



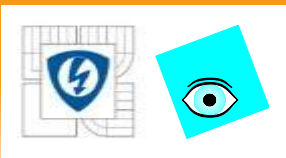
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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 2.2

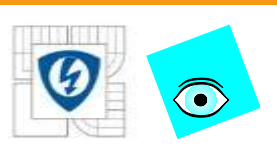
Určení minimální velikosti výběru

OVĚŘENÍ PŘEDPOKLADŮ O DATECH

31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

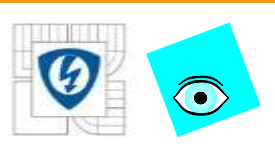




Určení minimální velikosti výběru

Rozsah výběrů n ovlivňuje přesnost odhadů parametrů polohy.

Rozptyly odhadů jsou funkcí n^{-1} .



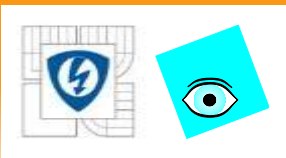
Metoda volby šíře intervalu přesnosti d

1. Z n_1 předběžných hodnot určí odhad rozptylu $s_0^2(x)$.
2. Zvolí se číslo d tak, aby s pravděpodobností $(1 - \alpha)$ platilo $\mu - d \leq \bar{x} \leq \mu + d$.

3. Minimální velikost výběru se vyčíslí dle

$$n_{min} = \left[\frac{t_{1-\frac{\alpha}{2}}(n_1 - 1)}{d} \right]^2 s_0^2(x)$$

1. Kde $t_{1-\frac{\alpha}{2}}(n_1 - 1)$ je kvantil Studentova rozdělení s $n_1 - 1$ stupni volnosti.

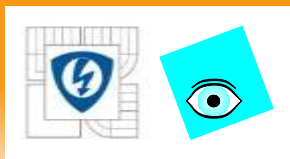


Metoda volby relativní směrodatné odchylky

Relativní chyba směrodatné odchylky $\delta(s)$ má předepsanou hodnotu

$$n_{min} = \frac{g_2(x) - 1}{4\delta^2(s)} + 1$$

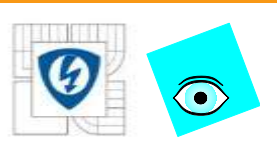
kde $g_2(x)$ je špičatost. Při $\delta(s) = 0.1$ (tj. 10%) je $n_{min} \approx 50$.



Minimální rozsahy výběru zajišťující relativní chybu směrodatné odchylky

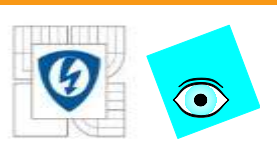
$$\delta(s) = 10\%$$

Rozdělení		Špičatost g_2	Minimální rozsah $n_{\min}$
(A)	rovnoměrné	1.8	21
(B)	normální	3.0	51
(C)	exponenciální	6.0	126
(D)	Laplaceovo	9.0	176
(E)	lognormální	15.0	351



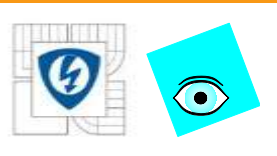
Klasické parametry

Název sloupce :	P226
Průměr :	10,40588235
Spodní mez :	7,742475919
Horní mez :	13,06928879
Rozptyl :	26,83433824
Směr. odchylka :	5,180187085
Šikmost	1,399077155
Odchylka od 0 :	Významná
Špičatost :	4,272081867
Odchylka od 3 :	Nevýznamná
Polosuma	13,5
Modus :	7,611764706



Robustní parametry

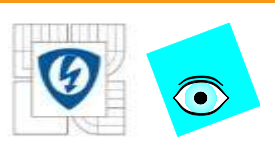
Název sloupce :	P226		
Medián :	9,4		
IS spodní :	0,206364272		
IS horní :	18,59363573	40% Průměr :	9,233333333
Medianová směr. odchylka :	4,336814353	40% IS spodní :	6,987013966
Medianový rozptyl :	18,80795873	40% IS horní :	11,4796527
10% Průměr :	9,993333333	40% Směr. odchylka :	1,065414535
10% IS spodní :	7,057957526	40% Rozptyl :	1,135108131
10% IS horní :	12,92870914		
10% Směr. odchylka :	3,788983074		
10% Rozptyl :	14,35639273		
20% Průměr :	9,453846154		
20% IS spodní :	7,788062387		
20% IS horní :	11,11962992		
20% Směr. odchylka :	1,845382576		
20% Rozptyl :	3,405436851		



Označení nezávislosti prvků výběru

Závislost měření je způsobena

- a) nestabilitou měřicího zařízení,
- b) nekonstantností podmínek měření,
- c) zanedbáním faktorů: objem vzorků, teplota, nečistota, ...
- d) nesprávným, nenáhodným výběrem vzorků k měření,
- e) časová závislost mezi prvky výběru.



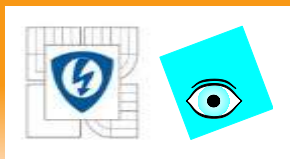
Testy nezávislosti

Testují H_0 :

Všechny prvky výběru jsou NAVZÁJEM nezávislé, ve výběru není autokorelace.

Používají se:

- testy autokorelace určitého řádu, např. pro autokorelaci I. řádu von Neumannův test
- testy významnosti autokorelačních koeficientů



Test významnosti autokorelačního koeficientu

Hypotéza: nulová $H_0: \rho_a = 0$, a alternativní $H_A: \rho_a \neq 0$

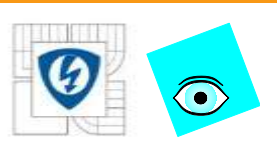
Testační kritérium:

$$t_n = \frac{T_1 \sqrt{n+1}}{\sqrt{1-T_1}} \quad \text{kde } T_1 = \left(1 - \frac{T}{2}\right) \sqrt{\frac{n^2-1}{n^2-4}} \quad \text{a}$$

T je von Neumannův poměr

$$T = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Testování: je-li $|t_n| > t_{1-\frac{\alpha}{2}}(n+1)$ je nutno hypotézu o nezávislosti prvků výběru na hladině významnosti α zamítnout.



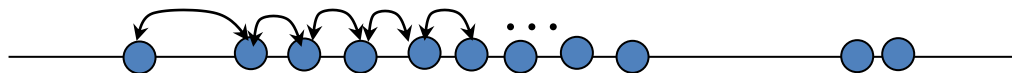
Závislost a autokorelace

Obecná definice závislosti: $x_i = kF(x_1, x_2, \dots, x_{i-1}) + e_i$
pokud platí $k = 0$, jedná se o data nezávislá.

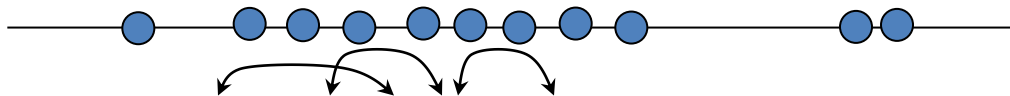
Lineární závislost prvků jednoho souboru -

AUTOKORELACE

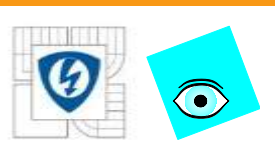
$x_i = \rho_k x_{i-k} + e_i$ kde ρ_k je autokorelační koeficient k -tého řádu.



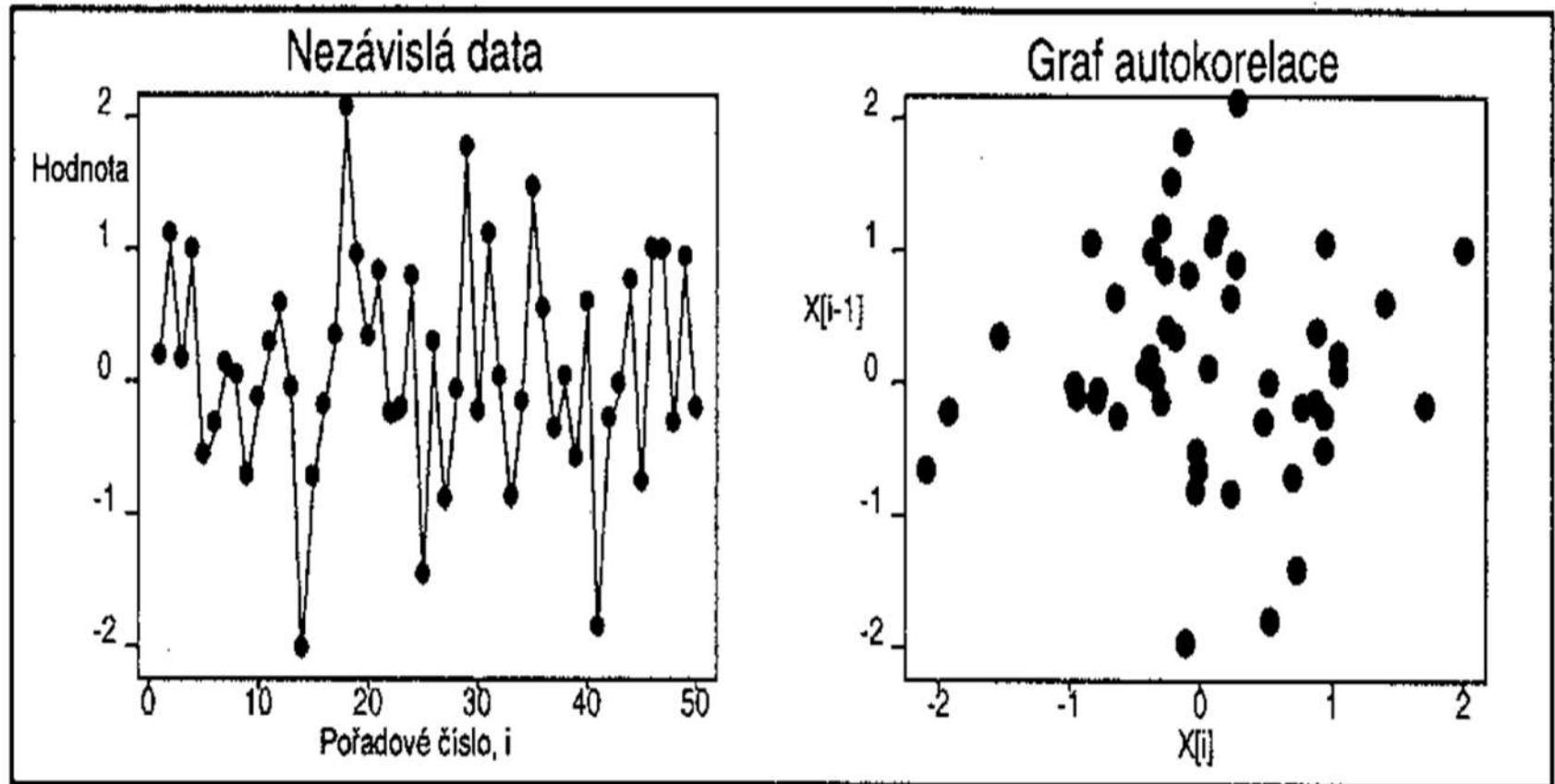
autokorelace I. řádu
sousední hodnoty

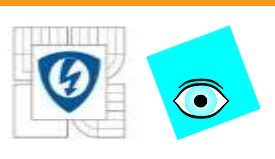


autokorelace II. řádu
hodnoty „přes jednu“

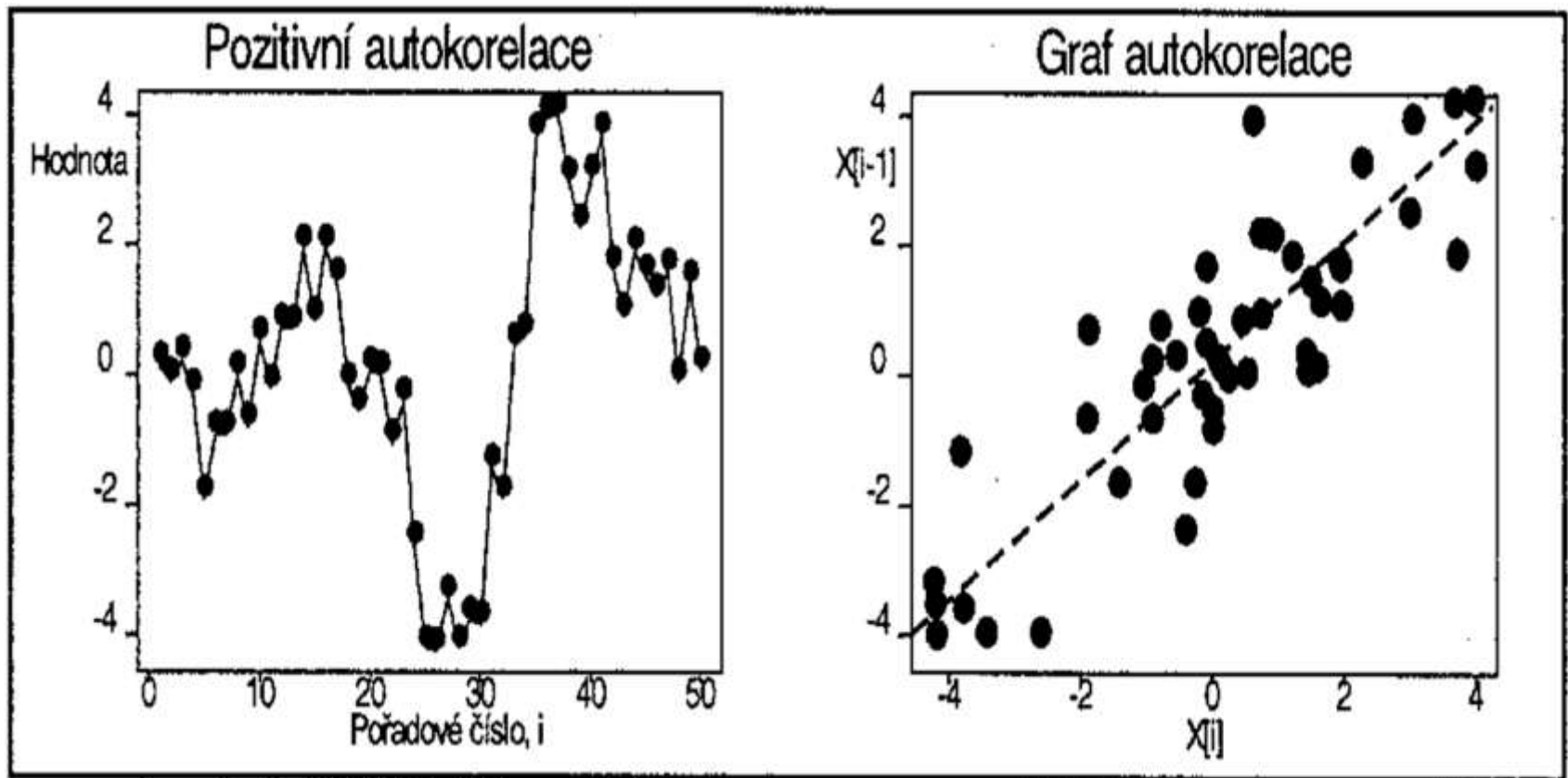


Příklady autokorelace





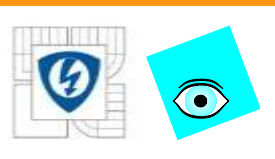
Příklady autokorelace



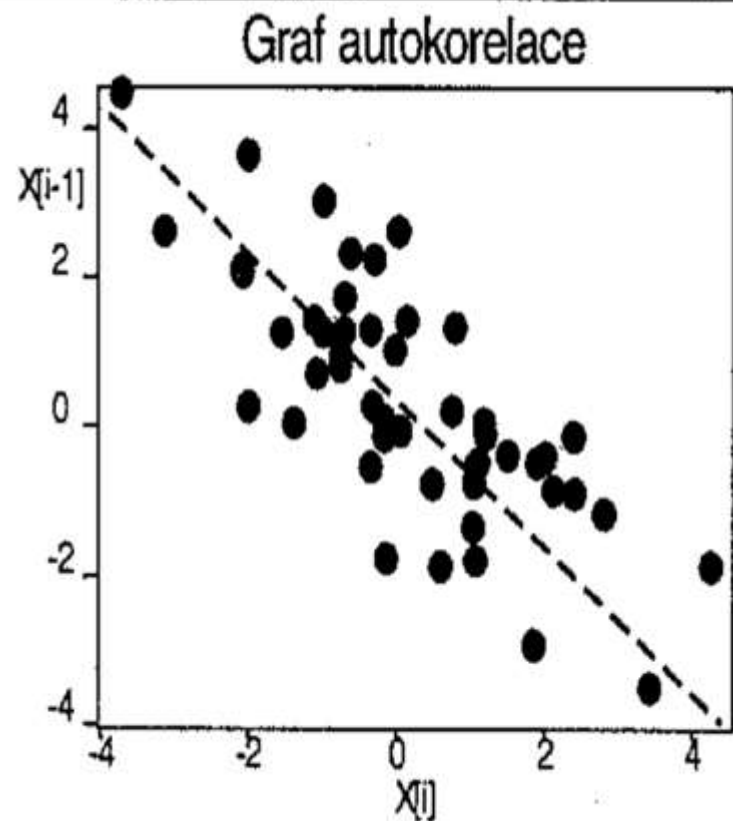
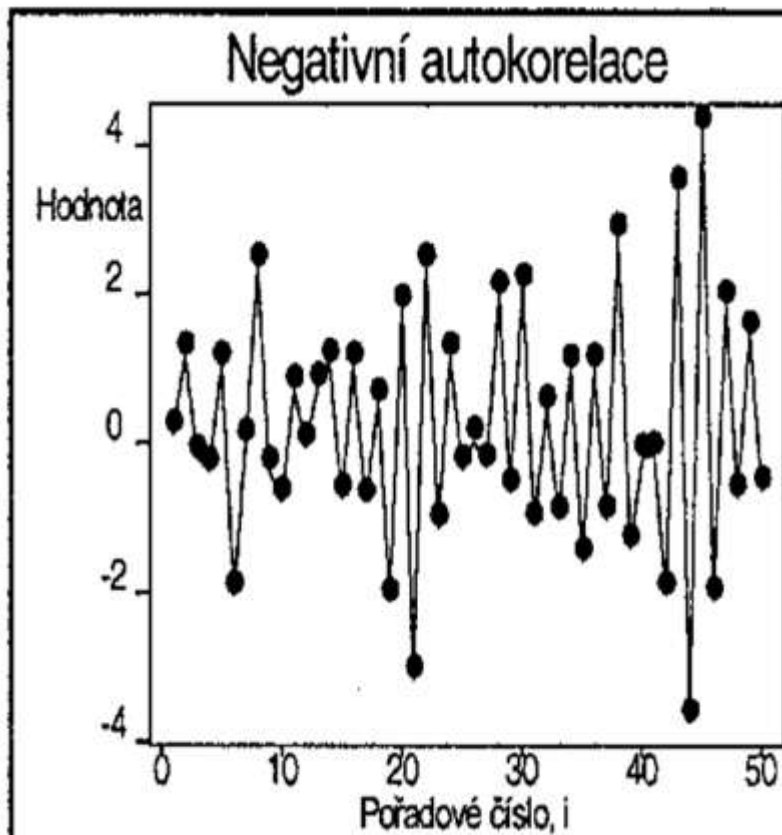
31.1.2011

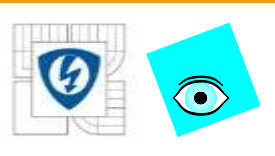
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Příklady autokorelace





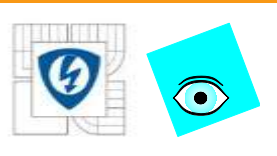
Ověření normality výběru

Existují dva základní typy testů normality:

1. Je-li typ odchylek od normality předem znám, používají se takzvané **směrové testy**.
2. Není-li typ odchylek od normality předem znám, používají se takzvané **omnibus testy**.

1. pravidlo: *Testy jsou obecně vždy méně citlivé na odchylky od normality než diagnostické grafy.*

2. pravidlo: *Když není normalita rozdělení prokázána, je nutné hlouběji analyzovat data.*



1. Test kombinace výběrové šikmosti a špičatosti

Hypotéza: nulová H_0 : normalita rozdělení výběru, vs.
 H_A : ...

Testovací kritérium: je definováno

$$C_1 = \frac{\hat{g}_1^2}{D(\hat{g}_1)} + \frac{[\hat{g}_2 - E(\hat{g}_2)]^2}{D(\hat{g}_2)}$$

kde výběrová šikmost a její rozptyl $\hat{g}_1$, $D(\hat{g}_1)$,
respektive výběrová špičatost a její střední hodnota,
respektive rozptyl $\hat{g}_2$, $E(\hat{g}_2)$, $D(\hat{g}_2)$.

Testování: při $C_1 > \chi^2_{1-\alpha}(2)$, je nutno hypotézu o normalitě rozdělení výběru zamítnout.



Title: p226

I N P U T

(1) CONDITIONS:

Sample size, n : 17
Significance level, alpha: 0.050
Output file name : P226A.TXT

(2) INPUT OF ORIGINAL DATA:

(1) 4.0000E+00 (6) 8.0000E+00 (11) 1.0000E+01 (16) 2.2000E+01
(2) 7.0000E+00 (7) 8.3000E+00 (12) 1.0500E+01 (17) 2.3000E+01
(3) 7.0000E+00 (8) 8.4000E+00 (13) 1.2000E+01
(4) 7.0000E+00 (9) 9.4000E+00 (14) 1.2800E+01
(5) 5.0000E+00 (10) 9.5000E+00 (15) 1.3000E+01

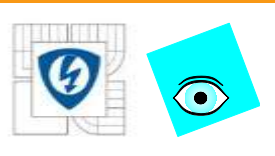
O U T P U T

(1) CLASSICAL ESTIMATORS OF LOCATION AND SHAPE:

Mean	: 1.0406E+01	Variance	: 2.6834E+01
Standard deviation	: 5.1802E+00	Skewness	: 1.3991E+00
Kurtosis	: 4.2721E+00		

(2) TEST FOR NORMALITY:

Quantile, $\chi^2(1-\alpha, 2)$	: 5.9915E+00
χ^2 statistics	: 7.9084E+00
Conclusion: Normality rejected	
Computed significance level	: 1.9174E-02



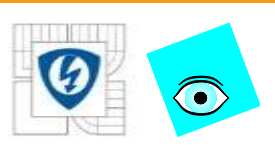
Ověření homogenity výběru

Modifikace vnitřních hradeb B_D^* a B_H^*

$$B_D^* = \tilde{x}_{0.25} - K(\tilde{x}_{0.75} - \tilde{x}_{0.25})$$

$$B_H^* = \tilde{x}_{0.75} + K(\tilde{x}_{0.75} - \tilde{x}_{0.25})$$

Parametr K : volí se tak, aby pravděpodobnost $P(n, K)$, že z výběru velikosti n pocházejícího z normálního rozdělení nebude žádný prvek mimo vnitřní hradby $[B_D^*, B_H^*]$, byla dostatečně vysoká, např. 0.95.

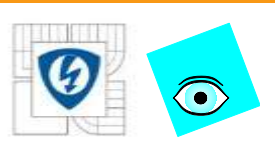


Ověření homogenity výběru

Při volbě $P(n, K) = 0.95$ lze v rozmezí $8 \leq n \leq 100$ použít aproximace

$$K \approx 2.25 - \frac{3.6}{n}$$

Pro takto určený parametr K se všechny prvky výběru ležící mimo hradby $[B_D^*, B_H^*]$ považují za vybočující.



Testy odlehlých hodnot

H_0 : Odchylka extrémní hodnoty je náhodná

GRUBBSŮV TEST (předpokládá normální rozdělení)

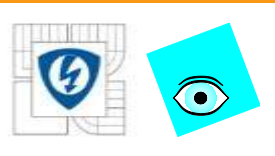
$$T_n = \frac{x_n - \bar{x}}{s} \qquad T_1 = \frac{\bar{x} - x_1}{s}$$

Hypotéza je přijata, když $T_1 < T_{1,\alpha}$, resp. $T_n < T_{n,\alpha}$.

DIXONŮV TEST (nepředpokládá normální rozdělení)

$$Q_n = \frac{x_n - x_{n-1}}{x_n - x_1} \qquad Q_1 = \frac{x_2 - x_1}{x_n - x_1}$$

Hypotéza je přijata, když $Q_1 < Q_{1,\alpha}$ resp. $Q_n < Q_{n,\alpha}$.



Test odlehlých hodnot v EDA

Metoda **modifikovaných vnitřních hradeb**

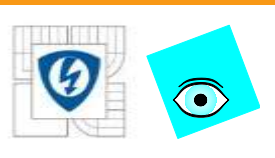
$$B_D^* = F_D - KR_F$$

$$B_H^* = F_H - KR_F$$

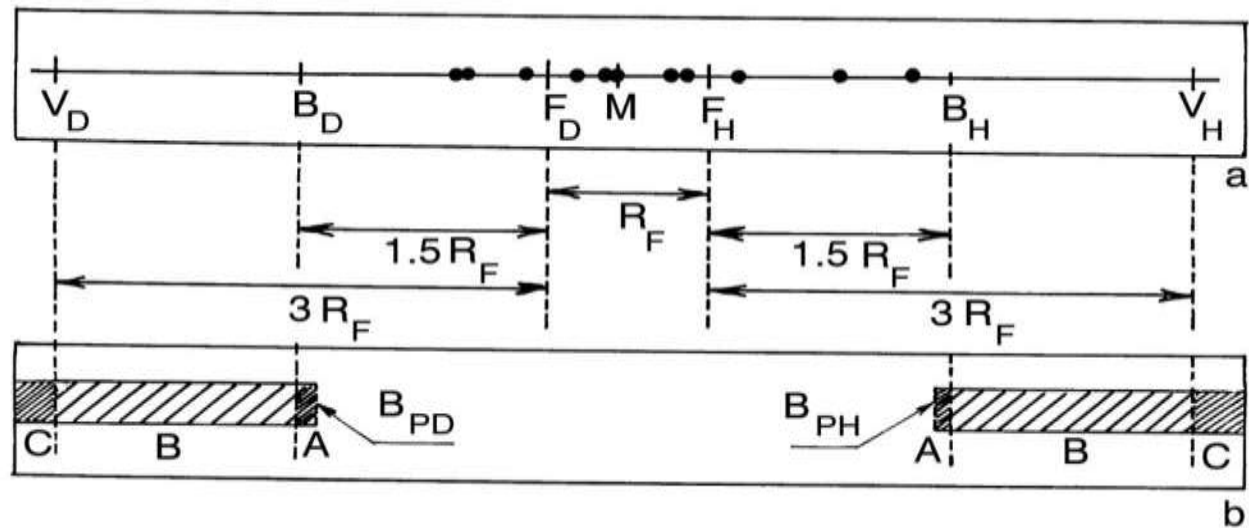
Při volbě $P(n, K) = 0.95$ lze v rozmezí $8 \leq n \leq 100$ použít aproximace

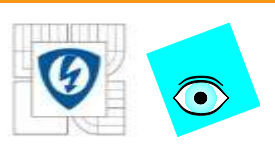
$$K \approx 2.25 - \frac{3.6}{n}$$

Pro takto určený parametr K se všechny prvky výběru ležící mimo hradby $[B_D^*, B_H^*]$ považují za vybočující.



Test odlehlých hodnot v EDA





Přilehlé hodnoty:

$B_H = F_H + 1.5R_F$ horní přilehlá hodnota

$B_D = F_D - 1.5R_F$ dolní přilehlá hodnota

Vnitřní hradby:

$V_H = F_H + 3R_F$ horní vnitřní hradba

$V_D = F_D - 3R_F$ dolní vnitřní hradba

„Podezřelá měření“ (=vybočující hodnoty):

Body mimo interval $\{V_D, V_H\}$ vnitřních hradeb.



(3) TEST FOR INDEPENDENCE:

Quantile, $t(1-\alpha/2, n+1)$: 2.1009E+00
t statistics : 1.0381E+00
Conclusion: Independence accepted
Computed significance level : 1.5649E-01

Assumption of sample homogeneity:

Arithmetic mean : 1.0406E+01
Variance : 2.6834E+01
Standard deviation : 5.1802E+00

Internal bounds:

Lower bound: -3.1909E+00

Upper bound: 2.2191E+01

(4) MINIMAL SAMPLE SIZES:

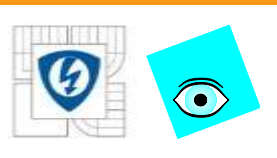
for 25% relative error of standard deviation: $n = 14$
for 10% relative error of standard deviation: $n = 83$
for 5% relative error of standard deviation: $n = 328$

(5) OUTLIERS DETECTION:

point number 17 (high) : 2.3000E+01
Number of outliers : 1

Measures with excluded outliers

Mean	: 9.6188E+00	Variance	: 1.7388E+01
Standard deviation	: 4.1699E+00	Skewness	: 1.6101E+00
Kurtosis	: 6.3399E+00		

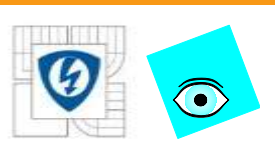


Klasické parametry

Název sloupce :	P226
Průměr :	10,40588235
Spodní mez :	7,742475919
Horní mez :	13,06928879
Rozptyl :	26,83433824
Směr. odchylka :	5,180187085
Šikmost	1,399077155
Odchylka od 0 :	Významná
Špičatost :	4,272081867
Odchylka od 3 :	Nevýznamná
Polosuma	13,5
Modus :	7,611764706

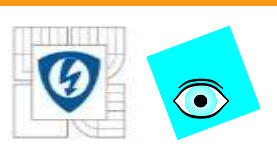
t-test

Testovaná hodnota :	0
Rozdíl :	Významný
Vypočtený :	8,282432925
Teoretický :	2,119905299
Pravděpodobnost :	1,76E-07
Konfidenční interval levý:	8,212389173
Konfidenční interval pravý:	12,59937553



Robustní parametry

Název sloupce :	P226		
Medián :	9,4		
IS spodní :	0,206364272		
IS horní :	18,59363573		
Medianová směr. odchylka :	4,336814353		
Medianový rozptyl :	18,80795873		
10% Průměr :	9,993333333		
10% IS spodní :	7,057957526	20% IS spodní :	7,788062387
10% IS horní :	12,92870914	20% IS horní :	11,11962992
10% Směr. odchylka :	3,788983074	20% Směr. odchylka :	1,845382576
10% Rozptyl :	14,35639273	20% Rozptyl :	3,405436851
20% Průměr :	9,453846154	40% Průměr :	9,233333333
		40% IS spodní :	6,987013966
		40% IS horní :	11,4796527
		40% Směr. odchylka :	1,065414535
		40% Rozptyl :	1,135108131



Znaménkový test

Závěr : Data jsou závislá

Analýza malých výběrů

N : 17

Střední hodnota : 9,5

Spodní mez (5%) : 7,395

Horní mez (95%) : 11,605

Spodní mez (2.5%) : 6,99

Horní mez (97.5%) : 12,01

Pivotové rozpětí : 5

Test normality :

Název sloupce : P226

Průměr : 10,40588235

Rozptyl : 26,83433824

Šikmost : 1,399077155

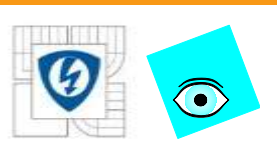
Špičatost : 4,272081867

Normalita : Přijata

Vypočtený : 5,900841254

Teoretický : 5,991464547

Pravděpodobnost : 0,052317695



Vybočující body

Název sloupce :	P226
Homogenita :	Zamítnuta
Počet vybočujících bodů :	2
Spodní mez :	-0,133823529
Horní mez :	17,63382353

Autokorelace :

Řád autokorelace 1

Korelační koeficient : 0,659732703

Pravděpodobnost : 0,002712181

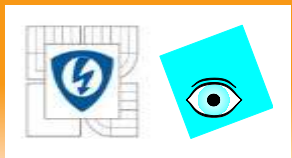
Závěr : Významný

Řád autokorelace 2

Korelační koeficient : 0,327058288

Pravděpodobnost : 0,117047366

Závěr : Nevýznamný



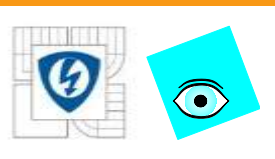
Kapitola 2.3

TRANSFORMACE DAT

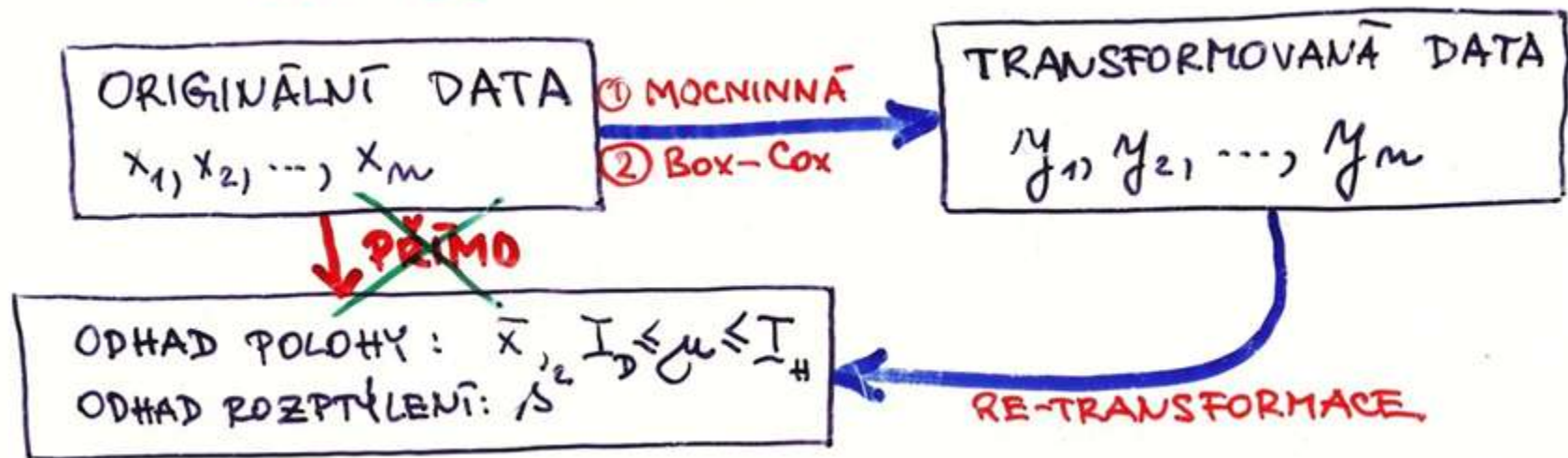
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





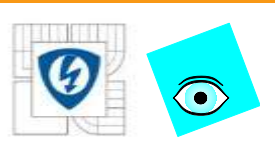
Transformace dat



(1) Transformace pro konstantní rozptyl: $\theta^2(x) \neq \textit{kons.}$

stabilizace rozptylu $\theta^2(y) \neq \textit{kons.}$

$$\theta^2(y) \approx \left(\frac{\delta g(x)}{\delta x} \right)^2 \cdot f_1(x) = C$$

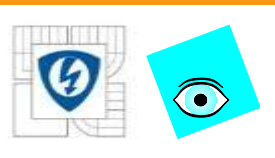


Transformace dat

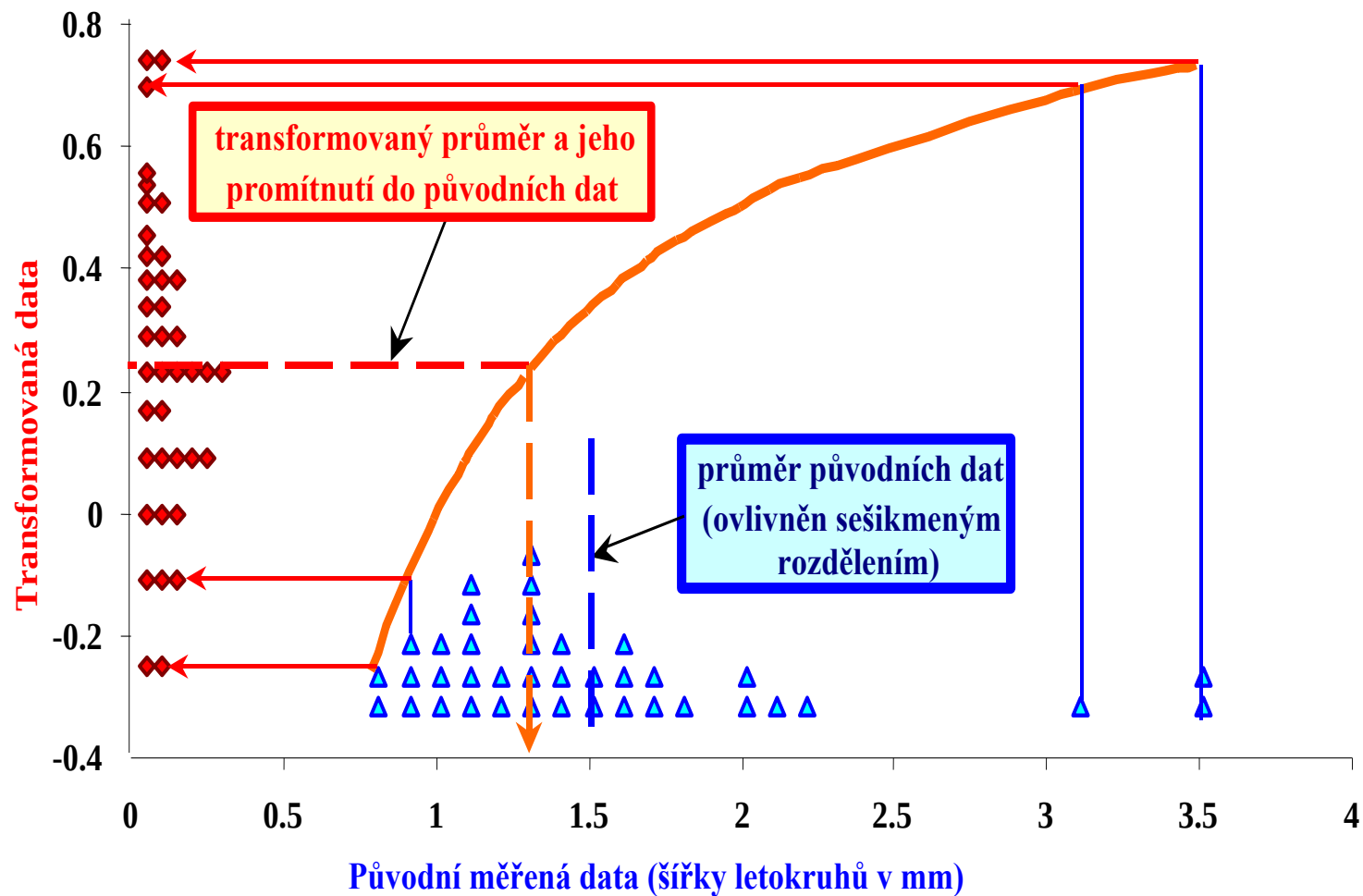
(2) Transformace pro symetrii:

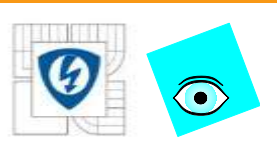
(a) zesimetričtění rozdělení: mocninné transformace

$$y = g(x) = \begin{cases} x^\lambda & \text{pro } \lambda > 0 \\ \ln x & \text{pro } \lambda = 0 \\ -x^\lambda & \text{pro } \lambda < 0 \end{cases}$$



Transformace dat - ukázka



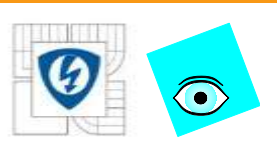


Symetrizující mocnná transformace

Zesymetričtění rozdělení výběru provede

$$y = g(x) = \begin{cases} x^\lambda & \text{pro } \lambda > 0 \\ \ln x & \text{pro } \lambda = 0 \\ -x^\lambda & \text{pro } \lambda < 0 \end{cases}$$

1. nezachovává měřítko,
2. není k λ všude spojitá,
3. zachovává pořadí dat ve výběru,
4. hodí se pouze pro kladná data.



Normalizační Box-Coxova transformace

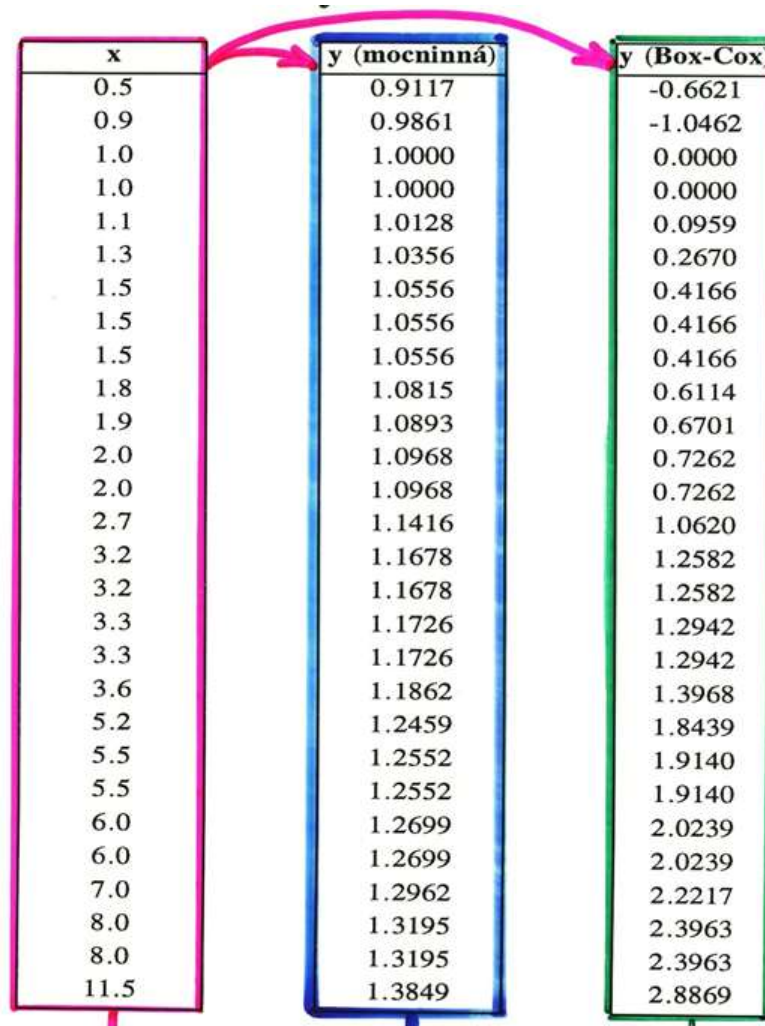
přiblížení rozdělení výběru k normálnímu

$$y = g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{pro } \lambda \neq 0 \\ \ln x & \lambda = 0 \end{cases}$$

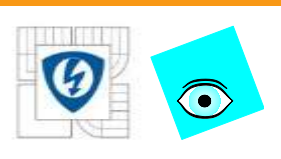
Transformace

1. je vzhledem k λ spojitá,
2. je použitelná pouze pro kladná data, jinak $(x - x_0)$.

Schéma mocninné Box-Coxovy transformace



x	y (mocninná)	y (Box-Cox)
0.5	0.9117	-0.6621
0.9	0.9861	-1.0462
1.0	1.0000	0.0000
1.0	1.0000	0.0000
1.1	1.0128	0.0959
1.3	1.0356	0.2670
1.5	1.0556	0.4166
1.5	1.0556	0.4166
1.5	1.0556	0.4166
1.8	1.0815	0.6114
1.9	1.0893	0.6701
2.0	1.0968	0.7262
2.0	1.0968	0.7262
2.7	1.1416	1.0620
3.2	1.1678	1.2582
3.2	1.1678	1.2582
3.3	1.1726	1.2942
3.3	1.1726	1.2942
3.6	1.1862	1.3968
5.2	1.2459	1.8439
5.5	1.2552	1.9140
5.5	1.2552	1.9140
6.0	1.2699	2.0239
6.0	1.2699	2.0239
7.0	1.2962	2.2217
8.0	1.3195	2.3963
8.0	1.3195	2.3963
11.5	1.3849	2.8869

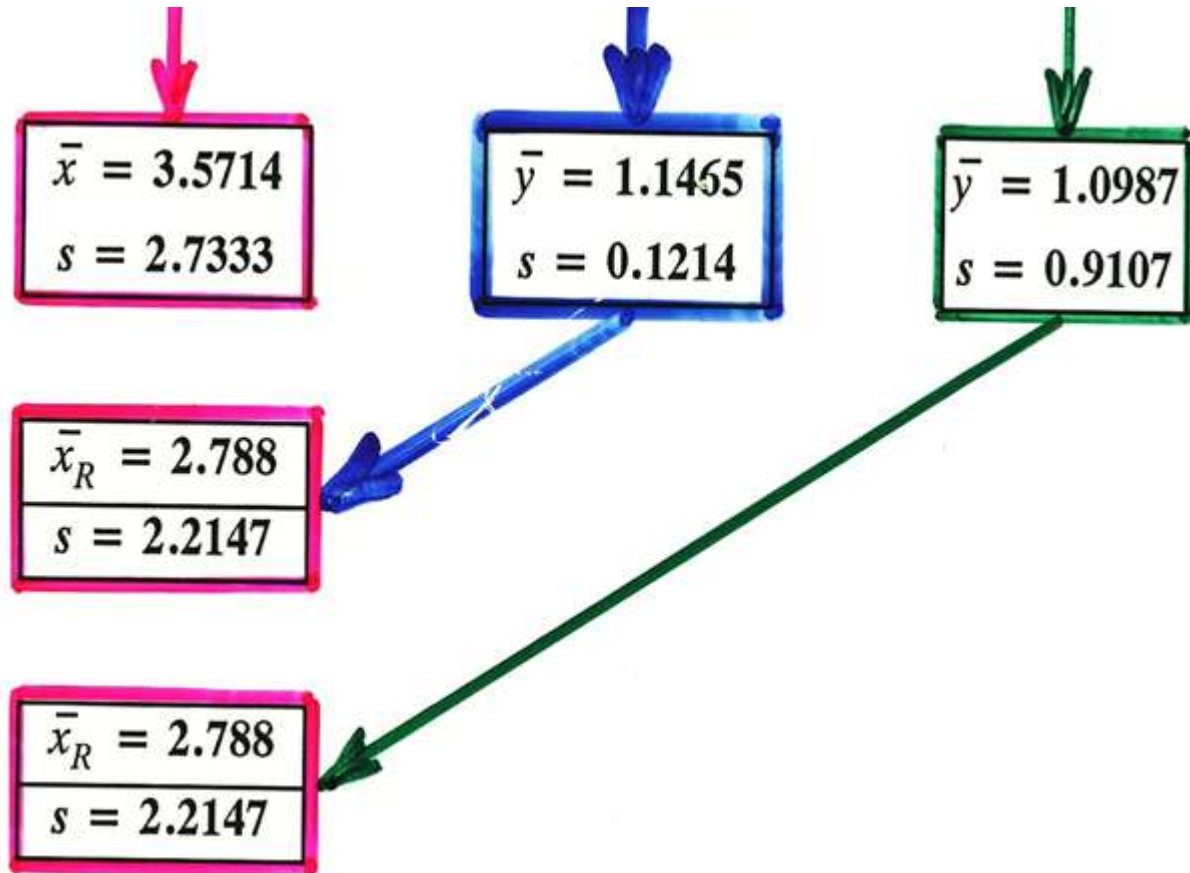


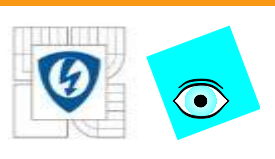
Průměr

Směrodatná odchylka

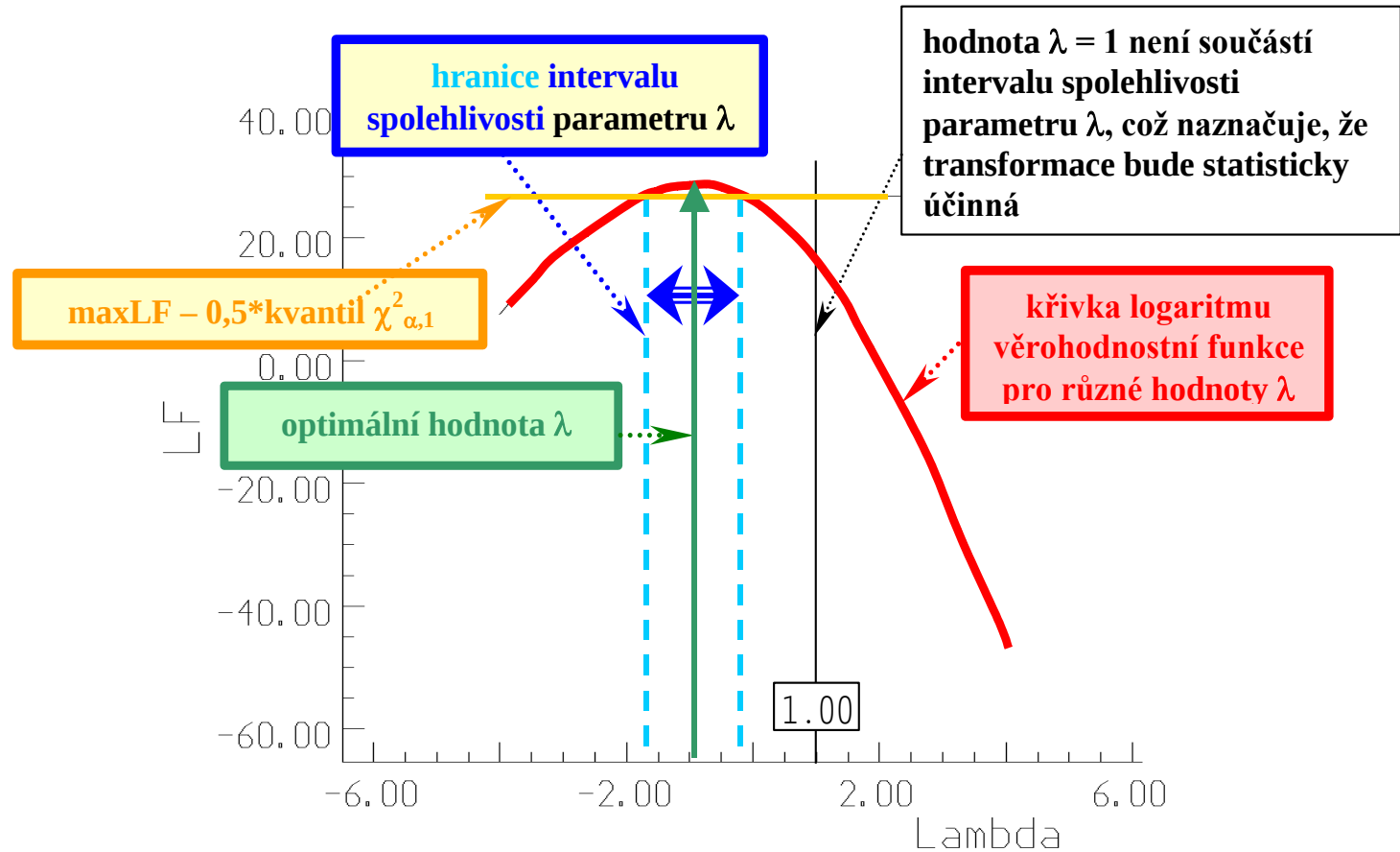
Retransform. průměr (mocnina)

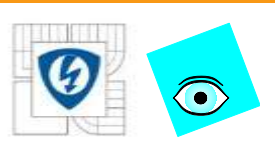
Retransform. průměr (Box-Cox)





Transformace dat – odhad optimálního λ





Zpětná transformace

$$\bar{y}, s^2(y), \bar{y} \pm t_{1-\frac{\alpha}{2}}(n-1) \cdot s(y)/\sqrt{n}$$

1. Nekorektní přístup

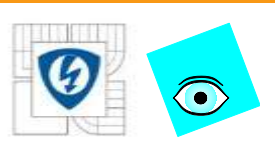
$$\bar{x}_R = \bar{x}_\lambda = \left[\frac{\sum_{i=1}^n x_i^\lambda}{n} \right]^{\frac{1}{\lambda}}$$

$\bar{x}_R = \bar{x}_{-1}$ představuje *harmonický* průměr

$\bar{x}_R = \bar{x}_0$ představuje *geometrický* průměr

$\bar{x}_R = \bar{x}_1$ představuje *aritmetický* průměr

$\bar{x}_R = \bar{x}_2$ představuje *kvadratický* průměr



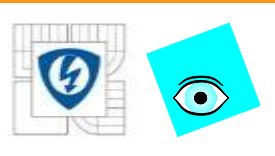
2. Korektní přístup

z Taylorova rozvoje funkce $y = g(x)$ v okolí y ,

$$\bar{x}_R = g^{-1} \left[\bar{y} - \frac{1}{2} \frac{d^2 g(x)}{dx^2} \left(\frac{d g(x)}{dx} \right)^{-2} s^2(y) \right]$$

Pro rozptyl vyjde

$$s^2(x_R) \approx \left(\frac{d g(x)}{dx} \right)^{-2} s^2(y)$$



Zpětná transformace

(a) Pro speciální případ $\lambda = 0$, tzn. logaritmickou transformaci typu $g(x) = \ln x$, bude

$$\bar{x}_R \approx \exp[\bar{y} + 0.5s^2(y)]$$

a rozptyl

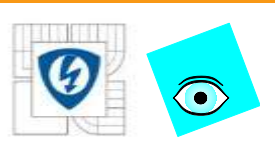
$$s^2(x_R) \approx \bar{x}_R^2 s^2(y).$$

(b) Pro případ $\lambda \neq 0$:

$$\bar{x}_{R,1,2} = \left[0.5(1 + \lambda \bar{y}) \pm 0.5 \sqrt{1 + 2\lambda(\bar{y} + s^2(y)) + \lambda^2(\bar{y}^2 - 2s^2(y))} \right]^{\frac{1}{\lambda}}$$

a rozptyl

$$s^2(x_R) \approx \bar{x}_R^{-2\lambda+2} s^2(y).$$



Zpětná transformace

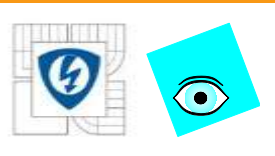
Rozptyl se vyčíslí

$$s^2(x_R) \approx \left(\frac{d g(x)}{dx} \right)^{-2} s^2(y)$$

kde derivace jsou vyčísleny v bodě $x = \bar{x}_R$.

100(1 - α)%ní interval spolehlivosti se vyčíslí dle

$$\bar{x}_R - I_D \leq \mu \leq \bar{x}_R + I_D$$



Zpětná transformace

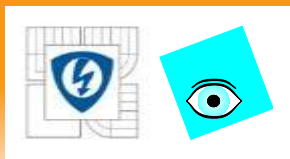
kde

$$I_D = g^{-1}(\bar{y} + G - t_{1-\frac{\alpha}{2}}(n-1) \frac{s(y)}{\sqrt{n}})$$

$$I_H = g^{-1}\left(\bar{y} + G + t_{1-\frac{\alpha}{2}}(n-1) \frac{s(y)}{\sqrt{n}}\right)$$

$$G = -0.5 \frac{d^2 g(x)}{dx^2} \left(\frac{d g(x)}{dx} \right)^{-2} s^2(y)$$

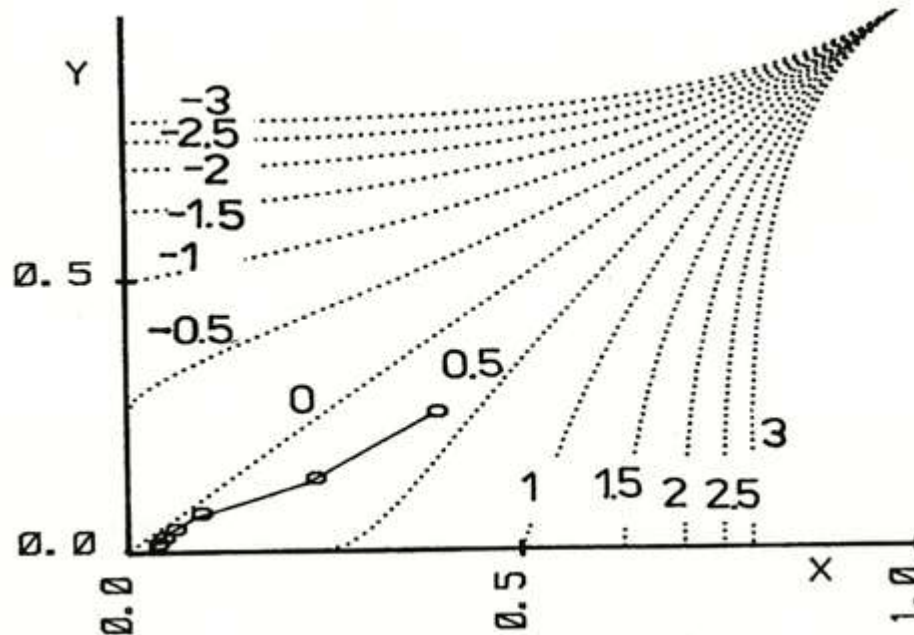
kde $t_{1-\frac{\alpha}{2}}(n-1)$ je 100(1- α /2)%ní kvantil Studentova rozdělení s $(n-1)$ stupni volnosti.



Hines - Hinesův selekční graf

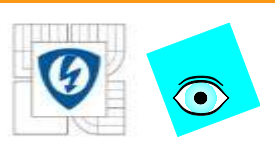
Osa x : $\tilde{x}_{0.5}/x_{1-P_i}$, Osa y : $\tilde{x}_{P_i}/\tilde{x}_{0.5}$

Selekční graf pro výběr z lognormálního rozdělení



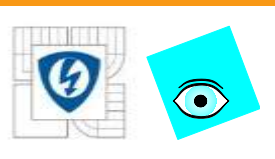
Požadavek symetrie kvantilů kolem mediánu

$$\left(\frac{\tilde{x}_{P_i}}{\tilde{x}_{0.5}}\right)^{\lambda} + \left(\frac{\tilde{x}_{0.5}}{\tilde{x}_{1-P_i}}\right)^{-\lambda} = 2$$

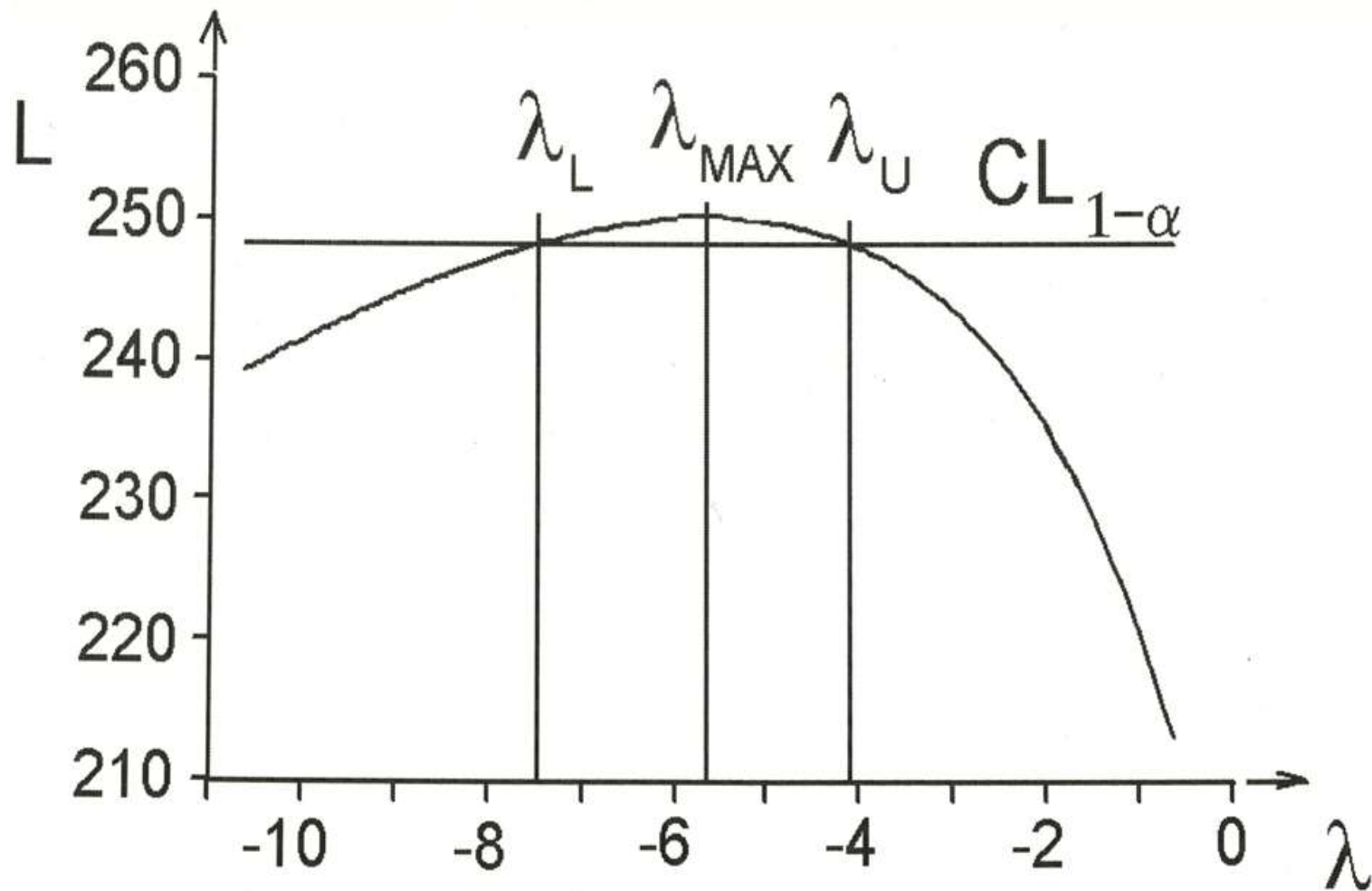


Hines - Hinesův selekční graf

kde pro pořadové pravděpodobnosti jsou voleny písmenové hodnoty $P_i = 2^{-i}, i = 2, 3$.



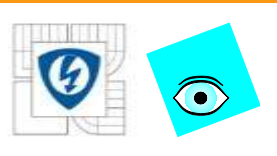
Graf maximální věrohodnosti



31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Příklad: Atestace stopového Cu v kaolinu

Ve vzorku kaolinu byl stanoven obsah Cu (ppm) a hodnoty ve vzestupném pořadí jsou 4, 5, 7, 7, 7, 8, 8.3, 8.4, 9.4, 9.5, 10.0, 10.5, 12.0, 12.8, 13.0, 22.0, 23.0.

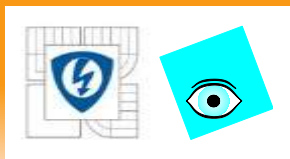
Stanovte míry polohy a rozptýlení uvedeného výběru.

Řešení:

1. Výpočet na kalkulačce

aritmetický průměr $\bar{x} = 10.406$

medián $\tilde{x}_{0.5} = 9.400$



1. Průzkumová analýza dat (EDA):

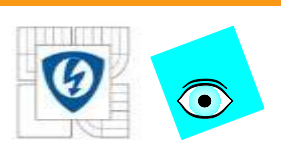
Diagnostické grafy:

- stupeň symetrie rozdělení
- lokální koncentrace
- vybočující data

2. Ověření předpokladů výběru dat:

Diagnostiky:

- ověření normality
- ověření nezávislosti
- ověření homogeneity
- ověření minimální četnosti



3. Transformace dat: pro symetrii a pro heteroskedasticitu výběru Box-Coxova transformace, mocninná transformace.

4. Odhady parametrů polohy, rozptýlení a tvaru:

Klasické odhady: aritmetický průměr, rozptyl.

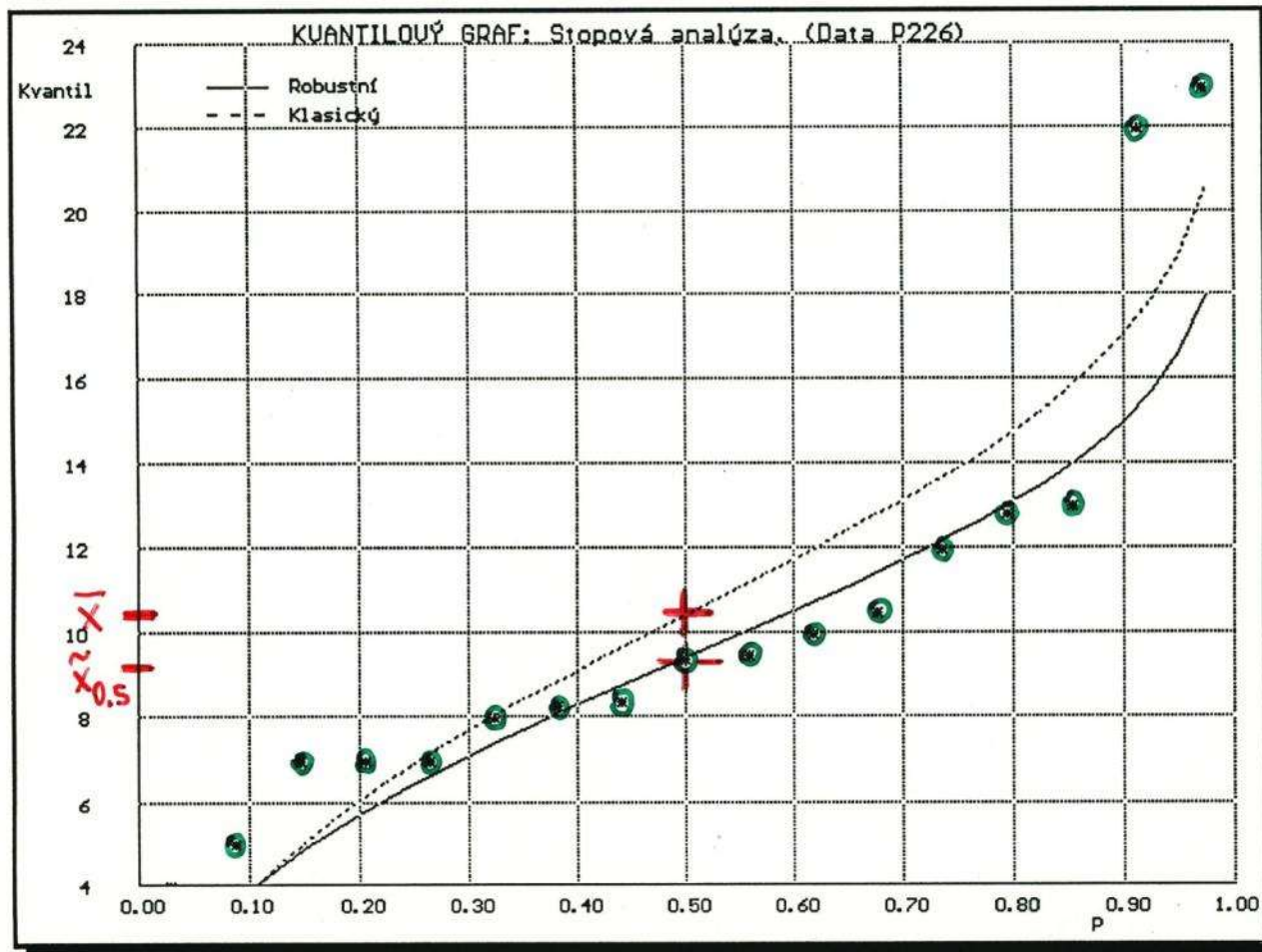
Robustní odhady: medián, uřezané průměry, neparametrické odhady rozptylu.

Adaptivní odhady polohy a rozptylu.



Kvantilový graf: stopová analýza

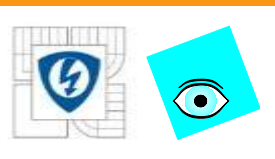
(G1)



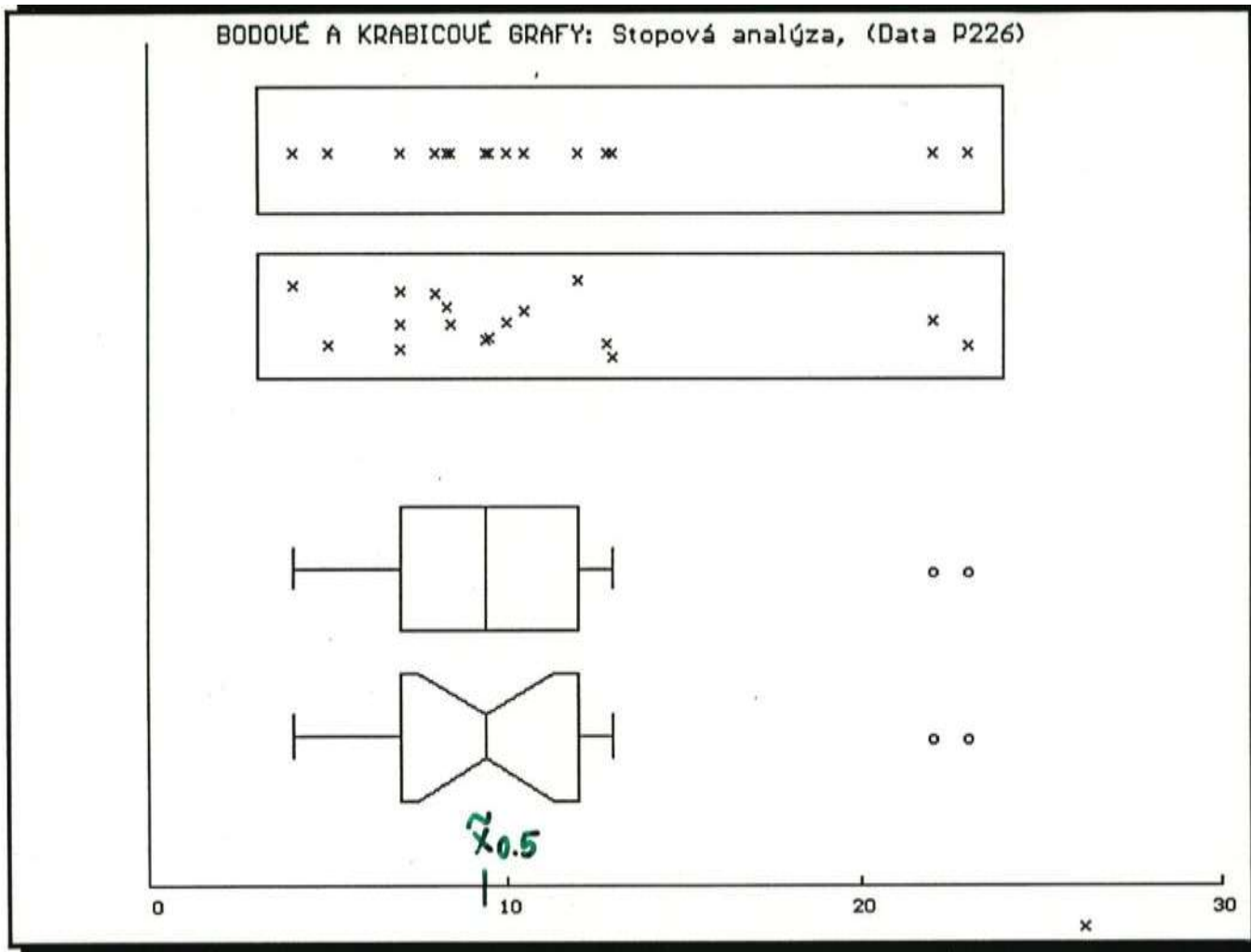
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



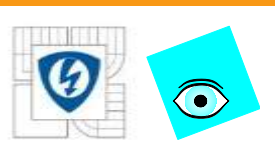


Bodové a krabicové grafy: stopová analýza



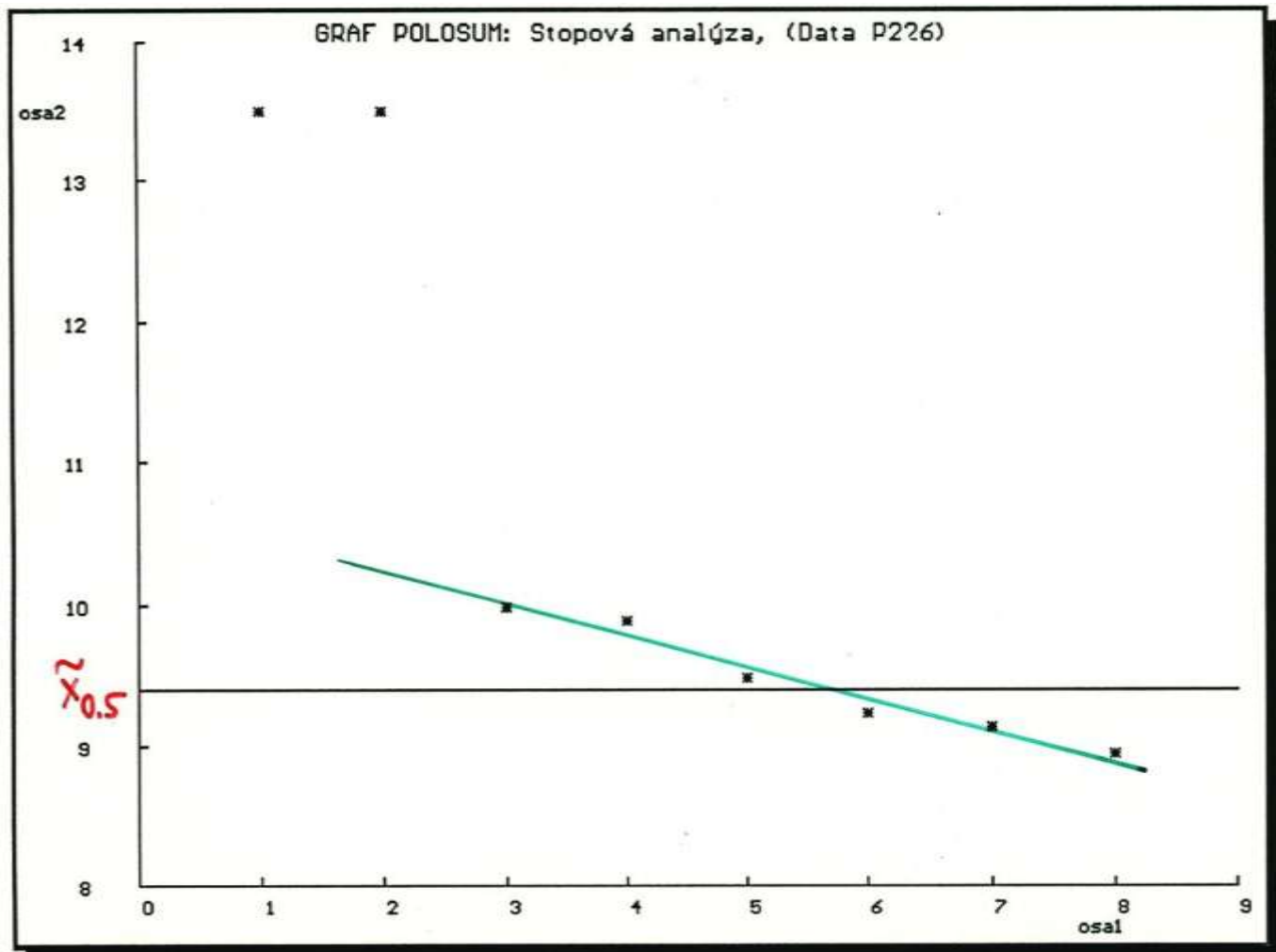
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



Graf polosum: stopová analýza

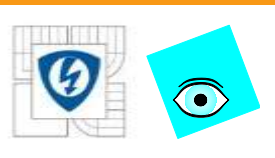
(G6)



31.1.2011

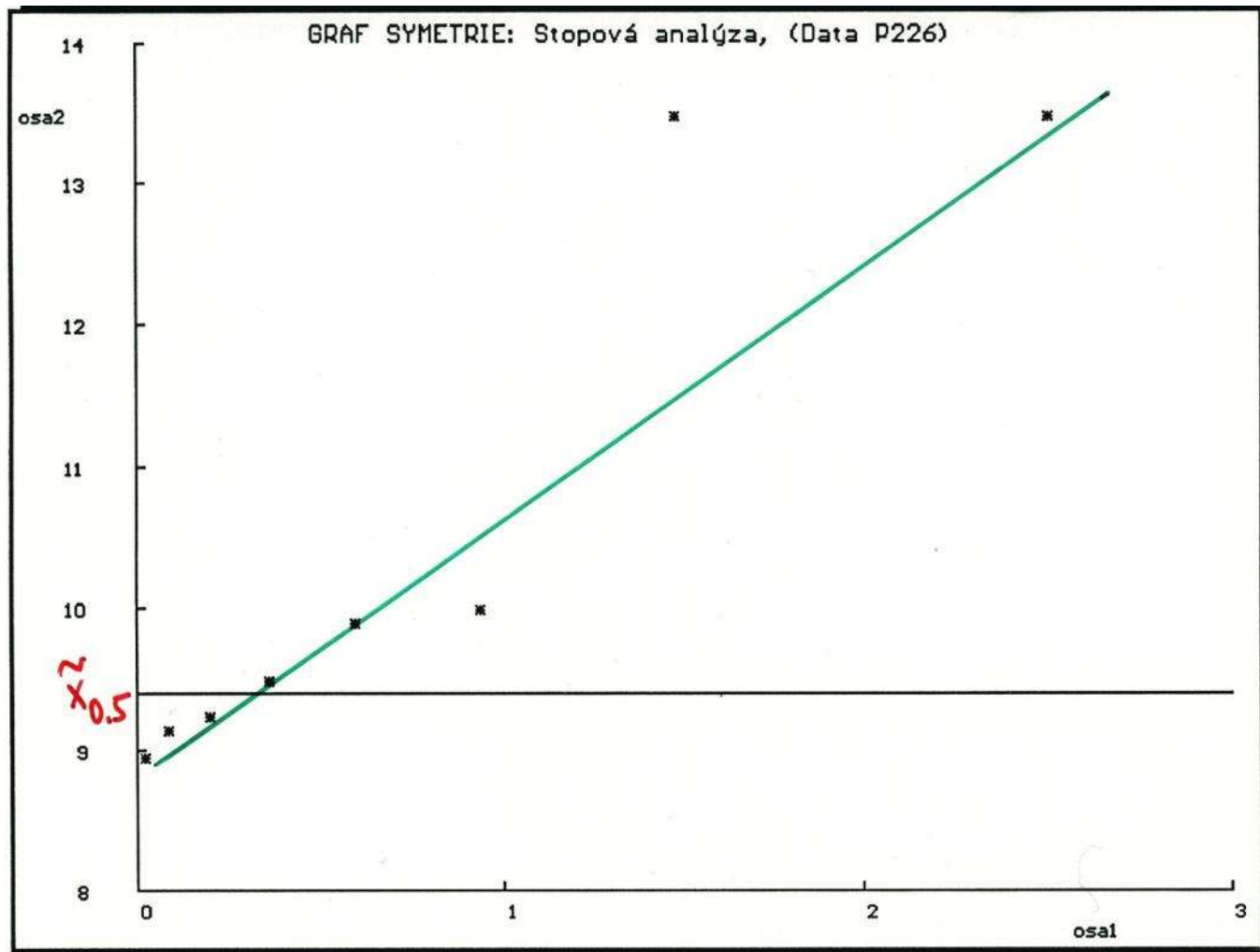
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Graf symetrie: stopová analýza

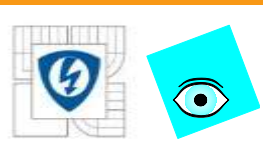
(G7)



31.1.2011

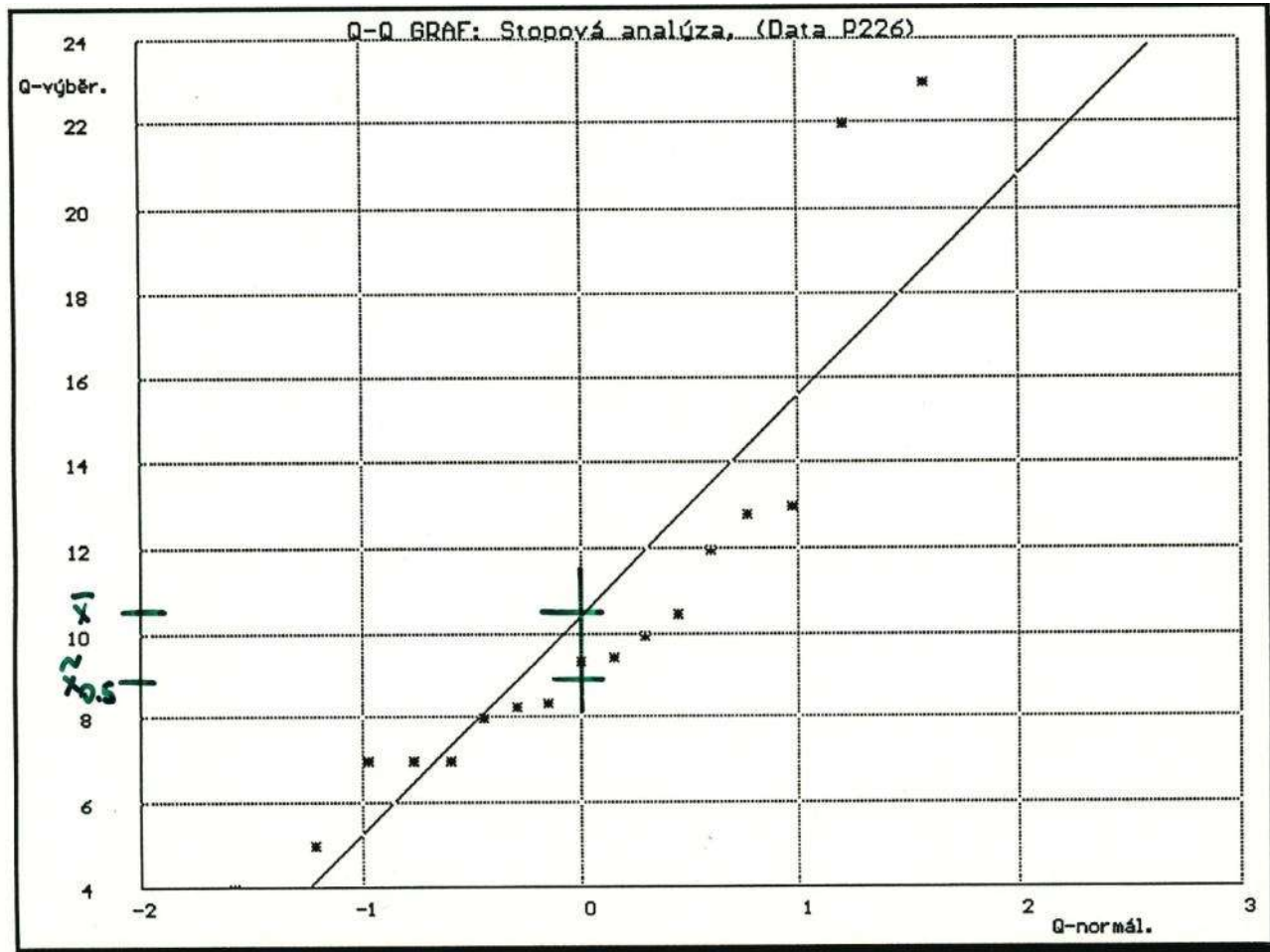
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Q-Q graf: stopová analýza

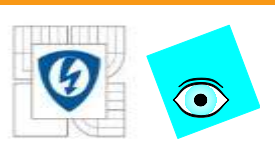
(G13)



31.1.2011

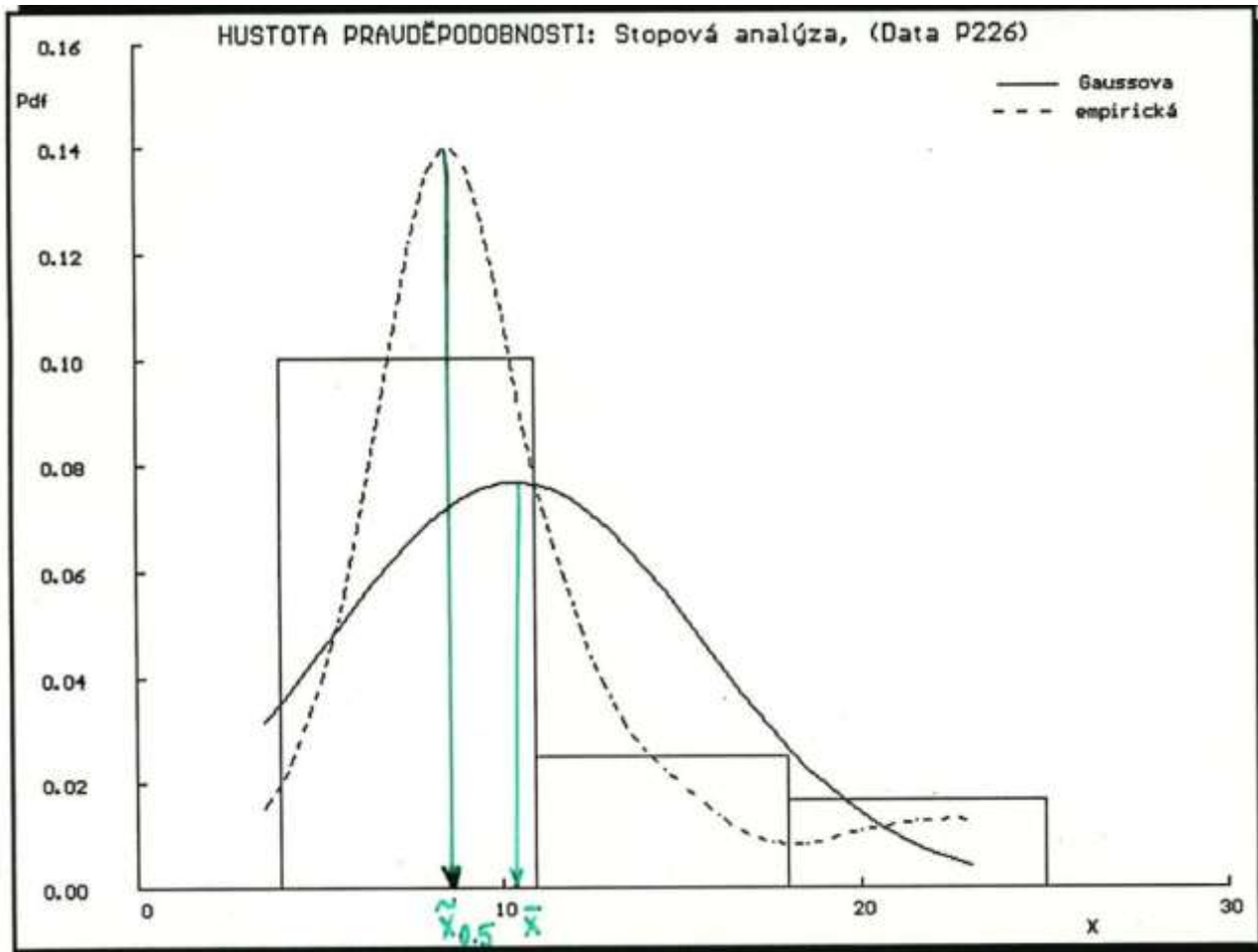
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Q-Q graf: stopová analýza

(G11)



31.1.2011

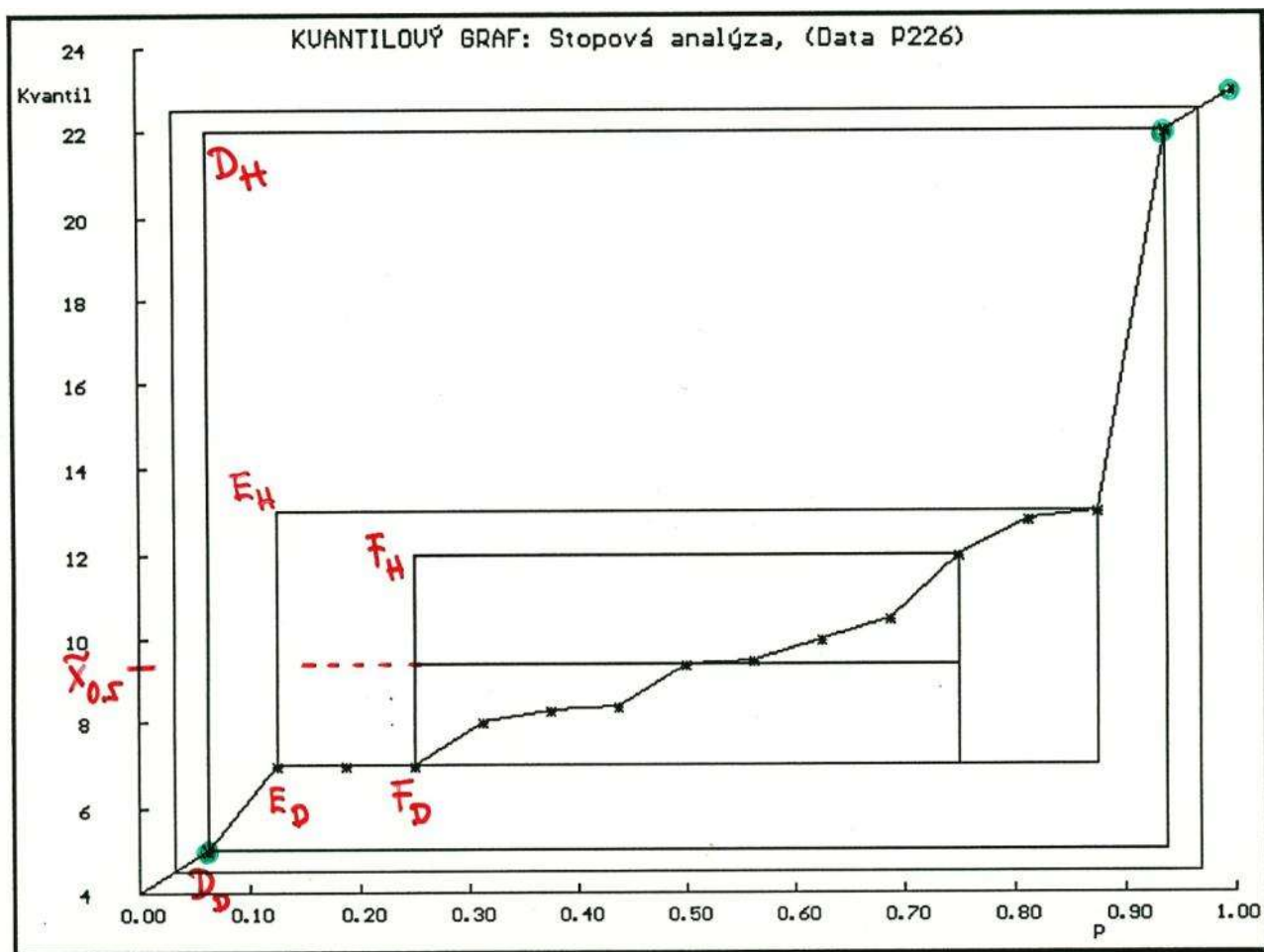
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Kvantilový graf: stopová analýza

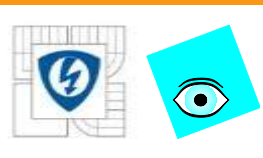
(G10)



31.1.2011

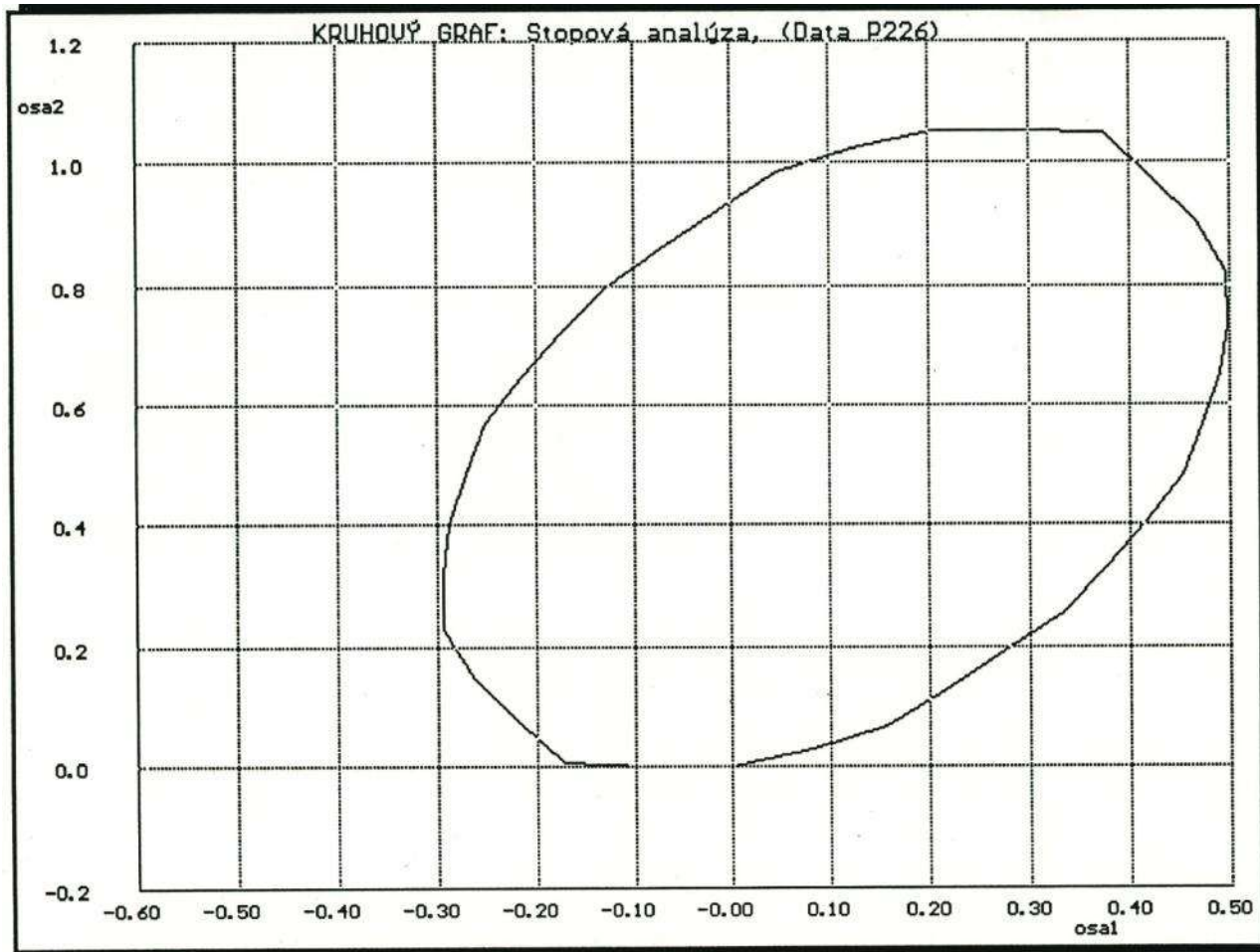
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Kvantilový graf: stopová analýza

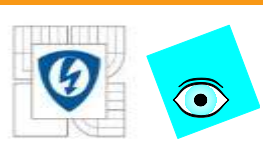
(G17)



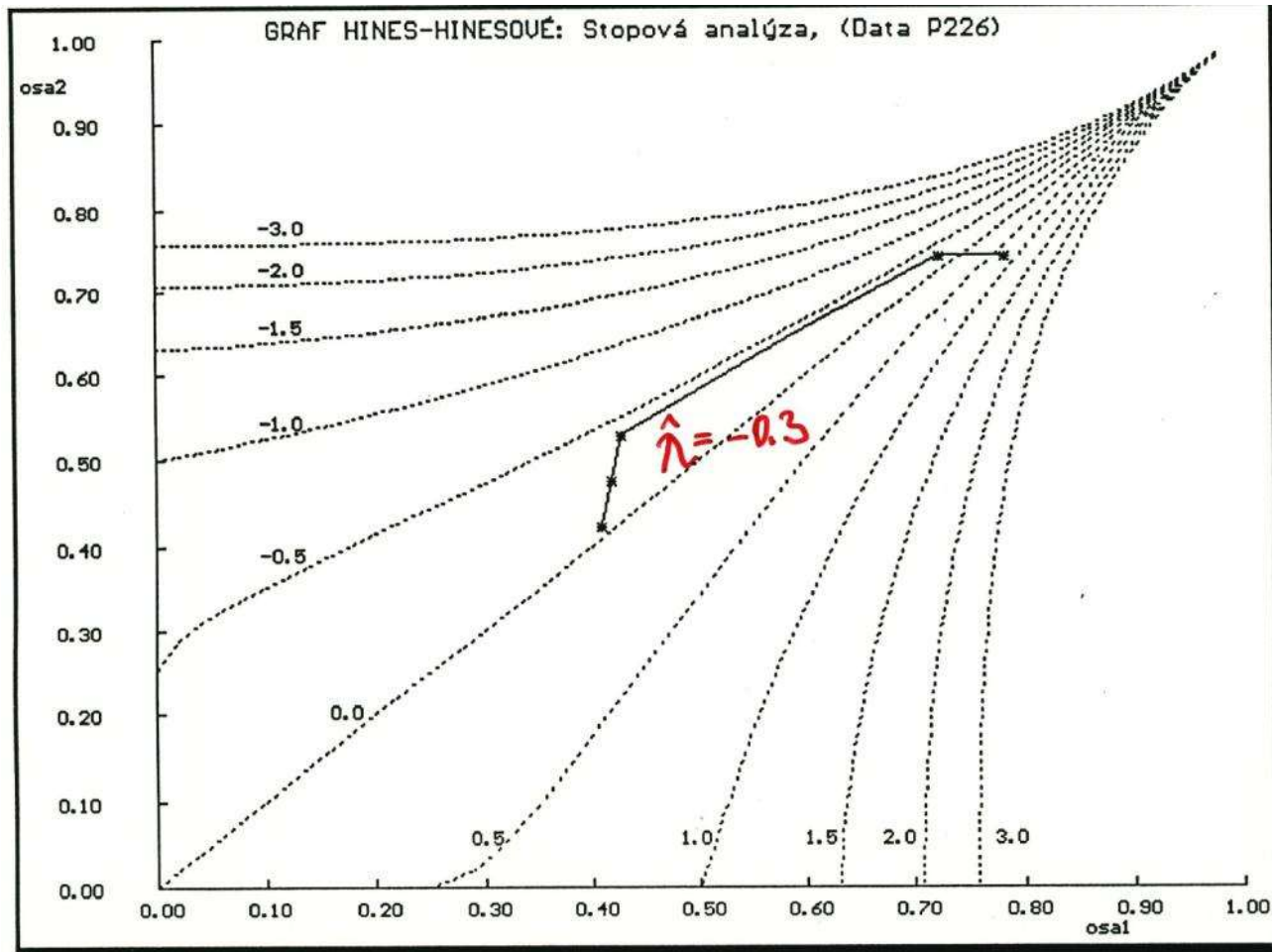
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





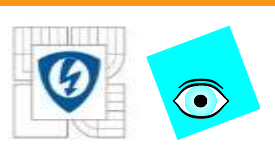
Graf Hines-Hinesové: stopová analýza (G21)



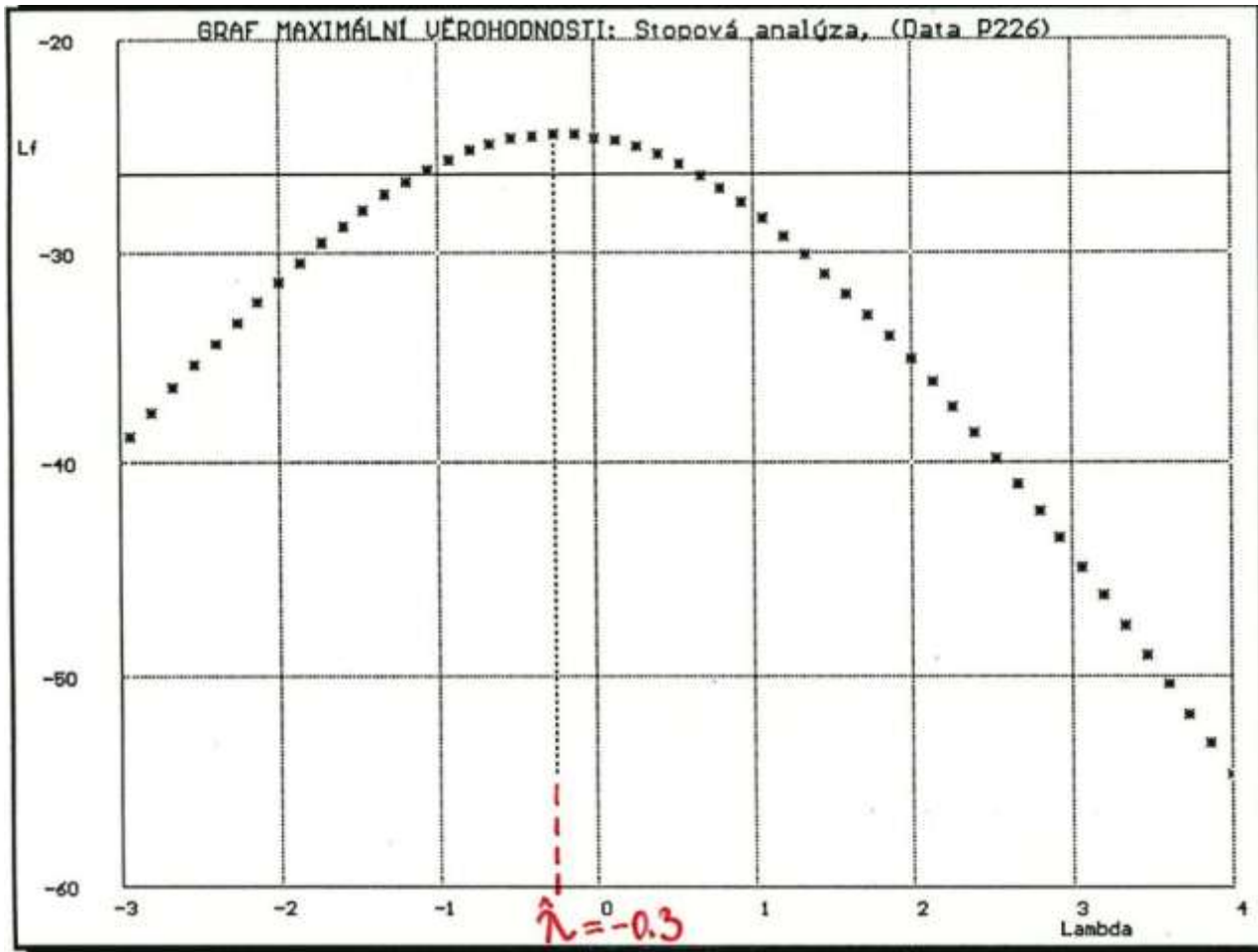
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





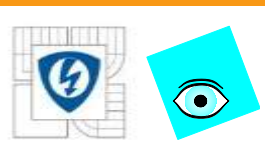
Graf maximální věrohodnosti: stopová analýza (G22) analýza



31.1.2011

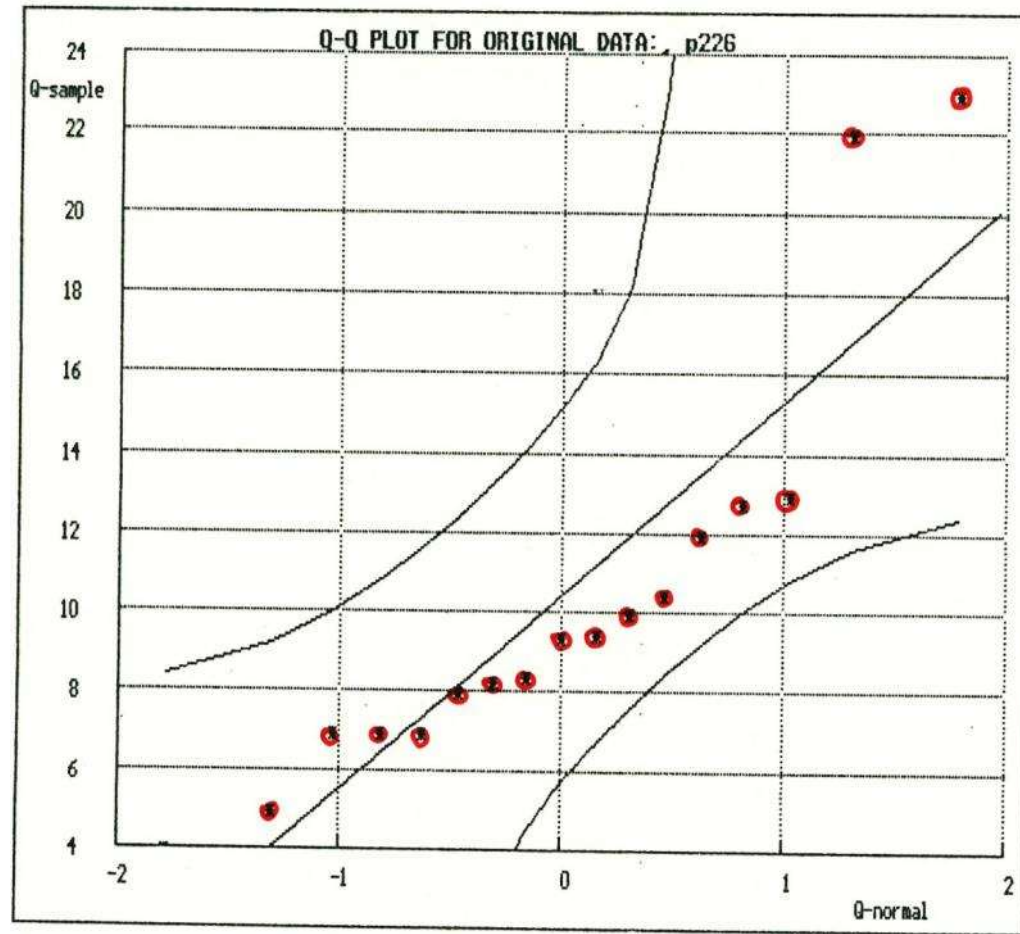
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Q-Q graf pro původní data

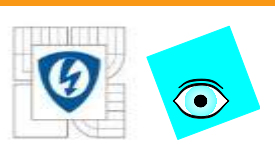
(G13)



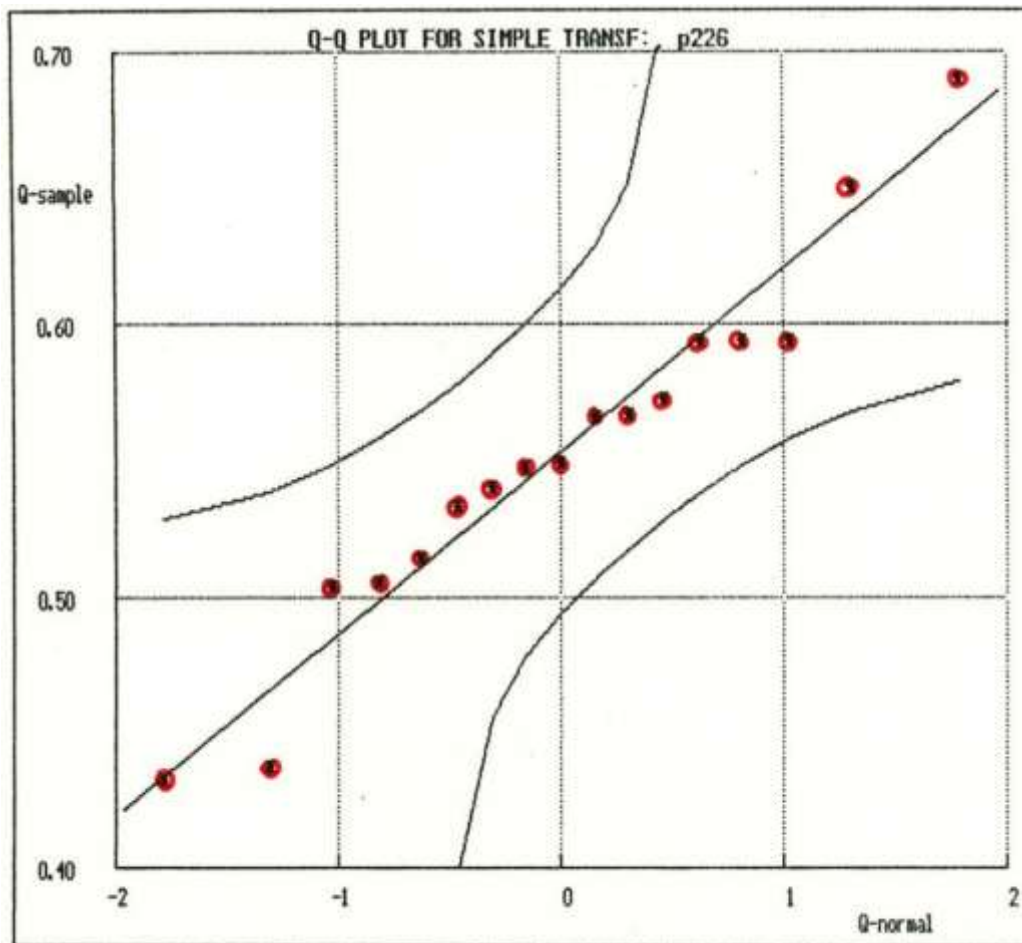
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





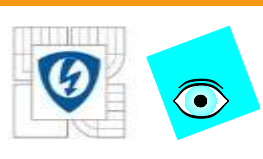
Q-Q graf pro jednoduchou transformaci (G13)



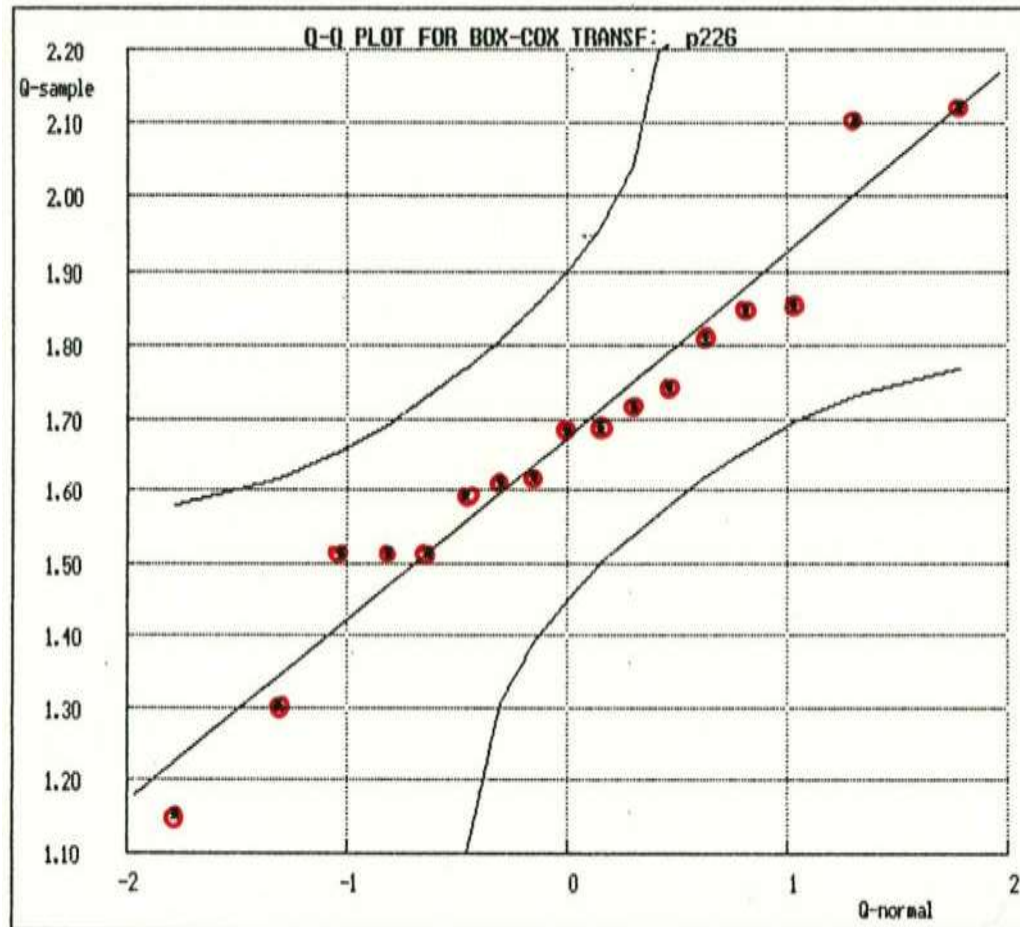
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





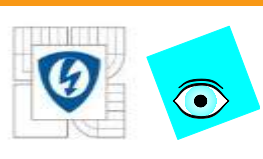
Q-Q graf pro Box-Coxovu transformaci (G13)



31.1.2011

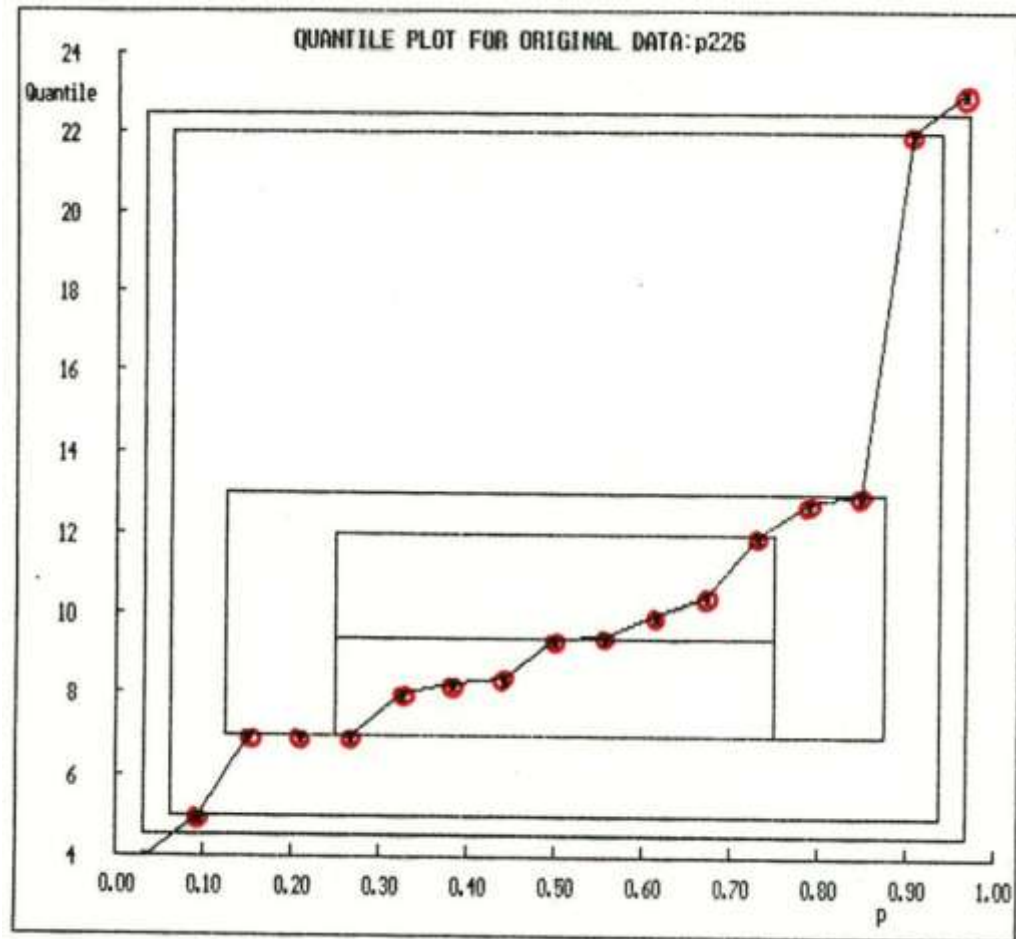
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Kvantilový graf pro původní data

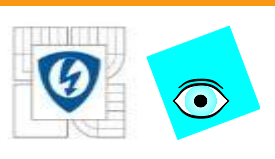
(G10)



31.1.2011

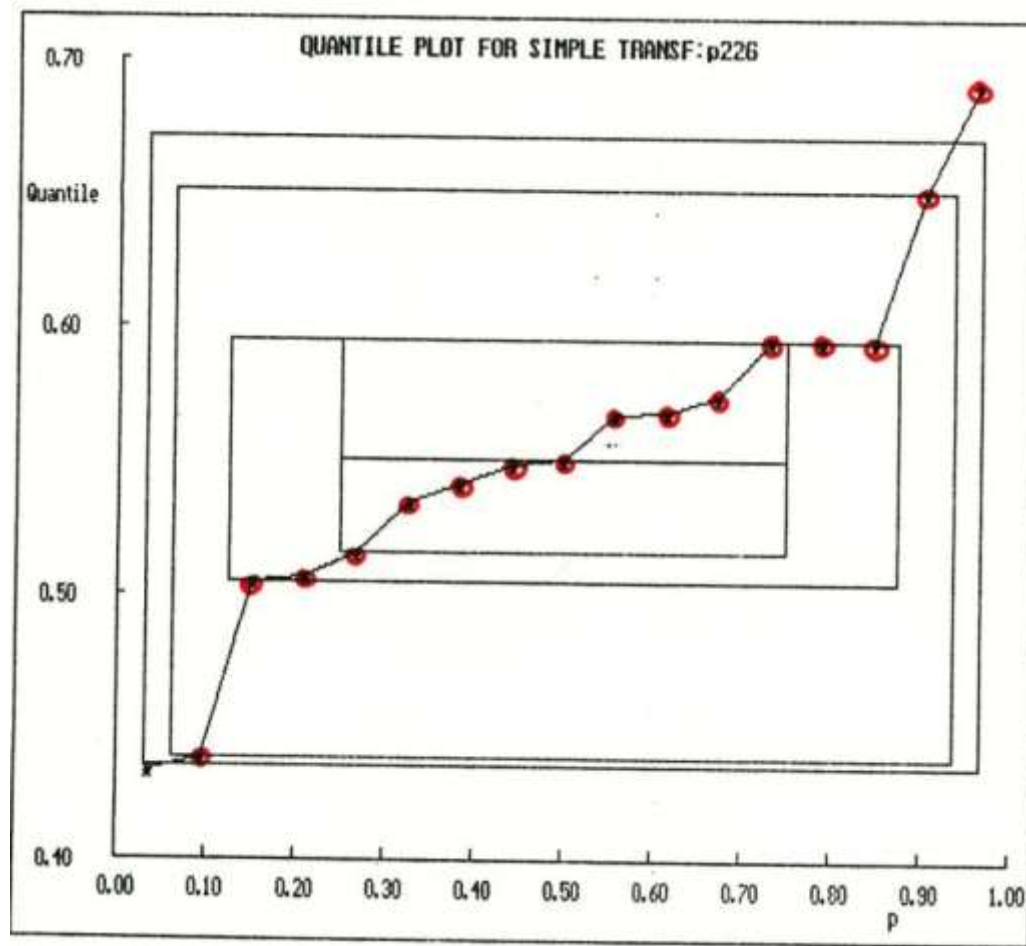
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Kvantilový graf pro jednoduchou transformaci

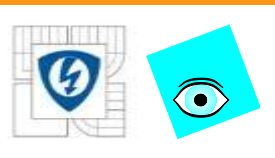
(G10)



31.1.2011

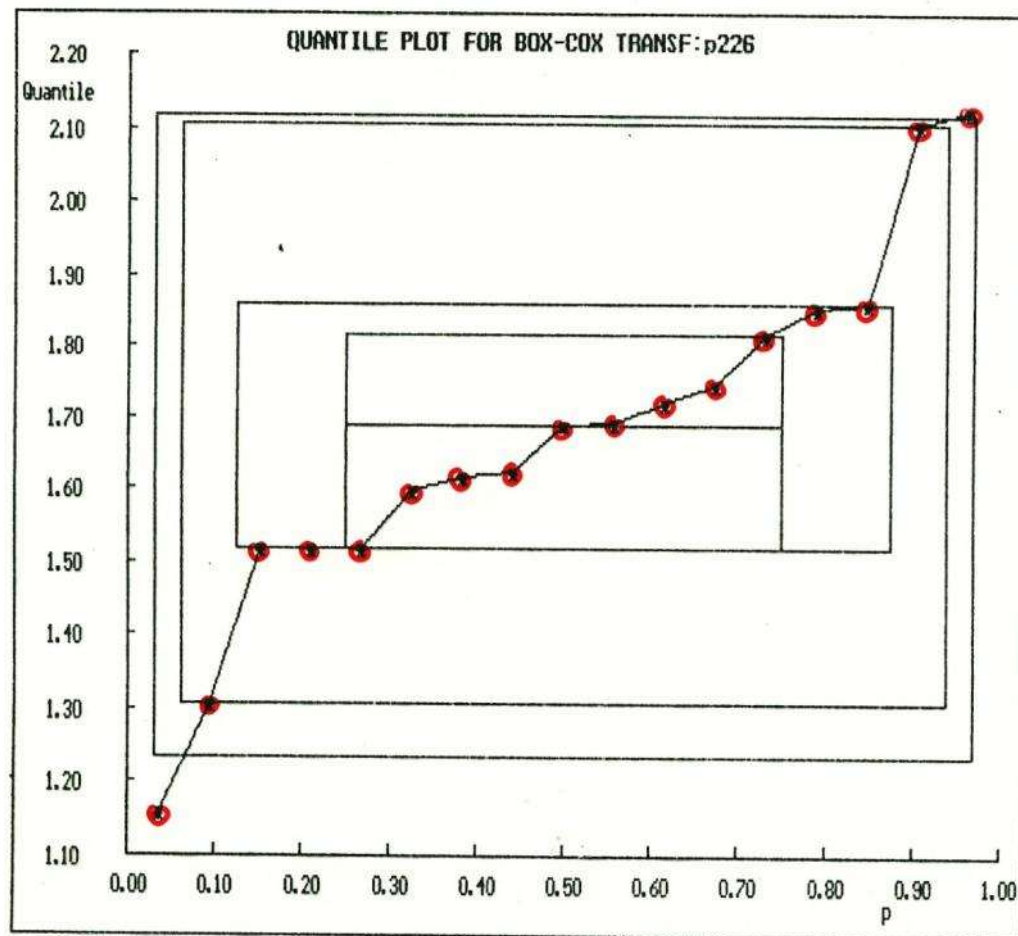
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Kvantilový graf pro Box-Coxovy transformaci

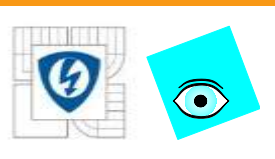
(G10)



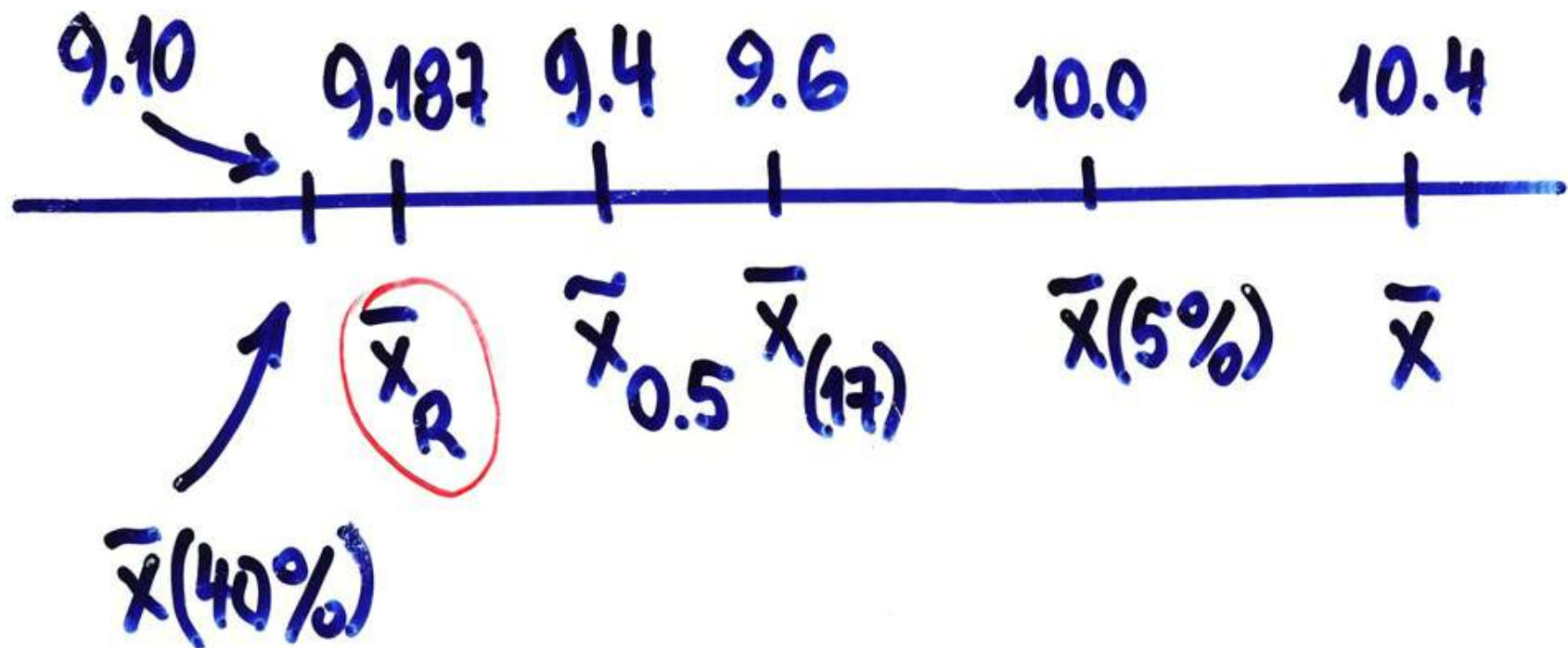
31.1.2011

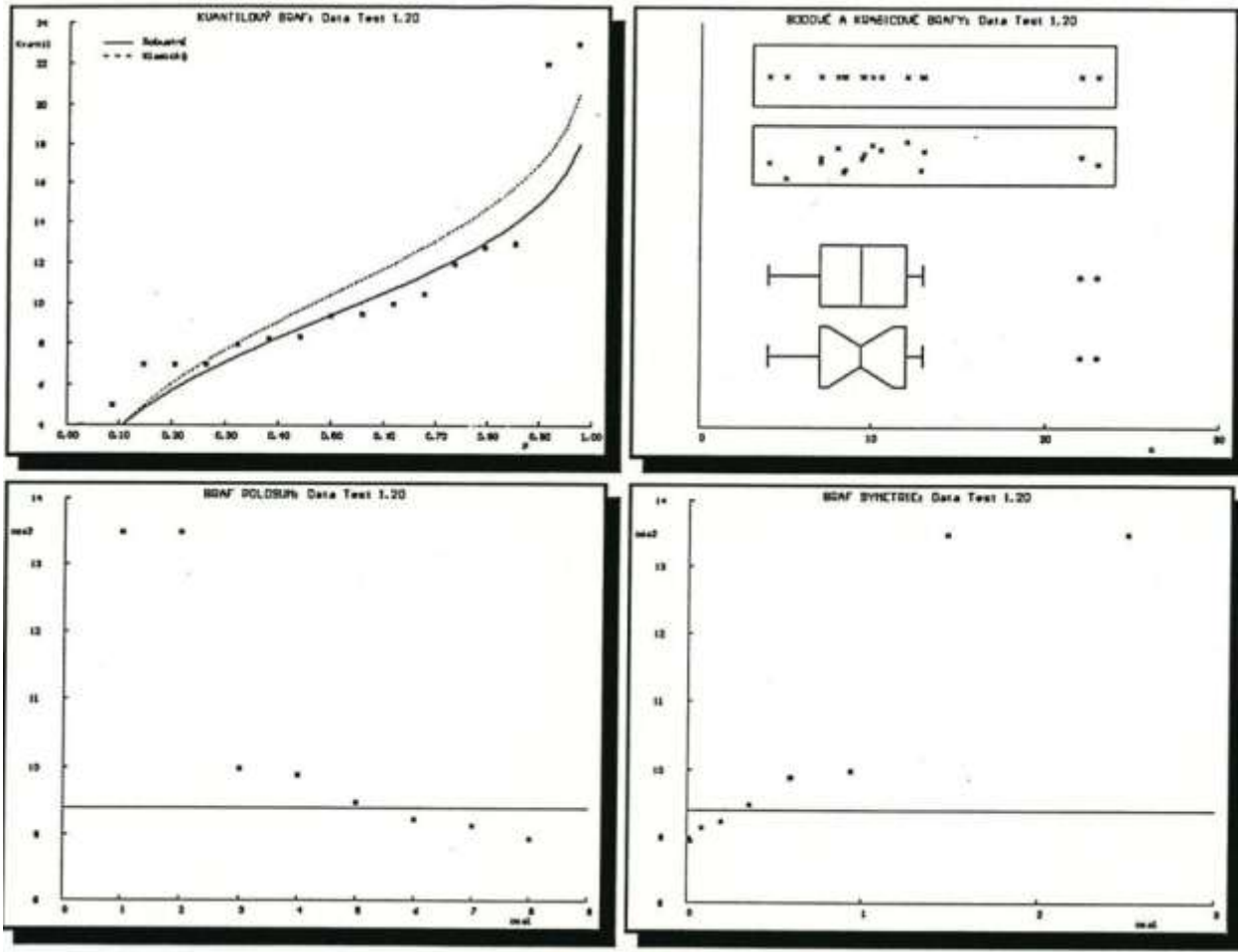
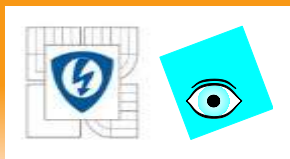
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



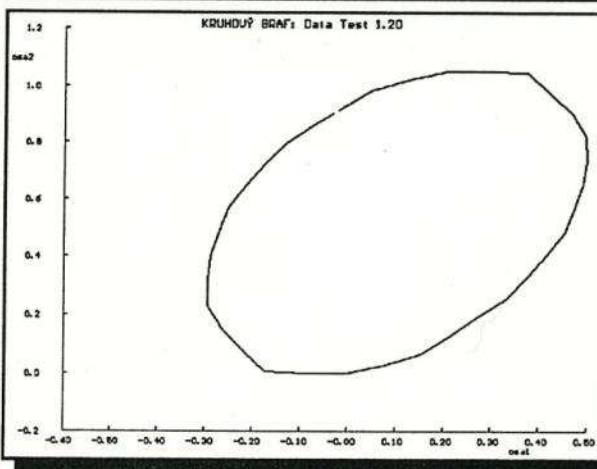
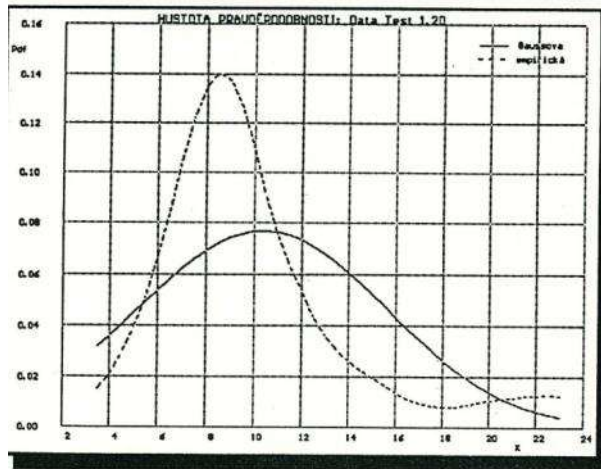
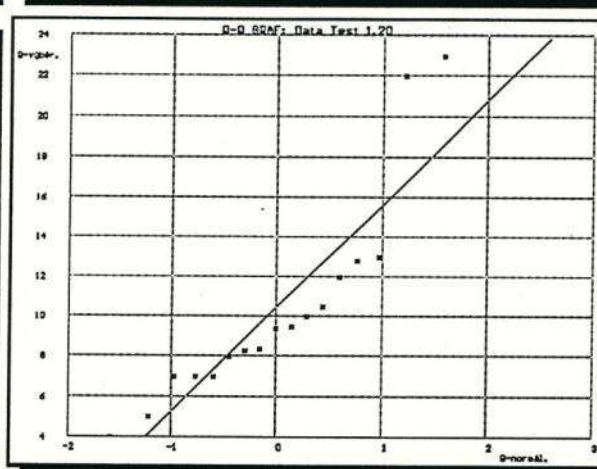
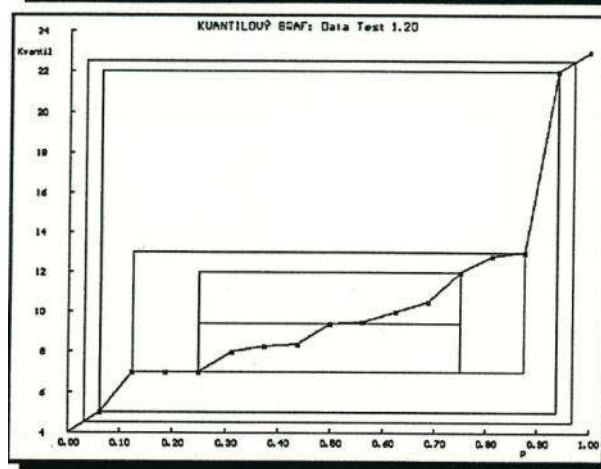
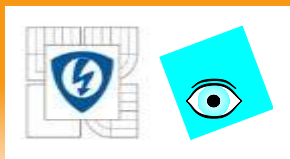


Střední hodnota





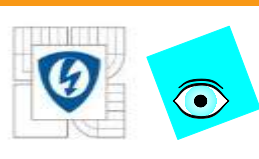
31.1.2011



31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Příklad

Basic Assumptions

Title: p226

I N P U T

(1) CONDITIONS:

Sample size, n : 17
Significance level, alpha: 0.050
Output file name : P226A.TXT

(2) INPUT OF ORIGINAL DATA:

```
( 1) 4.0000E+00 ( 6) 8.0000E+00 ( 11) 1.0000E+01 ( 16) 2.2000E+01
( 2) 7.0000E+00 ( 7) 8.3000E+00 ( 12) 1.0500E+01 ( 17) 2.3000E+01
( 3) 7.0000E+00 ( 8) 8.4000E+00 ( 13) 1.2000E+01
( 4) 7.0000E+00 ( 9) 9.4000E+00 ( 14) 1.2800E+01
( 5) 5.0000E+00 ( 10) 9.5000E+00 ( 15) 1.3000E+01
```

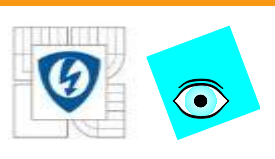
O U T P U T

(1) CLASSICAL ESTIMATORS OF LOCATION AND SHAPE:

Mean	: 1.0406E+01	Variance	: 2.6834E+01
Standard deviation	: 5.1802E+00	Skewness	: 1.3991E+00
Kurtosis	: 4.2721E+00		

(2) TEST FOR NORMALITY:

Quantile, $\chi^2(1-\alpha, 2)$	: 5.9915E+00
χ^2 statistics	: 7.9084E+00
Conclusion: Normality rejected	
Computed significance level	: 1.9174E-02



Pokračování příkladu

(3) TEST FOR INDEPENDENCE:

Quantile, $t(1-\alpha/2, n+1)$	:	2.1009E+00
t statistics	:	1.0381E+00
Conclusion: Independence accepted		
Computed significance level	:	1.5649E-01

Assumption of sample homogeneity:

Arithmetic mean	:	1.0406E+01
Variance	:	2.6834E+01
Standard deviation	:	5.1802E+00

Internal bounds:

Lower bound:	-3.1909E+00
Upper bound:	2.2191E+01

(4) MINIMAL SAMPLE SIZES:

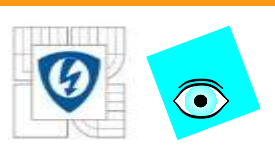
for 25% relative error of standard deviation:	$n = 14$
for 10% relative error of standard deviation:	$n = 83$
for 5% relative error of standard deviation:	$n = 328$

(5) OUTLIERS DETECTION:

point number 17 (high)	:	2.3000E+01
Number of outliers	:	1

Measures with excluded outliers

Mean	:	9.6188E+00	Variance	:	1.7388E+01
Standard deviation	:	4.1699E+00	Skewness	:	1.6101E+00
Kurtosis	:	6.3399E+00			



Pokračování příkladu

ADSTAT (c) Trilobyte
UNIVARIATE DATA ANALYSIS
One Sample Analysis

Title: p226

I N P U T

(1) CONDITIONS:

Sample size, n : 17
Significance level, alpha: 0.050
Output file name : P226.TXT

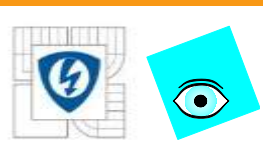
(2) INPUT OF ORIGINAL DATA:

(1)	4.0000E+00	(6)	8.0000E+00	(11)	1.0000E+01	(16)	2.2000E+01
(2)	7.0000E+00	(7)	8.3000E+00	(12)	1.0500E+01	(17)	2.3000E+01
(3)	7.0000E+00	(8)	8.4000E+00	(13)	1.2000E+01		
(4)	7.0000E+00	(9)	9.4000E+00	(14)	1.2800E+01		
(5)	5.0000E+00	(10)	9.5000E+00	(15)	1.3000E+01		

O U T P U T

(1) ESTIMATORS OF DISTRIBUTION SHAPE:

Skewness : 1.3991E+00
Kurtosis : 4.2721E+00



Pokračování příkladu

(2) CLASSICAL ESTIMATORS OF LOCATION AND SPREAD (Normality assumed):

Mean : 1.0406E+01
Standard deviation : 5.1802E+00
Variance : 2.6834E+01
95.0 % confidence interval:
Lower limit: 7.7425E+00 Upper limit: 1.3069E+01

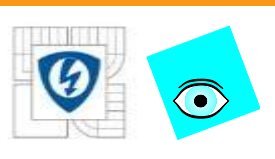
(3) OTHER ESTIMATORS OF LOCATION:

Mode estimate : 7.5000E+00
Mid-Sum estimate : 1.3500E+01

(4) ROBUST STATISTICS OF LOCATION AND SPREAD (Normality not assumed):

Median measures:

Median : 9.4000E+00
Standard deviation : 4.4596E+00
Variance : 1.9888E+01
Standard deviation : 2.0494E+00
Variance (nonpar.) : 1.4316E+00
Standard deviation : 1.0003E+00
Variance (Marritz) : 1.0001E+00
95.0 % confidence interval:
Lower limit: 7.2798E+00 Upper limit: 1.1520E+01



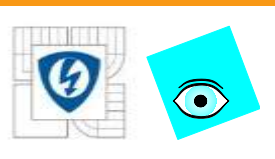
Pokračování příkladu

Trimming 5% (for $P=0.05$):

Mean : 1.0062E+01
Standard deviation : 5.5644E+00
Variance : 3.0963E+01
Winsorized mean : 1.0406E+01
Winsorized
Standard deviation : 5.1209E+00
Winsorized variance: 2.6223E+01
95.0 % confidence interval:
Lower limit: 7.1450E+00 Upper limit: 1.2979E+01

Trimming 10% (for $P=0.10$):

Mean : 9.6324E+00
Standard deviation : 3.8390E+00
Variance : 1.4738E+01
Winsorized mean : 1.0406E+01
Winsorized
Standard deviation : 5.1209E+00
Winsorized variance: 2.6223E+01
95.0 % confidence interval:
Lower limit: 6.7152E+00 Upper limit: 1.2549E+01



Pokračování příkladu

Trimming 40% (for $P=0.40$):

Mean : 9.1059E+00
Standard deviation : 3.5294E+00
Variance : 1.2457E+01
Winsorized mean : 9.1412E+00
Winsorized
Standard deviation : 1.4739E+00
Winsorized variance: 2.1722E+00
95.0 % confidence interval:
Lower limit: 7.2116E+00 Upper limit: 1.1000E+01

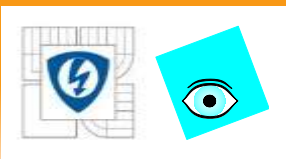
Biweight:

Mean : 8.8034E+00
Standard deviation : 2.5962E+00
Variance : 6.7403E+00
Weight sqrt(w) : 3.7120E+00
95.0 % confidence interval:
Lower limit: 7.3207E+00 Upper limit: 1.0286E+01

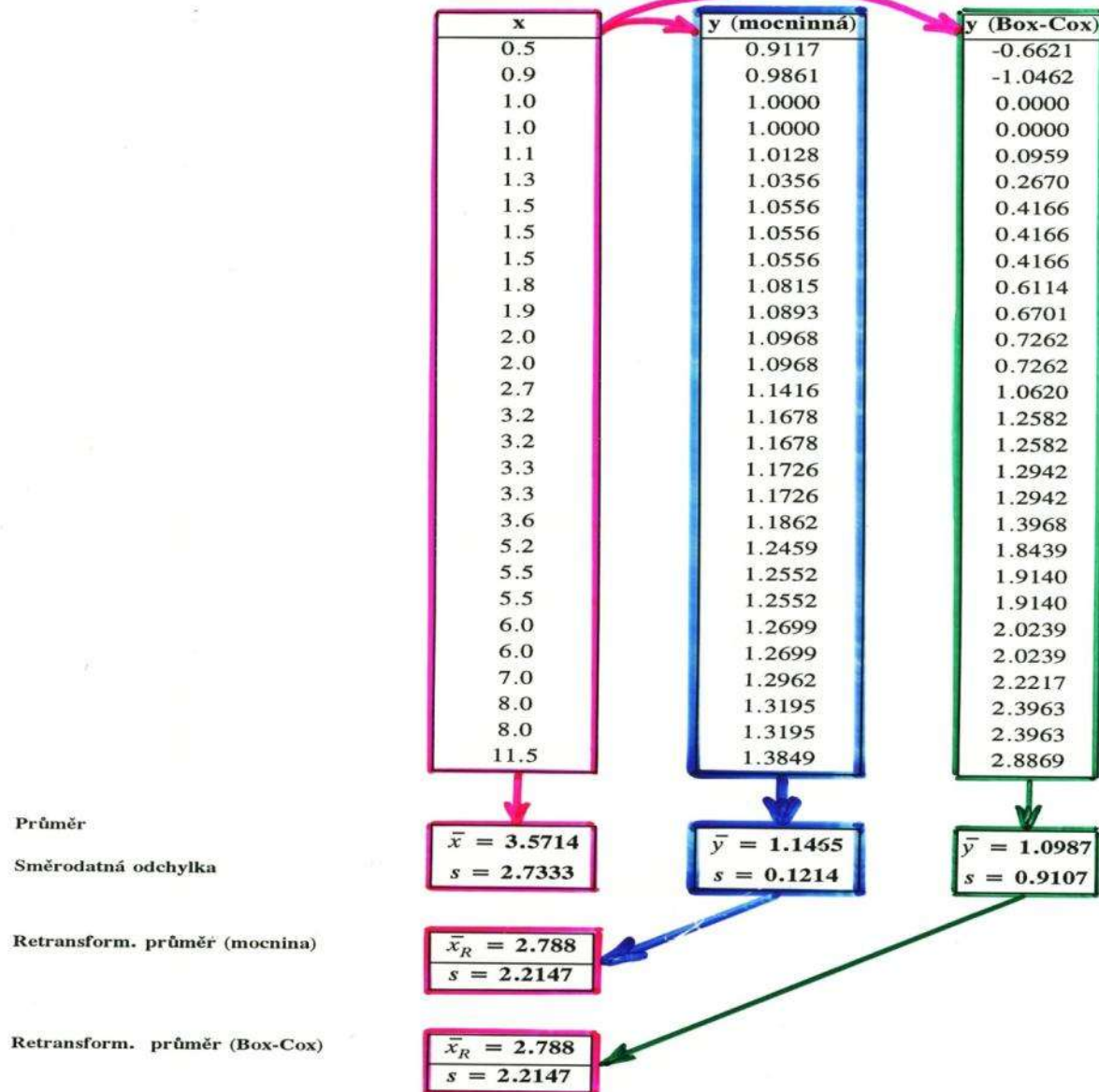
(5) ADAPTIVE ESTIMATORS OF LOCATION AND SPREAD:

Hoog's estimates:

Relative tail length: 2.7797E+00
Mean : 9.3353E+00
Standard deviation : 3.7223E+00
Variance : 1.3855E+01
95.0 % confidence interval:
Lower limit: 7.4215E+00 Upper limit: 1.1249E+01



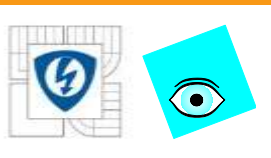
Schema mocninné a Box-Coxovy transformace





Schema mocninné a Box-Coxovy transformace

x	y (mocninná)	y (Box-Cox)
0.5	0.9117	-0.6621
0.9	0.9861	-1.0462
1.0	1.0000	0.0000
1.0	1.0000	0.0000
1.1	1.0128	0.0959
1.3	1.0356	0.2670
1.5	1.0556	0.4166
1.5	1.0556	0.4166
1.5	1.0556	0.4166
1.8	1.0815	0.6114
1.9	1.0893	0.6701
2.0	1.0968	0.7262
2.0	1.0968	0.7262
2.7	1.1416	1.0620
3.2	1.1678	1.2582
3.2	1.1678	1.2582
3.3	1.1726	1.2942
3.3	1.1726	1.2942
3.6	1.1862	1.3968
5.2	1.2459	1.8439
5.5	1.2552	1.9140
5.5	1.2552	1.9140
6.0	1.2699	2.0239
6.0	1.2699	2.0239
7.0	1.2962	2.2217
8.0	1.3195	2.3963
8.0	1.3195	2.3963
11.5	1.3849	2.8869

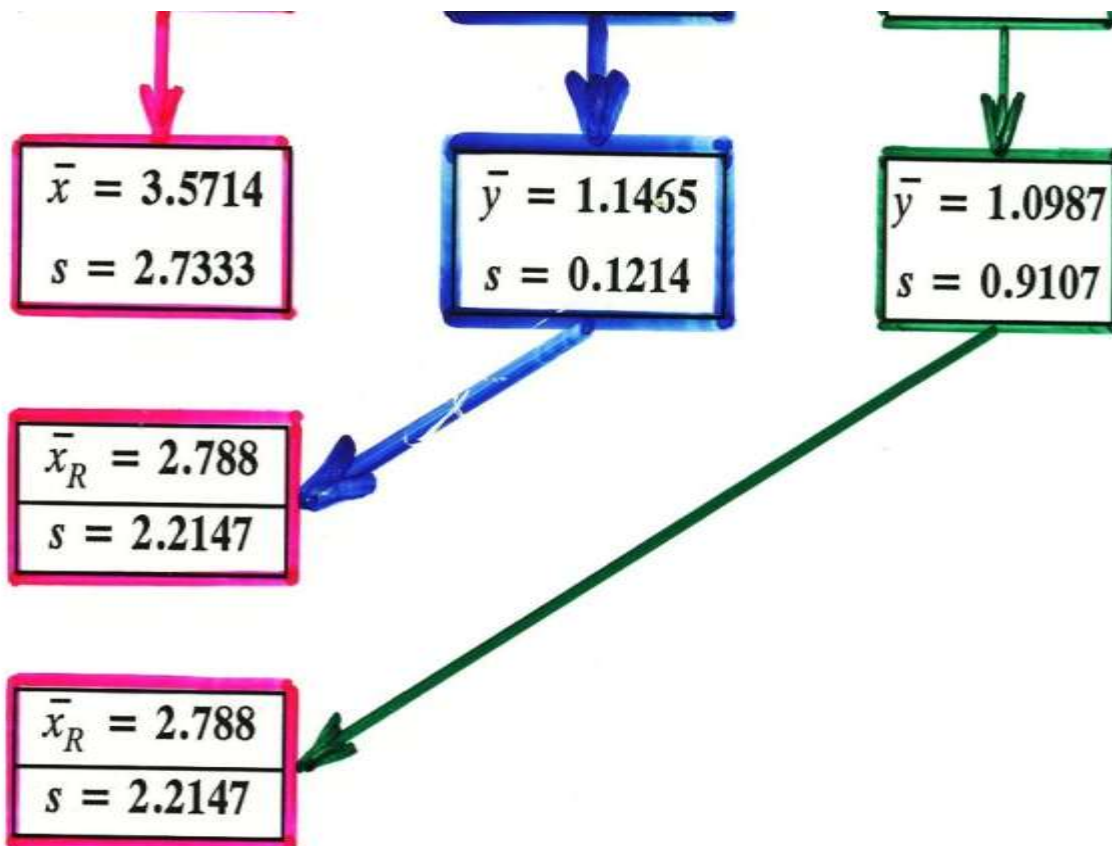


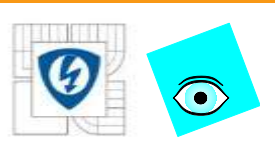
Průměr

Směrodatná odchylka

Retransform. průměr (mocnina)

Retransform. průměr (Box-Cox)





Závěr

(určení obsahu kadmia v bramborách, odhady nutno vynásobit rozměrem 10^{-2} [mg/kg])

Klasické odhady: (aritmetický průměr a směrodatná odchylka) $\bar{x} = 3.57, s = 2.73, L_D = 2.51, L_H = 4.63$

Retransformované odhady:

1. Symetrizující mocninná transformace

$\hat{\lambda} = 0.13$ dává $\bar{x}_R = 2.79, s_R = 2.21, L_D = 2.04, L_H = 3.77$

2. Normalizující Box-Coxova transformace

$\hat{\lambda} = 0.13$ dává $\bar{x}_R = 2.79, s_R = 2.21, L_D = 2.04, L_H = 3.77$

3. Exponenciální transformace

$\hat{\lambda} = 1.07$ dává $\bar{x}_R = 2.67, s_R = *.**, L_D = 2.03, L_H = 3.53$

31.1.2011 (*.** značí nevyčísleno)