



Vysoká škola ekonomická v Praze

Fakulta managementu v Jindřichově Hradci

Diplomová práce

2008

Martin Myšička

Vysoká škola ekonomická v Praze

Fakulta managementu

Jindřichův Hradec

Diplomová práce

Martin Myšička

2008



Vysoká škola ekonomická v Praze

Fakulta managementu v Jindřichově Hradci

Katedra managementu informací

Analýza sociální situace domácností v České republice

Vypracoval:

Martin Myšička

Vedoucí diplomové práce:

RNDr. Jitka Bartošová, Ph.D.

Praha, duben 2008

Prohlášení

Prohlašuji, že diplomovou práci na téma
» **Analýza sociální situace domácností v České republice**«
jsem vypracoval samostatně.

Použitou literaturu a podkladové materiály
uvádím v příloženém seznamu literatury.

Praha, duben 2008

podpis studenta

Anotace

Analýza sociální situace domácností v České republice

Diplomová práce je zaměřena na podrobnou analýzu příjmů českých domácností v závislosti na různých společensko-ekonomických a demografických faktorech (oporou práce budou informace zveřejňované Českým statistickým úřadem na webových stránkách).

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
A TEORETICKO – METODOLOGICKÁ ČÁST	3
1 ZÁKLADNÍ POPULACE, CENZUS A VÝBĚROVÝ SOUBOR DAT	3
2 DATA	4
3 ROZDĚLENÍ PŘÍJMŮ DOMÁCNOSTÍ	10
4 MODEL ROZDĚLENÍ PŘÍJMŮ	15
5 ODHAD CHARAKTERISTIK EMPIRICKÉHO ROZDĚLENÍ A PARAMETRŮ JEHO TEORETICKÉHO MODELU	18
6 TESTOVÁNÍ SHODY MODELU S EMPIRICKÝM ROZDĚLENÍM	21
7 REGRESNÍ ANALÝZA	23
8 POUŽITÝ SOFTWARE	28
B PRAKTICKÁ ČÁST	30
1 VÝBĚROVÉ SOUBORY	30
2 ČLENĚNÍ PODLE POČTU EKONOMICKY AKTIVNÍCH ČLENŮ DOMÁCNOSTI	32
3 ČLENĚNÍ DOMÁCNOSTÍ PODLE POČTU NEZAOPATŘENÝCH DĚTÍ	37
4 DOMÁCNOSTI ČLENĚNÉ DLE VZDĚLÁNÍ OSOBY V ČELE	43
5 ČLENĚNÍ VÝBĚROVÉHO SOUBORU DO SKUPIN PODLE KRAJŮ	49
6 ČLENĚNÍ VÝBĚROVÉHO SOUBORU PODLE POHLAVÍ OSOBY V ČELE	57
7 ČLENĚNÍ DOMÁCNOSTÍ PODLE SOCIÁLNÍ SKUPINY OSOBY V ČELE	60
8 DOMÁCNOSTI DŮCHODCŮ BEZ EKONOMICKY AKTIVNÍHO ČLENA	66
9 ANALÝZA ZÁVISLOSTÍ	71
ZÁVĚR	75
SEZNAM POUŽITÉ A PROSTUDOVANÉ LITERATURY	77
PŘÍLOHY	78

Úvod

Příjmy domácností patří všude na světě k těm nejpodstatnějším statistickým ukazatelům. Často se síla ekonomiky porovnává pomocí velikosti HDP, deficitu/přebytku obchodní bilance nebo jiných, jistě také velmi důležitých ukazatelů. O skutečném bohatství občanů, žijících společně v jednom státu (regionu, společenství) však mnohem více vypovídá právě výše peněžních příjmů domácností. Tento fakt velice dobře pociťují právě občané Česka. Příjmová konvergence k úrovni zemí západní Evropy zaostává za přibližováním nejen ekonomického výkonu (HDP), ale také bohužel cenové hladiny. Neméně důležitá je také struktura rozdělení příjmů mezi jednotlivé skupiny obyvatel. Aproximací takového rozdělení vhodným teoretickým modelem získáváme nástroj, který nejen vypovídá o celkové úrovni distribuce bohatství, může také sloužit dalším analýzám. Pro státní správu mohou být takovéto modely podkladem pro nastavení parametrů daňového zatížení občanů nebo určování výše sociálních dávek a podpor. Pro ziskové i neziskové organizace soukromého sektoru zase znamenají podporu při zavádění výrobků speciálně zacílených na vybrané skupiny obyvatel či při expanzi do nových regionů.

Ve své diplomové práci jsem si vytyčil dva hlavní cíle; prvním je testování vhodnosti logaritmicko-normálního rozdělení jako teoretické aproximace rozdělení příjmu domácnosti v České republice (dále již jen ČR) na empiricky zjištěných datech roku 2002 a 2005. Druhým cílem je pak odhad míry závislosti příjmů na skupině sociologických, ekonomických, společenských a demografických faktorech.

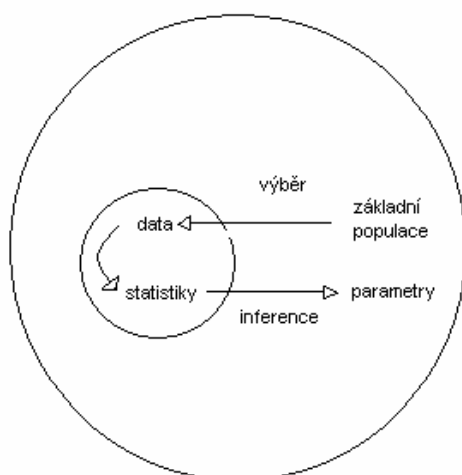
V rámci první části této práce je zpracován detailní popis charakteristik empirického rozdělení příjmů, jehož podkladem se stala dvě výběrová šetření provedená Českým statistickým úřadem (dále jen ČSÚ), metodik Mikrocensus pro rok 2002, resp. SILC v případě roku 2005. Druhým dílčím úkolem je volba modelu rozdělení a odhad parametrů modelu empirického rozdělení. To bych chtěl provést vhodnou metodou, zajišťující co možná nejlepší výsledky. V závěrečné části se pokusím srovnat obě rozdělení, tj. teoretické a empirické, a posoudit vhodnost zvolené aproximace. Ve své práci se chci zaměřit nejen na modelování příjmů celé populace, ale také na zjišťování vhodnosti použití výše uvedených teoretických rozdělení i pro jednotlivé subpopulace. Prostředkem k dosažení tohoto cíle jsou právě data z Mikrocenzu 2002 a SILCu 2005, kde je vedle výše příjmu zjišťováno nejméně 40 dalších společenských, demografických a sociálních faktorů. Pro účely této práce byl počet faktorů, a z toho vyplývajících podskupin snížen na šest, a to s důrazem na výběr těch nejdůležitějších.

Druhým cílem mé diplomové práce je odhad míry závislosti příjmů domácností v ČR v uvedených dvou letech na vybraných faktorech. Chtěl bych k tomu použít metodu regresní analýzy s využitím modelu s umělými proměnnými. Ten umožňuje transformaci kvalitativních proměnných na proměnné binární. Příjmy každé domácnosti ovlivňuje celá řada proměnných, jako vzdělání, počet dětí, sociální dávky apod. Cílem je redukovat původní počet faktorů na ty nejpodstatnější.

A Teoreticko – metodologická část

1 Základní populace, cenzus a výběrový soubor dat

Základní populací rozumíme množinu všech objektů, které můžeme teoreticky uvažovat jako zdroj dat pro testování statistických závislostí. V tomto případě je základní populací množina všech domácností v ČR. Cenzus je pak taková metoda sběru dat, kdy se do zkoumaných objektů zahrnou všechny jednotky základní populace, tj. provádí se totální výběr. Ten se používá při sčítání lidu domů.¹ Takový to projekt je velice nákladný, časově náročný a složitý na provedení, a proto se praxi pro potřebu statistických zjišťování používá výběru (vzorku), který je podmnožinou základní populace, jak dokresluje obr. 1.



obrázek 1: Výběrový soubor. Zdroj: Hendl, Jan: Přehled statistických metod zpracování dat

Výběr má z důvodu úspory nákladů na provedení sběru dat mnohem menší rozsah než totální výběr². Přesto je možné výsledky takového výzkumu za předpokladu dodržení stanovených postupů použít pro odhad základních charakteristik (průměru, mediánu, apod.) celé populace. Vedle velikosti výběrového souboru hraje rozhodující roli také jeho reprezentativnost³.

¹ Jedná se o vyčerpávající (úplné) šetření sociálních, demografických nebo ekonomických charakteristik domácností; provádí se jednou za deset let, přičemž poslední sčítání proběhlo v roce 2001

² Obvykle mezi 0,1 – 2%, ale např. u průzkumů veřejných mínění bývají vzorky ještě podstatně menší; zde bývá kladen důraz na jejich reprezentativnost

³ Reprezentativnost náhodného výběru znamená, že každý jeden prvek základního souboru měl stejnou pravděpodobnost, že bude vybrán. Pokud tento princip není dodržen, můžeme i při relativně velkém výběrovém vzorku docházet k mylným závěrům.

2 Data

Mikrocenzus - metodika

Od konce 50. let 20. stol. probíhalo v Československu, resp. v České republice výběrové šetření obyvatelstva za účelem zjištění jejich příjmového rozdělení technikou tzv. Mikrocenzů. Tato šetření se prováděla jednou za 3 až 5 let. Do roku 1989 probíhala na vybraném souboru 1 až 2 % domácností, informace byly získány od státních orgánů, státních podniků, popř. poštovních orgánů (důchody). Po sametové revoluci se uskutečnil Mikrocenzus ještě třikrát, a to v letech 1992 (Československo), 1996 a 2002 (pouze ČR). Změnou oproti dřívějším výběrovým zjišťováním bylo zmenšení počtu šetřených domácností na 0,25% - 1%⁴ a vedle toho také změna v získávání údajů, kdy byly získávány dotazováním přímo v jednotlivých domácnostech. Po vstupu země do Evropské Unie (dále již jen EU) v roce 2004 harmonizovala ČR legislativu v oblasti statistiky s příslušnými zákony EU a nahradila nepravdělná šetření prostřednictvím Mikrocenzů každoročním zjišťováním metodou SILC (viz. text níže), přičemž první data pocházejí z roku 2005.

Mikrocenzus používá výběrové šetření prováděné na základě dvoustupňového výběru. Tento systém třídění zjišťovaných subjektů se používá, pokud je základní soubor příliš velký a prostorově rozptýlený. V těchto případech by pouhý prostý náhodný výběr nedokázal zabránit nad- či podhodnocování některých skutečností, což by následně vedlo ke zkreslení celkových výsledků. Postup je tedy takový, že ze základního souboru vybereme nejprve tzv. primární jednotky (např. obce), a v druhém kole ve vybraných primárních jednotkách vybereme tzv. sekundární jednotky (např. domácnosti)⁵.

Vysvětlivky k základním údajům

Základním členěním subjektů je třídění na tzv. hospodařící domácnosti. Definicí tohoto pojmu se rozumí dobrovolné prohlášení osob bydlících ve vybraném bytě, že společně žijí, hospodaří, tj. hradí výdaje za stravu, ubytování apod.⁶ Pro účel této práce bylo vybráno těchto šest faktorů (z celkového počtu 40 zjišťovaných)⁷:

⁴ 1992: 0,5%; 1996: 1%; 2002: 0.25%

⁵ Hindls – Hronová – Seger, Statistika pro ekonomy (2004)

⁶ Mikrocensus 2002

⁷ Podrobnější členění sledovaných znaků viz. příloha

- počet ekonomicky aktivních pracujících členů,
- kraj,
- subpopulace členěny podle počtu nezaopatřených (závislých) dětí,
- pohlaví osoby v čele,
- vzdělání osoby v čele,
- sociální skupina osoby v čele.

Mikrocensus 2002

Výběr probíhal na základě informací z registru sčítacích obvodů⁸ (SO), a to pro každý kraj nezávisle, aby bylo dosaženo rovnoměrného rozdělení. Obvody s méně než 24 byty nebyly do výběru zařazeny. Nejprve bylo vybráno metodou znáhodněného systematického výběru s pravděpodobnostmi zahrnutí přímo úměrnými počtu trvale obydlených bytů 50% z počtu plánovaných SO, což činilo celkem 460. Ke každému byl následně vybrán ještě jeden SO ze stejné sídelní jednotky, příp. katastru (pokud na vybraném územním celku existoval pouze jeden SO). V jednotlivých SO pak byl proveden v druhém kole prostý náhodný výběr 12 bytů.

Údaje získané šetřením Mikrocenzu 2002 představují údaje za tzv. hospodařící domácnosti. Ta vychází z dobrovolného prohlášení osob užívajících společně vybraný byt. Jedná se o skupinu lidí, kteří se společně podílejí na hrazení základních výdajů domácnosti, jako je jídlo, náklady na bydlení, služby apod. Údaje o demografických faktorech (vzdělání, rodinný stav), stejně jako údaje o peněžních a naturálních příjmech, byly zjišťovány podle stavu ke konci roku 2002. Sledované znaky, tj. ekonomická aktivita, druh zaměstnání a

⁸ Sčítací obvod je podle příslušné metodiky ČSÚ z roku 2001 „statistická územní jednotka s prvky organizačního charakteru, která zahrnuje stavební objekty vymezené číslem domovním, která se vyznačuje územní celistvostí, skladebností do základních sídelních jednotek včetně dílu, jedinečností co do příslušností objektů v daném obvodě k jediné části obce a která má kvantitativní limit, tzn. obsahuje maximálně 140 bytů nebo 400 obyvatel. Hranice sčítacího obvodu jsou odvozené od hranic základních sídelních jednotek a hranic katastrálních území, jsou skladebné do hranic základních sídelních jednotek a jsou vedeny po vlastnických hranicích, osách komunikací a dalších liniových prvcích včetně přírodních“. Na základě novely zákona o státní statistické službě roku 2006 se sčítací obvody transformují na statistické obvody; nejvýznamnější změnou bude potlačení organizačního principu (tvorby samostatných obvodů s hromadnými ubytovacími kapacitami nebo tvorba více obvodů v jedné budově v závislosti na počtu bytů).

odvětví se posuzovaly podle převažujícího stavu⁹. V případě, že osoba ukončila v průběhu roku 2002 školní vzdělání, zaznamenával se stav ve 2. pololetí.¹⁰

V případě, že ve společné domácnosti žila některá osoba jen část sledovaného období (z důvodů stěhování, narození, delší nepřítomnost daná pobytem v zahraničí, vykonávání základní vojenské služby či výkonu trestu), byla pak do počtu osob žijících v domácnosti započítána jen z části.

V úplných rodinách¹¹ se jako osoba v čele domácnosti bere vždy muž, a to bez ohledu na jeho ekonomickou aktivitu. Žádnou roli tedy nehrálo, zda jeho příjmy tvoří opravdu větší (příp. alespoň podstatnější než u partnerky) část rodinných příjmů, či dokonce zda nebyl nezaměstnaný. V neúplných rodinách (tj. v domácnostech s jedním rodičem s dětmi) a v nerodinných domácnostech se posuzovala osoba v čele na základě její ekonomické aktivity popř. výše příjmu.

Šetření Mikrocensus 2002 bylo provedeno na 11 040 bytech v České republice, což představuje přibližně 0,25% z celkového počtu všech trvale obydlených bytů. Z tohoto počtu se ukázalo 351 bytů (tj. 3,2%) jako nebydlených. Ve zbylých 10 689 jednotkách bylo dosaženo těchto výsledků¹²

Tabulka 1: Úspěšnost sběru dat Mikrocensus 2002 (Zdroj: www.czso.cz)

Počet vyšetřených bytových domácností	7678	71,78%
Počet nevyšetřených bytových domácností	3011	28,22%
Celkem. Z toho:		100,00%
domácnost nezastižena		22,90%
cizí státní příslušník (jazyková bariéra)		1,30%
odmítnutí z objektivních příčin (vysoký věk, zdravotní nebo rodinné problémy)		7,20%
odmítnutí zúčastnit se šetření		68,60%

V jednotlivých krajích se lišila úspěšnost vyšetření jednotek zhruba o +/- 10%. Nejnižšího procenta vyšetření bylo dosaženo u domácností v Praze (61,9%), naopak nejlepšími výsledky se prezentovali tazatelé v Karlovarském kraji (81,3%). Ve výsledcích bylo také oproti původním předpokladům zastoupeno více domácností s členy v důchodovém věku, byla zjištěna také nižší průměrná velikost domácnosti, než jakou přinesl výsledek sčítání lidu, bytů a domů (SLBD), provedený v roce 2001. Z těchto důvodů není možné

⁹ V případě rovnosti podle stavu k 31.12.2002

¹⁰ Zdroj: Metodické vysvětlivky Mikrocensus 2002

¹¹ Manželé, popř. druh - družka

¹² Zdroj: Mikrocensus 2002

provést přepočítání na celou populaci pomocí koeficientů, které poměří počet vyšetřených domácností v kraji s jeho celkovým počtem obyvatel. Přesné hodnoty uvádí tabulka 2.

Tabulka 2: Úspěšnost vyšetření v krajích, Mikrocensus 2002 (Zdroj: www.czso.cz)

Kraj	Bytů v šetření	z toho vyšetřeno		Kraj	Bytů v šetření	z toho vyšetřeno	
		počet	%			počet	%
Hl. město Praha	2460	1523	61,9	Královeský	478	315	65,9
Středočeský	971	658	67,8	Pardubický	471	380	80,7
Jihočeský	564	430	76,2	Vysočina	479	377	78,7
Plzeňský	517	398	77	Jihomoravský	978	683	69,8
Karlovarský	428	348	81,3	Olomoucký	555	430	77,5
Ústecký	744	625	84	Zlínský	491	386	78,6
Liberecký	410	282	68,8	Moravskoslezský	1143	843	73,8

Proto byla data přepočítána pomocí iterační metody kalibrace vah, minimalizující rozdíl mezi odhadnutými a přepočítanými výběrovými charakteristikami, vybranými takto pro každý kraj zvlášť. Bylo využito těchto charakteristik:¹³

- počet trvale obydlených bytů - odhad stanovený na základě výsledků SLDB 2001 a přírůstků resp. úbytků počtu bytů za roky 2001 a 2002,
- počet osob bydlících v bytech - odvozený ze středního stavu obyvatelstva ke 30. 6. 2002 podle demografické statistiky (protože šetření podléhaly pouze osoby žijící v bytech, byly od údajů z demografie odečteny počty osob žijících v tzv. ústavních domácnostech podle údajů statistiky sociálního zabezpečení za rok 2002),
- počet důchodců (pracujících i nepracujících) odvozený z údajů Ministerstva práce a sociálních věcí a České správy sociálního zabezpečení podle stavu ke konci 1. pololetí 2002, přičemž byl odečten počet osob žijících v domovech důchodců apod.,
- počet nezaměstnaných - údaje z evidence MPSV za rok 2002 byly povýšeny odhadem neregistrované nezaměstnanosti na základě výsledků VŠPS,
- počet samostatně činných osob – odhad stanovený na základě výsledků VŠPS za rok 2002 a výsledků SLDB 2001.

V šetření také dochází k asi 10% podhodnocení příjmů, a to jednak proto, že dotazovaní si na všechny své příjmy nevzpomenou nebo mají snahu udávat nižší příjmy než odpovídají skutečnosti. Toto zkreslení se velice obtížně kvantifikuje a proto jsou korigovány

¹³ doslovná citace z dokumentu Metodika Mikrocensus 2002

po porovnání s údaji o průměrných hrubých mzdách, podobně se postupovalo v případě sociálních dávek, kde uváděné hodnoty naopak překračují skutečnost.¹⁴

SILC 2005

Po vstupu ČR do Evropské unie nahradila šetření Mikrocenzů každoroční zjišťování příjmů a životních podmínek EU-SILC¹⁵, poprvé provedená jako Životní podmínky 2005. Hlavním rozdílem oproti dříve prováděným šetřením metodikou Mikrocenzu je vedle menšího vzorku domácností především větší detailnost zjišťovaných informací a zpracování výstupů za jednotlivce. Výhodou však zůstává velmi podobná metodika výběru domácností¹⁶, i následná korekce dat přepočítáním dle metod uvedených výše¹⁷ u Mikrocenzů, včetně odhadů podcenění příjmů a eliminace chybějících příjmů. SILC 2005 byl proveden na vzorku 7000 bytů, tj. asi 0,16% z celkového počtu všech obydlených bytů. 354 jednotek se ukázalo jako neobydlených, příp. adresa nebyla nalezena nebo nebyla dostupná. Úspěšnost dotazování v ostatních bytových jednotkách uvádí tabulka.¹⁸

Tabulka 3: Úspěšnost sběru dat SILC 2005 (Zdroj: www.czso.cz)

Počet vyšetřených bytových domácností	4351	64,8%
Počet nevyšetřených bytových domácností	2363	35,2%
Celkem. Z toho:		100,00%
domácnost nezastižena		19,6%
cizí státní příslušník (jazyková bariéra)		0,8%
odmítnutí z objektivních příčin (vysoký věk, zdravotní nebo rodinné problémy)		4,1%
odmítnutí zúčastnit se šetření		75,5%

¹⁴ Především proto, že někteří respondenti do těchto dávek nesprávně zahrnují dávky sociální potřeby. Protože však v této práci pracuji s příjmy jako celkem, nikoliv dle podrobného členění (příjmy ze závislé činnosti, sociální dávky, příjmy z pronájmů atd.), nejsou tyto nepřesnosti pro účely této práce podstatné.

¹⁵ European Union – Statistics on Income and Living Conditions, tato povinnost každoročního zjišťování vyplývá z novelizace Nařízení (EC) 1177/2003 a navazujících nařízení Evropské komise.

¹⁶ Hlavním rozdílem při výběru vzorku domácností mezi zjišťováním EU – SILC a Mikrocenzus byla skutečnost, že celkový počet plánovaných SO byl nižší (700 oproti 960), nižší byla i velikost výběru jednotek v druhém kole (10 oproti 12). V případě zjišťování SILC také nebyly vypouštěny malé sčítací obvody s méně než 24 jednotkami, ale tyto byly slučovány, přičemž hranice pro tuto operaci se snížila na 20 jednotek.

¹⁷ Oproti výše uvedeným charakteristikám bylo pro přepočet použito rovněž členění dle věku a velikostní skupiny obcí.

¹⁸ Zdroj SILC 2005

Z výsledků vyplývá v podstatě stejná struktura neúspěšných odpovědí v rámci zjišťování domácností, jakou mělo šetření Mikrocenzu. Stejně jako u něho se také opět lišila úspěšnost tazatelů v jednotlivých krajích, která se pohybovala od 51,1% (Praha) do 73,9% (Moravskoslezský kraj). To ukazuje bohužel na ještě nižší úspěšnost než v roce 2002. Především výsledky z Prahy jsou už povážlivě nízké, protože se snižujícím se počtem vyšetřených domácností klesá pochopitelně kvalita výběrových dat. I v následných analýzách na těchto datech provedených pak dosahujeme nepřesných výsledků. Konkrétní definice pěti zjišťovaných faktorů a jejich případné rozdíly oproti metodice roku 2002 uvádí příloha.

Tabulka 4: Úspěšnost vyšetření v krajích, SILC 2005 (Zdroj: www.czso.cz)

Kraj	Bytů v šetření	z toho vyšetřeno		Kraj	Bytů v šetření	z toho vyšetřeno	
		počet	%			počet	%
Hl. město Praha	917	469	51,1	Královeshradecký	364	229	62,9
Středočeský	721	59	63,7	Pardubický	304	207	68,1
Jihočeský	396	249	62,9	Vysočina	317	233	73,5
Plzeňský	375	275	73,3	Jihomoravský	708	425	60,0
Karlovarský	193	118	61,1	Olomoucký	414	308	74,4
Ústecký	560	362	64,6	Zlínský	358	241	67,3
Liberecký	272	174	64,0	Moravskoslezský	815	602	73,9

3 Rozdělení příjmů domácností

Výsledkem šetření příjmů domácností je obsáhlý datový soubor s konkrétním údajem pro každou sledovanou jednotku (domácnost). Abychom mohli tato data spolehlivě interpretovat a pokusit se poté o navržení teoretického modelu rozdělení základního souboru, je nejprve třeba prozkoumat získané empirické rozdělení výběrového souboru. K tomuto účelu velice dobře slouží matematické výpočty jeho významných charakteristik, přinášející tak ucelenější pohled na zprvu nepřehlednou masu dat. Dalším velmi dobrým způsobem, jak poznat získané údaje, jsou nejrůznější grafické nástroje, zobrazující průběh rozdělení, na jehož základě můžeme potom posoudit vhodnost uvažovaného teoretického modelu.

Charakteristiky empirického rozdělení

Míry úrovně (polohy)

Základní vlastností rozdělení je jeho úroveň, kterou můžeme měřit pomocí jednak středních hodnot, založených na výpočtu pomocí všech jednotek statistického zjišťování (průměrů), a vedle toho středních hodnot vypočítaných pouze na základě význačných hodnot souboru (modus, medián). Pro účely této práce jsem zvolil prostý aritmetický průměr a medián. Prostý aritmetický průměr je elementární charakteristikou výběru, kterou můžeme definovat podle jednoduchého vztahu

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (3.1)$$

Aritmetický průměr má dobrou vypovídací schopnost zejména u symetrických rozdělení. V případě příjmových rozdělení, které bývá nesymetrické (obvykle s lognormálním průběhem křivky), však výstižnější pohled na získaná data přináší medián \tilde{x} . Ten rozděluje data na dvě stejně četné části, z nichž každá zahrnuje 50% jednotek, a je tedy určen jako 50% kvantil. Jeho odchylka od prostého aritmetického průměru potom ukazuje na nerovnoměrnost v rozdělení příjmů domácností, a to tím vyšší, čím větší je rozdíl mezi oběma hodnotami. Medián definujeme tedy jako¹⁹

¹⁹ Bartošová, Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v ČR po roce 1990 (2006)

$$\tilde{x} = x\left(\frac{n+1}{2}\right), \quad (3.2)$$

pro n liché, resp.

$$\tilde{x} = \frac{1}{2} \left(x\left(\frac{n}{2}\right) + x\left(\frac{n+2}{2}\right) \right), \quad (3.3)$$

pro n sudé. Dalším výstižnými ukazateli polohy jsou dolní a horní kvartil, tedy hodnota v 25% souboru, resp. 75% souboru, které lze zapsat

$$\tilde{x}_{0,25} = x\left(\frac{n}{4}\right)_{+}, \quad (3.4)$$

$$\tilde{x}_{0,75} = x\left(\frac{3n+4}{4}\right)_{-}, \quad (3.5)$$

kde $+$ v prvním případě značí zaokrouhlení na nejbližší vyšší celé číslo a $-$ zaokrouhlení na nejbližší nižší celé číslo. Souhrnným ukazatelem, který používá poslední tři uvedené charakteristiky, je odhad BES²⁰, který definujeme vztahem

$$BES = 0,25\tilde{x}_{0,25} + 0,5\tilde{x} + 0,25\tilde{x}_{0,75}, \quad (3.6)$$

kde $\tilde{x}_{0,75}$, $\tilde{x}_{0,25}$, \tilde{x} jsou hodnoty horního kvartilu, dolního kvartilu a mediánu.

Míry variability

Variabilita udává různost hodnot zkoumané proměnné. Při nízké variabilitě dat se od sebe hodnoty liší jen málo, čímž stoupají na významu charakteristiky polohy, jako medián či aritmetický průměr. Pokud je však variabilita souboru vysoká, jejich vypovídací hodnota ztrácí na síle. Hlavními momentovými charakteristikami variability pro vzorek hodnot ze základního souboru je výběrový rozptyl s^2 , definovaný²¹

²⁰ Bartošová, Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v ČR po roce 1990 (2006)

²¹ Cyhelský – Kahounová – Hindls, Elementární statistická analýza (1999), str. 67

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}, \quad (3.7)$$

a z rozptylu odvozená směrodatná odchylka s

$$s = \sqrt{s^2}, \quad (3.8)$$

Relativní variabilitu vystihuje variační koeficient

$$V = \frac{s}{\bar{x}}. \quad (3.9)$$

Pro posouzení variability výběrového souboru, které není zkreslené odlehlými hodnotami (zde především hodnoty v pravé části rozdělení příjmů), se používají vybrané kvantilové charakteristiky variability. Kvartilové rozpětí

$$QR = \tilde{x}_{0,75} - \tilde{x}_{0,25}, \quad (3.10)$$

a poměrná kvartilová odchylka

$$RQD = \frac{\tilde{x}_{0,75} - \tilde{x}_{0,25}}{\tilde{x}_{0,75} + \tilde{x}_{0,25}}, \quad (3.11)$$

kde $\tilde{x}_{0,75}, \tilde{x}_{0,25}$ jsou hodnoty horního a dolního kvartilu.

Šikmost a špičatost

Šikmost vypovídá o rozdělení poloviny malých hodnot vůči druhé polovině hodnot větších. Pokud zabírá první polovina hodnot větší část variačního rozpětí (a je tedy méně nahuštěna než polovina druhá), vykazuje soubor zápornou šikmost a aritmetický průměr je v takovém případě menší než medián hodnot. V příjmových rozděleních je obvyklá situace opačná, kdy medián má nižší hodnotu než aritmetický průměr, z čehož vyplývá, že polovina větších čísel souboru je oproti první polovině méně nahuštěna. Rozdělení tak vykazuje kladnou šikmost. V případě rovnosti obou charakteristik polohy je šikmost nulová. Obecně

definujeme šikmost α nejčastěji jako třetí normovaný moment rozdělení četností proměnné x daný vztahem²²

$$\alpha = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{ns^3}, \quad (3.12)$$

kde s je rozptyl, a dále používáme kvartilový koeficient šikmosti

$$\tau = \frac{\tilde{x}_{0,75} - 2\tilde{x} + \tilde{x}_{0,25}}{\tilde{x}_{0,75} - \tilde{x}_{0,25}}, \quad (3.13)$$

kde $\tilde{x}_{0,75}$, $\tilde{x}_{0,25}$, \tilde{x} jsou hodnoty horního kvartilu, dolního kvartilu a mediánu.

Špičatost rozdělení je tím větší, čím více jsou nahuštěny hodnoty prostřední velikosti ve srovnání s hodnotami ostatními, nebo-li velká špičatost rozdělení značí vysokou koncentraci hodnot v blízkosti středních hodnot. Pro určení špičatosti β se nejvíce používá čtvrtý normovaný moment zmenšený o 3 dle vzorce²³

$$\beta = \frac{\sum_{i=1}^k (x_i - \bar{x})^4 n_i}{ns^4} - 3, \quad (3.14)$$

a Moorsův koeficient (kvantilová charakteristika)

$$b_Q = \frac{\tilde{x}_{0,875} - \tilde{x}_{0,625} + \tilde{x}_{0,375} - \tilde{x}_{0,125}}{\tilde{x}_{0,75} - \tilde{x}_{0,25}}, \quad (3.15)$$

kde $\tilde{x}_{0,125}$, $\tilde{x}_{0,25}$, $\tilde{x}_{0,375}$, $\tilde{x}_{0,625}$, $\tilde{x}_{0,75}$, $\tilde{x}_{0,875}$ jsou hodnoty 12,5%, 25%, 37,5%, 62,5%, 75% a 87,5% kvantilu příslušného rozdělení.

²² Bartošová, Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v ČR po roce 1990

²³ tamtéž

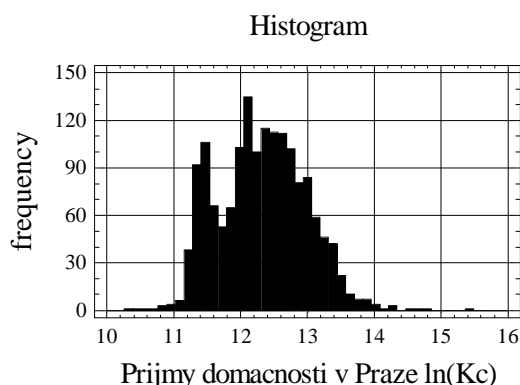
Grafické nástroje zobrazení rozdělení

Histogram

Histogram je jedním z nejstarších a nejpoužívanějších odhadů hustoty pravděpodobnosti.²⁴ Jde o sloupcový graf, zobrazující na ose x jednotlivé třídy (neboli intervaly hodnot), které definují šířku sloupců. Osa y pak definuje hustotu pravděpodobnosti jednotlivých tříd, jíž také odpovídá výška příslušného sloupce grafu. Celková plocha sloupců histogramu je rovna jedné. Důležitou proměnou pro správnou vypovídací schopnost grafu je vhodné stanovení šířky tříd (resp. jejich celkového počtu na intervalu daném rozpětím maximální a minimální hodnoty). Pokud bychom totiž stanovili počet tříd příliš malý, mohlo by dojít ke ztrátě informací o původním rozdělení; jejich příliš velký počet naopak způsobí neobsazenost některých z nich. Pro přibližně symetrické rozdělení se lze při určení počtu tříd histogramu řídit vzorcem²⁵

$$m = 15\sqrt{\left(\frac{n}{100}\right)^2}, \quad (3.16)$$

kde m udává počet tříd. Z toho plyne, že histogram dobře zobrazuje rozdělení hustot pravděpodobnosti pro dostatečně velké výběrové soubory, kdežto při menším počtu hodnot výběru není vhodný. Pro vykreslení histogramu jsem použil statistický program StatGraphics for Windows ve své aplikaci Plot > Exploratory plots > Frequency histogram. Příklad průběhu histogramu znázorňuje obr. 1 (data Mikrocenzus 2002, logaritmus příjmů domácností v Hl. městě Praha, velikost výběrového souboru 1587, počet tříd 47).



Graf 1: Histogram příjmů domácností

²⁴ Meloun - Militký, Statistické zpracování experimentálních dat (1998), str. 68

²⁵ Williams, Weighing the Odds (2001)

4 Model rozdělení příjmů

Konstrukce teoretického modelu rozdělení příjmů se skládá ze dvou fází. V první fázi je třeba nalézt teoretickou distribuční funkci, umožňující dobré vystižení empirického rozdělení četností. Druhým krokem je volba vhodné metody odhadu parametrů tohoto modelu.

Volba modelu

Při volbě vhodného modelu umožňující dobrou aproximaci tvaru rozdělení četnosti základního souboru je velmi výhodné vycházet z histogramu. Toto řešení je často nezbytné především v případech, kdy toho o základním souboru příliš mnoho nevíme. Grafická vizualizace nám poskytuje první komplexnější náhled na sledovaná data.

Při analýze rozdělení příjmů domácností můžeme vycházet z historických zkušeností s modelováním v této oblasti. Postupem času bylo zjištěno, že velmi dobré výsledky dává varianta logaritmicko-normálního modelu. Jak jsem uvedl výše, cílem práce je ověřit platnost tohoto modelu na výběrovém souboru dat příjmů českých domácností z let 2002 a 2005. Zaměřit bych se chtěl nejen na modelování celého výběrového souboru, ale především na jednotlivé podskupiny. Ty získám rozdělením výběrového souboru podle významných ekonomických, sociálních či demografických faktorů.

Logaritmicko-normální rozdělení

Lognormální model se dvěma parametry je definován jako model nezáporné náhodné veličiny X , jejíž logaritmy (a to jak přirozené, tak i dekadické) jsou rozděleny normálně. Toto rozdělení je definováno vztahem²⁶

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \left[-\frac{\ln(x-\mu)^2}{2\sigma^2} \right] \quad \text{pro } 0 < x < \infty, \quad (4.1)$$

= 0 jinak

kde μ je střední hodnota a σ rozptyl náhodné veličiny $\ln X$. Obecně se tento model považuje za nejvhodnější aproximaci příjmového rozdělení obyvatel. Rozhodně však neplatí obecně pro všechny podskupiny domácností. V některých případech, zvláště u souborů s odlehlými

²⁶ Hátle – Kahounová, Úvod do teorie pravděpodobnosti (1987), str. 129

hodnotami, vykazuje velkou nepřesnost. Proto rozšířím počet zvolených teoretických rozdělení o další typy²⁷.

Gamma rozdělení

Rozdělení gama je speciálním případem exponenciálního rozdělení²⁸ $E(0,\delta)$, které je definováno jako rozdělení náhodná veličiny X , jejíž hustota pravděpodobnosti má tvar²⁹

$$f(x) = \frac{1}{\Gamma(m)\delta^m} e^{-x/\delta} x^{m-1} \quad \text{pro } x > 0, \quad m > 0, \quad \delta > 0, \quad (4.2)$$

$$= 0 \quad \text{jinak}$$

obecně ho značíme jako $\Gamma(m,\delta)$, kde m je parametr tvaru a δ parametrem měřítka. Gamma rozdělení se obvykle používá v teoriích hromadné obsluhy, životnosti apod.

Weibullovo rozdělení

Hustota pravděpodobnosti Weibullova rozdělení je dána předpisem

$$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \quad \text{pro } x \geq 0, \quad (4.3)$$

$$= 0 \quad \text{jinak}$$

kde $k > 0$ je parametrem tvaru a $\lambda > 0$ je parametrem míry. Používá se při analýzách bezporuchovosti.

Normální rozdělení

Normální rozdělení je symetrické kolem své střední hodnoty, což není typický předpoklad pro příjmová rozdělení. Příjmy obyvatel jsou z principu omezeny zdola minimální hodnotou nula³⁰, naproti tomu maximální hodnota příjmů není nijak omezena. Zařazení tohoto typu rozdělení má smysl pro situace, kdy získaná data neodpovídají klasickému,

²⁷ Tato rozdělení se obvykle aplikují na jiné případy než na modelování příjmů

²⁸ Exponenciální rozdělení $E(\alpha,\delta)$, kde $x > \alpha$, $-\infty < \alpha < \infty$, $\delta > 0$; gamma rozdělení je speciálním typem tohoto případu, kdy $\alpha = 0$, a tedy $E(0,\delta)$

²⁹ Cyhelský – Kahounová – Hindls, Elementární statistická analýza (1999), str. 178

³⁰ Lze ovšem předpokládat, že každá domácnost má nějaký příjem, a proto bude minimum vyšší

nesymetricky vychýlenému průběhu příjmů domácností. Předpokládám, že toto rozdělení se může uplatnit ve specifických případech, pokud modelujeme příjmy úžeji vymezené skupiny obyvatel (důchodci, nezaměstnaní). Hustota pravděpodobnosti modelu má tvar

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{pro } x \in R, \quad (4.4)$$

kde μ je střední hodnota a σ je rozptyl. Normální rozdělení má široké použití v mnoha vědních oborech od biologie přes fyziku k ekonomii, např. při měření náhodných chyb.

Laplaceovo rozdělení

Laplaceovo rozdělení je odvozeno z normálního rozdělení. Je také souměrné, má však špičatější průběh. Model je definován hustotou pravděpodobnosti

$$f(x) = \frac{1}{2b} e^{-\left|\frac{x-\mu}{b}\right|}, \quad (4.5)$$

kde μ je střední hodnota a b parametr tvaru. Laplaceovo rozdělení by mohlo dobře vystihovat průběh příjmů u domácností bez ekonomicky aktivních členů (především důchodců). Tyto výběry mají nahuštěn velký počet jednotek kolem střední hodnoty. Používá se tam, kde má rozdělení výběrového souboru příliš špičatý průběh na to, aby bylo použito normální rozdělení.

5 Odhad charakteristik empirického rozdělení a parametrů jeho teoretického modelu

Předpoklady

Vzhledem k rozsahu výběrového souboru čistých ročních příjmů domácností lze pro odhad charakteristik empirického rozdělení a parametrů zvoleného teoretického modelu použít s poměrně velkou přesností bodový odhad (tj. když parametr rozdělení náhodné veličiny nenáhodného vektoru Θ odhadneme výběrovou charakteristikou, tedy vypočteným bodovým odhadem $\hat{\Theta}$). V praxi známe celou řadu metod, jak získat bodový odhad. Důležitými kritérii při posuzování vhodnosti každé z nich jsou³¹:

- **Konzistence odhadu**

Odhad $\hat{\Theta}$ je konzistentní, když pro daný rozsah výběru pravděpodobnost toho, že jeho vzdálenost od skutečné hodnoty Θ je libovolně malá, je rovna jedné. Vyšší konzistence bodového odhadu je častější při výběrových souborech většího rozsahu, př. pro nestranné odhady.

- **Nestrannost odhadu**

Jako nestranný odhad označujeme takový bodový odhad, kdy střední hodnota výběrového souboru je rovna parametru základního souboru.

- **Vydatnost odhadu**

Vydatnost odhadu určuje minimální rozptyl odhadu okolo skutečné hodnoty Θ . Vzhledem ke všem odhadům tohoto parametru. Ze dvou bodových odhadů $\hat{\Theta}_1$ a $\hat{\Theta}_2$ téhož parametru Θ stejného výběrového souboru má větší vydatnost ten s menším rozptylem.

Pokud odhad obsahuje všechny informace o výběru, říkáme, že se jedná o postačující odhad základního souboru. Jako nejlepší odhad pokládáme ten, který je současně nestranný, vydatný i postačující.

³¹ Meloun - Militký, Statistické zpracování experimentálních dat (1998), str. 128

Metoda maximální věrohodnosti

Vzhledem k rozsahu výběrového souboru jsem pro svou práci vybral jako nástroj bodového odhadu parametrů modelu rozdělení příjmů metodu maximální věrohodnosti. Výhodou této metody je její konzistentnost a vydatnost.³²

Maximálně věrohodný odhad $\hat{\Theta} = (\hat{\Theta}_1, \dots, \hat{\Theta}_k)$ vektoru parametrů $\Theta = (\Theta_1, \dots, \Theta_k)$ předpokládaného rozdělení, z něhož pochází výběr $\{X_1, \dots, X_n\}$, je argumentem suprema věrohodnostní funkce.

Pro spojitě náhodné veličiny je věrohodnostní funkce $L(\Theta)$ sdružená hustota pravděpodobnosti $f(\hat{\Theta}, x_1, \dots, x_n)$. Pokud jsou všechny prvky výběru nezávislé, pak platí³³, že

$$L(\Theta) = \prod_{i=1}^n f(x_i, \Theta). \quad (5.1)$$

Pokud když se místo $L(\Theta)$ používá jejího logaritmu, který přitom zachovává polohu extrému. Maximálně věrohodný odhad $\hat{\Theta}$ vektoru parametrů Θ odpovídá maximu věrohodnostní funkce $L(\Theta)$. Po derivaci logaritmu věrohodnostní funkce vychází rovnice

$$\frac{\delta \ln L(\Theta)}{\delta \Theta_j} = \sum_{i=1}^n \frac{\delta \ln f(x_i, \Theta)}{\delta \Theta_j} = 0, \quad (5.2)$$

pro $j = 1, \dots, k$, kde k je počet parametrů.

Odhad pro dvouparametrické lognormální rozdělení

Dvouparametrické logaritmicko-normální rozdělení, jinak také LN (μ, σ^2) , má hustotu pravděpodobnosti určenou rovnicí (5.1). Logaritmus věrohodnostní funkce je definován rovnicí³⁴

³² Bartošová, Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v ČR po roce 1990 (2006)

³³ Meloun - Militký, Statistické zpracování experimentálních dat (1998), str. 129

³⁴ tamtéž, str. 162

$$\ln L = -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^n \ln x_i - \frac{1}{2\sigma^2} \sum_{i=1}^n [\ln(x_i - \mu)]^2. \quad (5.3)$$

Maximálně věrohodné odhady veličin μ a σ^2 , tj. výběrový průměr \hat{x} , resp. výběrový rozptyl $\hat{\sigma}$, stanovíme odvodíme z rovnic výše jako³⁵

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln x_i, \quad (5.4)$$

resp.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (\ln x_i - \hat{\mu})^2. \quad (5.5)$$

Analogicky se odhadují parametry u ostatních rozdělení (přesné odvození není předmětem této práce, a proto ho zde neuvádím), přesný výpočet parametrů provede statistický software. Výjimkou je odhad parametrů pro Laplaceovo rozdělení, jehož výpočet provedu na základě odhadů

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (5.6)$$

resp.

$$\hat{b} = \frac{1}{n} \sum_{i=1}^n |x_i - \mu|. \quad (5.7)$$

³⁵ Bartošová, Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v ČR po roce 1990 (2006)

6 Testování shody modelu s empirickým rozdělením

Volbu teoretického modelu, který by vhodně vystihoval průběh příjmů vybraných skupin, provedeme na základě obecných znalostí, či úvah o sledovaném jevu (příjmech domácností), popř. prostřednictvím grafického zobrazení průběhu příjmů výběrového souboru, získaného např. z histogramu. K posouzení, zda je tato volba úspěšná, či nikoliv, slouží řada testů k ověření vhodnosti předpokládaného rozdělení. Jedním z nejužívanějších je χ^2 test dobré shody, který jsem také zvolil pro svou práci. Vizualně lze posoudit shodu modelu s empirickým rozdělením některou z grafických metod, např. P-P grafů.

Numerické ověření shody: χ^2 test dobré shody

V χ^2 testu dobré shody obecně testujeme předpoklad (hypotézu) H_0 : zvolený model je vhodnou aproximací rozdělení základního (úplného) souboru, proti alternativě H_1 : model není vhodný. Nulová hypotéza H_0 předpokládá, že v základním souboru roztríděném podle kvalitativního či kvantitativního znaku do k vzájemně se nepřekrývajících skupin, jsou podíly variant v základním souboru rovny číslům³⁶ $\pi_{0,1}, \pi_{0,2}, \dots, \pi_{0,k}$. Protože neznáme parametry rozdělení (pouze jeho model), a musíme je odhadnout z výběrového souboru. Jde tedy o tzv. neúplně specifikovaný model. Testové kritérium má tvar

$$G = \sum_{i=1}^k \frac{(n_i - n\pi_{0,i})^2}{n\pi_{0,i}}, \quad (6.1)$$

kde n_i jsou výběrové četnosti a $n\pi_{0,i}$ jsou teoretické četnosti v i -té skupině. Za předpokladu velkého výběru má statistika G přibližně χ^2 rozdělení s $v = k - m - 1$ stupni volnosti, kde m je počet neznámých parametrů. Za kritické hodnoty volíme $100(1-\alpha)\%$ kvantily χ^2 rozdělení s v stupni volnosti, kdy kritický obor nepřijetí hypotézy H_0 je určen nerovností

$$G \geq \chi_{1-\alpha}^2. \quad (6.2)$$

Uvedený test dobře indikuje shodu modelu se základním souborem za splnění podmínky dostatečného rozsahu souboru. Je nutné zajistit dostatečný počet hodnot v každé skupině tak, že by platilo

³⁶ Hindls - Hronová - Seger, Statistika pro ekonomy (2004), str. 152

$$n\pi_{0,i} > 5 \quad \text{pro } i = 1, 2, \dots, k. \quad (6.3)$$

Ze vzorce (6.3) vyplývá, že na test lze pohlížet také tak, že se budeme snažit najít P-hodnotu, pro kterou dosahuje součet testového kritéria přes všechny hodnoty právě kritické hranice. Ta potom určuje, na kolik procent lze daným modelem vyjádřit studovaná data. K výpočtu této hodnoty slouží distribuční funkce χ^2 rozdělení. Pro určení počtu tříd v testu vycházím z předpisu (4.16). Test shody především v případě velkých souborů ($n > 100$) nemusí hrát rozhodující roli při přijímání (či zamítání) hypotézy o vhodnosti aproximace daného empirického rozdělení zvoleným modelem. Někdy totiž i přes nízkou P-hodnotu testu vykazuje model dobrou shodu v grafickém testu. Pro přesnější závěr lze tedy použít některou z grafických metod.

Grafické nástroje ověření shody: P-P grafy

Probability – probability plot, neboli P-P graf, je grafickým prostředkem k posouzení míry shody teoretického a výběrového rozdělení, kde na jedné straně máme distribuční funkci, danou předpisem zvoleného teoretického rozdělení

$$F(x) = P(X \leq x), \quad (6.4)$$

a na druhé straně distribuční funkci dat výběrového souboru, seříděných podle velikosti, danou předpisem

$$F(x_i) = \frac{(i - 0,5)}{n}. \quad (6.5)$$

Při shodě obou rozdělení by měla mít tato závislost zhruba lineární průběh.

7 Regresní analýza

Regresní analýza je nástrojem pro vysvětlení vztahu mezi závislou proměnnou Y a jednou či několika vysvětlujícími proměnnými X_1, X_2, \dots, X_n . Regresní model³⁷

$$y_i = \eta_i + \varepsilon_i, \quad (7.1)$$

udává i -tou hodnotu y_i vysvětlované proměnné Y jako součet podmíněné střední hodnoty η_i proměnné Y při kombinaci vysvětlujících hodnot $x_{1i}, x_{2i}, \dots, x_{ki}$ a náhodného vlivu ε_i .

Bodové odhady regresních parametrů bývají zjišťovány metodou nejmenších čtverců, tj. hledáme taková a_0, a_1, \dots, a_n , které minimalizují reziduální součet čtverců

$$S_R = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (7.2)$$

kde $\hat{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_k x_{ki}$ je bodový odhad regresní funkce při i -té kombinaci proměnných. Rovnice pro výpočet odhadů parametrů dostaneme z parciálních derivací S_R podle jednotlivých proměnných. Část variability, která je vysvětlená regresním modelem, nazýváme teoretický součet čtverců

$$S_T = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2. \quad (7.3)$$

Celkový součet čtverců je pak dán součtem S_R a S_T , takže

$$S_y = \sum_{i=1}^n (y_i - \bar{y})^2 = S_T + S_R. \quad (7.4)$$

Jako jedna z nejdůležitějších charakteristik vhodnosti regresního modelu bývá označován koeficient determinace, který získáme z výše uvedených vztahů jako podíl

³⁷Jarošová – Pecáková, Příklady k předmětu Statistika B (2004)

$$R^2 = \frac{S_T}{S_y}. \quad (7.5)$$

F-Test

F-testem testujeme statistickou hypotézu $H_0 : a_1 = a_2 = \dots a_3 = 0$ oproti alternativě H_1 , kde alespoň jeden parametr je nenulový, tj. alespoň jedna z proměnných je vysvětlena modelem Testovací statistika je dána vztahem

$$F = \frac{\frac{S_T}{p-1}}{\frac{S_R}{n-p}}. \quad (7.6)$$

Tato statistika má za platnosti nulové hypotézy Fischerovo rozdělení s $(p-1)$ a $(n-p)$ stupni volnosti, kde n je rozsah výběru a $p = k + 1$ je počet regresních parametrů. Hypotézu H_0 zamítáme, je-li minimální hladina významnosti P-value menší než hladina významnosti α . To nastane pokud platí, že hladina testované statistiky je dána nerovností

$$F > F_{1-\alpha}(p-1, n-p). \quad (7.7)$$

Modely s umělými proměnnými

Pro vysvětlení vztahu vysvětlované proměnné a jedné nebo více vysvětlujících proměnných můžeme použít regresní analýzu. Toto použití má však svá omezení a specifika. Regresní analýza velmi dobře slouží v případě, kdy vysvětlovaná proměnná a nezávislé vysvětlující proměnné mají kvantitativní spojitý charakter, tedy pokud se jedná o spojitě číselné řady dat. V případě analýz závislosti příjmů na jednotlivých demografických faktorech, které jsou předmětem této práce, však tato podmínka v případě vysvětlujících proměnných splněna není. Příjmy obyvatelstva sice mají spojitou kvantitativní řadu dat, ale jednotlivé faktory mají kvalitativní charakter. Hodnoty statistického souboru pak nabývají jedné z nominální (slovní³⁸) varianty zkoumaného znaku. Mezi těmito variantami je buď nemožné určit jejich jednotlivé pořadí, nebo je jen velmi obtížné kvantifikovat vzdálenost

³⁸ dle Hindls – Hronová - Seger, Statistika pro ekonomy (2004) str. 15

jednotlivých variant.³⁹ Pro práci s kvalitativními proměnnými není možné použít regresní analýzu bez jejich patřičné úpravy.

Tento problém řeší zavedení umělých proměnných do modelu. Obvykle se používá binární proměnná, která nabývá hodnot 0 (pokud podmínku nesplňuje) a 1 (pokud ji splňuje). Při zkoumání modelu s kategoriálním znakem, který nabývá více než dvou variant, je třeba použít více umělých proměnných tak, aby umělých proměnných byl o jedničku menší než počet variací sledovaného znaku. Pokud tedy zkoumáme závislost příjmů na vzdělání, které rozdělujeme na čtyři úrovně (vysokoškolské, střední s maturitou, vyučen, základní), je třeba použít tři umělých proměnných.⁴⁰ Bez respektování této podmínky by v modelu došlo k tzv. perfektní multikolinearitě⁴¹, při které je potom regresní analýza nepoužitelná.

Multikolinearita

Multikolinearita je negativní jev, který v modelu nastává, pokud jsou hodnoty některých vysvětlujících proměnných vzájemně lineárně závislé. Pokud je tato lineární závislost úplná, hovoříme o úplné, tedy perfektní multikolinearitě, kterou definujeme⁴²

$$c_0x_0 + c_1x_1 + \dots + c_px_p = 0, \quad (7.7)$$

kde x_1, x_2, \dots, x_p jsou hodnoty nezávislé proměnné a c_1, c_2, \dots, c_p koeficienty, z nichž hodnota alespoň jednoho z nich je různá od nuly. Pak odhad modelu metodou nejmenších čtverců⁴³ neexistuje. Taková úplná lineární závislost se v praxi příliš nevyskytuje. Pro nespolehlivost modelu bohatě stačí, pokud mezi proměnnými existuje dostatečně silná závislost. Pak hovoříme o multikolinearitě. K tomuto jevu dochází jednak (jak jsem již uvedl výše) při nesprávném stanovení počtu umělých proměnných, jednak také při lineární závislosti

³⁹ Např. pro znak pohlaví dostáváme varianty muž nebo žena, přičemž jejich vzájemné pořadí nelze určit. Vedle toho můžeme třeba u znaku dosažené vzdělání určit pořadí variant (vysokoškolské vzdělání je jistě vyšší než středoškolské), ale odstup mezi jednotlivými variantami nelze kvantifikovat (nelze určit, jak velký rozdíl mezi vysokoškolským diplomem a maturitou, ve srovnání s maturitou a výučním listem)

⁴⁰ Stuchlý, Ekonometrie (2000)

⁴¹ více k tomuto pojmu v kap. 9

⁴² Stuchlý, Ekonometrie (2000), str. 63

⁴³ dále již jen MNC

vybraných sledovaných znaků. Multikolinearita je právě při zkoumání závislosti ekonomických veličin poměrně častým jevem.⁴⁴

Multikolinearita má velmi vážné důsledky pro výsledky, které nám model může o sledovaných datech poskytnout. Odhady parametrů modelu jsou velice nepřesné a intervaly spolehlivosti pro získané výsledky jsou příliš široké.

Nejběžnějším způsobem zjištění multikolinearity v modelu je využití korelační analýzy. Kolinearitu mezi dvěma proměnnými odhalíme, pokud mají párové korelační koeficienty vysokou hodnotu (tedy pokud $r_{i,j} > 0,8$). Nejběžnějším způsobem odstranění multikolinearity je vynechání některé ze závislých proměnných z modelu. Existují však i další způsoby, pomocí kterých se lze s multikolinearitou vyrovnat⁴⁵

Autokorelace

Pro lineární regresní model musí být splněn předpoklad, že náhodné složky ε_i jsou nekorelované (nezávislé), tj. platnost vztahu⁴⁶

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{pro } i \neq j. \quad (7.8)$$

Při autokorelaci dochází k porušení této podmínky tak, že mezi po sobě jdoucími náhodnými složkami vzniká sériová závislost, kdy

$$\text{cov}(\varepsilon_i, \varepsilon_j) \neq 0 \quad \text{alespoň pro jednu dvojici } i, j. \quad (7.11)$$

⁴⁴ Příkladem může být zkoumání znalostí z matematiky studentů gymnázií v závislosti na dvou sledovaných znacích, a to na věku a váze studenta. Je zřejmé, že znalosti matematiky v průměru budou vyšší u starších studentů vyšších ročníků než u mladších, ale není logické, aby se matematiky odvíjely od toho, kdo kolik váží. V takto nastaveném modelu by vznikla falešná závislost mezi váhou a úrovní znalostí, a to právě kvůli multikolinearitě mezi vysvětlujícími proměnnými, kdy váha a věk studentů jsou lineárně závislé (a tedy váha studentů je v takovémto modelu redundantní, tedy přebytečnou vysvětlující proměnnou a je třeba ji z modelu vypustit)

⁴⁵ Např. Stuchlý v Ekonometrii uvádí jako další možné postupy získání nových dat, transformace modelu použitím proměnných v jiných diferencích nebo metodu hřebenové regrese. Pro účely této práce však postačí metoda uvedená výše.

⁴⁶ Stuchlý, Ekonometrie (2000), str. 83

K autokorelaci dochází nejčastěji u modelu s časovými řadami (což není předmětem této práce), nicméně ji nelze vyloučit ani u modelů s výběrovými daty. Při výskytu autokorelace pak v modelu dochází k tomu, že odhady jsou nestranné a konzistentní, ale nikoliv optimální. Dochází k odchylce od skutečné hodnoty u rozptylu vypočteného metodou nejmenších čtverců (MNČ), což vede k špatným výsledkům statistické analýzy a k nesprávným závěrům. Jedním z nejpoužívanějších testů na odhalení autokorelace je Durbin – Watsonův test. Testové kritérium je zde tvořeno statistikou⁴⁷

$$d = \frac{\sum_{i=1}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}, \quad (7.12)$$

kde $i=1, \dots, n$ a e_i jsou rezidua. Při pozitivní autokorelaci jsou hodnoty čitatele oproti jmenovateli výrazně menší, a proto se hodnota d blíží k 0. Při negativní autokorelaci mají difference reziduí naopak asi dvakrát větší hodnotu než samotná rezidua, a proto se d blíží k 4. Na nepřítomnost autokorelace ukazují hodnoty d okolo 2.

⁴⁷ tamtéž, str. 85

8 Použitý software

Výpočetní část

Pro výpočty hodnot charakteristik rozdělení jsem použil částečně tabulkový kalkulátor MS Excel 2003. Další jsem provedl v programu Statgraphics for Windows verze 3.0. Přesné názvy funkcí a postupy zadávání příkazů jsou uvedeny v tabulce 5.

Tabulka 5: Výpočet charakteristik

Charakteristika	Program	Příkaz
průměr, medián	MS Excel	=PRŮMĚR, =MEDIAN
dolní a horní kvartil, směrodatná odchylka, šikmost a špičatost	Statgraphics for Windows	Describe – Numeric Data – One Variable Analysis Tabular options – Summary statistics

Parametry rozdělení, při jejichž výpočtu byla použita metoda maximální věrohodnosti, byly získány prostřednictvím Statgraphics. Pro výpočty kvantilů, hodnoty testového kritéria, kritické hodnoty a p-hodnoty použitých rozdělení jsem použil funkce programu MS Excel (tabulka 6).

Tabulka 6: Modelování rozdělení

Funkce	Program	Příkaz
kvantily rozdělení normální lognormální gamma	MS Excel	=NORMINV =LOGINV =GAMMAINV
hodnoty distribuční funkce normální lognormální gamma Weibull	MS Excel	=NORMINV =NORMINV ⁴⁸ =GAMMAIN =WEIBULL
parametry rozdělení	Statgraphics	Describe – Distribution Fitting – Uncensored Data Tabular options – Analysis summary

Všechny ostatní výpočty, stejně jako pomocné výpočty s mezivýsledky (jako výpočet počtu tříd histogramu, určení počtu tříd χ^2 testu dobré shody), jsem provedl pomocí MS Excel podle vzorců uvedených v této funkci. Pro analýzu závislostí metodou lineární regrese byl použit též Statgraphics, stejně jako pro testování na přítomnosti multikolinearity a autokorelace (tabulka 7).

⁴⁸ hodnoty X se zde vkládají zlogaritmované, tedy $X_{LN} = \ln(X)$

Tabulka 7: Analýza závislostí

Funkce	Program	Příkaz
Lineární regrese	Statgraphics	Relate – Multiple Regression
Test multikolinearity		Describe – Numeric data – Multiple Variable analysis - Correlations
Test autokorelace		Relate – Multiple Regression – Durbin–Watson test

Grafické výstupy

Grafické výstupy jsem získal výhradně pomocí programu Statgraphics for Windows. Tento program jsem použil pro modelování průběhu empirického rozdělení pomocí histogramu (se současným zakreslením průběhu teoretické funkce jednoho z modelů rozdělení). Dalším grafem získaným z programu byl P-P graf míry shody empirického rozdělení s teoretickým rozdělením. Pro jeho vykreslení byla použita funkce pro Q-Q graf, kde místo hodnot kvantilů byly zadány hodnoty distribuční funkce.

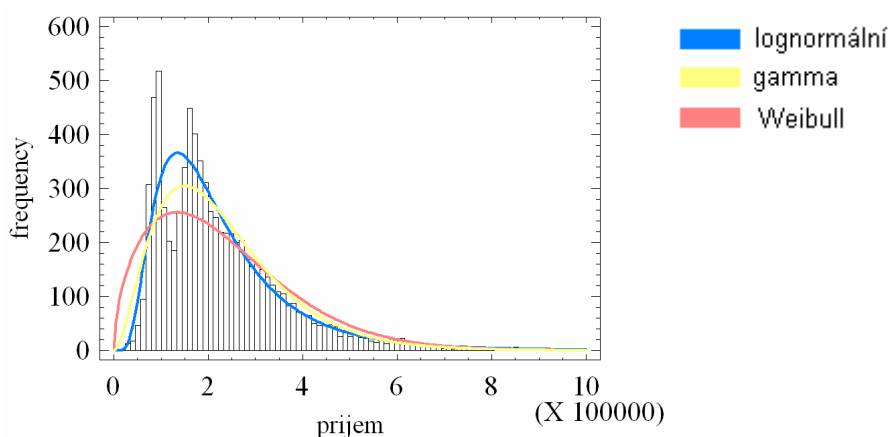
Tabulka 8: Grafické výstupy

Graf	Program	Příkaz
histogram	Statgraphics	Describe – Distribution Fitting – Uncensored Data Graphics options – Frequency histogram
P-P graf	Statgraphics	Compare – Two samples – Two samples comparison Graphics options – Q-Q plot

B Praktická část

1 Výběrové soubory

Průběh histogramu výběrového souboru jako celku má tvar, který nelze s úspěchem aproximovat jedním modelem. Problém tvoří především jeho dvouvrcholová část v první polovině rozpětí příjmů. Jak je naznačeno na grafu 2, průběh žádného z používaných jednovrcholových rozdělení nedokáže takový tvar dostatečně věrně kopírovat. Právě v místě, kde je rozdělení dvouvrcholové, se chybovost modelu prudce zvyšuje.

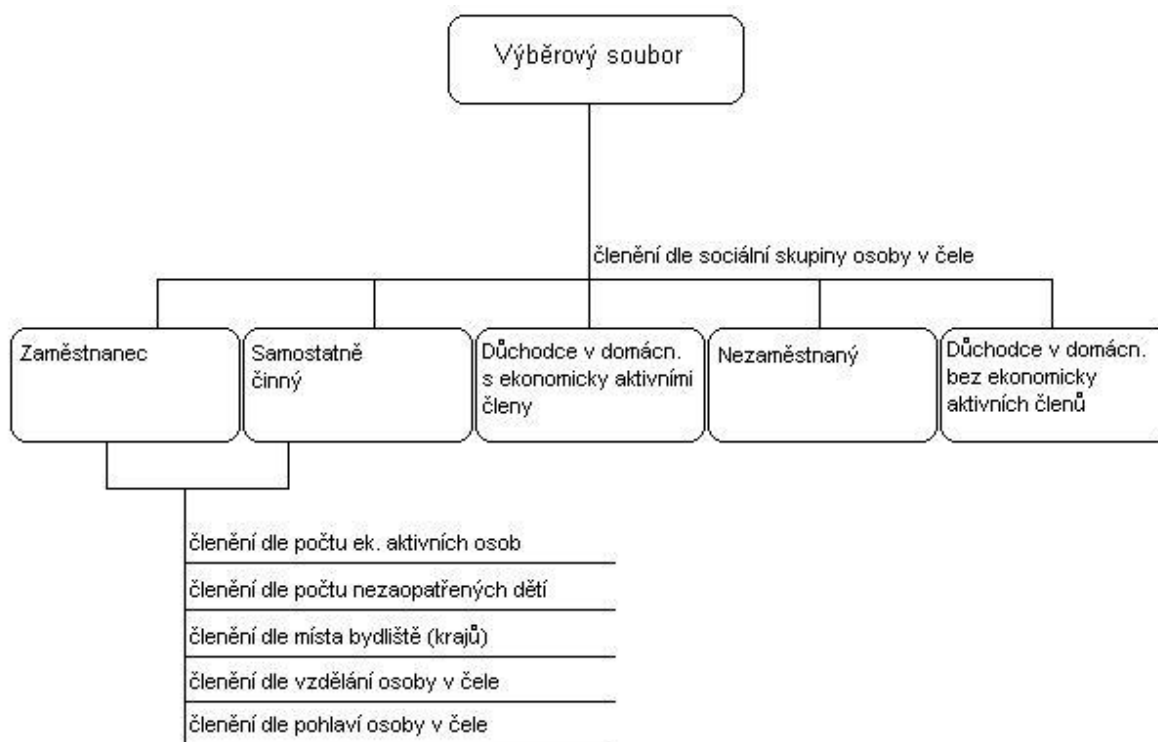


Graf 2: Rozdělení příjmů domácností v ČR 2002

V testu shody pak tato skutečnost způsobuje, že hodnoty testového kritéria značně překračují kritickou hodnotu, a to i několikanásobně. Ve světle těchto výsledků se pak modely jeví jako naprosto nepoužitelné. Proto je nutné výběrový soubor dále rozdělit na více podskupin (výběrových subpopulací), a zkoumat každou z nich zvlášť. Postup rozdělení výběrového souboru naznačuje obrázek 2. Dvouvrcholový průběh rozdělení výběrového souboru je způsoben především skupinou důchodců bez ekonomické aktivity. Tyto domácnosti získávají hlavní část svých příjmů z výplat státních sociálních dávek (důchodů), které jsou zpravidla podstatně nižší než mzdy.

Tabulka 9: χ^2 test shody

model	testové kritérium	kritická hodnota	p hodnota
lognormální	903,0857	108,6479	0,0000
gamma	1352,9672	107,5217	0,0000
Weibullovo	1986,5672	107,5217	0,0000



obrázek 2: Členění výběrového souboru na subpopulace

Metodika jejich výpočtu také způsobuje, že velká část adresátů těchto příjmů dostává přibližně stejné peníze, a rozdělení příjmů subpopulace má mnohem rovnoměrnější průběh, než u domácností vykazující ekonomickou aktivitu. Proto jsem vyčlenil tuto skupinu k samostatnému zkoumání. Vedle toho podrobím samostatné analýze také domácnosti v čele s nezaměstnaným jako speciální případ rozdělení.

Zúžený výběrový soubor, ve kterém chybí právě výše dvě zmiňované kategorie, tvoří především domácnosti v čele se zaměstnanci a samostatně činnými osobami.⁴⁹ Ten pak následně budu zkoumat rozčleněný na další podskupiny, a to z pěti různých hledisek.⁵⁰

⁴⁹ Vedle toho je sem náleží také skupina ostatní, tvořící nezařazené domácnosti; svojí velikostí nehraje ovšem zásadní roli.

⁵⁰ Důležité je ovšem upozornit na to, že osoba v čele domácnosti se určuje nikoliv primárně podle ekonomické aktivity nebo dalších, kvalitativních hledisek, ale většinou podle pohlaví (viz. kap.). Pokud tvoří domácnost tři ekonomicky aktivní osoby a jeden nezaměstnaný, který z definice odpovídá osobě v čele, zařadí se domácnost do kolonky domácnosti s nezaměstnaným v čele. Stejně tak pokud má osoba v čele základní vzdělání, patří domácnost do této skupiny (bez ohledu na vzdělání ostatních členů). Tento aspekt, který je dán použitou metodikou Mikrocenzu, resp. SILCu, vede k tomu, že některé výsledky charakteristik (průměrů, mediánů apod.) mohou být poněkud zkresleny.

2 Členění podle počtu ekonomicky aktivních členů domácnosti

Výběrový soubor jsem v tomto případě rozdělil na tři podskupiny, a to na domácnosti s jedním členem, se dvěma a na domácnosti s třemi a více ekonomicky aktivními členy. Domácnosti bez ekonomicky aktivních členů (tj. důchodci, nezaměstnaní) nebyly do tohoto výběru začleněny, jejich rozdělení bude zkoumáno samostatně. Dominantní podíl na vzorku mají první dvě skupiny; pokud je v domácnosti nějaký člen ekonomicky aktivní, bývá to obvykle jedna, maximálně dvě osoby.

Tabulka 10: Rozdělení skupin dle ek. aktivity

skupina	počet ek. aktivních členů
1	1
2	2
3	3 a více

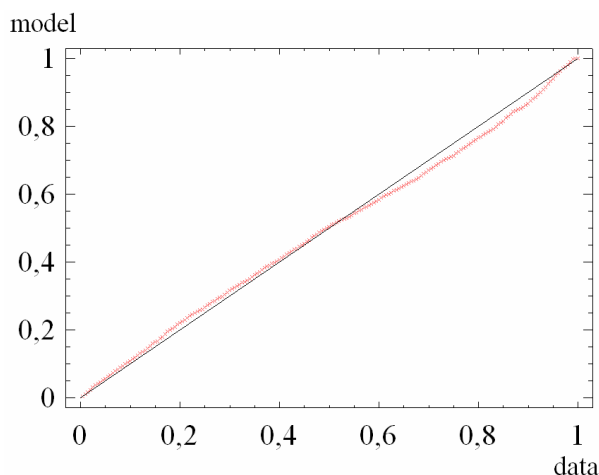
Rok 2002

Tabulka výsledků potvrzuje předpokládaný nárůst průměrných příjmů u domácností s více aktivními členy. Tento nárůst ovšem není proporční, průměrný příjem domácnosti na jednoho ekonomicky aktivního člena s jejich vzrůstajícím počtem klesá. To souvisí s příjmem ze sociálních transferů, které závisí na počtu členů domácnosti (a u domácnosti s více ekonomicky neaktivními členy bývá naopak vyšší).

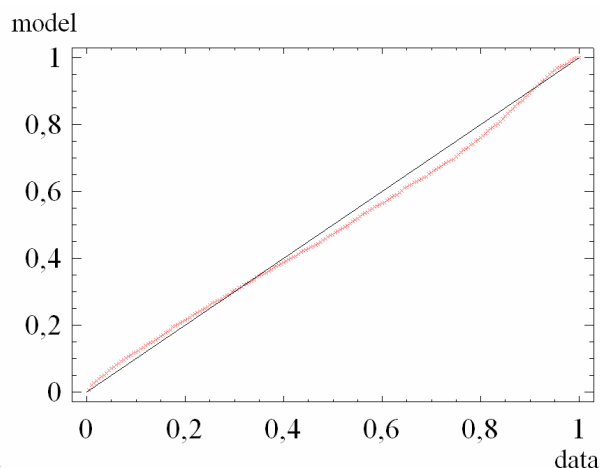
Tabulka 11: Charakteristiky výběrového souboru 2002 (členění dle ek. aktivity)

skupina	1	2	3
počet	2354	2322	449
podíl ve vzorku	45,93%	45,31%	8,76%
průměr	217 133 Kč	326 039 Kč	455 984 Kč
medián	193 695 Kč	294 129 Kč	409 749 Kč
medián / průměr	89,21%	90,21%	89,86%
dolní kvartil	145 352 Kč	237 317 Kč	334 120 Kč
horní kvartil	248 455 Kč	369 828 Kč	522 044 Kč
BES	195 299 Kč	298 851 Kč	418 916 Kč
sm. odchylka	178 740 Kč	148 110 Kč	211 514 Kč
variační koeficient	0,823	0,454	0,464
QR	103 103 Kč	132 511 Kč	187 924 Kč
RQD	0,262	0,218	0,219
šikmost	14,319	55,939	3,933
t	0,062	0,143	0,195
špičatost	321,472	142,204	28,425
bq	1,355	1,444	1,367

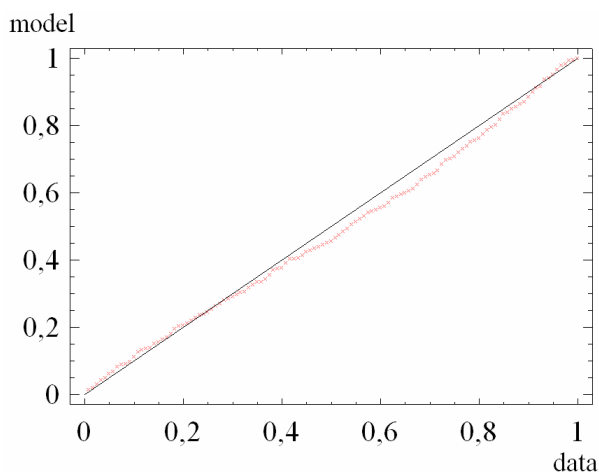
Vedle toho zde také hraje roli fakt, že ženy mají obvykle nižší plat (a tím i čistý příjem) než muži, a pokud v domácnosti pracuje jen jedna osoba, bývá to obvykle muž.⁵¹ První příjem domácnosti bývá tedy vyšší než druhý. Vzájemný poměr mediánu a průměru zůstává téměř konstantní (okolo 90%), nicméně první skupina má značně vyšší hodnotu variačního koeficientu. To naznačuje vyšší nerovnoměrnost v rozdělení příjmů u jednočlenných domácností oproti ostatním. Tomuto závěru nasvědčuje také vyšší hodnota poměrné kvartilové odchylky (RQD).



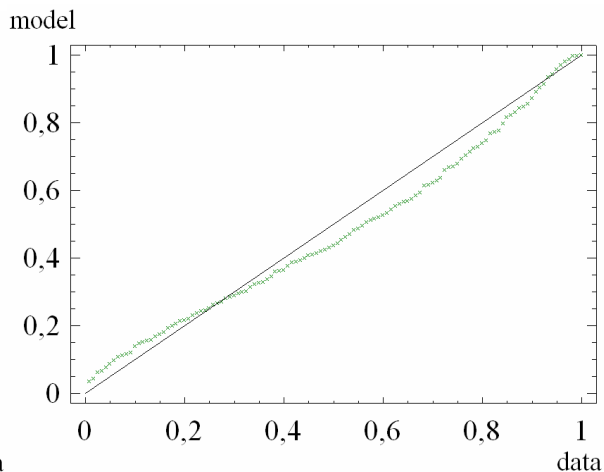
Graf 3: Shoda dat a ln modelu, skupina 1



Graf 4: Shoda dat a ln modelu, skupina 2



Graf 5: Shoda dat a ln modelu, skupina 3



Graf 6: Shoda dat a gamma modelu, skupina 3

⁵¹ Zde vycházím čistě ze statistických údajů ČSÚ

Tabulka 12: Výsledky χ^2 testu

model	testové kritérium	kritická hodnota	p hodnota	závěr
1				
lognormální	85,1424	37,2759	0,0034	ano
gamma	224,7275	36,4371	0,0000	ne
Weibullovo	958,1776	36,4371	0,0000	ne
2				
lognormální	105,0041	36,4371	0,0000	ano
gamma	234,0790	35,5999	0,0000	ne
Weibullovo	671,4621	35,5999	0,0000	ne
3				
lognormální	26,3658	16,1514	0,4984	ano
gamma	52,2079	15,3792	0,0025	ne
Weibullovo	159,7598	15,3792	0,0000	ne

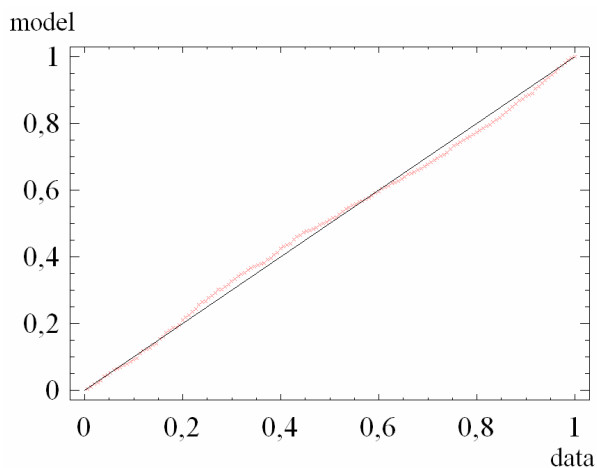
Jak vyplývá z grafů 3-6, nejlépe odpovídá průběhu rozdělení výběrového souboru lognormální model, a to pro skupinu 1 a 2. U obou dvou skupin bylo sice překročeno testové kritérium v testu shody, což vzhledem k velkému rozsahu skupin (více než 2000 hodnot) není neobvyklé. Proto upřednostním grafický test shody provedený P-P grafem. Poměrně dobře vyšel gamma model u skupiny 3, ale graf naznačuje velké odchylky ve střední části pravděpodobnostního rozpětí, proto hypotézu o shodě modelu s výběrovým rozdělení zamítám.

Rok 2005

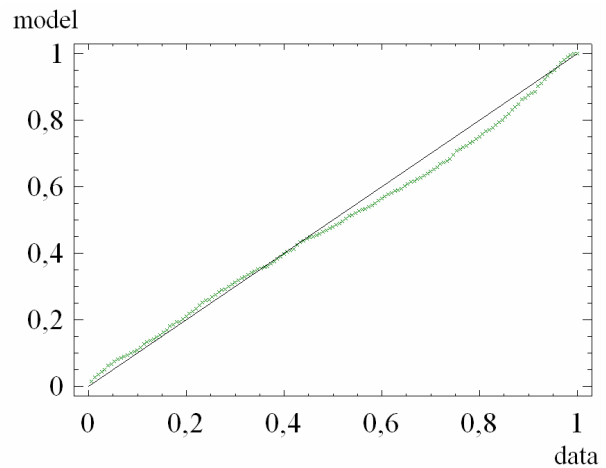
Tabulka 13: Charakteristiky výběrového souboru 2005 (členění dle ek. aktivity)

skupina	1	05/02	2	05/02	3	05/02
počet	1281	54,4	1198	51,6	247	55,0
podíl ve vzorku	46,99%	102,3	43,95%	97,0	9,06%	103,4
průměr	231 719 Kč	106,7	356 156 Kč	109,2	494 039 Kč	108,3
medián	210 424 Kč	108,6	317 460 Kč	107,9	470 316 Kč	114,8
medián / průměr	90,81%	101,8	89,13%	98,8	95,20%	105,9
dolní kvartil	158 558 Kč	109,1	255 500 Kč	107,7	383 574 Kč	114,8
horní kvartil	275 009 Kč	110,7	398 164 Kč	107,7	574 331 Kč	110,0
BES	213 604 Kč	109,4	322 146 Kč	107,8	474 634 Kč	113,3
sm. odchylka	137 660 Kč	77,0	191 749 Kč	129,5	208 454 Kč	98,6
variační koeficient	0,594	72,2	0,538	118,5	0,422	91,0
QR	116 451 Kč	112,9	142 664 Kč	107,7	190 757 Kč	101,5
RQD	0,269	102,6	0,218	100,0	0,199	90,7
šikmost	7,879	55,0	6,132	11,0	4,993	127,0
t	0,109	175,5	0,131	92,2	0,091	46,4
špičatost	139,199	43,3	69,064	48,6	44,929	158,1
bq	1,353	99,9	1,397	96,8	1,137	83,1

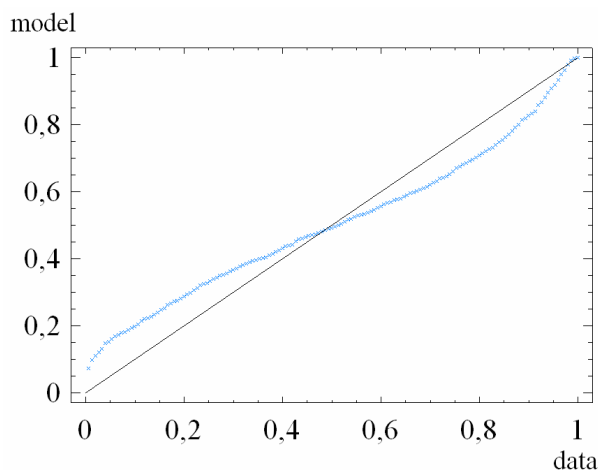
V roce 2005 došlo proti roku 2002 k jen nepatrným posunům mezi vybranými třemi skupinami hodnot. Drtivá většina domácností má opět jednoho, či maximálně dva ekonomicky aktivní členy. Také podíl průměrného příjmu a mediánu zůstává víceméně stejný, výraznější nárůst nastal pouze u třetí skupiny (na 95,2%). V charakteristikách variability se však výsledky liší více; u první skupiny se variabilita spíše snižuje, u druhé je tomu právě naopak. U třetí skupiny zůstávají hodnoty téměř shodné s rokem 2002. Obecný trend v tomto případě nelze vysledovat.



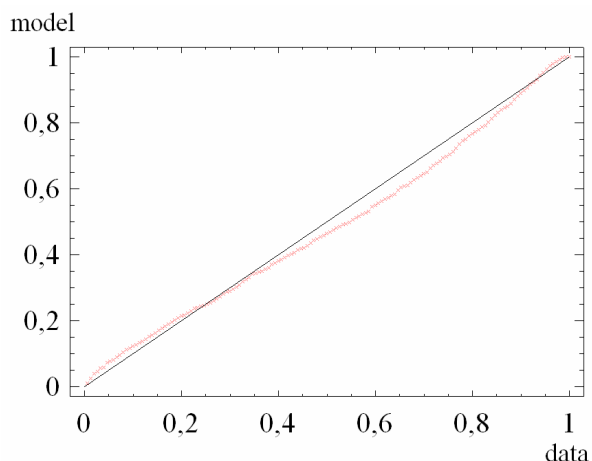
Graf 7: Shoda dat a ln modelu, skupina 1



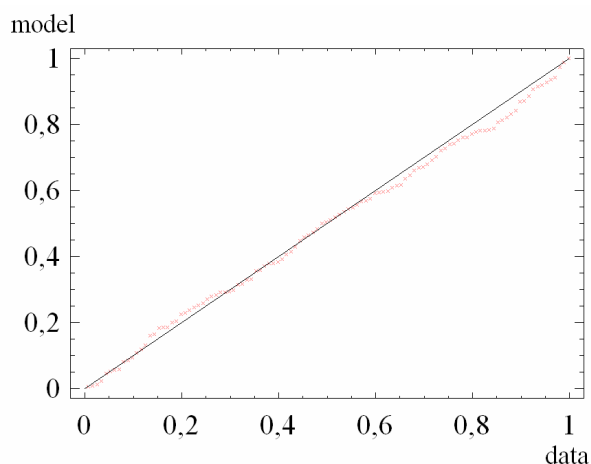
Graf 8: : Shoda dat a gamma modelu, skupina 1



Graf 9: Shoda dat a Weibull modelu, skupina 1



Graf 10: Shoda dat a ln modelu, skupina 2



Graf 11: Shoda dat a ln modelu, skupina 3

Při testování shody nejlépe modelovalo výběrová data u první skupiny lognormální rozdělení. Tento model však nevychází dle grafického výstupu špatně ani u zbylých dvou vzorků. Naopak naprosto nejhorší výsledky byly dosaženy při aproximaci dat Weibullovým rozdělením. Testové kritérium zde mnohonásobně překročilo kritickou mez, velké odchylky odhalil také P-P graf (příklad graf 9).

Tabulka 14: Výsledky χ^2 testu

model	testové kritérium	kritická hodnota	p hodnota	závěr
1				
lognormální	57,3773	27,3256	0,0461	ano
gamma	81,9142	26,5093	0,0002	ne
Weibullovo	309,6004	26,5093	0,0000	ne
2				
lognormální	73,5375	26,5093	0,0010	ano
gamma	166,7800	25,6954	0,0000	ne
Weibullovo	535,4433	25,6954	0,0000	ne
3				
lognormální	32,1871	11,5913	0,0560	ano
gamma	39,5281	10,8508	0,0085	ne
Weibullovo	120,0690	10,8508	0,0000	ne

3 Členění domácností podle počtu nezaopatřených dětí

Podle údajů z šetření o sociální situaci domácností z roku 2001⁵² bylo v ČR 38,4% domácností, ve kterých žilo alespoň jedno nezaopatřené dítě. Z celkového počtu čtyř milionů domácností tak jde zhruba o půldruhý milion případů. V tomto oddíle bych chtěl domácnosti rozdělit právě z tohoto hlediska. Nezaopatřené děti patří ke skupině obyvatel, která je státem zvýhodňována formou nejružnějších sociálních dávek, úspor na daních či jednorázových příspěvků při zvláštních událostech (porodné, pastelkovné apod.). Stát také vyplácí příspěvky na děti, které se odvozují od výše příjmů rodiny. Tento fakt by měl vést k tomu, že u domácností s dětmi by měla být nerovnost mezi příjmy menší, než je tomu v případě domácností bez dětí. S přihlédnutím ke zdrojovým datům jsem výběrový vzorek rozdělil do čtyř skupin, které uvádí tabulka níže.

Tabulka 15: Rozdělení skupin (domácnosti členěné dle počtu dětí)

skupina	počet nezaopatřených dětí
0	0
1	1
2	2
3	3 a více

Co se týče obecného trendu, tak v tomto případě jde o snižování celkového počtu domácností, ve kterých žije minimálně jedno nezaopatřené dítě (viz tabulka 16). Výrazně také ubývá domácností s více dětmi. Tento trend souvisí jednak s tím, že se obecně rodí čím dál méně dětí⁵³, a také s tím, že kvůli vysoké rozvodovosti stoupá počet dětí žijících v neúplných rodinách.

Tabulka 16: Počty nezaopatřených dětí 1991 – 2001 (ČSÚ)

Rok	1991	2001	2001/1991 (%)
1 dítě	731 442	695 654	95,1
2 děti	744 545	618 905	83,1
3 děti	149 191	99 507	66,7
4 děti a více	24 773	20 109	81,2
CELKEM	1 649 951	1 434 175	86,9

⁵² zdroj: ČSÚ

⁵³ Tento trend kulminoval mezi lety 1996 – 2001, kdy se ročně rodilo jen něco okolo 90 tis. dětí. Silné populační ročníky ze 70. let odložily plánování svých potomků o několik let, mnozí se dokonce rozhodly děti vůbec nemít. V posledních několika letech celkový počet narozených dětí opět stoupl (v roce 2006 okolo 105 tis.). Do budoucna se však předpokládá snížení porodnosti, z důvodů odeznění vlivu silných populačních ročníků.

Tabulka 17: Charakteristiky výběrového souboru 2002 (členění podle počtu dětí)

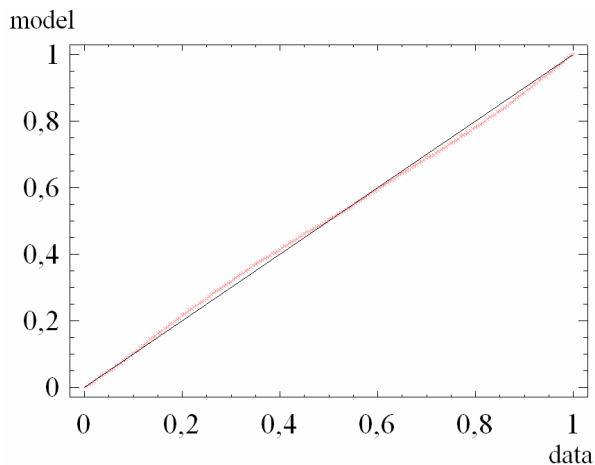
skupina	0	1	2	3
počet	2713	1107	1127	198
podíl ve	52,73%	21,52%	21,90%	3,85%
průměr	270 725 Kč	286 838 Kč	305 640 Kč	299 634 Kč
medián	233 764 Kč	248 911 Kč	275 565 Kč	273 350 Kč
medián /	86,35%	86,78%	90,16%	91,23%
dolní kvartil	164 478 Kč	181 347 Kč	211 360 Kč	213 696 Kč
horní kvartil	326 118 Kč	343 893 Kč	352 006 Kč	342 014 Kč
BES	239 531 Kč	255 766 Kč	278 624 Kč	275 603 Kč
sm. odchylka	184 771 Kč	218 191 Kč	151 593 Kč	144 905 Kč
variační	0,683	0,761	0,496	0,484
QR	161 640 Kč	162 546 Kč	140 646 Kč	128 318 Kč
RQD	0,329	0,309	0,250	0,231
šířkost	5,495	11,168	2,497	3,165
t	0,143	0,169	0,087	0,070
špičatost	67,436	223,561	11,122	16,181
bq	1,339	1,334	1,430	1,170

Výsledky charakteristik těchto podskupin, které přináší tabulka, nejsou jednoznačné. Domácnost s jedním či dvěma dětmi má vyšší průměrný příjem oproti té bez dětí, naproti tomu průměrný příjem domácností s více dětmi se oproti předchozí skupině snížil. Rozdíl je však v obou případech poměrně malý a naznačuje, že děti jako faktor výše příjmu domácnosti nebudou hrát zřejmě významnější roli. Patrné je snížení variability příjmů u domácností s více dětmi; dokládá to jak růst hodnoty poměru mediánu k průměru, tak i snížení variačního koeficientu.

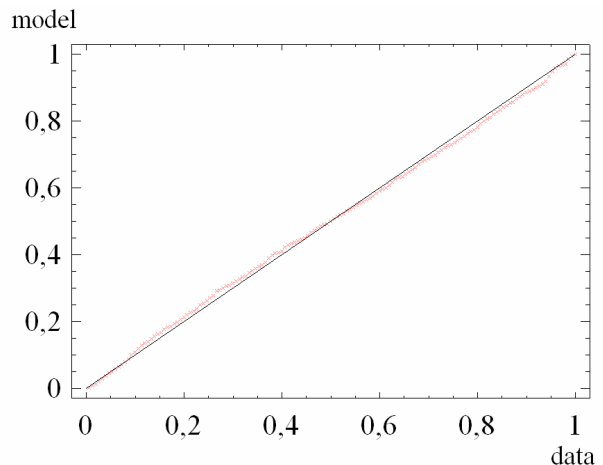
Tabulka 18: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
0				
lognormální	47,5782	39,8013	0,7810	ano
gamma	116,4721	38,9580	0,0000	ne
Weibullovo	425,0656	38,9580	0,0000	ne
1				
lognormální	44,4040	25,6954	0,2545	ano
gamma	71,8659	24,8839	0,0010	ne
Weibullovo	266,6372	24,8839	0,0000	ne
2				
lognormální	52,1212	25,6954	0,0779	ano
gamma	97,4633	24,8839	0,0000	ne
Weibullovo	254,0897	24,8839	0,0000	ne
3				
lognormální	29,3412	10,1170	0,0608	ne
gamma	38,9019	9,3905	0,0045	ne
Weibullovo	88,6592	9,3905	0,0000	ne

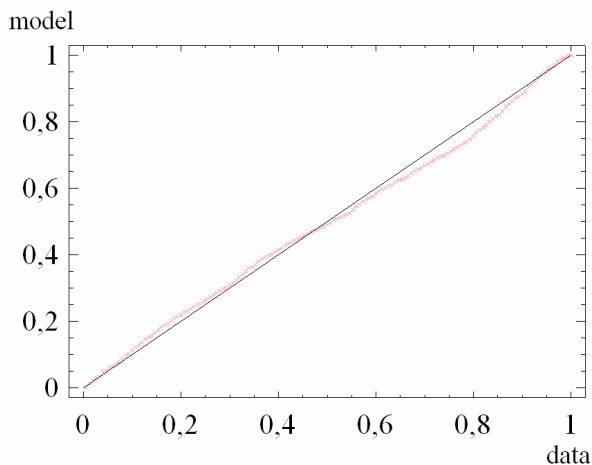
Podíl v počtu vyšetřených domácností s alespoň jedním dítětem vůči celkovému počtu zkoumaných domácností je nižší, než uvádí data z ČSÚ uvedená výše, protože předmětem zkoumání ani v tomto případě nebyly domácnosti důchodců bez ekonomické aktivity (z důvodů uvedených v první kapitole praktické části).



Graf 12: Shoda dat a ln modelu, skupina 0



Graf 13: Shoda dat a ln modelu, skupina 1



Graf 14: Shoda dat a ln modelu, skupina 2

Grafické výstupy 12-14 naznačují dobré výsledky lognormálního modelu, především pak u skupiny bez dětí. S tím také koresponduje výsledek testu shody, kdy se dosažená P-hodnota blíží 0,8. U třetí skupiny byly odchylky u lognormálního modelu příliš velké, stejně jako u ostatních modelů.

Rok 2005

Charakteristiky podskupin z roku 2005, uvedené v následující tabulce, ukazují na stejné proporcionalní uspořádání jednotlivých podskupin v obou letech. V jiných charakteristikách se však výrazněji liší. Zatímco u prvních dvou skupin se variabilita rozdělení vyjádřená variačním koeficientem snížila, u třetí zůstala více méně stejná a u poslední skupiny se prudce zvýšila. Tato skutečnost může být způsobena rozsahem skupiny, který byl pouhých 107 hodnot, a může zde tedy snáze dojít ke zkreslení.

Tabulka 19: Charakteristiky výb. souboru 2005 (členění podle počtu dětí), část první

skupina	0	podíl	1	podíl
počet	1461	53,85%	593	53,57%
podíl ve vzorku	52,24%	99,07%	21,20%	98,55%
průměr	287 433 Kč	106,17%	311 775 Kč	108,69%
medián	255 500 Kč	109,30%	274 015 Kč	110,09%
medián / průměr	88,89%	102,94%	87,89%	101,28%
dolní kvartil	182 923 Kč	111,21%	199 393 Kč	109,95%
horní kvartil	350 976 Kč	107,62%	372 807 Kč	108,41%
BES	261 225 Kč	109,06%	280 058 Kč	109,50%
sm. odchylka	176 078 Kč	95,30%	213 604 Kč	97,90%
variační koeficient	0,613	89,76%	0,685	90,07%
QR	168 053 Kč	103,97%	173 414 Kč	106,69%
RQD	0,315	95,53%	0,303	97,93%
šikmost	5,196	94,56%	5,626	50,38%
t	0,136	95,48%	0,139	82,63%
špičatost	64,369	95,45%	64,625	28,91%
bq	1,407	105,07%	1,491	111,80%

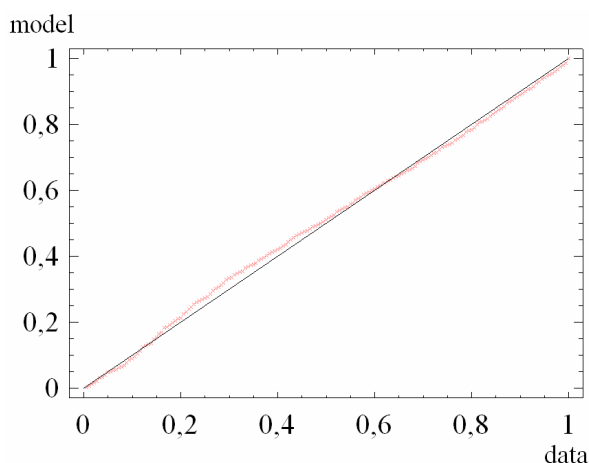
Tabulka 20: Charakteristiky výb. souboru 2005 (členění podle počtu dětí), část druhá

skupina	2	podíl	3	podíl
počet	619	54,92%	122	61,62%
podíl ve vzorku	22,13%	101,05%	4,36%	113,36%
průměr	336 257 Kč	110,02%	329 965 Kč	110,12%
medián	297 973 Kč	108,13%	281 911 Kč	103,13%
medián / průměr	88,61%	98,29%	85,44%	93,65%
dolní kvartil	228 464 Kč	108,09%	238 304 Kč	111,52%
horní kvartil	395 456 Kč	112,34%	360 456 Kč	105,39%
BES	304 967 Kč	109,45%	290 646 Kč	105,46%
sm. odchylka	177 405 Kč	117,03%	252 721 Kč	174,40%
variační koeficient	0,528	106,37%	0,766	158,37%
QR	166 992 Kč	118,73%	122 152 Kč	95,19%
RQD	0,268	107,21%	0,204	88,35%
šikmost	2,670	106,90%	7,120	225,00%
t	0,168	192,55%	0,286	407,34%
špičatost	11,688	105,09%	65,084	402,22%
bq	1,237	86,48%	1,791	153,09%

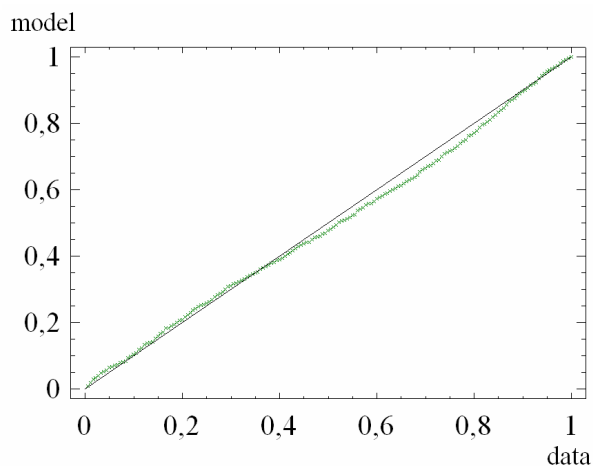
Z hlediska vhodnosti jednotlivých modelů bylo dosaženo jednoznačně nejlepšího výsledku u logaritmicko-normálního modelu u skupiny s 2 dětmi. Zde p-hodnota dokonce překročila 90%, což je výborný výsledek (graf 15). Dobrou aproximací dat jsou také modely gamma a lognormální u skupiny domácností bez dětí. Naopak u skupin s jedním dítětem a se třemi a více dětmi se ukázalo, že žádné z použitých typů rozdělení není použitelné.

Tabulka 21: Výsledky χ^2 testu shody

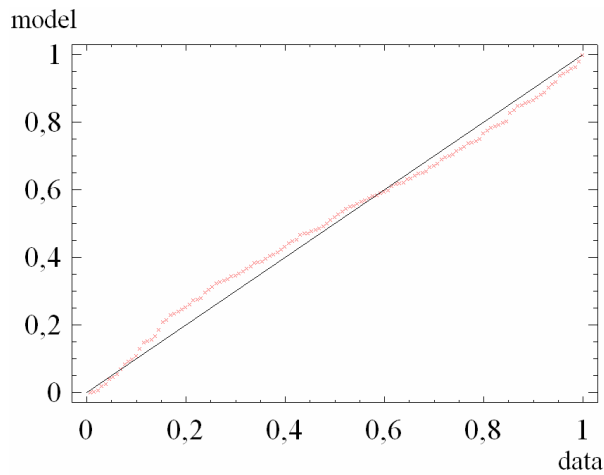
model	testové kritérium	kritická hodnota	p hodnota	závěr
0				
lognormální	61,4263	28,9647	0,0338	ano
gamma	59,3185	28,1440	0,0499	ano
Weibullovo	196,9337	28,1440	0,0000	ne
1				
lognormální	54,7397	18,4927	0,0038	ne
gamma	47,4265	17,7084	0,0226	ne
Weibullovo	112,4292	17,7084	0,0000	ne
2				
lognormální	18,9008	19,2806	0,9565	ano
gamma	39,1898	18,4927	0,1483	ne
Weibullovo	127,2392	18,4927	0,0000	ne
3				
lognormální	34,6708	7,9616	0,0044	ne
gamma	51,0937	7,2609	0,0000	ne
Weibullovo	104,6454	7,2609	0,0000	ne



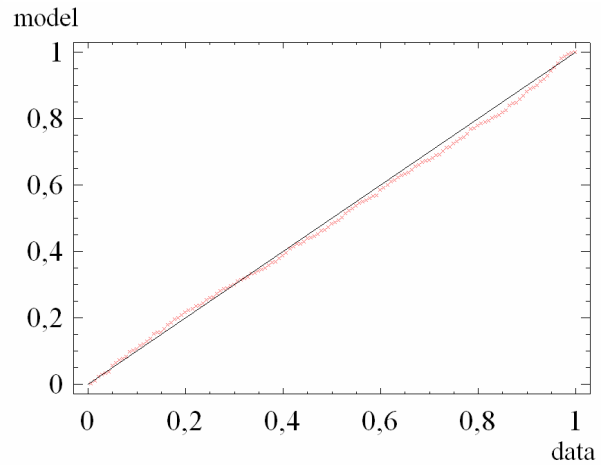
Graf 15: Shoda dat a ln modelu, skupina 0



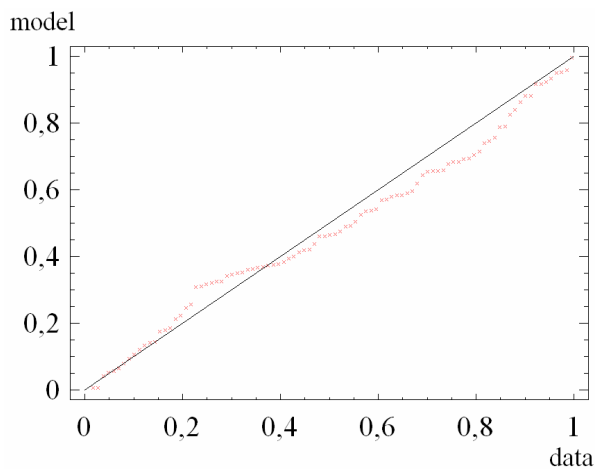
Graf 16: Shoda dat a gamma modelu, skupina 0



Graf 17: Shoda dat a ln modelu, skupina 1



Graf 18: Shoda dat a ln modelu, skupina 2



Graf 19: Shoda dat a ln modelu, skupina 3

Pokud srovnáme výsledky z obou let, nemůžeme vysledovat jasný společný znak. Vždy vycházel nejlépe model logaritmicko-normální, ovšem s velkým rozptylem výsledků. Zajímavým zjištěním je nízká závislost výše průměrného příjmu na počtu dětí, což naznačuje, že daňové, sociální a jiné úlevy zmíněné výše nehrají v příjmech domácnosti příliš významnou roli. To také možná souvisí s menší ochotou dnešních mladých lidí pořizovat si více dětí, protože by tím omezily svůj životní standart.

4 Domácnosti členěné dle vzdělání osoby v čele

Česká republika patří v rámci zemí OECD k zemím, které mají nadprůměrný počet obyvatel s ukončeným středním vzděláním (tj. lidí kteří absolvovali učiliště či střední školu), naopak výrazněji zaostává v počtu vysokoškolsky vzdělaných lidí⁵⁴. Tyto statistiky mají však velké úskalí v tom, že v každé zemi se jsou pojmy vysokoškolské a středoškolské vzdělání definovány trochu jinak. Některé profese (např. zdravotní sestry) bývají v některých státech zařazovány do terciárního vzdělávání, v jiných ne. V každém případě se stalo zvyšování počtu vysokoškoláků jednou z hlavních priorit posledních deseti let. Vývoj celkového počtu absolventů vysokých škol tomu také nasvědčuje. Z údajů vyplývá také větší příklon k dálkovému studiu, neboť vzrostl počet absolventů tohoto studia mezi roky 1997 – 2005 téměř čtyřnásobně.

Tabulka 22: Vývoj počtu absolventů vysokých škol v ČR⁵⁵

rok	prezenčně	kombinovaně	celkem
1997	20 942	2 447	23 389
1999	23 582	3 864	27 446
2001	24 866	4 290	29 156
2003	26 440	5 754	32 194
2004	29 950	7 444	37 394
2005	32 738	9 530	42 268

Výběrový soubor lze rozdělit do 5 skupin z hlediska nejvyššího dosaženého vzdělání. Počet domácností s osobou v čele bez vzdělání je však příliš nízký, proto došlo ke sloučení této se skupinou domácností s osobou v čele se základním vzděláním. Výsledné členění uvádí tabulka 23.

Tabulka 23: Rozdělení skupin (členění dle vzdělání)

skupina	Mikrocensus	SILC	dosažená úroveň vzdělání
1	1	1	bez vzdělání nebo nejvýše základní
2	2	2	vyučen
3	3	3,4	úplné střední, pomaturitní studium
4	4	5,6	vysokoškolské

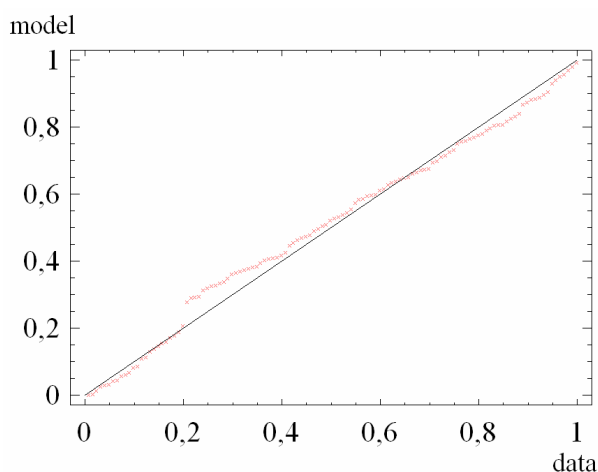
⁵⁴ zdroj: http://ihned.cz/3-22047730-oecc-000000_d-87, podle údajů z roku 2005 mělo v ČR 13% dospělé populace vysokoškolský titul, kdežto v průměr OECD byl 26%

⁵⁵ [http://www.czso.cz/csu/2007edicniplan.nsf/t/610030925C/\\$File/403607k03.pdf](http://www.czso.cz/csu/2007edicniplan.nsf/t/610030925C/$File/403607k03.pdf)

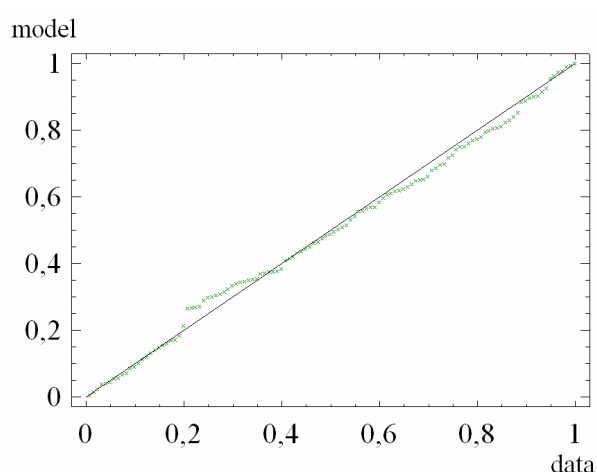
Tabulka 24: Charakteristiky výběrového souboru 2002 (členění podle vzdělání)

skupina	1	2	3	4
počet	294	2254	1936	784
podíl ve vzorku	5,58%	42,79%	36,75%	14,88%
průměr	206 595 Kč	249 855 Kč	291 214 Kč	387 272 Kč
medián	188 142 Kč	228 725 Kč	262 709 Kč	336 632 Kč
medián / průměr	91,07%	91,54%	90,21%	86,92%
dolní kvartil	143 553 Kč	175 146 Kč	187 667 Kč	242 401 Kč
horní kvartil	259 492 Kč	303 630 Kč	345 863 Kč	461 744 Kč
BES	194 832 Kč	234 057 Kč	264 737 Kč	344 352 Kč
sm. odchylka	106 749 Kč	116 987 Kč	188 687 Kč	283 092 Kč
variační koeficient	0,517	0,468	0,648	0,731
QR	115 939 Kč	128 484 Kč	158 196 Kč	219 343 Kč
RQD	0,288	0,268	0,297	0,312
šikmost	1,685	1,941	6,071	7,269
t	0,231	0,166	0,051	0,141
špičatost	5,626	9,733	78,921	103,214
bq	1,260	1,341	1,328	1,392

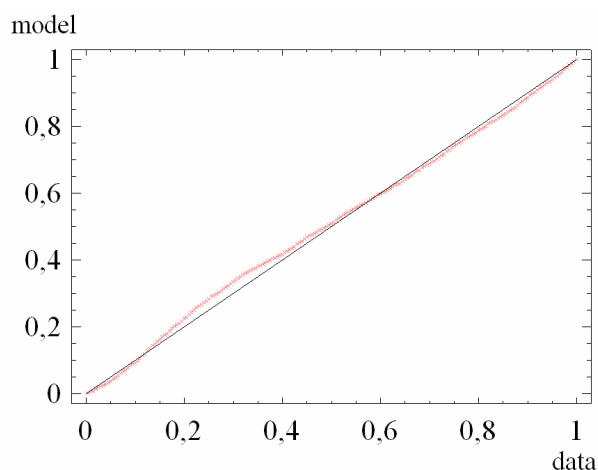
Charakteristiky rozdělení výběrových podskupin naznačují, že se zvyšujícím se vzděláním stoupá příjem domácnosti a tím i její životní úroveň. Markantní odstup vykazuje především průměr domácností s vysokoškolsky vzdělanou osobou v čele. I když samozřejmě příjmy osoby, která je podle pravidel Mikrocenzu uváděna jako osoba v čele, nemusí být největší (dokonce ani nejpodstatnější) částí příjmu dané domácnosti. Přesto však, i když připustíme možné zkreslení, je zde odstup tak velký, že vzdělání členů pravděpodobně bude mít na příjmy domácnosti zásadní vliv.



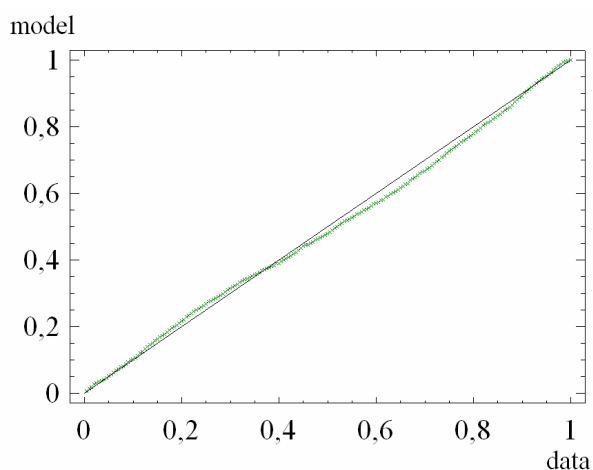
Graf 20: Shoda dat a ln modelu, skupina 1



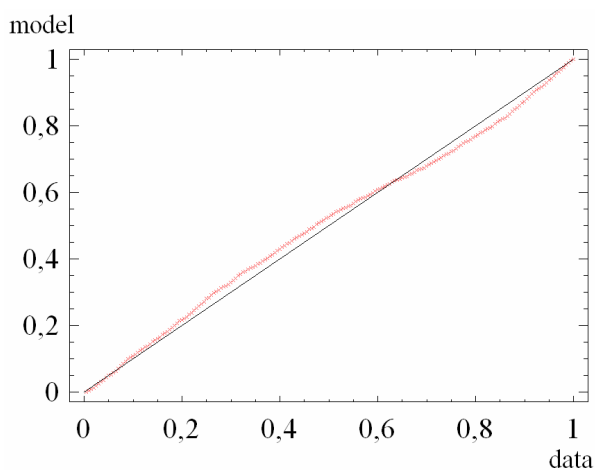
Graf 21: Shoda dat a gamma modelu, skupina 1



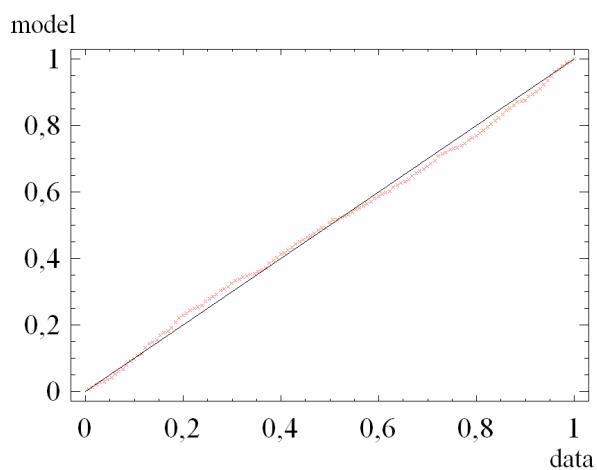
Graf 22: Shoda dat a ln modelu, skupina 2



Graf 23: Shoda dat a gamma modelu, skupina 2



Graf 24: Shoda dat a ln modelu, skupina 3



Graf 25: Shoda dat a ln modelu, skupina 4

Příjmy domácností s vysokoškolačkem v čele také vykazují největší variabilitu, což nejspíš způsobují vysoké příjmy části této podskupiny. Tomuto také napovídá nižší poměr mediánu a průměru. Podle výsledků v tabulce 25 nejsou zvolené modely příliš dobrou aproximací sledovaných dat. Oproti obvyklým výsledkům nejlepší míry shody dosáhl gamma model u skupiny 1. Grafické výsledky ovšem ukazují dobrou shodu u 2 a 4 skupiny, u skupiny 2 dokonce v případě dvou rozdělení (grafy 22 a 23). Dobře vyšel také lognormální model skupiny 3. V ostatních případech zvolené modely nevystihují vhodně výběrová data.

Tabulka 25: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
1				
lognormální	38,1556	13,0905	0,0245	ne
gamma	30,9559	12,3380	0,1239	ne
Weibullovo	46,0958	12,3380	0,0029	ne
2				
lognormální	77,1722	36,4371	0,0133	ano
gamma	75,1974	35,5999	0,0194	ano
Weibullovo	254,6623	35,5999	0,0000	ne
3				
lognormální	92,7285	33,9303	0,0002	ano
gamma	97,1725	33,0981	0,0001	ne
Weibullovo	352,7652	33,0981	0,0000	ne
4				
lognormální	50,3691	21,6643	0,0350	ano
gamma	73,3817	20,8665	0,0001	ne
Weibullovo	194,1942	20,8665	0,0000	ne

Rok 2005

Tabulka 26: Charakteristiky výb. souboru 2005 (členění podle vzdělání), část první

skupina	1	podíl	2	podíl
počet	166	56,46%	1258	55,81%
podíl ve vzorku	5,94%	106,36%	44,98%	105,13%
průměr	239 421 Kč	115,89%	278 871 Kč	111,61%
medián	208 830 Kč	111,00%	255 116 Kč	111,54%
medián / průměr	87,22%	95,78%	91,48%	99,93%
dolní kvartil	134 128 Kč	93,43%	195 589 Kč	111,67%
horní kvartil	277 600 Kč	106,98%	333 544 Kč	109,85%
BES	207 347 Kč	106,42%	259 841 Kč	111,02%
sm. odchylka	218 875 Kč	205,04%	140 973 Kč	120,50%
variační koeficient	0,914	176,92%	0,506	107,97%
QR	143 472 Kč	123,75%	137 955 Kč	107,37%
RQD	0,348	121,14%	0,261	97,15%
šikmost	41,952	355,66%	38,802	103,15%
t	-0,041	-17,91%	0,137	82,54%
špičatost	222,422	1129,46%	130,005	137,83%
bq	1,457	115,64%	1,399	104,27%

Ve výsledcích z roku 2005 došlo ke zmenšení skupiny domácností s osobou v čele se středním vzděláním a naopak ke zvětšení skupiny s vysokoškolským vzděláním. Tyto výsledky korespondují s trendem zvyšování počtu absolventů VŠ, který byl uveden výše. Zatímco v roce 2002 byl poměr medián/průměr nejnižší u poslední skupiny, což naznačovala větší nerovnost v příjmech oproti ostatním, v roce 2005 je tento poměr na úrovni ostatních skupin nebo mírně nad ním. To souvisí nejspíše s absencí odlehlých hodnot, jak jsem ji zmiňoval v předchozích kapitolách.

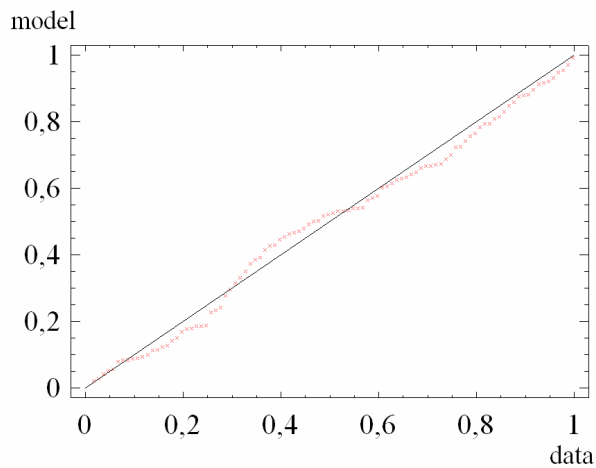
Tabulka 27: Charakteristiky výb. souboru 2005 (členění podle vzdělání), část druhá

skupina	3	podíl	4	podíl
počet	943	48,71%	428	54,59%
podíl ve vzorku	33,72%	91,75%	15,30%	102,84%
průměr	305 000 Kč	104,73%	408 977 Kč	105,60%
medián	271 088 Kč	103,19%	369 648 Kč	109,81%
medián / průměr	88,88%	98,53%	90,38%	103,98%
dolní kvartil	193 968 Kč	103,36%	273 296 Kč	112,75%
horní kvartil	363 150 Kč	105,00%	482 811 Kč	104,56%
BES	274 824 Kč	103,81%	373 851 Kč	108,57%
sm. odchylka	188 374 Kč	99,83%	256 534 Kč	90,62%
variační koeficient	0,618	95,32%	0,627	85,81%
QR	169 182 Kč	106,94%	209 515 Kč	95,52%
RQD	0,304	102,42%	0,277	88,95%
šikmost	5,395	88,85%	4,812	66,20%
t	0,088	172,21%	0,080	56,99%
špičatost	68,236	86,46%	40,863	39,59%
bq	1,422	107,08%	1,323	95,06%

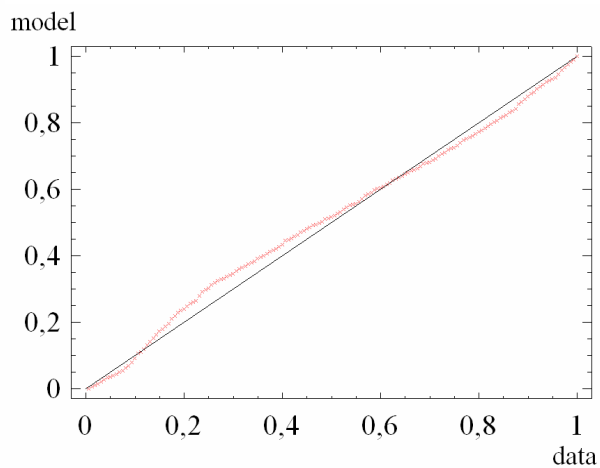
Tabulka 28: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
1				
lognormální	26,0146	9,3905	0,0994	ne
gamma	34,9652	8,6718	0,0095	ne
Weibullovo	66,8942	8,6718	0,0000	ne
2				
lognormální	96,4256	27,3256	0,0000	ano
gamma	70,2536	26,5093	0,0030	ano
Weibullovo	173,4330	26,5093	0,0000	ne
3				
lognormální	57,6857	23,2686	0,0124	ano
gamma	61,3212	22,4650	0,0053	ne
Weibullovo	173,2075	22,4650	0,0000	ne
4				
lognormální	31,9264	15,3792	0,1956	ne
gamma	27,6008	14,6114	0,3784	ne
Weibullovo	88,5434	14,6114	0,0000	ne

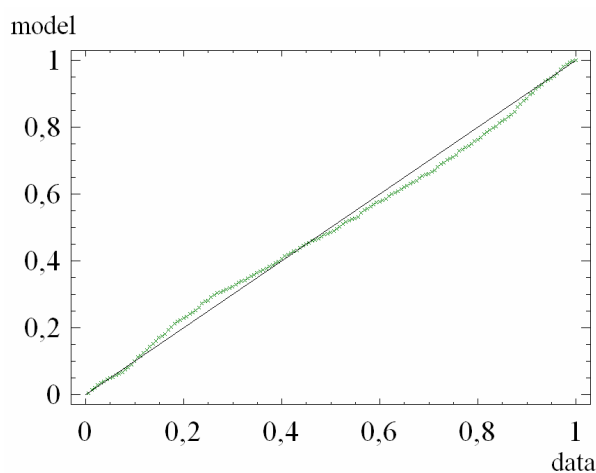
Výsledky testu shody se velmi blíží těm z roku 2002, a to jak v interpretaci restu shody, tak v grafické části.



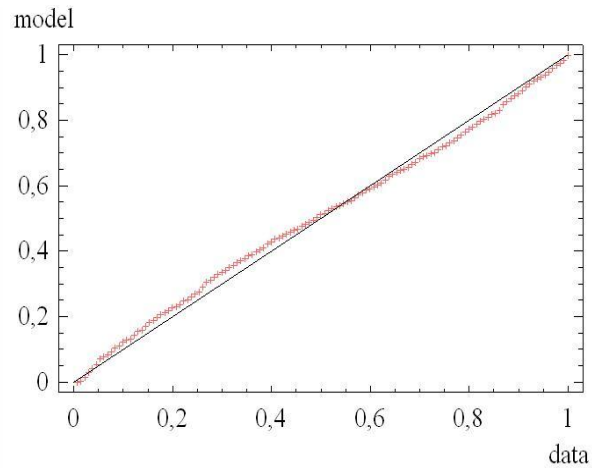
Graf 26: Shoda dat a ln modelu, skupina 1



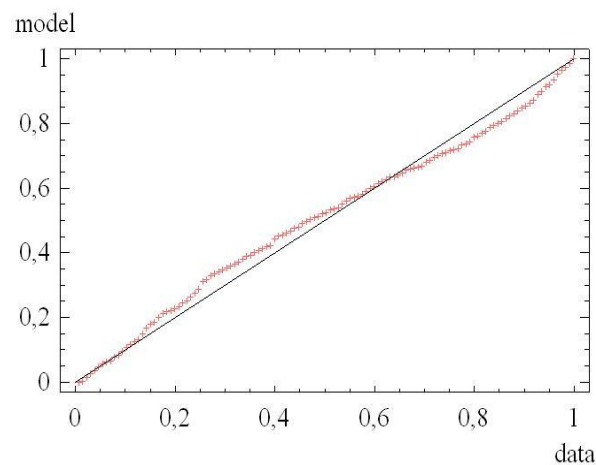
Graf 27: Shoda dat a ln modelu, skupina 2



Graf 28: Shoda dat a gamma modelu, skupina 2



Graf 29: Shoda dat a ln modelu, skupina 3



Graf 30: Shoda dat a ln modelu, skupina 4

5 Členění výběrového souboru do skupin podle krajů

Česká republika se od roku 2000 územně člení na 14 vyšších samosprávných celků, krajů. Toto členění je odlišné od předchozího krajského uspořádání, které zahrnovalo 8 oblastí⁵⁶. Hlavním důvodem k tomuto kroku bylo „uspokojení“ požadavků některých větších měst, ve kterých zatím krajské sídlo nebylo⁵⁷. Tímto zákonem vzniklo 14 krajů s velmi různou velikostí a u některých z nich také s velice problematicky vymezeným územím (Královéhradecký, Pardubický, Olomoucký). Zde krajské sídlo rozhodně netvoří přirozené epicentrum, což vedlo a stále vede k rozhořčení u obyvatel vzdálenějších obcí, kteří si (mnohdy oprávněně) stěžují na nutnost dlouhého cestování za vyřizováním svých úředních povinností. Druhým, zatím nedořešeným problémem, jsou zatím chybějící krajské státní zastupitelství a krajské soudy v některých krajích.



Obrázek 3 Kraje ČR, zdroj: www.kraj-jihocesky.cz

⁵⁶ Středočeský kraj se sídlem v Praze, Jihočeský kraj se sídlem v Českých Budějovicích, Západočeský kraj se sídlem v Plzni, Severočeský kraj se sídlem v Ústí nad Labem, Východočeský kraj se sídlem v Hradci Králové, Jihomoravský kraj se sídlem v Brně, Severomoravský kraj se sídlem v Ostravě, samostatnou jednotkou v mnoha ohledech postavenou na stejnou úroveň bylo území Hl. města Prahy.

⁵⁷ Především Liberec, Pardubice, Jihlava.

Přes tyto objektivní problémy se krajské uspořádání v České republice za sedm let od jeho vzniku zavedlo poměrně hladce a nově vzniklé vyšší samosprávné celky jsou dnes samozřejmou součástí samosprávné struktury. Jednotlivé kraje mají velmi různou ekonomickou úroveň a tím i rozdílnou výši příjmů obyvatel. Svou roli v tom hraje geografická poloha kraje, jeho napojení na dálniční tahy do západoevropských zemí a také podíl zemědělství, průmyslu a služeb na tvorbě HDP.

Vzhledem k velkému počtu krajů budou pro účely této práce vybrány jen čtyři kraje, a to zástupce menších krajů (Karlovarský), jeden středně velký (Olomoucký), jeden kraj ze skupiny velkých s více než milionem obyvatel (Moravskoslezský) a konečně kraj Praha jako speciální případ.

Kraj Praha má v rámci všech ostatních krajů v ČR výjimečné postavení. Jako jediný region vykazuje HDP na hlavu v paritě kupní síly⁵⁸ větší než průměr Evropské unie⁵⁹. V tomto kraji sídlí centrály většiny velkých a nadnárodních společností, většina velkých centrálních úřadů (Statistický úřad, Policejní prezidium, ministerstva apod.), a je zde více než kde jinde rozvinut sektor služeb na úkor především průmyslu. Všechny tyto aspekty vedou k tomu, že v Praze pobírají lidé mnohem vyšší příjmy, než v ostatních částech Česka⁶⁰. Příjmy jsou zde rozděleny také mnohem nerovnoměrněji (je zde poměrně velká skupina lidí s velmi vysokými příjmy). To vede jednak k tomu, že i celkové příjmy domácností jsou vysoko nad celkovým průměrem ČR, a dále k tomu, že medián příjmů výrazně zaostává za průměrným příjmem. Na druhou stranu mají ovšem domácnosti v Praze vyšší náklady (výdaje), především pak za bydlení. Ceny nových bytů v Praze se pohybují na začátku roku 2008 od 40000 Kč/m² výše a jsou tak nejdražší v České republice. Stejně tak neregulované nájemné strojů poměrně vysoko, což vysoké příjmy obyvatel v tomto kraji poněkud relativizuje.

Karlovarský kraj patří spolu s krajem Libereckým k nejmenším v České Republice. Ekonomiku zde ovlivňuje především dominantní obor lázeňství (především Karlovy Vary, Františkovy Lázně, Mariánské Lázně a lázně v Jáchymově) a z toho vyplývající příjmy

⁵⁸ PPP – Purchasing power parity, metodika měření HDP na hlavu, která zohledňuje různou úroveň cen ve srovnávaných státech, čímž daleko lépe vystihuje rozdíly v životní úrovni obyvatel

⁵⁹ podle údajů Eurostatu byl HDP v PPP v roce 2005 na úrovni 160% průměru EU-27

⁶⁰ Tato vysoká nerovnost mezi regiony vedla dokonce k tomu, že Praha, která představuje zhruba 15% pracovní síly v ČR, jako jediná vykazuje průměrnou mzdu vyšší než je celková průměrná mzda v České Republice

z cestovního ruchu. Častými návštěvníky západních Čech jsou turisté z Ruska⁶¹ a sousedícího Německa. Ve statistikách ČSÚ evidovaným HDP na obyvatele a průměrnou mzdou v kraji je Karlovarský kraj na jednom z posledních míst spolu s krajem Olomouckým (dosahuje cca. 76% (HDP), resp. 85% (mzdy) celorepublikové úrovně těchto veličen)

Kraj Olomoucký zahrnuje nejenom území Hané, rozprostírající se kolem centrálního města Olomouce, ale také Jeseníky u hranic s Polskem, které představují jednu z nejméně hospodářsky rozvinutých částí České republiky. K jeho specifikám také patří skutečnost, že zde nejméně lidí (společně s krajem Vysočina) žije ve městech⁶². Společně s výše zmíněným Karlovarským krajem má olomoucko nejnižší HDP na obyvatele a nejnižší průměrné mzdy⁶³, přičemž nejhůře je na tom právě okres Jeseník. V tomto okrese také přes značnou ekonomickou konjunkturu českého hospodářství stále zůstává dodnes (tj. do začátku roku 2008) značně vysoká míra nezaměstnanosti (kolem 11%).

Moravskoslezský kraj náleží k největším a nejlidnatějším krajům u nás. Před rokem 1989 zde byla silná tradice hutního průmyslu a těžkého strojírenství, což se stalo základem hospodářských problémů v druhé pol. 90. let. Tehdy se tyto obory hromadně restrukturalizovaly a došlo k masivnímu propouštění z dříve státních podniků. Kraj byl dlouho sužován vysokou nezaměstnaností a chybějící infrastrukturou, která by umožnila napojení na západoevropskou dálniční síť. Tyto problémy kulminovaly právě mezi sledovanými roky 2002 a 2005, kdy se míra registrované nezaměstnanosti v kraji pohybovala okolo 16%. V současnosti však zde vyrůstá několik nových průmyslových zón⁶⁴, vyrůstají zde nové dálniční úseky a zdá se, že obrovské problémy předchozích deseti let jsou minulostí a celková situace se začíná zlepšovat.

⁶¹ Že jsou západní Čechy častým cílem ruských turistů, je patrné zejména v Karlových Varech, kde má Rusko vlastní konzulát, místní letiště je spojeno pravidelnými lety s Moskvou.

⁶² podle údajů ČSÚ z roku 2002 jen něco kolem 58%

⁶³ čísla jsou přibližně totožná s Karlovarským krajem, tedy 85% průměru mezd a 76% průměru HDP na osobu

⁶⁴ V zónách investují především firmy s oborem lehkého strojírenství, výroba automobilových dílů či výroba spotřební elektroniky. Největší investicí se stala stavba automobilky Hyundai v obci Nošovice, která sama zaměstná několik tisícovek lidí.

Tabulka 29: Charakteristiky výběrového souboru 2002 (členění podle krajů)

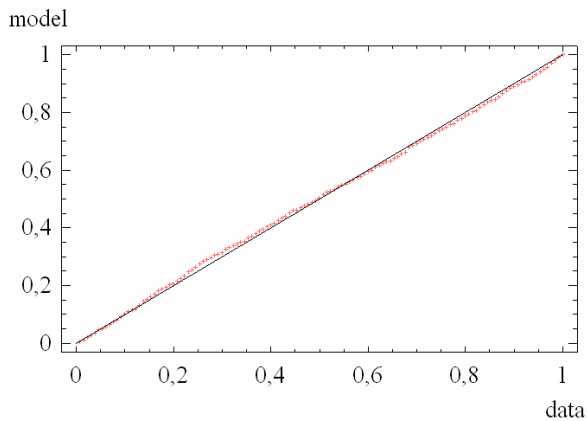
skupina	Praha	Karlovy Vary	Olomouc	Moravskoslezský
počet	1132	235	276	553
podíl ve vzorku	51,55%	10,70%	12,57%	25,18%
průměr	340 029 Kč	267 537 Kč	253 082 Kč	252 348 Kč
medián	287 456 Kč	228 018 Kč	226 193 Kč	236 808 Kč
medián / průměr	84,54%	85,23%	89,38%	93,84%
dolní kvartil	201 825 Kč	164 036 Kč	168 370 Kč	176 485 Kč
horní kvartil	415 557 Kč	299 599 Kč	308 455 Kč	309 245 Kč
BES	298 074 Kč	229 918 Kč	232 303 Kč	239 837 Kč
sm. odchylka	261 034 Kč	267 232 Kč	133 506 Kč	116 121 Kč
variační koeficient	0,768	0,999	0,528	0,460
QR	213 732 Kč	135 563 Kč	140 085 Kč	132 760 Kč
RQD	0,346	0,292	0,294	0,273
šikmost	7,290	9,479	2,267	1,603
t	0,199	0,056	0,174	0,091
špičatost	108,053	116,490	9,270	6,670
bq	1,384	1,404	1,254	1,253

Podrobnější ukazatele uvádí tabulka charakteristik těchto podskupin. Patrný je zde odstup příjmů obyvatel Prahy od ostatních oblastí. Překvapivě nízké hodnoty dosahuje kraj Moravskoslezský, který ve statistikách ČSÚ o průměrné hrubé mzdě zaměstnanců stojí výše než dva zbylé kraje. Ukazatel o poměru mediánu a průměru příjmů domácností také potvrzuje, že v Praze jsou největší rozdíly mezi chudými a bohatými, největší variabilita příjmů byla zjištěna v kraji Karlovarském.

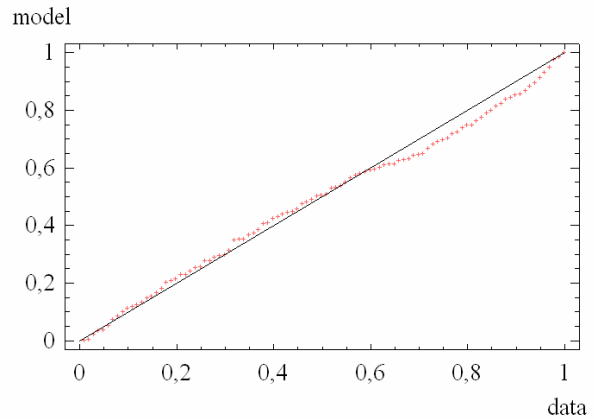
Tabulka 30: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
Praha				
lognormální	41,2549	25,6954	0,3723	ano
gamma	63,1817	24,8839	0,0085	ne
Weibullovo	195,4090	24,8839	0,0000	ne
Karlovy Vary				
lognormální	21,1857	11,5913	0,4477	ano
gamma	37,1517	10,8508	0,0162	ne
Weibullovo	102,2246	10,8508	0,0000	ne
Olomouc				
lognormální	23,7545	12,3380	0,3602	ano
gamma	21,3923	11,5913	0,4966	ne
Weibullovo	46,3164	11,5913	0,0018	ne
Moravskoslezský				
lognormální	48,4769	17,7084	0,0131	ano
gamma	31,8481	16,9279	0,3266	ano
Weibullovo	54,6820	16,9279	0,0027	ne

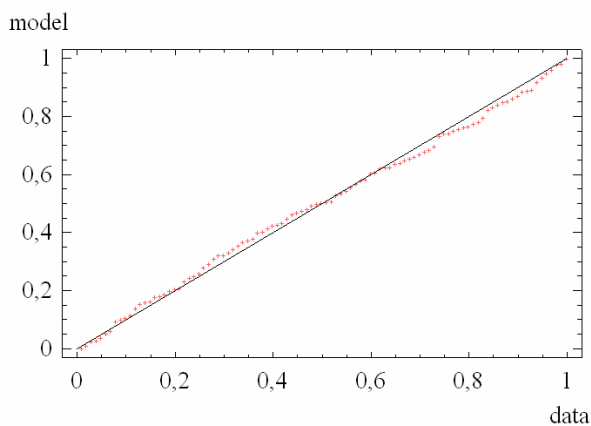
Testy shody hody naznačují rozdílné výsledky pro dvojice krajů. Zatímco v Praze a v Karlových Varech vychází nejlépe logaritmicko-normální model, pro Olomouc a severní Moravu vyšel jako nejvhodnější gamma model, a to s docela vysokou p-hodnotou. V případě grafického testu ovšem dobrou shodu dokazuje jen u Moravskoslezského kraje (průběh grafu 35), naopak v případě Olomouckého kraje se v grafické analýze ukázal jako lepší model lognormálního rozdělení.



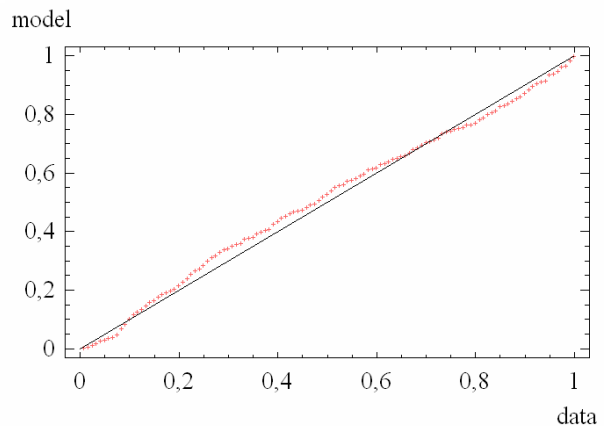
Graf 31: Shoda dat a ln modelu, Praha



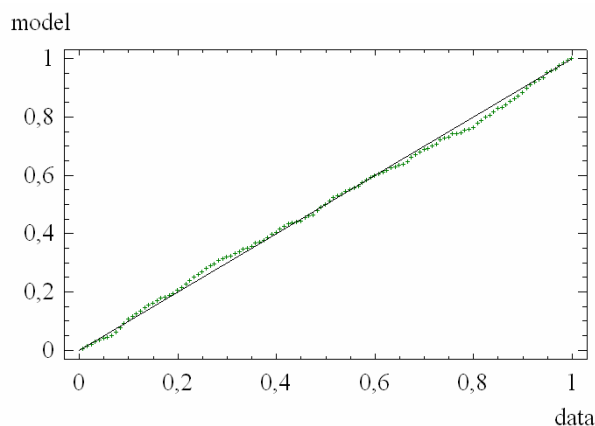
Graf 32: Shoda dat a ln modelu, Karlovy Vary



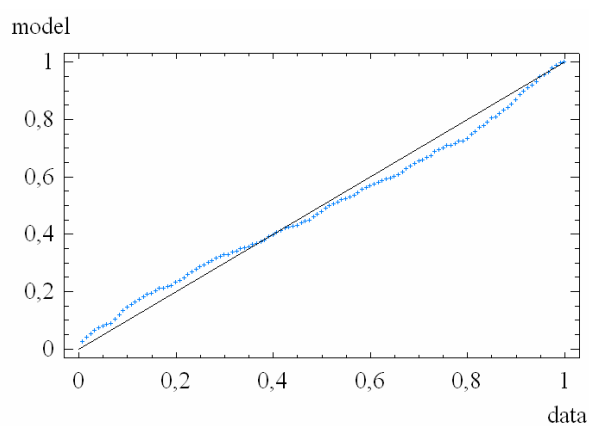
Graf 33: Shoda dat a ln modelu, Olomouc



Graf 34: Shoda dat a ln modelu, Moravs.



Graf 35: Shoda dat a gamma modelu, Moravs.



Graf 36: Shoda dat a Weibull modelu, Moravs.

Rok 2005

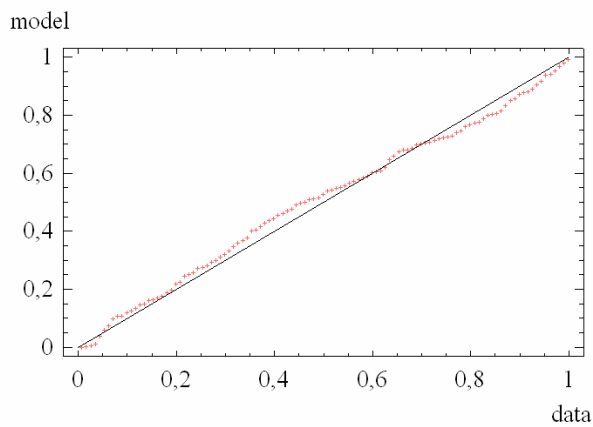
Tabulka 31: Charakteristiky výběrového souboru 2005 (členění podle krajů)

skupina	Praha	podíl	Olomoucký	podíl	Moravskoslezský	podíl
počet	327	28,89%	207	75,00%	366	66,18%
podíl ve vzorku	36,29%	70,40%	22,97%	182,79%	40,62%	161,30%
průměr	357 491 Kč	105,14%	297 816 Kč	117,68%	297 760 Kč	118,00%
medián	325 444 Kč	113,22%	250 140 Kč	110,59%	254 704 Kč	107,56%
medián / průměr	91,04%	107,69%	83,99%	93,98%	85,54%	91,15%
dolní kvartil	223 780 Kč	110,88%	195 589 Kč	116,17%	195 640 Kč	110,85%
horní kvartil	433 730 Kč	104,37%	347 000 Kč	112,50%	358 782 Kč	116,02%
BES	327 100 Kč	109,74%	260 717 Kč	112,23%	265 958 Kč	110,89%
sm. odchylka	195 780 Kč	75,00%	226 996 Kč	170,03%	178 297 Kč	153,54%
variační koeficient	0,548	71,34%	0,762	144,49%	0,599	130,13%
QR	209 950 Kč	98,23%	151 411 Kč	108,09%	163 142 Kč	122,88%
RQD	0,319	92,24%	0,279	94,98%	0,294	107,66%
šíkmost	2,000	27,44%	6,081	268,23%	2,958	184,51%
t	0,032	15,87%	0,279	160,17%	0,276	302,39%
špičatost	8,348	7,73%	55,654	600,33%	14,760	221,31%
bq	1,324	95,67%	1,371	109,33%	1,539	122,86%

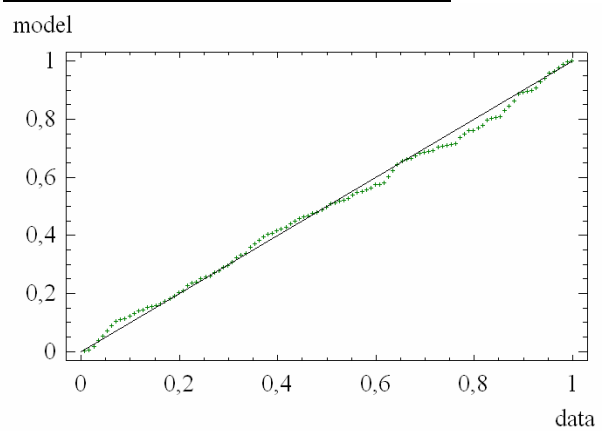
Výběrový soubor za rok 2005 pro Karlovarský kraj byl příliš malý (pouze 76 hodnot), a proto nebylo možné jej dobře statisticky zpracovat používanými metodami. V tomto roce se také výrazně snížil podíl vyšetřených domácností (na počtu předpokládaných) v Praze a tím i podíl tohoto kraje na celkovém výběru. Oproti výsledkům z roku 2002 došlo především ke snížení variability příjmů u prvních dvou krajů. Snížila se zde také šířka a špičatost, naopak u Moravskoslezského regionu se obě tyto charakteristiky zvýšily. Zajímavé jsou také poměry mediánu a průměru, které vyšly přesně v opačném pořadí než v předchozím šetření. Tento fakt je zřejmě způsoben tím, že u údajů sebraných v Praze chybí právě odlehlé hodnoty (zprava), tj. údaje za domácnosti s velmi vysokými příjmy.

Tabulka 32: Výsledky χ^2 testu shody

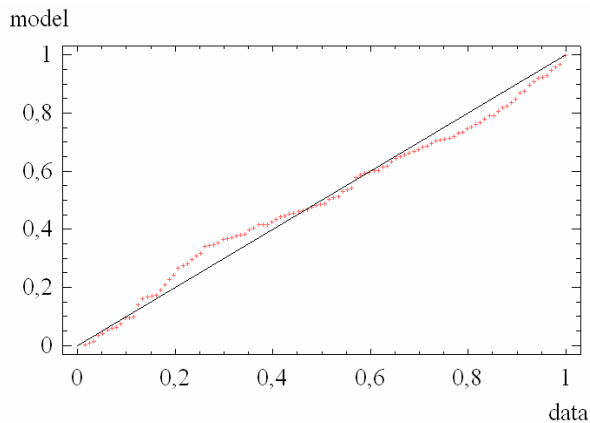
model	testové kritérium	kritická hodnota	p hodnota	závěr
Praha				
lognormální	38,5135	13,8484	0,0307	ano
gamma	36,0949	13,0905	0,0537	ano
Weibullovo	57,4738	13,0905	0,0001	ne
Olomouc				
lognormální	27,5515	10,8508	0,1204	ne
gamma	29,1907	10,1170	0,0841	ne
Weibullovo	67,4628	10,1170	0,0000	ne
Moravskoslezský				
lognormální	45,2205	14,6114	0,0079	ano
gamma	50,7087	13,8484	0,0017	ne
Weibullovo	99,0653	13,8484	0,0000	ne



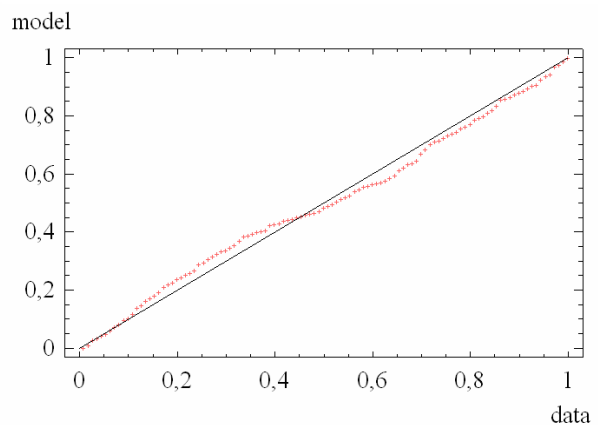
Graf 37: Shoda dat a ln modelu, Praha



Graf 38: Shoda dat a gamma modelu, Praha



Graf 39: Shoda dat a ln modelu, Olomouc



Graf 40: Shoda dat a ln modelu, Moravs.

To si však rozhodně nelze vykládat tím, že se v Praze během tří let snížila nerovnost v příjmech domácností (spíše tomu ve skutečnosti bude naopak). Respondenti s vysokými příjmy se v kraji Praha pravděpodobně zařadili mezi ty, co dotazník vůbec nevyplnili, nebo možná také úmyslně uvedli nižší částky než jaké odpovídají skutečnosti.⁶⁵

Změny v rovnoměrnosti rozdělení příjmů v případě Prahy ovlivnily také výsledky testu shody. Z grafů vyplývá, že dobrou aproximací výběrových dat je v tomto případě nejen lognormální rozdělení, ale i rozdělení gamma. To se osvědčilo také u Moravskoslezského kraje, naopak v případě kraje Olomouckého nevyhovuje žádný z použitých modelů.

⁶⁵ toto ČSÚ považuje za vážný problém zkreslující výsledky, nelze mu však nijak zabránit

6 Členění výběrového souboru podle pohlaví osoby v čele

Členění výběrového souboru podle pohlaví osoby v čele vychází ze zásad, které určil ČSÚ při sběru dat. V úplných rodinách (tedy tam, kde jsou manželé, nebo druh a družka) se bere jako osoba v čele domácnosti vždy muž, a to bez ohledu na to, zda má vyšší příjmy jako osoba než jeho manželka/družka. Také v případě, pokud je muž nezaměstnaný, nebo dokonce z nějakého důvodu nemá vůbec žádné příjmy, se označuje jako osoba v čele domácnosti. U neúplných rodin (jeden rodič s dítětem) se hodnotí v první řadě ekonomická aktivita osob, případně pak výše jejich příjmů (tato zásada platí i pro osoby nespojené manželským/partnerským vztahem, vztahem rodič – dítě nebo při společném hospodaření více rodin). Z toho plyne, že skupina domácností, v jejichž čele stojí muž (podle tohoto pravidla) je mnohem početnější než skupina domácností v čele se ženou.

Rok 2002

Tabulka 33: Charakteristiky výběrového souboru 2002 (členění podle pohlaví)

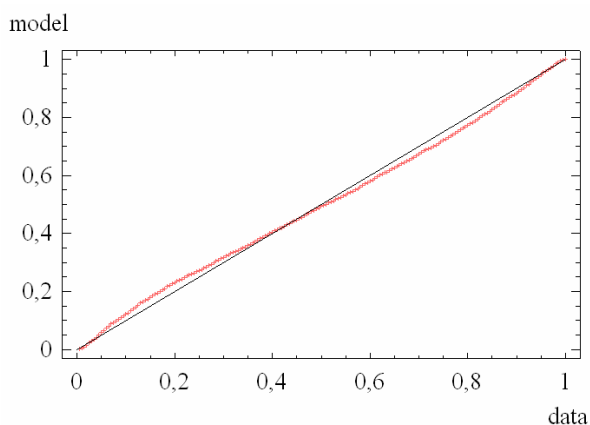
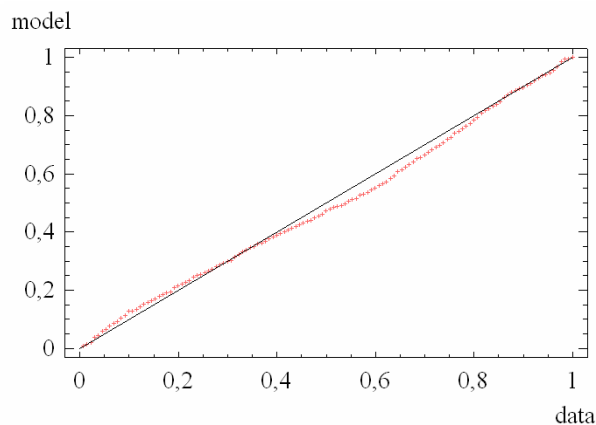
skupina	muž	žena
počet	4348	920
podíl ve vzorku	82,54%	17,46%
průměr	303 389 Kč	187 163 Kč
medián	267 752 Kč	157 336 Kč
medián / průměr	88,25%	84,06%
dolní kvartil	202 976 Kč	117 234 Kč
horní kvartil	355 511 Kč	219 915 Kč
BES	273 498 Kč	162 955 Kč
sm. odchylka	188 056 Kč	123 360 Kč
variační koeficient	0,620	0,659
QR	152 535 Kč	102 681 Kč
RQD	0,273	0,305
šíkmost	7,602	5,026
t	0,151	0,219
špičatost	138,385	51,239
bq	1,367	1,532

Z tabulky vyplývá, že domácnosti, v jejímž čele podle systému uvedeném výše stojí muž, mají podstatně vyšší průměrné příjmy než ty, v jejímž čele stojí žena. Jako možný důvod vidím fakt, že (právě vzhledem k definici osoby v čele) domácnosti v čele se ženou jsou obvykle neúplné rodiny s jednou ekonomicky aktivní osobou, kdežto v první skupině jsou obvykle dva ekonomicky aktivní členové. Míra nerovnosti, vyjádřená variačním koeficientem, se u obou skupin pohybuje v podstatě na stejné úrovni. Soubor domácností s mužem v čele překvapivě vykazuje vyšší poměr mediánu a průměru.

Tabulka 34: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
muž				
lognormální	145,7620	49,1623	0,0000	ano
gamma	291,8119	48,3054	0,0000	ne
Weibullovo	1094,5807	48,3054	0,0000	ne
žena				
lognormální	36,3298	23,2686	0,4533	ano
gamma	91,0746	22,4650	0,0000	ne
Weibullovo	245,3297	22,4650	0,0000	ne

Ani jedno z použitých rozdělení nevyhovuje podle testu shody u skupiny první, tedy domácností v čele s mužem. Na základě grafického testu však lze přijmout lognormální rozdělení. Dobré výsledky přinesla také aproximace lognormálním modelem u skupiny domácností v čele se ženou, jak naznačuje graf 42.

**Graf 41: Shoda dat a ln modelu, muž****Graf 42: Shoda dat a ln modelu, žena**

Rok 2005

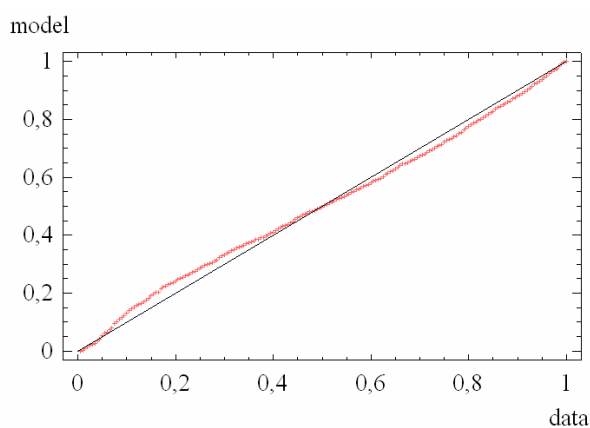
Údaje charakteristik za rok 2005 se v podstatě shodují s výsledky z roku 2002. Velmi podobné jsou poměry obou dvou skupin na celkovém počtu hodnot, nárůsty u průměrů a mediánů, velmi podobná zůstala i variabilita. Odlišné jsou především hodnoty šikmosti a špičatosti a s nimi související charakteristik. Míra shody empirického rozdělení hodnot s jednotlivými modely vyšla v roce 2005 také velmi podobně jako v roce 2002. Opět se potvrdilo, že jediný vyhovující model pro obě skupiny je logaritmicke-normální.

Tabulka 35: Charakteristiky výbě. souboru 2005 (členění podle pohlaví)

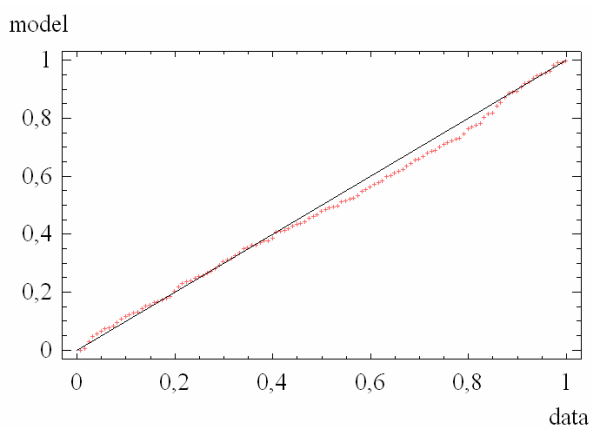
skupina	muž	podíl	žena	podíl
počet	2327	53,52%	469	50,98%
podíl ve vzorku	83,19%	100,80%	16,77%	96,01%
průměr	325 679 Kč	107,35%	203 577 Kč	108,77%
medián	289 547 Kč	108,14%	171 796 Kč	109,19%
medián / průměr	88,91%	100,74%	84,39%	100,39%
dolní kvartil	220 915 Kč	108,84%	125 740 Kč	107,26%
horní kvartil	385 406 Kč	108,41%	234 186 Kč	106,49%
BES	296 354 Kč	108,36%	175 880 Kč	107,93%
sm. odchylka	188 851 Kč	100,42%	159 252 Kč	129,10%
variační koeficient	0,580	93,55%	0,782	118,69%
QR	164 491 Kč	107,84%	108 446 Kč	105,61%
RQD	0,271	99,33%	0,301	98,93%
šikmost	5,156	67,82%	8,155	162,26%
t	0,166	109,85%	0,151	68,80%
špičatost	58,798	42,49%	113,015	220,56%
bq	1,386	101,33%	1,604	104,72%

Tabulka 36: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
muž				
lognormální	127,2453	36,4371	0,0000	ano
gamma	165,3966	35,5999	0,0000	ne
Weibullovo	529,4281	35,5999	0,0000	ne
žena				
lognormální	24,7411	16,1514	0,5890	ano
gamma	59,2835	15,3792	0,0003	ne
Weibullovo	150,0277	15,3792	0,0000	ne



Graf 43: Shoda dat a ln modelu, muž



Graf 44: Shoda dat a ln modelu, žena

7 Členění domácností podle sociální skupiny osoby v čele

Výběrové šetření Mikrocenzu z roku 2002 se v tomto sledovaném znaku liší od pozdějšího šetření SILC 2005. Zatímco u prvního případu bylo nadefinováno 9 sociálních skupin, v šetření SILC se jedná o 7 sociálních skupin, z nichž některé agregují skupiny předchozí, jiné naopak dále člení dříve jednotnou skupinu. Aby mohla být data porovnána, seskupil jsem hodnoty do 5 nových skupin, sestavených podle klíče v tabulce 34. Sociální skupina „důchodce v domácnosti bez ekonomicky aktivních členů“ bude analyzována podrobněji v následující kapitole.

Tabulka 37: Rozdělení skupin

skupina	skupiny Mikrocenzus	skupiny SILC
1	dělník, zaměstnanec, zemědělec, ostatní	nižší zaměstnanec, vyšší zaměstnanec, ostatní
2	samostatně činný – mimo zemědělství samostatně činný – v zemědělství	samostatně činný
3	nezaměstnaný	nezaměstnaný
4	důchodce v domácnosti s ekonomicky aktivními členy	důchodce v domácnosti s ekonomicky aktivními členy
	důchodce v domácnosti bez ekonomicky aktivních členů	důchodce v domácnosti bez ekonomicky aktivních členů

Toto členění domácností se oproti ostatním, dále v textu uvedeným, liší ve výchozích předpokladech o tvaru (průběhu) výběrového rozdělení u jednotlivých skupin. Zatímco se při členění domácností podle vzdělání osoby v čele, počtu nezaopatřených dětí atd. rozdělení příjmu modeluje rozdělením logaritmicko-normálním (lognormální), gamma a Weibullovým, zde se v případě sociální skupiny nezaměstnaný volí mezi modelem lognormálním, gamma, a normálním. U této skupiny se totiž předpokládá rovnoměrnější průběh hodnot příjmů, protože dominantní část příjmů tvoří sociální dávky. Ty se sice odvíjejí od výše předešlého příjmu, nicméně má určité stropy.⁶⁶ Tento předpoklad má ovšem úskalí v tom, že příjmy osoby v čele domácnosti nemusí tvořit největší část jejího disponibilního důchodu.⁶⁷

⁶⁶ Vyplácená částka je omezena výší 0,58 násobku průměrné mzdy v 1 – 3 čtvrtletí předcházejícího kalendářního roku (zdroj: MPSV)

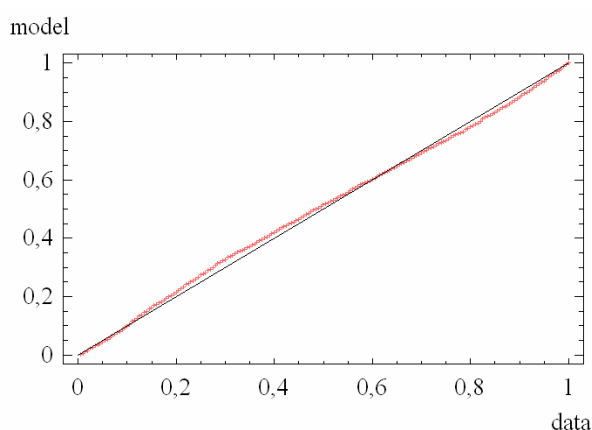
⁶⁷ Jak bylo uvedeno výše, osoba v čele domácnosti se neurčuje na základě ek. aktivity, nýbrž většinou podle pohlaví. Proto může mít i zde rozdělení průběh odpovídající spíše lognormálnímu modelu. Z tohoto důvodu byly pro modelování vybrány jen ty domácnosti s nezaměstnaným v čele, které zároveň nemají žádnou ekonomicky aktivní osobu

Rok 2002

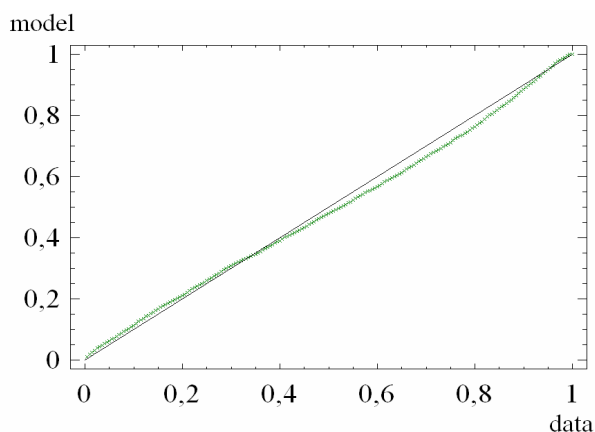
Tabulka 38: Charakteristiky výb. souboru 2002 (členění podle soc. skupiny)

skupina	1	2	1	4
počet	4233	757	139	278
podíl ve vzorku	78,29%	14,00%	2,57%	5,14%
průměr	273 465 Kč	340 373 Kč	96 547 Kč	273 682 Kč
medián	244 530 Kč	285 476 Kč	88 813 Kč	252 305 Kč
medián / průměr	89,42%	83,87%	91,99%	92,19%
dolní kvartil	176 697 Kč	205 890 Kč	58 140 Kč	203 088 Kč
horní kvartil	330 662 Kč	391 939 Kč	126 888 Kč	312 972 Kč
BES	249 105 Kč	292 195 Kč	90 664 Kč	255 168 Kč
sm. odchylka	173 136 Kč	245 684 Kč	49 525 Kč	108 650 Kč
variační koeficient	0,633	0,722	0,513	0,397
QR	153 965 Kč	186 049 Kč	68 748 Kč	109 884 Kč
RQD	0,303	0,311	0,372	0,213
šikmost	8,427	3,882	1,414	2,051
t	0,119	0,144	0,108	0,104
špičatost	185,422	24,850	4,377	6,462
bq	1,317	1,604	0,961	1,315

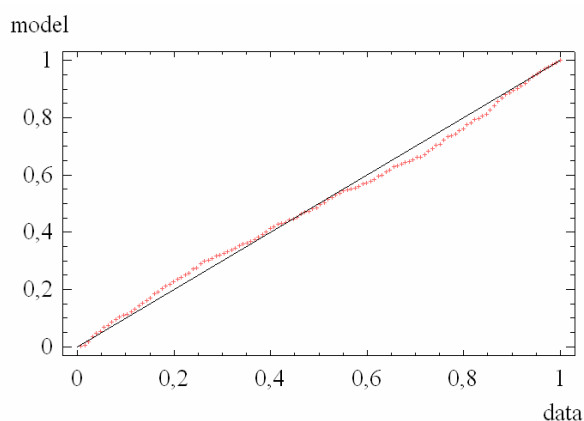
Největší podíl mezi vyšetřenými hodnotami zaujímají domácnosti v čele se zaměstnancem. Výsledky čtvrté skupiny, tj. důchodce v čele domácnosti hospodařící společně s jinými, ekonomicky aktivními členy, jsou v charakteristikách polohy velice obdobné. Liší se pouze ve vyšší hodnotě mediánu a z toho vyplývajícím užším kvartilovém rozpětí (což naznačuje rovnoměrnější rozdělení příjmů). Průměrem i mediánem ztelně zaostávají domácnosti v čele s nezaměstnaným, což se ovšem předpokládá, vzhledem k tomu že pro hlavní část této skupiny tvoří nejpodstatnější část příjmu sociální dávky.



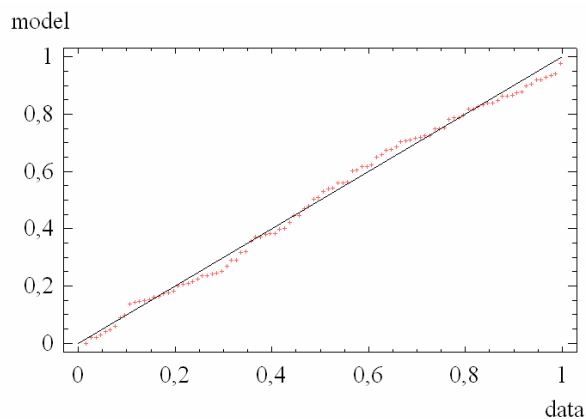
Graf 45: Shoda dat a ln modelu, skupina 1



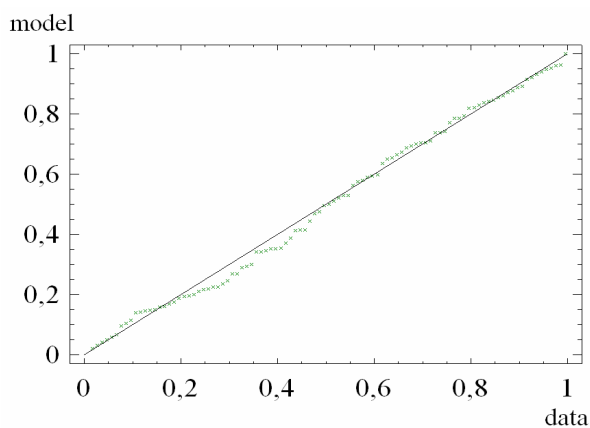
Graf 46: Shoda dat a gamma modelu, skupina 1



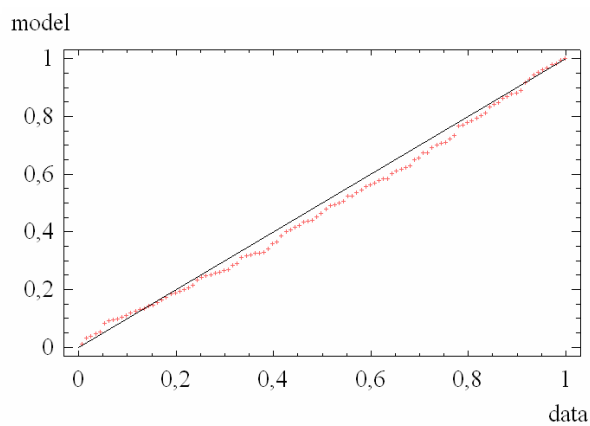
Graf 47: Shoda dat a ln modelu, skupina 2



Graf 48: Shoda dat a ln modelu, skupina 3



Graf 49: Shoda dat a gamma modelu, skupina 3



Graf 50: Shoda dat a ln modelu, skupina 4

Tabulka 39: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
1				
lognormální	113,5364	49,1623	0,0003	ano
gamma	153,5334	48,3054	0,0000	ano
Weibullovo	683,3889	48,3054	0,0000	ne
2				
lognormální	55,7538	20,8665	0,0079	ano
gamma	97,2938	20,0719	0,0000	ne
Weibullovo	195,2247	20,0719	0,0000	ne
3				
lognormální	21,9626	8,6718	0,1862	ano
gamma	22,7178	7,9616	0,1587	ano
normální	49,3698	27,5871	0,0001	ne
4				
lognormální	27,5622	12,3380	0,1906	ano
gamma	40,2691	11,5913	0,0101	ne
Weibullovo	91,5396	11,5913	0,0000	ne

Nejlepší výsledky zde prokázalo rozdělení lognormální. P-hodnoty v testu shody vycházejí sice opět velice nízké, nicméně průběhy grafů naznačují dobrou shodu teoretického modelu a dat. Ostatní typy rozdělení se ukázaly jako nevhodné, a to i rozdělení normální v případě skupiny nezaměstnaný. Nepotvrdil se tak můj původní předpoklad o tom, že právě tento model bude pro tuto homogenní příjmovou skupinu charakteristický.

Rok 2005

Tabulka 40: Charakteristiky výb. souboru 2005 (členění podle soc. skupiny), část první

skupina	1	podíl	2	podíl
počet	2226	52,59%	391	51,65%
podíl ve vzorku	76,90%	98,23%	13,51%	96,48%
průměr	289 690 Kč	105,93%	393 784 Kč	115,69%
medián	259 058 Kč	105,94%	321 380 Kč	112,58%
medián / průměr	89,43%	100,01%	81,61%	97,31%
dolní kvartil	190 440 Kč	107,78%	227 086 Kč	110,29%
horní kvartil	359 336 Kč	108,67%	454 506 Kč	115,96%
BES	266 973 Kč	107,17%	331 088 Kč	113,31%
sm. odchylka	157 838 Kč	91,16%	320 449 Kč	130,43%
variační koeficient	0,545	86,06%	0,814	112,74%
QR	168 896 Kč	109,70%	227 420 Kč	122,24%
RQD	0,307	101,23%	0,334	107,21%
šikmost	3,974	47,16%	4,216	108,61%
t	0,187	157,72%	0,171	118,20%
špičatost	57,620	31,08%	25,566	102,88%
bq	1,355	102,83%	1,398	87,15%

Podíl jednotlivých skupin na celkovém výběru se změnil minimálně. Nárůst u domácností s nezaměstnaným v čele souvisí spíše s vyšším procentem úspěšnosti šetření u této skupiny než s tím, že by se během období 2002 – 2005 zvyšoval počet nezaměstnaných. Spíše tomu bylo naopak, nezaměstnanost se díky ekonomickému růstu snižovala.

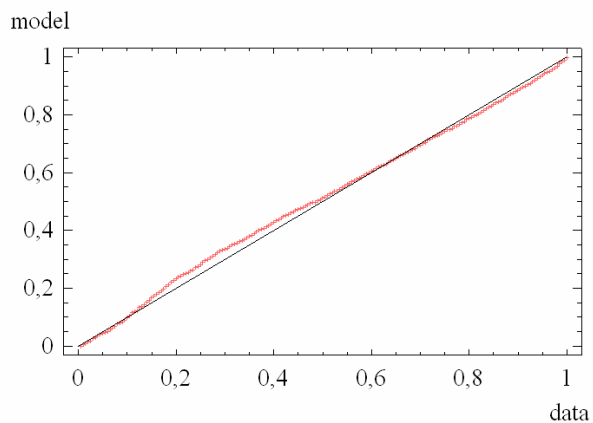
Tabulka 41: Charakteristiky výb. souboru 2005 (členění podle soc. skupiny), část druhá

skupina	3	podíl	4	podíl
počet	98	70,50%	178	64,03%
podíl ve vzorku	3,39%	131,69%	6,15%	119,60%
průměr	98 932 Kč	102,47%	305 632 Kč	111,67%
medián	85 810 Kč	96,62%	295 364 Kč	117,07%
medián / průměr	86,74%	94,29%	96,64%	104,83%
dolní kvartil	55 354 Kč	95,21%	245 211 Kč	120,74%
horní kvartil	129 112 Kč	101,75%	347 381 Kč	110,99%
BES	89 022 Kč	98,19%	295 830 Kč	115,94%
sm. odchylka	57 989 Kč	117,09%	92 955 Kč	85,55%
variační koeficient	0,586	114,27%	0,304	76,61%
QR	73 758 Kč	107,29%	102 170 Kč	92,98%
RQD	0,400	107,61%	0,172	80,97%
šikmost	1,692	119,66%	1,299	63,33%
t	0,174	161,76%	0,018	17,51%
špičatost	4,823	110,19%	2,872	44,45%
bq	1,005	104,59%	1,430	108,68%

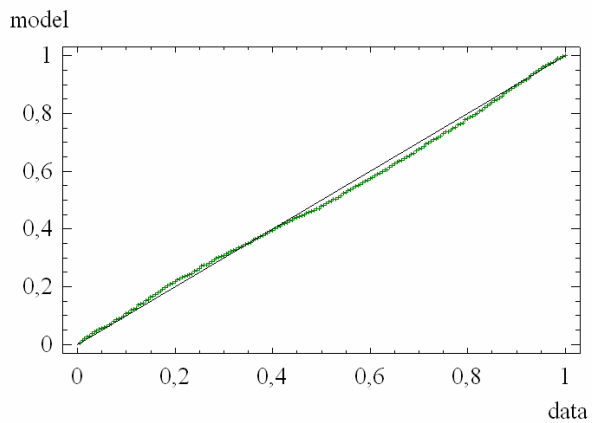
Tabulka 42: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
1				
lognormální	82,0798	35,5999	0,0038	ano
gamma	69,1708	34,7643	0,0459	ano
Weibullovo	231,8324	34,7643	0,0000	ne
2				
lognormální	31,7572	14,6114	0,1652	ano
gamma	56,7837	13,8484	0,0003	ne
Weibullovo	116,8015	13,8484	0,0000	ne
3				
lognormální	10,3046	6,5706	0,7396	ano
gamma	12,3764	5,8919	0,5761	ne
normální	33,7957	23,6848	0,0022	ne
4				
lognormální	26,6363	9,3905	0,0861	ano
gamma	26,0653	8,6718	0,0983	ne
Weibullovo	49,1647	8,6718	0,0001	ne

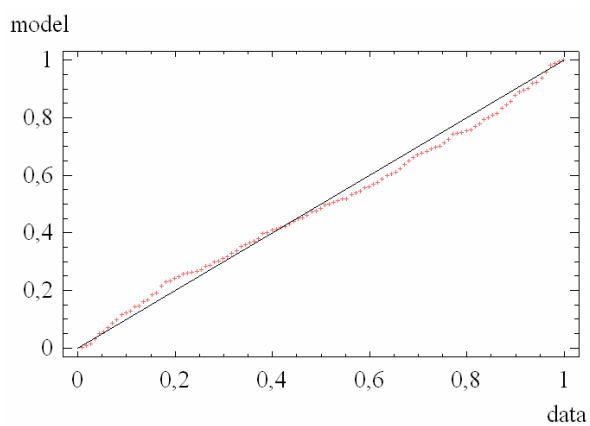
Výsledky korelují s výsledky z roku 2002. U třetí skupiny, tj. u domácností v čele s nezaměstnaným, bylo dokonce dosaženo vynikajícího výsledků shody modelu, a to tradičně lognormálního modelu.



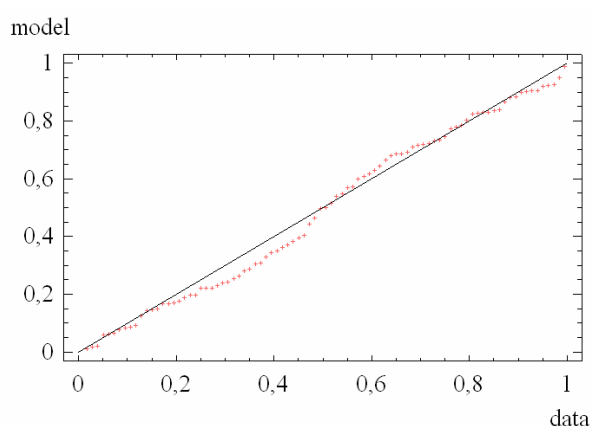
Graf 51: Shoda dat a ln modelu, skupina 1



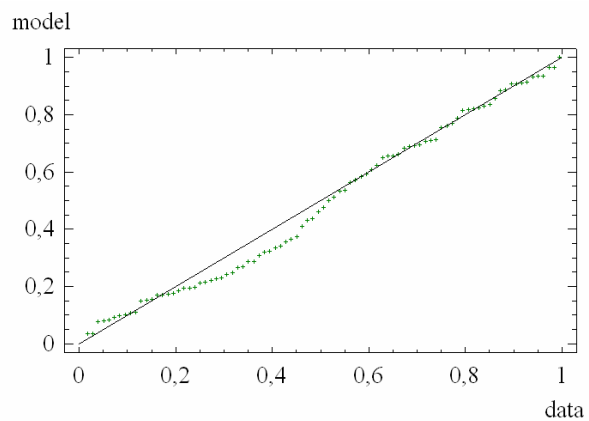
Graf 52: Shoda dat a gamma modelu, skupina 1



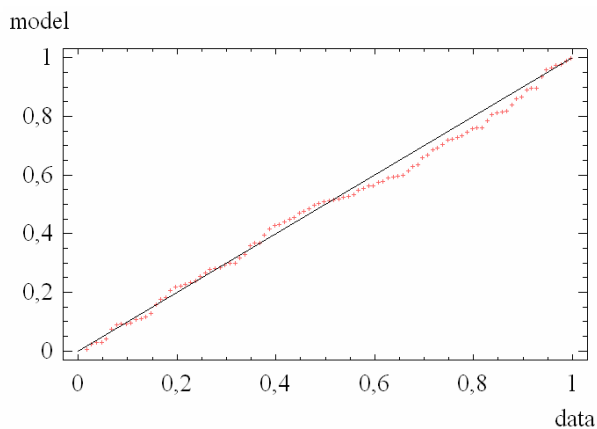
Graf 53: Shoda dat a ln modelu, skupina 2



Graf 54: Shoda dat a ln modelu, skupina 3



Graf 55: Shoda dat a gamma modelu, skupina 3



Graf 56: Shoda dat a ln modelu, skupina 4

8 Domácnosti důchodců bez ekonomicky aktivního člena

Tato skupina patří v posledních letech k nejsledovanějším a nejdiskutovanějším skupinám obyvatelstva nejenom u nás, ale i v ostatních zemích Evropské unie. Nepříznivý demografický vývoj Evropy poslední dekády staví vlády těchto zemí před problém, jak nalézt udržitelný systém financování důchodů při zachování solidní životní úrovně penzistů. V České republice se zdá být tento problém o to vážnější, že ani dnes není životní standart velké většiny důchodců nikterak vysoký. Průměrné důchody se pohybují okolo 55% čisté průměrné mzdy, což je mnohem méně než ve vyspělých západních zemích, kde není neobvyklý poměr okolo 70 – 80%⁶⁸. U nás jsou ovšem důchody mnohem rovnostářštější, jejich rozdělení nekopíruje rozdělení mezd. To je dáno především způsobem, jakým se důchody vypočítávají⁶⁹.

Z tohoto důvodu má také rozdělení příjmů domácností důchodců bez ekonomicky aktivního člena jiný průběh, než bývá obvyklé u rozdělení příjmů domácností se členy s ekonomickou aktivitou. Oproti výše používaným třem rozdělením proto nahradíme v následující analýze model lognormální modelem normálním. Vypustíme také model Weibullův a přibude naopak model Laplacův.

Počet členů domácností důchodců, kteří žijí v domácnosti bez dalších ekonomicky aktivních členů, je obvykle 1 nebo 2. Z tohoto důvodu rozčlením výběrový soubor na skupinu jednočlenných domácností a dvou- a více členných domácností. V prvním případě ještě rozliším muže a ženy. Výsledek uvádí tabulka 43.

Tabulka 43: Rozdělení skupin

skupina	definice
1	jednočlenná domácnost - muž
2	jednočlenná domácnost - žena
3	dvou a vícečlenná domácnost

⁶⁸ www.mesec.cz

⁶⁹ Pro rok 2008 se vypočte z měsíčního průměru průměrných mezd z let 1986 – 2007, upravených o koeficient přepočtu zohledňující inflaci; z této částky se do výpočtového základu počítá 100% částky do 10000 Kč, 30% z 10000 – 24800Kč, a pouze 10% z částky nad 24800Kč. Takto vypočtený základ se poté násobí počtem celých odpracovaných let a koeficientem 0,015. K takto vypočtenému důchodu se přičte pevná částka 1700 Kč. (zdroj: ČSSZ)

Tabulka 44: Charakteristiky výběrového souboru 2002

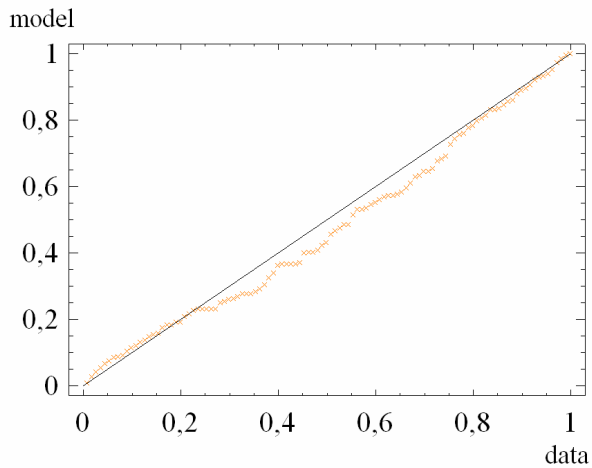
skupina	1	2	3
počet	331	1037	1089
podíl ve vzorku	13,47%	42,21%	44,32%
průměr	93 979 Kč	89 949 Kč	166 545 Kč
medián	92 400 Kč	87 000 Kč	164 040 Kč
medián / průměr	98,32%	96,72%	98,50%
dolní kvartil	84 000 Kč	78 000 Kč	151 560 Kč
horní kvartil	101 808 Kč	96 000 Kč	179 286 Kč
BES	92 652 Kč	87 000 Kč	164 732 Kč
sm. odchylka	19 255 Kč	27 482 Kč	51 418 Kč
variační koeficient	0,205	0,306	0,309
QR	17 808 Kč	18 000 Kč	27 726 Kč
RQD	0,096	0,103	0,084
šikmost	1,732	8,688	17,850
t	0,057	0,000	0,100
špičatost	9,665	108,608	459,162
bq	1,413	1,194	1,286

Charakteristiky polohy, zjištěné analýzou, odpovídají původním předpokladům o relativně rovnoměrném rozdělení příjmů důchodců. Svědčí o tom především hodnota poměru mediánu a průměru blížící se 100%. Dalším specifikem, podporujícím toto tvrzení, je nízká hodnota směrodatné odchylky a z toho vyplývající nízký variační koeficient. Konečně jako poslední indikátor této skutečnosti se ukazuje extrémně nízké kvartilové rozpětí. Ve vzorku se objevilo třikrát více jednočlenných domácností žen než mužů. To ukazuje na to, že ženy se obvykle dožívají vyššího věku a jejich zastoupení mezi jednočlennými domácnostmi důchodců je tedy vyšší než u mužů. Především u domácnosti s dvěma důchodci (tj. skupina 3) dosahuje velmi vysokých hodnot špičatost, což naznačuje značné nahuštění hodnot okolo střední hodnoty (jak už také vyplývá právě z charakteristik popsanych výše).

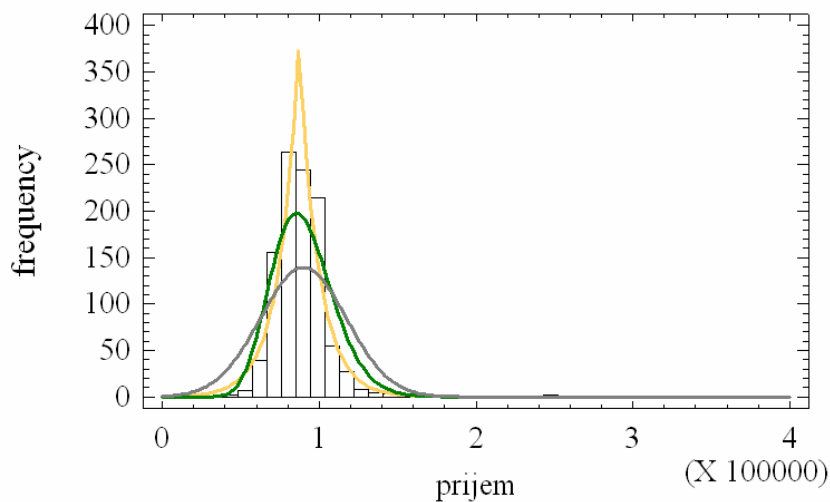
Tabulka 45: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
1				
normální	846,2904	13,8484	0,0000	ne
gamma	741,5962	13,0905	0,0000	ne
laplace	804,3135	13,0905	0,0000	ne
2				
normální	1209,9886	24,8839	0,0000	ne
gamma	623,5014	24,0749	0,0000	ne
laplace	507,0643	24,0749	0,0000	ne
3				
normální	1735,3634	24,8839	0,0000	ne
gamma	861,1841	24,0749	0,0000	ne
laplace	562,4106	24,0749	0,0000	ne

Ani jeden ze zvolených modelů nevystihuje dostatečně přesně některý z výběrového souboru dat. Kritické hodnoty v testu shody byly překročeny mnohonásobně, p-hodnoty u všech testů nulové. Tyto výsledky tentokrát plně potvrdila grafická analýza pomocí P-P grafů. Tento výsledek způsobuje především extrémní špičatost rozdělení (viz graf 58 histogramu), které nelze dostatečně přesně aproximovat zvolenými modely.



Graf 57: Shoda dat a Laplaceova modelu, skupina 1



Graf 58: Histogram skupiny 2

Tabulka 46: Charakteristiky výběrového souboru 2005

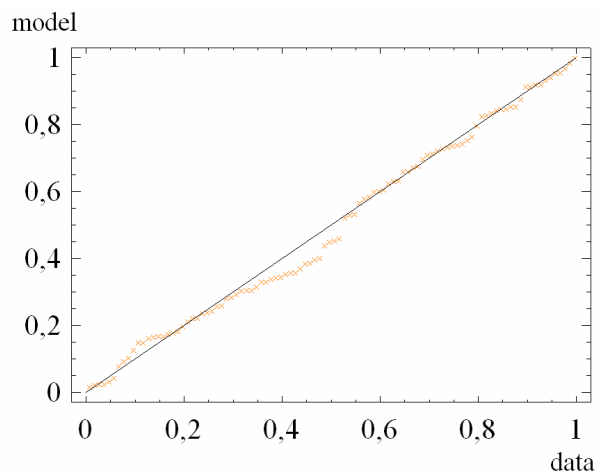
skupina	1	podíl	2	podíl	3	podíl
počet	167	50,45%	597	57,57%	661	60,70%
podíl ve vzorku	11,71%	86,89%	41,85%	99,15%	46,33%	104,53%
průměr	103 300 Kč	109,92%	97 240 Kč	108,11%	188 937	113,45%
medián	101 800 Kč	110,17%	95 800 Kč	110,11%	182 000	110,95%
medián / průměr	98,55%	100,23%	98,52%	101,86%	96,33%	97,80%
dolní kvartil	92 200 Kč	109,76%	86 200 Kč	110,51%	165 464	109,17%
horní kvartil	113 000 Kč	110,99%	105 028 Kč	109,40%	198 000	110,44%
BES	102 200 Kč	110,31%	95 707 Kč	110,01%	181 866	110,40%
sm. odchylka	20 590 Kč	106,93%	18 818 Kč	68,47%	79 113	153,86%
variační koeficient	0,199	97,29%	0,194	63,34%	0,419	135,63%
QR	20 800 Kč	116,80%	18 828 Kč	104,60%	32 536	117,35%
RQD	0,101	105,76%	0,098	95,18%	0,090	106,82%
šikmost	0,411	23,74%	2,170	24,97%	16,750	93,84%
t	0,077	135,90%	-0,020	-	-0,016	-
špičatost	2,424	25,08%	13,378	12,32%	366,057	79,72%
bq	1,200	84,95%	1,406	117,75%	1,460	113,52%

Výsledné charakteristiky vzorku z roku 2005, uvedené v tabulce 46, korespondují s předchozími z roku 2002. Podíl tří vytvořených podskupin skupin na výběrovém souboru se v podstatě výrazněji nezměnil, stejně jako podíl mediánu a průměru (došlo k poklesu pouze u třetí skupiny). Ke změnám došlo u charakteristik variability v domácnostech důchodkyň-žen a v domácnostech důchodců s více osobami. U prvně jmenované skupiny poklesla a u druhé vzrostla. U skupiny 2 se také velmi výrazně snížila špičatost.

Tabulka 47: Výsledky χ^2 testu shody

model	testové kritérium	kritická hodnota	p hodnota	závěr
1				
normální	50,0803	9,3905	0,0001	ne
gamma	68,8575	8,6718	0,0000	ne
laplace	37,1602	8,6718	0,0050	ano
2				
normální	104,5586	18,4927	0,0000	ne
gamma	69,5090	17,7084	0,0001	ne
laplace	73,3883	17,7084	0,0000	ne
3				
normální	864,3045	19,2806	0,0000	ne
gamma	283,8339	18,4927	0,0000	ne
laplace	160,0982	18,4927	0,0000	ne

Závěry analýzy dat z roku 2005 se téměř shodují s těmi z roku 2002. Opět se ukázalo, že modely spíše nevyhovují. Pouze v případě skupiny 1 lze použít model Laplaceův, jak naznačuje grafický průběh shody (graf 59).



Graf 59: Shoda dat a Laplaceova modelu, skupina 1

Vyhodnocení

Souhrnný přehled výsledků statistického zjišťování, popsaného na stránkách výše, uvádím v následující tabulce. Jak je vidět na první pohled, nejlepších výsledků bylo dosaženo při aproximaci dat lognormálním rozdělením.

Tabulka 48: Souhrnné výsledky testů χ^2 shody

model	gamma	lognormální	Weibull	normální	Laplace
Ano	6	34	0	0	1
Ne	41	7	39	8	5
%	14,89%	82,92%	0,00%	0,00%	16,67%

9 Analýza závislostí

Úkolem další části mé diplomové práce je provedení regresní analýze závislosti příjmů domácností v České republice na vybraných faktorech. Cílem je zjistit, které ze zvolených faktorů ovlivňují výši příjmů, a které nikoliv. Celkem jsem zvolil sedm faktorů, které uvádí tabulka 49. Všechny tyto faktory mají kvalitativní nebo diskrétní kvantitativní charakter, a proto bylo nutné provést binarizaci původních proměnných a vytvořit celkem 21 nových, umělých proměnných. Počet umělých proměnných vychází u každého faktoru z počtu znaků, které proměnná nabývá, zmenšených o 1⁷⁰.

Tabulka 49

faktor	znak	umělá proměnná
pohlaví	muž	a
	žena	
vzdělání	nejvýše základní	b ₃ b ₂ b ₁
	vyučen	
	úplné střední	
	vyšší a vysokoškolské	
počet dětí	0	c ₁
	1	c ₂
	2	c ₃
	3 a více	
oblasti ⁷¹	Střed	d ₁
	Jihozápad	d ₂
	Severozápad	d ₃
	Východ	d ₄
	Morava – jih Morava - sever	d ₅
typ obce	krajská města	e ₁
	městské obce	e ₂
	vesnické obce	
sociální skupina osoby v čele	zaměstnanec	f ₁
	samostatně činný	f ₂
	důchodce s EA	f ₃
	důchodce bez EA nezaměstnaný	f ₄
počet EA osob v domácnosti	0	g ₁
	1	g ₂
	2	g ₃
	3 a více	

⁷⁰ Při stejném počtu umělých proměnných jako znaků faktoru by v modelu nastala lineární závislost

⁷¹ Jednotlivé oblasti zahrnují následující kraje: Střed (Praha, Středočeský), Jihozápad (Jihočeský, Plzeňský), Severozápad (Karlovarský, Ústecký, Liberecký), Východ (Pardubický, Vysočina, Královehradecký), Morava – jih (Zlínský, Jihomoravský), Morava – sever (Moravskoslezský, Olomoucký)

Rok 2002

Tabulka 50: Kvalita dat

Porucha modelu	Test	závěr
Multikolinearita	Correlations (-0,70102;0,499491)	OK
Autokorelace	Durbin – Watson statistic = 1,91518	OK

Nejprve jsem provedl ověření kvality dat, tj. zjišťování, zda v souboru dat nefiguruje některá z poruch, uvedená v teoretické části. V modelu se vyskytla pozitivní multikolinearita mezi faktory f_4 a g_1 , a proto jsem faktor g_1 vypustil. Porucha byla způsobena velkou blízkostí skupin, které se z velké části překrývají. Domácnost v čele s důchodcem bez ekonomicky aktivních členů (f_4) se řadí vždy také do skupiny g_1 (což je skupina, zahrnující domácnosti bez členů s ekonomickou aktivitou). Výsledky upravené, které uvádí tabulka 50, prokázaly, že data jsou v pořádku.

Tabulka 51: Výstup Statgraphics pro rok 2002

Multiple Regression Analysis, dependent variable: příjem

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	121286,0	12417,3	9,76747	0,0000
a	57991,6	3740,82	15,5024	0,0000
b1	108265,0	6338,74	17,0799	0,0000
b2	36965,9	5178,28	7,13864	0,0000
b3	-3114,4	4880,21	-0,638169	0,5234
c1	-42266,0	9140,12	-4,62422	0,0000
c2	-26359,9	9535,29	-2,76446	0,0057
c3	-20402,8	9549,58	-2,13651	0,0326
d1	38130,9	4965,72	7,67882	0,0000
d2	6958,51	5827,59	1,19406	0,2325
d3	8118,62	5211,42	1,55785	0,1193
d4	2698,93	5420,54	0,497909	0,6185
d5	-3977,03	5405,81	-0,735696	0,4619
e1	7990,0	4319,64	1,84969	0,0644
e2	-1059,33	3729,22	-0,284061	0,7764
f1	240850,0	9213,57	26,1408	0,0000
f2	284586,0	10220,5	27,8445	0,0000
f3	283949,0	12456,0	22,7962	0,0000
f4	-13301,8	8242,65	-1,61378	0,1066
g2	-200707,0	6525,22	-30,7586	0,0000
g3	-114125,0	6623,98	-17,229	0,0000

R-squared = 40,0468 percent

R-squared (adjusted for d. f.) = 39,896 percent

Tabulka 52: Výstup Statgraphics pro rok 2002, forward selection

Multiple Regression Analysis, dependent variable: příjem

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	89017,1	4730,45	18,8179	0,0000
a	58155,9	3669,8	15,8471	0,0000
b1	109912,0	4866,57	22,5851	0,0000
b2	39348,2	3401,84	11,5667	0,0000
c1	-21818,1	3562,37	-6,1246	0,0000
d1	35702,1	3804,26	9,38476	0,0000
e1	7740,04	3714,55	2,08371	0,0372
f1	250113,0	6493,43	38,5179	0,0000
f2	293852,0	7880,24	37,2897	0,0000
f3	292983,0	10517,5	27,8566	0,0000
g2	-198720,0	6467,02	-30,7282	0,0000
g3	-112261,0	6562,97	-17,1052	0,0000

R-squared = 39,9126 percent

R-squared (adjusted for d. f.) = 39,8296 percent

Výsledky regresní analýzy prokázaly, že výše příjmu domácnosti nejvíce závisí na pohlaví osoby v čele, dále na jejím vzdělání, sociální skupině a počtu ekonomicky aktivních členů. Příjmy také významně ovlivňuje to, zda domácnost sídlí v Praze či nikoliv. Naopak jako nevýznamná se ukázala být adresa domácnosti (s výjimkou Prahy), a to jak z pohledu oblasti, tak z pohledu velikosti (typu) obce. Výši příjmů také nijak významně neovlivňuje počet dětí (pokud v domácnosti nějaké žijí).

Rok 2005

Tabulka 53: Kvalita dat

Porucha modelu	Test	závěr
Multikolinearita	Correlations (-0,6891;0,4541)	OK
Autokorelace	Durbin – Watson statistic = 1,89268	OK

Také v datech z roku 2005 se prokázala závislost mezi f_4 a g_1 příliš silná (92%), a proto jsem opět vynechal faktor g_1 (domácnosti se členy bez ekonomické aktivity).

Tabulka 54: Výstup Statgraphics pro rok 2005

Multiple Regression Analysis, dependent variable: příjem

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	161640,0	16244,6	9,95037	0,0000
a	58680,2	5171,29	11,3473	0,0000
b1	113478,0	8905,56	12,7424	0,0000
b2	32817,5	7255,77	4,52295	0,0000
b3	1449,97	6814,99	0,212762	0,8315
c1	-56502,4	12192,8	-4,63409	0,0000
c2	-34326,6	12802,5	-2,68123	0,0073
c3	-19864,3	12793,5	-1,55268	0,1205
d1	22799,7	6467,71	3,52516	0,0004
d2	-4835,92	7397,85	-0,653693	0,5133
d3	3155,67	6893,83	0,457754	0,6471
d4	-16650,3	6858,2	-2,42779	0,0152
d5	-14832,2	6873,59	-2,15785	0,0309
e1	-3078,72	5713,25	-0,538873	0,5900
e2	-6122,7	4834,9	-1,26635	0,2054
f1	253291,0	11574,5	21,8836	0,0000
f2	331357,0	13135,9	25,2252	0,0000
f3	319664,0	15692,5	20,3706	0,0000
f4	-10813,7	10544,7	-1,02551	0,3051
g2	-217934,0	8847,11	-24,6334	0,0000
g3	-118640,0	9032,13	-13,1353	0,0000

R-squared = 42,9133 percent

R-squared (adjusted for d. f.) = 42,6497 percent

Tabulka 55: Výstup Statgraphics pro rok 2005, forward selection

Multiple Regression Analysis, dependent variable: příjem

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	134502,0	7608,43	17,6781	0,0000
a	59596,2	5026,05	11,8574	0,0000
b1	110951,0	6603,33	16,8024	0,0000
b2	30766,9	4656,96	6,60664	0,0000
c1	-41818,2	5936,28	-7,0445	0,0000
c2	-17865,8	7203,12	-2,48028	0,0131
d1	23129,5	5354,55	4,31959	0,0000
d4	-16001,6	5981,47	-2,67519	0,0075
d5	-13867,7	5977,81	-2,31986	0,0203
f1	261315,0	8708,48	30,0069	0,0000
f2	338814,0	10736,6	31,5569	0,0000
f3	327751,0	13605,3	24,0899	0,0000
g2	-216352,0	8715,13	-24,8249	0,0000
g3	-118114,0	8957,95	-13,1853	0,0000

R-squared = 42,8343 percent

R-squared (adjusted for d. f.) = 42,663 percent

Ve výsledcích roku 2005 bylo dosaženo v podstatě obdobných výsledků jako na datech o tři roky starších.

Závěr

Domácnosti jako celek tvoří velkou a současně značně nehomogenní skupinu. Aproximace takového nesourodého výběrového souboru pomocí zvolených modelu se ukázala jako velmi nepraktická. Průběh hodnot empirického souboru dat tvoří v histogramu dvouvrcholové, obtížně vystihnutele rozdělení. Hlavním „viníkem“ tohoto jevu je skupina domácností důchodců bez ekonomické aktivity. Příjmy této skupiny samostatně mají oproti ostatním subpopulacím (skupinám výběrového souboru) netradiční, velmi špičatý průběh rozdělení, charakteristický úzkým variačním rozpětím. Tato vlastnost příjmových rozdělení byla dokázána v šetřeních i ve vyspělých státech západní Evropy. V České republice tento stav zvýrazňuje velice solidární výpočet starobních důchodů (velká skupina lidí pobírá přibližně stejný důchod) a fakt, že starobní penze tvoří dominantní část příjmů těchto domácností (narozdíl právě od některých vyspělých zemí, které mají mnohem více rozvinutý systém penzijního připojištění, spořicíh programů apod.).

Členění výběrového souboru na jednotlivé, homogennější podskupiny, se ukázalo jako cesta k možnosti použití některého ze známých teoretických modelů rozdělení. V této práci jsem se zaměřil především na ověření vhodnosti logaritnicko-normálního modelu se dvěma parametry, alternativně pak na rozdělení gamma a Weibullovo. U některých skupin jsem použil speciálně model normální a Laplacův. Jako nejlepší aproximace výběrových dat se ukázalo rozdělení lognormální, a potvrdilo tak původní očekávání, neboť v současné době právě tento model patří v těchto analýzách k těm nejpreferovanějším.

K ověřování shody jsem použil test χ^2 shody a grafické porovnání pomocí P-P grafů (probability-probability), s důrazem na grafickou shodu. Z celkového počtu 41 testů vyhovělo lognormální rozdělení ve 34 případech, tj. téměř v 83%. Ostatní používaná rozdělení se ukázala jako vhodná jen ve velmi malém počtu případů (gamma rozdělení), popř. dokonce vůbec v žádném (normální a Weibullovo rozdělení). U skupiny domácností v čele s důchodcem bez ekonomické aktivity nenabízí ani jeden z těchto modelů dostatečně dobré vystižení datového vzorku. Relativně nejlépe vystihuje tento případ model Laplacův. Pokud je tedy dodržen předpoklad vyčlenění skupiny domácností se členy bez ekonomické aktivity ze základního výběrového souboru, lze použít logaritnicko-normální rozdělení jako dobrý nástroj pro další analýzy.

V druhé části této práce, tj. při analýze faktorů ovlivňujících výši celkových příjmů domácností, jsem použil metodu regresní analýzy s binarizací nezávislých (vysvětlujících) proměnných. Zatímco v plánovaně řízených ekonomikách závisela výše příjmu domácností

kvůli nízké mzdové diferenciaci především na počtu ekonomicky aktivních členů (popř. celkovém počtu členů domácnosti a počtu dětí), tak ve vyspělých ekonomikách západní Evropy převládají faktory kvalitativní (především vzdělání). V České ekonomice bude s růstem příjmové diferenciaci růst i význam těchto faktorů. Podle provedené analýzy je významným faktorem počet ekonomicky aktivních osob v domácnosti, ale výše příjmů se také odvíjí od pohlaví osoby v čele, jejího dosaženého vzdělání a sociální skupiny. Naopak celkem nevýznamným se ukázal počet dětí a adresa bydliště domácnosti (tj. kraj, popř. typ obce), s výjimkou Prahy.

Závěrem bych se chtěl zmínit o zdrojových datech jako nástroji používaném ve výše uvedených postupech. Kvalita výsledku a přesnost odhadů jednotlivých modelů závisí vedle dodržení všech stanovených postupů také na kvalitě výběrového souboru. Jak vyplývá ze statistik míry úspěšnosti ve sběru dat v jednotlivých krajích, ochota občanů sdělovat citlivé informace o svých příjmech není na optimální úrovni (někdy jen okolo 60%). Navíc srovnání mezi lety 2002 a 2005 naznačuje mírně sestupný trend počtu vyšetřených domácností. Proto se domnívám, že velkou výzvou pro Český statistický úřad v nejbližších letech bude přesvědčit občany k uvádění co nejúplnějších a pravdivých informací o výši a struktuře jejich příjmů i ostatních sledovaných statistik.

Seznam použité a prostudované literatury

Knihy

1. **HENDL, J.:** *Přehled statistických metod zpracování dat: Analýza a metaanalýza dat.* Praha, Portál 2006. ISBN 80-7367-123-9.
2. **BARTOŠOVÁ, J.:** *Volba a aplikace metod analýzy stavu rozdělení příjmů v České republice po roce 1990.* Praha, Fakulta informatiky a statistiky VŠE 2006.
3. **CYHELSKÝ, L. – KAHOUNOVÁ, J. – HINDLS, R.:** *Elementární statistická analýza.* Praha, Management Press 1999. ISBN 80-7261-003-1
4. **MELOUN, M. – MILITKÝ, J.:** *Statistické zpracování experimentálních dat v chemometrii, biometrii, ekonometrii a v dalších oborech přírodních, technických a společenských věd.* Praha, East Publishing 1998. ISBN 80-7219-003-2.
5. **HINDLS, R. – HRONOVÁ, S. – SEGER, J.:** *Statistika pro ekonomy.* Praha, Professional Publishing 2004. ISBN 80-86419-59-2
6. **HÁTLE, J. – KAHOUNOVÁ, J.:** *Úvod do teorie pravděpodobnosti.* Praha, SNTL/Alfa 1987. 04-319-87
7. **JAROŠOVÁ, E. – PECÁKOVÁ, I.:** *Příklady k předmětu Statistika B.* Praha, Oeconomica 2004. ISBN 80-245-0680-7
8. **STUHLÝ, J.:** *Ekonometrie.* Jindřichův Hradec, Fakulta Managementu VŠE 2000.
9. **HEBÁK, P.:** *Vícerozměrné statistické metody.* Praha, Informatorium 2004. ISBN 80-7333-025-3
10. **WILLIAMS, D.:** *Weighing the odds.* Cambridge, Cambridge University Press 2001. ISBN 0-521-00618-X

Internetové zdroje

11. MIKROCENZUS 2002, dostupný z: www.czso.cz
12. SILC 2005, dostupný z: www.czso.cz
13. **GOLA, P.:** Důchody v Evropské unii – můžeme se těšit?, článek dostupný z: www.mesec.cz

Přílohy

Příloha 1: Podrobné členění sledovaných znaků Mikrocenzus

- **EA** Počet ekonomicky aktivních pracujících členů
- **KROK** Číslo kraje a okresu

11 - Hl. m. Praha	21 - Středočeský kraj	31 - Jihočeský kraj
32 - Plzeňský kraj	41 - Karlovarský kraj	42 - Ústecký kraj
51 - Liberecký kraj	52 - Královéhradecký kraj	53 - Pardubický kraj
61 - Vysočina	62 - Jihomoravský kraj	71 - Olomoucký kraj
72 - Zlínský kraj	81 - Moravskoslezský kraj	
- **DETI** Subpopulace členěny podle počtu nezaopatřených (závislých) dětí
- **POHL_P** Pohlaví osoby v čele

1 muž	2 žena
-------	--------
- **VZD_P** Vzdělání osoby v čele

0 bez vzdělání, nedokončené	1 základní
2 vyučení, nižší střední	3 úplné střední, vč. pomaturitního studia
4 vysokoškolské (vč. bakalářského)	
- **SKUP** Sociální skupina osoby v čele

1 dělník	2 samostatně činný - mimo zemědělství
3 zaměstnanec	4 samostatně činný - v zemědělství
5 zemědělec 6 důchodce v dom. s EA členy	7 důchodce v dom. bez EA členů
8 nezaměstnaný	0 ostatní

Příloha 2: Podrobné členění sledovaných znaků SILC

- **EA** Počet ekonomicky aktivních pracujících členů
- **KROK** Číslo kraje a okresu

11 - Hl. m. Praha	21 - Středočeský kraj	31 - Jihočeský kraj
32 - Plzeňský kraj	41 - Karlovarský kraj	42 - Ústecký kraj
51 - Liberecký kraj	52 - Královéhradecký kraj	53 - Pardubický kraj
61 - Vysočina	62 - Jihomoravský kraj	71 - Olomoucký kraj
72 - Zlínský kraj	81 - Moravskoslezský kraj	
- **DETI** Subpopulace členěny podle počtu nezaopatřených (závislých) dětí
- **POHL_P** Pohlaví osoby v čele

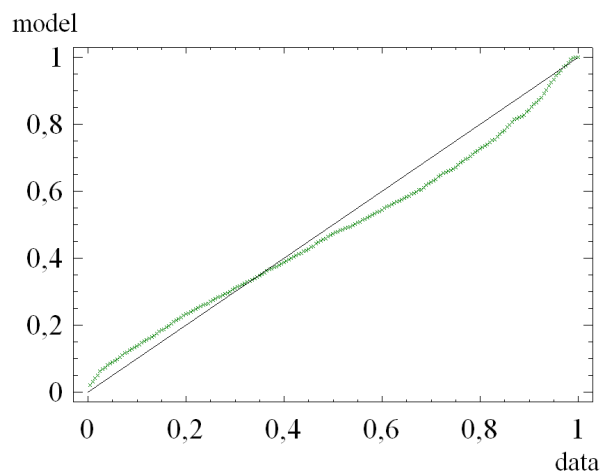
1 muž	2 žena
-------	--------
- **VZD_P** Vzdělání osoby v čele

0 bez vzdělání, nedokončené	1 základní
2 vyučení, nižší střední	3 úplné střední
4 nástavbové studium, pomaturitní kurzy, vyšší odborné	5 vysokoškolské
6 doktorské	
- **SKUP** Sociální skupina osoby v čele

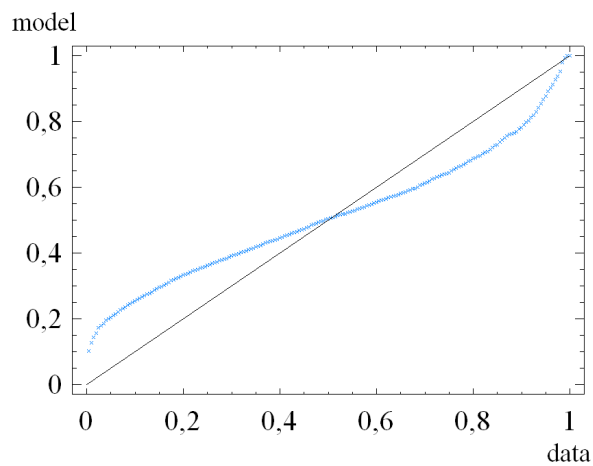
1 nižší zaměstnanec	2 samostatně činný
3 vyšší zaměstnanec	6 důchodce v dom. s EA členy
7 důchodce v dom. bez EA členů	8 nezaměstnaný
9 ostatní	

Příloha 3: Grafová příloha

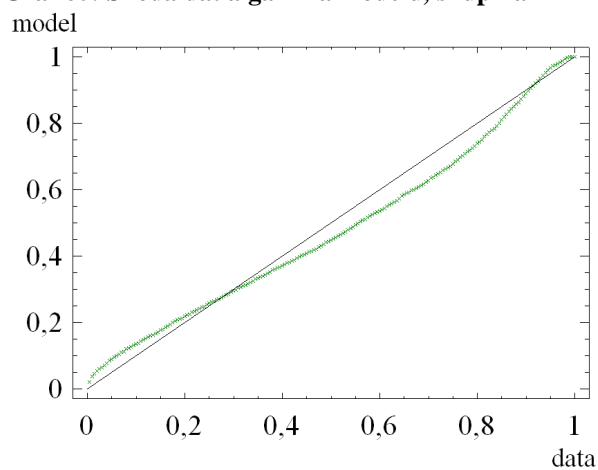
EA 2002



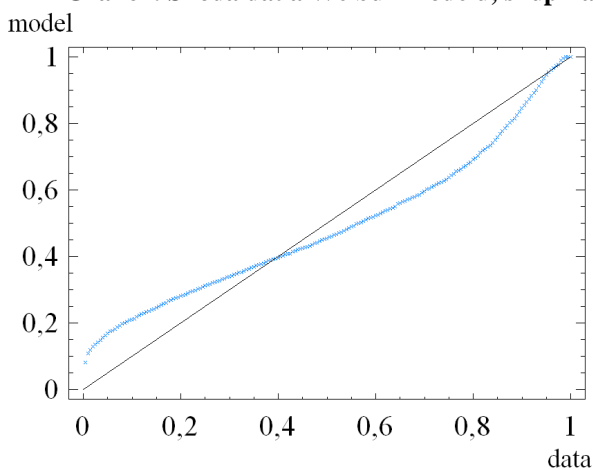
Graf 60: Shoda dat a gamma modelu, skupina 1



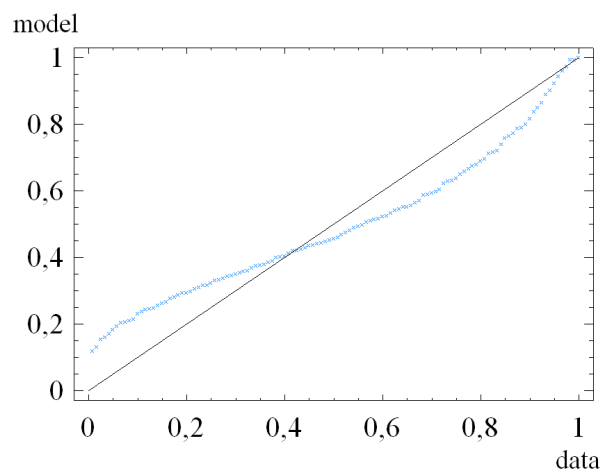
Graf 61: Shoda dat a Weibull modelu, skupina 1



Graf 62: Shoda dat a gamma modelu, skupina 2

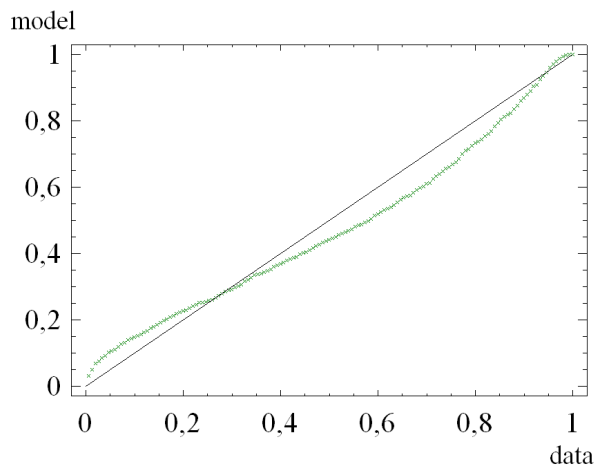


Graf 63: Shoda dat a Weibull modelu, skupina 2

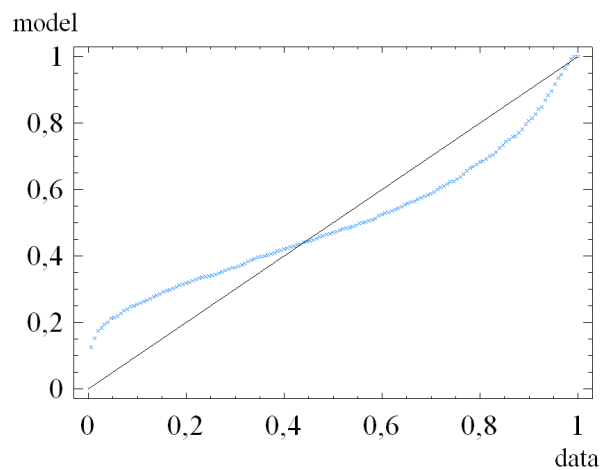


Graf 64: Shoda dat a Weibull modelu, skupina 3

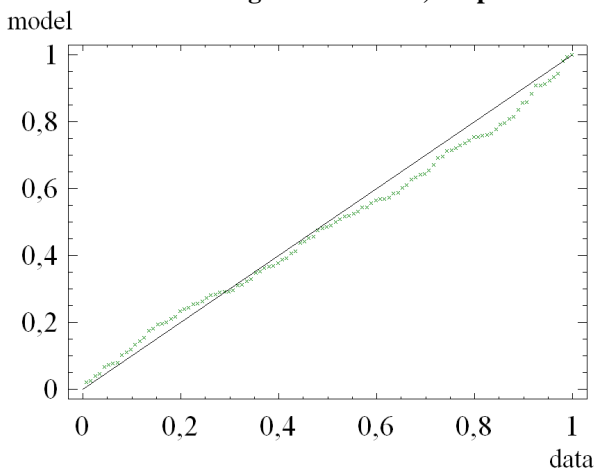
EA 2005



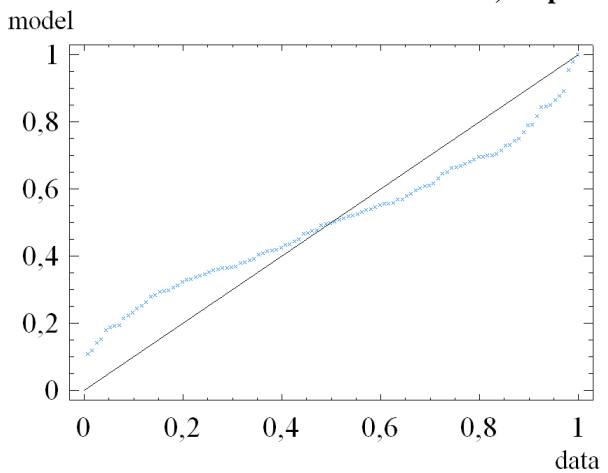
Graf 65: Shoda dat a gamma modelu, skupina 2



Graf 66: Shoda dat a Weibull modelu, skupina 2

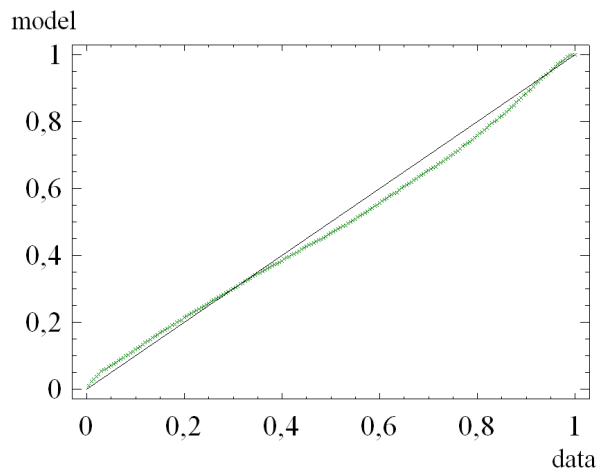


Graf 67: Shoda dat a gamma modelu, skupina 3

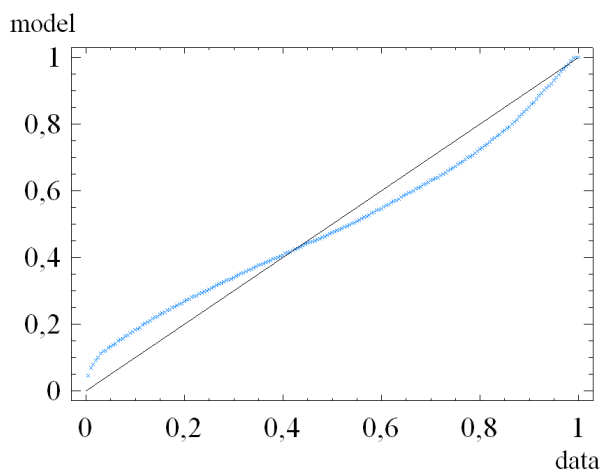


Graf 68: Shoda dat a Weibull modelu, skupina 3

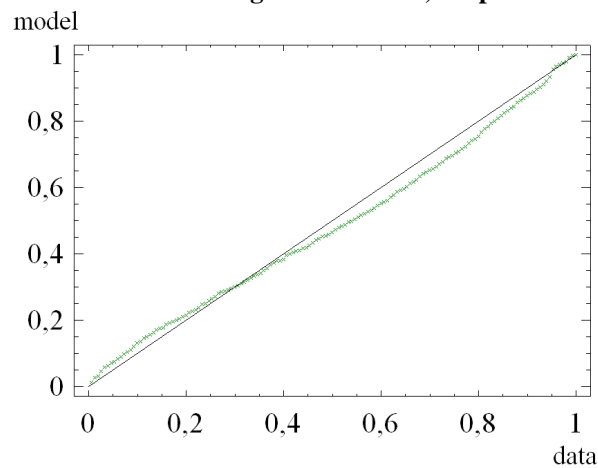
DĚTI 2002



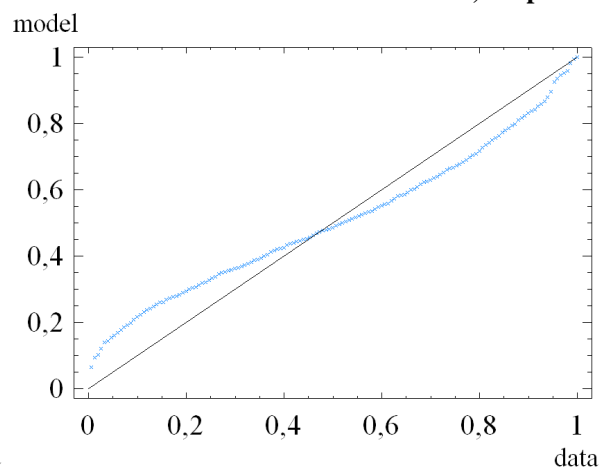
Graf 69: Shoda dat a gamma modelu, skupina 0



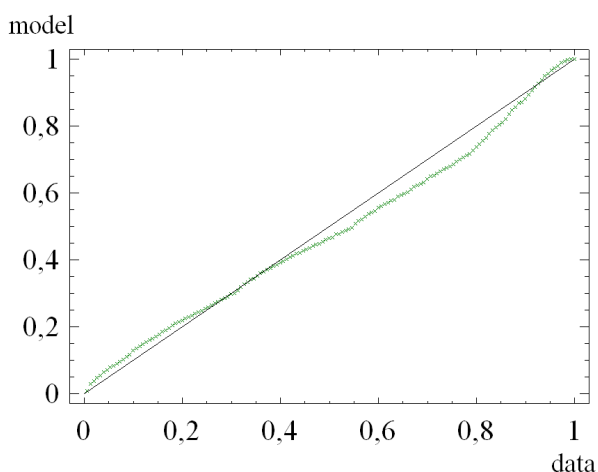
Graf 70: Shoda dat a Weib. modelu, skupina 0



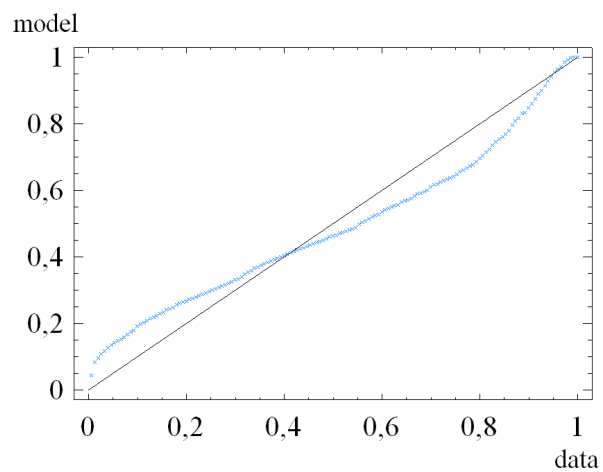
Graf 71: Shoda dat a gamma modelu, skupina 1



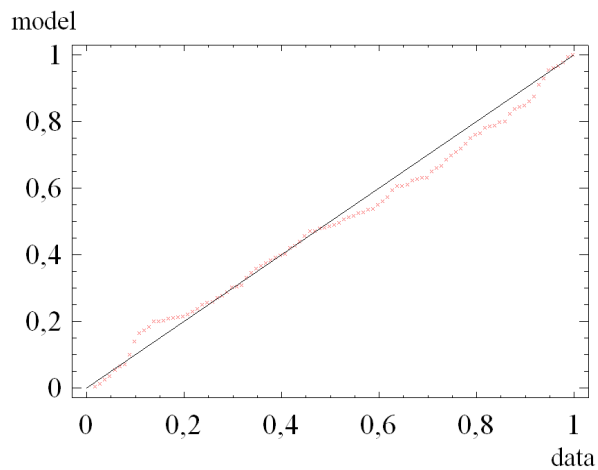
Graf 72: Shoda dat a Weib. modelu, skupina 1



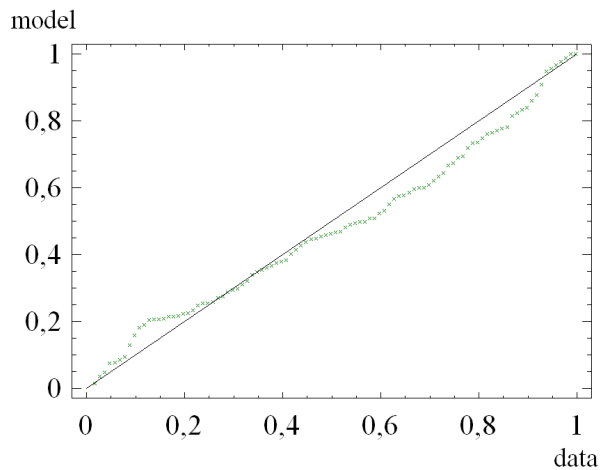
Graf 73: Shoda dat a gamma modelu, skupina 2



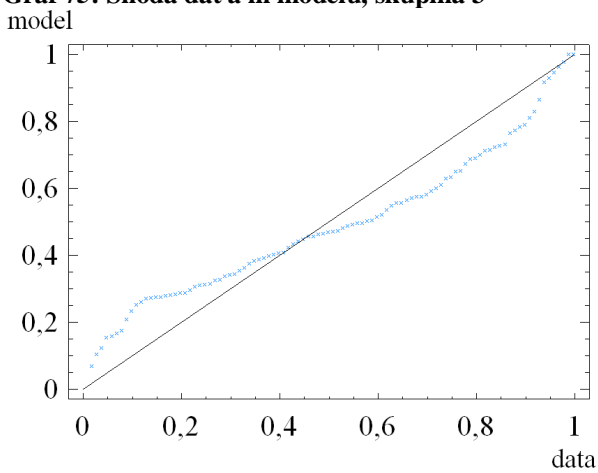
Graf 74: Shoda dat a Weib. modelu, skupina 2



Graf 75: Shoda dat a ln modelu, skupina 3

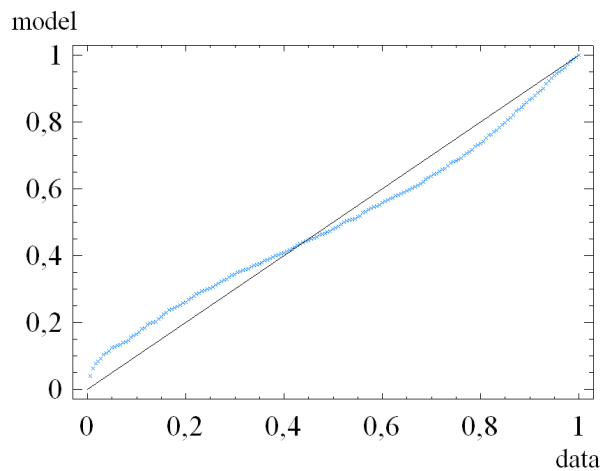


Graf 76: Shoda dat a gamma modelu, skupina 3

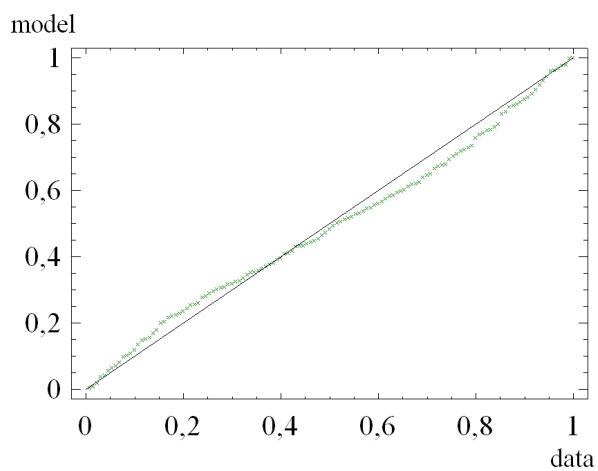


Graf 77: Shoda dat a Weib. modelu, skupina 3

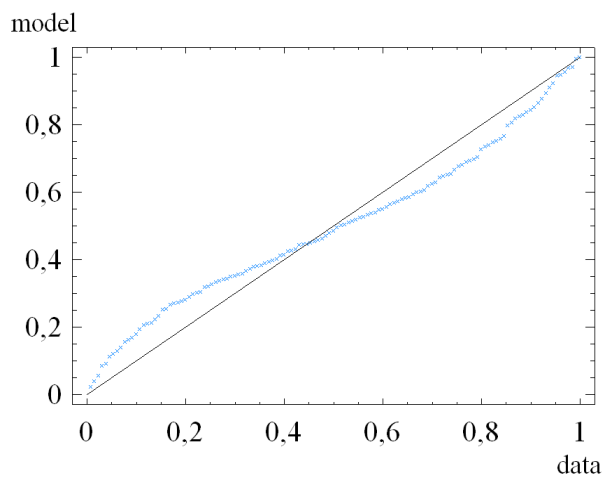
DĚTI 2005



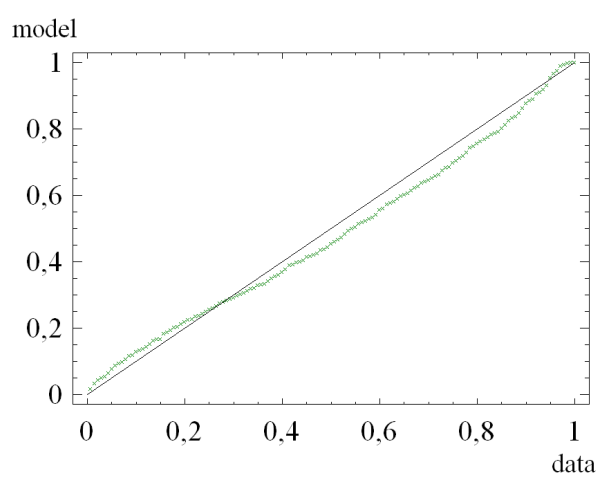
Graf 78: Shoda dat a Weib. modelu, skupina 0



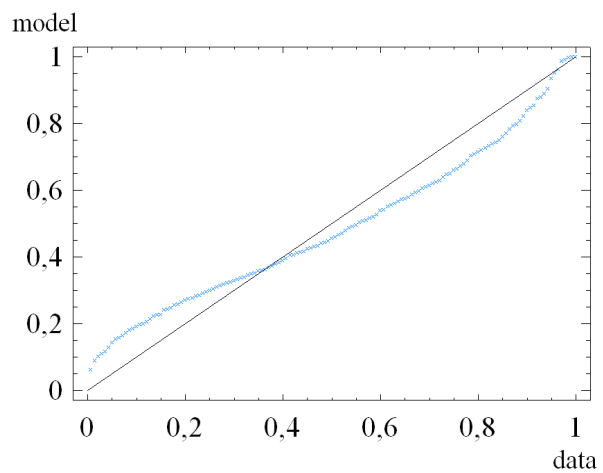
Graf 79: Shoda dat a gamma modelu, skupina 1



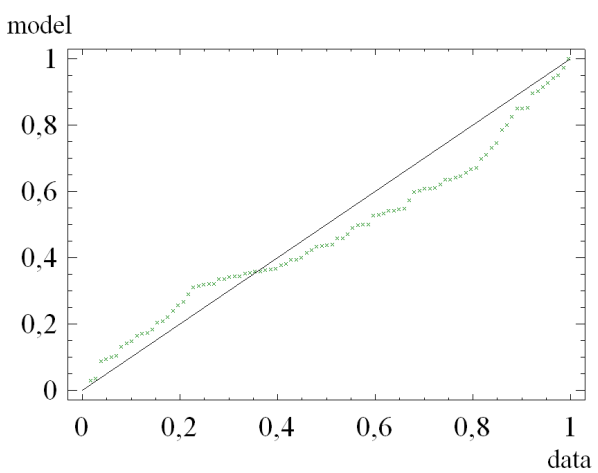
Graf 80: Shoda dat a Weib. modelu, skupina 1



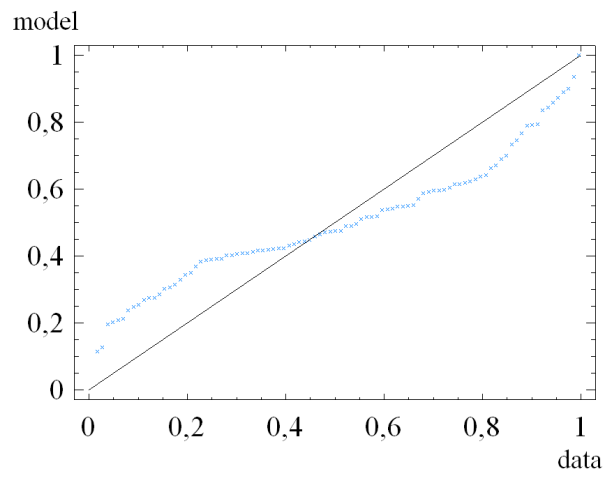
Graf 81: Shoda dat a gamma modelu, skupina 2



Graf 82: Shoda dat a Weib. modelu, skupina 2

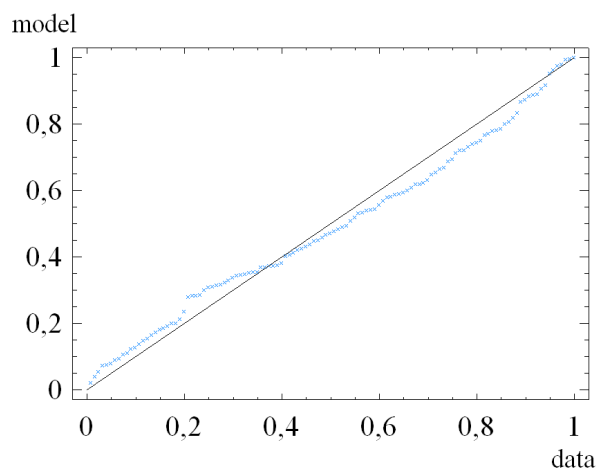


Graf 83: Shoda dat a gamma modelu, skupina 3

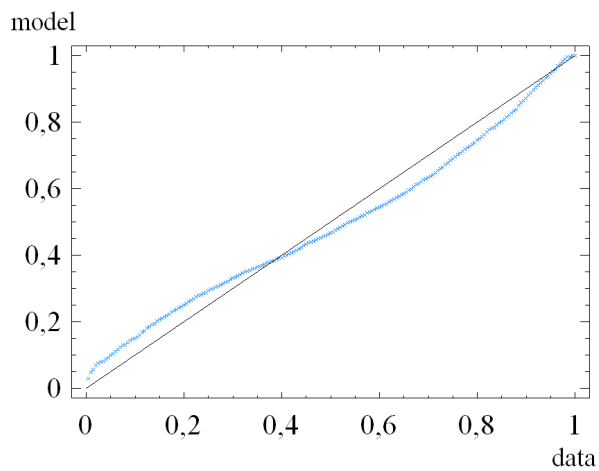


Graf 84: Shoda dat a Weib. modelu, skupina 3

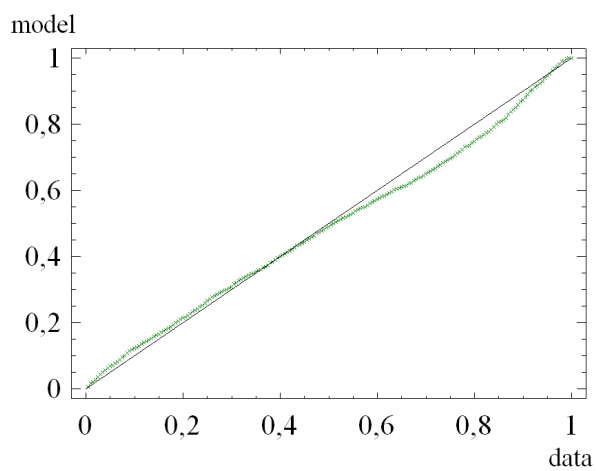
VZDĚLÁNÍ 2002



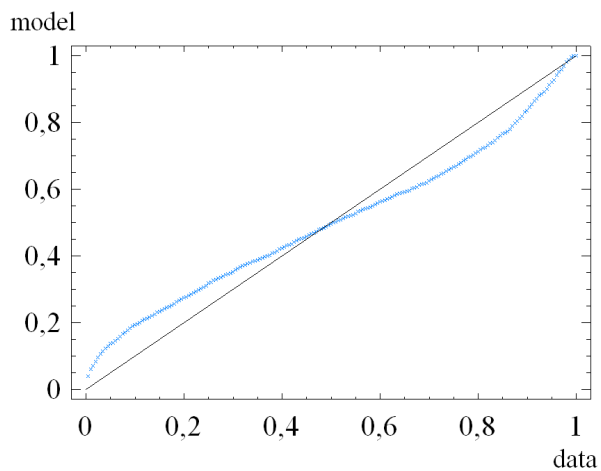
Graf 85: Shoda dat a Weib. modelu, skupina 1



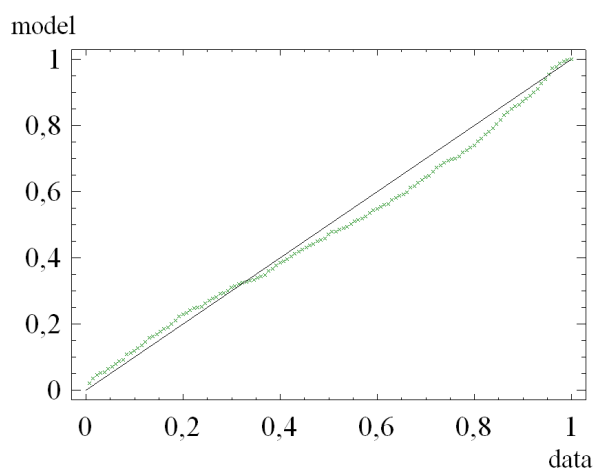
Graf 86: Shoda dat a Weib. modelu, skupina 2



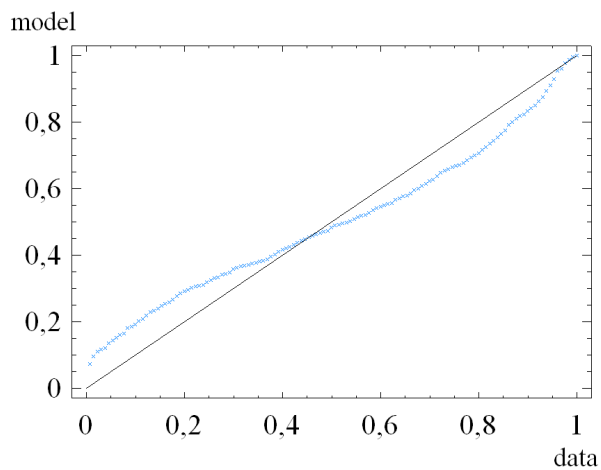
Graf 87: Shoda dat a gamma modelu, skupina 3



Graf 88: Shoda dat a Weib. modelu, skupina 3

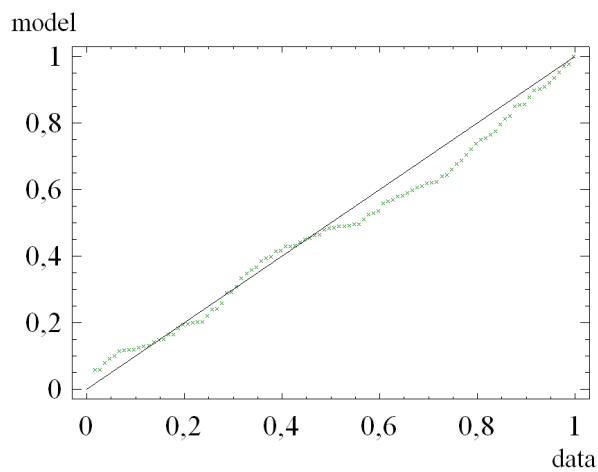


Graf 89: Shoda dat a gamma modelu, skupina 4

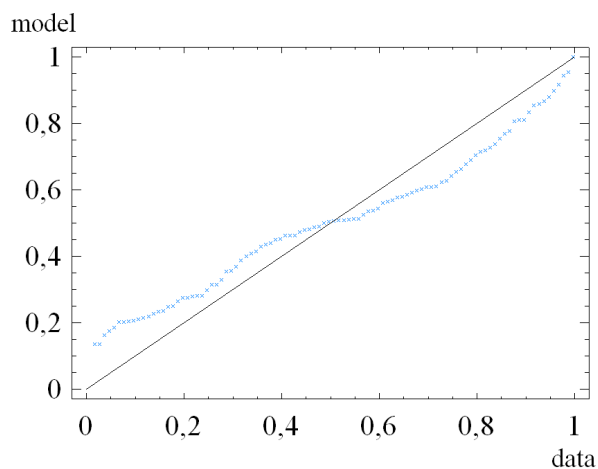


Graf 90: Shoda dat a Weib. modelu, skupina 4

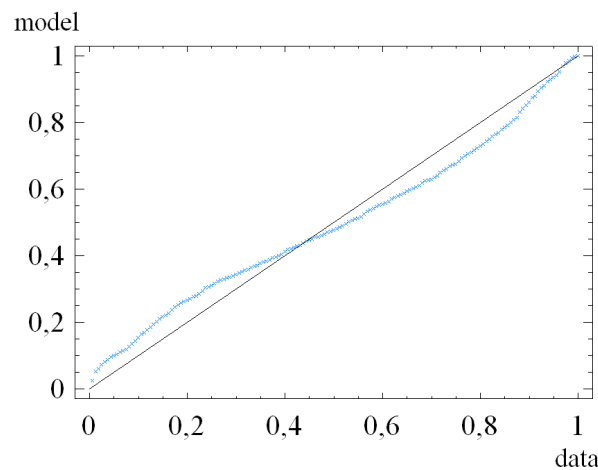
VZDĚLÁNÍ 2005



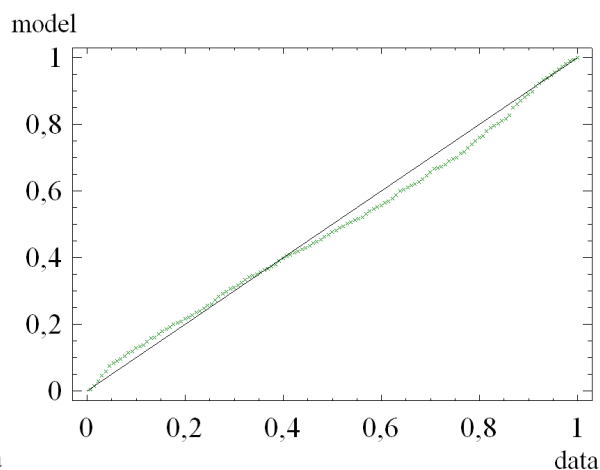
Graf 91: Shoda dat a gamma modelu, skupina 1



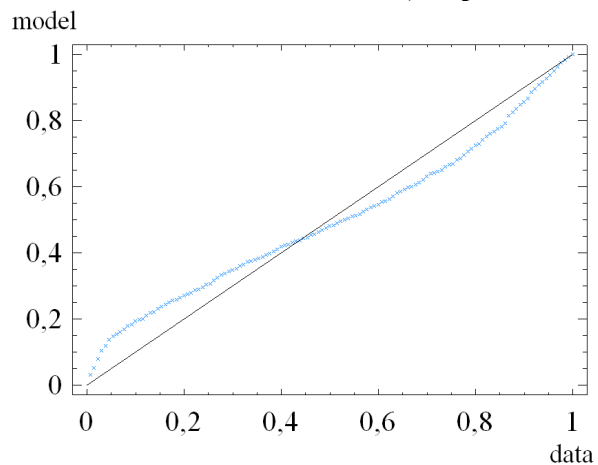
Graf 92: Shoda dat a Weib. modelu, skupina 1



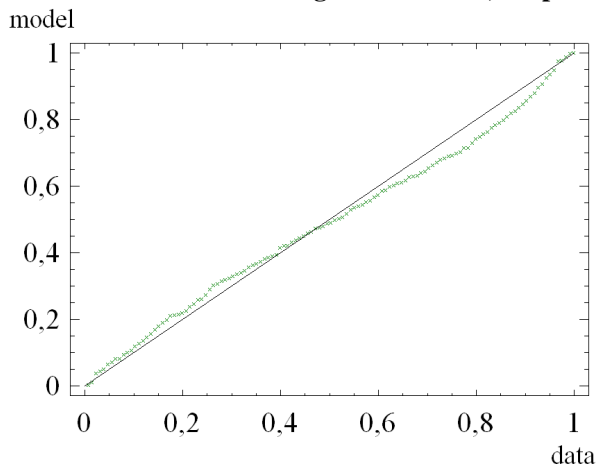
Graf 93: Shoda dat a Weib. modelu, skupina 2



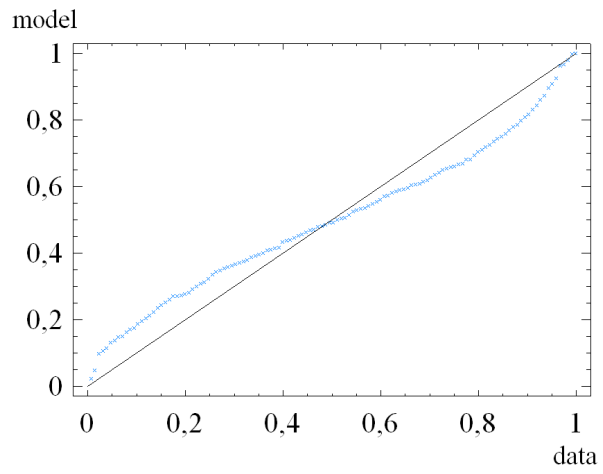
Graf 94: Shoda dat a gamma modelu, skupina 3



Graf 95: Shoda dat a Weib. modelu, skupina 3

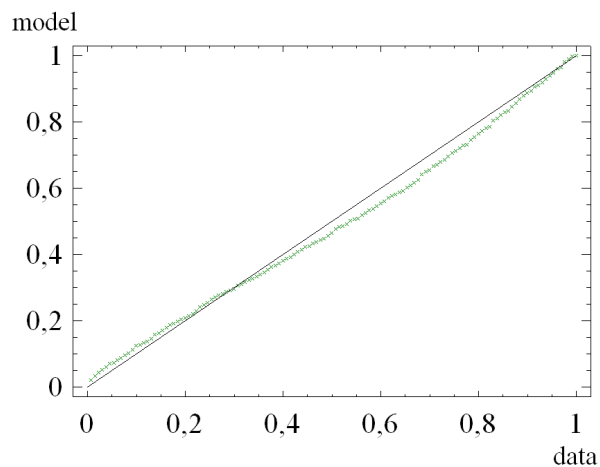


Graf 96: Shoda dat a gamma modelu, skupina 4

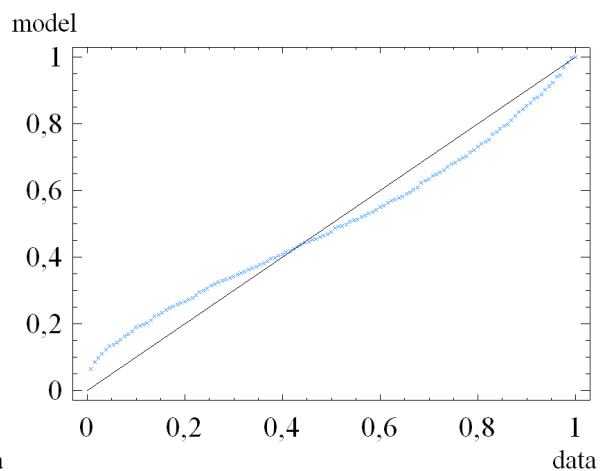


Graf 97: Shoda dat a Weib. modelu, skupina 4

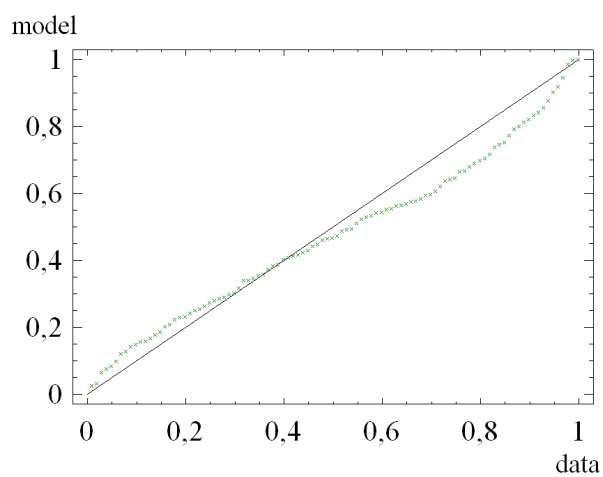
KRAJE 2002



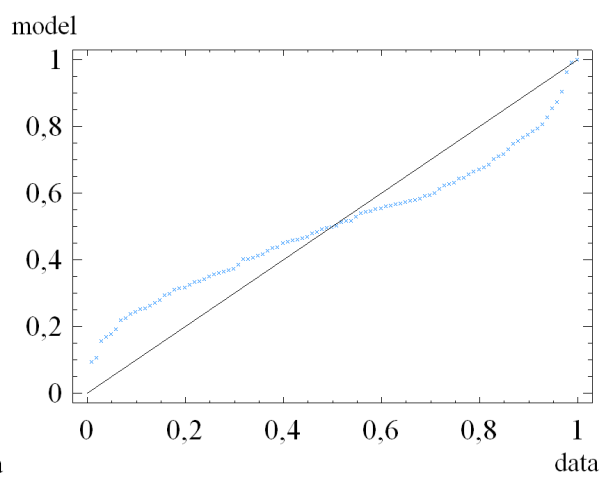
Graf 98: Shoda dat a gamma modelu, Praha



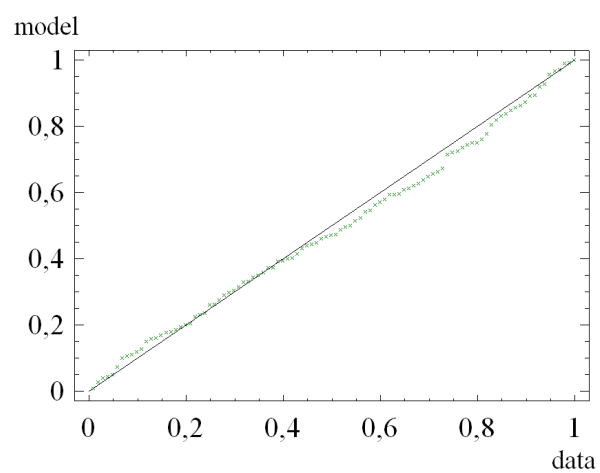
Graf 99: Shoda dat a Weibull modelu, Praha



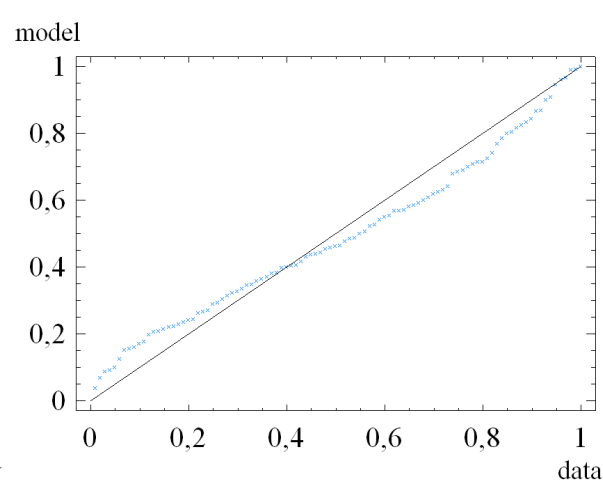
Graf 100: Shoda dat a gamma modelu, K. Vary



Graf 101: Shoda dat a Weibull modelu, K. Vary

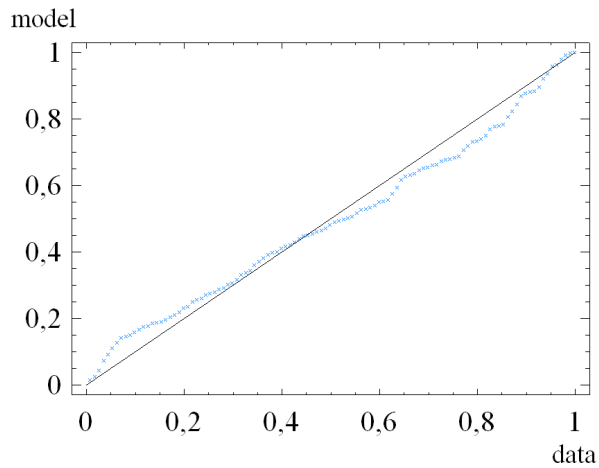


Graf 102: Shoda dat a gamma modelu, Olomouc

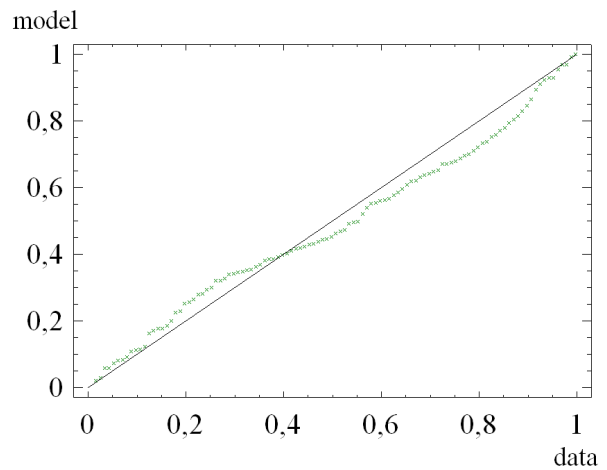


Graf 103: Shoda dat a Weibull modelu, Olomouc

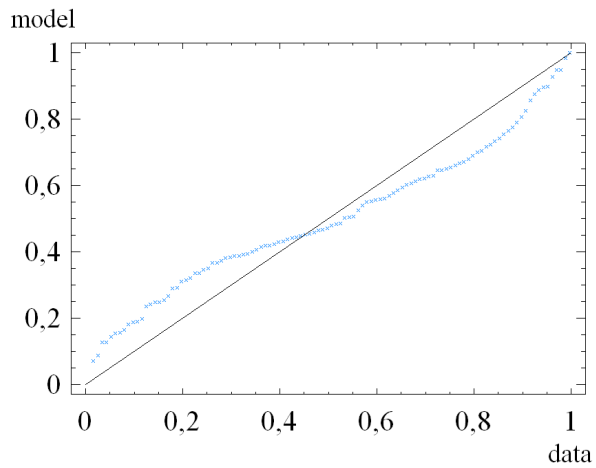
KRAJE 2005



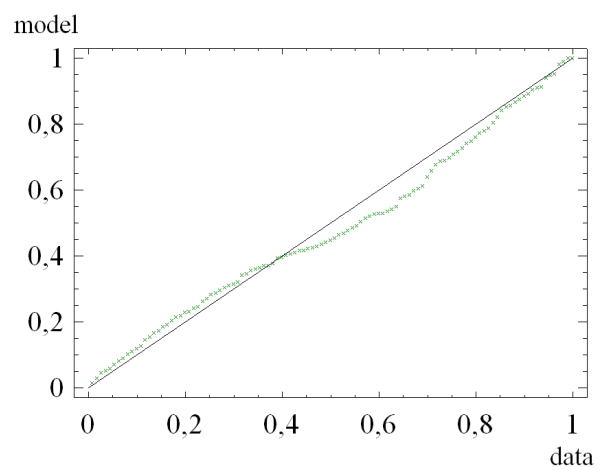
Graf 104: Shoda dat a Weibull modelu, Praha



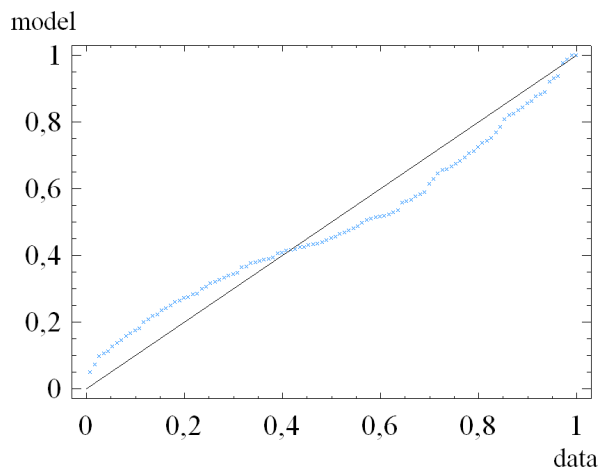
Graf 105: Shoda dat a gamma modelu, Olomouc



Graf 106: Shoda dat a Weibull modelu, Olomouc

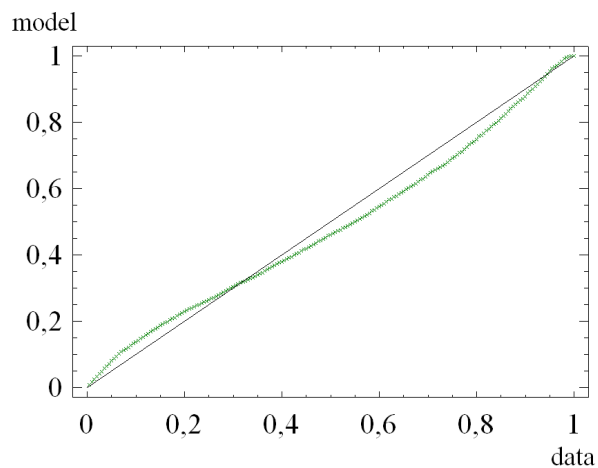


Graf 107: Shoda dat a gamma modelu, Moravs.

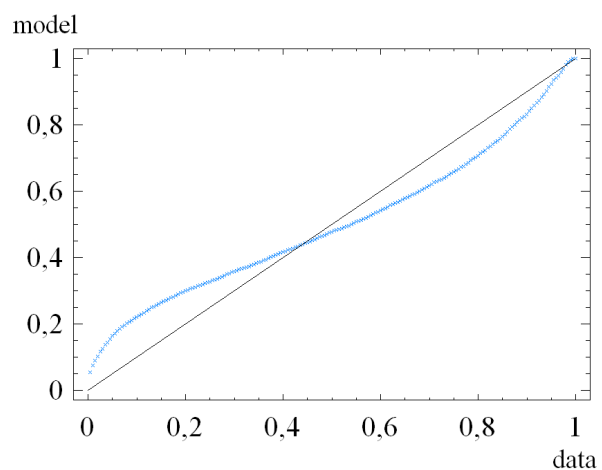


Graf 108: Shoda dat a Weibull modelu, Moravs.

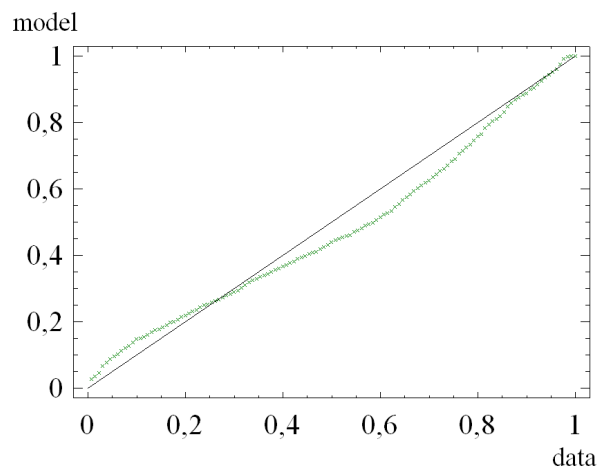
POHLAVÍ 2002



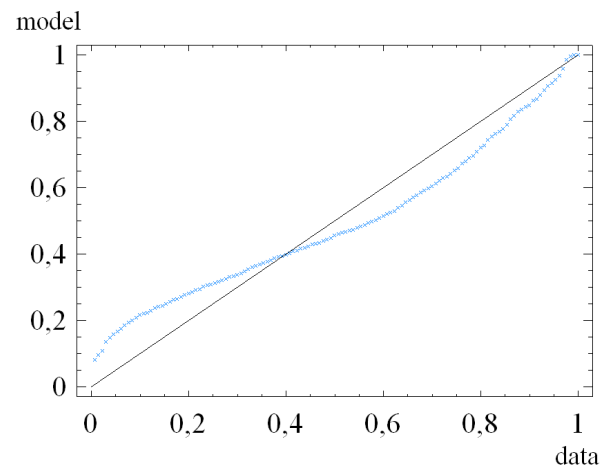
Graf 109: Shoda dat a gamma modelu, muž



Graf 110: Shoda dat a Weibull modelu, muž

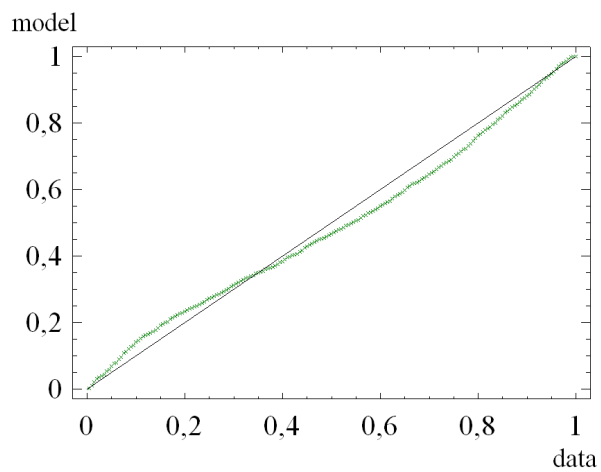


Graf 111: Shoda dat a gamma modelu, žena

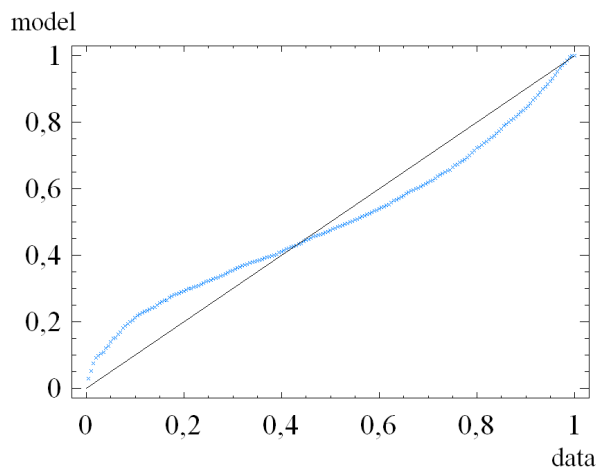


Graf 112: Shoda dat a Weibull modelu, žena

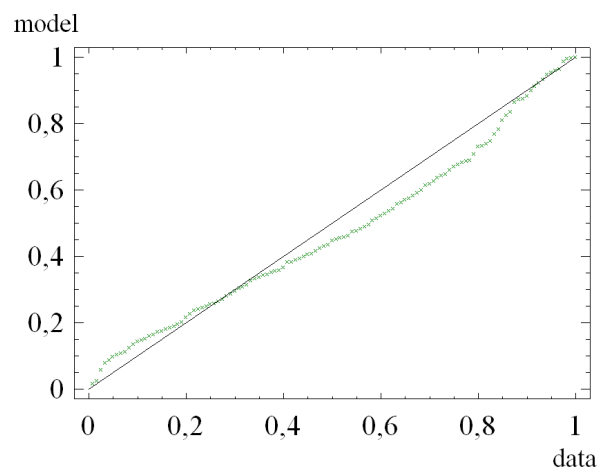
POHLAVÍ 2005



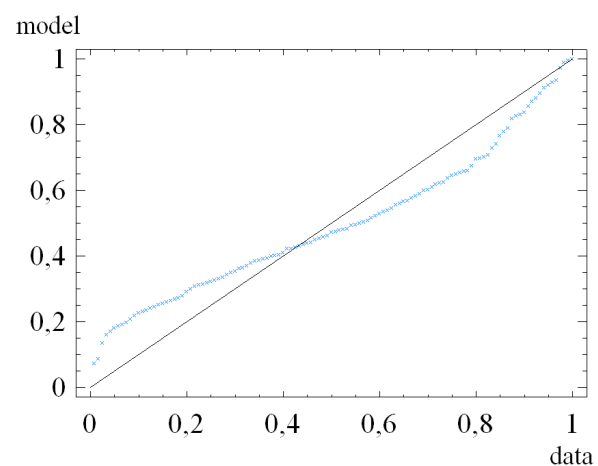
Graf 113: Shoda dat a gamma modelu, muž



Graf 114: Shoda dat a Weibull modelu, muž

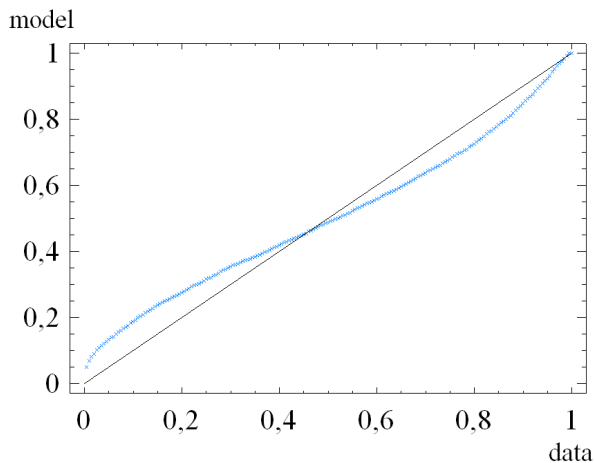


Graf 115: Shoda dat a gamma modelu, žena

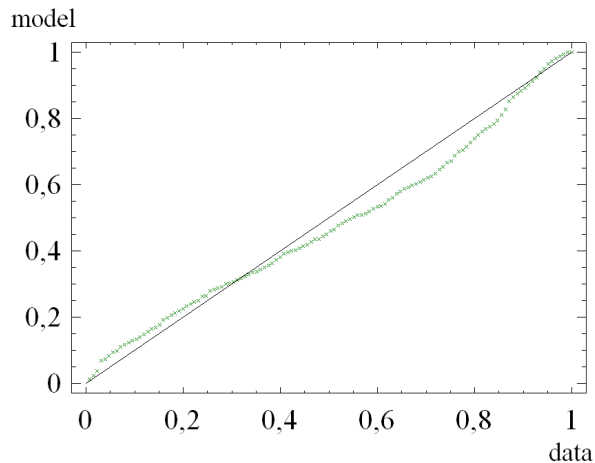


Graf 116: Shoda dat a Weibull modelu, žena

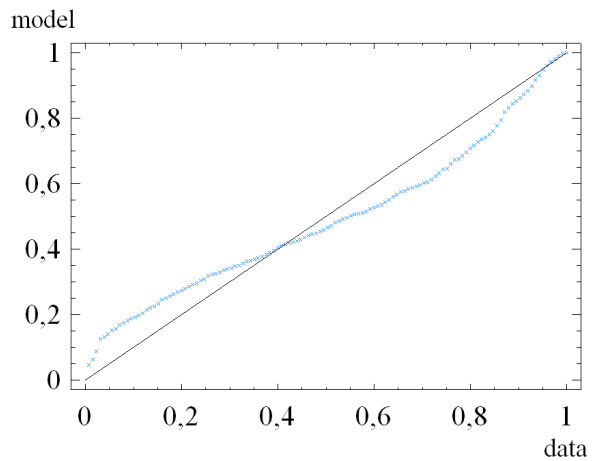
SKUPINY 2002



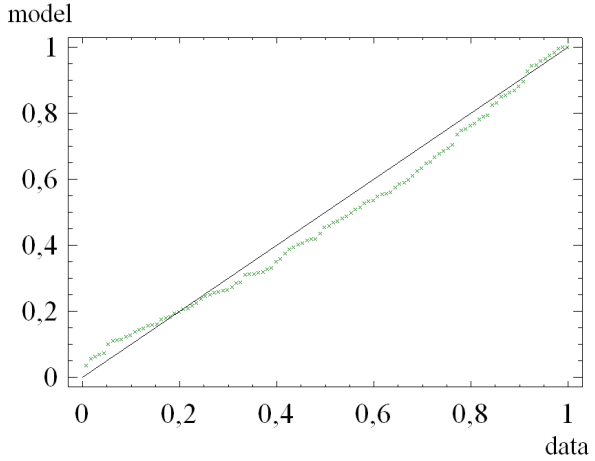
Graf 117: Shoda dat a Weibull modelu, skupina 1



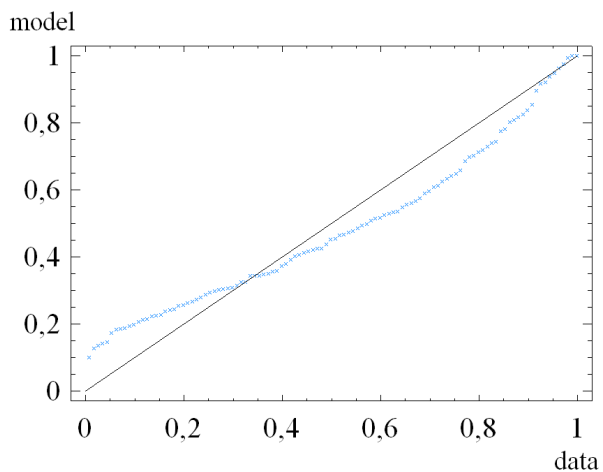
Graf 118: Shoda dat a gamma modelu, skupina 2



Graf 119: Shoda dat a Weibull modelu, skupina 2 skupina 4

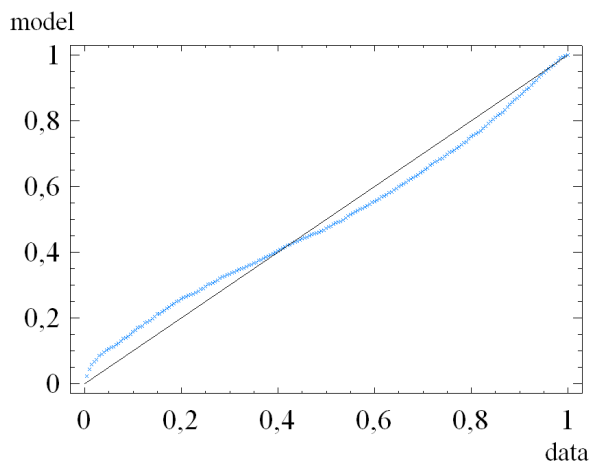


Graf 120: Shoda dat a gamma modelu,

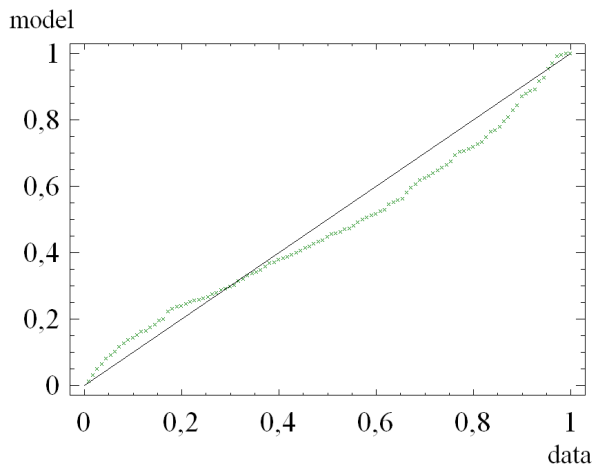


Graf 121: Shoda dat a Weib. modelu, skupina 4

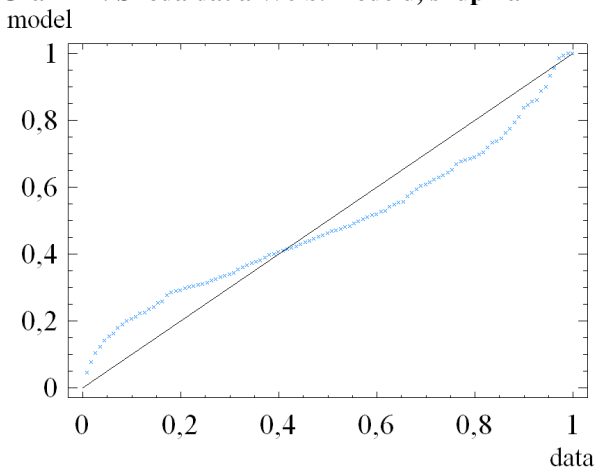
SKUPINY 2005



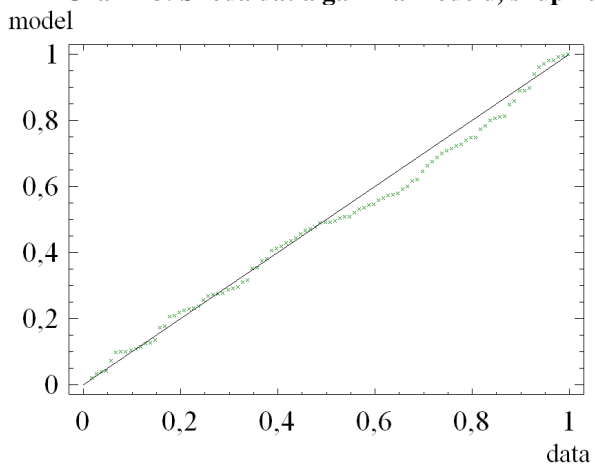
Graf 122: Shoda dat a Weib. modelu, skupina 1



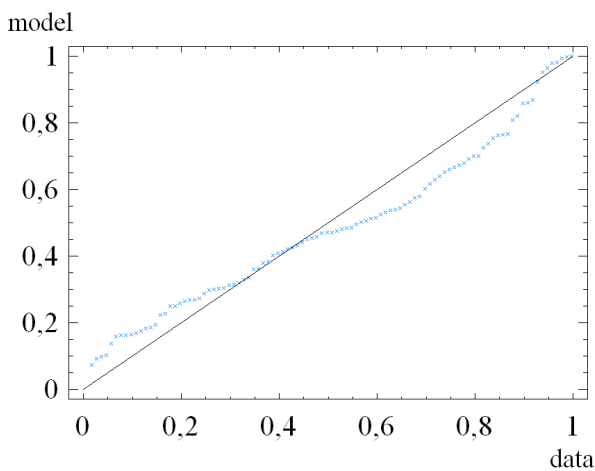
Graf 123: Shoda dat a gamma modelu, skupina 2



Graf 124: Shoda dat a Weib. modelu, skupina 2

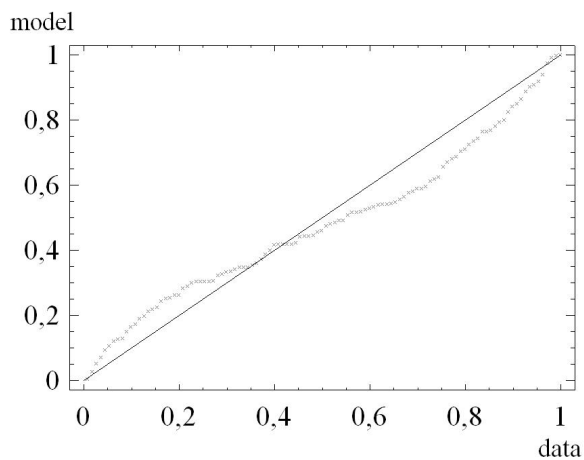


Graf 125: Shoda dat a gamma modelu, skupina 4

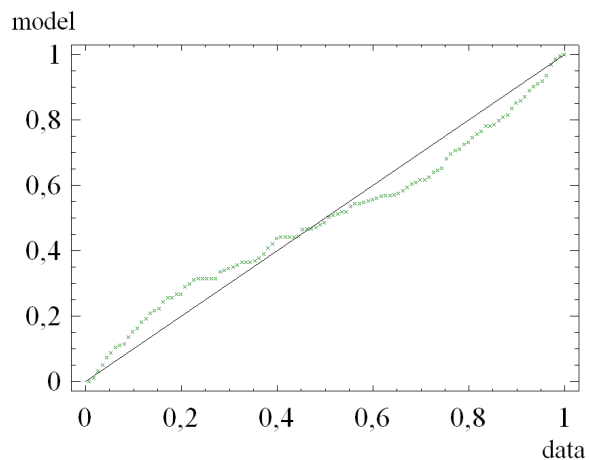


Graf 126: Shoda dat a Weib. modelu, skupina 4

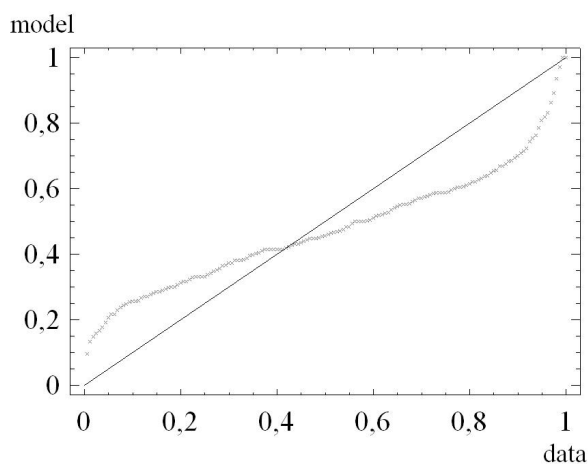
DŮCHODCI 2002



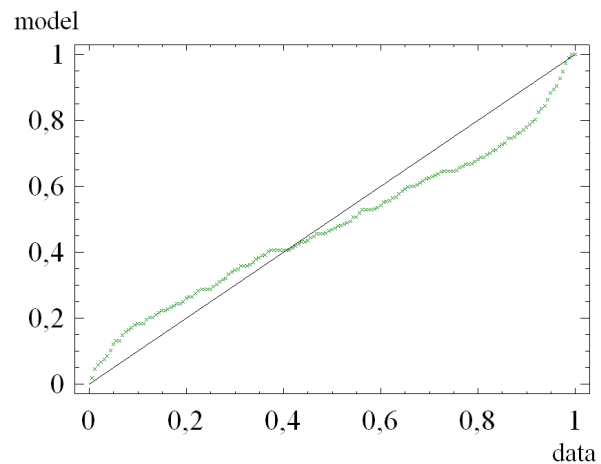
Graf 127: Shoda dat a normálního modelu, skupina 1



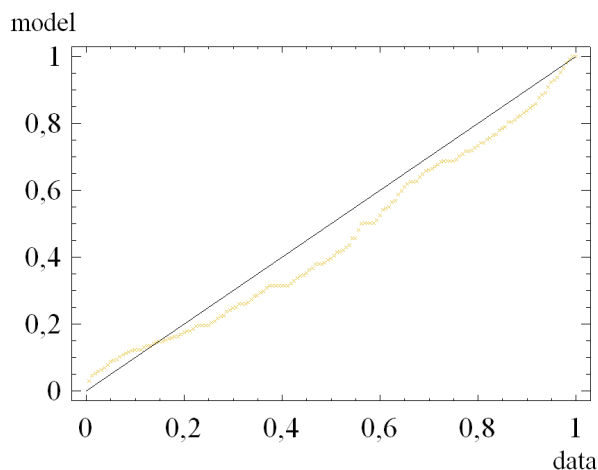
Graf 128: Shoda dat a gamma modelu, skupina 1



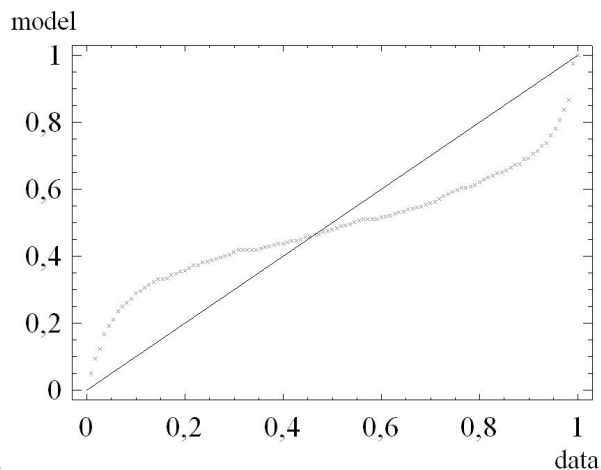
Graf 129: Shoda dat a normálního modelu, skupina 2



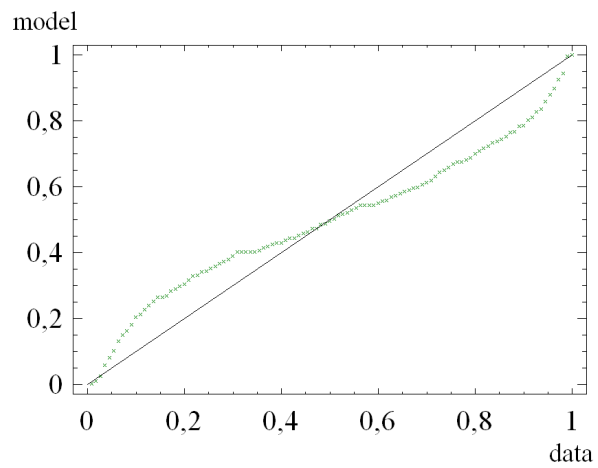
Graf 130: Shoda dat a gamma modelu, skupina 2



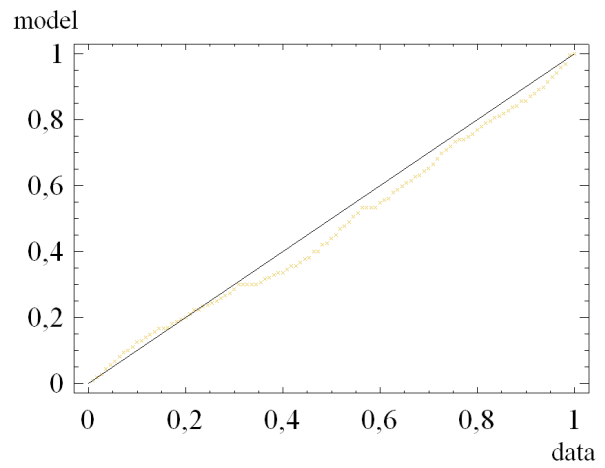
Graf 131: Shoda dat a Laplace modelu, skupina 2



Graf 132: Shoda dat a norm. modelu, skupina 3

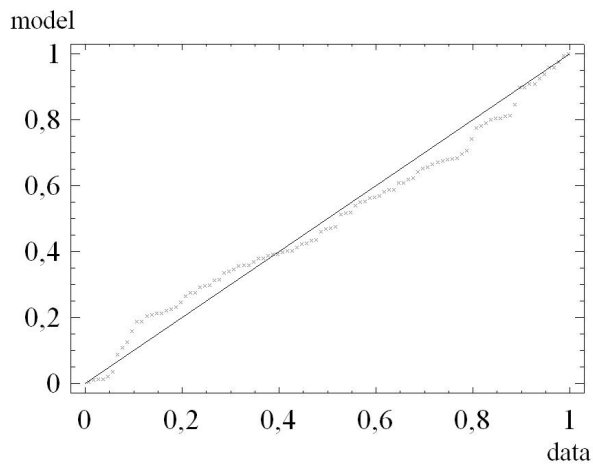


Graf 133: Shoda dat a gamma modelu, skupina 3

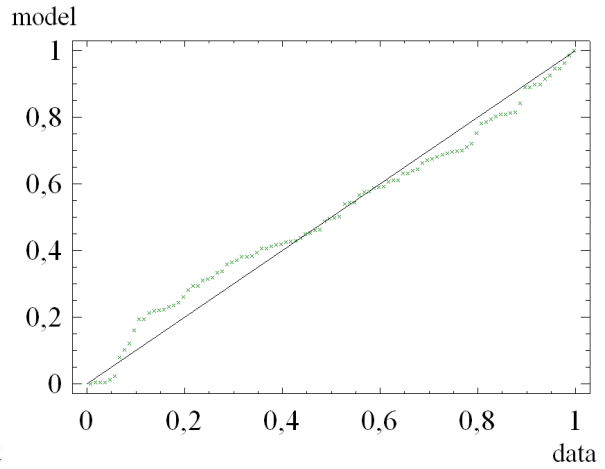


Graf 134: Shoda dat a Lapl. modelu, skupina 3

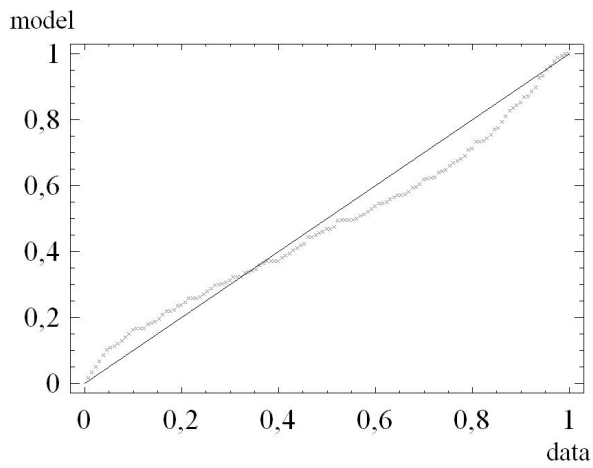
DŮCHODCI 2005



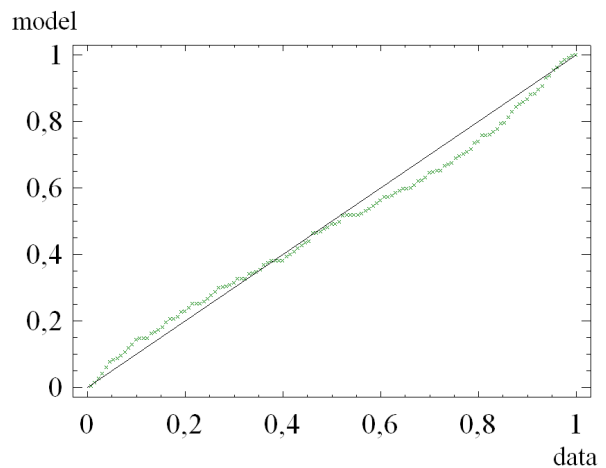
Graf 135: Shoda dat a norm. modelu, skupina 1



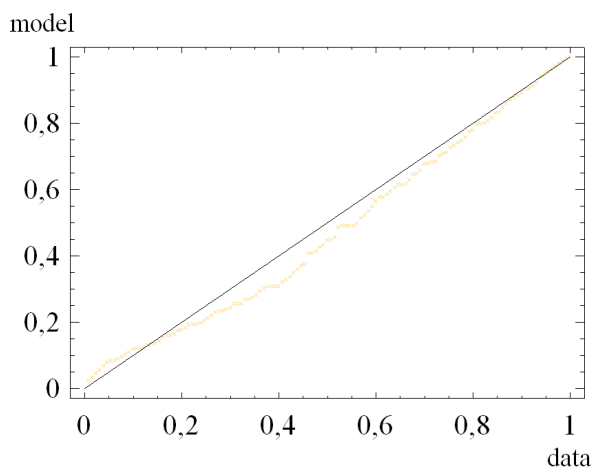
Graf 136: Shoda dat a gamma modelu, skupina 1



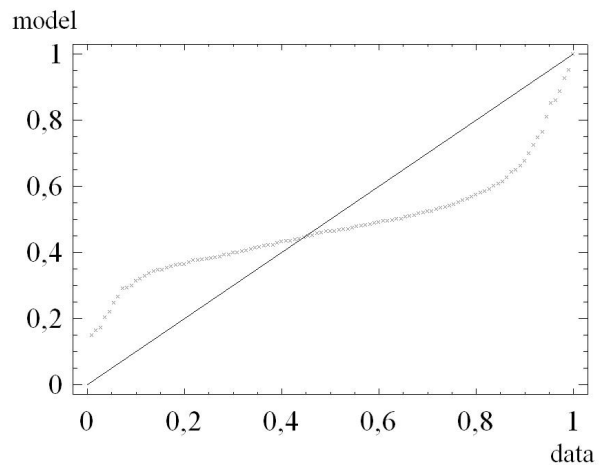
Graf 137: Shoda dat a norm. modelu, skupina 2



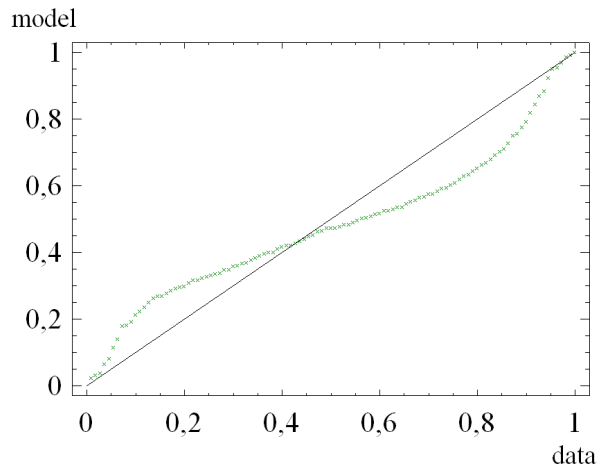
Graf 138: Shoda dat a gamma modelu, skupina 2



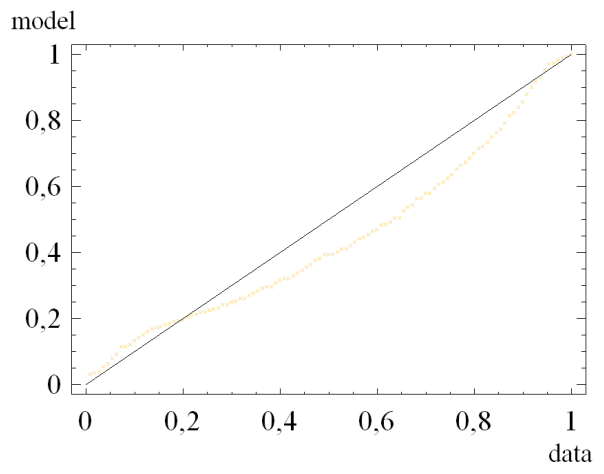
Graf 139: Shoda dat a Lapl. modelu, skupina 2



Graf 140: Shoda dat a norm. modelu, skupina 3



Graf 141: Shoda dat a gamma modelu, skupina 3



Graf 142: Shoda dat a Lapl. modelu, skupina 3