Katedra štatistiky,
Fakulta hospodárskej informatiky,
Ekonomická univerzita v Bratislave

Pravdepodobnostné modelovanie inverznými distribučnými funkciami : Identifikácia kvantilového modelu

Ľubica SIPKOVÁ

Apríl 2009

3. z cyklu prezentácií

Fázy kvantilového modelovania

Identifikácia

Vystihnutie tvaru empirického rozdelenia kvantilovým tvarom

Estimácia

Odhad parametrov kvantilového modelu

Verifikácia

Overenie vhodnosti kvantilového modelu

Proces interaktívny, revidujúci predchádzajúce výsledky novými zisteniami s vrátením sa s novými informáciami

k predchádzajúcim fázam, postupom a technikám.

Identifikácia

Vystihnutie tvaru empirického rozdelenia

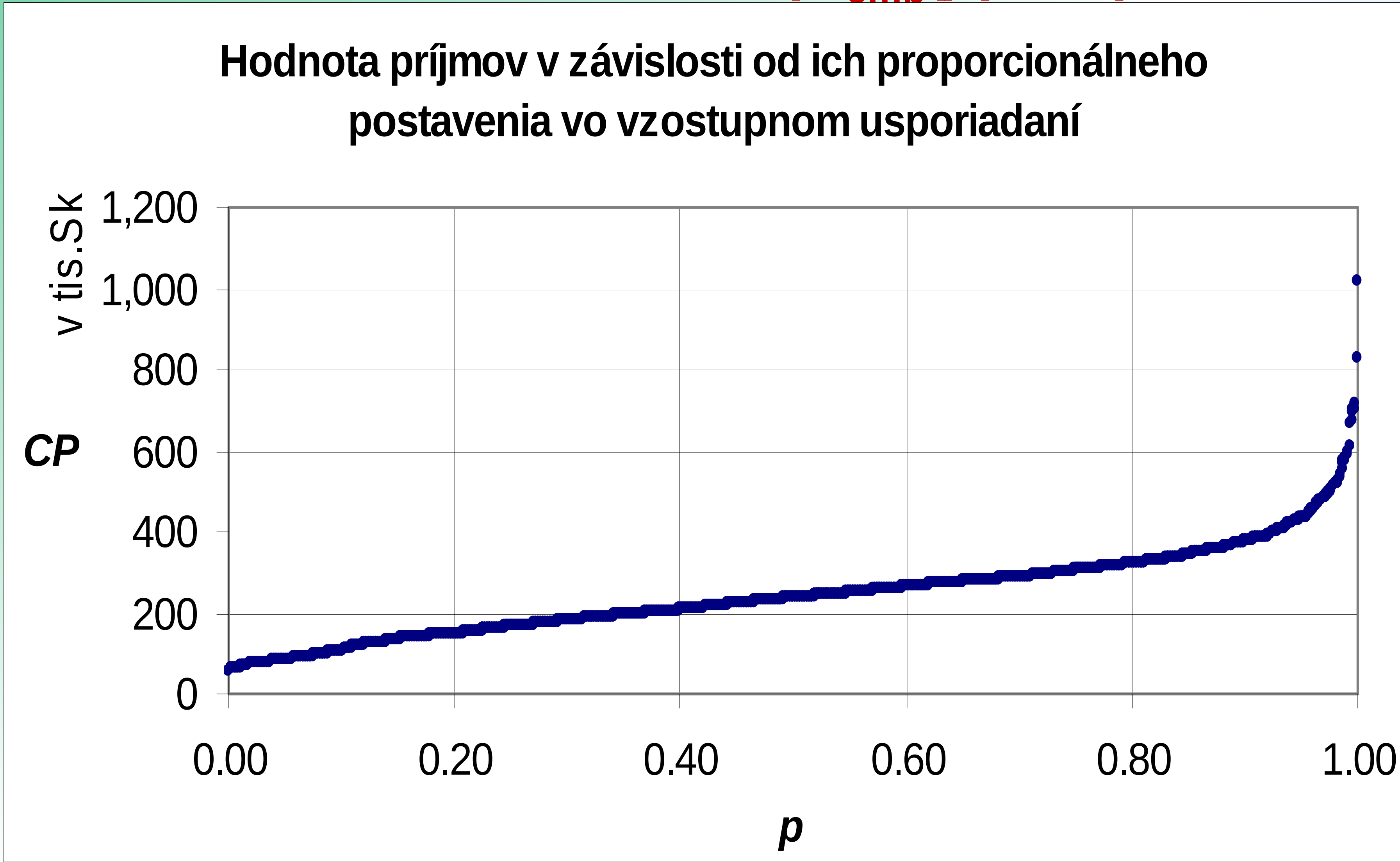
- Grafická identifikácia empirického rozdelenia

- Kvantitatívna analýza empirického rozdelenia:
  - na momentovom základe
  - na kvantilovom základe

- Identifikácia rozdelenia jednoduchými tvarmi
  - Celého rozdelenia
  - Zvlášť pre dolný a horný koniec

- Voľba tvaru váh pre konce rozdelenia
  - Ako funkcia *p*
  - Ako parametre modelu

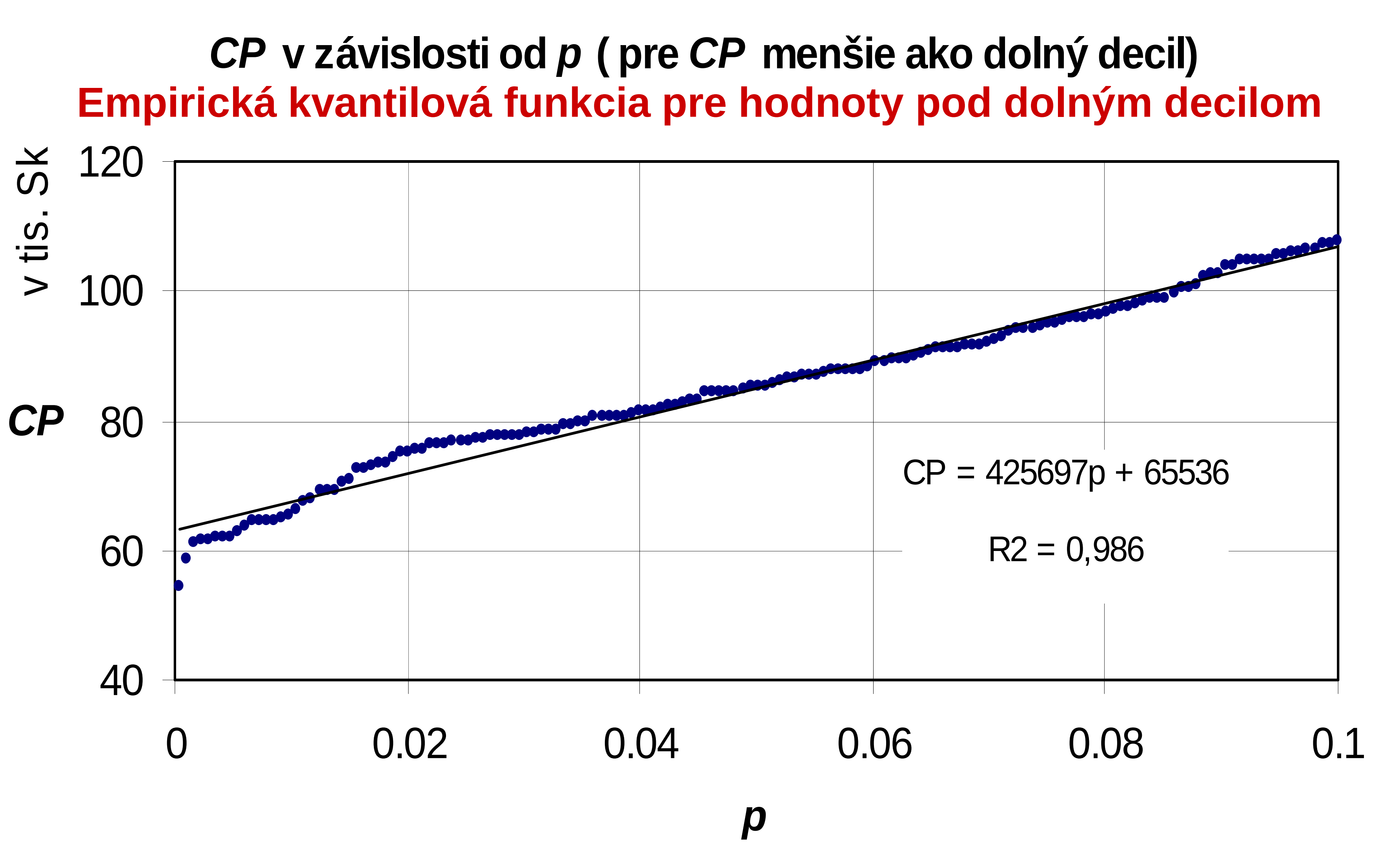
Grafická analýza celého aj častí empirického rozdelenia ($Q_{emp}(p)=CP$)



Grafická analýza empirického rozdelenia

Príjmy domácností SR v roku 2003

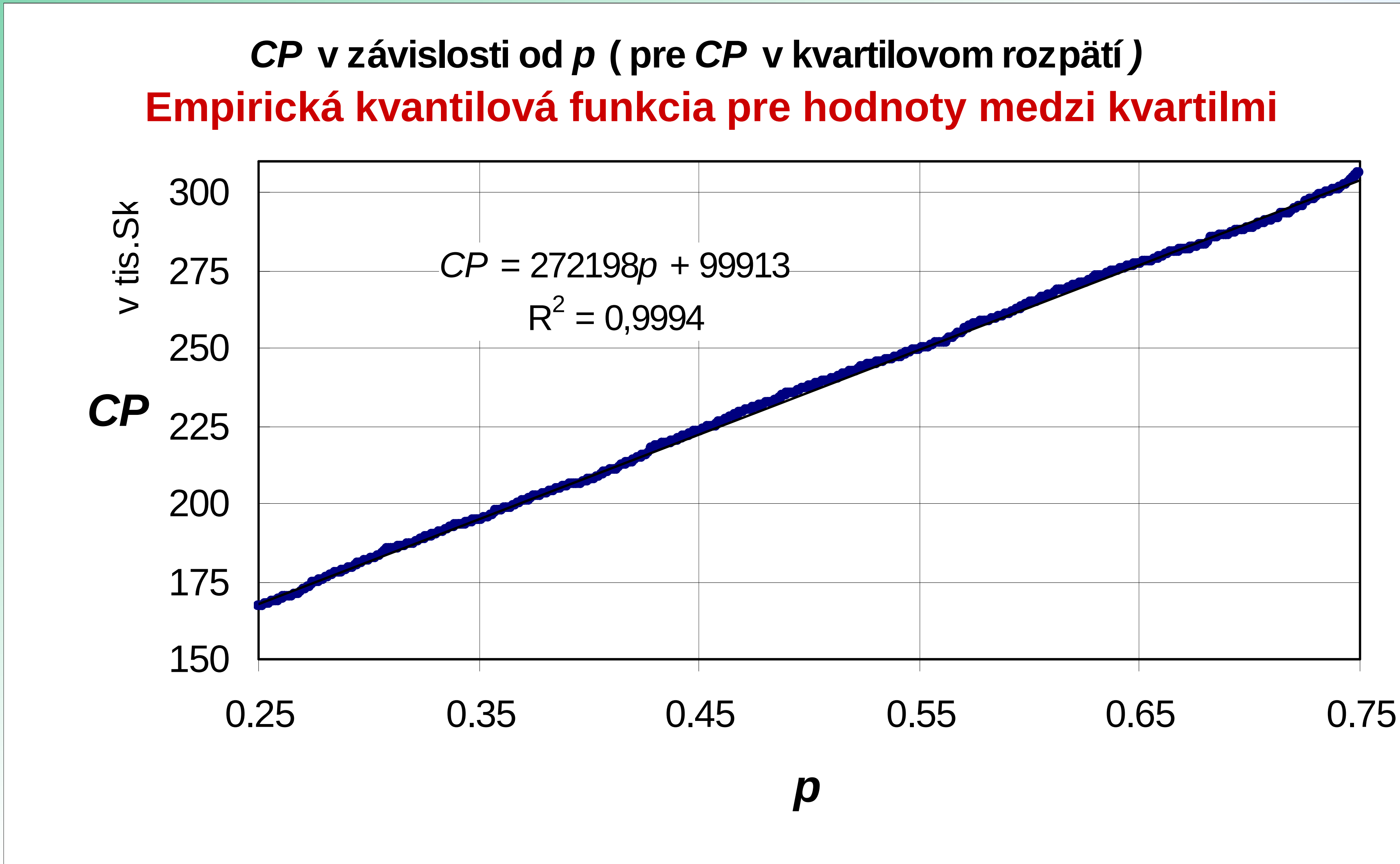
($Q_{emp}(p)=CP$)



Grafická analýza empirického rozdelenia

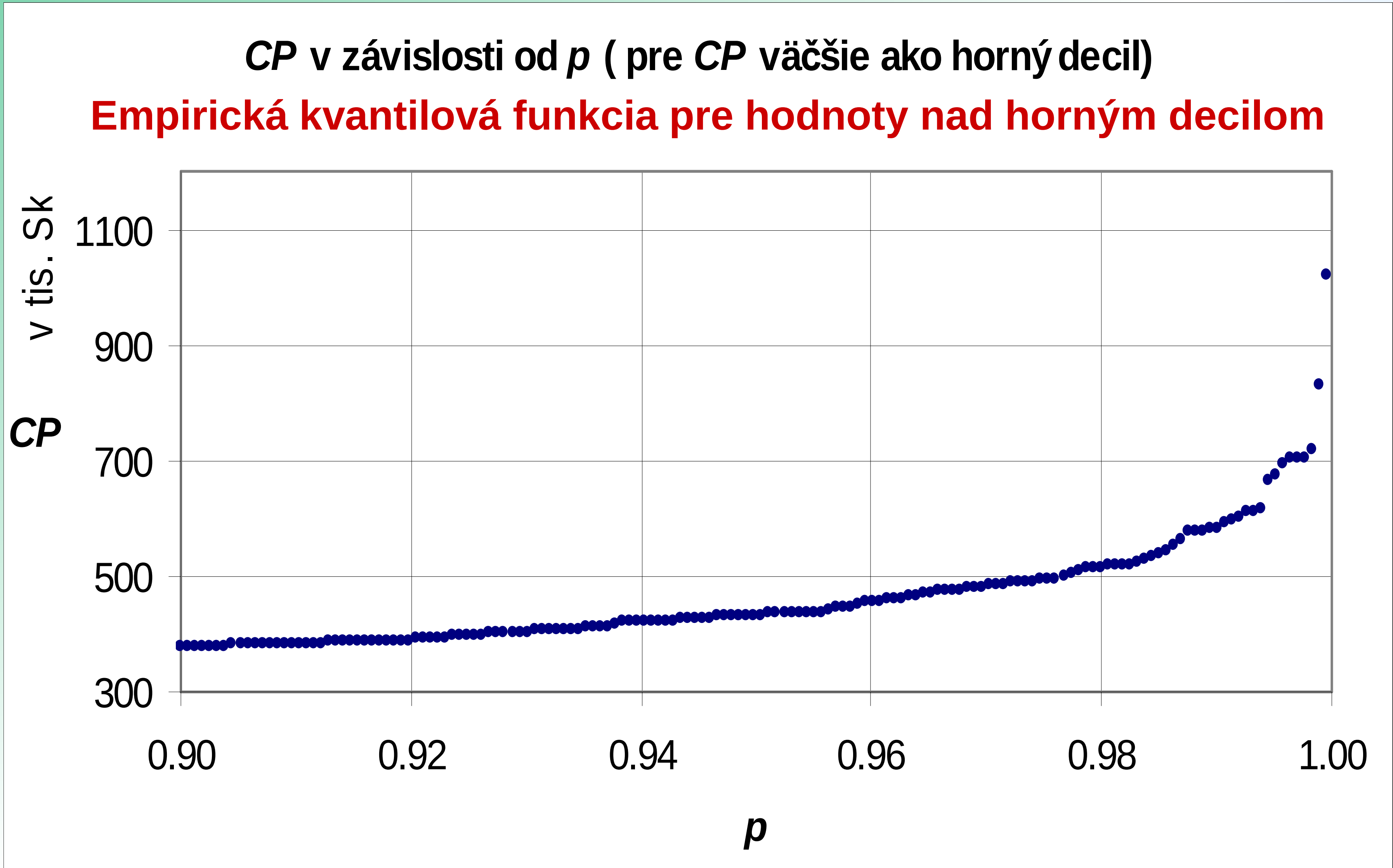
Príjmy domácností SR v roku 2003

($Q_{emp}(p)=CP$)



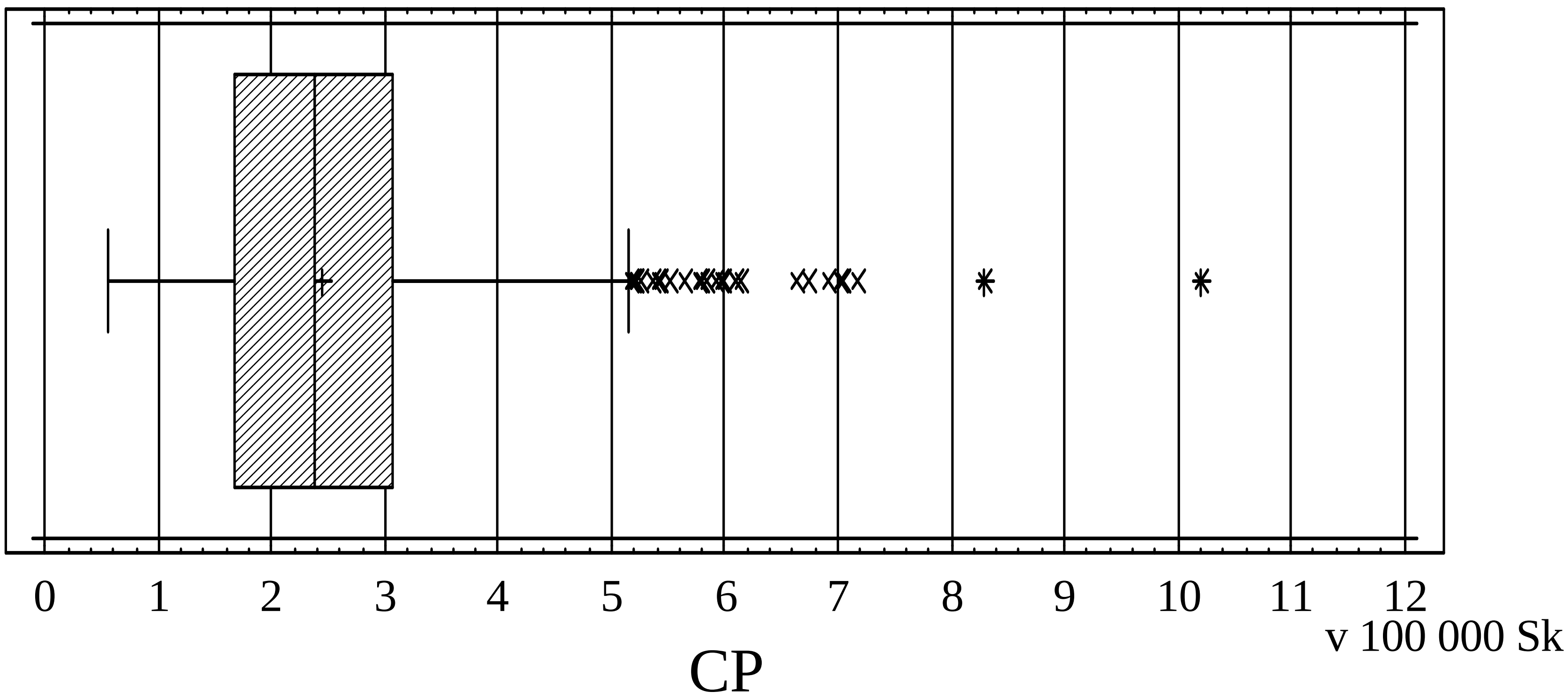
Grafická analýza empirického rozdelenia

Príjmy domácností SR v roku 2003 **($Q_{emp}(p)=CP$)**



Grafická analýza celého aj častí empirického rozdelenia (Box-Plot)

Box-and-Whisker graf pre CP

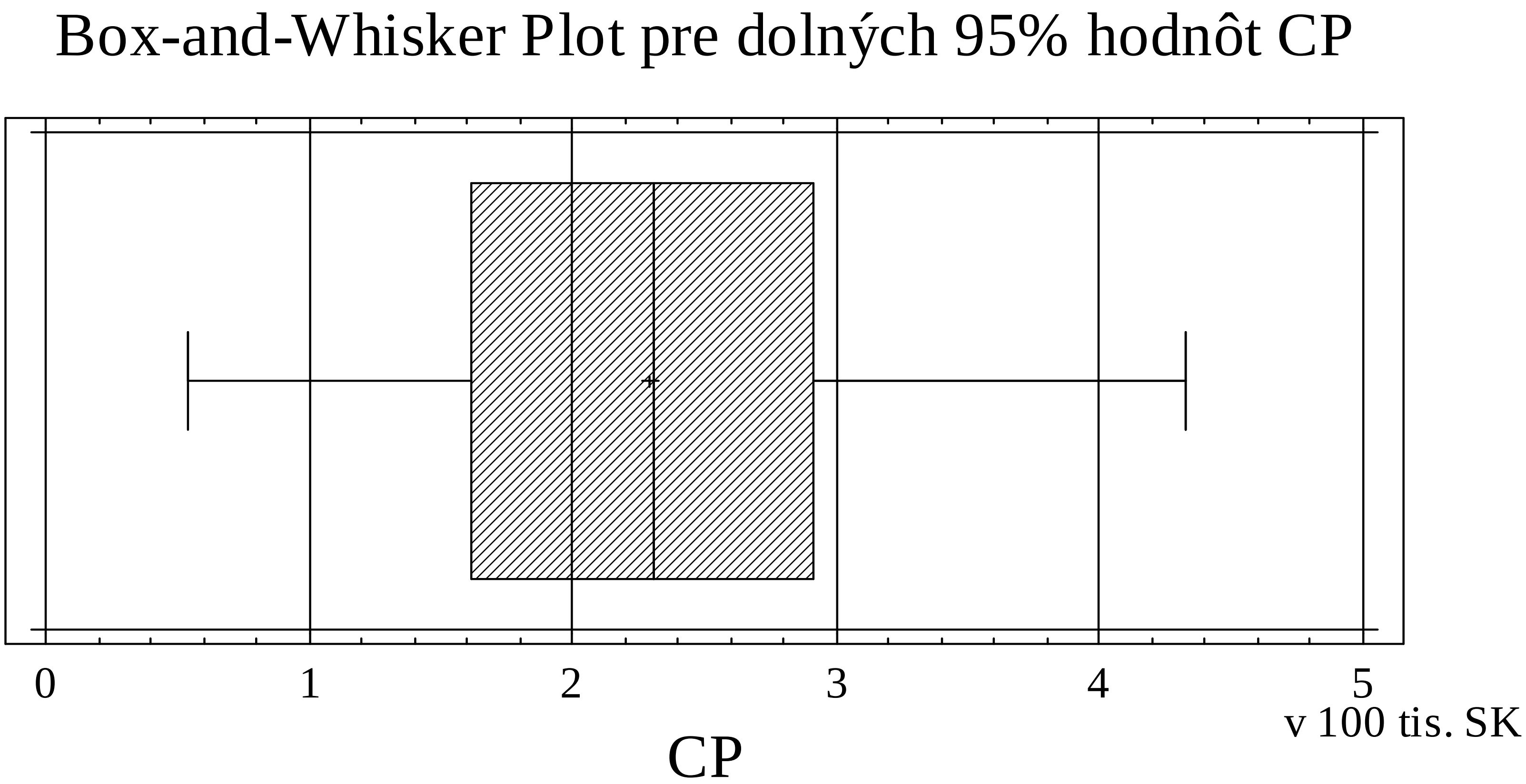


© Ľubica Sipková

Grafická analýza empirického rozdelenia

Príjmy domácností SR v 2003

(Box-Plot)



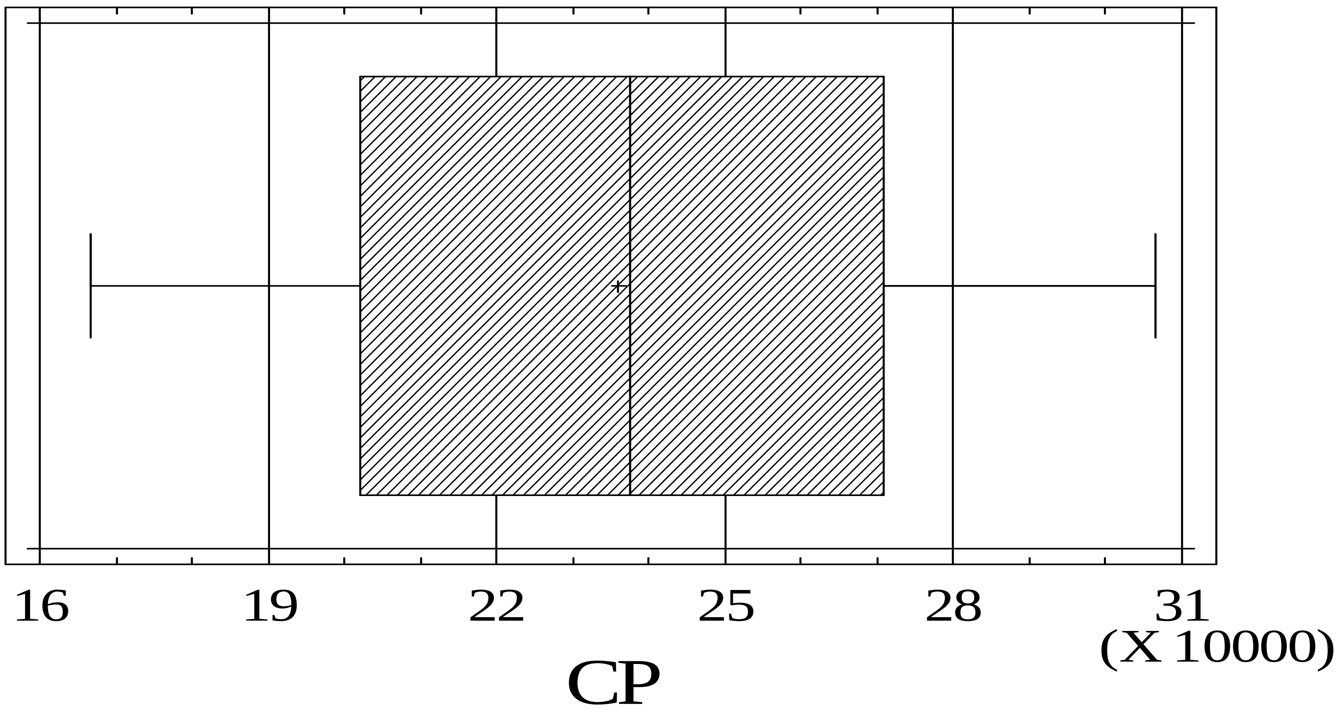
Takmer symetrické rozdelenie bez horných 5% hodnôt

Grafická analýza empirického rozdelenia

Príjmy domácností SR v 2003

(Box-Plot)

Box-and-Whisker Plot pre prostrednú polovicu hodnôt CP

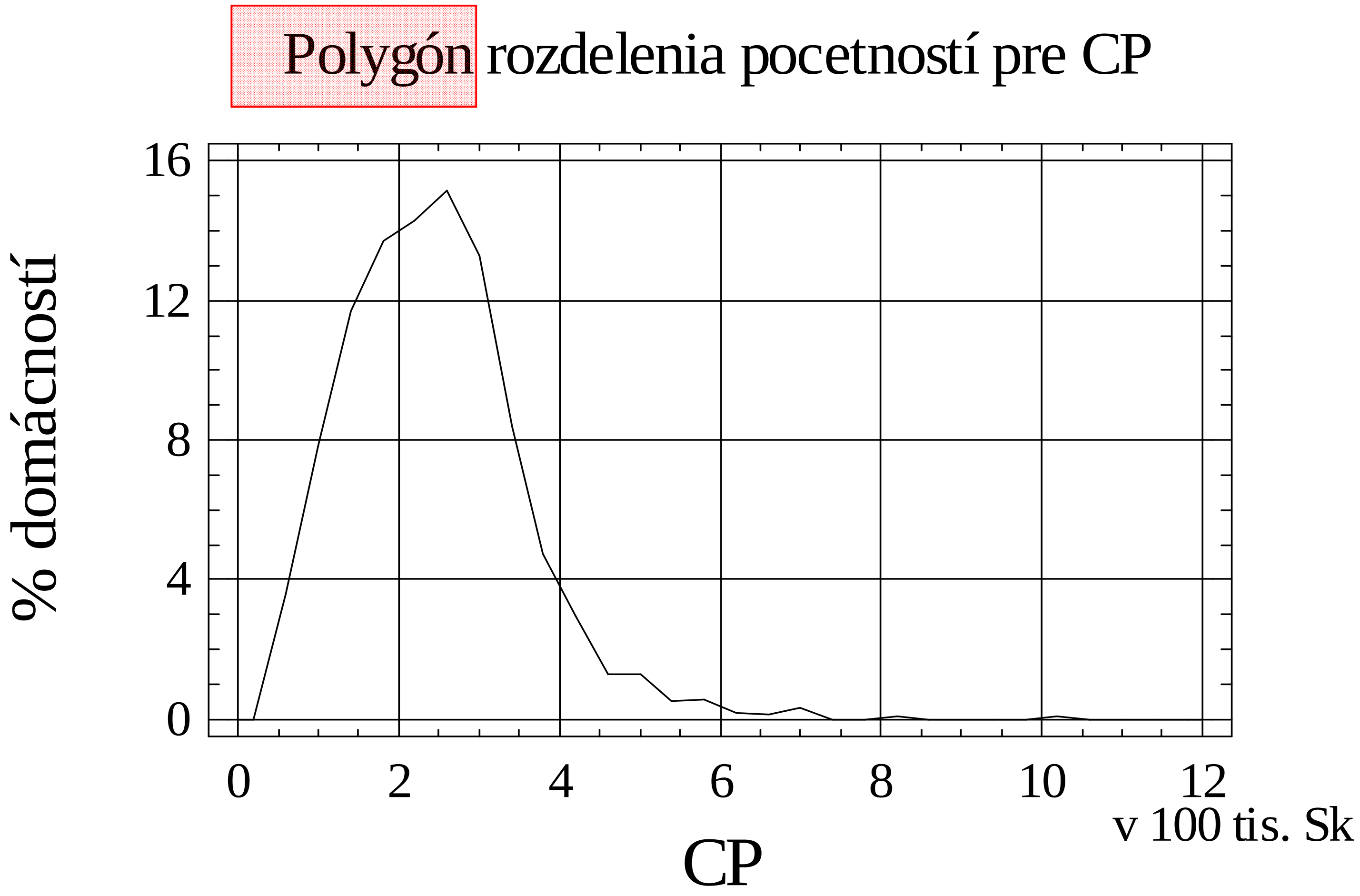


Symetrická prostredná časť medzi kvartilmi

Grafická analýza empirického rozdelenia

Príjmy domácností SR v 2003

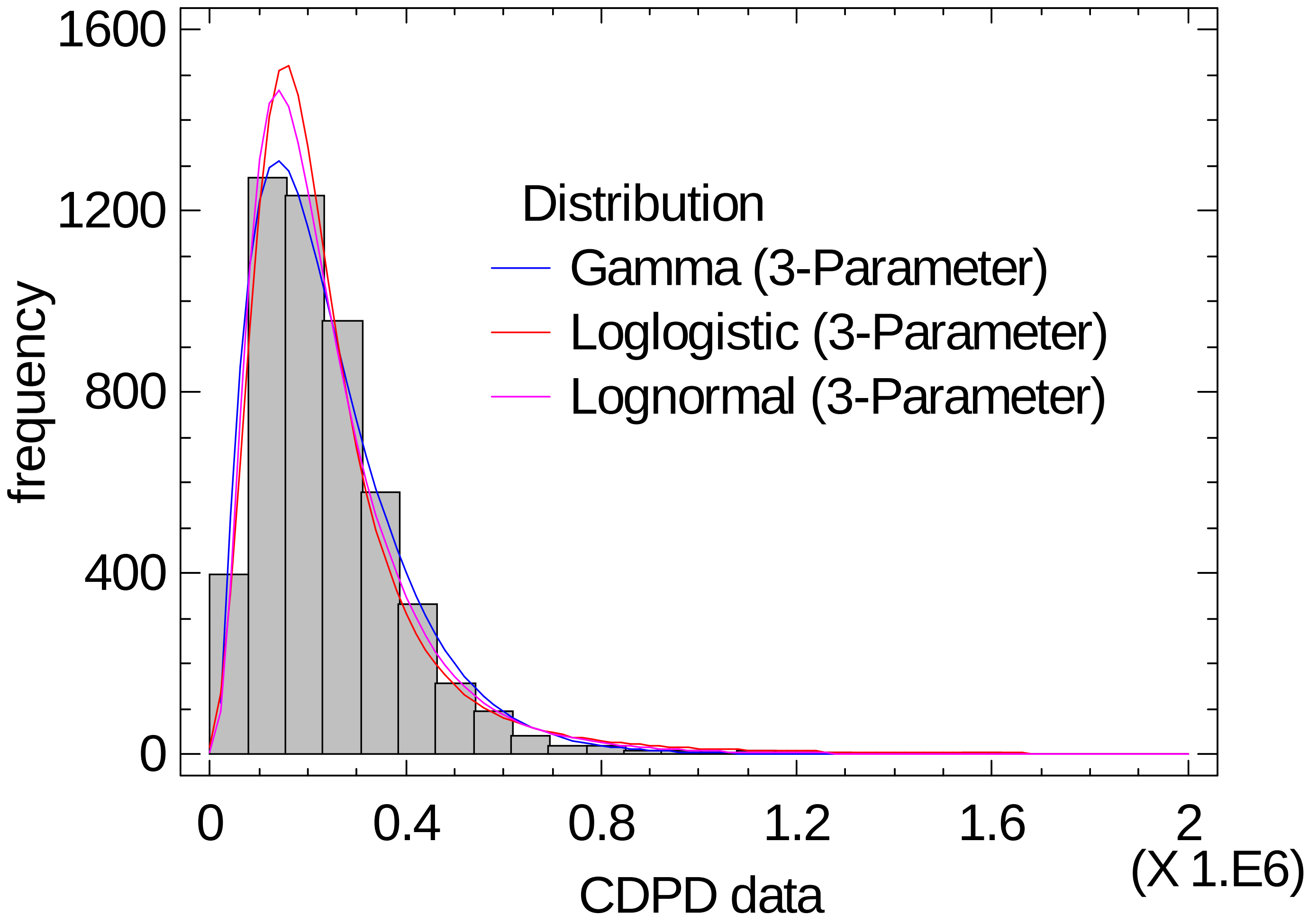
Polygón



Porovnanie zhody so známymi jednoduchými tvarmi pomocou procedúry: Comparison of Alternative Distributions

Distribution	*Est. Parameters*	*Log Likelihood*	*KS D*
Lognormal (3-Parameter)	**3**	**-67677.7**	**0.0362935**
Lognormal	**2**	**-67704.9**	**0.0397763**
Loglogistic	**2**	**-67730.3**	**0.0496562**
Gamma (3-Parameter)	**3**	**-67743.3**	**0.0244112**
Birnbaum-Saunders	**2**	**-67763.8**	**0.0479773**
Gamma	**2**	**-67775.9**	**0.0292224**
Inverse Gaussian	**2**	**-67789.4**	**0.0550491**
Largest Extreme Value	**2**	**-67873.8**	**0.0483855**
Weibull	**2**	**-68026.0**	**0.0646936**
Logistic	**2**	**-68477.7**	**0.0910885**
Laplace	**2**	**-68514.9**	**0.108414**
Exponential	**1**	**-68795.9**	**0.199527**
Normal	**2**	**-69321.9**	**0.120725**

Histogram for CDPD data



KS a CHI kv. Testy pre CDPD

KS Test	Erlang	Gamma (3-Parameter)	Loglogistic (3-Parameter)	Weibull (3-Parameter)	Lognormal (3-Parameter)
DPLUS	**0.0**	**0.0244112**	**0.0494693**	**0.0399911**	**0.0362935**
DMINUS	**1.0**	**0.0227598**	**0.0365427**	**0.0527996**	**0.0303411**
DN	**1.0**	**0.0244112**	**0.0494693**	**0.0527996**	**0.0362935**
P-Value	**0.0**	**0.00437615**	**0.0**	**0.0**	**0.00000264**

Chi kv. Test	Gamma	Gamma (3-Parameter)	Lognormal	Lognormal (3-Parameter)
Chi-Squared	399.179	379.992	394.717	354.507
D.f.	97	96	97	96
P-Value	0.0	0.0	0.0	0.0

Kvantitatívna analýza empirického rozdelenia

Príjmy domácností SR v 2003

Výberové charakteristiky

➢ na momentovom základe

Priemer	**= 244 284 Sk**	**Koeficient šikmosti**	**= +1.0498**
Rozptyl	**= 1.2012E10**	Štand. šikmosť	= 16.959
Štandardná odchýlka	= 109 598 Sk	**Pearsonova špicatosť**	**= 2.997**
Variačný koeficient	= 44.87 %	Štand. špicatosť	= 24.209

➢ na kvantilovom základe

Medián		**= 237 407 Sk**
Kvartilové rozpätie		**= 139 770 Sk**
Galtonov koeficient šikmosti	g=qd/iqr	**= -0.0123**
Kvartilová diferencia	qd=lq+uq-2	= -1718.50
Moorsova špicatosť	k=[(e7-e5)+(e3-e1)]/iqr	**= 1.174**

Čo by sme už mali vedieť na základe doterajšej identifikácie ?

- **nevhodnosť modelovať klasickým prístupom**

  (nedostatočná „dobrá zhoda“ jednoduchých rozdelení)

- **obor funkčných hodnôt** (hranice, limity rozdelenia)

- **symetria – asymetria** (smer a intenzita asymetrie empirického rozd. a jeho častí)

- **výsledky z histórie skúmania** (známe modely aplikované v minulosti u podobnej *NP* – overenie týchto tvarov pomocou Q-Q grafov)

Zopakujme všeobecné požiadavky na model

– kvalita a preferencie medzi modelmi sa budú posudzovať podľa ich splnenia

- vhodný na nadväzujúce analýzy, na vyjadrenie základných súhrnných charakteristík, má splniť svoj praktický účel
- „dobrá zhoda“ – štatisticky významný
- jednoduchosť (jednomodálnosť)
- možnosť aktualizácie v čase
- aplikovateľnosť viacnásobného prístupu – modelovanie v podsúboroch, v štruktúrach
- možnosť aplikácie simulačných metód

Dva prístupy ku kvantilovému modelovaniu

1. Aplikovať „odskúšané“ komplexné elastické tvary

– **použiť známe zovšeobecnené tvary**

2. Kombinovať jednoduché kvantilové tvary pomocou matematického aparátu a dať im príslušnú váhu v častiach rozdelenia

– **skladaním kvantilového tvaru „na mieru“**

1. Zovšeobecnené tvary

- definované **kvantilovou funkciou** (inverznou distribučnou funkciou)
- veľmi **elastické** (flexibilné) ľahko menia tvar, obor funkčných hodnôt (max, min), limity, inflexné body, rozloženie, asymetriu...
- sú **viacparametrické**, často s parametrom polohy, rozloženia, parametrami pre horný a dolný koniec

Zovšeobecnený tvar RS GLD

Tvar štvor-parametrického asymetrického ***zovšeobecného lambda rozdelenia*** *podľa Ramberga a Schmeisera* :

$$F^{-1}(p) = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2}\ ;\ 0 \leq p \leq 1$$

používa dva lineárne a dva nelineárne parametre

Základný tvar má dva parametre pre dolný a horný koniec, ktoré spoločne modelujú šikmosť a špicatosť

je extrémne elastický

2. Skladanie kvantilových tvarov

- **rozdelenia v základnom tvare** známeho jednoduchého kvantilového rozdelenia
  - bez parametrov polohy a stupnice (variability, rozloženia)

- potrebné **identifikovať tvar zvlášť dolného a zvlášť horného konca** pomocou empirického rozdelenia

- možno modifikovať kvantilové funkcie pomocou matematických operácií (**základné pravidlá modifikácie kvantilových funkcií**)

- zvoliť **vhodné váhy** týmto dvom tvarom

$$X \sim Q[p;\, \Theta]$$

Definovanie kvantilového modelu

$$Q(p) = F^{-1}(p) = x, \qquad 0 \le p \le 1$$

Jednoduchý kvantilový model v tvare :

$$Q(p) = \lambda + \eta \boxed{S(p)} \qquad 0 \le p \le 1$$

môže byť lineárnym, alebo semilineárnym
dvoj a viacparametrickým
kvantilovým distribučným modelom

v závislosti od jeho **základného kvantilového tvaru** :

$$S(p)$$

Základné kvantilové tvary známych pravdepodobnostných rozdelení

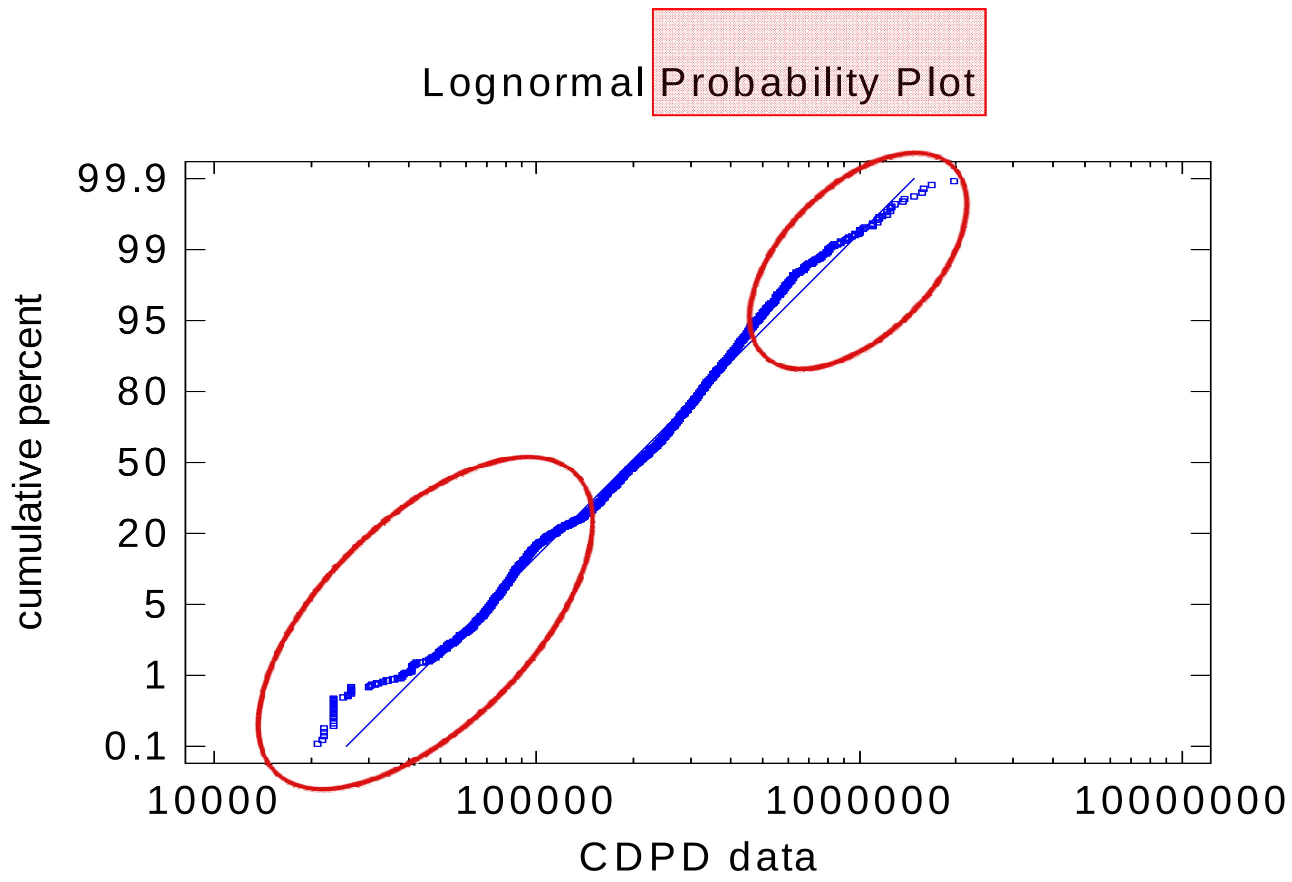
Distribution	F(x)	S(p)	Dist. range
Exponential	$1-e^{-x}$	$-\ln(1-p)$	$(0,\infty)$
Uniform	x	p	$(0,1)$
Power	$x^{\frac{1}{\alpha}}$	p^{α}	$(0,1)$, $\beta>0$
Pareto	$1-\dfrac{1}{(1+x)^{\beta}}$	$\dfrac{1}{(1-p)^{\beta}}$	$(0,\infty)$, $\beta>0$
Weibull	$1-e^{-x^{\frac{1}{\beta}}}$	$[-\ln(1-p)]^{\beta}$	$(0,\infty)$, $\beta>0$
Extreme value	$e^{-e^{-x}}$	$-\ln(-\ln p)$	$(-\infty,+\infty)$
Logistic	$\dfrac{1}{1+e^{-x}}$	$\ln\dfrac{p}{1-p}$	

Vystihnutie tvaru empirického rozdelenia - **identifikácia**

- ~~Grafická analýza empirického rozdelenia príjmov~~
- ~~Kvantitatívna analýza empirického rozdelenia príjmov:~~
  - ~~na momentovom základe~~
  - ~~na kvantilovom základe~~
- Identifikácia rozdelenia jednoduchými tvarmi
  - Celého rozdelenia
  - Zvlášť pre dolný a horný koniec
- Voľba tvaru váh pre konce rozdelenia
  - Ako funkcia *p*
  - Ako parametre modelu

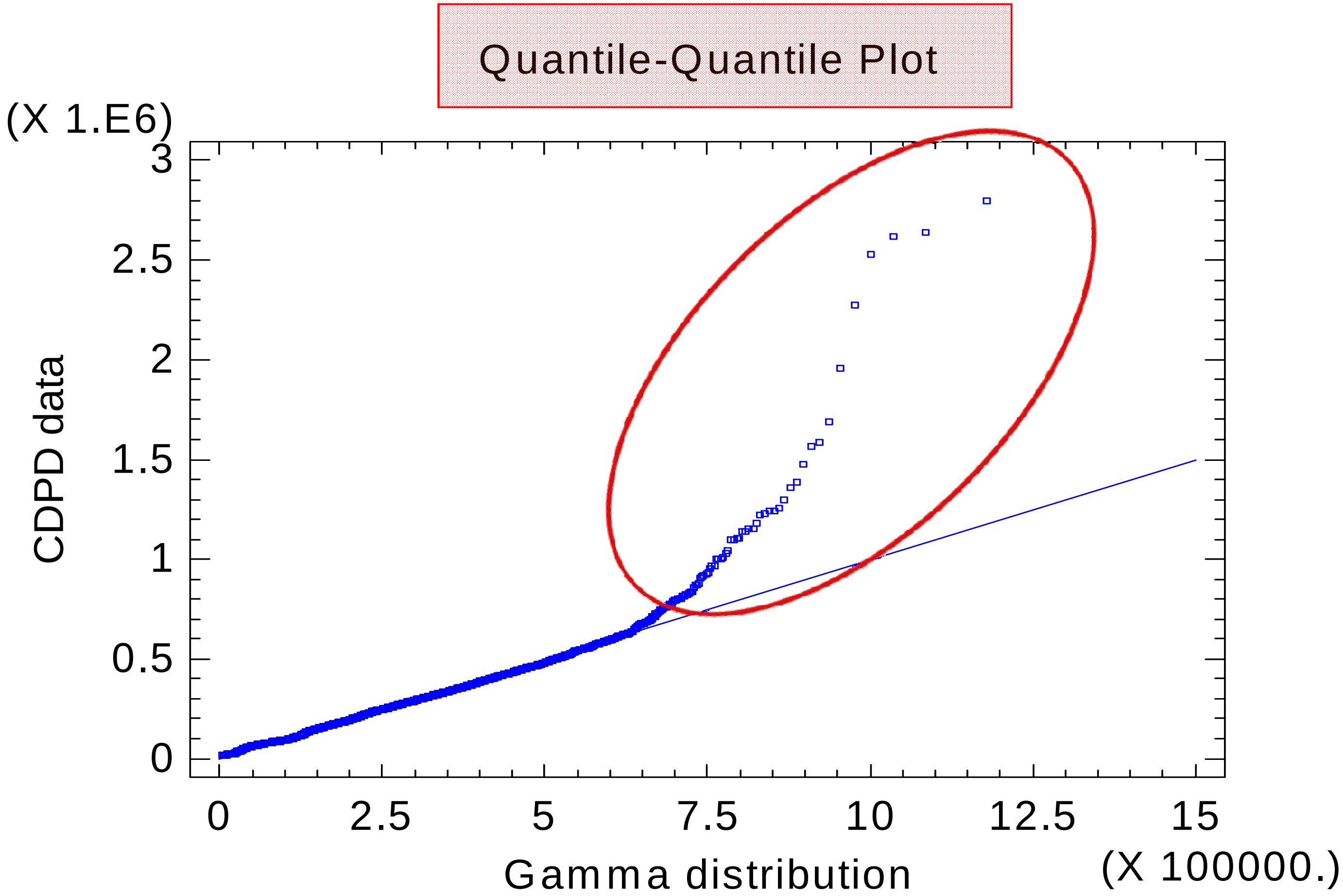
Identifikácia rozdelenia pomocou pravdepodobnostných grafov

napr. v STATGRAPHICS CENTURION XV pre Lognormálne 3-parametrické rozdelenie



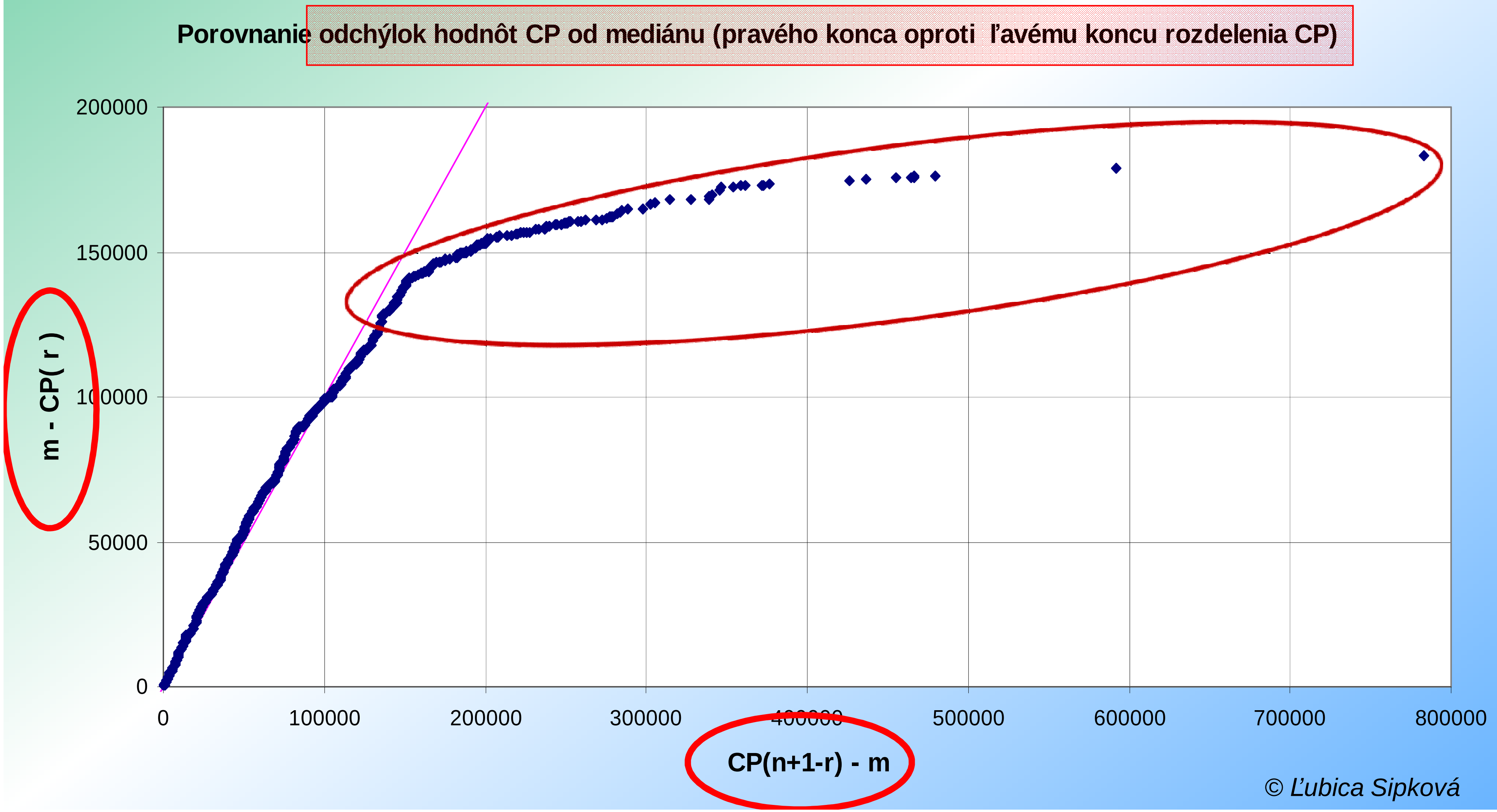
Identifikácia rozdelenia Q-Q grafmi

napr. pomocou STATGRAPHICS CENTURION XV
pre Gamma 3-parametrické rozdelenie



Identifikácia rozdelenia Q-Q grafmi porovnania vzdialenosti horných a dolných kvantilov od mediánu

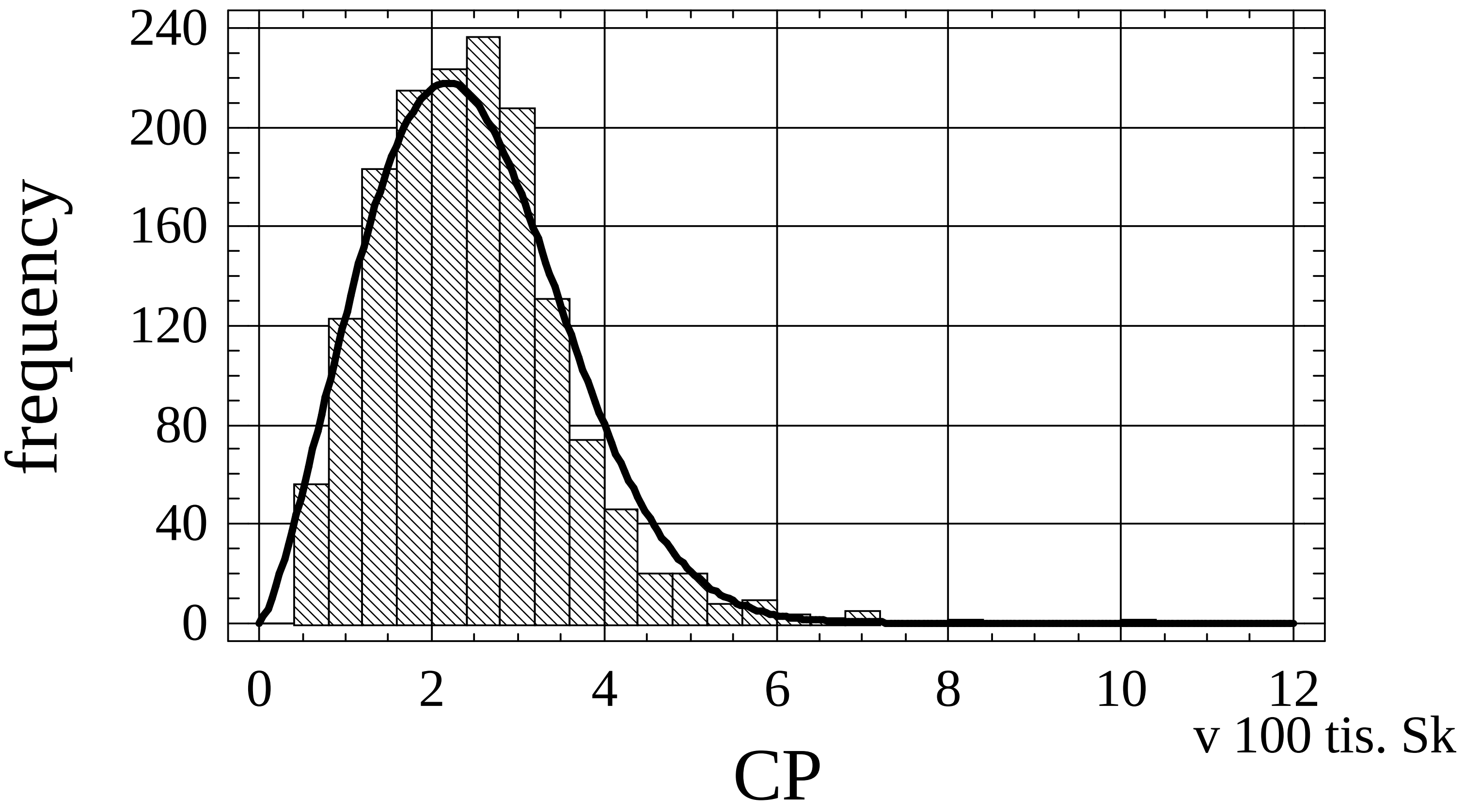
(v Exceli, ale možno aj pomocou STATGRAPHICS CENTURION XV)



Identifikácia len koncov rozdelenia jednoduchými tvarmi

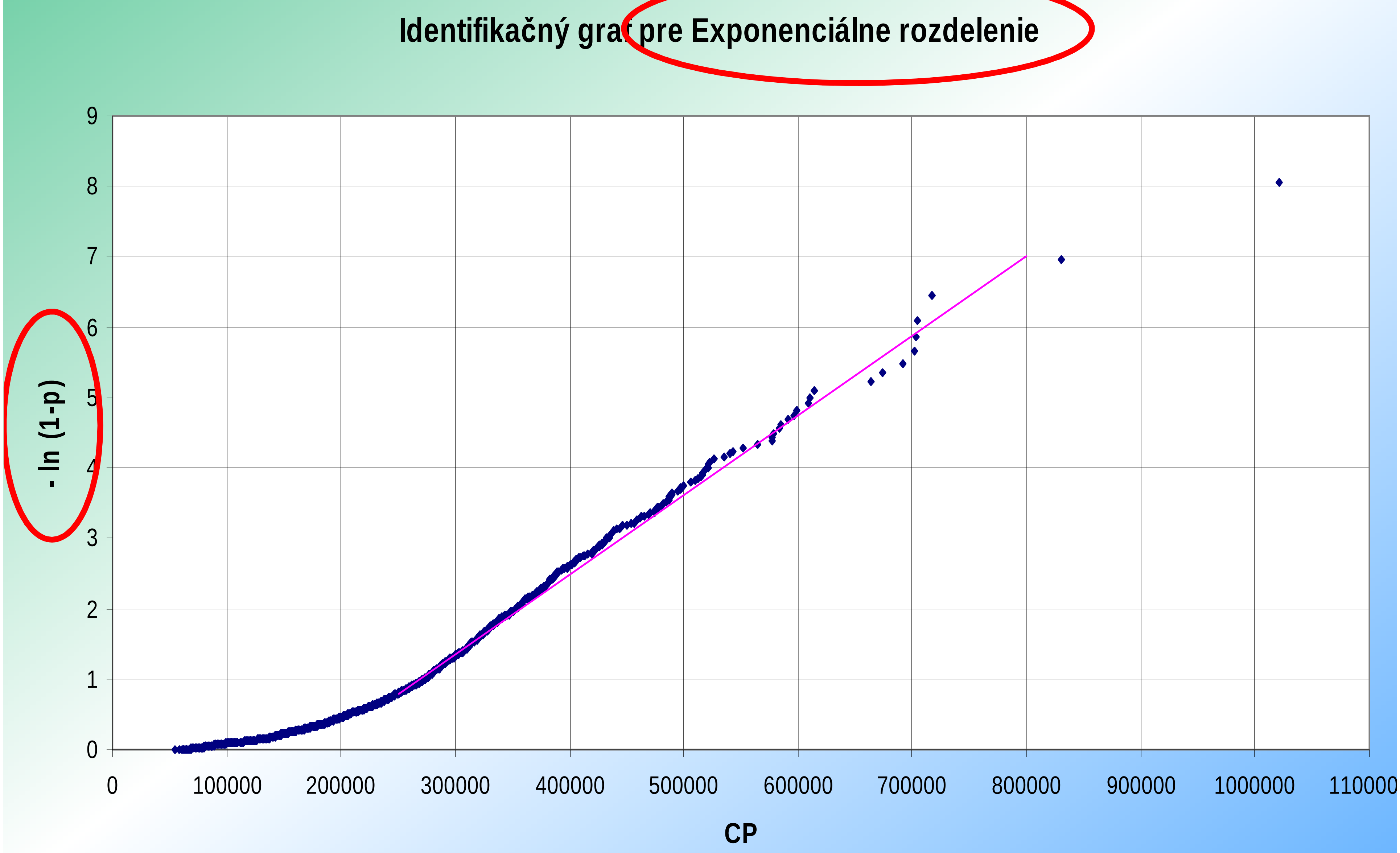
Useknutý Weibullov tvar pre dolný koniec

Histogram pre CP s funkciou hustoty Weibullovho rozdelenia

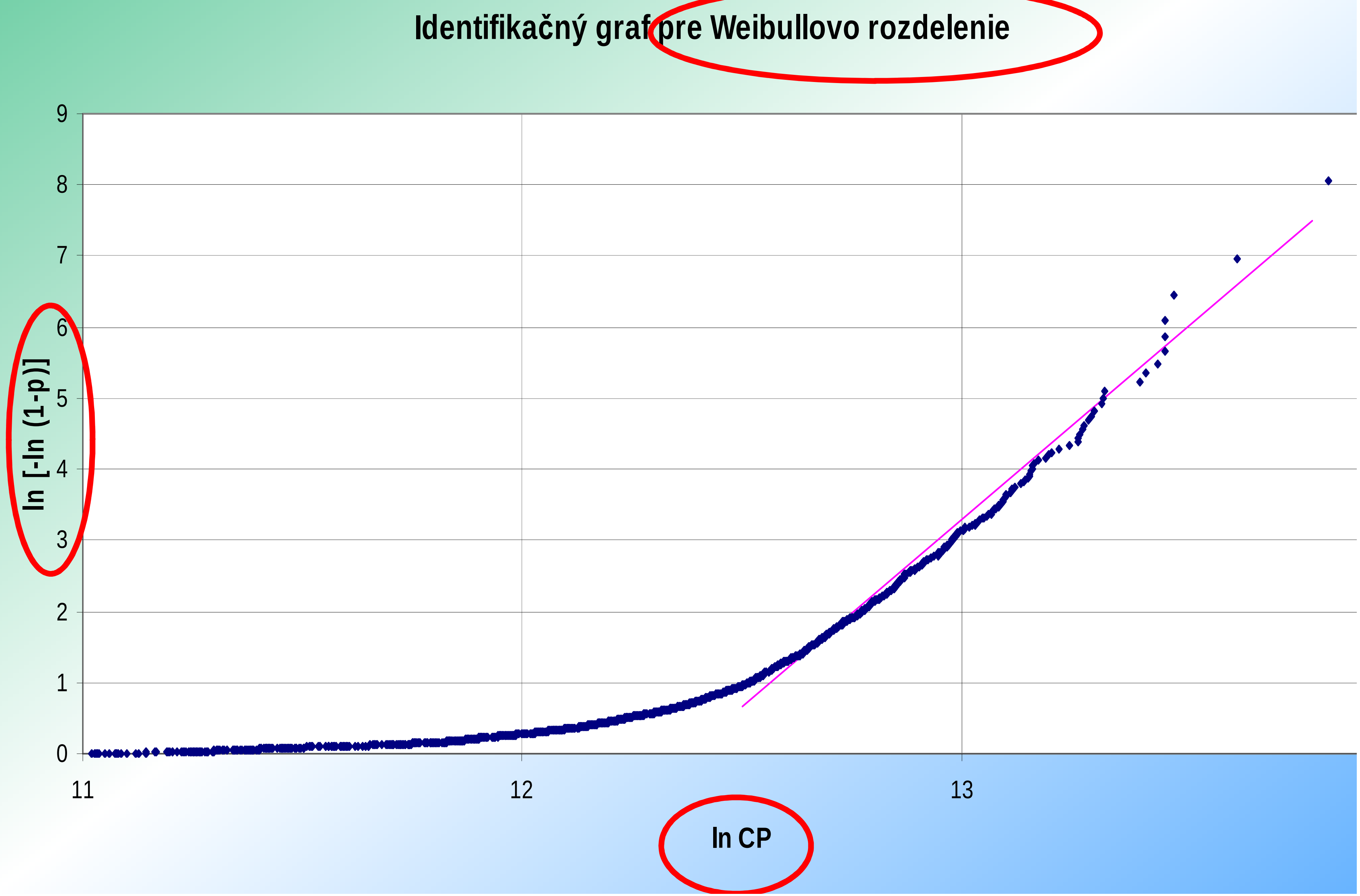


Identifikácia rozdelenia jednoduchými tvarmi

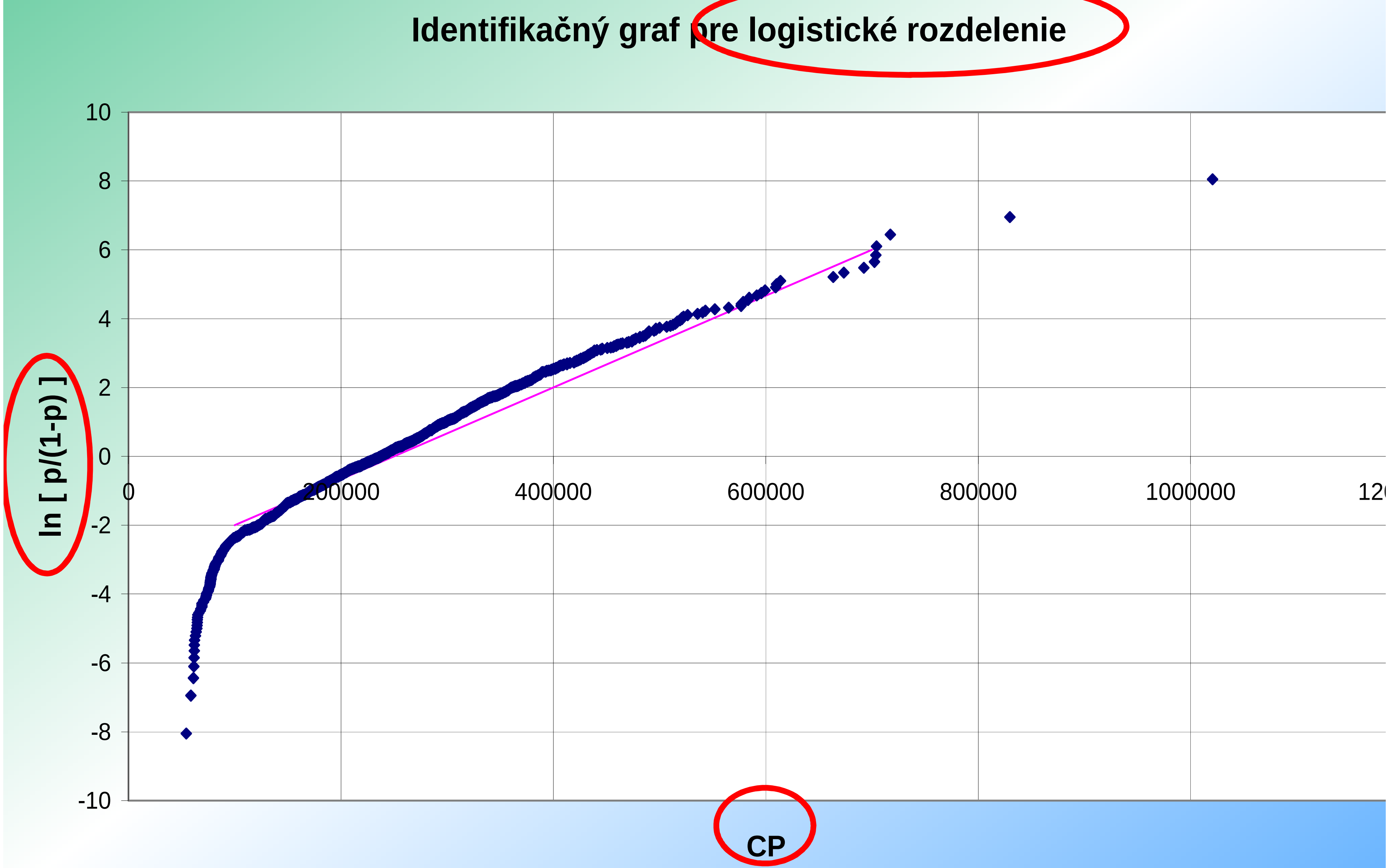
Aké rozdelenie použiť pre horný koniec?



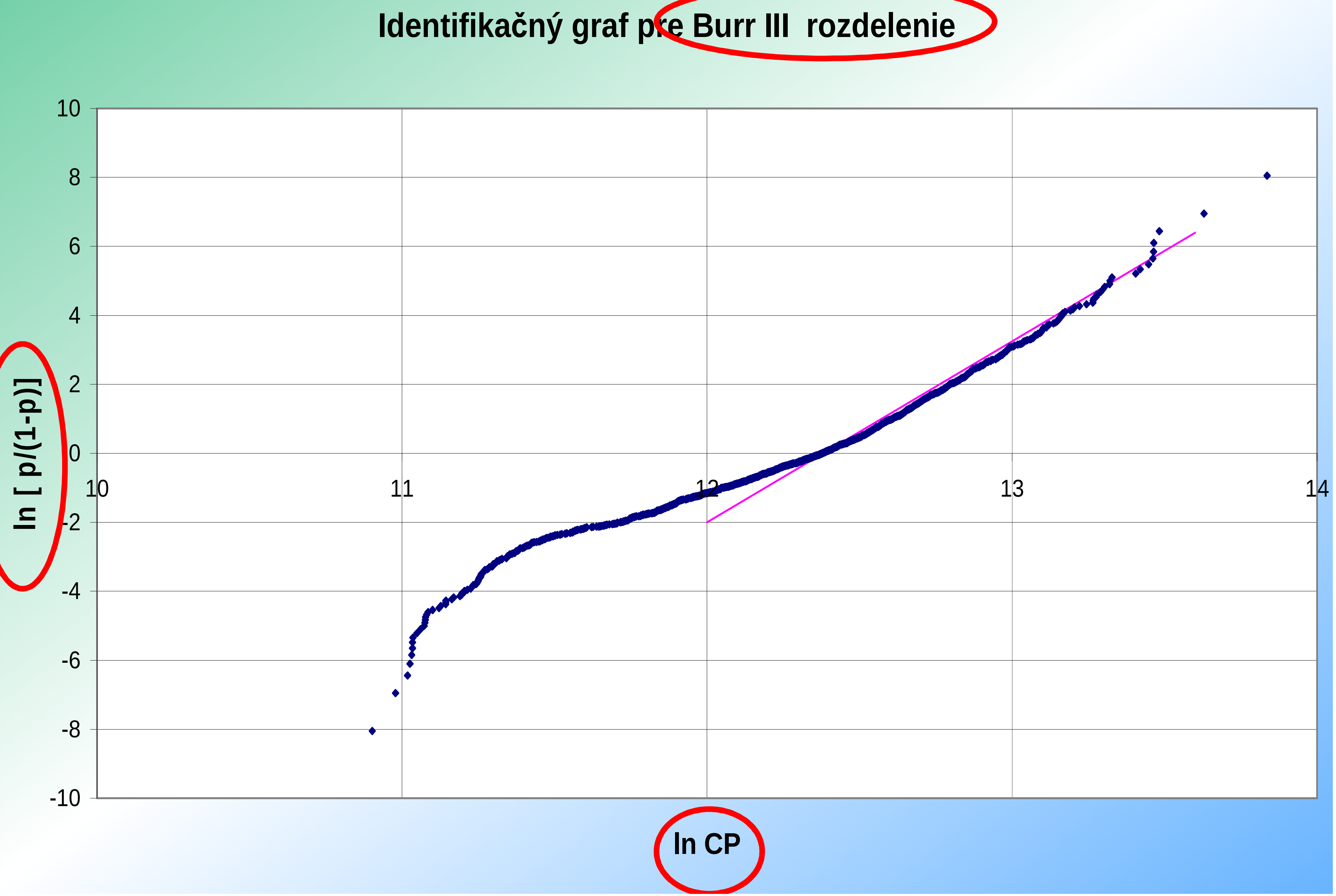
Identifikácia rozdelenia jednoduchými tvarmi



Identifikácia rozdelenia jednoduchými tvarmi



Identifikácia rozdelenia jednoduchými tvarmi



Identifikácia rozdelenia jednoduchými tvarmi

V identifikačných grafoch nesledujeme sklon preloženej priamky

Prečo sa v nich zameriame len na šikmosť a špicatosť rozdelení?

- možnosť ľahko zmeniť polohu a variabilitu kvantilových funkcií:
  - **posunutím**
  - **centrovaním**
  - **normovaním**

$$Q(p) = \lambda + \eta\, S(p)$$

- $\boldsymbol{S}(\boldsymbol{p})$ má vlastné parametre šikmosti, špicatosti, alebo parametre koncov rozdelenia, ktoré spolu definujú jeho šikmosť a špicatosť

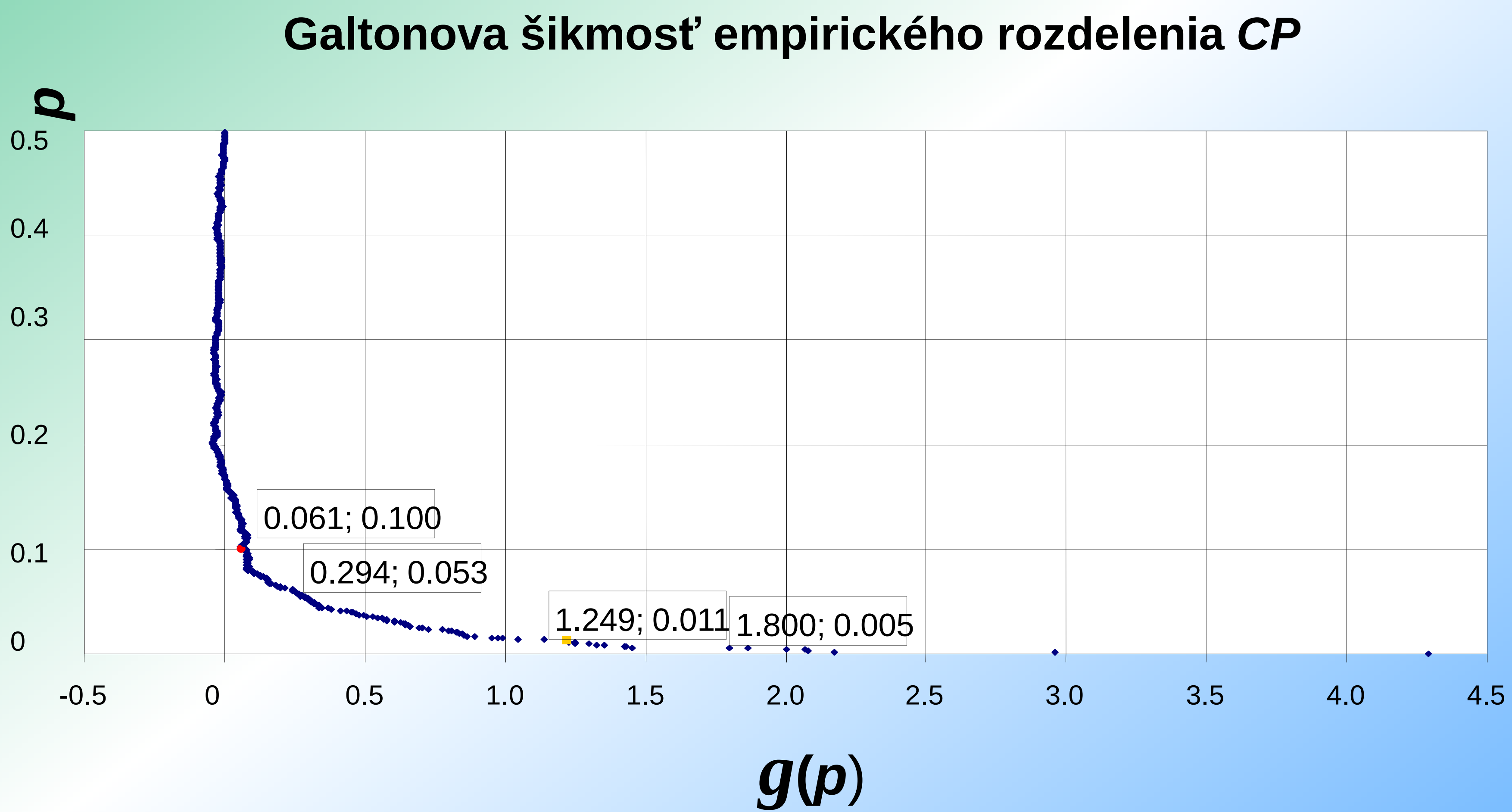
Identifikačné grafy

pre znázornenie šikmosti a špicatosti v empirickom súbore

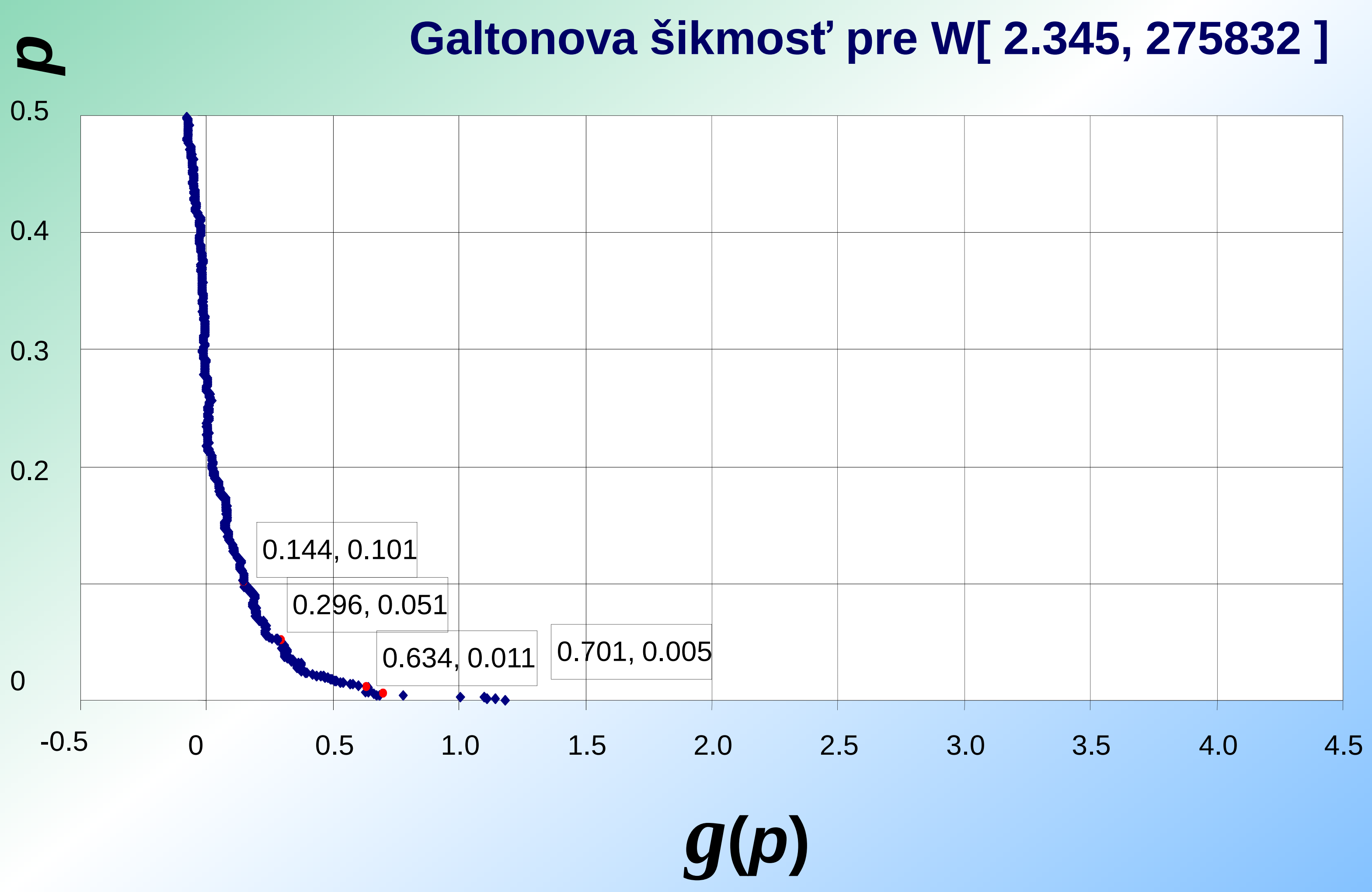
- Galtonova šikmosť $\boldsymbol{g(p)}$ oproti $\boldsymbol{p_r}$
- Galtonova p-šikmosť $\boldsymbol{g^*(p)}$ oproti $\boldsymbol{p_r}$
- indexu špicatosti $\boldsymbol{t(p)}$ oproti $\boldsymbol{p_r}$
- horného indexu špicatosti $\boldsymbol{ut(p)}$ oproti $\boldsymbol{p_r}$
- dolného indexu špicatosti $\boldsymbol{lt(p)}$ oproti $\boldsymbol{p_r}$

a ich porovnaním s tvarmi pre simulované hodnoty teoretických rozdelení

Graf Galtonovej šikmosti pre empirické dáta



Graf Galtonovej šikmosti pre teoretické rozdelenie



Tvar váh pre dva konce rozdelenia – možné voliť ako funkciu pravdepodobnosti pre plynulý prechod z jedného tvaru do druhého

Ako funkcia *p*

$$(1-p) \quad \text{a} \quad p$$

$$\left[1-p^2(3-2p)\right] \quad \text{a} \quad p^2(3-2p)$$

$$\left[1-p^3\left(10-15p+6p^2\right)\right] \quad \text{a} \quad p^3\left(10-15p+6p^2\right)$$

$$1-\omega(p) \quad \text{a} \quad \omega(p)$$

Ako parametre modelu

$$(1-\omega) \quad \text{a} \quad \omega$$

$$\frac{(1+\omega)}{2} \quad \text{a} \quad \frac{(1-\omega)}{2}$$

$$\omega_1(1-p) \quad \text{a} \quad \omega_2 p$$

Ako može vyzerať poskladaný (zložený) kvantilový model?

$$Q_{CP}(p) = \lambda + \eta\left\{(1-p)\left[-\ln(1-p)\right]^{\beta} + p\left[\frac{1}{(1-p)^{\gamma}}\right]\right\}$$

$$, 0 < p < 1, \beta > 0, \gamma > 0$$

Weibullov-Paretov tvar:

- parameter polohy
- parameter stupnice, variability
- kvantilový základný tvar rozdelenia pre dolný koniec

  s jeho parametrami
- kvantilový základný tvar rozdelenia pre horný koniec

  s jeho parametrami
- váhy jednotlivým rozdeleniam

Pravdepodobnostné modelovanie inverznými distribučnými funkciami: Identifikácia kvantilového modelu

Spracovanie tretej z cyklu prezentácií o kvantilovom modelovaní.

Podrobnejšie možno nájsť v monografii:

Sipková, Ľ; Sodomová, E.: Modelovanie kvantilovými funkciami, Vydavateľstvo EKONÓM, Bratislava, 2007; 175 s.

ISBN 978-80-225-2346-2

Ľubica SIPKOVÁ

apríl 2009