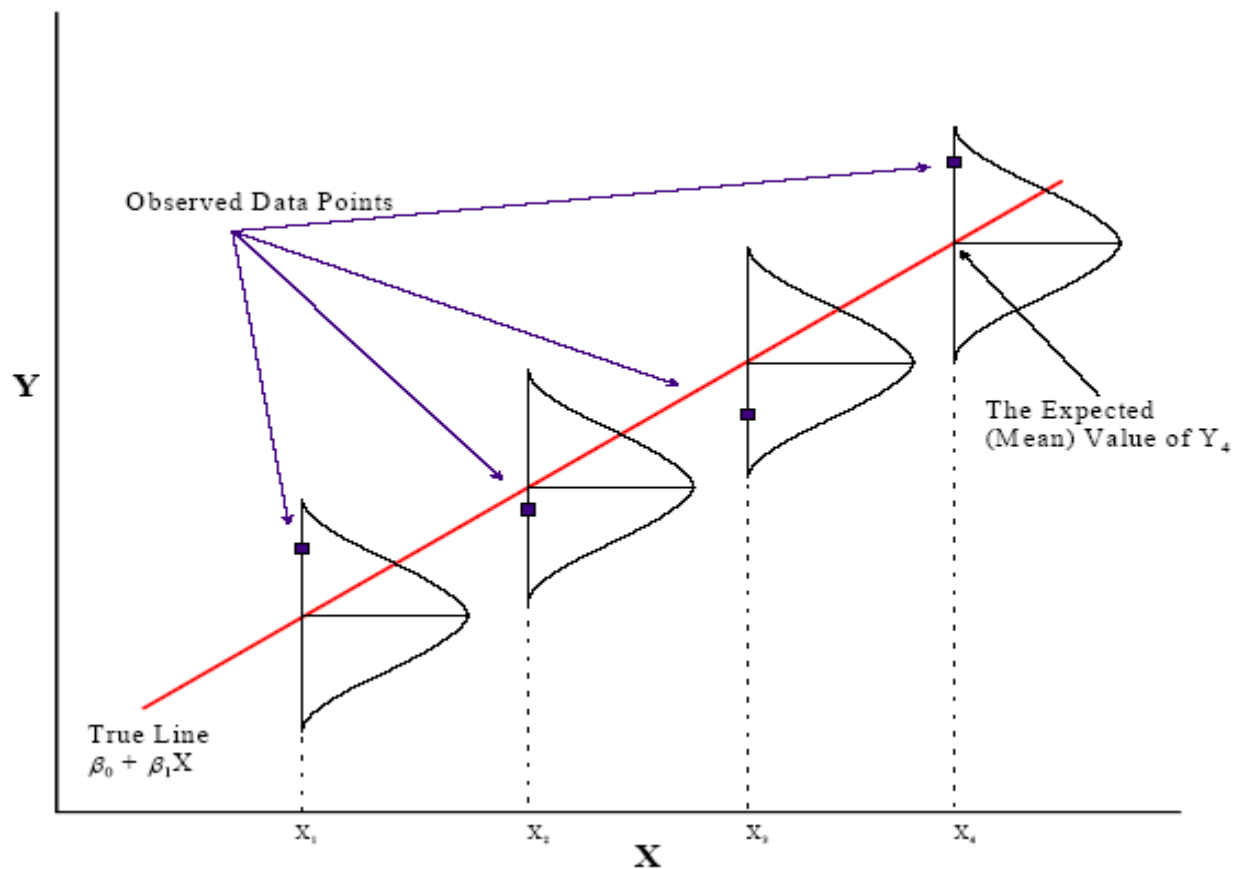


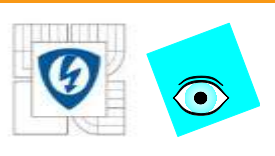
Heteroskedasticita



Homoskedasticita

- grafické vysvětlení principu





Homoskedasticita

- testování

Test trendu reziduí

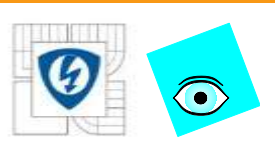
$$D = \sum_{i=1}^n [R(\hat{e}_i) - i]^2$$

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

Testujeme významnost Spearmanova korelačního koeficientu ρ_s



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



Homoskedasticita

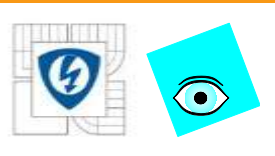
- testování

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

Cookův - Weisbergův test homoskedasticity

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

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



Homoskedasticita

- testování

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

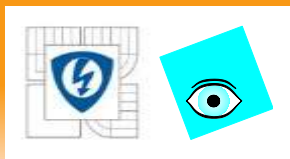
Cook-Weisbergovo testační kritérium

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

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

Testování:

1. Je-li S_f , H_0 (homoskedasticita) je přijata.
2. Pro homoskedasticitu tvoří diagnostický graf $\hat{e}_{Si}^2$ na $(1 - H_{ii})\hat{y}_i$ náhodný mrak bodů. Pro heteroskedasticitu vznikne typický klínový obrazec.



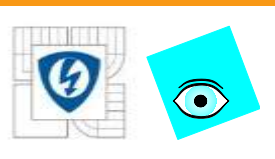
Homoskedasticita

- řešení jednoho z nesplněných předpokladů klasické MNČ

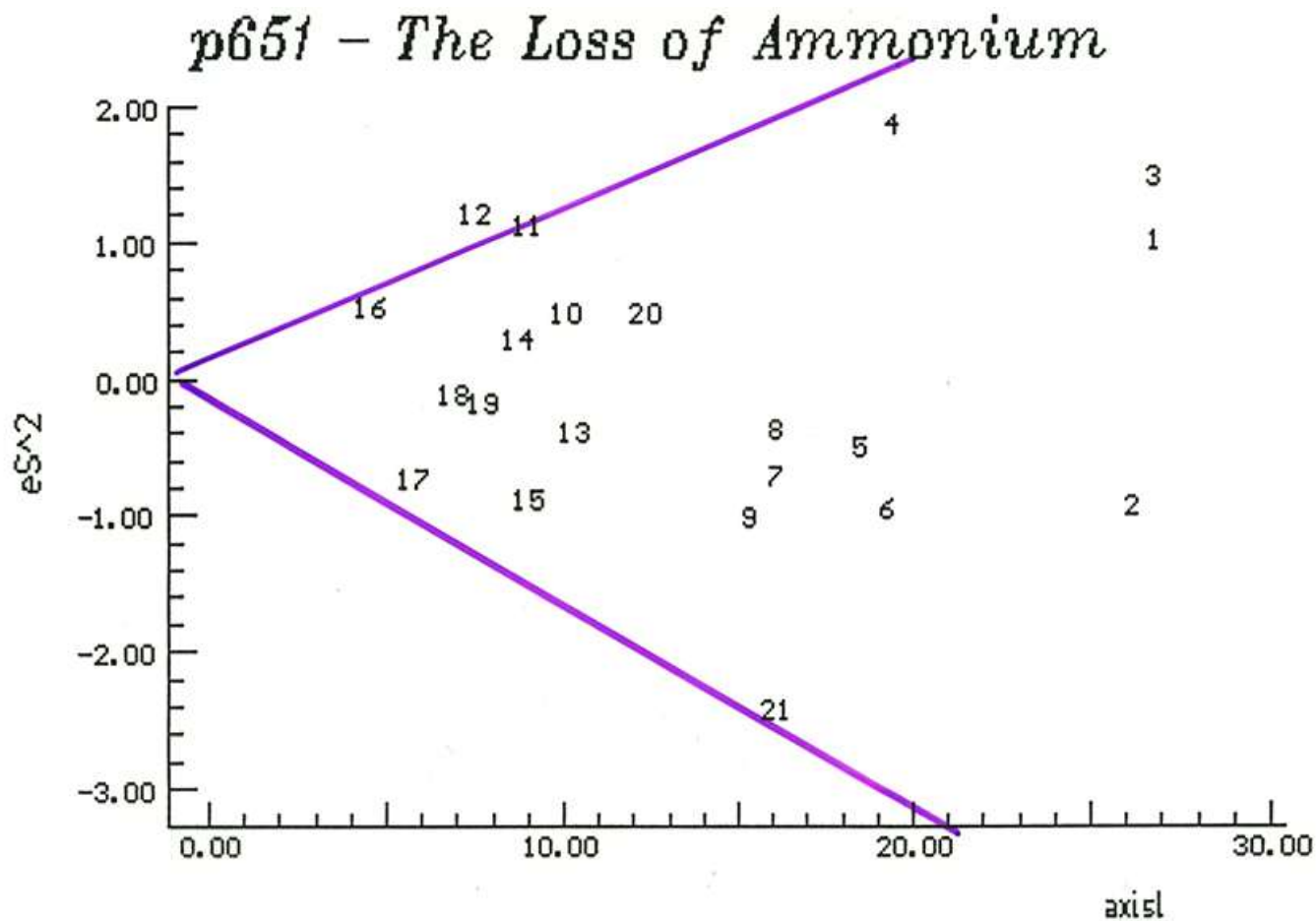
Nejčastějším způsobem je užití metody **vážených nejmenších čtverců**, kdy se účelová funkce čili suma čtverců reziduí násobí **vhodně zvolenými váhami**

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

V běžných případech je možné jako váhy volit hodnoty $1/y_i$ nebo $1/y_i^2$.



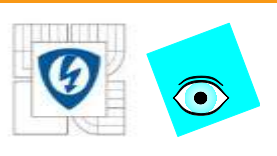
Graf homoskedasticity



31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





2. Autokorelace

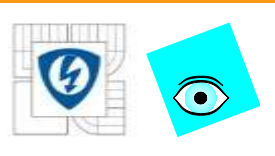
Data časových řad mají chyby ε_i vzájemně korelované.

Nejčastější je případ autokorelace prvního řádu

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

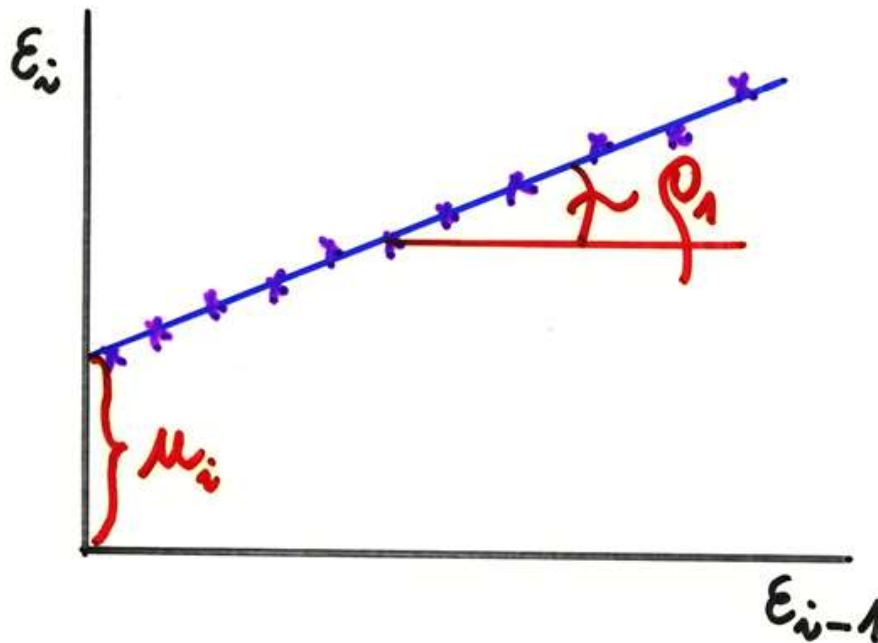
kde $u_i \sim N(0, \sigma^2)$.

- 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.



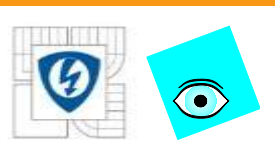
Autokorelace

1. Grafická indikace autokorelace



2. Waldův test pro $\rho_1: H_0: \rho_1 = 0$ 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.



3. Normalita chyb

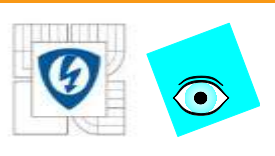
a) **Rankitový (Q-Q) graf**: $\hat{e}_{(i)}$ nebo $\hat{e}_{Ri}$ na u_{Pi} pro $P_i = i/(n + 1)$

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



Normalita chyb

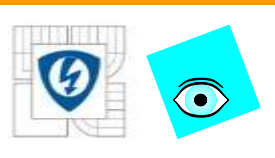
Test: Je-li $L(\hat{e}) > \chi^2_{1-\alpha}(2) = 5.99$, je H_0 (normalita) zamítnuta.

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

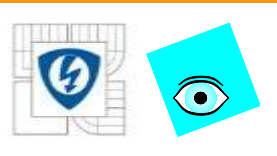
(test není vhodný pro malé výběry)



Normalita chyb

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



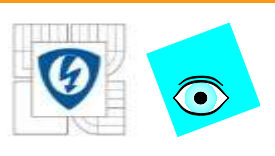


Znaménkový test vhodnosti modelu

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, 2 je počet sekvencí roven $n_U = 4$)
2. Stanoví se počet reziduí kladných (n_+) a záporných (n_-).
3. Teoretický počet sekvencí n_t a jeho rozptyl D_t je



Znaménkový test vhodnosti modelu

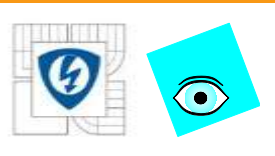
$$n_t = 1 + \frac{2n_+n_-}{n_+ + n_-} \approx 1 + \frac{n}{2}$$

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

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

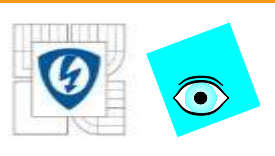
$$U = \frac{n_u - n_t + C}{\sqrt{D_t}}$$

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



Znaménkový test vhodnosti modelu

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



Testování regresního tripletu

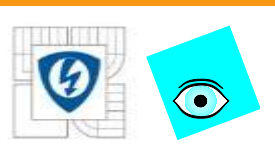
(6) TESTOVÁNÍ REGRESNÍHO TRIPLETU (DATA + MODEL + METODA):

Fisher-Snedocorův test významnosti regrese, F : 5.9574E+01
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 : 5.9840E-01
Závěr: Navržený model není korektní.

Cook-Weisbergův test heteroskedasticity, Sf : 1.5496E+01
Tabulkový kvantil, Chi^2(1-alpha,1) : 3.8415E+00
Závěr: Rezidua vykazují heteroskedasticitu.
Spočtená hladina významnosti : 0.000



Testování regresního tripletu

Jarque-Berraův test normality reziduí, $L(e)$: 8.0308E-02
Tabulkový kvantil, $\chi^2(1-\alpha, 2)$: 5.9915E+00

Závěr: Normalita je přijata.

Spočtená hladina významnosti : 0.961

Waldův test autokorelace, W_a : 4.2542E-03
Tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3.8415E+00

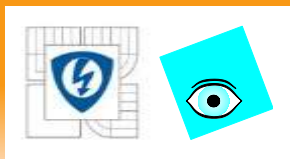
Závěr: Rezidua nejsou autokorelována.

Spočtená hladina významnosti : 0.948

Znaménkový test, D_t : 9.8058E-02
Tabulkový kvantil, $N(1-\alpha/2)$: 1.6449E+00

Závěr: Rezidua nevykazují trend.

Spočtená hladina významnosti : 0.461

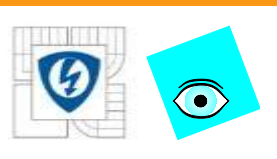


Problem 6.51 Operation of a plant for the oxidation of NH_3 to HNO_3

A set of data from 21 days of operation was collected. The y is the percentage of the input ammonia that is lost by escaping as unabsorbed nitric acid and x_1 is the rate of operation, x_2 is the temperature of the cooling water in the coils, and x_3 is the concentration of HNO_3 in the absorbing liquid. Suppose the linear model

$$E(y/x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4$$

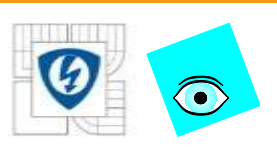
investigate the regression triplet and confirm the model proposed.



Problem 6.51 Operation of a plant for the oxidation of NH_3 to HNO_3

Data: (*) denotes the strongly influential point

Day	Stack Loss y [%]	Air Flow x_1	Temp. x_2	$[\text{HNO}_3]x_3$	Residual ε_{L1}	Concl.
1	42	80	27	89	2.53	(*)
2	37	80	27	88	$1.85 \cdot 10^{-6}$	
3	37	75	25	90	2.715	(*)
4	28	62	24	87	3.814	(*)
5	18	62	22	87	-0.61	
6	18	62	23	87	-0.89	
7	19	62	24	93	-0.50	
8	20	62	24	93	0	
9	15	58	23	87	-0.73	
10	14	58	18	80	$-2.8 \cdot 10^{-6}$	
11	14	58	18	89	0.279	
12	13	58	17	88	0.0389	
13	11	58	18	82	-1.43	
14	12	58	19	93	-0.887	
15	8	58	18	89	0.601	
16	7	50	18	86	0.0077	
17	8	50	19	72	-0.217	
18	8	50	19	79	$9.1 \cdot 10^{-5}$	
19	9	50	20	80	0.241	
20	15	56	20	82	0.812	
21	15	70	20	91	-4.722	(*)



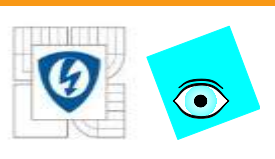
Problem 6.51 Operation of a plant for the oxidation of NH_3 to HNO_3

Program: ADSTAT

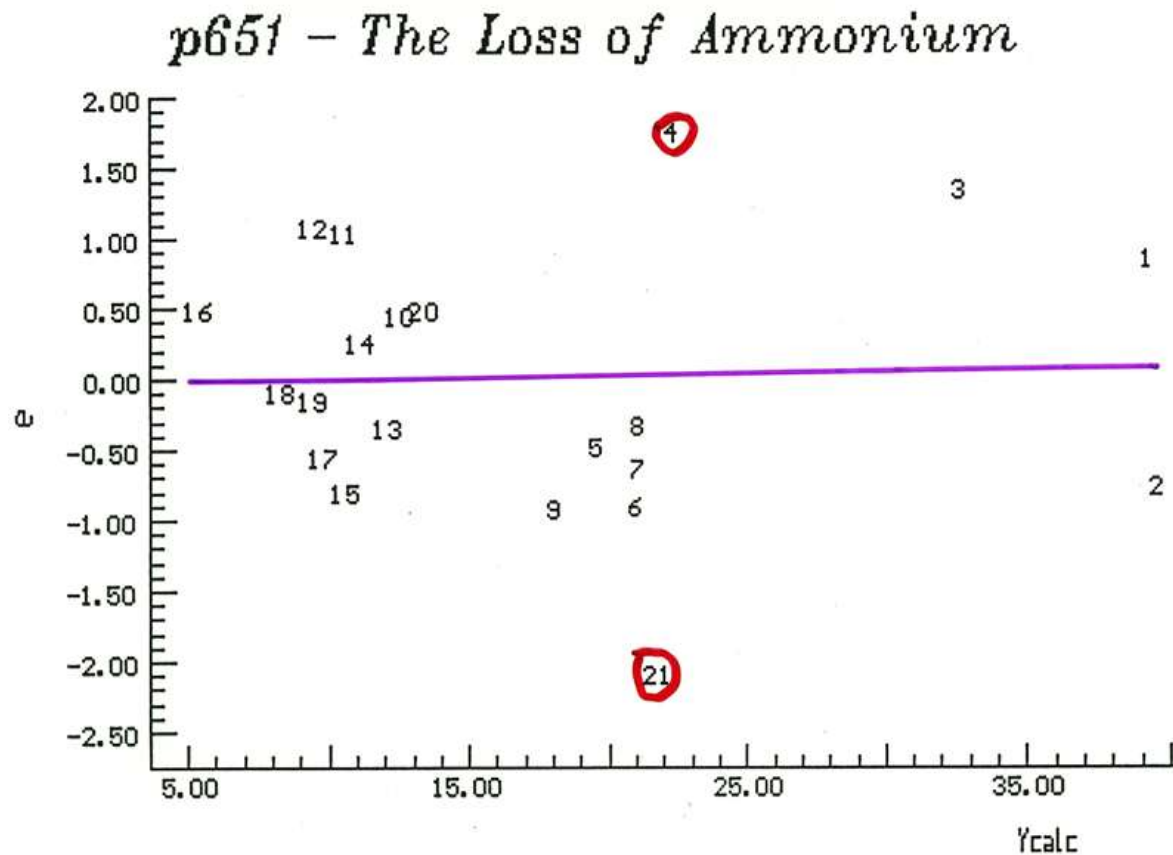
Solution: The classical LS method finds the regression equation

$$\hat{y} = -37.68 + 0.7336x_1 + 1.3883x_2 - 0.2164x_3$$

s $\hat{R}^2 = 0.913$ a $\hat{\sigma} = 3.251$.



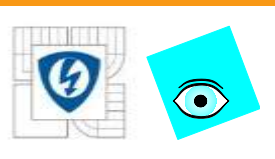
Prediction –Residual plot



31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Problem 6.51 Operation of a plant for the oxidation of NH_3 to HNO_3

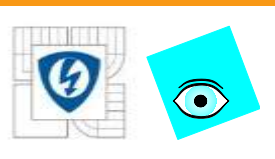
Statistical test $H_0: B[i] = 0$ vs. $H_A: B[i] \neq 0$

Original regression model:

Beta	Value	Stand. dev.	Student t	Hypothesis is	Alpha
B[0]	-3.7677E+01	1.2010E+01	-3.1372E+00	Rejected	0.006
B[1]	7.3357E-01	1.3879E-01	5.2853E+00	Rejected	0.000
B[2]	1.3883E+00	3.5650E-01	3.8943E+00	Rejected	0.001
B[3]	-2.1665E-01	1.6127E-01	-1.3434E+00	Accepted	0.197
Statistics:	R	: 9.5558E-01			
	R ²	: 9.1314E-01			
	Rp ²	: 9.2836E-01			
	MEP	: 1.3613E+01			
	AIC	: 5.3086E+01			

Modified regression model: *without outliers "4" and "21"*

Beta	Value	Stand. dev.	Student t	Hypothesis is	Alpha
B[0]	-5.2279E+01	3.8580E+00	-1.3551E+01	Rejected	0.000
B[1]	8.8786E-01	1.0619E-01	8.3607E+00	Rejected	0.000
B[2]	7.5615E-01	2.9675E-01	2.5481E+00	Rejected	0.021
Statistics:	R	: 9.7695E-01			
	R ²	: 9.5444E-01			
	Rp ²	: 9.6560E-01			
	MEP	: 6.9379E+00			
	AIC	: 3.5304E+01			



Problem 6.51 Operation of a plant for the oxidation of NH_3 to HNO_3

(2) INDIKACE MULTIKOLINEARITY:

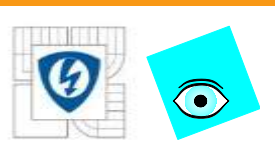
Č. [j]	Vlastní čísla korel. maticel[j]	Čísla podmí- něnosti K[j]	Variance inflation factor VIF[j]	Vícenás. korel. koef pro X[j]
1	2.1461E-01	9.9896E+00	2.8703E+00	0.8072
2	6.4148E-01	3.3421E+00	2.4019E+00	0.7640
3	2.1439E+00	1.0000E+00	1.4127E+00	0.5405
Maximální číslo podmíněnosti K			: 9.9896E+00	

(K[j], $K > 1000$ indikuje silnou multikolinearitu)

(VIF[j] > 10 indikuje silnou multikolinearitu)

(3) ODHADY PARAMETRU A TESTY VÝZNAMNOSTI:

Parametr	Odhad	Směrodatná odchylka	Test $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$ t-kriterium	hypoteza H_0 je
B[0]	-3.7677E+01	1.2010E+01	-3.1372E+00	Zamítnuta
B[1]	7.3357E-01	1.3879E-01	5.2853E+00	Zamítnuta
B[2]	1.3883E+00	3.5650E-01	3.8943E+00	Zamítnuta
B[3]	-2.1665E-01	1.6127E-01	-1.3434E+00	Akceptována



Problem 6.51 Operation of a plant for the oxidation of NH_3 to HNO_3

(4) STATISTICKÉ CHARAKTERISTIKY REGRESE:

Vícenásobný korelační koeficient, R	: 9.5558E-01
Koeficient determinace, R^2	: 9.1314E-01
Predikovaný korelační koeficient, R_p^2	: 9.2836E-01
Střední kvadratická chyba predikce, MEP	: 1.3613E+01
Akaikeho informační kritérium, AIC	: 5.3086E+01



(5) ANALÝZA KLASICKÝCH REZIDUÍ:

Bod i	Meřená hodnota $y_{exp}[i]$	Predikovaná hodnota $y_{vyp}[i]$	Směrodatná odchylka $s(y_{vyp}[i])$	Klasické reziduum $e[i]$	Relativní reziduum $er[i]$
1	4.2000E 01	3.9211E 01	1.8302E 00	2.7891E 00	6.6407E 00
2	3.7000E 01	3.9428E 01	1.8859E 00	-2.4275E 00	-6.5609E 00
3	3.7000E 01	3.2550E 01	1.3698E 00	4.4502E 00	1.2028E 01
4	2.8000E 01	2.2275E 01	1.1688E 00	5.7250E 00	2.0446E 01
5	1.8000E 01	1.9498E 01	7.4798E-01	-1.4984E 00	-8.3247E 00
6	1.8000E 01	2.0887E 01	9.1416E-01	-2.8867E 00	-1.6037E 01
7	1.9000E 01	2.0975E 01	1.5726E 00	-1.9751E 00	-1.0395E 01
8	2.0000E 01	2.0975E 01	1.5726E 00	-9.7514E-01	-4.8757E 00
9	1.5000E 01	1.7952E 01	1.2450E 00	-2.9525E 00	-1.9683E 01
10	1.4000E 01	1.2528E 01	1.4328E 00	1.4725E 00	1.0518E 01
11	1.4000E 01	1.0578E 01	1.2458E 00	3.4224E 00	2.4445E 01
12	1.3000E 01	9.4060E 00	1.4518E 00	3.5940E 00	2.7646E 01
13	1.1000E 01	1.2094E 01	1.2560E 00	-1.0942E 00	-9.9473E 00
14	1.2000E 01	1.1099E 01	1.5001E 00	9.0065E-01	7.5054E 00
15	8.0000E 00	1.0578E 01	1.2458E 00	-2.5776E 00	-3.2221E 01
16	7.0000E 00	5.3591E 00	1.2493E 00	1.6409E 00	2.3442E 01
17	8.0000E 00	9.7805E 00	2.0860E 00	-1.7805E 00	-2.2256E 01
18	8.0000E 00	8.2639E 00	1.3163E 00	-2.6391E-01	-3.2989E 00
19	9.0000E 00	9.4356E 00	1.3770E 00	-4.3556E-01	-4.8396E 00
20	1.5000E 01	1.3404E 01	9.2420E-01	1.5963E 00	1.0642E 01
21	1.5000E 01	2.1724E 01	1.6619E 00	-6.7238E 00	-4.4825E 01

Rezidualní součet čtverců, RSC	: 1.7973E 02
Průměr absolutních hodnot reziduí, Me	: 2.4372E 00
Průměr relativních reziduí, Mer	: 1.5551E 01
Odhad reziduálního rozptylu, $s^2(e)$	: 1.0572E 01
Odhad směrodatné odchylky reziduí, $s(e)$	: 3.2515E 00
Odhad šikmosti reziduí, $g1(e)$	: -1.4968E-02
Odhad špičatosti reziduí, $g2(e)$	: 2.6985E 00



(7) INDIKACE VLIVNÝCH BODŮ:

(* indikuje vlivný bod)

Bod	Standardizované reziduum	<u>Jackknife</u> <u>reziduum</u>	Predikované reziduum	<u>Diagonální</u> <u>prvky</u>
i	eS[i]	eJ[i]	eP[i]	H[i,i]
1	1.0378E 00	1.0403E 00	4.0826E 00	3.1682E-01
2	-9.1649E-01	-9.1194E-01	-3.6581E 00	3.3640E-01
3	1.5091E 00	1.5732E 00	5.4105E 00	1.7749E-01
4	1.8868E 00	2.0587E 00	6.5745E 00	1.2921E-01
5	-4.7354E-01	-4.6247E-01	-1.5822E 00	5.2919E-02
6	-9.2513E-01	-9.2099E-01	-3.1345E 00	7.9046E-02
7	-6.9403E-01	-6.8305E-01	-2.5782E 00	2.3392E-01
8	-3.4265E-01	-3.3357E-01	-1.2729E 00	2.3392E-01
9	-9.8294E-01	-9.8190E-01	-3.4597E 00	1.4661E-01
10	5.0448E-01	4.9313E-01	1.8273E 00	1.9417E-01
11	1.1395E 00	1.1503E 00	4.0112E 00	1.4681E-01
12	1.2353E 00	1.2561E 00	4.4889E 00	1.9935E-01
13	-3.6484E-01	-3.5534E-01	-1.2861E 00	1.4921E-01
14	3.1221E-01	3.0376E-01	1.1442E 00	2.1285E-01
15	-8.5825E-01	-8.5127E-01	-3.0212E 00	1.4681E-01
16	5.4663E-01	5.3503E-01	1.9252E 00	1.4763E-01
17	-7.1385E-01	-7.0315E-01	-3.0259E 00	4.1158E-01*
18	-8.8765E-02	-8.6135E-02	-3.1564E-01	1.6389E-01
19	-1.4787E-01	-1.4355E-01	-5.3075E-01	1.7934E-01
20	5.1207E-01	5.0066E-01	1.7366E 00	8.0790E-02
21	-2.4059E 00	-2.8741E 00*	-9.1015E 00	2.6124E-01

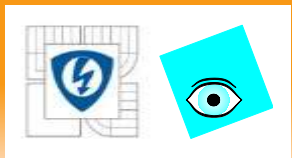


Bod i	Zobecněné diag. prvky Hm[i,i]	Cookova vzdálenost D[i]	Atkinsonova vzdálenost A[i]	Vliv na predikci DF[i]
1	3.6010E-01	1.2487E-01	1.4605E 00	7.0844E-01
2	3.6918E-01	1.0645E-01	1.3385E 00	-6.4929E-01
3	2.8768E-01	1.2286E-01	1.5066E 00	7.3082E-01
4	3.1157E-01	1.3207E-01	1.6349E 00	7.9304E-01
5	6.5412E-02	3.1324E-03	2.2536E-01	-1.0932E-01
6	1.2541E-01	1.8365E-02	5.5625E-01	-2.6982E-01
7	2.5562E-01	3.6769E-02	7.7811E-01	-3.7744E-01
8	2.3921E-01	8.9624E-03	3.7999E-01	-1.8432E-01
9	1.9511E-01	4.1496E-02	8.3901E-01	-4.0698E-01
10	2.0623E-01	1.5331E-02	4.9903E-01	2.4206E-01
11	2.1197E-01	5.5856E-02	9.8366E-01	4.7715E-01
12	2.7122E-01	9.4986E-02	1.2922E 00	6.2679E-01
13	1.5587E-01	5.8358E-03	3.0677E-01	-1.4881E-01
14	2.1737E-01	6.5895E-03	3.2564E-01	1.5796E-01
15	1.8377E-01	3.1686E-02	7.2797E-01	-3.5312E-01
16	1.6261E-01	1.2938E-02	4.5903E-01	2.2266E-01
17	4.2922E-01	8.9110E-02	1.2124E 00	-5.8808E-01
18	1.6428E-01	3.8611E-04	7.8617E-02	-3.8135E-02
19	1.8040E-01	1.1946E-03	1.3834E-01	-6.7106E-02
20	9.4968E-02	5.7616E-03	3.0599E-01	1.4843E-01
21	5.1278E-01* σ	5.1172E-01* σ	3.5234E 00* σ	-1.7091E 00* σ



Cookova-Weisbergova věrohodnostní vzdálenost

i	LD(b)[i]	LD(s ²)[i]	LD(b,s ²)[i]
1	6.0810E-01	2.9307E-03	6.3029E-01
2	5.1950E-01	3.7178E-05	5.2700E-01
3	5.9848E-01	9.7927E-02	7.6555E-01
4	6.4264E-01	3.8771E-01	1.1738E 00
5	1.5472E-02	1.3086E-02	2.8024E-02
6	9.0549E-02	8.6374E-05	9.1091E-02
7	1.8090E-01	4.1918E-03	1.8227E-01
8	4.4238E-02	1.8145E-02	6.0614E-02
9	2.0404E-01	9.9571E-04	2.0804E-01
10	7.5617E-02	1.1796E-02	8.5039E-02
11	2.7419E-01	9.9753E-03	2.9456E-01
12	4.6418E-01	2.1837E-02	5.1291E-01
13	2.8816E-02	1.7352E-02	4.5032E-02
14	3.2534E-02	1.9174E-02	5.0362E-02
15	1.5598E-01	2.1176E-04	1.5607E-01
16	6.3832E-02	1.0024E-02	7.1998E-02
17	4.3576E-01	3.5165E-03	4.3582E-01
18	1.9077E-03	2.4132E-02	2.5950E-02
19	5.9021E-03	2.3323E-02	2.8952E-02
20	2.8450E-02	1.1478E-02	3.9016E-02
21	2.3875E 00	1.6087E 00	5.2601E 00



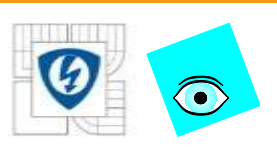
Kapitola 6.7

KALIBRACE

31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Skládá se z fází:

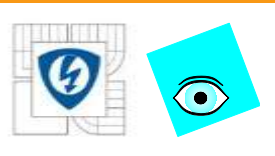
- a) sestavení kalibračního modelu,
- b) použití kalibračního modelu.

Druhy kalibrace a kalibrační metody

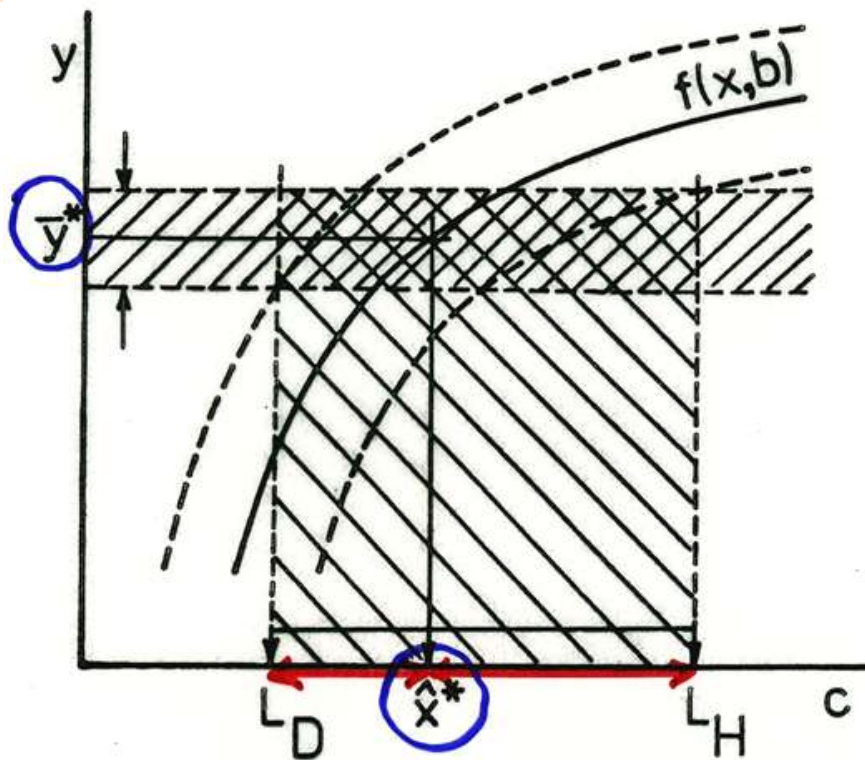
1. **Absolutní kalibrace**: vztah signálem y a veličinou x , kde

y je měřený signál (potenciál E , napětí článku U , proud I , elektrický odpor R , pH, absorbance A , atd.) a

x je stav nebo vlastnost systému (obsah m , objem V , koncentrace c , teplota t , čas τ , atd.)



Kalibrace

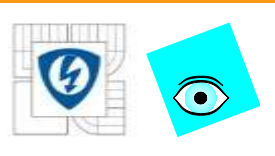


2. **Komparativní kalibrace**: přístroj se kalibruje vůči druhému, např. absorbance (1. metoda) a titračně (2. metoda)



Dle využití kalibračního modelu

1. **Jedno použití:** z modelu o bodech $\{x_i, y_i\}, i = 1, n$, se určí odhad $\hat{x}^*$ se svým intervalem spolehlivosti,
2. **Vícenásobné použití:** pro různé hodnoty signálů se určují odhady $\hat{x}^*$ na základě jednoho kalibračního modelu,
3. **Použití v kombinaci:** s dalšími měřeními.



Hledání kořene rovnice $\hat{x}^* = f^{-1}(y^*)$

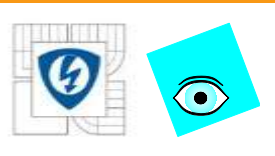
Na základě Taylorova rozvoje lze nalézt pro rozptyl

$$D(\hat{x}^*) \approx \left[\frac{\partial f(x, b)}{\partial x} \right]^{-2} \left[\frac{D(y^*)}{M} + D(f(x, b)) \right]$$

kde $D(y^*)$ je rozptyl y^* ových hodnot, obvykle σ^2 ,
 $D(f(x, b)) = D(\hat{y})$ je rozptyl predikce.

Pro lineární regresní modely je rozptyl predikce roven

$$\begin{aligned} D(\hat{y}_P) &= \sigma^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \\ &= \sigma^2 \left[\frac{1}{n} + \frac{(y^* - \bar{y})^2}{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2} \right] \end{aligned}$$



Hledání kořene rovnice $\hat{x}^* = f^{-1}(y^*)$

kde b_1 představuje odhad směrnice regresní přímky.

Po dosazení vyjde

$$D(\hat{x}^*) \approx \frac{\sigma^2}{b_1^2} \left[\frac{1}{M} + \frac{1}{n} + \frac{(y^* - \bar{y})^2}{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

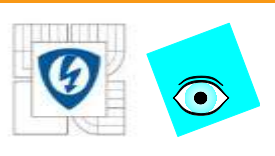
Speciálně pro model kali brační přímky

$$y = b_1(x - \bar{x}) + \bar{y}$$

a dosazením vyjde

$$D(\Delta) = \hat{\sigma}^2 \left[\frac{1}{M} + \frac{1}{n} + \frac{(\hat{x}^* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

Při znalosti $D(\Delta)$ lze již nalézt krajní meze
konfidenčního intervalu pro $\hat{x}^*$.



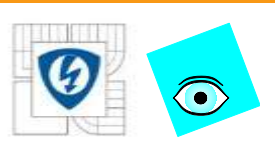
Kalibrační přímka

V laboratořích nejpoužívanější model, (např. Lambertův-Beerův zákon, $A = \varepsilon dc$)

Kalibrační přímka: $y_i = \beta_2 + \beta_1 x + \varepsilon_i, \quad i = 1, \dots, n$

Neznámý vzorek: $y_j^* = \beta_2 + \beta_1 \kappa + \varepsilon_j^*, \quad i = 1, \dots, M$

nalezení odhadu $\hat{x}^*$ parametru κ



Odhad $\hat{x}^*$

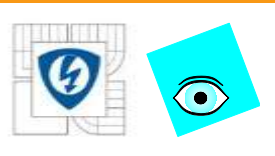
(1) **Přímý odhad:** $\hat{x}^* = \bar{x} + \frac{y^* - \bar{y}}{b_1}$

kde y^* je měřený signál (resp. průměr $\bar{y}^*$ pro $M > 1$ opakovaných měření), b_1 je odhad směrnice kalibrační přímky.

Přímý odhad je obecně vychýlený.

(2) **Naszodiho modifikovaný odhad:** korekce na vychýlení

$$\hat{x}_B^* = \bar{x} + \frac{(y^* - \bar{y})b_1}{b_1^2 + \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$



(3) **Krutchhoftův inverzní odhad:**

$$\hat{x}_I^* = \bar{x} + (y^* - \bar{y}) \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

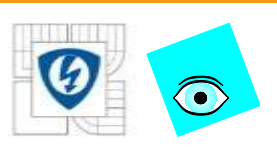
který vychází z inverzního regresního modelu

$$E(x/y) = \alpha_1(y - \bar{y}) + \alpha_2$$

jde také o vychýlený odhad.

(4) **Schwartzův nelineární odhad:** (předpoklad normality reziduí)

$$\hat{x}_N^* = \frac{\sum_{i=1}^n x_i \exp \left[\frac{-(y^* - b_2 - b_1 x_i)^2}{2\hat{\sigma}^2} \right]}{\sum_{i=1}^n \exp \left[\frac{-(y^* - b_2 - b_1 x_i)^2}{2\hat{\sigma}^2} \right]}$$



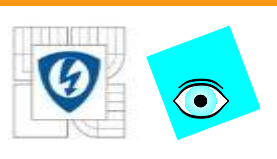
Příklad 6.53 Bodové odhady koncentrace metodou AAS

Metodou AAS byla změřena absorbance roztoků o různé koncentraci lithia. Z kalibrační přímky odhadněte koncentraci lithia pro $A_1 = 0.0002$, $A_2 = 0.5$, $A_3 = 1.0$.

Data: $n = 16$

A	0.063	0.120	0.189	0.251	0.316	0.393	0.442	0.502
c [g Li (25 cm) ⁻³]	2.5	5.0	7.5	10.0	12.5	15.0	17.5	20.0

A	0.568	0.639	0.694	0.749	0.821	0.884	0.947	1.010
c [g Li (25 cm) ⁻³]	22.5	25.0	27.5	30.0	32.5	35.0	37.5	40.0



Příklad 6.53 Bodové odhady koncentrace metodou AAS

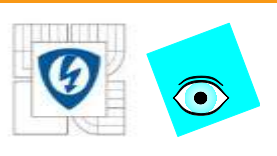
Řešení: Klasickou metodou MNČ určena

$$A = 0.02525(\pm 1.138 \cdot 10^{-4}) + 0.0002(\pm 2.753 \cdot 10^{-3})$$

s korelačním koeficientem $\hat{R} = 0.9999$.

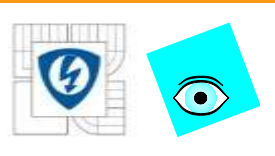
Odhady koncentrace rozličnými metodami

Absorbance	$\hat{x}^*$	$\hat{x}_B^*$	$\hat{x}_I^*$	$\hat{x}_N^*$
$2 \cdot 10^{-4}$	0	$4.32 \cdot 10^{-4}$	$6.0499 \cdot 10^{-3}$	2.5
0.5	19.795	19.795	19.795	20
1.0	39.597	39.597	39.592	40



Příklad 6.53 Bodové odhady koncentrace metodou AAS

Závěr: Pro přesná data s malým rozptylem kolem regresní přímky postačuje jednoduchý klasický odhad $\hat{x}^*$.



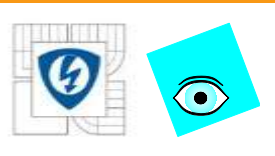
Interval spolehlivosti odhadu $\hat{x}^*$

$P(L_D < \kappa < L_H) = 1 - \alpha$, kde α je koeficient významnosti a $1 - \alpha$ je statistická pravděpodobnost, statistická jistota.

S $(1 - \alpha)$ ní statistickou jistotou tvrdíme, že „pravda κ “ leží v intervalu $\langle L_D; L_H \rangle$.

Meze 95%ního intervalu spolehlivosti se vypočtou

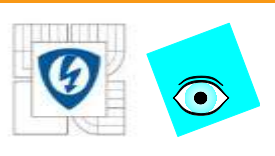
$$L_{D,H} = \hat{x}^* \pm \text{i. s.}$$



Pás spolehlivosti y hodnot

Udává rozpětí, ve kterém se budou v základním souboru nacházet hodnoty závisle proměnné se zvolenou pravděpodobností $1 - \alpha$

$$y_{i(min,max)} = y'_i \pm t_{\frac{\alpha}{2}; n-m} \cdot \sigma$$



IS modelových hodnot

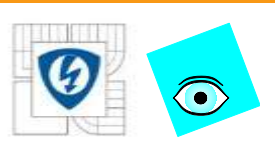
Pro model přímky

Směrodatná odchylka reziduí

$$\mu_{y'} = y'_i \pm t_{\frac{\alpha}{2}; n-m} \cdot \frac{\sigma}{\sqrt{n-2}} \cdot \sqrt{1 + \frac{n(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Modelová hodnota

Polovina IS modelu přímky



IS modelových hodnot

1. Při asymptotické normalitě:

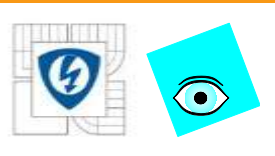
$$\text{dolní mez } L_D = \hat{x}^* - 1.96\sqrt{D(\hat{x}^*)}$$

$$\text{horní mez } L_H = \hat{x}^* + 1.96\sqrt{D(\hat{x}^*)}$$

2. Pro méně přesná data:

$$L_{D,H} = \bar{x} + \frac{(y^* - \bar{y}) \mp \hat{\sigma}\sqrt{F_{1-\alpha}(1, n-2)} \left[\frac{1+\lambda}{n} + \frac{(y^* - \bar{y})^2}{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2} \right]^{\frac{1}{2}}}{b_1(1-\lambda)}$$

kde poměr $\lambda = \frac{\hat{\sigma}^2 F_{1-\alpha}(1, n-2)}{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}$ je variační koeficient parametru β_1 .

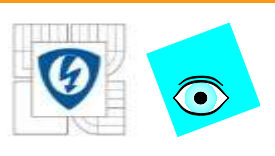


3. **Odhad směrnice přesný:** pro $\lambda < 0.1$ je odhad směrnice dostatečně přesný a lze použít

$$L_{D,H} = x^* \mp t_{1-\frac{\alpha}{2}}(n-2) \frac{\hat{\sigma}}{|b_1|} \sqrt{\frac{1}{n} + \frac{(y^* - \bar{y})^2}{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}}$$

Scheffého interval spolehlivosti predikované hodnoty y^* při $\hat{x}^*$ platí

$$R_{D,H} = y^* \mp \sqrt{2F_{1-\alpha}(2, n-2)} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(\hat{x}^* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$



IS modelových hodnot

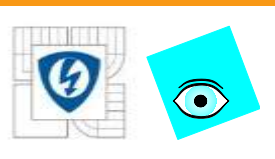
Úpravou dostaneme $100(1 - \alpha)$ ní interval
spolehlivosti x^*

$$L_{D,H} \\ = \hat{x}^*$$

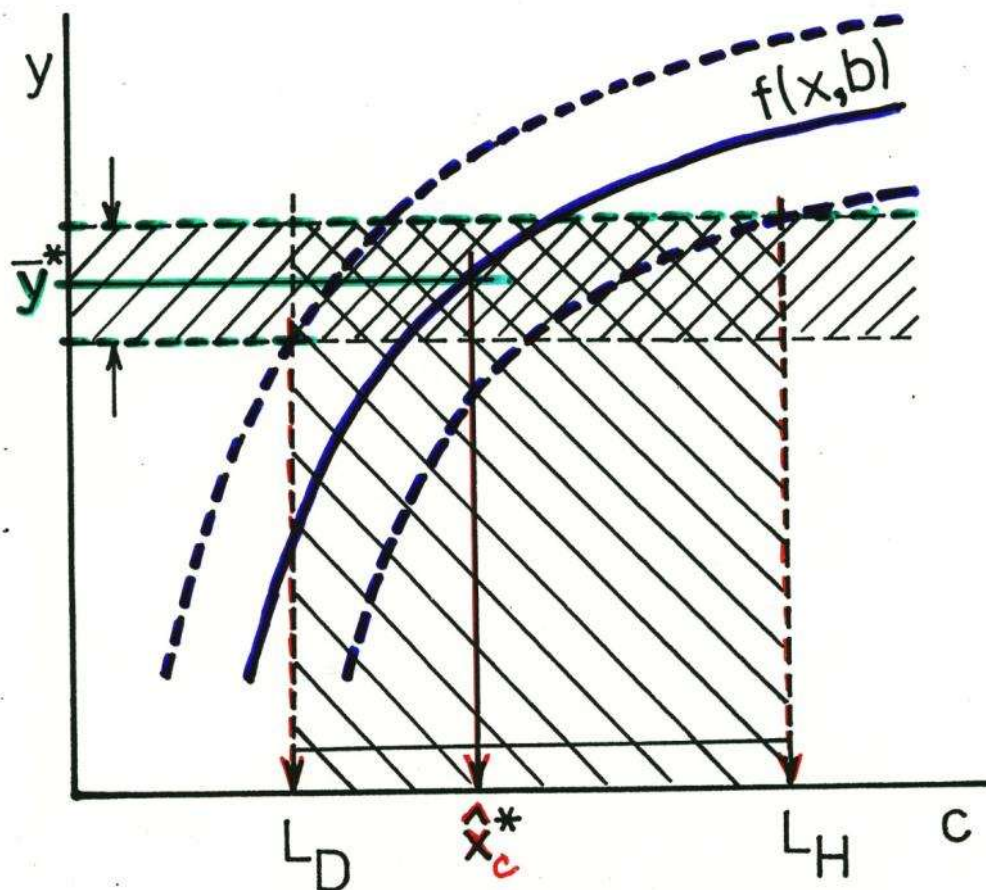
$$\pm \frac{\hat{\sigma} \sqrt{2F_{1-\alpha}(2, n-2)}}{\lambda_1} \sqrt{\lambda_1 \left(1 - \frac{1}{n}\right) + \frac{(y^* - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

kde

$$\lambda_1 = b_1^2 - \frac{\sqrt{2F_{1-\alpha}(2, n-2)}}{\sum_{i=1}^n (x_i - \bar{x})^2}$$



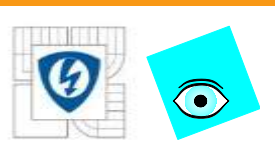
IS modelových hodnot



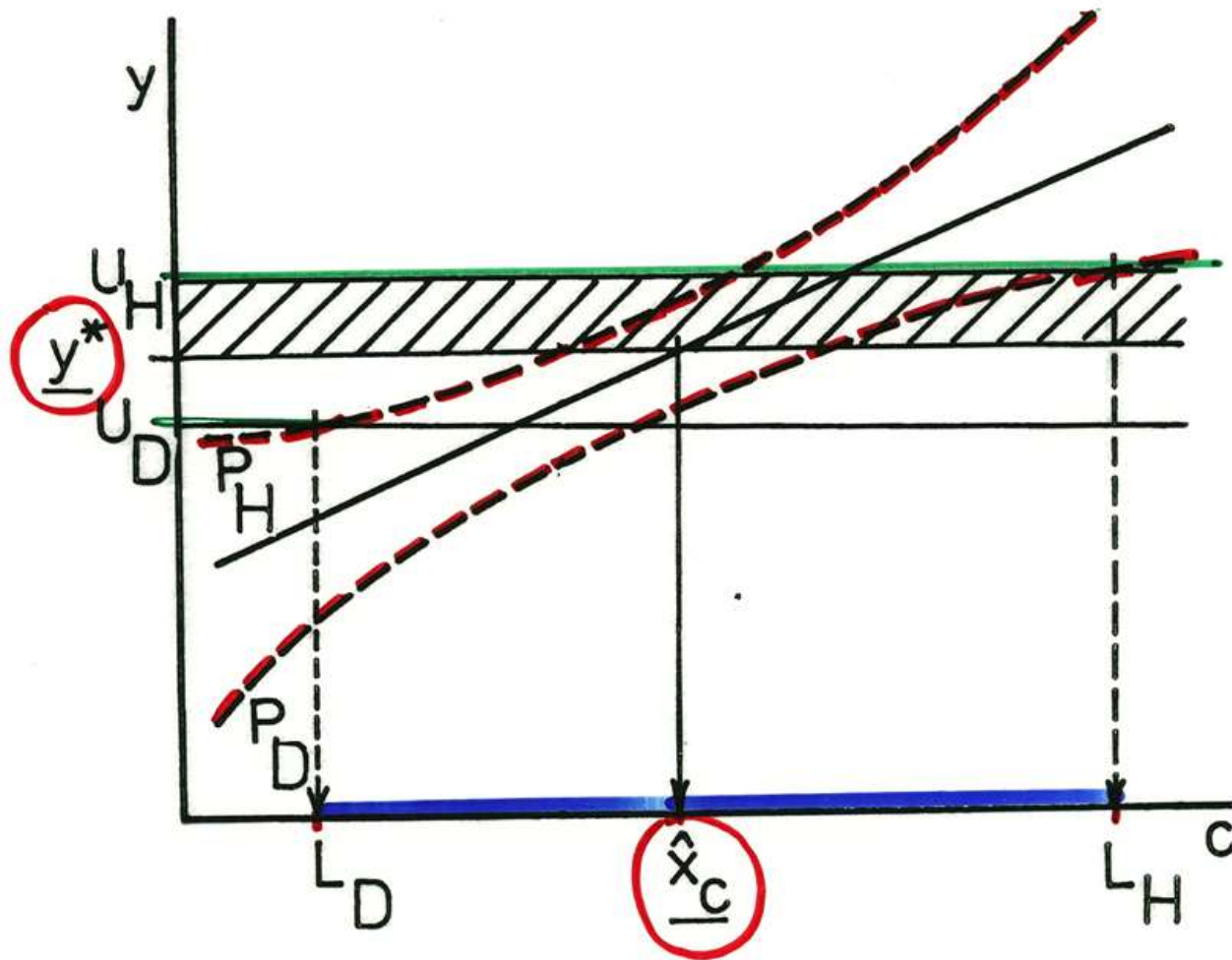
31.1.2011

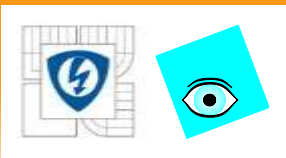
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Konstrukce intervalu spolehlivosti $\hat{x}^*$ z kalibrační přímky





Konstrukce intervalu spolehlivosti $\hat{x}^*$ z kalibrační přímky

100(1 - α)%ní interval spolehlivosti pro signál y^*

$$U_{D,H} = \bar{y}^* \mp u_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{M}} \sqrt{\frac{n-2}{\chi_{\frac{\alpha}{2}}^2(n-2)}}$$

a pro regresní přímku hraniční 100(1 - α)%ní paraboloidy

$$P_{D,H} = b_1 x + b_2 \mp \hat{\sigma} \{ 2F_{1-\alpha}(2, n-2) \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \}$$



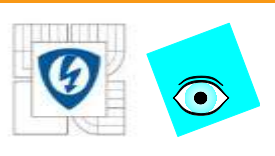
Konstrukce intervalu spolehlivosti $\hat{x}^*$ z kalibrační přímky

Dolní hraniční hodnota L_D je řešením rovnice $U_H = P_D$

Horní hraniční hodnota L_H je řešením rovnice $U_D = P_H$

Diskuze: Příliš velký rozptyl v datech, (kalibrační přímka není vhodná) způsobuje:

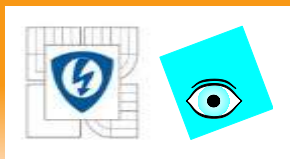
1. v některých případech nemusí průsečík přímky s parabolou existovat,
2. v jiném případě protne konfidenční přímka signálu parabolou kalibrační přímky ve dvou bodech.



Vlastnosti intervalu spolehlivosti

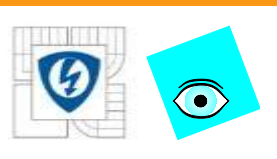
Zúžení intervalu spolehlivosti neznámé $\hat{x}^*$ lze způsobit

1. opakováním měření signálu y^* , čili růstem M ,
2. zúžením konfidenčních parabol lze dosáhnout eliminací vlivných bodů,
3. zmenšením rozptylu $\hat{\sigma}^2$, tj. zpřesněním měření.



Příklad 6.54 Intervaly spolehlivosti koncentrace z kalibrační přímky

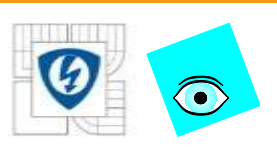
Pro kalibrační data Lambert.-Beerova zákona vypočtete 95%ní interval spolehlivosti koncentrace pro absorbance $A_1^* = 0.0002$, $A_2^* = 0.5$, $A_3^* = 1.0$, $A^* = 0.51$ (vzniklé jako průměr z 0.5 a 0.52, kde $M = 2$), $A^* = 0.977$ (vzniklé jako průměr z 0.95, 0.98, 1.00, kde $M = 3$)



Příklad 6.54 Interval spolehlivosti koncentrace z kalibrační přímky

Interval spolehlivosti stanovené koncentrace, vyčíslený z rozličných vztahů

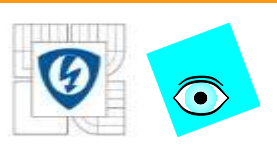
		Užitý vztah pro interval spolehlivosti L_1				
M	A^*	(6.218)	(6.220)	(6.221)	(6.223)	(6.227)
		dolní L_D	L_D	L_D	L_D	L_D
		horní L_H	L_H	L_H	L_H	L_H
1	0.0002	-0.46	-0.3018	-0.299	-0.646	-0.398
		0.46	0.295	0.299	0.639	0.417
1	0.50	19.37	19.65	19.65	19.20	19.39
		20.22	19.94	19.94	19.94	20.20
2	(0.50; 0.52)	19.89	20.05	20.05	19.76	19.90
		20.50	20.33	20.33	20.62	20.48
1	1.00	39.16	39.33	39.33	38.97	39.19
		40.05	39.87	39.87	40.23	40.01
3	(0.95; 0.98; 1.00)	38.37	38.42	38.42	38.26	38.43
		38.97	38.93	38.93	39.09	38.91



Příklad 6.54 Intervaly spolehlivosti koncentrace z kalibrační přímky

Řešení:

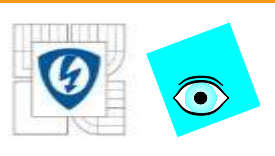
- a) I. s. z (6.220) a (6.221) neberou větší přesnost stanovení $\bar{y}^*$.
- b) Při výpočtu $U_{D,H}$ dle (6.227a,b) byl využit výraz pro konfidenční přímky signálu $U_{D,H} = \bar{y}^* \mp \frac{u_{1-\frac{\alpha}{2}} \hat{\sigma}}{\sqrt{M}}$ při volbě $\alpha = 0.05$.
- c) Vzhledem k dostatečné přesnosti dat poskytuje (6.220) i (6.221) stejné výsledky. Ostatní jsou však odlišné.
- d) Aproximace (6.218) poskytuje blízké hodnoty i.s. vzhledem k hodnotám určeným z (6.227).



Příklad 6.54 Intervaly spolehlivosti koncentrace z kalibrační přímky

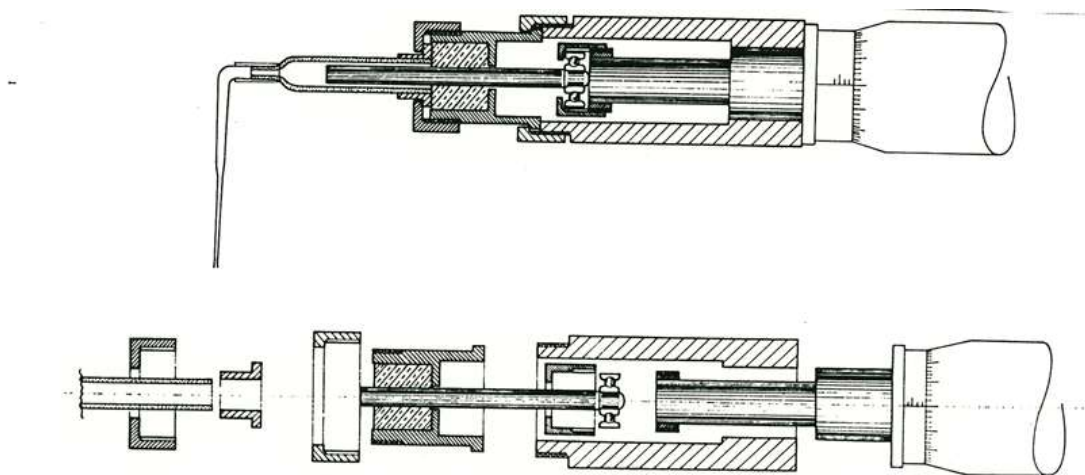
Závěr:

1. Pro data s malým rozptylem kolem regresní přímky je výhodné použít jednodušší aproximace (6.218).
2. U opakovaných měření signálu nejsou vhodné rovnice (6.220) a (6.221).



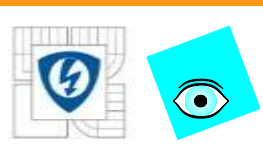
Super – přesné dávkování

1. Koncentrovaných anorganických kyselin a zásad,
2. Organických rozpouštědel

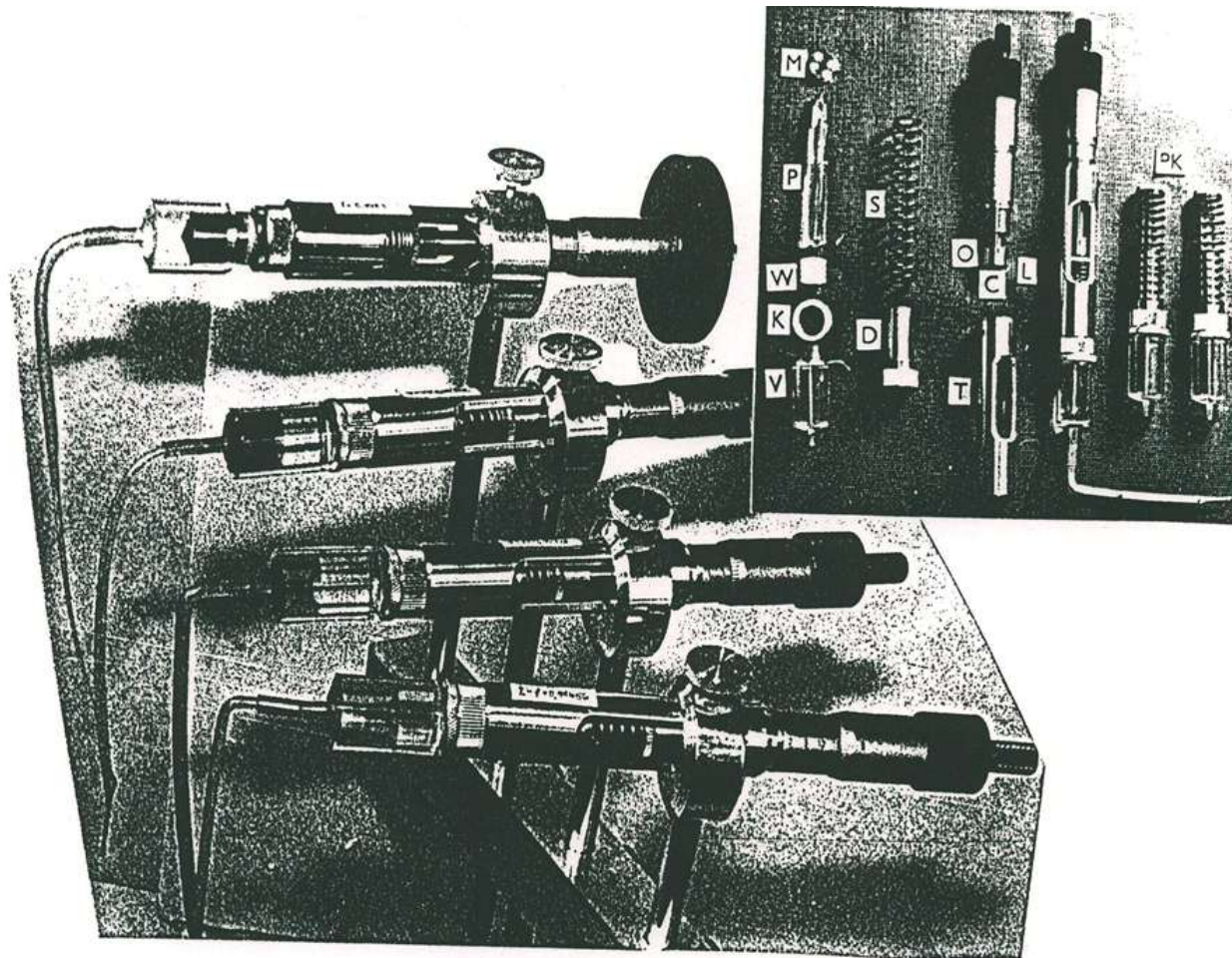


Pístová mikrobyreta (M. Meloun, V. Říha, J. Žáček: ZN. 11/1985)

PK2500:	$V = 2500 \mu\text{l}$,	$s(V) = 0.23 \mu\text{l}$,	$s_{\text{rel}}(V) = 0.012 \%$
PK1250:	$V = 1250 \mu\text{l}$,	$s(V) = 0.22 \mu\text{l}$,	$s_{\text{rel}}(V) = 0.024 \%$
PK0500:	$V = 500 \mu\text{l}$,	$s(V) = 0.25 \mu\text{l}$,	$s_{\text{rel}}(V) = 0.050 \%$
PK0250:	$V = 250 \mu\text{l}$,	$s(V) = 0.25 \mu\text{l}$,	$s_{\text{rel}}(V) = 0.100 \%$



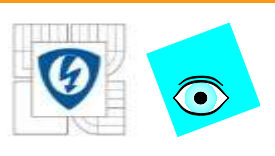
Stavebnice pístové mikrobyrety



31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

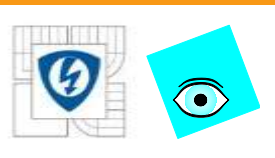




Limitní hodnoty, pro které je ještě signál odlišný od šumu.

1. **Kritická úroveň y_c** : představuje horní mez 100(1 – α)%ního intervalu spolehlivosti predikce signálu pro koncentraci rovnou nule, tzv. slepý pokus. Nad hodnotou y_c lze signál odlišit od šumu.

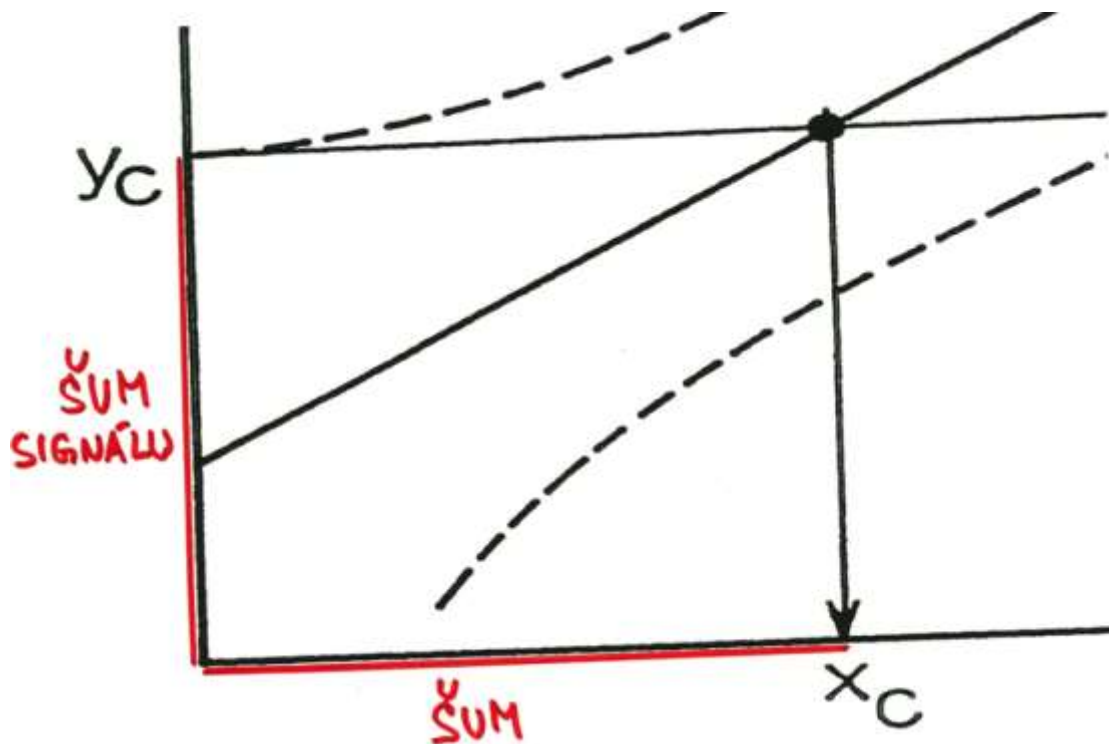
$$y_c = \bar{y} - b_1 \bar{x} + t_{1-\frac{\alpha}{2}}(n-2)\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

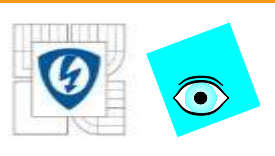


Přesnost kalibrace

Koncentrace x_c , odpovídající hodnotě kritické úrovně

$$x_c = \frac{y_c - \bar{y}}{b_1} + \bar{x}$$

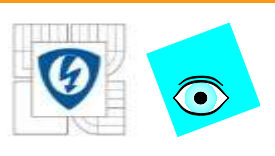




2. **Limita detekce** y_D : odpovídá koncentraci, pro kterou je dolní mez $100(1 - \alpha)$ ního intervalu spolehlivosti predikce signálu rovna y_c . Limita detekce udává úroveň signálu, která umožňuje ještě detekci koncentrace, tj. odlišení od šumu.

$$y_D = y_c + \hat{\sigma} t_{1-\frac{\alpha}{2}}(n-2) \sqrt{1 + \frac{1}{n} + \frac{(x_D - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$



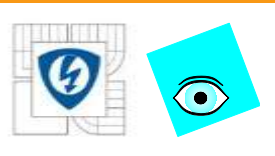


Limita detekce

koncentrace limity detekce x_D se vypočte

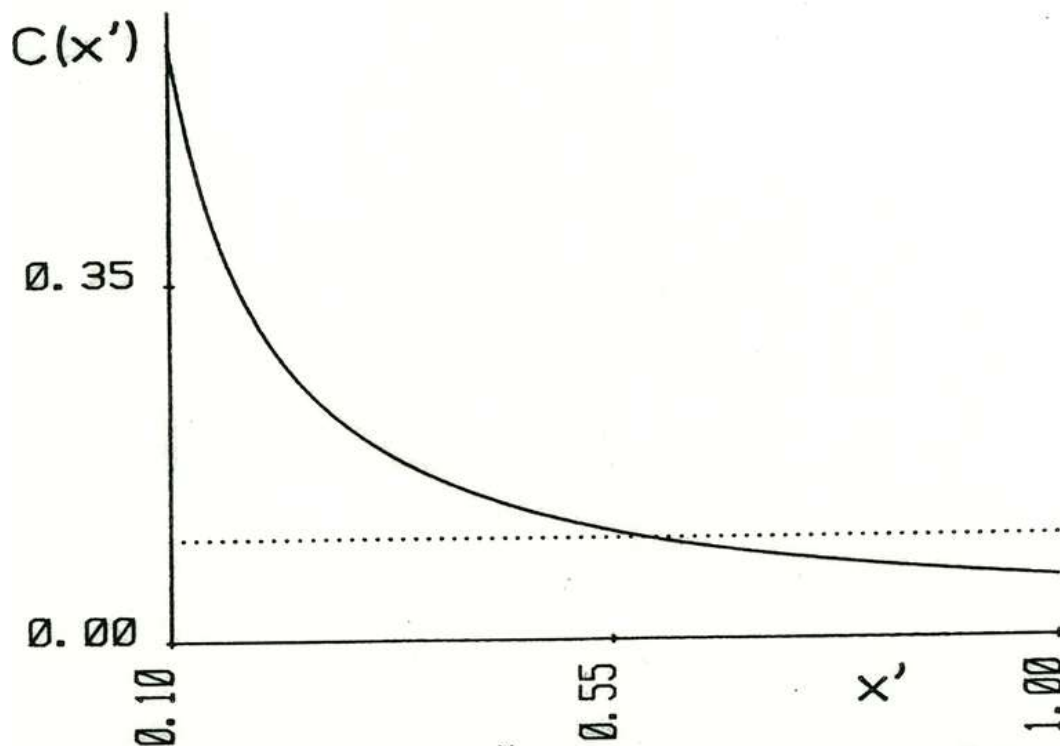
$$x_D = \frac{y_D - \bar{y}}{b_1} + \bar{x}$$

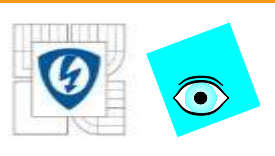
x_D udává minimální koncentraci, kterou lze s pravděpodobností $(1 - \alpha)$ ještě odlišit od nulové hodnoty.



Modifikovaná limita stanovení

4. **Modifikovaná limita stanovení** y'_s je hodnota, pro kterou je $C(x') = C$. Hledání limity stanovení. Čárkovaně je úroveň 0.1.

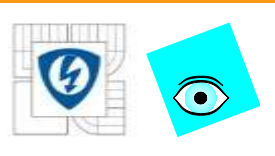




Přehled limit přesnosti kalibrace

Obecně platí pravidlo, že

$$y_c \leq y_D \leq y_s$$



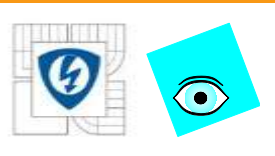
3. **Limita stanovení** y_s je nejmenší hodnota signálu, pro kterou je relativní směrodatná odchylka predikce dodatečně rovna číslu C . (C se volí 10%, tj. $C=0.1$).

Je-li predikce v místě x_s rovna $y(x_s) = \bar{y} + b_1(x_s - \bar{x})$, je podmínka k určení y_s

$$\frac{\sqrt{D(y(x_s))}}{\hat{y}(x_s)} = C$$

Dosazením

$$y_s = \frac{\hat{\sigma}}{C} \sqrt{1 + \frac{1}{n} + \frac{(x_s - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$



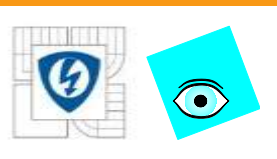
Limita stanovení

K praktickým výpočtům se užívá aproximace

$$y_s \approx \frac{\hat{\sigma}}{C} \sqrt{1 + \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

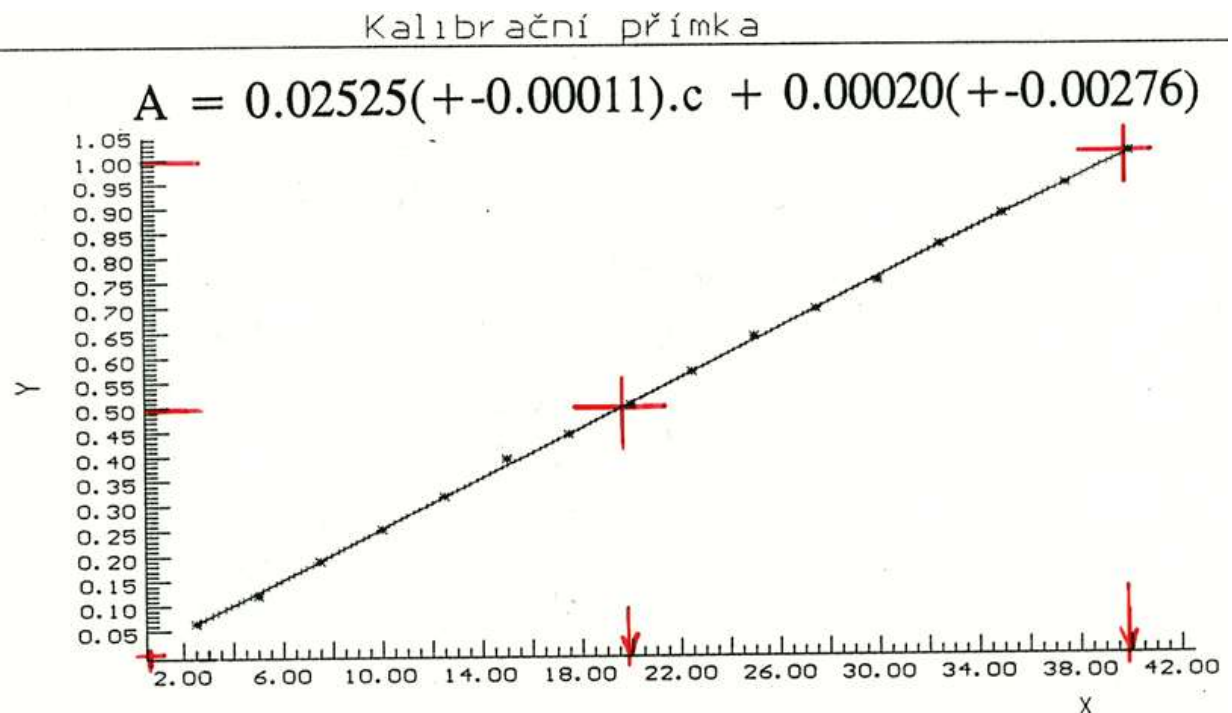
a koncentrace limity stanovení x_s je

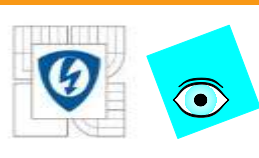
$$x_s = \frac{y_s - \bar{y}}{b_1} + \bar{x}$$



Příklad 6.53 Odhady koncentrace z kalibrační přímky AAS

Odhadněte z kalibrační přímky koncentraci lithia [g Li/25cm³] pro naměřené absorbance $A^* = 0.0002, 0.5, 1.0$.





Výstupy

O U T P U T

(2) CALIBRATION PARAMETERS:

Parameter	Estimate	Standard deviation	Test of H0: B[j] = 0 vs. HA: B[j] <> 0 t-Criterion	H0 hypothesis is	Sig.level
Intercept	2.0000E-04	2.7529E-03	7.2650E-02	Accepted	0.943
Slope	2.5249E-02	1.1388E-04	2.2172E+02	Rejected	0.000

(3) ANALYSIS OF RESIDUALS:

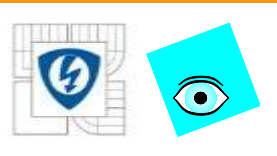
Residual sum of squares, RSS	: 3.8581E-04
Mean of absolute values of residuals, Me	: 2.7588E-03
Mean of relative residuals, Mer[%]	: 0.855
Estimate of residual variance, s ² (e)	: 2.7558E-05
Estimate of residual standard deviation, s(e)	: 5.2496E-03

(4) CALIBRATION LIMITS:

Critical level,	yc:	6.1044E-03	xc:	2.3384E-01
Detection limit,	yd:	1.1909E-02	xd:	4.6375E-01
Determination limit,	ys:	5.9276E-02	xs:	2.3397E+00

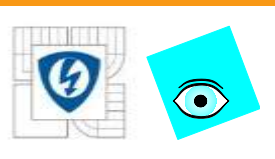
(5) CALIBRATION TABLE:

Response measured	Direct estimation	Naszodi estimation	Confidence interval	
yexp[i]	xcal[i]	xcal[i]	lower limit	upper limit
2.0000E-04	9.4602E-08	4.3235E-04	-5.0351E-01	5.0351E-01
5.0000E-01	1.9795E+01	1.9795E+01	1.9335E+01	2.0254E+01
1.0000E+00	3.9597E+01	3.9597E+01	3.9104E+01	4.0090E+01

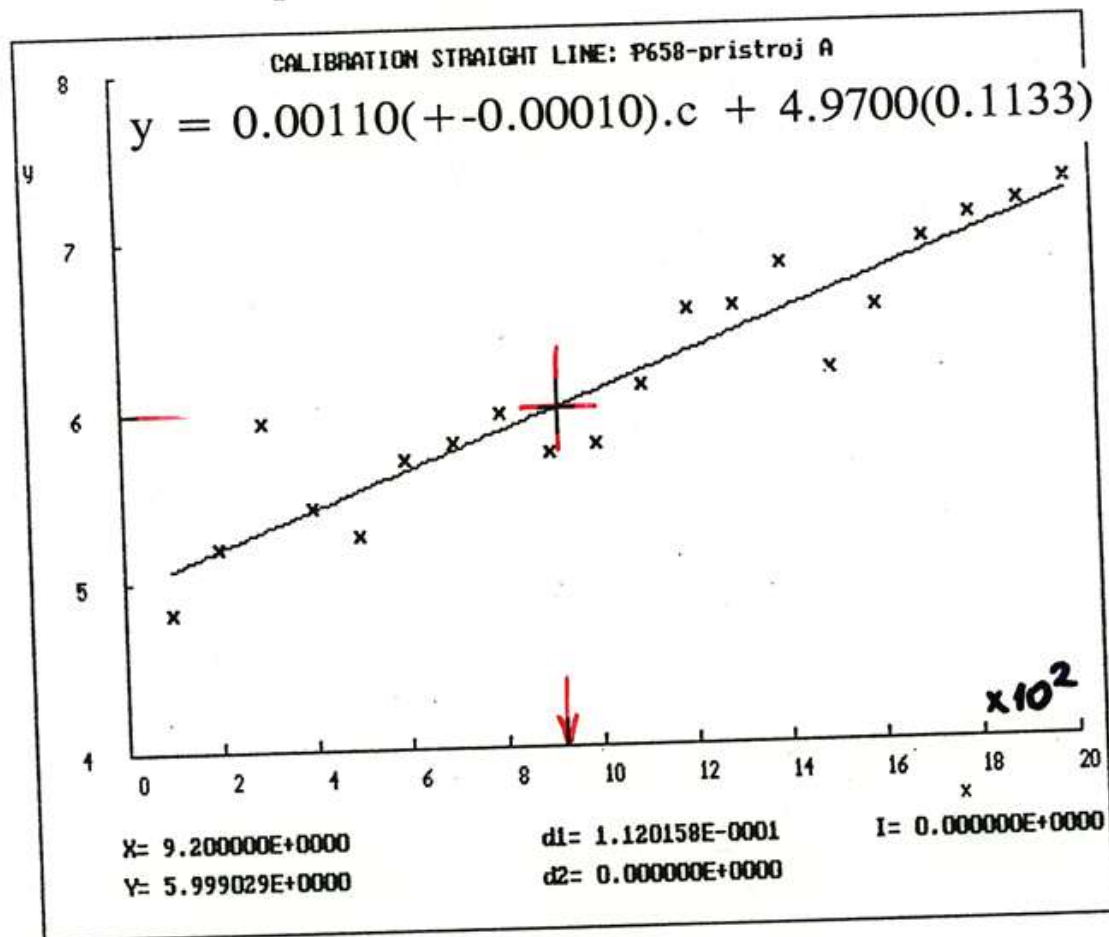


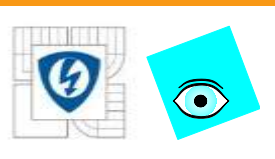
Příklad 6.58 Vliv přesnosti přístroje na mez detekce

Na třech rozličně přesných přístrojích A, B, a C byly změřeny signály pro 20 úrovní koncentrace. Určete regresní model a přesnost kalibrace pomocí meze detekce a kritické úrovně.



Příklad 6.58 Vliv přesnosti přístroje na mez detekce





Příklad 6.58 Vliv přesnosti přístroje na mez detekce

(2) CALIBRATION PARAMETERS:

Parameter	Estimate	Standard deviation	Test of $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$	t-Criterion	H_0 hypothesis is	Sig. level
Intercept	4.9700E+00	1.1334E-01		4.3852E-01	Rejected	0.000
Slope	1.1045E-03	9.4611E-05		1.1674E-01	Rejected	0.000

(3) ANALYSIS OF RESIDUALS:

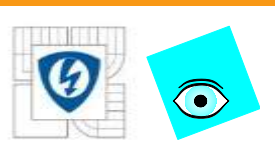
Residual sum of squares, RSS	: 1.0715E+00
Mean of absolute values of residuals, Me	: 1.7367E-01
Mean of relative residuals, Mer[%]	: 2.896
Estimate of residual variance, $s^2(e)$	: 5.9526E-02
Estimate of residual standard deviation, $s(e)$	: 2.4398E-01

(4) CALIBRATION LIMITS:

Critical level,	yc:	5.2082E+00	xc:	2.1558E+02
Detection limit,	yd:	5.3846E+00	xd:	3.7530E+02
Determination limit,	ys:	2.6902E+00	xs:	-2.0641E+03

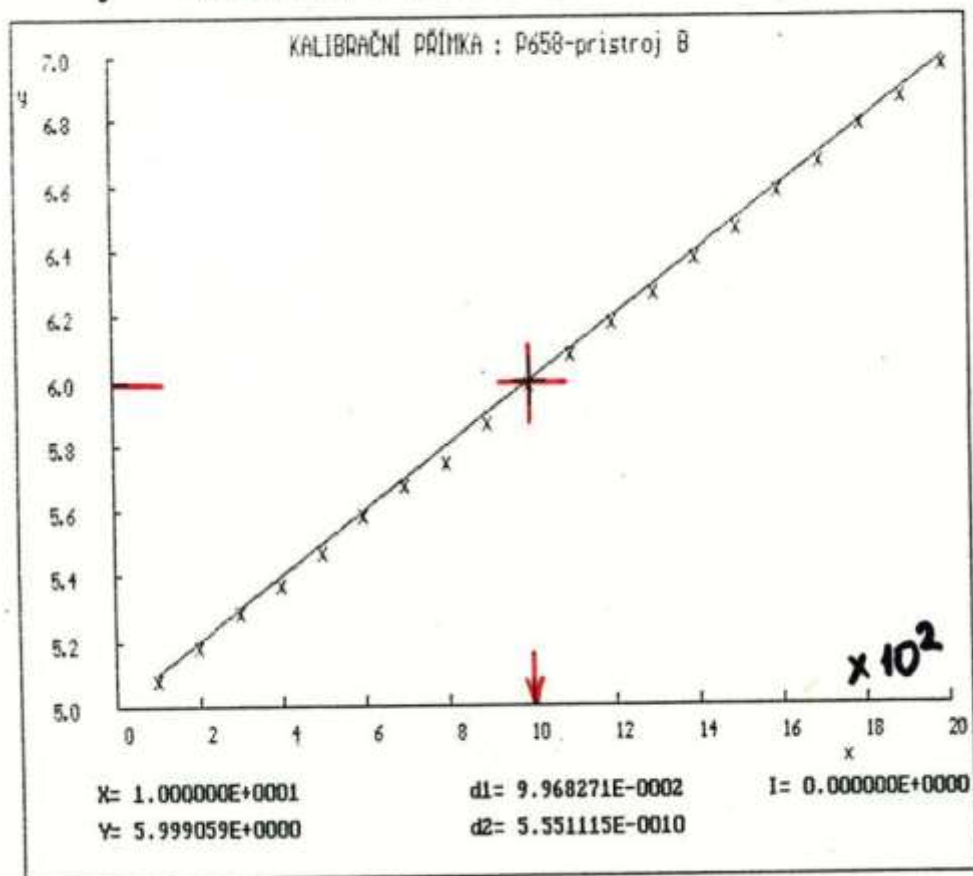
(5) CALIBRATION TABLE:

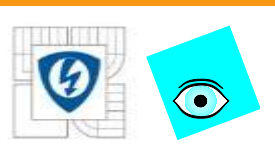
Response measured	Direct estimation	Naszodi estimation	Confidence interval lower limit	Confidence interval upper limit
yexp[i]	xcal[i]	xcal[i]	Llxcal[i]	Luxcal[i]
6.0000E+00	9.3248E+02	9.3334E+02	4.5648E+02	1.4085E+03



Příklad 6.58 Vliv přesnosti přístroje na mez detekce

$$y = 0.00100(+/-0.00000).x + 5.0022(0.0044)$$





Příklad 6.58 Vliv přesnosti přístroje na mez detekce

(2) CALIBRATION PARAMETERS:

Parameter	Estimate	Standard deviation	Test of $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$ t-Criterion	H_0 hypothesis is	Sig.level
Intercept	5.0022E+00	4.4460E-03	1.1251E+03	Rejected	0.000
Slope	9.9683E-04	3.7114E-06	2.6858E+02	Rejected	0.000

(3) ANALYSIS OF RESIDUALS:

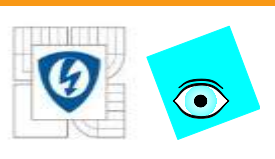
Residual sum of squares, RSS	: 1.6488E-03
Mean of absolute values of residuals, Me	: 6.1355E-03
Mean of relative residuals, Mer[%]	: 0.104
Estimate of residual variance, $s^2(e)$	: 9.1602E-05
Estimate of residual standard deviation, $s(e)$	: 9.5709E-03

(4) CALIBRATION LIMITS:

Critical level,	yc:	5.0116E+00	xc:	9.3704E+03
Detection limit,	yd:	5.0208E+00	xd:	1.8613E+01
Determination limit,	ys:	1.0553E-01	xs:	-4.9123E+03

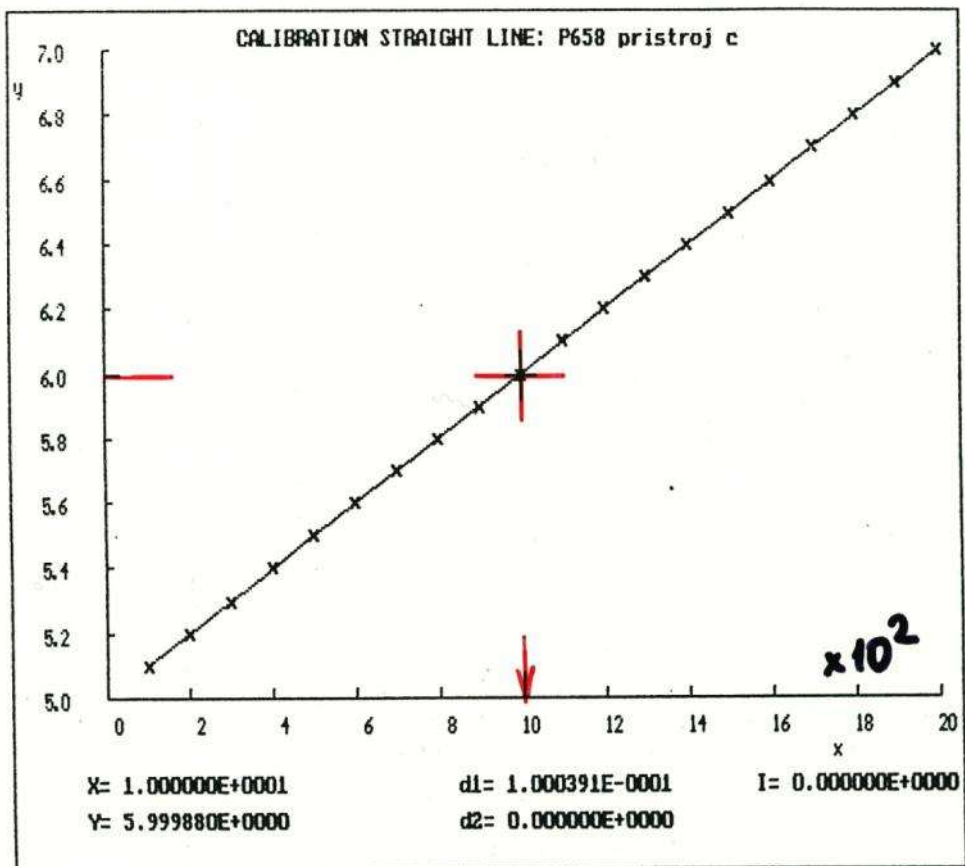
(5) CALIBRATION TABLE:

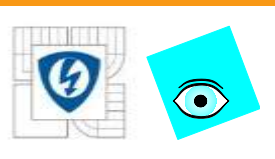
Response measured yexp[i]	Direct estimation xcal[i]	Naszodi estimation xcal[i]	Confidence interval lower limit Llxcal[i]	upper limit Luxcal[i]
6.0000E+00	1.0009E+03	1.0009E+03	9.8027E+02	1.0216E+03



Příklad 6.58 Vliv přesnosti přístroje na mez detekce

$$y = 0.00100(+/-0.00000).c + 5.0000(0.0004)$$





Příklad 6.58 Vliv přesnosti přístroje na mez detekce

(2) CALIBRATION PARAMETERS:

Parameter	Estimate	Standard deviation	Test of $H_0: B[j] = 0$ vs. $H_A: B[j] \neq 0$	t-Criterion	H_0 hypothesis is	Sig.level
Intercept	4.9995E+00	4.4856E-04		1.1146E+04	Rejected	0.000
Slope	1.0004E-03	3.7445E-07		2.6716E+03	Rejected	0.000

(3) ANALYSIS OF RESIDUALS:

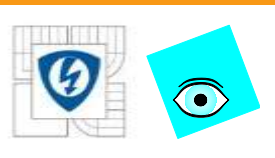
Residual sum of squares, RSS	: 1.6784E-05
Mean of absolute values of residuals, Me	: 7.2347E-04
Mean of relative residuals, Mer[%]	: <u>0.012</u>
Estimate of residual variance, $s^2(e)$	: 9.3242E-07
Estimate of residual standard deviation, $s(e)$	: 9.6562E-04

(4) CALIBRATION LIMITS:

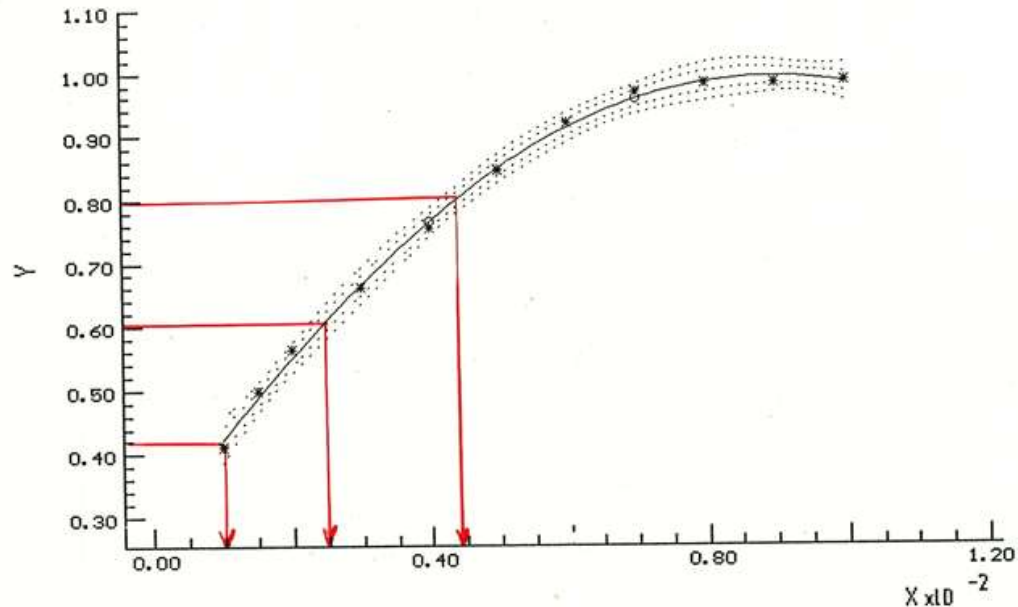
Critical level,	yc:	5.0004E+00	xc:	<u>9.4202E-01</u>
Detection limit,	yd:	5.0014E+00	xd:	<u>1.8827E+00</u>
Determination limit,	ys:	1.0647E-02	xs:	-4.9869E+03

(5) CALIBRATION TABLE:

Response measured	Direct estimation	Naszodi estimation	Confidence interval lower limit	Confidence interval upper limit
yexp[i]	xcal[i]	xcal[i]	Llxcal[i]	Luxcal[i]
6.0000E+00	<u>1.0001E+03</u>	<u>1.0001E+03</u>	<u>9.9804E+02</u>	<u>1.0022E+03</u>



Příklad: Stanovení Cd v máku metodou AAS (kalibrační přímka v mg/dm³)

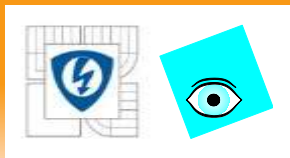


limity přesnosti kalibrační křivky:

Critical level,	yc:	2.794826E-01	xc:	1.927841E-04
Detection limit,	yd:	3.055490E-01	xd:	3.354795E-04

Neznámá koncentrace:

yexp[i]	xvyp[i]	Llxvyp[i]	Luxvyp[i]
4.130000E-01	9.904229E-04	9.517953E-04	1.024776E-03
6.050000E-01	2.323985E-03	2.275290E-03	2.373009E-03
7.970000E-01	4.422155E-03	4.376491E-03	1.763684E-02



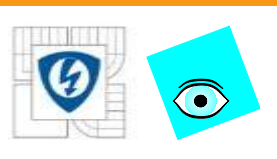
Úloha K6.04 Kalibrační model mědi v oceli rtg. – fluorescenční metodou

Rtg.-fluorescenční metodou byl stanovován obsah mědi v rozsahu 0.01 až 0.47% ve standardních vzorcích oceli.

Neznámé vzorky vykazovaly $y^* = 555, 1005, 1505$ impulzů signálu.

Úkoly:

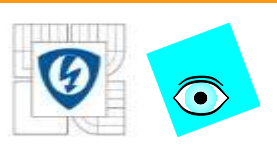
- (1) Nalezněte kalibrační model metodou regresního tripletu a vyšetřete vlivné body.
- (2) Sestrojte konfidenční interval kalibrační křivky a stanovte míry přesnosti kalibrace.
- (3) Vyčíslete bodové a intervalové odhady koncentrace neznámých vzorků.



Úloha K6.04 Kalibrační model mědi v oceli rtg. – fluorescenční metodou

Data: Obsah mědi c [%], y [impulzy]

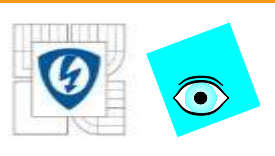
x :	0.10	y :	694	685	685	697	684,
	0.23		1211	1228	1220	1209	1212,
	0.17		953	951	960	968	950,
	0.30		1503	1515	1510	1498	1499,
	0.015		339	352	345	345	349,
	0.43		2028	2012	2031	2026	2028,
	0.73		3252	3247	3264	3244	3253,
	0.23		1244	1250	1239	1234	1240,
	0.47		2175	2172	2186	2175	2171,



Úloha K6.04 Kalibrační model mědi v oceli rtg. – fluorescenční metodou

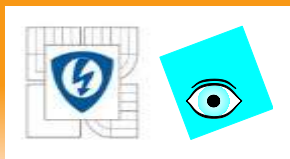
Postup výpočtu:

1. Výstavba kalibračního modelu metodou regresního tripletu (data, model, metoda) diagnostikou modelem **Lineární regrese**.
2. Zadávání neznámých vzorků M a jejich hodnot $y_j^*, j = 1, \dots, M$.
3. Program **Kalibrace**



Program kalibrace

- (a) Vypočte průměry, rozptyly, součty čtverců a kovarianci.
- (b) Určí odhady směrnice a úseku spolu s odpovídajícími rozptyly, korelační koeficient, predikce $\hat{y}_i$ s absolutními a relativními odchylkami.
- (c) Vyčíslí kritickou úroveň (x_c, y_c) , limitu detekce (x_D, y_D) a limitu stanovení (x_S, y_S) .
- (d) Kreslí graf kalibrační přímky spolu s experimentálními body.
- (e) Pro zadané hodnoty signálu $\{y_j^*\}, j = 1, \dots, M$, vyčíslí bodové a intervalové odhady pro $x_j^*, j = 1, \dots, M$.



Příklad 6.55 Přesnost kalibrace Lambertova-Beerova zákona

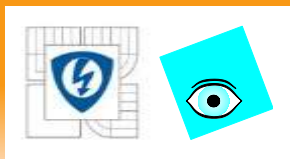
Pro kalibrační model Lambertova-Beerova zákona vypočtete kritickou úroveň y_c , limitu detekce y_D a limitu stanovení y_s pro relativní směrodatnou odchylku $C = 0.1$.

Řešení:

Z hodnot	$\bar{x} = 21.25$	$\bar{y} = 0.5368$	$\sum_{i=1}^{16} (x_i - \bar{x})^2 = 2125$
	$K = 0.0252$	$\hat{\sigma}^2 = 1.722 \cdot 10^{-6}$	$t_{1-\alpha/2}(16 - 2) = 2.14$

Ize dosazením určit

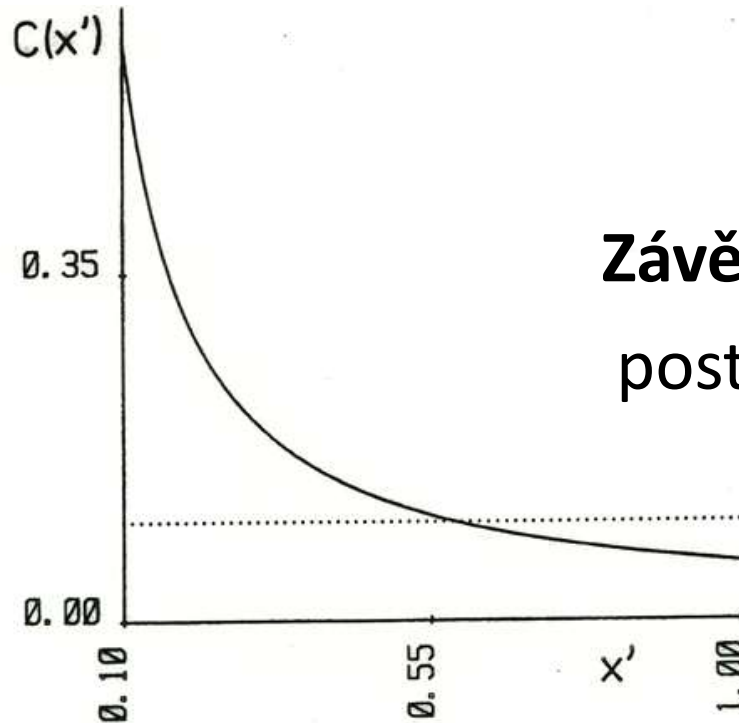
$y_c = 0.0129$	$x_c = 0.504$	$y_D = 0.0257$
$x_D = 1.008$	$y_s = 0.0593$	$x_s = 2.339.$



Příklad 6.55 Přesnost kalibrace Lambertova-Beerova zákona

Závislost $C(x')$ na x' v intervalu $0.1 \leq x' \leq 1$.

Hodnotě $C(x') = 0.1$ zde odpovídá $x'_s \approx 0.6$ a
 $y'_s = \bar{y} + (0.6 - \bar{x})b_1 = 0.0164$.



Závěr: K určení metody kalibrace
postačuje určení limity detekce.