

Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra matematiky

Diplomová práce

Analýza příjmových nerovností

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracovala samostatně a použila pouze uvedené prameny a literaturu.

V Plzni dne

.....

Blanka Sudová

Poděkování

V první řadě bych chtěla touto cestou poděkovat za pomoc, cenné rady a trpělivost vedoucímu mé diplomové práce panu doc. RNDr. Petru Stehlíkovi, Ph.D. Dále bych chtěla vyjádřit díky rodině a blízkým, kteří mě po celou dobu studia podporovali.

Abstrakt

Tato práce je zaměřena na problematiku příjmové nerovnosti ve společnosti. Teoretická a metodologická část popisuje charakteristiky, kterými lze měřit příjmové nerovnosti. Giniho index a Theilův index jsou představeny a analyzovány v první části práce. Jsou definovány důležitější nové indexy nerovnosti, které následně použijeme pro porovnání skupin obyvatelstva. Pak se snažíme objasnit, zda a jak věk, pohlaví, region a počet let ve firmě ovlivňují výši příjmů. Dělíme vzorek populace do skupin a měříme nerovnost uvnitř a mezi skupinami. Výstup této práce je založen na datovém souboru zaměstnanců v Argentině.

Klíčová slova: míry nerovnosti, Giniho index, Theilův index, příjmová mezera, procentuální poměr, logistická regrese

Abstract

This thesis focuses on questions of income inequality in the society. The theoretical and methodological section describes characteristics by which income inequalities can be measured. Gini index and Theil index are introduced and analysed in first part. More importantly new measures of inequality are defined. They are consequently used for comparison of population groups. We then try to clarify if and how age, gender, region and number of years in the firm influence the income. We divide the population sample to groups and measure the inequality within and among groups. Output of this thesis is based on a dataset of employees in Argentina.

Keywords: measures of inequality, Gini index, Theil index, pay gap, percentage ratio, logistic regression

Obsah

Úvod	1
1 Popis dat	2
1.1 Základní představa o datovém souboru	2
1.2 Úprava dat	3
2 Nerovnosti rozdělení příjmů	6
2.1 Ukazatele míry nerovnosti rozdělení příjmů v rámci skupiny	6
2.1.1 Lorenzova křivka	6
2.1.2 Giniho index	8
2.1.3 Theilův index	9
2.1.4 Vztah Giniho a Theilova indexu	10
2.2 Ukazatele míry nerovnosti rozdělení příjmů mezi skupinami	11
2.2.1 Příjmová mezera	11
2.2.2 Příjmová mezera pro x -tý percentil	12
2.2.3 Index příjmové mezery	13
2.2.4 Procentuální poměr	17
2.2.5 Index procentuálního poměru	18
2.2.6 Kolmogorovův-Smirnovův dvouvýběrový test	20
3 Zpracování dat	21
3.1 Obecná nerovnost v zemi	21
3.2 Regionální příjmové nerovnosti	22
3.3 Příjmové nerovnosti věkových skupin	25
3.4 Příjmové nerovnosti způsobené dobou výkonu zaměstnání	28
3.5 Příjmové nerovnosti mezi obory zaměstnání	31
3.6 Příjmové nerovnosti mezi pohlavími	33
3.6.1 Nerovnost pohlaví v rámci regionů	35
3.6.2 Jednoduchý model logistické regrese	42
3.6.3 Model logistické regrese se zahrnutím dalších proměnných	44
3.7 Přehled nerovností	48
Závěr	49
Použitá literatura	51

Úvod

Příjmová nerovnost, tedy rozdíl v obdržení příjmech, posuzuje a měří chudobu ve společnosti, ale také míru přerozdělení v rámci daňových či sociálních systémů. Otázkou nerovností se zabývají spíše sociologové, ale pokud se jedná o příjmy, mají své slovo také ekonomové a matematici, kteří se snaží nerovnost měřit. K určení míry nerovnosti byly v posledních desetiletích odvozeny ukazatelé, které vycházejí z poznatků ekonomie, matematiky nebo teorie informace. Toto téma se stává velice aktuálním a zajímavým. Vzniká spousta prací, které se příjmové nerovnosti věnují.

Prvotní rozdělení důchodů, realizované na základě tržního mechanismu, není zpravidla společností vnímáno jako spravedlivé. Tento fenomén je obecně vnímán spíše negativně. Mnozí se obávají, že přinese škodlivé vedlejší účinky. Existují například studie [1], které prokazují, že nerovnost příjmů v zemi ovlivňuje negativně pocit štěstí obyvatel. Čím větší se vyskytuje nerovnost v příjmech, tím jsou lidé méně šťastní. Jak lze předpokládat, nespokojení jsou hlavně ti s nižším příjmem. Jejich nespokojenost ale nevyplyvá z nízkého příjmu, ale z pocitu nespravedlnosti.

Cílem této práce je na základě statistických metod popsat nejen obecné trendy, ale také odhalit, ve kterých skupinách se nerovnosti objevují. Pro tento účel byl získán datový soubor, který obsahuje informace o příjmech obyvatel jednoho státu a další jejich charakteristiky jako je pohlaví, rok narození, datum nástupu do firmy, obor, ve kterém pracují. Jeho popisu se věnuje první kapitola, která nás seznámí se základními statistikami, data graficky znázorní a popíše, jaké úpravy byly provedeny.

Ve druhé kapitole je popsána teorie související s měřením příjmové nerovnosti. Nejprve jsou definovány vlastnosti indexů měřící nerovnost, dále jsou uvedeny známé míry nerovnosti Giniho a Theilův index. Abychom lépe popsali nerovnost příjmů mezi skupinami, nově zavádíme některé indexy, jejich popis je taktéž zařazen do této kapitoly.

Stávající i nově definované míry nerovnosti jsou aplikovány na získané příjmy, výsledky shrnuje třetí kapitola, která zkoumá nerovnosti v různých skupinách obyvatel. Pozornost je věnována hlavně příjmové nerovnosti v rámci pohlaví, která je ve společnosti často diskutovaná. Pro toto dělení je definována a následně použita i logistická regrese, která určuje pravděpodobnost, s jakou se jedná o ženu, pokud o zkoumané osobě známe podrobné informace. Kvantifikované nerovnosti jsou poté přehledně zaznamenány a porovnány.

1 Popis dat

Cílem práce je analyzovat nerovnosti v příjmech, prvním úkolem proto bylo získat reálná data, která by popisovala rozdělení příjmů ve společnosti. Peněžním příjmem se rozumí jakákoliv přijatá platba, jedná se o mzdy, platy, sociální podpory, úroky, dividendy atd. Výše příjmů je soukromá informace, není proto snadné získat data obsahující tyto údaje. Pro velkou část obyvatelstva je největší částí příjmu mzda. Po dlouhém vyjednávání souhlasila s poskytnutím dat firma, která se celosvětově zabývá analýzou mezd zaměstnanců. Příjmem se proto bude v dalším textu rozumět pouze mzda. Z důvodu citlivosti dat si firma nepřejde, aby její jméno bylo v textu práce uvedeno.

Získaná data jsou vybraní obyvatelé Argentiny, kteří jsou zaměstnaní na plný úvazek v některé z firem, která poslala data k analýze. Nebylo možné zpracovávat údaje o zaměstnancích v České republice ani v jiné evropské zemi, protože se společnost bála o ohrožení dobrých vztahů s klienty. Informace o mzdách jsou snadno zneužitelné a klientská data jsou chráněna, aby nemohla být dále šířena. Podařilo se domluvit kompromis, že data mohou pocházet z jihoamerické země. Vzhledem k tomu, že s firmou zpracovávající data mám uzavřenou dohodu o pracovní činnosti a znám proces sběru dat a pečlivost, s jakou přistupují k vyplnění údajů, byla jako nejzornější zástupce vybrána Argentina. U zaměstnanců není uvedeno jejich jméno ani jméno společnosti, ve které pracují, ale i přes všechna tato opatření nesmějí být data dále šířena, proto je nebylo možno uložit na přiložené CD, které tak obsahuje jen zdrojové kódy.

Výběr obsahuje celkem 78 154 záznamů. Roční hrubá mzda je pro každý záznam uvedená v národní měně, argentinském pesu (ARS). Údaje jsou platné pro březen 2015. Kromě mzdy byly dostupné i následující informace: obor, pracovní pozice, pohlaví, region, rok narození, datum nástupu.

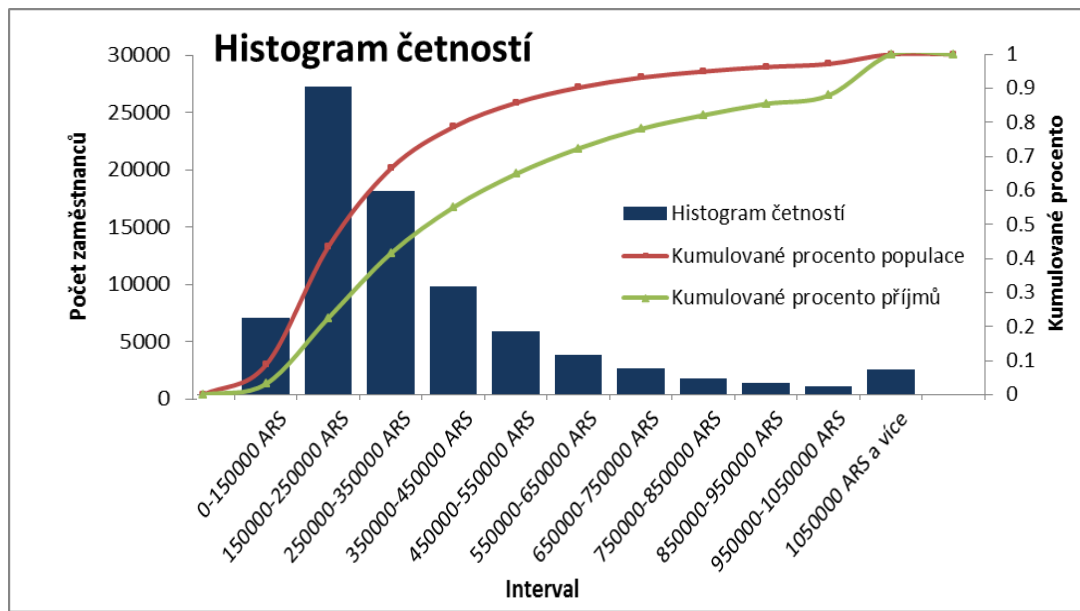
1.1 Základní představa o datovém souboru

Prvním krokem pro zřehlednění datového souboru je výpočet statistických charakteristik polohy a variability, Tabulka 1.1 udává přehled hodnot základních ukazatelů.

Minimum	42 029 (ARS)
Maximum	10 517 430 (ARS)
Průměr	356 574 (ARS)
Výběrová směrodatná odchylka	285 147 (ARS)
1. kvartil	194 442 (ARS)
medián	273 241 (ARS)
3. kvartil	413 400 (ARS)
Počet	78 154

Tabulka 1.1: Základní charakteristiky

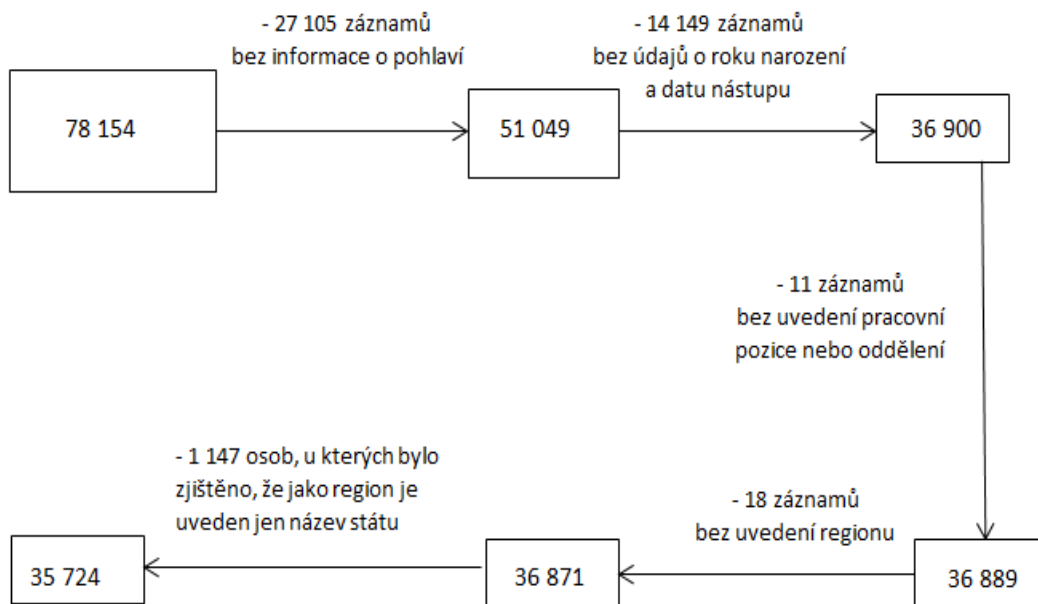
Rychlou vizuální informaci o výběrovém rozdělení četností nám podává histogram, jeho tvar může velmi dobře posloužit jako první informace pro odhad rozdělení pravděpodobností. Histogram četnosti rozdělení příjmů ve zpracovávaném náhodném výběru je uveden na Obrázku 1.1. Kromě histogramu je v obrázku vykresleno kumulované procento příjmů a kumulované procento populace, aby bylo možné získat lepší představu, jak jsou v populaci rozděleny příjmy. Je očividné, že nejvyšší procento příjmů patří nejbohatší skupině, jejíž roční příjem je více než 1 050 000 ARS, přestože tato skupina není nejpočetnější.



Obrázek 1.1: Histogram četností příjmů

1.2 Úprava dat

Jelikož data nebyla úplná, musela být nejprve protříděna a upravena. Z původních 78 154 osob byly odstraněny ti jedinci, u kterých některá informace chyběla nebo neměla žádnou vypovídací hodnotu. Postup třídění je znázorněn pomocí diagramu na následujícím obrázku. Po vytřídění nekompletních záznamů zbylo ve výběru 35 724 zaměstnanců.



Obrázek 1.2: Diagram postupu třídění záznamů

Vzhledem k dostupným informacím lze data rozdělit podle několika kritérií do skupin. Nejprve bylo určeno, o jaké skupiny se jedná a dále četnosti v jednotlivých skupinách. Jedním z možných dělení zaměstnanců je na základě pohlaví. Výběr obsahuje příjmy 10 964 žen a 24 760 mužů. Zároveň byly spočítány základní charakteristiky obou skupin, ale i výběru po odstranění nekompletních záznamů. Výsledky jsou uvedeny v Tabulce 1.2.

	Výběr po třídění	Ženy	Muži
Minimum	58523 (ARS)	59827 (ARS)	58523 (ARS)
Maximum	8015726 (ARS)	3650647 (ARS)	8015726 (ARS)
Průměr	332607 (ARS)	257466 (ARS)	365880 (ARS)
Výběrová směrodatná odchylka	276204 (ARS)	171588 (ARS)	305659 (ARS)
1. kvartil	194936 (ARS)	158057 (ARS)	186550 (ARS)
medián	286577 (ARS)	206700 (ARS)	270400 (ARS)
3. kvartil	453936 (ARS)	289848 (ARS)	435510 (ARS)
Počet	35724	10964	24760

Tabulka 1.2: Základní charakteristiky

Argentina je rozdělena do 24 oblastí (23 provincií a 1 federální distrikt). Aby mohlo dojít k analýzám nerovností v rámci regionů, bylo nutné ověřit, zda se ve výběru vyskytují zástupci ze všech oblastí. U každého zaměstnance je uvedeno v kolonce region jméno provincie, ze které zaměstnanec pochází, název ve výběru doprovází i přesnější určení lokality např. město nebo název oblasti. Bylo potvrzeno, že každý region je zastoupen, četnosti zastoupení jednotlivých regionů uvádí Tabulka 1.3, do které byly přidány informace HDP a počet obyvatel, které by mohly být v analýzách příjmových nerovností využity.

Provincie	HDP (v mld. USD) [2]	Počet obyvatel [3]	Četnost	Ženy	Muži
Buenos Aires	118.000	16 289 599	5 551	1408	4 143
Buenos Aires, federální distrikt	161.000	3 044 076	15 265	6468	8 797
Catamarca	2.331	389 256	309	59	250
Chaco	2.120	1 117 953	115	28	87
Chubut	7.110	545 656	938	188	750
Córdoba	33.239	3 489 669	1 008	338	670
Corrientes	4.053	1 049 325	156	42	114
Entre Ríos	7.137	1 295 121	400	129	271
Formosa	1.555	568 331	21	5	16
Jujuy	2.553	710 121	480	77	403
La Pampa	2.000	336 706	127	24	103
La Rioja	1.419	357 516	37	10	27
Mendoza	18.800	1 841 813	1 474	294	1 180
Misiones	4.044	1 159 445	119	33	86
Neuquén	14.398	601 003	1 815	233	1 582
Río Negro	4.924	678 797	129	21	108
Salta	5.165	1 295 944	264	46	218
San Juan	3.927	721 830	427	76	351
San Luis	2.444	463 411	153	29	124
Santa Cruz	6.892	302 420	1 695	207	1488
Santa Fé	37.500	3 341 228	3 942	960	2 982
Santiago del Estero	2.598	908 268	58	14	44
Tierra del Fuego	2.606	143 987	188	23	165
Tucumán	5.807	1 551 460	1 053	252	801
Celkem	451.622	42 202 235	35 724	10 964	24 760

Tabulka 1.3: Četnosti zastoupení regionů ve výběru a informace o regionu

Zaměstnanci pocházejí z různých oborů, celkem se objevuje 61 různých profesí. Nejčetnější skupinou jsou zaměstnanci ve výrobě (6 513 zástupců), dalšími obory s vysokým počtem zástupců jsou například logistika (3 117 zástupců) nebo administrativa (2 940 zástupců). Dále jsou mimo jiné dostupné informace z dalších profesí jako jsou právníci, lékaři, bankéři, manažeři.

2 Nerovnosti rozdělení příjmů

O nerovnosti (respektive rovnosti) můžeme mluvit tehdy, pokud porovnáváme jedince nebo skupiny jedinců mezi sebou. Příjmová nerovnost určuje, jak se reálné příjmy jednotlivců odlišují od situace, kdy by všichni měli stejné příjmy. Pokud je ve dvou případech rozhodnuto, že jde o nerovnost, není lehké určit, která z nerovností je větší. Nerovnost musí být měřitelná, aby mohla být porovnávána.

Indexy určené k měření příjmové nerovnosti by měly splňovat několik požadovaných vlastností. [4]

- V1) V případě, kdy mají všichni stejný příjem, měl by být výsledek indexu 0.
- V2) Hodnoty indexů by měly být bezrozměrné, mohou tak být porovnávány nerovnosti různých veličin.
- V3) Míra nerovnosti by měla být neměnná, pokud jsou příjmy všech vynásobeny stejnou konstantou. Tato podmínka zaručí, že se nerovnost nezmění, když jsou příjmy převedeny do jiné měny. Méně očividné je, že se míra nerovnosti zachová, pokud všem příjmem vzroste o stejné procento. Podíl příjmu jednotlivce na celkovém příjmu zůstává stejný.
- V4) Indexy určující míru nerovnosti závisí více na relativních než na absolutních rozdílech.

K měření příjmové nerovnosti slouží například indexy, které budou popsány v této kapitole. Základní míry variability jako je rozptyl a směrodatná odchylka nejsou k určení míry nerovnosti vhodné, protože nesplňují uvedené požadavky na ukazatele míry nerovnosti.

2.1 Ukazatele míry nerovnosti rozdělení příjmů v rámci skupiny

Jedním ze způsobů, jak měřit nerovnost rozdělení příjmů, je porovnávání jedinců v rámci jedné skupiny, která je určena na základě nějaké společné vlastnosti jejich zástupců (např. stejné pohlaví, stejná věková skupina nebo stejný region).

Pro ukazatele míry nerovnosti v rámci skupiny platí kromě obecně požadovaných vlastností V1 až další vlastnost. Pokud je všem k příjmu přičtena stejná kladná konstanta, hodnoty indexů se změní, sníží se. Pokud například příjem jednoho jedince je 1 000 Kč a druhého 2 000 Kč, je nerovnost mezi příjmy patrná, jeden má dvojnásobný příjem oproti druhému. Pokud je oběma navýšen příjem o 1 000 000 Kč, rozdíl mezi příjmy je vzhledem k jejich výši téměř zanedbatelný. [4]

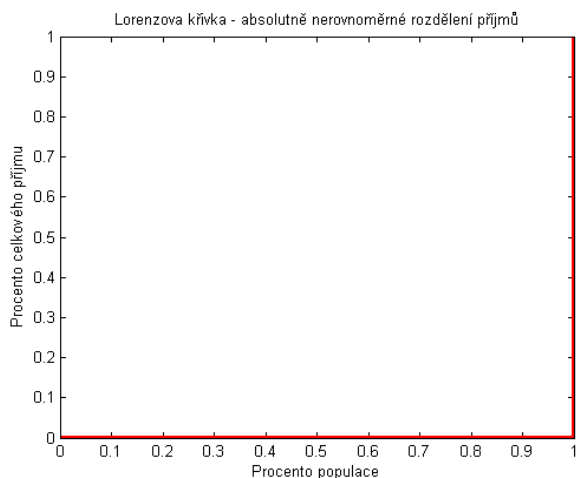
2.1.1 Lorenzova křivka

Lorenzova křivka (= *LC* z anglického názvu Lorenz Curve) slouží k vykreslení nerovnoměrnosti rozdělení příjmů. Horizontální osa *x* zobrazuje kumulované procento obyvatel, které je seřazeno podle výše příjmů nesestupně.

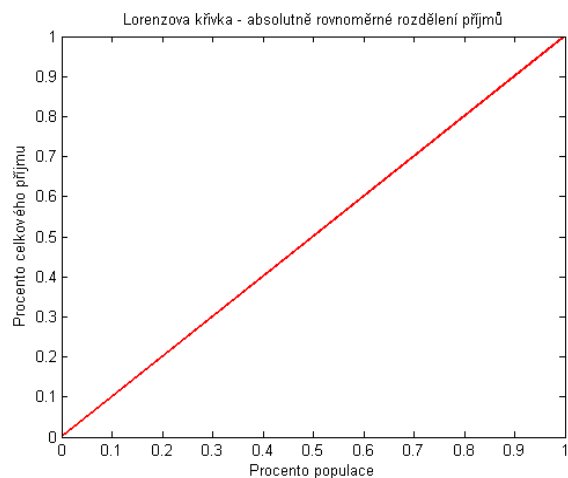
Vertikální osa *y* určuje podíl příjmů na celkové sumě. Lorenzova křivka říká, kolik procent příjmů přísluší určitému procentu jedinců. Dosahuje-li tedy kupříkladu funkce v bodě 40 % hodnoty 20 %, znamená to, že 40 % nejchudších jedinců disponuje celkem 20 % z celkového příjmu všech obyvatel.

Mohou nastat dva extrémní případy:

- Absolutně nerovnoměrné rozdělení příjmů, kdy veškerý příjem patří jedné osobě, zatímco příjem ostatních je nulový. Tomuto rozdělení příjmů odpovídá Lorenzova křivka na Obrázku 2.1.
- Absolutně rovnoměrné rozdělení příjmů, které odpovídá rovnosti v příjmech, tedy všichni pobírají stejný příjem. Tvar Lorenzovy křivky pro tento případ je zobrazen na Obrázku 2.2.

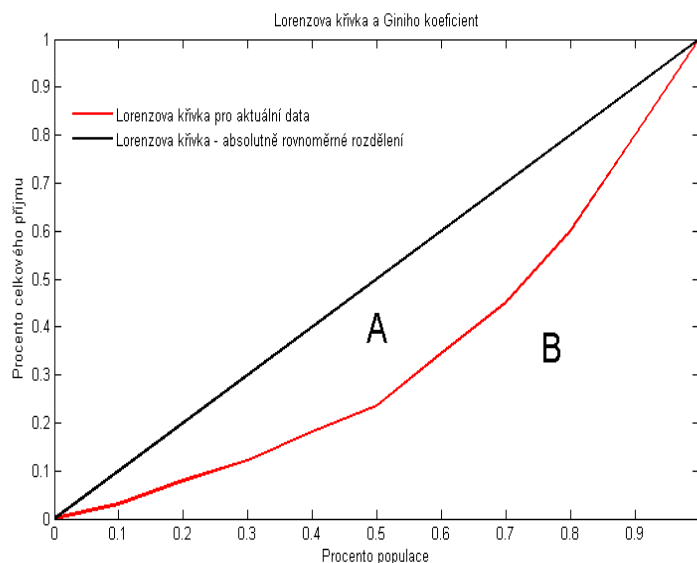


Obrázek 2.1: LC – absolutně nerovnoměrné rozdělení



Obrázek 2.2: LC – absolutně rovnoměrné rozdělení

Obecně platí ve všech případech, že Lorenzova křivka prochází body $[0,0]$ a $[1,1]$ a pohybuje se mezi zmíněnými extrémními případy. Čím více se Lorenzova křivka prohýbá (čím více se blíží ke křivce absolutně nerovnoměrného rozdělení příjmů), tím větší je míra nerovnosti. Průběh Lorenzovy křivky v obecném případě je znázorněn na Obrázku 2.3. Pro názornost, je v obrázku také zakreslena Lorenzova křivka pro případ absolutně rovnoměrného rozdělení.



Obrázek 2.3: Lorenzova křivka

Pro teoretické účely je vhodné považovat příjmy jednotlivců y_i považovat za náhodný výběr z rozdělení s distribuční funkcí $F(y)$. Nechť $F(y)$ je funkce rostoucí a střední hodnota příjmů existuje a značí se μ . První podmínka zaručuje existenci inverzní funkce $F^{-1}(x)$ k distribuční funkci $F(y)$. Odpovídající Lorenzova křivka (podíl x nejchudších jedinců na celkovém příjmu) má pak tvar

$$LC(x) = \mu^{-1} \int_0^x F^{-1}(t) dt, \quad 0 \leq x \leq 1. \quad (2.1)$$

Pokud rozdělení příjmů ve společnosti odpovídá některému ze známých rozdělení pravděpodobnosti, je tedy možné Lorenzovu křivku matematicky popsat s využitím předpisu pro distribuční funkci tohoto rozdělení. [5]

2.1.2 Giniho index

Giniho index (GI) je standartní ekonomická charakteristika příjmové nerovnosti založená na Lorenzově křivce. Vyčísluje, jak moc se Lorenzova křivka liší od absolutně rovnoměrného rozdělení příjmů. [5] Nabývá hodnot od 0 do 1. Společnosti, jejichž Giniho index se rovná 0, mají dokonalou rovnost v rozdělení příjmů (viz. Lorenzova křivka na Obrázku 2.2). Naopak hodnota 1 značí nerovnost, kdy veškerý příjem patří jedné osobě, Lorenzovu křivku pro tento případ znázorňuje Obrázek 2.1.

Obecně je Giniho index definován jako

$$GI = \frac{A}{A + B}, \quad (2.2)$$

kde A je plocha mezi úhlopříčkou (případ absolutně rovnoměrného rozdělení příjmů) a Lorenzovo křivkou, B je plocha pod Lorenzovo křivkou. Pro ilustraci jsou plochy graficky znázorněny na Obrázku 2.3. Je zřejmé, že součet obsahů obou ploch je roven $\frac{1}{2}$, jelikož Lorenzova křivka v případě absolutní rovnosti příjmů, která plochu $A + B$ ohraničuje, odpovídá úhlopříčce jednotkového čtverce. Pokud je Lorenzova křivka vyjádřena pomocí funkce $Y = LC(x)$, lze obsah plochy B vyjádřit pomocí určitého integrálu, tedy

$$B = \int_0^1 LC(x) dx. \quad (2.3)$$

Pokud $LC(X)$ je určena vzorcem využívající známý průběh distribuční funkce (viz. (2.1)), postačí k výpočtu Giniho indexu vzorec (2.2). Pro obecný případ je tento postup zdouhavý. Pomineme-li grafické znázornění pomocí Lorenzovy křivky, je možné koeficient spočítat dle následujícího matematického vzorce jako průměr rozdílů v příjmech y mezi všemi možnými páry jednotlivců [4]

$$GI = \frac{\sum_{i=1}^N \sum_{j=1}^N |y_i - y_j|}{2N^2 \bar{y}}. \quad (2.4)$$

Pokud se Giniho indexem bude měřit nerovnost, kde jsou hodnoty příjmů y_i , $i = 1$ až N indexovány v neklesajícím pořadí ($y_i \leq y_{i+1}$), vyhneme se některým porovnáváním a výpočet se zrychlí. Je použit následující vzorec [4]

$$GI = \frac{2}{N^2 \bar{y}} \sum_{i=1}^N i \cdot (y_i - \bar{y}) \quad (2.5)$$

Ekvivalentní vzorec, který je možné dále matematicky upravit a je vhodný, pokud jsou dostupná data pro jednotlivce, lze zapsat jako [6]

$$GI = \frac{2}{N^2 \bar{y}} \left(\sum_{i=1}^N i \cdot y_i \right) - \frac{N+1}{N}. \quad (2.6)$$

Giniho index sleduje i Světová banka (The World Bank).

2.1.3 Theilův index

K popsání nerovností jsou využívány i indexy související s entropií. Entropie je míra nejistoty. Je možné na ni pohlížet a měřit ji jako chaos, dezorganizaci systému. Tento pojem pochází z fyziky, kde se entropie zavádí v souvislosti se zákony termodynamiky, ale objevuje se i v jiných oblastech, jako například v teorii informace nebo v pravděpodobnosti. Lze se s ní setkat všude, kde se zkoumá pravděpodobnost možných stavů systému. [4]

Entropie, od které se odvozují indexy pro měření nerovností, vycházejí z teorie informace, která entropii definuje jako míru neurčitosti ve zprávě. Jinak také bývá informační entropie nazývána Shannonovou entropií (S). Pokud existuje celkem N navzájem se vylučujících stavů i , pravděpodobnost i -tého stavu je p_i , pak při platnosti vztahu $\sum_{i=1}^N p_i = 1$ lze Shannonovu entropii systému spočítat dle vztahu [7]

$$S = - \sum_{i=1}^N p_i \cdot \ln p_i. \quad (2.7)$$

Entropie je kladné číslo, v krajním případě se rovná 0, pokud pravděpodobnost jednoho z možných stavů je 100%, zatímco ostatní stavy mohou nastat s nulovou pravděpodobností. Naopak maximální entropie nastává, jakmile jsou pravděpodobnosti všech N jevů stejné, tedy pokud pravděpodobnost každého stavu je $\frac{1}{N}$. Maximální entropii lze pak jednoduše vyjádřit jako

$$S_{max} = -N \cdot \frac{1}{N} \cdot \ln \frac{1}{N} = \ln N \quad (2.8)$$

Pokud je vzorec použit v souvislosti s vyšší příjmů, jsou pravděpodobnost jednotlivých stavů p_i ve vzorci nahrazeny výrazem $\frac{y_i}{\sum_{i=1}^N y_i}$, kde N je počet dat (osob), y_i je příjem i - té osoby. Vzorec pro Shannonovu entropii S pak má tvar [7]

$$S = - \sum_{i=1}^N \frac{y_i}{\sum_{i=1}^N y_i} \ln \left(\frac{y_i}{\sum_{i=1}^N y_i} \right). \quad (2.9)$$

Theilův index je definován jako rozdíl maximální možné entropie a Shannonovy entropie pro analyzovaná data. To lze zapsat jako

$$TI = S_{max} - S, \quad (2.10)$$

S využitím vztahů (2.8) a (2.9) je pak odvozen vzorec

$$TI = \frac{1}{N} \cdot \sum_{i=1}^N \left(\frac{y_i}{\bar{y}} \cdot \ln \frac{y_i}{\bar{y}} \right), \quad (2.11)$$

kde \bar{y} je průměrný příjem použitých dat.

Pokud všichni dostávají stejný příjem, což odpovídá maximálnímu řádu v datech, tedy i entropie je maximální, Theilův index nabývá hodnoty 0, protože dle (2.10) je rozdíl od maximální možné entropie roven 0, to je nejmenší možný výsledek. Hodnota indexu je tím větší, čím větší je nerovnoměrnost příjmů jedinců. Pokud v datech žádný řád není, jedinec přijímá veškerý příjem populace, entropie analyzovaných dat je rovna 0. V tomto případě je Theilův index dle formulí (2.10) a (2.8) roven $S_{max} = \ln N$. Theilův index je touto hodnotou shora omezen.

Výsledek Theilova indexu lze znormalizovat, aby dosahoval výsledků 0 až 1, pokud je vydělen $\ln N$. Jedná se pak o tzv. normovaný Theilův index (NTI), který je určován dle vztahu

$$NTI = \frac{1}{N \cdot \ln N} \cdot \sum_{i=1}^N \left(\frac{y_i}{\bar{y}} \cdot \ln \frac{y_i}{\bar{y}} \right). \quad (2.12)$$

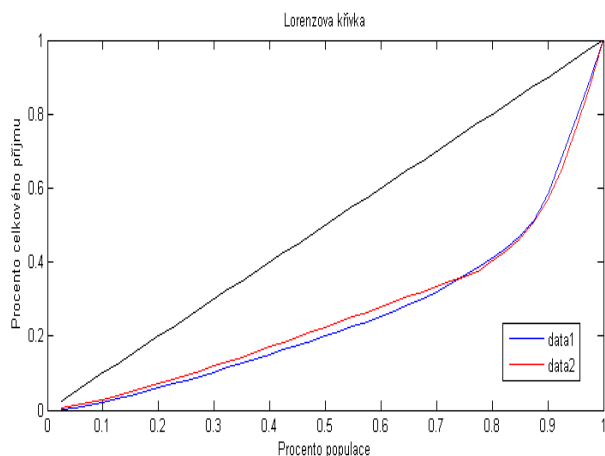
2.1.4 Vztah Giniho a Theilova indexu

Protože Giniho index a Theilův index přistupují k měření nerovností různým způsobem, nastávají situace, kdy pro různá data vychází jeden z indexů stejně, zatímco hodnoty druhého z nich se mohou lišit. To je také důvod, proč v práci budou použity oba tyto indexy.

Byly nasimulovány 2 datové sady, jejichž Theilův index je stejný. V tabulce 2.1 jsou uvedeny hodnoty obou indexů. Jak již bylo zmíněno, hodnota Theilova indexu je shodná pro obě, ale je patrné, že hodnoty Giniho indexu se liší. Pro názornost, Obrázek 2.4 ukazuje, jak by vypadaly Lorenzovy křivky rozdělení příjmů. Plocha mezi křivkami a úhlopříčkou jednotkového čtverce jsou různé.

	Theilův index	Giniho index
1. datová sada	0.4635	0.4854
2. datová sada	0.4635	0.4681

Tabulka 2.1: Přehled hodnot indexů

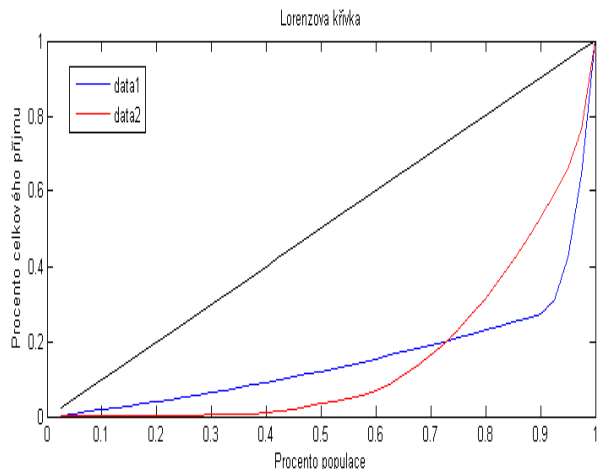


Obrázek 2.4: Grafické znázornění

Může nastat i situace opačná, kdy dvě různá rozdělení příjmů mají stejný Giniho index, ale různý Theilův index. Příklad možných výsledků pro tento případ jsou uvedeny v následující tabulce. Opět se jedná o dvě datové sady, které byly upraveny tak, aby se hodnoty Giniho indexů shodovaly. Ani jedna nebyla použita v předchozím případě. Lorenzovy křivky pro nasimulovaná data jsou na Obrázku 2.5. Obsah ploch, které definují Giniho indexy, jsou pro data stejné. Rozdíl v Lorenzových křivkách, který způsobuje nerovnost Theilových indexů, je, že se křivky nejvíce prohýbají pro jiné procento populace.

	Theilův index	Giniho index
1. datová sada	1.3007	0.6918
2. datová sada	0.9229	0.6918

Tabulka 2.5: Přehled hodnot indexů



Obrázek 2.5: Grafické znázornění

Kromě rovnosti pouze jednoho z výše uvedených indexů se můžeme setkat s jevem, kdy indexy nerozhodnou souhlasně, ve které ze skupin se objevuje větší nerovnost. Příklad můžeme najít v Kapitole 3.4, kde Theilův index dosahuje vyšších hodnot u méně loajálních, zatímco Giniho index určil větší nerovnost ve skupině zaměstnanců, kteří pracují ve firmě 8 let a více.

2.2 Ukazatele míry nerovnosti rozdělení příjmů mezi skupinami

V předchozí kapitole byly popsány indexy, které měřily nerovnost v rámci skupin. Jedinou možností, jak šlo porovnat nerovnost ve dvou skupinách, bylo určit, ve které skupině je nerovnost větší. Lze ale zavést indexy, které pracují se dvěma nebo více skupinami, a jejich výsledek pak dává představu o nerovnosti mezi skupinami navzájem.

2.2.1 Příjmová mezera

Dále popsaný ukazatel je možné použít i pro měření nerovnosti jiné veličiny, mohl by se tedy obecně nazývat např. nerovnostní mezera. Vzhledem k tomu, že se ale nejčastěji objevuje při měření nerovnosti v příjmech, bude v textu používán termín příjmová mezera.

Příjmová mezera (*PG* z anglického Pay Gap) se nejčastěji objevuje v souvislosti ve srovnávání příjmů žen a mužů (tzv. Gender Pay Gap - *GPG*) a je definována jako podíl mediánů (popřípadě průměru) příjmů žen a mediánu (průměru) příjmů mužů, je tedy určen vztahem [10]

$$GPG = \frac{Q_{\text{ž},50}}{Q_{\text{m},50}}, \quad (2.13)$$

kde $Q_{\text{ž},50}$ je medián příjmů žen a $Q_{\text{m},50}$ je medián příjmů mužů. Hodnota *GPG* bývá vyjádřena v procentech a obvykle je nižší než 100 %. Výsledek pak bývá interpretován, kolika procentům mužských příjmů odpovídá příjem žen. [9]

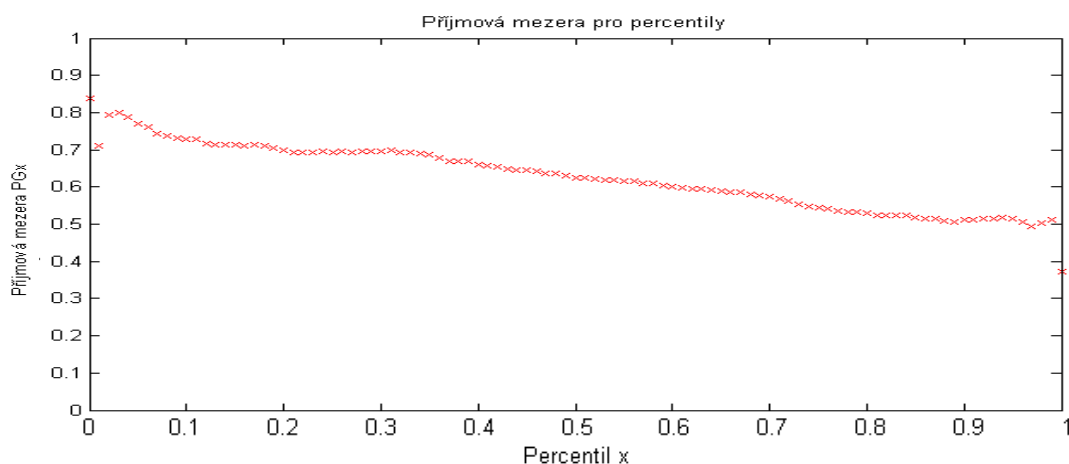
2.2.2 Příjmová mezera pro x -tý percentil

Myšlenku GPG ale lze aplikovat i na skupiny, které vznikly na základě jiného dělení. Pokud bude uvažován podíl pouze mediánů nebo průměrů, jedná se o velmi všeobecný údaj, který bývá nesprávně interpretován. Ukazatel by bylo možné určit jako podíl jakýchkoliv percentilů příjmů jakýchkoliv dvou skupin. Příjmová mezera PG_x pro x -tý percentil by se pak vypočítala jako

$$PG_x = \frac{Q_{1,x}}{Q_{2,x}}, \quad (2.14)$$

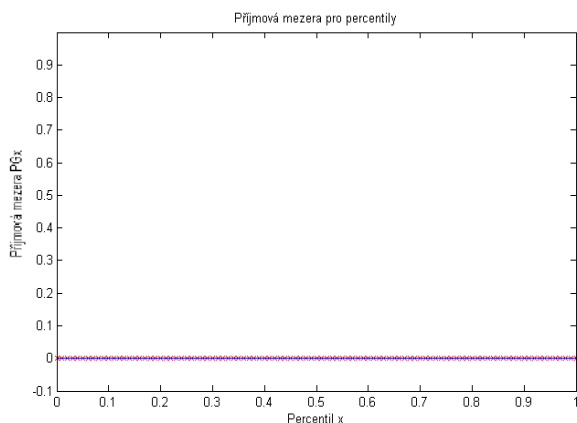
kde $Q_{1,x}$ označuje x -tý percentil první skupiny, kde je předpokládáno, že jde o skupinu s nižšími příjmy než příjmy druhé skupiny, aby PG_x skutečně měla význam mezery mezi příjmy a hodnota PG_x byla nižší než 1. Obdobně $Q_{2,x}$ značí druhou skupinu, jejíž příjmy jsou oproti první vyšší. V literatuře [9] se uvádí, že $0 < x < 1$, pro účely výpočtu PG_x lze dodefinovat i $x = 0$ a $x = 1$. Nulový percentil je pak určen nejnižším číslem ve výběru a stý percentil nejvyšším. Zavedení těchto percentilů bude odůvodněno dále.

Pokud by byla spočítána příjmová mezera pro všechny percentily, bylo by možné získat představu o tom, jak se příjmová mezera mění v závislosti na percentilu. Příkladem možného výsledku výpočtu příjmové mezery pro všechny percentily je graf na následujícím obrázku.

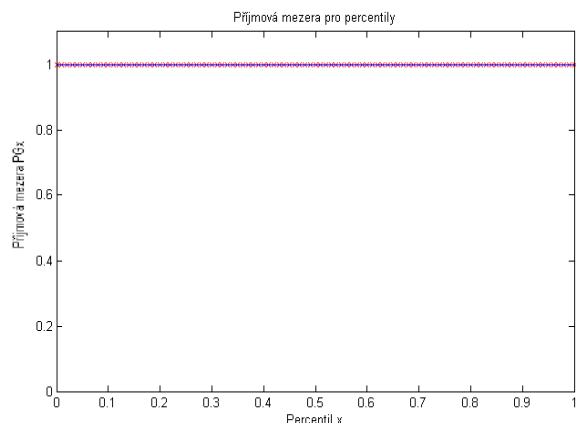


Obrázek 2.6: Příjmová mezera pro percentily

Jelikož se jedná o podíl dvou nezáporných čísel, minimální možná hodnota příjmové mezery je 0. Tuto hodnotu PG_x nabývá, pokud jedna skupina získává veškerý příjem společnosti a příjmy druhé skupiny jsou nulové, příjmové mezery pro každý z percentilů jsou nulové a body grafu leží ve vodorovné přímce $PG_x = 0$. Tuto situaci znázorňuje graf na Obrázku 2.7. Pokud by příjmy byly rozděleny v obou skupinách stejně, mezi skupinami by tedy nastala rovnost, pak by příjmové mezery byly rovné 1 pro všechny percentily, body grafu by ležely ve vodorovné přímce odpovídající $PG_x = 1$, jak znázorňuje Obrázek 2.8



Obrázek 2.7: PG_x -absolutně rovnoměrné rozdělení



Obrázek 2.8: PG_x -absolutně nerovnoměrné rozdělení

2.2.3 Index příjmové mezery

Plocha, kterou vymezuje graf znázorňující příjmové mezery pro percentily je 0 v případě Obrázku 2.7, na Obrázku 2.8 je obsah plochy roven 1. Tento poznatek nás přivádí k myšlence, že bychom mohli zavést index, který míru nerovnosti měří pomocí obsahu plochy, která leží pod grafem příjmové mezery

$$PGI = 1 - \int_0^1 PG_x dx. \quad (2.15)$$

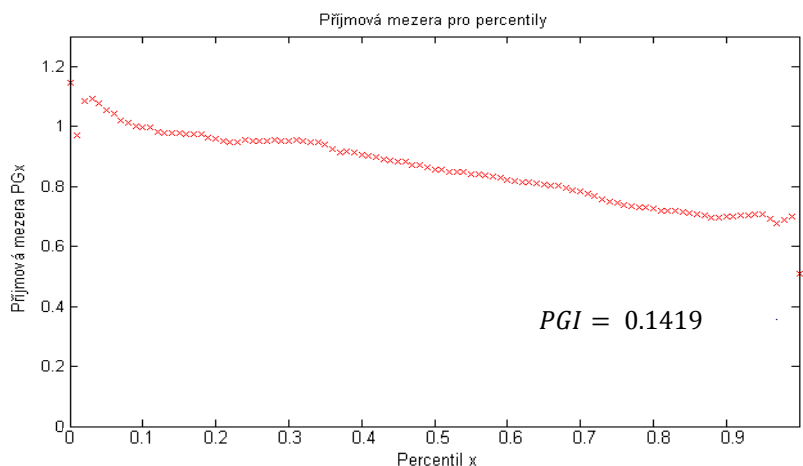
Aby nebyla porušena vlastnost V1, že se ukazatel rovná 0, když v populaci není žádná nerovnost, museli jsme ukazatel nerovnosti pracující s obsahem plochy vymezené grafem příjmové nerovnosti PGI (Pay Gap Index) upravit do výsledného tvaru (2.15).

Ze stejného důvodu byl při zavedení (2.14) určen význam nultého a stého percentilu, pokud by byl obsah počítán pouze od 1. do 99. percentilu, došlo by k porušení vlastnosti V1, přestože došlo k úpravě vzorce.

Ve (2.15) se objevuje integrál funkce PG_x , kterou lze pro reálná data vyčíslit pouze v konečně mnoha bodech. Při výpočtu se uchýlíme k numerickým metodám, kde je interval, na kterém je funkce integrována, rozdělen na několik podintervalů, které se nepřekrývají, ale zároveň pokrývají celý interval. Na takto vzniklých podintervalech je původní funkce nahrazena polynomem, je tedy aproximována funkcí po částech konstantní (obdélníková metoda), lineární (lichoběžníková metoda) nebo kvadratickou (Simpsonova metoda). V práci bude použita lichoběžníková metoda a příjmová mezera bude určena pro celé percentily od $x=0$ do $x=100$.

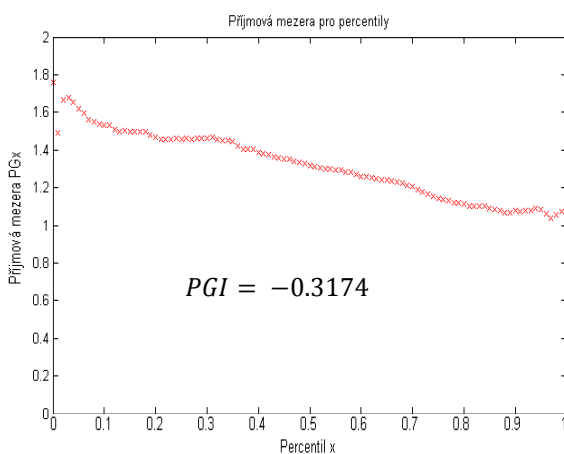
Při zavedení (2.14) je předpokládáno, že se poměří podíl percentilu skupiny s nižšími příjmy a percentilu skupiny s vyššími příjmy. Mohou ale nastat situace, kdy pro některé percentily tento předpoklad neplatí. Na Obrázku 2.9 je vykreslen průběh příjmové mezery pro percentily pro smyšlená data, která znázorňují popsáný případ. Pokud by byly využity výše popsané numerické metody a (2.15) by byl aplikován a modifikován na každý podinterval a hodnota PGI pak dána součtem hodnot na všech podintervalech, hodnota PGI na těch podintervalech, kdy oba krajní body mají hodnotu vyšší než 1, by byla záporná. Na podintervalech, kde pouze jeden z krajních bodů převyšuje hodnotu 1, tento jev může nastat také, záleží na hodnotě druhého krajního bodu. Výsledek PGI tohoto případu, ale bude stále kladný a v intervalu od 0 do 1, konkrétně pro data na Obrázku 2.9 vychází $PGI = 0.1419$.

Zmíněné intervaly způsobí snížení hodnoty indexu. Míra nerovnosti je dle výsledku nižší. Index zohledňuje, že některé percentily finančně méně ohodnocené skupiny jsou vyšší než u druhé skupiny, skupinu jako celek tedy nelze považovat za výrazně diskriminovanou.

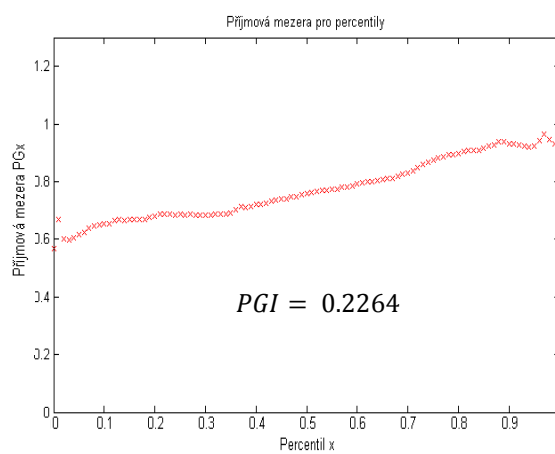


Obrázek 2.9: PG_x – příklad průběhu, kdy jsou některé podíly percentilů větší než 1

Pro Obrázek 2.10 byla vymyšlena data, která ukazují příklad, pokud je percentilů s hodnotami $PG_x > 1$ velké množství a plocha pod grafem funkce větší než 1. V tomto případě by hodnota PGI vyšla záporná. Aby byl počet bodů s hodnotou PG_x nižší než jedna více než s hodnotou PG_x vyšší než jedna, musí být vyměněn čitatel a jmenovatel v (2.14) pro dosažení do (2.15). Obrázek 2.10 tedy zaznamenává případ, kdy je špatně očekáváno, která ze skupin dosahuje vyšších příjmů. PGI pro použitá data dosahuje hodnoty -0.3174 . Při správném použití vzorce pro výpočet PG_x , jsou výsledky vykresleny na Obrázku 2.11, výsledek již není záporný, konkrétně $PGI=0.2264$.



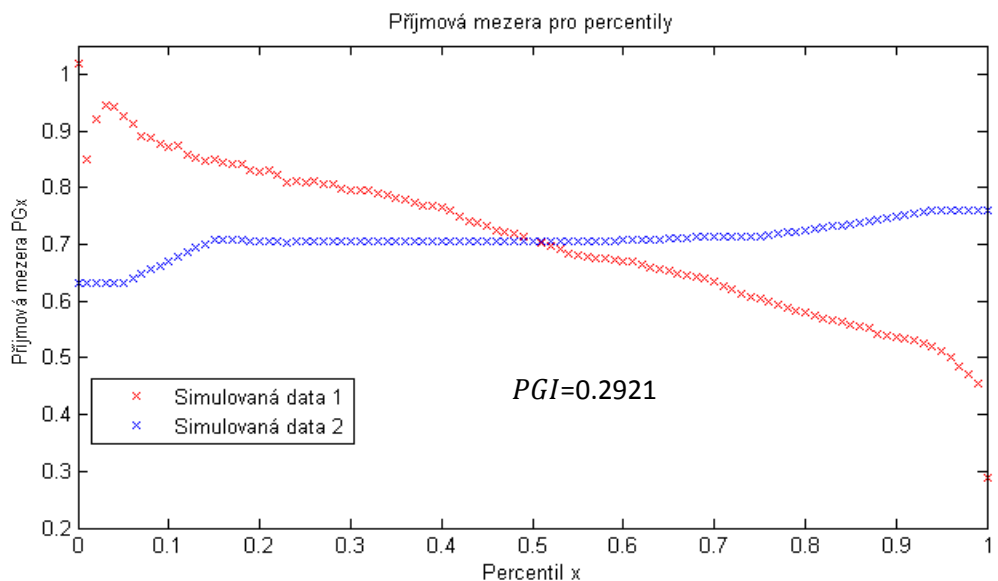
Obrázek 2.10: Příklad špatného užití vzorce pro PG_x



Obrázek 2.11: Korekce použití PG_x z Obrázku 2.10

Nově definovaný vzorec index PGI splňuje všechny vlastnosti, k výpočtu jeho hodnoty poslouží numerické metody a jsou vysvětleny hodnoty indexu, pokud pro všechny percentily neplatí předpokládané rozdělení příjmů mezi skupinami. Na Obrázku 2.9 je názorně ukázáno, jaké má výhody počítat příjmovou mezera pro více percentilů. Pokud by byla příjmová mezera počítána pouze

pro medián, jak je definováno ve vzorci (2.13), který se běžně používá, hodnota příjmové mezery pro obě datové sady na Obrázku 2.12 by byly 0.7042. Tento údaj ale nepodává žádnou informaci o tom, že se hodnota PG_x pro ostatní percentily výrazně liší. Data byla simulována, aby tuto vlastnost ukázala, přestože se nejedná o reálná data, je tento jev možný.



Obrázek 2.12: Příklad různého průběhu PG_x

Simulovaná data také ukazují případ, kdy PGI vychází v obou případech 0.2921. Na první pohled je patrné, že grafy příjmové mezery mají různý průběh. Modře znázorněný graf PG_x na rozdíl od grafu druhé simulované datové sady roste s rostoucím x . Oba grafy svým tvarem připomínají přímku, to nás přivádí k myšlence proložit body přímku a bylo by vhodné pro data uvádět směrnici této přímky, která nejlépe vystihuje průběh příjmové mezery pro percentily.

Směrnici lze určit pomocí lineární regrese [15], kde je vysvětlovanou proměnnou PG_x , snažíme se vysvětlit její závislost na percentilu x . Vzhledem k charakteru vysvětlující proměnné lze předpokládat, že nebudou splněny všechny předpoklady lineární regrese, objevuje se problém s autokorelací. Při nesplnění předpokladů přichází regrese o některé své vlastnosti, intervalové odhady nejsou směrodatné a statistické testy ztrácejí na síle. V práci ale slouží k odhadnutí směrnice, která je pouze doplňkovou informací k PGI , nevyužijeme všechny nástroje lineární regrese, proto ji i přes předpokládané nesplnění předpokladů v práci použijeme.

Pokud by bylo žádoucí porovnat nejbohatší a nejchudší zástupce skupin mezi sebou, mohla by být příjmová mezera nejchudších a nejbohatších kvantifikována a uváděna společně s PGI . Jelikož hodnoty pro první a poslední percentily se často odchyľují od přímky, na které leží ostatní body, použijeme vyhlazenou hodnotu, kterou získáme pomocí průměru PG_x prvních n a posledních n percentilů. Pro počáteční bod pak získáme jeho odhad vyjádřením

$$\hat{A}_n = \frac{\sum_{i=0}^{n-1} PG_i}{n}. \quad (2.16)$$

Pro odhad koncového bodu, který měří příjmovou mezeru nejbohatších, pak bude použit vzorec

$$\hat{A}_{-n} = \frac{\sum_{i=100-n+1}^{100} PG_i}{n} \quad (2.17)$$

V práci budou pro odhady krajních bodů využity předchozí vzorce, kde $n = 10$. Tato hodnota se jeví jako vhodná, nedochází k zprůměrování velkého množství členů, v odhadu krajních bodů tedy nehrají roli body, které leží od krajního bodu daleko. Pro $n = 10$ ve většině případů nedojde ke strhnutí odhadu krajního bodu hodnotou PG_0 nebo PG_{100} . Pokud se ale jedná o odlehlá pozorování, je žádoucí upravit (2.16) a (2.17), aby nezahrnovaly $i=0$ a $i=100$. Dojde-li k modifikaci vzorců, bude pro odhad krajních bodů použito značení \hat{A}_n' popřípadě \hat{A}_{-n}' . Následně také přepočítáme směrnici proložené přímky, kterou odlehlé pozorování taktéž ovlivňuje. Příkladem zlepšení odhadů po nezahrnutí odlehlého pozorování lze nalézt v Kapitole 3.4.

Výsledky pro data použitá na Obrázku 2.12 jsou uvedeny v Tabulce 2.3. Kromě PGI a směrnice přímky jsou zaznamenány také odhady počátečního a koncového bodu.

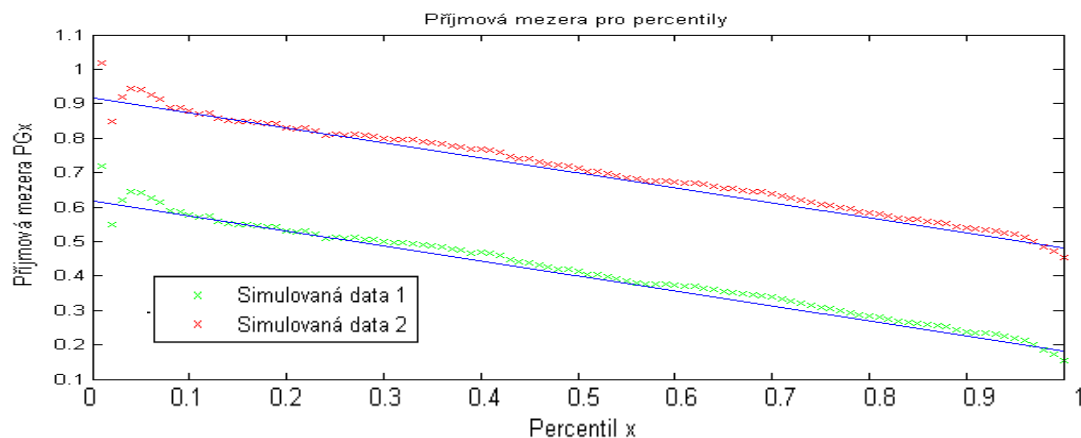
	PGI	\hat{A}_{10}	\hat{A}_{-10}	směrnice
Data 1	0.2921	0.9169	0.4822	-0.4346
Data 2	0.2921	0.6402	0.7584	0.1182

Tabulka 2.2: PGI a směrnice přímky pro data na Obrázku 2.12

Pokud by naopak byla uvedena pouze směrnice, nebylo by možné získat kompletní představu o datech. Na Obrázku 2.13 je vykreslen případ, kdy pro dvojce různá data je směrnice přímky, na které leží, stejná, přestože je patrné, že obsahy pod těmito přímkami jsou různé. Výsledky PGI a koeficienty odhadované přímky a shrnuje Tabulka 2.4.

	PGI	\hat{A}_{10}	\hat{A}_{-10}	směrnice
Data 1	0.2921	0.9169	0.4822	-0.4346
Data 2	0.5921	0.6169	0.1822	-0.4346

Tabulka 2.3: PGI a směrnice přímky pro data na Obrázku 2.13



Obrázek 2.13: Příklad různého průběhu PG_x

Odhad směrnice má smysl, pokud body grafu příjmové mezery pro percentily leží v přímce. Na dosud uvedených obrázcích lze průběh funkce přímkou dobře aproximovat. Rozdělení příjmů ve společnosti ale může být takové, že průběh této funkce připomíná parabolu, popřípadě jinou křivku, v tom případě výpočet směrnice přináší pouze informaci o rozdílu hodnoty n prvních percentilů a n posledních percentilů. Představu o průběhu funkce získáme pouze z vykresleného grafu, proto je vhodné nejprve sestavit graf a zvážit výpovědní hodnotu směrnice.

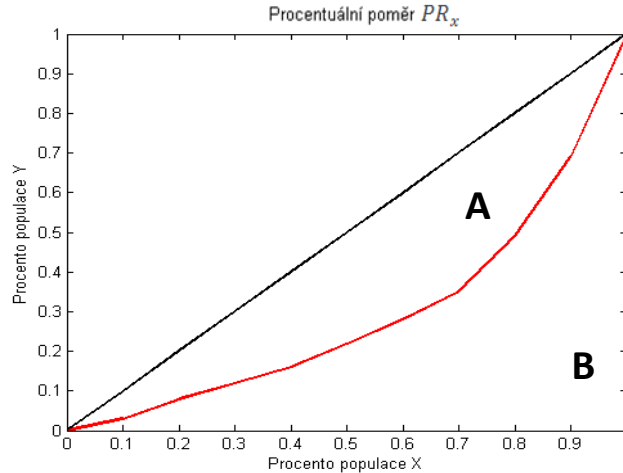
2.2.4 Procentuální poměr

Předchozí index porovnává percentily dvou skupiny navzájem, získáváme představu, jak se liší hodnoty percentilů ve zkoumaných skupinách. Nevýhodu tohoto indexu lze vidět v tom, že nedává představu o tom, zda někdo, popřípadě jaká část hůře ohodnocené skupiny získává příjem, který odpovídá percentilu bohatší skupiny. Lepší představu bychom získali, pokud by index vyjadřoval, kolik procent populace skupiny má příjem menší nebo rovný určitému percentilu skupiny s vyššími příjmy. Takto definovaný index pojmenujeme procentuální poměr (PR - Percentage Ratio). Index by tedy například říkal, že 20 % jedinců z první skupiny má nižší nebo stejný příjem než 5 % jedinců druhé populace. To by znamenalo, že příjmy v první skupině jsou nižší než příjmy ve druhé skupině, jelikož pro stejnou hodnotu příjmu má větší část populace nižší příjem. Hodnotu indexu tedy tvoří uspořádaná dvojice hodnot. Opět by bylo možné nově zavedený index použít i k posuzování nerovností nejen u příjmů.

Podobně jako při definování předchozího indexu je důležité před výpočtem určit, u které skupiny jsou předpokládány vyšší hodnoty zkoumané veličiny. Pro tuto skupinu se určí hodnoty nultého až stého percentilu. Pro každý percentil je pak spočítáno, kolik procent populace druhé skupiny získává nižší příjem. Nejdříve je tedy potřebné určit $x \cdot (100)$ -tý percentil $Q_{Y,x}$ lépe ohodnocené populace, kterou označíme Y . Pro $x \cdot (100)$ -tý percentil se stanoví, jaká část populace s nižším ohodnocením, která bude nazývána X , nabývá nižší příjem, než je hodnota x -tého percentilu populace, získáme tak hodnotu PR_x , která je počítána dle vztahu

$$PR_x = \frac{\text{počet jedinců s hodnotou} \leq Q_{Y,x}}{\text{počet jedinců v populaci } X}. \quad (2.18)$$

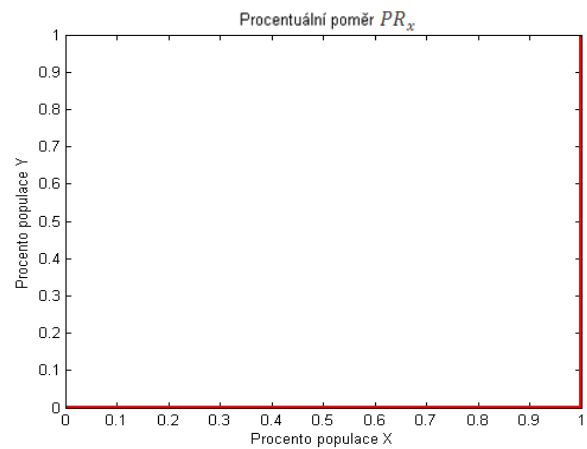
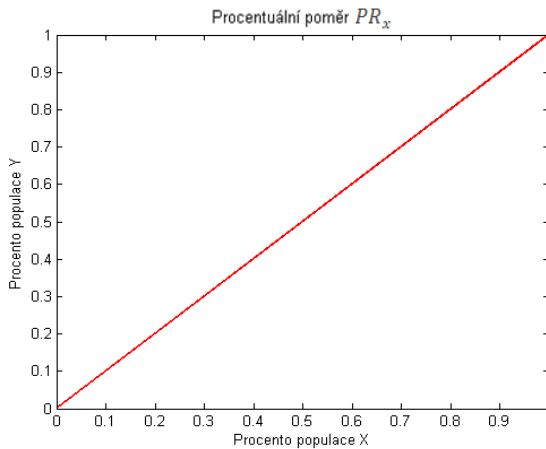
Je vhodné použít tento výpočet pro $x=0$ až $x=1$. O hodnotách indexu více vypoví vykreslení závislosti. Na vodorovné ose je vykresleno procento populace X (PR_x), na svislé ose procento populace Y ($Q_{Y,x}$). Průběh by potom mohl vypadat například jako graf na Obrázku 2.14.



Obrázek 2.14: Příklad průběhu funkce

Pokud by rozdělení příjmů ve společnosti bylo absolutně rovnoměrné, znamenalo by to, že jsou si percentily obou skupin rovny. Interpretace ve smyslu tohoto indexu by pak říkala, že PR_x % populace X má tedy rovný nebo nižší příjem než $Q_{Y,X}$ % populace Y . Grafem by byla přímka *Procento populace X = Procento populace Y* . Průběh funkce pro absolutně rovnoměrné rozdělení příjmu ve společnosti ukazuje Obrázek 2.15.

Druhým extrémním případem je absolutní nerovnoměrné rozdělení příjmů ve společnosti, kde skupina Y pobírá veškeré příjmy, zatímco X má příjmy nulové. Pro všechny percentily skupiny Y by pak platilo, že 100 % populace X pobírá nižší příjmy. Graf tohoto rozdělení příjmů ve společnosti je na Obrázku 2.16.



Obrázek 2.15: Absolutně rovnoměrné rozdělení příjmů

Obrázek 2.16: Absolutně nerovnoměrné rozdělení příjmů

2.2.5 Index procentuálního poměru

Křivka znázorňující průběh funkce procentuálního poměru prochází body $[0,0]$ a $[1,1]$. Čím více se graf prohýbá, o to větší nerovnost se projevuje mezi porovnávanými skupinami. Nelze si nevšimnout určité analogie s Lorenzovou křivkou, která se chová podobně. Na základě Lorenzovy křivky je pak definován Giniho index, který vyjadřuje, jak se zkoumaná data liší od případu absolutně rovnoměrného rozdělení. Proto i pro tuto křivku zavádíme index, který vyčíslí nerovnost v datech. Charakteristiku založenou

na procentuálním poměru pojmenujeme index procentuálního poměru a budeme jej značit PRI (Percentual Ratio Index) a zavedeme ji podobně jako Giniho index pomocí ploch v grafu

$$PRI = \frac{A}{A + B}, \quad (2.19)$$

kde A je plocha vymezena grafem aktuálních dat a případu absolutní rovnosti, B je plocha pod grafem aktuálních dat. Názorně je vše zakresleno na Obrázku 2.14.

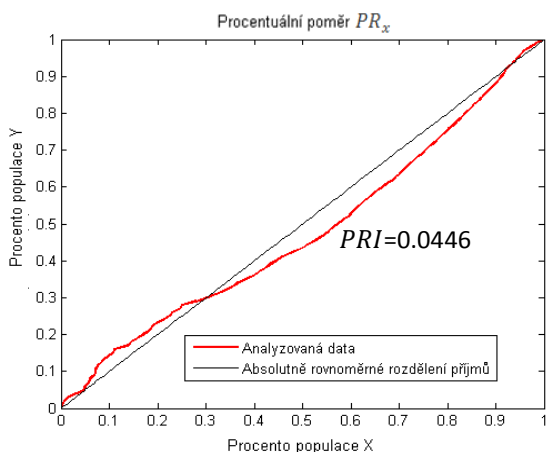
Protože se jedná o jednotkový čtverec, který je předělen úhlopříčkou. Je zřejmé, že součet ploch $A + B = \frac{1}{2}$.

Odtud lze vyjádřit $A = \frac{1}{2} - B$, dosazením do původního vzorce pro PRI . Získáme tak tvar

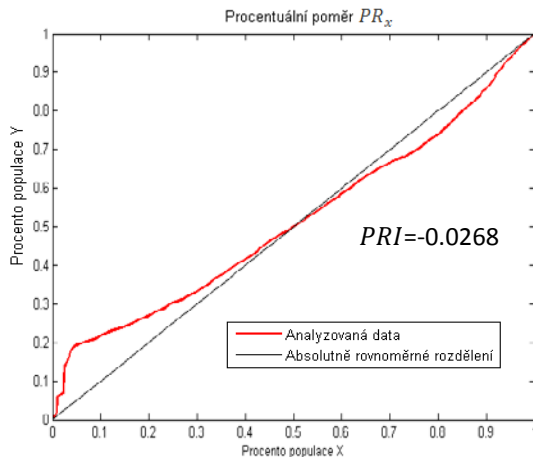
$$PRI = \frac{\frac{1}{2} - B}{\frac{1}{2}} = (1 - 2 \cdot B). \quad (2.20)$$

K výpočtu indexu tedy postačí vypočítat obsah plochy B . K výpočtu plochy pod grafem slouží integrál, podobně jako u předchozího indexu se při výpočtu tohoto integrálu uchýlíme k numerickým metodám.

U nově definovaného indexu PRI může nastat podobně jako u PGI , popsaného v předchozí kapitole, zvláštní případ, kdy nelze jednoznačně určit, která z porovnávaných skupin dosahuje vyšších příjmů. Může se stát, že pro některé percentily se jeví skupina X jako bohatší, ale pro další percentily se nerovnost změní a skupina X se stává pro jiné percentily hůře oceňovanou skupinou. Příklad tohoto průběhu je znázorněn na následujících obrázcích, které pracují se smyšlenými daty, jejichž úkolem je demonstrovat popisovanou situaci, která se může v datech vyskytnout.



Obrázek 2.17: Příklad nejednoznačné nerovnosti

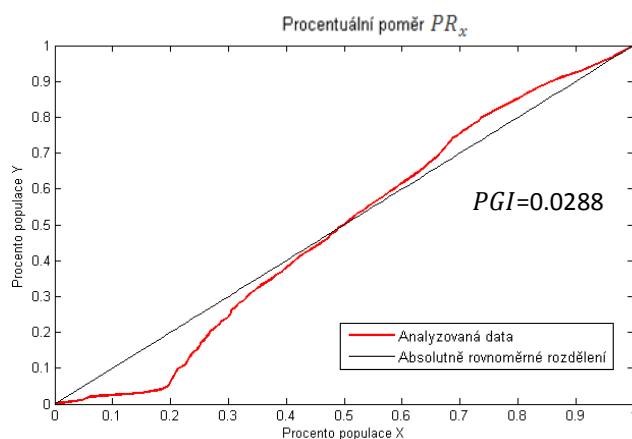


Obrázek 2.18: Příklad špatného užití PR_x

Na Obrázku 2.17 smyšlená data ukazují případ, kdy populace X je pro více než polovinu případů hůře ohodnocenou skupinou, body grafu PR_x leží pod grafem absolutně rovnoměrného rozdělení znázorněného v obrázku černě. Výsledek indexu PRI je 0.0446, je tedy kladný. Při výpočtu dle (2.20) s užitím výpočtu integrálu pomocí lichoběžníkové metody vychází na podintervalech, jejichž krajní body leží nad úhlopříčkou jednotkového čtverce, záporný výsledek, jelikož plocha B neleží pod zmíněnou

úhlopříčkou, jak je ve vzorci předpokládáno. Plocha nad úhlopříčkou je ale nižší než plocha pod ní. Výsledek tedy zůstává kladný, skupina X je tedy spíše hůře oceňovanou společenskou skupinou.

Na Obrázku 2.18 je vyobrazen příklad dat, kde bylo předpokládáno, že skupina X je diskriminována. Z obrázku je ale patrné, že plocha vymezená grafem nad úhlopříčkou má větší obsah než plocha pod ní a zároveň počet percentilů, pro které je skupina X lépe ohodnocená je vyšší než počet percentilů, pro které je ohodnocená hůře. Špatný předpoklad se projeví také na výsledku PRI , který pro tento případ vychází záporně, konkrétně je výsledkem hodnota $PRI = -0.0268$. Pokud je výsledek záporný, je třeba provést výpočet indexu znovu, ale vyměnit skupiny. Průběh grafu po korekci zachycuje Obrázek 2.19, PRI již nevychází záporně, jeho výsledek po úpravě je 0.0288. Aby bylo možné rozpoznat, ve které skupině se objevují vyšší příjmy, nepracuje (2.20) s absolutní hodnotou, pomocí které by se vzdálenost od grafu absolutně rovnoměrného rozdělení dala vyjádřit.



Obrázek 2.19: Korekce použití PR_x

2.2.6 Kolmogorovův-Smirnovův dvouvýběrový test

K ověření, zda výběry příjmů v obou skupinách, pochází ze stejného rozdělení pravděpodobnosti, slouží Kolmogorovův-Smirnovův test pro dva výběry. Nulová hypotéza říká, že dva výběry jsou ze stejného rozdělení pravděpodobnosti. Alternativní hypotézou je tvrzení, že výběry nejsou ze stejného rozdělení.

Provedení testu pro dva nezávislé výběry X_1, \dots, X_n a Y_1, \dots, Y_m spočívá v tom, že se vypočítají empirické distribuční funkce $F_n(x)$ a $G_m(y)$, pro které platí

$$F_n(x) = \frac{\text{počet } x_i \leq x}{n}, G_m(y) = \frac{\text{počet } y_i \leq y}{m}. \quad (2.21)$$

Testovací kritérium $D_{m,n}$ je dáno jako supremum absolutní hodnoty rozdílu empirických funkcí $F_n(x)$ a $G_m(y)$.

Nulová hypotéza se zamítá na hladině významnosti α , jestliže je hodnota kritéria $D_{n,m}$ větší než kritická hodnota $D_{(1-\alpha)}(n, m)$, v opačném případě se přijímá, přičemž $D_{(1-\alpha)}(n, m)$ je tabelovaná hodnota. [16]

3 Zpracování dat

Předchozí kapitola definovala indexy měřící nerovnosti v populaci, nyní je aplikujeme v praxi a popíšeme nerovnosti, které se objevují v příjmech obyvatel Argentiny. Získaný datový soubor byl již popsán v první kapitole, kde je rovněž zmíněno, že příjmy lze rozdělit na základě dostupných informací do několika skupin a zkoumat tak i nerovnosti mezi skupinami navzájem.

K výpočtům byly použity kódy softwaru Matlab, které data třídí do skupin a spočítají hodnoty indexů sloužících k měření nerovnosti nebo používají vhodné obecně známé statistické nástroje. Kódy jsou uvedeny na příloženém CD.

3.1 Obecná nerovnost v zemi

Giniho index pro zpracovávaná data vychází 0.3524 (35.24 %). Vypočítanou hodnotu můžeme porovnat s výsledkem, který určila World Bank (Světová banka), jedna z organizací OSN. Odhady této organizace Giniho indexu z posledních dostupných let jsou uvedeny v následující tabulce. Námi získaná data jsou platná k březnu 2015. World Bank výsledky za roky 2014 ani 2015 zatím nezveřejnila, z hodnot předchozích let ale lze získat určitou představu, kolem kterých hodnot by se odhad z roku 2015 pohyboval.

Rok	2011	2012	2013
Hodnota Giniho indexu	43.6 %	42.5 %	42.3 %

Tabulka 3.1: Hodnoty Giniho indexu dle World Bank, Zdroj [11]

Jak je z tabulky patrné, hodnota určená ze zpracovávaných dat je nižší než hodnoty World Bank. Rozdíl mohl vzniknout již zmíněným nesouladem platnosti dat. Z údajů v tabulce lze usoudit, že hodnoty v průběhu uvedených let klesaly. Pokud by se tento trend uchoval, byla by hodnota z roku 2015 také nižší než hodnota z roku 2013. Dalším důvodem může být fakt, že data pocházejí z vybraných soukromých firem. Rozhodně se v datech nevyskytují informace o příjmech státních zaměstnanců, jako jsou například poslanci, ministři atd. Chybí ale také údaje o příjmech nezaměstnaných, žen na mateřské dovolené, osob pracujících jen na částečný úvazek. World Bank zjišťuje i výši příjmů skupin osob, které v získaném výběru chybí.

Dalším možným indexem nerovnosti, který lze celkově pro všechna data použít je Theilův index. Hodnota tohoto indexu pro používaná data je 0.2293. Hodnota normovaného Theilova indexu je 0.0219 (2.19 %). Výsledky Theilova indexu lze opět porovnat s výsledky organizace World Bank, hodnoty, které zveřejnila na svých stránkách, jsou uvedeny v následující Tabulce 3.2.

Rok	2010	2012
Hodnota Theilova indexu	0.3460	0.3120

Tabulka 3.2: Hodnoty Theilova indexu dle World Bank, Zdroj [12]

Vypočítaná hodnota a hodnota World Bank se liší, získaná hodnota je nižší než hodnota uvedená na oficiálních stránkách organizace. K rozdílu mohlo dojít z obdobných důvodů jako u Giniho indexu.

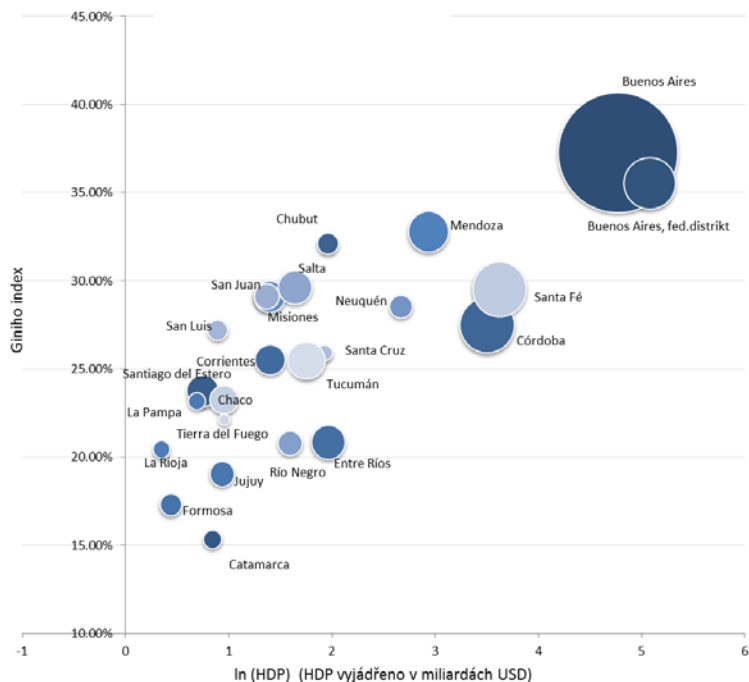
3.2 Regionální příjmové nerovnosti

V úvodní kapitole již bylo řečeno, že Argentina se dělí na 24 regionů, v Tabulce 1.3 jsou uvedené četnosti jednotlivých oblastí ve zkoumaném výběru. Výsledky základních indexů jako je Giniho index a Theilův indexy v regionech jsou uvedeny v následující tabulce. Hodnoty, které nabývají vyšších hodnot než výsledky pro zemi jako celek uvedené v Kapitole 3.1, jsou červeně podtrženy.

Provincie	Giniho index	Theilův index
Buenos Aires, federální distrikt	<u>0.3731</u>	<u>0.2646</u>
Buenos Aires	<u>0.3552</u>	<u>0.2455</u>
Catamarca	0.1535	0.0435
Chaco	0.2373	0.1008
Chubut	0.3212	0.1848
Córdoba	0.2746	0.1339
Corrientes	0.2556	0.1190
Entre Ríos	0.2087	0.0794
Formosa	0.1730	0.0504
Jujuy	0.1907	0.0602
La Pampa	0.2319	0.0983
La Rioja	0.2046	0.0734
Mendoza	0.3280	0.1857
Misiones	0.2911	0.1633
Neuquén	0.2856	0.1350
Río Negro	0.2082	0.0785
Salta	0.2964	0.1897
San Juan	0.2911	0.1589
San Luis	0.2721	0.1264
Santa Cruz	0.2591	0.1147
Santa Fé	0.2954	0.1596
Santiago del Estero	0.2325	0.0909
Tierra del Fuego	0.2215	0.0887
Tucumán	0.2548	0.1252

Tabulka 3.3: Výsledky Giniho a Theilova indexu v jednotlivých regionech

Nejnižších hodnot nabývají hodnoty indexů v regionu s názvem Catamarca, jedná se o jednu z ekonomicky méně významných oblastí Argentiny, která se zaměřuje výhradně na zemědělství. Druhé nejnižší hodnoty byly vypočítány pro region Formosa. Tato provincie dlouhou dobu patřila k nejchudším regionům v Argentině, její ekonomika je založená na chovu dobytka a dalších zemědělských aktivitách. Oba regiony se nacházejí na severu Argentiny, tato oblast je obecně nedostatečně industrializovaná. Lepší představu udává bublinový graf na Obrázku 3.1, který vizualizuje závislost Giniho indexu na HDP regionu (vyjádřeno v miliardách USD), pro lepší přehlednost je na vodorovné ose zvoleno logaritmické měřítko. Velikost bublin je určena počtem obyvatel regionu. Použité informace o regionech lze nalézt v Tabulce 1.3.



Obrázek 3.1: Závislost Giniho indexu na ln(HDP), kde velikost bublin závisí na počtu obyvatel regionu

Nejvyšších hodnot nabývají oba indexy ve federálním distriktu Buenos Aires a v provincii Buenos Aires. V těchto 2 regionech se rozkládá hlavní město Buenos Aires a jeho aglomerace. Souhrnně se tato oblast nazývá El Gran Buenos Aires (GBA) a lze rozdělit na 4 celky: Capital federal, Zona Norte, Zona Oeste a Zona Sur, pouze první z nich leží v regionu Buenos Aires, federální distrikt, ostatní patří do regionu Buenos Aires. V datech lze u zaměstnanců z provincie Buenos Aires rozpoznat, zda pochází z některé z těchto zón, nebo zda se jedná o jinou oblast tohoto regionu.

GBA je zdaleka nejvýznamnější průmyslové a dopravní centrum Argentiny. Lze předpokládat rozdílnou strukturu zaměstnanosti v tomto centru a v ostatních regionech. V hlavním městě a v jeho aglomeraci je umístěno velké množství podnikatelských subjektů se zahraniční kapitálovou účastí, centrály významných společností (např. bank, pojišťoven), centrální úřady, které se mimo hlavní město objevují jen zřídka. Zaměstnanci v takových firmách mohou pracovat na vysokých pozicích, které jsou dobře ohodnoceny, zároveň se ale v hlavním městě vyskytují pracovní místa s nízkým příjmem.

Uvedené důvody nás motivují k ověření, zda se nerovnosti v hlavním městě a mimo něj liší. Protože umíme rozlišit osoby zaměstnané v této oblasti hlavního města a mimo tuto oblast, rozdělíme obyvatele do 2 skupin podle tohoto kritéria.

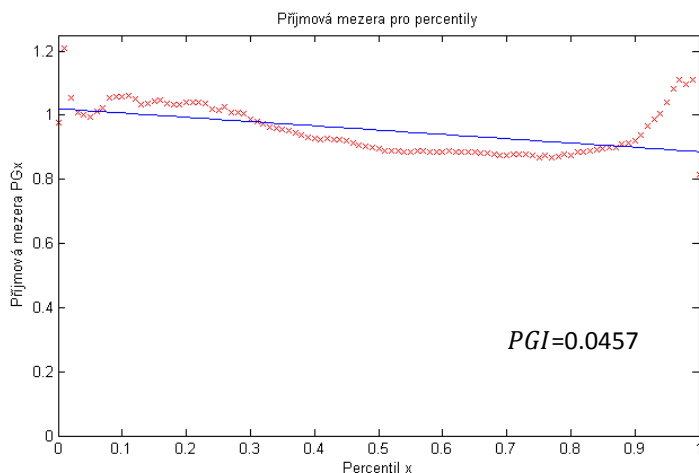
	Giniho index	Theilův index
V hlavním městě	<u>0.3538</u>	<u>0.2430</u>
Mimo hlavní město	0.3495	0.2170

Tabulka 3.4: Výsledky Giniho a Theilova indexu příjmů v hlavním městě a mimo něj

V hlavním městě je dle Tabulky 3.4 míra nerovnosti vyšší, hodnoty indexů převyšují i dříve uvedené výsledky pro zemi jako celek.

Pro porovnání příjmů obyvatel v hlavním městě a mimo něj lze také použít příjmovou mezeru pro percentily PG_x , která dává do poměru stejné percentily příjmů obyvatel v zavedených skupinách. Z tohoto ukazatele je dále odvozen Pay Gap Index PGI , vyrovnané hodnoty krajních bodů a směrnice přímkou, která je proložena body PG_x . (Kapitola 2.2.2)

Nejprve bylo předpokládáno, že v GBA dosahují obyvatelé vyšších příjmů než obyvatelé žijící v jiných obcích. Jelikož ale PGI při tomto předpokladu vycházel záporně ($PGI = -0.0541$), bylo nutné změnit předpoklad a index přepočítat. Na následujícím obrázku je zobrazen průběh PG_x , kde bohatší skupinou jsou obyvatelé mimo hlavní město.



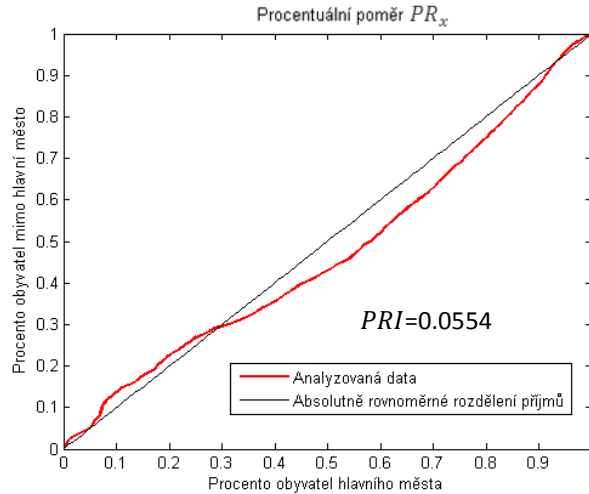
Obrázek 3.2: Příjmová mezera PG_x příjmů v hlavním městě a mimo něj

Obrázek 3.2 ukazuje, že nejbohatší a nejchudší obyvatelé hlavního města mají vyšší příjmy než osoby pocházející z jiných oblastí. Příjmová mezera pro všechny percentily od 90. percentilu a pro většinu do 30. percentilu vychází větší než 1, to znamená, že pro tyto percentily mají obyvatelé hlavního města vyšší příjmy než osoby ve druhé porovnávané skupině. Velká část percentilů se ale pohybuje pod hodnotou 1. Hodnota indexu PGI a doplňujících ukazatelů je uvedena v Tabulce 3.5.

PGI	\hat{A}_{10}	\hat{A}_{-10}	směrnice
0.0457	1.0396	1.0157	-0.1340

Tabulka 3.5: Výsledky PGI příjmů v hlavním městě a mimo něj

Dalším nově zavedeným indexem byl index procentuálního poměru PRI . (Kapitola 2.2.5) Na základě výsledků předchozího indexu již bylo předpokládáno, že obyvatelé mimo hlavní město jsou lépe ohodnocenou skupinou. Graf procentuálních poměrů, ze kterého je index odvozen, uvádí Obrázek 3.3.



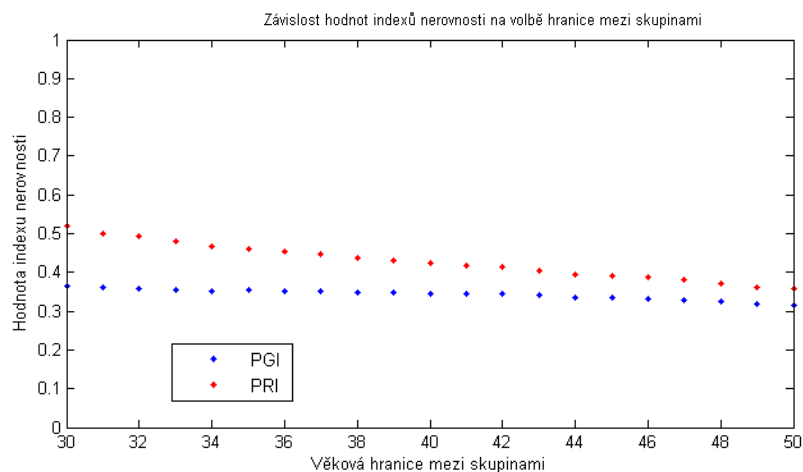
Obrázek 3.3: Procentuální poměr PR_x příjmů v hlavním městě a mimo něj

Analyzovaná data se od případu, kdy by byly příjmy absolutně rovnoměrně rozdělené, výrazněji neliší, o čemž také vypovídá hodnota indexu PRI , která je pro uvažované dělení skupin 0.0554.

Oba indexy nabývají nízkých hodnot. Obrázek 3.3 naznačuje, že jsou příjmy v hlavním městě a mimo město rozděleny rovnoměrně, proto byl aplikován Kolmogorovův-Smirnovův test. Na hladině významnosti 5 % zamítáme nulovou hypotézu, výběry tedy nepochází ze stejného rozdělení pravděpodobnosti, p -hodnota testu je $2.11 \cdot 10^{-54}$.

3.3 Příjmové nerovnosti věkových skupin

Vliv na výši příjmu má také věk zaměstnance. Firmy do vyšších funkcí jen málokdy hledají mladé absolventy, naopak sázejí na zkušené zájemce s praxí. Věkovou hranici, kdy lze o zaměstnanci tvrdit, že již dosáhl věku, kdy by mohl být považován za zkušeného, lze určit velmi obtížně. Budeme tedy uvažovat různý věk, který je předělem mezi skupinou starších a mladších. Na Obrázku 3.4 jsou zaznamenány hodnoty indexů, které měří nerovnost mezi skupinami navzájem, v závislosti na hranici, kde dochází k předělu starší a mladší věkové skupiny. Je patrné, že pro uvažované předěly je vždy skupinou s většími příjmy starší z těchto skupin. Nerovnost klesá s rostoucí hranicí mezi skupinami.



Obrázek 3.4: Závislost hodnot indexů na volbě hranice mezi skupinami, četnosti skupin uvedeny v Tabulce 3.6

Tabulka 3.6 shrnuje výsledky, které zpracovává Obrázek 3.4, navíc uvádí i počet osob ve skupinách.

Hranice	<i>PGI</i>	<i>PRI</i>	Četnost mladších nebo stejně starých	Četnost starších
30	0.3651	0.5188	8366	27358
31	0.3603	0.5015	9787	25937
32	0.3596	0.4924	11219	24505
33	0.3563	0.4794	12662	23062
34	0.3527	0.4663	14078	21646
35	0.3536	0.4614	15619	20105
36	0.3530	0.4552	17043	18681
37	0.3508	0.4464	18317	17407
38	0.3484	0.4378	19580	16144
39	0.3471	0.4304	20892	14832
40	0.3454	0.4230	22109	13615
41	0.3444	0.4178	23204	12520
42	0.3445	0.4141	24216	11508
43	0.3403	0.4058	25147	10577
44	0.3346	0.3947	26090	9634
45	0.3355	0.3925	26932	8792
46	0.3331	0.3871	27648	8076
47	0.3300	0.3813	28322	7402
48	0.3235	0.3699	28989	6735
49	0.3193	0.3629	29588	6136
50	0.3153	0.3587	30122	5602

Tabulka 3.6: Hodnoty indexů pro různou hranici mezi skupinami a četnosti v obou skupinách

Následující tabulka zaznamenává, jak by se změnilы hodnoty indexů nerovnosti, pokud by byla měřena nerovnost mezi skupinami po vynechání jedinců, jejichž věk leží v rozhraní 30 až 50 let, a proto je obtížné určit, do které věkové skupiny patří. Po porovnání s Tabulkou 3.6 lze konstatovat, že nerovnost mezi mladší a starší generací se projevuje více, pokud do zkoumaných skupin není zahrnuta populace tzv. středního věku.

<i>PRI</i>	<i>PGI</i>	\hat{A}_{10}	\hat{A}_{-10}	směrnice
0.6882	0.4887	0.8442	0.2611	-0.5831

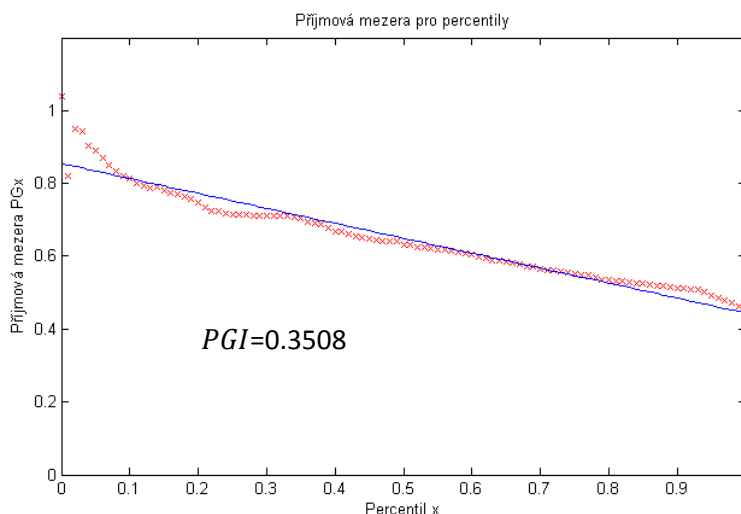
Tabulka 3.7: Výsledky *PRI* a *PGI* příjmů pro starší 50 let a ve věku méně než 30

Na základě Tabulky 3.6 byl jako věková hranice, která dělí mladší a starší věkovou kategorii, určen věk 37, jelikož četnosti obou skupin nabývají nejnižšího rozdílu. Pro vzniklé věkové skupiny byly určeny hodnoty Giniho a Theilova indexu. Dle Tabulky 3.8 lze konstatovat, že oba indexy dosahují vyšších hodnot pro starší věkovou skupinu, výsledky jsou dokonce vyšší než pro zemi jako celek, proto jsou tyto hodnoty červeně podtrženy.

	Giniho index	Theilův index
37 let a mladší	0.2670	0.1287
Starší 37 let	<u>0.3625</u>	<u>0.2327</u>

Tabulka 3.8: Výsledky Giniho a Theilova indexu příjmů věkových skupin

Příjmová mezera pro percentily PG_x dává v tomto případě do poměru percentily příjmu věkové skupiny 37 let a mladší a percentily starší věkové skupiny. Průběh PG_x ukazuje Obrázek 3.5.



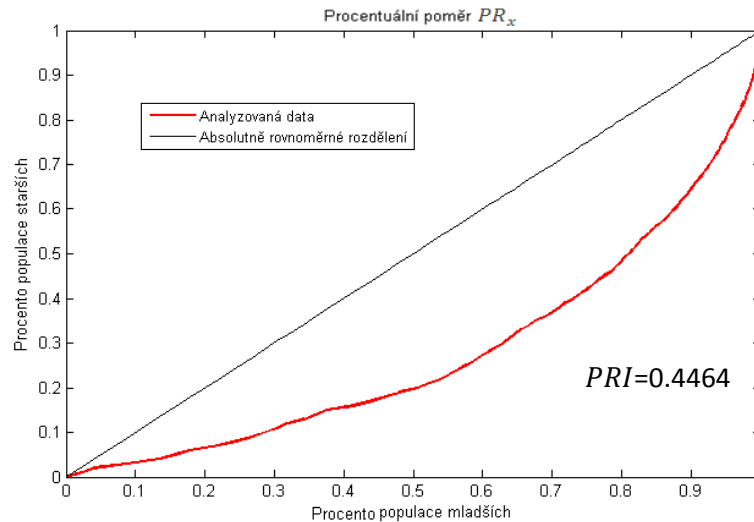
Obrázek 3.5: Příjmová mezera PG_x pro starší 37 let a ve věku 37 a méně

Zatímco nejhůře oceňovaný zástupce mladší věkové kategorie ročně získá 105 % příjmu osoby ze starší věkové kategorie s nejnižším příjmem, osoba s nejvyšším příjmem v rámci skupiny mladších pobírá 30 % příjmu nejbohatšího zaměstnance staršího 37 let. Jak Obrázek 3.5 napovídá, PG_x klesá s rostoucím percentilem x , z tohoto důvodu dosahuje PGI vycházející z obsahu plochy vymezené grafem funkce PG_x poměrně vysoké hodnoty. Příjmová mezera se prohlubuje s rostoucími příjmy, největší rozdíly mezi skupinami se objevují u osob s nejvyššími příjmy, což potvrzuje příjmová mezera nejbohatších \hat{A}_{-10} , nejbohatší mladí dosahují 47.19 % příjmu nejbohatších starších. Ukazatele popisující chování PG_x jsou uvedeny v následující tabulce.

<i>PGI</i>	\hat{A}_{10}	\hat{A}_{-10}	směrnice
0.3508	0.8925	0.4719	-0.4104

Tabulka 3.9: Výsledky *PGI* příjmů pro starší 37 let a ve věku 37 a méně

Dalším způsobem měření nerovnosti, který lze pro věkové skupiny použít, je procentuální poměr PR_x . Uvažujeme, že vyšších příjmů dosahují zástupci starší věkové kategorie. Na následujícím obrázku je znázorněn průběh zmiňovaného ukazatele. Dále je určena hodnota Percentage Ratio Index PRI , která je v tomto případě rovna 0.4464.



Obrázek 3.6: Procentuální poměr PR_x pro starší 37 let a ve věku 37 a méně

3.4 Příjmové nerovnosti způsobené dobou výkonu zaměstnání

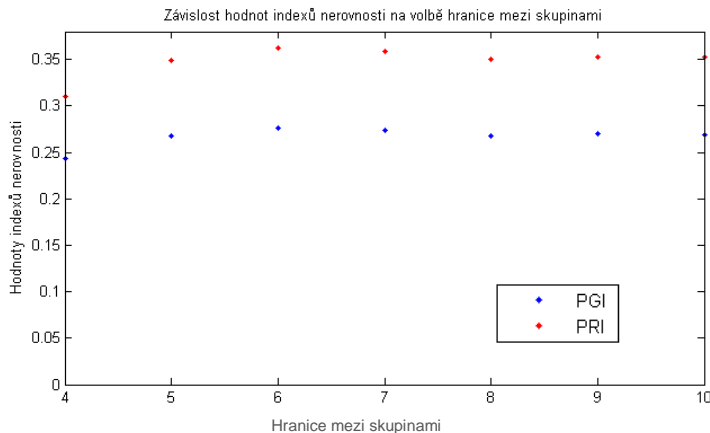
Personální politika firmy často dokáže ocenit loajalitu kvalitních zaměstnanců, po několika letech u jednoho zaměstnavatele může práce přerůst ve stereotypní činnost. Případná změna místa může přinést nejen novou motivaci ale i platový postup.

Doba zaměstnání u jedné firmy je ale zároveň ovlivněna věkem uchazeče, dvacetiletý zaměstnanec nemůže mít odpracováno 15 let u jedné firmy, zatímco padesátiletý ano. Korelaci mezi věkem a počtem let v jedné firmě společně s korelací uvedených ukazatelů a výší příjmu uvádí Tabulka 3.10, jež má strukturu korelační matice. Všechny korelační koeficienty jsou kladné, což značí přímou závislost, pokud jedna veličina roste, poroste i druhá.

	Věk	Počet let ve firmě	Příjem
Věk	1.0000	0.6875	0.3588
Počet let ve firmě	0.6875	1.0000	0.2113
Příjem	0.3588	0.2113	1.0000

Tabulka 3.10: Korelační matice

Podobně jako v předchozí kapitole závisí hranice loajálnosti na tom, kým je posuzována. Nemůžeme bez jakéhokoliv seznámení s daty konstatovat, kde bude hranice mezi skupinou loajálních a méně loajálních zaměstnanců. Hodnoty indexů nerovnosti v závislosti na předělovém počtu let ve firmě, po kterém lze zaměstnance pokládat za loajálního, ukazuje Obrázek 3.7. Na základě vykreslené závislosti lze říci, že hodnoty ukazatelů se výrazně nemění a loajální skupina zaměstnanců ve všech znázorněných případech získává vyšší příjmy než méně loajální.



Obrázek 3.7: Závislost hodnot indexů nerovnosti na volbě hranice mezi skupinami, četnosti skupin uvedeny v Tabulce 3.11

Výsledky, které zobrazuje Obrázek 3.7, jsou zaznamenány v Tabulce 3.11, která uvádí i četnosti v jednotlivých skupinách. Hodnoty *PGI* se pohybují v rozmezí od 0.2432 do 0.2757, hodnoty *PRI* dosahují hodnot od 0.3103 do 0.3622. Oproti Tabulce 3.6, která shrnuje výsledky indexů pro různou věkovou hranici mezi skupinami, neobjevuje se růst indexů s rostoucí hodnotou, která určuje předěl mezi skupinami.

Hranice	<i>PGI</i>	<i>PRI</i>	Četnost zaměstnaných kratší dobu	Četnost zaměstnaných daný počet let nebo déle
4	0.2432	0.3103	9415	26309
5	0.2672	0.3489	12441	23283
6	0.2757	0.3622	14721	21003
7	0.2736	0.3592	16115	19609
8	0.2680	0.3500	18162	17562
9	0.2702	0.3526	20313	15411
10	0.2694	0.3526	22102	13622

Tabulka 3.11: Hodnoty indexů pro různou hranici mezi skupinami a četnosti v obou skupinách

Výsledky indexů nerovnosti, které by porovnávaly část populace, kterou zaměstnává 1 zaměstnavatel více než 10 let, a skupinu, která je zaměstnána méně než 4 roky, shrnuje Tabulka 3.12. Nerovnost je vyšší než v případě, kdy je mezi skupiny zahrnuta i část zaměstnanců, kteří pracují 4 až 10 let u jedné společnosti.

<i>PRI</i>	<i>PGI</i>	\hat{A}_{10}	\hat{A}_{-10}	směrnice
0.4579	0.3353	0.7845	0.6432	-0.1433

Tabulka 3.12: Výsledky *PRI* a *PGI* příjmů obyvatel zaměstnaných více než 10 let a méně než 4

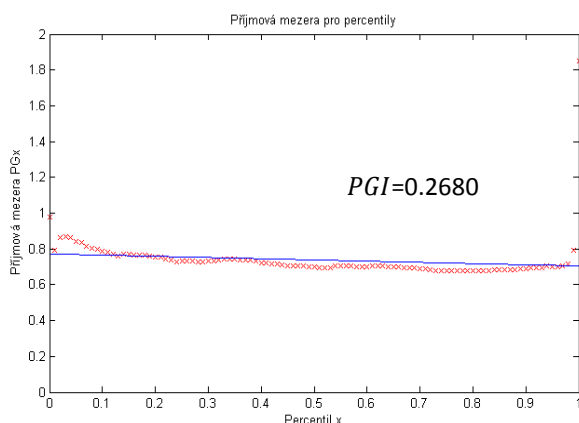
Zaměstnanci jsou rozděleni do 2 skupin z hlediska toho, jak dlouho setrvávají v jedné firmě. První skupinu tvoří loajální zaměstnanci, kteří ve firmě pracují 8 let a více, ve výběru se jich objevuje 17 562. Druhou skupinou jsou zaměstnanci, kteří v jedné firmě pracují méně než 8 let, pro ty máme 18 162 zástupců. Tato hranice byla opět vybrána s ohledem na četnosti obou skupin, jejichž rozdíl je v uvedeném případě nejmenší.

Hodnoty Giniho a Theilova indexu jsou uvedeny v následující tabulce. Hodnoty se od sebe liší v rámci setin. Větších nerovností dle uvedených výsledků dosahují příjmy loajálních zaměstnanců, pro obě skupiny však nabývají indexy nižších hodnot, než které byly určeny pro zemi jako celek.

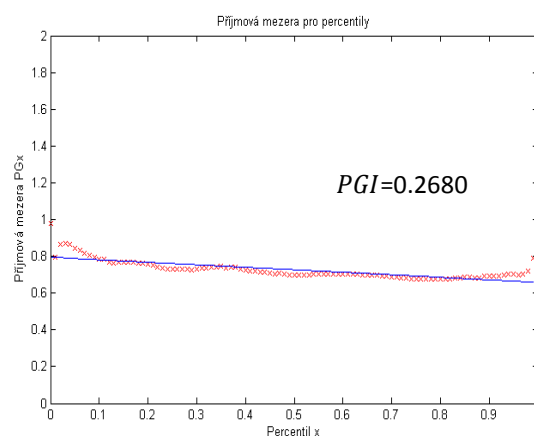
	Giniho index	Theilův index
Loajální	0.3435	0.2095
Méně loajální	0.3359	0.2250

Tabulka 3.13: Výsledky Giniho a Theilova indexu příjmů loajálních a nestálých zaměstnanců

Pro výpočet příjmové mezery je nejprve třeba určit, ve které skupině lze očekávat vyšší příjmy. Použijeme předpoklad, že loajální zaměstnanci jsou lépe oceňováni. Průběh příjmové mezery pro percentily je pak znázorněn na následujícím Obrázku 3.8, ze kterého je patrné, že příjmová mezera pro stý percentil lze považovat za odlehle pozorování, které výrazně ovlivňuje odhad koncového bodu \hat{A}_{-10} a tím i směrnice proložené přímkou. Na Obrázku 3.9 je pro odhad tohoto bodu použit průměr pouze 91. až 99. percentilu, jak je z obrázku patrné, proložená přímkou lépe vystihuje průběh dat.



Obrázek 3.8: PG_x příjmů loajálních a nestálých zaměstnanců



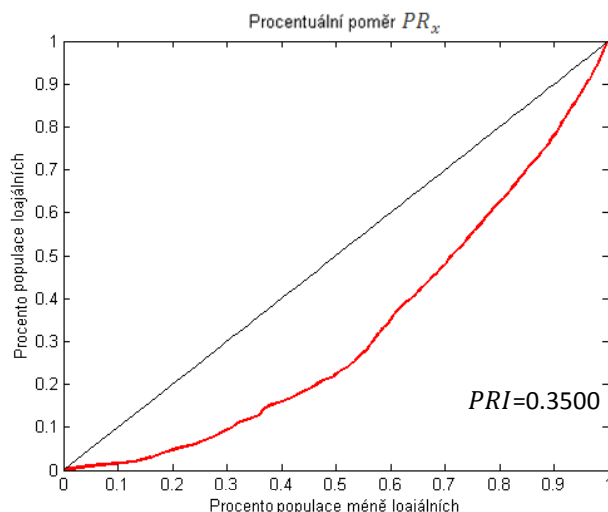
Obrázek 3.9: PG_x příjmů loajálních a nestálých zaměstnanců po nezahrnutí odlehleho pozorování

Obrázek 3.8 a 3.9 ukazuje průběh příjmové mezery PG_x . Předpoklad, že hůře placení jsou neloajální zaměstnanci, byl správný, jedinou výjimkou je stý percentil. Osoba s úplně nejvyšším příjmem ve výběru je zaměstnána méně než 8 let a její příjem je 1.85 krát vyšší než nejlépe placený loajální zaměstnanec. Ukazatele spojené s příjmovou mezerou pro uvedené obrázky jsou uvedeny v Tabulce 3.14. Vyšší hodnota PGI je v případě této nerovnosti způsobena hodnotami poměrů percentilů skupin, které se ale s výjimkou krajních hodnot pohybují mezi hodnotami 0.9 a 0.65, od hodnoty 1 jsou vzdálené příjmové mezery pro všechny percentily. Výsledky tabulky také potvrzují, že nezahrnutí odlehleho pozorování výrazně ovlivnilo odhady s ním spojených ukazatelů.

	PGI	\hat{A}_{10}	\hat{A}_{-10} (popřípadě \hat{A}_{-10}')	směrnice
Při zahrnutí odlehleho pozorování	0.2680	0.8467	0.8106	-0.0672
Po nezahrnutí odlehleho pozorování	0.2680	0.8467	0.7114	-0.1366

Tabulka 3.14: Výsledky PGI příjmů loajálních a nestálých zaměstnanců

Procentuální poměr PR_x , který udává představu o tom, zda někdo, popřípadě jaká část méně loajálních zaměstnanců získává příjem, který odpovídá percentilu loajálních zaměstnanců, je vykreslen na Obrázku 3.10. Hodnota PRI se rovná 0.3500. Největší zlom nastává kolem 55. percentilu méně loajálních zaměstnanců, od kterého dochází k prudkému nárůstu hodnot, naopak pro nižší percentily byl růst pouze pozvolný, například stejně nebo méně než vydělává 10 % loajálních zaměstnanců, přijímá 32 % neloajálních zaměstnanců.



Obrázek 3.10: Procentuální poměr PR_x příjmů loajálních a nestálých zaměstnanců

3.5 Příjmové nerovnosti mezi obory zaměstnání

Dalším faktorem, který může způsobovat nerovnost v příjmech zaměstnanců je obor, ve kterém pracují. Celkem se ve zkoumaném datovém souboru objevuje 62 profesí, které byly sloučeny a vzniklo tak 9 skupin, kterými jsou management, finance a bankovníctví, zdravotnictví, administrativa, výroba a těžba, zákaznický servis, vzdělávání a výzkum, logistika a odbyt, IT specialisté. V Tabulce 3.15 lze nalézt kromě četností jednotlivých skupin a průměrného příjmu také hodnoty indexů nerovnosti v rámci oborů.

Obor	Četnost ve výběru	Průměrný příjem (v ARS)	Giniho index	Theilův index
Management	2 050	601 136	0.3787	0.2625
Finance a bankovníctví	9 874	290 365	0.3047	0.1796
Zdravotnictví	752	421 164	0.3187	0.1851
Výroba a těžba	7 063	331 628	0.3327	0.1955
Vzdělávání a výzkum	1 792	456 478	0.3444	0.1983
Zákaznický servis	1 691	248 499	0.3327	0.2351
Logistika a odbyt	6 209	292 304	0.3527	0.2303
Administrativa	4 037	275 420	0.3093	0.1982
IT	2 256	424 929	0.2930	0.1414

Tabulka 3.15: Hodnoty Giniho a Theilova indexu v rámci oborů zaměstnání

Další míry nerovnosti, které práce využívá, porovnávají nerovnost ve dvou skupinách. Byl vybrán obor management, jehož průměrný příjem je nejvyšší, nerovnosti mezi ním a ostatními obory budou změřeny pomocí Pay Gap Index (*PGI*) a Percentage Ratio Index (*PRI*), výsledky jsou zapsány v Tabulce 3.16.

	<i>PRI</i>	<i>PGI</i>	\hat{A}_{10}	\hat{A}_{-10}	směrnice
Finance a bankovníctví	0.5893	0.4886	0.6880	0.4402	-0.2903
Zdravotnictví	0.2653	0.2424	0.9185	0.6218	-0.2992
Výroba a těžba	0.4768	0.4133	0.7048	0.5007	-0.1676
Vzdělávání a výzkum	0.2020	0.2040	0.8175	0.7104	-0.0937
Zákaznický servis	0.7139	0.5557	0.6529	0.4173	-0.2639
Logistika a odbyt	0.5900	0.4920	0.6492	0.5086	-0.1311
Administrativa	0.6437	0.4968	0.6830	0.4362	-0.3052
IT	0.2161	0.2166	0.8709	0.5427	-0.2877

Tabulka 3.16: Výsledky indexů nerovnosti porovnávající příjmy management a ostatních oborů

Směrnice příjmové mezery je záporná, nerovnost se tedy s rostoucím percentilem prohlubuje, nejvíce je patrný klesající trend u oboru administrativa, u kterého jsou navíc indexy *PGI* a *PRI* druhé nejvyšší. Nejnižších hodnot indexů nerovnosti dosahuje obor vzdělávání a výzkum, jehož směrnice je taktéž nejnižší. Skupinu zaměstnanců, jejichž příjmy se nejvíce odlišují od manažerů, určila tabulka zaměstnance pracující v oboru zákaznický servis, kde jsou například zahrnuti pracovníci call center.

Porovnání nerovností všech oborů na základě hodnot *PRI* nalezneme v Tabulce 3.17, užití tohoto indexu je podmíněno určením, která ze dvou pozorovaných skupin je znevýhodněna a která je naopak oceňována lépe. Řádky tabulky určují Skupinu Y, která je zvýhodněná, a sloupce definují Skupinu X, která je oproti druhé skupině podhodnocená. Byly určeny hodnoty indexu pro všechny kombinace dvou různých oborů. Výsledky, jejichž předpoklad o rozdělení příjmu mezi skupinami, byl správný, jsou kladné, pro lepší názornost jsou červeně podtrženy.

Skupina X Skupina Y	Management	Finance a bankovníctví	Zdravotnictví	Výroba a těžba	Vzdělávání a výzkum	Zákaznický servis	Logistika a odbyt	Administrativa	IT
Management	X	<u>0.5893</u>	<u>0.2653</u>	<u>0.4768</u>	<u>0.2020</u>	<u>0.7139</u>	<u>0.5900</u>	<u>0.2161</u>	<u>0.2161</u>
Finance a bankovníctví	-0.5643	X	-0.4082	-0.1061	-0.4216	<u>0.2884</u>	<u>0.0841</u>	<u>0.0933</u>	-0.4396
Zdravotnictví	-0.2614	<u>0.4116</u>	X	<u>0.2744</u>	-0.0525	<u>0.6026</u>	<u>0.4304</u>	<u>0.4959</u>	-0.0591
Výroba a těžba	-0.4670	<u>0.1068</u>	-0.2712	X	-0.2991	<u>0.3526</u>	<u>0.1765</u>	<u>0.1783</u>	-0.3127
Vzdělávání a výzkum	-0.1959	<u>0.4228</u>	<u>0.0551</u>	<u>0.3004</u>	X	<u>0.5999</u>	<u>0.4444</u>	<u>0.4903</u>	<u>0.0007</u>
Zákaznický servis	-0.7006	-0.2836	-0.6012	-0.3531	-0.5974	X	-0.1588	-0.2302	-0.6094
Logistika a odbyt	-0.5873	-0.0830	-0.4304	-0.1762	-0.4432	<u>0.1578</u>	X	-0.0158	-0.4556
Administrativa	-0.6319	-0.0918	-0.4931	-0.1778	-0.4888	<u>0.2320</u>	<u>0.0166</u>	X	-0.5076
IT	-0.2006	<u>0.4400</u>	<u>0.0616</u>	<u>0.3126</u>	<u>0.0020</u>	<u>0.6117</u>	<u>0.4562</u>	<u>0.5081</u>	X

Tabulka 3.17: Index nerovnosti *PRI* porovnávající obory navzájem

Tabulka 3.17 potvrzuje, že nejlépe ohodnocenou skupinou jsou manažeři, jelikož v porovnání se všemi ostatními obory se jedná o skupinu s vyššími příjmy. Naopak nejpodceňovanější skupinou je obor Zákaznický servis. Mezi těmito obory panuje největší nerovnost, *PRÍ* je v tomto případě roven 71.39 %. Nejnížší nerovnost se objevuje mezi obory IT a Vzdělávání a výzkum.

3.6 Příjmové nerovnosti mezi pohlavími

Ve společnosti často diskutované téma, rozdíl v odměňování mužů a žen, bylo inspirací pro zkoumání nerovností mezi pohlavími. Na základě definovaných indexů změříme nerovnost příjmů žen a příjmů mužů. Prvního náznaku nerovného rozdělení příjmů mezi pohlaví si můžeme všimnout již v Tabulce 1.2 v úvodu této práce, kde se značně liší základní charakteristiky.

Pro každou skupinu zvlášť byly určeny hodnoty Giniho indexu a Theilova indexu. Výsledky obou skupin jsou uvedeny v následující tabulce, která pro lepší představu navíc obsahuje i hodnoty indexů pro celý výběr.

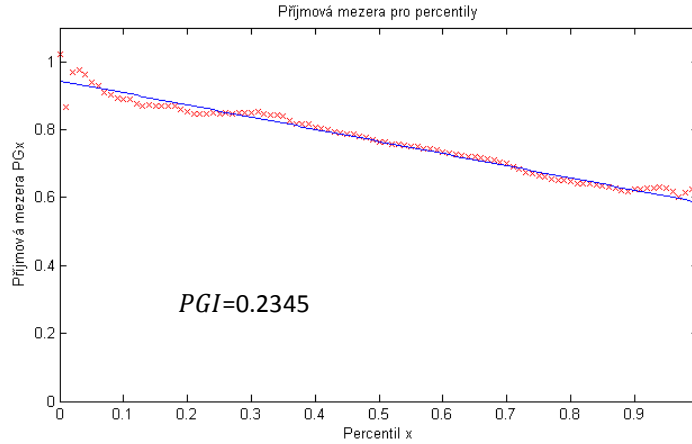
	Giniho index	Theilův index
Muži	<u>0.3616</u>	<u>0.2366</u>
Ženy	0.2877	0.1559

Tabulka 3.18: Theilův a Giniho index měřící nerovnost v rámci pohlaví

Z výsledků indexů uvedených v Tabulce 3.16 lze dojít k závěru, že větší nerovnost se objevuje ve skupině mužů. Nerovnost v rámci této skupiny je dokonce větší než nerovnost v zemi celkem, výsledky skupiny žen jsou pro oba indexy téměř o 10 procentních bodů nižší.

Nejnižších příjmů ve výběru dosahují, jak muži, tak ženy, jsou tedy zastoupena obě pohlaví. Osoby s nejnižšími příjmy ve výběru jsou zaměstnány ve výrobě, jako vzdělávací asistenti nebo jako administrativní pracovníci. Ženy navíc pracují jako služebné nebo uklízečky. Naopak mezi 10 nejbohatšími se ve výběru neobjevuje žena. Jako ředitelé, vedoucí oddělení a manažeři jsou zaměstnáváni častěji muži. Muži tedy pracují na hůře placených pozicích, ale i jako vedoucí pracovníci, což způsobuje větší nerovnost v rozdělení příjmů mužů.

Předchozí indexy porovnávají pouze nerovnost v rámci skupiny, o nerovnosti rozdělení příjmů mezi pohlaví prozradí více nově zavedené indexy měřící nerovnost. Na Obrázku 3.11 je zobrazen průběh příjmové mezery pro percentily.



Obrázek 3.11: Příjmová mezera PG_x příjmů mužů a žen

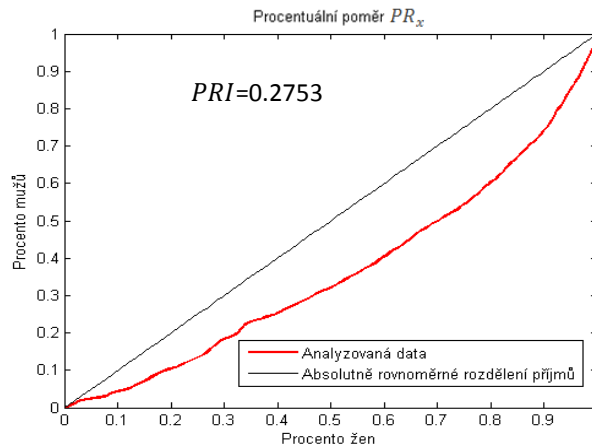
Z obrázku je zřetelné, že příjmová mezera se s rostoucím percentilem prohlubuje. Podíl percentilů žen a mužů se snižuje, zvětšuje se tedy rozdíl mezi mužským a ženským příjmem. Výsledky sledovaných ukazatelů jsou uvedeny v Tabulce 3.19.

PGI	\hat{A}_{10}	\hat{A}_{-10}	směrnice
0.2345	0.9367	0.6052	-0.3584

Tabulka 3.19: PGI srovnávající příjmy mužů a žen

\hat{A}_{10} pro data vychází 0.9367, tento ukazatel říká, že nejchudší ženy (prvních 10 percentilů) mají v průměru 93.67 % příjmu nejchudších mužů. $\hat{A}_{-10} = 0.6052$ znamená, že nejbohatší ženy (91. až 100. percentil) vydělávají v průměru jen 60.52 % toho, co nejbohatší muži. Směrnice přímky, která je proložena body PG_x , je -0.3355. Záporná hodnota potvrzuje klesající tendenci PG_x .

Procentuální poměr PR_x pro muže a ženy je zaznamenán na následujícím obrázku. PRI pro data nabývá hodnoty 0.2753. Jak je na Obrázku 3.12 patrné, křivka, na které body grafu leží, nejstrměji roste pro procento žen 0.9 až 1. Přibližně 90 % žen vydělává stejně nebo méně než je hodnota příjmu, který pobírá 70 % mužů.



Obrázek 3.12: Procentuální poměr PR_x příjmů mužů a žen

3.6.1 Nerovnost pohlaví v rámci regionů

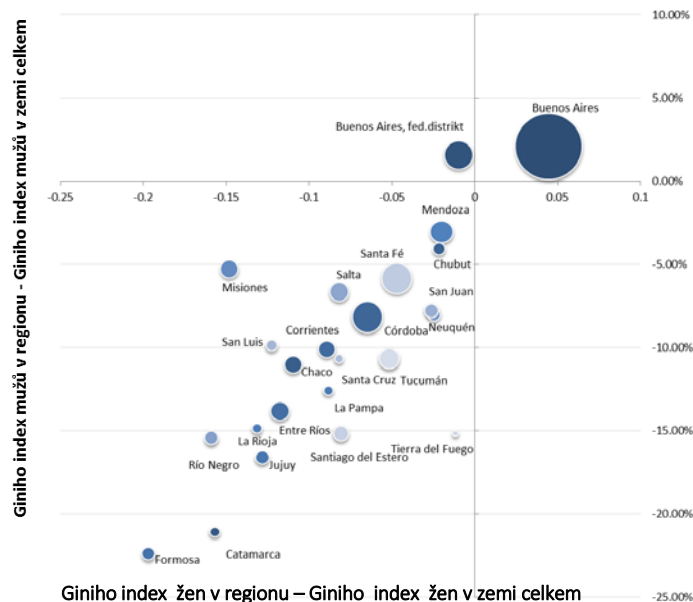
Nerovnost mezi pohlavími může být ovlivněna i regionem, ve kterém jsou osoby zaměstnávány. Ve velkých městech je předpokládána větší různorodost pracovních míst, tedy větší nerovnoměrnost v rozdělení příjmů. Výsledky základních indexů jako je Giniho index a Theilův index jsou uvedeny v následující tabulce. Hodnoty, které nabývají vyšších hodnot, než jsou výsledky pro zemi jako celek uvedené v Kapitole 3.1, jsou červeně podtrženy.

Hodnoty indexů se v regionech liší až v rámci desetin. Ve většině případech je hodnota Giniho i Theilova koeficientu vyšší pro muže, výjimkou je region Tierra del Fuego. Jak je vidět v Tabulce 1.3, ve které je mimo jiné zahrnuta informace o četnosti mužů a žen v regionu, z tohoto regionu pochází jen malé množství zástupců.

Provincie	Giniho index		Theilův index	
	Muži	Ženy	Muži	Ženy
Buenos Aires	<u>0.3825</u>	<u>0.3326</u>	<u>0.2785</u>	<u>0.2018</u>
Buenos Aires, federální distrikt	<u>0.3774</u>	0.2780	<u>0.2665</u>	0.1488
Catamarca	0.1509	0.1307	0.0418	0.0419
Chaco	0.2514	0.1781	0.1129	0.0499
Chubut	0.3208	0.2662	0.1848	0.1193
Córdoba	0.2799	0.2229	0.1356	0.0923
Corrientes	0.2607	0.1984	0.1229	0.0680
Entre Ríos	0.2235	0.1700	0.0904	0.0500
Formosa	0.1376	0.0910	0.0336	0.0163
Jujuy	0.1954	0.1596	0.0628	0.0457
La Pampa	0.2358	0.1993	0.1018	0.0672
La Rioja	0.2127	0.1562	0.0758	0.0543
Mendoza	0.3311	0.2679	0.1871	0.1246
Misiones	0.3088	0.1396	0.1740	0.0323
Neuquén	0.2818	0.2631	0.1313	0.1131
Río Negro	0.2072	0.1288	0.0775	0.0276
Salta	0.2952	0.2059	0.1909	0.0756
San Juan	0.2840	0.2617	0.1522	0.1345
San Luis	0.2627	0.1652	0.1185	0.0419
Santa Cruz	0.2550	0.2058	0.1114	0.0759
Santa Fé	0.3034	0.2405	0.1645	0.1102
Santiago del Estero	0.2100	0.2069	0.0952	0.0714
Tierra del Fuego	0.2100	0.2761	0.0816	0.1344
Tucumán	0.2547	0.2360	0.1259	0.1051

Tabulka 3.20: Výsledky Giniho a Theilova indexu pro různé regiony

Diference regionálních hodnot Giniho indexu mužů a žen od obecných hodnot je znázorněna pomocí bublinového grafu na Obrázku 3.13. Velikost bublin je dána počtem obyvatel regionu.



Obrázek 3.13: Vizualizace rozdílů Giniho indexů mužů a žen v regionech oproti zemi jako celku

Ověříme podobně jako v Kapitole 3.1, zda se nerovnosti v hlavním městě a mimo něj liší. Výsledky Giniho a Theilova indexu uvádí Tabulka 3.21, ze které je patrné, že příjmy mužů vykazují větší nerovnost v hlavním městě, naopak příjmy žen jsou více nerovnoměrné mimo hlavní město.

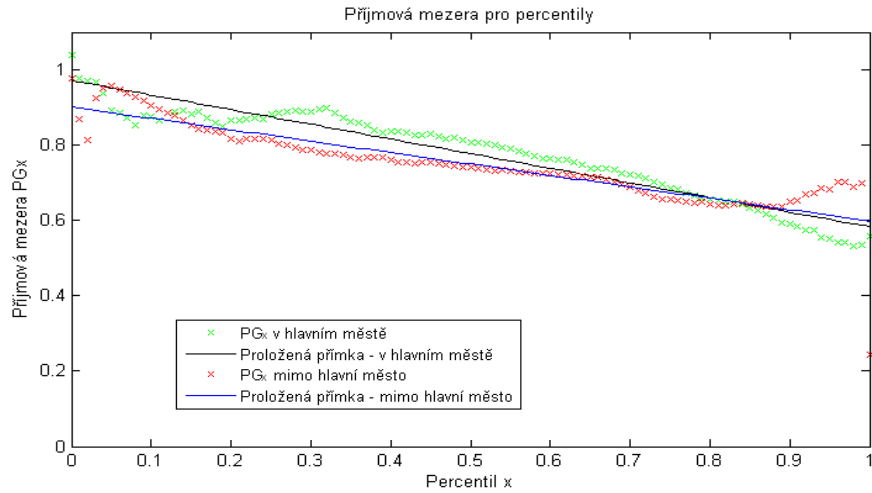
	V hlavním městě		Mimo hlavní město	
	Giniho index	Theilův index	Giniho index	Theilův index
Muži	<u>0.3760</u>	<u>0.2650</u>	0.3503	0.2167
Ženy	0.2774	0.1474	0.3034	0.1697

Tabulka 3.21: Výsledky Theilova a Giniho indexu v hlavním městě a mimo něj

Dále kvantifikované míry nerovnosti již porovnávají muže a ženy v rámci jednoho indexu. Na následujícím obrázku je vykreslen průběh PG_x pro muže a ženy v hlavním městě a pro muže a ženy mimo hlavní město. V obou případech jsou skupinou s vyššími příjmy muži. Na obrázku pozorujeme, že PG_x v hlavním městě je pro větší počet percentilů vyšší než PG_x mimo hlavní město. Výjimkou je několik hodnot kolem $x=0.1$ a hodnoty pro x větší než 0.8. Hodnoty PG_x pro data hlavního města se kromě krajních bodů výrazně oddalují od proložené přímky. Výsledky ukazatelů, které souvisí s PGI jsou uvedeny v Tabulce 3.22.

	PGI	\hat{A}_{10}	\hat{A}_{-10}	směrnice
V hlavním městě	0.2228	0.9272	0.5548	-0.3724
Mimo hlavní město	0.2489	0.9222	0.6399	-0.3889

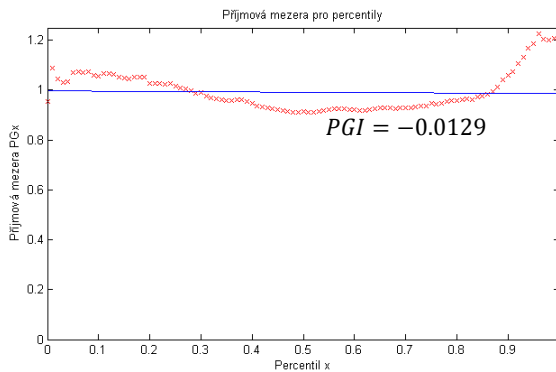
Tabulka 3.22: PGI pro příjmy mužů a žen



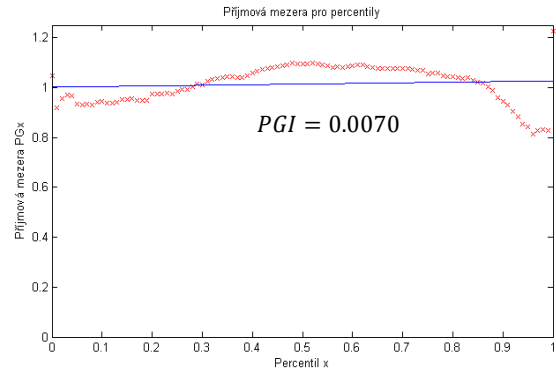
Obrázek 3.14: PG_x pro muže a ženy v hlavním městě a mimo něj

Na základě výsledků z předchozí tabulky lze tedy usuzovat, že nerovnost mezi pohlavími je obecně v hlavním městě nižší. V hlavním městě je ale větší rozdíl mezi muži a ženami s vyššími příjmy, muži pracující na vysokých pozicích jsou ve městě lépe placeni než ženy.

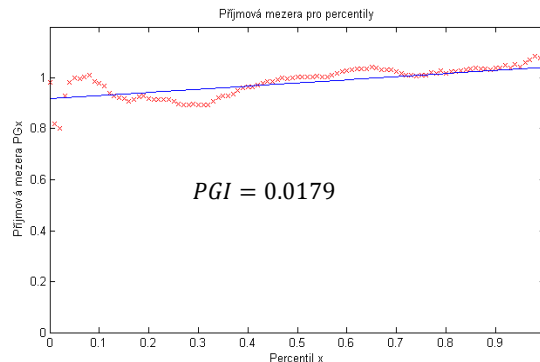
Pomocí PGI lze také analyzovat ženy i muže v hlavním městě a mimo něj. Bylo předpokládáno, že v hlavním městě jsou soustředěny vyšší příjmy. U žen se tento předpoklad potvrdil, PGI dosahuje kladné hodnoty 0.0179, průběh PG_x pro ženy zaznamenává Obrázek 3.17. Pro muže toto rozložení příjmů ale neplatí, PGI vychází -0.0129, z důvodu správného výpočtu zmíněného ukazatele je tedy nutné skupiny vyměnit a uvažovat, že vyšší příjmy jsou soustředěny mimo hlavní město. Obrázek 3.15 ukazuje průběh PG_x při užití špatné domněnky, správný předpoklad rozložení příjmů vede k průběhu funkce, který zaznamenává Obrázek 3.16.



Obrázek 3.15: PG_x pro muže v hlavním městě a mimo něj



Obrázek 3.16: PG_x pro muže mimo hlavní město a v hlavním městě



Obrázek 3.17: : PG_x pro ženy v hlavním městě a mimo něj

Z Obrázků 3.16 a 3.17 můžeme vyčíst, že důvodem, proč je důležité očekávat, že u mužů jsou vyšší příjmy mimo hlavní město. Rozhodující jsou hodnoty percentilů $x=0.3$ až $x=0.85$, které jsou vyšší mimo hlavní město. Naopak pro $x < 0.3$ a $x > 0.85$ dosahují percentily příjmů vyšších čísel v hlavním městě, například 98. percentil mužských příjmů v Gran Buenos Aires je 1.2 krát vyšší než v jeho vnější oblasti. U žen dosahuje od $x=0.7$ k pozvolnému růstu PG_x , hodnoty převýšily i hodnotu 1, což znamená, že skupinou s vyššími příjmy se pro tyto percentily stala skupina žen pracujících mimo hlavní město.

Uvedené jevy souvisí s Obrázkem 3.14, kde se s rostoucím percentilem začínaly objevovat větší rozdíly mezi muži a ženami právě v hlavním městě. Výsledky ukazatelů jsou uvedeny v Tabulce 3.23.

	PGI	\hat{A}_{10}	\hat{A}_{-10}	směrnice
Ženy	0.0179	0.9511	1.0056	0.1240
Muži	0.0070	1.0504	1.1326	-0.0107

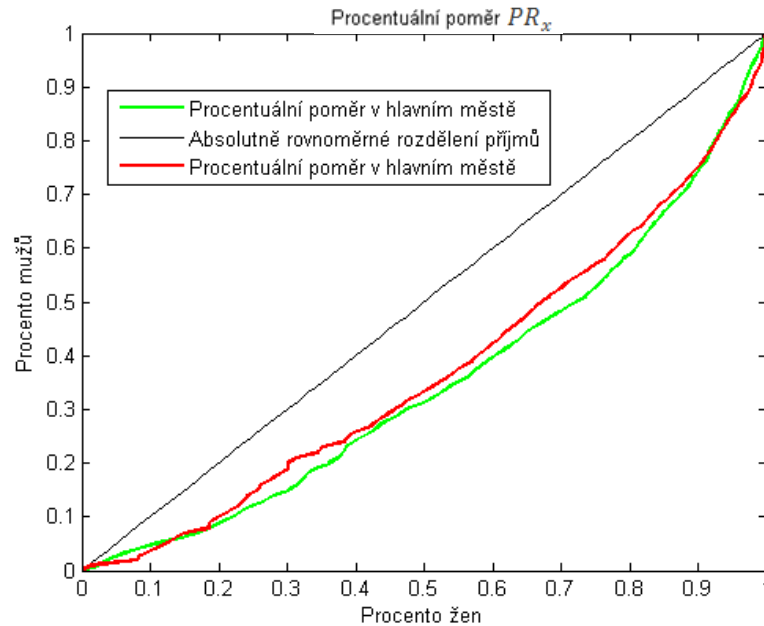
Tabulka 3.23: PGI a s ním související ukazatele porovnávající nerovnost příjmů v hlavním městě a mimo něj

Na základě uvedených obrázků a výsledků lze říci, že nerovnosti v rámci jednoho pohlaví v Gran Buenos Aires a mimo něj jsou podstatně nižší než rozdíly mezi pohlavími. PG_x se pohybuje kolem hodnoty 1, od které se v obou případech nejvíce odchyľují body pro $x > 0.95$. V obou případech jsou tedy osoby s nejvyššími příjmy v dané skupině lépe oceňované ve skupině, která je obecně považována za chudší.

Poslední index, který bude k měření nerovnosti v hlavním městě a mimo něj použit je index PRI . Výsledky uvádí Tabulka 3.24, kde se opět ukazuje, že větší nerovnost mezi pohlavími se vyskytuje mimo hlavní město a jeho aglomeraci. Průběh PR_x pro oba případy vykresluje Obrázek 3.18.

	PRI
V hlavním městě	0.2593
Mimo hlavní město	0.2936

Tabulka 3.24: PRI pro měření nerovnosti pohlaví v hlavním městě a mimo něj



Obrázek 3.18: Průběh procentuálního poměru pro pohlaví v hlavním městě a mimo něj

Tabulka 3.25 obsahuje výsledky tohoto indexu, pokud srovnává jedince stejného pohlaví z Gran Buenos Aires a regionů mimo něj. Větší rozdíl se objevuje u žen. Podobně jako u předchozího indexu, jsou ženy lépe placené v hlavním městě, naopak muži jsou lépe oceňováni mimo hlavní město.

	<i>PRI</i>
Ženy	0.0357
Muži	0.0237

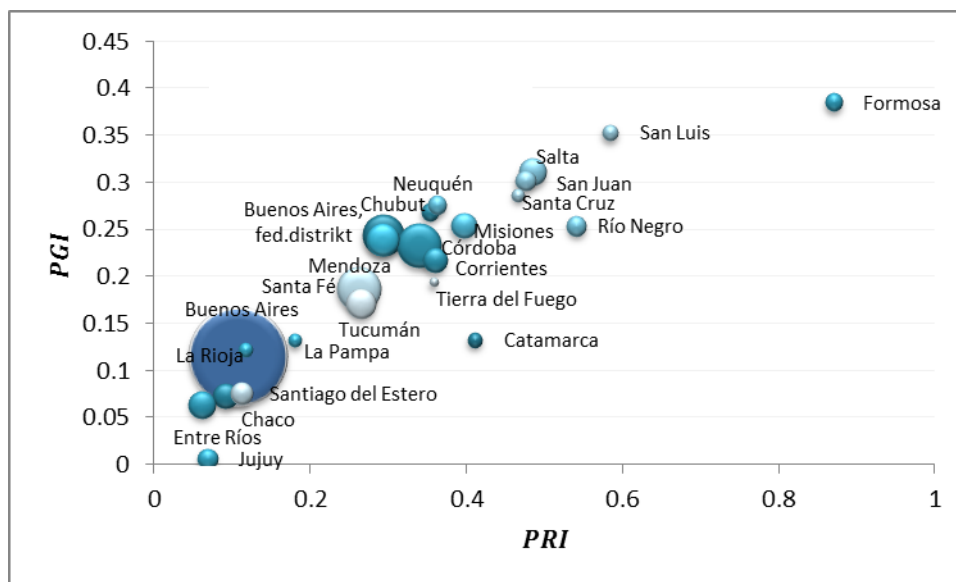
Tabulka 3.25: *PRI* pro příjmy v hlavním městě a mimo něj

Stejně tak jako Giniho a Theilův index byl určen pro muže a ženy nejen v hlavním městě, ale také pro každý region zvlášť, budou určeny hodnoty i zbylých dvou indexů, aby nerovnost mezi muži a ženami v rámci regionů byla co nejlépe popsána. Výsledky *PGI* a *PRI* společně s dalšími ukazateli, které souvisí s tímto indexem, shrnuje následující Tabulka 3.26. Ve všech regionech bylo uvažováno, že jsou muži skupinou s vyššími příjmy. Hodnoty ukazatelů vychází kladně, v žádném regionu se nemusel tento předpoklad měnit.

Provincie	<i>PRI</i>	<i>PGI</i>	\hat{A}_{10}	\hat{A}_{-10}	směrnice
Buenos Aires	0.1071	0.1152	0.9447	0.7251	-0.2404
Buenos Aires, federální distrikt	0.2926	0.2438	0.9206	0.5406	-0.3870
Catamarca	0.4103	0.1328	1.0505	0.8846	-0.1743
Chaco	0.0907	0.0735	1.0018	0.6257	-0.2389
Chubut	0.3528	0.2686	0.8900	0.5756	-0.2582
Córdoba	0.3391	0.2336	0.8148	0.6271	-0.2477
Corrientes	0.3586	0.2174	0.9571	0.6028	-0.2441
Entre Ríos	0.0607	0.0639	0.9700	0.7276	-0.2238
Formosa	0.8700	0.3853	0.7325	0.4485	-0.1553
Jujuy	0.0674	0.0065	1.2181	0.9278	-0.3125
La Pampa	0.1792	0.1324	0.6347	0.6756	-0.0534
La Rioja	0.1160	0.1228	0.8922	0.8431	-0.2989
Mendoza	0.2922	0.2401	0.8996	0.6022	-0.2779
Misiones	0.3961	0.2545	0.9114	0.3668	-0.5557
Neuquén	0.3616	0.2762	0.7199	0.6526	-0.0832
Río Negro	0.5405	0.2538	0.8322	0.5049	-0.2752
Salta	0.4843	0.3110	0.7902	0.5570	-0.2689
San Juan	0.4745	0.3020	0.8013	0.6373	-0.1125
San Luis	0.5831	0.3534	0.9964	0.4292	-0.3916
Santa Cruz	0.4658	0.2869	0.8309	0.6468	-0.2697
Santa Fé	0.2611	0.1873	0.9016	0.6876	-0.2675
Santiago del Estero	0.1114	0.0762	0.9381	0.7909	-0.1473
Tierra del Fuego	0.3570	0.1940	0.7457	1.0101	0.2997
Tucumán	0.2642	0.1717	0.8280	0.7547	-0.0599

Tabulka 3.26: Hodnoty ukazatelů *PGI* a *PRI* pro muže a ženy v regionech

Většina směrnic je záporné číslo, výjimkou je region Tierra del Fuego, jehož směrnice je kladná, což znamená, že rozdíl mezi nejbohatšími ženami a muži je nižší než mezi nejchudšími. Více prozradí další ukazatele. Pro region Tierra del Fuego vychází \hat{A}_{-10} větší než 1, to by znamenalo, že bohaté ženy jsou v průměru dokonce bohatší než nejbohatší muži. Vizualizace výsledků indexů nerovnosti na Obrázku 3.19 nabízí přehlednější shrnutí, navíc bublinový graf na zmíněném obrázku znázorňuje závislost obou indexů, velikost bublin odpovídá počtu obyvatel v regionu.



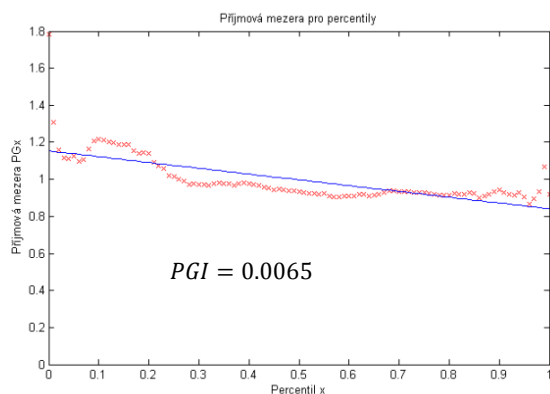
Obrázek 3.19: Závislost hodnot *PGI* na výsledcích *PRI* pro region, velikost bubliny značí počet obyvatel regionu

Na rozdíl od Giniho a Theilova indexu se neobjevují nejvyšší hodnoty indexů v regionech, na kterých se rozkládá hlavní město. Indexy *PRI* a *PGI* vychází nejvyšší pro region Formosa, který v Tabulce 3.20 nijak význačně nevybočoval. Stal se zajímavý jev, kdy Giniho index a Theilův index, které měří nerovnost v rámci jedné skupiny, neodhalily významnější nerovnost, ale indexy, které pracují se dvěma skupinami, nerovnost objevily.

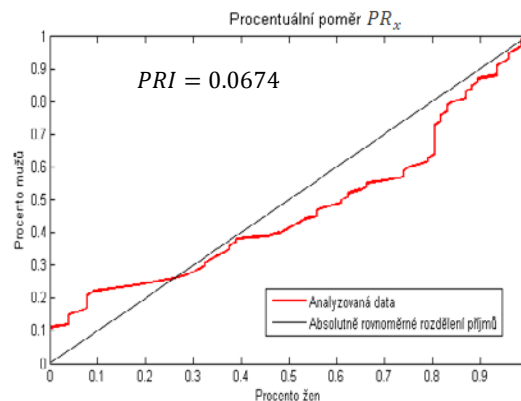
Pokud seřadíme zástupce tohoto regionu podle příjmu, osoby s nejnižšími příjmy jsou častěji ženy, naopak vyšší příjmy pobírají výhradně muži, to způsobilo, že výsledky indexů v rámci pohlaví neukazují výrazné nerovnosti, hodnoty Giniho indexu a Theilova indexu nejsou vysoké, ale nerovnost mezi skupinami navzájem posouzená mírami *PGI* a *PRI* je značná. Důvodem tohoto rozložení v regionu může být použitý výběr, který dle Tabulky 1.3 obsahuje pouze 21 zástupců. Podobný jev se ale projevuje např. u regionu Rio Negro nebo Salta, do kterých patří vyšší počet osob.

Opačný jev lze pozorovat v regionu Jujuy, do kterého lze z výběru zařadit 77 žen a 403 mužů. Zatímco Giniho a Theilův index nebyly v porovnání s ostatními regiony nízké, *PGI* a *PRI* dosahovali nejnižších hodnot. Na následujících obrázcích je zaznamenán průběh PG_x a PR_x , ze kterých indexy vycházejí. Z obrázků lze usuzovat, že ženy s nižšími příjmy v rámci skupiny žen dosahují vyšších platů než stejně postavení muži. Například 10. Percentil příjmů žen je 1.2 krát vyšší než příjem mužů. S rostoucím percentilem má PG_x spíše klesající tendenci, nikdy ale neklesne pod hodnotu 0.85. Ženy s vysokým příjmem v porovnání s muži získávají téměř totožné příjmy.

Graf průběhu PR_x potvrzuje dobré postavení žen, žádná nemá horší příjem, než který pobírá 12 % mužů. Až do 25. percentilu jsou lépe ohodnocované ženy, později dojde ke změně a muži se stávají skupinou, která ve společnosti získává vyšší příjmy. Nerovnost mezi skupinami se tak mění, nelze říci, že jedna z nich má obecně vyšší příjmy, proto indexy nabývají nízkých hodnot.



Obrázek 3.20: PG_x pro muže a ženy v regionu Jujuy



Obrázek 3.21: PR_x pro muže a ženy v regionu Jujuy

Pro muže a ženy v tomto regionu použijeme Kolmogorovův-Smirnovův dvouvýběrový test. Na hladině významnosti 5 % testujeme, zda se výběry řídí stejným rozdělením pravděpodobnosti oproti alternativě, že se neřídí stejným rozdělením. Kolmogorovův-Smirnovův test umí provést a vyhodnotit software Matlab. Po zadání příkazu $kstest2(x1,x2)$, kde $x1$ a $x2$ jsou porovnávané výběry, program vrací rozhodnutí, zda je nulová hypotéza na hladině významnosti 5 % přijata, nebo zamítnuta. Pro výběr mužů a žen z regionu Jujuy byla nulová hypotéza na určené hranici významnosti zamítnuta, p -hodnota testu je 0.0354. Přestože indexy dosahují nízkých hodnot, výběr mužů a žen nepochází ze stejného rozdělení pravděpodobnosti.

3.6.2 Jednoduchý model logistické regrese

V kapitole vyjádříme pomocí vhodného modelu závislost pohlaví na různých faktorech. Jinými slovy se pokusíme zjistit, zda z dostupných údajů jako je například příjem osoby a dalších osobních informací lze určit, jestli je daná osoba muž nebo žena. Budeme se snažit ověřit, že pohlaví zaměstnance lze poznat na základě jiných informací.

Případy, kdy vysvětlovaná proměnná Y_i je dichotomická (tj. veličina nabývá pouze dvou hodnot) nebo nabývá hodnot z intervalu $[0,1]$, modeluje logistická regrese. V našem případě budeme uvažovat možnosti $Y_i = 0$ (muž) a $Y_i = 1$ (žena).

V logistickém regresním modelu budeme nejprve předpokládat lineární závislost logitu $\eta_i = \ln \frac{P(Y_i=1)}{P(Y_i=0)}$ pouze na vysvětlované proměnné $příjem_i$. Model pak bude mít tvar

$$\eta_i = \beta_0 + \beta_{příjem} \cdot příjem_i, \quad i = 1, \dots, N, \quad (3.1)$$

kde β_0 a β_1 jsou odhadované regresní koeficienty.

Logistickou regresí se více zabývá např. [15], kde také lze najít odvození odhadu parametrů logistické regrese. K odhadu regresních koeficientů v modelu byl použit software Matlab, kde pomocí funkcí `glmval` a `glmfit` byl určen nejen odhad koeficientů, ale také p -hodnota t -testu pro logistickou regresí, který testuje významnost odhadů. Výsledky programu jsou zapsány do Tabulky 3.27, ve které se objevuje také Odds Ratio (Poměr šancí). [14]

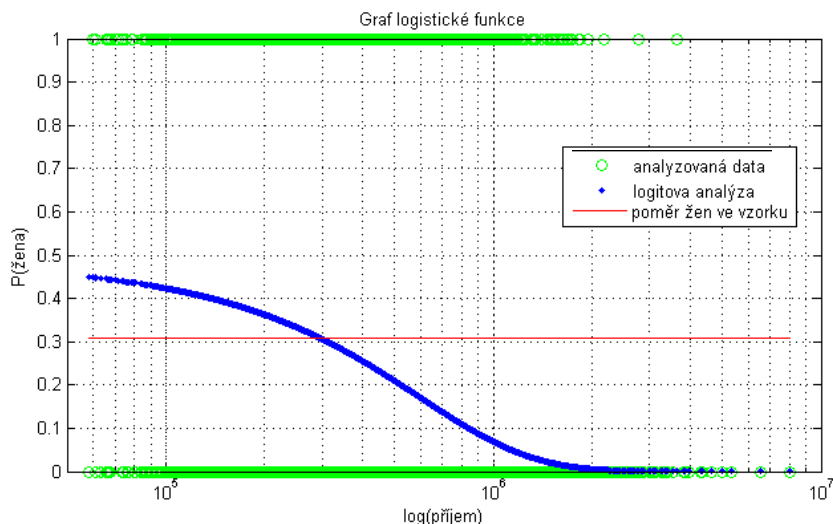
Šancí se při zavedení logistické regrese považuje podíl pravděpodobnosti, že jev nastal, a pravděpodobnosti, že jev nenastal. Po provedení odhadu regresních koeficientů, lze určit, jak se změní

šance, pokud se jedna ze spojitých vysvětlujících proměnných zvýší o jednotku a ostatní hodnoty zůstanou nezměněny. Míra této změny (Odds Ratio) je rovna $e^{\widehat{\beta}_j}$, kde $\widehat{\beta}_j$ je odhad regresního koeficientu změněného regresoru. Pro uvedená data tedy platí, že pokud je příjem zvýšen o 1, poměr $\frac{P(Y_i=\text{žena})}{P(Y_i=\text{muž})}$ se změní 0.99999 krát. Odds Ratio je nižší než 1, šance bude nižší. S ohledem na vysoké hodnoty částek ve zkoumaném vzorku bude vhodnější uvést příklad s větší změnou vysvětlující proměnné. Pokud se příjem zvýší o 10 000 argentinských pesos, Odds Ratio bude $e^{\widehat{\beta}_{\text{příjem}} \cdot 10\,000} = 0.9748$, šance se tedy sníží.

Regresní koeficienty	Odhady regresních koeficientů	Odds Ratio	p-hodnota
β_0	$-5.3524 \cdot 10^{-2}$	-	$2.1630 \cdot 10^{-2}$
$\beta_{\text{příjem}}$	$-2.5505 \cdot 10^{-6}$	0.99999	$6.9409 \cdot 10^{-260}$

Tabulka 3.27: Odhady regresních koeficientů, Odds Ratio a p-hodnoty t-testu

Vykreslení závislosti a skutečných hodnot je na Obrázku 3.22, kde hodnota 1 znamená, že osoba je žena, 0 jsou značení muži. V grafu je červeně znázorněn poměr žen ve vzorku, který se rovná $\frac{10\,964}{35\,724} \cong 0.3069$.



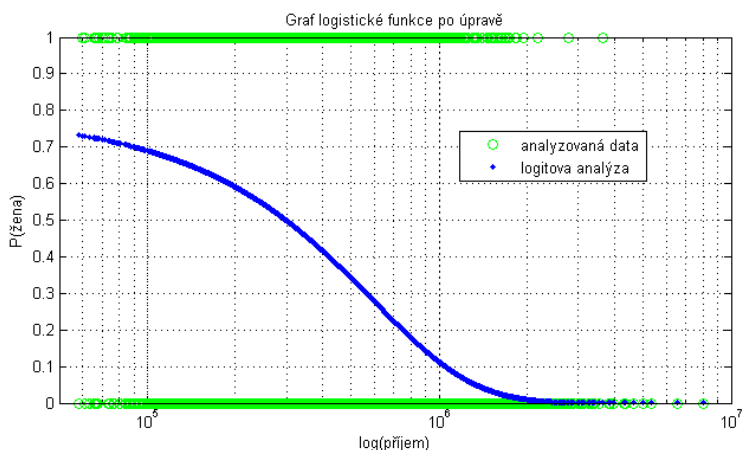
Obrázek 3.22: Logistická funkce

Lze předpokládat, že podíl žen v zemi je 0.5, v analyzovaném datovém souboru však je jejich poměr nižší, což může být způsobeno faktem, že se jedná o údaje o příjmech zaměstnaných osob. V Argentině je zaměstnáváno pouze 41.3 % žen [13], zaměstnanost mužů je vyšší. Pravděpodobnost, že se jedná o ženu, je podhodnocená. V reálném životě předpoklad, že je žen výrazně méně neplatí, proto byl výsledek logistické regrese upraven, aby zohlednil fakt, že ženy tvoří polovinu populace. Pravděpodobnost, že $Y_i = \text{žena}$ byla znormována dle vzorce¹

$$P(Y_i = \text{žena})_{\text{po úpravě}} = \frac{0.5}{\text{podíl žen ve vzorku}} \cdot P(Y_i = \text{žena}) = \frac{0.5}{\frac{10\,964}{35\,724}} \cdot P(Y_i = \text{žena}). \quad (3.2)$$

¹ Normalizaci lze použít pouze po pečlivém zvážení. Teoreticky by se mohlo stát, že by pravděpodobnost po úpravách přesáhla hodnotu 1.

Závislost pravděpodobnosti ženy na příjmu po transformaci (3.2) uvádí Obrázek 3.23.



Obrázek 3.23: Logistická funkce po úpravě

3.6.3 Model logistické regrese se zahrnutím dalších proměnných

Dalšími faktory, které určují strukturu zaměstnanosti mužů a žen, mohou být kromě výše příjmu také věk, doba zaměstnání u jedné společnosti, region a profese. Po začlenění těchto proměnných budeme uvažovat následující model

$$\eta_i = \beta_0 + \beta_{příjem} \cdot příjem_i + \beta_{věk} \cdot věk_i + \beta_{doba} \cdot doba_i + \beta_{region} \cdot region_i + \sum_{o=1}^9 \beta_{obor_o} \cdot obor_{o_i}, \quad (3.3)$$

kde i určuje i -tou osobu ve výběru a nabývá hodnot 1 až N (rozsah výběru). Proměnnou příjem lze považovat za spojitou veličinu, proměnné věk a doba jsou multinomické. Z hlediska velkého množství regionů v Argentině a nízké četnosti některých provincií ve výběru, jsou rozlišení obyvatelé hlavního města a mimo něj, proměnná region je tedy binární a nabývá hodnoty 1, pokud i -tý zaměstnanec pochází z hlavního města. Každý z oborů, které byly použity v Kapitole 3.5, lze považovat taktéž za binární. Obor i -tého zaměstnance $obor_{o_i}$ se rovná 1, vykonává-li i -tá osoba $obor_o$, ostatní obory jsou pro i -tou osobu rovny 0. Je zvoleno následující číslování: management $o=1$, finance a bankovníctví $o=2$, zdravotnictví $o=3$, těžební průmysl a výroba $o=4$, vzdělávání a výzkum $o=5$, zákaznický servis $o=6$, logistika a obchod $o=7$, administrativa $o=8$, IT specialisté $o=9$.

Hodnoty oboru IT specialisté byly parametrizovány jako

$$\beta_{obor_9} = 0. \quad (3.4)$$

Odhady zbylých regresních koeficientů a příslušná p -hodnota jsou uvedeny v následující tabulce.

Koeficient	Význam	Odhadovaná hodnota	Odds Ratio	p-hodnota
β_0	Absolutní člen	$-4.4576 \cdot 10^{-1}$	-	$6.7946 \cdot 10^{-8}$
$\beta_{příjem}$	Příjem	$-2.1290 \cdot 10^{-6}$	0.99999	$7.9486 \cdot 10^{-146}$
$\beta_{věk}$	Věk	$-1.0726 \cdot 10^{-2}$	0.98933	$1.7397 \cdot 10^{-8}$
β_{doba}	Doba zaměstnání	$2.6765 \cdot 10^{-3}$	1.0027	$2.0508 \cdot 10^1$
β_{region}	V hlavním městě	$4.6592 \cdot 10^{-1}$	1.5935	$4.8607 \cdot 10^{-71}$
β_{obor_1}	Management	$2.8444 \cdot 10^{-1}$	1.3290	$3.7594 \cdot 10^{-4}$
β_{obor_2}	Finance a bankovníctví	$9.606 \cdot 10^{-1}$	2.6133	$1.6521 \cdot 10^{-59}$
β_{obor_3}	Zdravotnictví	$3.9838 \cdot 10^{-1}$	1.4894	$1.0104 \cdot 10^{-4}$
β_{obor_4}	Výroba a těžba	-1.7270	0.17782	$5.9953 \cdot 10^{-104}$
β_{obor_5}	Vzdělávání a výzkum	$8.5051 \cdot 10^{-1}$	2.3408	$1.5275 \cdot 10^{-29}$
β_{obor_6}	Zákaznický servis	1.4299	4.1785	$3.5691 \cdot 10^{-80}$
β_{obor_7}	Logistika a odbyt	$-3.299 \cdot 10^{-2}$	0.96755	$6.0176 \cdot 10^{-1}$
β_{obor_8}	Administrativa	1.1258	3.0827	$3.8785 \cdot 10^{-69}$

Tabulka 3.28: Odhady regresních koeficientů, odds ratio a p-hodnota t-testu

Naším cílem je vybrat model, který nezahrnuje příliš mnoho proměnných, ale zároveň co nejlépe vystihuje závislost.

Logistický regresní model bude označen M , odhady jeho parametrů jako b . Testování podmodelu \tilde{M} , který vznikl z modelu M např. vynecháním některé vysvětlující proměnné a jehož koeficienty označíme jako \tilde{b} , může být provedeno pomocí testu poměru věrohodnosti (Wilksův test). Příléhavost modelu b lze posuzovat podle deviance D . [15]

Pro zavedení deviance je nejprve nutné uvažovat nejbohatší možný model, kde všechny nezávislé proměnné jsou významné a přispívají k vysvětlení hodnoty závislé proměnné. Tento model se nazývá saturevaný model. Deviance D je pak definována

$$D(b) = 2(l(b^0) - l(b)), \quad (3.5)$$

kde $l(b^0)$ je maximální hodnota logaritmické věrohodnostní funkce v saturevaném modelu. Pokud je předpokládáno, že všechny hodnoty vysvětlující proměnné jsou v tomto modelu různé, pak $l(b^0) = 0$ a devianci lze vyjádřit jako

$$D(b) = -2l(b) = -2 \sum_{i=1}^n (y_i \ln \hat{p}_i + (1 - y_i) \ln(1 - \hat{p}_i)). \quad (3.6)$$

Pokud bude uvažován základní model s odhady parametrů, které nesou označení b , a zároveň bude určen podmodel vzniklý z původního modelu s odhady parametrů b^1 , pak lze testovou statistiku vycházející z testu poměrem věrohodnosti vyjádřit jako

$$D(b) - D(b^1), \quad (3.7)$$

kteřá při platnosti podmodelu má asymptoticky χ_q^2 rozdělení, kde q je počet stupňů volnosti, který je roven rozdílu počtu vysvětlujících proměnných v porovnávaných modelech.

Podmodel může být vybrán dopřednou nebo zpětnou metodou, popřípadě kombinací zpětné a dopředné metody. Při dopředné metodě se začíná s nejmenším možným počtem regresorů a jsou přidávány regresory, které co nejpříznivěji ovlivní model, naopak zpětná metoda nejprve zahrne všechny dostupné regresory a na základě hodnot vybraného kritéria jsou z modelu odstraňovány nepřínosné proměnné.

Nejprve budeme uvažovat model v nejbohatším tvaru (3.3). Deviance tohoto modelu je 36 625. Podmodely vznikly vynecháním jedné proměnné. Následující tabulka udává přehled podmodelů, jejich deviance a také p -hodnotu testovací statistiky. Z tabulky lze vyčíst, že se deviance nejvíce změnila vynecháním proměnné obor, naopak k nejnižší změně došlo po vynechání proměnné loajalita, p -hodnota této proměnné navíc říká, že výrazně nepřispívá ke zlepšení modelu.

Podmodel vzniklý nezahrnutím proměnné	Deviance	p -hodnota
Příjem	37491	<0.0001
Věk	36657	<0.0001
Doba zaměstnání	36626	0.2050
Region	36943	<0.0001
Obory²	40973	<0.0001

Tabulka 3.29: Podmodely, jejich deviance a p -hodnota testovací statistiky

Vzhledem k výsledkům z Tabulky 3.29, budeme nyní uvažovat model bez proměnné doba zaměstnání

$$\eta_i = \beta_0 + \beta_{příjem} \cdot příjem_i + \beta_{věk} \cdot věk_i + \beta_{region} \cdot region_i + \sum_{o=1}^9 \beta_{oboro} \cdot oboro_i, \quad (3.8)$$

Deviance tohoto modelu je tedy 36626. Pro podmodely tohoto modelu jsou opět určeny deviance a p -hodnota testové statistiky.

Podmodel vzniklý nezahrnutím proměnné	Deviance	p -hodnota
Příjem	37494	<0.0001
Věk	36670	<0.0001
Region	36948	<0.0001
Obory	40993	<0.0001

Tabulka 3.30: Podmodely po vynechání proměnné loajálnost, deviance a p -hodnota testovací statistiky

Na základě Tabulky 3.30 můžeme dojít k závěru, že vynechání i dalších proměnných by významně zhoršilo kvalitu modelu. Odhady koeficientů, jejich p -hodnotu a hodnoty Odds Ratio prezentuje následující tabulka.

² Pokud by byla deviance určena pro každý obor zvlášť, bylo by na základě testovacího kritéria rozhodnuto, že vypuštění oboru logistika a odbyt, jehož p -hodnota činí 0.6020, nezhorší kvalitu modelu. Podobný výsledek by byl získán i při testování rozdílu deviancí podmodelů (3.8), v tomto případě by p -hodnota podmodelu nezahrnující zmíněný obor byla 0.6073. Po provedení testu podmodelu bez této proměnné a jeho podmodelů nebyla již žádná proměnná vyhodnocena jako nadbytečná.

Kladná hodnota odhadu regresního koeficientu znamená, že příslušná proměnná přispívá ke zvýšení pravděpodobnosti, že osoba je žena. Příjem a věk dle výsledků Tabulky 3.31 snižují pravděpodobnost ženy, s rostoucí hodnotou tedy pravděpodobnost $Y_i = \text{žena}$ klesá. Negativní vliv mají také obory výroba a těžba, logistika a odbyt. Ostatní působí opačně. V případě některých proměnných může být výsledek ovlivněn datovým vzorkem.

Koeficient	Význam	Odhadovaná hodnota	Odds Ratio	<i>p</i> -hodnota
β_0	Absolutní člen	$-4.8664 \cdot 10^{-1}$		$1.5074 \cdot 10^{-10}$
$\beta_{\text{příjem}}$	Příjem	$-2.1272 \cdot 10^{-6}$	0.99999	$5.5680 \cdot 10^{-146}$
$\beta_{\text{věk}}$	Věk	$-9.0712 \cdot 10^{-3}$	0.99097	$5.1038 \cdot 10^{-11}$
β_{region}	V hlavním městě	$4.6767 \cdot 10^{-1}$	1.5963	$9.5267 \cdot 10^{-72}$
β_{obor_1}	Management	$2.8486 \cdot 10^{-1}$	1.3296	$3.6730 \cdot 10^{-4}$
β_{obor_2}	Finance a bankovníctví	$9.6669 \cdot 10^{-1}$	2.6298	$1.0765 \cdot 10^{-60}$
β_{obor_3}	Zdravotnictví	$3.9445 \cdot 10^{-1}$	1.4836	$1.1717 \cdot 10^{-4}$
β_{obor_4}	Výroba a těžba	-1.7248	0.17820	$9.3222 \cdot 10^{-104}$
β_{obor_5}	Vzdělávání a výzkum	$8.4947 \cdot 10^{-1}$	2.3384	$1.7352 \cdot 10^{-29}$
β_{obor_6}	Zákaznický servis	1.4328	4.1904	$1.4492 \cdot 10^{-80}$
β_{obor_7}	Logistika a odbyt	$-3.2532 \cdot 10^{-2}$	0.96799	$6.0674 \cdot 10^{-1}$
β_{obor_8}	Administrativa	1.1262	3.0840	$3.3539 \cdot 10^{-69}$

Tabulka 3.31: Odhady regresních koeficientů, odds ratio a *p*-hodnota *t*-testu

Další nástroje, které jsou určeny pro výběr modelu zejména při stanovení počtu a výběru vysvětlujících proměnných, jsou tzv. informační kritéria. S množstvím proměnných přidávaných do regresního modelu se sice může zvýšit přesnost odhadu, roste ale nebezpečí nadhodnocení modelu. Tato kritéria penalizují narůstající počet vysvětlujících proměnných, jejich hodnoty zpravidla závisí na počtu pozorování N , počtu regresorů k a maximálně věrohodném odhadu modelu $\ln(L)$. Při výběru hledáme minimum informačního kritéria.

Jedním z těchto kritérií je Akaikeho informační kritérium (*AIC*), které je dáno vztahem

$$AIC(k) = 2k - 2\ln(L). \quad (3.9)$$

Příbuzným kritériem *AIC* je Bayesovské informační kritérium (*BIC*), které opět může sloužit k odhadu počtu parametrů v modelu. Míra penalizace vzniklá přidáním parametrem je zde však ještě vyšší, než tomu bylo u *AIC*. Kritérium je dáno vzorcem

$$BIC(k) = k \cdot \ln(N) - 2\ln(L). \quad (3.10)$$

Vzhledem k vysokým hodnotám $\ln(L)$ a nedostatečné penalizaci počtu použitých regresorů, nebyly tyto indexy použity. Navíc se od sebe výsledky *AIC* a *BIC* příliš nelišily, rozdíl v rámci desítek byl vzhledem k vysokým hodnotám zanedbatelný.

3.7 Přehled nerovností

V předchozích kapitolách byly kvantifikovány nerovnosti mezi skupinami, které byly určeny na základě různého dělení populace. Výsledky shrnuje následující tabulka. Vzhledem k velkému množství variant oborů, je pro toto dělení uvedena jen nejvyšší nerovnost, která se objevila.

Nerovnost v oblasti (skupina s vyššími příjmy, skupina s nižšími příjmy)	<i>PGI</i>	<i>PRI</i>
Region (mimo hlavní město, v hlavním městě)	4.57 %	5.54 %
Věk (starší 37, 37 a mladší)	35.08 %	44.64 %
Loajalita (více než 8 let, 8 a méně let)	26.80 %	35.00 %
Pohlaví (muži, ženy)	23.45 %	27.53 %
Největší nerovnost mezi obory (manažeři, zákaznický servis)	55.57 %	71.39 %

Tabulka 3.32: Porovnání nerovností

Z tabulky je patrné, že největší nerovnost byla změřena při porovnávání příjmů v profesních oborech. Podrobnější výsledky přinesla Tabulka 3.16 a Tabulka 3.17. Nízkých hodnot dosahují indexy v případě porovnávání regionů. Překvapivě nízkou příjmovou nerovnost vykazuje dělení na základě pohlaví. Pro věk a loajalitu byly zkoumány i různé hranice určující věkové skupiny a dělicí populaci na loajální a méně loajální zaměstnance. Hodnoty *PGI* věkových skupin se v závislosti na volbě dělení skupin pohybovaly v rozmezí od 31.51 % do 36.61 %, *PRI* od 35.87 % do 51.88 %. Při měření nerovností dle loajálnosti se pak *PGI* rovnal od 24.32 % do 26.57 % a *PRI* od 31.03 % do 36.22 %. Po porovnání s Tabulkou 3.32 lze konstatovat, že pro všechny uvažované možnosti vykazovaly oba indexy vyšších hodnot, než jakých dosahovaly při měření nerovnosti mužů a žen. Nerovnost mezi věkovými skupinami byla vyšší než nerovnost, která byla změřena u skupin loajálních a méně loajálních zaměstnanců.

Závěr

Cílem práce bylo analyzovat příjmovou strukturu obyvatelstva, zejména pak popis a studium nerovností, za tímto účelem byl po složitém vyjednávání získán datový soubor obsahující příjmy zaměstnanců soukromých firem v Argentině.

Kromě známých postupů a indexů měřících nerovnost byly nově definovány nové míry nerovnosti, které určují nerovnost mezi skupinami. Příjmová mezera, která porovnává percentily dvou skupin, byla rozšířena, aby nepracovala pouze s mediánem, popřípadě s průměrem, ale aby bylo možné ji použít pro jakýkoliv percentil. Z vypočtených hodnot příjmové mezery od nultého po stý percentil byl odvozen index příjmové mezery *PGI*, který určuje, jak moc se rozdělení příjmů ve skupinách liší od případu, kdy jsou percentily obou skupin shodné. Ukazatel, který říká, kolik procent populace skupiny, ve které lze očekávat nižší příjmy, má příjem menší nebo rovný určitému percentilu skupiny s vyššími příjmy, byl pojmenován Percentage Ratio. Z jeho grafického znázornění bylo zřejmé podobné chování jako u Lorenzovy křivky, ze které je odvozen Giniho index, proto byl i v tomto případě zaveden index, který jsme označili jako index procentuálního poměru *PRI* a který kvantifikuje, jak moc se data liší od absolutně rovnoměrného rozdělení příjmů mezi skupinami.

Nejdříve byla zpracována nerovnost v zemi jako celku, vypočítali jsme Giniho index a Theilův index a porovnali je s výsledky World Bank. Dále byly osoby rozděleny do skupin dle regionu, věku, délky zaměstnání v jedné firmě, profese a pohlaví. Skupiny byly navzájem porovnávány a nerovnosti mezi skupinami vyčísleny. Ze zkoumaných oborů byli nejlépe oceňovanými manažeři, mezi hůře oceňované profese lze zařadit zákaznický servis nebo administrativu. Nerovnosti mezi obory byly velmi variabilní, rozdělení příjmů některých oborů byly velmi blízké, indexy nerovnosti dosahovaly hodnot blízké 0, naopak některé se velmi lišily. Největších nerovností dosahovaly příjmy manažerů a zaměstnanců zákaznického servisu, jedná se o nejvyšší výsledky indexů v rámci všech zkoumaných skupin.

Poměrně vysoké nerovnosti naznačují výsledky skupin dle věku a loajality. Byla zkoumána různá hranice předělu mezi skupinami starších a mladších, loajálních a neloajálních. Pro věkové skupiny byly měřeny nerovnosti pro hranice od 30 do 50 let, přičemž platilo, že s vyšší hodnotou předělu míra nerovnosti klesala. Pro loajalitu byla volena hranice od 4 do 10 let. U obou se největší nerovnosti projeví v případě, kdy byly porovnávány osoby, které se nepohybovaly na předělu mezi skupinami, tedy pokud byla určována nerovnost osob starších 50 let a mladších 30 let, zaměstnaných méně než 4 roky a více než 10 let.

Vzhledem k velkému množství regionů v Argentině, bylo pouze určeno, zda zaměstnanec pochází z hlavního města a jeho aglomerace, druhou skupinou byli obyvatelé mimo něj. Způsob tohoto dělení byl motivován vyššími výsledky Giniho a Theilova indexu a vysokými hodnotami HDP. Mimo hlavní město a jeho aglomeraci jsou zaměstnanci odměňováni nižšími příjmy než mimo tuto oblast. *PGI* nabývá nízkých hodnot, nerovnosti oblastí, na kterých se rozkládá hlavní město, a ostatních se výrazněji neprojevily, což ale nepotvrdil Kolmogorovův-Smirnovův dvouvýběrový test, který nulovou hypotézu, že výběry mají stejnou distribuční funkci, zamítl.

Překvapivě nízké nerovnosti dosahovaly příjmy mužů a žen. Byly porovnány příjmy mužů v hlavním městě a mimo něj a také příjmy žen v hlavním městě a mimo něj. Ženy dosahovaly vyšších příjmů v oblasti Gran

Buenos Aires (hlavní město včetně aglomerace), naopak muži měli v této oblasti příjmy nižší než mimo ni. Nerovnost mezi pohlavími je obecně v hlavním městě nižší, ukazuje se ale větší rozdíl mezi muži a ženami s vyššími příjmy, muži pracující na vysokých pozicích jsou ve městě lépe placeni než ženy.

Logistickou regresí byla modelována pravděpodobnost, že zkoumaná osoba je žena, na základě dostupných informací. Model nejprve zahrnoval jen proměnnou příjem, s jeho rostoucí hodnotou pravděpodobnost, že se jedná o ženu, klesá. Byly zahrnuty i další dostupné proměnné, po testování podmodelů se ukázala nadbytečnost zahrnutí počtu let u jednoho zaměstnavatele. Výsledný model tedy obsahoval proměnné příjem, věk, region a obory. Příjem a věk a obory logistika a odbyt, výroba a těžba snižují pravděpodobnost ženy, ostatní ji ovlivňují pozitivně.

Lze tedy shrnout, že největší nerovnost byla změřena při porovnávání příjmů v profesních oborech. Nízkých hodnot dosahují indexy v případě porovnávání regionů. Nerovnosti věkových skupin byly vyšší než u loajálních a neloajálních zaměstnanců. Pro věk a loajalitu byly zkoumány i různé hranice určující starší věkovou skupinu nebo loajální zaměstnance, pro všechny uvažované možnosti vykazovaly oba indexy vyšší hodnoty, než jakých dosahovaly při měření nerovnosti mužů a žen.

Použitá literatura

- [1] Oishi S., Kesebir S., Diener E. (2011), Income Inequality and Happiness, *Psychological Science* 22(9) 1095–1100. Dostupné z http://pss.sagepub.com/content/22/9/1095.abstract?ikey=fef169f923bb594826cb7600c5efd5f484058d2c&keytype2=tf_ipsecsha
- [2] Počet obyvatel dle regionů, El Instituto Nacional de Estadística y Censos de la República Argentina INDEC [online]. 2016 [cit. 2016-01-10]. Dostupné z: http://www.indec.gov.ar/nivel4_default.asp?id_tema_1=2&id_tema_2=41&id_tema_3=135
- [3] List of Argentine provinces by gross domestic product. Wikipedia, The Free Encyclopedia. [online]. 2016 [cit. 2016-01-10]. Dostupné z: https://en.wikipedia.org/w/index.php?title=List_of_Argentine_provinces_by_gross_domestic_product&oldid=702188243
- [4] Allison, P. D.. (1978). Measures of Inequality. *American Sociological Review*, 43(6), 865–880. Dostupné z <http://www.jstor.org/stable/2094626>
- [5] Gastwirth, J. L.. (1972). The Estimation of the Lorenz Curve and Gini Index. *The Review of Economics and Statistics*, 54(3), 306–316. Dostupné z <http://doi.org/10.2307/1937992>
- [6] Statistická příručka. *Stats Direct* [online]. 2016 [cit. 2016-02-15]. Dostupné z: http://www.statsdirect.com/help/default.htm#nonparametric_methods/gini.htm
- [7] Zvára, K. Biomedicínská statistika IV: Základy statistiky v prostředí R. Karolinum, 2013
- [8] Yitzhaki, S. (1983). On an Extension of the Gini Inequality Index. *International Economic Review*, 24(3), 617–628. Dostupné z <http://doi.org/10.2307/2648789>
- [9] Newell, A., Reilly, B. (2001) The gender pay gap in the transition from communism: some empirical evidence. *Economic systems*, Volume 25, Issue 4, 287–304. Dostupné z <http://www.sciencedirect.com/science/article/pii/S0939362501000280>
- [10] Mzdové statistiky. Český statistický úřad [online]. 2016 [cit. 2016-02-29]. Dostupné z: <https://www.czso.cz/documents/10180/20568983/311110a04.pdf/aff2944d-1279-4f36-bde7-5501184bea29?version=1.0>
- [11] Výsledky Giniho indexu. World Bank [online]. 2016 [cit. 2016-03-20]. Dostupné z: <http://data.worldbank.org/indicator/SI.POV.GINI/>
- [12] Inequality Stagnation in Latin America in the Aftermath of the Global Financial Crisis. World Bank [online]. 2016 [cit. 2016-03-20]. Dostupné z: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2014/12/29/000158349_20141229220501/Rendered/PDF/WPS7146.pdf
- [13] Argentina, Doing business in. Reference for Business [online]. 2016 [cit. 2016-04-29]. <http://www.referenceforbusiness.com/encyclopedia/A-Ar/Argentina-Doing-Business-in.html>
- [14] Procházka, B. Biostatistika pro lékaře: Principy základních metod a jejich interpretace s využitím statistického systému R, Karlova univerzita v Praze, 2015
- [15] Zvára, K. Regrese. Matfyzpress, 2008
- [16] Budíková, M., Lerch, T., Mikoláš, Š. Základní statistické metody. Masarykova univerzita, 2005.
- [17] Reif, J. Metody matematické statistiky. Západočeská univerzita v Plzni, 2004
- [18] Farris, A. F. (2010). The Gini Index and Measures of Inequality. *The American Mathematical Monthly*, 117(10), 851–864. Dostupné z <http://doi.org/10.4169/000298910x523344>

- [19] Corak M. (2013), Income Inequality, Equality of Opportunity, and Intergenerational Mobility, *The Journal of Economic Perspectives*, Vol. 27, No. 3, pp. 79-102 Dostupné z <http://www.jstor.org/stable/41955546>
- [20] Durlauf, S. N. and Quah, D. T. (1999). The new empirics of economic growth. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*: Volume 1A, pages 235–308. Elsevier, Amsterdam. Dostupné z <http://www.sciencedirect.com/science/article/pii/S1574004899010071>
- [21] Freund R.J., Wilson W.J. (2002), *Statistical Methods*, Associated Press.