



**Vysoká škola ekonomická v Praze**

**Fakulta managementu v Jindřichově Hradci**

# **Diplomová práce**

**Pavel Jauker**

*2008*

**Vysoká škola ekonomická v Praze**

**Fakulta managementu**

**Jindřichův Hradec**

# **Diplomová práce**

**Pavel Jauker**

*2008*



**Vysoká škola ekonomická v Praze**

**Fakulta managementu v Jindřichově Hradci**

*Katedra informatiky*

# **Modelování sociální situace domácností v Čechách a na Slovensku**

**Vypracoval:**

*Pavel Jauker*

**Vedoucí diplomové práce:**

*RNDr. Jitka Bartošová, Ph.D.*

*České Budějovice, duben 2008*

# Prohlášení

Prohlašuji, že diplomovou práci na téma  
»**Modelování sociální situace domácností v Čechách a na Slovensku**«  
jsem vypracoval samostatně.

Použitou literaturu a podkladové materiály  
uvádím v příloženém seznamu literatury.

*České Budějovice, duben 2008*

---

podpis studenta

# **Anotace**

## **Modelování sociální situace domácností v Čechách a na Slovensku**

Diplomová práce je zaměřena na modelování příjmů českých a slovenských domácností v závislosti na vybraných společensko-ekonomických a demografických faktorech (oporou práce budou informace zveřejňované Českým statistickým úřadem a jeho slovenským protějškem).

*duben 2008*

# Poděkování

Za cenné rady, náměty a inspiraci

bych chtěl poděkovat

**RNDr. Jitce Bartošové, Ph.D.**

z Vysoké školy ekonomické v Praze,

Fakulty managementu v Jindřichově Hradci.

# Obsah

Úvod .....	1
<b>1 Sociální situace občana po roce 1989 .....</b>	<b>3</b>
<b>1.1 Růst mzdových nerovností a určující faktory</b>	<b>3</b>
1.1.1 Celková nerovnost	4
1.1.2 návratnost vzdělání	4
1.1.3 Mzdy podle odvětví	4
1.1.4 Demografické a kvalifikační vlivy	5
1.1.5 Vnímání nerovnosti občany	5
<b>1.2 Příjmová nerovnost domácností</b>	<b>5</b>
1.2.1 Změny v příjmech domácností	6
1.2.2 Struktura příjmů a příčiny nerovnosti	7
<b>2 Statistická teorie.....</b>	<b>8</b>
<b>2.1 Zpracování statistických dat</b>	<b>8</b>
<b>2.2 Rozdělení příjmů domácností</b>	<b>9</b>
<b>2.3 Ukazatele vlastností rozdělení</b>	<b>9</b>
2.3.1 Momentové a kvantilové charakteristiky rozdělení	10
2.3.2 Ukazatele polohy	10
2.3.3 Ukazatele variability	11
2.3.4 Charakteristiky tvaru rozdělení	15
<b>2.4 Rozdělení náhodné veličiny</b>	<b>17</b>
<b>2.5 Pravděpodobnostní modelování rozdělení</b>	<b>17</b>
2.5.1 Grafické metody	18
<b>2.6 Spojitá rozdělení</b>	<b>21</b>
2.6.1 Normální rozdělení	21
2.6.2 Logaritmicke normální rozdělení	21
2.6.3 Paretovo rozdělení	22
2.6.4 Weibullovo rozdělení	23
2.6.5 Gama rozdělení	24
<b>2.7 <math>\chi^2</math> test dobré shody</b>	<b>24</b>
<b>2.8 Odhady parametrů základního modelu</b>	<b>26</b>
2.8.1 Metoda maximální věrohodnosti	26
2.8.2 Metoda momentů	27
2.8.3 Metoda kvantilů	28

<b>3</b>	<b>Data .....</b>	<b>30</b>
3.1	<b>Mikrocensus ČR 2002</b>	<b>30</b>
3.2	<b>Mikrocensus SR 2003</b>	<b>30</b>
3.3	<b>Silc ČR 2005</b>	<b>31</b>
3.4	<b>Silc SR 2005</b>	<b>31</b>
<b>4</b>	<b>Použité programy .....</b>	<b>32</b>
4.1	<b>Microsoft Excel</b>	<b>32</b>
4.2	<b>Statgraphics</b>	<b>32</b>
<b>5</b>	<b>Praktická část .....</b>	<b>33</b>
<b>5.1</b>	<b>Analýza a modelování celkových příjmů domácností v České Republice za rok 2002</b>	<b>33</b>
5.1.1	Analýza a interpretace kvantilových grafů	33
5.1.2	Statistické ukazatele	40
5.1.3	Modelování jednoduchými pravděpodobnostními funkcemi	41
<b>5.2</b>	<b>Analýza a modelování celkových příjmů domácností ve Slovenské republice za rok 2003</b>	<b>50</b>
5.2.1	Analýza a interpretace kvantilových grafů	50
5.2.2	Statistické ukazatele	54
5.2.3	Modelování jednoduchými pravděpodobnostními funkcemi	56
<b>5.3</b>	<b>Analýza a modelování příjmů na jednoho člena domácnosti v České republice za rok 2002</b>	<b>62</b>
5.3.1	Analýza a interpretace kvantilových grafů	62
5.3.2	Statistické ukazatele	65
5.3.3	Modelování jednoduchými pravděpodobnostními funkcemi	67
<b>5.4</b>	<b>Analýza a modelování příjmů na jednoho člena domácnosti ve Slovenské republice za rok 2003</b>	<b>73</b>
5.4.1	Analýza a interpretace kvantilových grafů	73
5.4.2	Statistické ukazatele	77
5.4.3	Modelování jednoduchými pravděpodobnostními funkcemi	78
<b>5.5</b>	<b>Tříparametrické lognormální rozdělení</b>	<b>84</b>
5.5.1	Tříparametrické lognormální rozdělení pro příjmy domácností v České republice za rok 2002	84
5.5.2	Tříparametrické lognormální rozdělení pro příjmy domácností ve Slovenské republice za rok 2003	86
5.5.3	Tříparametrické lognormální rozdělení pro příjmy na jednoho člena domácnosti v České republice za rok 2002	87
5.5.4	Tříparametrické lognormální rozdělení pro příjmy na jednoho člena domácnosti ve Slovenské republice za rok 2003	89



<b>5.6</b>	<b>Dvouparametrické Paretovo rozdělení</b>	<b>91</b>
5.6.1	Dvouparametrické Paretovo rozdělení pro příjmy domácností v České republice za rok 2002	91
5.6.2	Dvouparametrické Paretovo rozdělení pro příjmy domácností ve Slovenské republice za rok 2003	93
5.6.3	Dvouparametrické Paretovo rozdělení pro příjmy na jednoho člena domácnosti v České republice za rok 2002	94
5.6.4	Dvouparametrické Paretovo rozdělení pro příjmy na jednoho člena domácnosti ve Slovenské republice za rok 2003	95
<b>5.7</b>	<b>Modelování celkových příjmů domácností podle počtu členů domácnosti</b>	<b>97</b>
5.7.1	Modelování celkových příjmů domácností v ČR podle počtu členů domácnosti za rok 2002	97
5.7.2	Modelování celkových příjmů domácností v SR podle počtu členů domácnosti za rok 2003	101
<b>5.8</b>	<b>Modelování celkových příjmů domácností ve Slovenské republice za rok 2003 podle sociálních skupin.</b>	<b>106</b>
5.8.1	Zaměstnanci	107
5.8.2	Samostatně činní	109
5.8.3	Nezaměstnaní	112
5.8.4	Důchodci	114
5.8.5	Ostatní	119
<b>5.9</b>	<b>Silec 2005</b>	<b>121</b>
5.9.1	Modelování celkových příjmů domácností v ČR podle počtu členů domácnosti za rok 2005	121
5.9.2	Modelování příjmů na člena domácnosti v ČR za rok 2005	124
5.9.3	Modelování celkových příjmů domácností v SR podle počtu členů domácnosti za rok 2005	124
5.9.4	Modelování příjmů na člena domácnosti v SR za rok 2005	127
<b>Závěr .....</b>		<b>129</b>
<b>Literatura .....</b>		<b>132</b>
<b>Přílohy</b>		

# Úvod

Tématem mé diplomové práce je modelování sociální situace domácností v Čechách a na Slovensku. Sociální situace je velmi široký pojem. V širším pojetí je možné ji charakterizovat jakou souhrn všech životních podmínek, a to jak hmotných, tak společenských. Podle této definice je však velmi obtížné ji kvantifikovat. Proto se při statistické analýze životní úrovně používají pouze měřitelné složky. Nejčastěji se pak využívá příjmů obyvatel. Výběr příjmů obyvatel právě těchto dvou zemí není náhodou.

Po roce 1990 můžeme sledovat v Československu první známky nerovnoměrného rozdělování důchodů. Důvodem byla transformace na plně tržní ekonomický systém. Po padesáti letech stability v rozdělování příjmů se začínají objevovat nerovnosti zapříčiněné tržními principy. Během této transformace se do popředí dostává otázka ekonomických nerovností. Tyto změny mají samozřejmě za následek narušení stability statistických modelů a z nich plynoucích závěrů. Ekonomická reforma přinesla mnohým lidem finanční potíže, některým dokonce chudobu, což způsobilo v otázce nerovnosti velmi citlivé téma. Nejnižší příjmy jsou sice zabezpečeny zavedením minimální mzdy, ale čím dál tím více rostou nejvyšší příjmy, což je způsobeno možností soukromého podnikání a přílivem silných zahraničních firem a jejich kapitálu. Privatizace, restituce a soukromé podnikání tyto sociální rozdíly ještě zvyšují. Velice zajímavé jsou údaje z výzkumu „Ekonomická očekávání a postoje“ organizovaným týmem pro socioekonomii Sociologického ústavu AV ČR. Na otázku: „Rozdíly ve mzdách a platech by se měly zvětšovat“ odpovědělo „určitě ano“ v květnu 1990 57,8% dotázaných, zatímco v listopadu 1994 to bylo už jen 18,9%.

Transformace nepřinesla pouze nerovnosti v rozdělování důchodů. V počátcích docházelo také k poklesu reálných příjmů, což bylo zapříčiněno uvolněním cen. Tento stav trval však jen krátkou dobu a již od roku 1992 začaly mzdy opět růst rychleji než inflace, což trvá do dnešní doby. V roce 1993 se bývalé Československo rozdělilo na dva samostatné státy – Českou a Slovenskou republiku, což znamenalo i vznik dvou nových a nezávislých ekonomik. To samozřejmě vedlo k odlišným ekonomickým reformám, odlišným makroekonomickým

ukazatelům a tudíž i k odlišným příjmům obyvatel. Příjmy obyvatel obou zemí budou vstupními daty pro vypracování mé diplomové práce, jelikož znalost jejich rozdělení, a tedy i možnost jejich porovnání z různých sociálně-ekonomických a časově prostorových hledisek, je východiskem pro posouzení životní úrovně.

Ve své teoretické práci uvedu nejen statistickou teorii, kterou pak použiji pro zpracování praktické části, ale také se pokusím vystihnout sociální situaci občanů po roce 1989.

Cílem praktické části je nalézt vhodné pravděpodobnostní modely příjmových rozdělení pro jednotlivé sociální skupiny i pro obyvatelstvo jako celek, bez ohledu na sociální skupinu. Pro nalezení vhodného modelu je možné si vybrat ze dvou cest. Využít hodnoty distribuční funkce nebo kvantilové funkce (případně jejich derivace). Já jsem si vybral první možnost. Ve své práci však využiji i kvantilových funkcí, zejména při zjišťování vhodnosti modelu pomocí Q-Q grafů. Na základě relativních ukazatelů porovnáám některé údaje v České a Slovenské republice. To však jen tam, kde výsledky budou mít dostatečnou vypovídací schopnost. Problém v tomto případě je totiž v různých měnách, kupních silách a různých letech pozorování. Vstupními daty pro mě budou statistická šetření z Českého a Slovenského statistického úřadu. Konkrétně jde o statistické šetření Mikrocensus a Silc.

# 1 Sociální situace občana po roce 1989

Jelikož sociální situace a tedy i životní úroveň velmi závisí na příjmech obyvatel, objasním v této části práce situaci na pracovním trhu po roce 1989 a s ní spjaté zvyšování nerovnosti v oblasti příjmů jednotlivců i domácností.

## 1.1 *Růst mzdových nerovností a určující faktory*

Československo za socialismu patřilo mezi země s nejmenšími mzdovými rozdíly nejen v Evropě, ale i ve světě. To bylo zapříčiněno zakořeněnými rovnostářskými tradicemi a později zejména politicko-ekonomickým tlakem socialistického systému.

„V tržní ekonomice lze žádoucí nerovnost vyjádřit „funkcionální triádou“ vzdělání, profese a příjmu. Kvalifikace pracovníka by měla odpovídat vykonávané profesi a její význam by se zase měl promítnout do odpovídající příjmové úrovně.“<sup>1</sup> Příjmová nerovnost přispívá k optimální alokaci práce a podmiňuje ekonomickou efektivnost a růst. Ekonomická reforma sice liberalizovala trh práce, ale zpočátku však zůstaly průměrné mzdy ve výrobním sektoru pod státní kontrolou prostřednictvím mzdové regulace. Tato mzdová regulace se týkala podniků s více než 25 zaměstnanci, které musely v případě překročení stanoveného limitu průměrné nominální mzdy zaplatit sankční odvody. To zapříčinilo, že v roce 1991 sice rostla nominální mzda, ale reálná poklesla o 26%. V roce 1992 došlo k další liberalizaci a už rostla i reálná mzda a to o 10%. Úplná regulace mezd byla zrušena v polovině roku 1995. Mzdové chování se nijak dramaticky nezměnilo.

„Se zahájením transformace se otevřely nové zdroje nerovnosti v odměňování, a to lepším ohodnocením dovedností, manažerské odpovědnosti, nasazování rizika a celkové výkonnosti pracovníka“<sup>2</sup> Také nové příležitosti, které přinesl tržní systém přispěl k příjmovému vzestupu

---

<sup>1</sup> J. Večerník, Občan a tržní ekonomika, str. 41

<sup>2</sup> J. Večerník, Občan a tržní ekonomika, str. 46

mnoha pracovníků. Zde můžeme zařadit možnost soukromého podnikání, zaměstnání v zahraničí nebo u západních společností.

Velice důležité je také odpovědět na otázky, jak se změnilo postavení nejvyšších a nejnižších příjmových kategorií. Možná ještě důležitější je ukázat měnící se váhu faktorů mzdové diferenciace, jako jsou rozdíly mezi muži a ženami, věkové rozpětí, návratnost lidského kapitálu a změny ve mzdách podle odvětví. Nezanedbatelné jsou však také demografické a vzdělanostní ukazatele.

### **1.1.1 Celková nerovnost**

Po roce 1992 se nerovnost výdělků zvětšuje. Zhoršuje se relace nejnižších příjmů vůči průměru a naopak příjmový vrchol roste výrazně. „Otevření prostoru pro vysoké výdělky a platové možnosti soukromého sektoru tedy vyvolaly u nejvyšší jedné desetiny odměňovaných výrazný posun.“<sup>3</sup> To způsobilo celkové rozšíření mzdového rozpětí v krajních hodnotách.

### **1.1.2 Návratnost vzdělání**

Do roku 1989 nebyla životní úroveň téměř diferencována podle vzdělání dospělých členů. Transformace přinesla příležitosti především pro osoby s nižším vzděláním. A to především díky ohromnému rozvoji služeb a stavebnictví. Teprve až v druhém plánu přinesla lepší zhodnocení vyšších stupňů vzdělání. Velký zvrat však přineslo šetření v roce 1996. Podle tohoto výzkumu získávají středoškolsky vzdělaní 105% průměrné mzdy a vysokoškolsky vzdělaní dokonce 172% průměrné mzdy.

### **1.1.3 Mzdy podle odvětví**

Po roce 1989 došlo výrazným změnám v oblasti mezd. „K relativnímu poklesu mezd vůči průměru došlo v zemědělství, průmyslu a dopravě. K nejdynamičtějším růstům mezd došlo v obchodu a zásobování. Nejviditelnější byl strmý růst mezd finančního sektoru.“<sup>4</sup>

---

<sup>3</sup> J. Večerník, Občan a tržní ekonomika, str. 49

<sup>4</sup> J. Večerník, Občan a tržní ekonomika, str. 51

### **1.1.4 Demografické a kvalifikační vlivy**

Během transformace se také změnil poměr demografických a kvalifikačních vlivů. Těsně před transformací vysvětlovaly pohlaví a věk dohromady 36% rozdílu ve mzdách, zatímco vzdělání pouze 8%. V roce 1992 už byl však vliv pohlaví a věku 19%, zatímco vliv vzdělání vzrostl na 12%.

### **1.1.5 Vnímání nerovnosti občany**

Za socialismu byla mzdová rovnost předmětem nespokojenosti. Avšak souhlas s rostoucími příjmovými rozdíly po transformaci rychle vyprchává. Od roku 1990 do roku 1996 se rozpětí mezi deseti procenty nejnižších a nejvyšších výdělků zvýšilo zhruba o polovinu. Ochota občanů přijmout toto rostoucí rozpětí poklesla z 90% na 45%. Je pochopitelné, že lidé vzdělanější a podnikatelé považují tuto změnu za sociálně spravedlivější.

## **1.2 *Příjmová nerovnost domácností***

Samozřejmě nelze předpokládat, že nerovnost v příjmech domácností se bude odvíjet pouze od příjmů jednotlivců. Působí zde i další faktory. Celkový příjem domácností ovlivňují například příjmy ze zaměstnání dalších členů, společné rodinné prostředky, daně a sociální dávky. Poslední dva jmenované faktory působí jako vyrovnávací prostředek mzdových rozdílů. Sociální dávky mají nárok čerpat většinou domácnosti s nižším příjmem a mzdové daně jsou v České Republice zatím nastaveny progresivně. Mluví se však o rovné dani, která by podle mého názoru příjmovou nerovnost domácností zvýšila. Dalším faktorem který ovlivňuje celkový rodinný příjem je bezpochyby velikost a složení domácností z hlediska poměru aktivních a závislých členů. Proto je důležité si ujasnit s jakými údaji pracovat. „Pro srovnatelné vystižení životní úrovně domácností je proto vhodnější uvažovat nikoli jejich celkový příjem, ale přepočtený na osobu. V zájmu vystižení domácnosti jako ekonomické jednotky je naopak vhodnější ukazatel jejího celkového příjmu.“<sup>5</sup>

---

<sup>5</sup> J. Večerník, *Občan a tržní ekonomika*, str. 66

Jak jsem uvedl v předchozí kapitole, po transformaci se výrazně změnila struktura mezd, což nepochybně vedlo i ke změnám nerovnosti v příjmech domácností, a to bylo dáno právě nestejným vzestupem různých skupin obyvatelstva. Určité skupiny obyvatelstva v novém tržním prostředí získávají, jiní zase ztrácejí. Tak vzniká nové společenské rozvrstvení. „Období transformace je velmi příznivé pro dosahování mimořádně vysokých příjmů a zakládání rozsáhlých majetků.“<sup>6</sup>

### **1.2.1 Změny v příjmech domácností**

Transformace umožnila nejen větší rozdíly z pracovních příjmů, ale také umožnila a posílila váhu i jiných příjmových zdrojů pocházející z majetku a finančních operací. Změna režimu zapříčinila i změnu metodiky statistických šetření. V Mikrocensech za socialismu byly částky mezd a důchodů sdělovány zaměstnavateli. Tato metodika byla se vznikem soukromého sektoru změněna. Již v Mikrocensu 1992 byly všechny příjmy sděleny samotnými respondenty, což zvětšuje i možnost jejich zkreslení. Je možno předpokládat, že výše nepřiznaných příjmů bude větší u osob nacházející se ve vrchní části příjmového žebříčku. Zavedení a zvýšení minimální mzdy, opakovaná valorizace důchodů a státní podpora pro děti a rodiny způsobilo to, že postavení domácností v nejnižší příjmové skupině zůstalo ve vztahu k průměru po transformaci v zásadě stejné. Rozpětí rodinných příjmů se rozšířilo hlavně ve prospěch domácností v nejvyšší příjmové skupině. S ohledem na tyto dvě fakta se postavení rodin ve střední příjmové skupině vůči průměru zhoršilo. „Výrazná státní intervence ve prospěch nízkopříjmových domácností zajistila udržení pouze jejich relativního postavení vůči průměru. Naproti tomu rodinám v nejvyšší příjmové kategorii se dostalo i výrazně většího dílu absolutního objemu úhrnných příjmů, čímž se výrazně posílila jejich koupěschopnost.“<sup>7</sup>

---

<sup>6</sup> J. Večerník, Občan a tržní ekonomika, str. 67

<sup>7</sup> J. Večerník, Občan a tržní ekonomika, str. 70

## 1.2.2 Struktura příjmů a příčiny nerovnosti

„Privatizace a otevření trhů vedly ke značným změnám ve struktuře příjmů domácností.“<sup>8</sup>

Podíl příjmů z mezd v úhrnných příjmech obyvatelstva poklesl a na druhé straně vzrostl podíl ostatních příjmů. Hlavní díl těchto ostatních příjmů patří příjmům z podnikání a ze zahraničí. Mezi další složky bychom mohli zařadit například přijaté půjčky, úroky z vkladů, pojistné platby, dary nebo dokonce stipendia. Významnou změnu přineslo zavedení nového systému sociálních dávek a daní v roce 1993. Daň ze mzdy byla nahrazena daní z příjmu fyzických osob a o dva roky později byl zaveden nový systém sociálních dávek. Jaké byly výsledky této změny? Průměrné daňové zatížení zůstává stejné, ale změnil se přerozdělovací vliv. V porovnání s minulostí získávají mladé bezdětné páry a páry a rodiny se čtyřmi a více dětmi. Naopak ztrácejí rodiny s jedním dítětem nebo s jedním rodičem. Tyto změny měly za úkol k příjmovému vzestupu nižších příjmových skupin.

Po transformaci se snížil vliv počtu ekonomicky aktivních osob na výši celkového příjmu domácnosti a snížil se i vliv věku hlavy rodiny. To znamená, že se zvýšily příjmy nejmladších domácností, zatímco příjmy nejstarších skupin poklesly. Naopak rostl význam vzdělání hlavy rodiny a velikost obce. „V nerovnosti příjmu přepočteného na spotřební jednotku je nejpatrnější změnou zmenšující se význam počtu nezaopatřených dětí a věku hlavy rodiny. Podobně jako v případě celkového příjmu domácností platí i pro přepočtený příjem, že vliv vzdělání hlavy rodiny a také velikosti obce se zvyšuje.“<sup>9</sup>

---

<sup>8</sup> J. Večerník, Občan a tržní ekonomika, str. 71

<sup>9</sup> J. Večerník, Občan a tržní ekonomika, str. 76



## 2 Statistická teorie

Statistika je velmi důležitou vědní disciplínou, která je založená na práci s daty. Mezi hlavní fáze patří sběr dat, jejich následné zpracování a shrnutí výsledků potřebné pro kvalifikovaná rozhodnutí. Počátky statistiky sahají až do starověkých říší, kde se jednalo o různé soupisy obyvatelstva využívané primárně pro daňové účely. „Od počátku 20. století se statistika rychle rozvíjela již jako samostatná vědní disciplína. K její další akceleraci od 70. let nepochybně přispěl také rychlý rozvoj výpočetní techniky a vznik statistického výpočetního prostředí, bez kterého je dnes moderní statistická analýza nemyslitelná.“<sup>10</sup>

Je nesporné, že význam statistiky v současném světě a hlavně v ekonomice je značný. Mimořádně významná role je připisována statistice, která poskytuje soustavu číselných informací o národním hospodářství. Realizace kvalitních a úspěšných ekonomických reforem není možná bez kvalitní statistiky a existence ekonomů nemajících specializované statistické vzdělání. Tato vědní disciplína je však potřebná např. i v procesu tvorby manažerských rozhodnutí, analýz trhu nebo managementu jakosti. Ani já se ve své praktické části neobejdu bez některých statistických metod, a proto se je pokusím definovat a vysvětlit v této teoretické části mé práce. [3]

### 2.1 Zpracování statistických dat

Výsledkem statistického šetření je většinou velké množství číselných údajů, které jsou nepřehledné a na první pohled nám nic nenapoví. Podmínkou vzniku charakteristických rysů a zákonitostí zkoumaného datového souboru je jejich setřídění. „Tříděním tedy rozumíme rozdělení jednotek souboru do takových skupin, aby co nejlépe vynikly charakteristické vlastnosti zkoumaných jevů.“<sup>11</sup> Třídění pouze podle obměn jednoho statistického znaku, čili

---

<sup>10</sup> Richard Hindls, Stanislava Hronová, Jan Seger, Jakub Fischer, Statistika pro ekonomy, Praha, Professional Publishing, 2006, str. 11

<sup>11</sup> Richard Hindls, Stanislava Hronová, Jan Seger, Jakub Fischer, Statistika pro ekonomy, Praha, Professional Publishing, 2006, str. 17

jedné vlastnosti, se nazývá jednostupňové třídění. Třídění podle více statistických znaků najednou se nazývá vícestupňové. [3]

## **2.2 Rozdělení příjmů domácností**

„K účelům statistického zkoumání příjmových rozdělení máme k dispozici pestrou škálu statistických metod a matematických i grafických postupů, které jsou dobře prozkoumány a propracovány jak po stránce teoretické, tak po stránce jejich praktického využití.“<sup>12</sup> Vlastnosti rozdělení příjmů lze popsat pomocí různých typů charakteristik. Pro vystižení empirického rozdělení máme možnost využít řadu grafických a matematických postupů. „A konečně – empirické rozdělení, které je mnohdy značně složité, můžeme aproximovat jednoduchými teoretickými modely a tím si usnadnit popis aktuálního stavu rozdělení i jeho predikci do budoucna.“<sup>13</sup>

## **2.3 Ukazatele vlastností rozdělení**

Mezi hlavní vlastnosti každého rozdělení náhodné proměnné podle Václava Čermáka [1] můžeme zařadit:

- polohu neboli obecnou úroveň hodnot  $X$  proměnné  $x$
- variabilitu těchto hodnot, kterou můžeme chápat jako vzájemnou rozdílnost všech možných párů hodnot, nebo jako odlišnost všech hodnot od nějaké konstanty (jako konstanta se zpravidla používá některá z charakteristik polohy).
- stupeň šikmosti daného rozdělení, který nám přináší informaci o tom, zda hodnoty větší než průměr (popř. medián) a hodnoty menší než průměr (popř. medián) jsou od nich stejně nebo nesterjně vzdáleny.
- stupeň špičatosti, který je v poslední době spíše nazýván kurtoza či exces.

---

<sup>12</sup> Jitka Bartošová, *Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990*, Disertační práce. Praha: Fakulta informatiky a statistiky VŠE v Praze 2006

<sup>13</sup> Jitka Bartošová, *Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990*, Disertační práce. Praha: Fakulta informatiky a statistiky VŠE v Praze 2006

### 2.3.1 Momentové a kvantilové charakteristiky rozdělení

Momentové charakteristiky jsou v dnešní době nejčastěji používanými způsobem, jak charakterizovat vlastnosti dat a identifikovat jejich rozdělení. Tyto charakteristiky, které vycházejí ze všech empirických dat, jsou často upřednostňované před kvantilovými. V posledních letech už není problém zpracovávat větší objem dat a tak můžeme najít ve více odborných pracích také kvantilové charakteristiky.

Kvantilové charakteristiky, které závisí pouze na jedné či několika empirických hodnotách jsou „imunní“ vůči extrémním hodnotám z konců rozdělení. Jsou vhodné jako doplňující pohled k momentovým charakteristikám hlavně při asymetrickém rozdělení s větším výskytem hodnot v koncích rozdělení. [4]

### 2.3.2 Ukazatele polohy

Mírou polohy se pro daný znak rozumí jednočíselná charakteristika zobecňující v jistém smyslu co možná nejlépe pozorované hodnoty.

#### 2.3.2.1 *Momentové charakteristiky*

##### **Průměry**

Průměry jsou střední hodnoty ze všech jednotek statistického souboru. Průměrů existuje několik, např. aritmetický, harmonický, geometrický, kvadratický. Já ve své práci budu pracovat pouze s průměrem aritmetickým. Ze zjištěných hodnot označených  $x_1, x_2, \dots, x_n$ , kde  $n$  znamená celkový počet pozorování, lze prostý aritmetický průměr vyjádřit jako

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

#### 2.3.2.2 *Kvantilové charakteristiky*

##### **Medián**

Medián je prostřední hodnota uspořádaného souboru. Je definován vztahem

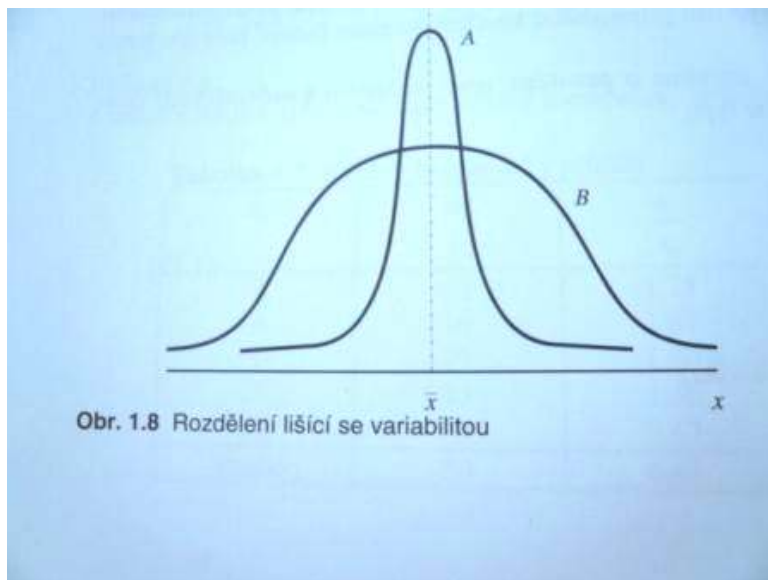
$$Q_{0,5} = x_{\left(\frac{n+1}{2}\right)}, \text{ pro } n \text{ liché} \quad (2)$$

$$Q_{0,5} = \frac{1}{2} \left( x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)} \right), \text{ pro } n \text{ sudé.} \quad (3)$$

Jeho výhodou je, že je očištěn od extrémně malých a extrémně velkých hodnot, což u rozdělení příjmů hraje významnou roli.

### 2.3.3 Ukazatele variability

Střední hodnoty, o kterých je psáno v předchozím odstavci v sobě shrnují informaci pouze o úrovni. Může se však stát, že rozdělení četností mají shodnou polohu, ale přesto se od sebe výrazně liší. Příkladem je následující obrázek.



Obr. č. 1 převzat z knihy Statistika pro ekonomy<sup>14</sup>

Hodnoty statistického souboru A jsou koncentrovány blíže okolo aritmetického průměru než hodnoty souboru B. To znamená, že vypovídací schopnost průměru je větší u statistického souboru A. Obecně je možné říci, že vypovídací schopnost aritmetického průměru je tím větší, čím je variabilita sledovaného souboru menší.

---

<sup>14</sup> Richard Hindls, Stanislava Hronová, Jan Seger, Jakub Fischer, Statistika pro ekonomy, Praha, Professional Publishing, 2006, str. 34

Měří variabilitu je celá řada. Ty míry, které charakterizují měnlivost v absolutní velikosti nazýváme mírami absolutní variability a jejich charakteristikou je, že vyjadřují variabilitu ve stejných měrových jednotkách jako sledovaný znak. Pro srovnávání variability souborů lišících se svojí úrovní jsou vhodné míry relativní variability. Ty pro změnu měří variabilitu v poměru k úrovni sledovaného znaku v souboru. Jejich výsledkem jsou bezrozměrná čísla. [3]

### 2.3.3.1 *Momentové charakteristiky*

#### **Míry absolutní variability**

Mezi nejvýznamnější míry absolutní variability patří rozptyl, který měří současně variabilitu hodnot kolem aritmetického průměru a také variabilitu vzájemných odchylek jednotlivých hodnot znaku. Je dán vztahem

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 . \quad (4)$$

Určitou nevýhodou rozptylu je, že vždy vyjadřuje hodnoty ve čtvercích použité měrné jednotky. Proto se variabilita často popisuje pomocí směrodatné odchylky, která je odmocninou rozptylu., tj.

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} . \quad (5)$$

[3]

#### **Míry relativní variability**

Jak už bylo uvedeno, míry relativní variability mají tu výhodu, že lze srovnávat variabilitu statistického znaku u dvou nebo více souborů, které se liší úrovní znaku nebo jsou vyjádřeny v jiných měrových jednotkách.

Nejznámější mírou relativní variability je variační koeficient

$$V_x = \frac{s_x}{\bar{x}} . \quad (6)$$

Variační koeficient je tedy definován jako poměr směrodatné odchylky a aritmetického průměru. Výsledkem je bezrozměrné číslo, které po vynásobení 100, udává variabilitu v procentech.

[3]

### 2.3.3.2 *Kvantilové charakteristiky*

„Kvantilové míry variability mají obecně tu nevýhodu, že nezachycují variabilitu všech hodnot znaku a nelze je hlouběji analyzovat a rozkládat, mají však i mnoho dobrých vlastností a jsou v mnoha případech snadno interpretovatelné.“<sup>15</sup>

#### **Míry absolutní variability**

mezikvartilové rozpětí:

$$iqr = Q_{0,75} - Q_{0,25} \quad (7)$$

Mezikvartilové rozpětí je rozdíl horního a dolního kvartilu. Jeho hodnota se využívá při výpočtu Galtonovi šikmosti a Moorsovi špičatosti.

horní kvartilová diference a dolní kvartilová diference:

$$uqd = Q_{0,75} - Q_{0,5} \quad (8)$$

$$lqd = Q_{0,5} - Q_{0,25} \quad (9)$$

Horní kvartilová diference poukazuje na rozdíl horního kvartilu od mediánu, dolní pak rozdíl mediánu od dolního kvartilu. Tyto hodnoty jsou důležité zejména pro výpočet kvantilových indexů špičatosti. Stejně tak se pro výpočet indexů špičatosti využívá dolní a horní p-diference, která nám říká rozdíl námi zvoleného kvantilu od mediánu.

dolní p-diference a horní p-diference:

$$ld(p) = Q_{0,5} - Q_p, \quad (10)$$

---

<sup>15</sup> Jitka Bartošová, *Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990*, Disertační práce. Praha: Fakulta informatiky a statistiky VŠE v Praze 2006

$$ud(p) = Q_{1-p} - Q_{0,5} \quad (11)$$

kde  $Q_p$  je  $p$ -ty kvantil a  $Q_{1-p}$  je  $(1-p)$ -tý kvantil a  $0 \leq p \leq 0,5$

kvantilové rozpětí a kvantilová odchylka:

$$ipr(p) = ud(p) + ld(p) = Q_{1-p} - Q_p \quad (12)$$

$$qo(p) = \frac{ipr(p)}{2} \quad (13)$$

Kvantilové rozpětí se vypočte jako rozdíl horní a dolní  $p$ -difference. Tyto hodnoty se pak využívají při výpočtu Galtonova  $p$ -indexu šikmosti. [4]

### Míry relativní variability

Mezi kvantilové relativní ukazatele variability se řadí relativní kvartilová odchylka, která se vypočítá podle vzorce

$$Q_{rel} = \frac{Q_{0,75} - Q_{0,25}}{Q_{0,75} + Q_{0,25}}. \quad (14)$$

### 2.3.3.3 Další charakteristiky

Nejjednodušší, ale i nejhrubší mírou v tomto případě je variační rozpětí. Můžeme ho definovat jako rozdíl největší a nejmenší hodnoty znaku tj.

$$R = x_{\max} - x_{\min} \quad (15)$$

Mezi přednosti variačního rozpětí můžeme zařadit snadnost a rychlost výpočtu a jednoduchou interpretaci. Na druhé straně jsou zde i nevýhody. Krajní hodnoty, na nichž je variační rozpětí založeno mohou být nahodilé. Výskyt jediné nahodilé extrémní hodnoty vyvolá ztrátu vypovídací schopnosti tohoto ukazatele. Další jeho nevýhodou je to, že nám neřekne nic o variabilitě hodnot uvnitř variačního rozpětí. Proto se dává přednost takovým mírám variability, které jsou závislé na variabilitě všech hodnot statistického souboru. [3]

## 2.3.4 Charakteristiky tvaru rozdělení

### 2.3.4.1 Ukazatele šikmosti

Ukazatele šikmosti vypovídají o tom, zda jsou hodnoty statistického souboru rozloženy souměrně kolem svého středu, nebo jestli je rozdělení hodnot zešikmeno na jednu stranu. Rozeznáváme opět dva druhy ukazatelů šikmosti – momentové a kvantilové.

Je-li rozdělení souměrné, pak je ukazatel roven nule. Čím více je rozdělení šikmější, tím se tento ukazatel více vzdaluje od nuly. Je-li koeficient  $> 0$ , je rozdělení zešikmeno kladně a znamená to, že ve statistickém souboru je větší koncentrace menších hodnot oproti hodnotám vyšších než je střední hodnota. V druhém případě je to opačně. [7]

#### Momentové charakteristiky

Mezi nepoužívanější momentové charakteristiky patří **momentový koeficient šikmosti**, který má tvar

$$a_x = \frac{1}{ns^3} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (16)$$

I přesto, že tato charakteristika má nepříznivé vlastnosti (není normovaná a je nerobustní), setkáváme se s ní nejčastěji.

#### Kvantilové charakteristiky

Kvartilová diference je dána jako

$$qd = uqd - lqd \quad (17)$$

a kvantilovou diferencí (p-diferencí) získáme jako

$$pd(p) = ud(p) - ld(p) = Q_{1-p} + Q_p - 2Q_{0,5}. \quad (18)$$

Kvantilová diference se vypočte jako rozdíl horní a dolní p-diference a využívá se zejména pro výpočet Galtonovi šikmosti.



Dalšími známými kvantilovými ukazateli šikmosti jsou:

Galtonova šikmost a Galtonův p-index šikmosti:

$$g(p) = \frac{pd(p)}{iqr}, \quad q^*(p) = \frac{pd(p)}{ipr(p)}, \quad (19), (20)$$

Galtonova šikmost upozorňuje na zešikmení prostředních 50% hodnot. Šikmost v určitých velikostech prostřední části souboru je možné měřit Galtonovo p-indexem šikmosti. Pokud za p dosadíme 0,375, pak měříme šikmost 25% prostředních hodnot. Pokud 0,125, pak měříme šikmost pro 75% prostředních hodnot. [4]

### 2.3.4.2 *Ukazatele špičatosti*

Tato charakteristika udává, jaký průběh má rozdělení hodnot statistického souboru kolem zvoleného středu. Vychází se ze srovnání s normovaným normálním rozdělením. U normovaného normálního rozdělení je míra špičatosti rovna nule. “Je-li u popisovaného rozdělení koeficient  $> 0$ , potom je toto rozdělení špičatější než normované normální rozdělení. A tedy, je-li koeficient  $< 0$ , je popisované rozdělení plošší než normované normální rozdělení.”<sup>16</sup> Čím více je míra koeficientu vzdálená nule, tím je rozdělení špičatější resp. plošší.

#### **Momentové charakteristiky**

momentový koeficient špičatosti:

$$b_x = \frac{1}{ns_x^4} \sum_{i=1}^n (x_i - \bar{x})^4 - 3 \quad (21)$$

#### **Kvantilové charakteristiky**

Moorsova špicatost:

$$k = \frac{[(e_1 - e_5) + (e_3 - e_1)]}{iqr}, \quad (22)$$

---

<sup>16</sup> M. Artlová, D. Bílková, E. Jarošová, Z. Pourová, Příklady k předmětu statistika A, Praha, OEconomica. 2004

kde  $e_1 \dots e_7$  jsou hodnoty jednotlivých oktilů. „Podle Moorsův medián, kvartilové rozpětí, Galtonův koeficient šikmosti a Moorsova špičatost ( $Q_{0,5}, iqr, g, k$ ) poskytují jednoduchou sumarizaci o tvaru rozdělení na bázi kvantilů, která je adekvátní Pearsonovo sumarizaci – průměr, rozptyl, koeficient šikmosti a špičatosti na bázi momentových charakteristik.“<sup>17</sup>

[4]

## 2.4 *Rozdělení náhodné veličiny*

Náhodnou veličinu pokládáme za danou, pokud známe všechny její možné hodnoty a pravděpodobnosti výskytu. Zákon rozdělení náhodné veličiny je pravidlo, které říká, že každé hodnotě nebo množině hodnot z každého intervalu je přiřazena pravděpodobnost, kdy náhodná veličina nabude právě této hodnoty, nebo hodnoty z určitého intervalu.

Existují čtyři způsoby funkčního definování rozdělení pravděpodobnosti náhodné proměnné.

Je to:

- distribuční funkce
- funkce hustoty pravděpodobnosti
- kvantilová distribuční funkce
- kvantilová funkce hustoty.

[3]

## 2.5 *Pravděpodobnostní modelování rozdělení*

„Pravděpodobnostní modely rozdělení sledované veličiny umožňují jednoduchou aproximaci často značně komplikovaného výběrového rozdělení.“<sup>18</sup> Typ modelu by měl vyplývat logicky z charakteru sledovaného znaku nebo z dlouhodobé zkušenosti. Pokud logická kritéria chybějí, je vhodné použít takovou distribuční funkci, která maximalizuje shodu empirického a

---

<sup>17</sup> L'ubica Sipková, Eva Sodomová, Modelovanie kvantilovými funkciami, Bratislava, Ekonom 2007, str. 38

<sup>18</sup> Acta Oeconomica Pragensia, roč. 15, č. 1, 2007, str. 7

teoretického rozdělení četností. „Ke konstrukci pravděpodobnostních modelů lze přistupovat dvěma principiálně odlišnými způsoby. Můžeme:

- použít některé známé rozdělení, jako je rozdělení logaritmicko normální, Paterovo, Weibullovo, gama apod,
- aproximovat empirické rozdělení některou teoretickou křivkou z Pearsonova či Johnsonova systému nebo namodelovat jeho tvar vhodnou rychle konvergující řadou.“<sup>19</sup>

„Při aproximaci empirického rozdělení pomocí známých rozdělení si můžeme vybrat jednu ze dvou cest vedoucích k cíli. Můžeme využít hodnoty:

- distribuční funkce  $F(x)$  (případně její derivace),
- kvantilové funkce  $Q(p)$  (popřípadě její derivace).“<sup>20</sup>

Já ve své práci použiji prvně zmiňované známé rozdělení viz. kapitola 2.6. a pro aproximaci distribuční funkci, avšak méně známou a používanou kvantilovou funkci využiji při zjišťování vhodnosti modelu pomocí Q-Q grafů. [6]

## 2.5.1 Grafické metody

### 2.5.1.1 Grafické kvantilové metody

Základem všech kvantilových metod jsou vzestupně seřazené empirické údaje. V dalším textu budeme vycházet z předpokladu: „Příslušná  $i$ -tá  $p$ -hodnota odpovídá  $i$ -tému pořadí zjištěné hodnoty  $(x_i)$  náhodné proměnné  $X$  ve výběrovém souboru.“<sup>21</sup>

Definiční obor hodnot pro  $p$  je interval  $(0,1)$ . Pro systematické umístění hodnot  $p$  v tomto uvedeném intervalu a jeho rozdělení na  $n$  stejných částí, kde  $n$  znamená rozsah souboru použijeme vztah:

---

<sup>19</sup> Acta Oeconomica Pragensia, roč. 15, č. 1, 2007, str. 7

<sup>20</sup> Acta Oeconomica Pragensia, roč. 15, č. 1, 2007, str. 8

<sup>21</sup> L'ubica Sipková, Eva Sodomová, Modelovanie kvantilovými funkciami, Bratislava, Ekonóm 2007, str. 18

$$p = \frac{(i-0,5)}{n} \quad (23)$$

Uspořádané dvojice  $x_i, p_i$  je možno znázornit graficky jako  $p$  v závislosti na  $x$ , nebo jako  $x$  v závislosti na  $p$ . První zmíněný graf, tedy  $p$  v závislosti na  $x$  poukazuje na zvláštnosti rozmístění hodnot ve výběru. Odpovídá distribuční funkci rozdělení a ukazuje nám, jak se mění proporcionální postavení proměnné v souvislosti s její velikostí. Druhý graf, tedy  $x$  v závislosti na  $p$  odpovídá grafu kvantilové funkce. Poukazuje na to, jak se mění hodnota proměnné v závislosti na proporcionálním postavení. Kombinace těchto dvou grafů často odhalí rozdílné črty rozložení empirických hodnot a intenzitu růstu.

„Diferenci, tj. vzdálenost mezi sousedními hodnotami proměnné  $X$  udává rozdíl  $D_x = x_i - x_{i-1}$  a  $D_p = p_i - p_{i-1}$ , což je rozdíl jejich  $p$ -hodnot.“<sup>22</sup> Pomocí těchto hodnot můžeme sestrojít další dva grafy, které poukazují na místa výskytu nejvyšších změn, posunutí

vrcholu a přiblížení konců rozdělení. Grafickým znázorněním směrnice růstu  $\frac{D_p}{D_x}$  vzhledem

ke středu příslušné difference  $D_x$ , dostaneme tvar, charakterizující funkci hustoty. Nejméně

obvyklé je grafické znázornění směrnice růstu  $\frac{D_x}{D_p}$  vzhledem ke středu  $D_p$ , což charakterizuje

kvantilovou funkci hustoty. [4]

### 2.5.1.2 *Ostatní grafické metody*

K těmto grafickým kvantilovým metodám je dobré také sestrojít klasické grafy, kterými mohou být např. histogram, krabičkový diagram či polygon.

„**Krabičkový diagram** je alternativní typ grafického zobrazení, které jednoduše znázorňuje odlehlé extrémní hodnoty a poskytuje přehled o středu a rozptýlenosti dat v souboru.“<sup>23</sup>

Krabičkový diagram ukazuje hodnotu

---

<sup>22</sup> L'ubica Sipková, Eva Sodomová, Modelovanie kvantilovými funkciami, Bratislava, Ekonóm 2007, str. 19

<sup>23</sup> Jitka Bartošová, Základy statistiky pro manažery, Oeconomica 2006, str. 5

- variačního rozpětí,
- kvantilového rozpětí,
- mediánu.

„Tento graf se vyskytuje i v různých obměnách, kdy pozorování, která lze považovat za odlehlá, či extrémní, jsou znázorněna zvláštními body.“<sup>24</sup> Za hodnotu extrémně nízkou je považována hodnota, která odpovídá nerovnosti  $x < Q_{0,25} - 1,5iqr$ . Na druhé straně za extrémně vysokou je považována hodnota, která splňuje nerovnost  $x > Q_{0,25} + 1,5iqr$ .

„**Histogram**, neboli sloupcový diagram, používáme pro grafické znázornění intervalového rozdělení četností. Šířka sloupce odpovídá šířce intervalu a jeho výška je dána zjištěnou četností.“<sup>25</sup> Pro zjištění, kolik sloupců v histogramu vytvořit, je vhodné použít vztah

$$L = [2,46(n-1)^{0,4}], \quad (24)$$

kde [ ] znamená celočíselnou část

**Lorenzova křivka** znázorňuje graficky stupeň příjmové nerovnosti (příjmové diferenciaci).

„Tato grafická metoda spočívá v tom, že se vyznačí body, které odpovídají kumulativním podílům příjmů na jedné straně a kumulovaným podílům příjemců na druhé straně.“<sup>26</sup>

Lorenzovu křivku jsem sestrojil v programu Microsoft Excel. Pro kumulovaný podíl příjemců jsem použil vztah

$$p = \frac{i - 0,5}{n}. \quad (25)$$

Pro kumulativní podíl příjmů jsem použil vzorec

$$i = \frac{1}{\sum_{i=1}^n x_{(i)}} \sum_{j=1}^n \left( \sum_{x_{(i)} \leq x_{(j)}} x_{(i)} \right), \quad (26)$$

---

<sup>24</sup> Jitka Bartošová, Základy statistiky pro manažery, Oeconomica 2006, str. 5

<sup>25</sup> Jitka Bartošová, Základy statistiky pro manažery, Oeconomica 2006, str. 5

<sup>26</sup> Jiřina Moravcová, Sociální statistika, Vysoká škola ekonomická, 2004, str. 72

kde  $x_{(i)}$  je pořadová statistika, kde  $x_{(i)} \leq x_{(i+1)} \leq \dots x_{(n)}$ .

V případě ideální rovnoměrnosti příjmů by body ležely na diagonále. To znamená, že např. 10% příjemců by se podílelo na objemu příjmů rovněž 10-ti procenty. V praxi tomu však v naprosté většině případů není, a proto se křivka prohnutě odklání od diagonály směrem dolů. [4], [5]

## 2.6 Spojitá rozdělení

### 2.6.1 Normální rozdělení

„Parametrizované normální rozdělení se definuje vždy se dvěma parametry: s parametrem polohy, který se místo  $a$  značí zpravidla  $\mu$ , a s parametrem škály, který se místo  $b$  značí zpravidla  $\sigma$ .“<sup>27</sup> Toto rozdělení je charakterizováno hustotou pravděpodobnosti

$$f_X(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (27)$$

kde  $-\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma < \infty$ ,

popř. distribuční funkcí

$$F_X(x, \mu, \sigma) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dt \quad (28)$$

Distribuční funkce, kvantilové funkce, ani funkce kvantilové hustoty nemají explicitní analytické vyjádření. Jsou vyjádřeny tabulkově. [1]

### 2.6.2 Logaritmicko normální rozdělení

„V nejobecnější podobě se lognormální rozdělení definuje zpravidla se třemi parametry: s parametrem polohy ( $a$ ), škály ( $b$ ) a tvaru ( $c$ ). Někdy se jako parametrizovaný tvar uvádí rozdělení s parametry škály a tvaru, tj. parametr polohy se položí roven nule.“<sup>28</sup>

---

<sup>27</sup> Václav Čermák, Diskrétní a spojitá rozdělení, vzorce, grafy, tabulky, Praha, VŠE, 1993

Funkce hustoty pravděpodobnosti je dána vztahem

$$f_X(x, a, b, c) = \frac{1}{\sqrt{2\pi c}} \frac{1}{x-a} \exp\left\{-\frac{1}{2c^2} \left(\log \frac{x-a}{b}\right)^2\right\}, \quad (29)$$

kde  $x > a, -\infty < a < \infty, b < 0, c > 0$

Distribuční funkce, kvantilová funkce ani funkce kvantilové hustoty nemají explicitní analytické tvary. [1]

### 2.6.3 Paretovo rozdělení

Rozdělení se dvěma parametry (s parametry polohy a tvaru).

Funkce hustoty pravděpodobnosti je dána vztahem

$$f_X(x, x_0, c) = \frac{c}{x_0} \left(\frac{x_0}{x}\right)^{c+1}, \quad (30)$$

kde  $x > x_0 > 0, c > 0$ ,

distribuční funkce má tvar

$$F_X(x, x_0, c) = 1 - \left(\frac{x_0}{x}\right)^c. \quad (31)$$

Kvantilovou funkci a funkci kvantilové hustoty lze vyjádřit jako

$$Q_U(u, x_0, c) = x_0(1-u)^{-1/c} \quad (32)$$

$$q_U(u, x_0, c) = \frac{c}{x_0} (1-u)^{1+1/c} \quad (33)$$

---

<sup>28</sup> Václav Čermák, Diskrétní a spojitá rozdělení, vzorce, grafy, tabulky, Praha, VŠE, 1993

„Speciální, ale velmi často uváděný parametrizovaný tvar se získá volbou  $x_0 = 1$ . Jedná se o rozdělení definované hustotou  $f_X(x,1,c) = \frac{x}{x^{c+1}}$ . Střední hodnota je u takto parametrizovaného tvaru rovněž o 1 větší.“<sup>29</sup> [1]

## 2.6.4 Weibullovo rozdělení

„V nejobecnější podobě se také rozdělení Weibullovo – stejně jako většina ostatních – definuje zpravidla se třemi parametry: s parametrem polohy (a), škály (b) a tvaru (c). Někdy se jako parametrizovaný tvar uvádí rozdělení s parametry škál a tvaru, tj. parametr polohy se předpokládá rovný nule.“<sup>30</sup>

Funkce hustoty pravděpodobnosti je dána jako

$$f_X(x,a,b,c) = \frac{c}{b} \left( \frac{x-a}{b} \right)^{c-1} \exp \left\{ - \left( \frac{x-a}{b} \right)^c \right\}, \quad (34)$$

kde  $x \geq a, -\infty < a < \infty$ , (většinou se používá  $a \geq 0$ ),  $b, c > 0$

Distribuční funkce tohoto rozdělení má tvar

$$F_X(x,a,b,c) = 1 - \exp \left\{ - \left( \frac{x-a}{b} \right)^c \right\}, x \geq a, \quad (35)$$

kvantilová funkce

$$Q_U(u,a,b,c) = a + b[-\ln(1-u)]^{1/c}, 0 < u < 1 \quad (36)$$

a funkce kvantilové hustoty:

$$q_U(u,a,b,c) = \frac{c}{b}(1-u)[-\ln(1-u)]^{1-1/c}, 0 \leq u \leq 1 \quad (37)$$

[1]

---

<sup>29</sup> Václav Čermák, Diskrétní a spojitá rozdělení, vzorce, grafy, tabulky, Praha, VŠE, 1993

<sup>30</sup> Václav Čermák, Diskrétní a spojitá rozdělení, vzorce, grafy, tabulky, Praha, VŠE, 1993



## 2.6.5 Gama rozdělení

„V nejobecnější podobě se i rozdělení gama definuje zpravidla se třemi parametry: s parametrem polohy ( $a$ ), škály ( $b$ ) a tvaru ( $c$ ). Někdy se při definování parametrizovaného tvaru zavedou jen parametr škály a parametr tvaru, tj. parametr polohy se položí roven nule.“<sup>31</sup>

Funkce hustoty pravděpodobnosti je dána obecně vztahem

$$f_X(x, a, b, c) = \frac{1}{b\Gamma(c)} \left(\frac{x-a}{b}\right)^{c-1} \exp\left\{-\left(\frac{x-a}{b}\right)\right\}, \quad x > a, b, c > 0 \quad (38)$$

a distribuční funkce vztahem

$$F_X(x, a, b, c) = \frac{\gamma(c, \frac{x-a}{b})}{\Gamma c} \quad (39)$$

Kvantilové funkce ani funkce kvantilové hustoty nemají explicitní analytické tvary. „Standardní tvar rozdělení gama získáme, když položíme  $a=0$  a  $b=1$ . Když položíme  $c=1$  dostaneme rozdělení exponenciální. Je-li ve standardním tvaru  $c$  číslo přirozené, jedná se o rozdělení Erlangovo.“<sup>32</sup> [1]

## 2.7 $\chi^2$ test dobré shody

„ $\chi^2$  testem dobré shody zjišťujeme, zda výběrový soubor pochází z populace s určitým typem rozdělení či nikoliv.“<sup>33</sup> Tento test se vždy provádí jako oboustranný. Nulová a alternativní hypotézy jsou formulovány takto:

$$H_0 : n_i = n\pi_{0i} \text{ versus } H_1 : n_i \neq n\pi_{0i} \quad (40)$$

Nulová hypotéza říká, že empirické a teoretické četnosti je možné považovat za shodné.

---

<sup>31</sup> Václav Čermák, Diskrétní a spojitá rozdělení, vzorce, grafy, tabulky, Praha, VŠE, 1993

<sup>32</sup> Václav Čermák, Diskrétní a spojitá rozdělení, vzorce, grafy, tabulky, Praha, VŠE, 1993

<sup>33</sup> Jitka Bartošová, Základy statistiky pro manažery, Oeconomica 2006, str. 172

K testování nulové hypotézy se používá statistika:

$$G = \sum_{i=1}^k \frac{(n_i - n\pi_{0i})^2}{n\pi_{0i}}, \quad (41)$$

kde  $n_i$  jsou absolutní četnosti výběrového rozdělení,  $n\pi_{0i}$  jsou absolutní četnosti předpokládaného teoretického rozdělení.

„Statistika  $G$  se za platnosti  $H_0$  řídí  $\chi^2$  rozdělením s  $(k-h-1)$  stupni volnosti, kde  $k$  je počet skupin, vzniklých setříděním výběru podle nějakého kritéria, a  $h$  je počet neznámých (odhadovaných) parametrů předpokládaného teoretického rozdělení. Pokud jsou všechny parametry teoretického rozdělení známy, je toto rozdělení plně specifikované a  $h=0$ .“<sup>34</sup>

Jedná se o asymptotický test, a proto je potřeba dodržet podmínku, která je nutná pro dosažení dostatečné přesnosti výsledku. Tou je setřídění dat do dostatečně velkých tříd. Za dostatečně velkou třídu je považována  $i$ -tá třída, která splňuje nerovnost

$$n\pi_{0i} > 5, \quad (42)$$

kde  $\pi_{0i}$  je pravděpodobnostní obsazení  $i$ -té třídy za platnosti nulové hypotézy.

„Oborem nezamítání hypotézy  $H_0$  na hladině významnosti  $\alpha$  je množina

$$V = \{G : G < \chi_{1-\alpha}^2(k-h-1)\}, \quad (43)$$

kritickým oborem je množina

$$W = \{G : \chi_{1-\alpha}^2(k-h-1) \leq G\} \quad (44)$$

[5]

---

<sup>34</sup> Jitka Bartošová, Základy statistiky pro manažery, Oeconomica 2006, str. 173

<sup>35</sup> Jitka Bartošová, Základy statistiky pro manažery, Oeconomica 2006, str. 173

## 2.8 Odhady parametrů základního modelu

Pro odhad parametrů základního modelu existuje několik metod. Já jsem ve své práci použil tři metody. Konkrétně to je:

- metoda maximální věrohodnosti
- metoda momentů
- metoda kvantilů

„Každá z výše uvedených metod odhadu má své přednosti a nedostatky. O tom, která z nich je v dané situaci nejlepší, rozhodují vlastnosti metody odhadu a charakter výběrových souborů.“<sup>36</sup>

### 2.8.1 Metoda maximální věrohodnosti

Metoda maximální věrohodnosti je jedna z nejpoužívanějších. „Máme-li výběr  $\{X_1, \dots, X_n\}$ , pak maximálně věrohodný odhad  $\vec{\hat{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$  vektoru parametrů  $\vec{\theta} = (\theta_1, \dots, \theta_k)$  předpokládaného rozdělení, z něhož výběr pochází, spočívá ve vyhledání argumentu suprema věrohodnostní funkce (likelihood function), tj. hodnoty  $\arg \sup_{(\theta)} L(\vec{\theta} | x_1, \dots, x_n)$ , kde  $\{x_1, \dots, x_n\}$  je realizace náhodného výběru.“<sup>37</sup> V případě náhodného výběru  $\{x_1, \dots, x_n\}$  je věrohodnostní funkce dána vztahem

$$L(\vec{\theta} | x_1, \dots, x_n) = \prod_{i=1}^n [f(x_i; \vec{\theta})] . \quad (45)$$

Metoda maximální věrohodnosti je podle konzistentní a vydatná, avšak pro výběrové soubory s omezeným rozsahem nemusí být odhady kvalitní. Rovnice pro výpočet parametrů jednotlivých rozdělení zde neuvádím z důvodu použití statistického programu Statgraphics.

---

<sup>36</sup> Jitka Bartošová, *Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990*, Disertační práce. Praha: Fakulta informatiky a statistiky VŠE v Praze 2006

<sup>37</sup> Jitka Bartošová, *Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990*, Disertační práce. Praha: Fakulta informatiky a statistiky VŠE v Praze 2006

Tato metoda je použita u odhadů parametrů Weibullova, gama, Erlangova, jednoparametrického Paretova a dvouparametrického lognormálního rozdělení.

[8]

## 2.8.2 Metoda momentů

„Za předpokladu, že náhodný výběr  $\{X_1, \dots, X_n\}$  pochází z rozdělení s vektorem parametrů

$\vec{\theta} = (\theta_1, \dots, \theta_k)$ , získáme momentový odhad  $\vec{\hat{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$  řešením soustavy

$$\mu'_r(\hat{\theta}_1, \dots, \hat{\theta}_k) = m'_r, \quad r = 1, 2, \dots, k, \quad (46)$$

kde  $\mu'_r(\hat{\theta}_1, \dots, \hat{\theta}_k) = E(X^r)$ ,  $r = 1, \dots, k$  je prvních  $k$  obecných teoretických momentů,

$m'_r = \frac{1}{n} \sum_{i=1}^n X_i^r$  jsou odpovídající výběrové momenty.<sup>38</sup>

Výhodou této metody je její jednoduchost rovnic i to, že výběrové momenty jsou konzistentními odhady teoretických momentů. Na druhou stranu tyto odhady nejsou robustní a jejich vydatnost klesá s růstem variability, což může způsobovat nepřesnost odhadu.

[8]

### 2.8.2.1 Odhady parametrů dvouparametrického Paretova rozdělení

Odhad parametrů tohoto modelu vychází z těchto rovnic

$$\bar{x} = E(X) = \frac{c}{c-1} x_0, \quad \text{pro } c > 1 \quad (47)$$

$$s^2 = D(X) = E(X^2) - [E(X)]^2 = \frac{c}{(c-1)^2(c-2)} x_0^2, \quad (48)$$

---

<sup>38</sup> Jitka Bartošová. *Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990*, Disertační práce. Praha: Fakulta informatiky a statistiky VŠE v Praze 2006

neboli

$$s = \sqrt{D(X)} = \sqrt{\frac{c}{c-2} \frac{x_0}{c-1}} \quad (49)$$

Řešením soustavy první a třetí rovnice pak dostáváme vzorec pro výpočet parametru  $c$ .

$$c = 1 + \frac{\sqrt{\bar{x}^2 + s^2}}{s}, \quad (50)$$

parametr  $x_0$  pak získáme dosazením do rovnice (47).

### 2.8.3 Metoda kvantilů

Kvantilová metoda vychází ze shody teoretických a výběrových kvantilů. „Hledané odhady jsou tedy kořeny soustavy

$$Q_{Pr0} = Q_{Pr}, \quad r = 1, 2, \dots, k,$$

kde  $Q_{Pr0}$ ,  $r = 1, \dots, k$  jsou teoretické kvantily, takže platí  $0 \leq \int_{-\infty}^{Q_{Pr0}} f(x, \vec{\theta}) dx = p_{r0} \leq 1$ ,

$Q_{Pr}$  jsou odpovídající kvantily výběrové<sup>39</sup>

Kvantilové odhady jsou jednoduché na výpočet a nevyžadují velké výběrové soubory. Další jejich výhodou je konzistentnost a robustnost. Nevýhodou je jejich malá vydatnost.

[8]

#### 2.8.3.1 Odhady parametrů tříparametrického lognormálního rozdělení

Kvantilová metoda byla použita k odhadu parametrů lognormálního modelu.

$$Q_{0,5} = e^{\mu} + \hat{\gamma}, \quad (51)$$

$$Q_p = e^{\hat{\mu} + \hat{\sigma}t_p} + \hat{\gamma}, \quad (52)$$

---

<sup>39</sup> Jitka Bartošová. *Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990*, Disertační práce. Praha: Fakulta informatiky a statistiky VŠE v Praze 2006

$$Q_{1-p} = e^{\hat{\mu} + \hat{\sigma}u_{1-p}} + \hat{\gamma}. \quad (53)$$

$u_p$  je 100.p% kvantil normovaného normálního rozdělení  $N(0;1)$ .

[9]

Po úpravě předchozích rovnic dostáváme

$$\hat{\mu} = \ln \frac{Q_p - Q_{1-p}}{e^{\hat{\sigma}u_p} - e^{-\hat{\sigma}u_p}}, \quad (54)$$

$$\hat{\sigma}^2 = \left( \frac{\ln \frac{Q_p - Q_{0,5}}{Q_{0,5} - Q_{1-p}}}{u_p} \right)^2, \quad (55)$$

$$\hat{\gamma} = Q_{0,5} - e^{\hat{\mu}}. \quad (56)$$

## **3 Data**

Pro následující analýzu a modelování příjmových ukazatelů jsem se rozhodl použít data z Českého a Slovenského statistického úřadu. Jako nejlépe zpracované, věrohodné a pro moji práci nejužitečnější jsem zvolil údaje z podrobných statistických šetření nazvaných Mikrocensus a Silc.

### **3.1 *Mikrocensus ČR 2002***

Účelem tohoto šetření bylo získat reprezentativní údaje o úrovni a struktuře příjmů a základní sociálně demografické charakteristiky nejen domácností, ale také jejich členů. Příjmové údaje jsou publikovány za nejmenší ekonomickou jednotku, čili domácnosti. Pro Mikrocensus 2002 bylo vybráno 11 040 bytů, tj. zhruba 0,25% (v hl. m. Praze 0,5%) z celkové počtu trvale obydlených bytů. Za neobydlené byty se náhradní nevybírají. Jako důležité je třeba také zmínit, že výběrový plán byl založen na náhodném dvoustupňovém výběru. V průběhu šetření bylo zjištěné, že z 11 040 bytů bylo 351 bytů neobydlených. Jak je výše zmíněno, za neobydlené byty se náhradní nevybírají, proto šetření proběhlo v 10 689 bytech. Jelikož se 3011 bytových domácností nepodařilo vyšetřit, konečné číslo vyšetřených bytových domácností je 7678.

### **3.2 *Mikrocensus SR 2003***

Při výběru bytových domácností pro Mikrocensus 2003 se použil dvojstupňový stratifikovaný systematický náhodný výběr. V prvním stupni byly vybrány jednotlivé obce. V druhém stupni se z vybraných obcí vykonal výběr bytů. Jednotkou výběru byl obývaný byt, jednotkou zjišťování byly všechny osoby, které bydlely v bytě k 31.12. 2002. Jednotkou zpracování pak byla hospodařící domácnost. Rozdílem oproti českému Mikrocenzu bylo to, že pokud domácnost odmítla odpovědět na zjišťované údaje, tazatel měl možnost při dodržení

předepsaných pokynů najít náhradní domácnost. Z celkového počtu 19 569 vyšetřených domácností bylo 5365 náhradních, tj. 27,4%.

### **3.3 *Silc ČR 2005***

Šetření Silc ČR má velmi podobnou metodiku jako Mikrocensus. Rozdíl je pouze v menším vzorku šetřených domácností a ve větším počtu zjišťovaných údajů. Šetření proběhlo v dubnu a květnu roku 2005. „Zjišťování se uskutečnilo ve všech krajích. Celkem bylo pro šetření vybráno 7000 bytů. Výběrový plán byl založen na náhodném dvoustupňovém výběru pro každý kraj nezávisle tak, aby celkový počet vybraných bytů byl úměrný velikosti jednotlivých krajů.“<sup>40</sup> Za neobydlené byty se náhradní nevybíraly. Ze 7000 vybraných domácností se podařilo vyšetřit 4351, což je 64,8%.

### **3.4 *Silc SR 2005***

Výběrové šetření EU SILC 2005 bylo realizované na vzorku 6016 domácností s použitím jednostupňového stratifikovaného výběru. Z celkového počtu 6016 domácností se jich podařilo vyšetřit 5147, což je 85,5%.

---

<sup>40</sup> [http://www.czso.cz/csu/2007edicniplan.nsf/t/6F002A030E/\\$File/301207mc.pdf](http://www.czso.cz/csu/2007edicniplan.nsf/t/6F002A030E/$File/301207mc.pdf)



## 4 Použité programy

### 4.1 *Microsoft Excel*

Microsoft Excel je tabulkový procesor od firmy Microsoft. Je možné ho charakterizovat jako účinný nástroj pro analýzy, sdílení a správu informací napomáhající činit informovanější rozhodnutí. Tento program jsem použil, protože ho používám již několik let a mám s ním velmi dobré zkušenosti. Oproti ostatním specializovaným statistickým programům má další výhodu v jeho velké dostupnosti. Využil jsem hlavně jeho statistické funkce a grafické výstupy.

### 4.2 *Statgraphics*

Statgraphics je specializovaný statistický program. V současné době se analýza dat v praxi provádí téměř výhradně pomocí statistického softwaru. Tento program jsem si opět vybral z důvodu jeho znalosti, jelikož při výuce statistických předmětů na fakultě managementu se používal právě tento software. „Pomocí tohoto balíčku lze sumarizovat a prezentovat data v podobě různých tabulek, grafů a charakteristik, hledat kvantily či hodnoty distribuční funkce nebo provádět indukční výsledky.“<sup>41</sup> Já jsem tento software využil zejména pro výpočet parametrů jednotlivých rozdělení, pro výpočet kvantilů, a poté při prezentaci výsledků ve formě Q-Q grafů.

---

<sup>41</sup> M. Artlová, D. Bílková, E. Jarošová, Z. Pourová, Příklady k předmětu statistika A, Praha, Oeconomica. 2004

## 5 Praktická část

### 5.1 Analýza a modelování celkových příjmů domácností v České Republice za rok 2002

#### 5.1.1 Analýza a interpretace kvantilových grafů

Vstupními daty jsou celkové příjmy domácností z Mikrocenzu 2002. Nejdříve jsem seřadil vzestupně všechny hodnoty příjmů a postupně vypočetl hodnoty ukazatelů uvedených v oddílu 2.5.1.1 Pro všechny výpočty jsem použil tabulkový program Excel. Všechny následující grafické výstupy jsou vytvořeny v tomto programu. V tabulce č. 1 uvádím ukázkou výsledných výpočtů.

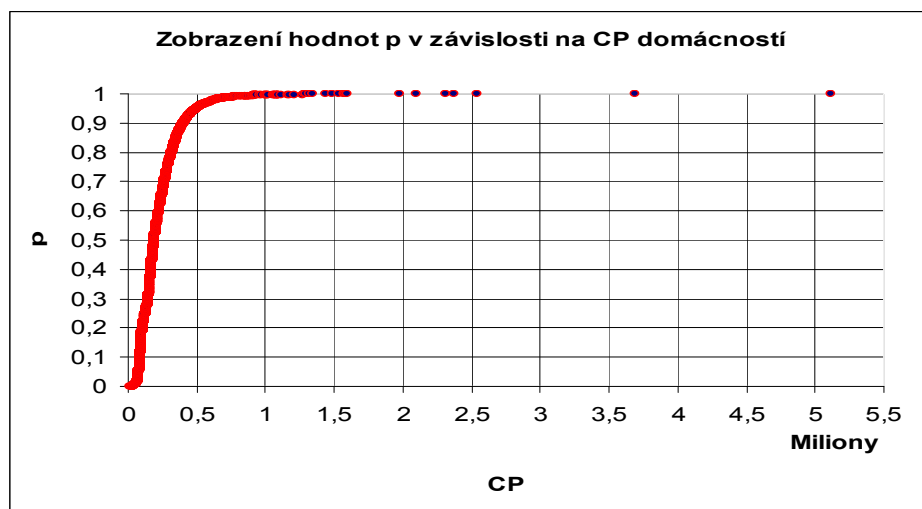
Tabulka č. 1 – ukázkou výpočtu pro grafickou analýzu

1	i	cp_prij=x	$p=(i-0,5)/n$	Dx	Dp	Midx	Midp	Dx/Dp	Dp/Dx
2	1	13500	6,27E-05	x	x	x	x	x	x
3	2	25001	0,000188	11501	0,000125	19250,5	0,000125	91697473	1,09054E-08
4	3	25200	0,000314	199	0,000125	25100,5	0,000251	1586627	6,30268E-07
5	4	28800	0,000439	3600	0,000125	27000	0,000376	28702800	3,48398E-08
6	5	28920	0,000564	120	0,000125	28860	0,000502	956760	1,04519E-06
7	6	29400	0,00069	480	0,000125	29160	0,000627	3827040	2,61299E-07
8	7	29436	0,000815	36	0,000125	29418	0,000753	287028	3,48398E-06
9	8	31008	0,000941	1572	0,000125	30222	0,000878	12533556	7,97858E-08
10	9	31320	0,001066	312	0,000125	31164	0,001003	2487576	4,01998E-07
11	10	33415	0,001192	2095	0,000125	32367,5	0,001129	16703435	5,98679E-08
12	11	34200	0,001317	785	0,000125	33807,5	0,001254	6258805	1,59775E-07

#### 5.1.1.1 Distribuční funkce

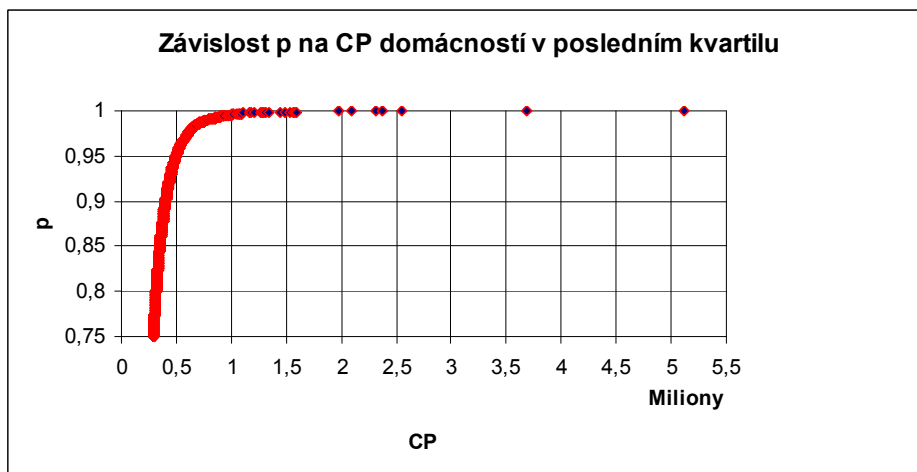
První částí grafické analýzy budou grafy ukazující závislost p na celkových příjmech (dále jen CP) – tedy znázornění distribuční funkce.

Graf č. 1



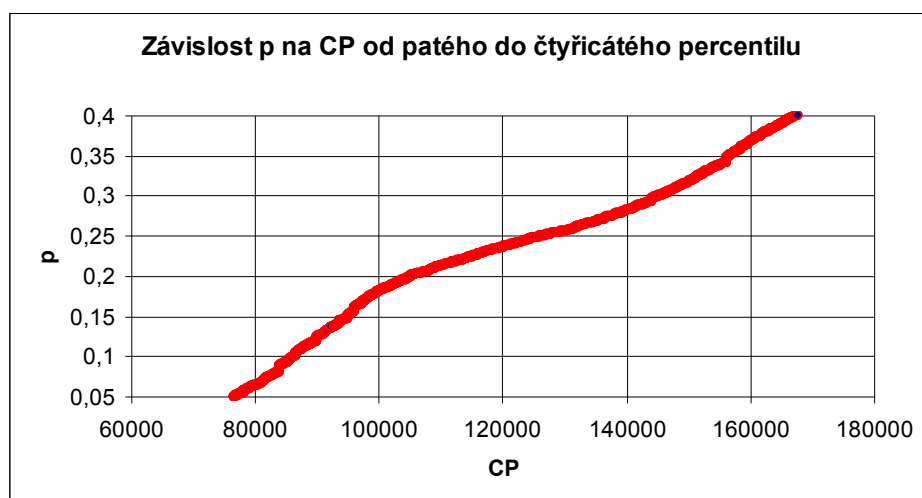
Z grafu je možné zpozorovat velmi krátký levý konec rozdělení oproti konci pravému. To je možné zdůvodnit minimální výší mzdy, důchodu a také různými příspěvky státu osobám, které jsou v těžké sociální situaci. Naopak pravý konec je velmi dlouhý a v jeho koncích je možné vidět velké diference. Před transformací před rokem 1989 zcela určitě takové rozdíly nebyly a postupem času lze předpokládat prodlužování pravého konce. Největší hodnotu směrnice je možné najít mezi 70 až 500 tis. Kč, která poukazuje na nejvyšší pravděpodobnost hodnot v tomto intervalu. Od hodnoty 500 tis. Kč hodnoty narůstají, hodnota směrnice klesá a hodnoty jsou čím dál tím méně pravděpodobné. Pro podrobnější znázornění a důkladnější analýzu jsem sestrojil grafy jen v určitých kvantilech – grafy č. 1.1 – 1.3

Graf č. 1.1 a 1.2



Zobrazení prvního kvartilu prokázalo menší pravděpodobnost hodnot do 70 tis. Kč, ale zároveň také poukázalo na opětovný pokles směrnice od cca 100 tis. Kč, na který se zaměřím v následujícím grafu. Obrovské rozdíly v CP nám ukazuje graf č. 3. Mezi 1,5 až 5,1 mil. Kč se vešlo pouze 0,15% domácností. Jsou to tedy extrémní hodnoty, které budou výrazně zkreslovat např. průměr.

Graf č. 1.3

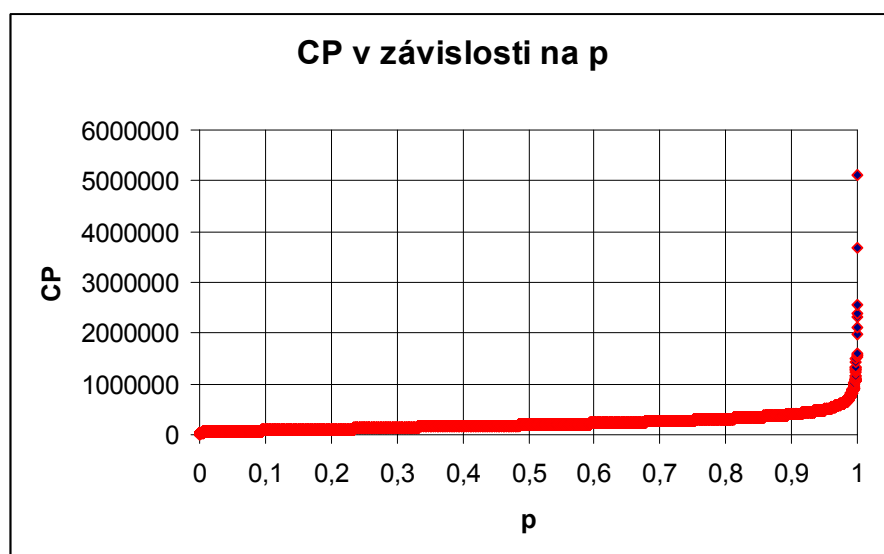


Zobrazení od pátého do čtyřicátého percentilu je záměrné z důvodu opětovného poklesu směrnice, jak jsem naznačil výše. Graf č. 1.3 nám ukazuje menší pravděpodobnost hodnot od 100 do 160 tis. Kč. Je těžké tento fakt zdůvodnit, ale podle mého osobního názoru se v tomto intervalu nachází jednočlenné a dvoučlenné domácnosti důchodců a jednočlenné domácnosti zaměstnaných.

### 5.1.1.2 *Kvantilová funkce*

Druhou částí grafické analýzy je výstup, který ukazuje závislost celkových příjmů na hodnotě p. Poukazuje na to, jak se mění hodnota celkových příjmů v závislosti na proporčním postavení.

Graf č. 2

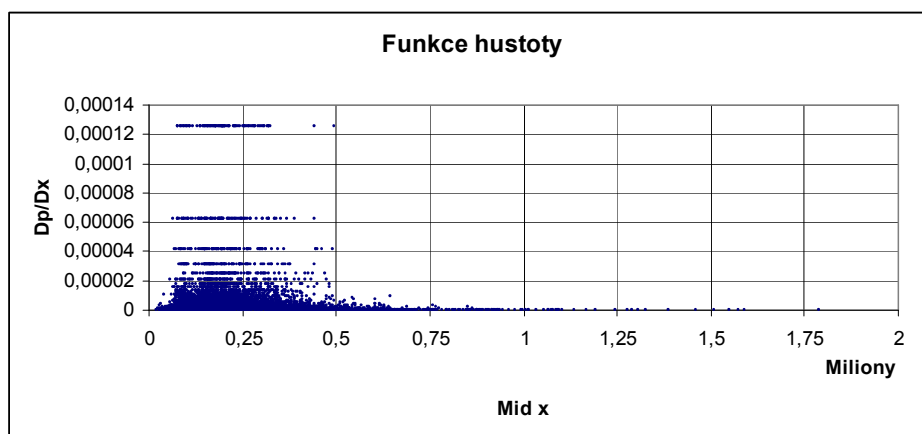


Graf č. 2 nám pouze dokazuje, že pokud bychom neudělali podrobnější zobrazení v dolním intervalu, ani bychom nepostřehli levostranné zakřivení, které je opravdu extrémně krátké. Na druhé straně pravostranné zakřivení je velmi zřetelně vidět i bez detailnějšího náhledu, který už nebudu pro kvantilovou funkci dále rozvádět, jelikož by potvrdil závěry z funkce distribuční.

### 5.1.1.3 *Funkce hustoty*

Dalším grafickým výstupem je zobrazení směrnic růstu  $Dp/Dx$  vzhledem ke středu příslušné difference  $Dx$ . Rychlost změny  $Dp$  je nejvyšší v intervalu od 80 do 500 tisíc Kč. Většina relativních přírůstků je do 0,002 % celkových příjmů. Ojediněle se v tomto intervalu objevují hodnoty dosahující až 0,013%. Velice zajímavé je, že větší přírůstky jsou v dolní a střední části rozdělení, zatímco u vyšších hodnot jsou přírůstky skoro nulové. Pro lepší zobrazení jsem omezil výběr  $Mid\ x$  do 2 milionů Kč.

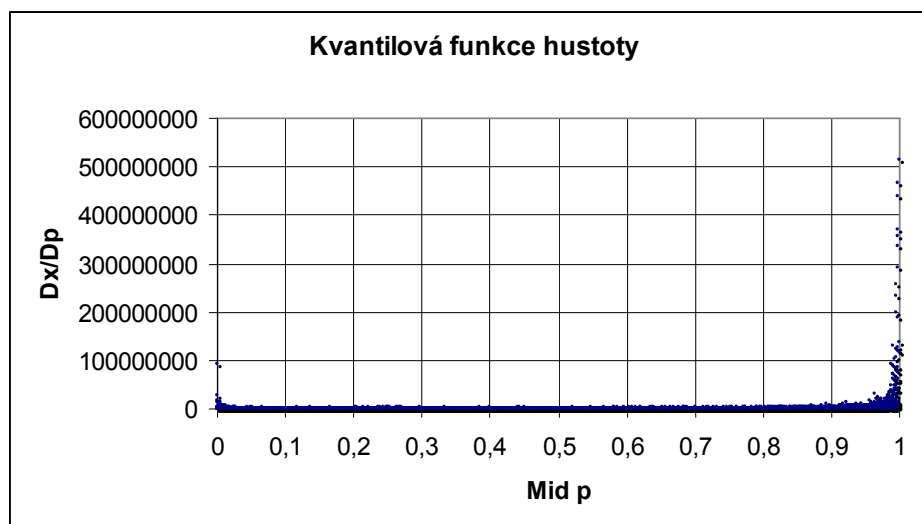
Graf č. 3



#### 5.1.1.4 *Kvantilová funkce hustoty*

Posledním grafem tohoto druhu je zobrazení směrnice  $Dx/Dp$  vzhledem ke středu příslušné difference  $Dp$  (graf č. 4). Pokud se podíváme na oba konce rozdělení, je patrné, že v pravém konci jsou hodnoty relativních přírůstků o dost vyšší než v levém. Dále je možné postřehnout, že ve střední části jsou tyto přírůstky menší a s vyšší intenzitou než v obou koncích. Proto je možné považovat rozdělení za asymetrické s prodlouženým pravým koncem. Pro lepší znázornění jsem opět omezil výběr, tentokrát na ose Y do 600 000 000.

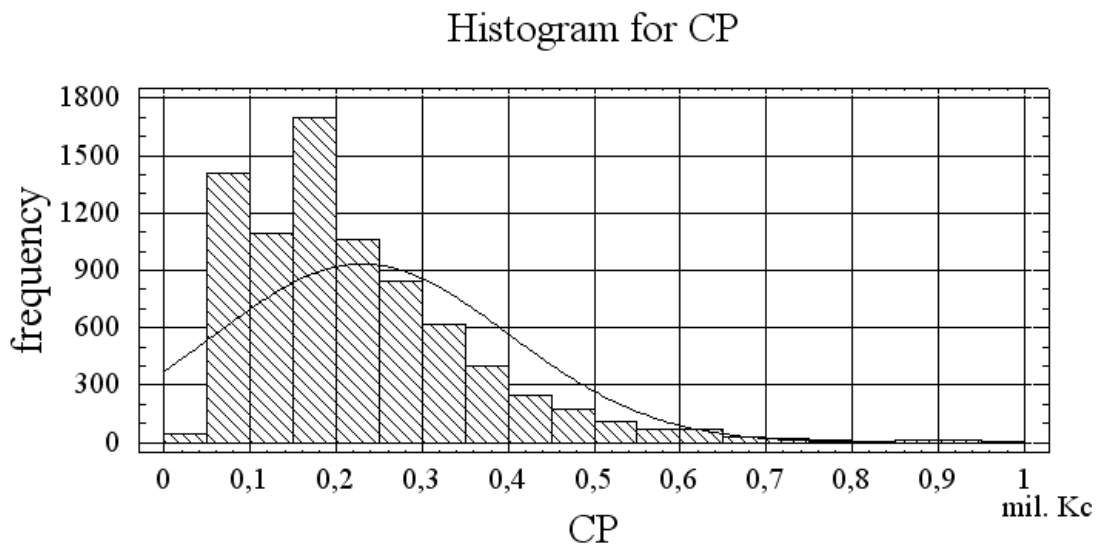
Graf č. 4



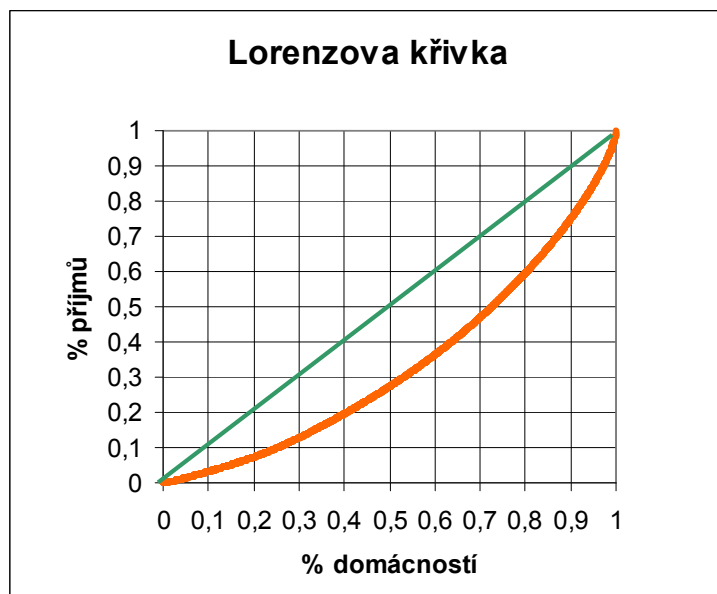
### 5.1.1.5 *Ostatní grafické analýzy*

Histogram pouze potvrdil výsledky dřívějších analýz. Opět zde vidíme „useknutou“ levou část rozdělení a ubývání četností s výškou čistých příjmů. Nejvíce domácností se nachází v intervalu od 150 do 200 tis. a překvapivě druhým nejčetnějším intervalem je 50-100 tis. Kč. Podle doplňující křivky normálního rozdělení v histogramu bych se přiklonil k levostrannému zešikmení. Ale to prokážou až výpočty v následující kapitole. Pro lepší zobrazení jsem použil data čistých příjmů do 1 mil. Kč.

Graf č. 5



Graf č. 6 – Lorenzova křivka





Tento graf nám ukazuje, jak se skutečná křivka rozdělení příjmů domácností v ČR odchyľuje od ideální rovnoměrnosti. Můžeme zde vyzorovat, že např. 10% domácností s nejmenšími příjmy se podílí na celkových příjmech pouze ze 3,2%. A 20% domácností s nejnižšími příjmy nedostává zdaleka ani 10% (přesně to je 7,3%). Na druhé straně 10% nejvíce vydělávajících domácností inkasuje 25% celkových příjmů. Zajímavé bude určitě srovnání s hodnotami ve SR.

## 5.1.2 Statistické ukazatele

Tabulka č. 2

<b>Průměr</b>	230004,0	<b>Horní kvartilová dif.</b>	97420
<b>Medián</b>	190923,0	<b>Dolní kvartilová dif.</b>	64900
<b>Modus</b>	84000,0	<b>Galtonův koef. šikmosti</b>	0,20
<b>Rozptyl</b>	2,91187E10	<b>Moorsova špičatost</b>	0,81
<b>Směrodatná odchylka</b>	170642,0	<b>Galtonova p šikmost</b>	
<b>Minimum</b>	13500,0	<b>p= 0,375</b>	0,08
<b>Maximum</b>	5,11063E6	<b>p= 0,125</b>	0,51
<b>Koeficient šikmosti</b>	6,57632		
<b>Koeficient špičatosti</b>	123,169		

Kromě standardních charakteristik vypočítaných v programu Statgraphics Plus jsem vypočítal i méně známé kvantilové charakteristiky v programu Microsoft Excel. Z důvodu obsáhlosti těchto charakteristik uvádím jen některé jejich hodnoty, ostatní jsou uveřejněny v příloze č. 1.

### 5.1.2.1 Interpretace charakteristik

Nejnižší zjištěný příjem u 7972 domácností byl 13 500 Kč, nejvyšší pak 5 110 628 Kč. Variační rozpětí je tedy pak 5 097 128 Kč. Tato výše je však ovlivněna extrémními příjmy a je pro nás nevyovídající. Průměrný příjem domácností za rok 2002 byl 230 004 Kč. Na průměr dosáhlo 38% domácností. Proto je reprezentativnější hodnotou medián, který činil 190 923 Kč. To znamená, že přesně polovina domácností má příjem menší než je tato hodnota. Medián je tedy oproti průměru nižší o 39 081 Kč. První decil má hodnotu 86 398 Kč, což znamená, že 10% domácností má nižší příjem než je tato hodnota. Pokud se podíváme na opačný konec můžeme říct, že v České Republice je 10% domácností, které mají větší příjem než 402 132

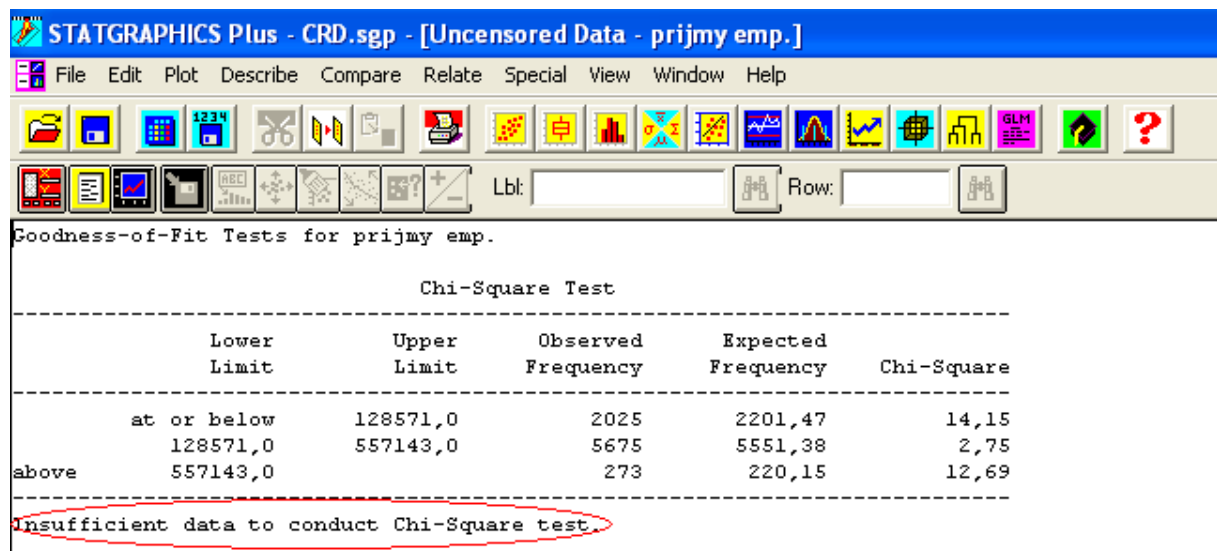
Kč, což odpovídá devátému decilu. Směrodatná odchylka, která je 170 642 Kč nám říká, že většina domácností se neodchýlila od průměru o více než tuto hodnotu v obou směrech. Koeficient šikmosti je 6,58 a je tedy kladný. Z toho můžeme usuzovat, že rozdělení není symetrické a jedná se o kladné zešikmení, což znamená, že převažují domácnosti s menšími příjmy než domácnosti s příjmy vyššími. Z vypočítané hodnoty míry špičatosti, která je 123,17 můžeme usuzovat, že rozdělení je výrazně špičatější než normované normální rozdělení. Rozdíl horního kvartilu a mediánu – horní kvartilová diference ( $u_{qd} = 97\,420$ ) je větší než dolní kvartilová diference ( $l_{qd} = 64\,900$ ). Z toho vyplývá, že kvartilový podíl je větší než 1. Galtonův koeficient šikmosti je tedy pak kladný a má hodnotu 0,20 což potvrzuje kladné, pravostranné zešikmení příjmů domácností. Tento ukazatel se liší oproti výše vypočtenému koeficientu šikmosti tím, že měří zešikmení prostřední poloviny s vrcholem rozdělení. Nebere tedy v úvahu extrémní hodnoty a je přesnější. Tato velikost střední části lze měnit. Galtonova p-šikmost pro prostředních 25% hodnot ( $p=0,375$ ) je kladná a má hodnotu ( $g(p)=0,08$ ). Galtonova p-šikmost pro prostředních 75% hodnot ( $p=0,125$ ) má pak hodnotu ( $g(p)=0,51$ ) Čím větší část rozdělení bereme v úvahu, tím je Galtonův koeficient šikmosti větší. Moorsova špičatost je chápána jako vlastnost výšky vrcholu rozdělení s vyloučením vlivu konců rozdělení od dolního oktilu dolů a horního oktilu nahoru. Spočtenou nízkou hodnotu ( $k=0,81$ ) je tedy možné charakterizovat jako malou odlišnost špičatosti vrcholu empirického rozdělení příjmů domácností v porovnání se špičatostí normálního rozdělení. Zde je tedy možné vidět rozdíl s vypočtenou hodnotou koeficientu špičatosti z programu Statgraphics, kde tento koeficient evidentně ovlivňují odlehlé hodnoty.

### **5.1.3 Modelování jednoduchými pravděpodobnostními funkcemi**

Při hledání jednoduchého tvaru rozdělení čistých reálných ročních příjmů domácností jsem použil programy Statgraphics a Microsoft Excel. Při hledání parametrů jednotlivých rozdělení byla použita metoda maximální věrohodnosti. Poté jsem testoval shodu empirického a teoretického rozdělení pomocí  $\chi^2$  testu a zejména pomocí grafických metod, které nám často napoví více.

Při výpočtu  $\chi^2$  testu se vyskytla komplikace. Program Statgraphics nedovedl u většiny rozdělení tento test vypočítat. (obr. č. 1) Při tomto testu se totiž jednotlivé hodnoty seskupují do tříd a porovnává se četnost empirického rozdělení s četností s teoretickým tvarem. Program Statgraphics seskupuje třídy podle velikosti příjmu a ne podle četnosti. Pak vzniká problém, že není splněna podmínka dostatečně velkých tříd (42). Důvodem byl velký počet odlehlých hodnot, jak ukazuje krabíčkový diagram. (příloha č. 2). Pokusil jsem se tedy odlehlé hodnoty vyloučit a vypočítat  $\chi^2$  test bez nich. Po té sice Statgraphics u většiny rozdělení test vypočetl, ale nebylo možné měnit počet stupňů volnosti. Jelikož byl podle mého názoru počet tříd pro tak velkou oblast hodnot příliš nízký, rozhodl jsem se vypočítat  $\chi^2$  test dobré shody v programu Microsoft Excel. Pro počet tříd jsem použil vzorec (24).

Obr. č. 1



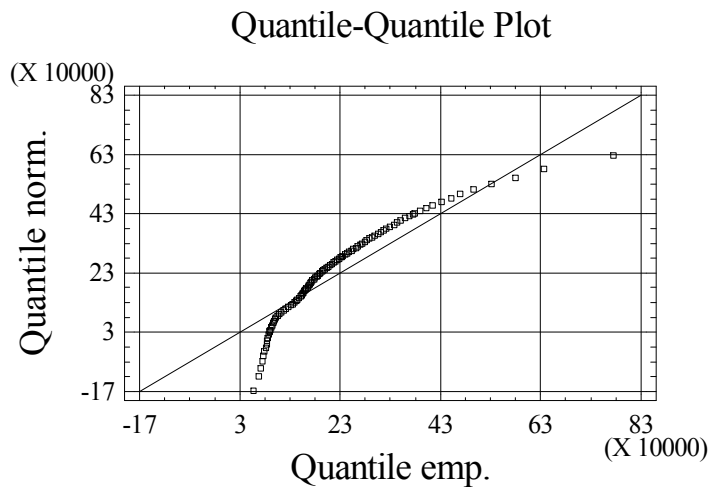
Tabulka č. 3 – Porovnání empirického rozdělení CP domácností s jednoduchými teoretickými tvary.

tvar rozdělení	empirické	Normální	Lognorm.	Weibullovo	Paretovo	Gama	Erlang
parametry odh. met. max věrohodnosti	x	230004	228526	1,56	0,08	2,89766	3
	x	170642	148381	257948	x	1,26E-05	1,30E-05
Chi-kvadrát	x	5589,83	902,72	2057,34	x	1752,36	1715,36
kritická hodnota	x	69,13	69,13	67,37	x	67,37	67,37
p-hodnota	x	0	2,5E-134	0	x	1,8E-307	7,74E-300
kvantily							
0,01	57351	-166969	48230	13579	1,13	31858	33431
0,02	67195	-120453	56694	21230	1,28	41748	43487
0,03	71901	-90939	62819	27611	1,45	49143	50969
0,04	74400	-68737	67858	33302	1,64	55329	57208
0,05	76800	-50678	72254	38543	1,87	60776	62691
0,25	126023	114907	128473	116204	33,09	130696	132249
0,5	190923	230604	191667	204012	4587,58	204155	205015
0,75	288343	345100	285947	317924	21000000	301381	300569
0,95	495967	510685	508434	520601	6,68E+15	487540	482686
0,96	530990	528745	541371	545098	1,6E+17	511347	505925
0,97	580024	550947	584801	575796	3,33E+18	541589	535434
0,98	635700	580466	647976	617566	4,62E+20	583473	576283
0,99	775566	626977	761865	685528	2,12E+24	653433	644467

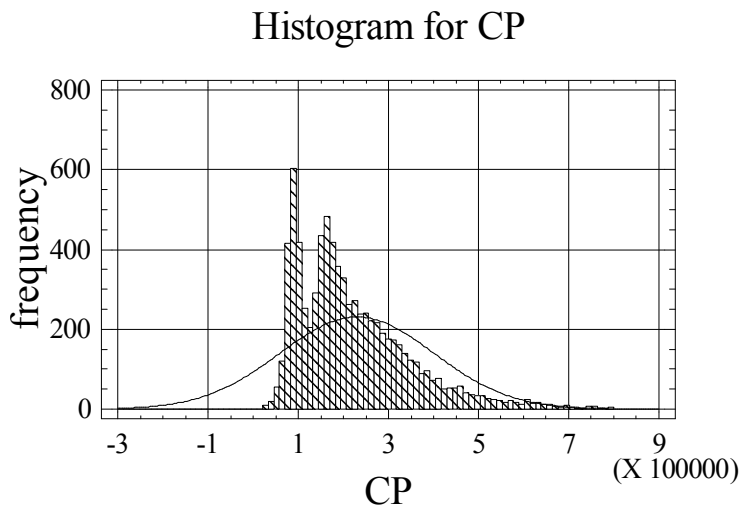
Jak ukazuje tabulka č. 3, podle chi-kvadrát testu dobré shody ani jeden jednoduchý teoretický tvar nemodeluje dostatečně dobře celé rozdělení CP. Nejnižší hodnotu testovacího kritéria má lognormální rozdělení, ale i tato hodnota výrazně převyšuje kritický obor. U Paretova rozdělení jsem testové kritérium nepočítal z důvodu obtížného seskupování do tříd a navíc je předem jisté, že toto rozdělení nemůže modelovat teoretický tvar v celém funkčním oboru. V další části práce se tedy zaměřím na modelování pouze jednotlivých částí empirického rozdělení.

### 5.1.3.1 Normální rozdělení

Graf č. 7



Graf č. 8

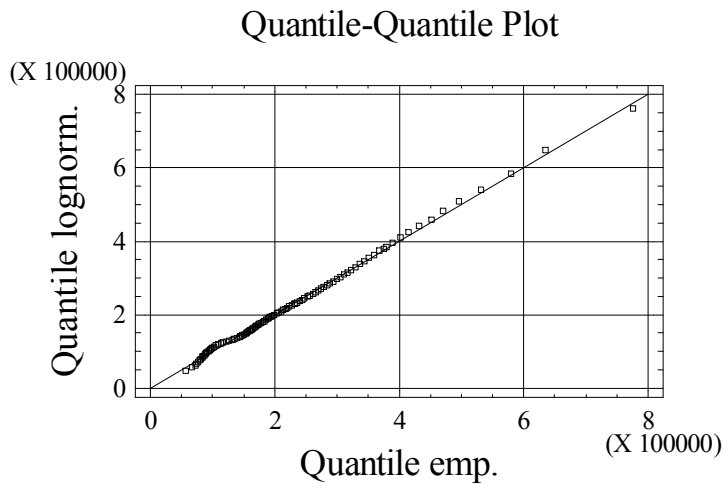


Tato grafická analýza potvrdila vysokou hodnotu  $\chi^2$  testu dobré shody. Graf č. 7 ukazuje porovnání empirického a teoretického normálního rozdělení s linií 45 stupňů. Čím vyšší je shoda empirického a teoretického rozdělení, tím jsou body blíže k přímce. V dolní a horní části normální rozdělení podhodnocuje skutečně naměřené CP domácností. Naopak ve střední část je nadhodnocuje. Můžeme tedy říct, že normální rozdělení není možné považovat za vhodné, a to ani pro některou jeho část. Pokud se podíváme na histogram v grafu č. 8, tak od 80-250 tis. Kč je skutečných četností daleko více, než teoretických četností normálního

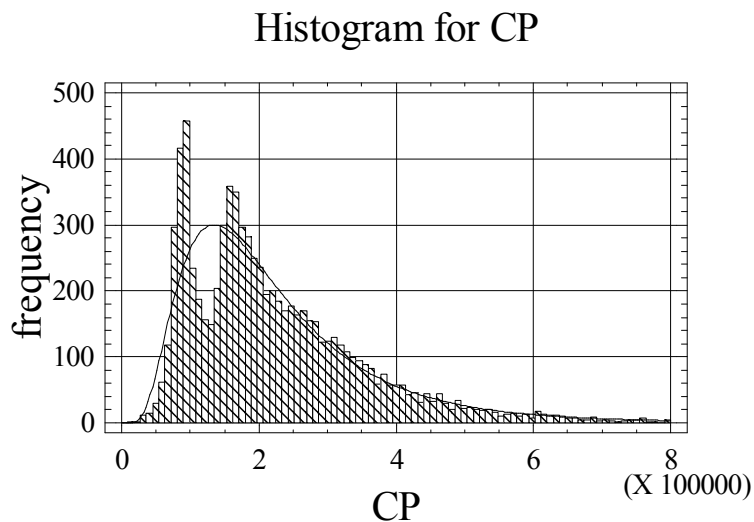
rozdělení, od 250 tis. Kč je tomu naopak. To jen potvrzuje závěr, že normální rozdělení je pro CP domácností nevhodné.

### 5.1.3.2 *Lognormální rozdělení*

Graf č. 9



Graf č. 10

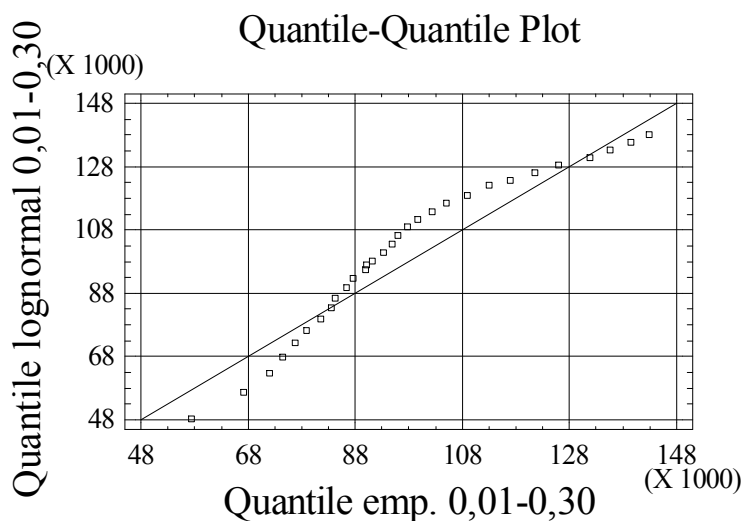


Pokud se podíváme na graf č. 9, který zobrazuje porovnání empirického a teoretického normálního rozdělení, můžeme vidět, že většina bodů se nachází velmi blízko linie 45 stupňů.

To vyvrací výsledek  $\chi^2$  testu dobré shody, který výrazně překročil kritickou hodnotu. Pokud se podíváme na graf č. 10, je vidět, že tento výsledek ovlivnily rozdíly četností, a to

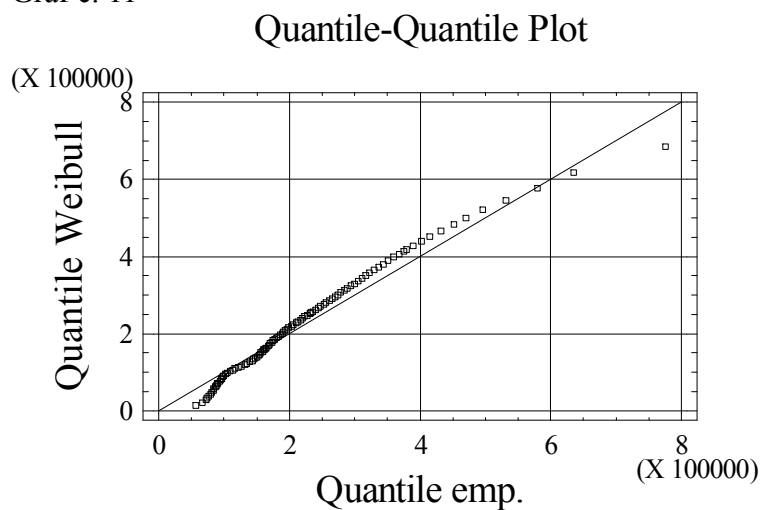
v intervalu od 70-170 tis. Kč, kde nejdříve lognormální rozdělení podhodnocuje počet tříd do 100 tis. Kč a následně nadhodnocuje. V ostatních intervalech se však empirické četnosti velmi shodují s četnostmi lognormálního rozdělení. Graf č. 9 ukazuje, že je možné považovat lognormální rozdělení za vhodné v úseku od dolního kvartilu a to až do devátého decilu. Velkou dolní část empirického rozdělení nemodeluje přesně. To ukazuje detailnější porovnání empirického a lognormálního rozdělení (graf č. 10).

Graf č. 10

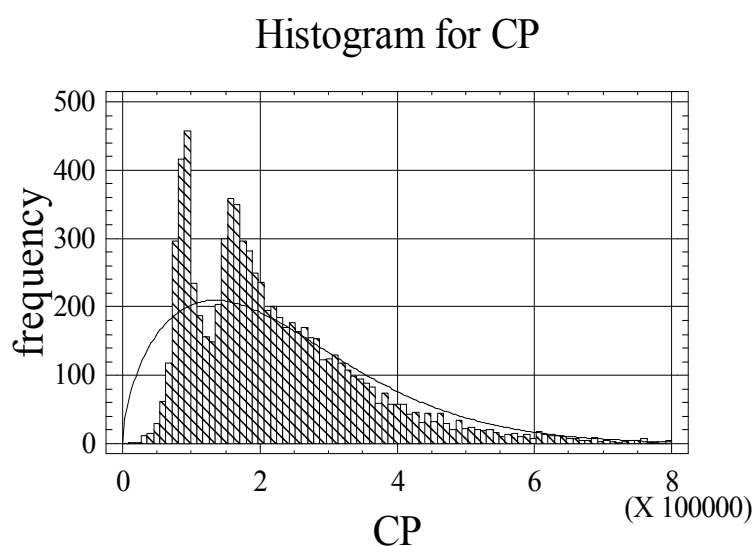


### 5.1.3.3 Weibullovo rozdělení

Graf č. 11



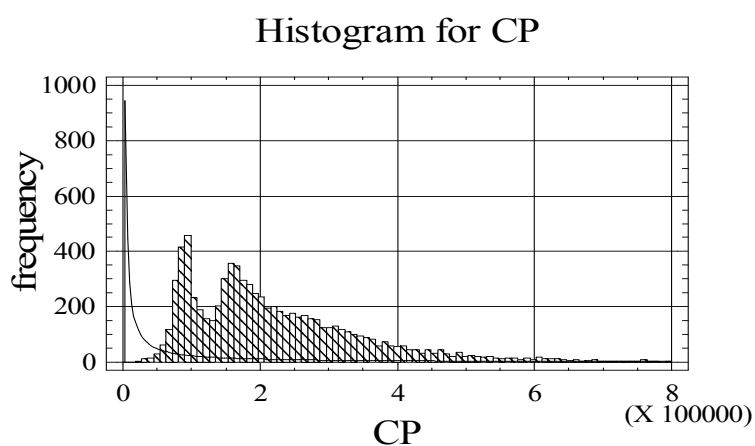
Graf č. 12



Pokud se podíváme na porovnání empirického a teoretického Weibullova tvaru v grafu č. 11, je vidět, že v dolní části toto rozdělení výrazně podhodnocuje empiricky zjištěná data, ve střední části je naopak nadhodnocuje a v horní části rozdělení je opět podhodnocuje. Není tedy možné považovat Weibullovo rozdělení za vhodné a to ani v některé jeho části.

#### 5.1.3.4 *Paretovo rozdělení*

Graf č. 13



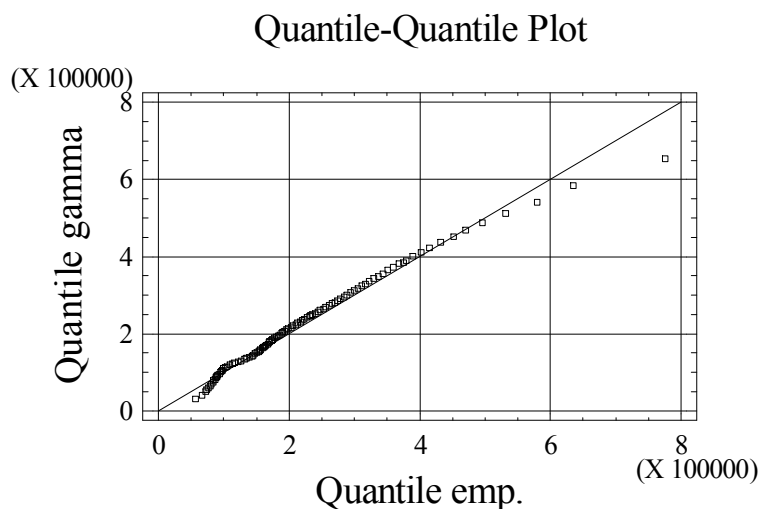
Porovnání empirického a teoretického Paretova rozdělení zde neuvádím z důvodu velké variability hodnot. První percentil má hodnotu 1,13 a devětadesátý má hodnotu  $2,12E24$ . Paretova křivka je často vhodným modelem rozdělení příjmů na okrajích. Ve středu svého intervalu je prakticky neuplatnitelná. Bohužel, ale v případě CP domácností v ČR není vhodná



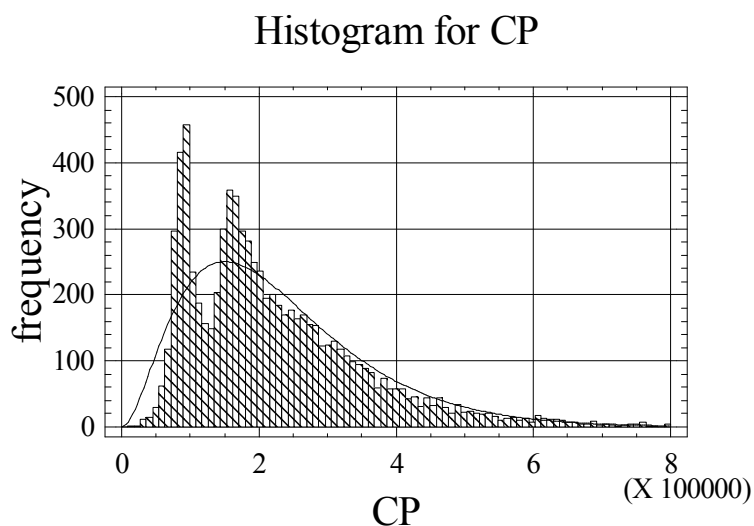
ani pro krajní oblasti. Obrovský rozdíl empirických a teoretických četností ukazuje histogram (graf č. 13).

### 5.1.3.5 *Gama rozdělení*

Graf č. 14



Graf č. 15

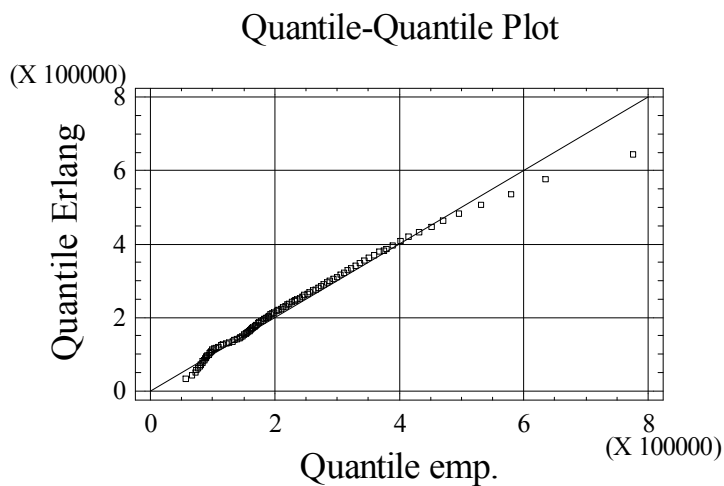


Porovnání empirického a teoretického gama rozdělení (graf č. 14) se velmi podobá porovnání výše uvedeného empirického a teoretického lognormálního rozdělení (graf č. 9). Bohužel stejně jako u něj nelze považovat v dolním kvartilu z důvodu podhodnocení skutečně naměřených hodnot do 12 percentilu a následného nadhodnocení do 25 percentilu gama rozdělení za vhodné. Na první pohled se zdá tento model pro střední část jako odpovídající,

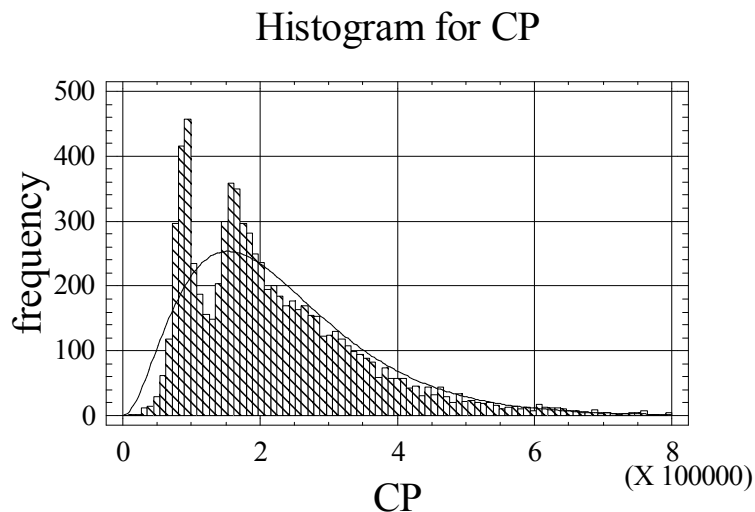
ale detailnější analýzy uvedené v příloze č. 3 přinesly jiný závěr. Od dolního kvartilu se teoretické kvantily začínají shodovat s empirickými, to však lze tvrdit pouze do 35. percentilu. Poté se opět začínají objevovat větší rozdíly a další dostatečnou shodu lze najít už jen v několika percentilech v okolí devátého decilu.

### 5.1.3.6 Erlangovo rozdělení

Graf č. 15



Graf č. 16



Erlangovo rozdělení je typem gama rozdělení, proto jsou rozdíly v grafických výstupech nepatrné. I hodnota  $\chi^2$  testu dobré shody je velmi podobná. U Erlangova rozdělení je nepatrně nižší. Bohužel však tímto typem rozdělení není možné modelovat větší část.

### 5.1.3.7 *Shrnutí*

Po důkladné výpočetní a grafické analýze je možné říct, že ani jeden jednoduchý tvar není možné považovat za vhodný v celé délce empirického rozdělení. Důkladnější grafické analýzy však přinesly zjištění, že střední část je možné namodelovat lognormálním rozdělením. Konkrétně je to oblast od dolního kvartilu až po devátý decil. Bohužel ani jeden jednoduchý tvar nemodeluje dostatečně dobře dolní část.

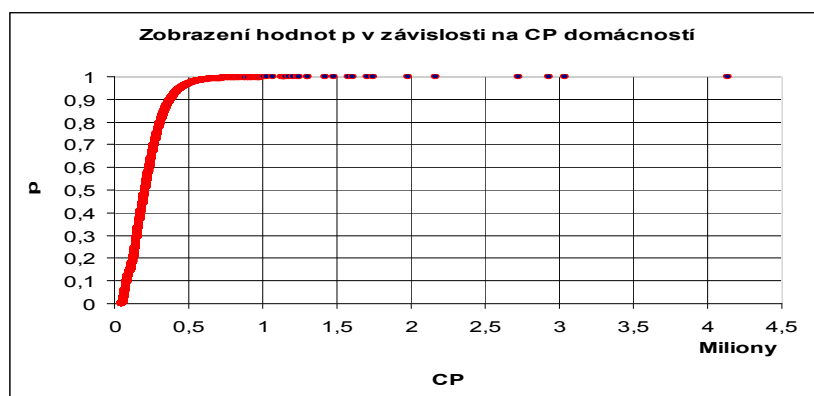
## 5.2 *Analýza a modelování celkových příjmů domácností ve Slovenské republice za rok 2003*

V následující kapitole budu vycházet z dat Mikrocenzu SR 2003. Vypočtu stejné ukazatele a tam kde to půjde, tak je porovnáám s ukazateli zjištěných z dat v České Republice. Data nebudu přepočítávat na české koruny z několika důvodů. Tím hlavním důvodem je to, že by se skutečně naměřená data po přepočtení zkreslila, a tím by se následně zkreslily i výsledky hledání vhodného modelu. Druhým důvodem je to, že i kdybychom přepočtli hodnoty na české koruny a poté je porovnávali s absolutními čísly v České republice, jejich vypovídací hodnota by byla omezená z důvodu odlišné kupní síly v obou zemích.

### 5.2.1 *Analýza a interpretace kvantilových grafů*

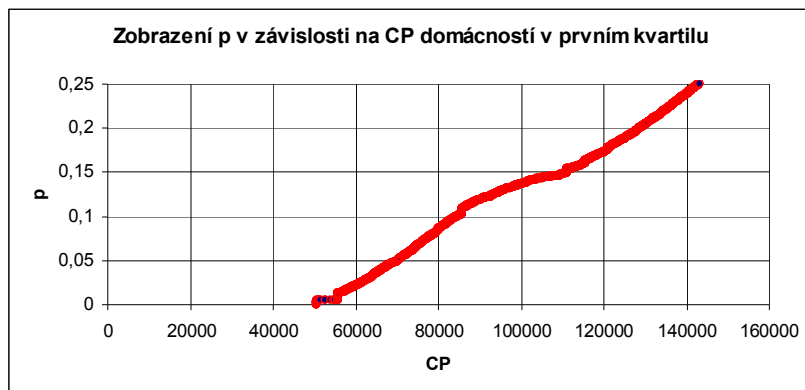
#### 5.2.1.1 *Distribuční funkce*

Graf č. 17

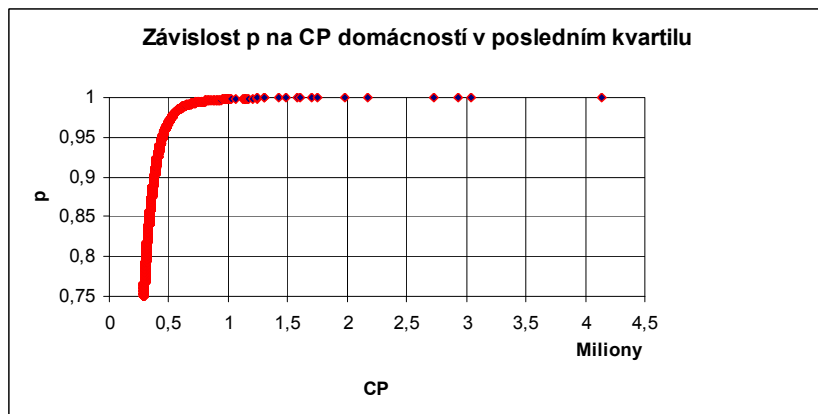


Distribuční funkce celkových příjmů domácností ve SR je velmi podobná funkci zjištěné z dat českých domácností. I zde můžeme vidět velmi krátký levý konec rozdělení oproti konci pravému. V pravém konci jsou opět velké diference. Největší pravděpodobnost hodnot je mezi 60 – 450 tis. Sk, kde je největší sklon směrnice. Pro podrobnější analýzu jsem opět sestrojil grafy v určitých kvantilech – grafy č. 17.1 – 17.3

Graf č. 17.1

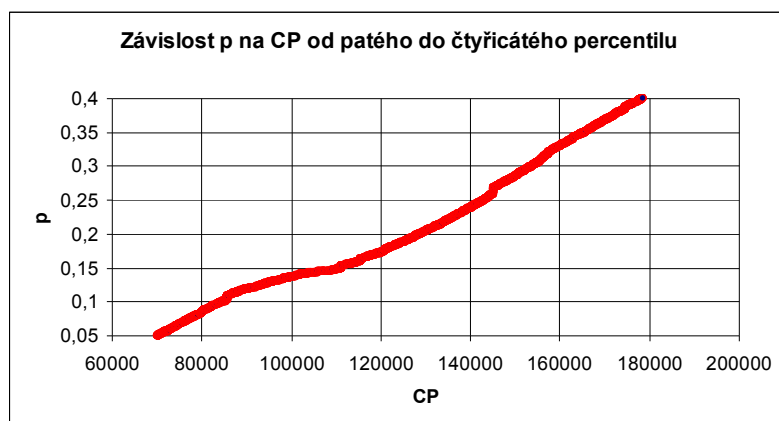


Graf č. 17.2



Zobrazení prvního kvartilu přineslo poznání, že levý konec rozdělení je o mnoho kratší než v České republice. Velké rozdíly příjmů jsou patrné již od 1 mil. Sk, nad tuto hodnotu se dostalo pouze 0,2% domácností. Výraznější je i „zub“, který začíná v hodnotě 110 tis. Sk a končí v hodnotě 145 tis. Sk. V tomto intervalu je menší pravděpodobnost hodnot. Od hodnoty 145 tis. Sk směrnice opět roste a až do výše 450 tis. Sk je stejná.

Graf č 17.3



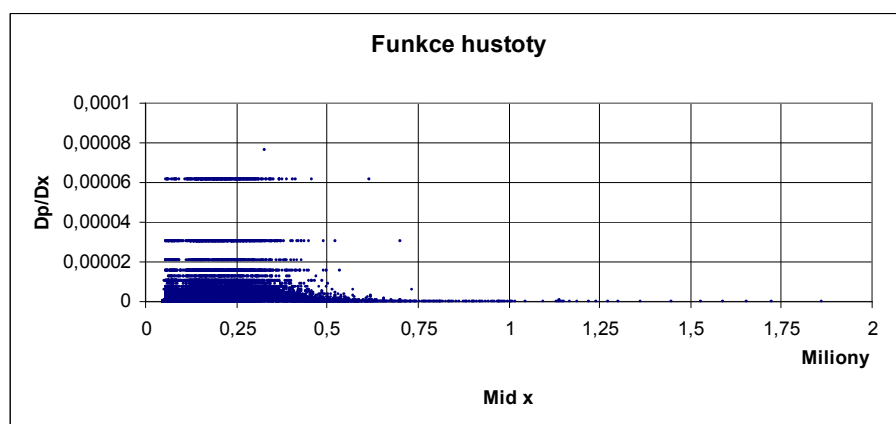
### 5.2.1.2 *Kvantilová funkce*

Kvantilová funkce nepřinesla žádné nové poznatky. Pouze potvrzuje závěry z funkce distribuční. Z tohoto důvodu zde neuvádím ani její grafický výstup

### 5.2.1.3 *Funkce hustoty*

Rychlost změny  $Dp$  je nejvyšší v intervalu od 60 do 460 tis. Sk. Většina relativních přírůstků je do 0,001% celkových příjmů. Ojedinele se objevují hodnoty dosahující až 0,006%. To je dvakrát méně než v České republice. To podle mého názoru způsobuje větší počet naměřených dat. Stejně jako v České republice jsou větší přírůstky v dolní a střední části rozdělení. U vyšších hodnot jsou skoro nulové.

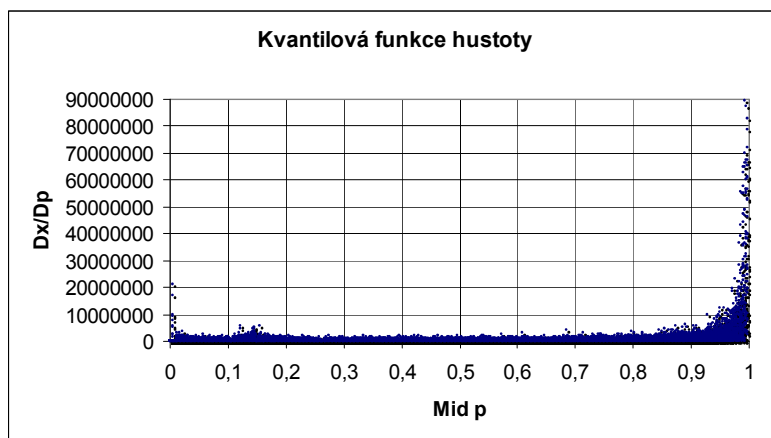
Graf č. 18



### 5.2.1.4 *Kvantilová funkce hustoty*

Stejně tak, jako tomu bylo u dat z ČR, v pravém konci jsou hodnoty relativních přírůstků o dost vyšší než v levém konci. Ve střední části jsou přírůstky menší než v obou koncích. Rozdělení lze tudíž opět považovat za asymetrické s prodlouženým pravým koncem. Také zde bylo potřeba omezit výběr a to na ose Y do 900 000 000.

Graf č. 19

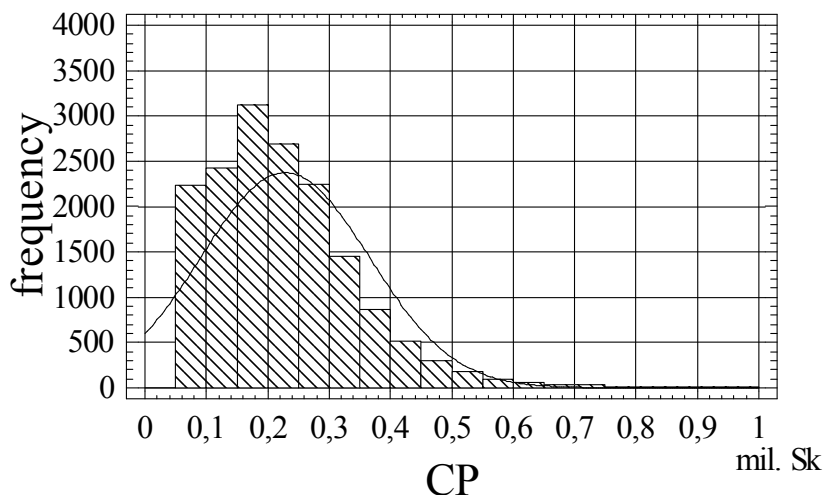


### 5.2.1.5 *Ostatní grafické analýzy*

Histogram ukazuje, že nejvíce domácností na Slovensku má celkové příjmy v rozmezí 150 – 200 tis. Sk. Tak tomu bylo i v Česku, přestože se jedná o dva různé státy s různými měnami a poměrem směnného kursu a navíc měření na Slovensku proběhlo o rok později. Opět jsem byl nucen omezit výběr, tentokrát data CP do 1 mil. Sk.

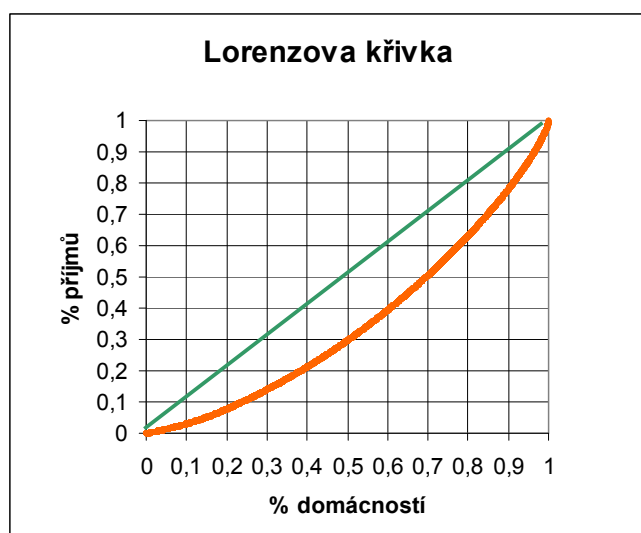
Graf č. 20

### Histogram for CP



Lorenzova křivka nám ukazuje, že 10% domácností s nejnižšími příjmy se podílí na celkových příjmech pouze ze 3%, což je o 0,2% méně, než v České Republice. A 20% domácností s nejnižšími příjmy inkasuje 7,7%, což je na druhé straně o 0,4% více než české domácnosti. Desetina nejlépe vydělávajících domácností na Slovensku pobírá 22% celkových příjmů, což je o 3% méně než v Česku. Z toho plyne, že v České republice je větší příjmová nerovnost. Po letošním zavedení rovné daně z příjmu fyzických osob se tato diference pravděpodobně ještě prohloubí.

Graf č. 21



## 5.2.2 Statistické ukazatele

Tabulka č. 4

<b>Průměr</b>	228285	<b>Horní kvartilová dif.</b>	81939
<b>Medián</b>	207096	<b>Dolní kvartilová dif.</b>	63965
<b>Modus</b>	145080	<b>Galtonův koef. šikmosti</b>	0,12
<b>Rozptyl</b>	1,88525E10	<b>Moorsova špičatost</b>	0,90
<b>Směrodatná odchylka</b>	137305	<b>Galtonova p šikmost</b>	
<b>Minimum</b>	50520	<b>p= 0,375</b>	0,02
<b>Maximum</b>	4,13479E6	<b>p= 0,125</b>	0,24
<b>Koeficient šikmosti</b>	4,65705		
<b>Koeficient špičatosti</b>	77,7244		

Kromě standardních charakteristik vypočítaných v programu Statgraphics Plus jsem vypočítal i méně známé kvantilové charakteristiky v programu Microsoft Excel. Z důvodu obsáhlosti těchto charakteristik uvádím jen některé jejich hodnoty, ostatní jsou uveřejněny v příloze č. 4.

### **5.2.2.1**            *Interpretace charakteristik*

Nejnižší zjištěný příjem u 16349 domácností byl 50 520 Sk, nejvyšší 4 134 788 Sk. Variační rozpětí je pak v tomto případě 4 084 268 Sk. Průměrný příjem domácnosti na Slovensku byl 228 285 Sk. Takto vysoký příjem pobíralo necelých 43% domácností, což je skoro stejné, jako tomu je v České republice. Medián, čili hodnota, na kterou dosahuje polovina domácností je 207 096 Sk. Medián je oproti průměru nižší o 21 162 Sk. První decil je 84 571 Sk. To tedy znamená, že 10% domácností má nižší příjem, než je tato hodnota. Na druhé straně devátý decil je 378 443 Sk. Směrodatná odchylka je 137 305. Čili většina domácností se neodchyluje od průměru o více než je tento ukazatel a to v obou směrech. Koeficient šikmosti je 4,66. Rozdělení není tedy symetrické a jde o kladné zešikmení. Převažuje počet domácností s nižšími příjmy než domácností s příjmy vyššími. Tento koeficient je menší než tomu bylo v České republice. Koeficient špičatosti je 77,72. To znamená, že rozdělení je špičatější než normované normální rozdělení, avšak méně, než tomu bylo v případě ČR. Kvartilový podíl je větší než 1, což způsobuje větší hodnota horní kvartilové difference než dolní. Z toho vyplývá, že Galtonův koeficient šikmosti je kladný a má hodnotu 0,12. To potvrzuje kladné, pravostranné zešikmení příjmů domácností. Galtonova p-šikmost pro prostředních 25% hodnot je kladná a má hodnotu 0,02, pro prostředních 75% hodnot je pak 0,24. Všechny tyto tři koeficienty jsou menší, než v ČR. To znamená, že rozdělení na Slovensku je méně zešikmené. Naopak Moorsova špičatost 0,90 nám říká, že rozdělení bez vlivů konců je špičatější než v ČR. I tak můžeme toto rozdělení charakterizovat jako malou odlišnost špičatosti od normálního rozdělení. Stejně tak, jak tomu bylo v ČR, koeficienty šikmosti a špičatosti spočtené ze všech hodnot se výrazně liší od koeficientů spočtených bez odlehlých hodnot.



### 5.2.3 Modelování jednoduchými pravděpodobnostními funkcemi

U hledání jednoduchého tvaru rozdělení ze slovenských dat jsem postupoval stejným způsobem jako tomu bylo z dat ČR z důvodu odlehklých hodnot. Krabičkový diagram je uveden v příloze č. 5.

Tabulka č. 5 – Porovnání empirického rozdělení CP domácností s jednoduchými teoretickými tvary.

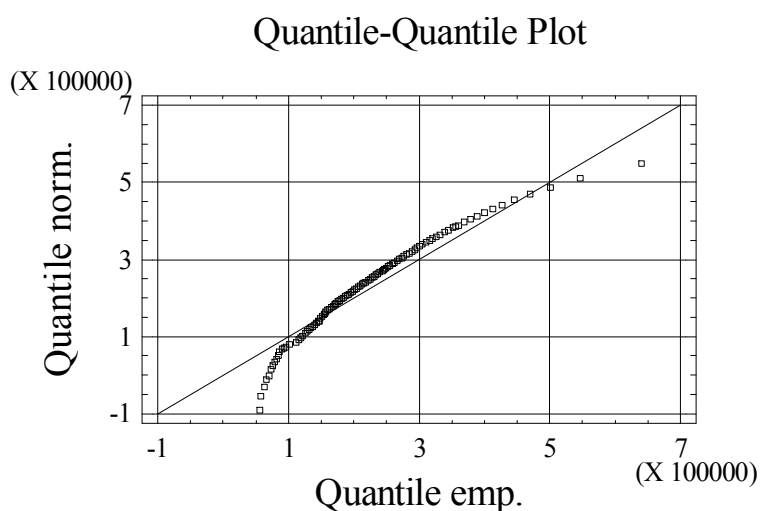
tvar rozdělení	empirické	Normální	Lognorm.	Weibullovo	Paretovo	Gama	Erlang
parametry odh. met. max. věrohodnosti	x	228285	229440	1,79	0,08	3,53	4
	x	137305	137430	257283	x	1,54E-05	1,75E-05
Chi-kvadrát	x	7536,04	2494,82	3213,76	x	1965,70	2630,48
kritická hodnota	x	94,81	94,81	93,92	x	93,92	93,92
p-hodnota	x	0	0	0	x	0	0
kvantily							
0,01	55572	-91133	54285	19749	1,13	40822	46984
0,02	58217	-53704	63129	29158	1,28	51459	57598
0,03	63277	-29957	69474	36666	1,45	59213	65920
0,04	66420	-12092	74664	43175	1,64	65596	72385
0,05	70200	-2439	72170	49042	1,87	71152	77978
0,25	143131	135675	135488	128369	33,34	139152	144694
0,5	207096	228285	196832	209693	4672,88	207129	209570
0,75	289035	320896	285950	308726	21835837	294521	291602
0,95	446402	454132	483364	474593	7,24E+15	457680	445512
0,96	469843	468663	518897	494007	1,1E+17	478297	461444
0,97	501484	486527	557659	518178	3,66E+18	504427	485406
0,98	547215	510275	613704	550805	5,14E+20	540515	518443
0,99	639988	547704	713691	603298	2,4E+24	600565	573288

Tak jak se dalo očekávat a jak ukazuje tabulka č. 5, podle  $\chi^2$  testu dobré shody ani jeden jednoduchý teoretický tvar nemodeluje dostatečně dobře celou délku rozdělení CP. Narozdíl od dat z ČR, nejnižší hodnotu tentokrát nemá lognormální rozdělení, jak by se dalo očekávat, ale je to rozdělení gama. Z tohoto důvodu se v následující analýze zaměřím opět na modelování pouze jednotlivých částí empirického rozdělení a pokusím se pro ně najít nejvhodnější model.

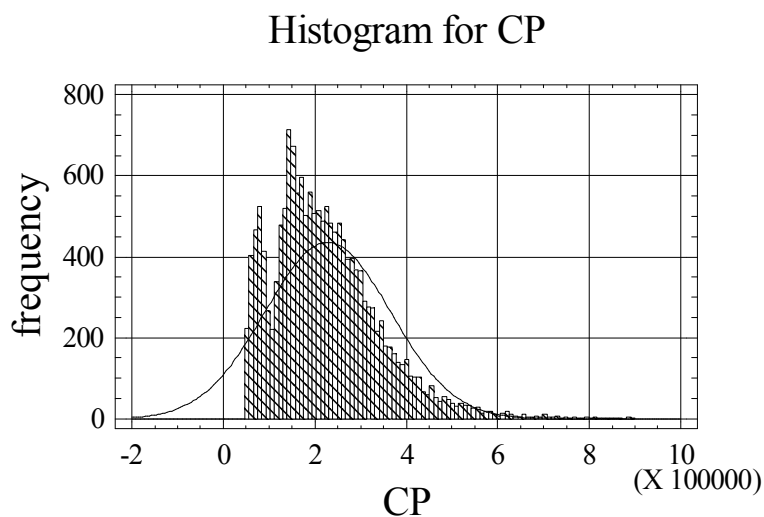
### 5.2.3.1 Normální rozdělení

Graf č. 21, který nám ukazuje porovnání kvantilů z empirických dat a kvantilů na základě normálního rozdělení potvrzuje závěry, které jsem provedl na základě  $\chi^2$  testu dobré shody. Nejen, že normální rozdělení nelze považovat za vhodné po celé jeho délce, ale na základě tohoto grafu je možné konstatovat, že není vhodné v žádném jeho úseku. V dolní a horní části totiž podhodnocuje skutečně naměřená data, v prostřední je naopak nadhodnocuje. Z histogramu (graf č. 22) je možné vyčíst, že v intervalu od 50 – 250 tis. Sk normální rozdělení podhodnocuje počet skutečných četností a v intervalu od 250 tis. je tomu naopak. Histogram opět pouze potvrdil výsledky, které je možno uskutečnit na základě předchozího grafu.

Graf č. 21

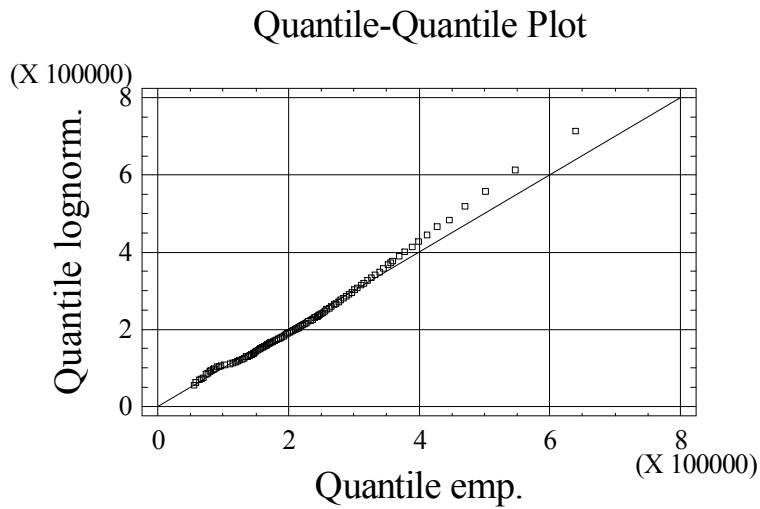


Graf č. 22

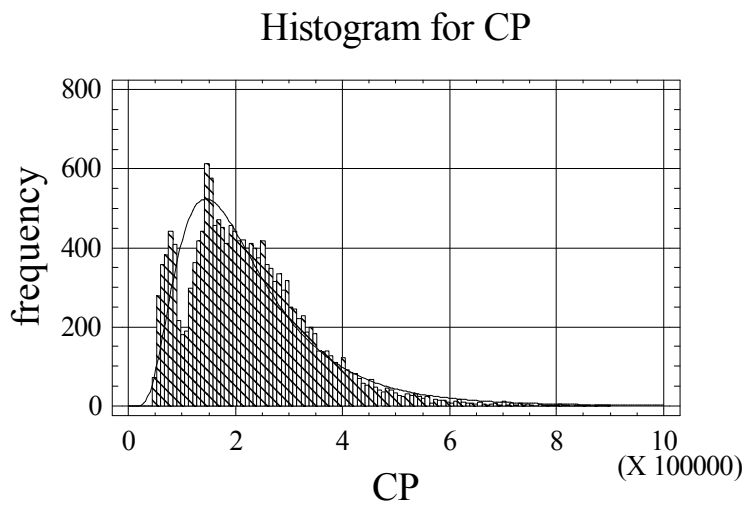


### 5.2.3.2 *Lognormální rozdělení*

Graf č. 23

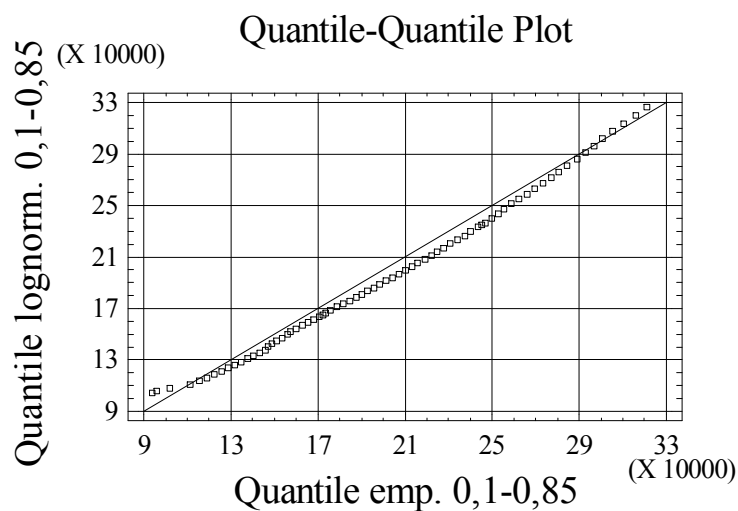


Graf č. 24



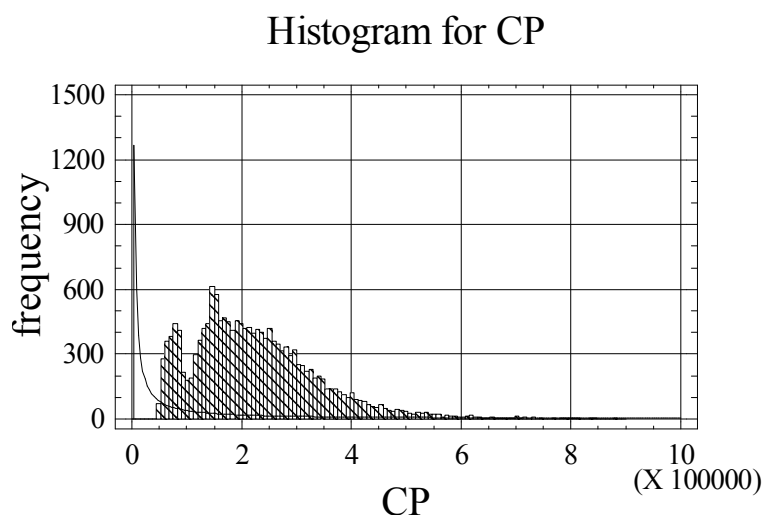
Graf č. 23 ukazuje, že lognormální rozdělení je určitě nevhodné v dolní a horní částí, kde skutečně naměřená data nadhodnocuje. Pro prostřední část jsem sestrojil detailnější graf (graf č. 25). Ten přinesl nečekané zjištění a to, že ani v prostřední části nemodeluje lognormální rozdělení dostatečně přesně. Skutečně naměřená data totiž podhodnocuje. O nevhodnosti tohoto rozdělení se můžeme opět přesvědčit na základě histogramu (graf č. 24).

Graf č. 25



### 5.2.3.3 *Paretovo rozdění*

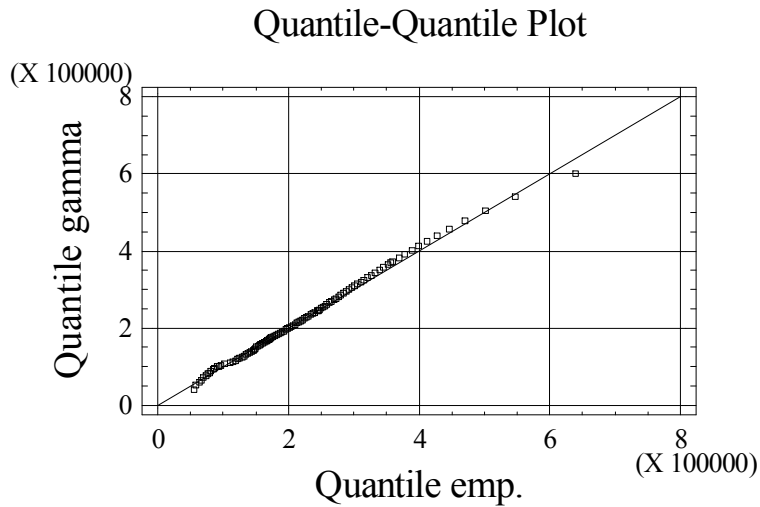
Graf č. 26



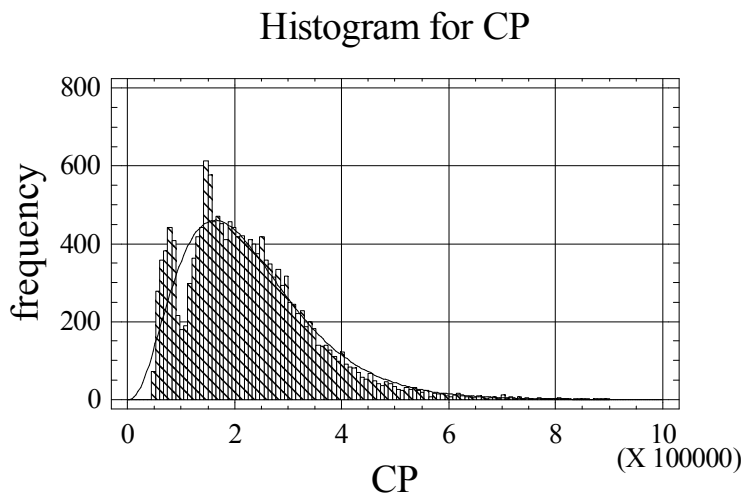
Ani v případě slovenských dat neuvádím porovnání empirických kvantilů a kvantilů teoretického Paretova rozdělení z důvodu velkého rozsahu dat a na první pohled zřejmě nevhodnosti tohoto rozdělení ve střední části. Tak jak tomu bylo u českých dat, tak ani tady není toto rozdělení vhodné ani v jeho krajních částech. To je možné vidět na histogramu (graf č. 26), nebo v tabulce č. 5.

### 5.2.3.4 *Gama rozdělení*

Graf č. 27



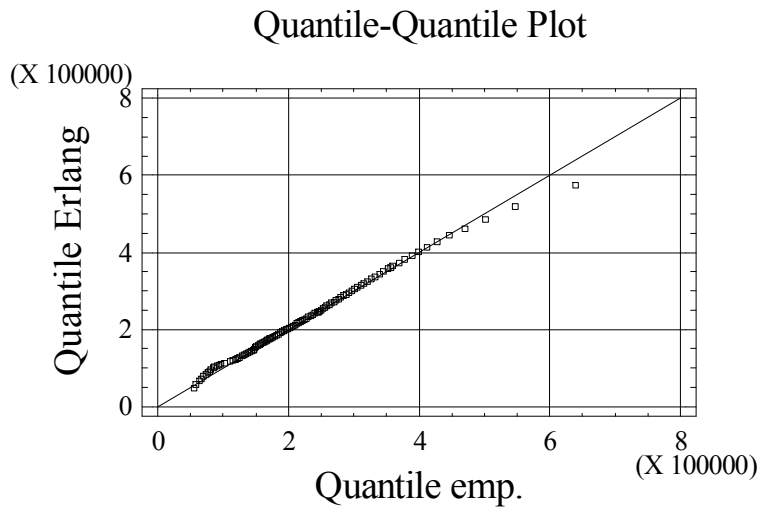
Graf č. 28



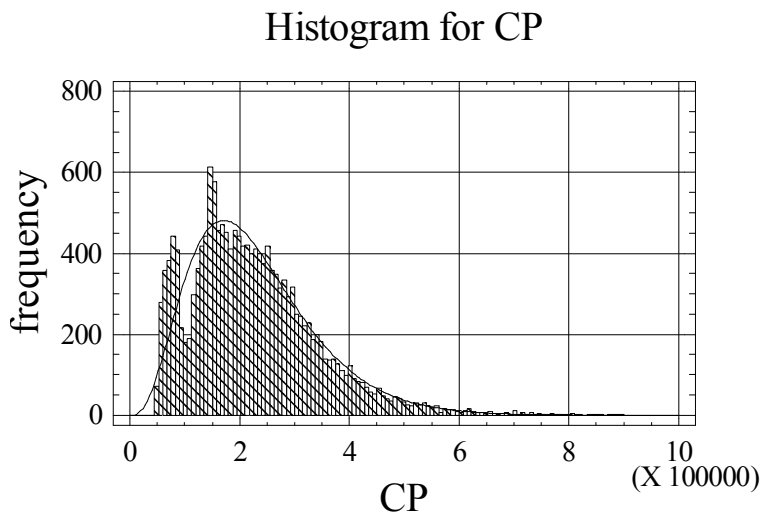
Jak jsem již zmínil výše, gama rozdělení má nejmenší hodnotu  $\chi^2$  testu dobré shody. Tato hodnota sice několikanásobně překračuje kritickou hodnotu pro přijetí tohoto rozdělení v celé jeho části, ale lze předpokládat, že v jeho středu by mohlo být vhodné. To potvrzuje graf č. 27. V dolním a horním konci se kvantily gama rozdělení neshodují s empirickými, ale ve střední části ano. Vhodnost gama rozdělení bych vymezil 15. až 75. percentilem. Poté se začínají objevovat větší rozdíly, které nejdříve data nadhodnocují a v úplném konci rozdělení podhodnocují. Nevhodnost v levém konci a vhodnost v prostřední části dokazuje i histogram (graf č. 28). Pravý konec zde není opět zobrazen z důvodu velkého rozsahu CP.

### 5.2.3.5 Erlangovo rozdělení

Graf č. 28



Graf č. 29



Pokud opět porovnáme grafy gama a Erlangovo rozdělení, najdeme mezi nimi pouze malé rozdíly. Opět můžeme konstatovat, že v levém ani v pravém konci není toto rozdělení možno považovat za vhodné. Po důkladnější analýze kvantilů je možno říci, že Erlangovo rozdělení je vhodné od 60 – 95 percentilu.

### 5.2.3.6 Shrnutí

Po výpočetní i grafické analýze je možné konstatovat, že ani jeden jednoduchý tvar nemodeluje dostatečně dobře celou část rozdělení. Na rozdíl od dat v ČR není možné považovat nejčastěji používaný a nejčastěji vyhovující lognormální model za vhodný v žádné části rozdělení. Také se nepodařilo najít shodu žádného jednoduchého tvaru v dolní a horní

části. Prostřední oblast je možné namodelovat gama rozdělením a to od 15 – 75 percentilu a Erlangovým rozdělením a to v intervalu od 60 – 95 percentilu.

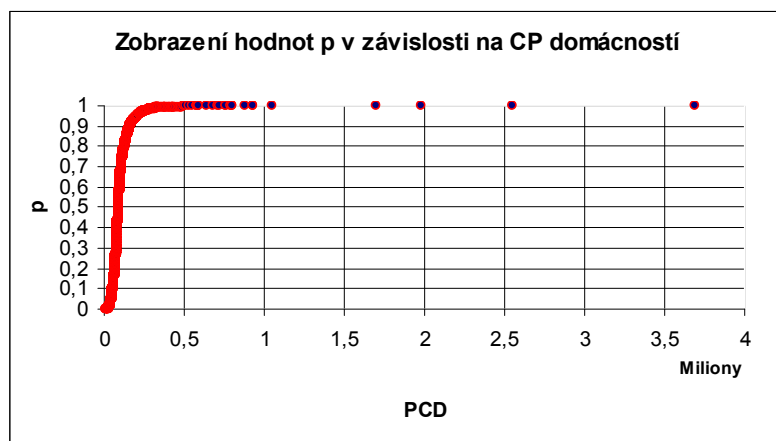
### 5.3 *Analýza a modelování příjmů na jednoho člena domácnosti v České republice za rok 2002*

Vstupními daty jsou údaje z Mikrocenzu ČR 2002. Příjem na jednoho člena domácnosti (dále PCD) jsem získal podílem celkových příjmů domácností a počtem členů dané rodiny.

#### 5.3.1 **Analýza a interpretace kvantilových grafů**

##### 5.3.1.1 *Distribuční funkce*

Graf č. 30

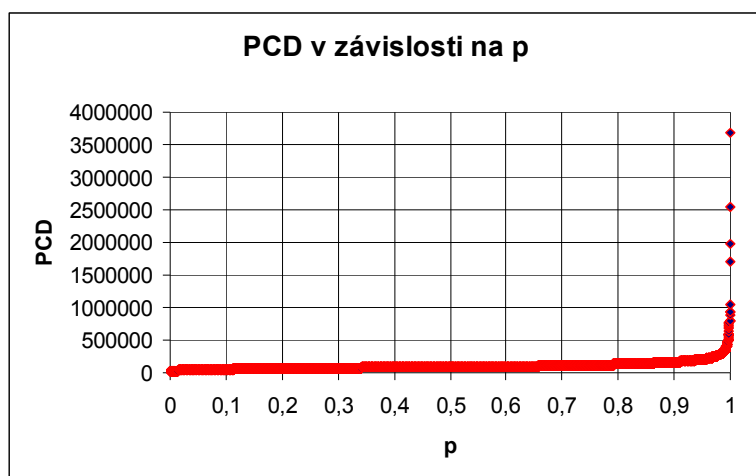


I po přepočtení příjmu na jednoho člena domácnosti má empirické rozdělení velmi krátký levý konec oproti konci pravému. Po přepočtení na jednoho člena už není na křivce patrný „zub“, který byl patrný při zobrazení CP jak v České, tak Slovenské republice. Proto jsem tentokrát detailnější grafy prvního a posledního kvantilu uveřejnil pouze v příloze č. 6.

Největší hodnota směrnice je mezi 40 – 270 tis. Kč. V tomto intervalu jsou tedy příjmy nejpravděpodobnější.

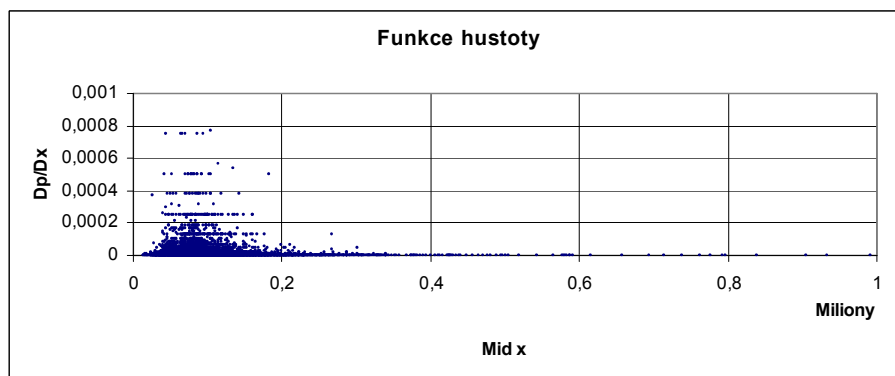
### 5.3.1.2 *Kvantilová funkce*

Graf č. 31



### 5.3.1.3 *Funkce hustoty*

Graf č. 32

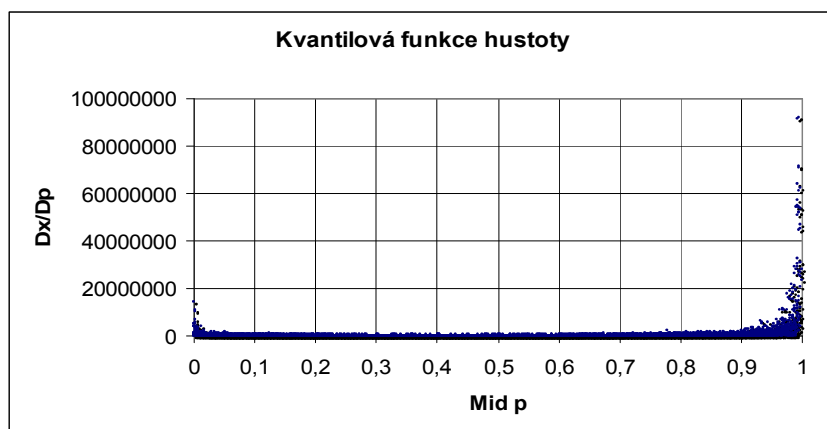


Rychlost změny  $Dp$  je nevyšší v intervalu od 40 – 170 tis. Kč. Většina relativních přírůstků je do 0,02% příjmu na člena domácnosti. Větší přírůstky jsou v dolní a střední části rozdělení. Pro lepší zobrazení jsem omezil výběr Mid x do 1 milionu Kč.



### 5.3.1.4 *Kvantilová funkce hustoty*

Graf č. 33



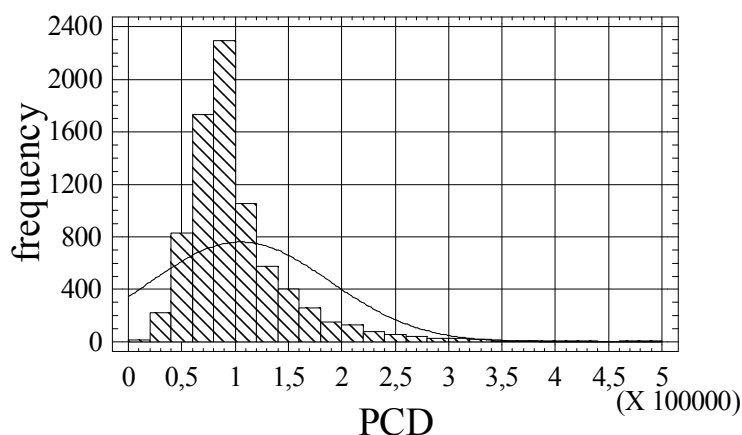
Kvantilová funkce hustoty nám ukazuje o dost vyšší relativní přírůstky v pravém konci rozdělení oproti konci levému, což je v případě příjmů logické. Detailnější zobrazení ukázalo vyšší intenzitu ve středu intervalu oproti oběma koncům. Pro lepší znázornění jsem omezil výběr na ose Y do 100 000 000.

### 5.3.1.5 *Ostatní grafické analýzy*

Histogram potvrzuje velmi krátkou levou část rozdělení a plynulé ubývání četností v pravé části rozdělení. Nejčetnější příjem na jednoho člena domácnosti je v intervalu 80 -100 tis Kč, druhým nejčetnějším intervalem pak je 60 -80 tis. Kč. Podle doplňující křivky normálního rozdělení bych se opět tak jako tomu bylo u příjmů domácností přiklonil k levostrannému zešikmění. PCD jsem omezil do 500 tis. Kč.

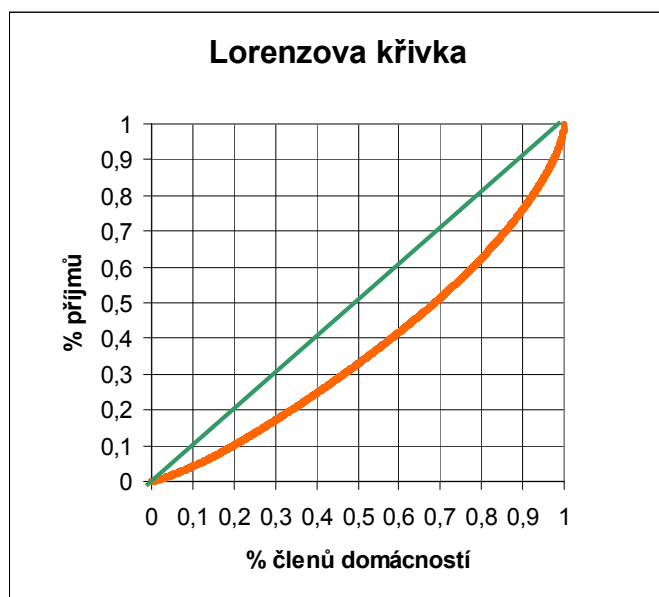
Graf č. 34

Histogram for PCD



Pokud porovnáme Lorenzovu křivku CP a PCD zjistíme, že u příjmů na člena domácnosti došlo k přiblížení k diagonále a tedy ke zmenšení nerovnosti. Např. 10% členů domácností s nejnižšími příjmy se podílí na celkových příjmech ze 4,15%, zatímco u CP domácností to bylo pouze ze 3%. Na druhé straně 10% členů s nejvyššími výdělky inkasuje 24% celkových příjmů.

Graf č. 35



### 5.3.2 Statistické ukazatele

Tabulka č. 6

<b>Průměr</b>	105030,0	<b>Horní kvartilová dif.</b>	26762,6
<b>Medián</b>	89204,0	<b>Dolní kvartilová dif.</b>	15746,0
<b>Modus</b>	84000,0	<b>Galtonův koef. šikmosti</b>	0,26
<b>Rozptyl</b>	6,98865E9	<b>Moorsova špičatost</b>	0,97
<b>Směrodatná odchylka</b>	83598,2	<b>Galtonova p šikmost</b>	
<b>Minimum</b>	13500,0	<b>p= 0,375</b>	0,04
<b>Maximum</b>	3,69091E6	<b>p= 0,125</b>	0,77
<b>Koeficient šikmosti</b>	17,1419		
<b>Koeficient špičatosti</b>	574,227		

Kromě standardních charakteristik vypočítaných v programu Statgraphics Plus jsem vypočítal i méně známé kvantilové charakteristiky v programu Microsoft Excel. Z důvodu obsáhlosti těchto charakteristik uvádím jen některé jejich hodnoty, ostatní jsou uveřejněny v příloze č. 7.

### **5.3.2.1**            *Interpretace charakteristik*

Nejnižší zjištěný příjem jednotlivých členů domácností byl 13 500 Kč, nejvyšší 3 690 914 Kč. Variační rozpětí je v tomto případě tedy 3 677 414 Kč. Průměrný příjem jednotlivých členů byl 105 030 Kč. Tuto a vyšší částku však pobíralo pouze 32% jednotlivých členů. Medián, čili příjem, který pobíralo 50% jednotlivých členů, činí 89 204 Kč. Medián je tedy oproti průměru nižší o 15 826 Kč. Hodnota prvního decilu nám říká, že 10% domácností má nižší příjem, než 54 355Kč, naopak poslední decil poukazuje na to, že 10% členů s nejvyššími příjmy pobírá více než 163 317 Kč ročně. Většina domácností se neodchýlila od průměru o více než 83 598 Kč, což je směrodatná odchylka. Koeficient šikmosti je 17,14. Jedná se tedy o kladné zešikmení a převažují tedy členové domácností s nižšími příjmy. Koeficient šikmosti poukazuje na výrazně špičatější rozdělení v porovnání s normovaným normálním. Horní kvartilová diference je větší než dolní kvartilová diference a kvartilový podíl je tedy větší než 1. Proto je Galtonův koeficient šikmosti pro 50% prostředních hodnot kladný. To potvrzuje kladné, pravostranné zešikmení příjmů jednotlivých členů domácností. Galtonova p-šikmost pro prostředních 25% hodnot má hodnotu 0,04 a pro prostředních 75% hodnot 0,77. Moorsova špičatost je 0,97, což lze ještě charakterizovat jako malou odlišnost špičatosti vrcholu empirického rozdělení v porovnání s normálním rozdělením. Opět jsou patrné výrazné rozdíly mezi koeficienty šikmosti a špičatosti vypočítané v programu Statgraphics a kvantilovými ukazateli vypočítanými ručně v Excelu, které neberou v úvahu odlehlé hodnoty.

### 5.3.3 Modelování jednoduchými pravděpodobnostními funkcemi

Krabičkový diagram je uveden v příloze č. 8.

Tabulka č. 7 – Porovnání empirického rozdělení PCD s jednoduchými teoretickými tvary.

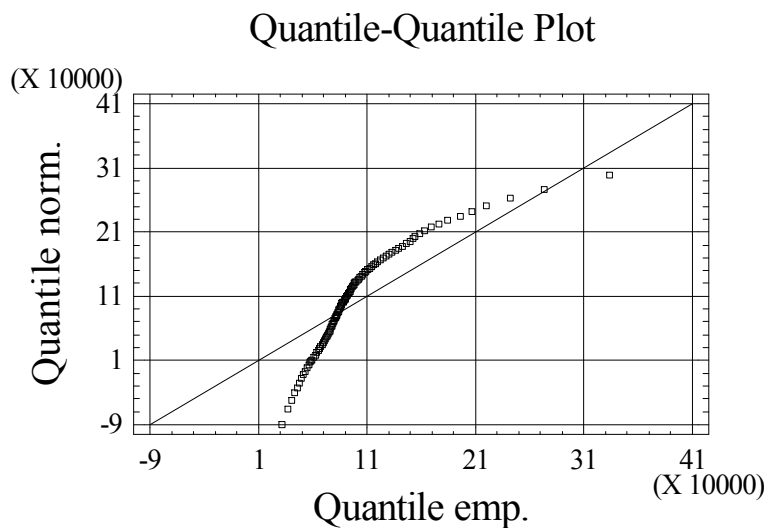
tvar rozdělení	empirické	Normální	Lognormální	Weibullovo	Paretovo	Gama	Erlang
parametry odh. met. max věrohodnosti	x	105030	103176	1,58139	0,08742	4,22467	4
	x	83598,2	49921,4	117588	x	4,022E-05	3,808E-05
Chi-kvadrát	x	9729,43	1842,28	5679,06	x	3073,37	3075,18
kritická hodnota	x	69,13	69,13	67,37	x	67,37	67,37
p-hodnota	x	0	0	0	x	0	0
kvantily							
0,01	31185	-89449	31955	6413	1,12	22884	21617
0,02	37094	-66660	36210	9972	1,26	28012	26684
0,03	40122	-52201	39200	12928	1,42	31681	30328
0,04	43205	-41325	41610	15558	1,6	34665	33303
0,05	45757	-32477	43679	17975	1,8	37239	35876
0,25	73458	48644	68164	53482	26,86	67640	66571
0,5	89204	105030	92876	93263	2776,43	96869	96419
0,75	115967	161416	126547	144566	7708580	133577	134161
0,95	206672	242537	197484	235333	7,63E+14	200689	203591
0,96	220090	251384	207306	246270	9,8E+15	209083	212302
0,97	242349	262261	220053	259969	2,63E+17	219701	223326
0,98	273215	276720	238219	278594	2,72E+19	234329	238526
0,99	334113	299509	269944	308866	7,55E+22	258589	263759

Tak jako u domácností, tak ani u jednotlivých jejich členů ani jeden teoretický tvar

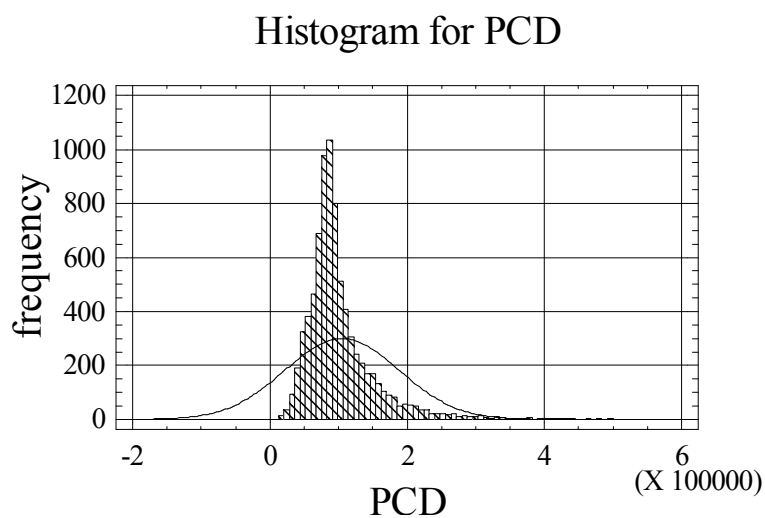
nemodeluje dostatečně dobře celé rozdělení příjmů. To ukazují výsledky  $\chi^2$  testu dobré shody. Navíc jsou všechny hodnoty tohoto testu výrazně vyšší, než tomu bylo u celkových příjmů domácností. Nejlépe vychází lognormální rozdělení, ale i to překračuje přístupnou kritickou hodnotu více než 20 krát. Uvidíme, jaké výsledky přinesou grafické analýzy srovnání empirických a teoretických kvantilů, a jestli bude možné přijmout některý z modelů alespoň v některé části rozdělení.

### 5.3.3.1 Normální rozdělení

Graf č. 35



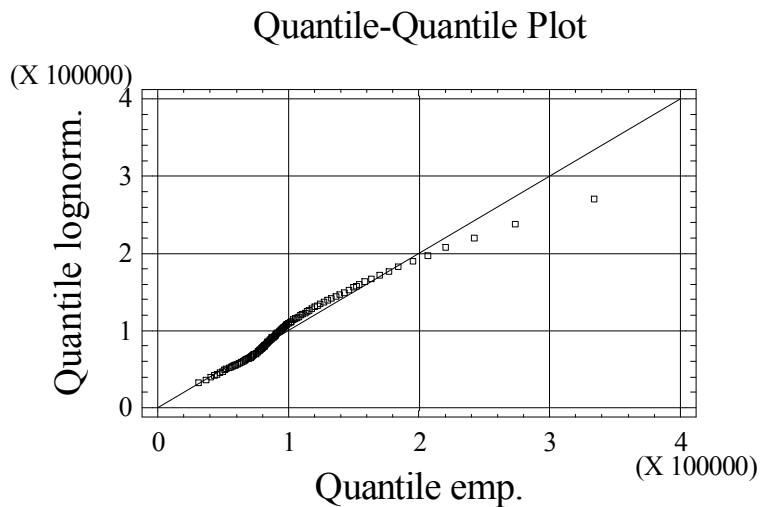
Graf č. 36



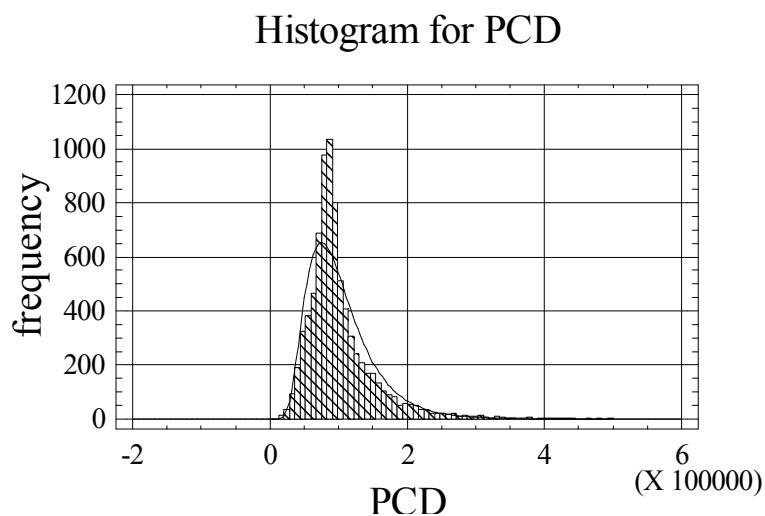
Z těchto obou grafů a také z předchozí analýzy  $\chi^2$  testu dobré shody je možné konstatovat, že normální rozdělení je nevhodné pro rozdělení PCD v České Republice. Dokonce v žádné jeho části není možné najít dostatečnou shodu. V dolní části, tak jak tomu v empirickém rozdělení pouze s kladnými hodnotami bývá, normální rozdělení zjištěná data podhodnocuje. Ve střední části a horní části až do 98 percentilu je tomu naopak, tedy teoretická data normálního rozdělení jsou výrazně nadhodnocena.

### 5.3.3.2 Lognormální rozdělení

Graf č. 37



Graf č. 38

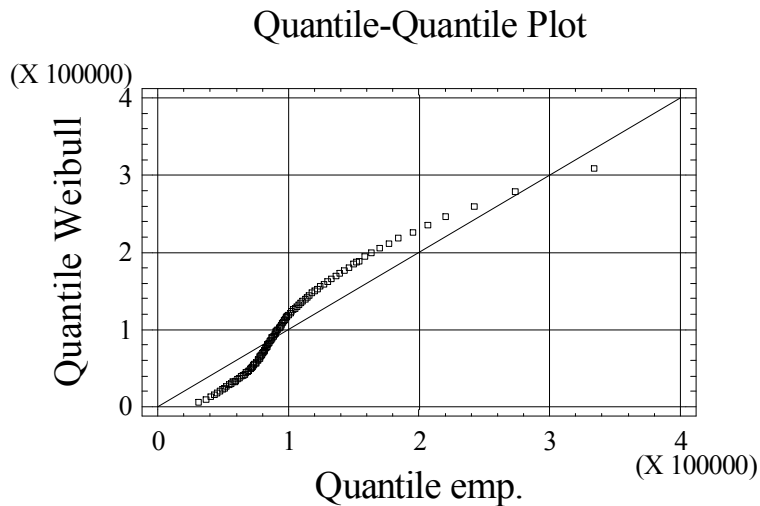


Pokud porovnáme graf č. 9 a graf č. 37, je na první pohled vidět, že lognormální rozdělení více odpovídá celkovým příjmům domácností, než příjmům jednotlivých členů. Ostatně to také říká výsledek  $\chi^2$  testu dobré shody, který je u jednotlivých členů domácností dvojnásobný. Pokud se podíváme na tabulku č. 7, můžeme říct, že ze všech rozdělení by právě lognormální mělo nejvíce odpovídat empirickým datům, a to hlavně v dolní a střední části rozdělení. Jelikož to ale není patrné z grafu č. 37, sestrojil jsem detailnější grafy, které je možné prohlédnout v příloze č. 9. Tyto grafy však přinesly zjištění, že lognormální rozdělení

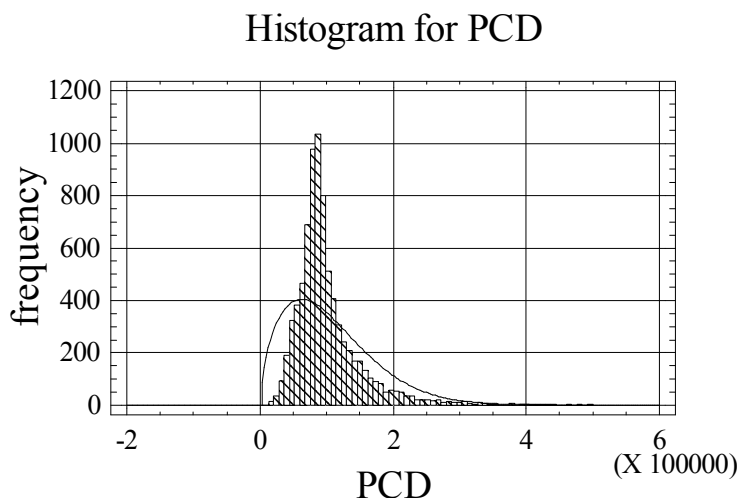
je možné považovat za vhodné pouze ve velmi krátké dolní části rozdělení. Poté už nelze najít dostatečně velkou shodu s empirickými daty.

### 5.3.3.3 Weibullovo rozdělení

Graf č. 39



Graf č. 40



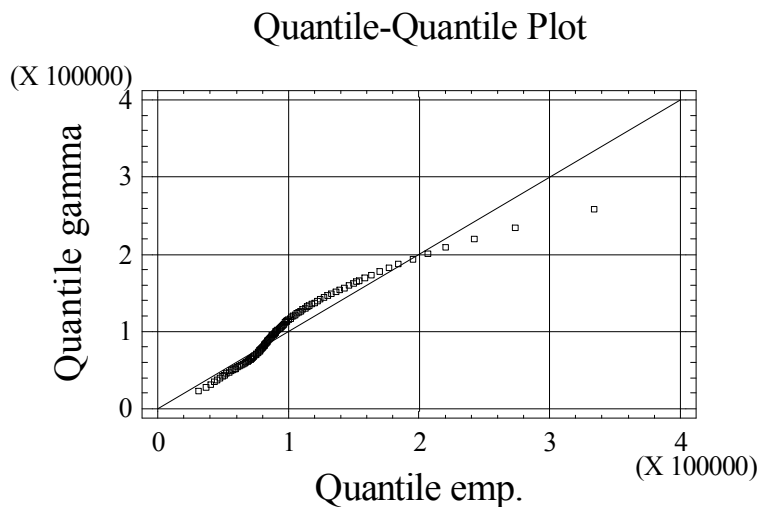
Velmi vysoká a po normálním rozdělení nejvyšší hodnota  $\chi^2$  testu dobré shody již napovídá o nevhodnosti Weibullova rozdělení na data PCD v České republice. Dodatkové grafické analýzy tento výsledek opět potvrdily a navíc ukázaly, že není možné toto teoretické rozdělení přijmout v žádné jeho části. V dolní části a v samotném konci Weibullova rozdělení podhodnocuje skutečně naměřená data, zatímco ve střední a v horní části do 99. percentilu je tomu naopak.

### 5.3.3.4 *Pareto rozdělení*

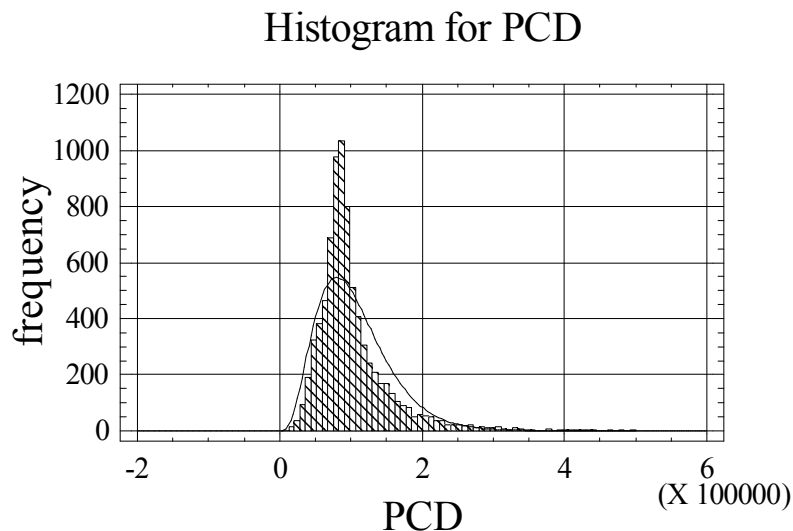
Pokud porovnáme jednotlivé kvantily empirického a Paretova rozdělení v tabulce č. 7, je možné vidět na první pohled výrazné rozdíly, a proto není potřeba sestavit dodatečné grafické analýzy. Nevhodnost v celé části je možné konstatovat i bez nich.

### 5.3.3.5 *Gama rozdělení*

Graf č. 41



Graf č. 42



Hodnota  $\chi^2$  testu dobré shody překračuje skoro třicetkrát přístupnou kritickou hodnotu. Tato hodnota je v porovnání s  $\chi^2$  testem lognormálního rozdělení skoro dvojnásobná. Proto se již dopředu dalo očekávat, že i gama rozdělení nebude příliš vhodné. To také ukázaly grafické

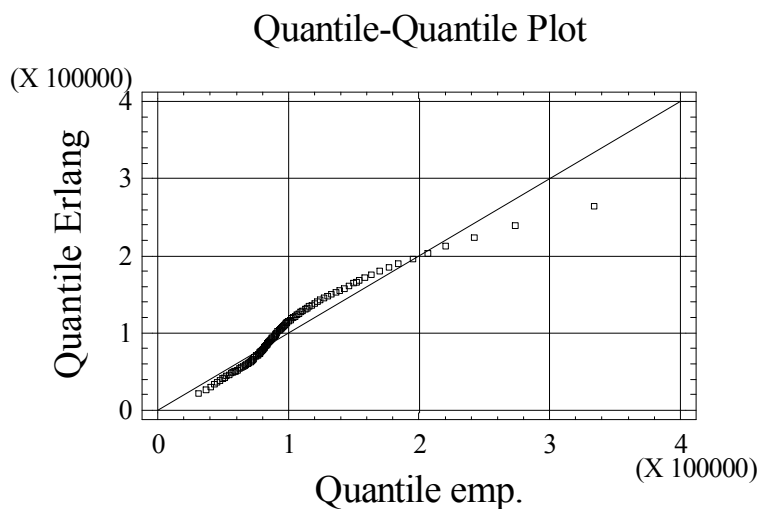


analýzy, kde z grafů č. 41 a 42 je patrné, že ve střední a horní části není gama rozdělení možné považovat za přípustné. Pro dolní část bylo potřeba udělat detailnější analýzu, která je uvedena v příloze č. 10. Tato analýza přinesla zjištění, že ani v dolní části není možné považovat gama rozdělení za vhodné.

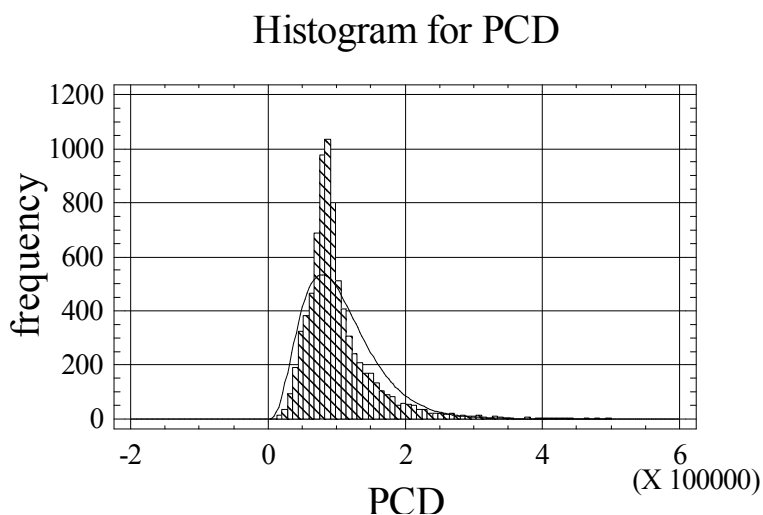
### 5.3.3.6 Erlangovo rozdělení

Tak jak se dalo očekávat, závěry z grafických analýz Erlangova rozdělení jsou totožné ze závěry rozdělení gama. V žádné části nelze najít dostatečnou shodu empirických a teoretických hodnot.

Graf č. 43



Graf č. 44



### **5.3.3.7            *Shrnutí***

Tak jako u obou analýz CP domácností, tak ani u analýzy PCD v České republice není možné přijmout ani jeden jednoduchý v celé části rozdělení. Výsledky grafických analýz u PCD však dopadly daleko hůře než u CP domácností. V horní a dokonce ani ve střední části rozdělení nebyl vhodný ani jeden jednoduchý teoretický model. Pouze ve velmi krátké spodní části by bylo možné za vhodné považovat lognormální rozdělení.

## **5.4            *Analýza a modelování příjmů na jednoho člena domácnosti ve Slovenské republice za rok 2003***

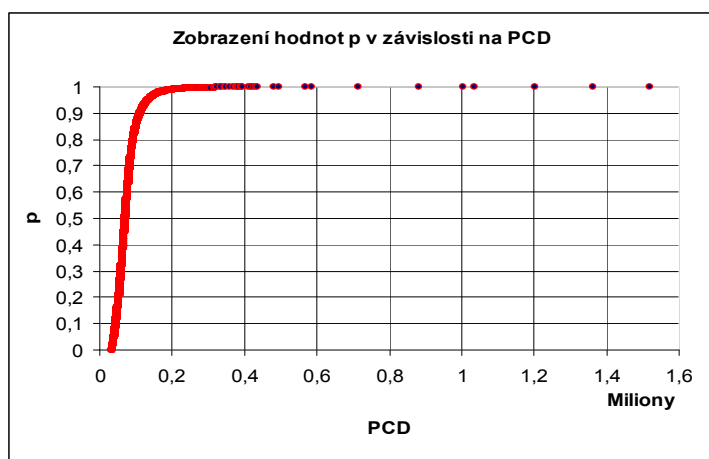
Vstupními daty jsou údaje z Mikrocenzu SR 2003. Příjem na jednoho člena domácnosti (dále PCD) jsem získal podílem celkových příjmů domácností a počtem členů dané rodiny.

### **5.4.1            *Analýza a interpretace kvantilových grafů***

#### **5.4.1.1            *Distribuční funkce***

Stejně jako tomu bylo ve třech předchozích analýzách, tak i u PCD ve SR se dal očekávat krátký dolní konec rozdělení oproti konci pravému. Dá se říci, že při zkoumání všech příjmů, a tudíž nevytváření specifických skupin, je to pro empirické rozdělení charakteristické. V porovnání s rozdělením PCD v ČR je však pravý konec o dost kratší. Nejpravděpodobnější hodnoty jsou v intervalu 40 – 150 tis. Sk. Poté začíná hodnota směrnice klesat. Jak jsem již dříve psal, nemůžeme srovnávat absolutní hodnoty jednotlivých šetření z důvodu různé měny, ale v CP domácností si byly částky velmi podobné, zatímco u PCD jsou výrazně odlišné. Proto jsem vypočítal průměrný počet členů domácností v ČR a SR. V české domácnosti je průměrně 2,36 členů, zatímco ve slovenské 3,15. To je velmi vysoký a překvapující rozdíl.

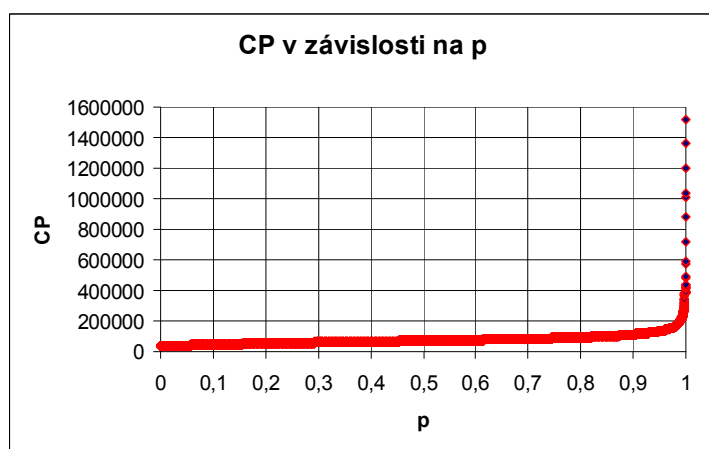
Graf č. 45



Podrobnější analýzy jednotlivých kvartilů tentokrát neodhalily žádné významné skutečnosti a proto je neuvádím.

#### 5.4.1.2 *Kvantilová funkce*

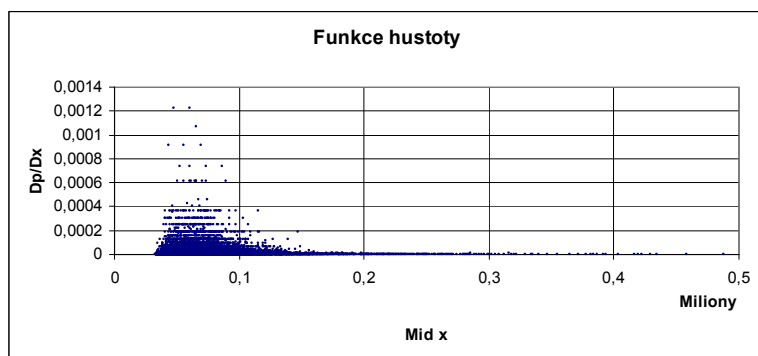
Graf č. 46



#### 5.4.1.3 *Funkce hustoty*

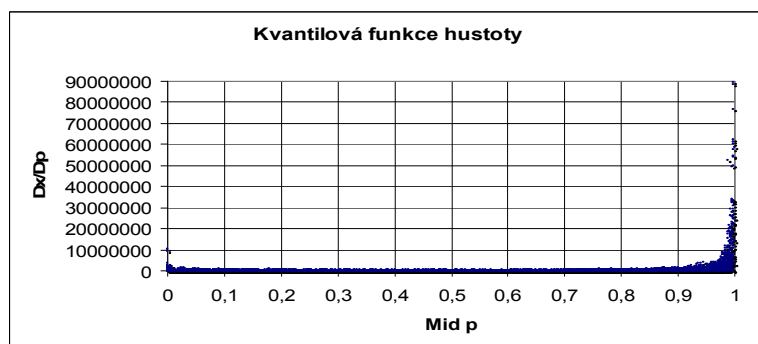
Rychlost změny  $D_p$  je nejvyšší v intervalu od 40 – 140 tis. Sk. Většina relativních přírůstků je do 0,02 % PCD. Ojedinele se objevují hodnoty, které dosahují 0,13%. Tak jak tomu bylo ve všech předchozích případech, tak i zde jsou největší přírůstky v dolní a střední části rozdělení. Funkce hustoty PCD ve SR je velmi podobná funkci PCD v ČR.

Graf č. 47



#### 5.4.1.4 *Kvantilová funkce hustoty*

Graf č. 48

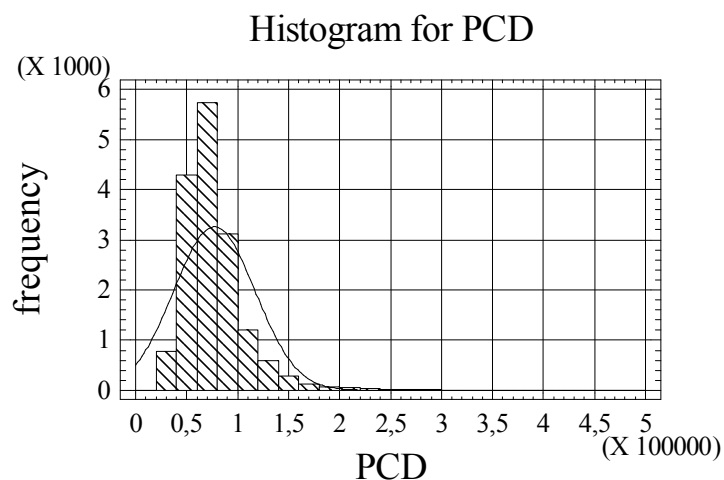


Samozřejmě, že podobnost s PCD v ČR lze najít i u kvantilové funkce hustoty. Opět, tak jako tomu bylo ve všech případech, vyšší relativní přírůstky jsou v obou koncích rozdělení, zatímco ve střední části jsou menší a s vyšší intenzitou. Avšak výrazně vyšší jsou v pravém konci. Proto je možné považovat rozdělení za asymetrické s prodlouženým pravým koncem. Na ose Y bylo potřeba omezit výběr do 90 000 000.

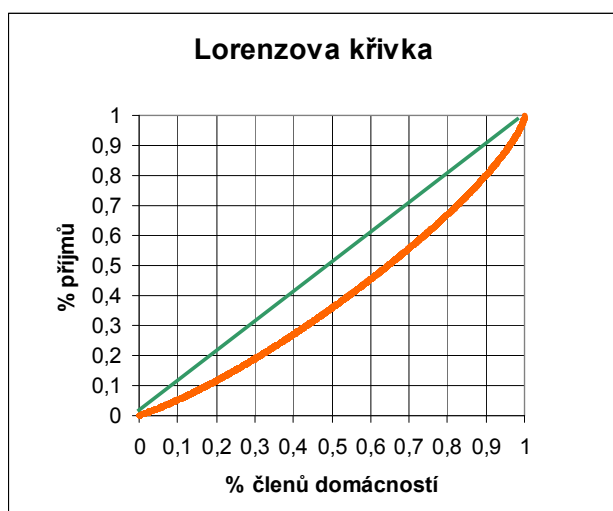
#### 5.4.1.5 *Ostatní grafické analýzy*

Histogram pro PCD potvrdil dřívější závěry z grafických analýz o „useknuté“ levé části rozdělení. Nejvíce příjmů jednotlivých členů domácnosti je v intervalu 60 – 80 tis. Sk, druhým nejčetnějším intervalem je 40 – 60 tis. Sk. Podle doplňující křivky normálního rozdělení se dá opět předpokládat levostranné zešíkmení. To však s jistotou budeme moci říci až po výpočetní analýze.

Graf č. 49



Graf č. 50



Tak jako tomu bylo u českých dat, tak i slovenských došlo v případě PCD k přiblížení skutečné Lorenzovy křivky k diagonále oproti CP domácností. To v realitě znamená nižší diferenciaci příjmů. Např. 10% členů domácností s nejnižšími příjmy se podílí na příjmech celkových z 4,54%. Na opačném konci, 10% s nejvyššími příjmy si rozdělí 20% celkových příjmů. V ČR to bylo 4,15 a 24%. Proto můžeme i u jednotlivých členů potvrdit nižší diferenciaci, než která je v České republice.

## 5.4.2 Statistické ukazatele

Tabulka č. 8

<b>Průměr</b>	77115,0	<b>Horní kvartilová dif.</b>	15959,4
<b>Medián</b>	70920,0	<b>Dolní kvartilová dif.</b>	14510,4
<b>Modus</b>	55572,0	<b>Galtonův koef. šikmosti</b>	0,05
<b>Rozptyl</b>	1,60619E9	<b>Moorsova špičatost</b>	0,86
<b>Směrodatná odchylka</b>	40077,3	<b>Galtonova p šikmost</b>	
<b>Minimum</b>	32355,0	<b>p= 0,375</b>	-0,003
<b>Maximum</b>	1,5187E6	<b>p= 0,125</b>	0,35
<b>Koeficient šikmosti</b>	10,460		
<b>Koeficient špičatosti</b>	266,202		

Kromě standardních charakteristik vypočítaných v programu Statgraphics Plus jsem vypočítal i méně známé kvantilové charakteristiky v programu Microsoft Excel. Z důvodu obsáhlosti těchto charakteristik uvádím jen některé jejich hodnoty, ostatní jsou uveřejněny v příloze č. 11.

### 5.4.2.1 Interpretace charakteristik

Nejnižší zjištěný příjem jednotlivých členů domácností v SR je 32 355 Sk, nejvyšší pak 1 518 705 Sk. Variační rozpětí je v tomto případě 1 486 350 Sk. Průměrný příjem na člena domácnosti byl 77 115 Sk. Na průměr dosáhlo 47,6% jednotlivých členů. To je u měření příjmů neobvykle vysoké číslo, které není příliš vzdálené od mediánu, který byl 70 920 Sk, čili pouze o 6 795 Sk nižší. Na průměr zde tedy dosáhlo o 15% více lidí, než tomu bylo v České republice. První decil má hodnotu 45 014 Sk, poslední pak 111 980 Sk. Směrodatná odchylka je 40 077. Většina lidí se neodchýlila od průměru o tuto hodnotu v obou směrech. Koeficient šikmosti je 10,46. Rozdělení je tedy asymetrické a jde o kladné zešikmení. Převažují tedy členové s menšími příjmy oproti členům s vyššími příjmy. Jde však o méně zešikmené rozdělení, než v ČR, kde byl tento koeficient roven 17. Rozdělení je výrazně špičatější než normované normální rozdělení. To je možné konstatovat z pohledu na koeficient špičatosti, který je 266,2. Opět však méně špičatější než v ČR. Horní kvartilová

diference ( $u_{qd} = 15\,959$ ) je vyšší než dolní kvartilová diference ( $l_{qd} = 14\,510$ ), což zapříčiní kvartilový podíl větší než 1. Proto je Galtonův koeficient šikmosti kladný. Jeho hodnota je 0,05. To je velmi nízká hodnota blížící nule. Proto bychom mohli říct, že pro prostředních 50% hodnot je rozdělení skoro symetrické. Galtonova p-šikmost pro prostředních 25% hodnot ( $p=0,375$ ) je záporná a má hodnotu ( $g(p)=-0,003$ ) Galtonova p-šikmost pro prostředních 75% hodnot ( $p=0,25$ ) má hodnotu ( $g(p)=0,35$ ). Moorsova špičatost je 0,86. Podle tohoto ukazatele můžeme říct, že jde o malou odlišnost špičatosti vrcholu empirického rozdělení v porovnání se špičatostí normálního rozdělení. V porovnání s ČR jsou všechny tyto kvantilové ukazatele nižší.

### 5.4.3 Modelování jednoduchými pravděpodobnostními funkcemi

Krabičkový diagram je uveden v příloze č. 12.

Tabulka č. 7 – Porovnání empirického rozdělení PCD s jednoduchými teoretickými tvary.

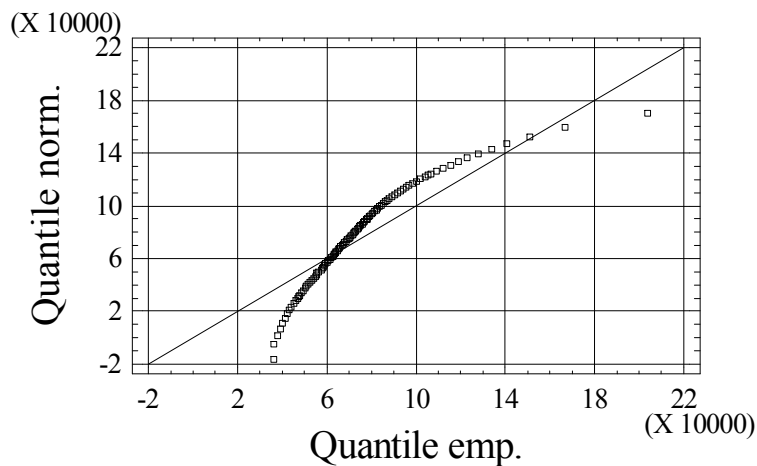
tvar rozdělení	empirické	Normální	Lognormální	Weibullovo	Paretovo	Gama	Erlang
parametry odh. met. max věrohodnosti	x	77115	76449	1,95943	0,08946	6,83824	7
	x	40077,3	28768,3	86317,3	x	8,86759E-05	9,077E-05
Chi-kvadrát	x	18735,53	5946,46	13571,92	x	9229,82	9321,94
kritická hodnota	x	94,81	94,81	93,92	x	93,92	93,92
p-hodnota	x	0	0	0	x	0	0
kvantily							
0,01	36072	-16119	30686	8251	1,12	25258	25671
0,02	36304	-5194	33886	11783	1,25	29160	29569
0,03	38139	1738	36088	14530	1,41	31851	32254
0,04	39121	6952	37837	16872	1,58	33989	34386
0,05	40187	11194	39323	18957	1,77	35803	36193
0,25	56410	50083	55977	45704	24,92	55734	55993
0,5	70920	77115	71551	71592	2317,18	73390	73476
0,75	86879	104147	91457	101975	5369329	94450	94284
0,95	133833	143036	130191	151106	3,49E+14	131142	130461
0,96	140724	147278	135303	156749	4,23E+15	135619	134871
0,97	150751	152492	141863	163749	1,05E+17	141256	140422
0,98	166780	159424	151079	173154	9,8E+18	148976	148021
0,99	204000	170349	166836	188186	2,27E+22	161673	160516

Podle předchozích studií porovnávání empirického rozdělení příjmů s jednoduchými teoretickými tvary se dalo předpokládat, že ani jeden teoretický tvar nebude s dostatečnou kvalitou modelovat celou část rozdělení PCD ve SR. Pokud se podíváme na výsledky  $\chi^2$  testu dobré shody v tabulce č. 7, vidíme, že všechny hodnoty mnohonásobně překračují kritické hodnoty. Tyto hodnoty jsou nejvyšší ze všech předchozích studií. Nejnižší hodnotu  $\chi^2$  testu má lognormální rozdělení, nejvyšší pak normální. Uvidíme, jaké výsledky přinese grafická analýza srovnání empirických a teoretických kvantilů v další části práce.

### 5.4.3.1 *Normální rozdělení*

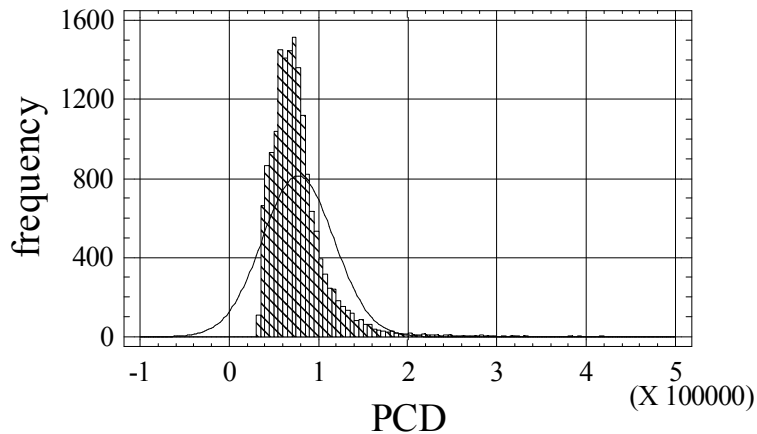
Graf č. 51

Quantile-Quantile Plot



Graf č. 52

Histogram for PCD



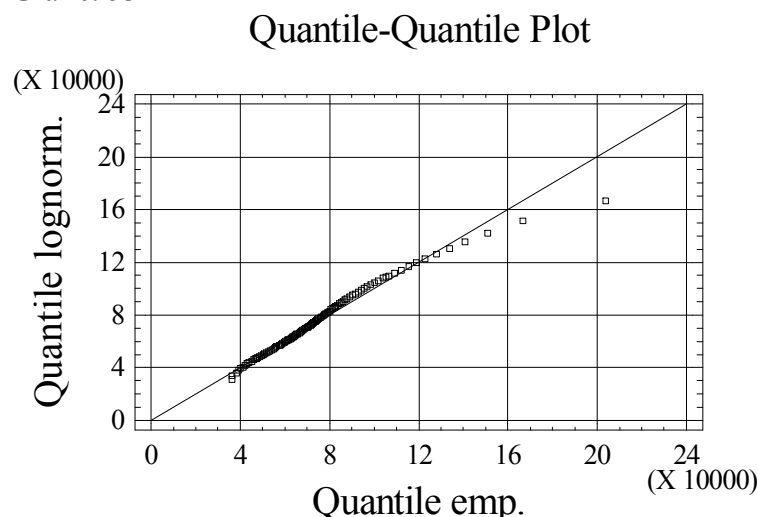


Obě tyto grafické analýzy potvrdily předchozí závěry o nevhodnosti normálního rozdělení v celé jeho části. Pokud se podíváme na srovnání empirických a teoretických kvantilů v grafu č. 51, je možné vidět, že v dolní a horní části normální rozdělení podhodnocuje skutečně naměřené příjmy, zatímco ve střední části je tomu naopak. Histogram zase ukazuje, že v dolní a horní části je skutečných četností daleko méně než teoretických četností normálního rozdělení, ve střední části je tomu naopak.

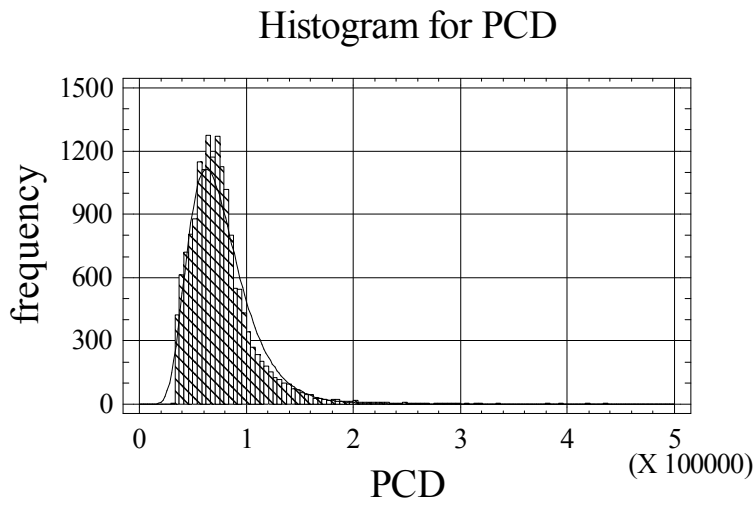
### 5.4.3.2 *Lognormální rozdělení*

Grafické analýzy sice potvrdily, že lognormální rozdělení není vhodné v celé délce rozdělení, ale určitě bude možné hledat shodu alespoň v některé jeho části. Pokud se podíváme na histogram, je vidět, že empirické četnosti se většinou shodují s četnostmi lognormálního rozdělení. To na první pohled vyvrací vysokou hodnotu  $\chi^2$  testu dobré shody. Je třeba však upozornit, že program Statgraphics, ze kterého pochází graf č. 54 používá jinou metodu pro rozdělení tříd, než jsem použil já při výpočtu  $\chi^2$  testu v programu Microsoft Excel. Tato metoda byla vysvětlena v kapitole 4.2.2. Porovnání empirických a teoretických kvantilů v grafu č. 53 ukazuje nevhodnost lognormálního rozdělení v horní části rozdělení. Pro dolní a střední část nelze podle tohoto výstupu rozhodnout, a proto jsem se rozhodl vytvořit detailnější grafy, které jsou uveřejněny v příloze č. 13. Z těchto grafů můžeme usuzovat, že lognormální rozdělení je pro PCD ve SR vhodné od 5. percentilu po medián. V krajních dolních a horních 5-ti percentilech teoretické rozdělení podhodnocuje skutečně naměřená data. Naopak od mediánu po 95. percentil nadhodnocuje skutečně vyšetřené hodnoty.

Graf č. 53

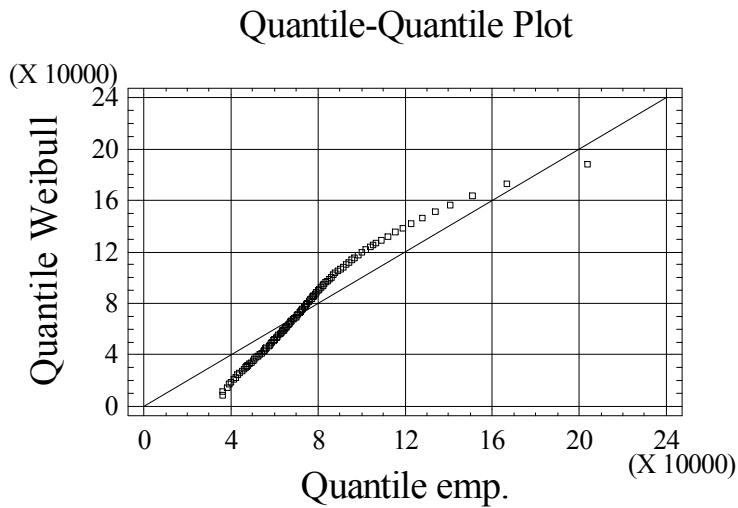


Graf č. 54

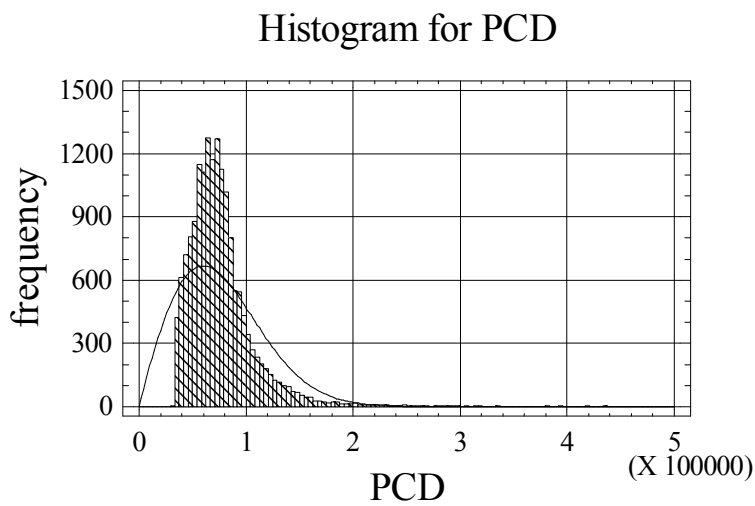


#### 5.4.3.3 Weibullovo rozdělení

Graf č. 55



Graf č. 56



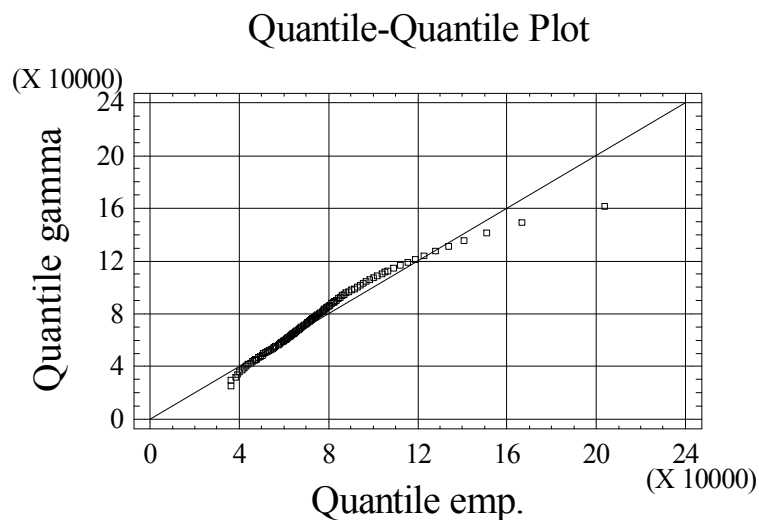
Na základě výpočetních i grafických analýz je možné u Weibullova rozdělení přijmout stejné závěry jako u normálního. Po srovnání empirických a teoretických kvantilů není možné najít shodu ani v některé části rozdělení. V dolní a horní části teoretický tvar podhodnocuje a v prostřední části nadhodnocuje skutečně naměřená data. Histogram pak tyto závěry o nevhodnosti potvrzuje.

#### 5.4.3.4 *Pareto rozdělení*

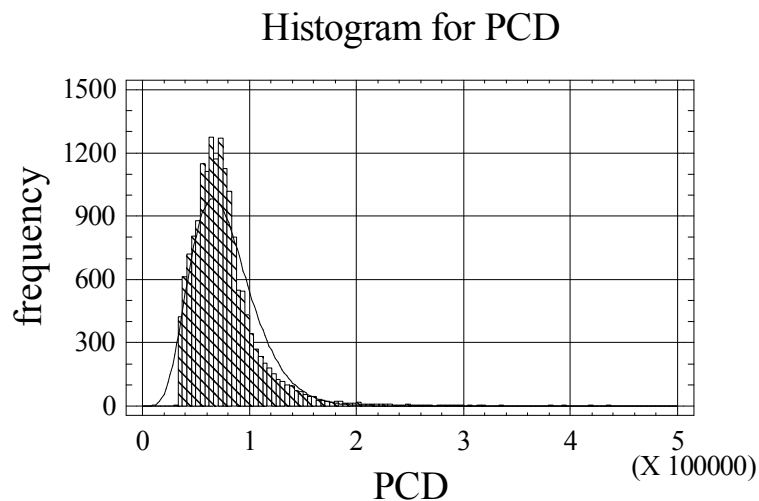
Pokud porovnáme jednotlivé kvantily empirického a Paretova rozdělení v tabulce č. 8, je možné vidět na první pohled výrazné rozdíly, a proto není potřeba sestavit dodatečné grafické analýzy. Nevhodnost v celé části je možné konstatovat i bez nich.

#### 5.4.3.5 *Gama rozdělení*

Graf č. 57



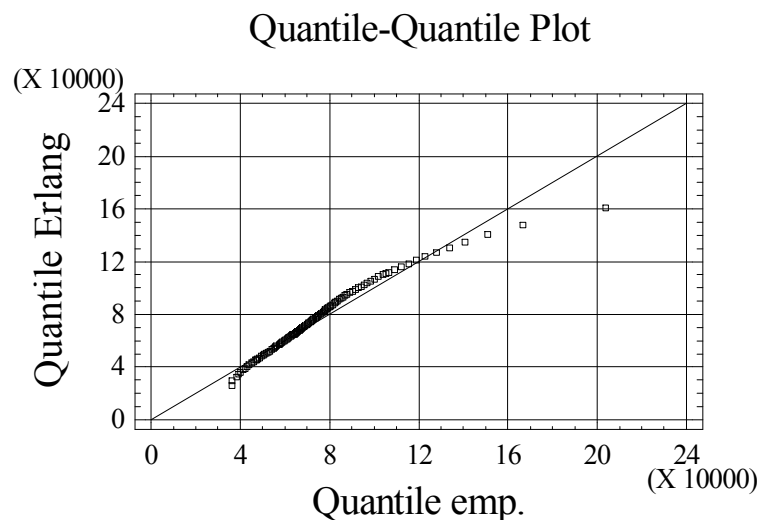
Graf č. 58



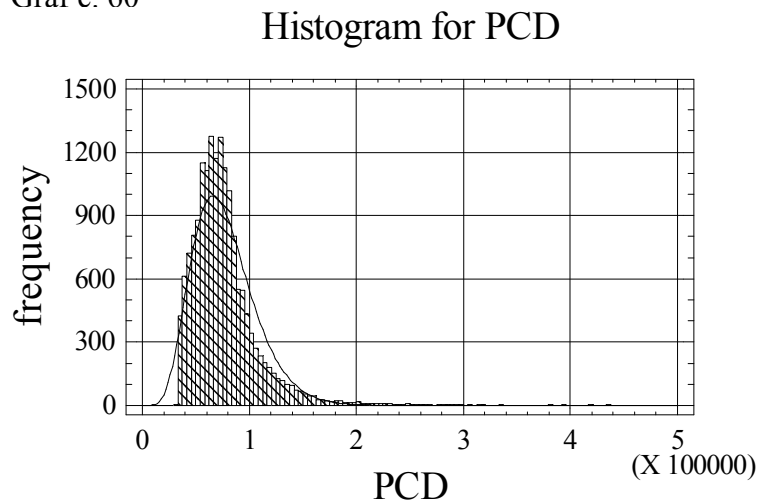
Gama rozdělení má druhou nejnižší hodnotu  $\chi^2$  testu. To je také vidět na grafických analýzách. Na první pohled je možné říct, že v dolní a horní části teoretické rozdělení nevyhovuje datům. V obou těchto koncích podhodnocuje data. To také ukazuje histogram, kde v obou koncích teoretický tvar nadhodnocuje skutečně naměřené četnosti. U vrcholu rozdělení je pak podhodnocuje. Pro porovnání empirických a teoretických kvantilů ve střední části je potřeba se podívat na detailnější grafické analýzy, které jsou uvedeny v příloze č. 14. Tyto grafy ukázaly, že shodu lze najít pouze od cca 25 – 35. percentilu, což je příliš malá část.

#### 5.4.3.6 Erlangovo rozdělení

Graf č. 59



Graf č. 60



Na první pohled je opět těžké najít v těchto grafických analýzách rozdíl od gama rozdělení. V obou koncích Erlangovo tvar nemodeluje dobře naměřená data. Ve střední části bychom

shodu hledali opět ve velmi malé části, proto jsem tentokrát detailnější grafické analýzy nezveřejňoval.

#### **5.4.3.7            *Shrnutí***

Tak jak tomu bylo u všech třech předchozích studiích, žádný jednoduchý teoretický tvar nemodeluje dostatečně kvalitně celou část rozdělení. To už napověděly velmi vysoké výsledky  $\chi^2$  testu dobré shody, které mnohonásobně překročily kritickou hodnotu. V dolní a horní části se opět nepodařilo najít vhodný model. Gama a Erlangovo rozdělení modeluje kvalitně oblast v rozmezí jednoho decilu, což je příliš málo. Nejlépe opět vyšel lognormální model, který lze považovat za přípustný od 5. percentilu po medián.

### **5.5            *Tříparametrické lognormální rozdělení***

Jelikož ve všech předchozích analýzách, a to jak u celkových příjmů domácností, tak u příjmů na jednoho člena domácnosti, vyšly nejnižší hodnoty  $\chi^2$  testu dobré shody u jednoduchého dvouparametrického lognormálního rozdělení, rozhodl jsem se použít a ověřit vhodnost tříparametrického lognormálního rozdělení. Jelikož je má práce založená na kvantilových funkcích, odhady parametrů provedu kvantilovou metodou. Pro výpočet jsem použil 25% a 75% kvantil, jelikož hodnoty vycházely přesněji, než s doporučovaným 30% a 70% kvantilem.

#### **5.5.1            *Tříparametrické lognormální rozdělení pro příjmy domácností v České republice za rok 2002***

Na základě vzorců popsaných v teoretické části jsem provedl odhad tříparametrického lognormálního rozdělení. Výsledky můžeme vidět v následující tabulce.

Tabulka č. 8 – Parametry lognormálního rozdělení.

$\mu$	12,17778
$\sigma^2$	0,362656
$\gamma$	-3497,6

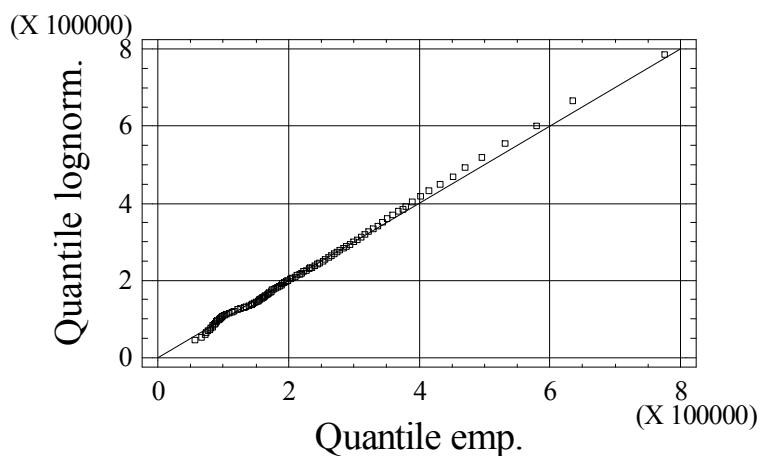
Tabulka č. 9 – Porovnání empirického rozdělení příjmů domácností s tříparametrickým lognormálním rozdělením.

tvar rozdělení	empirické	lognormální
průměr	230004	229575
kvantily	x	x
0,01	57351	44401
0,02	67195	52946
0,03	71901	59142
0,04	74400	64247
0,05	76800	68705
0,25	126023	126023
0,5	190923	190923
0,75	288343	288343
0,95	495967	520020
0,96	530990	554472
0,97	580024	599948
0,98	635700	666190
0,99	775566	785665

Z tabulky je patrné, že průměr vypočtený z tříparametrického lognormálního rozdělení se téměř shoduje s průměrem skutečným. Hodnota mediánu je totožná se skutečně naměřeným. Abychom však mohli analyzovat vhodnost či nevhodnost lognormálního rozdělení, sestojím opět graf, kde porovnam jednotlivé empirické a teoretické kvantily.

Graf č. 61

### Quantile-Quantile Plot



Graf tříparametrického lognormálního rozdělení (graf č. 61) je samozřejmě velmi podobný grafu dvouparametrického rozdělení (graf č. 9). Po důkladnější analýze je možné konstatovat, že ani tříparametrický model není vhodný v celé délce rozdělení. Shodu lze opět hledat až od dolního kvartilu po devátý decil. To je stejně velká oblast. Ovšem kvantily tříparametrického tvaru vystihují skutečně naměřená data lépe.

### 5.5.2 Tříparametrické lognormální rozdělení pro příjmy domácností ve Slovenské republice za rok 2003

Tabulka č. 10 – Parametry lognormálního rozdělení.

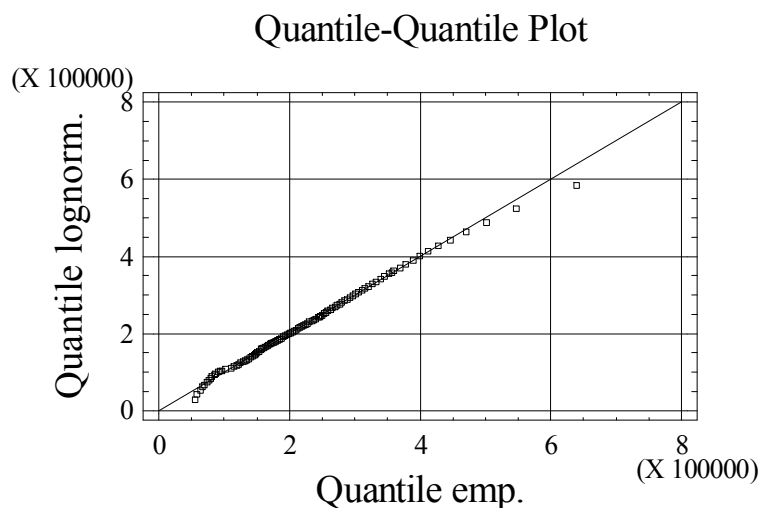
$\mu$	12,73191
$\sigma^2$	0,103362
$\gamma$	-131278

Tabulka č. 11 – Porovnání empirického rozdělení příjmů domácností s tříparametrickým lognormálním rozdělením.

tvar rozdělení	empirické	lognormální
průměr	228285	225043
kvantily	x	x
0,01	55572	28891
0,02	58217	43562
0,03	63277	53560
0,04	66420	61455
0,05	70200	68126
0,25	143131	141131
0,5	207096	207096
0,75	289035	289035
0,95	446402	442918
0,96	469843	462791
0,97	501484	488168
0,98	547215	523588
0,99	639988	583570

Průměr lognormálního rozdělení je opět velmi blízký skutečně naměřenému průměru a hodnota mediánu je u empirického a teoretického rozdělení stejná. To, jak odpovídá tříparametrické lognormální rozdělení skutečně naměřeným hodnotám je možné vidět v následujícím grafu.

Graf č. 62



Z grafu je na první pohled vidět, že teoretické rozdělení neodpovídá skutečně naměřeným datům v dolní a horní části. Skutečné hodnoty jsou zde podhodnoceny. Pro střední část je třeba udělat detailnější analýzu. Tu uvádím v příloze č. 15. Na základě těchto detailnějších grafů je možné konstatovat, že tříparametrické lognormální rozdělení modeluje dostatečně kvalitně empirická data od 16. do 95. percentilu. To je tedy velký rozdíl od dvouparametrického tvaru. Zde je tedy názorný příklad, jak přidání jednoho parametru zvyšuje pružnost přizpůsobení a zkvalitňuje výsledný model.

### 5.5.3 Tříparametrické lognormální rozdělení pro příjmy na jednoho člena domácnosti v České republice za rok 2002

Tabulka č. 12 – Parametry lognormálního rozdělení.

$\mu$	10,55194
$\sigma^2$	0,618426
$\gamma$	50952,28

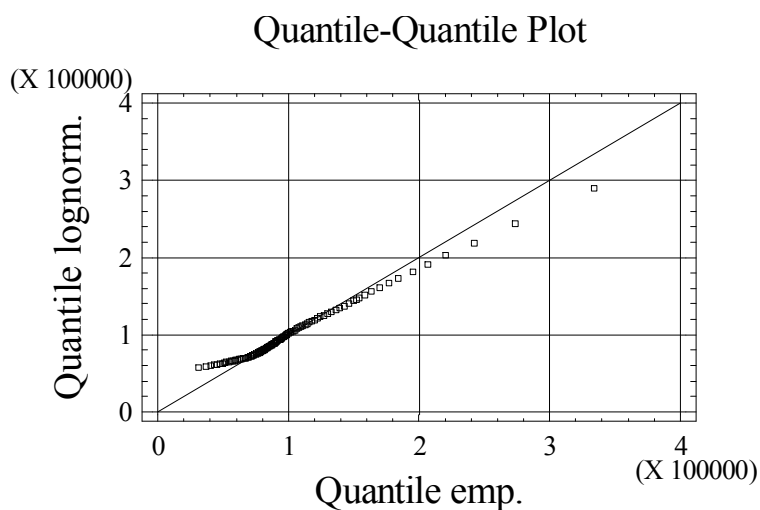


Tabulka č. 13 – Porovnání empirického rozdělení PCD s tříparametrickým lognormálním rozdělením.

tvar rozdělení	empirické	lognormální
průměr	105030	103065
kvantily	x	x
0,01	31185	57092
0,02	37094	58560
0,03	40122	59668
0,04	43205	60607
0,05	45757	61445
0,25	73458	73458
0,5	89204	89204
0,75	115967	115967
0,95	206672	190402
0,96	220090	202504
0,97	242349	218832
0,98	273215	243291
0,99	334113	289276

Tak jak tomu bylo v obou případech u příjmů domácností, tak i pro příjem na jednotlivého člena domácnosti v České republice se průměr lognormálního rozdělení blíží k průměru skutečnému a medián je dokonce totožný. Opět provedu grafickou analýzu srovnání skutečných a teoretických kvantilů tříparametrického lognormálního rozdělení.

Graf č. 63



U příjmů domácností v ČR i ve SR se grafy porovnání skutečných a teoretických kvantilů dvouparametrického a tříparametrického lognormálního rozdělení velmi podobaly. U příjmů na jednoho člena domácnosti v České republice tomu tak není. U rozdělení se dvěma parametry se dala najít shoda ve velmi malé dolní části. To na první pohled u tvaru se třemi parametry neplatí. Toto rozdělení nelze považovat za vhodné v dolní ani horní části. V prvním případě jsou skutečná data nadhodnocena a ve druhém případě je tomu naopak. Pro střední část jsem vytvořil detailnější graf, který je uveden v příloze č. 16. Ten přinesl poznatky, na základě kterých lze konstatovat, že tříparametrické lognormální rozdělení je vhodné od druhého do osmého decilu. Přidáním jednoho parametru se tedy model opět velmi zkvalitnil.

#### **5.5.4 Tříparametrické lognormální rozdělení pro příjmy na jednoho člena domácnosti ve Slovenské republice za rok 2003**

Pouze u rozdělení příjmů na jednoho člena domácnosti ve Slovenské republice bylo pro výpočet přesnější použít doporučený 30% a 70% kvantil.

Tabulka č. 14 – Parametry lognormálního rozdělení.

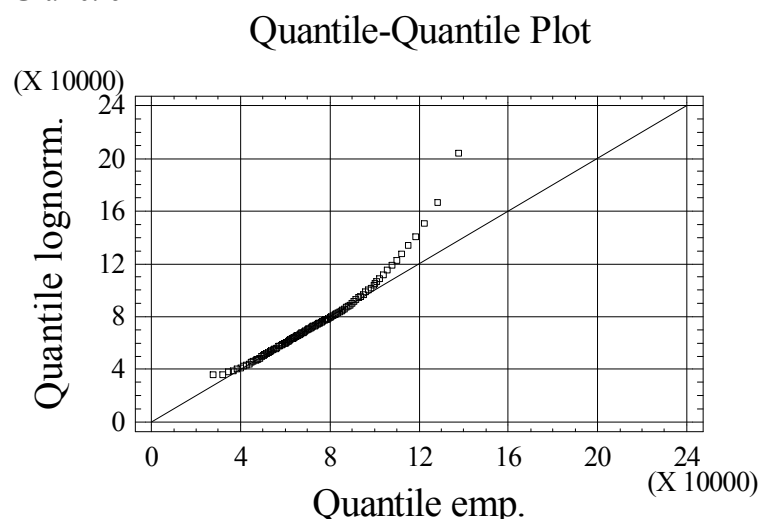
$\mu$	11,72347
$\sigma^2$	0,0345
$\gamma$	-52515,5

Tabulka č. 15 – Porovnání empirického rozdělení PCD s tříparametrickým lognormálním rozdělením.

tvar rozdělení	empirické	lognormální
průměr	77115	73068
kvantily	x	x
0,01	36072	27612
0,02	36304	31774
0,03	38139	34525
0,04	39121	36654
0,05	40187	38425
0,25	56410	56385
0,5	70920	70920
0,75	86879	87395
0,95	133833	115027
0,96	140724	118353
0,97	150751	122532
0,98	166780	128247
0,99	204000	137635

Ani tento případ není výjimkou a teoretický průměr je velmi blízký skutečnému. I medián je jako ve všech předchozích případech totožný. Opět následně porovnáme skutečné a teoretické kvantily.

Graf č. 64



I v ČR se pro rozdělení příjmů na jednoho člena domácnosti liší grafy porovnání skutečných a teoretických kvantilů pro dvouparametrické a tříparametrické lognormální rozdělení. Dolní a horní konec nemodeluje model se třemi parametry dobře. V obou případech výrazně nadhodnocuje skutečně naměřené hodnoty. Pro střední část jsem opět vytvořil detailnější graf, který je uveden v příloze č. 17. Na jeho základě je možné učinit závěr, že tříparametrické lognormální rozdělení modeluje dostatečně kvalitně skutečná data od 8. do 80 percentilu. To je výrazně větší část, než u modelu se dvěma parametry.

## **5.6 Dvouparametrické Paretovo rozdělení**

U modelování pomocí jednoduchých pravděpodobnostních funkcí jsem již Paretovo rozdělení použil, ale pouze za pomoci programu Statgraphics, který počítá pouze s jedním parametrem. Rozhodl jsem se tedy sestrojít dvouparametrický tvar, kde parametry tohoto rozdělení budou vypočítány momentovou metodou.

### **5.6.1 Dvouparametrické Paretovo rozdělení pro příjmy domácností v České republice za rok 2002**

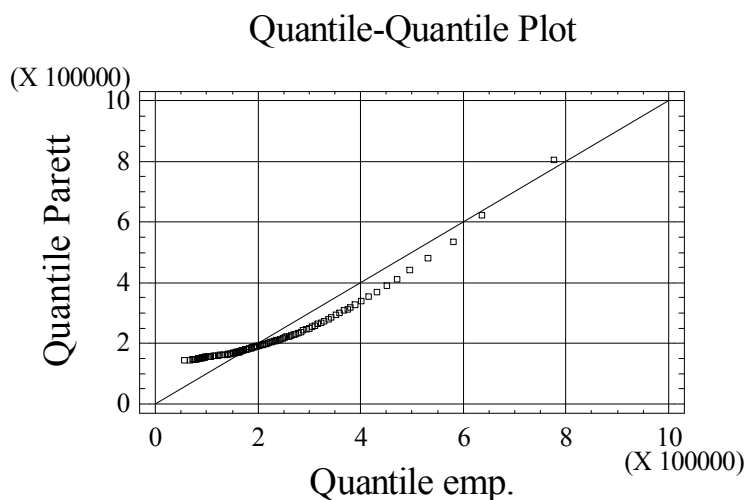
Tabulka č. 16 – Parametry Paretova rozdělení

c	2,67839
$x_0$	144129,9

Tabulka č. 17 – Porovnání empirického rozdělení příjmů domácností s dvouparametrickým Paretovým rozdělením

tvar rozdělení	empirické	Pareto
kvantily		
0,01	57351	144672
0,02	67195	145221
0,03	71901	145778
0,04	74400	146343
0,05	76800	146917
0,25	126023	160473
0,5	190923	186701
0,75	288343	241846
0,95	495967	441067
0,96	530990	479388
0,97	580024	533745
0,98	635700	620982
0,99	775566	804399

Graf č. 65



Tak jak ukazuje tabulka č. 17 i graf č. 65, ani dvouparametrické Pareto rozdělení není vhodné pro rozdělení příjmů domácností v České republice. V obou koncích tento teoretický model nadhodnocuje skutečně naměřená data, zatímco ve středu rozdělení je tomu naopak.

## 5.6.2 Dvouparametrické Pareto rozdělení pro příjmy domácností ve Slovenské republice za rok 2003

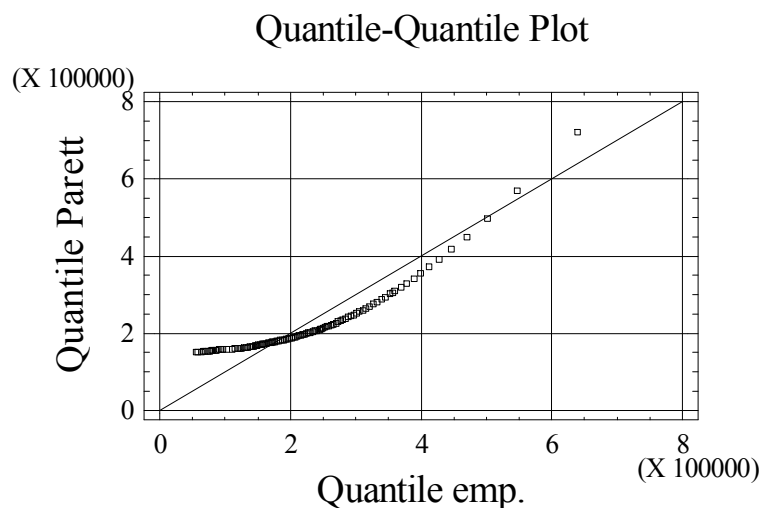
Tabulka č. 18 – Parametry Pareto rozdělení

c	2,940226
$x_0$	150643,3

Tabulka č. 19 – Porovnání empirického rozdělení příjmů domácností s dvouparametrickým Paretovým rozdělením

tvar rozdělení	empirické	Pareto
kvantily		
0,01	55572	151159
0,02	58217	151682
0,03	63277	152212
0,04	66420	152749
0,05	70200	153294
0,25	143131	166128
0,5	207096	190692
0,75	289035	241388
0,95	446402	417295
0,96	469843	450197
0,97	501484	496473
0,98	547215	569884
0,99	639988	721389

Graf č. 66



Bohužel ani pro data domácností ve Slovenské republice není Paretovo dvouparametrické rozdělení vhodné. Jednotlivé kvantily se velmi liší od skutečně naměřených a vzdálenost od linie 45 stupňů je velmi podobná jako v České republice.

### 5.6.3 Dvouparametrické Paretovo rozdělení pro příjmy na jednoho člena domácnosti v České republice za rok 2002

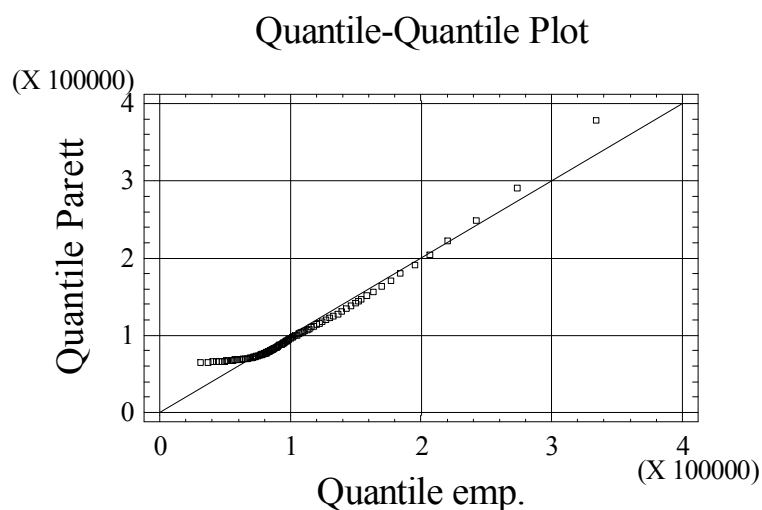
Tabulka č. 20 – Parametry Paretova rozdělení

c	2,605819
$x_0$	64723,98

Tabulka č. 21 – Porovnání empirického rozdělení příjmů na člena domácnosti s dvouparametrickým Paretovým rozdělením

tvar rozdělení	empirické	Paretovo
kvantily		
0,01	31185	64974
0,02	37094	65228
0,03	40122	65485
0,04	43205	65746
0,05	45757	66011
0,25	73458	72279
0,5	89204	84448
0,75	115967	110182
0,95	206672	204336
0,96	220090	222604
0,97	242349	248588
0,98	273215	290440
0,99	334113	378947

Graf č. 67



Jak je vidět z předchozího grafu a tabulky, dvouparametrické Pareto rozdělení modeluje lépe rozdělení příjmů na jednoho člena domácnosti. Shodu lze najít od druhého decilu do 95 percentilu. V dolní části a v posledních pěti percentilech teoretický model nadhodnocuje skutečně naměřená data.

#### 5.6.4 Dvouparametrické Pareto rozdělení pro příjmy na jednoho člena domácnosti ve Slovenské republice za rok 2003

Tabulka č. 22 – Parametry Pareto rozdělení

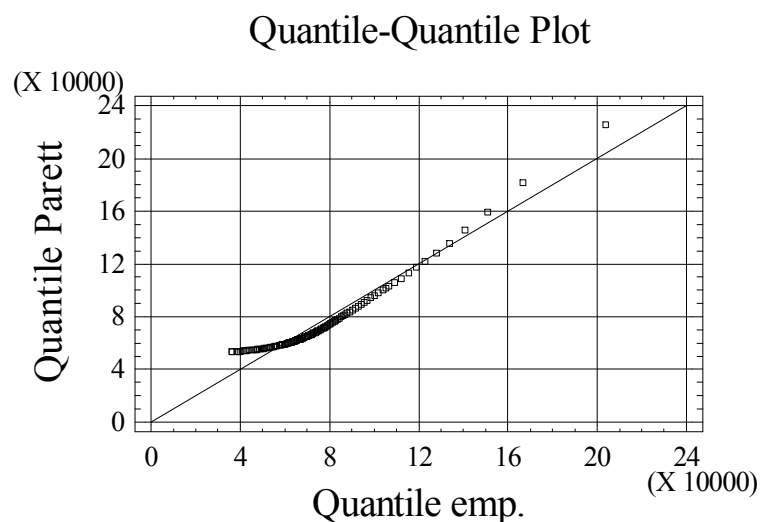
c	3,168549
$x_0$	52777,34



Tabulka č. 23 – Porovnání empirického rozdělení příjmů na člena domácnosti s dvouparametrickým Paretovým rozdělením

tvar rozdělení	empirické	Paretovo
kvantily		
0,01	36072	52945
0,02	36304	53115
0,03	38139	53287
0,04	39121	53462
0,05	40187	53639
0,25	56410	57793
0,5	70920	65683
0,75	86879	81745
0,95	133833	135849
0,96	140724	145761
0,97	150751	159614
0,98	166780	181403
0,99	204000	225762

Graf č. 68



Tak jako tomu bylo u jednotlivých členů domácnosti v České republice, tak i u slovenských dat lze dvouparametrické Paretovo rozdělení možné považovat za vhodné od druhého decilu do 95. percentilu. U českých dat je však model o něco přesnější.

## **5.7 Modelování celkových příjmů domácností podle počtu členů domácnosti**

Poté, co jsem prověřil všechny potencionálně vhodné jednoduché modely, rozhodl jsem se modelovat celkové příjmy domácností v ČR a v SR podle počtu členů domácnosti. K tomuto rozhodnutí mě vedl závažný důvod. Pokud se podíváme zpět na histogramy, zjistíme, že empirické rozdělení má dva vrcholy. Avšak všechny teoretické tvary jsou jednovrcholové. Proto je v následující analýze nutné domácnosti dále rozdělit. To je možné podle dvou hledisek. První je počet členů domácnosti. Druhé hledisko je sociální skupina. Rozhodl jsem se ve své práci použít obě hlediska. Podle sociálních skupin však budu modelovat pouze data SR. Důvod je uveden v části 5.8. Podle prvního hlediska rozdělím domácnosti na dvě skupiny. Na jednočlenné a vícečlenné domácnosti. V následujících analýzách vždy použiji model, který vyšel v předchozích studiích jako vhodný, pokud nebude odpovídat, pokusím se najít jiný.

### **5.7.1 Modelování celkových příjmů domácností v ČR podle počtu členů domácnosti za rok 2002**

#### **5.7.1.1 Vícečlenné domácnosti**

##### **5.7.1.1.1 Tříparametrické lognormální rozdělení**

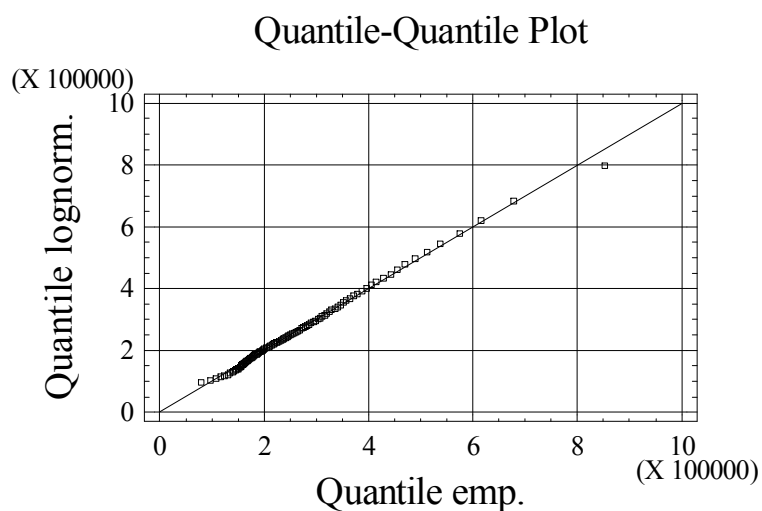
Tabulka č. 24 – Parametry lognormálního rozdělení.

$\mu$	12,10757
$\sigma^2$	0,370579
$\gamma$	51491,36

Tabulka č. 25 – Porovnání empirického rozdělení tříparametrickým lognormálním rozdělením.

tvar rozdělení	empirické	lognormální
průměr	271768	269625
kvantily	x	x
0,01	79248	95468
0,02	97379	103406
0,03	107556	109170
0,04	116202	113924
0,05	123557	118078
0,25	171699	171699
0,5	232731	232731
0,75	324749	324749
0,95	536611	544795
0,96	575121	577623
0,97	616560	620988
0,98	678330	684219
0,99	852774	798432

Graf č. 69



Z grafu č. 69 je vidět, že výsledný model je ještě přesnější, než tomu bylo u domácností jako celku. Průměr lognormálního rozdělení je téměř totožný s průměrem skutečných dat. Pouze v dolní části neodpovídají teoretické kvantily skutečně naměřeným. Shodu lze hledat od druhého decilu a to až do 99. percentilu, kdy všechny body leží téměř na přímce s linií 45 stupňů. Proto nebudu dále hledat jiné vhodnější rozdělení.

## 5.7.1.2 *Jednočlenné domácnosti*

### 5.7.1.2.1 **Tříparametrické lognormální rozdělení**

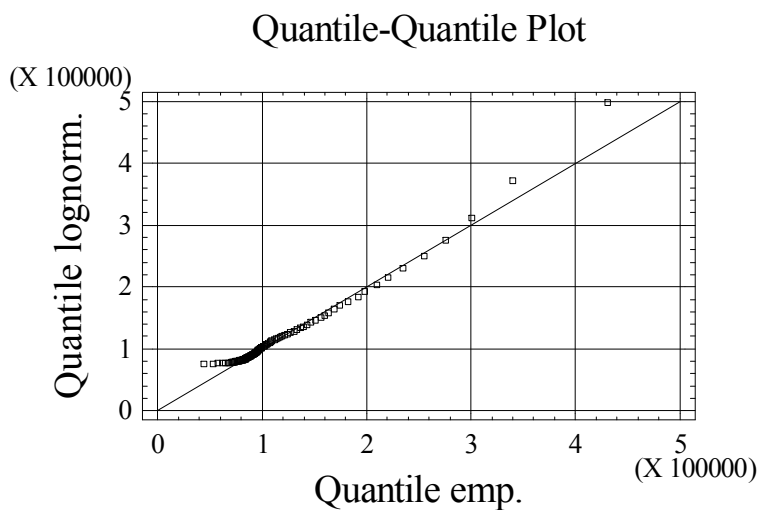
Tabulka č. 26 – Parametry lognormálního rozdělení.

$\mu$	9,94279
$\sigma^2$	1,678381
$\gamma$	74478,3

Tabulka č. 27 – Porovnání empirického rozdělení tříparametrickým lognormálním rozdělením.

tvar rozdělení	empirické	lognormální
průměr	121582	122624
kvantily	x	x
0,01	44282	75500
0,02	52694	75932
0,03	57859	76298
0,04	62076	76632
0,05	64294	76948
0,25	83160	83160
0,5	95280	95280
0,75	124320	124320
0,95	255729	249688
0,96	275681	275435
0,97	300850	312330
0,98	340324	372068
0,99	431308	498114

Graf č. 70



Jak je vidět v tabulce č. 27, průměry teoretického a empirického rozdělení se opět velmi podobají. Tentokrát lze však tříparametrickým lognormálním rozdělením modelovat menší část, než tomu bylo u vícečlenných domácností. Konkrétně se jedná o oblast od druhého do osmého decilu. Jelikož u modelování příjmů na jednoho člena domácnosti vycházelo velmi dobře dvouparametrické Paretovo rozdělení, pokusím se ho sestavit i v tomto případě.

#### 5.7.1.2.2 Dvouparametrické Paretovo rozdělení

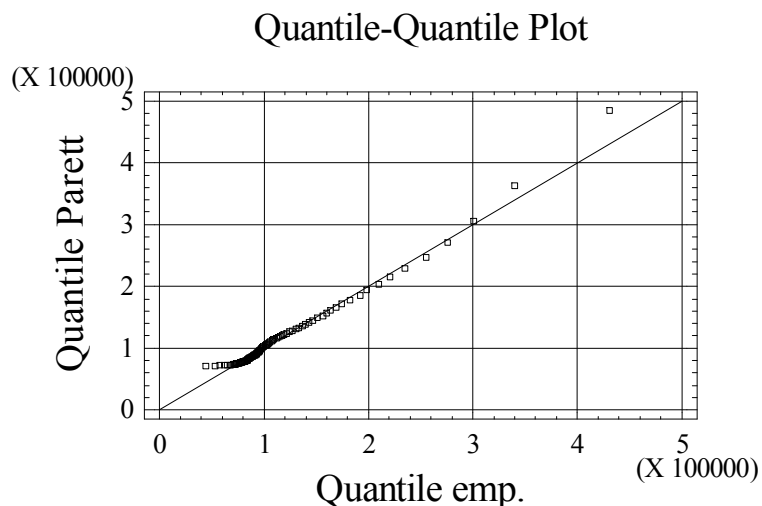
Tabulka č. 28 – Parametry Paretova rozdělení

c	2,394449
$x_0$	70833,73

Tabulka č. 29 – Porovnání empirického rozdělení s dvouparametrickým Paretoovým rozdělením

tvar rozdělení	empirické	Paretovo
0,01	44282	71132
0,02	52694	71434
0,03	57859	71741
0,04	62076	72052
0,05	64294	72367
0,25	83160	79876
0,5	95280	94615
0,75	124320	126380
0,95	255729	247510
0,96	275681	271685
0,97	300850	306368
0,98	340324	362899
0,99	431308	484736

Graf č. 71



Jak jsem předpokládal, tak i Paretovo rozdělení je možné považovat ve střední části za vhodné. Dokonce modeluje dostatečně kvalitně větší část, než tomu bylo u lognormálního rozdělení. Konkrétně jde o oblast od prvního do devátého decilu.

## 5.7.2 Modelování celkových příjmů domácností v SR podle počtu členů domácnosti za rok 2003

### 5.7.2.1 *Vícečlenné domácnosti*

#### 5.7.2.1.1 **Tříparametrické lognormální rozdělení**

Tabulka č. 30 – Parametry lognormálního rozdělení.

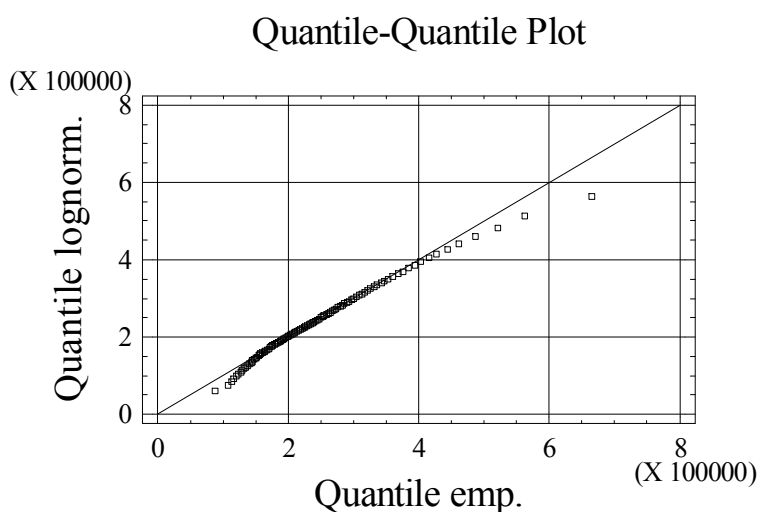
$\mu$	12,73539
$\sigma^2$	0,087426
$\gamma$	-110733

Tabulka č. 31 – Porovnání empirického rozdělení tříparametrickým lognormálním rozdělením.

tvar rozdělení	empirické	lognormální
průměr	253154	243994
kvantily	x	x
0,01	88359	59946
0,02	107077	74273
0,03	112899	83980
0,04	116257	91616
0,05	121013	98048
0,25	167429	167429
0,5	228822	228822
0,75	303765	303765
0,95	461993	441508
0,96	487580	459062
0,97	521329	481409
0,98	562327	512478
0,99	665116	564790

Jak je vidět v tabulce č. 31, tak i v následujícím grafu, tříparametrické lognormální rozdělení opět modeluje velmi dobře střed rozdělení. Konkrétně se jedná o oblast od druhého decilu do 85. percentilu. To je menší část, než tomu bylo na českých datech. V obou koncích jsou teoretické kvantily výrazně podhodnoceny v porovnání s empirickými. U českých dat byl tříparametrický model pro vícečlenné domácnosti přesnější než u domácností jako celku, ale zde je tomu bohužel naopak. Jelikož u domácností jako celku bylo ve Slovenské republice nejvhodnější gama rozdělení, ověřil jsem jeho vhodnost i u této skupiny. Bohužel jeho vhodnost nebyla prokázána, kvalitně modeluje pouze velmi malou část.

Graf č. 72



### 5.7.2.2 *Jednočlenné domácnosti*

#### 5.7.2.2.1 **Tříparametrické lognormální rozdělení**

Tabulka č. 32 – Parametry lognormálního rozdělení.

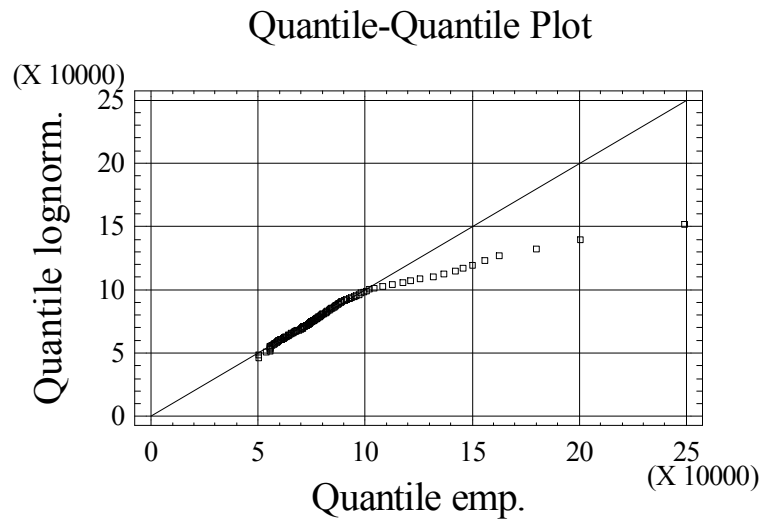
$\mu$	10,88616
$\sigma^2$	0,140438
$\gamma$	24022,23

Tabulka č. 33 – Porovnání empirického rozdělení s lognormálním rozdělením.

tvar rozdělení	empirické	lognormální
průměr	88215	81341
kvantily	x	x
0,01	50520	46367
0,02	50520	48771
0,03	53996	50428
0,04	55572	51747
0,05	55572	52869
0,25	65520	65520
0,5	77454	77454
0,75	92820	92820
0,95	156000	122992
0,96	162467	126996
0,97	179989	132142
0,98	200536	139381
0,99	249129	151789



Graf č. 73



Z předchozího grafu vidíme velký rozdíl v horní části rozdělení. Teoretické kvantily jsou výrazně odlišné od skutečně naměřených. Proto bych nepovažoval tříparametrické lognormální rozdělení v tomto případě za vhodné, i když mezi druhým a osmým decilem jsou si empirické a teoretické kvantily velmi blízké. Rozdíl v horním konci je příliš velký. Proto opět sestrojím dvouparametrické rozdělení.

#### 5.7.2.2.2 Dvouparametrické Pareto rozdělení

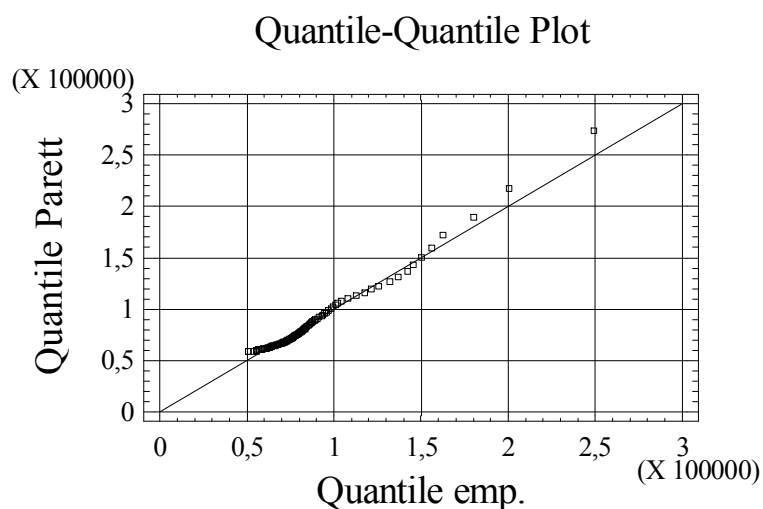
Tabulka č. 34 – Parametry Pareto rozdělení

c	2,991005
$x_0$	58732,02

Tabulka č. 35 – Porovnání empirického rozdělení s dvouparametrickým Paretoovým rozdělením

tvar rozdělení	empirické	Paretovo
0,01	50520	58930
0,02	50520	59130
0,03	53996	59333
0,04	55572	59539
0,05	55572	59748
0,25	65520	64662
0,5	77454	74049
0,75	92820	93361
0,95	156000	159903
0,96	162467	172289
0,97	179989	189683
0,98	200536	217221
0,99	249129	273871

Graf č. 74



Bohužel ani dvouparametrické Paretovo rozdělení není v tomto případě možné považovat za vhodné. To je vidět v grafu č. 74, kde křivka jednotlivých kvantilů tvoří jakousi „vlnovku“. Teoretické kvantily odpovídají tedy skutečně naměřeným vždy pouze v krátkých úsecích.

## **5.8 Modelování celkových příjmů domácností ve Slovenské republice za rok 2003 podle sociálních skupin.**

V další praktické části budu hledat vhodný model pro jednotlivé sociální skupiny. Na základě předchozích analýz se dají považovat některé funkce za nevhodné. Proto v následující části již použiji pouze tříparametrické lognormální rozdělení, dvouparametrické gama rozdělení a dvouparametrické Pareto rozdělení. Modelování jednotlivých sociálních skupin jsem se rozhodl vypracovat pouze ze slovenských dat, jelikož českými daty se zabývá Bc. Martin Myšička ve své diplomové práci na téma: Analýza sociální situace domácností v České republice. Jednotlivé skupiny a jejich počet je uveden v následující tabulce.

Tabulka č. 36 – Sociální skupiny a jejich zastoupení v celkovém počtu

Sociální skupina	Počet	%
Zaměstnanci	8027	49,1
Samostatně činní	1043	6,4
Nezaměstnaní	898	5,5
Důchodci	6084	37,2
Ostatní	297	1,8
Celkem	16349	100

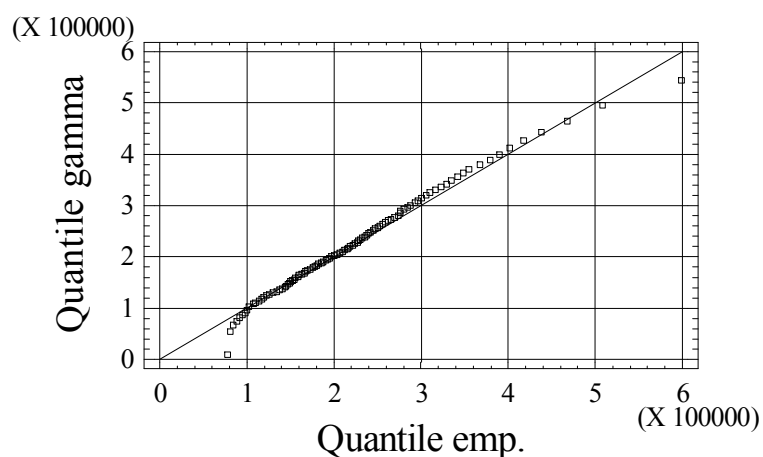
## 5.8.1 Zaměstnanci

Tabulka č. 37 – Parametry rozdělení a jejich kvantily

tvar rozdělení	empirické	Gama	Lognorm.	Paretovo
parametry rozdělení	x	4,66254	0,004346	2,986
	x	0,0000204	14,15937	152070,7
	x	x	-1198507	x
kvantily	x	x	x	x
0,01	77446	54871	11342	152583
0,02	80468	66224	33281	153103
0,03	83916	74273	47406	153630
0,04	88223	80784	58138	154164
0,05	91327	86378	66937	154706
0,25	150533	151294	150533	167452
0,5	211872	212525	211872	191807
0,75	276000	288519	276000	241926
0,95	414956	425970	373407	414740
0,96	435478	443069	384413	446923
0,97	462272	464675	398048	492125
0,98	501982	494405	416356	563704
0,99	597900	543624	445639	710999

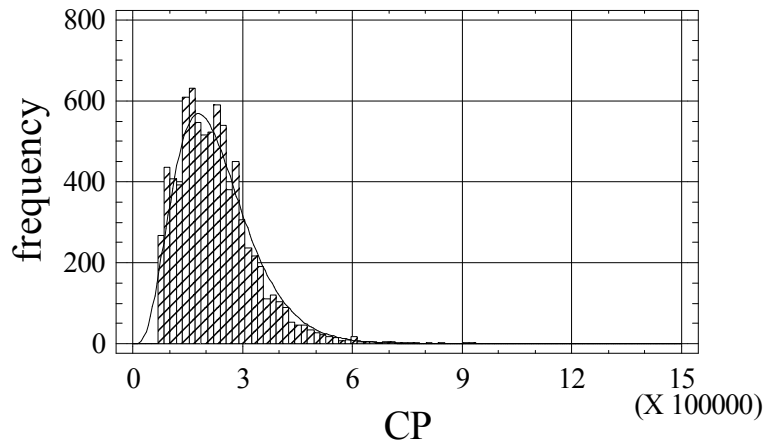
Graf č. 75

### Quantile-Quantile Plot



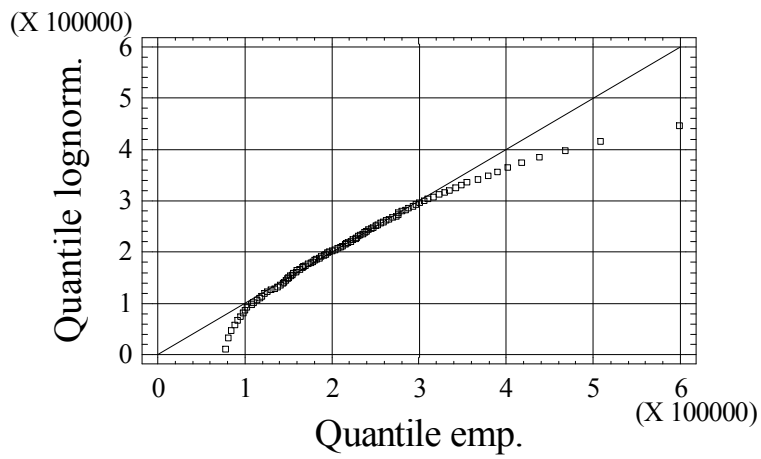
Graf č. 76

Histogram for CP



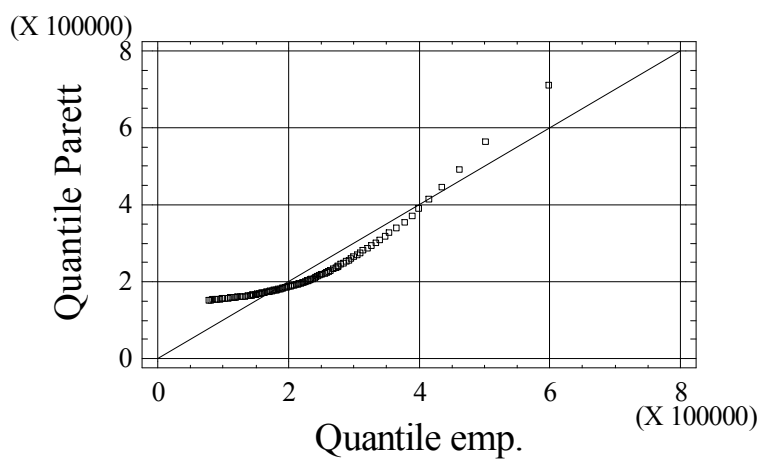
Graf č. 77

Quantile-Quantile Plot



Graf č. 78

Quantile-Quantile Plot



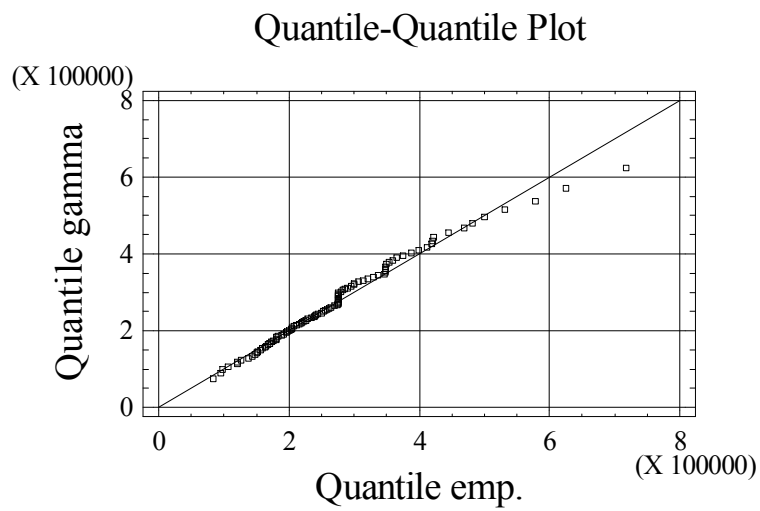
Jak je vidět v tabulce č. 37 a v grafu č. 78, Paretovo dvouparametrické rozdělení není vhodné pro skupinu zaměstnanců. V obou koncích nadhodnocuje skutečně naměřená data. Teoretické kvantily neodpovídají empirickým ani ve středu rozdělení, zde jsou naopak podhodnoceny. Lognormální tříparametrické rozdělení také není vhodné v krajních koncích, ale ve středu ano. Přesně jde o část mezi patnáctým a osmdesátým percentilem. Jako nejvhodnější je však na základě výše uvedených analýz možné považovat gama rozdělení, kde lze hledat shodu již od sedmého percentilu. Od šestého do devátého decilu se sice křivka trochu odchyluje od linie 45 stupňů, ale v posledním decilu opět modeluje skutečně naměřená data lépe, než je tomu u lognormálního rozdělení.

## 5.8.2 Samostatně činní

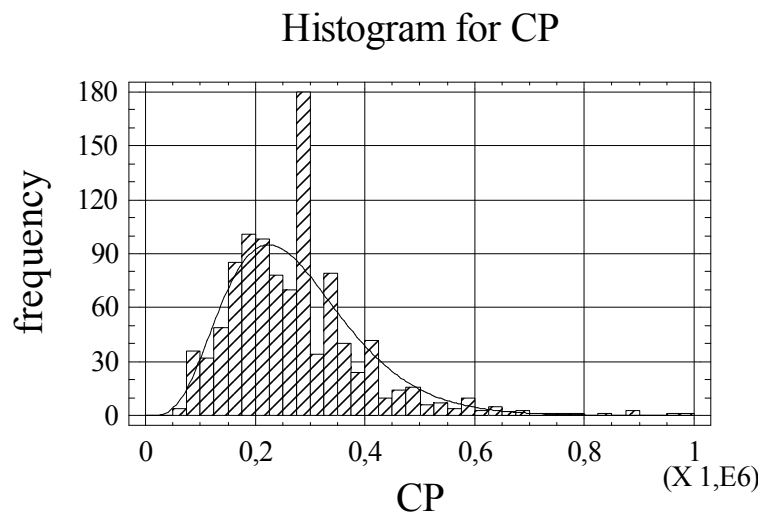
Tabulka č. 38 – Parametry rozdělení a jejich kvantily

tvar rozdělení	empirické	Gama	Lognorm.	Paretovo
parametry rozdělení	x	5,34058	0,003483	3,330605
	x	1,94055E-05	14,43596	192579,5
	x	x	-1598760	x
kvantily	x	x	x	x
0,01	84389	74543	22429	193162
0,02	94840	88407	48720	193751
0,03	98052	98129	65622	194349
0,04	106592	105937	78450	194954
0,05	120000	112612	88958	195568
0,25	188427	188433	188427	209120
0,5	261000	258237	261000	234358
0,75	336520	343570	336520	282227
0,95	500000	495761	450579	396821
0,96	531308	514560	463419	411105
0,97	577709	538281	479313	427922
0,98	625182	570865	500633	448193
0,99	717377	624681	534679	473412

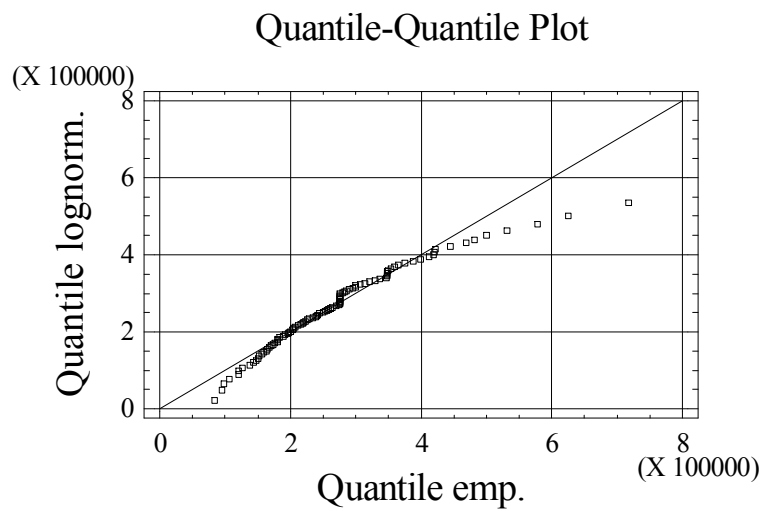
Graf č. 79



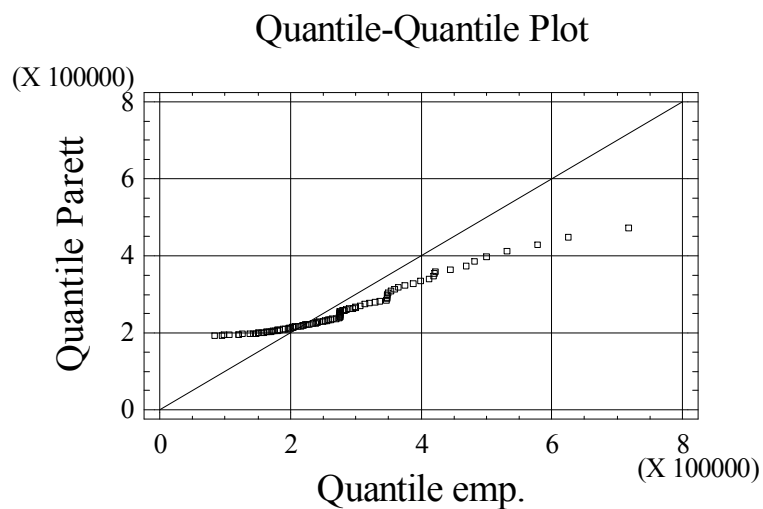
Graf č. 80



Graf č. 81



Graf č. 82



Bohužel ani jedno teoretické rozdění nemodeluje kvalitně dostatečně velkou oblast. Největší část, kterou lze považovat za přípustnou modeluje gama rozdění a to od druhého do pátého decilu. Pak se začínají objevovat odchylky, které lze pozorovat i u ostatních funkcí. To je způsobeno vysokou četností hodnot v intervalu 275-300 tis. Sk. V tomto intervalu je 180 hodnot. To způsobuje stejnou hodnotu několika empirických kvantilů, což způsobuje poté odchylky od kvantilů teoretických.



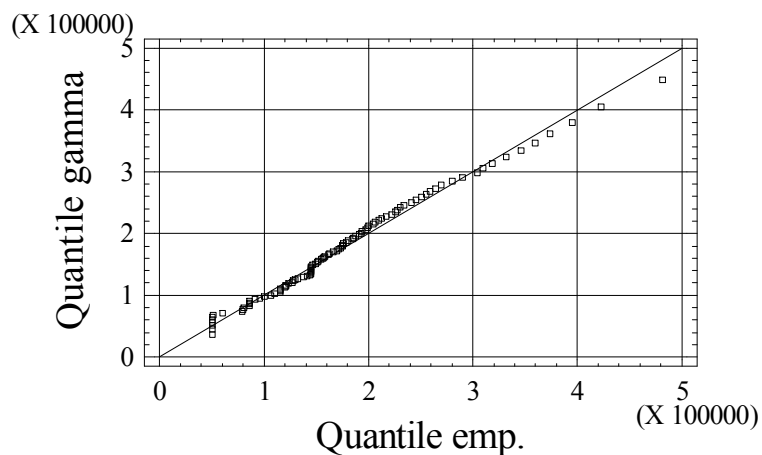
### 5.8.3 Nezaměstnaní

Tabulka č. 39 – Parametry rozdělení a jejich kvantily

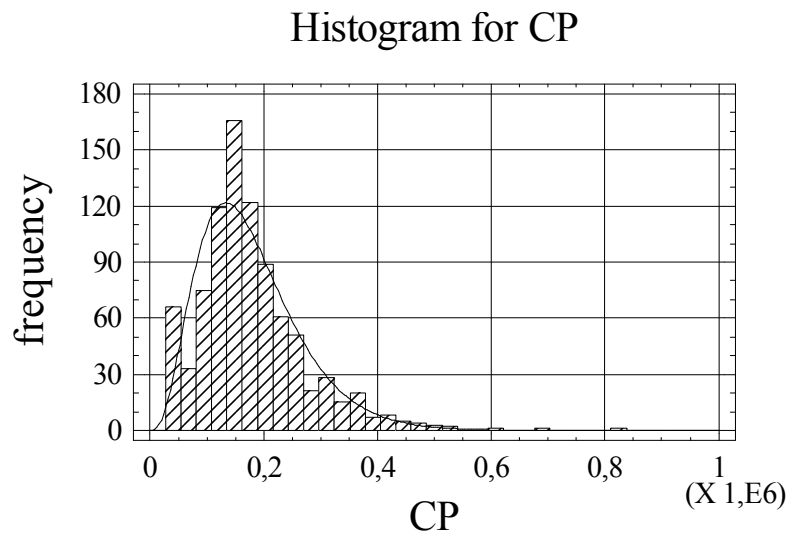
tvar rozdělení	empirické	Gama	Lognorm.	Paretovo
parametry rozdělení	x	3,97184	0,465749	3,152409
	x	2,22971E-05	11,5311	121625,9
	x	x	55941,59	x
kvantily	x	x	x	x
0,01	50520	36386	76757	122014
0,02	50520	44966	81014	122408
0,03	50520	51141	84155	122807
0,04	50520	56184	86774	123211
0,05	50520	60547	89084	123621
0,25	120208	112688	120208	132689
0,5	157776	163427	157776	149663
0,75	217305	227667	217305	182137
0,95	359316	345943	368844	261071
0,96	373628	360787	392280	271010
0,97	395530	379576	423511	282736
0,98	423269	405483	469561	296905
0,99	481685	448498	554132	314583

Graf č. 83

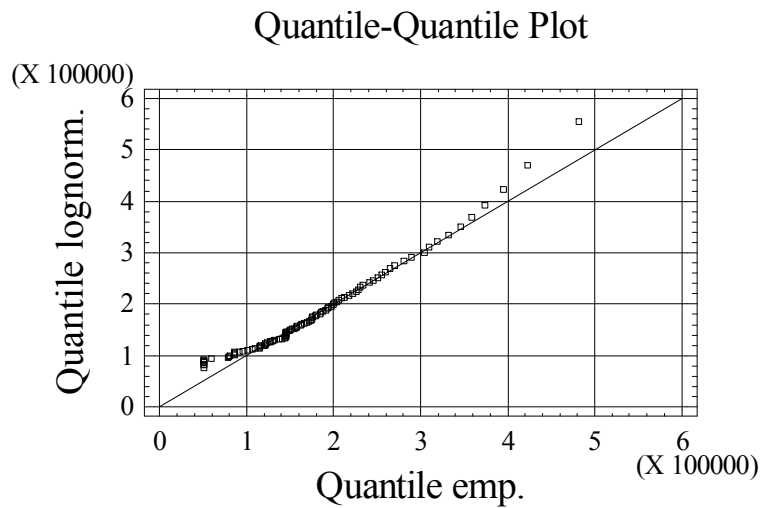
Quantile-Quantile Plot



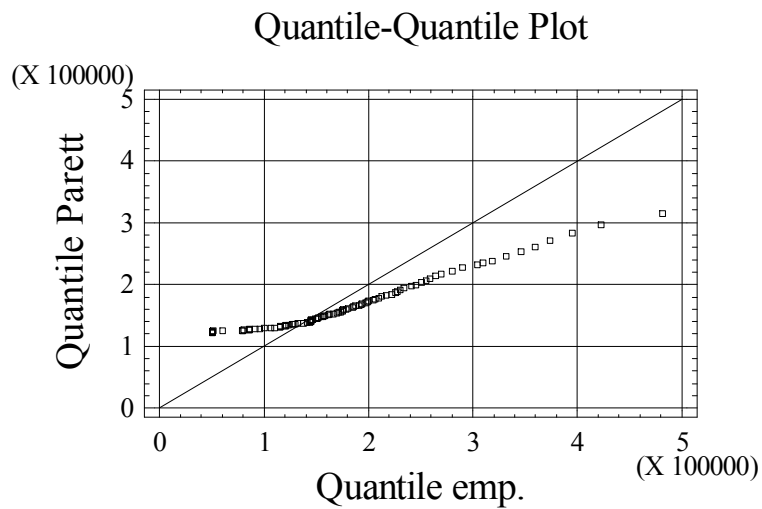
Graf č. 84



Graf č. 85



Graf č. 86



U skupiny nezaměstnaných opět nelze použít Paretovo rozdělení. Ani gama funkce neodpovídá skutečně naměřeným datům, v obou koncích podhodnocuje a ve středu nadhodnocuje empirická data. Jako nejlepší je možné považovat lognormální rozdělení. Shodu lze najít od druhého decilu do přibližně 94. percentilu.

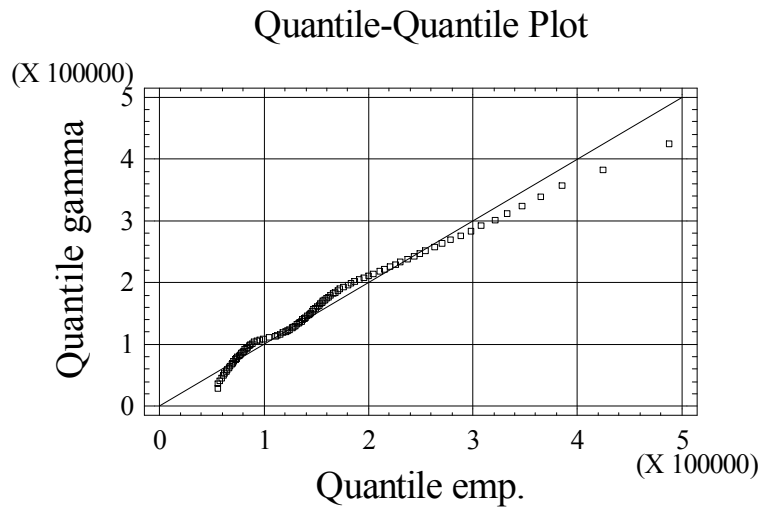
#### 5.8.4 Důchodci

U skupiny důchodců jsem byl nucen použít pro výpočet parametrů lognormálního rozdělení 30% a 70% kvantil.

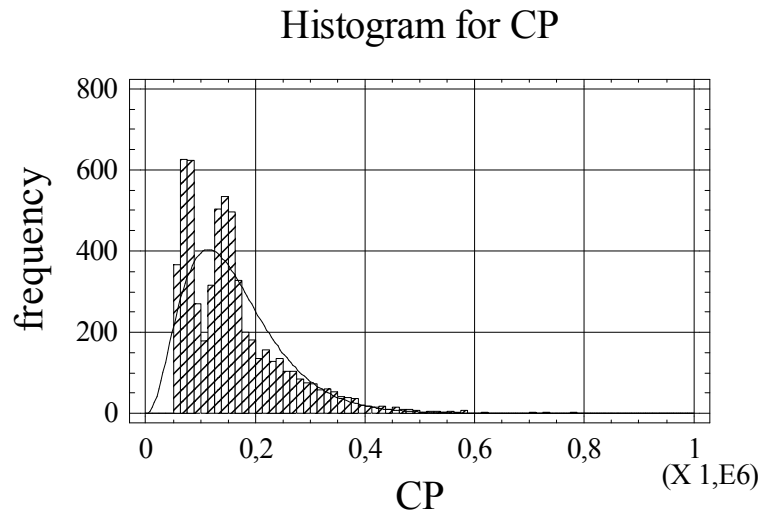
Tabulka č. 40 – Parametry rozdělení a jejich kvantily

tvar rozdělení	empirické	Gama	Lognorm.	Paretovo
parametry rozdělení	x	3,48539	0,256271	2,927594
	x	2,16738E-05	11,90949	105881,6
	x	x	-7537,36	X
kvantily	x	x	x	X
0,01	55572	28322	38252	106246
0,02	55572	35784	45028	106615
0,03	57671	41231	49838	106989
0,04	59179	45720	53744	107368
0,05	60979	49629	57117	107753
0,25	85800	97612	98129	116287
0,5	141133	145721	141133	132382
0,75	195639	207679	201639	163553
0,95	346854	323537	334325	241006
0,96	364902	338188	353141	250900
0,97	385128	356759	377696	262609
0,98	424445	382412	412946	276807
0,99	487768	425109	475167	294593

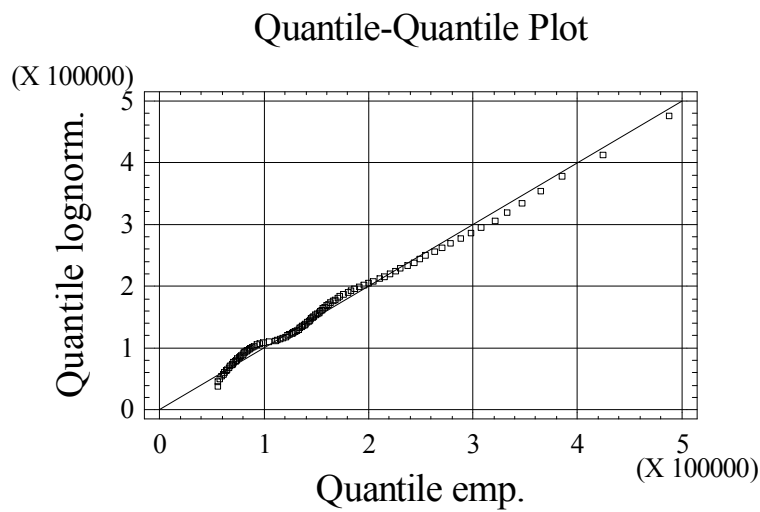
Graf č. 87



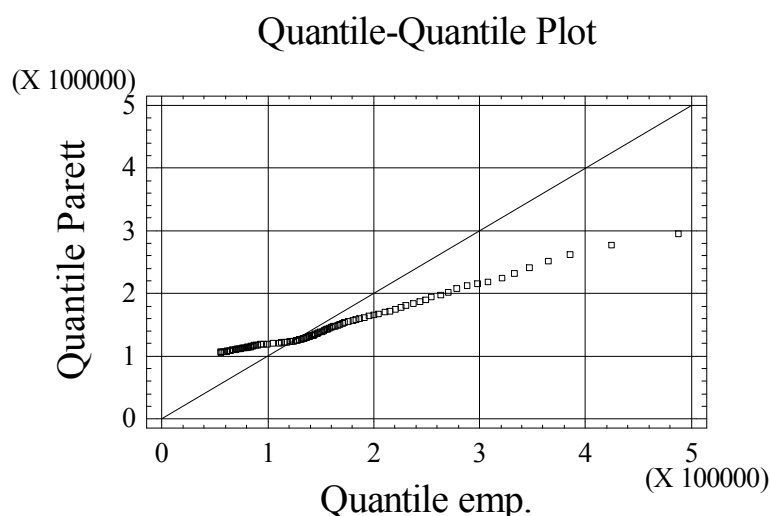
Graf č. 88



Graf č. 89



Graf č. 90



Jak je vidět v tabulce č. 40, tak i v grafech č. 87, 89 a 90, žádný teoretický model nevyhovuje skutečně naměřeným datům. U všech rozdělení se u kvantilů vytvořila „vlnovka“. To je způsobené tím, že empirické rozdělení má dva vrcholy. To je vidět v grafu č. 88. Proto nemůže žádný teoretický tvar odpovídat datům. Všechny teoretické tvary mají jeden vrchol. Proto je potřeba důchodce rozdělit do dalších skupin.

Nejdříve jsem se pokusil rozdělit důchodce podle počtu členů na jednočlenné domácnosti a vícečlenné domácnosti. Podle tohoto hlediska se bohužel nepodařilo najít vyhovující model. Proto jsem se je rozhodl rozdělit na pracující a nepracující. U jednotlivých skupin už uvedu pouze vyhovující model.

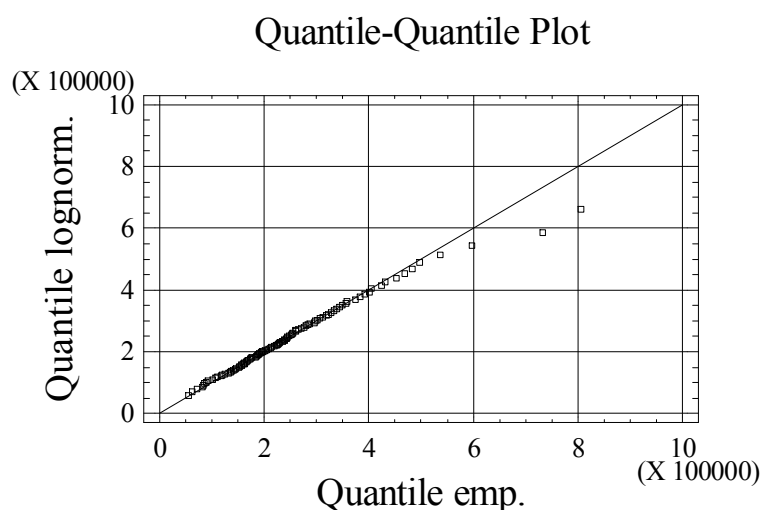
### 5.8.4.1 *Důchodci – pracující*

#### 5.8.4.1.1 **Tříparametrické lognormální rozdělení**

Tabulka č. 41 – Parametry lognormálního rozdělení.

$\mu$	12,50728
$\sigma^2$	0,170781
$\gamma$	-44905,5

Graf č. 91

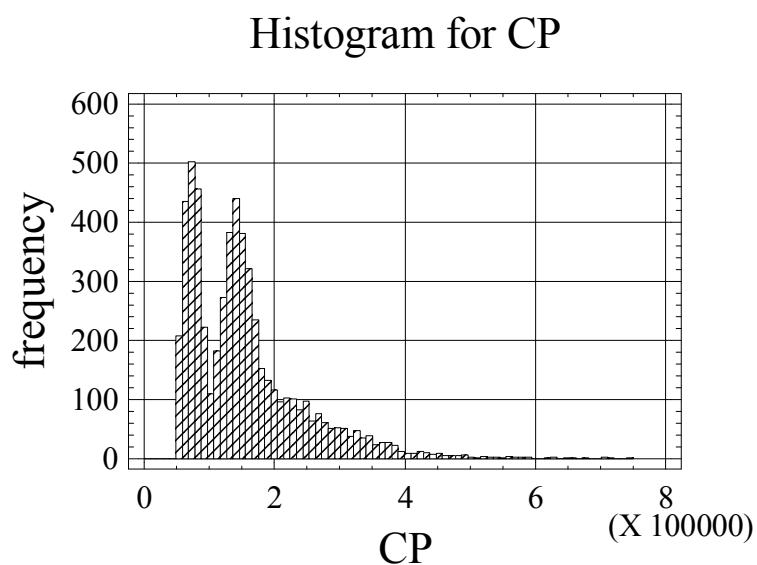


U pracujících důchodců vyšlo jako nejkvalitnější tříparametrické lognormální rozdělení. Shodu lze najít v úseku od patnáctého percentilu do devátého decilu. I teoretický průměr, který je 249 487 Sk je velmi blízký skutečně naměřenému (253 753 Sk).

#### 5.8.4.2 *Důchodci – nepracující*

Rozdělení nepracujících důchodců je dvouvrcholové. Proto zde neodpovídá žádný teoretický model, tak jak je vidět v příloze č. 18.

Graf č. 92



Proto jsem se rozhodl dále rozdělit tuto skupinu na jednočlenné a vícečlenné domácnosti nepracujících důchodců. U vícečlenných domácností se nepodařilo najít vhodný model.

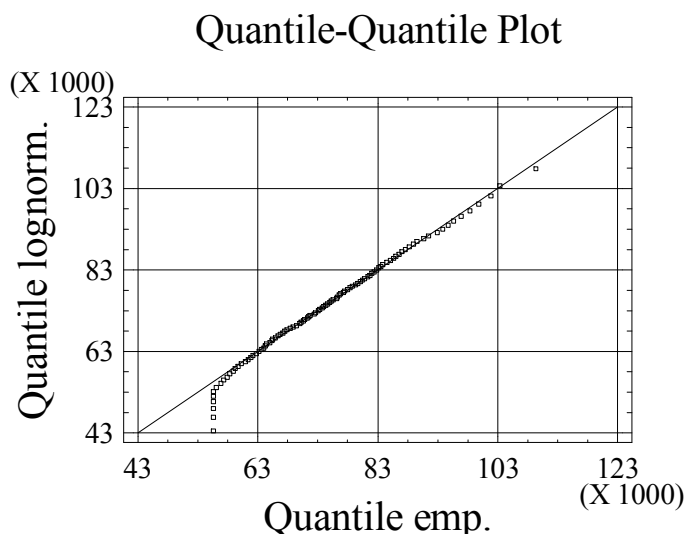
Q-Q grafy jsou uvedeny v příloze č. 19. U jednočlenných domácností je možné považovat za vhodné tříparametrické lognormální rozdělení, jak je uvedeno níže.

#### 5.8.4.2.1 Tříparametrické lognormální rozdělení

Tabulka č. 42 – Parametry lognormálního rozdělení.

$\mu$	12,62926
$\sigma^2$	0,002051
$\gamma$	-231398

Graf č. 93



Jak je vidět z předchozího grafu, tříparametrické lognormální rozdělení je možné považovat pro skupinu jednočlenných domácností nepracujících důchodců v intervalu od prvního decilu až po úplný konec. To je velmi velká oblast.

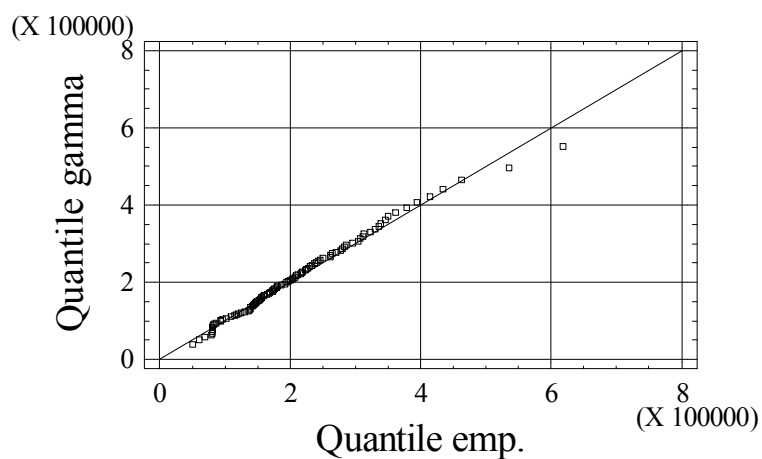
## 5.8.5 Ostatní

Tabulka č. 43 – Parametry rozdělení a jejich kvantily

tvar rozdělení	empirické	Gama	Lognorm.	Paretovo
parametry rozdělení	x	3,67376	0,558983	2,993265
	x	1,72637E-05	11,6864	141708,6
	x	x	67365,41	x
kvantily	x	x	x	x
0,01	50520	39865	88257	142185
0,02	61308	49903	92980	142668
0,03	69450	57188	96516	143158
0,04	78584	63168	99494	143654
0,05	80160	68362	102139	144158
0,25	139200	131396	139200	155315
0,5	186309	196483	186309	176309
0,75	264312	277897	264312	216816
0,95	414190	422019	474208	316786
0,96	433990	440720	507708	329500
0,97	462328	464412	552695	344532
0,98	535442	497113	619691	362740
0,99	618825	551488	744552	385521

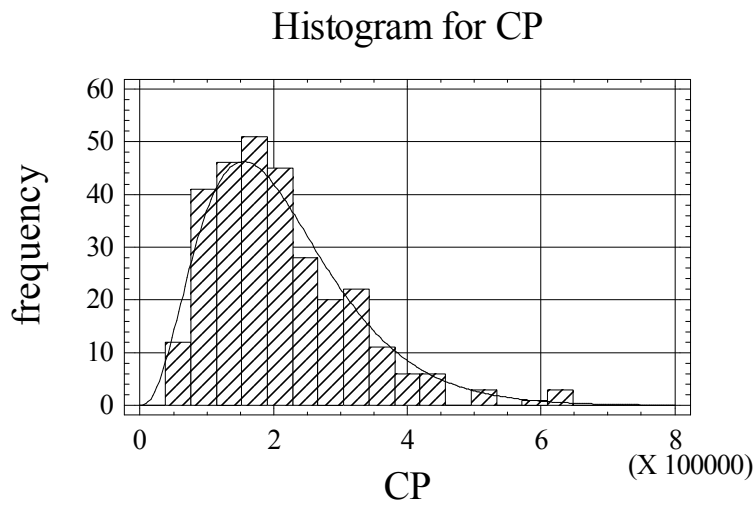
Graf č. 94

### Quantile-Quantile Plot

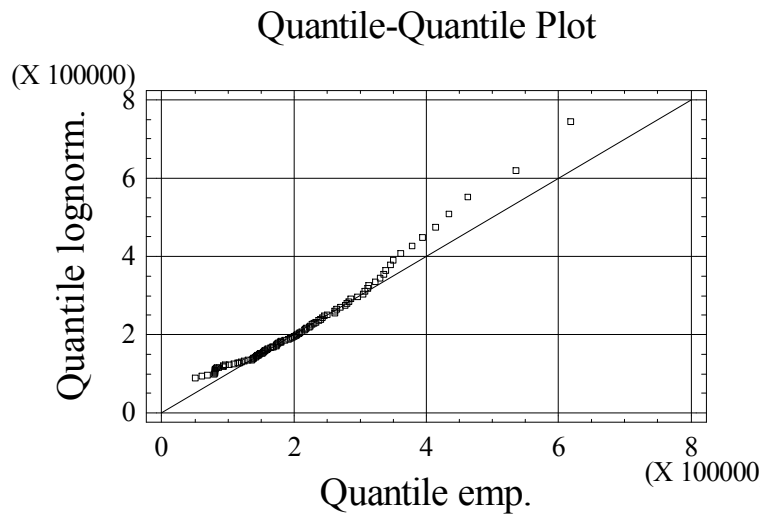




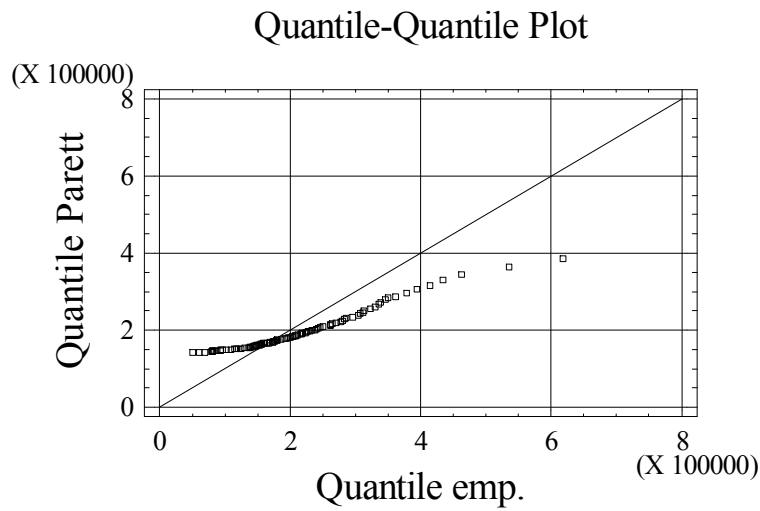
Graf č. 95



Graf č. 96



Graf č. 97



I u skupiny ostatních je nejvíce vyhovující tříparametrické lognormální rozdělení a to v části od druhého do osmého decilu. V pravém i v levém konci jsou teoretické kvantily nadhodnoceny. Paretovo rozdělení opět nevyhovuje v celé části. Z grafu č. 94 by se zdálo, že gama rozdělení odpovídá ve velké části, ale po důkladnější analýze je možné toto vyvrátit. Nejde zde najít dostatečně velkou oblast, kde by se teoretické a empirické kvantily shodovaly, vždy jde pouze o malé úseky.

## 5.9 *Silc 2005*

Pro data z roku 2005 již nebudu dělat tak důkladné analýzy, jako tomu bylo u dat z Mikrocenzů. Z předchozích analýz je možné shrnout, že u modelování celkových příjmů vícečlenných domácností je vhodné lognormální a gama rozdělení, zatímco u jednočlenných domácností a u modelování příjmů na jednoho člena domácnosti je nejvhodnější dvouparametrické Paretovo rozdělení. V následující části tedy pouze ověřím tuto skutečnost. Bohužel není možné modelovat jednotlivé skupiny ve SR jak jsem původně chtěl. Důvodem je neporovnatelnost se šetřením Mikrocenzu SR z roku 2003. Zatím co u českých dat je tato porovnatelnost zajištěna, u slovenských dat tomu tak z nepochopitelných důvodů není. Přitom by šetření Silc mělo být v jednotlivých státech EU stejné.

### 5.9.1 **Modelování celkových příjmů domácností v ČR podle počtu členů domácnosti za rok 2005**

#### 5.9.1.1 *Vícečlenné domácnosti*

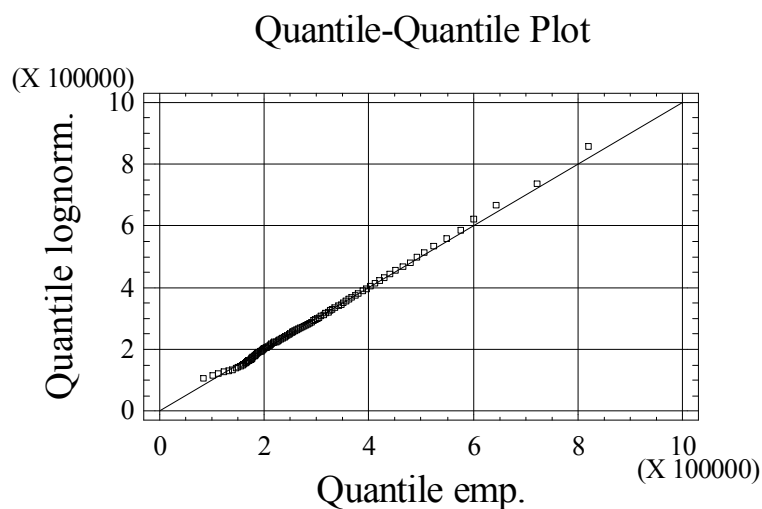
##### 5.9.1.1.1 **Tříparametrické lognormální rozdělení**

Tabulka č. 44 – Parametry lognormálního rozdělení.

$\mu$	12,16852
$\sigma^2$	0,372976
$\gamma$	60197,8

V následujících analýzách již nebudu uvádět tabulky srovnání empirického a teoretického rozdělení, jelikož grafické srovnání považuji za dostačující.

Graf č. 98



Předchozí graf potvrdil vhodnost tříparametrického lognormálního rozdělení pro vícečlenné domácnosti v České republice. Jak tomu bylo ve všech předchozích analýzách, kde žádný teoretický model nevyhovoval dolní části rozdělení, tak ani zde tomu není jinak. Za shodnou lze považovat část od druhého do devátého decilu. I průměr lognormálního rozdělení (292 317 Kč) je velmi blízký průměru skutečnému (293 305 Kč).

### 5.9.1.2 *Jednočlenné domácnosti*

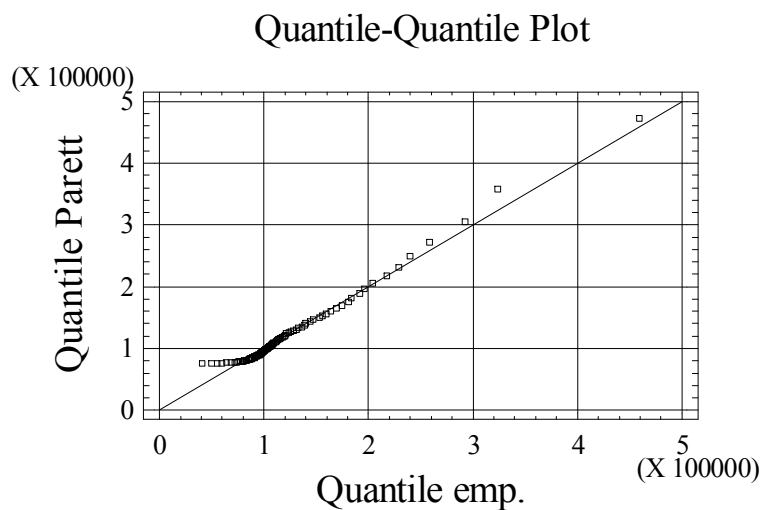
#### 5.9.1.2.1 **Dvouparametrické Pareto rozdělení**

Tabulka č. 45 – Parametry Pareto rozdělení

C	2,507912
$x_0$	75313,42

Vhodnost Pareto rozdělení je možné považovat od prvního do devátého decilu s tím, že v rozmezí od druhého do pátého decilu jsou v kvantilech trochu větší odchylky. Proto ještě sestrojím i tříparametrické lognormální rozdělení.

Graf .č 99

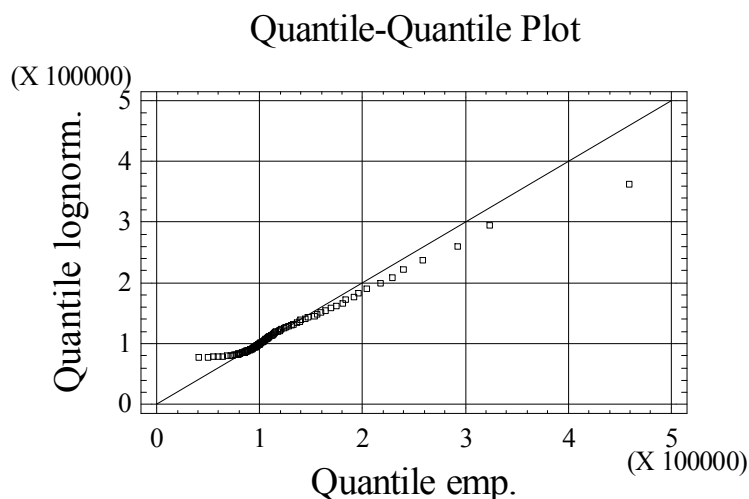


### 5.9.1.2.2 Tříparametrické lognormální rozdělení

Tabulka č. 46 – Parametry lognormálního rozdělení.

$\mu$	10,28501
$\sigma^2$	0,968758
$\gamma$	73709,98

Graf č. 100



Z grafu je vidět, že v dolní a horní části neodpovídají teoretické kvantily empirickým. Zatímco v dolní části jsou nadhodnoceny, v horní je tomu naopak. Shodu lze najít od druhého do osmého decilu, což je menší část než u Paretova rozdělení, ale jednotlivé kvantily jsou v intervalu od druhého do pátého decilu shodnější, než u Paretova rozdělení.

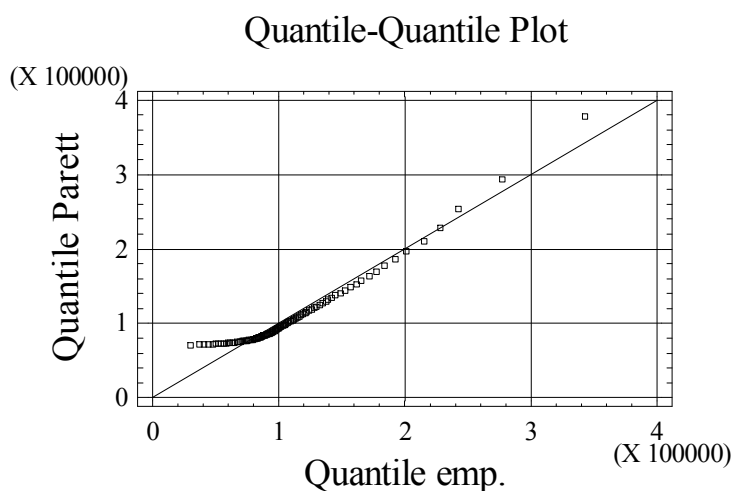
## 5.9.2 Modelování příjmů na člena domácnosti v ČR za rok 2005

### 5.9.2.1 Dvouparametrické Paretovo rozdělení

Tabulka č. 47 – Parametry Paretova rozdělení

C	2,74444
$x_0$	70571,55

Graf č. 101



Jak je vidět v grafu č. 101, Paretovo dvouparametrické rozdělení se opět ukázalo jako vyhovující pro příjmy na jednoho člena domácnosti. Opět neodpovídá skutečným datům v horní a dolní části rozdělení, kde je nadhodnocuje.

## 5.9.3 Modelování celkových příjmů domácností v SR podle počtu členů domácnosti za rok 2005

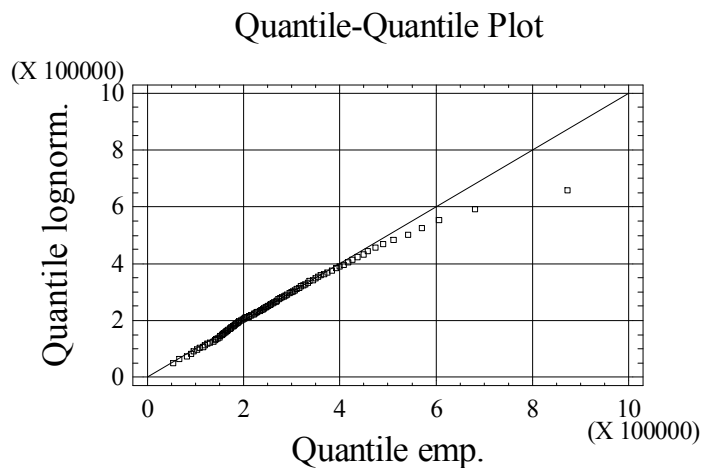
### 5.9.3.1 Vícečlenné domácnosti

#### 5.9.3.1.1 Tříparametrické lognormální rozdělení

Tabulka č. 48 – Parametry lognormálního rozdělení.

$\mu$	12,81136
$\sigma^2$	0,106387
$\gamma$	-123758

Graf č. 102



Analýza na slovenských datech opět potvrdila vhodnost tříparametrického lognormálního rozdělení v jeho velké části, konkrétně jde o část od druhého do osmého decilu. To je menší část než tomu bylo u českých dat z roku 2005. I průměry jsou zde vzdálenější. Skutečný průměr má hodnotu 273 520 Sk, teoretický pak 262 613 Sk.

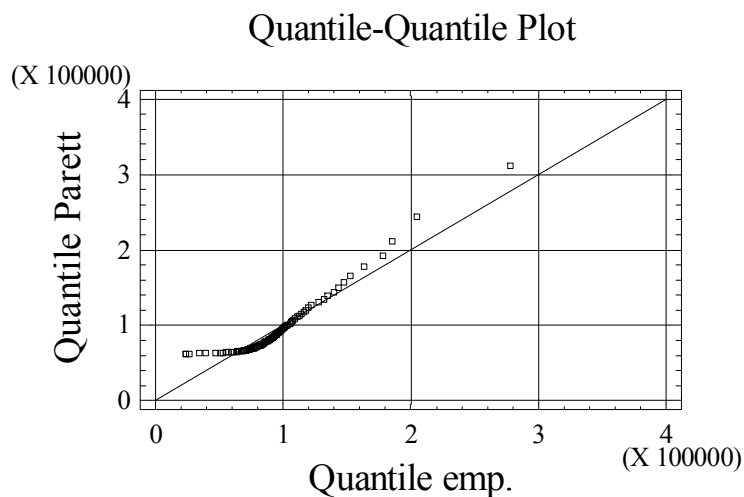
### 5.9.3.2 *Jednočlenné domácnosti*

#### 5.9.3.2.1 **Dvouparametrické Paretovo rozdělení**

Tabulka č. 49 – Parametry Paretova rozdělení

C	2,842928
$x_0$	61755,97

Graf č. 103



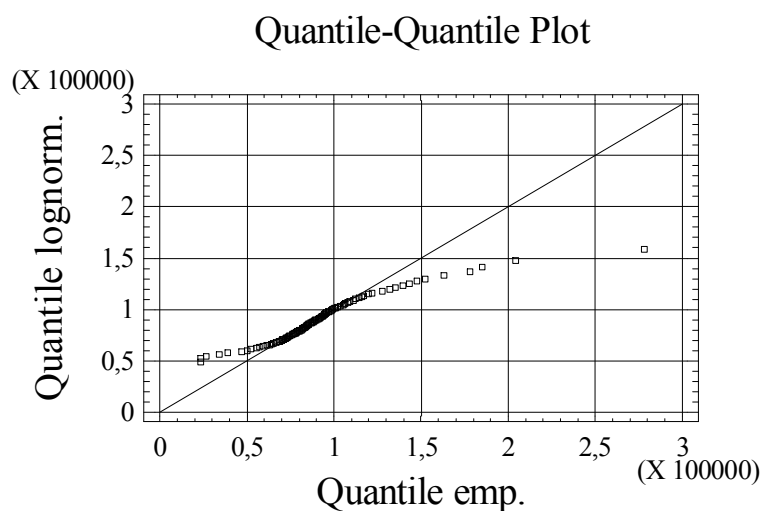
Bohužel jak je vidět na grafu č. 103, Paretovo rozdělení nemodeluje kvalitně skutečně naměřená data ve SR. Tak tomu ale bylo i u slovenských dat z roku 2003. Proto ještě vyzkouším tříparametrické lognormální rozdělení.

### 5.9.3.2.2 Tříparametrické lognormální rozdělení

Tabulka č. 50 – Parametry lognormálního rozdělení.

$\mu$	11,30597
$\sigma^2$	0,073977
$\gamma$	5949,882

Graf č. 104



Bohužel ani tříparametrické lognormální rozdělení není vhodné. Ve středu rozdělení je sice možné najít shodu, ale v koncích jsou až moc výrazné rozdíly.

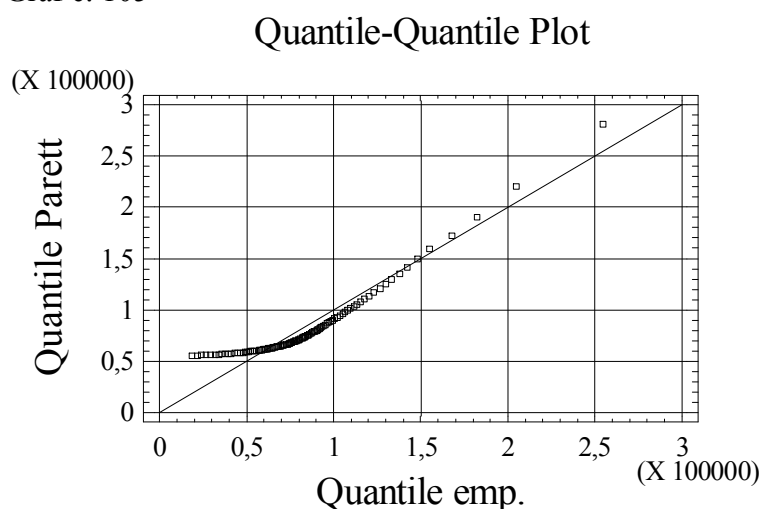
## 5.9.4 Modelování příjmů na člena domácnosti v SR za rok 2005

### 5.9.4.1 Dvouparametrické Pareto rozdělení

Tabulka č. 51 – Parametry Pareto rozdělení

C	2,835841
$x_0$	55369,34

Graf č. 105



Jak je vidět z předchozího grafu, bohužel se neprokázala vhodnost dvouparametrického Pareto rozdělení pro příjmy na člena domácnosti v SR za rok 2005. Proto ještě sestrojím tříparametrické lognormální rozdělení, které se dalo považovat v roce 2002 vedle Pareto tvaru za vhodné.

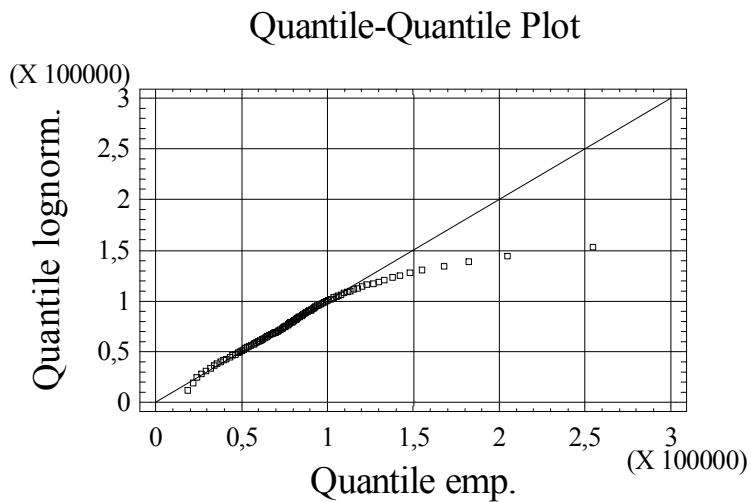
### 5.9.4.2 Tříparametrické lognormální rozdělení

Tabulka č. 52 – Parametry lognormálního rozdělení.

$\mu$	13,60223
$\sigma^2$	0,001397
$\gamma$	-728510



Graf č. 106



Tříparametrické lognormální rozdělení se nakonec ukázalo u příjmů na člena domácnosti v SR v roce 2005 jako vhodné od druhého do osmdesátého percentilu. U žádné jiné skupiny v předchozích analýzách nevyhovoval žádný model v levé krajní části tak, jak je tomu zde. Pokud se však podíváme na druhý konec, zjistíme, že v posledním decilu teoretický model výrazně podhodnocuje skutečně naměřená data. To se pak odráží v průměru, kdy teoretický je 79 983 Sk a skutečný 85 531 Sk.

# Závěr

Ve své diplomové práci zaměřené na modelování životní úrovně domácností v Čechách a na Slovensku jsem nejdříve v teoretické části nastínil situaci v oblasti příjmů po roce 1989. Transformace na plně tržní ekonomický systém s sebou přinesla první známky nerovnoměrného rozdělování důchodů. To samozřejmě přináší narušení stability dříve platných statistických modelů a z nich plynoucích závěrů. Proto je potřeba ověřit jejich platnost a v případě neshody hledat vhodnější modely. Pro nalezení vhodného modelu jsem se rozhodl využít distribuční funkce. Využil jsem však i hodnoty kvantilové funkce, a to zejména při zjišťování vhodnosti modelu v podobě Q-Q grafů. Na závěr teoretické části jsem uvedl základní údaje a metodiku ke statistickým šetřením, které byly vstupními daty pro praktickou část.

Ve své praktické části jsem se pokusil najít vhodné modely pro rozdělení celkových příjmů domácností v Čechách a na Slovensku, příjmů domácností přepočtených na jednu osobu v Čechách a na Slovensku a pro příjmy domácností různých sociálních skupin na Slovensku.

Před samotným hledáním vhodných teoretických modelů jsem provedl důkladnou analýzu empirického rozdělení. V této analýze jsem použil i méně známých kvantilových grafů a charakteristik, které přinesly zajímavé výsledky.

Jak u celkových příjmů domácností, tak u příjmů přepočtených na jednoho člena domácnosti v obou zemích má empirické rozdělení velmi krátký levý konec oproti konci pravému. Krátký levý konec lze zdůvodnit minimální výší mzdy a sociálními příspěvky státu osobám v těžké sociální situaci. Pravý konec je velmi dlouhý a v jeho samotném konci jsou velké diference. Do budoucna se dá očekávat zvyšování diferenciace mezd a tudíž prodlužování pravého konce, což má pak za následek neshodu empirického rozdělení s pravděpodobnostním modelem především na jeho okrajích.

U obou výše zmíněných skupin v obou zemích můžeme na základě momentových ukazatelů říci, že jde o kladně zešikmené rozdělení. To znamená, že převažují domácnosti s menšími příjmy oproti domácnostem s příjmy vyššími. Momentový ukazatel poukazuje na výrazně

špičatější rozdělení oproti normovanému normálnímu tvaru. Tak jak se však později ukázalo, tyto hodnoty výrazně ovlivňují odlehlé příjmy. Proto bych u analyzování příjmů doporučoval použít kvantilové ukazatele – Galtonův koeficient šikmosti a Moorsův koeficient špičatosti.

Po důkladné analýze empirického rozdělení jsem začal hledat vhodné teoretické tvary pro různé skupiny obyvatelstva. Pro posouzení vhodnosti jsem použil  $\chi^2$  test dobré shody a grafických analýz v podobě histogramů s křivkou teoretického rozdělení a zejména Q-Q grafů.  $\chi^2$  test dobré shody se později ukázal jako ne příliš vhodný. Podle tohoto testu by žádný teoretický tvar nemohl být přijat, jelikož hodnoty testu výrazně překračovaly kritické hodnoty. To je však možné zdůvodnit velmi vysokým počtem dat, kdy pak tento test odhalí i velmi malé nesrovnalosti a výsledná hodnota testu je velmi vysoká. Proto bych opět spíše doporučil využívat Q-Q grafů.

Ze všech typů rozdělení, které jsem ve své práci použil, bych jako vhodné označil dvouparametrické gama rozdělení, dvouparametrické Paretovo rozdělení a tříparametrické lognormální rozdělení. U posledně dvou zmiňovaných jsem si ověřil, jak přidání jednoho parametru výrazně zkvalitní model.

Jelikož rozdělení celkových příjmů domácností je dvouvrcholové, bylo potřeba vytvořit detailnější skupiny. Východiskem bylo rozdělení na jednočlenné a vícečlenné domácnosti. U vícečlenných domácností je možné považovat ve všech případech za vhodné tříparametrické lognormální rozdělení. U jednočlenných je možné považovat za vhodné dvouparametrické Paretovo rozdělení, avšak pouze v České Republice. U slovenských dat se nepodařilo najít vhodný tvar.

U modelování příjmů na jednoho člena domácnosti se jako nejlepší ukázalo dvouparametrické Paretovo rozdělení. Pouze v jediném případě, a to u příjmů ve Slovenské republice v roce 2005 se ukázalo jako vhodnější tříparametrické lognormální rozdělení. Je tedy vidět, že dvouparametrické Paretovo rozdělení je vhodné tam, kde je empirické rozdělení užší, tedy tam, kde nejsou velké odchylky příjmů. Právě v nalezení vhodnosti Paretova rozdělení u těchto skupin bych viděl velký přínos mé práce, jelikož v mnoha publikacích se píše o vhodnosti tohoto tvaru zejména v koncích rozdělení.

U slovenských dat jsem se pokusil najít vhodné modely i pro jednotlivé sociální skupiny. Skupin jsem vytvořil pět a to: zaměstnaní, samostatně činní, nezaměstnaní, důchodci, ostatní. Bohužel jsem zde nemohl pak ověřit závěry na datech z roku 2005. Důvodem je neporovnatelnost se šetřením Mikrocenzu SR z roku 2003. Zatím co u českých dat je tato porovnatelnost zajištěna, u slovenských dat tomu tak z nepochopitelných důvodů není. Přitom by šetření Silc mělo být v jednotlivých státech EU stejné.

U skupiny zaměstnanců se ukázalo jako nejvhodnější gama rozdělení. U nezaměstnaných a ostatních je to tříparametrické lognormální rozdělení. U skupiny důchodců bylo z důvodu dvou vrcholů nutné vytvořit podskupiny. To jsem vyřešil rozdělením na důchodce pracující a důchodce nepracující. U pracujících je nejvhodnější tříparametrické rozdělení. U nepracujících bylo opět nutné vytvořit podskupiny z důvodu dvou vrcholů. Příjmy jednočlenných domácností je možné ve velké části namodelovat tříparametrickým lognormálním rozdělením. Pro vícečlenné domácnosti se bohužel nepodařilo najít vhodný model. Stejně tak se nepodařilo najít vhodný model u skupiny samostatně činných.

Věřím, že se mi podařilo ve své práci dosáhnout všech cílů, které jsem si stanovil, a že má práce bude přínosná, jelikož pravděpodobnostní modely jsou jednoduchou aproximací složitého empirického rozdělení. Proto je znalost pravděpodobnostním modelů důležitá např. při odhadu spotřeby v budoucnosti, nebo zjišťování odbytových možností pro výroby dlouhodobé i krátkodobé spotřeby.

# Literatura

## *Knihy*

- [1] ČERMÁK, V.: *Diskrétní a spojitá rozdělení, vzorce, grafy, tabulky*, Vysoká škola ekonomická, Praha 1992, ISBN: 80-7079-711-8
- [2] VEČERNÍK, J.: *Občan a tržní ekonomika, příjmy, nerovnost a politické postoje v české společnosti*, Nakladatelství Lidové noviny, 1998, ISBN: 80-7106-255-9
- [3] HINDLS, R., HRONOVÁ, S., SEGER, J.: *Statistika pro ekonomy, 7. vydání*, Vysoká škola ekonomická, Professional Publishing, Praha 2006, ISBN: 80-86946-16-9
- [4] SIPKOVA, L., SODOMOVÁ, E.: *Modelovanie kvantilovými funkciami*, Ekonomická univerzita v Bratislavě, Vydavateľství Ekonom 2007, ISBN: 978-80-225-2346-2
- [5] BARTOŠOVÁ, J.: *Základy statistiky pro manažery*, Vysoká škola ekonomická, nakladatelství Oeconomica, 2006, ISBN: 80-245-1019-7
- [6] ŠAUEROVÁ, A.: *Vědecký časopis Vysoké školy ekonomické v Praze 1/2007 – Statistické a matematické metody v ekonomii*, Oeconomica, 2007, ISSN: 0572-3043
- [7] ARTLOVÁ, M., BÍLKOVÁ, D., JAROŠOVÁ, D., POUROVÁ, Z.: *Příklady k předmětu statistika A*, Praha, Oeconomica. 2004, ISBN 80-245-0178-3
- [8] BARTOŠOVÁ, J.: *Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990*, Disertační práce. Praha: Fakulta informatiky a statistiky VŠE v Praze 2006

## *Bakalářské práce*

- [9] ICHA, P.: *Vývoj rozdělení příjmů obyvatelstva v ČR*, 2005

## *Internetové zdroje*

<http://www.czso.cz>

<http://www.statistics.sk>

# Seznam příloh

**Příloha č. 1** - Kvantilové charakteristiky příjmů domácností v ČR za rok 2002

**Příloha č. 2** - Krabičkový diagram příjmů domácností v ČR za rok 2002

**Příloha č. 3** - Detailnější Q-Q grafy gama rozdělení pro příjmy domácností v ČR

**Příloha č. 4** - Kvantilové charakteristiky příjmů domácností v SR za rok 2003

**Příloha č. 5** - Krabičkový diagram příjmů domácností v SR za rok 2003

**Příloha č. 6** - Distribuční funkce příjmů domácností v SR za rok 2003 v prvním a posledním kvartilu

**Příloha č. 7** - Kvantilové charakteristiky příjmů na člena domácnosti v ČR za rok 2002

**Příloha č. 8** - Krabičkový diagram příjmů na člena domácnosti v ČR za rok 2002

**Příloha č. 9** - Detailnější Q-Q grafy dvouparametrického lognormálního rozdělení pro příjmy na člena domácnosti v ČR za rok 2002

**Příloha č. 10** - Detailnější Q-Q graf gama rozdělení pro příjmy na člena domácnosti v ČR za rok 2002

**Příloha č. 11** - Kvantilové charakteristiky příjmů na člena domácnosti v SR za rok 2003

**Příloha č. 12** - Krabičkový diagram příjmů na člena domácnosti v SR za rok 2003

**Příloha č. 13** - Detailnější Q-Q grafy dvouparametrického lognormálního rozdělení pro příjmy na člena domácnosti v SR za rok 2003

**Příloha č. 14** - Detailnější Q-Q grafy gama rozdělení pro příjmy na člena domácnosti v SR za rok 2003

**Příloha č. 15** - Detailnější Q-Q grafy tříparametrického lognormálního rozdělení pro příjmy domácností v SR za rok 2003

**Příloha č. 16** - Detailnější Q-Q graf tříparametrického lognormálního rozdělení pro příjmy na člena domácnosti v ČR za rok 2002

**Příloha č. 17** - Detailnější Q-Q graf tříparametrického lognormálního rozdělení pro příjmy na člena domácnosti v SR za rok 2003

**Příloha č. 18** – Q-Q grafy tříparametrického lognormálního a Paretova rozdělení pro příjmy domácností nepracujících důchodců v SR za rok 2003

**Příloha č. 19** – Q-Q grafy tříparametrického lognormálního a Paretova rozdělení pro příjmy vícečlenných domácností nepracujících důchodců v SR za rok 2003

# Přílohy

Příloha č. 1

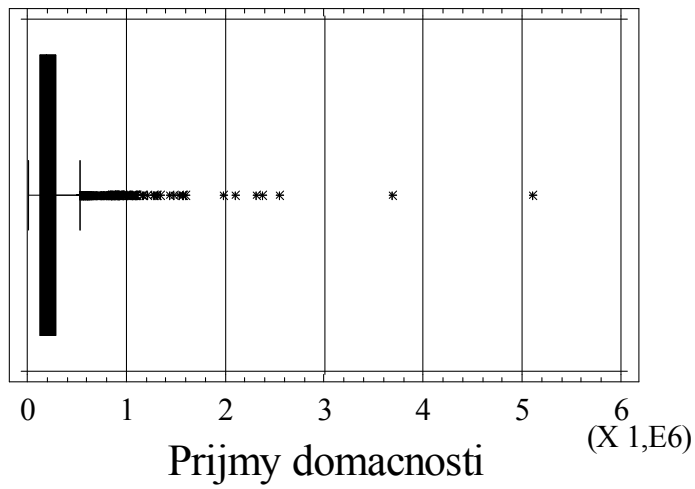
Výběrové charakteristiky	značení	hodnota
<b>polohy:</b>		
<b>minimální hodnota</b>	s	13500
<b>maximální hodnota</b>	l	5110628
<b>dolní kvartil</b>	$Q_{0,25}$	126023,000
<b>medián</b>	$Q_{0,5}$	190923,000
<b>horní kvartil</b>	$Q_{0,75}$	288343,000
<b>variability:</b>		
<b>variační rozpětí</b>	R	5097128
<b>kvartilové rozpětí</b>	iqr	162320,000
<b>kvartilová odchylka</b>	qo	81160
<b>horní kvartilová diference</b>	uqd	97420,000
<b>dolní kvartilová diference</b>	lqd	64900,000
<b>dolní p-diference</b>	ld(p)	
<b>p=</b>	0,01	133572,480
	0,05	114123,000
	0,1	104525,400
	0,125	99723,000
	0,375	29353,125
<b>horní p-diference</b>	udp(p)	
<b>1-p=</b>	0,99	584642,830
	0,95	305043,850
	0,9	211208,800
	0,875	183526,250
	0,625	42666,125
<b>kvantilové rozpětí</b>	ipr(p)	
<b>p=</b>	0,01	718215,310
	0,05	419166,850
	0,1	315734,200
	0,125	283249,250
	0,375	72019,250
<b>kvantilová odchylka</b>	qo(p)	
<b>p=</b>	0,01	359107,655



	0,05	209583,425
	0,1	157867,100
	0,125	141624,625
	0,375	36009,625
<b>šikmosti:</b>		
<b>kvartilová diference</b>	qd	32520
<b>kvantilová diference</b>	pd(p)	
<b>p=</b>	0,01	451070,35
	0,05	188520,85
	0,1	106683,4
	0,125	82711,5
	0,375	13313
<b>Galtonův koeficient šikmosti</b>	g	0,200345
<b>Galtonova p-šikmost</b>	g(p)	
<b>p=</b>	0,01	2,7788957
	0,05	1,1614148
	0,1	0,6572413
	0,125	0,5095583
	0,375	0,082017
<b>galtnův p-index šikmosti</b>	g*(p)	
<b>p=</b>	0,01	0,6280434
	0,05	0,4497513
	0,1	0,3378899
	0,125	0,2920096
	0,375	0,1848534
<b>kvantilový podíl</b>	sr(p)	
<b>p=</b>	0,01	4,3769707
	0,05	2,6729393
	0,1	2,0206457
	0,125	1,8403603
	0,375	1,4535463
<b>špičatosti:</b>		
<b>Moorsova špičatost</b>	k	0,8185338

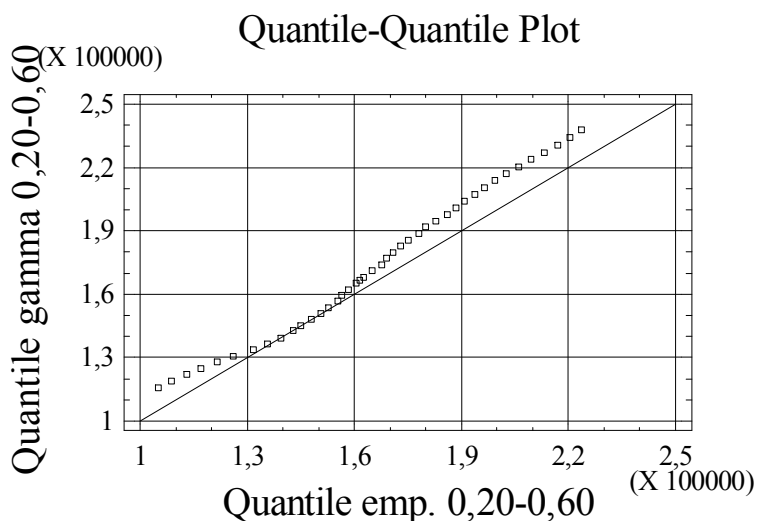
Příloha č. 2

### Box-and-Whisker Plot

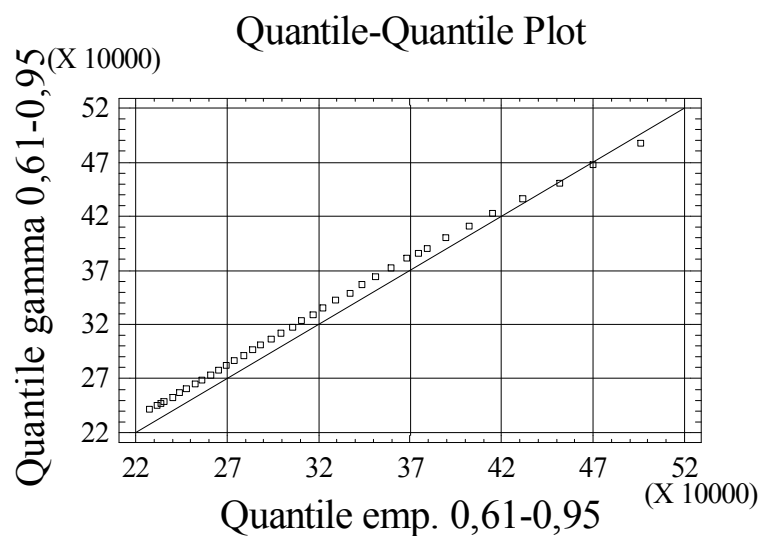


Příloha č. 3

Graf č. 14.1



Graf č. 14.2



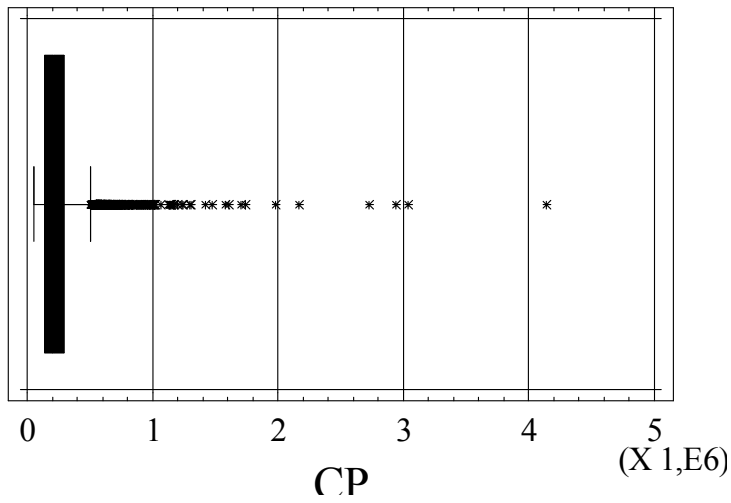
Příloha č. 4

Výběrové charakteristiky	značení	
<b>polohy:</b>		
<b>minimální hodnota</b>	s	50520
<b>maximální hodnota</b>	l	4134788
<b>dolní kvartil</b>	$Q_{0,25}$	143130,750
<b>medián</b>	$Q_{0,5}$	207096,000
<b>horní kvartil</b>	$Q_{0,75}$	289035,000
<b>variability:</b>		
<b>variační rozpětí</b>	R	4084268
<b>kvartilové rozpětí</b>	iqr	145904,250
<b>kvartilová odchylka</b>	qo	72952,125
<b>horní kvartilová diference</b>	uqd	81939,000
<b>dolní kvartilová diference</b>	lqd	63965,250
<b>dolní p-diference</b>	ld(p)	
<b>p=</b>	0,01	151524,000
	0,05	136896,000
	0,1	122524,800
	0,125	111128,780
	0,375	35037,750
<b>horní p-diference</b>	udp(p)	
<b>1-p=</b>	0,99	432891,850
	0,95	239305,600

	0,9	171346,800
	0,875	148909,375
	0,625	38024,000
<b>kvantilové rozpětí</b>	$ipr(p)$	
<b>p=</b>	0,01	584415,850
	0,05	376201,600
	0,1	293871,600
	0,125	260038,155
	0,375	73061,750
<b>kvantilová odchylka</b>	$qo(p)$	
<b>p=</b>	0,01	292207,925
	0,05	188100,800
	0,1	146935,800
	0,125	130019,078
	0,375	36530,875
<b>šikmosti:</b>		
<b>kvartilová diference</b>	$qd$	17973,75
<b>kvantilová diference</b>	$pd(p)$	
<b>p=</b>	0,01	281367,85
	0,05	98629,6
	0,1	48822
	0,125	35411,625
	0,375	2986,25
<b>Galtonův koeficient šikmosti</b>	$g$	0,1231887
<b>Galtonova p-šikmost</b>	$g(p)$	
<b>p=</b>	0,01	1,9284418
	0,05	0,6759885
	0,1	0,3346167
	0,125	0,2427045
	0,375	0,0204672
<b>galtnův p-index šikmosti</b>	$g^*(p)$	
<b>p=</b>	0,01	0,4814514
	0,05	0,2621722
	0,1	0,1661338
	0,125	0,1361786
	0,375	0,040873
<b>špičatosti:</b>		
<b>Moorsova špičatost</b>	$k$	0,9013116

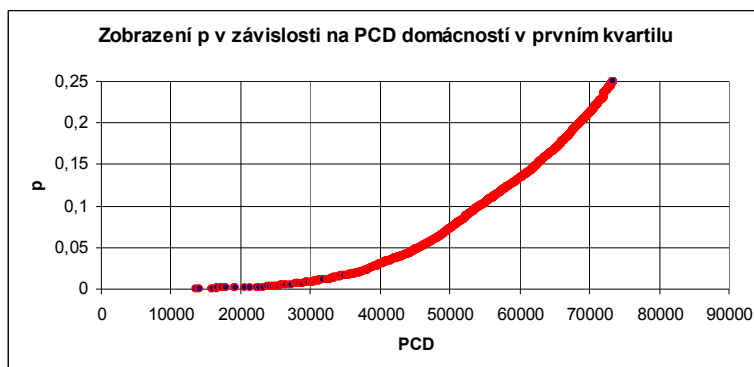
Příloha č. 5

## Box-and-Whisker Plot

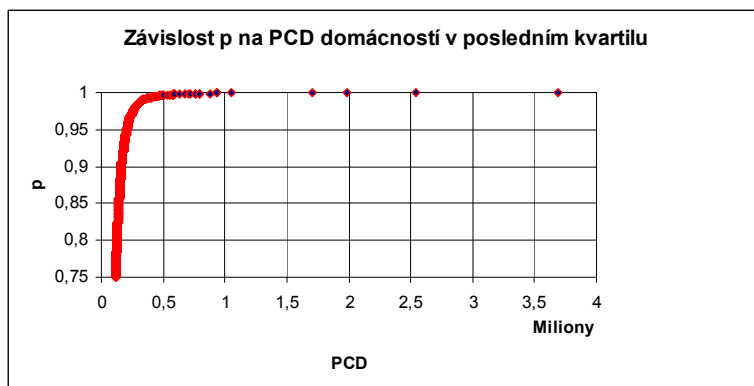


Příloha č. 6

Graf č. 30.1



Graf č. 30.2



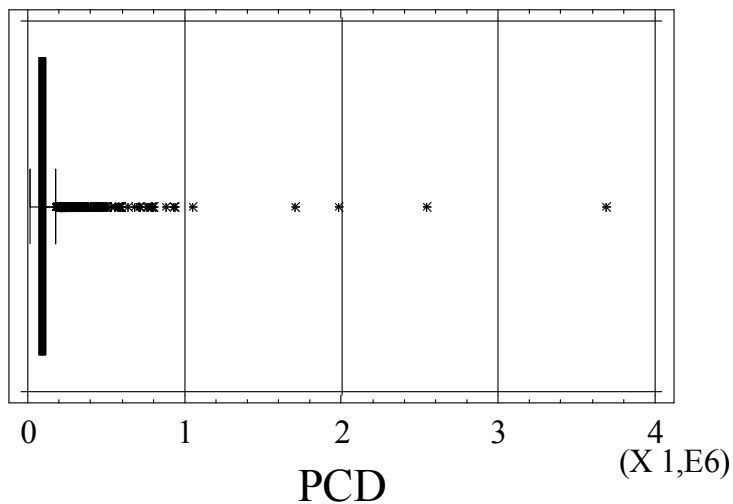
## Příloha č. 7

Výběrové charakteristiky	značení	
<b>polohy:</b>		
<b>minimální hodnota</b>	s	13500
<b>maximální hodnota</b>	l	3690914
<b>dolní kvartil</b>	$Q_{0,25}$	73458,000
<b>medián</b>	$Q_{0,5}$	89204,000
<b>horní kvartil</b>	$Q_{0,75}$	115966,563
<b>variability:</b>		
<b>variační rozpětí</b>	R	3677414
<b>kvartilové rozpětí</b>	iqr	42508,563
<b>kvartilová odchylka</b>	qo	21254,281
<b>horní kvartilová diference</b>	uqd	26762,563
<b>dolní kvartilová diference</b>	lqd	15746,000
<b>dolní p-diference</b>	ld(p)	
<b>p=</b>	0,01	58018,686
	0,05	43447,325
	0,1	34848,520
	0,125	29667,158
	0,375	7604,000
<b>horní p-difference</b>	udp(p)	
<b>p=</b>	0,99	244909,235
	0,95	117467,525
	0,9	74113,400
	0,875	63087,875
	0,625	9427,250
<b>kvantilové rozpětí</b>	ipr(p)	
<b>p=</b>	0,01	302927,921
	0,05	160914,850
	0,1	108961,920
	0,125	92755,033
	0,375	17031,250
<b>kvantilová odchylka</b>	qo(p)	
<b>p=</b>	0,01	151463,961
	0,05	80457,425
	0,1	54480,960
	0,125	46377,516
	0,375	8515,625

<b>šikmosti:</b>		
<b>kvartilová diference</b>	qd	11016,563
<b>kvantilová diference</b>	pd(p)	
<b>p=</b>	0,01	186890,55
	0,05	71468,565
	0,1	39264,88
	0,125	32576,406
	0,375	1823,25
<b>Galtonův koeficient šikmosti</b>	g	0,259161
<b>Galtonova p-šikmost</b>	g(p)	
<b>p=</b>	0,01	4,3965389
	0,05	1,6812746
	0,1	0,9236934
	0,125	0,7663493
	0,375	0,0428914
<b>galtnův p-index šikmosti</b>	g*(p)	
<b>p=</b>	0,01	0,6169473
	0,05	0,444139
	0,1	0,3603541
	0,125	0,351209
	0,375	0,1070532
<b>špičatosti:</b>		
<b>Moorsova špičatost</b>	k	0,971706

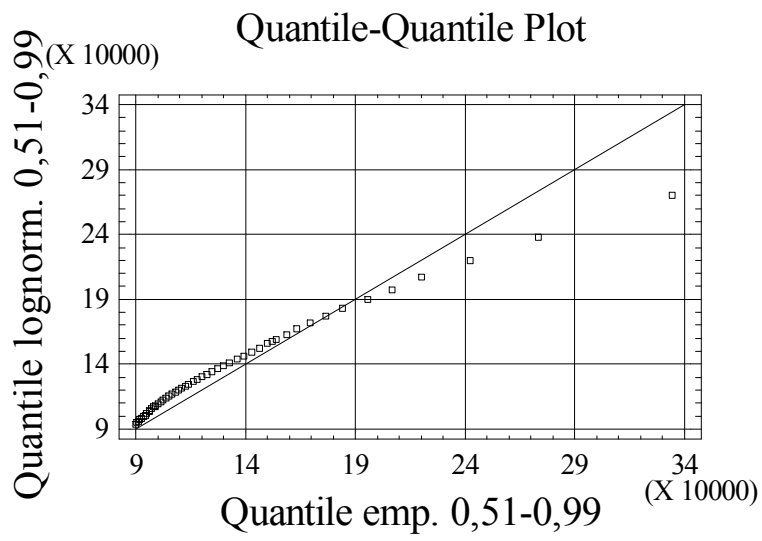
Příloha č. 8

## Box-and-Whisker Plot

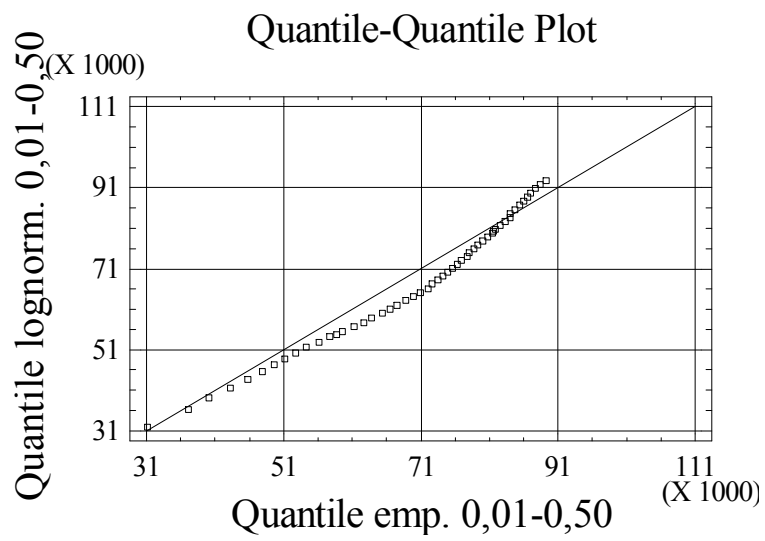


Příloha č. 9

Graf č. 37.1



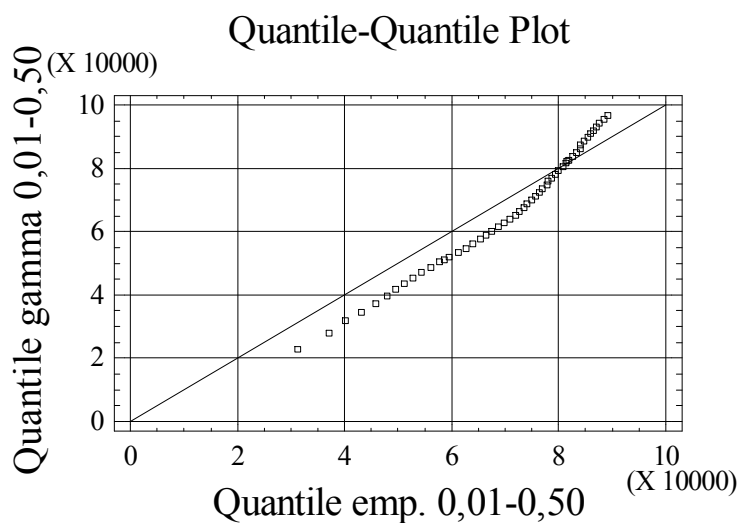
Graf č. 37.2





Příloha č. 10

Graf č. 41.1



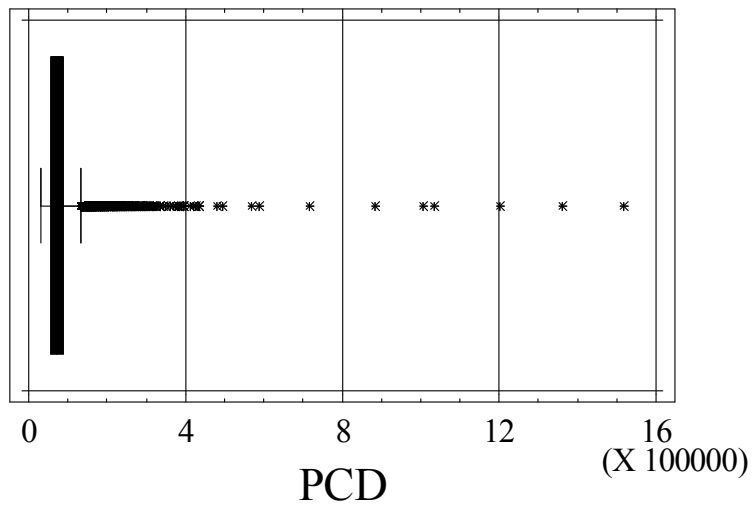
Příloha č. 11

<b>Výběrové charakteristiky</b>	<b>značení</b>	
<b>polohy:</b>		
<b>minimální hodnota</b>	s	32355
<b>maximální hodnota</b>	l	1518704,5
<b>dolní kvartil</b>	$Q_{0,25}$	56409,625
<b>medián</b>	$Q_{0,5}$	70920,000
<b>horní kvartil</b>	$Q_{0,75}$	86879,400
<b>variability:</b>		
<b>variační rozpětí</b>	R	1486349,5
<b>kvartilové rozpětí</b>	iqr	30469,775
<b>kvartilová odchylka</b>	qo	15234,888
<b>horní kvartilová diference</b>	uqd	15959,400
<b>dolní kvartilová diference</b>	lqd	14510,375
<b>dolní p-diference</b>	ld(p)	
<b>p=</b>	0,01	34848,000
	0,05	30733,050
	0,1	25906,440
	0,125	23131,066
	0,375	7012,167
<b>horní p-diference</b>	udp(p)	
<b>1-p=</b>	0,99	133080,000

	0,95	62913,467
	0,9	41060,160
	0,875	34317,156
	0,625	6905,625
<b>kvantilové rozpětí</b>	ipr(p)	
<b>p=</b>	0,01	167928,000
	0,05	93646,517
	0,1	66966,600
	0,125	57448,222
	0,375	13917,792
<b>kvantilová odchylka</b>	qo(p)	
<b>p=</b>	0,01	83964,000
	0,05	46823,258
	0,1	33483,300
	0,125	28724,111
	0,375	6958,896
<b>šikmosti:</b>		
<b>kvartilová diference</b>	qd	1449,025
<b>kvantilová diference</b>	pd(p)	
<b>p=</b>	0,01	98232
	0,05	31113,99
	0,1	15153,72
	0,125	10814,773
	0,375	-106,54167
<b>Galtonův koeficient šikmosti</b>	g	0,0475561
<b>Galtonova p-šikmost</b>	g(p)	
<b>p=</b>	0,01	3,2239162
	0,05	1,0211427
	0,1	0,4973361
	0,125	0,3549345
	0,375	-0,0034966
<b>galtnův p-index šikmosti</b>	g*(p)	
<b>p=</b>	0,01	0,584965
	0,05	0,3322493
	0,1	0,2262877
	0,125	0,1882525
	0,375	-0,0076551
<b>špičatosti:</b>		
<b>Moorsova špičatost</b>	k	0,8639263

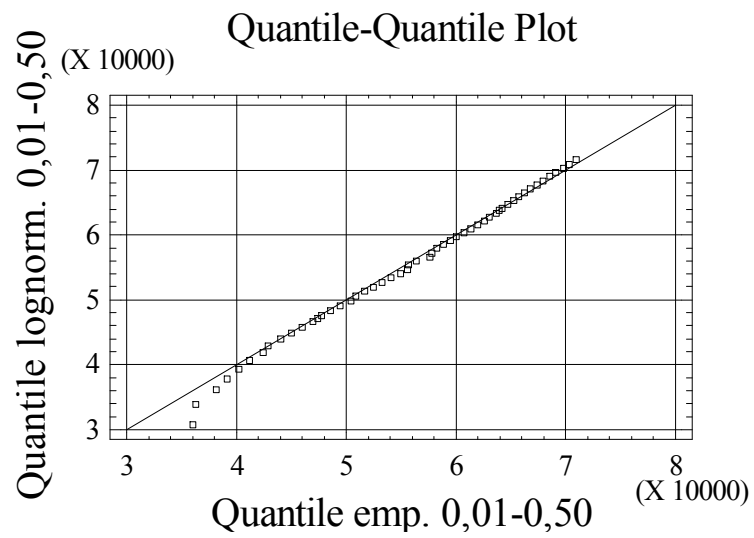
Příloha č. 12

### Box-and-Whisker Plot

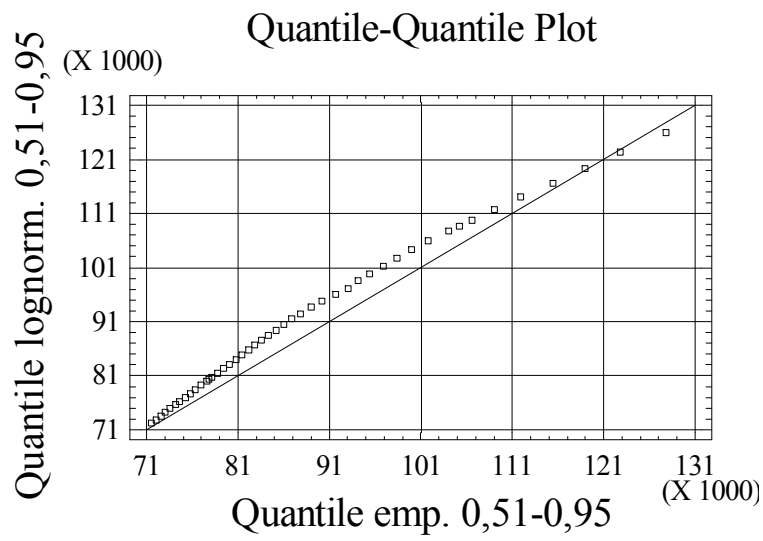


Příloha č. 13

Graf č. 53.1

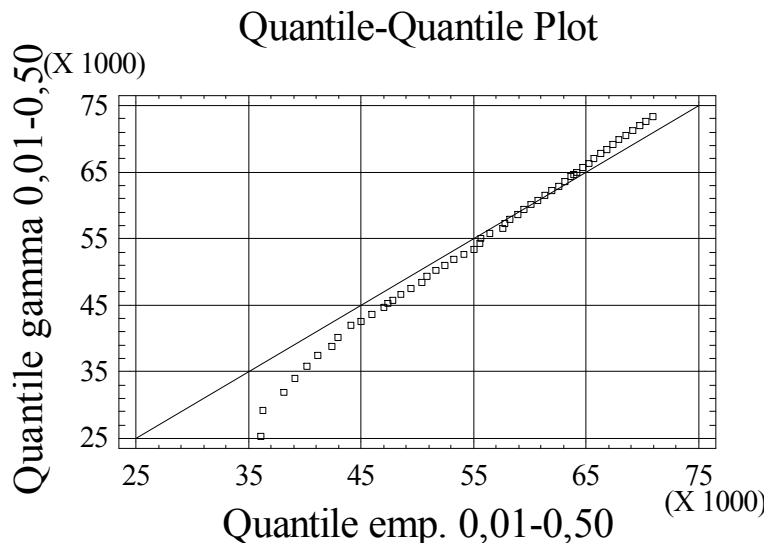


Graf č. 53.2

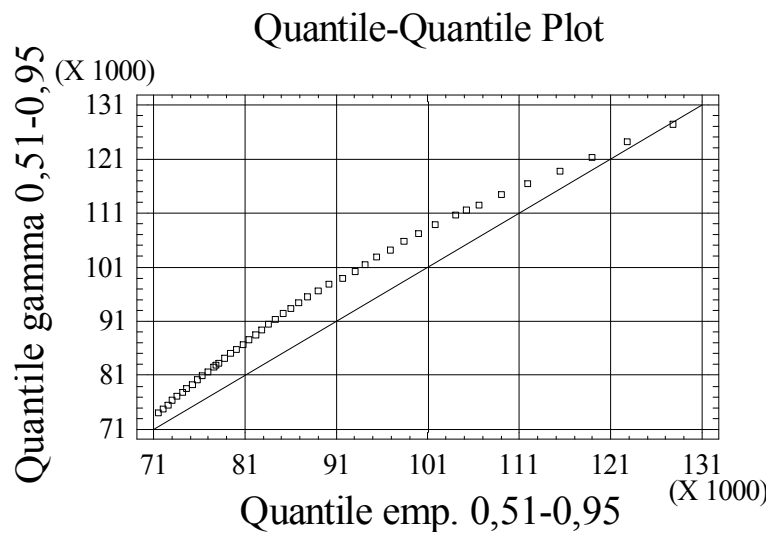


Příloha č. 14

Graf č. 57.1

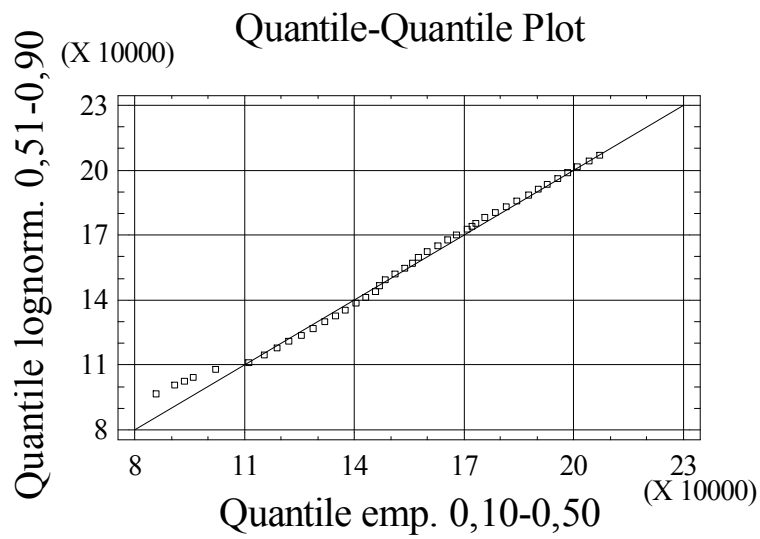


Graf č. 57.2

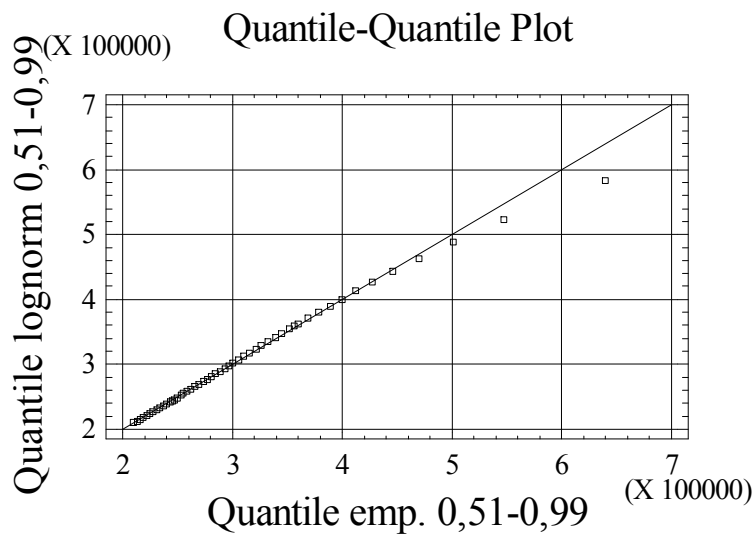


Příloha č. 15

Graf č. 62.1

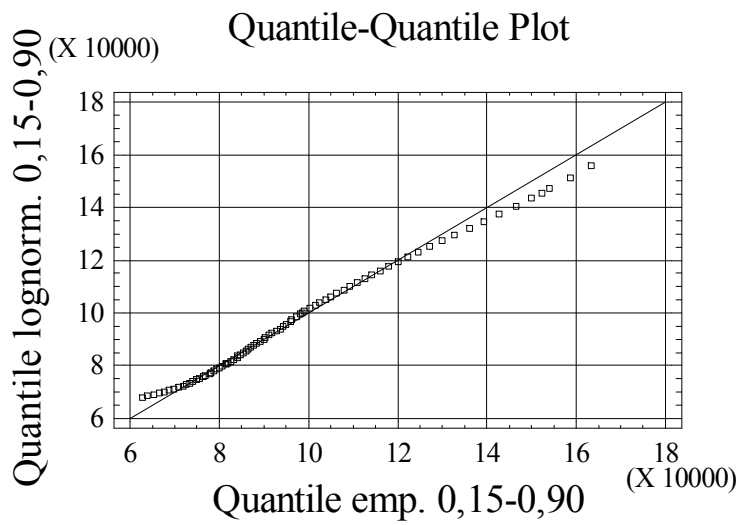


Graf č. 62.2



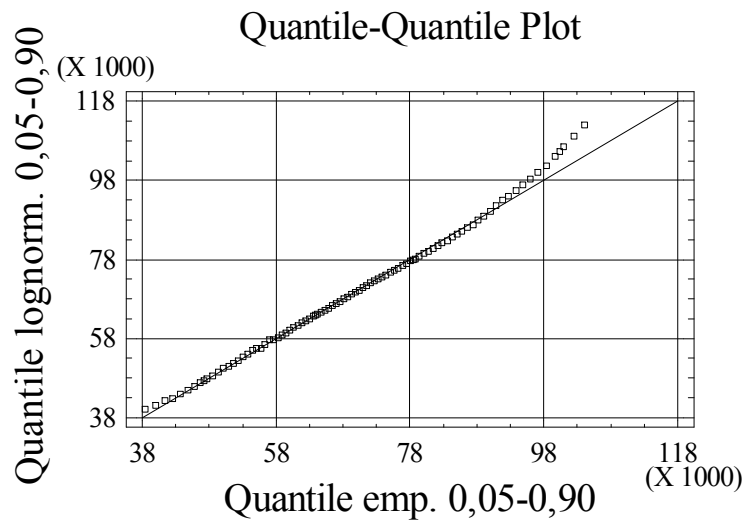
Příloha č. 16

Graf č. 63.1



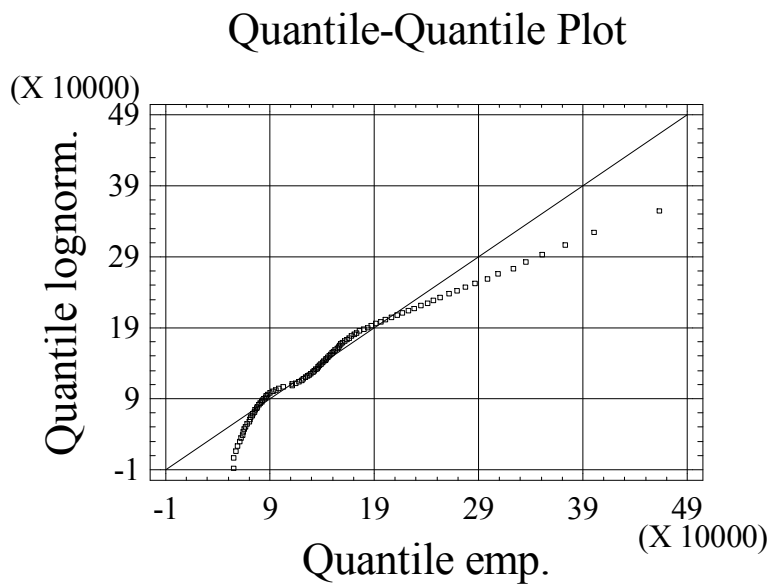
Příloha č. 17

Graf č. 64.1

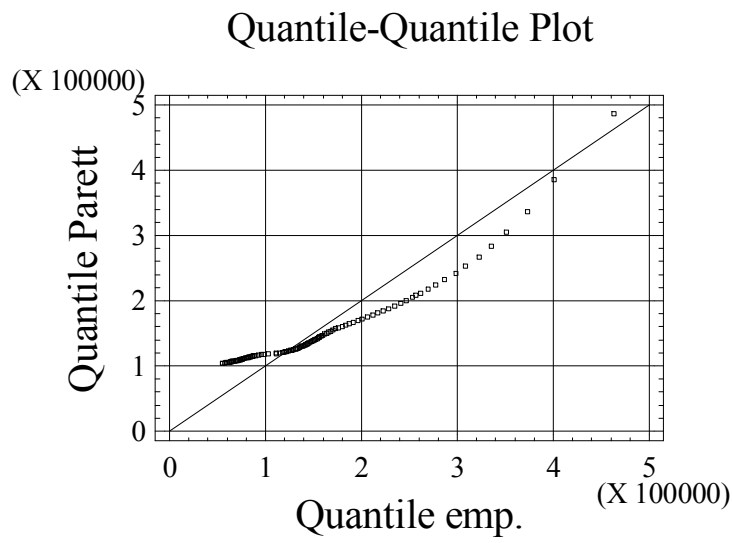


Příloha č. 18

Graf č. 92.1

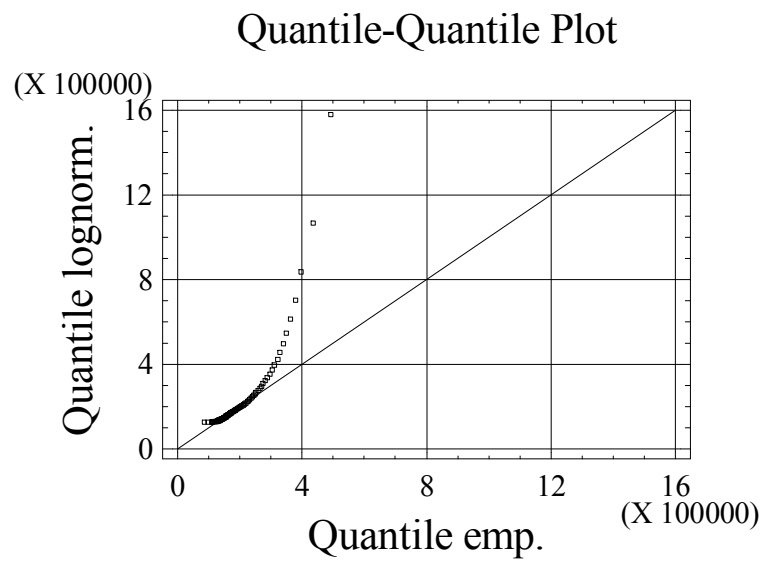


Graf č. 92.2



Příloha č. 19

Graf č. 92.3





Graf č. 92.4

### Quantile-Quantile Plot

